UNIVERSITÉ de Picardie Jules Verne

LAMFA

Thèse de Doctorat

Spécialité Sciences Mathématiques

présentée à

L'Université de Picardie Jules Verne

par

Julien LEROY

pour obtenir le grade de Docteur de l'Université de Picardie Jules Verne

Contribution à la résolution de la conjecture S-adique

Soutenue le 18 janvier 2012, après avis des rapporteurs, devant le jury d'examen :

M. S. Ferenczi, Directeur de Recherches	**Rapporteur**
M. L. Q. Zamboni, Professeur	**Rapporteur**
M. J. Cassaigne, Chargé de Recherches	**Examinateur**
M. F. Durand, Professeur	**Examinateur**
M. A. Fan, Professeur	**Examinateur**
M. B. Host, Professeur	**Examinateur**
M. G. Richomme, Professeur	**Examinateur**
M. M. Rigo, Professeur	**Examinateur**

Thèse de Doctorat

Spécialité Sciences Mathématiques

présentée à

L'Université de Picardie Jules Verne

par

Julien LEROY

pour obtenir le grade de Docteur de l'Université de Picardie Jules Verne

Contribution to the resolution of the S-adic conjecture

Soutenue le 18 janvier 2012, après avis des rapporteurs, devant le jury d'examen :

M. S. Ferenczi, Directeur de Recherches	**Rapporteur**
M. L. Q. Zamboni, Professeur	**Rapporteur**
M. J. Cassaigne, Chargé de Recherches	**Examinateur**
M. F. Durand, Professeur	**Examinateur**
M. A. Fan, Professeur	**Examinateur**
M. B. Host, Professeur	**Examinateur**
M. G. Richomme, Professeur	**Examinateur**
M. M. Rigo, Professeur	**Examinateur**

Remerciements

Mes premiers remerciements vont tout naturellement à Fabien Durand et à Gwénaël Richomme qui ont encadré mon travail durant ces trois années. Par leur incroyable disponibilité, de près comme de loin, ils ont su me guider dans mes recherches tout en me laissant beaucoup de liberté quant à la direction qu'elles prenaient. Les nombreuses discussions, mathématiques ou non, que nous avons partagées ont été un plaisir sans cesse renouvelé et ont largement contribué à l'aboutissement de ce travail. Je les en remercie vivement.

Je suis extrêmement reconnaissant envers Michel Rigo non seulement pour avoir accepté de faire partie de mon jury, mais surtout pour m'avoir toujours soutenu dans mes projets de doctorat et grâce à qui cette aventure amiénoise a pu commencer.

J'adresse mes remerciements à Julien Cassaigne, Ai-Hua Fan et Bernard Host pour avoir accepté de faire partie du jury de ma thèse ainsi qu'à Sébastien Ferenczi et Luca Zamboni pour me faire l'honneur d'en être les rapporteurs.

Merci aussi à Cassy Gentile, et à Anne et Dominique Lacroix pour les heures de relecture qu'elles ont passées sur ce travail.

Durant ma thèse, j'ai rencontré énormément de gens intéressants avec qui j'ai pu partagé une discussion, un tableau noir ou tout simplement un verre. Je leur adresse à tous toute ma reconnaissance. En particulier, je souhaite remercier Narad Rampersad et Samuel Petite pour toutes les discussions qu'ils m'ont accordées ainsi qu'Alexandre Blondin Massé et Sébastien Labbé pour m'avoir régulièrement aidé dans l'utilisation du logiciel mathématique SAGE.

Je tiens ensuite à exprimer toute ma sympathie aux membres du LAMFA pour leur disponibilité et leur hospitalité ainsi qu'à mes collègues doctorants ou post-doctorant de l'Université de Liège et de l'Université de Picardie Jules Verne.

Ces remerciements seraient incomplets si je n'y mentionnais pas le soutien et les encouragements de ma famille. En particulier, merci à Anne pour le stress qu'elle accumule à ma place.

Résumé

Cette thèse concerne la Conjecture S-adique qui stipule l'existence d'une version forte de S-adicité dans les suites qui serait équivalente à une complexité p (en facteurs) sous-linéaire. Une suite $\mathbf{w}$ à valeurs dans un alphabet fini A est dite S-adique si S est un ensemble de morphismes permettant de désubstituer indéfiniment $\mathbf{w}$. Sans condition supplémentaire, la complexité en facteurs d'une suite S-adique peut être arbitrairement grande. Cependant, de nombreuses familles de suites bien connues admettent des développements S-adiques avec $\mathrm{Card}(S) < +\infty$ et sont également de complexité sous-linéaire. La conjecture S-adique apparaît alors naturellement comme une tentative de relier ces deux notions.

Dans cette thèse, nous étudions en détails une méthode constructive basée sur les graphes de Rauzy et qui produit un développement S-adique des suites uniformément récurrentes de complexité sous-linéaire. Par ce biais, nous exhibons certaines propriétés nécessaires (mais pas suffisantes) du développement obtenu. Dans le cas particulier des suites uniformément récurrentes dont la différence première de complexité est majorée par deux, cette méthode est poussée à l'extrême, si bien que les conditions nécessaires obtenues en deviennent suffisantes.

Mots-clés : S-adique, complexité, système dynamique symbolique, sous-shift, graphe de Rauzy

Abstract

This thesis is about the S-adic conjecture which suppose the existence of a stronger notion of S-adicity that would be equivalent to having a sub-linear factor complexity. A sequence $\mathbf{w}$ over a finite alphabet A is said to be S-adic if S is a set of morphisms that allows to indefinitely de-substitute $\mathbf{w}$. Without additional condition, the factor complexity of an S-adic sequence might be arbitrarily large. However, many well-known families of sequences have a sub-linear complexity and admit some S-adic expansions with $\text{Card}(S) < +\infty$. The S-adic conjecture is therefore a natural attempt to link these two notions.

In this thesis, we study in detail a method based on Rauzy graphs that provides an S-adic expansion of uniformly recurrent sequences with sub-linear complexity. By this way we are able to determine some necessary (but not sufficient) conditions of these expansions. In the particular case of uniformly recurrent sequences with first difference of complexity bounded by two, the method is studied with even much more details, which makes the necessary conditions sufficient.

Keywords: S-adic, factor complexity, symbolic dynamical system, subshift, Rauzy graph

Contents

Introduction (version française)

Un outil classique dans l'étude des suites (ou mots infinis) à valeurs dans un ensemble fini A (généralement appelé *alphabet*) est la *fonction de complexité* p qui compte le nombre $p(n)$ de blocs (généralement appelés *facteurs*) de longueur n qui apparaissent dans la suite. Cette fonction permet de mesurer le désordre de la suite. Par exemple, elle permet de caractériser l'ensemble des suites ultimement périodiques, celles-ci étant exactement celles pour lesquelles $p(n) \leq n$ pour une longueur n (voir [MH40]). Par extension, cette fonction peut évidemment se définir pour n'importe quel ensemble de mots (généralement appelé *langage*) ou n'importe quel système dynamique symbolique (ou encore *sous-shift*). Pour des survols sur la complexité, voir [All94, Fer99] ou le Chapitre 4 de [BR10].

La fonction de complexité permet également de définir la classe des suites *sturmiennes* comme étant l'ensemble des suites apériodiques de complexité minimale $p(n) = n+1$ pour toute longueur n ; il s'agit donc de suites binaires ($p(1) = 2$). Celles-ci apparaissent dans divers domaines des mathématiques et une grande littérature leur est consacrée (voir le Chapitre 1 de [Lot02] et le Chapitre 6 de [Fog02] pour des survols). Elles possèdent notamment plusieurs définitions équivalentes : elles sont par exemple les suites obtenues par un codage naturel de rotations d'angle irrationnel ou encore les suites apériodiques équilibrées. Par ailleurs, il est bien connu que les sous-shifts qu'elles engendrent peuvent être obtenus par itérations successives des deux morphismes (ou *substitutions*) R_0 et R_1 définis, si l'alphabet A est $\{0, 1\}$, par $R_0(0) = 0$, $R_0(1) = 10$, $R_1(0) = 01$ et $R_1(1) = 1$ (voir [MH40]). Pour obtenir non pas les sous-shifts, mais les suites elles-mêmes, il est nécessaire de considérer les deux morphismes supplémentaires L_0 et L_1 définis par $L_0(0) = 0$, $L_0(1) = 01$, $L_1(0) = 10$ et $L_1(1) = 1$ (voir [MS93, BHZ06]). De manière générale, une suite (resp. un sous-shift) obtenue par un tel procédé, c'est-à-dire par itérations successives de morphismes appartenant à un ensemble S, est appelée suite (resp. sous-shift) *S-adique*, en rapport avec la terminologie des systèmes adiques introduite par Vershik (voir par exemple [VL92]).

L'utilisation de morphismes dans l'étude des suites ou, plus généralement, en combinatoire des mots n'est pas nouvelle. Au début du 20$^{\text{ème}}$ siècle, A. Thue utilisait déjà ce procédé, principalement afin d'étudier les répétitions dans les mots (voir [Thu06, Thu12]). Par ailleurs, le cas où S contient un unique morphisme (auquel cas on parle de suite *purement substitutive* ou de suite *purement morphique*) a été largement étudié, aussi bien par rapport aux propriétés combinatoires des suites ainsi obtenues (voir entre autres [Cas97, Cas03, CN03, Dev08, Dur98a, Dur98b, Dur02, ELR75, ER81, ER83, Fer95, NP09, Sie05, Pan84, RW02, Hon10]) que par rapport aux propriétés ergodiques et topologiques des sous-shifts engendrés (voir entre autres [DL06, Dur00, DHS99, Hos86, HP89, Hos00, Que87]. Par exemple, J.-J. Pansiot [Pan84] a complétement caractérisé les comportements asymptotiques de la complexité de ces suites. J. Cassaigne a également développé des techniques plus fines basées sur certains facteurs (appelés facteurs spéciaux) et permettant de calculer leur complexité exacte (voir [Cas97, Klo11]).

À propos de la conjecture S-adique

Il existe bien d'autres catégories de suites qui sont classiquement étudiées. Parmi celles-ci, on trouve des généralisations des suites sturmiennes, telles que les codages de rotations (voir par exemple [Ada02, Ada05, AS07, AB98, Did98a, Did98b, Rot94]), les codages d'échanges d'intervalles (voir par exemple [Daj02, Did97, FHZ01, FHZ03, FHZ04, FZ08, FZ10, GMP03, KBC10, LN98, LN00, LN01, Rau79, Vui07]), les suites d'Arnoux-Rauzy (voir par exemple [AR91, CFZ00, CC06, CFM08, Che09, MZ02]) ou encore les suites épisturmiennes (voir par exemple [Ber07, BdLDLZ08, GJ09, GLR09, JP02, JV00, PV07, Ric03, Ric07]). On peut également rencontrer des suites *automatiques* (voir entre autres [AS03, ARS09, Mos96, NR07, RM02, Sha88, Tap94, Tap96]), liées à la théorie des automates et aux morphismes ou encore des codages de rotations sur d'autres groupes compacts que $\mathbb{R}/\mathbb{Z}$ (voir par exemple [AB92, CK97, JK69, KP11, Kos98, RA96, Wil84]) ou encore des suites de Kolakoski (voir [Dek97]). Un point intéressant est qu'une grande partie de ces suites ont une complexité sous-linéaire, i.e., il existe une constante D telle que pour tout $n \geq 1$, $p(n) \leq Dn$. De plus, pour ces dernières, on peut trouver un ensemble (généralement fini) S de morphismes tel que la suite est S-adique (voir le Chapitre 2 pour plus de détails). Il est alors naturel de se demander s'il existe un lien entre le fait d'être S-adique et le fait d'avoir une complexité sous-linéaire. Ces deux notions ne peuvent clairement pas être équivalentes puisque, grâce au travail de Pansiot, on sait qu'il existe des suites purement substitutives de complexité quadratique. On peut cependant imaginer une notion plus forte de S-adicité qui serait équivalente à la complexité sous-linéaire. En d'autres termes, il faut trouver une condition

C telle qu'*une suite est de complexité sous-linéaire si et seulement si elle est S-adique satisfaisant la condition C*. Il s'agit là de la *conjecture S-adique*, conjecture due à B. Host. À l'heure actuelle, nous ignorons totalement la nature de la condition C. Il peut s'agir d'une condition sur l'ensemble S des morphismes, ou d'une condition sur la manière dont celles-ci doivent se succéder dans la représentation S-adique. Dans cette thèse, nous donnerons des exemples étayant l'idée que la réponse est très certainement une combinaison des deux (voir aussi [DLR]), confirmant ainsi la difficulté intrinsèque de la conjecture.

Le but de cette thèse est précisément l'étude de cette conjecture. La conjecture S-adique est étayée par l'existence de représentations S-adiques de certaines suites bien connues (notamment pour les suites sturmiennes, codages de rotations, codages d'échanges d'intervalles, etc.). Cependant, ces dernières dépendent fortement de la nature des suites initiales et il est donc difficile d'extraire des propriétés générales à partir de celles-ci. De plus, la caractérisation des suites de complexité sous-linéaire qui sont purement substitutives (obtenue par Pansiot) ne se généralise qu'en une condition suffisante pour les suites S-adiques (voir [Dur00, Dur03]) et bon nombre de conditions qu'on voudrait naturelles ne sont mêmes pas des conditions suffisantes à garantir une complexité sous-linéaire (voir la Section 2.3.2 pour plus de détails). Néanmoins, il existe un résultat dû à S. Ferenczi fournissant une méthode générale qui, étant donnée une suite uniformément récurrente de complexité sous-linéaire, permet de construire successivement les morphismes apparaissant dans la représentation S-adique (voir [Fer96]). Hormis le fait que le nombre de morphismes ainsi créés est fini, nous ne savons presque rien de ceux-ci. L'objectif premier de cette thèse était, dans le but de mieux cerner la condition C, l'étude de ces morphismes et un de nos résultats est la détermination de certaines de leurs propriétés (voir Chapitre 3).

L'algorithme produisant les morphismes est basé sur une utilisation massive des *graphes de Rauzy*. Ceux-ci sont des outils puissants pour étudier la combinatoire des suites ou des sous-shifts. Par exemple, ils sont à la base d'un puissant résultat de Cassaigne prouvant qu'une suite est de complexité sous-linéaire si et seulement si la différence première de sa complexité est borné (voir [Cas96]). Ils ont également permis à T. Monteil d'améliorer un résultat de M. Boshernitzan (voir [Bos85]) en donnant une meilleur borne sur le nombre de mesures ergodiques invariantes du système (voir le Chapitre 5 de [Mon05] ou le chapitre 7 de [BR10]). Cependant, ces graphes sont souvent difficiles à décrire dès que la complexité dépasse un niveau vraiment bas. Pour cette raison, l'extraction de propriétés générales se révèle être un problème des plus complexes. En appliquant ces mêmes méthodes pour les sous-shifts dont la différence première de complexité $p(n+1)-p(n)$ est inférieure à 2 pour tout n, Ferenczi a tout de même prouvé que le nombre de morphismes ainsi créés était inférieur à 3^{27}.

Grâce à une étude détaillée des graphes de Rauzy possibles pour ces complexités, nous améliorons cette borne et montrons l'existence d'un ensemble S de 5 morphismes tels que tout sous-shift minimal dont la différence première de complexité est majoré par 2 est S-adique (voir Chapitre 4). Plus précisément, nous donnons une condition nécessaire et suffisante sur les compositions d'éléments de S pour obtenir un tel sous-shift (voir Chapitre 5). Cette caractérisation contient celle des sous-shifts minimaux de complexité $2n$, dont certains avaient été étudié par G. Rote [Rot94].

Au delà de la conjecture

Un des grands intérêts des représentations S-adiques est qu'elles fournissent une interprétation arithmétique des suites étudiées et, dans de nombreux cas, un développement généralisé en fractions continues. Par exemple, la suite de morphismes qui apparaît dans le cas des suites sturmiennes dépend du développement en fractions continues classique de l'angle de la rotation correspondante. Ainsi, cela permet par exemple de caractériser les suites sturmiennes *primitives morphiques* (voir Definition 1.3.2) comme étant exactement celles codant des rotations d'angles quadratiques α de points de $\mathbb{Q}(\alpha)$ (voir [Par99]). Ce développement en fractions continues permet également de calculer, par exemple, la fréquences des facteurs de la suite (voir [AB98]) ou encore l'exposant critique de celle-ci, i.e., la plus grande puissance fractionnaire qui peut apparaître dans la suite (voir [Van00]).

Dans le cas d'une rotation d'angle α dont le codage est réalisé par rapport à la partition $[0, 1-\beta)$, $[1-\beta, 1)$ de $[0, 1)$, la représentation S-adique dépend également d'un développement généralisé en fractions continues de (α, β) (voir [Did98a]). Par ailleurs, comme expliqué dans [BCF99] (voir également [Ada02]), ces suites sont intimement liées aux codages d'échanges de trois intervalles. Ces suites dépendent de deux paramètres α et β (la longueur de deux des intervalles) qui peuvent être approximés simultanément via un algorithme basé sur l'induction de Rauzy (voir [Rau79, Rau77]). Dans [FHZ01, FHZ03, FHZ04], les auteurs développent également un autre algorithme permettant, par exemple, de donner une caractérisation combinatoire des suites de complexité $2n+1$ qui sont des codages naturels d'échanges de trois intervalles. Comme dans le cas des codages de rotations, cet algorithme est ultimement périodique si et seulement si les deux paramètres de l'échange appartiennent au même corps quadratique.

Les suites dites d'*Arnoux-Rauzy* sont un autre exemple de suites de complexité $2n+1$. Celles-ci jouissent de propriétés combinatoires supplémentaires généralisant celles des suites sturmiennes. La plus célèbre suite d'Arnoux-Rauzy est sans aucun doute la *suite de Tribonacci*, point fixe du morphisme τ défini par $\tau(0) = 01$, $\tau(1) = 02$ et $\tau(2) = 0$. Cette suite est liéée à une rotation sur le tore $\mathbb{T}^2$ et il a été conjecturé qu'il en était de même pour toute suite d'Arnoux-Rauzy. Dans [CFZ00], les auteurs donnent un contre-exemple à cette conjecture. Dans [CFM08], les auteurs exhibent une classe de suites d'Arnoux-Rauzy dont les sous-shifts associés sont (en mesure) faiblement mélangeants, ceux-ci ne pouvant alors pas être conjugués à des rotations. Par contre, il est prouvé dans [AR91] que toutes les suites d'Arnoux-Rauzy peuvent être interprétées comme des codages d'échanges de 6 intervalles. Par une étude de leurs graphes de Rauzy, les auteurs ont également obtenu un développement S-adique de celles-ci (voir aussi [RZ00]). Celui-ci permet par exemple, comme pour les suites sturmiennes, de calculer les fréquences des facteurs (voir [WZ01]) de la suite ainsi que la fonction de récurrence quotient (voir [CC06]).

Une autre classe de suites S-adiques est la classe des suites linéairement récurrentes dont font partie les suites primitives substitutives (voir [Dur98a, DHS99]). Ces suites sont de complexité sous-linéaire et F. Durand a montré dans [Dur00, Dur03] que ces suites correspondent exactement aux suites S-adiques primitives et propres (voir les Definitions 1.3.10 et 1.3.11). En particulier, une suite sturmienne est linéairement récurrente si et seulement si les coefficients de son développement en fraction continue sont bornés.

S-adicité à la Bratteli-Vershik

Dans [Bra72], O. Bratteli a introduit des graphes infinis découpés en niveaux (désormais appelés *diagrammes de Bratteli*) permettant l'approximation de C^*-algèbres. Dans une optique dynamique (*transformation adique*), A. Vershik eut l'idée dans [Ver82] d'associer à un tel diagramme un ordre lexicographique sur les chemins infinis dans ces diagrammes. Cet ordre est induit par un ordre partiel sur les arcs entre deux niveaux consécutifs, ce dernier pouvant alors être décrit par une matrice d'adjacence entre les deux niveaux considérés, i.e., par un morphisme. Pour plus de détails, voir le Chapitre 6 de [BR10] et voir [War02] pour le lien entre les diagrammes de Bratteli et les systèmes S-adiques.

Par un raffinement des constructions de Vershik, les auteurs de [HPS92] ont démontré que tout système de Cantor minimal est topologiquement isomorphe à un système de Bratteli-Vershik (résultat déjà obtenu en mesure par Vershik dans [Ver82]). Ces représentations à la Bratteli-Vershik sont intéressantes en dynamique, surtout dans les problèmes liés à la récurrence. Mais, étant donné un système minimal de Cantor, il est en général difficile d'en trouver une représentation de Bratteli-Vershik "canonique" (pour des exemples, voir [DHS99]). Cependant, Ferenczi a montré que pour les sous-shifts de complexité sous-linéaire, le nombre de morphismes lus dans le diagramme de Bratteli correspondant est fini (dans un contexte mesuré à la Vershik). En particulier, il a obtenu une majoration explicite du rang et démontré l'absence de mélange fort. Par ailleurs, Durand a montré que, dans le cas des sous-shifts linéairement récurrents, la suite de morphismes apparaissant dans la représentation S-adique est exactement la suite de morphismes lus sur le diagramme de Bratteli. De plus, contrairement au cas sous-linéaire de Ferenczi, la conjugaison entre le sous-shift et le système de Bratteli-Vershik se fait de manière topologique.

Organisation de la thèse

Après l'établissement des notations et le rappel des définitions au Chapitre 1, le Chapitre 2 a pour but de faire un tour d'horizon de la S-adicité en général. Ainsi, nous rappelons et comparons les résultats connus pour les suites purement substitutives, les suites substitutives (c'est-à-dire images par un morphisme d'une suite purement substitutive) et les suites S-adiques. Nous y présentons également quelques représentations S-adiques bien connues, ainsi que des résultats connus fournissant des conditions suffisantes à une complexité sous-linéaire. Nous considérons enfin une liste d'exemple permettant d'emblée de rejeter certaines idées "naïves" à propos de la conjecture.

Le Chapitre 3 attaque la conjecture dans le cas général. Ainsi, nous y étudions les morphismes construits sur base des graphes de Rauzy, ce qui nous permet par exemple de donner une caractérisation S-adique des sous-shifts minimaux. Par ailleurs, nous y explicitons quelques conditions nécessaires sur ces morphismes et démontrons à travers des exemples que celles-ci ne pas suffisantes. La majorité de ce chapitre se trouve également dans [Ler12].

Dans le Chapitre 4, nous étudions en détails les graphes de Rauzy et leurs évolutions correspondant à une différence première de complexité majorée par 2. Ceci nous permet de calculer explicitement tous les morphismes ainsi obtenues et nous montrons en fait que tous peuvent se décomposer en des produits de morphismes et que seuls 5 morphismes de base sont nécessaires à ces décompositions. Au Chapitre 5, nous améliorons le résultat obtenu au Chapitre 4 par une étude encore plus poussée des évolutions de graphes. Ceci nous permet d'obtenir la caractérisation annoncée. Les développements S-adiques obtenus permettent également de décrire explicitement leurs représentations de Bratteli-Vershik. Cependant, le nombre de notions à introduire pour présenter ce résultat paraît trop important par rapport à la portée du résultat. Plus de détails pourront être trouvés dans [DL].

Introduction (english version)

A classical tool in the study of sequences (or infinite words) with values in a finite set A (generally called *alphabet*) is the *complexity function* p which counts the number $p(n)$ of blocks (generally called *factors*) of length n that appear in the sequence. Thus this function allows to measure the regularity in the sequence. For example, it allows to describe all ultimately periodic sequences as exactly being those for which $p(n) \leq n$ for a length n (see [MH40]). By extension, this function can obviously be defined for any set of words (generally called *language*) or any symbolic dynamical system (or *subshift*). For surveys over the complexity function, see [All94, Fer99] or Chapter 4 of [BR10].

The complexity function can also be used to define the class of *Sturmian sequences*: it is the family of aperiodic sequences with minimal complexity $p(n) = n + 1$ for all lengths n. Those sequences are therefore defined over a binary alphabet (because $p(1) = 2$) and a large literature is devoted to them (see Chapter 1 of [Lot02] and Chapter 6 of [Fog02] for surveys). In particular, these sequences admit several equivalent definitions such as natural codings of rotations with irrational angle or aperiodic balanced sequences. Moreover, it is well known that the subshift they generate can be obtained by successive iterations of two morphisms (or substitutions) R_0 and R_1 defined (when the alphabet A is $\{0, 1\}$) by $R_0(0) = 0$, $R_0(1) = 10$, $R_1(0) = 01$ and $R_1(1) = 1$ (see [MH40]). To generate not all Sturmian subshifts but all sturmian sequences it is necessary to consider two additional morphisms L_0 and L_1 defined by $L_0(0) = 0$, $L_0(1) = 01$, $L_1(0) = 10$ and $L_1(1) = 1$ (see [MS93, BHZ06]). In general, a sequence (or subshift) obtained by such a method, that is, obtained by successive iterations of morphisms belonging to a set S, is called an S-*adic* sequence (or subshift), accordingly to the terminology of adic systems introduced by Vershik (see for instance [VL92]).

Using morphisms in the study of sequences, or more generally in combinatorics of words, is far from being new. At the beginning of 20^{th} century, A. Thue already used them, mainly in order to study repetitions in words (see [Thu06, Thu12]). Moreover, the case where S contains a unique morphism (in which case we talk about *purely morphic* or *purely substitutive* sequence) has been extensively studied both with respect to combinatorial properties of these sequences (see for instance [Cas97, Cas03, CN03, Dev08, Dur98a, Dur98b, Dur02, ELR75, ER81, ER83, Fer95, NP09, Sie05, Pan84, RW02, Hon10]) and with respect to ergodic and topological properties of the generated subshifts (see for instance [DL06, Dur00, DHS99, Hos86, HP89, Hos00, Que87]). For example, J.-J. Pansiot [Pan84] completely characterized all possible asymptotic behaviours of the complexity of these sequences. J. Cassaigne also developed thinner techniques (based on some particular factors called *special factors*) to compute their exact complexity (see [Cas97, Klo11]).

About the S-adic conjecture

There are many other families of sequences which are usually studied in the literature. Among them one can find generalizations of Sturmian sequences, such as codings of rotations (see for instance [Ada02, Ada05, AS07, AB98, Did98a, Did98b, Rot94]), codings of intervals exchanges (see for instance [Daj02, Did97, FHZ01, FHZ03, FHZ04, FZ08, FZ10, GMP03, KBC10, LN98, LN00, LN01, Rau79, Vui07]), Arnoux-Rauzy sequences (see for instance [AR91, CFZ00, CC06, CFM08, Che09, MZ02]) or episturmian sequences (see for instance [Ber07, BdLDLZ08, GJ09, GLR09, JP02, JV00, PV07, Ric03, Ric07]). One can also talk about *automatic sequences* (see [AS03, ARS09, Mos96, NR07, RM02, Sha88, Tap94, Tap96]) linked to automata theory and morphisms or about codings of rotations over other compact groups than $\mathbb{R}/\mathbb{Z}$ (see for instance [AB92, CK97, JK69, KP11, Kos98, RA96, Wil84]) or also about Kolakoski sequences (see [Dek97]). An interesting point is that much of these sequences have a sub-linear complexity, i.e., there exist a constant D such that for all positive integers n, $p(n) \leq Dn$. In addition, we can usually associate a (generally finite) set S of morphisms to these sequences in such a way that they are S-adic (see Chapter 2 for more details). It is then natural to ask whether there is a connection between the fact of being S-adic and the fact of having a sub-linear complexity. Both notions clearly cannot be equivalents as, thanks to Pansiot's work, there exist purely morphic sequences with a quadratic complexity. However, we can imagine a stronger notion of S-adicity that would be equivalent to having a sub-linear complexity. In other words, we would like to find a condition C such that *a sequence has a sub-linear complexity if and only if it is S-adic satisfying the condition C*. This problem is called the *S-adic conjecture* and is due to B . Host. Up to now, we have no idea about the nature of the condition C. It may be a condition on the set S of morphisms, or a condition on the way in which they must occur in the sequence of morphisms.

In this thesis, we give examples supporting the idea that the answer should be a combination of both (see also [DLR]), supporting the difficulty of the conjecture.

The purpose of this thesis is precisely to study this conjecture. This conjecture is supported by the existence of S-adic representations of many well-known sequences (such as Sturmian sequences, codings of rotations, codings of intervals exchanges, etc.). However, these representations strongly depend on the nature of the sequences which makes general properties difficult to extract. In addition, the characterization of purely morphic sequences with sub-linear complexity (obtained by Pansiot) can only be generalized into a sufficient condition for S-adic sequences (see [Dur00, Dur03]) and many (*a priori* natural) conditions over S-adic sequences are even not sufficient to guarantee a sub-linear complexity (see Section 2.3.2 for more details). Nevertheless, S. Ferenczi provided a general method that, given any uniformly recurrent sequence with sub-linear complexity, produces an S-adic representation (see [Fer96]). Except that the number of morphisms occurring in that S-adic representation is finite, we know almost nothing about them. The primary purpose of this thesis was, in order to better understand the condition C, the study of these morphisms and one of our results is the determination of some of their properties (see Chapter 3).

The algorithm that produces the morphisms is based on an extensive use of *Rauzy graphs*. These graphs are powerful tools to study combinatorial properties of sequences or subshifts. For example, they are the basis of a strong Cassaigne's result proving that a sequence has a sub-linear complexity if and only if the first difference of its complexity $p(n+1)-p(n)$ is bounded (see [Cas96]). They also allowed T. Monteil to improve a result due to M. Boshernitzan (see [Bos85]) by giving a better bound on the number of ergodic invariant measures of a subshift (see Chapter 7 of [BR10] or Chapter 5 of [Mon05]). However, these graphs are usually difficult to compute as soon as the complexity exceeds a very low level. For this reason, the extraction of properties of the S-adic representation from these graphs is usually hard. Anyway, applying these methods to subshifts for which the difference of complexity $p(n+1)-p(n)$ is no more than to 2 for every n, Ferenczi succeeded to prove that the number of morphisms built in such a way is less than 3^{27}.

By analysing all possible Rauzy graphs, we managed to strongly improve this bound and show the existence of a set $\mathcal{S}$ of 5 morphisms such that any minimal subshift with first difference of complexity bounded by 2 is $\mathcal{S}$-adic (see Chapter 4). More precisely, we give a necessary and sufficient condition on sequences in $\mathcal{S}^{\mathbb{N}}$ to be an $\mathcal{S}$-adic representation of such a subshift (see Chapter 5). This characterization contains the subshifts with complexity $2n$, some of which were studied by G. Rote [Rot94].

Beyond conjecture

An interesting point of S-adic representations is that they provide an arithmetical interpretation of the sequences and, in many cases, a generalized continued fraction development. For example, the sequence of morphisms that occur in the case of Sturmian sequences is governed by the classical continued fraction expansion of the angle of the corresponding rotation. Thus, this allows for example to characterize the Sturmian sequences that are *primitive* morphic (see Definition 1.3.2 page 31) as being exactly those coding rotations of quadratic angle α of points in $\mathbb{Q}(\alpha)$ (see [Par99]). This continued fraction expansion can also be used, for example, to compute frequencies of factors (see [AB98]) or also the critical exponent of the sequence, i.e., the largest fractional power that occurs in the sequence (see [Van00]).

In the case of a rotation of angle α whose coding is realized with respect to the partition $[0, 1-\beta)$, $[1-\beta, 1)$ of $[0, 1)$, the S-adic representation also depends on a generalized continued fraction development of (α, β) (see [Did98a]). Moreover, as explained in [BCF99] (see also [Ada02]), these sequences are intimately linked to codings of three intervals exchanges. These sequences depend of two parameters α and β (the length of the two intervals) that can be simultaneously approximated via an algorithm based on Rauzy induction (see [Rau79, Rau77]). In [FHZ01, FHZ03, FHZ04], the authors have developed another algorithm that allows for instance to give a combinatorial characterization of sequences with complexity $2n+1$ that are natural codings of three intervals exchanges. As with codings of rotations, this algorithm is ultimately periodic if and only if both parameters of the intervals exchange belong to the same quadratic field.

Another example of sequences with complexity $2n+1$ are the so-called *Arnoux-Rauzy sequences*. They satisfy additional combinatorial properties generalizing those of Sturmian sequences. There is no doubt that the most famous Arnoux-Rauzy sequence is the *Tribonacci sequence*, fixed point of the morphism τ defined by $\tau(0) = 01$, $\tau(1) = 02$ and $\tau(2) = 0$. This sequence is linked to a rotation on the torus $\mathbb{T}^2$ and it has been conjectured that it was the case of all Arnoux-Rauzy sequences. In [CFZ00], the authors provide a counter-example to that conjecture. In [CFM08], the authors exhibit a class of Arnoux-Rauzy sequences whose associated subshifts are weakly mixing and so that cannot be conjugated to rotations. By contrary, it is proved in [AR91] that all Arnoux-Rauzy sequences (over a three letters-alphabet) can be interpreted as codings of six intervals exchanges. By studying their Rauzy graphs, the authors also provided an S-adic representation of them (see also [RZ00]). This allows for instance, as for Sturmian sequences, to compute the frequencies of factors (see [WZ01]) of the sequence and the recurrence quotient of it (see [CC06]).

Another class of S-adic sequences is the class of linearly recurrent sequences that includes primitive substitutive sequences (see [Dur98a, DHS99]). These sequences have a sub-linear complexity and Durand proved in [Dur00, Dur03] that they exactly correspond to primitive and proper S-adic sequences with $\mathrm{Card}(S) < +\infty$ (see Definitions 1.3.10 and 1.3.11). In particular, a Sturmian sequence is linearly recurrent if and only if the coefficients of its continued fraction expansion are bounded.

Bratteli-Vershik S-adicity

In [Bra72] Bratteli introduced infinite graphs (subsequently called *Bratteli diagrams*) partitioned in levels in order to approximate C^*-algebras. With other motivations, Vershik thought in [Ver82] to associate dynamics (*adic transformations*) to these diagrams by introducing a lexicographic ordering on the infinite paths of the diagrams. This ordering is induced by a partial order on the arcs between two consecutive levels, it can then be defined by an adjacent matrix between the two considered levels and thus by a morphism. For more details see Chapter 6 of [BR10] and see [War02] for the link between Bratteli diagrams and S-adic systems.

By a refinement of Vershik's constructions, the authors of [HPS92] have proved that any minimal Cantor system is topologically isomorphic to a Bratteli-Vershik system (Vershik already obtained this result in [Ver82] in a measure theoretical context). These Bratteli-Vershik representations are helpful in dynamics, mainly with problems about recurrence. But, being given a minimal Cantor system, it is generally difficult to find a "canonical" Bratteli-Vershik representation (see [DHS99] for examples). However, Ferenczi proved that for minimal subshift with sub-linear complexity, the number of morphisms read on the associated Bratteli diagram (in a measure theoretical context) is finite [Fer96]. In particular, he obtained an upper bound on the rank of these systems and proved that they cannot be strongly mixing. In addition, Durand showed that, in the case of linearly recurrent subshifts, the morphisms appearing in the S-adic representation are exactly those read on the Bratteli diagram. In addition, unlike in Ferenczi's result, the subshift is topologically conjugated to the Bratteli-Vershik system.

Organization of the thesis

Chapter 1 contains all needed definitions and backgrounds. Chapter 2 is designed to make an overview of S-adicity. Thus, we recall and compare the well-known results about purely morphic sequences, morphic sequences (that are images by a morphism of a purely morphic sequence) and S-adic sequences. We also give some well-known S-adic representations of some families of sequences (such as the Sturmian ones) and recall some known results providing sufficient conditions to a sub-linear complexity. We finally give examples of S-adic sequences that allow us to reject some "naive" ideas about the S-adic conjecture.

Chapter 3 attack the conjecture in the general case. We study the morphisms constructed on the basis of Rauzy graphs which, in particular, allows us to give an S-adic characterization of minimal subshifts. In addition, we give some necessary conditions on these morphisms and prove through examples that they do not suffice. The majority of this chapter can also be found in [Ler12].

In Chapter 4, we start a detailed description of Rauzy graphs corresponding to minimal subshifts with first difference of complexity bounded by 2. This allows us to explicitly compute all needed morphisms and we show that they all can be decomposed into compositions of only five morphisms. In Chapter 5, we improve the result obtained in Chapter 4 by studying even more the sequences of possible evolutions of Rauzy graphs. This allows us to obtain an S-adic characterization, hence the condition C of the conjecture for this particular case. The obtained S-adic representations can also be used to explicitly give the Bratteli-Vershik representations of the systems. However, the amount of notions that would be needed to that aim seems to be too big compared to the importance of the result. More details will be found in [DL].

Résumé en français

Chapitre 1 : préliminaires

Nous supposons le lecteur familier avec les notions de base de combinatoire des mots et de systèmes dynamiques symboliques. Rappelons simplement que si S est un ensemble de morphismes, une suite $\mathbf{w}$ est S-adique s'il existe une suite de morphismes $(\sigma_n : A_{n+1}^* \to A_n^*)$ telle que $\min_{a \in A_{n+1}} |\sigma_0 \cdots \sigma_n(a_{n+1})|$ converge vers l'infini lorsque n augmente et telle que

$$\mathbf{w} = \lim_{n \to +\infty} \sigma_0 \cdots \sigma_n(a_{n+1}^{\omega}).$$

Chapitre 2 : tour d'horizon

Le but de ce chapitre est de recenser les résultats connus sur la S-adicité.

Suites (purement) morphiques et suites S-adiques

Dans un premier temps, nous comparons les suites purement morphiques, les suites morphiques et les suites S-adiques, les deux premières familles étant des cas particuliers de la troisième puisque les suites purement morphiques sont des suites S-adiques avec $\mathrm{Card}(S) = 1$ et les suites morphiques sont les images morphiques de suites purement morphiques.

Depuis le travail de Pansiot [Pan84], il est bien connu que la complexité des suites purement morphiques est extrêmement contrainte. En effet, pour des morphismes non-effaçant, celle-ci ne peut prendre que cinq comportements asymptotiques, ceux-ci étant $\Theta(1)$, $\Theta(n)$, $\Theta(n \log \log n)$, $\Theta(n \log n)$ et $\Theta(n^2)$ (voir Théorème 2.1.2 page 42) et ces différents comportements dépendent uniquement de la longueur des images des itérés du morphisme. Par ailleurs, il existe certains critères combinatoires qui contraignent encore plus la complexité de ces suites. Par exemple, il est également bien connu que si une suite purement morphique est uniformément récurrente, alors sa complexité est forcément sous-linéaire (voir Proposition 2.1.4 page 43).

En ce qui concerne la complexité des suites morphiques, il n'existe à l'heure actuelle aucune caractérisation similaire à celle obtenue par Pansiot. Il est cependant clair que les cinq comportements asymptotiques ne suffisent plus puisqu'il existe des suites de complexité $p(n) \in \Theta(n \sqrt[k]{n})$ pour tout entier $k \geq 1$. Le nombre de comportements asymptotiques possibles devient donc infini dénombrable. R. Deviatov a récemment conjecturé que ces comportements asymptotiques supplémentaires étaient les seuls possibles (voir Théorème 2.1.10 page 45).

Pour les suites S-adiques en général, le problème devient bien plus complexe. En effet, Cassaigne a démontré que toute suite pouvait être obtenue de manière S-adique (voir Proposition 2.1.15 page 46). Par conséquent, un autre résultat de Cassaigne (voir Proposition 2.1.17 page 47) implique que le nombre de comportements asymptotiques pour la complexité des suites S-adiques est indénombrable.

Résultats partiels pour la sous-linéarité de la complexité

En observant les représentations S-adiques des familles bien connues de suites de complexité sous-linéaire (suites sturmiennes, codages de rotations, codages d'échanges d'intervalles, etc.), on s'aperçoit que la longueur de toutes les images $\sigma_0\sigma_1 \cdots \sigma_n(a)$ croissent indéfiniment lorsque n tend vers l'infini. Un morphisme jouissant de cette propriété est appelée *morphisme partout croissant* (voir Definition 1.3.5 page 31). Par extension, nous dirons qu'une suite de morphismes $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ est *croissante partout*[1] si la longueur minimale de $\sigma_0\sigma_1 \cdots \sigma_n(a)$ pour a dans A_{n+1} croît à l'infini lorsque n tend vers l'infini.

Comme mentionné plus haut, la croissance des images joue un rôle capital dans la complexité des suites purement morphiques. En effet, pour un morphisme croissant partout, la complexité de ses points fixes est de complexité sous-linéaire si et seulement si toutes les images $\sigma^n(a)$ ont le même ordre de croissance.

Dans [Fer96], Ferenczi a démontré que pour toute suite $\mathbf{w}$ uniformément récurrente de complexité sous-linéaire, il existe une représentation S-adique croissante partout de $\mathbf{w}$ (voir Théorème 2.3.5 page 63). Cependant, la condition sur l'ordre de croissance des images obtenue par Pansiot dans le cas des suites purement morphiques ne se généralise qu'en une condition suffisante (voir Proposition 2.3.1 page 62) puisqu'il existe même des suites sturmiennes ne satisfaisant pas cette condition (celles pour lesquelles la suite $(a_k)_{k\geq 1}$ du

[1]Dans l'article [Ler12], cette propriété porte le nom de *ω-growth property* car *croissant partout* a déjà une autre signification dans le cadre des DT0L (voir [ELR76]) qui, ici, correspond à la notion d'*expansivité* (voir Définition 1.3.12 page 33).

Théorème 2.2.1 page 50 est non-bornée). Cette condition suffisante permet néanmoins de déterminer certaines familles d'ensembles S pour lesquels toute suite S-adique est de complexité sous-linéaire (voir Corollaire 2.3.2 et Corollaire 2.3.4 page 63).

Fausses bonnes idées sur la conjecture

Une des grandes difficultés de la conjecture est que bon nombre des conditions naturelles qu'on peut imaginer en connaissant les résultats sur les complexités des suites purement morphiques ne sont même pas suffisantes à assurer une complexité sous-linéaire. Par exemple, Boshernitzan a "légitimement" conjecturé que si un ensemble S_1 ne contient que des morphismes dont les points fixes sont de complexité sous-linéaire, alors toute suite S_1-adique est de complexité sous-linéaire. Il a par la suite prouvé qu'il n'en était rien (voir Proposition 2.3.9 page 65). Une idée similaire est de penser que si S_2 contient un morphisme σ ayant des points fixes de complexité élevée, alors les suites S_2-adiques correspondantes devraient également avoir une complexité élevée. Dans cette thèse, nous donnons des exemples contredisant cette idée et cela même lorsque σ apparaît très souvent dans la suite de morphismes (voir Proposition 2.3.12 page 67 et Proposition 2.3.14 page 69).

Chapitre 3 : progrès réalisés dans le cas général de la conjecture

Une approche pour résoudre la conjecture est de renforcer les conditions nécessaires jusqu'à les rendre suffisantes[2]. Sous la condition supplémentaire d'uniforme récurrence de la suite, nous y parvenons grâce à une étude poussée des graphes de Rauzy de la suite.

Méthode de dé-substitution

La méthode de construction de la représentation S-adique d'une suite $\mathbf{w}$ est basée sur les graphes de Rauzy. Dans le graphe de Rauzy $G_n(\mathbf{w})$, les sommets sont les facteurs de longueurs n de la suite et il existe une flèche de u vers v étiquetée par une lettre a s'il existe une lettre b telle que $ub = av$ est un facteur de $\mathbf{w}$. Les flèches dans $G_n(\mathbf{w})$ correspondent donc exactement aux sommets dans $G_{n+1}(\mathbf{w})$. Par ailleurs, les facteurs spéciaux (gauches ou droits)

[2]l'autre approche étant d'affaiblir les conditions suffisantes connues pour les rendre nécessaires

de $\mathbf{w}$, i.e. ceux qui peuvent se prolonger de plusieurs façons (à gauche ou à droite) dans $\mathbf{w}$ se repèrent directement dans les graphes de Rauzy puisqu'il s'agit des sommets ayant plusieurs flèches entrantes ou sortantes. Un résultat célèbre de Cassaigne implique directement que le nombre de tels sommets dans $G_n(\mathbf{w})$ est borné si et seulement si $\mathbf{w}$ est de complexité sous-linéaire.

En étudiant la suite $(G_n(\mathbf{w}))_{n\in\mathbb{N}}$, on peut remarquer que certains chemins particuliers dans $G_m(\mathbf{w})$ sont en fait des concaténations de ces mêmes chemins particuliers dans $G_n(\mathbf{w})$ pour $n < m$. Par exemple, dans le cas des suites sturmiennes, si $G_n(\mathbf{w})$ contient un sommet bispécial, i.e., un sommet avec deux flèches entrantes et deux flèches sortantes, alors $G_n(\mathbf{w})$ a nécessairement la forme représentée à la Figure A.1 page 183 et si m est le plus petit entier $m > n$ tel que $G_m(\mathbf{w})$ a la même forme que $G_n(\mathbf{w})$, alors une des deux boucles de $G_m(\mathbf{w})$ a la même étiquette qu'une des boucles de $G_n(\mathbf{w})$ et l'étiquette de l'autre est la concaténation des étiquettes des deux boucles de $G_n(\mathbf{w})$.

Dans le cas général, ces chemins particuliers peuvent être définis de différentes façons. Il peut soit s'agir de l'ensemble des chemins simples entre les facteurs spéciaux gauches (ou droits), soit de l'ensemble des chemins d'un sommet spécial gauche jusque lui-même dont l'étiquette est effectivement un facteur de la suite. Dans $G_n(\mathbf{w})$ les chemins de la première possibilité sont appelés les *n-segments* (voir Définition 3.1.2 page 78) et ceux de la deuxième sont appelés *n-circuits* (voir Définition 3.1.11 page 81). Dans un cas comme dans l'autre, les étiquettes de ces chemins sont toujours des facteurs de la suite et les longueurs du plus grand n-segment et du plus petit n-circuit tendent vers l'infini lorsque n augmente (voir Remarque 3.1.10 page 80 et Corollaire 3.1.17 page 83).

Pour chacune de ces deux familles de chemins, nous pouvons définir des morphismes. En effet, nous montrons que pour tout n, tout $(n+1)$-segment se décompose dans $G_n(\mathbf{w})$ en une concaténation bornée de n-segments (voir Lemme 3.2.6 page 86). Nous pouvons donc définir un morphisme σ_n défini sur l'ensemble des $(n+1)$-segments et ayant pour images des concaténations de n-segments (voir Définition 3.2.1 page 83). Les ensembles étant bornés et les images étant de longueurs bornées également, le nombre de morphismes ainsi créés est donc fini. Pour les n-circuits, les morphismes se définissent de la même façon (voir Lemme 3.1.13 page 82 et Définition 3.3.1 page 87), à la différence qu'ils dépendent du choix d'une suite de facteurs spéciaux gauches de longueurs croissantes et emboîtés en préfixes (voir Lemme 3.1.14 page 82). De plus, dans ce cas, le nombre de morphismes peut être infini. En effet, l'uniforme récurrence assure la finitude des alphabets et des longueurs des images, mais pas que ceux-ci soient bornés. Pour les deux familles de chemins, la construction définit une suite de morphismes $(\sigma_n : A_{n+1}^* \to A_n^*)$ dont les images $\sigma_0 \cdots \sigma_n(a)$ sont les étiquettes des $(n+1)$-segments (ou circuits) et produit donc une représentation S-adique du langage de la suite. Afin d'obtenir une représentation de la suite elle-même, il suffit de considérer que pour tout n, le préfixe de longueur n fait partie des extrémités des chemins (voir Section 3.2).

Un résultat intéressant sur la décomposition S-adique obtenue sur base des n-circuits est qu'elle fournit une caractérisation des suites uniformément récurrentes : il s'agit exactement des suites S-adiques *primitives* et *propres à gauche* (voir Définition 1.3.10 et Définition 1.3.11 page 33), S pouvant être de cardinal infini. Rappelons que dans le cas où S est fini, ces conditions sont équivalentes à la récurrence linéaire de la suite.

La conjecture étant définie pour des ensembles S de cardinal fini, il est naturel de se concentrer sur la décomposition S-adique obtenue sur base des n-segments. Afin d'en extraire des conditions nécessaires (notamment la *croissance presque partout* déjà mentionnée plus haut), une idée développée par Ferenczi dans [Fer96] est de considérer les morphismes σ_n créés, non pas sur l'ensemble des $(n+1)$-segments, mais sur un ensemble particulier de concaténations de $(n+1)$-segments défini comme suit. Même si la longueur maximale des n-segments tend vers l'infini lorsque n croît, il peut exister des n-segments qui sont courts et ce, même pour n très grand. Ceci nous pousse à partitionner l'ensemble des n-segments en ceux dits *courts* dont la longueur est bornée par une constante indépendante de n et ceux dits *longs* (voir Définition 3.4.2 page 92). Une conséquence directe de la définition est que le plus petit long n-segment a une longueur qui tend vers l'infini lorsque n augmente. Pour obtenir la croissance presque partout, il suffit donc de considérer qu'un long n-segment apparaît dans chacune des concaténations particulières de n-segments choisies. Plus précisément, sous la condition supplémentaire que l'étiquette correspondante soit un facteur de la suite, les concaténations que nous considérons sont de la forme suivante : une concaténation de n-segments cours suivie d'un long n-segment suivi d'une concaténation de n-segments courts. L'ensemble de ces concaténations étant de cardinal fini et même borné indépendamment de n (voir Lemme 3.4.3 page 93), il forme le nouvel alphabet B_n sur lequel nous définissons le morphisme τ_{n-1}, ce dernier exprimant l'action du morphisme σ_{n-1} (voir Définition 3.4.4 page 93).

Du fait de sa dépendance en l'existence des segments courts, la suite de morphismes $(\tau_n)_{n\in\mathbb{N}}$ ne peut être définie qu'à partir d'un certain rang N. Pour en déduire une représentation S-adique, il est alors nécessaire de considéré un morphisme supplémentaire κ associant à une lettre de B_N l'étiquette du chemin correspondant dans $G_N(\mathbf{w})$ (voir Proposition 3.4.7 page 94). En ce qui concerne les conditions nécessaires obtenues sur la suite $(\tau_n)_{n\geq N}$, elles découlent principalement de l'observation suivante : *si le mot de longueur 2 ab apparaît dans une image* $\tau_n(c)$*, alors, dans* $G_{n+1}(\mathbf{w})$*, le sous-chemin* *qui correspond au mot ab du chemin correspondant à la lettre c ne contient aucun facteur spécial gauche* (voir Lemme 3.4.8 page 95). Ce résultat nous permet de prouver qu'une même lettre de peut apparaître deux fois dans une même image (voir Proposition 3.4.9 page 96) et, plus généralement, qu'il ne peut pas exister de "cycles" dans l'ensemble des images, i.e., un ensemble de mots de la forme $\{a_1u_1a_2, a_2u_2a_3, \ldots, a_ku_ka_1\}$ ne peut pas être un ensemble de facteurs des mots dans $\tau_n(B_{n+1})$ (voir Proposition 3.4.11 page 96). L'observation mentionnée plus haut permet également de prouver que dans toute image $\tau_n(b)$, une même lettre a est toujours précédée par les suffixes d'un même mot, sauf éventuellement la première lettre de l'image (voir Proposition 3.4.10 page 96). Enfin, de ces trois propriétés découle une quatrième donnant une décomposition en morphismes "simples" de tout morphisme τ_n (voir Proposition 3.4.15 page 98) et sous la condition supplémentaire de non-existence de segments courts, nous montrons également que la suite $(\tau_n)_{n\geq N}$ est presque primitive (voir Proposition 3.4.12 page 97).

Malheureusement, toutes ces conditions nécessaires ne sont pas suffisantes à garantir une complexité sous-linéaire. En effet, Section 3.5, nous exhibons un exemple de suite S-adique satisfaisant toutes celles-ci, mais dont la complexité n'est pas sous-linéaire.

Chapitres 4 et 5 : résolution de la conjecture pour les complexités inférieures à $2n + 1$

L'étude des graphes de Rauzy réalisée au chapitre 3 ne s'étant pas révélée suffisamment fructueuse pour résoudre la conjecture, nous nous attaquons dans les chapitres 4 et 5 au cas particulier des suites uniformément récurrentes et dont la différence première de complexité $p(n + 1) - p(n)$ est majorée par 2. Notons tout de même que l'ensemble de ces suites contient une grande partie des suites étudiées dans la littérature. Dans ce cas, nous connaissons déjà toutes les formes de graphes de Rauzy qui peuvent apparaître (voir [Rot94]), ce qui en rend l'étude plus simple, bien que très technique. Pour ces complexités, nous parvenons à déterminer des conditions nécessaires fortes sur les représentations S-adiques et en fait suffisamment fortes pour qu'elles soient suffisantes, résolvant ainsi la conjecture dans ce cas particulier.

Une première remarque est que la caractérisation S-adique obtenue ne l'est que pour les sous-shifts et non pour les suites. Au Chapitre 3, nous avons obtenu les représentations S-adiques des suites grâce à une petite astuce qui consiste simplement à ajouter une flèche entrante au sommet de $G_n(\mathbf{w})$ qui correspond au préfixe de longueur n de $\mathbf{w}$. Cette astuce pourrait sans doute être reproduite dans le cas particulier qui nous intéresse, mais au prix de développements encore bien plus techniques et les difficultés supplémentaires que cela engendrerait pèseraient bien lourd par rapport à l'intérêt des nouveaux résultats.

Une deuxième remarque est que, contrairement aux représentations S-adiques obtenues au Chapitre 3, celles-ci sont basées sur les n-circuits. Cela nous fournit notamment un moyen de distinguer les représentations S-adiques *valides* au moyen de la presque primitivité de la suite de morphismes. De plus, comme extrémités des circuits, nous considérons une suite de spéciaux droits emboîtés en suffixes plutôt qu'une suites de facteurs spéciaux gauches emboîtés en préfixes. La raison en est que le travail très technique concernant ces complexités a été réalisé avant de découvrir l'intérêt de travailler avec les facteurs spéciaux gauches[3]. Par conséquent, les étiquettes des flèches dans les graphes de Rauzy étudiés dans ce chapitre sont, contrairement à ce qui a été fait jusqu'à présent dans ce document, les prolongements droits des sommets dans la suite. Les résultats "gauches" du Chapitre 3 ont donc besoin d'un équivalent "droit" pour ce chapitre ; ceux-ci sont listés dans la section 4.1.

Première étape : déterminer l'ensemble $\mathcal{S}$

Comme mentionné plus haut, les formes (ou *types*) de graphes de Rauzy qui peuvent apparaître pour les complexités qui nous intéressent sont connues et au nombre de 10 (si on suppose que tous les graphes contiennent un sommet bispécial). Ces types de graphes sont représentés à la Figure 4.5 page 109. Connaissant cela, il est alors possible d'étudier quel type de graphe peut évoluer vers quel type de graphe et de calculer explicitement les morphismes correspondant à ces évolutions. Pour cela, nous avons évidemment besoin de connaître les alphabets sur lesquels nous travaillons. Dans le chapitre 4, nous montrons que considérer des alphabets à trois lettres est toujours suffisant (voir Lemme 4.3.4 page 114) et choisissons une correspondance entre ces trois lettres et les circuits dans les graphes (voir page 117). Les évolutions de graphes sont alors représentées dans l'annexe A et les morphismes correspondants à ces évolutions se trouvent dans la section 4.5.

Il apparaît rapidement que le nombre de morphismes codant les évolutions de graphes est infini (à cause de leur dépendance en des puissances k et ℓ). Cependant, nous prouvons dans la section 4.6 que tous les morphismes obtenus peuvent en fait être vus comme des compositions de cinq morphismes particuliers, notés D, G, M, E_{01} et E_{12} et définis page 103. L'ensemble $\mathcal{S}$ de

[3] Celui-ci étant justement l'obtention des représentations S-adiques pour les suites et non pour les sous-shifts

ces cinq morphismes permet donc de donner une représentation $\mathcal{S}$-adique de n'importe quel sous-shift minimal dont la différence première de complexité est majorée par 2. Pour ce qui est des conditions, deux conditions nécessaires restent évidemment la presque primitivité et la propreté de la suite de morphismes de $\mathcal{S}$. De plus, nous pouvons ajouter une condition supplémentaire sur la manière dont les morphismes doivent être composés. En effet, les morphismes codent des évolutions de graphes de différents types et il est donc évident que si un morphisme code une évolution vers un graphe de type 3, le morphisme suivant ne peut coder une évolution d'un graphe d'un type autre que 3. Cette condition peut être exprimée par l'obligation d'étiqueter un chemin infini dans le graphe (appelé *graphe des graphes*) représenté à la figure 4.8 page 112. Les sommets de celui-ci correspondent aux différents type de graphes et les flèches représentent les évolutions possibles. Le résultat exprimant ces conditions est le théorème 4.0.1.

Deuxième étape : déterminer toutes les suites de $\mathcal{S}^{\mathbb{N}}$ qui sont des représentations $\mathcal{S}$-adiques valides

Au début du chapitre 5, nous définissons la notion de *chemin étiqueté valide* (voir Définition 5.1.1 page 138). Ces chemins sont exactement ceux dont l'étiquette est une représentation $\mathcal{S}$-adique d'un chemin dans le graphe des graphes d'un sous-shift minimal dont la différence première de complexité est majorée par deux. Donner une caractérisation $\mathcal{S}$-adique de ces sous-shifts revient donc à déterminer exactement les chemins étiquetés qui sont valides. Cette notion est indispensable, car certains chemins dans le graphe des graphes ne peuvent correspondre à une représentation $\mathcal{S}$-adique satisfaisant les conditions voulues. En effet, les exemples 5.1.2, 5.1.3 et 5.1.4 page 138 sont des illustrations de chemins non valides. Nous pouvons observer que ces chemins ne sont pas valides pour différentes raisons. Dans les deux premiers cas, la non-validité provient de la non-presque primitivité (autrement dit, un problème global de la suite de morphismes). Dans le troisième exemple, le problème n'est pas global, mais local. En effet, lors de certaines évolutions (dans ce cas, d'un graphe de type 1 vers un graphe de type 7 ou 8), le choix du morphisme γ_n codant l'évolution[4] peut induire certaines restrictions sur une suite finie $\gamma_{n+1}\gamma_{n+2}\cdots\gamma_{n+k}$ et ces restrictions constituent donc des conditions nécessaires supplémentaires sur $(\gamma_n)_{n\in\mathbb{N}}$.

Forts de ces deux observations, nous parvenons à caractériser les chemins valides par le biais de deux conditions : une locale (les restrictions finies cau-

[4] L'évolution d'un type de graphe vers un autre type de graphe peut se faire de plusieurs façons.

sées par le choix de certains morphismes) et une globale (la presque primitivité et la propreté) (voir Proposition 5.1.5 page 140). Reste donc à déterminer les chemins dans le graphe des graphes qui satisfont ces deux conditions.

Tout d'abord, il convient de remarquer que le graphe des graphes contient quatre composantes fortement connexes :

$$C_1 = \{2\},\ C_2 = \{3\},\ C_3 = \{4\} \text{ et } C_4 = \{1, 5, 6, 7, 8, 9, 10\}.$$

Par conséquent, il est suffisant d'étudier la condition globale séparément dans chacune de ces composantes. Si en plus nous étudions la condition locale dans ces composantes, il suffira ensuite d'étudier la condition locale pour les flèches entre les composantes.

Les trois premières composantes se traitent relativement facilement. En effet, la composante C_1 correspond aux sous-shifts dits d'*Arnoux-Rauzy* et a déjà été largement étudiée. Il n'existe pas de contrainte locale pour cette composante et les morphismes étiquetant les flèches rendent la condition globale particulièrement facile à déterminer (voir Proposition 5.2.1 page 142).

La composante C_2 ne se révèle pas beaucoup plus compliquée. Dans celle-ci, il existe une condition locale sur la suite $(\gamma_n)_{n\in\mathbb{N}}$, condition qui s'exprime au moyen du graphe représenté à la figure 5.3. La condition globale se lit alors directement sur ce graphe (voir Proposition 5.3.1 page 143). Dans le résultat final, il suffira alors de remplacer le sommet 3 du graphe des graphes par le graphe en question.

La composante C_3 est un peu plus complexe. En effet, le choix de certains morphismes γ_n détermine un nombre fini d'évolutions, donc un nombre fini de morphismes $\gamma_{n+1} \cdots \gamma_{n+k}$ (voir Lemme 5.5.1 page 149). Une fois que ces comportements sont déterminés, il suffit alors de remplacer, dans le graphe des graphes, la flèche étiquetée par γ_n par une flèche étiquetée par $\gamma_n\gamma_{n+1} \cdots \gamma_{n+k}$ du sommet de départ de la flèche étiquetée par γ_n jusqu'au sommet d'arrivée de la flèche étiquetée par γ_{n+k}. Déterminer les suites de morphismes qui satisfont la condition globale n'est alors pas difficile (voir Proposition 5.5.2 page 151).

La majeure difficulté de la caractérisation $\mathcal{S}$-adique obtenue relève de la composante C_4. En effet, d'une part, la condition locale est bien plus difficile à gérer à cause de la nature des types de graphes qui constituent cette composante et, d'autre part, le nombre plus élevé de sommets et d'étiquettes rend la condition globale plus compliquée à déterminer. Tout d'abord, nous parvenons à gérer les graphes de type 1 très facilement (ceux-ci correspondant au cas bien connu des sturmiens), puis nous parvenons à traiter les graphes de type 9 et 10 de façon assez similaire à ce que nous avons fait pour la composante C_3. La grosse difficulté concerne les graphes de type 5, 6, 7 et 8 et plus particulièrement les graphes de type 7 et 8. En effet, vu leurs formes, ces quatre types de graphes peuvent être regroupés en deux catégories : 5 et 6 d'une part et 7 et 8 d'autre part. Lorsqu'un graphe de Rauzy G_n évolue vers une de ces catégories, le type exact du nouveau graphe dépend de la longueur de certains chemins dans G_n (voir Figure 5.7 page 153) et ces longueurs requièrent des calculs très techniques (donnés dans l'annexe B). De plus, les morphismes codant des évolutions vers ces catégories peuvent également induire des suites finies d'évolutions qui, elles aussi, dépendent de ces longueurs (voir Lemme 5.6.2 page 154 et Lemme 5.6.4 page 155).

Une fois ces longueurs et évolutions calculées, nous modifions la composante C_4 du graphe des graphes de manière à pouvoir exprimer plus facilement les conditions locale et globale pour qu'un chemin soit valide (voir Proposition 5.6.8 page 166). Enfin, il suffit de regrouper toutes les composantes et conditions obtenues : il s'agit du théorème 5.8.1 page 175.

Conclusions et perspectives

La caractérisation S-adique obtenue aux chapitres 4 et 5 représente une avancée considérable dans l'étude des suites de très faible complexité. Celle-ci sera sans doute d'une grande aide dans la résolution d'autres problèmes liés à ces suites, notamment dans l'étude de leur propriétés géométriques ou arithmétiques. Cependant, les méthodes et constructions utilisées se révèlent beaucoup trop techniques pour espérer les utiliser dans un cadre plus général. En effet, même pour une différence première de complexité majorée par trois (au lieu de deux), les calculs s'alourdissent déjà considérablement. De plus, certains résultats cruciaux semblent intimement liés aux faibles complexités (voir Lemme 4.3.4 page 114 et Exemple 4.3.5 page 115).

Par contre, il serait peut-être intéressant d'étudier le sous-shift engendré par les suites valides dans $\mathcal{S}^{\mathbb{N}}$. Nous pouvons démontrer qu'il n'est pas sofique (voir Proposition 5.8.3 page 181), mais jouit-il d'autres propriétés intéressantes ? Par extension, ces possibles propriétés pourraient-elles se généraliser au cas général et quelles en seraient les retombées sur la conjecture ?

Une autre idée serait de chercher à renforcer les conditions nécessaires obtenues au chapitre 3. Dans un premier temps, il serait intéressant de rendre nécessaire la condition de presque primitivité et ce, même lorsque nous travaillons avec les concaténations de n-segments. Même si nous sommes incapables de le prouver actuellement, nous pensons qu'il est possible de considérer une suite de sous-alphabets $(\tilde{B}_n)$ qui rendraient la suite de morphismes $(\tau_n : \tilde{B}_{n+1}^* \to \tilde{B}_n^*)$ presque primitive.

Une autre question qui généraliserait le travail de Durand (voir [Dur98a] et aussi [HZ99]) est de déterminer pour quelles suites la décomposition obtenue au théorème 3.0.3 (page 76) est ultimement périodique. Des discussions sont en cours avec Štěpán Starosta pour répondre à cette question.

Au delà de la conjecture, il serait également intéressant de poursuivre l'étude initiée par la proposition 2.4.1 (page 72). Plus précisément, cette proposition donne une borne sur la complexité d'une suite S-adique *expansive* (voir Définition 1.3.12 page 33) avec $\mathrm{Card}(S) < +\infty$. Qu'en est-il si la suite S-adique n'est pas expansive, mais croissante partout ? Pourrait-on dire par exemple que, dans ce cas, la complexité est nécessairement sous-polynomiale ?

Chapter 1

Backgrounds

1.1 Words, sequences and languages

Words and languages

An *alphabet* is a finite set A whose elements are called *letters* (or *symbols*). In all what follows (except if it is explicitly stated), we always suppose that $A = \{0, 1, \dots, k-1\}$ for some $k \geq 1$. A *word* u over A is a finite sequence of elements of A. The *length* of a word $u = u_1 \cdots u_\ell$, $u_i \in A$, is the number of letters of u; it is denoted by $|u|$. The unique word of length 0 is called the *empty word* and is denoted by ε. For any word u over A and all letters $a \in A$, we let $|u|_a$ denote the number of occurrences of the letter a in u, i.e., the number of integers $i \in [1, |u|]$ such that $u_i = a$. The set of words of length ℓ over A is denoted by A^ℓ and $A^* = \bigcup_{\ell \in \mathbb{N}} A^\ell$ denotes the set of words over A. We let A^+ denote the set $A^* \setminus \{\varepsilon\}$ of non-empty words over A. The *concatenation* of two words u and v is simply uv and u^n is the concatenation of n copies of u. Endowed with concatenation, A^* is the free monoid generated by A. A *language* over A is a subset L of A^*. If L and M are languages, ML denotes the set of words uv with $u \in M$ and $v \in L$ and L^n denotes the set of words that are concatenations of n words of L.

Sequences

The elements of $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ are respectively called *one-sided sequences* and *two-sided sequences*; they are both denoted by bold letters. For a given two-sided sequence $\mathbf{w}$ over an alphabet A, we write $\mathbf{w} = \cdots \mathbf{w}_{-2}\mathbf{w}_{-1}.\mathbf{w}_0\mathbf{w}_1\mathbf{w}_2 \cdots$ with $\mathbf{w}_i \in A$ for all i. We also write $\mathbf{w}^+ = \mathbf{w}_0\mathbf{w}_1\mathbf{w}_2 \cdots$ and $\mathbf{w}^- = \cdots \mathbf{w}_{-2}\mathbf{w}_{-1}$ and for any non-empty word u over A, the two-sided sequence (resp. one-sided sequence) $\mathbf{w}$ composed of consecutive copies of u is denoted by $\mathbf{w} = u^\infty$

(resp. $\mathbf{w} = u^\omega$). Given two non-empty words u and v, we also let ${}^\omega u.v^\omega$ denote the two-sided sequence $\mathbf{w} = \cdots uuu.vvv \cdots$. A two-sided sequence (resp. a one-sided sequence) $\mathbf{w}$ is *periodic* if there is a word u such that $\mathbf{w} = u^\infty$ (resp. $\mathbf{w} = u^\omega$). A one-sided sequence $\mathbf{w}$ is *ultimately periodic* if there are two words u and v, $v \neq \varepsilon$ such that $\mathbf{w} = vu^\omega$. For a one-sided sequence (resp. two-sided sequence) $\mathbf{w}$ and a language L, we also write $\mathbf{w} \in L^\omega$ (resp. $\mathbf{w} \in L^\infty$) whenever $\mathbf{w}$ is composed of consecutive copies of words of L.

Prefixes, suffixes and factors

For a word $u = u_1 \cdots u_\ell$, we write $u_{[i,j]} = u_i \cdots u_j$ for $1 \leq i \leq j \leq \ell$. A word v is a *factor* of a word u (or *occurs at position* i in u) if $u_{[i,j]} = v$ for some integers i and j. It is a *prefix* (resp. *suffix*) if $i = 1$ (resp. $j = |u|$) and a *proper factor* if it is different from u. Given a language L, the language $\mathrm{Pref}(L)$ (resp. $\mathrm{Suff}(L)$, $\mathrm{Fact}(L)$) is the set of prefixes (resp. suffixes, factors) of words in L. If L contains a unique element u, we respectively write $\mathrm{Pref}(u)$, $\mathrm{Suff}(u)$ and $\mathrm{Fact}(u)$ instead of $\mathrm{Pref}(\{u\})$, $\mathrm{Suff}(\{u\})$ and $\mathrm{Fact}(\{u\})$. All these notions can be extended to one-sided sequences (resp. two-sided sequences): in the definition of prefixes, suffixes and factors, all we have to do is to put $i, j \in \mathbb{N}$ (resp. $i, j \in \mathbb{Z}$), $i \leq j$, $i = 0$ (resp. $i = -\infty$) for prefixes and $j = +\infty$ for suffixes. In particular, when $\mathbf{w}$ is a (one-sided or two-sided) sequence, the set $\mathrm{Fact}(\mathbf{w})$ is called the *language* of the sequence and is usually denoted by $L(\mathbf{w})$. For each $n \in \mathbb{N}$, we also let $L_n(\mathbf{w})$ denote the set of factors of length n in $\mathbf{w}$, i.e., $L_n(\mathbf{w}) = L(\mathbf{w}) \cap A^n$.

Definition 1.1.1. A language $L \subset A^*$ is *prolongable* if for all words $u \in L$, there are two letters $a, b \in A$ such that $aub \in L$.

Definition 1.1.2. A language $L \subset A^*$ is *factorial* if $\mathrm{Fact}(L) \subset L$.

Return words

Given a sequence $\mathbf{w}$ and a factor u of $\mathbf{w}$, a *left return word* to u in $\mathbf{w}$ is a word v such that $vu \in L(\mathbf{w})$, u is prefix of vu and u occurs only twice in vu. We can similarly define the notion of *right return word* by exchanging vu by uv and supposing that u is suffix of uv. Note that v is a left return word to u in $\mathbf{w}$ if and only if there exists a right return word v' to u in $\mathbf{w}$ such that $vu = uv'$. The set of left return words (resp. right return words) to u in $\mathbf{w}$ is denoted by $LRW_{\mathbf{w}}(u)$ (resp. $RRW_{\mathbf{w}}(u)$).

We can also extend these two notions to languages. let L be a subset of $L(\mathbf{w})$. A *left return word* to L in $\mathbf{w}$ is a word r such that there are two words u and v in L such that rv is a factor of $\mathbf{w}$, rv admits u as a prefix and u and v

are the only words of L that occur in rv. Similarly, a *right return word* to L in $\mathbf{w}$ is a word r such that there are two words u and v in L such that ur is a factor of $\mathbf{w}$, ur admits v as a suffix and u and v are the only words of L that occur in ur. The set of left return words to L in $\mathbf{w}$ is denoted by $\mathrm{LRW}_{\mathbf{w}}(L)$ and the set of right return words to L in $\mathbf{w}$ is denoted by $\mathrm{RRW}_{\mathbf{w}}(L)$.

Recurrence and uniform recurrence

A one-sided (resp. two-sided) sequence $\mathbf{w}$ is *recurrent* if all factors u of $\mathbf{w}$ occur infinitely often in $\mathbf{w}$ (resp. in $\mathbf{w}^+$ and in $\mathbf{w}^-$). It is *uniformly recurrent* if it is recurrent and every factor occurs with bounded gaps, i.e., if u is a factor of $\mathbf{w}$, there is a constant K_u such that for all integers i, j such that $\mathbf{w}_{[i,i+|u|-1]}$ and $\mathbf{w}_{[j,j+|u|-1]}$ are two consecutive occurrences of u in $\mathbf{w}$, then $|i-j| \leq K_u$. In particular, a sequence $\mathbf{w}$ is uniformly recurrent if and only if it is recurrent and any factor of $\mathbf{w}$ has a finite number of return words.

Remark 1.1.3. In the sequel, we sometimes do not explicitly say if the considered sequence is one-sided or two-sided. This is either when the context is clear enough or when what is said holds for both kind of sequences.

1.2 Factor complexity

The *complexity function* of a sequence $\mathbf{w}$ is the function $p_{\mathbf{w}}$ (or simply p) that counts the number of factors of a given length in $\mathbf{w}$:

$$p_{\mathbf{w}} : \mathbb{N} \to \mathbb{N} : n \mapsto \# L_n(\mathbf{w}).$$

It is obvious that if $\mathbf{w}$ is a sequence over an alphabet A, its complexity function $p_{\mathbf{w}}(n)$ is non-decreasing (any factor being prolongable to the right in $L(\mathbf{w})$) and bounded by $\mathrm{Card}(A)^n$. Moreover, the following trivially holds.

Proposition 1.2.1. *For any sequence* $\mathbf{w}$ *and all non-negative integers* m *and* n*, we have* $p_{\mathbf{w}}(m+n) \leq p_{\mathbf{w}}(m)p_{\mathbf{w}}(n)$.

However, not every function satisfying these properties can be a complexity function. For instance, the following result is well known.

Theorem 1.2.2 (Morse and Hedlund [MH38])**.** *Let* $\mathbf{w}$ *be a one-sided sequence over* A*. The following are equivalent.*

1. $\mathbf{w}$ *is ultimately periodic;*

2. $p_{\mathbf{w}}(n) \leq n$ *for some* $n \geq 1$*;*

3. $p_{\mathbf{w}}(n+1) = p_{\mathbf{w}}(n)$ *for some* $n \geq 1$.

In particular, this implies that if a one-sided sequence $\mathbf{w}$ is not ultimately periodic, then $p_{\mathbf{w}}(n) \geq n+1$ for all n. Sequences with minimal complexity $p(n) = n+1$ for all n exist and are called *Sturmian sequences*. They admit several equivalent definitions and a huge literature is devoted to them. See Chapter 2 of [Lot02] and Chapter 6 of [Fog02] for surveys on these sequences. See [All94, Fer99] and Chapter 4 of [BR10] for surveys on the complexity of sequences.

A sequence $\mathbf{w}$ has a *sub-linear complexity* (or an *at most linear complexity*) if there is a constant D such that

$$\forall n \geq 1, \quad p_{\mathbf{w}}(n) \leq Dn.$$

One could equivalently say that $\mathbf{w}$ has a sub-affine complexity if for all $n \in \mathbb{N}$, one has $p_{\mathbf{w}}(n) \leq Dn + C$ with $C \geq 1$. Indeed if for all n we have $p_{\mathbf{w}}(n) \leq Dn + C$, then for all $n \geq 1$ we have $p_{\mathbf{w}}(n) \leq (D+C)n$.

In order to compute the complexity function of a sequence, it is natural to study its first difference $p(n+1) - p(n)$ since it represents the growth rate of complexity. A first easy result is that if the first difference of complexity is bounded, say by a constant K, then $p(n) \leq 1 + Kn$. Indeed,

$$p(n) = 1 + \sum_{i=0}^{n-1} p(i+1) - p(i) \leq 1 + Kn.$$

Cassaigne proved that the converse is also true.

Theorem 1.2.3 (Cassaigne [Cas96]). *A sequence* $\mathbf{w}$ *has a sub-linear complexity if and only if the first difference of its complexity* $p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n)$ *is bounded.*

The first difference of complexity is also closely related to special factors that were first introduced by Rauzy in [Rau83] (see also [Cas97]). A factor u of a sequence $\mathbf{w}$ is a *right special factor* (resp. a *left special factor*) if there are at least two letters a and b in A such that ua and ub (resp. au and bu) are factors of $\mathbf{w}$. It is a *bispecial factor* if it is right and left special. For all n, we let $LS_n(\mathbf{w})$ (resp. $RS_n(\mathbf{w})$) denote the set of left (resp. right) special factors of length n in $\mathbf{w}$. For u in $L(\mathbf{w})$, we also let $\delta^+(u)$ (resp. $\delta^-(u)$) denote the number of letters a in A such that ua (resp. au) is in $L(\mathbf{w})$. For all n we have

$$p(n+1) - p(n) = \sum_{u \in RS_n(\mathbf{w})} \underbrace{(\delta^+(u) - 1)}_{\geq 1} \tag{1.1}$$

$$\leq \sum_{u \in LS_n(\mathbf{w})} \underbrace{(\delta^-(u) - 1)}_{\geq 1} \tag{1.2}$$

and if $\mathbf{w}$ is recurrent or two-sided, then the equality holds in (1.2). Theorem 1.2.3 can therefore be rephrased as follows.

Corollary 1.2.4. *A sequence* $\mathbf{w}$ *has a sub-linear complexity if and only if it has a bounded number of left and right special factors of each length.*

For the same reason that it was convenient to study the first difference of complexity to compute $p(n)$, we may want to study the second difference of complexity to compute $p(n+1) - p(n)$. More precisely, if for all n we write $s(n) = p(n+1) - p(n)$, we may want to study $s(n+1) - s(n) = p(n+2) - 2p(n+1) + p(n)$. Similarly to the link between special factors and the first difference of complexity, there are some particular factors that are linked to the second difference of complexity: the bispecial factors.

Definition 1.2.5. Let u be a bispecial factor of a sequence $\mathbf{w}$. The *bilateral order* of u is defined by

$$m(u) = \operatorname{Card}(L(\mathbf{w}) \cap AuA) - \delta^+ u - \delta^- u + 1.$$

A bispecial factor u is said to be *weak* (resp. *ordinary*, *strong*) whenever $m(u) < 0$ (resp. $m(u) = 0$, $m(u) > 0$). Observe that since

$$\#(L(\mathbf{w}) \cap AuA) = \sum_{aB \in L(\mathbf{w})} \delta^+(au),$$

we have

$$m(u) > 0 \Leftrightarrow \sum_{au \in L(\mathbf{w})} (\delta^+(au) - 1) > \delta^+(u) - 1 \qquad (1.3)$$

For sequences over a binary alphabet A, we have $m(u) \in \{-1, 0, 1\}$ for all factors u and a bispecial factor u is weak (resp. ordinary, strong) if $\#(L(X) \cap AuA) = 2$ (resp. 3, 4). Observe that for non-bispecial factors u, we always have $m(u) = 0$.

Proposition 1.2.6 (Cassaigne [Cas97])**.** *Let* $\mathbf{w}$ *be a recurrent sequence over an alphabet* A*. If for all* n*, we write* $s_{\mathbf{w}}(n) = p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n)$*, we have*

$$s_{\mathbf{w}}(n+1) - s_{\mathbf{w}}(n) = \sum_{u \in L_n(\mathbf{w})} m(u).$$

According to what is said in [Cas97], it seems to be difficult to find other particular factors that would be linked to differences of complexity of larger order.

1.3 S-adicity

Morphisms

Let A and B be two alphabets. A *morphism* (or a *substitution*) σ is a map from A^* to B^* such that $\sigma(uv) = \sigma(u)\sigma(v)$ for all words u and v over A; it is completely determined by the images of letters. When σ is *non-erasing* (i.e., when $\sigma(a) \neq \varepsilon$ for all a in A), it can be extended to a map from $A^{\mathbb{N}}$ to $B^{\mathbb{N}}$ by putting $\sigma(\mathbf{w}_0\mathbf{w}_1\mathbf{w}_2\cdots) = \sigma(\mathbf{w}_0)\sigma(\mathbf{w}_1)\sigma(\mathbf{w}_2)\cdots$ and, similarly, to a map from $A^{\mathbb{Z}}$ to $B^{\mathbb{Z}}$ by considering $\sigma(\cdots\mathbf{w}_{-2}\mathbf{w}_{-1}.\mathbf{w}_0\mathbf{w}_1\mathbf{w}_2\cdots) = \cdots\sigma(\mathbf{w}_{-2})\sigma(\mathbf{w}_{-1}).\sigma(\mathbf{w}_0)\sigma(\mathbf{w}_1)\sigma(\mathbf{w}_2)\cdots$. These maps are still denoted by σ. A morphism $\sigma : A^* \to A^*$ is *uniform* if there exists a positive integer k such that for all letters $a \in A$, we have $|\sigma(a)| = k$; it is *letter-to-letter* if $k = 1$.

When $B = A$, we say that $\sigma : A^* \to A^*$ is an endomorphism and we still abbreviate this by *morphism over A*. In this case, any word or sequence x such that $\sigma(x) = x$ is called a *fixed point* of σ. A morphism σ over A is *right prolongable* (resp. *left prolongable*) if there is a letter a in A such that $\sigma(a) = au$ (resp. $\sigma(a) = ua$) with $u \in A^+$ and $\lim_{n\to+\infty} |\sigma^n(a)| = +\infty$. It is *bi-prolongable* if it is left and right prolongable. If the morphism σ is right prolongable on the letter a (resp. bi-prolongable to the right on the letter a and to the left on the letter b), the sequence $(\sigma^n(a^\omega))_{n\in\mathbb{N}}$ (resp. $(\sigma^n({}^\omega b.a^\omega))_{n\in\mathbb{N}}$) converges in $A^{\mathbb{N}}$ (resp. in $A^{\mathbb{Z}}$) to a limit denoted by $\sigma^\omega(a)$ (resp. $\sigma^\omega(b.a)$) and this limit is a fixed point of σ. A one-sided sequence $\mathbf{w} \in A^{\mathbb{N}}$ is *purely morphic* (or *purely substitutive*) if there is a morphism σ over A prolongable on a such that $\mathbf{w} = \sigma^\omega(a)$. It is *morphic* (or *substitutive*) if it is the image under a morphism of a purely morphic sequence. We could easily extend these notions to two-sided sequences by replacing $\sigma^\omega(a)$ by $\sigma^\omega(b.a)$.

Example 1.3.1. The next two well-known morphisms will occur several times in the sequel. The first one is the *Fibonacci morphism* φ defined by

$$\varphi : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 0 \end{cases} .$$

It is right prolongable on the letter 0 and the corresponding fixed point $\varphi^\omega(0)$ is called the *Fibonacci sequence* and is denoted by $\mathbf{f}$; it is a Sturmian sequence.

The second one is the *Thue-Morse morphism* μ defined by

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

It is also right prolongable on 0 and on 1. The fixed point $\mu^\omega(0)$ is called the *Thue-Morse sequence* and is denoted by $\mathbf{t}$; it has a sub-linear complexity (see [Brl89, dLV89]).

In the sequel we will need the following definitions (that will be sometimes recalled).

Definition 1.3.2. A morphism $\sigma : A^* \to A^*$ is *primitive* if there is an integer n such that all letters in A occurs in all images $\sigma_n(b)$; it is *strongly primitive* if $n = 1$. The strong primitivity can be extended to morphisms $\sigma : A^* \to B^*$ by saying that all letters of B occur in all images $\sigma(a)$ for $a \in A$.

Definition 1.3.3. A morphism $\sigma : A^* \to B^*$ is said to be *left proper* (resp. right proper) if there exists a letter $b \in B$ such that $\sigma(A) \subset bB^*$ (resp. $\sigma(A) \subset B^*b$); it is said to be *proper* if it is left and right proper. Observe that if $\sigma : A^* \to A^*$ is proper, then $(\sigma^n({}^\omega c.d^\omega))_{n\in\mathbb{N}}$ converges in $A^\mathbb{Z}$ to the same limit $\mathbf{w}$ for all $c, d \in A$.

Definition 1.3.4. A morphism $\sigma : A^* \to B^*$ is said to be *expansive* if for all letters a in A, $|\sigma(a)| \geq 2$.

Definition 1.3.5. A morphism $\sigma : A^* \to A^*$ is said to be *everywhere growing* if for all letters a in A, the length of $\sigma^n(a)$ tends to infinity when n increases. A letter $a \in A$ such that the sequence $(|\sigma^n(a)|)_{n\in\mathbb{N}}$ is bounded is called a *bounded letter*, otherwise it is said to be *growing*. We let $A_{\mathfrak{B},\sigma}$ (or $A_{\mathfrak{B}}$ when the context is clear) denote the set of bounded letters. Observe that if σ is everywhere growing, there exists an integer k and a letter $a \in A$ such that σ^k is prolongable on a. There is also an integer k' such that $\sigma^{kk'}$ is expansive. Moreover, $\sigma^{kk'}$ is obviously still prolongable on a and $(\sigma^{kk'})^\omega(a) = (\sigma^k)^\omega(a)$.

S-adicity

Definition 1.3.6. The notion of S*-adic* sequence generalizes the notion of morphic sequence. Let $\mathbf{w}$ be a one-sided sequence over A. If S is a set of morphisms, an S*-adic representation* of $\mathbf{w}$ is given by a sequence $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ of morphisms in S and a sequence $(a_n)_{n\in\mathbb{N}}$ of letters, $a_i \in A_i$ for all i such that $A_0 = A$, $\lim_{n\to+\infty} |\sigma_0\sigma_1\cdots\sigma_n(a_{n+1})| = +\infty$ and

$$\mathbf{w} = \lim_{n\to+\infty} \sigma_0\sigma_1\cdots\sigma_n(a_{n+1}^\omega).$$

The sequence $(\sigma_n)_{n\in\mathbb{N}} \in S^\mathbb{N}$ is the *directive word* of the representation and we say that $\mathbf{w}$ *is directed* by $(\sigma_n, a_n)_{n\in\mathbb{N}}$. In the sequel, we will say that a sequence $\mathbf{w}$ is S-adic if S is a set of morphisms such that $\mathbf{w}$ is directed by

$(\sigma_n, a_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}} \times \prod_{n=0}^{\infty} A_n$. For a given set S of morphisms, a sequence might admit several directive words. However in most of the cases, the directive word is given by the context and it will always supposed to be fixed. In other words, when talking about an S-adic sequence, its directive word is always implicitly fixed.

Observe that we still suppose that all alphabets A_n are $\{0,\ldots,k_n-1\}$ for some integers k_n. Consequently, when $\mathrm{Card}(S) < +\infty$ (which will be often the case in the sequel), we have $\mathrm{Card}\left(\bigcup_{n\in\mathbb{N}} A_n\right) < +\infty$. In the sequel, we let $\mathbf{A}$ denote the set $\bigcup_{n\in\mathbb{N}} A_n$.

Example 1.3.7. Let us define the following four morphisms:

$$L_0 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \end{cases} \qquad R_0 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 10 \end{cases}$$

$$L_1 : \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \end{cases} \qquad R_1 : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 1 \end{cases}$$

Since the work of Hedlund and Morse [MH40] (see also for instance [BHZ06]) it is well known that for any Sturmian sequence $\mathbf{w}$, there is a sequence $(k_n)_{n\in\mathbb{N}}$ of non-negative integers such that

$$\mathbf{w} = \lim_{n\to+\infty} L_0^{k_0} R_0^{k_1} L_1^{k_2} R_1^{k_3} L_0^{k_4} R_0^{k_5} \cdots L_1^{k_{4n+2}} R_1^{k_{4n+3}}(0^\omega).$$

Remark 1.3.8. All these definitions can easily be adapted to two-sided sequences: as for morphic sequences, we have to consider two sequences of letters $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, $a_i, b_i \in A_i$, such that $\lim_{n\to+\infty} |\sigma_0\sigma_1\cdots\sigma_n(a_{n+1})| = +\infty$, $\lim_{n\to+\infty} |\sigma_0\sigma_1\cdots\sigma_n(b_{n+1})| = +\infty$ and

$$\mathbf{w} = \lim_{n\to+\infty} \sigma_0\sigma_1\cdots\sigma_n({}^\omega b.a^\omega).$$

In this case, we say that $\mathbf{w}$ is directed by $(\sigma_n, b_n.a_n)_{n\in\mathbb{N}}$.

In the sequel we will use the following definitions (that will sometimes be recalled). Note that some of these definitions already have another signification in terms of *DT0L* languages (see [ELR76]). Roughly speaking, an everywhere growing *DT0L* language is more or less the same as an expansive S-adic sequence (Definition 1.3.12).

Definition 1.3.9. We say that a sequence of morphisms $(\tau_n : B_{n+1}^* \to B_n^*)_{n\in\mathbb{N}}$ is a *contraction* of $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ if there is an increasing sequence of integers $(i_n)_{n\in\mathbb{N}}$ such that $i_0 = 0$ and for all n in $\mathbb{N}$, $B_n = A_{i_n}$ and

$$\tau_n = \sigma_{i_n}\sigma_{i_n+1}\cdots\sigma_{i_{n+1}-1}.$$

Definition 1.3.10. As for morphisms, we say that a directive word $(\sigma_n)_{n\in\mathbb{N}}$ is *primitive* if there exists a non-negative integer s_0 such that for all non-negative integers s, the morphism $\sigma_s \cdots \sigma_{s+s_0}$ is strongly primitive.

Definition 1.3.11. We say that a directive word $(\sigma_n)_{n\in\mathbb{N}}$ is *left or right proper* if all its morphisms are respectively left or right proper. It is proper S-adic if it is left and right proper.

Definition 1.3.12. We say that a directive word $(\sigma_n)_{n\in\mathbb{N}}$ is *expansive* if all its morphisms are expansive.

Definition 1.3.13. We say that a directive word $(\sigma_n)_{n\in\mathbb{N}}$ is *everywhere growing* if for all sequences of letters $(a_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} A_n$, the length of $\sigma_0\sigma_1\cdots\sigma_n(a_{n+1})$ tends to infinity when n increases. A sequence $(a_n)_{n\in\mathbb{N}}$ for which $(|\sigma_0\sigma_1\cdots\sigma_n(a_{n+1})|)_{n\in\mathbb{N}}$ is bounded is called a *bounded sequence* and the set of such sequences is denoted by $\mathcal{A}_{\mathfrak{B},(\sigma_n)}$ (or $\mathcal{A}_{\mathfrak{B}}$). When the sequence $(a_n)_{n\in\mathbb{N}}$ is simply a^ω we talk about *bounded letter*.

Definition 1.3.14. We say that a directive word $(\sigma_n)_{n\in\mathbb{N}}$ is *almost primitive* if it is everywhere growing and if for all sequences of letters $(a_n)_{n\in\mathbb{N}} \in \left(\prod_{n\in\mathbb{N}} A_n\right)$ and all integers r, there is an integer $s > r$ such that all letters of A_r occur in $\sigma_r \cdots \sigma_s(a_{s+1})$.

Remark 1.3.15. By abuse of language, we will say that an S-adic sequence has the property P (P being one of the previous definition) if its directive word has it. For instance, we will say that a sequence $\mathbf{w}$ is *primitive S-adic* if its directive word is primitive.

1.4 Topological dynamical systems

A *(topological) dynamical system* (X, T) is defined as a compact metric space X together with a continuous and onto map $T : X \to X$. Given a point $x \in X$, the *orbit* of x is the set $O(x) = \{T^n x \mid n \in \mathbb{Z}\}$.

Example 1.4.1. Let $X = \mathbb{R}/\mathbb{Z}$, $\alpha \in \mathbb{R}$ and $R_\alpha : X \to X$ be the rotation of angle α defined by

$$R_\alpha(x) = x + \alpha \mod 1.$$

Then the couple (X, R_α) is a topological dynamical system whose distance is given by

$$d(x, y) = \min\{|x - y|, |x - y + 1|\}$$

When $\alpha \in \mathbb{R}\setminus\mathbb{Q}$, it is closely related to Sturmian sequences (see Section 2.2.1 for more details).

Two dynamical systems (X_1, T_1) and (X_2, T_2) are said to be *topologically conjugate* (or *topologically isomorphic*) if there is a homeomorphism $\phi : X \to Y$ that conjugates T_1 and T_2, i.e., such that

$$\phi \circ T_1 = T_2 \circ \phi.$$

Minimality

A dynamical system (X, T) is *minimal* if the only closed T-invariant subsets of X are X and $\emptyset$. Another equivalent definition is that for all $x \in X$, the orbit of x is dense in X. A minimal dynamical system (X, T) is said to be *periodic* whenever X is finite. In particular, if (X, T) is not minimal, there is a subset $Y \subset X$ such that $(Y, T_{|Y})$ is minimal where $T_{|Y}$ is the restriction of T to Y.

Let us consider the dynamical system (X, R_α) of Example 1.4.1. It is well known that for $x \in \mathbb{R}$, the sequence $(\{x+n\alpha\})_{n\in\mathbb{N}}$ is dense in $[0, 1[$ if and only if α is irrational. Consequently, the dynamical system (X, R_α) is minimal if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Subshifts

First, recall that with the product topology of the discrete topology over A, both sets $A^{\mathbb{Z}}$ and $A^{\mathbb{N}}$ are compact metric spaces. The following metrics respectively on $A^{\mathbb{Z}}$ and $A^{\mathbb{N}}$ define the same topology.

$$\begin{aligned} d_{\mathbb{Z}}(\mathbf{x}, \mathbf{y}) &= 2^{-n} \text{ for } n = \inf\{i \in \mathbb{N} \mid \mathbf{x}_i \neq \mathbf{y}_i \text{ or } \mathbf{x}_{-i} \neq \mathbf{y}_{-i}\} \\ d_{\mathbb{N}}(\mathbf{x}, \mathbf{y}) &= 2^{-n} \text{ for } n = \inf\{i \in \mathbb{N} \mid \mathbf{x}_i \neq \mathbf{y}_i\}. \end{aligned}$$

The *shift transformation* T is defined over $A^{\mathbb{Z}}$ (or $A^{\mathbb{N}}$) by

$$T : \mathbf{w} = (\mathbf{w}_i)_{i\in\mathbb{Z}} \mapsto T(\mathbf{w}) = (\mathbf{w}_{i+1})_{i\in\mathbb{Z}}$$

(where we replace $\mathbb{Z}$ by $\mathbb{N}$ when working on $A^{\mathbb{N}}$). It is a continuous and onto map over $A^{\mathbb{Z}}$ (or $A^{\mathbb{N}}$) so that both $(A^{\mathbb{Z}}, T)$ and $(A^{\mathbb{N}}, T)$ are topological dynamical systems, respectively called *two-sided full shift* and *one-sided full shift*. Observe that for a two-sided full shift $(A^{\mathbb{Z}}, T)$, T is also one-to-one although it is not the case for one-sided subshifts.

If X is a closed T-invariant subset of $A^{\mathbb{Z}}$ or $A^{\mathbb{N}}$, then $(X, T_{|X})$ is also a dynamical system and is called a (*two-sided* or *one-sided*) subshift.

The *language* of a subshift X is the union of the languages of its elements; we denote it by $L(X)$ and we write $L_n(X) = L(X) \cap A^n$ for all $n \geq 0$. Observe that a subshift (X, T) is completely determined by its language. Indeed, a sequence $\mathbf{w}$ belongs to (X, T) if and only if $L(\mathbf{w}) \subset L(X)$.

Let $\mathbf{w}$ be a sequence over A. We denote by $X_{\mathbf{w}}$ the set $\{\mathbf{x} \in A^{\mathbb{Z}} \mid L(\mathbf{x}) \subset L(\mathbf{w})\}$. Then $(X_{\mathbf{w}}, T)$ is a two-sided subshift called the *subshift generated by* $\mathbf{w}$ and when $\mathbf{w}$ is a two-sided sequence, we have $X_{\mathbf{w}} = \overline{O(\mathbf{w})}$. If $\mathbf{w}$ is a purely morphic sequence $\sigma^{\omega}(a)$, then $(X_{\mathbf{w}}, T)$ is a *substitutive subshift.* If moreover σ is primitive, then all its fixed points generate the same minimal subshift which is denoted by (X_σ, T). Similarly, a minimal subshift X is *S-adic* if it is generated by an S-adic sequence. In that case, any directive word of elements of the subshift is a directive word of X.

Observe that if a minimal subshift (X, T) is periodic, then X contains only periodic sequences. Moreover, the following are equivalent:

- (X, T) is minimal,
- for all $\mathbf{w}$ in X, $X = X_{\mathbf{w}} = X_{\mathbf{w}^+}$,
- for all $\mathbf{w}$ in X, $L(X) = L(\mathbf{w}) = L(\mathbf{w}^+)$.

We also have that $(X_{\mathbf{w}}, T)$ is minimal if and only if $\mathbf{w}$ is uniformly recurrent.

As for sequences, we can define the *complexity function* of a subshift (X, T) as the function p_X (or simply p) that counts the number of factors of a given length in $L(X)$:

$$p_X : \mathbb{N} \to \mathbb{N} : n \mapsto \#L_n(X).$$

Observe that for minimal subshifts (X, T), since $L(X) = L(\mathbf{w})$ for all $\mathbf{w} \in X$, we have $p_{\mathbf{w}}(n) = p_X(n)$ for all $\mathbf{w} \in X$ and Equalities (1.1) and (1.2) also hold for this case.

1.5 Rauzy graphs

In the sequel, Rauzy graphs are widely used. The S-adic representations that we get are based on them and it is therefore crucial to really understand what they are and how they evolve. However they will only be needed in Chapter 3, Chapter 4 and Chapter 5. In other words, this section is not necessary to understand Chapter 2 and could therefore be read later. First let us recall some definitions of graph theory.

A *directed graph* G is a couple (V, E) where V is the set of vertices and $E \subset V \times V$ is the set of edges. Edges may be labelled by elements of a set A and then $E \subset V \times A \times V$. If $e = (u, a, v)$ is an edge of G, we let $o(e) = u$ denote its starting vertex (o for *outgoing*) and $i(e) = v$ its ending vertex (i for *incoming*). A *path* p in G is a sequence $(v_0, a_1, v_1)(v_1, a_2, v_2) \ldots (v_{\ell-1}, a_\ell, v_\ell)$ of consecutive edges. The label of p is the ℓ-tuple $(a_1, a_2, \ldots, a_\ell)$. However in the sequel we will simply denote it by concatenating the labels of each edge. We also let $o(p)$ denote the starting vertex v_0 of p and by $i(p)$ its ending vertex v_ℓ; they are called the *extremities* of p and $v_1, \ldots, v_{\ell-1}$ are called *interior vertices*. The *length* of a path is the number of edges composing it. A *subpath* of $p = (v_0, a_1, v_1)(v_1, a_2, v_2) \dots (v_{\ell-1}, a_\ell, v_\ell)$ is a path $q = (u_0, b_1, u_1)(u_1, b_2, u_2) \dots (u_{k-1}, b_k, u_k)$ such that $k \leq \ell$ and there exists an integer $i \in [0, \ell - k]$ such that $(v_{i+j}, a_{i+j+1}, v_{i+j+1}) = (u_j, b_{j+1}, u_{j+1})$ for all integers $j \in [0, k-1]$. It is a *proper subpath* if $k < \ell$.

All the notions of this section are defined with respect to sequences; they can easily be adapted to subshifts.

1.5.1 Rauzy graphs and allowed paths

Let $\mathbf{w}$ be a sequence over an alphabet A. For each non-negative integer n, we define the *Rauzy graph of order* n of $\mathbf{w}$ (also called *graph of words of length* n), denoted by $G_n(\mathbf{w})$ (or simply G_n) as the directed graph $(V(n), E(n))$, where

- the set $V(n)$ of vertices is the set L_n of factors of length n of $\mathbf{w}$ and

- there is an edge from u to v if there are two letters a and b in A such that $ub = av \in L_{n+1}$.

In the literature, there are different ways of labelling the edges. Indeed, the edges are sometimes labelled by the letter a, by the letter b, by the couple (a, b) or by the word av, i.e., the following four notations exist:

$$u \xrightarrow{b} v \qquad u \xrightarrow[a]{} v \qquad u \xrightarrow[a]{b} v \qquad u \xrightarrow{av} v.$$

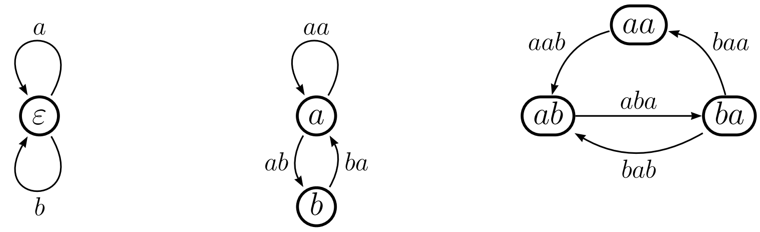
For an edge $e = (u, (a, b), v) = u \xrightarrow[a]{b} v$, let us call $\lambda_L(e) = a$ its *left label*, $\lambda_R(e) = b$ its *right label* and $\lambda(e) = ub = av$ its *full label*. Same definitions hold for labels of paths (left and right labels being words of same length as the considered path) where we naturally extend the map λ to the set of paths by $\lambda\left((u_0, (a_1, b_1), u_1)(u_1, (a_2, b_2), u_2) \cdots (u_{\ell-1}, (a_\ell, b_\ell), u_\ell)\right) = u_0 b_1 b_2 \cdots b_\ell = a_1 a_2 \cdots a_\ell u_\ell$.

Example 1.5.1. Let $\mathbf{f}$ be the *Fibonacci sequence* (see Example 1.3.1). Figure 1.1 represents the first three Rauzy graphs of $\mathbf{f}$ (with full labels on the edges).

Remark 1.5.2. A sequence is recurrent if and only if all its Rauzy graphs are *strongly connected* (that is for all vertices u and v of G_n there is a path p from u to v, i.e., $o(p) = u$ and $i(p) = v$.).

We say that a vertex v is *right special* (resp. *left special*, *bispecial*) if it corresponds to a right special (resp. left special, bispecial) factor.

Remark 1.5.3. By definition of Rauzy graphs, $(u, (a, b), v)$ is an edge in $G_n(\mathbf{w})$ if and only if the word ub is in the language $L(\mathbf{w})$. It is also clear that for any



(a) $G_0(\mathbf{f})$ (b) $G_1(\mathbf{f})$ (c) $G_2(\mathbf{f})$

Figure 1.1: First Rauzy graphs of the Fibonacci sequence.

word u in $L(\mathbf{w})$, for all non-negative integers $n < |u|$ there is a non-empty path p in $G_n(\mathbf{w})$ whose full label $\lambda(p)$ is u. The contrary is not true: not every path in $G_n(\mathbf{w})$ has a full label that is a factor of $\mathbf{w}$. Indeed, in the Rauzy graph $G_1(\mathbf{f})$ of the Fibonacci sequence (see Figure 1.1(b)), the full label of the path $(a,(a,a),a)^n$ is a^{n+1} for each n and this word is not in the language as soon as $n \geq 2$. The reason is that once we have reached the vertex a coming from some edge, we have two possibilities: either we stay in this vertex passing through the loop $(a,(a,a),a)$, or we go in the vertex b with the edge $(a,(a,b),b)$. These possibilities exist because the word a is a right special factor of the Fibonacci sequence, but this particularity only implies that, starting at vertex a, we can read a a or a b. In other words, it does not take care of what happened before (i.e., from which edge we arrived in this vertex) although we have to. Indeed, if we come from the loop, this means that the previous vertex of the path was the vertex a and the full label of this path is aa. Then the only possibility that we really have is to go into the vertex b (because $aaa \notin L(\mathbf{f})$).

Definition 1.5.4. A path in a Rauzy graph $G_n(\mathbf{w})$ is said to be *allowed* if its full label is a word in $L(\mathbf{w})$.

Note that, by definition, any path $p = (v_0,(a_1,b_1),v_1)\cdots(v_{\ell-1},(a_\ell,b_\ell),v_\ell)$ that does not contain any subpath $(v_i,(a_{i+1},b_{i+1}),v_{i+1})\cdots(v_{j-1},(a_j,b_j),v_j)$, $1 \leq i \leq j \leq \ell-1$ with v_i left special and v_j right special is allowed. Moreover, the following result trivially holds true.

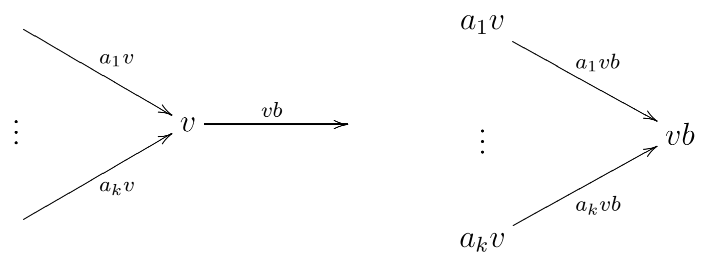
Proposition 1.5.5. *Let G_n be a Rauzy graph of order n. For all paths p of length $\ell \leq n$ in G_n, the left label of p is a prefix of $o(p)$ and the right label of p is a suffix of $i(p)$. Similarly, for any path p of length $\ell \geq n$, $i(p)$ is equal to the suffix of length n of $\lambda_R(p)$ and $o(p)$ is equal to the prefix of length n of $\lambda_L(p)$.*

1.5.2 Evolution of Rauzy graphs

In the sequel we will need to let the Rauzy graphs evolve, i.e., we will need to go from $G_n(\mathbf{w})$ to $G_{n+1}(\mathbf{w})$. Let us see how it goes. As the set of edges of G_n is in bijection with L_{n+1}, we can write G_n as the directed graph $(L_n(\mathbf{w}), L_{n+1}(\mathbf{w}))$. Then to get the Rauzy graph of order $n+1$, we only have to replace each edge of $G_n(\mathbf{w})$ by a vertex and to define the edges in the following way:

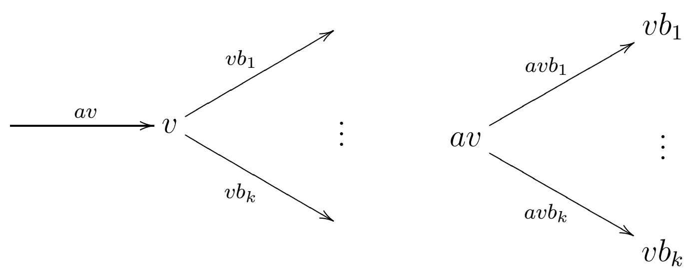
- for each non special vertex v in $G_n(\mathbf{w})$, we replace $\xrightarrow{av} v \xrightarrow{vb}$ by $av \xrightarrow{avb} vb$;

- for each left special vertex v in $G_n(\mathbf{w})$ that is not right special we make the following changes



Transitions in G_n Transitions in G_{n+1}

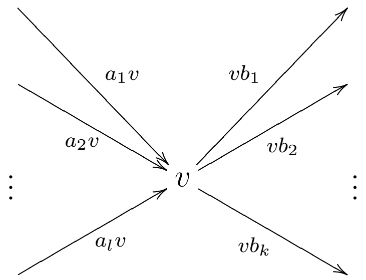
- for each right special vertex v in $G_n(\mathbf{w})$ that is not left special, we make the following changes



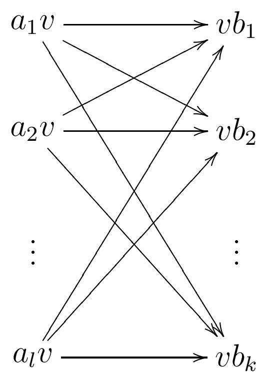
Transitions in G_n Transitions in G_{n+1}

- finally, for each bispecial vertex v in $G_n(\mathbf{w})$, we have among the transitions in $G_{n+1}(\mathbf{w})$ represented here below, only those whose label a_ivb_j

is a factor of $\mathbf{w}$.



Transitions in G_n



Possible transitions in G_{n+1}

Remark 1.5.6. It is a direct consequence of what precedes that for each non-negative integer n, if there is no bispecial factor in $L_n(\mathbf{w})$, then the Rauzy graph of order n determines exactly the Rauzy graph of order $n+1$. Moreover, in this case the length of the smallest path from a left special vertex to a right special vertex decreases by 1 as n increases by 1. Consequently, there exists a smallest non-negative integer k_n such that the Rauzy graph G_{n+k_n} contains a bispecial vertex v and we have to check which labels a_ivb_j belongs to $L(\mathbf{w})$ to construct the Rauzy graph $G_{n+k_n+1}(\mathbf{w})$.

1.5.3 Languages defined upon Rauzy graphs

It is possible to define languages upon Rauzy graphs. Indeed, given a Rauzy graph G_n, we can define the language $L_L(G_n)$ (resp. $L_R(G_n)$) respectively as the set of left labels (resp. right labels) of paths in G_n. In the sequel, we will mostly deal with minimal subshifts and uniformly recurrent sequences thus with strongly connected Rauzy graphs. In that case, the following results trivially holds true.

Fact 1.5.7. *Let (X,T) be a minimal subshift over an alphabet A. For all n, let $L_L(G_n)$ and $L_R(G_n)$ respectively denote the set of left labels and the set of right labels of all finite path in $G_n(X)$. Then, for all n, we have*

$$L_L(G_n) = L_R(G_n)$$

and

$$L(X) = \bigcap_{n\in\mathbb{N}} L_L(G_n).$$

Chapter 2

Overview of S-adicity

Recall that the S-adic conjecture states that *one can find a condition C such that a sequence has a sub-linear complexity if and only if it is an S-adic sequence satisfying the Condition C.*

In this chapter, we present some known results about the complexity of some S-adic sequences. First, we compare the case of (purely) morphic sequences with the case of S-adic sequences. Then we present some families of sequences (such as Sturmian sequences, Episturmian sequences, linearly recurrent sequences, codings of rotations,...) for which the S-adic representations are well known. Finally, we present some sufficient conditions for an S-adic sequence to have a sub-linear complexity and we give some examples that allow to reject some "naive ideas" that one could have about the Condition C.

2.1 Comparison between morphic and S-adic sequences

The complexity function of (purely) morphic sequences has already been extensively studied (see for instance [Cas97, Cas03, CN03, Dev08, Dur98a, ELR75, ER81, ER83, Fer95, NP09, Pan84]). In this section we present some known results about those sequences and we check whether they have a generalization for S-adic sequences. In most cases, they don't. Actually, even if things are already significantly more complicated for morphic sequences than for purely morphic ones, many results about purely morphic sequences can be generalized to morphic ones. On the opposite, we will see that very few of them still hold true for S-adic sequences.

2.1.1 The case of purely morphic sequences

Purely morphic sequences correspond to S-adic sequences with $\mathrm{Card}(S) = 1$. In that case, the complexity functions that can occur have been completely determined by Pansiot in [Pan84]. Indeed, he proved that for purely morphic sequences $\mathbf{w} = \sigma^\omega(a)$ with σ non-erasing, the complexity function $p_\mathbf{w}(n)$ can have only five asymptotic behaviors that are[1] $\Theta(1)$, $\Theta(n)$, $\Theta(n \log n)$, $\Theta(n \log\log n)$ and $\Theta(n^2)$. Moreover, when the sequence $\mathbf{w}$ is aperiodic, Theorem 1.2.2 implies that its complexity function cannot be $\Theta(1)$ and the class of complexity of the sequence only depends on the growth rate of images.

Definition 2.1.1. Recall that a morphism $\sigma : A^* \to A^*$ is said to be *everywhere growing* if it does not admit bounded letters (Definition 1.3.5 on page 31). Since for all letters a, we have $|\sigma^n(a)| \in \Theta(n^{\alpha_a}\beta_a^n)$ for some α_a in $\mathbb{N}$ and $\beta_a \geq 1$ (see [RS80]), any everywhere growing morphism satisfies exactly one of the following three definitions:

1. a morphism $\sigma : A^* \to A^*$ is *quasi-uniform* if there exists $\beta \geq 1$ such that for all letters $a \in A$, $|\sigma^n(a)| \in \Theta(\beta^n)$;

2. a morphism $\sigma : A^* \to A^*$ is *polynomially diverging* if there exists $\beta > 1$ and a function $\alpha : A \to \mathbb{N}$, $\alpha \neq 0$, such that for all letters $a \in A$, $|\sigma^n(a)| \in \Theta(n^{\alpha(a)}\beta^n)$;

3. a morphism $\sigma : A^* \to A^*$ is *exponentially diverging* if there exist $a_1, a_2 \in A$, $\alpha_1, \alpha_2 \in \mathbb{N}$ and $\beta_1, \beta_2 > 1$ with $\beta_1 \neq \beta_2$ such that for each $i \in \{1, 2\}$, $|\sigma^n(a_i)| \in \Theta(n^{\alpha_i}\beta_i^n)$.

Theorem 2.1.2 (Pansiot [Pan84])**.** *Let $\sigma : A^* \to A^*$ be a non-erasing morphism prolongable on $a \in A$ and let us consider the fixed point $\mathbf{w} = \sigma^\omega(a)$.*

1. *If σ is everywhere growing and*

   i. *quasi-uniform, then*[2] *$p_\mathbf{w}(n) \in O(n)$;*

   ii. *polynomially diverging, then $p_\mathbf{w}(n) \in \Theta(n \log\log n)$;*

   iii. *exponentially diverging, then $p_\mathbf{w}(n) \in \Theta(n \log n)$.*

2. *If σ is not everywhere growing and if there are infinitely many factors of $\mathbf{w}$ in $A^*_\mathfrak{B}$, then $p_\mathbf{w}(n) = \Theta(n^2)$.*

[1] $f(n) \in \Theta(g(n))$ if $\exists C_1, C_2 > 0, n_0\ \forall n \geq n_0\ |C_1 g(n)| \leq |f(n)| \leq |C_2 g(n)|$.
[2] $f(n) \in O(g(n))$ if $\exists C > 0, n_0\ \forall n \geq n_0\ |f(n)| \leq |Cg(n)|$.

3. *If σ is not everywhere growing and if there are only finitely many factors of $\mathbf{w}$ in $A_{\mathfrak{B}}^*$, then there exists an everywhere growing morphism $\tau : B^* \to B^*$ prolongable on $b \in B$ and a non-erasing morphism $\lambda : B^* \to A^*$ such that $\mathbf{w} = \lambda(\tau^\omega(b))$. In this case, we have $p_{\mathbf{w}}(n) \in \Theta(p_{\tau^\omega(b)}(n))$.*

One can regret that Theorem 2.1.2 only holds for non-erasing morphisms. However, the following result states that when the morphism is erasing, one can see the purely morphic sequence as a morphic sequence with non-erasing morphisms. The result is due to Cobham [Cob68] and has been recovered later by Pansiot [Pan83]. It can also be found in Cassaigne and Nicolas's survey [CN03].

Theorem 2.1.3 (Cobham [Cob68] and Pansiot [Pan83]). *If $\mathbf{w}$ is a morphic sequence, it is the image under a letter-to-letter morphism of a purely morphic word $\sigma^\omega(a)$ with σ a non-erasing morphism.*

In addition to the type of morphism (Definition 2.1.1), there exist some combinatorial criteria that have an influence on the complexity of purely morphic sequences. We give here three examples of such criteria — being uniformly recurrent, avoiding large powers and having a constant distribution — and will compare their consequences on the complexity for purely morphic, morphic and S-adic sequences. First, the following result concerns uniform recurrence and can be deduced from Theorem 2.1.2 (at least for everywhere growing morphisms). It can also be seen as a direct consequence of theorems 2.2.22 and 2.2.23.

Proposition 2.1.4. *Let $\mathbf{w} = \sigma^\omega(a)$ be a purely morphic sequence. If $\mathbf{w}$ is uniformly recurrent, then $p_{\mathbf{w}}(n) \in O(n)$. Moreover, if σ is everywhere growing, $\mathbf{w}$ is uniformly recurrent if and only if σ is primitive.*

Sketch of the proof: Let us prove it for everywhere growing morphisms. If $p_{\mathbf{w}}(n)$ is not sub-linear, there are two letters $b, c \in A$ such that the sequences $(|\sigma_n(b)|)_{n\in\mathbb{N}}$ and $(|\sigma_n(c)|)_{n\in\mathbb{N}}$ have different growth rates. Consequently, at least one of the following statements holds true:

1. for all n, the letter b does not occur in $\sigma^n(c)$;
2. for all n, the letter c does not occur in $\sigma^n(b)$.

Since all words $\sigma^n(b)$ and $\sigma^n(c)$ occur in $\mathbf{w}$ and since σ is everywhere growing, this implies that at least one of the letters b and c does not occur with bounded gaps in $\mathbf{w}$ and this contradict the uniform recurrence. □

Now, a sequence $\mathbf{w}$ is said to be *k-power free*, $k \geq 2$, if no factors of $\mathbf{w}$ can be written as u^k for some word $u \neq \varepsilon$. For instance, it is well known that the Thue-Morse sequence (see Example 1.3.1 on page 30) is cube-free.

Proposition 2.1.5 (Ehrenfeucht and Rozenberg [ER83]). *If* $\mathbf{w}$ *is a* k*-power free purely morphic sequence over an alphabet* A*, then its complexity function grows at least linearly and at most as* $n \log n$*. Moreover, if* $k = 2$ *and* $\mathrm{Card}(A) = 3$ *or if* $k \geq 3$ *and* $\mathrm{Card}(A) = 2$*, we have* $p_{\mathbf{w}}(n) \in \Theta(n)$*. Finally, there exist cube-free purely morphic sequences over 3 letters alphabets and square-free purely morphic sequences over 4 letters alphabets such that* $p(n) \in \Theta(n \log n)$*.*

Finally, a sequence is said to have a *constant distribution* if there is a length ℓ such that all factors of length ℓ of $\mathbf{w}$ contains all letters of $\mathbf{w}$. One can easily check that the Thue-Morse sequence $\mathbf{t}$ has a constant distribution (with $\ell = 3$). Actually, any k-power-free sequence over a binary alphabet has a constant distribution with $\ell = k + 1$.

Proposition 2.1.6 (Ehrenfeucht and Rozenberg [ER83]). *If* $\mathbf{w}$ *is purely morphic and has a constant distribution, then* $p_{\mathbf{w}}(n) \in O(n \log n)$*.*

Remark 2.1.7. In [Cas97], Cassaigne gave some methods (using bispecial factors) to compute the exact complexity of purely morphic sequences. In Chapter 4 of [BR10], he also shows how these methods can be used to compute the complexity of some particular S-adic sequences (see also the unpublished paper [Cas02]). However, the methods seem to be too much complicated to hope using them in a general case.

2.1.2 The case of morphic sequences

Theorems 2.1.2 and 2.1.3 show that to compute the complexity function of a purely morphic sequence, it is sometimes necessary to see it as a morphic sequence. It is therefore natural to be interested in the complexity function of such sequences. By definition, it is obvious that any purely morphic sequence is morphic. Next result shows that the converse is false.

Proposition 2.1.8 (Cassaigne and Nicolas [CN03]). *If* $\mathbf{w}$ *is a morphic sequence and if the letter* a *does not occur in* $\mathbf{w}$*, then the one-sided sequence* $aa\mathbf{w}$ *is morphic but not purely morphic.*

Moreover, not only the class of morphic sequences strictly contains the class of purely morphic sequences, but also the asymptotic behaviors of the complexity functions are different. Indeed, Example 2.1.9 shows that the classes of complexity given by Pansiot are not sufficient anymore.

Example 2.1.9 (Deviatov [Dev08]). Let $\mathbf{w}$ be the morphic sequence $\tau(\sigma^\omega(0))$ where σ and τ are defined by

$$\sigma : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 12 \\ 2 \mapsto 23 \\ 3 \mapsto 3 \end{cases} \quad \text{and} \quad \sigma : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \\ 3 \mapsto 2 \end{cases}$$

We have $p_{\mathbf{w}} \in \Theta(n\sqrt{n})$.

Other examples can be found in [Pan85]. Indeed, for all $k \geq 1$, Pansiot explicitly built a morphic sequence $\mathbf{w}$ whose complexity function satisfies $p_{\mathbf{w}}(n) \in \Theta(n \sqrt[k]{n})$. Consequently, the number of different asymptotic behaviors for the complexity function of morphic sequences is at least countably infinite. However, the behaviors $\Theta(n \sqrt[k]{n})$ seem to be the only new behaviors with respect to purely morphic sequences. Indeed, in [Dev08] Deviatov proved the next result and conjectured an equivalent result of Pansiot's Theorem (Theorem 2.1.2) for morphic sequences.

Theorem 2.1.10 (Deviatov [Dev08]). *Let $\mathbf{w}$ be a morphic sequence. Then, either $p_{\mathbf{w}}(n) \in \Theta(n^{1+\frac{1}{k}})$ for some $k \in \mathbb{N}^*$, or $p_{\mathbf{w}}(n) \in O(n \log n)$.*

Conjecture (Deviatov [Dev08]). *The complexity function of any morphic sequence only adopts one of the following asymptotic behaviors: $\Theta(1)$, $\Theta(n)$, $\Theta(n \log\log n)$, $\Theta(n \log n)$, $\Theta(n^{1+\frac{1}{k}})$ for some $k \in \mathbb{N}$.*

In particular, Theorem 2.1.10 implies that the highest complexity that one can get is the same for morphic sequences and for purely morphic sequences. This can be explained by the following result.

Proposition 2.1.11 (Cassaigne and Nicolas [CN03]). *Let $\mathbf{w}$ be a one-sided sequence over A and $\sigma : A^* \to B^*$ be a non-erasing morphism. If $M = \max_{a \in A} |\sigma(a)|$, for all n we have $p_{\sigma(\mathbf{w})}(n) \leq M p_{\mathbf{w}}(n)$. Moreover, if $\mathbf{w}$ is purely morphic and σ is injective, then $p_{\sigma(\mathbf{w})}(n) \in \Theta(p_{\mathbf{w}}(n))$.*

For purely morphic sequences, we have seen in Section 2.1.1 that some combinatorial criteria influence the complexity (uniform recurrence, k-power free and constant distribution). For morphic sequences, things are a little bit different.

First, the next result is rather similar to Proposition 2.1.4. It can be easily obtained with techniques developed by Durand in [Dur98a] and a detailed proof can also be found in [NP09].

Proposition 2.1.12. *If* $\mathbf{w} = \tau(\sigma^\omega(a))$ *is a uniformly recurrent morphic sequence with* τ *a letter-to-letter morphism, then* $p_{\mathbf{w}}(n) = \Theta(n)$.

Then, Proposition 2.1.5 can also be partially extended to morphic sequences (see Proposition 2.1.13 below).

Proposition 2.1.13 (Pansiot [Pan85])**.** *If* $\mathbf{w}$ *is a* k*-power free morphic sequence, then* $p_{\mathbf{w}}(n) \in O(n \log n)$.

Finally, Example 2.1.14 shows that Proposition 2.1.6 does not hold anymore for morphic sequences.

Example 2.1.14 (Pansiot [Pan85])**.** Let $\mathbf{w} = \tau(\sigma^\omega(a))$ be a morphism sequence where σ and τ are defined by

$$\sigma : \begin{cases} a \mapsto a1 \\ 1 \mapsto 01 \\ 0 \mapsto 0 \end{cases} \quad \text{and} \quad \tau : \begin{cases} a \mapsto 000 \\ 1 \mapsto 010 \\ 0 \mapsto 011 \end{cases}$$

Then $\mathbf{w}$ has a constant distribution and $p_{\mathbf{w}} \in \Theta(n^2)$.

2.1.3 The case of S-adic sequences

As we will see, S-adic sequences are considerably more complicated than morphic ones. Indeed, in this case the set of asymptotic behaviors of the complexity function is uncountably. Moreover, the combinatorial criteria given for (purely) morphic sequences (uniform recurrence, k-power free, constant distribution) do not influence the complexity anymore.

Nothing works fine

A first important result is the following.

Proposition 2.1.15 (Cassaigne [Fog11])**.** *Let* A *be an alphabet and* $l \notin A$*. There is a finite set* S *of morphisms over* $A' = A \cup \{l\}$ *such that any one-sided sequence over* A *is* S*-adic.*

Proof. Let $\mathbf{w} = \mathbf{w}_0\mathbf{w}_1 \cdots$ be a one-sided sequence over a finite alphabet A and let l be a letter that does not belong to A. For each letter a in A we define the morphism σ_a over $A \cup \{l\}$by

$$\sigma_a : \begin{cases} l \mapsto la \\ b \mapsto b \text{ if } b \neq l \end{cases}$$

We also define the morphism ϕ from $A \cup \{l\}$ to A by

$$\phi : \begin{cases} l \mapsto \mathbf{w}_0 \\ b \mapsto b \text{ if } b \neq l \end{cases}$$

Then we have

$$\mathbf{w} = \lim_{n \to +\infty} \phi \sigma_{\mathbf{w}_1} \sigma_{\mathbf{w}_2} \cdots \sigma_{\mathbf{w}_n}(l^\omega).$$

□

This results can be extended to two-sided sequences as follows.

Corollary 2.1.16. *Let A be an alphabet and $l, k \notin A$. There is a finite set S of morphisms over $A' = A \cup \{l, k\}$ such that any one-sided sequence over A is S-adic.*

Proof. Indeed, consider $\mathbf{w} = \cdots \mathbf{w}_{-2}\mathbf{w}_{-1}.\mathbf{w}_0\mathbf{w}_1 \cdots \in A^{\mathbb{Z}}$. Let us consider the morphisms defined in the proof of Proposition 2.1.15 and the following morphisms

$$\psi : \begin{cases} k \mapsto \mathbf{w}_{-1} \\ b \mapsto b \text{ if } b \neq k \end{cases} \qquad \text{and} \qquad \forall a \in A \ \tau_a : \begin{cases} k \mapsto ak \\ b \mapsto b \text{ if } b \neq k \end{cases} .$$

Then we have

$$\mathbf{w} = \lim_{n \to +\infty} \phi \psi \sigma_{\mathbf{w}_1} (\tau_{\mathbf{w}_{-2}} \sigma_{\mathbf{w}_2})(\tau_{\mathbf{w}_{-3}} \sigma_{\mathbf{w}_3}) \cdots (\tau_{\mathbf{w}_{-n}} \sigma_{\mathbf{w}_n})({}^\omega k.l^\omega).$$

□

In particular, this implies that one can get any high complexity with S-adic sequences, which is strongly different from what can be observed for morphic sequences. Moreover, the following proposition implies that the set of possible asymptotic behaviors for the complexity function of S-adic sequences is uncountable.

Proposition 2.1.17 (Cassaigne [Cas03])**.** *Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a function such that*

i. $\lim_{t \to +\infty} \frac{f(t)}{\log t} = +\infty$;

ii. *f is differentiable, except possibly at* 0;

iii. $\lim_{t \to +\infty} f'(t) t^\beta = 0$ *for some* $\beta > 0$;

iv. *f' is decreasing.*

Then there exists a uniformly recurrent sequence $\mathbf{w}$ *over* $\{0,1\}$ *such that* $\log(p_{\mathbf{w}}(n)) \sim f(n)$.

Moreover, the proof of this proposition is constructive (see also [MM10] for other constructions of sequences with complexity close to a given function). In particular, the function $f(n)$ in the previous proposition can be taken equal to n^α for any α with $0 < \alpha < 1$.

For (purely) morphic sequences, Propositions 2.1.4 and 2.1.12 imply that for uniformly recurrent (purely) morphic sequences $\mathbf{w}$, we have $p_{\mathbf{w}}(n) \in O(n)$. For S-adic sequences, the following theorem together with Proposition 2.1.15 imply that this is not true anymore. Recall that the *topological entropy* of a sequence (or a subshift) over an alphabet A is the real number h with $0 \leq h \leq \log(\mathrm{Card}(A))$ defined by

$$h = \lim_{n\to\infty} \frac{\log(p(n))}{n}.$$

Observe that h is well defined due to Fekete's Lemma (see [Fek23]) and to the inequality $p(m+n) \leq p(m)p(n)$ (see Chapter 4 of [BR10] for a proof).

Theorem 2.1.18 (Grillenberger [Gri73])**.** *Let A be an alphabet with $d = \mathrm{Card}(A) \geq 2$ and $h \in [0, \log(d)[$. There exists a uniformly recurrent one-sided sequence* $\mathbf{w}$ *over A with topological entropy h.*

For the other combinatorial criteria (k-power free and constant distribution), it is clear that Proposition 2.1.6 does not hold for S-adic sequences since it does not even hold for morphic ones (see Example 2.1.14). For the last one (avoiding large powers) Proposition 2.1.19 shows that, once again, nothing works fine for S-adic sequences (the proof follows a construction of Currie and Rampersad in [CR10]).

Proposition 2.1.19. *There exist some uniformly recurrent S-adic sequences that are cubefree and have an exponential complexity.*

Proof. From Proposition 2.1.15, we only have to prove the existence of uniformly recurrent sequences that are cubefree and have an exponential complexity. First, let us give some definitions. Given two sequences $\mathbf{x} \in A^{\mathbb{N}}$ and $\mathbf{y} \in B^{\mathbb{N}}$, the *direct product* of $\mathbf{x}$ and $\mathbf{y}$ is the sequence $\mathbf{x} \otimes \mathbf{y} \in (A \times B)^{\mathbb{N}}$ such that $(\mathbf{x} \otimes \mathbf{y})_i = (\mathbf{x}_i, \mathbf{y}_i)$ for all $i \in \mathbb{N}$. Then, we say that a sequence $\mathbf{x}$ is *strongly recurrent* if for all uniformly recurrent sequences $\mathbf{y}$, the sequence $\mathbf{x} \otimes \mathbf{y}$ is uniformly recurrent. In [Sal10], Salimov proved (in particular) that the Thue-Morse sequence $\mathbf{t}$ is strongly recurrent.

Now let us complete the proof. From Theorem 2.1.18 we can consider a uniformly recurrent sequence $\mathbf{x}$ with exponential complexity. Let also $\mathbf{t}$ be the Thue-Morse sequence and define $\mathbf{w} = \mathbf{t} \otimes \mathbf{x}$. It is obvious that $\mathbf{w}$ is also cubefree and has an exponential complexity so the result holds. □

Remark 2.1.20. In the proof of the previous result, one has to consider sequences over at least 4-letters alphabets. This can be improved to 2-letters alphabets by replacing the direct product by the perfect shuffle, i.e., $\text{Shuffle}(\mathbf{x}, \mathbf{y}) = \mathbf{x}_0\mathbf{y}_0\mathbf{x}_1\mathbf{y}_1\mathbf{x}_2\mathbf{y}_2\cdots$.

Nevertheless, two positive results

Up to now, it seems that, compared to (purely) morphic sequences, nothing works with S-adic sequences. However, we can still give the following two results. The first one deals with uniform recurrence and the second one is a generalization of Theorem 2.1.3.

Proposition 2.1.21 (Durand [Dur00]). *Let $\mathbf{w}$ be an S-adic sequence directed by the couple $(\sigma_n, a_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}} \times \prod_{n=0}^{\infty} A_n$. If for all $r \in \mathbb{N}$, there exists $s > r$ such that all letters of A_r occur in $\sigma_r \cdots \sigma_s(a)$ for all a in A_{s+1}, then $\mathbf{w}$ is uniformly recurrent.*

Proposition 2.1.22 (Cassaigne [Fog11]). *Let $\mathbf{w}$ be an S-adic sequence directed by the couple $(\sigma_n, a_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}} \times \prod_{n=0}^{\infty} A_n$. There exists a set $\tilde{S}$ of morphisms and an $\tilde{S}$-adic representation $(\tilde{\sigma}_n : \tilde{A}_{n+1}^* \to \tilde{A}_n^*, \tilde{a}_n)_{n\in\mathbb{N}} \in \tilde{S}^{\mathbb{N}} \times \prod_{n=0}^{\infty} \tilde{A}_n$ of $\mathbf{w}$ such that for all n:*

1. *$\tilde{A}_n \subset A_n$ and*
2. *$\tilde{\sigma}_n$ is non-erasing.*

Moreover, if S is finite, so is $\tilde{S}$.

2.2 Some well-known S-adic representations

In the literature, some results are already well known about S-adicity. For instance, some families of sequences admit an S-adic characterization. In other words, *there is a condition C* for those sequences. The most famous class of sequences that admit an S-adic characterization is the class of Sturmian sequences. As already mentioned, these sequences have been widely studied. In particular, they have been generalized into several directions (such as codings of rotations, codings of interval exchange transformations, episturmian sequences) and some of these generalizations also yield to S-adic characterizations. In this section we (partially) present what is known about these ones.

2.2.1 Sturmian sequences

Recall that a one-sided sequence $\mathbf{w}$ is *Sturmian* if $p_{\mathbf{w}}(n) = n + 1$ for all n. Example 1.3.7 (on page 32) shows that any Sturmian sequence is S-adic with $S = \{L_0, L_1, R_0, R_1\}$. Using another definition of Sturmian sequences (using codings of rotations), one can actually say more (see Theorem 2.2.1 below).

Theorem 2.2.1 (Berthé, Holton, Zamboni [BHZ06]). *Let S be the set of morphisms $\{L_0, R_0, L_1, R_1\}$ as in Example 1.3.7. A sequence $\mathbf{w}$ is Sturmian if and only if there exist two sequences of integers $(a_k)_{k\geq 1}$ and $(c_k)_{k\geq 1}$ such that*

$$\mathbf{w} = \lim_{n\to\infty} L_0^{a_1-1-c_1} R_0^{c_1} L_1^{a_2-c_2} R_1^{c_2} \cdots L_0^{a_{n-1}-1-c_{n-1}} R_0^{c_{n-1}} L_1^{a_n-c_n} R_1^{c_n}(0^\omega).$$

with for all $k \geq 1$, $a_k \geq 1$, $c_k \geq 0$ and for all $k \geq 2$, $c_k = a_k \Rightarrow c_{k-1} = 0$. Moreover, two different couples of sequences $(a_k, c_k)_{k\geq 1}$ satisfying the above conditions provide two differents Sturmian sequences.

To briefly explain that theorem, it is convenient to see Sturmian sequences as codings of rotations (or as mechanical words, but this direction will not be followed here). Indeed, as we will see, the sequence $(a_k)_{k\geq 1}$ is related to the *continued fraction* of the angle α and the sequence $(c_k)_{k\geq 1}$ is related to the *Ostrowski representation* [Ost22] of the point $x \in \mathbb{R}/\mathbb{Z}$ whose orbit under the rotation is coded by the Sturmian sequence. First, let us recall the link between Sturmian sequences and codings of rotations.

Formally, for $\alpha \in \mathbb{R}$, the *rotation of angle α* is $r_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by

$$r_\alpha(x) = x + \alpha \mod 1.$$

As already mentioned in Example 1.4.1, the couple $(\mathbb{R}/\mathbb{Z}, r_\alpha)$ is a topological dynamical system.

A one-sided sequence $\mathbf{w}$ over $A = \{0, \ldots, k-1\}$ is a *coding of the rotation* r_α if there exists $x \in \mathbb{R}/\mathbb{Z}$ and a partition $\mathcal{P}$ of the unit circle $\mathbb{R}/\mathbb{Z}$ into k intervals $\{I_0, I_1, \ldots, I_{k-1}\}$ such that for all $k \in \mathbb{N}$,

$$\mathbf{w}_k = i \quad \text{if } r_\alpha^k(x) \in I_i.$$

The set of codings of rotations of r_α with respect to a partition $\mathcal{P}$ is a subshift denoted by $(X_{\alpha,\mathcal{P}}, T)$.

We can suppose without loss of generality that α belongs to $]0, \frac{1}{2}[$. Indeed, when $\alpha > 1$ or $\alpha < 0$, the dynamical system $(\mathbb{R}/\mathbb{Z}, r_\alpha)$ is the same as $(\mathbb{R}/\mathbb{Z}, r_\beta)$ with $\beta \in]0, 1[$ and $\beta \equiv \alpha \mod 1$ and for $\alpha \in]\frac{1}{2}, 1[$ we have to consider the rotation in the opposite direction of angle $\alpha' = 1 - \alpha$. In the sequel we will always suppose that α is irrational, otherwise the orbit of any point $x \in [0, 1[$ under r_α is periodic (so is its coding).

Proposition 2.2.2 ([MH40] and [CH73]). *A sequence* $\mathbf{w}$ *is Sturmian if and only if there is an irrational number* α *and a point* $x \in \mathbb{R}/\mathbb{Z}$ *such that* $\mathbf{w}$ *is the coding of* x *under the rotation* r_α *with respect to the partition* $\{I_0, I_1\}$ *of* $\mathbb{R}/\mathbb{Z}$ *with* $I_0 = [0, 1-\alpha[$ *and* $I_1 = [1-\alpha, 1[$.

Now let us explain how we can obtain exactly the directive word of the S-adic representation of a Sturmian sequence.

For $\alpha \in \mathbb{R}$, let $[a_0; a_1, a_2, \dots]$ denote its *regular continued fraction*, i.e.,

$$\alpha = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_n + \cdots}}}, \quad a_i \in \mathbb{N}^*.$$

Observe that the continued fraction of α is finite if and only if α is rational. Moreover, we have $a_0 = \lfloor \alpha \rfloor$ and for all $n \geq 1$, a_n is called *partial quotient* and is obtained as follows. We define the *regular continued fraction operator* $F : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ by

$$F(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor.$$

Then, to find the partial quotients of α, we put $f_0 = \alpha - \lfloor \alpha \rfloor$ and for all $n \geq 1$, $f_n = F(f_{n-1})$. For all $n \geq 1$ we have

$$a_n = \left\lfloor \frac{1}{f_{n-1}} \right\rfloor.$$

Lemma 2.2.3 (Morse and Hedlund [MH40]). *Let* $\mathbf{w}$ *be a Sturmian sequence related to the rotation* r_α. *The sequence* $(a_k)_{k\geq 1}$ *of Theorem 2.2.1 is the sequence of partial quotients of* α.

Now, for all $n \in \mathbb{N}$ let us define p_n and q_n by $\gcd(p_n, q_n) = 1$ and

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_n}}}. \tag{2.1}$$

We have $\lim_{n\to+\infty} \frac{p_n}{q_n} = \alpha$ and the sequence $\left(\frac{p_n}{q_n}\right)_{n\in\mathbb{N}}$ represents the best approximation of α, i.e., for all rational number $\frac{r}{s}$ with $\gcd(r, s) = 1$, we have

$$\forall n, \quad \frac{r}{s} \neq \frac{p_n}{q_n} \text{ and } 0 < s \leq q_n \Rightarrow |s\alpha - r| > |q_n\alpha - p_n|.$$

Then, we consider a particular numeration system based on the sequence $(|q_n\alpha - p_n|)_{n\in\mathbb{N}}$ called *Ostrowski numeration system* (see [Ost22, Ber01]).

Proposition 2.2.4. *Let $\alpha = [0; a_1, a_2, \dots]$ be irrational. Any real number $x \in \mathbb{R}/\mathbb{Z}$ can be uniquely written as*

$$x = \sum_{n=1}^{+\infty} c_n |q_{n-1}\alpha - p_{n-1}|, \tag{2.2}$$

with

i. p_n *and* q_n *as in* (2.1)*;*

ii. for all n, $0 \leq c_n \leq a_n$*;*

iii. $c_n = 0 \Rightarrow c_{n+1} = a_{n+1}$*;*

iv. for infinitely many n, $c_{2n} \neq a_{2n}$ *and* $c_{2n+1} \neq a_{2n+1}$.

Lemma 2.2.5 (Berthé, Holton, Zamboni [BHZ06])**.** *Let* $\mathbf{w}$ *be a Sturmian sequence related to the rotation* r_α *and to the point* $x \in \mathbb{R}/\mathbb{Z}$*. The sequence* $(c_k)_{k\geq 1}$ *of Theorem 2.2.1 is the sequence of coefficients of the Ostrowski representation of* x *in Proposition 2.2.*

Remark 2.2.6. Let $\mathbf{w}$ be a Sturmian sequence coding the orbit of x under the rotation r_α, $\alpha = [0; a_1, a_2, \dots]$. The equidistribution of $(x + n\alpha \mod 1)_{n\in\mathbb{N}}$ in $\mathbb{R}/\mathbb{Z}$ implies that the subshift $(X_\mathbf{w}, T)$ is minimal and that the sequence $\mathbf{x}$ coding the orbit of 0 is in $X_\mathbf{w}$. Then, since the sequence of coefficient $(c_n)_{n\geq 1}$ of the Ostrowski representation of 0 is only composed of zero's, the sequence $\mathbf{x}$ is directed $L_0^{a_1-1} L_1^{a_2} L_0^{a_3} L_1^{a_4} \cdots$. Finally, since by definition, an S-adic representation of $X_\mathbf{w}$ is given by any S-adic representation of a sequence of $X_\mathbf{w}$, the subshift $X_\mathbf{w}$ admits $L_0^{a_1-1} L_1^{a_2} L_0^{a_3} L_1^{a_4} \cdots$ as an S-adic representation.

2.2.2 Codings of rotations

A natural way to generalize Sturmian sequences is to consider codings of rotations of irrational angle α but with a different partition of the unit circle $\mathbb{R}/\mathbb{Z}$. In the sequel we will only consider *non-degenerate* rotations, i.e., rotations of irrational angle α such that there are some real numbers $l_0, \dots, l_k$ verifying $0 = l_0 < l_1 < \cdots < l_k = 1$ and for all $i \in \{0, \dots, k-1\}$, we have $I_i = [l_i, l_{i+1}[$ and $l_{i+1} - l_i \geq \alpha$.

Let α and β be irrational numbers, $\alpha, \beta \in \mathbb{R}/\mathbb{Z}$. For any $x \in \mathbb{R}/\mathbb{Z}$, we consider the coding of rotation $\mathbf{w} \in \{0,1\}^\mathbb{N}$ defined by

$$\mathbf{w}_k = \begin{cases} 0 & \text{if } r_\alpha^k(x) \in [0, 1-\beta[\\ 1 & \text{if } r_\alpha^k(x) \in [1-\beta, 1[\end{cases}$$

and call it the *rotation of parameters* (α, β). Those codings of rotations are strongly related to Sturmian sequences, as shown by Proposition 2.2.9. First we need to define the notion of *sliding block code*.

Definition 2.2.7. Let A and B be alphabets. A *block map* is an application $\Phi : A^{m+n+1} \to B$ with $m, n \in \mathbb{Z}$, $m + n \geq 0$. A *sliding block code* is a map ϕ from $A^{\mathbb{Z}}$ to $B^{\mathbb{Z}}$ (or from $A^{\mathbb{N}}$ to $B^{\mathbb{N}}$) for which there exist two integers m and n, $-m \leq n$, and a block map $\Phi : A^{m+n+1} \to B$ such that for all sequences $\mathbf{w}$ over A and all i,

$$(\phi(\mathbf{w}))_i = \Phi(\mathbf{w}_{[i-m,i+n]}).$$

Obviously, for one-sided sequences, we must have $m = 0$ and $n \geq 0$.

Example 2.2.8. Let $\mathbf{t}$ be the Thue-Morse sequence (see Example 1.3.1 on page 30) and let f_0 and f_1 be the sliding block codes respectively related to the block maps F_0 and F_1 of Proposition 2.2.9 below. We have

$$\begin{aligned} f_0(\mathbf{t}) &= F_0(01)F_0(11)F_0(10)F_0(01)F_0(10)F_0(00)F_0(01)F_0(11)F_0(10)\cdots \\ &= 001010001\cdots \end{aligned}$$

and

$$\begin{aligned} f_1(\mathbf{t}) &= F_1(01)F_1(11)F_1(10)F_1(01)F_1(10)F_1(00)F_1(01)F_1(11)F_1(10)\cdots \\ &= 100100100\cdots \end{aligned}$$

Proposition 2.2.9 (Didier [Did98b])**.** *A sequence $\mathbf{w} \in \{0,1\}^{\mathbb{N}}$ codes a non-degenerate rotation of parameters (α, β) if and only if the sequences $f_0(\mathbf{w})$ and $f_1(\mathbf{w})$ are Sturmian, where f_0 and f_1 are the sliding block codes related to the block maps F_0 and F_1 defined by*

$$F_0 : \begin{cases} 00 \mapsto 0 \\ 01 \mapsto 0 \\ 10 \mapsto 1 \\ 11 \mapsto 0 \end{cases} \quad \textit{and} \quad F_1 : \begin{cases} 00 \mapsto 0 \\ 01 \mapsto 1 \\ 10 \mapsto 0 \\ 11 \mapsto 0 \end{cases}$$

Rote also proved in [Rot94] that these sequences have a sub-linear complexity.

Proposition 2.2.10 (Rote [Rot94])**.** *Let $\mathbf{w} \in \{0,1\}^{\mathbb{N}}$ be a coding of rotation of parameters (α, β). For all n we have $p_{\mathbf{w}}(n) \leq 2n$ and if for all $k \in \mathbb{N}$, $k\alpha \neq \beta \mod 1$, then $p_{\mathbf{w}}(n) = 2n$.*

Finally, Didier proved in [Did98a] that the subshifts generated by those sequences are S-adic for a particular set S.

Theorem 2.2.11 (Didier [Did98a]). *Let $\mathbf{w} \in \{0,1\}^{\mathbb{N}}$ be a coding of a non-degenerate rotation of parameters (α, β) and let $(i_n)_{n\in\mathbb{N}}$ and $(\alpha_n, \beta_n)_{n\in\mathbb{N}}$ be the sequences defined by $(\alpha_0, \beta_0) = (\alpha, \beta)$ and for all n,*

1. *$i_n = 0$ and $(\alpha_{n+1}, \beta_{n+1}) = \left(1 - \left\{\frac{1}{\alpha_n}\right\}, \left\{\frac{\beta_n}{\alpha_n}\right\}\right)$ if $\left\{\frac{1}{\alpha_n}\right\} \geq \left\{\frac{\beta_n}{\alpha_n}\right\}$ and*
2. *$i_n = 1$ and $(\alpha_{n+1}, \beta_{n+1}) = \left(\left\{\frac{1}{\alpha_n}\right\}, 1 - \left\{\frac{\beta_n}{\alpha_n}\right\}\right)$ otherwise.*

The subshift $(X_{\mathbf{w}}, T)$ admits the following S-adic representation

$$\lim_{n\to+\infty} \phi \zeta_A^{\left\lfloor \frac{1-\beta_0}{\alpha_0} \right\rfloor -1} \zeta_B^{\left\lfloor \frac{\beta_0}{\alpha_0} \right\rfloor -1} \sigma_{i_0} \zeta_A^{\left\lfloor \frac{1-\beta_1}{\alpha_1} \right\rfloor -1} \zeta_B^{\left\lfloor \frac{\beta_1}{\alpha_1} \right\rfloor} \sigma_{i_1} \cdots \zeta_A^{\left\lfloor \frac{1-\beta_n}{\alpha_n} \right\rfloor -1} \zeta_B^{\left\lfloor \frac{\beta_n}{\alpha_n} \right\rfloor} \sigma_{i_n} \zeta_A^{\left\lfloor \frac{1-\beta_{n+1}}{\alpha_{n+1}} \right\rfloor -1} \zeta_B^{\left\lfloor \frac{\beta_{n+1}}{\alpha_{n+1}} \right\rfloor} (bd)$$

where ϕ, ζ_A, ζ_B, σ_0 and σ_1 are defined by

$$\zeta_A : \begin{cases} a \mapsto a \\ b \mapsto bc \\ c \mapsto c \\ d \mapsto d \end{cases} \qquad \zeta_B : \begin{cases} a \mapsto a \\ b \mapsto b \\ c \mapsto c \\ d \mapsto da \end{cases}$$

$$\phi : \begin{cases} a \mapsto 1 \\ b \mapsto 1 \\ c \mapsto 0 \\ d \mapsto 0 \end{cases} \qquad \sigma_0 : \begin{cases} a \mapsto bda \\ b \mapsto b \\ c \mapsto dbc \\ d \mapsto d \end{cases} \qquad \sigma_1 : \begin{cases} a \mapsto bcd \\ b \mapsto bc \\ c \mapsto dab \\ d \mapsto da \end{cases}$$

Remark 2.2.12. Observe that in [Did98b], Didier actually gave a generalization of Proposition 2.2.9 for codings of rotation over arbitrarily large alphabets. Moreover, he ensured in [Did98a] that Theorem 2.2.11 can also be extended to them.

Remark 2.2.13. One can also note that, similarly to the Sturmian case, the S-adic representation given in Theorem 2.2.11 is related to continued fractions. Indeed, for irrational numbers α, the sequence $(\alpha_n, \beta_n)_{n\in\mathbb{N}}$ of Theorem 2.2.11 is infinite. If for all $k \in \mathbb{N}$ we define the integers a_k and b_k by

$$(a_k, b_k) = \begin{cases} \left(\left\lfloor \frac{1}{\alpha_k} \right\rfloor + 1, \left\lfloor \frac{\beta_k}{\alpha_k} \right\rfloor\right) & \text{if } i_k = 0 \\ \\ \left(\left\lfloor \frac{1}{\alpha_k} \right\rfloor, \left\lfloor \frac{\beta_k}{\alpha_k} \right\rfloor + 1\right) & \text{if } i_k = 1 \end{cases},$$

we get

$$\alpha_k = \frac{1}{a_k + (-1)^{i_k+1}\alpha_{k+1}} \quad \text{and} \quad \beta_k = \alpha_k b_k + (-1)^{i_k}\alpha_k\beta_{k+1}.$$

This provides a kind of *generalized continued fraction* of (α, β) which allows to write

$$\alpha = \frac{1}{a_0 + \frac{(-1)^{i_0+1}}{a_1 + \frac{(-1)^{i_1+1}}{\ddots}}}$$

and

$$\beta = \sum_{n=0}^{\infty} \left((-1)^{\sum_{k=0}^{n-1} i_k} b_n \prod_{k=0}^{n} \alpha_k \right).$$

2.2.3 Codings of interval exchange transformations

Another generalization of Sturmian sequences are the codings of interval exchange transformations. Indeed, for Sturmian sequences, the action of the rotation r_α (see Section 2.2.1) on the intervals I_0 and I_1 is simply a permutation of them (Figure 2.1). Interval exchange transformations (IET) have first been introduced by Oseledec [Ose66] (see also [KS67]) and have been extensively studied since then (see for instance [Ada02, AR91, Did97, FHZ01, FHZ03, FHZ04, Kea75, LN98, LN00, LN01, Rau79, Vee84a, Vee84b, Vee84c] or [Via06] for a survey)

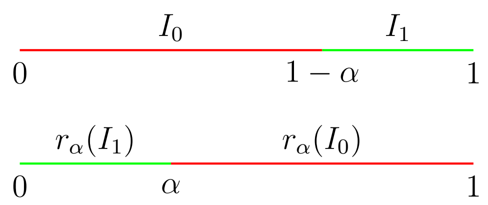


Figure 2.1: For Sturmian sequences, the action of r_α on I_0 and I_1 is simply a permutation.

Generalities

Let $\lambda = (\lambda_0, \ldots, \lambda_{k-1})$ be a k-dimensional positive vector ($k \geq 2$) such that $\sum_{j=0}^{k-1} \lambda_j = 1$ and let π be a permutation of $\{0, \ldots, k-1\}$. For all $i \in \{0, \ldots, k-1\}$, we let I_i denote the semi-interval $[b_i, b_{i+1}[$ with $b_i = \sum_{j<i} \lambda_j$.

A *k-interval exchange transformation* is a function $T_{\lambda,\pi} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ such that for all $i \in \{0, \ldots, k-1\}$ and all $x \in I_i$,

$$T_{\lambda,\pi}(x) = x - \sum_{j<i} \lambda_j + \sum_{\pi(j)<\pi(i)} \lambda_{\pi(j)}.$$

For $k = 3$ and $\lambda = \left(\frac{1}{3}, \frac{1}{2}, \frac{1}{6}\right)$ and $\pi = (2, 1, 0)$, the action of $T_{\lambda,\pi}$ on $\mathbb{R}/\mathbb{Z}$ is represented at Figure 2.2. With the same distance d as for rotations (Section 2.2.2), the couple $(\mathbb{R}/\mathbb{Z}, T_{\lambda,\pi})$ is a topological dynamical system.

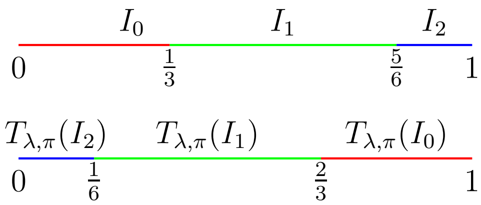


Figure 2.2: 3-IET with $\lambda = \left(\frac{1}{3}, \frac{1}{2}, \frac{1}{6}\right)$ and $\pi = (3, 2, 1)$.

Remark 2.2.14. Since [Kea75] it is well-known that rotations are closely linked to 3-IET (see also [Ada02] for a detailed proof). In particular, Adamczewski proved in [Ada02] that codings of rotations can be obtained as images by a morphisms of S-adic sequences where S contains four morphisms over $\{0, 1, 2\}$.

Let $A = \{0, \ldots, k-1\}$. A sequence $\mathbf{w} \in A^{\mathbb{N}}$ is the *coding of a k-interval exchange transformation* if there is a *k-interval exchange transformation* $T_{\lambda,\pi}$ and a point $x \in \mathbb{R}/\mathbb{Z}$ such that for all $j \in \mathbb{N}$,

$$\mathbf{w}_j = i \quad \text{if } T^j_{\lambda,\pi}(x) \in I_i.$$

Minimality and i.d.o.c.

A k-IET $T_{\lambda,\pi}$ is said to be *irreducible* if its permutation π is irreducible, i.e., if for all $j \in \{0, \ldots, k-2\}$, one has $\pi(\{0, \ldots, j\}) \neq \{0, \ldots, j\}$. One also says that $T_{\lambda,\pi}$ satisfies the *infinite distinct orbit condition (i.d.o.c.)* if the $k-1$ negative trajectories $\{T^{-n}_{\lambda,\pi}(b_i) \mid n \in \mathbb{N}\}$, $1 \leq i \leq k-1$, are infinite disjoint sets.

Proposition 2.2.15 (Keane [Kea75])**.** *If $T_{\lambda,\pi}$ is irreducible and satisfies the i.d.o.c., then $(\mathbb{R}/\mathbb{Z}, T_{\lambda,\pi})$ is minimal.*

A k-IET $T_{\lambda,\pi}$ is said to be *irrational* if λ is rationally independent, i.e., for all non-zero integer vectors $(n_0, n_1, \ldots, n_{k-1}) \in \mathbb{Z}^k$ we have $\sum_{i=0}^{k-1} n_i \lambda_i \neq 0$.

Proposition 2.2.16 (Keane [Kea75])**.** *If $T_{\lambda,\pi}$ is irreducible and irrational, then it satisfies the i.d.o.c.*

It is well known that for codings of irreducible k-IET satisfying the i.d.o.c., we have $p(n) = (k-1)n+1$ for all n (see for instance [AR91]) and the converse is also true. It is also well known that they are S-adic. There actually exist several ways to obtain an S-adic representation of these codings. The most famous is probably using the *Rauzy induction* (see [Rau79]).

In [FZ08], the authors gave a combinatorial characterization of codings of irreducible k-IET satisfying the i.d.o.c. (see also [KBC10]). Then, completing a work initialized in [FHZ01, FHZ03, FHZ04], they also provided another induction process in [FZ10] for symmetric k-IET, i.e., IET such that the permutation π is defined by $\pi(i) = k-1-i$ for all $i \in \{0,\dots,k-1\}$. In particular, this provided another S-adic representation of these IET. For instance, for $k = 3$, if $T_{(\alpha,\beta)}$ denotes the symmetric 3-IET with $\lambda = (\alpha, \beta, 1-\alpha-\beta)$, they obtain the following S-adic representation using return words.

Theorem 2.2.17 (Ferenczi, Holton and Zamboni [FHZ03])**.** *Let $(\alpha,\beta) \in\,]0,1[^2$ such that $2\alpha < 1$ and $2\alpha + \beta > 1$ and such that $(\mathbb{R}/\mathbb{Z}, T_{(\alpha,\beta)})$ is a symmetric 3-IET satisfying the i.d.o.c. Let also $(X_{(\alpha,\beta)}, T)$ be the corresponding subshift and for $m,n \in \mathbb{N}$, we let $\sigma_{(0,m,n)}$ and $\sigma_{(1,m,n)}$ respectively denote the morphisms*

$$\sigma_{(0,m,n)} : \begin{cases} 0 \mapsto 0^m 2 1^{n-1} \\ 1 \mapsto 1 0^{m-1} 2 1^{n-1} \\ 2 \mapsto 0^{m-1} 2 1^{n-1} \end{cases} \quad \text{and} \quad \sigma_{(1,m,n)} : \begin{cases} 0 \mapsto 1 0^{m-1} 2 1^{n-1} \\ 1 \mapsto 0^m 2 1^{n-1} \\ 2 \mapsto 1 0^m 2 1^{n-1} \end{cases} .$$

Then, there exist two sequences of positive integers $(m_k)_{k\in\mathbb{N}}$ and $(n_k)_{k\in\mathbb{N}}$ and a sequence $(i_k)_{k\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ such that $(\sigma_{(i_k,m_k,n_k)})_{k\in\mathbb{N}}$ is an S-adic representation of $(X_{(\alpha,\beta)}, T)$.

Observe that both morphisms $\sigma_{(0,m,n)}$ and $\sigma_{(1,m,n)}$ are actually compositions of the following four morphisms so this provides an S-adic representation with $\mathrm{Card}(S) = 4$.

$$D : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases} \qquad G : \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases}$$

$$E_{0,1} : \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases} \qquad E_{1,2} : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 2 \\ 2 \mapsto 1 \end{cases} .$$

Moreover, if (α, β) does not satisfy the conditions $2\alpha < 1$ and $2\alpha + \beta > 1$, there exists $(\bar{\alpha}, \bar{\beta})$ satisfying them and a finite sequence of integers $l_0, \dots, l_k$ such that

$$X_{(\alpha,\beta)} = \sigma^{l_0} E_{0,2} \sigma^{l_1} E_{0,2} \cdots \sigma^{l_{k-1}} E_{0,2} \sigma^{l_k} \left(X_{(\bar{\alpha},\bar{\beta})} \right),$$

where σ and $E_{0,2}$ are defined by

$$\sigma : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 10 \\ 2 \mapsto 20 \end{cases} \quad \text{and} \quad E_{0,2} : \begin{cases} 0 \mapsto 2 \\ 1 \mapsto 1 \\ 2 \mapsto 0 \end{cases}.$$

For other S-adic representations, see also [LN98, LN00, LN01].

Another class of sequences related to IET is the class of *Arnoux-Rauzy sequences.* They are defined as the uniformly recurrent sequences such that $p(n) = (k-1)n + 1$ for all n and such that for all n, there is a unique right special factor r and a unique left special factor l such that $\delta^+(r) = \delta^-(l) = k$. The link with IET is the following.

Proposition 2.2.18 (Arnoux and Rauzy [AR91]). *Let $\mathbf{w} \in \{0, \dots, k-1\}^{\mathbb{N}}$ be an Arnoux-Rauzy sequence. There exists a point $x \in \mathbb{R}/\mathbb{Z}$, an interval exchange transformation $T_{(\lambda,\pi)}$ over $2k$ intervals $A_1, \dots, A_k, B_1, \dots, B_k$ and a partition of $\mathbb{R}/\mathbb{Z}$ into k intervals $I_i = A_i \cup B_i$ such that for all $i \in \mathbb{N}$,*

$$\mathbf{w}_i = j \quad \text{if } T^i_{(\lambda,\pi)}(x) \in I_j.$$

Moreover, the corresponding subshifts (called *Arnoux-Rauzy subshifts*) admit the following S-adic characterization. Let $A = \{0, \dots, k-1\}$. For all $a \in A$, let $R_a : A^* \to A^*$ be the morphism

$$R_a : \begin{cases} a \mapsto a \\ b \mapsto ba \text{ if } b \neq a \end{cases}$$

Theorem 2.2.19 (Arnoux-Rauzy [AR91]). *Let $A = \{0, \dots, k-1\}$. A subshift (X, T) over A is an Arnoux-Rauzy subshift if and only if there is a sequence $(a_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}$, each value of A occurring infinitely often in $(a_n)_{n\in\mathbb{N}}$, such that $(R_{a_n}, 0)_{n\in\mathbb{N}}$ is an S-adic representation of (X, T).*

2.2.4 Episturmian sequences

In addition to the combinatorial similarity between Arnoux-Rauzy sequences and Sturmian sequences (in terms of complexity and special factors), a property shared by both type of sequences is that their languages are closed under *reversal*, i.e., for all words $u = u_1 \cdots u_\ell$ in $L(\mathbf{w})$, the *reversal* $\tilde{u} = u_\ell \cdots u_1$ of u also belongs to $L(\mathbf{w})$. The class of *episturmian sequences* introduced in [DJP01] (see also [GJ09] for a recent survey) generalizes these properties. Formally, a sequence $\mathbf{w}$ over $A = \{0, \ldots, k-1\}$ is *episturmian* if $L(\mathbf{w})$ is closed under reversal and if there is at most one right (or, equivalently, left) special factor of each length in $\mathbf{w}$. Consequently, Arnoux-Rauzy sequences over $A = \{0, \ldots, k-1\}$ are episturmian sequences such that for all right special factors r, one has $\delta^+(r) = k$ (and so $p(n) = (k-1)n+1$). Observe that an episturmian sequence might be periodic which is not the case of Sturmian and Arnoux-Rauzy sequences and it is a direct consequence of the definition that all episturmian sequences have sub-linear complexity (see Theorem 1.2.3). One could also show that any episturmian sequence is uniformly recurrent.

To study episturmian sequences, Justin and Pirillo introduced episturmian morphisms (see [JP02]) that are exactly the morphisms that preserve the family of aperiodic episturmian sequences. They consist of the compositions of the permutation morphisms (i.e., morphisms σ such that $\sigma(A) = A$) and the morphisms L_a and R_a where, for all $a \in A$,

$$L_a : \begin{cases} a \mapsto a \\ b \mapsto ab \text{ if } b \neq a \end{cases} \quad \text{and} \quad R_a : \begin{cases} a \mapsto a \\ b \mapsto ba \text{ if } b \neq a \end{cases}$$

They obtained the following S-adic characterization.

Theorem 2.2.20 (Justin and Pirillo [JP02]). *Let $A = \{0, \ldots, k-1\}$ be an alphabet and $S = \{L_a \mid a \in A\} \cup \{R_a \mid a \in A\}$. A one-sided sequence $\mathbf{w} \in A^{\mathbb{N}}$ is episturmian if and only if $\mathbf{w}$ is S-adic.*

Contrary to Sturmian sequences, the S-adic decomposition is not unique for episturmian sequences. In [GLR09], the authors defined the notion of *normalized directive word* such that any episturmian sequence admits a unique normalized directive word, i.e., a unique *normalized S-adic representation*. As an application of it, they gave a characterization of episturmian sequences having a unique S-adic representation.

2.2.5 Linearly recurrent sequences

A last type of sequences for which the S-adic representations are well known is the set of linearly recurrent sequences. Formally, a sequence $\mathbf{w}$ is *linearly recurrent* if it is uniformly recurrent and if there is a constant K such that for all factors u of $\mathbf{w}$ and all integers i and j such that u successively occurs in $\mathbf{w}$ at positions i and j, we have $|i - j| \leq K|u|$.

Proposition 2.2.21 (Durand [Dur98a]). *A purely morphic sequence $\sigma^\omega(a)$ with σ everywhere growing is linearly recurrent if and only if σ is a primitive morphism.*

Then, Damanik and Lenz improved this result as follows.

Theorem 2.2.22 (Damanik and Lenz [DL06]). *Let $\mathbf{w} = \sigma^\omega(a)$ be a purely morphic sequence over A. The following are equivalent:*

1. *there is a growing letter $b \in A$ (Definition 1.3.5) that occurs with bounded gaps in $\mathbf{w}$ and such that for all letters $c \in A$ there is an integer n such that $|\sigma^n(b)|_c > 0$;*

2. *$\mathbf{w}$ is uniformly recurrent;*

3. *$\mathbf{w}$ is linearly recurrent.*

Durand, Host and Skau proved (in particular) that these sequences have a sub-linear complexity[3].

Theorem 2.2.23 (Durand, Host and Skau [DHS99]). *Let $\mathbf{w}$ be a linearly recurrent sequence (with constant K). Then:*

1. *for all $n \in \mathbb{N}$, all factors of length n occur in all factors of length $(K+1)n$;*

2. *$p_{\mathbf{w}}(n) \leq Kn$;*

3. *$\mathbf{w}$ is $(K+1)$-power free;*

4. *for all $u \in L(\mathbf{w})$ and all $v \in RRW_{\mathbf{w}}(u)$ (or $LRW_{\mathbf{w}}(u)$), we have $\frac{1}{K}|u| < |v|$;*

5. *for all $u \in L(\mathbf{w})$, $\mathrm{Card}(RRW_{\mathbf{w}}(u)) \leq K(K+1)^2$.*

Then, using return words, Durand proved the following result.

Theorem 2.2.24 (Durand [Dur03]). *A sequence $\mathbf{w}$ is linearly recurrent if and only if it is primitive and proper S-adic (see Definition 1.3.10 and Definition 1.3.11 on page 33) with $\mathrm{Card}(S) < +\infty$.*

[3]Recall that $RRW_{\mathbf{w}}(u)$ is the set of right return words to u in $\mathbf{w}$ and that $LRW_{\mathbf{w}}(u)$ is the set of left return words to u in $\mathbf{w}$ (see Section 1.1.)

Remark 2.2.25. Durand also proved that a Sturmian sequence (corresponding to a rotation r_α) is linearly recurrent if and only if the partial quotients of α are bounded. Consequently, a Sturmian sequence is linearly recurrent if and only if its S-adic representation is primitive which is a kind of generalization of Proposition 2.2.21. This is not true in general. Indeed, the next example provides a primitive S-adic sequence with $\mathrm{Card}(S) < +\infty$ which is not linearly recurrent.

Example 2.2.26 (Durand [Dur03])**.** Let $S = \{\sigma, \tau\}$ where σ and τ are defined by

$$\sigma : \begin{cases} 0 \mapsto 021 \\ 1 \mapsto 101 \\ 2 \mapsto 212 \end{cases} \quad \text{and} \quad \tau : \begin{cases} 0 \mapsto 012 \\ 1 \mapsto 021 \\ 2 \mapsto 002 \end{cases}$$

The sequence

$$\mathbf{w} = \lim_{n \to +\infty} \sigma\tau\sigma^2\tau \cdots \sigma^n\tau(0^\omega)$$

is primitive S-adic but not linearly recurrent. Indeed, for all k, let us define $\rho_k = \sigma\tau\sigma^2\tau \cdots \sigma^k\tau$ and $\mathbf{w}_k$ by

$$\mathbf{w}_k = \lim_{n \to +\infty} \sigma^{k+1}\tau\sigma^{k+2}\tau \cdots \sigma^{k+n}\tau(a^\omega).$$

We have $\mathbf{w} = \rho_k(\mathbf{w}_k)$ for all k. Now let v be a return word to 20 in $\mathbf{w}_k$. We have $|v| \geq 3^{k+2}$ (indeed, one can check that this is true if we replace $\mathbf{w}_k$ by any sequence $\sigma^{k+1}\tau(\mathbf{x})$ for $\mathbf{x} \in A^{\mathbb{N}}$). Moreover, the word $\rho_k(v)$ is also a return word to $\rho_k(20)$ in $\mathbf{w}$. Finally, we have

$$\frac{|\rho_k(v)|}{|\rho_k(20)|} = \frac{|v|}{|20|} \geq \frac{3^{k+2}}{2},$$

which contradicts the definition of linear recurrence.

The next result provides a sufficient condition for a primitive S-adic sequence to be linearly recurrent.

Lemma 2.2.27. *Let* $\mathbf{w}$ *be a primitive* S*-adic sequence whose directive word is* $(\sigma_n, a_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}} \times \prod_{n=0}^{+\infty} A_n$ *with* $\mathrm{Card}(S) < +\infty$*. For all* k*, let* $\mathbf{w}_k$ *be the sequence directed by* $(\sigma_n, a_n)_{n\geq k}$ *and let* D_k *be the length of the largest gap between two successive occurrences of a word of length 2 in* $\mathbf{w}_k$*. If* $(D_k)_{k\in\mathbb{N}}$ *is bounded, then* $\mathbf{w}$ *is linearly recurrent.*

2.3 S-adicity and sub-linear complexity

In the previous section, we presented the S-adic representations of some well-known sequences that have a sub-linear complexity. However, all these representations strongly depend on the nature of the corresponding sequences and very few things are known in the general case. In this section, we give some partial results about S-adicity and sub-linear complexity. There are some sufficient (but not necessary) conditions for an S-adic sequence to have a sub-linear complexity (they are due to Durand). There is also a necessary (but not sufficient) condition due to Ferenczi. Next, we present some examples that allow to reject some naive ideas that one could have when trying to work on the conjecture.

2.3.1 Partial results

Some sufficient conditions

In [Dur00] and [Dur03], Durand gave some sufficient conditions for an S-adic sequence to have a sub-linear complexity. These conditions are generalizations of what exists for purely morphic sequences (see Theorem 2.1.2 and Proposition 2.1.4). However, even some Sturmian sequences do not satisfy them (those with unbounded partial quotients).

Proposition 2.3.1 (Durand [Dur03]). *Let* $\mathbf{w}$ *be an* S*-adic sequence with* $\mathrm{Card}(S) < +\infty$ *and whose directive word is* $(\sigma_n, a_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}} \times \prod_{n=0}^{\infty} A_n$*. If there is a constant* D *such that for all* n*, all letters* $a \in A_{n+1}$ *and* $b \in A_{n+2}$*, we have*

$$|\sigma_0 \cdots \sigma_{n+1}(b)| \leq D|\sigma_0 \cdots \sigma_n(a)|,$$

then $p_{\mathbf{w}}(n) \leq D(\mathrm{Card}(A))^2 n$ *with* $A = \cup_{n\in\mathbb{N}} A_n$*.*

Corollary 2.3.2 (Durand [Dur03]). *If* $\mathbf{w}$ *is* S*-adic with* $\mathrm{Card}(S) < \infty$ *and all morphisms in* S *are uniform, then we have* $p_{\mathbf{w}}(n) \leq l(\mathrm{Card}(A))^2 n$ *with* $A = \cup_{n\in\mathbb{N}} A_n$ *and* $l = \max_{\sigma\in S, a\in A(\sigma)} |\sigma(a)|$*.*

Proposition 2.3.3 (Durand [Dur00]). *If* $\mathbf{w}$ *is a primitive* S*-adic sequence with constant* s_0 *(Definition 1.3.10) directed by* $(\sigma_n, a_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}} \times \prod_{n=0}^{\infty} A_n$ *with* $\mathrm{Card}(S) < +\infty$*, then there exists a constant* D *such that for all non-negative integers* r *and all letters* $a, b \in A_{r+s_0+1}$*, we have*

$$\frac{|\sigma_r \cdots \sigma_{r+s_0}(a)|}{|\sigma_r \cdots \sigma_{r+s_0}(b)|} \leq D.$$

Corollary 2.3.4. *Let S be a set of non-erasing morphisms and $\tau \in S$ be strongly primitive. Any S-adic sequence for which τ occurs infinitely often with bounded gaps in the directive word is uniformly recurrent and has a sub-linear complexity.*

Proof. First, the uniform recurrence is a consequence of Proposition 2.1.21.

Let $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ be a directive word in which the morphism τ occurs infinitely often with bounded gaps. We consider the set $LRW_{\mathbf{s}}(\tau)$ of left return words to τ in $\mathbf{s}$. Since τ occurs with bounded gaps in $\mathbf{s}$, this set is finite. Moreover, all morphisms in it are strongly primitive (as τ is) and the directive word $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ is equal to

$$\phi\tau_0\tau_1\cdots\tau_n\cdots$$

with $(\tau_n)_{n\in\mathbb{N}} \in LRW_{\mathbf{s}}(\tau)^{\mathbb{N}}$ and ϕ non-erasing. We conclude the proof using Propositions 2.3.3, 2.3.1 and 2.1.11. □

A necessary condition

In [Fer96], Ferenczi provided a general method to build an S-adic representation of any minimal subshift of sub-linear complexity. We will develop this approach in details in Chapter 3, Chapter 4 and Chapter 5. This will significantly improve Theorem 2.3.5 and Proposition 2.3.6 below.

Theorem 2.3.5 (Ferenczi [Fer96])**.** *Let (X,T) be an aperiodic minimal subshift over an alphabet A with sub-linear complexity. There is a finite set S of morphisms over an alphabet $D = \{0,1,\ldots,k-1\}$, a sequence $(\sigma_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$ and a non-erasing morphism $\tau : D^* \to A^*$ such that*

1. *for all letters $d \in D$, the length of $\sigma_0\sigma_1\cdots\sigma_n(d)$ tends to infinity with n;*

2. *for all words u in $L(X)$, there is an integer n such that u occurs in $\tau\sigma_0\sigma_1\cdots\sigma_n(0)$.*

Proposition 2.3.6 (Ferenczi [Fer96])**.** *Let (X,T) be a minimal subshift over a three-letters alphabet such that for all $n \geq 0$,*

$$1 \leq p_X(n+1) - p_X(n) \leq 2.$$

Then Theorem 2.3.5 holds for $k = 3$ and $\mathrm{Card}(S) < 3^{27}$.

2.3.2 Naive ideas about the conjecture

A natural idea to try to understand the conjecture is to consider examples composed of well-known morphisms. For instance, one could consider the *Fibonacci morphism* φ whose fixed point $\varphi^\omega(0)$ is a Sturmian sequence and the *Thue-Morse morphism* μ whose both fixed points $\mu^\omega(0)$ and $\mu^\omega(1)$ have a sub-linear complexity (Example 1.3.1). We have:

Proposition 2.3.7. *If $S = \{\varphi, \mu\}$ where φ and μ are respectively the Fibonacci morphism and the Thue-Morse morphism, any S-adic sequence is linearly recurrent.*

Proof. Let $\mathbf{w}$ be an S-adic sequence directed by $(\sigma_n, a_n)_{n\in\mathbb{N}}$ and for all $k \in \mathbb{N}$, let $\mathbf{w}^{(k)}$ be the S-adic sequence directed by $(\sigma_n, a_n)_{n\geq k}$. It is a direct consequence of the choice of S that $\mathbf{w}$ is primitive S-adic. From Lemma 2.2.27 it is therefore sufficient to prove that the sequence $(D_k)_{k\in\mathbb{N}}$ is bounded, where D_k is the length of the largest gap between two successive occurrences of a word of length 2 in $\mathbf{w}^{(k)}$.

First, let us prove that the words 0000 and 111 do not occur in any sequence $\mathbf{w}^{(k)}$. From the definition of φ and μ, any word of $L_4(\mathbf{w}^{(k)})$ occurs in a word of $\sigma(\{0,1\}^2)$ for $\sigma \in \{\mu^2, \varphi\mu^2, \varphi^4, \varphi^2\mu, \varphi^3\mu, \mu\varphi\mu, \varphi\mu\varphi\mu, \mu\varphi^2, \varphi\mu\varphi^2\}$. Indeed, for all these morphisms, $\sigma(0)$ and $\sigma(1)$ have length greater than 4. Moreover, observe that for all k, $\mathbf{w}^{(k)}$ is equal to one of the following sequences:

$$\mu^2(\mathbf{w}^{(k+2)}), \varphi\mu^2(\mathbf{w}^{(k+3)}), \varphi^4(\mathbf{w}^{(k+4)}), \varphi^2\mu(\mathbf{w}^{(k+3)}), \varphi^3\mu(\mathbf{w}^{(k+4)})$$
$$\mu\varphi\mu(\mathbf{w}^{(k+3)}), \varphi\mu\varphi\mu(\mathbf{w}^{(k+4)}), \mu\varphi^2(\mathbf{w}^{(k+3)}), \varphi\mu\varphi^2(\mathbf{w}^{(k+4)})$$

and we have

$$\begin{aligned}
\mu^2 &= [0110, 1001] \\
\varphi\mu^2 &= [010001, 001010] \\
\varphi^4 &= [01001010, 01001] \\
\varphi^2\mu &= [01001, 01010] \\
\varphi^3\mu &= [01001010, 01001001] \\
\mu\varphi\mu &= [011001, 010110] \\
\varphi\mu\varphi\mu &= [010001010, 010010001] \\
\mu\varphi^2 &= [011001, 0110] \\
\varphi\mu\varphi^2 &= [010001010, 010001]
\end{aligned}$$

From the shape of these morphisms, we are therefore ensured that the words 0000 and 111 do not belong to $L(\mathbf{w}_k)$ for all k.

Now let us bound the sequence $(D_k)_{k\in\mathbb{N}}$. Let $k \in \mathbb{N}$ and consider $\mathbf{w}_k$. The language $L_2(\mathbf{w}_k)$ is equal to $\{00, 01, 10, 11\}$ or $\{00, 01, 10\}$ depending that $\sigma_k = \mu$ or $\sigma_k = \varphi$. Suppose that $\sigma_k = \mu$ (the other case is similar) and let us give an upper bound for D_k. We have to show that all words in $L_2(\mathbf{w}_k)$ occur in $\mathbf{w}_k$ with bounded gaps and that the upper bound does not depend on k. Let us prove it for the word 00, the other cases being similar. We have $\mathbf{w}_k = \mu(\mathbf{w}_{k+1})$ and the word 00 occurs in $\mu(10)$. Moreover, any factor u of $\mathbf{w}_{k+1}$ in which 10 does not occur is such that 00 does not occur in $\mu(u)$. Furthermore, the gap between two occurrences of 10 in $\mathbf{w}_{k+1}$ is at most 5 (since 0000 and 111 do not occur in $\mathbf{w}_{k+1}$). Consequently, the gap between two occurrences of 00 in $\mathbf{w}_k$ is at most 10 ($= |\mu(v)|$ for any word v of length 5). □

Then, one could try to generalize the previous result by saying that if we take only "good morphisms" (i.e., morphisms that can only yields to sequences with a sub-linear complexity if they are considered alone), we should only get S-adic sequences with a sub-linear complexity. This is reinforced by the fact that all morphisms previously considered are "good morphisms".

Following Pytheas Fogg's members, Boshernitzan conjectured the following statement.

Conjecture (Boshernitzan)**.** *If S contains only morphisms that can only yield to sequences with sub-linear complexity, then any S-adic sequence has a sub-linear complexity.*

But, he eventually provided the following counter-example to that conjecture. Since we did not find any detailed proof of it, we provided it.

Example 2.3.8. Let γ and E be the morphisms over $\{0, 1\}$ defined by

$$\gamma : \begin{cases} 0 \mapsto 001 \\ 1 \mapsto 1 \end{cases} \quad \text{and} \quad E : \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0 \end{cases} .$$

Observe that both morphisms γE and $E\gamma$ are *primitive*. Consequently, their respective subshifts are minimal and have a sub-linear complexity. We consider the sequence

$$\mathbf{w}_{\gamma,E} = \lim_{n\to+\infty} \gamma E \gamma^2 E \gamma^3 E \cdots \gamma^{n-1} E \gamma^n (0^\omega).$$

Proposition 2.3.9 (Boshernitzan)**.** *The sequence $\mathbf{w}_{\gamma,E}$ is S-adic for $S = \{\gamma E, E\gamma\}$, is uniformly recurrent and does not have a sub-linear complexity.*

Proof. First it is obvious that $\mathbf{w}_{\gamma,E}$ is indeed S-adic for $S = \{\gamma E, E\gamma\}$.

Next, the composition $\gamma \circ E \circ \gamma$ is strongly primitive and occurs infinitely often in the directive word of $\mathbf{w}_{\gamma,E}$. It is therefore a consequence of Proposition 2.1.21 that $\mathbf{w}_{\gamma,E}$ is uniformly recurrent.

Finally, from Theorem 1.2.3 and Equation (1.1), we have to prove that the number of right special factors of length n of $\mathbf{w}_{\gamma,E}$ is unbounded.

For all $k \in \mathbb{N}^*$, let us define the morphism $\Gamma_k = \gamma E \gamma^2 E \cdots \gamma^{k-1} E \gamma^k E$ and, for all $k \in \mathbb{N}$, the sequence

$$\mathbf{w}^{(k)} = \lim_{n \to +\infty} \gamma^{k+1} E \gamma^{k+2} E \cdots \gamma^{k+n-1} E \gamma^{k+n}(0^\omega).$$

For all k we then have $\mathbf{w}_{\gamma,E} = \mathbf{w}^{(0)} = \Gamma_k(\mathbf{w}^{(k)})$. For all $i \geq 1$ we also define the word $u_i = \gamma^i(10) = 1\gamma^i(0)$.

Any sequence of the form $\gamma(\mathbf{x})$ with 0 and 1 recurrent in $\mathbf{x}$ contains both words 00 and 01. Observing that $\gamma(0)$ and $\gamma(1)$ start with different letters and end with same letter 1, we deduce that for all integers i, $1 \leq i \leq k+1$, the word u_i is a right special factor of $\mathbf{w}^{(k)}$.

Now let us prove that the number of right special factor of a given length of $\mathbf{w}_{\gamma,E}$ is unbounded. One can check that for all $k \geq 1$, the words $\Gamma_k(0)$ and $\Gamma_k(1)$ start with different letters. Consequently, for all integers i such that $1 \leq i \leq k+1$, the word $\Gamma_k(u_i)$ is a right special factor of $\mathbf{w}_{\gamma,E}$. These factors do not have the same length so we cannot immediately conclude. However, all suffixes of these factors are obviously right special and we will show that the number of suffixes of the words $\Gamma_k(u_i)$ increases with the length of these suffixes, which will conclude the proof.

First, let us compute the length of $\Gamma_k(u_i)$ for all k and i. We can easily see that $|u_i| = 2^{i+1}$. Indeed, we have $u_i = 1\gamma^i(0)$ and, by induction, we get $|\gamma^i(0)|_0 = 2^i$ and $|\gamma^i(0)|_1 = 2^i - 1$. This also proves that for all i, we have $|u_i|_0 = |u_i|_1 = 2^i$. Then, we can deduce from the shape of γ that if $v \in \{0,1\}^*$ is such that $|v|_0 = |v_1|$, then $|\gamma(v)| = 2|v|$ and $|\gamma(v)|_0 = |\gamma(v)|_1 = |v|$. Consequently, we obtain

$$|\Gamma_k(u_i)| = 2^{i+1} 2^{\sum_{j=1}^{n} j} = 2^{i+1} 2^{\frac{k(k+1)}{2}}.$$

Now let us study the suffixes of the words $\Gamma_k(u_i)$ for $k \geq 1$ and $1 \leq i \leq k+1$. It is easily seen that for all i, the largest common suffix of u_i and u_{i+1} is 1^i. We need to compute the length of $\Gamma_k(1^k)$ to determine a lower bound on the number of right special factors of $\mathbf{w}_{\gamma,E}$. Indeed, all right special factors $\Gamma_k(u_i)$ whose length are greater than $|\Gamma_k(1^k)|$ have a distinct suffix of length $|\Gamma_k(1^k)| + 1$ and we will show that the set of integers i such that u_i satisfies this property is increasing with k. We have $|\Gamma_k(1^k)| = k|\Gamma_k(1)|$ and:

$$\begin{aligned}|\Gamma_k(1)| &= \left|\gamma E\gamma^2 E\cdots\gamma^k E(1)\right| \\ &= \left|\gamma^k(0)\right|_0 \left|\gamma E\gamma^2 E\cdots\gamma^{k-1}E(0)\right| + \left|\gamma^k(0)\right|_1 \left|\gamma E\gamma^2 E\cdots\gamma^{k-1}E(1)\right| \\ &= 2^k\left(|\Gamma_{k-1}(01)| - |\Gamma_{k-1}(1)|\right) + \left(2^k-1\right)|\Gamma_{k-1}(1)| \\ &= 2^{\frac{k^2+k+2}{2}} - |\Gamma_{k-1}(1)|\,.\end{aligned}$$

Consequently we obtain

$$\begin{aligned}|\Gamma_k(1^k)| &= k\left(2^{\frac{k^2+k+2}{2}} - |\Gamma_{k-1}(1)|\right) \\ &\leq k2^{f(k)}\end{aligned}$$

with $f(k) = \frac{k^2+k+2}{2}$.

Now we can conclude the proof. For all i with $\log_2 k < i \leq k+1$, the word 1^i is suffix of u_i, hence $|\Gamma_k(u_i)| > k2^{f(k)}$. As the longest common suffix of $\Gamma_k(u_i)$ and $\Gamma_k(u_j)$ when $i < j$ is $\Gamma_k(1^i)$, we deduce the existence of $k+1-\lceil\log_2 k\rceil$ right special factors of $\mathbf{w}$ (as u_i is right special, also are its suffixes) of length $\left\lceil \Gamma_k(1^k)\right\rceil + 1$. □

Remark 2.3.10. The previous result is even stronger than just considering sets S of morphisms with fixed points of sub-linear complexity. Indeed, the sequence also has *bounded partial quotients*, i.e., all morphisms occur with bounded gaps in the directive word (over $\{\gamma E, E\gamma\}$).

An opposite question of the previous one is to ask whether S-adic sequences can have a sub-linear complexity when S contains a morphism that admits a fixed point that does not have a sub-linear complexity. The next example positively answers that question.

Example 2.3.11. Let γ be the morphism defined in Example 2.3.8. From Theorem 2.1.2 we known that the sequence

$$\gamma^\omega(0) = 001001^2001001^3001001^2001001^4\cdots$$

has a quadratic complexity.

Proposition 2.3.12. *Let* $(k_n)_{n\in\mathbb{N}}$ *be a sequence of non-negative integers. The sequence*

$$\mathbf{w} = \lim_{n\to+\infty}\gamma^{k_0}\mu\gamma^{k_1}\mu\gamma^{k_2}\mu\cdots\gamma^{k_n}\mu(0^\omega)$$

is uniformly recurrent. Moreover, $\mathbf{w}$ *has an at most linear complexity if and only if the sequence* $(k_n)_{n\in\mathbb{N}}$ *is bounded. Finally, for all* n *we have*

$$|\gamma^{k_0}\mu\gamma^{k_1}\mu\gamma^{k_2}\mu\cdots\gamma^{k_n-1}\mu(0)| = |\gamma^{k_0}\mu\gamma^{k_1}\mu\gamma^{k_2}\mu\cdots\gamma^{k_n-1}\mu(1)|.$$

and denoting

$$\ell_n = |\gamma^{k_0}\mu\gamma^{k_1}\mu\gamma^{k_2}\mu\cdots\gamma^{k_n-1}\mu(0)|.$$

we have

$$p_{\mathbf{w}}(\ell_n) \leq 4\ell_n - 2.$$

Proof. First, as μ occurs infinitely often in the directive word, it is a consequence of Proposition 2.1.21 that $\mathbf{w}$ is uniformly recurrent.

Now let us study the complexity depending on the sequence $(k_n)_{n\in\mathbb{N}}$. The case of a bounded sequence is a direct consequence of Corollary 2.3.4. Hence let us consider that the sequence $(k_n)_{n\in\mathbb{N}}$ is unbounded and let us show that the complexity is not at most linear. Using Theorem 1.2.3 and Equation (1.1), we only have to prove that the number of right special factors of length n of $\mathbf{w}$ is unbounded.

As said in Example 2.3.11, the fixed point $\gamma^\omega(0)$ has a quadratic complexity. Consequently the number of right special factors of $\gamma^\omega(0)$ of a given length is unbounded (Corollary 1.2.4). Moreover it is easily seen that all the right special factors of length n of $\gamma^\omega(0)$ occurs in $\gamma^{n+1}(0)$. Now let us show that if u is a right special factor of length n in $\gamma^{k_n}(a)$, then $\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(u)$ is a right special factor of $\mathbf{w}$ of length $n2^q$ with $q = \sum_{i=0}^{n-1}(k_i+1)$. Indeed, as $\mu(0)$ and $\gamma(0)$ start with 0 and $\mu(1)$ and $\gamma(1)$ start with 1, the image of u is still a right special factor. Moreover, $\mu(u)$ contains exactly n occurrences of the letter 0 and n occurrences of the letter 1, and both γ and μ map a word with the same number of 0 and 1 to a word of double length with the same number of 0 and 1. Hence $|\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(u)| = |u|2^q$ with q defined as previously. Now, if u and v are two distinct right special factors of length n of $\gamma^\omega(0)$, then $\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(u)$ and $\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(v)$ are two distinct special factors of length $n2^q$ of $\mathbf{w}$. As the number of right special factors of a given length of $\gamma^\omega(0)$ is unbounded, the number of right special factors of a given length of $\mathbf{w}$ is also unbounded which concludes the first part of the proof.

The last step is to show that, for all integers ℓ_n, we have $p_{\mathbf{w}}(\ell_n) \leq 4\ell_n$. For all non-negative integers n, we already know that

$$|\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(0)| = |\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(1)| = \ell_n = 2^q$$

with q as defined previously by $\sum_{i=0}^{n-1}(k_i+1)$. Consequently, all factors u of length ℓ_n are factors of $|\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(v)|$ for some words v of length 2. As there are only 4 possible binary words of length 2 and as there are less than $\ell_n + 1$ distinct factors of length ℓ_n in a word of length $2\ell_n$, we obtain $p_{\mathbf{w}}(\ell_n) \leq \ell_n + 4$. However, among the $\ell_n + 4$ words, both words $\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(0)$ and $\gamma^{k_0}\mu\gamma^{k_1}\mu\cdots\gamma^{k_{n-1}}\mu(1)$ have been counted 4 times, hence $p_{\mathbf{w}}(\ell_n) \leq \ell_n - 2$. □

The previous example provides an example of S-adic sequence with a "bad morphism" in S and it is shown that when there are only bounded powers of that bad morphism in the directive word, then the sequence has a sub-linear complexity, which is actually not very surprising. The following example shows that there even exist some S-adic sequences with sub-linear complexity such that there are arbitrarily large powers of a "bad morphism" in the directive word.

Example 2.3.13. Let us consider the morphisms

$$\beta : \begin{cases} 0 \mapsto 010 \\ 1 \mapsto 1112 \\ 2 \mapsto 2 \end{cases} \qquad \text{and} \qquad M : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 1 \end{cases}$$

and the sequence

$$\mathbf{w}_{\beta,M} = \lim_{n \to +\infty} M\beta M\beta^2 M\beta^3 M \cdots \beta^{n-1} M\beta^n(0^\omega).$$

Proposition 2.3.14. *The sequence $\mathbf{w}_{\beta,M}$ defined just above has a sub-linear complexity. More precisely, for all n we have $p(n+1) - p(n) \in \{1, 2\}$.*

Proof. Let S be the set of morphisms $\{M\beta^n \mid n \geq 1\}$. All morphisms $M\beta^n$ are defined over $\{0, 1\}$ and the sequence $\mathbf{w}_{\beta,M}$ is obviously S-adic. It is also non-periodic so $p(n+1) - p(n) \geq 1$ for all n.

For all n, let us consider $s(n) = p(n+1) - p(n)$. We have $s(0) = 1$ and, by Proposition 1.2.6,

$$s(n+1) - s(n) = \sum_{u \in L_n(\mathbf{w})} m(u)$$

where $m(u)$ denotes the bilateral order of u (see Definition 1.2.5). As $\mathbf{w}_{\beta,M}$ is a binary sequence, for all its factors u we have $m(u) \in \{-1, 0 - 1\}$ and we therefore have to compute the bilateral orders of strong and weak bispecial factors of length n to obtain $s(n+1) - s(n)$.

For all integers $k \geq 1$ let us consider the morphism $\mathcal{B}_k = M\beta M\beta^2 \cdots M\beta^k$ and the sequence $\mathbf{w}^{(k)}$ directed by $(M\beta^{k+1} M\beta^{k+1} \cdots, 0)$. We also define $\mathcal{B}_0 = id$ and we $\mathbf{w}^{(0)} = \mathbf{w}_{\beta,M}$. For all $k \geq 0$ we therefore have

$$\mathbf{w}_{\beta,M} = \mathcal{B}_k(\mathbf{w}^{(k)})$$

and $L_2(\mathbf{w}^{(k)}) = \{01, 10, 11\}$. Moreover, for all k the image of $\mathcal{B}_k(0)$ starts and ends with 0 and the image of $\mathcal{B}_k(1)$ contains no occurrences of the letter 0. Consequently, if v is a strong (resp. weak) bispecial factor in

$M\beta^{k+1}(L_2(\mathbf{w}^{(k+1)}))$, then $\mathcal{B}_k(v)$ is a strong (resp. weak) bispecial factor of $\mathbf{w}_{\beta,M}$.

It is an easy computation that for all k, the strong and weak bispecial factors in $M\beta^{k+1}(L_2(\mathbf{w}^{(k+1)}))$ are respectively

$$\begin{aligned} M\beta^i(1) &\quad \text{for } i \in \{0,1,\dots,k+1\} \text{ and} \\ M\beta^i(101) &\quad \text{for } i \in \{0,1,\dots,k\} \end{aligned}$$

and the ordinary bispecial factors are

$$1^j \qquad \text{for } i \in \{0,\dots,M\beta^{k+1}(1)-1\} \text{ such that } 1^j \neq M\beta^i(1)\ \forall i.$$

Note that for all k, the strong bispecial factor $M\beta^{k+1}(1)$ in the language $M\beta^{k+1}(L_2(\mathbf{w}^{(k+1)}))$ is the image under the morphism $M\beta^{k+1}$ of the bispecial factor $1 = M\beta^0(1)$ in $M\beta^{k+2}(L_2(\mathbf{w}^{(k+2)}))$.

Then, for all i, $M\beta^i(1)$ is a factor of $M\beta^{i+1}(0)$, hence of $M\beta^{i+1}(101)$. Since for all $k \geq 1$ we also have $|\mathcal{B}_k(0)| < |\mathcal{B}_k(1)|$, we deduce that for all k and all i, $0 \leq i \leq k$, we have

$$|\mathcal{B}_k(M\beta^i(1))| < |\mathcal{B}_k(M\beta^i(101))| < |\mathcal{B}_k(M\beta^{i+1}(1))|.$$

To conclude the proof, we have to show that the words $\mathcal{B}_k(M\beta^i(1))$ and $\mathcal{B}_k(M\beta^i(101))$ are respectively the only strong and weak bispecial factors of $\mathbf{w}_{\beta,M}$. Indeed, in that case there is an increasing sequence $(\ell_n)_{n\in\mathbb{N}}$ such that $l_0 = 1$ and for all n, there are two integers $k \geq 0$ and i, $0 \leq i \leq k$ such that

1. $|\mathcal{B}_k(M\beta^i(1))|$ has length ℓ_{2n} and
2. $|\mathcal{B}_k(M\beta^i(1))|$ has length ℓ_{2n+1}.

Consequently we have $s(\ell_{2n}+1) - s(\ell_{2n}) = 1$, $s(\ell_{2n+1}+1) - s(\ell_{2n+1}) = -1$ and for all integers $j \geq 1$ that does not occur in $(\ell_n)_{n\in\mathbb{N}}$, $s(j+1) - s(j) = 0$ so $s(n+1) - s(n) \in \{1,2\}$ for all n.

Consider a bispecial factor u of $\mathbf{w}_{\beta,M}$ and let k denote the unique integer such that $2|\mathcal{B}_k(0)| \leq |u| < 2|\mathcal{B}_{k+1}(0)|$. If $u \neq 1^j$ for some integer j, the word $\mathcal{B}_k(0)$ is a factor of u. Let v be the longest word in $\{0,1\}^*$ such that $\mathcal{B}_k(v)$ is factor of u. From the shapes of $\mathcal{B}_k(0)$ and $\mathcal{B}_k(1)$, if v' is such that $\mathcal{B}_k(v')$ is a factor of u, then v' is a factor of v. Consequently, the word v has to be bispecial in $\mathbf{w}^{(k)}$, hence in $M\beta^{k+1}(L_2(\mathbf{w}^{(k+1)})$ (from the length of u). Then, since v contains an occurrence of the letter 0, it is weakly bispecial and equal to some $M\beta^i(101)$. Therefore u is weak bispecial.

Now let $u = 1^j$ be a factor of $\mathbf{w}_{\beta,M}$. Let us prove that it is strongly bispecial if and only if $u = \mathcal{B}_k(M\beta^i(1))$ for some integers k and i. Let k be the greatest integer such that $|\mathcal{B}_k(1)| \leq |u| < |\mathcal{B}_{k+1}(1)|$; $\mathcal{B}_k(1)$ is therefore a factor of u. Let n be the greatest integer such that $\mathcal{B}_k(1^n)$ is a factor of u. We obviously have $1 \leq n < |M\beta^{k+1}(1)|$ (from the length of u). If $u \neq \mathcal{B}_k(1^n)$, then u is not strong bispecial. Indeed, we can decompose u into either $1^{m_1}\mathcal{B}_k(1^n)1^{m_2}$ or $1^{m_1}\mathcal{B}_k(1^n)$ or $1^{m_1}\mathcal{B}_k(1^n)$ for some integers m_1 and m_2. Then, since $\mathcal{B}_k(0) \in 0\{0,1\}^*0$, the word 1^{m_1} and 1^{m_2} are respectively proper prefix and proper suffix of $\mathcal{B}_k(1)$. Consequently the word $1^{m_1}\mathcal{B}_k(1^n)1^{m_2}$ can only be extended to the left and to the right by 1, the word $1^{m_1}\mathcal{B}_k(1^n)$ can be extended to the left by 1 and the right by 0 and by 1 (so it is right special) and the word $\mathcal{B}_k(1^n)1^{m_2}$ can be extended to the right by 1 and to the left by 0 and by 1 (so it left special). Consequently the word $0u0$ is not a factor of $\mathbf{w}_{\beta,M}$ and u is not strong bispecial. We therefore have $u = \mathcal{B}_k(1^n)$ and 1^n has to be strong bispecial in $\mathbf{w}^{(k)}$ for u to be strong bispecial in $\mathbf{w}_{\mathbf{w},M}$. Since this can happen only if $n = |M\beta^i(1)|$ for some integer i, the result holds true. □

As a first conclusion, finding the condition C of the conjecture seems to be a really hard problem. Indeed, Proposition 2.3.12 shows that it is not enough to put some conditions on the morphisms in S to determine the condition of the conjecture and that we also have to take care of the directive word. Moreover, considering only "good morphisms" can provide too high complexity (Propositions 2.3.9) and even when arbitrarily large powers of a "bad morphism" occur in the directive word, the complexity still might be sub-linear (Proposition 2.3.14).

2.4 Beyond linearity

For purely morphic sequences, the complexity function can have only 5 asymptotic behaviours and only depends on the growth rate of images (see Theorem 2.1.2). For S-adic sequences we have seen in previous sections (for instance in Section 2.1.3) that things are highly more complicated. However, in Theorem 2.3.5, Ferenczi showed that if a minimal subshift has a sub-linear complexity, then it is S-adic and the length of all images tends to infinity (first point of the theorem). This is a kind of generalization of the third point of Theorem 2.1.2. Moreover, that property (i.e., the fact that the length of all images tends to infinity) is satisfied by most of the examples considered in previous sections. It is also interesting to note that for purely morphic sequences, the class of highest complexity $\Theta(n^2)$ can be reached only by morphisms with bounded letters (still Theorem 2.1.2). Furthermore, up to now, Cassaigne's constructions (Proposition 2.1.15) are the only ones that allow to build S-adic sequences with arbitrarily high complexity and they admit several bounded letters. Consequently, the fact that the length of all images tends to infinity with n seems to be important to get a *reasonably low complexity*. We propose to say that such an S-adic sequence is *everywhere growing* (Definition 1.3.13).

It is obvious that the everywhere growing Property is not a necessary condition for an S-adic sequence to have a sub-linear complexity since Cassaigne's constructions also hold for sequences with low complexity. One can also think to the *Chacon substitution* ζ defined by $\zeta(0) = 0010$ and $\zeta(1) = 1$ whose fixed point $\zeta^\omega(0)$ has complexity $p(n) = 2n + 1$ for all n (see [Fer95]). It is neither a sufficient condition since the sequence $\mathbf{w}_{\gamma,E}$ of Example 2.3.11 satisfies it and does not always have a sub-linear complexity. However, one could ask whether any high complexity can be reached by S-adic sequences satisfying it. This question seems to be a new non-trivial problem. Proposition 2.4.1 below provides a partial answer to that question. Indeed, it deals with expansive S-adic sequences (see Definition 1.3.12 on page 33), i.e., with S-adic sequences such that for all morphisms σ in S and all letters a, we have $|\sigma(a)| \geq 2$. Techniques are similar to those used in [ELR75] for D0L systems.

Recall that a *D0L system* (which means *deterministic L-system without interaction*) is essentially equivalent to a morphism $\sigma : A^* \to A^*$. Roughly speaking, the main difference is that for D0L systems, we are only interested in the language of the fixed point. In the same way that S-adic sequences are a generalization of (purely) morphic sequences, *DT0L systems* (which means *deterministic table system without interaction*) are a generalization of D0L systems. However there is a more important difference between DT0L and S-adic sequences than between D0L and substitutive sequences. Indeed, for DT0L systems, the language one is usually interested in is the set of words occurring in $\sigma_0\sigma_1 \cdots \sigma_k(a)$ for any finite sequence in S^* (where S denotes also the set of rules of the system). In other words, we consider the language of all S-adic sequences (i.e., we consider all directive words). It is proved in [ELR76] that *everywhere growing* DT0L systems (which means $|\sigma(a)| \geq 2$ for all σ and a, i.e., which is equivalent to expansivity for S-adic sequences) with a finite number of substitution rules have an at most polynomial complexity. For S-adic sequences built upon the same hypothesis, we have a better upper bound as it is shown by Proposition 2.4.1 below.

Proposition 2.4.1. *If* $\mathbf{w}$ *is an expansive S-adic sequence (Definition 1.3.12) such that* $\mathrm{Card}(S) < +\infty$*, then* $p_{\mathbf{w}}(n) \in O(n \log n)$.

Proof. First let us recall the definition of the *radix order* $\preceq^*$. Let $\preceq$ be an order on the alphabet A and let u and v be in A^*, $u \neq v$. We have $u \prec^* v$ if either $|u| < |v|$ or $|u| = |v|$ and there is a smallest integer $i \in [1, |u|]$ such that $u_i \prec v_i$.

Let $\mathbf{w}$ be an S-adic sequence with directive word $(\sigma_n, a_n)_{n\in\mathbb{N}}$. We let ℓ denote the maximal length of $\sigma(a)$ for σ in S and a in $A(\sigma)$. Consider an integer n greater than 2ℓ. For all words u in $L_n(\mathbf{w})$, we construct a sequence $(u_k)_{k\in\mathbb{N}}$ of words in the following way:

- $u_0 = u$;

- for all non-negative integers k, $\mathbf{w}^{(k)}$ is the S-adic sequence directed by $(\sigma_n, a_n)_{n\geq k}$ and u_{k+1} is the smallest word in $L(\mathbf{w}^{(k+1)})$ (with respect to the radix order) such that $u_k \in L(\sigma_k(u_{k+1}))$

We can easily see that the sequence $(|u_k|)_{k\in\mathbb{N}}$ is decreasing until a smallest integer r such that $|u_r| \leq 2$. We have $2 < r < 1 + C\log n$ for a constant C, the first inequality being trivial from the choice of n. For the second one, observe that $|u_{r-1}|$ is at least 3. Then, writing $u_{r-1} = av_{r-1}b$ with $a, b \in A$, we see that $\sigma_0\sigma_1\cdots\sigma_{r-2}(v_{r-1})$ is a proper factor of length at least 2^{r-1} of u. Therefore we have $n > 2^{r-1}$ and then $r < C\log n + 1$.

Now for all words u in $\mathbf{A}^*$, $\mathbf{A} = \bigcup_{n\in\mathbb{N}} A_n$, of length smaller than or equal to 2, we define $W_n(u)$ as the set of words of length n in $L(\mathbf{w})$ such that the construction previously described gives $u_r = u$. Obviously, $\bigcup_{u\in\mathbf{A}^{\leq 2}} W_n(u) = L_n(\mathbf{w})$. Then, each word $u \in \mathbf{A}^{\leq 2}$ provides at most $r-1$ factors of $\mathbf{w}$ that are $\sigma_0\sigma_1(u), \sigma_0\sigma_1\sigma_2(u), \ldots, \sigma_0\sigma_1\cdots\sigma_{r-1}(u)$ (maybe some of them are not well defined) and we have $r < 1 + C\log n$. To conclude the proof, we only have to check that there are no more than n words of length n in $\sigma_0\sigma_1\cdots\sigma_{r-1}(u_r)$ that admit $\sigma_0\sigma_1\cdots\sigma_{r-2}(v_{r-1})$ as a factor. □

Example 2.4.2 shows that this bound is the best one we can obtain.

Example 2.4.2. Let β be the morphism

$$\vartheta : \begin{cases} 0 \mapsto 0120 \\ 1 \mapsto 11 \\ 2 \mapsto 222 \end{cases}$$

and consider its fixed point $\mathbf{w} = \vartheta^\omega(a)$. It can be seen as an everywhere growing $\{\vartheta\}$-adic sequence and we know from Theorem 2.1.2 that $p_{\mathbf{w}}(n) = \Theta(n\log n)$.

Chapter 3

Some improvements of the S-adic conjecture

In this chapter, we present a general method to build S-adic representations of uniformly recurrent sequences (or minimal subshifts). The main idea is to use Rauzy graphs to build a sequence of morphisms $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ and then to consider the action of these morphisms over subsets of A_n^*. The way to construct $(\sigma_n)_{n\in\mathbb{N}}$ has first been introduced by Rauzy in [Rau83] and then in [AR91] for the particular case of Arnoux-Rauzy sequences (see Section 2.2.3). Then the idea of considering these morphisms over subsets of A_n^* is due to Ferenczi in [Fer96]. This chapter is mostly based on that last paper since we essentially present the method using the same two different kinds of subsets of A_n^*, but with much more details than in [Fer96]. In particular, this allows us to significantly improve Theorem 2.3.5.

Depending on the complexity of the sequence and on the chosen subsets of A_n^*, we of course get different properties of the S-adic representation. For instance, for sequences with sub-linear complexity, one of the two choices (based on particular concatenations of *n-segments* which we will define in Section 3.1.1) always provides a finite set S although it might be infinite for other subsets which are based on *n-circuits* (Section 3.1.2). On the other hand, that last method always provides strongly primitive morphisms that are also proper and we known from Theorem 2.2.24 that when $\mathrm{Card}(S) < +\infty$, these properties imply that the corresponding sequence is linearly recurrent.

In this Chapter, we always work with one-sided sequences. However, we will see that all methods can be adapted to two-sided sequences. The main results of this chapter are the following.

Theorem 3.0.1. *A one-sided sequence* $\mathbf{w}$ *is uniformly recurrent if and only* *if it is primitive and left proper S-adic*[1]*. Moreover, if* $\mathbf{w}$ *does not have a sub-linear complexity, then* $\mathrm{Card}(S) = +\infty$*. When dealing with subshifts instead of sequences, we can moreover replace "left proper" by "proper".*

From the definition, we can directly deduce the following corollary.

Corollary 3.0.2. *A one-sided sequence* $\mathbf{w}$ *is uniformly recurrent if and only if it is almost primitive and left proper S-adic. When dealing with subshifts instead of sequences, we can moreover replace "left proper" by "proper".*

Theorem 3.0.3. *Let A be an alphabet and $\mathbf{w} \in A^{\mathbb{N}}$ be a one-sided uniformly recurrent sequence with sub-linear complexity. There is a finite set S of morphisms such that $\mathbf{w}$ is S-adic and such that its directive word is everywhere growing*[2] *and satisfies Properties 1–3 of Definition 3.0.4 below.*

Let $\sharp \notin A$ and $\mathbf{w}' = \sharp\mathbf{w}$. When $\mathbf{w}'$ do not admit constant segments[3]*, the directive word is furthermore almost primitive*[4]*, satisfies Property 4 of Definition 3.0.4 and we can also replace Property 3 by (with the same notations)*

$$\sigma_n(d) \in (A_n^* \setminus A_n^* a A_n^*) \cup (u_1 u_2 \cdots u_\ell a A_n^*)$$

Definitions 3.0.4 (Properties)**.** A directive word $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ satisfies

1. *Property 1* if there is a non-negative integer N such that for all $n \geq N$, all letters $a \in A_n$ and all letters $c \in A_{n+1}$, we have $\sigma_n(c) \notin A_n^* a A_n^* a A_n^*$;

2. *Property 2* if there is a non-negative integer N such that for all $n \geq N$, all letters $a_1 \dots a_k$ in A_n and all letters $c_1, \dots, c_k$ in A_{n+1} with $k \geq 2$, we have

$$(\sigma_n(c_1), \dots, \sigma_n(c_k)) \notin \left(\prod_{i=1}^{k-1} A_n^* a_i A_n^* a_{i+1} A_n^* \right) \times A_n^* a_k A_n^* a_1 A_n^*;$$

3. *Property 3* if there is a non-negative integer N such that for all $n \geq N$, if $\sigma_n(c) \in u a A_n^*$ for $u = u_1 \cdots u_\ell \in A_n^+$, $a \in A_n$ and $c \in A_{n+1}$, then for all letters $d \in A_{n+1}$, we have

$$\sigma_n(d) \in (A_n^* \setminus A_n^* a A_n^*) \cup (A_n u_2 \cdots u_\ell a A_n^*)\,;$$

[1]See Definition 1.3.10 and Definition 1.3.11 on page 33.
[2]See Definition 1.3.13 on page 33.
[3]See Definition 3.1.9 on page 80.
[4]See Definition 1.3.14 on page 33.

4. *Property 4* if for all n, σ_n belongs to T^* with $T = \{G\} \cup \{E_{ij} \mid i,j \in \mathbf{A}\} \cup \{M_i \mid i \in \mathbf{A}\}$ a set of morphisms such that:
   ▶ $G(0) = 10$ and $G(i) = i$ for all letters $i \neq 0$;
   ▶ E_{ij} exchange i and j and fix the other letters;
   ▶ M_i maps i to 0 and fix the other letters.

3.1 Rauzy graphs: n-segments and n-circuits

To compute the S-adic representations of Theorem 3.0.1 and 3.0.3 we need to consider some particular paths in the Rauzy graphs (see Section 1.5) of the considered sequence or subshift. The main idea is that those paths are labelled by words of $L(\mathbf{w})$ and have larger and larger lengths when they are chosen in Rauzy graphs of larger and larger orders. Then, we show that such paths in a Rauzy graph of order $n+1$ are composed of paths in the Rauzy graph of order n; this will provide the morphisms of the directive word.

To explicitly formulate what happens to these paths, we need to define the following function. First, let $\mathcal{P}_n$ denote the set of paths in a Rauzy graph G_n. To be coherent with some definitions that will occur later, we need to consider the concatenation on $\mathcal{P}_n$. Observe that some concatenations of paths might not be a path in G_n, i.e., $\mathcal{P}_n \subsetneq \mathcal{P}_n^*$.

Definition 3.1.1. For all n, we let $\psi_{n,L}$ denote the function defined on $\mathcal{P}_{n+1}$ such that if $p \in \mathcal{P}_{n+1}$ is such that $\lambda_L(p) = u$, then $\psi_{n,L}(p)$ is the unique path q in $\mathcal{P}_n$ such that $\lambda_L(q) = u$, $o(q)$ is the prefix of length n of $o(p)$ and $i(q)$ is the prefix of length n of $i(p)$.

Roughly speaking, $\psi_{n,L}(p)$ is the corresponding path in $G_n(\mathbf{w})$ of the path p in $G_{n+1}(\mathbf{w})$. Observe that $\psi_{n,L}$ is not one-to-one. Indeed, if for example $\mathrm{Card}(A) = 2$ and p is a path in $G_n(\mathbf{w})$ that does not go through any bispecial vertex and such that $i(p)$ is strong bispecial, then the two right extensions of $i(p)$ are left special and both of them are therefore an extremity of a path q in $\mathcal{P}_{n+1}$ such that $\psi_{n,L}(q) = p$. Consequently, we have $\mathrm{Card}(\psi_{n,L}^{-1}(p)) = 2$. In Chapter 4 and Chapter 5, we will similarly define a function $\psi_{n,R}$ (Definition 4.1.1).

3.1.1 n-segments

The base of the S-adic representations of Theorem 3.0.1 and 3.0.3 is defined upon some paths in Rauzy graphs that are called n-segments. They were first introduced by Rauzy in [Rau83] and then used in [AR91] and [Fer96].

Definition 3.1.2. Let $n \in \mathbb{N}$ and G_n be a Rauzy graph. A *left n-segment* (resp. *right n-segment*) is a non-empty path $p \in \mathcal{P}_n$ whose only left (resp. right) special vertices are its extremities $o(p)$ and $i(p)$. In this chapter, we will mostly use left n-segments. Consequently, if not explicitly stated, *n-segment* means *left n-segment*. However, this only holds for this chapter. Take care that in Chapter 4 and Chapter 5, we will mostly work with right n-segments.

Example 3.1.3. Let $\mathbf{t}$ be the Thue-Morse sequence (Example 1.3.1). The Rauzy graph $G_3(\mathbf{t})$ labelled with left labels is represented in Figure 3.1. The left special factors are 010, 100, 101 and 011 and the 3-segments are the paths

$$\begin{array}{ll} 010 \to 101 & 100 \to 001 \to 010 \\ 010 \to 100 & 100 \to 001 \to 011 \\ 101 \to 010 & 011 \to 110 \to 101 \\ 101 \to 011 & 011 \to 110 \to 100 \end{array}$$

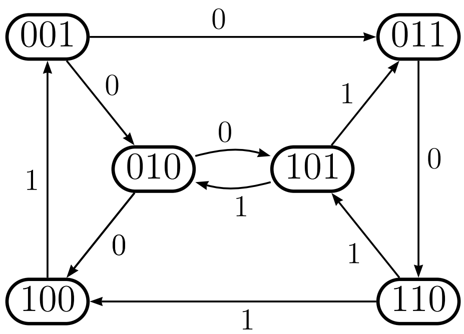


Figure 3.1: Rauzy graph of order 3 (with left labels) of the Thue-Morse sequence.

Remark 3.1.4. Observe that any (left or right) n-segment is trivially an allowed path (see Definition 1.5.4 on page 37). By definition, its full label is therefore a word of $L(\mathbf{w})$. Moreover, as the Rauzy graphs of recurrent sequences are strongly connected (see Remark 1.5.2), the set of n-segments is a covering of the set of edges of G_n in the sense that each edge belongs to at least one n-segment. Furthermore, for each n, as there exists only a finite (possibly unbounded) number of left special vertices in G_n, there exists only a finite (possibly unbounded) number of n-segments. Actually, it is easily seen that an n-segment p is completely determined by its ending vertex $i(p)$ and by the left label of its last edge (i.e., of the edge that arrives in $i(p)$). Consequently, the number of n-segments is exactly

$$\sum_{u \in LS_n(\mathbf{w})} \delta^-(u) = p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n) + \mathrm{Card}(LS_n(\mathbf{w}))$$

and this number is bounded by $\text{Card}(A)\text{Card}(LS_n(\mathbf{w}))$, hence by

$$\text{Card}(A)\left(p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n)\right).$$

The following corollary is a direct consequence of the previous remark and of Theorem 1.2.3.

Corollary 3.1.5. *A sequence* $\mathbf{w}$ *has a sub-linear complexity if and only if there is a constant* K *such that for all* n*, the number of* n*-segments in* $G_n(\mathbf{w})$ *is less than* K*.*

Remark 3.1.6. The notion of n-segment is related to the notion of *return word* (see Section 1.1). Indeed, it is easily seen that the set of left labels of n-segments is exactly $LRW_{\mathbf{w}}(LS_n(\mathbf{w}))$. Observe that some n-segments might have the same left label so we have

$$\text{Card}(LRW_{\mathbf{w}}(LS_n(\mathbf{w}))) \leq \text{Card}(\{n\text{-segments in } G_n(\mathbf{w})\}).$$

Remark 3.1.7. If the alphabet of $\mathbf{w}$ is $A = \{0, \ldots, k-1\}$, the Rauzy graph G_0 is as in Figure 3.2 so that for all 0-segments p, we have $\lambda(p) = \lambda_L(p) = \lambda_R(p) \in A$.

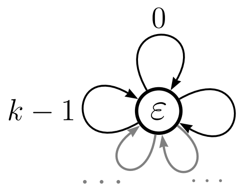


Figure 3.2: Rauzy graph G_0 of any sequence over $\{0, \ldots, k-1\}$.

For uniformly recurrent sequences with a "reasonably low" complexity, the number of left special factors increases much more slowly than the complexity. Consequently, we expect that the maximal length of n-segments will grow to infinity. Then, due to the uniform recurrence, all factors of $\mathbf{w}$ of length smaller than some ℓ will be factors of the label of the longest n_ℓ-segment for some n_ℓ large enough. Therefore, our aim is to study the behaviour of n-segments as n increases.

Lemma 3.1.8 here below — and also Lemmas 3.1.16, 3.2.6 and 3.4.1 in next sections — was already proved in [Fer96]. All these lemmas were parts of the proof of Theorem 2.3.5, but without being stated explicitly. Here, we decided to structure the proof in several lemmas.

Lemma 3.1.8 (Ferenczi [Fer96])**.** *Let* $\mathbf{w}$ *be a sequence over an alphabet* A*. For any* $(n+1)$*-segment* p *of* $\mathbf{w}$*,* $\psi_{n,L}(p)$ *is a concatenation of* n*-segments of* $\mathbf{w}$*. Moreover, the decomposition of* $\psi_{n,L}(p)$ *into* n*-segments is unique.*

Proof. Let p be a $(n+1)$-segment in $G_{n+1}(\mathbf{w})$ and $p' = \psi_{n,L}(p)$. As a prefix of a left special factor is still a left special factor, $o(p')$ and $i(p')$ are left special. Hence p' is a concatenation of n-segments. The uniqueness of the decomposition is obvious. Indeed, for two different n-segments q_1 and q_2, we have either $i(q_1) \neq i(q_2)$ or $i(q_1) = i(q_2)$ and $\lambda_L(q_1) \neq \lambda_L(q_2)$. Consequently, starting from $i(p')$, there is a unique n-segment q such that $\lambda_L(q)$ is suffix of $\lambda_L(p')$. Then if $q \neq p'$, there is a unique n-segment q' such that $i(q') = o(q)$ and $\lambda_L(q'q)$ is suffix of $\lambda_L(p')$. Continuing this way, we see that the decomposition is unique. □

Definition 3.1.9. From the previous lemma, the minimal length among all n-segments is non-decreasing. If it is bounded, there is an integer N and an N-segment p such that for all integers $n > N$, there is an n-segment q such that $p = \psi_{N,L}\psi_{N+1,L}\cdots\psi_{n-1,L}(q)$. Such a segment is said to be *constant*. Another equivalent definition is to say that an n-segment p is constant if there are two one-sided sequences $\mathbf{x}$ and $\mathbf{y}$, such that for all i, both $o(p)\mathbf{x}_{[0,i]}$ and $i(p)\mathbf{y}_{[0,i]}$ are left special factors of $\mathbf{w}$ and there is a path from $o(p)\mathbf{x}_{[0,i]}$ to $i(p)\mathbf{y}_{[0,i]}$ in $G_{n+i}(\mathbf{w})$ with left label $\lambda_L(p)$.

Remark 3.1.10. For aperiodic sequences with sub-linear complexity, for all n large enough there is at least one n-segment which is not constant. Indeed, the length of a constant segment is fixed and by Corollary 3.1.5, the number of n-segments is bounded by a constant K. Consequently, if all n-segments are constant, they all have length bounded by ℓ. Thus, the graph would have less than $K\ell$ edges. Since the number of edges in G_n is exactly $p_{\mathbf{w}}(n+1)$, this cannot happen for n large enough. As a consequence, we have

$$\lim_{n\to+\infty} \max\{|p| \mid p \text{ is an } n\text{-segment}\} = +\infty. \tag{3.1}$$

Also, since the number of n-segment is bounded and since two distinct constant n-segments give rise to distinct constant m-segments, $m > n$, there can exist only a bounded number of constant segments[5]. Consequently, there is an integer ℓ such that any constant segment has length bounded by ℓ.

3.1.2 n-circuits

The S-adic representation of Theorem 3.0.1 is based on n-circuits. They are also widely used in Chapter 4 and Chapter 5.

[5] We of course only consider the "initial" constant segments, i.e., if p is a constant n-segment, we do not consider the constant m-segments q, $m > n$, such that $p = \psi_{n,L}\cdots\psi_{m-1,L}(q)$.

Definition 3.1.11. Let $n \in \mathbb{N}$ and G_n be a Rauzy graph. A *left n-circuit* (resp. *right n-circuit*) is a non-empty path $p \in \mathcal{P}_n$ such that $o(p) = i(p)$ is a left (resp. right) special vertex and no interior vertex of p is $o(p)$. As for n-segments, we will mostly use left n-circuits in this chapter. Consequently, if not explicitly stated, *n-circuit* means *left n-circuit.* Once again, this only holds for this chapter since in Chapter 4 and Chapter 5 we will mostly work with right n-circuits.

Observe that, contrary to the n-segments, an n-circuit is not always an allowed path. Indeed, consider the path

$$010 \to (101 \to 011 \to 110 \to 101)^3 \to 010$$

in Figure 3.1 (on page 78). It is a 3-circuit and its full label contains the word $(101)^3$ which is not a factor of $\mathbf{t}$ since the Thue-Morse sequence is cube-free. However, when a Rauzy graph G_n is strongly connected, the set of allowed n-circuits is a covering of its edges in the sense that any edge occurs in at least one n-circuit. Furthermore, even if we fix a left special vertex l, the set of allowed n-circuits starting from l is still a covering of the edges. One can also note that for all sequences $\mathbf{w}$ over A, the set of 0-circuits is exactly the set of 0-segments. Therefore we have $\lambda(p) = \lambda_L(p) = \lambda_R(p) \in A$ for all 0-circuits p (see Remark 3.1.7).

Remark 3.1.12. Like for n-segments, the notion of n-circuit is closely related to the notion of return word. Indeed, if l is a left special vertex in a Rauzy graph $G_n(\mathbf{w})$, then the left labels of the n-circuits starting from l are exactly the elements of $LRW_{\mathbf{w}}(l)$. Moreover we have a one-to-one correspondence between n-circuits and return words, i.e., $\mathrm{Card}(LRW_{\mathbf{w}}(l)) = \mathrm{Card}(\{n\text{-circuits strating from } l\})$.

In particular, the above remark implies that the set of allowed n-circuits starting from a given left special vertex might be infinite. Indeed, a sequence is uniformly recurrent if and only if for all its factors u, the number of return words to u is finite. Moreover, the number of return words to a factor u is equal to the number of return words to the smallest bispecial factor v containing u as a factor. Consequently, if $\mathbf{w}$ is recurrent but not uniformly recurrent, there is a bispecial (hence left special) factor v such that the number of allowed $|v|$-circuits is infinite.

Like for n-segments, for uniformly recurrent sequences with a "reasonably low" complexity, we expect that the maximal length of n-circuits will grow to infinity. Then, due to the uniform recurrence, all factors of $\mathbf{w}$ of length smaller than some ℓ will be factors of the label of the longest n_ℓ-circuit for some n_ℓ large enough. Our aim is therefore to study the behaviour of the

n-circuits as n increases. The proof of the next lemma is exactly the same as the proof of Lemma 3.1.8

Lemma 3.1.13 (Ferenczi [Fer96]). *Let* $\mathbf{w}$ *be a sequence over an alphabet* A *and let* v *be a left special factor of length* $n+1$ *in* $\mathbf{w}$. *For all* $(n+1)$*-circuit* p *starting from* v, $\psi_{n,L}(p)$ *is a concatenation of* n*-circuit starting from the prefix of length* n *of* v. *Moreover, the decomposition of* $\psi_{n,L}(p)$ *into* n*-circuits is unique.*

The next lemma is well known and defines a sequence of left special factors $(v_n)_{n\in\mathbb{N}}$ to which we will apply Lemma 3.1.13. In the sequel, this sequence will be widely used, especially in Chapter 4 and Chapter 5 with the difference that instead of left special factors, we will consider right special factors.

Lemma 3.1.14. *Let* $\mathbf{w}$ *be an aperiodic sequence over an alphabet* A. *There exists an infinite sequence* $(v_n)_{n\in\mathbb{N}}$ *of words over* A *such that for each* $n \in \mathbb{N}$,

- v_n *is of length* n;
- v_n *is a left special factor of* $\mathbf{w}$;
- v_n *is a prefix of* v_{n+1}.

Proof. Let $\mathcal{T}$ be the directed graph whose vertices are the left special factors in $L(\mathbf{w})$ and such that there is an edge from u to v if u is a prefix of length $|v|-1$ of v. The sequence being aperiodic, there is at least one left special factor of each length so $\mathcal{T}$ is an infinite tree with, for all vertices, a bounded number of outgoing edges. We conclude the proof using König's Lemma (see Proposition 1.2.3 in [Lot02]). ☐

Definition 3.1.15. Like for n-segments, Lemma 3.1.13 implies that the minimal length of n-circuits is non-decreasing. If it is bounded, there is an integer N and a N-circuit p such that for all integers $n > N$, there is a n-circuit q such that $p = \psi_{N,L}\psi_{N+1,L}\cdots\psi_{n-1,L}(q)$. Such a circuit is said to be *constant.*

The next lemma states that for uniformly recurrent sequences, there is no constant n-circuits.

Lemma 3.1.16 (Ferenczi [Fer96]). *Let* $\mathbf{w}$ *be a uniformly recurrent sequence over an alphabet* A. *For any non-negative integer* n, *there is no constant* n*-circuit in* $G_n(\mathbf{w})$.

Proof. As the sequence $\mathbf{w}$ is uniformly recurrent, if it is ultimately periodic, it is periodic. Hence, in this case, there is no left special factor of length greater than some N and so no n-circuit for $n > N$. Now suppose that $\mathbf{w}$

is aperiodic and let p be a constant n-circuit of left label u in $G_n(\mathbf{w})$. By definition, for all positive integers k, there is an $(n+k)$-circuit q_k such that $p = \psi_{n,L} \cdots \psi_{n+k-1,L}(q_k)$. As the left label of q_k is u by definition, from Proposition 1.5.5 we deduce that, for all k large enough, $o(q_k)$ is equal to u followed by a prefix of itself. So, u^{e_k} is a prefix of $o(q_k)$ with $e_k = \left\lfloor \frac{|o(q_k)|}{|u|} \right\rfloor$. Since $o(q_k)$ is a factor of $\mathbf{w}$ for all k and $|o(q_k)|$ tends to infinity with k, there are arbitrarily large powers of u in $L(\mathbf{w})$ and this contradicts aperiodicity and uniform recurrence. □

The next corollary is a direct consequence of the previous lemma.

Corollary 3.1.17. *Let* $\mathbf{w}$ *be a uniformly recurrent sequence over an alphabet* A. *For any non-negative integer* ℓ, *there is an integer* n_ℓ *such that any* n_ℓ-*circuit has length greater than* ℓ.

3.2 Base of S-adic representations

In this section, we provide a general method to build morphisms. These morphisms will be the base of those considered for the S-adic representations of Theorem 3.0.1 and 3.0.3. Indeed, the method we give here provides some morphisms $\sigma_n : A_{n+1}^* \to A_n^*$. Then, the S-adic representations of both cited theorems will be obtained by considering some subsets of A_n^* for all n as new alphabets and to consider the morphisms σ_n over these new alphabets.

Definition 3.2.1 (Definition of the morphisms σ_n)**.** Lemma 3.1.8 allows us to define some morphisms σ_n over the alphabets of n-segments. Indeed, for each non-negative integer n, let $\mathcal{A}_n$ be the set of n-segments, A_n be the set $\{0, 1, \ldots, \mathrm{Card}(\mathcal{A}_n) - 1\}$ and let us consider a bijection $\theta_n : A_n \to \mathcal{A}_n$. We can extend θ_n to an isomorphism $\theta_n : A_n^* \to \mathcal{A}_n^*$ putting $\theta_n(ab) = \theta_n(a)\theta_n(b)$ for all $a, b \in A_n$. Now for all n, we define the morphism $\sigma_n : A_{n+1}^* \to A_n^*$ as the unique map that satisfies

$$\theta_n \circ \sigma_n = \psi_{n,L} \circ \theta_{n+1}.$$

Remark 3.2.2. If the alphabet of $\mathbf{w}$ is $A = \{0, \ldots, k-1\}$, the Rauzy graph G_0 is as in Figure 3.2 so that for all 0-segments p, we have $\lambda_L(p) \in A$. Consequently, we have $A_0 = A$ and we consider that for all $a \in A_0$, we have

$$\lambda_L \circ \theta_0(a) = a.$$

Corollary 3.1.5 implies that when $\mathbf{w}$ has a sub-linear complexity, there is an integer k such that $\mathbf{A} = \bigcup_{n \in \mathbb{N}} A_n = \{0, 1, \ldots, k-1\}$.

Example 3.2.3. Let $\mathbf{t}$ be the Thue-Morse sequence. The Rauzy graph $G_2(\mathbf{t})$ is represented in Figure 3.3 (labelled with left labels). The 2-segments are

$$\begin{array}{ll} \theta_2(0) = 01 \to 11 \to 10 & \theta_2(2) = 01 \to 10 \\ \theta_2(1) = 10 \to 00 \to 01 & \theta_2(3) = 10 \to 01 \end{array}$$

and if we define θ_3 (see Example 3.1.3) by

$$\begin{array}{ll} \theta_3(0) = 011 \to 110 \to 101 & \theta_3(4) = 010 \to 101 \\ \theta_3(1) = 011 \to 110 \to 100 & \theta_3(5) = 010 \to 100 \\ \theta_3(2) = 100 \to 001 \to 010 & \theta_3(6) = 101 \to 010 \\ \theta_3(3) = 100 \to 001 \to 011 & \theta_3(7) = 101 \to 011 \end{array}$$

we have

$$\begin{aligned} \psi_{2,L} \circ \theta_3(0) = \psi_{2,L} \circ \theta_3(1) = \theta_2(0) \\ \psi_{2,L} \circ \theta_3(2) = \psi_{2,L} \circ \theta_3(3) = \theta_2(1) \\ \psi_{2,L} \circ \theta_3(4) = \psi_{2,L} \circ \theta_3(5) = \theta_2(2) \\ \psi_{2,L} \circ \theta_3(6) = \psi_{2,L} \circ \theta_3(7) = \theta_2(3) \end{aligned}$$

and so

$$\sigma_2 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 0 \\ 2 \mapsto 1 \\ 3 \mapsto 1 \\ 4 \mapsto 2 \\ 5 \mapsto 2 \\ 6 \mapsto 3 \\ 7 \mapsto 3 \end{cases}$$

Remark 3.2.4. It is a consequence of the constructions described above that $|\sigma_n(i)| \geq 2$ means that there are at least two n-segments occurring in $\psi_{n,L} \circ \theta_{n+1}(i)$. Suppose that p and q are such n-segments with $i(p) = o(q)$. Then $\sigma_n(i) \in A_n^* \theta_n^{-1}(p) \theta_n^{-1}(q) A_n^*$ and as any interior vertex of a $(n+1)$-segment cannot be left special, the only possibility is that the vertex $i(p) = o(q)$ is a bispecial vertex such that its right extension which is an interior vertex of $\theta_{n+1}(i)$ is not left special. Hence if a Rauzy graph $G_n(\mathbf{w})$ does not contain any bispecial vertex, we have $\mathcal{A}_n = \psi_{n,L}(\mathcal{A}_{n+1})$ and the morphism σ_n is simply a bijective and letter-to-letter morphism.

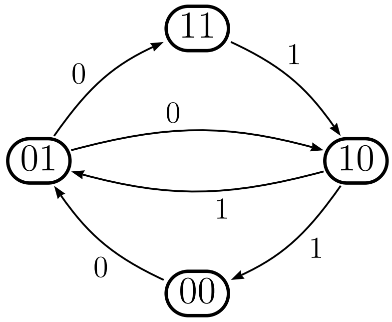


Figure 3.3: Rauzy graph of order 2 (with left labels) of the Thue-Morse sequence.

Also, observe that if p is a constant n-segment then for all positive integers k and all $(n+k)$-segment q_k such that $\psi_{n,L} \cdots \psi_{n+k-1,L}(q_k) = p$ we have $\sigma_{n+k-1}(\theta_{n+k}^{-1}(q_k))| = \theta_{n+k-1}^{-1}(q_{k-1})$ so the sequence $(\theta_{n+k}^{-1}(q_k))_{k\in\mathbb{N}}$ is the tail of a bounded sequence of $(\sigma_n)_{n\in\mathbb{N}}$ (Definition 1.3.13).

Remark 3.2.5. In the general case, morphisms in Definition 3.2.1 might be uninteresting. Indeed consider the case of sequences with maximal complexity (like the *Champernowne sequence*; see [IS75] for instance). As $L(\mathbf{w}) = A^*$ for these sequences, all factors are left special and so all edges of G_n are n-segments. For all n, the morphism σ_n is therefore uniform of length 1 so $|\sigma_0\sigma_1 \cdots \sigma_n(a)| = 1$ for all n and all letters a. However the construction of Definition 3.2.1 makes sense as soon as there is a sequence $(a_n)_{n\in\mathbb{N}}$ of letters $a_n \in A_n$ such that $|\sigma_0 \cdots \sigma_n(a_{n+1})|$ tends to infinity as n increases. Indeed, in this case, $L(\mathbf{w}) = \bigcup_{n\in\mathbb{N}} L(\sigma_0 \cdots \sigma_n(a_{n+1}))$ (due to the uniform recurrence). We can easily see that for $|\sigma_0 \cdots \sigma_n(a_{n+1})|$ to converge to infinity for at least one sequence of letters $(a_n)_{n\in\mathbb{N}}$, $a_n \in A_n$, it is sufficient that the sequence $\left(\frac{p_{\mathbf{w}}(n)}{\mathrm{Card}(\mathcal{A}_n)}\right)_{n\in\mathbb{N}}$ is unbounded. Since $\mathrm{Card}(\mathcal{A}_n) \leq \mathrm{Card}(A)\,(p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n))$ (see Remark 3.1.4), it is also sufficient that $\limsup_{n\to+\infty} \frac{p(n)}{p(n+1)-p(n)} = +\infty$ and so that $\liminf_{n\to+\infty} \frac{p(n+1)}{p(n)} = 1$. Note that sequences with an at most polynomial complexity satisfy this property although for sequences with higher complexity, it is not always the case.

The next lemma shows that for sequences with sub-linear complexity, the construction of Definition 3.2.1 is particularly efficient since it always provides a finite set of morphisms. Indeed, the lemma improves Lemma 3.1.8 stating that when the sequence has a sub-linear complexity, the number of n-segments occurring in an $(n+1)$-segment is bounded. In this case, we will construct only a finite number of morphisms σ_n because this only gives rise to morphisms of bounded length over bounded alphabets A_n (Corollary 3.1.5). Consequently, this will prove that the set $S = \{\sigma_n \mid n \in \mathbb{N}\}$ is finite.

Lemma 3.2.6 (Ferenczi [Fer96]). *Let* $\mathbf{w}$ *be an aperiodic sequence over an alphabet* A. *If* $\mathbf{w}$ *has an at most linear complexity, then for any* $(n+1)$-*segment* p *of* $\mathbf{w}$, $\psi_{n,L}(p)$ *is a bounded concatenation of* n-*segments and the decomposition is unique.*

Proof. The uniqueness of the decomposition has already been proved in Lemma 3.1.8. Let us prove that it is bounded. Let K be such that $p_{\mathbf{w}}(n+1)-p_{\mathbf{w}}(n) \leq K$ for all n (Theorem 1.2.3). Consider a $(n+1)$-segment $p \in \mathcal{A}_{n+1}$ (we know it exists since $\mathbf{w}$ is aperiodic). The number of n-segments in $\psi_{n,L}(p)$ is equal to 1 plus the number of vertices va in p, $a \in A$, such that v is a left special factor of $\mathbf{w}$ and va not. Moreover, as these vertices are not left special, the path p cannot pass through one of them more than once. Since there exist at most K left special vertices v in $G_n(\mathbf{w})$, there exist at most $K\mathrm{Card}(A)$ vertices va as considered just above. Consequently, the number of n-segments in $\psi_{n,L}(p)$ is bounded by $1 + K\mathrm{Card}(A)$. □

Introduction of a new symbol ♯

Let $\mathbf{w}$ be a one-sided uniformly recurrent sequence over A with sub-linear complexity. To get the S-adic representation of Theorem 3.0.3 we need to consider a new symbol $\sharp \notin A$ and the one-sided sequence $\mathbf{w}' = \sharp\mathbf{w}$. It is obvious that $\mathbf{w}'$ is not recurrent and that for all n we have $p_{\mathbf{w}'}(n) = p_{\mathbf{w}}(n)+1$. Moreover, since $\mathbf{w}$ is recurrent, all prefixes of $\mathbf{w}$ are left special factors of $\mathbf{w}'$ so for all n, $\mathrm{Card}(LS_n(\mathbf{w}')) \geq \mathrm{Card}(LS_n(\mathbf{w}))$. However we still have

$$p_{\mathbf{w}'}(n+1) - p_{\mathbf{w}'}(n) = \sum_{u \in LS_n(\mathbf{w})} (\delta^-(u) - 1).$$

Remark 3.2.7. Considering the sequence $\mathbf{w}'$ instead of $\mathbf{w}$ does not change much the shape of the Rauzy graphs (hence neither the sets of n-segments and of n-circuits). Indeed, it simply corresponds to highlighting a particular vertex (the prefix of $\mathbf{w}$ of each length n) by adding to it an incoming edge. Consequently, that can only split some n-segments of $\mathbf{w}$ into two n-segments of $\mathbf{w}'$ and it can add some possibilities for the choice of the sequence $(v_n)_{n\in\mathbb{N}}$ of Lemma 3.1.14.

However, considering $\mathbf{w}'$ instead of $\mathbf{w}$ has a significant consequence on the morphisms σ_n of Definition 3.2.1. Indeed, it implies that if $A = \{0,1,\dots,k-1\}$ we do not have $A_0 = A$ anymore but $A_0 = \{0,1,\dots,k\}$ and there is a letter $a_\sharp$ in A_0 such that $\lambda_L \circ \theta_0(a_\sharp) = \sharp$. But, since the symbol $\sharp$ does not occur in the label of any 1-segment, we have $\sigma_0(A_1) \subset A_0 \setminus A_0^* a_\sharp A_0^*$ so we can suppose that for all $a \in A_0 \setminus \{a_\sharp\}$, we have

$$\lambda_L \circ \theta_0(a) = a.$$

Example 3.2.8. If we consider the Thue-Morse sequence $\mathbf{t}$, the Rauzy graph $G_3(\mathbf{t}')$ is represented in Figure 3.4.

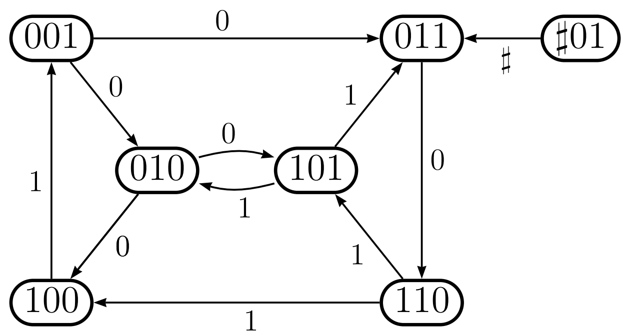


Figure 3.4: Rauzy graph of order 3 (with left labels) of $\sharp\mathbf{t}$ where $\mathbf{t}$ is the Thue-Morse sequence.

In all what follows we consider the notations introduced in Definition 3.2.1 such that the sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}}$ is built upon the set of n-segments of $\mathbf{w}'$.

3.3 S-adicity using n-circuits

In this section we prove Theorem 3.0.1. To that aim, we consider for all n the subsets C_n of A_n^* such that for all $c \in C_n$, $\theta_n(c)$ is an n-circuit starting from a particular vertex. This kind of choice of subsets of A_n^* will also be used in Chapter 4 and Chapter 5 to obtain an S-adic characterization of minimal subshift with first difference of complexity bounded by 2.

3.3.1 Morphisms over the set of n-circuits

In this section we explicitly define the morphisms of the S-adic representation of Theorem 3.0.1.

Definition 3.3.1 (Definition of the morphisms γ_n)**.** Let A be an alphabet, $\sharp \notin A$, $\mathbf{w}$ be a one-sided sequence over A and $\mathbf{w}' = \sharp\mathbf{w}$. For all n, we also let p_n denote the prefix of length n of $\mathbf{w}$. Since $\mathbf{w}$ is recurrent, all its prefixes are left special in $\mathbf{w}'$ thus $(p_n)_{n\in\mathbb{N}}$ corresponds to a sequence of left special factors of $\mathbf{w}'$ as in Lemma 3.1.14. For each non-negative integers n, let $\mathcal{C}_n$ be the set of allowed n-circuits starting from p_n. Now define the alphabet $C_n = \{0, 1, \ldots, \mathrm{Card}(\mathcal{C}_n) - 1\}$ and consider a bijection $\vartheta_n : C_n \to \mathcal{C}_n$. We can extend ϑ_n to an isomorphism by putting $\vartheta_n(ab) = \vartheta_n(a)\vartheta_n(b)$ for all letters

a, b in C_n. Then, for all n, Lemma 3.1.13 allows us define $\gamma_n : C_{n+1}^* \to C_n^*$ as the unique morphism satisfying

$$\vartheta_n \circ \gamma_n = \psi_{n,L} \circ \vartheta_{n+1}.$$

Observe that for all n we actually have

$$\gamma_n = \vartheta_n^{-1} \circ \theta_n \circ \sigma_n \circ \theta_{n+1}^{-1} \circ \vartheta_{n+1}.$$

Remark 3.3.2. As for n-segments, it is a direct consequence of Definition 3.3.1 that if a Rauzy graph G_n does not contain any bispecial vertices, the morphism γ_n is simply a bijective and letter-to-letter morphism. This morphism only depends on the differences that could exist between ϑ_n and ϑ_{n+1}. Furthermore, it is easily seen that when p_n is not a bispecial vertex, the morphism γ_n is a bijective and letter-to-letter morphism.

Remark 3.3.3. It is easily seen that the 0-circuits of $\mathbf{w}'$ correspond to its 0-segments. Thus, as for n-segments we have $C_0 = \{0, 1, \ldots, k\}$ whenever $A = \{0, \ldots, k-1\}$ and no letter c of C_1 is such that the letter $c_\sharp$ occurs in $\gamma_0(c)$ where $\lambda_L \circ \vartheta_0(c_\sharp) = \sharp$. Consequently, we consider that ϑ_0 is such that for all $c \in C_0 \setminus \{c_\sharp\}$,

$$\lambda_L \circ \vartheta_0(c) = c.$$

Example 3.3.4. Let $\mathbf{t}$ be the Thue-Morse sequence. The Rauzy graphs $G_2(\mathbf{t})$ and $G_3(\mathbf{t})$ are represented at Figures 3.3 and 3.1. Let us compute the morphism γ_2 of Definition 3.3.1 for this particular sequence. Since the left return words to 01 and 011 are respectively

$$\begin{aligned} \mathrm{LRW}_{\mathbf{t}}(01) &= \{0110, 01, 010, 011\} \\ \mathrm{LRW}_{\mathbf{t}}(011) &= \{011010, 011001, 0110, 01101001\} \end{aligned}$$

the allowed 2-circuits starting from 01 are

$$\begin{aligned} &\vartheta_2(0) = 01 \to 10 \to 00 \to 01 \qquad &\vartheta_2(2) = 01 \to 11 \to 10 \to 00 \to 01 \\ &\vartheta_2(1) = 01 \to 11 \to 10 \to 01 \qquad &\vartheta_2(3) = 01 \to 10 \to 01 \end{aligned}$$

and the allowed 3-circuits starting from 011 are

$$\begin{aligned} &\vartheta_3(0) = 011 \to 110 \to 100 \to 001 \to 010 \to 101 \to 011 \\ &\vartheta_3(1) = 011 \to 110 \to 101 \to 010 \to 100 \to 001 \to 011 \\ &\vartheta_3(2) = 011 \to 110 \to 100 \to 001 \to 011 \\ &\vartheta_3(3) = 011 \to 110 \to 101 \to 010 \to 100 \to 001 \to 010 \to 101 \to 011 \end{aligned}$$

By applying Definition 3.3.1 we obtain

$$\gamma_2 : \begin{cases} 0 \mapsto 23 \\ 1 \mapsto 10 \\ 2 \mapsto 2 \\ 3 \mapsto 103 \end{cases} .$$

3.3.2 Proof of Theorem 3.0.1

Let $(\gamma_n)_{n\in\mathbb{N}}$ be the sequence of morphisms of Definition 3.3.1. We first prove that it is indeed a directive word of $\mathbf{w}$. Then, we prove that there is a contraction of $(\gamma_n)_{n\in\mathbb{N}}$ (Definition 1.3.9) that contains only strongly primitive morphisms that are also left proper which is the first part of the theorem. Then, we show how we can slightly modify the contraction to get morphisms that are left and right proper.

Lemma 3.3.5. *Let* $\mathbf{w}$ *be a uniformly recurrent sequence. Then, the sequence of morphisms* $(\gamma_n)_{n\in\mathbb{N}}$ *of Definition 3.3.1 is a directive word of* $\mathbf{w}$*.*

Proof. By construction, for all n and all letters $c \in C_{n+1}$ the word $\gamma_0 \cdots \gamma_n(c)$ belongs to $L(\mathbf{w})$. Moreover, since for all n and $c \in C_n$, $o(\vartheta_n(c)) = p_n$ there is a sequence of letters $(c_n)_{n\in\mathbb{N}}$, $c_n \in \mathcal{C}_n$, such that $\vartheta_n(c_n)$ is labelled by a prefix of $\mathbf{w}$. Consequently, for such a sequence the word $\gamma_0 \cdots \gamma_n(c_{n+1})$ is a prefix of $\mathbf{w}$ for all n. To conclude the proof, we only have to notice that Corollary 3.1.17 implies that $(\gamma_n)_{n\in\mathbb{N}}$ is everywhere growing. ☐

Proposition 3.3.6. *A one-sided sequence* $\mathbf{w}$ *over an alphabet* A *is uniformly recurrent if and only if it is primitive and left proper* S*-adic. In particular, if* $\mathbf{w}$ *does not have a sub-linear complexity, then* $\mathrm{Card}(S) = +\infty$*.*

Proof. The sufficient part is simply a consequence of Proposition 2.1.21 and the "in particular" part is a consequence of Proposition 2.3.1 page 62 and Proposition 2.3.3 page 62. Let us prove that the condition is necessary.

Let $(\gamma_n)_{n\in\mathbb{N}}$ be the sequence of morphisms as defined in Definition 3.3.1 and let us prove that there is a contraction $(\Gamma_n)_{n\in\mathbb{N}}$ of $(\gamma_n)_{n\in\mathbb{N}}$ such that for all n, Γ_n is strongly primitive and left proper.

First, let us prove the strong primitivity. Let r be an integer and let ℓ_r be the maximal length of a r-circuit. Since $\mathbf{w}$ is uniformly recurrent, there is an integer $M_r > r$ such that all factors of $\mathbf{w}$ of length at least M_r contain all factors of $\mathbf{w}$ of length at most $r + \ell_r$. Let $s > r$ be an integer such that all s-circuits have length at least M_r. For all letters c, d in C_r, $\lambda \circ \vartheta_r(c)$ is not a factor of $\lambda \circ \vartheta_r(d)$. Consequently, for all s-circuits q, all r-circuits occur in

$\psi_{r,L} \circ \cdots \circ \psi_{s-1,L}(q)$ so all letters of C_r occur in $\gamma_r \cdots \gamma_{s-1}(d)$ for all $d \in C_s$. Let us denote by $(\gamma'_n : \tilde{C}^*_{n+1} \to \tilde{C}^*_n)_{n\in\mathbb{N}}$ a contraction of $(\gamma_n)_{n\in\mathbb{N}}$ such that all morphisms are strongly primitive.

Now let us prove that there is a contraction of $(\gamma'_n)_{n\in\mathbb{N}}$ such that all morphisms are left proper. Let r be a positive integer. By construction, there is a unique letter $c \in \tilde{C}_r$ such that $\gamma'_0 \cdots \gamma'_{r-1}(c)$ is a prefix of $\mathbf{w}$. Let $s > r$ be such that the alphabet $\tilde{C}_s$ corresponds to m-circuits with $m > |\gamma'_0 \cdots \gamma'_{r-1}(c)|$. By definition, these circuits are starting from a prefix of $\mathbf{w}$ of length m. From Proposition 1.5.5, we deduce that the image of all these circuits under the appropriate composition of function $\psi_{n,L}$ admit the circuit corresponding to c as a prefix. Consequently, we have $\gamma'_r \cdots \gamma'_{s-1}(\tilde{C}_s) \subset c\tilde{C}^*_r$ and this concludes the proof. □

To end the proof of Theorem 3.0.1, we have to introduce the following trick. If $\sigma : A^* \to B^*$ is a left proper morphism such that $\sigma(A) \subset bB^*$ for a letter $b \in B$, we let $\sigma^{(R)} : A^* \to B^*$ denote the right proper morphism such that for all $a \in A$, $\sigma^{(R)}(a) = ub$ whenever $\sigma(a) = bu$. We call $\sigma^{(R)}$ the *right conjugate* of σ.

Lemma 3.3.7. *Let $\sigma : A^* \to B^*$ be a left proper morphism such that $\sigma(A) \subset bB^*$ for a letter $b \in B$. Let also $\mathbf{w}$ be a sequence in $A^{\mathbb{N}}$. Then we have*

$$\sigma(\mathbf{w}) = b\sigma^{(R)}(\mathbf{w}).$$

In particular, if $\sigma(\mathbf{w})$ is recurrent, then $L(\sigma(\mathbf{w})) = L\left(\sigma^{(R)}(\mathbf{w})\right)$.

Proof. Indeed, from the shape of σ we have

$$\sigma(\mathbf{w}) = \underbrace{bu_1}_{\sigma(\mathbf{w}_0)} \underbrace{bu_2}_{\sigma(\mathbf{w}_1)} \underbrace{bu_3}_{\sigma(\mathbf{w}_2)} \cdots$$

for some word $u_1, u_2, u_3, \cdots \in B^*$. This sequence can then be decomposed into the images of $\sigma^{(R)}$ by

$$b \underbrace{u_1 b}_{\sigma^{(R)}(\mathbf{w}_0)} \underbrace{u_2 b}_{\sigma^{(R)}(\mathbf{w}_1)} \underbrace{u_3 b}_{\sigma^{(R)}(\mathbf{w}_2)} \cdots$$

so $\sigma(\mathbf{w}) = b\sigma^{(R)}(\mathbf{w})$ and we obviously have $L\left(\sigma^{(R)}(\mathbf{w})\right) \subset L(\sigma(\mathbf{w}))$. From $\sigma(\mathbf{w}) = b\sigma^{(R)}(\mathbf{w})$, we know that the only factors of $\sigma(\mathbf{w})$ that might not occur in $\sigma^{(R)}(\mathbf{w})$ are the prefixes. But, for recurrent sequences, all prefixes also occur later in the sequence. Thus, all prefixes of $\sigma(\mathbf{w})$ occur in some $\sigma^{(R)}(\mathbf{w}_{[i,j]})$ for some i, j such that $0 < i < j$ so the result holds. □

The next lemma ends the proof of Theorem 3.0.1. It states that for subshifts, we can replace the condition "left proper" of the theorem by the condition "proper".

Proposition 3.3.8. *An aperiodic subshift* (X,T) *over an alphabet* A *is minimal if and only if it is primitive and proper* S*-adic. In particular, if* (X,T) *does not have a sub-linear complexity, then* $\mathrm{Card}(S) = +\infty$.

Proof. The proof of the sufficient part and of the "in particular part" is the same as in Proposition 3.3.6. Moreover, we can also use that proposition to consider a directive word $(\Gamma_n)_{n\in\mathbb{N}}$ of (X,T) such that all morphisms Γ_n are strongly primitive and left proper.

Now let us consider the sequence of morphisms $(\varrho_n)_{n\in\mathbb{N}}$ such that for all n

$$\varrho_n = \Gamma_{2n}\Gamma_{2n+1}^{(R)}.$$

For all n, ϱ_n is clearly strongly primitive and proper. Therefore we only have to prove that $(\varrho_n)_{n\in\mathbb{N}}$ is a directive word of (X,T) which is obvious since $(\Gamma_n)_{n\in\mathbb{N}}$ is a directive word of (X,T) and Lemma 3.3.7 states that replacing a left proper morphism by its right conjugate does not change the language. ☐

Remark 3.3.9. Propositions 3.3.6 and 3.3.8 could be obtained easily using return words. For two-sided sequences, return words would also allow us to replace "left proper" by "proper" in the theorem. However, we think that Rauzy graphs can provide much more information than return words and our goal is therefore to understand how they evolve to get properties on the S-adic representations. Consequently, we prefer to keep working with them.

3.4 S-adicity using bounded concatenations of n-segments

Although some properties of the S-adic representation of Theorem 3.0.1 seem to be interesting, a bad thing is that the construction often yields to infinite sets of morphisms (even for sequences with sub-linear complexity). In this section we consider the action of the morphisms σ_n of Definition 3.2.1 over other subsets of A_n^* that allow us to prove Theorem 3.0.3. This makes us lose the almost primitivity and the left proper property of the directive word (as it is the case in Theorem 3.0.1) but this provides other interesting properties. In particular, with these subsets, we are always ensured to build a finite set of morphisms for sequences with sub-linear complexity.

3.4.1 Some preliminary lemmas

Before proving Theorem 3.0.3, we need some lemmas about the sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}}$ that will allow us to consider some particular subsets of A_n^*. The first one states that in any allowed path of a Rauzy graph of large order n, the number of consecutive constant n-segments is bounded.

Lemma 3.4.1 (Ferenczi [Fer96]). *Let $\mathbf{w}$ be a uniformly recurrent sequence over an alphabet A. If $\mathbf{w}$ has an at most linear complexity, then for n large enough, in any path in $G_n(\mathbf{w})$, the number of consecutive constant n-segments is bounded by a constant $C_{\mathbf{w}}$. In particular, this also holds for $\mathbf{w}' = \sharp\mathbf{w}$ with $\sharp \notin A$.*

Proof. If the result holds for $\mathbf{w}$, it is a direct consequence of Remark 3.2.7 that it also holds for $\mathbf{w}'$. Let K be such that $p(n+1) - p(n) \leq K$ for all n (Theorem 1.2.3). As any edge of $G_n(\mathbf{w})$ occurs in at least one n-segment, any finite path in $G_n(\mathbf{w})$ can be decomposed into a finite number of n-segments, the first one and the last one being possibly truncated. In this decomposition, some segments may be constant and so have bounded length, say by ℓ (Remark 3.1.10). Now if a path p composed of consecutive constant n-segments has length greater than $K\ell$, the path contains at least $K+1$ occurrences of left special vertices. Consequently, some vertices v_i and v_j of p are equal and the graph contains an n-circuit whose length is smaller than $K\ell$. As $\mathbf{w}$ is uniformly recurrent, by Corollary 3.1.17, this is impossible for n large enough. □

The previous lemma allows us to define new families of n-segments depending on their length. Indeed, we already know that constant n-segments have bounded length, say by ℓ and that some non-constant n-segment become very long (Remark 3.1.10). But, there might also exist some non-constant n-segment with very short length. Indeed, if for instance $\mathrm{Card}(A) = 2$ and p is a constant n-segment such that $o(p)$ is not bispecial and $i(p)$ is a strong bispecial vertex, the two right extensions of $i(p)$ are left special. Consequently, this provides two $(n+1)$-segments p_1 and p_2 such that $\psi_{n,L}(p_1) = \psi_{n,L}(p_2) = p$. But, the definition of constant n-segment only implies that one of these $(n+1)$-segment is constant so one of them might not be constant but having a small length. To build the S-adic representation of Theorem 3.0.3, we need to determine a family of segments that are always "long".

Definition 3.4.2. Let $\mathbf{w}$ be a sequence with sub-linear complexity and let $C_{\mathbf{w}}$ be the constant of Lemma 3.4.1. Let also ℓ be the maximal length of a constant segment (all orders n included). An n-segment is said to be *short* if it has length at most $C_{\mathbf{w}}\ell$, otherwise it said to be *long*. We also let N denote the smallest integer such that all constant segments already exist (i.e., for all $n \geq N$ and all constant n-segments p there is a constant N-segment q such that $\psi_{N,L} \circ \cdots \psi_{n-1,L}(p) = q$) and such that there exist some long n-segments (such an integer exists by Remark 3.1.10).

The next lemma states that Lemma 3.4.1 also holds for short segments; the proof is based on the fact that for n large enough, short n-segment can only arise from concatenation of constant segments. Thus, if there is an allowed concatenation of short m-segments (for $m > n$) which is long, its projection into G_n (through the functions $\psi_{k,L}$) has the same length and is an allowed concatenation of constant n-segments.

Lemma 3.4.3 (Ferenczi [Fer96]). *Let $\mathbf{w}$ be a uniformly recurrent sequence over an alphabet A. If $\mathbf{w}$ has an at most linear complexity, then for n large enough, any allowed concatenation of short n-segments has length bounded by $C_{\mathbf{w}}\ell$ where $C_{\mathbf{w}}$ is the constant of Lemma 3.4.1 and ℓ is the maximal length of a constant segment. In particular, this also holds for $\mathbf{w}' = \sharp\mathbf{w}$ with $\sharp \notin A$.*

Now we can define the directive word of Theorem 3.0.3.

Definition 3.4.4 (Definition of the morphisms τ_n). Let $\mathbf{w}$ be a one-sided sequence over A and let $\mathbf{w}' = \sharp\mathbf{w}$ with $\sharp \notin A$. For all $n \geq N$, we let $\mathcal{A}_{n,\text{short}}$ and $\mathcal{A}_{n,\text{long}}$ respectively denote the set of short and long n-segments of $\mathbf{w}'$. Then, we define the set $\mathcal{B}_n$ as the set of allowed path in $\mathcal{P}_n \cap \mathcal{A}^*_{n,\text{short}}\mathcal{A}_{n,\text{long}}\mathcal{A}^*_{n,\text{short}}$. From Lemma 3.4.3, we deduce that there is a constant C such that

$$\mathcal{B}_n \subset \mathcal{A}^{\leq C}_{n,\text{short}}\mathcal{A}_{n,\text{long}}\mathcal{A}^{\leq C}_{n,\text{short}}$$

so we have $\text{Card}(\mathcal{B}_n) < +\infty$. Observe that Lemma 3.1.8 implies that any path in $\mathcal{B}_{n+1}$ can be decomposed into paths of $\mathcal{B}_n$. However, the decomposition might not be unique in this case. Indeed, consider a path $p \in \mathcal{B}_{n+1}$ such that $\psi_{n,L}(p)$ can be decomposed into $l_1s_1s_2l_2$ with $l_1, l_2 \in \mathcal{A}_{n,\text{long}}$ and $s_1, s_2 \in \mathcal{A}_{n,\text{short}}$. This means that we can decompose $\psi_{n,L}(p)$ into two elements q_1 and q_2 of $\mathcal{B}_n$ in three different ways: $(q_1, q_2) \in \{(l_1, s_1s_2l_2), (l_1s_1, s_2l_2), (l_1s_1s_2, l_2)\}$. This is a problem when we want to define morphisms because this means that the letter that corresponds to the path p of $\mathcal{B}_{n+1}$ admits several images. However, any choice of decomposition yields to a morphism, hence to an S-adic representation.

We then consider a bijection $\Theta_n : B_n \to \mathcal{B}_n$ with $B_n = \{0, \ldots, \text{Card}(\mathcal{B}_n) - 1\}$ and define $\tau_n : B^*_{n+1} \to B^*_n$ as a morphism (there might exist several available morphisms) such that

$$\Theta_n \circ \tau_n = \psi_{n,L} \circ \Theta_{n+1}.$$

As explained just above, several choices can be made when decomposing a path of $\mathcal{B}_{n+1}$ into paths of $\mathcal{B}_n$. This comes from the fact that Θ_n is a bijection between B_n and $\mathcal{B}_n$ but not between B_n^* and $\mathcal{B}_n^*$. Consequently, we cannot write $\tau_n = \Theta_n^{-1} \circ \theta_n \circ \sigma_n \circ \theta_{n+1}^{-1} \circ \Theta_{n+1}$ as we did for morphisms σ_n and γ_n.

For all n, we let $A_{n,\text{long}}$ and $A_{n,\text{short}}$ respectively denote $\theta_n^{-1}(A_{n,\text{long}})$ and $\theta_n^{-1}(A_{n,\text{short}})$.

Remark 3.4.5. The first morphism in the new directive word is τ_N and except if there do not exist any constant segments (in that case we have $N = 0$ and $\sigma_n = \tau_n$ for all $n \geq 0$), we obviously might have $B_N \neq A$. Consequently, the sequence $(\tau_n)_{n \geq N}$ cannot be a directive word of $\mathbf{w}$. We need to consider a new morphism $\kappa = \lambda_L \circ \Theta_N$ and the sequence of morphisms $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$.

The next lemma shows that the directive word $(\kappa, \tau_N, \tau_{N+1}, \dots)$ is everywhere growing.

Lemma 3.4.6 (Ferenczi [Fer96])**.** *Let A be an alphabet and $\mathbf{w} \in A^{\mathbb{N}}$ be a uniformly recurrent sequence with sub-linear complexity. Let $(\kappa, \tau_N, \tau_{N+1}, \dots)$ be the directive word of Definition 3.4.4. For all integers ℓ, there is an integer $n_\ell \geq N$ such that for all $n \geq n_\ell$ and all letters $b \in B_{n+1}$, $|\kappa \tau_N \cdots \tau_n(b)| \geq \ell$.*

Proof. By definition, for all letters $b \in B_{n+1}$, $\Theta_{n+1}(b)$ contains a long $(n+1)$-segment so has length greater than $C_{\mathbf{w}'}\ell$. We also have $|\kappa \tau_N \cdots \tau_n(b)| = |\Theta_{n+1}(b)|$. Moreover, long segments can only arise from long segments of smaller order; otherwise that would contradict Lemma 3.4.3. Finally, the fact that the long segments cannot be constant (due to their length) ensures that they will occur as proper subpaths of long segments of larger order. These long segments will therefore have length greater than $2C_{\mathbf{w}'}\ell$. Continuing this way, for all k we can find larger and larger orders m such that all long m-segments have length greater than $kC_{\mathbf{w}'}\ell$ which concludes the proof. □

3.4.2 Proof of Theorem 3.0.3

Let us split the proof of Theorem 3.0.3 into several propositions. The first one proves that the sequence of morphisms $(\kappa, \tau_N, \tau_{N+1}, \dots)$ of Definition 3.4.4 is indeed a directive word of $\mathbf{w}$ and that the set of morphisms occurring in it is finite. Then we give one proposition for each property of Definition 3.0.4. When the sequence $\mathbf{w}' = \sharp\mathbf{w}$ does not admit constant segments (for the last part of Theorem 3.0.3), we obviously have $N = 0$ and $\tau_n = \sigma_n$ for all n.

Proposition 3.4.7 (S-adicity)**.** *Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over A with a sub-linear complexity. The set of morphisms $\{\tau_n \mid n \geq N\}$ of Definition 3.4.4 is finite and $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ is a directive word of $\mathbf{w}$.*

Proof. First let us prove that the set of morphisms in $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ is finite. We already know by definition that for all $n \geq N$, the alphabet B_n is finite so we have to prove that the images of letters under τ_n have bounded length. This is obvious since for any letter $b \in B_{n+1}$, $\Theta_{n+1}(b)$ can be decomposed into a bounded number of $(n+1)$-segments and Lemma 3.2.6 implies that each of them can be decomposed into a bounded number of n-segments.

Now let us prove that $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ is a directive word of $\mathbf{w}$. For all n, the prefix of $\mathbf{w}$ is left special in $\mathbf{w}'$. Consequently, for all n there are some n-segments p of $\mathbf{w}'$ such that $o(p) = \mathbf{w}_{[0,n-1]}$ and some of these n-segments have a left label which is a prefix of $\mathbf{w}$. For all $n \geq N$, let $P_n \subset B_n$ denote the set of letters p such that $o(\Theta_n(p)) = \mathbf{w}_{[0,n-1]}$. We have $\tau_n(P_{n+1}) \in P_n^+ B_n^*$ for all $n \geq N$. Let $(p_n)_{n \geq N}$ be a sequence of letters $p_n \in P_n$ such that $\lambda_L \circ \Theta_n(p_n)$ is a prefix of $\mathbf{w}$ and $\tau_n(p_{n+1}) \in p_n B_n^*$ (it is a consequence of the constructions that such a sequence exists). Since the directive word $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ is everywhere growing (Lemma 3.4.6), we have

$$\mathbf{w} = \lim_{n \to +\infty} \kappa \tau_N \tau_{N+1} \cdots \tau_n(p_{n+1}^\omega)$$

and this concludes the proof. □

The proofs of the properties in Theorem 3.0.3 are mostly based on the following lemma.

Lemma 3.4.8. *Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over A with a sub-linear complexity and let $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ be the directive word of Definition 3.4.4. Let $n \geq N$ and $b \in B_n$ such that $\Theta_n(b)$ can be decomposed into $p_1 \cdots p_i l p_{i+1} \cdots p_k$ with $l \in \mathcal{A}_{n,\text{long}}$ and $p_i \in \mathcal{A}_{n,\text{short}}$ for all i. If there is a letter $c \in B_{b+1}$ such that $\tau_n(c) \in B_n^+ b B_n^*$, then the right extension of $o(l)$ that is an interior vertex of $\Theta_{n+1}(c)$ is not left special and for all $j = 1, \dots, i$, the right extension of $o(p_i)$ that is an interior vertex of $\Theta_{n+1}(c)$ is not left special either. Similarly, if $\tau_n(c) \in B_n^* b B_n^+$, then the right extension of $i(l)$ that is an interior vertex of $\Theta_{n+1}(c)$ is not left special and for all $j = i+1, \dots, k$, the right extension of $i(p_i)$ that is an interior vertex of $\Theta_{n+1}(c)$ is not left special either.*

Proof. The result is almost trivial. Indeed, if $\tau_n(c) \in B_n^* b_1 b_2 B_n^*$ for some letters $c \in B_{n+1}$ and $b_1 b_2 \in B_n$, the definition of the alphabet B_n implies that the short $(n+1)$-segment that might occur in $\Theta_{n+1}(c)$ can only occur at the extremities. Consequently, the short n-segments at the beginning of $\Theta_n(b_2)$ and at the end of $\Theta_n(b_1)$ cannot keep being short $(n+1)$-segments. In other words, their extremities in $\Theta_{n+1}(c)$ must be non left special. □

Proposition 3.4.9 (Property 1). *Let* $\mathbf{w}$ *be an aperiodic and uniformly recurrent sequence over* A *with a sub-linear complexity and let* $(\kappa, \tau_N, \tau_{N+1}, \dots)$ *be the directive word of Definition 3.4.4. For all integers* $n \geq N$ *and all letters* $b \in B_n$ *and* c *in* B_{n+1}*,* $\tau_n(c) \notin B_n^* b B_n^* b B_n^*$.

Proof. This is an almost direct consequence of Lemma 3.4.8. Suppose $\tau_n(c) = ubvbw$ with $b \in B_n$ and $u, v, w \in B_n^*$. From Lemma 3.4.8, the subpath q of $\Theta_{n+1}(c)$ such that $\psi_{n,L}(q) = \Theta_n(bvb)$ does not contain any left special vertex. This path is therefore inaccessible from vertices that do not compose it so G_{n+1} is not strongly connected which is a contradiction with the recurrence of $\mathbf{w}$ (see Remark 1.5.2). □

Proposition 3.4.10 (Property 2). *Let* $\mathbf{w}$ *be an aperiodic and uniformly recurrent sequence over* A *with a sub-linear complexity and let* $(\kappa, \tau_N, \tau_{N+1}, \dots)$ *be the directive word of Definition 3.4.4. For all integers* $n \geq N$*, if there is* $b \in B_n$*,* $u = u_1 u_2 \cdots u_\ell \in B_n^+$ *and* $c \in B_{n+1}$ *such that* $\tau_n(c) \in ubB_n^*$*, then for all letters* $d \in B_{n+1}$*,* $\tau_n(d) \in (B_n^+ \setminus B_n^* b B_n^*) \cup (B_n u_2 \cdots u_\ell b B_n^*)$*. Moreover, if* $\Theta_n(u_1) \in \mathcal{A}_{n,\text{long}} \mathcal{A}_{n,\text{short}}^*$ *and if there is no letter* $a \in B_n$ *such that* $\Theta_n(a) \in \mathcal{A}_{n,\text{short}}^* \Theta_n(u_1)$*, then* $\tau_n(d) \in (B_n^+ \setminus B_n^* b B_n^*) \cup (u_1 \cdots u_\ell b B_n^*)$ *for all letters* $d \in B_{n+1}$.

Proof. As Proposition 3.4.9, this is an consequence of Lemma 3.4.8. Indeed, the fact that $\tau_n(c) \in ubB_n^*$ implies that the subpath q of $\Theta_{n+1}(c)$ such that $\psi_{n,L}(p) = l_{u_1} p_1 \cdots p_k \Theta_n(u_2 u_3 \cdots u_\ell) p_{k+1} \cdots p_m l_b$ does not contain any left special vertex, where $l_{u_1}, l_b \in \mathcal{A}_{n,\text{long}}$, $p_i \in \mathcal{A}_{n,\text{short}}$ for all i and $l_{u_1} p_1 \cdots p_k$ and $p_{k+1} \cdots p_m l_b$ are respectively suffix and prefix of $\Theta_n(u_1)$ and $\Theta_n(b)$. Consequently, q is the only path of G_{n+1} from $o(q)$ to $i(q)$ (supposing that we do not consider paths containing twice the vertex $o(q)$). In other word, the suffix of q that is mapped to l_b through $\psi_{n,L}$ can be uniquely extended to the left in G_{n+1} by the subpath of q that is mapped to p_m through $\psi_{n,L}$ and this one can also be uniquely extended to the left in G_{n+1} and so on until we reach the prefix of q. Any letter d in B_{n+1} such that $\Theta_{n+1}(d)$ contains the path l_b as a subpath contains also the path q. Consequently, if b occurs in $\tau_n(d)$, then $u_2 u_3 \cdots u_\ell b$ also occurs in it. The first letter of $\tau_n(d)$ might be different from u_1 because there might be different letters a in B_n such that $\Theta_n(a)$ admits $l_{u_1} p_1 \cdots p_k$ as a suffix. □

Proposition 3.4.11 (Property 3). *Let* $\mathbf{w}$ *be an aperiodic and uniformly recurrent sequence over* A *with a sub-linear complexity and let* $(\kappa, \tau_N, \tau_{N+1}, \dots)$ *be the directive word of Definition 3.4.4. For all integers* $n \geq N$*, all letters* $b_1, \dots, b_k$ *in* B_n *and all letters* $c_1, \dots, c_k$ *in* B_{n+1}*,* $(\tau_n(c_1), \dots, \tau_n(c_k)) \notin B_n^* b_1 B_n^* b_2 B_n^* \times B_n^* b_2 B_n^* b_3 B_n^* \times \cdots \times B_n^* b_{k-1} B_n^* b_k B_n^* \times B_n^* b_k B_n^* b_1 B_n^*$.

Proof. This is again a consequence of Lemma 3.4.8. Indeed, suppose by contrary that

$$(\tau_n(c_1), \dots, \tau_n(c_k)) \in B_n^* b_1 B_n^* b_2 B_n^* \times B_n^* b_2 B_n^* b_3 B_n^* \times \cdots \\ \times B_n^* b_{k-1} B_n^* b_k B_n^* \times B_n^* b_k B_n^* b_1 B_n^*.$$

For all letters b_i, $i = 1, \dots, k$, we let l_i denote the long n-segment of $\Theta_n(b_i)$. We also let q_k denote the subpath of $\Theta_{n+1}(c_k)$ such that $\psi_{n,L}(q_k) = l_1$ and for $i = 1, \dots, k-1$, q_i is the subpath of $\Theta_{n+1}(c_i)$ such that $\psi_{n,L}(q_i) = l_{i+1}$.

Consider the path q_k. Lemma 3.4.8 and $\tau_n(c_k) \in B_n^* b_k B_n^* b_1 B_n^*$ imply that q_k can be uniquely extended to the left in G_{n+1} until we reach q_{k-1} (i.e., there is no left special vertex between $i(q_k)$ and $o(q_{k-1})$). Then, Lemma 3.4.8 and $\tau_n(c_{k-1}) \in B_n^* b_{k-1} B_n^* b_k B_n^*$ imply that q_{k-1} (and so q_k) can be uniquely extended to the left in G_{n+1} until we reach the q_{k-2}. Continuing this way, we see that q_k can be uniquely extended to the left in G_{n+1} until we reach q_k again. Thus this provides a loop in G_{n+1} that is inaccessible from vertices that do not belong to it and G_{n+1} is not strongly connected: a contradiction. □

Proposition 3.4.12 (Almost primitivity). *Let* $\mathbf{w}$ *be an aperiodic and uniformly recurrent sequence over* A *with a sub-linear complexity and let* $(\sigma_n)_{n\in\mathbb{N}}$ *be the directive word of Definition 3.2.1. For all non-negative integers* r *and all letters* $b \in A_r$*, there is an integer* $s > r$ *such that for all* $a \in A_{s,\text{long}}$*,* b *occurs in* $\sigma_r \cdots \sigma_{s-1}(a)$*. In particular, if* $(\sigma_n)_{n\in\mathbb{N}}$ *is everywhere growing, then it is almost primitive.*

Proof. The proof is exactly the same as for the S-adic representation using n-circuits (see Proposition 3.3.6). For any n-segments p, there is no n-segment q such that $\lambda(p)$ is a factor of $\lambda(q)$. But, thanks to the uniform recurrence, the full label of any n-segment is factor of any sufficiently long word in $L(\mathbf{w})$, hence of the full label of any long m-segment for m large enough.

The particular case is a direct consequence of the definitions. □

Remark 3.4.13. One can regret that the almost primitivity does not hold in general. But, for the directive word $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$, there can exist some letters in B_n that are useless. For instance, if there is a path p in a Rauzy graph G_n such that $p = l_1 s_1 s_2 l_2$ where $l_1, l_2 \in \mathcal{A}_{n,\text{long}}$ and $s_1, s_2 \in \mathcal{A}_{n,\text{short}}$, the definition of B_n states that there are 6 letters $b_1, b_2, \dots, b_6$ in B_n such that $\Theta_n(b_1) = l_1$, $\Theta_n(b_2) = l_1 s_1$, $\Theta_n(b_3) = l_1 s_1 s_2$, $\Theta_n(b_4) = l_2$, $\Theta_n(b_5) = s_2 l_2$ and $\Theta_n(b_6) = s_1 s_2 l_2$. Now consider $q \in \mathcal{P}_{n+1}$ such that $\psi_{n,L}(q) = p$ and suppose that the left special vertex $i(s_1) = o(s_2)$ is bispecial. Suppose moreover that its left extension which is an interior vertex of q is not right special and that its right extension which is an interior vertex of q is not left special.

Finally, we suppose that $i(l_2)$ is not right special. All this implies that if $q' \in \mathcal{P}_{n+1}$ is such that $\psi_{n,L}(q')$ contains $l_1c_1c_2l_2$ as a subpath, then q is a subpath of q'. Moreover, this also implies that q can be decomposed into $l'_1sl'_2$ with $l'_1, l'_2 \in \mathcal{A}_{n+1,\text{long}}$, $s \in \mathcal{A}_{n+1,\text{short}}$, $\psi_{n,L}(l'_1) = l_1$, $\psi_{n,L}(l'_2) = l_2$ and $\psi_{n,L}(s) = s_1s_2$. Consequently, there is no letter b in B_{n+1} such that $\tau_n(b)$ contains b_2 or b_5. So, both of them can be removed from B_n without loss of generality.

An idea to obtain the almost primitivity in the general case would be to prove that we can choose some sub-alphabets $\tilde{B}_n \subset B_n$ such that the restriction of $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ to these alphabet is still a directive word of $\mathbf{w}$ and is almost primitive. However, we still have some troubles with the fact that Θ_n is not an isomorphism between B_n^* and $\mathcal{B}_n^*$.

Remark 3.4.14. It is easily seen that Propositions 3.4.9, 3.4.10 and 3.4.11 still hold true if we replace $(\kappa, \tau_N, \tau_{N+1}, \tau_{N+2}, \dots)$ by the directive word $(\sigma_n)_{n\in\mathbb{N}}$. Indeed, their proofs are always based on Lemma 3.4.8 and this result still holds true when working with $(\sigma_n)_{n\in\mathbb{N}}$. In particular, the last part of Proposition 3.4.10 (with the letter u_1) is always true.

For the proof of next property, we need to recall some basic notions of graph theory. Let G be a graph. A path p in G is a *cycle* if its extremities are equal. Let v be a vertex of graph G. The *neighbours* of v are the vertices u such that there is an edge between u and v.

A *tree* is an undirected graph in which any two vertices are connected by exactly one simple path, i.e., a path that does not pass twice through a same vertex. In other words, any connected graph with no cycle (except the cycles $(u, v)(v, u)$ where u and v are vertices) is a tree. A tree is said to be *rooted* if one particular vertex v_0 is designated the *root*. In this case, the vertices v can be ordered with respect to the length of the unique simple path between v_0 and v. If the length of the simple path between v_0 and v is i, we say that v is a vertex of *level* i. The *children* of a vertex v of level i are the neighbours of level $i + 1$ of v. A vertex u is a *successor* of a vertex v if there is a sequence of vertices $v = v_1, v_2, \dots, v_k = u$ such that v_{i+1} is a child of v_i for all i, $1 \leq i \leq k - 1$. The set of successors of v in G is denoted by $succ_G(v)$. In the same idea, the *parent* of v is the neighbour of level $i - 1$ of v and the *ancestors* of v are the vertices u such that $v \in succ_G(u)$. A vertex v is a *leaf* if it has no child.

A *forest* is an undirected graph whose connected component are trees. When the trees of a forest F are rooted, the *roots* (resp. the *leaves*) of F are the respective roots (resp. the respective leaves) of its connected components.

Proposition 3.4.15 (Property 4)**.** *Let* $\mathbf{w}$ *be an aperiodic and uniformly recurrent sequence over* A *with a sub-linear complexity and let* $(\sigma_n)_{n\in\mathbb{N}}$ *be* *the directive word of Definition 3.2.1. For all n, σ_n belongs to T^* with $T = \{G\} \cup \{E_{ij} \mid i,j \in \mathbf{A}\} \cup \{M_i \mid i \in \mathbf{A}\}$ a set of morphisms such that:*

1. *$G(0) = 10$ and $G(i) = i$ for all letters $i \neq 0$;*
2. *E_{ij} exchange i and j and fixes the other letters;*
3. *M_i maps i to 0 and fix the other letters.*

Proof. Let n be an integer. The main idea to decompose the morphism σ_n is the following. Let F be the graph whose set of vertices are the couples (a, n) with a in B_n and the couples $(c, n+1)$ with c in B_{n+1} and whose set of edges is defined as follows:

- for $c \in B_{n+1}$ and a in B_n, there is an edge between $(c, n+1)$ and (a, n) if $\sigma_n(c) \in A_n^* a$;
- for $a, b \in B_n$, there is an edge between (a, n) and (b, n) whenever there is a letter c in B_{n+1} such that ba occurs in $\sigma_n(c)$.

We already know that the last part of Proposition 3.4.10 and that Proposition 3.4.11 hold true for the directive word $(\sigma_n)_{n\in\mathbb{N}}$ (see Remark 3.4.14). Moreover, they imply that F is a forest such that the number of connected components (that are trees) of F is the number of letters a in A_n such that $\sigma_n(c) \in aA_n^*$ for some letter c in A_{n+1}. We suppose that the root of such a tree is the vertex (a, n). Consequently, the leaves of F are the vertices $(c, n+1)$ and we can check that the set of images in $\sigma_n(A_{n+1})$ is the set of words $a_1 \cdots a_k$, $k \geq 0$, $a_1, \ldots, a_k \in B_n$ being the respective first components of the vertices of a simple path in F from a root to the parent of a leaf.

Now let us explain how we can build σ_n with F. The idea is to start from the leaves, to move towards the roots and to build σ_n reading the letters on the vertices, i.e., the first components of them. The first step (from the leaves to their respective parents) is simply to map each letter c in A_{n+1} to the last letter of $\sigma_n(c)$. This can be realized with the morphisms E_{ij} and M_i. Indeed, for any n-segment p, let $\chi(p) = \{Xx \mid X = i(p) \text{ and } x \in A \text{ such that } Xx \in L(\mathbf{w})\}$. As a segment is completely determined by its last edge, there is a bijection between the set $\mathcal{A}_{n+1}$ of $(n+1)$-segments and the set $\{Xx \in \chi(p) \mid Xx \text{ is left special and } p \in \mathcal{A}_n\}$. We write

$$\mathcal{A}_{n+1} \cong \{Xx \in \chi(p) \mid Xx \text{ is left special and } p \in \mathcal{A}_n\}. \tag{3.2}$$

Let p be a n-segment and let $k(p)$ be the number of vertices Xx in $\chi(p)$ that are left special. If $k(p) = 1$, we deduce from Equation 3.2 that there is a unique $(n+1)$-segment q_p such that

$$\sigma_n \circ \theta_{n+1}^{-1}(q_p) \in A_n^* \theta_n^{-1}(p).$$

Consequently, there is a bijection between P^* and $\{q_p \mid p \in P\}^*$ with

$$P = \{p \in \mathcal{A}_n \mid \exists! Xx \in \chi(p) \text{ that is left special}\}.$$

This bijection is realized by a bijective and letter-to-letter morphism $\mathcal{E}$ and it is clear that such a morphism can be decomposed in a finite product of morphisms E_{ij} (see for instance Lemma 2.2 in [Ric03]).

Now, if $k(p) > 1$, Once again we deduce from Equation 3.2 that there are $k(p)$ $(n+1)$-segments $q_{p,1}, \ldots, q_{p,k(p)}$ such that

$$\sigma_n \circ \theta_{n+1}^{-1}(q_{p,i}) \in A_n^* \theta_n^{-1}(p)$$

for all i, $1 \leq i \leq k(p)$. For all i, $1 \leq i \leq k(p)$, the letter $\theta_{n+1}^{-1}(q_{p,i})$ must be mapped to $\theta_n^{-1}(p)$. This is realized by the following product of morphisms:

$$\mathcal{M} = \prod_{\substack{p \in \mathcal{A}_n \text{ such} \\ \text{that } k(p)>1}} E_{0\theta_n^{-1}(p)} \left(\prod_{1 \leq i \leq k(p)} M_{\theta_{n+1}^{-1}(q_{p,i})} \right) E_{0\theta_n^{-1}(p)}.$$

Observe that, by construction, the morphisms $\mathcal{E}$ and $\mathcal{M}$ respectively act on disjoints subsets of B_{n+1}. Consequently, we have

$$\mathcal{E} \circ \mathcal{M}(A_{n+1}) = \mathcal{M} \circ \mathcal{E}(A_{n+1})$$

and this morphism realizes the step from the leaves of F to their respective parents.

Now let us show that we can keep moving towards the roots of F and build σ_n reading the letters on the vertices. Let us define the morphism $\sigma_{\text{temp}} = \mathcal{E} \circ \mathcal{M}$ and the graph $F_{\text{temp}} = F$. Since we have already built the morphism realizing the step from the leaves to their respective parents, we remove them (the leaves) from F_{temp}. Once this is done, there might be some new leaves in F_{temp} that are also roots of F_{temp}. For these vertices (a, n), this means that for any child $(c, n+1)$ of (a, n) in F we have $\sigma_{\text{temp}}(c) = \sigma_n(c) = a$ (otherwise there would be an edge between (a, n) and another vertex (b, n)). Hence the work is done for these letters so we remove the corresponding vertices from F_{temp}. Consequently, the remaining vertices in F_{temp} correspond to the letters a in A_n that occur in images $\sigma_n(c)$ of length at least 2. Observe that since we have only removed some leaves from F_{temp}, the graph is still a forest and we can repeat the process until F_{temp} is empty. This is formalized by the algorithm below.

Algorithm:
While F_{temp} is not empty:

1. Consider a leaf (a, n) in F_{temp}. Let (b, n) be the parent of (a, n) in F_{temp}. Remove (a, n) from F_{temp}.
2. Replace σ_{temp} by $\varrho \circ \sigma_{\text{temp}}$ where $\varrho = E_{0a} \circ E_{1b} \circ G \circ E_{1b} \circ E_{0a}$ maps a to ba and fixes the other letters.
3. If (b, n) is a root of F_{temp}, remove (b, n) from F_{temp}.

This algorithm clearly stops since any vertex of F can be reached (so removed from F_{temp}) in a finite number of steps. Moreover, when it stops, we have $\sigma_{\text{temp}} = \sigma_n$ (by construction of F). □

3.5 First conclusions

In this chapter, we presented a method to build S-adic representations of sequences or subshifts. The main idea is to consider the morphisms σ_n of Definition 3.2.1 but instead of considering them on their canonical alphabet A_n, we study their action on subsets of A_n^*. Depending on the chosen subsets, we of course get different properties. But, none of the choices we made led us to a good candidate for the condition C of the S-adic conjecture. Indeed, the conditions in Theorem 3.0.1 are clearly a bad candidate: take for instance the set of morphisms $\{\gamma^{k_n}\mu \mid n \in \mathbb{N}\}$ of Example 2.3.11 (page 67) and define for all n the morphism

$$\Delta_n : \begin{cases} 0 \mapsto 1\gamma^{k_n}\mu(0) \\ 1 \mapsto 1\gamma^{k_n}\mu(1) \end{cases}$$

All morphisms are proper and strongly primitive and the sequence

$$\mathbf{w} = \lim_{n \to +\infty} \Delta_0 \cdots \Delta_n(0^\omega)$$

does not have a sub-linear complexity as soon as the sequence $(k_n)_{n \in \mathbb{N}}$ is unbounded (see Proposition 2.3.12).

Example 2.3.11 can also be slightly modified in such a way that it satisfies all conditions of Theorem 3.0.3 so this last result does not solve the conjecture either. Indeed, we can decompose the 2 morphisms γ and μ with 3 morphisms that satisfy the conditions of the theorem: $\gamma = \beta \circ \alpha'$ and $\mu = \beta \circ \mu'$ with

$$\gamma' : \begin{cases} 0 \mapsto 012 \\ 1 \mapsto 3 \end{cases} \qquad \mu' : \begin{cases} 0 \mapsto 02 \\ 1 \mapsto 31 \end{cases} \qquad \beta : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 0 \\ 2 \mapsto 1 \\ 3 \mapsto 1 \end{cases} .$$

It is easily seen that the new directive word is almost primitive and satisfies Properties 1–3. Moreover, we have

$$\begin{aligned} \gamma' &= E_{02}GE_{12}GE_{13} \\ \mu' &= E_{12}GE_{02}E_{13}GE_{03}E_{13} \\ \beta &= M_1E_{01}M_2M_3E_{01} \end{aligned}$$

Now a natural question is to ask whether there exist some other subsets of A_n^* that could be considered as alphabets and would lead to S-adic representations with other properties. One also could try to extend these results to non-uniformly recurrent sequences. For instance, it seems that when $\mathbf{w}$ has a sub-linear complexity and is k-power-free, the set of morphisms occurring in the directive word $(\kappa, \tau_N, \tau_{N+1}, \dots)$ is also finite. But, it is not clear that all properties still hold true. Another natural idea is to try to describe exactly for which sequences this directive word is ultimately periodic. Some discussions are currently ongoing with S. Starosta in that last direction.

Remark 3.5.1. We can extend Theorem 3.0.3 to two-sided sequences $\mathbf{w} = {}^{-}\mathbf{w}.\mathbf{w}^{+}$. Indeed, in the proof of Proposition 3.4.7, we only have to find a sequence of letters $(s_n)_{n \geq N}$ such that for all n, $i(\Theta_n(s_n)) = \mathbf{w}_{[0,n-1]}$, $\kappa\tau_N \cdots \tau_{n-1}(s_n)$ is suffix of ${}^{-}\mathbf{w}$ and $\tau_n(s_{n+1}) \in B_n^* s_n$. Such a sequence exists by construction so we obtain

$$\mathbf{w} = \lim_{n \to +\infty} \kappa\tau_N\tau_{N+1} \cdots \tau_n({}^{\omega}s_{n+1}.p_{n+1}^{\omega}).$$

Chapter 4

$\mathcal{S}$-adicity of minimal subshifts with complexity $2n$

In this chapter we consider the S-adic representation of Theorem 3.0.1 in the particular case of minimal subshifts with first difference of complexity bounded by 2. In that particular case, we are able to give much more details on the representations. In particular, we prove that such subshifts are $\mathcal{S}$-adic with $\text{Card}(S) = 5$ which is a considerable improvement of Proposition 2.3.6. In all this chapter, the set $\mathcal{S}$ is the set of 5 morphisms $\{G, D, M, E_{01}, E_{12}\}$ where

$$G : \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases} \qquad D : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases} \qquad M : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 1 \end{cases}$$

$$E_{01} : \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases} \qquad E_{12} : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 2 \\ 2 \mapsto 1 \end{cases}$$

Theorem 4.0.1. *Let $\mathcal{G}$ be the graph represented in Figure 4.8. By adding one edge from 7 to 10 and one edge from 8 to 10, there is a non-trivial way to label the edges of $\mathcal{G}$ with morphisms in $\mathcal{S}^*$ such that for any minimal subshift (X, T) such that $1 \leq p_X(n+1) - p_X(n) \leq 2$ for all n, there is an infinite path p in $\mathcal{G}$ whose label $(\sigma_n)_{n \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ is a directive word of (X, T). Furthermore, $(\sigma_n)_{n \in \mathbb{N}}$ is almost primitive and admits a contraction that contains only proper morphisms*[1]*.*

[1]It was already the case in Theorem 3.0.1.

The proof of this theorem is based on a detailed description of all possible Rauzy graphs of minimal subshifts with the considered complexity. The Rauzy graphs of such subshifts can have only 10 different shapes. These shapes correspond to vertices of **G** and the edges of **G** are given by the possible evolutions of these graphs. We then compute explicitly the morphisms representing these evolutions and show that they belong to $\mathcal{S}^*$. In the next chapter, we will study even more the evolutions of Rauzy graphs in order to obtain an $\mathcal{S}$-adic characterization of these subshifts.

In all this chapter, (X, T) satisfies the conditions of Theorem 4.0.1, i.e., it is minimal and is such that $1 \leq p_X(n+1) - p_X(n) \leq 2$ for all n. Consequently, we have $p_X(n) \leq 2n$ for all $n \geq 1$ when $\mathrm{Card}(A) = 2$ and $p_X(n) \leq 2n + 1$ for all n when $\mathrm{Card}(A) = 3$.

4.1 Some preliminary lemmas

As already mentioned in Chapter 3, in this chapter we deal with right n-segments and right n-circuits (see Definition 3.1.2 and Definition 3.1.11). Consequently, some lemmas of Chapter 3 cannot be directly applied and need to be rephrased. In this section we quickly present the "right version" of the results of Chapter 3 we need for the considered particular case. Proofs are similar to those of Chapter 3.

First we define the function $\psi_{n,R}$ similarly to $\psi_{n,L}$.

Definition 4.1.1. Given a path $p \in \mathcal{P}_{n+1}$, $\psi_{n,R}(p)$ is the unique path q in $\mathcal{P}_n$ such that $\lambda_R(q) = \lambda_R(p)$ and $o(q)$ and $i(q)$ are suffixes of $o(p)$ and $i(p)$ respectively.

Lemma 4.1.2 (Ferenczi [Fer96])**.** *Let (X, T) be a subshift over an alphabet A and let v be a right special factor of length $n+1$ of X. For all $(n+1)$-circuits p starting from v, $\psi_{n,R}(p)$ is a concatenation of n-circuits starting from the suffix of length n of v. Moreover, the decomposition of $\psi_{n,R}(p)$ into n-circuits is unique.*

Lemma 4.1.3. *Let (X, T) be a minimal and aperiodic subshift over an alphabet A. There exists an infinite sequence $(v_n)_{n \in \mathbb{N}}$ of words over A such that for each $n \in \mathbb{N}$,*

- *v_n is of length n;*
- *v_n is a right special factor of X;*
- *v_n is a suffix of v_{n+1}.*

Definition 4.1.4. Let (X, T) be a minimal and aperiodic subshift over an alphabet A. Let also $(v_n)_{n\in\mathbb{N}}$ be a sequence of right special factors of X as in Lemma 4.1.3. For each non-negative integers n, let $\mathcal{C}'_n$ be the set of allowed n-circuits starting from v_n. Now define the alphabet $C'_n = \{0, 1, \ldots, \mathrm{Card}(\mathcal{C}_n) - 1\}$ and consider a bijection $\vartheta_n : C'_n \to \mathcal{C}'_n$. We can extend ϑ_n to an isomorphism by putting $\vartheta_n(ab) = \vartheta_n(a)\vartheta_n(b)$ for all letters a, b in C_n. Then, for all n, Lemma 4.1.2 allows us to define $\gamma'_n : C'^*_{n+1} \to C'^*_n$ as the unique morphism satisfying

$$\vartheta_n \circ \gamma'_n = \psi_{n,L} \circ \vartheta_{n+1}.$$

Remark 4.1.5. As for left n-circuits, when a Rauzy graph G_n does not contain any bispecial vertices, the morphism γ'_n is simply a bijective and letter-to-letter morphism. This morphism only depends on the differences that could exist between ϑ_n and ϑ_{n+1}. Moreover, since G_n does not contain any bispecial vertex, the shape of G_{n+1} is the same as the one of G_n. Consequently, we can suppose without loss of generality that ϑ_n and ϑ_{n+1} satisfy $\psi_{n,L} \circ \vartheta_{n+1}(i) = \vartheta_n(i)$ for all letters i in C_{n+1} so that γ'_n is the identity morphism. As a consequence, to build the $\mathcal{S}$-adic representation of a subshift, we only have to consider the subsequence $(\gamma'_{i_n})_{n\in\mathbb{N}}$ of $(\gamma'_n)_{n\in\mathbb{N}}$ where $(i_n)_{n\in\mathbb{N}}$ is the growing sequence of integers such that for all n, either G_n does not contain any bispecial vertex, or $n = i_k$ for some integer k. We therefore have $\gamma'_{i_n} = \gamma'_{i_n} \cdots \gamma'_{i_{n+1}-1}$.

Definition 4.1.6. Let $(\gamma'_n)_{n\in\mathbb{N}}$ be the sequence of morphisms as defined in Definition 4.1.4 and let $(i_n)_{n\in\mathbb{N}}$ be the sequence of integer as defined in Remark 4.1.5. For all n we let C_n denote the alphabet C_{i_n} and γ_n denote the morphism $\gamma'_{i_n}\gamma'_{i_n+1} \cdots \gamma'_{i_{n+1}-1}$. For all n, the morphism γ_n is therefore defined from C^*_{n+1} to C^*_n.

Remark 4.1.7. It is easily seen that, as for left n-circuits, we have $C_0 = A$ so we can suppose that ϑ_0 is such that for all $c \in C_0$,

$$\lambda_R \circ \vartheta_0(c) = c.$$

Lemma 4.1.8. *Let (X, T) be a minimal and aperiodic subshift over an alphabet A. Then the sequence of morphisms $(\gamma_n)_{n\in\mathbb{N}}$ of Definition 4.1.6 is a directive word of X.*

Lemma 4.1.9. *Let (X, T) be a minimal and aperiodic subshift over an alphabet A and let $(\gamma_n)_{n\in\mathbb{N}}$ be the directive word of Definition 4.1.6. There is a contraction $(\Gamma_n)_{n\in\mathbb{N}}$ of $(\gamma_n)_{n\in\mathbb{N}}$ such that all morphisms Γ_n are right proper and strongly primitive.*

Like for Theorem 3.0.1, we need to introduce the following trick. If $\sigma : A^* \to B^*$ is a right proper morphism such that $\sigma(A) \subset B^*b$ for a letter $b \in B$, we let $\sigma^{(L)} : A^* \to B^*$ denote the left proper morphism such that for all $a \in A$, $\sigma^{(L)}(a) = bu$ whenever $\sigma(a) = ub$. We call $\sigma^{(L)}$ the *left conjugate* of σ.

Lemma 4.1.10. *Let $\sigma : A^* \to B^*$ be a right proper morphism such that $\sigma(A) \subset B^*b$ for a letter $b \in B$. Let also $\mathbf{w}$ be a sequence in $A^{\mathbb{Z}}$. Then we have*

$$\sigma(\mathbf{w}) = T\left(\sigma^{(L)}(\mathbf{w})\right).$$

Proposition 4.1.11. *Let (X, T) be a minimal and aperiodic subshift over an alphabet A and let $(\Gamma_n)_{n\in\mathbb{N}}$ be the directive word of Lemma 4.1.9. The sequence of morphism $(\varrho_n)_{n\in\mathbb{N}}$ is a primitive and proper directive word of X where for all n,*

$$\varrho_n = \Gamma_{2n}\Gamma_{2n+1}^{(L)}.$$

4.2 10 shapes of Rauzy graphs

In this section we describe the possible shapes of Rauzy graphs for the considered class of complexity. To that aim we define the following notion of *reduced Rauzy graph.*

Definition 4.2.1. Let G_n be a Rauzy graph, the corresponding *reduced Rauzy graph* is the directed graph g_n such that

- the vertices are the vertices of G_n that are either special such that at least one value in $\{\delta^+ v, \delta^- v\}$ is null and
- there is an edge from u to v if there is a path p in G_n from u to v such that all interior vertices of p are not special.

The (left, right and full) labels of an edge in g_n are the (left, right and full) labels of the corresponding path in G_n. Of course, for all subshifts X, $g_n(X)$ denotes the reduced Rauzy graph corresponding to $G_n(X)$.

To avoid any confusion, edges of reduced Rauzy graphs are represented by double lines. Figure 4.1 represents the reduced Rauzy graph $g_2(\mathbf{f})$ with full labels on the edges where $\mathbf{f}$ is the Fibonacci sequence (Example 1.3.1). The graph $G_2(\mathbf{f})$ is represented at Figure 1.1(c).

From Equation (1.1) (on page 28) the hypothesis on the complexity implies that for all integers n, there are either one right special factor u of length n with $\delta^+(u) \in \{2, 3\}$ or two right special factors v_1 and v_2 with

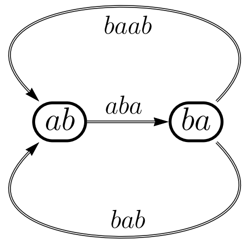


Figure 4.1: $g_2(\mathbf{f})$ with full labels on the edges.

$\delta^+(v_1) = \delta^+(v_2) = 2$. From Equation (1.2) we can make a similar observation for the left special factors. Hence for all integers n, we have the following possibilities:

1. there is one right special factor r and one left special factor l of length n with $\delta^+(r) = \delta^-(l) \in \{2, 3\}$ (Figure 4.2);

2. there is one right special factor r and two left special factors l_1 and l_2 of length n with $\delta^+(r) = 3$ and $\delta^-(l_1) = \delta^-(l_2) = 2$ (Figure 4.3(a));

3. there are two right special factors r_1 and r_2 and one left special factor l of length n with $\delta^+(r_1) = \delta^+(r_2) = 2$ and $\delta^-(l) = 3$ (Figure 4.3(b));

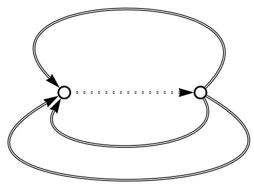
4. there are two right special factors r_1 and r_2 and two left special factors l_1 and l_2 of length n with $\delta^+(r_1) = \delta^+(r_2) = \delta^-(l_1) = \delta^-(l_2) = 2$ (Figure 4.4).

From these possibilities we can deduce that for all n, $g_n(X)$ only has eight possible shapes: those represented in Figures 4.2 to 4.4. Reduced Rauzy graphs in Figure 4.2 are well-known: they correspond to reduced Rauzy graphs of Sturmian sequences (Figure 4.2(a)) or of Arnoux-Rauzy sequences (Figure 4.2(b)). Reduced Rauzy graphs in Figure 4.4 have also been studied by Rote in [Rot94].

Observe that in the above figures, the edges represented by dots may have length 0. In this case, the two vertices they link are merged to one vertex.

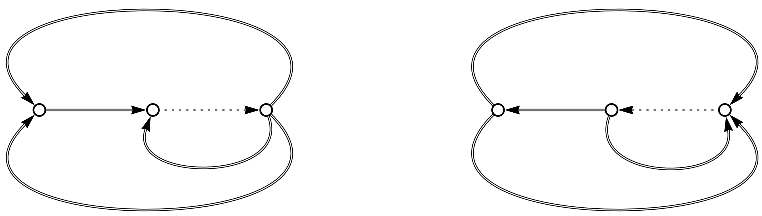
From Remark 4.1.5 we only have to consider Rauzy graphs containing at least one bispecial factor. To this aim, we have to merge the vertices that are linked by dots in Figures 4.2 to 4.4. Observe that both Figures 4.4(a) and 4.4(b) give rise to two different graphs: one with one bispecial vertex and one right special vertex and one with two bispecial vertices. This gives rise to 10 different type of graphs. They are represented in Figure 4.5.

Remark 4.2.2. In the sequel, we sometimes talk about the type of a Rauzy graph G_n without any bispecial vertex. In that case, the type of that graph



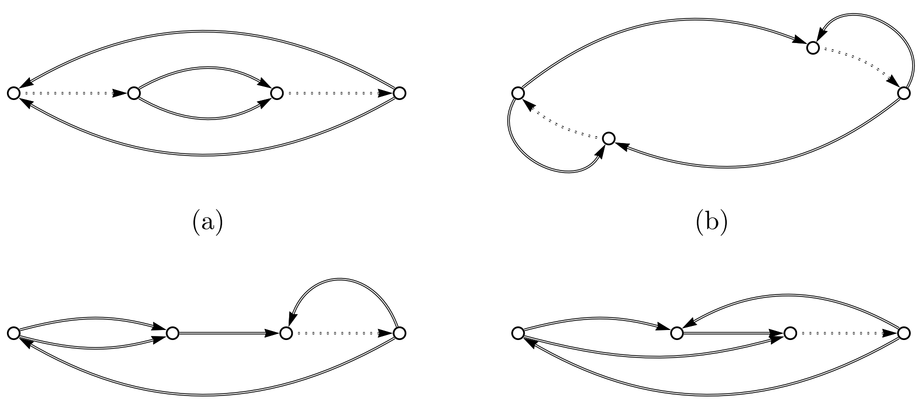
(a) (b)

Figure 4.2: Reduced Rauzy graphs with one left special factor and one right special factor.



(a) (b)

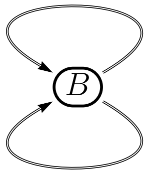
Figure 4.3: Reduced Rauzy graphs with different numbers of left and right special factors.



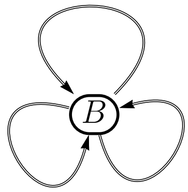
(a) (b)

(c) (d)

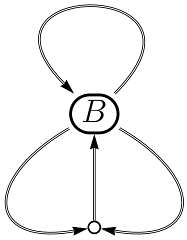
Figure 4.4: Reduced Rauzy graphs with two left and two right special factors.



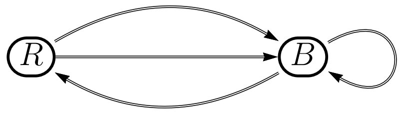
(a) Type 1



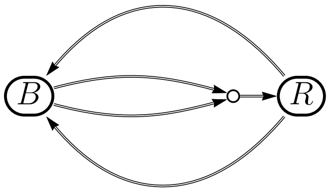
(b) Type 2



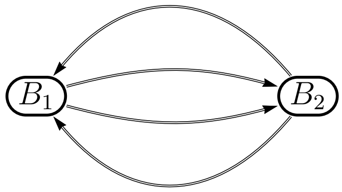
(c) Type 3



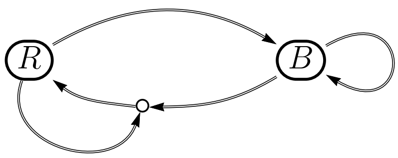
(d) Type 4



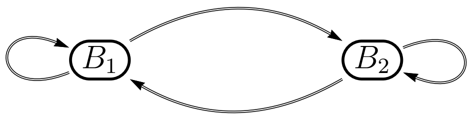
(e) Type 5



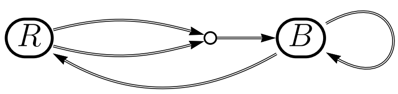
(f) Type 6



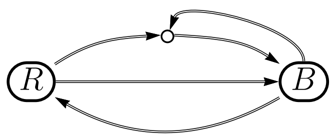
(g) Type 7



(h) Type 8



(i) Type 9



(j) Type 10

Figure 4.5: Reduced Rauzy graphs with at least one bispecial vertex.

is simply the type of G_n where i_n is the smallest integer greater than n such that G_n contains a bispecial vertex. Also it is obvious that if R is a right special vertex in a Rauzy graph, the circuits starting from it have the same full labels of those starting from the smallest bispecial vertex (in a Rauzy graph of larger order) containing R as a suffix.

Now that we have defined all types of graphs, we can check which evolutions are available, i.e., which type of graphs can evolve to which type of graphs. It is clear that a given Rauzy graph cannot evolve to any type of Rauzy graphs. For example, if G_n is a graph of type 4, both right special vertices can be extended by only two letters. Since for any word u and for any suffix v of u, we have $\delta^+(v) \geq \delta^+(u)$, the graph G_n will never evolve to a graph of type 2 or 3. Let us explain with an example how we can compute the possible evolutions.

An example

Consider a graph of type 1 as represented in Figure 4.6 and let us give all possible evolutions from it. The letters a and b (resp. α and β) represent the right (resp. left) extending letters of B.

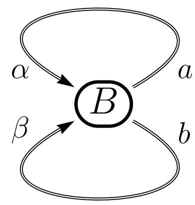
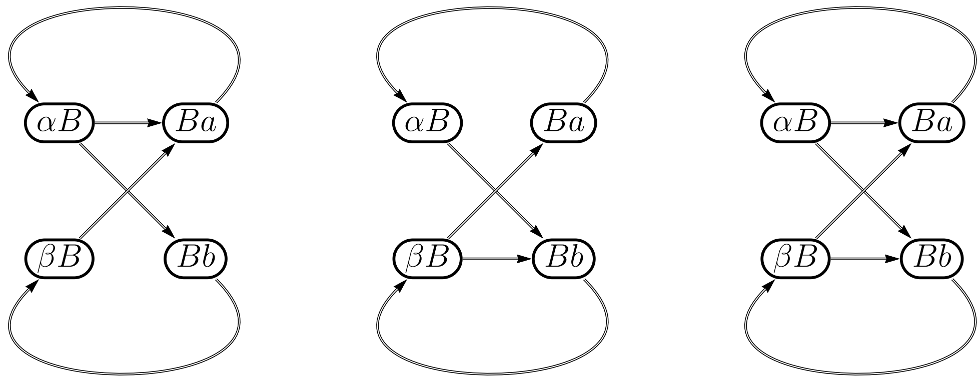


Figure 4.6: Reduced Rauzy graph of type 1 with some additional labels.

By definition of the Rauzy graph, the words αB, βB, Ba and Bb are vertices of G_{n+1}. Since the subshifts we are considering satisfy $p(n+1) - p(n) \geq 1$ for all n, at least one of the vertices αB and βB is right special and at least one of the vertices Ba and Bb is left special. Moreover, by definition of the reduced Rauzy graphs, the two loops of g_n become edges respectively from Ba to αB and from Bb to βB and the last thing we have to do is to decide which edges are starting from αB and βB and which edges are arriving to Ba and Bb. Except if a loop has length 1, there is obviously no other edge but the loops starting from Ba and Bb or arriving to αB and βB. For instance, we cannot have an edge from βB top αB (when the loops are longer than one). By minimality, we know that we have only three possibilities (2 of them being symmetric). The possible evolutions are represented at Figure 4.7. This shows that a graph of type 1 can evolve only to a graph of type 1, 7 or 8.



(a) B is ordinary bispecial (b) B is ordinary bispecial (c) B is strong bispecial

Figure 4.7: Possible evolutions of the graph represented in Figure 4.6.

Graph of graphs

Making an analogous reasoning starting from any type of Rauzy graph, we can compute which evolutions are available. Then, we can define the *graph of graphs* as the directed graph with 10 vertices (one for each type of Rauzy graph) such that there is an edge from i to j if a Rauzy graph of type i can evolve to a Rauzy graph of type j. This graph is represented in Figure 4.8 and all possible evolutions are given in Appendix A.

4.3 A critical result

Now that we know all possible Rauzy graphs we have to deal with, we can define the bijections ϑ_n of Definition 4.1.4. As already mentioned in Chapter 3, the alphabets C_n might be unbounded in the general case. In this section we prove that when the first difference of complexity is bounded by 2, they always contain 2 or 3 letters. This result seems to be inherent to that class of complexity. Actually, Example 4.3.5 at the end of the section shows that it cannot be extended to the general case of sub-linear complexity. It would be interesting to have a similar example with the first difference of complexity bounded by 3.

We need two technical lemmas to simplify the proof that $\mathrm{Card}(C_n) \in \{2, 3\}$ for all n.

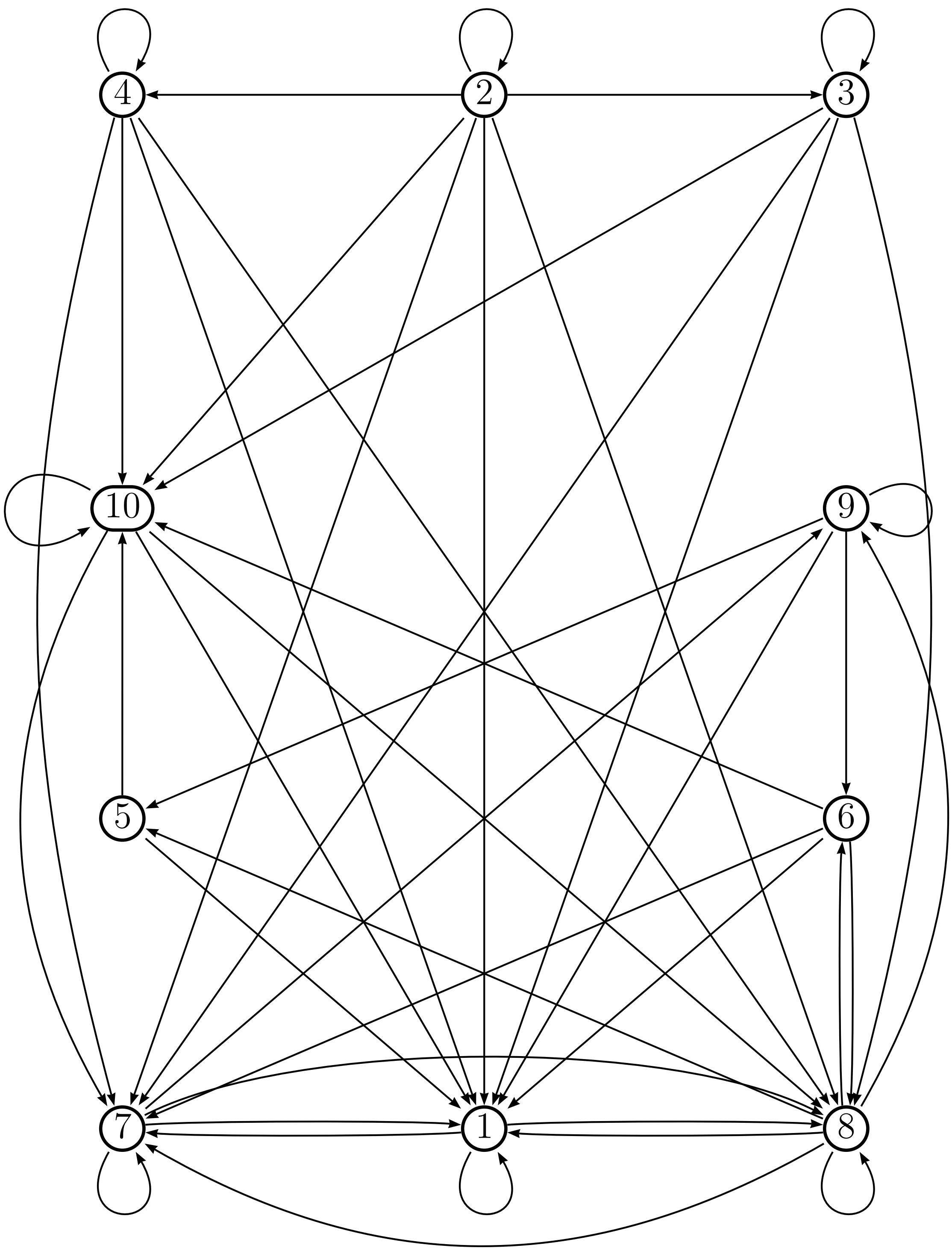


Figure 4.8: Graph of graphs.

Lemma 4.3.1. *Let A be an alphabet. If (X,T) is a minimal subshift over A satisfying $p(n+1)-p(n) \leq 2$ for all n and if B is a strong bispecial factor of X, then any right special factor of length $\ell > |B|$ admits B as a suffix.*

Proof. Indeed, B being supposed to be strong bispecial, we have $m(B) > 0$. Then, Equation (1.3) on page 29 shows that this is equivalent to

$$\sum_{aB \in L(X)} (\delta^+(aB) - 1) > \delta^+(B) - 1$$

where the second inequality is true only if there are at least two letters a and b in A such that aB and bB are right special (since $\delta^+(aB) \leq \delta^+(B)$). As there can exist at most 2 right special factors of each length (because $p(n+1)-p(n) \leq 2$) and as any suffix of a right special factor is still a right special factor, the result holds. □

The following result is a direct consequence of Lemma 4.3.1.

Corollary 4.3.2. *Let (X,T) be a minimal subshift satisfying $1 \leq p(n+1) - p(n) \leq 2$ for all n and let $(v_n)_{n\in\mathbb{N}}$ be a sequence of right special factors of X fulfilling conditions of Lemma 4.1.3. For any strong bispecial factor B of length n of X, we have $B = v_n$. In particular, if there are infinitely many strong bispecial factors in $L(X)$, there is a unique sequence $(v_n)_{n\in\mathbb{N}}$ fulfilling conditions of Lemma 4.1.3.*

Lemma 4.3.3. *Let G_n be a Rauzy graph. If there is a right special vertex R in G_n with $\delta^+(R) = 2$, an n-circuit q starting from R, two paths p and s in G_n and two integers k_1 and k_2, $k_1 < k_2 - 1$, such that*

1. *$i(p) = o(s) = R$;*
2. *p is not a suffix of q;*
3. *q is not a suffix of p;*
4. *the first edge of s is not the first edge of q;*
5. *both paths $pq^{k_1}s$ and $pq^{k_2}s$ are allowed;*

then there is a strong bispecial factor B that admits R as a suffix.

Proof. Since $i(p) = o(q) = R$ but p and q are not suffix of each other, there is a left special vertex L in G_n and two edges e_1 in p and e_2 in q such that p and q agree on a path q' from L to R and $i(e_1) = i(e_2) = L$. Let α and β be the respective left labels of e_1 and e_2. Let also a and b respectively denote the right labels of the first edge of q and of s. By hypothesis we have $a \neq b$.

Now let us prove that the word $\lambda(q'q^{k_1})$ is strong bispecial. As the paths $pq^{k_1}s$ and $pq^{k_2}s$ are allowed, the four words $\alpha\lambda(q'q^{k_1})a$, $\alpha\lambda(q'q^{k_1})b$, $\beta\lambda(q'q^{k_1})a$ and $\beta\lambda(q'q^{k_1})b$ belong to $L(X)$. Consequently we have

$$\delta^+(\alpha\lambda(q'q^{k_1})) + \delta^+(\beta\lambda(q'q^{k_1})) = 4.$$

Moreover, as the word $\lambda(q'q^{k_1})$ admits R as a suffix, we have $\delta^+(\lambda(q'q^{k_1})) \leq \delta^+(R) = 2$ and this implies that $m(\lambda(q'q^{k_1})) > 0$ (from Equation (1.3) on page 29). □

Proposition 4.3.4. *Let (X, T) be a minimal subshift satisfying $1 \leq p(n+1) - p(n) \leq 2$ for all n and let $(v_n)_{n \geq N}$ be a sequence of right special factors fulfilling the conditions of Lemma 4.1.3. Then for all right special factors v_n, there are at most 3 allowed n-circuits starting from v_n.*

Proof. Suppose that there exist 4 allowed n-circuits starting from the vertex v_n in the graph $G_n(X)$ and let us have a look at all possible reduced Rauzy graphs. We see that this is possible only if there exist two right special factors of length n. More precisely, this is only possible if v_n corresponds to the leftmost right special vertex in Figures 4.3(b), 4.4(c) and 4.4(d) or to any right special vertex in Figures 4.4(a) and 4.4(b) (as these two graphs present a kind of "symmetry"). We will show that for each of these graphs, the existence of 4 n-circuits starting from the described vertices implies that the other right special factor R of length n is a suffix of a strong bispecial factor B of length $m \geq n$ in $L(X)$. Then, due to Corollary 4.3.2, $v_m = B$ so v_n is not a suffix of v_m which contradicts the hypothesis.

The result clearly holds for graphs as represented in Figure 4.4(a) and it is a direct consequence of Lemma 4.3.3 for graphs as represented at Figure 4.4(b) (since the existence of 4 n-circuits implies that 3 of them goes through the loop respectively k_1, k_2 and k_3 times, $k_1 < k_2 < k_3$).

For graphs as represented in Figure 4.4(c), we have to consider several cases. To be clearer, Figure 4.9 represents the same graph with some labels. The letters α and β are the left extending letters of L_1 in $L(X)$ and the letters a and b are the right extending letters of R_2 in $L(X)$. If there are three n-circuits starting from R_1, going through a same simple path from R_1 to L_1 and passing through the loop $p = L_2 \to R_2 \to L_2$ respectively k_1, k_2, and k_3 times, $k_1 < k_2 < k_3$, then we can conclude using Lemma 4.3.3. Otherwise, for both simple paths from R_1 to L_1, there are two n-circuits passing through it. Let $k_{\alpha,1}$ and $k_{\alpha,2}$, $k_{\alpha,1} < k_{\alpha,2}$ (resp. $k_{\beta,1}$ and $k_{\beta,2}$, $k_{\beta,1} < k_{\beta,2}$) be the number of times that the two circuits passing through the edge with left label α (resp.

β) can pass through the loop p. If $k_{\alpha,1} < k_{\alpha,2} - 1$ or if $k_{\beta,1} < k_{\beta,2} - 1$ or if $k_{\alpha,1} \neq k_{\beta,1}$, we conclude using Lemma 4.3.3. Otherwise, we have $k_{\alpha,1} = k_{\beta,1}$ and $k_{\alpha,2} = k_{\beta,2} = k_{\alpha,1} + 1$ and we can easily check that the full label of the path $q = L_1 (\to L_2 \to R_2)^{k_{\alpha,1}}$ is a strong bispecial factor.

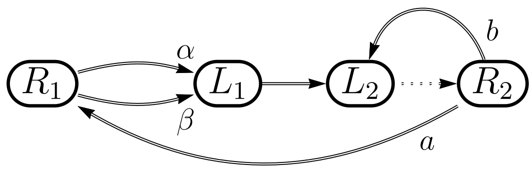


Figure 4.9: Graph as in Figure 4.4(c) with some labels.

The cases of graphs as represented at Figures 4.3(b) and 4.4(d) can be treated in a similar way. □

The next example shows that Proposition 4.3.4 cannot be extended to the general case. Indeed, it provides a uniformly recurrent sequence (hence a minimal subshift) with sub-linear complexity and such that the number of return words (hence of n-circuits) to any factor of length n increases with n.

Example 4.3.5. For all n let us define the morphism π_n over $\{0, 1\}$ by

$$\pi_n : \begin{cases} 0 \mapsto 010^2 10^4 1 \cdots 0^{2^n} = 0 \left(\prod_{i=0}^{n} 1\ 0^{2^i} \right) \\ 1 \mapsto 101^2 01^4 0 \cdots 1^{2^n} = 1 \left(\prod_{i=0}^{n} 0\ 1^{2^i} \right) \end{cases}$$

and consider the sequence

$$\mathbf{w}_{pi} = \lim_{n \to +\infty} \pi_1 \pi_2 \cdots \pi_n (0^\omega).$$

Proposition 4.3.6. *The sequence $\mathbf{w}_\pi$ defined above is uniformly recurrent, has a sub-linear complexity and for all integers k, there is a length ℓ_k such that all factors of $\mathbf{w}_\pi$ of length at least ℓ_k have at least k return words in $\mathbf{w}_\pi$.*

Proof. The uniform recurrence is a direct consequence of Proposition 2.1.21. With an analogous reasoning as in Proposition 2.3.14, we can also show that $\mathbf{w}_\pi$ has a sub-linear complexity.

Let us prove that all sufficiently long factors of $\mathbf{w}_\pi$ have many return words. For all $k \geq 1$, we let $\mathbf{w}_k$ denote the sequence

$$\mathbf{w}_k = \lim_{n \to \infty} \pi_k \pi_{k+1} \cdots \pi_n (0^\omega).$$

We obviously have $\mathbf{w}_1 = \mathbf{w}_\pi$ and for all $k \geq 1$, $\mathbf{w}_\pi = \pi_1 \cdots \pi_k(\mathbf{w}_{k+1})$. We also have

$$|\pi_1\pi_2 \cdots \pi_k(0)| = |\pi_1\pi_2 \cdots \pi_k(1)|$$

for all k so we let l_k denote this length.

Let n be a positive integer (suppose it is large). The sequence $(l_k)_{k\geq 1}$ is increasing so there is a unique positive integer k such that $l_{k-1} \leq n < l_k$. Consequently, all factors of length n of $\mathbf{w}_\pi$ belong to $\text{Fact}\left(\pi_1 \cdots \pi_k\left(\{0,1\}^2\right)\right)$. From the shape of π_k, for all $u \in L_n(\mathbf{w}_\pi)$ there is a unique word $v \in \text{Fact}\left(\pi_k\left(\{0,1\}^2\right)\right)$ such that $u \in \text{Fact}\left(\pi_1 \cdots \pi_{k-1}(v)\right)$ and any word v' such that $u \in \text{Fact}\left(\pi_1 \cdots \pi_{k-1}(v')\right)$ contains v as a factor. By unicity of v, the number of return words to u in $\mathbf{w}_\pi$ is equal to the number of return words to v in $\mathbf{w}_k$ so we only have to show that v has many return words. Let us show that the number of return words to v in $\mathbf{w}_k$ is at least linear in k. This will prove the result since k increases with n.

If both words 00 and 11 belong to $\text{Fact}(v)$, there are two possibilities:

- either $v = xy$ where x is a suffix of length at least 2 of $\pi_k(0)$ and y is a prefix of length at least 4 of $\pi_k(1)$;
- or $v = xy$ where x is a suffix of length at least 2 of $\pi_k(1)$ and y is a prefix of length at least 4 of $\pi_k(0)$.

In both cases, the number of return words to v in $\mathbf{w}_k$ is equal to the number of return words to 01 or 10 (depending on the case) in $\mathbf{w}_{k+1}$. To conclude this case, we have to check in the images of π_{k+1} that both 01 and 10 have at least $k+2$ return words in $\mathbf{w}_{k+1}$.

Now suppose that 00 belongs to $\text{Fact}(v)$ but that 11 does not (the opposite case can be proved similarly). This implies that v does not occur in $\pi_k(1)$ so v has at least as much return words in $\mathbf{w}_k$ as the number of distinct positive powers of 1 in $\mathbf{w}_{k+1}$, i.e., $k+1$.

Finally, suppose that 00 and 11 do not belong to $\text{Fact}(v)$. Then v can only belong to $\{0, 1, 01, 10, 010, 101\}$ and we can see in the images of π_k that the number of return words in $\mathbf{w}_k$ to each factor is at least k. This completes the proof. □

4.4 A procedure to assign letters to circuits

Now let us explicitly determine the bijections ϑ_k. From Remark 4.2.2 we only have to define ϑ_{i_n} for all n. We would like to define them for each graph represented at Figure 4.5 in such a way that two Rauzy graphs of same type provide the same bijection ϑ_k. In that case, a given evolution (from G_{i_n}

to $G_{i_{n+1}}$) would always provide the same morphism γ_n of Definition 4.1.6. However, we will see that it is sometimes impossible to give enough details about ϑ_k so that the morphisms γ_n are sometimes defined up to permutations of the letters (see Section 4.5).

From Lemma 4.3.4 we know that $\mathrm{Card}(C_n) \in \{2,3\}$ for all n (1 is not enough since the number of i_n-circuits is at least $\delta^+(v_{i_n}) \geq 2$). From Definition 4.1.4 we then have $C_n \in \{\{0,1\},\{0,1,2\}\}$ depending on n.

Observe that, in the description of the bijections ϑ_{i_n} below, we sometimes express some restrictions on the number of times that one can pass through a loop in the consider type of Rauzy graph. The reason for this is that if the circuits do not satisfy those restrictions, the right special factor that is not v_{i_n} is a suffix of a strong bispecial factor (by Lemma 4.3.3) so this contradicts Corollary 4.3.2.

1. **Type 1**: there exists only one right special vertex and the two possible circuits are the two loops. One is $\vartheta_{i_n}(0)$ and the other is $\vartheta_{i_n}(1)$ and we cannot be more precise (like we are for graphs of type 2 or 3 here below).

2. **Type 2 and 3**: also here there exists only one right special vertex and the three possible circuits are the three possible loops $\vartheta_{i_n}(0)$, $\vartheta_{i_n}(1)$ and $\vartheta_{i_n}(2)$. However, as shown by Figure 4.8, the only graphs that can evolve to a graph of type 2 (resp. of type 3) are the graphs of type 2 (resp. of type 2 and 3). Moreover after such an evolution, the right labels of the three loops start with the same letter as before the evolution. Consequently we suppose that for all $i \in \{0,1,2\}$, i is prefix of $\lambda_R \circ \vartheta_{i_n}(i)$.

3. **Type 4**: first consider $v_{i_n} = R$. There exist two segments from R to B. Consequently, there exist at least two circuits $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$, each of them passing through one of the two segments and looping respectively k and ℓ times, $k+\ell \geq 1$, in the loop $B \to B$ before coming back to R. If there exists a third circuit, then we suppose it starts with the same segment as the circuit $\vartheta_{i_n}(0)$ does, and then goes through the loop exactly $k-1$ times. In this case, we must have $\ell \leq k$. If the third circuit does not exist, then we suppose that $k \geq \ell$ so we have $k \geq \ell \geq 0$ and $k+\ell \geq 1$.

   Now consider $v_{i_n} = B$. There exist exactly three circuits: the circuit that does not pass through the vertex R is denoted by $\vartheta_{i_n}(0)$ and the two others, $\vartheta_{i_n}(1)$ and $\vartheta_{i_n}(2)$, are going to the vertex R and then are coming back to B with one of the two segments from R to B.

4. **Type 5 and 6**: as a consequence of Remark 4.2.2, the circuits are the same whatever the type of graphs is. Moreover, from the symmetry of theses graphs, it is useless to make a distinction between the two right special vertices. Suppose $v_{i_n} = R$ for a graph of type 5. There exist four possible circuits (but Proposition 4.3.4 implies that only three among them are allowed) and we only impose some restrictions to their labels: the circuits $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$ must pass through two different segments from R to B and through two different segments from B to R. If the third circuit $\vartheta_{i_n}(2)$ exists, then it pass through the same segment from R to B as $\vartheta_{i_n}(0)$ does and through the same segment from B to R as $\vartheta_{i_n}(1)$ does.

5. **Type 7 and 8**: like for graphs of type 5 or 6, the starting vertex and the type of the graph does not change anything to the definition of the circuits. Suppose $v_{i_n} = R$ for a graph of type 7. We consider that $\vartheta_{i_n}(0)$ is the circuit that does not pass through the vertex B. The circuit $\vartheta_{i_n}(1)$ goes to B, passes through the loop $B \to B$ k times, $k \geq 1$, and then comes back to R. The circuit $\vartheta_{i_n}(2)$, if it exists, is the same as $\vartheta_{i_n}(1)$ but passes through the loop $B \to B$ $k-1$ times instead of k times.

6. **Type 9**: suppose $v_{i_n} = R$. Like for graphs of type 4, we consider the two circuits $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$, each of them going through different segments from R to B and looping respectively k and ℓ times in the loop $B \to B$, $k + \ell \geq 1$, before coming back to R. However for these graphs, k and ℓ must satisfy $k - \ell \leq 1$ otherwise the vertex B would become strong bispecial (see Lemma 4.3.3). Moreover, if the third circuit $\vartheta_{i_n}(2)$ exists, we suppose it starts like $\vartheta_{i_n}(0)$ does and passes through the loop exactly $k-1$ times. In this case, the circuit $\vartheta_{i_n}(1)$ cannot go through the loop $k+1$ times otherwise B would again become strong bispecial. Hence we always suppose $k \geq \ell$. Consequently, ℓ can only take the values $k-1$ and k even if the circuit $\vartheta_{i_n}(2)$ does not exist.

   Now suppose $v_{i_n} = B$. There exist exactly three circuits: the circuit that does not pass through the vertex R is $\vartheta_{i_n}(0)$ and the two other circuits, $\vartheta_{i_n}(1)$ and $\vartheta_{i_n}(2)$, are going to the vertex R and then are coming back to B with one of the two segments from R to B.

7. **Type 10**: suppose $v_{i_n} = R$. Let x denote the segment from R to B that passes only through non-left-special vertices; y is the other segment from R to B. We consider that $\vartheta_{i_n}(0)$ (resp. by $\vartheta_{i_n}(1)$) is the circuit that starts with y (resp. with x), passes k times (resp. ℓ times) through

the loop $B \to B$, $k + \ell \geq 1$, and then comes back to R. If the third circuit $\vartheta_{i_n}(2)$ exists, then it starts with x or y and loops respectively $k-1$ or $\ell-1$ times before coming back to R. Moreover, if $\vartheta_{i_n}(2)$ starts with x, then we must have $k \leq \ell - 1$ and if $\vartheta_{i_n}(2)$ starts with y, then we must have $\ell \leq k$ (because of Lemma 4.3.3).

Now suppose $v_{i_n} = B$. There are exactly three circuits. The loop $B \to B$ is $\vartheta_{i_n}(0)$, the circuit passing through the segment y is $\vartheta_{i_n}(1)$ and the circuit passing through x is $\vartheta_{i_n}(2)$.

4.5 Computation of the morphisms γ_n

Now that we know the bijections ϑ_{i_n}, we can compute the morphisms γ_n as in Definition 4.1.4 (and Definition 4.1.6). In this section we only present the method on the same example as in Section 4.2. The entire list (from page 122 to 125) can be computed in the same way using graphs represented in Appendix A so it is left to the reader. However, pay attention that when the graph obtained after evolution contains two right special vertices, there are often at least two morphisms coding the evolution: one for each choice of v_{i_n+1}.

Suppose G_n is a graph of type 1 as in Figure 4.6 (on page 110). By definition of ϑ_{i_n} for this type of graphs, $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$ are the two loops of the graph. Suppose that ϑ_{i_n} maps 0 to the i_n-circuit starting with an a and 1 to the i_n-circuit starting with a b. For the two first evolutions (Figure 4.7(a) and 4.7(b)), $g_{i_{n+1}}$ is again of type 1. By definition of $\vartheta_{i_{n+1}}$ for this type of graphs, we therefore have two possibilities for each evolution. Indeed, in Figure 4.7(a) we have either

$$(\psi_{i_n,R} \circ \vartheta_{i_n+1}(0), \psi_{i_n,R} \circ \vartheta_{i_n+1}(1)) = (\vartheta_{i_n}(0), \vartheta_{i_n}(10))$$

or

$$(\psi_{i_n,R} \circ \vartheta_{i_n+1}(0), \psi_{i_n,R} \circ \vartheta_{i_n+1}(1)) = (\vartheta_{i_n}(10), \vartheta_{i_n}(0))$$

and in Figure 4.7(b) we have either

$$(\psi_{n,R} \circ \vartheta_{n+1}(0), \psi_{n,R} \circ \vartheta_{n+1}(1)) = (\vartheta_{i_n}(01), \vartheta_{i_n}(1))$$

or

$$(\psi_{n,R} \circ \vartheta_{n+1}(0), \psi_{n,R} \circ \vartheta_{n+1}(1)) = (\vartheta_{i_n}(1), \vartheta_{i_n}(01)).$$

The four morphisms labelling the edge from 1 to 1 in the graph of graphs are therefore

$$\begin{cases} 0 \mapsto 0 \\ 1 \mapsto 10 \end{cases} \qquad \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 0 \end{cases}$$
$$\begin{cases} 0 \mapsto 01 \\ 1 \mapsto 1 \end{cases} \qquad \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 01 \end{cases} \tag{4.1}$$

For the third evolution (Figure 4.7(c)), the bijection ϑ_{i_n+1} (hence $\vartheta_{i_{n+1}}$) depends on the choice of v_{i_n+1}. If $v_{i_n+1} = \alpha B$ we have

$$(\psi_{i_n,R}\vartheta_{i_n+1}(0), \psi_{i_n,R}\vartheta_{i_n+1}(1), \psi_{i_n,R}\vartheta_{i_n+1}(2)) = (\vartheta_{i_n}(0), \vartheta_{i_n}(1^k0), \vartheta_{i_n}(1^{k-1}0))$$

for an integer $k \geq 2$ (remember that the circuit $\vartheta_{i_n+1}(2)$ might not exist). Similarly, if $v_{i_n+1} = \beta B$ we have

$$(\psi_{i_n,R}\vartheta_{i_n+1}(0), \psi_{i_n,R}\vartheta_{i_n+1}(1), \psi_{i_n,R}\vartheta_{i_n+1}(2)) = (\vartheta_{i_n}(1), \vartheta_{i_n}(0^k1), \vartheta_{i_n}(0^{k-1}1))$$

for an integer $k \geq 2$. Consequently, there are infinitely many morphisms labelling the edges from 1 to 7 and from 1 to 8 (one for each $k \geq 2$) but they all have one of the following two shapes:

$$\begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1^k0 \\ 2 \mapsto 1^{k-1}0 \end{cases} \quad \text{and} \quad \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0^k1 \\ 2 \mapsto 0^{k-1}1 \end{cases} \quad . \tag{4.2}$$

Still remember that we possibly have to consider their restriction to the alphabet $\{0, 1\}$.

In this example we see that an edge in the graph of graphs might be labelled by several morphisms. This is due not only to a lack of precision in the definition of the bijections ϑ_{i_n} but also to the number of possibilities that exist for a given Rauzy graph to evolve to a given type of Rauzy graph. For example, consider a graph of type 8 as in Figure 4.10.

This graph can evolve to a graph of type 7 or 8 (depending on the length of some paths) in two different ways:

- either one of the bispecial factors B_1 and B_2 is a strong bispecial factor and the other one is a weak bispecial factor;

- or both of them are ordinary bispecial factors and the two new right special factors are αB_1 and δB_2.

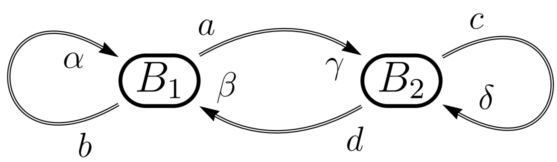


Figure 4.10: Rauzy graph of type 8 with some labels.

Indeed, the two other cases do not satisfy the hypothesis on the subshift: two weak bispecial factors delete all right special factors so the subshift is either not minimal (when the graph is not strongly connected anymore) or periodic (when the graph keeps being strongly connected) and two strong bispecial factors provide 4 right special factors so we do not have $p(n+1) - p(n) \leq 2$ anymore.

The Rauzy graphs obtained in both available cases are represented at Figure 4.11. They are of type 7 or 8 depending on the respective length of the paths $B_1 b \to \alpha B_1$ and $B_1 a \to \beta B_1$ for Figure 4.11(a) and on the respective length of the paths $B_1 b \to \alpha B_1$ and $B_2 c \to \delta B_2$ for Figure 4.11(b). These two possibilities of evolution to a same type of graphs imply that the edges $8 \to 7$ and $8 \to 8$ in $\mathcal{G}$ are labelled by several morphisms.

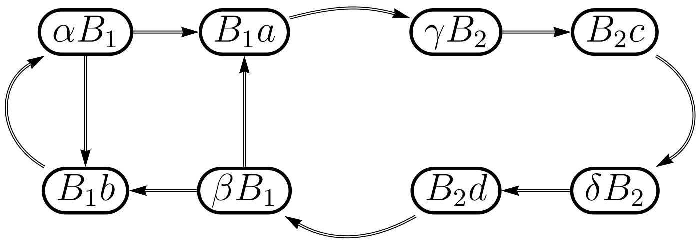
Now let us present all morphisms. To alleviate notations we let $[u, v, w]$ denote the morphism

$$\begin{cases} 0 \mapsto u \\ 1 \mapsto v \\ 2 \mapsto w \end{cases}$$

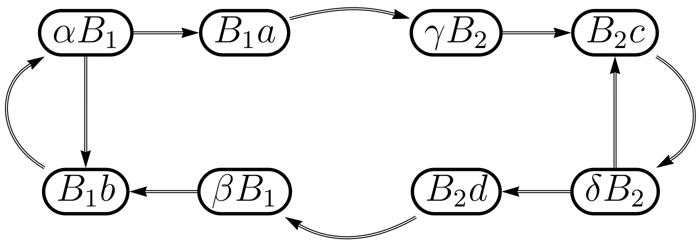
and when some letters are not completely determined (that is if some circuits can play the same role), we use the letters x, y and z.

For example, the morphisms in Equation (4.2) will be denoted by one morphism: $[x, y^{k_1}x, y^{k_1-1}x]$ and it is understood that $\{x, y\} = \{0, 1\}$. Observe that x and y depend on the type of graphs we come from. Indeed, coding the evolution of a graph of type 1, we cannot have $\{x, y\} = \{0, 2\}$ since there are only two circuits in a graph of type 1. Moreover, if for example letters 0, x and y occur in an image, it is understood that 0, x, and y are pairwise distinct.

Also, as explained in Section 4.4, the letter 2 might sometimes not exist in C_n and its existence may change the conditions that exist on the morphism (for example the number of times that one can pass through a loop as for graphs of type 10). Consequently, when the existence of 2 in C_{n+1} does not change anything, we simply put the third component of γ_n into parentheses



(a) B_1 is strong and B_2 is weak



(b) Both B_1 and B_2 are ordinary

Figure 4.11: Evolutions from 8 to 7 or 8.

and when it changes some conditions, we consider 2 morphisms: one with 3 components and one with 2. One last thing is that, for some graphs, we have to determine which right special vertex is the starting vertex of the n-circuits. To this aim, we simply replace in $(v_{i_n}, v_{i_{n+1}})$ what is needed to. When both choices for $v_{i_{n+1}}$ give rise to the same morphisms, we simply replace $v_{i_{n+1}}$ by the symbol $\star$, meaning that $v_{i_{n+1}}$ can be one of the two right special vertices.

Morphisms starting from a graph of type 1

1 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(B, B)	$[x, yx]$, $[yx, x]$	
7 or 8	$(B, \star)$	$[x, y^k x, (y^{k-1}x)]$	$k \geq 2$

Morphisms starting from a graph of type 2

2 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(B, B)	$[x, yzx]$, $[yzx, x]$, $[xy, zy]$ $[xy, zxy]$, $[zxy, xy]$	
2	(B, B)	$[0, 10, 20]$, $[01, 1, 21]$ $[02, 12, 2]$	
3	(B, B)	$[0, 10, 210]$, $[0, 120, 20]$ $[01, 1, 201]$, $[021, 1, 21]$ $[02, 102, 2]$, $[012, 12, 2]$	
4	(B, R)	$[xy^k z, y^\ell z, (xy^{k-1}z)]$ $[y^k z, xy^\ell z, (y^{k-1}z)]$	$k \geq \ell \geq 1$, $k + \ell \geq 3$
	(B, B)	$[x, yx, yzx]$, $[x, yzx, yx]$	
7 or 8	$(B, \star)$	$[x, y^k zx, (y^{k-1}zx)]$ $[x, zy^k x, (zy^{k-1}x)]$ $[x, (yz)^k x, ((yz)^{k-1}x)]$ $[xy, z^k xy, (z^{k-1}xy)]$ $[xy, z^k y, (z^{k-1}y)]$	$k \geq 2$
		$[x, (yz)^k yx, ((yz)^{k-1}yx)]$	$k \geq 1$
10	(B, R)	$[(xy)^k z, y(xy)^\ell z]$	$k \geq 1, \ell \geq 0$ $k + \ell \geq 2$
		$[(xy)^k z, y(xy)^\ell z, (xy)^{k-1}z]$	$k \geq 2, k > \ell \geq 0$
		$[(xy)^k z, y(xy)^\ell z, y(xy)^{\ell-1}z]$	$\ell \geq k \geq 1$
	(B, B)	$[xy, zxy, zy]$	

Morphisms starting from a graph of type 3

3 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(B, B)	$[xy, zy]$, $[xy, z]$, $[x, yz]$	
3	(B, B)	$[0, 10, 20]$, $[0, 10, 2]$, $[0, 1, 20]$ $[01, 1, 21]$, $[01, 1, 2]$, $[0, 1, 21]$ $[02, 12, 2]$, $[02, 1, 2]$, $[0, 12, 2]$	
7 or 8	$(B, \star)$	$[x, yz^k x, (yz^{k-1}x)]$	$k \geq 1$
		$[x, y^k z, (y^{k-1}z)]$	$k \geq 2$
10	(B, B)	$[x, yx, yz]$	
	(B, R)	$[x^k y, zx^\ell y]$	$k \geq 1$, $\ell \geq 0$, $k + \ell \geq 2$
		$[x^k y, zx^\ell y, (x^{k-1}y)]$	$k \geq 2$, $k > \ell \geq 0$
		$[x^k y, zx^\ell y, (zx^{\ell-1}y)]$	$\ell \geq k \geq 1$

Morphisms starting from a graph of type 4

4 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(R, B)	$[x, y]$	$\mathrm{Card}(C_n) = 2$
4	(R, R)	$[0, 1, (2)]$	
	(B, B)	$[0, 10, 20]$, $[0, 20, 10]$	
	(R, B)	$[1, 0, 2]$, $[1, 2, 0]$	
	(B, R)	$[0x^k y, x^\ell y, (0x^{k-1}y)]$ $[x^k y, 0x^\ell y, (x^{k-1}y)]$	$k \geq 1$, $k \geq \ell \geq 0$
7 or 8	$(R, \star)$	$[1, 0, (2)]$	
	$(B, \star)$	$[0, x^k y0, (x^{k-1}y0)]$	$k \geq 1$
10	(R, B)	$[1, 0, 2]$	
	(B, R)	$[0(x0)^k y, (x0)^\ell y]$	$k, \ell \geq 0$, $k + \ell \geq 1$
		$[0(x0)^k y, (x0)^\ell y, 0(x0)^{k-1} y]$	$k \geq 1$, $k \geq \ell \geq 0$
		$[0(x0)^k y, (x0)^\ell y, (x0)^{\ell-1} y]$	$\ell > k \geq 0$

Morphisms starting from a graph of type 5

5 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(R, B)	$[x, y]$	$\mathrm{Card}(C_n) = 2$
10	(R, B)	$[1, 2, 0]$	
	(B, R)	$[1, 01, 2]$	
		$[0^k 2, 1, (0^{k-1}2)]$	$k \geq 1$
		$[2^k 0, 12^\ell 0]$	$k, \ell \geq 0$, $k + \ell \geq 1$
		$[2^k 0, 12^\ell 0, 2^{k-1} 0]$	$k \geq \ell \geq 0$, $k \geq 1$
		$[2^k 0, 12^\ell 0, 12^{\ell-1} 0]$	$\ell > k \geq 0$

Morphisms starting from a graph of type 6

6 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	$(\star, B)$	$[x, yx]$, $[yx, x]$	$\mathrm{Card}(C_n) = 2$
7 or 8	$(\star, \star)$	$[1, 0^k 2, (0^{k-1}2)]$	$k \geq 1$
		$[x, y^k x, (y^{k-1}x)]$	$k \geq 2$ and $\mathrm{Card}(C_n) = 2$
10	$(\star, B)$	$[1, 01, 2]$	
	$(\star, R)$	$[12^k 0, 2^\ell 0]$	$k, \ell \geq 0$, $k + \ell \geq 1$
		$[12^k 0, 2^\ell 0, 12^{k-1} 0]$	$k \geq \ell \geq 0$, $k \geq 1$
		$[12^k 0, 2^\ell 0, 2^{\ell-1} 0]$	$\ell > k \geq 0$

Morphisms starting from a graph of type 7

7 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(R, B)	$[x, y]$	$\text{Card}(C_n) = 2$
7 or 8	$(R, \star)$	$[0, 1, (2)]$	
	$(B, \star)$	$[0, 10, (20)]$	
9	(R, B)	$[0, x, y]$	
	(B, R)	$[01, 1, (02)]$, $[1, 01, (2)]$	
		$[01, 2, (02)]$, $[1, 02, (2)]$	$\text{Card}(C_n) = 3$

Morphisms starting from a graph of type 8

8 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	$(\star, B)$	$[x, yx]$, $[yx, x]$	$\text{Card}(C_n) = 2$
5 or 6	$(\star, \star)$	$[0x, y, (0y)]$, $[x, 0y, (y)]$	$\text{Card}(C_n) = 3$
7 or 8	$(\star, \star)$	$[0, 10, (20)]$	
		$[x, y^k x, (y^{k-1}x)]$	$k \geq 2$, $\text{Card}(C_n) = 2$
9	$(\star, B)$	$[0, x0, y0]$	
	$(\star, R)$	$[01, 1, (02)]$, $[1, 01, (2)]$	$\text{Card}(C_n) = 3$
		$[01, 2, (02)]$, $[1, 02, (2)]$	

Morphisms starting from a graph of type 9

9 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(R, B)	$[x, y]$	$\text{Card}(C_n) = 2$
5 or 6	$(R, \star)$	$[0, 1, (2)]$, $[2, 1, 0]$	
	$(B, \star)$	$[0x, y, (0y)]$, $[x, 0y, (y)]$	
9	(R, R)	$[0, 1, (2)]$	
	(B, B)	$[0, x0, y0]$	

Morphisms starting from a graph of type 10

10 to	$(v_{i_n}, v_{i_{n+1}})$	Morphisms	Conditions
1	(R, B)	$[x, y]$	$\text{Card}(C_n) = 2$
7 or 8	$(R, \star)$	$[1, 0, (2)]$	
	$(B, \star)$	$[0, 2^k 1, (2^{k-1}1)]$	$k \geq 1$
10	(R, R)	$[1, 0, (2)]$	
	(B, B)	$[0, 20, 1]$	
	(R, B)	$[0, 1, 2]$	$\text{Card}(C_n) = 3$
	(B, R)	$[01^k 2, 1^\ell 2]$	$k, \ell \geq 0$, $k + \ell \geq 1$
		$[01^k 2, 1^\ell 2, 01^{k-1} 2]$	$k \geq 1$, $k \geq \ell \geq 0$
		$[01^k 2, 1^\ell 2, 1^{\ell-1} 2]$	$\ell > k \geq 0$

4.6 Proof of Theorem 4.0.1

In the previous section we computed the morphisms γ_n of Definition 4.1.6 when the first difference of complexity is bounded by 2. As expected, this provided an infinite set of morphisms but with a finite number of shapes. In this section, we prove that all these morphisms are actually compositions of morphisms in $\mathcal{S}$ where $\mathcal{S} = \{G, D, M, E_{01}, E_{12}\}$ (see the beginning of the chapter). If $(\Gamma_n)_{n\in\mathbb{N}}$ is the contraction of $(\gamma_n)_{n\in\mathbb{N}}$ as in Lemma 4.1.9 (i.e., for all n Γ_n is strongly primitive and right proper), then we also prove that for all n, $\Gamma_n^{(L)}$ also belongs to $\mathcal{S}^*$. In particular, this will prove Theorem 4.6.1 below. This theorem will then be improved in Chapter 5 to become Theorem 5.8.1 (page 175).

Theorem 4.6.1. *Let $\mathcal{G}$ be the graph represented in Figure 4.8 and let*

$$\mathcal{S} = \{G, D, M, E_{01}, E_{12}\}$$

as defined at the beginning of the chapter. If we add two edges in $\mathcal{G}$ – one from 7 to 10 and one from 8 to 10 –, then we can label the edges of $\mathcal{G}$ by morphisms in $\mathcal{S}^*$ such that for all minimal and aperiodic subshift (X, T) with first difference of complexity bounded by 2, there is a path p in $\mathcal{G}$ labelled by $(\sigma_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that $(\sigma_n)_{n\in\mathbb{N}}$ is a directive word of (X, T) and there is a contraction of it that contains only strongly primitive and proper morphisms.

We need three results to simplify its proof. The first one is a direct consequence of the definitions.

Fact 4.6.2. *If $(\sigma_n)_{n\in\mathbb{N}}$ is an almost primitive directive word, then it is everywhere growing.*

Lemma 4.6.3. *Let (X, T) be a minimal and aperiodic subshift with first difference of complexity bounded by 2. Let $(\gamma_n)_{n\in\mathbb{N}}$ be the directive word of Definition 4.1.6. Suppose that both γ_n and γ_{n+1} are coding an evolution from a graph of type 3 to a graph of type 3. Then if γ_n is equal to*

$$\begin{cases} x \mapsto x \\ y \mapsto yx \\ z \mapsto zx \end{cases} \qquad \left(\textit{resp.} \quad \begin{cases} x \mapsto xy \\ y \mapsto y \\ z \mapsto z \end{cases} \right)$$

for $\{x, y, z\} = \{0, 1, 2\}$, then γ_{n+1} can only be one of the three following *morphisms*

$$\begin{cases} x \mapsto x \\ y \mapsto yx \\ z \mapsto zx \end{cases} \quad \begin{cases} x \mapsto xy \\ y \mapsto y \\ z \mapsto z \end{cases} \quad \begin{cases} x \mapsto xz \\ y \mapsto y \\ z \mapsto z \end{cases}$$

$$\left(\textit{resp.} \quad \begin{cases} x \mapsto xz \\ y \mapsto yz \\ z \mapsto z \end{cases} \quad \begin{cases} x \mapsto x \\ y \mapsto y \\ z \mapsto zx \end{cases} \quad \begin{cases} x \mapsto x \\ y \mapsto y \\ z \mapsto zy \end{cases} \right)$$

Proof. We only have to look at the behaviour of the Rauzy graph when it evolves. Figure 4.12 shows the two possibilities for a graph of type 3 to evolve to a graph of type 3. When computing the morphisms coding these evolutions, we see that what is important to know is which letter corresponds to the top loop in Figure 4.12(a). Indeed, if $\vartheta_{i_n}(x)$ corresponds to the top loop in Figure 4.12(a), the three available morphisms are (the second must be counted twice since y can be replaced by z)

$$\begin{cases} x \mapsto x \\ y \mapsto yx \\ z \mapsto zx \end{cases} \quad \text{and} \quad \begin{cases} x \mapsto xy \\ y \mapsto y \\ z \mapsto z \end{cases} .$$

The evolution represented in Figure 4.12(b) is coded by the first morphism and the evolution represented in Figure 4.12(c) is coded by the second one (where $\vartheta_{i_n}(y)$ is the leftmost loop in Figure 4.12(a)).

After the first evolution, the graph becomes again a graph as in Figure 4.12(a) where the circuit $\vartheta_{i_{n+1}}(x)$ still corresponds to the top loop. The available morphisms are therefore the same as before the evolution.

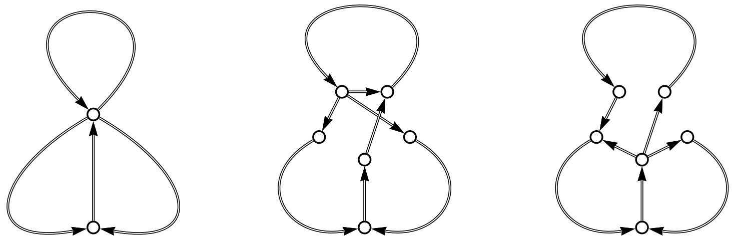
After the second evolution, the graph becomes again a graph as in Figure 4.12(a) but the top loop is the circuit $\vartheta_{i_{n+1}}(z)$. The available morphisms are therefore the same as before the evolution but with x and z exchanged. □

Lemma 4.6.4. *Let $\mathcal{G}$ be the graph of graphs represented at Figure 4.8. The sets of products of morphisms coding the sequences of evolutions in*

$$7 \to (9 \to 9)^+ \to 5 \textit{ or } 6 \quad \textit{and} \quad 8 \to (9 \to 9)^* \to 5 \textit{ or } 6$$

are the same and equal to

$$\left\{ [0x0^j, y0^j, (0y0^j)], [x0^j, 0y0^j, (y0^j)] \mid \{x,y\} = \{1,2\},\ j \geq 0 \right\}.$$



(a) Rauzy graph of type 3 (b) Evolution 1 (c) Evolution 2

Figure 4.12: Evolutions of a graph of type 3 to a graph of type 3.

Proof. This is simply a computation. When the graph is of type 7 and $(v_{i_n}, v_{i_{n+1}}) = (R, B)$, we have $\gamma_n = [0, x, y]$. Then the path can stay in the vertex 9 for a while with the morphism $[0, x0, y0]$ (which creates the power of 0 at the end of the images). Finally, the graph evolves to a graph of type 5 or 6 with the morphism $[0x, y, (0y)]$ or $[x, 0y, (y)]$. We only have to make the product to see that this corresponds to the morphisms of the lemma.

When the graph is of type 7 and $(v_{i_n}, v_{i_{n+1}}) = (B, R)$, we have

$$\gamma_n \in \{[01, 1, (02)], [1, 01, (2)], [01, 2, (02)], [1, 02, (2)]\}.$$

Then the path can stay in the vertex 9 for a while with the identity morphism and finally, the graph evolve to a graph of type 5 or 6 with a morphism in

$$\{[0, 1, (2)], [2, 1, 0]\}.$$

However, the definition of ϑ_n for graphs of type 5 and 6 implies that when the morphisms from 7 to 9 belongs to

$$\{[01, 1, (02)], [1, 01, (2)]\} \quad (\text{resp. } \{[01, 2, (02)], [1, 02, (2)]\})$$

then the morphism from 9 to 5 or 6 can only be $[2, 1, 0]$ (resp. $[0, 1, (2)]$). Once again we only have to make the product to see that this corresponds to the morphisms of the lemma.

When the graph is of type 8, the morphisms coding an evolution to a graph of type 5 or 6 already correspond to the morphisms of the lemma. If the graph first evolves to a graph of type 9, we only have to repeat the computation made starting from a graph of type 7. □

Now we can prove Theorem 4.6.1.

Proof of Theorem 4.6.1. First, Lemma 4.1.8 states that the sequence $(\gamma_n)_{n \in \mathbb{N}}$ of Definition 4.1.6 is a directive word of (X, T). Then, Lemma 4.1.9 ensures that one can find a contraction of that directive word such that all morphisms are strongly primitive[2] and right proper. Finally, Lemma 4.1.10 and Proposition 4.1.11 show how we can modify the obtained contraction in such a way that all morphisms are proper.

Now, it is a consequence of the construction that if we label the edge of $\mathcal{G}$ with the morphisms γ_n given in Section 4.5, the sequence of morphisms $(\gamma_n)_{n\in\mathbb{N}}$ labels an infinite path in $\mathcal{G}$. When looking at all these morphisms (in the previous section), we see that a large majority of them are already right proper. Moreover, when considering the left conjugate of some morphisms in $(\gamma_n)_{n\in\mathbb{N}}$, the almost primitivity still holds true. Also, if γ_n is a right proper morphism, then for all non-negative integers $i \leq n$, $\gamma_i\gamma_{i+1}\cdots\gamma_n^{(L)}$ is a left proper morphism. Consequently, if there is a sub-sequence $(\gamma_{m_n})_{n\in\mathbb{N}}$ of $(\gamma_n)_{n\in\mathbb{N}}$ that contains only right proper morphisms, we only have to consider the sequence of morphisms $(\beta_n)_{n\in\mathbb{N}}$ such that for all n,

$$\beta_n = \begin{cases} \gamma_n & \text{if } \forall k \in \mathbb{N},\ n \neq m_{2k} \\ \gamma_n^{(L)} & \text{otherwise} \end{cases}.$$

Then, $(\beta_n)_{n\in\mathbb{N}}$ is almost primitive and all products of morphisms

$$\beta_{m_{2n}}\beta_{m_{2n}+1}\cdots\beta_{m_{2n+1}}\cdots\beta_{m_{2(n+1)}-1}$$

are proper. Observe that γ_0 is always a right proper morphism. Indeed, if $p(1)-p(0)=2$ (resp. 3) then G_0 is of type 1 (resp. 2) and all morphisms starting from a graph of type 1 or 2 are right proper.

To complete the proof, there are two steps left:

1. show that all morphisms γ_n belong to $\mathcal{S}^*$ and that their left conjugate (when they are right proper) also belong to $\mathcal{S}^*$;

2. study what happens when there are only finitely many right proper morphisms in $(\gamma_n)_{n\in\mathbb{N}}$.

The first point will be done after the end of the proof (from page 133 to 136). For the second point, let us decompose the problem. The graph $\mathcal{G}$ has four strongly connected components that are $C_1 = \{2\}$, $C_2 = \{3\}$, $C_3 = \{4\}$ and $C_4 = \{1,5,6,7,8,9,10\}$. We can study them separately, i.e., for all $i \in \{1,2,3,4\}$, we study paths that ultimately stay in C_i and such that the label $(\gamma_n)_{n\in\mathbb{N}}$ contains only finitely many right proper morphisms. Then, in the final contraction $(\Gamma_n)_{n\in\mathbb{N}}$ (to get only strongly primitive and proper morphisms), we only have to put all morphisms $\gamma_0\cdots\gamma_N$ in Γ_0 where N is

[2]This comes from the almost primitivity of $(\gamma_n)_{n\in\mathbb{N}}$.

the greatest integer such that γ_N does not code an evolution from a vertex in C_i.

All morphisms labelling the unique edge of Component C_1 are right proper so there is nothing to do for that component.

The component C_2 contains only the vertex 3. If $(\gamma_n)_{n\in\mathbb{N}}$ contains only finitely many right proper morphisms, there is an integer k such that

$$(\gamma_n)_{n\geq k} \in \{[0,10,2],[0,1,20],[01,1,2],[0,1,21],[02,1,2],[0,12,2]\}^{\mathbb{N}}.$$

But, Lemma 4.6.3 implies that if γ_n is

$$\begin{cases} x \mapsto xy \\ y \mapsto y \\ z \mapsto z \end{cases}$$

for $n \geq k$ and $\{x,y,z\} = \{0,1,2\}$, then γ_{n+1} can only be one of the two following morphisms

$$\begin{cases} x \mapsto x \\ y \mapsto y \\ z \mapsto zx \end{cases} \qquad \begin{cases} x \mapsto x \\ y \mapsto y \\ z \mapsto zy \end{cases}.$$

Consequently, the morphism $\gamma_n \circ \gamma_{n+1}$ can be one of the two following morphisms

$$\begin{cases} x \mapsto xy \\ y \mapsto y \\ z \mapsto zxy \end{cases} \qquad \begin{cases} x \mapsto xy \\ y \mapsto y \\ z \mapsto zy \end{cases}.$$

These morphisms are right proper and it is easily seen that they belong to $\mathcal{S}^*$ (and that so do their respective left conjugates).

The component C_3 contains only the vertex 4. Among the morphisms from 4 to 4, the only ones that are not right proper are those in $Q = \{[0,1,(2)],[1,0,2],[1,2,0]\}$ and for all $m \geq 0$, $(\gamma_n)_{n\geq m}$ cannot belong to $Q^{\mathbb{N}}$ otherwise this would contradict Lemma 4.6.2. Consequently, there are infinitely many right proper morphisms in $(\gamma_n)_{n\in\mathbb{N}}$.

The component C_4 is $\{1,5,6,7,8,9,10\}$. First, as mentioned earlier, all morphisms coding an evolution from a graph of type 1 are right proper. Consequently, we have to consider paths in $\mathcal{G}$ that ultimately stay into $\{5,6,7,8,9,10\}$. We can see in $\mathcal{G}$ that if p does not go infinitely often through a vertex in $\{7,8\}$, then it ultimately stays either in the vertex 9 or in the vertex 10. Moreover, for all morphisms γ_n coding an evolution from a graph of type 9 to a graph of type 9, we have $\gamma_n(0) = 0$. Thus, from Lemma 4.6.2, either a path p ultimately stays in the vertex 10, or it goes infinitely often through a vertex in $\{7, 8\}$. In the first case, the morphism such that $(v_{i_n}, v_{i_{n+1}}) = (R, R)$ is $[1, 0, (2)]$ and Lemma 4.6.2 implies that for all $m \geq 0$, we cannot have $(\gamma_n)_{n \geq m} = [1, 0, (2)]^\omega$ otherwise all m-circuits would be constant. Consequently, the sequence $(v_{i_n})_{n \in \mathbb{N}}$ has to contain infinitely many occurrences of the vertex B. It also has to contain infinitely many occurrences of the vertex R because the morphism such that $(v_{i_n}, v_{i_{n+1}}) = (B, B)$ is $[0, 20, 1]$ and this would again contradict Lemma 4.6.2. Consequently, since all morphisms such that $(v_{i_n}, v_{i_{n+1}}) = (B, R)$ are right proper, there are infinitely many right proper morphisms in $(\gamma_n)_{n \in \mathbb{N}}$.

Now we still have to study the paths p that goes infinitely many times through a vertex in $\{7, 8\}$ but only finitely many times through 1. Suppose that we are starting from 7 or 8 in $\mathcal{G}$ and let us show that we always have to consider sequences of evolutions whose corresponding product of morphisms is right proper. The idea is to try to avoid right proper morphisms γ_n and to show that this always yields to build products of morphisms that are right proper.

First, the only non-right proper morphism coding an evolution from $\{7, 8\}$ to $\{7, 8\}$ is the identity morphisms from 7 to $\{7, 8\}$ and for all K, Lemma 4.6.2 implies that $(\gamma)_{n \geq K} \neq [0, 1, (2)]^\omega$. To avoid right proper morphisms, the path has therefore to leave $\{7, 8\}$. When looking at all possible evolutions, we see that the only possibility is to eventually evolve to a graph of type 5 or 6 (by possibly first evolving to a graph of type 9). Then, Lemma 4.6.4 ensures that the product of morphisms $\gamma_n \cdots \gamma_m$ coding this sequence of evolutions is the same whatever it started from 7 or from 8. That lemma also provides the possible morphisms that are

$$[0x0^j, y0^j, (0y0^j)] \quad \text{and} \quad [x0^j, 0y0^j, (y0^j)].$$

Since we want to avoid right proper morphisms, we consider that $j = 0$ so we obtain the morphisms

$$\eta_1 = [0x, y, (0y)] \quad \text{and} \quad \eta_2 = [x, 0y, (y)].$$

We see in these morphisms that the image of 1 and of 2 end with the same letter. Then, we can see that for all non-right proper morphisms coding an evolution from a graph of type 5 or 6 and avoiding the vertex 1 (except for the morphism $[1, 2, 0]$ from 5 to 10), all images ends with 1 or 2. Consequently, the composition of η_1 or η_2 with one of these morphisms provides a right proper morphism. The list of all possibilities is given below. Their decomposition into morphisms of $\mathcal{S}$ (and the decompositions of their respective left conjugates) are given after the end of the proof (on page 136).

$$\begin{aligned}
\eta_1 \circ [1, 01, 2] &= [y, 0xy, 0y] \\
\eta_1 \circ [0^k2, 1, (0^{k-1}2)] &= [(0x)^k0y, y, ((0x)^{k-1}0y)] \\
\eta_1 \circ [1, 0^k2, (0^{k-1}2)] &= [y, (0x)^k0y, ((0x)^{k-1}0y)] \\
\eta_2 \circ [1, 01, 2] &= [0y, x0y, y] \\
\eta_2 \circ [0^k2, 1, (0^{k-1}2)] &= [x^ky, 0y, (x^{k-1}y)] \\
\eta_2 \circ [1, 0^k2, (0^{k-1}2)] &= [0y, x^ky, (x^{k-1}y)]
\end{aligned}$$

Therefore, we only have to add an edge in $\mathcal{G}$ from $\{7, 8\}$ to 10 labelled by $[y, 0xy, 0y]$, $[(0x)^k0y, y, ((0x)^{k-1}0y)]$, $[0y, x0y, y]$ and $[x^ky, 0y, (x^{k-1}y)]$ and also to add the morphisms $[y, 0xy, 0y]$, $[y, (0x)^k0y, ((0x)^{k-1}0y)]$, $[0y, x0y, y]$ and $[0y, x^ky, (x^{k-1}y)]$ to the label of each edge from $\{7, 8\}$ to $\{7, 8\}$.

The last remaining case is when the graph has evolved to a graph of type 5 with η_1 or η_2 and then to a graph of type 10 with $[1, 2, 0]$. In $\eta_1 \circ [1, 2, 0]$ and $\eta_2 \circ [1, 2, 0]$, we see that the images of 0 and 1 end with the same letter. Then, from a graph G_m of type 10 with $v_m = B$, for all non-right proper morphisms coding an evolution avoiding the vertex 1 in G, all images end with 0 or 1. Therefore, the product provides again a right proper morphism. The list of possibilities is given below and the decompositions into morphisms in $\mathcal{S}$ are given after the end of the proof (on page 136).

$$\begin{aligned}
\eta_1 \circ [1, 2, 0] \circ [0, 2^k1, (2^{k-1}1)] &= [y, (0x)^k0y, ((0x)^{k-1}0y)] \\
\eta_1 \circ [1, 2, 0] \circ [0, 20, 1] &= [y, (0x)y, 0y] \\
\eta_2 \circ [1, 2, 0] \circ [0, 2^k1, (2^{k-1}1)] &= [0y, x^ky, (x^{k-1}y)] \\
\eta_2 \circ [1, 2, 0] \circ [0, 20, 1] &= [0y, x0y, y]
\end{aligned}$$

Once again we would have to add the two morphisms (depending on a parameter k) $[y, (0x)^k0y, ((0x)^{k-1}0y)]$ and $[0y, x^ky, (x^{k-1}y)]$ to the label of each edge from $\{7, 8\}$ to $\{7, 8\}$ and to add the morphisms $[y, (0x)y, 0y]$ and $[0y, x0y, y]$ to the label of each edge from $\{7, 8\}$ to 10. This is actually already done by the previous cases. □

To complete the proof of Theorem 4.6.1, let us show that all morphisms γ_n belong to $\mathcal{S}^*$. To avoid long decompositions, we define the morphism $E_{0,2} = [2, 1, 0] = E_{0,1}E_{1,2}E_{0,1}$. We also define the following morphisms. For $G_{x,y}$ (resp. $D_{x,y}$), read "add y to the left (resp. right) of x". For $M_{x,y}$, read "map x to y".

$$\begin{array}{ll}
G_{0,1} = [10,1,2] = G & D_{0,1} = [01,1,2] = D \\
G_{0,2} = [20,1,2] = E_{1,2}GE_{1,2} & D_{0,2} = [02,1,2] = E_{1,2}DE_{1,2} \\
G_{1,0} = [0,01,2] = E_{0,1}GE_{0,1} & D_{1,0} = [0,10,2] = E_{0,1}DE_{0,1} \\
G_{1,2} = [0,21,2] = E_{0,1}G_{0,2}E_{0,1} & D_{1,2} = [0,12,2] = E_{0,1}D_{0,2}E_{0,1} \\
G_{2,0} = [0,1,02] = E_{0,2}G_{0,2}E_{0,2} & D_{2,0} = [0,1,20] = E_{0,2}D_{0,2}E_{0,2} \\
G_{2,1} = [0,1,12] = E_{1,2}G_{1,2}E_{1,2} & D_{2,1} = [0,1,21] = E_{1,2}D_{1,2}E_{1,2} \\
M_{0,1} = [1,1,2] = E_{0,2}ME_{0,2} & M_{1,0} = [0,0,2] = E_{0,1}M_{0,1} \\
M_{0,2} = [2,1,2] = E_{0,1}E_{1,2}ME_{0,1} & M_{2,0} = [0,1,0] = E_{0,2}M_{0,2} \\
M_{1,2} = [0,2,2] = E_{1,2}M & M_{2,1} = [0,1,1] = M
\end{array}$$

Now we can decompose all morphisms γ_n. Here, we only present the decompositions of the morphisms depending of some exponents k or ℓ; the reader is invited to check the conditions that exist on k and ℓ in Section 4.5 (from page 122 to 125). When a morphism γ_n is right proper, we also give the decomposition of $\gamma_n^{(L)}$ into morphisms of $\mathcal{S}$.

Remark 4.6.5. To get simplest decompositions of $\gamma_n^{(L)}$, we sometimes consider another definition of it. Indeed, when γ_n is right proper and such that $\gamma_n(C_{n+1}) \subset C_n^* xy$ for two letters x, y, we define $\gamma_n^{(L)}$ as the morphism such that $\gamma_n^{(L)}(a) = xyu$ whenever $\gamma_n(a) = uxy$ for $u \in C_n^*$ and $x, y \in C_n$. We only have to adapt Lemma 4.1.10 (page 106) to keep all results true.

Remark 4.6.6. In the decompositions given below, some morphisms can commute and some other cannot. Consequently, the $\mathcal{S}$-adic characterization that we will get in the next chapter is defined up to some commutations of morphisms in the directive word. Maybe it would be interesting to see if there is a way to define a *normalized* $\mathcal{S}$*-adic representation* as it is done for episturmian sequences (see [GLR09]).

Decomposition of morphisms starting from a graph of type 1

1 to	Morphisms	Decomposition
7 or 8	$\gamma_n = [x, y^k x, y^{k-1} x]$	$M_{2,x} G_{2,y}^{k-1} D_{y,2} [x, y, 2]$
	$\gamma_n^{(L)} = [x, xy^k, xy^{k-1}]$	$M_{2,x} D_{2,y}^{k-1} G_{y,2} [x, y, 2]$

Decomposition of morphisms starting from a graph of type 2

2 to	Morphisms	Decomposition
4	$\gamma_n = [xy^k z, y^\ell z, xy^{k-1}z]$ $\gamma_n^{(L)} = [zxy^k, zy^\ell, zxy^{k-1}]$	$D_{x,y}^{k-\ell} G_{z,y}^{\ell-1} D_{y,z} G_{z,x} D_{x,y}[x,y,z]$ $D_{x,y}^{k-1} G_{x,z} D_{z,y}^{\ell} G_{y,x}[y,z,x]$
	$\gamma_n = [y^k z, xy^\ell z, y^{k-1}z]$ $\gamma_n^{(L)} = [zy^k, zxy^\ell, zy^{k-1}]$	$D_{x,y}^{\ell} D_{x,z} G_{z,y}^{k-1} D_{y,z}[y,x,z]$ $D_{x,y}^{\ell} G_{x,z} D_{z,y}^{k-1} G_{y,z}[y,x,z]$
7 or 8	$\gamma_n = [x, y^k zx, y^{k-1}zx]$ $\gamma_n^{(L)} = [x, xy^k z, xy^{k-1}z]$	$G_{z,y}^{k-1} D_{z,x} D_{y,z}[x,y,z]$ $G_{z,y}^{k-1} D_{y,z} G_{y,x} G_{z,x}[x,y,z]$
	$\gamma_n = [x, zy^k x, zy^{k-1}x]$ $\gamma_n^{(L)} = [x, xzy^k, xzy^{k-1}]$	$D_{z,y}^{k-1} G_{y,z} D_{y,x} D_{z,x}[x,y,z]$ $D_{z,y}^{k-1} G_{z,x} G_{y,z}[x,y,z]$
	$\gamma_n = [x, (yz)^k x, (yz)^{k-1}x]$ $\gamma_n^{(L)} = [x, x(yz)^k, x(yz)^{k-1}]$	$D_{y,z} M_{z,x} G_{z,y}^{k-1} D_{y,z}[x,y,z]$ $D_{y,z} M_{z,x} D_{z,y}^{k-1} G_{y,z}[x,y,z]$
	$\gamma_n = [x, (yz)^k yx, (yz)^{k-1}yx]$ $\gamma_n^{(L)} = [x, x(yz)^k y, x(yz)^{k-1}y]$	$G_{z,y} G_{y,z}^{k-1} D_{y,x} D_{z,y}[x,z,y]$ $G_{z,y} G_{y,z}^{k-1} D_{z,y} G_{y,x} G_{z,x}[x,z,y]$
	$\gamma_n = [xy, z^k xy, z^{k-1}xy]$ $\gamma_n^{(L)} = [xy, xyz^k, xyz^{k-1}]$	$D_{x,y} M_{y,x} G_{y,z}^{k-1} D_{z,y}[x,z,y]$ $D_{x,y} M_{y,x} D_{y,z}^{k-1} G_{z,y}[x,z,y]$
	$\gamma_n = [xy, z^k y, z^{k-1}y]$ $\gamma_n^{(L)} = [yx, yz^k, yz^{k-1}]$	$D_{x,y} G_{y,z}^{k-1} D_{z,y}[x,z,y]$ $G_{x,y} D_{y,z}^{k-1} G_{z,y}[x,z,y]$
10	$\gamma_n = [(xy)^k z, y(xy)^\ell z, (xy)^{k-1}z]$ $\gamma_n^{(L)} = [z(xy)^k, zy(xy)^\ell, zy(xy)^{\ell-1}]$	$D_{x,y} E_{x,y}[y^k z, xy^\ell z, y^{k-1}z]$ (see 2 to 4) $D_{x,y} E_{x,y}[zy^k, zxy^\ell, zxy^{\ell-1}]$ (see 2 to 4 with k and ℓ exchanged)

Decomposition of morphisms starting from a graph of type 3

3 to	Morphisms	Decompositions
7 or 8	$\gamma_n = [x, zy^k x, zy^{k-1}x]$ $\gamma_n^{(L)} = [x, xzy^k, xzy^{k-1}]$	see 2 to 7 or 8
	$\gamma_n = [x, y^k z, y^{k-1}z]$	$G_{z,y}^{k-1} D_{y,z}[x,y,z]$
10	$\gamma_n = [x^k y, zx^\ell y, x^{k-1}y]$ $\gamma_n^{(L)} = [yx^k, yzx^\ell, yx^{k-1}]$	see 2 to 4
	$\gamma_n = [x^k y, zx^\ell y, zx^{\ell-1}y]$ $\gamma_n^{(L)} = [yx^k, yzx^\ell, yzx^{\ell-1}]$	see 2 to 4 with k and ℓ exchanged

Decomposition of morphisms starting from a graph of type 4

4 to	Morphisms	Decompositions
4	$\gamma_n = [0x^k y, x^\ell y, 0x^{k-1}y]$ $\gamma_n^{(L)} = [y0x^k, yx^\ell, y0x^{k-1}]$	see 2 to 4
	$\gamma_n = [x^k y, 0x^\ell y, x^{k-1}y]$ $\gamma_n^{(L)} = [yx^k, y0x^\ell, yx^{k-1}]$	see 2 to 4
7 or 8	$\gamma_n = [0, x^k y0, x^{k-1}y0]$ $\gamma_n^{(L)} = [0, 0x^k y, 0x^{k-1}y]$	see 2 to 7 or 8
10	$\gamma_n = [0(x0)^k y, (x0)^\ell y, 0(x0)^{k-1}y]$ $\gamma_n^{(L)} = [y0(x0)^k, y(x0)^\ell, y0(x0)^{k-1}]$	see 2 to 10
	$\gamma_n = [0(x0)^k y, (x0)^\ell y, (x0)^{\ell-1}y]$ $\gamma_n^{(L)} = [y0(x0)^k, y(x0)^\ell, y(x0)^{\ell-1}]$	see 2 to 10

Decomposition of morphisms starting from a graph of type 5

5 to	Morphisms	Decompositions
10	$\gamma_n = [0^k 2, 1, 0^{k-1}2]$	see 3 to 7 or 8
	$\gamma_n = [2^k 0, 12^\ell 0, 2^{k-1}0]$ $\gamma_n^{(L)} = [02^k, 012^\ell, 02^{k-1}]$	see 2 to 4 see 2 to 4
	$\gamma_n = [2^k 0, 12^\ell 0, 12^{\ell-1}0]$ $\gamma_n^{(L)} = [02^k, 012^\ell, 012^{\ell-1}]$	see 2 to 4 with k and ℓ exchanged see 2 to 4 with k and ℓ exchanged

Decomposition of morphisms starting from a graph of type 6

6 to	Morphisms	Decompositions
7 or 8	$\gamma_n = [1, 0^k 2, 0^{k-1}2]$	see 3 to 7 or 8
	$\gamma_n = [x, y^k x, y^{k-1}x]$ $\gamma_n^{(L)} = [x, xy^k, xy^{k-1}]$	see 1 to 7 or 8
10	$\gamma_n = [12^k 0, 2^\ell 0, 12^{k-1}0]$ $\gamma_n^{(L)} = [012^k, 02^\ell, 012^{k-1}]$	see 2 to 4
	$\gamma_n = [12^k 0, 2^\ell 0, 12^{\ell-1}0]$ $\gamma_n^{(L)} = [012^k, 02^\ell, 012^{\ell-1}]$	see 2 to 4

Decomposition of morphisms starting from a graph of type 8

8 to	Morphisms	Decompositions
7 or 8	$\gamma_n = [x, y^k x, y^{k-1}x]$ $\gamma_n^{(L)} = [x, xy^k, xy^{k-1}]$	see 1 to 7 or 8

Decomposition of morphisms starting from a graph of type 10

10 to	Morphisms	Decompositions
10	$\gamma_n = [01^k2, 1^\ell 2, 01^{k-1}2]$ $\gamma_n^{(L)} = [201^k, 21^\ell, 201^{k-1}]$	see 2 to 4
	$\gamma_n = [01^k2, 1^\ell 2, 1^{\ell-1}2]$ $\gamma_n^{(L)} = [201^k, 21^\ell, 21^{\ell-1}]$	see 2 to 4

Decomposition of additional morphisms occurring in the proof of Theorem 4.6.1

7 or 8 to	Morphisms	Decompositions
7 or 8	$\gamma_n = [y, (0x)^k 0y, ((0x)^{k-1}0y)]$ $\gamma_n^{(L)} = [y, y(0x)^k 0, (y(0x)^{k-1}0)]$	see 2 to 7 or 8
	$\gamma_n = [0y, x^k y, (x^{k-1}y)]$ $\gamma_n^{(L)} = [y0, yx^k, (yx^{k-1})]$	see 2 to 7 or 8
10	$\gamma_n = [(0x)^k 0y, y, ((0x)^{k-1}0y)]$ $\gamma_n^{(L)} = [y(0x)^k 0, y, (y(0x)^{k-1}0)]$	see just above (7 or 8 to 7 or 8)
	$\gamma_n = [x^k y, 0y, (x^{k-1}y)]$ $\gamma_n^{(L)} = [yx^k, y0, (yx^{k-1})]$	see just above (7 or 8 to 7 or 8)

Chapter 5

$\mathcal{S}$-adic characterization of minimal subshifts with complexity $2n$

In the previous chapter we showed that the graph of graphs $\mathcal{G}$ (see Figure 4.8 on page 112) can be slightly modified in such a way that for any minimal and aperiodic subshift with first difference of complexity bounded by 2, there is a path in $\mathcal{G}$ that describes its directive word. As explained in Section 5.1 below, the converse is false. In this section we show that if we modify even more the graph of graphs, then we can obtain an $\mathcal{S}$-adic characterization of minimal subshifts with the considered complexity. In other words, we manage to determine the condition C of the S-adic conjecture for that particular case. This is Theorem 5.8.1 on page 175. In all this section, $\mathcal{S}$ is still the set of 5 morphisms defined at the beginning of Chapter 4.

Remark 5.0.1. In this section we will have to give many details on Rauzy graphs and on their evolutions. Let us recall all needed notations.

1. $(i_n)_{n\in\mathbb{N}}$ is the growing sequence of integers such that for all integers $m \geq 0$, $G_m(X)$ contains a bispecial vertex if and only if $m = i_n$ for some integer $n \geq 0$;

2. $(v_n)_{n\in\mathbb{N}}$ is a sequence of right special vertices as in Lemma 4.1.3 (so $(v_{i_n})_{n\in\mathbb{N}}$ is the sub-sequence of $(v_n)_{n\in\mathbb{N}}$ that corresponds to graphs with at least one bispecial vertex);

3. for all $n \in \mathbb{N}$, γ_n codes the evolution from G_{i_n} to $G_{i_{n+1}}$;

4. $(\sigma_n)_{n\in\mathbb{N}}$ is the directive word of Theorem 4.6.1; it is composed of the decompositions into $\mathcal{S}^*$ of morphisms γ_n or $\gamma_n^{(L)}$ (or even sometimes of $(\gamma_n \cdots \gamma_{n+k})^{(L)}$ for some integer k). Observe that we still sometimes

consider the definition of $\gamma^{(L)}$ given in Remark 4.6.5. The decomposition considered actually always corresponds to the one given from Page 133 to Page 136.

5.1 Valid paths

To get the $\mathcal{S}$-adic characterization of Theorem 5.8.1, we need to be able to explicitly describe all paths in $\mathcal{G}$ that correspond to the sequence of evolutions of a minimal and aperiodic subshift with first difference of complexity bounded by two. Therefore, our aim is to modify the graph $\mathcal{G}$ in such a way that these paths can be easily described.

The first step is to understand how we can describe the "good labelled paths" in $\mathcal{G}$, hence the good sequences of evolutions. To this aim, we introduce the notions of *valid directive word* and of *valid path*.

Definition 5.1.1. A sequence of morphisms $(\sigma_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ is said to be a *valid directive word* if it is an $\mathcal{S}$-adic representation of a minimal subshift with first difference of complexity bounded by 2. We also say that a finite sequence of morphisms $\sigma_0\sigma_1\cdots\sigma_k \in \mathcal{S}^*$ is *valid* if it is the prefix of a valid directive word $(\sigma_n)_{n\in\mathbb{N}}$.

Since all valid sequences of morphisms in $\mathcal{S}$ describe paths in $\mathcal{G}$, we also say that a labelled (finite or infinite) path p in $\mathcal{G}$ is *valid* if we can modify its label by contracting it and by replacing some right proper morphisms by their left conjugates such that the decomposition of the modified label into elements of $\mathcal{S}$ is valid.

There exist several reasons for which a given labelled path in $\mathcal{G}$ is not valid: two conditions (due to Proposition 3.3.8) are that the directive word $(\sigma_n)_{n\in\mathbb{N}}$ has to be almost primitive and must admit a contraction that contains only proper morphisms. Example 5.1.2 and Example 5.1.3 below show two sequences of evolutions which are forbidden because their respective directive words do not satisfy the almost primitivity.

Example 5.1.2. Sturmian subshifts have Rauzy graphs of type 1 for all n. However if, for instance, we consider that for all n, the morphism γ_n coding the evolution of G_{i_n} is $[0, 10]$, the directive word is not almost primitive and the sequence of Rauzy graphs $(G_{i_n})_{n\in\mathbb{N}}$ is such that for all n, $i_n = n$ and $\lambda_R(\vartheta_n(0)) = 0$ and $\lambda_R(\vartheta_n(1)) = 10^n$ (the reduced Rauzy graph g_n is represented in Figure 5.1). By Fact 1.5.7 (page 39), the language of the obtained subshift X is

$$L(X) = \left\{0^j \mid j \in \mathbb{N}\right\} \cup \left\{0^j10^k \mid j, k \in \mathbb{N}\right\}$$

so it has complexity $p_X(n) = n + 1$ for all n but it is not minimal, so not S-adic. One can easily check that it actually corresponds to the subshift generated by the sequence $\mathbf{w} = \cdots 000.1000 \cdots$.

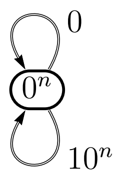
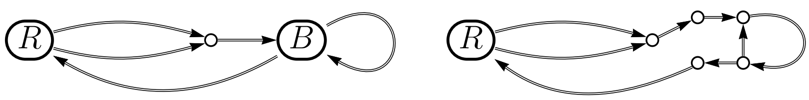


Figure 5.1: Reduced Rauzy graph g_n of $\cdots 000.1000 \cdots$.

Example 5.1.3. Let us consider a path in $\mathcal{G}$ that ultimately stays in the vertex 9. Figure 5.2 represents the only way for a Rauzy graph G_{i_n} of type 9 to evolve to a Rauzy graph of type 9. We can see that in this evolution, the i_n-circuit $\vartheta_{i_n}(0)$ starting from the vertex B (i.e., the loop that does not pass through the vertex R) "stays unchanged" in G_{i_n+1}, i.e., $\psi_{i_n,R}(\vartheta_{i_n+1}(0)) = \vartheta_{i_n}(0)$. Consequently, there is an integer n such that $\vartheta_{i_n+1}(0)$ is a constant circuit which is forbidden for minimal subshift (Lemma 3.1.16). One can also check on page 125 that for all morphisms γ_n coding such an evolution, we have $\gamma_n(0) = 0$. As there is no other evolution from a Rauzy graph of type 9 to a Rauzy graph of type 9, the directive word cannot be almost primitive.



(a) Before evolution (b) After evolution

Figure 5.2: Evolution of a graph of type 9 to a graph of type 9.

The two previously given conditions (being almost primitivity and proper) are not sufficient: there is also a "local condition" that has to be satisfied. Indeed, Example 5.1.4 below shows that for some prefixes $\gamma_0 \cdots \gamma_k$ labelling a finite path p in $\mathcal{G}$, not every edge starting from $i(p)$ is allowed.

Example 5.1.4. Consider a graph G_{i_n} of type 1 that evolves to a graph as in Figure 4.7(c) (Page 111), hence to a graph of type 7 or 8. We write $R_1 = \alpha B$

and $R_2 = \beta B$ and suppose that $v_{i_n+1} = R_1$. The morphism coding this evolution is $[x, y^k x, y^{k-1}x]$ for some integer $k \geq 2$. If we suppose $k \geq 3$, this means that the circuits $\vartheta_{i_n+1}(1)$ and $\vartheta_{i_n+1}(2)$ respectively go through $k-1$ and $k-2$ times in the loop $R_2 \to R_2$. By construction of the Rauzy graphs, this means that the shortest bispecial factor B' (with respect to the radix order) admitting R_2 as a suffix is an ordinary bispecial factor. Let $m > n$ be an integer such that B' is a bispecial vertex in G_{i_m}. Since B' is ordinary bispecial, there is a right special factor R' of length $i_m + 1$ that admits B' as a suffix. Moreover, since v_{i_m} is not B' (as R_1 has to be a suffix of v_{i_m}), the right special factor v_{i_m+1} is not R'. Consequently there are two right special factors in G_{i_m+1} so G_{i_m+1} is not of type 1.

To be a valid labelled path in $\mathcal{G}$ the three previous examples show that a given path p must necessary satisfy at least two conditions: a local one about its prefixes (Example 5.1.4) and a global one about almost primitivity (Example 5.1.2 and Example 5.1.4). The next result states that the converse is true.

Proposition 5.1.5. *A path p in $\mathcal{G}$ labelled by $(\sigma_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ is valid if and only if both following conditions are satisfied.*

1. *All prefixes of p are valid*[1]*;*

2. *$(\sigma_n)_{n\in\mathbb{N}}$ is almost primitive and a contraction of it contains only proper morphisms*[2]*.*

Proof. The first condition is obviously necessary and the second condition comes from Theorem 4.0.1 (since the $\mathcal{S}$-adic representation of that theorem is obtained by n-circuits, like in this chapter). For the sufficient part, if all prefixes of $(\sigma_n)_{n\in\mathbb{N}}$ are allowed, it implies that we can build a sequence of Rauzy graphs $(G_n)_{n\in\mathbb{N}}$ such that for all n, G_n is as represented in Figure 4.2 to Figure 4.4 and evolves to G_{n+1}. To these Rauzy graphs we can associate a sequence of languages $(L(G_n))_{n\in\mathbb{N}}$ defined as the set of finite words labelling paths in G_n. By construction we obviously have $L(G_{n+1}) \subset L(G_n)$ and the language

$$L = \bigcap_{n\in\mathbb{N}} L(G_n)$$

is factorial, prolongable (Definition 1.1.2 and Definition 1.1.1) and such that $1 \leq p_L(n+1) - p_L(n) \leq 2$ for all n (where p_L is the complexity function of

[1] a local condition
[2] a global condition

the language). It therefore defines a subshift (X,T) whose language is L. It remains therefore to prove that (X,T) is minimal.

By hypothesis, $(\sigma_n : A_{n+1}^* \to A_n^*)_{n\in\mathbb{N}}$ is almost primitive and admits a contraction that contains only proper morphisms. This implies that for all sequences of letters $(a_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} A_n$, the sequence

$$\big(\sigma_0\sigma_1\cdots\sigma_n(a_{n+1}^\infty)\big)_{n\in\mathbb{N}}$$

converges to the same limit $\mathbf{w}$. By construction of the morphisms, we have $L = L(\mathbf{w})$ so X is generated by $\mathbf{w}$. By Proposition 2.1.21 (page 49), $\mathbf{w}$ is uniformly recurrent so (X,T) is minimal. □

Our aim is now to describe exactly the set of all valid paths in $\mathcal{G}$. The idea is to modify the graph of graphs $\mathcal{G}$ in such a way that the "local condition" to be a valid path (the first point of Proposition 5.1.5) is treated by the graph. In other words, we would like to modify $\mathcal{G}$ in such a way that all finite paths are valid. In that case, we will only have to take care at the global condition, which is rather easy to check. But, we actually will see that modifying the graph $\mathcal{G}$ as wanted will not be possible. There will still remain some vertices v such that for some finite paths arriving in v, some edge e starting from v make the path pe not valid. However, we will manage to describe the local condition for these vertices so this will still provide an $\mathcal{S}$-adic characterization.

As in the proof of Theorem 4.6.1, we will split the proof of our characterization, Theorem 5.8.1 into several parts. The graph of graphs $\mathcal{G}$ contains 4 strongly connected components:

$$C_1 = \{2\},\ C_2 = \{3\},\ C_3 = \{4\},\ C_4 = \{1,5,6,7,8,9,10\}.$$

First, we will separately study the valid paths in each strongly connected component C_i of $\mathcal{G}$ and modify them as explained above. We will end the proof by linking them together.

Remark 5.1.6. As mentioned earlier, a path p in $\mathcal{G}$ always starts from the vertex 1 or from the vertex 2 (depending on the size of the alphabet: 2 or 3). When studying the validity of a path in the component C_2, C_3 or C_4, we only study the validity of its suffix that always stays in that component (even for C_4 since a path ultimately staying in the component C_4 might start in the vertex 2). The validity of the prefixes that correspond to edges that are not in the final component will be treated at the end of the proof, while linking the different components.

5.2 Valid paths in C_1

This component corresponds to the class of Arnoux-Rauzy subshifts and has already been studied in [AR91]. The morphisms γ_n that code an evolution in that component are

$$\begin{array}{rcl} [0,10,20] & = & D_{1,0} \circ D_{2,0} \\ [01,1,21] & = & D_{0,1} \circ D_{2,1} \\ [02,12,2] & = & D_{0,2} \circ D_{1,2} \end{array}$$

and their respective left conjugates are

$$\begin{array}{rcl} [0,01,02] & = & G_{1,0} \circ G_{2,0} \\ [10,1,12] & = & G_{0,1} \circ G_{2,1} \\ [20,21,2] & = & G_{0,2} \circ G_{1,2} \end{array} .$$

In [AR91], the authors only consider the morphisms $[0,10,20]$, $[01,1,21]$ and $[02,12,2]$. They proved (see Theorem 2.2.19 on Page 58) that a sequence of such morphisms is valid if and only if every morphism occurs infinitely often in the sequence (otherwise the subshift obtained would not be minimal). Adapting this result to our case (with the left conjugates), we have the following.

Proposition 5.2.1. *Let* $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ *be a sequence of morphisms in* $\mathcal{S}$. *Then* $\mathbf{s}$ *is a valid directive word corresponding to a subshift whose Rauzy graphs are all of type 2 if and only if there is a sequence of morphisms* $(\gamma_n)_{n\in\mathbb{N}}$ *in* $\{[0,10,20],[01,1,21],[02,12,2]\}$ *and a contraction* $(\Gamma_n)_{n\in\mathbb{N}}$ *of* $(\sigma_n)_{n\in\mathbb{N}}$ *such that*

1. *the three morphisms* $[0,10,20]$, $[01,1,21]$ *and* $[02,12,2]$ *occur infinitely often in* $(\gamma_n)_{n\in\mathbb{N}}$*;*

2. *for all non-negative integers* n, Γ_n *is either* γ_n *or* $\gamma_n^{(L)}$ *and there are infinitely many right proper morphisms and infinitely many left proper morphisms in* $(\Gamma_n)_{n\in\mathbb{N}}$.

Proof. Indeed, the validity of all prefixes of p can easily be checked and can also be found in [AR91]. Then, the first condition is necessary and sufficient for $(\sigma_n)_{n\in\mathbb{N}}$ to be almost primitive and the second condition is necessary and sufficient to obtain a contraction that contains only proper morphisms. $\square$

5.3 Valid paths in C_2

This component contains only the vertex 3 of $\mathcal{G}$ and the morphisms that code an evolution in this component are

$[0,10,20] = D_{1,0} \circ D_{2,0}$	$[01,1,21] = D_{2,1} \circ D_{0,1}$	$[02,12,2] = D_{0,2} \circ D_{1,2}$
$[0,10,2] = D_{1,0}$	$[01,1,2] = D_{0,1}$	$[02,1,2] = D_{0,2}$
$[0,1,20] = D_{2,0}$	$[0,1,21] = D_{2,1}$	$[0,12,2] = D_{1,2}$

Proposition 5.3.1. *Let* $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ *be a sequence of morphisms in* $\mathcal{S}$. *Then there is an integer* $N \geq 0$ *such that* $(\sigma_n)_{n\geq N}$ *is a suffix of a valid directive word corresponding to a minimal subshift whose Rauzy graphs are ultimately of type 3 if and only if there is a non-negative integers* $N' \leq N$, *a contraction* $(\Gamma_n)_{n\in\mathbb{N}}$ *of* $(\sigma_n)_{n\in\mathbb{N}}$ *and a sequence of morphisms* $(\gamma_n)_{n\geq N'}$ *such that*

1. $(\gamma_n)_{n\geq N'}$ *labels an infinite path in the graph represented in Figure 5.3 with*
   (a) *for all* $x \in \{0,1,2\}$, *the loop on* V_x *is labelled by morphisms in*
   $$F_x = \{D_{y,x}D_{z,x},\ D_{x,y}D_{z,y} \mid \{x,y,z\} = \{0,1,2\}\}\,;$$
   (b) *for all* $x, y \in \{0,1,2\}$, $x \neq y$, *the edge from* V_x *to* V_y *is labelled by morphisms in*
   $$F_{x\to y} = \{D_{x,z},\ D_{x,y}D_{z,x} \mid z \notin \{x,y\}\}\,;$$
2. *for all integers* $n \geq N'$, Γ_n *is either* γ_n *or* $\gamma_n^{(L)}$ *and there are infinitely many right proper morphisms and infinitely many left proper morphisms in* $(\Gamma_n)_{n\geq N'}$;
3. *for all* $x \in \{0,1,2\}$, *there are infinitely many integers* $n \geq N'$ *such that* $D_{y,x}$ *is a factor of* γ_n *for some* $y \in \{0,1,2\}$.

Proof. First let us define the integers N and N' of the result. Our aim is to study the validity of the suffix of $\mathbf{s}$ that corresponds to evolutions of Rauzy graphs of type 3. Consequently, if $(\gamma_n)_{n\in\mathbb{N}}$ is the sequence of morphisms coding the evolutions of Rauzy graphs then we let N' denotes the smallest integer such that $\gamma_{N'}$ codes the evolution of a Rauzy graph of type 3. The integer N is therefore the integer such that $(\Gamma_n)_{n\geq N'}$ is a contraction of $(\sigma_n)_{n\geq N}$, where $(\Gamma_n)_{n\geq N'}$ is obtained from $(\gamma_n)_{n\in\mathbb{N}}$ by contraction and by replacing some right proper morphisms by their left conjugate.

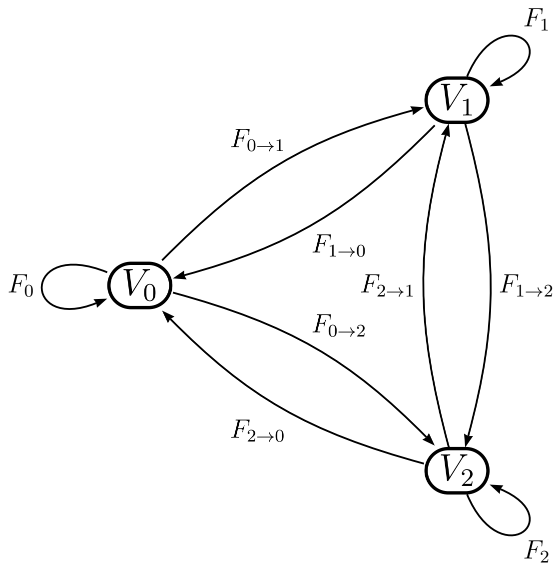


Figure 5.3: Graph corresponding to component C_2 in G.

Now we have to characterize sequences of morphisms that satisfy conditions 1 and 2 of Proposition 5.1.5 (only for the suffix $(\sigma_n)_{n\geq N}$).

Let us start with condition 1 (i.e., the local one). The morphisms that code an evolution from a graph of type 3 to a graph of type 3 (and their decomposition into $\mathcal{S}^*$) are listed above. However, Lemma 4.6.3 shows that they cannot be composed in every way. When computing the morphisms coding the different evolutions (see Figure 4.12 on page 128), we see that what is important is which letter corresponds to the top loop in Figure 4.12(a). Consequently, we can "split" the vertex 3 in G into 3 vertices V_0, V_1 and V_2, each V_x corresponding to the fact that the circuit $\vartheta_{i_n}(x)$ only goes through non-left special vertices (i.e., corresponds to the top loop in Figure 4.12(a)) and we put some edges between these vertices if the corresponding evolution is available. Then we label the graph as follows: for all $x, y \in \{0,1,2\}$ such that $x \neq y$, we let F_x denote the set of morphisms labelling the loop on V_x and we let $F_{x\to y}$ denote the set of morphisms labelling the edge from V_x to V_y. Of course, F_x and $F_{x\to y}$ contain the morphism corresponding to the evolution, i.e., F_x contains the morphism

$$D_{y,x}D_{z,x} = \begin{cases} x \mapsto x \\ y \mapsto yx \\ z \mapsto zx \end{cases}$$

and $F_{x\to y}$ contains the morphism

$$D_{x,z} = \begin{cases} x \mapsto xz \\ y \mapsto y \\ z \mapsto z \end{cases} .$$

Defining F_x and $F_{x\to y}$ this way ensures that the local condition is satisfied.

Now let us consider the second condition of Proposition 5.1.5. We have to describe all paths in Figure 5.3 whose label $(\sigma_n)_{n\geq N}$ is almost primitive and admits a contraction that contains only proper morphisms.

Let us start with the proper property. Up to now, there are only non-right proper morphisms in $F_{x\to y}$, $x, y \in \{0, 1, 2\}$, $x \neq y$ and Lemma 4.6.3 does not force a valid path in Figure 5.3 to be labelled by infinitely many right proper morphisms (we can for instance consider a path that does not go through any loop in 5.3). Consequently, for a valid path labelled by $(\sigma_n)_{n\geq N}$, even if we are ensured (by Proposition 5.1.5) to get a contraction of $(\sigma_n)_{n\geq N}$ which contains only right proper morphisms, this contraction might not label a path in Figure 5.3. Our aim is therefore to modify sets F_x and $F_{x\to y}$ in such a way that any valid path labelled by $(\sigma_n)_{n\geq N}$ admits a contraction with infinitely many right proper morphisms that labels a path in Figure 5.3. Then it will be enough to consider left conjugates of infinitely many right proper morphisms (but leaving infinitely many right proper morphisms unchanged).

As all non-right proper morphisms belong to some set $F_{x\to y}$, this can easily be done as follows: for all $x, y, z \in \{0, 1, 2\}$, $x \neq y$, $y \neq z$, one can check that the morphism $D_{x,z}D_{y,x} \in F_{x\to y}F_{y\to z}$ is right proper and labels a finite path from V_x to V_z. Consequently, for all x and all y, z such that $\{x, y, z\} = \{0, 1, 2\}$ we can add in F_x the morphism $D_{x,z}D_{y,z}$ and we add in $F_{x\to z}$ the morphism $D_{x,z}D_{y,x}$. Now, if a contraction $(\gamma'_n)_{n\geq N''}$ of $(\sigma_n)_{n\geq N}$ labels a valid path in Figure 5.3, if γ'_n and γ'_{n+1} are not right proper and if x, y, z belong to $\{0, 1, 2\}$ are such that $y \notin \{x, z\}$ and such that $\gamma'_n\gamma'_{n+1}$ labels a finite subpath of length 2 starting in V_x, going through V_y and ending in V_z, then there is contraction $(\gamma_n)_{n\geq N'}$ of $(\gamma'_n)_{n\geq N''}$ that labels a path in Figure 5.3 in which $\gamma'_n\gamma'_{n+1}$ is replaced by some right proper morphism γ_m labelling the edge from V_x to V_z.

Now let us describe all labelled paths in Figure 5.3 with almost primitive label. Morphisms in sets F_x and $F_{x\to y}$, $x, y \in \{0, 1, 2\}$, are composed of morphisms $D_{u,v}$ for some $u, v \in \{0, 1, 2\}$. Let us prove that the label $(\gamma_n)_{n\geq N}$ of a path in Figure 5.3 is almost primitive if and only if for all $x \in \{0, 1, 2\}$, there are infinitely many integers such that $D_{y,x}$ is a factor of γ_n for some $y \in \{0, 1, 2\}$, $y \neq x$. The condition is trivially necessary since if for all y, $D_{y,x}$ is not a factor of γ_n for n not smaller than some integer $m \geq N'$, then

x does not belong to $\gamma_m \cdots \gamma_m + k(z)$ for all $z \neq x$ and all integers $k \geq 0$. It is also sufficient. Indeed, from the way the morphisms can be composed (governed by Figure 5.3), the condition implies that $(\gamma_n)_{n \geq N'}$ is everywhere growing: for all x, y, z such that $\{x, y, z\} = \{0, 1, 2\}$, one cannot make morphisms in $\{D_{y,x}, D_{z,x}\}$ and morphisms in $\{D_{z,y}, D_{y,z}\}$ infinitely often occur as factors of some γ_n without making $D_{x,y}$ or $D_{x,z}$ occurring infinitely often too. Therefore, if $(\gamma_n)_{n \geq N'}$ is not almost primitive, there are letters x and y such that x does not occur in $\gamma_r \cdots \gamma_s(y)$ for some integers r and s, $s > r$. Consequently, $D_{y,x}$ cannot occur as a factor of γ_m for all $m \geq r$. Since $(\gamma_n)_{n \geq N'}$ is everywhere growing, this implies that $D_{y,z}$ infinitely often occurs as factor of morphisms γ_n. Thus, if x does not occur in $\gamma_r \cdots \gamma_s(y)$, the morphism $D_{z,x}$ cannot occur as factor of morphisms γ_n. This contradicts the fact that either $D_{y,x}$ or $D_{z,x}$ occurs infinitely often as factor of morphisms γ_n. □

5.4 Preliminary lemmas for C_3 and C_4

In both types of graphs of component C_1 and C_2, there is only one right special vertex. This makes the computation of valid paths easier to compute than when there are two right special factors. Indeed, if R_1 and R_2 are two bispecial factors in a Rauzy graph G_{i_n}, the circuits starting from R_1 impose some restrictions on the behaviour of R_2, i.e., on the way it will make the graph evolves when it will become bispecial (see Example 5.1.4 where the explosion of the bispecial vertex B' is governed by $\vartheta_{i_n}(1)$ and $\vartheta_{i_n}(2)$). Such a thing cannot happen for graphs of type 2 and 3, i.e., the local condition of Proposition 5.1.5 can be easily expressed. In this section, we introduce some notations and we give some lemmas that will be helpful to study valid paths in components C_3 and C_4.

First, let us briefly explain what we will mean when talking about *explosion* of a bispecial factor. Roughly speaking, "explosion" describes the behaviour of a bispecial vertex when the Rauzy graph evolves. These vertices are of a particular interest since those are the only ones that can change the shape of a graph (hence they are the only ones that determine the morphisms γ_n since they depend on the shape of the graphs). See Section 1.5.2 for more details on the behaviours of vertices when Rauzy graphs evolve.

The next lemma gives a method to build a contraction $(\eta_n)_{n\in\mathbb{N}}$ of $(\sigma_n)_{n\in\mathbb{N}}$ which is a little bit different from $(\gamma_n)_{n\in\mathbb{N}}$ and that will help us to describe the valid paths in C_3 and C_4.

Lemma 5.4.1. *Let (X, T) be a minimal subshift with first difference of complexity satisfying $1 \leq p(n+1) - p(n) \leq 2$ for all n. There is a non-decreasing sequence $(j_n)_{n\in\mathbb{N}}$ of integers and a contraction $(\eta_n)_{n\in\mathbb{N}}$ of $(\sigma_n)_{n\in\mathbb{N}}$ such that* *or all n, η_n codes the explosion of a unique bispecial factor of length j_n in $G_{j_n}(X)$.*

Proof. First it is obvious that if a Rauzy graph G_{i_n} contains two bispecial vertices, making them explode at the same time or separately produces the same graph G_{i_n+1} (hence $G_{i_{n+1}}$). Consequently, since γ_n describes how a graph evolves to the next one, we can decompose it into two morphisms $\gamma_n^{(1)}$ and $\gamma_n^{(2)}$ such that $\gamma_n = \gamma_n^{(1)}\gamma_n^{(2)}$, each one describing the explosion of one of the two bispecial vertices. Then it suffices to show that we can decompose $\gamma_n^{(1)}$ and $\gamma_n^{(2)}$ into morphisms of $\mathcal{S}$. This is actually obvious. Indeed, if there are two bispecial vertices, the graph can only be of type 6 or of type 8. Then, making only one bispecial vertex explode corresponds to considering that it is actually respectively of type 5 or 7 and we know that these morphisms belong to $\mathcal{S}^*$. However, we have to make it carefully: if B_1 and B_2 are the two bispecial vertices in G_{i_n} and if, for instance, B_1 is strong, we have to make B_2 explode before B_1 otherwise the explosion of B_1 would yield a graph with 3 right special vertices and this does not correspond to any type of graphs as considered in Figure 4.5. In other words, $\gamma_n^{(1)}$ has to correspond to the explosion of B_2 and $\gamma_n^{(2)}$ has to correspond to the explosion of B_1.

To conclude the proof, it suffices to build the sequences $(j_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$. From what precedes, the first one is simply the sequence $(i_n)_{n\in\mathbb{N}}$ but such that when G_{i_n} contains two bispecial factors, then i_n occurs twice in $(j_n)_{n\in\mathbb{N}}$. The second one is the sequence $(\gamma_n)_{n\in\mathbb{N}}$ (still with some morphisms $\gamma_n^{(L)}$ instead of γ_n) but such that when G_{i_n} contains two bispecial vertices, we split γ_n into $\gamma_n^{(1)}$ and $\gamma_n^{(2)}$. □

Example 5.4.2. Let us consider a path p in $\mathcal{G}$ that ultimately stays in the set of vertices $\{7, 8\}$. When the Rauzy graph G_{i_n} is of type 7, there is a unique bispecial factor so the morphism γ_n satisfies the conditions of the Lemma, i.e., it corresponds to a morphism in $(\eta_m)_{m\in\mathbb{N}}$. On the other hand, when G_{i_n} is of type 8, its two possible evolutions are represented at Figures 4.11(a) and 4.11(b) on page 122. Suppose that the starting vertex v_{i_n} corresponds to the vertex B_1 in Figure 4.10 (page 121) and suppose that G_{i_n} evolves as in Figure 4.11(a) with v_{i_n+1} equals to αB_1; the others cases are analogous. We have $\gamma_n = [0, 1^k0, (1^{k-1}0)]$. To decompose it as announced in Lemma 5.4.1, it suffices to consider that G_{i_n} is of type 7 with B_2 as bispecial vertex. We make this bispecial vertex explode like it is supposed to do (i.e. like a weak bispecial factor). This makes the graph evolving to a graph G'_{i_n} of type 1 (whose bispecial vertex is B_1) and we consider that the morphism coding this evolution is $\eta_m = [0, 1]$. Now it suffices to make this new graph G'_{i_n} evolve to a graph of type 7 or 8 with the morphism $\eta_{m+1} = [0, 1^k0, (1^{k-1}0)]$.

We then have $\gamma_n = \eta_m \circ \eta_{m+1}$ and these new morphisms satisfy condition 2 in Lemma 5.4.1. They can easily be decomposed by morphisms in $\mathcal{S}$ since $\eta_m = id$ and $\eta_{m+1} = \gamma_n$.

Definition 5.4.3. Let $(j_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ be as in Lemma 5.4.1. For all n we let B_{j_n} denotes the bispecial factor of length j_n whose explosion is coded by η_n.

The following result is a direct consequence of Definition 4.1.4.

Lemma 5.4.4. *Let $(j_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ be as in Lemma 5.4.1. The morphism η_n is a letter-to-letter morphism if and only if $B_{j_n} \neq v_{j_n}$ (where $(v_n)_{n\in\mathbb{N}}$ is the sequence of starting vertices of the circuits).*

Remark 5.4.5. Observe that, as illustrated by Example 5.1.4, when $B_{j_n} \neq v_{j_n}$, the evolution of G_{j_n} is influenced by the last morphism η_k, $k < n$, such that $B_{j_k} = v_{j_k}$. Indeed, as we have seen in Section 4.4, the circuits starting from v_{j_k} may depend on some parameters (the number of loops they contain for instance) and there exist some restrictions to these parameters[3]. Actually, considering a particular morphism η_k corresponds to determining these parameters. Since some of these circuits go through the other right special vertex in G_{j_k} (if it exists), these parameters influence the behaviour of this right special vertex.

On the other hand, when $B_{j_n} = v_{j_n}$, there are no restrictions on the possibilities for η_n since we do not have any information on the circuits starting from the right special vertex that is not v_{j_n}. Also, for graphs in components C_3 and C_4 there are no restrictions on the labels of the circuits like there are for Rauzy graphs of type[4] 2 or 3. Consequently, all possible morphisms are allowed. However, some of these morphisms are only *locally* allowed, i.e., even if a morphism is allowed, some "infinite choices" containing it may be forbidden. Indeed, Example 5.1.3 shows that a graph of type 9 can evolve to a graph of type 9 (so there is an allowed evolution) but it cannot ultimately keep being a graph of type 9 otherwise $(\gamma_n)_{n\in\mathbb{N}}$ would not be everywhere growing. To be clearer, the circuits starting in the right special vertex that is not v_{j_n} also depend on parameters and, as for the circuits starting from v_{j_n}, there are some restrictions on them. Those parameters are *partially* determined by the morphism η_n. For instance let us consider the evolution of a graph of type 9 as in Figure 5.2 (Page 139) such that v_{j_n} corresponds to the vertex B in Figure 5.2(a). This evolution implies that all circuits starting from the vertex R in Figure 5.2(a) go into the loop $B \to B$ at least once.

[3]For instance, when there are two parameters k and ℓ, one of them can sometimes not be greater than the other one.

[4]For those graphs, the right label of $\vartheta_n(x)$ starts with x for all $x \in \{0, 1, 2\}$.

5.5 Valid paths in C_3

This component only contains the vertex 4 in $\mathcal{G}$ and this type of graphs contains two right special vertices. Moreover, these two right special vertices cannot be bispecial at the same time since there is only one left special factor of each length. Consequently, we have $\eta_n = \gamma_n$ for all n and, as explained in Remark 5.4.5, we can *locally* choose any morphism we want when $v_{j_n} = B_{j_n}$ and we have to be careful when $v_{j_n} \neq B_{j_n}$. In other words, when v_{j_n} is the vertex R in Figure 5.4, the choice of the morphism γ_n is restrained by the latest morphism γ_m, $m < n$, such that v_{j_m} is the vertex B. From Section 4.5 this morphism γ_m is either

$$[0x^k y, x^\ell y, (0x^{k-1}y)] \quad \text{or} \quad [x^k y, 0x^\ell y, (x^{k-1}y)]$$

with $\{x, y\} = \{1, 2\}$, $k \geq 1$ and $k \geq \ell \geq 0$.

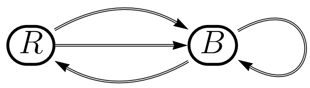


Figure 5.4: Rauzy graph of type 4.

Lemma 5.5.1 below expresses the consequences of this morphism γ_m.

Lemma 5.5.1. *Let $n \in \mathbb{N}$ and G_{i_n} be a Rauzy graph of type 4.*
Suppose that $v_{i_n} = R$ and that the two i_n-circuits $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$ pass respectively through the loop k and ℓ times with $k \geq 1$ and $k \geq \ell \geq 0$.
If the circuit $\vartheta_{i_n}(2)$ exists:

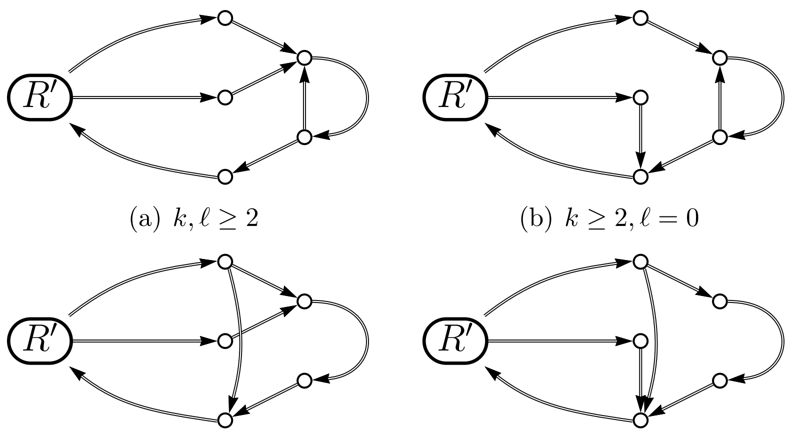
- *i if $\ell = k$, the Rauzy graph will evolve to a graph G_{i_m}, $m > n$ of type 10 such that v_{i_m} corresponds to the vertex B in Figure 4.5(j) (page 109) and the evolution from G_{i_n} to G_{i_m} is coded by the morphism $[1, 0, 2]$;*
- *ii if $\ell = k - 1$, the Rauzy graph will evolve to a graph G_{i_m}, $m > n$ of type 4 such that v_{i_m} corresponds to the vertex B in Figure 5.4 just above and the evolution from G_{i_n} to G_{i_m} is coded by a morphism in $\{[1, 0, 2], [1, 2, 0]\}$;*
- *iii if $\ell < k - 1$, the Rauzy graph will evolve to a graph G_{i_m}, $m > n$ of type 7 or 8 such that v_{i_m} corresponds to one of the vertices R and B in Figure 4.5(g) and to one of the vertices B_1 and B_2 in Figure 4.5(h). The evolution from G_{i_n} to G_{i_m} is coded by the morphism $[1, 0, 2]$ and we refer to Lemma 5.6.4 with $k := k - \ell - 1$ to know what will next happen.*

If the circuit $\vartheta_{i_n}(2)$ *does not exist:*

i *if* $\ell = k$ *or* $\ell = k-1$*, the graph will evolve to a graph* G_{i_m}*,* $m > n$ *of type 1 such that* v_{i_m} *corresponds to the vertex* B *in Figure 4.5(a) and the evolution from* G_{i_n} *to* G_{i_m} *is coded by in morphism in* $\{[0,1],[1,0]\}$*;*

ii *if* $\ell < k-1$*, the graph will evolve to a graph* G_{i_m}*,* $m > n$ *of type 7 or 8 such that* v_{i_m} *corresponds to one of the vertices* R *and* B *in Figure 4.5(g) and to one of the vertices* B_1 *and* B_2 *in Figure 4.5(h). The evolution from* G_{i_n} *to* G_{i_m} *is coded by the morphism* $[1,0]$ *and we refer to Lemma 5.6.4 with* $k := k - \ell - 1$ *to know what happens next.*

Proof. It suffices to see how the graph evolves. Indeed, when the vertex B explodes, we have eight possibilities represented at Figures 5.5 and 5.6. The main thing to notice is that if both circuits[5] $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$ can go through the loop $B \to B$ respectively k and ℓ times with k and ℓ greater than 1 (observe that in this case, the circuit $\vartheta_{i_n}(2)$ goes into that loop $k-1$ times), the graph will evolve as in Figure 5.5(a) and the new circuits $\vartheta_{i_n+1}(0)$ $\vartheta_{i_n+1}(1)$ will go into the loop respectively $k-1$ and $\ell-1$ times (so $k-2$ times for $\vartheta_{i_n+1}(2)$). The computation of the morphisms is left to the reader. ☐



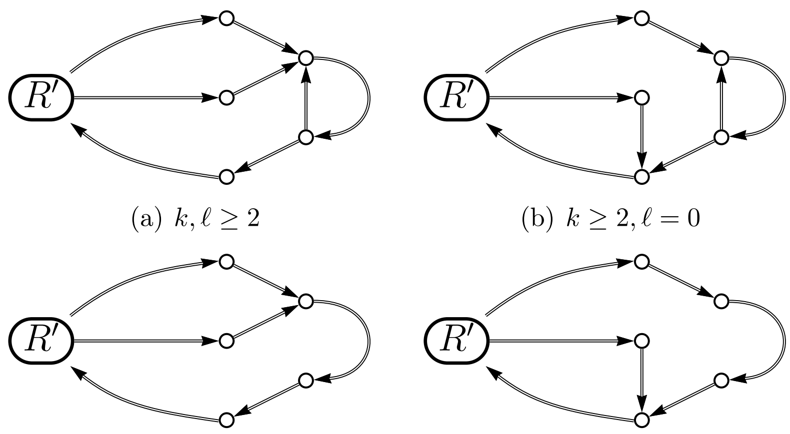
(a) $k, \ell \geq 2$ (b) $k \geq 2, \ell = 0$

(c) $k = \ell = 1$ (d) $k = 1, \ell = 0$

Figure 5.5: Evolutions of a graph of type 4 with 3 circuits starting from R.

Now we can determine which are the labels of the valid paths in the component C_3. Moreover, in $\mathcal{G}$ we can rename the vertex 4 by $4B$, meaning that we always have $v_{i_n} = B$.

[5]The reader is invited to check the definition of ϑ_{i_n} for such graphs on page 117.



(a) $k, \ell \geq 2$ (b) $k \geq 2, \ell = 0$

(c) $k = \ell = 1$ (d) $k = 1, \ell = 0$

Figure 5.6: Evolutions of a graph of type 4 with 2 circuits starting from R.

Proposition 5.5.2. *Let $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ be a sequence of morphisms in $\mathcal{S}$. Then there is an integer $N \geq 0$ such that $(\sigma_n)_{n\geq N}$ is a suffix of a valid directive word corresponding to a minimal subshift whose Rauzy graphs are ultimately of type 4 if and only if there is a non-negative integers $N' \leq N$, a contraction $(\Gamma_n)_{n\in\mathbb{N}}$ of $(\sigma_n)_{n\in\mathbb{N}}$ and a sequence of morphisms $(\gamma_n)_{n\geq N'}$ such that*

1. *for all $n \geq N'$,*

$$\gamma_n \in \{[0, 10, 20], [0, 20, 10], [x^{k-1}y, 0x^k y, 0x^{k-1}y], [x^{k-1}y, 0x^{k-1}y, 0x^k y], [0x^{k-1}y, x^k y, x^{k-1}y], [0x^{k-1}y, x^{k-1}y, x^k y] \mid k \geq 1\big\}$$

   with $\{x, y\} = \{1, 2\}$;

2. *for all $r \geq N'$,*

$$(\gamma_n)_{n\geq r} \notin \{[0, 10, 20], [0, 20, 10]\}^\omega$$

   and

$$(\gamma_n)_{n\geq r} \notin \big\{[0x^{k-1}y, x^k y, x^{k-1}y], [0x^{k-1}y, x^{k-1}y, x^k y] \mid k \geq 1\big\}^\omega$$

3. *for all integers $n \geq N'$, Γ_n is either γ_n or $\gamma_n^{(L)}$ and there are infinitely many right proper morphisms and infinitely many left proper morphisms in $(\Gamma_n)_{n\geq N'}$;*

Proof. First let us define the integers N and N' of the result. Our aim is to study the validity of the suffix of $\mathbf{s}$ that corresponds to evolutions of Rauzy graphs of type 4. From Lemma 5.5.1 we can suppose that N' is the smallest integer such that γ'_N codes an evolution from a graph of type 4 such that the starting vertex is the vertex B in Figure 5.4. Indeed, if $v_{i_{N'}}$ is not B, there is a smallest integer $k < N'$ such that G_{i_k} is of type 4 and v_{i_k} is the vertex R in Figure 5.4. Thus, the morphism γ_{k-1} codes an evolution from a graph of type 2 to a graph of type 4 (check in Figure 4.8 on page 112) and then Lemma 5.5.1 determines the sequence of morphisms $\gamma_k\gamma_{k+1}\cdots$ until v_{i_n} is B.

Then, the integer N is the integer such that $(\Gamma_n)_{n\geq N'}$ is a contraction of $(\sigma_n)_{n\geq N}$, where $(\Gamma_n)_{n\geq N'}$ is obtained from $(\gamma_n)_{n\in\mathbb{N}}$ by contraction and by replacing some right proper morphisms by their left conjugate.

We have to characterize sequences of morphisms that satisfies conditions 1 and 2 of Proposition 5.1.5 (only for the suffix $(\sigma_n)_{n\geq N}$).

Let us start with condition 1. Given a graph G_{i_n} of type 4 with $v_{i_n} = B$, the morphism γ_n coding the evolution to a graph of type 4 and such that

- $v_{i_{n+1}} = B$ are $[0, 10, 20]$ and $[0, 20, 10]$;
- $v_{i_{n+1}} = R$ are $[0x^ky, x^\ell y, 0x^{k-1}y]$ and $[x^ky, 0x^\ell y, x^{k-1}y]$.

When $v_{i_{n+1}} = R$, Lemma 5.5.1 impose some conditions on k and ℓ to evolves to a graph G_{i_m} of type 4 with $v_{i_m} = B$. Indeed, the exponent k (resp. ℓ) corresponds to the number of times the circuit $\vartheta_{i_{n+1}}(0)$ (resp. $\vartheta_{i_{n+1}}(0)$) goes into the loop $B \to B$. Consequently, we must have $\ell = k - 1$ and then the evolution from $G_{i_{n+1}}$ to G_{i_m} is coded by a morphism in

$$\{[1, 0, 2],\ [1, 2, 0]\}.$$

By composing these morphisms with the previous ones, we obtain all morphisms coding evolutions of graphs of type 4 to graph of type 4 such that all vertices v_{i_n} correspond to vertex B in Figure 5.4 so compositions of these morphisms provides valid prefixes of $(\sigma_n)_{n\geq N}$.

Now let us consider condition 2. It is evident that the third condition of the result is necessary and sufficient to obtain proper morphisms. Then, $(\sigma_n)_{n\geq N}$ is almost primitive if and only if so is $(\gamma_n)_{n\geq N'}$ and this is equivalent to impose that for all $r \geq N$,

$$(\gamma_n)_{n\geq r} \notin \{[0, 10, 20], [0, 20, 10]\}^\omega$$

and

$$(\gamma_n)_{n\geq r} \notin \left\{[0x^{k-1}y, x^ky, x^{k-1}y], [0x^{k-1}y, x^{k-1}y, x^ky] \mid k \geq 1\right\}^\omega$$

with $\{x, y\} = \{1, 2\}$. ☐

5.6 Valid paths in C_4

This component of $\mathcal{G}$ contains the vertices 1, 5, 6, 7, 8, 9 and 10. As for component C_3, we need some lemmas to determine the consequences of some morphisms γ_n on the sequence $(\gamma_k)_{k\geq n+1}$. The difficulty in determining the valid paths in this component is in the fact that we have to take care of the length of some paths in the Rauzy graphs to know which morphism we can choose. Indeed, the morphisms that code the evolutions to Rauzy graphs of type 5 or 6 (and 7 or 8) are the same and the precise type depends on the lengths of the path p_1 and p_2 in Figure 5.7(a) (and of the lengths of the paths u_1, u_2, v_1 and v_2 in Figure 5.7(b). When the Rauzy graph G_{i_n} is of type 6 or 8 (i.e., when $|p_1| = |p_2|$ or when $|u_1| = |u_2|$), we know from Lemma 5.4.1 that we can decompose the morphism γ_n into two morphisms, each one corresponding to the explosion of one bispecial vertex. On the other hand, if for example $|u_1| >> |u_2|$ in a graph of type 7 and if we denote by $B_1(1), B_1(2), \dots$ (resp. $B_2(1), B_2(2), \dots$) the bispecial vertices (ordered by increasing length) in the Rauzy graphs or larger order that admit R_1 (resp. R_2) as a suffix, we will see that many vertices $B_1(i)$ will explode before that $B_2(1)$ explodes. Consequently we cannot choose any morphisms we want.

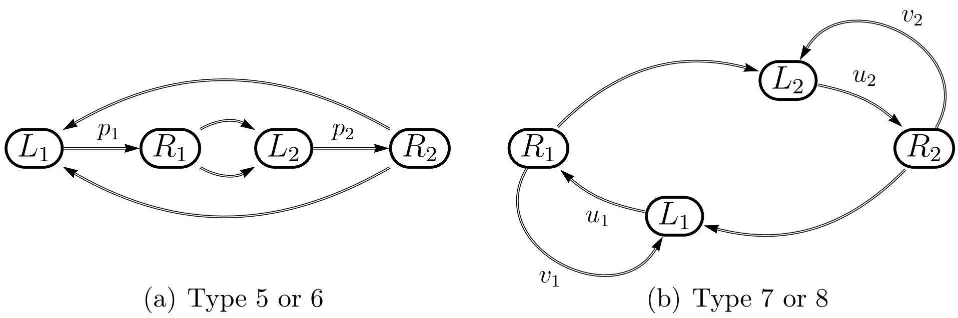


Figure 5.7: Rauzy graphs of type 5 or 6 and 7 or 8.

First, the following result will be helpful to characterize valid paths that goes infinitely often through the vertex 1 in the graph of graphs.

Fact 5.6.1. *We can suppose without loss of generality that the evolution of a Rauzy graph of type 1 to a Rauzy graph of type 1 is coded by* $[0, 10]$ *or by* $[01, 1]$.

Proof. On page 122, we see that the morphisms coding such an evolution are $[0, 10] = D_{1,0}$, $[10, 0] = D_{1,0}E_{0,1}$, $[01, 1] = D_{0,1}$ and $[1, 01] = D_{0,1}E_{0,1}$ and that the morphisms coding an evolution from a graph of type 1 to a graph of type 7 or 8 are $[0, 1^k0, 1^{k-1}0]$ and $[1, 0^k1, 0^{k-1}1] = E_{0,1}[0, 1^k0, 1^{k-1}0]$.

By induction, it is easily seen that for all integers $n \geq 0$, we have

$$E_{0,1}\left\{D_{0,1}, D_{1,0}\right\}^n E_{0,1} = \left\{D_{0,1}, D_{1,0}\right\}^n.$$

To conclude the proof of the result, we have to consider several possibilities.

1. If for all n, γ_n codes an evolution from a graph of type 1 to a graph of type 1 and if $(\gamma_n)_{n\in\mathbb{N}}$ contains infinitely many occurrences of $D_{1,0}E_{0,1}$ and/or of $D_{0,1}E_{0,1}$, then the result trivially holds.

2. If for all n, γ_n codes an evolution from a graph of type 1 to a graph of type 1 and if $(\gamma_n)_{n\in\mathbb{N}}$ contains a finite and even number of occurrences of $D_{1,0}E_{0,1}$ and/or of $D_{0,1}E_{0,1}$, then the result trivially holds too.

3. If for all n, γ_n codes an evolution from a graph of type 1 to a graph of type 1 and if $(\gamma_n)_{n\in\mathbb{N}}$ contains a finite and odd number of occurrences of $D_{1,0}E_{0,1}$ and/or of $D_{0,1}E_{0,1}$, then it suffices to insert in $(\gamma_n)_{n\in\mathbb{N}}$ infinitely many occurrences of the morphism $id = E_{0,1}^2$ and the result holds.

4. Finally, if $\gamma_r \cdots \gamma_s \in \{D_{1,0}, D_{1,0}E_{0,1}, D_{0,1}, D_{0,1}E_{0,1}\}^*$ codes a finite sequence of evolutions from graphs of type 1 to graphs of type 1 and if $\gamma_{s+1} \in \{[0, 1^k0, 1^{k-1}0], E_{0,1}[0, 1^k0, 1^{k-1}0]\}$ codes an evolution to a graph of type 7 or 8, then $\gamma_r \cdots \gamma_s\gamma_{s+1}$ can be replaced by $\gamma'_r \cdots \gamma'_s\gamma'_{s+1}$ with $\gamma'_r \cdots \gamma'_s \in \{D_{0,1}, D_{1,0}\}^*$ and $\gamma'_{s+1} \in \{[0, 1^k0, 1^{k-1}0], E_{0,1}[0, 1^k0, 1^{k-1}0]\}$, depending on the number of occurrences of $D_{1,0}E_{0,1}$ and of $D_{1,0}E_{0,1}$ in $\gamma_r \cdots \gamma_s$.

□

Next, Lemma 5.6.2 implies that we can merge the vertices 5 and 6 to one vertex denoted by 5/6 in $\mathcal{G}$ and that the outgoing edges of that vertex are the same as the outgoing edges of the vertex 6 in $\mathcal{G}$. However, we have to take care of the lengths of p_1 and p_2 in Figure 5.7(a) to know which morphism in the labels of the edges can be applied.

Lemma 5.6.2. *Let G_k be a Rauzy graph as in Figure 5.7(a) and let i_n be the smallest integer in $(i_n)_{n\in\mathbb{N}}$ such that $i_n \geq k$. We have*

$$\{\textit{Type of } G_{i_{n+1}} \mid G_{i_n} \textit{ is of type } 6\} = \{\textit{Type of } G_{i_{n+2}} \mid G_{i_n} \textit{ is of type } 5 \textit{ and } v_{i_n} \textit{ is not strong bispecial}\}$$

and

$$\{\gamma_n \mid G_{i_n} \textit{ is of type } 6\} = \{\gamma_n \circ \gamma_{n+1} \mid G_{i_n} \textit{ is of type } 5 \textit{ and } v_{i_n} \textit{ is not strong bispecial}\}.$$

Proof. It suffices to look at the graph of graphs (Figure 4.8 on page 112) and at the lists of morphisms in Section 4.5 on page 124. The only thing to observe is that when a graph G_{i_n} is of type 5 and if v_{i_n} corresponds to the vertex B in Figure 4.5(e) (page 109), then v_{i_n} cannot be a strong bispecial factors, otherwise there would be 3 right special vertices in G_{i_n+1} and this does not correspond to any considered type of graphs. □

Remark 5.6.3. In order to describe all valid paths in the component C_4, we sometimes have to know the precise type of a graph corresponding to the vertex 5/6. Indeed, when going to that vertex in the modified component (suppose the label of the edge is γ_n and that $v_{i_{n+1}}$ corresponds to the vertex R_1 in Figure 5.7(a)), we may want to leave it using the morphism $\gamma_{n+1} = [x, y^k x, (y^{k-1}x)]$ (see Section 4.5). However, the evolution corresponding to that morphism is such that the smallest bispecial factor that admits v_{i_n+1} as a suffix is strong (the other right special vertex is therefore suffix of a weak bispecial factor). Consequently, we can leave the vertex 5/6 with that morphism only if $v_{i_{n+1}}$ is not bispecial, i.e., the other right special vertex becomes bispecial before v_{i_n+1}. In other words, we must have $|p_1| \geq |p_2|$ in Figure 5.7(a).

Next lemma deals with the same kind of stuffs as in Lemma 5.6.2 but for Rauzy graphs of type 7 and 8. As for graphs of type 5 and 6, it allows us to merge the vertices 7 and 8 to one vertex denoted 7/8 in $\mathcal{G}$.

Lemma 5.6.4. *Let G_t be a Rauzy graph as in Figure 5.7(b) and let i_n be the smallest integer in $(i_m)_{m\in\mathbb{N}}$ such that $i_n \geq t$. Suppose that v_t is the vertex R_1 and that $\vartheta_t(1)$ goes k times through the loop v_2u_2. Let $\ell \in \mathbb{Z}$ such that*

$$|u_1|+(\ell-1)(|u_1|+|v_1|) < |u_2|+(k-1)(|u_2|+|v_2|) \leq |u_1|+\ell(|u_1|+|v_1|). \quad (5.1)$$

Then, the graph can evolve to a graph of type

i. 1 and the composition of morphisms coding this evolution is in

$$\left\{[0,10]^h\{[01,1],[1,01]\} \mid 0 \leq h < \max\{1,\ell\}\right\} \cup \left\{[0,10]^h[x,y] \mid \{x,y\} = \{0,1\}, h = \max\{0,\ell\}\right\}$$

ii. 5 or 6 as in Figure 5.7(a) and the composition of morphisms coding this evolution is in

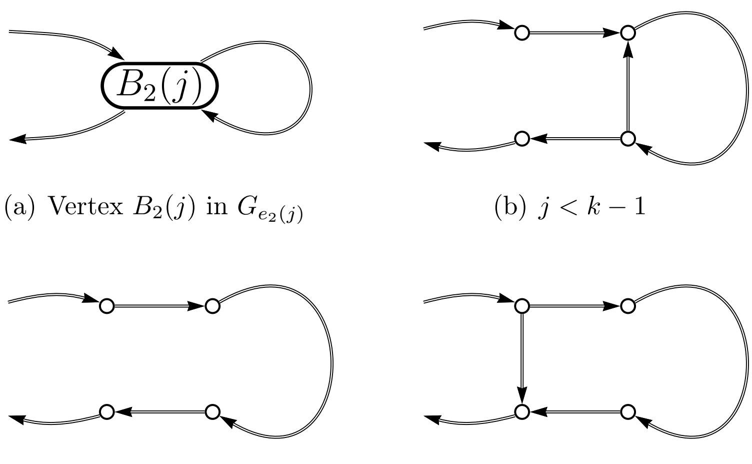
$$\left\{[0,10,20]^h\{[0x,y,(0y)],[x,0y,(y)]\} \mid \{x,y\} = \{1,2\}, 0 \leq h < \max\{1,\ell\}\right\};$$

iii. 9 with the starting vertex v_m, $m > i_n$, corresponding to the vertex B in Figure 4.5(i) and the composition of morphisms coding this evolution is in

$$\left\{[0, 10, 20]^h[0, x, y] \mid \{x, y\} = \{1, 2\}, h = \max\{0, \ell\}\right\}.$$

Proof. First let us study which are the bispecial vertices we have to deal with. It is a direct consequence of the definition of Rauzy graphs that for i and j in $\mathbb{N}$, the words $B_1(i) = \lambda\left(u_1(v_1u_1)^i\right)$ and $B_2(j) = \lambda\left(u_2(v_2u_2)^j\right)$ respectively admit L_1 and L_2 as prefixes and R_1 and R_2 as suffixes. For all i, j, we write $e_1(i) = |B_1(i)| = t + |u_1| + i(|u_1| + |v_1|)$ and $e_2(j) = |B_2(j)| = t + |u_2| + j(|u_2| + |v_2|)$. Inequality (5.1) therefore provides some information on the order the bispecial vertices $B_1(\ell - 1)$, $B_2(k - 1)$ and $B_1(\ell)$ (if they exist) explode.

By hypothesis, the path $u_2(v_2u_2)^k$ is allowed in G_t (since it is a subpath of a t-circuit). This implies that $B_2(j)$ is a bispecial factor in $L(X)$ for all $j \in \{0, 1, \ldots, k-1\}$ and this also gives us some information on the way they explode in their respective Rauzy graphs. Indeed, if there are 2 (resp. 3) t-circuits starting from R_1 in G_t, then in the Rauzy graph $G_{e_2(j)}$, the vertex $B_2(j)$ explodes as in Figure 5.8(b) if $j < k - 1$ and as in Figure 5.8(c) (resp. in Figure 5.8(d)) if $j = k - 1$.



(a) Vertex $B_2(j)$ in $G_{e_2(j)}$

(b) $j < k - 1$

(c) $j = k - 1$ and 2 circuits

(d) $j = k - 1$ and 3 circuits

Figure 5.8: Explosion of the vertex $B_2(j)$ in $G_{e_2(j)}$.

As $v_t = R_1$, we know from Lemma 5.4.4 and from Section 4.4 that the explosion of the vertices $B_2(j)$ are coded by the identity morphism for $j \in \{0, \ldots, k - 2\}$ and by a letter-to-letter morphism for $j = k - 1$.

Now let us study the behaviour of the vertex R_1. As we do not have any information about the circuits starting from R_2, there are several possibilities for the explosion of the vertices $B_1(i)$. First, we can observe that, if for some integer $i < \ell$, the word $B_1(i)$ belongs to $L(X)$, then for all $h < i$, the word $B_1(h)$ is a bispecial factor in $L(X)$ and it explodes like $B_2(j)$ in Figure 5.8(b). Each of these evolutions is coded by $[0, 10, 20]$ (or by $[0, 10]$ if there are only 2 circuits). On the other hand, if $B_1(i)$ is a bispecial factor of length $l < e_2(k-1)$ in $L(X)$ and if it explodes in G_l similarly to $B_2(j)$ in Figure 5.8(d), then G_l evolves to a graph of type 9 such that the starting vertex of the circuits corresponds to the vertex R in Figure 4.5(i). Consequently, the right special vertex in G_{l+1} that arises from $B_1(i)$ will not become bispecial until $B_2(k-1)$ has exploded. The evolution from G_l to G_{l+1} is coded by the morphism $[01, 1]$ or $[1, 01]$ if there are only 2 l-circuits and by one of the four following morphisms if there are three l-circuits: $[01, 1, 02]$, $[1, 01, 2]$, $[01, 2, (02)]$ and $[1, 02, (2)]$. Observe that $B_1(i)$ cannot explode similarly to $B_2(j)$ in Figure 5.8(c) as that would imply that the sequence of right special vertices $(v_n)_{n\in\mathbb{N}}$ is finite.

To conclude the proof, it suffices to list all the possibilities for the explosions of the vertices $B_1(i)$. By hypothesis, ℓ is an integer such that

$$|u_1| + (\ell-1)(|u_1| + |v_1|) < |u_2| + (k-1)(|u_2| + |v_2|) \leq |u_1| + \ell(|u_1| + |v_1|)$$

and we know that the vertices $B_1(i)$ and $B_2(j)$ respectively have length $e_1(i) = t + |u_1| + i(|u_1| + |v_1|)$ and $e_2(j) = t + |u_1| + j(|u_2| + |v_2|)$ for all non-negative integers i and j. Consequently, while $B_2(k-1)$ has not exploded yet, the vertex $B_1(i)$ (if it exists) has two possibilities: either it makes the graph evolving to a graph of type 7 or 8 with the morphism $[0, 10, (20)]$, or it makes it evolving to a graph of type 9 with one of the morphisms $[01, 1, (02)]$, $[1, 01, (2)]$, $[01, 2, (02)]$ and $[1, 02, (2)]$.

First suppose that the graph is not of type 7 or 8 anymore when the vertex $B_2(k-1)$ explodes. The only possibility is that $\ell \geq 1$ and that a vertex $B_1(i)$, $0 \leq i \leq \ell-1$, has exploded as in Figure 5.8(d), making the graph evolving to a graph of type 9 with one of the morphisms $[01, 1, (02)]$, $[1, 01, (2)]$, $[01, 2, (02)]$ and $[1, 02, (2)]$. Observe that each of the explosions of $B_1(0), B_1(1), \ldots, B_1(i-1)$ is coded by $[0, 10, 20]$. Then, the only bispecial vertices that occur in the next Rauzy graphs are vertices $B_2(j)$ for $j \in \{l', \ldots, k-1\}$ and l' the smallest integer such that $e_2(l') \geq e_1(i)$. They imply the following behaviours: for $j < k-1$, the explosions of $B_2(j)$ are coded by the identity morphism. For $j = k-1$, if there are three circuits starting from $B_1(i)$ and if its explosion is coded by the morphism $[01, 1, 02]$ or $[1, 01, 2]$ (resp. $[01, 2, (02)]$ or $[1, 02, (2)]$), then the explosion of $B_2(k)$ is coded by $[2, 1, 0]$ (resp. $[0, 1, 2]$). Consequently, the graph eventually evolves to a graph of type 5 or 6 and the composition of the morphisms is in

$$\left\{[0,10,20]^h \{[0x,y,(0y)],[x,0y,(y)]\} \mid \{x,y\}=\{1,2\}, 0 \leq h < \max\{1,\ell\}\right\}. \quad (5.2)$$

Still for $j = k-1$, if there are 2 circuits starting from $B_1(i)$, then the morphism coding its explosion is $[01,1]$ or $[1,01]$ and then the graph will evolve to a graph of type 1 with the morphism $[0,1]$ or $[1,0]$ (by exploding vertices $B_2(j)$). Consequently, the composition of morphisms coding this sequence of evolutions is in

$$\left\{[0,10]^h \{[01,1],[1,01]\} \mid 0 \leq i < \max\{1,\ell\}\right\}. \quad (5.3)$$

Now suppose that the graph is still of type 7 or 8 when the vertex $B_2(k-1)$ has exploded. If $\ell \geq 1$, this implies that the vertices $B_1(i)$ have exploded with the morphism $[0,10,(20)]$ for $i = 0,\dots,\ell-1$ (so we have $[0,10,(20)]^\ell$). Then, when the vertex $B_2(k-1)$ explodes, it makes the graph evolving to a graph G_{i_m} of type 1 or 9 depending on the number of circuits (2 or 3 respectively). If the vertex $B_1(\ell)$ has the same length, we can suppose from Lemma 5.4.1 that it does not explode at the same time so we can suppose that the graph does not evolve to a graph of type 7 or 8 (like it actually could with the morphism $[x,y^m x,(y^{m-1}x)]$). Consequently, we only have to consider the evolutions to graphs of type 1 or 9. They are respectively coded by $[0,1]$ or $[1,0]$ and by $[0,1,2]$ or $[0,2,1]$ and once this evolution is done, the next bispecial vertex is in $(v_n)_{n\in\mathbb{N}}$. □

The next lemma will allow us to delete the vertex 9 in $\mathcal{G}$. Indeed, we can see in Figure 4.8 (page 112) that the only types of graphs that can evolve to a graph of type 9 are types 9 and 7 or 8. The next lemma states that we can modify the outgoing edges of the vertex 7/8 such that the vertex 9 is isolated in $\mathcal{G}$.

Lemma 5.6.5. *In Lemma 5.6.4, we can delete the third case of all possible evolutions (the one to graphs of type 9) by replacing the set of morphisms coding the evolutions to graphs of type 5 or 6 (the second case) by*

$$\left\{[0,10,20]^h\{[0x,y,(0y)],[x,0y,(y)]\} \mid \{x,y\}=\{1,2\}, h\in\mathbb{N}\right\}.$$

We can also replace the morphisms coding the evolution to graphs of type 1 (the first case) by

$$\left\{[0,10]^h \{[01,1],[1,01]\} \mid h\in\mathbb{N}\right\} \cup \left\{[0,10]^h[x,y] \mid \{x,y\}=\{0,1\}, h \geq \max\{0,\ell\}\right\}$$

Proof. Indeed, in Lemma 5.6.4 the morphisms coding the evolution to a graph of type 9 are in

$$\left\{[0,10,20]^h[0,x,y] \mid \{x,y\}=\{1,2\}, h=\max\{0,\ell\}\right\}.$$

But, once the graph is of type 9 with $v_{i_n}=B$, it can only evolve either to a graph of type 9 with $v_{i_{n+1}}=B$, or to a graph of type 5 or 6 with a morphism in $\{[0x,y,(0y)],[x,0y,(y)] \mid \{x,y\}=\{1,2\}\}$. Consequently, the composition of evolution

$$7/8(\to 9)^j \to 5/6$$

is coded by a morphism in

$$\left\{[0,10,20]^h[0,x,y][0,x0,y0]^j\{[0x,y,(0y)],[x,0y,(y)]\} \mid \right. \\ \left. \{x,y\}=\{1,2\}, h=\max\{0,\ell\}\right\}.$$

Since j can be arbitrarily large, this set is equal to

$$\left\{[0,10,20]^h\{[0x,y,(0y)],[x,0y,(y)]\} \mid \{x,y\}=\{1,2\}, h\in\mathbb{N}\right\}.$$

For the second part (evolution to graphs of type 1), it suffices to observe that all considered morphisms also code evolutions from a graph of type 1 to a graph of type 1. Consequently, if h is chosen greater than $\max\{0,\ell\}$, the morphism $[0,10]^{h-\max\{0,\ell\}}$ is simply coding $h-\max\{0,\ell\}$ evolutions from 1 to 1. □

The last type of graph that has not been treated yet is the type 10. The next lemma does it.

Lemma 5.6.6. *Let G_{i_n} be a Rauzy graph of type 10. Suppose that v_{i_n} corresponds to the vertex R in Figure 4.5(j) and that the two i_n-circuits $\vartheta_{i_n}(0)$ and $\vartheta_{i_n}(1)$ respectively go through the loop k and ℓ times with $k,\ell\geq 0$ and $k+\ell\geq 1$.*
If the circuit $\vartheta_{i_n}(2)$ exists and starts like $\vartheta_{i_n}(0)$ does (recall that $\ell\leq k$ in this case), then

i. if $\ell=k$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m>n$, of type 10 such that v_{i_m} corresponds to the vertex B in Figure 4.5(j). This evolution is coded by the morphism $[1,0,2]$;

ii. if $\ell<k$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m>n$, of type 7 or 8 such that the i_m-circuit $\vartheta_{i_n}(1)$ starting from v_{i_m} goes through the loop $k'=k-\ell$ times. This evolution is also coded by the morphism $[1,0,2]$.

If the circuit $\vartheta_{i_n}(2)$ exists and starts like $\vartheta_{i_n}(1)$ do (recall that $k \leq \ell - 1$ in this case), then

i. if $k = \ell-1$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m > n$, of type 10 such that v_{i_m} corresponds to the vertex B in Figure 4.5(j). This evolution is coded by the morphism $[0, 1, 2]$*;*

ii. if $k < \ell - 1$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m > n$, of type 7 or 8 such that the i_m-circuit $\vartheta_{i_n}(1)$ starting from v_{i_m} goes through the loop $k' = \ell - k - 1$ times. This evolution is again coded by the morphism $[0, 1, 2]$.

If the circuit $\vartheta_{i_n}(2)$ does not exist, then

i. if $\ell \in \{k, k+1\}$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m > n$, of type 1. This evolution is coded by a morphism in $\{[0, 1], [1, 0]\}$*;*

ii. if $\ell < k$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m > n$, of type 7 or 8 such that the i_m-circuit $\vartheta_{i_n}(1)$ starting from v_{i_m} goes through the loop $k' = k - \ell$ times. This evolution is coded by the morphism $[1, 0]$.

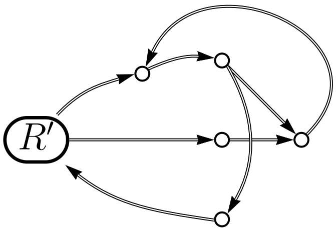
iii. if $\ell > k + 1$, G_{i_n} will evolve to a Rauzy graph G_{i_m}, $m > n$, of type 7 or 8 such that the i_m-circuit $\vartheta_{i_n}(1)$ starting from v_{i_m} goes through the loop $k' = \ell - k - 1$ times. This evolution is coded by the morphism $[0, 1]$.

Proof. Indeed, if the vertex B in Figure 4.5(j) explodes as in Figure 5.9(a), the new graph is still of type 10. This evolution is coded by the morphism $[1, 0, (2)]$. Moreover, if we denote by $k_{i_n}(0)$ (resp. $k_{i_n}(1)$, $k_{i_n}(2)$) the number of times that the i_n-circuit $\vartheta_{i_n}(0)$ (resp. $\vartheta_{i_n}(1)$, $\vartheta_{i_n}(2)$) goes through the loop, then we have $k_{i_n+1}(0) = k_{i_n}(1) - 1$ and $k_{i_n+1}(1) = k_{i_n}(0)$. We also have $k_{i_n+1}(2) = k_{i_n}(2)$ if the i_n-circuit $\vartheta_{i_n}(2)$ starts like $\vartheta_{i_n}(0)$ does and $k_{i_n+1}(2) = k_{i_n}(2) - 1$ if the i_n-circuit $\vartheta_{i_n}(2)$ starts like $\vartheta_{i_n}(1)$ does. Consequently, this evolution is repeated until either $k_{i_{n'}}(1) = 0$ or $k_{i_{n'}}(0) = 0$ and $k_{i_{n'}}(1) = 1$ for some $n' \geq n$. Then the graph $G_{i_{n'}}$ evolves to a Rauzy graph of type 1, 7, 8 or 9 depending on $k_{i_{n'}}(0)$, $k_{i_{n'}}(1)$ and $k_{i_{n'}}(2)$ (if the circuit $\vartheta_{i_n}(2)$ exists). The computation of the morphism coding this last evolution is left to the reader. □

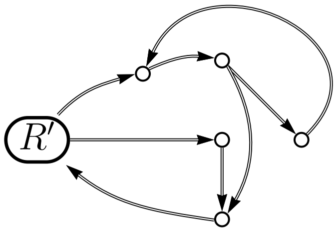
Modification of Component C_4

Now we can modify the component C_4 of $\mathcal{G}$.

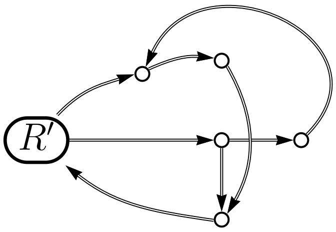
First let us modify the vertices. Lemmas 5.6.2 and 5.6.4 respectively allow to merge the vertices 5 and 6 to one vertex 5/6 and the vertices 7 and 8 to one



(a) either $\vartheta_{i_n}(2)$ starts like $\vartheta_{i_n}(0)$ does and $k, \ell \geq 1$, or $\vartheta_{i_n}(2)$ starts like $\vartheta_{i_n}(1)$ does and $\ell \geq 2$

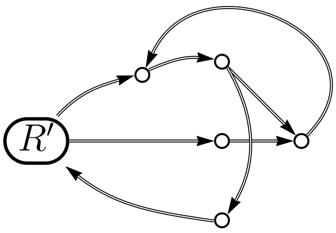


(b) $\vartheta_{i_n}(2)$ starts like $\vartheta_{i_n}(0)$ does and $k \geq 1$, $\ell = 0$

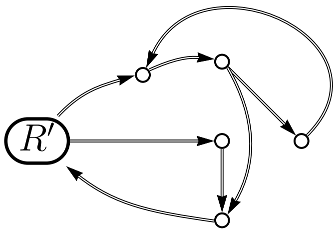


(c) $\vartheta_{i_n}(2)$ starts like $\vartheta_{i_n}(1)$ does and $k = 0$, $\ell = 1$

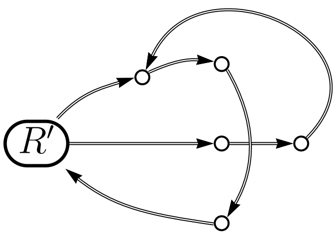
Figure 5.9: Evolutions of a graph of type 10 with 3 circuits starting from R.



(a) $k, \ell \geq 1$ or $\ell \geq 2$



(b) $k \geq 1$, $\ell = 0$



(c) $k = 0$, $\ell = 1$

Figure 5.10: Evolutions of a graph of type 10 with 2 circuits starting from R.

vertex 7/8. As already mentioned, the vertex 9 can also be deleted (thanks to Lemma 5.6.5). Finally, Lemma 5.6.6 describes the sequence of evolutions while v_{i_n} corresponds to the vertex R in a graph of type 10. Consequently, if a graph evolves to a graph of type 10 such that $v_{i_n} = R$, there is only one possible finite sequence of evolutions, the one given by Lemma 5.6.6. Consequently, we can simply treat these evolutions by modifying the edges in C_4 as explained just below and we rename vertex 10 by $10B$, meaning that the vertex v_{i_n} always corresponds to the vertex B in Figure 4.5(j).

Now let us modify the edges and/or their labels. All modifications are direct consequences of Fact 5.6.1, Lemma 5.6.2, Lemma 5.6.4, Lemma 5.6.5 and Lemma 5.6.6:

- Fact 5.6.1 implies that we can consider only two morphisms to label the loop on vertex 1.
- Lemma 5.6.2 implies that the edges starting from 5/6 are the same as those starting from 6 in $\mathcal{G}$.
- By Lemma 5.6.6, we can replace each morphism γ_n labelling an edge coming to 10 in $\mathcal{G}$ such that $v_{i_{n+1}} = R$ by the corresponding behaviour given in that lemma. For instance, in $\mathcal{G}$, the morphism $\gamma_n = [12^k0, 2^\ell 0, 12^{k-1}0]$ labels an edge from 6 to 10. By Lemma 5.6.6, this morphisms makes the graph of type 10 evolving to a graph of type 7 or 8 or 10 depending on k and ℓ. Consequently, we delete this morphism and add two morphisms: the morphism $\gamma_n \circ [1,0,2]$ from 5/6 to $10B$ with $k = \ell$ (case *i.*) and the morphism $\gamma_n \circ [1,0,2]$ from 5/6 to 7/8 with $\ell < k$. To keep working with the same notation, this new morphisms are still denoted by γ_n.
- In Lemma 5.6.4 (so also in Lemma 5.6.5), as the behaviours depend on some lengths in Rauzy graphs, we simply consider the needed outgoing edges of the vertex 7/8 to be able to follow all described behaviours and put some restrictions on the choices in Proposition 5.6.8.

We then obtain the modified component C_4 represented in Figure 5.11 with labels as given below; those are trivially compositions of morphisms of $\mathcal{S}$. We will also see that it is more convenient to modify a bit more that component.

1. *Morphisms starting from the vertex* 1*:*

From 1 to	Morphisms γ_n	Conditions
1	$[0, 10]$, $[01, 1]$	
7/8	$[x, y^k x, (y^{k-1}x)]$	$k \geq 2$

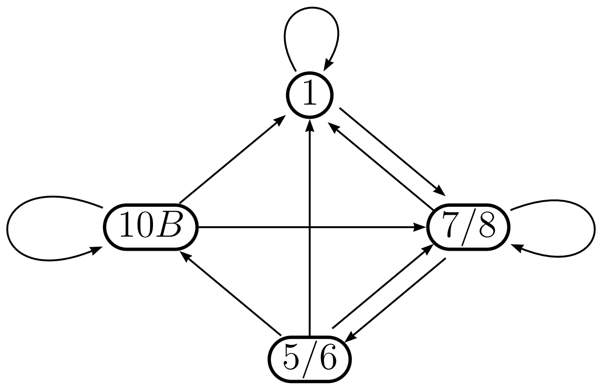


Figure 5.11: First attempt to modify the component C_4 in $\mathcal{G}$.

2. *Morphisms starting from the vertex* 5/6*:*

From 5/6 to	Morphisms γ_n	Conditions
1	$[x, yx], [yx, x]$	
	$[12^k0, 2^k0]$, $[2^k0, 12^k0]$	$k \geq 1$
	$[12^k0, 2^{k+1}0]$, $[2^{k+1}0, 12^k0]$	$k \geq 0$
7/8	$[1, 0^k2, (0^{k-1}2)]$	$k \geq 1$
	$[x, y^kx, (y^{k-1}x)]$	$k \geq 2$
	$[2^\ell0, 12^k0, (12^{k-1}0)]$	$k > \ell \geq 0$
	$[12^k0, 2^\ell0, (2^{\ell-1}0)]$	$\ell > k+1 \geq 1$
10B	$[1, 01, 2]$	
	$[2^k0, 12^k0, 12^{k-1}0]$	$k \geq 1$
	$[12^k0, 2^{k+1}0, 2^k0]$	$k \geq 0$

3. *Morphisms starting from the vertex* 7/8*:*

From 7/8 to	Morphisms γ_n	Conditions
1	$[01, 1]$, $[1, 01]$, $[x, y]$	
5 or 6	$[0x, y, (0y)], [x, 0y, (y)]$	
7/8	$[0, 10, (20)]$	

4. *Morphisms starting from the vertex* $10B$*:*

From 10B to	Morphisms γ_n	Conditions
1	$[01^k2, 1^k2]$, $[1^k2, 01^k2]$	$k \geq 1$
	$[01^k2, 1^{k+1}2]$, $[1^{k+1}2, 01^k2]$	$k \geq 0$
7/8	$[0, 2^k1, 2^{k-1}1]$	$k \geq 1$
	$[1^\ell 2, 01^k2, (01^{k-1}2)]$	$k > \ell \geq 0$
	$[01^k2, 1^\ell 2, (1^{\ell-1}2)]$	$\ell > k+1 \geq 1$
10B	$[0, 20, 1]$	
	$[1^k2, 01^k2, 01^{k-1}2]$	$k \geq 1$
	$[01^k2, 1^{k+1}2, 1^k2]$	$k \geq 0$

The next lemma describes paths in Figure 5.11 whose label are almost primitive.

Lemma 5.6.7. *A sequence of morphisms $(\gamma_n)_{n\geq N}$ labelling an infinite path p in Figure 5.11 is almost primitive if and only if one of the following conditions is satisfied:*

1. *p ultimately stays in vertex 1 and both morphisms $[0, 10]$ and $[01, 1]$ occur infinitely often in $(\gamma_n)_{n\geq N}$;*
2. *p ultimately stays in the subgraph $\{1, 7/8\}$, goes through both vertices infinitely often and for all suffixes p' of p starting in vertex 7/8, the label of p' is not only composed of finite sub-sequences of morphisms in*

$$\big([0, 10]^*[0, 1][0, 10]^*\{[0, 1^k0] \mid k \geq 2\}\big) \cup \big([0, 10]^*[1, 0][01, 1]^*\{[1, 0^k1] \mid k \geq 2\}\big)\,;$$

3. *p contains infinitely many occurrences of sub-paths q that start in vertex 1 and end in vertex 5/6.*
4. *p ultimately stays in the subgraph $\{5/6, 7/8, 10B\}$ and does not ultimately correspond to one of the two following configurations:*
   - (a) *the path ultimately stays in vertex 7/8;*
   - (b)
     - *the edge from 7/8 to 5/6 is labelled by $[1, 02, 2]$ or by $[01, 2, 02]$;*
     - *the edge from 5/6 to 7/8 is labelled by $[1, 02, 2]$;*
     - *the edge from 5/6 to 10B is labelled by $[1, 01, 2]$;*
     - *for all sub-paths q uniquely composed of loops over 10B, the label of q contains only occurrences of morphisms in*

$$\left\{[0, 20, 1]^{2n}, [02, 12, 2] \mid n \in \mathbb{N}\right\};$$

- *for all finite sub-paths q composed of loops over $10B$ and followed by the edge from $10B$ to $7/8$, the label of q is in*

$$\left\{[0,20,1]^{2n},[02,12,2]\mid n\in\mathbb{N}\right\}^{*}\{[2,012,02],[0,20,1][0,21,1]\};$$

(c)
- *the paths does not go through the loop over vertex $7/8$;*
- *the loop over vertex $10B$ is labelled by $[12^k0,2^{k+1}0,2^k0]$ for some integer $k\geq 0$;*
- *the edge from $5/6$ to $7/8$ is labelled either by $[1,0^k2,0^{k-1}2]$ for some integer $k\geq 1$ or by $[12^k0,2^\ell 0,2^{\ell-1}0]$ for some integers k and ℓ such that $\ell > k+1\geq 1$;*
- *the edge from $7/8$ to $5/6$ is labelled by $[1,02,2]$ or by $[2,01,1]$;*
- *the edge from $10B$ to $7/8$ is labelled by $[0,2^k1,2^{k-1}1]$ for some integer $k\geq 1$.*

Proof. The proof of this lemma is not really hard, but quite long so it is given in Appendix C page 213. □

Of course, we will also have to consider left conjugates of right proper morphisms. In the above list, their decompositions into compositions of morphisms in $\mathcal{S}$ can be easily computed using those given in Section 4.6 to Section 4.6 (from page 133 to page 136). But, it is still possible to consider valid directive words containing only non-right proper morphisms (so that make left conjugates impossible to compute directly). For instance, any path oscillating between $5/6$ and $7/8$ such that the edge from $5/6$ to $7/8$ is labelled by $[1,0^k2,0^{k-1}2]$ can be a suffix of a valid path: Lemma 5.6.2 and Lemma 5.6.4 ensure that the local condition of Proposition 5.1.5 is satisfied and if the morphism labelling the edge from $7/8$ to $5/6$ is $[01,2,02]$, then the composition of it with $[1,0^k2,0^{k-1}2]$ provides

$$[1,0^k2,0^{k-1}2]\circ[01,2,02]=[10^k2,0^{k-1}2,10^{k-1}2].$$

This last morphism is right proper and would label a loop on $5/6$ in Figure 5.11. Moreover, it can be trivially decomposed into morphisms in $\mathcal{S}$ since so are $[1,0^k2,0^{k-1}2]$ and $[01,2,02]$. It is also right proper, primitive and its left conjugate admits the following decomposition:

$$[210^k,20^{k-1},210^{k-1}]=D_{1,0}^{k-1}\circ G_{1,2}\circ D_{2,0}^{k-1}\circ G_{0,1}\circ[0,2,1].$$

As proved by Proposition 5.6.8, this kind of problem can be solved by adding two edges in Figure 5.11 labelled by the following additional morphisms. We then obtain the modified component as represented in Figure 5.12.

1. *additional loop on* 5/6 *labelled by the following morphisms:*

Morphisms γ_n	Conditions
$[10^k2, 0^{k-1}2, 10^{k-1}2]$	$k \geq 1$
$[10^{k-1}2, 0^k2, 10^k2]$	
$[0^k2, 10^{k-1}2, 0^{k-1}2]$	
$[0^{k-1}2, 10^k2, 0^k2]$	

2. *additional edge from* $10B$ *to* 5/6 *labelled by the following morphisms:*

Morphisms γ_n	Conditions
$[02^k1, 2^{k-1}1, 02^{k-1}1]$	$k \geq 1$
$[02^{k-1}1, 2^k1, 02^k1]$	
$[2^k1, 02^{k-1}1, 2^{k-1}1]$	
$[2^{k-1}1, 02^k1, 2^k1]$	

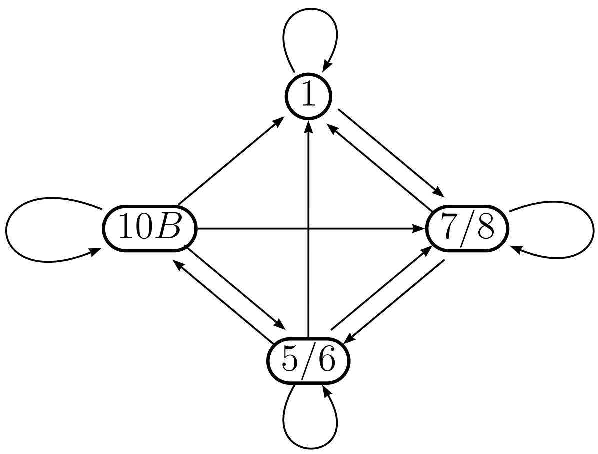


Figure 5.12: Graph corresponding to the component C_4 in $\mathcal{G}$.

Proposition 5.6.8. *Let* $\mathbf{s} = (\sigma_n)_{n\in\mathbb{N}}$ *be a sequence of morphisms in* $\mathcal{S}$*. Then there is an integer* $N \geq 0$ *such that* $(\sigma_n)_{n\geq N}$ *is a suffix of a valid directive word corresponding to a minimal subshift whose Rauzy graphs are ultimately of type 1, 5, 6, 7, 8, 9 or 10 if and only if there is a non-negative integers* $N' \leq N$*, a contraction* $(\Gamma_n)_{n\in\mathbb{N}}$ *of* $(\sigma_n)_{n\in\mathbb{N}}$ *and a sequence of morphisms* $(\gamma_n)_{n\geq N'}$ *such that*

1. *there are infinitely many right proper morphisms in* $(\gamma_n)_{n\geq N'}$ *and for all integers* $n \geq N'$*,* Γ_n *is either* γ_n *or* $\gamma_n^{(L)}$ *and there are infinitely many right proper morphisms and infinitely many left proper morphisms in* $(\Gamma_n)_{n\geq N'}$*;*

2. $(\gamma_n)_{n\geq N'}$ *labels an infinite path* p *in the graph represented in Figure 5.12 (whose labels are given on page 162 and on page 165) such that*

   (A) *if for some integer* $n \geq N'$, γ_n *labels an edge to* $5/6$, *then* γ_{n+1} *can be in* $\{[x, y^k x, (y^{k-1}x)] \mid \{x, y\} = \{0, 1\}, k \geq 2\}$ *only if* $|p_1| \geq |p_2|$ *where* $|p_1|$ *and* $|p_2|$ *are computed in Section B.2 (on page 209);*

   (B) *if for some integer* $n \geq N'$, γ_n *labels an edge to* $7/8$ *but not from* $7/8$, *then it is equal to* $[w_1, w_2 w_3^k w_4, w_2 w_3^{k-1} w_4]$ *for some words* w_1, w_2, w_3 *and* w_4 *and for an integer* $k \geq 1$ *which corresponds to the number of times that the* $(i_n + 1)$*-circuit* $\vartheta_{i_n+1}(1)$ *goes through the loop* $v_2 u_2$ *in Figure 5.7(b). Then, if* h *is the greatest integer such that* $\gamma_{n+i} = [0, 10]$ *for all* $i = 1, \ldots, h$, *then* h *is finite and* γ_{n+h+1} *can be in* $\{[0, 1], [1, 0]\}$ *if only if* $|u_1| + h(|u_1| + |v_1|) \geq |u_2| + (k-1)(|u_2| + |v_2|)$ *where* $|u_1|$, $|v_1|$, $|u_2|$ *and* $|v_2|$ *are computed in Section B.1 (on page 200);*

   and such that one of the following conditions is satisfied

   (i) p *ultimately stays in vertex 1 and both morphisms* $[0, 10]$ *and* $[01, 1]$ *occur infinitely often in* $(\gamma_n)_{n\geq N}$*;*

   (ii) p *ultimately stays in the subgraph* $\{1, 7/8\}$, *goes through both vertices infinitely often and for all suffixes* p' *of* p *starting in vertex* $7/8$, *the label of* p' *is not only composed of finite sub-sequences of morphisms in*

   $$\big([0, 10]^*[0, 1][0, 10]^*\{[0, 1^k 0] \mid k \geq 2\}\big) \cup \big([0, 10]^*[1, 0][01, 1]^*\{[1, 0^k 1] \mid k \geq 2\}\big) ;$$

   (iii) p *contains infinitely many occurrences of sub-paths* q *that start in vertex* 1 *and end in vertex* $5/6$.

   (iv) p *ultimately stays in the subgraph* $\{5/6, 7/8, 10B\}$ *and does not ultimately correspond to one of the two following configurations:*

      (a) *the path ultimately stays in vertex* $7/8$*;*

      (b)
      - *the loop over* $5/6$ *is labelled by* $[02, 12, 2]$ *or by* $[102, 2, 12]$*;*
      - *the edge from* $5/6$ *to* $7/8$ *is labelled by* $[1, 02, 2]$*;*
      - *the edge from* $5/6$ *to* $10B$ *is labelled by* $[1, 01, 2]$*;*
      - *the edge from* $7/8$ *to* $5/6$ *is labelled by* $[1, 02, 2]$ *or by* $[01, 2, 02]$*;*

- *for all sub-paths q uniquely composed of loops over $10B$, the label of q contains only occurrences of morphisms in*

$$\left\{[0,20,1]^{2n},[02,12,2] \mid n \in \mathbb{N}\right\};$$

- *for all finite sub-paths q composed of loops over $10B$ and followed by the edge from $10B$ to $5/6$, the label of q is in*

$$\left\{[0,20,1]^{2n},[02,12,2] \mid n \in \mathbb{N}\right\}^* [0,20,1]\{[21,01,1],[021,1,01]\};$$

(c)
- *the paths does not go through the loop over vertex $7/8$;*
- *the loop over $5/6$ is labelled by $[0^k2, 10^{k-1}2, 0^{k-1}2]$ or by $[0^{k-1}2, 10^k2, 0^k2]$ for some integer $k \geq 1$;*
- *the loop over $10B$ is labelled by $[12^k0, 2^{k+1}0, 2^k0]$ for some integer $k \geq 0$;*
- *the edge from $5/6$ to $7/8$ is labelled either by $[1, 0^k2, 0^{k-1}2]$ for some integer $k \geq 1$ or by $[12^k0, 2^\ell 0, 2^{\ell-1}0]$ for some integers k and ℓ such that $\ell > k+1 \geq 1$;*
- *the edge from $7/8$ to $5/6$ is labelled by $[1,02,2]$ or by $[2,01,1]$;*
- *the edge from $10B$ to $5/6$ is labelled by $[2^k1, 02^{k-1}1, 2^{k-1}1]$ or by $[2^{k-1}1, 02^k1, 2^k1]$ for some integer $k \geq 1$;*
- *the edge from $10B$ to $7/8$ is labelled by $[0, 2^k1, 2^{k-1}1]$ for some integer $k \geq 1$.*

Proof. First let us define the integers N and N' of the result. As in Proposition 5.3.1 and in Proposition 5.5.2, our aim is to study the validity of the suffix of $\mathbf{s}$ that corresponds to evolutions of Rauzy graphs of type 1, 5, 6, 7, 8, 9 or 10. From all previous modifications of the component C_4, we consider that N' is the smallest integer such that γ'_N codes the evolution of a Rauzy graph of type 1, 5, 6, 7, 8 or 10 and, for Rauzy graphs of type 10, we also suppose that the vertex $v_{i_{N'}}$ is the vertex B in Figure A.21.

Then, the integer N is the integer such that $(\Gamma_n)_{n \geq N'}$ is a contraction of $(\sigma_n)_{n \geq N}$, where $(\Gamma_n)_{n \geq N'}$ is obtained from $(\gamma_n)_{n \in \mathbb{N}}$ by replacing some right proper morphisms by their left conjugate.

We have to characterize sequences of morphisms that satisfies conditions 1 and 2 of Proposition 5.1.5 (only for the suffix $(\sigma_n)_{n \geq N}$). The proper property in Proposition 5.1.5 is equivalent to the last condition of the result. But, this last one suppose that $(\gamma_n)_{n \geq N'}$ contains infinitely many occurrences of right proper morphisms (which is actually supposed in the first condition of the result). As explained above, working with Figure 5.11 does not ensure that this condition is satisfied and this lead us to modify that graph as explained just below.

We would like that any valid labelled path p in Figure 5.11, there is a valid labelled path in Figure 5.12 whose label is a contraction of the label of p and contains infinitely many right proper morphisms. In Figure 5.11, the valid labelled path that contains only non-right proper morphisms are paths represented in Figure 5.13 where

1. the edge from 5/6 to $10B$ is labelled by $[1, 01, 2]$;
2. the edge from 5/6 to 7/8 is labelled by $[1, 0^k 2, 0^{k-1} 2]$;
3. the edge from 7/8 to 5/6 is labelled by $[0x, y, 0y]$ and $[x, 0y, x]$;
4. the edge from $10B$ to 7/8 is labelled by $[0, 2^k 1, 2^{k-1} 1]$;
5. the loop on $10B$ is labelled by $[0, 20, 1]$.

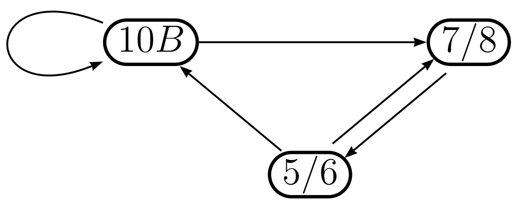


Figure 5.13: Part of Figure 5.11 where there might be some valid labelled path with only non-right proper morphisms as labels.

It is easily seen that labelled path in Figure 5.13 that ultimately stay in vertex $10B$ are not valid. Moreover, the labels of the path of length 2 from 5/6 to 5/6 (passing through 7/8) are right proper and equal to

$$[1, 0^k 2, 0^{k-1} 2] \circ [01, 2, 02] = [10^k 2, 0^{k-1} 2, 10^{k-1} 2]$$
$$[1, 0^k 2, 0^{k-1} 2] \circ [02, 1, 01] = [10^{k-1} 2, 0^k 2, 10^k 2]$$
$$[1, 0^k 2, 0^{k-1} 2] \circ [1, 02, 2] = [0^k 2, 10^{k-1} 2, 0^{k-1} 2]$$
$$[1, 0^k 2, 0^{k-1} 2] \circ [2, 01, 1] = [0^{k-1} 2, 10^k 2, 0^k 2]$$

Similarly, the labels of the path of length 2 from $10B$ to $5/6$ (passing through $7/8$) are right proper and equal to

$$\begin{aligned}
[0, 2^k1, 2^{k-1}1] \circ [01, 2, 02] &= [02^k1, 2^{k-1}1, 02^{k-1}1] \\
[0, 2^k1, 2^{k-1}1] \circ [02, 1, 01] &= [02^{k-1}1, 2^k1, 02^k1] \\
[0, 2^k1, 2^{k-1}1] \circ [1, 02, 2] &= [2^k1, 02^{k-1}1, 2^{k-1}1] \\
[0, 2^k1, 2^{k-1}1] \circ [2, 01, 1] &= [2^{k-1}1, 02^k1, 2^k1]
\end{aligned}$$

To our aim, it suffices therefore to add two edges in Figure 5.11: one loop on $5/6$ labelled by the first four morphisms above and one edge from $10B$ to $5/6$ labelled by the last four morphisms above. Indeed, if p is a valid labelled path in Figure 5.13 that contains only non-right proper morphisms in its label, it suffices to replace each subpath of length 2 from $5/6$ to $5/6$ by the new loop on $5/6$ and each subpath of length 2 from $10B$ to $5/6$ by the new edge from $10B$ to $5/6$.

We still have to prove that the left conjugates of the new 8 morphisms above can be decomposed into elements of $\mathcal{S}$. One can check that the following decompositions hold.

$$\begin{aligned}
[210^k, 20^{k-1}, 210^{k-1}] &= G_{1,2} \circ D_{1,0}^{k-1} \circ D_{2,0}^{k-1} \circ G_{0,1} \circ [0, 2, 1] \\
[210^{k-1}, 20^k, 210^k] &= G_{1,2} \circ D_{1,0}^{k-1} \circ D_{2,0}^{k} \circ G_{0,1} \circ [1, 2, 0] \\
[20^k, 210^{k-1}, 20^{k-1}] &= G_{1,2} \circ D_{2,0}^{k-1} \circ D_{1,0}^{k-1} \circ G_{0,2} \circ [0, 1, 2] \\
[20^{k-1}, 210^k, 20^k] &= G_{1,2} \circ D_{2,0}^{k-1} \circ D_{1,0}^{k} \circ G_{0,2} \circ [2, 1, 0] \\
[102^k, 12^{k-1}, 102^{k-1}] &= [2, 0, 1] \circ [210^k, 20^{k-1}, 210^{k-1}] \\
[102^{k-1}, 12^k, 102^k] &= [2, 0, 1] \circ [210^{k-1}, 20^k, 210^k] \\
[12^k, 102^{k-1}, 12^{k-1}] &= [2, 0, 1] \circ [20^k, 210^{k-1}, 20^{k-1}] \\
[12^{k-1}, 102^k, 12^k] &= [2, 0, 1] \circ [20^{k-1}, 210^k, 20^k]
\end{aligned}$$

With that modification of Figure 5.11, the proper condition of Proposition 5.1.5 is equivalent to the condition 1 of the result. For the first condition of Proposition 5.1.5 (the local one), it is a direct consequence of all previous lemmas and modifications of C_4:

1. any finite path passing only through the vertex 1 is trivially valid;
2. the condition 2A of the result summarizes what is allowed according to Lemma 5.6.2 for vertex $5/6$;
3. the condition 2B summarizes what is allowed with vertex $7/8$ according to Lemma 5.6.4 and Lemma 5.6.5;

4. the edges going to the vertex 10 in Figure 4.8 (page 112) have been modified according to Lemma 5.6.6.

It remains therefore to check the almost primitive property. It is easily seen that conditions 2i to 2iv are exactly those obtained in Lemma 5.6.7, but modified according to the added edges. □

5.7 Links between components

Now that we know how the suffixes of valid paths in each component must behave, it remains to describe all links between them. To this aim, it suffices to look at the graph of graphs $\mathcal{G}$ (Figure 4.8 page 112) and, like we did in each component, to study the consequences of a given morphism γ_n on the sequel in the directive word. For instance, in $\mathcal{G}$ there is an edge from 2 to 4 which is labelled by morphisms γ_n depending on some exponents k and ℓ (see on page 123). Then, Lemma 5.5.1 (page 149) states that, depending on k and ℓ, the graph will evolve to a graph of type 1, 4, 7 or 8 and 10 (with $v_{i_m} = B$) and it provides the morphism τ coding this evolution. Consequently, we add edges (if necessary) from 2 to $\{1, 4B, 7/8, 10B\}$ labelled by $\gamma_n \circ \tau$. This yield to the *modified graph of graphs* $\mathcal{G}'$ represented in Figure 5.14 (gray edges are simply those inner components). Labels of black edges are given below. In the list of morphisms, we express in the column "Trough" if the morphism is the result of a contraction like just explained. In the previous example, we would write $4R$ in the column "Through", meaning that the morphisms is a composition of γ_n and τ and that γ_n codes an evolution to a Rauzy graph of type 4 such that v_{i_n+1} corresponds to the vertex R in Figure 4.5(d).

Observe that, since black edges can only occur in a finite prefix of any valid path in $\mathcal{G}'$, we do not have to compute left conjugates of morphisms.

Remark 5.7.1. It is important to notice that the exponents k and ℓ in morphisms γ_n do not always correspond to the integers k and ℓ in Lemma 5.5.1, Lemma 5.6.4 and Lemma 5.6.6. Indeed, if for instance we consider the evolution of a Rauzy graph of type 2 to a Rauzy graph of type 4 as represented in Figure 5.15. The morphism coding this evolution is either $[yz^kx, z^\ell x, yz^{k-1}x]$ or $[z^kx, yz^\ell x, z^{k-1}x]$ for some integers k and ℓ. But, the circuits $\vartheta_{i_n+1}(0)$ and $\vartheta_{i_n+1}(1)$ go respectively $k-1$ and $\ell-1$ times through the loop.

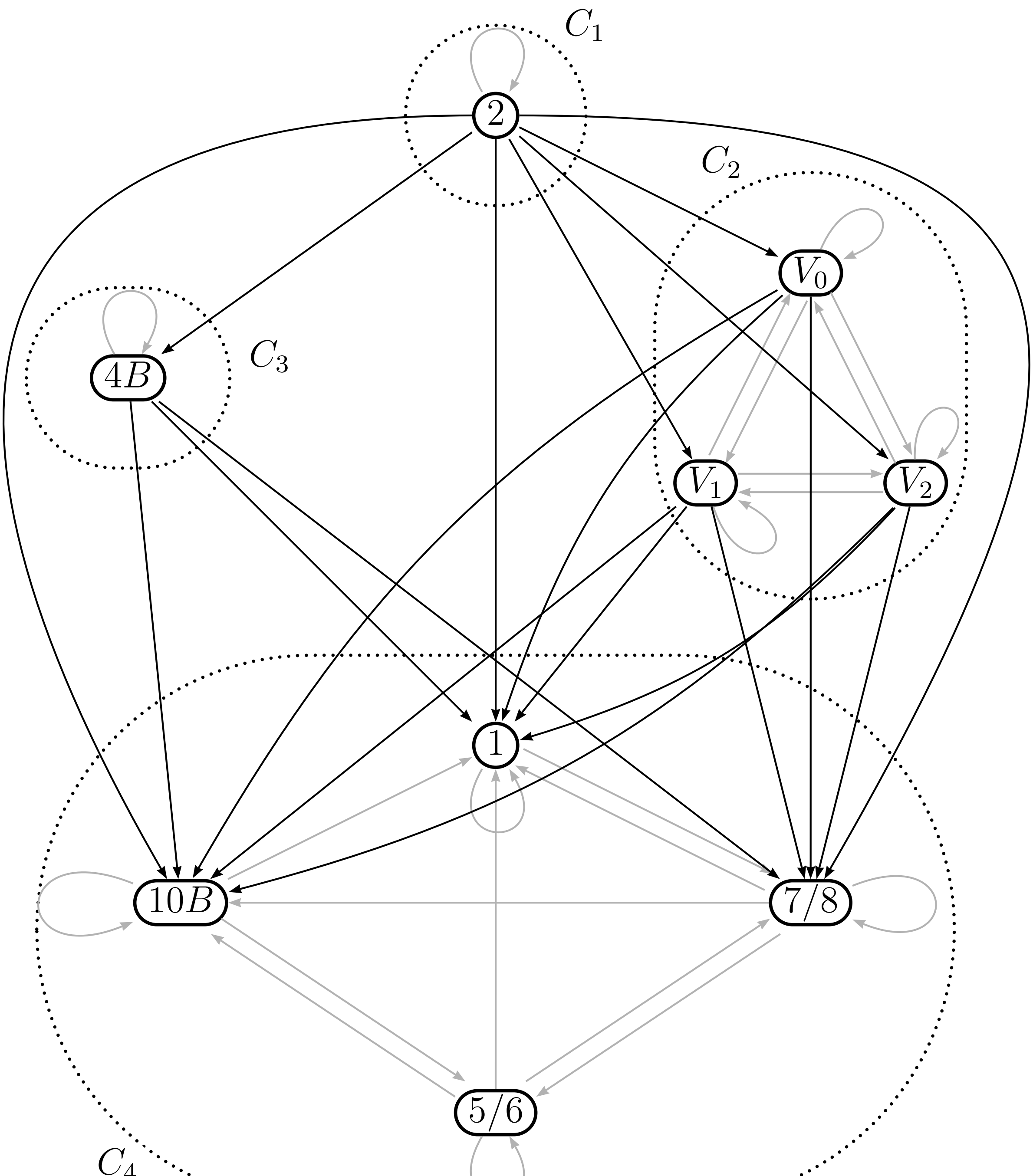
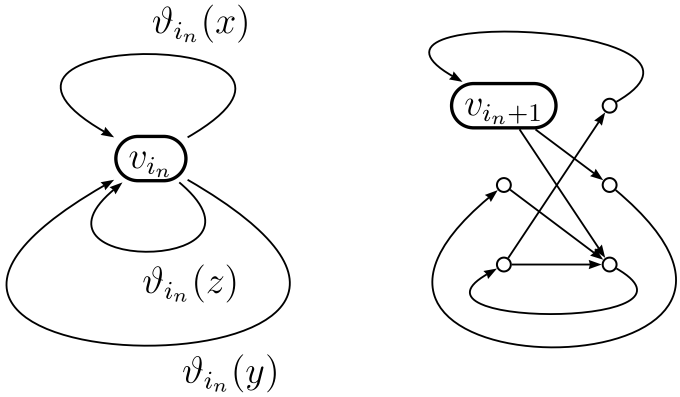


Figure 5.14: Modified graph of graphs.



(a) Before evolution (b) After evolution

Figure 5.15: Evolution of a graph of type 2 to a graph of type 4.

Morphisms labelling the black edge from 2 to $4B$ in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[x, yx, yzx]$, $[y, yzx, yx]$	
$4R$	$[y^{k-1}z, xy^kz, xy^{k-1}z]$ $[y^{k-1}z, xy^{k-1}z, xy^kz]$ $[xy^{k-1}z, y^kz, y^{k-1}z]$ $[xy^{k-1}z, y^{k-1}z, y^kz]$	$k \geq 2$

Morphisms labelling the black edge from 2 to V_i, $i \in \{0, 1, 2\}$, in $\mathcal{G}'$

Through	To	Morphisms	Conditions
/	V_0	$[0, 120, 20]$, $[0, 10, 210]$	
	V_1	$[01, 1, 201]$, $[021, 1, 21]$	
	V_2	$[02, 102, 2]$, $[012, 12, 2]$	

Morphisms labelling the black edge from 2 to 1 in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[x, yzx]$, $[yzx, x]$, $[xy, zy]$ $[xy, zxy]$, $[zxy, xy]$	
$4R$	$[yz^kx, z^kx]$, $[z^kx, yz^kx]$ $[yz^kx, z^{k-1}x]$, $[z^{k-1}x, yz^kx]$ $[yz^{k-1}x, z^kx]$, $[z^kx, yz^{k-1}x]$	$k \geq 2$
$10R$	$[(xy)^kz, y(xy)^kz]$, $[y(xy)^kz, (xy)^kz]$	$k \geq 1$
	$[(xy)^kz, y(xy)^{k-1}z]$, $[y(xy)^{k-1}z, (xy)^kz]$	$k \geq 2$

Morphisms labelling the black edge from 2 to $10B$ in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[xy, zxy, zy]$	
$4R$	$[z^kx, yz^kx, yz^{k-1}x]$ $[yz^kx, z^kx, z^{k-1}x]$	$k \geq 2$
$10R$	$[y(xy)^{k-1}z, (xy)^kz, (xy)^{k-1}z]$	$k \geq 2$
	$[(xy)^kz, y(xy)^kz, y(xy)^{k-1}z]$	

Morphisms labelling the black edge from 2 to $7/8$ in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[x, y^kzx, (y^{k-1}zx)]$ $[x, zy^kx, (zy^{k-1}x)]$ $[x, (yz)^kx, ((yz)^{k-1}x)]$ $[xy, z^kxy, (z^{k-1}xy)]$ $[xy, z^ky, (z^{k-1}y)]$	$k \geq 2$
	$[x, (yz)^kyx, ((yz)^{k-1}yx)]$	$k \geq 1$
$4R$	$[z^\ell x, yz^kx, yz^{k-1}x]$ $[yz^\ell x, z^kx, z^{k-1}x]$	$k-1 > \ell \geq 1$
$10R$	$[y(xy)^\ell z, (xy)^kz, (xy)^{k-1}z]$	$k-1 > \ell \geq 0$
	$[(xy)^kz, y(xy)^\ell z, y(xy)^{\ell-1}z]$	$\ell > k \geq 1$

Morphisms labelling the black edge from V_i, $i \in \{0,1,2\}$ to 1 in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[x, iy]$, $[iy, x]$, $[xi, yi]$	
$10R$	$[xy^ki, y^ki]$, $[y^ki, xy^ki]$	$k \geq 1$
	$[xy^ki, y^{k-1}i]$, $[y^{k-1}i, xy^ki]$	$k \geq 2$

Morphisms labelling the black edge from V_i, $i \in \{0,1,2\}$ to $10B$ in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[x, ix, iy]$	
$10R$	$[xy^{k-1}i, y^ki, y^{k-1}i]$	$k \geq 2$
	$[y^ki, xy^ki, xy^{k-1}i]$	$k \geq 1$

Morphisms labelling the black edge from V_i, $i \in \{0,1,2\}$ to 7/8 in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[i, xy^k i, xy^{k-1} i]$	$k \geq 1$
	$[x, i^k y, i^{k-1} y]$	$k \geq 2$
$10R$	$[xy^\ell i, y^k i, y^{k-1} i]$	$k \geq 2$, $k-1 > \ell \geq 0$
	$[y^k i, xy^\ell i, xy^{\ell-1} i]$	$\ell > k \geq 1$

Morphisms labelling the black edge from $4B$ to 1 in $\mathcal{G}'$

Through	Morphisms	Conditions
$4R$	$[x^k y, 0x^k y]$, $[0x^k y, x^k y]$ $[x^{k-1} y, 0x^k y]$, $[0x^k y, x^{k-1} y]$ $[x^k y, 0x^{k-1} y]$, $[0x^{k-1} y, x^k y]$	$k \geq 1$
$10R$	$[0(x0)^k y, (x0)^k y]$, $[(x0)^k y, 0(x0)^k y]$ $[0(x0)^{k-1} y, (x0)^k y]$, $[(x0)^k y, 0(x0)^{k-1} y]$	$k \geq 1$

Morphisms labelling the black edge from $4B$ to $10B$ in $\mathcal{G}'$

Through	Morphisms	Conditions
$4R$	$[x^k y, 0x^k y, 0x^{k-1} y]$ $[0x^k y, x^k y, x^{k-1} y]$	$k \geq 1$
$10R$	$[(x0)^k y, 0(x0)^k y, 0(x0)^{k-1} y]$ $[0(x0)^{k-1} y, (x0)^k y, (x0)^{k-1} y]$	$k \geq 1$

Morphisms labelling the black edge from $4B$ to 7/8 in $\mathcal{G}'$

Through	Morphisms	Conditions
/	$[0, x^k y0, x^{k-1} y0]$	$k \geq 1$
$4R$	$[x^\ell y, 0x^k y, 0x^{k-1} y]$ $[0x^\ell y, x^k y, x^{k-1} y]$	$k-1 > \ell \geq 0$
$10R$	$[(x0)^\ell y, 0(x0)^k y, 0(x0)^{k-1} y]$	$k > \ell \geq 0$
	$[0(x0)^k y, (x0)^\ell y, (x0)^{\ell-1} y]$	$\ell - 1 > k \geq 0$

5.8 Final Result

Now we can give an $\mathcal{S}$-adic characterization of minimal and aperiodic subshift with first difference of complexity bounded by 2. It suffices to put together all what we proved until now.

Theorem 5.8.1. *Let (X, T) be a subshift over an alphabet A and let*

$$\mathcal{S} = \{G, D, M, E_{01}, E_{12}\}$$

be the set of 5 morphisms as defined on page 103. Then, (X,T) is minimal and satisfies $1 \leq p_X(n+1) - p_X(n) \leq 2$ for all n if and only if (X,T) is $\mathcal{S}$-adic such that there exists a contraction $(\Gamma_n)_{n\in\mathbb{N}}$ of its directive word $(\sigma_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ and a sequence of morphisms $(\gamma_n)_{n\in\mathbb{N}}$ that labels an infinite path p in the graph represented at Figure 5.14 and such that

1. *there are infinitely many right proper morphisms in $(\gamma_n)_{n\in\mathbb{N}}$ and for all integers $n \geq 0$, Γ_n is either γ_n or $\gamma_n^{(L)}$ and there are infinitely many right proper morphisms and infinitely many left proper morphisms in $(\Gamma_n)_{n\in\mathbb{N}}$;*

2. *if p ultimately stays in the component C_1, then the three morphisms $[0,10,20]$, $[01,1,21]$ and $[02,12,2]$ occur infinitely often in $(\gamma_n)_{n\in\mathbb{N}}$;*

3. *if p ultimately stays in the component C_2, then the edges in C_2 are the following*

   (a) *for all $x \in \{0,1,2\}$, the loop on V_x is labelled by morphisms in*

   $$F_x = \{D_{y,x}D_{z,x},\ G_{y,x}G_{z,x},\ G_{x,y}G_{z,y} \mid \{x,y,z\} = \{0,1,2\}\};$$

   (b) *for all $x,y \in \{0,1,2\}$, $x \neq y$, the edge from V_x to V_y is labelled by morphisms in*

   $$F_{x\to y} = \{D_{x,z},\ D_{y,x}G_{x,z}G_{z,y} \mid z \notin \{x,y\}\};$$

   and if N is the smallest integer such that γ_N labels an edge in C_2, then for all $x \in \{0,1,2\}$, there are infinitely many integers $n \geq N$ such that $D_{y,x}$ is a factor of γ_n for some $y \in \{0,1,2\}$;

4. *if p ultimately stays in the component C_3 and if N is the smallest integer such that γ_N labels an edge in C_3, then*

   (a) *for all $n \geq N$,*

   $$\gamma_n \in \big\{[0,10,20],[0,20,10],[x^{k-1}y,0x^ky,0x^{k-1}y], [x^{k-1}y,0x^{k-1}y,0x^ky],[0x^{k-1}y,x^ky,x^{k-1}y], [0x^{k-1}y,x^{k-1}y,x^ky] \mid \{x,y\}=\{1,2\}, k\geq 1\big\};$$

   (b) *for all $r \geq N$,*

   $$(\gamma_n)_{n\geq r} \notin \{[0,10,20],[0,20,10]\}^{\mathbb{N}}$$

and

$$(\gamma_n)_{n\geq r} \notin \{[0x^{k-1}y, x^ky, x^{k-1}y], [0x^{k-1}y, x^{k-1}y, x^ky] \mid \{x,y\} = \{1,2\}, k \geq 1\}^{\mathbb{N}}$$

5. *if* p *ultimately stays in the component* C_4 *and if* N *is the smallest integer such that* γ_N *labels an edge in* C_4*, then*

   (A) *if for some integer* $n \geq N$*,* γ_n *labels an edge to* $5/6$*, then* γ_{n+1} *can be in* $\{[x, y^kx, (y^{k-1}x)] \mid \{x,y\} = \{0,1\}, k \geq 2\}$ *only if* $|p_1| \geq |p_2|$ *where* $|p_1|$ *and* $|p_2|$ *are computed in Section B.2 (on page 209);*

   (B) *if for some integer* $n \geq N$*,* γ_n *labels an edge to* $7/8$ *but not from* $7/8$*, then it is equal to* $[w_1, w_2w_3^kw_4, w_2w_3^{k-1}w_4]$ *for some words* w_1*,* w_2*,* w_3 *and* w_4 *and for an integer* $k \geq 1$ *which corresponds to the number of times that the* (i_n+1)*-circuit* $\vartheta_{i_n+1}(1)$ *goes through the loop* v_2u_2 *in Figure 5.7(b). Then, if* h *is the greatest integer such that* $\gamma_{n+i} = [0, 10]$ *for all* $i = 1, \ldots, h$*, then* h *is finite and* γ_{n+h+1} *can be in* $\{[0,1],[1,0]\}$ *if only if* $|u_1| + h(|u_1| + |v_1|) \geq |u_2| + (k-1)(|u_2| + |v_2|)$ *where* $|u_1|$*,* $|v_1|$*,* $|u_2|$ *and* $|v_2|$ *are computed in Section B.1 (on page 200);*

   and one of the following conditions is satisfied

   (i) p *ultimately stays in vertex 1 and both morphisms* $[0,10]$ *and* $[01,1]$ *occur infinitely often in* $(\gamma_n)_{n\geq N}$*;*

   (ii) p *ultimately stays in the subgraph* $\{1, 7/8\}$*, goes through both vertices infinitely often and for all suffixes* p' *of* p *starting in vertex* $7/8$*, the label of* p' *is not only composed of finite sub-sequences of morphisms in*

$$\big([0,10]^*[0,1][0,10]^*\{[0,1^k0] \mid k \geq 2\}\big) \cup \big([0,10]^*[1,0][01,1]^*\{[1,0^k1] \mid k \geq 2\}\big)\,;$$

   (iii) p *contains infinitely many occurrences of sub-paths* q *that start in vertex* 1 *and end in vertex* $5/6$*.*

   (iv) p *ultimately stays in the subgraph* $\{5/6, 7/8, 10B\}$ *and does not ultimately correspond to one of the two following configurations:*

      (a) *the path ultimately stays in vertex* $7/8$*;*

      (b) • *the loop over* $5/6$ *is labelled by* $[02,12,2]$ *or by* $[102,2,12]$*;*

- *the edge from* $5/6$ *to* $7/8$ *is labelled by* $[1, 02, 2]$*;*
- *the edge from* $5/6$ *to* $10B$ *is labelled by* $[1, 01, 2]$*;*
- *the edge from* $7/8$ *to* $5/6$ *is labelled by* $[1, 02, 2]$ *or by* $[01, 2, 02]$*;*
- *for all sub-paths* q *uniquely composed of loops over* $10B$*, the label of* q *contains only occurrences of morphisms in*

$$\left\{[0, 20, 1]^{2n}, [02, 12, 2] \mid n \in \mathbb{N}\right\};$$

- *for all finite sub-paths* q *composed of loops over* $10B$ *and followed by the edge from* $10B$ *to* $5/6$*, the label of* q *is in*

$$\left\{[0, 20, 1]^{2n}, [02, 12, 2] \mid n \in \mathbb{N}\right\}^* [0, 20, 1]\{[21, 01, 1], [021, 1, 01]\};$$

(c)
- *the paths does not go through the loop over vertex* $7/8$*;*
- *the loop over* $5/6$ *is labelled by* $[0^k2, 10^{k-1}2, 0^{k-1}2]$ *or by* $[0^{k-1}2, 10^k2, 0^k2]$ *for some integer* $k \geq 1$*;*
- *the loop over* $10B$ *is labelled by* $[12^k0, 2^{k+1}0, 2^k0]$ *for some integer* $k \geq 0$*;*
- *the edge from* $5/6$ *to* $7/8$ *is labelled either by* $[1, 0^k2, 0^{k-1}2]$ *for some integer* $k \geq 1$ *or by* $[12^k0, 2^\ell0, 2^{\ell-1}0]$ *for some integers* k *and* ℓ *such that* $\ell > k + 1 \geq 1$*;*
- *the edge from* $7/8$ *to* $5/6$ *is labelled by* $[1, 02, 2]$ *or by* $[2, 01, 1]$*;*
- *the edge from* $10B$ *to* $5/6$ *is labelled by* $[2^k1, 02^{k-1}1, 2^{k-1}1]$ *or by* $[2^{k-1}1, 02^k1, 2^k1]$ *for some integer* $k \geq 1$*;*
- *the edge from* $10B$ *to* $7/8$ *is labelled by* $[0, 2^k1, 2^{k-1}1]$ *for some integer* $k \geq 1$*.*

To obtain the exact complexities $p(n) = 2n$ or $p(n) = 2n + 1$, it suffices to impose respectively that $p(1) = 2$ or $p(1) = 3$ and that for all $n \geq 1$, $p(n + 1) - p(n) = 2$. This can be expressed by the fact the Rauzy graphs cannot be of type 1 (because these graphs are such that $p(n+1)-p(n) = 1$). Consequently, one just has to impose that the path p of the theorem does no go through vertex 1 except in some particular cases depending on the lengths $|u_1|$, $|u_2|$, $|v_1|$, $|v_2|$, $|p_1|$ and $|p_2|$.

Corollary 5.8.2. *A subshift* (X, T) *is minimal and has complexity* $p(n) = 2n$ *(resp.* $p(n) = 2n + 1$*) for all* $n \geq 1$ *if and only if it is an* $\mathcal{S}$*-adic subshift satisfying Theorem 5.8.1 and the following additional conditions:*

1. *the path p of Theorem 5.8.1 starts in vertex 1 (resp. vertex 2);*
2. *in Condition 5B of Theorem 5.8.1, the inequality*

$$|u_1| + h(|u_1| + |v_1|) \geq |u_2| + (k-1)(|u_2| + |v_2|)$$

is replaced by

$$|u_1| + h(|u_1| + |v_1|) = |u_2| + (k-1)(|u_2| + |v_2|)$$

and in that case, γ_{n+h+2} must label the edge from 1 *to* 7/8*;*

3. *in Condition 5A of Theorem 5.8.1, the inequality $|p_1| \geq |p_2|$ is replaced by $|p_1| = |p_2|$.*

Conclusions

The $\mathcal{S}$-adic characterization obtained in Chapter 4 and Chapter 5 is obviously a valuable improvement in the study of subshifts with very low complexity. The amount of given details will probably be very helpful to solve other problems related to these complexities such as giving a geometrical representation of subshifts with those complexities. On the other hand, the involved methods are too much technical to hope using them in a more general case. Indeed, even for minimal subshifts with a first difference of complexity bounded by 3 (instead of 2), computations are getting considerably more difficult. Furthermore, some crucial results seem to be closely linked to these low complexities (see Lemma 4.3.4 page 114 and Example 4.3.5 page 115).

However, one could try to highlight some general properties of the obtained $\mathcal{S}$-adic representations and check whether they can be generalized in a more general case. For instance, the next result states that the subshift generated by the set of all $\mathcal{S}$-adic representations of minimal subshifts with first complexity bounded by 2 is not *sofic*, i.e., $L(X_{\mathcal{S}})$ is not a regular language. But, one could prove that $L(X_{\mathcal{S}})$ is computable. A natural question is therefore whether there are some other properties that can be satisfied and, if yes, whether they are generalizable to a more general case.

Proposition 5.8.3. *The subshift $X_{\mathcal{S}}$ generated by all $\mathcal{S}$-adic representations of minimal subshifts such that $1 \leq p(n+1) - p(n) \leq 2$ is not sofic.*

Proof. Let us define the notion of *follower set*. If X is a subshift and u belongs to $L(X)$, then the *follower set* $F_X(u)$ is the set of all words in $L(X)$ that can follow u in X, i.e.,

$$F_X(u) = \{v \in L(X) : uv \in L(X)\}.$$

Follower sets can be used to characterize sofic subshifts: *a subshift is sofic if and only if it has a finite number of follower sets* (see Theorem 3.2.10 in [LM95]).

Let $u = u_1 \cdots u_{|u|}$ be a words in $L(X_{\mathcal{S}})$. Thus u is a finite sequence of morphisms labelling a finite path q in Figure 5.14 page 172. If $i(q)$ is the vertex 7/8 and $u_{|u|} \neq [0, 10, 20]$, then Theorem 5.8.1 implies that the family of words that can follow u in $X_{\mathcal{S}}$ depend on some lengths in Rauzy graphs (Condition 5B of the result). More precisely, u can be followed in $X_{\mathcal{S}}$ by $[0, 10]^h[0, 1]$ if and only if h is greater than a constant C that depends on some lengths in Rauzy graphs. When we consider a finite path in Figure 5.14 that starts and stays in vertex 1 before going to vertex 7/8, we see that the constant C can be chosen arbitrarily large. Indeed, it suffices to always consider the same morphism to label the loop over vertex 1 because in Rauzy graphs of type 1, this makes one loop becoming much longer than the other one. The set of follower sets is therefore infinite for $X_{\mathcal{S}}$. □

Another idea is to try to make stronger the necessary conditions obtained in Chapter 3. A first important work would be to make the almost primitivity necessary in all cases and not only when there are no constant segments. With notations of Chapter 3, we think (although we have no proof of it) that it should be possible to consider a sequence of sub-alphabets $(\tilde{B}_n \subset B_n)_{n \in \mathbb{N}}$ such that the directive word $(\tau_n : \tilde{B}_{n+1}^* \to \tilde{B}_n^*)$ is almost primitive.

An additional result that would generalize Durand's work would be to characterize the set of sequences for which the S-adic representation of Theorem 3.0.3 (page 76) is ultimately periodic. We are currently trying to solve this question with Štěpán Starosta.

Finally, beyond the conjecture, it would be interesting to improve the work initiated by Proposition 2.4.1 (page 72). This result provides a bound $(n \log n)$ over the complexity of expansive S-adic sequences with $\mathrm{Card}(S) < +\infty$. What could we say if we replace the expansivity by the condition of being everywhere growing? Could we say for instance that the complexity will be at most polynomial?

Appendix A

Evolution of Rauzy graphs such that $1 \leq p(n+1) - p(n) \leq 2$

A.1 Evolution of a Rauzy graph of type 1

A graph of type 1 is represented in Figure A.1. The possible evolutions are represented in Figure A.2.

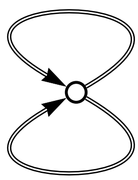
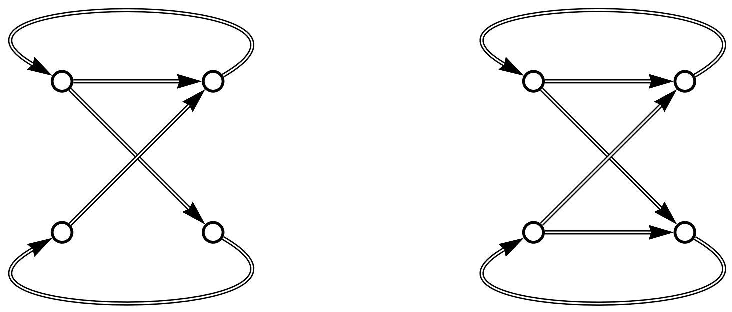


Figure A.1: Graph of type 1



(a) To a graph of type 1 (b) To a graph of type 7 or 8

Figure A.2: Possible evolutions for a graph of type 1

A.2 Evolution of a Rauzy graph of type 2

A graph of type 2 is represented in Figure A.3. The possible evolutions are represented in Figure A.4, Figure A.5 and Figure A.6.

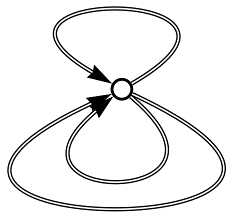
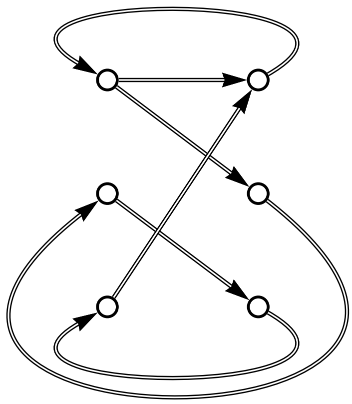
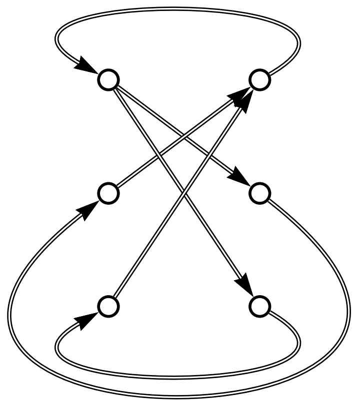


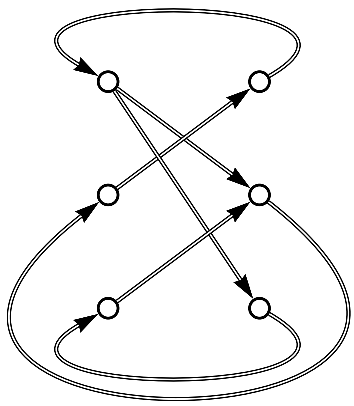
Figure A.3: Graph of type 2



(a) To a graph of type 1

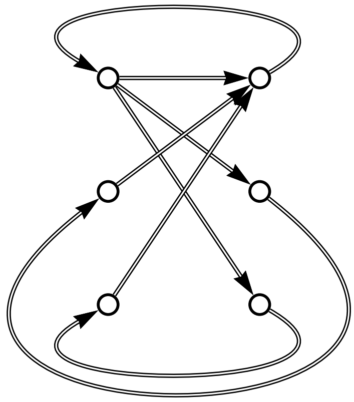


(b) To a graph of type 1

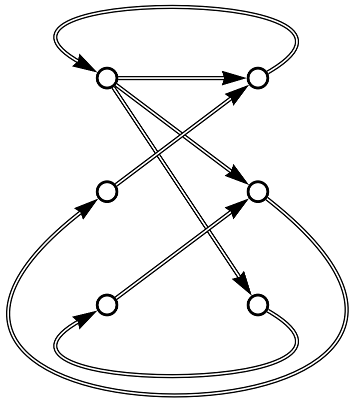


(c) To a graph of type 1

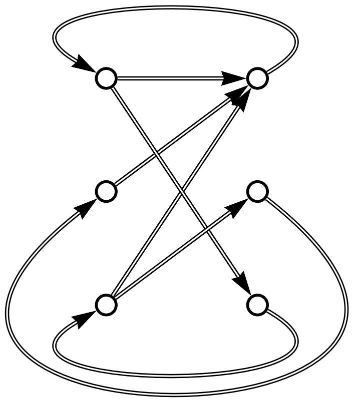
Figure A.4: Evolutions from 2 to 1



(a) To a graph of type 2

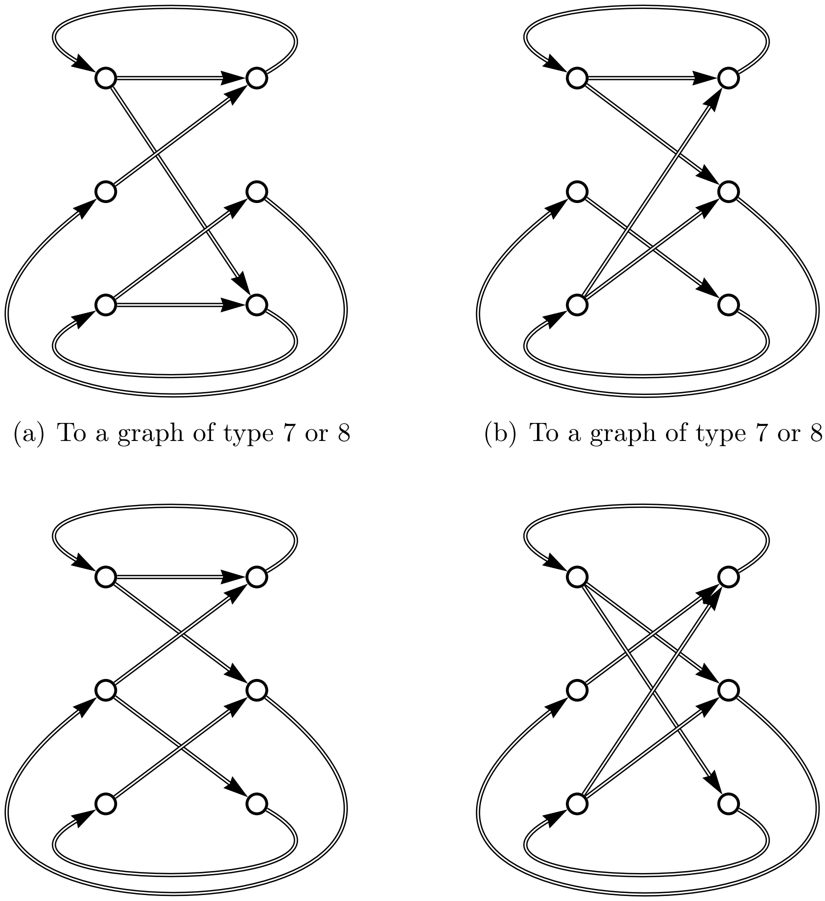


(b) To a graph of type 3



(c) To a graph of type 4

Figure A.5: Evolutions from 2 to $\{1, 2, 3, 4\}$



(a) To a graph of type 7 or 8

(b) To a graph of type 7 or 8

(c) To a graph of type 7 or 8

(d) To a graph of type 10

Figure A.6: Evolutions from 2 to $\{7, 8, 10\}$

A.3 Evolution of a Rauzy graph of type 3

A graph of type 3 is represented in Figure A.7. The possible evolutions are represented in Figure A.8.

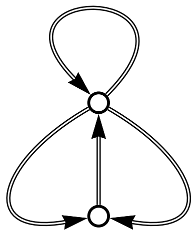
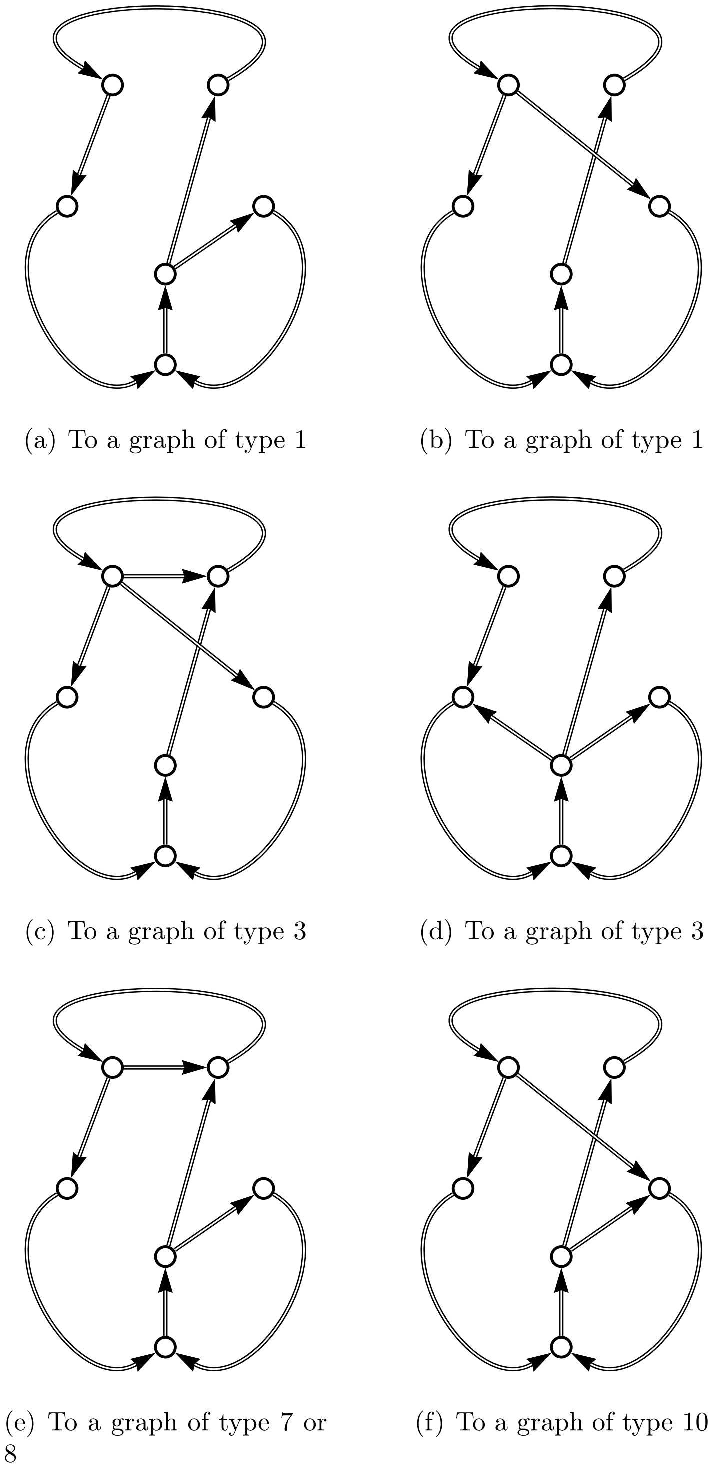


Figure A.7: Graph of type 3



(a) To a graph of type 1

(b) To a graph of type 1

(c) To a graph of type 3

(d) To a graph of type 3

(e) To a graph of type 7 or 8

(f) To a graph of type 10

Figure A.8: Possible evolutions of a graph of type 3

A.4 Evolution of a Rauzy graph of type 4

A graph of type 3 is represented in Figure A.9. The possible evolutions are represented in Figure A.10.

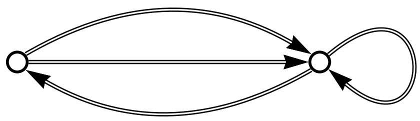
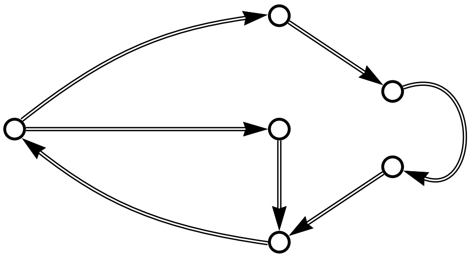
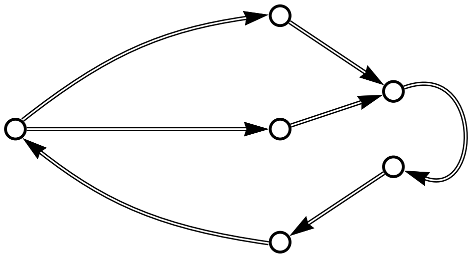


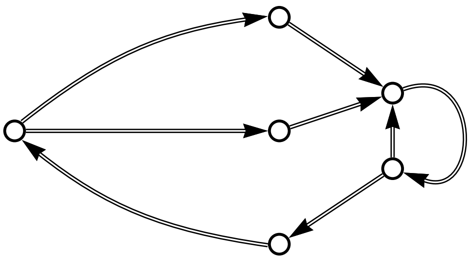
Figure A.9: Graph of type 4



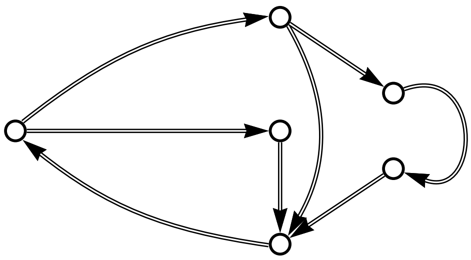
(a) To a graph of type 1



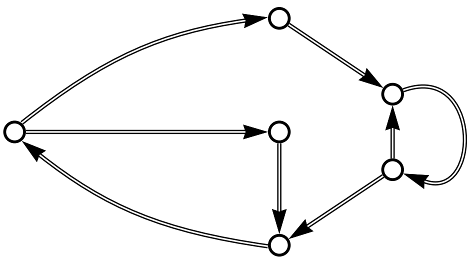
(b) To a graph of type 1



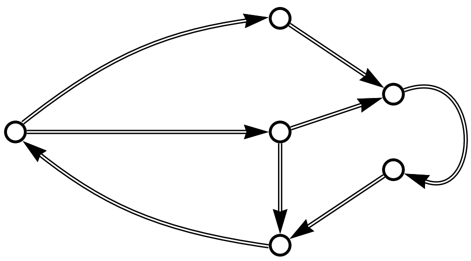
(c) To a graph of type 4



(d) To a graph of type 4



(e) To a graph of type 7 or 8



(f) To a graph of type 10

Figure A.10: Possible evolutions of a graph of type 4

A.5 Evolution of a Rauzy graph of type 5

A graph of type 3 is represented in Figure A.11. The possible evolutions are represented in Figure A.12.

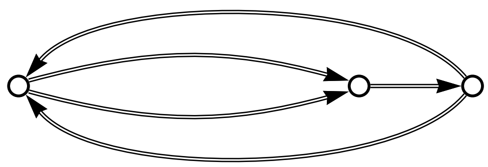
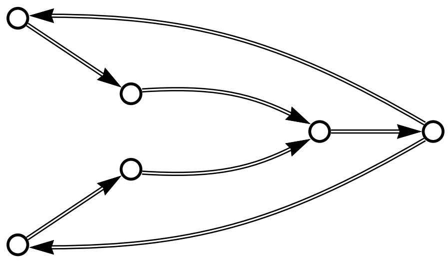
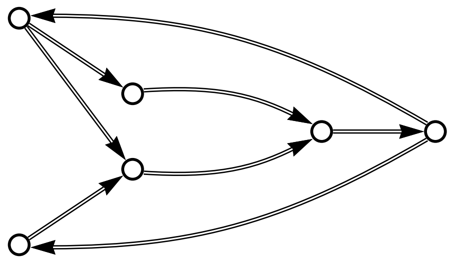


Figure A.11: Graph of type 5



(a) To a graph of type 1



(b) To a graph of type 10

Figure A.12: Possible evolutions of a graph of type 5

A.6 Evolution of a Rauzy graph of type 6

A graph of type 3 is represented in Figure A.13. The possible evolutions are represented in Figure A.14.

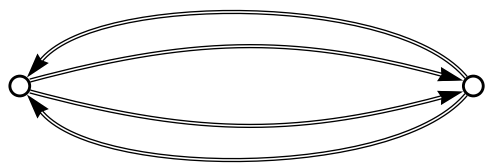
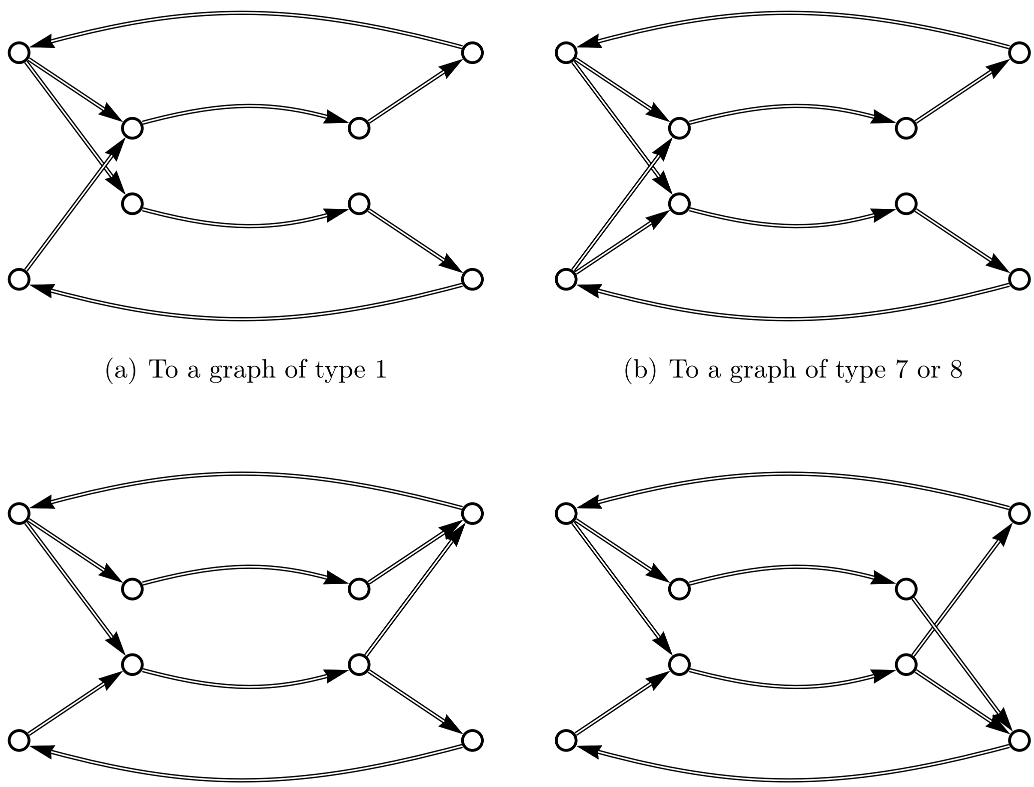


Figure A.13: Graph of type 6



(a) To a graph of type 1

(b) To a graph of type 7 or 8

(c) To a graph of type 7 or 8

(d) To a graph of type 10

Figure A.14: Possible evolutions of a graph of type 6

A.7 Evolution of a Rauzy graph of type 7

A graph of type 3 is represented in Figure A.15. The possible evolutions are represented in Figure A.16.

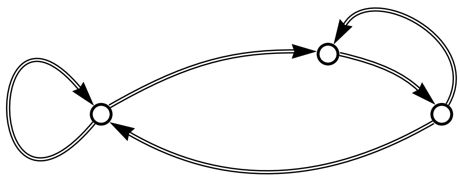
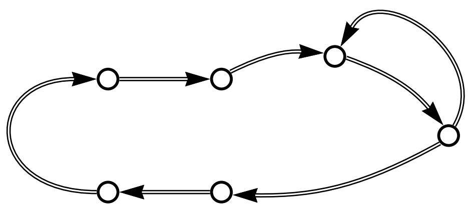
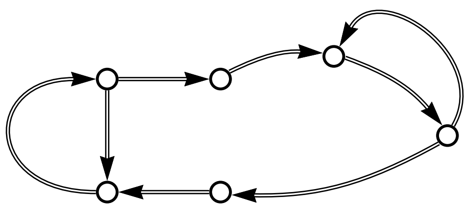


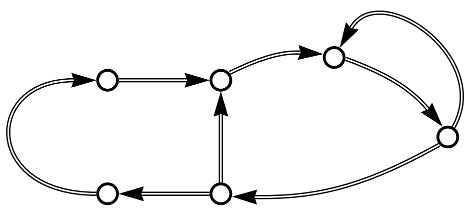
Figure A.15: Graph of type 7



(a) To a graph of type 1



(b) To a graph of type 7 or 8



(c) To a graph of type 9

Figure A.16: Possible evolutions of a graph of type 7

A.8 Evolution of a Rauzy graph of type 8

A graph of type 3 is represented in Figure A.17. The possible evolutions are represented in Figure A.18.

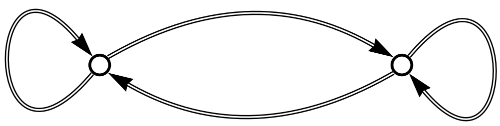
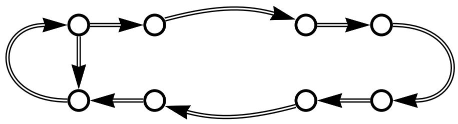
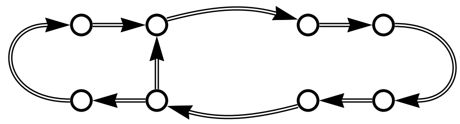


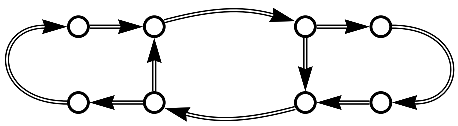
Figure A.17: Graph of type 8



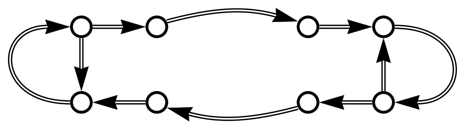
(a) To a graph of type 1



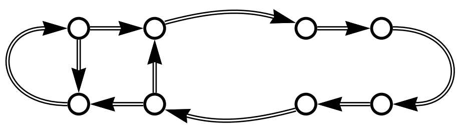
(b) To a graph of type 1



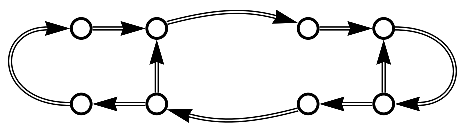
(c) To a graph of type 5 or 6



(d) To a graph of type 7 or 8



(e) To a graph of type 7 or 8



(f) To a graph of type 9

Figure A.18: Possible evolutions of a graph of type 7

A.9 Evolution of a Rauzy graph of type 9

A graph of type 3 is represented in Figure A.19. The possible evolutions are represented in Figure A.20.

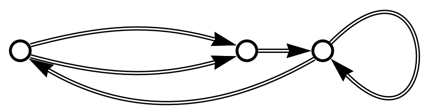
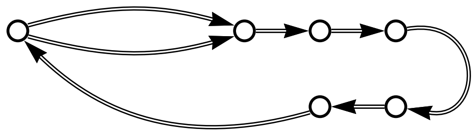
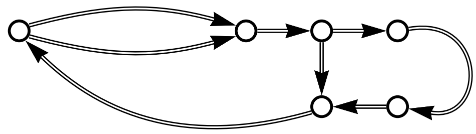


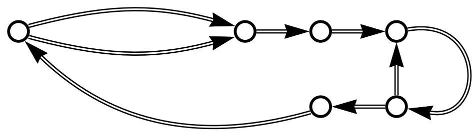
Figure A.19: Graph of type 9



(a) To a graph of type 1



(b) To a graph of type 5 or 6



(c) To a graph of type 9

Figure A.20: Possible evolutions of a graph of type 9

A.10 Evolution of a Rauzy graph of type 10

A graph of type 3 is represented in Figure A.21. The possible evolutions are represented in Figure A.22.

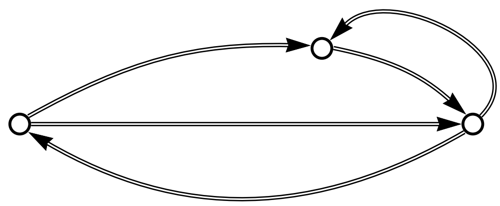
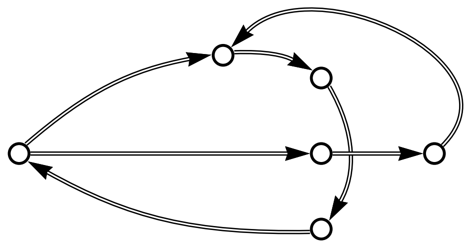
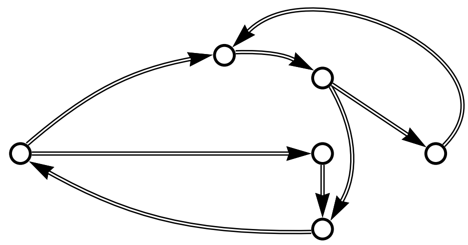


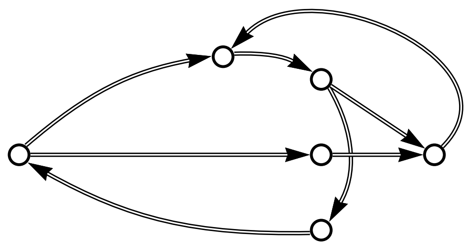
Figure A.21: Graph of type 10



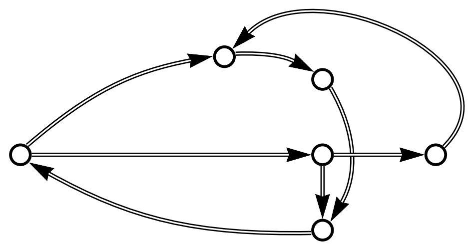
(a) To a graph of type 1



(b) To a graph of type 7 or 8



(c) To a graph of type 10



(d) To a graph of type 10

Figure A.22: Possible evolutions of a graph of type 10

Appendix B

Computation of length of paths in Rauzy graphs

To complete the proof of Theorem 5.8.1, we need to be able to compute some lengths in Rauzy graphs. However, when computing the $\mathcal{S}$-adic representation of our subshifts, we do not keep track of the order n of G_n. Consequently, we cannot simply compute the desired Rauzy graph and count the number of edges in the paths we are interested in. Moreover, that technique would not be efficient since the Rauzy graphs are getting bigger and bigger, making them harder to compute. To avoid this problem, we will compute lengths using the morphisms already computed. In other words, if for instance τ is a morphism labelling an edge to the vertex 7/8 and coding a loop (i.e., containing an exponent k or ℓ), we will express the lengths $|u_1|$, $|u_2|$, $|v_1|$ and $|v_2|$ using τ and morphisms preceding τ in the directive word.

Let us introduce some notations. We consider that $(\gamma_n)_{n\in\mathbb{N}}$ is the sequence of morphisms as in Theorem 5.8.1 and for all $n \geq 0$, we let $\gamma_{[0,n]}$ denote the morphism $\gamma_0 \cdots \gamma_n$. For any two words (or paths) u and v, we also let $\mathrm{CP}(u, v)$ and $\mathrm{CS}(u, v)$ respectively denote the longest common prefix and suffix of u and v.

The computation of lengths in Rauzy graphs is based on the following fact which is a direct consequence of the constructions.

Fact B.0.1. *Let $G_{i_{n+1}}$ be a Rauzy graph of a minimal subshift whose first difference of complexity satisfies $1 \leq p(n+1) - p(n) \leq 2$ for all n. If $\gamma_{[0,n]}$ is the morphism coding the evolution from G_0 to $G_{i_{n+1}}$, then for all $x \in \{0, 1, 2\}$, we have*

$$\gamma_{[0,n]}(x) = \lambda_{R,i_{n+1}} \circ \vartheta_{i_{n+1}}(x).$$

Observe that this result does not hold anymore if we replace $\gamma_{[0,n]}$ by $\Gamma_{[0,n]}$. We will also need the following lemma.

Lemma B.0.2. *Let (X,T) be a subshift over A. For all words $u \in L(X)$, there is at most one return word r to u such that $|w| \leq \frac{|u|}{2}$. As a corollary, for all n at most one n-circuit has for length at most $\frac{n}{2}$.*

Proof. The last part of the lemma is a direct consequence of Remark 3.1.12 (page 81). Let us recall that $LRW_X(u)$ and $RRW_X(u)$ respectively denote the set of left and right return words to u. Since for all $u \in L(X)$ we have

$$\{|r| \mid r \in LRW_X(u)\} = \{|r| \mid r \in RRW_X(u)\},$$

it is sufficient to prove this for left return words.

Let $u \in L(X)$ and let r be a return word to u with minimal length. By definition, u is prefix of ru. Therefore, if $|r| \leq \frac{|u|}{2}$, r is a prefix of u and we can write $u = r^k r_{[1,j]}$ with $k \in \mathbb{N}$, $k \geq 1$ and $j \in \{0, \dots, |r|-1\}$. Consequently, u is $|r|$-periodic, i.e., $u_{i+|r|} = u_i$ for all $i \in \{1, \dots, |u|-|r|\}$.

If there is another return word s to u such that $|s| \leq \frac{|u|}{2}$, we deduce similarly that s is a prefix of u and that u is $|s|$-periodic. Moreover, since $|s| \geq |r|$, we have $s = r^q r_{[1,t]}$ with $q \in \mathbb{N}$, $q \geq 1$ and $t \in \{0, \dots, |r|-1\}$. By Fine and Wilf's Theorem (see Theorem 8.1.4 in [Lot02]) the word u is therefore also p-periodic with $p = gcd(|r|, |s|)$. Consequently, there is a word v of length p such that $u = v^l v_{[1,i]}$ with $l \geq 1$ and $i \in \{0, \dots, p-1\}$. We also have $r = v^m$ for an integer $m \geq 1$. Therefore, the word u is prefix of vu and does not occur more than twice in vu. So, by definition v is a return word to u and, by hypothesis on the length of r, we have $v = r$ hence $p = |r|$. Thus $s = r^q$ so there are $q+1$ occurrences of u in su (because $u = r^k r_{[1,j]}$). Consequently, s is a return word to u if and only if $s = r$. □

B.1 Computation of $|u_1|$, $|u_2|$, $|v_1|$ and $|v_2|$

First let us compute the length of paths u_1, u_2, v_1 and v_2 in Rauzy graphs as represented in Figure B.1. We let $\mathfrak{k}$ denote the exponent k of Lemma 5.6.4, i.e., $\mathfrak{k}$ is the number of times the circuit $\vartheta_{i_n+1}(1)$ goes through the loop $v_2 u_2$.

B.1.1 Coming from C_1

In the modified graph of graphs (Figure 5.14 on page 172), the unique vertex in C_1 is the vertex 2 and the corresponding graph is represented in Figure B.2.

1. $\gamma_n = [x, y^k zx, (y^{k-1} zx)]$ with $k \geq 2$ coming from the vertex 2. The evolution corresponding to this morphism is represented in Figure A.6(a) (page 187) with v_{i_n+1} corresponding to the right special vertex on the

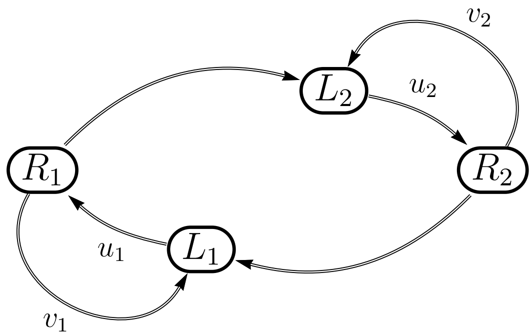


Figure B.1: Rauzy graphs of type 7 or 8

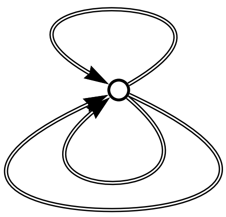


Figure B.2: Graph of type 2

top. We immediately obtain $|v_1| = |v_2| = 1$, $|u_1| = |\gamma_{[0,n-1]}(x)| - 1$, $|u_2| = |\gamma_{[0,n-1]}(y)| - 1$ and $\mathfrak{k} = k - 1$.

2. $\gamma_n = [x, zy^k x, (zy^{k-1}x)]$ with $k \geq 2$ coming from the vertex 2. The evolution corresponding to this morphism is represented in Figure A.6(a) (page 187) with v_{i_n+1} corresponding to the right special vertex at the bottom. We immediately obtain $|v_1| = |v_2| = 1$, $|u_1| = |\gamma_{[0,n-1]}(x)| - 1$, $|u_2| = |\gamma_{[0,n-1]}(y)| - 1$ and $\mathfrak{k} = k - 1$.

3. $\gamma_n = [x, (yz)^k x, ((yz)^{k-1}x)]$ with $k \geq 2$ coming from the vertex 2. The evolution corresponding to this morphism is represented in Figure A.6(b) (page 187) with v_{i_n+1} corresponding to the right special vertex on the top. We immediately obtain $|v_1| = |v_2| = 1$, $|u_1| = |\gamma_{[0,n-1]}(x)| - 1$, $|u_2| = |\gamma_{[0,n-1]}(yz)| - 1$ and $\mathfrak{k} = k - 1$.

4. $\gamma_n = [xy, z^k xy, (z^{k-1}xy)]$ with $k \geq 2$ coming from the vertex 2. The evolution corresponding to this morphism is represented in Figure A.6(b) (page 187) with v_{i_n+1} corresponding to the right special vertex at the bottom. We immediately obtain $|v_1| = |v_2| = 1$, $|u_1| = |\gamma_{[0,n-1]}(xy)| - 1$, $|u_2| = |\gamma_{[0,n-1]}(z)| - 1$ and $\mathfrak{k} = k - 1$.

5. $\gamma_n = [x, (yz)^k yx, ((yz)^{k-1}yx)]$ with $k \geq 1$ coming from the vertex 2. The evolution corresponding to this morphism is represented in Fig-

ure A.6(c) (page 187) with v_{i_n+1} corresponding to the right special vertex on the top. We immediately obtain $|v_1| = 1$, $|v_2| = |\gamma_{[0,n-1]}(z)| + 1$, $|u_1| = |\gamma_{[0,n-1]}(x)| - 1$, $|u_2| = |\gamma_{[0,n-1]}(y)| - 1$ and $\mathfrak{k} = k$.

6. $\gamma_n = [xy, z^k y, (z^{k-1}y)]$ with $k \geq 2$ coming from the vertex 2. The evolution corresponding to this morphism is represented in Figure A.6(c) (page 187) with v_{i_n+1} corresponding to the right special vertex at the bottom. We immediately obtain $|v_1| = |\gamma_{[0,n-1]}(x)| + 1$, $|v_2| = 1$, $|u_1| = |\gamma_{[0,n-1]}(y)| - 1$, $|u_2| = |\gamma_{[0,n-1]}(z)| - 1$ and $\mathfrak{k} = k - 1$.

7. $\gamma_n = [z^\ell x, yz^k x, yz^{k-1}x]$ with $k - 1 > \ell \geq 1$ coming from the vertex 2. The sequence of evolutions corresponding to that morphisms is the following. First, the graph evolves to a graph of type 4 as in Figure A.5(c) (page 186) such that $\vartheta_{i_n+1}(0)$ and $\vartheta_{i_n+1}(1)$ go respectively $k-1$ and $\ell - 1$ times through the loop. Then, the graph becomes a graph as in Figure A.9 and it evolves $\ell - 1$ times as represented in Figure A.10(c). Finally, it evolves to a a graph of type 7 or 8 as in Figure A.10(e). It is obviously seen that we have $|v_2| = 1$, $|u_2| = |\gamma_{[0,n-1]}(z)| - 1$ and $|u_1| + |v_1| = |\gamma_{[0,n-1]}(z^\ell x)|$. Moreover, the path in Figure A.9 that will become u_1 corresponds to the segment which is not curved. After the first evolution (from 2 to 4), this path has for length $|\gamma_{[0,n-1]}(z)|$ (check in Figure A.5(c)) and at each evolution to a graph of type 4 (as in Figure A.10(c)), its length increases by $|\gamma_{[0,n-1]}(z)|$. With the last evolution, we obtain $|u_1| = \ell|\gamma_{[0,n-1]}(z)| + 1$. Finally we can check that $\mathfrak{k} = k - \ell - 1$.

8. $\gamma_n = [yz^\ell x, z^k x, z^{k-1}x]$ with $k - 1 > \ell \geq 1$ coming from the vertex 2. The computation is the same as for the previous morphism. In this case we obtain $|v_2| = 1$, $|u_2| = |\gamma_{[0,n-1]}(z)| - 1$, $|u_1| + |v_1| = |\gamma_{[0,n-1]}(yz^\ell x)|$, $|u_1| = |\gamma_{[0,n-1]}(y)| + \ell|\gamma_{[0,n-1]}(z)| + 1$ and $\mathfrak{k} = k - \ell - 1$.

9. $\gamma_n = [y(xy)^\ell z, (xy)^k z, (xy)^{k-1}z]$ with $k - 1 > \ell \geq 1$ coming from the vertex 2. The sequence of evolutions corresponding to that morphisms is the following. First, the graph evolves to a graph of type 10 as in Figure A.6(d) (page 187) such that $\vartheta_{i_n+1}(0)$ and $\vartheta_{i_n+1}(1)$ go respectively $k-1$ and ℓ times through the loop. Then, the graph becomes a graph as in Figure A.21 and it evolves 2ℓ times as represented in Figure A.22(c). Finally, it evolves to a a graph of type 7 or 8 as in Figure A.22(b). It is obviously seen that we have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$ and $|u_2| + |v_2| = |\gamma_{[0,n-1]}(xy)|$. In Figure A.21, the path that will become u_1 is the segment from the bispecial vertex to the right special vertex. Once the graph has evolved as in Figure A.6(d), it has for length $|\gamma_{[0,n-1]}(z)|$ and

we can see in Figure A.22(c) that, during the 2ℓ evolutions to graphs of type 10, it keeps the same length. With the final evolution as in Figure A.22(b), we obtain $|u_1| = |\gamma_{[0,n-1]}(z) - 1|$. For $|u_2|$ and $|v_2|$, we see in Figure A.21 that the path that will become u_2 is the path from the left special vertex to the bispecial vertex. Once the graph has evolved as in Figure A.6(d), we also see that it has for length $|\gamma_{[0,n-1]}(x)|$. Then, when the graph evolves as in Figure A.22(c), we see that the path that will become u_2 and v_2 always keep the same length but are exchanged at each time. However, since this evolution occurs 2ℓ times, we obtain (with the last evolution) $|u_2| = |\gamma_{[0,n-1]}(x) - 1|$. We finally have $\mathfrak{k} = k - \ell - 1$.

10. $\gamma_n = [(xy)^k z, y(xy)^\ell z, y(xy)^{\ell-1} z]$ with $\ell > k \geq 1$ coming from the vertex 2. The computation is the same as for the previous morphism. We still have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$, $|u_2| + |v_2| = |\gamma_{[0,n-1]}(xy)|$ and $|u_1| = |\gamma_{[0,n-1]}(z) - 1|$. However, once the graph has evolved as in Figure A.6(d), it evolves an odd number of times as in Figure A.22(c) $(2(k-1)+1$ times). Consequently we have $|v_2| = |\gamma_{[0,n-1]}(x) - 1|$ instead of $|u_2|$. We also have $\mathfrak{k} = \ell - k$.

B.1.2 Coming from C_2

For that kind of evolutions, we need to know the length of the path from the left special vertex to the right special vertex in Figure A.7. Indeed, for instance in Figure A.8(e), we see that this path will become either u_1 or u_2, depending on the choice of the starting vertex v_{i_n+1}. This is achieved by the following lemma.

Lemma B.1.1. *Let G_{i_n} be a Rauzy graph of type 3 and let $\gamma_{[0,n-1]}$ be the morphism coding the evolution from G_0 to G_{i_n}. Suppose that $\{x, y, z\} = \{0, 1, 2\}$ and that $\vartheta_{i_n}(x)$ is the top loop in Figure A.7. Let also M be the length of the longest i_{n+1}-circuit. If i and j are such that $\min\{|\gamma_{[0,n-1]}(x^i)|, |\gamma_{[0,n-1]}(y^j)|\} \geq 2M$, then the path from the left special vertex to the bispecial vertex has for length*

$$|\mathrm{CS}(\gamma_{[0,n-1]}(y), \gamma_{[0,n-1]}(z))| - |\mathrm{CS}(\gamma_{[0,n-1]}(x^i), \gamma_{[0,n-1]}(y^j))|.$$

Proof. Indeed, by Proposition 1.5.5 (page 37) we immediately deduce that the length of the path from the left special vertex to the bispecial vertex is

$$|\mathrm{CS}(\gamma_{[0,n-1]}(y), \gamma_{[0,n-1]}(z))| - i_n.$$

Consequently, it suffices to prove that $i_n = |\mathrm{CS}(\gamma_{[0,n-1]}(x^i), \gamma_{[0,n-1]}(y^j))|$. By Lemma B.0.2 we know that $2M$ is greater than i_n and that so are $|\gamma_{[0,n-1]}(x^i))|$ and $|\gamma_{[0,n-1]}(y^j))|$. Consequently, Proposition 1.5.5 implies that both $\gamma_{[0,n-1]}(x^i))$ and $\gamma_{[0,n-1]}(y^j))$ admit the bispecial vertex B as a suffix. Moreover, it is easily seen that if they have a longer common suffix, B would not be bispecial so the result holds. □

In this section, we let q denote the path from the left special vertex to the bispecial vertex in Figure A.7.

1. $\gamma_n = [i, xy^k i, xy^{k-1} i]$ with $k \geq 1$ coming from the vertex V_i, $i \in \{0, 1, 2\}$. The evolution corresponding to that morphism is represented in Figure A.8(e) with vertex v_{i_n+1} corresponding to the right special vertex on the top. In that case we immediately have $|u_1| = |\gamma_{[0,n-1]}(i)| - 1$, $|v_1| = 1$, $|u_2| + |v_2| = |\gamma_{[0,n-1]}(y)|$ and $|u_2| = |q| - 1$. We also have $\mathfrak{k} = k$.

2. $\gamma_n = [x, i^k y, i^{k-1} y]$ with $k \geq 2$ coming from the vertex V_i, $i \in \{0, 1, 2\}$. The evolution corresponding to that morphism is represented in Figure A.8(e) with vertex v_{i_n+1} corresponding to the right special vertex at the bottom. In that case we immediately have $|u_2| = |\gamma_{[0,n-1](i)}| - 1$, $|v_2| = 1$, $|u_1| + |v_1| = |\gamma_{[0,n-1]}(x)|$ and $|u_1| = |q| - 1$. We also have $\mathfrak{k} = k - 1$.

3. $\gamma_n = [xy^\ell i, y^k i, y^{k-1} i]$ with $k - 1 > \ell \geq 0$ coming from the vertex V_i, $i \in \{0, 1, 2\}$. The sequence of evolutions corresponding to that morphism is the following. First the graph evolves to a graph of type 10 as in Figure A.8(f) with starting vertex corresponding to the right special vertex on the top. Then, the graph becomes a graph as in Figure A.21 and evolves 2ℓ times to graphs of type 10 as in Figure A.22(c). Finally, the graph evolves as in Figure A.22(b). For this morphism, we directly see that $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$ and that $|u_2| + |v_2| = |\gamma_{[0,n-1]}(y)|$. We also see in Figure A.21 that the path that will become u_2 is the path from the left special vertex to the bispecial vertex. Once the graph has evolved as in Figure A.8(f), we see that this path has for length $|\gamma_{[0,n-1]}(y)| - |q| - 1$. Then, we see that its length is unchanged after 2 evolutions as in Figure A.22(c) (such an evolution exchanged the curved part of the loop in Figure A.21 with the other part). Consequently, we obtain $|u_2| = |\gamma_{[0,n-1]}(y)| - |q| - 1$. Next, in Figure A.21 we see that the path that will become u_1 is the segment from the bisepcial vertex to the right special vertex. Once the graph has evolved as in Figure A.8(f), we see that it has for length $|\gamma_{[0,n-1]}(i)|$. We also see in Figure A.22(c) that it keeps the same length while these 2ℓ evolutions. While the last

evolution as in Figure A.22(b), we have $|u_1| = |\gamma_{[0,n-1]}(i)| - 1$. Finally, we have $\mathfrak{k} = k - \ell - 1$.

4. $\gamma_n = [y^k i, xy^\ell i, xy^{\ell-1} i]$ with $\ell > k \geq 1$ coming from the vertex V_i, $i \in \{0,1,2\}$. The computation is the same as for the previous morphism. In this case we still have $|u_1|+|v_1| = |\gamma_{[0,n]}(0)|$, $|u_2|+|v_2| = |\gamma_{[0,n-1]}(y)|$ and $|u_1| = |\gamma_{[0,n-1]}(i)| - 1$. However, in this case the graph evolves an odd number of times as in Figure A.22(c) ($2(k-1)+1$ times) so we have $|v_2| = |\gamma_{[0,n-1]}(y)| - |q| - 1$ instead of $|u_2|$. We also have $\mathfrak{k} = \ell - k$.

B.1.3 Coming from C_3

1. $\gamma_n = [0, x^k y0, x^{k-1} y0]$ with $k \geq 1$ coming from the vertex $4B$. The evolution corresponding to that morphism is represented in Figure A.10(e). In this case we immediately obtain $|u_1| = |\gamma_{[0,n-1]}(0) - 1|$, $|v_1| = 1$, $|u_2| + |v_2| = |\gamma_{[0,n-1]}(x)|$, $|u_2| = |\mathrm{CP}(\gamma_{[0,n-1]}(x), \gamma_{[0,n-1]}(y))| - 1$ and $\mathfrak{k} = k$.

2. $\gamma_n = [x^\ell y, 0x^k y, 0x^{k-1} y]$ with $k - 1 > \ell \geq 0$ coming from the vertex $4B$. The sequence of evolutions corresponding to that morphism is the following: first the graph evolves to graph of type 4 as in Figure A.10(d). Then it becomes a graph as in Figure A.9 such that the starting vertex is not the bispecial vertex. It then evolves ℓ times as in Figure A.10(c) and finally evolves as in Figure A.10(e). It is obviously seen that we have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$, $|u_2| = |\gamma_{[0,n-1]}(x)| - 1$ and that $|v_2| = 1$. We also see that the path in Figure A.9 that will become u_1 is the segment from the bispecial vertex to the right special vertex. We see in Figure A.10(c) that, during this evolution, it always keeps the same length. So, its has the same length than the path in Figure A.10(d) from the leftmost right special vertex to the right special vertex on the top. This path has for length $|\gamma_{[0,n-1]}(y)| - |\mathrm{CP}(\gamma_{[0,n-1]}(x), \gamma_{[0,n-1]}(y))|$. With the last evolution (as in Figure A.10(e)), we finally obtain $|u_1| = |\gamma_{[0,n-1]}(y)| - |\mathrm{CP}(\gamma_{[0,n-1]}(x), \gamma_{[0,n-1]}(y))| - 1$. We also have $\mathfrak{k} = k - 1 - \ell$.

3. $\gamma_n = [0x^\ell y, x^k y, x^{k-1} y]$ with $k - 1 > \ell \geq 0$ coming from the vertex $4B$. The computation and the lengths are exactly the same as for the previous morphism.

4. $\gamma_n = [(x0)^\ell y, 0(x0)^k y, 0(x0)^{k-1} y]$ with $k > \ell \geq 0$ coming from the vertex $4B$. The sequence of evolutions corresponding to that morphism is the following. First the graph evolves to a graph of type 10 as in Figure A.10(f) and becomes a graph as in Figure A.21 such that the

starting vertex is not the bispecial one. Then, the graph evolves 2ℓ times as in Figure A.22(c) and it finally evolves as in Figure A.22(b). We immediately have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$ and $|u_2| + |v_2| = |\gamma_{[0,n-1]}(x0)|$. Moreover, we see that the path in Figure A.21 that will become u_1 is the segment from the bispecial vertex to the right special vertex. Once the graph has evolved as in Figure A.10(f), we see that this path has for length $|\mathrm{CP}(\gamma_{[0,n-1]}(x), \gamma_{[0,n-1]}(y))|$. Then, we see in Figure A.22(c) that after 2 such evolutions, this path still have the same length (the two segments starting from the right special vertex which is not bispecial get simply exchanged). Consequently, it still have the same length after the 2ℓ evolutions to graphs of type 10. With the last evolution as in Figure A.22(b) we obtain $|u_1| = |\mathrm{CP}(\gamma_{[0,n-1]}(x), \gamma_{[0,n-1]}(y))| - 1$. We see that the paths in Figure A.21 that will become u_2 and v_2 are respectively the path q from the left special vertex to the bispecial vertex and the path q' from the bispecial vertex to the left special vertex. Once the graph has evolved as in Figure A.10(f), the path that will become q has for length $|\gamma_{[0,n-1]}(0)|$. Then, at each evolution as in Figure A.22(c), q and q' are exchanged. As there is an even number of such evolutions, we finally get (after the last evolution as in Figure A.22(b)) $|u_2| = |\gamma_{[0,n-1]}(0)| - 1$. We also have $\mathfrak{k} = k - \ell$.

5. $\gamma_n = [0(x0)^k y, (x0)^\ell y, (x0)^{\ell-1} y]$ with $\ell - 1 > k \geq 0$ coming from the vertex $4B$. The computation is the same as for the previous morphism. In this case we still have $|u_1|+|v_1| = |\gamma_{[0,n]}(0)|$, $|u_2|+|v_2| = |\gamma_{[0,n-1]}(x0)|$ and $|u_1| = |\mathrm{CP}(\gamma_{[0,n-1]}(x), \gamma_{[0,n-1]}(y))| - 1$. For u_2, in this case the graph evolves an odd number of times as in Figure A.22(c) so we have $|v_2| = |\gamma_{[0,n-1]}(0)| - 1$ instead of $|u_2|$. We also have $\mathfrak{k} = \ell - k - 1$.

B.1.4 Coming from C_4

To compute lengths in this component, we have to be careful with the vertex 5/6. Indeed, this vertex corresponds to the evolution from a graph of type 5 or 6 depending on the length of p_1 and p_2 in Figure 5.7(a) (page 153). To clearly explain how graphs evolve and how we compute lengths, we will always consider that the starting graph is of type 6. The reader is invited to check that all computations also hold when the graph is of type 5.

In the computations given below, we sometimes need to know the order of the starting Rauzy graph when it is of type 10. For this type of graph, we also need to know the length of the simple path from the left special vertex to the bispecial vertex. These information are given in the following lemma whose proof is similar to the proof of Lemma B.1.1 and left to the reader.

Lemma B.1.2. *Let G_{i_n} be a Rauzy graph of type 10 as in Figure A.21. Let $\gamma_{[0,n-1]}$ be the morphism coding the evolution from G_0 to G_{i_n} and suppose that v_{i_n} is the bispecial vertex. If $x \in \{0,1,2\}$ is such that $|\vartheta_{i_n}(x)| = \max\{|\vartheta_{i_n}(i)| \mid i \in \{0,1,2\}\}$ and if l_0, l_1 and l_2 are the smallest positive integers such that*

$$\min\{l_i|\gamma_{[0,n-1]}(i)| \mid i \in \{0,1,2\}\} \geq 2|\gamma_{[0,n-1]}(x)|,$$

then we have

$$i_n = \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(1^{l_1}), \gamma_{[0,n-1]}(2^{l_2})\right)\right|.$$

Moreover, the simple path from the left special vertex to the bispecial vertex in G_{i_n} has for length

$$\left|\mathrm{CS}\left(\gamma_{[0,n-1]}(0^{l_0}), \gamma_{[0,n-1]}(1^{l_1})\right)\right| - i_n.$$

Now let us compute the lengths $|u_1|$, $|u_2|$, $|v_1|$ and $|v_2|$.

1. $\gamma_n = [x, y^k x, y^{k-1}x]$ with $k \geq 2$ coming from the vertex 1 or from the vertex 5/6. The evolutions corresponding to that morphism are represented in Figure A.2(b) and in Figure A.14(b). We can easily see that $|u_1| = |\gamma_{[0,n-1]}(x)| - 1$, $|v_1| = 1$, $|u_2| = |\gamma_{[0,n-1]}(y)| - 1$ and $|v_2| = 1$. We also have $\mathfrak{k} = k - 1$.

2. $\gamma_n = [1, 0^k 2, (0^{k-1}2)]$ with $k \geq 1$ coming from the vertex 5/6. For this evolution, we directly have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$, $|u_2| + |v_2| = |\gamma_{[0,n-1]}(0)|$, $|u_2| = |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))| - 1$, $|u_1| = |\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))| - 1$ and $\mathfrak{k} = k$.

3. $\gamma_n = [2^\ell 0, 12^k 0, (12^{k-1}0)]$ with $k > \ell \geq 0$ coming from the vertex 5/6. The sequence of evolutions corresponding to that morphism is the following. First the graph evolves to a graph of type 10 as in Figure A.14(d) and becomes a graph as in Figure A.21 such that the starting vertex is not the bispecial one. Then, the graph evolves 2ℓ times as in Figure A.22(c) and it finally evolves as in Figure A.22(b). We immediately have $|u_1|+|v_1| = |\gamma_{[0,n]}(0)|$ and $|u_2|+|v_2| = |\gamma_{[0,n-1]}(2)|$. Moreover, we see that the path in Figure A.21 that will become u_1 is the segment from the bispecial vertex to the right special vertex. Once the graph has evolved as in Figure A.14(d), we see that this path has for length $|\gamma_{[0,n-1]}(0)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(1))|$. Then, we see in Figure A.22(c) that after two such evolutions, this path still have the same length (because with such an evolution, the two segments starting from the right special vertex which is not bispecial simply get exchanged). Consequently, it still have the same length after the 2ℓ evolutions to

graphs of type 10. With the last evolution as in Figure A.22(b) we obtain $|u_1| = |\gamma_{[0,n-1]}(0)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))| - 1$. For u_2 and v_2 we see that the paths in Figure A.21 that will become them are respectively the path q from the left special vertex to the bispecial vertex and the path q' from the bispecial vertex to the left special vertex. Once the graph has evolved as in Figure A.14(d), the path that will become q has for length $|\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))|$. Then, at each evolution as in Figure A.22(c), q and q' are exchanged. Since there are an even number of such evolutions, we finally get (after the last evolution as in Figure A.22(b)) $|u_2| = |\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))| - 1$. We also have $\mathfrak{k} = k - \ell$.

4. $\gamma_n = [12^k0, 2^\ell 0, (2^{\ell-1}0)]$ with $\ell > k + 1 \geq 1$ coming from the vertex 5/6. The computation is the same as for the previous morphism. In this case we still have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$, $|u_2| + |v_2| = |\gamma_{[0,n-1]}(2)|$ and $|u_1| = |\gamma_{[0,n-1]}(0)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))| - 1$. For u_2, in this case the graph evolves an odd number of times as in Figure A.22(c) so we have $|v_2| = |\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(0), \gamma_{[0,n-1]}(2))| - 1$ instead of $|u_2|$. We also have $\mathfrak{k} = \ell - k - 1$.

5. $\gamma_n = [0, 2^k1, 2^{k-1}1]$ with $k \geq 1$ coming from the vertex $10B$. The evolution corresponding to that morphism is represented in Figure A.22(b). We immediately see that $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$, $|u_1| + |v_1| = |\gamma_{[0,n-1]}(2)|$, $|u_2| = |\mathrm{CP}(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2))| - 1$. Moreover, by Lemma B.1.2 we have (with the same notation) $|u_1| = \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(0^{l_0}), \gamma_{[0,n-1]}(1^{l_1})\right)\right| - \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(1^{l_1}), \gamma_{[0,n-1]}(2^{l_2})\right)\right| - 1$. We also have $\mathfrak{k} = k$.

6. $\gamma_n = [1^\ell 2, 01^k2, (01^{k-1}2)]$ with $k > \ell \geq 0$ coming from the vertex $10B$. The sequence of evolutions corresponding to that morphism is the following. First the graph evolves to a graph of type 10 as in Figure A.22(d) and becomes a graph as in Figure A.21 such that the starting vertex is not the bispecial one. Then, the graph evolves 2ℓ times as in Figure A.22(c) and it finally evolves as in Figure A.22(b). We immediately have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$ and $|u_2| + |v_2| = |\gamma_{[0,n-1]}(1)|$. Moreover, we see that the path in Figure A.21 that will become u_1 is the segment from the bispecial vertex to the right special vertex. Once the graph has evolved as in Figure A.22(d), we see that this path has for length $|\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2))|$. Then, we see in Figure A.22(c) that after two such evolutions, this path still have the same length (because with such an evolution, the two segments starting from the right special vertex which is not bispecial simply get exchanged). Consequently, it still has the same length after the 2ℓ evolutions to

graphs of type 10. With the last evolution as in Figure A.22(b) we obtain $|u_1| = |\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2))| - 1$. We see in Figure A.21 that the paths that will become u_2 and v_2 are respectively the path q from the left special vertex to the bispecial vertex and the path q' from the bispecial vertex to the left special vertex. Once the graph has evolved as in Figure A.22(d), we know from Lemma B.1.2 that q has for length $\left|\mathrm{CS}\left(\gamma_{[0,n-1]}(0^{l_0}), \gamma_{[0,n-1]}(1^{l_1})\right)\right| - \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(1^{l_1}), \gamma_{[0,n-1]}(2^{l_2})\right)\right|$. Then, at each evolution as in Figure A.22(c), q and q' are exchanged. As there are an even number of such evolutions, we finally get (after the last evolution as in Figure A.22(b)) $|u_2| = \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(0^{l_0}), \gamma_{[0,n-1]}(1^{l_1})\right)\right| - \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(1^{l_1}), \gamma_{[0,n-1]}(2^{l_2})\right)\right| - 1$. We also have $\mathfrak{k} = k - \ell$.

7. $\gamma_n = [01^k2, 1^\ell 2, (1^{\ell-1}2)]$ with $\ell > k+1 \geq 1$ coming from the vertex $10B$. The computation is the same as for the previous morphism. In this case we still have $|u_1| + |v_1| = |\gamma_{[0,n]}(0)|$, $|u_2| + |v_2| = |\gamma_{[0,n-1]}(1)|$ and $|u_1| = |\gamma_{[0,n-1]}(2)| - |\mathrm{CP}(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2))| - 1$. For u_2, in this case the graph evolves an odd number of times as in Figure A.22(c) so we have $|v_2| = \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(0^{l_0}), \gamma_{[0,n-1]}(1^{l_1})\right)\right| - \left|\mathrm{CS}\left(\gamma_{[0,n-1]}(1^{l_1}), \gamma_{[0,n-1]}(2^{l_2})\right)\right| - 1$ instead of $|u_2|$. We also have $\mathfrak{k} = \ell - k - 1$.

B.2 Computation of $|p_1|$ and $|p_2|$

The aim of this section is to compute the length of the paths p_1 and p_2 of Figure 5.7(a) when evolving to such a graph, i.e., when considering an edge to the vertex 5/6 in Figure 5.12. These lengths do not only depend on the last morphism applied but on a finite number of morphisms. First, the next lemma shows how to compute these lengths when passing through the vertex 7/8 in Figure 5.12. The other cases will be particular cases of this one. Indeed, morphisms labelling the loop on vertex 5/6 in Figure 5.12 are simply compositions of the morphism $[1, 0^k2, 0^{k-1}2]$ (labelling the edge from 5/6 to 7/8) with a morphism in $\{[0x, y, 0y], [x, 0y, y]\}$ (labelling the edge from 7/8 to 5/6). In other words, it simply corresponds to the case $h = 0$ in Lemma B.2.1 below. For morphisms labelling the edge from $10B$ to 5/6 in Figure 5.12, the reasoning is the same but this time, the morphisms labelling the edge from $10B$ to 5/6 are compositions of the morphism $[0, 2^k1, 2^{k-1}1]$ (labelling the edge from $10B$ to 7/8) with a morphism in $\{[0x, y, 0y], [x, 0y, y]\}$ (labelling the edge from 7/8 to 5/6).

Lemma B.2.1. *Let G_{i_n+1} be a Rauzy graph as represented in Figure 5.7(b) (page 153) and let $\gamma_{[0,n-1]}$ be the morphism coding the evolution from G_0 to G_{i_n+1} (so to $G_{i_{n+1}}$). Suppose that v_{i_n+1} corresponds to the vertex R_1 in*

Figure 5.7(b) and that the circuit $\vartheta_{i_n+1}(1)$ *goes exactly* k *times through the loop* v_2u_2*.*

Let ℓ *be the unique integer such that*

$$|u_1| + (\ell-1)(|u_1|+|v_1|) < |u_2| + (k-1)(|u_2|+|v_2|) \leq |u_1| + \ell(|u_1|+|v_1|)$$

and let h *be the greatest integer such that for all* $i \in \{0,\dots,h-1\}$, $\gamma_{n+i} \in= [0,10,20]$*. Suppose that* γ_{n+h} *labels the edge from 7/8 to 5/6 (so belongs to* $\{[0x,y,(0y)],[x,0y,(y)] \mid \{x,y\}=\{1,2\}\}$*), then* $G_{i_{n+h}+1}$ *is a graph as represented in Figure 5.7(a) (page 153) and the lengths of* p_1 *and* p_2 *are as follows.*

If $h < \ell$*, we have*

$$\begin{aligned}
|p_1| &= \left|\mathrm{CP}\left(\gamma_{[0,n-1]}(1),\gamma_{[0,n-1]}(2)\right)\right| - (k-1-k')(|u_2|+|v_2|) \\
&\quad -(|u_2| + k'(|u_2|+|v_2|) - (|u_1| + h(|u_1|+|v_1|))) - 1 \\
|p_2| &= |\gamma_{[0,n-1]}(2)| - \left|\mathrm{CP}\left(\gamma_{[0,n-1]}(1),\gamma_{[0,n-1]}(2)\right)\right| - 1
\end{aligned}$$

and if $h \geq \ell$*, we have*

$$\begin{aligned}
|p_1| &= \left|\mathrm{CP}(\gamma_{[0,n-1]}(1),\gamma_{[0,n-1]}(2))\right| - 1 \\
|p_2| &= \left|\gamma_{[0,n-1]}(2)\right| - \left|\mathrm{CP}(\gamma_{[0,n-1]}(1),\gamma_{[0,n-1]}(2))\right| \\
&\quad -(|u_1| + \ell(|u_1|+|v_1|) - (|u_2| + (k-1)(|u_2|+|v_2|))) - 1.
\end{aligned}$$

Proof. Let us recall notation introduced in the proof of Lemma 5.6.4. For all non-negative integers i and j, $B_1(i)$ and $B_2(j)$ are respectively the words $\lambda(u_1(v_1u_1)^i)$ and $\lambda(u_2(v_2u_2)^j)$. For $j \in \{0,\dots,k-1\}$, $B_2(j)$ is a bispecial vertex in $G_{|B_2(j)|}$ and $B_2(k)$ does not belong to the language of the considered subshift. Also, for all non-negative integers i, if $B_1(i)$ is in the language of the considered subshift, then it is a bispecial vertex in $G_{|B_1(i)|}$.

Now let us determine the sequence of evolutions corresponding to the sequence of morphisms $(\gamma_m)_{n\leq m\leq n+h}$. The graph G_t will evolve to a graph of type 7 or 8 depending on $|u_1|$ and $|v_1|$. Thanks to Lemma 5.4.1 we can suppose without loss of generality that it evolves to a graph of type 7.

Let us start studying the behaviours of vertices $B_2(j)$. The hypothesis on $\vartheta_t(1)$ implies that for all $j \in \{0,\dots,k-2\}$, $B_2(j)$ will explode as represented in Figure 5.8(b) (page 156). Then, the hypothesis on γ_{n+h} implies that $B_2(k-1)$ will explode as in Figure 5.8(d) (because there are three distinct letters in its images).

Now let us study the behaviours of vertices $B_1(i)$. By constructions of the morphisms γ_m, for $i \in \{0,\dots,h\}$, the hypothesis on γ_{n+i} implies that $B_1(i)$ is a bispecial vertex of the subshift and that for $i \in \{0,\dots,h-1\}$,

$B_1(i)$ explodes like $B_2(j)$ does in Figure 5.8(b). However, the hypothesis on ℓ implies that at most the first ℓ vertices among $B_1(0), \dots, B_1(h)$ can explode strictly before that $B_2(k-1)$ explodes. Also, the hypothesis on γ_{n+h} implies that $B_1(h)$ explodes like $B_2(j)$ does in Figure 5.8(d).

Now let us exactly describe the sequence of evolution depending on h and ℓ.

When $h < \ell$, vertex $B_1(h)$ explodes before $B_2(k-1)$. Let k' be the smallest integer such that $|B_2(k')| \geq |B_1(h)|$. We obviously have $k' \leq k-1$. Then, all bispecial vertices $B_1(0), \dots, B_1(h-1), B_2(0), \dots, B_2(k'-1)$ explode and make the graph keeping type 7 or 8. Then, the explosion of $B_1(h)$ makes the graph $G_{|B_1(h)|}$ evolve as represented in[1] Figure A.16(c) (page 194) so the graph evolves to a graph of type 9 as in Figure A.19. Then, the explosions of $B_2(k'), \dots, B_2(k-2)$ make the graph evolve as in Figure A.20(c). Finally, the explosion of $B_2(k-1)$ makes the graph evolve as in Figure A.20(b).

When $h \geq \ell$, it means that vertex $B_1(h)$ will not explode strictly before that $B_2(k-1)$ explodes. In that case, Lemma 5.4.1 allows us to suppose that $B_1(\ell)$ explodes strictly after that $B_2(k-1)$ has exploded and, as a consequence, that so does $B_1(h)$. Consequently, vertices $B_1(0), \dots, B_1(\ell-1), B_2(0), \dots, B_2(k-2)$ explode and make graphs keeping type 7 or 8. Then, the explosion of $B_2(k-1)$ makes the graph $G_{|B_2(k-1)|}$ evolve as in Figure A.16(c) so it evolves to a graph of type 9 as in Figure A.19. Then, vertices $B_1(\ell), \dots, B_1(h-1)$ make graphs keeping type 9 as in Figure A.20(c). Finally, the explosion of $B_1(h)$ makes the graph $G_{|B_1(h)|}$ evolve as in Figure A.20(b).

Now let us compute $|p_1|$ and $|p_2|$. In Figure A.20(b), we see that the two paths in Figure A.19 that will become p_1 and p_2 are the path from the left special vertex to the bispecial vertex and the path from the bispecial vertex to the right special vertex[2]. In Figure A.20(c), we also see that, while graphs keep being graphs of type 9, these paths always have the same length (because, in Figure A.20(c), they are paths from a left special vertex to a left special vertex and from a right special vertex to a right special vertex to a right special vertex). Consequently, the lengths of the paths in Figure A.19 that will become p_1 and p_2 can be computed in the evolution from the last graph of type 7 to the first graph of type 9, i.e., in the evolution of $G_{|B_1(h)|}$ when $h < \ell$ and of $G_{|B_2(k-1)|}$ otherwise.

Suppose that h is smaller than ℓ. It means that $G_{|B_1(h)|}$ is a graph of type 7 as represented in Figure A.15 $v_{|B_1(h)|} = B_1(h)$ is the bispecial vertex. It is easily seen that in Figure A.15, the path from the left special vertex to the

[1]Thanks to Lemma 5.4.1, we can still suppose that the graph os of type 7.

[2]Which one is p_1 depends on the starting vertex for the circuits.

right special vertex has for length

$$|B_2(k')| - |B_1(h)| = |u_2| + k'(|u_2| + |v_2|) - (|u_1| + h(|u_1| + |v_1|)).$$

We also see in Figure A.16(c) that the path in $G_{|B_1(h)|}$ that will become p_1 (resp. that will become p_2) is the path from $B_1(h)$ to the left special vertex (resp. from the right special vertex to $B_1(h)$). Consequently, we directly have

$$|p_2| = |\gamma_{[0,n-1]}(2)| - \left|\mathrm{CP}\left(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2)\right)\right| - 1.$$

To compute, $|p_1|$, we can notice that the longest common prefix of $\vartheta_t(1)$ and $\vartheta_t(2)$ has the same length as the path starting from $B_1(h)$, going $k-1-k'$ times through the loop with label $\lambda_L(v_2u_2)$ and ending in the right special vertex which is not $B_1(h)$. Consequently, the path from $B_1(h)$ to the left special vertex has for length

$$\left|\mathrm{CP}\left(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2)\right)\right| - (k-1-k')(|u_2| + |v_2|) - (|B_2(k')| - |B_1(h)|)$$

so

$$\begin{aligned} |p_1| \;&=\; \left|\mathrm{CP}\left(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2)\right)\right| - (k-1-k')(|u_2| + |v_2|) \\ &\quad -(|u_2| + k'(|u_2| + |v_2|) - (|u_1| + h(|u_1| + |v_1|))) - 1 \end{aligned}$$

Now suppose that h is not smaller than ℓ. It means that $G_{|B_2(k-1)|}$ is a graph of type 7 as represented in Figure A.15 $v_{|B_2(k-1)|}$ is not the bispecial vertex. It is easily seen that in Figure A.15, the path from the left special vertex to the right special vertex has for length

$$|B_1(\ell)| - |B_2(k-1)| = |u_1| + \ell(|u_1| + |v_1|) - (|u_2| + (k-1)(|u_2| + |v_2|)).$$

From what precedes, we know that the paths in $G_{|B_2(k-1)|}$ that will become p_1 and p_2 are respectively the segment from $v_{|B_2(k-1)|}$ to $B_2(k-1)$ and the path from $B_2(k-1)$ to the left special vertex. Consequently, we directly have

$$|p_1| = \left|\mathrm{CP}(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2))\right| - 1$$

and

$$\begin{aligned} |p_2| \;&=\; \left|\gamma_{[0,n-1]}(2)\right| - \left|\mathrm{CP}(\gamma_{[0,n-1]}(1), \gamma_{[0,n-1]}(2))\right| \\ &\quad -(|u_1| + \ell(|u_1| + |v_1|) - (|u_2| + (k-1)(|u_2| + |v_2|))) - 1. \end{aligned}$$

□

Appendix C

Proof of Lemma 5.6.7

Let us prove the following result which is equivalent to Lemma 5.6.7 but with more details.

Lemma C.0.2. *A sequence of morphisms $(\gamma_n)_{n\geq N}$ labelling an infinite path p in Figure 5.11 is almost primitive if and only if one of the following conditions is satisfied:*

1. *p ultimately stays in vertex 1 and both morphisms $[0,10]$ and $[01,1]$ occur infinitely often in $(\gamma_n)_{n\geq N}$;*
2. *p ultimately stays in vertex $10B$ and for all integers $r \geq N$, $(\gamma)_{n\geq r}$ does not only contain occurrences of $[0,20,1]$, neither of $[01^k2, 1^{k+1}2, 1^k2]$ for $k \in \mathbb{N}$ and is not only composed of finite sub-sequences of morphisms in*
$$\left\{[0,20,1]^{2n}, [02,12,2]^n \mid n \in \mathbb{N} \setminus \{0\}\right\};$$
3. *p ultimately stays in the subgraph $\{1, 7/8\}$, goes through both vertices infinitely often and for all suffixes p' of p starting in vertex $7/8$, the label of p' is not only composed of finite sub-sequences of morphisms in*
$$\big([0,10]^*[0,1][0,10]^*\{[0,1^k0] \mid k \geq 2\}\big) \cup \big([0,10]^*[1,0][01,1]^*\{[1,0^k1] \mid k \geq 2\}\big);$$
4. *p ultimately stays in the subgraph $\{5/6, 7/8\}$, goes through both vertices infinitely often and for all suffixes p' of p starting in vertex $7/8$, the label of p' is not only composed of finite sub-sequences of morphisms in*
$$[0,10,20]^* \{[1,02,2],[01,2,02]\} [1,02,2]$$
and not only composed of finite sub-sequences of morphisms in
$$\{[2,01,1],[1,02,2]\} \left\{[1,0^k2,0^{k-1}2], [12^{k-1}0, 2^\ell 0, 2^{\ell-1}0] \mid \ell > k \geq 1\right\};$$

5. *p ultimately stays in the subgraph* $\{5/6, 7/8, 10B\}$*, goes through the three vertices infinitely often and if* $(q_n)_{n\in\mathbb{N}}$ *(resp.* $(t_n)_{n\in\mathbb{N}}$*) is the sequence of finite sub-paths or p that start and end in* $7/8$ *and go through* $10B$ *(resp. that start and end in* $7/8$ *and do not go through* $10B$*), then for all integers* $r \geq N$*, the following holds true:*

   - *if for all* $n \geq r$*, the label of* q_n *is in*

   $$\{[1,02,2],[01,2,02]\}[1,01,2]\{[0,20,1]^{2n},[02,12,2] \mid n \in \mathbb{N}\}^* \{[2,012,02],[0,20,1][0,21,1]\},$$

   then the sequence $(t_n)_{n\in\mathbb{N}}$ *is infinite and contains infinitely many occurrences of finite paths whose label is not in*

   $$\{[1,02,2],[01,2,02]\}\,[1,02,2];$$

   - *if for all* $n \geq r$*, the label of* q_n *is in*

   $$\{[1,02,2],[2,01,1]\}\{[12^k0,2^{k+1}0,2^k0] \mid k \geq 0\} \{[01^k2,1^{k+1}2,1^k2] \mid k \geq 0\}\{[0,2^k1,2^{k-1}1] \mid k \geq 2\},$$

   then the path p goes infinitely often through the loop on $7/8$ *or, the sequence* $(t_n)_{n\in\mathbb{N}}$ *is infinite and contains infinitely many occurrences of finite paths whose label is not in*

   $$\{[2,01,1],[1,02,2]\} \big\{[1,0^k2,0^{k-1}2],[12^{k-1}0,2^\ell 0,2^{\ell-1}0] \mid \ell > k \geq 1\big\};$$

6. *p contains infinitely many occurrences of sub-paths q that start in* 1 *and end in* $5/6$*.*

Proof. The method to prove this result is to study the almost primitivity in each subgraph of Figure 5.11. Among all these subgraphs, those in which there exist some infinite paths are

$$\{1\},\{7/8\},\{10B\},\{1,7/8\},\{5/6,7/8\}, \{1,5/6,7/8\},\{5/6,7/8,10B\} \text{ and } \{1,5/6,7/8,10B\}.$$

It is easily seen that all valid paths in the subgraph $\{7/8\}$ do not have almost primitive labels. Also, for the subgraphs $\{1\}$, $\{10B\}$, the given conditions of the result are trivially equivalent to the almost primitivity.

Let us study the subgraph $\{1, 7/8\}$. If q is a path starting in vertex 7/8, going through vertex 1, possibly staying in it for a while and then coming back to vertex 7/8, then its label belongs to the set

$$Q = \{[x,y][x,yx],[xy,y] \mid \{x,y\} = \{0,1\}\}\{[0,10],[01,1]\}^* \\ \left\{[0,1^k0,1^{k-1}0],[1,0^k1,0^{k-1}1] \mid k \geq 2\right\}.$$

If p ultimately stays in the subgraph $\{1, 7/8\}$, it means that its label is ultimately composed of finite subsequences of morphisms in that set and of occurrences of the morphism $[0, 10, 20]$ labelling the loop on vertex 7/8. However, morphisms labelling the edge from 7/8 to 1 do not contain the letter 2 in their images. Consequently, the third component of all morphisms can be ignored. Now it can be checked that for all finite sequences of morphisms $\gamma_1 \cdots \gamma_m$ in Q, $\gamma_1 \cdots \gamma_m(1)$ contains some occurrences of both 0 and 1. Since the morphism labelling the loop on 7/8 is $[0, 10]$, the label $(\gamma_n)_{n \geq N}$ of any infinite path p in $\{1, 7/8\}$ is not almost primitive if and only if there is an integer $r \geq N$ such that for all $n \geq r$, $\gamma_r \gamma_{r+1} \cdots \gamma_n(0) = 0$. To conclude the proof for the subgraph $\{1, 7/8\}$, it suffices to notice that the finite sequences of morphisms $\gamma'_1 \cdots \gamma'_m$ in

$$[0,1][0,10]^*[0,1^k0] \cup [1,0][01,1]^*[1,0^k1]$$

are the only ones in Q such that $\gamma'_1 \cdots \gamma'_m(0) = 0$.

Let us study the subgraph $\{5/6, 7/8\}$. For any word u over $\{0, 1, 2\}$ we let $\mathrm{Alph}(u)$ be the smallest lexicographic word over $\{0, 1, 2\}$ such that all letters occurring in u occur in $\mathrm{Alph}(u)$ too. By abuse of notation, for any path q with label $\sigma = \gamma_1 \cdots \gamma_m$ we write

$$\mathrm{Alph}(q) = (\mathrm{Alph}(\sigma(0)), \mathrm{Alph}(\sigma(1)), \mathrm{Alph}(\sigma(2))).$$

It can be algorithmically checked that, if q is a path of length two that starts in 7/8 and goes through 5/6 before coming back to 7/8, then $\mathrm{Alph}(q)$ is one of the following:

(01,12,1)	(01,12,12)	(012,12,12)	(02,12,12)	(02,12,2)
(012,012,012)	(01,012,012)	(02,012,012)	(12,012,012)	(1,012,012)
(2,012,012)	(1,012,01)	(2,012,02)		

Table C.1: List of $\mathrm{Alph}(q)$ for $q = 7/8 \to 5/6 \to 7/8$.

We let Q_1 denote the set of paths q of length 2 that start in 7/8, go through 5/6 and come back to 7/8 and such that $\mathrm{Alph}(q)$ is one of the following:

(012,012,012)	(01,012,012)	(02,012,012)	(12,012,012)
(1,012,012)	(2,012,012)	(1,012,01)	

Obviously, the label $(\gamma_n)_{n\in\mathbb{N}}$ of any infinite path p in the subgraph $\{5/6, 7/8\}$ that contains infinitely many occurrences of sub-paths q in Q_1 is almost primitive. Indeed, if p is a finite path in the subgraph $\{5/6, 7/8\}$ that contains two occurrences of paths in Q_1, then the letter 1 occurs in the three components of $\mathrm{Alph}(p)$ which makes $(\gamma_n)_{n\in\mathbb{N}}$ almost primitive because for all paths q in Q_1, the second component of $\mathrm{Alph}(q)$ contains occurrences of the three letters.

Let us consider an infinite path p such that all sub-paths q of length 2 that start in 7/8 and go through 5/6 do not belong to Q_1, so are such that $\mathrm{Alph}(q)$ is one of the following:

(01,12,1)	(01,12,12)	(012,12,12)
(02,12,12)	(02,12,2)	(2,012,02)

For suc paths q, we can see two problems for the almost primitivity:

- except for paths q such that $\mathrm{Alph}(q) = (2, 012, 02)$, the letter 0 never occurs in the two last components of $\mathrm{Alph}(q)$;

- for paths q such that $\mathrm{Alph}(q) \in \{(02, 12, 2), (2, 012, 02)\}$, the letter 1 never occurs in the first and in the last component of $\mathrm{Alph}(q)$.

Consequently, the following holds true: the label of any infinite path p in $\{5/6, 7/8\}$ such that all sub-paths $q : 7/8 \to 5/6 \to 7/8$ are such that

1. $\mathrm{Alph}(q) \in \{(02, 12, 2), (2, 012, 02)\}$ cannot be almost primitive;

2. $\mathrm{Alph}(q) \in \{(01, 12, 1), (01, 12, 12), (012, 12, 12), (02, 12, 12), (02, 12, 2)\}$ is almost primitive if and only if $\mathrm{Alph}(q)$ is not ultimately $(02, 12, 2)$ and the path p goes infinitely often through the loop on 7/8 (because it is labelled by $[0, 10, 20]$).

One can also check that if there are infinitely many occurrences of paths q and q' in p such that $\mathrm{Alph}(q) = (2, 012, 02)$ and

$$\mathrm{Alph}(q') \in \{(01, 12, 1), (01, 12, 12), (012, 12, 12), (02, 12, 12)\},$$

then the label of p is almost primitive.

To conclude the proof for the subgraph $\{5/6, 7/8\}$, it suffices now to study which labelled paths $q = 7/8 \to 5/6 \to 7/8$ correspond to the "forbidden cases" listed just above. If q is such a path and if γ_1 (resp. γ_2) labels the edge $7/8 \to 5/6$ (resp. $5/6 \to 7/8$), then we have

$$\mathrm{Alph}(q) = (02, 12, 2) \Leftrightarrow \begin{cases} \gamma_1 = [1, 02, 2] \\ \gamma_2 = [1, 02, 2] \end{cases}$$

$$\mathrm{Alph}(q) = (2, 012, 02) \Leftrightarrow \begin{cases} \gamma_1 = [01, 2, 02] \\ \gamma_2 = [1, 02, 2] \end{cases}$$

and

$$\mathrm{Alph}(q) \in \{(01, 12, 1), (01, 12, 12), (012, 12, 12), (02, 12, 12), (02, 12, 2)\}$$

$$\Updownarrow$$

$$\begin{cases} \gamma_1 \in \{[1, 02, 2], [2, 01, 1]\} \\ \gamma_2 \in \{[1, 0^k 2, 0^{k-1}2] \mid k \geq 1\} \cup \{[12^k 0, 2^\ell 0, 2^{\ell-1}0] \mid \ell > k+1 \geq 1\} \end{cases}$$

Let us study the subgraph $\{5/6, 7/8, 10B\}$. As for $\{5/6, 7/8\}$, it can be algorithmically checked that, if q is a finite path in $\{5/6, 7/8, 10B\}$ that starts and ends in 7/8 and that goes through $10B$, then $\mathrm{Alph}(q)$ is one of the following:

(01,012,01)	(01,012,012)	(012,012,012)	(012,12,12)	(02,012,012)
(02,012,02)	(1,012,01)	(1,012,012)	(2,012,012)	(2,012,02)

Table C.2: List of $\mathrm{Alph}(q)$ for $q = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$.

Let us start by determining some non-almost primitive infinite labelled paths. First, it is easily seen that if p_1 is an infinite path in $\{5/6, 7/8, 10B\}$ whose sub-paths $q_{1,1} = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$ are ultimately such that $\mathrm{Alph}(q_{1,1}) \in \{(2, 012, 02), (02, 012, 02)\}$, then the label of p_1 is almost primitive if and only if p_1 contains infinitely many occurrences of sub-paths $q_{1,2} = 7/8 \to 5/6 \to 7/8$ such that[1]

$$\mathrm{Alph}(q_{1,2}) \notin \{(02, 12, 2), (2, 012, 02)\}.$$

Next, one can also see that if p_2 is an infinite path in $\{5/6, 7/8, 10B\}$ whose sub-paths $q_{2,1} = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$ are ultimately such that $\mathrm{Alph}(q_{2,1}) = (012, 12, 12)$, then the label of p_2 is almost primitive

[1]The problem is the same as the one met in the subgraph $\{5/6, 7/8\}$: the letter 1 never occurs in the image of 02.

if and only if p_2 contains infinitely many occurrences of loops $7/8 \to 7/8$ or of sub-paths $q_{2,2} = 7/8 \to 5/6 \to 7/8$ such that[2]

$$\mathrm{Alph}(q_{2,2}) \notin \{(01,12,1),(01,12,12),(012,12,12),(02,12,12),(02,12,2)\}.$$

Now let us show that all other infinite paths p_3 in $\{5/6, 7/8, 10B\}$ that goes infinitely often through the three vertices have an almost primitive label. We can see that in all remaining values of $\mathrm{Alph}(q)$, i.e., for all paths $q = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$ with

$$\mathrm{Alph}(q) \notin \{(2,012,02),(02,012,02),(012,12,12)\},$$

the second component of $\mathrm{Alph}(q)$ is 012. This makes the label of p_3 almost primitive becaus if p' is a finite path in $\{5/6, 7/8, 10B\}$ that contains two occurrences of paths $q = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$ with

$$\mathrm{Alph}(q) \notin \{(2,012,02),(02,012,02),(012,12,12)\},$$

then each component of $\mathrm{Alph}(p')$ contains an occurrence of the letter 1.

To conclude the proof for the subgraph $\{\{5/6, 7/8, 10B\}$, it suffices (like for the subgraph $\{5/6, 7/8\}$) to study which labelled paths $q = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$ correspond to the "forbidden cases", i.e., which ones are such that

$$\mathrm{Alph}(q) \in \{(2,012,02),(02,012,02),(012,12,12)\}.$$

If the label of $q = 7/8 \to 5/6 \to 10B(\to 10B)^* \to 7/8$ is $\gamma_1\gamma_2\cdots\gamma_m$ with $m \geq 3$ such that γ_1 (resp. γ_2, γ_m) labels the edge $7/8 \to 5/6$ (resp. $5/6 \to 10B$, $10B \to 7/8$) and $\gamma_3\cdots\gamma_{m-1}$ labels the loop $10B \to 10B$, then it is not difficult (though a bit long) to check that the following holds true:

$$\mathrm{Alph}(q) \in \{(2,012,02),(02,012,02)\}$$
$$\Updownarrow$$
$$\begin{cases} \gamma_1\gamma_2 \in \{[1,02,2],[01,2,02]\}[1,01,2] \\ \gamma_3\cdots\gamma_{m-2} \in \left\{[0,20,1]^{2n},[02,12,2]^n \mid n \in \mathbb{N}\right\}^* \\ \gamma_m = [2,012,02] \text{ or } (m \geq 4 \text{ and } \gamma_{m-1}\gamma_m = [0,20,1][0,21,1]) \end{cases}$$

and

$$\mathrm{Alph}(q) = [012,12,12]$$
$$\Updownarrow$$
$$\begin{cases} \gamma_1\gamma_2 \in \{[1,02,2],[2,01,2]\}\left\{[12^k0,2^{k+1}0,2^k0] \mid k \geq 0\right\} \\ \gamma_3\cdots\gamma_{m-1} \in \left\{[01^k2,1^{k+1}2,1^k2] \mid k \geq 0\right\} \\ \gamma_m \in \left\{[0,2^k1,2^{k-1}1] \mid k \geq 2\right\} \end{cases}.$$

[2]This is again a problem met in the subgraph $\{5/6, 7/8\}$: the letter 0 never occurs in the image of 12.

To conclude the whole proof, it remains to show that the label of any path that goes infinitely often through the four vertices or that ultimately stays in the subgraph $\{1, 5/6, 7/8\}$ is almost primitive. This can be easily seen: any such path must contain infinitely many occurrences of finite paths $1 \to 7/8 \to 5/6$ and all these paths have a strongly primitive label. $\square$

List of Figures

Bibliography

[AB92] J.-P. Allouche and R. Bacher. Toeplitz sequences, paperfolding, Towers of Hanoi and progression-free sequences of integers. *Enseign. Math. (2)*, 38(3-4):315–327, 1992.

[AB98] P. Alessandri and V. Berthé. Three distance theorems and combinatorics on words. *Enseign. Math. (2)*, 44(1-2):103–132, 1998.

[Ada02] B. Adamczewski. Codages de rotations et phénomènes d'autosimilarité. *J. Théor. Nombres Bordeaux*, 14(2):351–386, 2002.

[Ada05] B. Adamczewski. On powers of words occurring in binary codings of rotations. *Adv. in Appl. Math.*, 34(1):1–29, 2005.

[All94] J.-P. Allouche. Sur la complexité des suites infinies. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2):133–143, 1994. Journées Montoises (Mons, 1992).

[AR91] P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexité $2n+1$. *Bull. Soc. Math. France*, 119:199–215, 1991.

[ARS09] J.-P. Allouche, N. Rampersad, and J. Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theoret. Comput. Sci.*, 410(30-32):2795–2803, 2009.

[AS03] J.-P. Allouche and J. Shallit. *Automatic sequences: Theory, applications, generalizations.* Cambridge University Press, Cambridge, 2003.

[AS07] S. Akiyama and M. Shirasaka. Recursively renewable words and coding of irrational rotations. *J. Math. Soc. Japan*, 59(4):1199–1234, 2007.

[BCF99] V. Berthé, N. Chekhova, and S. Ferenczi. Covering numbers: arithmetics and dynamics for rotations and interval exchanges. *J. Anal. Math.*, 79:1–31, 1999.

[BdLDLZ08] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On different generalizations of episturmian words. *Theoret. Comput. Sci.*, 393(1-3):23–36, 2008.

[Ber01] V. Berthé. Autour du système de numération d'Ostrowski. *Bull. Belg. Math. Soc. Simon Stevin*, 8:209–239, 2001. Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000).

[Ber07] J. Berstel. Sturmian and episturmian words (a survey of some recent results). In *Algebraic informatics*, volume 4728 of *Lecture Notes in Comput. Sci.*, pages 23–47. Springer, Berlin, 2007.

[BHZ06] V. Berthé, C. Holton, and L. Q. Zamboni. Initial powers of Sturmian sequences. *Acta Arith.*, 122:315–347, 2006.

[Bos85] M. Boshernitzan. A unique ergodicity of minimal symbolic flows with linear block growth. *J. Analyse Math.*, 44:77–96, 1984/85.

[BR10] V. Berthé and M. Rigo, editors. *Combinatorics, Automata and Number Theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.

[Bra72] O. Bratteli. Inductive limits of finite dimensional C^*-algebras. *Trans. Amer. Math. Soc.*, 171:195–234, 1972.

[Brl89] S. Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Appl. Math.*, 24(1-3):83–96, 1989. First Montreal Conference on Combinatorics and Computer Science, 1987.

[Cas96] J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in language theory, II (Magdeburg, 1995)*, pages 25–34. World Sci. Publ., River Edge, NJ, 1996.

[Cas97] J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4:67–88, 1997. Journées Montoises (Mons, 1994).

[Cas02] J. Cassaigne. Computing the subword complexity of a s-adic sequence: application to a family of interval translation maps. 2002.

[Cas03] J. Cassaigne. Constructing infinite words of intermediate complexity. In *Developments in language theory*, volume 2450 of *Lecture Notes in Comput. Sci.*, pages 173–184. Springer, Berlin, 2003.

[CC06] J. Cassaigne and N. Chekhova. Fonctions de récurrence des suites d'Arnoux-Rauzy et réponse à une question de Morse et Hedlund. *Ann. Inst. Fourier (Grenoble)*, 56(7):2249–2270, 2006. Numération, pavages, substitutions.

[CFM08] J. Cassaigne, S. Ferenczi, and A. Messaoudi. Weak mixing and eigenvalues for Arnoux-Rauzy sequences. *Ann. Inst. Fourier (Grenoble)*, 58(6):1983–2005, 2008.

[CFZ00] J. Cassaigne, S. Ferenczi, and L. Q. Zamboni. Imbalances in Arnoux-Rauzy sequences. *Ann. Inst. Fourier (Grenoble)*, 50(4):1265–1276, 2000.

[CH73] E. M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7:138–153, 1973.

[Che09] N. Chevallier. Coding of a translation of the two-dimensional torus. *Monatsh. Math.*, 157(2):101–130, 2009.

[CK97] J. Cassaigne and J. Karhumäki. Toeplitz words, generalized periodicity and periodically iterated morphisms. *European J. Combin.*, 18(5):497–510, 1997.

[CN03] J. Cassaigne and F. Nicolas. Quelques propriétés des mots substitutifs. *Bull. Belg. Math. Soc. Simon Stevin*, 10(suppl.):661–676, 2003.

[Cob68] A. Cobham. On the hartmanis-stearns problem for a class of tag machines. In *Proceedings of the 9th Annual Symposium on Switching and Automata Theory (swat 1968)*, pages 51–60, Washington, DC, USA, 1968. IEEE Computer Society.

[CR10] J. Currie and N. Rampersad. Cubefree words with many squares. *Discrete Math. Theor. Comput. Sci.*, 12(3):29–34, 2010.

[Daj02] K. Dajani. A note on rotations and interval exchange transformations on 3-intervals. *Int. J. Pure Appl. Math.*, 1(2):151–160, 2002.

[Dek97] F. M. Dekking. What is the long range order in the Kolakoski sequence? In *The mathematics of long-range aperiodic order (Waterloo, ON, 1995)*, volume 489 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 115–125. Kluwer Acad. Publ., Dordrecht, 1997.

[Dev08] R. Deviatov. On subword complexity of morphic sequences. In *Computer science—theory and applications*, volume 5010 of *Lecture Notes in Comput. Sci.*, pages 146–157. Springer, Berlin, 2008.

[DHS99] F. Durand, B. Host, and C. Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergodic Theory Dynam. Systems*, 19:953–993, 1999.

[Did97] G. Didier. Échanges de trois d'intervalles et suites sturmiennes. *J. Théor. Nombres Bordeaux*, 9(2):463–478, 1997.

[Did98a] G. Didier. Codages de rotations et fractions continues. *J. Number Theory*, 71(2):275–306, 1998.

[Did98b] G. Didier. Combinatoire des codages de rotations. *Acta Arith.*, 85(2):157–177, 1998.

[DJP01] X. Droubay, J. Justin, and G. Pirillo. Epi-Sturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.

[DL] F. Durand and J. Leroy. S-adic Bratteli diagrams. preprint.

[DL06] D. Damanik and D. Lenz. Substitution dynamical systems: characterization of linear repetitivity and applications. *J. Math. Anal. Appl.*, 321(2):766–780, 2006.

[DLR] F. Durand, J. Leroy, and G. Richomme. Some examples and counter-examples about the S-adic conjecture. In *Numération 2011*.

[dLV89] A. de Luca and S. Varricchio. Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups. *Theoret. Comput. Sci.*, 63(3):333–348, 1989.

[Dur98a] F. Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179:89–101, 1998.

[Dur98b] F. Durand. A generalization of Cobham's theorem. *Theory Comput. Syst.*, 31(2):169–185, 1998.

[Dur00] F. Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory Dynam. Systems*, 20:1061–1078, 2000.

[Dur02] F. Durand. A theorem of Cobham for non-primitive substitutions. *Acta Arith.*, 104(3):225–241, 2002.

[Dur03] F. Durand. Corrigendum and addendum to: "Linearly recurrent subshifts have a finite number of non-periodic subshift factors". *Ergodic Theory Dynam. Systems*, 23:663–669, 2003.

[ELR75] A. Ehrenfeucht, K. P. Lee, and G. Rozenberg. Subword complexities of various classes of deterministic developmental languages without interactions. *Theor. Comput. Sci.*, 1:59–75, 1975.

[ELR76] A. Ehrenfeucht, K. P. Lee, and G. Rozenberg. On the number of subwords of everywhere growing DT0L languages. *Discrete Math.*, 15:223–234, 1976.

[ER81] A. Ehrenfeucht and G. Rozenberg. On the subword complexity of square-free DOL languages. *Theoret. Comput. Sci.*, 16(1):25–32, 1981.

[ER83] A. Ehrenfeucht and G. Rozenberg. On the subword complexity of m-free DOL languages. *Inform. Process. Lett.*, 17(3):121–124, 1983.

[Fek23] M. Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, 1923.

[Fer95] S. Ferenczi. Les transformations de Chacon : combinatoire, structure géométrique, lien avec les systèmes de complexité $2n+1$. *Bull. Soc. Math. France*, 123:271–292, 1995.

[Fer96] S. Ferenczi. Rank and symbolic complexity. *Ergodic Theory Dynam. Systems*, 16:663–682, 1996.

[Fer99] S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Math.*, 206(1-3):145–154, 1999. Combinatorics and number theory (Tiruchirappalli, 1996).

[FHZ01] S. Ferenczi, C. Holton, and L. Q. Zamboni. Structure of three interval exchange transformations. I. An arithmetic study. *Ann. Inst. Fourier (Grenoble)*, 51(4):861–901, 2001.

[FHZ03] S. Ferenczi, C. Holton, and L. Q. Zamboni. Structure of three-interval exchange transformations. II. A combinatorial description of the trajectories. *J. Anal. Math.*, 89:239–276, 2003.

[FHZ04] S. Ferenczi, C. Holton, and L. Q. Zamboni. Structure of three-interval exchange transformations III: ergodic and spectral properties. *J. Anal. Math.*, 93:103–138, 2004.

[Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[Fog11] N. Pytheas Fogg. Terminologie S-adique et propriétés. https://www2.lirmm.fr/~monteil/hebergement/pytheas-fogg/terminologie_s_adique.pdf, 2011.

[FZ08] S. Ferenczi and L. Q. Zamboni. Languages of k-interval exchange transformations. *Bull. Lond. Math. Soc.*, 40(4):705–714, 2008.

[FZ10] S. Ferenczi and L. Q. Zamboni. Structure of K-interval exchange transformations: induction, trajectories, and distance theorems. *J. Anal. Math.*, 112:289–328, 2010.

[GJ09] A. Glen and J. Justin. Episturmian words: a survey. *Theor. Inform. Appl.*, 43(3):403–442, 2009.

[GLR09] A. Glen, F. Levé, and G. Richomme. Directive words of episturmian words: equivalences and normalization. *Theor. Inform. Appl.*, 43(2):299–319, 2009.

[GMP03] L.-S. Guimond, Z. Masáková, and E. Pelantová. Combinatorial properties of infinite words associated with cut-and-project sequences. *J. Théor. Nombres Bordeaux*, 15(3):697–725, 2003.

[Gri73] C. Grillenberger. Constructions of strictly ergodic systems. I. Given entropy. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 25:323–334, 1972/73.

[Hon10] J. Honkala. The equality problem for purely substitutive words. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 505–529. Cambridge Univ. Press, Cambridge, 2010.

[Hos86] B. Host. Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable. *Ergodic Theory Dynam. Systems*, 6(4):529–540, 1986.

[Hos00] B. Host. Substitution subshifts and Bratteli diagrams. In *Topics in symbolic dynamics and applications (Temuco, 1997)*, volume 279 of *London Math. Soc. Lecture Note Ser.*, pages 35–55. Cambridge Univ. Press, Cambridge, 2000.

[HP89] B. Host and F. Parreau. Homomorphismes entre systèmes dynamiques définis par substitutions. *Ergodic Theory Dynam. Systems*, 9(3):469–477, 1989.

[HPS92] R. H. Herman, I. F. Putnam, and C. F. Skau. Ordered Bratteli diagrams, dimension groups and topological dynamics. *Internat. J. Math.*, 3(6):827–864, 1992.

[HZ99] C. Holton and L. Q. Zamboni. Descendants of primitive substitutions. *Theory Comput. Syst.*, 32(2):133–157, 1999.

[IS75] S. Ito and I. Shiokawa. A construction of β-normal sequences. *J. Math. Soc. Japan*, 27:20–23, 1975.

[JK69] K. Jacobs and M. Keane. $0-1$-sequences of Toeplitz type. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 13:123–131, 1969.

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

[JV00] J. Justin and L. Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.

[KBC10] A. Ya. Kanel-Belov and A. L. Chernyat′ev. Describing the set of words generated by interval exchange transformation. *Comm. Algebra*, 38(7):2588–2605, 2010.

[Kea75] M. Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975.

[Klo11] K. Klouda. *Non-standart numeration systems and combinatorics on words*. PhD thesis, Université Paris 7 Denis Diderot and České vysoké učení technické v Praze, 2011.

[Kos98] M. Koskas. Complexités de suites de Toeplitz. *Discrete Math.*, 183(1-3):161–183, 1998.

[KP11] H. Kim and S. Park. Toeplitz sequences of intermediate complexity. *J. Korean Math. Soc.*, 48(2):383–395, 2011.

[KS67] A. B. Katok and A. M. Stepin. Approximations in ergodic theory. *Uspehi Mat. Nauk*, 22(5 (137)):81–106, 1967.

[Ler12] J. Leroy. Some improvements of the S-adic conjecture. *Advances in Applied Mathematics*, 48(1):79 – 98, 2012.

[LM95] D. Lind and B. Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

[LN98] L.-M. Lopez and P. Narbel. D0L-systems and surface automorphisms. In *Mathematical foundations of computer science, 1998 (Brno)*, volume 1450 of *Lecture Notes in Comput. Sci.*, pages 522–532. Springer, Berlin, 1998.

[LN00] L.-M. Lopez and P. Narbel. Substitutions from Rauzy induction (extended abstract). In *Developments in language theory (Aachen, 1999)*, pages 200–209. World Sci. Publ., River Edge, NJ, 2000.

[LN01] L.-M. Lopez and P. Narbel. Substitutions and interval exchange transformations of rotation class. *Theoret. Comput. Sci.*, 255(1-2):323–344, 2001.

[Lot02] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.

[MH38] M. Morse and G. A. Hedlund. Symbolic Dynamics. *Amer. J. Math.*, 60(4):815–866, 1938.

[MH40] M. Morse and G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.

[MM10] C. Mauduit and C. G. Moreira. Complexity of infinite sequences with zero entropy. *Acta Arith.*, 142(4):331–346, 2010.

[Mon05] T. Monteil. *Illumination in polygonal billiards and symbolic dynamics*. PhD thesis, Université Aix-Marseille II, 2005.

[Mos96] B. Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France*, 124(2):329–346, 1996.

[MS93] F. Mignosi and P. Séébold. Morphismes sturmiens et règles de Rauzy. *J. Théor. Nombres Bordeaux*, 5(2):221–233, 1993.

[MZ02] F. Mignosi and L. Q. Zamboni. On the number of Arnoux-Rauzy words. *Acta Arith.*, 101(2):121–129, 2002.

[NP09] F. Nicolas and Y. Pritykin. On uniformly recurrent morphic sequences. *Internat. J. Found. Comput. Sci.*, 20(5):919–940, 2009.

[NR07] S. Nicolay and M. Rigo. About frequencies of letters in generalized automatic sequences. *Theoret. Comput. Sci.*, 374(1-3):25–40, 2007.

[Ose66] V. I. Oseledec. The spectrum of ergodic automorphisms. *Dokl. Akad. Nauk SSSR*, 168:1009–1011, 1966.

[Ost22] A. Ostrowski. Bemerkungen zur theorie der diophantischen approximationen. *Abhandlungen aus dem Mathematischen Seminar der UniversitÃd't Hamburg*, 1:77–98, 1922. 10.1007/BF02940581.

[Pan83] J.-J. Pansiot. Hiérarchie et fermeture de certaines classes de tag-systèmes. *Acta Inform.*, 20(2):179–196, 1983.

[Pan84] J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. In *Automata, languages and programming (Antwerp, 1984)*, volume 172 of *Lecture Notes in Comput. Sci.*, pages 380–389. Springer, Berlin, 1984.

[Pan85] J.-J. Pansiot. Subword complexities and iteration. *Bulletin of the EATCS*, pages 55–62, 1985.

[Par99] B. Parvaix. Substitution invariant Sturmian bisequences. *J. Théor. Nombres Bordeaux*, 11(1):201–210, 1999. Les XXèmes Journées Arithmétiques (Limoges, 1997).

[PV07] G. Paquin and L. Vuillon. A characterization of balanced episturmian sequences. *Electron. J. Combin.*, 14(1):Research Paper 33, 12 pp. (electronic), 2007.

[Que87] M. Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1987.

[RA96] D. Razafy Andriamampianina. Suites de Toeplitz, p-pliage, suites automatiques et polynômes. *Acta Math. Hungar.*, 73(3):179–190, 1996.

[Rau77] G. Rauzy. Une généralisation du développement en fraction continue. In *Séminaire Delange-Pisot-Poitou, 18e année: 1976/77, Théorie des nombres, Fasc. 1*, pages Exp. No. 15, 16. Secrétariat Math., Paris, 1977.

[Rau79] G. Rauzy. Échanges d'intervalles et transformations induites. *Acta Arith.*, 34(4):315–328, 1979.

[Rau83] G. Rauzy. Suites à termes dans un alphabet fini. 1983.

[Ric03] G. Richomme. Conjugacy and episturmian morphisms. *Theoret. Comput. Sci.*, 302:1–34, 2003.

[Ric07] G. Richomme. A local balance property of episturmian words. In *Developments in language theory*, volume 4588 of *Lecture Notes in Comput. Sci.*, pages 371–381. Springer, Berlin, 2007.

[RM02] M. Rigo and A. Maes. More on generalized automatic sequences. *J. Autom. Lang. Comb.*, 7(3):351–376, 2002.

[Rot94] G. Rote. Sequences with subword complexity $2n$. *J. Number Theory*, 46:196–213, 1994.

[RS80] G. Rozenberg and A. Salomaa. *The mathematical theory of L systems*, volume 90 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980.

[RW02] G. Richomme and F. Wlazinski. Some results on k-power-free morphisms. *Theoret. Comput. Sci.*, 273(1-2):119–142, 2002. WORDS (Rouen, 1999).

[RZ00] R. N. Risley and L. Q. Zamboni. A generalization of Sturmian sequences: combinatorial structure and transcendence. *Acta Arith.*, 95(2):167–184, 2000.

[Sal10] P. V. Salimov. On uniform recurrence of a direct product. *Discrete Math. Theor. Comput. Sci.*, 12(4):1–8, 2010.

[Sha88] J. Shallit. A generalization of automatic sequences. *Theoret. Comput. Sci.*, 61(1):1–16, 1988.

[Sie05] A. Siegel. Spectral theory for dynamical systems arising from substitutions. In *European women in mathematics—Marseille 2003*, volume 135 of *CWI Tract*, pages 11–26. Centrum Wisk. Inform., Amsterdam, 2005.

[Tap94] T. Tapsoba. Automates calculant la complexité de suites automatiques. *J. Théor. Nombres Bordeaux*, 6(1):127–134, 1994.

[Tap96] T. Tapsoba. Special factors of automatic sequences. *J. Pure Appl. Algebra*, 108(3):301–313, 1996.

[Thu06] A. Thue.
"Uber unendliche Zeichenreihen. Skrifter udgivne af Videnskabsselskabet i Christiania. 1906. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell *et al.*, editors, Universitetsforlaget, Oslo, 1977, pp. 139– 158.

[Thu12] A. Thue.
"Uber die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Skrifter utg. av Videnskapsselskapet i Kristiania. I, Mat.-naturv. klasse. 1912. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell *et al.*, editors, Universitetsforlaget, Oslo, 1977, pp. 413–477.

[Van00] D. Vandeth. Sturmian words and words with a critical exponent. *Theoret. Comput. Sci.*, 242(1-2):283–300, 2000.

[Vee84a] W. A. Veech. The metric theory of interval exchange transformations. I. Generic spectral properties. *Amer. J. Math.*, 106(6):1331–1359, 1984.

[Vee84b] W. A. Veech. The metric theory of interval exchange transformations. II. Approximation by primitive interval exchanges. *Amer. J. Math.*, 106(6):1361–1387, 1984.

[Vee84c] W. A. Veech. The metric theory of interval exchange transformations. III. The Sah-Arnoux-Fathi invariant. *Amer. J. Math.*, 106(6):1389–1422, 1984.

[Ver82] A. M. Vershik. A theorem on Markov periodic approximation in ergodic theory. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 115:72–82, 306, 1982. Boundary value problems of mathematical physics and related questions in the theory of functions, 14.

[Via06] M. Viana. Ergodic theory of interval exchange maps. *Rev. Mat. Complut.*, 19(1):7–100, 2006.

[VL92] A. M. Vershik and A. N. Livshits. Adic models of ergodic transformations, spectral theory, substitutions, and related topics. In *Representation theory and dynamical systems*, volume 9 of *Adv. Soviet Math.*, pages 185–204. Amer. Math. Soc., Providence, RI, 1992.

[Vui07] L. Vuillon. On the number of return words in infinite words constructed by interval exchange transformations. *Pure Math. Appl. (PU.M.A.)*, 18(3-4):345–355, 2007.

[War02] K. Wargan. *S-adic Dynamical Systems and Bratteli Diagrams.* PhD thesis, George Washington University, 2002.

[Wil84] S. Williams. Toeplitz minimal flows which are not uniquely ergodic. *Z. Wahrsch. Verw. Gebiete*, 67(1):95–107, 1984.

[WZ01] N. Wozny and L. Q. Zamboni. Frequencies of factors in Arnoux-Rauzy sequences. *Acta Arith.*, 96(3):261–278, 2001.