Université de Liège
Faculté des Sciences
Unité de Recherche *Mathematics*

A study of dendricity through the lens of morphisms

France Gheeraert

Dissertation présentée
en vue de l'obtention du grade académique de
Docteur en Sciences
Année académique 2023 – 2024

Université de Liège
Faculté des Sciences
Unité de Recherche *Mathematics*

A study of dendricity through the lens of morphisms

France Gheeraert

Dissertation présentée
en vue de l'obtention du grade académique de
Docteur en Sciences
Année académique 2023 – 2024

Abstract

Dendric languages were introduced a decade ago as a generalization of both Arnoux-Rauzy languages and codings of regular interval exchange transformations. Right away, it was proved that dendric languages possess strong algebraic properties and are preserved under some fundamental operations, namely derivation and decoding with respect to a bifix code. A few years later, Dolce and Perrin studied the more general concept of eventual dendricity and proved that this notion is stable under topological conjugacy.

In this thesis, we explore another aspect of (eventual) dendricity and delve deeper into the link between dendricity and morphisms. We mainly answer four questions.

We first look at the evolution of the factor complexity when applying a morphism and show that for many languages, including the eventually dendric ones, it grows at most by an additive constant.

We then turn to the preservation of dendricity and show that the morphisms for which the image of a dendric language is always dendric are precisely those generated by the Arnoux-Rauzy morphisms.

Continuing on this idea of describing when the image of a dendric language under some morphism is dendric, we focus on particular morphisms related to return words, and for these so-called *return morphisms*, we obtain a practical characterization.

Finally, given any set of return morphisms, we show how to characterize the sequences of morphisms in this set generating a dendric language, obtaining an S-adic characterization of (eventual) dendricity. Consequently, we prove that (eventual) dendricity is decidable for morphic languages.

Keywords: Combinatorics on words · Symbolic dynamics · Dendric words · Neutral words · Morphisms · Factor complexity · S-adic representations · Return words

Résumé

Les langages dendriques ont été introduits il y a dix ans comme étant à la fois une généralisation des langages d'Arnoux-Rauzy et des codages d'échanges d'intervalles réguliers. Dès le départ, leurs propriétés algébriques ainsi que leur stabilité pour des opérations telles que la dérivation ont été mises en évidence. Quelques années plus tard, Dolce and Perrin ont étudié la notion plus générale de dendricité ultime et ont notamment montré que cette propriété était stable pour la conjugaison topologique.

Dans cette thèse, nous nous intéressons à un autre aspect de la dendricité (ultime) et explorons plus en profondeur le lien entre dendricité et morphismes. Plus précisément, nous répondons à quatre questions.

Nous commençons par regarder l'évolution de la complexité lors de l'application d'un morphisme et montrons que pour de nombreux langages, y compris les ultimement dendriques, elle augmente d'au plus une constante additive.

Nous nous tournons ensuite vers la préservation de la dendricité et montrons que les morphismes pour lesquels l'image d'un langage dendrique est toujours dendrique sont engendrés par les morphismes d'Arnoux-Rauzy.

Toujours dans cette idée de décrire quand l'image d'un langage dendrique est elle-même dendrique, nous nous intéressons à des morphismes particuliers liés aux mots de retour. Pour ces morphismes dit *de retour*, nous obtenons une caractérisation effective.

Enfin, étant donné un ensemble de morphismes de retour, nous montrons comment caractériser les suites de morphismes dans cet ensemble qui engendrent un langage dendrique, obtenant ainsi une caractérisation S-adique de la dendricité (ultime). En conséquence, nous montrons que la dendricité (ultime) est décidable pour les langages morphiques.

Keywords: Combinatoire des mots · Dynamique symbolique · Mots dendriques · Mots neutres · Morphismes · Complexité · Représentations S-adiques · Mots de retour

Remerciements

Je ne peux commencer ces remerciements sans adresser un immense merci à Julien Leroy pour m'avoir fait confiance et avoir accepté de m'accompagner dans cette grande aventure qu'est le doctorat, pour les discussions et séances de recherches qui ont mené à une très grande partie des résultats présentés ici, pour ses nombreuses questions me poussant toujours plus loin et ses réponses à mes questions encore plus nombreuses, et enfin pour sa relecture attentive de ce travail.

Je tiens à remercier Marie-Pierre Béal, Julien Cassaigne, Emilie Charlier, Fabien Durand, Arnaud Hilion et Michel Rigo de me faire l'honneur d'être membres de mon jury de thèse et j'espère que ce travail en sera à la hauteur.

Ces années de thèse auraient été bien moins agréables sans les nombreux conseils et encouragements qui m'ont été donnés par les membres du département ainsi que les temps de midi whist qu'on a partagés, et pour cela je les remercie. Un merci plus particulier va aux nombreux membres passé et présent du "couloir de math discrètes" (including the non french-speaking ones) pour leur accueil malgré des conditions loin d'être idéales ainsi que pour les moments passés en conférence lorsqu'elles ont pu reprendre. Concernant ce travail plus spécifiquement, je souhaite remercier Manon pour ses innombrables conseils et sa relecture détaillée de l'introduction, et Pierre pour les discussions sur certaines questions ouvertes.

My next thanks go to the people I met during various conferences or seminars and with whom I had the chance to have passionate discussions on mathematics (but not only) around a blackboard, a drink, or a meal, sometimes leading to further collaborations.

En particulier, je remercie Sébastien Labbé pour l'accueil chaleureux à Bordeaux à plusieurs reprises ainsi que pour ses nombreux conseils sur les bonnes pratiques en SageMath. I also want to thank Giuseppe Romana and Marinella Sciortino for welcoming me to Palermo and introducing me to a new research topic, but also to the delicious Sicilian food.

Toutes ces rencontres et ces échanges n'auraient pas pu avoir lieu sans le FNRS qui m'a donné l'opportunité de faire un doctorat.

Enfin, je souhaite adresser un remerciement bien moins mathématique mais tout aussi sincère à mes parents pour leur soutien moral lors de ces années de doctorat, particulièrement pendant ces derniers mois.

Contents

Introduction

Combinatorics on words is a fairly recent field at the intersection of mathematics and theoretical computer science. Indeed, while the notion of words already implicitly appeared in works of adjacent domains, the first combinatorics on words paper is often attributed to Thue at the beginning of the 20^{th} century [Thu06, Thu12] and the first reference book of the field dates back to 1983 [Lot97]. Since then, Combinatorics on words has been vastly developed due to its strong links with other domains such as symbolic dynamics, group theory, number theory, and graph theory for example.

A perfect example of this connection to various fields is the *Sturmian words (or sequences)*. They were introduced by Morse and Hedlund in 1940 [MH40], although they can be related to the much earlier work of Bernoulli III [Ber71]. They admit the following combinatorial definition: they are the (right) infinite words containing $n+1$ distinct length-n words for all $n \geq 0$. Sturmian words can also be defined in symbolic dynamics as the codings of irrational rotations, or from a number theory point of view using continued fractions, for example.

This diversity of approaches and applications made (and still makes) the Sturmian words some of the most studied sequences in combinatorics on words. Two well-known surveys ([Lot02, Chapter 2] and [Fog02, Chapter 6]) list some of their numerous properties.

By definition, Sturmian sequences are on a two-letter alphabet and it was then natural to search for generalizations on larger alphabets. Due to the different definitions of Sturmian words, several versions were studied. We focus here on two that can, in some sense, be considered to be antagonistic while still keeping many properties. The first one is the family of *Arnoux-Rauzy languages*. They originated in a combinatorial characterization of Sturmian words and were first introduced on a three-letter alphabet in [AR91], then studied in a more general setting under the name of *strict episturmian sequences* in [DJP01, JP02]. The second family is the set of languages of interval exchange transformations (or IET), which extend the symbolic dynamics point of view of Sturmian words and were extensively studied in [Kea75, Vee78, Rau79, FZ08] for example.

Inspired by a study of Arnoux-Rauzy languages and the properties of their bifix codes in [BDFP$^+$12], a group of researchers introduced a new generalization of Sturmian languages unifying both Arnoux-Rauzy languages and languages of regular interval exchange transformations: the dendric languages [BDFD$^+$15a].

With each element w of a language $\mathcal{L}$ on the alphabet $\mathcal{A}$, we can associate three sets that represent the different ways of extending the word w. More precisely, we have the set $E^L_{\mathcal{L}}(w)$ of letters preceding w in $\mathcal{L}$, the set $E^R_{\mathcal{L}}(w)$ of letters following w and the set $E_{\mathcal{L}}(w)$ of pairs of letters surrounding w. We then define a graph whose left (resp., right) vertices are the elements of $E^L_{\mathcal{L}}(w)$ (resp., $E^R_{\mathcal{L}}(w)$) and a left vertex a is connected to a right vertex b whenever $(a, b) \in E_{\mathcal{L}}(w)$. We then say that w is *dendric* if this graph is a tree and the language $\mathcal{L}$ itself is *dendric* if all of its elements are.

The initial study of dendric languages done in [BDFD$^+$15a, BDFD$^+$15c, BDFD$^+$15d] showed that they possessed remarkable properties in terms of return words and bifix codes, as well as being stable under two classical operations: derivation and maximal bifix decoding. Since then, several papers have studied other aspects of dendric languages. We mention here [AC16] focusing on the Schützenberger group, [BDD$^+$18] studying the morphisms stabilizing dendric sequences and the continuous eigenvalues, [BCB19] proving that balancedness on letters implies balancedness on words, and [BCBD$^+$21] in which the authors extend this last result and study dimension groups.

Relaxing the hypothesis on small words, the notion of *eventual dendricity* was studied in [DP21] where the authors prove that the class of eventually dendric shift spaces is stable under topological conjugacy. Independently, they were also introduced by Damron and Fickenscher [DF22] who provided bounds on their number of ergodic measures. Note that the notion of eventual dendricity also contains a third natural generalization of Sturmian languages: the recurrent balanced languages [DDP23]. Let us also mention the suffix-connected languages [GO22], another generalization of dendric languages.

The main goal of this thesis is to study an aspect of (eventually) dendric languages that, while related to some of the existing results, has not been fully explored yet: their behavior with respect to morphisms.

The concept of morphisms, i.e., applications preserving some pre-defined structure, is fundamental in most branches of mathematics, and combinatorics on words is no exception. Indeed, (word) morphisms appear as early as in Thue's work [Thu06, Thu12]. The morphic sequences, generated by two morphisms, form a classically studied family, see [AS03] for a survey.

Here, we mainly explore three questions on (eventually) dendric languages regarding morphisms: the properties of their images under a morphism, the preservation of dendricity when applying a morphism, and S-adic representations generalizing the idea of a morphism generating a language.

This work is divided into five chapters detailed below. Moreover, at the end of each chapter (except the first one), we list some open questions and perspectives of research related to the results of the corresponding chapter. The content of this work is mainly extracted from the following three papers [GLL22, GL22, Ghe23] but also contains new results.

The first chapter recalls the basic notions of combinatorics on words found in any book of the domain ([Lot97, AS03, Fog02] for example). We particularly emphasize the notions of languages (i.e., factorial bi-extendable sets of finite words) and morphisms since both concepts are central throughout this work.

We then turn to the three well-known families mentioned earlier: Sturmian languages, Arnoux-Rauzy languages, and languages of regular interval exchange transformations. In particular, we give a combinatorial characterization of languages of regular interval exchange transformations, which will be more practical than the usual definition in the following chapters.

We end this chapter with weak, neutral, and strong languages. A word w is *weak* (resp., *neutral*; resp., *strong*) if $E_{\mathcal{L}}(w) - E^L_{\mathcal{L}}(w) - E^R_{\mathcal{L}}(w) + 1$ is negative (resp., zero; resp., positive). These notions are fundamental since the study of weak and strong words allows to completely describe the factor complexity of the language [Cas96, Cas97]. In particular, neutral languages (i.e., exclusively containing neutral words) have affine factor complexity. Since dendricity implies neutrality, some results presented in this work are stated in the more general setting of neutral languages.

In Chapter 2, we introduce (eventual) dendricity and familiarize ourselves with some basic properties. As stated earlier, dendric languages are strongly related to the families of languages mentioned so far. Namely, Sturmian languages, Arnoux-Rauzy languages, and languages of regular interval exchange transformations (or RIET) are dendric [BDFD$^+$15a, BDFD$^+$15b], and dendric languages are neutral [BDFD$^+$15a]. A complete diagram synthesizing the interactions between these families and a few more is represented in Figure 2.2.

Requiring a property to be true "eventually" often implies some additional closure property. In the case of dendricity, this is the stability under topological conjugacy. This operation is the dynamical version of an isomorphism and is thus fundamental. A property stable under topological conjugacy is often said to be *dynamical*. Therefore, eventual dendricity being a dynamical property while dendricity is not was the original motivation of Dolce and Perrin for the introduction of eventually dendric languages in [DP21].

The family of eventually dendric languages seems particularly important since it admits many alternative definitions. Indeed, we show in Section 2.3 that being eventually dendric is equivalent to being eventually neutral (resp., eventually acyclic; resp., eventually ordinary) even though these four properties are all distinct when removing the adverb "eventually".

We devote the last section of Chapter 2 (before the open questions) to some new graphs. Indeed, while the dendricity of a language is defined using graphs associated with each word, we provide an alternative characterization by looking at two graphs per length. More precisely, for a given language $\mathcal{L}$ and an integer n, we define $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) as the graph whose vertices are the letters, and two letters a and b are connected by an edge labeled by length-n word v if $a, b \in E_{\mathcal{L}}^L(v)$ (resp., $a, b \in E_{\mathcal{L}}^R(v)$). While all these graphs are needed to characterize dendricity, we will more specifically be interested in their limit behavior when n tends to infinity. The two obtained graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are then closely related to asymptotic pairs, a classical tool in symbolic dynamics. In the following chapters, we will use both the original definition of dendricity and this new approach in the results and proofs.

The core concept behind the families of languages mentioned in this work is the notion of left, right, and bilateral extensions of words inside the language defined at the beginning of this introduction. As stated earlier, they determine the factor complexity [Cas96, Cas97] but can also be used to count various other objects. In the third chapter, we assess the influence of extensions on three different aspects of languages. We will mostly be interested in the number of extensions and less in their interactions. In other words, most of the results of this chapter are stated in terms of neutrality and not of dendricity.

We first look at the classical notion of codes [BPR10]. These objects, at the intersection between combinatorics on words and algebra, were central in the first papers on dendricity [BDFD+15a, BDFD+15c, BDFD+15d]. They were also the topic of the paper [DP17] on neutral languages. Therefore, many properties of codes included in a neutral or dendric language are known. We focus here on a refinement of the link between neutrality and factor complexity. More precisely, we obtain a description of the total number of left (resp., right) extensions of the elements of a prefix (resp., suffix) code.

The second object we consider is the factor complexity, not of a neutral or dendric language as this is already known, but of the image of such a language under a non-erasing morphism. A classical result [AS03, CN10] states that, when applying such a morphism to a language, the factor complexity cannot increase by more than a multiplicative constant. The proof of this result relies on the simple observation that any length-n word in the image appears in the image of a length-n word w of the initial language. It is sometimes sufficient to look at a suffix of w. Therefore, instead of considering all length-n words of the initial language, we look at the elements of a code. Using the result proved in the first part of the chapter, we show that, for a large family of languages (including eventually dendric ones), the factor complexity can only increase by an additive constant when applying a non-erasing morphism. However, this is not the case in general for the Thue-Morse language.

Finally, we turn to the properties of return words in dendric and neutral languages. Introduced by Durand in [Dur98], they are strongly related to topological induction in dynamical systems. Return words can be seen as natural blocks to decompose an infinite word, the operation consisting of looking at the sequence of blocks instead is then called *derivation*. The most famous result on dendric languages to date, called the Return Theorem [BDFD$^+$15a], states that the sets of return words in a dendric language are bases of the free group over the alphabet. In particular, there are as many return words as letters in the alphabet. More generally, the relation between neutrality and the number of return words was studied in [BPS08, DP17]. We recall these results and slightly extend them in Subsection 3.3.1. Another famous result on dendric languages is their stability under derivation [BDFD$^+$15d]. We extend it to eventually dendric languages in Subsection 3.3.2.

Two of the first results on dendricity were about the stability under some operations: derivation [BDFD$^+$15d] and maximal bifix decoding [BDFD$^+$15a, BDFD$^+$15d]. These operations can be seen as desubstitution (or taking the pre-image) under some particular morphisms. The opposite question of the preservation of dendricity when applying a morphism is the theme of the fourth chapter. However, this question is too broad so we focus on three partial answers.

We first obtain large families of morphisms that never preserve dendricity based on the following simple idea: dendric languages have a restricted factor complexity, strongly linked to the size of their underlying alphabets. Moreover, using the results of Chapter 3, we obtain a bound on the factor complexity of the image of a dendric language. Consequently, if the image alphabet is larger than the initial alphabet, the image of a dendric language is never dendric.

We can also look at the morphisms for which the image of a dendric language is always dendric. It is well-known that the only morphisms preserving Arnoux-Rauzy languages are the so-called *Arnoux-Rauzy morphisms* [JP02]. It turns out that we obtain the same morphisms for dendricity. On the other hand, the morphisms preserving languages of regular interval exchange transformation, starting from an alphabet of size at least three, are all trivial.

We then focus on preservation for specific morphisms called *return morphisms*. We chose them as they are behind the derivation operation and will be the building blocks of the results of Chapter 5. Moreover, these morphisms have interesting recognizability properties which allow us to obtain (reasonably) simple results. Indeed, we can fully characterize, for each return morphism, for which languages it preserves dendricity. This result is especially interesting in the case of a return morphism associated with a word or a set of letters. Indeed, in these cases, this characterization only depends on the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ introduced in the second chapter (and sometimes on the length-2 elements of $\mathcal{L}$).

On the other hand, the question of preserving eventual dendricity may admit a much simpler answer since we are not aware of any example of an eventually dendric language whose image under a morphism is not eventually dendric. As we show in Section 4.5, this is closely related to a question of Dolce and Perrin [DP21] on the stability of eventual dendricity under topological factorization. We also show that eventual dendricity is closed under the application of (resp., desubstitution with respect to) a recognizable morphism, which includes the case of return morphisms.

Chapter 5 focuses on a powerful tool to understand the structure and properties of a language: the *S-adic representations*. The terminology first appears in [Fer96] and is inspired by the work of Vershik [VL92] due to the historical link between S-adic representations and Bratelli-Vershik diagrams [Dur10]. S-adic representations quickly gained traction in symbolic dynamics as they seemed to be an appropriate tool to study other objects classically associated with topological spaces: dimension groups [BCBD$^+$21], topological rank [DDMP21] and ergodic measures [BHL23], for example.

They are also of interest for the combinatorics on words community since they generalize the idea of morphic (or substitutive) languages, and many known families of languages can be characterized by their S-adic representations: Sturmian [Fog02], linearly recurrent [Dur00, Dur03] and of at most linear complexity [Esp23], to mention a few.

The goal of the last chapter of this thesis is to show that recurrent (eventually) dendric languages can also be characterized by their S-adic representations. To obtain this result, we use theorems coming from every other chapter: the properties of the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ for (eventually) dendric languages from Chapter 2, the stability of (eventual) dendricity under derivation from Chapter 3 and the preservation of (eventual) dendricity when applying a return morphism from Chapter 4.

The characterization we obtain uses a theoretical graph that we effectively build in the case of a ternary alphabet (Figure 5.3). We also explain how to deduce an S-adic characterization of the languages of regular interval exchange transformations in Subsection 5.3.3.

In fact, given any set of return morphisms, we characterize all the dendric languages having an S-adic representation using exclusively morphisms from this set. We exploit this stronger formulation to characterize the dendric languages having exactly one right special word of each length in Subsection 5.3.2.

Finally, this stronger version is also fundamental to show that (eventual) dendricity is decidable for uniformly recurrent morphic languages. Indeed, using results of Durand [Dur98, Dur13b], we first obtain a constructive S-adic representation using only two return morphisms, then apply our characterization to show that dendricity is equivalent to the existence of some finite paths in a constructive graph.

Let me end this introduction by mentioning that, while dendric languages were my main focus of research during my PhD years, I also had the opportunity to work on other topics, mainly on string attractors on which I have co-authored a paper [GRS23].

Chapter 1

Preliminaries

This chapter contains the background definitions which will be fundamental in the following chapters. While most notions presented here are classical, we would like to emphasize two possibly unusual conventions used throughout this work: a language is always factorial and biextendable, and Sturmian and Arnoux-Rauzy languages are not necessarily recurrent.

We start by recalling in Section 1.1 the usual notions of combinatorics on words, focusing on languages and morphisms. We also recall the definition of the extensions of a word as it is a central concept for dendric languages.

In the following sections, we then give an overview of four families of languages: Sturmian, Arnoux-Rauzy, regular interval exchange transformations and neutral languages. The study of these families is at the origin of the definition of dendricity.

In Section 1.2, we start with the Sturmian words which are in some sense the simplest sequences whose properties are not trivial. In fact, most of the tools presented in this work were first introduced for or inspired by properties of Sturmian languages.

We then turn to two different generalizations of Sturmian languages in Sections 1.3 and 1.4: the Arnoux-Rauzy languages and the languages of regular interval exchange transformations. Indeed, Sturmian words are always on a binary alphabet and the idea was to generalize to larger alphabets one of the many alternative ways of defining Sturmian words. As we will see in Section 2.1, these are particular examples of dendric languages.

Finally, in Section 1.5, we introduce the notions of weak, neutral and strong words leading to the definition of neutral languages who, once again, are a generalization of Sturmian languages but which includes the dendric languages this time.

1.1 Words, languages and morphisms

We recall some fundamental notions of combinatorics on words that will be used throughout this work. This section is relatively brief as we assume that the reader is already familiar with most of these concepts. A more complete introduction can be found in [Fog02] for example.

Words and shift spaces

Definition 1.1. An *alphabet* is a finite non empty set $\mathcal{A}$ whose elements are called *letters*. A *(finite) word* on the alphabet $\mathcal{A}$ is a finite sequence of letters in $\mathcal{A}$. The *length* of a word $w = w_1 w_2 \cdots w_n$, $w_i \in \mathcal{A}$, is the number n of letters and is denoted $|w|$. The only word of length 0 is the *empty word* ε. We denote $\mathcal{A}^n$ the set of words of length $n \geq 0$ on the alphabet $\mathcal{A}$, $\mathcal{A}^* = \cup_{n \geq 0} \mathcal{A}^n$ the set of finite words and $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$ the set of non-empty finite words. The set $\mathcal{A}^*$ is a monoid when endowed with the concatenation, the empty word being the neutral element. Formally, the *concatenation* of $u = u_1 u_2 \cdots u_n$ and $v = v_1 v_2 \cdots v_m$ is defined as the word $u \cdot v = u_1 \cdots u_n v_1 \cdots v_m$. We will denote it uv instead most of the time.

A *right* (resp., *left*; resp., *bi-*) *infinite word* on the alphabet $\mathcal{A}$ is an element of $\mathcal{A}^{\mathbb{N}}$ (resp., $\mathcal{A}^{-\mathbb{N}}$; resp., $\mathcal{A}^{\mathbb{Z}}$) where $\mathbb{N}$ denotes the set of non-negative integers. Right infinite words are also sometimes called *(one-sided) sequences*. When representing a bi-infinite word, we will sometimes add a $\cdot$ between the letters at positions -1 and 0.

In this work, if the alphabet plays a role, we will always assume that we are working with the minimal alphabet, i.e., the set of letters that appear in the words we are working with. To make the distinction clearer we will sometimes say that a word is *over* the alphabet $\mathcal{A}$ if all the letters of $\mathcal{A}$ appear in the word.

We can define relations between the words as follows.

Definition 1.2. A finite word u is a *factor* of $w \in \mathcal{A}^*$ if there exist $p, s \in \mathcal{A}^*$ such that $w = pus$. If $p = \varepsilon$ (resp., $s = \varepsilon$), we say that u is a *prefix* (resp., *suffix*) of w. If $u \neq w$, we moreover say that u is a *proper factor* (resp., *proper prefix*; resp., *proper suffix*). We denote $\mathrm{Fac}(w)$ (resp., $\mathrm{Pref}(w)$; resp., $\mathrm{Suff}(w)$) the set of factors (resp., prefixes; resp., suffixes) of w, and $\mathrm{Fac}^*(w)$ (resp., $\mathrm{Pref}^*(w)$; resp., $\mathrm{Suff}^*(w)$) the set of proper factors (resp., proper prefixes; resp., proper suffixes) of w. We similarly define the notion of factor of a (one sided or two sided) infinite word by allowing p and/or s to be infinite.

An *occurrence* of a factor u of a (finite or infinite) word w is an index i such that $u = w_i w_{i+1} \cdots w_{i+|u|-1} = w_{[i,i+|u|-1]}$. We denote $|w|_u$ the number of occurrences of u in w.

We will also sometimes take a more dynamical point of view.

Definition 1.3. The *shift map* is the map $S\colon \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ such that for all $x \in \mathcal{A}^{\mathbb{Z}}$ and for all $n \in \mathbb{Z}$,

$$S(x)_n = x_{n+1}.$$

A *shift space* on the alphabet $\mathcal{A}$ is a closed subset X of $\mathcal{A}^{\mathbb{Z}}$ (if $\mathcal{A}^{\mathbb{Z}}$ is endowed with the product topology of the discrete topology over $\mathcal{A}$) which is S-invariant, i.e., $S(X) \subseteq X$. Shift spaces are particular examples of topological dynamical systems.

Languages

In this work, some of the concepts we consider are typically studied from the combinatorics on words point of view while others originate in dynamical systems. As most of the properties of this work depend only on the associated language, we attempt to unify both points of view and hopefully make the results accessible to a larger panel of readers by talking mostly about languages, as was done in the original paper on dendricity [BDFD$^+$15a].

Definition 1.4. The *language* of a (one-sided or two-sided) infinite word x is the set of its factors. We denote it $\mathcal{L}(x)$. The *language* of a shift space X is the union of the languages of its elements and we denote it $\mathcal{L}(X)$.

We will consider languages without necessarily caring about a corresponding word or shift space. In what follows and contrary to what is often found in the literature, a language is not any set of finite words as we will assume that it possesses some additional properties by definition.

Warning. In this work, we always assume that a *language* is a set of finite words which satisfies the following two hypotheses:

- the language is *factorial*, i.e., it contains the factors of its elements;
- the language is *biextendable*, i.e., for each of its elements w, there exist two letters a and b (not necessarily different) such that awb is in the language.

In particular, a language is always infinite.

These are two natural assumptions which are automatically satisfied if we consider the language of a bi-infinite word, of a recurrent one-sided infinite word (see Definition 1.6) or of a shift space.

Moreover, a language is *over* $\mathcal{A}$ if each letter of $\mathcal{A}$ appears in (at least) one word of the language.

We will then abusively assimilate an infinite word or shift space with its language as explained below.

Definition 1.5. Let P be a property defined on languages. An infinite word x satisfies P if and only if $\mathcal{L}(x)$ satisfies P. Similarly, a shift space X satisfies P if and only if $\mathcal{L}(X)$ does.

To simplify the notations, we also assume that if a notation is defined for languages, then there are corresponding notations for infinite words and shift spaces by replacing $\mathcal{L}(x)$ by x and $\mathcal{L}(X)$ by X. The only exceptions are the notations defined in the following paragraph.

As we are sometimes only interested in words of particular lengths in the language, we will denote $\mathcal{L}_n$ (resp., $\mathcal{L}_{\geq n}$; resp., $\mathcal{L}_{\leq n}$) the set of words of length n (resp., at least n; resp., at most n) in $\mathcal{L}$. We similarly define $\mathcal{L}_{>n}$ and $\mathcal{L}_{<n}$.

We mentioned above the notion of a recurrent one-sided infinite word. We properly define it here.

Definition 1.6. A language $\mathcal{L}$ is *recurrent* if for all $u, v \in \mathcal{L}$ there exists w such that $uwv \in \mathcal{L}$. For infinite words, recurrence is often understood as "each factor appears infinitely many times". A language $\mathcal{L}$ is *uniformly recurrent* if for all $u \in \mathcal{L}$, there exists $n \geq 0$ such that u is a factor of all $v \in \mathcal{L}_n$. Once again, there is an equivalent approach for infinite words by saying that every factor appears infinitely often with bounded gaps between the occurrences.

For shift spaces, we use a different terminology.

Definition 1.7. A shift space X is *minimal* if it is non-empty and its only closed S-invariant subsets are $\emptyset$ and X itself.

The following characterizations are also well-known: a shift space X is minimal if and only if, for all $x \in X$, we have $\mathcal{L}(x) = \mathcal{L}(X)$, if and only if the language $\mathcal{L}(X)$ is uniformly recurrent.

Morphisms

A classical way of generating a language is through the use of morphisms.

Definition 1.8. Throughout this work, a *morphism* is a monoid morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, i.e., $\sigma(uv) = \sigma(u)\sigma(v)$ for all $u, v \in \mathcal{A}^*$. It is therefore sufficient to give the images of the letters. We then naturally extend the morphism to infinite words by saying that

$$\sigma(x_0x_1x_2\cdots) = \sigma(x_0)\sigma(x_1)\sigma(x_2)\cdots$$

and

$$\sigma(\cdots x_{-2}x_{-1}\cdot x_0x_1\cdots) = \cdots\sigma(x_{-2})\sigma(x_{-1})\cdot\sigma(x_0)\sigma(x_1)\cdots.$$

We now define particular types of morphisms.

Definition 1.9. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism. We say that σ is

1. *non-erasing* if $\sigma(\mathcal{A}) \subseteq \mathcal{B}^+$;
2. *letter-to-letter* (or a *coding*) if $\sigma(\mathcal{A}) \subseteq \mathcal{B}$;
3. *primitive* if $\mathcal{B} \subseteq \mathcal{A}$ and there exists $n \geq 1$ such that, for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$, b is a factor of $\sigma^n(a)$;
4. *prolongable on* $a \in \mathcal{A}$ if $\mathcal{B} \subseteq \mathcal{A}$ and $\sigma(a) \in a\mathcal{B}^+$.

A morphism can also be extended to languages and shift spaces, or more precisely, a morphism induces a map from languages to languages and from shift spaces to shift spaces as described below. Observe that, if σ is erasing, then $\{\sigma(w) : w \in \mathcal{L}\}$ might not be infinite, and $\sigma(x)$, $x \in \mathcal{A}^{\mathbb{Z}}$, is potentially a finite word. Therefore, we will restrict ourselves to the non-erasing case.

Definition 1.10. Let σ be a non-erasing morphism. The *image of a language* $\mathcal{L}$ is defined as

$$\bigcup_{w\in\mathcal{L}} \mathrm{Fac}(\sigma(w)).$$

It is a language (in the sense given in the Warning of page 3) and we will abusively denote it $\sigma(\mathcal{L})$.

Similarly, the *image (induced by σ) of a shift space* X is the shift space

$$\{S^k(\sigma(x)) : x \in X, 0 \leq k < |\sigma(x_0)|\}$$

that we will abusively denote $\sigma(X)$.

Note that $\sigma(\mathcal{L}(x)) = \mathcal{L}(\sigma(x))$ and $\sigma(\mathcal{L}(X)) = \mathcal{L}(\sigma(X))$ for any infinite word x and any shift space X.

Definition 1.11. Let $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ be a primitive morphism. The *language generated by* σ is the set of factors of the words $\sigma^n(a)$, $n \geq 0$, $a \in \mathcal{A}$. It is a language by hypothesis on σ (if $\#\mathcal{A} \geq 2$). Moreover, if σ is prolongable on a, then the language generated by σ is the language of the fixed point

$$\lim_{n\to\infty} \sigma^n(a) = au\sigma(u)\sigma^2(u)\cdots$$

where $\sigma(a) = au$.

The two previous definitions can be slightly more general (by requiring weaker hypotheses on the morphism σ) but they will suffice in this work.

Let us consider the following language which will be a running example throughout this work.

Definition 1.12. Let $\varphi\colon \{0,1,2\}^* \to \{0,1,2\}^*$ be such that $\varphi(0) = 0012$, $\varphi(1) = 12$ and $\varphi(2) = 012$. The *(ternary) Chacon language* is the language generated by the morphism φ.

Since the morphism φ is primitive and prolongable on 0, the Chacon language is also the language of the sequence

$$\lim_{n\to\infty} \varphi^n(0).$$

Example 1.13. The first elements of the Chacon language $\mathcal{L}$ are described below:

$$\mathcal{L}_0 = \{\varepsilon\}, \quad \mathcal{L}_1 = \{0,1,2\}, \quad \mathcal{L}_2 = \{00, 01, 12, 20, 21\},$$

$$\mathcal{L}_3 = \{001, 012, 120, 121, 200, 201, 212\},$$

$$\mathcal{L}_4 = \{0012, 0120, 0121, 1200, 1201, 1212, 2001, 2012, 2120\},$$

and

$$\begin{aligned}\mathcal{L}_5 = \{&00120, 00121, 01200, 01212, 12001,\\ &12012, 12120, 20012, 20120, 20121, 21201\}.\end{aligned}$$

Factor complexity and extensions

Definition 1.14. The *factor complexity* of a language $\mathcal{L}$ is the map

$$p_{\mathcal{L}}(n)\colon \mathbb{N} \to \mathbb{N} \quad n \mapsto \#\mathcal{L}_n.$$

We will often also use the *first difference of complexity* which is the map $s_{\mathcal{L}}(n)\colon \mathbb{N} \to \mathbb{N}$ such that, for all $n \geq 0$,

$$s_{\mathcal{L}}(n) = p_{\mathcal{L}}(n+1) - p_{\mathcal{L}}(n).$$

Note that, by Example 1.13, we have $p_{\mathcal{L}}(n) = 2n+1$ for all $n \in \{0, \dots, 5\}$ for the Chacon language. This observation is more general as stated in the following result due to Ferenczi [Fog02, Chapter 5].

Proposition 1.15. *Let $\mathcal{L}$ be the Chacon language. For all $n \geq 0$, $s_{\mathcal{L}}(n) = 2$ and $p_{\mathcal{L}}(n) = 2n + 1$.*

The notion of extensions will be fundamental in this work. Indeed, all of the families of languages that we study can be defined using properties of their words' extensions.

Definition 1.16. Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and let $w \in \mathcal{L}$. The set of *left*, *right* and *bi-extensions* of w in $\mathcal{L}$ are defined respectively as follows

$$\begin{aligned} E^L_{\mathcal{L}}(w) &= \{a \in \mathcal{A} : aw \in \mathcal{L}\} \\ E^R_{\mathcal{L}}(w) &= \{a \in \mathcal{A} : wa \in \mathcal{L}\} \\ E_{\mathcal{L}}(w) &= \{(a, b) \in \mathcal{A} \times \mathcal{A} : awb \in \mathcal{L}\}. \end{aligned}$$

The word w is *left* (resp., *right*) *special* if $\#E^L_{\mathcal{L}}(w) \geq 2$ (resp., $\#E^R_{\mathcal{L}}(w) \geq 2$). It is *bispecial* if it is both left and right special.

Observe that the set of left (resp., right) extensions of the empty word in $\mathcal{L}$ will always be equal to $\mathcal{L}_1$, and the bi-extensions of ε correspond to $\mathcal{L}_2$.

Let us continue Example 1.13.

Example 1.17. Let $\mathcal{L}$ be the Chacon language. We describe below the left, right and bi-extensions of three particular (and well-chosen as will become clear in Example 1.45) words. For ε, we have

$$E^L_{\mathcal{L}}(\varepsilon) = \{0, 1, 2\}, \quad E^R_{\mathcal{L}}(\varepsilon) = \{0, 1, 2\},$$

and

$$E_{\mathcal{L}}(\varepsilon) = \{(0, 0), (0, 1), (1, 2), (2, 0), (2, 1)\}.$$

For 012, we have

$$E^L_{\mathcal{L}}(012) = \{0,2\}, \quad E^R_{\mathcal{L}}(012) = \{0,1\}$$

and

$$E_{\mathcal{L}}(012) = \{(0,0),(0,1),(2,0),(2,1)\}.$$

Lastly, for 120, we have

$$E^L_{\mathcal{L}}(120) = \{0,2\}, \quad E^R_{\mathcal{L}}(120) = \{0,1\}, \quad E_{\mathcal{L}}(120) = \{(0,0),(2,1)\}.$$

Studying the extensions of words also gives precious information on the factor complexity, as shown by Cassaigne [Cas97] (or [Cas96] in English).

Proposition 1.18. *Let $\mathcal{L}$ be a language. For all $n \geq 0$, we have*

$$s_{\mathcal{L}}(n) = \sum_{w \in \mathcal{L}_n} (\#E^R_{\mathcal{L}}(w) - 1) = \sum_{w \in \mathcal{L}_n} (\#E^L_{\mathcal{L}}(w) - 1).$$

1.2 Sturmian languages

The following fundamental result in combinatorics on words by Morse and Hedlund is at the origin of the study of Sturmian words.

Theorem 1.19 (Morse-Heldund [MH38])**.** *Let x be a right infinite word. The following are equivalent:*

1. *x is eventually periodic;*
2. *$p_x(n)$ is bounded;*
3. *there exists n such that $p_x(n) = p_x(n+1)$.*

In other words, if an infinite word x is not eventually periodic, then its factor complexity satisfies

$$p_x(n) \geq n + 1$$

for all $n \geq 0$. This led to the definition of Sturmian sequences as the aperiodic infinite words with the smallest factor complexity, i.e., the words x such that $p_x(n) = n + 1$ for all n. In particular, these words are on two letters and the Morse-Hedlund Theorem implies that they are recurrent.

The most famous Sturmian sequence is the Fibonacci word defined in the example below.

Example 1.20. Let $\sigma \colon \{0,1\}^* \to \{0,1\}^*$ be the morphism such that $\sigma(0) = 01$, $\sigma(1) = 0$. As σ is prolongable on 0, it admits a fixed point

$$f = \lim_{n\to\infty} \sigma^n(0).$$

This infinite word is called the *Fibonacci word* due to its link with the sequence of Fibonacci numbers.

Sturmian words are sometimes also defined using balancedness of their factors, as mechanical words or as codings of rotations for example. They admit many characterizations and have been vastly studied, see [Lot02, Chapter 2] and [Fog02, Chapter 6] for partial surveys on this topic.

By extension, we will define a Sturmian language as follows.

Definition 1.21. A language $\mathcal{L}$ is *Sturmian* if, for all $n \geq 0$, it satisfies $p_{\mathcal{L}}(n) = n + 1$.

Every recurrent Sturmian language corresponds to a Sturmian sequence. However, a Sturmian language is not necessarily recurrent and might therefore not be the language of a right infinite (Sturmian) sequence. This is the case of the language below which is also a recurring (counter-)example in this work.

Example 1.22. Let $\mathcal{L} = \{0^n : n \geq 0\} \cup \{0^n 1 0^m : n, m \geq 0\}$. It is a language and, for all $n \geq 1$, we have

$$\mathcal{L}_n = \{0^n\} \cup \{0^i 1 0^{n-1-i} : 0 \leq i \leq n-1\}$$

so $p_{\mathcal{L}}(n) = n+1$. By definition $\mathcal{L}$ is a Sturmian language. It is not recurrent since it is impossible to find a word containing two 1's. It is therefore not the language of a right infinite word. It is however the language of the bi-infinite word $\cdots 0001000 \cdots = {}^{\omega}010^{\omega}$.

Using Cassaigne's result on the first difference of complexity (Proposition 1.18), we obtain the following alternative definition.

Proposition 1.23. *Let $\mathcal{L}$ be a language. The following are equivalent:*

1. *$\mathcal{L}$ is Sturmian;*
2. *for all $n \geq 0$, $\mathcal{L}$ contains exactly one left special word of length n and it has two left extensions;*
3. *for all $n \geq 0$, $\mathcal{L}$ contains exactly one right special word of length n and it has two right extensions.*

1.3 Arnoux-Rauzy languages

In [AR91], Arnoux and Rauzy introduced a family of sequences on an alphabet of size 3 inspired by the Sturmian sequences, and more specifically, by Proposition 1.23. The definition of the so-called Arnoux-Rauzy sequences has been extended for an alphabet of any size as follows.

Definition 1.24. A language $\mathcal{L}$ over $\mathcal{A}$ is *Arnoux-Rauzy* if, for all $n \geq 0$, $\mathcal{L}$ contains exactly one left special factor of length n and this factor has $\#\mathcal{A}$ left extensions, and exactly one right special factor of length n and this factor has $\#\mathcal{A}$ right extensions.

A right infinite word x is an *Arnoux-Rauzy sequence* if it is recurrent and $\mathcal{L}(x)$ is Arnoux-Rauzy.

Arnoux-Rauzy sequences are also sometimes called strict episturmian sequences [JP02].

Note that, as in the definition of Sturmian languages, recurrence is not required for Arnoux-Rauzy languages.

Similarly to the Fibonacci language of Example 1.20, we can define the Tribonacci language over 3 letters. It is an Arnoux-Rauzy language.

Example 1.25. Let $\sigma\colon \{0,1,2\}^* \to \{0,1,2\}^*$ be the morphism such that $\sigma(0) = 01$, $\sigma(1) = 02$ and $\sigma(2) = 0$. The Tribonacci language is the language generated by σ, or equivalently, it is the language of the fixed point

$$t = \lim_{n\to\infty} \sigma^n(0).$$

Observe that

$$t = 0102010010201010 \cdots .$$

By definition, the Arnoux-Rauzy languages over an alphabet of size 2 are exactly the Sturmian languages by Proposition 1.23. More generally, if $\mathcal{L}$ is an Arnoux-Rauzy language over $\mathcal{A}$, then $p_{\mathcal{L}}(n) = (\#\mathcal{A} - 1)n + 1$ by Proposition 1.18. This is however not a characterization as the Chacon language has complexity $2n+1$ (Proposition 1.15) but is not Arnoux-Rauzy since, for example, it has two left special words of length 3 by Example 1.17.

To end this section, we recall the notion of ordinary words. It was introduced by Cassaigne in [Cas96] on an alphabet of size 2 and was later generalized in [CN10] for an alphabet of any size.

Definition 1.26. Let $\mathcal{L}$ be a language. A word $w \in \mathcal{L}$ is *ordinary* if there exist a, b such that

$$E_{\mathcal{L}}(w) = (E^L_{\mathcal{L}}(w) \times \{b\}) \cup (\{a\} \times E^R_{\mathcal{L}}(w)).$$

We can use this definition to give the following characterization of Arnoux-Rauzy languages.

Proposition 1.27. *A language $\mathcal{L}$ is Arnoux-Rauzy if and only if every $w \in \mathcal{L}$ is ordinary.*

1.4 Interval exchange transformations

Interval exchange transformations originate in the dynamical point of view of Sturmian words. Indeed, a Sturmian word can be seen as the coding of a rotation of irrational angle on the torus [MH40]. The idea is to replace this simple rotation by an exchange of several intervals, usually given by the interval lengths and a permutation of the set of intervals. We give here the (equivalent) definition based on interval lengths and two orders over the set of intervals due to its closer link with the combinatorial characterization that we detail in this section.

Definition 1.28. Let $\leq$ and $\preceq$ be two total orders on $\mathcal{A} = \{a_1, \dots, a_k\}$ and let k lengths $\lambda_{a_1}, \dots, \lambda_{a_k} > 0$ be such that $\sum_{i=1}^{k} \lambda_i = 1$. The associated *interval exchange transformation* (or IET for short) is the bijective map $T\colon [0,1[\to [0,1[$ such that

$$T(z) = z - \sum_{a_j < a_i} \lambda_{a_j} + \sum_{a_j \prec a_i} \lambda_{a_j} \quad \text{if } z \in \left[\sum_{a_j < a_i} \lambda_{a_j}, \sum_{a_j \leq a_i} \lambda_{a_j} \right[.$$

In other words, it is the piecewise translation mapping the length-λ_{a_i} interval $I_{a_i} := \left[\sum_{a_j < a_i} \lambda_{a_j}, \sum_{a_j \leq a_i} \lambda_{a_j} \right[$ to the interval $\left[\sum_{a_j \prec a_i} \lambda_{a_j}, \sum_{a_j \preceq a_i} \lambda_{a_j} \right[$.

Example 1.29. In Figure 1.1, we have represented the first few (positive and negative) iterations starting from the point $z = 0.07$ for an interval exchange transformation T corresponding to the orders $0 < 1 < 2$ and $2 \prec 1 \prec 0$, and some rationally independent lengths $\lambda_0 \sim 0.2$, $\lambda_1 \sim 0.27$ and $\lambda_2 \sim 0.53$. The second line represents an intermediary step to see the translations of intervals. This representation is the reason behind the notation $\binom{\leq}{\preceq}$ which is sometimes used to represent the two orders giving an IET.

In this work, we will focus on interval exchange transformations that satisfy an additional property.

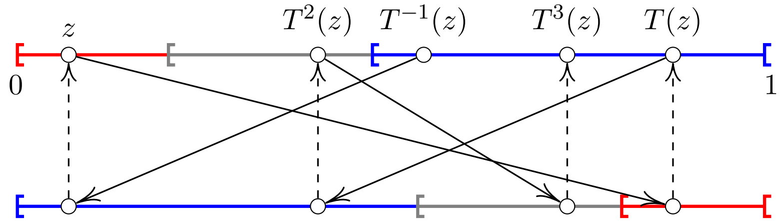


Figure 1.1: Example of the first iterations of an interval exchange transformation.

Definition 1.30. Let T be an interval exchange transformation and let us reuse the notations from Definition 1.28. If the orbits (under T) of the non-zero $\sum_{a_j<a_i} \lambda_{a_j}$, $i \in \{1, \dots, k\}$, are infinite and disjoint, then we say that T is a *regular interval exchange transformation*, or RIET for short.

Not every pair of orders can correspond to an RIET, as shown below.

Remark 1.31. If T is regular, then for all $0 < n < k$, the n smallest elements for the order $\leq$ cannot coincide with the n smallest elements for $\preceq$. Indeed, by contradiction, if these elements are denoted $b_1, \dots, b_n$ and if c (resp., d) is the $n+1$ smallest element for $\leq$ (resp., $\preceq$), then

$$\sum_{a_i<c} \lambda_{a_i} = \sum_{b_i} \lambda_{b_i} = \sum_{a_i \prec d} \lambda_{a_i}.$$

Therefore the orbits of the discontinuity points are not disjoint (if $c \neq d$) or infinite (if $c = d$). A pair of orders satisfying this condition is called *irreducible*.

To each interval exchange transformation, we can associate a language by coding the orbits of all points.

Definition 1.32. Let T be an IET and let us reuse the notations from Definition 1.28. The *(natural) coding* of a point $z \in [0, 1[$ is the bi-infinite word $x \in \mathcal{A}^{\mathbb{Z}}$ such that, for all $n \in \mathbb{Z}$, $x_n = a_i$ if and only if $T^n(z)$ is in the interval I_{a_i}. The *language* $\mathcal{L}(T)$ of the transformation T is the union of the languages of the codings of the points $z \in [0, 1[$.

Example 1.33. Let us continue Example 1.29. In Figure 1.1, the red interval corresponds to I_0, the gray one to I_1 and the blue one to I_2. Therefore, the coding x of $z = 0.07$ is such that $x_0 = 0$, $x_1 = 2$, $x_2 = 1$, $x_3 = 2$ and $x_{-1} = 2$.

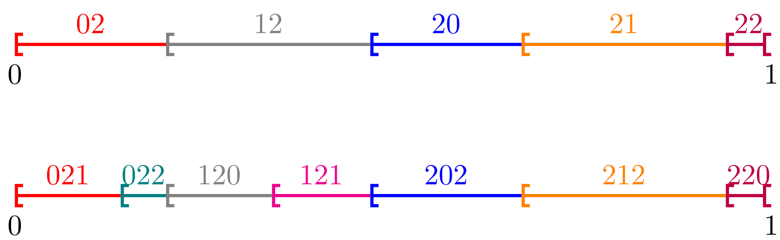


Figure 1.2: Partitions of $[0,1[$ given by the intervals I_w for w of length 2 (above) and length 3 (below).

Using the notations of Definition 1.32, we see that $w = x_{[m,m+n-1]}$ if and only if

$$T^m(z) \in I_{w_1} \cap T^{-1}(I_{w_2}) \cap \cdots T^{-n+1}(I_{w_n}).$$

For a length-n word w, we will thus denote

$$I_w = I_{w_1} \cap T^{-1}(I_{w_2}) \cap \cdots T^{-n+1}(I_{w_n}).$$

It is a (possibly empty) sub-interval of $[0,1[$. We then easily see that

$$\mathcal{L}(T) = \{w : I_w \neq \emptyset\}.$$

Observe that, for a fixed length n, the non-empty intervals I_w, $|w| = n$, form a partition of $[0,1[$.

Example 1.34. Starting once again from Example 1.29, we have represented in Figure 1.2 the partitions given by the intervals corresponding to words of length 2 and 3. We then have

$$\mathcal{L}(T)_2 = \{02, 12, 20, 21, 22\}$$

and

$$\mathcal{L}(T)_3 = \{021, 022, 120, 121, 202, 212, 220\}.$$

A famous result by Keane [Kea75] states that if an interval exchange transformation is regular, then the orbit of each $z \in [0,1[$ is dense. In that case, the language of the RIET is given by the language of the coding of any point. Moreover, the language is then uniformly recurrent.

For this work, it will often be more practical to study regular interval exchanges through a more combinatorial point of view. Namely, we will use the following characterization of languages of regular interval exchange transformations by Ferenczi and Zamboni [FZ08].

Theorem 1.35. *A language $\mathcal{L}$ over $\mathcal{A}$ is the language of a regular interval exchange transformation over $\mathcal{A}$ with the orders $\binom{\leq}{\preceq}$ if and only if $\mathcal{L}$ is uniformly recurrent and, for each $w \in \mathcal{L}$, we have*

1. *$E^L_{\mathcal{L}}(w)$ is an interval for $\preceq$ and $E^R_{\mathcal{L}}(w)$ is an interval for $\leq$;*
2. *for all $(a_1, b_1), (a_2, b_2) \in E_{\mathcal{L}}(w)$, if $a_1 \prec a_2$, then $b_1 \leq b_2$;*
3. *for all $a_1, a_2 \in E^L_{\mathcal{L}}(w)$, if a_1, a_2 are consecutive for $\preceq$, then $E^R_{\mathcal{L}}(a_1 w) \cap E^R_{\mathcal{L}}(a_2 w)$ is a singleton.*

An interval for $\preceq$ (resp., for $\leq$) is understood here as a set I for which there exist two letters $a, b \in \mathcal{A}$ such that $I = \{c \in \mathcal{A} : a \preceq c \preceq b\}$ (resp., $I = \{c \in \mathcal{A} : a \leq c \leq b\}$).

Example 1.36. We consider the interval exchange transformation given in Example 1.29. Recall that the orders are $0 < 1 < 2$ and $2 \prec 1 \prec 0$. Since the pair of orders is irreducible and we assumed that the interval lengths were rationally independent, the interval exchange transformation T is regular. The small words in its language are given by Example 1.34. Let us check Theorem 1.35 for the empty word. Observe that Condition 1 is always satisfied for ε, no matter the language $\mathcal{L}$. For Condition 2, we need to check that $E^R_{\mathcal{L}(T)}(0) \preceq E^R_{\mathcal{L}(T)}(1) \preceq E^R_{\mathcal{L}(T)}(2)$ in the sense that the inequality is true for any choice of elements in the sets. Since $E^R_{\mathcal{L}(T)}(0) = \{2\}$, $E^R_{\mathcal{L}(T)}(1) = \{2\}$ and $E^R_{\mathcal{L}(T)}(2) = \{0, 1, 2\}$, this is satisfied. Finally, for Condition 3, we can see that 2 is in all the sets of right extensions and is the only common letter for any two sets. We can similarly check the conditions for the words 0, 1 and 2 using $\mathcal{L}(T)_3$. Notice that, for 0 and 1, the conditions are trivially satisfied since they are neither left nor right special.

Observe that Condition 1 is redundant as it is implied by Condition 2 if we consider all $w \in \mathcal{L}$. Indeed, assume by contrary that w is the shortest word such that $E^L_{\mathcal{L}}(w)$ is not an interval for $\preceq$ and let $a_1 \prec a_2 \prec a_3$ be such that $a_1, a_3 \in E^L_{\mathcal{L}}(w)$ but $a_2 \notin E^L_{\mathcal{L}}(w)$. As $w \neq \varepsilon$, let us denote $w = w'b$ for some letter b. By minimality of w, we have $a_2 \in E^L_{\mathcal{L}}(w')$. But using Condition 2 for w', the only possible right extension of $a_2 w'$ is b. So, a_2 is in $E^L_{\mathcal{L}}(w'b)$, which contradicts its definition. The proof of the condition for $E^R_{\mathcal{L}}(w)$ is similar.

Observe also that, for Condition 3, it is sufficient to ask that, for all $a_1, a_2 \in E^L_{\mathcal{L}}(w)$ consecutive, the set $E^R_{\mathcal{L}}(a_1 w) \cap E^R_{\mathcal{L}}(a_2 w)$ is non-empty. Indeed, by Condition 2, this set can never contain two or more elements.

We can therefore replace Theorem 1.35 by the following simplified version.

Theorem 1.37. *A language $\mathcal{L}$ over $\mathcal{A}$ is the language of a regular interval exchange transformation over $\mathcal{A}$ with the orders $\binom{\leq}{\preceq}$ if and only if $\mathcal{L}$ is uniformly recurrent and, for each $w \in \mathcal{L}$, we have*

1. *for all $(a_1, b_1), (a_2, b_2) \in E_{\mathcal{L}}(w)$, if $a_1 \prec a_2$, then $b_1 \leq b_2$;*
2. *for all $a_1, a_2 \in E_x^L(w)$, if a_1, a_2 are consecutive for $\preceq$, then $E_x^R(a_1 w) \cap E_x^R(a_2 w)$ is non-empty.*

A well-known property of languages of RIET is that the long enough left (resp., right) special factors only have two left (resp., right) extensions. In fact, we show here that it is sufficient to satisfy Condition 1 of Theorem 1.37 eventually to obtain this result.

Proposition 1.38. *Let $\mathcal{L}$ be a uniformly recurrent language and let $N \geq 0$ such that every $w \in \mathcal{L}_{\geq N}$ satisfies Condition 1 of Theorem 1.37. There are only finitely many words $w \in \mathcal{L}$ such that $\#E_{\mathcal{L}}^L(w) \geq 3$ (resp., $\#E_{\mathcal{L}}^R(w) \geq 3$).*

Proof. By contrary, assume that there are infinitely many words having k left extensions for some $k \geq 3$ and let us denote W the set of such words. Let $u \in \mathcal{L}_{\geq N}$ be a prefix of infinitely many elements of W. By definition of u, it has at least k left extensions and, up to taking a longer u, we can assume that it has exactly k left extensions. Therefore, there exists a right extension b of u such that ub is a prefix of infinitely many elements of W, and in particular, ub has at least k left extensions. This implies that $E_{\mathcal{L}}^L(ub) = E_{\mathcal{L}}^L(u)$. As $k \geq 3$, let $a_1 \prec a_2 \prec a_3$ be three left extensions of u. In particular, u has the bi-extensions (a_1, b), (a_2, b) and (a_3, b). By Condition 1 on u, the only right extension of $a_2 u$ is b. In particular, $a_2 u$ is not right special.

We can iterate the reasoning on ub, then ubb_2, etc. to show that no word beginning with $a_2 u$ is right special. As $\mathcal{L}$ is uniformly recurrent, any long enough word contains the factor $a_2 u$ and is therefore not right special. Using Proposition 1.18, this implies that long enough factors are also not left special, which contradicts the hypothesis on W. □

As a consequence, we have a complete description of the sets of left (resp., right) extensions of the long enough words in the language of an RIET.

Proposition 1.39. *Let $\mathcal{L}$ be the language of a regular interval exchange transformation with the orders $\binom{\leq}{\preceq}$.*

1. *There exists N such that for every left special word $w \in \mathcal{L}_{\geq N}$, we have $E_{\mathcal{L}}^L(w) = \{a, b\}$ for two letters a, b consecutive for $\preceq$.*

2. *For every two letters a, b consecutive for $\preceq$ and for all $n \geq 0$, there exists a (unique) word $w \in \mathcal{L}_n$ such that $a, b \in E^L_{\mathcal{L}}(w)$.*

There is a similar result for right extensions and the order $\leq$.

Proof. The first claim is a direct consequence of Proposition 1.38 and of Condition 1 from Theorem 1.35. The second claim can be shown by induction. It is directly true for $n = 0$. For the induction, if w is the unique length-n word such that $a, b \in E^L_{\mathcal{L}}(w)$, then by Condition 3 from Theorem 1.35, there exists c such that $E^R_{\mathcal{L}}(aw) \cap E^R_{\mathcal{L}}(bw) = \{c\}$, or in other words, wc is the unique length-$(n+1)$ word such that $a, b \in E^L_{\mathcal{L}}(wc)$. □

We can use this to show that the pair of orders given in Theorem 1.37 is unique.

Proposition 1.40. *Let $\mathcal{L}$ be a uniformly recurrent language. If $\mathcal{L}$ satisfies Conditions 1 and 2 of Theorem 1.37, then the corresponding pair of orders $\binom{\leq}{\preceq}$ is unique, up to reversal.*

Proof. Assume that $\mathcal{L}$ satisfies Conditions 1 and 2 of Theorem 1.37 for a pair of orders. Using Theorem 1.37 and Proposition 1.39, the pairs of consecutive letters for the orders are entirely determined by $\mathcal{L}$. In other words, for each order, there are two candidates which are reversal of one another. Let us denote $\leq^*$ the reversal of an order $\leq$, i.e.,

$$a \leq b \iff b \leq^* a.$$

Assume that $\mathcal{L}$ satisfies the conditions for $\binom{\leq}{\preceq}$. We therefore have only four candidates for the pairs of orders: $\binom{\leq}{\preceq}$, $\binom{\leq^*}{\preceq}$, $\binom{\leq}{\preceq^*}$, and $\binom{\leq^*}{\preceq^*}$. It is easy to check that if we reverse both orders, then the conditions are still true, i.e., $\mathcal{L}$ satisfies the conditions for $\binom{\leq^*}{\preceq^*}$. However, if we only reverse one order they are not. Indeed, let a_m (resp., b_m) denote the smallest letter for $\preceq$ (resp., $\leq$) and a_M (resp., b_M) denote the largest letter for the same order. By Condition 1 of Theorem 1.37 for $(\leq, \preceq)$, we have $(a_m, b_m) \in E_{\mathcal{L}}(\varepsilon)$. Similarly, we have $(a_M, b_M) \in E_{\mathcal{L}}(\varepsilon)$. This shows that Condition 1 is not satisfied for $\binom{\leq^*}{\preceq}$ and $\binom{\leq}{\preceq^*}$. □

1.5 Weak, strong and neutral words

The families introduced in the previous sections are some of the most well-known. We now turn to slightly lesser-known families defined using properties of their elements' extensions. These notions were studied in [CN10] for their link with the factor complexity.

Definition 1.41. Let $\mathcal{L}$ be a language and let $w \in \mathcal{L}$. The *multiplicity* of w (in $\mathcal{L}$) is defined as

$$m_{\mathcal{L}}(w) = \#E_{\mathcal{L}}(w) - \#E^L_{\mathcal{L}}(w) - \#E^R_{\mathcal{L}}(w) + 1.$$

We say that w is *weak* if $m_{\mathcal{L}}(w) < 0$, *neutral* if $m_{\mathcal{L}}(w) = 0$ and *strong* if $m_{\mathcal{L}}(w) > 0$.

Example 1.42. Let us go back to the Chacon language $\mathcal{L}$. Using Example 1.17, we see that

$$\begin{aligned} m_{\mathcal{L}}(\varepsilon) &= 5 - 3 - 3 + 1 = 0 \\ m_{\mathcal{L}}(012) &= 4 - 2 - 2 + 1 = 1 \\ m_{\mathcal{L}}(120) &= 2 - 2 - 2 + 1 = -1 \end{aligned}$$

therefore ε is neutral, 012 is strong and 120 is weak.

The motivation behind the study of the multiplicity is its close link with the first difference of complexity, as obtained as a direct corollary of Proposition 1.18.

Corollary 1.43. *Let $\mathcal{L}$ be a language. For all $n \geq 0$, we have*

$$s_{\mathcal{L}}(n+1) - s_{\mathcal{L}}(n) = \sum_{w \in \mathcal{L}_n} m_{\mathcal{L}}(w).$$

This statement is more practical when multiplicities cannot cancel each other out, meaning that we will be interested in languages whose elements are all neutral (resp., weak or neutral; resp., strong or neutral). By extension, such a language will also be called *neutral* (resp., *weak or neutral*; resp., *strong or neutral*). This hypothesis is sometimes too strong and we will instead require that only elements of length at least N are neutral (resp., weak or neutral; resp., strong or neutral) and call the corresponding language *eventually neutral* (resp., *eventually weak or neutral*; resp., *eventually strong or neutral*). The minimal such N is then called the *threshold*.

Using Corollary 1.43, we can directly deduce information on the factor complexity of such languages. We only give here the most useful ones for us, as will become clear later on.

Corollary 1.44. *Let $\mathcal{L}$ be a language over an alphabet of size k.*

1. *If $\mathcal{L}$ is neutral then, for all $n \geq 0$,*

$$p_{\mathcal{L}}(n) = (k-1)n + 1.$$

2. *If $\mathcal{L}$ is eventually neutral of threshold N then there exist $K \geq 0$ and $C \in \mathbb{Z}$ such that, for all $n \geq N$*

$$p_{\mathcal{L}}(n) = Kn + C.$$

The previous result is not a characterization as shown by the Chacon language. Indeed, recall that, by Proposition 1.15, its factor complexity is given by $2n+1$ for all $n \geq 0$. However, the Chacon language is not neutral by Example 1.42, and it is not even eventually neutral, nor is it eventually weak or neutral, or eventually strong or neutral as shown below.

Example 1.45 (Dolce–Perrin [DP21, Example 3.4])**.** Let us show that the Chacon language $\mathcal{L}$ contains infinitely many weak (resp., strong) words. By Example 1.42, 120 is weak and 012 is strong. Their extensions are given in Example 1.17. Observe in particular that 1 (resp., 2) is a left (resp., right) extension of neither of them. However, a simple computation shows that if $E^L_{\mathcal{L}}(w) \subseteq \{0, 2\}$ and $E^R_{\mathcal{L}}(w) \subseteq \{0, 1\}$, then $E_{\mathcal{L}}(w) = E_{\mathcal{L}}(012\varphi(w))$ where φ is the morphism defining the Chacon language, i.e., $\varphi(0) = 0012$, $\varphi(1) = 12$ and $\varphi(2) = 012$.

This implies that, from the weak word 120, we can find infinitely many weak words in $\mathcal{L}$ by iterating $w \mapsto 012\varphi(w)$. We reach the same conclusion for strong words. Observe that, as 012 and 120 share the same letters, each weak word built like this has the same length as a strong word defined using this construction. This shows that there is indeed a cancellation of the multiplicities when taking the sum over the words of a given length.

Chapter 2

Dendricity and co

We now turn to the main actors of this work, namely the dendric and eventually dendric languages. The study of dendric words began a decade ago in a series of papers [BDFD$^+$15a, BDFD$^+$15c, BDFD$^+$15d] under the name of tree words. In this chapter, we only give a soft introduction to these words. Some of the main properties of dendric languages such as the Return Theorem (Theorem 3.33) and the stability under derivation (Corollary 3.44) will be recalled in Chapter 3. The curious reader can of course find many other interesting results in the original papers such as the stability under bifix decoding.

Instead of just studying the number of left, right and bi-extensions as done through the multiplicity, one can sometimes want more information on the relation between these sets. Indeed, for any language $\mathcal{L}$ and any word w, we have by definition $E_{\mathcal{L}}(w) \subseteq E^L_{\mathcal{L}}(w) \times E^R_{\mathcal{L}}(w)$. This means that the bi-extensions define a relation between the left and the right extensions. In mathematics, there are many ways of representing a relation between two finite sets, one of them being the use of a bipartite graph. This led to the definition of extension graphs.

Definition 2.1. Let $\mathcal{L}$ be a language and let $w \in \mathcal{L}$. The *extension graph* of w (in $\mathcal{L}$) is the bipartite graph $\mathcal{E}_{\mathcal{L}}(w)$ whose set of vertices is the disjoint union of $E^L_{\mathcal{L}}(w)$ and $E^R_{\mathcal{L}}(w)$, and such that there is an edge between $a \in E^L_{\mathcal{L}}(w)$ and $b \in E^R_{\mathcal{L}}(w)$ if and only if $(a, b) \in E_{\mathcal{L}}(w)$.

These graphs will be represented with the vertices from $E^L_{\mathcal{L}}(w)$ in a column on the left (they will sometimes be called *left vertices*) and the vertices from $E^R_{\mathcal{L}}(w)$ in a column on the right (we will call them *right vertices*). When needed, we use the terminologies of left vertex a (denoted a^L) and

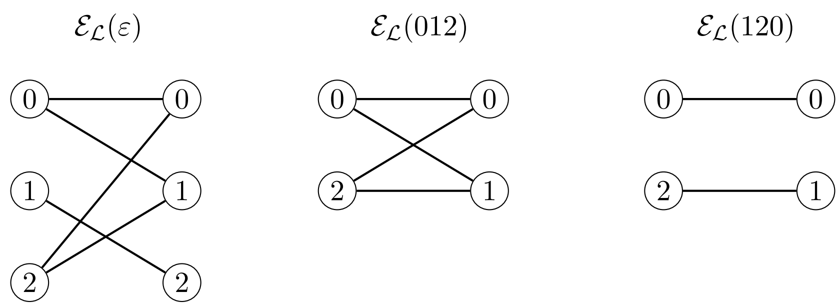


Figure 2.1: The extension graphs of ε (on the left), 012 (in the center) and 120 (on the right) in the Chacon language $\mathcal{L}$.

right vertex a (denoted a^R) to distinguish the two possible vertices labeled by the letter a.

Example 2.2. Let $\mathcal{L}$ be the Chacon language. Using the extensions of ε, 012 and 120 found in Example 1.17, we obtain the extension graphs of Figure 2.1.

As we have now associated a graph with each word, we can define families of words based on the properties of their corresponding graphs. These notions were introduced in [BDFD$^+$15a].

Definition 2.3. Let $\mathcal{L}$ be a language. A word $w \in \mathcal{L}$ is *acyclic* (resp., *connected*; resp., *dendric*) if its extension graph is acyclic (resp., connected; resp., a tree).

Example 2.4. Coming back to Example 2.2, we see that ε is neither acyclic nor connected in the Chacon language, but 012 is connected (not acyclic) and 120 is acyclic (not connected).

Observe that, as a language is always assumed to be biextendable, the extension graphs cannot have isolated vertices, which would not be the case when considering the factors of a non-recurrent one-sided infinite word for example. In particular, this implies that if $\#E^L_{\mathcal{L}}(w) = 1$ or $\#E^R_{\mathcal{L}}(w) = 1$, then the extension graph $\mathcal{E}_{\mathcal{L}}(w)$ is a tree, or in other words, we have the following direct result.

Lemma 2.5. *Let $\mathcal{L}$ be a language and $w \in \mathcal{L}$. If w is not bispecial in $\mathcal{L}$, then w is dendric in $\mathcal{L}$.*

As with neutral languages, we extend these notions to languages by saying that a language is *dendric* (resp., *acyclic*; resp., *connected*) if all of its elements are.

This chapter is organized as follows.

In Section 2.1, we study the link between dendric languages and the languages introduced in Chapter 1. We summarize the different relations in a diagram (Figure 2.2). In particular, many of the families defined before provide examples of dendric languages. We will give a method to generate infinitely many other examples in Chapter 5.

Afterwards, we turn to the link between dendricity and topological conjugacy and show that dendricity is not a dynamical property in Section 2.2. This was the main motivation for the introduction of eventually dendric languages by Dolce and Perrin.

The family of eventually dendric languages appears as the natural "eventual" closure of many other properties already mentioned in this work, or small variations of them. In Section 2.3, we study the many alternative ways of defining eventually dendric languages.

Finally, in Section 2.4, we introduce a new family of graphs which will be a key concept behind many results in Chapter 4 and Chapter 5. These graphs provide a new way of defining dendric languages (Corollary 2.44). They can also be used to characterize eventually dendric languages as shown in Proposition 2.61.

2.1 Dendricity and the families of Chapter 1

The purpose of this section is to list the inclusions between dendric languages and the other families of languages defined in Chapter 1. These relations are summarized in Figure 2.2.

Dendric languages were originally introduced to generalize some properties of Sturmian and Arnoux-Rauzy languages. It is therefore quite natural that these families are examples of dendric languages. To convince ourselves of this, it suffices to observe that any ordinary word is dendric. In fact, the extension graph of a bispecial ordinary word is a particular type of tree as its diameter is equal to 3, such a graph is called a *simple tree* in [DP21]. By Proposition 1.27, we conclude that any Sturmian or Arnoux-Rauzy language is dendric.

For languages of regular interval exchanges, it might not be as easy to see that they are dendric from their original definition. However, it can be deduced from the combinatorial characterization of Ferenczi and Zamboni.

Indeed, let us recall Condition 1 of Theorem 1.37. If $\mathcal{L}$ is the language of an RIET with orders $\binom{\leq}{\preceq}$, then for all $w \in \mathcal{L}$ and for all $(a_1, b_1), (a_2, b_2) \in E_{\mathcal{L}}(w)$, if $a_1 \prec a_2$, then $b_1 \leq b_2$. In other words, if we order the left vertices of $\mathcal{E}_{\mathcal{L}}(w)$ according to $\preceq$ and the right vertices according to $\leq$, then we can draw the edges as straight non-crossing segments. This leads to the following definition.

Definition 2.6. A bipartite graph is *planar for the orders* $\leq_1$ *and* $\leq_2$ if, whenever we place the left vertices on a line according to $\leq_1$, the right vertices on a parallel line according to $\leq_2$, and we draw the edges as straight segments, then the edges do not cross in this representation.

By extension, we will say that a word is *planar* for some orders if its extension graph is planar for these orders.

Example 2.7. Among the extension graphs represented in Figure 2.1, the first two contain a cycle and can therefore not be planar. The extension graph of 120 however is planar for any pair of orders $(\leq_1, \leq_2)$ such that $0 <_1 2$ if and only if $0 <_2 1$.

Theorem 1.37 can then be restated as follows.

Proposition 2.8. *A language $\mathcal{L}$ over $\mathcal{A}$ is the language of a regular interval exchange transformation over $\mathcal{A}$ with the orders $\binom{\leq}{\preceq}$ if and only if $\mathcal{L}$ is uniformly recurrent and, for each $w \in \mathcal{L}$, we have*

1. *w is planar for $\preceq$ and $\leq$;*
2. *w is connected.*

We invite the reader to pay attention to the orders here. Indeed, while we talk about an RIET for the orders $\binom{\leq}{\preceq}$, the words are planar for the pair of orders $(\preceq, \leq)$.

As planarity of a bipartite graph implies in particular that this graph is acyclic, we deduce that languages of RIET are indeed dendric.

Observe that while Arnoux-Rauzy languages and languages of RIET are both dendric generalizations of Sturmian languages, they are disjoint families as soon as the alphabet is of size at least 3. This can be seen by looking at the number of left (or right) extensions of the long enough special words. Indeed, in an Arnoux-Rauzy language, they will always have a number of extensions equal to the size of the alphabet whereas, in the language of an RIET, they will have two extensions by Proposition 1.38.

There is also a link between acyclic, connected and dendric words, and weak, strong and neutral words. This is a consequence of the following classical result from graph theory.

Lemma 2.9. *Let G be a graph with v vertices and e edges. If G has c connected components, then*

$$e - v + c \geq 0$$

and the equality occurs exactly when G is acyclic.

Proposition 2.10. *Let $\mathcal{L}$ be a language and $w \in \mathcal{L}$.*

1. *If w is acyclic, then it is weak or neutral.*

2. *If w is connected, then it is strong or neutral.*

3. *If w is dendric, then it is neutral.*

4. *The word w is weak or neutral, and connected, if and only if it is dendric.*

5. *The word w is strong or neutral, and acyclic, if and only if it is dendric.*

Proof. The result directly follows from Lemma 2.9 and the observation that, in the extension graph $\mathcal{E}_{\mathcal{L}}(w)$, the number of vertices is given by $\#E^L_{\mathcal{L}}(w) + \#E^R_{\mathcal{L}}(w)$ and the number of edges is $\#E_{\mathcal{L}}(w)$. □

The converses of Assertions 1, 2 and 3 of the previous proposition are true on an alphabet of size 2 but are false if the alphabet contains at least three letters. For exemple, in the Chacon language, the empty word is neutral but is neither acyclic, nor connected (Examples 1.42 and 2.4).

Using all of the observations made in this section, we can represent the interactions between the families in a diagram as done in Figure 2.2.

There is one last known relation between dendric languages and other families of languages, this time it is a "negative" result.

Proposition 2.11 (Berthé *et al.* [BDD$^+$18])**.** *Let $\mathcal{L}$ be a recurrent dendric language over an alphabet of size at least 2. Then $\mathcal{L}$ cannot be generated by a primitive k-uniform morphism, nor can it be the language of a Toeplitz bi-infinite word.*

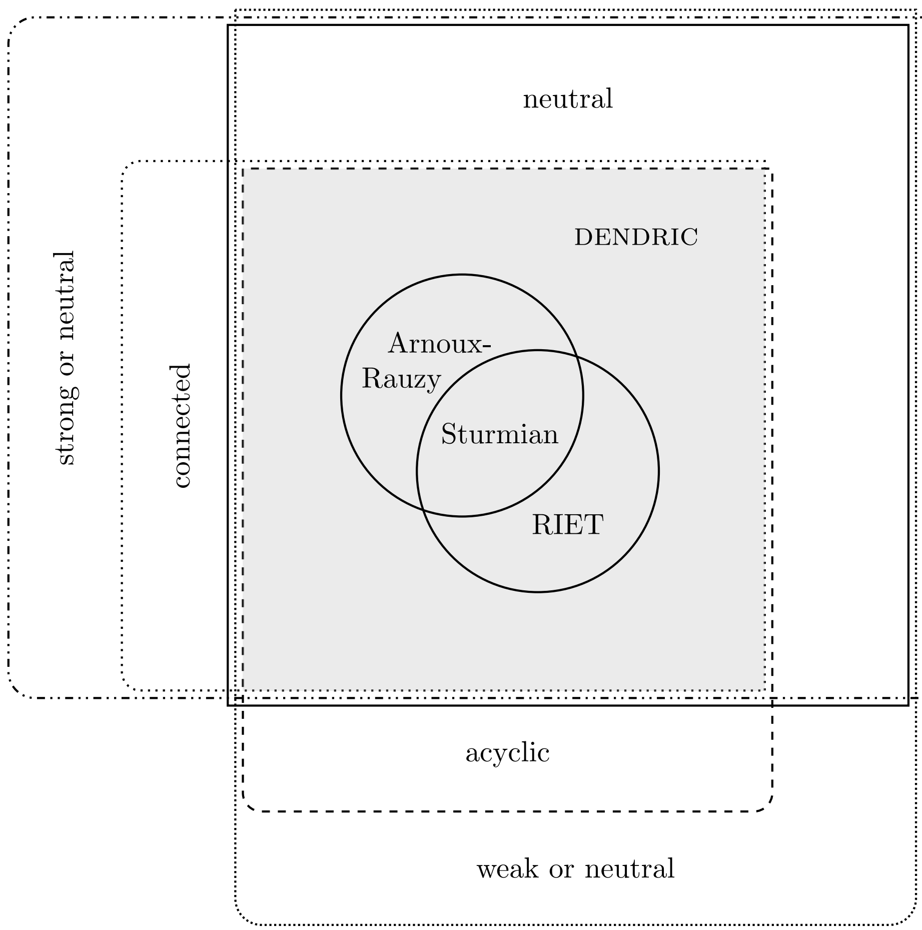


Figure 2.2: Interactions between the families defined in Chapter 1 and in Chapter 2. For readability, lines that are close together should be understood as overlapping.

A morphism is *k-uniform* if the images of the letters are all of length k. The fixed points of such morphisms are particular examples of *k-automatic sequences* which form a deeply studied families of words, see [AS03] for example.

On the other hand, a bi-infinite word x is *Toeplitz* if for all $n \in \mathbb{Z}$, there exists $p \geq 1$ such that $x_n = x_{n+kp}$ for all $k \in \mathbb{Z}$. Conceptually, every finite factor of x appears periodically. This does not however necessarily imply that x itself is periodic. For a survey on Toeplitz sequences, see [Dow05].

2.2 Dendricity and conjugacy

Shortly after its introduction, people realized that dendricity is not a dynamical property as the family of dendric shift spaces is not stable under topological conjugacy. We detail this observation here.

Conjugacy is to shift spaces what isomorphism is to groups, or homeomorphism is to topological spaces. In other words, if we can go from one shift space to another with a conjugacy then the two shift spaces can be seen as essentially equal. The question of whether two given shift spaces are conjugate or not has interested researchers for many years and, in most cases, only admits partial answers using invariants. It was however recently shown that this question is decidable for substitutive shift spaces [DL22].

Definition 2.12. Let X and Y be two shift spaces. A map $\varphi\colon X \to Y$ is a *(topological) factor map* if it is continuous, surjective and commutes with the dynamics, i.e., $\varphi \circ S_X = S_Y \circ \varphi$ where S_X (resp., S_Y) is the shift map on X (resp., Y). We then say that Y is a *(topological) factor* of X.

If, moreover, φ is a bijection, then we say that it is a *conjugacy* and that X and Y are *(topologically) conjugate*.

From the symbolic dynamics viewpoint, a "good" property is then a property such that, if a shift space satisfies it, then so does any of its conjugates. In that case, we say that it is a *dynamical property*. Unfortunately, one can easily observe that dendricity is not a dynamical property. To see it, we use the other point of view of conjugacies given by the so-called Curtis-Hedlund-Lyndon Theorem (see [LM95] for example).

Theorem 2.13. *Let $X \subseteq \mathcal{A}^{\mathbb{Z}}$ and $Y \subseteq \mathcal{B}^{\mathbb{Z}}$ be two shift spaces. A map $\varphi\colon X \to Y$ is a factor map if and only if there exist $s, r \geq 0$ and a map $f\colon \mathcal{L}_{s+r+1}(X) \to \mathcal{B}$ such that, for all $x \in X$ and for all $n \in \mathbb{Z}$, we have $\varphi(x)_n = f(x_{[n-s,n+r]})$.*

Since $\varphi(X) = S^{-s} \circ \varphi(X)$ and φ is a factor map if and only if $S^{-s} \circ \varphi$ is, we will assume that $s = 0$ to ease the notations.

In some way, the Curtis-Hedlund-Lyndon Theorem states that a factor map can be decomposed into two steps: first we map X to $\gamma(X)$ where, for all $x \in X$ and $n \in \mathbb{Z}$, $\gamma(x)_n = x_{[n,n+r]}$ seen as a letter of the alphabet $\mathcal{A}^{r+1}$, then we apply the map f, now seen as a letter-to-letter morphism. This intermediary shift space is called a *higher block shift space*.

Definition 2.14. Let $X \subseteq \mathcal{A}^{\mathbb{Z}}$ be a shift space and let $N \geq 1$. The *N-th higher block shift space* of X is the shift space $\gamma(X)$ over the alphabet $\mathcal{A}^N$ where

$$\gamma(x)_n = \begin{pmatrix} x_n \\ \vdots \\ x_{n+N-1} \end{pmatrix}$$

for all $x \in X$, $n \in \mathbb{Z}$. We then denote $X^{(N)} = \gamma(X)$.

It is well known that $X^{(N)}$ is indeed a shift space and, moreover, it is a conjugate of X. In some sense, the maps γ defined as in Definition 2.14 are the simplest conjugacies. We also directly see that $X^{(1)} = X$.

Example 2.15. Let X be the Tribonacci shift space (corresponding to the Tribonacci language of Example 1.25). The 2-nd higher block shift space is on the alphabet $\left\{\binom{0}{0}, \binom{0}{1}, \binom{0}{2}, \binom{1}{0}, \binom{2}{0}\right\}$. Let γ denote the map of Definition 2.14. We then have, for example,

$$\gamma(\cdots 10 \cdot 01020 \cdots) = \cdots \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} \cdots .$$

The extension graphs in the higher block shift spaces are closely related to the extension graphs in the original shift space. This link is described in the following result.

Proposition 2.16. *Let X be a shift space, let $N \geq 1$ and let $w \in \mathcal{L}(X^{(N)})$.*

1. *If $w = \varepsilon$, then*
$$\mathcal{E}_{X^{(N)}}(w) \cong \bigsqcup_{v \in \mathcal{L}_{N-1}(X)} \mathcal{E}_X(v).$$

2. *If* $w = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \begin{pmatrix} v_2 \\ \vdots \\ v_{N+1} \end{pmatrix} \cdots \begin{pmatrix} v_n \\ \vdots \\ v_{N+n-1} \end{pmatrix}$ *for $n \geq 1$, then*
$$\mathcal{E}_{X^{(N)}}(w) \cong \mathcal{E}_X(v_1 v_2 \cdots v_{N+n-1}).$$

Proof. Let $\left(\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}, \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}\right)$ be a bi-extension of w.

1. If $w = \varepsilon$, then it is equivalent to say that $\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}\begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}$ is a factor of $X^{(N)}$, or in other words that $a_2 \cdots a_N = b_1 \cdots b_{N-1} \in \mathcal{L}_{N-1}$ and $(a_1, b_N) \in \mathcal{E}_X(a_2 \cdots a_N)$. Therefore, $\mathcal{E}_{X^{(N)}}(\varepsilon)$ is made of copies of $\mathcal{E}_X(u)$, $u \in \mathcal{L}_{N-1}$, and these copies are disjoint as u is the suffix (resp., prefix) of the corresponding left (resp., right) vertices.

2. If $w = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}\begin{pmatrix} v_2 \\ \vdots \\ v_{N+1} \end{pmatrix} \cdots \begin{pmatrix} v_n \\ \vdots \\ v_{N+n-1} \end{pmatrix}$, then $\left(\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}, \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}\right)$ is an extension of w if and only if we have $a_2 \cdots a_N = v_1 \cdots v_{N-1}$, $b_1 \cdots b_{N-1} = v_{n+1} \cdots v_{N+n-1}$ and $(a_1, b_N) \in \mathcal{E}_X(v_1 \cdots v_{N+n-1})$. We directly deduce the link between the extension graphs.

□

As a consequence, if X is a shift space over an alphabet of size at least 2 and $N \geq 2$, then ε is not connected in $X^{(N)}$. This shows that the family of dendric shift spaces is not closed under conjugacy.

2.3 Eventual dendricity and other eventual properties

The observation done in the previous section was the main motivation for the study of eventual dendricity initiated by Dolce and Perrin [DP21], following a suggestion by Durand.

Definition 2.17. A language $\mathcal{L}$ is *eventually dendric* if there exists $N \geq 0$ such that every $w \in \mathcal{L}_{\geq N}$ is dendric. The minimal such N is called the *threshold*.

If the threshold is 0, we recover the notion of dendric languages.

Example 2.18. For the Chacon language, Example 1.45 and Proposition 2.10 show that it contains infinitely many words which are not acyclic (resp., connected). Therefore, this language is not eventually dendric.

Using Proposition 2.16, we then directly have the following observation.

Corollary 2.19. *Let X be an eventually dendric shift space of threshold N over an alphabet of size at least 2 and let $M \geq 1$. Then $X^{(M)}$ is eventually dendric of threshold* $\max\{1, N - M + 1\}$*. In particular, every eventually dendric shift space has an eventually dendric conjugate of threshold* 1*.*

Using this observation, Dolce and Perrin proved the following theorem stating that eventual dendricity is a dynamical property.

Theorem 2.20 (Dolce–Perrin [DP21])**.** *The family of eventually dendric shift spaces is closed under topological conjugacy.*

We will come back to this result in Section 4.5 as it is closely linked to the question of stability when taking the image under a morphism.

However, stability under conjugacy is far from being the only property of eventually dendric shift spaces, and eventual dendricity turned out to be much more natural than originally thought. Indeed, it coincides with many other "eventual" properties, showing that the diagram represented in Figure 2.2 becomes much simpler when relaxing the restrictions on the small words. We present here an overview of these equivalences.

The first known equivalence was proved by Dolce and Perrin [DP21] between eventual dendricity and what seems to be, at first glance, a much stronger property.

Proposition 2.21 (Dolce-Perrin [DP21])**.** *Let $\mathcal{L}$ be a language. The following are equivalent:*

1. *$\mathcal{L}$ is eventually dendric;*
2. *there exists $N \geq 0$ such that, for all left special $w \in \mathcal{L}_{\geq N}$ there exists exactly one $a \in E^R_{\mathcal{L}}(w)$ such that wa is left special;*
3. *there exists $N \geq 0$ such that, for all right special $w \in \mathcal{L}_{\geq N}$ there exists exactly one $a \in E^L_{\mathcal{L}}(w)$ such that aw is right special.*

Note that, Damron and Fickensher [DF22] call a word w *regular* if there exists exactly one $a \in E^R_{\mathcal{L}}(w)$ such that wa is left special and exactly one $b \in E^L_{\mathcal{L}}(w)$ such that bw is right special.

We will in fact re-obtain the equivalences of Proposition 2.21 as a consequence of other results of this section. To do so, we introduce an intermediary notion called right ordinary (resp., left ordinary) by analogy with the notion of ordinary words.

Definition 2.22. Let $\mathcal{L}$ be a language. A word $w \in \mathcal{L}$ is *right ordinary* if there exists $a \in E^R_{\mathcal{L}}(w)$ such that $E^L_{\mathcal{L}}(wa) = E^L_{\mathcal{L}}(w)$ and for all $b \in E^R_{\mathcal{L}}(w) \setminus \{a\}$, wb is not left special. Similarly, w is *left ordinary* if there exists $a \in E^L_{\mathcal{L}}(w)$ such that $E^R_{\mathcal{L}}(aw) = E^R_{\mathcal{L}}(w)$ and for all $b \in E^L_{\mathcal{L}}(w) \setminus \{a\}$, bw is not right special.

In some sense, we cannot distinguish a right ordinary (resp., left ordinary) word from an ordinary word if we only look at the degrees of the right (resp., left) vertices of the extension graph. More precisely, we have the following observation. We can easily check that a word is both right and left ordinary if and only if it is in fact ordinary.

Remark 2.23. A word is both left and right ordinary if and only if it is in fact ordinary. In particular, if a word is not bispecial, then it is trivially ordinary (resp., left ordinary; resp., right ordinary).

We also have the following direct link with dendricity.

Lemma 2.24. *Let $\mathcal{L}$ be a language and $w \in \mathcal{L}$. If w is right ordinary (resp., left ordinary), then it is dendric.*

Proof. Assume that w is right ordinary and let $a \in E^R_{\mathcal{L}}(w)$ be such that $E^L_{\mathcal{L}}(wa) = E^L_{\mathcal{L}}(w)$ and for all $b \in E^R_{\mathcal{L}}(w) \setminus \{a\}$, wb is not left special. Since $E^L_{\mathcal{L}}(wa) = E^L_{\mathcal{L}}(w)$, all the left vertices of $\mathcal{E}_{\mathcal{L}}(w)$ are connected but as $\mathcal{E}_{\mathcal{L}}(w)$ has no isolated vertices, this implies that $\mathcal{E}_{\mathcal{L}}(w)$ is connected. On the other hand, as wb is not left special for all $b \in E^R_{\mathcal{L}}(w) \setminus \{a\}$, the only possible right vertex of degree at least two is a. Since $\mathcal{E}_{\mathcal{L}}(w)$ is bipartite, this implies that it is acyclic. Therefore, w is dendric. The proof when w is left ordinary is symmetric. □

As explained earlier, we will look at "eventual" properties, formally defined as follows by analogy with eventual dendricity and eventual neutrality.

Definition 2.25. Let $\mathcal{L}$ be a language and P be a property defined on words inside a language. We say that $\mathcal{L}$ is *eventually* P if there exists $N \geq 0$ such that, for all $w \in \mathcal{L}_{\geq N}$, w satisfies the property P. The smallest such N is then called the *threshold*.

We now have a first set of equivalences [Ghe23]. The idea behind these equivalences is that, if a word is not strong, then there are only two possible cases: either it has a right extension with the same set of left extensions (and it is the only right extensions giving a left special word), or it has several right extensions giving left special words but each of them have strictly fewer left extensions than the original word. This second case can then only happen a finite number of time.

Proposition 2.26. *Let $\mathcal{L}$ be a language. The following are equivalent:*

1. *$\mathcal{L}$ is eventually right ordinary with threshold N;*
2. *$\mathcal{L}$ is eventually dendric with threshold M;*
3. *$\mathcal{L}$ is eventually acyclic with threshold L;*
4. *$\mathcal{L}$ is eventually neutral with threshold L';*
5. *$\mathcal{L}$ is eventually weak or neutral with threshold K.*

Moreover, $K \leq L \leq M \leq N$ and $K \leq L' \leq M \leq N$.

Proof. The implication $1 \implies 2$ follows from Lemma 2.24. By Proposition 2.10, we have the implications $2 \implies 3 \implies 5$, and $2 \implies 4 \implies 5$. Moreover, we also deduce the inequalities for the thresholds. Therefore, it only remains to prove that, if $\mathcal{L}$ is eventually weak or neutral, then it is eventually right ordinary.

Assume that $\mathcal{L}$ is eventually weak or neutral of threshold K but not eventually right ordinary. Thus, there exist infinitely many weak or neutral words which are not right ordinary. Let $W \subseteq \mathcal{L}$ denote the set of these words and let $u \in \mathcal{L}_{\geq K}$ be a prefix of an infinite number of elements in W. Assume also that $\#E^L_{\mathcal{L}}(u)$ is minimal among such words.

Using the pigeonhole principle, there exists a right extension a of u such that ua is a prefix of an infinite number of elements in W. By hypothesis on u, $\#E^L_{\mathcal{L}}(ua) \geq \#E^L_{\mathcal{L}}(u)$ but as $E^L_{\mathcal{L}}(ua) \subseteq E^L_{\mathcal{L}}(u)$, we must have the equality. This implies that every left vertex of $\mathcal{E}_{\mathcal{L}}(u)$ is connected to the right vertex a, and thus u is connected. Since $|u| \geq K$, u is also weak or neutral. By Proposition 2.10, it is dendric. In particular, a is the unique right extension of u such that ua is left special otherwise we have a cycle in $\mathcal{E}_{\mathcal{L}}(u)$. This shows that u is right ordinary, and in particular, $u \notin W$.

As the elements of W are not right ordinary, they are bispecial by Remark 2.23, thus left special. This shows that no element of W can begin with ub, $b \neq a$. We deduce that

$$W \cap u\mathcal{A}^* = W \cap ua\mathcal{A}^*$$

where $\mathcal{A}$ is the alphabet. By iterating the reasoning, we can find for each $n \geq 1$ a word $v^{(n)}$ of length n such that

$$W \cap u\mathcal{A}^* = W \cap uv^{(n)}\mathcal{A}^*.$$

In particular, this implies that $W \cap u\mathcal{A}^*$ does not contain any word of length $|u| + n - 1$. As it is true for all $n \geq 1$, this would mean that $W \cap u\mathcal{A}^*$ is empty, and contradict the definition of u. $\square$

We can easily replace right ordinary by left ordinary in the previous result. We then obtain the following corollary.

Corollary 2.27. *Let $\mathcal{L}$ be a language. The following are equivalent:*

1. *$\mathcal{L}$ is eventually dendric;*
2. *$\mathcal{L}$ is eventually left ordinary;*
3. *$\mathcal{L}$ is eventually ordinary.*

From Proposition 2.26, we can also deduce other equivalences with eventual properties that do not have a specific name. In particular, we re-obtain the equivalences given by Dolce and Perrin (Proposition 2.21).

Proposition 2.28. *Let $\mathcal{L}$ be a language. The following are equivalent:*

1. *$\mathcal{L}$ is eventually dendric;*
2. *there exists N such that, for all $w \in \mathcal{L}_{\geq N}$, there exists at most one $a \in E^R_{\mathcal{L}}(w)$ such that wa is left special;*
3. *there exists N such that, for all left special $w \in \mathcal{L}_{\geq N}$ there exists exactly one $a \in E^R_{\mathcal{L}}(w)$ such that wa is left special;*
4. *there exists N such that, for all left special $w \in \mathcal{L}_{\geq N}$, if there exists $a \in E^R_{\mathcal{L}}(w)$ such that $E^L_{\mathcal{L}}(wa) = E^L_{\mathcal{L}}(w)$, then for all $b \in E^R_{\mathcal{L}}(w)$, wb is not left special.*

Similarly, the following are also equivalent:

1. *$\mathcal{L}$ is eventually dendric;*
2. *there exists N such that, for all $w \in \mathcal{L}_{\geq N}$, there exists at most one $a \in E^L_{\mathcal{L}}(w)$ such that aw is right special;*
3. *there exists N such that, for all right special $w \in \mathcal{L}_{\geq N}$ there exists exactly one $a \in E^L_{\mathcal{L}}(w)$ such that aw is right special;*
4. *there exists N such that, for all right special $w \in \mathcal{L}_{\geq N}$, if there exists $a \in E^L_{\mathcal{L}}(w)$ such that $E^R_{\mathcal{L}}(aw) = E^R_{\mathcal{L}}(w)$, then for all $b \in E^L_{\mathcal{L}}(w)$, bw is not right special.*

Proof. We prove the first set of equivalences as the proof for the other one is symmetric. The second and third assertions imply that $\mathcal{L}$ is eventual acyclic and are implied by the fact that $\mathcal{L}$ is eventually right ordinary. Therefore they are equivalent to eventual dendricity by Proposition 2.26. For the last assertion, it is clearly true for an eventually right ordinary language. The converse follows from a careful analysis of the proof of Proposition 2.26. Indeed, we only use the hypothesis that u is weak or neutral to show that, since $a \in E^R_{\mathcal{L}}(u)$ is such that $E^L_{\mathcal{L}}(ua) = E^L_{\mathcal{L}}(u)$, then for all $b \in E^R_{\mathcal{L}}(u)$, ub is not left special. Therefore, we can replace the weak or neutral condition by the property of the last assertion. □

Warning. We recommend that the reader keeps the equivalences presented in this section, especially the ones in Proposition 2.26, in mind as, in the rest of this work, we will indifferently refer to any of the families, depending on the property which seems the most relevant in the proof or result.

Observe that the families of Proposition 2.26 have a linear factor complexity by Corollary 1.44 on the factor complexity of eventually neutral languages. This implies in particular that they do not also coincide with the family of eventually connected languages, or of eventually strong or neutral languages, as shown in the following example.

Example 2.29. Let $\mathcal{A}$ be an alphabet of size $k \geq 2$ and let $\mathcal{L} = \mathcal{A}^*$. The factor complexity of $\mathcal{L}$ is given by $p_{\mathcal{L}}(n) = k^n$ therefore $\mathcal{L}$ is not eventually dendric. On the other hand, for all $w \in \mathcal{A}^*$, its bi-extensions are given by $\mathcal{A} \times \mathcal{A}$. In particular, w is connected, and strong or neutral. In fact, w is always strong.

Under additional restriction on the factor complexity of the languages however, we recover an equivalence. This is the object of the following result. Recall that the notation $f \in O(g)$ means that there exist C and N such that $f(n) \leq Cg(n)$ for all $n \geq N$.

Proposition 2.30. *Let $\mathcal{L}$ be a language. The following are equivalent:*

1. *$\mathcal{L}$ is eventually neutral;*
2. *there exists $N \geq 0$ such that, for all $w \in \mathcal{L}_{\geq N}$, there exists at least one $a \in E^R_{\mathcal{L}}(w)$ such that $E^L_{\mathcal{L}}(wa) = E^L_{\mathcal{L}}(w)$, and $p_{\mathcal{L}}(n) \in O(n)$;*
3. *$\mathcal{L}$ is eventually connected and $p_{\mathcal{L}}(n) \in O(n)$;*
4. *$\mathcal{L}$ is eventually strong or neutral and $p_{\mathcal{L}}(n) \in O(n)$.*

Proof. As eventually neutral is equivalent to eventually right ordinary by Proposition 2.26, we have 1 $\implies$ 2 by Corollary 1.44 on the factor complexity. We trivially have 2 $\implies$ 3. Moreover, using Proposition 2.10, we know that 3 $\implies$ 4. It remains to show that, if $\mathcal{L}$ is eventually strong or neutral and has a linear complexity, then it is eventually neutral.

Assume that $\mathcal{L}$ is eventually strong or neutral of threshold N. Therefore, using Corollary 1.43, the sequence $s_{\mathcal{L}}(n)$ is non-decreasing starting at index N, and for all $n \geq N$, $s_{\mathcal{L}}(n) = s_{\mathcal{L}}(n+1)$ if and only if all length-n words are neutral. On the other hand, as $p_{\mathcal{L}}$ is at most linear, $s_{\mathcal{L}}$ is bounded by a result of Cassaigne [Cas96]. Therefore, $s_{\mathcal{L}}$ is eventually constant. This shows that $\mathcal{L}$ is eventually neutral. □

We end this section with one last equivalence.

Proposition 2.31. *A language $\mathcal{L}$ is eventually right ordinary if and only if there exists $N \geq 0$ and right infinite words $u^{(1)}, \dots, u^{(k)}$ such that their length-N prefixes are distinct and, for all $n \geq N$, their length-n prefixes are exactly the length-n left special words of $\mathcal{L}$.*

Similarly, a language $\mathcal{L}$ is eventually left ordinary if and only if there exists $N \geq 0$ and left infinite words $u^{(1)}, \dots, u^{(k)}$ such that their length-N suffixes are distinct and, for all $n \geq N$, their length-n suffixes are exactly the length-n right special words of $\mathcal{L}$.

Proof. We prove the first claim as the other one is symmetric. If $\mathcal{L}$ is eventually right ordinary of threshold N, then every left special word of length n is prefix of exactly one length-$(n+1)$ left special word for all $n \geq N$. In other words, the left special words of length at least N form k families according to their length-N prefix, each family containing exactly one word of each length $n \geq N$. We can then take the words $u^{(1)}, \dots, u^{(k)}$ as limits for each of these families.

For the converse, observe that this condition on the right infinite words implies that every left special word w of length $n \geq N$ has only one right extension a such that wa is left special. By Proposition 2.28, this implies that $\mathcal{L}$ is eventually right ordinary. □

2.4 Dendricity and some other graphs of extensions

In this section, we introduce a new family of graphs defined using extensions of words in the language and that we first studied with J. Leroy in [GL22].

By definition, these graphs are closely related to the usual extension graphs. We explore this link in Subsection 2.4.1 and show that the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ can be used to obtain an alternative definition of dendric languages. Subsection 2.4.2 is more focused on graph theory results to study the shape of these graphs. Finally, we look at the stabilization of the shapes of the graphs when n tends to infinity and show the link with eventually dendric languages in Subsection 2.4.3. Most of the results presented in this section come from [GL22].

Let us first define these graphs. Contrary to the classical extension graphs, we look at the left and right extensions separately. The idea is that, instead of considering extensions for just one word, we look at all the words of a given length.

Definition 2.32. Let $\mathcal{L}$ be a language over $\mathcal{A}$ and $n \geq 0$. The graph $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) is the multi-graph with labeled edges such that

- its vertices are the elements of $\mathcal{A}$,
- for any $w \in \mathcal{L}_n$ and any distinct $a, b \in E_{\mathcal{L}}^L(w)$ (resp., $a, b \in E_{\mathcal{L}}^R(w)$) there is an (undirected) edge labeled by w between the vertices a and b.

Observe that the edges of $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) are only labeled by left (resp., right) special words of length n. Moreover, for each special word w, the edges that it labels form a complete subgraph, or *clique*, whose vertices are the elements of $E_{\mathcal{L}}^L(w)$ (resp., $E_{\mathcal{L}}^R(w)$). The graph $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) can therefore also be seen as a union of cliques, potentially with some isolated vertices. This point of view will be central in Subsection 2.4.2.

We can also notice that the graphs $G_0^L(\mathcal{L})$ and $G_0^R(\mathcal{L})$ are always complete graphs with edges labeled by ε. Therefore, they only depend on the alphabet and not on the language $\mathcal{L}$.

We describe below the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ for two famous languages already mentioned in this work: the Tribonacci language and the Chacon language.

Example 2.33. In the Tribonacci language $\mathcal{L}$ (Example 1.25), there is exactly one left (resp., right) special factor of each length and it can be extended by all of the letters. Therefore, the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ are complete graphs and they only differ by the label of their edges. This observation is true for any Arnoux-Rauzy language. The first few graphs for the Tribonacci language are represented in Figure 2.3.

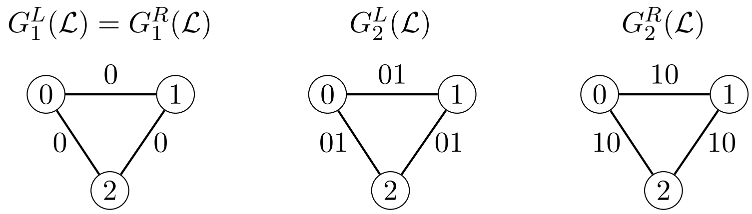


Figure 2.3: Graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$, $n \in \{1, 2\}$, for the Tribonacci language $\mathcal{L}$.

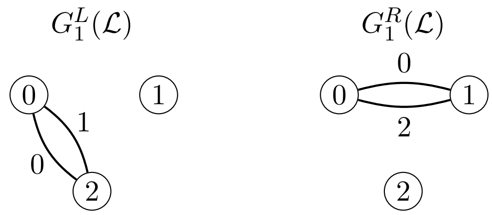


Figure 2.4: Graphs $G_1^L(\mathcal{L})$ and $G_1^R(\mathcal{L})$ for the Chacon language $\mathcal{L}$.

Example 2.34. Using Example 1.13, the graphs $G_1^L(\mathcal{L})$ and $G_1^R(\mathcal{L})$ for the Chacon language $\mathcal{L}$ are represented in Figure 2.4.

In particular, we observe that any left special word of length at least 1 will have 0 and 2 as its only left extensions. Moreover, as the Chacon language has complexity $2n+1$, there must be exactly two left special factors of each length $n \geq 1$. This shows that, for all $n \geq 1$, the graph $G_n^L(\mathcal{L})$ has exactly two edges (with different labels), and they are between 0 and 2. We can similarly show that $G_n^R(\mathcal{L})$, $n \geq 1$, has two edges between 0 and 1.

2.4.1 A characterization of dendric languages

Using Examples 2.2 and 2.34 with the Chacon language, we can see that paths in $\mathcal{E}_{\mathcal{L}}(\varepsilon)$ are translated into edges in $G_1^L(\mathcal{L})$, as represented in Figure 2.5.

This link is more general, as stated in the following lemma. Recall that, in a bipartite graph, a^L denotes the left vertex labeled a and a^R the right vertex labeled a.

Lemma 2.35. *Let $\mathcal{L}$ be a language and $w \in \mathcal{L}_n$. The graph $\mathcal{E}_{\mathcal{L}}(w)$ contains*

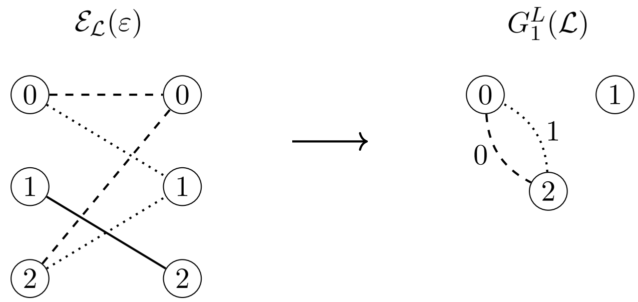


Figure 2.5: Example of a correspondence between the paths in $\mathcal{E}_{\mathcal{L}}(\varepsilon)$ and in $G^L_{\mathcal{L}}(1)$.

the path $(a_1^L, b_1^R, a_2^L, \ldots, b_k^R, a_{k+1}^L)$ *if and only if* $G^L_{n+1}(\mathcal{L})$ *contains the path*

$$a_1 \xrightarrow{wb_1} a_2 \cdots \xrightarrow{wb_k} a_{k+1}.$$

Symmetrically, the graph $\mathcal{E}_{\mathcal{L}}(w)$ *contains the path* $(a_1^R, b_1^L, a_2^R, \ldots, b_k^L, a_{k+1}^R)$ *if and only if* $G^R_{n+1}(\mathcal{L})$ *contains the path*

$$a_1 \xrightarrow{b_1w} a_2 \cdots \xrightarrow{b_kw} a_{k+1}.$$

Proof. Let us prove the result for $k = 1$, the general case follows by simple induction. We have the path (a_1^L, b_1^R, a_2^L) in $\mathcal{E}_{\mathcal{L}}(w)$ if and only if a_1 and a_2 are two left extensions of wb_1, which is exactly the definition of having an edge labeled by wb_1 between a_1 and a_2 in $G^L_{n+1}(\mathcal{L})$. We similarly show that we have the path (a_1^R, b_1^L, a_2^R) in $\mathcal{E}_{\mathcal{L}}(w)$ if and only if there is an edge labeled by b_1w between a_1 and a_2 in $G^R_{n+1}(\mathcal{L})$. □

Therefore, we have a link between the paths in the extension graphs of the length-n words and the paths in $G^L_{n+1}(\mathcal{L})$ and $G^R_{n+1}(\mathcal{L})$. It is then natural to wonder if there is also a link between the properties defined using paths, and in particular acyclicity and connectedness.

However, we right away notice that, while all the extension graphs are acyclic in the Tribonacci language, the graphs $G^L_n(\mathcal{L})$ (and $G^R_n(\mathcal{L})$) contain a cycle by Example 2.33. We need to look at a weaker form of acyclicity. As we are using multi-graphs with labeled edges and no loops, a *simple path* is a non-empty path that does not go twice through the same vertex, except potentially for its beginning and its end that can coincide, and that does not use the same (labeled) edge consecutively. A *simple cycle* is a closed simple path. If we know that a cycle uses at least two different edges, then only the condition on the vertices is needed for it to be simple.

Definition 2.36. A multi-graph with labeled edges G is *acyclic for the labeling* if any simple cycle in G only uses edges with the same label.

Example 2.37. The graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ for the Tribonacci language (Figure 2.3) are acyclic for the labeling while the corresponding graphs for the Chacon language (Figure 2.4) are not (except for $n = 0$).

Proposition 2.38. *Let $\mathcal{L}$ be a language and $N \geq 0$. The following properties are equivalent.*

1. *Every word $w \in \mathcal{L}_{<N}$ is acyclic.*
2. *The graph $G_n^L(\mathcal{L})$ is acyclic for the labeling for all $n \leq N$.*
3. *The graph $G_n^R(\mathcal{L})$ is acyclic for the labeling for all $n \leq N$.*

Proof. Let us show the equivalence between the acyclicity of the words and the acyclicity of the graphs $G_n^L(\mathcal{L})$. The proof for the graphs $G_n^R(\mathcal{L})$ is similar. We proceed by contraposition and show that there exists $w \in \mathcal{L}_{<N}$ for which $\mathcal{E}_{\mathcal{L}}(w)$ is not acyclic if and only if there exists $n \leq N$ such that the graph $G_n^L(\mathcal{L})$ contains a simple cycle with different labels.

Assume that $\mathcal{E}_{\mathcal{L}}(w)$ is not acyclic for some $w \in \mathcal{L}_n$, $n < N$. Therefore, it contains a simple cycle going through (at least) two different right vertices. By Lemma 2.35, $G_{n+1}^L(\mathcal{L})$ contains a simple cycle whose edges do not all have the same label.

For the converse, let $n \leq N$ be such that $G_n^L(\mathcal{L})$ contains a simple cycle with at least two distinct labels. Let us denote this cycle

$$a_1 \xrightarrow{u^{(1)}} a_2 \cdots \xrightarrow{u^{(k)}} a_1.$$

In particular, $n \geq 1$. Let w be the longest common prefix to all of the $u^{(i)}$, $i \leq k$, and let b_i be the letter such that wb_i is a prefix of $u^{(i)}$, $i \leq k$. By definition of $G_n^L(\mathcal{L})$ and $G_{|w|+1}^L(\mathcal{L})$, we also have the cycle

$$a_1 \xrightarrow{wb_1} a_2 \cdots \xrightarrow{wb_k} a_1.$$

in $G_{|w|+1}^L(\mathcal{L})$. It is a simple cycle as the a_i's are distinct and, by definition of w, the cycle uses at least two different edges. We can in fact assume that no two consecutive edge labels (including wb_k and wb_1) are equal. Indeed, if v labels an edge between a and b, and between b and c in $G^L_{|w|+1}(\mathcal{L})$, then it labels an edge between a and c by definition of $G^L_{|w|+1}(\mathcal{L})$. Making that replacement in the cycle does not impact the fact that it is a simple cycle with at least two distinct labels.

By Lemma 2.35, the graph $\mathcal{E}_{\mathcal{L}}(w)$ contains the cycle $(a_1^L, b_1^R, \dots, b_k^R, a_1^L)$ which is non-trivial as the a_i's are distinct and consecutive b_i's are also distinct. □

We obtain the following direct corollary.

Corollary 2.39. *Let $\mathcal{L}$ be a language. The following are equivalent.*

1. *The language $\mathcal{L}$ is acyclic.*
2. *The graph $G^L_n(\mathcal{L})$ is acyclic for the labeling for all $n \geq 0$.*
3. *The graph $G^R_n(\mathcal{L})$ is acyclic for the labeling for all $n \geq 0$.*

Proposition 2.38 is false if we only look at the properties locally and not for all $n \leq N$. Even though we can see in the proof that, if $G^L_N(\mathcal{L})$ is acyclic, then the elements of $\mathcal{L}_{N-1}$ are acyclic, the converse if false, as shown in the following example.

Example 2.40. In the Chacon language (Example 1.13), we can see that the words of length 1 and 2 are acyclic but, by Example 2.34, the graphs $G^L_2(\mathcal{L})$, $G^L_3(\mathcal{L})$, $G^R_2(\mathcal{L})$ and $G^R_3(\mathcal{L})$ are not acyclic for the labeling. As ε is not acyclic, this does not contradict Proposition 2.38 however.

We now turn to the connectedness properties. Unfortunately, we do not have a result as in Proposition 2.38 without additional restrictions. This can be seen in the following example.

Example 2.41. Let $\mathcal{L}$ be a language such that $\mathcal{L}_3 = \{001, 010, 011, 100, 110\}$, then the graph $\mathcal{E}_{\mathcal{L}}(0)$ is not connected. However, both $G^L_1(\mathcal{L})$ and $G^L_2(\mathcal{L})$ are (and so are $G^R_1(\mathcal{L})$ and $G^R_2(\mathcal{L})$). This situation is represented in Figure 2.6.

However, we see in this example that we had two paths in $G^L_1(\mathcal{L})$ and only one in $G^L_2(\mathcal{L})$. This loss is in fact caused by the disconnection in $\mathcal{E}_{\mathcal{L}}(0)$. More generally, if we want a link between disconnected words and a disconnection in $G^L_n(\mathcal{L})$, we need to ensure that there are no double paths in $G^L_{n-1}(\mathcal{L})$. In other words, we need acyclicity to have an equivalence between connectedness of the words and of the graphs $G^L_n(\mathcal{L})$.

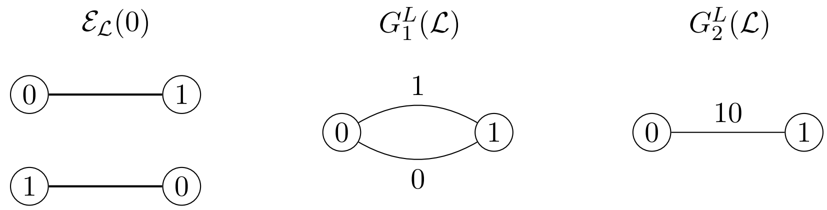


Figure 2.6: Example showing that a disconnected word does not always imply a disconnection in the graph $G_n^L(\mathcal{L})$.

Proposition 2.42. *Let $\mathcal{L}$ be a language and $N \geq 0$. If the words $w \in \mathcal{L}_{<N}$ are acyclic, then the following properties are equivalent.*

1. *Every word $w \in \mathcal{L}_{<N}$ is connected.*
2. *The graph $G_n^L(\mathcal{L})$ is connected for all $n \leq N$.*
3. *The graph $G_n^R(\mathcal{L})$ is connected for all $n \leq N$.*
4. *The graph $G_N^L(\mathcal{L})$ is connected.*
5. *The graph $G_N^R(\mathcal{L})$ is connected.*

We first show the following stronger result which will be key in Section 4.4.

Proposition 2.43. *Let $\mathcal{L}$ be a language over $\mathcal{A}$ and $N \geq 0$. If the words $w \in \mathcal{L}_{<N}$ are acyclic, then the following properties are equivalent for all $C \subseteq \mathcal{A}$.*

1. *For all $w \in \mathcal{L}_{<N}$ and all $a, b \in E_{\mathcal{L}}^L(w) \cap C$, the vertices a^L and b^L are connected in $\mathcal{E}_{\mathcal{L}}(w)$ by a path avoiding vertices c^L, $c \notin C$.*
2. *The subgraph of $G_n^L(\mathcal{L})$ generated by the vertices in C is connected for all $n \leq N$.*
3. *The subgraph of $G_N^L(\mathcal{L})$ generated by the vertices in C is connected.*

Similarly, the following are equivalent.

1. *For all $w \in \mathcal{L}_{<N}$ and all $a, b \in E_{\mathcal{L}}^R(w) \cap C$, the vertices a^R and b^R are connected in $\mathcal{E}_{\mathcal{L}}(w)$ by a path avoiding vertices c^R, $c \notin C$.*

2. *The subgraph of* $G_n^R(\mathcal{L})$ *generated by the vertices in* C *is connected for all* $n \leq N$*.*

3. *The subgraph of* $G_N^R(\mathcal{L})$ *generated by the vertices in* C *is connected.*

Proof. We only show the first set of equivalences, the other one being symmetric. Let H_n denote the subgraph of $G_n^L(\mathcal{L})$ generated by the vertices in C.

Assume that the first property is satisfied and let us prove that H_n is connected by induction on $n \leq N$. For $n = 0$, as the graph $G_0^L(\mathcal{L})$ is a complete graph, any of its subgraph, and H_0 in particular, is connected. Assume now that H_n is connected for $n < N$ and let us prove that H_{n+1} is also connected. It suffices to show that any two vertices $a, b \in C$ that were connected by an edge in H_n are connected by a path in H_{n+1}. Let $a, b \in C$ be two such vertices and w be the label of an edge between them. As a and b are left extensions of w and $w \in \mathcal{L}_{<N}$, a^L and b^L are connected in $\mathcal{E}_{\mathcal{L}}(w)$ by a path avoiding vertices c^L, $c \notin C$. This path in $\mathcal{E}_{\mathcal{L}}(w)$ corresponds to a path between a and b in H_{n+1} by Lemma 2.35. Observe that the acyclic hypothesis is not required for this implication.

Let us now prove the converse, i.e., if H_n is connected for all $n \leq N$, then the first claim is satisfied. We proceed by contradiction, so assume that, for $w \in \mathcal{L}_n$, $n < N$, there exist two letters $a, b \in E_{\mathcal{L}}^L(w) \cap C$ such that the vertices a^L and b^L are not connected in $\mathcal{E}_{\mathcal{L}}(w)$ by a path avoiding vertices c^L, $c \notin C$. By definition, there is an edge labeled by w between a and b in $G_n^L(\mathcal{L})$. As this graph is acyclic for the labeling by Proposition 2.38, any simple path between a and b in $G_n^L(\mathcal{L})$ uses exclusively edges labeled by w. By hypothesis, a and b are connected by a path P in H_{n+1}. This induces a path in H_n (so in G_n^L) by taking the length-n prefixes of the edges' labels. This implies that the path P only uses edges labeled by words in $w\mathcal{A}$. By Lemma 2.35, P then induces a path in $\mathcal{E}_{\mathcal{L}}(w)$ between a^L and b^L avoiding vertices c^L, $c \notin C$. This is a contradiction and ends the equivalence between the first two claims.

We now show that it is sufficient to look at H_N. Indeed, for all $n \leq N$, a path in $G_N^L(\mathcal{L})$ induces a path in $G_n^L(\mathcal{L})$ by taking the prefixes of the edges' labels. Therefore, if H_N is connected, then so is H_n for all $n \leq N$. □

Proof of Proposition 2.42. Recall that extension graphs do not have any isolated vertices. Therefore, they are connected if and only if all of their left vertices are connected, if and only if all of their right vertices are connected. The conclusion then follows from Proposition 2.43 for $C = \mathcal{A}$. □

Observe that the acyclicity is only needed to show that if the graphs $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) are connected, then the words are connected. The other implication is always true. We can however show that looking at properties locally is not sufficient, just as in Example 2.40.

Similarly to what we did for acyclicity, we can deduce the following corollary which gives an alternative definition of dendric languages.

Corollary 2.44. *Let $\mathcal{L}$ be a language. The following are equivalent.*

1. *The language $\mathcal{L}$ is dendric.*
2. *The graph $G_n^L(\mathcal{L})$ is acyclic for the labeling and connected for all $n \geq 0$.*
3. *The graph $G_n^R(\mathcal{L})$ is acyclic for the labeling and connected for all $n \geq 0$.*

2.4.2 Colors and cliques

We now take a slight step away from dendricity to study properties of the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ to get a better idea of what they can look like. In the rest of this work, we will be interested in whether two edges of $G_n^L(\mathcal{L})$ have the same label or not more than in the actual label of the edges. In other words, we will consider graphs with *colored* edges instead of graphs with *labeled* edges. Formally, we have the following definition.

Definition 2.45. Two multi-graphs G and G' with edges labeled by elements of C and C' respectively are *equivalent* if they both have the same set of vertices and if there exists a bijection $\varphi\colon C \to C'$ such that there are k edges labeled by $c \in C$ between a and b in G if and only if there are k edges labeled by $\varphi(c)$ between a and b in G'. A *(multi-)graph with colored edges* is an equivalence class for this relation.

We will however right-away drop these considerations of equivalence classes and identify the labeled graphs with their colored counterpart. In particular, we write $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ to denote both the labeled and the colored graphs, depending on the properties needed locally.

Example 2.46. Using Example 2.33, we see that, for the Tribonacci language $\mathcal{L}$, we have $G_n^L(\mathcal{L}) = G_m^R(\mathcal{L})$ for all $m, n \geq 0$, and by Example 2.34, $G_n^L(\mathcal{L}') = G_m^L(\mathcal{L}')$, $G_n^R(\mathcal{L}') = G_m^R(\mathcal{L}')$ for all $m, n \geq 1$ for the Chacon language $\mathcal{L}'$. This is represented in Figure 2.7.

The notion of acyclic for the labeling can easily be translated into acyclic for the coloring as follows.

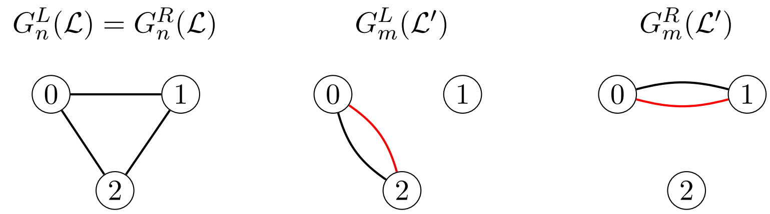


Figure 2.7: Graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$, $n \geq 0$, for the Tribonacci language $\mathcal{L}$ (on the left) and graphs $G_m^L(\mathcal{L}')$ and $G_m^R(\mathcal{L}')$, $m \geq 1$, for the Chacon language $\mathcal{L}'$ (on the right).

Definition 2.47. A multi-graph with colored edges G is *acyclic for the coloring* if any simple cycle in G only uses edges with the same color.

It is clear that a labeled graph is acyclic for the labeling if and only if its colored counterpart is acyclic for the coloring.

As explained after Definition 2.32, the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ can be seen as unions of cliques, each with a different label (or color). Such a graph will be called a multi-clique.

Definition 2.48. A multi-graph G with colored edges and with set of vertices V is a *multi-clique* if there exist subsets $C_1, \ldots, C_k$ of V such that the set of edges of G is the union, for $i \leq k$, of the sets of edges $(C_i \times C_i) \setminus \mathrm{diag}(C_i)$ with the color c_i, where $c_1, \ldots, c_k$ are distinct colors. The multi-clique G is then denoted $G_V(\{C_1, \ldots, C_k\})$[1]. When the context is clear or if we do not worry about isolated vertices, we omit the subscript V.

Example 2.49. If $V = \{0, 1, 2, 3\}$, $C_1 = \{0, 1\}$, $C_2 = \{1, 2, 3\}$, $C_3 = \{0, 3\}$, the multi-clique $G_V(\{C_1, C_2, C_3\})$ is represented in Figure 2.8. It is not acyclic for the coloring because of the cycle $(0, 1, 3, 0)$.

Remark 2.50. Given a language $\mathcal{L}$ over $\mathcal{A}$, we have the following alternative definition for the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$:

$$G_n^L(\mathcal{L}) = G_{\mathcal{A}}(\{E_{\mathcal{L}}^L(w) : w \in \mathcal{L}_n\})$$

and

$$G_n^R(\mathcal{L}) = G_{\mathcal{A}}(\{E_{\mathcal{L}}^R(w) : w \in \mathcal{L}_n\}).$$

[1]The notation $\{C_1, \ldots, C_k\}$ should be understood here as a multi-set, i.e., if $C_i = C_j$ with $i \neq j$ and $\#C_j \geq 2$, then $G_V(\{C_1, \ldots, C_k\}) \neq G_V(\{C_1, \ldots, C_k\} \setminus \{C_j\})$.

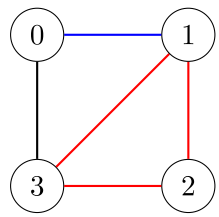


Figure 2.8: Multi-clique $G(\{\{0,1\},\{1,2,3\},\{0,3\}\})$.

Observe that if $\#C_i \leq 1$, then $G(\{C_1,\dots,C_k\}) = G(\{C_1,\dots,C_k\} \setminus \{C_i\})$ as C_i does not generate any edge. In particular, this shows once again that it suffices to consider the left special words to build $G_n^L(\mathcal{L})$ and the right special words for $G_n^R(\mathcal{L})$.

Since this work focuses on dendric languages, we will mostly be interested in multi-cliques that are acyclic for the coloring and connected by Corollary 2.44. These multi-cliques can be obtained by an iterative process starting from the complete graph and successively splitting cliques. This is based on the following lemma.

Lemma 2.51. *Let $C_1,\dots,C_k \subseteq V$ and let D,E be such that*

$$D \cup E = C_1 \quad \text{and} \quad \#(D \cap E) = 1.$$

The multi-clique $G = G_V(\{C_1,\dots,C_k\})$ is acyclic for the coloring (resp., connected) if and only if the multi-clique $G' = G_V(\{D,E,C_2,\dots,C_k\})$ is.

Proof. Let $\{c\} = D \cap E$. We first show the equivalence for the acyclicity. Assume that G is acyclic for the coloring and let us prove that G' also is. Any simple cycle of G' corresponds to cycle of G thus only uses edges corresponding to one of the C_i, $i \leq k$, in G. The problem can therefore only arise if the path uses edges from C_1. However, c is the only vertex with both ingoing edges corresponding to D and ingoing edges corresponding to E. Since a simple path cannot go twice through the same vertex, this show that a simple cycle in G' cannot use both edges corresponding to D and edges corresponding to E. This ends the proof that G' is acyclic for the coloring.

For the converse, if a simple cycle of G is in G', then it clearly can only use edges of the same color. Thus, assume that we have a cycle of G that uses an edge from C_1 which disappears when splitting the clique. We can replace each such edge by a length-2 path (corresponding to C_1) going through c. We then obtain one or several simple cycles going through c and corresponding to cycles in G'. Moreover, each cycle uses at least one edge from D (resp., E). Therefore, the cycles only use edges from D and E in G'. This shows that this modified path is unicolor in G, and so is the original path.

We now turn to connectedness. As G contains the edges of G', if G' is connected then so is G. For the converse, each edge of G either corresponds to an edge of G' or can be replaced by a length-2 path going through c. Therefore we do not lose connectedness when splitting the clique C_1. □

Remark 2.52. A multi-clique $G(\{C_1, \dots, C_k\})$ is acyclic for the coloring and connected if and only if it can be obtained by applying a succession of $k-1$ splittings as in Lemma 2.51 to a complete unicolor graph. Indeed, by Lemma 2.51, any graph built this way is acyclic for the coloring and connected. Conversely, if $G(\{C_1, \dots, C_k\})$, $k \geq 2$, is connected, then there exist two cliques with a non empty intersection. Without loss of generality, assume that it is C_1 and C_2. If moreover the graph is acyclic for the coloring then their intersection contains exactly one element. By Lemma 2.51, $G(\{C_1 \cup C_2, C_3, \dots, C_k\})$ is acyclic for the coloring and connected. We can then proceed by induction to show that $G(\{C_1, \dots, C_k\})$ is obtained by applying $k-1$ splittings.

In fact, one can show that if the words of $\mathcal{L}_n$ are dendric, then we can go from $G^L_n(\mathcal{L})$ to $G^L_{n+1}(\mathcal{L})$ (resp., from $G^R_n(\mathcal{L})$ to $G^R_{n+1}(\mathcal{L})$) with a succession of splittings as in Lemma 2.51. In other words, the sequence of splittings used to obtain $G^L_N(\mathcal{L})$ and $G^R_N(\mathcal{L})$ can be governed by the extensions of the words of $\mathcal{L}_{<N}$.

Example 2.53. Let $\mathcal{L}'$ be an Arnoux-Rauzy language over the alphabet $\{0, 1, 2, 3\}$ and let σ be the morphism

$$\sigma\colon \begin{cases} 0 \mapsto 20 \\ 1 \mapsto 20221 \\ 2 \mapsto 202 \\ 3 \mapsto 20203 \end{cases}.$$

As we will show later, the image $\mathcal{L} := \sigma(\mathcal{L}')$ is dendric (see Example 4.80). The right special words of length 1 are 0 with extensions $\{2, 3\}$ and 2 with extensions $\{0, 1, 2\}$. Therefore, $G^R_1(\mathcal{L})$ is obtained from $G^R_0(\mathcal{L})$ by splitting $\{0, 1, 2, 3\}$ into $\{0, 1, 2\}$ and $\{2, 3\}$. Similarly, the right special words of length 2 are 02 with extensions $\{0, 2\}$, 20 with extensions $\{2, 3\}$ and 22 with

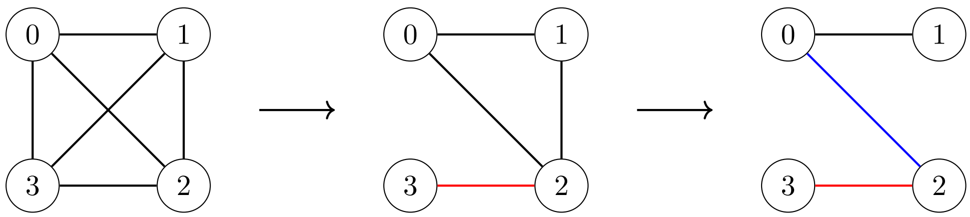


Figure 2.9: Splittings to go from the graph $G_0^R(\mathcal{L})$ (on the left) to the graph $G_2^R(\mathcal{L})$ (on the right) via the graph $G_1^R(\mathcal{L})$ (in the middle) for the language $\mathcal{L}$ of Example 2.53.

extensions $\{0,1\}$. Therefore, $G_2^R(\mathcal{L})$ is obtained from $G_1^R(\mathcal{L})$ by splitting $\{0,1,2\}$ into $\{0,1\}$ and $\{0,2\}$. The situation is represented in Figure 2.9.

We end this subsection with some considerations on how to color a graph. Observe that any multi-graph (with uncolored edges) G can be colored into a multi-clique. Indeed, we can color each edge with a different color. On the other hand, any multi-graph G can be colored into a graph which is acyclic for the coloring. It suffices to color all the edges with the same color. However, not every graph can be colored into a multi-clique which is acyclic for the coloring. This is the object of the following definition.

Definition 2.54. A multi-graph G is *acyclicly colorable* if it can be colored into an acyclic for the coloring multi-clique.

These graphs are also called *block graphs* [Har63]. By Corollary 2.44, if $\mathcal{L}$ is dendric then, for all $n \geq 0$, the uncolored version of $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) is acyclicly colorable. Moreover, the knowledge of this uncolored version is sufficient to recover $G_n^L(\mathcal{L})$ as stated in the following result.

Lemma 2.55. *If G is acyclicly colorable, the coloring giving a multi-clique which is acyclic for the coloring is unique.*

Proof. It suffices to observe that each color corresponds to a maximal clique in G. Indeed, each color of a multi-clique corresponds to a clique, and this clique is maximal otherwise we have a simple cycle using edges of different colors. $\square$

We now give an alternative method to check if a graph is acyclicly colorable.

Proposition 2.56. *A simple graph G is acyclicly colorable if and only if, for any distinct vertices a and b, if there is a simple cycle passing through a and b, then G contains the edge $\{a, b\}$.*

Proof. Assume that G can be colored into $G(\{C_1, \ldots, C_k\})$ which is acyclic for the coloring. If there is a simple cycle passing through a and b, then that cycle is unicolor and there exists i such that $a, b \in C_i$. This shows that G contains the edge $\{a, b\}$ since C_i is a clique.

Let us prove the existence of a coloring when, for any distinct vertices a and b, if there is a simple cycle passing through a and b, then G contains the edge $\{a, b\}$. If G is not connected, it is sufficient to prove the existence of a coloring for each of its connected components. We therefore assume that G is connected. As G can be written as a union of (uncolored) cliques, we proceed by induction on the number of maximal cliques.

If G itself is a clique then we can simply color all the edges with the same color. If G contains at least two maximal cliques, then there exist two maximal cliques C_1 and C_2 with a non-empty intersection since G is connected. Moreover, their intersection contains exactly one vertex. Indeed, assume by contradiction that it contains two different vertices c and d. As C_1 is a maximal clique, there exists $a \in C_1 \setminus C_2$. For all $b \in C_2 \setminus \{c, d\}$, we have the simple cycle (a, c, b, d, a) passing through a and b. By hypothesis on G, the pair $\{a, b\}$ is an edge of G. This shows that a has an edge to each vertex in C_2 and contradicts the maximality of C_2.

Let us consider the graph G' which is a copy of G where we added all the edges between vertices in C_1 and C_2. In other words, we merge C_1 and C_2 to form a bigger maximal clique $C_1 \cup C_2$. The graph G' has strictly fewer maximal cliques than G therefore, by induction hypothesis it is acyclicly colorable. Moreover, by Lemma 2.55, the edges of $C_1 \cup C_2$ correspond to one color, i.e., G' can be colored into $G(\{C_1 \cup C_2, C_3, \ldots, C_k\})$ which is acyclic for the coloring, for some $C_3, \ldots, C_k$. By Lemma 2.51, G can be colored into $G(\{C_1, C_2, \ldots, C_k\})$ which is acyclic for the coloring. $\square$

2.4.3 Stabilization and eventually dendric languages

By definition, if there is no edge between a and b in $G_N^L(\mathcal{L})$, then there won't be any edge between these two vertices in $G_n^L(\mathcal{L})$ for all $n \geq N$. This seems to suggest that the graphs $G_n^L(\mathcal{L})$ converge when n tends to infinity. However, as they are multi-graphs, we could have a growing number of edges between two given vertices. This behavior does not appear when $\mathcal{L}$ is eventually dendric.

Proposition 2.57. *Let $\mathcal{L}$ be an eventually right ordinary (resp., left ordinary) language of threshold N. For all $n \geq N$,*

$$G_n^L(\mathcal{L}) = G_N^L(\mathcal{L}) \quad \textit{(resp., } G_n^R(\mathcal{L}) = G_N^R(\mathcal{L})\textit{)}.$$

Proof. This is a direct consequence of Proposition 2.31. □

We will then drop the subscript n for the stabilized graph, as detailed below.

Definition 2.58. Let $\mathcal{L}$ be a language. If there exists $N \geq 0$ such that, for all $n \geq N$, we have $G_n^L(\mathcal{L}) = G_N^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L}) = G_N^R(\mathcal{L})$), then we denote

$$G^L(\mathcal{L}) = G_N^L(\mathcal{L}) \quad \text{(resp., } G^R(\mathcal{L}) = G_N^R(\mathcal{L})\text{)}.$$

If $\mathcal{L}$ is eventually dendric, we have no prior condition on the multi-cliques $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$. However, by Corollary 2.44, if $\mathcal{L}$ is dendric, then $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are multi-cliques which are acyclic for the coloring and connected. This is unfortunately not an equivalence as shown in the example below.

Example 2.59. Let $\mathcal{L}$ be the Fibonacci language (or any Sturmian language containing 00) and let σ be the morphism defined by $\sigma(0) = 0110$ and $\sigma(1) = 011$. We consider the image $\mathcal{L}'$ of $\mathcal{L}$ under σ. We will prove in Chapter 4 (Corollary 4.78) that $\mathcal{L}'$ is eventually dendric of threshold at most 5. In fact, we can check that the only non-dendric elements are ε and 1 (see Example 4.50) so $\mathcal{L}'$ is eventually dendric of threshold 2. As we are on an alphabet of size 2, any dendric word is right ordinary and left ordinary. This implies that $G^L(\mathcal{L}') = G_2^L(\mathcal{L}')$ and $G^R(\mathcal{L}') = G_2^R(\mathcal{L}')$. We can then easily check that 01 (resp., 10) is the only left special (resp., right special) length-2 word. This shows that $G^L(\mathcal{L}')$ and $G^R(\mathcal{L}')$ are both acyclic for the coloring and connected.

Proposition 2.57 is not an equivalence. Indeed, we saw in Example 2.34 that the graphs $G_n^L(\mathcal{L})$ (resp., $G_n^R(\mathcal{L})$) are equal starting from $n = 1$ for the Chacon language, but this language is not eventually dendric since we showed in Example 1.45 that it is not eventually neutral. Observe however that stabilization of the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$ implies that the first difference of complexity $s_{\mathcal{L}}(n)$ is eventually constant.

If we want to distinguish the case of the Chacon language from the eventually dendric languages, we need to look at the edges labels. By Proposition 2.31, if $\mathcal{L}$ is eventually dendric, with each edge color in $G^L(\mathcal{L})$, we can associate a (right) infinite word x such that, for all large enough n, the label corresponding to that color in $G^L_n(\mathcal{L})$ is the length-n prefix of x. In some sense, this word x can be seen as a left special infinite word if we generalize the notion of extensions.

Definition 2.60. Let $\mathcal{L}$ be a language and let x be a right (resp., left) infinite word. We extend the notion of left (resp., right) extensions by defining

$$E^L_{\mathcal{L}}(x) = \{a : \mathrm{Fac}(ax) \subseteq \mathcal{L}\}$$

$$(\text{resp., } E^R_{\mathcal{L}}(x) = \{a : \mathrm{Fac}(xa) \subseteq \mathcal{L}\}).$$

We then have the following characterization of eventually dendric languages.

Proposition 2.61. *Let $\mathcal{L}$ be a language over $\mathcal{A}$. The following are equivalent:*

1. *$\mathcal{L}$ is eventually dendric;*
2. *$G^L(\mathcal{L})$ is defined and $G^L(\mathcal{L}) = G(\{E^L_{\mathcal{L}}(x) : x \in \mathcal{A}^{\mathbb{N}}\})$;*
3. *$G^R(\mathcal{L})$ is defined and $G^R(\mathcal{L}) = G(\{E^R_{\mathcal{L}}(x) : x \in \mathcal{A}^{-\mathbb{N}}\})$.*

Proof. We only prove the equivalence between the first two claims. The link with the third claim is symmetric. If $\mathcal{L}$ is eventually dendric then the conclusion directly follows from Proposition 2.57 and Proposition 2.31 as explained above.

For the converse, if $G^L(\mathcal{L})$ exists then it is finite as it corresponds to $G^L_N(\mathcal{L})$ for some N. Moreover, it implies that $s_n(\mathcal{L}) = s_N(\mathcal{L})$ for all $n \geq N$, which in turn implies that the set of left special infinite words is finite (and bounded by $s_N(\mathcal{L})$). Indeed, otherwise we can find $s_N(\mathcal{L}) + 1$ different left special infinite words. There exists n such that their length-n prefixes are all distinct, showing that we have strictly more that $s_n(\mathcal{L})$ left special words of length n, which is a contradiction. As the set of left special infinite words is finite, we can find $M \geq N$ such that the length-M prefixes are distinct and each infinite word has the same extensions as its length-M prefix. In other words, $G(\{E^L_{\mathcal{L}}(x) : x \in \mathcal{A}^{\mathbb{N}}\})$ can also be defined using the length-m prefixes of the left special infinite words for all $m \geq M$. Since this graph is also equal to $G^L(\mathcal{L}) = G^L_m(\mathcal{L})$, this shows that the only left special words of length $m \geq M$ are the prefixes of the left special infinite words. We conclude that $\mathcal{L}$ is eventually dendric by Proposition 2.31. $\square$

This proof also shows that, if $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$) is defined, then we can see $G(\{E^L_{\mathcal{L}}(x) : x \in \mathcal{A}^{\mathbb{N}}\})$ as a subgraph of $G^L(\mathcal{L})$ (resp., $G(\{E^R_{\mathcal{L}}(x) : x \in \mathcal{A}^{-\mathbb{N}}\})$ as a subgraph of $G^R(\mathcal{L})$). For the Chacon language, we can show that there is a left special infinite word given by

$$\lim_{n\to\infty} 012\varphi(012)\cdots\varphi^n(012).$$

This follows from the study of extensions done in Example 1.45. As $G^L(\mathcal{L})$ has edges of two colors and the Chacon language is not eventually dendric, this must be the only left special infinite word.

We end this subsection with a comment on the link with asymptotic pairs which are a well-known tool in symbolic dynamics.

Definition 2.62. Two bi-infinite words x, x' are *right asymptotic equivalent* if there exist a right infinite word y and two integers $m, n \in \mathbb{Z}$ such that

$$S^m(x) = z \cdot y \quad \text{and} \quad S^n(x') = z' \cdot y,$$

where z, z' are left infinite words. If $z \neq z'$, we then say that $\{x, x'\}$ is a *right asymptotic pair*. We can similarly define left asymptotic equivalence.

While every right asymptotic pair in a shift space corresponds to a unique pair of extensions of a (also unique) left special infinite word, the converse is not necessarily true. For example, if we have the words $x = za \cdot y$, $x' = z'ab \cdot y$ and $x'' = z''bb \cdot y$, then $\{x, x'\}$ and $\{x, x''\}$ are two asymptotic pairs corresponding to the same extensions $\{a, b\}$ of the same right infinite word y.

2.5 Open questions

The first questions of this chapter concern the conjugacy classes of eventually dendric languages. Indeed, by Proposition 2.16, if X is an eventually dendric shift space of threshold M, then its higher block shift space $X^{(N)}$ is of threshold $M - N + 1$ for all $1 \leq N \leq M$. This implies that X has an eventually dendric conjugate of threshold n for all $1 \leq n \leq M$. In most cases however, no higher block shift space of X is dendric. We therefore ask the following question.

Question 2.1. Is every infinite eventually dendric shift space conjugate to a dendric shift space?

The infinity hypothesis is necessary here. Indeed, if X is a finite eventually dendric shift space, then its conjugates have the same cardinality. Since the only finite dendric shift spaces contain exactly one element, we directly conclude that a finite eventually dendric shift space containing at least two elements has no dendric conjugate. Observe that, in the case of minimal shift spaces, infinity is equivalent to aperiodicity (i.e., not containing any periodic element).

One can also ask a weaker version of the previous question.

Question 2.2. Is every aperiodic eventually dendric shift space a factor of a minimal dendric shift space?

Once again, the answer to this question is no if we remove the aperiodicity hypothesis. Indeed, minimal dendric shift spaces are known to have no rational continuous eigenvalues [BDD$^+$18]. Therefore, they have no (non-trivial) periodic factor by a result of [DG19]. In other words, a finite eventually dendric shift space containing at least two elements cannot be a factor of a minimal dendric shift space.

Observe that the existence of a minimal dendric factor is however always guaranteed as a shift space on a unary alphabet is dendric. We could therefore ask about the existence of a minimal aperiodic dendric factor. We then clearly also have to restrict ourselves to infinite eventually dendric shift spaces.

Question 2.3. Does every infinite eventually dendric shift space have a minimal aperiodic dendric factor?

On the other hand, if X is an eventually dendric shift space of threshold M, no higher block shift space of X will be eventually dendric of threshold $n > \max\{1, M\}$ either. So the same questions can be asked, replacing dendric shift spaces by eventually dendric shift spaces of some fixed threshold n.

Question 2.4. Is every aperiodic eventually dendric shift space conjugate to (resp., factor of) an eventually dendric shift space of threshold n for all $n \geq 1$? Does every aperiodic eventually dendric shift space have an eventually dendric factor of threshold n for all $n \geq 1$?

Observe that it in fact suffices to answer these questions for arbitrarily large values of n since we can then reduce the threshold by looking at higher block shift spaces.

The last question on the conjugacy classes concerns the uniqueness of the dendric shift spaces.

Question 2.5. Can two dendric shift spaces which are not image of one another via a letter-to-letter morphism be conjugate?

We now turn to a question on Section 2.3. Indeed, we showed that eventual dendricity was equivalent to many other eventual properties. In particular, if we restrict ourselves to linear factor complexity, being eventually dendric is equivalent to being eventually connected or eventually strong or neutral (see Proposition 2.30). However, as shown in Example 2.29 with the language $\mathcal{A}^*$, it is not true in general. This example is however both (eventually) connected and (eventually) strong or neutral, leaving therefore the possibility for these two properties to be equivalent without restriction on the factor complexity.

Question 2.6. Is a language eventually connected if and only if it is eventually strong or neutral?

Clearly, if a language is eventually connected, then it eventually is strong or neutral by Proposition 2.10. Therefore, a way to answer (by the negative) the previous question would be to find an eventually strong or neutral language $\mathcal{L}$ containing infinitely many non-connected words. Observe that this language $\mathcal{L}$ will not have a linear complexity and these strong or neutral but not connected words must have at least three left and right extensions.

Chapter 3

Extensions' impact on complexity and return words

In Chapter 1 and Chapter 2, we presented general tools as well as started to familiarize ourselves with the languages that will follow us throughout this work. It is now time to introduce more specific concepts and start to go deeper into the results. This chapter is centered around extensions of words and how they can help us count other objects, namely the words' extensions in the language, the words in the image under a morphism and the return words.

Recall that Proposition 1.18 shows the strong link between the words' extensions and the language's complexity. This result was the main motivation behind the study of neutral languages and, as explained in Chapter 1, is the key to get a complete description of the factor complexities of the languages that we study in this work.

We start this chapter with Section 3.1 in which we prove a generalization of Proposition 1.18. More precisely, after recalling the classical notions of prefix and suffix codes, we show that, if we consider words forming a maximal prefix (or suffix) code, then we can similarly relate the number of left (or right) extensions with the first difference of complexity. This result will be particularly useful in Sections 3.2 and 3.3.

Section 3.2 then focuses on the evolution of the factor complexity when applying a non-erasing morphism. It is well known that the complexity can grow at most by a multiplicative constant [AS03, CN10]. Using the same underlying tools but with a more careful analysis, we show here that, for some languages, we can vastly improve this result and replace the multiplicative constant by an additive constant (Corollary 3.14). This is the case for eventually strong or neutral languages but not only. We also use the tools developed in this section to look at the factor complexity of the images of the Thue-Morse language.

We then turn to the notion of return words in Section 3.3. This concept was the main focus of the authors in the original paper on dendric languages [BDFD+15a] and is arguably the main reason for studying dendric languages due to the particularly strong properties of their return words. We determine the number of return words in an eventually neutral language (Corollary 3.32). We also recall the famous Return Theorem and prove that eventually dendric languages are stable under derivation.

3.1 Codes and extensions

A set S of words is a *code* if no word can be factorized in two different ways over S. In other words, if we associate a letter with each element of S as a sort of "antecedent", then each word corresponds to at most one sequence of such letters. This is the reason behind the terminology of code as any word which can be decoded is uniquely decodable. This is also the idea behind recognizability, a well-known concept in symbolic dynamics that we will mention again later (in Section 4.5). For more results on codes, we invite the reader to look in [BPR10].

A particular example of a code is given by the set of words of the same length. It is then natural to wonder if a result using this sort of set can be extended to other kinds of codes. In this section, we focus on Cassaigne's result on the link between complexity and number of extensions (Proposition 1.18). We show that it can be extended by summing over words who form what is called an $\mathcal{L}$-maximal prefix code (or suffix code).

Definition 3.1. Two words u and v are *prefix comparable* if u is a prefix of v or v is a prefix of u. A *prefix code* is a set of words S such that for all distinct $u, v \in S$, u and v are not prefix comparable. It is moreover said to be *$\mathcal{L}$-maximal* for a language $\mathcal{L}$ if $S \subseteq \mathcal{L}$ and any $w \in \mathcal{L}$ is prefix comparable with an element of S.

We similarly define the notions of *suffix comparable*, *suffix code* and *$\mathcal{L}$-maximal suffix code* by replacing "prefix" with "suffix".

Remark 3.2. Distinct words of the same length cannot be prefix comparable. This implies that any set containing only words of a given length $n \neq 0$ is both a prefix and a suffix code (we then say that it is a *bifix code*). Another consequence is that the only candidate for an $\mathcal{L}$-maximal prefix code included in $\mathcal{L}_n$ is $\mathcal{L}_n$ itself. As every element of $\mathcal{L}$ can either be extended on the right in a element of $\mathcal{L}_n$ ($\mathcal{L}$ is biextendable) or has an element of $\mathcal{L}_n$ as a prefix ($\mathcal{L}$ is factorial), this shows that $\mathcal{L}_n$ is indeed en $\mathcal{L}$-maximal prefix code. We similarly reach the same conclusion for suffix codes.

The trade-off when we replace words of the same length by words in an $\mathcal{L}$-maximal prefix code in Proposition 1.18 is that we have to take into account the multiplicities of some words. This is detailed in the technical lemma below.

Lemma 3.3. *Let $\mathcal{L}$ be language and let $S \subseteq \mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M}$ be an $\mathcal{L}$-maximal prefix code. Then*

$$s_{\mathcal{L}}(N) + \sum_{\substack{w \in \mathcal{L}_{\geq N} \\ w \in \mathrm{Pref}^*(S)}} m_{\mathcal{L}}(w) = \sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1) = s_{\mathcal{L}}(M) - \sum_{\substack{w \in \mathcal{L}_{\leq M-1} \\ \mathrm{Pref}(w) \cap S \neq \emptyset}} m_{\mathcal{L}}(w).$$

Similarly, if $S \subseteq \mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M}$ is an $\mathcal{L}$-maximal suffix code, then

$$s_{\mathcal{L}}(N) + \sum_{\substack{w \in \mathcal{L}_{\geq N} \\ w \in \mathrm{Suff}^*(S)}} m_{\mathcal{L}}(w) = \sum_{w \in S} (\#E^R_{\mathcal{L}}(w) - 1) = s_{\mathcal{L}}(M) - \sum_{\substack{w \in \mathcal{L}_{\leq M-1} \\ \mathrm{Suff}(w) \cap S \neq \emptyset}} m_{\mathcal{L}}(w).$$

Proof. We only prove the case where S is a prefix code. The case of a suffix code is symmetric.

To prove the left equality, let us denote $n = \max\{|w| : w \in S\}$ and let us proceed by induction on $n \in [N, M]$. If $n = N$, then S is an $\mathcal{L}$-maximal prefix code included in $\mathcal{L}_N$ which implies, as explained above, that $S = \mathcal{L}_N$. Using Proposition 1.18, we directly conclude that

$$s_{\mathcal{L}}(N) = \sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1),$$

which is the desired equality as no word of $\mathcal{L}_{\geq N}$ can be a proper prefix of a word in S.

Assume now that $n > N$ and that the equality is true for any $\mathcal{L}$-maximal prefix code $S' \subseteq \mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M}$ such that $\max\{|w| : w \in S'\} = n - 1$. We define the new set

$$S' = (S \cap \mathcal{L}_{<n}) \cup \{w \in \mathcal{L}_{n-1} : \exists a \in E^R_{\mathcal{L}}(w) \text{ st. } wa \in S\}.$$

In other words, we truncate on the right the elements of length n in S. By construction, $\max\{|w| : w \in S'\} = n - 1$ and $S' \subseteq \mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M}$. Let us show that it is an $\mathcal{L}$-maximal prefix code. As S is a prefix code, so is $S' \cap S$, and no element of $S' \cap S$ can be a prefix of an element of $S' \setminus S$. Since $S' \cap S \subseteq \mathcal{L}_{\leq n-1}$ and $S' \setminus S \subseteq \mathcal{L}_{n-1}$ this is sufficient to conclude that S' is a prefix code. It is $\mathcal{L}$-maximal as every element of $\mathcal{L}$ either had a prefix in S meaning that it also has a prefix in S', or it was a proper prefix of an element in S in which case it is a prefix of an element in S'.

We now look at the link between the sum over S and the sum over S'. By definition of S', we have

$$\begin{aligned}&\sum_{w\in S}(\#E^L_{\mathcal{L}}(w)-1)-\sum_{w\in S'}(\#E^L_{\mathcal{L}}(w)-1)\\ =&\sum_{w\in S\cap\mathcal{L}_n}(\#E^L_{\mathcal{L}}(w)-1)-\sum_{w\in S'\setminus S}(\#E^L_{\mathcal{L}}(w)-1).\end{aligned}$$

Let $w \in S' \setminus S$, let a such that $wa \in S$ and let $b \in E^R_{\mathcal{L}}(w)$. By $\mathcal{L}$-maximality of S, wb is prefix comparable with an element of S. However, $|wb| = n$ and no prefix of w is in S since $wa \in S$. By definition of n, this implies that wb is in S. In other words,

$$S \cap \mathcal{L}_n = \{wa : w \in S' \setminus S, a \in E^R_{\mathcal{L}}(w)\}.$$

We then have

$$\begin{aligned}&\sum_{w\in S}(\#E^L_{\mathcal{L}}(w)-1)-\sum_{w\in S'}(\#E^L_{\mathcal{L}}(w)-1)\\ =&\sum_{w\in S'\setminus S}\left(\sum_{a\in E^R_{\mathcal{L}}(w)}(\#E^L_{\mathcal{L}}(wa)-1)\quad-\#E^L_{\mathcal{L}}(w)+1\right)\\ =&\sum_{w\in S'\setminus S}m_{\mathcal{L}}(w).\end{aligned}$$

Observe also that the elements of $S' \setminus S$ are proper prefixes of elements of S but are not proper prefixes of elements of S', and these are the only such words, i.e., $S' \setminus S = \mathrm{Pref}^*(S) \setminus \mathrm{Pref}^*(S')$. Moreover, by definition of S', we have $\mathrm{Pref}^*(S') \subseteq \mathrm{Pref}^*(S)$. Therefore, with the induction hypothesis on S', we conclude that

$$\begin{aligned}
&\sum_{w\in S}(\#E^L_{\mathcal{L}}(w)-1)\\
&=\sum_{w\in S'}(\#E^L_{\mathcal{L}}(w)-1)+\sum_{w\in \mathrm{Pref}^*(S)\setminus \mathrm{Pref}^*(S')} m_{\mathcal{L}}(w)\\
&=s_{\mathcal{L}}(N)+\sum_{w\in \mathrm{Pref}^*(S)} m_{\mathcal{L}}(w)
\end{aligned}$$

We now turn to the right equality. The idea of the proof is quite similar to the one we just did so we will not give as much detail. The main difference is that we do a decreasing induction on $m=\min\{|w| : w\in S\}$. The case $m=M$ is a direct consequence of Proposition 1.18. For the induction step, we define

$$S'=(S\cap\mathcal{L}_{>m})\cup\{wa\in\mathcal{L}_{m+1} : w\in S, a\in E^R_{\mathcal{L}}(w)\}.$$

It is an $\mathcal{L}$-maximal prefix code and by construction $\min\{|w| : w\in S'\}=m+1$. Moreover, we can show that

$$\sum_{w\in S}(\#E^L_{\mathcal{L}}(w)-1)-\sum_{w\in S'}(\#E^L_{\mathcal{L}}(w)-1)=-\sum_{w\in S\setminus S'} m_{\mathcal{L}}(w).$$

To conclude the induction step, it then suffices to observe that $S\setminus S'$ is exactly the set of words having a prefix in S but no prefix in S'. □

We make several remarks about this result.

First, if w is a proper prefix of an element of S, then $w\in\mathcal{L}_{\leq M-1}$. Similarly, if w has a prefix in S, then $w\in\mathcal{L}_{\geq N}$. In particular, this shows that when $N=M$, the sums of the multiplicities disappear and we obtain exactly Proposition 1.18.

Second, as S is an $\mathcal{L}$-maximal prefix code, every word of $\mathcal{L}_{\geq N}\cap\mathcal{L}_{\leq M-1}$ is either a proper prefix of an element of S or has a prefix in S, and not both. Therefore, each word of $\mathcal{L}_{\geq N}\cap\mathcal{L}_{\leq M-1}$ contributes either to the sum of multiplicities on the left or on the right. This is coherent with the fact that $s_{\mathcal{L}}(M)-s_{\mathcal{L}}(N)$ is the sum of the multiplicities of all the words in $\mathcal{L}_{\geq N}\cap\mathcal{L}_{\leq M-1}$.

Third, due to what we just observed, it was not necessary to prove the right equality as soon as we knew the left equality to be true. We still gave a sketch of proof to show that the same ideas could be used for both equalities.

If we are not interested in the actual value of the sum and if the multiplicities have the same sign, we obtain the simpler statement proved in [Ghe23].

Proposition 3.4. *Let $\mathcal{L}$ be a language and let $S \subseteq \mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M}$ be an $\mathcal{L}$-maximal prefix code.*

1. *If the elements of $\mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M-1}$ are weak or neutral, then*
$$s_{\mathcal{L}}(N) \geq \sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1) \geq s_{\mathcal{L}}(M).$$
2. *If the elements of $\mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq M-1}$ are strong or neutral, then*
$$s_{\mathcal{L}}(N) \leq \sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1) \leq s_{\mathcal{L}}(M).$$

If S is an $\mathcal{L}$-maximal suffix code, we have the same inequalities by considering the right extensions instead.

We will also need a version of Proposition 3.4 for a potentially infinite prefix (or suffix) code. This is the object of the following result, which can be found with a slightly different statement in [DP21].

Proposition 3.5. *Let $\mathcal{L}$ be an eventually neutral language of threshold N. If $S \subseteq \mathcal{L}_{\geq N}$ is a prefix code, then*
$$\sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1) \leq s_{\mathcal{L}}(N)$$
and we have the equality whenever S is $\mathcal{L}$-maximal and finite. If S is a suffix code, we have the same result by considering the right extensions instead.

Proof. We only prove the case of a prefix code. For all $n \geq N$, let us consider the set $S_n = S \cap \mathcal{L}_{\leq n}$. By construction, we have
$$\sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1) = \lim_{n \to \infty} \sum_{w \in S_n} (\#E^L_{\mathcal{L}}(w) - 1).$$
Moreover, for each n, S_n is a prefix code included in $\mathcal{L}_{\geq N} \cap \mathcal{L}_{\leq n}$ and it can be completed into an $\mathcal{L}$-maximal prefix code S'_n by adding the elements of $\mathcal{L}_n$ which have no prefix in S_n. Therefore, we can apply Proposition 3.4 to obtain
$$\sum_{w \in S_n} (\#E^L_{\mathcal{L}}(w) - 1) \leq \sum_{w \in S'_n} (\#E^L_{\mathcal{L}}(w) - 1) = s_{\mathcal{L}}(N)$$
since $s_{\mathcal{L}}(n) = s_{\mathcal{L}}(N)$ by Corollary 1.43. As it is true for all $n \geq N$, this implies that
$$\sum_{w \in S} (\#E^L_{\mathcal{L}}(w) - 1) \leq s_{\mathcal{L}}(N).$$
If S is finite and $\mathcal{L}$-maximal, there exists n such that $S = S_n = S'_n$ and the equality follows. $\square$

3.2 Factor complexity of morphic images

Factor complexity is one of the fundamental notions in combinatorics on words and is usually one of the first questions we ask ourselves when presented with a new language. Among the most well-known results on this topic, we can cite Pansiot's work ([Pan84]) on the complexity of purely morphic words. We will however be more interested in a folklore result (found in [AS03] or in [CN10] for example) stating that, when applying a non-erasing morphism, the factor complexity grows at most by a multiplicative constant, which is in fact given by the width of the morphism.

Definition 3.6. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be morphism. The *width* of σ is

$$\|\sigma\| = \max_{a\in\mathcal{A}} |\sigma(a)|.$$

Proposition 3.7. *Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. For all $n \geq 0$, we have*

$$p_{\sigma(\mathcal{L})}(n) \leq \|\sigma\| \cdot p_{\mathcal{L}}(n).$$

In this section, we refine this result to prove that, for some starting languages (including eventually dendric languages), the complexity grows at most by an additive constant. The main ideas of the proofs and preliminary versions of the results can be found in [Ghe23].

The common idea behind the proofs of Proposition 3.7 and what we do in this section is that we do not directly look at the words in the image but at something that we call here "covering" and which is only implicitly mentioned in the proofs of the multiplicative result.

Definition 3.8. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. A *covering* of a non empty word $u \in \mathcal{B}^+$ is a pair $(w, k) \in \mathcal{A}^+ \times \mathbb{N}$ such that $u = \sigma(w)_{[k+1,k+|u|]}$ and w is minimal, i.e., $k + 1 \leq |\sigma(w_1)|$ and $k + |u| > \left|\sigma(w_{[1,|w|-1]})\right|$.

For a language $\mathcal{L}$ and for $n \geq 1$, we denote $C_{\mathcal{L},\sigma}(n)$ the set of coverings (w, k) of length-n words such that $w \in \mathcal{L}$. We then denote $c_{\mathcal{L},\sigma}(n) = \#C_{\mathcal{L},\sigma}(n)$. As we usually consider only one morphism at a time, we will drop the subscript σ.

Example 3.9. Let σ be such that $\sigma(a) = ab$ and $\sigma(b) = abb$. The coverings of $babb$ are given by $(ab, 1)$ and $(bb, 2)$ as $babb$ can be seen in the image of both words, skipping the first (resp., the first two) letter(s). If $\mathcal{L}$ is a Sturmian language over the alphabet $\{a, b\}$ containing the word aa (and therefore, not the word bb), we have $(ab, 1) \in C_{\mathcal{L}}(4)$ and $(bb, 2) \notin C_{\mathcal{L}}(4)$.

If (w, k) is a covering of a length-n word then, by minimality of w, we directly conclude that $|w| \leq n$ since we only consider non-erasing morphisms. In particular, we can precisely describe the coverings of the letters.

Remark 3.10. If (w, k) is a covering of a letter a, w is also a letter and k can then vary between 0 and $|\sigma(w)| - 1$ thus

$$C_{\mathcal{L}}(1) = \{(a, k) : a \in \mathcal{A}, k \in \{0, 1, \ldots, |\sigma(a)| - 1\}\}$$

and

$$c_{\mathcal{L}}(1) = \sum_{a \in \mathcal{A}} |\sigma(a)|,$$

independently of the language $\mathcal{L}$ over $\mathcal{A}$.

The reason why we look at coverings is the clear relation with words of the image: a length-n word is in the image of $\mathcal{L}$ if and only if it has a covering in $C_{\mathcal{L}}(n)$. This gives a trivial link between the number of coverings and the factor complexity of the image.

Lemma 3.11. *Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. For all $n \geq 1$, we have*

$$p_{\sigma(\mathcal{L})}(n) \leq c_{\mathcal{L}}(n).$$

Proof. The map

$$\varphi\colon C_{\mathcal{L}}(n) \to (\sigma(\mathcal{L}))_n, \quad (w, k) \mapsto \sigma(w)_{[k+1, k+n]}$$

is well defined and surjective. This directly gives a inequality between the cardinals of the two sets. □

Remark 3.12. The previous result is only an inequality in general since a word can have multiple coverings. For example, a letter has as many coverings as the number of its occurrences in images of letters. However, if we apply a morphism such that no letter appears twice in the image of a letter or in the images of two different letters, then every word of the image has exactly one covering. In that case, we have $p_{\sigma(\mathcal{L})}(n) = c_{\mathcal{L}}(n)$ for all $n \geq 1$ and for any language $\mathcal{L}$. Such a morphism can be defined for any given vector of lengths $(|\sigma(a)|)_{a \in \mathcal{A}}$.

Instead of directly bounding the complexity of the image, it suffices to bound the number of coverings. A simple bound on the number of coverings can be obtained by saying that if $(w, k) \in C_{\mathcal{L}}(n)$, then w is a prefix of a length-n word and $k < |\sigma(w_1)| \leq \|\sigma\|$ so there are at most $p_{\mathcal{L}}(n)$ possibilities for w, and at most $\|\sigma\|$ choices for k for each w. This shows that

$$p_{\sigma(\mathcal{L})}(n) \leq c_{\mathcal{L}}(n) \leq \|\sigma\| \cdot p_{\mathcal{L}}(n),$$

which is exactly Proposition 3.7.

We can do a more careful analysis of the number of coverings, leading to better bounds on the complexity. This is the object of the following result.

Theorem 3.13. *Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. For all $n \geq 1$,*

$$c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) = s_{\mathcal{L}}(n) - \sum_{\substack{w \in \mathcal{L}_{\leq n-1} \\ |\sigma(w)| \geq n}} m_{\mathcal{L}}(w).$$

Proof. Observe that the elements of $C_{\mathcal{L}}(n)$ and of $C_{\mathcal{L}}(n+1)$ are linked. Indeed, each $(w,k) \in C_{\mathcal{L}}(n)$ is related to one or several elements of $C_{\mathcal{L}}(n+1)$ in exactly one of the following ways:

- if $|\sigma(w)| = k+n$, then $(w,k) \notin C_{\mathcal{L}}(n+1)$ but for all $a \in E^R_{\mathcal{L}}(w)$, (wa,k) is an element of $C_{\mathcal{L}}(n+1)$;

- otherwise, we have $|\sigma(w)| \geq k+n+1$ thus (w,k) itself is in $C_{\mathcal{L}}(n+1)$.

Moreover, with this technique, we obtain every element of $C_{\mathcal{L}}(n+1)$ exactly once. Indeed, if $(wa,k) \in C_{\mathcal{L}}(n+1)$ where a is a letter, then we have two cases:

- if $|\sigma(w)| < k+n$, then (wa,k) is an element of $C_{\mathcal{L}}(n)$ as wa is still minimal, i.e., the inequalities of Definition 3.8 are satisfied;

- if $|\sigma(w)| \geq k+n$, then (w,k) is an element of $C_{\mathcal{L}}(n)$ as we must have $|\sigma(w)| = k+n$ so (wa,k) is not minimal anymore but (w,k) is.

In other words, each $(w,k) \in C_{\mathcal{L}}(n)$ corresponds to exactly one element of $C_{\mathcal{L}}(n+1)$ if $|\sigma(w)| \geq k+n+1$ and to exactly $\#E^R_{\mathcal{L}}(w)$ elements of $C_{\mathcal{L}}(n+1)$ if $|\sigma(w)| = k+n$.

This implies that, for all $n \geq 1$,

$$c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) = \sum_{\substack{(w,k) \in C_{\mathcal{L}}(n) \\ |\sigma(w)| = k+n}} (\#E^R_{\mathcal{L}}(w) - 1).$$

However, for a given word w, the value of k such that $|\sigma(w)| = k + n$ is unique. Moreover, if (w, k) is a covering in $C_{\mathcal{L}}(n)$, we must have $0 \leq k < |\sigma(w_1)|$ so such a k exists if and only if

$$|\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|.$$

Let W_n be the set of words satisfying this inequality, i.e.,

$$W_n = \{w \in \mathcal{L} : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}.$$

We then have

$$c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) = \sum_{w \in W_n} (\# E^R_{\mathcal{L}}(w) - 1)$$

for all $n \geq 1$.

By definition and since σ is non-erasing, the set W_n is an $\mathcal{L}$-maximal suffix code included in $\mathcal{L}_{\leq n} \cap \mathcal{L}_{\geq \lceil \frac{n}{\|\sigma\|} \rceil}$. By Lemma 3.3 we deduce that

$$c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) = s_{\mathcal{L}}(n) - \sum_{\substack{w \in \mathcal{L}_{\leq n-1} \\ \mathrm{Suff}(w) \cap W_n \neq \emptyset}} m_{\mathcal{L}}(w).$$

The conclusion then follows from the observation that a word w has a suffix in W_n if and only if $|\sigma(w)| \geq n$. □

By definition, the elements of $C_{\mathcal{L}}(n)$ only depend on the initial language and the lengths of the images of the letters. This is coherent with the previous result. This observation should be compared to Remark 3.12. Indeed, we then see that the language and the lengths of the images determine the number of coverings, and the content of the images then determine the link with the factor complexity of the image.

We can use Theorem 3.13 to obtain bounds on the evolution of the complexity for some particular languages.

Corollary 3.14. *Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. If there exists $N \geq 0$ such that, for all $n \geq N$,*

$$\sum_{\substack{w \in \mathcal{L}_{\leq n-1} \\ |\sigma(w)| \geq n}} m_{\mathcal{L}}(w) \geq 0,$$

then there exists $C \in \mathbb{Z}$ such that

$$p_{\sigma(\mathcal{L})}(n) \leq C + p_{\mathcal{L}}(n)$$

for all $n \geq 0$.

Proof. By Theorem 3.13, we have

$$c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) \leq s_{\mathcal{L}}(n) = p_{\mathcal{L}}(n+1) - p_{\mathcal{L}}(n)$$

for all $n \geq N$. Therefore,

$$c_{\mathcal{L}}(n) \leq p_{\mathcal{L}}(n) + c_{\mathcal{L}}(N) - p_{\mathcal{L}}(N)$$

for all $n \geq N$. The conclusion then follows from Lemma 3.11. □

If we apply these results to some of the families introduced in Chapter 1, we obtain the following result [Ghe23].

Proposition 3.15. *Let $\mathcal{L}$ be a language over $\mathcal{A}$ and let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism.*

1. *If $\mathcal{L}$ is eventually neutral with threshold N then there exists $C \in \mathbb{Z}$ such that, for any $n \geq \max\{1, N\|\sigma\|\}$,*

   $$c_{\mathcal{L}}(n) = C + p_{\mathcal{L}}(n).$$

   In particular, if $\mathcal{L}$ is neutral, we have

   $$c_{\mathcal{L}}(n) = \sum_{a\in\mathcal{A}} |\sigma(a)| \ + (\#\mathcal{A} - 1)(n-1)$$

   for all $n \geq 1$.

2. *If $\mathcal{L}$ is eventually strong or neutral then there exists $C \in \mathbb{Z}$ such that, for any $n \geq 0$,*

   $$p_{\sigma(\mathcal{L})}(n) \leq C + p_{\mathcal{L}}(n).$$

Proof. The number of coverings in the eventually neutral case is obtained using Theorem 3.13. In the neutral case, since the growth rate of the factor complexity is given by $\#\mathcal{A} - 1$, we obtain the exact number of coverings by Remark 3.10.

For eventually strong or neutral languages, it is easy to check that they satisfy the hypothesis of Corollary 3.14. We then obtain the same bound on the factor complexity of the image. □

We can also apply Corollary 3.14 for a slightly more general family of languages which includes the Chacon language.

Proposition 3.16. *Let $\mathcal{L}$ be a language for which there exists N such that there is a injective map φ from the weak words of $\mathcal{L}_{\geq N}$ to the strong words of $\mathcal{L}_{\geq N}$ satisfying the following conditions for all weak $w \in \mathcal{L}_{\geq N}$:*

1. *$\varphi(w)$ is at least as strong as w is weak, i.e., $m_{\mathcal{L}}(w) + m_{\mathcal{L}}(\varphi(w)) \geq 0$;*
2. *the words w and $\varphi(w)$ have the same Parikh vector, i.e., for each letter a, w and $\varphi(w)$ contain the same number of occurrences of a.*

For every non-erasing morphism σ, there exists $C \in \mathbb{Z}$ such that

$$p_{\sigma(\mathcal{L})}(n) \leq C + p_{\mathcal{L}}(n)$$

for all $n \geq 0$.

Proof. For every weak $w \in \mathcal{L}_{\geq N}$ and for all $n \geq 0$, we have

$$|\sigma(w)| \geq n \iff |\sigma(\varphi(w))| \geq n$$

since w and $\varphi(w)$ have the same Parikh vector. Moreover, w and $\varphi(w)$ also have the same length. By assumption on the multiplicities, the language $\mathcal{L}$ then satisfies the hypothesis of Corollary 3.14 for every non-erasing morphism, which ends the proof. □

We would like to make a comment on the constant C of Corollary 3.14. In Proposition 3.7, the multiplicative constant is given by $\|\sigma\|$ and only depends on the morphism and not on the language. Here however, the constant C depends both on the language and on the morphism. One way to see that it is unavoidable is to look at images of periodic languages (languages with bounded complexity). Indeed, let $\mathcal{L}$ be a language with complexity eventually equal to P and let σ be a k-uniform morphism such that no letter appears twice in the image of a letter or appears in two images of letters. We can easily see that $\sigma(\mathcal{L})$ has a complexity eventually equal to $k \cdot P$. This is an example of a case where the multiplicative bound is tight, which does not contradict the existence of a additive bound. Indeed, we can see $k \cdot P$ as $P + (k-1) \cdot P$. In other words, the constant C is here equal to $(k-1) \cdot P$ which depends both on the language and on the morphism.

We end this section with an application of Theorem 3.13 to show that Corollary 3.14 is not true for any language. We study here the case of the Thue-Morse language.

Example 3.17. The *Thue-Morse* language is the language generated by the morphism $\tau\colon \{0,1\}^* \to \{0,1\}^*$ such that $\tau(0) = 01$ and $\tau(1) = 10$. Equivalently, it is generated by the Thue-Morse sequence

$$\lim_{n\to\infty} \tau^n(0) = 0110100110010110 \cdots .$$

Proposition 3.18. *Let $\mathcal{L}$ be the Thue-Morse language and let $k \geq 1$. If $\sigma\colon \{0,1\}^* \to \mathcal{B}^*$ is a non-erasing morphism such that $|\sigma(0)| + |\sigma(1)| = 2k$, then for all $n \geq 3\|\sigma\| + 1$*

$$c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) = s_{\mathcal{L}}(n) + \begin{cases} 2 & \text{if } \exists \ell \text{ st. } n \in]\max\{\frac{3}{2}, r\}2^\ell, \min\{\frac{3}{2}r, 2\}2^\ell] \\ -2 & \text{if } \exists \ell \text{ st. } n \in]\max\{\frac{3}{2}r, 2\}2^\ell, \min\{\frac{3}{2}, r\}2^{\ell+1}] \\ 0 & \text{otherwise} \end{cases}$$

where $k = 2^j \cdot r$ with $j \in \mathbb{N}$, $r \in [1, 2[$.

Proof. Let τ denote the morphism generating $\mathcal{L}$ (see Example 3.17). It is well known that the strong words in $\mathcal{L}$ are ε, and $\tau^i(0)$ and $\tau^i(1)$ (of length 2^i) for $i \geq 1$. The weak words are $\tau^i(010)$ and $\tau^i(101)$ (of length $3 \cdot 2^i$) for $i \geq 0$ (see [Cas97] for example). In particular, by definition of τ, all the non-neutral words of length at least 4 contain as many 0's as 1's. This shows that, if $|w| \geq 4$ and $m_{\mathcal{L}}(w) \neq 0$, then $|\sigma(w)| = k|w|$. Observe also that, since we are on an alphabet of size 2, $m_{\mathcal{L}}(w) \in \{-1, 0, 1\}$ for all $w \in \mathcal{L}$. Using these observations and Theorem 3.13, we deduce that, for all $n \geq 3\|\sigma\| + 1$, we have

$$\begin{aligned} c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) &= s_{\mathcal{L}}(n) \\ &\quad - \#\{w \in \mathcal{L}_{\leq n-1} : |\sigma(w)| \geq n \text{ and } w \text{ is strong}\} \\ &\quad + \#\{w \in \mathcal{L}_{\leq n-1} : |\sigma(w)| \geq n \text{ and } w \text{ is weak}\} \\ &= s_{\mathcal{L}}(n) \\ &\quad - 2\#\{i \geq 1 : \frac{n}{k} \leq 2^i \leq n-1\} \\ &\quad + 2\#\{i \geq 0 : \frac{n}{k} \leq 3 \cdot 2^i \leq n-1\}. \end{aligned}$$

Observe that, for all $n \geq 3\|\sigma\| + 1$,

$$\begin{aligned} \#\{i \geq 1 : \frac{n}{k} \leq 2^i \leq n-1\} &= \#\{i \geq 1 : 2^i < n \leq 2^i \cdot k\} \\ &= \begin{cases} j+1 & \text{if } \exists \ell \text{ st. } n \in]2^\ell, r \cdot 2^\ell] \\ j & \text{if } \exists \ell \text{ st. } n \in]r \cdot 2^\ell, 2^{\ell+1}] \end{cases} \end{aligned}$$

and

$$\begin{aligned} \#\{i \geq 0 : \frac{n}{k} \leq 3 \cdot 2^i \leq n-1\} &= \#\{i \geq 0 : 3 \cdot 2^i < n \leq 3 \cdot 2^i \cdot k\} \\ &= \begin{cases} j+1 & \text{if } \exists \ell \text{ st. } n \in]3 \cdot 2^\ell, 3 \cdot r \cdot 2^\ell] \\ j & \text{if } \exists \ell \text{ st. } n \in]3 \cdot r \cdot 2^\ell, 3 \cdot 2^{\ell+1}] \end{cases}. \end{aligned}$$

Combining the two and using the fact that $r \in [1, 2[$, we obtain

$$c_{\mathcal{L}}(n+1)-c_{\mathcal{L}}(n) = s_{\mathcal{L}}(n)+\begin{cases} 2 & \text{if } \exists \ell \text{ st. } n \in]\max\{\frac{3}{2}, r\}2^{\ell}, \min\{\frac{3}{2}r, 2\}2^{\ell}] \\ -2 & \text{if } \exists \ell \text{ st. } n \in]\max\{\frac{3}{2}r, 2\}2^{\ell}, \min\{\frac{3}{2}, r\}2^{\ell+1}] \\ 0 & \text{otherwise} \end{cases}$$

which ends the proof. $\square$

Corollary 3.19. *Let $\mathcal{L}$ be the Thue-Morse language and let $k \geq 1$. If k is a power of 2, then for all non-erasing morphisms $\sigma\colon \{0,1\}^* \to \mathcal{B}^*$ such that $|\sigma(0)| + |\sigma(1)| = 2k$, there exists $C \in \mathbb{Z}$ such that*

$$p_{\sigma(\mathcal{L})}(n) \leq C + p_{\mathcal{L}}(n)$$

for all $n \geq 0$. If k is not a power of 2, then there exists a non-erasing morphism $\sigma\colon \{0,1\}^* \to \mathcal{B}^*$ such that $|\sigma(0)|+|\sigma(1)| = 2k$ and $p_{\sigma(\mathcal{L})}(n)-p_{\mathcal{L}}(n)$ is unbounded.

Proof. Let us use the notations of Proposition 3.18. If k is a power of 2 then $r = 1$. We then have $c_{\mathcal{L}}(n+1) - c_{\mathcal{L}}(n) = s_{\mathcal{L}}(n)$ for any large enough n so the conclusion follows as in Corollary 3.14.

If k is not a power of 2, then let us denote $c = \min\{\frac{3}{2}r, 2\} - \max\{\frac{3}{2}, r\}$ and $d = \min\{3, 2r\}-\max\{\frac{3}{2}r, 2\}$. Since $r \in]1, 2[$, we can show that $2c > d$ by considering the three cases $r \geq \frac{3}{2}$, $r \in [\frac{4}{3}, \frac{3}{2}[$ and $r < \frac{4}{3}$. Using Remark 3.12, there exists σ such that $p_{\sigma(\mathcal{L})}(n) = c_{\mathcal{L}}(n)$ for all $n \geq 0$. The function

$$N \mapsto \sum_{n=0}^{N-1} (s_{\sigma(\mathcal{L})}(n) - s_{\mathcal{L}}(n)) = p_{\sigma(\mathcal{L})}(N) - p_{\mathcal{L}}(N)$$

is unbounded. Indeed, if we look at the non-zero values of the function $s_{\sigma(\mathcal{L})}(n) - s_{\mathcal{L}}(n)$ for n large enough, by Proposition 3.18, we have $c \cdot 2^{\ell}$ times the value 2, then $d \cdot 2^{\ell}$ times the value -2 , then 2 again $c \cdot 2^{\ell+1}$ times and so on. Since $2c > d$, the partial sums are unbounded. $\square$

3.3 Extensions and return words

Some of the first, and arguably most important, results on dendric languages are related to return words. In fact, these results (that we recall in this section) were the starting point of the work presented in Chapter 5.

Return words for a given factor can be summarized as what one will read starting from that factor before seeing the same factor again. This idea of returning to the same factor is related to induction, an important operation in dynamical systems.

For this work, we mostly use return words as the basic blocs needed to derive a word or shift space. The idea is that, starting from a bi-infinite word, one can cut it before (or after) each occurrence of a given factor. The small blocs obtained are return words, and if we code each return word by a different letter, we obtain an associated sequence called the derived word.

In this section, we essentially present already known results or slight generalizations of them. We start by giving a more precise definition of return words for a set or a word. In Subsection 3.3.1, we then look at the number of return words in an (eventually) neutral language. Afterwards, in Subsection 3.3.2, we turn to the concept of derivation of languages for which both the families of dendric and of eventually dendric languages are stable.

While generally studied with respect to a single word, return words can be defined for a set of words, as long as this set has the following property: no element of the set is a factor of another element. Conceptually, the problem when this condition is not satisfied arises when starting with the longer word as we have in some sense already returned to the shorter one (and maybe even more). We introduce the following terminology, by analogy with the notions of prefix and suffix codes.

Definition 3.20. A set S is a *factor code* if no element of S is a factor of another element of S.

We will moreover assume that S is not empty and $S \neq \{\varepsilon\}$, therefore $\varepsilon \notin S$.

Example 3.21. The set $\{001, 010, 1100\}$ is a factor code but $\{01, 1010\}$ is not as 01 is a factor of 1010.

Observe that the notion of factor code is strictly stronger than that of bifix code (i.e., both a prefix and a suffix code) as shown by the previous example with $\{01, 1010\}$. On the other hand, if $S = \{w\}$ with $w \neq \varepsilon$, then S is trivially a factor code.

Remark 3.22. If $S \subseteq \mathcal{L}$ is a factor code and $\mathcal{L}$ is uniformly recurrent, then S is finite. Indeed, for $w \in S$, there exists n such that w is a factor of every word of $\mathcal{L}_{\geq n}$ therefore $S \subseteq \mathcal{L}_{\leq n}$.

Let us now properly define the central notion of this section: return words.

Definition 3.23. Let $\mathcal{L}$ be a language and $S \subseteq \mathcal{L}$ be a factor code. A *complete return word* to S is a word of $\mathcal{L}$ which has a proper prefix in S, a proper suffix in S and which contains no other occurrence of words of S. We denote $\mathrm{CR}_{\mathcal{L}}(S)$ the set of such words, or $\mathrm{CR}_{\mathcal{L}}(w)$ when $S = \{w\}$.

In this work, when we say "return word" it will however mean "left return word" (unless specified otherwise).

Definition 3.24. Let $\mathcal{L}$ be a language and $S \subseteq \mathcal{L}$ be a factor code. A *(left) return word* to S is a word u for which there exists $w \in S$ such that $uw \in \mathrm{CR}_{\mathcal{L}}(S)$. We denote $\mathrm{R}_{\mathcal{L}}(S)$ the set of such words, or $\mathrm{R}_{\mathcal{L}}(w)$ when $S = \{w\}$.

We could have just as well defined *right return words* by asking that $wu \in \mathrm{CR}_{\mathcal{L}}(S)$. The results presented in this section can easily be adapted to this alternative notion.

Example 3.25. In the Tribonacci language (Example 1.25), the words 0102, 01001, 0101, and 201 are complete return words for $\{01, 2\}$. As we will show later (Corollary 3.32), these are the only ones. The (left) return words are then given by 010, 01 and 2.

The two notions of return words are often indifferently considered in the literature as authors restrict themselves to the case where $S = \{w\}$ most of the time. This is because of the trivial link explained below.

Remark 3.26. Let $w \in \mathcal{L} \setminus \{\varepsilon\}$. There is bijection between $\mathrm{R}_{\mathcal{L}}(w)$ and $\mathrm{CR}_{\mathcal{L}}(w)$ given by $u \mapsto uw$. When S is a factor code containing two or more words, we can similarly construct a map between the complete return words and the (left) return words to S but it is not necessarily injective, as shown in Example 3.25.

We now make the following simple observation.

Remark 3.27. Let $w \in \mathcal{L} \setminus \{\varepsilon\}$. The set $\mathrm{R}_{\mathcal{L}}(w)$ is a suffix code. Indeed, assume by contradiction that $u, v \in \mathrm{R}_{\mathcal{L}}(w)$ are such that u is a (proper) suffix of v. Then uw is a proper suffix of vw, which implies that vw contains 3 occurrences of w and contradicts the definition of return word. We can similarly show that the set $\mathrm{CR}_{\mathcal{L}}(w)$ is a bifix code.

Return words are also related to recurrence in a language. Indeed, we have the following folklore result.

Proposition 3.28. *Let $\mathcal{L}$ be a recurrent language. The following are equivalent.*

1. *The language $\mathcal{L}$ is uniformly recurrent.*
2. *For all $w \in \mathcal{L} \setminus \{\varepsilon\}$, the set $\mathrm{R}_{\mathcal{L}}(w)$ is finite, or equivalently, $\mathrm{CR}_{\mathcal{L}}(w)$ is finite.*
3. *There exists $N \geq 1$ such that, for all $w \in \mathcal{L}_{\geq N}$, the set $\mathrm{R}_{\mathcal{L}}(w)$ is finite, or equivalently, $\mathrm{CR}_{\mathcal{L}}(w)$ is finite.*

3.3.1 Number of return words

The multiplicity of words does not only give information on the number of factors as seen in Corollary 1.44 but also on the number of return words. This link has been studied by several authors. In [BPS08], the authors gave a link between weak words, factor complexity and return words. In [DP17] and [DP21], the authors study the number of return words in an eventually neutral language of threshold at most 1, and in an eventually dendric language respectively.

The results and proofs presented in this subsection are merely a rewritting or slight generalization of the ideas and theorems of these papers.

The first part of the following lemma can be adapted from [BPS08], and the second one from [DP17]. We only give here an idea of the proof.

Lemma 3.29. *Let $\mathcal{L}$ be a recurrent language and let $S \subseteq \mathcal{L}$ be a factor code. Let $P = \mathrm{Pref}^*(\mathrm{CR}_{\mathcal{L}}(S)) \setminus \mathrm{Pref}^*(S)$, i.e., P is the set of words which are proper prefix of an element of $\mathrm{CR}_{\mathcal{L}}(S)$ and have an element of S as a prefix. Then*

1. *we have*
$$\# \mathrm{CR}_{\mathcal{L}}(S) = \#S + \sum_{w \in P} (\#E^R_{\mathcal{L}}(w) - 1);$$
2. *P is an $\mathcal{L}$-maximal suffix code.*

Proof. Let us prove the first claim. For each $w \in S$, we define a rooted tree with root w and we iteratively build the other vertices as follows: if a vertex v is a complete return word for S, then it is a leaf, otherwise va is a child of v for all $a \in E^R_{\mathcal{L}}(v)$. We then consider the union of these trees for all $w \in S$. Observe that the vertices all have different labels and as $\mathcal{L}$ is recurrent, the internal vertices are exactly the elements of P. Moreover, for each $v \in P$, there are exactly $\#E^R_{\mathcal{L}}(v)$ edges leaving the vertex (i.e., between

v and one of its children). The equality of the first claim then follows from graph theory as $\#\,\mathrm{CR}_{\mathcal{L}}(S)$ is the number of leaves and $\#S$ the number of roots.

We now turn to the second claim. The proof that P is a suffix code is the same as the one used to show that $\mathrm{R}_{\mathcal{L}}(w)$ is a suffix code (Remark 3.27). The $\mathcal{L}$-maximality follows from the fact that $\mathcal{L}$ is recurrent. Indeed, by recurrence, any element of $\mathcal{L}$ is a suffix of a word $v \in \mathcal{L}$ containing (at least) one element of S, and there exists u such that $vuw \in \mathcal{L}$ with $w \in S$. The word v then has a suffix in P by definition of P. □

Remark 3.30. For the previous result, only the recurrence with respect to the elements of S is needed. In other words, we reach the same conclusion for any language $\mathcal{L}$ and any factor code $S \subseteq \mathcal{L}$ such that, for all $u \in \mathcal{L}$, there exist $v, v' \in S$ and $w, w' \in \mathcal{L}$ such that $vwuw'v' \in \mathcal{L}$. In fact, in most of the results from this subsection and from the following one, we can replace recurrence by this weaker hypothesis.

Notice that the sets $\mathrm{CR}_{\mathcal{L}}(S)$ and P need not to be finite in the previous result. Therefore, to exploit the first claim of Lemma 3.29, we will need the infinite version of Proposition 3.4 given in Proposition 3.5. Putting it together with Proposition 3.28, we obtain the following result. The first part and most of the second part were proved in [DP21].

Proposition 3.31. *Let $\mathcal{L}$ be an eventually neutral language of threshold N. If $\mathcal{L}$ is recurrent, then*

1. *$\mathcal{L}$ is uniformly recurrent;*
2. *for every factor code $S \subseteq \mathcal{L}_{\geq N}$, we have*

$$\#\,\mathrm{CR}_{\mathcal{L}}(S) = \#S + s_{\mathcal{L}}(N).$$

Proof. Let us prove the first claim and let $w \in \mathcal{L}_{\geq N}$. By Lemma 3.29, we have

$$\#\,\mathrm{CR}_{\mathcal{L}}(w) = 1 + \sum_{w \in P} (\#E^R_{\mathcal{L}}(w) - 1)$$

where $P = \mathrm{Pref}^*(\mathrm{CR}_{\mathcal{L}}(w)) \setminus \mathrm{Pref}^*(w)$ and P is a suffix code. Moreover, by definition, w is a prefix of all the elements of P so $P \subseteq \mathcal{L}_{\geq N}$. By Proposition 3.5, we deduce that

$$\#\,\mathrm{CR}_{\mathcal{L}}(w) \leq 1 + s_{\mathcal{L}}(N).$$

This is true for all $w \in \mathcal{L}_{\geq N}$ thus $\mathcal{L}$ is uniformly recurrent by Proposition 3.28.

We now turn to the second claim. Let $S \subseteq \mathcal{L}_{\geq N}$ be a factor code. Using Lemma 3.29 and Proposition 3.5 once again, we have

$$\#\,\mathrm{CR}_{\mathcal{L}}(S) = \#S + \sum_{w \in P} (\#E^R_{\mathcal{L}}(S) - 1) \leq \#S + s_{\mathcal{L}}(N)$$

where $P = \mathrm{Pref}^*(\mathrm{CR}_{\mathcal{L}}(S)) \setminus \mathrm{Pref}^*(S)$. Observe that S is finite by Remark 3.22 and the first claim. Consequently, $\mathrm{CR}_{\mathcal{L}}(S)$ is finite, which implies that P is also finite by definition. We then deduce the equality from Proposition 3.5. $\square$

The first claim of the previous result allows us to make no distinction between recurrence and uniform recurrence in this work as we almost exclusively work with eventually neutral (or dendric) languages.

We now turn to the number of (left) return words and deduce the following corollary using Remark 3.26 on the link between return words and complete return words.

Corollary 3.32. *Let $\mathcal{L}$ be a recurrent eventually neutral language of threshold N and let $S \subseteq \mathcal{L}_{\geq N}$ be a factor code. Then*

$$\#\,\mathrm{R}_{\mathcal{L}}(S) \leq \#\,\mathrm{CR}_{\mathcal{L}}(S) = \#S + s_{\mathcal{L}}(N).$$

In particular, if $\mathcal{L}$ is neutral then

$$\#\,\mathrm{R}_{\mathcal{L}}(S) \leq \#\,\mathrm{CR}_{\mathcal{L}}(S) = \#S + \#\mathcal{A} - 1$$

where $\mathcal{A}$ is the alphabet of $\mathcal{L}$, i.e., $\mathcal{A} = \mathcal{L}_1$.

If $S = \{w\}$, the inequalities above become equalities.

When $\mathcal{L}$ is dendric, not only are there as many return words for a word as letters but the return words generate the free group over the alphabet. This result is often called Return Theorem and was proved in [BDFD$^+$15a]. For more information on free groups, the reader can consult [LS01].

Theorem 3.33 (Return Theorem)**.** *Let $\mathcal{L}$ be a recurrent dendric language over $\mathcal{A}$. For all non-empty $w \in \mathcal{L}$, the set $\mathrm{R}_{\mathcal{L}}(w)$ is a basis of the free group over $\mathcal{A}$.*

In the case of a return morphism for a set, we can deduce the following result.

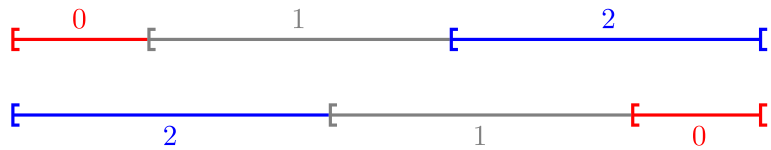


Figure 3.1: Representation of an RIET for the orders $\binom{0<1<2}{2\prec 1\prec 0}$ whose language $\mathcal{L}$ is such that $\#\,\mathrm{R}_{\mathcal{L}}(\{0,2\}) = 4$.

Corollary 3.34. *Let $\mathcal{L}$ be a recurrent dendric language over $\mathcal{A}$ and let $S \subseteq \mathcal{L}$ be a factor code.*

1. *The set $\mathrm{R}_{\mathcal{L}}(S)$ generates the free group over $\mathcal{A}$.*
2. *The set $\mathrm{R}_{\mathcal{L}}(S)$ is free if and only if $\#\,\mathrm{R}_{\mathcal{L}}(S) = \#\mathcal{A}$. In that case, it is then a basis of the free group over $\mathcal{A}$.*

Proof. Let $w \in S$. Any return word for w is a concatenation of return words for S. Therefore, any element of the free group over $\mathcal{A}$ generated by $\mathrm{R}_{\mathcal{L}}(w)$ is generated by $\mathrm{R}_{\mathcal{L}}(S)$. By Theorem 3.33, we conclude that $\mathrm{R}_{\mathcal{L}}(S)$ generates the free group over $\mathcal{A}$.

Since a set is free if and only if it is a basis of the free subgroup it generates, and a set generating the free group over $\mathcal{A}$ is a basis if and only if it contains $\#\mathcal{A}$ elements, the second item follows from the first one. □

In particular, it is impossible to have strictly fewer return words than the number of letters but it is entirely possible to have more return words.

Example 3.35. Let $\mathcal{L}$ be the language of the RIET for the orders $\binom{0<1<2}{2\prec 1\prec 0}$ represented in Figure 3.1. The complete return words for $\{0,2\}$ are 02, 20, 212 and 2112. We then see that $\mathrm{R}_{\mathcal{L}}(\{0,2\}) = \{0, 2, 21, 211\}$.

3.3.2 Stability under derivation

In [BDFD+15d], the authors prove that the family of recurrent dendric languages is stable under derivation. This result is fundamental to obtain particular S-adic representations of recurrent dendric languages, which in turn can be used to obtain a characterization of recurrent dendric languages as we will see in Chapter 5.

With the same techniques as first used by the authors of [BDFD+15d], we prove here a more general result stating that the family of recurrent eventually dendric languages is stable under derivation. This result was first stated in [GL22].

Let us first define the notion of derived language.

Definition 3.36. Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a uniformly recurrent language and let $w \in \mathcal{L} \setminus \{\varepsilon\}$. A *derived language* of $\mathcal{L}$ with respect to w is a language

$$D_f(\mathcal{L}) = \{u \in \mathcal{B}^* : f(u)w \in \mathcal{L}\}$$

for a morphism $f\colon \mathcal{B}^* \to \mathcal{A}^*$ where $f_{|\mathcal{B}}$ is a bijection between $\mathcal{B}$ and $\mathrm{R}_{\mathcal{L}}(w)$.

While a derived language not only depends on $\mathcal{L}$ and w but also on f, we will often talk about *the* derived language of $\mathcal{L}$ with respect to w and denote it $D_w(\mathcal{L})$. Indeed, modifying the morphism f does not impact the structure of the language and only renames the letters (by applying a bijective letter-to-letter morphism).

Example 3.37. Let us consider the Tribonacci language $\mathcal{L}$ (Example 1.25) and derive it with respect to 010. We can quickly check that 01, 010 and 0102 are return words for 010, and these are the only ones by Corollary 3.32. We consider the morphism f such that $f(0) = 0102$, $f(1) = 010$ and $f(2) = 01$. For example, since 0102010010 is an element of $\mathcal{L}$, we have $01 \in D_{010}(\mathcal{L})$. On the other hand, $0101010 \notin \mathcal{L}$ so $22 \notin D_{010}(\mathcal{L})$. We similarly see that $(D_{010}(\mathcal{L}))_2 = \{00, 01, 02, 10, 20\} = \mathcal{L}_2$. In fact, we can show that $D_{010}(\mathcal{L}) = \mathcal{L}$ in this particular case since f is a power of the morphism generating $\mathcal{L}$.

As we assume $\mathcal{L}$ uniformly recurrent, the derived language is indeed a language in the sense that it is factorial and biextendable. Moreover, it is also uniformly recurrent. Let us now look at the dendricity of this language.

Remark 3.38. Using the notations of the previous definition, for all $u \in \mathcal{B}^*$ and $a, b \in \mathcal{B}$, we have

$$\begin{aligned} aub \in D_w(\mathcal{L}) &\iff f(a)f(u)f(b)w \in \mathcal{L} \\ &\iff f(a) \cdot f(u)w \cdot g(b) \in \mathcal{L} \end{aligned}$$

where $g\colon \mathcal{B}^* \to \mathcal{A}^*$ is such that $wg(c) = f(c)w$ for all letters c. This is well defined as $f(c)$ is a (left) return word for w by definition. Observe that $g(c)$ is then a right return word for w, i.e., $wg(c) \in \mathrm{CR}_{\mathcal{L}}(w)$.

Using the previous remark, to understand the bi-extensions in the derived language, we need to look at extensions in the original language but these extensions might be of length greater than 1. This motivates the following definition.

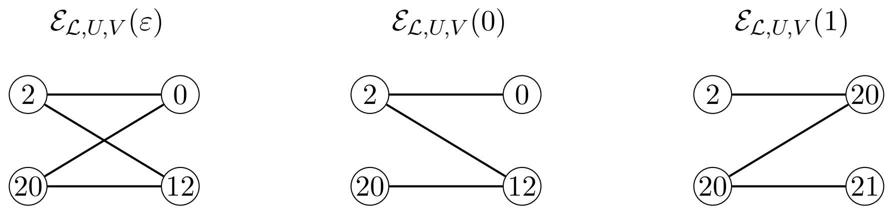


Figure 3.2: The generalized extension graphs of ε (on the left), 0 (in the center) and 1 (on the right) in the Chacon language $\mathcal{L}$, for $U = \{2, 20\}$ and $V = \{0, 12, 20, 20\}$.

Definition 3.39. Let $\mathcal{L}$ be a language, let U be a suffix code and V be a prefix code. We denote

$$E_{\mathcal{L},U,V}(w) = \{(u, v) \in U \times V : uwv \in \mathcal{L}\}.$$

We then define the *generalized extension graph* of w (with respect to U and V) as the bipartite graph $\mathcal{E}_{\mathcal{L},U,V}(w)$ whose edges are given by $E_{\mathcal{L},U,V}(w)$.

When $U = \mathcal{A} = V$, we recover the classical extension graphs.

Example 3.40. Let $\mathcal{L}$ be the Chacon language (Definition 1.12). We consider the suffix code $U = \{2, 20\}$ and the prefix code $V = \{0, 12, 20, 21\}$. Observe that, by Example 1.13, V is $\mathcal{L}$-maximal but U is not. The corresponding generalized extension graphs for ε, 0, 1 can be obtained using Example 1.13 and are represented in Figure 3.2. Observe that since neither 22 nor 202 are in $\mathcal{L}$, the word 2 cannot be extended on the left by elements of U therefore its generalized extension graph is empty.

While these graphs are more general than the usual extension graphs, they remain trees when the language is (eventually) dendric and U and V are maximal.

Proposition 3.41 (Dolce–Perrin [DP21])**.** *Let $\mathcal{L}$ be an eventually dendric language of threshold N, let U be an $\mathcal{L}$-maximal suffix code and V be an $\mathcal{L}$-maximal prefix code. Then for all $w \in \mathcal{L}_{\geq N}$, the graph $\mathcal{E}_{\mathcal{L},U,V}(w)$ is a tree.*

Coming back to Remark 3.38, we can now use the previous result to show the stability of the family of eventually dendric languages under derivation.

Theorem 3.42. *Let $\mathcal{L}$ be a recurrent eventually dendric language of threshold N and let $w \in \mathcal{L}\setminus\{\varepsilon\}$. The derived language $D_w(\mathcal{L})$ is eventually dendric of threshold at most* $\max\{0, N-|w|\}$.

Proof. Let us denote $\mathrm{R}'_{\mathcal{L}}(w)$ the set of right return words for w, i.e.,

$$\mathrm{R}'_{\mathcal{L}}(w) = \{u : wu \in \mathrm{CR}_{\mathcal{L}}(w)\}.$$

A similar proof as the one in Remark 3.27 shows that $\mathrm{R}'_{\mathcal{L}}(w)$ is a prefix code.

Using the notations of Remark 3.38, since $f_{|\mathcal{B}}$ and $g_{|\mathcal{B}}$ are bijections between $\mathcal{B}$ and $\mathrm{R}_{\mathcal{L}}(w)$ (resp., $\mathrm{R}'_{\mathcal{L}}(w)$), we have that, for all $u \in D_w(\mathcal{L})$, the extension graphs $\mathcal{E}_{D_w(\mathcal{L})}(u)$ and $\mathcal{E}_{\mathcal{L},\mathrm{R}_{\mathcal{L}}(w),\mathrm{R}'_{\mathcal{L}}(w)}(f(u)w)$ are isomorphic.

However, to apply Proposition 3.41, we need $\mathcal{L}$-maximal prefix and suffix codes. This can be solved as follows. Let us denote

$$n = \max\{|u| : u \in \mathrm{R}_{\mathcal{L}}(w)\} = \max\{|u| : u \in \mathrm{R}'_{\mathcal{L}}(w)\}.$$

We consider

$$U = \mathrm{R}_{\mathcal{L}}(w) \cup \{u \in \mathcal{L}_n : \mathrm{Suff}(u) \cap \mathrm{R}_{\mathcal{L}}(w) = \emptyset\}.$$

By construction and by Remark 3.27, it is an $\mathcal{L}$-maximal suffix code. Moreover, for each word u, if $uw \in \mathcal{L}$, then u is suffix comparable with an element of $\mathrm{R}_{\mathcal{L}}(w)$ as $\mathcal{L}$ is recurrent. This implies that such a word u cannot be in $U \setminus \mathrm{R}_{\mathcal{L}}(w)$. We similarly define

$$V = \mathrm{R}'_{\mathcal{L}}(w) \cup \{u \in \mathcal{L}_n : \mathrm{Pref}(u) \cap \mathrm{R}'_{\mathcal{L}}(w) = \emptyset\}$$

which is an $\mathcal{L}$-maximal prefix code such that, if $wv \in \mathcal{L}$, then $v \notin V \setminus \mathrm{R}'_{\mathcal{L}}(w)$. In the end, we have

$$\mathcal{E}_{\mathcal{L},U,V}(x) = \mathcal{E}_{\mathcal{L},\mathrm{R}_{\mathcal{L}}(w),\mathrm{R}'_{\mathcal{L}}(w)}(x)$$

for all $x \in w\mathcal{A}^* \cap \mathcal{A}^* w$. Observe that, for all $u \in D_w(\mathcal{L})$, $f(u)w \in w\mathcal{A}^* \cap \mathcal{A}^* w$ so $\mathcal{E}_{\mathcal{L},U,V}(f(u)w)$ is isomorphic to $\mathcal{E}_{D_w(\mathcal{L})}(u)$. Therefore, using Proposition 3.41, the extension graph $\mathcal{E}_{D_w(\mathcal{L})}(u)$ is a tree if $|f(u)w| \geq N$. As f is non-erasing, this shows in particular that any $u \in D_w(\mathcal{L})$ such that $|u| \geq N - |w|$ is dendric. □

Remark 3.43. In the previous proof, we show something slightly stronger: every word $u \in D_w(\mathcal{L})$ such that $|f(u)w| \geq N$ is dendric in $D_w(\mathcal{L})$. This implies that the bound given in the theorem is not optimal. In fact, if we denote $M = \min\{|v| : v \in R_{\mathcal{L}}(w)\}$, then we can directly deduce that $D_w(\mathcal{L})$ is in fact eventually dendric of threshold at most $\max\{0, \frac{N-|w|}{M}\}$.

We easily recover in particular the original result about dendric languages.

Corollary 3.44 (Berthé *et al.* [BDFD$^+$15d]). *Let $\mathcal{L}$ be a recurrent dendric language and let $w \in \mathcal{L} \setminus \{\varepsilon\}$. The derived language $D_w(\mathcal{L})$ is dendric.*

This also shows that any recurrent eventually dendric language has a derived language which is dendric since it suffices to derive with respect to a long enough word.

We can also deduce some additional information on derived languages using the number of return words.

Corollary 3.45. *Let $\mathcal{L}$ be a recurrent eventually dendric language of threshold N. If $w \in \mathcal{L}\setminus\{\varepsilon\}$ is of length at least N, then the derived language $D_w(\mathcal{L})$ is over an alphabet of size $s_{\mathcal{L}}(N) + 1$.*

More generally, for any $w \in \mathcal{L} \setminus \{\varepsilon\}$, we have $s_{D_w(\mathcal{L})}(n) = s_{\mathcal{L}}(N)$ for any large enough n.

Proof. By Corollary 3.32, if $|w| \geq N$, then $\# \mathrm{R}_{\mathcal{L}}(w) = s_{\mathcal{L}}(N) + 1$, which directly implies that $D_w(\mathcal{L})$ is over an alphabet of size $s_{\mathcal{L}}(N) + 1$.

Assume now that we take any $w \in \mathcal{L} \setminus \{\varepsilon\}$. Using the first part of the claim, there exist infinitely many words for which the corresponding derived language of $\mathcal{L}$ is on a size-k alphabet if and only if $k = s_{\mathcal{L}}(N) + 1$. Since $D_w(\mathcal{L})$ is eventually dendric of threshold M by Theorem 3.42, there exist infinitely many words for which the corresponding derived language of $D_w(\mathcal{L})$ is on an alphabet of size $s_{D_w(\mathcal{L})}(M)+1$. As any derived language of $D_w(\mathcal{L})$ is a derived language of $\mathcal{L}$, we conclude that $s_{D_w(\mathcal{L})}(n) = s_{D_w(\mathcal{L})}(M) = s_{\mathcal{L}}(N)$ for all $n \geq M$. □

3.4 Open questions

In Section 3.2, and in most of this work, we restrict ourselves to the case of non-erasing morphisms. A natural first question is therefore the following one.

Question 3.1. Let σ be a (potentially erasing) morphism, what are the restrictions on $p_{\sigma(\mathcal{L})}(n)$ if $\mathcal{L}$ is (eventually) neutral?

The other questions presented here are centered around Section 3.3 and return words. The Return Theorem (Theorem 3.33) is one of the most important result for the study of dendric languages, and could gain even more importance depending on the answer to the following open question.

Question 3.2. Is the Return Theorem a characterization of recurrent dendric languages? In other words, if, for all $w \in \mathcal{L} \setminus \{\varepsilon\}$, the set $\mathrm{R}_{\mathcal{L}}(w)$ is a basis of the free group over $\mathcal{L}_1$, does it imply that $\mathcal{L}$ is dendric?

The proof of the Return Theorem can in fact be decomposed into two steps. Indeed the authors first proved in [BDFD$^+$15a] that, if $\mathcal{L}$ is a recurrent connected language over $\mathcal{A}$, then $\mathrm{R}_{\mathcal{L}}(w)$ generates the free group $F_{\mathcal{A}}$ over $\mathcal{A}$ for all $w \in \mathcal{L} \setminus \{\varepsilon\}$. In this case, $\mathrm{R}_{\mathcal{L}}(w)$ is a basis of $F_{\mathcal{A}}$ if and only if $\#\mathcal{R}_{\mathcal{L}}(w) = \#\mathcal{A}$. Since a connected language has no weak element by Proposition 2.10, this equality is true for all $w \in \mathcal{L} \setminus \{\varepsilon\}$ if and only if $p_{\mathcal{L}}(n) = (\#\mathcal{A} - 1)n + 1$ by [BPS08, Theorem 4.5]. Using Corollary 1.43 and Proposition 2.10, this occurs exactly when $\mathcal{L}$ is dendric.

In other words, among the recurrent connected languages, the only ones such that $\mathrm{R}_{\mathcal{L}}(w)$ is a basis of the free group over the alphabet for all $w \in \mathcal{L} \setminus \{\varepsilon\}$ are the dendric languages.

In [GO22], Goulet-Ouellet introduced the notion of *suffix-connected languages* and showed that, if $\mathcal{L}$ is a uniformly recurrent suffix-connected language over $\mathcal{A}$ such that ε is connected, then $\mathrm{R}_{\mathcal{L}}(w)$ generates $F_{\mathcal{A}}$ for all $w \in \mathcal{L} \setminus \{\varepsilon\}$. Therefore, a negative anwser to Question 3.2 would follow from a positive answer to the question below.

Question 3.3. Does there exist a uniformly recurrent suffix-connected language $\mathcal{L}$ over $\mathcal{A}$ such that ε is connected and, for all $w \in \mathcal{L} \setminus \{\varepsilon\}$, $\#\,\mathrm{R}_{\mathcal{L}}(w) = \#\mathcal{A}$ but $\mathcal{L}$ is not dendric?

Clearly, sets of return words in high factor complexity languages are likely to contain more words and therefore generate the free group. However, they are less likely to be free. We will therefore focus on low-complexity languages, and it is then natural to ask if suffix-connectedness captures all such languages for which the sets of return words generate the free group on the alphabet. More generally, we have the following question.

Question 3.4. What are the uniformly recurrent languages $\mathcal{L}$ over $\mathcal{A}$ with factor complexity in $O(n)$ such that, for all $w \in \mathcal{L} \setminus \{\varepsilon\}$, the set $\mathrm{R}_{\mathcal{L}}(w)$ generates $F_{\mathcal{A}}$?

We end with some questions on the stability of (eventual) dendricity under derivation with respect to a set of words. Indeed, we saw in Theorem 3.42 and Corollary 3.44 that these properties were stable when deriving with respect to a word. There are two approaches to define the derived language of $\mathcal{L}$ with respect to a set S of words.

The first method is similar to the derivation with respect to a word: we fix a morphism $f\colon \mathcal{B}^* \to \mathcal{A}^*$ such that $f_{|\mathcal{B}}$ is a bijection between $\mathcal{B}$ and $\mathrm{R}_{\mathcal{L}}(S)$ and consider the language

$$D_S(\mathcal{L}) = \{u \in \mathcal{B}^* : \exists\, w \in S \text{ st. } f(u)w \in \mathcal{L}\}.$$

Unfortunately, one can find non-dendric languages whose image under an elementary morphism (see Definition 4.43) is dendric. Since these morphisms define a bijection between the initial alphabet and the return words for a set of letters in the image, dendricity is not stable under derivation with respect to a set of words for this definition. Still, the preservation of eventual dendricity is open.

Question 3.5. If $\mathcal{L}$ is a recurrent eventually dendric language and $S \subseteq \mathcal{L}$ is a factor code, is $D_S(\mathcal{L})$ eventually dendric?

The second approach was chosen in [Dur13b] and consists in considering complete return words instead. More precisely, we define

$$\mathcal{B}_S = \{(u, w) \in \mathrm{R}_{\mathcal{L}}(S) \times S : uw \in \mathrm{CR}_{\mathcal{L}}(S)\},$$

fix a bijective map $f\colon \mathcal{B} \to \mathcal{B}_S$ and consider the language

$$\begin{aligned} D'_S(\mathcal{L}) = \{u \in \mathcal{B}^* : f(u_i) = (v_i, w_i), v_1 v_2 \cdots v_n w_n \in \mathcal{L}, \\ w_i \in \mathrm{Pref}(v_{i+1} \cdots v_n w_n)\, \forall\, i < n\}. \end{aligned}$$

Observe that, if S contains a unique element, these two approaches are equivalent. However, in the case of a general factor code S, we ask the following question.

Question 3.6. If $\mathcal{L}$ is a recurrent (eventually) dendric language and $S \subseteq \mathcal{L}$ is a factor code, is $D'_S(\mathcal{L})$ (eventually) dendric?

Chapter 4

Preserving dendricity

This chapter is dedicated to the preservation of (eventual) dendricity when applying a morphism. It revolves around the following open question:

Given a dendric language $\mathcal{L}$ and a morphism σ can we determine whether $\sigma(\mathcal{L})$ is dendric or not?

More generally, given a family $\mathcal{F}$ of languages and a language $\mathcal{L}$ in this family, it is natural to wonder which morphisms satisfy $\sigma(\mathcal{L}) \in \mathcal{F}$. The answer to this question is well known for the family of Sturmian languages [Lot02], and more generally for Arnoux-Rauzy sequences [JP02]. Indeed, the morphisms who preserve these families for one language or, equivalently in this case, for all languages over a given alphabet are the elements of a finitely generated monoid. The precise statement of this result will be recalled in Subsection 4.2.1.

On the other hand, there are families preserved under any (non-erasing) morphism. Indeed, by Proposition 3.7 on the evolution of the factor complexity when applying a non-erasing morphism, sub-linear complexity is preserved, i.e. if $\mathcal{L}$ is such that $p_{\mathcal{L}}(n) \in O(n)$, then $p_{\sigma(\mathcal{L})} \in O(n)$ as soon as σ is non-erasing. It is also well known that if the image of a morphic sequence is infinite, then it is also a morphic sequence [AS03].

In general, preservation of a property for one language is not equivalent to preservation for all languages. We also mention some results on the second type of preserving morphism for lesser-known families which are related to the ones studied in this work. Partial answers have been given for example for interval exchanges associated with the orders $\binom{1<2<3}{3\prec 2\prec 1}$ [AP07] or for rich sequences, i.e. sequences whose length-n prefix contains $n+1$ palindroms for all n [GJWZ09].

In this chapter, we once again restrict ourselves to non-erasing morphisms. Using results of Chapter 3, we start by studying in Section 4.1 the possible relations between the sizes of the alphabets $\mathcal{A}$ and $\mathcal{B}$ if $\mathcal{L}$ is a neutral language over $\mathcal{A}$ and $\sigma(\mathcal{L})$ is a neutral language over $\mathcal{B}$. This gives us some first easy conditions to know that a morphism will never preserve dendricity.

In Section 4.2, we turn to the question of the morphisms preserving dendricity for all languages. We show that these morphisms are exactly the elements of the monoid generated by Arnoux-Rauzy morphisms and permutations. In particular, this shows that preserving dendricity is equivalent to preserving Arnoux-Rauzy languages. The intermediary steps to reach this result are in fact quite similar to the study of Sturmian morphisms done in [Lot02, Section 2.3.1].

When trying to determine if a morphism preserves dendricity for a specific dendric language, we essentially need to understand the extension graphs in the image, based on the morphism and the initial language. Understanding the extensions in the image can however become particularly tricky if there is no clear description of the possible coverings of a word. To avoid this situation, we restrict ourselves to morphisms having some form of recognizability property.

Defining some well-behaved morphisms is precisely the object of Section 4.3 in which we introduce the so-called *return morphisms*. These morphisms are related to return words and derivation and are therefore the building blocks of the S-adic representations considered in Chapter 5. This is in fact the main motivation behind the study of the link between dendricity and return morphisms started in [GLL22].

We then show in Section 4.4 that, for return morphisms, we can precisely describe the extensions of words in the image and therefore characterize when the image of a dendric language is dendric.

We then turn to the question of preserving eventual dendricity for which only partial answers are known. We present these results in Section 4.5.

4.1 Restrictions on the alphabets

One simple technique to right-away be able to tell that the image under some morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ of a neutral language $\mathcal{L}$ over $\mathcal{A}$ is not neutral is to look at the alphabet sizes. Indeed, we show in this section that, for many pairs $(\#\mathcal{A}, \#\mathcal{B})$, no morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ will preserve neutrality, even for one language.

In neutral languages, the size of the alphabet is closely related to the factor complexity. It is therefore quite natural that the results of this section are mostly obtained as consequences of the study of coverings and factor complexity done in Section 3.2.

The first observation we make follows from Proposition 3.15.

Proposition 4.1. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism with $\#\mathcal{B} \geq \#\mathcal{A}$ and let $\mathcal{L}$ be a neutral language over $\mathcal{A}$. If $\sigma(\mathcal{L})$ is a strong or neutral language over $\mathcal{B}$, then $\#\mathcal{B} = \#\mathcal{A}$ and $\sigma(\mathcal{L})$ is neutral. In particular, if $\sigma(\mathcal{L})$ is connected, then $\sigma(\mathcal{L})$ is dendric.*

Proof. By Corollary 1.43 on the link between neutrality and the first difference of complexity, we have $s_{\mathcal{L}}(n) = \#\mathcal{A} - 1$ for all $n \geq 0$, and the sequence $(s_{\sigma(\mathcal{L})}(n))_{n\geq 0}$ is non-decreasing and such that $s_{\sigma(\mathcal{L})}(0) = \#\mathcal{B} - 1$. However, by Proposition 3.15, there exists C such that, for all $n \geq 0$,

$$p_{\sigma(\mathcal{L})}(n) \leq C + p_{\mathcal{L}}(n) = C + (\#\mathcal{A} - 1)n + 1 \leq C + (\#\mathcal{B} - 1)n + 1$$

since $\#\mathcal{B} \geq \#\mathcal{A}$. Using the observation above, the only possibility is to have $s_{\sigma(\mathcal{L})}(n) = \#\mathcal{B} - 1$ for all $n \geq 0$ and $\#\mathcal{A} = \#\mathcal{B}$. This then implies that $\sigma(\mathcal{L})$ is neutral by Corollary 1.43.

If $\sigma(\mathcal{L})$ is connected, then it is strong or neutral, which in turn implies that it is neutral as shown above. We conclude that it is dendric by Proposition 2.10. □

Corollary 4.2. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. If there exists a neutral language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a neutral language over $\mathcal{B}$, then $\#\mathcal{B} \leq \#\mathcal{A}$.*

We cannot however give a lower bound on $\#\mathcal{B}$. Indeed, if σ maps all the letters of $\mathcal{A}$ to powers of some letter b, then $\sigma(\mathcal{L}) = \{b\}^*$ for any language $\mathcal{L}$. The language $\{b\}^*$ is neutral as it does not contain any bispecial word. This shows that, if $\sigma\colon \mathcal{A}^* \to \{b\}^*$ is non-erasing, for all (neutral) language $\mathcal{L}$ over $\mathcal{A}$, $\sigma(\mathcal{L})$ is a neutral language over $\{b\}$.

In fact, we have the following stronger result showing that, for all choices of the alphabets $\mathcal{A}$ and $\mathcal{B}$ with $\#\mathcal{B} \leq \#\mathcal{A}$, we can find a morphism preserving neutrality (and even dendricity) for infinitely many languages. This shows that, without additional restriction on the morphism, the inequality of Corollary 4.2 is optimal.

Proposition 4.3. *Let $\mathcal{A}$, $\mathcal{B}$ be two alphabets such that $\#\mathcal{A} \geq 2$ and $\#\mathcal{B} \leq \#\mathcal{A}$. There exists a morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ such that, for infinitely many dendric languages $\mathcal{L}$ over $\mathcal{A}$, $\sigma(\mathcal{L})$ is a dendric language over $\mathcal{B}$.*

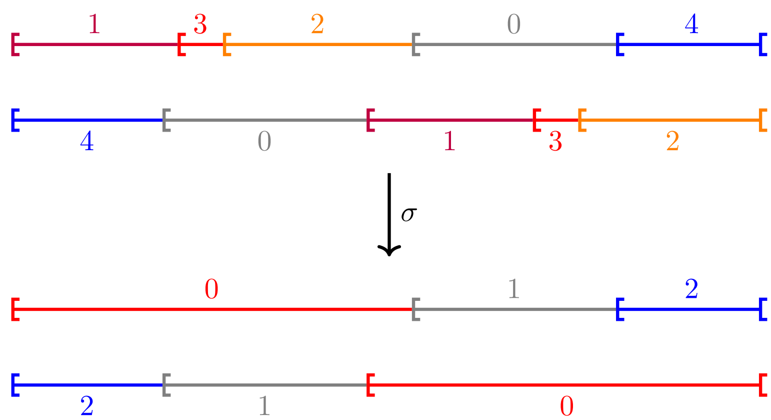


Figure 4.1: Example of the construction of Proposition 4.3 for $\sigma\colon \{0,1,2,3,4\}^* \to \{0,1,2\}^*$ such that $\sigma(0) = 1$, $\sigma(1) = \sigma(2) = \sigma(3) = 0$, $\sigma(4) = 2$.

Proof. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a letter-to-letter morphism such that there exists $b \in \mathcal{B}$ for which $\#\sigma^{-1}(b) = \#\mathcal{A} - \#\mathcal{B} + 1$, and for all $c \neq b$, $\#\sigma^{-1}(c) = 1$. Let T be a regular interval exchange transformation over $\mathcal{A}$ associated with a pair of orders $\binom{\leq}{\preceq}$ such that the elements of $\sigma^{-1}(b)$ are the smallest (resp., largest) for $\leq$ (resp., $\preceq$) and are in the same order for both $\leq$ and $\preceq$. Observe that any pair of such orders is irreducible so any set of rationally independent lengths will give an RIET. In other words, there are infinitely many choices for T since $\#\mathcal{A} \geq 2$.

Let $\mathcal{L}$ be the language of T. Then $\sigma(\mathcal{L})$ is the language of the interval exchange transformation T' over $\mathcal{B}$ where the intervals corresponding to elements of $\sigma^{-1}(b)$ have been merged and the intervals are renamed according to σ. We illustrate this construction on an example in Figure 4.1. This is still an RIET, in fact, the dynamical systems corresponding to T and T' are equal. This shows that both $\mathcal{L}$ and $\sigma(\mathcal{L})$ are dendric, and they are over the alphabets $\mathcal{A}$ and $\mathcal{B}$ respectively. $\square$

More precisely, the previous result shows that for any alphabets $\mathcal{A}$ and $\mathcal{B}$ such that $\#\mathcal{B} \leq \#\mathcal{A}$, we can find infinitely many languages of RIET over $\mathcal{A}$ having as image the language of an RIET over $\mathcal{B}$. On the other hand, Justin and Pirillo showed that, if $\mathcal{L}$ and $\sigma(\mathcal{L})$ are recurrent Arnoux-Rauzy languages over $\mathcal{A}$ and $\mathcal{B}$ respectively and if $\#\mathcal{B} \geq 2$, then $\#\mathcal{A} = \#\mathcal{B}$ [JP02]. In particular, a recurrent Arnoux-Rauzy over an alphabet of size at least 3 will not have any dendric image over a binary alphabet. In other words, the sizes $\#\mathcal{B} \leq \#\mathcal{A}$ for which there exists a dendric image $\sigma(\mathcal{L})$ over $\mathcal{B}$ depend on the dendric language $\mathcal{L}$ over $\mathcal{A}$.

As shown above, without additional restriction on the morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, the only information we have on $\#\mathcal{A}$ and $\#\mathcal{B}$ if σ preserves neutrality for a language is that $\#\mathcal{B} \leq \#\mathcal{A}$. We show below that, if σ is injective, we have more restrictions on $\mathcal{A}$ and $\mathcal{B}$. Observe that an injective morphism is, in particular, non-erasing and the image alphabet cannot be of size 1.

The study of injective morphisms allows to avoid the construction of Proposition 4.3 and is inspired by the following result from [CN10]. While the second part of the claim was not explicitly stated, it can easily be deduced from the authors' proof.

Proposition 4.4. *Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be an injective morphism. For all $n \geq 0$, we have*

$$p_{\mathcal{L}}(n) \leq \|\sigma\| \cdot p_{\sigma(\mathcal{L})}(\|\sigma\| n).$$

Moreover, if σ is $\|\sigma\|$-uniform, i.e. $|\sigma(a)| = \|\sigma\|$ for all $a \in \mathcal{A}$, then

$$p_{\mathcal{L}}(n) \leq p_{\sigma(\mathcal{L})}(\|\sigma\| n).$$

In the case of a morphism which preserves neutrality for at least one language, we obtain the following corollary.

Corollary 4.5. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be an injective morphism. If there exists a neutral language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a neutral language over $\mathcal{B}$, then*

$$\frac{\#\mathcal{A} - 1}{\#\mathcal{B} - 1} \leq \|\sigma\|^2.$$

Moreover, if σ is $\|\sigma\|$-uniform, then

$$\frac{\#\mathcal{A} - 1}{\#\mathcal{B} - 1} \leq \|\sigma\|.$$

Proof. Assume that we have such a language $\mathcal{L}$. Then for all $n \geq 0$

$$(\#\mathcal{A} - 1)n + 1 \leq \|\sigma\| \cdot ((\#\mathcal{B} - 1)\|\sigma\| n + 1).$$

As it is true for all n, we must have

$$\#\mathcal{A} - 1 \leq \|\sigma\|^2 (\#\mathcal{B} - 1)$$

which ends the proof of the general case. If σ is uniform, then we similarly conclude. □

We will now show that this bound is reached for infinitely many examples in the uniform case.

Proposition 4.6. *For all $k \in \mathbb{N} \setminus \{0\}$ and for all alphabets $\mathcal{A}$, $\mathcal{B}$ such that*

$$\frac{\#\mathcal{A} - 1}{\#\mathcal{B} - 1} = k$$

there exist an injective k-uniform morphism $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ and infinitely many dendric languages $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a dendric language over $\mathcal{B}$.

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be two such alphabets. Let $\mathcal{L}'$ be a recurrent dendric language over $\mathcal{B}$. By Corollary 1.44, we have $p_{\mathcal{L}'}(k) = (\#\mathcal{B} - 1)k + 1 = \#\mathcal{A}$. Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be such that $\sigma_{|\mathcal{A}}$ is a bijection between $\mathcal{A}$ and $\mathcal{L}'_k$. By definition, σ is k-uniform. Let also

$$\mathcal{L} = \{w \in \mathcal{A}^* : \sigma(w) \in \mathcal{L}'\}.$$

In other words, $\mathcal{L}$ is a decoding of $\mathcal{L}'$ with respect to the $\mathcal{L}'$-maximal bifix code $\mathcal{L}'_k$. Since the family of recurrent dendric languages is closed under maximal bifix decoding [BDFD+15d], $\mathcal{L}$ is dendric and such that $\sigma(\mathcal{L}) = \mathcal{L}'$ is dendric.

This construction can be done starting from any recurrent dendric language $\mathcal{L}'$ over $\mathcal{B}$ (and there are infinitely many of such languages). On the other hand, there are only finitely many possibilities for the morphism σ. The conclusion follows by the pigeonhole principle. □

The previous construction shows that we can build infinitely many examples when $\frac{\#\mathcal{A}-1}{\#\mathcal{B}-1}$ is an integer. We show below that it is however impossible to find examples when $\frac{\#\mathcal{A}-1}{\#\mathcal{B}-1} \in (1, 2)$ and σ is injective.

Proposition 4.7. *Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. If there exists a neutral language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a neutral language over $\mathcal{B}$ and σ is injective over $\mathcal{L}_{\geq N}$ for some $N \geq 0$, then either $\#\mathcal{B} = \#\mathcal{A}$ or $2(\#\mathcal{B} - 1) \leq \#\mathcal{A} - 1$.*

Proof. Let us show that, if $\#\mathcal{B} \neq \#\mathcal{A}$, then all the long enough words in $\sigma(\mathcal{L})$ have at least two coverings. By contradiction, assume that $\#\mathcal{B} \neq \#\mathcal{A}$ and $u \in (\sigma(\mathcal{L}))_{\geq \|\sigma\| N}$ has a unique covering (w, k). Observe that, by Corollary 4.2, since $\mathcal{L}$ and $\sigma(\mathcal{L})$ are neutral, we have in fact $\#\mathcal{B} < \#\mathcal{A}$.

Let us look at the return words for u in $\sigma(\mathcal{L})$. Since the only way of seeing u in $\sigma(\mathcal{L})$ is in the image of w after k letters, we have

$$R_{\sigma(\mathcal{L})}(u) = \{\sigma(v)_{[k+1, |\sigma(v)|]}\sigma(w_1)_{[1,k]} : v \in R_{\mathcal{L}}(w)\}.$$

In particular, $\#R_\sigma(\mathcal{L})(u) \leq \#R_\mathcal{L}(w)$. Since $\mathcal{L}$ and $\sigma(\mathcal{L})$ are neutral, by Corollary 3.32, we have $\#R_{\sigma(\mathcal{L})}(u) = \#\mathcal{B} < \#\mathcal{A} = \#R_\mathcal{L}(w)$. Therefore, there exists $v, v' \in R_\mathcal{L}(w)$ such that

$$\sigma(v)_{[k+1,|\sigma(v)|]}\sigma(w_1)_{[1,k]} = \sigma(v')_{[k+1,|\sigma(v')|]}\sigma(w_1)_{[1,k]},$$

or equivalently, $\sigma(v) = \sigma(v')$ since v and v' begin with w_1. As σ is non-erasing, v and v' cannot be prefix comparable. However, as return words v and v', are prefix comparable with w. This shows that w is a prefix of both v and v'. As (w, k) is a covering of a word of length $\|\sigma\|N$, we have $|w| \geq N$. By injectivity of σ on the long words, it is therefore impossible to find such v and v'.

We have therefore shown that, if $\#\mathcal{B} < \#\mathcal{A}$, all the words of $(\sigma(\mathcal{L}))_{\geq \|\sigma\| N}$ have at least two coverings. By Proposition 3.15, there exists $C \in \mathbb{Z}$ such that

$$C + (\#\mathcal{A} - 1)n = c_\mathcal{L}(n) \geq 2p_{\sigma(\mathcal{L})}(n) = 2(\#\mathcal{B} - 1)n + 2$$

for all $n \geq \|\sigma\|N$. The conclusion follows by looking at the linear coefficients. □

Observe that this does not contradict the construction of Proposition 4.3. Indeed, a morphism identifying letters coding consecutive intervals is not eventually injective.

4.2 Dendric preserving morphisms

In Section 4.1, we gave examples of morphisms which preserve dendricity for infinitely many dendric languages (Proposition 4.3 and Proposition 4.6). However these morphisms do not preserve dendricity for *all* dendric languages, which is what we are interested in for this section.

Definition 4.8. A morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ is *dendric preserving* if, for any dendric language $\mathcal{L}$ over $\mathcal{A}$, the language $\sigma(\mathcal{L})$ is dendric.

Warning. Throughout this section, we always implicitly assume that the image alphabet $\mathcal{B}$ is minimal, i.e. for all $b \in \mathcal{B}$, there exists $a \in \mathcal{A}$ such that b is in $\sigma(a)$. In other words, the image of a language over $\mathcal{A}$ is a language over $\mathcal{B}$.

When the initial alphabet or the image alphabet is unary, the description of dendric preserving morphisms becomes trivial as explained in the following remark.

Remark 4.9. Let us consider the dendric preserving morphisms $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$. If $\mathcal{B} = \{b\}$, then any such morphism is dendric preserving as the image of any language is $\{b\}^*$ which is dendric. On the other hand, if $\mathcal{A} = \{a\}$, then σ is dendric preserving if and only if $\sigma(\{a\}^*)$ is dendric. However, the image $\sigma(\{a\}^*)$ is periodic (i.e. of bounded complexity). Therefore, σ is dendric preserving if and only if $\mathcal{B}$ is also unary.

In what follows, we will therefore only consider morphisms where both the initial and the image alphabets are of size at least 2. Moreover, as often in this work, we restrict ourselves to the case of non-erasing morphisms.

Together with J. Leroy, we began the study of these morphisms in [GLL22] where we characterized the dendric preserving morphisms inside a particular sub-family of morphisms. Generalizing some of the ideas, a complete characterization of non-erasing dendric preserving morphisms was then given in [Ghe23]. We present the results leading to this characterization in this section.

We first recall the well-known result for Sturmian and, more generally, for Arnoux-Rauzy languages in Subsection 4.2.1. We also give the definition and first properties of the famous Arnoux-Rauzy morphisms which, as we will show later, generate the dendric preserving morphisms.

In Subsection 4.2.2, we prove the central result of this section which is the complete description of the dendric preserving morphisms.

We then do a more careful analysis of the proofs and obtain in Subsection 4.2.3 as a direct consequence that the only morphisms preserving languages of RIET are trivial (except on a binary alphabet).

4.2.1 Arnoux-Rauzy morphisms

On a two letters alphabet, the dendric languages are exactly the Sturmian languages. In this case, the dendric preserving morphisms are called *Sturmian morphisms* and it is well known that they are exactly the morphisms generated by

$$L_0\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \end{cases} \qquad L_1\colon \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \end{cases} \qquad R_0\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 10 \end{cases} \qquad R_1\colon \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 1 \end{cases} .$$

More precisely, we have the following well-known result (see [Lot02] for example).

Proposition 4.10. *Let $\sigma\colon \{0,1\}^* \to \{0,1\}^*$ be a morphism. The following are equivalent:*

1. *σ is a composition of L_0, L_1, R_0, R_1 and the morphism E such that $E(0) = 1$ and $E(1) = 0$;*

2. *there exists a recurrent Sturmian*[1] *language $\mathcal{L}$ over $\{0, 1\}$ such that $\sigma(\mathcal{L})$ is Sturmian;*

3. *for all Sturmian languages $\mathcal{L}$ over $\{0, 1\}$, $\sigma(\mathcal{L})$ is Sturmian.*

The four morphisms L_0, L_1, R_0, R_1 can be generalized to larger alphabets. They are then called Arnoux-Rauzy morphisms.

Definition 4.11. The *Arnoux-Rauzy morphisms* over $\mathcal{A}$ are defined by

$$L_\ell\colon \begin{cases} \ell \mapsto \ell \\ a \mapsto \ell a \quad \forall a \in \mathcal{A} \setminus \{\ell\} \end{cases} \qquad R_\ell\colon \begin{cases} \ell \mapsto \ell \\ a \mapsto a\ell \quad \forall a \in \mathcal{A} \setminus \{\ell\} \end{cases}$$

for any given letter $\ell \in \mathcal{A}$.

Justin and Pirillo proved the following result [JP02].

Proposition 4.12. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism with $\#\mathcal{B} \geq 2$. The following are equivalent:*

1. *$\sigma = \gamma \circ \tau$ where $\gamma\colon \mathcal{A}^* \to \mathcal{B}^*$ is a bijective letter-to-letter morphism and τ is a composition of morphisms L_ℓ, R_ℓ, $\ell \in \mathcal{A}$;*

2. *there exists a recurrent Arnoux-Rauzy language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is an Arnoux-Rauzy language over $\mathcal{B}$;*

3. *for all Arnoux-Rauzy languages $\mathcal{L}$ over $\mathcal{A}$, $\sigma(\mathcal{L})$ is an Arnoux-Rauzy language over $\mathcal{B}$.*

The Arnoux-Rauzy morphisms have a particular behavior regarding extensions. Indeed, they preserve the extensions of the bispecial words, as described below.

Lemma 4.13. *Let $\mathcal{L}$ be a language over $\mathcal{A}$. For all $\ell \in \mathcal{A}$, we have*

1. *$L_\ell(\mathcal{L}) = R_\ell(\mathcal{L})$;*

[1] Recall that, in this work, a Sturmian language is simply a language of complexity $n+1$ and is therefore not necessarily recurrent. This is also true for Arnoux-Rauzy languages. However, the recurrent hypothesis is needed here to avoid morphisms of the form $0 \mapsto 0^k$, $1 \mapsto 1$, $k \geq 2$ which preserve the language $\mathcal{L} = \{0^n : n \geq 0\} \cup \{0^n 1 0^m : n, m \geq 0\}$ for example.

2. *$E_{L_\ell(\mathcal{L})}(\varepsilon) = (\mathcal{A} \times \{\ell\}) \cup (\{\ell\} \times \mathcal{A})$;*

3. *if $w \in L_\ell(\mathcal{L})$ is bispecial, then $w = \varepsilon$ or there exists $u \in \mathcal{L}$ such that $w = L_\ell(u)\ell = \ell R_\ell(u)$;*

4. *for all $u \in \mathcal{L}$, we have $E_{L_\ell(\mathcal{L})}(L_\ell(u)\ell) = E_{\mathcal{L}}(u)$.*

Proof. The first claim follows from the observation that $L_\ell(u)\ell = \ell R_\ell(u)$ for all $u \in \mathcal{A}^*$. Indeed, we can easily verify it for the empty word and for letters. The conclusion then follows by induction.

We now turn to the second claim. By definition of L_ℓ, ℓ can be followed by any letter in $L_\ell(\mathcal{L})$ and any letter other than ℓ is always preceded by ℓ. Similarly, by definition of R_ℓ, ℓ can be preceded by any letter in $R_\ell(\mathcal{L})$ and any letter other than ℓ is always followed by ℓ. The complete description of $E_{L_\ell(\mathcal{L})}(\varepsilon)$ then follows from the first claim.

For the third claim, since ℓ is the only left (resp., right) special letter by the second claim, any bispecial word is either empty or it begins and ends with ℓ. In the second case, due to the shape of the morphism L_ℓ, there exists $u \in \mathcal{L}$ such that the only coverings of w for the morphism L_ℓ are of the form $(ub, 0)$, $b \in E^R_{\mathcal{L}}(u)$. In particular, $w = L_\ell(u)\ell$.

Since the only coverings of $L_\ell(u)\ell$ are $(ub, 0)$, $b \in E^R_{\mathcal{L}}(u)$, and since $L_\ell(a)$ ends with a for all a and $L_\ell(b)\ell$ begins with ℓb for all b, we have

$$(a, b) \in E_{\mathcal{L}}(u) \iff (a, b) \in E_{L_\ell(\mathcal{L})}(L_\ell(u)\ell).$$

This proves the last claim. □

This preservation of the extensions of words directly implies that the Arnoux-Rauzy morphisms are dendric preserving. We have in fact the following stronger result.

Proposition 4.14. *For any language $\mathcal{L}$ over $\mathcal{A}$ and any letter $\ell \in \mathcal{A}$, $\mathcal{L}$ is dendric if and only if $L_\ell(\mathcal{L}) = R_\ell(\mathcal{L})$ is.*

In particular, a morphism τ is dendric preserving if and only if $L_\ell \circ \tau$ (resp., $R_\ell \circ \tau$) is.

Proof. This is a direct consequence of Lemma 4.13 and of the fact that ordinary words and non-bispecial words are always dendric. □

It is clear that bijective letter-to-letter morphisms also satisfy this kind of equivalences. We then have the following corollary.

Corollary 4.15. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism. If $\sigma = \gamma \circ \tau$ with $\gamma\colon \mathcal{A}^* \to \mathcal{B}^*$ a bijective letter-to-letter morphism and τ a composition of Arnoux-Rauzy morphisms over $\mathcal{A}$, then σ is dendric preserving.*

Observe that, instead of considering Arnoux-Rauzy morphisms for any letter $\ell \in \mathcal{A}$, we could have limited ourselves to Arnoux-Rauzy morphisms for a fixed letter ℓ and allow composition with permutations of $\mathcal{A}$ to obtain the other Arnoux-Rauzy morphisms. Indeed, for any permutation π of $\mathcal{A}$, if σ_π denotes the morphism $a \mapsto \pi(a)$, we have $L_{\pi(\ell)} = \sigma_\pi L_\ell \sigma_{\pi^{-1}}$ and $R_{\pi(\ell)} = \sigma_\pi R_\ell \sigma_{\pi^{-1}}$. Since we are composing with a bijective letter-to-letter morphism at the end, allowing permutations does not generate a larger family of morphisms.

4.2.2 Characterization of dendric preserving morphisms

The purpose of this section is to prove the converse of Corollary 4.15, i.e. the only dendric preserving morphisms are compositions of Arnoux-Rauzy morphisms and bijective letter-to-letter morphisms.

Right away, we see that for the morphisms of Corollary 4.15, the initial alphabet and the image alphabet have the same size. We can in fact make the following stronger observation.

Remark 4.16. Let us look at the images of the letters under the morphism R_ℓ. We see that, for all $a \in \mathcal{A}$, $R_\ell(a)$ begins with a so the images all start with a different letter. Moreover, the images all end with ℓ, and if we consider $\ell\sigma(a)$ (which makes sense since $\sigma(a)$ can only be preceded by the image of another letter in $\sigma(\mathcal{L})$), then the letters preceding this common suffix ℓ are all different.

Similarly, for L_ℓ, the letters following the common prefix $p_{L_\ell} = \ell$ in $\sigma(a)p_{L_\ell}$, $a \in \mathcal{A}$, are all different and the letters preceding the common suffix $s_{L_\ell} = \varepsilon$ in $s_{L_\ell}\sigma(a)$, $a \in \mathcal{A}$, are also all different. This property is stable when applying a bijective letter-to-letter morphism and we can also show that it is preserved when composing with an Arnoux-Rauzy morphism.

We will prove that this observation is in fact satisfied by any dendric preserving morphism. We first need to properly define these common prefix p_σ and common suffix s_σ. This is the object of the following definition.

Definition 4.17. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. We denote p_σ (resp., s_σ) the longest common prefix (resp., suffix) to all the $\sigma(a)^\omega = \sigma(a)\sigma(a)\cdots$ (resp., ${}^\omega\sigma(a) = \cdots\sigma(a)\sigma(a)$), $a \in \mathcal{A}$.

Observe that, p_σ and s_σ can be empty or infinite. However, in the case of an *aperiodic morphism*, i.e. if there exists a language whose image has unbounded complexity, p_σ and s_σ are always finite as stated by the following lemma. Note that it is in particular the case for dendric preserving morphisms with non-unary image alphabet.

Lemma 4.18. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. If p_σ (resp., s_σ) is infinite, then there exists a word $v \in \mathcal{B}^+$ such that $\sigma(\mathcal{L}) = \cup_{n\in\mathbb{N}} \mathrm{Fac}(v^n)$ for all languages $\mathcal{L} \subseteq \mathcal{A}^*$.*

Proof. Assume that p_σ is infinite, the proof for s_σ is similar. Thus $\sigma(a)^\omega = p_\sigma$ for all $a \in \mathcal{A}$. By Fine and Wilf's theorem, this implies that there exists a word v such that $\sigma(a)$ is a power of v for all $a \in \mathcal{A}$. This proves that $\sigma(u) \in \{v\}^*$ for all $u \in \mathcal{A}^*$ so the image of any language only contains factors of powers of v. □

The choice of using $\sigma(a)^\omega$ to define p_σ makes the previous lemma almost trivial but will not be the most useful in the following results. The next lemma provides different equivalent ways of defining p_σ as the longest word satisfying some property. We also have a similar result for s_σ using suffixes.

Lemma 4.19. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. For any word $p \in \mathcal{B}^*$ and any letter $a \in \mathcal{A}$, the following are equivalent:*

1. *p is a prefix of $\sigma(a)^\omega$;*
2. *p is a proper prefix of $\sigma(a)p$.*

Moreover, the following are also equivalent:

1. *p satisfies one of the (equivalent) properties above for every letter $a \in \mathcal{A}$;*
2. *p is a prefix of $\sigma(w)p$ for any $w \in \mathcal{A}^*$;*
3. *there exists $N \geq 0$ such that p is a prefix of $\sigma(w)$ for any $w \in \mathcal{A}^*$ such that $|w| \geq N$.*

Proof. Let us prove the first equivalence. If p is a prefix of $\sigma(a)^\omega$, then it is a prefix of $\sigma(a) \cdot \sigma(a)^\omega$ so it directly follows that it is a prefix of $\sigma(a)p$. For the converse, if p is a proper prefix of $\sigma(a)p$, then p is prefix comparable with $\sigma(a)$. Thus $\sigma(a)p$ is prefix comparable with $\sigma(a)^2$, and in particular, p is prefix comparable with $\sigma(a)^2$. We can iterate to show that p is prefix comparable with $\sigma(a)^k$ for any $k \geq 1$. In particular, since the morphism σ

is non-erasing, there exists k such that $k|\sigma(a)| \geq |p|$ so p is a prefix of $\sigma(a)^k$. This shows that p is a prefix of $\sigma(a)^\omega$.

We now turn to the second set of equivalences. Assume that p satisfies the previous properties for all the letters. We proceed by induction on the length of w to show that p is a prefix of $\sigma(w)p$. If $w = \varepsilon$, it is trivial. Assume that it is satisfied for w' and that $w = w'a$, $a \in \mathcal{A}$. By hypothesis, p is a prefix of $\sigma(a)p$ thus $\sigma(w')p$ is a prefix of $\sigma(w)p$. The conclusion follows by induction hypothesis.

Since the morphism σ is non-erasing, for any word w such that $|w| \geq |p|$, we have $|\sigma(w)| \geq |p|$ thus, for any such word w, p being a prefix of $\sigma(w)p$ directly implies that p is a prefix of $\sigma(w)$.

Finally, if p is a prefix of $\sigma(w)$ for any long enough w, then p is a prefix of $\sigma(a^k)$ for any large enough k thus p is a prefix of $\sigma(a)^\omega$ for any letter $a \in \mathcal{A}$. $\square$

We now show that the observation made in Remark 4.16 is true for any dendric preserving morphism.

Proposition 4.20. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing dendric preserving morphism. For each letter $b \in \mathcal{B}$, there exists at most one letter $a \in \mathcal{A}$ such that $p_\sigma b$ is a prefix of $\sigma(a)p_\sigma$ and at most one letter $a' \in \mathcal{A}$ such that bs_σ is a suffix of $s_\sigma\sigma(a')$.*

Proof. Assume by contradiction that there exist two letters $a, a' \in \mathcal{A}$ such that $p_\sigma b$ is a prefix of both $\sigma(a)p_\sigma$ and $\sigma(a')p_\sigma$. By maximality of p_σ, there also exists a letter $a'' \in \mathcal{A}$ and a letter $b' \neq b$ such that $p_\sigma b'$ is a prefix of $\sigma(a'')p_\sigma$. Similarly, by maximality of s_σ, there exist two distinct letters $d, d' \in \mathcal{B}$ and two letters $c, c' \in \mathcal{A}$ such that ds_σ is a suffix of $s_\sigma\sigma(c)$ and $d's_\sigma$ is a suffix of $s_\sigma\sigma(c')$.

Up to exchanging c and c', we can assume that $c \neq a$. We claim that there exists a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $ca, ca'', c'a'', c'a' \in \mathcal{L}_2$. We can build it as follows. Let $\leq$ be an order over $\mathcal{A}$ whose three smallest elements are $a \leq a'' \leq a'$. Let c'' be the largest letter for $\leq$.

- If $c'' \notin \{c, c'\}$, then let $\preceq$ be an order over $\mathcal{A}$ whose three smallest elements are $c'' \preceq c \preceq c'$. By construction, the pair $(\leq, \preceq)$ is irreducible, i.e. for all $1 \leq k \leq \#\mathcal{A} - 1$, the k smallest elements for both orders do not coincide. We then consider the interval exchange transformation T associated with $\binom{\leq}{\preceq}$ and some rationally independent lengths such that
$$\lambda_{c''} < \lambda_a < \lambda_{c''} + \lambda_c < \lambda_a + \lambda_{a''} < \lambda_{c''} + \lambda_c + \lambda_{c'}.$$
This is possible since $c'' \notin \{a, a''\}$.

- If $c'' \in \{c, c'\}$, then let $\preceq$ be an order over $\mathcal{A}$ whose two smallest elements are $c \preceq c'$. By construction and since $c \neq a$, the pair $(\leq, \preceq)$ is irreducible. We then consider the interval exchange transformation T associated with $\binom{\leq}{\preceq}$ and some rationally independent lengths such that
$$\lambda_a < \lambda_c < \lambda_a + \lambda_{a''} < \lambda_c + \lambda_{c'}.$$
This is possible as $c \neq a$ and $\{c, c'\} \neq \{a, a''\}$ since $c'' \in \{c, c'\} \setminus \{a, a''\}$.

By construction, T is in fact an RIET and its language $\mathcal{L}$ (which is dendric) contains the words ca, ca'', $c'a''$ and $c'a'$. In $\sigma(\mathcal{L})$, the extension graph of $s_\sigma p_\sigma$ then contains the cycle $(b^R, d^L, b'^R, d'^L, b^R)$, which contradicts the fact that σ is dendric preserving. □

Observe that the previous result is direct when $\#\mathcal{A} = 2$ by definition of p_σ. Moreover, the same conclusion can be reached with much simpler hypotheses by doing a careful analysis of the proof. We detail this in the next subsection.

Using Corollary 4.2, we obtain as a direct corollary that, if a morphism is dendric preserving then the initial alphabet and the image alphabet have the same number of letters.

Corollary 4.21. *Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing morphism. If σ is dendric preserving, then $\#\mathcal{A} = \#\mathcal{B}$.*

We now use Proposition 4.20 to show that the only dendric preserving morphisms are generated by Arnoux-Rauzy morphisms. To do so, we proceed by induction on $|s_\sigma p_\sigma|$. The base case is given by the following lemma.

Lemma 4.22. *Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing dendric preserving morphism. If $s_\sigma p_\sigma = \varepsilon$, then σ is a bijective letter-to-letter morphism.*

Proof. It suffices to prove that all the images of letters have length one as the images of the letters will then all be different by Proposition 4.20. First, observe that in the case of two letters alphabets, it directly follows from the study of Sturmian morphisms, or more precisely, from [Lot02, Lemma 2.3.8].

For larger alphabets, assume by contradiction that there exist $a \in \mathcal{A}$, $b, c \in \mathcal{B}$ such that bc is a factor of $\sigma(a)$. Let b' denote the letter such that $\sigma(b')$ ends with b and c' be the letter such that $\sigma(c')$ begins with c. Such letters exist by Proposition 4.20 and Corollary 4.21.

Since we are on an alphabet of size at least 3, we can find a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $b'c'$ is not in its language. For example, we can take an Arnoux-Rauzy language where the bispecial letter is not in $\{b', c'\}$. Observe that this would not be true on a binary alphabet if $b' \neq c'$, hence why we considered the case of binary alphabets separately.

Since $\mathcal{L}$ is dendric, the vertices b'^L and c'^R are connected by a unique path in $\mathcal{E}_{\mathcal{L}}(\varepsilon)$. Moreover, this path is not reduced to the edge (b', c') as this edge does not exist by definition of $\mathcal{L}$. By Proposition 4.20, any simple path in $\mathcal{E}_{\mathcal{L}}(\varepsilon)$ corresponds to a simple path of the same length in $\mathcal{E}_{\sigma(\mathcal{L})}(s_\sigma p_\sigma) = \mathcal{E}_{\sigma(\mathcal{L})}(\varepsilon)$ by simply renaming the left vertices according to the last letter of their image and the right vertices according to the first letter. This implies that b and c are connected by a path in $\mathcal{E}_{\sigma(\mathcal{L})}(\varepsilon)$ and this path is not reduced to the edge (b, c). However, bc is a factor of $\sigma(a)$ thus (b, c) is an edge of $\mathcal{E}_{\sigma(\mathcal{L})}(\varepsilon)$ and we have a cycle, a contradiction since σ is dendric preserving. $\square$

The induction step will require the following lemma.

Lemma 4.23. *Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing dendric preserving morphism. If $|s_\sigma p_\sigma| = n > 0$, then*

1. *$(s_\sigma p_\sigma)_1 = (s_\sigma p_\sigma)_n =: \ell$ and it is such that $E_{\sigma(\mathcal{L})}(\varepsilon) = (\{\ell\} \times \mathcal{B}) \cup (\mathcal{B} \times \{\ell\})$ for any dendric language $\mathcal{L}$ over $\mathcal{A}$;*

2. *there exists a dendric preserving morphism $\tau \colon \mathcal{A}^* \to \mathcal{B}^*$ such that $\sigma \in \{L_\ell \circ \tau, R_\ell \circ \tau\}$. Moreover, $|s_\tau p_\tau| < |s_\sigma p_\sigma|$.*

Proof. Let us prove the first claim. Let $\mathcal{L}$ be a dendric language over $\mathcal{A}$. By Proposition 4.20, $(s_\sigma p_\sigma)_n$ can be extended on the right by at least $\#\mathcal{A}$ different letters in $\sigma(\mathcal{L})$ so, by Corollary 4.21, $E^R_{\sigma(\mathcal{L})}((s_\sigma p_\sigma)_n) = \mathcal{B}$. Similarly, we show that $E^L_{\sigma(\mathcal{L})}((s_\sigma p_\sigma)_1) = \mathcal{B}$. This implies that $E_{\sigma(\mathcal{L})}(\varepsilon) \supseteq (\{(s_\sigma p_\sigma)_n\} \times \mathcal{B}) \cup (\mathcal{B} \times \{(s_\sigma p_\sigma)_1\})$, but as ε is dendric in $\sigma(\mathcal{L})$, this must be an equality.

Let us now show that $(s_\sigma p_\sigma)_1 = (s_\sigma p_\sigma)_n$. Assume that it is not the case. Given the extensions of ε that we have just found, $(s_\sigma p_\sigma)_1$ can then only be followed by itself in $\sigma(\mathcal{L})$. This contradicts the fact that $(s_\sigma p_\sigma)_n$ appears in every word of length $\|\sigma\|$ in $\sigma(\mathcal{L})$.

We now turn to the second claim. Assume first that $p_\sigma \neq \varepsilon$ and that $(p_\sigma)_1 = \ell$. In other words, for each letter $a \in \mathcal{A}$, $\sigma(a)$ begins with ℓ. By the first claim, any letter other than ℓ can only be followed by ℓ in $\sigma(a)$ thus we can find u such that $\sigma(a) = L_\ell(u)$. We then define τ such that $\sigma = L_\ell \circ \tau$. Note that, by maximality of s_σ and p_σ, we have $s_\sigma p_\sigma = L_\ell(s_\tau p_\tau)\ell$.

If $p_\sigma = \varepsilon$ or $(p_\sigma)_1 \neq \ell$, then we first show that, for each letter $a \in \mathcal{A}$, $\sigma(a)$ ends with ℓ. Let $a \in \mathcal{A}$. Using the first claim, for any dendric language

$\mathcal{L}$, the letter ℓ appears in every length-2 element of $\sigma(\mathcal{L})$. Therefore, if we can find a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $ab \in \mathcal{L}$ where $\sigma(b)$ does not begin with ℓ, this will imply that $\sigma(a)$ ends with ℓ. Let us find such a dendric language. If $p_\sigma = \varepsilon$, then by Proposition 4.20, there is exactly one letter $a_0 \in \mathcal{A}$ such that $\sigma(a_0)$ begins with ℓ so any $b \neq a_0$ will do. If $(p_\sigma)_1 \neq \ell$, then the image of the letters never begin with ℓ so any $b \in \mathcal{A}$ will do. We can therefore simply take $\mathcal{L}$ such that a is right special to conclude that $\sigma(a)$ ends with ℓ. Similarly to what we did in the first case, we can now define τ such that $\sigma = R_\ell \circ \tau$. Note that $s_\sigma p_\sigma = \ell R_\ell(s_\tau p_\tau)$.

In both cases, τ is dendric preserving by Proposition 4.14, and $|s_\tau p_\tau| < |s_\sigma p_\sigma|$. □

We can now prove the main result of this section.

Theorem 4.24. *A non-erasing morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, $\#\mathcal{B} \geq 2$, is dendric preserving if and only if it is, up to a bijective letter-to-letter morphism, a composition of Arnoux-Rauzy morphisms over $\mathcal{A}$.*

Proof. The fact that such a morphism is dendric preserving was proved in Corollary 4.15. Assume now that $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ is dendric preserving. Using Lemma 4.22 and Lemma 4.23, we can effectively obtain, by induction on $|s_\sigma p_\sigma|$, a composition τ of Arnoux-Rauzy morphisms over $\mathcal{B}$ and a bijective letter-to-letter morphism γ such that $\sigma = \tau \circ \gamma$. We can always modify the Arnoux-Rauzy morphisms so that the bijective letter-to-letter morphism is applied after the Arnoux-Rauzy morphisms. □

4.2.3 Stronger version and RIET

For most of the results of the previous subsection, the hypotheses can be largely weakened by looking closely at the proofs. We detail these changes below. In some sense, this stronger version shows that when looking at dendric preserving morphisms, what really matters is the connectedness of the empty word in the initial languages and the acyclicity of the small factors in the image languages.

Theorem 4.25. *Let $\mathcal{A}$ and $\mathcal{B}$ be such that $\#\mathcal{B} \leq \#\mathcal{A}$ and let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing aperiodic morphism. If there exists a family F of languages over $\mathcal{A}$ such that*

(H1) *for all distinct $a, a' \in \mathcal{A}$ and all distinct $b, b', b'' \in \mathcal{A}$, there exists $\mathcal{L} \in F$ such that $ab, ab', a'b', a'b''$ are in $\mathcal{L}$, or $a'b, a'b', ab', ab''$ are in $\mathcal{L}$;*

(H2) *for all distinct $a, a', a'' \in \mathcal{A}$ and all distinct $b, b' \in \mathcal{A}$, there exists $\mathcal{L} \in F$ such that $ab, a'b, a'b', a''b'$ are in $\mathcal{L}$, or $ab', a'b', a'b, a''b$ are in $\mathcal{L}$;*

(H3) *for all $a, b \in \mathcal{A}$ (non necessarily distinct), there exists $\mathcal{L} \in F$ such that ab is not in $\mathcal{L}$ but a^L and b^R are connected in $\mathcal{E}_{\mathcal{L}}(\varepsilon)$;*

(H4) *the factors of $s_\sigma p_\sigma$ are acyclic in $\sigma(\mathcal{L})$ for all $\mathcal{L} \in F$,*

then σ is, up to a bijective letter-to-letter morphism, a composition of Arnoux-Rauzy morphisms over $\mathcal{A}$.

Proof. As explained in Lemma 4.18, the finiteness of p_σ and s_σ is guaranteed as soon as σ is aperiodic, i.e. there exists a language $\mathcal{L}$ such that $\sigma(\mathcal{L})$ is not periodic.

In Proposition 4.20, we build a particular language which is then used to obtain a contradiction. Hypothesis (H1) on F guarantees the existence of such a language. The contradiction then comes from the fact that $s_\sigma p_\sigma$ is acyclic in $\sigma(\mathcal{L})$, which is implied by Hypothesis (H4). To do the symmetric proof where we look at the letters preceding s_σ, we use Hypothesis (H2) instead of Hypothesis (H1).

To deduce Corollary 4.21, we used Corollary 4.2 which we replace here by the hypothesis $\#\mathcal{B} \leq \#\mathcal{A}$.

For Lemma 4.22, we once again need to build a counter-example whose existence is guaranteed here by Hypothesis (H3). The contradiction then follows by acyclicity of $\varepsilon = s_\sigma p_\sigma$ which is implied by Hypothesis (H4).

Lemma 4.23 is slightly more tricky. For the first claim, we only need the acyclicity of ε which is once again a consequence of Hypothesis (H4). However, we need to slightly adapt the second claim. Indeed, we now want to show that $\sigma \in \{L_\ell \circ \tau, R_\ell \circ \tau\}$ where τ is aperiodic and satisfies Hypothesis (H4) (which are the only hypotheses involving σ in this theorem). For the existence of τ, we only need to know that, for all $a, b \in \mathcal{A}$, there exists $\mathcal{L} \in F$ such that b is not a right extension of a. This is a consequence of Hypothesis (H3). Since σ is aperiodic, τ also is. Let us now show that τ satisfies Hypothesis (H4). We prove it when $\sigma = L_\ell \circ \tau$, the other case is symmetric. We then have $s_\sigma p_\sigma = L_\ell(s_\tau p_\tau)\ell$. Let $\mathcal{L} \in F$ and let $u \in \mathrm{Fac}(s_\tau p_\tau)$. By Lemma 4.13, we have $\mathcal{E}_{\tau(\mathcal{L})}(u) = \mathcal{E}_{\sigma(\mathcal{L})}(L_\ell(u)\ell)$ but as $u \in \mathrm{Fac}(s_\tau p_\tau)$, $L_\ell(u)\ell$ is a factor of $s_\sigma p_\sigma$ so it is acyclic in $\sigma(\mathcal{L})$ by Hypothesis (H4) on σ. This shows that τ also satisfies Hypothesis (H4).

The conclusion then follows as in Theorem 4.24. $\square$

Observe that Hypothesis (H3) is not needed to conclude that $\#\mathcal{A} = \#\mathcal{B}$ and that Hypotheses (H1) and (H2) are trivially satisfied if $\#\mathcal{A} = 2$.

Theorem 4.25 also shows that, to reach the same conclusion as in Theorem 4.24, it suffices to preserve dendricity for a small well-chosen family of dendric languages. For example, in the case of the languages of RIET, we obtain the following result.

Corollary 4.26. *Let $\mathcal{A}$ and $\mathcal{B}$ be such that $\#\mathcal{A} \geq \max\{3, \#\mathcal{B}\}$, and let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing aperiodic morphism. If, for all languges $\mathcal{L}$ of RIET over $\mathcal{A}$, the factors of $s_\sigma p_\sigma$ are acyclic in $\sigma(\mathcal{L})$, then σ is, up to a bijective letter-to-letter morphism, a composition of Arnoux-Rauzy morphisms over $\mathcal{A}$.*

Proof. It suffices to show that the family of languages of RIET over $\mathcal{A}$ satisfies Hypotheses (H1), (H2) and (H3) of Theorem 4.25. We have in fact already shown in the proof of Proposition 4.20 that it satisfies (H1) and (H2). For (H3), we will use the assumption that $\#\mathcal{A} \geq 3$. We can then find an irreducible pair of orders $\binom{\leq}{\preceq}$ such that a is the minimum for $\preceq$ and b the maximum for $\leq$. The interval exchange transformation T associated with this pair of orders and with rationally independent lengths such that $\lambda_a + \lambda_b \leq 1$ is an RIET such that ab is not in its language $\mathcal{L}$ (but a^L and b^R are connected in $\mathcal{E}_\mathcal{L}(\varepsilon)$ since $\mathcal{L}$ is dendric). □

Observe however that the family of Arnoux-Rauzy languages over $\mathcal{A}$ does not satisfy Hypotheses (H1) and (H2). In fact, we will show in Section 4.4 that there are morphisms which are not generated by Arnoux-Rauzy morphisms but such that the image of any Arnoux-Rauzy language over $\mathcal{A}$ is dendric.

We end this subsection (and section) with a note on morphisms preserving RIET, i.e. the morphisms such that the image of a language of an RIET over $\mathcal{A}$ is the language of an RIET. For conciseness, we call such morphisms *RIET preserving*. As a consequence of the study done in this section, we can describe RIET preserving morphisms. We first need the following direct lemma.

Lemma 4.27. *Let $\mathcal{L}$ be a language. If $L_\ell(\mathcal{L})$ (resp., $R_\ell(\mathcal{L})$) is the language of an RIET associated with the orders $\binom{\leq}{\preceq}$, then $\mathcal{L}$ is the language of an RIET for the same orders.*

Proof. By Lemma 4.13, if $L_\ell(\mathcal{L})$ satisfies the conditions of the combinatorial characterization of languages of RIET (Theorem 1.37), then so does $\mathcal{L}$ for the same orders. □

Theorem 4.28. *A non-erasing morphism* $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, $\#\mathcal{B} \geq 2$, *is RIET preserving if and only if we are in one of the following cases:*

1. $\#\mathcal{A} = 2$ *and* σ *is a Sturmian morphism;*

2. $\#\mathcal{A} \geq 3$ *and* σ *is a bijective letter-to-letter morphism.*

Proof. If $\#\mathcal{A} = 2$, then σ satisfies all hypotheses of Theorem 4.25 except (H3) so we can still conclude that $\#\mathcal{B} = 2$. In other words, σ preserves in fact the recurrent Sturmian languages so it is a Sturmian morphism.

When $\#\mathcal{A} \geq 3$, by Corollary 4.26, it only remains to prove that the compositions of Arnoux-Rauzy morphisms are not RIET preserving. Moreover, using Lemma 4.27, it suffices to prove that the Arnoux-Rauzy morphisms themselves are not RIET preserving.

Let $\ell \in \mathcal{A}$ and let $\mathcal{L}$ be the language of an RIET over $\mathcal{A}$ associated with the orders $\binom{\leq}{\preceq}$ such that ℓ is neither the maximum nor the minimum for $\leq$. This is possible as we are on an alphabet of size at least 3. By Proposition 1.40, the pair of orders $\binom{\leq}{\preceq}$ is then the only one (up to reversal) satisfying the hypotheses of Theorem 1.37. By Lemma 4.13, these are also the only orders satisfying the hypotheses of Theorem 1.37 for all non-empty $w \in L_\ell(\mathcal{L}) = R_\ell(\mathcal{L})$. However, the empty word is not planar for these orders. Indeed, $E_{L_\ell(\mathcal{L})}(\varepsilon) = (\{\ell\} \times \mathcal{A}) \cup (\mathcal{A} \times \{\ell\})$ by Lemma 4.13 and ℓ is neither the minimum nor the maximum for $\leq$. This shows that no pair of orders satisfies Theorem 1.37 for $L_\ell(\mathcal{L}) = R_\ell(\mathcal{L})$, therefore it is not the language of an RIET. This ends the proof that neither L_ℓ nor R_ℓ are RIET preserving. $\square$

There are so few RIET preserving morphisms due to the fact that we want the morphism to preserve languages of RIET starting from RIET associated with any pair of irreducible orders. If we fix the orders, we obtain vastly different results. Indeed, in [AP07], the authors proved that, for the family of RIET associated with the orders $\binom{1<2<3}{3\prec 2\prec 1}$, the monoid of preserving morphisms is infinitely generated.

4.3 Return morphisms

Our original motivation with J. Leroy and M. Lejeune to look at preservation of dendricity when applying a morphism is the search of an S-adic characterization (see Chapter 5). In this context, we were working with very specific morphisms which essentially coded the return words for some given word w.

In this section, we introduce these so-called *return morphisms*. We first define them in Subsection 4.3.1 with respect to some word. As done for return words, we can in fact study return morphisms with respect to a factor code S. This is what we do in Subsection 4.3.2. If the image of a language under a return morphism for a word is dendric, then we can deduce properties of the morphism based on properties of return words in dendric languages. In particular, the return morphism is *tame* in the sense that it is generated by some extremely simple morphisms. We say a word on tame morphisms and their link with return morphism for a word in Subsection 4.3.3.

4.3.1 Return morphism for a word

We define here return morphisms for a word w as studied in [GLL22] (in the case where w is a letter) then in [GL22] (in the general case).

Definition 4.29. A *return morphism for a word* $w \in \mathcal{B}^+$ is a morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ injective on the letters and such that, for all $a \in \mathcal{A}$, $\sigma(a)w$ contains exactly two occurrences of w: one as a proper prefix and one as a proper suffix.

As in Section 3.3 on return words, we chose to consider left return morphisms but we could have just as well defined right return morphisms by considering the occurrences of w in $w\sigma(a)$. All the properties can easily be adapted.

Observe that a return morphism σ is non-erasing. Moreover, the set $\sigma(\mathcal{A})$ is a suffix code. Indeed, if $\sigma(a)$ was a proper suffix of $\sigma(b)$, then $\sigma(b)w$ would contain three occurrences of w. This implies in particular that injectivity on the letters is equivalent to injectivity on the words.

Example 4.30. The morphism σ such that $\sigma(0) = 01$, $\sigma(1) = 010$ and $\sigma(2) = 0102$ is a return morphism for 01. Note that it is also a return morphism for 010.

Before showing that return morphisms are indeed related to return words, we need the following lemma, which will also imply that return morphisms are well-behaved in terms of coverings in Section 4.4.

Lemma 4.31. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for $w \in \mathcal{B}^+$. For any language $\mathcal{L} \subseteq \mathcal{A}^*$ and for all $u = u_1 \cdots u_n \in \mathcal{L}$, $n \geq 0$, we have the following properties:*

1. *w is a prefix of $\sigma(u)w$, and there exists $N \geq 0$ such that, if $n \geq N$, w is a prefix of $\sigma(u)$;*

2. *if $\sigma(u)$ contains an occurrence of w, then there exists i such that this occurrence is a prefix of $\sigma(u_i \cdots u_n)$;*

3. *if w is a prefix of $\sigma(u_i \cdots u_n)$, then for all $1 \leq j \leq i$, w is a prefix of $\sigma(u_j \cdots u_n)$.*

Proof. We prove the first claim by induction on $|u|$. If $u = \varepsilon$, this is trivially true and if $u \in \mathcal{A}$, it follows by definition of a return morphism. Assume now that $u = va$, $a \in \mathcal{A}$, and that w is a prefix of $\sigma(v)w$. As w is a prefix of $\sigma(a)w$, this shows that w is a prefix of $\sigma(v)\sigma(a)w = \sigma(u)w$. The existence of N follows from the fact that σ is non-erasing.

For the second claim, assume that an occurrence of w starts in $\sigma(u_i)$. By the first claim, w is a prefix of $\sigma(u_{i+1} \cdots u_n)w$ and, up to extending u on the right, we can even assume that w is a prefix of $\sigma(u_{i+1} \cdots u_n)$. Therefore, $\sigma(u_i)w$ is a prefix of $\sigma(u_i \cdots u_n)$ and it has an occurrence of w starting in $\sigma(u_i)$. By definition of σ, this implies that this occurrence is a prefix.

We directly deduce the third claim by decreasing induction on $j \leq i$. Indeed, $\sigma(u_j)w$ is then a prefix of $\sigma(u_j \cdots u_n)$ so it has w as a prefix by the first claim. □

We can now motivate the "return morphism" terminology.

Proposition 4.32. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism and $w \in \mathcal{B}^+$. If σ is injective on the letters, then the following are equivalent:*

1. *σ is a return morphism for w;*

2. *for any language $\mathcal{L}$ over $\mathcal{A}$, $\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{L})}(w)$;*

3. *there exists a language $\mathcal{L}' \subseteq \mathcal{B}^*$ such that $\sigma(\mathcal{A}) = \mathrm{R}_{\mathcal{L}'}(w)$.*

Proof. Assume that σ is a return morphism for w. By the first claim of Lemma 4.31, $\sigma(a)w \in \sigma(\mathcal{L})$ for any language $\mathcal{L}$ over $\mathcal{A}$. Therefore, by definition of a return morphism, the elements of $\sigma(\mathcal{A})$ are return words for w in $\sigma(\mathcal{L})$. Using the second and third claims of Lemma 4.31, two consecutive occurrences of w in $\sigma(u)$ occur as prefixes of $\sigma(u_i \cdots u_n)$ and $\sigma(u_{i+1} \cdots u_n)$ for some i. This shows that the only return words for w are the elements of $\sigma(\mathcal{A})$.

The implication from the second to the third assertion is direct as it suffices to take $\mathcal{L}' = \sigma(\mathcal{L})$ for any language $\mathcal{L}$ over $\mathcal{A}$.

Assume now that $\sigma(\mathcal{A}) = \mathrm{R}_{\mathcal{L}}(w)$ for some language $\mathcal{L} \subseteq \mathcal{B}^*$. Then $\sigma(a)$ is a return word for w in $\mathcal{L}$ so $\sigma(a)w$ contains exactly two occurrences of w, one as a prefix and one as a suffix. Since σ is injective on the letters, this implies that σ is a return morphism. $\square$

As we saw in the Example 4.30, the word w for which σ is a return morphism is not always unique. However, if σ is a return morphism for several words, these words are related as described in the following result.

Proposition 4.33. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for two distinct words w and w'. If $|w| \leq |w'|$, then*

1. *w is a proper prefix of w';*
2. *σ is a return morphism for all prefixes of w' of length at least $|w|$;*
3. *for any language $\mathcal{L} \subseteq \mathcal{A}^*$, w is not right special in $\sigma(\mathcal{L})$;*
4. *if w' is of maximal length, it is right special in $\sigma(\mathcal{L})$ for all language $\mathcal{L} \subseteq \mathcal{A}^*$, and $w' = p_\sigma$.*

Proof. Let us prove the four claims separately.

1. Using the first claim of Lemma 4.31 for the language $\mathcal{A}^*$, there exists $u \in \mathcal{A}^*$ such that w and w' are prefixes of $\sigma(u)$. Therefore, w and w' are prefix comparable. Since $|w| \leq |w'|$ and $w \neq w'$, we conclude that w is a proper prefix of w'.

2. Let wu be a prefix of w'. Since σ is a return morphism for w', wu is a prefix and a suffix of $\sigma(a)wu$ for all $a \in \mathcal{A}$. Moreover, since w only has two occurrences in $\sigma(a)w$, wu only has two occurrences in $\sigma(a)wu$ thus σ is a return morphism for wu.

3. Since σ is a return morphism for w, w only occurs in $\sigma(\mathcal{L})$ as a proper prefix of some $\sigma(a)w$, $a \in \mathcal{A}$, by Lemma 4.31. However, for all $a \in \mathcal{A}$, w' is a prefix of $\sigma(a)w'$ thus $ww'_{|w|+1}$ is a prefix of $\sigma(a)w$. The only right extension of w is then $w'_{|w|+1}$.

4. Assume that w' only has one right extension b. Thus $w'b$ is a prefix (and a suffix) of each $\sigma(a)w'b$, $a \in \mathcal{A}$. As w' only has two occurrences in $\sigma(a)w'$, the word $w'b$ only has two occurrences in $\sigma(a)w'b$, which proves that σ is a return morphism for $w'b$, and contradicts the maximality of w'. This shows that w' is right special. By Lemma 4.31, w' is then the longest word satisfying the conditions of Lemma 4.19 so it corresponds to p_σ.

□

4.3.2 Return morphism for a set of words

As we saw in Section 3.3, return words can also be defined with respect to a factor code. We similarly generalize the notion of return morphism for a word to return morphism for a set of words.

Definition 4.34. Let $S \subseteq \mathcal{B}^*$ be a finite factor code. A non-erasing morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ is a *return morphism for* S if it is injective over the letters and, for all $a \in \mathcal{A}$ and $w \in S$, the word $\sigma(a)w$ has a proper prefix and a proper suffix in S but contains no other occurrence of words of S.

Factor codes are always implicitly assumed to be finite in this section or more generally, when we talk about a return morphism. Recall that it is always the case if the factor code is included in a uniformly recurrent language (Remark 3.22).

Of course, any return morphism for a word w is a return morphism for the set $\{w\}$. There are also return morphisms for sets of larger size, as shown in the example below.

Example 4.35. The morphism σ such that $\sigma(0) = 0$, $\sigma(1) = 0102$ and $\sigma(2) = 012$ is a return morphism for $\{00, 01\}$. Indeed, we can check that, 000, 001, 010200, 010201, 01200 and 01201 all have exactly two occurrences of words in $\{00, 01\}$, one as a prefix and one as a suffix. Observe however that it is not a return morphism for a word.

Return morphisms for a set share some properties with their "for a word" counterparts. For example, they are also injective on the words since the images of the letters form a suffix code. This is however not an equivalence as shown below.

Example 4.36. Let σ be the morphism such that $\sigma(0) = 0$ and $\sigma(1) = 101$ and assume that σ is a return morphism for S. Then S clearly contains a word w beginning with 0 but not 0 itself since it is a non-prefix factor of $\sigma(1)$. Since $\sigma(1)w = 101w$ has a prefix in S and no internal occurrence of a word of S, this shows that S contains the word $101w^{(1)}$ for some proper prefix $w^{(1)}$ of w. Similarly, $\sigma(0)101w^{(1)} = 0101w^{(1)}$ has a prefix in S so, if $|w^{(1)}| \geq 2$, then 0 is a proper prefix of $w^{(1)}$ and S contains $0101w^{(2)}$ for some proper prefix $w^{(2)}$ of $w^{(1)}$. We then consider $\sigma(1)0101w^{(2)} = 1010101w^{(2)}$, showing that S contains $1010101w^{(3)}$ if $|w^{(2)}| \geq 2$. We can iterate the process until $|w^{(i)}| \leq 1$, i.e., $w^{(i)} \in \{\varepsilon, 0\}$.

If i is even, then S contains $(01)^i w^{(i)}$ so we consider $\sigma(1)(01)^i w^{(i)}$ which is equal to $1(01)^{i+1}w^{(i)}$. We see that $(01)^i w^{(i)}$ is an internal factor, which contradicts the fact that σ is a return morphism for S.

If i is odd, then S contains $(10)^i 1w^{(i)}$ so we consider $\sigma(0)(10)^i 1w^{(i)}$ which is equal to $(01)^{i+1}w^{(i)}$. This shows that S contains $(01)^i v$ for $v \in \{\varepsilon, 0, 01\}$ (depending on $|w^{(i)}|$). The conclusion follows as in the previous case by showing that $(01)^i v$ is an internal factor of $\sigma(1)(01)^i v$. Therefore, it is impossible to find such S.

Lemma 4.31 giving properties of return morphisms for a word can also be adapted as follows.

Lemma 4.37. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for a factor code $S \subseteq \mathcal{B}^*$. For any language $\mathcal{L} \subseteq \mathcal{A}^*$ and all $u = u_1 \cdots u_n \in \mathcal{L}$, we have the following properties:*

1. *for all $w \in S$, $\sigma(u)w$ has a prefix in S, and there exists $N \geq 0$ such that, if $n \geq N$, $\sigma(u)$ has a prefix in S;*
2. *if $\sigma(u)$ contains an occurrence of a word in S, then there exists i such that this occurrence is a prefix of $\sigma(u_i \cdots u_n)$;*
3. *if $\sigma(u_i \cdots u_n)$ has a prefix in S, then for all $1 \leq j \leq i$, $\sigma(u_j \cdots u_n)$ has a prefix in S.*

Proof. The proof is very similar to the one of Lemma 4.31, only the first claim slightly differs. We give below the main ideas for this claim.

We proceed by induction on $|u|$. Indeed, the first claim is satisfied for ε and for the letters. For $u = va$, we have $\sigma(u)w = \sigma(v)\sigma(a)w$ so it has the prefix $\sigma(v)w'$ for some $w' \in S$ therefore, by induction hypothesis, it has a prefix in S. The existence of N follows from the fact that σ is non-erasing and S is finite. $\square$

Using the previous lemma, we can sometimes remove some elements of S and keep a return morphism for S.

Remark 4.38. If σ is a return morphism for S, then it is a return morphism for $S' = S \cap \sigma(\mathcal{A}^*)$. Indeed, S' is clearly a factor code and, for all $a \in \mathcal{A}$, $w \in S \cap \sigma(\mathcal{A}^*)$, $\sigma(a)w$ contains exactly two occurrences of elements of S, one as a prefix and one as a suffix. The previous lemma shows that, since $w \in \sigma(\mathcal{A}^*)$, there exists $u \in \mathcal{A}^*$ such that w is a prefix of $\sigma(u)$. Therefore, $\sigma(a)w \in \mathrm{Pref}(\sigma(au)) \subseteq \sigma(\mathcal{A}^*)$. This shows that the factors of $\sigma(a)w$ in S are in fact in S'. Observe that we also have $S \cap \sigma(\mathcal{A}^*) = S \cap \cup_{u \in \mathcal{A}^*} \mathrm{Pref}(\sigma(u))$.

Using Lemma 4.37, we deduce the following result which motivates the terminology, as in the case of a return morphism for a word.

Proposition 4.39. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism and let $S \subseteq \sigma(\mathcal{A}^*)$ be a factor code. If σ is injective on the letters and $\sigma(\mathcal{A})$ is a suffix code, then the following are equivalent:*

1. *σ is a return morphism for S;*
2. *for any language $\mathcal{L}$ over $\mathcal{A}$, $\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{L})}(S \cap \sigma(\mathcal{L}))$;*
3. *$\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{A}^*)}(S)$.*

Proof. Once again, the proof is very similar to the proof of Proposition 4.32. The only modification needed is to show that if $\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{A}^*)}(S)$ then σ is return morphism for S. Let $a \in \mathcal{A}$ and $w \in S$. Since $S \subseteq \sigma(\mathcal{A}^*)$, there exists $u \in \mathcal{A}^*$ such that $w \in \mathrm{Pref}(\sigma(u))$ by Lemma 4.37. Therefore $\sigma(a)w \in \sigma(\mathcal{A}^*)$ and, as $\sigma(\mathcal{A}^*)$ is recurrent, $\sigma(a)w$ is suffix comparable with a complete return word for S in $\sigma(\mathcal{A}^*)$. Since $\sigma(\mathcal{A})$ is a suffix code, this shows that $\sigma(a)w$ is a complete return word for S, i.e., it contains exactly two occurrences of elements of S, one as a prefix and one as a suffix. As it is true for any $a \in \mathcal{A}$ and $w \in S$, we conclude that σ is a return morphism for S. $\square$

Observe the difference with Proposition 4.32 in the third assertion. Indeed, the existence of a language $\mathcal{L}' \subseteq \mathcal{B}^*$ such that $\sigma(\mathcal{A}) = R_{\mathcal{L}'}(S)$ is not sufficient anymore for σ to be a return morphism for S. This can be seen in the proof as we would need that $\sigma(a)w \in \mathcal{L}'$ for all $a \in \mathcal{A}$, $w \in S$. This is confirmed by the following example.

Example 4.40. Let σ be the morphism such that $\sigma(0) = 0$, $\sigma(1) = 010$ and $\sigma(2) = 2$. As $\sigma(\mathcal{A})$ is not a suffix code, σ is not a return morphism for a factor set. However, let $\mathcal{L}$ be a language over $\{0, 1, 2\}$ such that $\mathcal{L}_2 = \{00, 01, 12, 20, 22\}$. Then $\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{L})}(\{00, 01, 2\})$. Indeed, it is clear that 01 (resp., 2) occurs only (and exactly) as the prefix of $\sigma(1)$ (resp., $\sigma(2)$) in $\sigma(\mathcal{L})$. Moreover, by definition of $\mathcal{L}$, 00 occurs only and exactly as the prefix of $\sigma(0a)$, $a \in E^R_{\mathcal{L}}(0)$. This shows that the image of any length-2 element of $\mathcal{L}$ has a prefix in $S = \{00, 01, 2\}$ and the elements of S occur only as prefixes of the image of some word. We conclude that $\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{L})}(\{00, 01, 2\})$.

Just as for return morphisms for a word, the factor code for which σ is a return morphism is not always unique. While the link between such factor codes is not as simple as in Proposition 4.33, we still have the following result.

Proposition 4.41. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for two factor codes $S, S' \subseteq \sigma(\mathcal{A}^*)$. Then,*

1. *for all $w \in S$, either it has a (unique) prefix in S', or it is a proper prefix of one or more elements of S' (and not both);*

2. *σ is a return morphism for*

$$S'' = (S \cap \mathrm{Pref}(S')) \cup (\mathrm{Pref}(S) \cap S');$$

3. *$\#S'' \leq \min\{\#S, \#S'\}$ and*

$$\sum_{w \in S''} |w| \leq \min\left\{\sum_{w \in S} |w|, \sum_{w \in S'} |w|\right\}.$$

Proof. Let $w \in S$. By Lemma 4.37 for S, there exists $u \in \mathcal{A}^*$ such that w is a prefix of $\sigma(u)$. By Lemma 4.37 for S' now, $\sigma(u)$ is prefix comparable with $w' \in S'$. This shows that w is prefix comparable with w'. The first claim then follows since S' is a factor code so, in particular, a prefix code.

Let

$$S'' = (S \cap \mathrm{Pref}(S')) \cup (\mathrm{Pref}(S) \cap S').$$

We first show that it is a factor code. It is clear that $S \cap \mathrm{Pref}(S')$ and $\mathrm{Pref}(S) \cap S'$ are both factor codes. Let $w \in S \cap \mathrm{Pref}(S')$ and $w' \in \mathrm{Pref}(S) \cap S'$. As $w \in S$ and $w' \in \mathrm{Pref}(S)$, w cannot be a proper factor of w' since S is a factor code. Similarly, w' cannot be a proper factor of w. This shows that S'' is a factor code.

Let us now prove that σ is a return morphism for S''. In other words, we need to show that, for all $a \in \mathcal{A}$ and all $w \in S''$, $\sigma(a)w$ contains exactly two words in S'', one as a prefix and one as a suffix. Assume that $w \in S \cap \mathrm{Pref}(S')$, the other case is symmetric. Let $w' \in S'$ be such that $w \in \mathrm{Pref}(w')$. Since $\sigma(a)w$ is a prefix of $\sigma(a)w'$ and σ is a return morphism for S and S', $\sigma(a)w$ has no internal factor in S nor in S' so it has no internal factor in S''. Let us show that it has a prefix in S''. As σ is a return morphism for S, $\sigma(a)w$ has a prefix w'' in S. If $w'' \notin S''$, then w'' has a prefix in S' by the first claim so $\sigma(a)w$ has a prefix in $\mathrm{Pref}(S) \cap S' \subseteq S''$. This ends the proof that σ is a return morphism for S''.

Observe that $S'' \subseteq \mathrm{Pref}(S)$ and, as S'' is a factor code (and in particular a prefix code), it contains at most one prefix of each $w \in S$. This shows that $\#S'' \leq \#S$ and $\sum_{w \in S''} |w| \leq \sum_{w \in S} |w|$. The proof for S' is similar. □

This proposition allows to single out a particular factor code associated with a return morphism. Indeed, if we define the partial order $\preceq$ on subsets of $\mathcal{B}^*$ as follows:

$$S \preceq S' \iff S \subseteq \mathrm{Pref}(S'),$$

then the previous result tells us that the set

$$\{S \subseteq \mathcal{B}^* : S \text{ is a factor code and } \sigma \text{ is a return morphism for } S\}$$

admits a global minimum S for this order. This factor code S also minimizes $\#S$ and $\sum_{w \in S} |w|$. In particular, $\#S = 1$ if and only if σ is a return morphism for a word.

4.3.3 Tame morphisms

As seen before, a return morphism is such that the images of the letters correspond to return words in the image. Using results on the number of return words obtained in Chapter 3 (Corollary 3.32), we obtain restrictions on the return morphisms whose images can be dendric.

Corollary 4.42. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S. If there exists a language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a recurrent neutral language, then*

$$\#\mathcal{A} \leq \#S + \#\mathcal{B} - 1.$$

In particular, if $S = \{w\}$, then $\#\mathcal{A} = \#\mathcal{B}$.

If we restrict ourselves to return morphism for a word, then we can obtain even more conditions. Indeed, as recalled in Chapter 3, if σ is a return morphism for a word and there exists $\mathcal{L}$ such that $\sigma(\mathcal{L})$ is recurrent dendric, then $\sigma(\mathcal{A})$ forms a basis of the free group over $\mathcal{B}$. It is furthermore a tame basis as proved in [BDFD$^+$15d].

Definition 4.43. Let $a, b \in \mathcal{A}$ distinct. We define the two following morphisms

$$L_{a,b}\colon \begin{cases} a \mapsto ba & \\ c \mapsto c & \forall c \in \mathcal{A} \setminus \{a\} \end{cases} \qquad R_{a,b}\colon \begin{cases} a \mapsto ab & \\ c \mapsto c & \forall c \in \mathcal{A} \setminus \{a\} \end{cases}.$$

The *elementary (auto)morphism* on $\mathcal{A}$ are the permutations and the morphisms defined above for all distinct $a, b \in \mathcal{A}$. A morphism is *tame* if it is a composition of elementary morphisms, and a base S of the free group over $\mathcal{A}$ is *tame* if there exists a tame morphism such that $S = \sigma(\mathcal{A})$.

The definition of the morphisms above should not be confused with the definition of Arnoux-Rauzy morphisms. They are however related as follows:

$$L_\ell = L_{a_1,\ell} \circ L_{a_2,\ell} \cdots \circ L_{a_k,\ell}, \quad R_\ell = R_{a_1,\ell} \circ R_{a_2,\ell} \cdots \circ R_{a_k,\ell}$$

if $\mathcal{A} = \{\ell, a_1, a_2, \ldots, a_k\}$.

As announced, we have the following result.

Proposition 4.44. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for a word. If there exists a language $\mathcal{L}$ such that $\sigma(\mathcal{L})$ is a recurrent dendric language, then σ is tame.*

Observe that $L_{a,b}$ is a return morphism for the set $\mathcal{A} \setminus \{a\}$, and $R_{a,b}$ can be seen as a right return morphism for the same set if we adapt all the definitions. This suggests that elementary morphisms are somehow linked to return morphisms for sets of letters. Indeed, we have the following result.

Proposition 4.45. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism where $\mathcal{B}$ is minimal (i.e., all the letters of $\mathcal{B}$ appear in at least one image). If $\sigma = \gamma \circ \tau$ where γ is an elementary morphism, then σ is a return morphism for $\mathcal{C} \subseteq \mathcal{B}$ if and only if we are in one of the following cases:*

1. *γ is a permutation on $\mathcal{B}$ and τ is a return morphism for $\gamma^{-1}(\mathcal{C})$;*
2. *$\gamma = L_{a,b}$ for $a \in \mathcal{B} \setminus \mathcal{C}$, $b \in \mathcal{C}$ and τ is a return morphism for $\mathcal{C} \cup \{a\}$;*
3. *$\gamma = L_{a,b}$ for distinct $a, b \in \mathcal{B} \setminus \mathcal{C}$ and τ is a return morphism for $\mathcal{C}$;*
4. *$\gamma = R_{a,b}$ for distinct $a \in \mathcal{B}$, $b \in \mathcal{B} \setminus \mathcal{C}$ and τ is a return morphism for $\mathcal{C}$.*

Proof. It is clear that τ is injective on the letters if and only if σ is. Moreover, as we are looking at sets of letters, σ is a return morphism for $\mathcal{C}$ if and only if $\sigma(a) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$ for all $a \in \mathcal{A}$ (and the same for τ with the corresponding set of letters). Let us consider the three cases for γ: permutation, $L_{a,b}$ and $R_{a,b}$ and characterize, for each case, when $\sigma(a) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$ for all $a \in \mathcal{A}$.

1. Assume that γ is a permutation over $\mathcal{B}$. It is clear that, for all $a \in \mathcal{A}$, $\sigma(a) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$ if and only if $\tau(a) \in \gamma^{-1}(\mathcal{C})(\mathcal{B} \setminus \gamma^{-1}(\mathcal{C}))^*$ so the conclusion is direct.
2. Assume that $\gamma = L_{a,b}$ for distinct $a, b \in \mathcal{B}$. Observe that, if σ is a return morphism for $\mathcal{C}$ then, since the images under σ do not begin with a and $\mathcal{B}$ is minimal, a appears as a non-prefix factor in some

image so $a \notin \mathcal{C}$. In what follows we therefore restrict ourselves to the case where $a \notin \mathcal{C}$.

If $b \notin \mathcal{C}$ then, for all $c \in \mathcal{A}$, we directly have $\sigma(c) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$ if and only if $\tau(c) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$. We now consider the case where $b \in \mathcal{C}$. Observe that $\sigma(c)$ contains a non-prefix occurrence of b if and only if $\tau(c)$ contains a non-prefix occurrence of a or b. Therefore, for all $c \in \mathcal{A}$, $\sigma(c) \in \mathcal{C}(\mathcal{B}\setminus\mathcal{C})^*$ if and only if $\tau(c) \in \mathcal{C}'(\mathcal{B}\setminus\mathcal{C}')^*$ for $\mathcal{C}' = \mathcal{C} \cup \{a\}$.

3. Assume now that $\gamma = R_{a,b}$ for distinct $a, b \in \mathcal{B}$. If σ is a return morphism for $\mathcal{C}$, then $b \notin \mathcal{C}$ since $\mathcal{B}$ is minimal. Assuming that $b \notin \mathcal{C}$, we directly have $\sigma(c) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$ if and only if $\tau(c) \in \mathcal{C}(\mathcal{B} \setminus \mathcal{C})^*$.

$\square$

The previous result can for example be used to define an automaton which will accept the elementary decompositions of tame return morphisms for some given set of letters.

Corollary 4.46. *Let $\mathcal{A}$ be an alphabet, $\mathcal{B} \subseteq \mathcal{A}$ be non-empty and let $\mathfrak{A}$ be the automaton defined as follows:*

- *the states are the non-empty subsets of $\mathcal{A}$;*
- *the initial state is $\mathcal{A}$;*
- *the final state is $\mathcal{B}$;*
- *for any state $\mathcal{C}$, the transitions leaving it are*
  - $(\mathcal{C}, \gamma, \gamma(\mathcal{C}))$ *for any permutation γ of $\mathcal{A}$;*
  - $(\mathcal{C}, L_{a,b}, \mathcal{C} \setminus \{a\})$ *for any distinct $a, b \in \mathcal{C}$;*
  - $(\mathcal{C}, L_{a,b}, \mathcal{C})$ *for any distinct $a, b \in \mathcal{A} \setminus \mathcal{C}$;*
  - $(\mathcal{C}, R_{a,b}, \mathcal{C})$ *for any distinct $a \in \mathcal{A}$, $b \in \mathcal{A} \setminus \mathcal{C}$.*

Then $\mathfrak{A}$ accepts exactly the elementary decompositions of the tame return morphisms for $\mathcal{B}$, i.e., $\mathfrak{A}$ accepts

$$\{\sigma_1 \cdots \sigma_k : \sigma_i \text{ elementary } \forall\, i \text{ and } \sigma_k \circ \ldots \circ \sigma_1 \text{ return morphism for } \mathcal{B}\}.$$

Observe that the number of elements in a state cannot increase when following a transition. In particular, this shows that we can remove all the states $\mathcal{C}$ such that $\#\mathcal{C} < \#\mathcal{B}$ without changing the language.

If we do not worry about permutations and only allow them as the last morphism, then we can get an even simpler automaton where the non-final states are of the form $\{a_1, \ldots, a_i\}$, $1 \leq i \leq n$, if $\mathcal{A} = \{a_1, \ldots, a_n\}$. This automaton is represented in Figure 4.2 for $\mathcal{A} = \{0, 1, 2\}$ and for $\mathcal{B} = \{1\}$.

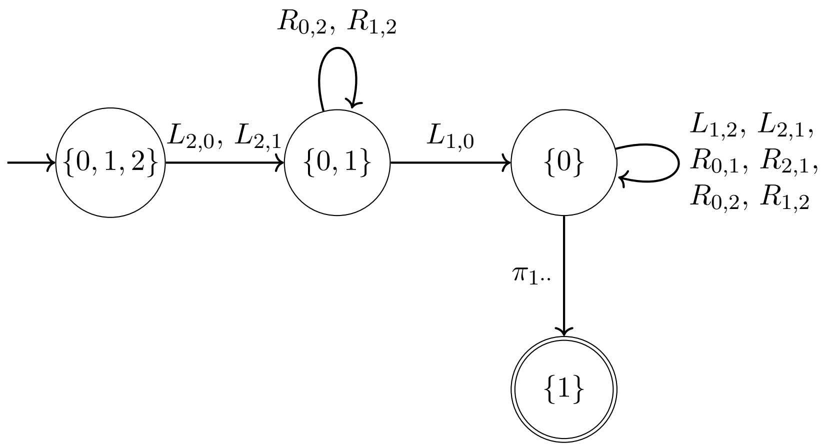


Figure 4.2: Automaton accepting the sequences $(\sigma_1, \ldots, \sigma_n, \pi)$ of elementary morphisms on $\{0, 1, 2\}$ where only π is a permutation and such that $\pi \circ \sigma_n \circ \ldots \circ \sigma_1$ is a return morphism for $\{1\}$, if $\pi_{i_0 i_1 i_2}$ is the morphism such that $\pi_{i_0 i_1 i_2}(j) = i_j$ and the dots are place-holders that can represent any value in $\{0, 1, 2\}$ giving a permutation.

4.4 Dendric images under return morphisms

Due to the properties of return morphisms proved in the previous section, we can easily describe the extensions of any word in the image of a language under a return morphism. This allows us to characterize when the image of a dendric language under a return morphism is also dendric.

This characterization was first stated for return morphisms for a letter in [GLL22] then was generalized in [GL22] to return morphisms for a word. In this second paper, we also showed a simplified characterization using the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$. We extend here some of these results to return morphisms for a factor code.

When looking at dendricity of the image under a return morphism, we distinguish two types of words: those who do not have a factor in S and those who do. The first ones are called initial factors and are the object of Subsection 4.4.1 while the second ones are called extended images for reasons that will become clear later on. We study the behavior of extensions of extended images in Subsection 4.4.2. We then present the global results and characterization in Subsection 4.4.3. When looking at return morphisms for a word, some of the statements can be simplified. We state these results in

Subsection 4.4.4 and also show a similar simplification for return morphisms for a set of letters in Subsection 4.4.5.

4.4.1 Initial factors

Given a language $\mathcal{L}$ and a return morphism σ, our goal is now to understand the extensions of the elements of $\sigma(\mathcal{L})$. As often, we begin our study with what can be considered, in some sense, as the small words.

Definition 4.47. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S and let $\mathcal{L}$ be a language over $\mathcal{A}$. A word $u \in \sigma(\mathcal{L})$ is an *initial factor* if $\mathrm{Fac}(u) \cap S = \emptyset$.

Observe that, since σ is a return morphism for S, any word of length at least $\|\sigma\| + \max_{w \in S} |w|$ in $\sigma(\mathcal{L})$ has a factor in S. In particular, there are only finitely many initial factors in $\sigma(\mathcal{L})$. Hence why initial factors can be seen as small words. This also implies that their extensions can be understood by looking only in small images. This is described more precisely in the following result.

Proposition 4.48. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S and let $\mathcal{L}$ be a language over $\mathcal{A}$. If $u \in \sigma(\mathcal{L})$ is such that $\mathrm{Fac}(u) \cap S = \emptyset$ then $aub \in \sigma(\mathcal{L})$, $a, b \in \mathcal{B}$, if and only if there exist $c \in \mathcal{A}$ and $w \in S$ such that $\sigma(c)w \in \sigma(\mathcal{L})$ and $aub \in \mathrm{Fac}(\sigma(c)w)$.*

Proof. It is clear that if aub is a factor of $\sigma(c)w \in \sigma(\mathcal{L})$, then $aub \in \sigma(\mathcal{L})$. Let us show the converse and assume that $aub \in \sigma(\mathcal{L})$. Let $v = v_1 \cdots v_n \in \mathcal{L}$ be such that $aub \in \mathrm{Fac}(\sigma(v))$ and aub starts in $\sigma(v_1)$. Up to extending v on the right, we can assume that $\sigma(v_2 \cdots v_n)$ has a prefix $w \in S$ by Lemma 4.37. Since w is not a factor of u by assumption on u, this implies that aub is a factor of $\sigma(v_1)w \in \sigma(\mathcal{L})$, which ends the proof. □

Example 4.49. Let σ be the morphism such that $\sigma(0) = 0$, $\sigma(1) = 0102$ and $\sigma(2) = 012$. It is a return morphism for $S = \{00, 01\}$ (see Example 4.35). Let $\mathcal{L}$ be the Tribonacci language. Since 1 and 2 are always followed by 0 in $\mathcal{L}$, the elements of $\sigma(\{0,1,2\})S \cap \sigma(\mathcal{L})$ are then $000, 001, 010200, 01200$. Therefore, the initial factors are ε, 0, 1, 2, 02, 10, 12, 20, 020, 102, 120 and 1020. We easily check that 1 is not left special and 2, 20 are not right special in $\sigma(\mathcal{L})$. Therefore, the only bispecial initial factors are ε and 0. Their extension graphs are represented in Figure 4.3. This shows that $\sigma(\mathcal{L})$ is not dendric.

In the particular case where σ is a return morphism for a word w, then for any language $\mathcal{L}$ over $\mathcal{A}$ and for all $c \in \mathcal{A}$, we have $\sigma(c)w \in \sigma(\mathcal{L})$. This

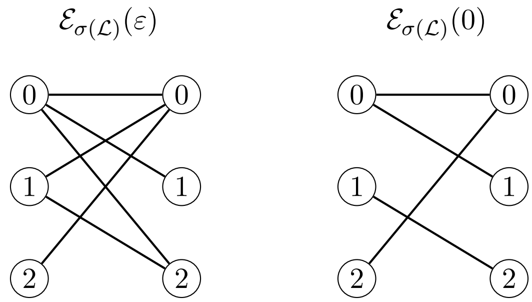


Figure 4.3: The extension graphs of ε (on the left) and 0 (on the right) in the image $\sigma(\mathcal{L})$ of the Tribonacci language under the morphism σ of Example 4.49.

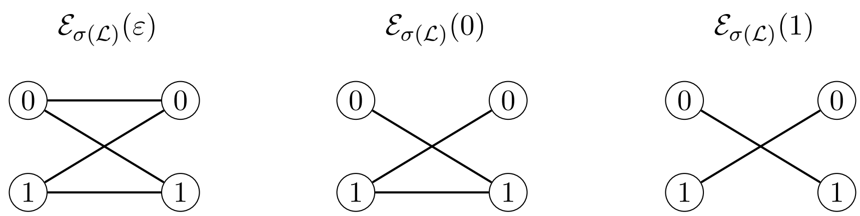


Figure 4.4: The extension graphs of ε (on the left), 0 (in the center) and 1 (on the right) in the image of any language $\mathcal{L}$ under the morphism σ of Example 4.50.

implies by Proposition 4.48 that the initial factors and their extensions only depend on σ and not on the language $\mathcal{L}$. This observation will be used in Subsection 4.4.4.

Example 4.50. Let σ be the morphism such that $\sigma(0) = 0110$ and $\sigma(1) = 011$. It is a return morphism for 01 so, no matter the language $\mathcal{L}$ over $\{0, 1\}$, the initial factors are the internal factors of 011001 and of 01101. Since 01 is the only left special word of length 2, the only bispecial initial words are ε, 0 and 1. Their extension graphs are represented in Figure 4.4. In particular, ε and 1 are the only non dendric initial factors.

4.4.2 Extended images

We now turn to the extensions of the other words, i.e., those having a factor in S. We also obtain a description of their extensions in the following proposition. This result relies heavily on the fact that σ is a return morphism and of the properties shown in Section 4.3.

Proposition 4.51. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S and let $\mathcal{L}$ be a language over $\mathcal{A}$. If $u \in \sigma(\mathcal{L})$ has a factor in S, then there exists a unique triplet $(s, v, p) \in \mathrm{Suff}^*(\sigma(\mathcal{A})) \times \mathcal{L} \times (S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S))$ such that $u = s\sigma(v)p$. Moreover, for all $a, b \in \mathcal{B}$, we have $aub \in \sigma(\mathcal{L})$ if and only if there exist $c, d \in \mathcal{A}$ and $w \in S$ such that $\sigma(cvd)w \in \sigma(\mathcal{L})$, $as \in \mathrm{Suff}(\sigma(c))$ and $pb \in \mathrm{Pref}(\sigma(d)w)$.*

Proof. Let us prove the existence and uniqueness of the triplet (s, v, p) at the same time. If $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$, then p has a prefix in S but no other factor in S. Therefore, if p is a suffix of u, it is precisely the shortest suffix of u having a prefix in S. It exists since u has a factor in S. Then, for any $v \in \mathcal{A}^*$, $\sigma(v)p$ has a prefix in S. In particular, if $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, then no occurrence of a word in S can start in s in $s\sigma(v)p$ since σ is a return morphism. This shows that if (s, v, p) is a triplet for u, then $\sigma(v)p$ must be the longest suffix of u having a prefix in S. By injectivity of σ, we conclude that, if it exists, the triplet is unique. Moreover, since $u \in \sigma(\mathcal{L})$, by Lemma 4.37, there exists $v \in \mathcal{L}$ such that $\sigma(v)p$ is the longest suffix of u having a prefix in S. This shows that the triplet (s, v, p) exists and is unique for $u \in \sigma(\mathcal{L})$.

We now look at the extensions of u. If $aub \in \sigma(\mathcal{L})$, then there exists a covering (v', k) of aub (see Definition 3.8 of coverings). By Lemma 4.37, there exists a non-empty suffix t of v' such that the occurrence of pb in $\sigma(v')$ is a prefix of $\sigma(t)$. By injectivity of σ, vt is then a suffix of v'. Let x be such that $v' = xvt$. Then as is a suffix of $\sigma(x)$ so, by definition of s and by definition of a covering, $x = c \in \mathcal{A}$. As t is not empty, let us denote d its first letter and t' such that $t = dt'$. By Lemma 4.37, we can extend t' on the right so that $cvdt'$ is still in $\mathcal{L}$ and $\sigma(t')$ has a prefix $w \in S$. By definition of p, pb contains no internal occurrences of elements of S so pb is in fact a prefix of $\sigma(d)w$. By construction, we also have $\sigma(cvd)w \in \sigma(\mathcal{L})$ which shows that c, d, w satisfy the statement for a, b.

Conversely, assume that there exist $c, d \in \mathcal{A}$ and $w \in S$ such that $\sigma(cvd)w \in \sigma(\mathcal{L})$, $as \in \mathrm{Suff}(\sigma(c))$ and $pb \in \mathrm{Pref}(\sigma(d)w)$. Then, we directly conclude that $aub = as\sigma(v)pb \in \sigma(\mathcal{L})$. □

Example 4.52. Let σ be the return morphism for $S = \{00, 01\}$ defined in Example 4.49 by $\sigma(0) = 0$, $\sigma(1) = 0102$ and $\sigma(2) = 012$ and let $\mathcal{L}$ be the Tribonacci language. Since $2010 \in \mathcal{L}$, the word $u = 2001020$ is in $\sigma(\mathcal{L})$. The triplet (s, v, p) associated with u can be obtained as follows: p is the shortest suffix of u having a prefix in S so $p = 01020$, $\sigma(v)p$ is the longest suffix of u having a prefix in S so $\sigma(v)p = 001020$ and $v = 0$, s is such that $u = s\sigma(v)p$ so $s = 2$.

Based on this result, we introduce the following terminology.

Definition 4.53. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S and let $\mathcal{L}$ be a language over $\mathcal{A}$. A word $u \in \sigma(\mathcal{L})$ is an *extended image* of $v \in \mathcal{L}$ if there exist $s \in \mathcal{B}^*$, $p \in S\mathcal{B}^*$ such that (s, v, p) is the triplet of Proposition 4.51 for u.

Proposition 4.51 essentially states that, if u is an extended image of v, we can obtain the extensions of u by doing some operations on well-chosen generalized extensions of v. In the rest of this subsection, we introduce some notations to precisely describe the extensions of v of interest and the operations on them.

The first step is to notice that if $c, d \in \mathcal{A}$ and $w \in S$ are such that $\sigma(cvd)w \in \sigma(\mathcal{L})$, then c itself is a left extension of v (since σ is a return morphism). However, we need to associate a particular (generalized) right extension of v to d and w. This is the purpose of the following notation.

Recall that if $(v, k) \in \mathcal{A}^+ \times \mathbb{N}$ is a covering of a word w, then an occurrence of w starts in $\sigma(v_1)$ after k letters. Therefore, by Lemma 4.37, if σ is a return morphism for S and $w \in S$, then for any covering (v, k) of w, we have $k = 0$. In this case, we will abusively say that v itself is a *covering* of w and denote $\mathcal{C}_{\mathcal{L}}(w)$ the set of such words v that are in $\mathcal{L}$. We also denote

$$\mathcal{C}_{\mathcal{L}}(S) = \bigcup_{w \in S} \mathcal{C}_{\mathcal{L}}(w).$$

Example 4.54. Let σ be the return morphism for $S = \{00, 01\}$ of Examples 4.49 and 4.52 (i.e., $\sigma(0) = 0$, $\sigma(1) = 0102$ and $\sigma(2) = 012$). We have $\mathcal{C}_{\mathcal{L}}(01) = \{1, 2\}$ independently of the language $\mathcal{L}$ over $\{0, 1, 2\}$, and $\mathcal{C}_{\mathcal{L}}(00) = \{00, 01, 02\} \cap \mathcal{L}$. So, if $\mathcal{L}$ is the Tribonacci language, then $\mathcal{C}_{\mathcal{L}}(S) = \{00, 01, 02, 1, 2\}$.

The following result is a direct consequence of Lemma 4.37 on the properties of return morphisms and of the fact that S is a prefix code.

Lemma 4.55. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S and let $\mathcal{L}$ be a language over $\mathcal{A}$. Then*

1. *$\mathcal{C}_{\mathcal{L}}(w) \cap \mathcal{C}_{\mathcal{L}}(w') = \emptyset$ for all distinct $w, w' \in S$;*

2. *$\mathcal{C}_\mathcal{L}(S)$ is an $\mathcal{L}$-maximal prefix code and therefore, $\mathcal{AC}_\mathcal{L}(S)$ also is;*

3. *for any non-empty $u \in \mathcal{L}$, $\sigma(u)w \in \sigma(\mathcal{L})$ if and only if there exists $v \in \mathcal{C}_\mathcal{L}(w)$ such that $uv \in \mathcal{L}$;*

4. *for all $c, d \in \mathcal{A}$, $v \in \mathcal{L}$ and $p' \in \mathrm{Pref}(\sigma(\mathcal{A})S)$, there exists $w \in S$ such that $\sigma(cvd)w \in \sigma(\mathcal{L})$ and $p' \in \mathrm{Pref}(\sigma(d)w)$ if and only if there exists $t \in \mathcal{C}_\mathcal{L}(S)$ such that $cvdt \in \mathcal{L}$ and $p' \in \mathrm{Pref}(\sigma(dt))$.*

In other words, using Proposition 4.51 and the last claim of Lemma 4.55, if u is an extended image of v, then the extensions of u can be obtained from the generalized extensions in $\mathcal{A} \times \mathcal{AC}_\mathcal{L}(S)$ of v. To go from $E_{\mathcal{L},\mathcal{A},\mathcal{AC}_\mathcal{L}(S)}(v)$ to $E_{\sigma(\mathcal{L})}(u)$, we can then use the following maps.

Definition 4.56. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a morphism. For all $s \in \mathcal{B}^*$, we define the partial map

$$\varphi^L_{\sigma,s}\colon \mathcal{A}^* \to \mathcal{B}, \quad v \mapsto a \text{ if } as \in \mathrm{Suff}(\sigma(v)).$$

It is therefore only defined on $\mathrm{dom}(\varphi^L_{\sigma,s}) = \{v \in \mathcal{A}^* : s \in \mathrm{Suff}^*(\sigma(v))\}$. Similarly, for all $p \in \mathcal{B}^*$, we define

$$\varphi^R_{\sigma,p}\colon \mathcal{A}^* \to \mathcal{B}, \quad v \mapsto b \text{ if } pb \in \mathrm{Pref}(\sigma(v))$$

which is defined on $\mathrm{dom}(\varphi^R_{\sigma,p}) = \{v \in \mathcal{A}^* : p \in \mathrm{Pref}^*(\sigma(v))\}$. If the context is clear, we will drop the subscript σ.

Using these notations, Proposition 4.51 can be restated as follows.

Proposition 4.57. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $u \in \sigma(\mathcal{L})$ have a factor in S. If (s, v, p) is the triplet of Proposition 4.51, then*

$$E_{\sigma(\mathcal{L})}(u) = (\varphi^L_s \times \varphi^R_p) E_{\mathcal{L},\mathcal{A},\mathcal{AC}_\mathcal{L}(S)}(v). \tag{4.1}$$

Proof. Indeed, Proposition 4.51 and Lemma 4.55 imply that $aub \in \sigma(\mathcal{L})$ if and only if there exists $(c, dt) \in E_{\mathcal{L},\mathcal{A},\mathcal{AC}_\mathcal{L}(S)}(v)$ such that $as \in \mathrm{Suff}(\sigma(c))$ and $pb \in \mathrm{Pref}(\sigma(dt))$. We conclude by definition of φ^L_s and φ^R_p. □

Example 4.58. Let σ be the return morphism for $S = \{00, 01\}$ defined by $\sigma(0) = 0$, $\sigma(1) = 0102$ and $\sigma(2) = 012$ and let $\mathcal{L}$ be the Tribonacci language. By Example 4.54, we know that $\mathcal{C}_\mathcal{L}(S) = \{00, 01, 02, 1, 2\}$. Let us describe the extensions of $u = 2001020$. By Example 4.52, the associated triplet (s, v, p) is $(2, 0, 01020)$ so the extensions of u depend on

$$E_{\mathcal{L},\mathcal{A},\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)}(0) = \{(0, 102), (1, 01), (1, 102), (1, 201), (2, 100), (2, 101), (2, 102)\}.$$

Since $\operatorname{dom}(\varphi_2^L) = \mathcal{A}^*\{1, 2\}$ and $\operatorname{dom}(\varphi_{01020}^R) = 1\mathcal{A}^+$, we obtain that

$$E_{\sigma(\mathcal{L})}(2001020) = \{(0, 0), (1, 0)\}.$$

We now take the viewpoint of extension graphs. Indeed, Equation (4.1) states that $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ can be obtained from $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)}(v)$ by doing some manipulations. We decompose these using two intermediary graphs.

Observe that if $t, t' \in \mathcal{C}_{\mathcal{L}}(w)$ for some $w \in S$ and if $p \in \operatorname{Pref}^*(\sigma(\mathcal{A})S)$, then for all $d \in \mathcal{A}$, $\varphi_p^R(dt)$ and $\varphi_p^R(dt')$ are either not defined or they both exist and are equal. Indeed, if $\varphi_p^R(dt) = b$, then $pb \in \operatorname{Pref}(\sigma(\mathcal{A})w)$ so pb is also a prefix of $\sigma(dt')$ and $\varphi_p^R(dt') = b$. This motivates the following notation.

Definition 4.59. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. The set $E_{\mathcal{L},\sigma}(v)$ denotes the bi-extensions of v in $\mathcal{A}$ on the left and $\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)$ on the right where we identify the elements of $a\mathcal{C}_{\mathcal{L}}(w)$ on the right for all $a \in \mathcal{A}$, $w \in S$. More precisely, if we define the equivalence relation $\equiv$ on $\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)$ by

$$t \equiv t' \iff t_1 = t'_1 \text{ and } \exists w \in S \text{ st. } t_2 \cdots t_{|t|}, t'_2 \cdots t'_{|t'|} \in \mathcal{C}_{\mathcal{L}}(w),$$

then

$$E_{\mathcal{L},\sigma}(v) = \{(a, [t]_\equiv) : (a, t) \in E_{\mathcal{L},\mathcal{A},\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)}(v)\}.$$

We then denote $\mathcal{E}_{\mathcal{L},\sigma}(v)$ the bipartite graph generated by the edges in $E_{\mathcal{L},\sigma}(v)$.

In other words, $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is the image of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)}(v)$ under the graph morphism induced by the quotient map $q_\equiv\colon t \mapsto [t]_\equiv$ acting on the right vertices.

As explained above, if $p \in \operatorname{Pref}^*(\sigma(\mathcal{A})S)$, we can extend φ_p^R by defining, for all $t \in \mathcal{A}\mathcal{C}_{\mathcal{L}}(S)$, $\varphi_p^R([t]_\equiv) = \varphi_p^R(t)$ since this does not depend on the choice of $t \in [t]_\equiv$. Equation (4.1) then implies that

$$E_{\sigma(\mathcal{L})}(u) = (\varphi_s^L \times \varphi_p^R) E_{\mathcal{L},\sigma}(v).$$

Therefore, $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is the first intermediary graph. To simplify notations, we will often abusively assimilate $[t]_\equiv$ with any of its elements since we are mostly interested in images under φ_p^R.

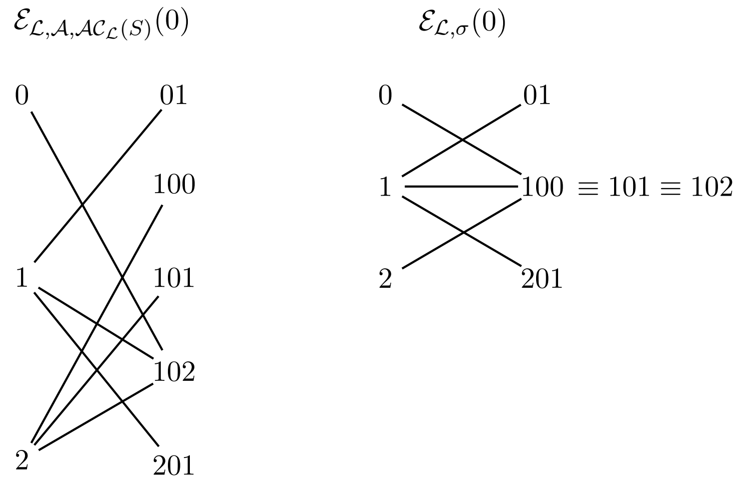


Figure 4.5: The graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_S(\mathcal{L})}(0)$ (on the left) corresponding to Example 4.60, and its image $\mathcal{E}_{\mathcal{L},\sigma}(0)$ (on the right) under the graph morphism identifying equivalent vertices in $\mathcal{AC}_S(\mathcal{L})$.

Example 4.60. Let σ be the return morphism defined by $\sigma(0) = 0$, $\sigma(1) = 0102$ and $\sigma(2) = 012$ and let $\mathcal{L}$ be the Tribonacci language. Using Example 4.58, we obtain the graphs $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(0)$ and $\mathcal{E}_{\mathcal{L},\sigma}(0)$ represented in Figure 4.5.

If σ is a return morphism for a word, the graph $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is not really new as explained below.

Remark 4.61. If $S = \{w\}$, then $\mathcal{C}_{\mathcal{L}}(S) = \mathcal{C}_{\mathcal{L}}(w)$ so, for all $t, t' \in \mathcal{AC}_{\mathcal{L}}(S)$, they are equivalent if and only if they begin with the same letter. Therefore, in this case, we have $E_{\mathcal{L},\sigma}(v) \cong E_{\mathcal{L}}(v)$ and $\mathcal{E}_{\mathcal{L},\sigma}(v) \cong \mathcal{E}_{\mathcal{L}}(v)$.

Since φ_s^L and φ_p^R are partial maps, when computing $(\varphi_s^L \times \varphi_p^R)E_{\mathcal{L},\sigma}(v)$, we first need to remove the elements of $E_{\mathcal{L},\sigma}(v)$ which are not in the domain of $\varphi_s^L \times \varphi_p^R$. To make this step clearer, we use the following notations.

Definition 4.62. Let G be a bipartite graph, E be its set of edges and U and V be two sets. We denote[2]

$$E^{U,V} = \{(u, v) \in E : u \in U, v \in V\}$$

and $G^{U,V}$ the subgraph of G generated by the edges in $E^{U,V}$.

[2]We assume here that, if $(u, v) \in E$, then u is a left vertex and v is a right vertex.

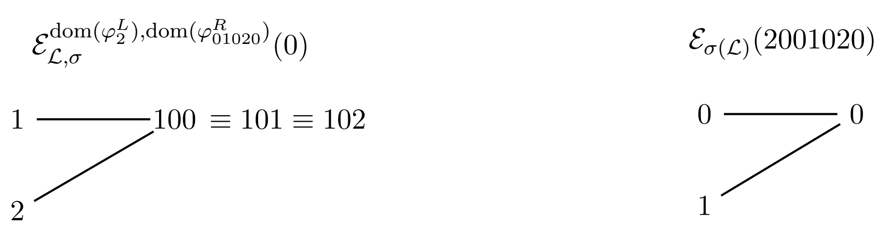


Figure 4.6: The graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_2^L),\mathrm{dom}(\varphi_{01020}^R)}(0)$ (on the left) corresponding to Example 4.60, and its image $\mathcal{E}_{\sigma(\mathcal{L})}(2001020)$ (on the right) under the graph morphism induced by φ_2^L acting on the left vertices and φ_{01020}^R acting on the right vertices.

We will combine this with other notations to write $E_{\mathcal{L},\sigma}^{U,V}(v)$ instead of $(E_{\mathcal{L},\sigma}(v))^{U,V}$ and $\mathcal{E}_{\mathcal{L},\sigma}^{U,V}(v)$ instead of $(\mathcal{E}_{\mathcal{L},\sigma}(v))^{U,V}$, for example.

It is clear that, for all $E \subseteq \mathcal{A}^* \times \mathcal{A}^*$, we have

$$(\varphi_s^L \times \varphi_p^R)E = (\varphi_s^L \times \varphi_p^R)E^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}$$

so Equation 4.1 now becomes

$$E_{\sigma(\mathcal{L})}(u) = (\varphi_s^L \times \varphi_p^R)E_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v).$$

Therefore, the second intermediary graph between $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ and $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$. We now have the following final statement for the extension graph of u.

Proposition 4.63. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $u \in \sigma(\mathcal{L})$ have a factor in S. If (s, v, p) is the corresponding triplet of Proposition 4.51, then $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is the image of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ under the graph morphism induced by φ_s^L acting on the left vertices and φ_p^R acting on the right vertices.*

Example 4.64. Let us continue Example 4.60. Starting from the graph $\mathcal{E}_{\mathcal{L},\sigma}(0)$ represented in Figure 4.5, we can obtain $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_2^L),\mathrm{dom}(\varphi_{01020}^R)}(0)$ by removing the left vertices ending with 0 and the right vertices that do not start with 1. The graphs $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_2^L),\mathrm{dom}(\varphi_{01020}^R)}(0)$ and $\mathcal{E}_{\sigma(\mathcal{L})}(2001020)$ are represented in Figure 4.6. In particular, we indeed obtain the same extensions for 2001020 as in Example 4.58

4.4.3 Dendric images

Using the results obtained in the previous subsections, we now look at the dendricity of the image. For the initial factors, by Proposition 4.48, their dendricity only depends on the words $\sigma(c)w$ for the pairs $(c, w) \in \mathcal{A} \times S$ for which there exists $t \in \mathcal{C}_{\mathcal{L}}(w)$ such that $ct \in \mathcal{L}$. This can then easily be checked.

We therefore focus on dendricity of the extended images in the remainder of this subsection. Instead of looking at one extended image at a time, we consider the set of extended images of a given word. We first give the following natural description of all the extended images of v.

Lemma 4.65. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. A word $u \in \sigma(\mathcal{L})$ is an extended image of v if and only if there exist $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$ such that $u = s\sigma(v)p$ and $E_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v) \neq \emptyset$.*

Proof. If u is an extended image, the existence of s and p directly follows from Propositions 4.51 and 4.63. For the converse, if $E_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v) \neq \emptyset$, then $u := s\sigma(v)p$ is in $\sigma(\mathcal{L})$. Moreover, by hypothesis on s and p, (s, v, p) is the triplet given by Proposition 4.51 so u is an extended image of v. □

We now look at the connectedness of the extended images.

Lemma 4.66. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. The extended images of v in $\sigma(\mathcal{L})$ are connected if and only if $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$.*

Proof. Let $u = s\sigma(v)p$ be an extended image of v. By Proposition 4.63, $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is the image of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ under the graph morphism induced by φ_s^L acting on the left vertices and φ_p^R acting on the right vertices. Therefore, if $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected, u is connected in $\sigma(\mathcal{L})$.

Let us prove the converse by contraposition so assume that there exist $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$ such that the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is not connected. Moreover, assume that among all such pairs (s, p), we have chosen one such that p is of maximal length and s is of maximal length for this p.

As $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is a bipartite graph with no isolated vertices since it is generated by its edges, this means that we can find a partition $\{A_1, A_2\}$

(resp., $\{B_1, B_2\}$) of the left (resp., right) vertices of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ into non-empty subsets such that the edges are all included in $(A_1 \times B_1) \cup (A_2 \times B_2)$.

By Lemma 4.65, (s, v, p) is the triplet corresponding to the extended image $u = s\sigma(v)p$ so $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is the image of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ under the graph morphism induced by φ_s^L acting on the left vertices and φ_p^R acting on the right vertices by Proposition 4.63. Therefore, the edges of $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ are included in $(\varphi_s^L(A_1) \times \varphi_p^R(B_1)) \cup (\varphi_s^L(A_2) \times \varphi_p^R(B_2))$, both products containing at least one edge.

However, $\varphi_p^R(B_1) \cap \varphi_p^R(B_2) = \emptyset$. Indeed, assume by contradiction that there exist $t \in B_1$, $t' \in B_2$ such that $\varphi_p^R(t) = \varphi_p^R(t') = b$, meaning by definition that $pb \in \mathrm{Pref}(\sigma(t)) \cap \mathrm{Pref}(\sigma(t'))$. As $p \in \mathrm{Pref}^*(\sigma(\mathcal{A})S)$ and pb is prefix of some image, we have $pb \in \mathrm{Pref}(\sigma(\mathcal{A})S)$. We cannot have $pb \in \sigma(\mathcal{A})S$ since, by definition of $\mathcal{E}_{\mathcal{L},\sigma}(v)$, there is at most one right vertex of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ whose image begins with $\sigma(a)w$ for all $a \in \mathcal{A}$, $w \in S$. Therefore, $pb \in \mathrm{Pref}^*(\sigma(\mathcal{A})S)$, and $pb \in \mathrm{Pref}^*(\sigma(t)) \cap \mathrm{Pref}^*(\sigma(t'))$. However, t and t' are then two right vertices of the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_{pb}^R)}(v)$ which is a subgraph of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$. This shows that $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_{pb}^R)}(v)$ is not connected and contradicts the maximality of p.

We can similarly show that $\varphi_s^L(A_1) \cap \varphi_s^L(A_2) = \emptyset$. Indeed, if $\varphi_s^L(c) = \varphi_s^L(c') = a$, $c \in A_1$, $c' \in A_2$, then as cannot be in $\sigma(\mathcal{A})$ since $\sigma(\mathcal{A})$ is a suffix code and σ is injective. Thus $as \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_{as}^L),\mathrm{dom}(\varphi_p^R)}(v)$ is not connected which contradicts the maximality of s.

This shows that $\{\varphi_s^L(A_1), \varphi_s^L(A_2)\}$ (resp., $\{\varphi_p^R(B_1), \varphi_p^R(B_2)\}$) is a partition of the left (resp., right) vertices of $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ and therefore, u is not connected in $\sigma(\mathcal{L})$. We conclude that the extended images of v are not all connected, which ends the contraposition. □

Using the same techniques we can also obtain one implication for the acyclicity of the extended images. We however also need an hypothesis on connectedness.

Lemma 4.67. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. If $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is a tree for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$, then the extended images of v are acyclic in $\sigma(\mathcal{L})$.*

Proof. Let $u = s\sigma(v)p$ be an extended image of v. By Proposition 4.63,

$\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is the image of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ under the graph morphism induced by φ_s^L acting on the left vertices and φ_p^R acting on the right vertices.

Let us show that, if $\varphi_p^R(t) = \varphi_p^R(t') = b$ for two right vertices t, t' of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$, then any path connecting them only uses right vertices t'' such that $\varphi_p^R(t'') = b$. As $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is a tree, there is a unique path connecting t and t'. Assume that this path uses a right vertex t'' such that $\varphi_p^R(t'') \neq b$. As in the proof of Lemma 4.66, we have $pb \in \mathrm{Pref}^*(\sigma(\mathcal{A})S)$ so t and t' are right vertices of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_{pb}^R)}(v)$ but they are not connected since t'' is not a right vertex of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_{pb}^R)}(v)$. This contradicts the fact that $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_{pb}^R)}(v)$ is connected.

Similarly, if $\varphi_s^L(c) = \varphi_s^L(c') = a$ for two left vertices c, c' of the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$, then the only path connecting them exclusively uses left vertices c'' such that $\varphi_s^L(c'') = a$. This shows that, if $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is acyclic, then so is its image $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ under the graph morphism induced by φ_s^L acting on the left vertices and φ_p^R acting on the right vertices. □

By combining these results, we obtain a first characterization of dendricity.

Proposition 4.68. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. If $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_w^R)}(v)$ is acyclic for all $w \in S$, then the extended images of v are dendric in $\sigma(\mathcal{L})$ if and only if the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$.*

Proof. If the extended images of v are dendric, then they are connected so, by Lemma 4.66, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$. This shows one of the implications.

Assume now that, for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected. If w is the prefix of p in S, then $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is a subgraph of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_w^R)}(v)$, which is acyclic by hypothesis. Therefore, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is a tree for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$. Using Lemmas 4.66 and 4.67, we conclude that the extended images of v are dendric. □

For this characterization however, we have an acyclicity hypothesis on $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_w^R)}(v)$ for all $w \in S$. It is implied by the acyclicity of $\mathcal{E}_{\mathcal{L},\sigma}(v)$ which we will now consider. So far, we do not have any specific hypothesis on the language $\mathcal{L}$ over $\mathcal{A}$. However, keep in mind that the original goal of this section is to characterize when the image of a dendric language is again dendric.

Observe that dendricity does not directly imply that the graph $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is acyclic, unless $S = \{w\}$. Indeed, $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is the image of the graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ under a graph morphism identifying some right vertices. Dendricity implies that $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ is acyclic (see Proposition 3.41) but the graph morphism could create cycles. We will therefore use the following lemma.

Lemma 4.69. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. If $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ is a tree and, for all $x \in \mathrm{Pref}^*(\mathcal{AC}_{\mathcal{L}}(S))$ and all $w \in S$, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_w^R)}(vx)$ is connected, then $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is acyclic.*

Proof. Recall that $q_\equiv$ is the quotient map which identifies $ct, c't' \in \mathcal{AC}_{\mathcal{L}}(S)$ if $c = c'$ and there exists $w \in S$ such that $t, t' \in \mathcal{C}_{\mathcal{L}}(w)$ (see Definition 4.59). By definition, $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is the image of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ under the graph morphism induced by $q_\equiv$ acting on the right vertices.

Assume that there exist two right vertices t, t' of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ such that $q_\equiv(t) = q_\equiv(t')$ and that the (unique) path connecting them in $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ uses one right vertex t'' such that $q_\equiv(t'') \neq q_\equiv(t)$. We show that this leads to a contradiction.

The graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ where we remove the right vertex t'' has at least two connected components, one containing t and another containing t'. Let us denote B_1 the right vertices connected to t in this graph, and B_1' the other right vertices. Let us also denote $b_1 = t_1 = t_1'$ and $w^{(1)} \in S$ such that $t_2 \cdots t_{|t|}, t_2' \cdots t_{|t'|}' \in \mathcal{C}_{\mathcal{L}}(w^{(1)})$.

In the graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}^{\mathcal{A},\mathrm{dom}(\varphi_{w^{(1)}}^R)}(vb_1)$, there is a partition $\{B_2, B_2'\}$ of the right vertices such that $u \in B_2$ (resp., $u \in B_2'$) if and only if $b_1 u$ has a prefix in B_1 (resp., B_1'). By definition of $w^{(1)}$, B_2 (resp., B_2') is not empty as it contains a vertex starting with $t_2 \cdots t_{|t|}$ (resp., $t_2' \cdots t_{|t'|}'$). Moreover, no vertex of B_2 is connected to a vertex of B_2' otherwise we have a corresponding path between vertices of B_1 and of B_1' in $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ and this path does not use t''. This would contradict the acyclicity of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$.

However, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_{w^{(1)}}^R)}(vb_1)$ is connected by hypothesis, and it is the image of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}^{\mathcal{A},\mathrm{dom}(\varphi^R_{w^{(1)}})}(vb_1)$ under the graph morphism induced by $q_\equiv$ on the right vertices. This implies that there exist $r \in B_2$, $r' \in B_2'$ such that $q_\equiv(r) = q_\equiv(r')$. Let us denote $b_2 = r_1 = r_1'$ and $w^{(2)} \in S$ such that $r_2 \cdots r_{|r|}, r_2' \cdots r'_{|r'|} \in \mathcal{C}_{\mathcal{L}}(w^{(2)})$. Then $b_1 b_2$ is prefix comparable with an element of B_1 and an element of B_1'. As $B_1, B_1' \subseteq \mathcal{AC}_{\mathcal{L}}(S)$ and $\mathcal{AC}_{\mathcal{L}}(S)$ is a prefix code, this shows that $b_1 b_2$ is a prefix of an element of B_1 and an element of B_1'. In particular, $b_1 b_2 \in \mathrm{Pref}^*(\mathcal{AC}_{\mathcal{L}}(S))$.

We iterate the process by looking at the graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}^{\mathcal{A},\mathrm{dom}(\varphi^R_{w^{(i)}})}(vb_1 \cdots b_i)$. We can define B_{i+1} (resp., B'_{i+1}) as the right vertices s such that $b_i s$ has a prefix in B_i (resp., B_i'). These sets are not empty and no vertex of B_{i+1} is connected to a vertex of B'_{i+1}. As $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi^R_{w^{(i)}})}(vb_1 \cdots b_i)$ is connected, $q_\equiv$ identifies an element of B_{i+1} with an element of B'_{i+1}. We then define b_{i+1} and $w^{(i+1)} \in S$ corresponding to these elements. Since $b_1 \cdots b_{i+1} \in \mathrm{Pref}^*(\mathcal{AC}_{\mathcal{L}}(S))$, we can keep iterating.

However, $\mathrm{Pref}^*(\mathcal{AC}_{\mathcal{L}}(S))$ is finite so this process should stop, which leads to a contradiction. This shows that, if $q_\equiv$ identifies two right vertices t, t' of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$, then the path connecting them only uses vertices t'' such that $q_\equiv(t'') = q_\equiv(t)$. Therefore, since $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ is acyclic, its image $\mathcal{E}_{\mathcal{L},\sigma}(v)$ under the graph morphism induced by $q_\equiv$ on the right vertices also is. $\square$

We can then replace the hypothesis on the acyclicity of the graphs $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi^R_w)}(v)$ in Proposition 4.68 by dendricity of the language $\mathcal{L}$. The price to pay however is that we need to look at the extended images of all the words $v \in \mathcal{L}$ at the same time.

Before stating this result, we give one last lemma showing that, instead of looking at the connectedness of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi^L_s),\mathrm{dom}(\varphi^R_p)}(v)$ for all pairs (s, p), we can look at the pairs where $s = \varepsilon$ or $p \in S$. This allows us to somehow split the connectedness condition in two: a condition on the left vertices and one on the right vertices. This will be fundamental in Chapter 5.

Lemma 4.70. *Let G be an acyclic bipartite graph, V^L (resp., V^R) be its set of left (resp., right) vertices and E be its set of edges. For all sets U, W, if the graphs G^{U,V^R} and $G^{V^L,W}$ are connected, then so is the graph $G^{U,W}$.*

Proof. Let x, y be two vertices of $G^{U,W}$. Since $G^{U,W}$ is a subgraph of the connected graph G^{U,V^R} (resp., $G^{V^L,W}$), there exists a path P (resp., P') connecting x and y in G^{U,V^R} (resp., $G^{V^L,W}$). The paths P and P' are then also paths of G which is acyclic. Therefore, $P = P'$. As P is a path in

G^{U,V^R}, it only uses left vertices in U. Similarly, as P' is a path in $G^{V^L,W}$, it only uses right vertices in W. This implies that the path $P = P'$ is a path of $G^{U,W}$. We conclude that any two vertices of $G^{U,W}$ are connected.

□

We can now prove the main result of this section.

Theorem 4.71. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for S and let $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The extended images of all $v \in \mathcal{L}$ are dendric in $\sigma(\mathcal{L})$ if and only if the following conditions are satisfied for all $v \in \mathcal{L}$:*

- *the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_w^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $w \in S$;*
- *the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$.*

The image $\sigma(\mathcal{L})$ is then dendric if and only if, moreover, the initial factors are dendric.

Proof. For all $v \in \mathcal{L}$, if the extended images of v are dendric, then the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$ by Lemma 4.66. This is in particular the case if $p \in S$, or if $s = \varepsilon$ so this concludes the first implication.

Assume now that, for all $v \in \mathcal{L}$, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_w^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $w \in S$, and the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$. We show that the conditions of Proposition 4.68 are satisfied.

Since $\mathcal{L}$ is dendric and $\mathcal{AC}_{\mathcal{L}}(S)$ is an $\mathcal{L}$-maximal prefix code, the graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{AC}_{\mathcal{L}}(S)}(v)$ is a tree for all $v \in \mathcal{L}$ by Proposition 3.41. Moreover, since $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $p \in S$ and all $v \in \mathcal{L}$, we can apply Lemma 4.69 to deduce that $\mathcal{E}_{\mathcal{L},\sigma}(v)$ is acyclic for all $v \in \mathcal{L}$.

This directly implies that $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_w^R)}(v)$ is acyclic for all $v \in \mathcal{L}$ and $w \in S$. It also implies, by Lemma 4.70, that $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $v \in \mathcal{L}$, all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and all $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$. This shows that, by Proposition 4.68, the extended images of v are dendric in $\sigma(\mathcal{L})$ for all $v \in \mathcal{L}$.

Finally, by definition, $\sigma(\mathcal{L})$ is dendric if and only if both the initial factors and the extended images are dendric. □

Remark 4.72. Observe that, if $\varphi_s^L(\mathcal{A}) = \{a\}$ and $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_w^R)}(v)$ has more than one left vertex, then $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_w^R)}(v) = \mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_{as}^L),\mathrm{dom}(\varphi_w^R)}(v)$. In other words, it is not necessary to check the connectedness of the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_w^R)}(v)$ for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ but only for those such that $\#\varphi_s^L(\mathcal{A}) \geq 2$. Similarly, it is sufficient to look at the connectedness of the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_p^R)}(v)$ for the words $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$ such that $\#\varphi_p^R(\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)) \geq 2$.

Remark 4.73. If $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ with $\mathcal{B}$ minimal and $\#\mathcal{B} \geq \#\mathcal{A}$ and if $\mathcal{L}$ is dendric, then by Proposition 4.1, $\sigma(\mathcal{L})$ is dendric if and only if it is connected. In that case, we can therefore replace dendricity of the initial factors by connectedness of the initial factors in Theorem 4.71. Moreover, since we are then only interested in the connectedness of the extended images, Lemma 4.67 and Proposition 4.68 are not needed anymore. However, the acyclicity of $\mathcal{E}_{\mathcal{L},\sigma}(v)$ (Lemma 4.69) is still needed to apply Lemma 4.70.

The previous result should be understood as follows: given a return morphism σ for the factor code S, there is a computable finite prefix code P, a property I of finite sets and two properties E_L, E_R on bipartite graphs such that if $\mathcal{L}$ is a dendric language, then $\sigma(\mathcal{L})$ is dendric if and only if $P \cap \mathcal{L}$ satisfies I and, for all $v \in \mathcal{L}$, the graph $\mathcal{E}_{\mathcal{L},\mathcal{A},P}(v)$ satisfies E_L and E_R.

Indeed, we can take $P = \mathcal{A}\mathcal{C}_{\mathcal{A}^*}(S)$. Since $P \subseteq \mathcal{A}^{\leq \max_{w\in S}|w|+1}$, it is computable. The condition I represents the dendricity of the initial factors. By Proposition 4.48, it only depends on $\sigma(P \cap \mathcal{L})$ so it suffices to say that a set $Q \subseteq P$ satisfies I if all $u \in \mathrm{Fac}(\sigma(Q))$ having no factor in S are dendric in $\mathrm{Fac}(\sigma(Q))$ (where we define the extension graph of a word in a finite factorial set as the bipartite graph generated by the biextensions).

The conditions E_L and E_R ensure that the extended images are dendric. For all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and $w \in S$, the construction to go from $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)}(v)$ to $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_w^R)}(v)$ only depends on σ. Therefore we can define it starting from any bipartite graph G with left vertices in $\mathcal{A}$ and right vertices in P. We then say that G satisfies E_L if, for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and $w \in S$ (note that the values of s and w also only depend on σ), the graph obtained with this construction is connected. We similarly define E_R using the graphs $\mathcal{E}_{\mathcal{L},\sigma}^{\mathcal{A},\mathrm{dom}(\varphi_p^R)}(v)$, $p \in S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S)$, instead.

Example 4.74. Let $\mathcal{A} = \{0, 1, 2\}$ and σ be such that $\sigma(0) = 0$, $\sigma(1) = 01$ and $\sigma(2) = 0102$. It is a return morphism for $S = \{00, 01\}$. We can easily see that $\mathcal{C}_{\mathcal{A}^*}(S) = \{00, 01, 02, 1, 2\}$ so we define $P = \mathcal{A}\{00, 01, 02, 1, 2\}$.

Let us define the property I. We need to look at the dendricity of the initial factors in $\sigma(Q)$ for $Q \subseteq P$. Observe that, in $\sigma(P)$, 1 and 2 are always preceded and followed by 0. This shows that the only potential bispecial initial factors in $\sigma(Q)$ are ε and 0. Assuming that Q contains at least one word beginning with 0 and one beginning with 2 (which is the case if $Q = P \cap \mathcal{L}$ for a language $\mathcal{L}$ over $\mathcal{A}$), then ε is ordinary so it is always dendric. The extensions of 0 however depend on $\mathcal{L}$. Indeed, $(1, 2)$ is always an extension but 0 has the extension $(a, 0)$ if and only if Q contains a word beginning with $a0$, and $(a, 1)$ if and only if Q contains $a1$ or $a2$. If $Q = P \cap \mathcal{L}$ for a dendric language, then one cannot have $a0, b0, a1, b1 \in \mathcal{L}$ or $a0, b0, a2, b2 \in \mathcal{L}$ for distinct a and b. The condition I is then defined as follows: Q satisfies I if, for all distinct $a, b \in \mathcal{A}$,

$$(a0, b0 \in \mathrm{Pref}(Q) \text{ and } a1 \in Q) \implies b2 \notin Q.$$

We now turn to the condition E_L. Since $\varphi^L_\varepsilon = \mathrm{id}$, E_L only depends on the connectedness for $s = \varepsilon$ and $w \in S$. By definition of σ, for any language $\mathcal{L}$ and word $v \in \mathcal{L}$, to go from $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)}(v)$ to $\mathcal{E}^{\mathcal{A},\mathrm{dom}(\varphi^R_{00})}_{\mathcal{L},\sigma}(v)$ (resp., $\mathcal{E}^{\mathcal{A},\mathrm{dom}(\varphi^R_{01})}_{\mathcal{L},\sigma}(v)$), we first merge, for all $a \in \mathcal{A}$, the right vertices in $a0\mathcal{A}$, and $a1$ with $a2$, then we remove the right vertices starting with 1 or 2 (resp., with 0). Starting from a bipartite graph G with left vertices in $\mathcal{A}$ and right vertices in $\mathcal{A}P$, we reproduce these constructions and say that G satisfies E_L if both obtained graphs are connected. An example is done in Figure 4.7.

For the condition E_R, we similarly only need to check the connectedness when $p \in \{00, 010, 01020\}$. Observe that the cases $p = 00$ and $p = 010$ are already considered in condition E_L through the cases $w = 00$ and $w = 01$ respectively (since $\varphi^R_{01}(\mathcal{A}\mathcal{C}_{\mathcal{L}}(S)) = \{0\}$).

4.4.4 The case of return morphisms for a word

As announced in the previous subsections, some of the results can be simplified when we consider a return morphism for a single word. One of the first observation made is that the extensions of the initial factors in $\sigma(\mathcal{L})$ only depends on the morphism σ and not on the language $\mathcal{L}$.

Another simplification is that $\mathcal{E}_{\mathcal{L},\sigma}(v) \cong \mathcal{E}_{\mathcal{L}}(v)$ (see Remark 4.61). More precisely, these graphs share the same left vertices and any right vertex $[t]_\equiv$ of $\mathcal{E}_{\mathcal{L},\sigma}(v)$ corresponds to the right vertex t_1 of $\mathcal{E}_{\mathcal{L}}(v)$. Recall that we are interested in the images of the right vertices under the map φ^R_p for all $p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)$ and, for such p, we have $\varphi^R_p([t]_\equiv) = b$ if and only

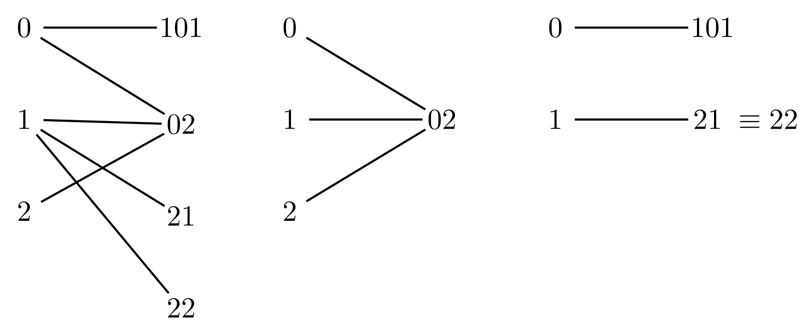


Figure 4.7: Graph G (on the left) and the corresponding graphs for $s = \varepsilon$, and $w = 00$ (in the center) or $w = 01$ (on the right). Since the graph on the right is not connected, G does not satisfy E_L for the morphism of Example 4.74.

if $pb \in \mathrm{Pref}(\sigma(t_1)w)$. We therefore define a new map depending only on t_1 as follows.

Definition 4.75. Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w. For all $p \in \mathcal{B}^*$, we define

$$\phi^R_{\sigma,p} \colon \mathcal{A}^* \to \mathcal{B}^* \quad v \mapsto b \text{ if } pb \in \mathrm{Pref}(\sigma(v)w)$$

on the set $\mathrm{dom}(\phi^R_{\sigma,p}) = \{v \in \mathcal{A}^* : p \in \mathrm{Pref}^*(\sigma(v)w)\}$. We will drop the subscript σ most of the time.

While the map $\phi^R_{\sigma,p}$ technically also depends on w, this will not actually impact the following results. Indeed, by Remark 4.72, we are only interested in this map when p is right special so, by Proposition 4.33, we can for example assume that we take w of minimal length.

For $p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)$, we have $[t]_\equiv \in \mathrm{dom}(\varphi^R_p)$ if and only if $t_1 \in \mathrm{dom}(\phi^R_p)$, and in this case, $\varphi^R_p([t]_\equiv) = \phi^R_p(t_1)$. We then obtain the following simpler statements of Proposition 4.63 and Proposition 4.68 as written in [GLL22, GL22].

Proposition 4.76. *Let $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w, $\mathcal{L}$ be a language over $\mathcal{A}$ and $u \in \sigma(\mathcal{L})$ be an extended image of $v \in \mathcal{L}$. If $u = s\sigma(v)p$, then $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is the image of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi^L_s),\mathrm{dom}(\phi^R_p)}(v)$ under the graph morphism induced by φ^L_s acting on the left vertices and ϕ^R_p acting on the right vertices.*

Proposition 4.77. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. If v is acyclic in $\mathcal{L}$, then the extended images of v are dendric in $\sigma(\mathcal{L})$ if and only if the graph $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\phi_p^R)}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and all $p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)$.*

We directly obtain the following corollary.

Corollary 4.78. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w, $\mathcal{L}$ be a language over $\mathcal{A}$ and $v \in \mathcal{L}$. If v is ordinary in $\mathcal{L}$, then the extended images of v are dendric in $\sigma(\mathcal{L})$.*

Proof. Indeed, if v is ordinary, any subgraph of $\mathcal{E}_{\mathcal{L}}(v)$ with no isolated vertices is connected. □

As a consequence, if the elements of $\mathcal{L}$ are ordinary, all of the extended images are dendric. In other words, if $\mathcal{L}$ is an Arnoux-Rauzy language and the initial factors are dendric, then $\sigma(\mathcal{L})$ is dendric. We have in fact the following stronger result.

Proposition 4.79. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for a word w. The following assertions are equivalent:*

1. *for all (resp., for one) language $\mathcal{L}$ over $\mathcal{A}$, the initial factors are dendric in $\sigma(\mathcal{L})$;*
2. *for all (resp., for one) Arnoux-Rauzy language $\mathcal{L}$ over $\mathcal{A}$, the image $\sigma(\mathcal{L})$ is dendric;*
3. *there exists a language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is dendric;*
4. *there exists a recurrent dendric language $\mathcal{L}$ such that $\sigma(\mathcal{A}) = \mathrm{R}_{\mathcal{L}}(w)$;*
5. *$\#\mathcal{A} = \#\mathcal{B}$ (assuming that $\mathcal{B}$ is minimal) and, for all (resp., for one) language $\mathcal{L}$ over $\mathcal{A}$, the initial factors are connected in $\sigma(\mathcal{L})$.*

Proof. Using Corollary 4.78 and the characterization of Arnoux-Rauzy languages with ordinary words (Proposition 1.27), we conclude that, for any Arnoux-Rauzy language $\mathcal{L}$ over $\mathcal{A}$, the image $\sigma(\mathcal{L})$ is dendric if and only if the initial factors are dendric. Since the initial factors and their extensions do not depend on $\mathcal{L}$, this shows the equivalence between the first two assertions (both in their universal and existential versions).

Let us show the equivalence with the third assertion. If dendricity is preserved for all Arnoux-Rauzy languages, then there clearly exists a language whose image is dendric. Conversely, if the image of a language is dendric, then the initial factors for this language are dendric.

We now turn to the fourth assertion. Let $\mathcal{L}$ be a recurrent Arnoux-Rauzy language over $\mathcal{A}$. If $\sigma(\mathcal{L})$ is dendric, then by Proposition 4.32 on the terminology of return morphisms, $\sigma(\mathcal{A}) = \mathrm{R}_{\sigma(\mathcal{L})}(w)$ and $\sigma(\mathcal{L})$ is a recurrent dendric language. Conversely, if $\sigma(\mathcal{A}) = \mathrm{R}_{\mathcal{L}}(w)$ for some recurrent dendric language $\mathcal{L}$, then let $\mathcal{L}'$ be its derived language with respect to w (and σ). Since $\sigma(\mathcal{L}') = \mathcal{L}$ is dendric, then the initial factors are dendric.

Finally, we turn to the last assertion. If σ satisfies the fourth assertion, then σ defines a bijection between $\mathcal{A}$ and the return words for w in a recurrent dendric language over $\mathcal{B}$ so, by Corollary 3.32 on the number of return words, $\#\mathcal{A} = \#\mathcal{B}$. By the first assertion, the initial factors are dendric so connected. For the converse, if the initial factors are connected, then the image of any Arnoux-Rauzy language is connected. Since $\#\mathcal{A} = \#\mathcal{B}$, this image is in fact dendric by Proposition 4.1. □

We then say that a return morphism for a word is *dendric* if it satisfies any of the equivalent assertions of Proposition 4.79.

Example 4.80. Let σ be the morphism such that $\sigma(0) = 20$, $\sigma(1) = 20221$, $\sigma(2) = 202$ and $\sigma(3) = 20203$. It is a return morphism for 202. Let us show that it is dendric. The only left special initial factors are ε, 2 and 20. Their extension graphs in $\mathrm{Fac}(\sigma(\{0,1,2,3\})202)$ are represented in Figure 4.8. These words are dendric so σ is a dendric morphism and the image of any Arnoux-Rauzy language over $\{0,1,2,3\}$ is dendric. Observe however that ε and 2 are not ordinary therefore σ is not generated by Arnoux-Rauzy morphisms so, by Theorem 4.24, there exists a language $\mathcal{L}$ such that $\sigma(\mathcal{L})$ is not dendric.

Let us now turn to the main theorem of the previous subsection (Theorem 4.71). It also admits an alternative statement in the case of a return morphism for a word, this is the original result from [GLL22] and [GL22].

Proposition 4.81. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The extended images of all $v \in \mathcal{L}$ are dendric in $\sigma(\mathcal{L})$ if and only if the following conditions are satisfied for all $v \in \mathcal{L}$:*

- *the graph $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$;*
- *the graph $\mathcal{E}_{\mathcal{L}}^{\mathcal{A},\mathrm{dom}(\phi_p^R)}(v)$ is connected for all $p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)$.*

The image $\sigma(\mathcal{L})$ is then dendric if and only if, moreover, σ is a dendric return morphism for a word.

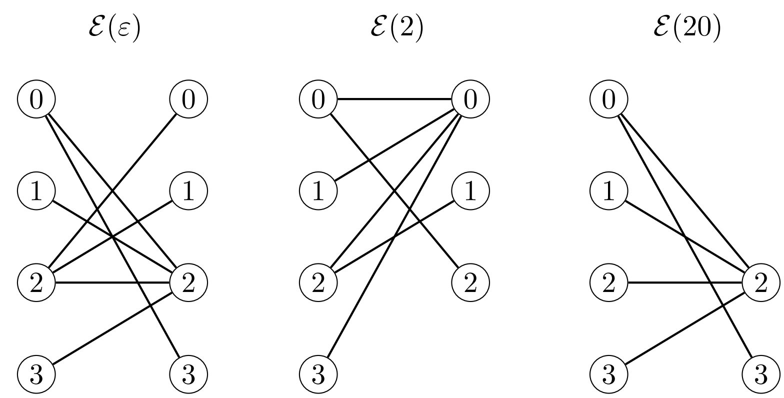


Figure 4.8: The extension graphs of ε (on the left), 2 (in the center) and 20 (on the right) in $\mathrm{Fac}(\sigma(\{0,1,2,3\})202)$ for the morphism σ of Example 4.80.

At first glance, the previous result might seem unpractical as we still need to check some condition for all $v \in \mathcal{L}$. However, when taking a closer look, we can see that only finitely many words can fail this condition. In fact, it is not the first time in this work that we look at extension graphs in which we remove some of the vertices. This was already implicitly done in Proposition 2.43 where we relate it to connectedness of the graphs $G_n^L(\mathcal{L})$ and $G_n^R(\mathcal{L})$. This gives us the following alternative characterization [GL22].

Proposition 4.82. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The extended images of all $v \in \mathcal{L}$ are dendric in $\sigma(\mathcal{L})$ if and only if the following conditions are satisfied:*

- *for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_s^L)$ is connected;*
- *for all $p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)$, the subgraph of $G^R(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\phi_p^R)$ is connected.*

The image $\sigma(\mathcal{L})$ is then dendric if and only if, moreover, σ is a dendric return morphism for a word.

Proof. By Proposition 2.57, there exists N such that, for all $n \geq N$, $G_n^L(\mathcal{L}) = G^L(\mathcal{L})$. Let $C \subseteq \mathcal{A}$. Then, by Proposition 2.43, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in C is connected if and only if, for all $v \in \mathcal{L}$, in the graph $\mathcal{E}_{\mathcal{L}}(v)$ where we removed the left vertices not in C, the left vertices are connected, or equivalently, the graph $\mathcal{E}_{\mathcal{L}}^{C,\mathcal{A}}(v)$ is connected. We similarly show that, for all $C \subseteq \mathcal{A}$, the subgraph of $G^R(\mathcal{L})$ generated by the vertices in C is connected if and only if, for all $v \in \mathcal{L}$, the graph $\mathcal{E}_{\mathcal{L}}^{\mathcal{A},C}(v)$ is connected. The conclusion then follows by Proposition 4.81. □

Example 4.83. Let σ be the morphism such that $\sigma(0) = 20$, $\sigma(1) = 20221$, $\sigma(2) = 202$ and $\sigma(3) = 20203$. By Example 4.80, σ is a dendric return morphism for 202.

Since the images of the letters all end with different letters, $\mathrm{dom}(\varphi_s^L) \cap \mathcal{A}$ is either $\mathcal{A}$, $\emptyset$ or one letter depending on s. In other words, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_s^L)$ is always connected if $\mathcal{L}$ is dendric by Corollary 2.44. On the other hand, $\mathrm{dom}(\phi_p^R)$ can take other values if $p = 2020$ or $p = 2022$. We then have $\mathrm{dom}(\phi_{2020}^R) = \{0, 3\}$ and $\mathrm{dom}(\phi_{2022}^R) = \{1, 2\}$.

Therefore, by Proposition 4.82, the image $\sigma(\mathcal{L})$ of a dendric language $\mathcal{L}$ is dendric if and only if $G^R(\mathcal{L})$ contains an edge between 0 and 3, and between 1 and 2. In other words, if and only if 0 and 3 (resp., 1 and 2) are right extensions of a right special infinite word in $\mathcal{L}$ by Proposition 2.61.

4.4.5 The case of return morphisms for a set of letters

We now consider another particular case of return morphisms for a set and look at return morphisms for a set of letters. We show below that we can obtain a simpler characterization, as in the case of return morphisms for a word.

In the previous subsection, we were able to obtain simpler results because we had $\mathcal{E}_{\mathcal{L},\sigma}(v) \cong \mathcal{E}_{\mathcal{L}}(v)$. While we do not have this isomorphism anymore when σ is a return morphism for a set of letters, we will show that we can still use $\mathcal{E}_{\mathcal{L}}(v)$ instead of $\mathcal{E}_{\mathcal{L},\sigma}(v)$ to obtain the characterizations.

To do this we introduce some notations.

Definition 4.84. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for $S \subseteq \mathcal{B}$. For all $p \in \mathrm{Pref}(\sigma(\mathcal{A}))$, we define

$$d_p = \{a \in \mathcal{A} : p \in \mathrm{Pref}(\sigma(a))\}.$$

Observe that, if $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ is a return morphism for $S \subseteq \mathcal{B}$, then

$$S\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})S) = \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A})).$$

Moreover, any $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$ has exactly one letter in S (its first letter), so

$$p \in \mathrm{Pref}(\sigma(t)) \iff p \in \mathrm{Pref}(\sigma(t_1)).$$

This shows that $\mathcal{C}_{\mathcal{L}}(S) = \mathcal{A}$ and, for all $t \in \mathcal{A}^2 = \mathcal{A}\mathcal{C}_{\mathcal{L}}(S)$,

$$t \in \mathrm{dom}(\varphi_p^R) \iff t_1 \in d_p.$$

Therefore, for any set U, the graph $\mathcal{E}_{\mathcal{L},\sigma}^{U,\mathrm{dom}(\varphi_p^R)}(v)$ is closely related to $\mathcal{E}_{\mathcal{L}}^{U,d_p}(v)$.

Using this observation, we obtain the following characterization.

Proposition 4.85. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for $S \subseteq \mathcal{B}$ and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The image $\sigma(\mathcal{L})$ is dendric if and only if the initial factors are dendric in $\sigma(\mathcal{L})$ and, for all $v \in \mathcal{L}$, the following conditions are satisfied:*

- *the graph $\mathcal{E}_{\mathcal{L}}^{\mathcal{A},d_p}(v)$ is connected for all $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$;*
- *the graph $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$.*

Proof. Assume that $\sigma(\mathcal{L})$ is dendric. Clearly, the initial factors are dendric thus let us look at the conditions on $v \in \mathcal{L}$.

The first condition is direct. More generally, for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ and $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$, the graph $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),d_p}(v)$ is the image of $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ under the graph morphism identifying right vertices beginning with the same letter. By Lemma 4.66, $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected thus so is $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),d_p}(v)$. This is in particular the case if $s = \varepsilon$.

Assume now by contradiction that there exists $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ of minimal length such that $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ is not connected. Since $\mathcal{L}$ is dendric, $s \neq \varepsilon$ so let $c \in \mathcal{B}$ and $s' \in \mathcal{B}^*$ be such that $s = cs'$. By minimality of s, $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_{s'}^L),\mathcal{A}}(v)$ is connected. Let a_1 and a_2 be two left vertices of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ that are not connected in $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ and such that the path P connecting them in $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_{s'}^L),\mathcal{A}}(v)$ is of minimal length. In particular, no intermediary left vertex of P is a vertex of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$. Let b_1, b_2 be the two right vertices of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ such that $\{a_1, b_1\}$ (resp., $\{a_2, b_2\}$) is the first (resp., last) edge of P. Then $\sigma(b_1)$ and $\sigma(b_2)$ cannot begin with the same letter. Indeed, otherwise the graph $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\sigma(b_1)_1}(v)$ is not connected, which contradicts what we have already proved since $\sigma(b_1)_1 \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$.

For every edge $\{a, b\}$ in P (where a is the left vertex), $s'\sigma(v)$ has a corresponding extension (a', b') in $\sigma(\mathcal{L})$ where $a's' \in \mathrm{Suff}(\sigma(a))$, $b' = \sigma(b)_1$. Therefore, P induces a cycle P' in $\mathcal{E}_{\sigma(\mathcal{L})}(s'\sigma(v))$ and this cycle reduces to a non trivial simple cycle since, by definition of a_1 and a_2, the only left vertices of P mapped to c are a_1 and a_2, and the vertices following and preceding c in P' are $\sigma(b_1)_1$ and $\sigma(b_2)_1$ which are different. This contradicts the fact that $\sigma(\mathcal{L})$ is dendric.

Let us now prove the converse. More specifically, we show that under the conditions of the statement for all $v \in \mathcal{L}$, the extended images are dendric. By Theorem 4.71, it suffices to show that the graph $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $v \in \mathcal{L}$, $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$.

Since $\mathcal{E}_{\mathcal{L},\sigma}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is the image of the graph $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ under the graph morphism induced by $q_{\equiv}$ acting on the right vertices, it suffices to show that $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected. Moreover, as this graph has no isolated vertex, it suffices to prove that any two right vertices are connected.

Let a_1a_2 and b_1b_2 be two distinct right vertices of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$. If $a_1 = b_1$, then a_2, b_2 are two distinct right vertices of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(va_1)$ which is connected by hypothesis. Moreover, since $p \in \mathrm{Pref}(\sigma(\mathcal{A})) \cap \mathrm{Pref}^*(\sigma(a_1a_2))$, p is a prefix of $\sigma(a_1)$ so the path connecting a_2, b_2 in $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(va_1)$ induces a path between a_1a_2 and b_1b_2 in $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$. In other words, any two right vertices of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ beginning with the same letter are connected.

Assume now that $a_1 \neq b_1$. Then a_1 and b_1 are two right vertices of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),d_p}(v)$ which is connected by hypothesis and by Lemma 4.70. Let P be a path connecting them. For any edge $\{a, b\}$ of P (where a is the left vertex), there is a corresponding edge $\{a, bb'\}$ in $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$. Since right vertices of $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ beginning with the same letter are connected, this shows that P induces a path between a_1a_2 and b_1b_2 in $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$.

This ends the proof that $\mathcal{E}_{\mathcal{L},\mathcal{A},\mathcal{A}^2}^{\mathrm{dom}(\varphi_s^L),\mathrm{dom}(\varphi_p^R)}(v)$ is connected for all $v \in \mathcal{L}$, $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$ and therefore concludes the proof of the second implication. □

As in Remark 4.73, if $\#\mathcal{A} = \#\mathcal{B}$, then we can replace dendricity of the initial factors by connectedness. In the case of return morphisms for a set of letters, we also have the following stronger result.

Proposition 4.86. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ (assuming that $\mathcal{A}$ is the smallest image alphabet) be a return morphism for $S \subseteq \mathcal{A}$ and $\mathcal{L}$ be a dendric language over*

$\mathcal{A}$. The image $\sigma(\mathcal{L})$ is dendric if and only if, for some $c \notin \mathcal{A} \setminus S$, the initial factors are connected in $\sigma(\mathcal{A})c$ and, for all $v \in \mathcal{L}$, the following conditions are satisfied:

- *the graph $\mathcal{E}_{\mathcal{L}}^{\mathcal{A},d_p}(v)$ is connected for all $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$;*
- *the graph $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_s^L),\mathcal{A}}(v)$ is connected for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$.*

Proof. Let us show that, in Proposition 4.85, we can replace the condition "the initial factors are dendric in $\sigma(\mathcal{L})$" by "the initial factors are connected in $\sigma(\mathcal{A})c$" with $c \notin \mathcal{A} \setminus S$.

Let u be an initial factor. If $\sigma(\mathcal{L})$ is dendric, then u is connected in $\sigma(\mathcal{L})$. Moreover, its extension graph in $\sigma(\mathcal{A})c$ is obtained by identifying in $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ the right vertices in S to $c \notin \mathcal{A} \setminus S$. In particular, this graph is also connected.

Conversely, assume that the conditions on $\mathcal{E}_{\mathcal{L}}(v)$ are satisfied for all $v \in \mathcal{L}$ and that the initial factors are connected in $\sigma(\mathcal{A})c$. As in the proof of Proposition 4.85, this implies that the extended images are dendric. To conclude that $\sigma(\mathcal{L})$ is dendric, it then suffices to show that the initial factors are connected in $\sigma(\mathcal{L})$ by Proposition 4.1.

Let u be an initial factor. Observe that u has a right extension in S if and only if $u \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$. Clearly, if $u \notin \mathrm{Suff}^*(\sigma(\mathcal{A}))$, then its extensions in $\sigma(\mathcal{L})$ and its extensions in $\sigma(\mathcal{A})c$ coincide so u is connected in $\sigma(\mathcal{L})$ by hypothesis. Assume now that $u \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$. Observe that $\mathcal{E}_{\sigma(\mathcal{L})}^{\mathcal{A},S}(u)$ is the image of $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_u^L),\mathcal{A}}(\varepsilon)$ under the graph morphism induced by φ_u^L acting on the left vertices and φ_ε^R acting on the right vertices. Since $\mathcal{E}_{\mathcal{L}}^{\mathrm{dom}(\varphi_u^L),\mathcal{A}}(\varepsilon)$ is connected by hypothesis, this shows that any two right vertices in S in $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ are connected. Recall that the graph $\mathcal{E}_{\sigma(\mathcal{A})c}(u)$ is obtained by identifying in $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ the right vertices in S to c. We then conclude that if two vertices are connected in $\mathcal{E}_{\sigma(\mathcal{A})c}(u)$, then they are connected in $\mathcal{E}_{\sigma(\mathcal{L})}(u)$. This ends the proof that u is connected in $\sigma(\mathcal{L})$. □

In the previous result, we are considering the extension graph of an initial factor w in a finite factorial set Q. Recall that it is understood as the bipartite graph generated by the extensions of w in Q, i.e., the pairs (a, b) such that $awb \in Q$. In particular, this graph has no isolated vertices meaning that its left (resp., right) vertices are not necessarily the left (resp., right) extensions of w in Q.

The idea of looking at the extensions of the initial factors in $\sigma(\mathcal{A})c$ comes from the study of Arnoux-Rauzy languages. Indeed, we have the following equivalences.

Proposition 4.87. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for $S \subseteq \mathcal{B}$, let $\mathcal{L}$ be an Arnoux-Rauzy language over $\mathcal{A}$ and let $c \notin \mathcal{B} \setminus S$. The initial factors are dendric (resp., connected) in $\sigma(\mathcal{L})$ if and only if they are dendric (resp., connected) in $\sigma(\mathcal{A})c$.*

Proof. Let u be an initial factor. If $u \notin \mathrm{Suff}^*(\sigma(\mathcal{A}))$, then its extensions in $\sigma(\mathcal{L})$ and $\sigma(\mathcal{A})c$ coincide so the conclusion is direct. Assume now that $u \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, i.e., u has at least one right extension in S.

Recall that $\mathcal{E}^{\mathcal{A},S}_{\sigma(\mathcal{L})}(u)$ is the image of $\mathcal{E}^{\mathrm{dom}(\varphi^L_u),\mathcal{A}}_{\mathcal{L}}(\varepsilon)$ under the graph morphism induced by φ^L_u on the left and φ^R_ε on the right. Since $\mathcal{L}$ is an Arnoux-Rauzy language, in $\mathcal{E}^{\mathrm{dom}(\varphi^L_u),\mathcal{A}}_{\mathcal{L}}(\varepsilon)$, there exists at most one right vertex having multiple left neighbors and all the right vertices have a common left neighbor ℓ. Therefore, the same can be said for the graph $\mathcal{E}^{\mathcal{A},S}_{\sigma(\mathcal{L})}(u)$, replacing ℓ by $\varphi^L_u(\ell)$. This shows that, in $\mathcal{E}_{\sigma(\mathcal{L})}(u)$, at most one right vertex in S is not of degree 1 and all the right vertices in S have a common neighbor. We conclude that $\mathcal{E}_{\sigma(\mathcal{L})}(u)$ is acyclic (resp., connected) if and only if its image $\mathcal{E}_{\sigma(\mathcal{A})c}(u)$ under the graph morphism identifying the right vertices in S is acyclic (resp., connected). □

Corollary 4.88. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ (assuming that $\mathcal{A}$ is the smallest image alphabet) be a return morphism for $S \subseteq \mathcal{A}$. The following assertions are equivalent:*

1. *for all (resp., for one) Arnoux-Rauzy language $\mathcal{L}$ over $\mathcal{A}$, the image $\sigma(\mathcal{L})$ is dendric;*
2. *there exists a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is dendric;*
3. *for any (resp., for one) $c \notin \mathcal{A} \setminus S$, the initial factors are connected in $\sigma(\mathcal{A})c$;*
4. *for any (resp., for one) $c \notin \mathcal{A} \setminus S$, the initial factors are dendric in $\sigma(\mathcal{A})c$.*

Proof. The first assertion directly implies the second one, which implies the third one by Proposition 4.86. Since the elements of an Arnoux-Rauzy language are ordinary, the third assertion implies the first one by Proposition 4.86 once again. Finally, the equivalence with the last assertion follows from Proposition 4.87. □

Example 4.89. Let σ be the morphism defined by $\sigma(0) = 01$, $\sigma(1) = 011$ and $\sigma(2) = 2$. It is a return morphism for $\{0, 2\}$ and the initial factors are

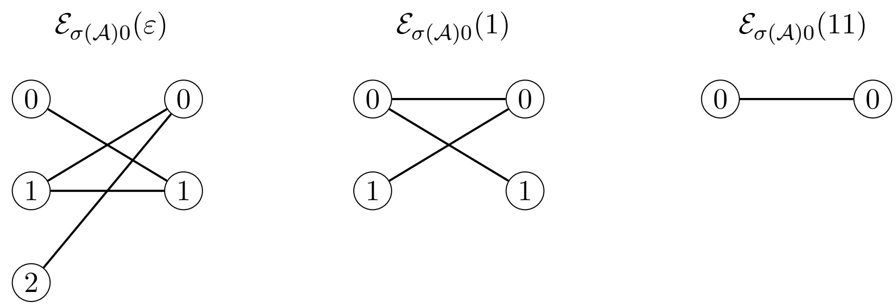


Figure 4.9: The extension graphs of ε, 1 and 11 in $\sigma(\mathcal{A})0$ for the morphism σ of Example 4.89 and $\mathcal{A} = \{0, 1, 2\}$.

ε, 1 and 11. Let us take $c = 0$ and look at the extension graphs of the initial factors in $\sigma(\{0, 1, 2\})0$. Their extension graphs are represented in Figure 4.9. We then see that σ satisfies the equivalent conditions of Corollary 4.88.

Using Proposition 4.82, or more specifically, its proof, we can restate the characterizations of Propositions 4.85 and 4.86 using the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ as follows.

Corollary 4.90. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for $S \subseteq \mathcal{B}$ and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The image $\sigma(\mathcal{L})$ is dendric if and only if the initial factors are dendric in $\sigma(\mathcal{L})$ and the following conditions are satisfied:*

- *for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_s^L)$ is connected;*
- *for all $p \in \mathcal{B}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$, the subgraph of $G^R(\mathcal{L})$ generated by the vertices in d_p is connected.*

Corollary 4.91. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ (assuming that $\mathcal{A}$ is the smallest image alphabet) be a return morphism for $S \subseteq \mathcal{A}$ and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The image $\sigma(\mathcal{L})$ is dendric if and only if the initial factors are connected in $\sigma(\mathcal{A})c$ for $c \notin \mathcal{A} \setminus S$ and the following conditions are satisfied:*

- *for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_s^L)$ is connected;*
- *for all $p \in \mathcal{A}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$, the subgraph of $G^R(\mathcal{L})$ generated by the vertices in d_p is connected.*

This final result shows that, if σ stays on the same alphabet (or more generally, an alphabet of the same size) the preservation of dendricity only depends on the morphism σ and on the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$, as in the case of a return morphism for a word. Without this condition on the alphabet sizes, we moreover need to know $\mathcal{L}_2 = \mathcal{AC}_{\mathcal{L}}(S) \cap \mathcal{L}$ to determine the dendricity of the initial factors.

Example 4.92. Let us continue Example 4.89 with the morphism σ such that $\sigma(0) = 01$, $\sigma(1) = 011$ and $\sigma(2) = 2$. We easily check that the only non trivial value of $\mathrm{dom}(\varphi_s^L)$ is when $s = 1$ and $\mathrm{dom}(\varphi_1^L) = \{0, 1\}$. Similarly, the only non trivial d_p is when $p \in \{0, 01\}$ and we then have $d_p = \{0, 1\}$. By Example 4.89, σ satisfies the conditions of Corollary 4.88 so, by Corollary 4.91, we conclude that, if $\mathcal{L}$ is a dendric language over $\{0, 1, 2\}$, then $\sigma(\mathcal{L})$ is dendric if and only if both $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ contain the edge $\{0, 1\}$.

4.5 What about eventual dendricity?

We now turn to the question of preserving eventual dendricity, for one or for all eventually dendric languages. Based on the previous section, we can directly deduce that return morphisms for a word preserve eventual dendricity.

Proposition 4.93. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w. If $\mathcal{L}$ is an eventually ordinary language of threshold N over $\mathcal{A}$, then $\sigma(\mathcal{L})$ is eventually dendric of threshold at most $\|\sigma\|(N + 1) + |w| - 1$.*

Proof. Observe that the initial factors are of length at most $\|\sigma\| + |w| - 2$. Therefore, if $u \in \sigma(\mathcal{L})$ is such that $|u| \geq \|\sigma\|(N + 1) + |w| - 1$, then u is an extended image of some $v \in \mathcal{L}$. By definition, u is an internal factor of $\sigma(avb)w$ for some $a, b \in \mathcal{A}$ so $|v| \geq N$. By Corollary 4.78, u is dendric. $\square$

Using Theorem 3.42 on the stability of eventual dendricity under derivation, we deduce the following equivalence.

Corollary 4.94. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w and $\mathcal{L}$ be a language over $\mathcal{A}$. The language $\mathcal{L}$ is recurrent eventually dendric if and only if $\sigma(\mathcal{L})$ is.*

In particular, this shows that, contrary to the results presented in Section 4.1 for dendricity, there is no constraint on the alphabet sizes to preserve eventual dendricity, even with the injectivity hypothesis.

Return morphisms for a word, and more generally, return morphisms for a set are particular examples of recognizable morphisms. Recognizability of a morphism can be understood as the fact that we can uniquely determine the pre-image of any bi-infinite word, sometimes restricting ourselves to pre-images in a given shift space. An important result on recognizability is Mossé's Theorem (see [BPR23] for the most general statement so far in the substitutive case, and [BPRS23] for the S-adic case).

Definition 4.95. Let $X \subseteq \mathcal{A}^{\mathbb{Z}}$ be a shift space. A morphism $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ is *recognizable on* X if, for all $y \in \sigma(X)$, there exists exactly one pair $(x, k) \in X \times \mathbb{N}$ such that $0 \leq k < |\sigma(x_0)|$ and $y = S^k(\sigma(x))$.

Observe that if $Y \subseteq X$ and if σ is recognizable on X, then σ is recognizable on Y. In particular, if σ is recognizable on $\mathcal{A}^{\mathbb{Z}}$, then it is recognizable on any shift space. We then say that σ is *recognizable*.

In this section, we therefore sometimes talk about shift spaces instead of languages. Recall that any language corresponds to a shift space and conversely. We will prove that non-erasing recognizable morphisms preserve eventual dendricity. More generally, we show that preservation of eventual dendricity can be reduced to preservation of eventual dendricity under a letter-to-letter morphism, and in the case of recognizable morphisms, this morphism induces a topological conjugacy. The results presented here were done in collaboration with J. Leroy and P. Stas.

Any non-erasing morphism can be written as the composition of a particular morphism and a letter-to-letter morphism. Indeed, if $\sigma \colon \{a_1, \ldots, a_k\}^* \to \mathcal{B}^*$, then $\sigma = \gamma \circ \tau$ if we define

$$\tau(a_i) = a_{i,1} \cdots a_{i,|\sigma(a_i)|}$$

and

$$\gamma(a_{i,j}) = \sigma(a_i)_j$$

where $a_{i,j} \neq a_{i',j'}$ if $i \neq i'$ or $j \neq j'$. Such a morphism τ preserves eventual dendricity as shown in the following lemma inspired by discussions with B. Espinoza.

Lemma 4.96. *Let $\mathcal{A} = \{a_1, \ldots, a_k\}$ and $(\ell_i)_{1 \leq i \leq k}$ be a vector of lengths such that $\ell_i \geq 1$ for all i. Let τ be a morphism such that, for all i, $\tau(a_i) = a_{i,1} \cdots a_{i,\ell_i}$ where $a_{i,j} \neq a_{i',j'}$ if $i \neq i'$ or $j \neq j'$. Then $\mathcal{L} \subseteq \mathcal{A}^*$ is an eventually dendric language if and only if $\tau(\mathcal{L})$ is eventually dendric. Moreover, if N and M are the threshold of $\mathcal{L}$ and $\tau(\mathcal{L})$ respectively, then* $M \leq \max\{1, \|\tau\|(N-1)+1\}$.

Proof. Let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language and let $u \in \tau(\mathcal{L}) \setminus \{\varepsilon\}$. If $u_1 = a_{i,j}$ with $j \neq 1$, then u is always preceded by $a_{i,j-1}$ in $\tau(\mathcal{L})$. Similarly, if $u_{|u|} = a_{i,j}$ with $j \neq \ell_i$, then u is always followed by $a_{i,j+1}$. In other words, if u is bispecial, there exists $v \in \mathcal{L}$ such that $u = \tau(v)$ and u can only be preceded (resp., followed) by letters of the form a_{i,ℓ_i} (resp.,$a_{i,1}$). Moreover,

$$(a_{i,\ell_i}, a_{j,1}) \in E_{\tau(\mathcal{L})}(u) \iff (a_i, a_j) \in E_{\mathcal{L}}(v)$$

so u is dendric if and only v is. This shows that $\mathcal{L}$ is eventually dendric if and only if $\tau(\mathcal{L})$ is. Moreover, if $\mathcal{L}$ is eventually dendric of threshold N and $|u| \geq \|\tau\|(N-1)+1$, then u is not bispecial or $u = \tau(v)$ with $|v| \geq N$ so u is dendric. □

The key of preserving dendricity resides then in the letter-to-letter morphism γ. Any letter-to-letter morphism induces a factor map between any shift space X and its image. However, when σ is recognizable on X, then the corresponding morphism γ induces a topological conjugacy between $\tau(X)$ and $\sigma(X)$. This implies that X is eventually dendric if and only if $\sigma(X)$ is. This is detailed in the following result.

Proposition 4.97. *Let $X \subseteq \mathcal{A}^{\mathbb{Z}}$ be a shift space and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a non-erasing recognizable morphism on X. Then X is eventually dendric if and only if $\sigma(X)$ is.*

Proof. Assume that $\mathcal{A} = \{a_1, \dots, a_k\}$ and let us define the morphisms τ and γ as follows:

$$\tau(a_i) = a_{i,1} \cdots a_{i,|\sigma(a_i)|}$$

and

$$\gamma(a_{i,j}) = \sigma(a_i)_j$$

where $a_{i,j} \neq a_{i',j'}$ if $i \neq i'$ or $j \neq j'$. Then $\sigma = \gamma \circ \tau$. Let us show that γ induces a topological conjugacy between $\tau(X)$ and $\sigma(X)$. We denote $\mathcal{C} = \cup_i \{a_{i,j} : j \leq |\sigma(a_i)|\}$ and $\varphi\colon \mathcal{C}^{\mathbb{Z}} \to \mathcal{B}^{\mathbb{Z}}$ such that for all $x \in \mathcal{C}^{\mathbb{Z}}$ and $n \in \mathbb{Z}$

$$\varphi(x)_n = \gamma(x_n).$$

By definition, for any $x \in \mathcal{C}^{\mathbb{Z}}$, we have $\varphi(x) = \gamma(x)$ so $\varphi(\tau(X)) = \sigma(X)$. Therefore, by the Curtis-Hedlund-Lyndon theorem (Theorem 2.13), $\varphi_{|\tau(X)}$ is a factor map between $\tau(X)$ and $\sigma(X)$. Let us show that it is a conjugacy. Let $y, y' \in \tau(X)$ with $y \neq y'$. By definition of τ, there exist unique $x \in X$, $k < |\tau(x_0)|$ such that $y = S^k(\tau(x_0))$ and unique $x' \in X$, $k' < |\tau(x'_0)|$ such that $y' = S^{k'}(\tau(x'_0))$. Moreover, $x \neq x'$ or $k \neq k'$. Since φ commutes with the dynamics, we then have

$$\begin{aligned}\varphi(y) = \varphi(y') &\iff \varphi(S^k(\tau(x))) = \varphi(S^{k'}(\tau(x')))\\ &\iff S^k(\varphi(\tau(x))) = S^{k'}(\varphi(\tau(x')))\\ &\iff S^k(\sigma(x)) = S^{k'}(\sigma(x')).\end{aligned}$$

This shows that φ is injective if and only if σ is recognizable on X. In this case, we then conclude that $\sigma(X)$ is eventually dendric if and only if $\tau(X)$ is by stability under topological conjugacy (Theorem 2.20), if and only if X is by Lemma 4.96. □

For non-recognizable morphisms on the other hand, $\sigma(X)$ can be eventually dendric even if X is not. Indeed, if there exists v such that $\sigma(a)$ is a power of v for all a, then $\sigma(X)$ is eventually dendric, independently of the shift space X.

It is not known however if the converse is true, i.e., if X can be eventually dendric but not $\sigma(X)$. This question is equivalent to stability under topological factorization as stated below.

Proposition 4.98. *The following assertions are equivalent.*

1. *Eventual dendricity is stable under topological factorization.*
2. *Eventual dendricity is stable under image by any non-erasing morphism.*
3. *Eventual dendricity is stable under image by any letter-to-letter morphism identifying two letters.*

Proof. Using the notations of the proof of Proposition 4.97, $\sigma(X)$ is a factor of $\tau(X)$. Therefore, if eventual dendricity is stable under topological factorization, then it is preserved when taking the image under a non-erasing morphism by Lemma 4.96.

If it is stable for any non-erasing morphism, it is clearly stable for any letter-to-letter morphism.

Conversely, if eventual dendricity is preserved by letter-to-letter morphisms identifying two letters, then it is preserved by any letter-to-letter morphism since they can be obtained as compositions of morphisms identifying two letters and permutations. Moreover, if Y is a factor of X, then Y is the image of the higher block shift space $X^{(N)}$, $N \geq 1$, under some letter-to-letter morphism. This implies that, if X is eventually dendric, so is

$X^{(N)}$ (Theorem 2.20) and Y by hypothesis. Therefore, eventual dendricity is stable under topological factorization. $\square$

While we do not know if the assertions of the previous proposition are (all) true or false, we end this section with a partial result obtained with J. Leroy and P. Stas.

Proposition 4.99. *Let $\mathcal{L} \subseteq \mathcal{A}^{\mathbb{Z}}$ be an eventually dendric language and $\gamma \colon \mathcal{A}^* \to \mathcal{B}^*$ be a letter-to-letter morphism.*

1. *The language $\gamma(\mathcal{L})$ is eventually dendric if and only if it is eventually connected.*
2. *There exists $M \geq 0$ such that if $v \in (\gamma(\mathcal{L}))_{\geq M}$ is not connected, then $\gamma^{-1}(v)$ contains no left special nor right special word in $\mathcal{L}$.*

Proof. Since $\mathcal{L}$ is eventually dendric, it has eventually affine factor complexity so the factor complexity of $\gamma(\mathcal{L})$ is at most linear (Proposition 3.7). By Proposition 2.30, this shows that $\gamma(\mathcal{L})$ is eventually dendric if and only if it is eventually connected.

We now turn to the second claim. Since $\mathcal{L}$ is eventually right ordinary by Proposition 2.26, there exists $N \geq 0$ and $u^{(1)}, \dots, u^{(k)} \in \mathcal{A}^{\mathbb{N}}$ such that for all $n \geq N$, their length-n prefixes are the k length-n left special words of $\mathcal{L}$.

Let $i \leq k$ and let us consider the left extensions of the images of the prefixes of $u^{(i)}$. Since $E^L_{\gamma(\mathcal{L})}(\gamma(u^{(i)}_{[0,n+1)})) \subseteq E^L_{\gamma(\mathcal{L})}(\gamma(u^{(i)}_{[0,n)}))$ for all $n \in \mathbb{N}$, there exists $N_i \geq N$ such that, for all $n \geq N_i$, we have $E^L_{\gamma(\mathcal{L})}(\gamma(u^{(i)}_{[0,n)})) = E^L_{\gamma(\mathcal{L})}(\gamma(u^{(i)}_{[0,N_i)}))$. In particular, $\gamma(u^{(i)}_{[0,n)})$ is then connected since its left extensions can all be extended on the right by $\gamma(u^{(i)}_n)$.

If we set $M^L = \max\{N_i : i \leq k\}$, then for all $v \in (\gamma(\mathcal{L}))_{\geq M^L}$, if $\gamma^{-1}(v)$ contains a left special word, then there exists $i \leq k$ such that $v = \gamma(u^{(i)}_{[0,|v|)})$ so v is connected by definition of M^L.

We can symmetrically define M^R such that, for all $v \in (\sigma(\mathcal{L}))_{\geq M^R}$, if $\gamma^{-1}(v)$ contains a right special word, then v is connected. We conclude by taking $M = \max\{M^L, M^R\}$. $\square$

4.6 Open questions

This chapter is centered around one broad open question stated in the introduction and that we naturally recall here.

Question 4.1. Given a dendric language $\mathcal{L}$, can we characterize the morphisms σ such that $\sigma(\mathcal{L})$ is dendric? Alternatively, given a morphism σ, can we characterize the dendric languages $\mathcal{L}$ such that $\sigma(\mathcal{L})$ is dendric?

While the techniques presented in Section 4.4 are quite specific to return morphisms, one could hope to use a similar approach for families of morphisms having nice properties with respect to recognizability.

We will however focus here on questions which seem simpler to tackle. The first questions are based on the study of possible alphabet sizes done in Section 4.1. As we already mentioned, if $\mathcal{L}$ is a recurrent Arnoux-Rauzy language on an alphabet of size at least 3 and σ is a non-erasing morphism such that $\sigma(\mathcal{L})$ is dendric, then $\sigma(\mathcal{L})$ cannot be on a binary alphabet [JP02]. We therefore ask the question below.

Question 4.2. Let $\mathcal{L}$ be a recurrent Arnoux-Rauzy language on an alphabet of size k and let σ be a non-erasing morphism. If $\sigma(\mathcal{L})$ is dendric, can $\mathcal{L}$ be over an alphabet of size $\ell \notin \{1, k\}$?

On the other hand, we showed in Proposition 4.3 that, if we consider languages of RIET for some specific orders in the previous question, any size $\ell \in [1, k]$ can be obtained. This leads to our next question.

Question 4.3. Given a dendric language $\mathcal{L}$ over an alphabet of size k, can we determine the sizes $\ell \in [1, k]$ for which there exists a non-erasing morphism σ such that $\sigma(\mathcal{L})$ is a dendric language over an alphabet of size ℓ?

Languages of RIET and Arnoux-Rauzy languages often present examples of antagonist extreme behaviors among dendric languages. In particular, they have completely different graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$. Indeed, for Arnoux-Rauzy languages, they are complete graphs (of minimal diameter) while, for languages of RIET, they are line graphs (of maximal diameter). Therefore, the answer to the previous question could be related to the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$.

We now restrict ourselves to injective morphisms. Indeed, we proved in Proposition 4.6 that, for any alphabets $\mathcal{A}$ and $\mathcal{B}$ such that $\frac{\#\mathcal{A}-1}{\#\mathcal{B}-1}$ is an integer, we could find an injective morphism σ and a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a dendric language over $\mathcal{B}$. However, if $1 < \frac{\#\mathcal{A}-1}{\#\mathcal{B}-1} < 2$, this is impossible by Proposition 4.7 (we can even slightly reduce the conditions). This raises the following question.

Question 4.4. Let σ be an injective morphism for which there exists a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $\sigma(\mathcal{L})$ is a dendric language over $\mathcal{B}$. Do we have $\frac{\#\mathcal{A}-1}{\#\mathcal{B}-1} \in \mathbb{N}$?

We now turn to some questions on Section 4.2. Indeed, we fully characterized the non-erasing morphisms which preserved dendricity for all dendric languages. The erasing case however has not been considered at all. Clearly, if a morphism is periodic (i.e., the image of any language is either not a language or it is a periodic language), then we can easily determine if it is dendric preserving. The question is therefore more interesting in the aperiodic case.

Question 4.5. Can an aperiodic erasing morphism be dendric preserving? More precisely, can we find an erasing morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, $\#\mathcal{B} \geq 2$, such that the image of any dendric language over $\mathcal{A}$ is either not a language or it is a dendric language over $\mathcal{B}$, and there exists a dendric language over $\mathcal{A}$ whose image is an aperiodic language?

In fact, even the following simpler version of this question is still open (at least, to our knowledge).

Question 4.6. Can an aperiodic erasing morphism preserve dendricity for one language? More precisely, can we find an erasing morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ and a dendric language over $\mathcal{A}$ whose image is an aperiodic dendric language over $\mathcal{B}$?

In Section 4.2, we also recalled that if σ is a dendric return morphism for a letter, then it is tame and therefore admits a decomposition into elementary morphisms. Moreover, by Proposition 4.45, all of the intermediary morphisms of this decomposition are return morphisms for a set of letters. It is likely that a similar result exists for dendric return morphisms for a word however the precise statement of the general version of Proposition 4.45 is unclear.

Question 4.7. Let γ be an elementary morphism and let τ be a tame morphism such that $\sigma = \gamma \circ \tau$ is a return morphism for the set S. Is τ a return morphism for a set S' and can we determine S' based on γ and S?

A positive answer would lead to the description of an automaton accepting the elementary decompositions of the tame return morphism for a given set S, as in Corollary 4.46.

We end with a question that we have already mentioned and discussed in Section 4.5. It was first asked by Dolce and Perrin in [DP21].

Question 4.8. Is the family of eventually dendric shift spaces stable under topological factorization? Equivalently, is eventual dendricity preserved when applying a letter-to-letter morphism?

Chapter 5

S-adic characterizations

The morphisms L_0 and L_1 presented in Subsection 4.2.1 do not only generate Sturmian morphisms (with their "R" counterparts) but they also generate the Sturmian languages. Indeed, we have the following well-known result (see [Fog02] for example).

Proposition 5.1. *A language $\mathcal{L} \subseteq \{0,1\}^*$ is recurrent Sturmian if and only if there exists a non-eventually constant sequence $(a_n)_{n\geq 0} \in \{0,1\}^{\mathbb{N}}$ such that*

$$\mathcal{L} = \bigcup_{n\geq 0} \mathrm{Fac}(L_{a_0} \cdots L_{a_n}(\{0,1\}))$$

where

$$L_0\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \end{cases} \quad \textit{and} \quad L_1\colon \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \end{cases} .$$

The sequence $(L_{a_n})_{n\geq 0}$ is then called an *S-adic representation* of $\mathcal{L}$. More generally, an *S-adic representation of a language* $\mathcal{L} \subseteq \mathcal{A}_0^*$ is a sequence of morphisms $\sigma_n\colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*$, $n \geq 0$, such that

$$\mathcal{L} = \bigcup_{n\geq 0} \mathrm{Fac}(\sigma_0 \cdots \sigma_n(\mathcal{A}_{n+1})).$$

If we moreover want to emphasize the fact that the morphisms σ_n, $n \geq 0$, all belong in some family $\mathfrak{S}$, we talk about $\mathfrak{S}$-adic representations.

While S-adic representations are mostly studied from the symbolic dynamics viewpoint due to their historical link with Bratelli-Vershik diagrams, we chose to speak here in terms of languages to be coherent with the rest of this work.

If all the morphisms of an S-adic representation are equal, we naturally recover the classical notion of a morphism generating a language. Observe that, in the definition of a language generated by a morphism (Definition 1.11), we chose to restrict ourselves to languages generated by *primitive* morphisms. Similarly, if $(\sigma_n)_{n\geq 0}$ is an S-adic representation of a language $\mathcal{L}$, we have some natural conditions on the growth of letters under the morphisms $(\sigma_0 \cdots \sigma_n)_{n\geq 0}$ since $\mathcal{L}$ is infinite. We do not detail these restrictions here since, as we mostly consider uniformly recurrent languages in this chapter, we will often restrict ourselves to *primitive* S-adic representations. This is detailed in Section 5.1.

In this chapter, not only are we interested in S-adic representations, we more specifically want to obtain results similar to Proposition 5.1 for dendric related languages. Indeed, Proposition 5.1 shows how S-adic representations can be used to characterize a family of languages, namely the recurrent Sturmian languages. Such a result is naturally called an *S-adic characterization*.

The Sturmian languages are not the only ones to admit a known S-adic characterization. We present some of these results in Section 5.1.

The main goal of this chapter is however to characterize (recurrent) dendric and eventually dendric languages. Therefore, in Section 5.2, we give a constructive characterization of the $\mathfrak{S}$-adic representations generating (eventually) dendric languages, for any set $\mathfrak{S}$ of return morphims for a word. These results were first stated for ternary alphabets in [GLL22] and for general alphabets in [GL22].

We then use the characterization obtained in Section 5.2 to give S-adic characterizations of some dendric related families in Section 5.3. Finally, we show how this can be used to decide dendricity of uniformly recurrent morphic languages in Section 5.4.

5.1 Quick overview of known S-adic characterizations

We present here several S-adic characterizations of known families of languages. The goal is to show the variety of results both in terms of families of languages and of conditions on the S-adic representations. This does not however claim to be an exhaustive list.

Like Sturmian languages, Arnoux-Rauzy languages admit a well-known characterization which naturally generalizes Proposition 5.1.

Proposition 5.2 (Arnoux–Rauzy [AR91], Justin–Pirillo [JP02])**.** *A language $\mathcal{L} \subseteq \mathcal{A}^*$ is recurrent Arnoux-Rauzy if and only if it has an S-adic representation $(L_{a_n})_{n\geq 0}$ where $(a_n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$ is a sequence in which each letter of $\mathcal{A}$ appears infinitely many times.*

We see here that the S-adic representations only use a finite set of morphisms: the Arnoux-Rauzy morphisms presented in Section 4.2. Moreover, there is a global condition (each Arnoux-Rauzy morphism appears infinitely many times in the S-adic representation) which essentially ensures primitivity.

Definition 5.3. A sequence $(\sigma_n \colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*)_{n\geq 0}$ of morphisms is *primitive* if, for all $n \geq 0$, there exists $K \geq 1$ such that all the letters of $\mathcal{A}_n$ appear in $\sigma_n \cdots \sigma_{n+K-1}(a)$ for all $a \in \mathcal{A}_{n+K}$.

Since we mostly consider uniformly recurrent languages, the S-adic representations presented in this section and in the following ones are primitive. In fact, it is well-known that if a language admits a primitive S-adic representation, then it is uniformly recurrent. The converse is also true as we will recall in Proposition 5.7.

Another classical S-adic characterization was given by Durand in [Dur00, Dur03]. It concerns the family of *linearly recurrent languages*, i.e., the languages $\mathcal{L}$ for which there exists a constant K such that each $u \in \mathcal{L}$ is a factor of all $v \in \mathcal{L}_{K|u|}$. It is a stronger version of uniform recurrence. We first state this characterization.

Proposition 5.4. *A language $\mathcal{L}$ is linearly recurrent if and only if it has a strongly primitive $\mathfrak{S}$-adic representation where $\mathfrak{S}$ is a finite set of proper morphisms.*

For clarity, we give a brief definition of the new concepts appearing in the previous statement.

- A sequence $(\sigma_n \colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*)_{n\geq 0}$ of morphisms is *strongly primitive* if there exists a constant K such that, for all n, all the letters of $\mathcal{A}_n$ appear in $\sigma_n \cdots \sigma_{n+K-1}(a)$ for all $a \in \mathcal{A}_{n+K}$. Note the difference with the classical notion of primitivity.
- A morphism $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$ is *proper* if there exist $b, c \in \mathcal{B}$ such that $\sigma(a) \in b\mathcal{B}^* \cap \mathcal{B}^* c$ for all $a \in \mathcal{A}$.

There is also an S-adic characterization of uniformly recurrent and aperiodic languages such that $s_n(\mathcal{L}) \leq 2$ given by Leroy in [Ler12]. While technical, the main idea behind this result is the study of the evolution of Rauzy graphs encoded into a "graph of graphs" in which we can read the S-adic representations. We will in fact do a similar construction in Section 5.2.

From a more dynamical viewpoint, we mention also the following S-adic characterization.

Proposition 5.5 (Donoso–Durand–Maass–Petite [DDMP21])**.** *A shift space X is conjugate to an expensive finite topological rank minimal Cantor system if and only if it has a primitive S-adic representation $(\sigma_n \colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*)_{n\geq 0}$ such that $\#\mathcal{A}_n$ is bounded and, for all $N \geq 0$, σ_N is recognizable on the shift space generated by $(\sigma_n)_{n>N}$.*

The most recent result on S-adic representations is the answer to a famous conjecture in the field called the *S-adic conjecture*. This conjecture is attributed to Host and claims the existence of an S-adic characterization of languages of complexity in $O(n)$. Espinoza recently proved it in [Esp23] in the case of uniformly recurrent languages.

Proposition 5.6. *A uniformly recurrent language $\mathcal{L}$ satisfies $p_{\mathcal{L}}(n) \in O(n)$ if and only if it has an S-adic representation $(\sigma_n \colon \mathcal{A}_{n+1}^* \to \mathcal{A}_n^*)_{n\geq 0}$ satisfying the following conditions:*

- *there exists c_1 such that $\|\sigma_n\| \leq c_1$ for all $n \geq 0$,*
- *there exists c_2 such that $\|\sigma_0 \cdots \sigma_n\| \leq c_2 \cdot \min_{a\in\mathcal{A}_{n+1}} |\sigma_0 \cdots \sigma_n(a)|$ for all $n \geq 0$,*
- *there exists c_3 such that, for all $n \geq 0$, there exist a set $C_n \subseteq \mathcal{A}_0^*$ of c_3 words such that $\sigma_0 \cdots \sigma_n(\mathcal{A}_{n+1}) \subseteq \bigcup_{w\in C_n} \{w\}^*$.*

An S-adic characterization also sometimes naturally follows from the definition of the languages as is the case for families related to continued fraction algorithms. We for example mention the uniformly recurrent Cassaigne languages (or sequences) which are known to be dendric [CLL22].

5.2 The case of (eventually) dendric languages

We now turn to the search of an S-adic characterization of dendric and eventually dendric languages. A classical method to obtain S-adic representations of a language is the iterated use of derivation. It was first implicitly mentioned by Durand for the study of substitutive sequences [Dur98]. Namely we have the following now folklore result.

Proposition 5.7. *Let $\mathcal{L}$ be a uniformly recurrent language. If the sequence $(\mathcal{L}^{(n)})_{n\geq 0}$ of languages and the sequence $(\sigma_n)_{n\geq 0}$ of morphisms are such that $\mathcal{L}^{(0)} = \mathcal{L}$ and for all $n \geq 0$, there exists $w^{(n)} \in \mathcal{L}^{(n)} \setminus \{\varepsilon\}$ such that $\mathcal{L}^{(n+1)} = D_{w^{(n)}}(\mathcal{L}^{(n)})$ and σ_n is the morphism such that $\mathcal{L}^{(n)} = \sigma_n(\mathcal{L}^{(n+1)})$, then $(\sigma_n)_{n\geq N}$ is a primitive S-adic representation of $\mathcal{L}^{(N)}$ for all $N \geq 0$.*

In fact, the construction of Proposition 5.7 corresponds to the S-adic representations used in the S-adic characterization of recurrent Sturmian languages (Proposition 5.1) and recurrent Arnoux-Rauzy languages (Proposition 5.2) for well-chosen words $w^{(n)}$ (more precisely, when $w^{(n)}$ is the unique bispecial letter of $\mathcal{L}^{(n)}$).

In the case where the language $\mathcal{L}$ is eventually dendric of threshold N, the intermediary languages $\mathcal{L}^{(n)}$ of Proposition 5.7 are eventually dendric as well by Theorem 3.42, and $\mathcal{L}^{(n)}$ is even dendric for all $n \geq N$. Because of the number of return words given in Corollary 3.32, we can also assume that all $\mathcal{L}^{(n)}$, $n \geq N$, are on the same alphabet.

In particular, this implies that every recurrent dendric language has an S-adic representation using exclusively dendric return morphisms for a word (recall that a return morphism for a word is dendric if the initial factors are dendric). However, not every sequence of dendric return morphisms for a word generates a dendric language. The goal of this section is to characterize the sequences that do.

The main idea is based on the following observation: if $\mathcal{L}$ is dendric and $(\sigma_n)_{n\geq 0}$ is an S-adic representation built as in Proposition 5.7, then for all n, σ_n is a return morphism for a word and both $\mathcal{L}^{(n+1)}$ and $\sigma_n(\mathcal{L}^{(n+1)}) = \mathcal{L}^{(n)}$ are dendric. Therefore, σ_n satisfies some conditions related to $G^L(\mathcal{L}^{(n+1)})$ and $G^R(\mathcal{L}^{(n+1)})$ given by Proposition 4.82.

In Subsection 5.2.1, we therefore take a closer look at the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ and their evolution when applying a return morphism. Using this, we then build a graph in Subsection 5.2.2 and show that it can be used to characterize the sequences of return morphisms generating recurrent (eventually) dendric languages. This graph can be simplified and this is what we do in Subsection 5.2.3.

In this section, we will often say "return morphism" instead of "return morphism for a word" as we exclusively work with these morphisms.

5.2.1 Back to the graphs G^L and G^R

Let us recall the statement of Proposition 4.82.

Proposition 4.82. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for w and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The extended images of all $v \in \mathcal{L}$ are dendric in $\sigma(\mathcal{L})$ if and only if the following conditions are satisfied:*

- *for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_s^L)$ is connected;*
- *for all $p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)$, the subgraph of $G^R(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\phi_p^R)$ is connected.*

The image $\sigma(\mathcal{L})$ is then dendric if and only if, moreover, σ is a dendric return morphism for a word.

While we can easily check if the conditions of this result are satisfied by a given return morphism σ and some given graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ for some language $\mathcal{L}$, the more complicated part is finding these two graphs corresponding to a given language $\mathcal{L}$.

Luckily, the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ have two major properties regarding return morphisms. The first one, as we saw in Proposition 4.82, is that they can be used to characterize dendric languages whose images under a given return morphism are dendric. The second one, as we will now show, is that their evolution when applying a return morphism is well understood. In other words, there exists a constructive method to obtain $G^L(\sigma(\mathcal{L}))$ from $G^L(\mathcal{L})$ and $G^R(\sigma(\mathcal{L}))$ from $G^R(\mathcal{L})$. This is described in the following definition.

Definition 5.8. Let G be the multi-clique $G_{\mathcal{A}}(\{C_1, \dots, C_k\})$ (see Definition 2.48) and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for a word w. The *left image* of G under σ is the multi-clique

$$\sigma^L(G) = G_{\mathcal{B}}(\{\varphi_s^L(C_i) : i \le k, s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))\})$$

and the *right image* of G under σ is the multi-clique

$$\sigma^R(G) = G_{\mathcal{B}}(\{\phi_p^R(C_i) : i \le k, p \in w\mathcal{B}^* \cap \mathrm{Pref}^*(\sigma(\mathcal{A})w)\}).$$

Recall that $\varphi_s^L(a) = b$ if $bs \in \mathrm{Suff}(\sigma(a))$ and $\phi_p^R(a) = b$ if $pb \in \mathrm{Pref}(\sigma(a)w)$.

Note that the left and right images do not depend on the choice of the word w for which σ is a return morphism by Proposition 4.33.

Example 5.9. Let $\mathcal{A} = \{0, 1, 2, 3\}$ and σ be the morphism of Example 4.83, i.e., $\sigma(0) = 20$, $\sigma(1) = 20221$, $\sigma(2) = 202$ and $\sigma(3) = 20203$. It is a dendric

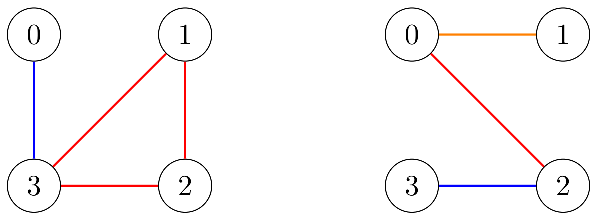


Figure 5.1: The graphs $G = G(\{\{0,3\},\{1,2,3\}\})$ (on the left) and $\sigma^R(G)$ (on the right) for the morphism σ of Example 5.9.

return morphism for 202. Since the only word $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$ such that $\#\varphi_s^L(\mathcal{A}) \geq 2$ is ε and φ_ε^L is the identity on $\mathcal{A}$, we have $\sigma^L(G) = G$ for any multi-clique G since the cliques of size 1 do not impact the resulting graph.

Let us turn to the right images. Using Example 4.83, the only words $p \in 202\mathcal{A}^*$ such that $\#\phi_p^R(\mathcal{A}) \geq 2$ are 202, 2020 and 2022 and the associated partial maps on $\mathcal{A}$ are given by

$$\phi_{202}^R\colon \begin{cases} 0,3 \mapsto 0 \\ 1,2 \mapsto 2 \end{cases}, \quad \phi_{2020}^R\colon \begin{cases} 0 \mapsto 2 \\ 3 \mapsto 3 \end{cases} \quad \text{and} \quad \phi_{2022}^R\colon \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 0 \end{cases}.$$

Observe that if C_i contains at most one element in the domain of ϕ_p^R, then $\phi_p^R(C_i)$ contains at most one element and does not impact the right image of a graph. Therefore, if $G = G_{\mathcal{A}}(\{\{0,3\},\{1,2,3\}\})$, then we have

$$\begin{aligned} \sigma^R(G) &= G_{\mathcal{A}}\left(\{\phi_{202}^R(\{0,3\}), \phi_{2020}^R(\{0,3\}), \phi_{202}^R(\{1,2,3\}), \phi_{2022}^R(\{1,2,3\})\}\right) \\ &= G_{\mathcal{A}}(\{\{2,3\},\{0,2\},\{0,1\}\}). \end{aligned}$$

These graphs are represented in Figure 5.1.

We now prove that this construction is indeed what allows us to obtain $G^L(\sigma(\mathcal{L}))$ based on $G^L(\mathcal{L})$ (resp., $G^R(\sigma(\mathcal{L}))$ based on $G^R(\mathcal{L})$).

Proposition 5.10. *Let $\mathcal{L}$ be an eventually dendric language over $\mathcal{A}$ and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ a return morphism for a word. If $\mathcal{L}' = \sigma(\mathcal{L})$, then $G^L(\mathcal{L}') = \sigma^L(G^L(\mathcal{L}))$ and $G^R(\mathcal{L}') = \sigma^R(G^R(\mathcal{L}))$.*

Proof. Let N be such that $\mathcal{L}$ is eventually ordinary of threshold N and let w be such that σ is a return morphism for w. By Proposition 4.93, $\mathcal{L}'$ is eventually dendric of threshold at most $M' := \|\sigma\|(N+1) + |w| - 1$. Let $M \geq M'$ be such that $\mathcal{L}'$ is eventually ordinary of threshold at most $M - \|\sigma\|$.

Let us prove the link between $G^L(\mathcal{L})$ and $G^L(\mathcal{L}')$, the proof for $G^R(\mathcal{L}')$ is symmetric. By Proposition 2.61, $G^L(\mathcal{L})$ (resp., $G^L(\mathcal{L}')$) is constructed with the left special elements in $\mathcal{L}_N$ (resp., $\mathcal{L}'_M$). Let

$$D = \{(s,v) \in \mathcal{B}^* \times \mathcal{L}_N : \#\varphi_s^L(E^L_{\mathcal{L}}(v)) \geq 2\}.$$

By definition, we then have

$$\begin{aligned}\sigma^L(G^L(\mathcal{L})) &= \sigma^L\left(G_{\mathcal{A}}(\{E^L_{\mathcal{L}}(v) : v \in \mathcal{L}_N \text{ left special}\})\right) \\ &= G_{\mathcal{B}}(\{\varphi_s^L(E^L_{\mathcal{L}}(v)) : (s,v) \in D\}).\end{aligned}$$

On the other hand, we have

$$G^L(\mathcal{L}') = G_{\mathcal{B}}(\{E^L_{\mathcal{L}'}(u) : u \in \mathcal{L}'_M \text{ left special}\}).$$

To conclude that $\sigma^L(G^L(\mathcal{L})) = G^L(\mathcal{L}')$, it then suffices to find a bijection f between the length-M left special words of $\mathcal{L}'$ and the elements of D such that, if $f(u) = (s,v)$, then

$$E^L_{\mathcal{L}'}(u) = \varphi_s^L(E^L_{\mathcal{L}}(v)).$$

We claim that the application f defined as follows satisfies these properties: for any left special $u \in \mathcal{L}'_M$, $f(u)$ is the pair $(s_u, v) \in \mathcal{B}^* \times \mathcal{L}_N$ where (s_u, v_u, p_u) is the triplet associated with u by Proposition 4.51 (i.e., u is an extended image of v_u and $u = s_u\sigma(v_u)p_u$) and v is the length-N prefix of v_u. Observe that u is indeed an extended image of a word of length at least N since $M \geq M' = \|\sigma\|(N+1) + |w| - 1$.

Note first that, if $f(u) = (s,v)$, then $s\sigma(v)w$ is a prefix of u by definition. Therefore, by Proposition 4.76 on the extensions of extended images,

$$E^L_{\mathcal{L}'}(u) \subseteq E^L_{\mathcal{L}'}(s\sigma(v)w) = \varphi_s^L(E^L_{\mathcal{L}}(v))$$

thus (s,v) is in D since u is left special.

Let us show that the application f is injective. If $f(u) = f(u') = (s,v)$, then $s_u = s = s_{u'}$ and v is a prefix of both v_u and $v_{u'}$. However, as u and u' are left special, v_u and $v_{u'}$ must be left special. As they have a common length-N prefix and $\mathcal{L}$ is eventually ordinary of threshold N, the only possibility is that one is prefix of the other. Let us assume without loss of generality that v_u is a prefix of $v_{u'}$. Then $s\sigma(v_u)w$ is a prefix of both u and u'. Since $u = s\sigma(v_u)p_u$ is of length M, $s\sigma(v_u)w$ is a left special factor of length at least $M - \|\sigma\|$. Since $\mathcal{L}'$ is eventually ordinary of threshold at most $M - \|\sigma\|$, $s\sigma(v_u)w$ is prefix of a unique left special factor of length M. We then conclude that $u = u'$ and that f is injective.

We now prove the surjectivity. For any $(s, v) \in D$, v is in particular left special so, by definition of N, there exists $v' \in v\mathcal{A}^*$ left special such that $|s\sigma(v')w| \geq M$ and $E^L_{\mathcal{L}}(v') = E^L_{\mathcal{L}}(v)$. If u is the length-M prefix of $s\sigma(v')w$, then $s_u = s$ and v_u is a prefix of v' of length at least N. Thus $s\sigma(v)w$ is a prefix of u. In addition, by Proposition 4.76, we have

$$E^L_{\mathcal{L}'}(u) \supseteq E^L_{\mathcal{L}'}(s\sigma(v')w) = \varphi^L_s(E^L_{\mathcal{L}}(v')) = \varphi^L_s(E^L_{\mathcal{L}}(v)).$$

In particular, u is left special by definition of D, so $f(u)$ is well-defined and $f(u) = (s, v)$. This proves that f is surjective. Moreover, we now have shown both inclusions thus $E^L_{\mathcal{L}'}(u) = \varphi^L_s(E^L_{\mathcal{L}}(v))$ if $f(u) = (s, v)$. $\square$

This result will play a major role in the next subsection but we first turn to another application of Proposition 5.10. Indeed, it can be used to prove the characterization of all possible graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ for dendric languages $\mathcal{L}$. We showed in Section 2.4 that, if $\mathcal{L}$ is dendric, then $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are two acyclic for the coloring and connected multi-cliques. We now show that these are the only restrictions on the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$, i.e., any pair of acyclic for the coloring and connected multi-cliques corresponds to a dendric language. We first need the following lemma.

Lemma 5.11. *Let $C_1, \ldots, C_k \subseteq \mathcal{A}$ and let D, E of cardinality at least 2 be such that*

$$D \cup E = C_1 \quad \textit{and} \quad \#(D \cap E) = 1.$$

There exists a return morphism σ for a letter such that, for any eventually dendric language $\mathcal{L}$, if $G^L(\mathcal{L}) = G_{\mathcal{A}}(\{C_1, \ldots, C_k\})$, then $G^L(\sigma(\mathcal{L})) = G_{\mathcal{A}}(\{D, E, C_2, \ldots, C_k\})$ and $G^R(\sigma(\mathcal{L})) = G^R(\mathcal{L})$, and if moreover $\mathcal{L}$ is dendric, then $\sigma(\mathcal{L})$ is dendric.

Similarly, there exists a return morphism σ for a letter such that, for any eventually dendric language $\mathcal{L}$, if $G^R(\mathcal{L}) = G_{\mathcal{A}}(\{C_1, \ldots, C_k\})$, then $G^L(\sigma(\mathcal{L})) = G^L(\mathcal{L})$ and $G^R(\sigma(\mathcal{L})) = G_{\mathcal{A}}(\{D, E, C_2, \ldots, C_k\})$, and if moreover $\mathcal{L}$ is dendric, then $\sigma(\mathcal{L})$ is dendric.

Proof. Let us denote $\{c\} = D \cap E$. We first introduce a classification of the cliques C_i, $i \geq 2$. The cliques *of type 1* are recursively defined as follows:

- if $C_i \cap D \neq \emptyset$, $i \geq 2$, then C_i is of type 1,
- if there exists C_j, $j \geq 2$, of type 1 such that $C_i \cap C_j \neq \emptyset$, $i \geq 2$, then C_i is of type 1.

We then say that C_i, $i \geq 2$, is *of type 2* if it is not of type 1. This defines a partition of $\mathcal{A}$ into four parts: $\{c\}$,

$$V := \left(D \cup \bigcup_{C_i \text{ of type } 1} C_i \right) \setminus \{c\}, \quad V' := \left(E \cup \bigcup_{C_i \text{ of type } 2} C_i \right) \setminus \{c\}$$

and the set V'' of the remaining (necessarily isolated) vertices. We then define the morphism σ as follows: let $b \in D \setminus \{c\}$, then

$$\sigma \colon \begin{cases} c \mapsto c \\ a \mapsto ca & \text{if } a \in V \cup V'' \\ a \mapsto cab & \text{if } a \in V' \end{cases}.$$

It is a return morphism for c. Moreover, we directly see that the only right special letter in $\sigma(\mathcal{A})c$ is c which implies that σ is a dendric return morphism.

Assume now that we have an eventually dendric language $\mathcal{L}$ such that $G^L(\mathcal{L}) = G_{\mathcal{A}}(\{C_1, \dots, C_k\})$. We first show that $G^R(\sigma(\mathcal{L})) = G^R(\mathcal{L})$. We easily see that the only word $p \in c\mathcal{A}^*$ such that $\#\phi_p^R(\mathcal{A}) \geq 2$ is c. Since we then have $\phi_c^R = \mathrm{id}$ over $\mathcal{A}$, we conclude that $G^R(\sigma(\mathcal{L})) = G^R(\mathcal{L})$ by Proposition 5.10.

We now turn to $G^L(\sigma(\mathcal{L}))$. The only words $s \in \mathcal{A}^*$ such that $\#\varphi_s^L(\mathcal{A}) \geq 2$ are ε and b. Therefore, by Proposition 5.10, we know that

$$G^L(\sigma(\mathcal{L})) = G_{\mathcal{A}}(\{\varphi_\varepsilon^L(C_i) : i \leq k\} \cup \{\varphi_b^L(C_i) : i \leq k\}).$$

If C_i is of type 1 then $C_i \subseteq \{c\} \cup V$ so $\#\varphi_b^L(C_i) \leq 1$ and $\varphi_\varepsilon^L(C_i) = C_i$. If C_i is of type 2 then $C_i \subseteq V'$ so $\#\varphi_\varepsilon^L(C_i) = 1$ and $\varphi_b^L(C_i) = C_i$. We now look at C_1. By definition, we have

$$\varphi_\varepsilon^L(C_1) = D \quad \text{and} \quad \varphi_b^L(C_1) = E$$

which shows that $G^L(\sigma(\mathcal{L})) = G_{\mathcal{A}}(\{D, E, C_2, \dots, C_k\})$.

Assume now that, moreover, $\mathcal{L}$ is dendric and let us show that $\sigma(\mathcal{L})$ is dendric. We already know that σ is a dendric morphism so let us turn to the other conditions of the characterization of dendric images (Proposition 4.82). We first look at the condition on $G^R(\mathcal{L})$. For all $p \in c\mathcal{A}^*$, the set $d_p = \{a \in \mathcal{A} : p \in \mathrm{Pref}^*(\sigma(a)c)\}$ is either $\mathcal{A}$ (if $p = c$) or contains at most one element. This shows that the second condition of Proposition 4.82 is directly satisfied.

We now turn to the condition on $G^L(\mathcal{L})$. It suffices to prove that the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_b^L) = \{a \in \mathcal{A} : b \in$

$\mathrm{Suff}^*(\sigma(a))\} = V' \cup \{b\}$ is connected. Let $e \in E \setminus \{c\}$, observe that $e \in V'$. We show that any $d \in V' \cup \{b\}$ is connected to e by a path entirely in $V' \cup \{b\}$. Clearly, if $d \in E$ or $d = b$, then $d \in C_1$ so $\{d, e\}$ is an edge of $G^L(\mathcal{L})$. We now turn to the other vertices. Since $\mathcal{L}$ is dendric, the graph $G^L(\mathcal{L})$ is acyclic for the coloring and connected by Corollary 2.44. This shows that d is connected to e. Moreover, since $d \in C_i$ of type 2, we can choose the path between d and e so that it does not use any vertex of D or of any clique of type 1. We then conclude that the subgraph of $G^L(\mathcal{L})$ generated by $V' \cup \{b\}$ is connected, so the conditions of Proposition 4.82 are satisfied and $\sigma(\mathcal{L})$ is dendric.

For the case where $G^R(\mathcal{L}) = G_{\mathcal{A}}(\{C_1, \ldots, C_k\})$, the proof is symmetric but we consider the return morphism

$$\sigma\colon \begin{cases} c \mapsto c & \\ a \mapsto ca & \text{if } a \in V \cup V'' \\ a \mapsto cba & \text{if } a \in V' \end{cases}$$

instead. □

Proposition 5.12. *Let G, G' be two graphs whose vertices are the elements of $\mathcal{A}$. There exists a dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $G^L(\mathcal{L}) = G$ and $G^R(\mathcal{L}) = G'$ if and only if the graphs G and G' are acyclic for the coloring and connected multi-cliques.*

Proof. If there exists such a dendric language, then G and G' are acyclic for the coloring and connected multi-cliques by Corollary 2.44.

Assume now that G is an acyclic for the coloring and connected multi-clique. By Remark 2.52, G can be obtained by applying a succession of splittings as in Lemma 2.51 (or Lemma 5.11) to the complete unicolor graph. Let $\mathcal{L}$ be an Arnoux-Rauzy language over $\mathcal{A}$. By definition, $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are complete unicolor graphs over $\mathcal{A}$. We can then iterate Lemma 5.11 to obtain a (return) morphism σ such that $\sigma(\mathcal{L})$ is dendric, $G^L(\sigma(\mathcal{L})) = G$ and $G^R(\sigma(\mathcal{L}))$ is the complete unicolor graph.

Using the same reasoning for G', we can find a (return) morphism τ such that $\tau(\sigma(\mathcal{L}))$ is dendric, $G^L(\tau(\sigma(\mathcal{L}))) = G$ and $G^R(\tau(\sigma(\mathcal{L}))) = G'$, which ends the proof. □

Example 5.13. Let us build a dendric language $\mathcal{L}$ such that $G^L(\mathcal{L}) = G(\{\{0,1,2,3\}\})$ and $G^R(\mathcal{L}) = G(\{\{0,1\},\{0,2\},\{2,3\}\})$. Let $\mathcal{L}'$ be an Arnoux-Rauzy language over $\{0,1,2,3\}$. We first build a dendric language $\mathcal{L}''$ such that $G^L(\mathcal{L}'') = G(\{\{0,1,2,3\}\})$ and $G^R(\mathcal{L}'') = G(\{\{0,1\},\{0,2,3\}\})$. To do so, we can define the morphism

$$\sigma\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 021 \\ 2 \mapsto 02 \\ 3 \mapsto 03 \end{cases}.$$

Using the notations of the proof of Lemma 5.11, it would correspond to taking $D = \{0,2,3\}$, $E = \{0,1\}$ and $b = 2$. We then take $\mathcal{L}'' = \sigma(\mathcal{L}')$ which satisfies the desired properties. We now define

$$\tau\colon \begin{cases} 0 \mapsto 20 \\ 1 \mapsto 21 \\ 2 \mapsto 2 \\ 3 \mapsto 203 \end{cases}.$$

This time, it corresponds to $D = \{0,2\}$, $E = \{2,3\}$ and $b = 0$. Finally, we take $\mathcal{L} = \tau(\mathcal{L}'') = \tau \circ \sigma(\mathcal{L}')$. It is a dendric language such that $G^L(\mathcal{L}) = G(\{\{0,1,2,3\}\})$ and $G^R(\mathcal{L}) = G(\{\{0,1\},\{0,2\},\{2,3\}\})$. Observe that

$$\tau \circ \sigma\colon \begin{cases} 0 \mapsto 20 \\ 1 \mapsto 20221 \\ 2 \mapsto 202 \\ 3 \mapsto 20203 \end{cases}$$

is the morphism of Example 2.53. However, the sequence of splitting of cliques to obtain $G^R(\mathcal{L})$ is different.

5.2.2 A graph of graphs

Based on Subsection 5.2.1, not only do we want σ_n to satisfy the conditions related to $G^L(\mathcal{L}^{(n+1)})$ and $G^R(\mathcal{L}^{(n+1)})$ so that $\mathcal{L}^{(n)} = \sigma_n(\mathcal{L}^{(n+1)})$ is dendric, but it must also give the correct graphs $G^L(\mathcal{L}^{(n)})$ and $G^R(\mathcal{L}^{(n)})$. This is formalized below using the notion of left (resp., right) valid triplets.

Definition 5.14. Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for a word w. The triplet (G', σ, G) is *left (resp., right) valid* if the following conditions are satisfied

1. σ is a dendric return morphism;

2. G is an acyclic for the coloring and connected multi-clique;
3. for all $x \in \mathcal{B}^*$, the subgraph of G generated by the vertices in $\mathrm{dom}(\varphi_x^L)$ (resp., $\mathrm{dom}(\phi_x^R)$) is connected;
4. $G' = \sigma^L(G)$ (resp., $G' = \sigma^R(G)$) is an acyclic for the coloring and connected multi-clique.

We would like to stress the fact that the first element of the triplet is the left (resp., right) image of the third one and not the converse. While it may seem unnatural at first, it will soon make perfect sense.

Example 5.15. Let σ be the morphism such that $\sigma(0) = 20$, $\sigma(1) = 20221$, $\sigma(2) = 202$ and $\sigma(3) = 20203$. By Examples 4.83 and 5.9, the triplet (G, σ, G) is left valid for any acyclic for the coloring and connected multi-clique G, and $(G(\{\{0,1\},\{0,2\},\{2,3\}\}), \sigma, G(\{\{0,3\},\{1,2,3\}\}))$ is right valid.

By definition, we directly have the following result which essentially implies that valid triplets do what is asked of them.

Proposition 5.16. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ be a return morphism for a word, let $\mathcal{L}$ be a dendric language over $\mathcal{A}$ and $\mathcal{L}' = \sigma(\mathcal{L})$. Then $\mathcal{L}'$ is dendric if and only if $(G^L(\mathcal{L}'), \sigma, G^L(\mathcal{L}))$ is left valid and $(G^R(\mathcal{L}'), \sigma, G^R(\mathcal{L}))$ is right valid.*

Proof. Since $\mathcal{L}$ is dendric, then $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are acyclic for the coloring and connected multi-cliques by Corollary 2.44. Moreover, since σ is a return morphism, we have $G^L(\mathcal{L}') = \sigma^L(G^L(\mathcal{L}))$ and $G^R(\mathcal{L}') = \sigma^R(G^R(\mathcal{L}))$ by Proposition 5.10.

By Proposition 4.82, $\sigma(\mathcal{L})$ is dendric if and only if σ is dendric and $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$) satisfies the third item of Definition 5.14 for left (resp., right) valid triplets. The fourth items are then satisfied by Corollary 2.44. $\square$

In the context of S-adic representations, we will mostly be interested in the following corollary.

Corollary 5.17. *Let $\mathcal{L}$ be a dendric language and $(\sigma_n)_{n\geq 0}$ be an S-adic representation of $\mathcal{L}$ where each σ_n is a return morphism. For all $N \geq 0$, let also $\mathcal{L}^{(N)}$ be the language generated by $(\sigma_n)_{n\geq N}$. Then for all $n \geq 0$, $(G^L(\mathcal{L}^{(n)}), \sigma_n, G^L(\mathcal{L}^{(n+1)}))$ is left valid and $(G^R(\mathcal{L}^{(n)}), \sigma_n, G^R(\mathcal{L}^{(n+1)}))$ is right valid.*

The previous result can be interpreted using graphs. Indeed, the fact that $(G^L(\mathcal{L}^{(n)}), \sigma_n, G^L(\mathcal{L}^{(n+1)}))$ is left valid for all n means that the sequence

$(\sigma_n)_{n\geq 0}$ labels a path going through the sequence of vertices $(G^L(\mathcal{L}^{(n)}))_{n\geq 0}$ in a particular graph. The same can be said using the graphs $G^R(\mathcal{L}^{(n)})$. We define below these two graphs of graphs.

Definition 5.18. Let $\mathfrak{S}$ be a set of return morphisms from $\mathcal{A}^*$ to $\mathcal{A}^*$. The graph $\mathcal{G}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$) is the graph whose vertices are the acyclic for the coloring and connected multi-cliques on $\mathcal{A}$ and such that there is an edge from G' to G labeled by $\sigma \in \mathfrak{S}$ if and only if (G', σ, G) is a left (resp., right) valid triplet.

If $\#\mathcal{A} = 2$, then $\mathcal{G}^L(\mathfrak{S})$ contains the unique vertex $G(\{\mathcal{A}\})$ and a loop over this vertex labeled by each dendric return morphism in $\mathfrak{S}$. In this particular case, we also have $\mathcal{G}^R(\mathfrak{S}) = \mathcal{G}^L(\mathfrak{S})$.

Remark 5.19. For any given return morphism, the set of corresponding left (resp., right) valid triplets is finite and computable. Therefore, if $\mathfrak{S}$ is finite, then the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ are finite and computable.

Indeed, we can easily determine if σ is dendric or not. If it is not, then there are no corresponding left (resp., right) valid triplets. If $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$ is dendric, then we can easily generate the acyclic for the coloring and connected acyclic multi-cliques on $\mathcal{A}$ since they correspond to particular simple graphs by Lemma 2.55 and Proposition 2.56. In particular, there are only finitely many of them. For each such multi-clique G, we can then check the third item of Definition 5.14 for left valid triplets and, if it is satisfied, compute $\sigma^L(G)$, which gives us the left valid triplet $(\sigma^L(G), \sigma, G)$.

Observe also that, once we know whether σ is dendric or not, this construction of all left valid triplets only depends on the set $\{\varphi_s^L : s \in \mathcal{A}^*, \#\varphi_s^L(\mathcal{A}) \geq 2\}$ of maps and not on the morphism σ. We stress the fact that only the maps matter, and not the corresponding words s.

We can similarly generate all the right valid triplets corresponding to σ using only the set $\{\phi_p^R : p \in \mathcal{A}^*, \#\phi_p^R(\mathcal{A}) \geq 2\}$.

We now show that, not only do the S-adic representations of dendric languages label infinite paths in these graphs, but these are the only S-adic representations to do so, and we therefore have an S-adic characterization. This is the main result of this section.

Theorem 5.20. *Let $\mathfrak{S}$ be a set of return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$. A language $\mathcal{L}$ having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma} = (\sigma_n)_{n\geq 0}$ is recurrent dendric if and only if $\boldsymbol{\sigma}$ is primitive and labels infinite paths in the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$.*

Proof. Assume that $\mathcal{L}$ is recurrent dendric. The morphisms of $\mathfrak{S}$ being return morphisms, the sequence $\boldsymbol{\sigma}$ is defined as in Proposition 5.7 and thus is primitive. By Corollary 5.17 and by definition of the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$, we directly conclude that $\boldsymbol{\sigma}$ labels infinite paths in the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$.

Assume now that $\boldsymbol{\sigma}$ is primitive and labels a path $(G_n^L)_{n\geq 0}$ in $\mathcal{G}^L(\mathfrak{S})$ and a path $(G_n^R)_{n\geq 0}$ in $\mathcal{G}^R(\mathfrak{S})$. Since $\boldsymbol{\sigma}$ is primitive, the language $\mathcal{L}$ is uniformly recurrent. We now show that it is dendric.

For all $N \geq 0$, let us denote $\mathcal{L}^{(N)}$ the language generated by $(\sigma_n)_{n\geq N}$. Recall that, by definition of an S-adic representation, $\mathcal{L}^{(n)} = \sigma_n(\mathcal{L}^{(n+1)})$ for all $n \geq 0$.

Let $u \in \mathcal{L}$. Since $\mathcal{L} = \sigma_0(\mathcal{L}^{(1)})$, either u is an initial factor (for σ_0) or it is an extended image of some $u^{(1)} \in \mathcal{L}^{(1)}$ and the extensions of u are entirely determined by the extensions of $u^{(1)}$ and σ_0 by Proposition 4.76. Moreover, $|u^{(1)}| < |u|$ so, by iterating this reasoning, there exist $k \geq 0$ and $u^{(k)} \in \mathcal{L}^{(k)}$ such that $u^{(k)}$ is an initial factor for σ_k. By Proposition 4.76, the extensions of u in $\mathcal{L}$ are then entirely determined by the extensions of $u^{(k)}$ in $\mathcal{L}^{(k)}$ and by the morphism $\sigma_0 \cdots \sigma_{k-1}$.

Since the extensions of u only depend on the morphism $\sigma_0 \cdots \sigma_{k-1}$ and the extensions of $u^{(k)}$, and the extensions of $u^{(k)}$ only depend on σ_k, to show that u is dendric in $\mathcal{L}$ it suffices to prove that there exists a language $\mathcal{L}'$ over $\mathcal{A}$ such that $\sigma_0 \cdots \sigma_k(\mathcal{L}')$ is dendric.

By definition of the vertices of the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ and by Proposition 5.12, there exists a dendric language $\mathcal{L}'$ such that $G^L(\mathcal{L}') = G_{k+1}^L$ and $G^R(\mathcal{L}') = G_{k+1}^R$. Since $(\sigma_n)_{0\leq n\leq k}$ labels a path in $\mathcal{G}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$) ending in G_{k+1}^L (resp., G_{k+1}^R), we deduce that $\sigma_0 \cdots \sigma_k(\mathcal{L}')$ is dendric by Proposition 5.16 and by definition of the edges of $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$.

We can then conclude that u is dendric for all $u \in \mathcal{L}$, which ends the proof that if $\mathcal{L}$ has a primitive $\mathfrak{S}$-adic representation labeling infinite paths in $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$, then $\mathcal{L}$ is recurrent dendric. □

Remark 5.21. We have the exact same characterization if $\mathfrak{S}$ contains (both or either) left and right return morphisms. It suffices to naturally define the edges labeled by right return morphisms: a right return morphism $\sigma \in \mathfrak{S}$ labels an edge from G to G' in $\mathcal{G}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$) if and only if the corresponding left return morphism τ labels an edge from G to G' in $\mathcal{G}^L(\{\tau\})$ (resp., $\mathcal{G}^R(\{\tau\})$). Indeed, for any word u, we have $w\sigma(u) = \tau(u)w$ if w is such that σ is a right return morphism for w. Therefore, for any language $\mathcal{L}$, $\sigma(\mathcal{L}) = \tau(\mathcal{L})$.

Observe that the path labeled by $\boldsymbol{\sigma}$ in $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ is not necessarily unique. Indeed, for example, a sequence of Arnoux-Rauzy morphisms in $\mathfrak{S}$ labels a path starting in any vertex of $\mathcal{G}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$).

Let us now turn to the case of eventually dendric languages. As explained in the introduction of this section, if $\boldsymbol{\sigma}$ (which only uses return morphisms) is an S-adic representation of a recurrent eventually dendric language, then there exists a suffix of $\boldsymbol{\sigma}$ which is an S-adic representation of a recurrent dendric language by Theorem 3.42. Moreover, this is a characterization by Corollary 4.94 so we directly obtain the following result.

Corollary 5.22. *For all $k \geq 1$, let $\mathcal{A}_k$ be an alphabet of size k and let $\mathfrak{S}$ be a set of return morphisms for a word such that, for all $\sigma \in \mathfrak{S}$, there exists $i, j \geq 1$ such that $\sigma \colon \mathcal{A}_i^* \to \mathcal{A}_j^*$, $\mathcal{A}_j$ being minimal. For each $k \geq 1$, let us denote $\mathfrak{S}_k$ the morphisms of $\mathfrak{S}$ going from $\mathcal{A}_k^*$ to $\mathcal{A}_k^*$.*

A language $\mathcal{L}$ having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$ is eventually dendric recurrent if and only if there exists a suffix $\boldsymbol{\sigma}'$ of $\boldsymbol{\sigma}$ which is primitive and labels infinite paths in $\mathcal{G}^L(\mathfrak{S}_k)$ and $\mathcal{G}^R(\mathfrak{S}_k)$ for some k.

Not only do we know that if $\mathcal{L}$ has an S-adic representation labeling infinite paths in the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$, then $\mathcal{L}$ is dendric, we can also explicitly give its graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$. This is the object of the following result.

Proposition 5.23. *Let $\mathfrak{S}$ be a set of return morphism for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$, let $\mathcal{L}$ be a recurrent dendric language having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma} = (\sigma_n)_{n \geq 0}$.*

1. *The sequence $\boldsymbol{\sigma}$ labels a path in $\mathcal{G}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$) starting in $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$).*
2. *If $\boldsymbol{\sigma}$ labels a path in $\mathcal{G}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$) starting in G, then G is a subgraph of $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$).*

Proof. The first claim is a direct consequence of Corollary 5.17. For the second claim, assume that $\boldsymbol{\sigma}$ labels the path $(G_n)_{n \geq 0}$ in $\mathcal{G}^L(\mathfrak{S})$ and let us show that G_0 is a subgraph of $G^L(\mathcal{L})$.

By Proposition 2.57, there exists N such that $G^L(\mathcal{L}) = G^L_N(\mathcal{L})$. As in the proof of Theorem 5.20, if u is a length-N element of $\mathcal{L}$, then its extensions are entirely determined by $\sigma_0 \cdots \sigma_k$ for some $k \leq N$. As it is true for all $u \in \mathcal{L}_N$, this implies that, for any language $\mathcal{L}'$, $G^L_N(\mathcal{L}) = G^L_N(\sigma_0 \cdots \sigma_N(\mathcal{L}'))$.

Let $\mathcal{L}'$ be a dendric language such that $G^L(\mathcal{L}') = G_{N+1}$. This exists by Proposition 5.12. By construction of the graph $\mathcal{G}^L(\mathfrak{S})$ and by Proposition 5.10, we have $G^L(\sigma_0 \cdots \sigma_N(\mathcal{L}')) = G_0$. Moreover, $G^L(\sigma_0 \cdots \sigma_N(\mathcal{L}'))$ is a subgraph of $G_N^L(\sigma_0 \cdots \sigma_N(\mathcal{L}'))$ so we conclude that G_0 is a subgraph of $G_N^L(\mathcal{L}) = G^L(\mathcal{L})$.

We similarly show that, if $\boldsymbol{\sigma}$ labels a path in $\mathcal{G}^R(\mathfrak{S})$ starting in G, then G is a subgraph of $G^R(\mathcal{L})$. $\square$

The previous result therefore states that, out of all the possible starting vertices for paths labeled by $\boldsymbol{\sigma}$ in $\mathcal{G}^L(\mathfrak{S})$, the graph $G^L(\mathcal{L})$ is the largest. Alternatively, $G^L(\mathcal{L})$ is the union of all possible starting vertices in the sense that its set of edges is the union of the sets of edges of the starting vertices. This second viewpoint will reappear in the following subsection. Of course, we have a similar interpretation for $G^R(\mathcal{L})$.

5.2.3 From graphs to trees

In the previous subsection, we defined the graphs $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ whose vertices are the acyclic for the coloring and connected multi-cliques. By Lemma 2.55, we have a bijection between these multi-cliques and the acyclicly colorable and connected graphs. Recall that, by Proposition 2.56, a graph is acyclicly colorable if and only if it is simple and for any two vertices a and b, if there is a cycle going through them, then (a, b) is an edge. In other words, we could have just as well said that the vertices of $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ are the acyclicly colorable and connected graphs. We will then abusively identify an acyclic for the coloring and connected multi-clique with its uncolored version and vice-versa.

In particular, among the acyclicly colorable and connected graphs, we have the trees. The purpose of this subsection is to show that these are the only important vertices of $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ in the sense that the corresponding subgraphs of $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ are sufficient to obtain Theorem 5.20.

Let us formally define these subgraphs.

Definition 5.24. Let $\mathfrak{S}$ be a set of return morphisms from $\mathcal{A}^*$ to $\mathcal{A}^*$. The graph $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$) is the graph whose vertices are the trees on $\mathcal{A}$ and such that there is an edge from T' to T labeled by $\sigma \in \mathfrak{S}$ if and only if (T', σ, T) is a left (resp., right) valid triplet.

We first need the following lemma. Observe that an acyclic for the coloring and connected multi-clique $G(\{C_1, \cdots, C_k\})$ is a tree if and only if $\#C_i \leq 2$ for all $i \leq k$.

Lemma 5.25. *Let (G', σ, G) be a left (resp., right) valid triplet. For any covering tree T' of G', there exists a covering tree T of G such that (T', σ, T) is left (resp., right) valid.*

Proof. We prove the result for a left valid triplet, the case of a right valid triplet is symmetric. Since G is a multi-clique, we set $C_1, \ldots, C_k$ such that $G = G(\{C_1, \ldots, C_k\})$. Then, as (G', σ, G) is left valid, one has

$$G' = \sigma^L(G) = G(\{\varphi_s^L(C_i) : i \leq k, s \in \mathcal{B}^*\})$$

where $\mathcal{B}$ is the image alphabet of σ.

If G and G' are trees, then $T' = G'$ and it suffices to take $T = G$. Assume thus that either G or G' is not a tree. We show that we can define a new left valid triplet (H', σ, H), where

- H is a connected strict subgraph of G (with the same vertices);
- T' is a covering tree of H'.

We can therefore iterate this construction as long as H or H' is not a tree. Since G is finite, this process must stop and this will therefore prove the existence of a covering tree T (given by the last graph H) of G such that (T', σ, T) is left valid.

We start by highlighting a word $s_0 \in \mathcal{B}^*$, some clique C_i, $i \leq k$, and two distinct vertices $a, b \in \varphi_{s_0}^L(C_i)$ as follows.

- If G' is not a tree, there exist $i \leq k$ and $s_0 \in \mathcal{B}^*$ such that $C = \varphi_{s_0}^L(C_i)$ contains at least 3 elements. The subgraph T'' of T' generated by the vertices of C is acyclic. Moreover, for any two vertices in C, any path connecting them in G' uses edges corresponding to C as G' is acyclic for the coloring. Thus the path connecting them in T' is in T'' and T'' is connected. This shows that T'' is a tree and thus, there exists a vertex $a \in C$ of degree 1 in T''. Let $b \in C$ be its (unique) neighbor in T''.
- If G' is a tree but G is not, there exists $i \leq k$ such that C_i contains at least 3 elements. Let $s_0 \in \mathcal{B}^*$ be the longest common suffix to all $\sigma(d)$, $d \in C_i$. By definition of s_0, the set $C = \varphi_{s_0}^L(C_i)$ contains at least two elements, but since G' is a tree, $\varphi_{s_0}^L(C_i)$ contains at most 2 elements. Therefore, let us write $C = \{a, b\}$ with b such that

$$\#(\{d \in C_i : \varphi_{s_0}^L(d) = b\}) \geq 2.$$

In both cases, let $c \in C_i$ be such that $\varphi_{s_0}^L(c) = b$ and let us denote

$$D = \{c\} \cup \{d \in C_i : \varphi_{s_0}^L(d) = a\}$$

and

$$E = (C_i \setminus D) \cup \{c\}.$$

We then define $\mathcal{C} = \{D, E, C_1, \dots, C_k\} \setminus \{C_i\}$, $H = G(\mathcal{C})$ and $H' = \sigma^L(H)$.

Let us prove that (H', σ, H) is left valid by checking the items of Definition 5.14.

1. As (G', σ, G) is left valid, σ is a dendric return morphism.

2. Since G is an acyclic for the coloring and connected multi-clique and $D \cup E = C_i$ and $D \cap E = \{c\}$, by Lemma 2.51, H is also an acyclic for the coloring and connected multi-clique.

3. We now show that for all $s \in \mathcal{B}^*$, the subgraph of H generated by the vertices in $\mathrm{dom}(\varphi_s^L) = \{a \in \mathcal{A} : \sigma(a) \in \mathcal{B}^+ s\}$ is connected. Let us denote H_s this graph and G_s the same graph defined starting from G. Since H is the subgraph of G obtained by removing the edges between $D \setminus \{c\}$ and $E \setminus \{c\}$, the same can be said about the link between H_s and G_s. Since (G', σ, G) is valid, the graph G_s is connected. Let us show that it is also the case of H_s by showing that, for any edge (d, e) in G_s but not in H_s, we have the path (d, c, e) in H_s. Indeed, by definition of D and E, if s is a proper suffix of both $\sigma(d)$ and $\sigma(e)$, then s is a suffix of s_0 so it is also a proper suffix of $\sigma(c)$. By definition of H, the edges (d, c) and (c, e) are then in H_s.

4. Finally, we show that H' is an acyclic for the coloring and connected multi-clique. Recall that by Proposition 5.10,

$$H' = G(\{\varphi_s^L(F) : F \in \mathcal{C}, s \in \mathcal{B}^*\}.$$

   We focus on $\varphi_s^L(D)$ and $\varphi_s^L(E)$, $s \in \mathcal{B}^*$. Observe that, by definition of D, if $\#\varphi_s^L(D) \geq 2$, then $s = s_0$ or as_0 is a suffix of s. In the first case, $\varphi_{s_0}^L(D) = \{a, b\}$ and in the second case, D contains all the elements of C_i whose images end with s so $\varphi_s^L(D) = \varphi_s^L(C_i)$. On the other hand, $\varphi_s^L(E) = \varphi_s^L(C_i)$ unless $s = s_0$ or as_0 is a suffix of s. In the first case, $\varphi_{s_0}^L(E) = \varphi_{s_0}^L(C_i) \setminus \{a\}$ and in the second case, $\varphi_{s_0}^L(E)$ is empty.

   This shows that, if $\mathcal{C}' = \{\varphi_s^L(C_j) : j \leq k, s \in \mathcal{B}^*\}$, then

$$H' = G\left(\left(\mathcal{C}' \setminus \{\varphi_{s_0}^L(C_i)\}\right) \cup \{\{a, b\}, \varphi_{s_0}^L(C_i) \setminus \{a\}\}\right).$$

By Lemma 2.51, the fact that G' is an acyclic for the coloring and connected multi-clique implies that H' also is.

We now show that the triplet (H', σ, H) satisfies the two desired conditions. Clearly, H is a subgraph of G with the same vertices, and it is connected since $H = H_\varepsilon$. Moreover, since $a \in \varphi_{s_0}^L(C_i)$, the set D contains at least two elements. The same can be said for E by definition of a and b. Indeed, if G' is not a tree, $\varphi_{s_0}^L(C_i)$ contains at least three elements, and if G' is a tree but G is not, there exists more than one element $d \in C_i$ such that $\varphi_{s_0}^L(d) = b$. This implies that H has strictly fewer edges than G.

We finally show that T' is still a subtree of H'. If G' is a tree, since H' is a connected subgraph of G', we directly have $T' = G' = H'$. If G' is not a tree, then a and b were chosen such that no edge of $\{a\} \times (\varphi_{s_0}^L(C_i) \setminus \{a, b\})$ is in T'. As these are the only edges lost when going from G' to H', this implies that T' is a subgraph of H'. $\square$

We now show a result similar to Theorem 5.20.

Theorem 5.26. *Let $\mathfrak{S}$ be a set of return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$. A language $\mathcal{L}$ having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$ is recurrent dendric if and only if $\boldsymbol{\sigma}$ is primitive and labels infinite paths in the graphs $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$.*

Proof. By Theorem 5.20, it suffices to prove that $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{G}^R(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$) contain the same infinite paths. Clearly, $\mathcal{T}^L(\mathfrak{S})$ is a subgraph of $\mathcal{G}^L(\mathfrak{S})$ so any path of $\mathcal{T}^L(\mathfrak{S})$ is a path of $\mathcal{G}^L(\mathfrak{S})$. Conversely, by Lemma 5.25, if $\boldsymbol{\sigma}$ labels a path in $\mathcal{G}^L(\mathfrak{S})$ starting in G, then $\boldsymbol{\sigma}$ labels a path in $\mathcal{T}^L(\mathfrak{S})$ starting in T for any covering tree T of G (such a tree exists since G is connected). We similarly prove the link between $\mathcal{G}^R(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$. $\square$

We naturally also have a characterization of eventual dendricity.

Corollary 5.27. *For all $k \geq 1$, let $\mathcal{A}_k$ be an alphabet of size k and let $\mathfrak{S}$ be a set of return morphisms for a word such that, for all $\sigma \in \mathfrak{S}$, there exists $i, j \geq 1$ such that $\sigma \colon \mathcal{A}_i^* \to \mathcal{A}_j^*$, $\mathcal{A}_j$ being minimal. For each $k \geq 1$, let us denote $\mathfrak{S}_k$ the morphisms of $\mathfrak{S}$ going from $\mathcal{A}_k^*$ to $\mathcal{A}_k^*$.*

A language $\mathcal{L}$ having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$ is eventually dendric recurrent if and only if there exists a suffix $\boldsymbol{\sigma}'$ of $\boldsymbol{\sigma}$ which is primitive and labels infinite paths in $\mathcal{T}^L(\mathfrak{S}_k)$ and $\mathcal{T}^R(\mathfrak{S}_k)$ for some k.

We can also deduce $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ from the paths labeled by $\boldsymbol{\sigma}$.

Proposition 5.28. *Let $\mathfrak{S}$ be a set of return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$, let $\mathcal{L}$ be a recurrent dendric language having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$.*

1. *For any covering tree T of $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$), $\boldsymbol{\sigma}$ labels a path in $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$) starting in T.*
2. *If $\boldsymbol{\sigma}$ labels a path in $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$) starting in T, then T is a covering tree of $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$).*

In particular, $G^L(\mathcal{L})$ (resp., $G^R(\mathcal{L})$) is the union of the starting vertices of all the paths labeled by $\boldsymbol{\sigma}$ in $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$), where the union of graphs is understood here as the graph obtained by taking the union of the sets of edges.

Proof. We prove the claims for $\mathcal{T}^L(\mathfrak{S})$, the proof is the same for $\mathcal{T}^R(\mathfrak{S})$. By Proposition 5.23, $\boldsymbol{\sigma}$ labels a path in $\mathcal{G}^L(\mathfrak{S})$ starting in $G^L(\mathcal{L})$ so, by Lemma 5.25, it labels a path in $\mathcal{T}^L(\mathfrak{S})$ starting in any covering tree of $G^L(\mathcal{L})$.

If $\boldsymbol{\sigma}$ labels a path in $\mathcal{T}^L(\mathfrak{S})$ starting in T, then this path is also in $\mathcal{G}^L(\mathfrak{S})$ so, by Proposition 5.23, T is a subgraph of $G^L(\mathcal{L})$. By definition of $\mathcal{T}^L(\mathfrak{S})$, T is then a covering tree of $G^L(\mathcal{L})$.

Since a connected graph is the union of its covering trees, the conclusion follows. □

Note that the vertices of $\mathcal{G}^L(\mathfrak{S})$, $\mathcal{G}^R(\mathfrak{S})$, $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$ only depend on the size of the alphabet $\mathcal{A}$ so we can study the sizes of these graphs for given values of $\#\mathcal{A}$.

The vertices of $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ are the (labeled) connected acyclicly colorable graphs (also known as block-graphs [Har63]) so the size of $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ as a function of $n = \#\mathcal{A}$ is given by the sequence A030019 on OEIS.

For the vertices of $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$ (i.e., the labeled trees with n vertices), there is a simple formula: n^{n-2} (see sequence A000272 on OEIS).

Since we want $\boldsymbol{\sigma}$ to label infinite paths in both $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^L(\mathfrak{S})$ and $\mathcal{G}^R(\mathfrak{S})$ (resp., $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$), it is equivalent to ask that it labels an infinite path in the product $\mathcal{G}^L(\mathfrak{S}) \times \mathcal{G}^R(\mathfrak{S})$ (resp., $\mathcal{T}^L(\mathfrak{S}) \times \mathcal{T}^R(\mathfrak{S})$) of these graphs.

Clearly, even for the smaller graph built using trees, the number of vertices grows so fast that, starting from alphabets of size 4, the graph becomes unpractical (see Table 5.1).

However, it is possible to reduce the number of vertices with a clever use of permutations. Indeed, the vertices of $\mathcal{T}^L(\mathfrak{S}) \times \mathcal{T}^R(\mathfrak{S})$ are pairs of labeled trees and some of these pairs are equivalent, in the sense that they are equal up to a permutation of the letters. The idea is then to only keep one vertex per equivalence class. To keep equivalent infinite paths, we must however increase the set of edge labels by composing the morphisms of $\mathfrak{S}$ with permutations.

Definition 5.29. Let $\mathfrak{S}$ be a set of return morphisms from $\mathcal{A}^*$ to $\mathcal{A}^*$. The graph $\mathcal{T}_p(\mathfrak{S})$ is a graph whose set of vertices contains one pair of labeled trees on $\mathcal{A}$ per equivalence class and such that there is an edge from (T'^L, T'^R) to (T^L, T^R) labeled by $\pi\sigma\psi$, $\sigma \in \mathfrak{S}$, π, ψ permutations on $\mathcal{A}$, if and only if $(T'^L, \pi\sigma\psi, T^L)$ is left valid and $(T'^R, \pi\sigma\psi, T^R)$ is right valid.

Proposition 5.30. *Let $\mathfrak{S}$ be a set of return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$ and let $\mathcal{L}$ be a language. If $\mathcal{L}$ has an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$, then $\mathcal{L}$ is recurrent dendric if and only if $\boldsymbol{\sigma}$ is primitive and there exists a sequence $(\pi_n)_{n\geq 0}$ of permutations on $\mathcal{A}$ such that $(\pi_n\sigma_n\pi_{n+1}^{-1})_{n\geq 0}$ labels a path in $\mathcal{T}_p(\mathfrak{S})$.*

This graph is particularly useful when the set $\mathfrak{S}$ is stable under composition by permutations since we then have the following characterization.

Proposition 5.31. *Let $\mathfrak{S}$ be a set of return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$ such that $\Sigma_{\mathcal{A}}\mathfrak{S}\Sigma_{\mathcal{A}} = \mathfrak{S}$ where $\Sigma_{\mathcal{A}}$ is the set of permutations over $\mathcal{A}$. A language $\mathcal{L}$ having an $\mathfrak{S}$-adic representation is recurrent dendric if and only if there exists a permutation π such that $\pi(\mathcal{L})$ has a primitive $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$ labeling an infinite path in $\mathcal{T}_p(\mathfrak{S})$.*

To give an idea of the improvement, in the case of a ternary alphabet, this allows to replace a 9-vertex graph by a graph with 2 vertices, as first done in [GLL22]. More generally, we give the graphs sizes for small alphabets in Table 5.1.

5.3 Some examples

In the previous section, we gave a method to characterize the recurrent dendric languages having an $\mathfrak{S}$-adic representation for any given set $\mathfrak{S}$ of return morphisms. We now use this result to characterize particular families of dendric languages by finding a suitable set $\mathfrak{S}$ of morphisms.

We first look at the original question on this topic which is the characterization of all dendric languages on a given alphabet in Subsection 5.3.1.

n	1	2	3	4	5
$\mathcal{G}^L(\mathfrak{S}) \times \mathcal{G}^R(\mathfrak{S})$	1	1	16	841	96721
$\mathcal{T}^L(\mathfrak{S}) \times \mathcal{T}^R(\mathfrak{S})$	1	1	9	256	15625
$\mathcal{T}_p(\mathfrak{S})$	1	1	2	14	141

Table 5.1: Number of vertices in the graphs $\mathcal{G}^L(\mathfrak{S})\times\mathcal{G}^R(\mathfrak{S})$, $\mathcal{T}^L(\mathfrak{S})\times\mathcal{T}^R(\mathfrak{S})$ and $\mathcal{T}_p(\mathfrak{S})$ for an alphabet of size n.

Due to the explosion in the number of morphisms, we focus on the case of a ternary alphabet which was first proved in [GLL22].

We show however in Subsection 5.3.2 that, if we restrict ourselves to the recurrent dendric languages having a unique right special word of each length, then we can easily build a graph giving an S-adic characterization on any given alphabet. This was first given as an example in [GL22].

Finally, we turn to languages of RIET in Subsection 5.3.3 and show how we can deduce an S-adic characterization for them from the techniques developed in this work.

5.3.1 Dendric languages on a given alphabet

Since any recurrent dendric language has an S-adic representation using only return morphisms, Theorem 5.26 implies the following S-adic characterization.

Proposition 5.32. *Let $\mathcal{A}$ be an alphabet. A language $\mathcal{L}$ over $\mathcal{A}$ is recurrent dendric if and only if it has a primitive $\mathfrak{S}$-adic representation labeling infinite paths in the graphs $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$ where $\mathfrak{S}$ is the set of all return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$.*

This characterization is however not practical as this set $\mathfrak{S}$ contains many non-dendric morphisms and each dendric language then has infinitely many $\mathfrak{S}$-adic representations. In this section, we therefore look for a smaller and more practical set of morphisms.

To have the existence of an $\mathfrak{S}$-adic representation, it suffices that for any dendric language $\mathcal{L}$ over $\mathcal{A}$, there exists a word $w \in \mathcal{L}$ and a morphism $\sigma \in \mathfrak{S}$ coding the return words for w in $\mathcal{L}$. In other words, we can build the set $\mathfrak{S}$ by choosing, for each dendric language $\mathcal{L}$, a word $w \in \mathcal{L}$ and add the corresponding return morphism to $\mathfrak{S}$. Since we can choose the word w, we will naturally pick the word giving us the simplest morphism in some sense.

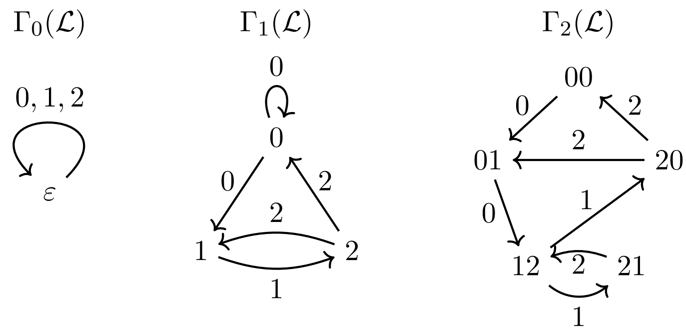


Figure 5.2: Rauzy graphs of order 0, 1 and 2 (from left to right) for the Chacon language $\mathcal{L}$.

This construction of the set $\mathfrak{S}$ relies heavily on the use of Rauzy graphs as they are strongly related to return words. We recall their definition below.

Definition 5.33. Let $\mathcal{L}$ be a language over $\mathcal{A}$. The *Rauzy graph of order* n, $n \geq 0$, of $\mathcal{L}$ is the graph $\Gamma_n(\mathcal{L})$ whose vertices are the elements of $\mathcal{L}_n$ and there is an edge from u to v labeled by $a \in \mathcal{A}$ if and only if there exists a letter b such that $ub = av \in \mathcal{L}_{n+1}$.

Example 5.34. Let $\mathcal{L}$ be the Chacon language, i.e., the language generated by the morphism σ such that $\sigma(0) = 0012$, $\sigma(1) = 12$ and $\sigma(2) = 012$. Using the first elements of $\mathcal{L}$ (Example 1.13), the Rauzy graphs of small orders of $\mathcal{L}$ are represented in Figure 5.2.

Since the Rauzy graph of order n is defined using $\mathcal{L}_n$ and $\mathcal{L}_{n+1}$, it is entirely determined by the extension graphs of the length-$(n-1)$ words. In particular $\Gamma_1(\mathcal{L})$ depends only on $\mathcal{E}_{\mathcal{L}}(\varepsilon)$, a fact that we will use later.

By definition of a Rauzy graph, if $uw \in \mathcal{L}$ with $|w| = n$, then u labels a path in $\Gamma_n(\mathcal{L})$ going from $(uw)_1 \cdots (uw)_n$ to w. The converse is not true however, if u labels a path in $\Gamma_n(\mathcal{L})$ ending in w, then uw is not necessarily in $\mathcal{L}$. This can easily be seen by looking at $\Gamma_0(\mathcal{L})$ in which all the words label a path.

In terms of return words, this means that the return words for w in $\mathcal{L}$ are among the labels of the paths of $\Gamma_{|w|}(\mathcal{L})$ starting in w, ending in w and for which w is not an intermediary vertex. We will call such a path a *return path to* w. This is why Rauzy graphs are useful when building the set $\mathfrak{S}$. Indeed, infinitely many dendric languages have the same graph $\Gamma_n(\mathcal{L})$, and this graph gives a regular language containing the return words for $w \in \mathcal{L}_n$.

Recall that given a dendric language, we can choose the non-empty word w of which we want to know the return words. The simplest Rauzy graph (other than $\Gamma_0(\mathcal{L})$) is $\Gamma_1(\mathcal{L})$ so we will look at return words for a letter. Moreover, we can choose which letter. As a rule of thumb, the fewer cycles not using this letter in $\Gamma_1(\mathcal{L})$, the simpler the return paths. Therefore, we will always choose a bispecial letter. This can be done because of the following result.

Proposition 5.35. *Let $\mathcal{L}$ be a language over an alphabet of size at least 2. If the order 1 Rauzy graph of $\mathcal{L}$ is strongly connected (meaning that, for any two vertices, there is a cycle passing through them) and if ε is connected in $\mathcal{L}$, then $\mathcal{L}$ has a bispecial letter.*

Proof. Let us first notice that, if a graph is strongly connected, then the only sets of vertices stable when taking successors (i.e., if (u, v) is an edge and u is in the set, so is v) are the empty set and the set of all vertices. In the case of the order 1 Rauzy graph, it means that any non-empty set of letters containing the right extensions of its elements is the whole alphabet.

Let us also observe that, as we are on an alphabet of size at least 2 and $\mathcal{E}_{\mathcal{L}}(\varepsilon)$ is connected, there is at least one right special letter and one left special letter. Moreover, if $(a, b) \in E_{\mathcal{L}}(\varepsilon)$ and a is not right special, then b has to be left special.

Assume now that there is no bispecial letter and let us consider the set of left special letters. Due to the observation above, any left special letter is not right special so its unique right extension is left special. This implies that the set of left special letters contains the right extensions of its elements. As it is not empty, it then corresponds to the whole alphabet, which contradicts the fact that there exists a right special letter. □

Observe that, if a language is recurrent, then its Rauzy graphs are strongly connected so in particular, we have the following consequence.

Corollary 5.36. *Any recurrent dendric language on an alphabet of size at least 2 has a bispecial letter.*

If $\mathfrak{S}$ then contains one return morphism per choice of $\#\mathcal{A}$ return paths for ℓ in $\Gamma_1(\mathcal{L})$ (i.e., per choice of $\#\mathcal{A}$ potential return words), then $\mathfrak{S}$ contains a morphism coding the return words for ℓ in $\mathcal{L}$. Doing this for any possible Rauzy graph of order 1 already gives a much smaller and manageable set $\mathfrak{S}$ of return morphisms.

However, this set is still infinite if $\#\mathcal{A} \geq 3$. In fact, it is the case of any set $\mathfrak{S}$ of return morphisms such that any dendric over $\mathcal{A}$ has an $\mathfrak{S}$-adic representation. Therefore, it will be more practical to group the morphisms of $\mathfrak{S}$ based on the edges they label in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$. Recall that, by Remark 5.19, given a return morphism σ, the edges it labels in $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$) are entirely determined by whether σ is dendric or not and by the corresponding set $\{\varphi_s^L : s \in \mathcal{A}^*, \#\varphi_s^L(\mathcal{A}) \geq 2\}$ (resp., $\{\phi_p^R : p \in \mathcal{A}^*, \#\phi_p^R(\mathcal{A}) \geq 2\}$).

Given a Rauzy graph and one of its vertices w, since we can describe using regular expressions the size $\#\mathcal{A}$ sets of return paths, the idea is to use these regular expressions to obtain a description of all size $\#\mathcal{A}$ sets of return paths whose corresponding morphisms have the same sets $\{\varphi_s^L : s \in \mathcal{A}^*, \#\varphi_s^L(\mathcal{A}) \geq 2\}$ and $\{\phi_p^R : p \in \mathcal{A}^*, \#\phi_p^R(\mathcal{A}) \geq 2\}$. Then, in each obtained family of morphisms, characterize the dendric return morphisms.

We use this technique to give the effective construction of a graph characterizing recurrent dendric languages on the ternary alphabet $\{0, 1, 2\}$. It is the simplest non trivial alphabet since the case of a binary alphabet corresponds to recurrent Sturmian languages and is well known (see Proposition 5.1).

We first need to find all possible Rauzy graphs of order 1. Since they are determined by the extension graphs of the empty word, we first list the bipartite trees whose left and right vertices are $\{0, 1, 2\}$. Moreover, since we restrict ourselves to recurrent languages, the corresponding Rauzy graphs must be strongly connected which can be seen as follows in $\mathcal{E}_{\mathcal{L}}(\varepsilon)$: for any non-empty set $\mathcal{B} \subsetneq \{0, 1, 2\}$, there exists an edge in $\mathcal{B} \times (\{0, 1, 2\} \setminus \mathcal{B})$ and an edge in $(\{0, 1, 2\} \setminus \mathcal{B}) \times \mathcal{B}$. We then obtain, up to permutation on $\{0, 1, 2\}$, the six possibilities described in Table 5.2.

For each of the obtained Rauzy graphs, we now choose a vertex with the simplest return paths. In the case of a ternary alphabet, any bispecial letter will do so, in what follows, we consider the return words (or paths) for 0 (which is bispecial in all 6 cases represented in Table 5.2). We now need to describe the return paths to 0. In cases (A), (B), (C) and (F), there are exactly three return paths to 0 meaning that the labels of these paths are exactly the return words. Let us consider the other two cases separately.

In the case (D) of Table 5.2, the labels of the return paths to 0 are 0 and the elements of 012^*2. Since every edge must be used by at least one return path, the possible sets of return words are of the form $\{0, 012^k2, 012^\ell 2\}$ for $0 \leq k < \ell$. The corresponding morphism is then given by

$$\sigma_{k,\ell} \colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 012^k2 \\ 2 \mapsto 012^\ell 2 \end{cases} .$$

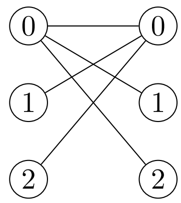
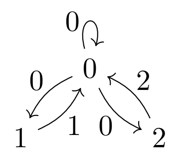
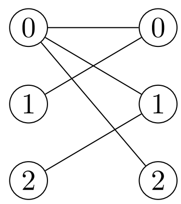
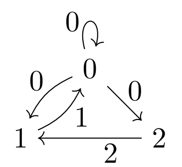
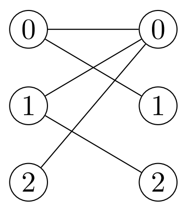
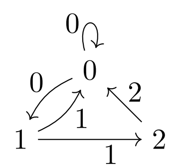
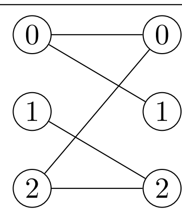
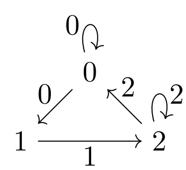
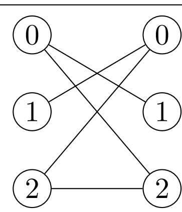
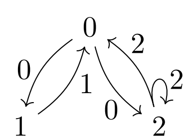
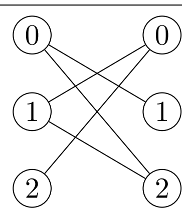
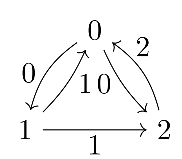
	$\mathcal{E}_{\mathcal{L}}(\varepsilon)$	$\Gamma_1(\mathcal{L})$
(A)		
(B)		
(C)		
(D)		
(E)		
(F)		

Table 5.2: Up to permutations, all the possible extension graphs of ε in a recurrent dendric language over $\{0,1,2\}$ and the corresponding Rauzy graphs of order 1.

Note that we can choose which letter is sent to which word since we only need one morphism per set of return words.

No matter the parameters $k < \ell$, we have

$$\{\varphi^L_{\sigma_{k,\ell},s} : s \in \mathcal{A}^*, \#\varphi^L_{\sigma_{k,\ell},s}(\{0,1,2\}) \geq 2\} = \left\{ \begin{cases} 0 \mapsto 0 \\ 1,2 \mapsto 2 \end{cases}, \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases} \right\}$$

and

$$\{\phi^R_{\sigma_{k,\ell},p} : p \in \mathcal{A}^*, \#\phi^R_{\sigma_{k,\ell},p}(\{0,1,2\}) \geq 2\} = \left\{ \begin{cases} 0 \mapsto 0 \\ 1,2 \mapsto 1 \end{cases}, \begin{cases} 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases} \right\}.$$

Therefore, the dendric morphisms of the form $\sigma_{k,\ell}$ all label the same edges in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$. Let us now characterize the parameters k and ℓ for which $\sigma_{k,\ell}$ is dendric.

Since 1 is neither left nor right special, the only bispecial initial factors are ε and 2^i for $1 \leq i \leq \ell$. Since $\ell \geq 1$, we easily check that ε is dendric. If $\ell > k+1$, then the extensions of 2^ℓ are $(1,2)$ and $(2,0)$ so 2^ℓ is not connected. Therefore, if $\sigma_{k,\ell}$ is dendric, then $\ell = k+1$. In that case, we then see that the extensions of 2^ℓ are $(1,0)$, $(1,2)$ and $(2,0)$, and the extensions of 2^i are $(1,2)$,$(2,2)$ and $(2,0)$ for all $1 \leq i < \ell$. This shows that $\sigma_{k,k+1}$ is indeed dendric for all $k \geq 0$. We will then denote $\sigma_{k,k+1} = \delta_k$ since it corresponds to case (D).

Similarly, for case (E), we can show that the dendric return morphisms form an infinite family depending on a unique parameter and they all label the same edges in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$.

The complete description of the morphisms for all cases and the corresponding maps φ^L_s and ϕ^R_p are given in Table 5.3.

Let us denote $\mathfrak{S}_3 = \{\alpha, \beta, \gamma, \eta\} \cup \{\delta_k : k \geq 0\} \cup \{\zeta_k : k \geq 0\}$. Let also Σ_3 be the set of permutations on $\{0,1,2\}$. By construction, any recurrent dendric language over $\{0,1,2\}$ has an $\Sigma_3\mathfrak{S}_3$-adic representation so in particular, it has an $\Sigma_3\mathfrak{S}_3\Sigma_3$-adic representation. We can then use the symmetries to replace $\mathcal{T}^L(\Sigma_3\mathfrak{S}_3\Sigma_3)$ and $\mathcal{T}^R(\Sigma_3\mathfrak{S}_3\Sigma_3)$ by a unique graph with two vertices as in Proposition 5.31. Indeed, the only tree on three vertices is the line tree so, when we look at pairs of trees, there are only two cases up to permutation: either the middle vertices are equal or they are different. To obtain an even simpler characterization, we will allow some of the morphisms in the S-adic characterization to be simple permutations instead of having to duplicate many edges. We finally obtain the following characterization.

	σ	$\{\varphi^L_{\sigma,s} : \#\varphi^L_{\sigma,s}(\mathcal{A}) \geq 2\}$	$\{\phi^R_{\sigma,p} : \#\phi^R_{\sigma,p}(\mathcal{A}) \geq 2\}$
(A)	$\alpha\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 02 \end{cases}$	id	id
(B)	$\beta\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 021 \end{cases}$	$\begin{cases} 0 \mapsto 0 \\ 1,2 \mapsto 1 \end{cases}$ $\begin{cases} 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$	id
(C)	$\gamma\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 012 \end{cases}$	id	$\begin{cases} 0 \mapsto 0 \\ 1,2 \mapsto 1 \end{cases}$ $\begin{cases} 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$
(D)	$\delta_k\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 012^k2 \\ 2 \mapsto 012^{k+1}2 \end{cases}$	$\begin{cases} 0 \mapsto 0 \\ 1,2 \mapsto 2 \end{cases}$ $\begin{cases} 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases}$	$\begin{cases} 0 \mapsto 0 \\ 1,2 \mapsto 1 \end{cases}$ $\begin{cases} 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$
(E)	$\zeta_k\colon \begin{cases} 0 \mapsto 02^k2 \\ 1 \mapsto 01 \\ 2 \mapsto 02^{k+1}2 \end{cases}$	$\begin{cases} 0,2 \mapsto 2 \\ 1 \mapsto 1 \end{cases}$ $\begin{cases} 0 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$	$\begin{cases} 0,2 \mapsto 2 \\ 1 \mapsto 1 \end{cases}$ $\begin{cases} 0 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$
(F)	$\eta\colon \begin{cases} 0 \mapsto 02 \\ 1 \mapsto 01 \\ 2 \mapsto 012 \end{cases}$	$\begin{cases} 0,2 \mapsto 2 \\ 1 \mapsto 1 \end{cases}$ $\begin{cases} 0 \mapsto 0 \\ 2 \mapsto 1 \end{cases}$	$\begin{cases} 0 \mapsto 2 \\ 1,2 \mapsto 1 \end{cases}$ $\begin{cases} 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$

Table 5.3: For each case of Table 5.2, the corresponding dendric return morphism(s) and the maps determining the edges they label in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$.

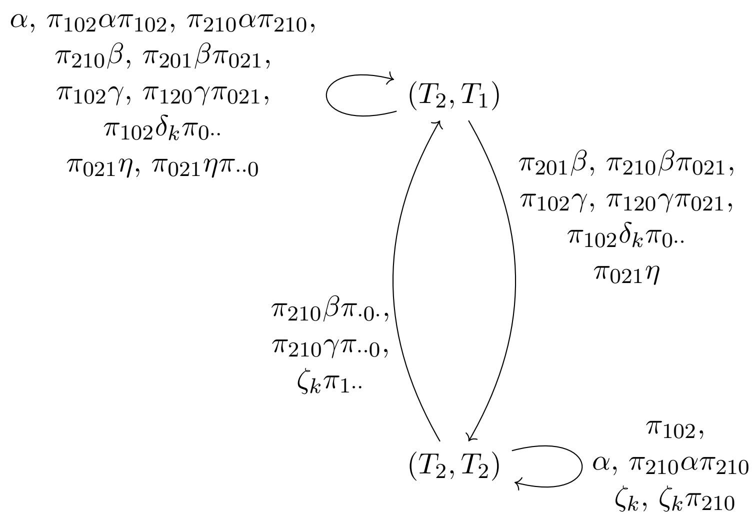


Figure 5.3: A language over $\{0, 1, 2\}$ is dendric recurrent if and only if it has a primitive S-adic representation labeling an infinite path in this graph where T_i is the tree with vertices 0,1,2 where i is the degree-2 vertex, $\pi_{i_0 i_1 i_2}$ is the morphism such that $\pi_{i_0 i_1 i_2}(j) = i_j$ and the dots are place-holders that can represent any value in $\{0, 1, 2\}$ giving a permutation.

Theorem 5.37. *A language $\mathcal{L}$ over $\{0, 1, 2\}$ is dendric recurrent if and only if, up to permutation, it has a primitive S-adic representation labeling an infinite path in the graph represented in Figure 5.3.*

The set $\Sigma_3\mathfrak{S}_3\Sigma_3$ is infinite due to the morphisms δ_k and ζ_k. It is however possible to obtain a similar characterization using only finitely many morphisms. Indeed, the elements of $\Sigma_3\mathfrak{S}_3\Sigma_3$ are dendric return morphisms, therefore, by Proposition 4.44, they are tame. This means that we can replace each edge of $\mathcal{T}_p(\Sigma_3\mathfrak{S}_3\Sigma_3)$ by a path corresponding to a factorization of the morphisms into elementary morphisms. More precisely, as a consequence of Proposition 4.45, we have the following result.

Corollary 5.38. *Let σ be a dendric return morphism for a letter and let $\sigma = \tau_1 \circ \cdots \circ \tau_n$ be an elementary decomposition of σ. Then, for all $i \leq n$, $\tau_i \circ \cdots \circ \tau_n$ is a return morphism for a set of letters.*

Since moreover, $\tau^{(i)} := \tau_i \circ \cdots \circ \tau_n$ is tame, the initial alphabet and the image alphabet are equal so, for any dendric language $\mathcal{L}$, the fact that $\tau^{(i)}(\mathcal{L})$ is dendric only depends on an easily checked property of $\tau^{(i)}$ and on the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ by Corollary 4.91 that we recall here.

Corollary 4.91. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ (assuming that $\mathcal{A}$ is the smallest image alphabet) be a return morphism for $S \subseteq \mathcal{A}$ and $\mathcal{L}$ be a dendric language over $\mathcal{A}$. The image $\sigma(\mathcal{L})$ is dendric if and only if the initial factors are connected in $\sigma(\mathcal{A})c$ for $c \notin \mathcal{A} \setminus S$ and the following conditions are satisfied:*

- *for all $s \in \mathrm{Suff}^*(\sigma(\mathcal{A}))$, the subgraph of $G^L(\mathcal{L})$ generated by the vertices in $\mathrm{dom}(\varphi_s^L)$ is connected;*
- *for all $p \in \mathcal{A}^+ \cap \mathrm{Pref}(\sigma(\mathcal{A}))$, the subgraph of $G^R(\mathcal{L})$ generated by the vertices in d_p is connected.*

The idea is then to replace an edge labeled by σ and ending in (T^L, T^R) in the graph of Figure 5.3 by a path labeled by $\tau_1, \ldots, \tau_n$ such that $\tau_1 \circ \cdots \circ \tau_n$ is an elementary decomposition of σ and, for all $i \leq n$, the initial factors (for $\tau^{(i)}$) are connected in $\tau^{(i)}(\{0,1,2\})3$ (recall that we ignore isolated vertices here) and the connectedness conditions of Corollary 4.91 are satisfied for T^L and T^R. The choice of such a path for each edge is described in Tables 5.4, 5.5, 5.6 and 5.7. Note that we wrote L_{ab} instead of $L_{a,b}$ and R_{ab} instead of $R_{a,b}$ to simplify the notations. We let the reader convince themselves that they indeed satisfy the conditions.

We make some remarks on these decompositions. First, to our knowledge, there is no prior guarantee that, for each edge, we can find an elementary decomposition satisfying the conditions for each intermediary morphism. This has simply been checked by hand. Second, even though we only give one decomposition per edge, there were many other possibilities, especially due to permutations. The choices made here are sometimes arbitrary. Third, observe that, for some morphisms labeling multiple edges, we used different decompositions. It is for example the case of the morphism α which labels a loop on (T_2, T_1) and on (T_2, T_2). In fact, one can easily check that it is impossible to find a unique decomposition of α that will satisfy the conditions for both loops.

We then have the following S-adic characterization.

Theorem 5.39. *Let $\mathfrak{S}_e$ denote the set of elementary morphisms over $\{0,1,2\}$. A language $\mathcal{L}$ over $\{0,1,2\}$ is dendric recurrent if and only if, up to permutation, it has a primitive $\mathfrak{S}_e$-adic representation labeling an infinite path in the graph represented in Figure 5.4.*

Proof. By construction and by Theorem 5.37, a language $\mathcal{L}$ over $\{0,1,2\}$ is dendric recurrent if and only if, up to permutation, it has a primitive $\mathfrak{S}_e$-adic representation labeling an infinite path starting in (T_2, T_1) or (T_2, T_2) in the graph of Figure 5.4.

morphism	decomposition	intermediary morphisms
α	$L_{20}L_{10}$	$\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 2\end{cases}$
$\pi_{102}\alpha\pi_{102}$	$L_{21}L_{01}$	$\begin{cases}0\mapsto 10\\1\mapsto 1\\2\mapsto 2\end{cases}$
$\pi_{210}\alpha\pi_{210}$	$L_{02}L_{12}$	$\begin{cases}0\mapsto 0\\1\mapsto 21\\2\mapsto 2\end{cases}$
$\pi_{210}\beta$	$\pi_{210}R_{21}L_{20}L_{10}$	$\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 2\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 02\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 021\end{cases}$
$\pi_{201}\beta\pi_{021}$	$\pi_{210}R_{12}L_{20}L_{10}$	$\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 2\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 02\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 012\\2\mapsto 02\end{cases}$
$\pi_{102}\gamma$	$\pi_{102}L_{10}L_{21}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 012\end{cases}$
$\pi_{120}\gamma\pi_{021}$	$\pi_{102}L_{20}L_{12}$	$\begin{cases}0\mapsto 0\\1\mapsto 21\\2\mapsto 2\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 021\\2\mapsto 02\end{cases}$
$\pi_{102}\delta_k$	$\pi_{102}L_{10}R_{12}^kR_{12}L_{21}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 12^i\\2\mapsto 12^{i+1}\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 012^{k+1}\\2\mapsto 012^{k+2}\end{cases}$
$\pi_{102}\delta_k\pi_{021}$	$\pi_{120}L_{20}R_{21}^kR_{21}L_{12}$	$\begin{cases}0\mapsto 0\\1\mapsto 21\\2\mapsto 2\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 21^{i+1}\\2\mapsto 21^i\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 021^{k+2}\\2\mapsto 021^{k+1}\end{cases}$
$\pi_{021}\eta$	$\pi_{021}L_{10}R_{02}L_{21}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 02\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 02\\1\mapsto 01\\2\mapsto 012\end{cases}$
$\pi_{021}\eta\pi_{120}$	$L_{20}R_{01}L_{12}\pi_{210}$	$\begin{cases}0\mapsto 2\\1\mapsto 21\\2\mapsto 0\end{cases}$ $\begin{cases}0\mapsto 2\\1\mapsto 21\\2\mapsto 01\end{cases}$
$\pi_{021}\eta\pi_{210}$	$L_{20}R_{01}L_{12}\pi_{120}$	$\begin{cases}0\mapsto 21\\1\mapsto 2\\2\mapsto 0\end{cases}$ $\begin{cases}0\mapsto 21\\1\mapsto 2\\2\mapsto 01\end{cases}$

Table 5.4: For each edge from (T_2, T_1) to (T_2, T_1) in the graph of Figure 5.3, we give a decomposition into elementary morphisms and the corresponding intermediary morphisms.

morphism	decomposition	intermediary morphisms
$\pi_{201}\beta$	$\pi_{201}R_{21}L_{10}L_{20}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 02\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 02\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 021\end{cases}$
$\pi_{210}\beta\pi_{021}$	$\pi_{201}R_{12}L_{10}L_{20}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 02\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 02\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 012\\2\mapsto 02\end{cases}$
$\pi_{102}\gamma$	$\pi_{102}L_{10}L_{21}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 01\\2\mapsto 012\end{cases}$
$\pi_{120}\gamma\pi_{021}$	$\pi_{102}L_{20}L_{12}$	$\begin{cases}0\mapsto 0\\1\mapsto 21\\2\mapsto 2\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 021\\2\mapsto 02\end{cases}$
$\pi_{102}\delta_k$	$\pi_{102}L_{10}R_{12}^kR_{12}L_{21}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 12^i\\2\mapsto 12^{i+1}\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 012^{k+1}\\2\mapsto 012^{k+2}\end{cases}$
$\pi_{102}\delta_k\pi_{021}$	$\pi_{120}L_{20}R_{21}^kR_{21}L_{12}$	$\begin{cases}0\mapsto 0\\1\mapsto 21\\2\mapsto 2\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 21^{i+1}\\2\mapsto 21^i\end{cases}$ $\begin{cases}0\mapsto 0\\1\mapsto 021^{k+2}\\2\mapsto 021^{k+1}\end{cases}$
$\pi_{021}\eta$	$\pi_{021}L_{10}R_{02}L_{21}$	$\begin{cases}0\mapsto 0\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 02\\1\mapsto 1\\2\mapsto 12\end{cases}$ $\begin{cases}0\mapsto 02\\1\mapsto 01\\2\mapsto 012\end{cases}$

Table 5.5: For each edge from (T_2,T_1) to (T_2,T_2) in the graph of Figure 5.3, we give a decomposition into elementary morphisms and the corresponding intermediary morphisms.

morphism	decomposition	intermediary morphisms
$\pi_{210}\beta\pi_{102}$	$\pi_{210}R_{21}L_{20}L_{10}\pi_{102}$	$\begin{cases}0 \mapsto 01\\1 \mapsto 0\\2 \mapsto 2\end{cases}$ $\begin{cases}0 \mapsto 01\\1 \mapsto 0\\2 \mapsto 02\end{cases}$ $\begin{cases}0 \mapsto 01\\1 \mapsto 0\\2 \mapsto 021\end{cases}$
$\pi_{210}\beta\pi_{201}$	$\pi_{210}R_{21}L_{20}L_{10}\pi_{201}$	$\begin{cases}0 \mapsto 2\\1 \mapsto 0\\2 \mapsto 01\end{cases}$ $\begin{cases}0 \mapsto 02\\1 \mapsto 0\\2 \mapsto 01\end{cases}$ $\begin{cases}0 \mapsto 021\\1 \mapsto 0\\2 \mapsto 01\end{cases}$
$\pi_{210}\gamma\pi_{120}$	$\pi_{210}L_{10}L_{21}\pi_{120}$	$\begin{cases}0 \mapsto 1\\1 \mapsto 12\\2 \mapsto 0\end{cases}$ $\begin{cases}0 \mapsto 01\\1 \mapsto 012\\2 \mapsto 0\end{cases}$
$\pi_{210}\gamma\pi_{210}$	$\pi_{210}L_{10}L_{21}\pi_{210}$	$\begin{cases}0 \mapsto 12\\1 \mapsto 1\\2 \mapsto 0\end{cases}$ $\begin{cases}0 \mapsto 012\\1 \mapsto 01\\2 \mapsto 0\end{cases}$
$\zeta_k\pi_{102}$	$L_{10}R_{02}^kR_{02}L_{20}\pi_{102}$	$\begin{cases}0 \mapsto 1\\1 \mapsto 0\\2 \mapsto 02\end{cases}$ $\begin{cases}0 \mapsto 1\\1 \mapsto 02^i\\2 \mapsto 02^{i+1}\end{cases}$
$\zeta_k\pi_{120}$	$L_{10}R_{02}^kR_{02}L_{20}\pi_{120}$	$\begin{cases}0 \mapsto 1\\1 \mapsto 02\\2 \mapsto 0\end{cases}$ $\begin{cases}0 \mapsto 1\\1 \mapsto 02^{i+1}\\2 \mapsto 02^i\end{cases}$

Table 5.6: For each edge from (T_2, T_2) to (T_2, T_1) in the graph of Figure 5.3, we give a decomposition into elementary morphisms and the corresponding intermediary morphisms.

morphism	decomposition	intermediary morphisms		
π_{102}	π_{102}	none		
α	$L_{10}L_{20}$	$\begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 02 \end{cases}$		
$\pi_{210}\alpha\pi_{210}$	$L_{12}L_{02}$	$\begin{cases} 0 \mapsto 20 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases}$		
ζ_k	$L_{10}R_{02}^{k}R_{02}L_{20}$	$\begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 02 \end{cases}$	$\begin{cases} 0 \mapsto 02^{i} \\ 1 \mapsto 1 \\ 2 \mapsto 02^{i+1} \end{cases}$	$\begin{cases} 0 \mapsto 012^{k+1} \\ 1 \mapsto 01 \\ 2 \mapsto 012^{k+2} \end{cases}$
$\zeta_k\pi_{210}$	$\pi_{210}L_{12}R_{20}^{k}R_{20}L_{02}$	$\begin{cases} 0 \mapsto 20 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases}$	$\begin{cases} 0 \mapsto 20^{i+1} \\ 1 \mapsto 1 \\ 2 \mapsto 20^{i} \end{cases}$	$\begin{cases} 0 \mapsto 20^{k+2} \\ 1 \mapsto 21 \\ 2 \mapsto 20^{k+1} \end{cases}$

Table 5.7: For each edge from (T_2, T_2) to (T_2, T_2) in the graph of Figure 5.3, we give a decomposition into elementary morphisms and the corresponding intermediary morphisms.

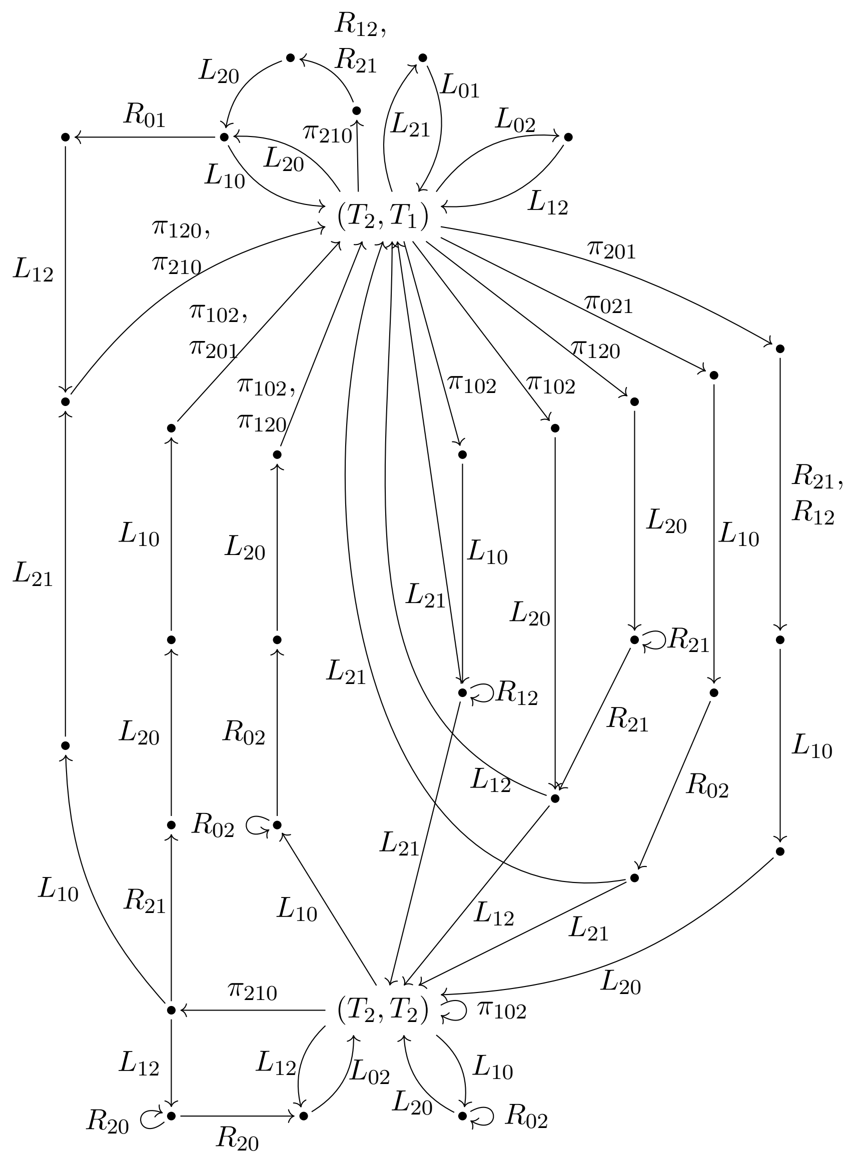


Figure 5.4: A language over $\{0, 1, 2\}$ is dendric recurrent if and only if it has a primitive $\mathfrak{S}_e$-adic representation labeling an infinite path in this graph, where $\mathfrak{S}_e$ is the set of elementary morphisms over $\{0, 1, 2\}$ and $\pi_{i_0 i_1 i_2}$ is the morphism such that $\pi_{i_0 i_1 i_2}(j) = i_j$.

Let us show that if a primitive sequence $\boldsymbol{\sigma} = (\sigma_n)_{n\geq 0}$ labels a path starting in one of the other vertices, then it also is an $\mathfrak{S}_e$-adic representation of a recurrent dendric language. By construction, there exists $N \geq 0$ such that $\boldsymbol{\sigma}' = (\sigma_n)_{n\geq N}$ labels a path starting in (T_2, T_1) (resp., (T_2, T_2)), and $\sigma_0 \cdots \sigma_{N-1}$ is a suffix of a well-chosen elementary decomposition of a morphism σ labeling an edge ending in (T_2, T_1) (resp., (T_2, T_2)) in the graph of Figure 5.3.

By Theorem 5.37, $\boldsymbol{\sigma}'$ is an S-adic representation of a recurrent dendric language $\mathcal{L}$ and, by Proposition 5.23, T_2 is a subgraph of $G^L(\mathcal{L})$ and T_1 (resp., T_2) is a subgraph of $G^R(\mathcal{L})$. By construction of the graph of Figure 5.4 and by Corollary 4.91, this implies that $\sigma_0 \cdots \sigma_{N-1}(\mathcal{L})$ is recurrent dendric. □

5.3.2 One sided Arnoux-Rauzy languages

The construction of the previous subsection can also be used to obtain a constructive S-adic characterization of other families of dendric languages. Indeed, we will now characterize the dendric languages having exactly one right special word of each length. Of course, we can similarly study dendric languages having exactly one left special word of each length. Let us first observe that this family is stable under derivation.

Proposition 5.40. *Let $\mathcal{L}$ be a recurrent dendric language over $\mathcal{A}$ having exactly one right special word of each length and let $w \in \mathcal{L} \setminus \{\varepsilon\}$. Then $D_w(\mathcal{L})$ is a recurrent dendric language having exactly one right special word of each length.*

Proof. Note that a recurrent dendric language over $\mathcal{A}$ has exactly one right special word of each length if and only if $G^R(\mathcal{L})$ is the complete unicolor graph over $\mathcal{A}$. Let us denote $\mathcal{L}' = D_w(\mathcal{L})$. Since the family of recurrent dendric languages is stable under derivation (Theorem 3.42), $\mathcal{L}'$ is recurrent dendric. Moreover, it is on an alphabet $\mathcal{B}$ such that $\#\mathcal{B} = \#\mathcal{A}$ by Corollary 3.32 on the number of return words. Let σ denote the return morphism for w such that $\sigma(\mathcal{L}') = \mathcal{L}$. By Proposition 4.93, $\sigma^R(G^R(\mathcal{L}')) = G^R(\mathcal{L})$ and it is the complete unicolor graph on $\mathcal{A}$ by hypothesis on $\mathcal{L}$. By definition of σ^R of a graph, this implies that $G^R(\mathcal{L}')$ contains a clique of size at least $\#\mathcal{A}$. Since $\#\mathcal{B} = \#\mathcal{A}$ and $G^R(\mathcal{L}')$ is acyclic for the coloring, we conclude that $G^R(\mathcal{L}')$ is the complete unicolor graph on $\mathcal{B}$, showing that $\mathcal{L}'$ has exactly one right special word of each length. □

Therefore, to build a sufficient set $\mathfrak{S}$ of return morphism for the $\mathfrak{S}$-adic characterization, we only need to consider the Rauzy graphs corresponding to the recurrent dendric languages having exactly one right special word of each length.

However, if a recurrent dendric language $\mathcal{L}$ over $\mathcal{A}$ has a unique right special letter ℓ, then there are exactly $\#\mathcal{A}$ return paths to ℓ. Indeed, ℓ has $\#\mathcal{A}$ right extensions, meaning that there are $\#\mathcal{A}$ ways of starting a return path, and since there are no other right special letters and $\Gamma_1(\mathcal{L})$ is strongly connected, each start gives exactly one return path.

This leads to the following S-adic characterization.

Proposition 5.41. *Let $\mathcal{A}$ be an alphabet. There exists a finite computable set $\mathfrak{S}$ of morphisms such that a language $\mathcal{L}$ is recurrent dendric and has exactly one right special word of each length if and only if it has a primitive $\mathfrak{S}$-adic representation labeling an infinite path in $\mathcal{T}^L(\mathfrak{S})$.*

Proof. As explained in Subsection 5.3.1, we can list all the order 1 Rauzy graphs of recurrent dendric languages having exactly one right special word of each length by first listing the corresponding extension graphs of ε. For each obtained Rauzy graph, the observation above implies that there is a unique possible set of return words for the bispecial letter. Moreover, any morphism corresponding to these return words is dendric since the only bispecial initial factor is ε. Thus, we can add these morphisms to $\mathfrak{S}$.

By construction of $\mathfrak{S}$ and by Proposition 5.40, any recurrent dendric language over $\mathcal{A}$ having exactly one right special word of each length has an $\mathfrak{S}$-adic representation. This representation is primitive and labels an infinite path in $\mathcal{T}^L(\mathfrak{S})$ (and $\mathcal{T}^R(\mathfrak{S})$) by Theorem 5.26.

Let us prove the converse now. Observe that, by construction, for any morphism $\sigma \in \mathfrak{S}$, we have $\phi_\ell^R(\mathcal{A}) = \mathcal{A}$ if σ is a return morphism for ℓ. This implies that $(K_\mathcal{A}, \sigma, K_\mathcal{A})$ is a right valid triplet, where $K_\mathcal{A}$ denotes the complete unicolor graph on $\mathcal{A}$. In other words, any sequence of morphisms in $\mathfrak{S}$ labels an infinite path in $\mathcal{G}^R(\mathfrak{S})$ which stays in $K_\mathcal{A}$. In particular, if $\mathcal{L}$ has a primitive $\mathfrak{S}$-adic representation labeling an infinite path in $\mathcal{T}^L(\mathfrak{S})$, then $\mathcal{L}$ is recurrent dendric by Theorem 5.20. Moreover, by Proposition 5.23, we have $G^R(\mathcal{L}) = K_\mathcal{A}$, or in other words, $\mathcal{L}$ has exactly one right special word of each length. □

We illustrate this result by building the set $\mathfrak{S}$ and the corresponding $\mathfrak{S}$-adic characterization for $\mathcal{A} = \{0, 1, 2, 3\}$.

Example 5.42. We first list the possible extension graphs of ε in a recurrent dendric language over $\{0, 1, 2, 3\}$ having exactly one right special letter. We will work up to permutation and assume that the right special letter is

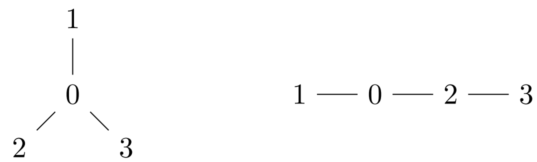


Figure 5.5: Graphs T_1 on the left and T_2 on the right.

0. These graphs $\mathcal{E}_{\mathcal{L}}(\varepsilon)$, the corresponding Rauzy graphs $\Gamma_1(\mathcal{L})$ and the associated return morphisms for 0 are given in Table 5.8.

Let us denote $\mathfrak{S}_4 = \{\alpha, \beta, \gamma, \delta\}$ and Σ_4 the set of permutations on $\{0, 1, 2, 3\}$. Since we are only interested in the paths in the graph $\mathcal{T}^L(\Sigma_4\mathfrak{S}_4)$, we can use symmetries and only consider one labeled tree per shape, i.e., the two vertices of our graph will be given by the graphs T_1 and T_2 of Figure 5.5.

To simplify the graph, we also allow the S-adic representation to contain permutations. This gives us the final graph represented in Figure 5.6.

5.3.3 Languages of RIET

The last example that we consider in this work is the family of languages of regular interval exchange transformations on a given alphabet $\mathcal{A}$. Indeed, these are particular recurrent dendric languages by Proposition 2.8. As in the dendric case, we will present a theoretical S-adic characterization from [GL22] then obtain the explicit graph in the case of the ternary alphabet $\{0, 1, 2\}$ which was first given in [GLL22].

The family of languages of RIET is stable under derivation. Therefore, by Proposition 5.7, any languages of an RIET on $\mathcal{A}$ admits an S-adic representation $(\sigma_n)_{n\geq 0}$ where the intermediary languages (generated by $(\sigma_n)_{n\geq N}$) are also languages of RIETs on $\mathcal{A}$.

Not every acyclic for the coloring and connected multi-clique can be the graph $G^L(\mathcal{L})$ or $G^R(\mathcal{L})$ of a language $\mathcal{L}$ of an RIET. This follows from the study of extensions of long enough words done in Proposition 1.39. We therefore give an alternative statement of this result below.

Definition 5.43. A *line graph* on $\mathcal{A}$ is a graph G such that, if $\mathcal{A} = \{a_1, \dots, a_n\}$, the edges are exactly the pairs (a_i, a_{i+1}), $i < n$. This graph is associated with the orders

$$a_1 < a_2 < \cdots < a_n \quad \text{and} \quad a_n <^* a_{n-1} <^* \cdots <^* a_1.$$

We then define $G(\leq)$ and $G(\leq^*)$ as the graph G.

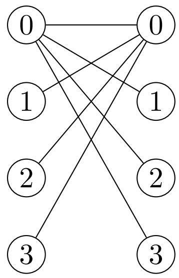
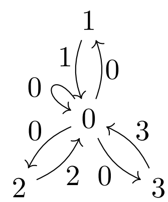
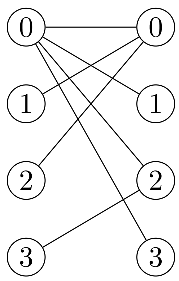
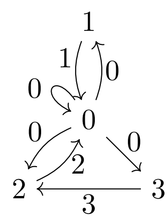
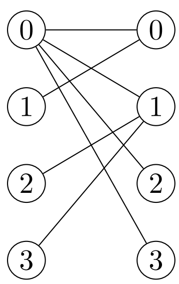
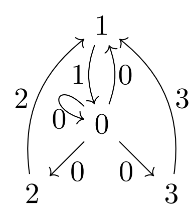
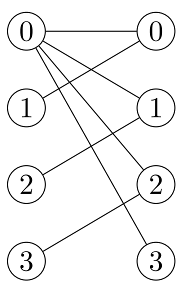
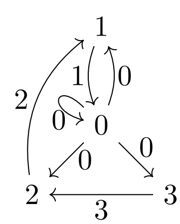
$\mathcal{E}_{\mathcal{L}}(\varepsilon)$	$\Gamma_1(\mathcal{L})$	σ
		$\alpha\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 02 \\ 3 \mapsto 03 \end{cases}$
		$\beta\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 02 \\ 3 \mapsto 032 \end{cases}$
		$\gamma\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 021 \\ 3 \mapsto 031 \end{cases}$
		$\delta\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \\ 2 \mapsto 021 \\ 3 \mapsto 0321 \end{cases}$

Table 5.8: Up to permutations, all the possible extension graphs of ε in a recurrent dendric language over $\{0,1,2,3\}$ with a unique right special letter, the corresponding Rauzy graphs of order 1 and the associated return morphism.

$$\begin{tikzcd}[row sep=huge]
T_1 \arrow[loop above, "{\begin{array}{c}\pi_{0\cdot\cdot\cdot}, \\ \alpha,\ \pi_{1023}\alpha\pi_{1023},\ \pi_{2103}\alpha\pi_{2103},\ \pi_{3120}\alpha\pi_{3120}, \\ \gamma\pi_{1023}, \\ \delta\pi_{\cdot 10\cdot}\end{array}}"] \arrow[d, bend left, "{\begin{array}{c}\beta\pi_{01\cdot\cdot}, \\ \gamma\pi_{\cdot 01\cdot}, \\ \delta\pi_{10\cdot\cdot},\ \delta\pi_{\cdot 1\cdot 0}\end{array}}"] \\
T_2 \arrow[u, bend left, "{\begin{array}{c}\pi_{2103}\beta\pi_{\cdot 10\cdot}, \\ \gamma\pi_{2103},\ \pi_{0132}\gamma\pi_{3120}\end{array}}"] \arrow[loop below, "{\begin{array}{c}\pi_{2301}, \\ \alpha,\ \pi_{1023}\alpha\pi_{1023},\ \pi_{2103}\alpha\pi_{2103},\ \pi_{3120}\alpha\pi_{3120}, \\ \pi_{2013}\beta\pi_{10\cdot\cdot},\ \pi_{2103}\beta\pi_{\cdot 1\cdot 0}, \\ \gamma\pi_{\cdot 02\cdot},\ \pi_{0132}\gamma\pi_{\cdot 03\cdot}\end{array}}"]
\end{tikzcd}$$

Figure 5.6: A language $\mathcal{L}$ over $\{0, 1, 2, 3\}$ is dendric recurrent and has exactly one right special word of each length if and only if it there exists a permutation π such that $\pi(\mathcal{L})$ has a primitive S-adic representation labeling an infinite path in this graph where the morphisms α, β, γ, δ are defined in Table 5.8, $\pi_{i_0 i_1 i_2 i_3}$ is the morphism such that $\pi_{i_0 i_1 i_2 i_3}(j) = i_j$ and the dots are place-holders that can represent any value in $\{0, 1, 2, 3\}$ giving a permutation.

Corollary 5.44. *If $\mathcal{L}$ is the language of an RIET for the orders $\binom{\leq}{\preceq}$, then $G^L(\mathcal{L}) = G(\preceq)$ and $G^R(\mathcal{L}) = G(\leq)$.*

However, the converse is false. The fact that $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are line graphs does not even imply that $\mathcal{L}$ is dendric, as seen in Example 2.59.

All of this leads to the following observation.

Remark 5.45. Let $\mathfrak{S}$ be a set of return morphisms for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$. If $\boldsymbol{\sigma}$ is an $\mathfrak{S}$-adic representation of the language of an RIET on $\mathcal{A}$, then $\boldsymbol{\sigma}$ labels an infinite path in the subgraph of $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$) generated by the vertices that are line graphs.

The converse is not necessarily true however, if $\boldsymbol{\sigma}$ is primitive and labels infinite paths in these subgraphs, then it does not always generate the language of an RIET. Indeed, if $\mathfrak{S}$ contains the Arnoux-Rauzy morphisms, then each of these morphisms labels a loop on every vertex of $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$. However, the languages they generate are the Arnoux-Rauzy languages by Proposition 5.2, which are not languages of RIET if the alphabet size is at least 3.

The goal of this section is then to characterize the paths corresponding to languages of RIET in the subgraph of $\mathcal{T}^L(\mathfrak{S})$ (resp., $\mathcal{T}^R(\mathfrak{S})$) generated by the line graph vertices.

To do so, we will heavily use the characterization of the languages of RIET as the recurrent planar dendric languages (Proposition 2.8). Recall that a word $w \in \mathcal{L}$ is planar for the orders $\preceq$ and $\leq$ if for all $(a_1, b_1), (a_2, b_2) \in E_{\mathcal{L}}(w)$ such that $a_1 \prec a_2$, we have $b_1 \leq b_2$. A language is planar for the orders $\preceq$ and $\leq$ if all of its elements are.

Therefore, the morphisms σ_n of the S-adic representations have to preserve dendricity and planarity for some orders. This is the object of the following definitions. Recall that if $\leq$ is an order, then $\leq^*$ denotes the inverse order, i.e., $a < b$ if and only if $b <^* a$.

Definition 5.46. Let $\preceq$ and $\leq$ be two total orders on $\mathcal{A}$.

- A partial map $\varphi\colon \mathcal{A} \to \mathcal{A}$ is *order preserving from $\preceq$ to $\leq$* if, for all $x, y \in \operatorname{dom}(\varphi)$, we have
$$x \prec y \Rightarrow \varphi(x) \leq \varphi(y).$$
- A return morphism $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ for w is *left order preserving from $\preceq$ to $\leq$* if, for all $s \in \operatorname{Suff}^*(\sigma(\mathcal{A}))$, $\varphi^L_{\sigma,s}$ is order preserving from $\preceq$ to $\leq$. Similarly, σ is *right order preserving from $\preceq$ to $\leq$* if, for all $p \in \operatorname{Pref}^*(\sigma(\mathcal{A})w)$, $\phi^R_{\sigma,p}$ is order preserving from $\preceq$ to $\leq$.

- A return morphism $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ is *planar preserving* from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$ if it is left order preserving from $\preceq^L$ to $\leq^L$ and right order preserving from $\preceq^R$ to $\leq^R$, or if it is left order preserving from $\preceq^L$ to $(\leq^L)^*$ and right order preserving from $\preceq^R$ to $(\leq^R)^*$.

Observe that, given a return morphism σ and some pairs of orders, we can easily determine if σ is planar preserving for these pairs of orders.

Example 5.47. Let us consider the morphism η of Subsection 5.3.1 given by $\eta(0) = 02$, $\eta(1) = 01$ and $\eta(2) = 012$. The corresponding left maps are

$$\varphi_\varepsilon^L\colon \begin{cases} 0, 2 \mapsto 2 \\ \quad 1 \mapsto 1 \end{cases} \quad \text{and} \quad \varphi_2^L\colon \begin{cases} 0 \mapsto 0 \\ 2 \mapsto 1 \end{cases}$$

(see Table 5.3). Let us consider the order $0 \prec 2 \prec 1$ and describe the orders $\leq$ such that η is left order preserving from $\preceq$ to $\leq$. Since $0, 2 \prec 1$, we must have $2 < 1$, and since $0 \prec 2$, we also have $0 < 1$. The two orders $0 < 2 < 1$ and $2 < 0 < 1$ satisfying these conditions are such that η is left order preserving from $\preceq$ to $\leq$.

Let us now show that the terminology planar preserving is indeed appropriate.

Proposition 5.48. *Let $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ be a return morphism which is planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$ and let $\mathcal{L} \subseteq \mathcal{A}^*$ be a language. A word $v \in \mathcal{L}$ is planar for $\preceq^L$ and $\preceq^R$ if and only if every extended image of v under σ is planar for $\leq^L$ and $\leq^R$.*

Proof. We prove that v has two bi-extensions breaking the planarity for $(\preceq^L, \preceq^R)$ if and only if there exists an extended image u of v having two bi-extensions breaking the planarity for $(\leq^L, \leq^R)$. Observe that being planar for $(\leq^L, \leq^R)$ or for $((\leq^L)^*, (\leq^R)^*)$ is equivalent. Therefore, it suffices to consider the case where σ is left order preserving from $\preceq^L$ to $\leq^L$ and right order preserving from $\preceq^R$ to $\leq^R$.

Let $(x_1, y_1), (x_2, y_2) \in E_{\mathcal{L}}(v)$ be such that $x_1 \neq x_2$ and $y_1 \neq y_2$. We define s as the longest common suffix between $\sigma(x_1)$ and $\sigma(x_2)$ and p as the longest common prefix between $\sigma(y_1)w$ and $\sigma(y_2)w$ (where σ is a return morphism for w). We then denote

$$x_1' = \varphi_s^L(x_1), \quad x_2' = \varphi_s^L(x_2), \quad y_1' = \phi_p^R(y_1) \quad \text{and} \quad y_2' = \phi_p^R(y_2).$$

Using Proposition 4.76, (x_1', y_1') and (x_2', y_2') are two bi-extensions of the extended image $u := s\sigma(v)p$ and are such that $x_1' \neq x_2'$ and $y_1' \neq y_2'$. Moreover, for any such pair of bi-extensions of an extended image u' of v, it is possible to find a corresponding pair of bi-extensions of v.

Let us show that (x_1, y_1) and (x_2, y_2) are crossing edges in $\mathcal{E}_{\mathcal{L}}(v)$ if and only if (x_1', y_1') and (x_2', y_2') are crossing edges in $\mathcal{E}_{\sigma(\mathcal{L})}(u)$. Without loss of generality, assume that $x_1 \prec^L x_2$. As σ is planar preserving, φ_s^L is order preserving from $\preceq^L$ to $\leq^L$ so we have $x_1' <^L x_2'$.

Since ϕ_p^R is order preserving from $\preceq^R$ to $\leq^R$, we conclude that

$$y_1 \prec^R y_2 \iff y_1' <^R y_2'.$$

This ends the proof by definition of planarity. □

The previous result therefore shows that being planar preserving handles the extended images. For the initial factors, we introduce a new notion.

Definition 5.49. A return morphism σ is $(\leq^L, \leq^R)$-*planar* if all of its initial factor are planar for $(\leq^L, \leq^R)$.

Example 5.50. Let us describe the pairs of orders for which the morphism η of Subsection 5.3.1 (or Example 5.47) is planar. The only bispecial initial factor is ε and its extension graph is given by case (F) of Table 5.2. One then easily checks that it is only planar for the pairs of orders $(0 <^L 1 <^L 2, 1 <^R 2 <^R 0)$ and $(2 <^L 1 <^L 0, 0 <^R 2 <^R 1)$ which are inverse of one another.

We then have the following direct consequence.

Corollary 5.51. *Let σ be a return morphism for a word and let $\mathcal{L}$ be a language. If σ is $(\leq^L, \leq^R)$-planar and planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$, then $\mathcal{L}$ is planar for $(\preceq^L, \preceq^R)$ if and only if $\sigma(\mathcal{L})$ is planar for $(\leq^L, \leq^R)$.*

We can then define a subgraph of $\mathcal{T}^L(\mathfrak{S}) \times \mathcal{T}^R(\mathfrak{S})$ using these additional properties. This does not directly give an $\mathfrak{S}$-adic characterization of languages of RIET however. Indeed, any pair of line graphs corresponds to four pairs of orders: $(\leq, \preceq)$, $(\leq, \preceq^*)$, $(\leq^*, \preceq)$ and $(\leq^*, \preceq^*)$. Clearly, reversing both orders does not impact planarity, reversing only one of the orders does however. Therefore, the information contained in the line graphs is not sufficient and we should split each pair of line graphs into two pairs of orders in which we assume for example that we always have $a < b$ for the left (or first) order for some fixed letters a, b. This would give us too many pairs of orders however. Indeed, if a pair of order corresponds to the language of an RIET, then it is irreducible meaning that we only need to consider irreducible pairs of orders. This leads to the following definition.

Definition 5.52. Let $\mathfrak{S}$ be a set of return morphism for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$ and let $a, b \in \mathcal{A}$ such that $a \neq b$. The graph $\mathcal{T}_{IET}(\mathfrak{S})$ is the graph whose vertices are the irreducible pairs of total orders $(\leq^L, \leq^R)$ on $\mathcal{A}$ such that $a <^L b$, and there is an edge from $(\leq^L, \leq^R)$ to $(\preceq^L, \preceq^R)$ labeled by $\sigma \in \mathfrak{S}$ if

- σ labels an edge from $G(\leq^L)$ to $G(\preceq^L)$ in $\mathcal{T}^L(\mathfrak{S})$ and an edge from $G(\leq^R)$ to $G(\preceq^R)$ in $\mathcal{T}^R(\mathfrak{S})$,
- σ is $(\leq^L, \leq^R)$-planar,
- σ is planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$.

Before proving that this graph indeed gives an S-adic characterization of languages of RIET, we give one last result stating that, if σ is planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$ and we know that $a \prec^L b$, then $(\preceq^L, \preceq^R)$ is entirely determined by $(\leq^L, \leq^R)$ and σ.

Lemma 5.53. *For every return morphism $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ and every total order $\leq$ on $\mathcal{A}$, there exists a unique total order $\preceq$ on $\mathcal{A}$ such that σ is left (resp., right) order preserving from $\preceq$ to $\leq$.*

Proof. We prove the result for a left order preserving morphism σ. Let us begin with the existence of such an order $\preceq$. For all $s \in \mathrm{Suff}(\sigma(\mathcal{A}))$, we will build an order $\preceq_s$ on $\mathcal{B}_s = \{a \in \mathcal{A} : s \in \mathrm{Suff}(\sigma(a))\}$[1] such that, for all $s' \in \mathcal{A}^* s$, the map $\varphi^L_{s'}$ is order preserving from $\preceq_s$ to $\leq$. The conclusion will follow with $s = \varepsilon$.

We iteratively build this order $\preceq_s$, starting with s maximal, i.e., such that for all $a \in \mathcal{A}$, $\mathcal{B}_{as}$ is empty. Then, since σ is a return morphism, $s \in \sigma(\mathcal{A})$ and $\mathcal{B}_s$ contains a unique element thus $\preceq_s$ is a trivial order.

Assume now that s is not maximal and that we have the orders $\preceq_{as}$ for all $a \in \varphi^L_s(\mathcal{A})$. Since $\sigma(\mathcal{A})$ is a suffix code, the sets $\mathcal{B}_{as}$ form a partition of $\mathcal{B}_s$ thus, we can define the order $\preceq_s$ on $\mathcal{B}_s$ by $x \prec_s y$ if

1. $x, y \in \mathcal{B}_{as}$ and $x \prec_{as} y$, or
2. $x \in \mathcal{B}_{as}$, $y \in \mathcal{B}_{bs}$ and $a < b$.

For all $s' \in \mathcal{A}^* as$, $\varphi^L_{s'}$ is order preserving from $\preceq_{as}$ to $\leq$ thus it is order preserving from $\preceq_s$ to $\leq$. Let us show that φ^L_s is also order preserving from $\preceq_s$ to $\leq$. If $x, y \in \mathcal{B}_s$ are as in case 1, then

$$\varphi^L_s(x) = a = \varphi^L_s(y)$$

[1] Notice the difference with $\mathrm{dom}(\varphi^L_s) = \{a \in \mathcal{A} : s \in \mathrm{Suff}^*(\sigma(a))\}$.

and if they are as in case 2, then

$$\varphi_s^L(x) = a < b = \varphi_s^L(y).$$

Thus, if $x \prec_s y$, we have $\varphi_s^L(x) \leq \varphi_s^L(y)$.

We now prove the uniqueness. Assume that σ is left order preserving from $\preceq$ to $\leq$ and from $\preceq'$ to $\leq$ and let $x, y \in \mathcal{A}$ be such that $x \prec y$ and $y \prec' x$. If s is the longest common suffix between $\sigma(x)$ and $\sigma(y)$, then $\varphi_s^L(x) \neq \varphi_s^L(y)$. Since σ is left order preserving from $\preceq$ to $\leq$, this implies that $\varphi_s^L(x) < \varphi_s^L(y)$. Since σ is also left order preserving from $\preceq'$ to $\leq$, we have the converse inequality, which is a contradiction. $\square$

Example 5.54. Since the morphism η of Example 5.47 is $(0 <^L 1 <^L 2, 1 <^R 2 <^R 0)$-planar by Example 5.50, let us find the pair $(\preceq^L, \preceq^R)$ of orders such that η is planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$ and $0 \prec^L 1$. Given the left maps recalled in Example 5.47, we first define $\preceq_2^L$ by saying that $0 \prec_2^L 2$ since $0 <^L 1$. Then $\preceq_\varepsilon^L$ is given by $1 \prec_\varepsilon^L 0 \prec_\varepsilon^L 2$ since $1 <^L 2$. As we want to have $0 \prec^L 1$, we will reverse $\preceq_\varepsilon^L$ to obtain $\preceq^L$. This means that we will also need to reverse $\preceq_\varepsilon^R$ to obtain $\preceq^R$.

The right maps are

$$\phi_0^R \colon \begin{cases} 0 \mapsto 2 \\ 1, 2 \mapsto 1 \end{cases} \quad \text{and} \quad \phi_{01}^R \colon \begin{cases} 1 \mapsto 0 \\ 2 \mapsto 2 \end{cases}$$

by Table 5.3. Therefore, we have $2 \prec_{01}^R 1$ since $2 <^R 0$, and $2 \prec_0^R 1 \prec_0^R 0$ since $1 <^R 2$. Since $\preceq_\varepsilon^R = \preceq_0^R$, we deduce that $0 \prec^R 1 \prec^R 2$.

Lemma 5.53 has two consequences. The first one is that we can use it to reprove the fact that languages of RIET are stable under derivation as explained below.

Remark 5.55. Let $\mathcal{L}$ be the language of an RIET for the orders $\binom{\leq^R}{\leq^L}$, let $w \in \mathcal{L} \setminus \{\varepsilon\}$ and let σ be the return morphism such that $\mathcal{L} = \sigma(D_w(\mathcal{L}))$. By definition, σ is $(\leq^L, \leq^R)$-planar and, by Lemma 5.53, there exists unique orders (up to reversal) such that σ is planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$. Then, by Corollary 5.51, we deduce that $D_w(\mathcal{L})$ is planar for $(\preceq^L, \preceq^R)$. Since, by Theorem 3.42, $D_w(\mathcal{L})$ is also a recurrent dendric language, we conclude that it is the language of an RIET for the orders $\binom{\preceq^R}{\preceq^L}$ by Proposition 2.8.

The second consequence of Lemma 5.53 is that the graph $\mathcal{T}_{IET}(\mathfrak{S})$ is deterministic in the sense that, for each vertex and each morphism, there is at most one edge labeled by the morphism leaving that vertex.

We can now prove a result similar to Theorem 5.20 but in the case of RIET.

Theorem 5.56. *Let $\mathfrak{S}$ be a set of return morphism for a word from $\mathcal{A}^*$ to $\mathcal{A}^*$. A language $\mathcal{L}$ having an $\mathfrak{S}$-adic representation $\boldsymbol{\sigma}$ is the language of an RIET if and only if $\boldsymbol{\sigma}$ is primitive and labels an infinite path in the graph $\mathcal{T}_{IET}(\mathfrak{S})$. Moreover, if this path starts in $(\leq^L, \leq^R)$, then $\mathcal{L}$ is the language of an RIET for the orders $\binom{\leq^R}{\leq^L}$.*

Proof. Assume first that $\mathcal{L}$ is the language of an RIET. By Theorem 5.26, $\boldsymbol{\sigma}$ is primitive and labels an infinite path in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$. Moreover, if $\mathcal{L}^{(i)}$ denotes the languages generated by $(\sigma_n)_{n\geq i}$ then by Remark 5.55, $\mathcal{L}^{(i)}$ is the language of an RIET for some orders $\binom{\leq_i^R}{\leq_i^L}$ such that $a <_i^L b$, where a and b are the letters used to define $\mathcal{T}_{IET}(\mathfrak{S})$. By uniqueness of these orders (Proposition 1.40) and by Remark 5.55 once again, σ_i is $(\leq_i^L, \leq_i^R)$-planar and planar preserving from $(\leq_{i+1}^L, \leq_{i+1}^R)$ to $(\leq_i^L, \leq_i^R)$. This shows that $\boldsymbol{\sigma}$ labels an infinite path in $\mathcal{T}_{IET}(\mathfrak{S})$ starting in $(\leq_0^L, \leq_0^R)$.

Assume now that $\boldsymbol{\sigma}$ is primitive and labels a path $((\leq_n^L, \leq_n^R))_{n\geq 0}$ in $\mathcal{T}_{IET}(\mathfrak{S})$. Since any path in $\mathcal{T}_{IET}(\mathfrak{S})$ is a path in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$, $\mathcal{L}$ is recurrent dendric by Theorem 5.26. Let us show that $\mathcal{L}$ is planar for the pair of orders $(\leq_0^L, \leq_0^R)$. The proof is in fact similar to the dendric case (Theorem 5.20).

Let $u \in \mathcal{L}$. Iterating Proposition 4.51, there exists $k \geq 0$ such that the extensions of u in $\mathcal{L}$ are entirely determined by the fact that $\mathcal{L}$ is the image of some language under the morphism $\sigma_0 \cdots \sigma_k$. Therefore, it suffices to prove that there exists a language $\mathcal{L}'$ such that $\sigma_0 \cdots \sigma_k(\mathcal{L}')$ is planar for $(\leq_0^L, \leq_0^R)$. By definition of $\mathcal{T}_{IET}(\mathfrak{S})$, the pair $(\leq_{k+1}^L, \leq_{k+1}^R)$ is irreducible thus there exists a language $\mathcal{L}'$ of an RIET for the orders $\binom{\leq_{k+1}^R}{\leq_{k+1}^L}$. Since $\sigma_0 \cdots \sigma_k$ labels a path from $(\leq_0^L, \leq_0^R)$ to $(\leq_{k+1}^L, \leq_{k+1}^R)$, by Corollary 5.51 and by definition of $\mathcal{T}_{IET}(\mathfrak{S})$, this implies that $\sigma_0 \cdots \sigma_k(\mathcal{L}')$ is planar for $(\leq_0^L, \leq_0^R)$. We can then conclude that any $u \in \mathcal{L}$ is planar for $(\leq_0^L, \leq_0^R)$. $\square$

As in the dendric case, we can moreover use permutations to reduce the number of vertices. For example, we can completely fix the left order.

Let us now give an example of a use of Theorem 5.56 by giving an S-adic characterization of languages of RIET on the alphabet $\{0, 1, 2\}$.

The first thing to notice is that, on an alphabet of size 3, any pair of line graphs corresponds to exactly one irreducible pair of orders $(\leq^L, \leq^R)$ such that $a <^L b$ for some fixed a and b. This implies that $\mathcal{T}_{IET}(\mathfrak{S})$ can in

σ	$\binom{\leq^L}{\leq^R}$	$\binom{\preceq^L}{\preceq^R}$
α	$\binom{0<a<b}{a'<b'<0}$	$\binom{0\prec a\prec b}{a'\prec b'\prec 0}$
β	$\binom{2<0<1}{1<2<0}$	$\binom{0\prec 2\prec 1}{1\prec 2\prec 0}$
γ	$\binom{0<2<1}{1<0<2}$	$\binom{0\prec 2\prec 1}{1\prec 2\prec 0}$
δ_k	$\binom{0<2<1}{1<0<2}$	$\binom{0\prec 2\prec 1}{1\prec 2\prec 0}$
ζ_k	$\binom{0<2<1}{1<2<0}$	$\binom{0\prec 2\prec 1}{1\prec 2\prec 0}$
η	$\binom{0<1<2}{1<2<0}$	$\binom{2\prec 0\prec 1}{0\prec 1\prec 2}$

Table 5.9: For each morphism σ of Table 5.3, $\binom{\leq^L}{\leq^R}$ are the pairs of orders for which σ is $(\leq^L, \leq^R)$-planar (for the morphism α, we assume that $\{a, b\} = \{1, 2\} = \{a', b'\}$) such that $0 <^L 1$, and $\binom{\preceq^L}{\preceq^R}$ are the corresponding pairs of orders such that $0 \prec^L 1$ and σ is planar preserving from $(\preceq^L, \preceq^R)$ to $(\leq^L, \leq^R)$.

fact be seen as a subgraph of $\mathcal{T}^L(\mathfrak{S}) \times \mathcal{T}^R(\mathfrak{S})$. The same can be said for the reduced graph after the use of permutations. In other words, to obtain an $\mathfrak{S}$-adic characterization of languages of RIET on the alphabet $\{0, 1, 2\}$, it suffices to start from the graph represented in Figure 5.3 and remove the edges that do not respect the planarity condition of Definition 5.52.

We summarize the information regarding planarity of the six families of morphisms of Subsection 5.3.1 in Table 5.9. This can be obtained as in Examples 5.50 and 5.54 where we handled the case of the morphism η. Using this information, we can select the appropriate edges of the graph of Figure 5.3 to obtain Figure 5.7. Recall that, at any point, we can reverse both orders in a pair since it does not impact planarity (or languages of RIET).

$$\begin{tikzcd}
\binom{0<2<1}{2<1<0} \arrow[loop left, "{\alpha,\ \pi_{021}\eta\pi_{210}}"] \arrow[rr, "{\pi_{201}\beta,\ \pi_{102}\gamma,\ \pi_{102}\delta_k}"] & & \binom{0<2<1}{1<2<0} \arrow[loop right, "{\pi_{102},\ \alpha,\ \zeta_k}"]
\end{tikzcd}$$

Figure 5.7: A language over $\{0, 1, 2\}$ is the language of an RIET if and only if, up to permutation, it has a primitive S-adic representation labeling an infinite path in this graph where where $\pi_{i_0 i_1 i_2}$ is the morphism such that $\pi_{i_0 i_1 i_2}(j) = i_j$ and the other morphisms are defined in Table 5.3.

5.4 Decidability of dendric substitutive languages

Some of the most studied languages are the morphic (or substitutive) languages, which are characterized by having an eventually periodic S-adic representation. More precisely, we have the following definition.

Definition 5.57. A language $\mathcal{L}$ is *morphic for* (τ, σ) if $\sigma\colon \mathcal{B}^* \to \mathcal{B}^*$ and $\tau\colon \mathcal{B}^* \to \mathcal{A}^*$ are such that

$$\mathcal{L} = \bigcup_{n \geq 0} \operatorname{Fac}(\tau(\sigma^n(\mathcal{B}))).$$

If there exists such a pair (τ, σ), then $\mathcal{L}$ is said to be *morphic*, and if moreover we can choose τ as the identity, then $\mathcal{L}$ is said to be *purely morphic*.

In this section, we show that, for uniformly recurrent morphic languages, we can decide (eventual) dendricity. This was first stated and proved in [GL22]. Since the results presented here rely heavily on the work of Durand on morphic sequences in [Dur98, Dur13b], we will take the viewpoint of morphic one-sided infinite word.

Definition 5.58. A right infinite word $x \in \mathcal{A}^{\mathbb{N}}$ is *morphic* if there exist $\sigma\colon \mathcal{B}^* \to \mathcal{B}^*$ prolongable on a and $\tau\colon \mathcal{B}^* \to \mathcal{A}^*$ such that

$$x = \tau(\lim_{n\to\infty} \sigma^n(a)) = \tau(\sigma^\omega(a)).$$

In the context of uniformly recurrent languages, looking at sequences instead is not a restriction as shown below.

Lemma 5.59. *If $\mathcal{L}$ is a uniformly recurrent morphic language for (τ, σ), then it is the language of a uniformly recurrent morphic sequence x and one can compute morphisms σ' and τ' such that $x = \tau'(\sigma'^\omega(b))$.*

Proof. Let B_g, B_b be the subalphabets of growing and bounded letters, i.e.,

$$\mathcal{B}_g = \{a \in \mathcal{B} : \lim_{n\to+\infty} |\sigma^n(a)| = +\infty\}$$
$$\mathcal{B}_b = \{a \in \mathcal{B} : (|\sigma^n(a)|)_{n\geq 0} \text{ is bounded}\}.$$

Since $\mathcal{L}$ is infinite, there exists a growing letter $a \in B_g$. Since a is growing, for all $n \geq 0$, $\sigma^n(a)$ contains a growing letter so let us define $u_n \in \mathcal{B}_b^*$ and $a_n \in \mathcal{B}_g$ such that $u_n a_n$ is a prefix of $\sigma^n(a)$.

By the pigeonhole principle, we can find $k < \ell$ such that $a_k = a_\ell = b$. If we define $\varphi = \sigma^{\ell-k}$, then $\varphi(u_k)\varphi(b)$ is a prefix of $\varphi(\sigma^k(a)) = \sigma^\ell(a)$ and is therefore prefix comparable with $u_\ell b$. Since the images of bounded letters only contain bounded letters, we have $\varphi(u_k) \in \mathcal{B}_b^*$, and since b is growing, $\varphi(b)$ contains a growing letter. This shows that $\varphi(b)$ has a prefix ub with $u \in \mathcal{B}_b^*$.

In particular, $\varphi^n(u)\cdots\varphi(u)ub$ is a prefix of $\varphi^{n+1}(b)$ for all $n \geq 0$. However, since $u \in \mathcal{B}_b^*$, the set $\{\varphi^n(u) : n \geq 0\}$ is finite so, by the pigeonhole principle, there exists $k' < \ell'$ such that $\varphi^{k'}(u) = \varphi^{\ell'}(u)$. Let $v = \varphi^{\ell'}(u)\varphi^{\ell'-1}(u)\cdots\varphi^{k'+1}(u)$. Then, for all $n \geq 0$, v^n is a prefix of $\varphi^{n(\ell'-k')+k'+1}(b)$ which implies that $\tau(v)^n \in \mathcal{L}$ for all $n \geq 0$.

Since σ and τ are non-erasing, u is empty if and only if $\tau(v)$ is.

If $u \neq \varepsilon$, then by uniform recurrence of $\mathcal{L}$, for all $m \geq 0$, there exists n such that $\mathcal{L}_m \subseteq \mathrm{Fac}(\tau(v)^n)$. This shows that $\mathcal{L} = \mathrm{Fac}(\tau(v)^\omega)$ and $\tau(v)^\omega = \tau(\sigma'^\omega(v_1))$ where $\sigma'(c) = v^2$ for all $c \in \mathcal{B}$.

If $u = \varepsilon$, then φ is prolongable on b. Since b is growing and σ is non-erasing, $\lim_{n\to\infty}\varphi^n(b)$ is an infinite word x such that $\mathcal{L}(\tau(x)) \subseteq \mathcal{L}$. Since $\mathcal{L}$ is uniformly recurrent, we then have $\mathcal{L} = \mathcal{L}(\tau(\varphi^\omega(b)))$. □

For non-recurrent languages however, the previous result is not necessarily true, as shown in the example below.

Example 5.60. Let $\mathcal{L} = \{0^n : n \geq 0\} \cup \{0^n 1 0^m : n, m \geq 0\}$. This language is purely morphic and generated by the morphism σ such that $\sigma(0) = 0$ and $\sigma(1) = 010$. It is however not the language of a right infinite word (see Example 1.22).

Some decidability properties are already well-known for morphic words.

Proposition 5.61. *Let $x = \tau(\sigma^\omega(a))$ be a morphic word in $\mathcal{A}^{\mathbb{N}}$.*

1. *There exists an algorithm to obtain a non-erasing morphism σ' and a letter-to-letter morphism τ' such that $x = \tau'(\sigma'^\omega(a))$.*

2. *We can decide if x is eventually periodic;*

3. *We can decide if σ is growing, i.e.,* $\lim_{n\to\infty} |\sigma^n(a)| = +\infty$ *for all* $a \in \mathcal{B}$*;*

4. *If x is aperiodic uniformly recurrent, there is an algorithm to obtain a growing morphism σ' and a morphism τ' such that $x = \tau'(\sigma'^\omega(a))$.*

Proof. The first claim is a famous result by Cobham, an algorithmic construction was for example given in [Hon09]. The second claim is proved in [Dur13a] while the third claim is a consequence of [Dur13b, Proposition 5]. Finally, the last claim follows from a result by Pansiot [Pan84]. □

By definition of a morphic language, we know one of its S-adic representations. However, the results presented in Section 5.2 require S-adic representations using return morphisms exclusively. While we know the existence of such a representation if $\mathcal{L}$ is uniformly recurrent by Proposition 5.7, we need to be more precise in the construction to obtain decidability.

The idea is to iteratively derive with respect to a prefix of the morphic word corresponding to $\mathcal{L}$. For this, we introduce some notations.

Definition 5.62. Let $x = \tau(\sigma^\omega(a)) \in \mathcal{A}^\mathbb{N}$ be a uniformly recurrent morphic word and let u be a non-empty prefix of x. We denote $\mathcal{R}(u) = \{0, \ldots, \#(\mathrm{R}_\mathcal{L}(u))-1\}$ and define $\theta_u \colon \mathcal{R}(u)^* \to \mathcal{A}^*$ so that $\theta_u(i)$ is the $(i+1)$-th return word for u appearing in x. In other words, for all $i \in \mathcal{R}(u)$, if the first occurrence of $\theta_u(i)u$ in x starts at index k, then $x_{[0,k)}$ belongs to $\theta_u(\{0, \ldots, i-1\})^*$.

In the rest of this section, we assume that, if u is a prefix of x, then $D_u(x)$ is the derived sequence corresponding to the morphism θ_u. We then have the following theorem which is a combination of results by Durand [Dur13b].

Theorem 5.63. *Let $x = \tau(\sigma^\omega(a))$ be an aperiodic uniformly recurrent morphic word in $\mathcal{A}^\mathbb{N}$.*

1. *For every non-empty prefix u of x, the morphism θ_u is computable and there exist some computable morphisms $\sigma_u \colon \mathcal{B}^* \to \mathcal{B}^*$ and $\tau_u \colon \mathcal{B}^* \to \mathcal{R}(u)^*$ such that $D_u(x) = \tau_u(\sigma_u^\omega(0))$.*

2. *There is a computable constant D such that the set of pairs (σ_u, τ_u) has cardinality at most D. In particular, the number of derived sequences of x (on its non-empty prefixes) is at most D.*

Proof. By Proposition 5.61, we can algorithmically modify τ and σ so that σ is growing. Then, the existence of σ_u and τ_u is given by [Dur13b, Proposition 28] and the fact that θ_u, σ_u and τ_u are computable is explained in [Dur13b, Section 4]. The bound on the number of possible pairs (σ_u, τ_u) is in [Dur13b, Theorem 29]. □

Using this result, we can then build an alternative S-adic representation of $\mathcal{L}(\tau(\sigma^\omega(a)))$ using only return morphisms.

Theorem 5.64. *Let $x = \tau(\sigma^\omega(a)) \in \mathcal{A}^\mathbb{N}$ be an aperiodic uniformly recurrent morphic word. One can algorithmically compute (from (σ, τ, a)) two return morphisms for a word $\theta\colon \mathcal{B}^* \to \mathcal{B}^*$ and $\lambda\colon \mathcal{B}^* \to \mathcal{A}^*$ such that θ is primitive and $(\lambda, \theta, \theta, \dots)$ is an S-adic representation of $\mathcal{L}(x)$.*

Proof. We can iteratively derive x with respect to the first letter. More precisely, we define $x^{(0)} = x$ and $x^{(n+1)} = D_{x_0^{(n)}}(x^{(n)})$ for all $n \geq 0$. By Theorem 5.63, $x^{(n)}$ is also an aperiodic uniformly recurrent morphic word and we can compute the morphism θ_n such that $x^{(n)} = \theta_n(x^{(n+1)})$. Note that $(\theta_n)_{n\geq 0}$ is a primitive S-adic representation of $\mathcal{L}(x)$ using only return morphisms.

Observe also that $x^{(n)}$ is then the derived sequence of x with respect to some prefix $u^{(n)}$. Therefore, by Theorem 5.63 again, we can compute some morphisms τ_n, σ_n such that $x^{(n)} = \tau_n(\sigma_n^\omega(0))$. Moreover, there exists $k < \ell < D$, where D is a computable constant, such that $(\sigma_k, \tau_k) = (\sigma_\ell, \tau_\ell)$.

This implies that $x^{(k)} = x^{(\ell)}$ and, the construction of $(x^{(n)})_{n\geq 0}$ being deterministic, that, for all $n \geq k$, $x^{(n)} = x^{(n-k+\ell)}$, so that $\theta_n = \theta_{n-k+\ell}$. We can then define $\lambda = \theta_0 \circ \cdots \circ \theta_{k-1}$ and $\theta = \theta_k \circ \cdots \circ \theta_{\ell-1}$. By construction, λ and θ are return morphisms and $(\lambda, \theta, \theta, \dots)$ is an S-adic representation of $\mathcal{L}(x)$ such that θ is primitive. □

Let us build this S-adic representation for the Thue-Morse language.

Example 5.65. Let $x = \sigma^\omega(0)$ be the Thue-Morse word, i.e., σ satisfies $\sigma(0) = 01$ and $\sigma(1) = 10$. Note that, since x is purely morphic, we have $\tau_n = \mathrm{id}$ for all n. To start the construction, let us find the return words for 0 (in order of appearance).

We first iterate σ on 0 to obtain two occurrences of 0. Since $\sigma^2(0) = 0110$, the first return word is 011. Moreover, since σ is prolongable on 0, $\sigma(011)$ is a concatenation of return words. Since $\sigma(011) = 011010$, the second and third return words are 01 and 0 respectively. Once again, $\sigma(01) = 0110$ and $\sigma(0) = 01$ are concatenations of return words, but there is no new return word this time. This shows that x can be written as a concatenation of 011, 01 and 0, i.e., we have found all return words. We then define $\theta_0(0) = 011$, $\theta_0(1) = 01$ and $\theta_0(2) = 0$.

Since $\sigma(\theta_0(0)) = \theta_0(0)\theta_0(1)\theta_0(2)$, $\sigma(\theta_0(1)) = \theta_0(0)\theta_0(2)$ and $\sigma(\theta_0(2)) = \theta_0(1)$, we have $D_0(x) = \sigma_1^\omega(0)$ where $\sigma_1(0) = 012$, $\sigma_1(1) = 02$ and $\sigma_1(2) = 1$.

We now repeat this construction for the sequence $x^{(1)} = \sigma_1^\omega(0)$. We skip the details here and obtain

$$\theta_1\colon \begin{cases} 0 \mapsto 012 \\ 1 \mapsto 021 \\ 2 \mapsto 0121 \\ 3 \mapsto 02 \end{cases} \quad \text{and} \quad \sigma_2\colon \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 23 \\ 2 \mapsto 013 \\ 3 \mapsto 2 \end{cases}.$$

As $\sigma_2 \notin \{\sigma, \sigma_1\}$, we keep iterating on $x^{(2)} = \sigma_2^\omega(0)$ to obtain

$$\theta_2\colon \begin{cases} 0 \mapsto 0123 \\ 1 \mapsto 0132 \\ 2 \mapsto 01232 \\ 3 \mapsto 013 \end{cases} \quad \text{and} \quad \sigma_3\colon \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 23 \\ 2 \mapsto 013 \\ 3 \mapsto 2 \end{cases}.$$

We see here that $\sigma_3 = \sigma_2$, showing that $(\theta_0 \circ \theta_1, \theta_2, \theta_2, \dots)$ is a primitive S-adic representation of the Thue-Morse language, using only return morphisms.

We finally deduce the decidability of (eventual) dendricity using the S-adic characterizations that we obtained in Section 5.2.

Theorem 5.66. *Let $x = \tau(\sigma^\omega(a))$ be an aperiodic uniformly recurrent morphic word and let λ, θ be the computable morphisms given by Theorem 5.64.*

1. *The word x is eventually dendric if and only if the sequence $(\theta, \theta, \dots)$ labels an infinite path in $\mathcal{T}^L(\{\theta\})$ and in $\mathcal{T}^R(\{\theta\})$.*

2. *The word x is dendric if and only if the sequence $(\lambda, \theta, \theta, \dots)$ labels an infinite path in $\mathcal{T}^L(\{\lambda, \theta\})$ and in $\mathcal{T}^R(\{\lambda, \theta\})$.*

In particular, if $\mathcal{L}$ is a uniformly recurrent morphic language for (τ, σ), one can decide (based on τ and σ) if $\mathcal{L}$ is dendric and/or eventually dendric. Moreover, if $\mathcal{L}$ is eventually dendric, the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ are then computable, as well as the dendricity threshold.

Proof. Since λ and θ are return morphisms for a word, θ is primitive and $(\lambda, \theta, \theta, \dots)$ is an S-adic characterization of $\mathcal{L}(x)$, the characterization of the (eventual) dendricity of x follows from Theorem 5.26 and Corollary 5.27.

If $\mathcal{L}$ is a uniformly recurrent morphic language, then we can construct morphisms τ' and σ' such that $\mathcal{L} = \mathcal{L}(x)$ with $x = \tau'(\sigma'^{\omega}(a)))$ by Lemma 5.59. By Proposition 5.61, we can decide if x is eventually periodic, in which case it is always eventually dendric, and it is dendric if and only if it is on a unary alphabet. Assuming therefore that x is aperiodic, the corresponding morphisms λ and θ are computable by Theorem 5.64. The above conditions for (eventual) dendricity are then decidable since the graphs $\mathcal{T}^L(\{\theta\})$, $\mathcal{T}^R(\{\theta\})$, $\mathcal{T}^L(\{\lambda, \theta\})$ and $\mathcal{T}^R(\{\lambda, \theta\})$ are computable by Remark 5.19.

In the case where x is eventually periodic, the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ have no edge, and the computability of the dendricity threshold follows from the fact that we can in fact compute the words u and v such that $x = uv^{\omega}$.

Assume that $\mathcal{L}$ is eventually dendric but x is not eventually periodic, and let $\mathcal{L}'$ be the language generated by θ. It is dendric so, by Proposition 5.28, its graph $G^L(\mathcal{L}')$ (resp., $G^R(\mathcal{L}')$) is entirely determined by the possible starting vertices of an infinite path in $\mathcal{T}^L(\{\theta\})$ (resp., $\mathcal{T}^R(\{\theta\})$). As this graph is finite, this shows that both $G^L(\mathcal{L}')$ and $G^R(\mathcal{L}')$ are computable, and so are $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ by Proposition 5.10.

Let us show that the dendricity threshold is computable as well. As $\mathcal{L}'$ is dendric, if $G^L_N(\mathcal{L}') = G^L(\mathcal{L}')$ and $G^R_N(\mathcal{L}') = G^R(\mathcal{L}')$ for some N, then all words of $\mathcal{L}'_{\geq N}$ are ordinary. By Proposition 4.93, $\mathcal{L}$ is then eventually dendric of threshold at most $\|\lambda\|(N+1)+|w|-1$ if λ is a return morphism for w. As x is not eventually periodic, we can moreover show that $|w| \leq 2\|\lambda\|$. Indeed, otherwise for every return word u for w in x, $|u|$ is a period of w smaller than $\frac{|w|}{2}$ so, by Fine and Wilf's Theorem, all return words for w are power of a common word and x is eventually periodic. In the end, this proves that $\mathcal{L}$ is eventually dendric of threshold at most $\|\lambda\|(N+3)-1$, where N is computable. The exact threshold can then be found by considering the extension graphs of all small words. $\square$

Example 5.67. Let us continue Example 5.65 on the Thue-Morse language. We can easily see that θ_2 is not dendric since ε is neither acyclic nor connected. This directly shows that $\theta = \theta_2$ will not label any edge in the graphs of Theorem 5.66 and that $\mathcal{L}(x)$ is neither dendric nor eventually dendric. Observe that this fact is not new and was already mentioned in Section 3.2 since $\mathcal{L}(x)$ contains infinitely many strong and weak words.

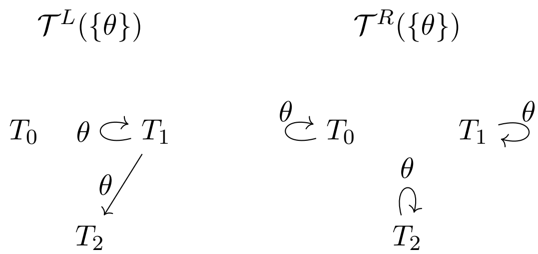


Figure 5.8: Graphs $\mathcal{T}^L(\{\theta\})$ and $\mathcal{T}^R(\{\theta\})$ for the morphism θ of Example 5.68, where T_i is the tree with vertices $0, 1, 2$ where i is the degree-2 vertex.

Example 5.68. Let us consider another example where $\mathcal{L}$ is morphic for (τ, σ) such that

$$\tau\colon \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 0 \\ 2 \mapsto 1 \end{cases} \quad \text{and} \quad \sigma\colon \begin{cases} 0 \mapsto 202 \\ 1 \mapsto 2102102102 \\ 2 \mapsto 2102102 \end{cases}.$$

Using a simple computation, we obtain the alternative S-adic representation $(\lambda, \theta, \theta, ...)$ where

$$\lambda\colon \begin{cases} 0 \mapsto 100 \\ 1 \mapsto 1001 \\ 2 \mapsto 1001101 \end{cases} \quad \text{and} \quad \theta\colon \begin{cases} 0 \mapsto 01002 \\ 1 \mapsto 0100201 \\ 2 \mapsto 01002010201 \end{cases}$$

are return morphisms for 100 and for 0100 respectively. Clearly, the morphism λ is not dendric as ε is not acyclic so, by Theorem 5.66, $\mathcal{L}$ is not dendric. However, the graphs $\mathcal{T}^L(\{\theta\})$ and $\mathcal{T}^R(\{\theta\})$ represented in Figure 5.8 contain infinite paths. This implies that $\mathcal{L}$ is eventually dendric.

From Figure 5.8, we also deduce that, if $\mathcal{L}'$ is the dendric language generated by θ, then $G^L(\mathcal{L}') = T_1$ and $G^R(\mathcal{L}') = K_3$ where T_1 is the tree with degree-2 vertex 1 and K_3 is the complete graph. This in turn implies that $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ both have two edges between 0 and 1 by Proposition 5.10 and, as $\mathcal{L}'$ is eventually ordinary of threshold 2, that $\mathcal{L}$ is eventually dendric of threshold at most $\|\lambda\|(2+1) + |0100| - 1 = 24$ by Proposition 4.93. One then easily checks that $\mathcal{L}$ is in fact eventually dendric of threshold 1.

5.5 Open questions

The questions presented here are mostly centered around the construction of an effective graph giving an S-adic characterization of dendric languages on a given alphabet, as done in Subsection 5.3.1 for the alphabet $\{0, 1, 2\}$.

The approach used in Subsection 5.3.1, while theoretically applicable to any alphabet, can soon become tedious. Indeed, the number of different Rauzy graphs grows with the alphabet size. For example, there are, up to permutations, 66 different Rauzy graphs of order 1 for recurrent dendric languages on an alphabet of size 4, and 1182 on an alphabet of size 5. Moreover, larger Rauzy graphs also mean potentially more complicated return paths.

To our knowledge, there is however no simple or automatic method to obtain a description of the dendric return morphisms corresponding to a given Rauzy graph and labeling the same edges in $\mathcal{T}^L(\mathfrak{S})$ and $\mathcal{T}^R(\mathfrak{S})$ so this must be done by hand for each Rauzy graph. Therefore, such a technique can only realistically be used on larger alphabets if the following question admits a positive answer.

Question 5.1. Given a graph G with set of vertices $\mathcal{A}$, can we give an effective description of the sets S of return words for which there exists a recurrent dendric language $\mathcal{L}$ over $\mathcal{A}$ such that $G = \Gamma_1(\mathcal{L})$ and $S = \mathrm{R}_{\mathcal{L}}(\ell)$ for some $\ell \in \mathcal{A}$?

An alternative approach would be to directly obtain an $\mathfrak{S}_e$-adic characterization where $\mathfrak{S}_e$ is the set of elementary morphisms, as in the graph of Figure 5.4. However, the main advantage of return morphisms is that they behaved well with respect to the graphs $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$. Indeed, we can decide the dendricity of the image and the graphs $G^L(\sigma(\mathcal{L}))$ and $G^R(\sigma(\mathcal{L}))$ if σ is a return morphism.

For elementary morphisms, we can still determine if the image is dendric based on $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ by Proposition 4.86 since elementary morphisms are left or right return morphisms for a set of letters (staying on the same alphabet). However, the information contained in $G^L(\mathcal{L})$ and $G^R(\mathcal{L})$ is not sufficient to determine the graphs $G^L(\sigma(\mathcal{L}))$ and $G^R(\sigma(\mathcal{L}))$ if $\sigma \in \mathfrak{S}_e$ (unless σ is a permutation). This leads to the following voluntarily vague question.

Question 5.2. Can we associate with each (eventually) dendric language $\mathcal{L}$ a finite object $o(\mathcal{L})$ similar to $(G^L(\mathcal{L}), G^R(\mathcal{L}))$ such that, if $\sigma \in \mathfrak{S}_e$, $o(\sigma(\mathcal{L}))$ only depends on σ and $o(\mathcal{L})$?

A method to define such an object would be to study the elementary graph obtained in the case of a ternary alphabet (Figure 5.4). More precisely, an answer to the following question could help.

Question 5.3. Let G be the graph of Figure 5.4. For each vertex, can we characterize the dendric languages having an $\mathfrak{S}_e$-adic representation labeling an infinite path starting in this vertex?

Another advantage of return morphisms was that, if $(\sigma_n)_{n\geq 0}$ is an S-adic representation of a dendric language, then $(\sigma_n)_{n\geq N}$ is also the S-adic representation of a dendric language for all $N \geq 0$. This is false if the morphisms σ_n are elementary. In other words, if we desubstitute a dendric language with respect to an elementary morphism, we might not get a dendric language. However, since the dendric return morphisms are tame, every recurrent dendric language has an $\mathfrak{S}_e$-adic representation $(\sigma_n)_{n\geq 0}$ such that, for infinitely many $N \geq 0$, $(\sigma_n)_{n\geq N}$ is an $\mathfrak{S}_e$-adic representation of a dendric language. This still leaves the following open question.

Question 5.4. Does every recurrent dendric language admit an $\mathfrak{S}_e$-adic representation $(\sigma_n)_{n\geq 0}$ such that, for all $N \geq 0$, $(\sigma_n)_{n\geq N}$ is an $\mathfrak{S}_e$-adic representation of a dendric language?

In the case of a dendric language over a ternary alphabet, the answer is yes as shown by Theorem 5.39. It followed from the fact that we could find appropriate decompositions of the return morphisms. The following question asks whether this is possible in general. A positive answer to this question would lead to a positive answer to the previous question.

Question 5.5. Let σ be a dendric return morphism and $\mathcal{L}$ be a dendric language such that $\sigma(\mathcal{L})$ is dendric. Does there exist an elementary decomposition $\tau_1 \circ \cdots \circ \tau_n$ of σ such that $\tau_i \circ \cdots \circ \tau_n(\mathcal{L})$ is dendric for all $i \leq n$?

In the case of a return morphism for a letter, we have the following alternative equivalent statement.

Question 5.6. Let σ be a dendric return morphism for a letter and let (G^L, G^R) be a pair of graphs such that the conditions of Proposition 4.82 are satisfied for σ. Does there exist an elementary decomposition $\tau_1 \circ \cdots \circ \tau_n$ of σ such that $\tau_i \circ \cdots \circ \tau_n$ satisfies the conditions of Corollary 4.91 for the graphs G^L and G^R for all $i \leq n$?

Finally, the existence of a finite graph giving an $\mathfrak{S}_e$-adic characterization of recurrent dendric languages is related to the following question.

Question 5.7. Given an alphabet $\mathcal{A}$, does there exist a regular language $\mathfrak{L}$ over $\mathfrak{S}_e$ containing at least one elementary decomposition of each dendric return morphism for a (bispecial) letter on $\mathcal{A}$ and such that the composition of any element of $\mathfrak{L}$ gives a dendric return morphism for a (bispecial) letter?

Observe that the key here is to be able to identify dendric return morphisms among tame return morphisms. Indeed, a regular language of elementary decompositions of tame return morphisms for a letter was given in Corollary 4.46. Moreover, note that we only ask $\mathfrak{L}$ to contain at least one decomposition per morphism and not all decompositions. Indeed, one can show that, on an alphabet of size at least 4, the language of all elementary decompositions of dendric return morphisms for a (bispecial) letter is not regular.

We end with some questions on Section 5.4. The following is quite natural.

Question 5.8. Is the (eventual) dendricity of a (potentially non uniformly recurrent) morphic language decidable?

Indeed, the construction of S-adic representation presented in this chapter relies on the possibility of iteratively deriving infinitely many times. While this can be done with a slightly weaker hypothesis than uniform recurrence, it seems particularly tricky to fully adapt Section 5.4 in the non-recurrent case, leading to the question above.

We also give another question related to decidability and recurrence.

Question 5.9. Is uniform recurrence decidable for morphic languages?

Uniform recurrence is decidable for morphic words [Dur13b] and for purely morphic languages [BPR21]. However, a morphic language is not necessarily the language of a morphic word (see Example 5.60) therefore it is not known, at least to our knowledge, if uniform recurrence is decidable for morphic languages. A positive answer to this question would lead to the following slightly different statement of Theorem 5.66: if $\mathcal{L}$ is a morphic language, one can decide if $\mathcal{L}$ is recurrent (eventually) dendric.

Bibliography

[AC16] Jorge Almeida and Alfredo Costa. A geometric interpretation of the Schützenberger group of a minimal subshift. *Ark. Mat.*, 54(2):243–275, 2016.

[AP07] Petr Ambrož and Edita Pelantová. A note on 3iet preserving morphisms, 2007. arXiv:math/0703792.

[AR91] Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité 2n+1. *Bull. Soc. Math. Fr.*, 119(2):199–215, 1991.

[AS03] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[BCB19] Valérie Berthé and Paulina Cecchi Bernales. Balancedness and coboundaries in symbolic systems. *Theoret. Comput. Sci.*, 777:93–110, 2019.

[BCBD+21] V. Berthé, P. Cecchi Bernales, F. Durand, J. Leroy, D. Perrin, and S. Petite. On the dimension group of unimodular $\mathcal{S}$-adic subshifts. *Monatsh. Math.*, 194(4):687–717, 2021.

[BDD+18] Valérie Berthé, Francesco Dolce, Fabien Durand, Julien Leroy, and Dominique Perrin. Rigidity and substitutive dendric words. *Internat. J. Found. Comput. Sci.*, 29(5):705–720, 2018.

[BDFD+15a] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176(4):521–550, 2015.

[BDFD$^+$15b] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and interval exchanges. *J. Pure Appl. Algebra*, 219(7):2781–2798, 2015.

[BDFD$^+$15c] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. *J. Pure Appl. Algebra*, 219(7):2521–2537, 2015.

[BDFD$^+$15d] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Math.*, 338(5):725–742, 2015.

[BDFP$^+$12] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.

[Ber71] Jean Bernoulli. *Recueil pour les astronomes*, volume 1. Desaint, Libraire rue du Foin, Paris, 1771.

[BHL23] Nicolas Bédaride, Arnaud Hilion, and Martin Lustig. Measure transfer and S-adic developments for subshifts, 2023. arXiv:2211.11235.

[BPR10] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and automata*, volume 129 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2010.

[BPR21] Marie-Pierre Béal, Dominique Perrin, and Antonio Restivo. Decidable problems in subtitution shifts, 2021. arXiv:2112.14499.

[BPR23] Marie-Pierre Béal, Dominique Perrin, and Antonio Restivo. Recognizability of morphisms. *Ergodic Theory Dyn. Syst.*, page 1–25, 2023.

[BPRS23] Marie-Pierre Béal, Dominique Perrin, Antonio Restivo, and Wolfgang Steiner. Recognizability in S-adic shifts, 2023. arXiv:2302.06258.

[BPS08] L'ubomíra Balková, Edita Pelantová, and Wolfgang Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4):251–263, 2008.

[Cas96] Julien 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] Julien Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc.*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994).

[CLL22] Julien Cassaigne, Sébastien Labbé, and Julien Leroy. Almost everywhere balanced sequences of complexity $2n+1$. *Mosc. J. Comb. Number Theory*, 11(4):287–333, 2022.

[CN10] Julien Cassaigne and François Nicolas. Factor complexity. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 163–247. Cambridge University Press, Cambridge, 2010.

[DDMP21] Sebastián Donoso, Fabien Durand, Alejandro Maass, and Samuel Petite. Interplay between finite topological rank minimal Cantor systems, $\mathcal{S}$-adic subshifts and their complexity. *Trans. Amer. Math. Soc.*, 374(5):3453–3489, 2021.

[DDP23] Francesco Dolce, L'ubomíra Dvořáková, and Edita Pelantová. On balanced sequences and their critical exponent. *Theoret. Comput. Sci.*, 939:18–47, 2023.

[DF22] Michael Damron and Jon Fickenscher. The number of ergodic measures for transitive subshifts under the regular bispecial condition. *Ergodic Theory Dyn. Syst.*, 42(1):86–140, 2022.

[DG19] Fabien Durand and Valérie Goyheneche. Decidability, arithmetic subsequences and eigenvalues of morphic subshifts. *Bull. Belg. Math. Soc. Simon Stevin*, 26(4):591–618, 2019.

[DJP01] Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.

[DL22] Fabien Durand and Julien Leroy. Decidability of the isomorphism and the factorization between minimal substitution subshifts. *Discrete Anal.*, 2022.

[Dow05] Tomasz Downarowicz. Survey of odometers and Toeplitz flows. In *Algebraic and topological dynamics. Proceedings of the conference, Bonn, Germany, May 1–July 31, 2004*, pages 7–37. Providence, RI: American Mathematical Society (AMS), 2005.

[DP17] Francesco Dolce and Dominique Perrin. Neutral and tree sets of arbitrary characteristic. *Theor. Comput. Sci.*, 658:159–174, 2017.

[DP21] Francesco Dolce and Dominique Perrin. Eventually dendric shift spaces. *Ergod. Theory Dyn. Syst.*, 41(7):2023–2048, 2021.

[Dur98] Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998.

[Dur00] Fabien Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory Dynam. Systems*, 20(4):1061–1078, 2000.

[Dur03] Fabien Durand. Corrigendum and addendum to: "Linearly recurrent subshifts have a finite number of non-periodic subshift factors" [*Ergodic Theory Dynam. Systems* 20(4):1061–1078, 2000]. *Ergodic Theory Dynam. Systems*, 23(2):663–669, 2003.

[Dur10] Fabien Durand. Combinatorics on Bratteli diagrams and dynamical systems. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 324–372. Cambridge Univ. Press, Cambridge, 2010.

[Dur13a] Fabien Durand. Decidability of the HD0L ultimate periodicity problem. *RAIRO Theor. Inform. Appl.*, 47(2):201–214, 2013.

[Dur13b] Fabien Durand. Decidability of uniform recurrence of morphic sequences. *Internat. J. Found. Comput. Sci.*, 24(1):123–146, 2013.

[Esp23] Bastián Espinoza. The structure of low complexity subshifts, 2023. arXiv:2305.03096.

[Fer96] Sébastien Ferenczi. Rank and symbolic complexity. *Ergodic Theory Dynam. Systems*, 16(4):663–682, 1996.

[Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002.

[FZ08] Sébastien Ferenczi and Luca Q. Zamboni. Languages of k-interval exchange transformations. *Bull. Lond. Math. Soc.*, 40(4):705–714, 2008.

[Ghe23] France Gheeraert. Some properties of morphic images of (eventually) dendric words. *Monatsh. Math.*, 202(2):335–351, 2023.

[GJWZ09] Amy Glen, Jacques Justin, Steve Widmer, and Luca Q. Zamboni. Palindromic richness. *European J. Combin.*, 30(2):510–531, 2009.

[GL22] France Gheeraert and Julien Leroy. $\mathcal{S}$-adic characterization of minimal dendric shifts, 2022. arXiv:2206.00333.

[GLL22] France Gheeraert, Marie Lejeune, and Julien Leroy. $\mathcal{S}$-adic characterization of minimal ternary dendric shifts. *Ergodic Theory Dynam. Systems*, 42(11):3393–3432, 2022.

[GO22] Herman Goulet-Ouellet. Suffix-connected languages. *Theoret. Comput. Sci.*, 923:126–143, 2022.

[GRS23] France Gheeraert, Giuseppe Romana, and Manon Stipulanti. String attractors of fixed points of k-Bonacci-like morphisms. In *Combinatorics on words*, volume 13899 of *Lecture Notes in Comput. Sci.*, pages 192–205. Springer, Cham, 2023.

[Har63] Frank Harary. A characterization of block-graphs. *Canadian Mathematical Bulletin*, 6(1):1–6, 1963.

[Hon09] Juha Honkala. On the simplification of infinite morphic words. *Theoret. Comput. Sci.*, 410(8-10):997–1000, 2009.

[JP02] Jacques Justin and Giuseppe Pirillo. Episturmian words and episturmian morphisms. *Theor. Comput. Sci.*, 276(1-2):281–313, 2002.

[Kea75] Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975.

[Ler12] Julien Leroy. Contribution to the resolution of the S-adic conjecture, 2012. PhD Thesis.

[LM95] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

[Lot97] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997. Corrected reprint of the 1983 original, with a new preface by Perrin.

[Lot02] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encycl. Math. Appl.* Cambridge University Press, Cambridge, 2002.

[LS01] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1977 edition.

[MH38] Marston Morse and Gustav A. Hedlund. Symbolic dynamics. *Am. J. Math.*, 60:815–866, 1938.

[MH40] Marston Morse and Gustav A. Hedlund. Symbolic dynamics. II: Sturmian trajectories. *Am. J. Math.*, 62:1–42, 1940.

[Pan84] Jean-Jacques Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. In *Automata, Languages and Programming*, pages 380–389. Springer Berlin Heidelberg, 1984.

[Rau79] Gérard Rauzy. Échanges d'intervalles et transformations induites. *Acta Arith.*, 34(4):315–328, 1979.

[Thu06] A. Thue. Über unendliche Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.*, 7:1–22, 1906.

[Thu12] A. Thue. Über die gegenseitige lage gleicher teile gewisser zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.*, 1:1–67, 1912.

[Vee78] William A. Veech. Interval exchange transformations. *J. Analyse Math.*, 33:222–272, 1978.

[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.

Notations

Index