

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COMBINATOIRE DES MOTS : MOTS PARFAITEMENT AMASSANTS,  
TRIPLETS DE MARKOFF ET GRAPHS CHENILLES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN MATHÉMATIQUES

PAR

MÉLODIE LAPOINTE

OCTOBRE 2020

## REMERCIEMENTS

Quatre ans d'efforts sont regroupés dans cette thèse. J'en suis l'auteure, mais je n'ai pas réussi ce projet seule. J'aimerais prendre le temps de remercier les personnes qui m'ont accompagnée tout au long de mon doctorat.

Merci à Christophe Reutenauer et Alexandre Blondin Massé, mes directeurs de recherche. Merci, Christophe, d'avoir partagé avec moi ton incroyable intuition des mathématiques. Merci, Alexandre, pour tes excellents conseils. Vous m'avez énormément appris sur la recherche et l'écriture des mathématiques au fil des années.

Un merci spécial à Pauline et Nadia, qui ont partagé mon bureau à l'UQAM. Votre amitié et votre soutien ont été essentiels à ma réussite. Merci d'avoir été là pour moi. Je n'oublierai pas les tomates, les séances de natation, de vélo ou de mots croisés. Merci aux professeurs, étudiants, postdocs et membres du personnel du LACIM. Vous faites de ce laboratoire un environnement de travail exceptionnel où tout le monde est bienvenu.

Certains des résultats présentés dans cette thèse n'auraient pas vu le jour sans de nombreux collaborateurs. Je voudrais remercier Alain, Élise, Émile, Antoine et Vincent.

Je remercie le CRSNG et l'ISM pour leur soutien financier. Grâce à ces organismes, j'ai eu la chance de me consacrer entièrement à mes études sans souci financier.

Merci à ma famille. Merci à Marc et Ginette, mes parents, pour leur soutien et leur présence. Merci à Martin, mon partenaire de vie. Discuter de maths, d'informatique ou de jeux de société avec toi me rend heureuse. Merci à Chloé, mon petit rayon de soleil.

## TABLE DES MATIÈRES

LISTE DES TABLEAUX . . . . .	v
LISTE DES FIGURES . . . . .	vi
RÉSUMÉ . . . . .	viii
ENGLISH ABSTRACT . . . . .	ix
INTRODUCTION . . . . .	1
CHAPITRE I COMBINATOIRE DES MOTS . . . . .	7
1.1 Mots . . . . .	7
1.2 Morphismes et antimorphismes . . . . .	10
1.3 Opérations classiques sur les mots . . . . .	12
1.4 Permutations et compositions . . . . .	14
CHAPITRE II MOTS DE CHRISTOFFEL . . . . .	16
2.1 Définitions géométriques . . . . .	16
2.2 Arbres . . . . .	21
2.3 Arbre dual des mots de Christoffel . . . . .	25
2.4 Palindromes et mots sturmiens . . . . .	28
2.5 Mots parfaitement amassants . . . . .	29
CHAPITRE III GÉNÉRALISATION DE L'ARBRE DE RANEY . . . . .	33
3.1 Échanges d'intervalles discrets . . . . .	33
3.2 Dénombrement des cycles . . . . .	41
3.3 Arbres des compositions circulaires . . . . .	52
3.4 Type cyclique . . . . .	62
CHAPITRE IV MOTS PARFAITEMENT AMASSANTS : ARBRES ET AUTOMORPHISMES . . . . .	66
4.1 Automorphismes du groupe libre . . . . .	66

4.2 Arbres des mots parfaitement amassants . . . . .	71
4.3 Positivité . . . . .	74
4.4 Mots parfaitement amassants . . . . .	75
4.5 Mots de Lyndon . . . . .	82
4.6 Démonstration du théorème 4.9 . . . . .	87
4.7 Primitivité sur le groupe libre . . . . .	90
CHAPITRE V MOTS DE CHRISTOFFEL ET CONJECTURE DE FRO- BENIUS . . . . .	94
5.1 Nombres de Markoff . . . . .	95
5.2 Une bijection entre le monoïde libre et les triplets de Markoff . . . . .	98
5.3 Mots de Christoffel spéciaux . . . . .	105
5.4 Généralisation de la conjecture de Frobenius . . . . .	112
CHAPITRE VI MOTS FEUILLUS DES GRAPHES CHENILLES . . . . .	121
6.1 Mots préfixes normaux . . . . .	122
6.2 Mots feuillus . . . . .	124
6.3 Graphes chenilles orientées . . . . .	134
6.4 Mots feuillus des graphes chenilles . . . . .	142
CONCLUSION . . . . .	149
RÉFÉRENCES . . . . .	154
INDEX . . . . .	159
INDEX DES NOTATIONS . . . . .	162

## LISTE DES TABLEAUX

Tableau		Page
3.1	Étape du calcul de la fonction $\mathcal{Z}(3, 5, 4, 2)$ . . . . .	43
4.1	Les différents morphismes nécessaires à la démonstration du théorème 4.12 sur le mot <i>acbcacc</i> . . . . .	78
6.1	Fonction feuille du graphe de la figure 6.3 . . . . .	128

## LISTE DES FIGURES

Figure	Page
1.1 La matrice des conjugués du mot <i>appartement</i> . . . . .	14
2.1 Le segment de droite de pente 5/8 et le chemin discret associé . .	17
2.2 La droite de pente 5/8 et le mot de Christoffel supérieur associé .	18
2.3 Règle de construction récursive de l'arbre des paires de Christoffel	21
2.4 Arbre des paires de Christoffel . . . . .	21
2.5 L'arbre des mots de Christoffel . . . . .	22
2.6 L'arbre de Stern-Brocot . . . . .	24
2.7 Règle récursive de construction de l'arbre de Raney . . . . .	24
2.8 L'arbre de Raney . . . . .	25
2.9 L'arbre dual des mots de Christoffel . . . . .	26
2.10 Règle de construction récursive de l'arbre dual des mots de Christoffel	27
2.11 Matrice des conjugués du mot de Christoffel <i>aabaabaabab</i> . . . . .	28
3.1 Un échange d'intervalles discrets de l'ensemble $\llbracket 1, 6 \rrbracket$ . . . . .	34
3.2 Chemin entre $(5, 2)$ et la racine sur l'arbre de Raney . . . . .	53
3.3 Les enfants de la composition $(1, 1, 2, 4)$ . . . . .	54
3.4 Échange d'intervalles discrets symétrique . . . . .	55
3.5 L'arbre des compositions circulaires . . . . .	56
4.1 L'arbre des mots parfaitement amassants . . . . .	73
5.1 Arbre des triplets de Markoff . . . . .	96
5.2 Règle de construction récursive de l'arbre des triplets de Markoff .	97

5.3	Arbre du monoïde libre . . . . .	98
5.4	L'arbre des triplets . . . . .	99
5.5	Le diagramme commutatif correspondant au théorème 5.11 . . . . .	105
5.6	Position des sommets $a\text{Pal}(u)b$ , $a\text{Pal}(ua)b$ et $a\text{Pal}(uab)b$ dans l'arbre des mots de Christoffel. . . . .	107
5.7	La matrice des conjugués de $aaabaab$ et la matrice des valeurs $\mu(u)$ . . . . .	114
6.1	Un graphe simple ayant 8 sommets et 8 arêtes et le sous-graphe induit par $Y = \{1, 2, 3, 7, 6\}$ coloré en bleu . . . . .	125
6.2	Différents types de graphes . . . . .	127
6.3	Un graphe et certains de ses sous-arbres induits . . . . .	127
6.4	Un arbre et sa fonction feuille . . . . .	133
6.5	Le graphe chenille $S$ avec sa colonne orientée en bleu et le graphe chenille $\tilde{S}$ . . . . .	134
6.6	La greffe des chenilles orientées $(5, 2)$ et $(4, 2, 2)$ . . . . .	135
6.7	La chenille orientée $(4, 2, 4, 6, 2, 2)$ et quelques-unes de ses sous-chenilles colorées en bleu . . . . .	137
6.8	Le diagramme de Hasse de l'ensemble partiellement ordonné $(CO, \leq)$ . . . . .	138
6.9	Deux sous-arbres pleinement feuillus de $\text{RC}(00110101100)$ . . . . .	143
6.10	Toutes les chenilles d'ordre 11 dont le mot feuille est 11010101 . . . . .	147
6.11	Un arbre dont le mot feuillu n'est pas préfixe normal . . . . .	148
6.12	La matrice des conjugués du mot $acbbcacbc$ . . . . .	150

## RÉSUMÉ

Dans cette thèse, nous étudions divers problèmes de combinatoire des mots. Ces problèmes sont en lien avec les systèmes dynamiques, le groupe libre, les nombres de Markoff et la théorie des graphes.

D'abord, nous étudions les mots parfaitement amassants, une généralisation des mots de Christoffel sur un alphabet ordonné quelconque. Un mot est parfaitement amassant si sa transformée de Burrows-Wheeler est un mot décroissant. Nous proposons deux nouvelles façons de construire ces mots. D'abord, nous généralisons l'arbre de Raney, c'est-à-dire l'arbre binaire complet infini étiqueté par les fractions réduites, pour obtenir un arbre infini étiqueté par les compositions correspondant à l'image commutative des mots parfaitement amassants. Ensuite, nous proposons un ensemble d'automorphismes du groupe libre engendrant l'ensemble des mots parfaitement amassants. Ces automorphismes permettent de construire l'arbre des mots parfaitement amassants. De plus, nous démontrons que les mots parfaitement amassants sont des éléments primitifs positifs du groupe libre.

Ensuite, nous étudions les triplets de Markoff et la conjecture de Frobenius. Nous proposons d'abord une bijection entre les mots binaires et les triplets de Markoff qui utilise des fonctions de palindromisation. La conjecture de Frobenius suggère que les nombres de Markoff apparaissent comme maximums d'un seul triplet de Markoff. Plusieurs formulations équivalentes sont connues, en particulier en combinatoire des mots. Nous généralisons cette conjecture et nous démontrons que ces généralisations sont valides sur certains sous-ensembles de mots.

Finalement, nous présentons une nouvelle relation entre les mots préfixes normaux et la théorie des graphes. Les mots préfixes normaux sont les mots binaires sur l'alphabet  $\{0, 1\}^*$  dont tout préfixe contient plus de 1 que tous ses facteurs de même longueur. Les mots préfixes normaux sont en bijection avec les mots feuillus des graphes chenilles, c'est-à-dire les graphes tels que le sous-graphe induit par tous les sommets sauf les feuilles est une chaîne simple. Le mot feuillu d'un graphe est la dérivée discrète de la fonction dont l'image est le nombre maximal de feuilles d'un sous-graphe induit ayant  $k$  sommets.

**Mots clés** : Mots de Christoffel, Mot parfaitement amassants, Échange d'intervalles discrets symétriques, Automorphismes, Triplets de Markoff, Conjecture de Frobenius, Mots préfixes normaux, Mots feuillus, graphes chenilles

## ENGLISH ABSTRACT

**English title** : Combinatorics on words : perfectly clustering words, Markoff triples and caterpillars graphs

In this thesis, we study different problems of combinatorics on words. These problems are linked with symbolic dynamics, free groups, Markoff numbers and graph theory.

First, we study perfectly clustering words, a generalization of Christoffel words on an ordered alphabet. A word is perfectly clustering if the Burrows-Wheeler transform of this word is a decreasing word. We give two ways to compute perfectly clustering words. The Raney tree is the complete infinite binary tree labelled by irreducible fractions. We give a generalization of the Raney tree to obtain an infinite tree labelled by all compositions corresponding to the commutative image of perfectly clustering words. We also construct the set of perfectly clustering words using a set of automorphisms of the free group. Moreover, we show that perfectly clustering words are positive primitive elements of the free group.

After, we study Markoff triples and the Frobenius conjecture. We give a bijection between binary words and Markoff triples which uses palindromization functions. The Frobenius conjecture states that every Markoff number appears exactly once as the maximum in a Markoff triple. This conjecture has many equivalent formulations, one of which is related to words. We give a generalization of the Frobenius conjecture on words and prove it in some particular cases.

Finally, we study a new relation between prefix normal words and graph theory. Prefix normal words are binary words such that for any length  $k$ , no factor of length  $k$  has more 1's than the prefix of the same length. We show that prefix normal words are in bijection with leaf words of caterpillar graphs. The induced subgraphs containing all vertices of degree greater than 1 of a caterpillar graph is a simple chain. The leaf word of a graph is the discrete derivative of a function called leaf function. The leaf function is the maximum numbers of leaves that can be realized by an induced subtree of order  $i$  for a given graph.

**Keywords** : Christoffel words, perfectly clustering words, symmetric discrete interval exchanges, automorphisms, Markoff triples, Frobenius conjecture, prefix normal words, leaf words, caterpillar graphs

## INTRODUCTION

La combinatoire des mots est l'étude des suites de symboles finies ou infinies communément appelés mots. De telles suites apparaissent naturellement dans de nombreux contextes tels que la représentation des nombres, la trajectoire d'un système dynamique et les fractions continues. Ainsi, la combinatoire des mots apparait implicitement dans les travaux de Bernoulli ([Bernoulli, 1771](#)), Christoffel ([Christoffel, 1875](#)) et Markoff ([Markoff, 1879, 1880](#)) bien avant de devenir une discipline mathématique à part entière au 20<sup>e</sup> siècle. Dans cette thèse, nous étudions trois problèmes à l'aide de la combinatoire des mots : l'énumération des mots parfaitement amassants, une bijection entre les triplets de Markoff et les mots binaires et les mots feuillus des graphes chenilles à l'aide des mots préfixes normaux.

### **Mots parfaitement amassants**

[Burrows et Wheeler \(1994\)](#) ont introduit la transformée de Burrows-Wheeler, une fonction inversible qui permet de regrouper les occurrences d'une même lettre afin de faciliter la compression des données. Pour obtenir la transformée de Burrows-Wheeler d'un mot, on énumère d'abord ses conjugués selon l'ordre lexicographique. Puis, on conserve la dernière lettre de chacun des mots énumérés précédemment. Un mot dont la transformée de Burrows-Wheeler est de la forme  $a_n^{i_n} a_{n-1}^{i_{n-1}} \dots a_1^{i_1}$  se nomme mot parfaitement amassant ([Simpson et Puglisi, 2008](#)). Ces mots sont remarquables, car ils ont une transformée de Burrows-Wheeler optimale. De plus, les mots parfaitement amassants sont une généralisation d'un autre ensemble de mots binaires connus, appelés mots de Christoffel. En effet, [Mantaci \*et al.\* \(2003\)](#)

ont démontré que sur l’alphabet binaire, la transformée de Burrows-Wheeler d’un mot est de la forme  $b^q a^p$  si et seulement si ce mot est une puissance ou un conjugué d’un mot de Christoffel.

Les mots de Christoffel sont parmi les ensembles de mots les plus étudiés. Aujourd’hui, plus de quinze définitions équivalentes de cet ensemble de mots sont connues utilisant entre autres la transformée de Burrows-Wheeler (Mantaci *et al.*, 2003), les facteurs (Borel et Reutenauer, 2005), les palindromes (de Luca, 1997; de Luca et Mignosi, 1994), les factorisations (Borel et Laubie, 1993; Chuan, 1997), les graphes de Cayley (Christoffel, 1875), les périodes (de Luca et Mignosi, 1994) et les mots de Lyndon (Melançon, 1999). Toutes ces définitions mènent à des généralisations naturelles des mots de Christoffel sur un alphabet fini quelconque. Certaines de ces généralisations ont généré une abondante littérature comme les mots d’Arnoux-Rauzy et les mots épisturmiens (Glen et Justin, 2009).

Une définition équivalente des mots parfaitement amassants a déjà été proposée par Ferenczi et Zamboni (2013). Un mot est parfaitement amassant si et seulement si son carré est un facteur de la trajectoire d’un échange d’intervalles discrets symétrique minimal (Ferenczi et Zamboni, 2013). Le résultat de Ferenczi et Zamboni permet de construire les mots parfaitement amassants directement à l’aide des échanges d’intervalles discrets symétriques minimaux.

Puisqu’un échange d’intervalles discrets symétrique est engendré par une suite d’entiers positifs appelée composition, la question de caractériser les compositions engendrant un échange d’intervalles discrets symétrique minimal est alors naturelle. Pour les compositions ayant 2 intervalles, la réponse est simple : si la composition est  $(c_1, c_2)$ , l’échange d’intervalles discrets symétrique est minimal si et seulement si  $c_1$  et  $c_2$  sont premiers entre eux. Cette question a aussi été résolue pour les compositions ayant 3 intervalles par Pak et Redlich. Plus précisément,

si la composition est  $(c_1, c_2, c_3)$ , alors l'échange d'intervalles discrets symétrique est minimal si et seulement si les entiers  $c_1 + c_2$  et  $c_2 + c_3$  sont premiers entre eux (Pak et Redlich, 2008). De plus, Karnauhova et Liebscher (2017) ont proposé une fonction récursive qui compte le nombre de cycles d'un échange d'intervalles discrets symétrique. Par conséquent, un échange d'intervalles discrets symétrique est minimal si et seulement si l'image de cette fonction est 1. Dans le chapitre 3, nous proposons une démonstration différente de celle de Karnauhova et Liebscher. Ensuite, nous construisons un arbre infini contenant toutes les compositions qui engendrent un échange d'intervalles discrets symétrique minimal. Cet arbre en est une généralisation de l'arbre de Raney (Raney, 1973; Calkin et Wilf, 2000) qui contient toutes les fractions réduites.

Pour les mots parfaitement amassants sur un alphabet ternaire, Simpson et Puglisi (2008) ont proposé une façon de les construire à l'aide de trois fonctions. De plus, l'arbre de Raney est en bijection avec un arbre permettant de construire les mots de Christoffel, appelé l'arbre dual des mots de Christoffel (Berthé *et al.*, 2008). L'arbre dual des mots de Christoffel est construit à l'aide de deux automorphismes du groupe libre. Dans le chapitre 4, nous proposons une généralisation de l'arbre dual des mots de Christoffel. Pour construire cet arbre, nous utilisons deux ensembles d'automorphismes du groupe libre ainsi que deux ensembles de morphismes. Nous étudions les propriétés de ces morphismes et démontrons que l'arbre ainsi obtenu contient tous les mots parfaitement amassants.

### Triplets de Markoff

Les mots de Christoffel apparaissent aussi dans les travaux de Markoff (1879, 1880). Un triplet de Markoff est un triplet d'entiers satisfaisant l'équation diophantienne  $x^2 + y^2 + z^2 = 3xyz$ . Un nombre de Markoff est simplement un élément d'un triplet de Markoff. Markoff a utilisé ces triplets pour définir une suite

de constantes, appelées constantes de Markoff, pour obtenir un théorème sur l'approximation des nombres réels irrationnels : le développement en fraction continue des constantes de Markoff est de la forme  $p^\infty$  où  $p$  est un mot de Christoffel sur l'alphabet  $\{11, 22\}$ .

Cent ans plus tard, Bombieri a décrit le morphisme  $\mu$  du monoïde libre sur deux lettres vers  $SL_2(\mathbb{N})$ . À l'aide de ce morphisme, il a établi une bijection entre un mot de Christoffel avec sa factorisation standard et les triplets de Markoff (Bombieri, 2007). Une autre relation entre les mots et les nombres de Markoff a été démontrée : il existe un mot binaire  $v \in \{a, b\}^*$  tel que la longueur du mot  $\text{Pal} \circ \text{Antipal}(av)$  correspond au nombre de Markoff  $m - 2$  (Reutenauer et Vuillon, 2017). La clôture palindromique  $w^{(+)}$  d'un mot  $w$  est le plus court palindrome pour lequel  $w$  est un préfixe. La palindromisation itérée  $\text{Pal}$  est définie récursivement par  $\text{Pal}(\varepsilon) = \varepsilon$  et pour toute lettre  $a \in \mathcal{A}$ ,  $\text{Pal}(wa) = (\text{Pal}(w)a)^{+}$  (de Luca, 1997). La fonction  $\text{Antipal}(w)$  est définie similairement à  $\text{Pal}$  en utilisant la notion d'antipalindrome au lieu de palindrome (de Luca et De Luca, 2006). L'utilisation de la fonction  $\text{Pal}$  dans ce contexte n'est pas surprenante puisque de Luca (1997) a démontré que  $\text{Pal}$  est une bijection entre les mots binaires et les facteurs propres centraux maximaux des mots de Christoffel.

Dans le chapitre 5, nous utilisons la fonction  $\text{Pal} \circ \text{Antipal}(av)$  pour définir une bijection entre les triplets de Markoff et les mots binaires. Nous étudions aussi les mots de Christoffel impliqués dans cette construction, ceux de la forme  $\text{Pal} \circ \text{Antipal}(av)$ . Ce sont entre autres des mots harmoniques, c'est-à-dire que leur pente est égale au ratio entre les longueurs des facteurs de leurs factorisations standard (Carpi et de Luca, 2004).

Il y a une conjecture célèbre de Frobenius sur les nombres de Markoff ouverte depuis plus de 100 ans : chaque nombre de Markoff apparaît exactement une fois

comme élément maximum d'un triplet de Markoff. Cette conjecture a été reformulée de nombreuses fois (Zhang, 2006; Berstel, 2007; Aigner, 2013; Reutenauer et Vuillon, 2017). Sur les mots, la conjecture de Frobenius revient à démontrer l'injectivité du morphisme  $\mu$  ou bien l'injectivité de la fonction  $\text{Pal} \circ \text{Antipal}(av)$ . Au chapitre 5, on étudie des généralisations de la conjecture de Frobenius basées sur le morphisme  $\mu$ . De plus, on démontre que ce morphisme est injectif sur certains sous-ensembles de mots qui correspondent à des cas particuliers de la conjecture.

### Graphes chenilles

Nous nous intéressons à un ensemble de mots construit à partir de graphes. Ce n'est pas une idée nouvelle, plusieurs notions étudiées sur les mots ont été adaptées aux arbres comme les facteurs, les carrés (Crochemore *et al.*, 2012), les palindromes (Brek *et al.*, 2015) et les mots sturmiens (Berstel *et al.*, 2007). Par exemple, un arbre étiqueté est sturmien s'il a exactement  $n + 1$  sous-arbres distincts de hauteur  $n$  pour tout  $n$ . Dans cette partie de la thèse, nous nous intéressons aux sous-arbres induits ayant un nombre maximal de feuilles. Plus précisément, la fonction feuille d'un graphe  $G$  associe un entier  $i$  au nombre maximal de feuilles d'un sous-arbre induit d'ordre  $i$  de  $G$ . Cette fonction apparaît implicitement dans Blondin Massé *et al.* (2018b) et est définie dans Blondin Massé *et al.* (2018a).

Nous souhaitons décrire l'ensemble des mots qui représentent l'image de fonctions feuilles. Plus spécifiquement, étant donné un mot  $l_1 l_2 \dots l_n$ , existe-il un graphe  $G$  tel que la dérivée discrète de la fonction feuille de  $G$  soit  $l_1 l_2 \dots l_n$ ? Les mots ayant cette propriété sont appelés mots feuillus. Cette question est inspirée des suites graphiques (Erdős et Gallai, 1961) : étant donné une suite d'entiers  $(l_1, \dots, l_n)$ , existe-t-il un graphe avec les sommets  $1, 2, \dots, n$  dont le sommet  $x_i$  soit de degré  $i$ ? Au chapitre 6, nous décrivons les mots feuillus d'un type de graphes ap-

pelé graphe chenille ([Harary et Schwenk, 1973](#)). Pour ce faire, nous orientons les graphes chenilles et introduisons une opération binaire sur les chenilles orientées qui transforme l'ensemble des chenilles orientées en un monoïde isomorphe au monoïde des mots sur l'alphabet binaire. Finalement, on démontre que les mots feuillus des graphes chenilles sont en bijection avec les mots préfixes normaux définis par [Fici et Lipták \(2011\)](#). Ces mots binaires sont définis par la propriété suivante : pour n'importe quelle longueur  $k$  d'un mot, aucun de ses facteurs de longueur  $k$  ne contient plus de 1 que son préfixe de longueur  $k$ .

Les deux premiers chapitres sont consacrés à définir les notions de base de la combinatoire des mots et les mots de Christoffel qui sont les principaux outils nécessaires à la compréhension des chapitres subséquents. On termine en présentant divers problèmes ouverts reliés à chacun de ces thèmes.

## CHAPITRE I

### COMBINATOIRE DES MOTS

Les notions de base de la combinatoire des mots sont présentées dans ce chapitre ainsi que des notions combinatoires classiques. Pour plus de détails, consulter ([Lothaire, 1997](#)).

#### 1.1 Mots

Un *alphabet* est un ensemble fini  $\mathcal{A}$ . Ses éléments sont appelés des *lettres*. Un *mot* sur l'alphabet  $\mathcal{A}$  est une suite finie d'éléments de  $\mathcal{A}$ ,  $(w_1, w_2, \dots, w_n)$ . La *longueur* d'un mot  $w = (w_1, w_2, \dots, w_n)$ , notée  $|w|$ , est  $n$ . L'unique mot de longueur 0, appelé *mot vide*, est dénoté par  $\varepsilon$ . La *concaténation* de deux mots  $u = (u_1, \dots, u_m)$  et  $v = (v_1, \dots, v_n)$ , notée  $uv$ , est obtenue en écrivant le mot  $u$  suivi du mot  $v$  :  $uv = (u_1, \dots, u_m, v_1, \dots, v_n)$ . La concaténation étant une opération associative, pour alléger le texte, les mots sont écrits  $w_1w_2 \dots w_n$  au lieu de  $(w_1, w_2, \dots, w_n)$ . L'ensemble des mots sur l'alphabet  $\mathcal{A}$  muni de la concaténation forme un *monoïde libre* dont l'élément neutre est le mot vide  $\varepsilon$  et est noté par  $\mathcal{A}^*$ .

Le nombre d'occurrences d'une lettre  $a$  dans le mot  $w$  est noté  $|w|_a$ . La fonction  $\text{Alph}$  est définie par  $\text{Alph}(w) = \{x \in \mathcal{A} \mid |w|_x \geq 1\}$ . Si  $w \in \mathcal{A}^*$ , alors  $\text{Alph}(w) \subseteq \mathcal{A}$ . Le mot  $w$  est dit *complet* si  $\text{Alph}(w) = \mathcal{A}$ . Par exemple, la phrase « Portez ce whisky au vieux juge blond qui fume. » est complète sur l'alphabet  $\{a, b, c, \dots, z\}$ .

Un mot  $u$  est un *facteur* de  $w$ , s'il existe des mots  $x, y$  appartenant à  $\mathcal{A}^*$  tels que  $w = xuy$ . Un facteur est dit *propre*, si  $u \neq w$ . De plus, si  $x = \varepsilon$ , alors  $u$  est appelé *préfixe* de  $w$ . De même, si  $y = \varepsilon$ , alors  $u$  est un *suffixe* de  $w$ . L'ensemble des facteurs d'un mot  $w$  est désigné par  $\text{Fact}(w)$ ; celui des préfixes par  $\text{Pref}(w)$  et celui des suffixes par  $\text{Suff}(w)$ . Pour préciser la longueur des facteurs, on ajoute celle-ci en indice sur les ensembles. Ainsi,  $\text{Fact}_n(w)$  désigne l'ensemble de tous les facteurs de longueur  $n$  du mot  $w$ .

**Exemple 1.1.** *Le mot muse est un facteur propre du mot amusement. Le mot amuse est un préfixe de ce mot et le mot ment en est un suffixe.*

Un mot  $u$  est la *puissance  $n^{\text{ième}}$*  du mot  $w$ , notée  $w^n$  si  $u = \underbrace{ww \dots w}_{n \text{ fois}}$ . Un mot  $w$  est *primitif* s'il n'est pas la puissance d'un autre mot, c'est-à-dire qu'il n'existe pas de  $n > 1$  tel que  $w = z^n$ .

**Exemple 1.2.** *Le mot coucou n'est pas primitif, car coucou = (cou)<sup>2</sup>.*

Le mot  $u$  est *conjugué* au mot  $w$ , s'il existe deux autres mots  $x$  et  $y$  tel que  $w = xy$  et  $u = yx$ . La relation «être conjugué», dénoté par  $\equiv$ , est une relation d'équivalence. La *classe de conjugaison* d'un mot est l'ensemble de ses conjugués. Le nombre de conjugués d'un mot primitif est égale à sa longueur. La classe de conjugaison d'un mot  $w$  est aussi appelée *mot circulaire* et est notée par  $(w)$ .

**Exemple 1.3.** *Le mot prendre est un conjugué du mot reprend, en prenant  $x = prend$  et  $y = re$ . Le mot prendre est primitif, alors sa classe de conjugaison contient 7 mots :*

$$(prendre) = \{prendre, rendrep, endrepr, ndrepre, drepren, reprend, eprendr\}.$$

Soit  $\mathcal{A}$  un alphabet totalement ordonné. L'*image commutative*<sup>1</sup> d'un mot  $w$  est

---

1. Aussi appelée vecteur de Parikh.

la suite d'entiers  $(|w|_a)_{a \in \mathcal{A}}$  qui indique le nombre d'occurrences de chacune des lettres dans le mot  $w$ . Deux mots  $u$  et  $v$  sont dits *abéliens équivalents* si leurs images commutatives sont égales, c'est-à-dire que  $|u|_a = |v|_a$  pour tout  $a \in \mathcal{A}$ .

**Exemple 1.4.** *L'image commutative du mot étincelle sur l'alphabet  $\{c, e, i, l, n, t\}$  tel que  $c < e < i < l < n < t$  est  $(1, 3, 1, 2, 1, 1)$ . Les mots clientèle et étincelle sont abéliens équivalents en ignorant les accents.*

Un mot  $w$  est dit *équilibré* si pour tous  $u$  et  $v$  facteurs de  $w$  tels que  $|u| = |v|$ ,

$$||u|_a - |v|_a| \leq 1, \text{ pour toute lettre } a \in \mathcal{A}.$$

Autrement dit, le nombre d'occurrences de la lettre  $a$  contenu dans deux facteurs  $u$  et  $v$  de même longueur diffère d'au plus 1. Par exemple, le mot *ananas* est équilibré.

L'*ordre lexicographique* est une extension de l'ordre total de l'alphabet  $\mathcal{A}$  sur le monoïde  $\mathcal{A}^*$  défini comme suit : soit  $u, v \in \mathcal{A}^*$ , on écrit  $u \prec v$  si  $u$  est un préfixe propre de  $v$  ou  $u = rxs$  et  $v = ryt$  avec  $x < y$  et  $x, y \in \mathcal{A}$  et  $r, s, t \in \mathcal{A}^*$ . Un mot est appelé *mot de Lyndon* s'il est primitif et minimal selon l'ordre lexicographique dans sa classe de conjugaison.

**Exemple 1.5.** *Les mots appartement et ananas sont des mots de Lyndon. Le mot coucou n'est pas un mot de Lyndon puisqu'il n'est pas primitif. De plus, le mot banane n'est pas un mot de Lyndon puisque ananeb est l'un de ses conjugués et que ananeb  $\prec$  banane.*

Les mots de Lyndon peuvent être définis de différentes façons.

**Proposition 1.6** (Chen *et al.* (1958)). *Les énoncés suivants sont équivalents :*

- $w$  est un mot de Lyndon ;
- pour tout  $u, v$  non vide tel que  $w = uv$ ,  $w \prec vu$  ;

- pour tous les suffixes propres  $s$  de  $w$ ,  $w \prec s$ .

La démonstration de ce résultat classique est disponible dans (Lothaire, 1997).

L'image miroir d'un mot  $w = w_1 \dots w_n$ , notée  $\tilde{w}$ , est obtenue en écrivant les lettres de  $w$  dans l'ordre inverse, c'est-à-dire  $\tilde{w} = w_n \dots w_2 w_1$ . Un mot  $w$  est appelé un *palindrome* si  $w = \tilde{w}$ . Par exemple, le plus long palindrome de la langue française est le mot « ressasser ».

## 1.2 Morphismes et antimorphismes

Un *morphisme de monoïde*, simplement appelé morphisme, est une fonction  $f$  entre deux monoïdes  $M$  et  $N$  compatible avec les opérations de  $M$  et  $N$ , c'est-à-dire  $f(uv) = f(u)f(v)$  pour tout  $u, v \in M$  et  $f(1_M) = 1_N$ . Un morphisme de  $M$  vers lui-même est appelé un *endomorphisme*. L'ensemble des endomorphismes de  $M$  est noté par  $\text{End}(M)$ . L'ensemble  $\text{End}(M)$  est un monoïde avec l'opération de composition et le morphisme identité est l'élément neutre. Un *automorphisme* est un endomorphisme bijectif.

Soit  $\mathcal{A} = \{a_1, \dots, a_r\}$  un alphabet totalement ordonné. Le monoïde  $\mathcal{A}^*$  est dit *libre* sur  $\mathcal{A}$ , c'est-à-dire que toute fonction de  $\mathcal{A}$  vers un monoïde a une unique extension en un morphisme de monoïde de  $\mathcal{A}^*$  vers ce monoïde. Par conséquent, sur le monoïde libre  $\mathcal{A}^*$ , il suffit de définir l'image des éléments de  $\mathcal{A}$  pour définir un morphisme. Pour simplifier la notation, on note  $(u_1, \dots, u_r)$  l'endomorphisme  $f$  de  $\mathcal{A}^*$  défini par  $f(a_i) = u_i$  pour tout  $i \in \{1, \dots, r\}$ .

**Exemple 1.7.** *Le morphisme de Thue-Morse  $\theta$  défini par  $\theta(a) = ab$  et  $\theta(b) = ba$  est un exemple d'endomorphisme sur  $\{a, b\}^*$ .*

Un morphisme  $f$  est *strictement alphabétique* si pour tout  $a \in \mathcal{A}$ ,  $f(a) \in \mathcal{A}$ . Une

fonction  $f$  telle que  $f \circ f = \text{Id}$  est appelée une *involution*. On observe alors qu'une fonction qui est à la fois un endomorphisme et un involution est un automorphisme.

Un *antimorphisme* est une fonction  $f$  de  $M$  vers  $N$  telle que pour tous  $u, v \in M$ ,  $f(uv) = f(v)f(u)$ .

**Exemple 1.8.** *L'image miroir est un antimorphisme de  $\mathcal{A}^*$ . En effet, posons  $u = u_1 \dots u_m$  et  $v = v_1 \dots v_n$ . Alors, on a que*

$$\widetilde{uv} = \overbrace{u_1 \dots u_m v_1 \dots v_n} = v_n \dots v_1 u_m \dots u_1 = \widetilde{v_1 \dots v_n} \widetilde{u_1 \dots u_m} = \widetilde{v} \widetilde{u}.$$

Certaines fonctions sont compatibles avec l'ordre sur l'alphabet  $\mathcal{A}$ . Une fonction  $f$  de  $\mathcal{A}^*$  est dite *croissante* si pour tous  $u, v \in \mathcal{A}^*$ ,  $u \prec v$  implique que  $f(u) \prec f(v)$ . De même, une fonction  $f$  de  $\mathcal{A}^*$  est dite *décroissante* si  $u \prec v$  implique  $f(u) \succ f(v)$  pour tous  $u, v \in \mathcal{A}^*$ .

**Exemple 1.9.** *Le morphisme de Thue-Morse  $\theta$  est croissant sur  $\{a, b\}^*$  où  $a \prec b$ . En effet, soit  $u \prec v$  deux mots de  $\mathcal{A}^*$ . Il y a deux cas : (i)  $u$  préfixe de  $v$  ou (ii)  $u$  n'est pas un préfixe de  $v$ . (i) Si  $u$  est un préfixe de  $v$ , alors  $\theta(u)$  est un préfixe de  $\theta(v)$ . (ii) Si  $u$  n'est pas un préfixe de  $v$ , alors il existe des mots  $r, s, t$  et deux lettres  $x \prec y$  tels que  $u = rxs$  et  $v = ryt$ . On a alors que  $\theta(u) = \theta(r)\theta(x)\theta(s)$  et  $\theta(v) = \theta(r)\theta(y)\theta(t)$ . On observe que  $\theta(x) \prec \theta(y)$  si  $x \prec y$ . Par conséquent,  $\theta(u) \prec \theta(v)$ .*

Un ensemble  $X$  est *fermé par* une fonction  $f$  si pour tout  $x \in X$ ,  $f(x) \in X$ .

**Exemple 1.10.** *L'ensemble des conjugués du mot ananas est*

$$X = \{\text{ananas}, \text{nanasa}, \text{anasan}, \text{nasana}, \text{asanan}, \text{sanana}\}.$$

*On peut vérifier aisément que  $X$  est fermé par l'image miroir.*

### 1.3 Opérations classiques sur les mots

La *clôture palindromique (à droite)* d'un mot  $w = ps$ , notée  $w^{(+)}$ , est le mot  $ps\tilde{p}$  où  $s$  est le plus long suffixe de  $w$  qui soit un palindrome. On peut démontrer que la clôture palindromique d'un mot est égale au plus court palindrome dont ce mot est un préfixe (de Luca, 1997). La *palindromisation itérée*, dénotée par  $\text{Pal}$ , est définie par  $\text{Pal}(\varepsilon) = \varepsilon$  et pour tout mot  $v$  et toute lettre  $x$ ,  $\text{Pal}(vx) = (\text{Pal}(v)x)^{(+)}$ . La clôture palindromique et la palindromisation itérée ont été introduites par de Luca (1997).

**Exemple 1.11.** *La clôture palindromique du mot  $essaya$  est  $essayasse$ . La palindromisation itérée du mot  $sas$  est  $\text{Pal}(sas) = sassas$ .*

Justin (2005) a exprimé la fonction  $\text{Pal}$  en fonction de morphismes. Plus précisément, soit le morphisme  $\kappa_a$  défini par  $\kappa_a(a) = a$  et  $\kappa_a(x) = ax$  pour tout  $x \neq a$ .

**Théorème 1.12** (Justin (2005)). *Pour tout mot  $w \in \mathcal{A}^*$  et toute lettre  $x \in \mathcal{A}$ , nous avons que*

$$\text{Pal}(xw) = \kappa_x(\text{Pal}(w))x.$$

Soit  $\{a, b\}$  un alphabet. Le *complément* du mot  $w \in \{a, b\}^*$ , noté  $\bar{w}$  est défini par  $\bar{a} = b$  et  $\bar{b} = a$ . La fonction  $\hat{\cdot}$  est définie par  $\hat{w} = \widetilde{\bar{w}}$ . Il est simple de vérifier que le complément est un morphisme et que la fonction  $\hat{\cdot}$  est un antimorphisme. Un *antipalindrome* est un mot  $w$  satisfaisant l'égalité  $w = \hat{w}$ . On note que la fonction complément est définie sur un alphabet binaire, alors la notion d'antipalindrome n'a de sens que sur les alphabets binaires.

**Exemple 1.13.** *Le mot  $abbaab$  est un antipalindrome, car  $\widehat{abbaab} = abbaab$ .*

En utilisant une approche similaire à la clôture palindromique et à la palindromisation itérée, de Luca et De Luca (2006) ont défini la clôture antipalindromique

et l'antipalindromisation itérée. La *clôture antipalindromique* d'un mot  $w = ps$ , noté  $w^\oplus$ , est le mot  $ps\hat{p}$  où  $s$  est le plus long suffixe de  $w$  qui est un antipalindrome. Ainsi, la clôture antipalindromique d'un mot correspond au plus court antipalindrome dont  $w$  est un préfixe. La fonction  $\text{Antipal}(w)$  est définie par  $\text{Antipal}(vx) = (\text{Antipal}(v)x)^\oplus$  pour tout  $v \in \{a, b\}^*$  et tout  $x \in \{a, b\}$ .

**Exemple 1.14.** *La clôture antipalindromique du mot  $abb$  est  $abbaab$ . De plus,  $\text{Antipal}(abb) = abbaabbaab$ .*

Soit  $w$  un mot de longueur  $n$  sur l'alphabet totalement ordonné  $\mathcal{A}$  et ses conjugués  $w_1 \preceq w_2 \preceq \dots \preceq w_n$ . La *matrice des conjugués* de  $w$  est la matrice  $M(w)$  dont l'élément sur la ligne  $i$  est  $w_i$ . La *transformée de Burrows-Wheeler*, notée  $\text{BW}$ , est une fonction du monoïde  $\mathcal{A}^*$  vers lui-même définie par  $\text{BW}(w) = d_1 \dots d_n$  où  $d_i$  est le  $n^{\text{ième}}$  élément de la  $i^{\text{ième}}$  ligne de la matrice  $M(w)$ .

**Exemple 1.15.** *La matrice des conjugués du mot appartement est illustré à la figure 1.1 et sa transformée de Burrows-Wheeler est  $tptmeepaanr$ .*

La transformée de Burrows-Wheeler n'est pas surjective, car les mots débutant par la plus petite lettre de l'alphabet, comme *ananas*, n'ont pas d'antécédent. Elle n'est pas injective, car deux conjugués ont la même matrice des conjugués et donc la même transformée de Burrows-Wheeler. En fait, seul les conjugués d'un mot ont la même image.

**Proposition 1.16** (Mantaci et al. (2003)). *Soit deux mots  $u, v \in \mathcal{A}^*$ . Les mots  $u$  et  $v$  sont conjugués si et seulement si  $\text{BW}(u) = \text{BW}(v)$ .*

Une démonstration de cette proposition est dans (Ferenczi et Zamboni, 2013).

**Proposition 1.17** (Mantaci et al. (2003)). *Soit  $u, v \in \mathcal{A}^*$ . Le mot  $u$  est conjugué à  $v^d$  avec  $d \in \mathbb{N}$  si et seulement si  $\text{BW}(v) = v_1 \dots v_n$  et  $\text{BW}(u) = v_1^d \dots v_n^d$ .*

<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>
<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>
<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>
<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>
<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>
<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>
<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>
<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>
<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>
<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>	<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>
<i>t</i>	<i>e</i>	<i>m</i>	<i>e</i>	<i>n</i>	<i>t</i>	<i>a</i>	<i>p</i>	<i>p</i>	<i>a</i>	<i>r</i>

Figure 1.1: La matrice des conjugués du mot *appartement*

La transformée de Burrows-Wheeler du mot *appartement* est *tptmeepaanr* et celle du mot  $(\textit{appartement})^3$  est  $t^3p^3t^3m^3e^6p^3a^6n^3r^3$ . La démonstration de cette proposition se trouve dans (Mantaci *et al.*, 2003).

#### 1.4 Permutations et compositions

On dénote par  $\llbracket a, b \rrbracket$  l'ensemble des entiers  $\{a, a + 1, \dots, b\}$  où  $a, b \in \mathbb{N}$ . Une *composition* de  $n$  est une suite d'entiers positifs non nuls  $(c_1, \dots, c_r)$  dont la somme est  $n$ . La longueur d'une composition est le nombre d'éléments de sa suite, c'est-à-dire que la longueur  $(c_1, \dots, c_r)$  est  $r$ . Une *composition faible* de  $n$  est une suite d'entiers positifs ou nuls  $(c_1, \dots, c_r)$  dont la somme est  $n$ .

Une *permutation* de l'ensemble  $\llbracket 1, r \rrbracket$  est une bijection de l'ensemble  $\llbracket 1, r \rrbracket$  vers

lui-même. Une permutation  $\sigma$  peut-être représentée par le mot

$$\sigma(1), \sigma(2), \dots, \sigma(r).$$

**Exemple 1.18.** La permutation symétrique de l'ensemble  $\llbracket 1, r \rrbracket$  est définie par  $\sigma(k) = r - k + 1$  pour tout  $k \in \llbracket 1, r \rrbracket$ . Le mot qui représente la permutation symétrique est  $r, r - 1, \dots, 1$ .

L'orbite de  $x$  sous l'action de la permutation  $\sigma$ , notée par  $O_x$ , est l'ensemble  $\{\sigma^k(x) \mid k \in \mathbb{N}\}$ . Le cycle de  $x$  est le mot circulaire  $(\sigma^0(x)\sigma^1(x)\dots\sigma^p(x))$  où  $p = |O_x|$ . Deux cycles sont *disjoints* si leurs mots circulaires respectifs ne sont pas égaux. On dénote par  $\gamma(\sigma)$  le nombre de cycles disjoints de  $\sigma$ . Une permutation  $\sigma$  est dite *circulaire* si elle est composée d'un seul cycle disjoint, c'est-à-dire  $\gamma(\sigma) = 1$ .

**Exemple 1.19.** Le mot  $5, 4, 1, 2, 9, 10, 11, 8, 3, 7, 6$  représente une permutation de l'ensemble  $\llbracket 1, 11 \rrbracket$ . L'orbite de 5 sous l'action de cette permutation est  $\{1, 3, 5, 9\}$ . Cette permutation a quatre cycles disjoints :  $(1, 5, 9, 3)$ ,  $(2, 4)$ ,  $(6, 10, 7, 11)$  et  $(8)$ .

La notation *cyclique* d'une permutation  $\sigma$  représente une permutation par le produit de ses cycles disjoints. Par exemple, la permutation symétrique est représentée par le produit de cycles  $(1, r)(2, r - 1) \dots (\lceil \frac{r}{2} \rceil, \lceil \frac{r+1}{2} \rceil)$  si  $r$  est pair et  $(1, r)(2, r - 1) \dots (\lceil \frac{r-1}{2} \rceil, \lceil \frac{r+2}{2} \rceil)(\lceil \frac{r}{2} \rceil)$  si  $r$  est impair. La permutation  $5, 4, 1, 2, 9, 10, 11, 8, 3, 7, 6$  est représentée par le produit de cycles  $(1, 5, 9, 3)(2, 4)(6, 10, 7, 11)(8)$ .

## CHAPITRE II

### MOTS DE CHRISTOFFEL

Dans ce chapitre, nous nous intéressons aux mots de Christoffel, les mots binaires correspondant à la discrétisation des droites de pentes rationnelles dans le plan à coordonnées entières. Plus précisément, nous présentons la définition géométrique de ces mots et les factorisations qui en découlent. Ensuite, nous présentons divers arbres binaires infinis étiquetés qui permettent de construire les mots de Christoffel et mettent en évidence leurs liens avec les fractions réduites. Nous enchainons avec des morphismes qui engendrent les mots de Christoffel et nous terminons en discutant d'une généralisation des mots de Christoffel appelés mots parfaitement amassants.

Ce chapitre ne donne pas une présentation exhaustive sur les mots de Christoffel, mais plutôt un survol des résultats que nous utilisons dans cette thèse. De plus, toutes les démonstrations sont omises. Les références suivantes font une présentation détaillée des mots de Christoffel ([Lothaire, 2002](#); [Berstel, 2007](#); [Berstel \*et al.\*, 2009](#); [Reutenauer, 2019](#)).

#### 2.1 Définitions géométriques

Un *pas élémentaire* est un segment de la forme  $[(x, y), (x+1, y)]$  ou  $[(x, y), (x, y+1)]$  avec  $x, y \in \mathbb{Z}$ . Un *chemin discret* est une suite de pas élémentaires consécutifs.

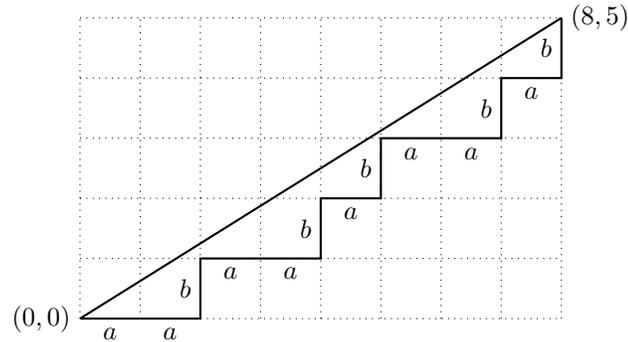


Figure 2.1: Le segment de droite de pente  $5/8$  et le chemin discret associé

Soit  $p$  et  $q$  deux entiers positifs ou nuls premiers entre eux. Soit le segment de droite compris entre les points à coordonnées entières  $A$  et  $B$  avec  $B - A = (q, p)$  et le chemin discret entre les points  $A$  et  $B$  passant en dessous de ce segment tel que l'intérieur du polygone formé par le segment et le chemin discret ne contienne aucun point à coordonnées entières.

**Définition 2.1.** Soit  $\{a, b\}$  un alphabet tel que  $a < b$ . Le mot de Christoffel inférieur de pente  $q/p$  est le mot appartenant à  $\{a, b\}^*$  représentant le chemin discret décrit ci-dessus où les pas élémentaires de la forme  $[(x, y), (x + 1, y)]$  sont identifiés par  $a$  et ceux de la forme  $[(x, y), (x, y + 1)]$  par  $b$ .

**Exemple 2.2.** Le mot de Christoffel inférieur de pente  $5/8$  est  $aabaababaabab$ . La figure 2.1 illustre le segment de droite de pente  $5/8$  ainsi que le chemin discret qui lui est associé.

**Définition 2.3.** Soit  $\{a, b\}$  un alphabet tel que  $a < b$ . Le mot de Christoffel supérieur de pente  $q/p$  est le mot appartenant à  $\{a, b\}^*$  représentant le chemin discret entre les points  $A$  et  $B$  passant au dessus du segment  $\overline{AB}$  tel que le polygone intérieur formé par le segment  $\overline{AB}$  et le chemin discret ne contienne aucun point à coordonnées entières où les pas élémentaires de la forme  $[(x, y), (x + 1, y)]$  sont identifiés par  $a$  et ceux de la forme  $[(x, y), (x, y + 1)]$  par  $b$ .

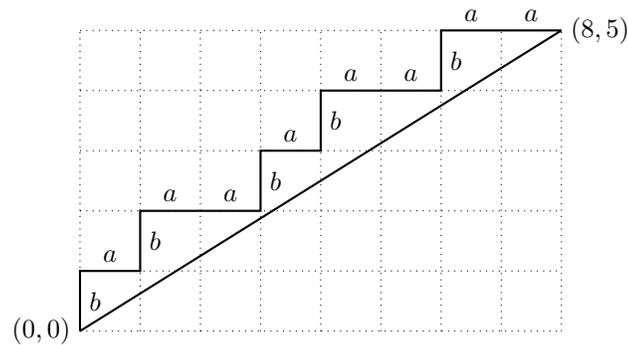


Figure 2.2: La droite de pente  $5/8$  et le mot de Christoffel supérieur associé

**Exemple 2.4.** *Le mot de Christoffel supérieur de pente  $5/8$  est  $babaababaabaa$ . La figure 2.2 illustre le segment de droite de pente  $5/8$  ainsi que le chemin discret qui lui est associé.*

On observe que le *mot de Christoffel supérieur* est obtenu en prenant le chemin discret au-dessus du segment de la droite au lieu de celui en dessous (voir la figure 2.2). Puisque qu'une rotation de  $180^\circ$  autour du centre d'un rectangle préserve celui-ci, un mot de Christoffel supérieur de pente  $q/p$  est égal à l'image miroir du mot de Christoffel inférieur de même pente. Nous nous concentrons sur les mots de Christoffel inférieurs et, pour alléger le texte, les désignons simplement par *mots de Christoffel*.

Les mots  $a$  et  $b$  sont les mots de Christoffel de pente  $0 = 0/1$  et  $\infty = 1/0$  respectivement. Ce sont les seuls mots de Christoffel qui sont à la fois des mots de Christoffel inférieurs et supérieurs. Les autres mots de Christoffel sont dits *propres*. La pente d'un mot de Christoffel est la pente du segment de droite qu'il discrétise. Ainsi, le nombre d'occurrences de la lettre  $a$  dans un mot de Christoffel est  $p$  et le nombre d'occurrences de la lettre  $b$  est  $q$ . La longueur d'un mot de Christoffel  $w$  est donc  $|w|_a + |w|_b = p + q$ .

**Proposition 2.5.** *Un mot de Christoffel et tous ses conjugués sont primitifs.*

Cette proposition se démontre par l'absurde à l'aide de la pente d'un mot de Christoffel. De plus, les symétries du rectangle déterminent la forme des mots de Christoffel.

**Proposition 2.6.** *Un mot de Christoffel propre est de la forme  $amb$  où  $m$  est un palindrome.*

Le palindrome  $m$  est appelé *mot central*. Par exemple,  $abaababaaba$  est le mot central du mot de Christoffel de pente  $5/8$ .

La *factorisation standard* d'un mot de Christoffel propre  $w$  est une paire de mots  $(u, v)$  telle que  $u$  et  $v$  sont des mots de Christoffel et  $w = uv$ . Toute paire de mots  $(u, v)$  telle que les mots  $u$ ,  $v$  et  $uv$  sont des mots de Christoffel est appelée *paire de Christoffel*. De même, un triplet  $(u, w, v)$  formé d'un mot de Christoffel  $w$  et de sa factorisation standard  $w = uv$  est appelé *triplet de Christoffel*. Le théorème suivant garantit l'existence et l'unicité de cette factorisation.

**Théorème 2.7** (Borel et Laubie (1993)). *Un mot de Christoffel propre a une unique factorisation en un produit de deux mots de Christoffel.*

**Exemple 2.8.** *La factorisation standard du mot de Christoffel  $aabaababaabab$  est  $(aabaabab, aabab)$ . Ces deux mots sont respectivement les mots de Christoffel de pente  $3/5$  et  $2/3$ .*

L'idée de la démonstration du théorème 2.7 consiste à diviser le chemin discret en deux à un point situé à distance minimale du segment sur le chemin discret et à démontrer que les mots ainsi obtenus sont des mots de Christoffel.

À partir de la factorisation standard, il est possible de définir une autre factorisation. En effet, les mots de la paire  $(u, v)$  sont soit des lettres, soit de la

forme  $au'b$  et  $av'b$  avec  $u'$  et  $v'$  des palindromes, par la proposition 2.6. Le mot  $w = au'bav'b = av'abu'b$  peut alors être factorisé en un produit de deux palindromes  $av'a$  et  $bu'b$ . La *factorisation palindromique* d'un mot de Christoffel  $w$  est la paire de palindromes  $(v'', u'')$  telle que  $w = v''u''$  (Chuan, 1997).

**Théorème 2.9** (Chuan (1997)). *Tout mot de Christoffel propre  $w$  a une unique factorisation palindromique  $w = v'u'$  avec  $v'$  et  $u'$  des palindromes. En plus, si  $(u, v)$  est la paire de Christoffel tel que  $w = uv$ , alors  $|u| = |u'|$  et  $|v| = |v'|$ .*

**Exemple 2.10.** *La factorisation palindromique du mot  $aabaababaabab$  est la paire  $(aaba, babaabab)$ .*

L'existence de la factorisation palindromique découle de la factorisation standard et de la proposition 2.6 tandis que l'unicité est une conséquence de la primitivité des mots de Christoffel grâce à la proposition suivante.

**Proposition 2.11** (de Luca et Mignosi (1994)). *Soit  $w$  un mot primitif qui est un produit de deux palindromes. Alors cette factorisation est unique.*

De plus, il est possible de construire récursivement les mots de Christoffel grâce à la factorisation standard.

**Proposition 2.12** (Borel et Laubie (1993)). *Soit  $u$  et  $v$  deux mots de Christoffel. La paire  $(u, v)$  est une paire de Christoffel si et seulement si*

$$\det \begin{pmatrix} |u|_a & |u|_b \\ |v|_a & |v|_b \end{pmatrix} = 1.$$

La condition sur le déterminant est nécessaire, car la concaténation des mots de Christoffel  $aab$  et  $abb$  est  $aababb$  n'est pas un mot de Christoffel. La démonstration de ce résultat est aussi géométrique (Borel et Laubie, 1993; Reutenauer, 2019).

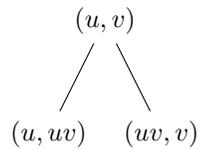


Figure 2.3: Règle de construction récursive de l'arbre des paires de Christoffel

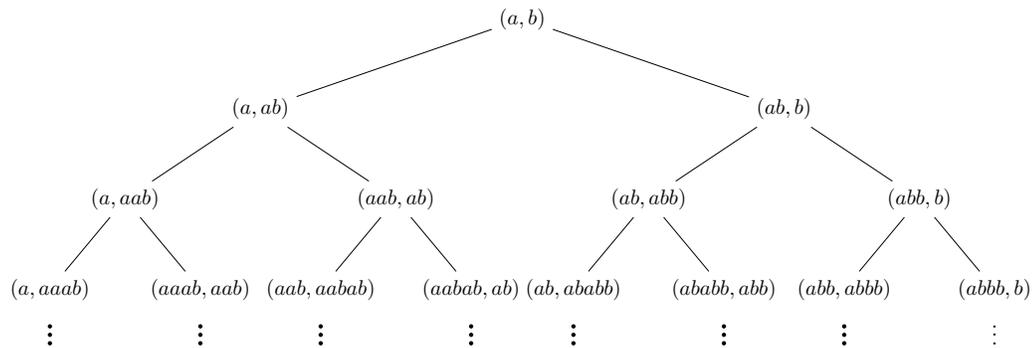


Figure 2.4: Arbre des paires de Christoffel

Une conséquence immédiate de la proposition est que les mots  $uuv$  et  $uvv$  sont des mots de Christoffel. On a donc une règle récursive qui est illustrée par la figure 2.3 pour construire les mots de Christoffel. L'*arbre des paires de Christoffel* est un arbre binaire infini dont la racine est étiquetée  $(a, b)$  et construit récursivement par la règle illustrée à la figure 2.3. L'arbre des paires de Christoffel est illustré à la figure 2.4. Cet arbre est pratique, car chacune des paires de Christoffel apparaît exactement une fois dans l'arbre (Borel et Laubie, 1993).

## 2.2 Arbres

L'arbre des paires de Christoffel n'est pas le seul étiquetage de l'arbre binaire infini complet permettant de construire les mots de Christoffel ou les fractions irréductibles correspondant à leur pente. Dans cette section, on présente quatre

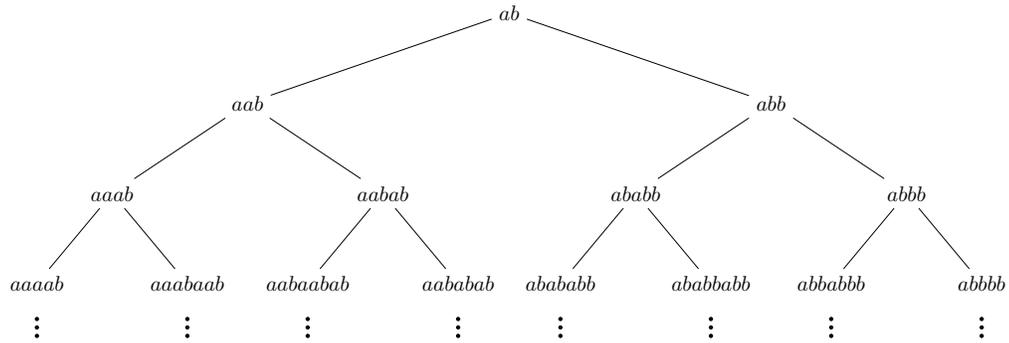


Figure 2.5: L'arbre des mots de Christoffel

autres étiquetages possibles ainsi que les relations entre les différents arbres ainsi obtenus. Ces arbres illustrent les liens nombreux et riches existant entre les mots de Christoffel et les fractions irréductibles.

Soit l'arbre binaire infini complet  $\mathcal{T}$ . Le chemin entre la racine et un sommet de l'arbre peut-être représenté par un mot  $w$  sur l'alphabet  $\{a, b\}$  où  $a$  représente un mouvement vers la gauche et  $b$  un mouvement vers la droite. Ainsi, un mot représente la suite des mouvements gauches et droits nécessaire pour atteindre un sommet à partir de la racine. Les sommets dans l'arbre sont donc identifiés de façon unique par les mots sur  $\{a, b\}^*$  puisqu'il y a un unique chemin entre la racine et chacun des sommets dans l'arbre  $\mathcal{T}$ . Par exemple, le sommet étiqueté par la paire de Christoffel  $(aab, aabab)$  dans l'arbre de la figure 2.4 est représenté par le mot  $aba$ .

L'*arbre des mots de Christoffel* est une variante de l'arbre des paires de Christoffel où la paire de Christoffel  $(u, v)$  est remplacée par le mot de Christoffel  $uv$  (voir figure 2.5). Par conséquent, chaque mot de Christoffel propre apparait une seule fois dans l'arbre des mots de Christoffel (Borel et Laubie, 1993). Le théorème suivant décrit explicitement la construction de l'arbre des mots de Christoffel.

**Théorème 2.13.** *Soit  $w$  un sommet dans l'arbre des mots de Christoffel et  $\ell$  le chemin entre le sommet  $w$  et la racine. Prenons  $u$  le sommet immédiatement après le premier pas nord-ouest sur le chemin  $\ell$  (s'il n'y a pas de pas nord-ouest, alors  $u = a$ ) et  $v$  le sommet immédiatement après le premier pas nord-est sur le chemin  $\ell$  (s'il n'y a pas de pas nord-est, alors  $v = b$ ). Alors le sommet  $w$  est étiqueté par le mot de Christoffel  $wv$ .*

Par exemple, le sommet  $aba$  est étiqueté par le mot  $aabaabab$  dont les sommets correspondant à  $u$  et  $v$  sont étiquetés par  $aab$  et  $aabab$  respectivement. La démonstration du théorème 2.13 se fait par récurrence en utilisant la règle illustrée à la figure 2.3 (Reutenauer, 2019).

Les mots de Christoffel sont étroitement liés aux fractions irréductibles par leur pente. En effet, chaque fraction réduite correspond à la pente d'un seul mot de Christoffel. En remplaçant les mots de Christoffel par leur pente dans l'arbre des mots de Christoffel, on construit un arbre où chaque fraction réduite apparaît une seule fois. Cet arbre illustré à la figure 2.6 est appelé l'*arbre de Stern-Brocot* (Stern, 1858; Brocot, 1861).

La construction directe de l'arbre de Stern-Brocot est analogue à la construction directe de l'arbre des mots de Christoffel. La racine de l'arbre de Stern-Brocot est la fraction réduite  $1/1$ . Soit  $\ell$  le chemin entre le sommet  $x$  et la racine. L'étiquette  $q/p$  du sommet  $x$  dans l'arbre de Stern-Brocot est  $(q' + q'')/(p' + p'')$  où  $q'/p'$  est l'étiquette du sommet immédiatement après le premier pas nord-ouest sur le chemin  $\ell$  (s'il n'y a pas de pas nord-ouest, alors  $q'/p' = 0/1$ ) et  $q''/p''$  est l'étiquette du sommet immédiatement après le premier pas nord-est sur le chemin  $\ell$  (s'il n'y a pas de pas nord-est, alors  $q''/p'' = 1/0$ ).

Il y a une autre façon d'étiqueter l'arbre binaire infini complet pour obtenir toutes les fractions irréductibles. On remplace la fraction  $q/p$  de l'arbre de Stern-Brocot

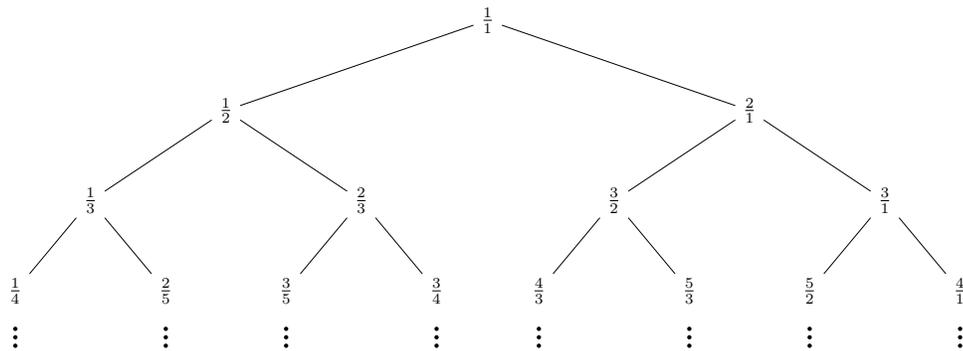


Figure 2.6: L'arbre de Stern-Brocot

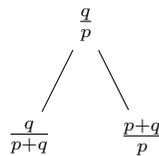


Figure 2.7: Règle récursive de construction de l'arbre de Raney

par la fraction  $q^*/p^*$  où  $p^*$  est l'inverse modulo  $p+q$  de  $p$  et  $q^*$  est l'inverse modulo  $p+q$  de  $q$  avec  $p^*, q^* \in \llbracket 0, p+q-1 \rrbracket$ . Cet arbre est appelé *arbre de Raney* (Raney, 1973) ou bien l'*arbre de Calkin-Wilf* (Calkin et Wilf, 2000). La racine de l'arbre de Raney est la fraction  $1/1$  et la règle récursive permettant de calculer les autres sommets est illustrée à la figure 2.7. On note qu'il est aussi possible de construire l'arbre de Raney directement à partir de l'arbre des mots de Christoffel : si  $w$  est un mot dans l'arbre des mots de Christoffel, alors on le remplace par la fraction  $|u|/|v|$  où  $(u, v)$  est la factorisation standard du mot  $w$ .

On peut aussi remplacer les fractions réduites  $q^*/p^*$  dans l'arbre de Raney par le mot de Christoffel de pente  $q^*/p^*$  tel qu'illustré à la figure 2.9. Cet arbre est appelé l'*arbre dual des mots de Christoffel*. Bien que cette construction n'est pas explicite dans la littérature, elle découle naturellement des mots de Christoffel

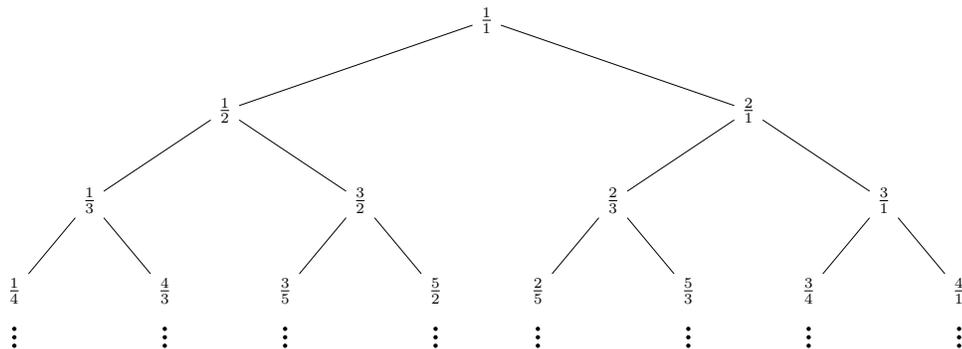


Figure 2.8: L'arbre de Raney

duals introduits par [Berthé et al. \(2008\)](#).

**Définition 2.14.** *Étant donné un mot de Christoffel  $w$  de pente  $q/p$ , on définit le mot de Christoffel dual  $w^*$  de pente  $q^*/p^*$  où  $p^*$  et  $q^*$  sont les inverses multiplicatifs modulo  $p + q$  de  $p$  et  $q$  dans  $\llbracket 0, p + q - 1 \rrbracket$ .*

On note que le dual d'un mot de Christoffel existe pour n'importe quel mot de Christoffel. En effet, les inverses de  $p$  et  $q$  existent, puisque  $p$  et  $q$  sont premiers entre eux. De plus,  $p$  et  $q$  sont tous deux premiers avec  $p + q$ . Donc, les entiers  $p^*$  et  $q^*$  sont aussi premiers entre eux. Par exemple, le dual du mot de Christoffel de pente  $4/7$ ,  $aabaabaabab$ , est le mot de Christoffel de pente  $3/8$ ,  $aaabaabaab$ .

### 2.3 Arbre dual des mots de Christoffel

Dans le chapitre 4, nous proposons une généralisation de l'arbre dual des mots de Christoffel. Pour bien introduire cette démarche, nous explorons dans cette section l'arbre dual des mots de Christoffel ainsi que sa construction directe à l'aide de morphismes.

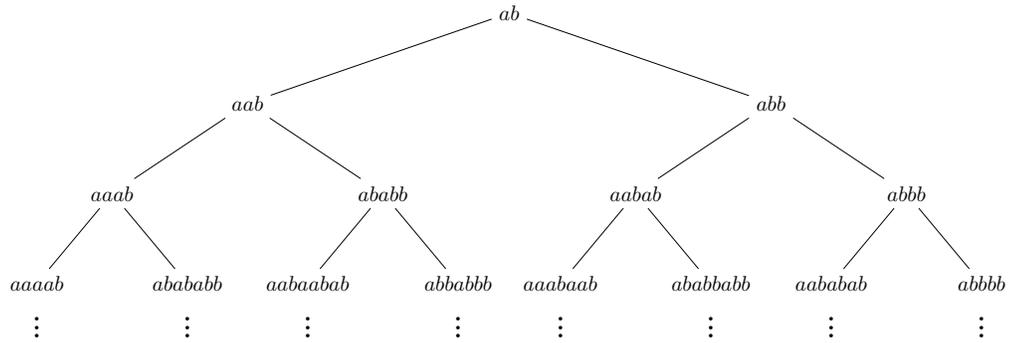


Figure 2.9: L'arbre dual des mots de Christoffel

Soit les endomorphismes  $\mathcal{G} = (a, ab)$  et  $\mathcal{D} = (ab, b)$ <sup>1</sup>. Ces endomorphismes préservent les mots de Christoffel.

**Proposition 2.15.** *L'image d'un mot de Christoffel par les endomorphismes  $\mathcal{G}$  et  $\mathcal{D}$  est un mot de Christoffel.*

**Exemple 2.16.** *L'image du mot de Christoffel  $ababbabb$  est*

$$\mathcal{G}(ababbabb) = aabaababaabab,$$

*le mot de Christoffel de pente  $5/8$ .*

La démonstration de cette proposition se fait par récurrence sur la longueur du mot  $w$  en utilisant la proposition 2.12 plutôt que la définition géométrique des mots de Christoffel. Les morphismes  $\mathcal{G}$  et  $\mathcal{D}$  peuvent être utilisés pour construire l'arbre dual des mots de Christoffel.

**Proposition 2.17.** *L'arbre dual des mots de Christoffel est construit directement par la règle récursive de la figure 2.10.*

---

1. Ces endomorphismes sont notés  $G$  et  $\tilde{D}$  dans Lothaire (2002).

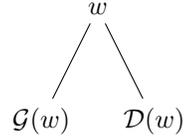


Figure 2.10: Règle de construction récursive de l'arbre dual des mots de Christoffel

Soit  $\mathcal{M}$  le plus petit ensemble contenant  $a$  et  $b$  qui est fermé par les morphismes  $\mathcal{G}$  et  $\mathcal{D}$ . Cet ensemble est l'ensemble des mots de Christoffel sur l'alphabet  $\{a, b\}$ . En effet, les morphismes  $\mathcal{G}$  et  $\mathcal{D}$  préservent les mots de Christoffel par la proposition 2.15. De plus, si  $w$  est un mot de Christoffel, alors il apparait dans l'arbre dual des mots de Christoffel. Par la proposition 2.17, il existe une composition de ces morphismes  $X_1, X_2, \dots, X_k$  telle que  $X_1 \circ X_2 \circ \dots \circ X_k(ab) = w$  et donc,  $w$  est dans  $\mathcal{M}$ .

Finalement, la généralisation des mots de Christoffel que nous étudions est basée sur la caractérisation suivante.

**Théorème 2.18** (Mantaci *et al.* (2003)). *Soit  $w \in \{a, b\}^*$ . Le mot  $w$  est conjugué à un mot de Christoffel si et seulement si  $\text{BW}(w) = b^q a^p$  avec  $\text{pgcd}(q, p) = 1$ .*

Par exemple, la transformée de Burrows-Wheeler du mot  $aaabaaabaab$  est  $b^3 a^8$  et c'est le mot de Christoffel de pente  $3/8$ . La forme de la matrice d'un mot de Christoffel est complètement décrite par la proposition suivante.

**Proposition 2.19** (Borel et Reutenauer (2006)). *Soit  $w = w_1 \prec w_2 \prec \dots \prec w_n$  les conjugués d'un mot de Christoffel  $w \in \{a < b\}^*$  de longueur  $n$ . Alors pour tout  $i \in \llbracket 2, n \rrbracket$ , on a  $w_{i-1} = uabv$  et  $w_i = ubav$  pour certains mots  $u$  et  $v$  tels que  $vu$  correspond le mot central de  $w$ .*

Par exemple, la matrice des conjugués du mot de Christoffel  $aabaabaabab$  est écrite à la figure 2.11 ainsi que le facteur  $ab$  à inverser pour obtenir la ligne suivante. La

$$\begin{array}{cccccccc}
a & a & b & a & a & b & a & \underline{a} & \underline{b} & a & b \\
a & a & b & a & \underline{a} & \underline{b} & a & \underline{b} & \underline{a} & a & b \\
a & \underline{a} & \underline{b} & a & \underline{b} & a & a & b & a & a & b \\
a & \underline{b} & \underline{a} & a & b & a & a & b & a & \underline{a} & \underline{b} \\
a & b & a & a & b & a & \underline{a} & \underline{b} & a & \underline{b} & \underline{a} \\
a & b & a & \underline{a} & \underline{b} & a & \underline{b} & \underline{a} & a & b & a \\
\underline{a} & \underline{b} & a & \underline{b} & a & a & b & a & a & b & a \\
\underline{b} & \underline{a} & a & b & a & a & b & a & a & \underline{a} & \underline{b} & a \\
b & a & a & b & a & \underline{a} & \underline{b} & a & \underline{b} & a & a \\
b & a & \underline{a} & \underline{b} & a & \underline{b} & \underline{a} & a & b & a & a \\
b & a & \underline{b} & \underline{a} & a & b & a & a & b & a & a
\end{array}$$

Figure 2.11: La matrice des conjugués du mot de Christoffel  $aabaabaabab$ . Les lettres soulignées sont celles à inverser pour obtenir la ligne suivante de la matrice.

ligne 3 contient le mot  $aababaabaab$  et la ligne 4 le mot  $abaabaabaab$ , alors  $u = a$  et  $v = abaabaab$ . On observe que  $vu = abaabaaba$  est le mot central du mot de Christoffel  $aabaabaabab$ .

## 2.4 Palindromes et mots sturmiens

Une autre caractérisation des mots de Christoffel est basée sur la palindromisation itérée.

**Théorème 2.20** (de Luca (1997)). *La fonction  $w \mapsto a\text{Pal}(w)b$  est une bijection du monoïde libre  $\{a, b\}^*$  vers l'ensemble des mots de Christoffel propres.*

Cette bijection implique entre autres que les mots centraux sont tous des palindromes itérés. Par exemple,  $a\text{Pal}(abab)b = a(abaababaaba)b$  est le mot de Christoffel de pente  $5/8$  que l'on a vu à l'exemple 2.4. La démonstration du théorème 2.20 utilise les observations suivantes sur les morphismes. Soit  $\zeta : \{a, b\}^* \rightarrow \text{End}(\{a, b\}^*)$  le morphisme de monoïde défini par  $\zeta(a) = \mathcal{G}$  et  $\zeta(b) = \mathcal{D}$ .

**Proposition 2.21** (Berstel et de Luca (1997)). *Soit  $(u, v)$  un sommet dans l'arbre des paires de Christoffel, représenté par le mot  $x \in \{a, b\}^*$ . Alors dans le monoïde des endomorphismes du monoïde libre  $\{a, b\}^*$ , on a l'égalité  $(u, v) = \zeta(x)$ .*

Par exemple, dans l'arbre des paires de Christoffel illustré à la figure 2.4, le sommet  $(aab, aabab)$  étiquette le sommet  $aba$  et  $\zeta(aba) = \mathcal{G} \circ \mathcal{D} \circ \mathcal{G} = (aab, aabab)$ , car  $\mathcal{G} \circ \mathcal{D} \circ \mathcal{G}(a) = \mathcal{G} \circ \mathcal{D}(a) = \mathcal{G}(ab) = aab$  et  $\mathcal{G} \circ \mathcal{D} \circ \mathcal{G}(b) = \mathcal{G} \circ \mathcal{D}(ab) = \mathcal{G}(abb) = aabab$ .

**Proposition 2.22** (Reutenauer (2019)). *Soit  $w \in \{a, b\}^*$ . Alors,  $\zeta(w)(ab) = a\text{Pal}(w)b$ . Par conséquent,  $a\text{Pal}(w)b$  est un mot de Christoffel.*

Un mot *sturmien* (fini) est un facteur d'un mot de Christoffel. Un mot binaire est sturmien si et seulement s'il est équilibré (Dulucq et Gouyou-Beauchamps, 1990).

**Remarque 2.23.** *Ce n'est pas la définition usuelle des mots sturmiens. Habituellement, les mots sturmiens sont définis comme les facteurs finis d'une suite sturmiennne, c'est-à-dire une suite infinie binaire  $s$  telle que  $|\text{Fact}_n(s)| = n + 1$  pour tout  $n \in \mathbb{N}$ . L'équivalence entre les deux notions est démontrée dans (Coven et Hedlund, 1973).*

## 2.5 Mots parfaitement amassants

Toutes les caractérisations des mots de Christoffel suggèrent une généralisation de cet ensemble de mots à un alphabet quelconque. Les mots parfaitement amassants, introduits par Simpson et Puglisi (2008), généralisent la caractérisation des mots de Christoffel basées sur la transformée de Burrows-Wheeler, c'est-à-dire celle faisant l'objet du théorème 2.18. Soit  $\mathcal{A} = \{a_1, \dots, a_r\}$  un alphabet totalement ordonné de cardinalité  $r$  où  $a_1 < a_2 < \dots < a_r$ .

**Définition 2.24.** *Un mot  $w$  appartenant à  $\mathcal{A}^*$  est parfaitement amassant<sup>2</sup> si*

$$\text{BW}(w) = a_r^{|w|_{a_r}} a_{r-1}^{|w|_{a_{r-1}}} \dots a_1^{|w|_{a_1}},$$

*c'est-à-dire que la transformée de Burrows-Wheeler regroupe toutes les occurrences d'une même lettre et ordonne les lettres en ordre décroissant.*

**Exemple 2.25.** *Les mots *ananas*, *garage* et *opposés* sont parfaitement amassants. De même, tous les conjugués et toutes les puissances des mots de Christoffel sont parfaitement amassants par le théorème 2.18 et la proposition 1.17.*

L'ensemble des mots parfaitement amassants est dénoté par  $\mathcal{P}$ . Tous les conjugués et toutes les puissances d'un mot parfaitement amassant sont aussi parfaitement amassants par les propositions 1.16 et 1.17. Nous étudions les mots de Lyndon parfaitement amassants complets puisqu'il est simple d'énumérer les conjugués et les puissances d'un mot. On définit alors  $\mathcal{P}_r$ , l'ensemble des mots de Lyndon parfaitement amassants complets sur l'alphabet  $\mathcal{A}$  de cardinalité  $r$ . Par exemple, l'ensemble  $\mathcal{P}_2$  est l'ensemble des mots de Christoffel. Puglisi et Simpson ont démontré que tous les mots de Lyndon parfaitement amassants sur un alphabet de cardinalité 3 sont des éléments de l'ensemble minimal contenant tous les mots de Christoffel et fermés par l'action de deux morphismes et d'une fonction (voir [Simpson et Puglisi \(2008\)](#)). Au chapitre 4, nous répondons à la question suivante : Comment engendrer les mots de Lyndon parfaitement amassants complets sur l'alphabet de taille  $|\mathcal{A}| \geq 2$  à l'aide de morphismes ?

Quelques propriétés des mots parfaitement amassants ont déjà été démontrées par d'autres auteurs. Voici leurs principales propriétés connues.

---

2. Parfaitement amassant est une traduction libre de la dénomination anglaise «perfectly clustering» introduite par [Ferenczi et Zamboni \(2013\)](#). Cette famille de mots est désignée par «words with simple Burrows-Wheeler transforms» par [Simpson et Puglisi \(2008\)](#).

**Proposition 2.26** (Simpson et Puglisi (2008)). *Soit  $w$  un mot parfaitement amassant et  $w_1 \prec w_2 \prec \dots \prec w_n$  les conjugués de  $w$  triés selon ordre lexicographique. Pour tout  $i \in \llbracket 1, n \rrbracket$ ,*

$$w_i = \widetilde{w_{n+1-i}}.$$

*En particulier, l'image miroir est une fonction décroissante sur l'ensemble des conjugués d'un mot parfaitement amassant.*

La démonstration de cette proposition est dans (Simpson et Puglisi, 2008). Cette proposition a plusieurs conséquences directes.

**Corollaire 2.27.** *La classe de conjugaison d'un mot parfaitement amassant est fermée par l'image miroir.*

**Corollaire 2.28.** *Soit  $w$  un mot parfaitement amassant. Le mot  $w$  est un produit de deux palindromes. De plus, cette factorisation est unique.*

Si  $w$  est un mot parfaitement amassant, alors

$$|\text{Fact}_n(w)| \leq (k - 1)r + 1$$

où  $k$  est la taille de l'alphabet du mot  $w$  (Simpson et Puglisi, 2008). De plus, si un mot  $w$  est parfaitement amassant, alors tous ses conjugués ont exactement  $|w| + 1$  facteurs qui sont des palindromes (Restivo et Rosone, 2009b). Par exemple, les facteurs du mot *ananas* qui sont des palindromes sont  $\varepsilon, a, n, s, ana, nan, anana$ .

De plus, deux caractérisations partielles des mots parfaitement amassants sont connues. La première a été démontrée par Restivo et Rosone (2009a, 2011). Si un mot  $w$  est équilibré, alors les trois propriétés suivantes sont équivalentes :  $w$  est parfaitement amassant, les conjugués de  $w$  ont exactement  $|w| + 1$  facteurs qui sont des palindromes et  $w$  est un conjugué d'une puissance d'un mot épisturmien fini

(les mots épisturmiens finis sont une généralisation des mots sturmiens basée sur le théorème 2.20). Comme les mots de Christoffel, ainsi que tous leurs conjugués, sont équilibrés les trois énoncés sont équivalents sur l'alphabet binaire. L'autre caractérisation est présentée à la fin de la section 3.1, car elle met en évidence un lien entre les mots parfaitement amassants et les échanges d'intervalles discrets symétriques.

## CHAPITRE III

### GÉNÉRALISATION DE L'ARBRE DE RANEY

Nous introduisons les échanges d'intervalles discrets symétriques et la bijection qui les lie aux mots parfaitement amassants tels que définis par [Ferenczi et Zamboni \(2013\)](#). Ensuite, nous présentons une fonction récursive permettant de compter le nombre de cycles d'un échange d'intervalles discrets symétrique. Cette fonction nous permet de généraliser l'arbre de Raney à des alphabets quelconques. Finalement, nous décrivons le type cyclique des échanges d'intervalles discrets symétriques de longueur 3. Les résultats de ce chapitre ont été publiés dans ([Lapointe, 2019](#)).

#### 3.1 Échanges d'intervalles discrets

La notion d'échange d'intervalles discrets a été introduite par [Ferenczi et Zamboni \(2013\)](#). C'est une version discrète de la notion d'échange d'intervalles ([Oseledec, 1966](#)). Un échange d'intervalles discrets est une fonction bijective de l'ensemble des entiers  $\llbracket 1, n \rrbracket$  vers lui-même obtenue en découpant en  $r$  intervalles (morceaux) de différentes longueurs l'ensemble  $\llbracket 1, n \rrbracket$  et en plaçant ces intervalles dans un nouvel ordre.

**Exemple 3.1.** *L'ensemble  $\llbracket 1, 6 \rrbracket$  peut être découpé en trois intervalles  $I_1 = \{1\}$ ,  $I_2 = \{2\}$  et  $I_3 = \{3, 4, 5, 6\}$  et réorganisé dans l'ordre  $I_{\sigma(3)}, I_{\sigma(1)}$  et  $I_{\sigma(2)}$ . L'échange*

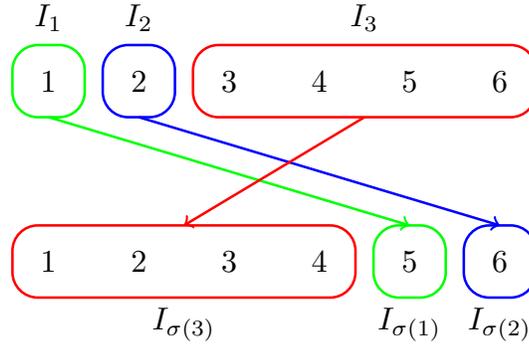


Figure 3.1: Un échange d'intervalles discrets de l'ensemble  $\llbracket 1, 6 \rrbracket$ .

d'intervalles discrets  $E$  construit de cette façon est illustré à la figure 3.1. Ainsi, on a que  $E(4) = 2$  et  $E(2) = 6$ .

Plus formellement, un échange d'intervalles discrets est défini de la façon suivante.

**Définition 3.2.** Soit  $\sigma$  une permutation de l'ensemble  $\llbracket 1, r \rrbracket$  et  $c = (c_1, \dots, c_r)$  une composition de  $n$ . Les intervalles  $I_i$  sont définis par

$$I_i = \left\{ x \mid \sum_{j < i} c_j < x \leq \sum_{j \leq i} c_j \right\}$$

pour tout  $i \in \llbracket 1, r \rrbracket$ . L'échange d'intervalles discrets  $E_{\sigma, c}$  est une fonction de l'ensemble  $\llbracket 1, n \rrbracket$  vers lui-même définie par

$$E_{\sigma, c}(x) = x + \sum_{\sigma^{-1}(j) < \sigma^{-1}(i)} c_j - \sum_{j < i} c_j, \quad (3.1)$$

où  $x \in I_i$ .

**Exemple 3.3.** L'exemple 3.1 est obtenu en prenant la permutation  $\sigma = 312$  et la

composition  $c = (1, 1, 4)$ . Par exemple, on remarque que

$$\begin{aligned}
 E_{312,(1,1,4)}(2) &= 2 + \sum_{\sigma^{-1}(j) < \sigma^{-1}(2)} c_j - \sum_{j < 2} c_j \\
 &= 2 + \sum_{\sigma^{-1}(j) < 3} c_j - 1 \\
 &= 2 + 5 - 1 \\
 &= 4,
 \end{aligned}$$

car  $2 \in I_2$ , puisque  $I_2 = \{2\}$ .

Un échange d'intervalles discrets  $E_{\sigma,c}$  est une permutation sur l'ensemble des entiers  $\llbracket 1, n \rrbracket$ , ainsi on utilise la notation en cycle des permutations pour le représenter.

Soit  $E$  une fonction sur un ensemble  $X$  partitionné en  $r$  ensembles  $I_i$ , où  $1 \leq i \leq r$ . La *trajectoire* de  $x$  sous l'action de  $E$  est la suite infinie  $(x_n)_{n \in \mathbb{N}}$  définie par  $x_n = a_k$  si  $E_{\sigma,c}^{(n)}(x) \in I_k$ , où  $1 \leq k \leq r$ . Un échange d'intervalles discrets est dit *minimal* si pour tout sous-ensemble fermé non vide  $Y$  de  $X$  tel que  $E^{-1}(Y) = Y$ , alors  $Y = X$ .

Dans le cas discret, la trajectoire d'un élément est une suite infinie périodique. Ainsi, la trajectoire d'un élément peut être décrite par des mots circulaires. Donc, un échange d'intervalles discrets  $E_{\sigma,c}$  est minimal, si  $E_{\sigma,c}$  est une permutation circulaire.

**Exemple 3.4.** En reprenant les valeurs de l'exemple 3.1, on a que l'échange d'intervalles discrets  $E_{312,(1,1,4)}$  est la permutation  $(1, 5, 3)(2, 6, 4)$ . Cet échange d'intervalles discrets a 2 cycles disjoints, alors il n'est pas minimal. La trajectoire de 3 sous l'action de  $E_{312,(1,1,4)}$  est  $(a_3 a_1 a_3)^\infty$ .

Un échange d'intervalle est dit *symétrique* si  $\sigma$  est une permutation symétrique.

L'expression décrivant la translation peut-être simplifiée pour les échanges d'intervalles discrets symétriques.

**Lemme 3.5.** *Soit  $c = (c_1, \dots, c_r)$  une composition de  $n$ . Les intervalles  $I_i$  sont définis par*

$$I_i = \left\{ x \mid \sum_{j < i} c_j < x \leq \sum_{j \leq i} c_j \right\}$$

pour tout  $i \in \llbracket 1, r \rrbracket$ . L'échange d'intervalles discrets symétrique  $E_c$  est une fonction de l'ensemble  $\llbracket 1, n \rrbracket$  vers lui-même définie par

$$E_c(x) = x + \sum_{j > i} c_j - \sum_{j < i} c_j, \quad (3.2)$$

où  $x \in I_i$ .

*Démonstration.* Prenons  $\sigma$  est la permutation symétrique dans la définition 3.2. De plus, on a que  $\sigma^{-1}(j) < \sigma^{-1}(i)$  si et seulement si  $j > i$  pour tout  $i, j \in \llbracket 1, r \rrbracket$ . ■

On note qu'un échange d'intervalles discrets symétrique agit sur ces éléments en leur appliquant une translation déterminée par  $\sum_{j > i} c_j - \sum_{j < i} c_j$ . Le vecteur de translation de  $E_c$  est  $t = (t_i)_{1 \leq i \leq r}$  où  $t_i = \sum_{j > i} c_j - \sum_{j < i} c_j$  pour tout  $i \in \llbracket 1, r \rrbracket$ .

**Exemple 3.6.** *Le mot circulaire qui correspond à l'échange d'intervalles discrets symétrique  $E_{(3,1,2,6)}$  est*

$$(1, 10, 4, 9, 3, 12, 6, 8, 2, 11, 5, 7).$$

Ces intervalles sont  $I_1 = \{1, 2, 3\}$ ,  $I_2 = \{4\}$ ,  $I_3 = \{5, 6\}$  et  $I_4 = \{7, 8, 9, 10, 11, 12\}$ . Son vecteur de translation est

$$t = (1 + 2 + 6, 2 + 6 - 3, 6 - 3 - 1, -3 - 1 - 2) = (9, 5, 2, -6).$$

On observe que dans l'exemple précédent le vecteur de translation de  $E_{(3,1,2,6)}$  est strictement décroissant, ce n'est pas une coïncidence.

**Lemme 3.7.** *Le vecteur de translation d'un échange d'intervalles discrets symétrique est strictement décroissant. De plus, si  $r > 1$ , il existe un entier  $i \in \llbracket 1, r \rrbracket$  tel que  $t_i$  est positif ou nul et  $t_{i+1}$  est négatif.*

*Démonstration.* Soit  $t$  le vecteur de translation d'un échange d'intervalles discrets symétrique  $E_c$ . Soit  $i$  et  $k$  deux entiers dans l'ensemble  $\llbracket 1, r \rrbracket$  tel que  $i > k$ . On a alors que  $\sum_{j>i} c_j < \sum_{j>k} c_j$  et  $\sum_{j<i} c_j > \sum_{j<k} c_j$ . Par conséquent,  $t_i = \sum_{j>i} c_j - \sum_{j<i} c_j < \sum_{j>k} c_j - \sum_{j<k} c_j = t_k$ . Donc, le vecteur de translation est strictement décroissant.

Supposons que  $r > 1$ . Le premier terme du vecteur de translation est  $t_1 = \sum_{j>1} c_j = \sum_{j<1} c_j - \sum_{j>1} c_j > 0$ , car les  $c_j$  sont non nuls. Le dernier terme du vecteur de translation est  $t_r = \sum_{j>r} c_j - \sum_{j<r} c_j = -\sum_{j<r} c_j < 0$ , car les  $c_j$  sont non nuls. Puisque le vecteur de translation est strictement décroissant, il existe un  $i$  tel que  $t_i$  est positif ou nul et  $t_{i+1}$  est négatif tel que souhaité. ■

Ici, nous présentons une version restreinte du Théorème 4 de (Ferenczi et Zamboni, 2013) établissant un lien entre les mots parfaitement amassants et les facteurs des trajectoires des échanges d'intervalles discrets symétriques.

**Théorème 3.8** (Ferenczi et Zamboni (2013)). *Un mot primitif complet  $w$  sur l'alphabet  $\mathcal{A} = \{a_1, \dots, a_r\}$  est parfaitement amassant si et seulement si  $ww$  est un facteur de la trajectoire d'un élément de l'échange d'intervalles discrets symétrique minimal  $E_c$ , où  $c$  correspond à l'image commutative de  $w$ .*

La démonstration que nous présentons ici est la même que celle de (Ferenczi et Zamboni, 2013). Elle repose sur les lemmes suivants. La *permutation standard*

d'un mot  $w$  sur un alphabet totalement ordonné notée  $\text{st}(w)$ . Si  $w = w_1 \dots w_n$ , c'est la permutation  $\sigma = \text{st}(w)$  telle que les énoncés suivants sont équivalents :

1.  $\sigma(i) < \sigma(j)$  ;
2.  $w_i < w_j$  ou  $w_i = w_j$  et  $i < j$ .

La permutation standard d'un mot est bien définie, car le deuxième énoncé définit un ordre total sur les lettres de  $w$ . La permutation standard d'un mot  $w$  s'obtient aussi en numérotant les lettres de  $w$  de 1 à  $n$  en commençant par la plus petite lettre et en numérotant ses occurrences de gauche à droite, ensuite on numérote les occurrences de la deuxième plus petite et ainsi de suite. Par exemple, la permutation standard du mot  $aacabac$  est  $\text{st}(aacabac) = 1263547$ . Soit  $g$  la fonction définie par  $\text{st}(a_1^{|w|_{a_1}} \dots a_n^{|w|_{a_n}})_i \mapsto \text{st}(\text{BW}(w))_i$ . Par exemple, la fonction  $g$  sur le mot  $aacabac$  est  $1234567 \mapsto 6751234$ .

**Lemme 3.9** (Ferenczi et Zamboni (2013)). *Si  $w$  est parfaitement amassant complet, alors  $g$  est l'inverse de l'échange d'intervalles discrets symétrique engendré par la composition correspondant à l'image commutative de  $w$ .*

*Démonstration.* Soit  $w$  un mot sur l'alphabet  $\{a_1, \dots, a_r\}$  et  $c = (c_1, \dots, c_r)$  son image commutative. Soit  $u$  le mot représentant la première colonne de la matrice des conjugués de  $w$ . On a alors que  $\text{st}(u)$  est le mot

$$1, 2, \dots, n.$$

Regardons la dernière colonne de la matrice des conjugués. On a que  $\text{BW}(w) = a_n^{c_n} a_{n-1}^{c_{n-1}} \dots a_1^{c_1}$ . Par conséquent, on a que

$$\begin{aligned} \text{st}(\text{BW}(w))_i &= |w| - \sum_{\ell \geq a_j} |w|_\ell + i - \sum_{\ell > a_j} |w|_\ell \\ &= i + \sum_{\ell < a_j} |w|_\ell - \sum_{\ell > a_j} |w|_\ell \end{aligned}$$

où  $a_j = \text{BW}(w)_i$ . La domaine de la fonction  $g$  est l'ensemble  $\{1 \dots n\}$ . Les intervalles  $I_j$  sont définis par les différentes lettres. L'entier  $x$  appartient à un certain intervalle  $I_j$  avec  $j \in \llbracket 1, r \rrbracket$ . L'ordre de la dernière colonne signifie que  $g(x) = x - \sum_{k>j} |w|_{a_k} + \sum_{k<j} |w|_{a_k}$ . Ainsi,  $g$  est bien l'inverse de l'échange d'intervalles discrets symétrique engendré par l'image commutative de  $w$ . ■

**Lemme 3.10** (Ferenczi et Zamboni (2013)). *Si l'échange d'intervalles discrets symétrique  $E_{(c_1, \dots, c_r)}$  n'est pas minimal, alors le mot  $r^{c_r} \dots 1^{c_1}$  n'a pas d'antécédent primitif par la transformée de Burrows-Wheeler.*

*Démonstration.* Soit  $w$  un tel antécédent. Par le lemme 3.9, la fonction  $g^{-1}$  correspond à  $E_{(c_1, \dots, c_r)}$ . Supposons que  $E_{(c_1, \dots, c_r)}$  ne soit pas minimal et donc qu'il existe un sous-ensemble propre  $A$  de  $\{1 \dots n\}$  qui est invariant sous l'action de  $g$ . Ainsi, dans les conjugués de  $w$ , précédant chacune des occurrences d'une lettre de  $A$ , il y a une lettre de  $A$ . Par conséquent,  $w$  ne contient que des lettres de  $A$ , contredisant le fait que  $w$  est complet. ■

*Démonstration du théorème 3.8.* Par hypothèse, il existe un  $x \in \llbracket 1, n \rrbracket$  dont la trajectoire initiale est le mot  $ww$ . Posons  $E = \{E_c(x), E_c^2(x), \dots, E_c^n(x)\}$ . Ainsi pour tout  $y \in E$ , la trajectoire initiale de  $y$  de longueur  $n$ , notée  $O(y)$  est un conjugué de  $w$ . Supposons que  $y, z \in E$  avec  $y < z$ . Soit  $j$  le plus petit entier naturel tel que  $E_c^j(y)$  et  $E_c^j(z)$  ne sont pas dans le même intervalle. Alors,  $E_c^j(y)$  est à la gauche de  $E_c^j(z)$ , car  $j = 0$  et  $E_c$  est croissante sur les intervalles. Par conséquent,  $O(y) \prec O(z)$ . Le mot  $\text{BW}(w)$  est obtenu des dernières lettres, notées  $O(y)_n$  de  $O(y)$  où les  $y$  sont ordonnés de droite à gauche. De plus,  $O(y)_n$  est associé à l'intervalle  $I_i$  de  $E_c^{n-1}(y) = E_c^{-1}(y)$ . Ainsi, par la définition de  $E$ , si  $y$  est à la gauche de  $z$ , alors  $O(y)_n \succeq O(z)_n$ , et si  $y'$  est entre  $y$  et  $z$  avec  $O(y)_n = O(z)_n$ , alors  $O(y')_n = O(y)_n = O(z)_n$ . Ainsi, le mot  $w$  est parfaitement amassant.

Soit  $w$  un mot primitif complet parfaitement amassant, la transformée de Burrows-Wheeler de  $w$  est  $r^{n_r} \dots 1^{n_1}$ . Le résultat de  $g$  est un échange d'intervalles discrets symétrique par le lemme 3.9. De plus,  $g$  est minimal par le lemme 3.10 et  $g$  a une trajectoire périodique  $w'w' \dots$  telle que l'image commutative de  $w'$  est  $(n_1, \dots, n_r)$ . Si  $w' = u^k$ , alors les  $n_i$  sont tous divisible par  $k$  et l'ensemble comprenant les  $|u|$  points les plus à gauche est invariant. Par conséquent,  $w'$  est primitif et donc parfaitement amassant. Puisque  $\text{BW}(w) = \text{BW}(w')$  par la proposition 1.16, les mots  $w$  et  $w'$  sont conjugués. ■

L'image commutative des mots parfaitement amassants correspond à une composition qui engendre les échanges d'intervalles discrets symétriques minimaux par le théorème 3.8. Les compositions engendrant un échange d'intervalles discrets symétrique minimal sont appelées *compositions circulaires*. Quelles sont les propriétés des compositions circulaires? Une réponse partielle à cette question est donnée dans (Pak et Redlich, 2008). En effet, une composition de la forme  $(c_1, c_2)$  est circulaire si et seulement si  $\text{pgcd}(c_1, c_2) = 1$ . De plus, une composition de la forme  $(c_1, c_2, c_3)$  est circulaire si et seulement si  $\text{pgcd}(c_1 + c_2, c_2 + c_3) = 1$  (Pak et Redlich, 2008). En revanche, Karnauhova et Liebscher (2017) ont démontré qu'il n'existe pas de polynôme homogène à coefficients entiers  $f_1, f_2$  tels que le  $\text{pgcd}(f_1(c_1, \dots, c_r), f_2(c_1, \dots, c_r))$  compte le nombre de cycles de l'échange d'intervalles discrets symétrique engendré par cette composition si  $r \geq 4$ . Dans la section suivante, nous considérons une fonction récursive permettant de compter ces cycles.

### 3.2 Dénombrement des cycles

Définissons la fonction  $\mathcal{Z}$  de l'ensemble des compositions vers  $\mathbb{N}$  de la façon suivante :

$$\mathcal{Z}(c_1, \dots, c_r) = \begin{cases} c_i + \mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) & \text{si } |t_i| = 0, \\ \mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) & \text{si } c_i = |t_i|, \\ \mathcal{Z}(c_1, \dots, c_i - |t_i|, \dots, c_r) & \text{sinon,} \end{cases} \quad (3.3)$$

où  $i$  est le plus petit entier satisfaisant la relation  $c_i \geq |t_i|$  et  $t_i = \sum_{j>i} c_j - \sum_{j<i} c_j$ . Le lemme suivant nous garantit qu'il existe un entier  $i$  tel que  $c_i \geq |t_i|$ .

**Lemme 3.11.** *Soit  $c = (c_1, \dots, c_r)$  une composition de  $n$  et  $t$  le vecteur de translation de l'échange d'intervalles discrets symétrique  $E_c$ . Alors, il existe un entier  $i \in \llbracket 1, r \rrbracket$  tel que  $c_i \geq |t_i|$ .*

**Remarque 3.12.** *Il peut y avoir plusieurs valeurs  $i$  telles que  $c_i \geq |t_i|$ . Par exemple, la relation est satisfaite pour  $i = 2$  et  $i = 3$  pour la composition  $(2, 1, 2, 1)$ . Pour que  $\mathcal{Z}$  soit une fonction, nous prenons systématiquement la plus petite valeur qui vérifie l'inégalité.*

*Démonstration.* Il existe un entier  $i$  satisfaisant les deux relations suivantes :

$$\begin{aligned} \sum_{j<i} c_j &\leq \sum_{j>i} c_j + c_i \\ \sum_{j<i} c_j + c_i &\geq \sum_{j>i} c_j, \end{aligned}$$

par le lemme 3.7. Isolons  $c_i$  dans les deux inégalités :

$$\begin{aligned} \sum_{j<i} c_j - \sum_{j>i} c_j &\leq c_i \\ c_i &\geq \sum_{j>i} c_j - \sum_{j<i} c_j. \end{aligned}$$

En les combinant, on a donc que

$$c_i \geq \left| \sum_{j>i} c_j - \sum_{j<i} c_j \right| = |t_i|,$$

par la définition du vecteur de translation et de la fonction valeur absolue. ■

La relation  $(c_1, \dots, c_r) \preceq (c'_1, \dots, c'_s)$  si  $r < s$  ou  $r = s$  et  $(c_1, \dots, c_r) \preceq (c'_1, \dots, c'_s)$  est un ordre total sur les compositions.

**Lemme 3.13.** *La fonction  $\mathcal{Z}$  est bien définie sur l'ensemble des compositions.*

*Démonstration.* Par le lemme 3.11, il existe un entier  $i \in \llbracket 1, r \rrbracket$  tel que  $c_i \geq |t_i|$  pour n'importe quelle composition  $c$ . De plus,  $(c_1, \dots, c_i - |t_i|, \dots, c_r)$  et  $(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r)$  sont des compositions, car  $c_i - |t_i| > 0$ . De plus, la fonction  $\mathcal{Z}$  termine, car l'étape récursive de la composition  $(c_1, \dots, c_r)$  plus petite que la composition elle-même, c'est-à-dire que  $(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) \triangleleft (c_1, \dots, c_i - |t_i|, \dots, c_r) \triangleleft (c_1, \dots, c_r)$ . Donc, la fonction  $\mathcal{Z}$  est bien définie. ■

Il reste à démontrer que cette fonction compte bel et bien le nombre de cycles de l'échange d'intervalles discrets symétrique engendré par la composition  $c$ .

**Exemple 3.14.** *L'échange d'intervalles discrets symétrique  $E_{(3,5,4,2)}$  a 3 cycles, car  $E_{(3,5,4,2)} = (1, 12, 6, 9, 3, 14, 2, 13)(4, 7, 10)(5, 8, 11)$ . L'image de la composition  $(3, 5, 4, 2)$  par la fonction  $\mathcal{Z}$  est*

$$\begin{aligned} \mathcal{Z}(3, 5, 4, 2) &= \mathcal{Z}(3, 2, 4, 2) = \mathcal{Z}(3, 2, 1, 2) \\ &= 2 + \mathcal{Z}(3, 1, 2) \\ &= 2 + \mathcal{Z}(1, 2) = 2 + \mathcal{Z}(1, 1) \\ &= 2 + \mathcal{Z}(1) \\ &= 3. \end{aligned}$$

Tableau 3.1: La première ligne contient les compositions aux étapes successives dans le calcul de  $\mathcal{Z}(3, 5, 4, 2)$ , la deuxième ligne représente l'indice  $i$  tel que  $c_i \geq |t_i|$  et la dernière ligne donne la valeur absolue du  $i^{\text{ième}}$  élément du vecteur de translation.

$c$	(3, 5, 4, 2)	(3, 2, 4, 2)	(3, 2, 1, 2)	(3, 1, 2)	(1, 2)	(1, 1)	(1)
$i$	2	3	2	1	2	1	
$ t_i $	3	3	0	3	1	1	

Les valeurs de  $i$  et de l'élément du vecteur de translation correspondant à chacune des étapes sont présentées dans le tableau 3.1.

Une fonction analogue a déjà été étudiée par [Karnauhova et Liebscher \(2017\)](#) sur un objet géométrique appelé méandre arc-en-ciel<sup>1</sup>. Ils ont démontré que cette fonction compte le nombre de cycles des échanges d'intervalles discrets symétriques. La démonstration présentée ici est différente, car elle est combinatoire au lieu de géométrique.

**Lemme 3.15.** *Soit  $c$  une composition. Si  $t_i = 0$ , alors l'échange d'intervalles discrets symétrique  $E_c$  a*

$$c_i + \gamma(E_{(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r)})$$

*cycles.*

**Exemple 3.16.** *L'échange d'intervalles discrets symétrique  $E_{(3,2,4,5)}$  a 5 cycles et  $t_3 = 0$ . De plus,  $E_{(3,2,5)}$  a un cycle, alors  $\gamma(E_{(3,2,4,5)}) = 5 = 4 + \gamma(E_{(3,2,5)})$ .*

---

1. bi-rainbow meander

*Démonstration.* Soit  $c = (c_1, \dots, c_r)$  une composition et  $i \in \llbracket 1, r \rrbracket$  tel que  $t_i = 0$ . Soit  $x \in \llbracket 1, n \rrbracket$ . L'échange d'intervalles discrets  $E_c$  est

$$E_c(x) = x + t_k \quad (3.4)$$

où  $x$  est un élément de  $I_k = \{x \mid \sum_{j < k} c_j < x \leq \sum_{j < k+1} c_j\}$  et  $t_i = \sum_{j > k} c_j - \sum_{j < k} c_j$  par le lemme 3.5.

Si  $x \in I_i$ , l'équation 3.4 devient  $E_c(x) = x$ , car  $t_i = 0$ . Par conséquent, chaque entier de l'ensemble  $I_i$  est un point fixe de  $E_c$  et son orbite contient un seul élément. Ainsi,  $E_c|_{I_i}$  est l'identité et donc  $E$  stabilise l'ensemble  $H = \llbracket 1, n \rrbracket \setminus I_i$ .

Il reste à démontrer que  $E_c|_H$  est conjugué à  $E_{c'}$ , c'est-à-dire que  $E_c|_H = \omega^{-1} \circ E_{c'} \circ \omega$ , où  $\omega$  est l'unique bijection décroissante telle que  $\omega(H) = \llbracket 1, n - c_i \rrbracket$ . Cette affirmation implique que  $E_c|_H$  et  $E_{c'}$  ont le même nombre de cycles.

Il ne reste que le cas  $x \notin I_i$ . La composition  $c' = (c'_1, \dots, c'_{r-1})$  est définie par  $c'_j = c_j$  si  $j < i$  et  $c'_j = c_{j+1}$  autrement. Observons que  $E_{c'}$  est un échange d'intervalles discrets symétrique avec les intervalles  $I'_1, \dots, I'_{r-1} \subseteq \llbracket 1, n - c_i \rrbracket$  définis par  $I'_m = \{x \mid \sum_{j < m} c'_j < x \leq \sum_{j < m+1} c'_j\}$ . Pour tout  $x \in I'_m$ ,  $E_{c'}(x) = x + t'_m$  où  $t'_m$  est le paramètre de translation associé à l'intervalle  $I'_m$ . On note que  $E_{c'} = E_{(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r)}$ .

D'abord on décrit les ensembles  $I_m$  à l'aide des ensembles  $I'_m$ . Il y a deux cas. Si  $m < i$ , alors

$$I_m = \left\{ x \mid \sum_{j < m} c_j < x \leq \sum_{j < m+1} c_j \right\} = \left\{ x \mid \sum_{j < m} c'_j < x \leq \sum_{j < m+1} c'_j \right\} = I'_m$$

puisque pour  $j \leq m < i$ , nous avons que  $c_j = c'_j$ .

Sinon, on a que  $m > i$ , alors

$$\begin{aligned}
I_m &= \left\{ x \mid \sum_{j < m} c_j < x \leq \sum_{j < m+1} c_j \right\} \\
&= \left\{ x \mid \sum_{j < m-1} c'_j + c_i < x \leq \sum_{j < m} c'_j + c_i \right\} \\
&= \{x + c_i \mid x \in I'_{m-1}\},
\end{aligned}$$

puisque  $m > i$  et que  $\sum_{j < m} c_j = \sum_{j < i} c_j + c_i + \sum_{i < j < m} c_j = c_i + \sum_{j < m-1} c'_j$ .

Ensuite, on veut montrer que  $E_c(x) = E_{c'}(x) + c_i$  ou  $E_c(x) = E_{c'}(x - c_i)$  pour tout  $x \notin I_i$ . Mettons en évidence  $c_i$  dans la définition de  $E_c$  :

$$E_c(x) = \begin{cases} x + \sum_{j > i} c_j + \sum_{i > j > m} c_j - \sum_{j < m} c_j + c_i & \text{si } m < i \\ x + \sum_{j > m} c_j - \sum_{j < i} c_j - \sum_{i < j < m} c_j - c_i & \text{si } m > i \end{cases}$$

En remplaçant  $c_j$  par  $c'_j$  si  $j < i$  et par  $c'_{j-1}$  autrement, on obtient

$$\begin{aligned}
E_c(x) &= \begin{cases} x + \sum_{j > i-1} c'_j + \sum_{i > j > m} c'_j - \sum_{j < m} c'_j + c_i & \text{si } m < i \\ x + \sum_{j > m-1} c'_j - \sum_{j < i} c'_j - \sum_{i < j < m-1} c'_j - c_i & \text{si } m > i \end{cases} \\
&= \begin{cases} x + \sum_{j > m} c'_j - \sum_{j < m} c'_j + c_i & \text{si } m < i \\ x + \sum_{j > m-1} c'_j - \sum_{j < m-1} c'_j - c_i & \text{si } m > i \end{cases} \\
&= \begin{cases} x + t'_m + c_i & \text{si } m < i \\ x + t'_{m-1} - c_i & \text{si } m > i \end{cases} \\
&= \begin{cases} E_{c'}(x) + c_i & \text{si } m < i \\ E_{c'}(x - c_i) & \text{si } m > i. \end{cases}
\end{aligned}$$

Par conséquent,  $E_c|_H$  est conjugué à  $E_{c'}$  et le nombre de cycles de  $E_c$  est  $c_i + \gamma(E_{(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r)})$ . ■

Les cycles de l'échange d'intervalles discrets symétrique engendré par la composition faible  $c$  sont les mêmes que les cycles de celui engendré par la composition faible où les éléments nuls ont été retirés, car les intervalles correspondant aux éléments nuls sont vides.

**Lemme 3.17.** *Soit  $c$  une composition faible. Les échanges d'intervalles discrets symétriques  $E_c$  et  $E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)}$  ont le même nombre de cycles.*

**Exemple 3.18.** *Les compositions  $(5, 1, 2)$  et  $(5, 4, 2)$  ont la forme souhaitée par le lemme 3.17, c'est-à-dire que  $(5, 4, 2) = (5, 1 + |t_2|, 2)$ . De plus, les échanges d'intervalles discrets symétriques  $E_{(5,1,2)}$  et  $E_{(5,4,2)}$  ont tous deux 3 cycles. De même, les compositions faibles  $(5, 0, 1, 2)$  et  $(5, 2, 1, 2)$  ont la forme souhaitée, c'est-à-dire que  $(5, 2, 1, 2) = (5, 0 + |t_2|, 1, 2)$  et ils engendrent toutes deux des échanges d'intervalles discrets symétriques ayant trois cycles.*

Avant de démontrer le lemme 3.17, décrivons les cycles de  $E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)}$  à partir des cycles de  $E_c$ . En utilisant la notation cyclique d'une permutation, on peut représenter un échange d'intervalle discrets symétrique par un ensemble de mots circulaires sur l'alphabet  $\llbracket 1, n \rrbracket$ . Exprimons l'ensemble des mots circulaires de  $E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)}$  à l'aide d'une substitution  $\psi_i$  sur l'ensemble des mots circulaires de  $E_c$  définie par

$$\psi_i(x) = \begin{cases} x & \text{si } x \in \llbracket 1, m \rrbracket, \\ x \cdot (x + |t_i|) & \text{si } x \in \llbracket m + 1, m + |t_i| \rrbracket \text{ et } t_i \geq 0, \\ (x + |t_i|) \cdot x & \text{si } x \in \llbracket m + 1, m + |t_i| \rrbracket \text{ et } t_i < 0, \\ x + |t_i| & \text{sinon} \end{cases} \quad (3.5)$$

où  $x \in \llbracket 1, n \rrbracket$  est une lettre,  $\cdot$  est la concaténation usuelle sur les mots et

$$m = \begin{cases} \sum_{j \leq i} c_j & \text{si } t_i \geq 0, \\ \sum_{j > i} c_j & \text{si } t_i < 0. \end{cases}$$

**Exemple 3.19.** Les cycles de l'échange d'intervalles discrets symétrique  $E_{(5,1,2)}$  sont  $(1, 4, 7)(2, 5, 8)(3, 6)$ . Les cycles de l'échange d'intervalles discrets symétrique  $E_{(5,4,2)}$  sont  $(1, 7, 4, 10)(2, 8, 5, 11)(3, 9, 6)$ . Appliquons  $\psi_2$  à  $E_{(5,1,2)}$ . La substitution est

$$\psi_2(x) = \begin{cases} x & \text{si } x \in \llbracket 1, 2 \rrbracket \\ (x+3) \cdot x & \text{si } x \in \llbracket 3, 5 \rrbracket \\ x+3 & \text{sinon} \end{cases}$$

puisque  $t_2 = -3 < 0$  et  $k = 2$ . Par conséquent,  $\psi_2(E_{(5,1,2)}) = \psi_2((1, 4, 7)(2, 5, 8)(3, 6)) = (1, 7, 4, 10)(2, 8, 5, 11)(6, 3, 9)$ .

**Lemme 3.20.** Soit  $i$  un entier dans  $\llbracket 1, r \rrbracket$ . Alors

$$\psi_i(E_c) = E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)} \quad (3.6)$$

sur les ensembles de mots circulaires où  $c = (c_1, \dots, c_r)$ .

Dans cette situation,  $\psi_i$  agit comme une substitution sur les cycles de  $E_c$ . Rappelons que les facteurs circulaires de longueur 2 d'un mot circulaire  $(w_1 \dots w_n)$  sont les mots  $w_i w_{(i+1) \bmod n}$  avec  $i \in \llbracket 1, n \rrbracket$ . L'observation suivante est nécessaire à la démonstration du lemme 3.20 : deux mots circulaires où chacune des lettres a une seule occurrence sont égaux si et seulement s'ils ont les mêmes facteurs circulaires de longueur 2.

*Démonstration.* Tout d'abord, supposons que  $t_k = 0$ . Alors,  $\psi_k(E_c) = E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)}$  puisque  $\psi_k$  est le morphisme identité lorsque  $t_k = 0$ .

Soit  $k$  un entier dans l'ensemble  $\llbracket 1, r \rrbracket$  et  $E_c$  un échange d'intervalles discrets symétrique défini comme au lemme 3.5. Soit  $c' = (c'_1, \dots, c'_r)$  une composition où  $c'_k = c_k + t_k$  et  $c'_i = c_i$  si  $i \neq k$ . Rappelons que  $E_{c'}$  est un échange d'intervalles discrets symétrique où  $I'_1, \dots, I'_r$  sont des sous-intervalles de  $\llbracket 1, n + |t_k| \rrbracket$  définis par  $I'_i = \llbracket 1 + \sum_{j < i} c'_j, c'_i + \sum_{j < i} c'_j \rrbracket$  et ses paramètres de translation sont  $t'_i = \sum_{j > i} c'_j - \sum_{j < i} c'_j$ . Pour tout  $x \in I'_i$ ,  $E_{c'}(x) = x + t'_i$ . Aussi,  $E_{c'} = E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)}$ .

Commençons par décrire les facteurs circulaires de longueur 2 de  $E_{c'}$  en utilisant  $c$  et  $I_i$  plutôt que  $c'$  et  $I'_i$ . Pour décrire  $I'_i$  en utilisant  $I_i$ , il y a trois cas :

1. Si  $i < k$ , alors

$$I'_i = \left[ \left[ 1 + \sum_{j < i} c'_j, \sum_{j \leq i} c'_j \right] \right] = \left[ \left[ 1 + \sum_{j < i} c_j, \sum_{j \leq i} c_j \right] \right] = I_i$$

car  $j \leq i < k$  implique que  $c'_j = c_j$  ;

2. Si  $i = k$ , alors

$$\begin{aligned} I'_k &= \left[ \left[ 1 + \sum_{j < k} c'_j, \sum_{j \leq k} c'_j \right] \right] \\ &= \left[ \left[ 1 + \sum_{j < k} c_j, |t_k| + \sum_{j \leq k} c_j \right] \right] \\ &= I_k \cup \left[ \left[ 1 + \sum_{j \leq k} c_j, |t_k| + \sum_{j \leq k} c_j \right] \right] \end{aligned}$$

car  $i = k$  implique que  $c'_j = c_j$  si  $j \neq k$  et  $c'_k = c_k + |t_k|$  ;

3. Si  $i > k$ , alors

$$\begin{aligned} I'_i &= \left[ \left[ 1 + \sum_{j < i} c'_j, \sum_{j \leq i} c'_j \right] \right] \\ &= \left[ \left[ 1 + |t_k| + \sum_{j < i} c_j, |t_k| + \sum_{j \leq i} c_j \right] \right] \\ &= \{x + |t_k| \mid x \in I_i\}. \end{aligned}$$

On a que  $E_{c'}(x) = x + \sum_{j>i} c'_j - \sum_{j<i} c'_j$ . En isolant  $c'_k$  dans cette expression, on obtient

$$E_{c'}(x) = \begin{cases} x + \sum_{j=i+1}^{k-1} c'_j + \sum_{j>k} c'_j - \sum_{j<i} c'_j + c'_k, & \text{si } i < k; \\ x + \sum_{j>i} c'_j - \sum_{j<i} c'_j, & \text{si } i = k; \\ x + \sum_{j>i} c'_j - \sum_{j<k} c'_j - \sum_{j=k+1}^{i-1} c'_j - c'_k, & \text{si } i > k. \end{cases}$$

En remplaçant,  $c'_i$  par  $c_i$  si  $i \neq k$  et par  $c_k + |t_k|$ , on a que

$$E_{c'}(x) = \begin{cases} x + \sum_{j=i+1}^{k-1} c_j + \sum_{j>k} c_j - \sum_{j<i} c_j + c_k + |t_k|, & \text{si } i < k; \\ x + \sum_{j>i} c_j - \sum_{j<i} c_j, & \text{si } i = k; \\ x + \sum_{j>i} c_j - \sum_{j<k} c_j - \sum_{j=k+1}^{i-1} c_j - c_k, & \text{si } i > k \end{cases}$$

$$= \begin{cases} x + t_i + |t_k|, & \text{si } i < k; \\ x + t_i, & \text{si } i = k; \\ x + t_i - |t_k|, & \text{si } i > k. \end{cases}$$

Par conséquent, les facteurs circulaires de longueur 2 de  $E_{c'}$  sont :

- $x \cdot (x + t_i + |t_k|)$ , si  $x \in I_i$  et  $i < k$ ;
- $x \cdot (x + t_i)$ , si  $x \in I_k$ ;
- $x \cdot (x + t_i - |t_k|)$ , sinon.

Maintenant, décrivons les facteurs de longueur 2 de  $\psi_k(E_c)$ . Soit  $x \in \llbracket 1, n \rrbracket$ . Alors, il existe  $i \in \llbracket 1, r \rrbracket$  tel que  $x \in I_i$ . Il y a deux cas à considérer.

Supposons que  $t_k$  est positif. Ainsi, la valeur de  $m$  dans l'équation (3.5) est  $\sum_{j \leq k} c_j$ . Si  $i < k$ , alors  $x \in \llbracket 1, \sum_{j < k} c_j \rrbracket$  et donc  $x < m$ . De plus, nous avons que  $E_c(x) > m + |t_k|$ , puisque  $\sum_{j > k} c_j = |t_k| + \sum_{j < k} c_j$ . Ainsi,  $\psi_k(x \cdot E_c(x)) = x \cdot (E_c(x) + |t_k|) = x \cdot (x + t_i + |t_k|)$ . Puisque  $i < k$ , on a que  $x \in I_i$  implique  $x \in I'_i$  et  $E'_c(x) = x + t_i + |t_k|$ . Par conséquent, l'élément  $x$  est envoyé sur le même élément par les deux échanges d'intervalles discrets symétriques tel que désiré.

Si  $i = k$ , alors  $x \in \llbracket 1 + \sum_{j < k} c_j, \sum_{j \leq k} c_j \rrbracket$  et  $x \leq m$ . De plus,  $E_c(x) \in \llbracket \sum_{j < k} c_j + 1, m + |t_k| \rrbracket$ , car  $\sum_{j > k} c_j = |t_k| + \sum_{j < k} c_j$ . Par conséquent,

$$\psi_k(x \cdot E_c(x)) = \begin{cases} x \cdot (x + t_i), & \text{si } x < m; \\ x \cdot (x + t_i) \cdot (x + t_i + |t_k|), & \text{sinon.} \end{cases}$$

Cela signifie que tous les éléments de  $I_k$  sont envoyés sur  $x + t_i$  tel que souhaité.

Si  $i > k$ , nous avons que  $x \in \llbracket 1 + \sum_{j \leq k} c_j, n \rrbracket$  et donc que  $x > m$ . Ainsi,  $E_c(x) < m + |t_k|$ , puisque  $\sum_{j \leq k} c_j > |t_k| + \sum_{j > k} c_j$ . Par conséquent, on a que

$$\psi_k(x \cdot E_c(x)) = \begin{cases} x \cdot (x + |t_k|) \cdot E_c(x) & \text{si } x \leq m + |t_k| \text{ et } E_c(x) \leq m \\ x \cdot (x + |t_k|) \cdot E_c(x) \cdot (E_c(x) + |t_k|) & \text{si } x \leq m + |t_k| \text{ et } E_c(x) > m \\ (x + |t_k|) \cdot E_c(x) & \text{si } x > m + |t_k| \text{ et } E_c(x) \leq m \\ (x + |t_k|) \cdot E_c(x) \cdot (E_c(x) + |t_k|) & \text{si } x > m + |t_k| \text{ et } E_c(x) > m. \end{cases} \quad (3.7)$$

Ainsi, nous avons que chaque élément  $x$  entre  $m + 1$  et  $m + |t_k|$  est envoyé sur  $x + |t_k|$ . Par conséquent, tous les éléments de  $I'_k$  sont envoyés sur la valeur désirée. De plus, tous les éléments de  $I'_i$  avec  $i > k$  sont de la forme  $x + |t_k|$  correspondant aux deux derniers cas de l'équation (3.7). Les facteurs circulaires de longueur 2 sont  $(x + |t_k|) \cdot E_c(x) = (x + |t_k|) \cdot (x + t_i)$ , ce qui peut être réécrit par  $y \cdot (y + t_i - |t_k|)$  en substituant  $x + |t_k|$  par  $y$ . Ceci complète la démonstration du cas  $t_k > 0$ .

Si  $t_k$  est négatif, une analyse cas par cas analogue démontre l'égalité des facteurs de longueur 2. ■

*Démonstration lemme 3.17.* Par le lemme 3.20, nous avons que

$$\psi_k(E_c) = E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)},$$

pour tout  $k \in \llbracket 1, r \rrbracket$ . Comme une substitution sur l'ensemble des mots circulaires

préserve le nombre de mots circulaires, nous avons que

$$\gamma(E_c) = \gamma(\psi_k(E_c)) = \gamma(E_{(c_1, \dots, c_i + |t_i|, \dots, c_r)})$$

tel qu'affirmé. ■

Il existe d'autres opérations sur les compositions préservant le nombre de cycles telles que les opérations de [Rauzy \(1979\)](#) ou bien des opérations sur le premier élément ou le dernier élément de la composition données par [Karnauhova et Liebscher \(2017\)](#). Les opérations présentées ici permettent de démontrer que la fonction  $\mathcal{Z}$  calcule le nombre de cycles de l'échange d'intervalles discrets symétrique associé à une composition.

**Théorème 3.21** ([Karnauhova et Liebscher \(2017\)](#)). *Le nombre de cycles de l'échange d'intervalles discrets symétrique  $E_c$  est donné par la fonction  $\mathcal{Z}$  définie par*

$$\mathcal{Z}(c) = \begin{cases} c_1 & \text{si } r = 1, \\ c_i + \mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) & \text{si } |t_i| = 0, \\ \mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) & \text{si } c_i = |t_i|, \\ \mathcal{Z}(c_1, \dots, c_i - |t_i|, \dots, c_r) & \text{sinon,} \end{cases} \quad (3.8)$$

avec  $i$  le plus petit entier satisfaisant la relation  $c_i \geq |t_i|$ .

La démonstration du théorème [3.21](#) n'est pas celle présentée dans [Karnauhova et Liebscher \(2017\)](#), mais plutôt celle présentée dans [Lapointe \(2019\)](#).

*Démonstration.* La fonction  $\mathcal{Z}$  est bien définie par le lemme [3.13](#). Analysons chacun des cas de la partie récursive de la fonction  $\mathcal{Z}$  afin de démontrer que l'image de la fonction  $\mathcal{Z}$  est le nombre d'orbites de la composition. Si  $r = 1$ ,

on a  $\mathcal{Z}(c_1) = c_1$ , car  $E_{(c_1)}(x) = x$  pour tout  $x \in \llbracket 1, c_1 \rrbracket$ . Si  $|t_i| = 0$ , alors nous avons que  $\gamma(c) = c_i + \gamma(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) = \mathcal{Z}(c)$  par le lemme 3.15. Si  $|t_i| = c_i$ , alors  $\gamma(c) = \gamma(c_1, \dots, c_i - |t_i|, \dots, c_r) = \gamma(c_1, \dots, c_{i-1}, 0, c_{i+1}, \dots, c_r) = \gamma(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) = \mathcal{Z}(c)$  par le lemme 3.17. Si  $|t_i| > c_i$  alors  $\gamma(c) = \gamma(c_1, \dots, c_i - |t_i|, \dots, c_r) = \mathcal{Z}(c)$  par le lemme 3.17. Donc, l'image de la fonction  $\mathcal{Z}$  est le nombre de cycles de l'échange d'intervalles discrets engendré par cette composition. ■

**Corollaire 3.22.** *Un échange d'intervalles discrets symétrique  $E_c$  est minimal si et seulement si  $\mathcal{Z}(c) = 1$ .*

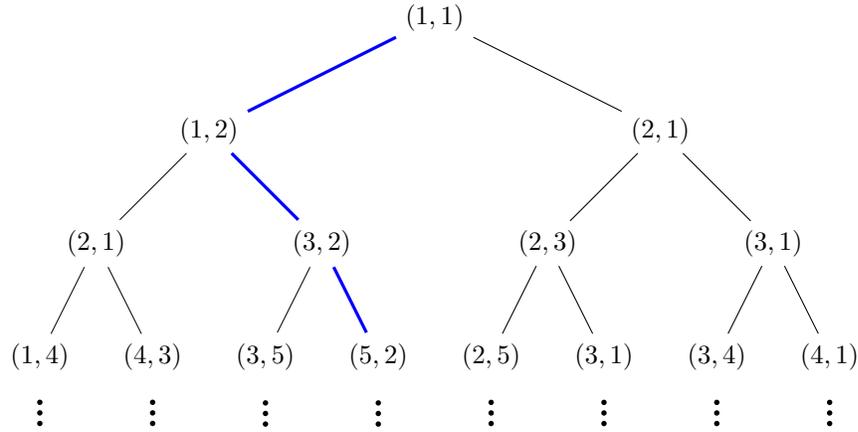
*Démonstration.* L'échange d'intervalles discrets symétrique  $E_c$  est minimal s'il a un seul cycle, ainsi par le théorème 3.21,  $\mathcal{Z}(c) = 1$  ■

### 3.3 Arbres des compositions circulaires

L'arbre de Raney, de même que celui de Stern-Brocot, contient toutes les fractions irréductibles et donc toutes les paires d'entiers premiers entre eux. Remplaçons les fractions  $\frac{i}{j}$  par les paires  $(i, j)$ . L'arbre ainsi obtenu contient toutes les compositions circulaires de longueur 2. La suite de compositions obtenue en appliquant la fonction  $\mathcal{Z}$  sur une composition de longueur 2 correspond aux étiquettes des noeuds sur le chemin entre ce sommet et la racine de l'arbre. Par exemple, la suite de compositions engendrée par  $(5, 2)$  est  $(5, 2), (3, 2), (1, 2), (1, 1)$ . Elle est identifiée en bleu dans la figure 3.2.

Généralisons l'arbre de Raney. L'*arbre des compositions circulaires* est un arbre infini dont la racine est la composition  $(1, 1)$ . Les enfants du noeud étiqueté par la composition circulaire  $c$  sont :

Forme 1 :  $(c_1, \dots, c_i + |t_i|, \dots, c_r)$  pour tout  $i \in \llbracket 1, r \rrbracket$  ;

Figure 3.2: Chemin entre  $(5, 2)$  et la racine sur l'arbre de Raney

Forme 2 :  $(c_1, \dots, c_i, |\delta_i|, c_{i+1}, \dots, c_r)$  pour tout  $i \in \llbracket 0, r \rrbracket$  avec

$$\delta_i = n - 2 \sum_{j \leq i} c_j, \text{ si } \delta_i > c_{i+1} \text{ ou } -\delta_i > c_i. \quad (3.9)$$

On note qu'un chemin dans l'arbre ne correspond pas nécessairement à la suite de composition engendrée par  $\mathcal{Z}$ , car la condition (3.9) ne choisit pas nécessairement le plus petit  $i$  tel que  $c_i \geq |t_i|$ . La figure 3.3 illustre tous les enfants de la composition  $(1, 1, 2, 4)$ . Tous les enfants d'une composition  $c$  de forme 1 sont distincts de  $c$ . En effet, la composition étant circulaire, on a que  $t_i$  est non nul par le lemme 3.15. Ainsi, chaque composition circulaire a exactement  $r$  enfants de forme 1. La situation est plus complexe dans le cas des enfants de forme 2. Par exemple à la figure 3.3,  $\delta_3 = 0$ , alors  $(1, 1, 2, 0, 4)$  n'est pas une composition. La condition (3.9) permet d'éliminer les enfants qui ne sont pas des compositions.

De plus, on souhaite que l'arbre des compositions circulaires contienne toutes les compositions circulaires une seule fois. Ce n'est pas le cas sans la condition (3.9). En effet, la composition circulaire  $(1, 1, 2, 4)$  peut être obtenue de deux façons : par la composition  $(1, 1, 4)$  et par la composition  $(1, 1, 2)$  (voir la figure 3.4). Nous démontrons que les compositions de la forme  $(c_1, \dots, c_i, |\delta_i|, c_{i+1}, \dots, c_r)$  ont deux

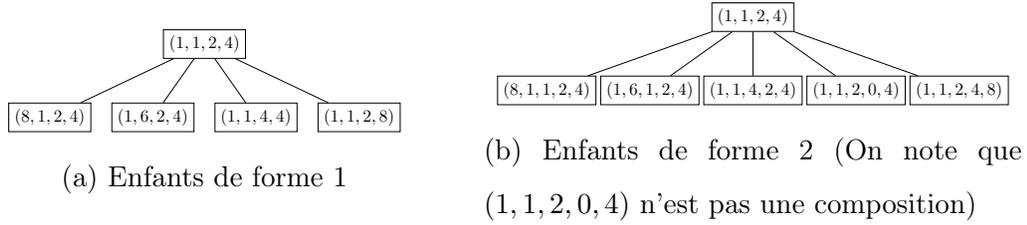


Figure 3.3: Les enfants de la composition  $(1, 1, 2, 4)$  dont le vecteur de translation est  $(7, 5, 2, -4)$  et  $\delta = (8, 6, 4, 0, -8)$

parents possibles. La condition (3.9) permet de choisir le parent d'une composition dans l'arbre.

**Théorème 3.23** (Lapointe (2019)). *Toutes les compositions circulaires apparaissent exactement une fois dans l'arbre des compositions circulaires.*

La figure 3.5 illustre le début de l'arbre des compositions circulaires. Avant de démontrer le théorème, il nous faut quelques résultats sur la forme des compositions circulaires.

**Lemme 3.24.** *Soit  $c$  une composition et  $i \in \llbracket 1, r \rrbracket$ . Si  $c_i > |t_i|$ , alors  $c_k < |t_k|$  pour tout  $k \in \llbracket 1, r \rrbracket \setminus \{i\}$ .*

*Démonstration.* Soit  $i$  un entier dans l'ensemble  $\llbracket 1, r \rrbracket$  tel que  $c_i > |t_i|$ . Dans un premier temps, on montre que  $i$  est l'indice autour duquel la fonction  $t_i$  change de signe, c'est-à-dire que si  $k < i$ ,  $t_k$  est positif et si  $k > i$ ,  $t_k$  est négatif. Soit  $k < i$ . Il y a deux cas à considérer pour démontrer que  $t_k$  est positif :

1. Supposons que  $t_i$  est positif. Par la définition de  $t_i$ ,  $\sum_{j>i} c_j > \sum_{j<i} c_j$ .

Puisque  $k < i$  et que tous les éléments de la composition sont des entiers

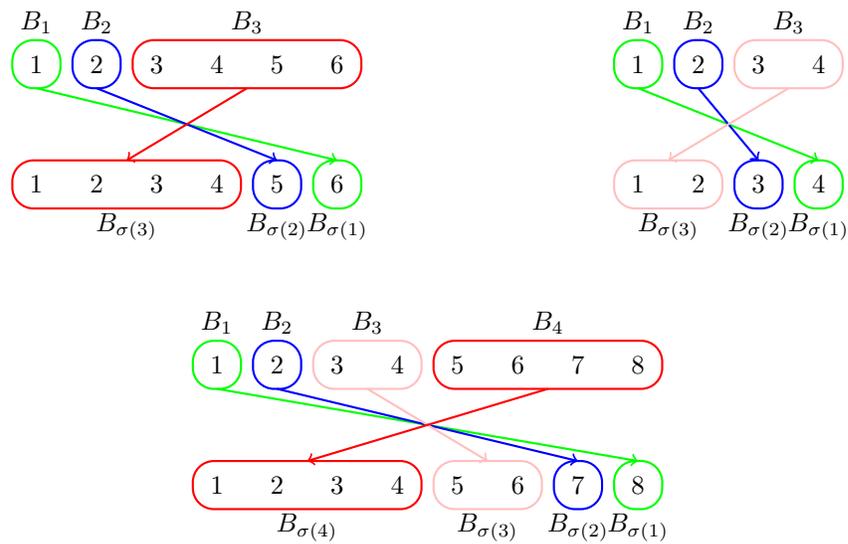


Figure 3.4: L'échange d'intervalles discrets symétrique en haut à droite est engendré par la composition  $(1, 1, 4)$  et celui en haut à gauche est engendré par la composition  $(1, 1, 2)$ . Ce sont les parents possibles de l'échange d'intervalles discrets symétrique engendré par la composition  $(1, 1, 2, 4)$ . Le  $\delta$  de la composition  $(1, 1, 4)$  est  $(6, 4, 2, -6)$  et celui de la composition  $(1, 1, 2)$  est  $(4, 2, 0, -4)$ . Par conséquent,  $(1, 1, |\delta_2|, 2) = (1, 1, 2, 4) = (1, 1, 2, |\delta_4|)$ .

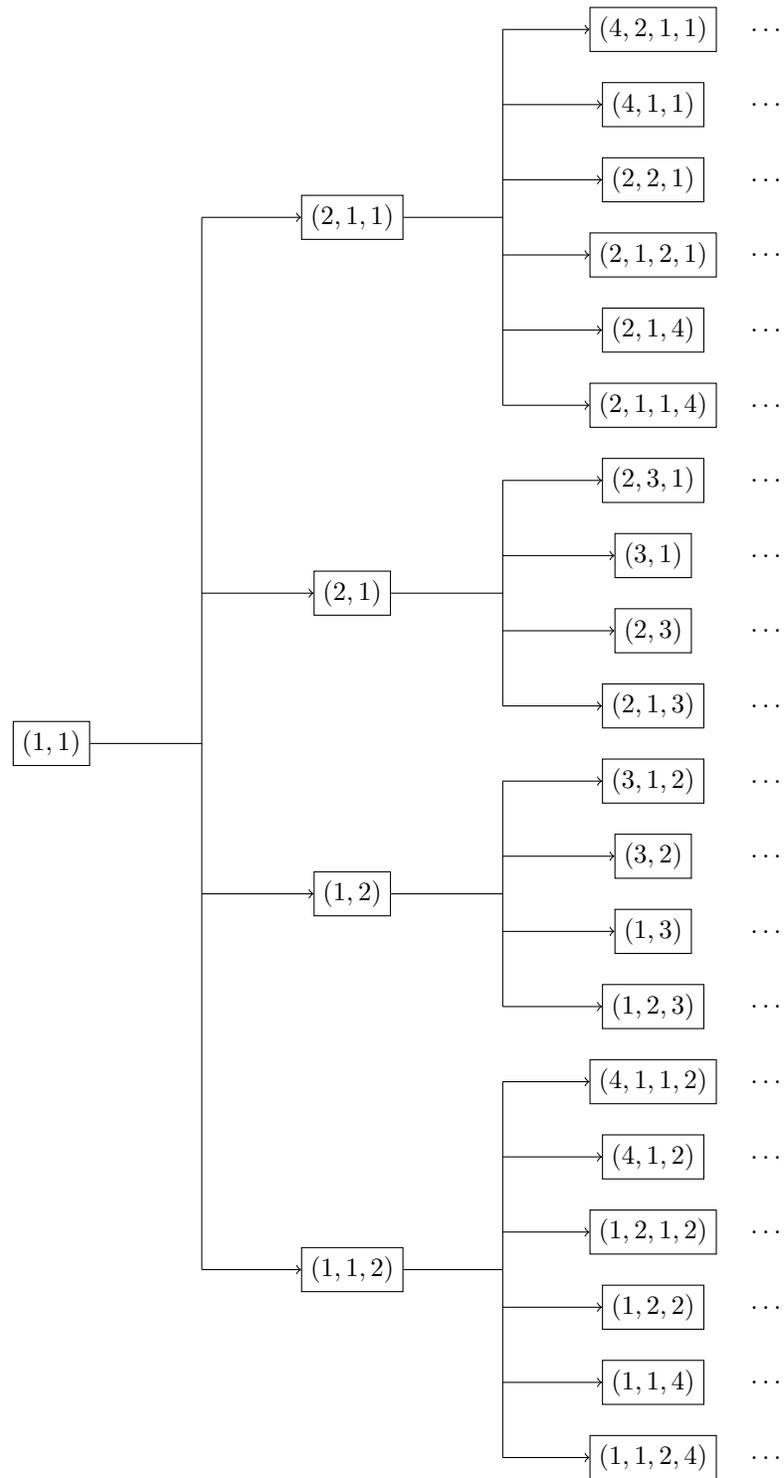


Figure 3.5: L'arbre des compositions circulaires

strictement positifs, on obtient les inégalités suivantes :

$$\sum_{j>k} c_j > \sum_{j>i} c_j > \sum_{j<i} c_j > \sum_{j<k} c_j.$$

Par conséquent,  $t_k = \sum_{j>k} c_j - \sum_{j<k} c_j > 0$ .

2. Supposons que  $t_i$  est négatif. La définition de  $t_i$  signifie que

$$c_i > |t_i| = -t_i = \sum_{j<i} c_j - \sum_{j>i} c_j.$$

En ajoutant  $\sum_{j>i} c_j$  de chaque côté de l'inégalité, on obtient que

$$c_i + \sum_{j>i} c_j > \sum_{j<i} c_j.$$

Cette inégalité implique que

$$\sum_{j>k} c_j \geq c_i + \sum_{j>i} c_j > \sum_{j<i} c_j > \sum_{j<k} c_j$$

puisque  $i < k$  et que tous les éléments de la composition sont positifs.

Donc,  $t_k = \sum_{j>k} c_j - \sum_{j<k} c_j > 0$  tel que souhaité. Un raisonnement similaire prouve que si  $k > i$ , alors  $t_k < 0$ .

Démontrons que si  $k \neq i$ , alors  $c_k < |t_k|$ . Il y a deux cas :  $t_i$  est positif ou négatif. Supposons que  $t_i$  est positif. Si  $k < i$ , alors le paramètre de translation  $t_k$  peut-être écrit comme une somme de  $t_i$  et d'éléments de la composition  $c$  :

$$\begin{aligned} t_i + c_i + 2 \sum_{j=k+1}^{i-1} c_j + c_k &= \sum_{j>i} c_j - \sum_{j<i} c_j + c_i + 2 \sum_{j=k+1}^{i-1} c_j + c_k \\ &= \sum_{j>k} c_j - \sum_{j<k} c_j \\ &= t_k. \end{aligned}$$

De plus,  $t_k = t_i + c_i + 2 \sum_{j=k+1}^{i-1} c_j + c_k > c_k$ , car  $t_i$  est positif. Par conséquent,  $|t_k| = t_k > c_k$  puisque  $k < i$ .

Si  $k > i$ , alors le paramètre de translation  $t_k$  peut-être écrit comme une somme de  $t_i$  et d'éléments de la composition  $c$  :

$$\begin{aligned} t_i - c_i - 2 \sum_{j=i+1}^{k-1} c_j - c_k &= \sum_{j>i} c_j - \sum_{j<i} c_j - c_i - 2 \sum_{j=i+1}^{k-1} c_j - c_k \\ &= \sum_{j>k} c_j - \sum_{j<k} c_j \\ &= t_k. \end{aligned}$$

De plus,  $t_k = -t_i + c_i + 2 \sum_{j=i+1}^{k-1} c_j + c_k$  est une somme d'entiers positifs, car  $c_i - t_i > 0$ . Ainsi,  $|t_k| > c_k$ . Un raisonnement similaire démontre le cas où  $t_i$  est négatif. ■

**Lemme 3.25.** *Soit  $c$  une composition circulaire et  $i \in \llbracket 1, r \rrbracket$ . Si  $c_i = |t_i|$ , alors il existe un entier  $d \in \{i-1, i+1\}$  tel que  $c_d = |t_d|$  et  $c_j < |t_j|$  pour tout  $j \in \llbracket 1, r \rrbracket - \{i, d\}$ . De plus, on a que  $c_d \neq c_i$ .*

*Démonstration.* Soit  $c$  une composition circulaire telle que  $c_i = |t_i|$ . D'abord, assurons-nous que  $c_d$  existe. Si  $t_i > 0$ , alors

$$c_i = t_i = \sum_{j>i} c_j - \sum_{j<i} c_j \Rightarrow -c_{i+1} = \sum_{j>i+1} c_j - \sum_{j<i+1} c_j = |t_{i+1}|.$$

De même, si  $t_i < 0$ , alors  $c_{i-1} = |t_{i-1}|$ .

Ensuite, démontrons que  $c_i \neq c_d$  par l'absurde. Supposons que  $d = i+1$  et que  $c_i = c_d$ . Puisque  $c_i = |t_i|$ , nous avons que  $\mathcal{Z}(c) = \mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r)$  par le théorème 3.21. De plus,  $c_i = \sum_{j>i} c_j - \sum_{j<i} c_j$  et  $c_i = c_{i+1}$ , ce qui signifie que  $\sum_{j>i+1} c_j - \sum_{j<i} c_j = 0$ . Ainsi, nous avons que  $\mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_r) = c_{i+1} + \mathcal{Z}(c_1, \dots, c_{i-1}, c_{i+2}, \dots, c_r)$  par le lemme 3.15. On a alors que  $\mathcal{Z}(c) > 1$ , car  $c_{i+1} > 0$ , contredisant l'hypothèse que  $c$  est un composition circulaire. La preuve est similaire pour le cas  $d = i-1$ .

Finalement, montrons que  $c_k < |t_k|$  si  $k \neq d$  et  $k \neq i$ . Supposons que  $t_i > 0$ , alors  $d = i + 1$ . Si  $k < i$ , nous avons que  $c_i = t_i = \sum_{j>i} c_j - \sum_{j<i} c_j$ . Isolons  $c_k$  dans l'équation pour obtenir  $c_k = -\sum_{j<k} c_j - \sum_{j=k+1}^i c_j + \sum_{j>k} c_j$ . De plus,

$$\begin{aligned} c_i &= -\sum_{j<k} c_j - \sum_{j=k+1}^i c_j + \sum_{j>k} c_j + t_i - t_i \\ &= t_k - 2 \sum_{j=k+1}^i c_j \\ &< t_k \end{aligned}$$

puisque tous les éléments de la composition sont strictement positifs. Une démonstration similaire prouve que si  $k > i$ , alors  $c_k < |t_k|$ . Pour terminer la preuve, il suffit d'utiliser un raisonnement analogue pour démontrer que  $c_k < |t_k|$  si  $t_i < 0$ . ■

**Corollaire 3.26.** *Soit  $c$  une composition circulaire. Si la composition  $c$  n'est pas  $(1, 1)$ , alors elle a un unique ancêtre dans l'arbre des compositions circulaires.*

*Démonstration.* Soit  $c$  une composition circulaire. Il existe  $k \in \llbracket 1, r \rrbracket$  tel que  $c_k \geq |t_k|$  par lemme 3.11. De plus,  $t_k$  est non nul par le lemme 3.15 puisque la composition est circulaire. Ainsi, la composition  $c$  est un enfant de forme 1 ou bien de forme 2 d'une composition  $c'$ .

Supposons que  $c$  est un enfant de forme 1, c'est-à-dire que  $c_k > |t_k|$ . Nous avons que  $c_i < |t_i|$  pour tout  $i \in \llbracket 1, r \rrbracket - \{k\}$  par le lemme 3.24. De plus,  $c$  n'est pas un enfant de forme 2, car pour tout  $j \in \llbracket 1, r \rrbracket$ ,  $c_j \neq |t_j|$ . Ainsi, la seule réduction possible de  $c$  est à la position  $k$ . Par conséquent, l'unique parent de  $c$  est  $c' = (c_1, \dots, c_k - |t_k|, \dots, c_r)$ .

Supposons que  $c$  est un enfant de forme 2, c'est-à-dire que  $c_k = |t_k|$ . Nous avons qu'il existe  $d \in \{k-1, k\}$  tel que  $c_d = |t_d|$  et  $c_i < |t_i|$  pour tout  $i \in \llbracket 1, r \rrbracket - \{k, d\}$  par le lemme 3.25. Ainsi,  $c$  n'est pas une enfant de forme 1, car  $c_i \leq |t_i|$

pour tout  $i \in \llbracket 1, r \rrbracket$ . Si  $\delta_k > c_{k+1}$  ou  $-\delta_k > c_k$ , alors le parent de  $c$  est  $c' = (c_1, \dots, c_k, c_{k+2}, \dots, c_r)$  sinon le parent de  $c$  est  $c' = (c_1, \dots, c_k, c_{k+2}, \dots, c_r)$  si  $\delta_k > 0$ , ou  $c' = (c_1, \dots, c_{k-2}, c_k, \dots, c_r)$  si  $\delta_k < 0$ . Par conséquent, chaque composition circulaire a un unique parent dans l'arbre des compositions circulaires. ■

*Démonstration du théorème 3.23.* Soit  $T$  l'arbre des compositions circulaires. Tout d'abord, démontrons que toutes les compositions dans l'arbre  $T$  sont circulaires. L'échange d'intervalles discrets symétrique  $E_{(1,1)}$  est minimal, car c'est la permutation  $(1, 2)$ . Soit  $c$  une composition apparaissant dans l'arbre  $T$  qui n'engendre pas un échange d'intervalles discrets symétrique minimal et telle que toutes les compositions apparaissant avant elles dans  $T$  sont circulaires. Il existe un entier  $i \in \llbracket 1, r \rrbracket$  tel que  $c_i \geq |t_i|$  et

$$\gamma(c_1, \dots, c_i - |t_i|, \dots, c_r) = \gamma(c) > 1$$

par le lemme 3.17. La composition  $(c_1, \dots, c_i - |t_i|, \dots, c_r)$  n'est pas circulaire et apparaît dans l'arbre  $T$  avant la composition  $c$ , contradiction. Par conséquent, toutes les compositions dans l'arbre  $T$  sont circulaires.

Ensuite, démontrons que toutes les compositions circulaires apparaissent dans l'arbre  $T$ . La composition  $(1, 1)$  est la racine de l'arbre. Soit  $S$  l'ensemble des compositions circulaires n'apparaissant pas dans l'arbre  $T$  et  $c$  la plus petite composition de  $S$  selon l'ordre  $\triangleleft$ . Il existe un entier  $i \in \llbracket 1, r \rrbracket$  tel que  $c_i \geq |t_i|$  et

$$\gamma(c_1, \dots, c_i - |t_i|, \dots, c_r) = \gamma(c)$$

par le lemme 3.17. Ainsi, la composition  $(c_1, \dots, c_i - |t_i|, \dots, c_r)$  n'apparaît pas dans l'arbre, autrement  $c$  apparaîtrait aussi dans l'arbre  $t$ . Il y a une contradiction avec la minimalité de  $c$ , car  $(c_1, \dots, c_i - |t_i|, \dots, c_r) \triangleleft c$ .

Finalement, il faut montrer qu'aucune composition circulaire n'est l'étiquette d'au plus un sommet de l'arbre  $T$ . Soit  $c$  la plus petite composition apparaissant plus

d'une fois dans l'arbre  $T$ . Par le corollaire 3.26, nous savons que  $c$  a pour unique parent  $c'$ . On observe que  $c$  ne peut pas être deux fois l'enfant du même noeud étiqueté par  $c'$  sinon  $c_d = c_i$  ce qui contredit le lemme 3.25. Par conséquent,  $c$  est l'enfant de deux sommets distincts tous deux étiquetés par  $c'$ , contredisant la minimalité de  $c$  puisque  $c' \triangleleft c$ . Donc, chaque composition circulaire apparaît une seule fois dans l'arbre des compositions circulaires. ■

Dans la section 4.6, on démontre que l'arbre étiqueté par l'image commutative des mots parfaitement amassants est isomorphe à l'arbre des compositions circulaires. Pour ce faire, un sommet dans ces arbres doit avoir le même nombre d'enfants. La proposition suivante dénombre les enfants d'un sommet quelconque dans l'arbre des compositions circulaires.

**Proposition 3.27.** *Soit  $c$  une composition circulaire de longueur  $r$ . La composition  $c$  a exactement  $2r$  enfants dans l'arbre des compositions circulaires.*

*Démonstration.* Une composition circulaire a exactement  $r$  enfants de forme 1. De plus, une composition circulaire a au plus  $r + 1$  enfants de forme 2. Ainsi, il existe un entier  $i$  tel que la composition  $(c_1, \dots, c_i, |\delta_i|, c_{i+1}, \dots, c_r)$  n'est pas un enfant de  $(c_1, \dots, c_r)$ , c'est-à-dire que  $\delta_i \leq c_{i+1}$  et  $-\delta_i \leq c_i$ . Posons  $\sum_{i=1}^r c_i = n$ . On a que

$$\delta_i = n - 2 \sum_{j \leq i} c_j = \sum_{j > i} c_j - \sum_{j \leq i} c_j.$$

Par conséquent, les conditions sont équivalentes à

$$\delta_i \leq c_{i+1} \Rightarrow \sum_{j > i+1} c_j - \sum_{j < i+1} c_j \leq 0 \quad (3.10)$$

$$-\delta_i \leq c_i \Rightarrow \sum_{j < i} c_j - \sum_{j > i} c_j \leq 0 \quad (3.11)$$

La fonction  $\sum_{j>k} c_j - \sum_{j<k} c_j$  est décroissante sur  $\llbracket 1, r \rrbracket$  et elle change de signe, par le lemme 3.7. Par conséquent, il existe un entier  $k$  tel que

$$\sum_{j>k} c_j - \sum_{j<k} c_j \geq 0 \Rightarrow -\sum_{j>k} c_j + \sum_{j<k} c_j \leq 0 \quad (3.12)$$

$$\sum_{j>k+1} c_j - \sum_{j<k+1} c_j \leq 0. \quad (3.13)$$

Les équations (3.12) et (3.13) signifie que  $\delta_k \leq c_{k+1}$  et que  $-\delta_k \leq c_k$ . Donc, la composition  $(c_1, \dots, c_k, |\delta_k|, c_{k+1}, \dots, c_r)$  n'est pas un enfant de la composition  $(c_1, \dots, c_r)$ . Pour tous les autres entiers  $j \neq k$ , soit  $\delta_j > c_{j+1}$  ou  $-\delta_j > c_j$  puisque la fonction est décroissante. Donc, la composition a  $r$  enfants de forme 2 et un total de  $2r$  enfants. ■

### 3.4 Type cyclique

Le *type cyclique* d'un échange d'intervalles discrets symétrique est un partage de  $n$  dénoté par  $\ell_1^{\alpha_1} \ell_2^{\alpha_2} \dots \ell_k^{\alpha_k}$  avec  $l_1 > l_2 > \dots > l_k$ , où  $l_i$  est la longueur d'un cycle de  $E_c$  et  $\alpha_i$  est le nombre de cycles de longueur  $l_i$ . Le type cyclique des échanges d'intervalles discrets symétriques de longueur 2 est déjà connu.

**Proposition 3.28** (folklore). *Soit  $(\lambda_1, \lambda_2)$  une composition. Le type cyclique de l'échange d'intervalles discrets symétrique  $E_{(\lambda_1, \lambda_2)}$  est le partage*

$$\left( \frac{\lambda_1 + \lambda_2}{\text{pgcd}(\lambda_1, \lambda_2)} \right)^{\text{pgcd}(\lambda_1, \lambda_2)}.$$

*Démonstration.* L'idée est de décrire l'orbite d'un élément  $x$  quelconque. Soit  $d = \text{pgcd}(c_1, c_2)$ . Démontrons par récurrence que  $E_{(c_1, c_2)}^{(k)}(x) = x + i \cdot d$ , où  $i \in \mathbb{N}$ .

On a que

$$\begin{aligned}
 E_{(c_1, c_2)}^{(k+1)}(x) &= E_{(c_1, c_2)}(E_{(c_1, c_2)}^{(k)}(x)) \\
 &= E_{(c_1, c_2)}(x + id) \\
 &= x + id + t_i \\
 &= x + kd,
 \end{aligned}$$

par l'hypothèse de récurrence. De plus,  $n = c_1 + c_2$ , alors  $d \mid n$ . Tous les éléments appartenant à l'orbite de  $x$  sont congrus à  $x$  modulo  $d$ , ainsi le nombre d'orbites est égal au nombre de classes de congruence c'est-à-dire  $d$ . De plus, la longueur des cycles est  $n/d$ . ■

De façon similaire, le type cyclique des échanges d'intervalles discrets symétriques de longueur 3 peut aussi être décrit en utilisant le pgcd.

**Lemme 3.29.** *Soit  $(c_1, c_2, c_3)$  une composition. Les paramètres de translations  $t_1, t_2, t_3$  sont des multiples de  $\text{pgcd}(c_1 + c_2, c_2 + c_3)$ .*

*Démonstration.* Le lemme est direct pour  $t_1 = c_2 + c_3$  et  $t_3 = -c_1 - c_2$ . En soustrayant  $c_2 + c_3$  de  $c_1 + c_2$ , nous avons que  $\text{pgcd}(c_1 + c_2, c_2 + c_3) = \text{pgcd}(c_1 - c_3, c_2 + c_3)$ . Par conséquent,  $t_2 = |c_3 - c_1|$  est un multiple de  $\text{pgcd}(c_1 + c_2, c_2 + c_3)$ . ■

Décrivons les orbites des échanges d'intervalles discrets symétriques de longueur 3.

**Théorème 3.30.** *L'orbite d'un entier  $x \in \llbracket 1, n \rrbracket$  dans l'échange d'intervalles discrets symétrique  $E_{(c_1, c_2, c_3)}$  est l'ensemble*

$$\left\{ x + kd \mid k \in \mathbb{N} \text{ et } \frac{1-x}{d} \leq k \leq \frac{n-x}{d} \right\}$$

avec  $d = \text{pgcd}(c_1 + c_2, c_2 + c_3)$ .

Par exemple, l'échange d'intervalles discrets symétrique  $E_{(9,1,4)}$  est

$$(1, 6, 11)(2, 7, 12)(3, 8, 14)(4, 9, 14)(5, 10).$$

Ainsi, l'orbite de 9 pour cet échange d'intervalles discrets symétrique est  $\{9 + 5k \mid -8/5 \leq k \leq 5/5\} = \{4, 9, 14\}$ .

*Démonstration.* Soit  $d = \text{pgcd}(c_1 + c_2, c_2 + c_3)$  et  $x$  un entier dans l'ensemble  $\llbracket 1, r \rrbracket$ . D'abord, démontrons par récurrence que  $E_{(c_1, c_2, c_3)}^{(k)}(x) = x + id$ . Nous avons que  $E_{(c_1, c_2, c_3)}^{(0)}(x) = x + 0 \cdot d$  et  $E_{(c_1, c_2, c_3)}^{(1)}(x) = x + t_i = x + i \cdot d$ , car  $t_i$  est un multiple de  $d$  par le lemme 3.29. Alors

$$E_{(c_1, c_2, c_3)}^{(k+1)}(x) = E_{(c_1, c_2, c_3)}(E_{(c_1, c_2, c_3)}^{(k)}(x)) = E_{(c_1, c_2, c_3)}(x + id) = x + id + t_j$$

par l'hypothèse de récurrence. De plus,  $t_j$  est un multiple de  $d$  par le lemme 3.29. Donc,  $E_{(c_1, c_2, c_3)}^{(k+1)}(x) = x + id + md = x + (i + m)d$ . Par la définition de  $E_{(c_1, c_2, c_3)}$ , on a que  $1 \leq x + kd \leq n$  ce qui signifie que  $(1 - x)/d \leq k \leq (n - x)/d$ . ■

**Corollaire 3.31.** *Soit  $(c_1, c_2, c_3)$  une composition de  $n = c_1 + c_2 + c_3$ . Posons  $d = \text{pgcd}(c_1 + c_2, c_2 + c_3)$ . Soit  $q$  et  $r$  deux entiers tels que  $n = qd + r$  et  $0 \leq r < d$ . Le type cyclique de  $E_{(c_1, c_2, c_3)}$  est  $q^d$  si  $r = 0$  et  $(q + 1)^r q^{d-r}$  autrement. De plus,  $d = \gamma(E_{(c_1, c_2, c_3)})$ .*

**Exemple 3.32.** *Les cycles de  $E_{(2,4,5)}$  sont  $(1, 10, 4, 7)(2, 11, 5, 8)(3, 6, 9)$ . Par conséquent, son type cyclique est  $4^2 3^1$ . Cela correspond à la formule proposée par le corollaire 3.31, car  $d = \text{pgcd}(2 + 4, 4 + 5) = 3$  et  $11 = 3 \cdot 3 + 2$ .*

*Démonstration.* Par le théorème 3.30, nous avons que les orbites de  $x$  sont

$$\left\{ x + kd \mid k \in \mathbb{N} \text{ et } \frac{1 - x}{d} \leq k \leq \frac{n - x}{d} \right\}.$$

Par conséquent, tous les éléments de l'orbite de  $x$  sont congrus à  $x$  modulo  $d$ . Ainsi, le nombre d'orbites est le nombre de classes de congruence modulo  $d$ , c'est-à-dire  $d$ . La cardinalité de  $\gamma(x)$  est  $\lfloor n/d \rfloor$  si  $x \bmod d < r$  et  $\lfloor n/d \rfloor + 1$  si  $x \bmod d \geq r$ . ■

Le type cyclique des échanges d'intervalles discrets symétriques engendré par une composition de longueur 2 ou 3 est lié au P.G.C.D. Pour les autres échanges d'intervalles discrets symétriques, il n'y a pas de description connue. Nous terminons par une conjecture sur la forme des échanges d'intervalles discrets symétriques quelconques.

**Conjecture 3.33.** *Soit  $c$  une composition. Si  $\ell_1^{\alpha_1} \ell_2^{\alpha_2} \dots \ell_k^{\alpha_k}$  est le partage décrivant le type cyclique de l'échange d'intervalles symétrique discrets  $E_c$ , alors  $k \leq \lceil \frac{r-1}{2} \rceil$ .*

## CHAPITRE IV

### MOTS PARFAITEMENT AMASSANTS : ARBRES ET AUTOMORPHISMES

Les mots parfaitement amassants généralisent les mots de Christoffel en se basant sur la transformée de Burrows-Wheeler. Les mots de Christoffel peuvent être construits à l'aide de certains automorphismes du groupe libre, appelés *morphismes sturmiens*. Une question naturelle est : existe-t-il des morphismes décrivant les mots parfaitement amassants ? [Simpson et Puglisi \(2008\)](#) ont partiellement répondu à cette question en proposant des fonctions sur le monoïde libre qui engendrent les mots parfaitement amassants sur un alphabet ternaire. Dans ce chapitre, nous proposons des morphismes sur le groupe libre qui engendrent les mots parfaitement amassants sur un alphabet quelconque. On construit aussi une généralisation de l'arbre dual des mots de Christoffel duals. Pour finir, on utilise ces morphismes pour démontrer que les mots parfaitement amassants sont des éléments primitifs du groupe libre.

#### 4.1 Automorphismes du groupe libre

Soit  $\mathcal{A}$  un alphabet fini totalement ordonné. Le *groupe libre* avec base  $\mathcal{A}$ , noté  $F(\mathcal{A})$ , est un groupe pour lequel toute fonction de l'ensemble  $\mathcal{A}$  vers le groupe  $G$  a une unique extension en un morphisme de groupe de  $F$  dans  $G$ . L'*inversion* est une fonction du groupe libre  $F$  vers lui-même qui associe un élément  $x$  à son

inverse  $x^{-1}$ . L'inversion est un antimorphisme et une involution.

Un *élément*  $x$  du groupe libre  $F(\mathcal{A})$  est représenté par un mot sur le monoïde libre  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$ . Un mot sur ce monoïde est dit *réduit* s'il ne contient pas de facteur de la forme  $xx^{-1}$  avec  $x$  appartenant à l'ensemble  $\mathcal{A} \cup \mathcal{A}^{-1}$ . Chaque élément de  $F(\mathcal{A})$  est représenté par un unique mot réduit sur le monoïde  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$ . La *longueur* d'un élément de  $F(\mathcal{A})$  est la longueur de son unique mot réduit. Un mot sur le monoïde  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$  est *circulairement réduit* s'il est réduit et qu'en plus la dernière lettre du mot n'est pas l'inverse de sa première lettre. Le mot *circulairement réduit* associé à  $w = xu$  où  $x \in \mathcal{A}$  est le mot  $w'$  tel que les mots  $w$  et  $ux$  sont réduits.

**Exemple 4.1.** Soit  $\mathcal{A} = \{a \prec b \prec c\}$ . Le mot  $abc^{-1}b^{-1}ca$  est circulairement réduit et il est de longueur 6. Le mot  $ab^{-1}bcc a^{-1}$  n'est pas réduit, car il contient le facteur  $b^{-1}b$ . Le mot réduit représentant  $ab^{-1}bcc a^{-1}$  est  $acca^{-1}$ , mais  $acca^{-1}$  n'est pas circulairement réduit. Le mot circulairement réduit associé à  $acca^{-1}$  est  $cc$ .

Soit  $\mathcal{A} = \{a_1 < a_2 < \dots < a_r\}$  un alphabet totalement ordonné de cardinalité  $r$ .

Pour chaque lettre  $\ell$  de  $\mathcal{A}$ , on définit les morphismes de groupes  $\lambda_\ell$  et  $\rho_\ell$  par

$$\lambda_\ell(a) = \begin{cases} a, & \text{si } a = \ell; \\ a\ell^{-1}, & \text{si } a \prec \ell; \\ \ell a, & \text{si } a \succ \ell; \end{cases} \quad \text{et} \quad \rho_\ell(a) = \begin{cases} a, & \text{si } a = \ell; \\ a\ell, & \text{si } a \prec \ell; \\ \ell^{-1}a, & \text{si } a \succ \ell; \end{cases}$$

pour toute lettre  $a \in \mathcal{A}$ . De plus, le morphisme  $\tau$  est défini par  $\tau(a_k) = a_{r-k+1}$  pour tout  $a_k \in \mathcal{A}$ . Le morphisme  $\tau$  est une involution. On définit aussi l'antimorphisme  $\tilde{\tau} = \tilde{\circ} \circ \tau = \tau \circ \tilde{\circ}$  qui est aussi une involution. Cet antimorphisme a une extension naturelle sur le groupe libre  $F(\mathcal{A})$  en posant que  $\tilde{\tau}(a^{-1}) = \tilde{\tau}(a)^{-1}$  pour tout  $a \in \mathcal{A}$ . Des relations entre les morphismes  $\rho_\ell$  et  $\lambda_\ell$  découlent directement de leurs définitions.

**Proposition 4.2.** *Soit  $\ell$  une lettre de  $\mathcal{A}$ . Le morphisme  $\lambda_\ell$  est l'inverse de  $\rho_\ell$ , c'est-à-dire que  $\lambda_\ell \circ \rho_\ell = \rho_\ell \circ \lambda_\ell = \text{Id}$ .*

*Démonstration.* Il suffit de vérifier que  $\lambda_\ell \circ \rho_\ell(a) = \rho_\ell \circ \lambda_\ell(a) = a$  pour toute lettre  $a \in \mathcal{A}$ , puisque les deux fonctions sont des morphismes. Il y a trois cas à considérer.

1. Si  $a \prec \ell$ , alors

$$\begin{aligned}\lambda_\ell \circ \rho_\ell(a) &= \lambda_\ell(a\ell) = a\ell^{-1}\ell = a \\ \rho_\ell \circ \lambda_\ell(a) &= \rho_\ell(a\ell^{-1}) = a\ell\ell^{-1} = a.\end{aligned}$$

2. Si  $a = \ell$ , alors

$$\lambda_\ell \circ \rho_\ell(a) = \lambda_\ell(a) = a = \rho_\ell(a) = \rho_\ell \circ \lambda_\ell(a).$$

3. Si  $a \succ \ell$ , alors

$$\begin{aligned}\lambda_\ell \circ \rho_\ell(a) &= \lambda_\ell(\ell^{-1}a) = \ell^{-1}\ell a = a \\ \rho_\ell \circ \lambda_\ell(a) &= \rho_\ell(\ell a) = \ell\ell^{-1}a = a.\end{aligned}$$

On en conclut que le morphisme  $\lambda_\ell$  est l'inverse de  $\rho_\ell$ . ■

Avant de décrire la deuxième involution, décrivons une propriété du morphisme  $\tau$ .

**Proposition 4.3.** *Soit  $u$  et  $v$  deux mots sur  $\mathcal{A}^*$  tel que  $|u| = |v|$ . Si  $u \prec v$ , alors  $\tau(u) \succ \tau(v)$ .*

*Démonstration.* Puisque  $u \prec v$  et  $|u| = |v|$ , il existe des mots  $r, s, t \in \mathcal{A}^*$  et des lettres  $x \prec y \in \mathcal{A}$  tels que

$$\begin{aligned}u &= rxt \\ v &= rys.\end{aligned}$$

Ainsi, on a que  $\tau(u) = \tau(r)\tau(x)\tau(t)$  et  $\tau(v) = \tau(r)\tau(y)\tau(t)$ . Il suffit alors de comparer les lettres  $\tau(x)$  et  $\tau(y)$ . Puisque  $\tau$  inverse l'ordre des lettres, on a que  $\tau(x) \succ \tau(y)$ . Donc,  $\tau(u) \succ \tau(v)$ . ■

**Proposition 4.4.** *Pour toute lettre  $\ell \in \mathcal{A}$ , les identités suivantes sont vérifiées*

$$\rho_\ell = \tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}$$

$$\lambda_\ell = \tilde{\tau} \circ \rho_{\tau(\ell)} \circ \tilde{\tau}.$$

**Remarque 4.5.** *L'antimorphisme  $\tilde{\tau}$  ne commute pas avec  $\rho_\ell$  et  $\lambda_\ell$ , alors les morphismes  $\rho_\ell$  et  $\lambda_{\tau(\ell)}$  ne sont pas égaux en général.*

*Démonstration.* Les fonctions  $\rho_\ell$ ,  $\lambda_\ell$ ,  $\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}$  et  $\tilde{\tau} \circ \rho_{\tau(\ell)} \circ \tilde{\tau}$  sont des morphismes. Ainsi, il suffit de montrer qu'ils sont égaux pour chacune des lettres  $a \in \mathcal{A}$ . Comme la preuve de chacune des relations est analogue, on le fait seulement pour la relation  $\rho_\ell = \tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}$ . Il y a trois cas à considérer selon la relation entre  $\ell$  et  $a$  :

1. Si  $a \prec \ell$ , alors

$$\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(a) = \tilde{\tau}(\lambda_{\tau(\ell)}(\tau(a))) = \tilde{\tau}(\tau(\ell)\tau(a)) = a\ell = \rho_\ell(a),$$

car  $a \prec \ell$  implique  $\tau(a) \succ \tau(\ell)$ , par le lemme 4.3.

2. Si  $a = \ell$ , alors

$$\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(a) = \tilde{\tau}(\lambda_{\tau(\ell)}(\tau(a))) = \tilde{\tau}(\tau(a)) = a = \rho_\ell(a).$$

3. Si  $a \succ \ell$ , alors

$$\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(a) = \tilde{\tau}(\lambda_{\tau(\ell)}(\tau(a))) = \tilde{\tau}(\tau(a)\tau(\ell)^{-1}) = \ell^{-1}a = \rho_\ell(a),$$

car  $a \succ \ell$  implique  $\tau(a) \prec \tau(\ell)$  par le lemme 4.3.

Par conséquent, les morphismes  $\rho_\ell$  et  $\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}$  sont égaux. ■

Un élément du groupe libre  $F(\mathcal{A})$  est dit *positif* si c'est un élément du monoïde libre  $\mathcal{A}^*$ . Un morphisme est dit *positif* si l'image des éléments de  $A$  appartient au monoïde libre  $\mathcal{A}^*$ . Par exemple sur le groupe libre  $F(\{a, b, c, d\})$ , les morphismes  $\lambda_a = (a, ab, ac, ad)$  et  $\rho_d = (ad, bd, cd, d)$  sont positifs. Sur  $F(\mathcal{A})$ , le morphisme  $\lambda_\ell$  est positif si  $\ell$  est la plus petite lettre de  $\mathcal{A}$  et le morphisme  $\rho_\ell$  est positif si  $\ell$  est la plus grande lettre de  $\mathcal{A}$ . Les autres morphismes ne sont pas positifs, mais le lemme suivant indique si l'image d'un mot par l'un des morphismes est positive.

**Lemme 4.6.** *Soit  $w$  un mot positif appartenant à  $F(\mathcal{A})$  de longueur  $n$ . Le mot circulairement réduit associé à  $\lambda_\ell(w)$  est positif si et seulement si pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $w_{i \bmod n} \prec \ell$  implique que  $w_{i+1 \bmod n} \succeq \ell$ . De même, le mot circulairement réduit associé à  $\rho_\ell(w)$  est positif si et seulement si pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $w_{i \bmod n} \prec \ell$  implique que  $w_{i-1 \bmod n} \succeq \ell$ .*

Par exemple, le mot  $\rho_b(\text{cababa}) = b^{-1}\text{cabbabbab}$ , alors le mot circulairement réduit qui lui est associé est  $\text{cabbabba}$  qui est un élément positif de  $F(\mathcal{A})$ .

*Démonstration.* La démonstration pour  $\lambda_\ell$  et  $\rho_\ell$  est analogue, alors nous présentons seulement celle pour  $\lambda_\ell$ .

Supposons que  $w_{i \bmod n} \prec \ell$  implique que  $w_{i+1 \bmod n} \succeq \ell$  pour tout  $i \in \llbracket 1, n \rrbracket$ . Le mot  $w$  étant positif, le seul élément inverse pouvant apparaître dans  $\lambda_\ell(w)$  est  $\ell^{-1}$ . De plus, toutes les occurrences de la lettre  $\ell^{-1}$  dans  $\lambda_\ell(w)$  sont suivies (circulairement) d'au moins une occurrence de la lettre  $\ell$  dans  $\lambda_\ell(w)$ . Dans le mot circulairement réduit associé à  $\lambda_\ell(w)$ , la lettre  $\ell^{-1}$  n'apparaît pas. Donc le mot circulairement réduit associé à  $\lambda_\ell(w)$  est positif.

Supposons maintenant que le mot circulairement réduit associé à  $\lambda_\ell(w)$  est positif. Alors la lettre  $\ell^{-1}$  n'apparaît pas dans le mot circulairement réduit associé à  $\lambda_\ell(w)$ . Ainsi, on a que toutes les lettres de  $w$  sont plus grandes que  $\ell$  ou que toutes les

occurrences de lettres plus petites que  $\ell$  sont suivies d'une occurrence de la lettre  $\ell$ . Cette condition est respectée si  $w_{i \bmod n} \prec \ell$  implique que  $w_{i+1 \bmod n} \succeq \ell$ . ■

## 4.2 Arbres des mots parfaitement amassants

L'arbre des mots parfaitement amassants est une généralisation de l'arbre dual des mots de Christoffel. Pour le construire, nous présentons une règle récursive utilisant les morphismes  $\lambda_\ell$  et  $\rho_\ell$  ainsi que des morphismes ajoutant une nouvelle lettre à un mot. Soit  $\mathcal{A}$  un alphabet totalement ordonné,  $b \notin \mathcal{A}$  et  $\mathcal{B} = \mathcal{A} \cup \{b\}$  l'alphabet totalement ordonné tel que pour tout  $\ell \in \mathcal{A}$ ,  $\ell \prec b$ . Le morphisme  $N$  de  $\mathcal{A}^*$  vers  $\mathcal{B}^*$  associe à une lettre  $\ell$  la lettre suivante selon l'ordre. On observe que si  $u \prec v$ , alors  $N(u) \prec N(v)$ . Pour toute lettre  $a \in \mathcal{A}$ , on définit un morphisme du monoïde  $\mathcal{A}^*$  vers  $\mathcal{B}^*$ , notés  $\Delta_\ell$  par

$$\Delta_\ell(a) = \begin{cases} a, & \text{si } a \prec \ell; \\ N(a), & \text{sinon;} \end{cases}$$

Soit  $T$  l'arbre infini où chaque sommet a  $\mathbb{N}$  enfants. Identifions les enfants de gauche à droite par les entiers naturels. Un chemin entre un sommet et la racine est représenté par un mot sur  $\mathbb{N}^*$ . Il y a un unique chemin entre un sommet et la racine, alors nous identifions chaque sommet par le mot qui représente ce chemin. L'*arbre des mots parfaitement amassants* est un sous-arbre de l'arbre  $T$ . La racine de l'arbre des mots parfaitement amassants est étiquetée par le mot  $ab$ . Soit  $w$  l'étiquette d'un sommet dans l'arbre des mots parfaitement amassants. Le sommet

$w$  a  $2|\text{Alph}(w)|$  enfants obtenus par la fonction suivante :

$$S_i(w) = \begin{cases} \lambda_\ell \circ \Delta_\ell(w), & \text{si } i \text{ est pair et } g > p; \\ \lambda_\ell(w), & \text{si } i \text{ est impair et } g > p; \\ \rho_\ell(w), & \text{si } i \text{ est pair et } g < p; \\ \rho_{N(\ell)} \circ \Delta_{N(\ell)}(w), & \text{si } i \text{ est impair et } g < p; \end{cases} \quad (4.1)$$

où  $\ell = a_{\lfloor i/2 \rfloor + 1}$ ,  $p = \sum_{j < \ell} |w|_j$  et  $g = \sum_{j > \ell} |w|_j$  pour tout  $i \in \llbracket 0, 2|\text{Alph}(w)| - 1 \rrbracket$ .

**Remarque 4.7.** *Le cas  $g = p$  n'est pas mentionné dans la formule, car l'image commutative des mots parfaitement amassants est une composition circulaire par le théorème 3.8.*

La figure 4.1 illustre le début de cet arbre dessiné de gauche à droite plutôt que de haut en bas et les mots sont écrits sur l'alphabet  $\{a, b, c, d, e, \dots\}$  plutôt que sur l'alphabet  $\{a_1, a_2, \dots\}$  pour en faciliter la lecture.

**Exemple 4.8.** *Soit  $\{a \prec b \prec c\}$  un alphabet totalement ordonné tel que  $a_1 = a$ ,  $a_2 = b$  et  $a_3 = c$ . Les enfants de  $ab$  sont*

- $S_0(ab) = \lambda_a \circ \Delta_a(ab) = \lambda_a(bc) = abac$ , car  $\ell = a$ ,  $p = 0$  et  $g = 1$  ;
- $S_1(ab) = \lambda_a(ab) = aab$ , car  $\ell = a$ ,  $p = 0$  et  $g = 1$  ;
- $S_2(ab) = \rho_b(ab) = abb$ , car  $\ell = b$ ,  $p = 1$  et  $g = 0$  ;
- $S_3(ab) = \rho_{N(b)} \circ \Delta_{N(b)}(ab) = \rho_c(ab) = acbc$ , car  $\ell = b$ ,  $p = 1$  et  $g = 0$ .

L'intérêt de cette construction est reflété dans le théorème suivant :

**Théorème 4.9.** *Tous les mots de Lyndon parfaitement amassants complets apparaissent exactement une fois dans l'arbre des mots parfaitement amassants.*

La démonstration de ce théorème est présentée à la section 4.6, car elle nécessite d'autres résultats préliminaires sur les morphismes.

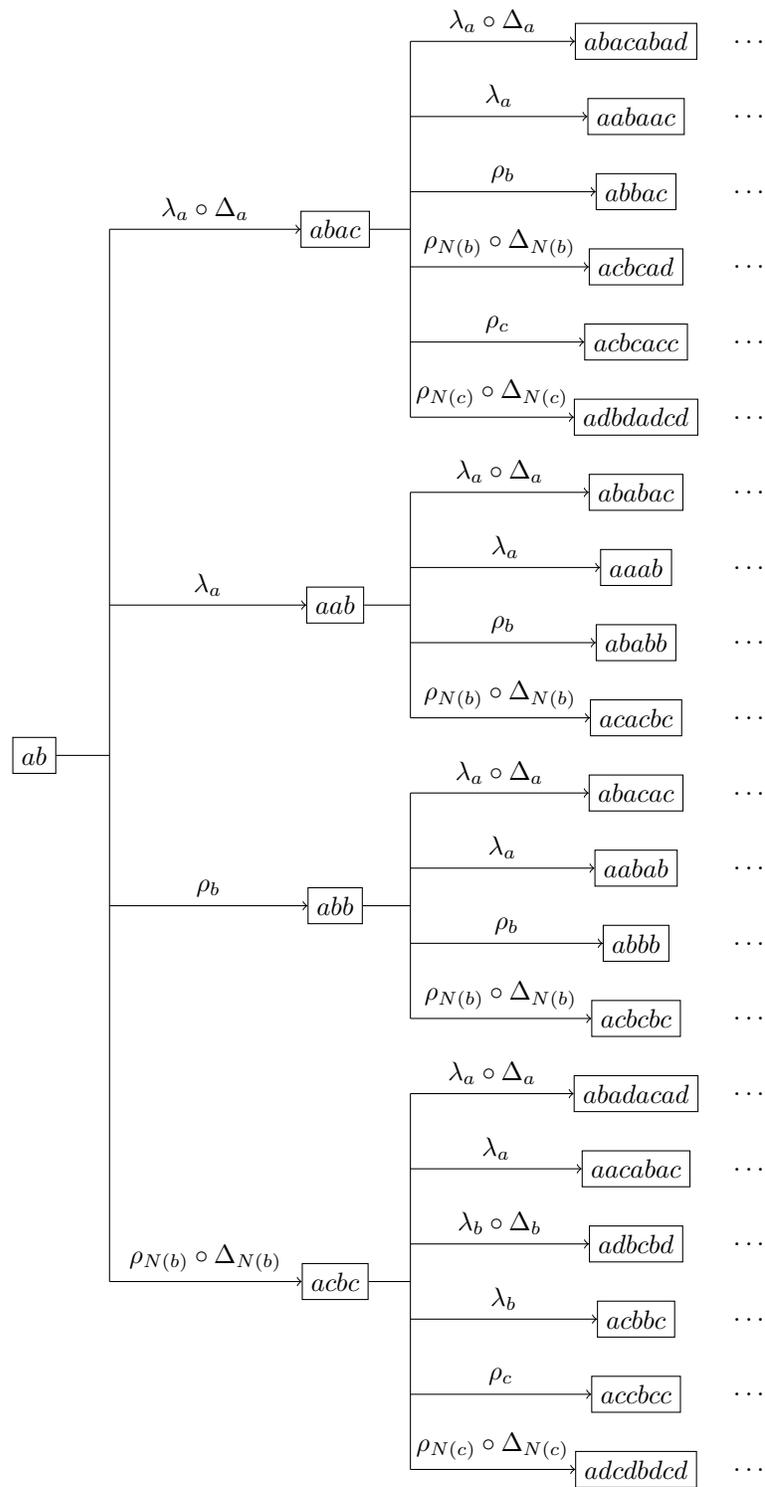


Figure 4.1: L'arbre des mots parfaitement amassants

### 4.3 Positivité

Comme les mots amassants sont des éléments positifs de  $F(\mathcal{A})$ , la première propriété qui nous intéresse est la positivité de l'image d'un mot parfaitement amassant par les morphismes  $\lambda_\ell$ ,  $\rho_\ell$  et  $\Delta_\ell$ . Le morphisme  $\Delta_\ell$  est positif, car l'image d'une lettre positive par l'un de ces morphismes est une lettre positive. Les morphismes  $\lambda_\ell$  et  $\rho_\ell$  ne sont pas des morphismes positifs. Par contre, l'image d'un mot parfaitement amassant par ces morphismes peut être positive.

**Lemme 4.10.** *Soit  $w$  un mot parfaitement amassant. Le mot circulairement réduit associé à  $\lambda_\ell(w)$  (resp.  $\rho_\ell(w)$ ) est positif si*

$$\sum_{j \succ \ell} |w|_j > \sum_{j \prec \ell} |w|_j \quad \left( \text{resp. } \sum_{j \succ \ell} |w|_j < \sum_{j \prec \ell} |w|_j \right)$$

où  $j \in \mathcal{A}$ .

**Exemple 4.11.** *Le mot  $acbcacc$  est parfaitement amassant et son image commutative est  $d = (2, 1, 4)$ . Puisque  $\sum_{j \succ b} |acbcacc|_j = 4 > \sum_{j \prec b} |acbcacc|_j = 2$ , le mot  $\lambda_b(w) = acbbcacbc$  est positif, mais le mot  $\rho_b(w) = accacb^{-1}c$  ne l'est pas.*

*Démonstration.* Soit  $w$  un mot parfaitement amassant de longueur  $n$ . L'ensemble des facteurs de longueur 2 avec répétitions de  $w$  est obtenu en concaténant la dernière colonne de la matrice de ses conjugués avec la première colonne. Puisque  $w$  est parfaitement amassant, la dernière colonne de la matrice de ses conjugués est  $a_n^{d_n} a_{n-1}^{d_{n-1}} \dots a_1^{d_1}$  et la première colonne est  $a_1^{d_1} \dots a_n^{d_n}$ .

Supposons que  $\sum_{j \succ \ell} |w|_j > \sum_{j \prec \ell} |w|_j$ . Alors, tous les facteurs de longueur 2, notés  $l_1 l_2$ , dont la dernière lettre est plus grande que  $\ell$ , débutent par une lettre plus petite que  $\ell$ . De plus, l'ensemble des facteurs du mot  $w$  est fermé par image miroir (corollaire 2.27), c'est-à-dire que  $l_1 l_2 \in \text{Fact}_2(w)$  si et seulement si  $l_2 l_1 \in \text{Fact}_2(w)$ .

Par conséquent,  $l_{i \bmod n} < \ell$  implique que  $l_{i+1 \bmod n} > \ell$ . Donc,  $\lambda_\ell(w)$  est positif, par le lemme 4.6.

Supposons que  $\sum_{j>\ell} |w|_j < \sum_{j<\ell} |w|_j$ . Alors tous les facteurs de longueur 2, notés  $l_1l_2$ , dont la première lettre est plus grande que  $\ell$ , terminent par une lettre plus petite que  $\ell$ . Par conséquent,  $l_{i \bmod n} > \ell$  implique que  $l_{i-1 \bmod n} < \ell$ . Donc,  $\rho_\ell(w)$  est positif, par le lemme 4.6. ■

#### 4.4 Mots parfaitement amassants

Les mots parfaitement amassants sont définis sur le monoïde libre. En identifiant les mots positifs, on cible des mots parfaitement amassants dont l'image par les automorphismes  $\lambda_\ell$  et  $\rho_\ell$  est aussi parfaitement amassante.

**Théorème 4.12.** *Soit  $w$  un mot parfaitement amassant. Le mot circulairement réduit associé à  $\lambda_\ell(w)$  est parfaitement amassant si  $\sum_{j>\ell} |w|_j > \sum_{j<\ell} |w|_j$ .*

**Exemple 4.13.** *Le mot  $acbcacc$  est parfaitement amassant et son image commutative respecte la condition du théorème 4.12. Par conséquent, le mot  $\lambda_b(acbcacc)$  est lui aussi parfaitement amassant. En effet, la transformée de Burrows-Wheeler de  $\lambda_b(acbcacc)$  est  $c^4b^3a^2$ .*

Avant de faire la démonstration du théorème 4.12, nous présentons une propriété des échanges d'intervalles discrets symétriques.

**Lemme 4.14.** *Soit  $c$  une composition de longueur  $r$  telle que  $\sum_{j>i} c_j > \sum_{j<i} c_j$  où  $i \in \llbracket 1, r \rrbracket$ . Si  $k \in \llbracket 1, \sum_{j<i} c_j \rrbracket$ , alors l'image de  $k$  par l'échange d'intervalles discret symétrique,  $E_c$ , satisfait l'inégalité suivante :*

$$E_c(k) > \sum_{j>i} c_j + c_i.$$

*Démonstration.* Il suffit que l'énoncé soit vrai pour la plus petite valeur de  $E_c(k)$ . Déterminons quelle valeur  $k \in \llbracket 1, \sum_{j < i} c_j \rrbracket$  minimise  $E_c(k)$ .

La valeur  $k$  est la plus petite valeur de l'un des intervalles, car  $E_c$  est croissante sur les intervalles. En effet, pour tout  $x, y \in I_m$ , si  $x < y$ , alors  $E_c(x) < E_c(y)$ , car  $E_c(x) = x + t_m$  et  $E_c(y) = y + t_m$ .

De plus,  $k$  est dans l'intervalle  $I_{i-1}$ , car pour tout  $x \in I_m$  et  $y \in I_n$ , l'inégalité  $m < n$  implique  $E_c(x) > E_c(y)$ . En effet,

$$E_c(x) = x + \sum_{j > m} c_j - \sum_{j < m} c_j > \sum_{j < m} c_j + \sum_{j > m} c_j - \sum_{j < m} c_j = \sum_{j > m} c_j,$$

car  $x \in I_m$  signifie que  $\sum_{j < m} c_j < x \leq \sum_{j \leq m} c_j$ . De plus,

$$E_c(y) = y + \sum_{j > n} c_j - \sum_{j < n} c_j \leq \sum_{j \leq n} c_j + \sum_{j > n} c_j - \sum_{j < n} c_j = c_n + \sum_{j > n} c_j,$$

car  $y \in I_n$  signifie que  $\sum_{j < n} c_j < y \leq \sum_{j \leq n} c_j$ . Par conséquent, on obtient

$$E_c(x) > \sum_{j > m} c_j \geq c_n + \sum_{j > n} c_j \geq E_c(y),$$

car  $m < n$  signifie que  $\sum_{j > m} c_j = c_{m+1} + \cdots + c_n + \sum_{j > n} c_j$ .

Par conséquent,  $\sum_{j < i-1} c_j + 1$  est la valeur que minimise  $E_c(k)$  pour tout  $k \in \llbracket 1, \sum_{j < i} c_j \rrbracket$ . Ainsi,

$$\begin{aligned} E_c \left( \sum_{j < i-1} c_j + 1 \right) &= \sum_{j < i-1} c_j + 1 + \sum_{j > i-1} c_j - \sum_{j < i-1} c_j \\ &= \sum_{j > i-1} c_j + 1 \\ &= c_i + \sum_{j > i} c_j + 1. \end{aligned}$$

Donc, pour  $k \in \llbracket 1, \sum_{j < i} c_j \rrbracket$ , on que  $E_c(k) > c_i + \sum_{j > i} c_j$ . ■

Le théorème 4.12 décrit la forme de la transformée de Burrows-Wheeler du mot  $\lambda_\ell(w)$ , mais la démonstration s'appuie sur la relation entre les échanges d'intervalles discrets symétriques minimaux et les mots parfaitement amassants. Soit  $M_c : \llbracket 1, n \rrbracket^* \mapsto \mathcal{A}^*$  qui envoie un entier sur une lettre représentant l'intervalle  $I_j$  dont cet entier est élément dans  $E_c$ , c'est-à-dire  $M_c(i) = a_j$  si  $i \in I_j$ .

**Proposition 4.15.** *Soit  $c = (c_1, c_2, \dots, c_r)$  une composition et  $E_c$  un échange d'intervalle discret symétrique minimal dont le cycle est  $(\alpha_1, \dots, \alpha_n)$ . Si  $k \in \llbracket 1, r \rrbracket$  et  $\sum_{j>k} c_j > \sum_{j<k} c_j$ , alors*

$$\lambda_\ell \circ M_c(\alpha_1, \dots, \alpha_n) \equiv M_d \circ \psi_k(\alpha_1, \dots, \alpha_n), \quad (4.2)$$

où  $\psi_k$  est défini à l'équation (3.5) et  $\ell = a_k$  et  $d = (c_1, \dots, c_k + |\sum_{j>i} c_j - \sum_{j<i} c_j|, \dots, c_r)$ .

On observe que les morphismes  $\lambda_\ell \circ M_d$  et  $M_c \circ \psi_k$  ne sont pas égaux, car l'image du premier est un élément du groupe libre  $F(\mathcal{A})$  tandis que l'image du second est un élément du monoïde libre  $\mathcal{A}^*$ .

**Exemple 4.16.** *Soit  $d = (2, 2, 5)$  et  $k = 2$ . Le cycle de  $E_{(2,2,5)}$  est  $(1, 8, 4, 7, 3, 6, 2, 9, 5)$ . Le tableau 4.1 définit les morphismes  $M_{(2,2,5)}$ ,  $M_{(2,5,5)}$ ,  $\lambda_b$  et  $\psi_2$ . Ainsi, on a que*

$$\lambda_b \circ M_{(2,2,5)}(1, 8, 4, 7, 3, 6, 2, 9, 5) = M_{(2,5,5)} \circ \psi_2(1, 8, 4, 7, 3, 6, 2, 9, 5).$$

*Démonstration.* Rappelons (voir l'équation (3.5)) que la fonction  $\psi_k$  est définie par

$$\psi_k(x) = \begin{cases} x, & \text{si } x \in \llbracket 1, \sum_{j \leq k} c_j \rrbracket; \\ x \cdot (x + |t_k|), & \text{si } x \in \llbracket 1 + \sum_{j \leq k} c_j, |t_k| + \sum_{j \leq k} c_j \rrbracket; \\ x + |t_k|, & \text{sinon;} \end{cases}$$

Tableau 4.1: Les différents morphismes nécessaires à la démonstration du théorème 4.12 sur le mot  $acbcacc$

$M_{(2,2,5)} : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ a & a & b & b & c & c & c & c & c \end{pmatrix}$	$M_{(2,5,5)} : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ a & a & b & b & b & b & b & c & c & c & c & c \end{pmatrix}$
$\lambda_b : \begin{pmatrix} a & b & c \\ ab^{-1} & b & bc \end{pmatrix}$	$\psi_2 : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 \cdot 8 & 6 \cdot 9 & 7 \cdot 10 & 11 & 12 \end{pmatrix}$
$\lambda_b \circ M_{(2,2,5)} : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ ab^{-1} & ab^{-1} & b & b & bc & bc & bc & bc & bc \end{pmatrix}$	$M_{(2,5,5)} \circ \psi_2 : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ a & a & b & b & bc & bc & bc & c & c \end{pmatrix}$

car  $\sum_{j>k} c_j > \sum_{j<k} c_j$  implique que  $t_k > 0$ .

Nous démontrons que l'équation (4.3) est vérifiée en étudiant les facteurs des mots de cette équation. Soit  $\alpha$  une lettre de  $(\alpha_1, \dots, \alpha_n)$ . Si  $\alpha \in I_k$ , alors  $\lambda_\ell \circ M_c(\alpha) = a_k$  et  $M_d \circ \psi_k(\alpha) = a_k$ . De même, si  $\alpha \in \llbracket 1 + \sum_{j \leq k} c_j, |t_k| + \sum_{j \leq k} c_j \rrbracket$ , alors  $\lambda_\ell \circ M_c(\alpha) = M_d \circ \psi_k(\alpha)$ , puisque  $\lambda_\ell \circ M_c(\alpha_k) = a_k a_j$  avec  $\alpha \in I_j$  et  $M_d \circ \psi_k(\alpha) = a_k a_j$ . Dans ces deux cas, les lettres de  $(\alpha_1, \dots, \alpha_k)$  ont la même image.

Par contre, si  $\alpha \in I_j$  avec  $j < k$ , alors

$$\lambda_\ell \circ M_c(\alpha) = a_j a_k^{-1} \neq M_d \circ \psi_t(\alpha) = a_j.$$

De plus, si  $\alpha > \sum_{j \leq k} c_j + |t_k|$ , il existe  $I_{j'}$  avec  $j' > k$  tel que  $\alpha \in I_{j'}$ , alors

$$\lambda_\ell \circ M_c(\alpha) = a_k a_{j'} \neq M_d \circ \psi_t(\alpha) = a_{j'}.$$

Le lemme 4.14 signifie que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $\alpha_m \in I_j$  avec  $j < k$  est suivi de  $\alpha_{m+1} > \sum_{j \leq k} c_j + |t_k|$ . Bien que les  $\alpha_m$  ne soient pas envoyés sur le même mot par les deux morphismes, la structure de  $\alpha_1, \dots, \alpha_n$  garantit que dans ce cas  $\lambda_\ell \circ M_c(\alpha_m \alpha_{m+1}) = a_j a_t^{-1} a_t a_{j'} = a_j a_{j'} = M_d \circ \psi_t(\alpha_m \alpha_{m+1})$ .  $\blacksquare$

*Démonstration du théorème 4.12.* Le mot  $\lambda_\ell(w)$  est positif par le lemme 4.10. Soit  $\mathcal{A}$  un alphabet totalement ordonné de cardinalité  $r$ . Reformulons le problème en termes d'équations sur les morphismes et les échanges d'intervalles discrets symétriques minimaux.

Soit  $c = (|w|_{a_1}, \dots, |w|_{a_n})$  l'image commutative de  $w$  et  $n = |w|$ . Le mot  $w$  étant parfaitement amassant, il existe un échange d'intervalles discrets symétrique minimal  $E_c$  dont le cycle est  $(\alpha_1, \dots, \alpha_n)$  par le théorème 3.8. Ainsi, on a que  $w \equiv M_c(\alpha_1, \dots, \alpha_n)$ .

Comme  $\ell \in \mathcal{A}$ , il existe un entier  $k$  tel que  $\ell = a_k$ . La composition  $d = (d_1, \dots, d_r)$  telle que  $d_i = c_i$  si  $i \neq k$  et  $d_k = c_k + |t_k|$  est circulaire par le lemme 3.17. Alors il existe un échange d'intervalles discret symétrique minimal  $E_d$  dont le cycle est  $(\beta_1, \dots, \beta_{n+|t_k|})$ . Il existe un mot  $w'$  tel que  $w' \equiv M_d(\beta_1, \dots, \beta_{n+|t_k|})$ .

Par conséquent, démontrer que  $\lambda_\ell(w)$  est parfaitement amassant est équivalent à démontrer que  $\lambda_\ell(w) \equiv w'$ , c'est-à-dire que  $\lambda_\ell \circ M_c(\alpha_1, \dots, \alpha_n) \equiv M_d(\beta_1, \dots, \beta_{n+|t_k|})$ . En utilisant le morphisme  $\psi_k$  défini à l'équation (3.5), on a que

$$\lambda_\ell \circ M_c(\alpha_1, \dots, \alpha_n) \equiv M_d \circ \psi_k(\alpha_1, \dots, \alpha_n), \quad (4.3)$$

puisque  $\psi_k(\alpha_1, \dots, \alpha_n) \equiv (\beta_1, \dots, \beta_{n+|t_k|})$  par le lemme 3.20.

Ainsi, les mots  $w'$  et  $\lambda_\ell(w)$  sont conjugués par la proposition 4.15. Par conséquent,  $\lambda_\ell(w)$  est un mot parfaitement amassant, car la transformée de Burrows-Wheeler d'un mot et de chacun de ces conjugués coïncide (proposition 1.16).  $\blacksquare$

Pour démontrer que le morphisme  $\rho_\ell$  préserve les mots parfaitement amassants, on utilise les relations entre  $\rho_\ell$  et  $\lambda_\ell$ . Avant tout, il faut que  $\tilde{\tau}$  préserve lui aussi les mots parfaitement amassants.

**Lemme 4.17.** *Le mot  $w$  est parfaitement amassant si et seulement si  $\tilde{\tau}(w)$  est*

*un mot parfaitement amassant.*

*Démonstration.* Démontrons que si  $w$  est parfaitement amassant, alors le mot  $\tilde{\tau}(w)$  est parfaitement amassant. L'antimorphisme  $\tilde{\tau}$  est la composition de  $\tilde{\cdot}$  et  $\tau$ . Il suffit de démontrer que  $\tau(w)$  est parfaitement amassant, puisque  $\tau(w)$  et  $\tilde{\tau}(w)$  sont alors des conjugués par le lemme 2.26.

Soit  $w_1 \prec w_2 \prec \dots \prec w_n$  les conjugués de  $w$  triés selon l'ordre lexicographique. Les conjugués de  $\tau(w)$  sont les mots  $\tau(w_1), \tau(w_2), \dots, \tau(w_n)$ , car  $\tau$  est un morphisme strictement alphabétique. Comme les mots  $\tau(w_i)$  ont tous la même longueur,  $w_i \prec w_j$  signifie que  $\tau(w_i) \succ \tau(w_j)$  par le lemme 4.3. Les conjugués de  $\tau(w)$  triés selon l'ordre lexicographique sont  $\tau(w_n) \prec \tau(w_{n-1}) \prec \dots \prec \tau(w_1)$ . Appelons  $d_i$  la dernière lettre du mot  $w_i$ . Le mot  $w$  est parfaitement amassant si

$$w_i \prec w_j \Rightarrow d_i \succeq d_j$$

pour tout  $i, j \in \llbracket 1, |w| \rrbracket$ . Par le lemme 4.3, la relation devient

$$\tau(w_i) \succ \tau(w_j) \Rightarrow \tau(d_i) \preceq \tau(d_j)$$

pour tout  $i, j \in \llbracket 1, |w| \rrbracket$ . Donc, le mot  $\tau(w)$  est parfaitement amassant, de même que  $\tilde{\tau}(w)$ .

Supposons que le mot  $\tilde{\tau}(w)$  est parfaitement amassant. On vient de démontrer que le mot  $\tilde{\tau} \circ \tilde{\tau}(w)$  est parfaitement amassant. Or,  $\tilde{\tau}$  est une involution. Ainsi, le mot  $w = \tilde{\tau} \circ \tilde{\tau}(w)$  est parfaitement amassant. ■

**Théorème 4.18.** *Soit  $w$  un mot parfaitement amassant. Le mot circulairement réduit associé à  $\rho_\ell(w)$  est parfaitement amassant si  $\sum_{j>\ell} |w|_j < \sum_{j<\ell} |w|_j$ .*

*Démonstration.* Soit l'alphabet totalement ordonné  $\mathcal{A} = \{a_1, \dots, a_r\}$ . Le mot circulairement réduit associé à  $\rho_\ell(w)$  est positif, par le lemme 4.10. De plus,

$\rho_\ell(w) = \tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(w)$  sont égaux par la proposition 4.4. Démontrons que le mot  $\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(w)$  est parfaitement amassant. Comme l'antimorphisme  $\tilde{\tau}$  préserve les mots parfaitement amassants, il suffit de démontrer que  $\lambda_{\tau(\ell)}$  est parfaitement amassant.

Posons  $d = (d_1, \dots, d_r) = (|w|_a)_{a \in \mathcal{A}}$  l'image commutative du mot  $w$ . L'image commutative de  $\tilde{\tau}(w)$  est  $d' = (d'_1, \dots, d'_r)$  avec  $d'_i = d_{r-i+1}$ . Soit  $k$  l'entier tel que  $\ell = a_k$ . Le mot  $\lambda_{\tau(\ell)} \circ \tilde{\tau}(w)$  est parfaitement amassant si  $\sum_{a_j > \tau(\ell)} d'_j > \sum_{a_j < \tau(\ell)} d'_j$ . Exprimons  $\sum_{a_j > \tau(\ell)} d'_j$  en fonction de l'image commutative de  $w$  :

$$\begin{aligned} \sum_{a_j > \tau(\ell)} d'_j &= \sum_{a_j > a_{r-k+1}} d'_j \\ &= d'_{r-k+2} + \dots + d'_r \\ &= d_{r-(r-k+2)+1} + \dots + d_{r-r+1} \\ &= d_{k-1} + \dots + d_1 \\ &= \sum_{a_j < \ell} d_j. \end{aligned}$$

Un calcul similaire montre que  $\sum_{a_j < \tau(\ell)} d'_j = \sum_{a_j > \ell} d_j$ . Par conséquent,

$$\sum_{a_j > \tau(\ell)} d'_j = \sum_{a_j < \ell} d_j > \sum_{a_j < \tau(\ell)} d'_j = \sum_{a_j > \ell} d_j.$$

Donc, le mot  $\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(w)$  est parfaitement amassant par le théorème 4.12. Le mot  $\rho_\ell(w)$  est parfaitement amassant puisqu'il est égal à  $\tilde{\tau} \circ \lambda_{\tau(\ell)} \circ \tilde{\tau}(w)$ .  $\blacksquare$

**Lemme 4.19.** *Si  $w$  est parfaitement amassant, alors le mot  $\Delta_\ell(w)$  est parfaitement amassant.*

*Démonstration.* Supposons que  $w$  est parfaitement amassant. Soit  $w_1 \prec \dots \prec w_n$  les conjugués de  $w$  selon l'ordre lexicographique. Les conjugués de  $\Delta_\ell(w)$  sont les mots  $\Delta_\ell(w_1), \dots, \Delta_\ell(w_n)$ , car  $\Delta_\ell$  est un morphisme strictement alphabétique.

De plus,  $w_i \prec w_j$  implique que  $\Delta_\ell(w_i) \prec \Delta_\ell(w_j)$ , car  $\Delta_\ell$  préserve l'ordre lexicographique. Les conjugués de  $\Delta_\ell(w)$  selon l'ordre lexicographique sont  $\Delta_\ell(w_1) \prec \Delta_\ell(w_2) \prec \dots \prec \Delta_\ell(w_n)$ . Appelons  $d_i$  la dernière lettre du mot  $w_i$ . Le mot  $w$  est parfaitement amassant si

$$w_i \prec w_j \Rightarrow d_i \succeq d_j$$

pour tout  $i, j \in \llbracket 1, |w| \rrbracket$ . Par conséquent, on a que

$$\Delta_\ell(w_i) \prec \Delta_\ell(w_j) \Rightarrow \Delta_\ell(d_i) \succeq \Delta_\ell(d_j)$$

pour tout  $i, j \in \llbracket 1, |w| \rrbracket$ . Donc, le mot  $\Delta_\ell(w)$  est parfaitement amassant.  $\blacksquare$

#### 4.5 Mots de Lyndon

Dans la construction de l'arbre des mots parfaitement amassants présentée à la section 4.2, les classes de conjugaison des mots sont représentées par l'unique mot de Lyndon appartenant à une classe. Ainsi, il semble naturel que les morphismes  $\lambda_\ell$ ,  $\rho_\ell$  et  $\Delta_\ell$  préservent les mots de Lyndon. Richomme (2003) a caractérisé les morphismes du monoïde  $\mathcal{A}^*$  vers le monoïde  $\mathcal{B}^*$  qui préserve les mots de Lyndon, qu'il appelle *morphisme de Lyndon*.

**Théorème 4.20** (Richomme (2003)). *Un morphisme  $f$  préserve les mots de Lyndon si et seulement s'il vérifie les deux propriétés suivantes :*

1.  $f(a)$  est un mot de Lyndon pour tout  $a \in \mathcal{A}$ ;
2.  $f$  est un morphisme croissant.

La démonstration de ce théorème est disponible dans Richomme (2003). Nous utilisons le théorème afin de démontrer que  $\Delta_\ell$  préserve les mots de Lyndon.

**Lemme 4.21.** *Si  $w$  est un mot de Lyndon, alors le mot  $\Delta_\ell(w)$  est un mot de Lyndon.*

*Démonstration.* Comme  $\Delta_\ell$  est un morphisme strictement alphabétique, le mot  $\Delta_\ell(a)$  est un mot de Lyndon pour tout  $a \in \mathcal{A}$ . De plus, la fonction  $\Delta_\ell(w)$  est croissante. En effet, ces fonctions préservent l'ordre total sur  $\mathcal{A}$ . Par conséquent,  $u \prec v$  implique  $\Delta_\ell(u) \prec \Delta_\ell(v)$ . Donc,  $\Delta_\ell(w)$  est un mot de Lyndon par le théorème 4.20. ■

Pour les morphismes  $\rho_\ell$  et  $\lambda_\ell$ , nous ne pouvons pas utiliser le théorème 4.20 puisque  $\rho_\ell$  et  $\lambda_\ell$  sont des morphismes de groupe et non pas de monoïde. Par contre, les idées pour démontrer ces résultats sont similaires à celles de Richomme (2003). Pour étudier cette question, nous restreignons les automorphismes  $\lambda_\ell$  et  $\rho_\ell$  de  $\mathcal{A}^*$  vers  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$ . Sur l'alphabet  $\mathcal{A} \cup \mathcal{A}^{-1}$ , considérons l'ordre total  $\{a_1 < a_1^{-1} < a_2 < a_2^{-1} < \dots < a_n < a_n^{-1}\}$ . Ainsi, l'ordre lexicographique sur le monoïde libre  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$  est une extension de l'ordre lexicographique sur le monoïde  $\mathcal{A}^*$ , c'est-à-dire que  $u \prec v$  sur  $\mathcal{A}^*$  implique  $u \prec v$  sur  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$ .

**Lemme 4.22.** *La fonction  $\rho_\ell$  de  $\mathcal{A}^*$  vers  $(\mathcal{A} \cup \mathcal{A}^{-1})^*$  est croissante.*

*Démonstration.* Soit  $u, v \in \mathcal{A}^*$  tel que  $u \prec v$ . Le mot  $u$  est un préfixe de  $v$  ou il existe des mots  $r, s, t \in \mathcal{A}^*$  et des lettres  $x \prec y \in \mathcal{A}$  tel que  $u = rxs$  et  $v = ryt$ .

1. Supposons que  $u$  est un préfixe de  $v$ , il existe un mot  $r$  appartenant  $\mathcal{A}^*$  tel que  $v = ur$ . L'image de  $\rho_\ell(v)$  est  $\rho_\ell(u)\rho_\ell(r)$ , car  $\rho_\ell$  est un morphisme. Si le produit  $\rho_\ell(u)\rho_\ell(r)$  est réduit, alors  $\rho_\ell(u)$  est un préfixe de  $\rho_\ell(v)$ , c'est-à-dire que  $\rho_\ell(u) \prec \rho_\ell(v)$ . Sinon,  $u = u'x$  et  $r = yr'$  avec  $x \preceq \ell$  et  $y \succ \ell$  des lettres de  $\mathcal{A}$  et  $u', r'$  sont des mots sur  $\mathcal{A}^*$ . Par conséquent, on a que  $\rho_\ell(u) = \rho_\ell(u')x\ell$  ou  $\rho_\ell(u) = \rho_\ell(u')\ell$  et  $\rho_\ell(v) = \rho_\ell(u')x\ell\ell^{-1}y\rho_\ell(r') = \rho_\ell(u')xy\rho_\ell(r')$  ou  $\rho_\ell(v) = \rho_\ell(u')\ell\ell^{-1}y\rho_\ell(r') = \rho_\ell(u')y\rho_\ell(r')$ . Les produits  $\rho_\ell(u')x\ell$ ,  $\rho_\ell(u')\ell$  et  $\rho_\ell(u')y$  sont réduits, car  $\rho_\ell(u')$  termine par une lettre positive. Le produit

$\ell^{-1}y\rho_\ell(r')$  est réduit, car  $y \neq \ell$ . Comme les produits  $\rho_\ell(u')x\ell$ ,  $\rho_\ell(u')\ell$   $\rho_\ell(u')y$  et  $\ell^{-1}y\rho_\ell(r')$  sont réduits et que  $x \prec y$ , on a que  $\rho_\ell(u) \prec \rho_\ell(v)$ .

2. Il existe des mots  $r, s, t \in \mathcal{A}^*$  et des lettres  $x \prec y$  appartenant à  $\mathcal{A}$  tel que  $u = rxs$  et  $v = ryt$ . On a alors que

$$\rho_\ell(u) = \rho_\ell(r)\rho_\ell(x)\rho_\ell(s) \quad (4.4)$$

$$\rho_\ell(v) = \rho_\ell(r)\rho_\ell(y)\rho_\ell(t). \quad (4.5)$$

La relation entre  $\rho_\ell(u)$  et  $\rho_\ell(v)$  dépend des valeurs de  $x$  et  $y$ .

- Si  $x \prec \ell$  et  $y \prec \ell$ , les produits  $\rho_\ell(u)\rho_\ell(x)$  et  $\rho_\ell(v)\rho_\ell(y)$  sont réduits. Par conséquent,  $\rho_\ell(x) = x\ell \prec y\ell = \rho_\ell(y)$  et  $\rho_\ell(u) \prec \rho_\ell(v)$ .
- Si  $x \prec \ell$  et  $y = \ell$ , les produits  $\rho_\ell(r)\rho_\ell(x)$  et  $\rho_\ell(r)\rho_\ell(y)$  sont réduits. Par conséquent,  $\rho_\ell(x) = x\ell \prec y = \rho_\ell(y)$  et donc  $\rho_\ell(u) \prec \rho_\ell(v)$ .
- Si  $x \preceq \ell$  et  $y \succ \ell$ , le produit  $\rho_\ell(r)\rho_\ell(x)$  est réduit, mais le produit  $\rho_\ell(r)\rho_\ell(y)$  n'est pas nécessairement réduit. Si  $\rho_\ell(r)\rho_\ell(y)$  est réduit, alors  $\rho_\ell(u) \prec \rho_\ell(v)$  par un raisonnement analogue au cas précédent. Sinon, il existe un mot  $r' \in \mathcal{A}^*$  et une lettre  $z \in \mathcal{A}$  tels que  $z \preceq \ell$ . Par conséquent,  $\rho_\ell(r) = \rho_\ell(r')\rho_\ell(z) = \rho_\ell(r')\ell$  ou  $\rho_\ell(r')z\ell$ . Le produit  $\rho_\ell(r')\rho_\ell(z)$  est réduit, car la dernière lettre de  $\rho_\ell(r')$  est positive. Si  $\rho_\ell(r) = \rho_\ell(r')\ell$ , alors l'équation 4.5 est  $\rho_\ell(v) = \rho_\ell(r')y\rho_\ell(t)$  et l'équation 4.4 est  $\rho_\ell(u) = \rho_\ell(r')\ell\rho_\ell(x)\rho_\ell(s)$ . Les lettres à comparer sont  $\ell$  et  $y$ . Si  $\rho_\ell(r) = \rho_\ell(r')z\ell$ , alors l'équation 4.5 est  $\rho_\ell(v) = \rho_\ell(r')zy\rho_\ell(t)$  et l'équation 4.4 est  $\rho_\ell(u) = \rho_\ell(r')z\ell\rho_\ell(x)\rho_\ell(s)$ . Les lettres à comparer sont aussi  $\ell$  et  $y$ . Donc,  $\rho_\ell(u) \prec \rho_\ell(v)$ , car  $y \succ \ell$ .
- Si  $x \succ \ell$  et  $y \succ \ell$ , alors les produits,  $\rho_\ell(r)\rho_\ell(x)$  et  $\rho_\ell(r)\rho_\ell(y)$  ne sont pas nécessairement réduits. S'ils sont réduits, alors  $\rho_\ell(u) \prec \rho_\ell(v)$  par le même raisonnement que les autres cas. Sinon il existe un mot  $r' \in \mathcal{A}^*$  et une lettre  $z \in \mathcal{A}$  tel que  $z \preceq \ell$ . Par conséquent,  $\rho_\ell(r) = \rho_\ell(r')\rho_\ell(z) = \rho_\ell(r')\ell$

ou  $\rho_\ell(r')z\ell$ . Les produits  $\rho_\ell(r')\ell$  et  $\rho_\ell(r')z\ell$  sont réduits, car la dernière lettre du mot  $\rho_\ell(r')$  est positive. Si  $\rho_\ell(r) = \rho_\ell(r')\ell$ , alors l'équation 4.5 est  $\rho_\ell(v) = \rho_\ell(r')y\rho_\ell(t)$  et l'équation 4.4 est  $\rho_\ell(u) = \rho_\ell(r')x\rho_\ell(x)\rho_\ell(s)$ . Les lettres à comparer sont  $x$  et  $y$ . Si  $\rho_\ell(r) = \rho_\ell(r')z\ell$ , alors l'équation 4.5 est  $\rho_\ell(v) = \rho_\ell(r')zy\rho_\ell(t)$  et l'équation 4.4 est  $\rho_\ell(u) = \rho_\ell(r')zx\rho_\ell(s)$ . Les lettres à comparer sont aussi  $x$  et  $y$ . Donc,  $\rho_\ell(u) \prec \rho_\ell(v)$  parce que  $x \prec y$ .

Donc, si  $u \prec v$  avec  $u, v \in \mathcal{A}^*$ ,  $\rho_\ell(u) \prec \rho_\ell(v)$ . ■

**Remarque 4.23.** *La restriction aux éléments positifs de  $F(\mathcal{A})$  est importante sinon le lemme 4.22 ne tient pas. En effet, posons  $u = ababc$  et  $v = aba^{-1}bc$  des éléments de  $F(\{a, b, c\})$  où  $a < b < c$ . Le mot  $u$  est plus petit que  $v$ , mais  $\rho_b(ababc) = abbabbc \succ \rho_b(aba^{-1}bc) = aba^{-1}bbc$ .*

**Proposition 4.24.** *Soit  $u \in \mathcal{A}^*$  un mot de Lyndon. Si  $\rho_\ell(u)$  est positif, alors  $\rho_\ell(u)$  est un mot de Lyndon.*

*Démonstration.* Désignons par  $S$  l'ensemble des suffixes propres de  $\rho_\ell(u)$ . À partir de  $u$ , on construit l'ensemble  $R$  de la façon suivante :  $r \in R$  s'il existe des mots  $u_1, u_2 \in \mathcal{A}^+$  et des lettres  $x, p, s \in \mathcal{A}$  tels que

- $r = \rho_\ell(u_2)$  avec  $u = u_1u_2$  ;
- $r = s\rho_\ell(u_2)$  avec  $u = u_1xu_2$  et  $\rho_\ell(x) = ps$ .

Le mot  $\rho_\ell(u)$  est un mot de Lyndon si pour tout  $s \in S$ ,  $\rho_\ell(u) \prec s$  par la proposition 1.6. Comme l'ensemble  $S$  est un sous-ensemble de  $R$ , il suffit de démontrer que pour tout  $r \in R$ , on a que  $\rho_\ell(u) \prec r$ .

- Supposons qu'il existe  $u_1, u_2 \in \mathcal{A}^+$  tel que  $r = \rho_\ell(u_2)$  avec  $u = u_1u_2$ . Puisque  $u$  est un mot de Lyndon, on a que  $u \prec u_2$ . Par le lemme 4.22, on que  $\rho_\ell(u) \prec \rho_\ell(u_2) = r$ .

- Supposons qu'il existe  $u_1, u_2 \in A^+$ ,  $x \in \mathcal{A}$  et  $p, s \in \mathcal{A} \cup \mathcal{A}^{-1}$  tel que  $r = s\rho_\ell(u_2)$  avec  $u = u_1xu_2$  et  $\rho_\ell(x) = ps$ . On a que  $\rho_\ell(x) \prec s$ , car  $\rho_\ell(x) = x\ell$  si  $x \prec \ell$  ou  $\rho_\ell(x) = \ell^{-1}x$  si  $x \succ \ell$ . Puisque  $s$  n'est pas un préfixe de  $\rho_\ell(x)$ , on a que  $\rho_\ell(x)\rho_\ell(u_2) \prec s\rho_\ell(u_2)$ . De plus, on a que  $u \prec xu_2$ , car c'est un mot de Lyndon. Par le lemme 4.22, on a que

$$\rho_\ell(u) \prec \rho_\ell(xu_2) \prec s\rho_\ell(u_2) = r.$$

Par conséquent, pour tout  $r \in R$ , on a que  $\rho_\ell(u) \prec r$ . Ainsi, pour tout  $s \in S \subseteq R$ , on a que  $\rho_\ell(u) \prec s$ . Donc,  $\rho_\ell(u)$  est bel et bien un mot de Lyndon. ■

Contrairement à la fonction  $\rho_\ell$ , la fonction  $\lambda_\ell$  n'est pas croissante. En effet, pour les mots  $u = abacaacab$  et  $v = abacabacb$ , on a que  $u \prec v$ , mais  $\lambda_b(u) = aacab^{-1}aca \succ \lambda_b(v) = aacaacb$ . Cette absence de symétrie entre  $\lambda_\ell$  et  $\rho_\ell$  vient de l'ordre lexicographique qui n'est pas symétrique. Malgré cela si  $u$  est un mot de Lyndon parfaitement amassant, alors  $\lambda_\ell(u)$  est un mot de Lyndon. Par exemple,  $acbcacc$  est un mot de Lyndon parfaitement amassant et  $\lambda_b(acbcacc) = acbbcacbc$  est un mot de Lyndon.

**Proposition 4.25.** *Si  $w$  est un mot de Lyndon parfaitement amassant, alors  $\tilde{\tau}(w)$  est un mot de Lyndon.*

*Démonstration.* Le mot  $w$  est un mot de Lyndon, alors pour tout  $u$  conjugué de  $w$  tel que  $w \neq u$ , on a que  $w \prec u$ . Pour tout  $\tau(u)$  conjugué à  $\tau(w)$ , on a que  $\tau(w) \succ \tau(u)$ , par le lemme 4.3. Par conséquent,  $\tau(w)$  est le plus grand mot de sa classe de conjugaison. L'image miroir est décroissante pour l'ordre alphabétique sur les mots parfaitement amassants, par le lemme 2.26, alors  $\tilde{\tau}(w)$  est un mot de Lyndon. ■

**Proposition 4.26.** *Soit  $w$  un mot de Lyndon parfaitement amassant. Si  $\lambda_\ell(w)$  est positif, alors  $\lambda_\ell(w)$  est un mot de Lyndon.*

*Démonstration.* Par la proposition 4.4, on a que  $\lambda_\ell(w) = \tilde{\tau} \circ \rho_{\tau(\ell)} \circ \tilde{\tau}(w)$ . De plus, le mot  $\tilde{\tau} \circ \rho_{\tau(\ell)} \circ \tilde{\tau}(w)$  est un mot de Lyndon par les propositions 4.24 et 4.25. Donc,  $\lambda_\ell(w)$  est un mot de Lyndon. ■

#### 4.6 Démonstration du théorème 4.9

Après avoir démontré diverses propriétés des mots engendrés par les différents morphismes nécessaires à la construction de l'arbre, on peut maintenant démontrer que l'arbre des mots parfaitement amassants contient tous les mots de Lyndon parfaitement amassants complets. Rappelons l'énoncé du théorème.

**Théorème 4.9.** *Tous les mots de Lyndon parfaitement amassants complets apparaissent exactement une fois dans l'arbre des mots parfaitement amassants.*

**Lemme 4.27.** *Si  $w$  est complet, alors  $S_i(w)$  est complet pour  $i \in \llbracket 0, 2|\text{Alph}(w)| - 1 \rrbracket$ .*

*Démonstration.* Posons  $(d_1, \dots, d_r)$  l'image commutative du mot  $w$ . L'image commutative de  $S_i(w)$  est

1.  $(d_1, \dots, d_{k-1}, \sum_{j \geq k} d_j - \sum_{j < k} d_j, d_k, \dots, d_r)$  si  $i$  est pair et  $g > p$ ;
2.  $(d_1, \dots, d_{k-1}, d_k + \sum_{j > k} d_j - \sum_{j < k} d_j, d_{k+1}, \dots, d_r)$  si  $i$  est impair et  $g > p$ ;
3.  $(d_1, \dots, d_{k-1}, d_k + \sum_{j < k} d_j - \sum_{j > k} d_j, d_{k+1}, \dots, d_r)$  si  $i$  est pair et  $g < p$ ;
4.  $(d_1, \dots, d_k, \sum_{j \leq k} d_j - \sum_{j > k} d_j, d_{k+1}, \dots, d_r)$  si  $i$  est impair et  $g < p$ ;

par l'équation (4.1) où  $\ell = a_k$ ,  $g = \sum_{j > k} d_j$  et  $p = \sum_{j < k} d_j$ . Les  $d_i$  sont non nuls, car  $w$  est complet. De plus, les conditions sur  $g$  et  $p$  assurent que la somme des  $d_i$  est positive et non nulle. Par conséquent, le mot  $S_i(w)$  est complet. ■

**Théorème 4.28.** *L'arbre des compositions circulaires et l'arbre des mots parfaitement amassants sont isomorphes.*

*Démonstration.* Soit  $K$  l'arbre des mots parfaitement amassants où chacun des mots est remplacé par son image commutative. Démontrons que l'arbre  $K$  est isomorphe à l'arbre des compositions circulaires. En effet, les deux arbres ont la même racine, car l'image commutative du mot  $ab$  est la composition  $(1, 1)$ . De plus, l'image commutative des mots parfaitement amassants est une composition circulaire par le théorème 3.8. Soit  $(d_1, \dots, d_r)$  une composition circulaire de longueur  $n$  apparaissant dans l'arbre  $K$ . Elle apparaît aussi dans l'arbre des compositions circulaires par le théorème 3.23. La composition  $(d_1, \dots, d_r)$  a  $2r$  enfants dans les deux arbres par la proposition 3.27. Il faut démontrer que les enfants de  $(d_1, \dots, d_r)$  dans  $K$  et dans l'arbre des compositions circulaires sont les mêmes. L'image commutative des enfants de  $w$  dans  $K$  correspond à l'image commutative de  $S_i(w)$  que nous avons décrite ci-dessus. Les cas 2 et 3 correspondent aux enfants de forme 1 de l'arbre des compositions circulaires tandis que les cas 1 et 4 correspondent aux enfants de forme 2 de l'arbre des compositions circulaires. Pour les cas 1 et 4, il faut aussi démontrer que  $\delta_i > d_{i+1}$  ou  $-\delta_i > d_i$  est équivalent à  $g > p$  ou  $g < p$ . Or, par la définition de  $\delta_i$ , on a que

$$\delta_i > d_{i+1} \Rightarrow \sum_{j>i} d_j - \sum_{j<i} d_j - d_i > d_{i+1} \quad (4.6)$$

$$\Rightarrow \sum_{j>i+1} d_j + d_{i+1} - \sum_{j<i+1} d_j > d_{i+1} \quad (4.7)$$

$$\Rightarrow \sum_{j>i+1} d_j > \sum_{j<i+1} d_j. \quad (4.8)$$

De plus, on a que  $g = \sum_{j>i} d_j = \sum_{j>i+1} d_j + d_{i+1} > \sum_{j>i+1} d_j$ , car  $d_{i+1} > 0$ . La  $\sum_{j<i+1} d_j = \sum_{j<i} d_j + d_i > p$ , car  $d_i > 0$ . Par conséquent, l'inégalité 4.8 devient

$$g > \sum_{j>i+1} d_j > \sum_{j<i+1} d_j > p.$$

Par la définition de  $\delta_i$ , on a que

$$\begin{aligned}
 -\delta_i > d_i &\Rightarrow \sum_{j < i} d_j + d_i - \sum_{j > i} d_j > d_i \\
 &\sum_{j < i} d_j - \sum_{j > i} d_j > 0 \\
 &\sum_{j < i} d_j > \sum_{j > i} d_j \\
 &g < p.
 \end{aligned}$$

Ainsi,  $\delta_i > d_{i+1}$  est équivalent à  $g > p$  et  $-\delta_i > d_i$  est équivalent à  $g < p$ . Donc, les enfants de  $w$  dans  $K$  et l'arbre des compositions circulaire sont les mêmes compositions. L'arbre  $K$  et celui des compositions circulaires sont isomorphes. ■

*Démonstration du théorème 4.9.* Le mot  $ab$  est un mot de Lyndon parfaitement amassant complet. En utilisant les théorèmes 4.12 et 4.18, on a que tous les mots de l'arbre sont parfaitement amassants. De plus, le mot  $ab$  est un mot de Lyndon. Ainsi, tous les mots de l'arbre sont des mots de Lyndon par les propositions 4.24 et 4.26. Tous les mots dans l'arbre des mots parfaitement amassants sont complets par le lemme 4.27. Pour démontrer que tous les mots de Lyndon parfaitement amassants apparaissent exactement une fois dans l'arbre, utilisons l'arbre des compositions circulaires puisque les deux arbres sont isomorphes par le théorème 4.28. Toutes les compositions circulaires apparaissent exactement une fois dans l'arbre des compositions circulaires par le théorème 3.23. Donc, tous les mots de Lyndon parfaitement amassants complets apparaissent exactement une fois dans l'arbre des mots parfaitement amassants. ■

Revenons à l'étude des mots à l'aide des morphismes.

**Théorème 4.29.** *Soit  $w$  un mot de Lyndon complet parfaitement amassant sur l'alphabet totalement ordonné  $\mathcal{A}$ . Il existe une suite de morphismes du groupe libre*

notée  $\chi = \chi_1 \circ \dots \circ \chi_k$  avec  $\chi_i \in \{\lambda_\ell, \rho_\ell, \lambda_\ell \circ \Delta_\ell, \rho_{N(\ell)} \circ \Delta_{N(\ell)}\}$  telle que  $\chi(ab) = w$ .

**Exemple 4.30.** Le mot  $adbbdadbcdbd$  est un mot de Lyndon parfaitement amassant complet sur l'alphabet  $\{a, b, c, d\}$  où  $a < b < c < d$ . La suite  $\chi$  telle que  $\chi(ab) = adbbdadbcdbd$  est  $\chi = \lambda_b \circ (\rho_{N(c)} \circ \Delta_{N(c)}) \circ (\lambda_a \circ \Delta_a)$ .

*Démonstration.* Le mot  $w$  apparaît dans l'arbre des mots de Lyndon parfaitement amassant par le théorème 4.9. Alors il existe une suite de morphismes  $\chi = \chi_1 \circ \dots \circ \chi_k$  avec  $\chi_i \in \{\lambda_\ell, \rho_\ell, \lambda_\ell \circ \Delta_\ell, \rho_{N(\ell)} \circ \Delta_{N(\ell)}\}$  telle que  $\chi(ab) = w$  par la définition de l'arbre des mots parfaitement amassants. ■

#### 4.7 Primitivité sur le groupe libre

Tout comme les mots de Christoffel, les mots parfaitement amassants sont des éléments primitifs positifs du groupe libre. La notion de primitivité sur le groupe libre n'est pas équivalente à celle sur le monoïde libre. En fait, un élément primitif sur le groupe libre est primitif sur le monoïde libre, mais le contraire n'est pas nécessairement vrai. Une *base* ordonnée de  $F(\mathcal{A})$  est un  $|\mathcal{A}|$ -uplet  $u = (u_1, \dots, u_{|\mathcal{A}|})$  dont la fonction de  $\mathcal{A}$  vers  $u$  associant  $a_i$  à  $u_i$  s'étend en un automorphisme de  $F(\mathcal{A})$ . Un *automorphisme* est une fonction bijective qui envoie une base du groupe libre sur une autre base. Par la propriété universelle du groupe libre, une fonction bijective de  $F(\mathcal{A})$  envoyant  $\mathcal{A}$  sur une autre base est un automorphisme. Un élément  $g$  du groupe libre  $F(\mathcal{A})$  est *primitif* s'il existe des éléments  $h_1, \dots, h_{|\mathcal{A}|-1}$  tels que  $(g, h_1, \dots, h_{|\mathcal{A}|-1})$  forme une base ordonnée de  $F(\mathcal{A})$ .

**Proposition 4.31.** Les morphismes  $\lambda_\ell$  et  $\rho_\ell$  sont des automorphismes du groupe libre  $F(\mathcal{A})$ .

*Démonstration.* Soit  $\mathcal{A}$  un alphabet totalement ordonné de cardinalité  $r$  et le groupe libre avec base  $\mathcal{A}$ ,  $F(\mathcal{A})$ . Le morphisme  $\lambda_\ell$  est une fonction bijective de

$F(\mathcal{A})$  par la proposition 4.2. De plus,  $u = (a_1\ell^{-1}, \dots, a_j\ell^{-1}, \ell, \ell a_{j+2} \dots, \ell a_n)$  est une base ordonnée de  $F(\mathcal{A})$ , car chacune des lettres  $a \in \mathcal{A}$  s'écrit comme un produit d'éléments de  $u$ . En effet, si  $x = \ell$ , alors  $x \in u$ ; si  $x \prec \ell$ , alors  $x = (x\ell^{-1})\ell$ ; sinon  $x \succ \ell$  et  $x = \ell^{-1}(\ell x)$ . Donc, le morphisme  $\lambda_\ell$  est un automorphisme du groupe libre  $F(\mathcal{A})$ . Un argument similaire prouve que  $\rho_\ell$  est aussi un automorphisme du groupe libre  $F(\mathcal{A})$ . ■

Le morphisme  $\Delta_\ell$  n'est pas un automorphisme, mais il préserve les éléments primitifs du groupe libre. On note que le morphisme  $\Delta_\ell$  est une restriction à  $\mathcal{A}^*$  d'un automorphisme de  $F(\mathcal{B})$ . Étudions quelques propriétés des éléments primitifs.

**Lemme 4.32.** *Tout élément primitif du groupe libre  $F(\mathcal{A})$  est un élément primitif du groupe libre  $F(\mathcal{B})$  avec  $\mathcal{A} \subseteq \mathcal{B}$ .*

*Démonstration.* Posons  $\mathcal{A} = \{a_1, \dots, a_r\}$ . Comme  $\mathcal{A} \subseteq \mathcal{B}$ , on peut écrire que  $\mathcal{B} = \{a_1, \dots, a_r, b_1, \dots, b_s\}$ . Soit  $u$  un élément primitif de  $F(\mathcal{A})$ , alors il existe des éléments  $h_1, \dots, h_{r-1}$  tels que  $(u, h_1, \dots, h_r)$  forme une base de  $F(\mathcal{A})$ . De plus, les éléments  $u$  et  $h_i$  sont tous des éléments de  $F(\mathcal{B})$  contenant seulement les lettres  $a_1, \dots, a_r$ , car  $\mathcal{A} \subseteq \mathcal{B}$ . Par conséquent,  $(u, h_1, \dots, h_{r-1}, b_1, \dots, b_s)$  forme une base de  $F(\mathcal{B})$ . Donc,  $u$  est un élément primitif de  $F(\mathcal{B})$ . ■

**Proposition 4.33.** *Soit  $\mathcal{A}$  l'alphabet totalement ordonné de cardinalité  $r$ ,  $b \notin \mathcal{A}$  et  $\mathcal{B} = \mathcal{A} \cup \{b\}$  tel que pour tout  $a \in \mathcal{A}$ ,  $a \prec b$ . Si  $w$  est un élément primitif de  $F(\mathcal{A})$ , alors les éléments  $\lambda_\ell(w)$ ,  $\rho_\ell(w)$ ,  $\lambda_\ell \circ \Delta_\ell(w)$  et  $\rho_{N(\ell)} \circ \Delta_{N(\ell)}(w)$  sont des éléments primitifs de  $F(\mathcal{B})$ .*

*Démonstration.* Puisque  $\lambda_\ell$  et  $\rho_\ell$  sont des automorphismes de  $F(\mathcal{A})$  par le corollaire 4.31, on a que  $\lambda_\ell(w)$  et  $\rho_\ell(w)$  sont des éléments primitifs de  $F(\mathcal{A})$  et de  $F(\mathcal{B})$  par le lemme 4.32.

Démontrons que  $\lambda_\ell \circ \Delta_\ell(w)$  est un élément primitif de  $F(\mathcal{B})$ . La fonction  $\lambda_\ell \circ \Delta_\ell$  de  $F(\mathcal{A})$  vers  $F(\mathcal{B})$  est définie par  $\lambda_\ell \circ \Delta_\ell(a) = a\ell^{-1}$  si  $a \prec \ell$  et  $\lambda_\ell \circ \Delta_\ell(a) = \ell N(a)$  sinon. Définissons  $\beta : F(\mathcal{B}) \rightarrow F(\mathcal{B})$  par  $\beta(a) = \lambda_\ell \circ \Delta_\ell(a)$  si  $a \neq b$  et  $\beta(b) = b$ . On a alors que  $\lambda_\ell \circ \Delta_\ell = \beta \circ i$  où  $i$  est l'inclusion naturelle de  $\mathcal{A}$  dans  $\mathcal{B}$ . Montrons que  $u = (\lambda_\ell \circ \Delta_\ell(a_1), \dots, \lambda_\ell \circ \Delta_\ell(a_r), b)$  forme une base ordonnée de  $F(\mathcal{B})$ . Si tous les éléments  $x \in \mathcal{B}$  peuvent s'écrire en un produit d'éléments de  $u$ , alors  $u$  est une base ordonnée de  $F(\mathcal{B})$ . Si  $x \prec \ell$ , on a que  $x = (\lambda_\ell \circ \Delta_\ell(x)) \cdot (\lambda_\ell \circ \Delta_\ell(a_r))(b^{-1})$ . Si  $x \preceq \ell$ , on a que  $x = (\lambda_\ell \circ \Delta_\ell(x)) \cdot (\lambda_\ell \circ \Delta_\ell(a_r))(b^{-1})$ . Donc,  $u$  est une base ordonnée de  $F(\mathcal{B})$ . Par conséquent,  $\beta$  est un automorphisme de  $F(\mathcal{B})$ , c'est-à-dire qu'il préserve les éléments primitifs de  $F(\mathcal{B})$ . Comme  $w$  est primitif sur  $F(\mathcal{A})$ , il est primitif sur  $F(\mathcal{B})$  par le lemme 4.32. Donc  $\Delta_\ell(w)$  est primitif, car  $\beta \circ i(w)$  est primitif. ■

**Théorème 4.34.** *Tout mot parfaitement amassant complet sur l'alphabet totalement ordonnée  $\mathcal{A}$  est un élément primitif de  $F(\mathcal{A})$ .*

Avant de faire la démonstration, introduisons la relation de conjugaison sur le groupe libre  $F(\mathcal{A})$ . Les éléments  $u, v \in F(\mathcal{A})$  sont dits *conjugués* s'il existe  $g \in F(\mathcal{A})$  tel que  $u = gv g^{-1}$ . Bien que la définition de conjugués sur le groupe libre est différente que celle sur le monoïde libre, on peut démontrer que les deux notions sont équivalentes.

*Démonstration du théorème 4.34.* Supposons que  $w$  soit un mot de Lyndon parfaitement amassant complet. Il existe une suite de morphismes du groupe libre  $F(\mathcal{A})$  telle que  $\chi(ab) = w$  par le théorème 4.29. De plus, le mot  $ab$  est un élément primitif de  $F(\{a, b\})$ , car  $(ab, b)$  est une base de ce groupe libre. Le mot  $ab$  est donc un élément primitif de  $F(\mathcal{A})$  en supposant que  $\{a < b\} \subseteq \mathcal{A}$ . De plus, tous les morphismes de  $\chi$  préservent les éléments primitifs, par le lemme 4.32. Donc,  $w$  est un élément primitif de  $F(\mathcal{A})$ .

Si le mot  $w$  n'est pas un mot de Lyndon, alors il est conjugué à un mot de Lyndon puisque les mots parfaitement amassants sont primitifs. On note  $v$  le conjugué de  $w$  qui est un mot de Lyndon. Comme  $w$  et  $v$  sont conjugués, il existe deux mots  $x$  et  $y$  appartenant à  $\mathcal{A}^*$  tel que  $w = xy$  et  $v = yx$ . Ainsi,  $ywy^{-1} = y(xy)y^{-1} = yx = v$ . Donc,  $w$  et  $v$  sont conjugués dans  $F(\mathcal{A})$ . De plus,  $v$  est un élément primitif de  $F(\mathcal{A})$ , alors il existe une base ordonnée  $(v, h_1, \dots, h_r)$  de  $F(\mathcal{A})$  avec  $h_i \in F(\mathcal{A})$ . Ainsi,  $(y^{-1}vy, y^{-1}h_1y, \dots, y^{-1}h_ry) = (w, y^{-1}h_1y, \dots, y^{-1}h_ry)$  (la conjugaison est un automorphisme de groupe) est une base ordonnée de  $F(\mathcal{A})$ . Donc,  $w$  est un élément primitif de  $F(\mathcal{A})$ . ■

Ce théorème est une généralisation d'une propriété des mots de Christoffel qui sont des éléments primitifs de  $F(\{a, b\})$ .

**Corollaire 4.35** (Osborne et Zieschang (1981); Kassel et Reutenauer (2007)).  
*Tous les conjugués des mots de Christoffel sont des éléments primitifs positifs du groupe libre  $F(\{a, b\})$ .*

Les articles de Osborne et Zieschang (1981) et Kassel et Reutenauer (2007) démontrent en plus que tous les éléments primitifs du groupe libre  $F(\{a, b\})$  sont des conjugués des mots de Christoffel. Est-ce que tous les éléments primitifs positifs du groupe libre  $F(\mathcal{A})$  sont parfaitement amassants? La réponse à cette question reste ouverte.

## CHAPITRE V

### MOTS DE CHRISTOFFEL ET CONJECTURE DE FROBENIUS

Les mots de Christoffel sont en bijection avec des triplets de nombres naturels appelés triplets de Markoff. De plus, les mots de Christoffel sont obtenus par la palindromisation itérée des mots sur le monoïde libre. Dans ce chapitre, une bijection entre le monoïde libre  $\{a, b\}^*$  et les triplets de Markoff basée sur les palindromes itérés est présentée. Cette bijection suggère une nouvelle notion de dualité pour les mots de Christoffel.

La conjecture de Frobenius porte sur l'unicité des nombres de Markoff. Des formulations équivalentes de cette conjecture sont connues dans différents contextes mathématiques, dont les mots de Christoffel. Nous présentons des généralisations de la conjecture de Frobenius sur les mots sturmiens et sur les conjugués des mots de Christoffel ainsi que des résultats partiels sur ces conjectures.

Les résultats des sections [5.2](#) et [5.3](#) ont été publiés dans [Abram \*et al.\* \(2020\)](#).

## 5.1 Nombres de Markoff

Un *triplet de Markoff* est un multi-ensemble à trois éléments,  $\{x, y, z\}$ , d'entiers positifs qui sont une solution de l'équation diophantienne

$$x^2 + y^2 + z^2 = 3xyz. \quad (5.1)$$

Un triplet de Markoff est dit *propre* si tous ses éléments sont distincts.

**Exemple 5.1.** *Les multi-ensembles  $\{1, 1, 1\}$ ,  $\{1, 1, 2\}$ ,  $\{1, 2, 5\}$  et  $\{5, 433, 29\}$  sont tous des triplets de Markoff, mais seuls les deux derniers sont des triplets de Markoff propres.*

La bijection qui nous intéresse entre les nombres de Markoff et les mots de Christoffel est construite à partir d'une fonction du monoïde libre  $\{a, b\}^*$  vers le groupe  $\mathrm{SL}_2(\mathbb{N})$ . Plus précisément, soit  $\mu$  le morphisme du monoïde libre  $\{a, b\}^*$  vers  $\mathrm{SL}_2(\mathbb{N})$  défini par

$$\mu(a) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mu(b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}.$$

Désignons par  $\mu(w)_{12}$  l'entrée sur la première ligne et la deuxième colonne de la matrice  $\mu(w)$ . L'image d'un mot de Christoffel par la fonction  $\mu(\cdot)_{12}$  est un nombre de Markoff.

**Théorème 5.2** (Bombieri (2007); Reutenauer (2009)). *La fonction qui associe à un mot de Christoffel  $w$  dont la factorisation standard est  $uv$  le multi-ensemble*

$$\{\mu(u)_{12}, \mu(w)_{12}, \mu(v)_{12}\}$$

*est une bijection entre l'ensemble des mots de Christoffel propres et l'ensemble des triplets de Markoff propres.*

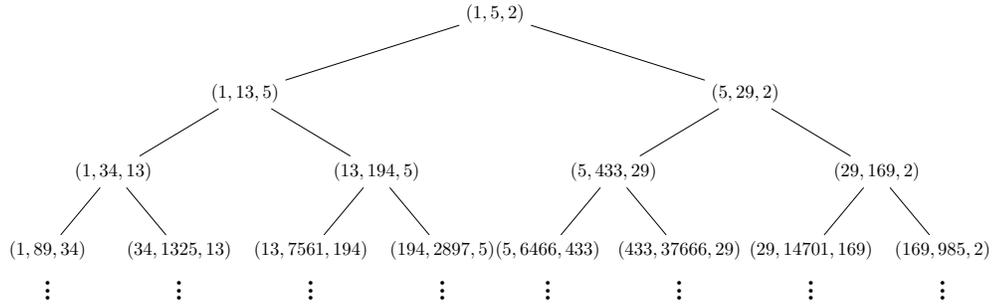


Figure 5.1: Arbre des triplets de Markoff

Le théorème est énoncé ici sans la démonstration. Celle-ci est présentée dans le livre de [Reutenauer \(2019\)](#).

**Exemple 5.3.** *Le mot  $ababb$  est un mot de Christoffel dont la factorisation standard est  $ab \cdot abb$ . L'image du mot  $ababb$  par la fonction  $\mu$  est*

$$\mu(ababb) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1045 & 433 \\ 613 & 254 \end{pmatrix}.$$

*Le triplet  $\{\mu(ab)_{12}, \mu(ababb)_{12}, \mu(abb)_{12}\} = \{5, 433, 29\}$  est un triplet de Markoff.*

Cette bijection suggère une autre façon d'étiqueter l'arbre binaire infini en remplaçant un mot de Christoffel  $w$  par le triplet de Markoff correspondant, c'est-à-dire  $(\mu(u)_{12}, \mu(w)_{12}, \mu(v)_{12})$  comme à la figure 5.1. L'arbre obtenu ainsi est appelé *l'arbre des triplets de Markoff*. Puisque les mots de Christoffel apparaissent une seule fois dans l'arbre des mots de Christoffel, chaque triplet de Markoff apparaît une fois dans l'arbre des triplets de Markoff. Il est facile de voir que, dans l'arbre des triplets de Markoff, la valeur au centre d'un triplet est le maximum du multi-ensemble. L'arbre des triplets de Markoff peut aussi être construit directement à partir de l'équation diophantienne (5.1). La racine est étiquetée par le triplet  $(1, 5, 2)$  et la règle de construction récursive de l'arbre est illustrée par la figure 5.2.

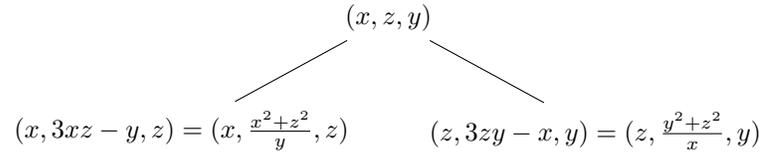


Figure 5.2: Règle de construction récursive de l'arbre des triplets de Markoff

Un *nombre de Markoff* est un élément d'un triplet de Markoff. Les 13 premiers nombres de Markoff sont

$$1, 2, 5, 13, 29, 34, 89, 169, 194, 233, 433, 610, 985$$

et la suite est disponible dans ([OEIS](#), 2020).

Le théorème 5.2 implique que pour chaque nombre de Markoff  $m$ , il existe un mot de Christoffel  $w$  tel que  $m = \mu(w)_{12}$ . Par conséquent, la fonction  $w \mapsto \mu(w)_{12}$  de l'ensemble des mots de Christoffel vers l'ensemble des nombres de Markoff est surjective. La *conjecture de Frobenius*, aussi appelé *conjecture d'injectivité de Markoff*, porte sur l'injectivité de cette fonction.

**Conjecture 5.4** (Frobenius 1913). *La fonction  $w \mapsto \mu(w)_{12}$  est injective sur l'ensemble des mots de Christoffel.*

Il convient de mentionner que ce n'est pas la formulation originale de la conjecture, mais une formulation équivalente découlant du théorème 5.2. Voici la conjecture originale : chaque nombre de Markoff est l'élément maximal d'un unique triplet de Markoff. Le livre d'([Aigner](#), 2013) présente plusieurs formulations équivalentes de cette conjecture.

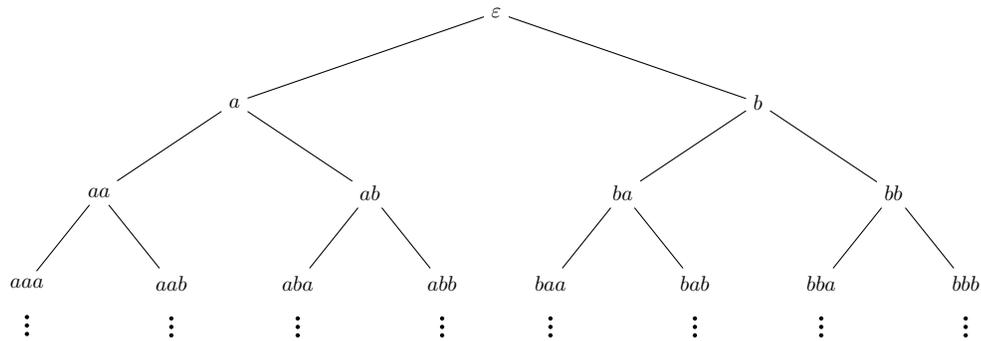


Figure 5.3: Arbres du monoïde libre

## 5.2 Une bijection entre le monoïde libre et les triplets de Markoff

On rappelle que  $\theta = (ab, ba)$  est le morphisme de Thue-Morse. [Reutenauer et Vuillon \(2017\)](#) ont démontré que pour chaque nombre de Markoff  $m$ , il existe un mot  $v \in \{a, b\}^*$  tel que l'image de la fonction  $\pi$  définie par  $\pi(v) = |\text{Pal} \circ \theta \circ \text{Pal}(av)| + 2$ , est le nombre  $m$ . Par exemple,  $\pi(a) = |\text{Pal} \circ \theta \circ \text{Pal}(aa)| + 2 = |\text{abaababaaba}| + 2 = 13$  qui est un nombre de Markoff. Dans cette section, une construction des triplets de Markoff inspirée par l'approche de Reutenauer et Vuillon est présentée. Cette construction se fait en trois étapes : sur l'arbre infini des mots binaires, on remplace chaque mot par un triplet de mots dont chacun est remplacé par un nombre de Markoff grâce à la fonction  $\pi$ .

L'*arbre du monoïde libre*  $\{a, b\}^*$  est l'arbre binaire infini complet où un sommet est identifié par le mot binaire représentant le chemin entre la racine et ce sommet. La figure 5.3 représente l'arbre du monoïde libre.

Soit  $X = \{a, b\}^* \cup \{a^{-1}, b^{-1}\}$  l'ensemble des mots sur l'alphabet  $\{a, b\}$  auquel on

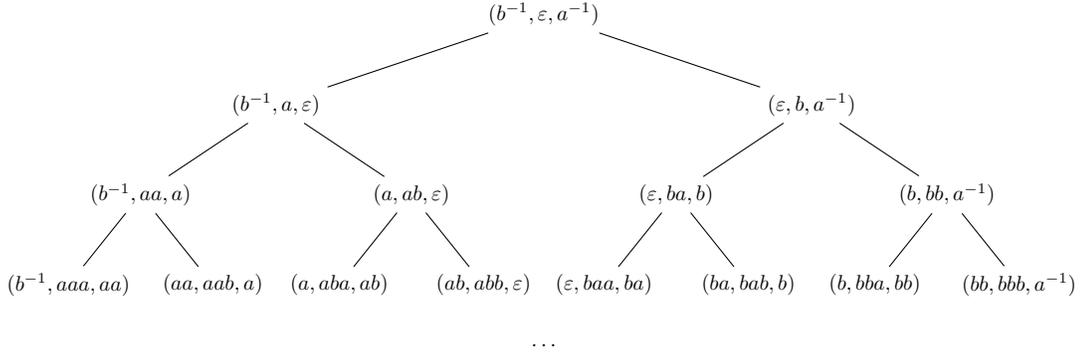


Figure 5.4: L'arbre des triplets

ajoute les symboles  $a^{-1}$  et  $b^{-1}$ . On définit la fonction  $\text{TRIPLET} : \{a, b\}^* \rightarrow X^3$  par

$$\text{TRIPLET}(w) = \begin{cases} (b^{-1}, a^k, a^{k-1}), & \text{si } w = a^k; \\ (b^{k-1}, b^k, a^{-1}), & \text{si } w = b^k; \\ (v, w, wa^{-1}), & \text{si } w = vba^{k+1}; \\ (wb^{-1}, w, v), & \text{si } w = vab^{k+1}; \end{cases}$$

avec  $k \geq 0$ . L'arbre obtenu en appliquant la fonction  $\text{TRIPLET}$  sur l'arbre du monoïde libre est appelé *l'arbre des triplets* (voir la figure 5.4). Il peut aussi être construit directement à partir de l'arbre du monoïde libre. Examinons le chemin  $\ell$  entre le sommet  $w$  et la racine. Soit  $u$  le sommet immédiatement après le premier pas nord-ouest sur le chemin  $\ell$  (s'il n'y a pas de pas nord-ouest, alors  $u = b^{-1}$ ) et  $v$  le sommet immédiatement après le premier pas nord-est sur le chemin  $\ell$  (s'il n'y a pas de pas nord-est, alors  $v = a^{-1}$ ). Alors, le sommet  $w$  est identifié par le triplet  $(u, w, v)$ .

**Exemple 5.5.** Prenons le mot  $abb$ , le sommet immédiatement après le premier pas nord-ouest est  $ab$  et celui après le premier pas nord-est est  $\varepsilon$ . Par conséquent,  $\text{TRIPLET}(abb) = (ab, abb, \varepsilon)$ . Prenons le mot  $bbb$ , le sommet immédiate-

ment après le premier pas nord-ouest est  $bb$  et il n'a pas de pas nord-est. Ainsi,  $\text{TRIPLET}(bbb) = (bb, bbb, a^{-1})$ .

La fonction  $\text{TRIPLET}$  est injective, car tout mot du monoïde libre apparaît au centre d'un unique triplet. De plus, la construction de l'arbre des triplets à l'aide de l'arbre du monoïde libre est analogue à la construction directe de l'arbre des mots de Christoffel.

Soit  $\xi$  la fonction du monoïde libre vers les mots de Christoffel propres définie par  $\xi(u) = a\text{Pal}(u)b$  avec  $u \in \{a, b\}^*$ . Elle peut être étendue à l'ensemble  $X$  en posant  $\xi(a^{-1}) = b$  et  $\xi(b^{-1}) = a$  et à  $X^3$  en posant  $\xi(u, w, v) = (\xi(u), \xi(w), \xi(v))$ .

**Lemme 5.6.** *La fonction  $\xi \circ \text{TRIPLET}$  est une bijection du monoïde libre  $\{a, b\}^*$  vers l'ensemble des triplets de Christoffel.*

*Démonstration.* Nous montrons que pour tout  $v \in \{a, b\}^*$ ,  $\xi \circ \text{TRIPLET}(v) = (w_1, w, w_2)$  où  $w = \xi(v)$  est un mot de Christoffel dont la factorisation standard est  $w_1w_2$ . Cette affirmation implique le lemme, car  $\xi|_{\{a, b\}^*}$  est une bijection de  $\{a, b\}^*$  sur l'ensemble des mots de Christoffel par le théorème 2.20.

La fonction  $\xi$  est une bijection de  $X$  dans l'ensemble des mots de Christoffel. Il y a trois cas à considérer : (i)  $v$  n'est pas une puissance de  $a$  ou  $b$ , (ii)  $v = a^k$  et (iii)  $v = b^k$ .

(i) Soit  $v$  un mot sur le monoïde libre. Supposons d'abord que  $v$  n'est pas une puissance de  $a$  ou de  $b$ . Alors,  $v_1$  identifie le sommet immédiatement après le premier pas nord-ouest et  $v_2$  identifie le sommet immédiatement après le premier pas nord-est. Par le théorème 2.13, on obtient que  $\xi(v) = \xi(v_1)\xi(v_2)$ .

(ii) Soit  $v = a^k$  pour un certain  $k \geq 0$ . Alors  $\text{TRIPLET}(v) = (b^{-1}, a^k, a^{k-1})$ . Donc,  $\xi \circ \text{TRIPLET}(v) = (a, aa^k b, aa^{k-1} b)$ , car  $\xi(a^k) = aa^k b$  est un mot de Christoffel

dont la factorisation standard est  $a \cdot a^k b$ .

(iii) Soit  $v = b^k$  pour un certain  $k \geq 0$ . Alors  $\text{TRIPLET}(v) = (b^{k-1}, b^k, a^{-1})$ . Donc,  $\xi \circ \text{TRIPLET}(v) = (ab^{k-1}b, ab^k b, b)$ , car  $\xi(b^k) = ab^k b$  est un mot de Christoffel dont la factorisation standard est  $ab^k \cdot b$ . ■

Soit  $\eta$  la fonction qui applique  $\mu(\cdot)_{12}$  sur des triplets de mots, c'est-à-dire

$$\eta(u, w, v) = (\mu(u)_{12}, \mu(w)_{12}, \mu(v)_{12}).$$

On obtient alors une autre bijection intéressante.

**Lemme 5.7.** *La fonction  $\eta \circ \xi \circ \text{TRIPLET}$  est une bijection du monoïde libre vers l'ensemble des triplets de Markoff.*

*Démonstration.* Par le théorème 5.2, la fonction  $\eta$  est une bijection de l'ensemble des mots de Christoffel sur l'ensemble des triplets de Markoff. De plus, par le lemme 5.6,  $\xi \circ \text{TRIPLET}$  est une bijection du monoïde libre dans l'ensemble précédent. ■

Rappelons que l'image d'un mot par la fonction  $\pi$  est un nombre de Markoff.

**Proposition 5.8** (Reutenauer et Vuillon (2017)). *Soit  $v$  un mot sur le monoïde libre  $\{a, b\}^*$ . Le nombre  $\pi(v)$  est le nombre de Markoff correspondant à  $\mu(w)_{12}$  où  $w$  est le mot de Christoffel  $a\text{Pal}(v)b$ .*

La démonstration présentée ici est basée sur celle de l'article Reutenauer (2018). Avant de démontrer la proposition, on définit le morphisme  $\beta$  du monoïde libre  $\{a, b\}^*$  vers  $\text{SL}_2(\mathbb{N})$  défini par

$$\beta(a) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \beta(b) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Ce morphisme est en relation avec le morphisme  $\mu$ .

**Lemme 5.9** (Reutenauer (2018)). *Pour tout mot  $w \in \{a, b\}^*$ ,*

$$\mu(w) = P^{-1}(\beta \circ \mathcal{G}(w))P,$$

avec  $P = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$  et  $\mathcal{G} = (a, ab)$ .

*Démonstration.* Puisque  $w \mapsto P^{-1}(\beta \circ \mathcal{G}(w))P$  et  $\mu$  sont des morphismes, il suffit de démontrer que les deux coïncident pour les lettres  $a$  et  $b$ . On a que

$$P^{-1}(\beta \circ \mathcal{G}(a))P = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \mu(a)$$

et

$$P^{-1}(\beta \circ \mathcal{G}(b))P = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} = \mu(b)$$

■

Soit  $\nu$  un autre morphisme du monoïde libre  $\{a, b\}^*$  vers  $\mathrm{SL}_2(\mathbb{N})$  défini par

$$\nu(a) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \nu(b) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Le morphisme  $\nu$  donne des informations sur les longueurs des mots de la factorisation standard d'un mot de Christoffel.

**Proposition 5.10** (Berthé *et al.* (2008)). *Soit  $w = a\mathrm{Pal}(v)b$  un mot de Christoffel et  $w = w_1w_2$  sa factorisation standard. Les longueurs des mots de Christoffel  $w$ ,  $w_1$  et  $w_2$  sont respectivement  $p + r$ ,  $q + s$  et  $p + q + r + s$  où*

$$\nu(v) = \begin{pmatrix} p & q \\ r & s \end{pmatrix}.$$

La démonstration de cette proposition est basée celle présentée dans (Reutenauer, 2019).

*Démonstration.* Soit  $w = a\text{Pal}(m)b$  un mot de Christoffel et  $w = uv$  sa factorisation standard. La paire  $(u, v)$ , vue comme un morphisme du monoïde libre  $\{a, b\}^*$ , est égale à  $\zeta(m)$  par la proposition 2.21. L'image commutative  $I$  est une fonction de  $\text{End}(\{a, b\}^*)$  vers  $\mathbb{Z}^{2 \times 2}$  qui associe un morphisme à une matrice d'incidence. Comme  $\mathcal{G} = (a, ab)$  et  $\mathcal{D} = (ab, b)$ , on a que  $I(\mathcal{G}) = \zeta(a)$  et  $I(\mathcal{D}) = \zeta(b)$ . Par conséquent,  $I \circ \zeta = \nu$ . Ainsi, l'image commutative de  $(u, v)$  est  $I \circ \zeta(m) = \nu(m)$ . On a donc que  $|u| = p + r$  et  $|v| = q + s$ , car l'image commutative de  $(u, v)$  est la matrice

$$\begin{pmatrix} |u|_a & |v|_a \\ |u|_b & |v|_b \end{pmatrix}.$$

■

*Démonstration de la proposition 5.8.* Soit  $v \in \{a, b\}^*$ . On veut démontrer que  $\pi(v) = \mu(a\text{Pal}(v)b)_{12}$ . Comme  $\pi(v) = |\text{Pal} \circ \theta \circ \text{Pal}(av)| + 2$ , il suffit de démontrer que  $\mu(a\text{Pal}(v)b)_{12}$  correspond à la longueur du mot de Christoffel  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$ . Par la proposition 5.10, on a que  $|a(\text{Pal} \circ \theta \circ \text{Pal}(av))b| = (S \circ \nu \circ \theta \circ \text{Pal})(av)$  où  $S$  est la fonction qui associe une matrice à la somme de ses entrées.

On note que  $\nu \circ \theta = \beta$ . En effet,

$$\nu \circ \theta(a) = \nu(ab) = \nu(a)\nu(b) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \beta(a)$$

et

$$\nu \circ \theta(b) = \nu(ba) = \nu(b)\nu(a) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \beta(b).$$

Par conséquent,  $(S \circ \nu \circ \theta \circ \text{Pal})(av) = (S \circ \beta \circ \text{Pal})(av)$ . Par la formule de Justin 1.12, on a que

$$S(\beta(\text{Pal}(av))) = S(\beta(\mathcal{G}(\text{Pal}(v))a)) = (1 \ 1)\beta \circ \mathcal{G}(\text{Pal}(v)) \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Posons  $P = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ . Pour tout  $v \in \{a, b\}^*$ , on a que

$$(1 \ 1)\beta \circ \mathcal{G}(\text{Pal}(v)) \begin{pmatrix} 3 \\ 2 \end{pmatrix} = (1 \ 1)PP^{-1}(\beta \circ \mathcal{G}(\text{Pal}(v)))PP^{-1} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

En appliquant le lemme 5.9, on obtient

$$(2 \ 1)P^{-1}(\beta \circ \mathcal{G}(\text{Pal}(v)))P \begin{pmatrix} 2 \\ 1 \end{pmatrix} = (2 \ 1)\mu(v) \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

De plus,

$$\mu(a\text{Pal}(v)b)_{12} = (1 \ 0) \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \mu(\text{Pal}(v)) \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (2 \ 1)\mu(\text{Pal}(v)) \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Donc,  $S(\nu(\theta(\text{Pal}(av)))) = \mu(a\text{Pal}(v)b)_{12}$  et  $\pi(v) = \mu(w)_{12}$  avec  $w = a\text{Pal}(v)b$ . ■

Une conséquence directe du lemme 5.7 est que  $\mu(\xi(v))_{12} = \pi(v)$  pour tout mot  $v \in \{a, b\}^*$ . La fonction  $\pi$  peut être étendue à l'ensemble  $X$  en posant  $\pi(b^{-1}) = 1$  et  $\pi(a^{-1}) = 2$  de même qu'à l'ensemble des triplets  $X^3$  en posant  $\pi(u, w, v) = (\pi(u), \pi(w), \pi(v))$ .

**Théorème 5.11** (Abram *et al.* (2020)). *La fonction  $\pi \circ \text{TRIPLET}$  est une bijection du monoïde libre  $\{a, b\}^*$  sur l'ensemble des triplets de Markoff.*

*Démonstration.* Par le lemme 5.6, la fonction  $\eta \circ \xi \circ \text{TRIPLET}$  est une bijection entre le monoïde libre et l'ensemble des triplets de Markoff. Grâce au lemme 5.8, on a alors  $\eta \circ \xi = \pi$ . Le diagramme commutatif 5.5 résume la démonstration de ce théorème. ■

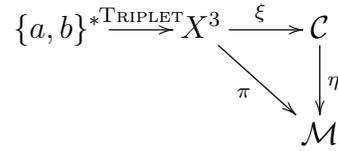


Figure 5.5: Le diagramme commutatif correspondant au théorème 5.11.  $\mathcal{C}$  désigne l'ensemble des triplets  $(u, w, v)$  où  $w$  est un mot de Christoffel dont la factorisation standard est  $uw$  et  $\mathcal{M}$  désigne l'ensemble des triplets de Markoff.

Cette construction met aussi en évidence une bijection entre l'arbre du monoïde libre, l'arbre des triplets, l'arbre des mots de Christoffel et l'arbre des triplets de Markoff.

### 5.3 Mots de Christoffel spéciaux

Pour tout mot  $v \in \{a, b\}^*$ , la fonction  $\pi(v)$  correspond à la longueur du mot de Christoffel  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$ , puisque que  $\pi(v) = |\text{Pal} \circ \theta \circ \text{Pal}(av)| + 2$ . Un mot de Christoffel est *spécial* s'il est de la forme  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b = \xi(\theta(\text{Pal}(av)))$  où  $v \in \mathcal{A}^*$ . Nous étudions d'abord les factorisations de ces mots de Christoffel spéciaux. Ensuite, nous définissons une autre notion de dualité pour les mots de Christoffel. Les mots de la forme  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  sont alors autoduals.

La factorisation standard des mots de Christoffel de la forme  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  est liée à la palindromisation itérée.

**Théorème 5.12** (Abram *et al.* (2020)). *Soit  $v \in \{a, b\}^*$ . Si  $uab = \theta(\text{Pal}(av))$ , alors*

$$\xi(uab) = \xi(ua) \cdot \xi(u)$$

*est la factorisation standard de  $\xi(uab)$ .*



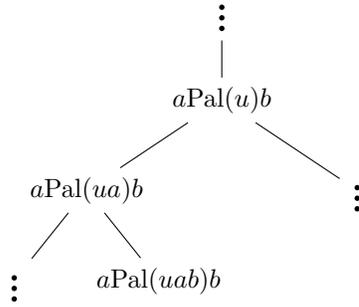


Figure 5.6: Position des sommets  $a\text{Pal}(u)b$ ,  $a\text{Pal}(ua)b$  et  $a\text{Pal}(uab)b$  dans l'arbre des mots de Christoffel.

Posons  $\theta(\text{Pal}(av))(ab) = uab$ . Donc, la factorisation standard de  $\xi(\theta(\text{Pal}(av))) = \xi(uab) = \xi(ua) \cdot (\xi(u))$  par la proposition 5.14. ■

La factorisation palindromique d'un mot de Christoffel peut être construite à partir de sa factorisation standard.

**Corollaire 5.15.** *La factorisation palindromique du mot de Christoffel  $w = a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  est*

$$a\text{Pal}(u)a \cdot b\text{Pal}(ua)b$$

où  $u = (\theta \circ \text{Pal}(av))(ab)^{-1}$ .

*Démonstration.* Par le théorème 5.12, on a que  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b = a\text{Pal}(ua)b \cdot a\text{Pal}(u)b$  avec  $u = (\theta \circ \text{Pal}(av))(ab)^{-1}$ . Comme  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  est un mot de Christoffel, le mot  $\text{Pal}(ua)ba\text{Pal}(u)$  est un palindrome. Par conséquent,

$$a\text{Pal}(ua)b \cdot a\text{Pal}(u)b = a\text{Pal}(u)a \cdot b\text{Pal}(ua)b.$$

Il s'agit de la factorisation palindromique de  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$ , car  $a\text{Pal}(u)a$  et  $b\text{Pal}(ua)b$  sont des palindromes et on sait que la factorisation palindromique d'un mot de Christoffel est unique par le théorème 2.9. ■

Nous introduisons maintenant une notion de dualité différente de celle présentée à la section 2. Cette variante est intéressante, car les mots de Christoffel de la forme  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  sont autoduals dans ce contexte. Soit  $w$  un mot de Christoffel propre de pente  $q/p$  et de longueur  $p + q$ . Le *dual* du mot de Christoffel  $w$ , noté  $w^\dagger$ , est le mot de Christoffel de pente  $q^\dagger/p^\dagger$ , où  $p^\dagger$  (respectivement  $q^\dagger$ ) est l'unique entier dans l'ensemble  $\{0, 1, \dots, n - 1\}$  satisfaisant l'équation  $pp^\dagger \equiv -1 \pmod{n}$  (resp.  $qq^\dagger \equiv -1 \pmod{n}$ ).

**Exemple 5.16.** *Le dual du mot de Christoffel de pente  $4/7$ ,  $w = aabaabaabab$  est le mot de Christoffel de pente  $8/3$ ,  $w^\dagger = abbabbbabbb$ , car  $4 \cdot 8 \equiv -1 \pmod{11}$  et  $7 \cdot 3 \equiv -1 \pmod{11}$ .*

Posons  $n = p + q$ . Le dual existe pour tous les mots de Christoffel de pente  $q/p$ . En effet, les entiers  $p$  et  $q$  sont premiers entres eux. Donc, les entiers  $n$  et  $p$  sont premiers entre eux, de même que les entiers  $n$  et  $q$ . Par conséquent, les entiers  $p^\dagger$  et  $q^\dagger$  existent et sont premiers entre eux et leur ratio est la pente d'un mot de Christoffel. De plus, un mot de Christoffel et son dual ont la même longueur. En effet, soit  $p^\dagger$  l'entier tel que  $pp^\dagger \equiv -1 \pmod{n}$  et  $p^\dagger \in \{0, \dots, n - 1\}$ . Alors,  $q(n - p^\dagger) \equiv -qp^\dagger \equiv -(n - p)p^\dagger \equiv -1 \pmod{n}$  avec  $n - p^\dagger \in \{0, \dots, n - 1\}$ . Par conséquent,  $q^\dagger = n - p^\dagger$ .

Il existe une autre involution naturelle sur les mots de Christoffel, celle qui inverse leur pente. Ces quatre involutions (l'identité, inverser la pente,  $w^{*1}$  et  $w^\dagger$ ) agissent sur l'ensemble des mots de Christoffel comme le groupe de Klein.

**Théorème 5.17** (Abram *et al.* (2020)). *Soit  $w = a\text{Pal}(u)b$  un mot de Christoffel propre. Le dual du mot de Christoffel  $w$  est  $w^\dagger = a\text{Pal}(\hat{u})b$  avec  $\hat{u} = \tilde{\tilde{u}}$ .*

---

1. voir définition 2.14

**Remarque 5.18.** *Un lien similaire entre  $w$  et son dual  $w^*$  a été mis en évidence dans [Berthé et al. \(2008\)](#) : Si  $w = a\text{Pal}(u)b$ , alors son dual est  $w^* = a\text{Pal}(\tilde{u})b$ .*

*Démonstration.* Nous utilisons une approche inspirée par [Berthé et al. \(2008\)](#).

Soit  $w = a\text{Pal}(u)b$  et  $w' = a\text{Pal}(\hat{u})b$  ainsi que

$$\nu(v) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ et } \nu(\hat{v}) = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}.$$

Par la proposition [5.10](#), nous avons que  $|w|_a = a + b$ ,  $|w|_b = c + d$ ,  $|w'|_a = a' + b'$  et  $|w'|_b = c' + d'$ . Pour démontrer que  $w' = w^\dagger$ , il suffit de démontrer que  $a + b + c + d = a' + b' + c' + d'$  et que  $(c + d)(c' + d') \equiv -1 \pmod{n}$ .

Nous montrons que  $\nu(\hat{v}) = \nu(v)^t$ , où  $\cdot^t$  est la transposition matricielle, par récurrence sur la longueur de  $v$ . Si  $|v| = 0$ , alors  $v = \varepsilon$ . On a alors que

$$\nu(\hat{\varepsilon}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \nu(\varepsilon)^t,$$

Si  $|v| = 1$ , alors  $v = a$  ou  $v = b$ . On a alors que

$$\nu(\hat{a}) = \nu(b) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^t = \nu(a)^t,$$

$$\nu(\hat{b}) = \nu(a) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^t = \nu(b)^t.$$

Supposons que pour  $v \in \{a, b\}^*$  de longueur au plus  $n$ , l'égalité  $\nu(\hat{v}) = \nu(v)^t$ . Soit  $v$  un mot de longueur  $n$ . Alors il existe une lettre  $x \in \{a, b\}$  et un mot  $v' \in \{a, b\}^*$  tels que  $v = v'x$ . On a que

$$\nu(\hat{v}) = \nu(\widehat{v'x}) = \nu(\hat{x})\nu(\hat{v}') = \nu(x)^t\nu(v')^t = (\nu(v')\nu(x))^t = \nu(v)^t$$

par l'hypothèse de récurrence.

On écrit  $\nu(\hat{v})$  en fonction de  $\nu(v)$  :

$$\nu(\hat{v}) = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^t = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Par conséquent, la pente de  $w'$  est  $\frac{c'+d'}{a'+b'} = \frac{b+d}{a+c}$ . De plus,  $a+b+c+d = a'+b'+c'+d' = n$ .

Nous montrons que  $(c+d)(c'+d') \equiv -1 \pmod{n}$ . Plus généralement, si

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{N}),$$

alors  $(b+d)(c+d) \equiv -1 \pmod{(a+b+c+d)}$ . En effet, nous avons que  $ad - bc = 1$ , alors  $(b+d)(c+d) = bc + bd + dc + d^2 = ad - 1 + bd + dc + d^2 = -1 + d(a+b+c+d) \equiv -1 \pmod{(a+b+c+d)}$ . Par conséquent, le dual de  $w = a\text{Pal}(v)b$  est  $w' = a\text{Pal}(\hat{v})b$ , tel que souhaité. ■

Ce résultat aussi peut être démontré à l'aide des fractions continues en suivant [Carpi et de Luca \(2004\)](#). À partir de la relation entre un mot et son dual, on peut caractériser les mots qui sont autoduals.

**Corollaire 5.19** ([Abram et al. \(2020\)](#)). *Soit  $w = a\text{Pal}(v)b$  un mot de Christoffel de pente  $q/p$ . Les énoncés suivants sont équivalents :*

1.  $w$  est autodual, c'est-à-dire, fixé par l'involution  $w \mapsto w^\dagger$  ;
2.  $v$  est un antipalindrome ;
3. la pente de  $w$  est égale à la pente de  $w^\dagger$  ;
4.  $p^2 \equiv -1 \pmod{(p+q)}$  ;
5.  $q^2 \equiv -1 \pmod{(p+q)}$ .

*Démonstration.* L'équivalence entre les énoncés découle directement du théorème [5.17](#) et de l'injectivité de la fonction Pal. ■

**Corollaire 5.20.** *Les mots de Christoffel de la forme  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  sont autoduals.*

**Remarque 5.21.** *Les mots de Christoffel spéciaux ne sont pas les seuls mots auto-duals, car il existe des antipalindromes comme  $ba$ , qui ne sont pas de la forme  $\theta(\text{Pal}(av))$ .*

Pour en faire la démonstration, nous utilisons un résultat de l'article [de Luca et De Luca \(2006\)](#).

**Lemme 5.22** ([de Luca et De Luca \(2006\)](#)). *Si  $w$  est un palindrome, alors  $\theta(w)$  est un antipalindrome.*

*Démonstration.* Soit  $w = w_1 \dots, w_n$  un palindrome. On a que  $w_i = w_{n-i+1}$  puisque  $w$  est un palindrome. Le morphisme  $\theta$  envoie  $x$  sur  $x\bar{x}$  pour tout  $x \in \{a, b\}$ , alors le mot  $\theta(w) = a_1 \dots a_{2n} = w_1 \bar{w}_1 \dots w_n \bar{w}_n$ . Ainsi,  $w_i = w_{n-i+1}$  implique que  $a_i = \overline{a_{n-i+1}}$ . Donc  $\theta(w) = \widehat{\theta(w)}$ , c'est-à-dire que  $\theta(w)$  est un antipalindrome. ■

*Démonstration du corollaire 5.20.* Comme l'image par  $\theta$  d'un palindrome est un antipalindrome par le lemme 5.22, l'image de la fonction  $\theta \circ \text{Pal}$  est un sous-ensemble de l'ensemble des antipalindromes. Par le corollaire 5.19,  $a(\text{Pal} \circ \theta \circ \text{Pal}(av))b$  est autodual. ■

Le corollaire 5.19 signifie que les mots auto-duals, en particulier les mots  $\text{Pal} \circ \theta \circ \text{Pal}(av)$ , sont *harmoniques* dans le sens de [Carpi et de Luca \(2004\)](#). Un mot est dit harmonique si la pente  $q/p$  du mot de Christoffel  $w$  est égale à  $|u|/|v|$  où  $w = uv$  est la factorisation standard de  $w$ .

#### 5.4 Généralisation de la conjecture de Frobenius

La conjecture de Frobenius a été beaucoup étudiée au cours des 100 dernières années comme l'illustre le livre d'Aigner (2013). Le but ici est de généraliser la conjecture de Frobenius sur les mots de Christoffel au plus grand sous-ensemble du monoïde libre  $\{a, b\}^*$  possible. Il est facile de voir que la fonction  $w \mapsto \mu(w)_{12}$  n'est pas injective sur le monoïde libre  $\{a, b\}^*$ . En effet, on a que  $\mu(aabbab)_{12} = \mu(abaabb)_{12} = 1130$ . En revanche, elle semble l'être si on se restreint aux conjugués des mots de Christoffel.

**Conjecture 5.23.** *La fonction  $w \mapsto \mu(w)_{12}$  est injective sur l'ensemble des conjugués des mots de Christoffel.*

**Remarque 5.24.** *L'ensemble  $Y = \{\mu(w)_{12} \mid w \text{ est conjugué à mot de Christoffel}\}$  contient tous les nombres de Markoff par le théorème 5.2, mais aussi d'autres entiers qui ne sont pas des nombres de Markoff.*

Cette conjecture a été vérifiée informatiquement sur les mots de Christoffel et leurs conjugués de longueur inférieure à 1000. Tous les nombres obtenus sont bel et bien différents. La suite des nombres  $\mu(w)_{12}$  inférieurs à 100 où  $w$  un conjugué d'un mot de Christoffel est :

$$1, 2, 5, 7, 13, 17, 19, 29, 31, 34, 41, 44, 46, 50, 89, \dots$$

Nous savons que nous ne manquons pas de termes car  $\mu(w)_{12} > 100$  pour tous les mots  $w$  de longueur 6. Par conséquent,  $\mu(w)_{12} > 100$  pour tous les mots  $w$  tel que  $|w| \geq 6$ , car les entrées des matrices  $\mu(w)$ ,  $\mu(a)$  et  $\mu(b)$  sont toutes strictement positives. Nous avons identifié un autre ensemble contenant les conjugués des mots de Christoffel sur lequel la fonction  $w \mapsto \mu(w)_{12}$  est injective.

**Conjecture 5.25.** *La fonction  $w \mapsto \mu(w)_{12}$  est injective sur l'ensemble des mots sturmiens.*

Ainsi, la conjecture 5.25 généralise la conjecture de Frobenius ainsi que la conjecture 5.23. Pour valider cette conjecture par ordinateur, nous avons besoin d'une façon de construire les mots sturmiens. Nous avons utilisé la *forme normale* des mots sturmiens introduite dans Reutenauer (2015). Soit  $m$  un mot sturmien qui n'est pas une puissance de  $a$  ou de  $b$ , alors il existe une unique factorisation de  $m$  de la forme

$$m = sw^n p$$

où  $n \geq 1$ ,  $w$  est un mot de Christoffel inférieur ou supérieur dont la factorisation palindromique est  $w = v'u'$ ,  $s$  est un suffixe propre de  $u'$  et  $p$  est un suffixe propre de  $w$ . Ainsi, nous avons pu calculer la fonction  $m \mapsto \mu(m)_{12}$  sur tous les mots sturmiens de longueur inférieure à 500 sans découvrir de paires de mots dont l'image coïncide. La suite des nombres  $\mu(m)_{12}$  inférieurs à 100 avec  $m$  un mot sturmien est :

1, 2, 3, 5, 7, 8, 12, 13, 17, 19, 21, 29, 31, 34, 41, 44, 46, 50, 55, 70, 75, 81, 89, 99, . . .

Nous ne savons pas démontrer les conjectures précédentes, mais nous pouvons démontrer que la fonction  $w \mapsto \mu(w)_{12}$  est injective sur certains sous-ensembles de mots sturmiens et de conjugués des mots de Christoffel.

On définit l'*ordre militaire*  $\leq_{\text{mil}}$  sur les mots par  $u \leq_{\text{mil}} v$  si  $|u| < |v|$  ou  $|u| = |v|$  et  $u \preceq v$ . Ainsi, l'ordre militaire ordonne d'abord les mots par longueur et ensuite lexicographiquement.

**Proposition 5.26.** *Soit  $w$  un mot de Christoffel et  $H$  l'ensemble de ses facteurs circulaires. La fonction  $w \mapsto \mu(w)_{12}$  est strictement croissante pour l'ordre militaire sur  $H$  ; en particulier, elle est injective sur  $H$ .*

$a$	$a$	$a$	$b$	$a$	$a$	$b$	1	3	8	34	115	311	1325
$a$	$a$	$b$	$a$	$a$	$a$	$b$	1	3	13	44	119	313	1327
$a$	$a$	$b$	$a$	$a$	$b$	$a$	1	3	13	44	119	507	1715
$a$	$b$	$a$	$a$	$a$	$b$	$a$	1	5	17	46	121	513	1735
$a$	$b$	$a$	$a$	$b$	$a$	$a$	1	5	17	46	196	663	1793
$b$	$a$	$a$	$a$	$b$	$a$	$a$	2	7	19	50	212	717	1939
$b$	$a$	$a$	$b$	$a$	$a$	$a$	2	7	19	81	274	741	1949

Figure 5.7: La matrice des conjugués de  $aaabaab$  et la matrice des valeurs  $\mu(u)_{12}$

La figure 5.7 illustre cette proposition. À droite de la figure 5.7, on voit la matrice des conjugués du mot de Christoffel  $aaabaab$  et à gauche le nombre à la position  $(i, j)$  correspond à  $\mu(u)_{12}$  où  $u$  est le préfixe terminant à la position  $j$  de la  $i^{\text{ième}}$  ligne de la matrice des conjugués de  $w$ . La matrice d'entiers ainsi obtenue est croissante du haut vers le bas et de gauche à droite.

**Lemme 5.27.** *Soit  $w$  un mot sur  $\{a, b\}^*$ . Si*

$$\mu(w) = \begin{pmatrix} p & q \\ r & s \end{pmatrix},$$

*alors, les entiers  $p - q + r$  et  $2p - 3q + r - s$  sont strictement positifs.*

*Démonstration.* Puisque les coefficients des matrices  $\mu(a)$  et  $\mu(b)$  sont strictement positifs, les coefficients des matrices de  $\mu(w)$  avec  $w \in \{a, b\}^*$  sont strictement positifs, sauf ceux de la matrice identité  $\mu(\varepsilon)$  pour laquelle  $p = s = 1 > 0$  et  $q = r = 0$ . Démontrons les deux égalités par récurrence sur la longueur de  $w$ . Si  $w$  est le mot vide, les inégalités sont évidentes.

Soit  $w'$  un mot de longueur  $n > 0$ . Alors  $w' = wa$  ou  $w' = wb$ . Posons

$$\mu(w') = \begin{pmatrix} p' & q' \\ r' & s' \end{pmatrix}.$$

Supposons d'abord que  $w' = wa$ , alors

$$\mu(w') = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2p + q & p + q \\ 2r + s & r + s \end{pmatrix}$$

Par conséquent,  $p' - q' + r' = 2p + q - p - q + 2r + s = p + 2r + s > 0$ , car  $r$  et  $s$  sont positifs ou nuls et  $p > 1$ . De plus,  $2p' - 3q' + r' - s' = 4p + 2q - 3p - 3q + 2r + s - r - s = p - q + r > 0$  par l'hypothèse de récurrence.

Supposons ensuite que  $w' = wb$ . Alors

$$\mu(w') = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5p + 2q & 2p + q \\ 5r + 2s & 2r + s \end{pmatrix}$$

Par conséquent,  $p' - q' + r' = 5p + 2q - 2p - q + 5r + 2s = 3p + q + 5r + 2s > 0$ , car  $p > 1$  et  $q, r$  et  $s$  sont positifs ou nuls. De plus,  $2p' - 3q' + r' - s' = 10p + 4q - 6p - 3q + 5r + 2s - 2r - s = 4p + q + 3r + s > 0$ , car  $p > 1$  et  $q, r$  et  $s$  sont positifs ou nuls. ■

**Lemme 5.28.** *Si  $u$  et  $v$  sont des préfixes de même longueur de deux conjugués  $x$  et  $y$  de  $w$  consécutifs selon l'ordre lexicographique et si  $u \neq v$ , alors  $\mu(u)_{12} < \mu(v)_{12}$ .*

*Démonstration.* Soit  $x$  et  $y$  des conjugués consécutifs de  $w$  dans la matrice des conjugués. Alors pour certains mots  $p, s$ , on a que  $x = pabs$  et  $y = pbas$  et  $sp$  est le palindrome qui est le facteur central maximal de  $w$  par la proposition 2.19. Supposons que  $u \neq v$ . Il y a deux cas :  $u = pabs_1$  et  $v = pbas_1$  avec  $s = s_1s_2$  et  $u = pa, v = pb$ .

Supposons que  $u = pabs_1$  et  $v = pbas_1$  avec  $s = s_1s_2$ . On a alors que  $|p| \leq |s_1|$  ou  $|p| > |s_1|$ . Supposons que  $|p| \leq |s_1|$ . Puisque  $s_1s_2p$  est un palindrome, on a que

$s_1 = \tilde{p}t$ . Par conséquent,  $u = pab\tilde{p}t$  et  $v = pba\tilde{p}t$ . Posons

$$\mu(p) = \begin{pmatrix} i & j \\ k & l \end{pmatrix} \text{ et } \mu(t) = \begin{pmatrix} x & y \\ z & m \end{pmatrix}.$$

Notons que  $\mu(\tilde{p})$  est la transposée de  $\mu(p)$ . En effet, posons  $p = a_1 \dots a_n$ . On a que

$$\mu(p)^t = \mu(a_1 \dots a_n)^t = \mu(a_n)^t \dots \mu(a_1)^t = \mu(a_n) \dots \mu(a_1) = \mu(\tilde{p}),$$

car les matrices  $\mu(a)$  et  $\mu(b)$  sont des matrices symétriques. Le déterminant de  $\mu(p) = il - jk = 1$ , car le déterminant des matrices  $\mu(a)$  et  $\mu(b)$  est 1. De plus,

$$\begin{aligned} \mu(ba) - \mu(ba) &= \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 12 & 7 \\ 5 & 3 \end{pmatrix} - \begin{pmatrix} 12 & 5 \\ 7 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}. \end{aligned}$$

Ainsi, on a que

$$\begin{aligned}
\mu(v) - \mu(u) &= \mu(pba\tilde{p}t) - \mu(pab\tilde{p}t) \\
&= \mu(p)(\mu(ba) - \mu(ab))\mu(\tilde{p})\mu(t) \\
&= \begin{pmatrix} i & j \\ k & l \end{pmatrix} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} i & k \\ j & l \end{pmatrix} \begin{pmatrix} x & y \\ z & m \end{pmatrix} \\
&= \begin{pmatrix} -2j & 2i \\ -2l & 2k \end{pmatrix} \begin{pmatrix} i & k \\ j & l \end{pmatrix} \begin{pmatrix} x & y \\ z & m \end{pmatrix} \\
&= \begin{pmatrix} 0 & -2jk + 2il \\ -2li + 2kj & 0 \end{pmatrix} \begin{pmatrix} m & n \\ o & p \end{pmatrix} \\
&= \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} x & y \\ z & m \end{pmatrix} \\
&= \begin{pmatrix} 2z & 2m \\ -2x & -2y \end{pmatrix},
\end{aligned}$$

car  $il - kj = \det \begin{pmatrix} i & k \\ j & l \end{pmatrix} = 1$ . Comme  $m > 0$ , on a que  $\mu(u)_{12} < \mu(v)_{12}$ .

Supposons que  $|p| > |s_1|$ . Comme  $s_1s_2p$  est un palindrome,  $p = t\tilde{s}_1$ . Par conséquent,  $u = t\tilde{s}_1abs_1$  et  $v = t\tilde{s}_1bas_1$ . Posons

$$\mu(s_1) = \begin{pmatrix} i & j \\ k & l \end{pmatrix} \text{ et } \mu(t) = \begin{pmatrix} x & y \\ z & m \end{pmatrix}$$

Ainsi,

$$\begin{aligned}
\mu(v) - \mu(u) &= \mu(t\tilde{s}_1bas_1) - \mu(t\tilde{s}_1abs_1) \\
&= \mu(t)\mu(\tilde{s}_1)(\mu(ba) - \mu(ab))\mu(s_1) \\
&= \begin{pmatrix} x & y \\ z & m \end{pmatrix} \begin{pmatrix} i & k \\ j & l \end{pmatrix} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} i & j \\ k & l \end{pmatrix} \\
&= \begin{pmatrix} x & y \\ z & m \end{pmatrix} \begin{pmatrix} i & k \\ j & l \end{pmatrix} \begin{pmatrix} 2k & 2l \\ -2i & -2j \end{pmatrix} \\
&= \begin{pmatrix} x & y \\ z & m \end{pmatrix} \begin{pmatrix} 0 & 2il - 2kj \\ 2kj - 2il & 0 \end{pmatrix} \\
&= \begin{pmatrix} x & y \\ z & m \end{pmatrix} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \\
&= \begin{pmatrix} -2y & 2x \\ -2m & 2z \end{pmatrix},
\end{aligned}$$

car  $il - kj = \det \begin{pmatrix} i & j \\ k & l \end{pmatrix} = 1$ . Puisque  $x > 0$ , nous avons que  $\mu(u)_{12} < \mu(v)_{12}$ .

Supposons que  $u = pa$  et  $v = pb$ . Alors,  $\mu(u)_{12} = \mu(p)_{11} + \mu(p)_{12} < 2\mu(p)_{11} + \mu(p)_{12} = \mu(v)_{12}$ , puisque  $\mu(p)_{11} > 0$ . Par conséquent,  $\mu(u)_{12} > \mu(v)_{12}$ .  $\blacksquare$

*Démonstration de la proposition 5.26.* Étudions la suite de mots suivante : soit  $w_1 \prec w_2 \prec \dots \prec w_n$  les conjugués de  $w$  ordonnés en ordre lexicographique ; ensuite prenons les préfixes de longueur 1 de  $w_1, \dots, w_n$  suivi des préfixes de longueur 2 et ainsi de suite jusqu'aux préfixes de longueur  $n$  qui sont les  $w_i$  eux-mêmes. Alors la suite comprenant  $n^2$  mots est l'ensemble des facteurs circulaires de  $w$  ordonnés en ordre militaire avec répétitions (voir figure 5.7). On a démontré que si  $u$  et  $v$  sont des préfixes de même longueur de deux conjugués  $x$  et  $y$  consécutifs selon l'ordre lexicographique et que  $u \neq v$ , alors  $\mu(u)_{12} < \mu(v)_{12}$  par le lemme 5.28. Il reste à

démontrer que si  $u$  est le préfixe de longueur  $k$  de  $w_n$  et  $v$  le préfixe de longueur  $k + 1$  de  $w_1$ , alors  $\mu(u)_{12} < \mu(v)_{12}$ .

Soit  $u$  un préfixe de longueur  $k$  de  $w_n$  et  $v$  le préfixe de longueur  $k + 1$  de  $w_1$ . Puisque  $w_1 = ahb$  et  $w_n = bha$ , nous avons que  $u = bz$  et  $v = azx$  pour un certain préfixe  $z$  de  $h$  et une lettre  $x$ . Démontrons que  $\mu(bz)_{12} < \mu(azx)_{12}$ . Il suffit de démontrer que  $\mu(bz)_{12} < \mu(aza)_{12}$ , car la matrice  $\mu(a)$  est plus petite que  $\mu(b)$  en comparant les coefficients.

On a que

$$\mu(bz) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} i & j \\ k & l \end{pmatrix} = \begin{pmatrix} * & 5j + 2l \\ * & * \end{pmatrix}$$

et

$$\begin{aligned} \mu(aza) &= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} i & j \\ k & l \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2i + k & 2j + l \\ * & * \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} * & 2i + k + 2j + l \\ * & * \end{pmatrix} \end{aligned}$$

Par conséquent,  $\mu(aza)_{12} - \mu(bz)_{12} = 2i + k + 2j + l - 5j - 2l = 2i - 3j + k - l > 0$  par le lemme 5.27. Donc, la fonction  $w \mapsto \mu(w)_{12}$  est strictement croissante pour l'ordre militaire sur  $H$ . ■

Nous terminons en présentant un cas particulier de la conjecture 5.23.

**Corollaire 5.29.** *Soit  $c$  un chemin infini dans l'arbre des mots de Christoffel débutant à la racine et  $L$  l'ensemble des conjugués des mots de Christoffel étiquetant un sommet sur le chemin  $c$ . Alors  $w \mapsto \mu(w)_{12}$  est injective sur  $L$ .*

**Exemple 5.30.** *Le mot de  $(ab)^\infty$  représente un chemin infini dans l'arbre de Christoffel. Les premiers mots de  $L$  pour ce chemin sont*

$$\begin{array}{cccccccccccc} w & ab & ba & aab & aba & baa & aabab & ababa & babaa & abaab & baaba & \dots \\ \mu(w)_{12} & 5 & 7 & 13 & 17 & 19 & 194 & 254 & 284 & 196 & 274 & \dots \end{array}$$

*Démonstration.* Soit  $w_1, w_2$  deux mots consécutifs sur le chemin  $c$  dans l'arbre des mots de Christoffel. Comme  $w_1$  est un mot de Christoffel, on a que  $\widetilde{w}_1$  est conjugué à  $w_1$  et que  $x \preceq \widetilde{w}_1$  pour tout conjugué  $x$  de  $w_1$ . Par la proposition 5.26, on a que  $\mu(x)_{12} \leq \mu(\widetilde{w}_1)_{12}$ . Il suffit donc de démontrer que  $\widetilde{w}_1$  est un facteur circulaire de  $w_2$  par la proposition 5.26.

Par la construction de l'arbre, on a que  $w_1 = uv$  et  $w_2 = uvv$  ou  $w_2 = uvv$ . Ainsi,  $\widetilde{w}_1 = \widetilde{v}u$  et  $\widetilde{w}_2 = \widetilde{v}u\widetilde{u}$  ou bien  $\widetilde{w}_2 = \widetilde{v}\widetilde{v}u$ . Dans les deux cas,  $\widetilde{w}_1$  est un facteur de  $\widetilde{w}_2$ . Comme  $w_2$  est un mot de Christoffel,  $\widetilde{w}_1$  est un facteur circulaire de  $w_2$ , car  $w_2$  et  $\widetilde{w}_2$  sont conjugués. Soit  $y$  un conjugué de  $w_2$ . On a que  $\mu(x)_{12} \leq \mu(\widetilde{w}_1)_{12} < \mu(w_2)_{12} \leq \mu(y)_{12}$ . Donc la fonction  $w \mapsto \mu(w)_{12}$  est injective sur  $L$ . ■

## CHAPITRE VI

### MOTS FEUILLUS DES GRAPHS CHENILLES

Dans ce chapitre, nous étudions les mots feuillus, des mots définis à partir de graphes. Ces mots sont en lien avec le nombre maximal de feuilles d'un sous-graphe induit d'ordre  $i$  d'un graphe. Plus précisément, la fonction feuille d'un graphe  $G$  associe à un entier  $i$  le nombre maximal de feuille d'un sous-graphe induit d'ordre  $i$  de  $G$ . Le mot feuillu de  $G$  est la dérivée discrète de la suite obtenue à partir de la fonction feuille de  $G$ . Nous cherchons un critère simple pour déterminer si un mot est feuillu. Nous démontrons que les mots feuillus d'un sous-ensemble de graphes, appelé graphes chenilles ([Harary et Schwenk, 1973](#)), sont en bijection avec les mots préfixe normaux ([Fici et Lipták, 2011](#)).

D'abord, nous introduisons les mots préfixes normaux. Ensuite, nous rappelons les notions de bases de la théorie des graphes et introduisons les mots feuillus. Nous étudions aussi les graphes chenilles et différentes façons de les représenter. Finalement, nous démontrons la bijection entre les mots feuillus des graphes chenilles et les mots préfixes normaux. Les résultats de ce chapitre ont déjà été publiés dans [Blondin Massé \*et al.\* \(2018a\)](#).

## 6.1 Mots préfixes normaux

Les mots préfixes normaux ont été introduits dans [Fici et Lipták \(2011\)](#). Soit  $w$  un mot binaire appartenant à  $\{0, 1\}^*$ . La fonction  $F_x$  définie par

$$F_x(w, k) = \max\{|v|_x \mid v \in \text{Fact}_k(w)\}$$

pour tout  $0 \leq k \leq |w|$  correspond au maximum d'occurrences de la lettre  $x$  dans un facteur de longueur  $k$ . De plus, on définit  $F_x(w)$  par  $k \mapsto F_x(w, k)$ . La *forme préfixe normale*  $\text{FPN}_x(w)$  de  $w$  selon  $x$  est l'unique mot  $w'$  tel que  $F_x(w, k) = F_x(w', k) = |\text{Pref}_k(w')|_x$  pour tout  $k$ . La forme préfixe normale existe pour tous les mots ([Fici et Lipták, 2011](#)). Un mot  $w$  est dit *préfixe normal* selon  $x$  si  $w = \text{FPN}_x(w)$ . L'ensemble des mots préfixes normaux selon  $x$  est noté  $\mathcal{N}_x$ . La forme préfixe normale définit une relation d'équivalence sur le monoïde  $\{0, 1\}^*$ . Soit  $u, v$  deux mots appartenant à  $\{0, 1\}^*$ , on a que

$$u \equiv_{\text{FPN}} v \Leftrightarrow \text{FPN}_x(u) = \text{FPN}_x(v).$$

[Fici et Lipták \(2011\)](#) ont démontré qu'il y a un unique mot préfixe normal dans chacune des classes d'équivalence.

**Exemple 6.1.** *La forme préfixe normale du mot 10100111 selon 1 est*

$$\text{FPN}(10100111) = 11100101.$$

*Ainsi, le mot 10100111 n'est pas préfixe normal selon 1, mais le mot 11100101 l'est.*

[Fici et Lipták \(2011\)](#) se sont aussi intéressés au lien entre les mots préfixes normaux selon  $a$  et les mots de Lyndon sur l'alphabet  $\{a < b\}$ . Il existe des mots préfixes normaux selon  $a$  qui sont des mots de Lyndon tels que le mot  $aabab$ , mais aussi des mots préfixes normaux selon  $a$  qui ne sont pas des mots de Lyndon, comme

$abab$ . De plus, il y a des mots de Lyndon qui ne sont pas préfixes normaux selon  $a$ , comme  $aabbabaabbb$ . Ils ont démontré que les mots préfixes normaux selon  $a$  sont des préfixes des puissances des mots de Lyndon. Les mots de Christoffel sont des mots de Lyndon qui sont tous des mots préfixes normaux selon  $a$ .

**Proposition 6.2.** *Tous les mots de Christoffel sont préfixes normaux selon  $a$ .*

*Démonstration.* Les mots de Christoffel  $a$ ,  $b$  et  $ab$  sont préfixes normaux selon  $a$ . Soit  $w$  un mot de Christoffel de longueur au moins 3. On a que  $w = amb$  où  $m$  est un palindrome par la proposition 2.6. Démontrons par l'absurde que  $w$  est préfixe normal selon  $a$ . Supposons que  $w$  ne soit pas préfixe normal selon  $a$ , alors il existe un facteur  $v$  de  $w$  tel que  $|\text{Pref}_{|v|}(w)|_a < |v|_a$ . Posons  $k = |v|_a$ . Comme un mot de Christoffel est équilibré (Dulucq et Gouyou-Beauchamps, 1990), on a que

$$\begin{aligned} ||\text{Pref}_{|v|}(w)|_a - |v|_a| \leq 1 &\Rightarrow ||\text{Pref}_{|v|}(w)|_a - k| < 1 \\ &\Rightarrow |\text{Pref}_{|v|}(w)|_a = k - 1. \end{aligned}$$

De plus,  $|\text{Pref}_{|v|}(w)|_a = au$  où  $u$  est un préfixe de  $m$ . Par conséquent, on a que  $\tilde{u}$  est un suffixe de  $m$  et  $\tilde{u}b$  est un facteur de  $w$ . De plus, on a que

$$|\tilde{u}b|_a = |\tilde{u}|_a = |au|_a - 1 = k - 2.$$

Par conséquent,

$$||\tilde{u}b|_a - |v|_a| = 2 > 1$$

ce qui est une contradiction. Donc, les mots de Christoffel sont préfixes normaux selon  $a$ . ■

Dans cette thèse, nous considérons uniquement les mots préfixes normaux selon 1 sur l'alphabet  $\{0, 1\}$ , qu'on appelle simplement mots préfixes normaux pour alléger le texte. Le lemme suivant qui a été démontré dans (Blondin Massé *et al.*,

(2018a) décrit une propriété des mots qui ne sont pas préfixes normaux qui sera utilisée pour démontrer la bijection entre les mots feuilles et les mots préfixes normaux.

**Lemme 6.3.** *Si  $w$  est un mot binaire sur  $\{0, 1\}^*$  qui n'est pas préfixe normal, alors il existe deux mots abéliens équivalents  $u$  et  $u'$  tels que  $u0$  est un préfixe de  $w$  et  $1u'$  est un facteur de  $w$ .*

*Démonstration.* Soit  $w$  un mot qui n'est pas préfixe normal. Il existe au moins un préfixe  $p$  et un facteur  $f$  de  $w$ , tous deux de même longueur telle que  $|p|_1 < |f|_1$ . Choisissons  $p$  et  $f$  le plus court possible. Soit  $p = ua$  et  $f = bu'$  pour certaines lettres  $a, b \in \{0, 1\}$ . Puisque  $p$  et  $f$  sont minimaux, nous avons que  $|u|_1 \geq |u'|_1$ . Par conséquent,

$$|u|_1 + a = |ua|_1 = |p|_1 < |f|_1 = |bu'|_1 = |u'|_1 + b \leq |u|_1 + b.$$

Cette inégalité est satisfaite seulement si  $a = 0$  et  $b = 1$ , auquel cas  $|u|_1 = |u'|_1$ . ■

Le résultat principal de ce chapitre est une bijection entre les mots préfixes normaux et les mots feuillus des graphes chenille.

**Théorème 6.4** (Blondin Massé *et al.* (2018a)). *Soit  $L : \llbracket 1, n \rrbracket \rightarrow \mathbb{N}$  une fonction telle que  $(L(0), L(1), L(2), L(3)) = (0, 0, 2, 2)$ . Alors, il existe une chenille  $C$  telle que  $L = L_C$  si et seulement si  $\Delta L$  est préfixe normal.*

Avant de démontrer ce théorème, nous allons rappeler plusieurs notions sur les graphes et les graphes chenilles.

## 6.2 Mots feuillus

Les mots feuillus sont construits à partir de graphes. Nous rappelons dans un premier temps la terminologie de base de la théorie des graphes (Labelle, 1981).

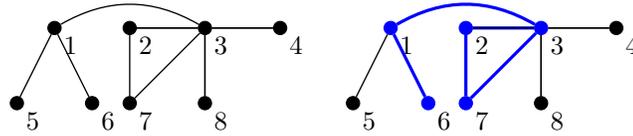


Figure 6.1: Un graphe simple ayant 8 sommets et 8 arêtes et le sous-graphe induit par  $Y = \{1, 2, 3, 7, 6\}$  coloré en bleu

Ensuite, on introduit la notion de mots feuillus.

Un *graphe simple*  $G = (V, E)$  est un couple formé d'un ensemble fini non vide  $V$  appelé l'*ensemble des sommets* de  $G$  et d'un ensemble  $E \subseteq \mathcal{P}_2(V)$  appelé l'*ensemble des arêtes* de  $G$ , où  $\mathcal{P}_2(V)$  est l'ensemble de tous les sous-ensembles de cardinalité 2 de  $V$ . Le nombre de sommets du graphe simple  $G = (V, E)$ , c'est-à-dire  $|V|$ , est appelé l'*ordre de  $G$*  et il est noté  $|G|$ . Soit  $e = \{x, y\} \in E$ . On dit que  $e$  est l'*arête* de  $G$  d'extrémités  $x$  et  $y$ . Deux sommets  $x$  et  $y$  sont dits *adjacents* dans  $G$  si  $\{x, y\} \in E$ . On dit aussi que les sommets  $x$  et  $y$  sont *incidents* à cette arête.

On représente habituellement les graphes simples par une figure dans le plan : les sommets sont représentés par des points et les arêtes par des traits entre les sommets. La figure 6.1 représente le graphe  $G = (V, E)$  avec  $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$  et  $E = \{\{1, 3\}, \{1, 5\}, \{1, 6\}, \{2, 3\}, \{2, 7\}, \{3, 4\}, \{3, 7\}, \{3, 8\}\}$ .

Le *degré* d'un sommet  $x$ , noté  $\deg(x)$  est le nombre d'arêtes de  $G$  qui sont incidentes à  $x$ . Un sommet de degré 1 est appelé un *sommet pendent* de  $G$ .

Soit  $G = (V, E)$  un graphe simple et  $Y \subseteq V$ . Le *sous-graphe induit par  $Y$*  de  $G$ , noté  $G_Y$ , est défini par  $G_Y = (Y, E \cap \mathcal{P}_2(Y))$ , où  $\mathcal{P}_2(Y)$ . Le sous-graphe induit par  $Y = \{1, 2, 3, 7, 6\}$  du graphe de la figure 6.1 est le graphe  $G_Y = (Y, \{\{1, 3\}, \{1, 6\}, \{2, 3\}, \{2, 7\}, \{3, 7\}\})$ .

Dans un graphe simple  $G = (V, E)$ , une *chaîne* est une suite finie de sommets  $x_1, x_2, \dots, x_m$  telle que pour tout  $1 \leq i \leq m$ ,  $\{x_i, x_{i+1}\}$  est une arête de  $G$ . La *longueur* d'une chaîne est l'entier  $m - 1$ . Dans le cas où  $x_1 = x_m$ , on dit que la chaîne est *chaîne fermée* et on appelle cette chaîne un *cycle* de  $G$ . Une chaîne de longueur supérieure à 1 dont toutes les arêtes sont distinctes est dite *simple*.

**Exemple 6.5.** La suite  $1, 3, 7, 2, 3, 4$  est une chaîne de longueur 5 du graphe de la figure 6.1. Elle est simple, car toutes les arêtes sont distinctes. La suite  $2, 3, 7, 2, 3, 4$  est une autre chaîne de ce graphe, mais elle n'est pas simple puisqu'elle contient deux fois l'arête  $\{2, 3\}$ . La suite  $2, 3, 7, 2$  est un cycle de ce graphe, car c'est une chaîne commençant et terminant au sommet 2.

Un graphe  $G = (V, E)$  est dit *connexe* si, pour tous  $x, y \in V$ , il existe une chaîne de  $x$  à  $y$ . Un graphe simple  $G = (V, E)$  qui est connexe et sans cycle simple est appelé un *arbre*. Un *arbre enraciné* est un arbre dans lequel on distingue un sommet qu'on appelle sa *racine*. Soit  $T$  un arbre et  $x$  un sommet de  $T$ , le sommet  $x$  est appelé une *feuille* si  $\deg(x) = 1$ . Le nombre de feuilles de l'arbre  $T$  est noté  $|T|_{\emptyset}$ . Une *chenille* est un arbre dont le sous-graphe induit par  $Y = \{x \in V \mid \deg(x) > 1\}$  est une chaîne simple, c'est-à-dire que toutes les feuilles du graphe sont adjacentes à la chaîne simple centrale appelée la *colonne* de la chenille. La figure 6.2 représente les différents types de graphes définis dans ce paragraphe.

Nous nous intéressons à des sous-graphes ayant une structure très particulière. Un *sous-arbre* de  $G$  est un sous-graphe induit connexe et acyclique de  $G$ , c'est-à-dire un sous-graphe induit qui est un arbre. L'ensemble des sous-arbres d'ordre  $i$  d'un graphe  $G$  est dénoté par  $\mathcal{T}_G^i$ . Un sous-arbre est dit *pleinement feuillu* si c'est un sous-arbre d'ordre  $i$  de  $G$  ayant le nombre maximal de feuilles, c'est-à-dire  $|T|_{\emptyset} = \max\{|T_1|_{\emptyset} : T_1 \in \mathcal{T}_G^i\}$ .

La *fonction feuille* d'un graphe  $G$ , dénoté  $L_G$ , de l'ensemble  $\llbracket 0, |G| \rrbracket$  vers  $\llbracket 0, |G| \rrbracket \cup$

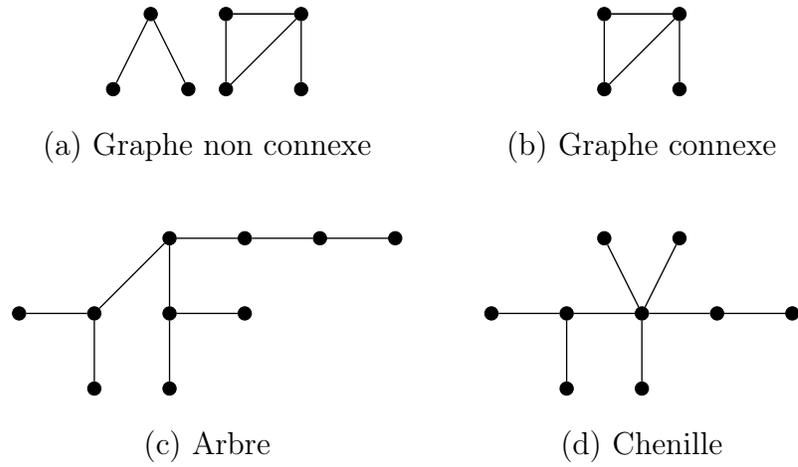


Figure 6.2: Différents types de graphes

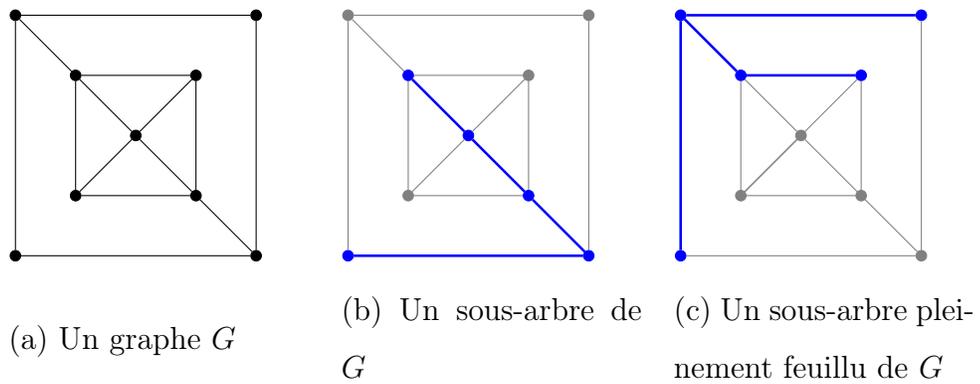


Figure 6.3: Un graphe et certains de ses sous-arbres induits

Tableau 6.1: Fonction feuille du graphe de la figure 6.3

$i$	0	1	2	3	4	5	6	7	8	9
$L_G(i)$	0	0	2	2	3	3	4	$-\infty$	$-\infty$	$-\infty$

$\{-\infty\}$  est définie par

$$L_G(i) = \max\{|T|_{\emptyset} : T \in \mathcal{T}_G^i\}.$$

On pose que  $\max \emptyset = -\infty$ . Ainsi, la fonction feuille calcule le nombre de feuilles d'un sous-arbre pleinement feuillu d'ordre  $i$  du graphe  $G$ .

**Exemple 6.6.** Soit  $G$  le graphe illustré à la figure 6.3. Sa fonction feuille est décrite dans le tableau 6.1. Le sous-arbre de la figure 6.3 (c) est pleinement feuillu, mais pas celui de la figure 6.3 (b). Pour calculer la fonction feuille, il suffit d'énumérer tous les sous-ensembles de sommets du graphe  $G$ , de vérifier si le sous-graphe induit par un sous-ensemble est un arbre et de calculer son nombre de feuilles.

Ce ne sont pas toutes les fonctions de  $\llbracket 1, n \rrbracket$  vers  $\mathbb{N} \cup \{-\infty\}$  qui sont des fonctions feuilles. On peut donc se demander s'il existe un critère simple pour identifier une fonction feuille. Plutôt que d'étudier directement les fonctions feuilles, nous associons un mot à une fonction feuille. La *dérivée discrète* aussi appelée *mot des premières différences*, associé à une fonction  $f : \llbracket 1, n \rrbracket \rightarrow \mathbb{N} \cup \{\omega\}$  le mot dont la  $i^e$  lettre est

$$\Delta f(i) = f(i+3) - f(i+2)$$

pour tout  $i \in \llbracket 1, n-3 \rrbracket$ . Par convention, on a que  $L(i+3) - L(i+2) = \omega$  si  $L(i+2) = -\infty$  ou  $L(i+3) = -\infty$ . Un *mot feuillu* est la dérivée discrète de la

fonction feuille  $L_G$  d'au moins un graphe simple  $G$ . Le mot feuillu du graphe  $G$  est dénoté par  $\Delta L_G$ .

Les premiers termes de la fonction feuille  $L_G$  sont égaux pour tous les graphes ayant au moins une arête et elle prend les valeurs  $L_G(0) = 0$ ,  $L_G(1) = 0$  et  $L_G(2) = 2$ . De plus,  $L_G(3) = 2$  pour tous les graphes connexes non isomorphes au graphe complet. Ainsi, pour la plupart des graphes connexes, le mot de différence  $(L_G(i) - L_G(i - 1))_{i \in \llbracket 1, n \rrbracket}$  des valeurs consécutives de  $L_G$  débute par le préfixe 020. Pour éviter de répéter cette information, nous avons choisi d'enlever les trois premières lettres de la dérivée discrète et de décaler les indices par 3 de gauche à droite dans la définition de mot feuillu. Par conséquent, le mot feuillu d'un graphe est défini seulement si le graphe est d'ordre au moins 4.

**Exemple 6.7.** Soit  $G$  le graphe à la figure 6.3. Le mot feuille de  $G$  est  $\Delta L_G = 101\omega\omega\omega$ .

Étudions la fonction feuille et le mot feuillu de différentes familles de graphes classiques. La fonction feuille des graphes complets  $K_n = (\llbracket 1, n \rrbracket, \mathcal{P}_2(\llbracket 1, n \rrbracket))$  est

$$L_{K_n}(i) = \begin{cases} 0, & \text{si } i = 0, 1; \\ 2, & \text{si } i = 2; \\ -\infty & \text{si } 3 \leq i \leq n; \end{cases}$$

car tous les sous-graphes induits de  $K_n$  d'ordre  $i \geq 3$  contiennent au moins un cycle. Ainsi le mot feuille de  $K_n$  est  $\omega^{n-3}$ .

La fonction feuille des cycles

$$C_n = (\llbracket 1, n \rrbracket, \{\{1, 2\}, \{2, 3\}, \dots, \{n-1, n\}, \{n, 1\}\})$$

est

$$L_{C_n}(i) = \begin{cases} 0, & \text{si } i = 0, 1; \\ 2, & \text{si } 2 \leq i < n; \\ -\infty & \text{si } i = n. \end{cases}$$

car la chaîne simple d'ordre  $i < n$  est un sous-arbre d'ordre  $i$  ayant deux feuilles et il n'y a pas de sous-arbre induit d'ordre  $n$ . Ainsi, le mot feuillu du graphe  $C_n$  est  $0^{n-4}\omega$  pour tout  $n \geq 4$ .

La fonction feuille des graphes roues  $R_n = (\llbracket 0, n \rrbracket, \{\{0, 1\}, \{0, 2\}, \dots, \{0, n\}\} \cup \{\{1, 2\}, \{2, 3\}, \dots, \{n-1, n\}, \{n, 1\}\})$ , c'est-à-dire le graphe cycle d'ordre  $n$  avec un sommet distingué qui est relié à chacun des sommets du cycle, est

$$L_{R_n}(i) = \begin{cases} 0, & \text{si } i = 0, 1; \\ 2, & \text{si } i = 2; \\ i - 1, & \text{si } 3 \leq i \leq \lfloor \frac{n}{2} \rfloor + 1; \\ 2, & \text{si } \lfloor \frac{n}{2} \rfloor + 2 \leq i \leq n - 1; \\ -\infty & \text{si } n \leq i \leq n + 1. \end{cases}$$

Pour  $i \leq \lfloor \frac{n}{2} \rfloor + 1$ , le sous-arbre incluant le sommet distingué et des sommets non adjacents du cycle est pleinement feuillu. Ensuite, les sous-arbres pleinement feuillus sont isomorphes à la chaîne simple d'ordre  $i$  et il n'y a pas de sous-arbre d'ordre  $n$ , ni  $n + 1$ . Ainsi, le mot feuillu du graphe  $R_n$  est  $1^k(-k)0^j\omega\omega$  avec  $k = \lfloor \frac{n}{2} \rfloor - 2$  et  $j = k$  si  $n$  est pair et  $j = k - 1$  sinon.

La fonction feuille du graphe biparti complet

$$K_{p,q} = (\{x_1, \dots, x_p, y_1, \dots, y_q\}, \{\{x_i, y_j\} | 1 \leq i \leq p, 1 \leq j \leq q\})$$

est

$$L_{K_{p,q}}(i) = \begin{cases} 0, & \text{si } i = 0, 1; \\ 2, & \text{si } i = 2; \\ i - 1, & \text{si } 3 \leq i \leq \max(p, q) + 1; \\ -\infty & \text{si } \max(p, q) + 2 \leq i \leq p + q, \end{cases}$$

car les sommets  $x_i$  sont non-adjacents, de même que les sommets  $y_i$ . Alors il suffit de prendre un sommet dans l'un des ensembles et tous les sommets restants dans l'autre ensemble pour obtenir un sous-arbre induit d'ordre  $i$  où  $i \leq \max(p, q) + 1$ . Pour tout  $i > \max(p, q) + 1$ , il n'y a pas de sous-arbres, car tous les sous-graphes induits contiennent au moins un cycle de longueur 4. Par conséquent, le mot feuillu du graphe  $K_{p,q}$  est  $1^{\max(p,q)-2}\omega^{p+q-\max(p,q)+1}$ .

La fonction feuille de la grille infinie  $\mathbb{Z}^2$ , le graphe dont les sommets sont les paires d'entiers  $(i, j)$  et l'ensemble des arêtes est

$$\bigcup_{(i,j) \in \mathbb{Z}^2} \{ \{(i, j), (i + 1, j)\} \cup \{(i, j), (i, j + 1)\} \}.$$

est

$$L_{\mathbb{Z}^2}(i) = \begin{cases} 0, & \text{si } i = 0, 1; \\ 2, & \text{si } i = 2; \\ i - 1, & \text{si } 3 \leq i \leq 5; \\ L_{\mathbb{Z}^2}(i - 4) + 2 & \text{si } i \geq 6. \end{cases}$$

tel que démontré dans (Blondin Massé *et al.*, 2018b). Ainsi, le mot feuillu du graphe  $\mathbb{Z}^2$  est  $(1100)^\infty$ . Dans (Blondin Massé *et al.*, 2018b), les fonctions feuilles de la grille hexagonale, triangulaire et cubique sont aussi calculées. L'alphabet des mots feuillus est simple à décrire.

**Lemme 6.8.** *L'ensemble  $\mathcal{A} = \{1, 0, -1, -2, \dots, \omega\}$  est le plus petit ensemble tel que pour tout graphe  $G$  d'ordre au moins 3,  $\Delta L_G \in \mathcal{A}^*$ .*

*Démonstration.* Les graphes  $C_n$  et  $\mathbb{Z}^2$  montrent les lettres 0, 1 et  $\omega$  apparaissent dans des mots feuillus. De plus, la lettre  $-k$  apparaît dans le graphe roue  $R_{2k+4}$ . Ainsi, il reste à démontrer que le mot feuille d'un graphe  $G$  quelconque ne peut pas contenir la lettre  $k$  où  $k$  est un entier supérieur à 1 par l'absurde.

Supposons qu'il existe un graphe  $G$  d'ordre au moins 3 et un entier  $i$  tel que  $L_G(i+3) - L_G(i+2) = k$ . Alors,  $G$  possède un sous-arbre  $T$  d'ordre  $i+3$  tel que  $L_G(i+3) = k$ . En enlevant une feuille à  $T$ , on obtient un sous-arbre  $T'$  d'ordre  $i+2$ . Toutes les feuilles de  $T$ , sauf celle qui a été enlevée, sont des feuilles de  $T'$ . Donc,  $G$  contient un sous-arbre d'ordre  $i+2$  ayant  $L_G(i+2) + k - 1$ , ce qui est impossible par la définition de la fonction feuille. ■

Une observation élémentaire, mais utile est que l'alphabet des mots feuillus indique si le graphe associé est un arbre.

**Lemme 6.9.** *Le graphe  $G$  est un arbre si et seulement si  $\Delta L_G \in \{0, 1\}^*$ .*

*Démonstration.* Soit  $T$  un arbre d'ordre  $n$ . Il suffit de démontrer que la fonction feuille de  $T$  est croissante, car  $\Delta L_T \in \{1, 0, -1, -2, \dots, \omega\}^*$  par le lemme 6.8. Posons  $l = L_T(i)$ . Alors, il existe un sous-arbre  $T'$  d'ordre  $i$  qui est pleinement feuillu. Si  $i < n$ , il existe un sommet  $v$  adjacent à un sommet de  $T'$  dans  $T$ . Le sous-graphe induit par  $T' \cup \{v\}$  est un sous-arbre d'ordre  $i+1$ . De plus, le sous-arbre  $T' \cup \{v\}$  a au moins  $l$  feuilles, car ajouter un sommet à  $T'$  ajoute forcément une feuille à  $T'$  et au plus une feuille de  $T'$  devient un sommet de degré supérieur à 1. Ainsi,  $L_T(i+1) \geq l \geq L_T(i)$ . Donc, la fonction feuille de  $T$  est croissante. Comme  $L_T(0) = 0$ , la fonction feuille ne prend pas la valeur  $-\infty$  et son mot feuillu ne contient pas  $\omega$ .

$i$	0	1	2	3	4	5	6	7	8	9
$L_T(i)$	0	0	2	2	3	4	4	5	5	6
$L_T(i+1) - L_T(i)$	0	2	0	1	1	0	1	0	1	

Figure 6.4: Un arbre et sa fonction feuille

Réciproquement, soit  $G$  un graphe d'ordre  $n$ . Supposons que  $\Delta L_G \in \{0, 1\}^*$ , alors  $L_G(n) - L_G(n-1) \neq \omega$ . Ainsi,  $L_G(n) \neq -\infty$ . Par conséquent,  $G$  contient un sous-arbre d'ordre  $n$ . Or le seul sous-arbre de  $G$  d'ordre  $n$  est le graphe  $G$  lui-même. Donc,  $G$  est forcément un arbre. ■

Il y a un lien direct entre la fonction feuille d'un arbre et le nombre de 1 apparaissant dans son mot feuillu : posons  $\Delta L_G = w_1 \dots w_n$ , alors

$$L_T(j) - L_T(i) = |w_{i-2} \dots w_{j-3}|_1 \quad (6.1)$$

pour  $3 \leq i \leq j \leq |T| = n$ . En effet, en appliquant le lemme 6.9, on a que

$$|w_{i-2} \dots w_{j-3}|_1 = \sum_{k=i-2}^{j-3} \Delta L_G(k) = \sum_{k=i-2}^{j-3} (L_T(k-3) - L_T(k-2)).$$

**Exemple 6.10.** Dans le graphe  $G$  de la figure 6.4, l'équation (6.1) avec  $i = 4$  et  $j = 7$  est satisfaite puisque  $L_G(7) - L_G(4) = 2 = |101|_1 = |w_2 \dots w_4|_1$ .

Une question naturelle est : est-ce que tout mot  $w \in \{1, 0, -1, \dots, \omega\}^*$  est un mot feuillu ? Autrement dit, existe-t-il un graphe  $G$  dont le mot feuillu  $\Delta L_G = w$  ? Nous ne connaissons pas de caractérisation des mots feuillus pour les graphes quelconques. Par contre, nous pouvons caractériser les mots feuillus des graphes chenilles. Le reste du chapitre est consacré à l'étude des mots feuillus de cette famille de graphe.

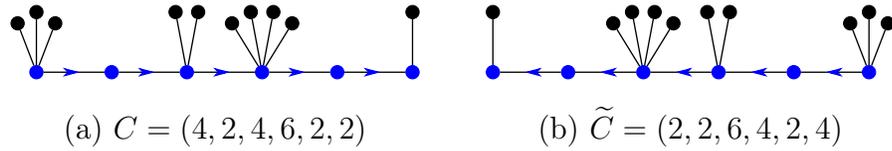


Figure 6.5: Le graphe chenille  $S$  avec sa colonne orientée en bleu et le graphe chenille  $\tilde{S}$ .

### 6.3 Graphes chenilles orientées

Nous décrivons les chenilles comme des objets orientés. Nous choisissons une orientation pour la colonne de la chenille  $C$ , car cela permet de représenter toutes les chenilles par une suite d'entiers  $(c_1, \dots, c_r)$  qui représentent la suite des degrés de chacun des sommets sur la colonne de la chenille. Une *chenille orientée* est un graphe chenille dont la colonne a une orientation. L'image miroir de la chenille orientée  $C$ , notée  $\tilde{C}$ , est la même chenille dont la colonne est orientée dans la direction opposée, c'est-à-dire que  $\tilde{C} = (c_r, \dots, c_1)$ . Dans tous les exemples dessinés, l'orientation de la colonne d'une chenille est fixée de gauche à droite.

**Exemple 6.11.** *Le graphe illustré à la figure 6.5 (a) est la chenille orientée  $(4, 2, 4, 6, 2, 2)$  et son image miroir est illustrée à la figure 6.5 (b).*

La représentation d'une chenille sous forme de suite de degrés permet de compter facilement son nombre de sommets et son nombre de feuilles grâce aux identités

$$|C| = \sum_{i=1}^r c_i - r + 2 \quad |C|_{\text{f}} = \sum_{i=1}^r c_i - 2r + 2 \quad (6.2)$$

où  $C = (c_1, \dots, c_r)$ .

L'ensemble des chenilles orientées est noté  $\mathcal{CO}$ . Soit  $C = (c_1, \dots, c_r)$  et  $C' = (c'_1, \dots, c'_r)$  deux chenilles. La *greffe* de  $C$  et  $C'$ , notée  $C \diamond C'$ , est la chenille

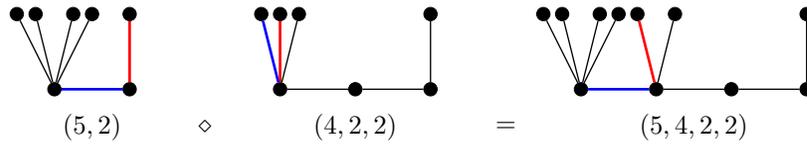


Figure 6.6: La greffe des chenilles orientées  $(5, 2)$  et  $(4, 2, 2)$ .

orientée

$$C \diamond C' = (c_1, \dots, c_{r-1}, c_r + c'_1 - 2, c'_2, \dots, c'_r).$$

**Exemple 6.12.** La greffe des chenilles orientées  $(5, 2)$  et  $(4, 2, 2)$  est la chenille orientée  $(5, 2) \diamond (4, 2, 2) = (5, 4, 2, 2)$  tel qu'illustré à la figure 6.6. Les arêtes colorées en bleu du côté gauche de l'équation sont fusionnées pour obtenir l'arête bleue de l'autre côté, de même les arêtes rouges sont fusionnées pour obtenir l'arête rouge.

La définition de la greffe peut sembler artificielle, mais elle permet à la plus petite chenille orientée,  $S = (2)$ , d'être l'élément neutre de cette opération. De plus, la greffe est compatible avec les fonctions qui indiquent le nombre de sommets et le nombre de feuilles d'une chenille. Soit  $C$  et  $C'$  deux chenilles orientées, on a que

$$|C \diamond C'| = |C| + |C'| - 3 \quad (6.3)$$

$$|C \diamond C'|_{\varnothing} = |C|_{\varnothing} + |C'|_{\varnothing} - 2. \quad (6.4)$$

L'ensemble des chenilles orientées  $\mathcal{CO}$  muni de la greffe est un monoïde dont l'élément neutre est la chenille orientée  $(2)$ . En plus, nous montrons dans la suite que le monoïde  $(\mathcal{CO}, \diamond)$  est isomorphe au monoïde libre  $(\{0, 1\}^*, \cdot)$ . Soit  $\text{RC}$  un morphisme du monoïde libre  $\{0, 1\}^*$  vers l'ensemble  $\mathcal{CO}$  défini par :

$$\text{RC}(x) = \begin{cases} (2, 2), & \text{si } x = 0 \\ (3), & \text{si } x = 1. \end{cases}$$

avec  $x \in \{0, 1\}$ . On note que  $\text{RC}(\varepsilon) = (2)$ . La chenille orientée obtenue à l'aide du morphisme RC est appelée la *chenille de lecture* du mot binaire  $w$ .

**Exemple 6.13.** *La chenille de lecture associée au mot 0110 est*

$$\text{RC}(0110) = (2, 2) \diamond (3) \diamond (3) \diamond (2, 2) = (2, 4, 2).$$

**Proposition 6.14.** *La fonction RC est un isomorphisme de monoïde. Ainsi, L'ensemble  $\mathcal{CO}$  muni de l'opération greffe est un monoïde libre.*

*Démonstration.* La fonction RC est un morphisme de monoïde, alors il suffit de montrer que RC est bijective.

D'une part, la fonction RC est surjective, car une chenille orientée  $C = (c_1, \dots, c_r)$  peut être décomposée en un nombre fini de chenilles orientées  $(2, 2)$  ou  $(3)$  de la façon suivante : si  $c_r > 2$ , alors  $(c_1, \dots, c_r) = (c_1, \dots, c_r - 1) \diamond (3)$  ou  $c_r = 2$  et  $(c_1, \dots, c_{r-1}) \diamond (2, 2)$ . Ensuite, il suffit de remplacer les chenilles orientées  $(2, 2)$  par 0, les chenilles orientées  $(3)$  par 1 et la greffe par la concaténation pour obtenir un mot sur le monoïde  $\{0, 1\}^*$ . Comme toutes les chenilles orientées ont une telle décomposition, il existe toujours un mot  $w$  sur  $\{0, 1\}^*$  tel que  $\text{RC}(w) = C$ .

D'autre part, la fonction RC est injective, car

$$\text{RC}(u) = \text{RC}(v)$$

$$C_1 \diamond C_2 \diamond \dots \diamond C_r = C'_1 \diamond C'_2 \diamond \dots \diamond C'_r$$

avec  $C_i = C'_i$  pour tout  $i \in \llbracket 1, r \rrbracket$  et  $C_i, C'_i \in \{(2, 2), (3)\}$ . En associant la chenille  $(2, 2)$  sur 0 et  $(3)$  sur 1 et on obtient que  $u = v$ . Par conséquent, le monoïde  $\mathcal{CO}$  est isomorphe au monoïde libre  $\{0, 1\}^*$ . ■

Soit  $r$  et  $r'$  deux entiers positifs non nuls tels que  $r' \leq r$  et  $C = (c_1, \dots, c_r)$  et  $C' = (c'_1, \dots, c'_{r'})$  deux chenilles orientées. La chenille orientée  $C'$  est une *sous-chenille* de  $C$  et on écrit alors  $C' \leq C$ , si  $C'$  est un sous-arbre de  $C$  et qu'il existe

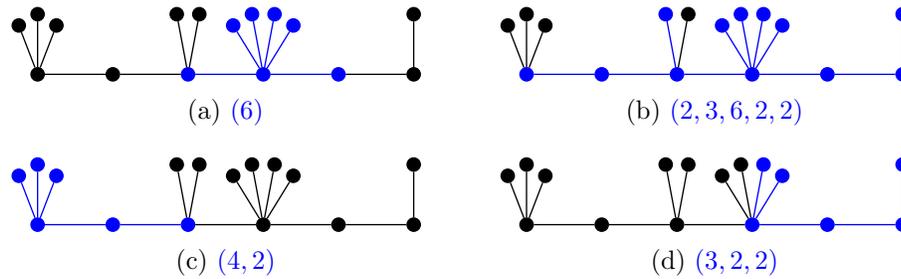


Figure 6.7: La chenille orientée  $(4, 2, 4, 6, 2, 2)$  et quelques-unes de ses sous-chenilles colorées en bleu

un entier  $k \in \llbracket 1, r - r' - 1 \rrbracket$  tel que

$$c'_j \leq c_j$$

pour tout  $j \in \llbracket 1, r' \rrbracket$ . Intuitivement, une chenille orientée  $C'$  est une sous-chenille de  $C$  si  $C'$  est un sous-arbre de  $C$  et que l'orientation de la colonne de  $C'$  coïncide avec celle de  $C$ .

Ainsi, la fonction feuille d'une chenille  $C = (c_1, \dots, c_r)$  satisfait l'équation

$$L_C(i) = \max\{|C'|_{\emptyset} : C' \leq C \text{ et } |C'| = i\}.$$

Une sous-chenille  $C'$  de  $C$  est dite *pleinement feuillue* si c'est un sous-arbre pleinement feuillu de  $C$ .

**Exemple 6.15.** *Les chenilles orientées  $(6)$ ,  $(2, 3, 6, 2, 2)$ ,  $(4, 2)$  et  $(3, 2, 2)$  sont des sous-chenilles de la chenille orientée  $(4, 2, 4, 6, 2, 2)$ . Elles sont toutes illustrées à la figure 6.7.*

La relation  $\leq$  est une relation d'ordre sur l'ensemble des chenilles orientées. La figure 6.8 illustre le diagramme de Hasse de la relation d'ordre sur les chenilles

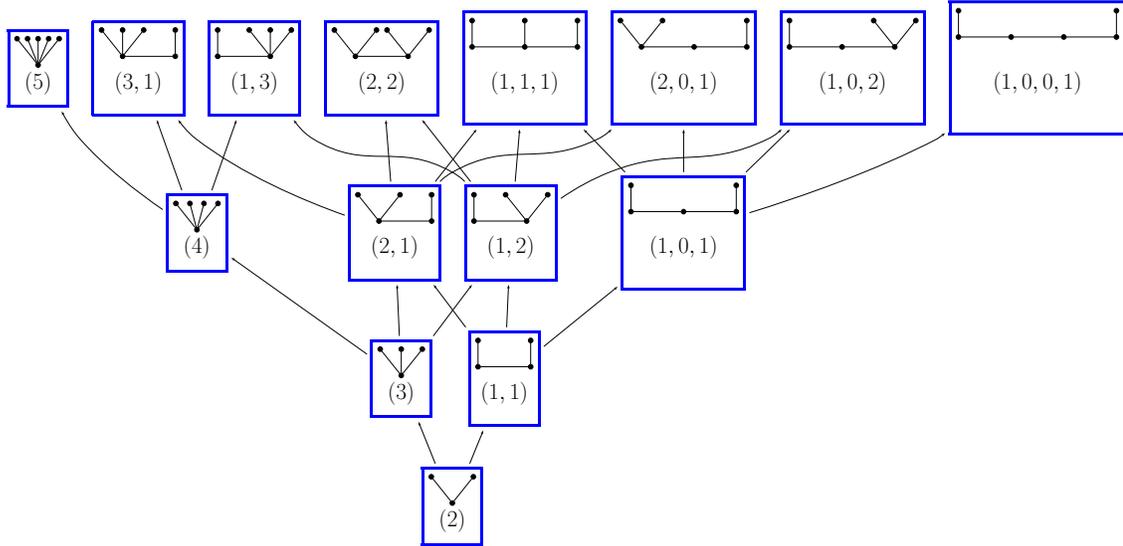


Figure 6.8: Le diagramme de Hasse de l'ensemble partiellement ordonné  $(CO, \leq)$

orientées. On observe que l'ensemble partiellement ordonné  $(CO, \leq)$  n'est pas un treillis, puisque les chenilles  $(1, 1)$  et  $(3)$  sont toutes deux couvertes par les chenilles  $(1, 2)$  et  $(2, 1)$  dans  $(CO, \leq)$ .

La bijection entre les chenilles orientées et les mots binaires permet de définir des notions analogues aux facteurs, préfixes et suffixes sur les chenilles de lecture.

**Proposition 6.16.** *Soit  $w \in \{0, 1\}^*$ . Si  $u$  est un facteur de  $w$ , alors  $RC(u)$  est une sous-chenille de  $RC(w)$ .*

**Remarque 6.17.** *Ce ne sont pas toutes les sous-chenilles de  $RC(w)$  qui correspondent à des facteurs de  $w$ . La chenille  $RC(010111100) = (2, 3, 6, 2, 2)$  est une sous-chenille de la chenille*

$$RC(1100110111100) = (4, 2, 4, 6, 2, 2),$$

*tel qu'illustré à la figure 6.7, mais le mot 010111100 n'est pas un facteur de 1100110111100.*

*Démonstration.* Soit  $w \in \{0, 1\}^*$  et  $u$  un facteur de  $w$ . Il existe des mots  $x, y \in \{0, 1\}^*$  tels que  $w = xuy$ . Ainsi,

$$\text{RC}(w) = \text{RC}(x) \diamond \text{RC}(u) \diamond \text{RC}(y).$$

Posons  $\text{RC}(x) = (x_1, \dots, x_r)$ ,  $\text{RC}(u) = (u_1, \dots, u_s)$  et  $\text{RC}(y) = (y_1, \dots, y_t)$ . On a alors que  $\text{RC}(w) = (w_1, \dots, w_k) = (x_1, \dots, x_r + u_1 - 2, u_2, \dots, u_r + y_1 - 2, y_2, \dots, y_r)$  par la définition de greffe. En posant  $i = r - 1$ , on a que  $u_j \leq w_{j+i}$  pour tout  $j \in \llbracket 1, s \rrbracket$ . Donc,  $\text{RC}(u) \leq \text{RC}(w)$ . ■

De même, les sous-chenilles situées aux extrémités d'une chenille correspondent aux préfixes et aux suffixes d'un mot. Soit  $i \in \llbracket 3, |C| \rrbracket$ . La *sous-chenille gauche* d'ordre  $i$  de la chenille  $C = (c_1, \dots, c_r)$  est la sous-chenille  $\text{Gauche}_i(C) = (g_1, \dots, g_k)$  où  $1 \leq k \leq r$ ,  $g_j = c_j$  pour tout  $j \in \llbracket 1, k - 1 \rrbracket$  et  $2 \leq g_k \leq c_k$  tel que

$$i = \sum_{j=1}^k g_j - k + 2.$$

De même, la *sous-chenille droite* d'ordre  $i$  de la chenille  $C = (c_1, \dots, c_r)$  est la sous-chenille  $\text{Droite}_i(C) = (d_k, \dots, d_r)$  où  $1 \leq k \leq r$ ,  $d_j = c_{j+k-1}$  pour tout  $j \in \llbracket 1, r - k + 1 \rrbracket$  et  $2 \leq g_k \leq c_k$  tel que

$$i = \sum_{j=k}^r d_j - (r - k) + 2.$$

**Exemple 6.18.** Soit  $C = (4, 2, 4, 6, 2, 2)$ . La sous-chenille gauche d'ordre 6 de  $C$  est  $\text{Gauche}_6(C) = (4, 2)$  et la sous-chenille droite d'ordre 6 est  $\text{Droite}_6(C) = (3, 2, 2)$ . La sous-chenille  $(2, 3, 6, 2, 2)$  n'est pas une sous-chenille droite de  $C$ , car  $d_2 \neq c_3$ . Toutes ces chenilles orientées sont illustrées à la figure 6.7

Une chenille  $C$  peut être décomposée en une sous-chenille droite et une sous-chenille gauche à l'aide de la greffe.

**Lemme 6.19.** *Si  $C$  est une chenille orientée, alors*

$$C = \text{Gauche}_i(C) \diamond \text{Droite}_{|C|+3-i}(C).$$

pour tout  $i \in \llbracket 3, |C| \rrbracket$ .

*Démonstration.* Soit  $C = (c_1, \dots, c_r)$  une chenille orientée. Posons,  $n = |C|$ . On a que  $\text{Gauche}_i(C) = (g_1, \dots, g_k)$  et  $\text{Droite}_{n+3-i}(C) = (d_{k'}, \dots, d_r)$  avec  $1 \leq k \leq r$ ,  $1 \leq k' \leq r$ ,  $g_j = c_j$  pour tout  $j \in \llbracket 1, k-1 \rrbracket$ ,  $g_k = c_k$ ,  $d_j = c_{j+k-1}$  pour tout  $j \in \llbracket 2, r-k'+1 \rrbracket$ ,  $d_1 = c'_k$  tel que

$$i = \sum_{j=1}^k g_j - k + 2 \text{ et } n + 3 - i = \sum_{j=k'}^r d_j - (r - k') + 2.$$

En additionnant les deux égalités ci-dessus, on obtient que

$$\begin{aligned} n + 3 &= \sum_{j=1}^k g_j + \sum_{j=1}^{r-k'+1} d_j - k + 2 - (r - k') + 2 \\ &= \sum_{j=1}^{k-1} g_j + g_k + \sum_{j=2}^{r-k'+1} c_{j+k'-1} + d_{k'} - k - r + k' + 4 \\ &= \sum_{j=1}^{k-1} c_j + g_k + \sum_{j=k'+1}^r c_j + d_1 - k - r + k' + 4 \\ &= \sum_{j=1}^{k-1} c_j + \sum_{j=k'+1}^r c_j + (g_k + d_1 - 2) + (-k + k') - r + 4 \\ &= \sum_{j=1}^k c_j - r + 2 \end{aligned}$$

car l'ordre des sous-chenilles gauches et droites implique que les deux sous-chenilles partagent 3 sommets, par conséquent,  $k = k'$  et  $c_k = g_k + d_1 - 2$ . Ainsi,  $\text{Gauche}_i(C) \diamond \text{Droite}_{|C|+3-i}(C) = C$ . ■

Les sous-chenilles droites et gauches d'une chenille de lecture correspondent aux préfixes et suffixes du mot qui engendre la chenille de lecture.

**Lemme 6.20.** *Si  $w$  est un mot binaire sur  $\{0, 1\}^*$ ,  $a \in \{0, 1\}$  et  $3 \leq i \leq |w| + 3$ , alors*

1.  $|\text{RC}(w)| = |w| + 3$ ;
2.  $|\text{RC}(wa)|_{\emptyset} = |\text{RC}(w)|_{\emptyset} + a$ ;
3.  $|\text{RC}(w)|_{\emptyset} = |w|_1 + 2$ ;
4.  $\text{Gauche}_i(\text{RC}(w)) = \text{RC}(\text{Pref}_{i-3}(w))$ ;
5.  $\text{Droite}_i(\text{RC}(w)) = \text{RC}(\text{Suff}_{i-3}(w))$ ;
6.  $|\text{Gauche}_i(\text{RC}(w))|_{\emptyset} = |\text{Pref}_{i-3}(w)|_1 + 2$ ;
7.  $|\text{Droite}_i(\text{RC}(w))|_{\emptyset} = |\text{Suff}_{i-3}(w)|_1 + 2$ ;

*Démonstration.* 1. Procédons par récurrence sur la longueur de  $w$ . Si  $|w| = 0$ , alors  $w = \varepsilon$  et  $\text{RC}(\varepsilon) = (2)$  est bien d'ordre  $3 = |w| + 3$ . Supposons que  $|\text{RC}(w)| = |w| + 3$  pour tout mot  $w$  tel que  $|w| = n$ . Soit  $w'$  un mot de longueur  $n + 1$ . On a alors que  $\text{RC}(w') = \text{RC}(wa)$  avec  $w \in \{0, 1\}^*$  et  $a \in \{0, 1\}$ . Par définition, on a que  $\text{RC}(wa) = \text{RC}(w) \diamond (2, 2)$  ou bien  $\text{RC}(w) \diamond (3)$ . Dans les deux cas  $|\text{RC}(w)| = |\text{RC}(w)| + 1 = |w| + 4 = |w'| + 3$  par l'hypothèse de récurrence.

2. On a que  $\text{RC}(wa) = \text{RC}(w) \diamond \text{RC}(a)$ . Par conséquent,

$$\begin{aligned} |\text{RC}(wa)|_{\emptyset} &= |\text{RC}(w) \diamond \text{RC}(a)|_{\emptyset} \\ &= |\text{RC}(w)|_{\emptyset} + |\text{RC}(a)|_{\emptyset} - 2 \\ &= |\text{RC}(w)|_{\emptyset} + a, \end{aligned}$$

par l'équation 6.4 et  $|\text{RC}(a)|_{\emptyset} = a + 2$ .

3. Procédons par récurrence sur la longueur de  $w$ . Si  $|w| = 0$ , alors  $|\text{RC}(\varepsilon)|_{\emptyset} = 2 = |w|_1 + 2$ . Supposons que pour tout mot  $w'$  tel que  $|w'| = n - 1$ , on a que  $|\text{RC}(w')|_{\emptyset} = |w'|_1 + 2$ . Soit  $w$  un mot de longueur  $n$ , alors  $w = w'a$  avec

$w' \in \{0, 1\}^*$  et  $a \in \{0, 1\}$ . Ainsi,  $|\text{RC}(w)|_{\emptyset} = |\text{RC}(w'a)|_{\emptyset} = |\text{RC}(w')|_{\emptyset+a} = |w'|_1 + 2 + |a|_1 = |w|_1 + 2$  par la propriété 3 et l'hypothèse de récurrence.

4. On a que Gauche construit une chenille comportant les  $i$  premiers sommets de  $\text{RC}(w)$ . De même, RC construit les chenilles de gauche à droite en ajoutant un sommet à la fois par la propriété 1. Comme chaque lettre correspond à un sommet de RC, on obtient l'égalité souhaitée.
5. La démonstration est analogue à celle de la propriété 4.

Les propriétés 6 et 7 sont obtenues en combinant les propriétés 3, 4 et 5. ■

#### 6.4 Mots feuillus des graphes chenilles

Dans cette section, nous démontrons que les mots préfixes normaux correspondent aux dérivées discrètes de la fonction feuille d'un graphe chenille. De plus, nous caractérisons les chenilles orientées ayant la même fonction feuille à l'aide de la forme préfixe normale. Par exemple, le mot feuillu de la chenille  $\text{RC}(00110101100)$  est 11010110000 qui est bel et bien un mot préfixe normal.

Dans la section 6.3, nous avons démontré que les chenilles de lecture des facteurs du mot  $w$  sont des sous-chenilles de  $\text{RC}(w)$ . Évidemment, ce ne sont pas toutes les sous-chenilles qui sont de  $\text{RC}(w)$  qui sont de la forme  $\text{RC}(u)$  où  $u$  est un facteur de  $w$ . De même, certains sous-arbres pleinement feuillus de la chenille  $\text{RC}(w)$  correspondent à des facteurs de  $w$ , mais pas tous. Par exemple, nous avons illustré à la figure 6.9 la chenille  $\text{RC}(00110101100)$  dont  $L_{\text{RC}(00110101100)}(8) = 5$  et deux de ses sous-arbres pleinement feuillus. Celui en bleu correspond à  $\text{RC}(110101)$  avec 110101 un facteur de 00110101100, mais celui en rouge ne correspond pas à la chenille de lecture d'un facteur du mot 00110101100. Comme les sous-chenilles gauches et droites correspondent aux préfixes et aux suffixes de  $w$ , un facteur de  $w$  peut être décrit à l'aide des sous-chenilles gauches et droites. Le lemme 6.21

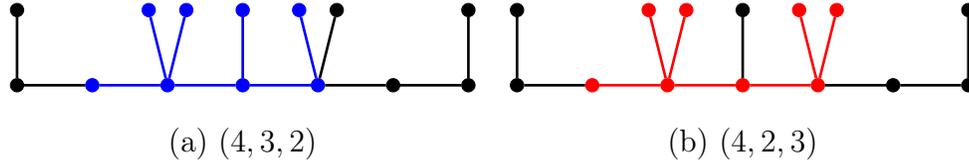


Figure 6.9: Deux sous-arbres pleinement feuillus de  $\text{RC}(00110101100)$ .

démontre qu'il existe toujours au moins un facteur  $u$  de  $w$  tel que  $\text{RC}(u)$  soit un sous-arbre pleinement feuillu de  $\text{RC}(w)$ .

**Lemme 6.21.** *Soit  $w$  un mot binaire. Pour tout entier  $i \in \llbracket 3, |w| + 3 \rrbracket$ , il existe des entiers  $j$  et  $j'$  tel que*

- $\text{Gauche}_i(\text{Droite}_j(\text{RC}(w)))$
- $\text{Droite}_i(\text{Gauche}_{j'}(\text{RC}(w)))$

*sont des sous-chenilles pleinement feuillues de  $\text{RC}(w)$ .*

*Démonstration.* Posons  $\text{RC}(w) = (s_1, \dots, s_r)$ . Soit une sous-chenille pleinement feuillue  $C = (c_1, \dots, c_k)$  de  $\text{RC}(w)$  d'ordre  $i$ . Puisque  $C$  est une sous-chenille de  $\text{RC}(w)$ , il existe un entier  $n \in \llbracket 0, r - k \rrbracket$  tel que  $c_m \leq s_{n+m}$  pour tout  $m \in \llbracket 1, k \rrbracket$ .

Posons  $d = \sum_{m=2}^{k-1} (s_{n+m} - c_m)$ . Démontrons par l'absurde que  $c_1 + c_k - 4 \geq d$ . En effet, supposons que  $c_1 + c_k - 4 < d$ . L'idée est qu'on peut construire une sous-chenille de  $\text{RC}(w)$  d'ordre  $|C|$ , mais ayant plus de feuilles. On enlève  $c_1 + c_k - 4$  sommets à la chenille  $C$  afin d'obtenir la chenille  $C' = (c_2, \dots, c_{k-1}, 2)$ . L'ordre de la chenille  $C'$  est  $|C| - c_1 - c_k + 3$  et elle a  $|C|_{\emptyset} - c_1 - c_k + 2$  feuilles. Par hypothèse, on peut ajouter  $c_1 + c_k - 3$  feuilles à la chenille  $C'$ . Formellement, on choisit un entier  $z \in \llbracket 2, k - 1 \rrbracket$  maximal tel que  $c_1 + c_k - 3 \geq \sum_{m=2}^z (s_{n+m} - c_m)$ . On prend  $t = c_1 + c_k - 3 - \sum_{m=2}^z (s_{n+m} - c_m)$ . On obtient alors la chenille

$$C'' = (s_{n+2}, s_{n+3}, \dots, s_{n+z}, c_{z+1} + t, c_{z+2}, \dots, c_{k-1}, 2)$$

qui est une sous-chenille de  $\text{RC}(w)$ . De plus,  $|C''| = |C|$  et  $|C''|_{\emptyset} = |C|_{\emptyset} + 1$ . Ce qui contredit le fait que  $C$  est pleinement feuillue.

Ainsi, nous avons que  $c_1 - c_k - 4 \geq d$ . Par conséquent,  $S = (s_1, c_{n+2}, \dots, c_{n+k-1}, s_k)$  avec  $s_1 = c_1 - \min(c_1 - 2, d)$  et  $s_k = c_1 + c_k - d - s_1$  est une sous-chenille pleinement feuillue de  $\text{RC}(w)$ , car  $|C| = |S|$  et  $|C|_{\emptyset} = |S|_{\emptyset}$ . De plus, la chenille  $S$  est de la forme  $\text{Gauche}_i(\text{Droite}_j(\text{RC}(w)))$  avec

$$j = n - \sum_{m=1}^n c_m - c_{k+1} + s_1.$$

La démonstration pour le cas  $\text{Droite}_i(\text{Gauche}_j(\text{RC}(w)))$  est symétrique. ■

**Corollaire 6.22.** *Soit  $w$  un mot binaire. Pour toute taille  $i \geq 3$ , il existe un facteur  $u$  de  $w$  de longueur  $i - 3$  tel que  $\text{RC}(u)$  est un sous-arbre pleinement feuillu de  $\text{RC}(w)$ .*

*Démonstration.* Posons  $n = |w|$  et  $w = w_1 \dots w_n$  avec  $w_i \in \{0, 1\}$ . Rappelons que  $\text{RC}(w)$  est une chenille d'ordre  $n + 3$ . Soit  $i \in \{3, \dots, n + 3\}$ . Par le lemme 6.21, il existe un indice  $j \geq i$  tel que  $\text{Gauche}_i(\text{Droite}_j(\text{RC}(w)))$  est un sous-arbre pleinement feuillu de  $\text{RC}(w)$  d'ordre  $i$ . Posons  $u = w_{n-j+4} \dots w_{n-j+i}$ . Par le lemme 6.20, nous avons que

$$\begin{aligned} \text{Gauche}_i(\text{Droite}_j(\text{RC}(w))) &= \text{Gauche}_i(\text{RC}(\text{Suff}_{j-3}(w))) \\ &= \text{RC}(\text{Pref}_{i-3}(\text{Suff}_{j-3}(w))) \\ &= \text{RC}(w_{n-j+4} \dots w_{n-j+i}) \\ &= \text{RC}(u) \end{aligned}$$

et  $|u| = i - 3$  tel que souhaité. ■

*Démonstration du théorème 6.4.* Nous démontrons d'abord que si  $w$  est un mot

préfixe normal, alors il existe une chenille  $C$  telle que  $\Delta L_C = w$ . Soit  $w$  un mot préfixe normal et  $3 \leq i \leq |w| + 3$ . Il faut démontrer que

1.  $L_{\text{RC}(w)}(i) = |\text{Gauche}_i(\text{RC}(w))|_{\emptyset}$  ;
2.  $\Delta L_{\text{RC}(w)} = w$ .

Tout d'abord,  $L_{\text{RC}(w)}(i) \geq |\text{Gauche}_i(\text{RC}(w))|_{\emptyset}$ , car  $\text{Gauche}_i(\text{RC}(w))$  est une sous-chenille d'ordre  $i$  de  $\text{RC}(w)$ . Par conséquent, il suffit de démontrer que  $L_{\text{RC}(w)}(i) \leq |\text{Gauche}_i(\text{RC}(w))|_{\emptyset}$ . Par le corollaire 6.22, il existe un facteur  $u$  de  $w$  de longueur  $i - 3$  tel que  $\text{RC}(u)$  est un sous-arbre pleinement feuillu de  $\text{RC}(w)$ . Par conséquent,

$$\begin{aligned}
 L_{\text{RC}(w)} &= |\text{RC}(u)|_{\emptyset} \\
 &= |u|_1 + 2 && \text{par lemme 6.20} \\
 &\leq F_1(w, i - 3) + 2 \\
 &= |\text{Pref}_{i-3}(w)|_1 + 2 && \text{car } w \text{ est préfixe normal} \\
 &= |\text{Gauche}_i(\text{RC}(w))|_{\emptyset} && \text{par le lemme 6.20.}
 \end{aligned}$$

Pour  $1 \leq i \leq |w|$ , nous avons que

$$\begin{aligned}
 \Delta L_{\text{RC}(w)}(i) &= L_{\text{RC}(w)}(i + 3) - L_{\text{RC}(w)}(i + 2) \\
 &= |\text{Gauche}_{i+3}(\text{RC}(w))|_{\emptyset} - |\text{Gauche}_{i+2}(\text{RC}(w))|_{\emptyset} \\
 &= (|\text{Pref}_i(w)|_1 + 2) - (|\text{Pref}_{i-1}(w)|_1 + 2) && \text{par le lemme 6.20} \\
 &= w_i
 \end{aligned}$$

tel que souhaité.

Ensuite, nous démontrons que si  $C$  est une chenille, alors  $\Delta L_C$  est préfixe normal par l'absurde. Supposons que  $\Delta L_C$  n'est pas préfixe normal. Par le lemme 6.3, il existe deux mots  $p$  et  $f$ , avec  $|p|_1 = |f|_1$  tel que  $p0$  est un préfixe de  $\Delta L_C$  et  $1f$  est un facteur de  $\Delta L_C$ . Posons  $\Delta L_C = w_1 \dots w_n$ . Soit  $i - 3$  l'indice de la dernière lettre d'une occurrence du facteur  $1f$  dans le mot  $\Delta L_C$ , c'est-à-dire l'entier  $i$  satisfaisant

$1f = w_{i-|1f|-2} \dots w_{i-3}$ . Soit  $C'$  une sous-chenille de  $C$  tel que  $|C'|_{\emptyset} = L_C(i)$  et  $|C'| = i$ . Posons  $A = \text{Gauche}_{i-|1f|}(C')$  et  $B = \text{Droite}_{|1f|+3}(C')$ . Par le lemme 6.19, nous avons que  $C' = A \diamond B$ . Par conséquent,

$$\begin{aligned}
|B|_{\emptyset} &= |C'|_{\emptyset} - |A|_{\emptyset} + 2 \\
&= L_C(i) - |A|_{\emptyset} + 2 \\
&\geq L_C(i) - L_C(i - |1f|) + 2 \\
&= |w_{i-|1f|-2} \dots w_{i-3}|_1 + 2 \\
&= |1f|_1 + 2 \\
&= |f|_1 + 3 \\
&= |p|_1 + 3 \\
&> |p|_1 + 2 \\
&= |p0|_1 + 2.
\end{aligned}$$

En prenant,  $i = 3$  et  $j = |p0| + 3$ , on obtient que  $|B|_{\emptyset} > L_C(|p0| + 3)$ . C'est une contradiction, puisque  $|B| = |p0| + 3$  et  $B \leq C'$ . Donc,  $\Delta L_C$  est un mot préfixe normal. ■

Il existe plusieurs chenilles de lecture ayant la même suite feuillue. La forme préfixe normale d'un mot permet de déterminer les chenilles de lecture ayant la même fonction feuille.

**Théorème 6.23** (Blondin Massé *et al.* (2018a)). *Soit deux mots binaires  $w$  et  $w'$ . On a que*

$$L_{RC(w)} = L_{RC(w')} \text{ si et seulement si } w \equiv_{\text{FPN}} w'.$$

Par exemple, les mots 11010101, 11001101, 10110101, 10110011, 10101101 et 10101011 ont tous la même forme préfixe normale, alors leur chenille de lecture respective ont la même fonction feuille. Ces chenilles sont illustrées à la figure 6.10.

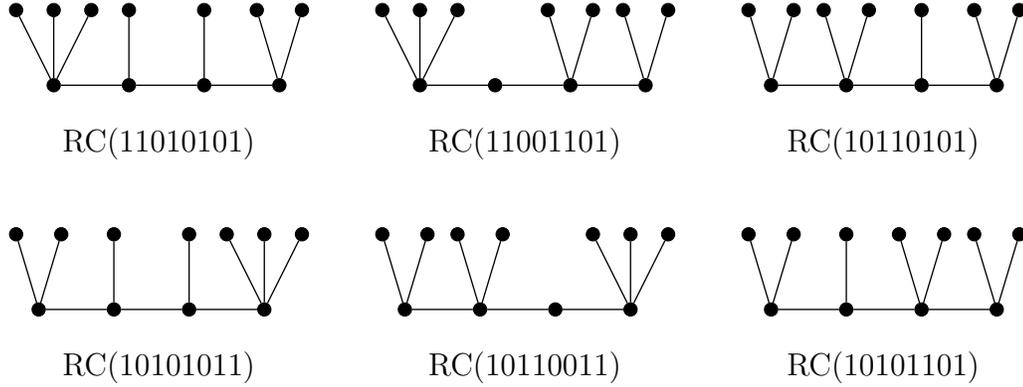


Figure 6.10: Toutes les chenilles d'ordre 11 dont le mot feuille est 11010101

*Démonstration du théorème 6.23.* Nous démontrons d'abord que  $L_{\text{RC}(w)} = L_{\text{RC}(w')}$  implique que  $F_1(w, i) = F_1(w', i)$  pour tout  $i \in \mathbb{N}$  par la contraposée. Supposons que  $F_1(w, i) > F_1(w', i)$  pour un certain  $i \in \mathbb{N}$ . Il existe un facteur  $u$  de  $w$  de longueur  $i$  tel que  $|u|_1 > |u'|_1$  pour tout facteur  $u'$  de  $w'$  de longueur  $i$ . Ainsi, par le lemme 6.16,  $\text{RC}(u)$  est une sous-chenille de  $\text{RC}(w)$  avec plus de feuilles que toutes les autres sous-chenilles de  $\text{RC}(w')$  de la forme  $\text{RC}(u')$  pour un certain facteur  $u'$  de  $w'$  de longueur  $i$ . On conclut que  $L_{\text{RC}(w)}(i) > L_{\text{RC}(w')}(i)$  par le corollaire 6.22.

Démontrons que  $F_1(w, i) = F_1(w', i)$  pour tout  $i \in \mathbb{N}$  implique  $L_{\text{RC}(w)} = L_{\text{RC}(w')}$  par la contraposée. Supposons que  $L_{\text{RC}(w)} \neq L_{\text{RC}(w')}$ . Posons  $n = |w|$  et  $w = w_1 w_2 \dots w_n$  avec  $w_j \in \{0, 1\}$ . Sans perte de généralité, il existe un entier  $i$  tel que  $L_{\text{RC}(w)}(i) > L_{\text{RC}(w')}(i)$ . Soit  $C_w$  une sous-chenille pleinement feuillue de  $\text{RC}(w)$  tel que  $|C_w| = i$  et  $|C_w|_{\emptyset} = L_{\text{RC}(w)}(i)$ . Par le lemme 6.21, nous pouvons supposer que  $C_w$  est tel qu'il existe un entier  $k$  tel que

$$\text{RC}(w) = \text{Gauche}_k(\text{RC}(w)) \diamond C_w \diamond \text{Droite}_{n+9-k-i}(\text{RC}(w)).$$

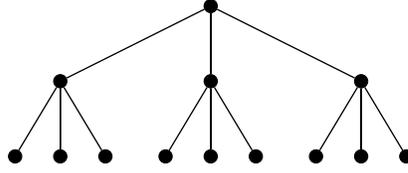


Figure 6.11: Un arbre dont le mot feuillu n'est pas préfixe normal

Le lemme 6.20, on obtient que

$$\begin{aligned}
 L_{\text{RC}(w)}(i) &= |C_w|_{\emptyset} \\
 &= |\text{RC}(w)|_{\emptyset} - |\text{Gauche}_k(\text{RC}(w))|_{\emptyset} - |\text{Droite}_{n+9-k-i}(\text{RC}(w))|_{\emptyset} + 4 \\
 &= |w|_1 + 2 - |\text{Pref}_{k-3}(w)|_1 - 2 - |\text{Suff}_{n+6-k-i}(w)|_1 - 2 + 4 \\
 &= |w_{k-2}w_{k-1} \dots w_{k+i-6}|_1 + 2.
 \end{aligned}$$

Ainsi, il y a un facteur  $u = w_{k-2}w_{k-1} \dots w_{k+i-6}$  de longueur  $i-3$  de  $w$  qui contient  $L_{\text{RC}(w)}(i) - 2$  occurrences de la lettre 1. Par conséquent,

$$|u|_1 = L_{\text{RC}(w)}(i) - 2 > L_{\text{RC}(w')}(i) - 2 \geq |u'|_1$$

pour tout facteur  $u'$  de  $w'$  tel que  $|u'| = i-3$ . Par conséquent,  $F_1(w, i-3) > F_1(w', i-3)$  et  $w \not\equiv_{\text{FPN}} w'$ . ■

Dans ce chapitre, nous avons démontré que les mots feuillus des graphes chenilles correspondent aux mots préfixes normaux. Appelons  $\mathcal{C}$  l'ensemble des mots feuillus des chenilles et  $\mathcal{T}$  l'ensemble des mots feuillus des arbres. Comme les graphes chenilles sont des arbres, on a que  $\mathcal{C} \subseteq \mathcal{T}$ . Par contre, les deux ensembles ne sont pas égaux. En effet, le mot feuillu du graphe de la figure 6.11 est 1101011011 et ce mot n'est pas préfixe normal, car

$$|\text{Pref}_5(11010110011)|_1 = |11010|_1 = 3 < 4 = |11011|_1 = |\text{Suff}_5(11010110011)|.$$

## CONCLUSION

Dans cette thèse, trois thématiques principales ont été abordées en utilisant la combinatoire. Plusieurs questions demeurent sans réponse. Nous résumons les résultats obtenus et présentons les principales pistes de recherche que ces problèmes nous suggèrent.

### **Mots parfaitement amassants**

Les mots parfaitement amassants sont définis par l'image de leur transformée de Burrows-Wheeler. Cette propriété est peu pratique pour énumérer ou construire des mots. Nous avons décrit deux nouvelles façons de construire les mots parfaitement amassants : d'abord en construisant les compositions qui représentent l'image commutative de ces mots à l'aide de l'arbre de composition circulaire ; ensuite, en engendrant ces mots à l'aide de quatre ensembles de morphismes. Ainsi, nous obtenons l'arbre des mots de Lyndon parfaitement amassants complets. L'étude des morphismes fournit une nouvelle approche pour étudier les propriétés combinatoires des mots parfaitement amassants. Voici quelques-unes des conjectures que nous avons sur les mots parfaitement amassants. Plusieurs d'entre elles sont inspirées des résultats connus sur les mots de Christoffel comme les mots parfaitement amassants coïncident avec les mots de Christoffel sur un alphabet binaire.

Les mots parfaitement amassants sont définis à l'aide de la transformée de Burrows-Wheeler, par conséquent la dernière colonne de leur matrice des conjugués a une forme particulière. Sur les mots de Christoffel, la proposition [2.19](#) décrit la relation entre deux lignes consécutives de la matrice des conjugués. Nous pensons qu'il est

$$\begin{array}{cccccccc}
a & c & b & b & c & a & c & b & c \\
a & c & b & c & a & c & b & b & c \\
b & b & c & a & c & b & c & a & c \\
b & c & a & c & b & b & c & a & c \\
b & c & a & c & b & c & a & c & b \\
c & a & c & b & b & c & a & c & b \\
c & a & c & b & c & a & c & b & b \\
c & b & b & c & a & c & b & c & a \\
c & b & c & a & c & b & b & c & a
\end{array}$$

Figure 6.12: La matrice des conjugués du mot  $acbbcacbc$ 

possible de décrire plus finement la relation entre deux lignes consécutives de la matrice des conjugués d'un mot parfaitement amassants.

**Conjecture 6.24.** *Soit  $w = w_1 \prec w_2 \prec \dots \prec w_n$  les conjugués d'un mot de Lyndon parfaitement amassant  $w$ . Alors, pour  $i \in \llbracket 2, n \rrbracket$ , il existe des mots  $u_i, s_i, p_i$  tels que*

$$w_{i-1} = s_i u_i p_i$$

$$w_i = s_i \tilde{u}_i p_i$$

et  $p_i s_i$  est un palindrome. De plus, le nombre de  $u_i$  distincts est  $|\text{Alph}(w)| - 1$ . Les palindromes  $p_i s_i$  distincts sont aussi de ce nombre, et pour un tel palindrome, appelons-le  $v$ , toutes les factorisations  $v = ps$  apparaissent dans la matrice des conjugués.

La matrice des conjugués du mot parfaitement amassant  $acbbcacbc$  est illustrée à la figure 6.12. Les  $u_i$  sont des éléments de  $\{bcac, acbcacbb\}$  et les  $p_i s_i$  sont des éléments de  $\{bcacb, c\}$ . Cette conjecture soulève plusieurs autres questions. Quelle est la longueur des mots  $u$ ? À quelle position commence le mot  $u$  dans chacun des conjugués? Pour les mots de Christoffel, on a que  $ps$  est le mot central, c'est-à-dire un palindrome itéré. Ce n'est pas le cas en général pour les mots parfaitement

amassants. En effet, l'un des palindromes pour le mot  $acbbcacbc$  est  $bcacb$  qui n'est pas un palindrome itéré. Quelles sont les propriétés de ces palindromes ?

Finalement, nous avons démontré que les mots parfaitement amassants sont des éléments positifs primitifs de  $F(\mathcal{A})$  (théorème 4.34). Les mots de Christoffel sont non seulement des éléments positifs du groupe libre  $F(\{a, b\})$ , mais ce sont les seuls (Osborne et Zieschang, 1981; Kassel et Reutenauer, 2007). Les mots parfaitement amassants sont-ils les seuls éléments positifs du groupe libre  $F(\mathcal{A})$  ? Peut-on factoriser les mots parfaitement amassants en un produit d'éléments positifs primitifs du groupe libre ? Une première étape est d'étudier les mots parfaitement amassants sur l'alphabet ternaire. Est-ce que les mots parfaitement amassants sur l'alphabet ternaire ont une factorisation  $w = xyz$  où  $x, y, z$  sont des mots parfaitement amassants et qu'ils forment une base de  $F(\{a, b, c\})$  ?

### Triplets de Markoff

Le second problème que nous avons abordé concerne les triplets de Markoff. Nous avons établi une bijection entre les mots binaires et les triplets de Markoff. Cette bijection implique des mots de Christoffel ayant une forme particulière. Nous avons explicité la factorisation standard et palindromique de ces mots, en plus de démontrer que ce sont des mots harmoniques. Ensuite, nous avons étudié deux généralisations de la conjecture de Frobenius sur les mots binaires concernant l'injectivité de la fonction  $\mu$ . Nous avons démontré que sur les facteurs d'un mot de Christoffel cette fonction est injective, ainsi que sur les mots de Christoffel apparaissant sur un même chemin dans l'arbre des mots de Christoffel ainsi que leurs conjugués la fonction est aussi injective. Les trois conjectures 5.4, 5.23 et 5.25 sont ouvertes. La résolution de la conjecture 5.23 ou conjecture 5.25 implique la résolution de la conjecture de Frobenius.

Un autre cas particulier est la conjecture de Frobenius ou de ses généralisations

est

**Conjecture 6.25.** *Soit  $u$  et  $v$  deux mots de Christoffel tels que  $|u| = |v|$ . Si  $u \prec v$ , alors  $\mu(u)_{12} < \mu(v)_{12}$ .*

La conjecture 6.25 est équivalente à la conjecture 3.11 dans Aigner (2013) utilisant les fractions réduites plutôt que les mots de Christoffel pour étiqueter les nombres de Markoff. Deux autres conjectures similaires y sont proposées. Tout d'abord, si  $u$  et  $v$  sont des mots de Christoffel tel que  $|u|_b = |v|_b$  et  $u \prec v$ , alors  $\mu(u)_{12} < \mu(v)_{12}$ . Cette conjecture a été démontrée récemment dans sa version originale (Rabideau et Schiffler, 2018). Ensuite, si  $u$  et  $v$  sont des mots de Christoffel tels que  $|u|_a = |v|_a$  et  $u \prec v$ , alors  $\mu(u)_{12} < \mu(v)_{12}$ .

### Graphes chenilles

Finalement, nous avons démontré que les mots feuillus des graphes chenilles sont en bijection avec les mots préfixes normaux. Cette bijection est un résultat partiel à des questions plus générales. Quels sont les mots feuillus des arbres et des graphes ? Puisque les graphes chenilles sont des arbres, on peut se demander si l'ensemble des mots feuillus des graphes chenilles et l'ensemble des mots feuillus des arbres coïncident. Or, ce n'est pas le cas. La figure 6.11 est le plus petit exemple d'arbre dont le mot feuillu n'est pas préfixe normal. En effet, le mot feuille de cet arbre est  $w = 1101011011$  qui n'est pas préfixe normal, car

$$|\text{Pref}_5(w)|_1 = |11010|_1 = 3 < 4 = |11011|_1 = |\text{Suff}_5(w)|_1.$$

Soit  $T$  ensemble des mots feuillus des arbres et  $R$  l'ensemble des mots feuillus des graphes. Le lemme 6.9 signifie que  $T \cap (R - T) = \emptyset$ . Ainsi, déterminer les mots appartenant à  $T$  et  $R - T$  est une question ouverte. Par conséquent, les questions suivantes sont toujours ouvertes : quels mots binaires correspondent au mot feuille

d'un arbre? Quels mots sur l'alphabet  $\{1, 0, -1, -2, \dots\} \cup \{\omega\}$  correspondent au mot feuille d'un graphe?

## RÉFÉRENCES

- Abram, A., Lapointe, M. et Reutenauer, C. (2020). Palindromization and construction of Markoff triples. *Theoret. Comput. Sci.*, 809, 21–29.
- Aigner, M. (2013). *Markov's theorem and 100 years of the uniqueness conjecture*. Springer, Cham. A mathematical journey from irrational numbers to perfect matchings.
- Bernoulli, J. (1771). Sur une nouvelle espèce de calcul. *Recueil pour les Astronomes*, 1, 255–284.
- Berstel, J. (2007). Sturmian and episturmian words (a survey of some recent results). In *Algebraic informatics*, volume 4728 de *Lecture Notes in Comput. Sci.* 23–47. Springer, Berlin.
- Berstel, J., Boasson, L., Carton, O. et Fagnot, I. (2007). A first investigation of Sturmian trees. In *STACS 2007*, volume 4393 de *Lecture Notes in Comput. Sci.* 73–84. Springer, Berlin.
- Berstel, J. et de Luca, A. (1997). Sturmian words, Lyndon words and trees. *Theoret. Comput. Sci.*, 178(1-2), 171–203.
- Berstel, J., Lauve, A., Reutenauer, C. et Saliola, F. V. (2009). *Combinatorics on words*, volume 27 de *CRM Monograph Series*. American Mathematical Society, Providence, RI. Christoffel words and repetitions in words.
- Berthé, V., de Luca, A. et Reutenauer, C. (2008). On an involution of Christoffel words and Sturmian morphisms. *European J. Combin.*, 29(2), 535–553.

- Blondin Massé, A., de Carufel, J., Goupil, A., Lapointe, M., Nadeau, E. et Vandomme, E. (2018a). Leaf realization problem, caterpillar graphs and prefix normal words. *Theoret. Comput. Sci.*, 732, 1–13.
- Blondin Massé, A., de Carufel, J., Goupil, A. et Samson, M. (2018b). Fully leafed tree-like polyominoes and polycubes. In *Combinatorial algorithms*, volume 10765 de *Lecture Notes in Comput. Sci.* 206–218. Springer, Cham.
- Bombieri, E. (2007). Continued fractions and the Markoff tree. *Expo. Math.*, 25(3), 187–213.
- Borel, J.-P. et Laubie, F. (1993). Quelques mots sur la droite projective réelle. *J. Théor. Nombres Bordeaux*, 5(1), 23–51.
- Borel, J.-P. et Reutenauer, C. (2005). Palindromic factors of billiard words. *Theoret. Comput. Sci.*, 340(2), 334–348.
- Borel, J.-P. et Reutenauer, C. (2006). On Christoffel classes. *Theor. Inform. Appl.*, 40(1), 15–27.
- Brlek, S., Lafrenière, N. et Provençal, X. (2015). Palindromic complexity of trees. In *Developments in language theory*, volume 9168 de *Lecture Notes in Comput. Sci.* 155–166. Springer, Cham.
- Brocot, A. (1861). Calcul des rouages par approximation, nouvelle méthode. *Revue Chronométrique*, 3, 186–194.
- Burrows, M. et Wheeler, D. J. (1994). A block sorting data compression algorithm. *Technical report, Digital System Research Center*, p. 18.
- Calkin, N. et Wilf, H. S. (2000). Recounting the rationals. *Amer. Math. Monthly*, 107(4), 360–363.

- Carpi, A. et de Luca, A. (2004). Harmonic and gold Sturmian words. *European J. Combin.*, 25(5), 685–705.
- Chen, K.-T., Fox, R. H. et Lyndon, R. C. (1958). Free differential calculus. IV. The quotient groups of the lower central series. *Ann. of Math. (2)*, 68, 81–95.
- Christoffel, E. B. (1875). Observatio arithmetica. *Ann. Mat. Pura Appl.*, 6(2), 148 – 152.
- Chuan, W.-F. (1997).  $\alpha$ -words and factors of characteristic sequences. *Discrete Math.*, 177(1-3), 33–50.
- Coven, E. M. et Hedlund, G. A. (1973). Sequences with minimal block growth. *Math. Systems Theory*, 7, 138–153.
- Crochemore, M., Iliopoulos, C. S., Kociumaka, T., Kubica, M., Radoszewski, J., Rytter, W., Tyczyński, W. et Waleń, T. (2012). The maximum number of squares in a tree. In *Combinatorial pattern matching*, volume 7354 de *Lecture Notes in Comput. Sci.* 27–40. Springer, Heidelberg.
- de Luca, A. (1997). Sturmian words : structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci.*, 183(1), 45–82.
- de Luca, A. et De Luca, A. (2006). Pseudopalindrome closure operators in free monoids. *Theoret. Comput. Sci.*, 362(1-3), 282–300.
- de Luca, A. et Mignosi, F. (1994). Some combinatorial properties of Sturmian words. *Theoret. Comput. Sci.*, 136(2), 361–385.
- Dulucq, S. et Gouyou-Beauchamps, D. (1990). Sur les facteurs des suites de Sturm. *Theoret. Comput. Sci.*, 71(3), 381–400.
- Erdős, P. et Gallai, T. (1961). Gráfok előírt fokú pontokkal (Graphs with prescribed degrees of vertices). *Mat. Lapok.*, 11, 264–274.

- Ferenczi, S. et Zamboni, L. Q. (2013). Clustering words and interval exchanges. *J. Integer Seq.*, 16(2), Article 13.2.1, 9.
- Fici, G. et Lipták, Z. (2011). On prefix normal words. In *Developments in language theory*, volume 6795 de *Lecture Notes in Comput. Sci.* 228–238. Springer, Heidelberg.
- Glen, A. et Justin, J. (2009). Episturmian words : a survey. *Theor. Inform. Appl.*, 43(3), 403–442.
- Harary, F. et Schwenk, A. J. (1973). The number of caterpillars. *Discrete Math.*, 6, 359–365.
- Justin, J. (2005). Episturmian morphisms and a Galois theorem on continued fractions. *Theor. Inform. Appl.*, 39(1), 207–215.
- Karnauhova, A. et Liebscher, S. (2017). Connected components of meanders : I. Bi-rainbow meanders. *Discrete Contin. Dyn. Syst.*, 37(9), 4835–4856.
- Kassel, C. et Reutenauer, C. (2007). Sturmian morphisms, the braid group  $B_4$ , Christoffel words and bases of  $F_2$ . *Ann. Mat. Pura Appl. (4)*, 186(2), 317–339.
- Labelle, J. (1981). *Théorie des graphes*. MODULO, Mont-Royal.
- Lapointe, M. (2019). Number of orbits of discrete interval exchanges. *Discrete Math. Theor. Comput. Sci.*, 21(3), Paper No. 17, 16.
- Lothaire, M. (1997). *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge.
- Lothaire, M. (2002). *Algebraic combinatorics on words*, volume 90 de *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

- Mantaci, S., Restivo, A. et Sciortino, M. (2003). Burrows-Wheeler transform and Sturmian words. *Inform. Process. Lett.*, 86(5), 241–246.
- Markoff, A. (1879). Sur les formes quadratiques binaires indéfinies. *Math. Ann.*, 15(3), 381–496.
- Markoff, A. (1880). Sur les formes quadratiques binaires indéfinies (second mémoire). *Math. Ann.*, 17(3), 379–399.
- Melançon, G. (1999). Lyndon words and singular factors of Sturmian words. volume 218 41–59. WORDS (Rouen, 1997).
- OEIS, F. I. (2020). The on-line encyclopedia of integer sequences
- Osborne, R. P. et Zieschang, H. (1981). Primitives in the free group on two generators. *Invent. Math.*, 63(1), 17–24.
- Oseledec, V. I. (1966). The spectrum of ergodic automorphisms. *Dokl. Akad. Nauk SSSR*, 168, 1009–1011.
- Pak, I. et Redlich, A. (2008). Long cycles in  $abc$ -permutations. *Funct. Anal. Other Math.*, 2(1), 87–92.
- Rabideau, M. et Schiffler, R. (2018). Continued fractions and orderings on the markov numbers.
- Raney, G. N. (1973). On continued fractions and finite automata. *Math. Ann.*, 206, 265–283.
- Rauzy, G. (1979). Échanges d’intervalles et transformations induites. *Acta Arith.*, 34(4), 315–328.
- Restivo, A. et Rosone, G. (2009a). Balanced words having simple Burrows-Wheeler transform. In *Developments in language theory*, volume 5583 de *Lecture Notes in Comput. Sci.* 431–442. Springer, Berlin.

- Restivo, A. et Rosone, G. (2009b). Burrows-Wheeler transform and palindromic richness. *Theoret. Comput. Sci.*, 410(30-32), 3018–3026.
- Restivo, A. et Rosone, G. (2011). Balancing and clustering of words in the Burrows-Wheeler transform. *Theoret. Comput. Sci.*, 412(27), 3019–3032.
- Reutenauer, C. (2009). Christoffel words and Markoff triples. *Integers*, 9, A26, 327–332.
- Reutenauer, C. (2015). Studies on finite Sturmian words. *Theoret. Comput. Sci.*, 591, 106–133.
- Reutenauer, C. (2018). Combinatorics on words and the theory of Markoff. In *Computation and combinatorics in dynamics, stochastics and control*, volume 13 de *Abel Symp.* 691–707. Springer, Cham.
- Reutenauer, C. (2019). *From Christoffel words to Markoff numbers*. Oxford University Press, Oxford.
- Reutenauer, C. et Vuillon, L. (2017). Palindromic closures and Thue-Morse substitution for Markoff numbers. *Unif. Distrib. Theory*, 12(2), 25–35.
- Richomme, G. (2003). Lyndon morphisms. *Bull. Belg. Math. Soc. Simon Stevin*, 10(suppl.), 761–785.
- Simpson, J. et Puglisi, S. J. (2008). Words with simple Burrows-Wheeler transforms. *Electron. J. Combin.*, 15(1), Research Paper 83, 17.
- Stern, M. A. (1858). Über eine zahlentheoretische funktion. *j. reine angew. Math.*, 55, 193–220.
- Zhang, Y. (2006). An elementary proof of uniqueness of markoff numbers which are prime powers.

## INDEX

- abélien équivalent, 9
- alphabet, 7
- antimorphisme, 11
- antipalindrome, 12
- arbre, 126
  - de Raney, 24
  - de Stern-Brocot, 23
  - des compositions circulaires, 52
  - des mots des Christoffel, 22
  - des mots parfaitement amassants, 71
  - des paires de Christoffel, 21
  - des triplets, 99
  - des triplets de Markoff, 96
  - dual mots de Christoffel, 24
  - enraciné, 126
  - monoïde libre, 98
- arête, 125
- automorphisme, 10
  - groupe libre, 90
- chaîne, 126
  - longueur, 126
  - simple, 126
- chenille, 126
  - colonne, 126
  - de lecture, 136
  - orientée, 134
  - pleinement feuillue, 137
- classe de conjugaison, 8
- clôture
  - antipalindromique, 13
  - palindromique, 12
- complément, 12
- composition, 14
  - circulaire, 40
  - faible, 14
- concaténation, 7
- conjecture de Frobenius, 97
- conjugué, 8, 92
- croissante, 11
- cycle
  - d'un graphe, 126
  - d'une permutation, 15
- disjoints, 15

- décroissante, 11
- dérivée discrète, 128
- endomorphisme, 10
- facteur, 8
- factorisation
  - palindromique, 20
  - standard, 19
- fermé, 126
- fermé par, 11
- feuille, 126
- fonction feuille, 126
- forme normale, 113
- forme préfixe normale, 122
- graphe
  - biparti complet, 130
  - complet, 129
  - connexe, 126
  - cycle, 129
  - grille infini, 131
  - roues, 130
  - simple, 125
- greffe, 134
- groupe libre, 66
  - base, 90
- image commutative, 8
- image miroir, 10
- inversion, 66
- involution, 11
- lettre, 7
- libre, 10
- longueur, 7
- matrice
  - des conjugués, 13
- morphisme
  - monoïde, 10
  - positif, 70
  - strictement alphabétique, 10
- mot, 7
  - de Christoffel propre, 18
  - de Christoffel spéciaux, 105
  - central, 19
  - circulaire, 8
  - complet, 7
  - de Christoffel, 18
  - de Christoffel dual, 25, 108
  - de Christoffel inférieur, 17
  - de Christoffel supérieur, 17
  - de Lyndon, 9
  - feuillu, 128
  - parfaitement amassant, 30
  - primitif, 8
  - préfixe normal, 122
  - sturmien, 29

- équilibré, 9
- nombre de Markoff, 97
- notation cyclique, 15
- orbite, 15
- ordre
  - d'un graphe, 125
  - lexicographique, 9
  - militaire, 113
- paire de Christoffel, 19
- palindrome, 10
- palindromisation itérée, 12
- permutation, 14
  - circulaire, 15
  - standard, 38
  - symétrique, 15
- préfixe, 8
- puissance, 8
- racine, 126
- sommet
  - adjacents, 125
  - degré, 125
  - incident, 125
  - pendant, 125
- sous-arbre, 126
- pleinement feuillu, 126
- sous-chenille, 136
  - droite, 139
  - gauche, 139
- sous-graphe induit, 125
- suffixe, 8
- trajectoire, 35
- transformée de Burrows-Wheeler, 13
- triplet
  - de Christoffel, 19
  - de Markoff, 95
  - de Markoff propre, 95
- type cyclique, 62
- vecteur de translation, 36
- échange d'intervalles discrets, 34
  - minimal, 35
  - symétrique, 35, 36
- élément, 67
  - circulairement réduit, 67
  - inverse, 67
  - longueur, 67
  - positif, 70
  - primitif, 90
  - réduit, 67

## INDEX DES NOTATIONS

$(w)$ , 8 $G = (V, E)$ , 125 $G_Y$ , 125 $O_x$ , 15 $S_i$ , 72 $\mathcal{A}$ , 7 $\mathcal{A}^*$ , 7 $\mathcal{M}$ , 27 $\mathcal{D}$ , 26 $\mathcal{G}$ , 26 $\mathcal{P}$ , 30 TRIPLET, 99 Alph, 7 Antipal( $w$ ), 13 $\beta$ , 101 BW, 13 $\cdot^\oplus$ , 13 RC, 135 $w^{(+)}$ , 12 $\mathcal{Z}$ , 41 $\Delta_\ell$ , 71 deg( $x$ ), 125	$w^*$ , 25 $w^\dagger$ , 108 $E_c$ , 36 $E_{\sigma,c}$ , 34 $\mathcal{CO}$ , 134 $[[a, b]]$ , 14 $\equiv_{\text{FPN}}$ , 122 $ T _{\emptyset}$ , 126 $L_G$ , 126 $\diamond$ , 134 $\tau$ , 67 $x^{-1}$ , 67 $\kappa_a$ , 12 $\leq$ , 136 $\preceq$ , 42 $\leq_{\text{mil}}$ , 113 $\prec$ , 9 $\mu$ , 95 $M(w)$ , 13 $\mathcal{P}_2(V)$ , 125 $\mathcal{T}$ , 22 Fact( $w$ ), 8
--	--

$\text{Pref}(w)$ , 8	$\text{Pal}$ , 12
$\text{Suff}(w)$ , 8	$\xi$ , 100
$F_x$ , 122	$\mathcal{N}_x$ , 122
$\rho_\ell$ , 67	$\psi_i$ , 46
$\Delta L_G$ , 129	$ G $ , 125
$\mu(w)_{12}$ , 95	$\mathcal{T}_G^i$ , 126
$\lambda_\ell$ , 67	$\text{Droite}_i(C)$ , 139
$\tilde{\cdot}$ , 10	$\text{Gauche}_i(C)$ , 139
$\text{End}(M)$ , 10	$\theta$ , 10
$\text{FPN}_x$ , 122	$t_i$ , 36
$\nu$ , 102	$\eta$ , 101
$\gamma(\sigma)$ , 15	$\hat{\cdot}$ , 12
$\bar{\cdot}$ , 12	$\zeta$ , 28