



**HAL**  
open science

# Algorithmique et combinatoire des mots par les représentations S-adiques

Pierre Béaur

► **To cite this version:**

Pierre Béaur. Algorithmique et combinatoire des mots par les représentations S-adiques. Mathématique discrète [cs.DM]. Université Paris-Saclay, 2024. Français. NNT : 2024UPASG033 . tel-04661982

**HAL Id: tel-04661982**

**<https://theses.hal.science/tel-04661982v1>**

Submitted on 25 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Algorithmique et combinatoire des mots par les représentations S-adiques

*Algorithms and combinatorics on words  
using S-adic representations*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n°580 :  
Sciences et technologies de l'information et de la communication (STIC)  
Spécialité de doctorat : Informatique mathématique  
Graduate School : Informatique et sciences du numérique  
Réfèrent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Laboratoire interdisciplinaire des sciences  
du numérique (Université Paris-Saclay, CNRS)**, sous la direction de  
**Nathalie AUBRUN**, Chargée de recherche et le coencadrement de  
**Benjamin HELLOUIN de MENIBUS**, Maître de conférences

**Thèse soutenue à Paris-Saclay, le 5 juillet 2024, par**

**Pierre BÉAUR**

## Composition du jury

Membres du jury avec voix délibérative

<b>Valérie BERTHÉ</b> Directrice de recherche, CNRS, Université Paris-Cité (IRIF)	Présidente
<b>Sylvain LOMBARDY</b> Professeur, Institut Polytechnique de Bordeaux (LABRI)	Rapporteur & Examinateur
<b>Gwenaél RICHOMME</b> Professeur, Université Paul-Valéry Montpellier 3 (LIRMM)	Rapporteur & Examinateur
<b>Marie-Pierre BÉAL</b> Professeure, Université Paris-Est Marne-la-Vallée (IGM)	Examinatrice
<b>Matthieu ROSENFELD</b> Maître de conférences, Université Paul-Valéry Montpellier 3 (LIRMM)	Examinateur
<b>Manon STIPULANTI</b> Chercheuse qualifiée, Université de Liège	Examinatrice

**Titre :** Algorithmique et combinatoire des mots par les représentations S-adiques

**Mots-clés :** combinatoire des mots, calculabilité, substitutions, dynamique symbolique.

**Résumé :** En combinatoire des mots, une méthode classique de construction de mots infinis est le modèle substitutif. Il consiste à itérer infiniment une transformation (une substitution) sur une lettre initiale. Le modèle substitutif a permis de créer et d'étudier des mots infinis possédant des structures répétitives fortes mais non périodiques. Introduites à la fin des années 1990, les représentations S-adiques forment une extension classique du modèle substitutif. Dans le modèle S-adique, plutôt qu'itérer une seule et même substitution, il est possible de choisir une substitution à chaque itération dans un ensemble fini. Les représentations S-adiques ont originellement été établies à des fins dynamiques, et caractérisent plusieurs familles classiques de mots comme les mots sturmiens qui n'étaient pas complètement capturées par le modèle substitutif. Cette thèse s'intéresse à l'utilisation des représentations S-adiques à des fins combinatoires et algorithmiques.

Dans une première partie, je propose une application des représentations S-adiques dans le cadre de la théorie des  $\omega$ -automates. L'objectif est de décider si un  $\omega$ -automate faible accepte un mot sturmien. Je développe une méthode, la désubstitution d'automates, qui permet de résoudre cette question, et de donner des propriétés combinatoires des  $\omega$ -automates acceptant un mot sturmien. Ces méthodes peuvent être généralisées à d'autres constructions substitutives (mot purement substitutif, point fixe d'une substitution) et aux autres familles de mots admettant une caractérisation S-adique. Il est possible d'utiliser ces méthodes pour résoudre différents problèmes annexes, comme le codage d'un mot sturmien, ou, en

géométrie discrète, le recollement de segments discrets.

La deuxième partie est consacrée à l'étude des pièges à facteurs sur les mots bi-infinis. Elle résulte d'un travail collaboratif avec Hellouin et Gheeraert. Les pièges à facteurs viennent de la théorie de la compression, et sont un objet combinatoire permettant de mesurer la répétitivité d'un mot. Dans le cas mono-infini, seuls les mots ultimement périodiques admettent des pièges à facteurs finis. Nous prouvons que dans le cas bi-infini, ce résultat n'est plus valable : nous exhibons et caractérisons les mots bi-infinis apériodiques admettant des pièges à facteurs finis. Il s'agit des mots sturmiens caractéristiques, de leurs décalages finis, et de leurs images par des substitutions apériodiques. Nos méthodes reposent sur la caractérisation S-adique des mots sturmiens, et consiste principalement en une adaptation de la désubstitution aux pièges à facteurs.

Dans la troisième et dernière partie, j'explore les possibilités combinatoires des représentations S-adiques dans de nouveaux espaces. Je prouve que deux modèles exotiques de représentations S-adiques, les modèles d'Aubrun-Sablik et de Baraviera-Leplaideur, respectivement sur  $\mathbb{N}^d$  et sur  $\mathbb{F}_2^+$ , ne peuvent pas représenter toute configuration : ils ne sont pas universels. Enfin, j'étudie une variante du problème du domino, appelée le problème du  $X$ -domino, paramétré par un sous-shift ou une famille de mots  $X$ . Le but est d'appréhender la frontière d'indécidabilité entre une et deux dimensions. Je m'intéresse au cas où  $X$  est un sous-shift minimal, puis j'explore le cas des mots sturmiens et des mots sans carré.

**Title :** Algorithms and combinatorics on words using S-adic representations

**Keywords :** combinatorics on words, calculability, substitutions, symbolic dynamics.

**Abstract :** In combinatorics on words, a classical method to construct infinite words is the substitutive model. It consists in iterating infinitely an operation (a substitution) on an initial letter. The substitutive model has made it possible to create and study infinite words exhibiting some repetitive structures that are still aperiodic. Introduced in the late 1990s, S-adic representations are a classical extension of the substitutive model. In the S-adic model, at each iteration, rather than always iterate a single substitution, one can iterate a substitution chosen in a finite set. S-adic representations were originally established for dynamic purposes, and characterize several classical word families such as Sturmian words, that were not fully captured by the substitutive model. This thesis focuses on the use of S-adic representations for combinatorial and algorithmic purposes.

In a first part, I propose an application of S-adic representations in the framework of  $\omega$ -automata theory. The objective is to decide whether a weak  $\omega$ -automaton accepts a Sturmian word. I develop a method, called automata desubstitution, that solves this question, and gives combinatorial properties of  $\omega$ -automata accepting a Sturmian word. These methods can be generalized to other substitutive constructions (purely substitutive word, fixed point of a substitution) and to other word families admitting an S-adic characterization. These methods can be used to solve a number of related problems, such as the encoding of a

Sturmian word or, in discrete geometry, the reconnection of discrete segments.

The second part is devoted to the study of string attractors on bi-infinite words. It is the result of a collaboration with Hellouin and Gheeraert. String attractors come from text compression theory, and are a combinatorial object for measuring the repetitiveness of a word. In the mono-infinite case, only ultimately periodic words admit finite string attractors. We prove that in the bi-infinite case, this result no longer holds : we exhibit and characterize bi-infinite aperiodic words admitting finite string attractors. These are the characteristic Sturmian words, their finite shifts, and their images by aperiodic substitutions. Our methods are based on the S-adic characterization of Sturmian words, and consist mainly in an adaptation of desubstitution to string attractors.

In the third and final part, I explore the combinatorial possibilities of S-adic representations in new spaces. I prove that two exotic models of S-adic representations, the Aubrun-Sablik and Baraviera-Leplaideur models, on  $\mathbb{N}^d$  and the free monoid on two elements respectively, cannot represent every configuration : they are not universal. Finally, I study a variant of the domino problem, called the  $X$ -domino problem, parameterized by a subshift or a family of words  $X$ . The aim is to understand the undecidability boundary between one and two dimensions. I focus on the case where  $X$  is a minimal subshift, and then explore the case of Sturmian words and squarefree words.

# Remerciements

Dans un souci de respect de l'ordre protocolaire, je commence par remercier Marie-Pierre Béal, Matthieu Rosenfeld et Manon Stipulanti d'avoir accepté d'être jury pour ma soutenance de thèse; Valérie Berthé d'avoir accepté de présider ledit jury; Sylvain Lombardy et Gwenaël Richomme d'avoir accepté de rapporter cette thèse. Je remercie tout particulièrement Marie-Pierre, Matthieu et Gwenaël pour les échanges scientifiques durant ce doctorat, qui m'auront permis d'éclairer des zones scientifiques qui m'étaient abscones; Manon pour les échanges lors des événements scientifiques; Sylvain pour avoir accepté de relire un manuscrit assez éloigné de son domaine; et Valérie pour m'avoir encadré lors de mon stage de master, pour son soutien scientifique et moral qui fut critique avant le début de cette thèse, et pour m'avoir donné le goût de la combinatoire des mots et des mots Sturmien.

Je remercie aussi ma direction de thèse, Nathalie Aubrun et Benjamin Hellouin de Menibus, pour leur soutien scientifique, personnel et académique lors de cette thèse. Votre accueil après mon débarquement inopiné et votre soutien indéfectible m'ont permis de profiter de trois ans de science, de conférences et de relative insouciance. Vous m'avez laissé la liberté du sujet, la liberté de la forme et la liberté des projets, et je vous en suis très reconnaissant.

Je remercie ensuite les différentes communautés scientifiques qui m'ont accueilli le temps de cette thèse.

D'abord, je pense à mon laboratoire, le LISN, et tout particulièrement à l'équipe GALaC. Même si je milite depuis 2 ans pour son changement de nom <sup>1</sup>, cette équipe m'a apporté un soutien émotionnel important qui consistait principalement à me laisser râler, mais aussi à me rassurer lors de mes crises de doute, à rire à la meilleure blague apériodique, à se détendre au CESFO ou au billard et à manger mes biscuits trop cuits. Je m'attelle à une périlleuse liste exhaustive de ses membres, dont certains m'auront d'autant plus marqué : les doctorant·e·s et postdoctorant·e·s du présent et du passé de l'équipe, Hugo, Daniel, Noémie, Balthazar, Tianjao, Marc, Hugo, Philippe, Loïc, Valentin, Léo, Quentin, Manon, Pacôme, Clément et Nicolás, un réseau de soutien et d'amitié qui me fut précieux aux différentes étapes de ces trois années; les différents stagiaires avec lesquels j'ai interagi, Dhia, Rémi, Hoan, Thomas, Victorin, Tilmann, Hugo et Raphaël, pour m'avoir rappelé mon âge avancé; et les différents membres (plus ou moins) permanents de l'équipe, François, Hoang, Viviane, Joanna, Francesca, Johanne, Antoine, Florent, Chiara, Nathalie et Benjamin, pour les discussions endiablées sur le monde académique, le Scrabble et l'huile d'olive. Je remercie par ailleurs les différents enseignants avec lesquels j'ai travaillé durant cette thèse, Jean-Claude, Yacine, Jacques, Maxime et Nicolas, pour l'opportunité qu'ils m'ont offerte d'enseigner et d'apprendre des compétences très variées; et Olga et Sonia, pour leur travail administratif précieux qui m'aura permis de découvrir de nombreux recoins de la France (et de la Suède).

Ensuite, je pense aux communautés francophone et internationale de dynamique symbolique et de combinatoire des mots, que j'ai pu rencontrer aux cours des conférences, des journées nationales et des séminaires. Je les remercie de m'avoir accepté avec mes couleurs flashy et de m'avoir intégré dans ce réseau d'humains exceptionnels. Je pense tout particulièrement à l'équipe des Liégeois, France, Manon, Pierre, Antoine et Savinien, pour leur énergie, leur humour décapant et leur tolérance de mon accent de frouze; aux (plus ou moins) Caennais, Antonin, Léo, Victor et Pascal, pour leur soutien, leurs blagues et les whiskys; et aux autres, Jana, Bastien, Gabriele, Giuseppe, Anahi, Pierre G., Guillaume, Firas, Vivien, Raphaël, Emmanuel, Étienne, Carole, Daria, Matthieu, Julien L., Émilie, Julien C., Solène, Samuel N., Pierre P., Pierre-Adrien, Shuo, Francesco, Ahmed, Valérie, Herman, Pierre A., Sebastián,

---

1. Pour un judicieux AAAAA : Algorithmes, Automates, combinAtoire, grAphes et Autres

Bastián, Guilhem, Léo G., Vincent et Svetlana, pour leurs sourires, leurs histoires et leurs amitiés parfois d'une semaine.

D'un point de vue plus large, je remercie aussi ma famille pour son soutien tout le long de cette (trop longue) scolarité : Maman, Papa, Julie et Othman <sup>2</sup>. Vous m'avez toujours soutenu, dans les années heureuses ou non, m'avez encouragé à aller plus loin, à poursuivre ces études comme un rêve insensé, à être qui je voulais; vous m'avez donné la sagesse, l'humour, le soutien, l'espoir, la raison et l'énergie d'aller au bout. Promis j'appellerai plus souvent, maintenant que je n'ai plus de chapitre à rédiger.

Je tiens enfin à remercier mes ami·e·s, que j'ai rencontré au fil du temps et des années, qui auront apporté leur pierre dans l'édifice de cette scolarité.

Je remercie le groupe Maths, Jeux, Vélibs et autres Joyeusetés pour les bons moments passés en M103, les plans sur la comète, les nuits sous les étoiles, les colles improvisées <sup>3</sup>, les voyages impromptus.

Je remercie les différent·e·s ami·e·s de l'ENS, Louis, Manon, Corto, Yoan, Valentin, Rémi, Alice, Camille, Guillaume, Xavier, Elyes, Vincent, Yehudi, Rémy, pour les accès de rire, les soirées potaches, les apports culturels variés, les éclats de joie. Je remercie Zéphyr, pour les conseils, les badges, les DM dans la Med, les potins et les cheveux; Antonin pour les films, les râleries, le  $\LaTeX$  par Discord interposé, les opinions et les discussions à cœur ouvert; Corentin pour les blagues, les jeux, les câlins, les soirées et les encouragements; Bastien pour la franchise, l'humour sardonique, la camaraderie, la sagesse et la gentillesse.

Je remercie Nicolás, mon cobureau devant l'éternel, avec qui j'ai traversé ces trois années de science, d'angoisses, de papiers, de joies et de dramas au coin des mails. En arrivant en 2021, nous étions les deux mousquetaires de l'ambiance de l'équipe, toujours fidèles au poste : nous avons parcouru toute la France, piétiné les langues française, anglaise et espagnole (enfin, surtout moi), tu m'as appris que les groupes, finalement, c'était pas si pire que ça, je t'ai appris tout ce que je pouvais de français (n'oublie pas : on te posera toutes ces questions pendant ta soutenance). Tu as été un soutien inestimable, toujours là pour m'écouter dans mes doutes et dans mes délires, pour m'aider à trouver des papiers obscurs de théorie des automates, pour me donner l'énergie de continuer à rédiger ces chapitres. Soy muy orgulloso de tener un amigo tan coolo como tú.

Je remercie enfin Benjamin, qui fut un pilier important de cette thèse. Nos chemins se sont croisés plus par une forme d'accident routier que par la légèreté du hasard, et pourtant je ne regretterais pour rien au monde de t'avoir connu. Que ce soit dans l'enseignement ou dans la recherche, tu as su me soutenir, me rassurer et parfois me réaligner pour aller dans la bonne direction; et, dans ta simple démarche, tu m'as appris que la discrétion n'était pas la timidité, et que le pacifisme n'est pas l'abandon. J'espère pouvoir te considérer comme un ami pour les années à venir, quand tu auras des mails de retard et des anecdotes à partager aux plus jeunes.

---

2. Et Mousse, Pacha et Eira.

3. Sachant que 7 est un nombre premier, déterminer le degré du  $n$ -ième polynôme cyclotomique.

# Table des matières

<b>0</b>	<b>Introduction</b>	<b>1</b>
0.1	Contexte scientifique . . . . .	1
0.2	Objectifs et organisation de la thèse . . . . .	3
<b>1</b>	<b>Préliminaires</b>	<b>5</b>
1.1	Définitions . . . . .	5
1.1.1	Combinatoire des mots élémentaire . . . . .	5
1.1.2	Topologie prodiscrète . . . . .	7
1.1.3	Substitutions . . . . .	8
1.1.4	Dynamique symbolique . . . . .	10
1.2	Représentations S-adiques et mots sturmiens . . . . .	11
1.2.1	Représentations S-adiques : cas général . . . . .	11
1.2.2	Mots sturmiens . . . . .	13
<b>2</b>	<b>Représentations S-adiques des mots acceptés par un <math>\omega</math>-automate</b>	<b>21</b>
2.1	Introduction . . . . .	21
2.2	Représentations S-adiques et $\omega$ -automates faibles . . . . .	22
2.2.1	Définitions . . . . .	22
2.2.2	Dés substitution des $\omega$ -automates faibles . . . . .	24
2.2.3	Algorithmes pour une substitution . . . . .	26
2.2.4	Algorithmes pour plusieurs substitutions : représentations S-adiques . . . . .	31
2.2.5	Un phénomène de totalité . . . . .	38
2.2.6	Applications de ces algorithmes . . . . .	41
2.3	Étude du cas des $\omega$ -automates de Büchi . . . . .	50
2.3.1	Définition d'un automate de Büchi . . . . .	50
2.3.2	Dés substitution d'un $\omega$ -automate de Büchi-arêtes . . . . .	51
2.3.3	Un problème topologique . . . . .	52
2.3.4	Acceptation d'un mot sturmien morphique par un $\omega$ -automate de Büchi-arêtes . . . . .	52
2.3.5	Cas des $\omega$ -automates de Büchi déterministes co-complets . . . . .	55
2.3.6	Croyons en nos rêves . . . . .	58
2.4	Conclusion . . . . .	59
<b>3</b>	<b>Pièges à facteurs pour les mots bi-infinis</b>	<b>61</b>
3.1	Introduction . . . . .	61
3.2	Définitions . . . . .	62
3.3	Existence et caractérisation des mots bi-infinis admettant un piège à facteurs fini . . . . .	65
3.3.1	Quels sont les candidats possibles? . . . . .	65
3.3.2	Premier résultat d'existence . . . . .	68

3.3.3	Caractérisation des mots sturmiens admettant un piège à facteurs fini . . . . .	70
3.3.4	Généralisation au cas d'envergure finie quelconque . . . . .	75
3.4	Étude des pièges à facteurs infinis . . . . .	85
3.5	Généralisation aux familles de mots : les pièges à facteurs universels . . . . .	89
3.6	Conclusion . . . . .	92
<b>4</b>	<b>Combinatoire des représentations S-adiques et problème du <math>X</math>-domino</b>	<b>95</b>
4.1	Introduction générale du chapitre . . . . .	95
4.2	Problème de l'universalité des représentations S-adiques . . . . .	95
4.2.1	Introduction . . . . .	95
4.2.2	Universalité des représentations S-adiques sur $\mathbb{N}$ . . . . .	96
4.2.3	Représentations S-adiques sur $\mathbb{N}^d$ . . . . .	97
4.2.4	Représentations S-adiques sur $\mathbb{F}_2^+$ . . . . .	101
4.3	Problème du $X$ -domino . . . . .	105
4.3.1	Introduction . . . . .	105
4.3.2	Problème du $X$ -domino et problème d'appartenance au langage de $X$ . . . . .	106
4.3.3	Plusieurs versions du problème du $X$ -domino . . . . .	107
4.4	Conclusion . . . . .	109



# 0 Introduction

## 0.1 Contexte scientifique

Qu'est-ce qu'un mot? Pour le linguiste, le philosophe ou le poète, c'est une question profonde sur l'identité, la culture et l'homme, c'est un combat épique entre le réel et l'intangible, entre le sens et la forme, entre le son et l'écrit. Pour le mathématicien, un mot est une suite de lettres; les lettres sont les éléments d'un alphabet; et un alphabet est un ensemble fini. Pas de discrimination, pas de dictionnaire officiel, pas d'orthographe, ni de sens, ni de conjugaison. *Rhabarberbarbarabarbarbarenbartbarbierbierbarbärbel* est un mot; le cycle de la *Recherche du Temps Perdu* de Proust est un mot, pour peu que les signes de ponctuations et les espacements soient considérés comme des lettres; le génome d'un zèbre de Burchell est un autre mot.

Malgré la simplicité de leur définition, les mots ne sont devenus un objet d'étude à part entière qu'assez récemment dans l'histoire des mathématiques, et on place usuellement la naissance de la combinatoire des mots au travail de Thue en 1906 [Thu06]. Il s'agit du domaine des mathématiques discrètes qui s'intéresse à la variété combinatoire des mots en tant qu'objet : à quoi peuvent-ils ressembler? Quelles propriétés doivent-ils respecter par essence, malgré la généralité de leur définition? Comment ces propriétés interagissent-elles, et comment sont-elles reliées aux méthodes de construction d'un mot?

Cette étude porte, le plus souvent, sur les mots *infinis*, c'est-à-dire les suites infinies de lettres. Le modèle infini recouvre des objets d'origines variées : un premier exemple classique concerne les systèmes de numération. Un nombre réel écrit en base  $b$  est, de fait, un mot infini; une partie de la combinatoire des mots s'attelle ainsi à faire le lien entre propriétés algébriques d'un nombre et propriétés combinatoires de son développement numérique [AB07]. De manière générale, la combinatoire des mots entretient des liens profonds avec les systèmes de numérations et la théorie des nombres [AS03; Cha23; LL23].

Un autre exemple éminent concerne la discrétisation des systèmes dynamiques. Considérez un ballon de football pendant un match : décrire complètement sa trajectoire à chaque instant et avec une précision spatiale infinie est non seulement techniquement impossible, mais aussi plutôt difficile mathématiquement. Pour avoir une approximation de la trajectoire du ballon, une méthode consiste à discrétiser l'espace et le temps : on quadrille le terrain, et on note, chaque seconde, dans quel carré se situe la balle. La description approchée de la dynamique du ballon se fait donc sous la forme d'un mot, et pour peu que le match s'éternise, le mot décrivant la trajectoire du ballon sera infini. Les propriétés combinatoires du mot peuvent alors traduire les propriétés de la dynamique du ballon : par exemple, si toutes les lettres apparaissent infiniment souvent dans le mot, alors le ballon parcourra infiniment souvent tous les carrés du terrain. Plus largement, l'étude des systèmes dynamiques *discrétisés* passe par l'étude des systèmes dynamiques *discrets* : cette étude est appelée dynamique symbolique, dont la combinatoire des mots est un outil essentiel.

Que veut-on savoir sur les mots? Comme son nom l'indique, la combinatoire des mots s'attache surtout à étudier les propriétés combinatoires des mots. Par exemple, on définit un *carré* comme une répétition d'un mot : le mot *coucou* est le carré du mot *cou*. Est-il possible de construire un mot arbitrairement long sans carré? Sur un alphabet à deux lettres, c'est impossible : on se rend vite compte en essayant sur une feuille qu'on ne peut pas faire plus long que le mot *aba*. Et sur un alphabet à 3 lettres? La situation est alors bien différente : l'article fondateur de Thue a

exhibé un mot infini sur un alphabet à trois lettres ne contenant aucun carré.

$$abcacbabcbacabcacbacabcacba \dots$$

Il existe aujourd'hui une littérature riche sur les mots évitant certains types de répétition : par exemple, le théorème de Dejean-Rao établit les puissances fractionnaires évitables maximales avec un alphabet donné [Dej72; Rao11]. L'absence de carré est une propriété combinatoire du mot de Thue. Il n'existe pas vraiment de définition formelle de ce qu'est une propriété combinatoire : il s'agit le plus souvent d'une propriété discrète décrite simplement à l'aide d'objets finis. D'autres propriétés combinatoires classiques des mots concernent l'inévitabilité d'une famille de motifs [Ros98] (étant donnée une famille de motifs, existe-t-il un mot infini ne contenant aucun de ses motifs?) ou les périodes possibles d'un mot fini, avec le théorème de Fine-Wilf [FW65].

Pour étudier les propriétés combinatoires des mots infinis, il existe plusieurs outils, comme les automates ou les langages formels, et surtout une méthode : la méthode substitutive. Une substitution est une opération de transformation des mots, qui consiste à remplacer chaque lettre constituant un mot par un autre mot, souvent plus grand. Par exemple, considérons la substitution  $\phi : a \mapsto ab, b \mapsto ba$ . On peut alors l'appliquer à un mot en le découpant lettre à lettre, puis en concaténant les images :

$$\begin{aligned} \phi(abbab) &= \phi(a b b a b) \\ &= \phi(a) \phi(b) \phi(b) \phi(a) \phi(b) \\ &= ab ba ba ab ba \\ &= abbabaabba \end{aligned}$$

La méthode substitutive consiste à itérer ce processus infiniment. Sous de bonnes conditions, cette méthode donnera un mot infini ; et selon les propriétés de la règle choisie, le mot infini résultant aura des propriétés combinatoires différentes. C'est par exemple ainsi que Thue a exhibé le premier mot uniformément mais non linéairement récurrent, et, à l'aide d'une autre règle<sup>1</sup>, le premier mot infini ternaire sans carré [Thu06; Thu12]. Depuis, la méthode substitutive a été étendue au-delà de la combinatoire des mots, notamment pour construire de nombreuses structures inhabituelles en plusieurs dimensions, comme les fractales, les pavages apériodiques ou les quasi-cristaux [Pen74; Bar88]. Au sein de la combinatoire des mots, la littérature sur les substitutions est particulièrement riche. Les travaux de Thue [Thu06; Thu12], Morse, Hedlund [MH38; MH40], Schützenberger [Sch64], Coven [CH73], Christol [Chr+80], Rauzy [Rau82], Pansiot [Pan84], Berstel, Séebold [BS94], Durand [Dur98], Mauduit [Mau01], parmi tant d'autres, explorent les propriétés combinatoires des substitutions.

Une approche dynamique des substitutions est aussi particulièrement féconde : dans [Que87], Quéffélec décrit comment a été établie une théorie spectrale et analytique sur les mots infinis pour en déduire certaines propriétés des systèmes substitutifs. C'est dans ce cadre qu'a été créée la notion de représentation S-adique. Il s'agit d'une variation de la méthode substitutive : plutôt que d'appliquer infiniment la même substitution, il est possible, dans une représentation S-adique, d'itérer une substitution arbitraire à chaque étape de calcul. L'idée, due à Host<sup>2</sup>, est décrite pour la première fois dans [Fer96] dans un contexte de théorie générale des systèmes dynamiques. Les représentations S-adiques permettent de faire un pont entre théorie des systèmes dynamiques et les méthodes plus combinatoires établies dans le cadre des systèmes substitutifs. Le pont est tout de même souvent emprunté unilatéralement : les représentations S-adiques servent le plus couramment d'outil de traduction, permettant de convertir

1. Ici un automate cellulaire.

2. Ferenczi a plusieurs fois insisté dessus lors de présentations orales, sans l'attester dans ses propres articles.

des résultats combinatoires en résultats dynamiques Elles permettent par exemple d'étudier certaines notions pré-existantes comme les diagrammes de Bratteli-Vershik ou les tours de Rokhlin (voir [BD14] pour des explications plus précises). Un autre exemple de propriété dynamique étudiée grâce aux représentations  $S$ -adiques est la conjecture  $S$ -adique : existe-t-il une condition combinatoire  $C$  telle qu'un mot est minimal et linéairement récurrent si, et seulement si, il admet une représentation  $S$ -adique remplissant la condition  $C$ ? Cette conjecture, posée dans [Fer96], a été résolue en 2023 par Espinoza [Esp23b].

Bien qu'elles soient originellement motivées par des questions dynamiques, quelques résultats combinatoires ont été établis concernant les représentations  $S$ -adiques. En particulier, l'une des démarches les plus productives des dernières décennies est la caractérisation de certaines familles classiques de mots par les représentations  $S$ -adiques : les mots sturmiens [Lot02], les mots d'Arnoux-Rauzy [AR91], les échanges d'intervalles [Rau79], et plus récemment les mots dendriques minimaux [GL22]. À l'aide de ces caractérisations, on peut alors en déduire des conséquences dynamiques et spectrales sur les familles classiques sus-citées : de nouveau, malgré leur caractère combinatoire, les représentations  $S$ -adiques sont fréquemment envisagées comme un outil dont le but est l'étude dynamique.

Parallèlement, lors de l'étude d'objets combinatoires, une question importante est la décidabilité de leurs propriétés. Dispose-t-on d'algorithmes pour savoir si un objet  $O$  a la propriété  $P$ ? En informatique théorique, la décidabilité est une question fondamentale, autour de la limite de la machine vis-à-vis des mathématiques [Tur36]. En l'occurrence, l'intersection entre dynamique symbolique et théorie de la décidabilité est un domaine très prolifique. L'un des résultats les plus fondamentaux est l'indécidabilité du problème du domino dans le plan [Ber66]. Cette indécidabilité a pour conséquence l'indécidabilité d'une grande variété de questions dans le cadre multidimensionnel, comme de nombreuses propriétés des automates cellulaires [Kar92; DFV03], des variantes du problème du domino [AES20; CH22; ABH23], ou le problème du serpent [AB23]. Mais la combinatoire des mots, elle, n'étudie que des objets unidimensionnels : les mots s'écrivent sur une ligne, et pas sur tout le plan. Elle échappe alors au « marais de l'indécidabilité » [Lin04].

Beaucoup de problèmes en combinatoire des mots peuvent être montrés décidables par la théorie des automates : un solveur de vérification formelle, Walnut, a été développé dans le cadre des suites automatiques [MSS16; Hie+22]. Dans le cadre du modèle substitutif, Durand a prouvé la décidabilité de plusieurs problèmes classiques, comme l'égalité entre deux mots morphiques [Dur12], l'ultime périodicité d'un mot morphique [Dur13a] et l'uniforme récurrence d'un mot morphique [Dur13b]. Les méthodes qu'il a développées, autour des mots de retour [Dur98], sont particulièrement puissantes, et ont initié de nouvelles questions de recherche [GL22; BG23]. Il est attendu, de manière générale, que la plupart des questions envisageables en combinatoire des mots soient décidables.

## 0.2 Objectifs et organisation de la thèse

Dans cette thèse, nonobstant leurs origines dynamiques, je m'intéresse aux propriétés combinatoires et algorithmiques des représentations  $S$ -adiques.

Les questions qui m'intéressent dans ce manuscrit sont les suivantes : comment exploiter, en tant qu'outil, les représentations  $S$ -adiques pour en déduire des résultats combinatoires? Peuvent-elles repousser et préciser les limites de ce qui est décidable en combinatoire des mots? Il s'agira de prendre un point de vue algorithmique et informatique sur les représentations  $S$ -adiques. L'objectif ne sera pas d'étudier les représentations  $S$ -adiques *in fine*, mais de les utiliser. Ma démarche est en cela distincte de celle développée dans [BFS16], où les représentations  $S$ -adiques sont l'objectif de l'étude. L'intérêt combinatoire des représentations  $S$ -adiques est qu'elles permettent de définir non seulement un mot, mais toute une famille de mots : si un algorithme réussit à exploiter les représentations  $S$ -adiques, il pourra alors sûrement résoudre des questions à l'échelle d'une classe de mots. Et pour des familles aussi largement étudiées que les mots sturmiens ou les mots d'Arnoux-Rauzy, disposer d'outils algorithmiques sera bénéfique pour

les futures recherches. Une autre question est l'utilisation combinatoire des représentations S-adiques, c'est-à-dire avec une démarche plus théorique. Un mot disposant d'une représentation S-adique est forcé d'avoir une certaine structure : comment exploiter cette structure pour déterminer des propriétés du mot? Une autre question qui sera partiellement soulevée dans cette thèse est la suivante : quelles sont les limites du modèle S-adique en tant que modèle combinatoire? En particulier, est-il possible d'étendre le modèle en dehors du cas unidimensionnel?

Parmi les familles caractérisées par les représentations S-adiques, la famille des mots sturmiens est d'autant plus remarquable qu'elle tient une figure centrale dans la littérature de la combinatoire des mots. C'est la raison pour laquelle, dans ce manuscrit, je m'attarderai d'autant plus sur le cas des mots sturmiens. Il faut les voir comme une famille-exemple particulièrement importante : mais certains résultats, comme plusieurs résultats du chapitre 2, peuvent être étendus à d'autres familles.

Le reste de ce manuscrit est organisé comme suit. Dans le premier chapitre, j'introduis les définitions formelles nécessaires à la compréhension du reste de cette thèse, principalement autour de la combinatoire des mots et de la dynamique symbolique. Je redémontre certains résultats classiques comme la représentation S-adique des mots sturmiens, et je prouve certaines autres propriétés sur les représentations S-adiques des mots sturmiens.

Dans le deuxième chapitre, je m'intéresse au problème de la détection d'une représentation S-adique dans un  $\omega$ -automate. Les  $\omega$ -automates faibles, notamment, peuvent être utilisés pour modéliser une famille classique de structures dynamiques, les sous-shifts sofiqes. Un objectif de ce chapitre est de relier algorithmiquement deux objets aux origines orthogonales : étant donné un sous-shift sofique, contient-il un mot sturmien? Je développe, à partir de techniques connues en théorie des langages formels et en employant le vocabulaire des  $\omega$ -automates, la méthode de la désubstitution qui, combinée avec l'idée des représentations S-adiques, permet de prouver que la question précédente est décidable. Ce problème d'apparence très théorique trouve en fait plusieurs applications en combinatoire des mots ou en géométrie discrète, comme le codage des mots sturmiens ou le recollement de segments discrets. J'évoque aussi l'extension du problème au cas des  $\omega$ -automates de Büchi.

Dans le troisième chapitre, j'étudie la notion introduite relativement récemment de pièges à facteurs. Plus précisément, je m'intéresse aux pièges à facteurs sur les mots infinis et les mots bi-infinis. J'établis alors des caractérisations combinatoires des mots bi-infinis admettant des pièges à facteurs d'envergure 1, puis d'envergure finie de manière plus générale. Les démonstrations reposent sur la désubstitution et les représentations S-adiques des mots sturmiens.

Dans le quatrième et dernier chapitre, je présente des résultats variés autour des représentations S-adiques et de la décidabilité en combinatoire des mots avec une approche plus exploratoire. D'abord, j'aborde le cas des représentations S-adiques dans le cadre multidimensionnel, et présente le problème de l'universalité. En particulier, j'explore deux modèles : celui d'Aubrun-Sablik sur  $\mathbb{Z}^d$ , et celui de Baraviera-Leplaideur sur le monoïde libre à deux éléments  $\mathbb{F}_2^+$ . Puis j'introduis une nouvelle question de décidabilité en combinatoire des mots, le problème du  $X$ -domino, dans l'objectif de mieux comprendre le fossé de décidabilité entre les dimensions 1 et 2.

# 1 Préliminaires

Avec des mots, on peut tout prouver ou tout réfuter.

---

Anton Tchekhov, *Les Feux*

## 1.1 Définitions

### 1.1.1 Combinatoire des mots élémentaire

Dans cette thèse,  $\mathcal{A}$  désignera un alphabet fini, c'est-à-dire un ensemble fini de lettres ou de symboles. Par exemple,  $\mathcal{A} = \{0, 1\}$  désigne l'alphabet binaire sur les lettres 0 et 1, et  $\mathcal{A} = \{a, b, c, \dots, y, z\}$  est l'alphabet latin usuel en français sur 26 lettres.

**Définition 1.1.1** (Mot fini, longueur et occurrences des lettres). *L'ensemble des mots finis sur  $\mathcal{A}$  est  $\mathcal{A}^* = \bigcup_{i=0}^{\infty} \mathcal{A}^i$ .*

*Pour  $m \in \mathcal{A}^*$  un mot fini, la longueur de  $m$  est l'entier  $|m| \in \mathbb{N}$  tel que  $m \in \mathcal{A}^{|m|}$ . Le seul mot de longueur 0, noté  $\varepsilon$ , est le mot vide, et l'ensemble des mots non vides est  $\mathcal{A}^+$ . Pour désigner les lettres de  $m$ , on écrit  $m(i) = m_i$  pour  $0 \leq i \leq |m| - 1$ , et on écrit généralement  $m = m_0 m_1 m_2 \dots m_{|m|-1}$ . Pour  $a \in \mathcal{A}$ , le nombre d'occurrences de la lettre  $a$  dans  $m$  est noté  $|m|_a$ .*

Par exemple, le mot *dixie* est de longueur 5 :  $|dixie| = 5$ . De plus, il contient deux occurrences de la lettre  $i$  : donc  $|dixie|_i = 2$ .

**Définition 1.1.2** (Mot infini, mono-infini, bi-infini). *L'ensemble des mots infinis (ou mots mono-infinis) sur  $\mathcal{A}$  est  $\mathcal{A}^{\mathbb{N}}$ , et pour  $x \in \mathcal{A}^{\mathbb{N}}$ , on écrit  $x = x_0 x_1 x_2 x_3 \dots$ . L'ensemble des mots bi-infinis sur  $\mathcal{A}$  est  $\mathcal{A}^{\mathbb{Z}}$ , et pour  $x \in \mathcal{A}^{\mathbb{Z}}$ , on écrit  $x = \dots x_{-3} x_{-2} x_{-1} \cdot x_0 x_1 x_2 x_3 \dots$ .*

Une opération essentielle sur les mots est la concaténation : il s'agit de coller un mot à un autre. Par exemple, la concaténation de *ka* et *va* est le mot *kava* : pour deux mots  $x$  et  $y$ , je noterai simplement leur concaténation  $xy$ . La concaténation peut s'étendre partiellement aux mots infinis, par exemple en concaténant un mot fini  $m$  et un mot mono-infini  $x$  pour former  $m_0 m_1 \dots m_{|m|-1} x_0 x_1 x_2 \dots$ .

Les définitions suivantes seront souvent données exclusivement dans le cas mono-infini. Sauf quelques exceptions, des définitions analogues existent dans le cas bi-infini.

Pour  $0 \leq i \leq j < \infty$  ( $-\infty < i$  si le mot est bi-infini),  $x_{[i,j]}$  désigne le mot fini  $x_i x_{i+1} x_{i+2} \dots x_{j-1} x_j$ .

**Définition 1.1.3** (Facteur et préfixe). *Soit  $x$  un mot (fini ou infini) et  $m$  un mot fini. On dit que  $m$  est un facteur de  $x$  s'il existe  $i \in \mathbb{N}$  tel que  $x_{[i, i+|m|-1]} = m$ , et on note cette relation  $m \preceq_f x$  (où  $f$  signifie facteur). On dit que  $m$  est un préfixe de  $x$  si  $x_{[0, |m|-1]} = m$ , et on note cette relation  $m \sqsubseteq_p x$  (où  $p$  signifie préfixe).*

Par exemple, le mot *joli* est un facteur du mot *enjoliveur*; tandis que le mot *lust* est un préfixe du mot *lustucru*. Je noterai alors  $joli \preceq_f enjoliveur$  et  $lust \sqsubseteq_p lustucru$ . Si  $m$  est un facteur de  $x$ , une occurrence de  $m$  est un entier  $i \in \mathbb{N}$  tel que  $x_{[i, i+|m|-1]} = m$  : pour  $joli \preceq_f enjoliveur$ , il y a une seule occurrence de *joli*, qui est la position  $i = 2$ .

**Définition 1.1.4** (Langage et complexité d'un mot infini). Soit  $x$  un mot infini. Pour tout  $n \geq 0$ , on note  $\mathcal{L}_n(x)$  l'ensemble de ses facteurs de longueur  $n$  :  $\mathcal{L}_n(x) = \{m \in \mathcal{A}^n \mid m \preceq_f x\}$ .

On définit alors comme suit la complexité d'un mot :

$$p_x : \begin{pmatrix} \mathbb{N} & \rightarrow & \mathbb{N} \\ n & \mapsto & \text{Card}(\mathcal{L}_n(x)) \end{pmatrix}$$

Par exemple, soit le mot infini  $x = abbbbbbb \dots$ . Alors  $x$  admet deux facteurs de longueur 1,  $a$  et  $b$  : donc  $p_x(1) = 2$ . Puis, pour  $n \geq 2$ ,  $x$  admet deux facteurs de longueur  $n$  :  $ab^{n-1}$  et  $b^n$  :  $p_x(n) = 2$ . Finalement, pour tout  $n \geq 1$ , on a  $p_x(n) = 2$ .

**Définition 1.1.5** (Mot périodique, ultimement périodique, apériodique). Soit  $x$  un mot mono-infini. On dit que  $x$  est :

- périodique s'il existe  $p \geq 1$  tel que pour tout  $n \geq 0$ ,  $x_n = x_{n+p}$  ;
  - $p$  est une période de  $x$  ;
  - la plus petite période de  $x$  est appelé la période de  $x$ .
- ultimement périodique s'il existe  $m \in \mathcal{A}^*$  et  $y \in \mathcal{A}^{\mathbb{N}}$  périodique tels que  $x = my$  ;
  - $m$  est appelé une pré-période de  $x$  ;
  - toute période de  $y$  est une période de  $x$  ;
  - lorsque  $m$  est de taille minimale, c'est la pré-période de  $x$  ;
  - la période de  $x$  est définie comme la période de  $y$ .
- apériodique si  $x$  n'est pas ultimement périodique.

Par exemple, le mot infini  $x = abababababab \dots$ , tel que  $x_n = a$  si  $n$  est pair, et  $x_n = b$  si  $n$  est impair, est périodique de période 2. Un exemple de mot ultimement périodique est le mot  $y = ccccababababababab \dots$ . Ce mot est ultimement périodique de période 2 et de pré-période  $cccc$ .

L'ultime périodicité étant plus délicate à définir dans le cas bi-infini, je précise ici comment elle est étendue.

**Définition 1.1.6** (Mot bi-infini ultimement périodique, à gauche, à droite). Soit  $x \in \mathcal{A}^{\mathbb{Z}}$  un mot bi-infini. On dit que  $x$  est :

- ultimement périodique à gauche, resp. à droite, si  $x_{\llbracket -\infty, 0 \rrbracket}$ , vu comme un mot mono-infini, est ultimement périodique (resp. si  $x_{\llbracket 0, \infty \rrbracket}$  est ultimement périodique) ;
- ultimement périodique si  $x$  est ultimement périodique à gauche et à droite ;
- apériodique si  $x$  n'est ni ultimement périodique à gauche, ni à droite.

En particulier, dans le cas bi-infini, un mot peut n'être ni ultimement périodique, ni apériodique, comme l'exemple suivant l'illustre.

**Exemple 1.1.7.** Soit  $x = \dots aaaaaaa \cdot baaaaa \dots$ . Alors  $x$  est ultimement périodique à gauche et à droite, donc  $x$  est ultimement périodique.

Soit  $y$  un mot infini apériodique, et soit la concaténation  $z = \dots aaaaaaaa \cdot y_0y_1y_2y_3y_4 \dots$  :  $z$  est ultimement périodique à gauche mais pas à droite. Donc  $z$  n'est ni ultimement périodique, ni apériodique.

Ces définitions constituent les briques fondamentales de la combinatoire des mots. Rien qu'avec elles, on peut déjà établir l'un des résultats les plus importants du domaine : le théorème de Morse-Hedlund.

**Théorème 1.1.8** ([MH40]). *Soit  $x$  un mot mono-infini. Alors il existe  $n \geq 1$  tel que  $p_x(n) \leq n$  si, et seulement si,  $x$  est ultimement périodique.*

*Soit  $x$  un mot bi-infini. Alors il existe  $n \geq 1$  tel que  $p_x(n) \leq n$  si, et seulement si,  $x$  est périodique.*

Une autre manière de mesurer le caractère répétitif d'un mot est d'étudier ses propriétés de récurrence. Ce critère est moins quantitatif et plus qualitatif. Beaucoup des résultats présentés dans ce manuscrit ne tiendront que sous des hypothèses de récurrence.

**Définition 1.1.9** (Récurrence et récurrence uniforme). *Soit  $x$  un mot infini. On dit que  $x$  est :*

- *récurent si, pour tout facteur  $f \preceq_f x$ ,  $f$  a une infinité d'occurrences dans  $x$ . De manière équivalente,  $x$  est récurent si, pour tout facteur  $f \preceq_f x$ ,  $f$  a au moins deux occurrences dans  $x$ .*
- *uniformément récurent si, pour tout facteur  $f$ , il existe une borne  $N_f \geq 1$  tel que pour tout  $i \in \mathbb{N}$ ,  $x_{[i, i+N_f-1]}$  contient une occurrence de  $f$ . De manière équivalente,  $x$  est uniformément récurent si, pour tout  $n \geq 1$ , il existe  $N \geq 1$  tel que pour tout  $i \in \mathbb{N}$ ,  $x_{[i, i+N]}$  contient au moins une occurrence de tout élément de  $\mathcal{L}_n(x)$ .*

**Exemple 1.1.10.** *Le mot  $abbbbb \dots$  n'est pas récurent : le facteur  $a$  n'y apparaît qu'une seule fois.*

Il existe des mots récurrents, mais non uniformément récurrents. Par exemple, construisons un mot  $x_C$  sur l'alphabet  $\{a, b\}$  de la manière suivante : on commence par  $x^{(0)} = \varepsilon$  le mot vide. Pour tout  $n$ , on obtient  $x^{(n+1)}$  en concaténant à droite de  $x^{(n)}$  la liste de tous les mots de longueur  $n + 1$ . Par exemple, l'ensemble des mots de longueur 1 est  $\{a, b\}$ . Donc  $x^{(1)} = ab$ . Puis l'ensemble des mots de longueur 2 est  $\{aa, ab, ba, bb\}$  : donc  $x^{(2)} = x^{(1)}aaababbb = abaaabbabb$ . On continuant ce processus à l'infini, on obtient un mot infini  $x_C$  connu sous le nom de *mot de Champernowne* [Cha33]. Ce mot est récurent : prenez n'importe quel mot  $m$ . Alors ce mot apparaît dans  $ma, maa, maaa, maaaa, \dots$  donc il sera concaténé à la fin de  $x^{(|m|+1)}, x^{(|m|+2)}, x^{(|m|+3)}, \dots$  : il apparaîtra une infinité de fois dans  $x_C$ . Mais il n'est pas uniformément récurent : considérons le facteur  $a$ . Comme le mot de Champernowne admet tout mot fini comme facteur, il admet tout  $b^N$  comme facteur. Donc pour tout  $N \geq 1$ , il existe une position  $i \in \mathbb{N}$  telle que  $x_{C[i, i+N]}$  ne contient aucune occurrence de la lettre  $a$ .

### 1.1.2 Topologie prodiscrète

Il est possible de munir l'ensemble des mots infinis d'une topologie. De nouveau, les définitions seront faites dans le cas des mots mono-infinis : il existe des définitions analogues dans le cas bi-infini.

**Définition 1.1.11** (Cylindre). *Soit  $m \in \mathcal{A}^*$  et  $i \geq 0$ . Le cylindre associé à  $m$  en position  $i$  est  $[m]_i = \{x \in \mathcal{A}^{\mathbb{N}} \mid x_{[i, i+|m|-1]} = m\}$ .*

La topologie prodiscrète sur  $\mathcal{A}^{\mathbb{N}}$  est alors la topologie définie par la base des ouverts  $\{[m]_i \mid m \in \mathcal{A}^*, i \geq 0\}$ . Cette topologie est équivalente à la topologie métrique définie par la distance de Cantor  $d_C(x, y) = \exp(-\inf\{n \geq 0 \mid x_n \neq y_n\})$ . Les détails plus précis de cette topologie sont expliqués dans [LM95] et [CC10]. Il est possible d'étendre cette topologie aux mots finis : pour cela, on utilise un symbole supplémentaire  $B \notin \mathcal{A}$ , et on assimile les mots finis  $m$  aux mots mono-infinis  $mBBBBBB \dots$ .

Avec cette topologie, il m'est désormais possible d'évoquer des notions topologiques classiques comme les limites, la continuité, les fermés, les ouverts ou les adhérences sur  $\mathcal{A}^{\mathbb{N}}$  (et  $\mathcal{A}^{\mathbb{Z}}$ ). Je préciserai plusieurs objets topologiques spécifiques à la dynamique symbolique dans la sous-section 1.1.4.

### 1.1.3 Substitutions

Je peux désormais introduire les transformations les plus fondamentales de cette thèse : les morphismes et les substitutions.

**Définition 1.1.12** (Morphisme). *Un morphisme (de mots) sur  $\mathcal{A}$  est une application  $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$  qui vérifie  $\sigma(uv) = \sigma(u)\sigma(v)$  pour tous  $u, v \in \mathcal{A}^*$ . Un morphisme est uniquement déterminé par l'image des lettres de l'alphabet.*

Par exemple, le morphisme de Thue-Morse est  $\sigma_{\text{TM}} : a \mapsto ab, b \mapsto ba$  étendu à l'ensemble des mots finis, et l'image du mot  $ab$  est alors  $\sigma_{\text{TM}}(ab) = \sigma_{\text{TM}}(a)\sigma_{\text{TM}}(b) = abba$ . Le morphisme de Thue-Morse n'efface aucune lettre : c'est ce qu'on appelle une substitution.

**Définition 1.1.13** (Substitution). *Un morphisme  $\sigma$  sur l'alphabet  $\mathcal{A}$  est effaçant s'il existe une lettre  $a \in \mathcal{A}$  telle que  $\sigma(a) = \varepsilon$ , où  $\varepsilon$  désigne le mot vide. Une substitution  $\sigma$  sur  $\mathcal{A}$  est un morphisme non-effaçant, c.-à-d. pour tout  $a \in \mathcal{A}$ ,  $\sigma(a) \neq \varepsilon$ .*

Un exemple simple de substitution est un codage  $\kappa : \mathcal{A} \rightarrow \mathcal{A}$ , c'est-à-dire une substitution qui transforme une lettre en une autre.

Il est à noter que dans cette thèse, je supposerai toujours qu'un morphisme est un *endomorphisme*, c.-à-d. que l'alphabet d'entrée et de sortie sont toujours identiques, ce qui n'est pas toujours le cas dans la littérature.

Un morphisme de mots  $\sigma$  sur  $\mathcal{A}$  peut être étendu aux mots mono-infinis de la manière suivante. En premier lieu, on définit  $E_\sigma = \{x \in \mathcal{A}^{\mathbb{N}} \mid \exists N \geq 0, \forall n \geq N, \sigma(x_n) = \varepsilon\}$ . Puis on pose  $\sigma^{(\mathbb{N})} : \mathcal{A}^{\mathbb{N}} \setminus E_\sigma \rightarrow \mathcal{A}^{\mathbb{N}}$  avec  $\sigma^{(\mathbb{N})}(x_0x_1x_2\dots) = \sigma(x_0)\sigma(x_1)\sigma(x_2)\dots$ . En excluant les éléments de  $E_\sigma$ , j'assure que  $\sigma^{(\mathbb{N})}$  est bien à valeurs dans  $\mathcal{A}^{\mathbb{N}}$ .

De manière analogue, ce morphisme peut aussi être étendu aux mots bi-infinis en définissant  $E'_\sigma = \{x \in \mathcal{A}^{\mathbb{Z}} \mid \exists N \geq 0, (\forall n \geq N, \sigma(x_n) = \varepsilon) \text{ ou } (\forall n \geq N, \sigma(x_{-n}) = \varepsilon)\}$  et en posant  $\sigma^{(\mathbb{Z})} : \mathcal{A}^{\mathbb{Z}} \setminus E'_\sigma \rightarrow \mathcal{A}^{\mathbb{Z}}$  avec  $\sigma^{(\mathbb{Z})}(\dots x_{-2}x_{-1}.x_0x_1x_2\dots) = \dots \sigma(x_{-2})\sigma(x_{-1}) \cdot \sigma(x_0)\sigma(x_1)\sigma(x_2)\dots$ . On remarque que, si  $\sigma$  est une substitution,  $E_\sigma = E'_\sigma = \emptyset$ .

En reprenant l'exemple du morphisme de Thue-Morse  $\sigma_{\text{TM}} : a \mapsto ab, b \mapsto ba$ , on peut alors déterminer l'image du mot périodique  $abababab\dots$  par  $\sigma_{\text{TM}}^{(\mathbb{N})}$  :

$$\begin{aligned} \sigma_{\text{TM}}^{(\mathbb{N})}(abababab\dots) &= \sigma_{\text{TM}}(a)\sigma_{\text{TM}}(b)\sigma_{\text{TM}}(a)\sigma_{\text{TM}}(b)\dots \\ &= ab\ ba\ ab\ ba\ ab\ ba\dots \\ &= abbaabbaabbaabba\dots \end{aligned}$$

Désormais, sauf mention précise, je confondrai  $\sigma$ ,  $\sigma^{(\mathbb{N})}$  et  $\sigma^{(\mathbb{Z})}$ .

La propriété de non-effacement est motivée principalement par des raisons topologiques. Une substitution est en effet continue pour la topologie prodiscrète, ce qui n'est pas le cas des morphismes effaçants (qui, d'ailleurs, ne sont pas définis sur l'entiereté de  $\mathcal{A}^{\mathbb{N}}$  et  $\mathcal{A}^{\mathbb{Z}}$ ). Le cas effaçant peut donc être interprété comme une version dégénérée des substitutions. Dans le modèle substitutif, des efforts ont été faits pour étendre les résultats des substitutions aux morphismes effaçants (par exemple dans [BPR23]). J'essaierai de préciser au cas par cas quand les résultats évoqués peuvent être généralisés aux morphismes effaçants.

À partir des morphismes et des substitutions, il est possible de définir tout un éventail de mots qui constituent les objets d'étude du modèle substitutif. Les définitions sont reprises de [BR10].



**Définition 1.1.14** (Mot purement substitutif, substitutif, morphique et point fixe). Soient  $\sigma$  et  $\tau$  deux morphismes sur  $\mathcal{A}$ . Un mot infini  $x$  est :

- un mot purement substitutif généré par  $\sigma$  s'il existe  $a \in \mathcal{A}$  tel que  $\lim_{n \rightarrow \infty} \sigma^n(a)$  existe et vaut  $x$ ;
- un mot substitutif généré par  $\sigma$  s'il existe un mot purement substitutif  $y$  généré par  $\sigma$  et un codage  $\kappa : \mathcal{A} \rightarrow \mathcal{A}$  tel que  $x = \kappa(y)$ ;
- un mot morphique généré par  $\sigma$  et  $\tau$  s'il existe  $y$  un mot purement substitutif généré par  $\sigma$  tel que  $x = \tau(y)$ ;
- un point fixe de  $\sigma$  si  $x = \sigma(x)$ .

**Remarque 1.1.15.** Si  $x$  un mot est purement substitutif généré par  $\sigma$ , alors  $x$  est un mot substitutif généré par  $\sigma$ ,  $x$  est aussi un mot morphique généré par  $\sigma$  et  $\text{id}$  (où  $\text{id}$  est l'identité, c.-à-d.  $\text{id} : a \mapsto a$  pour tout  $a \in \mathcal{A}$ ), et  $x$  est aussi un point fixe pour  $\sigma$ .

Pour un morphisme  $\sigma$  et une lettre  $a$ , si la limite  $\lim_{n \rightarrow \infty} \sigma^n(a)$  existe, je la noterai souvent  $\sigma^\infty(a)$ .

**Exemple 1.1.16.** Soit  $\mathcal{A} = \{a, b, c\}$  et deux morphismes  $\sigma : a \mapsto abc, b \mapsto bc, c \mapsto \varepsilon$  et  $\tau : a \mapsto aa, b \mapsto bb, c \mapsto cc$ . Alors :

- le mot  $x_1 = bcbcbcbcbcbcbcb \dots$  vérifie  $\sigma(x_1) = x_1$ , donc  $x_1$  est un point fixe de  $\sigma$ ;
- le mot  $x_2 = abcbcbcbcbcbcbcb \dots$  vérifie  $x_2 = \sigma^\infty(a)$ , donc  $x_2$  est un mot purement substitutif généré par  $\sigma$ ;
- le mot  $x_3 = ababababab \dots$  vérifie  $x_3 = \kappa(x_2)$  avec  $\kappa : a \mapsto a, b \mapsto b, c \mapsto a$ , donc  $x_3$  est un mot substitutif généré par  $\sigma$ ;
- le mot  $x_4 = aabbcbcbcbcbcbcb \dots$  vérifie  $x_4 = \tau(x_2)$ , donc  $x_4$  est un mot morphique généré par  $\sigma$  et  $\tau$ .

Les définitions ci-dessus ont été données dans le cas des morphismes, incluant donc le cas des morphismes effaçants. Le résultat suivant montre cependant qu'on peut obtenir tout mot morphique en utilisant uniquement des substitutions.

**Théorème 1.1.17** ([Cob68; CLR16]). Soit  $x$  un mot morphique. Alors il existe une substitution  $\sigma$  (c.-à-d. un morphisme non-effaçant), un codage  $\kappa$  et une lettre  $a \in \mathcal{A}$  tels que  $\sigma^\infty(a)$  existe et  $x = \kappa(\sigma^\infty(a))$ . De plus, étant donné les morphismes générant  $x$ , il est possible de construire  $\sigma$  et  $\kappa$  de manière effective.

Dans le cas d'un mot purement substitutif, purement morphique ou morphique<sup>1</sup>, cela signifie qu'on peut toujours se ramener au cas où il est généré par des substitutions.

Étant donnée une substitution  $\sigma$ , l'existence d'un mot purement substitutif généré par  $\sigma$  admet une caractérisation combinatoire simple.

**Théorème 1.1.18.** Une substitution  $\sigma$  génère un mot purement substitutif si, et seulement si, il existe  $a \in \mathcal{A}$  telle que  $\sigma(a) = au$  avec  $u \in \mathcal{A}^+$ .

Il sera enfin parfois pertinent d'utiliser l'identité suivante sur les mots purement substitutifs.

**Théorème 1.1.19.** Soit  $\sigma$  une substitution telle que  $\sigma(a) = au$  avec  $a \in \mathcal{A}$  et  $u \in \mathcal{A}^+$ .

$$\sigma^\infty(a) = au\sigma(u)\sigma^2(u)\sigma^3(u) \dots$$

1. Dans le cas d'un point fixe, la substitution identité convient toujours.

### 1.1.4 Dynamique symbolique

Pour compléter les définitions nécessaires à la compréhension de ce manuscrit, je définis ici une famille de structures topologiques qui seront évoquées à plusieurs reprises : les sous-shifts.

**Définition 1.1.20** (Action de décalage). *L'action de décalage sur  $\mathbb{N}$  est l'opération suivante :*

$$S : \left( \begin{array}{ccc} \mathcal{A}^{\mathbb{N}} & \rightarrow & \mathcal{A}^{\mathbb{N}} \\ x = x_0x_1x_2x_3\dots & \mapsto & S(x) = x_1x_2x_3x_4\dots \end{array} \right)$$

*Autrement écrit,  $S(x)_n = x_{n+1}$  pour  $n \in \mathbb{N}$ .*

Cette opération induit une dynamique sur  $\mathcal{A}^{\mathbb{N}}$  au sens de la théorie des systèmes dynamiques (voir [Wal82] pour une introduction à la théorie des systèmes dynamiques). Dans ce contexte, il est naturel de s'intéresser aux sous-systèmes dynamiques de  $\mathcal{A}^{\mathbb{N}}$ . Dans le cadre de la dynamique symbolique, ces sous-systèmes sont appelés des sous-shifts :

**Définition 1.1.21** (Sous-shift). *Un sous-shift de  $\mathcal{A}^{\mathbb{N}}$  est un sous-ensemble  $X \subseteq \mathcal{A}^{\mathbb{N}}$  vérifiant les deux propriétés suivantes :*

1.  *$X$  est fermé pour la topologie prodiscrète;*
2.  *$X$  est invariant par décalage, c.-à-d. pour  $x \in X$ ,  $S(x) \in X$ .*

On observe alors que  $\mathcal{A}^{\mathbb{N}}$  est lui-même un sous-shift : on l'appelle le shift plein<sup>2</sup>. Les sous-shifts bénéficient d'une caractérisation combinatoire plus accessible. Le lecteur intéressé pourra lire [LM95] pour une œuvre de référence sur la question.

**Théorème 1.1.22.** *Un sous-ensemble  $X \subseteq \mathcal{A}^{\mathbb{N}}$  est un sous-shift si, et seulement si, il existe  $\mathcal{F} \subseteq \mathcal{A}^*$  tel que*

$$X = \{x \in \mathcal{A}^{\mathbb{N}} \mid \forall f \in \mathcal{F}, f \not\subseteq_f x\}.$$

*Dans ce cas, on dit que  $\mathcal{F}$  est un ensemble de motifs interdits définissant  $X$ .*

**Exemple 1.1.23.** *Soit  $X_{\text{gm}} = \{x \in \{0, 1\}^{\mathbb{N}} \mid 11 \not\subseteq_f x\}$  le sous-shift de la moyenne d'or<sup>3</sup>. C'est un sous-shift défini par la famille  $\mathcal{F} = \{11\}$ .*

*Cependant,  $\mathcal{F}$  ne doit pas forcément être fini. Par exemple, le sous-shift de l'œuf au plat<sup>4</sup> est l'ensemble suivant :  $SSU = \{x \in \{0, 1\}^{\mathbb{N}} \mid \text{il y a au plus une occurrence de la lettre 1 dans } x\}$ . Il est possible de montrer que  $SSU$  est fermé pour la topologie prodiscrète et invariant par décalage, donc c'est bien un sous-shift. En l'occurrence,  $SSU$  est défini par l'ensemble de motifs interdits  $\mathcal{F}_{SSU} = \{10^i1 \mid i \geq 0\}$ .*

Il existe une classification des sous-shifts selon les ensembles de motifs interdits les définissant.

**Définition 1.1.24** (Sous-shift de type fini, sofique et effectif). *Soit  $X$  un sous-shift. On dit que  $X$  est :*

- de type fini s'il existe un ensemble fini de motifs interdits définissant  $X$ ;
- sofique s'il existe un ensemble régulier de motifs interdits définissant  $X$ ;
- effectif s'il existe un ensemble calculable de motifs interdits définissant  $X$ .

2. par opposition à  $\emptyset$ , qui est aussi un sous-shift, et est donc le shift vide.

3. en anglais *golden mean subshift*.

4. en anglais *sunny side up subshift*.

**Remarque 1.1.25.** Si  $X$  est un sous-shift de type fini,  $X$  est un sous-shift sofique; si  $X$  est un sous-shift sofique,  $X$  est un sous-shift effectif.

**Exemple 1.1.26.** Le sous-shift de la moyenne d'or  $X_{\text{gm}}$  (voir exemple 1.1.23) est un sous-shift de type fini. En revanche, on peut montrer que le sous-shift de l'œuf au plat n'est pas un sous-shift de type fini : c'est un sous-shift sofique.

Enfin, l'une des constructions dynamiques utiles en combinatoire des mots est celle de la clôture de l'orbite d'un mot. Il s'agit, en quelque sorte, du sous-shift « naturel » associé à un mot infini.

**Définition 1.1.27** (Orbite et clôture de l'orbite d'un mot). Soit  $x$  un mot infini. Son orbite est  $\mathcal{O}(x) = \{S^k(x) \mid k \in \mathbb{N}\}$ . La clôture de son orbite est  $\overline{\mathcal{O}(x)}$ , c.-à-d. l'adhérence de  $\mathcal{O}(x)$  pour la topologie prodiscrète.

Par construction,  $\overline{\mathcal{O}(x)}$  est fermé pour la topologie prodiscrète et invariant par décalage, il s'agit donc bien d'un sous-shift. Il est possible de relier propriétés combinatoires d'un mot et propriétés dynamiques de la clôture de son orbite : un sous-shift  $X \subseteq \mathcal{A}^{\mathbb{N}}$  est dit minimal si le seul sous-shift non vide  $Y$  inclus dans  $X$  est  $X$  lui-même.

**Théorème 1.1.28** ([Bir12]). Un mot  $x$  est uniformément récurrent si, et seulement si,  $\overline{\mathcal{O}(x)}$  est minimal.

## 1.2 Représentations S-adiques et mots sturmiens

### 1.2.1 Représentations S-adiques : cas général

Le modèle substitutif étant introduit, je vais maintenant définir le modèle S-adique. Il s'agit du modèle principal étudié dans cette thèse, introduit formellement par Ferenczi [Fer96]. L'idée générale est que, lors de la construction d'un mot purement substitutif, on itère une seule et même substitution; dans le modèle S-adique, on peut, à chaque étape de calcul, appliquer une substitution différente choisie dans un ensemble fini de substitutions. Dans la mise en œuvre pratique, il y a toutefois une difficulté importante à souligner pour le lecteur non averti : les substitutions sont appliquées « depuis l'intérieur ». Je commence par donner la définition formelle avant d'expliquer plusieurs propriétés des représentations S-adiques.

Dans ce manuscrit,  $\mathcal{S}$  désignera un ensemble fini de substitutions sur l'alphabet  $\mathcal{A}$ . Par extension de la notation sur les mots, pour une suite de morphismes  $(\sigma_n)_{n \in \mathbb{N}}$ , la notation  $\sigma_{\llbracket i, j \rrbracket}$  désigne le morphisme  $\sigma_i \circ \sigma_{i+1} \circ \cdots \circ \sigma_{j-1} \circ \sigma_j$ .

**Définition 1.2.1** (Représentation S-adique). Soit  $x$  un mot infini,  $(\sigma_n)_{n \in \mathbb{N}}$  une suite de substitutions de  $\mathcal{S}$  et  $(a_n)_{n \in \mathbb{N}}$  une suite de lettres de  $\mathcal{A}$ . On dit que  $x$  admet  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  comme représentation S-adique si la suite  $(\sigma_{\llbracket 0, n \rrbracket}(a_n))_{n \in \mathbb{N}}$  admet une limite et qu'elle vaut  $x$ , c.-à-d.

$$x = \lim_{n \rightarrow \infty} \sigma_{\llbracket 0, n \rrbracket}(a_n).$$

La suite  $(\sigma_n)_{n \in \mathbb{N}}$  est appelée suite directrice de la représentation S-adique.

Dans la littérature, il est courant de dire que  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  est une représentation S-adique du mot  $x$ , pour préciser que  $(\sigma_n)_{n \in \mathbb{N}}$  est inclus dans  $\mathcal{S}$ . Dans ce manuscrit, je me contenterai de parler de représentation S-adique lorsque  $\mathcal{S}$  est clair dans le contexte, et je préciserai explicitement qui est  $\mathcal{S}$  dans le cas contraire.

C'est souvent la suite  $(\sigma_n)_{n \in \mathbb{N}}$  plutôt que la suite  $(a_n)_{n \in \mathbb{N}}$  qui traduit les propriétés combinatoires et dynamiques d'un mot. De plus, le théorème 1.2.8 justifiera, sous certaines conditions, que si on sait que  $x$  admet  $(\sigma_n)_{n \in \mathbb{N}}$  comme suite directrice d'une représentation S-adique, alors on peut calculer assez simplement une suite  $(a_n)_{n \in \mathbb{N}}$  telle que  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  est une représentation S-adique de  $x$  : c'est pour cela que je me concentrerai sur les suites directrices dans ce manuscrit.

**Exemple 1.2.2.** Soit  $\sigma_{\text{TM}}$  le morphisme de Thue-Morse défini par  $\sigma_{\text{TM}} : a \mapsto ab, b \mapsto ba$ ; et soit  $\sigma_{\text{Fib}}$  le morphisme de Fibonacci défini par  $\sigma_{\text{Fib}} : a \mapsto ab, b \mapsto a$ . Ces deux morphismes sont bien des substitutions. Posons maintenant  $\sigma_n = \sigma_{\text{TM}}$  si  $n$  pair,  $\sigma_n = \sigma_{\text{Fib}}$  si  $n$  impair, et  $a_n = a$  pour tout  $n \geq 0$ . Soit  $x$  le mot infini admettant  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  comme représentation S-adique<sup>5</sup>.

On peut calculer  $x$  comme la limite du processus suivant :

$$\begin{aligned} \sigma_{\llbracket 0,0 \rrbracket}(a_0) &= \sigma_{\text{TM}}(a) = ab \\ \sigma_{\llbracket 0,1 \rrbracket}(a_1) &= \sigma_{\text{TM}} \circ \sigma_{\text{Fib}}(a) \\ &= \sigma_{\text{TM}}(\sigma_{\text{Fib}}(a)) = \sigma_{\text{TM}}(ab) \\ &= abba \\ \sigma_{\llbracket 0,2 \rrbracket}(a_2) &= \sigma_{\text{TM}} \circ \sigma_{\text{Fib}} \circ \sigma_{\text{TM}}(a) \\ &= \sigma_{\text{TM}}(\sigma_{\text{Fib}}(\sigma_{\text{TM}}(a))) = \sigma_{\text{TM}}(\sigma_{\text{Fib}}(ab)) = \sigma_{\text{TM}}(aba) \\ &= abbaab \\ \sigma_{\llbracket 0,3 \rrbracket}(a_3) &= \sigma_{\text{TM}} \circ \sigma_{\text{Fib}} \circ \sigma_{\text{TM}} \circ \sigma_{\text{Fib}}(a) \\ &= \sigma_{\text{TM}}(\sigma_{\text{Fib}}(\sigma_{\text{TM}}(\sigma_{\text{Fib}}(a)))) = \sigma_{\text{TM}}(\sigma_{\text{Fib}}(\sigma_{\text{TM}}(ab))) \\ &= \sigma_{\text{TM}}(\sigma_{\text{Fib}}(abba)) = \sigma_{\text{TM}}(abaaab) \\ &= abbaabababba \text{ et ainsi de suite.} \end{aligned}$$

Dans une représentation S-adique, la composition se fait « à l'intérieur ». Par exemple, à l'étape 3, on a  $\sigma_{\llbracket 0,2 \rrbracket}(a_2)$ , et à l'étape 4,  $\sigma_{\llbracket 0,3 \rrbracket}(a_3) = \sigma_{\llbracket 0,2 \rrbracket}(\sigma_3(a_3))$  : l'apport de la nouvelle étape se fait depuis l'intérieur, depuis l'argument donné aux substitutions.

**Exemple 1.2.3.** Un autre exemple à partir du morphisme de Thue-Morse : un mot purement substitutif généré par  $\sigma_{\text{TM}}$ , appelé le mot de Thue-Morse, est  $\text{TM} = \sigma_{\text{TM}}^\infty(a) = abbabaabbaababba \dots$ . Alors TM admet  $((\sigma_{\text{TM}})_{n \in \mathbb{N}}, (a)_{n \in \mathbb{N}})$  comme représentation S-adique, dont la suite directrice  $(\sigma_{\text{TM}})_{n \in \mathbb{N}}$  est la suite constante égale à  $\sigma_{\text{TM}}$ .

Ce dernier exemple illustre le fait que le modèle S-adique est bien une généralisation du modèle substitutif. La définition que je donne ici des représentations S-adiques n'est pas tout à fait la définition classique : j'en ai défini le modèle *finitaire* [BBY22]. Plus précisément, dans le modèle général, la représentation se fait potentiellement sur une infinité d'alphabets, avec  $\sigma_n : \mathcal{A}_{n+1} \rightarrow \mathcal{A}_n$  pour tout  $n \in \mathbb{N}$ , et avec une infinité de substitutions différentes. Cette généralisation est utile dans le cadre de la théorie des systèmes dynamiques, et permet par exemple de parler du rang alphabétique d'une représentation S-adique [Esp23a]. Je laisserai de côté cette version du modèle S-adique dans cette thèse.

Il existe une différence notable entre les modèles substitutif et S-adique quant aux mots constructibles. Tout mot infini n'est pas purement substitutif, ou même morphique : par exemple, la complexité d'un mot purement substitutif ou morphique est nécessairement polynomiale [Pan84]. En revanche, il existe un ensemble fini de substitutions  $\mathcal{C}$  tel que n'importe quel mot infini dispose d'une représentation S-adique dont la suite directrice est incluse dans  $\mathcal{C}$  (voir la section 4.2 pour plus de détails sur cette question). C'est une propriété propre à l'ensemble  $\mathcal{C}$  : la puissance du modèle S-adique dépend de l'ensemble de substitutions  $\mathcal{S}$  utilisé, qui paramétrise donc le modèle. Il dépend aussi de l'ensemble des suites directrices autorisées dans  $\mathcal{S}$ .

<sup>5</sup>. ce mot existe, car la représentation S-adique donnée ici est en fait celle d'un mot purement substitutif généré par la substitution  $\sigma_{\text{TM}} \circ \sigma_{\text{Fib}}$  qui vérifie bien les hypothèses du théorème 1.1.18.

Un modèle proche du modèle S-adique est celui des mots infiniment désubstituables, développé par Richomme [Ric21].

**Définition 1.2.4** (Mot infiniment désubstituable). *Soit  $x$  un mot infini et  $(\sigma_n)_{n \in \mathbb{N}}$  une suite de substitutions sur  $\mathcal{A}$ . On dit que  $x$  est infiniment désubstituable par  $(\sigma_n)_{n \in \mathbb{N}}$  s'il existe une suite de mots infinis  $(x^{(n)})_{n \in \mathbb{N}} \in (\mathcal{A}^{\mathbb{N}})^{\mathbb{N}}$  telle que :*

$$x = \sigma_{\llbracket 0, n \rrbracket}(x^{(n)}) \text{ pour tout } n \geq 0.$$

Les modèles sont proches, mais tout de même subtilement différents : en particulier, le modèle S-adique est inclus dans le modèle de la désubstitution infinie.

**Théorème 1.2.5** ([Ric21]). *Soit  $x$  un mot infini admettant comme représentation S-adique  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$ . Alors  $x$  est infiniment désubstituable par  $(\sigma_n)_{n \in \mathbb{N}}$ .*

Il peut arriver qu'un mot soit infiniment désubstituable par une suite  $(\sigma_n)_{n \in \mathbb{N}}$ , mais n'admette pas de représentation S-adique dont la suite directrice est  $(\sigma_n)_{n \in \mathbb{N}}$ .

**Exemple 1.2.6.** *Soit  $x$  un mot infini : alors  $x$  est infiniment désubstituable par  $(\text{id})_{n \in \mathbb{N}}$  la suite constante égale à l'identité. Mais pour  $n$  importe quelle suite  $(a_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ ,  $\text{id}^{n+1}(a_n) = a_n$  et donc  $x \neq \lim_{n \rightarrow \infty} \text{id}^{n+1}(a_n)$  (si tant est que la limite existe).*

L'exemple précédent montre la principale différence entre représentation S-adique et désubstitution infinie : il s'agit de la croissance des mots construits. Une manière de mesurer la croissance d'une représentation S-adique est l'ubicroissance (*everywhere growingness* en anglais).

**Définition 1.2.7** (Ubicroissance). *Soit  $(\sigma_n)_{n \in \mathbb{N}}$  une suite de morphismes. On dit que  $(\sigma_n)_{n \in \mathbb{N}}$  est ubicroissante si, pour tout  $(a_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ ,  $\lim_{n \rightarrow \infty} |\sigma_{\llbracket 0, n \rrbracket}(a_n)| = \infty$ .*

Avec cette définition, on peut alors obtenir une réciproque partielle du théorème 1.2.5.

**Théorème 1.2.8** ([Ric21]). *Soit  $x$  un mot infiniment désubstituable par  $(\sigma_n)_{n \in \mathbb{N}}$  ubicroissante. Alors il existe  $(a_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$  tel que  $x$  admet  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  comme représentation S-adique.*

En particulier, si  $x = \sigma_{\llbracket 0, n \rrbracket}(x^{(n)})$  pour tout  $n$ , alors  $((\sigma_n)_{n \in \mathbb{N}}, (x_0^{(n)})_{n \in \mathbb{N}})$  est une représentation S-adique de  $x$  : on peut calculer facilement une représentation S-adique à partir de la seule suite directrice si elle est ubicroissante.

Dans la plupart des cas étudiés ici, les suites de substitutions considérées seront ubicroissantes. C'est la raison pour laquelle, dans le contexte de ce manuscrit, représentation S-adique et désubstituabilité infinie auront souvent la même puissance : j'utiliserai le nom plus connu de représentation S-adique.

## 1.2.2 Mots sturmiens

Pour parler plus en détail des propriétés des représentations S-adiques, je vais les développer dans le cas d'un exemple bien connu en combinatoire des mots, celui des mots sturmiens.

Les mots sturmiens, introduits par Morse et Hedlund [MH40], ont été nommés en référence aux travaux de Sturm sur l'oscillation des solutions de certaines équations différentielles [Stu36]. Ils forment certainement la famille de mots la plus étudiée dans la littérature de la combinatoire des mots (par exemple [CH73; Sto76; Bru81; BF84; LP92] explorent des aspects très différents des mots sturmiens; des survols plus complets sur les mots sturmiens sont disponibles dans [Ber96; Lot02; Fog02]).

**Définition 1.2.9** (Mot sturmien sur  $\mathbb{N}$  et sur  $\mathbb{Z}$ ). Soit  $\text{Card}(\mathcal{A}) = 2$ . Un mot infini  $x \in \mathcal{A}^{\mathbb{N}}$  est sturmien si  $p_x(n) = n + 1$  pour tout  $n \geq 1$ . Un mot bi-infini  $x \in \mathcal{A}^{\mathbb{Z}}$  est sturmien si  $x$  est a périodique et  $p_x(n) = n + 1$  pour tout  $n \geq 1$ .

Les mots sturmiens bénéficient de très nombreuses caractérisations. Par exemple, d'après le théorème de Morse-Hedlund (voir le théorème 1.1.8), les mots sturmiens sont exactement les mots a périodiques de complexité minimale. Ils peuvent aussi être interprétés comme des mots mécaniques [Sto76] :

**Théorème 1.2.10.** Un mot  $x \in \{0, 1\}^{\mathbb{N}}$  est sturmien si, et seulement si, il existe  $\alpha \in \mathbb{R} \setminus \mathbb{Q}$  et  $\beta \in \mathbb{R}$  tels que  $x_n = \lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor$  pour tout  $n \geq 0$  ou  $x_n = \lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil$  pour tout  $n \geq 0$ .

Il existe aussi des interprétations à partir des rotations du cercle, des échanges d'intervalles ou des droites discrètes (voir [Lot02] pour plus de détails).

### Propriétés combinatoires des mots sturmiens

Dans ce manuscrit, j'utiliserai à plusieurs reprises les propriétés et définitions suivantes sur les mots sturmiens.

**Proposition 1.2.11** ([Pyt+02]). Les mots sturmiens sont uniformément récurrents.

**Définition 1.2.12** (Type d'un mot sturmien). Soit  $x \in \{0, 1\}^{\mathbb{N}}$  un mot sturmien. On dit que  $x$  est de type 0 si  $00 \preceq_f x$ ; on dit que  $x$  est de type 1 si  $11 \preceq_f x$ .

**Proposition 1.2.13.** Les mots sturmiens sont partitionnés par type : pour tout mot sturmien, il existe un unique  $i \in \{0, 1\}$  tel que  $x$  est de type  $i$ .

*Démonstration.*

Soit  $x$  un mot sturmien, alors  $p_x(2) = 3$ . Par lemme des tiroirs, l'un de 00 ou de 11 est un facteur de  $x$ . Si 00 et 11 étaient tous deux facteurs de  $x$ , alors l'un de 01 ou 10 ne serait pas facteur de  $x$  : supposons que les facteurs de longueur 2 de  $x$  soient 00, 11 et 01, l'autre cas se traitant de façon symétrique. Alors  $x$  serait de la forme  $0^i 1^{\mathbb{N}}$  et  $x$  serait ultimement périodique, ce qui est absurde.

Les mots sturmiens peuvent aussi être caractérisés comme l'ensemble des mots a périodiques équilibrés (voir [Pyt+02]) :

**Définition 1.2.14** (Équilibre). Un mot infini  $x \in \{0, 1\}^{\mathbb{N}}$  est équilibré si, pour tous facteurs  $f$  et  $g$  de  $x$  de même longueur,  $\|f\|_0 - \|g\|_0 \leq 1$ . Autrement écrit, le nombre de 0 dans  $f$  et dans  $g$  ne diffère que d'un.

**Lemme 1.2.15.** Un mot infini  $x \in \{0, 1\}^{\mathbb{N}}$  n'est pas équilibré si, et seulement si, il existe un mot  $m$  tel que  $0m0$  et  $1m1$  sont des facteurs de  $x$ .

*Démonstration.*

Le sens réciproque est évident.

Si  $x$  n'est pas équilibré, il existe  $f$  et  $g$  deux facteurs de  $x$  de même longueur tels que  $\|f\|_0 - \|g\|_0 > 1$  : pour simplifier les notations, notons cette inégalité (I). Supposons que  $f$  et  $g$  sont des exemples de facteurs de  $x$  de taille minimale vérifiant l'inégalité (I).

D'abord, si  $f_0 = g_0$ , alors le couple  $(f_{\llbracket 1, |f| - 1 \rrbracket}, g_{\llbracket 1, |g| - 1 \rrbracket})$  vérifierait aussi l'inégalité (I), et serait plus petit que  $(f, g)$ , ce qui est absurde. Donc  $f_0 \neq g_0$  : supposons  $f_0 = 0$  et  $g_0 = 1$ .

Si  $f_i = g_i$  pour tout  $1 \leq i \leq |f| - 1 = |g| - 1$ , alors  $f$  et  $g$  ne vérifieraient pas l'inégalité (I). Soit

alors  $p$  le plus petit entier tel que pour tout  $1 \leq i < p$ ,  $f_i = g_i$  et  $f_p \neq g_p$ . Si  $f_p = 1$  et  $g_p = 0$ , alors  $|f_{[0,p]}|_0 = |g_{[0,p]}|_0$ . Comme  $f$  et  $g$  vérifient l'inégalité (I), cela implique que  $p < |f| - 2$  et que  $(f_{[p+1,|f|-1]}, g_{[p+1,|g|-1]})$  vérifie aussi l'inégalité (I), ce qui est absurde par minimalité de  $f$  et  $g$ . Donc  $f_p = 0$  et  $g_p = 1$ . Donc  $(f_{[0,p]}, g_{[0,p]})$  vérifie l'inégalité (I), par minimalité  $p = |f| - 1$ , et on a alors bien  $f = 0m0$  et  $g = 1m1$  pour un certain mot  $m \in \{0, 1\}^*$ .

**Théorème 1.2.16** ([Lot02; Pyt+02]). *Un mot infini ou bi-infini  $x$  est sturmien si, et seulement si,  $x$  est aperiodique et équilibré.*

Une définition que j'utiliserai régulièrement est la notion d'écartement d'un mot sturmien. Cette notion n'est pas définie comme objet d'étude dans la littérature, bien qu'elle ne soit ni inconnue, ni fondamentalement originale : il s'agit du premier bit du codage multiplicatif d'un mot sturmien (voir [Pyt+02]).

**Proposition 1.2.17.** *Soit  $x$  un mot sturmien de type 0. Alors il existe un unique  $\lambda(x) \geq 1$  tel que  $x \in 0^*1(0^{\lambda(x)}1 + 0^{\lambda(x)+1}1)^{\mathbb{N}}$ .*

*Démonstration.*

Comme  $x$  est un mot sturmien,  $x$  est un mot uniformément récurrent : soit  $\ell$  la borne d'uniforme récurrence du facteur 1, et on suppose  $\ell$  minimal. Montrons que  $\lambda(x) = \ell - 2$ . Étudions alors  $\mathcal{L}_\ell(x)$ , dont on sait que le cardinal vaut  $\ell + 1$ . Par minimalité de  $\ell$ , il existe un facteur de  $x$  de la forme  $0^{\ell-1}1$  ou de la forme  $10^{\ell-1}$ . Par le théorème 1.2.16,  $x$  est équilibré : donc tout élément de  $\mathcal{L}_\ell(x)$  a zéro, une ou deux occurrences de la lettre 1. Comme  $\ell$  est une borne d'uniforme récurrence du facteur 1,  $x$  n'admet pas  $0^\ell$  comme facteur. Donc tout  $\mathcal{L}_\ell(x)$  élément a une ou deux occurrences de 1. Par cardinal de  $\mathcal{L}_\ell(x)$ , il doit y avoir au moins un facteur  $f$  admettant deux occurrences de 1 : si ce facteur n'était pas  $10^{\ell-2}1$ , alors il existerait  $k \leq \ell - 3$  tel que  $x$  admettrait deux facteurs de la forme  $00^k0$  (lui-même facteur de  $0^{\ell-1}1$  ou  $10^{\ell-1}$ ) et un autre de la forme  $10^k1$ , et ne serait donc pas équilibré. Donc  $\mathcal{L}_\ell(x)$  contient exactement un élément avec deux occurrences de la lettre 1. On en déduit alors la forme attendue de  $x$ .

Une autre notion importante en combinatoire des mots de manière générale est celle des facteurs spéciaux à gauche, spéciaux à droite et bispéciaux.

**Définition 1.2.18** (Facteurs spéciaux et bispéciaux). *Soit  $x$  un mot infini, et soit  $f \preceq_f x$  un facteur de  $x$ . Alors on dit que  $f$  est :*

- *spécial à gauche s'il existe  $a \neq b \in \mathcal{A}$  tels que  $af$  et  $bf$  sont des facteurs de  $x$ ;*
- *spécial à droite s'il existe  $a \neq b \in \mathcal{A}$  tels que  $fa$  et  $fb$  sont des facteurs de  $x$ ;*
- *bispécial si  $f$  est spécial à gauche et à droite.*

On peut montrer simplement qu'un mot dont la complexité s'écrit  $p_x(n) = n + k$  pour tout  $n \geq 1$  (et donc en particulier un mot sturmien) possède, pour toute longueur, exactement un facteur spécial à gauche (et exactement un facteur spécial à droite) [Cas97a]. Pour un tel mot  $x$ , je noterai  $l_n(x)$  son facteur spécial à gauche de longueur  $n$  (resp.  $r_n(x)$  son facteur spécial à droite de longueur  $n$ ).

Certains mots sturmiens sont particuliers : il s'agit des mots sturmiens caractéristiques, ou standard [Pyt+02].

**Définition 1.2.19** (Mot sturmien caractéristique). *Un mot sturmien infini  $x$  est caractéristique si  $x = \lim_{n \rightarrow \infty} l_n(x)$ .*

Le cas des mots sturmiens caractéristiques bi-infinis sera développé dans le chapitre 3.

## Représentations S-adiques des mots sturmiens

Il existe une autre caractérisation des mots sturmiens, qui passe par les représentations S-adiques. Cette caractérisation sera utilisée à de nombreuses reprises tout au long de ce manuscrit. Elle est due à Arnoux [Pyt+02], et j'en restitue une démonstration rapide par souci de complétude et de pédagogie dans ce chapitre introductif. Elle utilise quatre substitutions particulières, les substitutions sturmiennes élémentaires.

**Définition 1.2.20** (Substitutions sturmiennes élémentaires). *Soit  $\mathcal{A} = \{0, 1\}$ . Les quatre substitutions sturmiennes élémentaires sont les substitutions suivantes :*

$$L_0 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \end{cases} \quad L_1 : \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \end{cases} \quad R_0 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 10 \end{cases} \quad R_1 : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 1 \end{cases}$$

L'ensemble  $\mathcal{S}_{St} = \{L_0, L_1, R_0, R_1\}$  est l'ensemble des substitutions sturmiennes élémentaires. Les substitutions de type  $i$  sont les substitutions  $\{L_i, R_i\}$ .

J'utilise ici la notation de [Ric21] :  $L$  signifie *left*, car  $L_i$  fait de  $i$  la lettre à gauche de l'image. Symétriquement,  $R$  signifie *right*.

**Théorème 1.2.21** ([Pyt+02]). *Soit  $x \in \{0, 1\}^{\mathbb{N}}$ . Alors  $x$  est un mot sturmien si, et seulement si,  $x$  est infiniment désubstituable par une suite  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  telle que  $(\sigma_n)_{n \in \mathbb{N}}$  alterne infiniment de type : pour tout  $N \in \mathbb{N}$  et  $i \in \{0, 1\}$ , il existe  $n \geq N$  tel que  $\sigma_n$  est de type  $i$ .*

*Démonstration.*

Soit  $x$  un mot sturmien de type 0. Alors il existe un mot infini  $y$  tel que  $x = R_0(y)$  : en effet, comme  $x$  est de type 0,  $x$  ne contient pas d'occurrence du mot 11 ; on peut donc encoder  $x$  à l'aide de 0 et 10, d'où l'existence de  $y$ . De même, si  $x$  est un mot sturmien de type 1, alors il existe un mot infini  $y$  tel que  $x = R_1(y)$ .

Par ailleurs, soit  $x$  un mot sturmien de type 0 tel que  $x_0 = 0$ . Alors il existe  $y'$  tel que  $x = L_0(y')$  : il faut cette fois faire attention lors du codage par 0 et 01. Un résultat équivalent tient pour le type 1 : si  $x$  est un mot sturmien de type 1 tel que  $x_0 = 1$ , il existe  $y'$  tel que  $x = L_1(y')$ . Les semi-algorithmes présentés dans la figure 1.3 explicitent comment déduire  $y$  et  $y'$  de  $x$ .

---

```

i ← 0
j ← 0
tant que vrai :
  si  $x_i = 1$  :
    |  $y_j \leftarrow 1; i += 2; j += 1;$ 
  sinon
    |  $y_j \leftarrow 0; i += 1; j += 1;$ 
retourner y
    
```

---

**FIGURE 1.1** – Calcul de  $y$ .

---

```

i ← 0
j ← 0
tant que vrai :
  si  $x_{i+1} = 1$  :
    |  $y'_j \leftarrow 1; i += 2; j += 1;$ 
  sinon
    |  $y'_j \leftarrow 0; i += 1; j += 1;$ 
retourner y'
    
```

---

**FIGURE 1.2** – Calcul de  $y'$ .

**FIGURE 1.3** – Semi-algorithmes de calcul des antécédents de  $x$  par  $R_0$  et  $L_0$ .

Ces constructions permettent alors de déduire le lemme suivant :

**Lemme 1.2.22.** *Soit  $x$  de type 0 et  $y$  tel que  $x = R_0(y)$ . S'il existe, soit  $y'$  tel que  $x = L_0(y')$ . Alors l'un de  $y$  ou de  $y'$  (si ce dernier existe) est sturmien.*



*Démonstration.*

Supposons d'abord que  $y'$  existe. Montrons alors que  $y'$  est sturmien.

D'abord, si  $y'$  était ultimement périodique,  $x = L_0(y')$  serait aussi ultimement périodique, ce qui est absurde. Donc  $y'$  est apériodique.

Ensuite, montrons que  $y'$  est équilibré et procédons par l'absurde. D'après le lemme 1.2.15, il existerait  $m \in \{0, 1\}^*$  tel que  $0m0$  et  $1m1$  seraient des facteurs de  $y'$ . Il existerait  $a \in \{0, 1\}$  tel que  $0m0a$  serait un facteur de  $y'$ . Donc  $L_0(0m0a) = 0L_0(m)0L_0(a)$  et  $L_0(1m1) = 01L_0(m)01$  seraient des facteurs de  $x$ . En remarquant que  $L_0(a)$  commence toujours par la lettre 0, on en déduit alors que  $0L_0(m)00$  et  $1L_0(m)01$  seraient deux facteurs de  $x$  : d'après le lemme 1.2.15,  $x$  ne serait donc pas équilibré, ce qui est absurde.

Donc si  $y'$  existe,  $y'$  est apériodique et équilibré :  $y'$  est sturmien.

Supposons maintenant que  $y'$  n'existe pas : alors  $x$  commence par un 1. Montrons alors que  $y$  est sturmien. Par des arguments similaires au cas précédent,  $y$  est apériodique.

Montrons que  $y$  est équilibré et procédons de nouveau par l'absurde. D'après le lemme 1.2.15, il existe  $m \in \{0, 1\}^*$  tel que  $0m0$  et  $1m1$  sont des facteurs de  $y$ . Comme  $x$  commence par un 1,  $y$  aussi : donc  $0m0$  n'est pas un préfixe de  $y$ . Donc il existe une lettre  $a$  tel que  $a0m0$  est un facteur de  $y$ . Alors  $R_0(a0m0) = R_0(a)0R_0(m)0$  et  $R_0(1m1) = 10R_0(m)10$  sont des facteurs de  $x$ . Mais alors, en remarquant que  $R_0(a)$  termine par un 0,  $x$  admet comme facteurs  $00R_0(m)0$  et  $10R_0(m)1$  et n'est donc pas équilibré, ce qui est absurde.

Donc si  $y'$  n'existe pas,  $y$  est apériodique et équilibré :  $y$  est sturmien.

Il existe un analogue du lemme 1.2.22 pour les mots sturmiens de type 1, avec  $L_1$  et  $R_1$  à la place de  $L_0$  et  $R_0$  respectivement.

Soit  $x$  un mot sturmien. Alors le lemme 1.2.22 et son analogue nous donnent une suite  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  telle que  $x$  est infiniment désostituable par  $(\sigma_n)_{n \in \mathbb{N}}$ . Il reste à montrer que  $(\sigma_n)_{n \in \mathbb{N}}$  alterne infiniment de type. C'est une conséquence du lemme suivant :

**Lemme 1.2.23.** *Soit  $x, y$  deux mots sturmiens tels que  $x = L_0(y)$ . Alors soit  $y$  est de type 1, soit  $\lambda(y) = \lambda(x) - 1$ .*

*Démonstration du lemme 1.2.23.*

Si  $y$  est de type 0, alors  $10^{\lambda(y)}1$  est un facteur de  $y$  : donc son image par  $L_0$  est un facteur de  $x$ . Son image étant  $010^{\lambda(y)}01$ , on en déduit que  $10^{\lambda(y)+1}1$  est un facteur de  $x$ . Puis  $10^{\lambda(y)+1}1$  est un facteur de  $y$  : donc, par le même raisonnement,  $10^{\lambda(y)+2}1$  est un facteur de  $x$ . Donc d'après la proposition 1.2.17,  $\lambda(y) + 1 = \lambda(x)$ .

On peut alors appliquer un principe de descente infinie de Fermat : si  $x$  est de type 0, alors les désstitués successifs sont de  $\lambda$  strictement décroissants. Mais pour tout mot sturmien  $y$ ,  $\lambda(y) \geq 1$  : donc il existe un rang  $n$  tel que  $x = \sigma_{\llbracket 0, n \rrbracket}(x^{(n)})$  et  $x^{(n)}$  est de type 1. Donc  $\sigma_{n+1}$  est de type 1. En itérant le raisonnement, on en déduit que  $(\sigma_n)_{n \in \mathbb{N}}$  alterne infiniment de type.

Réciproquement, soit  $x$  un mot infini infiniment désstituable par  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  alternant infiniment en type, et notons  $x = \sigma_{\llbracket 0, n \rrbracket}(x^{(n)})$ .

Montrons que  $x$  est équilibré. On étudie le cas où  $\sigma_0 = L_0$ , les autres cas se traitent de manière analogue. Par l'absurde, supposons  $x$  non-équilibré. Alors il existe un mot  $m$  tel que  $0m0$  et  $1m1$  sont des facteurs

de  $x$ . Comme  $x = L_0(x^{(0)})$ , à gauche de toute occurrence de 1 dans  $x$  se trouve un 0 : donc  $m$  termine par un 0. Les facteurs déséquilibrés de  $x$  s'écrivent alors  $0m'00$  et  $1m'01$ . Par la forme de  $L_0$ , le fait que  $x$  admette  $1m'01$  comme facteur implique que  $m' = L_0(m^{(0)})$  pour un certain mot  $m^{(0)}$ . Donc dans  $x$  se trouvent deux facteurs  $0L_0(m^{(0)})00$  et  $1L_0(m^{(0)})01$ . La construction de  $x^{(0)}$  par le semi-algorithme implique alors que  $0m^{(0)}0$  et  $1m^{(0)}1$  sont des facteurs de  $x^{(0)}$  et que  $|m| > |m^{(0)}|$ . On peut répéter le processus et construire  $0m^{(n)}0$  et  $1m^{(n)}1$  des facteurs de  $x^{(n)}$  tels que  $|m| > |m^{(0)}| > \dots > |m^{(n)}|$  : par principe de descente infinie de Fermat, il existe  $n \geq 0$  tel que  $m^{(n)}$  est le mot vide. Alors dans ce cas,  $x^{(n)}$  aura pour facteurs  $00$  et  $11$  et ne sera désubstituable par aucune substitution sturmienne élémentaire, ce qui est une contradiction.

Montrons enfin que  $x$  est apériodique. Si  $x$  est équilibré et ultimement périodique, alors  $x$  est de la forme  $x = 0^i(10^k)^{\mathbb{N}}$  avec  $i \leq k$ . Donc  $\sigma_n$  est de type 0 pour  $0 \leq n \leq k-1$ , et on a  $x^{(k-1)} = 0^{\varepsilon_k}(10)^{\mathbb{N}}$  avec  $0 \leq \varepsilon_k \leq 1$ . Par conséquent, peu importe le choix de  $\sigma_k$ ,  $x^{(k)}$  est l'un des mots suivants :  $0^{\mathbb{N}}$ ,  $1^{\mathbb{N}}$ ,  $01^{\mathbb{N}}$  ou  $10^{\mathbb{N}}$ . En traitant les cas à la main, on montre alors que  $(\sigma_n)_{n \geq k}$  ne peut contenir que des substitutions du même type. Par exemple,  $0^{\mathbb{N}}$  n'est désubstituable que par  $L_0$  ou  $R_0$ , et la désubstitution donnera de nouveau  $0^{\mathbb{N}}$ . De même,  $10^{\mathbb{N}}$  n'est désubstituable que par  $R_0$  et donnera de nouveau  $10^{\mathbb{N}}$ . Cela contredit l'alternance infinie de type de la suite  $(\sigma_n)_{n \in \mathbb{N}}$ .

Donc  $x$  est équilibré et apériodique :  $x$  est sturmien.

Les suites directrices caractérisées par le théorème 1.2.21 sont aussi ubicroissantes, ce qui permet aussi de faire le lien avec les représentations S-adiques.

**Lemme 1.2.24.** *Soit  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  alternant infiniment en type. Alors  $(\sigma_n)_{n \in \mathbb{N}}$  est ubicroissante.*

*Démonstration.*

Je montre plus précisément par récurrence sur  $k \geq 1$  que si  $(\sigma_n)_{0 \leq n \leq N}$  contient  $k$  alternances de type, alors pour toute lettre  $a \in \{0, 1\}$ ,  $|\sigma_{\llbracket 0, N \rrbracket}(a)| \geq k + 1$  et toutes les lettres apparaissent dans  $\sigma_{\llbracket 0, N \rrbracket}$ .

Dans le cas  $k = 1$ , le tableau exhaustif suivant permet de conclure.

$$\begin{array}{cccc} L_0L_1(0) = 010 & L_0L_1(1) = 01 & L_0R_1(0) = 001 & L_0R_1(1) = 01 \\ R_0L_1(0) = 100 & R_0L_1(1) = 10 & R_0R_1(0) = 010 & R_0R_1(1) = 10 \end{array}$$

Maintenant, si  $(\sigma_n)_{0 \leq n \leq N}$  contient  $k + 1$  alternances de type, considérons  $i \geq 1$  tel que  $(\sigma_n)_{i \leq n \leq N}$  contient  $k$  alternances de type. Par hypothèse de récurrence,  $|\sigma_{\llbracket i, N \rrbracket}(a)| \geq k$  et le mot contient les deux lettres. Comme il existe une lettre  $b \in \mathcal{A}$  telle que  $\sigma_{\llbracket 0, i-1 \rrbracket}(b) > |b| = 1$ , on en déduit que

$$\begin{aligned} |\sigma_{\llbracket 0, N \rrbracket}(a)| &= |\sigma_{\llbracket 0, i-1 \rrbracket} \circ \sigma_{\llbracket i, N \rrbracket}(a)| \\ &> |\sigma_{\llbracket i, N \rrbracket}(a)| \geq k \end{aligned}$$

Donc  $|\sigma_{\llbracket 0, N \rrbracket}(a)| \geq k + 1$ .

**Corollaire 1.2.25.** *Un mot infini  $x$  est sturmien si, et seulement si,  $x$  admet une représentation S-adique  $((\sigma_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  avec  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  alternant infiniment de type.*

Les démonstrations précédentes sont similaires à des démonstrations existantes dans la littérature : je les ai réécrites par souci de complétude, et aussi pour souligner leur caractère purement combinatoire. Je n'utilise pas de propriétés dynamiques ou numériques des mots sturmiens, mais seulement des propriétés d'équilibre et d'apéro-

dicité. L'objectif de cette thèse sera dès lors de chercher comment exploiter cette construction des mots sturmiens pour en déduire des résultats combinatoires.

Il existe, en l'occurrence, d'autres manières d'obtenir des représentations des mots sturmiens. En particulier, cela peut se faire à l'aide de trois substitutions (voir [BS94]). La question de la caractérisation S-adique des mots sturmiens à l'aide de deux substitutions seulement est ouverte. Les représentations S-adiques à l'aide des substitutions de  $\mathcal{S}_{St}$  ont l'avantage d'être relativement simples à comprendre, complètement caractérisées, et encodent aussi la pente et l'ordonnée à l'origine de la droite associée à un mot sturmien [BHZ06]. C'est pour cela que je me concentrerai sur ce modèle de représentations S-adiques pour les mots sturmiens :

**Définition 1.2.26** (Représentation S-adique d'Arnoux). *Soit  $x$  un mot sturmien. Une représentation S-adique d'Arnoux de  $x$  est une suite  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  telle que  $x$  est infiniment désubstituable par  $(\sigma_n)_{n \in \mathbb{N}}$ .*

Cette définition est un abus de notation, car elle confond représentation S-adique et désubstitution infinie. Elle se justifie par l'importance de la suite  $(\sigma_n)_{n \in \mathbb{N}}$  par rapport à  $(a_n)_{n \in \mathbb{N}}$  dans la représentation S-adique d'un mot, et par la précédence historique des représentations S-adiques par rapport à la désubstitution infinie.

Les représentations S-adiques d'Arnoux bénéficient de plusieurs propriétés qui seront utilisées et détaillées le long de ce manuscrit. Par exemple, un mot sturmien donné peut avoir, a priori, plusieurs représentations S-adiques d'Arnoux, et même une infinité; mais elles vérifient toutes certaines propriétés. Par exemple, pour une suite  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}_{St}^{\mathbb{N}}$  alternant infiniment en type et commençant par le type 0, on peut définir une suite  $(k(\sigma)_n)_{n \in \mathbb{N}}$  d'entiers vérifiant :

$$k(\sigma)_0 = 0 \text{ et } \forall k(\sigma)_n \leq i \leq k(\sigma)_{n+1}, \sigma_i \text{ est de type } n \pmod{2}.$$

Alors pour un mot sturmien  $x$  de type 0, toutes ses représentations S-adiques d'Arnoux partagent le même découpage en type :  $\{(k(\sigma))_{n \in \mathbb{N}} \mid (\sigma_n)_{n \in \mathbb{N}} \text{ est une représentation S-adique d'Arnoux de } x\} = 1$ , de même symétriquement pour les mots sturmiens de type 1. Cette propriété est reliée à la valeur  $\lambda(x)$  : en effet,  $k(\sigma)_1 = \lambda(x)$ .

Certains mots sturmiens n'admettent qu'un nombre fini de représentations S-adiques d'Arnoux : par exemple, le mot de Fibonacci n'en admet qu'une seule. Comme les substitutions  $L_0$  et  $R_0$  commutent, le mot purement substitutif généré par  $L_0 R_0 L_1 R_1$  admet une quantité infinie indénombrable de représentations S-adiques d'Arnoux. La caractérisation des mots sturmiens n'admettant qu'un nombre fini de représentations S-adiques d'Arnoux reste toutefois ouverte, ainsi que l'existence d'un mot sturmien admettant exactement une quantité infinie dénombrable de représentations S-adiques d'Arnoux. L'une des difficultés autour de ces questions est que, dans chaque bloc de type, la quantité de  $L$  ou de  $R$  n'est pas assurée.



## 2 Représentations S-adiques des mots acceptés par un $\omega$ -automate

Êtes-vous soudainement un spécialiste de l'automate, Lewis? Votre mère était-elle une automate?

---

Criquette Rockwell, adapté de *Le Cœur a ses raisons*

### 2.1 Introduction

Dans ce chapitre, j'aborde les principaux résultats techniques et théoriques accomplis pendant cette thèse. Ceux-ci concernent la décidabilité de l'acceptation d'un mot Sturmien par un graphe, ou un  $\omega$ -automate faible.

Le problème de l'acceptation d'un mot par un  $\omega$ -automate est une question très classique en théorie des langages formels [HU79] : il s'agit de savoir si un  $\omega$ -automate donné accepte un mot. Les variantes du problème tournent souvent autour de restrictions sur le modèle de l' $\omega$ -automate :  $\omega$ -automates de Büchi,  $\omega$ -automates de Müller [Tho90],  $\omega$ -automates à compteurs [FS08] ... L'une des forces de la théorie des langages formels est sa capacité à modéliser différentes théories logiques, comme l'arithmétique de Presburger ou la logique monadique du second ordre [Tho90]. La recherche en théorie des langages formels cherche, entre autres questions, à préciser, à classer et à comparer les différents modèles d' $\omega$ -automates existants.

Du point de vue de la combinatoire des mots, la théorie des langages formels peut fournir des outils intéressants pour étudier les familles classiques de mots. Cette fois, la variante n'est pas sur le modèle de l' $\omega$ -automate étudié, mais sur le mot lui-même : étant donné un  $\omega$ -automate  $\mathcal{A}$ , accepte-t-il un mot vérifiant la propriété P? Les deux domaines ont longtemps été étudiés conjointement, par exemple par Perrin [PP04] ou Schützenberger [PS92]. Un résultat récent à l'interaction entre ces deux domaines est la création, autour de 2020, du logiciel de vérification formelle Walnut par Shallit [MSS16] : pour les suites automatiques, ce logiciel peut vérifier une propriété décrite par la logique du premier ordre. Ce résultat a été étendu par Oei aux mots Sturmiens [Hie+22].

Toutefois, la question de l'acceptation d'un mot par un  $\omega$ -automate est rarement étudiée directement par la combinatoire des mots. Un rare exemple est un article de Carton et Thomas [CT02], qui résout le problème de l'acceptation d'un mot morphique par un  $\omega$ -automate de Büchi, dont la méthode est directe et combinatoire. Peut-on étendre ce résultat au cas des mots Sturmiens?

Les approches logiques de Shallit et Oei donnent certes des méthodes efficaces pour répondre à certaines questions, mais ne sont pas satisfaisantes sur certains aspects. D'abord, la construction faite dans leurs articles est particulièrement complexe, et repose sur des résultats techniques sur les systèmes de numération. Walnut ne peut alors être interprété que comme une boîte noire, qui répond aux questions fournies sans les expliquer davantage : les différents articles publiés par Shallit depuis plusieurs années [BSS24; CS24; Sha23; SSZ23] ont répondu à plusieurs conjectures ou questions préexistantes sans mettre en lumière de phénomènes combinatoires sous-jacents.

Dans ce chapitre, j'apporte des méthodes combinatoires pour résoudre le problème de l'acceptation d'un mot Sturmien par un  $\omega$ -automate faible. Les algorithmes proposés reposent sur la caractérisation S-adique des mots Sturmiens. Même si le modèle d' $\omega$ -automates étudié dans ce chapitre est relativement peu puissant, il trouve néanmoins plusieurs applications intéressantes. Par exemple, du point de vue de la dynamique symbolique, les  $\omega$ -automates faibles représentent les shifts sofiqes : mes méthodes permettent alors de décider s'il existe un mot Sturmien dans un shift sofiq donné, et apportent des algorithmes à l'intersection entre dynamique symbolique et combinatoire des mots.

Ce chapitre se découpe en trois parties : dans un premier temps, je détaille les méthodes combinatoires que j'ai établies sur les  $\omega$ -automates faibles. Ce travail a été publié à la conférence *Combinatorics on Words* [BH23]. Je détaille en particulier une variété de questions reliées, comme l'acceptation d'un mot purement substitutif ou d'un point fixe d'une substitution. J'établis plusieurs résultats plus théoriques liant mots Sturmiens et  $\omega$ -automates faibles, en particulier sur les suites directrices des mots Sturmiens acceptés par un  $\omega$ -automate faible. Ensuite, j'introduis plusieurs applications des algorithmes établis, notamment en combinatoire des mots sur une question de codage, ou en géométrie discrète sur un problème de recollement de segments. Enfin, j'aborde des résultats partiels sur l'extension aux  $\omega$ -automates de Büchi.

## 2.2 Représentations S-adiques et $\omega$ -automates faibles

### 2.2.1 Définitions

Je commence par donner les définitions formelles des objets manipulés tout le long de ce chapitre. Ce formalisme est adapté de la théorie des langages formels, et plus précisément de [PP04].

**Définition 2.2.1** ( $\omega$ -automate). *Un  $\omega$ -automate  $\mathfrak{A}$  est un uplet  $(\mathcal{A}, Q, I, T, C)$  où  $\mathcal{A}$  est un alphabet fini,  $Q$  est un ensemble fini d'états,  $I \subseteq Q$  est un ensemble d'états initiaux,  $T \subseteq Q \times \mathcal{A} \times Q$  est un ensemble de transitions et  $C \subseteq T^{\mathbb{N}}$  est la condition d'acceptation d'une marche par  $\mathfrak{A}$ . Une transition  $(d, a, f) \in T$  sera généralement notée  $d \xrightarrow{a} f$  ( $d$  pour début,  $f$  pour fin).*

**Définition 2.2.2** (Marche). *Pour  $n \geq 0$  ou  $n = \infty$ , une suite  $(t_k)_{0 \leq k \leq n}$  avec  $t_k = d_k \xrightarrow{a_k} f_k \in T$  est une marche sur  $\mathfrak{A}$  si  $f_k = d_{k+1}$  pour tout  $0 \leq k < n$ . Le mot  $m = (a_k)_{0 \leq k \leq n}$  étiquette la marche. Pour deux états  $d$  et  $f \in Q$  et un mot fini  $w \in \mathcal{A}^*$ , la notation  $d \xrightarrow{m}^* f$  signifie une marche sur  $\mathfrak{A}$   $(t_k)_{0 \leq k \leq n}$  (avec  $n \in \mathbb{N}$ ) étiquetée par  $m$  et telle que  $d_0 = d$  et  $f_n = f$ . Plusieurs marches peuvent remplir cette condition, et la notation  $d \xrightarrow{m}^* f$  signifiera un choix arbitraire parmi l'une d'entre elles. Enfin, je considérerai que pour  $q \in Q$ ,  $q \xrightarrow{\varepsilon}^* q$  sera toujours une marche valide sur  $\mathfrak{A}$ , et représentera une marche de  $q$  à  $q$  étiquetée par le mot vide.*

**Définition 2.2.3** (Acceptation d'un mot infini par un  $\omega$ -automate). *Un mot infini  $x \in \mathcal{A}^{\mathbb{N}}$  est accepté par un  $\omega$ -automate  $\mathfrak{A} = (\mathcal{A}, Q, I, T, C)$  s'il existe une suite  $(t_k)_{k \in \mathbb{N}} \in T^{\mathbb{N}}$  qui vérifie les conditions suivantes :*

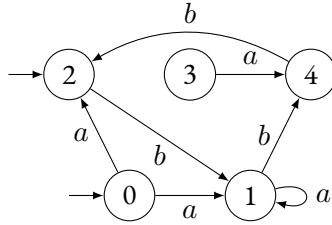
1.  $(t_k)_{k \in \mathbb{N}}$  est une marche sur  $\mathfrak{A}$ ;
2. en notant  $t_0 = d_0 \xrightarrow{a_0} f_0$ , on a  $d_0 \in I$ ;
3.  $x$  étiquette la marche  $(t_k)_{k \in \mathbb{N}}$ ;
4.  $(t_k)_{k \in \mathbb{N}} \in C$ .

*On écrit alors que  $(t_k)_{k \in \mathbb{N}}$  est un calcul acceptant pour  $x$ . Le langage de  $\mathfrak{A}$ , noté  $\mathcal{L}(\mathfrak{A})$ , est l'ensemble des mots infinis acceptés par  $\mathfrak{A}$ .*

Ces notations, très généralistes, ne sont pas toujours représentatives de l'utilisation qui en sera faite dans ce manuscrit. En particulier, la condition d'acceptation sera explicitée par le modèle d' $\omega$ -automate utilisé. Un premier exemple est le modèle des  $\omega$ -automates faibles, qui ont la condition d'acceptation la plus simple.

**Définition 2.2.4** ( $\omega$ -automate faible). *Un  $\omega$ -automate faible  $\mathfrak{A}$  est un  $\omega$ -automate  $(\mathcal{A}, Q, I, T, C)$  où  $C = T^{\mathbb{N}}$  : un mot infini  $x$  est accepté par  $\mathfrak{A}$  si et seulement si  $x$  étiquette une marche infinie sur  $\mathfrak{A}$  commençant par un état initial. Je noterai alors  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$ .*

**Exemple 2.2.5.** *Dans cette thèse, je représenterai les  $\omega$ -automates faibles comme dans le schéma représenté dans la figure 2.1.*



**FIGURE 2.1** – Exemple d' $\omega$ -automate faible  $\mathfrak{A}_1$ .

Les états sont les cercles, les transitions d'état en état sont les flèches étiquetées. Exceptionnellement dans cet exemple, j'ai nommé explicitement les états : je ne le ferai pas dans la suite de ce manuscrit. Les états initiaux de l' $\omega$ -automate faible  $\mathfrak{A}_1$  sont ceux pointés par une petite flèche sans étiquette : ici, il s'agit des états 0 et 2. On remarque, par exemple, que dans cet  $\omega$ -automate, l'état 3 n'est pas accessible depuis les états initiaux, et que l'une des arêtes de  $\mathfrak{A}_1$  est une boucle étiquetée par  $a$ . Une marche dans l' $\omega$ -automate correspond, graphiquement, à aller d'état en état, et le mot associé est la concaténation des étiquettes des transitions.

Par exemple, il existe une marche infinie sur  $\mathfrak{A}$  de la forme suivante :

$$0 \xrightarrow{a} 1 \xrightarrow{a} 1 \xrightarrow{a} 1 \xrightarrow{b} 4 \xrightarrow{b} 2 \xrightarrow{b} 1 \xrightarrow{a} 1 \xrightarrow{b} 4 \xrightarrow{b} 2 \xrightarrow{b} 1 \xrightarrow{a} 1 \xrightarrow{b} 4 \xrightarrow{b} 2 \xrightarrow{b} \dots$$

Cette marche fait un cycle sur les états 1, 1, 4, 2; elle est étiquetée par le mot infini ultimement périodique  $a a (a b b b)^{\mathbb{N}}$ . Comme  $\mathfrak{A}_1$  est un  $\omega$ -automate faible, n'importe quel mot infini étiquetant une marche infinie sur  $\mathfrak{A}_{\text{ex}}$  en commençant par un état initial est accepté par  $\mathfrak{A}_1$  : donc  $a a (a b b b)^{\mathbb{N}} \in \mathcal{L}(\mathfrak{A}_1)$ .

En utilisant cette notion d' $\omega$ -automate faible, il est possible de fournir une autre définition des sous-shifts so-fiques (voir la définition 1.1.24) :

**Théorème 2.2.6** ([LM95]). *Un shift  $\mathbb{X}$  est so-fique si et seulement s'il existe un  $\omega$ -automate faible  $\mathfrak{A}$  dont tous les états sont initiaux et tel que  $\mathcal{L}(\mathfrak{A}) = \mathbb{X}$ .*

Par ailleurs, il est possible d'étendre les propriétés topologiques des shifts so-fiques aux langages d' $\omega$ -automates faibles. La proposition suivante sera un outil puissant très important dans le reste de cette section :

**Proposition 2.2.7.** *Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  un  $\omega$ -automate faible. Alors son langage  $\mathcal{L}(\mathfrak{A}) \subseteq \mathcal{A}^{\mathbb{N}}$  est un compact pour la topologie prodiscrète.*

*Démonstration.*

Si on s'intéresse à  $\mathcal{C}$  l'ensemble des calculs acceptants sur  $\mathfrak{A}$ , on remarque qu'on peut les définir à partir

des cylindres de la topologie prodiscrète dans le shift plein  $T^{\mathbb{N}}$  :

$$\mathcal{C} = \left( \bigcup_{\substack{t \in T, \\ t=d \rightarrow f, \\ d \in I}} [t]_0 \right) \cap \bigcap_{n \in \mathbb{N}} \left( \bigcup_{\substack{t_1, t_2 \in T \\ t_1=d_1 \rightarrow f_1, t_2=d_2 \rightarrow f_2, \\ f_1=d_2}} [t_1 t_2]_n \right)$$

Le premier ensemble retranscrit qu'un calcul doit commencer par un état initial, et le deuxième ensemble que deux transitions successives doivent être ... successives. Donc  $\mathcal{C}$  est bien un compact de  $T^{\mathbb{N}}$ . Maintenant, si on définit  $\pi : T \rightarrow \mathcal{A}$  qui associe à une transition son étiquette, et en étendant  $\pi$  comme une substitution sur  $T^{\mathbb{N}}$ , on a que  $\mathcal{L}(\mathfrak{A}) = \pi(\mathcal{C})$ . Comme  $\pi$  est une substitution,  $\pi$  définit une application continue, et donc  $\mathcal{L}(\mathfrak{A})$  est bien un compact.

Le lecteur expert en dynamique symbolique remarquera que la première intersection de cylindres définit des conditions non invariantes par décalage, alors que la deuxième intersection définit une condition ressemblant à un shift de type fini.

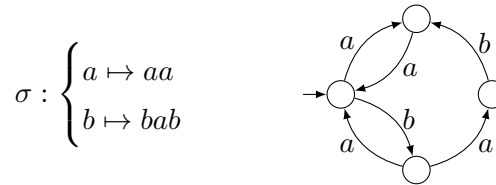
### 2.2.2 Désubstitution des $\omega$ -automates faibles

Le point de vue automatique des définitions précédentes n'est pas standard dans la littérature de la dynamique symbolique, et n'est pas innocent dans le contexte de ce chapitre de thèse. Lind et Marcus [LM95] ont en effet employé le vocabulaire des graphes, et il existe des définitions purement dynamiques des shifts sofiques [Wei73]. L'avantage du point de vue automatique est qu'il justifie l'utilisation d'outils propres à la théorie des langages formels, comme la désubstitution.

La désubstitution est bien connue dans la littérature des langages formels, en particulier dans le cas des automates finis [HU79; Car08]. Elle permet d'illustrer la stabilité des langages réguliers par morphisme inverse. Son utilisation dans le cadre de la dynamique symbolique et de la combinatoire des mots est en revanche beaucoup plus marginale (je n'ai trouvé comme seul exemple que [Sal17], et pas de manière explicite).

**Définition 2.2.8** (Désubstitution d'un  $\omega$ -automate faible). *Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  un  $\omega$ -automate faible, et  $\sigma$  un morphisme. La désubstitution de  $\mathfrak{A}$  par  $\sigma$  est l' $\omega$ -automate faible  $\sigma^{-1}(\mathfrak{A}) = (\mathcal{A}, Q, I, \sigma^{-1}(T))$  où  $d \xrightarrow{a} f \in \sigma^{-1}(T)$  si et seulement s'il existe une marche sur  $\mathfrak{A}$  de la forme  $d \xrightarrow{\sigma(a)^*} f$ .*

**Exemple 2.2.9.** *Considérons l' $\omega$ -automate faible  $\mathfrak{A}_2$  et la substitution  $\sigma$  suivantes :*



**FIGURE 2.2** – Exemple d'une substitution  $\sigma$  et d' $\omega$ -automate faible  $\mathfrak{A}_2$ .

Pour calculer l' $\omega$ -automate faible  $\sigma^{-1}(\mathfrak{A}_2)$ , on procède comme suit : d'abord, les états de  $\sigma^{-1}(\mathfrak{A}_2)$  sont les mêmes que ceux de  $\mathfrak{A}$ ; ensuite on calcule les arêtes de  $\sigma^{-1}(\mathfrak{A}_2)$ . Pour cela, on commence par identifier les marches dans  $\mathfrak{A}$  étiquetées par  $\sigma(a) = aa$ . On relie alors les débuts et les fins de toutes ces marches par  $a$ . On procède de même pour  $\sigma(b) = bab$ . Ainsi, on obtient  $\sigma^{-1}(\mathfrak{A})$  :



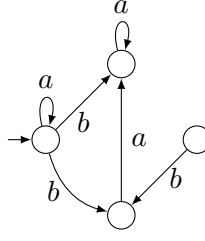


FIGURE 2.3 –  $L\omega$ -automate faible  $\sigma^{-1}(\mathfrak{A}_2)$ .

Dans le cas d'un morphisme effaçant, la convention de la définition 2.2.2 se gère en créant des boucles étiquetées par les lettres effacées sur tous les états : par exemple, en considérant le même  $\omega$ -automate faible  $\mathfrak{A}_2$  et le morphisme  $\tau : a \mapsto aab, b \mapsto \varepsilon$ , l' $\omega$ -automate  $\tau^{-1}(\mathfrak{A}_2)$  est le suivant :

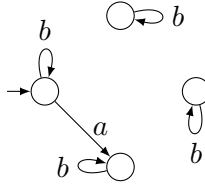


FIGURE 2.4 – Désubstitution de  $\mathfrak{A}_2$  par le morphisme effaçant  $\tau$ .

**Proposition 2.2.10.** Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  un  $\omega$ -automate faible,  $\sigma$  un morphisme,  $m \in \mathcal{A}^*$  et  $d, f \in Q$ . Alors  $d \xrightarrow{m}^* f$  est une marche de  $\sigma^{-1}(\mathfrak{A})$  ssi  $d \xrightarrow{\sigma(m)}^* f$  est une marche de  $\mathfrak{A}$ .

*Démonstration.*

On procède par récurrence sur la longueur de  $m$ . Le cas  $|m| = 0$  est traité par la convention de la définition 2.2.2.

Si  $d \xrightarrow{m}^* q \xrightarrow{a} f$  est une marche de  $\sigma^{-1}(\mathfrak{A})$ , alors par hypothèse de récurrence,  $d \xrightarrow{\sigma(m)}^* q$  est une marche de  $\mathfrak{A}$ . De plus, comme  $q \xrightarrow{a} f$  est une marche de  $\sigma^{-1}(\mathfrak{A})$ ,  $q \xrightarrow{\sigma(a)}^* f$  est une marche de  $\mathfrak{A}$ . Par concaténation des marches,  $d \xrightarrow{\sigma(ma)}^* f$  est une marche de  $\mathfrak{A}$ .

Réciproquement, si  $d \xrightarrow{\sigma(m)}^* q \xrightarrow{\sigma(a)}^* f$  est une marche de  $\mathfrak{A}$ , par hypothèse de récurrence,  $d \xrightarrow{m}^* q$  est une marche de  $\sigma^{-1}(\mathfrak{A})$ . De plus, comme  $q \xrightarrow{\sigma(a)}^* f$  est une marche de  $\mathfrak{A}$ ,  $q \xrightarrow{a} f$  est une marche de  $\sigma^{-1}(\mathfrak{A})$ . Par concaténation des marches,  $d \xrightarrow{ma}^* f$  est une marche de  $\sigma^{-1}(\mathfrak{A})$ .

**Proposition 2.2.11.** Pour un  $\omega$ -automate faible  $\mathfrak{A}$  et un morphisme  $\sigma$ ,  $\mathcal{L}(\sigma^{-1}(\mathfrak{A})) = \{w \in \mathcal{A}^{\mathbb{N}} \mid \sigma(w) \in \mathcal{L}(\mathfrak{A})\}$ .

*Démonstration.*

On pose  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$ . Soit  $x \in \mathcal{A}^{\mathbb{N}}$  un mot infini accepté par  $\sigma^{-1}(\mathfrak{A})$ , et soit  $(t_k)_{k \in \mathbb{N}}$  un calcul acceptant pour  $x$ , avec  $t_k = q_k \xrightarrow{x_k} q_{k+1}$ , et  $q_0 \in I$ . Par définition de  $\sigma^{-1}(\mathfrak{A})$ , pour tout  $k \in \mathbb{N}$ , il existe dans  $\mathfrak{A}$  une marche de la forme  $q_k \xrightarrow{\sigma(x_k)}^* q_{k+1}$ . En concaténant ces marches dans  $\mathfrak{A}$ , on obtient alors une marche sur  $\mathfrak{A}$  de la forme  $q_0 \xrightarrow{\sigma(x_0)}^* q_1 \xrightarrow{\sigma(x_1)}^* q_2 \xrightarrow{\sigma(x_2)}^* \dots$ . C'est un calcul acceptant pour  $\sigma(x)$  dans  $\mathfrak{A}$ , donc  $\sigma(x) \in \mathcal{L}(\mathfrak{A})$ .

Réciproquement, soit  $x \in \mathcal{A}^{\mathbb{N}}$  tel que  $\sigma(x)$  est accepté par  $\mathfrak{A}$ . Soit  $(t_k)_{k \in \mathbb{N}}$  un calcul acceptant pour  $\sigma(x)$  dans  $\mathfrak{A}$  : alors on peut l'écrire sous la forme  $q_0 \xrightarrow{\sigma(x)_0} q_1 \xrightarrow{\sigma(x)_1} q_2 \xrightarrow{\sigma(x)_2} \dots$ . On peut alors regrouper certaines transitions pour réécrire ce calcul sous la forme  $q_0 = q'_0 \xrightarrow{\sigma(x_0)}^* q'_1 \xrightarrow{\sigma(x_1)}^* q'_2 \xrightarrow{\sigma(x_2)}^* \dots$

\* ... Par définition de  $\sigma^{-1}(\mathfrak{A})$ , on a donc  $q'_0 \xrightarrow{x_0} q'_1 \xrightarrow{x_1} q'_2 \xrightarrow{x_2} \dots$  un calcul acceptant pour  $x$  dans  $\sigma^{-1}(\mathfrak{A})$ , donc  $x \in \mathcal{L}(\sigma^{-1}(\mathfrak{A}))$ .

Le principal intérêt des méthodes présentées dans ce chapitre ne réside pas dans l'originalité de leur construction, mais dans leur application à la combinatoire des mots, et à leur systématisation, qui permet d'établir de nouveaux outils algorithmiques.

Je donne ici une complexité de l'opération de désubstitution d'un  $\omega$ -automate faible afin de pouvoir, plus tard, estimer la complexité d'algorithmes qui l'utiliseront.

**Proposition 2.2.12.** *Étant donné un  $\omega$ -automate faible  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  et un morphisme  $\sigma$ , l'automate  $\sigma^{-1}(\mathfrak{A})$  peut se construire en temps  $\mathcal{O}(|Q|^3 |\mathcal{A}| \max_a |\sigma(a)|)$ .*

*Démonstration.*

Pour construire  $\sigma^{-1}(\mathfrak{A})$ , on cherche pour tout couple d'états  $(d, f)$  et lettre  $a$  s'il doit y avoir une transition  $d \xrightarrow{a} f$  dans  $\sigma^{-1}(\mathfrak{A})$ . Pour cela, on part de  $d$  et on cherche à rejoindre  $f$  en lisant le mot  $\sigma(a)$ , de longueur plus petite que  $\max_a |\sigma(a)|$ . Le parcours en largeur donne un algorithme en  $\mathcal{O}(|Q| \times \max_a |\sigma(a)|)$ . Il y a  $|Q|^2 \times |\mathcal{A}|$  possibilités pour  $d, f$  et  $a$ , d'où la complexité annoncée.

La borne proposée est sûrement améliorable<sup>1</sup> mais de manière générale, l'opération reste en temps polynomial en la taille de l'automate.

Une propriété importante de cette opération est sa compatibilité avec la composition :

**Proposition 2.2.13.** *Pour  $\mathfrak{A}$  un  $\omega$ -automate faible et  $\sigma$  et  $\tau$  deux morphismes, on a  $(\sigma \circ \tau)^{-1}(\mathfrak{A}) = \tau^{-1}(\sigma^{-1}(\mathfrak{A}))$ .*

*Démonstration.*

Les deux  $\omega$ -automates ont même alphabet, mêmes états et mêmes états initiaux : il suffit de montrer qu'ils ont les mêmes transitions. Notons  $T$  les transitions de  $(\sigma \circ \tau)^{-1}(\mathfrak{A})$  et  $T'$  les transitions de  $\tau^{-1}(\sigma^{-1}(\mathfrak{A}))$ .

Alors :

$$\begin{aligned} d \xrightarrow{a} f \in T &\iff d \xrightarrow{\sigma \circ \tau(a)}^* f \text{ est une marche de } \mathfrak{A}, \\ &\iff d \xrightarrow{\tau(a)}^* f \text{ est une marche de } \sigma^{-1}(\mathfrak{A}), \text{ par la proposition 2.2.10,} \\ &\iff d \xrightarrow{a} f \in T', \text{ par la définition 2.2.8.} \end{aligned}$$

Cette propriété de composition, qui résulte simplement des propriétés élémentaires des constructions précédentes, nous sera particulièrement utile par la suite, car elle nous permettra d'obtenir des désubstitutions de compositions par étapes successives.

### 2.2.3 Algorithmes pour une substitution

Maintenant que les principaux outils ont été expliqués, je développe les techniques utilisées pour décider plusieurs problèmes sur les  $\omega$ -automates. Les méthodes présentées ici reposent expressément sur les deux faits suivants :

1. par exemple, en modélisant l'automate par les matrices d'incidences des lettres, on peut directement calculer les arêtes de  $\sigma^{-1}(\mathfrak{A})$  en procédant à des multiplications matricielles, ce qui donne une complexité en  $\mathcal{O}(Q^\omega |\mathcal{A}| \max_a |\sigma(a)|)$  avec  $\omega$  un exposant de la multiplication matricielle, et il est connu que  $\omega < 3$  [Bos+17].

**Fait 2.2.14.** Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  un  $\omega$ -automate faible. Soit  $\mathfrak{S}(\mathfrak{A})$  l'ensemble des  $\omega$ -automates faibles ayant le même alphabet, les mêmes états et les mêmes états initiaux que  $\mathfrak{A}$ . Alors pour tout morphisme  $\sigma$ ,  $\sigma^{-1}(\mathfrak{A})$  est un élément de  $\mathfrak{S}(\mathfrak{A})$ .

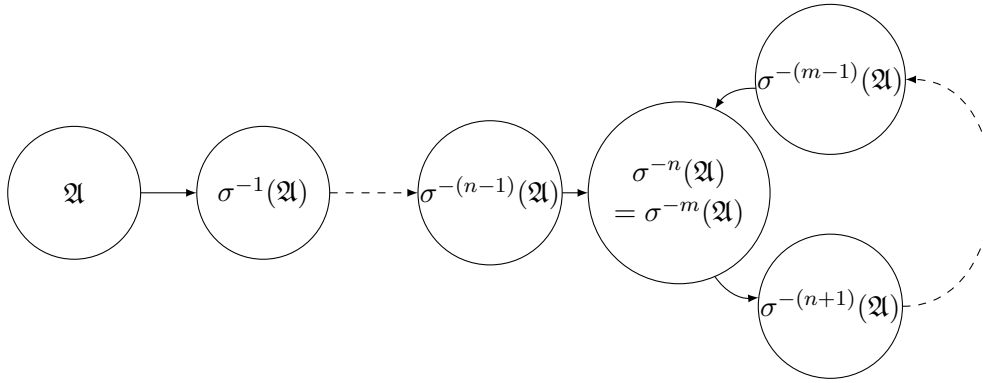
*Démonstration.*

C'est une conséquence directe de la définition 2.2.8.

Or, on observe que  $\mathfrak{S}(\mathfrak{A})$  est fini : tout élément  $\mathfrak{B}$  de  $\mathfrak{S}(\mathfrak{A})$  s'écrit comme  $\mathfrak{B} = (\mathcal{A}, Q, I, T')$ , avec  $T' \subseteq Q \times \mathcal{A} \times Q$ , donc  $\text{Card}(\mathfrak{S}(\mathfrak{A})) = 2^{|Q|^2 \times |\mathcal{A}|}$ . En pratique, dans ce chapitre, ce sont les langages desdits  $\omega$ -automates qui m'intéressent. Il serait donc possible de réduire l'espace de travail en identifiant les éléments de  $\mathfrak{S}(\mathfrak{A})$  qui ont le même langage [BDS12] : la finitude de l'ensemble est toutefois suffisante pour conclure sur la décidabilité des questions étudiées.

Soit  $\mathfrak{A}$  un  $\omega$ -automate faible et  $\sigma$  un morphisme : alors la désubstitution de  $\mathfrak{A}$  par  $\sigma$  définit une dynamique sur l'ensemble  $\mathfrak{S}(\mathfrak{A})$ . Par le lemme des tiroirs :

**Fait 2.2.15.** Soit  $\mathfrak{A}$  un  $\omega$ -automate faible, et soit  $\sigma$  un morphisme. Alors il existe  $0 \leq n < m \leq \text{Card}(\mathfrak{S}(\mathfrak{A}))$  tels que  $\sigma^{-n}(\mathfrak{A}) = \sigma^{-m}(\mathfrak{A})$ .



**FIGURE 2.5** – Schéma de l'idée du fait 2.2.15 : chaque flèche représente une opération de désubstitution par  $\sigma$ . Comme  $\mathfrak{S}(\mathfrak{A})$  est fini, on retombera à un moment sur un  $\omega$ -automate faible vu précédemment.

Dans la suite de cette partie, je montre que, étant donné un  $\omega$ -automate  $\mathfrak{A}$  et un morphisme  $\sigma$ , les problèmes de détection d'un point fixe de  $\sigma$  ou d'un mot purement substitutif généré par  $\sigma$  accepté par  $\mathfrak{A}$  sont décidables. Je vais me concentrer sur le cas où  $\sigma$  est une substitution, c'est-à-dire où  $\sigma$  est un morphisme non effaçant. Les résultats que je prouve dans cette section peuvent être élargis aux morphismes effaçants grâce au théorème 1.1.17 (voir [Cob68; CLR16]). L'hypothèse de non-effacement me permettra d'utiliser des arguments de continuité, et de pouvoir m'assurer de la construction de mots infinis. Désormais, tous les morphismes seront donc non effaçants.

### Acceptation d'un point fixe d'une puissance d'une substitution par un $\omega$ -automate faible

Je commence par démontrer un premier résultat sur la détection d'un point fixe pour une puissance d'une substitution. Il s'agit d'une première étape où la condition de détection est la plus simple : il suffira de vérifier qu'un langage est non vide. J'utiliserai plusieurs fois les notations du fait 2.2.15.

**Proposition 2.2.16.** Soit  $\mathfrak{A}$  un  $\omega$ -automate faible, soit  $\sigma$  une substitution, et soit  $0 \leq n < m \leq \text{Card}(\mathfrak{S}(\mathfrak{A}))$  tels que  $\sigma^{-n}(\mathfrak{A}) = \sigma^{-m}(\mathfrak{A})$ . Alors  $\mathfrak{A}$  accepte un point fixe de  $\sigma^k$  pour un certain  $k \geq 1$  si, et seulement si,  $\mathcal{L}(\sigma^{-n}(\mathfrak{A}))$  est non vide.

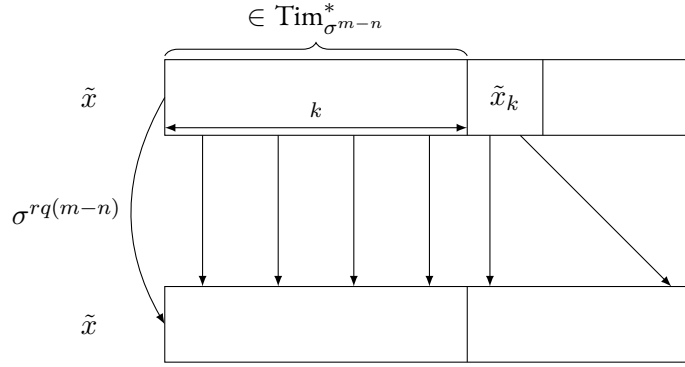
*Démonstration.*

Si  $\mathcal{L}(\sigma^{-n}(\mathfrak{A}))$  est vide, alors d'après les propositions 2.2.11 et 2.2.13,  $\mathcal{L}(\sigma^{-p}(\mathfrak{A}))$  est vide pour tout  $p \geq n$ . Soit  $x \in \mathcal{A}^{\mathbb{N}}$  tel que  $x = \sigma^k(x)$  pour un certain  $k \geq 1$ . Alors, par itération,  $x = \sigma^k(x) = \sigma^{kn}(x)$ . Si  $x$  était accepté par  $\mathfrak{A}$ , alors  $x \in \mathcal{L}(\sigma^{kn}(\mathfrak{A}))$  qui est vide. Par contradiction,  $\mathfrak{A}$  n'accepte aucun point fixe d'aucun  $\sigma^k$ .

Si  $\mathcal{L}(\sigma^{-n}(\mathfrak{A}))$  est non vide, soit  $x$  un mot accepté par  $\sigma^{-n}(\mathfrak{A})$ . De nouveau, d'après les propositions 2.2.11 et 2.2.13, comme  $\sigma^{-n}(\mathfrak{A}) = \sigma^{-m}(\mathfrak{A}) = \sigma^{-(m-n)}(\sigma^{-n}(\mathfrak{A}))$ ,  $x$  est accepté par  $\sigma^{-j(m-n)}(\sigma^{-n}(\mathfrak{A})) = \sigma^{-(n+j(m-n))}(\mathfrak{A})$  pour tout  $j \in \mathbb{N}$ . Cela signifie que  $\sigma^{n+j(m-n)}(x)$  est accepté par  $\mathfrak{A}$  pour tout  $j \in \mathbb{N}$ . Soit alors  $\tilde{x}$  une valeur d'adhérence de la suite  $(\sigma^{n+j(m-n)}(x))_{j \in \mathbb{N}}$ . Par compacité du langage d'un  $\omega$ -automate faible (voir la proposition 2.2.7),  $\tilde{x} \in \mathcal{L}(\sigma^{-n}(\mathfrak{A}))$ .

On définit alors  $\text{Tim}_{\sigma^{m-n}} \subseteq \mathcal{A}$  l'ensemble des lettres timides de  $\sigma^{m-n}$  :  $a \in \text{Tim}_{\sigma^{m-n}}$  si  $|\sigma^{j(m-n)}(a)| = 1$  pour tout  $j \geq 1$ . Soit  $k = \inf\{i \in \mathbb{N} \mid \tilde{x}_i \notin \text{Tim}_{\sigma^{m-n}}\}$  ( $k$  peut être infini).

Alors pour tout  $i < k$ , comme  $\sigma$  est une substitution non-effaçante et que toute lettre de  $\tilde{x}_{\llbracket 0, k-1 \rrbracket}$  est timide,  $\sigma^{m-n}(\tilde{x})_i = \sigma^{m-n}(\tilde{x}_i)$ . De plus, comme  $\tilde{x}$  est une valeur d'adhérence de  $(\sigma^{n+j(m-n)}(x))_{j \in \mathbb{N}}$ , il existe  $r(\tilde{x}_i) \geq 1$  tel que  $\sigma^{r(\tilde{x}_i) \times (m-n)}(\tilde{x}_i) = \tilde{x}_i$  pour toute position  $i < k$ . La valeur de  $r(\tilde{x}_i)$  ne dépend que de  $\tilde{x}_i$ , pas de sa position : si  $\tilde{x}_i = \tilde{x}_j$ , alors  $r(\tilde{x}_i) = r(\tilde{x}_j)$ . Par finitude de  $\mathcal{A}$ ,  $\{r(\tilde{x}_i)\}_{0 \leq i < k}$  ne contient qu'un nombre fini de valeurs distinctes, et on peut donc définir  $r = \text{PPCM}_{0 \leq i < k} \{r(\tilde{x}_i)\}$ . Lorsque  $k < \infty$ , il existe  $q \geq 1$  tel que  $|\sigma^{q(m-n)}(\tilde{x}_k)| > 1$  et  $\tilde{x}_k \sqsubseteq_p \sigma^{q(m-n)}(\tilde{x}_k)$ , aussi parce que  $\tilde{x}$  est une valeur d'adhérence de  $(\sigma^{n+j(m-n)}(x))_{j \in \mathbb{N}}$ . Si  $k = \infty$ , je pose  $q = 1$ .



**FIGURE 2.6** – Schéma du découpage de  $\tilde{x}$  : le préfixe ne croît pas, tandis que le reste du mot est généré par  $\tilde{x}_k$ . Alors, si  $k < \infty$ , par concaténation,  $\tilde{x}_{\llbracket 0, k \rrbracket}$  est un préfixe strict de  $\sigma^{rq(m-n)}(\tilde{x}_{\llbracket 0, k \rrbracket})$ ; si  $k = \infty$ ,  $\sigma^{2rq(m-n)}(\tilde{x}) = \sigma^{rq(m-n)}(\tilde{x})$ . Dans les deux cas, la suite  $(\sigma^{jrq(m-n)}(\tilde{x}))_{j \in \mathbb{N}}$  a une limite, qui est un point fixe de  $\sigma^{rq(m-n)}$ , et par compacité de  $\mathcal{L}(\mathfrak{A})$ , cette limite est acceptée par  $\mathfrak{A}$ .

**Corollaire 2.2.17.** *Le problème suivant est décidable :*

**Donnée :** un  $\omega$ -automate faible  $\mathfrak{A}$  et une substitution  $\sigma$

**Question :** est-ce que  $\mathfrak{A}$  accepte un point fixe d'un  $\sigma^k$  pour un certain  $k \in \mathbb{N}$  ?

*Démonstration.*

D'après la proposition 2.2.16, le problème revient à décider si  $\mathcal{L}(\sigma^{-n}(\mathfrak{A}))$  est vide. La valeur  $n$  est calculable : il suffit de désubstituer  $\mathfrak{A}$  autant que nécessaire pour retomber sur un  $\omega$ -automate obtenu précédemment. La vacuité du langage d'un  $\omega$ -automate faible est un problème décidable, d'où la conclusion.

La preuve présentée ici peut sembler complexe, car elle doit gérer différents cas sur la croissance des lettres. Mais derrière ces aspects techniques, le raisonnement principal est assez simple : l' $\omega$ -automate faible accepte des images « arbitrairement » grandes de  $\sigma$ , et on conclut par compacité. On remarquera ici l'utilisation de la compacité, qui permet d'obtenir une démonstration étonnamment précise et peu constructive. La section suivante sur les  $\omega$ -automates de Büchi illustrera qu'en fait il s'agit du meilleur outil disponible pour obtenir des résultats, et que sans la possibilité d'utiliser la topologie, nos résultats peuvent s'effondrer.

L'algorithme suggéré ci-dessus ne suffit pas à décider de l'acceptation par un  $\omega$ -automate faible d'un point fixe d'une substitution (sans puissance). Ce problème est aussi décidable, et est établi dans la proposition 2.2.20. Alors que la preuve pourrait donner l'impression que  $k$  est un multiple de  $m - n$ , ce n'est pas nécessaire : ci-dessous un exemple d' $\omega$ -automate faible où les valeurs minimales pour  $k$ ,  $m$  et  $n$  ne donnent pas de multiplicité entre  $k$  et  $m - n$ .

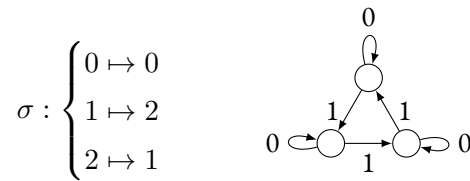


FIGURE 2.7 – Premier exemple d'absence de lien entre  $n$ ,  $m$  et  $k$ .

Dans ce premier exemple, les valeurs minimales de  $n$  et  $m$  sont 0 et 2 respectivement. Il faut désubstituer au moins deux fois  $\mathfrak{A}$  par  $\sigma$  pour retomber sur un  $\omega$ -automate faible déjà vu dans le processus. Mais en l'occurrence,  $\mathfrak{A}$  accepte un point fixe de  $\sigma : \sigma(0^{\mathbb{N}}) = 0^{\mathbb{N}}$ . Donc  $\mathfrak{A}$  accepte un point fixe de  $\sigma^k$  pour  $k = 1$ . Dans cet exemple,  $n - m \nmid k$ . La relation inverse n'est pas vraie non plus :

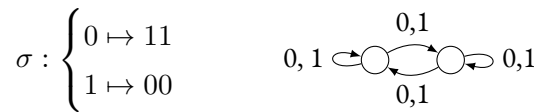


FIGURE 2.8 – Second exemple d'absence de lien entre  $n$ ,  $m$  et  $k$ .

Dans ce second exemple, les valeurs minimales de  $n$  et  $m$  sont 0 et 1 respectivement :  $\sigma^{-1}(\mathfrak{A}) = \mathfrak{A}$ . Mais la substitution  $\sigma$  n'admet pas de point fixe : seule  $\sigma^2$  en admet. Donc la valeur minimale de  $k$  est 2. Cet exemple montre qu'on peut avoir  $k \nmid n - m$ .

### Acceptation des mots purement substitutifs par un $\omega$ -automate faible

Je me tourne maintenant vers la question des mots purement substitutifs. Une remarque utile est qu'un mot purement substitutif généré par une substitution  $\sigma$  est aussi un point fixe pour un certain  $\sigma^k$  (en fait, c'est un point fixe pour toutes les puissances  $\sigma^k$ ,  $k \geq 1$ ). Je vais donc raffiner la méthode développée précédemment pour en déduire une méthode dans le cas des mots purement substitutifs.

**Proposition 2.2.18.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate faible,  $\sigma$  une substitution et  $0 \leq n < m \leq \text{Card}(\mathfrak{S}(\mathfrak{A}))$  tels que  $\sigma^{-n}(\mathfrak{A}) = \sigma^{-m}(\mathfrak{A})$ . Soit  $\text{PD}_\sigma \subseteq \mathcal{A}$  l'ensemble des lettres  $b$  qui sont strictement prolongeables à droite pour  $\sigma$ , c.-à-d.  $b \sqsubseteq_p \sigma(b)$  et  $b \neq \sigma(b)$ . Alors  $\mathfrak{A}$  accepte un mot purement substitutif généré par  $\sigma$  si, et seulement si,  $\sigma^{-n}(\mathfrak{A})$  accepte un mot infini commençant par un élément de  $\text{PD}_\sigma$ .*

*Démonstration.*

Si  $\mathfrak{A}$  accepte  $x$  un mot purement substitutif généré par  $\sigma$ , il existe  $b \in \mathcal{A}$  tel que  $x = \sigma^\infty(b)$ . Alors  $b \in \text{PD}_\sigma$  d'après le théorème 1.1.18. Comme  $\sigma(x) = x$ ,  $\sigma^n(x)$  est accepté par  $\mathfrak{A}$  donc  $x$  est accepté par  $\sigma^{-n}(\mathfrak{A})$  et commence par un élément de  $\text{PD}_\sigma$ .

Réciproquement, supposons que  $\sigma^{-n}(\mathfrak{A})$  accepte un mot infini commençant par  $b \in \text{PD}_\sigma$ . Alors  $\sigma^{m-n}(b)$  étiquette une marche sur  $\sigma^{-m}(\mathfrak{A}) = \sigma^{-n}(\mathfrak{A})$ . Par itération, pour tout  $k \geq 1$ , on a que  $\sigma^{k(m-n)}(b)$  étiquette une marche sur  $\sigma^{-n}(\mathfrak{A})$ , donc  $\sigma^{n+k(m-n)}(b)$  étiquette une marche sur  $\mathfrak{A}$ . Par compacité,  $x = \lim_{k \rightarrow \infty} \sigma^{n+k(m-n)}(b)$  est accepté par  $\mathfrak{A}$ . Maintenant, comme  $b \in \text{PD}_\sigma$ , le mot  $\lim_{j \rightarrow \infty} \sigma^j(b)$  est défini et égal à  $x$ . En conséquence,  $x$ , le mot purement substitutif généré par  $\sigma$  et la lettre  $b$ , est accepté par  $\mathfrak{A}$ .

Le résultat suivant apparaît aussi dans [CT02], avec une preuve complètement différente. L'approche présentée par Carton et Thomas repose sur la manipulation de semi-groupes de congruence, et sur la construction explicite d'un mot substitutif comme concaténation d'images de la substitution<sup>2</sup>. Le résultat est aussi apparu dans [Sal17], mais un erratum a clarifié que certains cas n'étaient pas couverts [Sal22]. La preuve ci-présente emploie le point de vue de la désubstitution, ce qui rend le résultat plus simple à étendre à d'autres problèmes.

**Corollaire 2.2.19.** *Le problème de l'acceptation d'un mot purement substitutif est décidable :*

**Donnée :** *un  $\omega$ -automate faible  $\mathfrak{A}$  et une substitution  $\sigma$*

**Question :** *est-ce que  $\mathfrak{A}$  accepte un mot purement substitutif généré par  $\sigma$  ?*

*Démonstration.*

D'après la proposition 2.2.18, il suffit de vérifier si  $\sigma^{-n}(\mathfrak{A})$  accepte un mot infini commençant par une lettre de  $\text{PD}_\sigma$  (voir notations de la proposition 2.2.18). La valeur  $n$  est calculable, tout comme l'ensemble  $\text{PD}_\sigma$ . Le problème se réduit alors à la recherche d'une marche infinie dans un  $\omega$ -automate faible en commençant par certaines arêtes, qui est un problème décidable.

Ce résultat s'étend aussi aux mots morphiques : pour trouver un mot morphique généré par  $\sigma$  et  $\tau$  accepté par  $\mathfrak{A}$ , il suffit de trouver un mot purement substitutif généré par  $\sigma$  accepté par  $\tau^{-1}(\mathfrak{A})$ .

De même, ce résultat s'étend aussi aux mots substitutifs : pour trouver un mot substitutif généré par  $\sigma$  accepté par  $\mathfrak{A}$ , il suffit de trouver un mot morphique généré par  $\sigma$  et  $\tau$  accepté par  $\mathfrak{A}$  avec  $\tau$  un codage, et il n'existe qu'un nombre fini de codages.

### Acceptation d'un point fixe par un $\omega$ -automate faible

J'étends enfin les méthodes précédentes pour montrer la décidabilité de l'acceptation d'un point fixe d'une substitution  $\sigma$  par un  $\omega$ -automate faible  $\mathfrak{A}$ . En traitant ce dernier cas, j'aurai traité la détection des quatre constructions de mots les plus classiques à partir d'un morphisme (voir la définition 1.1.14). Le résultat suivant améliore le corollaire 2.2.17, où on a montré la décidabilité de l'acceptation d'un point fixe d'une puissance d'une substitution  $\sigma$ .

**Proposition 2.2.20.** *Le problème de l'acceptation d'un point fixe est décidable :*

**Donnée :** *un  $\omega$ -automate faible  $\mathfrak{A}$ , une substitution  $\sigma$*

**Question :** *est-ce que  $\mathfrak{A}$  accepte un point fixe de  $\sigma$  ?*

<sup>2</sup>. Cette approche sera reprise dans la preuve de la proposition 2.3.16

*Démonstration.*

Soit  $x$  un point fixe de  $\sigma$ . Je définis  $\text{PF}_\sigma = \{b \in \mathcal{A} \mid \sigma(b) = b\} \subseteq \mathcal{A}$  l'ensemble des lettres points fixes de  $\sigma$ . Alors il y a deux possibilités :

1.  $x$  est un mot infini sur l'alphabet  $\text{PF}_\sigma$  ;
2. il existe une lettre  $a$  apparaissant dans  $x$  telle que  $a \neq \sigma(a)$ . Supposons que  $a$  soit la première lettre de  $x$  vérifiant cette propriété. Alors on peut écrire  $x = pax'$  où  $p \in \text{PF}_\sigma^*$ . On a donc  $x = \sigma(x) = \sigma(p)\sigma(a)\sigma(x') = p\sigma(a)\sigma(x')$ . Donc  $a \sqsubseteq_p \sigma(a)$  :  $a$  est donc une lettre strictement prolongeable à droite, et donc  $\lim_{n \rightarrow \infty} \sigma^n(a)$  existe. Comme  $x = \sigma^n(x) = p\sigma^n(a)\sigma^n(x')$  pour tout  $n \in \mathbb{N}$ , par compacité,  $x = p \lim_{n \rightarrow \infty} \sigma^n(a)$ .

L'algorithme est alors le suivant.

1. d'abord, on vérifie si  $\mathfrak{A}$  accepte un mot sur l'alphabet  $\text{PF}_\sigma$  ;
2. puis on définit un nouvel  $\omega$ -automate faible  $\mathfrak{A}'$  qui est identique à  $\mathfrak{A}$ , sauf pour l'ensemble de ses états initiaux : les états initiaux de  $\mathfrak{A}'$  sont les états accessibles depuis ceux de  $\mathfrak{A}$  en ne lisant que des lettres de  $\text{PF}_\sigma$ . On vérifie maintenant, en utilisant le corollaire 2.2.19, si  $\mathfrak{A}'$  accepte un mot purement substitutif généré par  $\sigma$ .

L'algorithme renvoie « oui » si l'un des cas précédents est vérifié et « non » sinon.

Il est remarquable de voir que les traitements des cas purement substitutifs, substitutifs, morphiques et des points fixes ne sont pas très éloignés les uns des autres. La difficulté réside dans le traitement de la croissance des mots.

## 2.2.4 Algorithmes pour plusieurs substitutions : représentations S-adiques

Les méthodes évoquées dans la section précédente se concentrent sur le cas d'une seule substitution : dans cette section, je les étends au cas où on dispose de plusieurs substitutions. En particulier, je montre dans cette section la décidabilité de l'acceptation d'un mot infiniment désubstituable par un  $\omega$ -automate faible.

Dans cette sous-section, je ferai de nouveau l'hypothèse que tous les morphismes sont des substitutions, c'est-à-dire que les morphismes sont non-effaçants. La motivation n'est pas très différente de la sous-section précédente : l'hypothèse de non-effacement est nécessaire pour pouvoir procéder à des arguments topologiques et pour s'assurer de la continuité des morphismes, comme par exemple la preuve de la proposition 2.2.24. Toutefois, on ne dispose pas d'un équivalent du théorème 1.1.17 dans le cas des mots infiniment désubstituables : on ne peut pas facilement étendre au cas où certains des morphismes seraient effaçants.

Pour obtenir des résultats de décidabilité, je commence par introduire l'une des notions phare de ce manuscrit : le méta- $\omega$ -automate de désubstitution.

**Définition 2.2.21** (Méta- $\omega$ -automate de désubstitution). *Soit  $\mathfrak{A}$  un  $\omega$ -automate faible sur l'alphabet  $\mathcal{A}$ , et  $\mathcal{S}$  un ensemble fini de substitutions sur l'alphabet  $\mathcal{A}$ . On définit l' $\omega$ -automate faible  $\mathcal{S}^{-\infty}(\mathfrak{A}) = (\mathcal{S}, \text{D}(\mathfrak{A}), \{\mathfrak{A}\}, T)$  avec pour alphabet  $\mathcal{S}$ , pour ensemble d'états  $\text{D}(\mathfrak{A}) = \{\sigma^{-1}(\mathfrak{A}), \sigma \in \mathcal{S}^*\}$ , pour seul état initial  $\mathfrak{A}$  et pour transitions l'ensemble  $T = \{\mathfrak{B} \xrightarrow{\sigma} \sigma^{-1}(\mathfrak{B})\}$ .*

Comme  $\text{D}(\mathfrak{A}) \subseteq \mathfrak{S}(\mathfrak{A})$  (voir les notations du fait 2.2.14), le nombre d'états de  $\mathcal{S}^{-\infty}(\mathfrak{A})$  est bien fini, et l' $\omega$ -automate faible  $\mathcal{S}^{-\infty}(\mathfrak{A})$  est bien calculable à partir de  $\mathfrak{A}$  et  $\mathcal{S}$ .

**Exemple 2.2.22.** *Considérons les deux substitutions suivantes sur  $\mathcal{A} = \{a, b\}$  :*

$$\sigma : \begin{cases} a \mapsto ab, \\ b \mapsto bba \end{cases} \quad \text{et} \quad \tau : \begin{cases} a \mapsto aba, \\ b \mapsto b \end{cases}$$

En notant  $\mathcal{S} = \{\sigma, \tau\}$ , on peut alors calculer les désstitutions de l' $\omega$ -automate  $\mathfrak{A}_3$  suivant par  $\mathcal{S}$ . Pour simplifier la lecture, les arêtes en noir sont étiquetées par  $a$ , celles en rouge sont étiquetées par  $b$ .

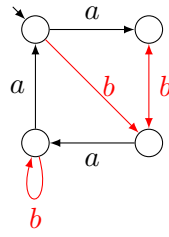


FIGURE 2.9 – Un  $\omega$ -automate faible exemple  $\mathfrak{A}_3$ .

En calculant les désstitutions successives de  $\mathfrak{A}_3$  par  $\mathcal{S}$ , on obtient alors le méta- $\omega$ -automate faible de désubstitution suivant :

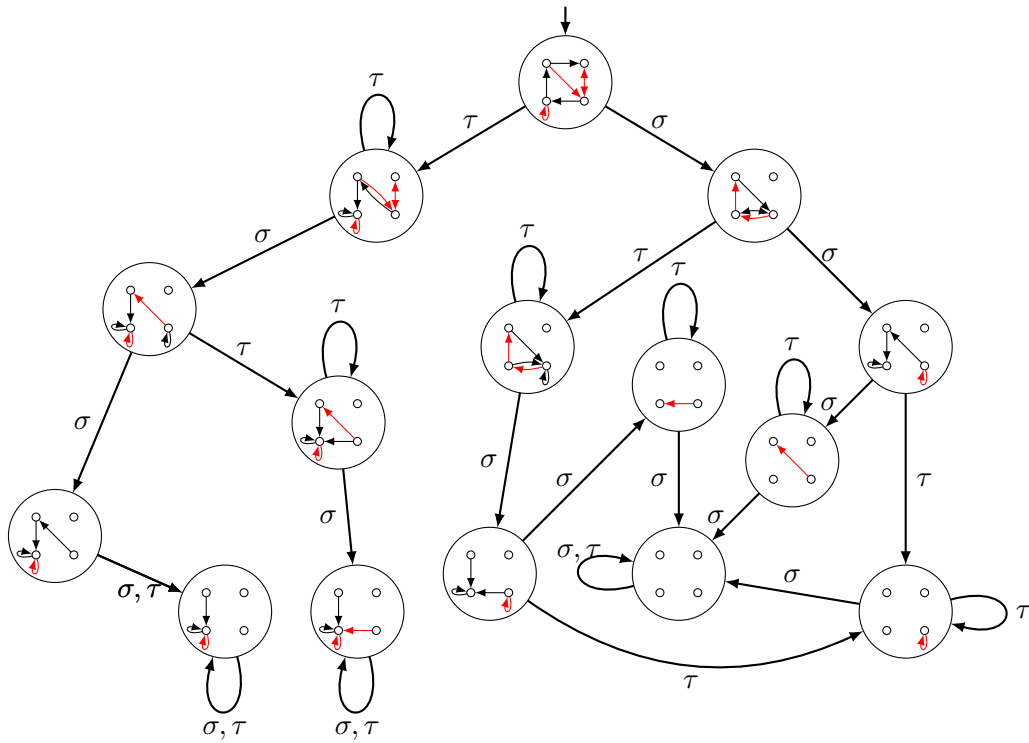


FIGURE 2.10 – Le méta- $\omega$ -automate de désubstitution  $\mathcal{S}^{-\infty}(\mathfrak{A}_3)$ .

Chaque état de ce méta- $\omega$ -automate de désubstitution est lui-même un  $\omega$ -automate faible. Une transition  $\mathfrak{B} \xrightarrow{\mu} \mathfrak{C}$  (pour  $\mathfrak{B}, \mathfrak{C}$  des  $\omega$ -automates et  $\mu$  un morphisme) signifie que  $\mathfrak{C} = \mu^{-1}(\mathfrak{B})$ . La finitude de  $D(\mathfrak{A}_3)$  assure que  $\mathcal{S}^{-\infty}(\mathfrak{A}_3)$  a un nombre fini d'états, ici 14. On remarque qu'ici, je n'ai pas quotienté par les langages : plusieurs  $\omega$ -automates présents ici ont le même langage (notamment, plusieurs ont pour langage le langage vide). En quotientant par le langage, le méta- $\omega$ -automate présenté ici aurait seulement 6 états.

On définit maintenant une notion qui permettra de caractériser les suites directrices acceptées par  $\mathfrak{A}$  :

**Définition 2.2.23** (Calcul non-nilpotent). Soit  $\mathfrak{A}$  un  $\omega$ -automate faible,  $\mathcal{S}$  un ensemble fini de substitutions, et  $\mathcal{S}^{-\infty}(\mathfrak{A})$  le méta- $\omega$ -automate de désubstitution associé. Soit  $y \in \mathcal{S}^{\mathbb{N}}$  accepté par  $\mathcal{S}^{-\infty}(\mathfrak{A})$  : on dit que  $y$  admet un calcul acceptant non-nilpotent s'il existe  $(\mathfrak{B}_i \xrightarrow{y_i} \mathfrak{B}_{i+1})_{i \in \mathbb{N}}$  sur  $\mathcal{S}^{-\infty}(\mathfrak{A})$  un calcul acceptant  $y$  tel que pour tout  $i \in \mathbb{N}$ ,  $\mathcal{L}(\mathfrak{B}_i)$  est non vide.



**Proposition 2.2.24.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate faible,  $\mathcal{S}$  un ensemble fini de substitutions. Alors il existe un mot  $x$  accepté par  $\mathfrak{A}$  et infiniment désubstituable par  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$  si, et seulement si,  $(\sigma_n)_{n \in \mathbb{N}}$  admet un calcul non-nilpotent dans  $\mathcal{S}^{-\infty}(\mathfrak{A})$ .*

*Démonstration de la proposition 2.2.24.*

Soit  $x$  un mot infiniment désubstituable de suite directrice  $(\sigma_n)_{n \in \mathbb{N}}$ , et soit  $(x^{(n)})_{n \in \mathbb{N}}$  la suite des mots désubstitués associés, c.-à-d.  $x = \sigma_{\llbracket 0, n \rrbracket}(x^{(n)})$  pour tout  $n \in \mathbb{N}$ . Alors, par la proposition 2.2.11,  $x^{(n)} \in \mathcal{L}((\sigma_{\llbracket 0, n \rrbracket})^{-1}(\mathfrak{A}))$ . Donc la suite de transitions  $((\sigma_{\llbracket 0, n \rrbracket})^{-1}(\mathfrak{A}) \xrightarrow{\sigma_{n+1}} (\sigma_{\llbracket 0, n+1 \rrbracket})^{-1}(\mathfrak{A}))_{n \in \mathbb{N}}$  définit bien une marche de  $\mathcal{S}^{-\infty}(\mathfrak{A})$ , donc chaque état est de langage non vide, et étiquetée par  $(\sigma_n)_{n \in \mathbb{N}}$ .

Maintenant, soit  $(\sigma_n)_{n \in \mathbb{N}}$  un mot accepté par  $\mathcal{S}^{-\infty}(\mathfrak{A})$  admettant un calcul non-nilpotent. Cela signifie que chaque langage  $\mathcal{L}(\sigma_{\llbracket 0, n \rrbracket}^{-1}(\mathfrak{A}))$  est non vide. Considérons dès lors la suite de langages  $(\sigma_{\llbracket 0, n \rrbracket}(\mathcal{L}(\sigma_{\llbracket 0, n \rrbracket}^{-1}(\mathfrak{A}))))_{n \in \mathbb{N}}$  :

1. tout élément de la suite est inclus dans  $\mathcal{L}(\mathfrak{A})$ ;
2. par continuité des substitutions et compacité du langage d'un  $\omega$ -automate faible, chaque élément de la suite est un compact non vide;
3. la suite est décroissante pour l'inclusion.

D'après le théorème des compacts emboîtés, il existe  $x$  un point dans l'intersection de tous les éléments de la suite. Ce point  $x$  est désubstituable par n'importe quel  $\sigma_{\llbracket 0, n \rrbracket}$ , donc est infiniment désubstituable par la suite  $(\sigma_n)_{n \in \mathbb{N}}$  et est accepté par  $\mathfrak{A}$ .

L'utilisation de la topologie, bien qu'elle puisse paraître gratuite, est bien une nécessité. En particulier, le résultat sera contredit dans le cas des  $\omega$ -automates de Büchi, où cette fois les arguments de compacité ne fonctionnent pas. De manière équivalente, on peut retirer comme sommets de  $\mathcal{S}^{-\infty}(\mathfrak{A})$  les  $\omega$ -automates faibles de langage vide, résultant en un nouveau méta- $\omega$ -automate faible  $\mathcal{S}_{\text{modif}}^{-\infty}(\mathfrak{A})$ . Alors un mot  $x$  accepté par  $\mathfrak{A}$  est infiniment désubstituable par  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$  ssi  $(\sigma_n)_{n \in \mathbb{N}}$  est accepté par  $\mathcal{S}_{\text{modif}}^{-\infty}(\mathfrak{A})$ .

**Corollaire 2.2.25.** *L'ensemble des suites directrices de mots infiniment désubstituables par  $\mathcal{S}$  acceptés par  $\mathfrak{A}$  est le langage d'un  $\omega$ -automate faible.*

Le corollaire 2.2.25 permet, par la proposition 2.2.7, d'établir que ce langage est un compact de l'ensemble  $\mathcal{S}^{\mathbb{N}}$ , et aussi d'établir le résultat suivant :

**Proposition 2.2.26.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate faible, et soit  $\mathcal{S}$  un ensemble fini de substitutions tels que  $\mathcal{S}^{-\infty}(\mathfrak{A})$  accepte au moins un mot. Alors  $\mathfrak{A}$  accepte au moins un mot morphique  $x$  généré par des substitutions  $\sigma$  et  $\tau$  telles que  $\sigma, \tau \in \mathcal{S}^+$ .*

*Démonstration.*

D'après le corollaire 2.2.25, l'ensemble des suites directrices de mots infiniment désubstituables par  $\mathcal{S}$  acceptés par  $\mathfrak{A}$  est le langage d'un  $\omega$ -automate faible. Comme  $\mathcal{S}^{-\infty}(\mathfrak{A})$  accepte au moins un mot, son langage est non vide. Or, si un  $\omega$ -automate faible accepte au moins un mot, alors il accepte au moins un mot ultimement périodique. Donc  $\mathfrak{A}$  accepte un mot infiniment désubstituable dont la suite directrice est ultimement périodique.

On peut aussi déduire de la proposition 2.2.24 une propriété de décidabilité :

**Proposition 2.2.27.** *Le problème de l'acceptation d'un mot infiniment désubstituable par un  $\omega$ -automate faible est décidable :*

**Donnée :** un  $\omega$ -automate faible  $\mathfrak{A}$ , un ensemble fini de substitutions  $\mathcal{S}$

**Question :** est-ce que  $\mathfrak{A}$  accepte un mot infiniment désostituable par une suite  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$  ?

*Démonstration.*

D'après la proposition 2.2.24, le problème revient à décider l'existence d'une marche non-nilpotente dans  $\mathcal{S}^{-\infty}(\mathfrak{A})$ . Pour cela, il suffit de retirer les états de  $\mathcal{S}^{-\infty}(\mathfrak{A})$  dont le langage est vide (la vacuité du langage d'un  $\omega$ -automate faible étant décidable), puis de vérifier si, une fois cette étape finie, le langage de  $\mathcal{S}^{-\infty}(\mathfrak{A})$  est vide.

La proposition 2.2.27 ne suffit toutefois pas à résoudre la question des mots sturmiens. La caractérisation classique des mots sturmiens par désstitution infinie présentée dans le théorème 1.2.21 demande en effet une condition en plus, celle d'alternance infinie entre les types, ce qui demande plus de vérifications sur le langage de  $\mathcal{S}^{-\infty}(\mathfrak{A})$  que simplement s'assurer de l'existence d'une marche non-nilpotente. Cette condition supplémentaire est heureusement plutôt aisément vérifiable, et permet d'affirmer le résultat suivant, qui constitue le principal apport technique de cette thèse.

**Proposition 2.2.28.** *Le problème de l'acceptation d'un mot sturmien par un  $\omega$ -automate faible est décidable.*

**Donnée :** un  $\omega$ -automate faible  $\mathfrak{A}$

**Question :** est-ce que  $\mathfrak{A}$  accepte un mot sturmien ?

*Démonstration.*

Soit  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  le méta- $\omega$ -automate de désstitution de  $\mathfrak{A}$ , avec  $\mathcal{S}_{St}$  comme défini dans la définition 1.2.20. Alors d'après la caractérisation des mots sturmiens et la proposition 2.2.27, il existe un mot sturmien accepté par  $\mathfrak{A}$  si et seulement s'il existe un mot  $(\sigma_n)_{n \in \mathbb{N}}$  accepté par  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  alternant infiniment en type et admettant un calcul non-nilpotent. On commence par retirer les états de langage vide de  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  : tous les mots acceptés restants sont bien des suites directrices de mots acceptés par  $\mathfrak{A}$ . Pour vérifier la condition d'alternance de type, on calcule maintenant les composantes fortement connexes de  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  et on vérifie s'il en existe au moins une contenant deux transitions étiquetées par des substitutions de type différents. Le calcul des composantes fortement connexes et la vérification de l'existence d'un couple de transitions de types distincts sont bien décidables.

En l'occurrence, cette condition d'alternance infinie de type est modélisable par un  $\omega$ -automate de Büchi, un autre modèle d' $\omega$ -automate. J'étudierai ce modèle plus en détail dans la sous-section 2.3, et je n'utilise pour le moment que leur définition : je laisse le lecteur néophyte voir la définition 2.3.1. Il est facile alors de généraliser le cas des mots sturmiens à toute famille admettant une caractérisation S-adique modélisable par un  $\omega$ -automate de Büchi.

**Définition 2.2.29** (Ensemble  $X_{\mathfrak{R}}$ ). *Soit  $\mathcal{A}$  un alphabet fini, soit  $\mathcal{S}$  un ensemble fini de substitutions sur  $\mathcal{A}$ , et soit  $\mathfrak{R}$  un  $\omega$ -automate de Büchi sur l'alphabet  $\mathcal{S}$ . On définit alors  $X_{\mathfrak{R}}$  comme  $\{x \in \mathcal{A} \mid \exists (\sigma_n)_{n \in \mathbb{N}} \in \mathcal{L}(\mathfrak{R}), x \text{ est infiniment désostituable par } (\sigma_n)_{n \in \mathbb{N}}\}$ .*

**Proposition 2.2.30.** *Le problème suivant est décidable.*

**Donnée :** un  $\omega$ -automate faible  $\mathfrak{A}$ , un ensemble fini de substitutions  $\mathcal{S}$ , un  $\omega$ -automate de Büchi  $\mathfrak{R}$

**Question :** est-ce que  $\mathfrak{A}$  accepte un mot de  $X_{\mathfrak{R}}$  ?

*Démonstration.*

Le problème est équivalent au suivant :  $\mathcal{L}(\mathfrak{R}) \cap \mathcal{L}(\mathcal{S}^{-\infty}(\mathfrak{A})) \neq \emptyset$  ? L'intersection entre un  $\omega$ -automate de Büchi et un  $\omega$ -automate faible est un  $\omega$ -automate de Büchi qui peut être construit de manière effec-

tive, et la vacuité du langage d'un  $\omega$ -automate de Büchi est décidable.

L'intérêt de la proposition 2.2.30 est qu'il existe plusieurs familles de mots admettant une caractérisation par infinie désubstitution, avec des contraintes  $\omega$ -régulières. De plus, la proposition 2.2.30 donne une nouvelle raison pour motiver la recherche de caractérisations S-adiques. Parmi les familles déjà caractérisées, il existe par exemple les mots d'Arnoux-Rauzy [AR91] et les mots minimaux ternaires dendriques [GLL22]. Un mot infini  $x$  sur l'alphabet  $\{0, 1, 2\}$  est d'Arnoux-Rauzy s'il est récurrent, si  $p_x(n) = 2n + 1$  et si  $x$  admet exactement un facteur spécial gauche et droit de chaque longueur. La définition d'un mot minimal ternaire dendrique étant bien plus complexe et n'étant pas nécessaire à la compréhension de cette thèse, le lecteur intéressé pourra se référer à la très complète thèse de Gheeraert pour des explications [Ghe23a].

Dans le cas des mots d'Arnoux-Rauzy, la caractérisation S-adique donnée par l'article d'origine remplit les conditions d' $\omega$ -régularité, et l'algorithme développé plus tôt permet de conclure directement. La seule difficulté réside dans le passage de représentation S-adique à mot infiniment désubstituable. Cette transition est une conséquence de l'ubicroissance de la représentation S-adique fournie par Arnoux et Rauzy, comme explicité dans [Ric21].

**Corollaire 2.2.31.** *Le problème d'acceptation d'un mot d'Arnoux-Rauzy par un  $\omega$ -automate faible est décidable.*

**Donnée :** un  $\omega$ -automate faible  $\mathcal{A}$

**Question :** est-ce que  $\mathcal{A}$  accepte un mot d'Arnoux-Rauzy?

Le résultat pour les mots minimaux ternaires dendriques est plus compliqué à établir. La caractérisation d'origine ne remplit en effet pas les conditions de la proposition 2.2.30, car la caractérisation d'origine requiert une infinité de morphismes distincts. Pour pouvoir conclure en utilisant les outils développés ici, il faut donc borner le nombre de morphismes utilisés. La proposition suivante explique comment modifier la caractérisation d'origine pour revenir aux conditions nécessaires à la proposition 2.2.30.

**Proposition 2.2.32.** *Le problème de l'acceptation d'un mot minimal ternaire dendrique par un  $\omega$ -automate faible est décidable.*

**Donnée :** un  $\omega$ -automate faible  $\mathcal{A}$

**Question :** est-ce que  $\mathcal{A}$  accepte un mot minimal ternaire dendrique?

*Démonstration.*

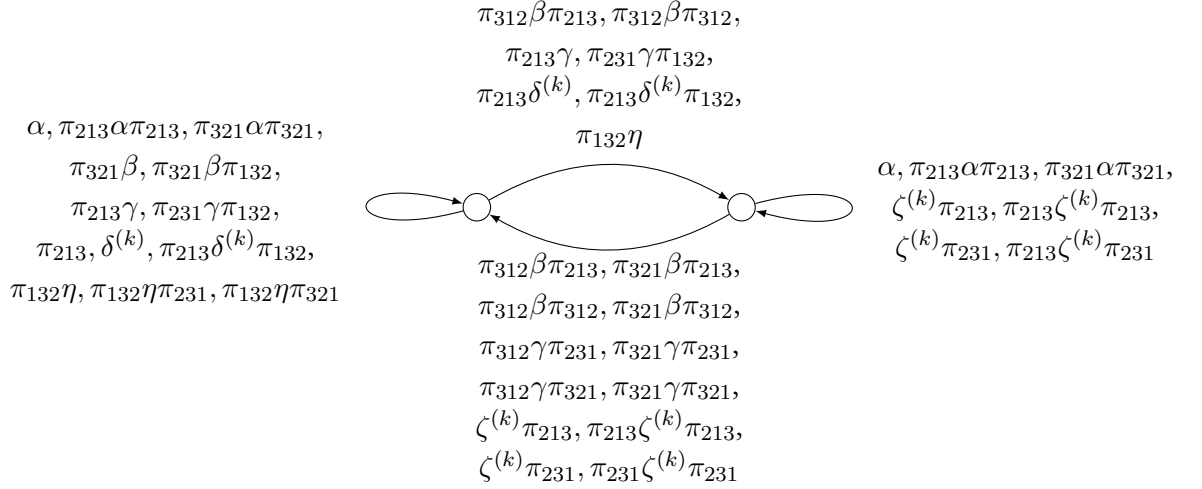
Dans l'article [GLL22], Gheeraert, Lejeune et Leroy proposent une caractérisation S-adique des mots minimaux ternaires dendriques à l'aide d'une infinité de morphismes. Ces représentations S-adiques étant ubicroissantes, elles remplissent bien les conditions nécessaires pour parler de mots infiniment désubstituables (voir [Ric21]). En reprenant leurs notations, ils utilisent les morphismes suivants (où  $k$  désigne un entier  $\geq 1$ ) :

$$\alpha : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 12, \\ 3 \mapsto 13 \end{cases}, \beta : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 12, \\ 3 \mapsto 132 \end{cases}, \gamma : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 12, \\ 3 \mapsto 123 \end{cases}$$

$$\delta^{(k)} : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 123^k, \\ 3 \mapsto 123^{k+1} \end{cases}, \zeta^{(k)} : \begin{cases} 1 \mapsto 13^k, \\ 2 \mapsto 12, \\ 3 \mapsto 13^{k+1} \end{cases}, \eta : \begin{cases} 1 \mapsto 13, \\ 2 \mapsto 12, \\ 3 \mapsto 123 \end{cases}$$

L'ensemble des suites directrices possibles est décrit par un  $\omega$ -automate faible étiqueté par  $\mathcal{S} =$

$\{\alpha, \beta, \gamma, \eta\} \cup \{\delta^{(k)}, \zeta^{(k)} \mid k \geq 1\}$ . Peut-on alors directement appliquer la proposition 2.2.30? Non, car la famille  $\mathcal{S}$  est une famille infinie de substitutions.



**FIGURE 2.II** – Caractérisation S-adique des mots minimaux ternaires dendriques de Gheeraert, Leroy et Lejeune : une représentation S-adique convenable décrit une marche alternant infiniment entre les deux sommets.

Il est toutefois possible de modifier la caractérisation de Gheeraert, Leroy et Lejeune : cette nouvelle caractérisation réduit le nombre de substitutions à un nombre fini en introduisant des lettres de travail. Ces lettres ne sont pas satisfaisantes pour une caractérisation S-adique propre<sup>3</sup>, mais elles nous permettront ici de faire fonctionner les algorithmes et de pouvoir appliquer la proposition 2.2.30.

La modification de la représentation S-adique de Gheeraert, Leroy et Lejeune est la suivante. Pour  $\delta^{(k)}$ , on introduit les trois substitutions suivantes :

$$\delta_d : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 124, \\ 3 \mapsto 1234 \end{cases}, \delta_p : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 2, \\ 3 \mapsto 3, \\ 4 \mapsto 43 \end{cases}, \delta_f : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 2, \\ 3 \mapsto 3, \\ 4 \mapsto 3 \end{cases}$$

On a alors :  $\forall w \in \{1, 2, 3\}^*, \delta^{(k)}(w) = \delta_f \circ \delta_p^{k-1} \circ \delta_d(w)$ .

De même, en posant

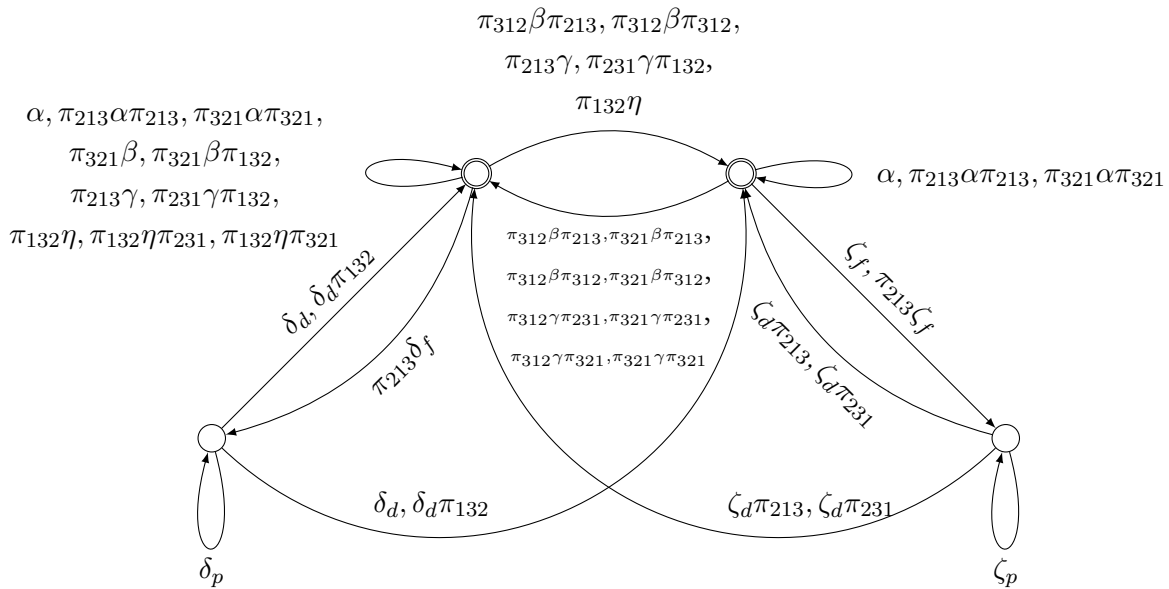
$$\zeta_d : \begin{cases} 1 \mapsto 14, \\ 2 \mapsto 12, \\ 3 \mapsto 134 \end{cases}, \zeta_p : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 2, \\ 3 \mapsto 3, \\ 4 \mapsto 43 \end{cases}, \zeta_f : \begin{cases} 1 \mapsto 1, \\ 2 \mapsto 2, \\ 3 \mapsto 3, \\ 4 \mapsto 3 \end{cases},$$

on arrive au résultat suivant :  $\forall w \in \{1, 2, 3\}^*, \zeta^{(k)}(w) = \zeta_f \circ \zeta_p^{k-1} \circ \zeta_d(w)$ .

3. Suivant la construction de Cassaigne (voir la Remark 3 de [BD14] ou la sous-section 4.2), tout mot admet une représentation S-adique en utilisant des lettres de travail.

En utilisant le découpage ci-dessus, la caractérisation S-adique proposée par Gheeraert, Leroy et Lejeune devient l' $\omega$ -automate de Büchi présenté dans la figure 2.12.

Les états doublement cerclés sont à la fois les états initiaux et les états acceptants. Les substitutions  $\pi$  représentent les permutations sur l'alphabet  $\{1, 2, 3\}$ . Un mot est ternaire minimal dendrique si, et seulement si, il est infiniment désubstituable par une suite de substitutions acceptée par cet  $\omega$ -automate de Büchi.



**FIGURE 2.12** – Nouvelle caractérisation S-adique des mots minimaux ternaires dendriques : on y ajoute deux états supplémentaires en bas à gauche et à droite, qui servent à décomposer les deux substitutions problématiques  $\delta^{(k)}$  et  $\zeta^{(k)}$ .

Il est primordial de noter que toutes les substitutions proposées ici n'ont pas le même alphabet. Certaines ont pour alphabet  $\{1, 2, 3\}$  (comme dans l'article d'origine), tandis que d'autres ont pour alphabet  $\{1, 2, 3, 4\}$ . La définition de désubstitution s'applique toujours : si on cherche à désubstituer un  $\omega$ -automate  $\mathfrak{B}$  sur l'alphabet  $\{1, 2, 3\}$  et qu'on cherche à le désubstituer par exemple par  $\delta_d$ , on ne pourra pas créer pas d'arête étiquetée par 2 ou 3. La conclusion sera alors que  $\mathfrak{B}$  conduira à une impasse pour accepter un mot minimal ternaire dendrique. J'insiste sur le fait qu'il ne faut pas étendre toutes les substitutions sur l'alphabet  $\{1, 2, 3, 4\}$  : si on le faisait, certaines substitutions devraient effacer la lettre 4, ce qui casserait plusieurs preuves précédentes, en particulier celles reposant sur la topologie.

J'ai donc fourni ici plusieurs résultats de décidabilité pour plusieurs familles de mots. L'étude des  $\omega$ -automates faibles, si elle peut paraître restreinte, permet en fait d'établir des résultats dans plusieurs contextes : l'un d'entre eux est celui des systèmes dynamiques. Comme rappelé dans le théorème 2.2.6, les  $\omega$ -automates faibles permettent en effet de caractériser les shifts sofiqes. La conséquence directe est alors la suivante :

**Proposition 2.2.33.** *La présence d'un mot infiniment désubstituable avec suite directrice  $\omega$ -régulière dans un shift sofique est décidable.*

**Donnée :** un ensemble fini de substitutions  $\mathcal{S}$ , un  $\omega$ -automate de Büchi  $\mathfrak{A}$  sur l'alphabet  $\mathcal{S}$  et un shift sofique  $\mathbb{S}$

**Question :** est-ce que  $X_{\mathfrak{A}} \cap \mathbb{S}$  est vide ?

Ces résultats fournissent donc des algorithmes de détection utiles pour l'étude algorithmique des sous-shifts so-

fiques. Toutefois, les complexités des algorithmes sont particulièrement hautes.

**Proposition 2.2.34.** *Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  un  $\omega$ -automate faible. Alors on peut décider si  $\mathfrak{A}$  accepte un mot sturmien en  $\mathcal{O}(|Q|^3 \times 2^{2 \times |Q|^2})$  opérations.*

*Démonstration.*

Comme on cherche à détecter un mot sturmien, je supposerai que  $\text{Card}(\mathcal{A}) = 2$ . La première opération de l'algorithme proposé ici consiste à construire  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$ . L'ensemble des états de  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  est inclus dans  $\mathfrak{S}(\mathfrak{A})$  : une borne grossière<sup>4</sup> de  $\text{Card}(\mathfrak{S}(\mathfrak{A}))$  est  $2^{2 \times |Q|^2}$ . La construction des transitions se fait en désubstituant : il faut faire 4 désubstitutions par sommet. Chaque désubstitution se fait en  $\mathcal{O}(|Q|^3)$  (voir la proposition 2.2.12). Donc au total, la construction des transitions de  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  se fait en  $\mathcal{O}(|Q|^3 \times 2^{2 \times |Q|^2})$ . Maintenant, il faut retirer tous les sommets ayant un langage vide. La vacuité du langage d'un  $\omega$ -automate faible se décide en temps linéaire en le nombre d'états. Donc retirer tous les états de langage vide de  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  se fait en temps  $\mathcal{O}(|Q| \times 2^{2 \times |Q|^2})$ . Maintenant, il s'agit de trouver un chemin alternant infiniment en type : on commence alors à calculer les composantes fortement connexes. Cette opération se fait en temps linéaire<sup>5</sup> sur le nombre d'états de  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$ , donc en  $\mathcal{O}(2^{2 \times |Q|^2})$ . Enfin, on vérifie s'il existe au moins une composante fortement connexe contenant au moins deux arêtes de types différents : sur chaque composante fortement connexe, cela se fait en temps linéaire sur le nombre d'arêtes. En sommant sur toutes les composantes fortement connexes, cela donne  $4 \times 2^{2 \times |Q|^2}$  opérations.

Finalement, c'est la construction du méta- $\omega$ -automate  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A})$  qui coûte le plus de temps.

### 2.2.5 Un phénomène de totalité

Une implémentation de l'algorithme précédent a été élaborée pendant cette thèse sous Sage par El Haq Ouerfelli lors de son stage de licence. L'un des objectifs était de tester la complexité en pratique des algorithmes proposés. Lors de ces tests, c'est en fait un autre résultat qui s'est imposé concernant la construction du méta- $\omega$ -automate dans le cas sturmien. Il existe, dans la littérature, plusieurs résultats clarifiant bien que les facteurs d'un mot sturmien ne forment pas un langage régulier [Fog02]. Dans cette sous-section, j'étends ce résultat au cas des  $\omega$ -automates faibles en le sens suivant : si un  $\omega$ -automate faible accepte un mot sturmien, alors cet  $\omega$ -automate faible accepte l'image du shift plein par un morphisme sturmien. Un  $\omega$ -automate faible acceptant un mot sturmien contiendra donc énormément de mots très proches, ce qui en fait un outil trop grossier pour étudier précisément les mots sturmiens.

**Proposition 2.2.35.** *Soit  $S_{St}$  les substitutions sturmiennes élémentaires (voir la définition 1.2.20), et soit  $\mathfrak{A}$  un  $\omega$ -automate faible. Si  $\mathfrak{A}$  accepte au moins un mot sturmien, alors il existe  $\sigma \in S_{St}^*$  tel que  $\sigma(\mathcal{A}^{\mathbb{N}}) \subseteq \mathcal{L}(\mathfrak{A})$ .*

Ce résultat est équivalent à la présence d'un sommet dans le méta- $\omega$ -automate qui accepte tout mot. Afin de faciliter les notations, j'introduis la définition suivante :

**Définition 2.2.36** (Totalité). *Un  $\omega$ -automate faible est total si  $\mathcal{L}(\mathfrak{A}) = \mathcal{A}^{\mathbb{N}}$ .*

**Proposition 2.2.37.** *Si  $\mathfrak{A}$  est un  $\omega$ -automate faible total, pour toute substitution  $\sigma$ ,  $\sigma^{-1}(\mathfrak{A})$  est un  $\omega$ -automate faible total.*

4. on pourrait améliorer cette borne en identifiant les  $\omega$ -automates faibles qui ont le même langage. Mais il n'y a pas de borne efficace dans ce cas, le meilleur résultat à l'heure actuelle est une fonction de récurrence difficile à calculer [BDS12].

5. Par exemple, avec l'algorithme de Tarjan.

*Démonstration.*

Il suffit de remarquer que  $\mathcal{L}(\sigma^{-1}(\mathfrak{A})) = \{w \in \mathcal{A}^{\mathbb{N}} \mid \sigma(w) \in \mathcal{L}(\mathfrak{A})\} = \{w \in \mathcal{A}^{\mathbb{N}} \mid \sigma(w) \in \mathcal{A}^{\mathbb{N}}\} = \mathcal{A}^{\mathbb{N}}$ .

Pour démontrer la proposition 2.2.35, j'introduis une propriété technique, appelée la propriété (H).

**Définition 2.2.38** (Propriété (H)). *Soit  $\mathfrak{A}$  un  $\omega$ -automate faible sur l'alphabet  $\mathcal{A} = \{0, 1\}$ . Un état  $q$  de  $\mathfrak{A}$  a la propriété (H) si*

$$\exists s \in Q, q \xrightarrow{0} s \rightarrow^{\omega} \dots \text{ est une marche sur } \mathfrak{A} \iff \exists t \in Q, q \xrightarrow{1} t \rightarrow^{\omega} \dots \text{ est une marche sur } \mathfrak{A}$$

Si tous les états d'un  $\omega$ -automate faible  $\mathfrak{A}$  ont la propriété (H), il y a alors deux possibilités : soit il n'y a pas de mots infinis étiquetant une marche commençant par un état initial, et alors le langage de  $\mathfrak{A}$  est vide; soit  $\mathfrak{A}$  est un  $\omega$ -automate faible total.

**Lemme 2.2.39.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate faible, et soit  $\phi \in \mathcal{S}_{St}^*$  commençant par  $L_0$ , terminant par  $L_1$ , et tel que  $\phi^{-1}(\mathfrak{A}) = \mathfrak{A}$ . Alors tout état de  $\mathfrak{A}$  a la propriété (H).*

*Démonstration.*

Soit  $\mathfrak{A} = (\{0, 1\}, Q, I, T)$  et soit  $q \in Q$ . D'abord, supposons que  $q \xrightarrow{0} s \rightarrow^{\omega} \dots$  est une marche sur  $\mathfrak{A}$ . Alors  $q \xrightarrow{0} s$  est aussi une transition de  $\phi^{-1}(\mathfrak{A})$ . Donc  $q \xrightarrow{\phi(0)} s$  est une marche sur  $\mathfrak{A}$ . Comme  $\phi$  termine par la substitution  $L_1, \phi(1) \sqsubseteq_p \phi(0)$ . Donc  $q \xrightarrow{\phi(1)} u \xrightarrow{m} s \rightarrow^{\omega} \dots$  est une marche sur  $\mathfrak{A}$ , avec un certain état  $u \in Q$  et  $\phi(0) = \phi(1)m$ . En utilisant de nouveau  $\mathfrak{A} = \phi^{-1}(\mathfrak{A})$ , on obtient donc que  $q \xrightarrow{1} u \xrightarrow{m} s \rightarrow^{\omega} \dots$  est une marche sur  $\mathfrak{A}$ .

Réciproquement, si  $q \xrightarrow{1} t \rightarrow^{\omega} \dots$  est une marche sur  $\mathfrak{A} = \phi^{-1}(\mathfrak{A})$ , alors  $q \xrightarrow{\phi(1)} t \rightarrow^{\omega} \dots$  est aussi une marche sur  $\mathfrak{A}$ . Parce que  $\phi$  commence par  $L_0$ , on sait alors que  $\phi(1) = 0m$  pour un certain mot  $m \in \{0, 1\}^*$ . Donc la dernière marche peut être écrite comme  $q \xrightarrow{0} u \xrightarrow{m} t \rightarrow^{\omega} \dots$

Maintenant que cette première étape technique est faite, je m'attaque à la preuve de la proposition 2.2.35.

*Démonstration de la proposition 2.2.35.*

On considère l'application amnésique  $\eta : (\mathcal{A}, Q, I, T) \mapsto (\mathcal{A}, Q, Q, T)$  qui prend un  $\omega$ -automate faible et rend tous ses états initiaux.

Soit  $x$  un mot sturmien accepté par  $\mathfrak{A}$ . Alors  $\eta(\mathfrak{A})$  accepte aussi  $x$ , et  $\mathcal{L}(\eta(\mathfrak{A}))$  est un shift sofique. Donc la clôture de l'orbite de  $x$  est aussi contenue dans  $\mathcal{L}(\eta(\mathfrak{A}))$ . Soit  $\chi(x)$  le mot sturmien caractéristique associé à  $x$  (voir [PR12] et la définition 1.2.19) :  $\chi(x)$  est dans la clôture de l'orbite de  $x$ , et est donc accepté par  $\eta(\mathfrak{A})$ . Donc la représentation S-adique d'Arnoux  $(\sigma_n)_{n \in \mathbb{N}}$  de  $\chi(x)$  est une suite directrice d'un mot sturmien accepté par  $\eta(\mathfrak{A})$  : comme  $\chi(x)$  est un mot caractéristique, il représente l'orbite de zéro du point de vue de la rotation de cercle (voir [PR12]). Combiné avec le corollaire 2.7 de [JP02], on en conclut que  $(\sigma_n)_{n \in \mathbb{N}}$  est exclusivement constitué des substitutions  $L_0$  et  $L_1$ . Par le lemme des tiroirs, il existe donc un  $\omega$ -automate faible  $\mathfrak{B}$  qui apparaît infiniment souvent dans la suite  $(\sigma_{\llbracket 0, n \rrbracket}^{-1}(\eta(\mathfrak{A})))_{n \in \mathbb{N}} \in \mathfrak{S}(\eta(\mathfrak{A}))^{\mathbb{N}}$ . On peut donc trouver une substitution  $\tau$  telle que  $\mathfrak{B} = \tau^{-1}(\mathfrak{B})$  avec  $\tau \in \{L_0, L_1\}^* \setminus L_0^* \cup L_1^*$ . Comme  $\tau$  contient à la fois  $L_0$  et  $L_1$ , il y a alors deux cas :

1.  $L_1 L_0 \preceq_f \tau$  : on peut alors écrire  $\tau = p_{\tau} L_1 L_0 s_{\tau}$ . Soit alors  $\mathfrak{B}' = (p_{\tau} \circ L_1)^{-1}(\mathfrak{B})$  et  $\tau' = L_0 \circ s_{\tau} \circ p_{\tau} \circ L_1$  : on a alors que  $\tau'^{-1}(\mathfrak{B}') = \mathfrak{B}'$ .

2.  $L_1 L_0 \not\leq_f \tau$  : alors  $\tau$  commence par  $L_0$  et termine par  $L_1$ .

Dans les deux cas, j'en suis revenu à étudier le cas où  $\tau$  commence par un  $L_0$  et termine par un  $L_1$ .

Maintenant, on applique le lemme 2.2.39 à  $\mathfrak{B}$ . On en déduit alors que tout état de  $\mathfrak{B}$  a la propriété (H).

On peut alors écrire  $\mathfrak{B}$  sous la forme  $\psi^{-1}(\eta(\mathfrak{A}))$  pour un certain morphisme  $\psi$ . Comme l'application  $\eta$  ne modifie pas les transitions de l' $\omega$ -automate en argument, on en déduit que tout état de  $\psi^{-1}(\mathfrak{A})$  a aussi la propriété (H). Or, par hypothèse,  $\psi^{-1}(\mathfrak{A})$  accepte un mot infini : donc c'est un  $\omega$ -automate faible total.

Ce résultat justifie que pour tout  $\omega$ -automate faible  $\mathfrak{A}$ , soit son méta- $\omega$ -automate de désubstitution associé aux mots sturmiens n'accepte aucune alternance infinie de type, soit il possède un sommet qui accepte tout mot infini. En particulier, le méta- $\omega$ -automate de désubstitution possède un état puits. Une autre manière de considérer le résultat précédent est la suivante :

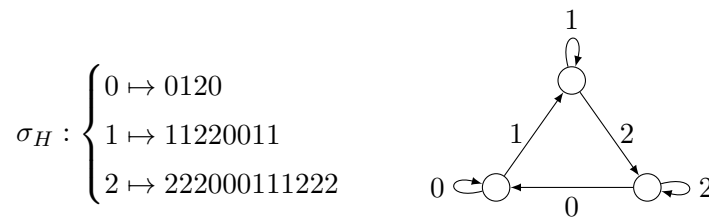
**Corollaire 2.2.40.** *Soit  $\sigma_f : 0 \mapsto 01, 1 \mapsto 0$  la substitution de Fibonacci, et  $f$  le mot purement substitutif généré par  $\sigma_f$  (il s'agit du mot de Fibonacci). Soit  $\mathfrak{A}$  un  $\omega$ -automate faible qui accepte  $f$ . Alors il existe  $n \in \mathbb{N}$  tel que  $\sigma_f^{-n}(\mathfrak{A})$  est un  $\omega$ -automate faible total.*

En termes dynamiques, le corollaire 2.2.40 se réécrit ainsi :

**Corollaire 2.2.41.** *Un sous-shift sofique  $\mathbb{S}$  contient le mot de Fibonacci  $f$  si, et seulement si,  $\mathbb{S}$  contient l'un des ensembles  $\sigma_f^n(\mathcal{A}^{\mathbb{N}})$ .*

Un résultat folklorique établit que si un sous-shift sofique  $\mathbb{S}$  contient un mot apériodique (comme le mot de Fibonacci), alors il existe une substitution  $\tau$  telle que  $\mathbb{S}$  contient l'ensemble  $\tau(\mathcal{A}^{\mathbb{N}})$ . Comme le mot de Fibonacci est un mot sturmien, Berstel et Séebold ont démontré [Lot02] que  $\tau$  devait être une substitution sturmiennne. Ce nouveau résultat précise qu'on peut toujours supposer qu'il s'agit d'une puissance de  $\sigma_f$ .

Les résultats que j'ai prouvés ici concernent tout particulièrement les mots sturmiens : on peut étendre la question à d'autres représentations S-adiques. Sous quelles conditions apparaît ce domaine de totalité? Déjà, il n'apparaît pas tout le temps : la figure 2.13 présente un  $\omega$ -automate faible  $\mathfrak{A}_H$  et une substitution  $\sigma_H$  tel que  $\sigma_H^{-1}(\mathfrak{A}_H) = \mathfrak{A}_H$ , mais  $\mathfrak{A}_H$  n'accepte pas tout mot infini : donc son méta- $\omega$ -automate de désubstitution associé à l'ensemble de substitutions  $\{\sigma_H\}$  ne contient pas d' $\omega$ -automate total.



**FIGURE 2.13** – Exemple d' $\omega$ -automate faible non total se désubstituant en lui-même.

Une caractérisation des ensembles de substitutions  $\mathcal{S}$  entraînant ce phénomène de totalité serait intéressante, pour mieux comprendre quels sont les arguments clés de la preuve précédente. La difficulté est néanmoins que, dans le cas des mots sturmiens, j'ai utilisé activement l'alternance infinie de type : il faudrait donc, pour pouvoir généraliser, prendre en compte ce genre de condition  $\omega$ -régulière.



## 2.2.6 Applications de ces algorithmes

### Application à la dynamique symbolique

Originellement, mon intérêt pour le problème de la détection d'un mot sturmien dans le langage d'un  $\omega$ -automate faible provient de questions soulevées pendant mon stage de master sous la direction de Berthé. Le contexte était alors l'étude de décidabilité de propriétés des représentations S-adiques elles-mêmes. Parmi les propriétés les plus étudiées du modèle substitutif se trouve la notion de primitivité.

**Définition 2.2.42** (Primitivité faible). *Soit  $(\sigma_n)_{n \in \mathbb{N}}$  une suite de substitutions sur un alphabet  $\mathcal{A}$ . La suite est dite faiblement primitive si pour tout  $n \in \mathbb{N}$ , il existe  $r \geq 0$  tel que  $\sigma_{\llbracket n, n+r \rrbracket}$  est une substitution strictement positive : pour tous  $a, b \in \mathcal{A}$ ,  $b \preceq_f \sigma_{\llbracket n, n+r \rrbracket}(a)$ .*

**Théorème 2.2.43** ([BD14]). *Un mot  $x$  est uniformément récurrent si, et seulement si, il admet une représentation S-adique faiblement primitive.*

On se donne un ensemble fini de substitutions  $\mathcal{S}$  sur un alphabet  $\mathcal{A}$ . L'objectif est d'étudier les propriétés des représentations S-adiques utilisant les éléments de  $\mathcal{S}$ . J'avais alors posé la définition suivante :

**Définition 2.2.44** ( $\mathcal{F}$ -primitivité). *Soit  $\mathcal{S} = \{s_i\}_{0 \leq i \leq k-1}$  une famille de substitutions. Soit  $\mathcal{F} \subseteq \llbracket 0, k-1 \rrbracket^{\mathbb{N}}$  une famille de suites. On dit que  $\mathcal{S}$  est  $\mathcal{F}$ -primitive si, pour toute suite  $(i_n)_{n \in \mathbb{N}} \in \mathcal{F}$ , la suite  $(s_{i_n})_{n \in \mathbb{N}}$  est faiblement primitive.*

Les méthodes développées plus tôt dans ce chapitre permettent d'étudier les notions que j'avais définies lors de ce stage, en particulier dans le cas où  $\mathcal{F}$  est l'ensemble des mots sturmiens. Il est en effet possible d'établir le théorème suivant :

**Définition 2.2.45** (Sturm-primitivité). *Un couple de substitutions  $\{s_0, s_1\}$  est dit Sturm-primitif si pour tout mot sturmien  $(x_n)_{n \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$ , la suite  $(s_{x_n})_{n \in \mathbb{N}}$  est faiblement primitive.*

**Proposition 2.2.46.** *La Sturm-primitivité est décidable.*

**Donnée :** *un couple de substitutions  $\{s_0, s_1\}$  sur un alphabet  $\mathcal{A}$*

**Question :** *est-ce que  $\{s_0, s_1\}$  est Sturm-primitive ?*

Pour prouver la proposition 2.2.46, je vais introduire plusieurs notations et notions standard en combinatoire des mots.

**Définition 2.2.47** (Matrice d'incidence). *Soit  $\sigma$  une substitution sur  $\mathcal{A}$ . On suppose que  $\mathcal{A} = \llbracket 1, m \rrbracket$  quitte à faire une bijection. On pose alors  $M_\sigma \in \mathcal{M}_m(\mathbb{N})$  la matrice carrée d'incidence de  $\sigma$  définie comme  $M_{i,j} = |\sigma(j)|_i$ .*

Pour rappel, pour  $\sigma$  et  $\tau$  deux substitutions sur un même alphabet  $\mathcal{A}$ , on a  $M_{\sigma \circ \tau} = M_\sigma \times M_\tau$ .

*Démonstration de la proposition 2.2.46.*

Soit  $m = \text{Card}(\mathcal{A})$ . On commence par construire un  $\omega$ -automate faible noté  $\mathfrak{A}_\mathcal{S}$ . Les états de  $\mathfrak{A}_\mathcal{S}$  sont les éléments de  $\mathcal{M}_m(\{0, 1\})$ . On définit ensuite l'opérateur  $\delta : 0 \mapsto 0, n \neq 0 \mapsto 1$  et on l'étend coefficient par coefficient aux matrices. On définit alors les transitions de  $\mathfrak{A}_\mathcal{S}$  de la manière suivante :  $M \xrightarrow{\sigma_i} \delta(M \times M_{\sigma_i})$  pour  $i \in \{0, 1\}$ . Le seul état initial de  $\mathfrak{A}_\mathcal{S}$  est la matrice identité  $I_m$ . Alors  $\mathcal{S}$  est Sturm-primitive si, et seulement si, toute marche sur  $\mathfrak{A}_\mathcal{S}$  étiquetée par un mot sturmien passe par l'état  $J$ , correspondant à la matrice dont tous les coefficients valent 1. Pour décider cette nouvelle caractérisation, on modifie  $\mathfrak{A}_\mathcal{S}$  en retirant l'état  $J$  : s'il existe une marche

étiquetée par un mot sturmien dans  $\mathfrak{A}_S \setminus \{J\}$ , alors  $\mathcal{S}$  n'est pas Sturm-primitif. Sinon,  $\mathcal{S}$  est Sturm-primitif. La proposition 2.2.28 permet de conclure.

### Application à la combinatoire des mots

Une notion classique de la combinatoire des mots est celle du codage. La question du codage d'un mot est étroitement liée à celle de la désubstitution de ce mot, et il est donc important d'avoir une compréhension des codages possibles. En l'appliquant à la famille des mots sturmiens, le codage s'écrit sous la forme suivante :

**Définition 2.2.48** (Problème du codage d'un mot sturmien). *Le problème du codage d'un mot sturmien est le suivant :*

**Donnée :** un nombre fini de mots  $m_1, \dots, m_k$

**Question :** est-ce que  $\left(\bigcup_{i=1}^k m_i\right)^{\mathbb{N}}$  contient un mot sturmien ?

Il est possible, quand  $k = 2$ , d'apporter une réponse complètement combinatoire à ce problème. Le résultat suivant est une conséquence d'un résultat de Séébold [Séé18] : j'en redonne une preuve rapide et combinatoire.

**Proposition 2.2.49.** *Soient  $u, v \in \{0, 1\}^+$ , et soit  $E : 0 \mapsto 1, 1 \mapsto 0$  la substitution d'échange. Alors  $(u + v)^{\mathbb{N}}$  contient un mot sturmien si, et seulement si, il existe  $\sigma \in (\mathcal{S}_{St} \cup \{E\})^*$  tel que  $\{u, v\} = \{\sigma(0), \sigma(1)\}$ .*

*Démonstration.*

D'abord, si  $\{u, v\} = \{\sigma(0), \sigma(1)\}$ , alors  $\sigma$  étant une substitution sturmiennne,  $\sigma(x_{\text{Fib}})$  est un mot sturmien (avec  $x_{\text{Fib}}$  le mot de Fibonacci) et est un élément de  $(u + v)^{\mathbb{N}}$ .

Maintenant, je prouve la réciproque : supposons que  $(u + v)^{\mathbb{N}}$  contienne un mot sturmien  $x$ . Alors je définis  $\sigma : 0 \mapsto u, 1 \mapsto v$  : il existe donc  $x' \in \{0, 1\}^{\mathbb{N}}$  tel que  $x = \sigma(x')$ .

Dans un premier temps, je suppose aussi que  $u$  et  $v$  ne sont ni préfixe ni suffixe l'un de l'autre.

J'affirme alors que  $x'$  est sturmien. En premier lieu, comme  $x$  n'est pas ultimement périodique,  $x'$  n'est pas ultimement périodique. Donc  $x'$  admet au moins un facteur spécial à gauche pour toute longueur.

Montrons que  $x'$  admet *exactement* un facteur spécial à gauche pour toute longueur. D'abord, comme  $u$  et  $v$  ne sont ni préfixe ni suffixe l'un de l'autre, ils admettent un suffixe strict commun (potentiellement vide)  $s$ , et quitte à étudier le cas symétrique, on peut écrire  $u = u'0s$  et  $v = v'1s$ . Soit  $f$  un facteur spécial à gauche de  $x'$  : alors  $0f$  et  $1f$  sont des facteurs de  $x'$ , donc leurs images par  $\sigma$ ,  $u'0z\sigma(f)$  et  $v'1z\sigma(f)$ , sont des facteurs de  $x$ . Donc  $s\sigma(f)$  est un facteur spécial à gauche de  $x$ . Si  $x'$  avait deux facteurs spéciaux différents  $f$  et  $g$  de même longueur,  $s\sigma(f)$  et  $s\sigma(g)$  seraient deux facteurs spéciaux de  $x$  (potentiellement de longueurs distinctes). Mais, en identifiant  $p$  le plus long préfixe commun de  $f$  et  $g$ , on a  $f = p0f'$  et  $g = p1g'$ , quitte à étudier le cas symétrique : alors  $s\sigma(p)\sigma(0) = s\sigma(p)u$  et  $s\sigma(p)\sigma(1) = s\sigma(p)v$  sont deux facteurs spéciaux à gauche de  $x$ ; en notant  $q$  le plus long préfixe commun de  $u = q0u''$  et  $v = q1v''$ , quitte à étudier le cas symétrique, on en conclut enfin que  $s\sigma(p)q1$  et  $s\sigma(p)q2$  sont deux facteurs spéciaux gauches de  $x$  de même longueur et distincts, ce qui est absurde pour un mot sturmien. Donc  $x'$  est un mot sturmien. On peut alors appliquer les résultats de Berstel et Séébold [Lot02] :  $\sigma$  est une substitution localement sturmiennne<sup>6</sup>, donc sturmiennne. D'après le théorème 2.3.7 de Berstel et Séébold,  $\sigma$  est alors un élément de  $(\mathcal{S}_{St} \cup \{E\})^*$ . Mais si  $\sigma \notin \{id, E\}$ , on peut montrer par récurrence sur la longueur de  $\sigma$  que  $\sigma(0)$  ou  $\sigma(1)$  est un préfixe ou un suffixe de l'autre : donc, dans notre sous-cas,  $\sigma = id$  ou  $\sigma = E$ , et  $\{u, v\} = \{0, 1\}$ .

6. C'est-à-dire qu'il existe au moins un mot sturmien dont le seul antécédent par  $\sigma$  est sturmien.

Maintenant, je relâche l'hypothèse sur  $u$  et  $v$ , et prouve le cas général par récurrence forte sur  $|u| + |v|$ . D'abord, si  $|u| + |v| = 2$ , comme  $u$  et  $v$  sont non vides, on a bien  $\{u, v\} = \{0, 1\}$ . Puis, supposons le résultat vrai jusqu'au rang  $|u| + |v| = n$ , et étudions le cas  $|u| + |v| = n + 1$ . Alors  $\{u, v\} \neq \{0, 1\}$ , et donc, par le raisonnement précédent, l'un de  $u$  ou de  $v$  est un préfixe ou un suffixe de l'autre. Je prouve le cas où  $u$  est un préfixe de  $v$ , les trois autres cas se traitent de manière analogue. Si  $u$  est un préfixe de  $v$ , alors  $v = uv'$ , et  $(u + v')^{\mathbb{N}}$  contient un mot sturmien. On peut appliquer l'hypothèse de récurrence forte, et on trouve  $\sigma \in (\mathcal{S}_{St} \cup \{E\})^*$  tel que  $u = \sigma(0)$  et  $v' = \sigma(1)$  : donc  $v = \sigma(0)\sigma(1) = \sigma(01) = \sigma \circ L_0(1)$ , et  $u = \sigma(0) = \sigma \circ L_0(0)$ . Donc on trouve bien  $\sigma' = \sigma \circ L_0$  tel que  $\{u, v\} = \{\sigma'(0), \sigma'(1)\}$ .

Dans le cas général, les méthodes proposées dans ce chapitre permettent d'obtenir un résultat de décidabilité, peu importe le nombre de mots avec lesquels coder.

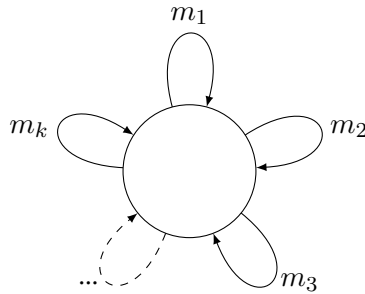
**Proposition 2.2.50.** *Le problème du codage d'un mot sturmien est décidable.*

**Donnée :**  $k$  un entier strictement positif,  $m_1, \dots, m_k$   $k$  mots finis

**Question :** est-ce que  $\left(\sum_{j=1}^k m_j\right)^{\infty}$  contient un mot sturmien ?

*Démonstration.*

Il suffit de remarquer que  $\left(\sum_{j=1}^k m_j\right)^{\infty}$  est le langage d'un  $\omega$ -automate faible  $\mathfrak{A}_m$  : il s'agit du langage d'un  $\omega$ -automate en fleur, avec un état central d'où partent  $k$  boucles étiquetées par les mots  $m_i$ . Le problème revient donc à savoir si  $\mathfrak{A}_m$  accepte un mot sturmien, ce qui est un problème décidable par la proposition 2.2.28.



**FIGURE 2.14** – Exemple d' $\omega$ -automate en fleur.

**Exemple 2.2.51.** Soit le triplet  $T = (01, 0010, 1001)$ . Ce triplet code le mot de Fibonacci : il suffit de remarquer que  $(L_0L_1)^{-3}(\mathfrak{A}_T)$  est un  $\omega$ -automate faible total, où  $\mathfrak{A}_T$  désigne l' $\omega$ -automate en fleur de  $T$ .

On remarque aussi que, pour tout  $\sigma \in \mathcal{S}_{St}^*$ ,  $\sigma(0)$  et  $\sigma(1)$  ne peuvent pas être tous les deux de longueurs paires. Donc, en particulier dans notre exemple, il n'existe pas de  $\sigma \in \mathcal{S}_{St}$  tel que  $\sigma(0), \sigma(1) \in T^+$  : les trois éléments de  $T$  sont nécessaires pour former un mot sturmien, et on ne peut pas les recombinaisonner en un couple de mots qui codera aussi un mot sturmien.

Ce résultat a néanmoins des limites. Il ne fournit en effet pas de caractérisation combinatoire des familles  $(m_i)_{1 \leq i \leq k}$  qui codent un mot sturmien pour  $k \geq 3$ . De manière générale, la recherche d'une telle caractérisation semble parti-

culièrement compliquée, car elle consiste en une généralisation des résultats de Berstel et Séébold sur les morphismes sturmiens au cas de morphismes provenant d'un alphabet plus grand, et on perd alors toute propriété de stabilité par composition. Une propriété toutefois intéressante est que la construction précédente fournit les suites directrices constructibles avec la famille  $(m_i)_{1 \leq i \leq k}$ . On peut donc obtenir une description des pentes constructibles. En particulier, comme le problème de l'inclusion entre deux  $\omega$ -automate de Büchi est un problème décidable [KV98], on peut savoir si  $(m_i)_{1 \leq i \leq k}$  peut coder tous les mots sturmiens. En réalité, ce problème peut être résolu plus simplement : si un  $\omega$ -automate faible accepte tous les mots sturmiens, par compacité de son langage, alors il accepte tous les mots équilibrés. Il suffira donc de vérifier que son méta- $\omega$ -automate de désubstitution accepte n'importe quel mot infini : et l'universalité d'un  $\omega$ -automate faible est un problème décidable.

### Application à la géométrie discrète

Les méthodes proposées plus haut ont aussi des applications à la géométrie discrète, et permettent de résoudre proprement une question naturelle sur les droites discrètes. Afin de garder la compatibilité avec les propositions précédentes, les résultats seront établis sur les demi-droites discrètes, car elles correspondront au cas des mots mono-infinis. Il est toutefois clair que toutes les méthodes établies précédemment tiennent dans le cas bi-infini, et permettront donc de résoudre le problème dans le cas des droites discrètes. Je vais d'abord introduire les définitions classiques de géométrie discrète. La définition suivante a été introduite par Reveillès [Rev91] :

**Définition 2.2.52** (Droite discrète). *Soit  $(a, b, \mu, \omega) \in \mathbb{R}^3 \times \mathbb{R}_+^*$ . La droite discrète associée est :*

$$D(a, b, \mu, \omega) = \{(x, y) \in \mathbb{Z}^2 \mid 0 \leq ax + by + \mu < \omega\}$$

Lorsque  $\omega = \|(a, b)\|_\infty$ , la droite  $D(a, b, \mu, \omega)$  est dite naïve; lorsque  $\omega = \|(a, b)\|_1$ , la droite  $D(a, b, \mu, \omega)$  est dite standard.

La pente de la droite est  $-\frac{b}{a}$  (et vaut  $\infty$  si  $a = 0$ ).

La littérature de la géométrie discrète s'est généralement concentrée sur le cas des droites de pente rationnelle (c.-à-d.  $a, b \in \mathbb{Z}$ ) [Deb95]. La définition ci-dessus en est une généralisation au cas d'une pente quelconque, qu'on retrouve par exemple dans [Ber09] ou [LD23]. Dans la suite de cette partie, je me concentrerai sur le cas naïf : les résultats annoncés fonctionneront aussi dans le cas standard. L'une des propriétés importantes des droites discrètes naïves est qu'elles sont 8-connexes :

**Définition 2.2.53** (Points 8-voisins, ensembles 8-connexes). *Soient  $u$  et  $v \in \mathbb{Z}^2$ . On dit que  $u$  et  $v$  sont 8-voisins si  $\|u - v\|_\infty \leq 1$ .*

*Soit  $E \subseteq \mathbb{Z}^2$ . On dit que  $E$  est 8-connexe si pour tous  $u, v \in E$ , il existe  $n \geq 0$  et  $(p_i)_{0 \leq i \leq n} \in E^{\mathbb{N}}$  tels que  $p_0 = u, p_n = v$  et tels que pour tout  $0 \leq i \leq n$ ,  $p_i$  et  $p_{i+1}$  sont 8-voisins.*

Avec ces définitions, on peut définir les notions de segment discret et de demi-droite discrète.

**Définition 2.2.54** (Segment et demi-droite discrètes). *Soit  $S \subseteq \mathbb{Z}^2$ . On dit que  $S$  est un segment discret si  $S$  est fini, 8-connexe et s'il existe une droite discrète  $D$  telle que  $S \subseteq D$ .*

*On dit que  $S$  est une demi-droite discrète si  $S$  est infini, 8-connexe et s'il existe une droite discrète  $D$  telle que  $S \subseteq D$  et  $S \neq D$ .*

Les définitions suivantes ne sont pas standard dans la littérature et sont propres à ce manuscrit.

**Définition 2.2.55** (Croissance et petite pente). *Soit  $S \subseteq \mathbb{Z}^2$  un segment, une droite ou une demi-droite discrète. On dit que  $S$  est croissant si pour  $u, v \in S$  deux points 8-voisins,  $u_1 < v_1 \implies u_0 \leq v_0$ . On dit que  $S$  est de petite pente si pour  $u, v \in S$  deux points 8-voisins,  $u_1 < v_1 \implies u_0 \neq v_0$ .*

Moins formellement, un segment est croissant de petite pente si c'est un morceau d'une droite de pente comprise entre 0 et 1.

Maintenant que les objets classiques (ou moins classiques) sont définis, j'introduis une nouvelle question : celle du recollement de segments. Il existe en géométrie discrète de nombreux problèmes algorithmiques, comme celui de la reconnaissance d'un segment discret [Deb95], celui de la transformée douce d'un segment [LAD22], ou celui de la préservation de la convexité des formes discrètes par l'homothétie [Tar+22]. S'il existe des approches utilisant la combinatoire des mots [Ber90], les méthodes algorithmiques les plus efficaces passent le plus souvent par des considérations arithmétiques plus directes. Les algorithmes proposés dans la sous-section précédente permettent néanmoins de résoudre des problèmes jusqu'alors non résolus (certes aussi non posés) par la géométrie discrète.

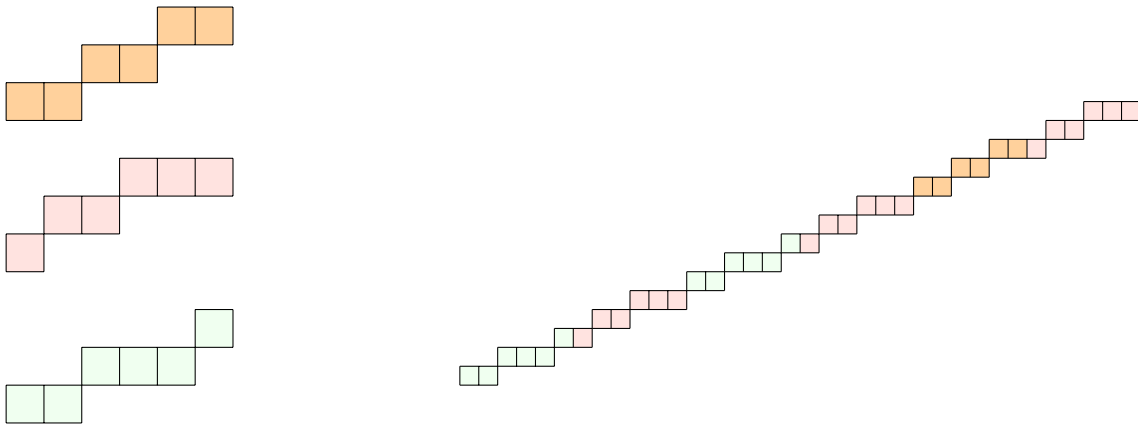
**Définition 2.2.56** (Équivalence sur  $\mathbb{Z}^2$ ). *On définit sur les parties finies de  $\mathbb{Z}^2$  la relation d'équivalence suivante :*

$$S \sim T \iff \exists v \in \mathbb{Z}^2, S = T + v$$

Cette relation d'équivalence traduit l'égalité de deux figures à décalage dans l'espace près, et peut se généraliser à toute dimension  $d \geq 2$ . Une classe d'équivalence de  $\sim$  peut être interprétée comme un *modèle* de l'objet étudié.

**Définition 2.2.57** (Recollement en une demi-droite discrète). *Soient  $k \geq 0$  et  $S_0, \dots, S_k$  des segments discrets. On dit que  $(S_i)_{0 \leq i \leq k}$  se recolle en une demi-droite discrète s'il existe une demi-droite discrète  $D$  telle qu'il existe une partition de  $D = \bigsqcup_{i \in I} R_i$  (c.-à-d.  $R_i \cap R_j \neq \emptyset$  ssi  $i = j$ ) telle que pour tout  $i \in I$ , il existe  $j \in \llbracket 0, k \rrbracket$  tel que  $R_i \sim S_j$ .*

Plus intuitivement, l'idée est qu'on dispose d'un nombre fini de modèles de segments  $S_0, \dots, S_k$  : ces segments se recollent en une demi-droite discrète si on peut les « concaténer » infiniment pour former une demi-droite discrète.



**FIGURE 2.15** – À gauche : trois modèles de segments ; à droite, un recollement partiel de ces segments pour former une droite de pente  $\varphi - 1$ , où  $\varphi$  est le nombre d'or.

**Remarque 2.2.58.** *Si une famille  $S_0, \dots, S_k$  peut se recoller en une demi-droite discrète, elle peut aussi se recoller en une droite discrète, et réciproquement.*

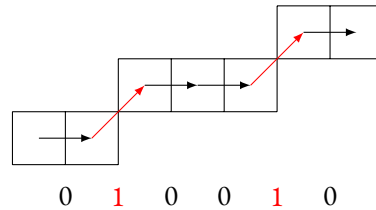
L'objectif de cette section est de montrer que le problème du recollement de segments discrets est décidable. Pour cela, je vais ramener ce problème de géométrie discrète à un problème de combinatoire des mots.

**Définition 2.2.59** (Énumération d'une demi-droite discrète). *Soit  $D$  une demi-droite discrète. L'énumération canonique des points de  $D$  est la fonction bijective  $\text{enum}_D : \mathbb{N} \rightarrow D$  telle que  $\text{enum}_D(n)$  et  $\text{enum}_D(n + 1)$  sont 8-voisins pour tout  $n \geq 0$ .*

**Définition 2.2.60** (Représentation d'une demi-droite discrète). Soit  $D$  une demi-droite discrète. On définit  $\text{rep}_D : \mathbb{N} \rightarrow \{0, 1\}$  de la manière suivante :

$$\text{rep}_D(n) = \begin{cases} 0 & \text{si } \|\text{enum}_D(n) - \text{enum}_D(n+1)\|_1 = 1, \\ 1 & \text{sinon.} \end{cases}$$

**Lemme 2.2.61** ([MH40; Ber09]). Un mot  $x \in \{0, 1\}^{\mathbb{N}}$  est équilibré de type 0<sup>7</sup> si, et seulement si, il existe une demi-droite discrète croissante de petite pente telle que  $x = (\text{rep}_D(n))_{n \in \mathbb{N}}$ .



**FIGURE 2.16** – Explication visuelle de la représentation d'un segment discret.

Cette représentation fait écho à la notion de mot mécanique [Lot02]. On a alors une représentation des objets de la géométrie discrète dans le contexte de la combinatoire des mots : une demi-droite discrète est un mot sturmien, et un segment discret est un facteur d'un mot sturmien, c.-à-d. un mot équilibré [Hei01]. À l'aide du lemme 2.2.61, je peux alors étudier le problème du recollement de segments discrets, et montrer sa décidabilité.

**Proposition 2.2.62.** Le problème du recollement de segments discrets dans le plan est décidable :

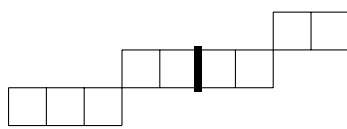
**Donnée :**  $k \geq 0$ , un ensemble fini de segments discrets  $S_0, \dots, S_k$

**Question :** est-ce que  $(S_i)_{0 \leq i \leq k}$  se recolle en une demi-droite discrète ?

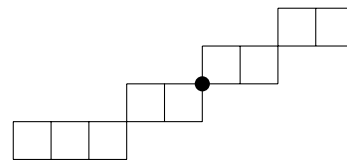
*Démonstration.*

Le recollement d'un segment discret croissant avec un segment qui ne l'est pas ne peut pas résulter en la création d'un segment discret : je supposerai donc, quitte à étudier le cas symétrique, que tous les segments  $S_0, \dots, S_k$  sont croissants. De même, si un segment discret est de petite pente et un autre non, il est impossible de les recoller ensemble pour former un segment discret : je supposerai donc, quitte à étudier le cas symétrique, que tous les segments sont de petite pente.

La première opération consiste à faire la traduction entre géométrie discrète et combinatoire des mots. Cette traduction est effectuée par une opération similaire à celle de la définition 2.2.60. Pour tout segment  $S_i$ , on dispose désormais d'un mot  $w_i$  sur l'alphabet  $\{0, 1\}$  pour le représenter. La dernière traduction à faire est la traduction du recollement de deux segments. On observe qu'il existe deux types de recollements possibles : le recollement par faces et le recollement par coins.



**FIGURE 2.17** – Recollement par faces.



**FIGURE 2.18** – Recollement par coins.

Pour deux segments  $S$  et  $T$ , je noterai  $S \diamond_f T$  le recollement par faces de  $S$  et  $T$ , et  $S \diamond_c T$  le recollement par coins de  $S$  et  $T$ . Pour traduire ces deux opérations sur les mots, une idée naturelle est la concaténation. On a alors les résultats suivants :

7. Comme pour les mots sturmiens, un mot équilibré est de type 0 s'il admet 00 comme facteur.

**Lemme 2.2.63.** *Pour tous segments discrets  $S$  et  $T$ , on a :*

$$\text{rep}_S \cdot 0 \cdot \text{rep}_T = \text{rep}_{S \circ_f T},$$

$$\text{rep}_S \cdot 1 \cdot \text{rep}_T = \text{rep}_{S \circ_c T}.$$

*Démonstration du lemme 2.2.63.*

Le mot  $\text{rep}_{S \circ_f T}$  admet pour préfixe  $\text{rep}_S$ . Après, le recollement par faces nous donne qu'entre le dernier point de  $S$  et le premier de  $T$ , la distance en norme 1 n'est que de 1 : donc  $\text{rep}_{S \circ_f T}$  admet pour préfixe  $\text{rep}_S \cdot 0$ . Le mot  $\text{rep}_{S \circ_f T}$  admet pour suffixe  $\text{rep}_T$ . Pour des raisons de longueur, on a donc bien  $\text{rep}_{S \circ_f T} = \text{rep}_S \cdot 0 \cdot \text{rep}_T$ .

Le cas du recollement par coins se traite de manière analogue.

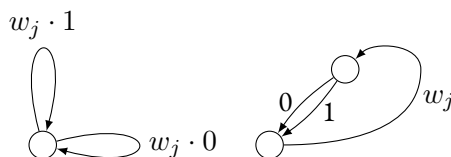
Il faut donc rajouter des 0 et des 1 à la fin des mots  $w_j$ . Soit alors la famille  $(w_0 \cdot 0, w_1 \cdot 0, \dots, w_k \cdot 0, w_0 \cdot 1, w_1 \cdot 1, \dots, w_k \cdot 1)$ . Le lemme 2.2.63 donne alors par induction que  $(S_i)_{0 \leq i \leq k}$  se recolle en une droite discrète bornée à gauche si, et seulement si,  $(\sum_{i=0}^k w_i \cdot 0 + w_i \cdot 1)^{\mathbb{N}}$  contient un mot équilibré. La proposition 2.2.50 règle le cas des mots sturmiens : il est en fait possible de l'étendre au cas des mots équilibrés, en remarquant que les mots équilibrés admettent des représentations S-adiques d'Arnoux sans alternance infinie de type, ce qui permet de conclure.

Maintenant que le problème a été montré décidable, je vais donner une idée de la complexité atteinte par l'algorithme proposé.

**Corollaire 2.2.64.** *Pour  $(S_i)_{0 \leq i \leq k}$  une famille de segments discrets, en notant  $L = \sum_{i=0}^k |S_i|$  la somme des longueurs des segments, le problème du recollement peut être résolu en  $\mathcal{O}(L^3 \times 2^{2 \times L^2})$ .*

*Démonstration.*

Je reprends ici les notations de la preuve de la proposition 2.2.62. D'abord, pour  $S_i$ , les mots  $w_i \cdot 0$  et  $w_i \cdot 1$  sont de longueur  $|S_i|$ . Maintenant, la construction d'un  $\omega$ -automate faible de langage  $(\sum_{i=0}^k w_i \cdot 0 + w_i \cdot 1)^{\mathbb{N}}$  consiste d'abord en la construction d'un  $\omega$ -automate faible en fleur, dont chaque pétale est étiqueté par  $w_i \cdot 0$  ou  $w_i \cdot 1$ . On peut en fait directement améliorer légèrement cette idée en utilisant l'idée suivante :



**FIGURE 2.19** – Légère amélioration de l' $\omega$ -automate en fleur.

Avec cette amélioration, cet  $\omega$ -automate faible aura alors  $1 + \sum_{i=0}^k |w_i|$  états, soit au plus  $L = \sum_{i=0}^k |S_i|$  états. En appliquant la proposition 2.2.34, on en déduit alors la complexité annoncée.

Bien que les bornes de complexité proposées ici soient grossières, elles sont particulièrement hautes par rapport aux valeurs attendues dans la littérature en géométrie discrète, où les problèmes sont le plus souvent résolus en temps

polynomial. Il devrait être possible, en utilisant l'interprétation des segments par l'arbre de Stern-Brocot [DL24], de résoudre ce même problème de manière plus efficace. Il s'agit toutefois de la première résolution connue du problème du recollement.

Je vais maintenant m'attacher à développer deux généralisations du problème du recollement, qui permettront de percevoir les limites théoriques des méthodes développées ici.

D'abord, il est possible de généraliser toutes les définitions précédentes pour étudier les droites discrètes dans un espace de dimension  $d \geq 3$ . Dans le cas des droites discrètes naïves, on parlera alors de points  $3^d - 1$ -voisins et  $3^d - 1$ -connexité, par exemple. Dans le contexte des méthodes que je développe ici, le plus important est la généralisation du lemme 2.2.61 dans le cas multidimensionnel, c.-à-d. la traduction entre géométrie discrète et combinatoire des mots. Cette généralisation devient la suivante :

**Lemme 2.2.65** ([Arn+94]). *Un mot  $x \in \llbracket 0, d - 1 \rrbracket^{\mathbb{N}}$  est un mot de billard hypercubique si, et seulement si, il existe une demi-droite discrète telle que  $x = (\text{rep}_D(n))_{n \in \mathbb{N}}$ .*

Les mots de billard hypercubique ont été introduits dans [Arn+94] et forment une généralisation des mots sturmiens à un alphabet plus grand. L'idée est qu'ils modélisent la trajectoire d'une balle de billard rebondissant sur les hyperfaces d'un hypercube : le lecteur intéressé pourra lire les travaux d'Andrieu pour plus de détails [AV22]. Toutefois, les méthodes que j'ai développées dans cette thèse ont peu de chance de résoudre ce problème. Les mots de billard hypercubiques ont en effet des propriétés assez différentes des mots sturmiens. Dans les méthodes que j'ai développées précédemment, je m'appuyais complètement sur la représentation S-adique d'un mot sturmien. Mais on ne connaît, à l'heure actuelle, aucune caractéristion S-adique des mots de billard hypercubiques. De plus, les principales familles de mots pour lesquelles on dispose d'une caractérisation S-adique (les mots sturmiens, les mots d'Arnoux-Rauzy et les mots minimaux ternaires dendriques) sont toutes de complexité linéaire, ce qui n'est pas le cas des mots de billard hypercubique, dont la complexité est en  $\theta(n^2)$  pour  $d = 2$  [Arn+94] et  $\theta(n^{3d-3})$  pour  $d \geq 4$  [BH07]. Cette haute complexité ne contredit donc pas l'existence d'une caractérisation, mais la rend plus difficile à établir.

Enfin, il est aussi possible de généraliser la question à des objets plus complexes que des segments. Considérons des morceaux de plans, que j'appellerai des surfaces.

**Définition 2.2.66** (Plan discret). *Soit  $(a, b, c, \mu, \omega) \in \mathbb{R}^4 \times \mathbb{R}_+^*$ . Le plan discret associé est :*

$$P_{a,b,c,\mu,\omega} = \{(x, y, z) \in \mathbb{Z}^3 \mid 0 \leq ax + by + cz + \mu < \omega\}$$

**Définition 2.2.67** (Surface discrète). *Soit  $S \subseteq \mathbb{Z}^3$ . On dit que  $S$  est une surface discrète si  $S$  est finie, 26-connexe et s'il existe un plan discret tel que  $S \subseteq P$ .*

La littérature usuelle s'est concentrée sur le cas des plans à vecteur normal rationnel. Les définitions utilisées ici, reprises de [LD23], permettent d'envisager le cas de plans à vecteur normal de coordonnées rationnellement indépendantes. Le problème de la reconnaissance des surfaces discrètes est un problème difficile de la géométrie discrète [KS91; Deb95], et plusieurs algorithmes de complexité quasi-linéaire ont été proposés ([GDZ05] ou [CB08]). Dans le cadre des résultats présentés ici, il suffit que la reconnaissance des surfaces discrètes soit décidable : je supposerai donc que les entrées de mon problème sont bien des surfaces discrètes.

On peut définir la notion de recollement de surfaces en un plan discret de manière analogue au recollement de segments en une demi-droite discrète.

**Définition 2.2.68** (Recollement en un plan discret). *Soient  $k \geq 0$  et  $S_0, \dots, S_k$   $k$  des surfaces discrètes. On dit que  $(S_i)_{0 \leq i \leq k}$  se recolle en un plan discret s'il existe un plan discret  $P$  et une partition de  $P = \bigsqcup_{i \in I} R_i$  (c.-à-d.  $R_i \cap R_j \neq \emptyset$  ssi  $i = j$ ) tel que pour tout  $i \in I$ , il existe  $j \in \llbracket 0, k \rrbracket$  tel que  $R_i \sim S_j$ .*



On peut remarquer que dans le cas des segments, je recollais les segments en une *demi*-droite, alors qu'ici je recolle les surfaces en un plan « entier ».

En passant des segments aux surfaces, on peut alors construire des objets plus complexes, et la conséquence en est sévère : il est impossible d'adopter les méthodes algorithmiques précédentes au problème du recollement de surfaces discrètes en un plan discret. Pour montrer cela, j'ai besoin d'introduire le modèle des tuiles de Wang et le problème du domino sur  $\mathbb{Z}^2$ .

**Définition 2.2.69** (Tuiles de Wang et problème du domino sur  $\mathbb{Z}^2$ ). Je note  $\Delta = \{N, S, E, O\}$  les quatre directions cardinales nord, sud, est et ouest. Soit  $\mathcal{C}$  un ensemble fini (assimilé à un ensemble de couleurs). Une tuile de Wang est un tuple  $t \in \mathcal{C}^\Delta$ .

Le problème du domino sur  $\mathbb{Z}^2$  est le suivant :

**Donnée :** un ensemble fini de couleurs  $\mathcal{C}$  et un ensemble de tuiles de Wang  $T \subseteq \mathcal{C}^\Delta$

**Question :** est-ce qu'il existe un pavage  $\pi : \mathbb{Z}^2 \rightarrow T$  valide, c.-à-d. vérifiant les conditions suivantes ?

$$\forall i, j \in \mathbb{Z}, \pi(i, j)_E = \pi(i + 1, j)_O \text{ et } \pi(i, j)_N = \pi(i, j + 1)_S.$$

**Théorème 2.2.70** ([Ber66]). Le problème du domino sur  $\mathbb{Z}^2$  est indécidable.

**Proposition 2.2.71.** Le problème de recollement de surfaces dans l'espace est indécidable.

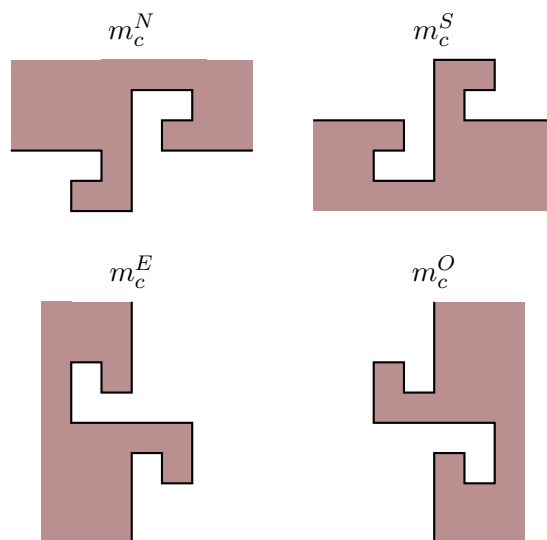
**Donnée :**  $k \geq 0$ , un ensemble fini de surfaces discrètes  $S_0, \dots, S_k$

**Question :** est-ce que  $(S_i)_{0 \leq i \leq k}$  se recolle en un plan discret ?

*Démonstration.*

On réduit le problème du domino au problème du recollement de surfaces dans l'espace. Soit  $\mathcal{C}$  un ensemble fini de couleurs, et  $T$  un ensemble fini de tuiles de Wang. Pour chaque  $c \in \mathcal{C}$ , on définit alors des motifs géométriques  $m_c^{(d)}$ , avec  $d \in \Delta$  vérifiant les symétries présentées dans la figure 2.20.

Soit  $L \in \mathbb{N}$  une borne de la largeur des motifs géométriques représentant les couleurs. Pour la tuile  $t$ , on définit alors la surface  $S_t$  de la manière suivante : il s'agit d'une surface d'épaisseur 1, ressemblant à un carré de largeur  $L$ , dont les bords Nord, Sud, Est et Ouest sont marqués par les motifs de la bonne couleur correspondants à leur direction.



**FIGURE 2.20** – Exemple des symétries à respecter pour une couleur  $i$  : les motifs de directions opposées doivent s'emboîter, et les formes utilisées doivent être non convexes.

J'affirme alors que  $\{S_t\}_{t \in T}$  se recolle en un plan discret si, et seulement si,  $T$  pave le plan. Si  $T$  pave le plan, alors l'implémentation géométrique permet à  $\{S_t\}_{t \in T}$  de se recoller en un plan discret. Réciproquement, si  $\{S_t\}_{t \in T}$  se recolle en un plan discret  $P$ , la platitude totale de tous les  $S_i$  et la non convexité des motifs géométriques contraignent alors  $P$  à être le plan plat. Maintenant, en partant d'un des éléments de la partition de  $P$ , la seule manière de combler les trous des quatre côtés consiste à recoller avec l'une des surfaces  $S_i$  qui encode la bonne couleur : on construit alors à partir de  $P$  un pavage à l'aide de  $T$  par itération à partir d'une première tuile.

Ce dernier résultat utilise des techniques assez classiques autour de l'encodage géométrique de couleurs ([Gea+16; AB23; HLN23]). Il est néanmoins remarquable, car il introduit une limite de décidabilité dimensionnelle au sein de la géométrie discrète, là où la littérature de la géométrie discrète se concentre surtout sur l'amélioration des complexités. Les méthodes de détection de mots sturmiens dans un  $\omega$ -automate faible prennent alors la place des algorithmes de décision du domino dans le cas 1D. Les complexités exponentielles du recollement de segments sont alors en quelque sorte excusées, puisque l'équivalent du recollement pour les surfaces est même indécidable en 2D.

## 2.3 Étude du cas des $\omega$ -automates de Büchi

Comme leur nom peut laisser l'entendre, les  $\omega$ -automates faibles sont un modèle faible d' $\omega$ -automates. D'autres modèles plus forts permettent de les généraliser, et un modèle en particulier est important en informatique théorique : celui des  $\omega$ -automates de Büchi (voir [PP04] pour un historique des  $\omega$ -automates de Büchi). Une motivation pour l'étude des  $\omega$ -automates de Büchi est logique. Dans [Hie+22], Oei prouve que les propriétés du premier ordre sur les mots sturmiens sont décidables. Cela a mené à la création du logiciel de vérification formelle Pecan sur les propriétés du premier ordre des mots sturmiens [Oei+21]. Les  $\omega$ -automates de Büchi, eux, permettent d'étudier les propositions de la logique monadique du second ordre (MSO) : il s'agit d'un modèle plus fort de logique. Si l'acceptation d'un mot sturmien par un  $\omega$ -automate de Büchi était décidable, on pourrait élargir les résultats de Oei à la logique MSO. Le reste de cette section établit des résultats partiels sur les liens entre les mots sturmiens et les  $\omega$ -automates de Büchi, et montre que les méthodes développées précédemment ne peuvent pas s'appliquer directement.

### 2.3.1 Définition d'un automate de Büchi

**Définition 2.3.1** ( $\omega$ -automate de Büchi). *Un  $\omega$ -automate  $\mathfrak{A} = (\mathcal{A}, Q, I, T, C)$  est un  $\omega$ -automate de Büchi s'il existe  $F \subseteq Q$  tel que  $C = \{(q_k)_{k \in \mathbb{N}} \in Q^{\mathbb{N}} \mid \text{il existe une infinité de } k \in \mathbb{N} \text{ tels que } q_k \in F\}$ . On décrira alors  $\mathfrak{A}$  par l'uplet  $(\mathcal{A}, Q, I, T, F)$ , et  $F$  est l'ensemble des états acceptants de  $\mathfrak{A}$ .*

Le modèle des  $\omega$ -automates de Büchi est un modèle plus fort que le modèle des  $\omega$ -automates faibles : en effet, tout  $\omega$ -automate faible  $\mathfrak{A} = (\mathcal{A}, Q, I, T)$  est aussi un  $\omega$ -automate de Büchi de la forme  $(\mathcal{A}, Q, I, T, Q)$ .

Le reste de la section va, en réalité, se concentrer sur un modèle légèrement différent, celui des  $\omega$ -automates de Büchi-arêtes.

**Définition 2.3.2** ( $\omega$ -automate de Büchi-arêtes). *Un  $\omega$ -automate  $\mathfrak{A} = (\mathcal{A}, Q, I, T, C)$  est un  $\omega$ -automate de Büchi-arêtes s'il existe  $F \subseteq T$  tel que  $C$  est l'ensemble des marches passant une infinité de fois par  $F$ .*

**Proposition 2.3.3** ([PP04]). *Les modèles des  $\omega$ -automates de Büchi et des  $\omega$ -automates de Büchi-arêtes sont équivalents.*

La raison de ce choix est purement pratique : l'opération de désubstitution est plus simple à manipuler pour des  $\omega$ -automates de Büchi-arêtes que pour des  $\omega$ -automates de Büchi. L'équivalence entre les deux modèles assure que la plupart des résultats établis ci-après resteront aussi valables dans le cas des  $\omega$ -automates de Büchi.

### 2.3.2 Désubstitution d'un $\omega$ -automate de Büchi-arêtes

**Définition 2.3.4** (Désubstitution d'un  $\omega$ -automate de Büchi-arêtes). Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T, F)$  un  $\omega$ -automate de Büchi-arêtes, et  $\sigma$  un morphisme. La désubstitution de  $\mathfrak{A}$  par  $\sigma$  est l' $\omega$ -automate de Büchi-arêtes  $\sigma^{-1}(\mathfrak{A}) = (\mathcal{A}, Q, I, \sigma^{-1}(T), \sigma^{-1}(F))$  où  $(q_s, a, q_t) \in \sigma^{-1}(T)$  si et seulement si  $(q_s, \sigma(a), q_t)$  est une marche dans  $\mathfrak{A}$  et  $(q_s, a, q_t) \in \sigma^{-1}(F)$  si et seulement si  $(q_s, \sigma, q_t)$  est une marche dans  $\mathfrak{A}$  passant par au moins un élément de  $F$ .

**Proposition 2.3.5.** Pour un  $\omega$ -automate de Büchi-arêtes  $\mathfrak{A}$  et un morphisme  $\sigma$ ,  $\mathcal{L}(\sigma^{-1}(\mathfrak{A})) = \{w \in \mathcal{A}^{\mathbb{N}} \mid \sigma(w) \in \mathcal{L}(\mathfrak{A})\}$ .

*Démonstration.*

On pose  $\mathfrak{A} = (\mathcal{A}, Q, I, T, F)$ . Soit  $x \in \mathcal{A}^{\mathbb{N}}$  un mot fini accepté par  $\sigma^{-1}(\mathfrak{A})$ . Soit  $(t_k)_{k \in \mathbb{N}}$  un calcul acceptant pour  $x$ , avec  $t_k = q_k \xrightarrow{x_k} q_{k+1}$ , et  $q_0 \in I$  et une infinité d'indices  $k$  tels que  $t_k$  est une arête acceptante de  $\sigma^{-1}(\mathfrak{A})$ . Par définition de  $\sigma^{-1}(\mathfrak{A})$ , pour tout  $k \in \mathbb{N}$ , il existe dans  $\mathfrak{A}$  une marche de la forme  $q_k \xrightarrow{\sigma(x_k)}^* q_{k+1}$ . En concaténant ces marches dans  $\mathfrak{A}$ , on obtient alors une marche sur  $\mathfrak{A}$  de la forme  $q_0 \xrightarrow{\sigma(x_0)}^* q_1 \xrightarrow{\sigma(x_1)}^* q_2 \xrightarrow{\sigma(x_2)}^* \dots$ . C'est une marche, et comme il y avait une infinité de  $t_k$  acceptant dans  $\sigma^{-1}(\mathfrak{A})$ , on peut choisir pour chacune de ses transitions un calcul  $q_k \xrightarrow{\sigma(x_k)}^* q_{k+1}$  passant par une arête acceptante de  $\mathfrak{A}$ . Donc  $q_0 \xrightarrow{\sigma(x_0)}^* q_1 \xrightarrow{\sigma(x_1)}^* q_2 \xrightarrow{\sigma(x_2)}^* \dots$  est bien un calcul acceptant de  $\mathfrak{A}$  étiqueté par  $\sigma(x)$ .

Réciproquement, soit  $x \in \mathcal{A}^{\mathbb{N}}$  tel que  $\sigma(x)$  est accepté par  $\mathfrak{A}$ . Soit  $(t_k)_{k \in \mathbb{N}}$  un calcul acceptant pour  $\sigma(x)$  dans  $\mathfrak{A}$  : alors on peut l'écrire sous la forme  $q_0 \xrightarrow{\sigma(x)_0} q_1 \xrightarrow{\sigma(x)_1} q_2 \xrightarrow{\sigma(x)_2} \dots$ . On peut regrouper certaines transitions pour réécrire ce calcul sous la forme  $q_0 = q'_0 \xrightarrow{\sigma(x_0)}^* q'_1 \xrightarrow{\sigma(x_1)}^* q'_2 \xrightarrow{\sigma(x_2)}^* \dots$ . Une infinité de ces morceaux contiennent des arêtes acceptantes. Par définition de  $\sigma^{-1}(\mathfrak{A})$ , on a donc  $q'_0 \xrightarrow{x_0} q'_1 \xrightarrow{x_1} q'_2 \xrightarrow{x_2} \dots$  une marche sur  $\sigma^{-1}(\mathfrak{A})$  : et toujours par définition de  $\sigma^{-1}(\mathfrak{A})$ , une infinité des transitions de cette marche sont acceptantes. Donc  $x \in \mathcal{L}(\sigma^{-1}(\mathfrak{A}))$ .

On retrouve alors plusieurs propriétés de cette désubstitution : le nombre d'états ne varie pas, l'ensemble des arêtes est borné, donc on pourra construire des méta- $\omega$ -automates de désubstitution.

L'une des difficultés dans l'étude des  $\omega$ -automates de Büchi-arêtes est que la notion de calcul doit être manipulée avec précaution. Pour cela, j'introduis un outil qui permet de construire de nouveaux calculs à partir des désubstitutions : les morphismes de calculs.

**Définition 2.3.6** (Morphisme de calculs). Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T, F)$  un  $\omega$ -automate de Büchi-arêtes et soit  $\sigma$  une substitution sur  $\mathcal{A}$ . Soit  $\sigma^{-1}(\mathfrak{A}) = (\mathcal{A}, Q, I, T', F')$  la désubstitution de  $\mathfrak{A}$  par  $\sigma$ . Un morphisme de calculs pour  $\mathfrak{A}$  et  $\sigma$  est un morphisme  $\phi_{\sigma, \mathfrak{A}} : T'^* \rightarrow T^*$  qui vérifie les conditions suivantes :

- pour tout  $t' = q_s \xrightarrow{a} q_t \in T'$ ,  $\phi_{\sigma, \mathfrak{A}}(t')$  est un calcul fini de  $\mathfrak{A}$  de la forme  $q_s \xrightarrow{\sigma(a)}^* q_t$ ;
- pour tout  $t' = q_s \xrightarrow{a} q_t \in F'$ ,  $\phi_{\sigma, \mathfrak{A}}(t')$  est un calcul fini de  $\mathfrak{A}$  de la forme  $q_s \xrightarrow{\sigma(a)}^* q_t$  qui passe par une arête acceptante de  $\mathfrak{A}$ .

J'attire ici l'attention du lecteur : pour  $\mathfrak{A}$  et  $\sigma$  donnés, il peut y avoir plusieurs morphismes de calculs associés. Si  $\mathfrak{A}$  est déterministe, pour toute substitution  $\sigma$ , alors il n'y a qu'un seul morphisme de calcul possible : mais le modèle

de  $\omega$ -automates de Büchi-arêtes déterministes est plus faible que le modèle des  $\omega$ -automates de Büchi-arêtes. Je ne ferai pas, sauf cas précis, cette hypothèse de déterminisme.

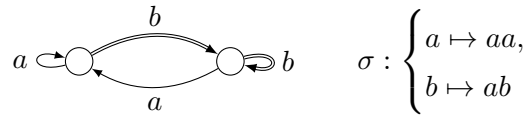
Dans la suite de la section, on ne procédera à un choix parmi les morphismes de calculs que lorsque précisé : sinon, c'est qu'ils font tous l'affaire. Les morphismes de calculs bénéficient de plusieurs propriétés qui nous intéresseront par la suite, et détaillées dans la proposition suivante :

**Proposition 2.3.7.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes et  $\sigma$  une substitution. Soit  $\phi_{\sigma, \mathfrak{A}}$  un morphisme de calculs pour  $\mathfrak{A}$  et  $\sigma$ .*

1. *si  $c' = (t'_k)_{k \in \mathbb{N}}$  est un calcul acceptant de  $\sigma^{-1}(\mathfrak{A})$ , alors  $\phi_{\sigma, \mathfrak{A}}(c')$  est un calcul acceptant de  $\mathfrak{A}$ ;*
2. *si un mot  $w$  tel que  $w = \sigma(w')$  pour un certain  $w'$  est accepté par  $\mathfrak{A}$ ,  $w'$  étant accepté par  $\sigma^{-1}(\mathfrak{A})$ , pour tout calcul  $c'$  acceptant  $w'$  dans  $\sigma^{-1}(\mathfrak{A})$ ,  $\phi_{\sigma, \mathfrak{A}}(c')$  est un calcul acceptant  $w$  dans  $\mathfrak{A}$ ;*
3. *si  $\sigma^{-1}(\mathfrak{A}) = \mathfrak{A}$ , et s'il existe  $t \in T$  telle que  $\phi_{\sigma, \mathfrak{A}}(t)$  commence par  $t$ , alors  $c = \phi_{\sigma, \mathfrak{A}}^\infty(t)$  est défini et est une marche infinie sur  $\mathfrak{A}$  (sans forcément être un calcul acceptant).*

### 2.3.3 Un problème topologique

Bien que j'ai pu établir directement la désubstitution pour les  $\omega$ -automates de Büchi, le reste des méthodes développées dans le cas des  $\omega$ -automates faibles ne peuvent pas s'appliquer directement au cas des  $\omega$ -automates de Büchi. Par exemple, la proposition 2.2.16 sur les points fixes de substitution n'est plus valide dans le cas des  $\omega$ -automates de Büchi :



**FIGURE 2.21** – Contre-exemple dans le cas des  $\omega$ -automates de Büchi :  $\sigma^{-1}(\mathfrak{A}) = \mathfrak{A}$ , mais  $\mathfrak{A}$  n'accepte aucun mot purement substitutif généré par  $\sigma$  ou une de ses puissances. Les arêtes acceptantes sont les arêtes doubles.

Dans cet exemple, on voit que la marche étiquetée par  $\sigma^\infty(a)$  est valide, mais ne passe pas par une infinité d'arêtes acceptantes, et est donc refusée.

La principale raison est topologique : le langage d'un  $\omega$ -automate de Büchi n'est pas nécessairement fermé pour la topologie prodiscrète, ni l'ensemble des calculs acceptants. Donc les arguments de compacité échouent dans le cas des  $\omega$ -automates de Büchi. C'est pour cela que, dans le reste de cette section, je me contenterai de prouver des résultats partiels ou de faire des hypothèses supplémentaires.

### 2.3.4 Acceptation d'un mot sturmien morphique par un $\omega$ -automate de Büchi-arêtes

Pour mieux comprendre les liens entre un  $\omega$ -automate de Büchi-arêtes et les mots sturmiens qu'il accepte, une idée est de généraliser les résultats établis dans le cas des  $\omega$ -automates faibles. Dans le cas des  $\omega$ -automates faibles, il est possible d'étendre la proposition 2.2.26 pour prouver que si un  $\omega$ -automate faible accepte un mot sturmien, alors il accepte un mot sturmien morphique. Il s'agit d'une propriété importante, car elle borne les possibilités des représentations S-adiques d'Arnoux des mots sturmiens acceptés : on est sûr qu'au moins l'une d'entre elles est ultimement périodique.

Dans cette sous-section, mon objectif est d'étendre ce résultat partiellement au cas des  $\omega$ -automates de Büchi-arêtes :

**Proposition 2.3.8.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes dont tous les états sont initiaux acceptant un mot sturmien. Alors  $\mathfrak{A}$  accepte un mot sturmien morphique.*

L'objectif de cette sous-section est la démonstration de la proposition 2.3.8. Pour cela, j'introduis deux nouvelles propriétés.

**Définition 2.3.9** (Relation de sous-automate). *Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T_{\mathfrak{A}}, F_{\mathfrak{A}})$  et  $\mathfrak{B} = (\mathcal{A}, Q, I, T_{\mathfrak{B}}, F_{\mathfrak{B}})$  deux  $\omega$ -automates de Büchi-arêtes partageant les mêmes alphabets, états et états initiaux :  $\mathfrak{B}$  est un sous-automate de  $\mathfrak{A}$  si  $T_{\mathfrak{B}} \subseteq T_{\mathfrak{A}}$  et  $F_{\mathfrak{B}} \subseteq F_{\mathfrak{A}}$ . Je note alors  $\mathfrak{B} \odot \mathfrak{A}$ .*

En particulier, si deux  $\omega$ -automates de Büchi-arêtes vérifient  $\mathfrak{B} \odot \mathfrak{A}$ ,  $\mathfrak{A}$  et  $\mathfrak{B}$  ont les mêmes états.

**Définition 2.3.10** (Sturm-minimalité). *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes qui accepte un mot sturmien. On dit que  $\mathfrak{A}$  est Sturm-minimal si le seul sous- $\omega$ -automate de  $\mathfrak{A}$  qui accepte aussi un mot sturmien est  $\mathfrak{A}$  lui-même.*

**Lemme 2.3.11.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes vérifiant les conditions suivantes :*

1. tous les états de  $\mathfrak{A}$  sont initiaux;
2.  $\mathfrak{A}$  accepte un mot sturmien et est Sturm-minimal;
3. il existe une substitution  $\sigma \in \mathcal{S}_{St}^* \setminus \{\{L_0, R_0\}^* \cup \{L_1, R_1\}^*\}$  tel que  $\sigma^{-1}(\mathfrak{A}) = \mathfrak{A}$ .

*Alors  $\mathfrak{A}$  accepte un mot sturmien purement substitutif.*

*Démonstration.*

L'hypothèse (3) signifie que  $\sigma$  contient des substitutions sturmiennes élémentaires de type 0 et de type 1. Notons  $\phi_{\sigma, \mathfrak{A}}$  un morphisme de calculs pour  $\mathfrak{A}$  et  $\sigma$ .

Par l'hypothèse (2), soit  $c$  un calcul acceptant un mot sturmien dans  $\mathfrak{A}$ . Alors, d'après l'hypothèse (3), pour tout  $n \geq 0$ ,  $\phi_{\sigma, \mathfrak{A}}^n(c)$  est un calcul acceptant de  $\mathfrak{A}$  étiqueté par un mot sturmien.

Considérons la suite  $(\phi_{\sigma, \mathfrak{A}}^n(c))_{n \in \mathbb{N}}$  : c'est une suite de calculs de  $\mathfrak{A}$ . Il en existe alors une infinité commençant par la même arête  $g$ . Quitte à modifier la définition de  $c$ , supposons que  $c$  commence aussi par l'arête  $g$ . Si  $c$  ne contenait qu'une seule fois l'arête  $g$ , alors en retirant cette première arête de  $c$ , par l'hypothèse (1), on obtiendrait un nouveau calcul acceptant de  $\mathfrak{A}$  étiqueté par un mot sturmien : donc  $\mathfrak{A}$  ne serait pas Sturm-minimal. Soit  $k \geq 1$  tel que  $c_{\llbracket 0, k \rrbracket}$  contient deux occurrences de l'arête  $g$  (dont une comme première arête) et au moins une arête acceptante.

Comme  $\sigma$  contient des substitutions sturmiennes élémentaires des deux types, la suite  $(\sigma)_{n \in \mathbb{N}}$  est ubicroissante. Donc la suite  $(\phi_{\sigma, \mathfrak{A}})_{n \in \mathbb{N}}$  est aussi ubicroissante. Donc il existe  $m \geq 1$  tel que  $|\phi_{\sigma, \mathfrak{A}}^m(g)| > k$ .

On choisit alors  $\phi_{\sigma, \mathfrak{A}}^m$  tel que  $c_{\llbracket 0, k \rrbracket} \sqsubseteq_p \phi_{\sigma, \mathfrak{A}}^m(g)$  : notons  $\phi_{\sigma, \mathfrak{A}}^m = \kappa$ .

Considérons alors le calcul  $c' = \kappa^\infty(g)$ . C'est bien une marche sur  $\mathfrak{A}$ ; et comme  $\sigma$  est des deux types,  $c'$  est étiqueté par un mot sturmien. Reste à montrer que  $c'$  est acceptant. Il suffit alors de remarquer que  $\kappa(g) = gu$ , avec  $u$  contenant une arête acceptante et une occurrence de  $g$ . D'après le théorème 1.1.19,  $c' = gu\kappa(u)\kappa^2(u)\kappa^3(u) \dots$ . Or,  $g$  apparaît dans  $\kappa(g)$ , et  $\kappa(g)$  contient une arête acceptante : donc par récurrence immédiate,  $\kappa^n(g)$  contient une arête acceptante. Et comme  $g$  apparaît dans  $u$ ,  $\kappa^n(u)$  contient une arête acceptante. Donc  $c'$  contient une infinité d'arêtes acceptantes, et est bien un calcul acceptant de  $\mathfrak{A}$ . Par construction,  $c'$  est étiqueté par un mot sturmien purement substitutif.

Je vais alors introduire une variante de la désubstitution que j'appelle la *désubstitution à pertes*.

**Définition 2.3.12** (Désubstitution à pertes). *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes et  $\sigma$  une substitution. Une désubstitution à pertes de  $\mathfrak{A}$  par  $\sigma$  est un  $\omega$ -automate de Büchi-arêtes  $\mathfrak{B}$  tel que  $\mathfrak{B} \odot \sigma^{-1}(\mathfrak{A})$ .*

On remarque que pour  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes et  $\sigma$  une substitution,  $\mathfrak{A}$  admet de nombreuses désubstitutions à pertes par  $\sigma$ . C'est une forme de non-déterminisation de l'opération de désubstitution. La désubstitution à pertes bénéficie de plusieurs propriétés :

**Proposition 2.3.13.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes,  $\sigma$  et  $\tau$  deux substitutions.*

Alors :

1. *si  $\mathfrak{B}$  est une désubstitution à pertes de  $\mathfrak{A}$  par  $\sigma$  et  $\mathfrak{C}$  est une désubstitution à pertes de  $\mathfrak{B}$  par  $\tau$ , alors  $\mathfrak{C}$  est une désubstitution à pertes de  $\mathfrak{A}$  par  $\tau \circ \sigma$ ;*
2. *si  $\phi_{\sigma, \mathfrak{A}}$  est un morphisme de calculs pour  $\mathfrak{A}$  et  $\sigma$ , pour toute  $\mathfrak{B} = (\mathcal{A}, Q, I, T_{\mathfrak{B}}, F_{\mathfrak{B}})$  désubstitution à pertes de  $\mathfrak{A}$  par  $\sigma$ , on peut restreindre  $\phi_{\sigma, \mathfrak{A}}$  à l'alphabet  $T_{\mathfrak{B}}$ .*

Il est alors possible d'étendre le lemme 2.3.11 au cas de la désubstitution à pertes.

**Lemme 2.3.14.** *Soit  $\mathfrak{A}$  un  $\omega$ -automate de Büchi-arêtes vérifiant les conditions suivantes :*

- *tous les états de  $\mathfrak{A}$  sont initiaux;*
- *$\mathfrak{A}$  accepte un mot sturmien et est Sturm-minimal;*
- *il existe une substitution  $\sigma \in \mathcal{S}_{St}^* \setminus \{\{L_0, R_0\}^* \cup \{L_1, R_1\}^*\}$  tel que  $\mathfrak{A} \otimes \sigma^{-1}(\mathfrak{A})$ .*

Alors  $\mathfrak{A}$  accepte un mot sturmien purement substitutif.

La preuve est essentiellement la même que celle du lemme 2.3.11, en utilisant les propriétés décrites dans la proposition 2.3.13. Cette généralisation me permet enfin de prouver le résultat attendu :

*Démonstration de la proposition 2.3.8.*

Pour  $\mathfrak{A}$ , on construit un nouveau méta- $\omega$ -automate qui retranscrit la notion de désubstitution à pertes : on définit  $\mathcal{M}_{\mathcal{S}_{St}, \text{pertes}}(\mathfrak{A}) = (\mathcal{S}_{St}, \mathfrak{S}(\mathfrak{A}), \{\mathfrak{A}\}, T)$  où :

- $\mathfrak{S}(\mathfrak{A})$  est l'ensemble des  $\omega$ -automates de Büchi-arêtes avec le même alphabet, les mêmes états et mêmes états initiaux que  $\mathfrak{A}$  acceptant au moins un mot sturmien;
- $\mathfrak{B} \xrightarrow{\sigma} \mathfrak{C} \in T$  ssi  $\mathfrak{C} \otimes \sigma^{-1}(\mathfrak{B})$ .

De manière équivalente, c'est le graphe représentant la relation  $\otimes$ , en marquant  $\mathfrak{A}$  comme l'état initial. Comme  $\mathfrak{A}$  accepte un mot sturmien,  $\mathcal{M}_{\mathcal{S}_{St}, \text{pertes}}(\mathfrak{A})$  accepte au moins une représentation S-adique d'Arnoux. Modifions le méta- $\omega$ -automate de désubstitution, et notons cette modification  $\mathcal{M}_{\mathcal{S}_{St}, \text{pertes}}^{\text{modif}}(\mathfrak{A}) = (\mathcal{A}, Q, I, T')$  avec :

- $Q \subseteq \mathfrak{S}(\mathfrak{A})$  est l'ensemble des  $\omega$ -automates de Büchi-arêtes appartenant à  $\mathfrak{S}(\mathfrak{A})$  qui sont Sturm-minimaux;
- $I$  est l'ensemble des  $\omega$ -automates  $\mathfrak{B}$  vérifiant  $\mathfrak{B} \otimes \mathfrak{A}$ ;
- $T' = (Q \times \mathcal{S}_{St} \times Q) \cap T$ .

Comme  $\mathcal{M}_{\mathcal{S}_{St}, \text{pertes}}(\mathfrak{A})$  accepte un mot infini avec une alternance infinie de type,  $\mathcal{M}_{\mathcal{S}_{St}, \text{pertes}}^{\text{modif}}(\mathfrak{A})$  aussi : donc il existe  $\mathfrak{B} \in Q$  qui boucle par un mot contenant les deux types. On peut alors appliquer le lemme 2.3.14 : donc  $\mathfrak{B}$  accepte un mot sturmien purement substitutif. En remontant à  $\mathfrak{A}$ ,  $\mathfrak{A}$  accepte donc un mot sturmien morphique.

Une conséquence importante de la proposition 2.3.8 est qu'on peut montrer que l'acceptation d'un mot sturmien par un  $\omega$ -automate de Büchi-arêtes est un problème semi-décidable sous certaines conditions :

**Proposition 2.3.15.** *Le problème de l'acceptation d'un mot sturmien par un  $\omega$ -automate de Büchi-arêtes d'états tous initiaux est un problème semi-décidable.*

**Donnée :** *un  $\omega$ -automate de Büchi-arêtes  $\mathfrak{A}$  dont tous les états sont initiaux*

**Question :** *est-ce que  $\mathfrak{A}$  accepte un mot sturmien ?*

*Démonstration.*

Le semi-algorithme est le suivant : on énumère tous les couples de substitutions sturmiennes  $(\sigma, \tau)$  avec  $\tau$  contenant des substitutions des deux types. En utilisant l'algorithme présenté dans [CT02], on peut tester si le mot morphique généré par  $\sigma$  et  $\tau$  est accepté par  $\mathfrak{A}$ . En procédant ainsi, on énumère tous les mots sturmiens morphiques. Donc  $\mathfrak{A}$  termine et dit « oui » si, et seulement si,  $\mathfrak{A}$  accepte un mot sturmien.

L'hypothèse qui contraint les états à tous être initiaux retranscrit l'idée que le langage de  $\mathfrak{A}$  est invariant par décalage. Elle est donc naturelle dans le cadre de la dynamique symbolique, mais peu satisfaisante du point de vue logique. De plus, le semi-algorithme précédent ne permet pas d'obtenir la co-semi-décidabilité du problème, car on ne peut pas donner de borne sur les tailles de  $\sigma$  et  $\tau$  énumérés.

### 2.3.5 Cas des $\omega$ -automates de Büchi déterministes co-complets

Dans cette sous-section, je retourne brièvement sur l'étude du modèle des  $\omega$ -automates de Büchi et non de Büchi-arêtes.

Une autre piste à envisager consiste à modifier les preuves de Carton et Thomas [CT02]. Je propose la méthode suivante, mais elle nécessite deux hypothèses très fortes : le déterminisme et la co-complétude. Pour rappel, étant donné un  $\omega$ -automate  $\mathfrak{A} = (\mathcal{A}, Q, I, T, C)$ , il est déterministe si pour tout  $a \in \mathcal{A}$  et  $d \in Q$ , il existe au plus un état  $f \in Q$  tel que  $d \xrightarrow{a} f \in T$ ; il est co-complet si pour tout  $a \in \mathcal{A}$  et  $f \in Q$ , il existe au moins un état  $d \in Q$  tel que  $d \xrightarrow{a} f \in T$ .

**Proposition 2.3.16.** *Le problème de l'acceptation d'un mot sturmien purement substitutif par un  $\omega$ -automate de Büchi déterministe co-complet est décidable.*

**Donnée :** un  $\omega$ -automate de Büchi déterministe et co-complet  $\mathfrak{A}$

**Question :** est-ce que  $\mathfrak{A}$  accepte un mot sturmien purement substitutif?

Ma preuve s'inspire beaucoup d'idées développées dans celle de Carton et Thomas, et je recommande donc au lecteur de d'abord s'informer sur leur preuve avant de lire la preuve suivante.

*Démonstration.*

Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T, F)$  un  $\omega$ -automate de Büchi-arêtes.

On définit la relation d'équivalence de Büchi sur les mots finis de la manière suivante :

$$u \equiv u'$$

$$\Updownarrow$$

$$\forall p, q \in Q, (p \xrightarrow{u}^* q \text{ est une marche de } \mathfrak{A} \iff p \xrightarrow{u'}^* q \text{ est une marche de } \mathfrak{A})$$

$$\text{et } (p \xrightarrow{u}^* q \text{ est une marche acceptante de } \mathfrak{A} \iff p \xrightarrow{u'}^* q \text{ est une marche acceptante de } \mathfrak{A})$$

Ici, par une marche acceptante, j'entends une marche passant par un état acceptant. Cette relation d'équivalence est une congruence : si  $u \equiv u'$  et  $v \equiv v'$ , alors  $uv \equiv u'v'$ . Elle vérifie aussi une propriété de congruence infinie : si  $x$  et  $y$  sont deux mots infinis se découpant en  $x = x^{(1)}x^{(2)}x^{(3)} \dots$  et  $y = y^{(1)}y^{(2)}y^{(3)} \dots$  avec  $x^{(i)}, y^{(i)} \in \mathcal{A}^+$  et  $x^{(i)} \equiv y^{(i)}$  pour tout  $i \geq 1$ , alors  $x$  est accepté par  $\mathfrak{A}$  ssi  $y$  est accepté par  $\mathfrak{A}$ . Le quotient de  $\mathcal{A}^+$  par  $\equiv$ , noté  $\mathcal{A}^+ / \equiv$ , vérifie de plus deux propriétés :

- comme  $\equiv$  préserve la concaténation, la concaténation est une opération interne sur les classes d'équivalence : donc  $\mathcal{A}^+ / \equiv$  est un semi-groupe;

- de plus, un élément de  $\mathcal{A}^+ / \equiv$  est identifié par une matrice  $(m_{p,q})_{p,q \in Q}$  où  $m_{p,q}$  indique si l'élément relie  $p$  à  $q$ , et si oui en passant par une arête acceptante : donc  $\text{Card}(\mathcal{A}^+ / \equiv) \leq 3^{|Q|^2}$ .  
Donc  $\mathcal{A}^+ / \equiv$  est un semi-groupe fini.

Étudions alors l'ensemble  $\mathcal{M}$  des morphismes de  $\mathcal{A}^+$  vers  $\mathcal{A}^+ / \equiv$ . Un élément de  $\mathcal{M}$  est déterminé par l'image des lettres : donc  $\mathcal{M}$  est un ensemble fini, de cardinal  $\text{Card}(\mathcal{M}) \leq |\mathcal{A}|^{|\mathcal{A}^+ / \equiv|}$ . Soit  $\pi$  la projection de  $\mathcal{A}^+$  vers  $\mathcal{A}^+ / \equiv$ , c.-à-d. l'application qui envoie un élément de  $\mathcal{A}^+$  vers sa classe d'équivalence par  $\equiv$ . Comme  $\equiv$  est une congruence,  $\pi$  est un morphisme, c.-à-d.  $\pi \in \mathcal{M}$ .

On en déduit que pour toute substitution  $\sigma$  sur  $\mathcal{A}$ ,  $\pi \circ \sigma \in \mathcal{M}$ . Plus encore : considérons la suite  $(\pi \circ \sigma^i)_{i \geq 1}$ . Cette suite poursuit une dynamique<sup>8</sup> dans  $\mathcal{M}$  qui est finie : donc  $(\pi \circ \sigma^i)_{i \geq 1}$  est une suite ultimement périodique pour toute substitution  $\sigma$ . De plus, la somme de la prépériode et de la période est plus petite que la taille de  $\mathcal{M}$  : il existe donc une borne uniforme,  $\text{Card}(\mathcal{M})$ , sur la prépériode et la période de ces suites. En notant  $\text{Seq}_{\mathcal{M}} = \{(\phi, \psi) \mid \phi, \psi \in \mathcal{M}^*, 1 \leq |\psi| \leq |\phi\psi| \leq \text{Card}(\mathcal{M}) \text{ et } \forall i \neq j, (\phi\psi)_i \neq (\phi\psi)_j\}$  l'ensemble des couples de mots sur  $\mathcal{M}$  sans répétition d'élément de  $\mathcal{M}$  dont la longueur totale est plus petite que  $\text{Card}(\mathcal{M})$ , on peut alors écrire que pour toute substitution  $\sigma$  sur  $\mathcal{A}$ , il existe un couple  $(\phi, \psi) \in \text{Seq}_{\mathcal{M}}$  tel que, pour  $i \geq 1$  :

$$\pi \circ \sigma^i = \begin{cases} \phi_{i-1} & \text{si } i \leq |\phi|, \\ \psi_{(i-|\phi|-1) \bmod |\psi|} & \text{sinon.} \end{cases}$$

Je dirai alors que la suite des puissances de  $\sigma$  s'identifie au couple  $(\phi, \psi)$ ; et pour  $(\phi, \psi) \in \text{Seq}_{\mathcal{M}}$ , j'écris  $\Sigma^{(\phi, \psi)} = \phi\psi^{\mathbb{N}}$ , c.-à-d. le mot infini ultimement périodique sur l'alphabet  $\mathcal{M}$  de prépériode  $\phi$  et de période  $\psi$ . Avec ces notations, la suite des puissances de  $\sigma$  s'identifie à  $(\phi, \psi)$  ssi  $\pi \circ \sigma^i = \Sigma_{i-1}^{(\phi, \psi)}$ .

Soit  $(\phi, \psi) \in \text{Seq}_{\mathcal{M}}$  : je définis  $\iota(\phi, \psi) = (\pi, \phi_0, \dots, \phi_{|\phi|-1}, \psi_0, \dots, \psi_{|\psi|-1}) \in \mathcal{M}^{|\phi\psi|+1}$  le découpage élément par élément de  $\phi\psi$  en ajoutant  $\pi$  au début; et  $\kappa(\phi, \psi) = (\phi_0, \dots, \phi_{|\phi|-1}, \psi_0, \dots, \psi_{|\psi|-1}, \psi_0)$  : c'est le vecteur  $\iota(\phi, \psi)$  décalé d'un cran à gauche en répétant  $\psi_0$  à la fin. Ces notations me permettent de définir un automate fini déterministe<sup>9</sup>  $\mathfrak{B}_{\phi, \psi} = (\mathcal{S}_{St}, \mathcal{M}^{|\phi\psi|+1}, \{\iota(\phi, \psi)\}, \{\kappa(\phi, \psi)\}, \delta_{\mathfrak{B}_{\phi, \psi}})$  tel que  $\delta_{\mathfrak{B}_{\phi, \psi}}((m_0, m_1, \dots, m_{|\phi\psi|}), \mu) = (m_0 \circ \mu, \dots, m_{|\phi\psi|} \circ \mu)$  pour  $\mu \in \mathcal{S}_{St}$ . Alors  $\sigma \in \mathcal{L}(\mathfrak{B}_{\phi, \psi})$  ssi  $\phi_0 = \pi \circ \sigma, \dots, \phi_{|\phi|-1} = \pi \circ \sigma^{|\phi|}, \psi_0 = \pi \circ \sigma^{|\phi|+1}, \dots, \psi_{|\psi|-1} = \pi \circ \sigma^{|\phi\psi|}$  et  $\pi \circ \sigma^{|\phi\psi|+1} = \pi \circ \sigma^{|\phi|}$  : on dispose d'un automate fini décrivant exactement les substitutions sturmiennes dont la suite des puissances s'identifie à  $(\phi, \psi)$ .

Pour  $\sigma \in \mathcal{S}_{St}^*$  et  $a \in \mathcal{A}$ , on note, s'il existe,  $u_{\sigma, a}$  le mot vérifiant  $\sigma(a) = au_{\sigma, a}$ . Mon objectif est maintenant d'étudier, pour  $(\phi, \psi) \in \text{Seq}_{\mathcal{M}}$  et  $a \in \mathcal{A}$ , l'ensemble  $\mathcal{I}_{\phi, \psi, a} = \{(\phi_0(u_{\sigma, a}), \dots, \phi_{|\phi|-1}(u_{\sigma, a}), \psi_0(u_{\sigma, a}), \dots, \psi_{|\psi|-1}(u_{\sigma, a})) \mid \exists \sigma \in \mathcal{S}_{St}^*, \text{ la suite des puissances de } \sigma \text{ s'identifie à } (\phi, \psi)\} \subseteq (\mathcal{A}^+ / \equiv)^{|\phi\psi|}$ . Pour cela, on fait le calcul suivant pour tout  $\sigma$  convenable :

$$\begin{aligned} \Sigma_i^{(\phi, \psi)}(a) &= \pi \circ \sigma^{i+1}(a) \\ &= \pi \circ \sigma^i(au_{\sigma, a}) \\ &= (\pi \circ \sigma^i(a)) \cdot (\pi \circ \sigma^i(u_{\sigma, a})) \\ &= \Sigma_{i-1}^{(\phi, \psi)}(a) \cdot \Sigma_{i-1}^{(\phi, \psi)}(u_{\sigma, a}) \end{aligned}$$

8. on passe d'un terme au suivant par une opération : c'est une suite de la forme  $u_{n+1} = f(u_n)$  avec  $f$  une fonction.

9. Ici, la notation sera de la forme  $\mathfrak{B} = (\mathcal{A}, Q, I, F, \delta)$  où  $\mathcal{A}$  est l'alphabet,  $Q$  l'ensemble des états,  $I$  les états initiaux,  $F$  les états finaux,  $\delta : Q \times \mathcal{A} \rightarrow Q$  la fonction partielle de transition.



Sans hypothèse supplémentaire, je ne sais pas comment aller plus loin. J'utilise alors maintenant l'hypothèse que  $\mathfrak{A}$  est déterministe et co-complet pour établir le lemme suivant :

**Lemme 2.3.17.** *Si  $\mathfrak{A}$  est déterministe et co-complet, alors  $\mathcal{A}^+ / \equiv$  est un semi-groupe annulable à gauche : si  $uv \equiv uw$ , alors  $v \equiv w$ .<sup>10</sup>*

*Démonstration.*

Soient  $u, v, w \in \mathcal{A}^+$  tels que  $uv \equiv uw$ , et soient  $p, q \in Q$ . Supposons que  $p \xrightarrow{v}^* q$  soit une marche dans  $\mathfrak{A}$  : alors, par co-complétude, il existe  $r \in Q$  tel que  $r \xrightarrow{u}^* p$  est une marche dans  $\mathfrak{A}$  : donc  $r \xrightarrow{u}^* p \xrightarrow{v}^* q$  est une marche dans  $\mathfrak{A}$ . Comme  $uv \equiv uw$ , on a alors  $r \xrightarrow{uv}^* q$  est une marche dans  $\mathfrak{A}$ . Par déterminisme de  $\mathfrak{A}$ , cette marche se découpe alors  $r \xrightarrow{u}^* p \xrightarrow{w}^* q$  : donc  $p \xrightarrow{w}^* q$  est une marche dans  $\mathfrak{A}$ . La démonstration est similaire dans le cas où  $p \xrightarrow{v}^* q$  est une marche acceptante de  $\mathfrak{A}$ .

Avec le lemme 2.3.17, on en déduit alors que  $\Sigma_i^{(\phi, \psi)}(u_{\sigma, a})$  est uniquement déterminé par  $\phi, \psi$  et  $a$  : pour tout  $\sigma$  dont la suite des puissances s'identifie à  $(\phi, \psi)$ , la suite  $(\Sigma_i^{(\phi, \psi)}(u_{\sigma, a}))_{i \in \mathbb{N}}$  est la même. Donc  $\mathcal{I}_{\phi, \psi, a}$  ne contient au plus qu'un élément.

Pour  $u \in \mathcal{A}^+$ , je note  $\bar{u}$  la classe d'équivalence de  $u$  pour la relation  $\equiv$ . Parallèlement, il est relativement facile de calculer  $\overline{\sigma(a)}$  en fonction de  $\sigma$  : on peut pour cela construire un nouvel automate  $\text{ImageSturm} = (\mathcal{S}_{St}, (\mathcal{A}^+ / \equiv)^2, (\bar{0}, \bar{1}), \delta_{\text{ImageSturm}})$  avec :

$$\begin{aligned} \delta_{\text{ImageSturm}}((u, v), L_0) &= (u, uv) & \delta_{\text{ImageSturm}}((u, v), L_1) &= (vu, v) \\ \delta_{\text{ImageSturm}}((u, v), R_0) &= (u, vu) & \delta_{\text{ImageSturm}}((u, v), R_1) &= (uv, v) \end{aligned}$$

L'automate  $\text{ImageSturm}$  vérifie par récurrence immédiate  $\delta_{\text{ImageSturm}}((\bar{0}\bar{1}), \sigma) = (\overline{\sigma(0)}, \overline{\sigma(1)})$  pour  $\sigma \in \mathcal{S}_{St}$ .

En combinant  $\mathfrak{B}_{\phi, \psi}$ ,  $\text{ImageSturm}$  et  $\mathcal{I}_{\phi, \psi, a}$ , on peut alors déterminer l'ensemble fini  $\{(\overline{\sigma(a)}, \Sigma_0^{(\phi, \psi)}(u_{\sigma, a}), \dots, \Sigma_{|\phi\psi|-1}^{(\phi, \psi)}(u_{\sigma, a})) \mid \exists \sigma \in \mathcal{S}_{St}^*$ , la suite des puissances de  $\sigma$  s'identifie à  $(\phi, \psi)\}$ . On définit enfin un ensemble fini de mots infinis ultimement périodiques  $\text{Mot}_{\phi, \psi, a} = \{x^{(0)}x^{(1)}x^{(2)} \dots \mid x^{(i)} \in \mathcal{A}^+ \text{ et } \exists \sigma \in \mathcal{S}_{St}^*$ , la suite des puissances de  $\sigma$  s'identifie à  $(\phi, \psi)$  et  $x^{(0)} \in \overline{\sigma(a)}$ ,  $x^{(i)} \in \Sigma_i^{(\phi, \psi)}(u_{\sigma, a})$  pour  $i \geq 1\}$ .

Après toutes ces constructions intermédiaires, je vais enfin conclure : soit  $\sigma \in \mathcal{S}_{St}^*$  et  $a \in \mathcal{A}$  tel que  $\sigma(a) = au_{\sigma, a}$  : par le théorème 1.1.19, ce mot s'écrit  $\sigma^\infty(a) = \sigma(a)\sigma^1(u_{\sigma, a})\sigma^2(u_{\sigma, a})\sigma^3(u_{\sigma, a}) \dots$ . Par congruence infinie de  $\equiv$ ,  $\sigma^\infty(a)$  est accepté par  $\mathfrak{A}$  ssi  $x = x^{(0)}x^{(1)}x^{(2)} \dots$  est accepté par  $\mathfrak{A}$  avec  $x^{(0)} \equiv \sigma(a)$  et  $x^{(i)} \equiv \sigma^i(u_{\sigma, a})$ . Soit  $(\phi, \psi) \in \mathcal{M}$  tel que la suite des puissances de  $\sigma$  s'identifie à  $(\phi, \psi)$  :  $\sigma^\infty(a)$  est accepté par  $\mathfrak{A}$  ssi  $x = x^{(0)}x^{(1)}x^{(2)} \dots$  est accepté par  $\mathfrak{A}$  avec  $x^{(0)} \equiv \sigma(a)$  et  $x^{(i)} \in \Sigma_{i-1}^{(\phi, \psi)}(u_{\sigma, a})$ . Donc  $\sigma^\infty(a)$  est accepté par  $\mathfrak{A}$  ssi le mot infini  $x$  de  $\text{Mot}_{\phi, \psi, a}$  vérifiant  $x^{(0)} \in \overline{\sigma(a)}$  est accepté par  $\mathfrak{A}$ . Donc  $\mathfrak{A}$  accepte un mot sturmien purement substitutif ssi il accepte l'un des mots ultimement périodique inclus dans l'un des  $\text{Mot}_{\phi, \psi, a}$ .

L'algorithme est alors le suivant : on construit l'ensemble fini  $E = \bigcup_{(\phi, \psi) \in \text{Seq}_{\mathcal{M}}, a \in \mathcal{A}} \text{Mot}_{\phi, \psi, a}$ . Dans cet ensemble fini, il n'y a que des mots ultimement périodique : on vérifie si l'un des éléments de  $E$  est ac-

10. Il existe des semi-groupes finis annulables à gauche qui ne sont pas des groupes : pour un ensemble fini  $E$ , si on considère la loi  $x + y = y$ , l'ensemble  $(E, +)$  est un semi-groupe fini annulable à gauche mais qui n'est pas un groupe, car  $x \neq y \implies x + y \neq y + x$ .

cepté par  $\mathfrak{A}$  (l'acceptation d'un mot ultimement périodique par un  $\omega$ -automate de Büchi est décidable). Si l'un d'entre eux est accepté,  $\mathfrak{A}$  accepte un mot sturmien purement substitutif; sinon,  $\mathfrak{A}$  n'accepte aucun mot sturmien purement substitutif.

Le résultat précédent, bien que résultant d'une preuve longue et complexe, est en fait plutôt décevant. Les deux hypothèses, celle de déterminisme et de co-complétude, sont particulièrement coûteuses. Les  $\omega$ -automates de Büchi déterministes sont par exemple un modèle strictement plus faible d' $\omega$ -automates que celui des  $\omega$ -automates de Büchi; donc, déjà, il n'est pas possible de directement généraliser le résultat obtenu ici au cas des  $\omega$ -automates de Büchi. La condition supplémentaire de co-complétude restreint encore plus le modèle : je ne sais pas s'il existe un  $\omega$ -automate de Büchi déterministe co-complet qui n'est équivalent, en langage, à aucun  $\omega$ -automate faible. L'objectif principal de la preuve précédente est de proposer une première piste pour de futures recherches sur le sujet.

### 2.3.6 Croyons en nos rêves

Comme vu précédemment, la décidabilité de l'acceptation d'un mot Sturmien par un  $\omega$ -automate de Büchi, de manière générale, ne découle pas directement de la désubstitution. En faisant des hypothèses supplémentaires, il est possible toutefois de trouver des méthodes algorithmiques. Ici, on fait une hypothèse sur la topologie du calcul acceptant :

**Proposition 2.3.18.** *Le problème suivant est décidable :*

**Donnée :** *un  $\omega$ -automate de Büchi-arêtes  $\mathfrak{A}$*

**Question :** *est-ce que  $\mathfrak{A}$  accepte un mot sturmien dont le calcul a une arête acceptante uniformément récurrente ?*

*Démonstration.*

Soit  $\mathfrak{A} = (\mathcal{A}, Q, I, T, F)$  un  $\omega$ -automate de Büchi-arêtes. Quitte à décrire  $\mathfrak{A}$  comme l'union d' $\omega$ -automates de Büchi-arêtes distincts, on peut supposer que  $|F| = 1$  : on note  $F = \{f\}$ . Supposons maintenant que  $\mathfrak{A}$  accepte un mot sturmien  $s$  accepté par un calcul  $c$  où  $f$  est uniformément récurrente. Soit  $(\sigma_n)_{n \in \mathbb{N}}$  la représentation S-adique d'Arnoux associée à  $s$ . Pour un morphisme  $\tau$ , on peut alors désubstituer  $c$  comme  $\tau^{-1}(c)$  qui est un calcul acceptant de  $\tau^{-1}(\mathfrak{A})$ . Soit  $r$  la constante d'uniforme récurrence associée à  $f$  dans  $c$  : dans toute fenêtre de taille au moins  $r$  dans  $c$ , on trouve au moins une occurrence de  $f$ . Par ubicroissance de la représentation S-adique d'Arnoux d'un mot sturmien, on sait que pour tout  $n$  assez grand,  $\min_{a \in \mathcal{A}} |\sigma_{[0,n]}^{-1}(a)| > r$ . Alors  $\sigma_{[0,n]}^{-1}(c)$  est un calcul de  $\sigma_{[0,n]}^{-1}(\mathfrak{A})$  acceptant, dont toutes les transitions sont acceptantes. En conclusion, si  $\mathfrak{A}$  accepte un mot sturmien  $s$  accepté par un calcul  $c$  où  $f$  est uniformément récurrente, alors il existe un désubstitué de  $\mathfrak{A}$  qui accepte un mot sturmien en ne lisant que des arêtes acceptantes. Réciproquement, s'il existe un morphisme sturmien  $\tau$  tel que  $\tau^{-1}(\mathfrak{A})$  accepte un mot sturmien en ne lisant que des arêtes acceptantes, alors  $\mathfrak{A}$  accepte l'image de ce mot par  $\tau$ .

L'algorithme est alors le suivant : on construit un  $\omega$ -automate de désubstitution de  $\mathfrak{A}$ ,  $\mathcal{S}_{St}^{-\infty}(\mathfrak{A}) = (\mathcal{S}_{St}, \mathfrak{S}(A), \{\mathfrak{A}\}, T)$ , où :

- $\mathfrak{S}(\mathfrak{A})$  est l'ensemble de tous les  $\omega$ -automates de Büchi-arêtes partageant les états et états initiaux de  $\mathfrak{A}$ ;
- pour tout  $\mathfrak{B} \in \mathfrak{S}(A)$  et  $\sigma \in \mathcal{S}_{St}$ ,  $B \xrightarrow{\sigma} \sigma^{-1}(\mathfrak{B}) \in T$ .

On modifie maintenant cet  $\omega$ -automate pour construire  $\mathcal{S}_{St}^{-\infty, UR}(\mathfrak{A})$  en retirant de tous les états (qui

sont des  $\omega$ -automates de Büchi-arêtes) les arêtes non acceptantes. On vérifie maintenant s'il existe un sommet qui accepte un mot sturmien avec les arêtes restantes. Cette vérification se fait avec l'algorithme connu sur les  $\omega$ -automates faibles.

Ce résultat, s'il peut paraître assez spécifique, a en fait sûrement du potentiel. J'ai en effet essayé de construire des contre-exemples d'automates de Büchi qui acceptent un mot sturmien, mais sans calcul uniformément récurrent et ai échoué. En raison de l'immensité de l'espace à explorer, je n'ai pas pu automatiser la recherche de contre-exemple : mais aucun  $\omega$ -automate de Büchi à moins de 3 états n'accepte un mot sturmien sans que le calcul associé ne soit uniformément récurrent. Le calcul résultant en un mot sturmien est, dans le shift plein  $T^{\mathbb{N}}$ , un antécédent par un codage : et je pense que cette propriété, en lien avec la structure d'un  $\omega$ -automate de Büchi, contraint l'un des calculs à être uniformément récurrent. Je n'ai toutefois pas réussi à écrire une preuve, et je laisse donc cette piste à l'état de conjecture.

**Conjecture 2.3.19.** *Soit  $\mathcal{A}$  un  $\omega$ -automate de Büchi qui accepte un mot sturmien. Alors  $\mathcal{A}$  accepte un mot sturmien dont le calcul a une arête acceptante uniformément récurrente.*

## 2.4 Conclusion

Dans ce chapitre, j'ai établi de nouvelles méthodes pour la décidabilité du problème de l'acceptation d'un mot par un  $\omega$ -automate tournées autour de la désubstitution. En particulier, dans le cas des  $\omega$ -automates faibles, j'ai établi que les problèmes de l'acceptation d'un mot purement substitutif, morphique, d'un point fixe ou d'un mot sturmien étaient tous décidables. Les méthodes algorithmiques sont combinatoires, et les preuves de correction sont topologiques. J'ai généralisé ces résultats aux représentations S-adiques; j'ai aussi pu établir un nouveau lien entre  $\omega$ -automates et mots sturmiens avec des résultats de totalité, prouvant une forme d'orthogonalité entre le modèle automatique et le modèle sturmien. À partir de ces constructions, j'ai pu étudier plusieurs questions de combinatoire des mots, de dynamique symbolique et de géométrie discrète. Ensuite, j'ai essayé d'étendre mes méthodes au cas des  $\omega$ -automates de Büchi. Dans ce cadre, j'ai fourni des pistes partielles de résolutions, et prouvé que la désubstitution avait le potentiel de répondre à la question.

La question des  $\omega$ -automates faibles est presque résolue. Il reste la question de la complexité de l'algorithme proposé : les bornes que je donne sont particulièrement grossières. Une question que j'ai étudiée sans succès est la taille maximale du méta- $\omega$ -automate de désubstitution : par exemple, parmi les  $\omega$ -automates faibles avec 4 états, en quotientant par le langage mais sans minimiser, le plus grand méta- $\omega$ -automate de désubstitution compte 53 états. Pourrait-on trouver une borne plus efficace? Concernant les applications, plusieurs questions restent en suspens. Par exemple, concernant le codage des mots sturmiens, la méthode que je rends possible ici peut paraître disproportionnée : existe-t-il un algorithme plus simple? Aussi, je ne donne qu'un critère de décidabilité : existe-t-il une caractérisation combinatoire complète des mots codant un mot sturmien? Pour la géométrie discrète, la question des droites discrètes dans  $\mathbb{Z}^d$  avec  $d \geq 3$  reste ouverte. Est-il possible de réduire la question au cas des droites dans  $\mathbb{Z}^2$ ?

Ce chapitre, ambitieux dans ses objectifs, n'a en fait pas tout à fait atteint sa cible : celle des  $\omega$ -automates de Büchi. Si la conjecture 2.3.19 devait être avérée, cela permettrait alors de nettement améliorer les résultats théoriques derrière le logiciel Walnut, en prouvant la décidabilité des propriétés monadiques du second ordre sur les mots sturmiens. De plus, les méthodes développées dans ce chapitre, combinatoires et reposant sur les représentations S-adiques, sont plus élémentaires que la théorie sur laquelle s'appuie Pecan, à savoir des systèmes de numération non standard et des algorithmiques de fractions continues.



## 3 Pièges à facteurs pour les mots bi-infinis

Quand tous les calculs compliqués s'avèrent faux, quand les philosophes eux-mêmes n'ont plus rien à nous dire, il est excusable de se tourner vers le babillage fortuit des oiseaux, ou vers le lointain contrepois des astres.

---

Marguerite Yourcenar, *Les Mémoires d'Hadrien*

### 3.1 Introduction

Ce chapitre traite une partie assez indépendante du reste de la thèse. Le travail qui y est présenté est une collaboration avec Hellouin et Gheeraert.

Les pièges à facteurs, ou *string attractors* en anglais, sont une notion introduite par Prezza en 2017 [Pre17] et étudiée de manière plus étendue par Prezza et Kempa [KP18]. À l'origine, la notion a été créée dans le contexte de l'algorithmique du texte, afin de pouvoir comparer différents algorithmes de compression et fournir une nouvelle mesure de la répétitivité d'un mot. Kempa et Prezza ont notamment montré que le calcul du piège à facteurs de taille minimale d'un mot est un problème NP-complet.

Ce concept a toutefois beaucoup intéressé la communauté de recherche en combinatoire des mots, car les pièges à facteurs peuvent être interprétés comme un nouvel outil combinatoire. Les pièges à facteurs ont originellement été développés pour étudier des mots finis, mais il est toutefois possible de généraliser la notion de piège à facteurs aux mots infinis. Le résultat obtenu est néanmoins décevant : Restivo, Romana et Sciortino ont montré que si un mot infini admet un piège à facteurs fini, alors ce mot est ultimement périodique (proposition 7 de [RRS22]). Cette contrainte a motivé l'étude des pièges à facteurs des préfixes d'un mot infini, par exemple dans le cas des suites automatiques [SS21], le cas du mot de Thue-Morse [Dol23] ou celui des mots sturmiens et épisturmiens [Dvo24]. Ce résultat s'applique au cas des mots *mono*-infinis, c'est-à-dire les mots indexés par  $\mathbb{N}$ . Une question naturelle est alors le cas des mots *bi*-infinis, c'est-à-dire les mots indexés par  $\mathbb{Z}$ . En combinatoire des mots, les deux cas ont souvent des comportements proches, comme la décidabilité du problème du domino [New77] ou l'existence d'une caractérisation S-adique des mots sturmiens [Pyt+02]. Les cas mono-infinis et bi-infinis ont cependant parfois des variations, comme le théorème de Morse-Hedlund [MH38] ou la caractérisation par complexité des mots sturmiens [Pyt+02].

Ce chapitre cherche ainsi à traiter le cas bi-infini : que dire des mots bi-infinis admettant un piège à facteurs fini ? Je montre que dans le cas bi-infini, le résultat de Restivo, Romana et Sciortino n'est plus valable : il existe des mots bi-infinis apériodiques admettant des pièges à facteurs finis. Les méthodes développées dans ce chapitre l'inscrivent dans le fil de cette thèse : les preuves utilisent en effet les propriétés combinatoires des représentations S-adiques des mots sturmiens. J'introduirai en particulier dans ce chapitre une notion de désubstitution des pièges à facteurs, qui jouera un rôle-clé dans la caractérisation complète des mots bi-infinis admettant un piège à facteurs fini. Ce chapitre illustre l'utilisation de deux notions (les pièges à facteurs et les représentations S-adiques) d'origines éloignées et dont

L'utilisation combinatoire n'a pas été entièrement explorée par la littérature. Les résultats présentés dans ce chapitre permettent, de plus, d'établir une nouvelle caractérisation combinatoire des mots sturmiens.

Dans une première partie, j'introduis les définitions nécessaires à la compréhension des objets et des résultats évoqués, ainsi que des résultats préliminaires sur la manipulation des pièges à facteurs. Puis je procède à une caractérisation des mots bi-infinis admettant  $\llbracket 0, 1 \rrbracket$  comme piège à facteurs : il s'agit exactement des mots sturmiens caractéristiques bi-infinis. Ensuite, je m'intéresse au cas des pièges à facteurs finis arbitraires ; à partir d'une étude des mots de complexité  $n + k$  pour  $n$  assez grand — appelés mots quasi-sturmiens —, je prouve que seules les images des mots sturmiens caractéristiques bi-infinis admettent des pièges à facteurs finis. J'établis de plus pour ces une relation entre la taille minimale d'un piège à facteurs qu'ils admettent et leur complexité. Enfin, je traite deux extensions sur les pièges à facteurs de mots bi-infinis : la première concerne l'étude des pièges à facteurs infinis, leur densité et le cas des progressions arithmétiques. La seconde étend la notion de piège à facteurs dans le cas des sous-shifts.

## 3.2 Définitions

D'abord, donnons la définition d'un piège à facteurs d'un mot bi-infini.

**Définition 3.2.1** (Piège à facteurs). *Soit  $x \in \mathcal{A}^{\mathbb{Z}}$  un mot bi-infini. Soit  $\Gamma \subseteq \mathbb{Z}$  et  $f$  un facteur non vide de  $x$ . On dit que  $\Gamma$  capture  $f$  s'il existe  $i \in \mathbb{Z}$  tel que  $x_{\llbracket i, i+|f|-1 \rrbracket} = f$  et  $\llbracket i, i+|f|-1 \rrbracket \cap \Gamma \neq \emptyset$ ; on dit que  $\Gamma$  est un piège à facteurs de  $x$  s'il capture tous les facteurs non vides de  $x$ .*

**Exemple 3.2.2.** *Soit  $x = \dots 000 \cdot 111 \dots$  un mot bi-infini ultimement périodique à gauche et à droite. Posons  $\Gamma = \llbracket -1, 0 \rrbracket$ , et montrons que  $\Gamma$  est un piège à facteurs de  $x$ . Soit  $f$  un facteur non vide de  $x$ . Alors  $f$  est de la forme  $f = 0^i 1^j$ , avec  $i+j \geq 1$ . On a en particulier  $x_{\llbracket -i, j-1 \rrbracket} = f$ . La condition  $i+j \geq 1$  assure que  $\llbracket -i, j-1 \rrbracket \cap \llbracket -1, 0 \rrbracket \neq \emptyset$ . Donc  $\Gamma$  capture tout facteur de  $x$  :  $\Gamma$  est bien un piège à facteurs de  $x$ .*

*En revanche,  $\Gamma' = \mathbb{N}$  ne capture pas tout facteur de  $x$  : le facteur  $0$  n'a pas d'occurrence capturée par  $\mathbb{N}$ . Et  $\Gamma'' = \mathbb{Z}$  est un piège à facteurs de  $x$ , et est même un piège à facteurs pour tout mot bi-infini.*

Des définitions analogues existent dans le cas des mots finis et des mots mono-infinis. Le résultat le plus marquant de ce chapitre est en effet le contraste entre le cas des mots mono-infinis et celui des mots bi-infinis établi par Restivo, Romana et Sciortino.

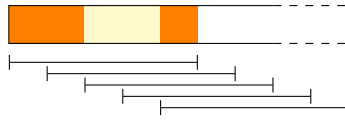
**Théorème 3.2.3** ([RRS22, Proposition 7]). *Soit  $x$  un mot mono-infini. Alors  $x$  admet un piège à facteurs de taille finie si, et seulement si,  $x$  est un mot ultimement périodique.*

*Démonstration.*

Soit  $x = uv^{\mathbb{N}}$  un mot ultimement périodique de pré-période  $u$  et de période  $v$  (avec  $v \neq \varepsilon$ ). Posons  $\Gamma = \llbracket 0, |u| + |v| - 1 \rrbracket$ . Soit  $f = x_{\llbracket n, n+|f|-1 \rrbracket}$  un facteur de  $x$ . Si  $n < |u| + |v|$ , alors  $f$  est capturé par  $\Gamma$ . Sinon, ou bien  $f$  est un facteur de  $v$  (et alors il est déjà capturé par  $\Gamma$ ), ou bien  $f$  s'écrit comme  $v_s v^k v_p$  avec  $v_s$  un suffixe de  $v$ ,  $k \geq 0$  et  $v_p$  un préfixe de  $v$ . Alors  $x_{\llbracket |u|+|v|-|v_s|, |u|+k|v|+|v_p|-1 \rrbracket} = f$  par ultime périodicité de  $x$ . Comme  $\Gamma \cap \llbracket |u| + |v| - |v_s|, |u| + k|v| + |v_p| - 1 \rrbracket \neq \emptyset$ ,  $\Gamma$  est un piège à facteurs de  $x$ .

Soit maintenant  $x$  un mot mono-infini admettant un piège à facteurs  $\Gamma$  fini. Quitte à rajouter des positions internes, on suppose que  $\Gamma$  est un segment de  $\mathbb{N}$ . Soit  $M = \max(\Gamma)$ . Comme  $\Gamma$  est un piège à facteurs de  $x$ ,  $\Gamma$  capture tous les facteurs de longueur  $M + 1$  de  $x$ . Donc si  $f$  est un facteur de  $x$  de longueur  $M + 1$ , il existe  $0 \leq n \leq M$  tel que  $x_{\llbracket n, n+M \rrbracket} = f$ . Donc  $x$  admet au plus  $M + 1$  facteurs

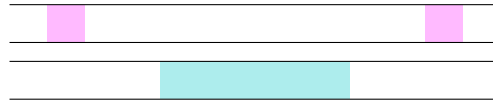
distincts de longueur  $M + 1$ . D'après le théorème de Morse-Hedlund dans le cas mono-infini,  $x$  est alors un mot ultimement périodique.



**FIGURE 3.1** – En orange, les positions initiales de  $\Gamma$  ; en jaune, les positions internes à  $\Gamma$  qu'on lui ajoute pour simplifier la preuve. En dessous, les traits représentent les occurrences possibles de facteurs de longueur  $M + 1$  capturées par  $\Gamma$ .

On remarque en particulier qu'il est facile de trouver un piège à facteurs d'un mot ultimement périodique à partir d'une pré-période et d'une période. La question de la minimisation du piège à facteurs d'un mot ultimement périodique semble intéressante, mais reposera plus sûrement sur l'étude des pièges à facteurs de mots finis : le cas des pièges à facteurs *circulaires* [Man+21] semble un bon point d'approche de la question. Dans le cadre général de la combinatoire des mots, la plupart des mots étudiés classiquement ne sont toutefois pas ultimement périodiques. C'est pour cela que la littérature ([SS21; Dol23; Dvo24]) se concentre sur l'étude de *fonctions de profil* de pièges à facteurs : pour  $x$  un mot mono-infini, il s'agit de  $s_x(n) = \min\{\text{Card}(\Gamma) \mid \Gamma \subseteq \llbracket 0, n \rrbracket \text{ est un piège à facteurs de } x_{\llbracket 0, n \rrbracket}\}$ . En cela, la démarche présentée ici diffère par rapport à la littérature : je vais garder l'idée d'un piège à facteurs pour un mot infini, je vais seulement modifier *quel* infini. Dans ce chapitre, je vais principalement traiter le cas des pièges à facteurs finis. Il sera alors important de manipuler la taille d'un piège à facteurs. Plus que la taille, la notion qui s'articulera bien avec nos résultats est celle du *diamètre* d'un piège à facteurs (en anglais le *span*) :

**Définition 3.2.4** (Diamètre). Soit  $\Gamma \subseteq \mathbb{Z}$ . Le diamètre de  $\Gamma$ , noté  $\text{diam}(\Gamma)$ , est  $\text{diam}(\Gamma) = \sup(\Gamma) - \inf(\Gamma)$ . En particulier, si  $\Gamma$  est un ensemble infini,  $\text{diam}(\Gamma) = \infty$ .



**FIGURE 3.2** – Le piège en rose contient moins de positions, mais a un diamètre plus grand que le piège en bleu.

L'idée est alors qu'un petit piège à facteurs est un piège « compact ». En pratique, on considérera alors toujours les pièges à facteurs finis comme des intervalles : on peut toujours transformer un piège à facteurs fini en intervalle de  $\mathbb{Z}$  sans changer son diamètre. Enfin, on peut comparer les pièges à facteurs possibles d'un mot en introduisant la notion d'envergure d'un mot bi-infini (en anglais *span*, de nouveau [SS21]) :

**Définition 3.2.5** (Envergure). Soit  $x \in \mathcal{A}^{\mathbb{Z}}$  un mot bi-infini. Alors l'envergure de  $x$  est  $\text{env}(x) = \inf\{\text{diam}(\Gamma) \mid \Gamma \text{ piège à facteurs de } x\}$ . En particulier,  $\text{env}(x) = \infty$  si, et seulement si,  $x$  n'admet aucun piège à facteurs fini.

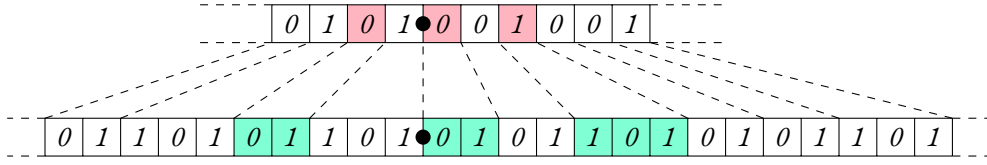
Je parlerai donc du diamètre d'un piège à facteurs ou d'un sous-ensemble de  $\mathbb{Z}$ , et de l'envergure d'un mot. Il me sera aussi nécessaire, dans la suite de ce chapitre, de manipuler les pièges à facteurs et de les transformer via des substitutions (voir la définition 1.1.13). En particulier, j'ai besoin de définir l'image et la préimage d'un sous-ensemble de  $\mathbb{Z}$  par une substitution.

**Définition 3.2.6** (Image d'un piège à facteurs). Soit  $\Gamma \subseteq \mathbb{Z}$ , soit  $x$  un mot bi-infini et soit  $\phi$  une substitution. On note alors  $\phi_x(\Gamma)$  l'ensemble des supports des images des éléments de  $\Gamma$  dans  $\phi(x)$ , c.-à-d. :

$$\phi_x(\Gamma) = \left( \bigcup_{i \in \Gamma \cap \mathbb{N}} \llbracket |\phi(x_{[0, i-1]})|, |\phi(x_{[0, i]})| - 1 \rrbracket \right) \cup \left( \bigcup_{i \in \Gamma \cap \mathbb{Z} \setminus \mathbb{N}} \llbracket -|\phi(x_{[i, -1]})|, -|\phi(x_{[i+1, -1]})| - 1 \rrbracket \right)$$

On notera  $\phi(\Gamma)$  lorsque, dans le contexte,  $x$  est clairement établi.

**Exemple 3.2.7.** Soit  $x = \dots 010101 \cdot 001001001 \dots$ ,  $\phi : 0 \mapsto 01, 1 \mapsto 101$  et  $\Gamma = \{-2, 0, 2\}$ . Alors  $\phi(x) = \dots 011010110101101 \cdot 010110101011010101101 \dots$  et  $\phi(\Gamma) = \{-5, -4, 0, 1, 4, 5, 6\}$ .



**FIGURE 3.3** – En haut, le mot bi-infini  $x$ ; en bas, son image  $\phi(x)$ . En rose,  $\Gamma$ ; son image  $\phi(\Gamma)$  est en aigue-marine.

**Définition 3.2.8** (Préimage d'un piège à facteurs). Soit  $\Gamma \subseteq \mathbb{Z}$ , soit  $x$  un mot bi-infini et soit  $\phi$  une substitution. On note alors  $\phi_x^{-1}(\Gamma)$  l'ensemble des positions de  $x$  dont l'image dans  $\phi(x)$  intersecte  $\Gamma$  :

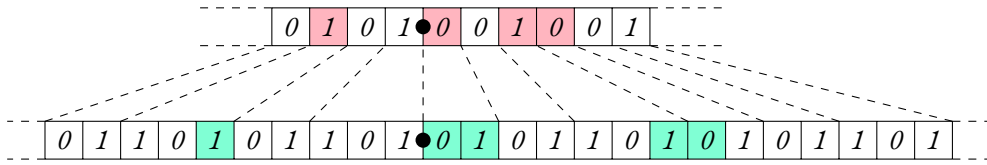
$$\begin{aligned} \phi_x^{-1}(\Gamma) = & \{i \geq 0 \mid [|\phi(x_{[0,i-1]})|, |\phi(x_{[0,i]})| - 1] \cap \Gamma \neq \emptyset\} \cup \\ & \{i < 0 \mid [ -|\phi(x_{[i,-1]})|, -|\phi(x_{[i+1,-1]})| - 1] \cap \Gamma \neq \emptyset\} \end{aligned}$$

De manière équivalente, c'est le plus petit (au sens de l'inclusion) sous-ensemble de  $\mathbb{Z}$  dont l'image par  $\phi$  contient  $\Gamma$  :

$$\phi_x^{-1}(\Gamma) = \inf\{\Delta \subseteq \mathbb{Z} \mid \phi_x(\Delta) \supseteq \Gamma\}$$

On notera  $\phi^{-1}(\Gamma)$  lorsque, dans le contexte,  $x$  est clairement établi.

**Exemple 3.2.9.** Soit  $x = \dots 010101 \cdot 001001001 \dots$ ,  $\phi : 0 \mapsto 01, 1 \mapsto 101$ ; l'image de  $x$  par  $\phi$  est  $\phi(x) = \dots 011010110101101 \cdot 010110101011010101101 \dots$ . Si on pose  $\Gamma = \{-6, 0, 1, 6, 7\}$ , alors  $\phi^{-1}(\Gamma) = \{-3, 0, 2, 3\}$ .

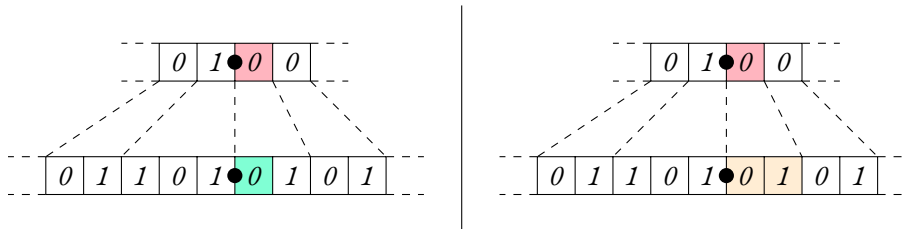


**FIGURE 3.4** – En haut, le mot bi-infini  $x$ ; en bas, son image  $\phi(x)$ . En aigue-marine,  $\Gamma$ ; en rose, sa préimage  $\phi^{-1}(\Gamma)$ .

Les deux notions définies ici donnent lieu aux deux relations suivantes :

$$\Gamma = \phi_x^{-1}(\phi_x(\Gamma)) \text{ et } \Gamma \subseteq \phi_x(\phi_x^{-1}(\Gamma))$$

**Exemple 3.2.10.** La deuxième inclusion peut être stricte : si  $x = \dots 010101 \cdot 001001001 \dots$  et  $\phi : 0 \mapsto 01, 1 \mapsto 101$ , pour  $\Gamma = \{0\}$ , on a  $\phi_x^{-1}(\Gamma) = \{0\}$  et  $\phi_x(\phi_x^{-1}(\Gamma)) = \{0, 1\} \neq \Gamma$ .





**FIGURE 3.5** – À gauche, une visualisation de pourquoi  $\phi_x^{-1}(\Gamma) = \{0\}$ ; à droite, une visualisation de pourquoi  $\phi_x(\phi_x^{-1}(\Gamma)) = \{0, 1\}$ .

L'une des principales propriétés relatives aux substitutions est que l'image d'un piège à facteurs d'un mot est un piège à facteurs pour l'image du mot :

**Proposition 3.2.11.** *Soit  $x$  un mot bi-infini, soit  $\Gamma$  un piège à facteurs de  $x$  et soit  $\phi$  une substitution. Alors  $\phi(\Gamma)$  est un piège à facteurs pour  $\phi(x)$ .*

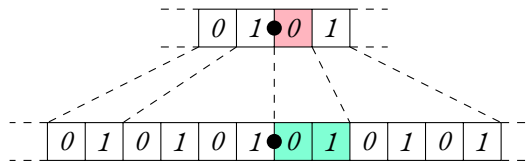
*De plus,  $\text{env}(\phi(x)) \leq (\text{env}(x) + 1) \times \max_{a \in \mathcal{A}} |\phi(a)| - 1$ .*

*Démonstration.*

Soit  $f$  un facteur de  $\phi(x)$ . Par définition,  $f$  est un facteur d'un certain  $\phi(u)$  avec  $u$  un facteur de  $x$ . On suppose que  $u$  est choisi minimal, c.à-d. que l'occurrence de  $f$  dans  $\phi(u)$  débute dans  $\phi(u_0)$  et termine dans  $\phi(u_{|u|-1})$ . Comme  $\Gamma$  est un piège à facteurs pour  $x$ ,  $\Gamma$  capture une occurrence de  $u$ . Donc, par définition de  $\phi(\Gamma)$ ,  $\phi(\Gamma)$  capture une occurrence de  $\phi(u)$ . Par minimalité de  $u$ ,  $f$  est donc bien capturé par  $\phi(\Gamma)$ . L'inégalité provient simplement d'une majoration de la taille de  $\phi(\Gamma)$  par le produit de la taille de  $\Gamma$  et de la taille de la plus grande image d'une lettre par  $\phi$ .

Un énoncé analogue à la proposition 3.2.11 avec l'opération de désubstitution est malheureusement faux dans le cas général : si  $\Gamma$  est un piège à facteurs de  $\phi(x)$ ,  $\phi^{-1}(\Gamma)$  n'est pas nécessairement un piège à facteurs de  $x$ .

**Exemple 3.2.12.** *Soit  $x = \dots 010101 \cdot 010101 \dots$  et  $\phi : 0 \mapsto 01, 1 \mapsto 0101$ . On remarque alors que  $\phi(x) = x = \dots 010101 \cdot 010101 \dots$ . Alors  $\Gamma = \{0, 1\}$  est un piège à facteurs de  $\phi(x)$ , mais  $\phi^{-1}(\Gamma) = \{0\}$  n'est pas un piège à facteurs de  $x$  (par exemple,  $\phi^{-1}(\Gamma)$  ne capture pas le facteur 1).*



**FIGURE 3.6** – En bas,  $\{0, 1\}$  est bien un piège à facteurs de  $\phi(x) = x$ ; mais en désubstituant par  $\phi$ , on tombe sur  $\{0\}$ , qui n'est pas un piège à facteurs de  $x$ .

Paradoxalement, c'est tout de même cette réciproque qui nous intéresse. On remarque notamment que dans l'exemple 3.2.12, la substitution n'est pas reconnaissable sur son point fixe (voir [BPR23] pour une explication plus approfondie sur la reconnaissabilité d'une substitution) : c'est donc principalement l'inadéquation de la substitution qui empêche de créer un piège à facteurs en désubstituant. Je démontrerai que la désubstitution préserve les pièges à facteurs dans le cas de certains morphismes spécifiques, ou sous certaines conditions supplémentaires. C'est en exploitant ces résultats partiels que je pourrai établir des bornes optimales liant complexité et envergure d'un mot.

## 3.3 Existence et caractérisation des mots bi-infinis admettant un piège à facteurs fini

### 3.3.1 Quels sont les candidats possibles ?

Pour commencer à étudier les pièges à facteurs finis, il faut commencer par réduire l'espace de recherche à explorer. C'est là qu'intervient la proposition 3.3.1 ci-après : elle permet d'obtenir des contraintes fortes sur les mots qui

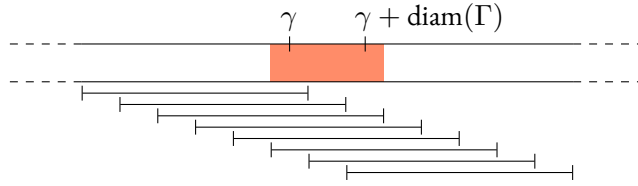
pourraient admettre un piège à facteurs fini. Elle permet en particulier de relier deux notions qui a priori émergent de questions combinatoires différentes : la complexité d'un mot (voir la définition 1.1.4) et l'envergure d'un mot bi-infini. Une version faible du résultat suivant a été démontré dans le cas des mots finis dans [Pre17]. Je fournis ici une nouvelle preuve d'un résultat plus fin dans le cas des mots bi-infinis (qui est aussi valable dans le cas mono-infini).

**Proposition 3.3.1.** *Pour tout mot bi-infini  $x$ , en notant  $p_x$  sa fonction de complexité (voir la définition 1.1.4), on a*

$$\forall n \geq 1, p_n(x) \leq n + \text{env}(x)$$

*Démonstration.*

Si  $\text{env}(x) = \infty$ , la proposition est vraie. Autrement, soit  $\Gamma$  un piège à facteurs fini de  $x$  de diamètre minimal :  $\text{diam}(\Gamma) = \text{env}(x)$ . Quitte à ajouter des valeurs, on suppose que  $\Gamma$  est un intervalle :  $\Gamma = \llbracket \gamma, \gamma + \text{diam}(\Gamma) \rrbracket$ . Soit  $n \geq 1$  et  $f \in \mathcal{L}_n(x)$ . Alors, par hypothèse,  $f$  est capturé par  $\Gamma$  : donc il a une occurrence commençant dans  $\llbracket \gamma - n + 1, \gamma + \text{diam}(\Gamma) \rrbracket$ . Donc  $p_x(n) \leq \gamma + \text{diam}(\Gamma) - (\gamma - n + 1) + 1 = n + \text{diam}(\Gamma) = n + \text{env}(x)$ .



**FIGURE 3.7** – Visualisation de la preuve : les traits représentent les facteurs de longueur  $n$  capturés par  $\Gamma$ , et qui doivent donc être les seuls facteurs distincts de  $x$ .

On en déduit que seuls des mots ayant une complexité de la forme  $p_x(n) \leq n + k$  peuvent admettre des pièges à facteurs finis. On peut alors traiter un premier cas simple : celui où le mot  $x$  n'est pas récurrent.

**Théorème 3.3.2** ([Cov75]). *Soit  $x$  un mot bi-infini tel qu'il existe  $k \geq 1$  vérifiant  $p_x(n) \leq n + k$  pour tout  $n \geq 1$ . Supposons que  $x$  n'est pas récurrent. Alors  $x$  est ultimement périodique à gauche et à droite.*

*Démonstration.*

Quitte à décaler, on suppose que  $x_{\llbracket 0, n \rrbracket}$  est un facteur de  $x$  qui n'apparaît qu'une seule fois. Alors tout facteur de  $x$  contenant  $x_{\llbracket 0, n \rrbracket}$  n'apparaît aussi qu'une seule fois dans  $x$ . Pour  $N \geq n$ , considérons alors les mots  $f_i = x_{\llbracket i - (N - n - 1), i + n \rrbracket}$  avec  $0 \leq i \leq N - n - 1$ . Ce sont des facteurs de  $x$  de longueur  $N$  qui contiennent le facteur  $x_{\llbracket 0, n \rrbracket}$ . Donc il n'apparaissent qu'une fois dans  $x$ . En particulier, considérons  $x' = x_{\llbracket 1, +\infty \rrbracket}$  : aucun  $f_i$  n'est facteur de  $x'$ . De plus, si  $i \neq j$ , alors  $f_i \neq f_j$  : sinon, on aurait deux occurrences de  $x_{\llbracket 0, n \rrbracket}$  dans  $x$ .

Si on calcule la fonction de complexité de  $x'$  (en tant que mot sur  $\mathbb{N}$ ), on a :

$$\begin{aligned} p_{x'}(N) &\leq p_x(N) - \text{Card}(\{f_i \mid 0 \leq i \leq N - n - 1\}) \\ &\leq N + k - (N - n - 1) \leq k + n + 1 \end{aligned}$$

Donc  $x'$  est un mot mono-infini de complexité bornée : d'après le théorème de Morse-Hedlund sur  $\mathbb{N}$ ,  $x'$  est donc ultimement périodique. Donc en particulier  $x$  est ultimement périodique à droite.

On peut faire un raisonnement symétrique à gauche, donc  $x$  est ultimement périodique à gauche aussi.

La construction d'un piège à facteurs pour un mot ultimement périodique à gauche et à droite est alors plutôt simple. Pour un tel  $x$ , quitte à décaler, on peut écrire  $x = v_g^{-\mathbb{N}} \cdot uv_d^{\mathbb{N}}$  avec  $v_g, u, v_d \in \mathcal{A}^+$ . Alors l'intervalle  $\llbracket -|v_g|, |u| + |v_d| - 1 \rrbracket$  est un piège à facteurs de  $x$  (voir preuve du théorème 3.2.3). Donc dans le cas non récurrent, le résultat se rapproche de ce qui est connu sur  $\mathbb{N}$ .

**Corollaire 3.3.3.** *Soit  $x$  un mot bi-infini non récurrent. Alors  $x$  admet un piège à facteurs fini si, et seulement si,  $x$  est ultimement périodique à gauche et à droite.*

Qu'en est-il dans le cas récurrent ? C'est ce cas qui est l'objet d'étude du reste de ce chapitre. D'abord, on peut montrer qu'un tel mot est non seulement récurrent, mais aussi uniformément récurrent.

**Théorème 3.3.4** ([Cov75]). *Soit  $x$  un mot bi-infini tel qu'il existe  $k \geq 1$  vérifiant  $p_x(n) \leq n + k$  pour tout  $n \geq 1$ . Si  $x$  est récurrent, alors  $x$  est uniformément récurrent.*

*Démonstration.*

On procède par l'absurde. Soit  $f$  un facteur de  $x$  tel qu'il existe des intervalles arbitrairement grands sans occurrence de  $f$ .

Comme  $x$  est récurrent, il existe  $n \geq |f| + 1$  et un facteur  $g_{|f|+k}$  de  $x$  de longueur  $n$  contenant exactement  $|f| + k$  occurrences de  $f$ . On peut de plus supposer que  $f$  est un préfixe de  $g_{|f|+k}$ .

Comme  $x$  n'est pas uniformément récurrent, il existe  $p \in \mathbb{Z}$  tel que  $x_{\llbracket p, p+n-1 \rrbracket}$  n'admet aucune occurrence de  $f$  : notons  $g_0 = x_{\llbracket p, p+n-1 \rrbracket}$ . Alors la suite  $(|x_{\llbracket p-i, p-i+n-1 \rrbracket}|_f)_{i \in \mathbb{Z}}$  vérifie :

$$||x_{\llbracket p-i, p-i+n-1 \rrbracket}|_f - |x_{\llbracket p-i-1, p-i-1+n-1 \rrbracket}|_f| \leq 1$$

Dans ce contexte,  $|m|_f$  désigne le nombre d'occurrences du facteur  $f$  dans le mot  $m$ . Cette inégalité découle du fait qu'en décalant la fenêtre de largeur  $n$  d'un cran à gauche, on fait apparaître une occurrence de  $f$ , disparaître une occurrence de  $f$ , ou les deux en même temps, mais rien d'autre : on ne peut pas faire apparaître d'un coup deux occurrences de  $f$ . L'existence du facteur  $m_{|f|+k}$  et la récurrence de  $x$  impliquent que le nombre d'occurrences atteindra  $|f| + k$ . Avec l'inégalité précédente, et en exploitant la récurrence de  $x$ , on peut utiliser une version discrète du théorème des valeurs intermédiaires et conclure, pour tout  $j \in \llbracket 0, |f| + k - 1 \rrbracket$ , à l'existence d'un facteur  $g_j$  de  $x$  de longueur  $n$  contenant exactement  $j$  occurrences de  $f$ . De plus, en utilisant ce procédé, chaque nouvelle valeur de  $j \neq 0$  est atteinte en ajoutant une nouvelle occurrence à gauche, c'est-à-dire comme préfixe : on peut alors supposer de plus que  $f$  est un préfixe des facteurs  $g_j$  pour  $j \neq 0$ .

Symétriquement à  $g_1$ , qui admet une seule occurrence de  $f$  comme préfixe, il existe  $r \in \mathbb{Z}$  tel que  $h_0 = x_{\llbracket r, r+n-1 \rrbracket}$  admette une seule occurrence de  $f$  comme suffixe :  $h_0$  étant trop long,  $h_0$  n'admet pas  $f$  comme préfixe. Notons alors  $h_i = x_{\llbracket r+i, r+i+n-1 \rrbracket}$  pour  $i \in \llbracket 0, n - |f| - 1 \rrbracket$ . Alors pour  $i \neq j$ ,  $h_i \neq h_j$  : sinon,  $h_0$  admettrait deux occurrences de  $f$ .

Enfin, aucun  $h_i$  n'admet  $f$  comme préfixe, avec  $i \in \llbracket 0, n - |f| - 1 \rrbracket$ . Donc tous les  $g_i$  et les  $h_j$  sont distincts. On a alors :

$$\begin{aligned} p_x(n) &\geq \text{Card}((g_i)_{0 \leq i \leq |f|+k}) + \text{Card}((h_j)_{0 \leq j \leq n-|f|-1}) \\ &\geq |f| + k + 1 + n - |f| = n + k + 1 \end{aligned}$$

Cela contredit la complexité initialement supposée de  $x$ .

Les théorèmes 3.3.2 et 3.3.4 donnent comme seuls candidats aperiodiques les mots quasi-sturmiens :

**Définition 3.3.5** (Mot quasi-sturmien). *Un mot bi-infini  $x$  est quasi-sturmien si  $x$  est a périodique et s'il existe  $n_0 \geq 0$  et  $k \geq 1$  tels que  $p_x(n) = n + k$  pour tout  $n \geq n_0$ .*

Cette classe a été caractérisée par Cassaigne dans le cas des mots mono-infinis [Cas97b]. Dans le cas des mots bi-infinis, la littérature parle plutôt des *words of minimal block growth* [Cov75; Bos85; Hei01]. Cette terminologie semble ne plus être utilisée à l'heure actuelle. C'est cette classe candidate qui sera l'objet d'étude du reste de cette section. Pour étudier cette question, je vais commencer par une sous-classe bien plus connue : celle des mots sturmiens.

### 3.3.2 Premier résultat d'existence

Pour traiter le cas des mots sturmiens, j'aurai besoin des notions de facteurs spéciaux à gauche, à droite et bispéciaux (voir définition 1.2.18). Dans le cadre bi-infini, la définition des mots sturmiens caractéristiques est la suivante :

**Définition 3.3.6** (Mot sturmien caractéristique bi-infini, [Pyt+02]). *Soit  $x$  un mot sturmien. On dit que  $x$  est un mot sturmien caractéristique si  $x = \lim_{n \rightarrow \infty} r_n(x) \cdot 01 \lim_{n \rightarrow \infty} l_n(x)$  ou si  $x = \lim_{n \rightarrow \infty} r_n(x) \cdot 10 \lim_{n \rightarrow \infty} l_n(x)$ .*

Le résultat sur lequel repose ce chapitre est alors le suivant :

**Proposition 3.3.7.** *Soit  $x \in \{0, 1\}^{\mathbb{Z}}$  un mot bi-infini de complexité  $p_x(n) = n + 1$ . Alors  $x$  admet  $\llbracket 0, 1 \rrbracket$  comme piège à facteurs si, et seulement si,  $x = \lim_{n \rightarrow \infty} r_n(x) \cdot 01 \lim_{n \rightarrow \infty} l_n(x)$  ou  $\lim_{n \rightarrow \infty} r_n(x) \cdot 10 \lim_{n \rightarrow \infty} l_n(x)$ .*

**Corollaire 3.3.8.** *Soit  $x \in \{0, 1\}^{\mathbb{Z}}$  un mot bi-infini a périodique de complexité  $p_x(n) = n + 1$ . Alors  $x$  admet  $\llbracket 0, 1 \rrbracket$  comme piège à facteurs si, et seulement si,  $x$  est un mot sturmien caractéristique.*

La preuve suivante de ce résultat est principalement l'œuvre de Gheeraert, dont je reprends ici les arguments.

*Démonstration de la proposition 3.3.7.*

D'abord, on remarque que  $\llbracket 0, 1 \rrbracket$  est un piège à facteurs de  $x$  si, et seulement si, pour tout  $n \geq 1$ ,  $x_{\llbracket -n+1, n \rrbracket}$  contient tout élément de  $\mathcal{L}_n(x)$ . De plus, comme  $p_x(n) = n + 1$ , tout élément de  $\mathcal{L}_n(x)$  apparaît alors exactement une fois dans  $x_{\llbracket -n+1, n \rrbracket}$ .

On introduit la propriété  $P_N$  :

$P_N$  : pour tout  $n \leq N$ , tout élément de  $\mathcal{L}_n(x)$  apparaît exactement une fois dans  $x_{\llbracket -n+1, n \rrbracket}$

Donc si  $\llbracket 0, 1 \rrbracket$  est un piège à facteurs de  $x$ , alors  $P_N$  est vraie pour tout  $N \geq 1$ . Comme  $x$  est a périodique et de complexité  $n + 1$ ,  $x$  admet exactement un facteur spécial à droite et un facteur spécial à gauche de toutes les longueurs, notés respectivement  $l_n(x)$  et  $r_n(x)$  pour la longueur  $n$ . On introduit la propriété  $Q_N$  :

$Q_N : x_{\llbracket -N+1, N \rrbracket} = r_{N-1}(x)01l_{N-1}(x)$  ou  $r_{N-1}(x)10l_{N-1}(x)$

Je montre maintenant que, pour tout  $N \geq 1$ ,  $P_N \iff Q_N$ . On commence par traiter le cas  $N = 1$  à part en remarquant que  $r_0(x) = l_0(x) = \varepsilon$  : il est donc clair que  $P_1 \iff Q_1$ .

**Lemme 3.3.9.**  $P_N \implies Q_N$ .

*Démonstration du lemme 3.3.9.*

Le cas  $N = 1$  étant déjà traité, je traite le cas  $N \geq 2$ .

D'abord, remarquons que si  $f \in \mathcal{L}_x(N-1)$ , il existe  $a \in \mathcal{A}$  tel que  $fa \in \mathcal{L}_x(N)$ . Donc, par  $P_N$ ,  $fa$  apparaît exactement une fois dans  $x_{[-N+1, N]}$ , donc  $f$  apparaît exactement une fois dans  $x_{[-N+1, N-1]}$ . En faisant le même raisonnement à gauche,  $f$  apparaît exactement une fois dans  $x_{[-N+2, N]}$  : on en déduit que  $f$  apparaît exactement une fois dans  $x_{[-N+2, N-1]}$ . Autrement écrit,  $P_N \implies P_{N-1}$ .

Par  $P_{N-1}$ , comme  $x_{[-N+2, N-1]}$  contient tout élément de  $\mathcal{L}_{N-1}(x)$  exactement une fois, c'est en particulier le cas de  $l_{N-1}(x)$ . Par définition,  $0l_{N-1}(x)$  et  $1l_{N-1}(x)$  sont aussi des facteurs de  $x$ , donc par l'hypothèse  $P_N$ , comme ce sont des éléments de  $\mathcal{L}_N(x)$ , ils apparaissent dans  $x_{[-N+1, N]}$ . Donc il y a deux occurrences de  $l_{N-1}(x)$  dans  $x_{[-N+2, N]}$ . Comme  $x_{[-N+2, N-1]}$  ne contient qu'une seule occurrence de  $l_{N-1}(x)$ , il en résulte que  $l_{N-1}(x)$  est un suffixe de  $x_{[-N+2, N]}$  : autrement écrit,  $l_{N-1}(x) = x_{[2, N]}$ . On montre de manière analogue que  $x_{[-N+1, -1]} = r_{N-1}(x)$ . Le rang  $N = 1$  nous donne que  $x_{[0, 1]} = 01$  ou  $10$ . En recollant toutes les techniques précédentes, on obtient  $x_{[-N+1, N]} = r_{N-1}(x)x_{[0, 1]}l_{N-1}(x)$  et on obtient bien  $Q_N$ .

**Lemme 3.3.10.**  $Q_N \implies P_N$ .

*Démonstration du lemme 3.3.10.*

On montre cette implication par récurrence sur  $N$ . On sait déjà que  $Q_1 \implies P_1$ . On suppose maintenant que  $Q_N \implies P_N$  : montrons que  $Q_{N+1} \implies P_{N+1}$ . Supposons donc que  $x_{[-N, N+1]} = r_N(x)0l_N(x)$  (l'autre cas se traite de manière symétrique). Par relations de suffixe et de préfixe entre les facteurs spéciaux des différentes longueur,  $Q_{N+1} \implies Q_N$ . Par hypothèse de récurrence,  $Q_N \implies P_N$  : donc  $x_{[-N+1, N]} = r_{N-1}(x)0l_{N-1}(x)$  contient tout élément de  $\mathcal{L}_N(x)$  exactement une fois. Soit  $f \in \mathcal{L}_{N+1}(x)$  dont le suffixe  $g$  de longueur  $N$  n'est pas spécial à gauche. Alors  $g$  est un facteur de  $x_{[-N+1, N]}$  et comme  $g$  n'est pas spécial à gauche, toute occurrence de  $g$  dans  $x$  est uniquement prolongeable à gauche en  $f$ . Donc  $f$  apparaît exactement une fois dans  $x_{[-N, N]}$ . Comme  $g \neq l_N(x) = x_{[2, N+1]}$ ,  $f$  apparaît exactement une fois dans  $x_{[-N, N+1]}$ . De manière analogue, si  $f' \in \mathcal{L}_{N+1}(x)$  a un préfixe de longueur  $N$  non spécial à droite,  $f'$  apparaît exactement une fois dans  $x_{[-N, N+1]}$ . Il nous reste à traiter le cas où  $f \in \mathcal{L}_{N+1}(x)$  admet  $l_N(x)$  comme suffixe de longueur  $N$ . Par hypothèse  $Q_{N+1}$ ,  $x_{[1, N+1]} = 1l_N(x)$  : il faut donc traiter le cas où  $f = 0l_N(x)$ . Symétriquement, en raisonnant sur les préfixes spéciaux à droite, il faut traiter le cas où  $f' = r_N(x)1$ . Supposons que  $f$  n'apparaît pas dans  $x_{[-N, N+1]}$ . Alors sur les  $N+2$  éléments de  $\mathcal{L}_{N+1}(x)$ , c'est le seul qui n'apparaît pas. Donc en raisonnant sur les préfixes spéciaux à droite, le seul cas inconnu doit aussi ne pas apparaître : cela signifie que  $f'$  n'apparaît pas non plus, et par cardinalité,  $f = f'$ . Donc  $0l_N(x) = r_N(x)1$ . Il faut alors trouver le coupable qui apparaîtrait deux fois dans  $x_{[-N, N+1]}$  : c'est l'autre mot qui a un suffixe spécial à gauche de longueur  $N$ ,  $1l_N(x)$ . Mais toujours par symétrie du raisonnement, c'est aussi l'autre mot qui a un préfixe spécial à droite de longueur  $N$ ,  $r_N(x)0$ . Et encore une fois, pour des raisons de cardinalité,

ces deux mots sont en fait un et un seul :  $1l_N(x) = r_N(x)0$ . Mais 
$$\begin{cases} 0l_N(x) = r_N(x)1, \\ 1l_N(x) = r_N(x)0 \end{cases}$$

n'est pas un système d'équations satisfiable :  $r_N(x)$  commencerait à la fois par un 0 et un 1.  
 Donc tous les éléments de  $\mathcal{L}_{N+1}(x)$  apparaissent exactement une fois dans  $x_{[-N, N+1]}$ , d'où  $P_{N+1}$ .

Donc  $\llbracket 0, 1 \rrbracket$  est un piège à facteurs pour  $x$  ssi  $P_N$  est vraie pour tout  $N \geq 1$  ssi  $Q_N$  est vraie pour tout  $N \geq 1$  ssi  $x = \lim_{n \rightarrow \infty} r_n(x) \cdot 01 \lim_{n \rightarrow \infty} l_n(x)$  ou  $\lim_{n \rightarrow \infty} r_n(x) \cdot 10 \lim_{n \rightarrow \infty} l_n(x)$ .

On peut alors étendre le corollaire 3.3.8 en la formulation suivante :

**Corollaire 3.3.11.** *Un mot bi-infini apériodique  $x$  est d'envergure 1 si, et seulement si,  $x$  est un mot sturmien caractéristique à décalage près.*

D'abord, ce résultat va à l'encontre du résultat dans le cas de  $\mathbb{N}$  : la famille caractérisée a beau être très contraignante, elle n'est tout de même pas vide. Ensuite, ce résultat aurait en fait pu être établi à partir de la littérature existante : dans [BLS21], Barbieri, Labbé et Starosta, dans le contexte des paires de mots asymptotiquement indistinguables, fournissent des preuves des propositions 3.3.1 et 3.3.8. Dans la preuve de la proposition 3.3.7, j'ai d'abord traité le cas des mots finis avant de passer à la limite : le cas des mots finis avait déjà été traité par [BR06]. Enfin, la forme générale de la preuve du lemme 3.3.9 ressemble à des techniques combinatoires développées dans [Zam18]. La preuve présentée ici a l'avantage de regrouper de manière compacte les arguments disséminés dans la littérature. Maintenant, on dispose d'exemples explicites de mots bi-infinis apériodiques admettant un piège à facteurs fini. La suite de ce chapitre a maintenant pour but de caractériser complètement les mots bi-infinis avec ces propriétés.

### 3.3.3 Caractérisation des mots sturmiens admettant un piège à facteurs fini

Bien que le corollaire 3.3.11 permette d'éclaircir le cas des mots sturmiens caractéristiques à décalage près, il ne traite toutefois pas le cas de *tous* les mots sturmiens : que dire des mots sturmiens non caractéristiques ? Certes, ils ne peuvent pas avoir de piège à facteurs de diamètre 1, mais peuvent-ils admettre un piège à facteurs fini plus grand ? L'objectif de cette sous-section est de montrer que non : un mot sturmien non caractéristique n'admet que des pièges à facteurs infinis.

Je vais pour cela m'appuyer sur la représentation S-adique des mots sturmiens. Tout comme dans le chapitre précédent, je vais présenter une application combinatoire des représentations S-adiques. En particulier, je vais utiliser une version bi-infinie de la représentation S-adique d'Arnoux, qui n'emploie que  $L_0$  et  $L_1$  : il s'agit de la représentation S-adique de Berthé, Holton et Zamboni.

**Théorème 3.3.12** ([BHZ06]). *Soit  $x$  un mot sturmien bi-infini. Alors il existe  $(a_i)_{i \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$  non stationnaire,  $(c_i)_{i \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$  et  $(x^{(i)})_{i \in \mathbb{N}}$  une suite de mots sturmiens bi-infinis telles que :*

$$\forall i \geq 0, x = S^{c_0} L_{a_0} \circ S^{c_1} L_{a_1} \circ \dots \circ S^{c_i} L_{a_i}(x^{(i)})$$

L'une des propriétés fortes de cette représentation est le lien entre  $x$  et les  $x^{(i)}$  : si l'un des  $x^{(i)}$  est un mot sturmien caractéristique, alors  $x$  est aussi un mot sturmien caractéristique, à décalage près.

**Proposition 3.3.13.** *Soit  $x$  un mot sturmien caractéristique. Alors  $SL_0(x)$  et  $SL_1(x)$  sont aussi des mots sturmiens caractéristiques.*

*Démonstration.*

On traite ici le cas où  $x = \lim_{n \rightarrow \infty} r_n(x) \cdot 01 \lim_{n \rightarrow \infty} l_n(x)$  et où on applique  $SL_0$ , les autres cas sont analogues.

D'abord,

$$\begin{aligned} SL_0(x) &= S[L_0(\lim_{n \rightarrow \infty} r_n(x)) \cdot L_0(01)L_0(\lim_{n \rightarrow \infty} l_n(x))] \\ &= L_0(\lim_{n \rightarrow \infty} r_n(x)) 0 \cdot 01 L_0(\lim_{n \rightarrow \infty} l_n(x)) \\ &= \lim_{n \rightarrow \infty} L_0(r_n(x)) 0 \cdot 01 \lim_{n \rightarrow \infty} L_0(l_n(x)) \text{ par continuité de } L_0. \end{aligned}$$

On montre maintenant que  $\lim_{n \rightarrow \infty} L_0(r_n(x))0 = \lim_{n \rightarrow \infty} r_n(L_0(x))$ . L'égalité a bien du sens, car comme  $L_0$  est une substitution sturmienne,  $L_0(x)$  est bien un mot sturmien bi-infini. D'abord, comme  $r_n(x)$  est spécial à droite dans  $x$ ,  $r_n(x)0$  et  $r_n(x)1$  sont des facteurs de  $x$ . Il existe même une lettre  $a \in \{0, 1\}$  telle que  $r_n(x)0a$  est un facteur de  $x$ . Alors  $L_0(r_n(x)0a) = L_0(r_n(x))0L_0(a)$  et  $L_0(r_n(x)1) = L_0(r_n(x))01$  sont des facteurs de  $L_0(x)$ . Comme  $L_0(a)$  commence toujours par un 0, on en conclut aussi que  $L_0(r_n(x))00$  est un facteur de  $x$ . Donc  $L_0(r_n(x))00$  et  $L_0(r_n(x))01$  sont des facteurs de  $L_0(x)$  : alors  $L_0(r_n(x))0$  est un facteur spécial à droite de  $L_0(x)$ . On en déduit que  $(L_0(r_n(x))0)_{n \in \mathbb{N}}$  est une sous-suite de  $(r_n(L_0(x)))_{n \in \mathbb{N}}$  : les deux suites partagent donc la même limite. Par un raisonnement similaire, on montre que  $\lim_{n \rightarrow \infty} L_0(l_n(x)) = \lim_{n \rightarrow \infty} l_n(L_0(x))$ . Enfin,  $SL_0(x)$  et  $L_0(x)$  partagent les mêmes facteurs spéciaux. On obtient :

$$\begin{aligned} SL_0(x) &= \lim_{n \rightarrow \infty} L_0(r_n(x)) 0 \cdot 01 \lim_{n \rightarrow \infty} L_0(l_n(x)) \\ &= \lim_{n \rightarrow \infty} r_n(L_0(x)) \cdot 01 \lim_{n \rightarrow \infty} l_n(L_0(x)) \\ &= \lim_{n \rightarrow \infty} r_n(SL_0(x)) \cdot 01 \lim_{n \rightarrow \infty} l_n(SL_0(x)) \end{aligned}$$

Donc  $SL_0(x)$  est bien un mot sturmien caractéristique.

On dispose d'une caractérisation des pièges à facteurs dans le cas des mots sturmiens caractéristiques et j'ai montré que ces derniers formaient une classe stable par  $L_0$  et par  $L_1$  : l'étape suivante consiste à étudier le comportement des pièges à facteurs vis-à-vis de  $L_0$  et  $L_1$ .

Pour ces résultats, une notion importante est celle de correspondance entre les occurrences d'un facteur dans un mot et d'un autre facteur dans son image. Il s'agit d'une forme de reconnaissabilité que j'appliquerai précisément pour les pièges à facteurs.

**Définition 3.3.14** (Correspondance des occurrences). *Soit  $x$  un mot bi-infini, et soit  $\phi$  une substitution. Soit  $f$  un facteur de  $x$  et  $g$  un facteur de  $\phi(x)$ . On dit que toute occurrence de  $g$  dans  $\phi(x)$  correspond à une occurrence de  $f$  dans  $x$  si pour tout  $m \in \mathbb{Z}$  tel que  $\phi(x)_{[m, m+|g|-1]} = g$ , il existe  $n \in \mathbb{Z}$  tel que  $x_{[n, n+|f|-1]} = f$  et  $\phi([n, n+|f|-1]) \supseteq [m, m+|g|-1]$ .*

Maintenant que j'ai introduit les notions et le vocabulaire nécessaires, lançons-nous dans la manipulation précise de désubstitution d'un piège à facteurs par  $L_0$ . Les deux propositions suivantes sont à la fois techniques et très précises : c'est la finesse des détails qui nous permettra de conclure plus loin.

**Proposition 3.3.15.** *Soit  $y \in \{0, 1\}^{\mathbb{Z}}$  et  $x = L_0(y)$ . Soit  $\Gamma = [n, n+k]$  un piège à facteurs de  $x$ . Alors :*

1. *Il existe  $m \in \mathbb{Z}$  tel que  $L_0^{-1}(\Gamma) = [m, m+\ell]$  avec*

$$\ell = k - |x_{\llbracket n+1, n+k \rrbracket}|_1$$

2. L'ensemble  $L_0^{-1}(\Gamma) \cup \{m-1\}$  est un piège à facteurs de  $y$ .

3. De plus,

- si  $x_{n-1}x_n \neq 00$ ,  $L_0^{-1}(\Gamma)$  est aussi un piège à facteurs de  $y$ ;
- de surcroît, si  $x_{n+k}x_{n+k+1} = 01$ , alors on peut aussi retirer la position  $m+\ell$  du piège à facteurs obtenu jusqu'ici et quand même former un piège à facteurs de  $y$ .

*Démonstration.*

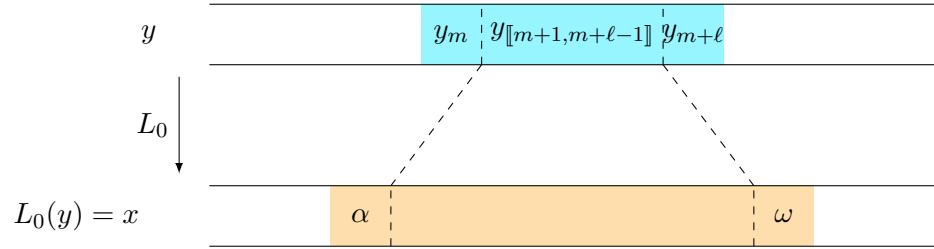
1. Comme  $L_0$  est un morphisme non-éffaçant,  $L_0^{-1}(\Gamma)$  est bien un segment de  $\mathbb{Z}$ , d'où l'existence de  $m$  et  $\ell$ .

Je prouve maintenant la valeur de  $\ell$ . Si  $\ell = 0$ ,  $y$  est un mot constant : l'égalité est alors vérifiée dans le cas  $y = (0)_{n \in \mathbb{Z}}$  et  $y = (1)_{n \in \mathbb{Z}}$ . Autrement, par définition de  $L_0$  et de la préimage d'un piège à facteurs, il existe deux mots non vides  $\alpha$  et  $\omega$  tels que  $x_{\llbracket n, n+k \rrbracket} = \alpha L_0(y_{\llbracket m+1, m+\ell-1 \rrbracket})\omega$ . Plus précisément, si  $x_n = 1$ , alors  $\alpha = 1$  : sinon,  $\alpha = L_0(y_m)$ . Si  $x_{n+k}x_{n+k+1} = 01$ , alors  $\omega = 0$  : sinon,  $\omega = L_0(y_{m+\ell})$ . Donc  $|\alpha| = 2$  ssi ( $\alpha = L_0(y_m)$  et  $y_m = 1$ ) ssi  $x_{n+1} = 1$ . Et  $|\omega| = 2$  ssi ( $\omega = L_0(y_{m+\ell})$  et  $y_{m+\ell} = 1$ ) ssi  $x_{n+k} = 1$ .

On obtient finalement que :

$$\begin{aligned} k+1 &= |x_{\llbracket n, n+k \rrbracket}| \\ &= |\alpha| + \sum_{i=1}^{\ell-1} |L_0(y_{m+i})| + |\omega| \\ &= |\alpha| + \ell - 1 + |x_{\llbracket n+|\alpha|, n+k-|\omega| \rrbracket}|_1 + |\omega|. \end{aligned}$$

Les caractérisations de  $|\alpha|$  et  $|\omega|$  données plus tôt permettent alors de déduire que  $k = \ell + |x_{\llbracket n+1, n+k \rrbracket}|_1$ .



**FIGURE 3.8** – Découpage de  $\Gamma$  et  $L_0^{-1}(\Gamma)$  :  $\alpha$  est bien un suffixe de l'image de  $y_m$  et  $\omega$  un préfixe de l'image de  $y_{m+\ell}$ .

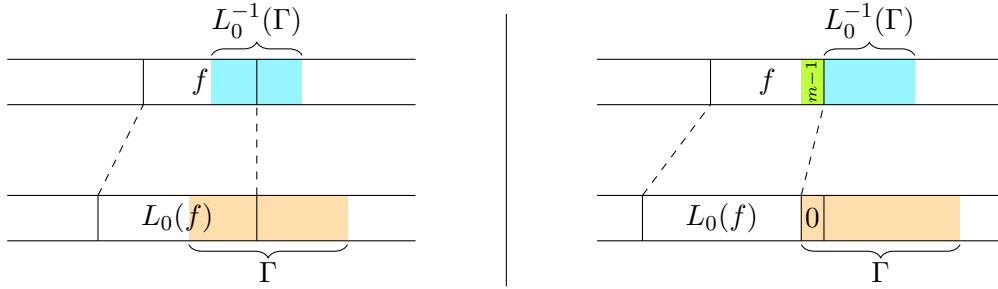
2. Soit  $f$  un facteur non vide de  $y$ . Montrons que  $f$  est capturé par  $L_0^{-1}(\Gamma) \cup \{m-1\}$ . Supposons d'abord que  $f$  termine par 1. Alors par la forme de  $L_0$  et de  $f$ , toute occurrence de  $L_0(f)$  dans  $x$  correspond à une occurrence de  $f$  dans  $y$ . Comme  $\Gamma$  est un piège à facteurs de  $x$ ,  $L_0(f)$  est capturé par  $\Gamma$ , donc l'une des occurrences de  $f$  est capturé par  $L_0^{-1}(\Gamma)$  dans  $y$ .

Supposons maintenant que  $f$  termine par un 0. Alors toujours par la forme de  $L_0$  et de  $f$ , toute occurrence de  $L_0(f)0$  dans  $x$  correspond à une occurrence de  $f$  dans  $y$ . Le facteur  $L_0(f)0$  est capturé par  $\Gamma$  dans  $x$ . Il y a deux maintenant deux possibilités : si  $L_0(f)0$  est capturé par  $\llbracket n, n+k \rrbracket$ , on conclut comme dans le premier cas et  $f$  est capturé par  $L_0^{-1}(\Gamma)$ . Sinon, c'est que  $L_0(f)0$



ne croise  $\Gamma$  qu'à la position  $n$  : alors l'occurrence de  $f$  correspondante dans  $y$  croise la position  $m - 1$ .

Dans les deux cas,  $L_0^{-1}(\Gamma) \cup \{m - 1\}$  capture tout facteur de  $y$ .



**FIGURE 3.9** – Visualisation de pourquoi  $L_0^{-1}(\Gamma) \cup \{m - 1\}$  est un piège à facteurs de  $y$  dans tous les cas : à gauche, le cas général ; à droite, le pire cas, où  $f$  termine par un 0 et  $L_0(f)0$  n'est capturé par  $\Gamma$  que par sa dernière position. Dans les deux cas,  $L_0^{-1}(\Gamma) \cup \{m - 1\}$  capture bien  $f$ .

3. On reprend le raisonnement suivi dans le point précédent.

La position  $m - 1$  n'est nécessaire que pour la capture de facteurs terminant par un 0. Alors si  $x_n = 1$ ,  $L_0(f)0$  ne peut être capturé par  $\Gamma$  uniquement par la position  $n$ , et donc il n'y a pas besoin de rajouter la position  $m - 1$  à  $L_0^{-1}(\Gamma)$  pour former un piège à facteurs de  $y$ . Puis si  $x_{n-1}x_n = 10$ , c'est aussi impossible que  $L_0(f)0$  ne soit capturé par  $\Gamma$  que par la position  $n$ , puisqu'on aurait alors que  $f$  termine par un 1. Donc si  $x_{n-1}x_n \neq 00$ ,  $L_0^{-1}(\Gamma)$  est déjà un piège à facteurs de  $x$ .

Enfin, supposons que  $x_{n+k}x_{n+k+1} = 01$ . Alors  $y_{m+\ell} = 1$ . Montrons alors que la position  $m + \ell$  est superflue, c.-à-d. que tout facteur  $y_{\llbracket m+\ell, m+\ell+j \rrbracket}$  a une occurrence dans  $y$  croisant  $\llbracket m, m + \ell - 1 \rrbracket$  ou  $\llbracket m - 1, m + \ell - 1 \rrbracket$  (selon le piège à facteurs obtenu jusqu'ici). Soit  $f$  un mot tel que  $y_{\llbracket m+\ell, m+\ell+j \rrbracket} = 1f$ . De par la forme de  $L_0$ , on peut affirmer deux choses : si  $1f$  termine par 1, alors toute occurrence de  $1L_0(f)$  dans  $x$  correspond à une occurrence de  $1f$  dans  $y$  ; si  $1f$  termine par 0, alors toute occurrence de  $1L_0(f)0$  correspond à une occurrence de  $1f$  dans  $y$ . Par hypothèse,  $1L_0(f)$  (resp.  $1L_0(f)0$ ) est couvert par  $\Gamma$  dans  $x$  et, comme  $x_{n+k} = 0$ , est en fait même couvert par  $\llbracket n, n + k - 1 \rrbracket$ . Donc  $1f$  est couvert par  $\llbracket m - 1, m + \ell - 1 \rrbracket$ , et la position  $m - 1$  n'est nécessaire que si  $x_{n-1}x_n = 00$ .

Le symétrique de la proposition 3.3.15 pour la substitution  $L_1$  est la proposition suivante :

**Proposition 3.3.16.** Soit  $y \in \{0, 1\}^{\mathbb{Z}}$  et  $x = L_1(y)$ . Soit  $\Gamma = \llbracket n, n + k \rrbracket$  un piège à facteurs de  $x$ . Alors :

1. Il existe  $m \in \mathbb{Z}$  tel que  $L_1^{-1}(\Gamma) = \llbracket m, m + \ell \rrbracket$  avec

$$\ell = k - |x_{\llbracket n+1, n+k \rrbracket}|_0$$

2. L'ensemble  $L_1^{-1}(\Gamma) \cup \{m - 1\}$  est un piège à facteurs de  $y$ .

3. De plus,

- si  $x_{n-1}x_n \neq 11$ ,  $L_1^{-1}(\Gamma)$  est aussi un piège à facteurs de  $y$  ;
- de surcroît, si  $x_{n+k}x_{n+k+1} = 10$ , alors on peut aussi retirer la position  $m + \ell$  du piège à facteurs obtenu jusqu'ici et quand même former un piège à facteurs de  $y$ .

La proposition 3.3.16 se démontre de la même manière que la proposition 3.3.15. Ces deux propositions sont

techniques, mais permettent enfin de complètement caractériser les mots sturmiens admettant des pièges à facteurs finis.

**Proposition 3.3.17.** *Soit  $x$  un mot sturmien bi-infini. Alors  $x$  admet un piège à facteurs fini si, et seulement si,  $x$  est un mot sturmien caractéristique à décalage près.*

*Démonstration.*

On a démontré (voir la proposition 3.3.8) qu'un mot sturmien  $x$  est d'envergure 1 si, et seulement si,  $x$  est un mot sturmien caractéristique à décalage près. Donc si  $x$  est un mot sturmien caractéristique à décalage près,  $x$  admet un piège à facteurs fini. Il reste à montrer la réciproque.

Soit  $x$  un mot sturmien admettant un piège à facteurs fini. D'après le théorème 3.3.12, il existe  $(a_i)_{i \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$  non stationnaire,  $(c_i)_{i \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$ ,  $(x^{(i)})_{i \in \mathbb{N}}$  une suite de mots sturmiens bi-infinis telle que  $x = S^{c_0} L_{a_0} \circ S^{c_1} L_{a_1} \circ \dots \circ S^{c_i} L_{a_i}(x^{(i)})$  pour tout  $i \geq 0$ . Cette représentation S-adique vérifie toujours l'alternance infinie de type : on définit alors  $\theta(x) \geq 0$  le plus petit entier tel que  $(a_i)_{0 \leq i \leq \theta(x)}$  contienne  $\text{env}(x) - 1$  alternances de type. Dans un premier temps, je ferai l'hypothèse suivante, nommée (H) :

$$c_i = 0 \text{ pour tout } 0 \leq i \leq \theta(x)$$

Afin de terminer la preuve de la proposition 3.3.17, je démontre le lemme suivant :

**Lemme 3.3.18.** *Soit  $1 \leq \eta(x) \leq \theta(x)$  tel que  $a_0 = a_1 = \dots = a_{\eta(x)-1} \neq a_{\eta(x)}$ . Si  $\text{env}(x) \neq 1$ , alors  $\text{env}(x^{(\eta(x))}) < \text{env}(x)$ .*

*Démonstration du lemme 3.3.18.*

Soit  $\Gamma = \llbracket n, n+k \rrbracket$  un piège à facteurs de  $x$  de diamètre minimal : je construis maintenant un piège à facteurs pour  $x^{(0)}$ , avec l'espoir qu'il soit plus petit que  $\Gamma$ . Comme  $\text{env}(x) \neq 1$ ,  $\theta(x) \geq 1$ , et de plus, par (H),  $c_0 = 0$ . Nous traiterons ici le cas  $x = L_0(x^{(0)})$ , le cas  $x = L_1(x^{(0)})$  se traite de manière analogue.

La proposition 3.3.15 permet de conclure directement, c.-à-d. de construire un piège à facteurs strictement plus petit que  $\Gamma$ , dans les cas suivants :

1.  $|x_{\llbracket n+1, n+k \rrbracket}|_1 \geq 2$ , ou
2.  $|x_{\llbracket n+1, n+k \rrbracket}|_1 = 1$  et  $(x_{n-1}x_n \neq 00$  ou  $x_{n+k}x_{n+k+1} = 01)$ , ou
3.  $|x_{\llbracket n+1, n+k \rrbracket}|_1 = 0$ ,  $x_{n-1} \neq 00$  et  $x_{n+k}x_{n+k+1} = 01$ .

En notant  $\llbracket m, m+l \rrbracket = L_0^{-1}(\Gamma)$ , il nous reste deux cas problématiques :

1. lorsque  $|x_{\llbracket n+1, n+k \rrbracket}|_1 = 1$  et  $x_{n-1}x_n = 00$  et  $x_{n+k}x_{n+k+1} \neq 01$ . D'après la proposition 3.3.15,  $x^{(0)}$  admet comme piège à facteurs  $\llbracket m-1, m+l \rrbracket$  avec  $\ell = k-1$ . De plus, on a  $x_{\llbracket n, n+k \rrbracket} = 0^i 10^j$ , et si  $j \geq 1$ ,  $x_{n+k+1} = 0$ . Cela implique alors que  $x_{\llbracket m-1, m+k-1 \rrbracket}^{(0)} = 0^i 10^j = x_{\llbracket n, n+k \rrbracket}$ .
2. lorsque  $|x_{\llbracket n+1, n+k \rrbracket}|_1 = 0$  et  $(x_{n-1}x_n = 00$  ou  $x_{n+k}x_{n+k+1} \neq 01)$ . Un piège à facteurs devant capturer toutes les lettres, on a  $x_{\llbracket n, n+k \rrbracket} = 10^k$ . Donc en fait,  $x_{n-1}x_n \neq 00$  et c'est la condition  $x_{n+k}x_{n+k+1} \neq 01$  qui est remplie : comme  $x_{n+k} = 0$ , on a donc  $x_{n+k+1} = 0$ . D'après la proposition 3.3.15,  $x^{(0)}$  admet comme piège à facteurs  $\llbracket m, m+l \rrbracket$  avec  $\ell = k$ . Par désubstitution, on a donc  $x_{\llbracket m, m+k \rrbracket}^{(0)} = 10^k = x_{\llbracket n, n+k \rrbracket}$ .

Dans ces deux cas problématiques, en notant  $\llbracket m', m' + k \rrbracket$  le piège à facteurs de  $x^{(0)}$  obtenu, on a  $x_{\llbracket m', m' + k \rrbracket}^{(0)} = x_{\llbracket n, n + k \rrbracket}$ . Par définition du type d'un mot sturmien (voir définition 1.2.12), on en déduit alors que si un cas problématique a lieu, c'est que  $x$  et  $x^{(0)}$  ont le même type. Donc par contraposée, si  $x$  et  $x^{(0)}$  ne sont pas de même type, c'est que les cas problématiques n'ont pas lieu. La représentation S-adique implique alors que  $x^{(\eta(x)-1)}$  et  $x^{(\eta(x))}$  n'ont pas même type : en construisant un piège à facteurs de  $x^{(\eta(x))}$  à partir d'un piège à facteurs de  $x^{(\eta(x)-1)}$ , les cas problématiques n'ont donc pas lieu. Donc  $\text{env}(x^{(\eta(x))}) < \text{env}(x^{(\eta(x)-1)}) \leq \text{env}(x)$  : on conclut bien que  $\text{env}(x^{(\eta(x))}) < \text{env}(x)$ .

Je reprends la preuve de la proposition 3.3.17. À partir du lemme 3.3.18, on peut maintenant amorcer un principe de descente infinie de Fermat, en remarquant que l'envergure décroît strictement à chaque alternance de type : et le raisonnement pourra toujours bien s'appliquer parce que  $c_i = 0$  pour  $0 \leq i \leq \theta(x)$ . En conclusion, on a  $\text{env}(x^{(\theta(x))}) = 1$ .

Si  $x$  ne respecte pas l'hypothèse (H), considérons le mot  $\tilde{x} = L_{a_0} \circ \dots \circ L_{a_{\theta(x)}}(x^{\theta(x)})$ . Alors d'après la représentation S-adique de Berthé, Holton et Zamboni (voir le théorème 3.3.12), en définissant  $\tilde{x}^{(i)}$  de manière analogue à  $x^{(i)}$ , on a  $\tilde{x}^{(\theta(x))} = x^{(\theta(x))}$  à décalage près; et  $\tilde{x}$  vérifie, lui, l'hypothèse (H). D'après le raisonnement précédent,  $\text{env}(\tilde{x}^{(\theta(x))}) = 1$  : on a bien  $\text{env}(x^{(\theta(x))}) = 1$ .

Autrement écrit, dans tous les cas, on dispose d'un indice  $i \geq 0$  tel que  $\text{env}(x^{(i)}) = 1$ . Donc, d'après le corollaire 3.3.11,  $x^{(i)}$  est un mot sturmien caractéristique. Et d'après la proposition 3.3.13, l'image d'un mot sturmien caractéristique à décalage près par les morphismes  $L_0$  et  $L_1$  est un mot sturmien caractéristique à décalage près. Donc  $x$  est un mot sturmien caractéristique à décalage près.

Autrement écrit, l'exemple exhibé par le corollaire 3.3.11 est la seule possibilité pour un mot sturmien d'admettre un piège à facteurs fini. À partir de cette caractérisation, on peut en déduire le corollaire suivant :

**Corollaire 3.3.19.** *Soit  $x$  un mot sturmien bi-infini. Alors :*

$$\text{env}(x) = \begin{cases} 1 & \text{si } x \text{ est un mot sturmien caractéristique à décalage près,} \\ \infty & \text{sinon.} \end{cases}$$

*De plus, si  $\mathbb{X}$  est un shift sturmien, alors il y a une quantité infinie dénombrable d'éléments de  $\mathbb{X}$  d'envergure 1, et une quantité infinie indénombrable d'éléments de  $\mathbb{X}$  d'envergure infinie.*

*Démonstration.*

Dans un shift sturmien, il y a exactement deux mots sturmiens caractéristiques : à décalage près, il y a donc une quantité infinie dénombrable de mots sturmiens caractéristiques à décalage près. On conclut en rappelant qu'un shift sturmien contient une quantité infinie indénombrable d'éléments.

Pour récapituler jusqu'ici, j'ai traité le cas des mots sturmiens, et le cas des mots apériodiques d'envergure 1. Il me reste alors à traiter le cas des pièges à facteurs plus grands, et donc étudier les mots quasi-sturmiens.

### 3.3.4 Généralisation au cas d'envergure finie quelconque

Maintenant qu'on a traité le cas des mots apériodiques bi-infinis d'envergure 1, je vais généraliser le résultat aux mots apériodiques bi-infinis d'envergure  $k$  pour  $k \geq 1$ .

D'abord, à partir des propositions précédentes, on peut établir un premier résultat d'existence :

**Proposition 3.3.20.** *Soit  $y$  un mot bi-infini sturmien caractéristique, et soit  $\phi$  une substitution aperiodique. Alors  $\phi(y)$  est un mot aperiodique et admet un piège à facteurs fini.*

*Démonstration.*

C'est une conséquence de la proposition 3.2.11, du corollaire 3.3.8 et du fait que les substitutions aperiodiques préservent l'aperiodicité.

Techniquement, il est possible d'élargir légèrement la proposition 3.3.20, en considérant le cas où  $y$  est un décalage d'un mot sturmien caractéristique. Encore une fois, ce résultat est en contraste avec le cas de  $\mathbb{N}$  établi dans [RRS22]. On exhibe ainsi en effet une infinité de mots aperiodiques admettant des pièges à facteurs finis. C'est le caractère *bilatéral* de  $\mathbb{Z}$  qui fait la différence avec le cas de  $\mathbb{N}$ .

Mon objectif désormais est de caractériser plus précisément quels mots admettent un piège à facteurs fini, et quelles sont les envergures possibles. D'après la proposition 3.3.1, les seuls candidats récurrents sont les mots quasi-sturmiens (voir la définition 3.3.5). Nous allons alors déterminer précisément quels mots quasi-sturmiens admettent des pièges à facteurs finis.

### Caractérisation des mots quasi-sturmiens

Pour cela, il faut d'abord mieux comprendre *ce que sont* les mots quasi-sturmiens. La première étude des mots quasi-sturmiens remonte à Hedlund, Coven et Paul [CH73; Cov75; Pau75], qui étudiaient de manière générale les *words of minimal block growth*, c.-à-d. les mots  $x$  vérifiant  $p_x(n+1) - p_x(n) = 1$ . La question a connu une résurgence vingt ans plus tard, avec notamment les travaux de Didier, Alessandri et Heinis [Did99; Ale96; Hei01]. On peut aussi citer un article de Ferenczi et Mauduit qui les utilisent comme un outil [FM97]. C'est pendant cette période qu'a émergé l'appellation de mots *quasi-sturmiens* par Cassaigne dans le cas mono-infini [Cas97b], repris depuis dans d'autres articles [KT07a; RRS22]. Je procède ici à une extension de la caractérisation de [Cas97b] au cas bi-infini, en raffinant certains détails : la caractérisation de [Cas97b] utilise les substitutions aperiodiques, et j'établis une caractérisation utilisant les *morphismes de retour*.

**Définition 3.3.21** (Morphisme de retour). *Soit  $w$  un mot fini non vide. Alors une substitution  $\phi$  est un morphisme de retour pour  $w$  si :*

1.  $\phi$  est injectif sur  $\mathcal{A}$ ;
2. pour tout  $a \in \mathcal{A}$ ,  $w\phi(a)$  contient exactement deux occurrences de  $w$ , une comme préfixe, l'autre comme suffixe.

*Une substitution  $\phi$  est un morphisme de retour si c'est un morphisme de retour pour un mot non vide  $w$ .*

Il s'agit d'une notion rattachée à celle de *mot de retour*, introduite en combinatoire des mots par Durand dans [Dur98] afin de, notamment, établir des résultats de décidabilité sur les mots purement substitutifs. Les morphismes de retour, eux, ont principalement été étudiés par Leroy, en particulier dans son travail avec Gheeraert [GLL22; GL22; Ghe23a]. Je vais utiliser dans le contexte de ce chapitre plusieurs de leurs propriétés. Ces propriétés ont été démontrées dans les références données plus tôt : je fournis ici une preuve des trois premières propriétés.

**Proposition 3.3.22.** *Soit  $w$  un mot fini non vide, et  $\phi$  un morphisme de retour pour  $w$ .*

1. Pour tout  $u \in \mathcal{A}^*$ ,  $w$  est un suffixe de  $w\phi(u)$ .
2. La substitution  $\phi$  est injective sur  $\mathcal{A}^*$ , sur  $\mathcal{A}^{\mathbb{N}}$  et sur  $\mathcal{A}^{\mathbb{Z}}$ .
3. Soit  $x \in \mathcal{A}^{\mathbb{Z}}$ , et soit  $u$  un facteur de  $x$ . Alors toute occurrence de  $u$  dans  $x$  correspond à une occurrence de  $w\phi(u)$  dans  $\phi(x)$  et réciproquement.
4. La substitution  $\phi$  est aperiodique, c.-à-d. que si  $x$  est un mot aperiodique, alors  $\phi(x)$  est aperiodique.

*Démonstration.*

1. Cette propriété se prouve par récurrence immédiate sur la longueur de  $u$ .
2. Soient  $x, y \in \mathcal{A}^*$  tels que  $\phi(x) = \phi(y)$ . On montre par récurrence immédiate que la longueur de  $x$  correspond au nombre d'occurrences moins 1 de  $w$  dans  $w\phi(x)$ . Donc  $|x| = |y|$ . L'injectivité de  $\phi$  sur  $\mathcal{A}$  implique alors l'injectivité sur  $\mathcal{A}^*$ . Soit  $x, y \in \mathcal{A}^{\mathbb{N}}$  tels que  $\phi(x) = \phi(y)$ . Alors  $w\phi(x) = w\phi(y)$ , et on peut découper  $w\phi(x) = w_{x,0}wx'$ , et de même  $w\phi(y) = w_{y,0}wy'$  en posant  $w_{x,0}$  comme le préfixe (non vide) avant la deuxième occurrence du mot  $w$  dans  $w\phi(x)$  (de même pour  $y$ ). Donc comme  $w\phi(x) = w\phi(y)$ ,  $|w_{x,0}| = |w_{y,0}|$ , et donc  $w_{x,0} = w_{y,0}$ . Par définition de morphisme de retour pour  $w$ , on a alors  $x_0 = y_0$ . En procédant ainsi sur  $wx' = wy'$ , on obtient par induction que  $x = y$ . Enfin, si  $x, y \in \mathcal{A}^{\mathbb{Z}}$  vérifient  $\phi(x) = \phi(y)$ , il suffit de découper  $x = x_- \cdot x_+$  et  $y = y_- \cdot y_+$  et d'utiliser le cas de  $\mathbb{N}$  pour conclure.
3. Supposons que  $x_{\llbracket n, n+|u|-1 \rrbracket} = u$ . On traite le cas où  $n \geq 0$ , les autres cas se traitent de façon analogue. En notant  $m = \sum_{i=0}^{n-1} |\phi(x_i)|$  et  $k = \sum_{i=n}^{n+|u|-1} |\phi(x_i)|$ , on a  $\phi(x)_{\llbracket m, m+k-1 \rrbracket} = \phi(u)$ . Par ailleurs,  $\phi(x)_{\llbracket m-|w|, m-1 \rrbracket}$  vaut  $w$  et est un suffixe de  $\phi(x_{\llbracket n-|w|, n-1 \rrbracket})$  : par le point 1,  $w\phi(x_{\llbracket n-|w|, n-1 \rrbracket})$  admet  $w$  comme un suffixe, mais comme  $\phi$  est non effaçant, ce suffixe est strictement inclus dans  $\phi(x_{\llbracket n-|w|, n-1 \rrbracket})$ . Donc  $\phi(x)_{\llbracket m-|w|, m+k-1 \rrbracket} = w\phi(u)$ . Dans l'autre sens, supposons que  $\phi(x)_{\llbracket m-|w|, m+k-1 \rrbracket} = w\phi(u)$  avec  $m, k \geq 0$  (le cas  $m < 0$  se traite de manière analogue). Si  $\phi(x)_{\llbracket m-|w|, m+k-1 \rrbracket}$  terminait à l'intérieur d'une image de  $\phi$ , alors on aurait une lettre  $a \in \mathcal{A}$  telle que  $w\phi(a)$  contient trois occurrences de  $w$ , ce qui contredit la définition 3.3.21. De même,  $\phi(x)_{\llbracket m, m+k-1 \rrbracket}$  ne peut commencer à l'intérieur d'une image de  $\phi$ . Donc il existe un mot  $v$  et  $n \geq 0$  tel que  $m = \sum_{i=0}^{n-1} |\phi(x_i)|$  et  $k = \sum_{i=n}^{n+|v|-1} |\phi(x_i)|$  et  $\phi(x)_{\llbracket m, m+k-1 \rrbracket} = \phi(x_{\llbracket n, n+|v|-1 \rrbracket}) = \phi(v)$ . Par injectivité de  $\phi$  sur les mots finis, on a  $v = u$  : donc on a bien que  $x_{\llbracket n, n+|u|-1 \rrbracket} = u$ .

Pour fournir une caractérisation des mots quasi-sturmiens, j'ai besoin de préciser le concept des graphes de Rauzy.

**Définition 3.3.23** (Graphe de Rauzy). *Soit  $x$  un mot bi-infini, et soit  $n \geq 1$ . Alors le graphe de Rauzy de  $x$  de rang  $n$  est l' $\omega$ -automate faible (voir la définition 2.2.4)  $\mathcal{G}_n(x) = (\mathcal{A}, \mathcal{L}_n(x), \mathcal{L}_n(x), T_n(x))$  dont les sommets sont les facteurs de  $x$  de longueur  $n$  et où  $f \xrightarrow{a} g \in T_n(x)$  si  $fa = bg \in \mathcal{L}_{n+1}(x)$  pour une lettre  $b \in \mathcal{A}$ .*

Le graphe de Rauzy  $\mathcal{G}_x(x)$  est traditionnellement un graphe, et l'interprétation que j'en fais ici comme un  $\omega$ -automate n'est pas standard. Les graphes de Rauzy (introduits par Rauzy [Rau83]) ont été amplement examinés en combinatoire des mots (par exemple dans [AR91; Rot94; Fri01; Abe03; RR22]) afin par exemple d'étudier plusieurs classes de mots caractérisées par leurs complexités.

L'une des propriétés importantes des graphes de Rauzy est que  $x \in \mathcal{L}(\mathcal{G}_n(x))$  pour tout  $n \geq 1$ . En combinant les propriétés des morphismes de retour et celles des graphes de Rauzy, on peut enfin fournir une caractérisation des mots quasi-sturmiens ressemblant à celle du cas mono-infini (voir [Cas97b]) :

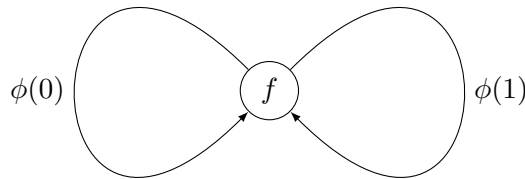
**Proposition 3.3.24.** *Soit  $x$  un mot quasi-sturmien. Alors il existe un facteur bispécial  $f$  de  $x$ , un morphisme de retour  $\phi$  pour  $f$ , un mot sturmien  $y$  et un entier  $m$  tels que  $x = S^m\phi(y)$ . De plus,  $p_x(n) = n + |\phi(0)| + |\phi(1)| - |f| - 1$  pour tout  $n \geq |f|$ .*

*Démonstration.*

Soient  $k \geq 1$  et  $n_0 \geq 0$  tels que  $p_x(n) = n + k$  pour tout  $n \geq n_0$ . Alors pour tout  $n \geq n_0$ ,  $x$  admet exactement un facteur spécial à gauche (et à droite) de longueur  $n$ ; comme  $p_x(n+1) - p_x(n) = 1$  pour  $n \geq n_0$ , ces facteurs spéciaux admettent deux extensions. Notons alors  $l_n(x)$  le facteur spécial à gauche de  $x$  de longueur  $n$  (et  $r_n(x)$  son facteur spécial à droite de longueur  $n$ ). On remarque que  $l_n(x)$  est un préfixe de  $l_{n+1}(x)$ , et symétriquement que  $r_n(x)$  est un suffixe de  $r_{n+1}(x)$ . De plus,  $l_n(x)$  est un facteur récurrent de  $x$  : sinon, il existerait un rang  $r$  tel que  $x_{\llbracket -\infty, r \rrbracket}$  ne contient pas le mot  $l_n(x)$ . Donc pour tout  $i \leq 0$ ,  $x_{\llbracket r+i-n+1, r+i \rrbracket} \neq l_n(x)$ . Donc chacun de ces facteurs est extensible de manière unique à gauche : par le lemme des tiroirs, on retomberait deux fois sur le même facteur, et l'unicité de l'extension à gauche impliquerait alors l'ultime périodicité de  $x$  à gauche, ce qui contredit l'apériodicité des mots quasi-sturmien. De même,  $r_n(x)$  est un facteur récurrent de  $x$ .

Montrons d'abord que  $x$  admet au moins un facteur bispécial : supposons qu'aucun facteur  $l_n(x)$  ne soit spécial à droite. Alors, pour tout  $n \geq n_0$ , l'extension à droite de  $l_n(x)$  à  $l_{n+1}(x)$  serait unique. Par récurrence de  $l_{n_0}(x)$  dans  $x$ , il y aurait alors un indice  $n \geq n_0$  tel que  $l_n(x)$  admet comme suffixe  $l_{n_0}(x)$ . Et par un raisonnement similaire au paragraphe précédent,  $x$  serait alors ultimement périodique à droite, en contradiction avec les hypothèses. Donc il existe  $n_1 \geq n_0$  tel que  $l_{n_1}(x)$  est spécial à droite : autrement écrit,  $x$  admet au moins un facteur bispécial (le même raisonnement justifie en fait qu'il en admet une infinité).

Soit donc  $f$  un facteur bispécial de  $x$ , qu'on suppose de longueur  $n_1 \geq n_0$ . On considère maintenant le graphe de Rauzy  $\mathcal{G}_{n_1}(x)$  :  $f$  est son seul sommet de degré sortant 2 (les autres ayant degré sortant 1), et  $f$  est aussi le seul sommet de degré entrant 2 (les autres ayant degré entrant 1). Le graphe de Rauzy  $\mathcal{G}_{n_1}(x)$  est donc un graphe en papillon, comme sur la figure 3.10.



**FIGURE 3.10** – La forme du graphe de Rauzy  $\mathcal{G}_{n_1}(x)$  : au centre, le sommet correspondant au facteur  $f$  ; à gauche et droite les deux extensions possibles distinctes.

En notant  $\phi(0)$  et  $\phi(1)$  les étiquettes des deux cycles de  $\mathcal{G}_{n_1}(x)$  partant de  $f$ , on en déduit alors qu'il existe  $y \in \{0, 1\}^{\mathbb{Z}}$  et  $m$  un entier tel que  $x = S^m \phi(y)$ . Je vais maintenant montrer que  $\phi$  est bien un morphisme de retour pour  $f$  et que  $y$  est un mot sturmien.

D'abord, montrons que  $\phi$  est bien un morphisme de retour pour  $f$ . Pour  $a \in \{0, 1\}$ , on remarque que  $f \xrightarrow{\phi(a)}^* f$  est une marche sur  $\mathcal{G}_n(x)$  qui ne passe par  $f$  qu'au début et à la fin. Par définition d'un graphe de Rauzy, on en déduit que  $f\phi(a)$  contient bien deux occurrences de  $f$ , une comme préfixe et une comme suffixe. De plus, les deux boucles commencent par des lettres différentes, donc  $\phi(0) \neq \phi(1)$ . Donc  $\phi$  est injectif et donc est bien un morphisme de retour pour  $f$ . Enfin, nous n'avons pas encore précisé comment faire le choix pour savoir quelle boucle sera l'image de quelle lettre : comme  $f$  a deux transitions entrantes étiquetées différemment, les lettres précédant  $f$  comme suffixe dans  $f\phi(0)$  et  $f\phi(1)$  sont différentes. On choisit de telle sorte que  $af$  est un suffixe de  $f\phi(a)$ , quitte à renommer les deux extensions à gauche de  $f$  selon  $\{0, 1\}$ .

Enfin, montrons que  $y$  est un mot sturmien. Comme  $x$  est apériodique et  $\phi$  est une substitution apé-

riodique,  $y$  doit aussi être un mot apériodique. Donc d'après le théorème de Morse-Hedlund pour les mots bi-infinis [MH40],  $p_y(n) \geq n + 1$  pour tout  $n \geq 1$ . Je montre alors que  $y$  admet au plus un facteur spécial à droite pour toute longueur : cela impliquera que  $p_y(n + 1) - p_y(n) \leq 1$  pour tout  $n \geq 1$ , et donc que  $p_y(n) = n + 1$  pour tout  $n \geq 1$ . On procède par l'absurde : soit  $t_0$  et  $t_1$  deux facteurs spéciaux à droite de  $y$  de longueur  $n$ , avec  $n$  le plus petit possible. On suppose, sans perte de généralité, qu'il existe un mot  $t'$  (potentiellement vide) tel que pour  $a \in \{0, 1\}$ ,  $t_a = at'$ . Alors par définition de  $\phi$ , pour tout facteur  $g$  de  $y$ ,  $f\phi(g)$  est un facteur de  $x$ . Donc pour tout  $a, a' \in \{0, 1\}$ ,  $f\phi(a)\phi(t')\phi(a')$  est un facteur de  $x$  (le premier  $a$  représentant quel facteur spécial à droite est choisi, et  $a'$  les deux extensions à droite possibles). Comme  $\phi(0)$  et  $\phi(1)$  commencent par des lettres distinctes, et comme  $af$  est un suffixe de  $f\phi(a)$ , on en conclut que  $af\phi(t')$  est un facteur spécial à droite de  $x$  pour tout  $a \in \{0, 1\}$ . On aurait donc deux facteurs spéciaux à droite de  $x$  de même longueur plus grande que  $n_1 \geq n_0$ , ce qui est absurde.

Enfin, pour calculer la complexité : il suffit de compter le nombre de sommets dans  $\mathcal{G}_{n_1}(x)$ . D'une part, il y en a autant que de facteurs de longueur  $n_1$  dans  $x$ , c.-à-d.  $n_1 + k = |f| + k$ . D'autre part, il y en a la somme des longueurs des deux boucles moins un (les deux boucles partageant le même sommet de départ). Donc il y en a  $|\phi(0)| + |\phi(1)| - 1$ . D'où  $k = |\phi(0)| - |\phi(1)| - |f| - 1$ .

On a donc montré que si un mot est quasi-sturmien, alors c'est l'image d'un mot sturmien par une substitution avec des propriétés précises. Pour conclure la caractérisation, il suffit de montrer que l'image d'un mot sturmien par un morphisme « raisonnable » est un mot quasi-sturmien : il s'agit d'un résultat de Heinis<sup>1</sup>.

**Théorème 3.3.25** ([Hei01, Theorem 3.1]). *Soit  $x$  un mot sturmien bi-infini, et soit  $\sigma$  une substitution telle que  $\sigma(0)$  et  $\sigma(1)$  ne sont pas puissances d'un même mot. Alors  $\sigma(x)$  est un mot quasi-sturmien.*

Le théorème 3.3.25 a aussi été démontré de façon plus générale par Gheeraert [Ghe23b] dans le cas des mots dendriques. Quoi qu'il en soit, on en déduit une caractérisation complète des mots quasi-sturmiens :

**Proposition 3.3.26.** *Un mot bi-infini  $x$  est quasi-sturmien si, et seulement si, il existe  $f$  un facteur de  $x$ ,  $\phi$  un morphisme de retour pour  $f$ ,  $m$  un entier et  $y$  un mot sturmien tels que  $x = S^m\phi(y)$ .*

### Caractérisation des mots bi-infinis admettant un piège à facteurs fini

Avec cette caractérisation des mots quasi-sturmiens, je reviens enfin au cœur du sujet : les pièges à facteurs finis. La proposition 3.3.20 nous a permis d'établir que l'image d'un mot sturmien caractéristique par un morphisme « raisonnable » admettait un piège à facteurs fini. Dans cette sous-section, l'objectif est d'étudier la réciproque : quels sont les mots qui admettent un piège à facteurs fini ? Pour cela, je commence par montrer une forme simplifiée de la proposition 3.3.15 dans le cas des morphismes de retour.

**Lemme 3.3.27.** *Soit  $x$  un mot bi-infini et  $\phi$  un morphisme de retour pour un mot  $f$ . Si  $\Gamma = \llbracket n, n + k \rrbracket$  est un piège à facteurs de  $\phi(x)$ , alors il existe  $m \in \mathbb{Z}$  et  $\ell \geq 0$  tels que  $\phi^{-1}(\Gamma) = \llbracket m, m + \ell \rrbracket$  et tels que  $\llbracket m, m + \ell + |f| \rrbracket$  est un piège à facteurs de  $x$ .*

1. Si Cassaigne écrivait simplement en 1997 « it can be shown that [l'image d'un mot sturmien par un morphisme apériodique est quasi-sturmienne] » comme un exercice au lecteur [Cas97b, Proposition 8], la preuve de ce théorème est en fait un vrai tour de force. La preuve de Heinis fait 4 pages, est particulièrement technique et emploie un vocabulaire qui n'est plus au jour de la recherche actuelle. J'espère que le lecteur saura me pardonner l'absence de la preuve de ce résultat dans ce manuscrit.

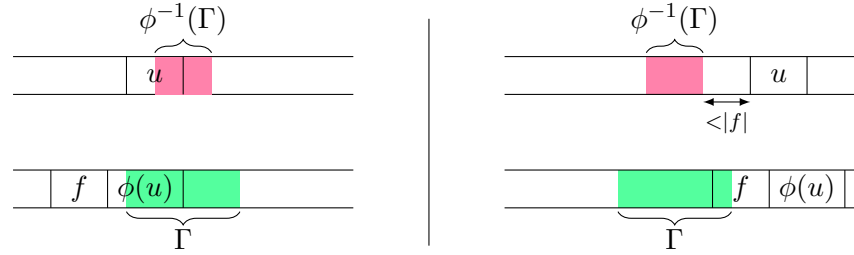
*Démonstration.*

L'existence de  $m$  et de  $\ell$  est due au caractère non-éffaçant de  $\phi$ . Montrons maintenant que  $\llbracket m, m + \ell + |f| \rrbracket$  est un piège à facteurs de  $x$ .

Soit  $u$  un facteur non vide de  $x$ . Comme  $\phi$  est un morphisme de retour pour  $f$ , toute occurrence de  $f\phi(u)$  dans  $\phi(x)$  correspond à une occurrence de  $u$  dans  $x$ . Comme  $\Gamma$  est un piège à facteurs de  $\phi(x)$ , il y a une occurrence de  $f\phi(u)$  capturée par  $\Gamma$ . On considère alors deux cas.

D'abord, si la capture se fait dans la partie  $\phi(u)$  de  $f\phi(u)$ , alors il y a directement une occurrence de  $u$  capturée par  $\phi^{-1}(\Gamma) \subseteq \llbracket m, m + \ell + |f| \rrbracket$ .

Ensuite, si la capture se fait dans la partie  $f$  de  $f\phi(u)$ , comme  $\phi$  est non-éffaçant, l'occurrence de  $u$  est au plus loin  $|f| - 1$  après  $\phi^{-1}(\Gamma)$ . Donc  $u$  est capturé par  $\llbracket m, m + \ell + |f| \rrbracket$ .



**FIGURE 3.11** – Visualisation de pourquoi  $\llbracket m, m + \ell + |f| \rrbracket$  est un piège à facteurs : à gauche, le cas où  $\phi(u)$  est la partie capturée par  $\Gamma$  ; à droite, celle où c'est  $f$  qui est capturé par  $\Gamma$ .

On peut alors maintenant caractériser complètement les mots bi-infinis apériodiques admettant des pièges à facteurs finis.

**Proposition 3.3.28.** *Soit  $x$  un mot bi-infini apériodique. Alors  $x$  admet un piège à facteurs fini si, et seulement si, il existe un morphisme de retour  $\phi$ , un entier  $m$  et un mot sturmien caractéristique  $y$  tel que  $x = S^m \phi(y)$ .*

*Démonstration.*

La proposition 3.3.20 montre que si  $x = S^m \phi(y)$  pour un morphisme de retour  $\phi$ , un entier  $m$  et un mot sturmien caractéristique  $y$ ,  $x$  a un piège à facteurs fini.

Soit  $x$  un mot bi-infini apériodique admettant un piège à facteurs fini. Alors  $x$  est un mot quasi-sturmien par la proposition 3.3.1. D'après la proposition 3.3.24, on a alors  $x = S^m \phi(y)$  avec  $\phi$  un morphisme de retour,  $m$  un entier et  $y$  un mot sturmien. Comme  $x$  a un piège à facteurs fini, le lemme 3.3.27 nous donne que  $y$  aussi admet un piège à facteurs fini. Donc, d'après la proposition 3.3.17,  $y$  est un mot sturmien caractéristique à décalage près. L'image d'un mot décalé étant le décalé de son image, on peut faire ressortir le décalage de  $y$  dans  $m$  pour conclure.

Nous avons enfin atteint notre but : nous avons établi quels mots bi-infinis apériodiques admettent des pièges à facteurs finis. La caractérisation fournie permet même de discriminer au sein des mots quasi-sturmiens lesquels admettent des pièges à facteurs finis et lesquels non. La proposition 3.3.28 fournit en particulier l'existence de mots bi-infinis apériodiques admettant des pièges à facteurs finis.

**Exemple 3.3.29.** *Soit  $x_{\text{Fib}} = \dots 0100101001010 \cdot 0100101001001 \dots$  le mot de Fibonacci bi-infini, et soit  $f$  le mot fini 0120. On pose comme substitution  $\phi : 0 \mapsto 120, 1 \mapsto 010120$ . C'est bien un morphisme de retour pour  $f$ . Alors  $\phi(x_{\text{Fib}}) = \dots 120010120120010120120 \cdot 120010120120120 \dots$  admet comme piège à facteurs  $\llbracket 0, 8 \rrbracket = \phi(\llbracket 0, 1 \rrbracket)$ .*



L'essentiel du travail effectué dans les derniers résultats est en fait la caractérisation des mots quasi-sturmien. Dans la preuve de la proposition 3.3.28, j'ai mis à profit le travail fait auparavant dans le cas sturmien. Il est intéressant de noter que les mots quasi-sturmiens admettant des pièges à facteurs finis sont les images des mots sturmiens caractéristiques (à décalage près) : cela pourrait motiver la notion de *mot quasi-sturmien caractéristique*. Malgré toute notre joie face à la dernière caractérisation, il reste quand même une question en suspens.

### Liens entre complexité et envergure d'un mot quasi-sturmien

Dans le cas sturmien, on avait prouvé dans le corollaire 3.3.19 que l'envergure d'un mot sturmien était soit 1, soit infinie. Intuitivement, la valeur 1 était une borne inférieure due à la proposition 3.3.1, et au fait que la complexité d'un mot sturmien est de la forme  $n + \underline{1}$ . Qu'en est-il dans le cas quasi-sturmien ? Je montre dans cette sous-section que le résultat sur les mots sturmiens se généralise aux mots quasi-sturmiens. Pour cela, j'ai besoin d'un résultat technique pour passer des pièges à facteurs des mots sturmiens à ceux des mots quasi-sturmiens.

**Lemme 3.3.30.** *Soit  $x$  un mot bi-infini, et  $\phi$  un morphisme de retour pour un mot  $f$  tel que  $\phi(x)$  est aperiodique. Si  $\Gamma = \llbracket n, n + k \rrbracket$  est un piège à facteurs de  $x$ , alors  $\phi(\Gamma)$  privé de ses  $|f|$  dernières positions est un piège à facteurs de  $\phi(x)$ .*

*Démonstration.*

D'abord, la proposition 3.2.11 implique que  $\phi(\Gamma)$  est bien un piège à facteurs de  $\phi(x)$ .

Puis dans un premier temps, montrons que  $\phi(\Gamma)$  contient au moins  $|f| + 1$  positions. Soit  $a \neq b \in \mathcal{A}$  : montrons que  $|\phi(a)\phi(b)| \geq |f| + 1$ . Considérons les deux mots  $f\phi(a)\phi(b)$  et  $f\phi(b)\phi(a)$ . Parce que  $\phi$  est un morphisme de retour pour  $f$ ,  $f\phi(a)\phi(b)$  et  $f\phi(b)\phi(a)$  ont  $f$  pour suffixe. Ces deux mots sont aussi de même longueur. Mais si  $|\phi(a)\phi(b)| \leq |f|$ , alors on aurait  $f\phi(a)\phi(b) = ff'$  avec  $f'$  un suffixe de  $f$  de longueur  $|\phi(a)\phi(b)|$ . Le même raisonnement tient sur  $f\phi(b)\phi(a)$ . Donc on aurait  $f\phi(a)\phi(b) = f\phi(b)\phi(a)$  : cela contredit l'injectivité du morphisme de retour  $\phi$  (voir la proposition 3.3.22). Donc on a bien  $|\phi(a)\phi(b)| \geq |f| + 1$ . Or, comme  $\Gamma$  est un piège à facteurs de  $x$ ,  $\Gamma$  doit capturer une occurrence de toutes les lettres de  $x$ , ce qui implique  $|\phi(\Gamma)| \geq \sum_{a \in \mathcal{A}} |\phi(a)|$  : on a bien  $|\phi(\Gamma)| \geq |f| + 1$ .

Comme  $\Gamma$  est un segment de  $\mathbb{Z}$ ,  $\phi(\Gamma)$  est aussi un segment de  $\mathbb{Z}$ . On écrit  $\phi(\Gamma) = \llbracket m, m + l \rrbracket$  (avec donc  $l \geq |f|$ ). Par le point 1 de la proposition 3.3.22,  $\phi(x)_{\llbracket m+l-|f|+1, m+l \rrbracket} = f$ . Soit  $\Delta = \llbracket m + l - |f| + 1, m + l \rrbracket$ , et soit  $\Gamma' = \llbracket m, m + l - |f| \rrbracket$  : on a bien  $\phi(\Gamma) = \Gamma' \sqcup \Delta$ . L'objectif final de cette preuve est de montrer que  $\Gamma'$  est un piège à facteurs de  $\phi(x)$ . Pour cela, montrons que tout facteur non vide  $g$  de  $\phi(x)$  ayant une occurrence commençant dans  $\Delta$  a une occurrence capturée par  $\Gamma'$ .

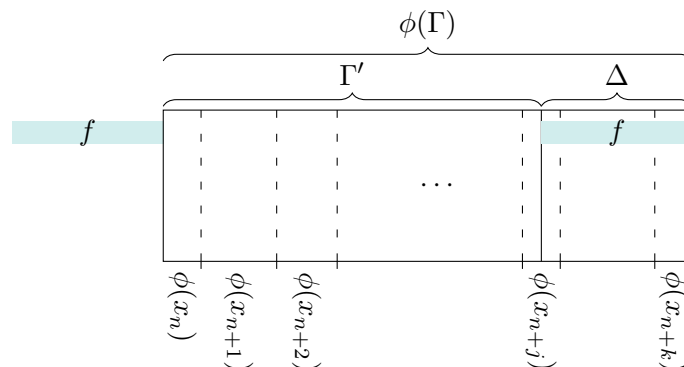
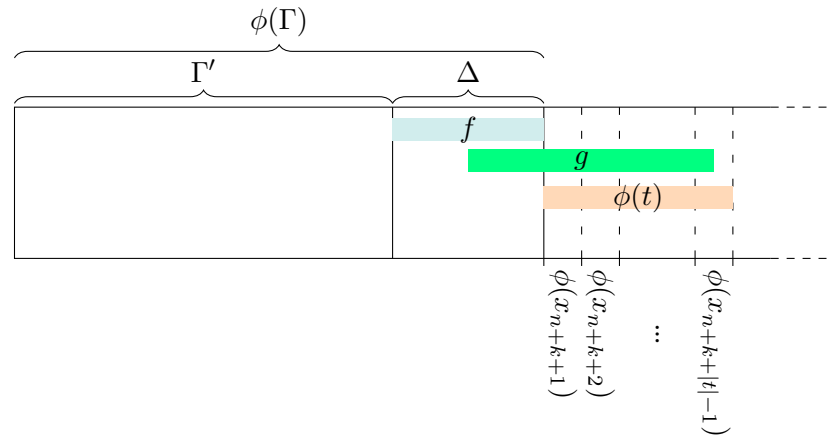


FIGURE 3.12 – Découpage commun le long de cette preuve.

J'introduis une remarque sur le découpage de  $\phi(\Gamma)$  en images de  $\phi$  : comme  $\phi$  est un morphisme de retour pour  $f$  et que  $\phi(\Gamma)$  tombe pile sur des images de  $\phi$ , en continuant à gauche de  $\phi(\Gamma)$ , on tombe sur une occurrence de  $f$ .

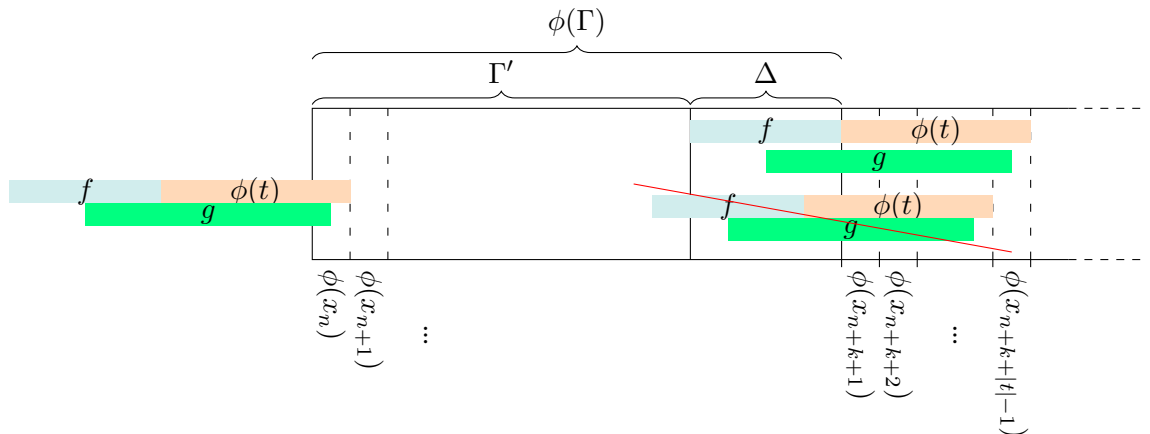
Je vais diviser la preuve en plusieurs cas à traiter.



**FIGURE 3.13** – Cas 1 :  $g$  n'est pas un facteur de  $f$ .

- Cas 1 :  $g$  n'est pas un facteur de  $f$  (voir figure 3.13).

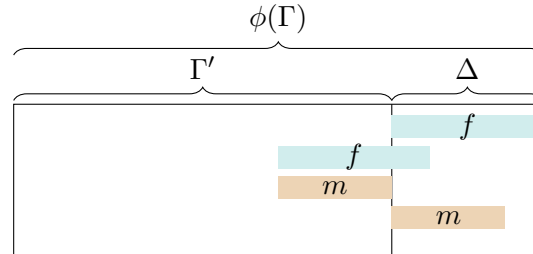
On note  $t$  le plus petit préfixe de  $x_{[n+k+1, +\infty[}$  tel que  $g$  est un facteur de  $f\phi(t)$ . Comme  $g$  n'est pas un facteur de  $f$ ,  $t$  est non vide. La minimalité de  $t$  assure que  $f\phi(t)$  contient l'occurrence la plus à gauche de  $g$  commençant dans  $\Delta$ . Comme  $\Gamma$  est un piège à facteurs de  $x$ ,  $t$  a une occurrence capturée par  $\Gamma$ . Cela correspond donc à une occurrence de  $\phi(t)$  dans  $\phi(x)$  capturée par  $\phi(\Gamma)$ . Comme  $\phi$  est un morphisme de retour pour  $f$ , cette occurrence est précédée par  $f$ . Comme  $t$  est minimal,  $g$  est aussi capturé par  $\phi(\Gamma)$ . Si cette nouvelle occurrence de  $g$  commençait elle aussi dans  $\Delta$ , alors on aurait trouvé une nouvelle occurrence de  $g$  commençant dans  $\Delta$  plus à gauche que celle de tout à l'heure, ce qui contredit la construction de  $t$  jusqu'alors. Donc cette occurrence de  $g$  est bien capturée par  $\Gamma'$ .



**FIGURE 3.14** – Cas 1 : en haut à droite l'occurrence de départ; à gauche, la minimalité de  $t$  assure que si  $f\phi(t)$  est capturé à gauche de  $\phi(\Gamma)$ ,  $g$  est capturé par  $\Gamma'$ ; en bas à droite,  $\phi(t)$  étant capturé par  $\phi(\Gamma)$ , la nouvelle occurrence de  $g$  est plus à gauche que l'occurrence de départ, qui était la plus à gauche commençant dans  $\Delta$ , donc le cas dessiné est impossible.

- Cas 2 :  $g$  est un facteur non vide de  $f$ , et une occurrence de  $f$  commence (resp. termine) dans  $\Gamma'$  (voir figure 3.15).

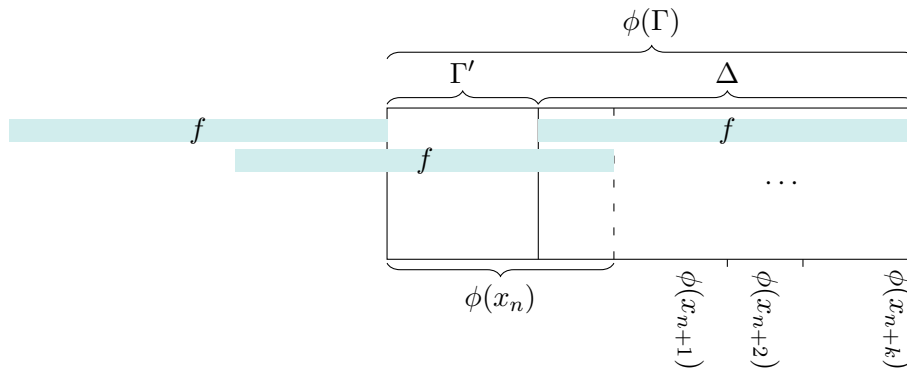
Si cette occurrence est complètement incluse dans  $\Gamma'$ ,  $g$  est capturé par  $\Gamma'$ . Sinon, on en déduit alors que  $\Gamma'$  contient à la fin un préfixe strict (resp. un suffixe strict)  $m$  de  $f$ . Donc  $f$  est un mot périodique de période  $|m|$ . Donc tout facteur de  $f$  a une occurrence dans  $f$  commençant dans son préfixe  $m$  qui est contenu dans  $\Gamma'$ , et donc  $g$  est couvert par  $\Gamma'$ .



**FIGURE 3.15** – Cas 2 :  $m$  est un préfixe de  $f$ , et par répétition de  $f$ ,  $f$  est un préfixe de  $m^2$ .

- Cas 3 :  $g$  est un facteur non vide de  $f$ , et aucune occurrence de  $f$  ne commence ni ne termine dans  $\Gamma'$  (voir figure 3.16).

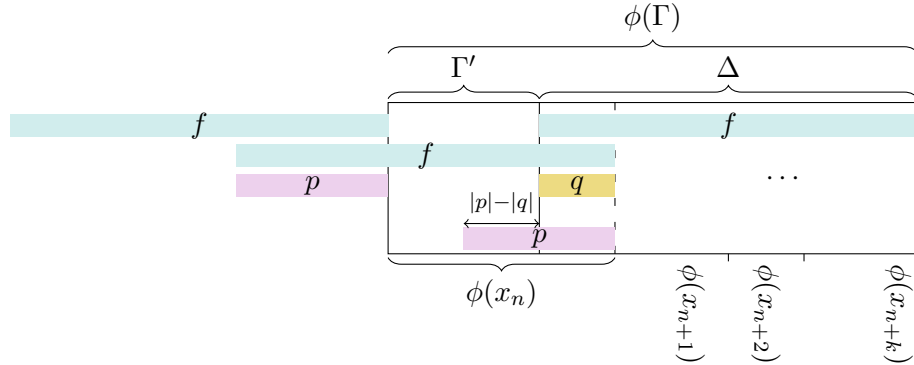
Comme  $f$  est un suffixe de  $f\phi(x_n)$ ,  $x_{\llbracket m, m+l-|f| \rrbracket} = x_{\Gamma'}$  est alors un préfixe strict de  $\phi(x_n)$ . Par construction de  $\Gamma'$ ,  $\phi(x_{\llbracket n+1, n+k \rrbracket})$  est donc un suffixe strict de  $f$ . De plus, comme  $f\phi(x_n)$  admet  $f$  comme suffixe,  $\phi(x_n)$  doit être un suffixe strict de  $f$  (sinon, on contredit qu'aucune occurrence de  $f$  ne commence dans  $\Gamma'$ ).



**FIGURE 3.16** – Cas 3 : résumé de la situation.

Dans ce cas, on note  $f = p\phi(x_n) = q\phi(x_{\llbracket n+1, n+k \rrbracket})$ . Si  $k = 1$ , comme  $\Gamma$  est un piège à facteurs de  $x$ , on a  $x_n \neq x_{\llbracket n+1, n+k \rrbracket}$ ; si  $k \neq 1$ , on a aussi trivialement  $x_n \neq x_{\llbracket n+1, n+k \rrbracket}$ , donc l'inégalité est vérifiée dans tous les cas. Et comme  $\phi$  est un morphisme de retour,  $\phi$  est injectif sur les mots finis, donc  $p \neq q$ , et comme ils sont tous les deux préfixes de  $f$ ,  $|p| \neq |q|$ . Supposons que  $|q| < |p|$ , l'autre cas se traite de manière analogue. Alors  $q$  est un préfixe de  $p$ . De plus,  $p$  et  $q$  sont aussi des suffixes de  $f$  (voir la figure 3.17). Donc  $q$  est un préfixe et un suffixe de  $p$ , donc  $p$  est périodique de période  $|p| - |q|$ . Par hypothèse du cas 3, les positions de  $\Gamma'$  correspondent à une occurrence de  $f'$  avec  $f = pf'q$ . Les facteurs de  $f$  touchant  $f'$  étant capturés par  $\Gamma'$ , il faut maintenant montrer que si  $u$  est facteur de  $p$  ou de  $q$ , alors  $u$  est quand même couvert par  $\Gamma'$ . Si  $u$  est un facteur de  $q$ , c'est aussi un facteur de  $p$ . Supposons donc que  $u$  est un facteur de  $p$ . Par périodicité de  $p$ , tout facteur de  $p$  a une occurrence commençant dans les

$|p| - |q|$  premières lettres de  $p$ . Comme  $p$  est aussi un suffixe de  $f$ , les  $|p| - |q|$  premières lettres de  $p$  sont incluses dans  $\Gamma'$ . Donc  $g$  est couvert par  $\Gamma'$ .



**FIGURE 3.17** – Cas 3 : visualisation de pourquoi  $u$  est bien couvert par  $\Gamma'$  dans ce cas.

On a couvert tous les cas :  $\Gamma'$  est bien un piège à facteurs de  $\phi(x)$ .

Je dispose maintenant des outils techniques nécessaires pour généraliser la proposition 3.3.17 et le corollaire 3.3.19.

**Proposition 3.3.31.** *Soit  $x$  un mot bi-infini apériodique. Alors  $x$  admet un piège à facteurs de diamètre  $k$  si, et seulement si,  $x$  est un mot quasi-sturmien vérifiant :*

1.  $x = S^m \phi(y)$  avec  $\phi$  un morphisme de retour,  $m$  un entier et  $y$  un mot sturmien caractéristique;
2.  $\exists N \geq 0, \forall n \geq N, p_x(n) = n + k'$  avec  $k' \leq k$ .

*Démonstration.*

Si  $x$  admet un piège à facteurs de diamètre  $k$ , les propositions 3.3.28 et 3.3.1 permettent de conclure. Soit  $x = S^m \phi(y)$  un mot quasi-sturmien de complexité  $p_x(n) = n + k'$  pour  $n$  assez grand, avec  $k' \leq k$ . Montrons que  $x$  admet un piège à facteurs de diamètre  $k$ . D'après la proposition 3.3.24, on peut supposer que  $\phi$  est un morphisme de retour pour un mot  $f$  tel que  $k' = |\phi(0)| + |\phi(1)| - |f| - 1$ . Le mot sturmien  $y$  admet un piège à facteurs de diamètre 1, contenant un 0 et un 1. Alors d'après le lemme 3.3.30,  $\phi(y)$  admet un piège à facteurs de diamètre  $|\phi(0)| + |\phi(1)| - |f| - 1 = k'$ . Donc  $x$  admet bien un piège à facteurs de diamètre  $k \geq k'$ .

**Corollaire 3.3.32.** *Soit  $x$  un mot bi-infini quasi-sturmien tel que sa complexité vérifie  $p_x(n) = n + k$  pour  $n$  assez grand. Alors :*

$$\text{env}(x) = \begin{cases} k & \text{si } x = S^m \phi(y) \text{ avec } y \text{ un mot sturmien caractéristique, } \phi \text{ une substitution et } m \in \mathbb{Z}, \\ \infty & \text{sinon.} \end{cases}$$

De plus, si  $\mathbb{X}$  est un shift quasi-sturmien de complexité  $n + k$  pour  $n$  assez grand, alors il y a une quantité infinie dénombrable d'éléments de  $\mathbb{X}$  d'envergure  $k$  et une quantité infinie indénombrable d'éléments d'envergure infinie.

*Démonstration.*

La dichotomie sur l'envergure se prouve de la même manière que la proposition 3.3.31.

En résumé, un mot bi-infini apériodique peut admettre un piège à facteurs fini, et j'en ai produit ici une caractérisation complète. Le lien entre complexité et envergure permet de voir que la technique développée dans la proposition 3.3.1 est en fait plus proche de la caractérisation que de la simple remarque. Les techniques employées dans cette section, si elles peuvent paraître absconses, sont en fait largement facilitées par l'utilisation des morphismes de retour et les propriétés qui leur sont propres. Il devrait être possible d'obtenir les mêmes résultats en n'utilisant que la propriété de reconnaissabilité [BPR23] des morphismes de retour, permettant de généraliser les résultats à tout morphisme apériodique.

### 3.4 Étude des pièges à facteurs infinis

Une fois la caractérisation complète des pièges à facteurs finis établie, il reste à traiter le cas des pièges à facteurs infinis. Le sujet est vaste, et la difficulté réside principalement dans la formulation de définitions intéressantes. Il est en effet facile d'obtenir des résultats qui peuvent paraître intéressants, mais ne sont pas vraiment pertinents :

**Proposition 3.4.1.** *Soit  $x$  un mot bi-infini récurrent. Alors  $x$  admet un piège à facteurs  $\Gamma$  mono-infini, c.-à-d.  $\inf(\Gamma) \neq -\infty$  ou  $\sup(\Gamma) \neq \infty$ .*

*Démonstration.*

Montrons d'abord que si  $x$  est récurrent, alors  $x$  est récurrent à gauche ou à droite. Pour cela, en notant  $f_n = x_{\llbracket -n, n \rrbracket}$  le facteur central de  $x$  pour  $n \geq 1$ ,  $f_n$  apparaît infiniment souvent dans  $x$ . On note  $\rho_n \subseteq \{-, +\}$  « l'ensemble des récurrences » de  $f_n$  :  $- \in \rho_n$  ssi  $f_n$  apparaît infiniment dans  $x_{\llbracket -\infty, 0 \rrbracket}$  ;  $+ \in \rho_n$  ssi  $f_n$  apparaît infiniment dans  $x_{\llbracket 0, \infty \rrbracket}$ .

Alors deux propriétés :

1. pour tout  $n \geq 1$ ,  $\rho_n \neq \emptyset$  par récurrence de  $x$  ;
2. pour tout  $n \geq 1$ ,  $\rho_n \supseteq \rho_{n+1}$  : si  $f_{n+1}$  apparaît infiniment souvent à gauche, alors  $f_n$ , qui est un de ses facteurs, aussi ; de même à droite.

Donc la suite  $(\rho_n)_{n \in \mathbb{N}}$  est une suite décroissante de  $\mathcal{P}(\{-, +\})$  dont aucun élément n'est l'ensemble vide : donc il existe  $\rho \in \bigcap_{n \in \mathbb{N}} \rho_n$ . Si  $\rho = -$ ,  $x$  est récurrent à gauche ; et si  $\rho = +$ ,  $x$  est récurrent à droite.

Pour en revenir à l'énoncé, il suffit alors de choisir  $\mathbb{Z}_- = \llbracket -\infty, 0 \rrbracket$  si  $x$  est récurrent à gauche, et  $\mathbb{N}$  si  $x$  est récurrent à droite pour obtenir un piège à facteurs mono-infini de  $x$ .

**Proposition 3.4.2.** *Il existe des mots bi-infinis n'admettant pas de piège à facteurs mono-infini.*

*Démonstration.*

Sur l'alphabet  $\{0, 1\}$ , considérons  $x = \dots 01^5 01^3 01^1 \cdot 01^2 01^4 01^6 0 \dots$ . Alors un facteur de la forme  $01^k 0$  n'apparaît qu'une seule fois dans  $x$ . Donc pour capturer le facteur  $01^k 0$ , un piège à facteurs de  $x$  sera obligé d'avoir une position au bon endroit, et donc d'aller arbitrairement loin à gauche et à droite.

Une idée naturelle pour obtenir des résultats intéressants sur les pièges à facteurs infinis est d'étudier leur densité. Si un mot n'admet pas de piège à facteurs finis, que peut-on affirmer sur la densité de ses pièges à facteurs infinis ? La réponse est nette : pour les mots récurrents, la densité minimale est nulle.

**Proposition 3.4.3.** *Soit  $x$  un mot bi-infini récurrent, et soit  $\eta$  une fonction croissante telle que  $\eta(n) \xrightarrow{n \rightarrow \infty} \infty$ . Alors  $x$  admet un piège à facteur  $\Gamma_\eta$  tel que  $\text{Card}(\Gamma_\eta \cap \llbracket -n, n \rrbracket) \leq \eta(n)$  pour tout  $n \in \mathbb{N}$ .*

*Démonstration.*

Soit  $\text{enum} : \mathbb{N} \rightarrow \mathcal{L}(x)$  une fonction énumérant les facteurs de  $x$ . Pour tout  $i \geq 0$ , soit  $\gamma_i \in \mathbb{Z}$  tel que  $\gamma_i$  capture  $\text{enum}(i)$  et  $\eta(|\gamma_i|) \geq i$ . Une telle position est assurée d'exister par récurrence de  $x$ . Alors  $\Gamma_\eta = \{\gamma_i \mid i \in \mathbb{N}\}$  est un piège à facteurs de  $x$ , et  $\Gamma_\eta \cap \llbracket -n, n \rrbracket \leq \eta(n)$ .

On remarque que cette preuve reste valable dans le cas mono-infini. Cette proposition montre qu'étudier l'ensemble des pièges à facteurs infinis d'un mot mono ou bi-infini ne donnera pas de résultat intéressant quant à la densité. Une possibilité est de s'intéresser à certains pièges à facteurs infinis particuliers : nous allons ici nous intéresser au cas des ensembles  $i + k\mathbb{Z}$  avec  $i \in \mathbb{Z}, k \geq 1$ , c.-à-d. les progressions arithmétiques. On peut dans ce cas relier une propriété dynamique aux pièges à facteurs.

**Définition 3.4.4** (Ensemble  $\text{Occ}_f^x(k)$ ). Soit  $k \geq 1, f \in \mathcal{A}^*$  et  $x \in \mathcal{A}^{\mathbb{Z}}$ . Alors l'ensemble des positions des occurrences de  $f$  dans  $x$  modulo  $k$  est :

$$\text{Occ}_f^x(k) = \{i \bmod k \mid x_{\llbracket i, i+|f|-1 \rrbracket} = f\}$$

On étend cette définition au cas où  $x$  est un mot fini.

**Lemme 3.4.5.** Soit  $x$  un mot bi-infini,  $k \geq 1, i \in \mathbb{Z}$ . Alors  $i + k\mathbb{Z}$  est un piège à facteurs de  $x$  si, et seulement si,  $\text{Occ}_f^x(k) \cap (\llbracket i - |f| + 1, i \rrbracket \bmod k) \neq \emptyset$  pour tout facteur non vide  $f$ .

*Démonstration.*

Si  $f$  est capturé par la position  $i + km$ , le début de l'occurrence capturée est en position  $p = i + km - j$  avec  $0 \leq j < |f|$ . Donc  $p \bmod k \in \text{Occ}_f^x(k)$  et  $p \bmod k = i - j \bmod k \in (\llbracket i - |f| + 1, i \rrbracket \bmod k)$ .  
Si  $\text{Occ}_f^x(k) \cap (\llbracket i - |f| + 1, i \rrbracket \bmod k) \neq \emptyset$ ,  $f$  est capturé par  $i + k\mathbb{Z}$ .  
On conclut en élargissant à tout facteur non vide de  $x$ .

Le lemme 3.4.5 exige alors une condition un peu pénible à remplir pour que  $i + k\mathbb{Z}$  soit un piège à facteurs de  $x$ . Pour l'étudier plus simplement, on introduit alors la définition suivante, due à Cassaigne, Kaboré et Tapsoba [CKT10] :

**Définition 3.4.6** (Récurrence modulaire). Soit  $x$  un mot bi-infini, alors  $x$  est modulairement récurrent si pour tout  $k \geq 1$  et tout facteur non vide  $f$ ,  $\text{Occ}_f^x(k) = \mathbb{Z}/k\mathbb{Z}$ .

La définition vient originellement du cas mono-infini. Kaboré et Tapsoba ont par exemple démontré que les mots sturmiens mono-infinis sont modulairement récurrents [KT07b].

**Remarque 3.4.7.** Si  $x$  est modulairement récurrent, alors  $x$  est récurrent : pour tout facteur  $f$  de  $x$ ,  $\text{Occ}_f^x(2) = \mathbb{Z}/2\mathbb{Z}$ , donc tout facteur de  $x$  a au moins deux occurrences dans  $x$ .

En revanche, un mot modulairement récurrent n'est pas forcément uniformément récurrent, et la réciproque aussi est fausse :

- Soit la suite  $(x_n)_{n \in \mathbb{N}}$  définie par  $x_0 = 01 \cdot 010$  et  $x_{n+1} = 0x_n 0^{a_n} x_n 0^{b_n} x_n 0$  avec le point central dans le  $x_n$  du milieu et des entiers  $a_n, b_n \geq 1$ ; on pose alors  $x = \lim_{n \rightarrow \infty} x_n$ . Alors, si  $a_n$  ou  $b_n \xrightarrow[n \rightarrow \infty]{} \infty$ ,  $x$  n'est pas uniformément récurrent ( $x$  possède des plages arbitrairement grandes de 0), mais, en manipulant les entiers  $a_n$  et  $b_n$ , on peut s'assurer que tous les facteurs de  $x$  ont des occurrences à toute valeur  $i$  modulo  $k$ .
- Soit  $\phi : 0 \mapsto 00, 1 \mapsto 01$  et  $x$  un mot uniformément récurrent sur  $\{0, 1\}$ . Alors  $\phi(x)$  est uniformément récurrent, mais  $\text{Occ}_1^{\phi(x)}(2) = \{1\} \neq \mathbb{Z}/2\mathbb{Z}$ .

**Corollaire 3.4.8.** *Soit  $x$  un mot bi-infini modulairement récurrent. Alors pour tout  $k \geq 1$ , pour tout  $i \in \mathbb{Z}/k\mathbb{Z}$ ,  $i + k\mathbb{Z}$  est un piège à facteurs de  $x$ .*

*Démonstration.*

Soit  $k \geq 1$  et  $i \in \mathbb{Z}/k\mathbb{Z}$ . Pour tout  $f$  facteur non vide de  $x$ ,  $i \in \text{Occ}_f^x(k)$ . Le lemme 3.4.5 permet de conclure.

Intuitivement, la récurrence modulaire signifie simplement que tout facteur a une occurrence commençant sur un point de toute progression arithmétique. Du point de vue des pièges à facteurs, il est clair qu'un mot modulairement récurrent admet toute progression arithmétique comme piège à facteurs. Du point de vue dynamique, la récurrence modulaire d'un mot retranscrit l'idée qu'il n'admet pas de structure périodique sous-jacente. On caractérise cette propriété à l'aide d'automates cellulaires :

**Définition 3.4.9** (Automate cellulaire). *Un automate cellulaire  $\pi : A^{\mathbb{Z}} \rightarrow B^{\mathbb{Z}}$  est une fonction continue telle que  $\pi \circ S = S \circ \pi$ .*

**Proposition 3.4.10.** *Soit  $x$  un mot uniformément récurrent. Alors  $x$  est modulairement récurrent si, et seulement si, pour tout automate cellulaire  $\pi$ ,  $\pi(x)$  est soit apériodique, soit un mot constant (c.-à-d. un mot de période 1).*

*Démonstration.*

D'abord, supposons qu'il existe  $\pi$  un automate cellulaire tel que  $\pi(x)$  n'est pas apériodique. Alors, comme  $x$  est uniformément récurrent,  $\pi(x)$  est uniformément récurrent; et donc comme  $\pi(x)$  n'est pas apériodique,  $\pi(x)$  est périodique. Supposons maintenant que  $\pi(x)$  est non constant. On écrit alors  $\pi(x) = w^{-\mathbb{N}} \cdot w^{\mathbb{N}}$ , avec  $w$  non constant (donc de longueur au moins 2). Quitte à décaler, on suppose que  $w = abw'$ . Soit  $f \in \mathcal{L}_n(x)$  tel que  $\pi(f) = b$  (il en existe au moins un). Alors  $0 \notin \text{Occ}_f^x(|w|)$ . Donc  $x$  n'est pas modulairement récurrent.

Supposons à présent que  $x$  n'est pas modulairement récurrent. Il existe  $f$  un facteur de  $x$ ,  $k \geq 1$  et  $i \in \mathbb{Z}/k\mathbb{Z}$  tel que  $i \notin \text{Occ}_f^x(k)$ . D'abord, pour tout  $a \in \text{Occ}_f^x(k)$ , il existe  $j_a \in \mathbb{Z}$  tel que  $a + kj_a$  est le début d'une occurrence de  $f$  dans  $x$ . Quitte à décaler  $x$ , on peut supposer que  $j_a \geq 0$ . On définit alors  $N = |f| + \max_{a \in \text{Occ}_f^x(k)} \{a + kj_a\}$ . On pose alors  $g = x_{[0, N]}$ . L'idée est que  $g$  contient tous les exemples d'occurrences de  $f$  modulo  $k$  dans  $x$ . En particulier,  $\text{Occ}_f^g(k) = \text{Occ}_f^x(k)$ .

Soit  $N_g$  la borne d'uniforme récurrence de  $g$  (c.-à-d. dans tout facteur de  $x$  de longueur  $N_g$ , il y a une occurrence de  $g$ ). Alors pour tout  $m \in \mathbb{Z}$ ,  $x_{[m, m+N_g-1]}$  contient une occurrence de  $g$ .

**Lemme 3.4.11.** *Pour  $m \in \mathbb{Z}$ ,*

$$\text{Occ}_f^{x_{[m, m+N_g-1]}}(k) = (\text{Occ}_f^x(k) + m) \pmod k.$$

*Démonstration.*

Soit  $m \in \mathbb{Z}$ . Si pour  $0 \leq i \leq N_g - |f|$ , on a  $x_{[m+i, m+i+|f|-1]} = f$ , alors  $i \pmod k \in \text{Occ}_f^{x_{[m, m+N_g-1]}}(k)$  et  $m + i \pmod k \in \text{Occ}_f^x(k)$ . Donc on a l'inclusion :

$$\text{Occ}_f^{x_{[m, m+N_g-1]}}(k) \subseteq (\text{Occ}_f^x(k) + m) \pmod k$$

Mais comme  $x_{[m, m+N_g-1]}$  contient une occurrence du mot  $g$ , qui lui-même contient autant

d'occurrence distinctes que  $x$ , on a les inégalités suivantes :

$$\begin{aligned} \text{Card}(\text{Occ}_f^{x_{\llbracket m, m+N_g-1 \rrbracket}}(k)) &\geq \text{Card}(\text{Occ}_f^g(k)) \\ &\geq \text{Card}(\text{Occ}_f^x(k)) = \text{Card}((\text{Occ}_f^x(k) + m) \bmod k) \end{aligned}$$

Donc on a bien égalité entre les ensembles comme annoncé.

Considérons alors la règle locale suivante :

$$\phi : \left( \begin{array}{l} \mathcal{A}^{N_g} \rightarrow \{0, 1\} \\ p \mapsto \begin{cases} 1 \text{ s'il y a une occurrence de } f \text{ dans } p \text{ dont la position est un multiple de } k, \\ 0 \text{ sinon} \end{cases} \end{array} \right)$$

Alors  $\phi(x_{\llbracket m, m+N_g-1 \rrbracket}) = 1$  ssi  $0 \in \text{Occ}_f^{x_{\llbracket m, m+N_g-1 \rrbracket}}(k)$  ssi  $0 \in (\text{Occ}_f^x(k) + m) \bmod k$ . Comme  $(\text{Occ}_f^x(k) + m) \bmod k = (\text{Occ}_f^x(k) + (m+k)) \bmod k$ , la suite  $((\text{Occ}_f^x(k) + m) \bmod k)_{m \in \mathbb{Z}}$  est  $k$ -périodique, et donc  $\phi(x)$  est aussi  $k$ -périodique.

Comme  $f$  est un facteur de  $x$ , il existe  $m \in \mathbb{Z}$  tel que  $0 \in (\text{Occ}_f^x(k) + m) \bmod k$ ; et comme  $i \notin \text{Occ}_f^x(k)$ , il existe  $m' \in \mathbb{Z}$  tel que  $0 \notin (\text{Occ}_f^x(k) + m') \bmod k$ . Donc  $\phi(x)$  est  $k$ -périodique et n'est pas un mot constant.

Il existe déjà dans la littérature plusieurs exemples de familles de mots modulairement récurrents.

**Théorème 3.4.12** ([Ber+18]). *Les mots dendriques récurrents sont modulairement récurrents.*

Ce dernier théorème inclut le cas des mots sturmiens déjà établi par Kaboré et Tapsoba. On peut aussi construire des mots quasi-sturmiens qui contredisent la réciproque du corollaire 3.4.8 :

**Exemple 3.4.13.** *Soit  $x_{\text{Fib}}$  le mot de Fibonacci bi-infini et soit  $\phi : 0 \mapsto 01, 1 \mapsto 10$ . Alors  $\phi(x_{\text{Fib}})$  n'est pas modulairement récurrent, mais admet tout  $i + k\mathbb{Z}$  comme piège à facteurs. D'après la proposition 3.4.10,  $\phi(x_{\text{Fib}})$  admet une image périodique non constante par un automate cellulaire : l'automate cellulaire  $\psi$  de règle locale suivante convient.*

$$\left( \begin{array}{l} \{0, 1\}^{\llbracket 0, 6 \rrbracket} \rightarrow \{0, 1\} \\ w \mapsto \begin{cases} 0 \text{ si } 11 \text{ apparaît uniquement dans des positions paires de } w, \\ 1 \text{ si } 11 \text{ apparaît uniquement dans des positions impaires de } w, \\ 2 \text{ sinon.} \end{cases} \end{array} \right)$$

Plus précisément,  $\psi(\phi(x_{\text{Fib}})) = (10)^{-\mathbb{N}} \cdot (10)^{\mathbb{N}}$ .

*Démonstration.*

Par le théorème 3.3.25,  $\phi(x_{\text{Fib}})$  est bien quasi-Sturmien. Comme  $x_{\text{Fib}}$  est de type 0,  $00 \preceq_f x$  et  $11 \not\preceq_f x_{\text{Fib}}$ . Alors  $01 \preceq_f \phi(x_{\text{Fib}})$  ne peut apparaître qu'à des positions paires : un 0 en position impaire ne peut pas être suivi d'un 1 (sinon, 11 apparaîtrait dans  $x_{\text{Fib}}$ ). Donc  $\text{Occ}_{01}^{\phi(x_{\text{Fib}})}(2) = \{0\} \neq \mathbb{Z}/2\mathbb{Z}$ ,  $\phi(x_{\text{Fib}})$  n'est pas modulairement récurrent.

Maintenant, soit  $i \in \mathbb{Z}$  et  $k \geq 1$ . Montrons que  $i + k\mathbb{Z}$  est un piège à facteurs de  $\phi(x_{\text{Fib}})$ . Prenons un



facteur  $f$  de  $\phi(x_{\text{Fib}})$  : montrons que  $f$  est capturé par  $i + k\mathbb{Z}$ .

D'abord on traite le cas  $|f| = 1$ . Considérons l'ensemble  $(i + k\mathbb{Z}) \cap 2\mathbb{Z}$  : si l'ensemble n'est pas vide, c'est une progression arithmétique, et comme on est dans  $2\mathbb{Z}$ , par la forme de  $\phi$ , cela correspond à une progression arithmétique dans  $x_{\text{Fib}}$ . Donc  $i + k\mathbb{Z}$  capture bien les deux lettres. Si l'ensemble est vide, c'est que  $i + k\mathbb{Z} \subseteq 2\mathbb{Z} + 1$ . On refait le même raisonnement, les deux lettres sont bien capturées.

Maintenant on traite le cas  $|f| > 1$ . En ajoutant au plus une lettre à gauche et à droite, on peut étendre  $f$  en  $f' = \phi(g)$  avec  $g$  un facteur de  $x_{\text{Fib}}$ . Comme, d'après le théorème 3.4.12,  $x_{\text{Fib}}$  est modulairement récurrent, pour tout  $k \geq 1$ ,  $\text{Occ}_g^{x_{\text{Fib}}}(k) = \mathbb{Z}/k\mathbb{Z}$ . Donc  $\text{Occ}_{f'}^{x_{\text{Fib}}}(k)$  contient au moins  $2\mathbb{Z}/k\mathbb{Z}$ . Selon qu'on a obtenu  $f'$  en étendant  $f$  d'une lettre à gauche ou pas, on obtient alors que  $\text{Occ}_f^{x_{\text{Fib}}}(k) = \text{Occ}_{f'}^{x_{\text{Fib}}}(k)$  ou  $(\text{Occ}_{f'}^{x_{\text{Fib}}}(k) + 1) \bmod k$ . Donc  $\text{Occ}_f^{x_{\text{Fib}}}(k) \supseteq 2\mathbb{Z}/k\mathbb{Z}$  ou  $\text{Occ}_f^{x_{\text{Fib}}}(k) \supseteq (2\mathbb{Z} + 1)/k\mathbb{Z}$ . Comme  $|f| > 1$ , on en déduit que tout  $i + k\mathbb{Z}$  capture  $f$ .

Enfin, montrons que  $\psi(\phi(x_{\text{Fib}})) = (10)^{-\mathbb{N}} \cdot (10)^{\mathbb{N}}$ . Comme  $\phi$  est continue,  $\phi(x_{\text{Fib}})$  est uniformément récurrent, et pour le facteur 11, la constante d'uniforme récurrence est 7. Si  $\phi(x_{\text{Fib}})_{[n, n+1]} = 11$ , alors de par la forme de  $\phi$ ,  $n$  est impair. Donc quand on lit  $\phi(x_{\text{Fib}})_{[n, n+6]}$ , on trouve forcément un 11, par définition de la constante d'uniforme récurrence; et tous les 11 auront une occurrence sur des nombres impairs dans  $[n, n+6]$ , donc  $\phi(x_{\text{Fib}})_{[n, n+6]} \neq 2$ . Donc, en considérant  $\phi(x_{\text{Fib}})_{[n, n+6]}$  comme un mot à part entière,  $\phi(x_{\text{Fib}})_{[n, n+6]}$  contient un 11 en position paire si, et seulement si,  $n$  est impair. D'où le fait que  $\psi(\phi(x_{\text{Fib}})) = (10)^{-\mathbb{N}} \cdot (10)^{\mathbb{N}}$ .

La récurrence modulaire et l'admission de toute progression arithmétique comme piège à facteurs ne sont donc pas équivalentes. C'est jusque-là notre meilleure tentative pour donner une caractérisation dynamique à une propriété relative aux pièges à facteurs. Il paraît toutefois difficile dans le contexte d'aller plus loin : j'ai déjà fourni ici plusieurs contre-exemples montrant que toutes les notions introduites ici ne coïncident pas.

### 3.5 Généralisation aux familles de mots : les pièges à facteurs universels

La proposition 3.4.3 montre qu'une étude plus poussée des pièges à facteurs infinis a un intérêt limité d'un point de vue purement combinatoire : je vais alors procéder à une transition vers un point de vue plus dynamique. Pour prolonger le travail sur les pièges à facteurs, une idée est plutôt d'étudier un piège à facteurs communs à plusieurs mots bi-infinis.

**Définition 3.5.1** (Piège à facteurs universel). Soit  $\mathcal{F} \subseteq \mathcal{A}^{\mathbb{Z}}$  un ensemble de mots. Alors  $\Gamma \subseteq \mathbb{Z}$  est un piège à facteurs universel de  $\mathcal{F}$  si  $\Gamma$  est un piège à facteurs pour tout  $x \in \mathcal{F}$ .

**Exemple 3.5.2.** Soit  $\mathcal{F} = \{ \dots 0000 \cdot 1111 \dots, \dots 1111 \cdot 0000 \dots \}$ , alors  $[-1, 0]$  est un piège à facteurs universel de  $\mathcal{F}$ .

Parmi les familles sur lesquels étudier les pièges à facteurs, les sous-shifts (voir les définitions 1.1.21) paraissent des candidats naturels, de par leurs propriétés dynamiques. Dans la transition de la combinatoire vers la dynamique, les clôtures des orbites permettent de généraliser l'étude des mots à celle des sous-shifts. En faisant cette généralisation, on obtient alors un résultat qui contraste avec l'étude des mots individuels.

**Proposition 3.5.3.** Soit  $x$  un mot bi-infini, et  $\overline{\mathcal{O}(x)}$  la clôture de son orbite. Alors  $\overline{\mathcal{O}(x)}$  admet un piège à facteurs universel fini si, et seulement si,  $x$  est un mot périodique.

*Démonstration.*

D'abord, si  $x = w^{-\mathbb{N}} \cdot w^{\mathbb{N}}$  pour un certain mot non-vide  $w$ , alors tout élément de  $\overline{\mathcal{O}(x)}$  est  $|w|$ -périodique, et  $\llbracket 0, |w| - 1 \rrbracket$  est un piège à facteurs universel de  $\overline{\mathcal{O}(x)}$ .

Réciproquement, supposons que  $\overline{\mathcal{O}(x)}$  admette un piège à facteurs fini universel  $\Gamma = \llbracket \gamma, \gamma + k \rrbracket$ . Alors tous les facteurs de longueur  $n$  apparaissent dans  $x$  dans les positions  $\llbracket \gamma - n, \gamma + k + n \rrbracket$ . Et comme  $\Gamma$  est un piège à facteurs universel, c'est vrai pour tout  $S^j(x)$ . On en déduit que  $x_{\llbracket j, j + \gamma + k + 2n \rrbracket}$  contient tout facteur de  $x$  de longueur  $n$  : donc  $x$  est un mot uniformément récurrent.

Supposons maintenant que  $x$  n'est pas périodique. Comme  $x$  est aussi uniformément récurrent,  $x$  n'est ni positivement ni négativement périodique. Donc on peut utiliser la proposition 3.3.28 et affirmer que  $x$  est un mot quasi-sturmien. Mais alors d'après le corollaire 3.3.32, il y aurait dans  $\overline{\mathcal{O}(x)}$  un élément n'admettant aucun piège à facteurs fini : absurde ! Donc  $x$  n'est pas apériodique, et par uniforme récurrence,  $x$  est périodique.

Ce résultat est à la fois triste et rassurant. D'abord, ce résultat montre que l'étude de piège à facteurs dans le cas des sous-shifts est limitée, au moins sur  $\mathbb{Z}$ . Il devrait être possible de généraliser la question des pièges à facteurs fini sur  $\mathbb{Z}^d$  (ou même sur d'autres structures plus compliquées, comme les groupes, qui sont étudiés dans le cadre du problème du domino [BJ08; AB23]). Savoir si la proposition 3.5.3 reste valable sur des structures plus complexes est intéressante, mais sort du cadre de cette thèse. Dans le cas de  $\mathbb{Z}$ , cela montre qu'une étude systématique des pièges à facteurs de sous-shifts n'est pas très féconde. D'un autre côté, cette généralisation aux sous-shifts permet de retrouver un résultat similaire au cas des pièges à facteurs pour les mots mono-infinis : on obtient alors une caractérisation des mots périodiques bi-infinis à l'aide des pièges à facteurs, comme dans le cas mono-infini. On peut étendre la proposition 3.5.3 au cas des sous-shifts de manière générale.

**Corollaire 3.5.4.** *Soit  $\mathbb{X}$  un sous-shift. Alors  $\mathbb{X}$  admet un piège à facteurs universel fini si, et seulement si,  $\mathbb{X}$  est une union finie de sous-shifts périodiques finis.*

*Démonstration.*

Si  $\mathbb{X}$  est une union finie de sous-shifts périodiques finis,  $\mathbb{X}$  est une union finie d'orbites de mots périodiques, et la proposition 3.5.3 permet de conclure en prenant l'union des pièges à facteurs de chaque orbite.

Si  $\mathbb{X}$  admet un piège à facteurs fini  $\Gamma$ , considérons  $x \in \mathbb{X}$ . Alors  $\overline{\mathcal{O}(x)} \subseteq \mathbb{X}$  admet un piège à facteurs fini, donc par la proposition 3.5.3,  $x$  est périodique. Donc  $\mathbb{X}$  est un sous-shift périodique. Donc  $\mathbb{X}$  est une union finie de sous-shifts périodiques finis par la proposition 3.8 de [BDJ08]<sup>2</sup>.

**Remarque 3.5.5.** *Un sous-shift est périodique si, et seulement si, il est fini.*

L'étude des pièges à facteurs universels finis dans les cas des sous-shifts est donc assez limitée. Nous pourrions alors nous tourner vers le rapport entre envergure de la clôture d'une orbite et période minimale : l'envergure est plus petite que la période minimale. Mais la réciproque n'est pas assurée.

**Exemple 3.5.6.** *Soit  $x = (abba)^{-\mathbb{N}} \cdot (abba)^{\mathbb{N}}$ . Alors  $\llbracket 0, 2 \rrbracket$  est un piège à facteurs de tout  $S^i(x)$ , donc de  $\overline{\mathcal{O}(x)}$ , mais  $x$  est de période minimale 4. Donc  $\text{env}(\overline{\mathcal{O}(x)}) < 4$ .*

On peut alors, de nouveau, étendre la question au cas des pièges à facteurs infinis. Mais de nouveau, nos espoirs sont douchés, et il est possible de diluer infiniment un piège à facteurs infini.

2. Une reformulation plus accessible à mon goût est disponible dans [Gay23, Proposition 2.9].

**Proposition 3.5.7.** *Soit  $\mathbb{X}$  un sous-shift minimal, et soit  $\eta : \mathbb{N} \rightarrow \mathbb{N}$  une fonction croissante telle que  $\eta(n) \xrightarrow[n \rightarrow \infty]{} \infty$ . Alors  $\mathbb{X}$  admet un piège à facteurs universel  $\Gamma$  tel que  $\text{Card}(\Gamma \cap \llbracket -n, n \rrbracket) \leq \eta(n)$  pour tout  $n \geq 0$ .*

*Démonstration.*

Soit  $\text{enum} : \mathbb{N} \rightarrow \mathcal{L}(\mathbb{X})$  une énumération des facteurs de  $\mathbb{X}$ . Par minimalité, en utilisant le théorème 1.1.28, pour tout  $i \in \mathbb{N}$ , on dispose d'une constante d'uniforme récurrence pour  $\text{enum}(i)$ , c.-à-d. que tout facteur de  $\mathbb{X}$  de longueur au moins  $c_i$  contient au moins une occurrence de  $\text{enum}(i)$ . On pose alors  $\Gamma = \bigcup_{i \in \mathbb{N}} \llbracket \gamma_i, \gamma_i + c_i \rrbracket$  avec  $\gamma_i$  choisi assez grand pour que  $\text{Card}(\Gamma \cap \llbracket -n, n \rrbracket) \leq \eta(n)$  pour tout  $n \geq 0$ . Par propriété de  $c_i$ ,  $\llbracket \gamma_i, \gamma_i + c_i \rrbracket$  capture une occurrence de  $\text{enum}(i)$  dans tout élément de  $\mathbb{X}$ . Donc  $\Gamma$  vérifie les conditions de densité imposées et est bien un piège à facteurs universel de  $\mathbb{X}$ .

C'est une généralisation de la proposition 3.4.3 au cas des sous-shifts. On est donc condamné à choisir entre les pièges à facteurs universels finis, qui ne conviennent que pour les sous-shifts périodiques, et les pièges à facteurs universels infinis, qu'on peut infiniment diluer.

Dans nos derniers efforts, il est enfin possible de combiner les deux extensions proposées jusqu'ici : on peut enfin étudier les progressions arithmétiques comme pièges à facteurs des sous-shifts. Plus précisément, je vais étudier des propriétés qui impliqueront qu'un sous-shift admet toutes les progressions arithmétiques comme pièges à facteurs.

**Définition 3.5.8** (Minimalité totale). *Soit  $\mathbb{X}$  un sous-shift. On dit que  $\mathbb{X}$  est totalement minimal si  $\mathbb{X}$  est minimal sous l'action de  $S^k$  pour tout  $k \geq 1$ .*

**Proposition 3.5.9.** *Un sous-shift  $\mathbb{X}$  est totalement minimal si, et seulement si,  $\mathbb{X}$  est la clôture de l'orbite d'un mot modulairement et uniformément récurrent; en particulier, tout mot de  $\mathbb{X}$  est modulairement et uniformément récurrent.*

*Démonstration.*

D'abord, supposons que  $\mathbb{X}$  est totalement minimal, et soit  $x \in \mathbb{X}$ . Comme  $\mathbb{X}$  est minimal,  $x$  est uniformément récurrent et  $\mathbb{X} = \overline{\mathcal{O}(x)}$ . Supposons alors par l'absurde qu'il existe  $k \geq 2$  et  $f$  un facteur de  $x$  tel que  $\text{Occ}_f^x(k) \neq \mathbb{Z}/k\mathbb{Z}$ . On a alors que  $\text{Occ}_f^{S^k(x)}(k) = (\text{Occ}_f^x + 1) \bmod k \neq \text{Occ}_f^x(k)$ . Soit alors  $\mathbb{Y} = \{y \in \mathbb{X} \mid \text{Occ}_f^y(k) \subseteq \text{Occ}_f^x(k)\}$ . L'ensemble  $\mathbb{Y}$  est invariant par  $S^k$  et  $S(x) \notin \mathbb{Y}$  donc  $\mathbb{Y} \neq \mathbb{X}$ ; de plus,  $x \in \mathbb{Y}$  donc  $\mathbb{Y}$  est non vide. Montrons que  $\mathbb{Y}$  est fermé. Soit  $(y^{(n)})_{n \in \mathbb{N}} \in \mathbb{Y}^{\mathbb{N}}$  telle que  $y^{(n)} \xrightarrow[n \rightarrow \infty]{} y$  : alors si  $y_{\llbracket p, p+|f|-1 \rrbracket} = f$ , il existe  $n \in \mathbb{N}$  tel que  $y_{\llbracket p, p+|f|-1 \rrbracket}^{(n)} = f$  : donc si  $p \in \text{Occ}_f^y(k)$ ,  $p \in \text{Occ}_f^{y^{(n)}}(k) \subseteq \text{Occ}_f^x(k)$ . Donc  $\mathbb{Y}$  est fermé : cela contredit la minimalité de  $\mathbb{X}$  par  $S^k$ . Donc  $x$  est aussi modulairement récurrent.

Soit maintenant  $x$  un mot modulairement et uniformément récurrent. Alors en notant  $\mathbb{X}$  la clôture de son orbite par  $S$ ,  $\mathbb{X}$  est minimal. En particulier, tous les éléments de  $\mathbb{X}$  partagent le même langage. Pour montrer que  $\mathbb{X}$  est totalement minimal, il suffit de montrer que  $\mathbb{X}$  est égal à la clôture de l'orbite par  $S^k$  de tout  $y \in \mathbb{X}$ . Soit  $f$  un facteur de  $x$ . Comme  $x$  est modulairement récurrent, il existe un facteur  $g$  de  $x$  dans lequel  $f$  apparaît à toutes les positions modulo  $k$ . Par minimalité,  $g$  est aussi un facteur de  $y$  : donc  $\text{Occ}_f^y(k) = \mathbb{Z}/k\mathbb{Z}$ . Donc  $S^{nk}(y)_{\llbracket 0, |f|-1 \rrbracket} = f$  pour un certain  $n \in \mathbb{Z}$ . Comme cela tient pour tout facteur de  $x$  et donc pour tout facteur de tout élément de  $\mathbb{X}$ , pour tout  $z \in \mathbb{X}$ ,  $z$  est dans la clôture de l'orbite de  $y$  sous  $S^k$ .

On peut en déduire alors des résultats sur les progressions arithmétiques comme pièges à facteurs universels de certains shifts.

**Proposition 3.5.10.** *Soit  $\mathbb{X}$  un shift totalement minimal. Alors toutes les progressions arithmétiques sont des pièges à facteurs universels de  $\mathbb{X}$ .*

*Démonstration.*

Dans un shift totalement minimal, tout mot est uniformément et modulairement récurrent par la proposition 3.5.9. Donc toute progression arithmétique est un piège à facteurs de tout élément de  $\mathbb{X}$ .

**Exemple 3.5.11.** *La réciproque de la proposition 3.5.10 est fautive : en reprenant les notations de l'exemple 3.4.13, la clôture de l'orbite de  $\phi(x_{\text{Fib}})$  n'est pas totalement minimale mais admet toute progression arithmétique comme piège à facteurs universel.*

Enfin, il est possible de donner une caractérisation des sous-shifts totalement minimaux : le sens indirect de la proposition suivante a été établi par Paul [Pau75, Corollary 3.3].

**Proposition 3.5.12.** *Un sous-shift minimal  $\mathbb{X}$  est totalement minimal si, et seulement si,  $\mathbb{X}$  est complètement apériodique, c.-à-d.  $\pi(\mathbb{X})$  est soit un mot constant, soit un sous-shift apériodique pour tout automate cellulaire  $\pi$ .*

*Démonstration.*

Soit  $\mathbb{X}$  un sous-shift totalement minimal et soit  $\pi$  un automate cellulaire. Soit  $x \in \mathbb{X}$ . Comme  $x$  est modulairement et uniformément récurrent par la proposition 3.5.9,  $\pi(x)$  est alors soit apériodique soit constant d'après la proposition 3.4.10. Si  $\pi(x)$  est constant, comme  $\mathbb{X}$  est la clôture de l'orbite de  $x$  sous l'action de  $S$  et que  $\pi$  commute avec  $S$ ,  $\mathbb{X}$  est un point constant. Donc soit  $\pi(\mathbb{X})$  est un mot constant, soit c'est un shift apériodique.

Réciproquement, supposons que  $\mathbb{X}$  est complètement apériodique. Alors tout élément  $x \in \mathbb{X}$  est aussi complètement apériodique. Donc d'après la proposition 3.4.10,  $x$  est modulairement récurrent. Comme  $\mathbb{X}$  est minimal, c'est bien la clôture de l'orbite de  $x$  qui est un mot modulairement et uniformément récurrent : donc d'après la proposition 3.5.9,  $\mathbb{X}$  est totalement minimal.

## 3.6 Conclusion

Dans ce chapitre, j'ai étudié et caractérisé les mots bi-infinis admettant des pièges à facteurs fini. Les résultats vont à l'encontre de ce qui existe sur  $\mathbb{N}$ , montrant la variété combinatoire qui peut exister rien qu'en dimension 1. Parmi les outils les plus utiles dans ce chapitre se trouvent la désubstitution d'un piège à facteurs (ou d'un sous-ensemble de  $\mathbb{Z}$  en général), ainsi que les caractérisations S-adiques des mots sturmiens. Avec Hellouin et Gheeraert, nous avons réétabli des caractérisations des mots quasi-sturmiens dans le cas bi-infini, remettant en lumière les travaux de Paul, Alessandri et Heinis.

J'ai beaucoup utilisé les propriétés des morphismes de retour, qui paraissent, à la lumière de ce qui est produit dans ce chapitre, un outil particulièrement précis et facile d'utilisation pour obtenir des résultats combinatoires. Il devrait cependant être possible de relâcher les conditions de plusieurs des propositions établies, et de se contenter de substitutions apériodiques, ou de n'utiliser que des propriétés de reconnaissabilité.

L'extension dynamique à la fin a permis de voir les limites de la notion de piège à facteurs : en tant qu'objet combinatoire, il permet plus de finesse qu'en tant qu'objet dynamique. Une généralisation intéressante des pièges à facteurs serait sur d'autres structures, par exemple sur les configurations de  $\mathbb{Z}^2$ . Une note intéressante à ce sujet est que plusieurs des preuves que nous faisons se reposent sur le théorème de Morse-Hedlund. Dans  $\mathbb{Z}^2$ , l'équivalent du théorème de Morse-Hedlund est la conjecture de Nivat : comme le dit bien son nom, c'est à l'heure actuelle une

conjecture qui n'a pas été prouvée. Plus intéressant encore, l'équivalent de la conjecture de Nivat sur  $\mathbb{Z}^3$  est faux : donc il est crédible que les preuves établies ici s'effondrent si on généralise la notion de piège à facteurs sur  $\mathbb{Z}^3$ .



# 4 Combinatoire des représentations S-adiques et problème du $X$ -domino

Cet endroit est étrange. [...] Le temps est figé. Et l'espace instable. C'est un monde qui enfreint les lois les plus élémentaires.

---

Cynthia, *Pokémon Platine*

## 4.1 Introduction générale du chapitre

Dans ce chapitre, j'examine des questions supplémentaires que j'ai abordées durant cette thèse, et qui n'ont pas conduit à des publications. L'approche y est plus ouverte et exploratoire que dans les chapitres précédents. Le chapitre se divise en deux sections, une sur les représentations S-adiques multidimensionnelles, et une autre sur une variante du problème du domino. Les deux sections suivantes pourraient être considérées comme des chapitres à part entière : à des fins de cohérence éditoriale de cette thèse, j'ai préféré les regrouper dans un petit chapitre à part. Je ferai donc des introductions séparées propres à chaque section.

## 4.2 Problème de l'universalité des représentations S-adiques

### 4.2.1 Introduction

Dans cette section, j'explore les limites combinatoires des représentations S-adiques. Comme expliqué dans le premier chapitre, les représentations S-adiques ont été étudiées le plus souvent à des fins dynamiques depuis leur création par Host et Ferenczi [Fer96]. Mais il est possible de les envisager comme une généralisation du modèle substitutif, et donc comme un modèle combinatoire de construction de mots infinis. Dans ce cas, la généralisation de ce modèle au-delà de  $\mathbb{N}$  ou  $\mathbb{Z}$  se pose : le modèle substitutif a déjà été généralisé par exemple dans le cas du plan et des pavages géométriques [Fra08]. Dans ce cadre, le modèle substitutif a été particulièrement prolifique : le premier exemple connu de pavage géométrique apériodique, le pavage de Penrose [Pen74; Pen84]<sup>1</sup>, bénéficie de descriptions substitutives permettant d'explorer ses propriétés [Gar89]; 50 ans plus tard, c'est en utilisant le modèle substitutif que Goodman-Strauss a prouvé qu'une tuile géométrique appelée le chapeau (en anglais *the hat*) pave apériodiquement le plan [Smi+23]. D'autres exemples incluent le pavage d'Ammann-Beenker [GS89] ou les pavages Planar Rosa de Kari et Lutfalla [KL23]. La généralisation du modèle S-adique à d'autres structures que les mots pourrait donc avoir un potentiel intéressant.

Mais avant de chercher des applications de l'extension du modèle, il est utile d'en comprendre les limites combinatoires. Et dans l'étude de ce modèle, la question de l'universalité est assez naturelle : toute configuration infinie dispose-t-elle d'une représentation S-adique ? Il est connu, par exemple, que sur  $\mathbb{N}$ , le modèle substitutif n'est pas

---

1. Le premier article a introduit les pavages de Penrose, mais je n'ai pas réussi à accéder à son contenu; le second est accessible plus facilement.

universel : tout mot infini n'est pas purement substitutif, ni même morphique<sup>2</sup>. Cette limitation du modèle substitutif, contrairement à ce qu'on pourrait imaginer, peut être perçue comme l'une de ses forces : étudier le modèle substitutif, c'est étudier un modèle contraint, où la substitution est un paramètre relativement facile à étudier et à moduler. Dans le modèle S-adique, on multiplie le nombre de paramètres en utilisant plusieurs substitutions : le modèle reste-t-il contraint de la même manière ? Qu'en est-il des généralisations du modèle S-adique à des structures plus complexes que  $\mathbb{N}$  ?

Dans la première sous-section, je résume le cas des représentations S-adiques sur  $\mathbb{N}$ , déjà expliqué dans [BD14]. Puis j'explore le cas de deux généralisations du modèle S-adique sur des structures plus complexes : d'abord, j'aborde le modèle d'Aubrun-Sablik sur  $\mathbb{N}^d$  [AS14]. Puis j'introduis un modèle de représentations S-adiques sur les arbres, inspiré du modèle substitutif de Baraviera-Leplaideur [BL22]. Dans ces deux modèles, j'explore le problème de l'universalité, et présente différents arguments justifiant que ces modèles sont contraints.

### 4.2.2 Universalité des représentations S-adiques sur $\mathbb{N}$

Dans le cas de  $\mathbb{N}$ , le modèle S-adique utilisé est celui défini dans le premier chapitre. Le résultat suivant est dû à Cassaigne<sup>3</sup> :

**Proposition 4.2.1** ([BD14]). *Soit  $\mathcal{A}$  un alphabet fini et  $\lambda \notin \mathcal{A}$  une lettre de travail.*

*Alors il existe  $\mathcal{C}$  une famille finie de substitutions sur l'alphabet  $\mathcal{A} \cup \{\lambda\}$  telle que pour tout mot  $x \in \mathcal{A}^{\mathbb{N}}$ , il existe  $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$  telle que  $x = \lim_{n \rightarrow \infty} \sigma_{\llbracket 0, n \rrbracket}(\lambda)$ .*

Autrement écrit, il existe une famille de substitutions qui génère le shift plein.

*Démonstration.*

Pour  $a \in \mathcal{A}$ , posons les substitutions suivantes :

$$\sigma_a : \begin{cases} b \in \mathcal{A} \mapsto b, \\ \lambda \mapsto \lambda a \end{cases}, \tau_a : \begin{cases} b \in \mathcal{A} \mapsto b, \\ \lambda \mapsto a \end{cases}$$

Alors pour tout mot fini  $m \in \mathcal{A}^+$ ,

$$m = \tau_{m_0} \circ \sigma_{m_1} \circ \sigma_{m_2} \circ \cdots \circ \sigma_{m_{|m|-1}}(\lambda)$$

On le prouve par récurrence sur la longueur de  $m$  : pour  $|m| = 1$ ,  $m = \tau_{m_0}(\lambda)$ . Puis si  $m$  est de longueur  $n + 1$ , on a :

$$\begin{aligned} \tau_{m_0} \circ \sigma_{m_1} \circ \cdots \circ \sigma_{m_{n-1}} \circ \sigma_{m_n}(\lambda) &= \tau_{m_0} \circ \sigma_{m_1} \circ \cdots \circ \sigma_{m_{n-1}}(\lambda m_n) \text{ par définition de } \sigma_{m_n} \\ &= \tau_{m_0} \circ \sigma_{m_1} \circ \cdots \circ \sigma_{m_{n-1}}(\lambda) \cdot \tau_{m_0} \circ \sigma_{m_1} \circ \cdots \circ \sigma_{m_{n-1}}(m_n) \\ &= m_{\llbracket 0, n-1 \rrbracket} \cdot \tau_{m_0} \circ \sigma_{m_1} \circ \cdots \circ \sigma_{m_{n-1}}(m_n) \text{ par hypothèse de récurrence} \\ &= m_{\llbracket 0, n-1 \rrbracket} \cdot m_n \text{ par définition des substitutions de } \mathcal{C} \\ &= m \end{aligned}$$

Soit  $x \in \mathcal{A}^{\mathbb{N}}$  : on a alors bien que  $x = \lim_{n \rightarrow \infty} \tau_{x_0} \circ \sigma_{x_1} \circ \cdots \circ \sigma_{x_n}(\lambda)$ .

2. Il existe une quantité infinie dénombrable de substitutions, mais une quantité infinie indénombrable de mots infinis.  
3. La source originale était une page du wiki du groupe de travail N. Pythas Fogg, qui ne semble plus accessible en ligne.



En dimension 1, cette propriété peut être perçue comme un défaut du modèle général des représentations S-adiques : le modèle est trop puissant, et peut décrire n'importe quel mot, en particulier des mots incalculables. Il faut reconnaître, par ailleurs, que la représentation S-adique associée à la proposition 4.2.1 n'est pas bien palpitante : un mot infini  $x$  est représenté par lui-même, avec seulement une distinction pour la première lettre. Ce résultat a conduit à l'élaboration de plusieurs limitations sur les représentations S-adiques. Par exemple, il est aujourd'hui couramment exigé que les représentations S-adiques soient *ubicroissantes*, c'est-à-dire qu'elles remplissent des conditions de croissance strictes (voir la définition 1.2.7).

Une autre exigence courante est de ne pas utiliser de lettre de travail. Dans la proposition 4.2.1, on note par exemple l'utilisation d'une lettre  $\lambda$  qui n'apparaît pas dans les mots à coder. Dans ce manuscrit, j'ai déjà utilisé un tel procédé pour prouver la proposition 2.2.32 : ma preuve revient à fournir une nouvelle caractérisation des mots minimaux ternaires dendriques. La caractérisation que je propose est néanmoins moins satisfaisante sur le plan théorique que celle de Gheeraert, Leroy et Lejeune, puisqu'elle utilise une lettre de travail supplémentaire.

La construction de Cassaigne ne s'applique toutefois directement que dans le cas de la dimension 1. Or, il est possible d'étendre les représentations S-adiques dans le cas multidimensionnel ou dans les monoïdes finiment engendrés : la construction de Cassaigne tient-elle alors dans ce cas ? Le shift plein admet-il une caractérisation S-adique ? Et quelles règles peut-on assouplir ?

### 4.2.3 Représentations S-adiques sur $\mathbb{N}^d$

Dans cette sous-section, je vais m'intéresser au cas des représentations S-adiques multidimensionnelles. La principale difficulté dans ce cadre est de poser les bonnes définitions pour les substitutions. La littérature regorge en effet de différents modèles plus ou moins génériques pour faire fonctionner les substitutions dans le cas multidimensionnel : aucun ne fonctionne de manière complètement satisfaisante. Un panel de différents modèles substitutifs est présenté dans [Fra08]. Dans cette sous-section, je me concentrerai sur le modèle d'Aubrun-Sablik [AS14]. Il s'agit d'un modèle de représentations S-adiques sur  $\mathbb{Z}^d$ , avec  $d \leq 2$  : je l'adapte ici à  $\mathbb{N}^d$ .

#### Modèle d'Aubrun-Sablik

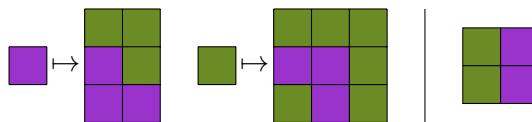
Pour définir un modèle S-adique multidimensionnel, il faut d'abord définir les substitutions multidimensionnelles ; dans le cas du modèle d'Aubrun-Sablik, on se retient au cas des substitutions rectangulaires.

Pour  $k \in \mathbb{N}^d$ , le rectangle  $R_k$  associé à  $k$  est  $\llbracket 0, k_0 - 1 \rrbracket \times \cdots \times \llbracket 0, k_{d-1} - 1 \rrbracket$ . Un motif rectangulaire sur l'alphabet  $\mathcal{A}$  est un élément de  $\mathcal{A}^{R_k}$  avec  $k \in \mathbb{N}^d$ . L'ensemble des motifs rectangulaires est  $M_{\text{rec}}(\mathcal{A}) = \bigcup_{k \in \mathbb{N}^d} \mathcal{A}^{R_k}$ . Lorsque l'alphabet est clair dans le contexte, j'écrirai simplement  $M_{\text{rec}}$ . Si  $m$  est un motif rectangulaire, on appelle grandeur de  $m$  l'uplet  $g(m) \in \mathbb{N}^d$  tel que  $m \in \mathcal{A}^{R_{g(m)}}$ .

Une substitution rectangulaire sur l'alphabet  $\mathcal{A}$  est une fonction  $\phi : \mathcal{A} \rightarrow M_{\text{rec}}(\mathcal{A})$ . Pour appliquer une substitution à un motif, on met alors une condition de compatibilité avec le motif. On se contentera ici de s'occuper des motifs rectangulaires. Soit  $m \in \mathcal{A}^R$ , avec  $R \subseteq \mathbb{N}^d$  un rectangle support du motif  $m$ . Alors une substitution rectangulaire  $\phi$  est compatible avec le motif  $m$  si pour tout  $i, j \in R$  et  $p \in \llbracket 0, d - 1 \rrbracket$ ,  $i_p = j_p \implies g(\phi(m_i))_p = g(\phi(m_j))_p$ .

Cette compatibilité s'étend naturellement aux configurations de  $\mathbb{N}^d$  : si  $x \in \mathcal{A}^{\mathbb{N}^d}$ , alors  $\phi$  est compatible avec  $x$  si pour tout  $i, j \in \mathbb{N}^d$  et  $p \in \llbracket 0, d - 1 \rrbracket$ ,  $i_p = j_p \implies g(\phi(x_i))_p = g(\phi(x_j))_p$ .

**Exemple 4.2.2.** Soit  $d = 2$ ,  $\mathcal{A} = \{\blacksquare, \blacksquare\}$  et  $\phi$  la substitution définie comme sur la figure 4.1.



**FIGURE 4.1** – À gauche, la substitution  $\phi$ ; à droite, le motif  $m$ .

Avec les notations,  $g(\phi(\blacksquare)) = (1, 2)$  et  $g(\phi(\blacklozenge)) = (2, 2)$ . La substitution  $\phi$  est bien compatible avec  $m$  : par exemple,  $i = (0, \underline{0})$  et  $j = (1, \underline{0})$  vérifient  $i_1 = j_1$ , et on a bien  $g(\phi(\blacksquare))_1 = g(\phi(\blacklozenge))_1$ . Pour vérifier la compatibilité, il faut vérifier tous les couples  $i, j \in \llbracket 0, 1 \rrbracket^2$ .

Maintenant que j'ai défini une notion de compatibilité, je vais pouvoir en déduire ce qu'est concrètement l'image d'un motif par une substitution. Pour continuer à décrire le modèle d'Aubrun-Sablík, j'ai besoin de définir un ordre (partiel) sur  $\mathbb{N}^d$  : c'est l'ordre induit par le produit des ordres naturels sur  $\mathbb{N}$ , c.-à-d.  $i < j$  si  $i_p < j_p$  pour tout  $p \in \llbracket 0, d-1 \rrbracket$  (même chose avec les inégalités larges).

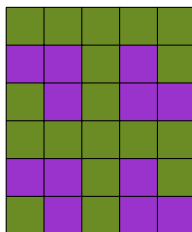
On définit comme suit l'image d'un motif rectangulaire par une substitution rectangulaire qui lui est compatible. Si  $\phi$  est une substitution rectangulaire et  $m$  est un motif rectangulaire tels que  $\phi$  est compatible avec  $m$ , alors  $\phi(m)$  est le motif rectangulaire vérifiant :

$$\forall i \leq g(m), \forall j \leq g(\phi(m_i)), \phi(m)_{(\sum_{e < i} g(\phi(m_e)))+j} = \phi(m_i)_j.$$

On étend cette définition à l'image d'une configuration  $x \in \mathbb{N}^d$  compatible avec  $\phi$ . On définit  $\phi(x)$  comme la seule configuration vérifiant :

$$\forall i \in \mathbb{N}^d, \forall j \leq g(\phi(x_i)), \phi(x)_{(\sum_{e < i} g(\phi(x_e)))+j} = \phi(x_i)_j.$$

**Exemple 4.2.3.** Je reprends les notations de l'exemple 4.2.2. Comme  $\phi$  est compatible sur  $m$ , l'image  $\phi(m)$  est définie : il s'agit du motif représenté sur la figure 4.2.



**FIGURE 4.2** – Le motif  $\phi(m)$ .

**Définition 4.2.4** (Représentation S-adique multidimensionnelle). Soit  $x \in \mathcal{A}^{\mathbb{N}^d}$  une configuration,  $(\phi_n)_{n \in \mathbb{N}}$  une suite de substitutions rectangulaires sur  $\mathcal{A}$  et  $(a_n)_{n \in \mathbb{N}}$  une suite de lettres de  $\mathcal{A}$ . On dit que  $x$  admet  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  comme représentation S-adique si pour tout  $n \in \mathbb{N}$ ,  $\phi_n$  est compatible avec le motif  $\phi_{\llbracket 0, n-1 \rrbracket}(a_n)$ , si  $\lim_{n \rightarrow \infty} \phi_{\llbracket 0, n \rrbracket}(a_n)$  existe et vaut  $x$ .

**Définition 4.2.5** (Désostituabilité infinie multidimensionnelle). Soit  $(\phi_n)_{n \in \mathbb{N}}$  une suite de substitutions. Une configuration  $x \in \mathcal{A}^{\mathbb{N}^d}$  est infiniment désostituable par  $(\phi_n)_{n \in \mathbb{N}}$  si pour tout  $n \in \mathbb{N}$ , il existe  $y_n \in \mathcal{A}^{\mathbb{N}^d}$  tel que  $\phi_{\llbracket 0, n \rrbracket}$  est compatible avec  $y_n$  et  $x = \phi_{\llbracket 0, n \rrbracket}(y_n)$ .

**Remarque 4.2.6.** Si une configuration  $x \in \mathcal{A}^{\mathbb{N}^d}$  admet une représentation S-adique  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$ , alors  $x$  est infiniment désostituable par  $(\phi_n)_{n \in \mathbb{N}}$ . La réciproque est fausse.

Le modèle substitutif utilisé par Aubrun et Sablik est plutôt contraignant de par sa restriction aux substitutions rectangulaires, et est connu dans la littérature avant le travail d'Aubrun et Sablik [Moz89]. En comparaison avec les substitutions unidimensionnelles, ce modèle peut être rapproché des substitutions uniformes. Le modèle d'Aubrun-Sablik est toutefois le premier modèle *S-adique* multidimensionnel. Dans l'article introduisant ce modèle, Aubrun et Sablik prouve la soficité des shifts S-adiques multidimensionnels [AS14].

### Non-universalité du modèle d'Aubrun-Sablik

Afin d'étudier les limites combinatoires du modèle d'Aubrun-Sablik, je me suis demandé si le shift plein disposait d'une caractérisation S-adique : la réponse est négative.

**Proposition 4.2.7.** *Soient  $d = 2$ ,  $\mathcal{A}$  un alphabet binaire, et  $\lambda \notin \mathcal{A}$ . Alors il n'existe pas d'ensemble fini de substitutions rectangulaires  $\mathcal{C}$  sur l'alphabet  $(\mathcal{A} \cup \{\lambda\})$  tel que toute configuration de  $\mathcal{A}^{\mathbb{N}^2}$  admette une représentation S-adique  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  avec  $(\phi_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ .*

*Autrement écrit, il n'existe pas de caractérisation S-adique du shift plein binaire bidimensionnel sous le modèle d'Aubrun-Sablik, même avec une lettre de travail.*

La preuve suivante est inspirée de la preuve de la proposition 3.7 de [BD14], qui prouve que le modèle S-adique sur  $\mathbb{N}$  sans lettre de travail n'est pas universel.

*Démonstration.*

On pose  $\mathcal{A} = \{0, 1\}$ . Soit  $\mathcal{C}$  un ensemble fini de substitutions sur l'alphabet  $\mathcal{A} \cup \{\lambda\}$ , et soit  $x$  une configuration vérifiant les conditions suivantes :

- $x$  est *univers*, sur l'alphabet  $\mathcal{A}$ , c'est-à-dire que pour tout motif sur l'alphabet  $\mathcal{A}$ , ce motif apparaît dans  $x$ ;
- $x$  n'a aucune ligne ni aucune colonne périodique.

Une telle configuration existe : en notant  $x_{\text{Fib}}$  le mot de Fibonacci sur  $\mathbb{N}$ , on peut déjà construire une configuration  $y \in \mathcal{A}^{\mathbb{N}^2}$  telle que  $y_{i,j} = x_{\text{Fib}, i+j}$  : toutes les lignes et les colonnes de  $y$  sont des mots infinis dans l'orbite du mot de Fibonacci, et sont donc toutes aperiodiques. Il suffit alors de forcer sur la diagonale de  $y$  tous les motifs carrés possibles sur l'alphabet  $\mathcal{A}$  : cette dernière configuration  $x$  vérifie bien les conditions demandées ci-dessus.

Supposons alors que  $x$  admette une représentation S-adique de la forme  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  avec  $(\phi_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ . Alors  $x$  est infiniment désubstituable par  $(\phi_n)_{n \in \mathbb{N}}$ , avec  $x = \phi_{\llbracket 0, n \rrbracket}(y_n)$  pour des  $y_n \in \mathcal{A}^{\mathbb{N}^2}$ . Soit  $n \in \mathbb{N}$  : alors pour tout motif  $m$  de  $x$ ,  $m$  peut se découper en images de  $\phi_{\llbracket 0, n \rrbracket}$ . En particulier, soit  $k \in \mathbb{N}^2$  tel que  $k > 4g(\phi_{\llbracket 0, n \rrbracket}(a))$  pour tout  $a \in \mathcal{A} \cup \{\lambda\}$ . Comme  $x$  est univers, tous les motifs rectangulaires de support  $R_k$  monochromes sont des motifs de  $x$ , et peuvent se découper en images de  $\phi_{\llbracket 0, n \rrbracket}$ . Comme  $k$  est assez grand, ces deux motifs monochromes (un rempli de 0, l'autre de 1) contiennent strictement au moins une image de  $\phi_{\llbracket 0, n \rrbracket}$ . Donc il existe  $a \neq b \in \mathcal{A} \cup \{\lambda\}$  tels que  $\phi_{\llbracket 0, n \rrbracket}(a)$  (resp.  $\phi_{\llbracket 0, n \rrbracket}(b)$ ) est un motif rectangulaire uniforme de seule lettre 0 (resp. de seule lettre 1). Soit  $c$  la dernière lettre de l'alphabet, c.-à-d.  $\mathcal{A} \cup \{\lambda\} = \{a, b, c\}$ . Les trois lettres définissent trois motifs rectangulaires  $\phi_{\llbracket 0, n \rrbracket}(a)$ ,  $\phi_{\llbracket 0, n \rrbracket}(b)$  et  $\phi_{\llbracket 0, n \rrbracket}(c)$ . Montrons que parmi ces trois motifs rectangulaires, ceux qui apparaissent dans  $x$  ont les mêmes grandeurs, c.-à-d. :

$$\text{Card}(\{g(\phi_{\llbracket 0, n \rrbracket}(\ell)) \mid \phi_{\llbracket 0, n \rrbracket}(\ell) \text{ constitue l'un des blocs de } x \text{ dans son découpage par } \phi_{\llbracket 0, n \rrbracket}\}) = 1.$$

Plus précisément, en notant  $\alpha_i = g(\phi_{\llbracket 0, n \rrbracket}(i))_1$  pour  $i \in \{a, b, c\}$ , je prouve que :

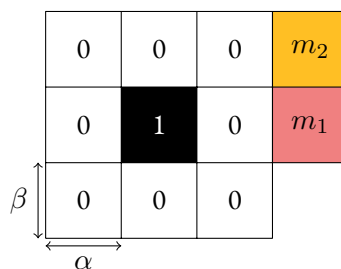
$$\text{Card}(\{\alpha_\ell \mid \phi_{\llbracket 0, n \rrbracket}(\ell) \text{ constitue l'un des blocs de } x \text{ dans son découpage par } \phi_{\llbracket 0, n \rrbracket}\}) = 1.$$

Le cas des secondes coordonnées des formes se prouve de manière similaire.

En considérant en effet  $x$  comme un pavage des trois motifs rectangulaires, on peut considérer le motif recouvrant la case  $(0, 0)$ . Ce motif est donc de la forme  $\phi_{\llbracket 0, n \rrbracket}(\ell)$  pour une certaine lettre  $\ell \in \{a, b, c\}$ .

Comme  $x$  est un pavage des trois motifs rectangulaires, on peut alors considérer le motif juste au-dessus de  $\phi_{\llbracket 0, n \rrbracket}(\ell)$  : si ce motif n'a pas de largeur  $\alpha_\ell$ , cela contredit la compatibilité de  $\phi_{\llbracket 0, n \rrbracket}$  avec  $y_n$ . Si aucun des deux autres motifs n'a la même largeur, on retrouve alors au-dessus le même motif rectangulaire  $\phi_{\llbracket 0, n \rrbracket}(\ell)$  : en itérant l'argument, on obtient alors que la première colonne de  $x$  est périodique, ce qui est absurde. Donc un autre motif  $\phi_{\llbracket 0, n \rrbracket}(l')$ , avec  $l' \neq l$ , a la même largeur ; si le dernier motif  $\phi_{\llbracket 0, n \rrbracket}(l'')$ , avec  $l'' \notin \{l, l'\}$ , n'a pas aussi cette largeur et qu'il apparaît dans le découpage de  $x$ , par un raisonnement similaire, la colonne où il apparaîtrait serait périodique. Donc tous les motifs qui apparaissent ont la même largeur.

Soit alors  $(\alpha, \beta) = f(\phi_{\llbracket 0, n \rrbracket}(\ell))$  la forme commune de tous les motifs rectangulaires. Comme  $x$  est universel, tout motif bidimensionnel apparaît dans  $x$ . Supposons que  $(\alpha, \beta) \neq (1, 1)$ . Considérons alors le motif avec les spécificités précisées dans la Figure 4.3 :



**FIGURE 4.3** – Les carrés blancs représentent le motif uniforme de 0, les noirs le motif uniforme de 1,  $m_1$  et  $m_2$  sont des motifs non uniformes distincts.

Par universalité de  $x$ , ce motif apparaît dans  $x$ . Cherchons alors à retrouver comment superposer le découpage de  $x$  par  $\phi_{\llbracket 0, n \rrbracket}$  à ce motif.

Si  $\alpha > 1$ , alors les blocs de 0 et de 1 contraignent le découpage à se superposer trivialement au motif. En particulier, les carrés blancs et noirs sont alors exactement des motifs rectangulaires  $\phi_{\llbracket 0, n \rrbracket}(a)$  et  $\phi_{\llbracket 0, n \rrbracket}(b)$  respectivement. Mais alors, par compatibilité, les motifs  $m_1$  et  $m_2$  valent tous les deux  $\phi_{\llbracket 0, n \rrbracket}(c)$ , ce qui est absurde puisqu'ils sont distincts. Donc  $\alpha = 1$ ; de manière analogue,  $\beta = 1$ .

Donc pour tout  $n \in \mathbb{N}$ ,  $\phi_{\llbracket 0, n \rrbracket}$  n'a pour images que des rectangles de dimension  $1 \times 1$ . Donc  $\mathcal{C}$  est constitué uniquement de substitutions triviales. Mais alors on ne peut pas avoir  $x = \lim_{n \rightarrow \infty} \phi_{\llbracket 0, n \rrbracket}(a_n)$ , car la suite  $(\phi_{\llbracket 0, n \rrbracket}(a_n))_{n \in \mathbb{N}}$  est une suite de motifs de taille  $1 \times 1$  et  $x$  est une configuration infinie.

La proposition 4.2.7 établit la non-universalité du modèle d'Aubrun-Sablik dans le cas restreint des configurations de  $\mathbb{N}^2$ , d'un alphabet binaire et d'une unique lettre de travail. Il est déjà simple de généraliser à tout nombre de dimension  $d \geq 2$ , en considérant les configurations contenant un plan vérifiant les propriétés demandées dans la preuve.

Toutefois, la démonstration que je propose ici ne permet a priori pas de généraliser à tout alphabet  $\mathcal{A}$ , ou à un nombre quelconque de lettres de travail. La part la plus importante de cette preuve est d'établir que toutes les images

des substitutions sont de mêmes grandeurs, et les arguments que j'utilise nécessitent qu'il y ait au plus 3 lettres au total, lettres de travail comprises. Je pense cependant que la non-universalité du modèle d'Aubrun-Sablík est indépendante de la taille de l'alphabet et du nombre de lettres de travail.

Par ailleurs, comme le principal argument de la preuve précédente est que toutes les substitutions  $\sigma_{[0,n]}$  sont uniformes, il devrait alors être possible d'adopter une démarche plus proche de la dynamique symbolique, et de prouver qu'une configuration qui admet une représentation S-adique admet une entropie nulle, comme cela a par exemple été prouvé dans le cas de  $\mathbb{Z}$  par Monteil [BD14].

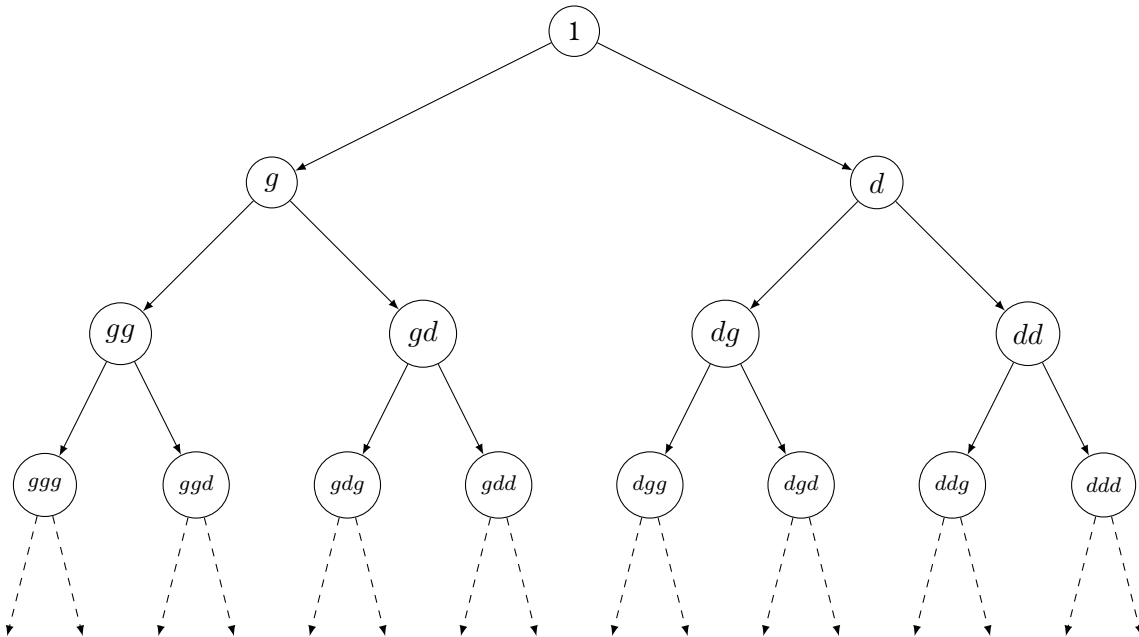
Il est clair que, dans le modèle d'Aubrun-Sablík, c'est la rectangularité des substitutions et la condition de compatibilité qui sont les arguments sous-tendant le résultat. La proposition 4.2.7 établit alors que le modèle d'Aubrun-Sablík généralise certes le modèle substitutif, mais reste contraint, même en se permettant des lettres de travail : il n'y a pas équivalence entre le cas unidimensionnel et le cas multidimensionnel.

#### 4.2.4 Représentations S-adiques sur $\mathbb{F}_2^+$

Une autre structure où il est possible de définir un modèle de représentation S-adique est le monoïde libre à deux éléments  $\mathbb{F}_2^+$ .

**Définition 4.2.8** (Monoïde  $\mathbb{F}_2^+$ ). *Le monoïde libre à deux éléments est le monoïde  $\mathbb{F}_2^+ = \langle g, d \rangle$  sans relation. L'opération interne est la concaténation, l'élément neutre est noté 1. Un élément de  $\mathbb{F}_2^+$  peut être interprété comme un mot sur l'alphabet  $\{g, d\}$ .*

Il s'agit du monoïde naturellement associé au groupe libre  $\mathbb{F}_2$ . Une visualisation naturelle de  $\mathbb{F}_2^+$  est le graphe de Cayley associé, qui correspond à un arbre binaire complet infini.



**FIGURE 4.4** – Graphe de Cayley de  $\mathbb{F}_2^+$ .

Sur cette structure, il est tout à fait possible de définir une notion de configurations et de motifs, comme cela a été fait sur les groupes [CC10].

**Définition 4.2.9** (Arbre fini, feuille, sous-arbre). *Un arbre (binaire) fini est un sous-ensemble fini  $A \subseteq \mathbb{F}_2^+$  tel que si  $p \in A$ , alors pour tout  $q \sqsubseteq_p p$ ,  $q \in A$ . L'ensemble des arbres finis est noté  $\mathcal{T}$ . Pour un arbre fini  $A \in \mathcal{T}$ ,  $p \in A$*

est une feuille de  $A$  s'il n'existe pas d'élément  $q \in A$  tel que  $p \sqsubseteq_p q$ . Le sous-arbre gauche (resp. sous-arbre droit) d'un arbre fini  $A$  est l'arbre fini  $\text{SsArb}^g(A) = \{p \in \mathbb{F}_2^+ \mid gp \in A\}$  (resp.  $\text{SsArb}^d(A) = \{p \in \mathbb{F}_2^+ \mid dp \in A\}$ ).

**Définition 4.2.10** (Motifs sur les arbres). Soit  $\mathcal{A}$  un alphabet. Une configuration de  $\mathbb{F}_2^+$  est un élément de  $\mathcal{A}^{\mathbb{F}_2^+}$ . Un motif fini est un élément de  $\bigcup_{A \in \mathcal{T}} \mathcal{A}^A$ . Si  $m \in \mathcal{A}^A$ ,  $A$  est appelé le support de  $m$ . Je noterai couramment  $m(p) = m_p$  pour  $p$  dans le support d'un motif  $m$ . L'ensemble des motifs finis sur l'alphabet  $\mathcal{A}$  est noté  $\mathcal{M}(\mathcal{A})$ . Pour un motif fini  $m$  de support  $A$ , le fils gauche (resp. le fils droit) de  $m$  est le motif fini  $\text{Fils}^g(m)$  de support  $\text{SsArb}^g(A)$  tel que pour  $p \in \text{SsArb}^g(A)$ ,  $\text{Fils}^g(m)_p = m_{gp}$  (resp. le motif fini  $\text{Fils}^d(m)$  de support  $\text{SsArb}^d(A)$  tel que pour  $p \in \text{SsArb}^d(A)$ ,  $\text{Fils}^d(m)_p = m_{dp}$ ).

On dispose aussi d'une topologie sur  $\mathcal{A}^{\mathbb{F}_2^+}$ , et on peut donc parler de limites et de continuité. Les configurations de  $\mathcal{A}^{\mathbb{F}_2^+}$  peuvent être particulièrement complexes, comme l'atteste le résultat suivant.

**Définition 4.2.11** (Sous-configuration sur les arbres). Soit  $x$  une configuration de  $\mathcal{A}^{\mathbb{F}_2^+}$ . On dit que  $y \in \mathcal{A}^{\mathbb{F}_2^+}$  est une sous-configuration de  $x$  s'il existe  $\delta \in \{g, d\}^*$  tel que  $y_k = x_{\delta k}$  pour tout  $k \in \mathbb{F}_2^+$ , et on appelle  $\delta$  le décalage de  $y$  par rapport à  $x$ .

**Lemme 4.2.12.** Il existe  $x \in \{a, b\}^{\mathbb{F}_2^+}$  telle que toutes les sous-configurations de  $x$  sont distinctes.

L'idée de la preuve suivante est due à Kreczman (communication personnelle).

*Démonstration.*

Soit  $(s^{(n)})_{n \in \mathbb{N}}$  une suite de mots infinis a périodiques d'orbites distinctes. Par exemple, on peut choisir des mots sturmiens de pentes toutes distinctes. Soit  $\text{enum} : \mathbb{N} \rightarrow \{g, d\}^* d \cup \{\varepsilon\}$  une énumération du langage régulier  $\{g, d\}^* d \cup \{\varepsilon\}$ .

On définit alors la configuration  $x$  de la manière suivante : pour  $p \in \mathbb{F}_2^+$ , on écrit  $p = p'g^k$  avec  $p' \in \{g, d\}^* d \cup \{\varepsilon\}$  et  $k \geq 0$ . On pose alors  $x_{p'g^k} = s_k^{\text{enum}(p')}$ .

Soient  $y$  et  $z$  deux sous-configurations de  $x$ , de décalages respectifs  $\delta_y \neq \delta_z \in \mathbb{F}_2^+$ . Soit  $\pi_y$  (resp.  $\pi_z$ ) le plus long élément de  $\{g, d\}^* d \cup \{\varepsilon\}$  préfixe de  $\delta_y$  (resp. de  $\delta_z$ ). Par définition, il existe donc  $j_y \geq 0$  tel que  $\delta_y = \pi_y g^{j_y}$  (resp.  $j_z \geq 0$  tel que  $\delta_z = \pi_z g^{j_z}$ ).

Alors il y a deux possibilités :

1. cas 1 :  $\pi_y = \pi_z$ . Quitte à étudier le cas symétrique, supposons  $j_y < j_z$ . Alors pour tout  $k \geq 0$ ,  $y_{g^k} = s_{j_y+k}^{\text{enum}(\pi_y)}$  et  $z_{g^k} = s_{j_z+k}^{\text{enum}(\pi_y)}$ . Comme  $s^{\text{enum}(\pi_y)}$  est a périodique, il existe donc  $k \geq 0$  tel que  $y_{g^k} \neq z_{g^k}$ .
2. cas 2 :  $\pi_y \neq \pi_z$ . Alors pour tout  $k \geq 0$ ,  $(y_{g^k})_{k \in \mathbb{N}} = S^{j_y} s^{\text{enum}(\pi_y)}$  et  $(z_{g^k})_{k \in \mathbb{N}} = S^{j_z} s^{\text{enum}(\pi_z)}$ . Comme les orbites de  $s^{\text{enum}(\pi_y)}$  et  $s^{\text{enum}(\pi_z)}$  sont distinctes, cela implique qu'il existe  $k \geq 0$  tel que  $y_{g^k} \neq z_{g^k}$ .

Donc en particulier  $y$  et  $z$  sont différentes.

## Modèle de Baraviera-Leplaideur

En 2021, Baraviera et Leplaideur [BL22] ont introduit un modèle de substitutions sur  $\mathbb{F}_2^+$  appelé les *substreeutations*. Un modèle similaire avait été établi par Damm en 1982 dans le cas des arbres finis et avec un point de vue plus informatique [Dam82]. L'objectif étant plutôt d'étudier les propriétés des substreeutations dans le cas d'arbres infinis, je reprendrai donc le modèle de Baraviera et Leplaideur. Il est à noter qu'un autre modèle encore plus récent a été développé par Bitar, Cabezas et Guillon dans le cas des groupes.

**Définition 4.2.13** (Motif étendu sur les arbres). Soit  $\mathcal{A}$  un alphabet, soit  $A$  un arbre fini et soient  $G, D \notin \mathcal{A}$ . Un motif fini étendu sur l'alphabet  $\mathcal{A}$  est un motif fini sur l'alphabet  $\mathcal{A} \cup \{G, D\}$  de support  $A$  tel que  $m_p \in \{G, D\}$ ssi  $p$  est une feuille de  $A$ . L'ensemble des motifs étendus sur l'alphabet  $\mathcal{A}$  est noté  $\mathcal{M}_e(\mathcal{A})$ .

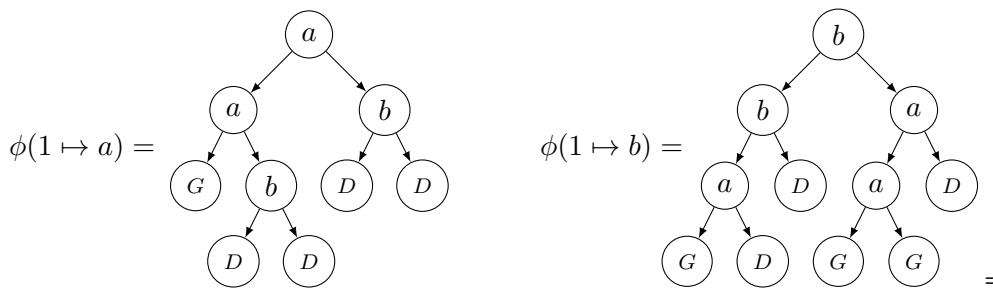
**Définition 4.2.14** (Substreetution). Une substreetution est une fonction  $\phi : (\mathcal{A} \cup \{G, D\})^{\{1\}} \rightarrow \mathcal{M}_e(\mathcal{A})^{\neq}$  telle que  $\phi(1 \mapsto G) = 1 \mapsto G$  et  $\phi(1 \mapsto D) = 1 \mapsto D$ .

On étend le domaine d'application d'une substreetution à  $\mathcal{M}_e(\mathcal{A})$  par induction. Soit  $m$  est un motif étendu, et soient  $\phi(\text{Fils}^g(m))$  et  $\phi(\text{Fils}^d(m))$  les images des fils gauche et droit de  $m$ , de supports  $A^g$  et  $A^d$  respectivement. On note  $m' = \phi(1 \mapsto m_1)$ , et  $A$  son support. Alors l'image de  $m$  par  $\phi$  est le motif fini étendu  $\phi(m)$  défini ainsi :

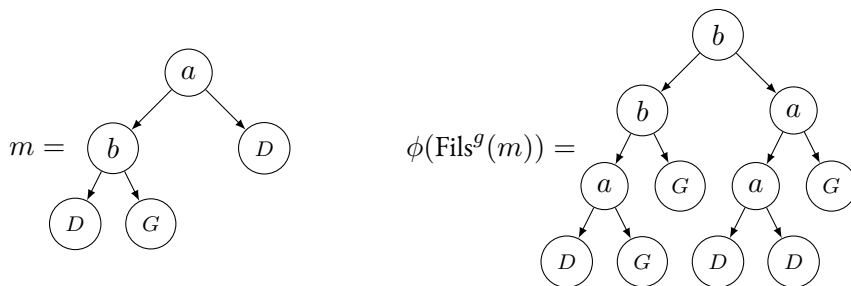
- son support est  $A \cup \{fp \mid m'_f = G \text{ et } p \in A^g\} \cup \{fp \mid m'_f = D \text{ et } p \in A^d\}$ ;
- si  $p \in A$  n'est pas une feuille de  $m$ , alors  $\phi(m)_p = m'_p$ ;
- si  $p = fp'$  avec  $f$  une feuille de  $A$ ,  $m'_f = G$  et  $p' \in A^g$ , alors  $\phi(m)_p = \phi(\text{Fils}^g(m))_{p'}$ ;
- si  $p = fp'$  avec  $f$  une feuille de  $A$ ,  $m'_f = D$  et  $p' \in A^d$ , alors  $\phi(m)_p = \phi(\text{Fils}^d(m))_{p'}$ .

On peut encore étendre le domaine d'application aux configurations de  $\mathbb{F}_2^+$ .

**Exemple 4.2.15.** Soit  $\phi$  la substreetution sur  $\{a, b\}$  définie par le schéma suivant :

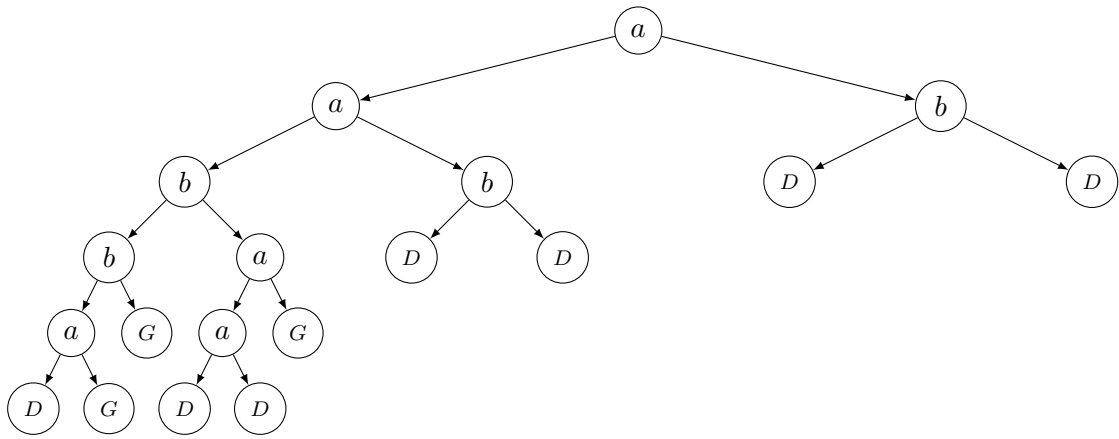


Considérons le motif  $m$  suivant. Je vais calculer son image par  $\phi$  : pour cela, je commence par déterminer l'image de son fils gauche par  $\phi$ .



Puis on calcule enfin l'image de  $m$  lui-même par  $\phi$ .

4.  $(\mathcal{A} \cup \{G, D\})^{\{1\}}$  désigne les motifs sur  $\mathcal{A} \cup \{G, D\}$  dont le support est l'arbre trivial  $\{1\}$ .



Les substrestitutions sont notamment continues pour la topologie prodiscrète sur  $\mathcal{A}^{\mathbb{F}_2^+}$ , et on peut donc définir des analogues des définitions sur  $\mathbb{N}$  au cas de  $\mathbb{F}_2^+$ , comme celles de point fixe ou de configuration purement substitutive. Il est enfin possible, dans le cadre de cette section, de définir une notion de représentation S-adique sur  $\mathbb{F}_2^+$  : il s'agit d'une définition analogue à celle sur  $\mathbb{N}$ .

**Définition 4.2.16** (Représentation S-adique sur  $\mathbb{F}_2^+$ ). Soit  $x \in \mathcal{A}^{\mathbb{F}_2^+}$  une configuration de  $\mathbb{F}_2^+$ . Si  $(\phi_n)_{n \in \mathbb{N}}$  est une suite de substrestitutions sur  $\mathcal{A}$  et  $(a_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ , on dit que  $x$  admet  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  comme une représentation S-adique si  $x = \lim_{n \rightarrow \infty} \phi_{[0, n]}(m^{a_n})$  où  $m^{a_n}$  est le motif de support  $\{1, g, d\}$  et vérifiant  $m_1^{a_n} = a_n, m_g^{a_n} = G, m_d^{a_n} = D$ .

**Définition 4.2.17** (Désostituabilité infinie sur  $\mathbb{F}_2^+$ ). Soit  $x \in \mathcal{A}^{\mathbb{F}_2^+}$  une configuration de  $\mathbb{F}_2^+$ , et soit  $(\phi_n)_{n \in \mathbb{N}}$  une suite de substrestitutions. Alors  $x$  est infiniment désostituable par  $(\phi_n)_{n \in \mathbb{N}}$  s'il existe une suite  $(y_n)_{n \in \mathbb{N}}$  de configurations de  $\mathbb{F}_2^+$  telle que  $x = \phi_{[0, n]}(y_n)$ .

Le modèle de Baraviera-Leplaidier étant récent, il ne dispose pas d'exploration combinatoire complète. Les derniers résultats concernent par exemple l'étude d'une configuration substitutive particulière appelée le Jacaranda [BL23].

### Non-universalité du modèle de Baraviera-Leplaidier

Tout comme dans le cas du modèle d'Aubrun-Sablik, il est important de savoir si le shift plein dispose d'une caractérisation S-adique : la réponse est, de nouveau, négative.

**Proposition 4.2.18.** Soit  $\mathcal{A} = \{a, b\}$  un alphabet binaire, et  $\lambda \notin \mathcal{A}$ . Alors il n'existe pas d'ensemble fini de substrestitutions  $\mathcal{C}$  sur l'alphabet  $\mathcal{A} \cup \{\lambda\}$  tel que toute configuration de  $\mathbb{F}_2^+$  admette une représentation S-adique  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  avec  $(\phi_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ .

Autrement écrit, il n'existe pas de caractérisation S-adique de  $\{a, b\}^{\mathbb{F}_2^+}$  sous le modèle de Baraviera-Leplaidier, même avec une lettre de travail.

*Démonstration.*

Soit  $x \in \mathcal{A}^{\mathbb{F}_2^+}$  une configuration qui ne contient aucune paire de sous-configurations identiques (voir le lemme 4.2.12). Supposons qu'il existe  $\mathcal{C}$  un ensemble fini de substrestitutions sur l'alphabet  $\mathcal{A} \cup \{\lambda\}$  tel que  $x$  admette une représentation S-adique de la forme  $((\phi_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}})$  avec  $(\phi_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ . Alors  $x$  est infiniment désostituable par  $(\phi_n)_{n \in \mathbb{N}}$  : il existe  $(y_n)_{n \in \mathbb{N}}$  une suite de configurations telles que  $x = \phi_{[0, n]}(y_n)$  pour tout  $n \geq 0$ . Supposons alors que, pour une lettre  $\alpha \in \mathcal{A} \cup \{\lambda\}$ , il existe  $p \neq 1$  tel que  $\phi_{[0, n]}(\alpha)_p \notin \{G, D\}$ . Alors,



en tant qu'arbre,  $\phi_{\llbracket 0, n \rrbracket}(\alpha)$  possède au moins deux nœuds internes, et donc au moins trois feuilles : il existe donc  $p_1, p_2$  deux positions telles que  $\phi_{\llbracket 0, n \rrbracket}(\alpha)_{p_1} = \phi_{\llbracket 0, n \rrbracket}(\alpha)_{p_2} \in \{G, D\}$ . Donc si  $\alpha$  apparaît dans  $y_n$ ,  $\phi_{\llbracket 0, n \rrbracket}(y_n)$  aura deux sous-configurations identiques : absurde!

Donc si  $\alpha$  apparaît dans  $y_n$ ,  $\phi_{\llbracket 0, n \rrbracket}(\alpha)$  est un arbre avec un seul nœud interne. Mais alors la suite  $\phi_{\llbracket 0, n \rrbracket}(a_n)$  est stationnaire, et ne donne que des arbres avec un seul nœud interne, et ne peut converger vers  $x$ . Absurde!

La proposition 4.2.18 est, de nouveau, généralisable. Par exemple, le nombre de lettres de travail peut être choisi arbitrairement grand, et le modèle restera non universel. De même, il est possible de généraliser le modèle de Baraviera-Leplaideur à tout monoïde libre à  $k$  éléments  $\mathbb{F}_k^+$  : la proposition précédente restera valable, et le shift plein ne disposera pas non plus d'une caractérisation S-adique.

Ici, la limite est à la fois dans le modèle et dans la complexité des configurations possibles. Une piste pour obtenir un modèle universel sur  $\mathbb{F}_2^+$  serait de reproduire la construction de Cassaigne « en parallèle » sur toutes les branches en même temps : mais par principe, il paraît difficile de le faire, car cela demanderait d'encoder à l'origine quelle branche on étendrait.

## 4.3 Problème du $X$ -domino

### 4.3.1 Introduction

Dans cette section, j'étends les questions du chapitre 2 à des conceptions plus générales de la dynamique symbolique. L'une des questions de décidabilité les plus fondamentales de la dynamique symbolique est le problème du domino, introduit par Wang [Wan61]. Le problème du domino est le problème suivant : étant donné  $X$  un sous-shift de type fini, est-ce que  $X$  est non vide ? Si, dans cette thèse, j'ai défini les sous-shifts sur  $\mathbb{N}$  ou  $\mathbb{Z}$ , ils bénéficient en fait de définitions sur tout groupe ou monoïde. Une particularité du problème du domino est en effet que sa décidabilité dépend du groupe sur lequel il est défini : en particulier, le problème est notoirement décidable sur  $\mathbb{Z}$  [New77], par un argument de lemme des tiroirs. Mais le problème devient indécidable sur  $\mathbb{Z}^2$ , d'après un résultat célèbre de Berger [Ber66]. L'une des questions les plus brûlantes en dynamique symbolique est l'étude de la frontière de décidabilité : à quelle condition sur le groupe ou le monoïde le problème du domino devient-il décidable ? Cette question fait l'objet d'une conjecture, la conjecture du domino [BS18], qui relie propriété algébrique du groupe concerné et décidabilité du problème du domino. Cette conjecture est toujours ouverte à l'heure actuelle.

Une approche de la littérature pour étudier cette frontière de la décidabilité est une approche *descendante* : on continue à étudier des configurations sur des structures multidimensionnelles, et on étudie la décidabilité de problèmes qui semblent plus simples. Par exemple, plutôt que de chercher à construire une configuration qui respecte partout des contraintes dans toutes les directions, on peut chercher à construire une configuration qui respecte partout des contraintes unidimensionnelles : il s'agit du problème du serpent [Kar03; AB23]. Dans cette section, j'aborde une approche *ascendante* : je vais plutôt chercher à complexifier le problème du domino sur  $\mathbb{N}$  en étudiant des variations arbitrairement complexes du problème. Pour cela, j'introduis le problème du  $X$ -domino.

**Définition 4.3.1** (Problème du  $X$ -domino). *Soit  $X \subseteq \mathcal{A}^{\mathbb{N}}$ . Le problème du  $X$ -domino est le suivant :*

**Donnée :** *un shift de type fini  $S$*

**Question :** *est-ce que  $X \cap S \neq \emptyset$  ?*

Par exemple, si  $X = \mathcal{A}^{\mathbb{N}}$ , le problème du  $X$ -domino est le problème du domino classique. Mais il est maintenant possible d'étendre ce problème à plusieurs classes d'ensembles, avec des résultats divers, comme cela sera fait dans

les sous-sections suivantes dans le cas où  $X$  est l'ensemble des mots sturmiens ou l'ensemble des mots sans carré.

### 4.3.2 Problème du $X$ -domino et problème d'appartenance au langage de $X$

L'objectif de cette sous-section est de faire le lien entre le problème du  $X$ -domino et le problème de l'appartenance au langage de  $X$ . Plus précisément, je parle du problème suivant :

**Définition 4.3.2** (Problème de l'appartenance au langage). *Soit  $X$  un sous-shift. Le problème de l'appartenance au langage de  $X$  est le suivant :*

**Donnée :** un mot  $m \in \mathcal{A}^*$

**Question :** est-ce que  $m \in \mathcal{L}(X)$  ?

Le problème de l'appartenance au langage de  $X$  est le problème de décision le plus naturel et le plus simple pour un sous-shift  $X$ . C'est d'ailleurs souvent à partir de ce problème que sont classifiées les principales classes de sous-shifts (voir définition 1.1.21) : un sous-shift est sofique si le problème de l'appartenance à son langage peut être décidé par un automate fini; un sous-shift est effectif si le problème de l'appartenance à son langage est décidable.

Dans le cas des sous-shifts minimaux, il est possible de rapprocher le problème du  $X$ -domino du problème de l'appartenance au langage de  $X$  :

**Proposition 4.3.3.** *Soit  $X$  un shift minimal. Alors le problème de l'appartenance au langage de  $X$  est décidable si, et seulement si, le problème du  $X$ -domino est décidable.*

*Démonstration.*

Supposons que le problème du  $X$ -domino est décidable et soit  $m$  un mot. Considérons  $S_m$  le sous-shift de type fini dont le seul facteur interdit est  $m$ . Alors  $X \cap S_m \subseteq X$  est un sous-shift de  $X$  : par minimalité de  $X$ , on a alors  $X \cap S_m = X$  ou  $X \cap S_m = \emptyset$ . Si  $X \cap S_m = \emptyset$ , c'est alors que toute configuration de  $X$  contient le facteur  $m$ , et alors  $m \in \mathcal{L}(X)$ . Sinon, par minimalité, on a donc  $X \cap S_m = X$  : donc aucune configuration de  $X$  ne contient le facteur  $m$ , et  $m \notin \mathcal{L}(X)$ . Donc le problème de l'appartenance au langage de  $X$  est décidable.

Réciproquement, supposons que le problème de l'appartenance au langage de  $X$  soit décidable. Soit  $S$  un sous-shift de type fini, avec  $F$  un ensemble fini de mots interdits caractérisant  $S$ . S'il existe  $f \in F$  tel que  $f \in \mathcal{L}(X)$ , alors il existe une configuration  $x \in X$  tel que  $f$  apparaît dans  $x$  : donc  $x \notin S$ . Donc  $X \cap S \neq X$  : par minimalité,  $X \cap S = \emptyset$ . Si pour tout  $f \in F$ ,  $f \notin \mathcal{L}(X)$ , alors aucune configuration de  $X$  ne contient de motif interdit : donc  $X \cap S = X$ .

Ce résultat général donne une piste d'étude entre la difficulté du langage d'un shift et la difficulté de son problème du domino. La preuve utilise toutefois abondamment l'hypothèse de minimalité. Cette hypothèse peut en fait être retranscrite dans le cadre de l'étude des sous-shifts minimaux d'un sous-shift :

**Proposition 4.3.4.** *Soit  $X$  un sous-shift et soit  $(X_i^m)_{i \in I}$  l'ensemble de ses sous-shifts minimaux, c.-à-d.  $X_i^m \subseteq X$  et  $X_i^m$  est minimal pour tout  $i \in I$ .*

*Alors pour  $S$  un sous-shift,  $X \cap S \neq \emptyset$  si, et seulement si,  $X_i^m \cap S \neq \emptyset$  pour un certain  $i \in I$ .*

*Démonstration.*

C'est une conséquence du fait que l'intersection de deux sous-shifts est un sous-shift.

En conséquence, on peut en déduire une décomposition du problème du  $X$ -domino en termes de sous-shifts minimaux :

**Corollaire 4.3.5.** *Soit  $X$  un sous-shift, et soit  $(X_i^m)_{i \in I}$  l'ensemble de ses sous-shifts minimaux. On a alors l'égalité suivante :*

$$\{S \text{ sous-shift de type fini} \mid X \cap S \neq \emptyset\} = \bigcup_{i \in I} \{S \text{ sous-shift de type fini} \mid X_i^m \cap S \neq \emptyset\}$$

Une approche pour étudier le problème du  $X$ -domino serait alors l'étude des sous-shifts minimaux, et obtenir ainsi une décomposition. Toutefois, cette décomposition ne donne pas plus de détails : si  $\bigcup_{i \in I} \mathcal{L}_i$  est décidable, ce n'est pas pour autant que chaque  $\mathcal{L}_i$  est décidable.

En fait, de manière générale, en abandonnant l'hypothèse de minimalité, il existe des contre-exemples de la proposition 4.3.3.

**Exemple 4.3.6.** *Soit  $\mathcal{E} \subseteq \mathbb{N}$  une famille d'entiers : on définit  $SSU(\mathcal{E}) = \{x \in \{0,1\}^{\mathbb{N}} \mid x \text{ contient au plus 3 occurrences de la lettre 1 et } \forall 0 \leq i < j, x_i = x_j = 1 \implies j - i - 1 \in \mathcal{E}\}$ . Alors  $SSU(\mathcal{E})$  est bien un sous-shift, défini par l'ensemble infini de facteurs interdits  $F_{\mathcal{E}} = \{10^i 10^j 1 \mid i, j \in \mathbb{N}\} \cup \{10^k 1 \mid k \notin \mathcal{E}\}$ . Pour tout  $\mathcal{E} \subseteq \mathbb{N}$ , le seul sous-shift minimal de  $SSU(\mathcal{E})$  est le sous-shift constant  $\{000\dots\}$ .*

*Le problème du  $SSU(\mathcal{E})$ -domino est donc décidable pour tout  $\mathcal{E} \subseteq \mathbb{N}$  : pour un sous-shift de type fini  $S$ ,  $S \cap SSU(\mathcal{E}) \neq \emptyset$  ssi  $S \cap SSU(\mathcal{E}) = \{000\dots\}$ , et cela revient à savoir si  $S$  contient la configuration constante égale à 0, ce qui est décidable.*

*En revanche, si  $\mathcal{E}$  est un sous-ensemble incalculable de  $\mathbb{N}$  (par exemple,  $\mathcal{E}$  est l'ensemble des codages en binaire des machines de Turing s'arrêtant sur le mot vide), alors le problème de l'appartenance au langage de  $SSU(\mathcal{E})$  est indécidable.*

*Il existe donc des sous-shifts  $X$  tels que le problème du  $X$ -domino est décidable, mais pas le problème de l'appartenance au langage.*

Existe-t-il un contre-exemple non minimal dans l'autre direction de la proposition 4.3.3, c.-à-d. existe-t-il  $X$  un sous-shift non minimal tel que le problème de l'appartenance au langage de  $X$  soit décidable, mais pas le problème du  $X$ -domino ? Il est remarquable qu'en étendant les questions de  $\mathbb{N}$  à  $\mathbb{N}^2$ , le shift plein est un tel contre-exemple. Mais en restant sur  $\mathbb{N}$ , le problème reste ouvert. Guillon et Hellouin ont proposé une piste par la construction de shifts de Toeplitz, mais la formalisation est difficile à établir.

### 4.3.3 Plusieurs versions du problème du $X$ -domino

#### Problème du domino sturmien

Un exemple possible de problème du  $X$ -domino est le cas où  $X$  désigne l'ensemble des mots sturmiens. Pour rappel, l'ensemble des mots sturmiens ne forme pas un sous-shift, car il n'est pas fermé : la clôture de l'ensemble des mots sturmiens est l'ensemble des mots équilibrés. Le cas des mots sturmiens justifie la généralisation du problème du  $X$ -domino au cas où  $X$  n'est pas un shift.

Les résultats du chapitre 2 donnent alors la décidabilité du problème du domino sturmien.

**Proposition 4.3.7.** *Le problème du domino sturmien est décidable.*

**Donnée :** *un sous-shift de type fini  $S$*

**Question :** *est-ce que  $S$  contient un mot sturmien ?*

*Démonstration.*

Un sous-shift de type fini  $S$ , comme tout sous-shift sofique, peut être décrit à l'aide d'un  $\omega$ -automate faible  $\mathfrak{A}_S : x \in S$  si, et seulement si,  $x \in \mathcal{L}(\mathfrak{A}_S)$ . Le problème revient alors à la détection d'un mot sturmien dans un  $\omega$ -automate faible, ce qui est un problème décidable par la proposition 2.2.28.

On remarque même que la proposition 4.3.7 peut être étendue au cas où  $S$  est un sous-shift sofique. Cela peut justifier l'idée d'un problème du  $X$ -domino sofique, où l'entrée est un sous-shift sofique.

### Problème du domino sans carré

Un autre exemple intéressant est celui du shift des mots ternaires sans carrés  $X_{\text{SC}}^3 = \{x \in \{a, b, c\}^{\mathbb{N}} \mid \forall f \in \{a, b, c\}^+, f^2 \not\leq_f x\}$ . Il s'agit, selon la tradition, de la première classe de mots étudiée par la combinatoire des mots par Thue [Thu06]. Dans le cadre de cette sous-section, l'objectif est l'étude du problème du domino sans carré. On a vu dans la sous-section précédente qu'il existe un lien entre le problème du  $X$ -domino et le problème de l'appartenance au langage de  $X$  : dans le cas du sous-shifts des mots ternaires sans carrés, le problème de l'appartenance au langage a été démontré décidable par Shelton et Soni [SS82] :

**Théorème 4.3.8.** *Le problème de l'appartenance au langage de  $X_{\text{SC}}^3$  est décidable.*

**Donnée :** *un mot fini  $m \in \{a, b, c\}^*$*

**Question :** *est-ce que  $m \in \mathcal{L}(X_{\text{SC}}^3)$  ?*

Les méthodes utilisées par Shelton et Soni reposent principalement sur l'étude d'un arbre, appelé arbre préfixe sans carré [PS15], introduit par Shelton dans un article précédent [She81a].

**Définition 4.3.9** (Arbre préfixe sans carré). *L'arbre préfixe sans carré est l'arbre infini  $\text{Arbres}_{\text{SC}}$  de racine le mot vide  $\varepsilon$ , dont les sommets sont les mots finis sans carré et tels que, pour  $u$  fini sans carré,  $u \xrightarrow{\ell} u\ell$  si  $u\ell$  est sans carré, avec  $\ell \in \{a, b, c\}$ .*

L'étude de  $\text{Arbres}_{\text{SC}}$  a fourni la première preuve de la stricte positivité de l'entropie de  $X_{\text{SC}}^3$  [She81b]. Dans le contexte de l'étude de l'arbre préfixe sans carré, le problème du domino sans carré revient à la question de savoir si, en interdisant certains motifs, on peut tout de même construire un arbre infini. Cette piste a été étudiée au cas par cas pour certains types de motifs [Thu06; Thu12], mais pas de manière systématique à ma connaissance. Mes tentatives sur la question ont été infructueuses.

Une autre piste de recherche possible concerne l'étude de la stricte positivité de l'entropie du domino sans carré. Cette approche est justifiée par la littérature riche sur l'étude de l'entropie des mots ternaires sans carrés, fournissant des outils avec lesquels attaquer la question.

**Définition 4.3.10** (Problème de la stricte positivité du domino sans carré). *Le problème de la stricte positivité du domino sans carré est le suivant :*

**Donnée :** *un sous-shift de type fini  $S$*

**Question :** *est-ce que  $X_{\text{SC}}^3 \cap S$  est d'entropie strictement positive ?*

Sur cette question, il existe deux résultats. D'abord, le problème du domino sans carré et celui de la stricte positivité du domino sans carré ne sont pas équivalents.

**Théorème 4.3.11** ([BO15]). *Il existe  $S$  un sous-shift de type fini tel que  $X_{\text{SC}}^3 \cap S$  est non vide, mais d'entropie nulle.*

Ce résultat contraint donc les éventuels algorithmes de décision à une certaine finesse. La méthode développée par Ochem repose sur la méthode substitutive, et elle fait écho à une conjecture de Cassaigne :

**Conjecture 4.3.12** (Conjecture de Cassaigne, [Och11]). *Soit  $S$  un sous-shift de type fini. Si  $X_{SC}^3 \cap S$  est non vide, alors  $X_{SC}^3 \cap S$  contient un mot substitutif.*

Si cette conjecture était vraie, on disposerait alors de l'arsenal d'outils fournis par la méthode substitutive. Par exemple, il existe toute une théorie des substitutions sans carré [Cro82; Zol15]. La conjecture reste toutefois ouverte à l'heure actuelle.

Des discussions avec Rosenfeld l'ont mené à prouver le résultat suivant non publié (communication personnelle) :

**Proposition 4.3.13.** *Le problème de la stricte positivité du domino sans carré est semi-décidable.*

La preuve utilise des arguments de comptage assez subtils et des manipulations de séries génératrices.

## 4.4 Conclusion

Dans ce chapitre, j'ai décrit diverses pistes de recherche que j'ai partiellement explorées au cours de ma thèse. Par exemple, j'ai commencé à explorer les possibilités combinatoires de deux extensions du modèle S-adique à  $\mathbb{N}^d$  et à  $\mathbb{F}_2^+$  respectivement; j'ai aussi amorcé l'étude du problème du  $X$ -domino, qui relie dynamique symbolique et combinatoire des mots.

Dans la première partie, les résultats que j'ai établi sur la non-universalité des modèles d'Aubrun-Sablik et de Baraviera-Leplaideur prouvent, surtout, que ces deux modèles sont assez rigides dans leurs constructions respectives, en particulier le modèle d'Aubrun-Sablik. La difficulté réside en fait dans une généralisation intéressante des substitutions dans des structures plus complexes que  $\mathbb{N}$  ou  $\mathbb{Z}$  : c'est un travail encore en cours dans le monde de la dynamique symbolique. Une question serait, si on réussit à trouver un modèle intéressant, s'il est possible d'établir une généralisation des mots sturmiens à ces structures plus complexes par une généralisation de leur représentation S-adique. Aussi, avec ces modèles combinatoires, est-il possible de déduire des propriétés dynamiques des systèmes étudiés? Il s'agit a priori partiellement de la démarche entamée par Baraviera et Leplaideur, qui cherchent à étudier des questions de physique mathématique sur les substitutions.

Dans la seconde partie, l'étude générale du problème du  $X$ -domino a laissé une grande part à la minimalité. J'ai proposé un lien partiel entre les problèmes du  $X$ -domino d'un sous-shift et de ses sous-shifts minimaux; est-il possible d'établir une hiérarchie plus claire? Sous quelles hypothèses peut-on comparer les difficultés des problèmes du  $X$ -domino d'un sous-shift et de ses sous-shifts minimaux? Sur le problème du domino sans carré, mon travail n'a pas porté de fruit, et a principalement consisté à essayer de comprendre la littérature. La conjecture de Cassaigne est intéressante, car elle permet d'entrevoir des solutions algorithmiques si elle devait être avérée : mais même dans ce cas, cela ne fournirait que la semi-décidabilité du problème. Si la conjecture de Cassaigne était vraie, cela entraînerait-il la décidabilité du domino sans carré? L'étude de l'arbre préfixe sans carré et les outils qui ont été développés autour bénéficient d'une littérature riche, mais très dispersée : le problème du domino sans carré pourrait avoir déjà été prouvé décidable. Il me paraît, en principe, improbable que la contrainte d'absence de carré puisse entraîner l'indécidabilité du problème, car les preuves d'indécidabilité en dynamique symbolique procèdent surtout par réduction du problème de l'arrêt des machines de Turing : encoder des machines de Turing via des sous-shifts de type fini en dimension 1 en utilisant l'absence de carrés me semble déraisonnable.



# Bibliographie

- [Abe03] Ali ABERKANE. « Words whose complexity satisfies  $\lim \frac{p(n)}{n} = 1$  ». In : t. 307. 1. Words. 2003, p. 31-46. DOI : 10.1016/S0304-3975(03)00091-4. URL : [https://doi.org/10.1016/S0304-3975\(03\)00091-4](https://doi.org/10.1016/S0304-3975(03)00091-4).
- [AB07] Boris ADAMCZEWSKI et Yann BUGEAUD. « On the complexity of algebraic numbers. I. Expansions in integer bases ». In : *Ann. of Math. (2)* 165.2 (2007), p. 547-565. ISSN : 0003-486X,1939-8980. DOI : 10.4007/annals.2007.165.547. URL : <https://doi.org/10.4007/annals.2007.165.547>.
- [Ale96] Pascal ALESSANDRI. « Codage de rotations et suites de basse complexité ». Thèse de doct. Université d'Aix-Marseille II, 1996.
- [AS03] Jean-Paul ALLOUCHE et Jeffrey SHALLIT. *Automatic sequences*. Theory, applications, generalizations. Cambridge University Press, Cambridge, 2003, p. xvi+571. ISBN : 0-521-82332-3. DOI : 10.1017/CB09780511546563. URL : <https://doi.org/10.1017/CB09780511546563>.
- [AV22] Mélodie ANDRIEU et Léo VIVION. « Imbalances in Hypercubic Billiard Words ». In : *18th Mons Theoretical Computer Science Days*. 2022.
- [AR91] Pierre ARNOUX et Gérard RAUZY. « Représentation géométrique de suites de complexité  $2n + 1$  ». In : *Bull. Soc. Math. France* 119.2 (1991), p. 199-215. ISSN : 0037-9484,2102-622X. URL : [http://www.numdam.org/item?id=BSMF\\_1991\\_\\_119\\_2\\_199\\_0](http://www.numdam.org/item?id=BSMF_1991__119_2_199_0).
- [Arn+94] Pierre ARNOUX et al. « Complexity of sequences defined by billiard in the cube ». In : *Bull. Soc. Math. France* 122.1 (1994), p. 1-12. ISSN : 0037-9484,2102-622X. URL : [http://www.numdam.org/item?id=BSMF\\_1994\\_\\_122\\_1\\_1\\_0](http://www.numdam.org/item?id=BSMF_1994__122_1_1_0).
- [ABH23] Nathalie AUBRUN, Nicol'as BITAR et Sacha HURIOT-TATTEGRAIN. « Strongly aperiodic SFTs on generalized Baumslag–Solitar groups ». In : *Ergodic Theory and Dynamical Systems* (2023). DOI : 10.1017/etds.2023.44.
- [AB23] Nathalie AUBRUN et Nicolás BITAR. « Domino snake problems on groups ». In : *Fundamentals of computation theory*. T. 14292. Lecture Notes in Comput. Sci. Springer, Cham, [2023] ©2023, p. 46-59. ISBN : 978-3-031-43586-7; 978-3-031-43587-4. DOI : 10.1007/978-3-031-43587-4\_4. URL : [https://doi.org/10.1007/978-3-031-43587-4\\_4](https://doi.org/10.1007/978-3-031-43587-4_4).
- [AES20] Nathalie AUBRUN, Solène ESNAY et Mathieu SABLİK. « Domino problem under horizontal constraints ». In : *37th International Symposium on Theoretical Aspects of Computer Science*. T. 154. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2020, Art. No. 26, 15. ISBN : 978-3-95977-140-5.
- [AS14] Nathalie AUBRUN et Mathieu SABLİK. « Multidimensional effective S-adic subshifts are sofic ». In : *Unif. Distrib. Theory* 9.2 (2014), p. 7-29. ISSN : 1336-913X,2309-5377.

- [BO15] Golnaz BADKOBEB et Pascal OCHEM. « Characterization of some binary words with few squares ». In : *Theoret. Comput. Sci.* 588 (2015), p. 73-80. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2015.03.044. URL : <https://doi.org/10.1016/j.tcs.2015.03.044>.
- [BDJ08] Alexis BALLIER, Bruno DURAND et Emmanuel JEANDEL. « Structural aspects of tilings ». In : *STACS 2008: 25th International Symposium on Theoretical Aspects of Computer Science*. T. 1. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2008, p. 61-72. ISBN : 978-3-939897-06-4.
- [BJ08] Alexis BALLIER et Emmanuel JEANDEL. « Tilings and model theory ». In : *JAC 2008*. Sous la dir. de Bruno DURAND. Uzès, France, avr. 2008, p. 29-39.
- [BS18] Alexis BALLIER et Maya STEIN. « The domino problem on groups of polynomial growth ». In : *Groups Geom. Dyn.* 12.1 (2018), p. 93-105. ISSN : 1661-7207,1661-7215. DOI : 10.4171/GGD/439. URL : <https://doi.org/10.4171/GGD/439>.
- [BL22] Alexandre BARAVIERA et Renaud LEPLAIDEUR. *Substreetutions and more on trees*. 2022. arXiv : 2112.05242 [math.DS].
- [BL23] Alexandre BARAVIERA et Renaud LEPLAIDEUR. *The Jacaranda tree is strongly aperiodic and has zero entropy*. 2023. arXiv : 2304.08039 [math.DS].
- [BLS21] Sebastián BARBIERI, Sébastien LABBÉ et Štěpán STAROSTA. « A characterization of Sturmian sequences by indistinguishable asymptotic pairs ». In : *European J. Combin.* 95 (2021), Paper No. 103318, 22. ISSN : 0195-6698,1095-9971. DOI : 10.1016/j.ejc.2021.103318. URL : <https://doi.org/10.1016/j.ejc.2021.103318>.
- [Bar88] Michael BARNESLEY. *Fractals everywhere*. Academic Press, Inc., Boston, MA, 1988, p. xii+396. ISBN : 0-12-079062-9.
- [BDS12] Frédérique BASSINO, Julien DAVID et Andrea SPORTIELLO. « Asymptotic enumeration of minimal automata ». In : *29th International Symposium on Theoretical Aspects of Computer Science*. T. 14. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2012, p. 88-99. ISBN : 978-3-939897-35-4.
- [BPR23] Marie-Pierre BÉAL, Dominique PERRIN et Antonio RESTIVO. « Recognizability of morphisms ». In : *Ergodic Theory Dynam. Systems* 43.11 (2023), p. 3578-3602. ISSN : 0143-3857,1469-4417. DOI : 10.1017/etds.2022.109. URL : <https://doi.org/10.1017/etds.2022.109>.
- [BH23] Pierre BÉAUR et Benjamin HELLOUIN DE MENIBUS. « Sturmian and infinitely desubstitutable words accepted by an  $\omega$ -automaton ». In : *Combinatorics on words*. T. 13899. Lecture Notes in Comput. Sci. Springer, Cham, [2023] ©2023, p. 104-116. ISBN : 978-3-031-33179-4; 978-3-031-33180-0. DOI : 10.1007/978-3-031-33180-0\_8. URL : [https://doi.org/10.1007/978-3-031-33180-0\\_8](https://doi.org/10.1007/978-3-031-33180-0_8).
- [BH07] Nicolas BEDARIDE et Pascal HUBERT. « Billiard complexity in the hypercube ». In : *Ann. Inst. Fourier (Grenoble)* 57.3 (2007), p. 719-738. ISSN : 0373-0956,1777-5310. URL : [http://aif.cedram.org/item?id=AIF\\_2007\\_\\_57\\_3\\_719\\_0](http://aif.cedram.org/item?id=AIF_2007__57_3_719_0).
- [BSS24] Jason BELL, Chris SCHULZ et Jeffrey SHALLIT. *Consecutive Power Occurrences in Sturmian Words*. 2024. arXiv : 2402.09597 [math.CO].
- [Ber66] Robert BERGER. « The undecidability of the domino problem ». In : *Mem. Amer. Math. Soc.* 66 (1966), p. 72. ISSN : 0065-9266,1947-6221.



- [Ber90] Jean BERSTEL. « Tracé de droites, fractions continues et morphismes itérés ». In : *Mots*. Lang. Raison. Calc. Hermès, Paris, 1990, p. 298-309. ISBN : 2-86601-206-2.
- [Ber96] Jean BERSTEL. « Recent results in Sturmian words ». In : *Developments in language theory, II (Magdeburg, 1995)*. World Sci. Publ., River Edge, NJ, 1996, p. 13-24. ISBN : 981-02-2682-9.
- [BS94] Jean BERSTEL et Patrice SÉEBOLD. « Morphismes de Sturm ». In : t. 1. 2. Journées Montoises (Mons, 1992). 1994, p. 175-189. URL : <http://projecteuclid.org/euclid.bbms/1103408546>.
- [Ber09] Valérie BERTHÉ. « Discrete geometry and symbolic dynamics ». In : *Complex analysis and digital geometry*. T. 86. Acta Univ. Upsaliensis Skr. Uppsala Univ. C Organ. Hist. Uppsala Universitet, Uppsala, 2009, p. 81-110. ISBN : 978-91-554-7672-4.
- [BBY22] Valérie BERTHÉ, Paulina Cecchi BERNALES et Reem YASSAWI. *Coboundaries and eigenvalues of finitary S-adic systems*. 2022. arXiv : 2202.07270 [math.DS].
- [BD14] Valérie BERTHÉ et Vincent DELECROIX. « Beyond substitutive dynamical systems : S-adic expansions ». In : *Numeration and substitution 2012*. T. B46. RIMS Kôkyûroku Bessatsu. Res. Inst. Math. Sci. (RIMS), Kyoto, 2014, p. 81-123.
- [BFS16] Valérie BERTHÉ, Thomas FERNIQUE et Mathieu SABLİK. « Effective S-adic symbolic dynamical systems ». In : *Pursuit of the universal*. T. 9709. Lecture Notes in Comput. Sci. Springer, [Cham], 2016, p. 13-23. ISBN : 978-3-319-40189-8; 978-3-319-40188-1. DOI : 10.1007/978-3-319-40189-8\_2. URL : [https://doi.org/10.1007/978-3-319-40189-8\\_2](https://doi.org/10.1007/978-3-319-40189-8_2).
- [BG23] Valérie BERTHÉ et Herman GOULET-OUELLET. « On substitutions preserving their return sets ». In : *Combinatorics on words*. T. 13899. Lecture Notes in Comput. Sci. Springer, Cham, [2023] ©2023, p. 77-90. ISBN : 978-3-031-33179-4; 978-3-031-33180-0. DOI : 10.1007/978-3-031-33180-0\_6. URL : [https://doi.org/10.1007/978-3-031-33180-0\\_6](https://doi.org/10.1007/978-3-031-33180-0_6).
- [BHZ06] Valérie BERTHÉ, Charles HOLTON et Luca Q. ZAMBONI. « Initial powers of Sturmian sequences ». In : *Acta Arith.* 122.4 (2006), p. 315-347. ISSN : 0065-1036,1730-6264. DOI : 10.4064/aa122-4-1. URL : <https://doi.org/10.4064/aa122-4-1>.
- [BR10] Valérie BERTHÉ et Michel RIGO, éd. *Combinatorics, automata and number theory*. T. 135. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2010, p. xx+615. ISBN : 978-0-521-51597-9. DOI : 10.1017/CB09780511777653. URL : <https://doi.org/10.1017/CB09780511777653>.
- [Ber+18] Valérie BERTHÉ et al. « Rigidity and substitutive dendric words ». In : *Internat. J. Found. Comput. Sci.* 29.5 (2018), p. 705-720. ISSN : 0129-0541,1793-6373. DOI : 10.1142/S0129054118420017. URL : <https://doi.org/10.1142/S0129054118420017>.
- [Bir12] G. David BIRKHOFF. « Quelques théorèmes sur le mouvement des systèmes dynamiques ». In : *Bull. Soc. Math. France* 40 (1912), p. 305-323. ISSN : 0037-9484. URL : [http://www.numdam.org/item?id=BSMF\\_1912\\_\\_40\\_\\_305\\_0](http://www.numdam.org/item?id=BSMF_1912__40__305_0).
- [BR06] Jean-Pierre BOREL et Christophe REUTENAUER. « On Christoffel classes ». In : *Theor. Inform. Appl.* 40.1 (2006), p. 15-27. ISSN : 0988-3754,1290-385X. DOI : 10.1051/ita:2005038. URL : <https://doi.org/10.1051/ita:2005038>.
- [Bos85] Michael BOSHERNITZAN. « A unique ergodicity of minimal symbolic flows with linear block growth ». In : *J. Analyse Math.* 44 (1984/85), p. 77-96. ISSN : 0021-7670,1565-8538. DOI : 10.1007/BF02790191. URL : <https://doi.org/10.1007/BF02790191>.

- [BF84] Michael BOSHERNITZAN et Aviezri S. FRAENKEL. « A linear algorithm for nonhomogeneous spectra of numbers ». In : *J. Algorithms* 5.2 (1984), p. 187-198. ISSN : 0196-6774. DOI : 10.1016/0196-6774(84)90026-9. URL : [https://doi.org/10.1016/0196-6774\(84\)90026-9](https://doi.org/10.1016/0196-6774(84)90026-9).
- [Bos+17] Alin BOSTAN et al. *Algorithmes Efficaces en Calcul Formel*. 686 pages. Imprimé par CreateSpace. Aussi disponible en version électronique. Palaiseau : Frédéric Chyzak (auto-édit.), sept. 2017. ISBN : 979-10-699-0947-2. URL : <https://hal.archives-ouvertes.fr/AECF/>.
- [Bru81] Nicolaas G. de BRUIJN. « Sequences of zeros and ones generated by special production rules ». In : *Nederl. Akad. Wetensch. Indag. Math.* 43.1 (1981), p. 27-37. ISSN : 0019-3577.
- [CH22] Antonin CALLARD et Benjamin HELLOUIN DE MENIBUS. « The aperiodic Domino problem in higher dimension ». In : *39th International Symposium on Theoretical Aspects of Computer Science*. T. 219. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022, Art. No. 19, 15. ISBN : 978-3-95977-222-8.
- [Car08] Olivier CARTON. *Langages formels, calculabilité et complexité*. Vuibert, 2008. ISBN : 9782711720774.
- [CT02] Olivier CARTON et Wolfgang THOMAS. « The monadic theory of morphic infinite words and generalizations ». In : *Inform. and Comput.* 176.1 (2002), p. 51-65. ISSN : 0890-5401,1090-2651. DOI : 10.1006/inco.2001.3139. URL : <https://doi.org/10.1006/inco.2001.3139>.
- [Cas97a] Julien CASSAIGNE. « Complexité et facteurs spéciaux ». In : t. 4. 1. Journées Montoises (Mons, 1994). 1997, p. 67-88. URL : <http://projecteuclid.org/euclid.bbms/1105730624>.
- [Cas97b] Julien CASSAIGNE. « Sequences with grouped factors ». In : *Developments in Language Theory*. 1997, p. 211-222.
- [CKT10] Julien CASSAIGNE, Idrissa KABORÉ et Théodore TAPSOBA. « On a new notion of complexity on infinite words ». In : *Acta Univ. Sapientiae Math.* 2.2 (2010), p. 127-136. ISSN : 1844-6094,2066-7752.
- [CC10] Tullio CECCHERINI-SILBERSTEIN et Michel COORNAERT. *Cellular automata and groups*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2010, p. xx+439. ISBN : 978-3-642-14033-4. DOI : 10.1007/978-3-642-14034-1. URL : <https://doi.org/10.1007/978-3-642-14034-1>.
- [Cha33] D. G. CHAMPERNOWNE. « The Construction of Decimals Normal in the Scale of Ten ». In : *J. London Math. Soc.* 8.4 (1933), p. 254-260. ISSN : 0024-6107,1469-7750. DOI : 10.1112/jlms/s1-8.4.254. URL : <https://doi.org/10.1112/jlms/s1-8.4.254>.
- [Cha23] Émilie CHARLIER. « Alternate base numeration systems ». In : *Combinatorics on words*. T. 13899. Lecture Notes in Comput. Sci. Springer, Cham, [2023] ©2023, p. 14-34. ISBN : 978-3-031-33179-4; 978-3-031-33180-0. DOI : 10.1007/978-3-031-33180-0\_2. URL : [https://doi.org/10.1007/978-3-031-33180-0\\_2](https://doi.org/10.1007/978-3-031-33180-0_2).
- [CLR16] Émilie CHARLIER, Julien LEROY et Michel RIGO. « Asymptotic properties of free monoid morphisms ». In : *Linear Algebra Appl.* 500 (2016), p. 119-148. ISSN : 0024-3795,1873-1856. DOI : 10.1016/j.laa.2016.02.030. URL : <https://doi.org/10.1016/j.laa.2016.02.030>.
- [CB08] Emilie CHARRIER et Lilian BUZER. « An Efficient and Quasi Linear Worst-Case Time Algorithm for Digital Plane Recognition ». In : *Discrete Geometry for Computer Imagery*. Sous la dir. de David COEURJOLLY et al. Berlin, Heidelberg : Springer Berlin Heidelberg, 2008, p. 346-357.

- [Chr+80] Gilles CHRISTOL et al. « Suites algébriques, automates et substitutions ». In : *Bull. Soc. Math. France* 108.4 (1980), p. 401-419. ISSN : 0037-9484. URL : [http://www.numdam.org/item?id=BSMF\\_1980\\_\\_108\\_\\_401\\_0](http://www.numdam.org/item?id=BSMF_1980__108__401_0).
- [CS24] Benoit CLOITRE et Jeffrey SHALLIT. *Some Fibonacci-Related Sequences*. 2024. arXiv : 2312.11706 [math.CO].
- [Cob68] Alan COBHAM. « On the Hartmanis-Stearns problem for a class of tag machines ». In : *9th Annual Symposium on Switching and Automata Theory (swat 1968)*. 1968, p. 51-60. DOI : 10.1109/SWAT.1968.20.
- [Cov75] Ethan M. COVEN. « Sequences with minimal block growth. II ». In : *Math. Systems Theory* 8.4 (1974/75), p. 376-382. ISSN : 0025-5661. DOI : 10.1007/BF01780584. URL : <https://doi.org/10.1007/BF01780584>.
- [CH73] Ethan M. COVEN et Gustav A. HEDLUND. « Sequences with minimal block growth ». In : *Math. Systems Theory* 7 (1973), p. 138-153. ISSN : 0025-5661. DOI : 10.1007/BF01762232. URL : <https://doi.org/10.1007/BF01762232>.
- [Cro82] Max CROCHEMORE. « Sharp characterizations of squarefree morphisms ». In : *Theoret. Comput. Sci.* 18.2 (1982), p. 221-226. ISSN : 0304-3975,1879-2294. DOI : 10.1016/0304-3975(82)90023-8. URL : [https://doi.org/10.1016/0304-3975\(82\)90023-8](https://doi.org/10.1016/0304-3975(82)90023-8).
- [Dam82] Werner DAMM. « The IO- and OI-hierarchies ». In : *Theoret. Comput. Sci.* 20.2 (1982), p. 95-207. ISSN : 0304-3975,1879-2294. DOI : 10.1016/0304-3975(82)90009-3. URL : [https://doi.org/10.1016/0304-3975\(82\)90009-3](https://doi.org/10.1016/0304-3975(82)90009-3).
- [Deb95] Isabelle DEBLED-RENNESON. « Étude et reconnaissance des droites et plans discrets ». Thèse de doct. Strasbourg : Université Louis Pasteur, 1995.
- [DL24] Isabelle DEBLED-RENNESON et Bastien LABOUREIX. « Recognition of arithmetic line segments and hyperplanes using the Stern-Brocot tree ». In : *Discrete Geometry and Mathematical Morphology* (2024). Sous la dir. d'Andrea FROSINI.
- [Dej72] Françoise DEJEAN. « Sur un théorème de Thue ». In : *J. Combinatorial Theory Ser. A* 13 (1972), p. 90-99. ISSN : 0097-3165. DOI : 10.1016/0097-3165(72)90011-8. URL : [https://doi.org/10.1016/0097-3165\(72\)90011-8](https://doi.org/10.1016/0097-3165(72)90011-8).
- [Did99] Gilles DIDIER. « Caractérisation des  $N$ -écritures et application à l'étude des suites de complexité ultimement  $n + c^{\text{ste}}$  ». In : *Theoret. Comput. Sci.* 215.1-2 (1999), p. 31-49. ISSN : 0304-3975,1879-2294. DOI : 10.1016/S0304-3975(97)00122-9. URL : [https://doi.org/10.1016/S0304-3975\(97\)00122-9](https://doi.org/10.1016/S0304-3975(97)00122-9).
- [Dol23] Francesco DOLCE. « String attractors for factors of the Thue-Morse word ». In : *Combinatorics on words*. T. 13899. Lecture Notes in Comput. Sci. Springer, Cham, [2023] ©2023, p. 117-129. ISBN : 978-3-031-33179-4; 978-3-031-33180-0. DOI : 10.1007/978-3-031-33180-0\_9. URL : [https://doi.org/10.1007/978-3-031-33180-0\\_9](https://doi.org/10.1007/978-3-031-33180-0_9).
- [DFV03] Bruno DURAND, Enrico FORMENTI et Georges VAROUCAS. « On undecidability of equicontinuity classification for cellular automata ». In : *Discrete models for complex systems, DMCS '03 (Lyon)*. T. AB. Discrete Math. Theor. Comput. Sci. Proc. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2003, p. 117-127.

- [Dur98] Fabien DURAND. « A characterization of substitutive sequences using return words ». In : *Discrete Math.* 179.1-3 (1998), p. 89-101. ISSN : 0012-365X,1872-681X. DOI : 10.1016/S0012-365X(97)00029-0. URL : [https://doi.org/10.1016/S0012-365X\(97\)00029-0](https://doi.org/10.1016/S0012-365X(97)00029-0).
- [Dur12] Fabien DURAND. « HD0L  $\omega$ -equivalence and periodicity problems in the primitive case ». In : *Unif. Distrib. Theory* 7.1 (2012), p. 199-215. ISSN : 1336-913X,2309-5377.
- [Dur13a] Fabien DURAND. « Decidability of the HD0L ultimate periodicity problem ». In : *RAIRO Theor. Inform. Appl.* 47.2 (2013), p. 201-214. ISSN : 0988-3754,1290-385X. DOI : 10.1051/ita/2013035. URL : <https://doi.org/10.1051/ita/2013035>.
- [Dur13b] Fabien DURAND. « Decidability of uniform recurrence of morphic sequences ». In : *Internat. J. Found. Comput. Sci.* 24.1 (2013), p. 123-146. ISSN : 0129-0541,1793-6373. DOI : 10.1142/S0129054113500032. URL : <https://doi.org/10.1142/S0129054113500032>.
- [Dvo24] L'ubomíra DVOŘÁKOVÁ. « String attractors of episturmian sequences ». In : *Theoret. Comput. Sci.* 986 (2024), p. 114341. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2023.114341. URL : <https://doi.org/10.1016/j.tcs.2023.114341>.
- [Esp23a] Bastián ESPINOZA. « Symbolic factors of S-adic subshifts of finite alphabet rank ». In : *Ergodic Theory Dynam. Systems* 43.5 (2023), p. 1511-1547. ISSN : 0143-3857,1469-4417. DOI : 10.1017/etds.2022.21. URL : <https://doi.org/10.1017/etds.2022.21>.
- [Esp23b] Bastián ESPINOZA. *The structure of low complexity subshifts*. 2023. arXiv : 2305.03096 [math.DS].
- [Fer96] Sébastien FERENCZI. « Rank and symbolic complexity ». In : *Ergodic Theory Dynam. Systems* 16.4 (1996), p. 663-682. ISSN : 0143-3857,1469-4417. DOI : 10.1017/S0143385700009032. URL : <https://doi.org/10.1017/S0143385700009032>.
- [FM97] Sébastien FERENCZI et Christian MAUDUIT. « Transcendence of numbers with a low complexity expansion ». In : *J. Number Theory* 67.2 (1997), p. 146-161. ISSN : 0022-314X,1096-1658. DOI : 10.1006/jnth.1997.2175. URL : <https://doi.org/10.1006/jnth.1997.2175>.
- [FS08] Henning FERNAU et Ralf STIEBE. « Blind counter automata on  $\omega$ -words ». In : *Fund. Inform.* 83.1-2 (2008), p. 51-64. ISSN : 0169-2968,1875-8681.
- [FW65] Nathan. J. FINE et Herbert. S. WILF. « Uniqueness theorems for periodic functions ». In : *Proc. Amer. Math. Soc.* 16 (1965), p. 109-114. ISSN : 0002-9939,1088-6826. DOI : 10.2307/2034009. URL : <https://doi.org/10.2307/2034009>.
- [Fog02] N. Pytheas FOGG. *Substitutions in dynamics, arithmetics and combinatorics*. Sous la dir. de V. BERTHÉ et al. T. 1794. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2002, p. xviii+402. ISBN : 3-540-44141-7. DOI : 10.1007/b13861. URL : <https://doi.org/10.1007/b13861>.
- [Fra08] Natalie Priebe FRANK. « A primer of substitution tilings of the Euclidean plane ». In : *Expo. Math.* 26.4 (2008), p. 295-326. ISSN : 0723-0869,1878-0792. DOI : 10.1016/j.exmath.2008.02.001. URL : <https://doi.org/10.1016/j.exmath.2008.02.001>.
- [Fri01] Anna E. FRID. « On factor graphs of DOL words [MR1760731 (2001d:05156)] ». In : t. 114. 1-3. Discrete analysis and operations research. 2001, p. 121-130. DOI : 10.1016/S0166-218X(00)00365-6. URL : [https://doi.org/10.1016/S0166-218X\(00\)00365-6](https://doi.org/10.1016/S0166-218X(00)00365-6).
- [Gar89] Martin GARDNER. *Penrose tiles to trapdoor ciphers*. Recreational Mathematics. . . and the return of Dr. Matrix. W. H. Freeman et Company, New York, 1989, p. x+311. ISBN : 0-7167-1986-X; 0-7167-1987-8.

- [Gay23] Léo GAYRAL. « Complexité et robustesse des pavages avec perturbations aléatoires ». Thèse de doct. Université de Toulouse, 2023.
- [Gea+16] Cody GEARY et al. « Programming biomolecules that fold greedily during transcription ». In : *41st International Symposium on Mathematical Foundations of Computer Science*. T. 58. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, Art. No. 43, 14. ISBN : 978-3-95977-016-3.
- [GDZ05] Yan GERARD, Isabelle DEBLED-RENNESON et Paul ZIMMERMANN. « An elementary digital plane recognition algorithm ». In : *Discrete Applied Mathematics* 151.1 (2005). IWCIA 2003-Ninth International Workshop on Combinatorial Image Analysis, p. 169-183. ISSN : 0166-218X. DOI : <https://doi.org/10.1016/j.dam.2005.02.026>.
- [Ghe23a] France GHEERAERT. « A study of dendricity through the lens of morphisms ». Thèse de doct. Université de Liège, 2023.
- [Ghe23b] France GHEERAERT. « Some properties of morphic images of (eventually) dendric words ». In : *Monatsb. Math.* 202.2 (2023), p. 335-351. ISSN : 0026-9255,1436-5081. DOI : 10.1007/s00605-023-01877-4. URL : <https://doi.org/10.1007/s00605-023-01877-4>.
- [GLL22] France GHEERAERT, Marie LEJEUNE et Julien LEROY. «  $S$ -adic characterization of minimal ternary dendric shifts ». In : *Ergodic Theory Dynam. Systems* 42.11 (2022), p. 3393-3432. ISSN : 0143-3857,1469-4417. DOI : 10.1017/etds.2021.84. URL : <https://doi.org/10.1017/etds.2021.84>.
- [GL22] France GHEERAERT et Julien LEROY. *S-adic characterization of minimal dendric shifts*. 2022. arXiv : 2206.00333 [math.DS].
- [GS89] Branko GRÜNBAUM et G. C. SHEPHARD. *Tilings and patterns*. A Series of Books in the Mathematical Sciences. An introduction. W. H. Freeman et Company, New York, 1989, p. xii+446. ISBN : 0-7167-1998-3.
- [Hei01] Alex HEINIS. « Arithmetics and Combinatorics of Words of Low Complexity ». Thèse de doct. Universiteit Leiden, 2001.
- [HLN23] Benjamin HELLOUIN DE MENIBUS, Victor H. LUTFALLA et Camille NOÛS. « The Domino problem is undecidable on every rhombus subshift ». In : *Developments in language theory*. T. 13911. Lecture Notes in Comput. Sci. Springer, Cham, [2023] ©2023, p. 100-112. ISBN : 978-3-031-33263-0; 978-3-031-33264-7. DOI : 10.1007/978-3-031-33264-7\_9. URL : [https://doi.org/10.1007/978-3-031-33264-7\\_9](https://doi.org/10.1007/978-3-031-33264-7_9).
- [Hie+22] Philipp HIERONYMI et al. « Decidability for Sturmian words ». In : *30th EACSL Annual Conference on Computer Science Logic*. T. 216. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022, Art. No. 24, 23. ISBN : 978-3-95977-218-1.
- [HU79] John E. HOPCROFT et Jeffrey D. ULLMAN. *Introduction to automata theory, languages, and computation*. Addison-Wesley Series in Computer Science. Addison-Wesley Publishing Co., Reading, MA, 1979, p. x+418. ISBN : 0-201-02988-X.
- [JP02] Jacques JUSTIN et Giuseppe PIRILLO. « Episturmian words and episturmian morphisms ». In : *Theoret. Comput. Sci.* 276.1-2 (2002), p. 281-313. ISSN : 0304-3975,1879-2294. DOI : 10.1016/S0304-3975(01)00207-9. URL : [https://doi.org/10.1016/S0304-3975\(01\)00207-9](https://doi.org/10.1016/S0304-3975(01)00207-9).

- [KT07a] Idrissa KABORÉ et Théodore TAPSOBA. « Combinatoire de mots récurrents de complexité  $n + 2$  ». fr. In : *RAIRO - Theoretical Informatics and Applications - Informatique Théorique et Applications* 41.4 (2007), p. 425-446. DOI : 10.1051/ita:2007027. URL : <http://www.numdam.org/articles/10.1051/ita:2007027/>.
- [KT07b] Idrissa KABORÉ et Théodore TAPSOBA. « Combinatoire de mots récurrents de complexité  $n + 2$  ». In : *Theor. Inform. Appl.* 41.4 (2007), p. 425-446. ISSN : 0988-3754,1290-385X. DOI : 10.1051/ita:2007027. URL : <https://doi.org/10.1051/ita:2007027>.
- [Kar92] Jarkko KARI. « The nilpotency problem of one-dimensional cellular automata ». In : *SIAM J. Comput.* 21.3 (1992), p. 571-586. ISSN : 0097-5397. DOI : 10.1137/0221036. URL : <https://doi.org/10.1137/0221036>.
- [Kar03] Jarkko KARI. « Infinite snake tiling problems ». In : *Developments in language theory*. T. 2450. Lecture Notes in Comput. Sci. Springer, Berlin, 2003, p. 67-77. ISBN : 3-540-40431-7. DOI : 10.1007/3-540-45005-X\_6. URL : [https://doi.org/10.1007/3-540-45005-X\\_6](https://doi.org/10.1007/3-540-45005-X_6).
- [KL23] Jarkko KARI et Victor H. LUTFALLA. « Planar Rosa : a family of quasiperiodic substitution discrete plane tilings with  $2n$ -fold rotational symmetry ». In : *Nat. Comput.* 22.3 (2023), p. 539-561. ISSN : 1567-7818,1572-9796. DOI : 10.1007/s11047-022-09929-8. URL : <https://doi.org/10.1007/s11047-022-09929-8>.
- [KP18] Dominik KEMPA et Nicola PREZZA. « At the roots of dictionary compression : string attractors ». In : *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, 2018, p. 827-840. ISBN : 978-1-4503-5559-9. DOI : 10.1145/3188745.3188814. URL : <https://doi.org/10.1145/3188745.3188814>.
- [KS91] Chul E KIM et Ivan STOJMENOVIĆ. « On the recognition of digital planes in three-dimensional space ». In : *Pattern Recognition Letters* 12.11 (1991), p. 665-669. ISSN : 0167-8655. DOI : [https://doi.org/10.1016/0167-8655\(91\)90003-5](https://doi.org/10.1016/0167-8655(91)90003-5). URL : <https://www.sciencedirect.com/science/article/pii/0167865591900035>.
- [KV98] Orna KUPFERMAN et Moshe Y. VARDI. « Verification of fair transition systems ». In : *Chicago J. Theoret. Comput. Sci.* (1998), Article 2, 37. ISSN : 1073-0486.
- [LL23] Sébastien LABBÉ et Jana LEPŠOVÁ. « A Fibonacci analogue of the two's complement numeration system ». In : *RAIRO Theor. Inform. Appl. (RAIRO:ITA)* 57 (2023), Paper No. 12, 23. ISSN : 2804-7346. DOI : 10.1051/ita/2023007. URL : <https://doi.org/10.1051/ita/2023007>.
- [LAD22] Bastien LABOUREIX, Eric ANDRES et Isabelle DEBLED-RENNESON. « Introduction to Discrete Soft Transforms ». In : *Discrete Geometry and Mathematical Morphology*. Sous la dir. de Étienne BAUDRIER et al. Cham : Springer International Publishing, 2022, p. 422-435. ISBN : 978-3-031-19897-7.
- [LD23] Bastien LABOUREIX et Éric DOMENJOUR. « On the connectedness of arithmetic hyperplanes ». soumis à Theoretical Computer Science. Avr. 2023. URL : <https://hal.science/hal-04310348>.
- [Lin04] Douglas LIND. « Multi-dimensional symbolic dynamics ». In : *Symbolic dynamics and its applications*. T. 60. Proc. Sympos. Appl. Math. Amer. Math. Soc., Providence, RI, 2004, p. 61-79. ISBN : 0-8218-3157-7. DOI : 10.1090/psapm/060/2078846. URL : <https://doi.org/10.1090/psapm/060/2078846>.

- [LM95] Douglas LIND et Brian MARCUS. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995, p. xvi+495. ISBN : 0-521-55124-2; 0-521-55900-6. DOI : 10.1017/CB09780511626302. URL : <https://doi.org/10.1017/CB09780511626302>.
- [Lot02] M. LOTHAIRE. *Algebraic combinatorics on words*. T. 90. Encyclopedia of Mathematics and its Applications. A collective work by Jean Berstel, Dominique Perrin, Patrice Seebold, Julien Cassaigne, Aldo De Luca, Steffano Varricchio, Alain Lascoux, Bernard Leclerc, Jean-Yves Thibon, Veronique Bruyere, Christiane Frougny, Filippo Mignosi, Antonio Restivo, Christophe Reutenauer, Dominique Foata, Guo-Niu Han, Jacques Desarmenien, Volker Diekert, Tero Harju, Juhani Karhumaki and Wojciech Plandowski, With a preface by Berstel and Perrin. Cambridge University Press, Cambridge, 2002, p. xiv+504. ISBN : 0-521-81220-8. DOI : 10.1017/CB09781107326019. URL : <https://doi.org/10.1017/CB09781107326019>.
- [LP92] W. Fred LUNNON et Peter A. B. PLEASANTS. « Characterization of two-distance sequences ». In : *J. Austral. Math. Soc. Ser. A* 53.2 (1992), p. 198-218. ISSN : 0263-6115.
- [Man+21] Sabrina MANTACI et al. « A combinatorial view on string attractors ». In : *Theoret. Comput. Sci.* 850 (2021), p. 236-248. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2020.11.006. URL : <https://doi.org/10.1016/j.tcs.2020.11.006>.
- [Mau01] Christian MAUDUIT. « Multiplicative properties of the Thue-Morse sequence ». In : *Period. Math. Hungar.* 43.1-2 (2001), p. 137-153. ISSN : 0031-5303,1588-2829. DOI : 10.1023/A:1015241900975. URL : <https://doi.org/10.1023/A:1015241900975>.
- [MH38] Marston MORSE et Gustav A. HEDLUND. « Symbolic Dynamics ». In : *Amer. J. Math.* 60.4 (1938), p. 815-866. ISSN : 0002-9327,1080-6377. DOI : 10.2307/2371264. URL : <https://doi.org/10.2307/2371264>.
- [MH40] Marston MORSE et Gustav A. HEDLUND. « Symbolic dynamics II. Sturmian trajectories ». In : *Amer. J. Math.* 62 (1940), p. 1-42. ISSN : 0002-9327,1080-6377. DOI : 10.2307/2371431. URL : <https://doi.org/10.2307/2371431>.
- [MSS16] Hamoon MOUSAVI, Luke SCHAEFFER et Jeffrey SHALLIT. « Decision algorithms for Fibonacci-automatic words, I : Basic results ». In : *RAIRO Theor. Inform. Appl.* 50.1 (2016), p. 39-66. ISSN : 0988-3754,1290-385X. DOI : 10.1051/ita/2016010. URL : <https://doi.org/10.1051/ita/2016010>.
- [Moz89] Shahar MOZES. « Tilings, substitution systems and dynamical systems generated by them ». In : *J. Analyse Math.* 53 (1989), p. 139-186. ISSN : 0021-7670,1565-8538. DOI : 10.1007/BF02793412. URL : <https://doi.org/10.1007/BF02793412>.
- [New77] Donald J. NEWMAN. « Tessellation of integers ». In : *J. Number Theory* 9.1 (1977), p. 107-111. ISSN : 0022-314X,1096-1658. DOI : 10.1016/0022-314X(77)90054-3. URL : [https://doi.org/10.1016/0022-314X\(77\)90054-3](https://doi.org/10.1016/0022-314X(77)90054-3).
- [Och11] Pascal OCHEM. *Pattern Avoidance and HDOL Words*. WORDS 2011 - 8th International Conference on Combinatorics on Words. Poster. Sept. 2011. URL : <https://hal-lirmm.ccsd.cnrs.fr/lirmm-00626116>.
- [Oei+21] Reed OEI et al. *Pecan : An Automated Theorem Prover for Automatic Sequences using Büchi Automata*. 2021. arXiv : 2102.01727 [cs.LG].

- [Pan84] Jean-Jacques PANSIOT. « Complexité des facteurs des mots infinis engendrés par morphismes itérés ». In : *Automata, languages and programming (Antwerp, 1984)*. T. 172. Lecture Notes in Comput. Sci. Springer, Berlin, 1984, p. 380-389. ISBN : 3-540-13345-3. DOI : 10.1007/3-540-13345-3\_34. URL : [https://doi.org/10.1007/3-540-13345-3\\_34](https://doi.org/10.1007/3-540-13345-3_34).
- [Pau75] Michael E. PAUL. « Minimal symbolic flows having minimal block growth ». In : *Math. Systems Theory* 8.4 (1974/75), p. 309-315. ISSN : 0025-5661. DOI : 10.1007/BF01780578. URL : <https://doi.org/10.1007/BF01780578>.
- [Pen84] R. PENROSE. « Pentaplexity : a class of nonperiodic tilings of the plane ». In : *Geometrical combinatorics (Milton Keynes, 1984)*. T. 114. Res. Notes in Math. Pitman, Boston, MA, 1984, p. 55-65. ISBN : 0-273-08675-8.
- [Pen74] Roger PENROSE. « The role of aesthetics in pure and applied mathematical research ». In : *Bulletin of the Institute of Mathematics and Its Applications* 10 (1974), p. 266-271.
- [PP04] Dominique PERRIN et Jean-Éric PIN. *Infinite words : automata, semigroups, logic and games*. Academic Press, 2004.
- [PR12] Dominique PERRIN et Antonio RESTIVO. « A note on Sturmian words ». In : *Theoret. Comput. Sci.* 429 (2012), p. 265-272. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2011.12.047. URL : <https://doi.org/10.1016/j.tcs.2011.12.047>.
- [PS92] Dominique PERRIN et Marcel-Paul SCHÜTZENBERGER. « Synchronizing prefix codes and automata and the road coloring problem ». In : *Symbolic dynamics and its applications (New Haven, CT, 1991)*. T. 135. Contemp. Math. Amer. Math. Soc., Providence, RI, 1992, p. 295-318. ISBN : 0-8218-5146-2. DOI : 10.1090/conm/135/1185096. URL : <https://doi.org/10.1090/conm/135/1185096>.
- [PS15] Elena A. PETROVA et Arseny M. SHUR. « On the tree of ternary square-free words ». In : *Combinatorics on words*. T. 9304. Lecture Notes in Comput. Sci. Springer, Cham, 2015, p. 223-236. ISBN : 978-3-319-23660-5; 978-3-319-23659-9. DOI : 10.1007/978-3-319-23660-5\_19. URL : [https://doi.org/10.1007/978-3-319-23660-5\\_19](https://doi.org/10.1007/978-3-319-23660-5_19).
- [Pre17] Nicola PREZZA. *String Attractors*. 2017. arXiv : 1709.05314 [cs.DS].
- [Pyt+02] N. PYTHEAS FOGG et al. « Sturmian sequences ». In : *Substitutions in Dynamics, Arithmetics and Combinatorics*. Springer Berlin Heidelberg, 2002. Chap. 6, p. 143-198. DOI : 10.1007/3-540-45714-3\_6.
- [Que87] Martine QUEFFÉLEC. *Substitution dynamical systems—spectral analysis*. T. 1294. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1987, p. xiv+240. ISBN : 3-540-18692-1. DOI : 10.1007/BFb0081890. URL : <https://doi.org/10.1007/BFb0081890>.
- [RR22] Praveen M RAJAVEL et R RAMA. *A Note On  $\ell$ -Rauzy Graphs for the Infinite Fibonacci Word*. 2022. arXiv : 2210.08629 [math.CO].
- [Rao11] Michaël RAO. « Last cases of Dejean's conjecture ». In : *Theoret. Comput. Sci.* 412.27 (2011), p. 3010-3018. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2010.06.020. URL : <https://doi.org/10.1016/j.tcs.2010.06.020>.
- [Rau79] Gérard RAUZY. « Échanges d'intervalles et transformations induites ». In : *Acta Arith.* 34.4 (1979), p. 315-328. ISSN : 0065-1036. DOI : 10.4064/aa-34-4-315-328. URL : <https://doi.org/10.4064/aa-34-4-315-328>.



- [Rau82] Gérard RAUZY. « Nombres algébriques et substitutions ». In : *Bull. Soc. Math. France* 110.2 (1982), p. 147-178. ISSN : 0037-9484. URL : [http://www.numdam.org/item?id=BSMF\\_1982\\_\\_110\\_\\_147\\_0](http://www.numdam.org/item?id=BSMF_1982__110__147_0).
- [Rau83] Gérard RAUZY. « Suites à termes dans un alphabet fini ». In : *Seminar on number theory, 1982–1983 (Talence, 1982/1983)*. Univ. Bordeaux I, Talence, 1983, Exp. No. 25, 16.
- [RRS22] Antonio RESTIVO, Giuseppe ROMANA et Marinella SCIORTINO. « String attractors and infinite words ». In : *LATIN 2022: theoretical informatics*. T. 13568. Lecture Notes in Comput. Sci. Springer, Cham, [2022] ©2022, p. 426-442. ISBN : 978-3-031-20623-8; 978-3-031-20624-5. DOI : 10.1007/978-3-031-20624-5\\_26. URL : [https://doi.org/10.1007/978-3-031-20624-5\\_26](https://doi.org/10.1007/978-3-031-20624-5_26).
- [Rev91] Jean-Pierre REVEILLÈS. « Géométrie discrète, calcul en nombres entiers et algorithmique ». Thèse de doct. Strasbourg : Université Louis Pasteur, 1991.
- [Ric21] Gwenaël RICHOMME. « On sets of indefinitely desubstitutable words ». In : *Theoret. Comput. Sci.* 857 (2021), p. 97-113. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2021.01.004. URL : <https://doi.org/10.1016/j.tcs.2021.01.004>.
- [Ros98] Laurent ROSAZ. « Inventories of unavoidable languages and the word-extension conjecture ». In : *Theoret. Comput. Sci.* 201.1-2 (1998), p. 151-170. ISSN : 0304-3975,1879-2294. DOI : 10.1016/S0304-3975(97)00031-5. URL : [https://doi.org/10.1016/S0304-3975\(97\)00031-5](https://doi.org/10.1016/S0304-3975(97)00031-5).
- [Rot94] Günter ROTE. « Sequences with subword complexity  $2n$  ». In : *J. Number Theory* 46.2 (1994), p. 196-213. ISSN : 0022-314X,1096-1658. DOI : 10.1006/jnth.1994.1012. URL : <https://doi.org/10.1006/jnth.1994.1012>.
- [Sal17] Ville SALO. « Decidability and universality of quasiminimal subshifts ». In : *J. Comput. System Sci.* 89 (2017), p. 288-314. ISSN : 0022-0000,1090-2724. DOI : 10.1016/j.jcss.2017.05.017. URL : <https://doi.org/10.1016/j.jcss.2017.05.017>.
- [Sal22] Ville SALO. *Notes and errata on "Decidability and Universality of Quasiminimal Subshifts"*. 2022. URL : <https://villesalo.com/notes/DaUoQSNNotes.html>.
- [SS21] Luke SCHAEFFER et Jeffrey SHALLIT. *String Attractors for Automatic Sequences*. 2021. arXiv : 2012.06840 [cs.FL].
- [Sch64] Marcel-Paul SCHÜTZENBERGER. « On the synchronizing properties of certain prefix codes ». In : *Information and Control* 7 (1964), p. 23-36. ISSN : 0019-9958,1878-2981.
- [Sée18] Patrice SÉÉBOLD. « Sturmian images of non Sturmian words and standard morphisms ». In : *Theoret. Comput. Sci.* 711 (2018), p. 92-104. ISSN : 0304-3975,1879-2294. DOI : 10.1016/j.tcs.2017.11.011. URL : <https://doi.org/10.1016/j.tcs.2017.11.011>.
- [Sha23] Jeffrey SHALLIT. *Proof of Irvine's Conjecture via Mechanized Guessing*. 2023. arXiv : 2310.14252 [math.CO].
- [SSZ23] Jeffrey SHALLIT, Arseny M. SHUR et Stefan ZORCIC. *Power-free Complementary Binary Morphisms*. 2023. arXiv : 2310.15064 [math.CO].
- [She81a] Robert SHELTON. « Aperiodic words on three symbols ». In : *J. Reine Angew. Math.* 321 (1981), p. 195-209. ISSN : 0075-4102,1435-5345. DOI : 10.1515/crll.1981.321.195. URL : <https://doi.org/10.1515/crll.1981.321.195>.

- [She81b] Robert SHELTON. « Aperiodic words on three symbols. II ». In : *J. Reine Angew. Math.* 327 (1981), p. 1-11. ISSN : 0075-4102,1435-5345. DOI : 10.1515/crll.1981.327.1. URL : <https://doi.org/10.1515/crll.1981.327.1>.
- [SS82] Robert O. SHELTON et Raj P. SONI. « Aperiodic words on three symbols. III ». In : *J. Reine Angew. Math.* 330 (1982), p. 44-52. ISSN : 0075-4102,1435-5345.
- [Smi+23] David SMITH et al. « An aperiodic monotile ». In : *arXiv preprint arXiv:2303.10798* (2023).
- [Sto76] Kenneth B. STOLARSKY. « Beatty sequences, continued fractions, and certain shift operators ». In : *Canad. Math. Bull.* 19.4 (1976), p. 473-482. ISSN : 0008-4395,1496-4287. DOI : 10.4153/CMB-1976-071-6. URL : <https://doi.org/10.4153/CMB-1976-071-6>.
- [Stu36] Jacques Charles François STURM. « Mémoire sur les équations différentielles linéaires du second ordre. » In : *Journal de Mathématiques Pures et Appliquées* (1836), p. 106-186. URL : <http://eudml.org/doc/235613>.
- [Tar+22] Lama TARSISSI et al. « Algorithms for Pixelwise Shape Deformations Preserving Digital Convexity ». In : *Discrete Geometry and Mathematical Morphology*. Sous la dir. de Étienne BAUDRIER et al. Cham : Springer International Publishing, 2022, p. 84-97.
- [Tho90] Wolfgang THOMAS. « Automata on infinite objects ». In : *Handbook of theoretical computer science, Vol. B*. Elsevier, Amsterdam, 1990, p. 133-191. ISBN : 0-444-88074-7.
- [Thu06] Axel THUE. *Über unendliche Zeichenreihen*. Skrifter udgivne af Videnskabssekabet i Christiania : Mathematisk-naturvidenskabelig Klasse. 1906.
- [Thu12] Axel THUE. *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*. Skrifter udgivne af Videnskabssekabet i Christiania. 1, Math. Nat. wiss. Kl. 1912, 1. Jacob Dybwad, 1912.
- [Tur36] Alan M. TURING. « On Computable Numbers, with an Application to the Entscheidungsproblem ». In : *Proc. London Math. Soc. (2)* 42.3 (1936), p. 230-265. ISSN : 0024-6115. DOI : 10.1112/plms/s2-42.1.230. URL : <https://doi.org/10.1112/plms/s2-42.1.230>.
- [Wal82] Peter WALTERS. *An introduction to ergodic theory*. T. 79. Graduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1982, p. ix+250. ISBN : 0-387-90599-5.
- [Wan61] Hao WANG. « Proving theorems by pattern recognition — II ». In : *The Bell System Technical Journal* 40.1 (1961), p. 1-41. DOI : 10.1002/j.1538-7305.1961.tb03975.x.
- [Wei73] Benjamin WEISS. « Subshifts of finite type and sofic systems ». In : *Monatsb. Math.* 77 (1973), p. 462-474. ISSN : 0026-9255,1436-5081. DOI : 10.1007/BF01295322. URL : <https://doi.org/10.1007/BF01295322>.
- [Zam18] Luca Q. ZAMBONI. « On the character of words of sublinear complexity ». In : *Acta Arith.* 184.3 (2018), p. 201-213. ISSN : 0065-1036,1730-6264. DOI : 10.4064/aa8577-3-2018. URL : <https://doi.org/10.4064/aa8577-3-2018>.
- [Zol15] Boris ZOLOTOV. *Another Solution to the Thue Problem of Non-Repeating Words*. 2015. arXiv : 1505.00019 [math.CO].