Iterated morphisms, combinatorics on words and symbolic dynamical systems

Revekka Kyriakoglou

Université Paris-Est

École Doctorale MSTIC

Mathématiques et Sciences et Techniques de l'Information et de la Communication

Thèse de doctorat

Spécialité Informatique

Présentée par:

Revekka Kyriakoglou

Morphismes itérés, combinatoire des mots et systèmes dynamiques symboliques

Iterated morphisms, combinatorics on words and symbolic dynamical systems

Thèse dirigée par Dominique Perrin

Soutenue le 30 octobre 2019

Devant le jury composé de

Valérie Berthé	Examinatrice
Anna Pappa	Examinatrice
Dominique Perrin	Directeur de thèse
George Rachonis	Rapporteur
Giuseppina Rindone	Examinatrice
Patrice Séébold	Rapporteur

Laboratoire d'informatique Gaspard-Monge

UMR 8049 LIGM

Résumé

La thèse actuelle porte sur la combinatoire des mots et les systèmes dynamiques symboliques. Les systèmes dynamiques symboliques sont des objets permettant de coder les trajectoires de mots dans des systèmes dynamiques de transformations d'espaces topologiques. Parmi ces systèmes dynamiques, des exemples bien connus sont donnés par les mots Sturmiens et par les échanges d'intervalles. Les mots Sturmiens sont liés à des algorithmes de géométrie discrète et les échanges d'intervalles forment une classe intéressante de systèmes dynamiques. En outre, nous remarquons que certaines familles d'échanges d'intervalles fournissent des généralisations prometteuses de mots Sturmiens.

Le sujet principal de cette thèse est la reconnaissabilité des mots engendrés par des morphismes primitifs. Le concept de reconnaissabilité des morphismes trouve son origine dans l'article de C. Martin [77] sous le terme de détermination. B. Host a utilisé ce terme pour la première fois dans son article sur la théorie ergodique des systèmes dynamiques [58]. La notion de reconnaissabilité est apparue après l'intérêt manifesté par de nombreux scientifiques pour ses diverses applications théoriques dans divers domaines, de la combinatoire des mots à la dynamique symbolique. Une notion similaire est celle de la circularité. Les deux termes sont souvent, mais pas toujours, utilisés comme synonymes. Ce manque de cohérence dans la littérature est parfois source de confusion. À la connaissance de l'auteur, il n'y a pas encore d'étude qui rassemble ces définitions et

prouve leur équivalence ou indique les différences qui existent entre elles. Dans cette thèse nous étudions ces différentes notions, en utilisant une définition cohérente de la reconnaissabilité et de la circularité.

Les ensembles de mots peuvent être définis à partir de propriétés sur les graphes d'extension de ses éléments, tels que les ensembles acycliques, les ensembles dendriques, les ensembles neutres, etc. Plus précisément, pour un ensemble de mots S, on peut associer à chaque mot w dans S son graphe d'extension qui décrit les extensions gauches et droites possibles de w dans S. Nous montrons comment obtenir la décidabilité de certaines propriétés (acyclique, dendrique, etc.) d'un langage en utilisant les notions de reconnaissabilité, de graphes d'extension des mots de ce langage et d'une technique présentée dans [65].

La notion de reconnaissabilité est egalement utilisée dans le domaine des semigroupes profinis. Nous décrivons la relation entre la reconnaissabilité des morphismes et les propriétés des semigroupes profinis libres [4].

Mots clés— Informatique théorique, combinatoire des mots, systèmes symboliques dymaniques, reconnaissabilité, circularité, ensembles dendrique, graphe d'extension, langage substitutive, groupe libre,semi-groups profinis

Abstract

The current thesis focuses on the topic of combinatorics on words and symbolic dynamical systems. The symbolic dynamical systems are objects for encoding word trajectories in dynamic systems of transformations in topological spaces. Among these dynamical systems, well-known examples are given by Sturmian words and by interval exchange. The Sturmian words are related to discrete geometry algorithms and the interval exchange forms an interesting class of dynamical systems. Furthermore, it should be mentioned that some exchange families provide promising generalizations of Sturmian words.

The main subject of the thesis is the recognizability of words generated by primitive morphisms. The concept of recognizability of morphisms originates in the paper of C. Martin [77] under the term of determinization. The term was first used by B. Host in his paper on the Ergodic theory of Dynamical Systems [58]. The notion of recognizability came in full bloom after the interest shown by many scientists due to its various theoretical applications in various topics, from combinatorics on words to symbolic dynamics. A similar notion is that of circularity. The two terms are often, but not always used as synonymous. This lack of consistency along the literature could result in confusion. To the best of the author's knowledge, there is not, as of yet, any study that collects those definitions and proves their equivalence or indicates the differences among them. In the current thesis, we study those different notions, using a coherent

definition of recognizability.

Sets of words can be defined from properties of the extension graphs of their elements, such as acyclic sets, dendric sets, neutral sets, etc. More precisely, given a set of words S, one can associate with every word w in S its extension graph. That graph describes the possible left and right extensions of w in the set S. We show how to obtain the decidability of certain properties using the notion of recognizability, the extension graphs of the words of a language and a technique presented in [65].

Furthermore, the notion of recognizability is used in the subject of Profinite Semigroups. We describe the relationship between the recognizability of morphisms and certain properties of the free profinite semigroups [4].

Keywords— Theoretical Computer Science, Combinatorics on Words, Symbolic Dynamical Systems, Recognizability, Circularity, Dendric Sets, Extension Graphs, Substitutive Language, Free Group, Profinite Semigroups.

Contents

1 Introduction

1.1 The notion of recognizability

The current thesis focuses on the topic of combinatorics on words and symbolic dynamical systems. In particular we study symbolic dynamical systems, and we apply the "natural coding" to trajectories in order to obtain infinite words and to connect the two theories. Among these dynamical systems, well-known examples are given by Sturmian words and by interval exchange. Sturmian words are related to discrete geometry algorithms and interval exchange forms an interesting class of dynamical systems. Promising generalizations of the Sturmian words can be provided by some exchange families.

The main subject of the thesis is the recognizability of words generated by primitive morphisms. Let φ be a morphism on a finite alphabet A. The ability to decompose in a unique way a long enough word in the language of the morphism and to recognize the preimage of that word is what described by the term *recognizability*. The concept of recognizability of morphisms originates from the paper of John C. Martin [77] under the term of determinization. The term was first used by B. Host in his paper "Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable" [58]. The notion of recognizability came in full bloom after the interest shown by many scientists due to its various theoretical applications in diverse topics, from combinatorics on words to symbolic dynamics. A similar notion that of *circularity* that appears in the work of many scientists as by Julien Cassaigne in [24] and by

Filippo Mignosi and Patrice Séébold in [79]. The two terms are often, but not always used as synonymous. John C. Martin worked on recognizability on binary alphabets in his work "Minimal flows arising from substitutions of non-constant length" [77] in 1973. Later, in 1992 Brigitte Mossé proved that recognizability holds in the case of primitive non-erasing morphisms in his work "Puissances de mots et reconnaissabilité des points fixes d'une substitution" [80]. That result played a pivotal role in the field given that prior to this recognizability was assumed.

1.2 Main results

Recognizability and circularity are two terms with various and often overlapping definitions. This inconsistency along the literature can act as a root of confusion within the field. The lack of clear distinction between the two terms, calls for a novel revision of the definitions that will allow the disentanglement of the respective notions and their potential misuse. To the best of the author's knowledge, there is not, as of yet, any study that collects those definitions and proves their equivalence or indicates the differences among them. In this thesis it has been studied in great detail the equivalence between those different notions. This clarifies the relation between the possible variance of the definition which have been introduced by many authors.

We also show that the notion of recognizability alongside with a technique introduced in [65] by Karel Klouda and Štěpán Starosta can be used in order to prove the decidability of different properties of a language.

In order to understand the combinatorial properties of a language $\mathcal{L}(\mathbf{u})$ of an infinite fixed point $\mathbf{u}$ generated by the morphism φ over some finite alphabet A, it is usually convenient to focus on the set of its bispecial factors, i.e., finite words that occur in $\mathbf{u}$ and that can be extended in $\mathbf{u}$ to the left and to the right by several letters. For instance, knowing the set of bispecial factors, it is possible to compute the factor complexity of $\mathbf{u}$, that is the number of distinct factors of a given length (see, for instance, [25]). Bispecial factors also allow for characterization of several families of words like Sturmian words, and more generally, (eventually) dendric words ([19, 33, 34]). In order to prove the decidability of such properties we used a fam-

ily of graphs known as extension graphs. More precisely, given a set of words S, one can associate with every word $u \in S$ a graph called *extension graph* that carries the information of which are the possible left and right extensions of the word u in S. The reason why the extension graphs are worth studying is that they allow the characterization of a set as acyclic, dendric set, neutral set, etc.

Normally, in order to decide the above properties we should construct the extension graph for all of the words of the set S and check if the properties hold for all of these graphs. However, since we may work on an infinite set S, such as substitutive languages, the above method cannot be applied. In order to overcome the previous physical constrain of constructing infinite many graphs, we make use of the notion of strong two-sided recognizability.

The process of describing bispecial factors is considered by Anna Frid and Sergey V. Avgustinovich in [48] for a particular class of recognizable morphisms. It is also studied by K. Klouda in "Bispecial factors in circular non-pushy D0L languages" [63]. In his work Karel Klouda provide an algorithm for finding all bispecial words arising from a specific family of morphisms, using pairs of bispecial extensions (i.e. pairs $(a, b) \in A \times A$ such that aub is a factor of $\mathbf{u}$). Inspired by this algorithm and using the notion of the extension graphs and recognizability, we designed an algorithm that provide us all the possible extension graphs of the bispecial words in the language of a morphism. The above mentioned algorithm is the result of a collaborative work amongst the present author, Francesco Dolce and Julien Leroy in [35].

Lastly, in the chapter Profinite Semigroups another application of the notion of recognizability is presented. In this chapter, the link between the recognizability of morphisms and some properties of the free profinite semigroups is described thoroughly. This chapter is based on a survey on the profinite semigroups written by the current author and Dominique Perrin [69] while most of the results that are presented are the work of Jorge Almeida and Alfredo Costa in [7] and [8]. The above mentioned authors along with the present author are currently preparing a book on the topic of the profinite semigroups.

Acknowledgement

I wish to express my sincere appreciation to my supervisor, Dominique Perrin, who guided and encouraged me during my doctoral studies. Without his help and support, the goal of this project would not have been realized.

I would also like to thank George Rachonis and Patrice Séébold who accepted to report my thesis, as well as Valérie Berthé, Anna Pappa and Giuseppina Rindone for accepting to be members of the thesis committee. Their feedback was invaluable.

I thank also my colleagues and friends for the stimulating discussions, Francesco Dolce, Pavel Heller, Andrew Ryzhinov, Pablo Rotondo, as well as the fellow doctoral students of LIGM.

Furthermore, I would like to express my gratitude to several members of the LIGM that helped me and advised me, including Marie-Pierre Béal, Claire David, Olivier Champalle, Tita Kyriakopoulou, Patrice Herault, Corinne Palescendo. A special thanks also to the secretary of the doctoral school Sylvie Cach who was always available.

I wish to acknowledge the support and great love of my family, my father, Dimitrios, my mother, Chrysoula, and my sister, Ifigeneia. Last but not least, I would like to thank my partner Christos-Nikolaos for his patience and support. He kept me going on and this work would not have been possible without his input.

2 Preliminaries

2.1 Words

Let A be a set, called *alphabet*, the elements of which are called *letters*. In the following chapters the alphabet is supposed to be finite. A finite sequence $a_0a_1 \dots a_n$, of elements $(a_i)_{0 \leq i \leq n}$ of A is called a *finite word* on A or just a word on A.

For two words $u = u_0u_1 \dots u_n$ and $y = y_0y_1 \dots y_m$ on the alphabet A, the *concatenation product*, or simply product, is defined as follows,

$$uy = u_0u_1 \dots u_ny_0y_1 \dots y_m.$$

The empty sequence is called the *empty word* and is denoted by ε. The set of words on A is denoted by A^* and the set of all the non-empty words on A is denoted by A^+. The set A^* is the free monoid on the set A. If a word w on the alphabet A can be written as a concatenation product of two words v and z on A (i.e. $w = vz$), then v is a *prefix* of w and z is a *suffix* of w.

Consider the word $u = u_0u_1 \dots u_n$ on the alphabet A. For all integers i, j such that $0 \leq i \leq j \leq n$, the word $u_iu_{i+1} \dots u_j$ is called a *factor* of u and is denoted $u_{[i,j]}$ (or $u_{[i,j+1)}$, with $u_{[i,i)} = \varepsilon$). If a factor of u is not u itself, it is called a *proper* factor of u. The *length* of u, denoted as $|u|$, is the number of letters in u. For example, the word *abcda* has length 5. The length of the empty word ε is zero.

A *right infinite* word on the alphabet A is an infinite sequence $x = x_0x_1 \dots x_n \dots$ of letters on A. The set of right infinite words on A is denoted by $A^{\mathbb{N}}$. For a word u in A^*, we denote by u^ω the

infinite sequence formed by u, i.e., $u^\omega = uuu\ldots$. The product of a finite word on A with a right infinite word on A is a right infinite word on A, as follows,

$$\begin{aligned} A^* \times A^{\mathbb{N}} &\to A^{\mathbb{N}} \\ (u, v) &\mapsto x = uv \end{aligned}$$

with $u = u_0u_1\ldots u_n$ a word in A^*, $v = v_0v_1\ldots v_m\ldots$ a word in $A^{\mathbb{N}}$ and $uv = u_0\ldots u_nv_0\ldots v_m\ldots$ in $A^{\mathbb{N}}$. For right infinite words, the notions of prefix and factor are defined similarly to the finite case. More precisely, for $x = x_0x_1\ldots$ in $A^{\mathbb{N}}$ the word $x_ix_{i+1}\ldots x_j$, with i, j integers with $0 \leq i \leq j$, is called a *factor* of x and is denoted by $x_{[i,j]}$ (or $x_{[i,j+1)}$ with $x_{[i,i)} = \varepsilon$). Any finite factor $x_0x_1\ldots x_j$ of x on A is called *prefix* of x. For any prefix $v = x_0x_1\ldots x_j$ of x there is right infinite word u on A such that $vu = x$. The language of an infinite word x on A denoted by $\mathcal{L}(x)$ is the set of all its factors.

A *left infinite* word on the alphabet A is an infinite sequence $y = \ldots y_{-n}\ldots y_{-1}$ of letters on A. The set of left infinite words on A is denoted by $A^{-\mathbb{N}}$. The product of a left infinite word on A with a finite word on A is a left infinite word on A, as follows,

$$\begin{aligned} A^{-\mathbb{N}} \times A^* &\to A^{-\mathbb{N}} \\ (u, v) &\mapsto y = uv \end{aligned}$$

with the word $u = \ldots u_{-m}\ldots u_{-1}$ in $A^{-\mathbb{N}}$, the word $v = v_0v_1\ldots v_n$ in A^* and $uv = \ldots u_{-m}\ldots u_{-1}v_0v_1\ldots v_n$ being a left infinite word. For left infinite words, the notions of suffix and factor are defined in a similar way as in the case of finite words. More precisely, let $y = \ldots y_{-n}\ldots y_{-1}$ be a left infinite word. The word $y_{-j}y_{-j+1}\ldots y_{-i-1}y_{-i}$ in A^*, with $0 \leq i \leq j$ is called a *factor* of y and is denoted by $y_{[-j,-i]}$ (or $y_{[-j,-i+1)}$ with $y_{[-i,-i)} = \varepsilon$). Any finite factor $v = y_{-n}\ldots y_{-1}$ is called a *suffix* of y. For any suffix v there is a left infinite word u such that $uv = y$. The *language* of a left infinite word y on A, denoted by $\mathcal{L}(y)$ is the set of all its factors.

A *bi-infinite* word or *two-sided infinite* word on an alphabet A is an infinite sequence $\ldots a_{-n}\ldots a_{-1}a_0a_1\ldots a_n\ldots$. The set of bi-infinite words on A is denoted by $A^{\mathbb{Z}}$. The product of a left infinite word on A with a right infinite word on A is a bi-infinite word on A, as in the following bijection,

$$\begin{aligned} A^{-\mathbb{N}} \times A^{\mathbb{N}} &\to A^{\mathbb{Z}} \\ (y, x) &\mapsto z = y \bullet x \end{aligned}$$

with $y = \dots y_{-n} \dots y_{-1}$ in $A^{-\mathbb{N}}$, $x = x_0 x_1 \dots x_m \dots$ in $A^{\mathbb{N}}$ and

$$z = y \bullet x = \dots y_{-n} \dots y_{-1} \bullet x_0 x_1 \dots x_m \dots$$

being a bi-infinite word, where $\bullet$ is the position zero. The bijection follows from the fact the position zero is fixed for a bi-infinite word. For $z = y \bullet x$, the word x is in $A^{\mathbb{N}}$ and is denoted z^+, and the word y is in $A^{-\mathbb{N}}$ and is denoted z^-. Let $z = \dots z_{-1} \bullet z_0 z_1 \dots$, the word $z_i \dots z_j$, with i, j integers with $i \leq j$, is called *factor* of z and is denoted by $z_{[i,j]}$ (or $z_{[i,j+1)}$ with $x_{[i,i)} = \varepsilon$). The language of a bi-infinite word z on A, denoted by $\mathcal{L}(z)$, is the set of all its factors.

From now on, we will often use the term infinite words instead of right infinite words, if it is clear from the context.

Let S be a set of words on the alphabet A. For $w \in S$, we denote

$$\begin{aligned} L_S(w) &= \{a \in A \mid aw \in S\} \\ R_S(w) &= \{a \in A \mid wa \in S\} \\ B_S(w) &= \{(a,b) \in A \times A \mid awb \in S\} \end{aligned}$$

and further

$$\ell_S(w) = \mathrm{Card}(L_S(w)),\ r_S(w) = \mathrm{Card}(R_S(w)),\ e_S(w) = \mathrm{Card}(B_S(w)).$$

We omit the subscript S when it is clear from the context. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$.

A set F of words on the alphabet A is said to be *factorial* if it contains all the factors of its elements. A factorial set S is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in S is right-extendable (resp. left-extendable, resp. biextendable).

Remark 2.1.1. If $x \in A^{\mathbb{N}}$ then the language $\mathcal{L}(x)$ is factorial and right-extendable.

Remark 2.1.2. If $y \in A^{-\mathbb{N}}$ then the language $\mathcal{L}(y)$ is factorial and left-extendable.

Remark 2.1.3. If $z \in A^{\mathbb{Z}}$ then the language $\mathcal{L}(z)$ is factorial and biextendable.

A word w is called *right-special* if $r(w) \geq 2$, *left-special* if $\ell(w) \geq 2$ and *bispecial* if it is both left-special and right-special.

For $w \in S$, we denote

$$m_S(w) = e_S(w) - \ell_S(w) - r_S(w) + 1.$$

We omit the subscript S when it is clear from the context. The word w is called *weak* if $m(w) < 0$, *neutral* if $m(w) = 0$ and *strong* if $m(w) > 0$. We say that a factorial set S is *neutral* if every word in S is neutral. Also, we say that a factorial set S is weak (resp. strong) if every word in S is neutral or weak (resp. strong or weak).

An integer $p > 0$ is a *period* of a finite word $u = a_0a_1 \dots a_n$ if for all k such that $k + p \leq n$, we have $a_k = a_{k+p}$. The smallest period of u is called *the period* of u.

Example 2.1.4. The word *abcaabcaab* has period $p = 4$, since $abcaabcaab = (abca)^2ab$.

The notion of period extends without difficulty to infinite words. An infinite word is called *periodic* if the set of periods is not empty. More precisely, a right infinite word u (resp. left infinite word y) is *periodic* if there is integer $p > 0$ such that for all $k, n \in \mathbb{N}$ with $0 \leq n < p$ holds that $u_{n+kp} = u_n$ (resp. $y_{-(n+kp)} = y_{-n}$). In an analogous way, a bi-infinite word z is *periodic* if there is p such that for all $k, n \in \mathbb{Z}$ with $0 \leq n < p$ holds that $z_{n+kp} = z_n$. The smallest period is called the period of u.

An integer $p > 0$ is an *ultimate period* of a right infinite word $u = a_0a_1 \dots$ if there is an integer $k_0 \geq 0$ such that for all $k \geq k_0$ we have $a_k = a_{k+p}$. The smallest ultimate period of an infinite word u is called the ultimate period of u. An infinite word which admits an ultimate period is called *ultimately periodic*. Similarly it can defined the ultimate periodicity for a left infinite word.

Example 2.1.5. The right infinite word $x = abcbcbcbcbc\cdots = a(bc)^\omega$ is ultimately periodic with $k_0 = 1$ and ultimate period $p = 2$, since it can be written as $a \cdot bc \cdot bc \cdot bc \cdot bc \dots$.

An infinite word u is *aperiodic* if it is not ultimately periodic.

Given a set of words S, subset of A^*, and a word $u \in A^*$, we let Su^{-1} and $u^{-1}S$ denote the sets

$$\begin{aligned} Su^{-1} &= \{v \in A^* \mid vu \in S\}, \\ u^{-1}S &= \{v \in A^* \mid uv \in S\}. \end{aligned}$$

A factorial set F is called *recurrent* if for all $u, v \in F$, there is word $w \in F$ such that $uwv \in F$. A right extendable factorial set is said to be *uniformly recurrent* if for all words $u \in F$, there is integer $n \geq 1$ such that the word u is factor of any word in F of length n.

Proposition 2.1.6. [17] A uniformly recurrent set is recurrent.

Proof. Let F be a uniformly recurrent set and w, u be words in F. Since F is uniformly recurrent there are integers n, m such that the word w is factor of any word in F of length n and the word u is factor of words with length m. Since the set is right-extendable, there exists a word v such that uv in F, with v longer than n. Since v is longer than n, it follows that w is factor of v, i.e., there are words r, s in F such that $v = rws$. Thus, $uv = urws$ is in F and since the set F is factorial, the factor urw is also in F. □

Given an one-sided infinite word $x \in A^{\mathbb{N}}$, an *occurrence* of $u \in \mathcal{L}(x)$ in x is an integer $i \geq 0$ such that $x_i \dots x_{i+|u|-1} = u$.

Definition 2.1.7. A right infinite word $x \in A^{\mathbb{N}}$ is *recurrent* if any word $u \in \mathcal{L}(x)$ has a second occurrence in x, i.e., if i is an occurrence of u in x, then there is another occurrence j, with $j > i$.

Observe that if every word u has a second occurrence in the right infinite recurrent word x, then it has an infinity of occurrences.

Proposition 2.1.8. A set F is recurrent and right-extendable if and only if there is a right infinite word $x \in A^{\mathbb{N}}$ such that $F = \mathcal{L}(x)$.

Proof. Let $x \in A^{\mathbb{N}}$ be a recurrent right infinite word and let us set $F = \mathcal{L}(x)$. It follows that F is a factorial right-extendable set from the definition of $\mathcal{L}(x)$ as the set of all possible factors of the right infinite word x. It remains to be proved that F is recurrent. Let $u, v \in F$. Let $i \in \mathbb{N}$ be an occurrence of u. Since there is an infinite number of occurrences of v, there is an occurrence j of v that is such $j > i + |v|$. Thus, $x_i \dots x_{j+|v|-1} = uwv$ and therefore F is recurrent.

Let us now prove the other direction. Let F be recurrent and right-extendable. The set F can be written with decreasing order with respect to the length of its elements as follows,

$$F = \{w_1, w_2, \dots\}, \text{ with } |w_1| \leq |w_2| \leq \dots$$

Since F is recurrent it holds that for all $i \in \mathbb{N}$, there is a word u_i such that $w_i u_i w_{i+1} \in F$. Let us now define the following right

infinite word,

$$x = w_1u_1w_2u_2w_3u_3 \cdots \in A^{\mathbb{N}}.$$

By construction of x it follows that $F \subset \mathcal{L}(x)$. Any word $u = w_iu_iw_{i+1}$ is a factor of x and is also a word in the set F. Since F is a factorial set, any factor of u is also in F, which proves that for all $u \in \mathcal{L}(x)$, $\mathcal{L}(u) \subset F$ and therefore $\mathcal{L}(x) \subset F$. Thus, $F = \mathcal{L}(x)$. It remains to prove that the right infinite word x is recurrent. Let $u \in \mathcal{L}(x)$. Because of the definition x it follows that $u \in F$ and since F is recurrent, there is some w such that $uwu \in F$ and therefore uwu is a factor x which means that $uwu \in \mathcal{L}(x)$. Thus x is recurrent. □

2.2 Morphisms

A *substitution* φ is a mapping from a set of letters A to a set of words B^*. A substitution can be extended to a monoid morphism from A^* to B^*, where $\varphi(\varepsilon) = \varepsilon$ and $\varphi(uv) = \varphi(u)\varphi(v)$. Then, any morphism φ is uniquely determined by the set of its images $\varphi(a)$ for all letters a in A. Many times we use the word substitution when we talk about monoid morphisms. A morphism φ is called *non-erasing* if $\varphi(a) \neq \varepsilon$ for all letters a in the alphabet A. From now on we consider all morphisms being non-erasing.

Let φ be the mapping $\varphi : A^* \to A^*$. The *matrix associated* to φ is the integer $A \times A$-matrix such that $M_{a,b}$ is the number of occurrences of b in $\varphi(a)$, that is

$$M_{a,b} = |\varphi(a)|_b.$$

It can be verified that $M(\varphi^n) = M(\varphi)^n$ for every integer $n \geq 1$.

Example 2.2.1. Let φ be the morphism on the alphabet $A = \{0, 1\}$ such that $\varphi(0) = 01$ and $\varphi(1) = 0$. The morphism φ has associated matrix

$$M(\varphi) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

We set $||\varphi|| = \max_{a \in A} |\varphi(a)|$ and $|\varphi| = \min_{a \in A} |\varphi(a)|$.

Any $\varphi : A^+ \to B^+$ can be extended to $\varphi : A^{\mathbb{N}} \to B^{\mathbb{N}}$ as follows,

$$\begin{array}{rcl} \varphi : A^{\mathbb{N}} & \to & B^{\mathbb{N}} \\ x & \mapsto & y = \varphi(x) \end{array}$$

where $x = x_0x_1\ldots$ an infinite word and $\varphi(x) = \varphi(x_0)\varphi(x_1)\ldots$ being also an infinite word. This extension is always well defined since the morphism is supposed to be non-erasing. In the case that the morphism is erasing, the extension of the morphism φ is well defined if $|\varphi(x_0\ldots x_n)| \xrightarrow[n\to\infty]{} +\infty$. Let $y = \varphi(x)$, then for all positive integers i there is a minimal integer j such that $y_i = \varphi(x_j)_{i-k}$, $k = |\varphi(x_0\ldots x_{j-1})|$ and $\ell = |\varphi(x_0\ldots x_j)|$ with,

$$k < i \leq \ell, \text{ and } \ell = k + |\varphi(x_j)|.$$

In other words, the letter y_i of y is the letter with index $i-k$ in the image of the letter x_j of x under the morphism φ. The existence of the integer j follows from the fact that the morphism φ is non-erasing. The extension to the infinite words corresponds to the relation $\varphi(uv) = \varphi(u)\varphi(v)$, for u in A^* and v in $A^{\mathbb{N}}$.

Similarly, any non-erasing morphism $\varphi : A^* \to B^*$ can be extended to $\varphi : A^{-\mathbb{N}} \to A^{-\mathbb{N}}$ as follows,

$$\begin{array}{rcl} \varphi : A^{-\mathbb{N}} & \to & B^{-\mathbb{N}} \\ x & \mapsto & y = \varphi(x) \end{array}$$

where $x = \ldots x_{-1}$ a left infinite word and $\varphi(x) = \ldots\varphi(x_{-1})$ being also a left infinite word. (Same remark as above for the morphisms that are not non-erasing.) Let $y = \varphi(x)$, then for all positive integers i there are integers k, j such that $y_{-i} = \varphi(x_{-j})_{-i+k}$, j minimal such that, for $k = |\varphi(x_{-j-1}\ldots x_{-1})|$ and $\ell = |\varphi(x_{-j}\ldots x_{-1})|$ with,

$$k < i \leq \ell, \text{ and } \ell = k + |\varphi(x_{-j})|.$$

In other words, the letter y_{-i} of y is the letter with index $-i+k$ in the image of the letter x_{-j} of x under the morphism φ. The extension of the morphism on the left infinite words corresponds to the relation $\varphi(uv) = \varphi(u)\varphi(v)$, for u in $A^{-\mathbb{N}}$ and v in A^*.

Using the previous two extensions of a morphism φ we have the following extension to the set of bi-infinite words,

$$\begin{array}{rcl} \varphi : A^{\mathbb{Z}} & \to & B^{\mathbb{Z}} \\ z & \mapsto & \varphi(z) = \varphi(z^-) \bullet \varphi(z^+) \end{array}$$

where z^- is in $A^{-\mathbb{N}}$ and z^+ in $A^{\mathbb{N}}$.

If a letter a in A has image under the morphism φ that begins with the letter a (i.e $\varphi(a)$ in aA^*) and if $|\varphi^n(a)|$ tends to infinity

with n, there is a unique right infinite word x that has all words $\varphi^n(a)$ as prefixes. Indeed, since x is a right infinite word such as $\varphi^n(a)$ is a prefix of this word for all integers n, then, $\varphi^{n+1}(a)$ (that is the image of $\varphi^n(a)$) is also a prefix of x. That word x, denoted as $\varphi^\omega(a)$, is a *right infinite fixed point* of the morphism φ since it satisfies that $\varphi(x) = x$. When it is clear from the context it may be called a *fixed point*.

Let $b \in A$ such that its image under the morphism φ ends with the letter b (i.e $\varphi(b)$ in A^*b). Then, there is a unique left infinite word y that has all words $\varphi^n(b)$ as suffixes. That word y, denoted as $\varphi^{\tilde{\omega}}(b)$, is called a *left infinite fixed point* of the morphism φ and satisfies that $\varphi(y) = y$. (The proof is the same as in the right infinite case.)

Let $x = \varphi^\omega(a)$ be a right infinite fixed point and $y = \varphi^{\tilde{\omega}}(b)$ a left infinite fixed point for the morphism φ defined above. The bi-infinite word $z = y \bullet x$ has all words $\varphi^n(a)$ and all words $\varphi^n(b)$ as factors, for all n. That word z, denoted as $\varphi^{\tilde{\omega}}(b) \bullet \varphi^\omega(a)$, is called *two-sided infinite fixed point* of the morphism φ and satisfies that $\varphi(z) = z$. Indeed,

$$\varphi(z) = \varphi(y) \bullet \varphi(x) = y \bullet x = z.$$

The *language* of a (right, left or two-sided) infinite fixed point x is the set of all its factors and it is denoted as $\mathcal{L}(x)$. This language is called *substitutive*. A morphism can have more than one infinite fixed point and not all of them need to have the same language.

Example 2.2.2 (Fibonacci word)**.** The Fibonacci morphism on the alphabet $A = \{0, 1\}$ is defined as:

$$\begin{aligned} \varphi_F : A^* &\to A^*, \\ 0 &\mapsto 01, \\ 1 &\mapsto 0. \end{aligned}$$

The Fibonacci word is the right infinite fixed point of φ_F, $\varphi_F^\omega(0) = 01001010010010100 1010\ldots$.

The Fibonacci word can also be defined as the limit of the sequence given by $u^{(0)} = 1$, $u^{(1)} = 0$ and $u^{(n+1)} = u^{(n)}u^{(n-1)}$ for $n \geq 1$.

Example 2.2.3. The morphism $\widetilde{\varphi_F} : \{0, 1\}^* \to \{0, 1\}^*$ with $\widetilde{\varphi_F}(0) = 10$ and $\widetilde{\varphi_F}(1) = 0$ has no right infinite fixed point, but it has left fixed point $\widetilde{\varphi_F}^{\tilde{\omega}}(0)$.

In the following examples there are presented morphisms with more than one fixed point. In Example 2.2.4 there exist two distinct right infinite fixed points with different languages, while in Example 2.2.5 there exist two different right infinite fixed points with the same language.

Example 2.2.4. Let φ be a morphism such that $\varphi(0) = 01$, $\varphi(1) = 0$ and $\varphi(2) = 212$. This morphism has two distinct right infinite fixed points, the Fibonacci word $x = \varphi^\omega(0) = 010010100100101\ldots$ and the word $y = \varphi^\omega(2) = 2120212012120212\ldots$. It can be easily verified that the languages $\mathcal{L}(x)$ and $\mathcal{L}(y)$ are different. For instance, $2 \in \mathcal{L}(y)$ but $2 \notin \mathcal{L}(x)$.

Example 2.2.5. The morphism $\widetilde{\varphi_F}^2 : \{0,1\}^* \to \{0,1\}^*$ with $\widetilde{\varphi_F}^2(0) = 010$ and $\widetilde{\varphi_F}^2(1) = 10$ has two distinct right infinite fixed points $x = (\widetilde{\varphi_F}^2)^\omega(0) = 0101001010010010\ldots$ and $y = (\widetilde{\varphi_F}^2)^\omega(1) = 1001001010010010\ldots$. The two languages $\mathcal{L}(x)$ and $\mathcal{L}(y)$ are the same and it can be justified with use of definition of primitivity, in the next subsection, and Proposition 2.3.7.

2.3 Primitive morphisms

A morphism $\varphi : A^* \to A^*$ is called *primitive* if there is an integer k such that for all letters a, b in A, the letter b appears in the word $\varphi^k(a)$. The smallest such integer k is called the *primitivity index* of φ.

A square non-negative real matrix M is said to be *irreducible* if for all i, j there is integer m such that $M^m_{i,j} > 0$.

A square non-negative real matrix M is said to be *primitive* when, for some positive integer n, all entries in M^n are positive (i.e. all $M^n_{i,j} > 0$). It can be seen that a morphism φ is primitive if and only if $M(\varphi)$ is primitive.

Example 2.3.1. The matrix,

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is irreducible but not primitive.

Example 2.3.2. The Fibonacci morphism φ_F (Example 2.2.2) is primitive since both letters $0, 1$ appear in each of the words $\varphi_F^2(0)$

and $\varphi_F^2(1)$. One has

$$M(\varphi_F) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad M(\varphi_F^2) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Remark 2.3.3. Every primitive morphism is non-erasing.

The following is a classical result known as *Perron-Frobenius Theorem* [87].

Theorem 2.3.4. *Let M be an $n \times n$-matrix with real non-negative coefficients. If M is primitive, then it has a positive eigenvalue ρ such that $|\lambda| < \rho$ for every other eigenvalue λ of M. Moreover, there corresponds to ρ an eigenvector with strictly positive coefficients.*

Proposition 2.3.5. Let φ be a primitive morphism on the alphabet A that is not a singleton. Then, $|\varphi^n(a)| \xrightarrow[n\to\infty]{} \infty$, for all letters a in A.

Proof. Since φ is primitive, there is an integer k such that for all letters a, b in A, the letter b appears in the word $\varphi^k(a)$ and thus, $|\varphi^k(a)| \geq |A|$. Applying this iteratively we obtain $|\varphi^{k^n}(a)| \geq n|A|$. It follows that for any integer c there is integer $N \in \mathbb{N}$ such that $|\varphi^n(a)| > c$, for all $n \geq N$. Hence $|\varphi^n(a)| \xrightarrow[n\to\infty]{} \infty$, for all letters a in A. □

Proposition 2.3.6. Let φ be a primitive morphism. The morphism φ^n is also primitive, for all positive integers n.

Proof. If k is the primitivity index of the morphism φ, then for the morphism φ^n primitivity holds for the integer $k' = \mathrm{lcm}(k, n)$. □

Proposition 2.3.7. For a primitive morphism φ, all the right infinite fixed points have the same language.

Proof. Let x, y words in $A^{\mathbb{N}}$ be fixed points of the primitive morphism φ, such that $x = \varphi^\omega(a)$ and $y = \varphi^\omega(b)$ for $a, b \in A$. Since φ is primitive there is an integer k such that b is a factor of $\varphi^k(a)$, from which follows that $\mathcal{L}(y) \subset \mathcal{L}(x)$. Similarly, since a is a factor of $\varphi^k(b)$ it follows that that $\mathcal{L}(x) \subset \mathcal{L}(y)$, which proves that the two languages are the same. □

Example 2.3.8. The Fibonacci morphism φ_F (Example 2.2.2) is primitive, while the morphism $\varphi : \{0,1\} \to \{0,1\}^*$ with $\varphi(0) = 010$ and $\varphi(1) = 11$ is not primitive since there is no integer k such that $0 \in \varphi^k(1)$.

Proposition 2.3.9. The language of a right infinite fixed point x of a primitive substitution φ is uniformly recurrent.

Proof. Let φ be a primitive morphism with fixed point $x = (x_i)_{i\geq 0}$. Since the morphism is primitive, there is integer k such that for every letter $b, a \in \mathcal{L}(x)$, b is factor of $\varphi^k(a)$.

Let us now consider a word $w \in \mathcal{L}(x)$. There must exist a letter in A (without any lose of generality let the letter be $b \in A$) and an integer $k' \in \mathbb{N}$ such that w is factor of $\varphi^{k'}(b)$. Hence, for the integer $m = k+k'$, the word w is factor of all $\varphi^m(x_i)$ for all $x_i \in \mathcal{L}(x)$, which implies that w is factor of all words with length at least $||\varphi^m||$. □

A primitive morphism with aperiodic right infinite fixed point is called *aperiodic*.

Example 2.3.10. The Fibonacci word (Example 2.2.2) is a right infinite aperiodic fixed point and since the Fibonacci morphism is primitive, the Fibonacci morphism and all the powers of this morphism are aperiodic.

Remark 2.3.11. In the non-primitive case it is possible to have a periodic fixed point and an aperiodic fixed point, in which case the morphism is neither periodic neither aperiodic.

A trivial example is the following,

Example 2.3.12. Let $\varphi : \{012\}^* \to \{012\}^*$ with $\varphi(0) = 01$, $\varphi(1) = 0$ and $\varphi(2) = 22$. This morphism is not primitive and has two distinct right infinite fixed points, the Fibonacci word $\varphi^\omega(0)$ that is known to be aperiodic and the word $\varphi^\omega(2) = 222\ldots$ which is periodic with period 1.

We are going to show several results due to Mossé, concerning the periodicity of fixed points of morphisms.

A morphism $\varphi : A^* \to A^*$ is called *elementary* if it cannot be written as $\varphi = \alpha \circ \beta$ with $\beta : A^* \to B^*$, $\alpha : B^* \to A^*$ and $\mathrm{Card}(B) < \mathrm{Card}(A)$.

Lemma 2.3.13. [42] An elementary morphism is injective as a mapping from $A^{\mathbb{N}}$ to $A^{\mathbb{N}}$.

Proof. Let $\varphi : A^* \to A^*$ be a morphism which is not injective as a mapping from $A^{\mathbb{N}}$ to itself. Set $X = \varphi(A)$ the image of the morphism φ. Let Y be the basis of the intersection of all free submonoids containing X^*, from which follows that $X \subset Y^*$, and let $\beta : B \to Y$ be a bijection from an alphabet B to the free monoid Y. Since Y is a free monoid, the image of a letter $a \in A$ under the morphism φ can be decomposed in a unique way in elements of Y, that is, for any $a \in A$, $\varphi(a)$ can be written as $\varphi(a) = y_1 y_2 \dots y_n$ for $y_i \in Y$. Then, we can define a morphism $\alpha : A^* \to B^*$ that maps each letter $a \in A$ to a word in B^*, i.e., $\alpha(a) = b_1 \dots b_n$ with $b_i \in B$. Since for any letter $a \in A$ we have that,

$$\varphi(a) = \beta \circ \alpha(a) = \beta(b_1 \dots b_n) = y_1 \dots y_n$$

it follows that $\varphi = \beta \circ \alpha$. Let λ be a morphism that maps each word $x \in X$ to the first symbol of its decomposition in words of Y, that is $x \in \lambda(x)Y^*$. If some $y \in Y$ does not appear as an initial symbol in the words of X, set $Z = (Y \setminus y)y^*$. Then Z^* is free and $X^* \subset Z^* \subset Y^*$. Thus $Y = Z$, a contradiction. Since φ is not injective on $A^{\mathbb{N}}$, there are distinct words x, x' such that $\varphi(x) = \varphi(x')$, and since the first letter of the two images is the same, the map λ is also not injective which implies that $\mathrm{Card}(Y) < \mathrm{Card}(X)$, showing that φ is not elementary. □

Lemma 2.3.14. Let $\varphi : A^* \to A^*$ be a periodic primitive elementary morphism with fixed point x, then for every letter $a \in A$ one has $r(a) \leq 1$ (i.e. the letter a can be followed by at most one letter in x) and thus the period of x is at most $\mathrm{Card}(A)$.

Before giving the proof, let us define $p(n)$ as the number of factors of length n of the fixed point x.

Proof. It will be proved that if $p(n) < p(n+1)$ for an integer $n \in \mathbb{N}$, then there is an integer $m > n$ such that $p(m) < p(m+1)$. Indeed if $p(n) < p(n+1)$, then there is word $u \in \mathcal{L}(x)$ with $|u| = n$ which is right-special, which means that there are two distinct letters $a, b \in A$ such that $ua, ub \in \mathcal{L}(x)$. Since φ is elementary, it follows by Lemma 2.3.13 that the morphism φ is injective on $A^{\mathbb{N}}$. Thus there are some words v, w with av, bw in $\mathcal{L}(x)$ such that $\varphi(av) \neq \varphi(bw)$.

If $|\varphi(u)| > |u|$ or if $\varphi(a), \varphi(b)$ begin by the same letter, the longest common prefix of $\varphi(uav), \varphi(ubw)$ (i.e. $\mathrm{lcp}(uav, ubw)$) is a right-special word with length $m > n$. Otherwise, if $|\varphi(u)| = |u|$ and the $\mathrm{lcp}(a, b) = \varepsilon$, we replace u by $\varphi(u)$ and we do the same thing until we will have right-special word longer than n. The existence of such a word is verified because of primitivity, since having an image of a word with the same length as the word itself can happen only a bounded number of times. However, if the morphism is periodic, i.e., the fixed point x is periodic, then $p(n)$ is bounded. Thus, by the previous argument, no letter in the fixed point can be right-special, which implies that the period of x is at most $\mathrm{Card}(A)$. □

The following Lemma gives an upper bound for the periodicity, which implies decidability.

Lemma 2.3.15. Let φ be a primitive morphism with fixed point x. If the morphism is periodic then the period of x is at most $||\varphi||^{\mathrm{Card}(A)-1}$.

Proof. The proof is done by induction on $\mathrm{Card}(A)$. If $\mathrm{Card}(A) = 1$, then the periodicity holds for a period $p = 1$. Let us now study the case of an elementary morphism φ. The morphism φ is injective as a mapping from $A^{\mathbb{N}}$ to $A^{\mathbb{N}}$ as it follows from Lemma 2.3.13. Since φ is also periodic, it follows from Lemma 2.3.14 that the period of x is at most $\mathrm{Card}(A)$.

Let us now consider the not elementary case. Since φ is not elementary, it holds that there are morphisms $\alpha : B^* \to A^*$ and $\beta : A^* \to B^*$ such that $\varphi = \alpha \circ \beta$ and $\mathrm{Card}(B) < \mathrm{Card}(A)$. Set $y = \beta(x)$ and $\psi = \beta \circ \alpha$. Then, $\psi(y) = \psi(\beta(x)) = \beta \circ \alpha(\beta(x)) = \beta(x) = y$. Thus y is a fixed point of the primitive morphism ψ and we may apply the induction hypothesis, that is, ψ is periodic. Last, since $y = \beta(x)$ and therefore $\alpha(y) = \alpha \circ \beta(x) = \varphi(x) = x$. This implies that $||\alpha|| \cdot ||\psi||^{\mathrm{Card}(B)-1} \leq ||\varphi||^{\mathrm{Card}(A)-1}$. □

The following result, as well as Theorem 2.3.17 are going to be used in the proof of Theorem 2.3.18.

Proposition 2.3.16. Let φ be a primitive morphism. There is a $k \geq 0$ such that $||\varphi^n|| \leq k|\varphi^n|$ for all $n \geq 1$.

Proof. Set $M = M(\varphi)$. By Theorem 2.3.4 there is a real number ρ and a strictly positive vector v such that $Mv = \rho v$. Let $\alpha =$

$\min\{v_a \mid a \in A\}$ and $\beta = \max\{v_a \mid a \in A\}$. Then for every $a \in A$,

$$\frac{\alpha}{\beta}\rho^n \leq \frac{\rho^n v_a}{\beta} = \sum_{b\in A} \frac{(M^n)_{a,b}v_b}{\beta} \leq \sum_{b\in A}(M^n)_{a,b} \leq \sum_{b\in A} \frac{(M^n)_{a,b}v_b}{\alpha} = \frac{\rho^n v_a}{\alpha} \leq \frac{\beta}{\alpha}\rho^n.$$

Since for every $a \in A$, $|\varphi^n(a)| = \sum_{b\in A}(M^n)_{a,b}$, this shows that

$$\frac{\alpha}{\beta}\rho^n \leq |\varphi^n| \leq ||\varphi^n|| \leq \frac{\beta}{\alpha}\rho^n$$

and, thus,

$$||\varphi^n|| \leq \frac{\beta}{\alpha}\rho^n \leq \frac{\beta^2}{\alpha^2}|\varphi^n|.$$

This proves the statement with $k = \beta^2/\alpha^2$. □

A word v in a language is called *primitive* if it cannot be written in the form u^i for any word u in the language and any number $i > 1$.

Lemma 2.3.17. Let φ be a primitive morphism with a fixed point x. If there exists a primitive word v and integers $n, p \geq 1$ such that

(i) for every $a, b \in A$ with $ab \in \mathcal{L}(x)$, $\varphi^p(ab)$ is a factor of v^n, and

(ii) $2|v| \leq |\varphi^p|$,

then $\mathcal{L}(x)$ is periodic.

Proof. For every letter $a \in A$, since $\varphi^p(a)$ is a factor of v^n, as factor of $\varphi^p(ab)$ that is factor of v^n, there is an integer n_a, a proper prefix w_a of v and a proper suffix v_a of v such that $\varphi^p(a) = v_a v^{n_a} w_a$. Similarly, there is an integer n_b and word w_b that is proper prefix of v such that $\varphi^p(b) = v_b v^{n_b} w_b$. If $ab \in \mathcal{L}(x)$, then $v^{n_a} w_a v_b v^{n_b} \in \mathcal{L}(x)$ and is a factor of v^n. Since v is primitive, this forces $w_a v_b = v$ or $w_a v_b = \varepsilon$. Thus the language $\mathcal{L}(x)$ is periodic. □

Theorem 2.3.18. *[80] Let $\varphi : A^* \to A^*$ be a primitive aperiodic morphism with right infinite fixed point x. Then there is $m > 0$, such that for any v in A^+, v^m does not belong in the language $\mathcal{L}(x)$.*

Proof. Since φ is primitive, $\mathcal{L}(\mathbf{u})$ is uniformly recurrent. Let $r \geq 1$ be such that every word $ab \in \mathcal{L}(\mathbf{u})$ of length 2 is a factor of every word of length r in $\mathcal{L}(\mathbf{u})$.

Assume that $w^n \in \mathcal{L}(\mathbf{u})$ for some nonempty $w \in \mathcal{L}(\mathbf{u})$. Let $p \geq 1$ be such that $|\varphi^{p-1}| \leq 2|w| < |\varphi^p|$. For every $a, b \in A$ with

$ab \in \mathcal{L}(\mathbf{u})$, since ab appears in all words of the language of length r it follows that the word $\varphi^p(ab)$ appears as factor in each word of $\mathcal{L}(\mathbf{u})$ of length $2r||\varphi^p||$. Hence, by Lemma 2.3.17, we have

$$|w^n| < 2r||\varphi^p|| \tag{2.3.1}$$

since otherwise every $\varphi^p(ab)$ would be factor of w^n with $2|w| < |\varphi^p|$ and thus $\mathcal{L}(x)$ would be periodic, that leads to a contradiction.

Since $|w^n| = n|w|$, we deduce from (2.3.1) the inequality

$$\begin{aligned} n &< \frac{2r||\varphi^p||}{\frac{1}{2}|\varphi^{p-1}|} \\ &< 4r\frac{||\varphi^p||}{|\varphi^{p-1}|} \leq 4r||\varphi||\frac{||\varphi^{p-1}||}{|\varphi^{p-1}|} \leq 4r||\varphi||k \end{aligned}$$

where the last inequality and the constant k follow from Proposition 2.3.16. This shows that for any v in $\mathcal{L}(x) \setminus \{\varepsilon\}$, there is no power v^m of v in $\mathcal{L}(x)$ with m larger that $4r||\varphi||k$. □

2.4 Language of a morphism

Let $\varphi : A^* \to A^*$ be a non-erasing morphism. The set of all the factors of the words $\varphi^n(a)$ for all integers n and all letters a in A is called the *language of the morphism* φ and is denoted by $\mathcal{L}(\varphi)$.

Proposition 2.4.1. Let φ be a non-erasing morphism with right infinite fixed point x. Then $\mathcal{L}(x) \subseteq \mathcal{L}(\varphi)$.

Proof. For a morphism φ with right infinite fixed point x, there is a letter a in A such that $x = \varphi^\omega(a)$. For any word w factor of x there is an integer n such that w is factor of the prefix $\varphi^n(a)$ and since $\varphi^n(a)$ is in $\mathcal{L}(\varphi)$ (because of definition), so is its factor w. □

A similar result holds for the language of a left fixed point of a morphism.

Proposition 2.4.2. Let φ be a non-erasing morphism with left fixed point y. Then $\mathcal{L}(y) \subseteq \mathcal{L}(\varphi)$.

The proof is similar to the proof of Proposition 2.4.1. It is enough to work with suffixes instead of prefixes.

Proposition 2.4.3. Let φ be a primitive morphism with a right infinite fixed point x. Then $\mathcal{L}(x) = \mathcal{L}(\varphi)$.

Proof. Let φ be a primitive morphism with a right infinite fixed point $x = \varphi^\omega(a)$. Because of primitivity, there is an integer k such that all letters of the finite alphabet A appear in $\varphi^k(a)$. Since $\varphi^\omega(a)$ is fixed point it holds that,

$$\varphi^k(\varphi^\omega(a)) = \varphi^\omega(a) = x$$

which proves that all letters appear in the fixed point x and thus all their images appear in x too. Hence, $\mathcal{L}(\varphi) \subseteq \mathcal{L}(x)$.

From Proposition 2.4.1 we have that $\mathcal{L}(x) \subseteq \mathcal{L}(\varphi)$, which proves the wanted equality. □

Proposition 2.4.4. Let φ be a primitive morphism with a left infinite fixed point y. Then $\mathcal{L}(y) = \mathcal{L}(\varphi)$.

The proof is similar to the proof of Proposition 2.4.3.

The previous results do not hold for the case of the two-sided infinite fixed point, as it can be seen in the following example, where $z = y \bullet x$ in $A^{\mathbb{Z}}$ with $\mathcal{L}(z) \not\subset \mathcal{L}(\varphi)$.

Example 2.4.5. Let φ be the primitive morphism such that $\varphi(0) = 02$, $\varphi(1) = 201$ and $\varphi(2) = 212$. Since $\varphi(0)$ is in $0A^*$, there is unique right infinite fixed point x and since $\varphi(1)$ in A^*1, there is unique left fixed point y as follows,

$$x = \varphi^\omega(0) = 022122122\ldots \text{ and } y = \varphi^{\tilde{\omega}}(1) = \ldots 21202201.$$

Let $z = y \bullet x$ be a two-sided infinite fixed point of φ. The word 10 is in language $\mathcal{L}(z)$ but not in the language of the morphism $\mathcal{L}(\varphi)$. (It can be proved using extension graph for the empty word, as defined in Section 4.1).

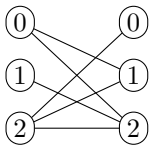


Figure 2.1: The extension graph of $\mathcal{E}(\varepsilon)$.

A two-sided fixed point $z = y \bullet x$ of a morphism φ is called *admissible* if $z_{-1}z_0 = y_{-1} \bullet x_0$ is in the language $\mathcal{L}(\varphi)$, where x is a right infinite fixed point and y is a left infinite fixed point of the morphism φ.

Example 2.4.6. The word z in Example 2.4.5 is not admissible. For the morphism $\varphi : \{0,1\}^* \to \{0,1\}^*$ with $\varphi(0) = 010$ and $\varphi(1) = 0$ the two-sided infinite fixed point $z = \varphi^{\tilde{\omega}}(0) \bullet \tilde{\varphi}^{\omega}(0)$ is admissible.

Observation 2.4.7. For a two-sided infinite fixed point $z = y \bullet x$ of a morphism φ it holds that $\mathcal{L}(x), \mathcal{L}(y) \subseteq \mathcal{L}(z)$.

Proposition 2.4.8. Let z be an admissible two-sided fixed point of a morphism φ, then $\mathcal{L}(z) \subseteq \mathcal{L}(\varphi)$.

Proof. Let z be an admissible two-sided fixed point. Since $z_{-1}z_0$ is in the language $\mathcal{L}(\varphi)$, it follows that $\varphi^n(z_{-1}z_0)$ is in $\mathcal{L}(\varphi)$, for all positive integers n. Thus, any factor of the two-sided fixed point z is in $\mathcal{L}(\varphi)$, which implies that $\mathcal{L}(z) \subseteq \mathcal{L}(\varphi)$. □

Example 2.4.9. Let φ be the morphism with $\varphi(0) = 01$, $\varphi(1) = 01$ and $\varphi(2) = 0$. The two-sided $z = \varphi^{\tilde{\omega}}(1) \bullet \varphi^{\omega}(0)$ is admissible fixed point since $10 \in \mathcal{L}(\varphi)$. It holds that $\mathcal{L}(z) \subset \mathcal{L}(\varphi)$ but the equality between the two languages does not hold since 2 is in $\mathcal{L}(\varphi)$ but not in $\mathcal{L}(z)$.

Proposition 2.4.10. Let z be an admissible two-sided infinite fixed point of a primitive morphism φ, then $\mathcal{L}(z) = \mathcal{L}(\varphi)$.

Proof. It follows from Proposition 2.4.8 that $\mathcal{L}(z) \subseteq \mathcal{L}(\varphi)$. In order to prove the other side, it is enough to recall that since the morphism is primitive, all letters of the alphabet can be found in $\varphi^i(z)$ for some integer $i > 0$. Since z is a fixed point it holds that $\varphi^i(z) = z$ which implies that all letters can be found in the fixed point. Since $\varphi^k(z) = z$ for all integers $k \geq 0$, it follows that all words $w \in \mathcal{L}(\varphi)$ are factors of z. Thus, $\mathcal{L}(\varphi) \subseteq \mathcal{L}(z)$ which proves that the two sets coincide. □

Example 2.4.11. The language $\mathcal{L}(z)$ from Example 2.4.6 coincides with the language $\mathcal{L}(\varphi)$ of the morphism.

A *connection* for a substitution φ is a pair (b, a) of letters in A such that the first letter of $\varphi^{\omega}(a)$ is a, and the last letter of $\varphi^{\tilde{\omega}}(b)$ is b and the word ba is in $\mathcal{L}(\varphi)$.

Remark 2.4.12. A two-sided infinite fixed point z of a morphism φ is admissible if and only if (z_{-1}, z_0) is a connection.

Proposition 2.4.13. Let z be an admissible two-sided fixed point of a primitive morphism φ such that $z = y \bullet x$, then $\mathcal{L}(z) = \mathcal{L}(x) = \mathcal{L}(y) = \mathcal{L}(\varphi)$.

Proof. The proof follows easily from Propositions 2.4.10, 2.4.3 and 2.4.4. □

Proposition 2.4.14. Every morphism on an alphabet A has a power with a right (resp. left) infinite fixed point in $A^{\mathbb{N}}$.

The following proof is for the right infinite fixed point. The proof for the left infinite fixed point is similar.

Proof. Let $\varphi : A^* \to A^*$ be a non-erasing morphism. Let A_1 be the set defined as follows,

$$A_1 = \{a \in A \mid |\varphi^n(a)| \to \infty\}.$$

Let A_0 be the complement of A_1, i.e., $A = A_0 \cup A_1$. For the set A_0 we have that,

$$\varphi(A_0) \subset A_0^*.$$

Indeed, for a letter a in A_0 let the image under the morphism φ be,

$$\varphi(a) = \beta_0\beta_1 \dots \beta_n \text{ , for } \beta_i \in A.$$

If there is a β_i, with $0 \leq i \leq n$, such that β_i in A_1, then $|\varphi^n(\beta_i)| \to \infty$, which implies that $|\varphi^n(a)| \to \infty$, that leads to a contradiction.

First we will examine the case where $A_0 \neq \emptyset$. There exists a letter a in A for which there is integer n such that $\varphi^n(a) = a$. Hence, $\varphi(a^\omega) = a^\omega$ is a right infinite fixed point.

Let us now examine the case where $A_0 = \emptyset$. It holds that for any letter a,

$$\varphi^n(a) \in a_n A^* \text{ and } \varphi(a_n) = a_{n+1}A^*, \text{ with } a_n, a_{n+1} \in A$$

for all integers n. Since the alphabet is finite, there exist integers i, j with $i < j$ such that $a_i = a_j$ and $\varphi^{j-i}(a_i)$ is in $a_i A^*$, which proves that there is a right infinite fixed point for the power $j - i$ of the morphism φ. □

Example 2.4.15. Let φ be the morphism with $\varphi(0) = 1$ and $\varphi(1) = 01$. For the morphism φ there is no letter in the alphabet A such that $\varphi(a) \in aA^*$. However, for the 2-power of the morphism holds that $\varphi^2(0) = 01 \in 0A^*$ and $\varphi^2(1) = 101 \in 1A^*1$. It follows that $(\varphi^2)^\omega(0)$ and $(\varphi^2)^\omega(1)$ are right infinite fixed points and a $(\varphi^2)^{\tilde{\omega}}(1)$ is left infinite fixed point of the morphism φ^2.

Proposition 2.4.16. Every primitive morphism has a power that has a connection.

Proof. Let a, b be letters in A such that $ba \in \mathcal{L}(\varphi)$. Since A is finite it follows that there are integers i, p and j, q such that $\varphi^i(a)$ and $\varphi^{i+p}(a)$ start with the same letter, without loss of generality let it be $c \in A$, and $\varphi^j(b)$ and $\varphi^{j+q}(b)$ end with the same letter, without loss of generality let it be $d \in A$. Let $k = \max\{i, j\}$, since $ba \in \mathcal{L}(\varphi)$ it follows that $\varphi^k(ba) \in \mathcal{L}(\varphi)$ and hence its factor dc is also in $\mathcal{L}(\varphi)$. We have that $\varphi^p(c)$ starts with the letter c and $\varphi^q(d)$ ends with the letter d. Let $\ell = \mathrm{lcm}(p, q)$, then $\varphi^r(c) \in cA^*$ and $\varphi^r(d) \in A^*d$ which implies that $(\varphi^\ell)^\omega(c) \in cA^*$ and $(\varphi^\ell)^{\tilde{\omega}}(d) \in A^*d$. Thus, (d, c) is a connection for φ^ℓ. □

Observation 2.4.17. Every morphism has a power with an admissible two-sided infinite fixed point.

Example 2.4.18. The morphism φ from Example 2.4.15 has no connection but the power φ^2 has connection $(1, 0)$ and two-sided infinite fixed point $z = (\varphi^2)^{\tilde{\omega}}(1) \bullet (\varphi^2)^\omega(0)$.

3 Recognizability

The recognizability of morphisms is a core notion underpinning various scientific fields, ranging from combinatorics on words to symbolic dynamics. In the nominal work of Martin [77], the author examines the concept of recognizability using the term of determinization. The goal of this chapter is to elucidate further the notion of recognizability of morphisms and bridge the various definitions found in the literature. The term first appears in the work of B. Host, "Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable" [58]. A notion that bears apparent similarity is that of circularity, to the extent where the two terms are used interchangeably in the literature. Nevertheless, these two terms are not entirely equivalent, and this inconsistency across definitions might act as a root of confusion in the field. This chapter aims to accumulate the existing definitions and bridge them under a common framework, where the equivalence of each is studied against that of the others.

3.1 Synchronization point

In this chapter we will denote by $\mathbf{u}$ the right infinite fixed point of a morphism and we will refer to it as fixed point. Also, when not specifically defined otherwise, we consider that a morphism is non-erasing.

Definition 3.1.1. Let $\varphi : A \to A^*$ be a morphism with fixed point $\mathbf{u} \in A^{\mathbb{N}}$. A triplet (p, v, s) where $p, s \in A^*$, $v = v_0 v' v_n \in \mathcal{L}(\mathbf{u}) \setminus \{\varepsilon\}$,

$v' = v_1 \dots v_{n-1}$, $(v_i)_{0 \le i \le n} \in A$ is an *external interpretation* of a word $u \in \mathcal{L}(\mathbf{u})$ if $\varphi(v) = pus$, where p is prefix of $\varphi(v_0) = pq$ and s is suffix of $\varphi(v_n) = rs$ and $u = q\varphi(v')r$, as shown in Figure 3.1, while the triplet (q, v', r) is an *internal interpretation* of the word u when it holds that $u = q\varphi(v')r$.

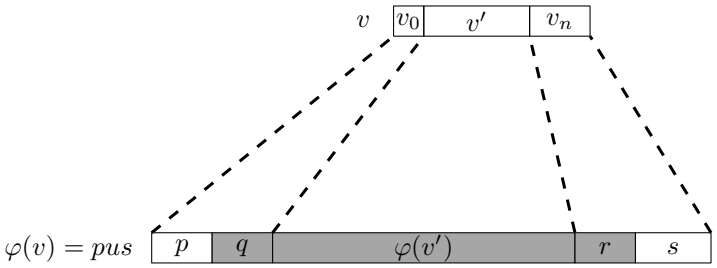


Figure 3.1: A graphical explanation of the definitions of external and internal interpretation.

Remark 3.1.2. If there is an external interpretation (p, v, s) of a word u in $\mathcal{L}(\mathbf{u})$, then there is an internal interpretation too.

Let φ be a morphism with fixed point $\mathbf{u}$. If there is an external interpretation (p, w, s) of u on $\mathbf{u}$, then we say that the word $w \in \mathcal{L}(\mathbf{u})$ *covers* the word $u \in \mathcal{L}(\mathbf{u})$ if it holds that u is a factor of $\varphi(w)$ but it is not a factor of $\varphi(v)$ for any proper factor v of w. Note that the number of words covering a finite word u is finite.

Example 3.1.3. Let φ_F be the Fibonacci morphism (Example 2.2.2). The word $w = 010$ covers the word $u = 0100$, as it can be seen in Figure 3.2. Indeed, there is no proper factor v of w such with u would be factor of $\varphi(v)$, while u is factor of $\varphi(w)$.

In the case of a primitive morphism, every word $w \in \mathcal{L}(\mathbf{u})$ can be extended both on the left and the right and therefore an internal interpretation implies the existence of at least one external interpretation.

A morphism φ is called *prefix (resp. suffix)* if φ is non-erasing and for each $a, b \in A$ if $\varphi(a)$ is prefix (resp. suffix) of $\varphi(b)$, then $\varphi(a) = \varphi(b)$. A morphism φ is called *bifix* if φ is simultaneously prefix and suffix.

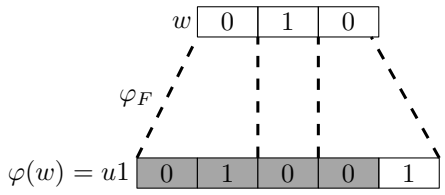


Figure 3.2: In the Fibonacci language the word $w = 010$ covers $u = 0100$.

Example 3.1.4. The Fibonacci morphism φ_F (Example 2.2.2) is suffix since no image of a letter is suffix of the image of another letter but it is not prefix since $\varphi_F(1) = 0$ is a prefix of $\varphi_F(0) = 01$.

Let $u \in \mathcal{L}(\mathbf{u})$. We say that (u_1, u_2) is a *synchronization point* of u for the morphism φ with fixed point $\mathbf{u}$, if $u = u_1u_2$ and

$$\forall p, s, v \in \mathcal{L}(\mathbf{u}) \text{ such that } \varphi(v) = pus \Rightarrow \exists v_1v_2 \text{ factorization of } v \text{ with } \varphi(v_1) = pu_1 \text{ and } \varphi(v_2) = u_2s.$$

A synchronization point is a boundary between two elements of $\varphi(\mathcal{L}(\mathbf{u}))$ whenever w appears as a factor of an element of $\varphi(A^*)$ and can be denoted by $|$, i.e $w = w_1|w_2$. It is permitted to have several synchronization points in a word, for example (w_1, w_2w_3) and (w_1w_2, w_3) are two synchronization points of the word $w = w_1w_2w_3$, i.e., $w = w_1|w_2|w_3$.

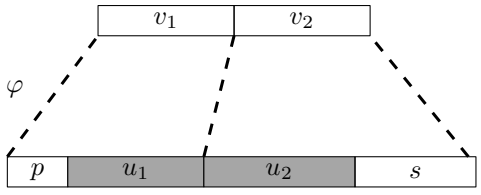


Figure 3.3: A graphical explanation of the definition of the synchronizing point.

An interpretation of $u = u_1u_2$ *passes* by (u_1, u_2) if there is $v = v_1v_2$ such that $\varphi(v_1) = pu_1$ and $\varphi(v_2) = u_2s$.
From the previous definition we can easily obtain the following proposition,

Proposition 3.1.5. Every external interpretation of $u \in \mathcal{L}(\mathbf{u})$ passes by (u_1, u_2) if and only if (u_1, u_2) is a synchronization point of u.

Example 3.1.6. Let φ be the Fibonacci morphism (Example 2.2.2) with fixed point the Fibonacci word $\mathbf{u}$. The word $00101 \in \mathcal{L}(\mathbf{u})$ has two synchronizing points $(0, 0101)$ and $(001, 01)$, i.e., $0|01|01$.

Using the word *cutting point* we will indicate the positions of a word w that are such that $w = w_1 w_2$ and there exist words u_1, u_2 with $pw_1w_2s \in \mathcal{L}(\mathbf{u})$ such that $\varphi(u_1) = pw_1$, $\varphi(u_2) = w_2 s$ with $p, s \in A^*$ and $u_1 u_2 \in \mathcal{L}(\mathbf{u})$.

Observation 3.1.7. A synchronization point is "fixed" for a word no matter where the word may appear as factor in the fixed point ("fixed" in the sense that if $v = v_0 \dots v_n \in \mathcal{L}(\mathbf{u})$ has synchronizing point at position k, with $0 \leq k < n$, then if $\mathbf{u}_{[i,i+n]} = v$, there is synchronization point in position $i + k$). A synchronization point is a cutting point while every cutting point is not always a synchronization point, as it is shown in the following example.

Example 3.1.8. Let φ_{TM} be the Thue-Morse morphism with $\varphi_{TM}(0) = 01$, $\varphi_{TM}(1) = 10$ and fixed point,

$$\mathbf{u} = \varphi_{TM}^{\omega}(0) = 01101001100101 \cdots = |01|10|10|01|10|01|01| \dots .$$

The word $010 \in \mathcal{L}(\mathbf{u})$ contains one cutting point because of its length ($|010| > ||\varphi_{TM}|| = 2$). Specifically, there are two cutting points, at position 1 and 2, but none of these is a synchronization point. Indeed, for the pair $(0, 10)$ there are words $v_1 = 1$ and $v_2 = \varepsilon$, such that $v_1 010 v_2 = 1010 \in \varphi(\mathcal{L}(u))$ and $10 \in \varphi(\mathcal{L}(u))$. However, if we change the extensions to the words $v_1' = \varepsilon$ and $v_2' = 1$ with $v_1' 010 v_2' = 0101 \in \varphi(\mathcal{L}(\mathbf{u}))$ with $101 \notin \varphi(\mathcal{L}(\mathbf{u}))$, which implies that $(0, 10)$ is not synchronization point. Similarly, it can be seen that the pair $(01, 0)$ is also not a synchronization point. However, the word $011 \in \mathcal{L}(\mathbf{u})$, has a synchronizing point $(01, 1)$ because every decomposition passes from the position between the two consecutive 1's, since they cannot appear as a factor of the image of any letter of the alphabet under the morphism φ_{TM}. Since the position 2 is synchronization point, it is also a cutting point as it follows from the above observation.

A morphism $\varphi : A^* \to A^*$ is called *k-uniform* if for all $a \in A$, we have $|\varphi(a)| = k$.

Example 3.1.9. The Thue-Morse morphism φ_{TM} (Example 3.1.8) is 2-uniform. Particularly it is prefix and suffix, thus it is bifix. The Fibonacci morphism φ_F (Example 2.2.2) is not k-uniform for any integer k.

Definition 3.1.10. Let φ be a morphism with fixed point $\mathbf{u}$ and let $u \in \mathcal{L}(\mathbf{u})$. We say that (u_1, u_2) is a *strong synchronization point* if there exists letter $a \in \mathcal{L}(\mathbf{u})$ such that ,

$$\forall p, s, v \in \mathcal{L}(\mathbf{u}) \text{ such that } \varphi(v) = pus \Rightarrow \exists v_1 v_2 \text{ factorization of } v \text{ with } \varphi(v_1) = pu_1, \varphi(v_2) = u_2 s \text{ and the first letter of } v_2 \text{ is } a.$$

Remark 3.1.11. Every strong synchronization point of a word $u \in \mathcal{L}(\mathbf{u})$ is a synchronization point but the reverse is not necessarily true, as shown bellow.

Example 3.1.12. Let φ_F be the Fibonacci morphism (Example 3.2.3) and $\mathbf{u}$ be the Fibonacci word. The word $10 \in \mathcal{L}(\mathbf{u})$ has a synchronization point $(1, 0)$. There are two external interpretations of u, $(0, 00, 1)$ and $(0, 01, \varepsilon)$. For the interpretation $(0, 00, 1)$ of u there is $v = 00$ with $v_1 = 0$ and $v_2 = 1$, while for $(0, 01, \varepsilon)$ there is $v' = 01$ with $v'_1 = 0$ and $v'_2 = 1$ that verify the definition of the synchronizing point. Since v_2 and v'_2 do not start with the same letter, the synchronization point $(1, 0)$ is not strong.

The word $w = 00 \in \mathcal{L}(\mathbf{u})$ has synchronization point $(0, 0)$ and unique external interpretation $(\varepsilon, 10, 1)$ and therefore there are unique $v_1 = 1$ and $v_2 = 0$, which means that the word v_2 starts always with the letter 0 and thus $(0, 0)$ is a strong synchronization point of w.

The word $w' = 0010 \in \mathcal{L}(\mathbf{u})$ has a synchronization point $(0, 010)$ and it also has two different external interpretations that pass from that position, $(\varepsilon, 100, 1)$ and $(\varepsilon, 101, \varepsilon)$. For the interpretation $(\varepsilon, 100, 1)$ there are $v_1 = 0$, $v_2 = 01$ and for the interpretation $(\varepsilon, 101, \varepsilon)$ there are $v'_1 = 0$, $v'_2 = 00$. The words v_2 and v'_2 start with the same letter and thus the synchronization point $(0, 010)$ is strong.

Observation 3.1.13. In the case of a prefix morphism every synchronization point is a strong synchronization point.

The above observation holds since if a word of a substitutive language of a prefix morphism has a synchronization point, then the decoding after that synchronization point is unique.

3.2 Determined words

The first time that the idea of identifying the interpretations of a word of a substitutive language took place was in the work of John C. Martin [77], where the following definition appears,

Definition 3.2.1 (Martin)**.** Let $\varphi : A^* \to A^*$ be a morphism with right infinite fixed point $\mathbf{u}$. Let $u \in \mathcal{L}(\mathbf{u})$ longer than $||\varphi(a)||$. The word u is *determined* if there is a unique internal interpretation (q, v, r) of u.

See Figure 3.4 for a graphical representation of the above definition.

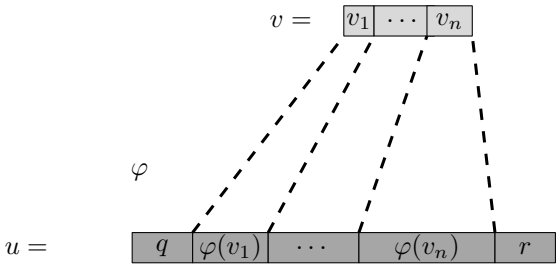


Figure 3.4: A graphical explanation of Definition 3.2.1.

Example 3.2.2. Let us work with the Thue-Morse morphism φ_{TM} (Example 3.1.8) and fixed point $\mathbf{u} = \varphi_{TM}^{\omega}(0)$. All words $u \in \mathcal{L}(\mathbf{u})$ with length at least 5 are determined. Specifically, all the words with such length having as a factor one of the words 00 or 11 (i.e. 0010, 0011, 0110, 1100, 1101) have a synchronization point between the 0's and the 1's, since there is no letter in the alphabet that under the morphism φ_{TM} has an image with factor 00 or 11. The only words of length 4 in the language that do not have two consecutive same letters are 0101 and 1010, which have unique decomposition, 01|01

and 10|10 respectively, and therefore unique internal interpretations $(\varepsilon, 00, \varepsilon)$ and $(\varepsilon, 11, \varepsilon)$ respectively. Indeed, if the previous is not true, then their preimages are factors of length more than 3 of the sequence 0^* or 1^* that is not in the language $\mathcal{L}(\mathbf{u})$ (we cannot have more than two consecutive 0's or 1's). Finally, every long enough word $u \in \mathcal{L}(\mathbf{u})$ ($|u| \geq 4$) has at least one synchronizing point. Since the Thue-Morse morphism φ_{TM} is bifix, it follows that the internal interpretation of the word u is also unique.

Later, Martin stated the following (Lemma 1 [77]):
If φ is primitive, there exists an integer t such that any factor of length t of the fixed point $\mathbf{u}$ is determined.

The proof of the above lemma was characterized as not convincing and the following counterexample (Example 3.2.3) proves that the above lemma is not true.

Example 3.2.3. Let φ_F be the Fibonacci morphism that is primitive, with fixed point $\mathbf{u} = \varphi^\omega(0)$ (Example 2.2.2). It is known that the prefixes of the Fibonacci word are left-special and that the set of factors of the Fibonacci word is closed by reversal. Thus, the reversal of any prefix of the Fibonacci word is right special. Hence, there are right special words of every length. Let us suppose that u is a right special word in $\mathcal{L}(\mathbf{u})$ (i.e. can be extended on the right by 0 and 1). Since there is no word in $\mathcal{L}(\mathbf{u})$ that has two consecutive 1's, the last letter of u has to be 0. Similarly, we conclude that $u = w10$ since otherwise we would had three consecutive 0's (that implies that the preimage of this factor has two consecutive 1's, that is not possible). It can be seen that there are two distinct internal interpretations of u, $(\varepsilon, v1, \varepsilon)$ and $(\varepsilon, v, 0)$, where $\varphi(v) = w1$. (See Figure 3.5)

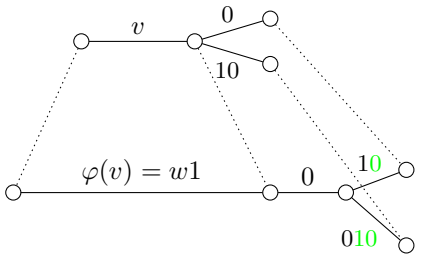


Figure 3.5: A graphiccal explanation of Example 3.2.3.

We introduce the following definition that describes the morphisms that satisfy Martin's lemma (Lemma 1 [77]).

Definition 3.2.4. Let $\varphi : A^* \to A^*$ be a morphism with a fixed point $\mathbf{u}$. We say that φ is *strongly circular* if there is an integer t such that any factor of $\mathbf{u}$ of length t is determined.

In other words, a morphism is strongly circular if there is a unique internal interpretation of any word of length at least t.

Example 3.2.5. As it has been explained in Example 3.2.2 the Thue-Morse morphism φ_{TM} is strongly-circular for $t = 4$, while according to Example 3.2.3 the Fibonacci morphism φ_F is not strongly-circular.

3.3 Right, left and two-sided recognizability

The term *recognizable* was introduced by B. Host [58] and M. Queffélec [87]. In Mossé's work [80] there is the notion of one-sided recognizability (right recognizable and left recognizable) and that of two-sided recognizability. The first one is the one used by B. Host (Definition 3.3.2).

3.3.1 Right recognizability

Let $\mathbf{u}$ be a fixed point of a morphism φ. We will define the set $E_{k,\mathbf{u}}$, where $k \geq 1$, as follows,

$$E_{k,\mathbf{u}} = \{0\} \cup \{|\varphi^k(\mathbf{u}_{[0,p-1]})|; p > 0\}.$$

We are going to use the notation $E_{\mathbf{u}}$ instead of $E_{1,\mathbf{u}}$. The set $E_{\mathbf{u}}$ gives all the cutting points of the fixed point.

It holds that $E_{k,\mathbf{u}} \subset E_{k-1,\mathbf{u}}$, $\forall k > 1$. Indeed, if $i \in E_{k,\mathbf{u}}$, then there is a word u prefix of $\mathbf{u}$ such that $i = |\varphi^k(u)| = |\varphi^{k-1}(\varphi(u))| = |\varphi^{k-1}(u')|$, with $u' = \varphi(u)$ prefix of the fixed point. Thus, $i \in E_{k-1,\mathbf{u}}$. Hence, $E_{k,\mathbf{u}} \subset E_{k-1,\mathbf{u}}$.

Example 3.3.1. Let φ be the Fibonacci morphism (Example 2.2.2) with fixed point,

$$\mathbf{u} = 0100101001\ldots.$$

It can be seen that $E_{\mathbf{u}} = \{0, 2, 3, 5, 7, 8, 10, \ldots\}$ and $E_{2,\mathbf{u}} = \{0, 3, 5, 8, \ldots\}$ for which holds that $E_{2,\mathbf{u}} \subset E_{\mathbf{u}}$ as it was expected.

Let $f_{\mathbf{u}}$ denote the function

$$\begin{aligned} f_{\mathbf{u}}: \ \mathbb{N} &\to \mathbb{N} \\ i &\mapsto f_{\mathbf{u}}(i) = \begin{cases} |\varphi(\mathbf{u}_{[0,i)})| & \text{if } i > 0, \\ 0 & \text{if } i = 0. \end{cases} \end{aligned}$$

The function $f_{\mathbf{u}}$ can be extended for the two-sided infinite word case as follows,

$$\begin{aligned} f_{\mathbf{u}}: \ \mathbb{Z} &\to \mathbb{Z} \\ i &\mapsto f_{\mathbf{u}}(i) = \begin{cases} |\varphi(\mathbf{u}_{[0,i)})| & \text{if } i > 0, \\ 0 & \text{if } i = 0, \\ -|\varphi(\mathbf{u}_{[i,0)})| & \text{if } i < 0. \end{cases} \end{aligned}$$

It is implied that $E_{\mathbf{u}} = f_{\mathbf{u}}(\mathbb{N})$. When it is clear from the context, we simply write f instead of $f_{\mathbf{u}}$.

The following definition appears in the work of B. Host [58].

Definition 3.3.2. [58] Let $\varphi : A^* \to A^*$ be a morphism with a fixed point $\mathbf{u}$. We say that φ is *right recognizable* on $\mathbf{u}$ if there is an integer $L > 0$ such that if $\mathbf{u}_{[i,i+L]} = \mathbf{u}_{[j,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$. (See Figure 3.6)

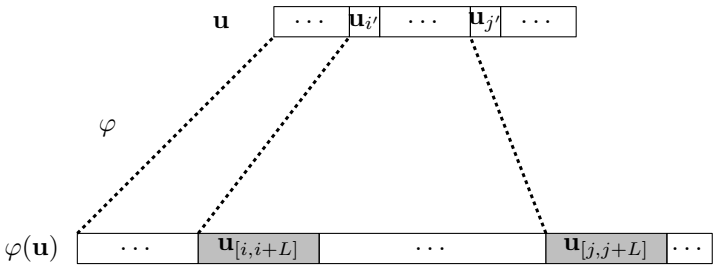


Figure 3.6: A graphical illustration of Definition 3.3.2.

Example 3.3.3. The Fibonacci morphism φ_F is right recognizable with $L = 1$. All the possible words of length 2 in the language $\mathcal{L}(\mathbf{u})$ are $00, 01, 10$. For the words $01, 00$ we know that there is always a synchronization point in the first position since the letter 0 is always

the beginning of an image of a letter, hence $i \in E_{\mathbf{u}}$ for all i such that $\mathbf{u}_{[i,i+1]} = 00$ or 01. Contrary to this, for the word 10 we are sure that $i \notin E_{\mathbf{u}}$ for all i such that $\mathbf{u}_{[i,i+1]} = 10$.

The following property shows that the recognizability on a fixed point $\mathbf{u}$ of a morphism φ depends only on the language $\mathcal{L}(\mathbf{u})$ of the fixed point.

Proposition 3.3.4. Let φ be a morphism with a fixed point $\mathbf{u}$. Then φ is right recognizable on $\mathbf{u}$ for some constant L, if and only if any word $u \in \mathcal{L}(\mathbf{u})$ of length $L+1$ with an external interpretation (ε, v, s) has a synchronization point (ε, u).

Proof. Let φ be a right recognizable morphism on a fixed point $\mathbf{u}$ for some constant $L > 0$. Let $u \in \mathcal{L}(\mathbf{u})$ with $|u| = L + 1$ and external interpretation (ε, v, s). There is an occurrence i of u such that $\varphi(v) = us = \mathbf{u}_{[i,i+L]}s$. Let us now take another external interpretation (p', v', s') of u on $\mathbf{u}$. There is an occurrence $j \neq i$ of u such that $\varphi(v') = p'\mathbf{u}_{[j,j+L]}s' = p'us'$ (See Figure 3.7). Since φ is right recognizable on $\mathbf{u}$ and $i \in E_{\mathbf{u}}$, it follows that $j \in E_{\mathbf{u}}$ and therefore $j \in E_{\mathbf{u}}$. Thus $p' = \varepsilon$ and (ε, u) is a synchronization point.

In order to prove the other direction, let $u \in \mathcal{L}(\mathbf{u})$ have an occurrence $i \in E_{\mathbf{u}}$. Then, there is external interpretation (ε, v, s) of u and (ε, u) is a synchronization point, from hypothesis. Thus (ε, u) is a synchronization point. Thus, for any other occurrence j of u with external interpretation (p', v', s') (i.e. $\varphi(v') = p'\mathbf{u}_{[j,j+L]}s'$ with $\mathbf{u}_{[j,j+L]} = u$) holds that $p' = \varepsilon$. Then $j \in E_{\mathbf{u}}$ which proves that right recognizability holds for $L > 0$. □

Proposition 3.3.5. Let φ be a morphism with fixed points $\mathbf{u}, \mathbf{v}$ such that $\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{v})$. The morphism φ is right recognizable on $\mathbf{v}$ if and only if it is right recognizable on $\mathbf{u}$.

Proof. This result follows from Proposition 3.3.4. □

Remark 3.3.6. The hypothesis $\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{v})$ is satisfied in the primitive case since the languages of all fixed points coincide (i.e. if $\mathbf{u}, \mathbf{v} \in A^{\mathbb{N}}$ are fixed points of the primitive morphism φ, then $\mathcal{L}(\mathbf{u}) = \mathcal{L}(\mathbf{v})$). In this case the above result is consequence of Mossé's Theorem (to be seen below).

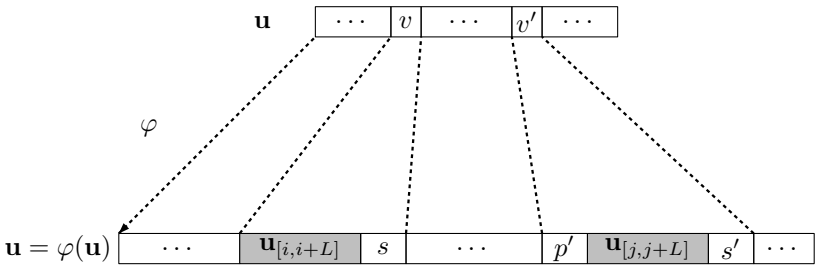


Figure 3.7: Occurrences i and j of $u \in \mathcal{L}(\mathbf{u})$ from the proof of Proposition 3.3.4.

Proposition 3.3.7. [80] Let φ be a primitive morphism with aperiodic fixed point. The morphism is not right recognizable if and only if for any integer $L > 0$ there is w with $|w| = L + 1$ and two letters $a, b \in A$ such that,

1. the word $\varphi(b)$ is a strict suffix of the word $\varphi(a)$,
2. the words $\varphi(a)w$ and $\varphi(b)w$ are in the language $\mathcal{L}(\mathbf{u})$ and they have the same decoding.

The following proposition indicates the relation between strongly circularity and right recognizability.

Proposition 3.3.8. If a morphism is strongly circular then it is right recognizable.

Proof. Let φ be a strongly circular morphism for some constant $t > 0$ and a fixed point $\mathbf{u}$. Let $u \in \mathcal{L}(\mathbf{u})$ with $|u| = t$ and let i be an occurrence of u in $\mathbf{u}$, i.e., $\mathbf{u}_{[i,i+|u|)} = u$, with $i \in E_{\mathbf{u}}$. Since φ is strongly circular for t, the word u has a unique internal interpretation (p, v, s) with $u = p\varphi(v)s$. Since $i \in E_{\mathbf{u}}$ the interpretation must pass by the position i which implies that $p = \varepsilon$. Let $j > 0$ be another occurrence of u, i.e., $\mathbf{u}_{[j,j+|u|)} = u$. Since the internal interpretation is always (ε, v, s) it holds that $j \in E_{\mathbf{u}}$ and therefore the morphism is right recognizable. □

The inverse of the previous proposition is not always true, as it is indicated in the following example.

Example 3.3.9. Let $\varphi : \{0,1,2\}^* \to \{0,1,2\}^*$ with $\varphi(0) = 0120$, $\varphi(1) = 12$, $\varphi(2) = 12$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. It can be verified that this morphism is right recognizable for the constant $L = 2$. However, this is not strongly circular. Indeed, for any integer $t > 1$, there is a word u,

$$u = \begin{cases} (12)^{\frac{t}{2}} & \text{if } t \text{ is even,} \\ (12)^{\frac{t-1}{2}}1 & \text{if } t \text{ is odd.} \end{cases}$$

that have at least two distinct internal interpretations. More precisely, in the case of $t = 2$, there is the word $u = 12$ with interpretations $(\varepsilon, 1, \varepsilon)$ and $(\varepsilon, 2, \varepsilon)$. Similarly, for any longer pair integer t such that $t \mod 4 = 0$ there is a word $w = (12)^{\frac{t}{2}}$ with interpretations $(\varepsilon, (12)^{\frac{t}{4}}, \varepsilon)$ and $(\varepsilon, (21)^{\frac{t}{4}}, \varepsilon)$. For any pair t with $t \mod 4 = 2$ the word w has interpretations $(\varepsilon, (12)^{\lfloor \frac{t}{4} \rfloor}1, \varepsilon)$ and $(\varepsilon, (21)^{\lfloor \frac{t}{4} \rfloor}1, \varepsilon)$. Finally, for an odd integer t there is the word $w' = (12)^{\frac{t-1}{2}}1$ with interpretations $(\varepsilon, (12)^{\lfloor \frac{t-1}{4} \rfloor}, 2)$ and $(\varepsilon, (21)^{\lfloor \frac{t-1}{4} \rfloor}, 2)$. Thus, the morphism is right recognizable but not strongly circular.

Remark 3.3.10. If the morphism is prefix and right recognizable then it is strongly circular.

3.3.2 Left recognizability

Definition 3.3.11. Let $\varphi : A^* \to A^*$ be a morphism with fixed point $\mathbf{u}$. We say that φ is *left recognizable* on $\mathbf{u}$ if there is an integer $L > 0$ such that for every $i > L$ if $\mathbf{u}_{[i-L,i]} = \mathbf{u}_{[j-L,j]}$ and $i + 1 \in E_{\mathbf{u}}$, then $j + 1 \in E_{\mathbf{u}}$.

Example 3.3.12. Let φ be the morphism on the alphabet $A = \{0,1\}$ such that $\varphi(0) = 010$, $\varphi(1) = 10$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. The morphism φ is left recognizable with $L = 1$. All the possible words of length 2 in the language $\mathcal{L}(\mathbf{u})$ are $00, 01, 10$. For the word 01 we know that there is never a synchronization point in the last position since the letter 1 is never the ending of an image of a letter, hence $i \notin E_{\mathbf{u}}$ for all i such that $\mathbf{u}_{[i,i+1]} = 01$. Similarly to this, for the word 00 it can be easily seen that there is a synchronization between the two 0's and never at the last position. Thus, $i \notin E_{\mathbf{u}}$ for all i such that $\mathbf{u}_{[i,i+1]} = 00$ Contrary to the previous cases, for the word 10 we are sure that $i \in E_{\mathbf{u}}$ for all i such that $\mathbf{u}_{[i,i+1]} = 10$.

Proposition 3.3.13. Let φ be a morphism with a fixed point $\mathbf{u}$. Then φ is left recognizable on $\mathbf{u}$ for some constant L, if and only if any word $u \in \mathcal{L}(\mathbf{u})$ of length $L+1$ with an external interpretation (p, v, ε) has a synchronization point (u, ε).

Proof. Let φ be a left recognizable morphism on a fixed point $\mathbf{u}$ for some constant $L > 0$. Let $u \in \mathcal{L}(\mathbf{u})$ with $|u| = L+1$ and external interpretation (p, v, ε). There is an occurrence $i - L$ of u such that $\varphi(v) = pu = \mathbf{u}_{[i-L,i]}$. Let us now take another external interpretation (p', v', s') of u on $\mathbf{u}$. There is an occurrence $j - L \neq i$ of u such that $\varphi(v') = p'\mathbf{u}_{[j-L,j]}s' = p'us'$ (See Figure 3.8). Since φ is left recognizable on $\mathbf{u}$ and $i+1 \in E_{\mathbf{u}}$, it follows that $j+1 \in E_{\mathbf{u}}$. Thus, $s' = \varepsilon$ and (u, ε) is a synchronization point of u on $\mathbf{u}$.

In order to prove the other direction, let $u \in \mathcal{L}(\mathbf{u})$ have an occurrence $i - L \in E_{\mathbf{u}}$ with interpretation (p, u, ε). Then, (u, ε) is a synchronization point (from hypothesis). For any other interpretation (p', v', s') for some occurrence $j - L$ of u holds that $s' = \varepsilon$. Then $j+1 \in E_{\mathbf{u}}$ which implies left recognizability for $L > 0$. □

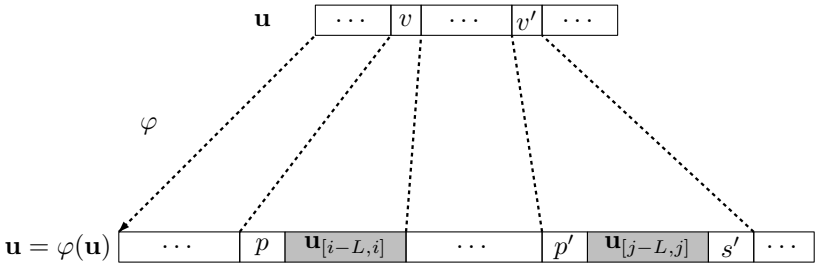


Figure 3.8: Occurrences $i - L$ and $j - L$ of $u \in \mathcal{L}(\mathbf{u})$ from the proof of Proposition 3.3.13.

The following result is similar to the one for the right recognizability given by Mossé in [80] and can be proved in an analogous way.

Proposition 3.3.14. Let φ be a primitive morphism with aperiodic fixed point. The morphism is not right recognizable if and only if for any integer $L > 0$ there is a word w with $|w| = L+1$ and two letters $a, b \in A$ such that,

1. the word $\varphi(b)$ is a strict prefix of the word $\varphi(a)$,
2. the words $w\varphi(a)$ and $w\varphi(b)$ are in the language $\mathcal{L}(\mathbf{u})$ and they have the same decoding.

The following proposition indicates the relation between strongly circularity and left recognizability.

Proposition 3.3.15. If a morphism is strongly circular then it is left recognizable.

Proof. The proof is analogous to the proof of Proposition 3.3.8 □

The inverse of the previous proposition is not always true, as it is indicated in the following example.

Example 3.3.16. Let $\varphi : \{0,1,2\}^* \to \{0,1,2\}^*$ with $\varphi(0) = 0120$, $\varphi(1) = 12$, $\varphi(2) = 12$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. It can be verified that this morphism is left recognizable for some constant $L = 2$. However, this is not strongly circular, from Example 3.3.9.

Remark 3.3.17. If the morphism is suffix and left recognizable then it is strongly circular.

3.3.3 Two-sided recognizability

Definition 3.3.18. [80] Let $\varphi : A^* \to A^*$ be a morphism with a fixed point $\mathbf{u}$. We say that φ is *two-sided recognizable* on $\mathbf{u}$ if there is an integer $L > 0$ such that for every $i > L$ if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$.

Example 3.3.19. Let us work with the Fibonacci morphism φ_F (Example 2.2.2) with fixed point $\mathbf{u} = \varphi_F^\omega(0)$. The words that belong in the language $\mathcal{L}(\mathbf{u})$ cannot have as factor two consecutive 1's nor three consecutive 0's (since that implies that their preimage would have as factor at least two consecutive 1's). Also, if there is a factor 00, there has to be a synchronization point in the position between the two 0's. Moreover, after each appearance of the letter 1 there is a synchronization point (since 1 cannot be anything but the end of an image of a letter). Because of the aboves, it can be seen that any word with length more that three ($L = 1$) has at least one synchronization point (because this word contains 1's or 00's factors) and is uniquely decomposed. Thus, the Fibonacci morphism is two-sided recognizable.

Let us introduce the following equivalence that can be used in order to reform the above definition.

For every $n \geq 1$, we define an equivalence relation on the set of integers at least equal to n as follows,

$$i \sim_{\mathbf{u},n} j \text{ if and only if } \mathbf{u}_{[i-n,i+n]} = \mathbf{u}_{[j-n,j+n]}.$$

Using the above equivalence relation, a morphism φ is said to be *two-sided recognizable* if there is an integer $n \geq 0$ such that for every $i, j \geq n$, whenever $i \sim_{\mathbf{u},n} j$ then $i \in E_{\mathbf{u}}$ if and only if $j \in E_{\mathbf{u}}$.

In other words, for a two-sided recognizable morphism, a long enough word of its substitutive language has synchronization point except maybe of a bounded suffix and prefix.

Proposition 3.3.20. Let φ be a morphism with a fixed point $\mathbf{u} \in A^{\mathbb{N}}$. Then φ is two-sided recognizable on $\mathbf{u}$ for some constant $L > 0$, if and only for any word $u \in \mathcal{L}(\mathbf{u})$ and for any words $u_1, u_2 \in \mathcal{L}(\mathbf{u})$ of length L and $L+1$ respectively with $u_1u_2 \in \mathcal{L}(\mathbf{u})$ a factorization of u such that there is an external interpretation (p, v, s) of u which passes by (u_1, u_2), then (u_1, u_2) is a synchronization point.

Proof. Let φ be a two-sided recognizable morphism on a fixed point $\mathbf{u}$ for some constant $L > 0$. Let $u \in \mathcal{L}(\mathbf{u})$ with $|u| = 2L + 1$ and external interpretation (p, v, s). There is an occurrence $i - L$ of u such that

$$\varphi(v) = pus = p\mathbf{u}_{[i-L,i+L]}s = p\mathbf{u}_{[i-L,i)}\mathbf{u}_{[i,i+L]}s = pu_1u_2s$$

with $u_1 = \mathbf{u}_{[i-L,i)}$, $u_2 = \mathbf{u}_{[i,i+L]}$ and $i \in E_{\mathbf{u}}$. For another external interpretation (p', v', s') of u on $\mathbf{u}$, there is a position $j \neq i$ in $\mathbf{u}$ such that $\varphi(v') = p'\mathbf{u}_{[j-L,j+L]}s' = p'us'$ (See Figure 3.9). Since φ is two-sided recognizable on $\mathbf{u}$ and $i \in E_{\mathbf{u}}$, it follows that $j \in E_{\mathbf{u}}$. Thus $(\mathbf{u}_{[i-L,i)}, \mathbf{u}_{[i,i+L]}) = (u_1, u_2)$ is a synchronization point.

In order to prove the other direction, let $u \in \mathcal{L}(\mathbf{u})$ have an occurrence $i - L$ in $\mathbf{u}$ such that $i \in E_{\mathbf{u}}$ (i.e. $\mathbf{u}_{[i-L,i+L]} = u$). Then, there is external interpretation (p, v, s) of u and the pair $(\mathbf{u}_{[i-L,i)}, \mathbf{u}_{[i,i+L]}) = (u_1, u_2)$ is a synchronization point, from hypothesis. Thus, for any other occurrence $j-L$ of u with external interpretation (p', v', s') (i.e. $\varphi(v') = p'\mathbf{u}_{[j-L,j+L]}s'$ with $\mathbf{u}_{[j-L,j+L]} = u$) holds that $j \in E_{\mathbf{u}}$ which proves two-sided recognizability for $L > 0$. □

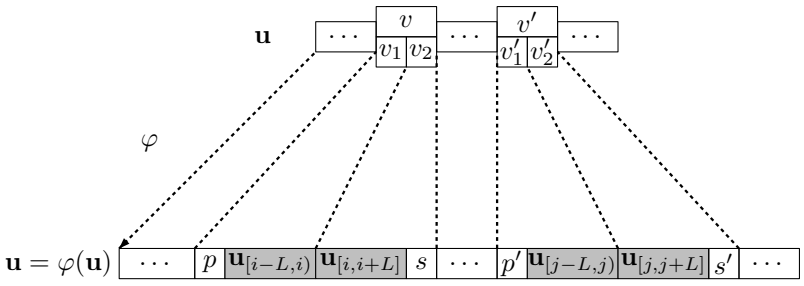


Figure 3.9: Occurrences $i - L$ and $j - L$ of $u \in \mathcal{L}(\mathbf{u})$ from the proof of Proposition 3.3.20.

Proposition 3.3.21. The definition of one-sided recognizability (resp. right, left) is stronger than the one of two-sided recognizability.

Proof. If a morphism is right recognizable (resp. left) for a constant $L > 0$ then for any word $u_{[i-K,i+K]}$ with $K \geq L$, if the position i is in $E_{\mathbf{u}}$ then this position is a synchronization point since for its factor $u_{[i,K]}$ (resp.$u_{[K,i]}$) if the position i is in $E_{\mathbf{u}}$ then it is a synchronization point. □

The following is an example of a morphism that is two-sided recognizable but not left recognizable.

Example 3.3.22. Let us work with the Fibonacci morphism φ_F that is two-sided recognizable according to Example 3.3.19 and right recognizable from Example 3.3.3.

However, Fibonacci morphism is not left recognizable. It can be seen that the image of a right special word extended by the longest common prefix of the images of its extensions is also right-special. Let us use the notation $\mathrm{lcp}(e_1, e_2)$ for the longest common prefix of the images of the words e_1, e_2. If for a right special word $w \in \mathcal{L}(\mathbf{u})$ with extensions $e_1, e_2 \in \mathcal{L}(\mathbf{u})$ we suppose that $\varphi_F(w)\mathrm{lcp}(e_1, e_2)$ is not right special, then it follows that two distinct words we_1, we_2 have the same images, i.e., $\varphi_F(we_1) = \varphi_F(we_2)$. That leads to a contradiction since the Fibonacci morphism is injective. For a right-special word, the last letter is 0 and could have been image of 1 or prefix of the image of 0, which implies that the position after

the 0 can be in $E_{\mathbf{u}}$ or not respectively. Thus, the morphism φ_F is not left recognizable.

Proposition 3.3.23. Every suffix two-sided recognizable morphism is right recognizable.

Proof. Let φ be a suffix two-sided recognizable morphism. Because of the definition of two-sided recognizability there is an integer $L > 0$ such that for every $i, j \in \mathbb{N}$ with $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$. Since the morphism is suffix, if $i \in E_{\mathbf{u}}$, then the decoding before this cutting point is unique. Hence, if $i - L \in E_{\mathbf{u}}$ then $j - L \in E_{\mathbf{u}}$. Consequently, there is an integer $L' = 2L > 0$ for which the morphism is right recognizable. □

Example 3.3.24. The Fibonacci morphism φ_F is suffix and two-sided recognizable (Example 3.3.19). We have that φ_F to be right recognizable (Example 3.3.3) which is in accord with Proposition 3.3.23.

Proposition 3.3.25. Every prefix two-sided recognizable morphism is left recognizable.

The proof is analogous to the proof for Proposition 3.3.23.

We have already seen that two-sided recognizability does not imply one-sided recognizability and as it can be seen in the following example, it is possible for a morphism φ to be two-sided recognizable but neither right recognizable nor left recognizable.

Example 3.3.26. Let us work with the morphism $\varphi : \{0,1\}^* \to \{0,1\}^*$, with $\varphi(0) = 010$, $\varphi(1) = 0$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. The morphism φ is two-sided recognizable with $L = 1$, since for all words $u_{[i-1,i+1]} \in \mathcal{L}(\mathbf{u})$ of length $3 = 2L + 1$ if the position $i \in E_{\mathbf{u}}$ then for any other position $[j-L, j+L]$ in the fixed point that the word may appear, holds that j is also in $E_{\mathbf{u}}$. Indeed, it can be seen that two consecutive 0's cannot appear as a factor of the image of any letter of the language under the morphism φ. Also, the letter 1 is never the first neither the last letter of an image of a letter, and therefore it cannot appear in the beginning or the end of a word. Because of the aboves, for any word $\mathbf{u}_{[i-1,i+1]}$ with $\mathbf{u}_i = 1$ follows that $i \notin E_{\mathbf{u}}$, while if $u_i = 0$, it comes from 01 or 10, then if $\mathbf{u}_{i-1} = 0$ then $i \in E_{\mathbf{u}}$ but if $\mathbf{u}_{i-1} = 1$ then $i \notin E_{\mathbf{u}}$ since it comes from 010. It can be verified easily for all words of length 3, 001, 010, 100 and 000. In order a morphism φ not to be left recognizable, we have to prove

that for all integers L there is long enough word about which it is not decidable whether its last position is in $E_{\mathbf{u}}$ or not. We can see that for morphism φ, the word $u = 0$ can be extended on the right by the words 10 and 0. In the first case, the last position of u is not in $E_{\mathbf{u}}$, while in the second case it is. Let us now take the image of them under the morphism φ (as seen in Figure 3.10). We know that $\varphi(0) = 010$ and the images of its extensions are $\varphi(0) = 010$, $\varphi(10) = 0010$. If we take the longest common prefix of them (the notation lcp is used for the longest common prefix of the images of the words) and add it after of the image of 0, we obtain the word $u' = \varphi(u)\,\mathrm{lcp}(0, 10) = 0100 \in \mathcal{L}(\mathbf{u})$ that can be extended on the right by 10 and 0. In the first case the last position of u' is not in $E_{\mathbf{u}}$ while in the second is. Every time that we repeat the same procedure we got longer words whose last position may be or not in the set $E_{\mathbf{u}}$ (See Figure 3.10). Thus, φ is not left recognizable.

The way to prove that φ is not right recognizable is similar to the one used for the left recognizability. Let us start with the word $u = 0 \in \mathcal{L}(\mathbf{u})$, that can be extended on the left by 01 and 0. In the first case the position on the left of $u = 0$ is in $E_{\mathbf{u}}$, while in the second case it is not in $E_{\mathbf{u}}$. Taking the image of u under φ extended by the longest common suffix of the images of the two possible extensions of its (the notation lcs() is used for the longest common suffix of the images of two words), we have the word $u' = \mathrm{lcs}(01, 0)\varphi(u) = 0010$ that can be extended on the left by 01 or 0 and in the first case the position on the left of u' is not in $E_{\mathbf{u}}$ while in the second case it is. Reapplying the same idea we obtain an arbitrary long word w for which we cannot decide whether the position on the left of the first letter is in $E_{\mathbf{u}}$ or not (See Figure 3.10). Hence, the morphism φ is not right recognizable.

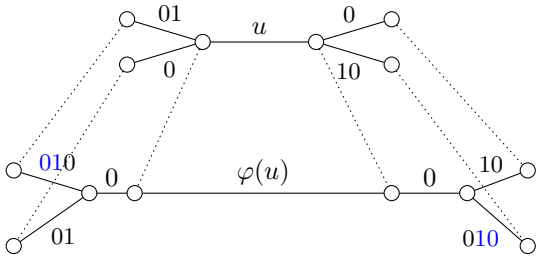


Figure 3.10: A graphiccal explanation of Example 3.3.26.

3.3.4 Strong two-sided recognizability

Let us now introduce the definition of strong recognizability. Some authors use the term recognizability when they refer to the term of strong two-sided recognizability.

Definition 3.3.27. A morphism φ with a fixed point $\mathbf{u}$ is called *strongly two-sided recognizable* if there exists an integer $l > 0$ such that for any word $u \in \mathcal{L}(\mathbf{u})$ of length at least $2l$, there are integers $i, j \in \mathbb{N}$ with $0 \leq i < l$ and $|u| - l \leq j < |u|$ and a unique word $v \in \mathcal{L}(\mathbf{u})$, such that $u_{[i,j)} = \varphi(v)$, and whenever $\mathbf{u}_{[m,m+|u|)} = u$, then there are positions $i', j' \in \mathbb{N}$ with $f_{\mathbf{u}}(i') = m + i$, $f_{\mathbf{u}}(j') = m + j$, $\mathbf{u}_{[i',j')} = v$.

Observation 3.3.28. In other words, if a morphism is strongly two-sided recognizable, it is two-sided recognizable and the preimage of every factor of the fixed point of this morphism is unique except for some bounded suffix and prefix. The two-sided recognizability follows from strong recognizability. Indeed, the fact that $u_{[i,j)} = \varphi(v)$ implies that the decoding of $u_{[i,j)}$ is known and thus if the position $\lfloor |u|/2 \rfloor \in E_{\mathbf{u}}$, then the position is a synchronization point, otherwise it can never be a cutting point whenever the word may appear in the fixed point.

The smallest integer l that satisfies Definition 3.3.27 is called *constant of recognizability of* φ.

Example 3.3.29. Let φ_F be the Fibonacci morphism (Example 2.2.2) with fixed point $\mathbf{u}$ being the Fibonacci word. This morphism is strongly two-sided recognizable with constant of recognizability $l = 2$. That can be checked for all words of length $2l = 4$ in the language $\mathcal{L}(\mathbf{u})$ which are the following ones, 0010, 0100, 0101, 1001 and 1010. For each one of them there are integers i, j and unique word $v \in \mathcal{L}(\mathbf{u})$ that satisfy Definition 3.3.27. Specifically, the words above can be written as $0\varphi_F(0)0$, $01\varphi_F(1)0$, $\varepsilon\varphi_F(0)01$, $1\varphi_F(1)01$ and $1\varphi_F(0)0$ respectively, no matter where they may appear in the fixed point $\mathbf{u}$. Finally, since the definition is satisfied for all words of length $2l$, then it is also satisfied for all the possible extensions of them and therefore for all words with length at least 2ℓ.

Observation 3.3.30. Strongly recognizability does not imply one-sided recognizability.

The following example indicates a case where the previous observation holds.

Example 3.3.31. Let φ_F be the Fibonacci morphism (Example 2.2.2). As it was shown in Example 3.3.29, the Fibonacci morphism is strongly two-sided recognizable, while from Example 3.3.22 it is not left recognizable.

Intuitively someone may wrongly assume that strongly recognizability is the same as injectivity, that is why we present the following remark.

Remark 3.3.32. Strongly recognizability does not imply injectivity.

An example that supports the previous remark is the following one:

Example 3.3.33. Let $\varphi : \{0,1,2\}^* \to \{0,1,2\}^*$ be the primitive morphism with $\varphi(0) = 01002$, $\varphi(1) = 010$ and $\varphi(2) = 010$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. This morphism can be obtained as the third power of the Fibonacci morphism φ_F (Example 2.2.2) if we replace the second appearance of the letter 1 in the word $\varphi_F^3(0) = 01001$ by the letter 2, which is going to have as image $\varphi(2) = \varphi_F^3(1) = 010$. The fixed point $\mathbf{u}$ is almost identical to the Fibonacci word (with some replacements of the letter 1 by 2) and thus it is aperiodic. For $l = 3$ we check all words $w \in \mathcal{L}(\mathbf{u})$ with $|w| = 2l = 6$, to see if they satisfy Definition 3.3.27. The only words of this length in $\mathcal{L}(\mathbf{u})$ are 001002, 002010, 010010, 010020, 020100, 100100, 100201, 201001 and 201002. All these words have synchronization points but they are not long enough in order to have integers i, j that satisfy the conditions of Definition 3.3.27. Indeed, the word $u = 010010 \in \mathcal{L}(\mathbf{u})$ has a synchronizing point in position 3 (in between the two 0's) and has two distinct internal interpretations $(\varepsilon, 1, 010)$ and $(\varepsilon, 2, 010)$. It can be seen that for the integers $i = 0$ and $j = 3$, there are words $v_1 = 1$ and $v_2 = 2$ such that $\varphi(v_1) = u_{[0,3)}$ and $\varphi(v_2) = u_{[0,3)}$ and thus the constant of recognizability cannot be equal 3. However, if the words are long enough we can obtain the wanted integers and the unique preimage of the factor in between the two integers. It can be verified that the morphism is strongly two-sided recognizable with constant of recognizability $l = 11$ but it is not injective. It

can be verified that all words with the needed with length at least $2l + 1 = 23$ verify the definition. An example is the word,

$$w = 01002010020100100201002$$

which can be decomposed uniquely as $w = \varphi(0)\varphi(0)\varphi(1)\varphi(0)\varphi(0)$. For the word w there are integers $i = 10 < l$ and $j = 13 \geq |w| - l = 12$ such that $w_{[i,j)} = \varphi(1)$ and it will be like that whenever w appears in the fixed point. However, if we had chosen a smaller constant, for example $l' = 10$ we have the word

$$w' = 0100201002010010020100$$

that can be decomposed in two different ways, $w' = \varphi(0)\varphi(0)\varphi(1)\varphi(0)010$ and $w' = \varphi(0)\varphi(0)\varphi(2)\varphi(0)\varphi(1)0$ and since the central part is different in the two decompositions, it follows that there cannot be found integers i, j that satisfy the definition of strongly recognizability.

Let us recall the equivalence relation on $\sim_{\mathbf{u},n}$ for $n \geq 1$ by

$$i \sim_{\mathbf{u},n} j \text{ if and only if } \mathbf{u}[i - n, i + n] = \mathbf{u}[j - n, j + n].$$

As a complement, the condition defining the two-sided recognizability can be formulated with the following stronger condition, called asymptotic injectivity: whenever $f_{\mathbf{u}}(i) \sim_{\mathbf{u},n} f_{\mathbf{u}}(j)$ then $\mathbf{u}_i = \mathbf{u}_j$. Such a condition is clearly satisfied when φ is injective.

Example 3.3.34. Let φ_F be the Fibonacci morphism (Example 2.2.2) with fixed point $\mathbf{u}$ the Fibonacci word.

$$\mathbf{u} = 01|0|01|01|0|01|0|01|01|0|01|01|0\ldots$$

The morphism φ_F does not satisfy the asymptotic injectivity for $n = 0$ on $\mathbf{u}$. Indeed, $f_{\mathbf{u}}(1) = 2$ and $f_{\mathbf{u}}(2) = 3$ with $\mathbf{u}_2 = \mathbf{u}_3 = 0$ although $\mathbf{u}_1 = 1$ and $\mathbf{u}_2 = 0$. It is satisfied however for $n = 1$. Indeed, $\mathbf{u}_k = 0$ if and only if $\mathbf{u}_{f(k)+1} = 1$.

Remark 3.3.35. Asymptotic injectivity holds if and only if the morphism is strongly two-sided recognizable.

Indeed, if a morphism φ is strongly two-sided recognizable for a constant of recognizability $l > 0$, then it can be verified that asymptotic injectivity holds for all integers n such that $n \geq l$, since the position that would be examined is in the part of the word that

has unique preimage. It is also easy to verify the other way. If asymptotic injectivity holds for some integer n, then the morphism φ is strongly recognizability for some constant $l > n$.

The following lemma proves that two-sided recognizability, under certain conditions other than asymptotic injectivity may also implies strongly recognizability.

Lemma 3.3.36. If φ is two-sided recognizable morphism and suffix (resp. prefix, injective) then it is strongly two-sided recognizable.

Proof. Suppose that φ is two-sided recognizable for a recognizability constant $L > 0$. Let $u \in \mathcal{L}(\mathbf{u})$ be a word such that $u = \mathbf{u}_{[i,i+2L+2||\varphi||]}$. It can be seen that there is a position n, $i + 1 \leq n < i + L + ||\varphi||$ such that $n \in E_{\mathbf{u}}$ and similarly there is a position $m \in E_{\mathbf{u}}$ with $i + |u| - L - ||u|| \leq m < i - L + ||u||$. Because of the two-sided recognizability, the positions n, m are synchronizing points. If the morphism is prefix (resp. suffix, injective), the preimage of $\mathbf{u}_{[n,m)}$ is unique. Hence, the morphism φ is strongly two-sided recognizable for a constant $\ell \leq L + ||\varphi||$. □

Let us now give the following proposition that provides us with another way of defining strongly two-sided recognizability. The proposition is based on Mossé's definition of two-sided recognizability that was presented earlier in this chapter (Definition 3.5.1).

Proposition 3.3.37. Let $\varphi : A^* \to A^*$ be a morphism with a fixed point $\mathbf{u}$. The morphism φ is *strongly two-sided recognizable* on $\mathbf{u}$ if and only if there is an integer $L > 0$ such that for every $i > L$ if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$ and $\mathbf{u}_{i'} = \mathbf{u}_{j'}$, for $f(i') = i$ and $f(j') = j$.

Proof. Let φ be a strongly two-sided recognizable for some constant $\ell > 0$, then the morphism is two-sided recognizable with unique preimage except of a suffix or prefix (Observation 3.3.28). The existance of a unique preimage for any long enough word (longer that 2ℓ) implies that there is an integer $L > 0$ such that for every $i > L$ if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$ and $\mathbf{u}_{i'} = \mathbf{u}_{j'}$, for $i' = f(i)$ and $j' = f(j)$..

In order to prove the other direction, let φ have a constant $L > 0$ such that for every $i > L$ if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$ and $\mathbf{u}_{i'} = \mathbf{u}_{j'}$, for $i' = f(i)$ and $j' = f(j)$. It can be verified that the morphism φ is strongly two-sided recognizable for a

constant $\ell = L + 2||\varphi||$. Indeed, for any word u with length 2ℓ there is a position $k \in E_{\mathbf{u}}$ in the factor $u_{[L,\ell)}$ and a position $n \in E_{\mathbf{u}}$ in the factor $u_{[|u|-\ell,|u|-L)}$ and there are possibly more positions in $E_{\mathbf{u}}$ in the interval $[k, m]$ of u. For any position i of the interval $[k, m]$ we can apply the hypothesis. Thus, for any such position i, if $i \in E_{\mathbf{u}}$ then the i is a synchronization point and $\mathbf{u}_{i'}$ is unique, for the position $i' = f(i)$. Thus, by concatenation there is unique $v = \mathbf{u}_{[k',m']}$, with $f(k') = k$ and $f(m') = m$, that is such that $\varphi(v) = u_{[k,m]}$. Since the positions k, m are synchronization points, if there is an occurrence $n \in \mathbb{N}$ of the word u in the fixed point $\mathbf{u}$, i.e., if $\mathbf{u}_{[n,n+|u|)} = u$, then there are positions $k_1, m_1 \in E_{\mathbf{u}}$ with $k_1 = n+k$ and $m_1 = n+m$ with $\mathbf{u}_{[k_1,m_1]} = \varphi(v)$, which proves strongly two-sided recognizability. □

Because of the above Proposition, it seems natural that there is some connection between the strong two-sided recognizability and the strong synchronization points, as it is indicated in the following proposition.

Proposition 3.3.38. Let φ be a morphism with a fixed point $\mathbf{u} \in A^{\mathbb{N}}$. Then φ is strongly two-sided recognizable on $\mathbf{u}$ for some constant $L > 0$, as described in Proposition 3.3.37, if and only for any word $u \in \mathcal{L}(\mathbf{u})$ and for any words $u_1, u_2 \in \mathcal{L}(\mathbf{u})$ of length L and $L + 1$ respectively with $u_1u_2 \in \mathcal{L}(\mathbf{u})$ a factorization of u such that there is an external interpretation (p, v, s) of u which passes by (u_1, u_2), then (u_1, u_2) is a strong synchronization point.

Proof. Let φ be a strongly two-sided recognizable morphism. From Observation 3.3.28 we have that φ is two sided recognizable and therefore by Proposition 3.3.20, any word $u \in \mathcal{L}(\mathbf{u})$ of length $2L+1$ with an external interpretation (p, v, s) such that $\varphi(v) = p\mathbf{u}_{[i-L,i+L]}s$ that passes by $(\mathbf{u}_{[i-L,i)}, \mathbf{u}_{[i,i+L]}) = (u_1, u_2)$ has a synchronization point (u_1, u_2). Let k be the integer such that $i = f(k)$. Since the morphism is strongly two-sided recognizable, the letter $\mathbf{u}_k$ is unique, which proves that (u_1, u_2) is a strong synchronization point.

In order to prove the other direction, let $u \in \mathcal{L}(\mathbf{u})$ of length $2L + 1$ with an external interpretation (p, v, s) such that $\varphi(v) = p\mathbf{u}_{[i-L,i+L]}s = pu_1u_2s$ that passes by $(\mathbf{u}_{[i-L,i)}, \mathbf{u}_{[i,i+L]}) = (u_1, u_2)$ and this point is a strong synchronization point. Since any strong synchronization point is always a synchronization point and also with the condition that the letter $\mathbf{u}_{i'}$ that corresponds to the position

$i' = f(i)$ is fixed, it follows from Proposition 3.3.20 that φ is two-sided recognizable. Therefore, for any other occurrence $j - L$ of u, i.e., $\mathbf{u}_{[j-L,j+L]} = u = u_1u_2$, since $i \in E_{\mathbf{u}}$ then $j \in E_{\mathbf{u}}$ and $\mathbf{u}_{j'} = \mathbf{u}_{i'}$, for $j' = f(j)$, which proves strong two-sided recognizability. □

3.4 Mossé's recognizability Theorem

The following theorem is known as B.Mossé's recognizability Theorem and it is a very important result of B. Mossé's [80], because before it the recognizability was supposed.

Theorem 3.4.1. *[80] Every primitive aperiodic morphism is strongly two-sided recognizable.*

We show here the proof given by P. Kůrka [68], which might be more comprehensible than the original one, given by B. Mossé, but let us first define the following equivalence relation.

Let φ be a morphism with a fixed point $\mathbf{u}$. Let $\mathbf{u}_{[i,i+n)} = \mathbf{u}_{[j,j+n)}$. The equivalence relation $\approx$ is defined between two intervals $[i, i+n)$ and $[j, j+n)$ when $f_{\mathbf{u}}(\mathbb{N}) \cap [j, j+n)$ can be obtained by $f_{\mathbf{u}}(\mathbb{N}) \cap [i, i+n)$ by a shift of $j - i$ positions. In other words, $[i, i+n) \approx [j, j+n)$ if the two intervals have cutting points in similar positions.

Example 3.4.2. Let φ_F be the Fibonacci morphism with fixed point $\mathbf{u} = \varphi_F^{\omega}(0) = 01|0|01|01|0|01|0|01|01|0|01|01|0|\ldots$, where $|$ is the notation for the positions in $E = f(\mathbb{N})$. The word $u = 0010 = \mathbf{u}_{[2,5]} = \mathbf{u}_{[7,10]} = \mathbf{u}_{[15,18]}$ and it can be verified that $[2, 5] \not\approx [7, 10]$ and $[2, 5] \approx [15, 18]$.

The following proof is divided in two parts. In the first part it is proved that every primitive aperiodic morphism is recognizable while in the second part it is proved the stronger result that states that under the same conditions the morphism is strongly two-sided recognizable. The fist part is used in order to prove the second part.

Proof. (Theorem 3.4.1)[68] *First part:* The theorem is going to be proved with use of contradiction. Let us fix integer $k > 0$, we shall later chose a large enough value for the integer k, and let us work with a morphism φ, assuming that the contrary of the theorem is true (i.e. the morphism φ is primitive aperiodic and not two-sided recognizable). For a given p, let $l_p = ||\varphi^p||k + 1$. Since φ is not two-sided recognizable, for each integer $p \geq 0$ there exist positions i_p, j_p

such that $i_p \in f(\mathbb{N})$ and $j_p \notin f(\mathbb{N})$ with $\mathbf{u}_{[i_p-l_p,i_p+l_p]} = \mathbf{u}_{[j_p-l_p,j_p+l_p]}$. Because of the size of the constant l_p, every word of length more than $2l_p$ has at least $2k+1$ cutting points, which implies that $\mathbf{u}_{[i_p-l_p,i_p+l_p]}$ has also at least $2k+1$ cutting points. More precisely, every word of the previous form has at least k cutting points before the position i_p and k after. Let r_p, s_p be the smallest integers such that,

$$\left(Card([i_p - r_p, i_p) \cap f^p(\mathbb{N})) = \lfloor \frac{k}{2} \rfloor\right) \wedge \left(Card([i_p, i_p + s_p) \cap f^p(\mathbb{N})) = k - \lfloor \frac{k}{2} \rfloor\right).$$

The choice of the integers r_p, s_p is such that the factor $\mathbf{u}_{[i_p-r_p,i_p+s_p]}$ has k cutting points and it is a factor of $\mathbf{u}_{[i_p-l_p,i_p+l_p]}$. It follows that $i_p - l_p < i_p - r_p$ and $i_p + s_p \leq i_p + l_p$ and $\mathbf{u}_{[i_p-r_p,i_p+s_p]} = \mathbf{u}_{[j_p-r_p,j_p+s_p]}$. Since r_p, s_p are chosen to be the smallest such integers, it follows that $i_p - r_p$ and $i_p + s_p$ are cutting points. Hence, there exists an integer i'_p such that $i_p - r_p = f^p(i'_p)$ and $i_p + s_p = f^p(i'_p + k)$ as in the upper part of Figure 3.11. Since $j_p \notin f(\mathbb{N})$, for $j_p - r_p$ and $j_p + s_p$ there exist integers j'_p, t_p such that,

$$\left(f^p(j'_p) \leq j_p - r_p < f^p(j'_p + 1)\right) \wedge \left(f^p(j'_p + t_p - 1) < j_p - s_p \leq f^p(j'_p + t_p)\right).$$

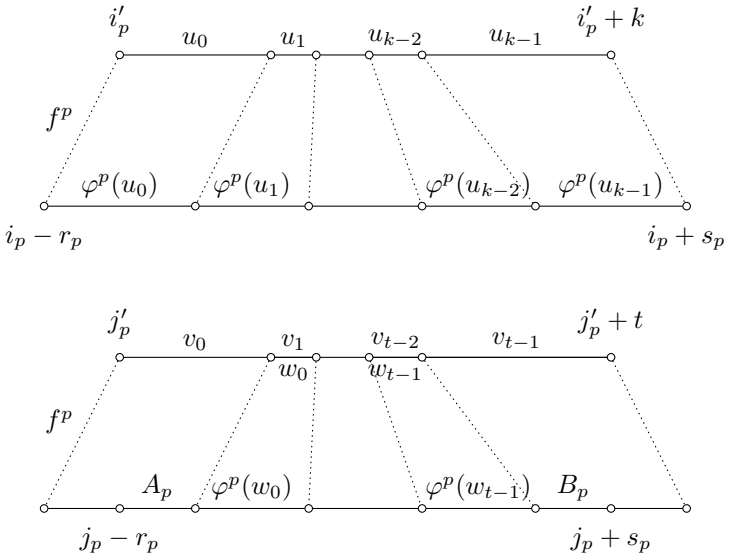


Figure 3.11: Decomposition of $\varphi^p(\mathbf{u})$.

It holds that,

$$|\mathbf{u}_{[j_p-r_p,j_p+s_p)}| = |\mathbf{u}_{[i_p-r_p,i_p+s_p)}| \leq k||\varphi^p||$$

and also,

$$|\mathbf{u}_{[j_p-r_p,j_p+s_p)}| \geq |\mathbf{u}_{[f^p(j'_p+1),f^p(j'_p+t_p-1)]}| \geq t_p|\varphi^p|$$

Thus,

$$t_p|\varphi^p| \leq k||\varphi^p|| \Rightarrow t_p \leq k\frac{||\varphi^p||}{|\varphi^p|}.$$

The above inequality means that the integer t_p is bounded, which implies that there exists $t > 0$, $u \in A^k$, $v \in A^t$ and an infinite set $E' \subseteq \mathbb{N}$ such that for any $p \in E', t_p = t, \mathbf{u}_{[i'_p,i'_p+k)} = u$ and $\mathbf{u}_{[j'_p,j'_p+t)} = v$ (See Figure 3.11 bottom). Let $w = v_{[1,t-2]}$. For any $p \in E'$ there exists A_p suffix of $\varphi^p(v_0)$ and B_p prefix of $\varphi^p(v_{t-1})$ such that,

$$\varphi^p(u) = \mathbf{u}_{[i_p-r_p,i_p+s_p]} = \mathbf{u}_{[j_p-r_p,j_p+s_p]} = A_p\varphi^p(w)B_p.$$

as in Figure 3.12. For $q \in E_{\mathbf{u}}$ such that $q > p$ we have,

$$A_q\varphi^q(w)B_q = \varphi^q(u) = \varphi^{q-p}(\varphi^p(u)) = \varphi^{q-p}(A_p)\varphi^q(w)\varphi^{q-p}(B_p).$$

Let us suppose that $A_q \neq \varphi^{q-p}(A_p)$, hence one of them has to be prefix of the other. Without loss of generality let us set $\varphi^{q-p}(A_p) = A_qz$, which implies that $|z|$ is a period of $\varphi^q(w)$. Thus, z^n is prefix of $\varphi^q(w)$ for n integer such that $n|z| \leq |\varphi^q(w)|$. For a k large enough , since $|A_q| + |B_q| + |w| \geq k$ and thus $|w| \geq k - |A_q| - |B_q|$, we can find a large enough integer n that contradicts Theorem 2.3.18. Thus, there are $p, q \in E'$ with $q > p$ such that $A_q = \varphi^{q-p}(A_p)$ and $B_q = \varphi^{q-p}(B_p)$ as in Figure 3.12.

Let m_p be the position in $\mathbf{u}$ such that $m_p - r_p - f^p(j'_p) = j_p - r - p - f^p(j'_p)$. Then, from $A_q = \varphi^{q-p}(A_q)$ and $B_q = \varphi^{q-p}(B_q)$ follows that,

$$\mathbf{u}_{[m_p-r_p,m_p+s_p)} = \mathbf{u}_{[j_p-r_p,j_p+s_p)} = A_p\varphi^p(w)B_p = \varphi^p(u) = \mathbf{u}_{[f^p(i'_p),f^p(i'_p+k))}.$$

Finally, applying φ^{q-p} we obtain,

$$\begin{aligned}[j_q - r_q, j_q + s_p) = [f^{q-p}(m_q - r_p), f^{q-p}(m_q + s_p)) &\approx [f^q(i'_p), f^q(i'_q + k)) \\ &= [i_q - r_q, i_q + s_q)\end{aligned}$$

which leads to a contradiction since $j_p \notin f(\mathbb{N})$ by assumption.

Second part: We define the following equivalence on the letters of an alphabet. Two letters $a, b \in A$ are equivalent ($a \sim b$) if and

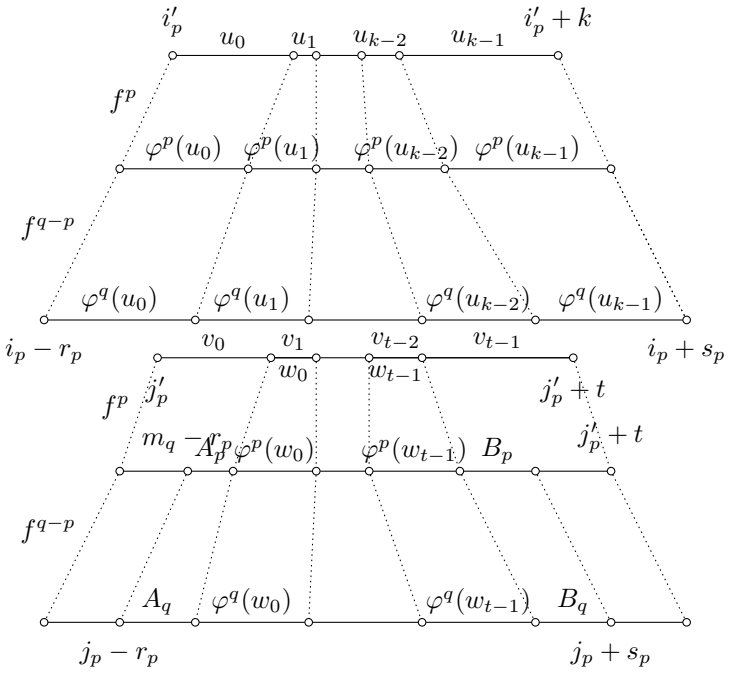


Figure 3.12: Decomposition of $\varphi^p(\mathbf{u})$.

only if there is an integer $n \in \mathbb{N}$ such that $\varphi^n(a) = \varphi^n(b)$. Since we work on a finite alphabet A, there is a maximal integer p such that two letters of the alphabet A will be equivalent (i.e if two letters are not equivalent up to the integer p, then there is no need to check larger integers cause we are sure that they are not equivalent). Thus, if $\varphi^p(a) \neq \varphi^p(b)$, then for all $k \geq 0$, $\varphi^k(a) \neq \varphi^k(b)$. From Proposition 2.3.6 we have that if φ is primitive, so is any power φ^i for $i \in \mathbb{N}$ of his. Also, since φ is primitive, the language $\mathcal{L}(\varphi)$ is the same with the language of all fixed points of φ and the ones of the powers of his. That being the case, the morphism φ^{p+1} is primitive with aperiodic fixed point. Applying Theorem 3.4.1 to φ^{p+1}, we have that two-sided recognizability holds, i.e., there exists an integer l_p such that for any integers $r, q > l_p$, if $\mathbf{u}_{[q-l_p,q+l_p]} = \mathbf{u}_{[r-l_p,r+l_p]}$ and $q \in f^{p+1}(\mathbb{N})$, then $r \in f^{p+1}(\mathbb{N})$. Let us set $l = l_p + ||\varphi^{p+1}||$. For $u \in \mathcal{L}(\mathbf{u})$, $|u| \geq 2l$ we are sure that there is at least a cutting point between the positions l_p and l (including the position l_p). Hence,

there is unique integer k and unique numbers $n_0, n_1, \dots n_k$ with,

$$l_p \leq n_0 < n_1 < \cdots < n_k \leq |u| - l_p,$$

such that for any $\mathbf{u}_{[q,q+|u|]} = u$,

$$[q + l_p, q + |u| - l_p] \cap f^{p+1}(\mathbb{N}) = \{q + n_0, q + n_1, \dots, q + n_k\}$$

as it can be seen in Figure 3.13. Hence, there is word $w \in A^k \cap \mathcal{L}(\mathbf{u})$ such that $\mathbf{u}_{[n_t,n_{t+1}]} = \varphi^{p+1}(w_t)$ for $0 \leq t < k$. Let us suppose that there is $w' \in A^k \cap \mathcal{L}(\mathbf{u})$ such that $\varphi^{p+1}(w) = \varphi^{p+1}(w')$. Since $p + 1 > p$ we deduce that $\varphi^p(w) = \varphi^p(w') = v$ that is unique. If we set $n_0 = i$ and $n_k = j$ we have that $i \leq l$ and $j \geq |u| - l$ and the word v id unique, which implies that the morphism is strongly two-sided recognizable, which proves the theorem. □

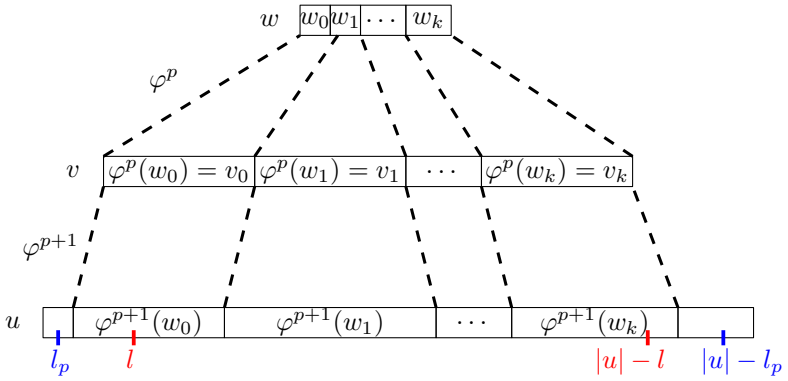


Figure 3.13: A graphical explanation of the proof of Theorem 3.4.1.

We should also mention that aperiodicity is an important condition in the above theorem, since there are primitive periodic morphisms that are not two-sided recognizable, as it can be seen in the following example.

Example 3.4.3. The morphism φ with $\varphi(0) = 010$, $\varphi(1) = 101$ and fixed point $\mathbf{u} = \varphi^\omega(0)$ is periodic with period $p = 2$ and primitive but not two-sided recognizable. For every $m \in \mathbb{N}$ the words $(10101010)^{2m}$ belong in $\mathcal{L}(\mathbf{u})$. However, there are two different external interpretations for each one of them, $(\varepsilon, \varphi((10)^m), 10)$ and $(1, \varphi((01)^m), 0)$, which implies that we can always find long enough word $u_{[i-L,i+L]} = u_{[j-L,j+L]}$ such that $i \in E_\mathbf{u}$ and $j \notin E_\mathbf{u}$.

Example 3.4.4. Let φ be the morphism with $\varphi(0) = 012$, $\varphi(1) = 0$, $\varphi(2) = 22$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. This morphism is non-primitive and it has aperiodic fixed point,

$$\mathbf{u} = 012|0|22|012|22|22|012|0|22|\dots$$

The morphism is not periodic since at every application of the morphism φ, the number of 2's increases. All words of the form $2(22)^k \in \mathcal{L}(\mathbf{u})$ with $k \in \mathbb{N}$ have two distinct external interpretations $(\varepsilon, 2^{k+1}, 2)$ and $(01, 02^k, \varepsilon)$. Let us suppose that the two interpretations correspond to the factors $\mathbf{u}_{[i-k,i+k]} = \mathbf{u}_{[j-k,j+k]}$, for which it can be seen that $i \in E_\mathbf{u}$ and $j \notin E_\mathbf{u}$, which implies that φ is not two-sided recognizable.

Theorem 3.4.1 does not give any information on the constant of recognizability. When we know that such an integer exists, it can be computed by checking all positive integers one by one until we reach it. For any integer l it is enough to verify if strongly recognizability holds for all the words of length $2l + ||\varphi||$, since if it is verified for them that means that it is for any longer word as well. The next result by F. Durand and J. Leroy gives a theoretical bound on that constant.

Theorem 3.4.5. *[40] Let $\varphi : A^* \to A^*$ be an aperiodic primitive morphism. The constant of recognizability of φ is bounded by*

$$2||\varphi||^{6(|A|)^2+6(|A|)||\varphi||^{28(|A|)^2}} + ||\varphi||^{|A|}.$$

According to the authors of [40], that bound could have been improved but the result would not have been optimal, while the difficulty would have been increased rapidly.

3.5 Recognizability on two-sided infinite fixed points

Everything that has been presented so far for the right infinite fixed points holds for the left fixed points. We should now present the recognizability in the case of two-sided infinite fixed points.

Definition 3.5.1. Let φ be a primitive morphism with admissible two-sided infinite fixed point $\mathbf{z}$. The morphism is *two-sided recognizable* if there is an integer $L > 0$ such that if $\mathbf{z}_{[i-L,i+L]} = \mathbf{z}_{[j-L,j+L]}$ and $i \in E_\mathbf{u}$, then $j \in E_\mathbf{u}$, where $E_\mathbf{u} = f_\mathbf{z}(\mathbb{Z})$.

The above definition is the same as the one for the one-sided infinite case (Definition 3.3.18), with the only difference that in this case the $i, j \in \mathbb{Z}$ are relative integers.

Example 3.5.2. Let φ_F be the Fibonacci morphism (Example 2.2.2) and φ^2 be the square Fibonacci morphism. Since $\varphi^2(0) = 010 \in 0A^*$ and $\varphi^2(1) = 01 \in A^*1$ it follows that $\mathbf{z} = (\varphi^2)^{\tilde{\omega}}(1) \bullet (\varphi^2)^{\omega}(0)$ is a two-sided fixed point,

$$\mathbf{z} = \cdots 0101 \bullet 01001 \cdots$$

such that $\mathbf{z}_{[-1,0]} = 10$. The morphism φ^2 is two-sided recognizable with $L = 1$. That can be seen by checking all words of length 3 in $\mathcal{L}(\mathbf{z})$ which are 001, 010, 100, 101. It can be verified that for all words $\mathbf{z}_{[i-1,i+1]}$ holds that $i \in E_{\mathbf{u}}$ (i.e. the position i is a cutting point) only in the cases $\mathbf{z}_{[i-1,i+1]} = 001$ and 101, while for the rest of the words of that length it holds that $i \notin E_{\mathbf{u}}$. Since it is decidable whether $i \in E_{\mathbf{u}}$ for all words $\mathbf{z}_{[i-1,i+1]}$, it follows that it is decidable for any larger words too.

Proposition 3.5.3. Let φ be a morphism with a admissible two-sided infinite fixed point $\mathbf{z} \in A^{\mathbb{Z}}$. Then φ is two-sided recognizable on $\mathbf{z}$ for some constant L, if and only if any word $u \in \mathcal{L}(\mathbf{z})$ of length $2L + 1$ with an external interpretation (p, v, s) such that $\varphi(v) = p\mathbf{z}_{[i-L,i+L]}s$ that passes by $(\mathbf{z}_{[i-L,i)}, \mathbf{z}_{[i,i+L]})$ has a synchronization point $(\mathbf{z}_{[i-L,i)}, \mathbf{z}_{[i,i+L]})$.

Proof. Let us suppose that φ is two-sided recognizable for the admissible two-sided infinite fixed point $\mathbf{z} \in A^{\mathbb{Z}}$. If $u \in \mathcal{L}(\mathbf{z})$ with $|u| = 2L + 1$ has external interpretation (p, v, s) that passes by $(u_{[0,\frac{|u|-1}{2})}, u_{[\frac{|u|-1}{2},|u|-1]})$, then v is such that $\varphi(v) = p\mathbf{z}_{[i-L,i+L]}s$ with $i \in E_{\mathbf{z}}$, for $i \in \mathbb{Z}$, as in Figure 3.14. Let us consider another interpretation (p', v', s') of u. Then there is a $j \in \mathbb{Z}$ such that $\varphi(v') = p'\mathbf{z}_{[j-L,j+L]}s' = p'us'$ as in Figure 3.14. Since φ is two-sided recognizable and $i \in E_{\mathbf{z}}$ for (p, v, s), then $j \in E_{\mathbf{z}}$. Thus, the interpretation (p', v', s') passes by $(\mathbf{z}_{[j-L,j)}, \mathbf{z}_{[j,j+L])}$ which proves that $(u_{[0,\frac{|u|-1}{2})}, u_{[\frac{|u|-1}{2}]})$ is a synchronization point of $|u|$.

In order to prove the other direction, let $u \in \mathcal{L}(\mathbf{z})$ with $|u| = 2L+1$ and external interpretation (p, v, s) with $(u_{[0,\frac{|u|-1}{2})}, u_{[\frac{|u|-1}{2},|u|-1)})$ a synchronization point of u in the language $\mathcal{L}(\mathbf{z})$, i.e., there is $i \in E_{\mathbf{z}}$ such that $\varphi(v) = p\mathbf{z}_{[i-L,i+L]}s = pus$. Then, since $(u_{[0,\frac{|u|-1}{2})}, u_{[\frac{|u|-1}{2},|u|-1)})$

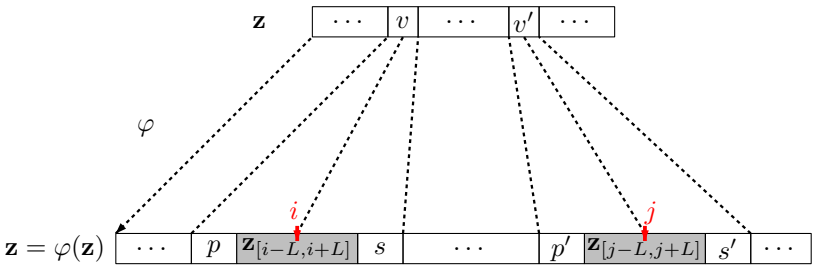


Figure 3.14: Occurrences i and j of $u \in \mathcal{L}(\mathbf{u})$ from the proof of Proposition 3.3.4.

a synchronization point of u, it follows that any other interpretation (p', v', s') of u passes by $(p'u_{[0,\frac{|u|-1}{2})}, u_{[\frac{|u|-1}{2},|u|-1)}s')$ and therefore if $j \in \mathbb{Z}$ is such that $\varphi(v') = p'\mathbf{z}_{[i-L,i+L]}s'$ then $j \in E_{\mathbf{z}}$ which proves two-sided recognizability. □

The *asymptotic injectivity* is defined similarly to the one-side infinite case: if there is integer $n > 0$ such that whenever $f_{\mathbf{u}}(k) \sim_{\mathbf{u},n} f_{\mathbf{u}}(\ell)$ then $\mathbf{u}_k = \mathbf{u}_\ell$ (where for an integer i, $\mathbf{u}_i$ is the letter $\mathbf{u}_{[i,i]}$). It implies that φ is injective on $\mathcal{L}(\mathbf{z})$, as it follows from Remark 3.5.7. Such a condition is clearly satisfied when φ is injective.

Example 3.5.4. Let φ be the Fibonacci morphism (Example 2.2.2). For $n = 0$ the morphism φ does not satisfy the asymptotic injectivity as it can be seen from the words $\mathbf{u}_2 = \mathbf{u}_3 = 0$. Since $f_{\mathbf{u}}(1) = 2$ and $f_{\mathbf{u}}(2) = 3$, the letters $\mathbf{u}_1 = 1$, $\mathbf{u}_2 = 0$ are prefixes of the images of the letters $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively. The asymptotic injectivity is satisfied for the integer $n = 1$ since $\mathbf{u}_k = 0$ if and only if $\mathbf{u}_{f(k)+1} = 1$ and $\mathbf{u}_k = 1$ if and only if $\mathbf{u}_{f(k)+1} = 0$ for any $k \in \mathbb{N}$.

Example 3.5.5. Let us work with the square of Fibonacci morphism φ^2, presented in Example 3.5.2, with fixed point the follows,

$$\mathbf{z} = (\varphi^2)^{\tilde{\omega}}(1)\bullet(\varphi^2)^{\omega}(0) = \ldots|010|01|010|010|01\bullet010|01|010|010|01|010|\ldots.$$

This morphism has asymptotic injectivity which requires $n = 3$. Indeed, for $n = 2$ asymptotic injectivity does not hold because of the word $10010 \in \mathcal{L}(\mathbf{z})$. More precisely, $f_{\mathbf{z}}(3) \sim_{\mathbf{z},2} f_{\mathbf{z}}(4)$, since $f(3) =$

$|\varphi^2(010)| = |01001010| = 8$, $f(4) = |\varphi^2(0100)| = |01001010010| = 11$ and

$$\mathbf{z}_{[f(3)-2,f(3)+2]} = \mathbf{z}_{[6,10]} = 10010 = \mathbf{z}_{[9,13]} = \mathbf{z}_{[f(4)-2,f(4)+2]}.$$

However, it can be seen that $\mathbf{z}_3 = 0$ while $\mathbf{z}_4 = 1$, which implies that asymptotic injectivity does not hold for $n = 2$.

Let us check if it holds for $n = 3$. Let $f_{\mathbf{z}}(k) \sim_{\mathbf{z},3} f_{\mathbf{z}}(\ell)$, i.e., $\mathbf{z}_{[f(k)-3,f(k)+3]} = \mathbf{z}_{[f(\ell)-3,f(\ell)+3]}$. All words in the language of length 7 with a cutting point in the third position are, 0100100, 0100101 and 0010100. Each of them has a unique external interpretation $(\varepsilon, 001, 1)$, $(\varepsilon, 010, 0)$ and $(01, 0100, 10)$ respectively, and therefore each of them has a unique preimage which implies that $\mathbf{z} = \ell$ and that asymptotic injectivity holds.

The following is a variant of Mossé's Theorem 3.4.1.

Theorem 3.5.6. *A primitive aperiodic substitution φ is recognizable on each of its two-sided admissible fixed-points.*

Proof. Let φ be primitive and non-periodic with two-sided infinite fixed point $\mathbf{z}$. From Mossé's Theorem 3.4.1 we know that this morphism is two-sided recognizable with some constant $L \geq 0$ for the one-sided infinite fixed point $\mathbf{z}^+ = z_0 z_1 \cdots$. We will show that φ is recognizable on the substitutive language of the two-sided infinite fixed point $\mathbf{z}$ and moreover that the recognizability constant is the same.

Let $i, j \in \mathbb{Z}$ be such that the equivalence $i \sim_{\mathbf{z},L} j$ holds. Without any loss of generality we may choose $i < j$. Let $k < \ell$ be an integer such that $f_{\mathbf{z}}(k) \leq i - L < j + L < f_{\mathbf{z}}(\ell)$ (see Figure 3.15).

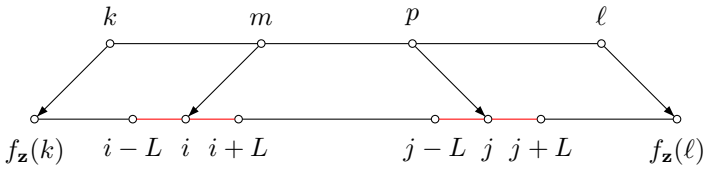


Figure 3.15: Locating k, m, p, ℓ.

Suppose that $i \in E_{\mathbf{z}}(\varphi)$. Let $m \in \mathbb{Z}$ be such that $f_{\mathbf{z}}(m) = i$. Then

$$\begin{aligned} f_z(m+t) &\sim_{z,L} f_z(m) = i \\ &\sim_{z,L} j \\ &\sim_{z,L} j - i + f(m+t) \end{aligned}$$

where the first and the last equivalence follow from the fact that $\mathbf{z}_{[k+t,\ell+t)} = \mathbf{z}_{[k,\ell)}$ and thus, $\mathbf{z}_{[f_{\mathbf{z}}(k+t),f_{\mathbf{z}}(\ell+t))} = \mathbf{z}_{[f_{\mathbf{z}}(k),f_{\mathbf{z}}(\ell))}$.

Since $f_z(m+t) \sim_{z^+,L} j - i + f_z(m+t)$ and $f_z(m+t) \in E_{\mathbf{z}}$ (from definition of $E_{\mathbf{z}}$), it follows from two-sided recognizability on z^+ that $j - i + f_z(m+t) \in E_{\mathbf{z}}$, i.e there is $q \geq 0$ such that $f_{z^+}(q) = f_z(q) = j - i + f_z(m+t)$. Set $p = q - t$. Then,

$$\begin{aligned} f_{\mathbf{z}}(p+t) - f_{\mathbf{z}}(m+t) &= |\varphi(\mathbf{z}_{[m+t,p+t)})| \\ &= |\varphi(\mathbf{z}_{[m,p)})| = f_{\mathbf{z}}(p) - f_{\mathbf{z}}(m). \end{aligned}$$

Since $f_{\mathbf{z}}(m) = i$, we conclude that $f_{\mathbf{z}}(p) = f_{\mathbf{z}}(m) + j - i = j$ and thus that $j \in E_{\mathbf{z}}(\varphi)$. The proof that $j \in E_{\mathbf{z}}(\varphi)$ implies $i \in E_{\mathbf{z}}(\varphi)$ is entirely analogous. □

Remark 3.5.7. Two-sided recognizability in the case of a substitutive language of an admissible two-sided infinite word $\mathbf{z} \in A^{\mathbb{Z}}$ with asymptotic injectivity implies injectivity.

Proof. Let φ be a primitive morphism with admissible two-sided fixed point $\mathbf{z}$. Let $X = \cup_{i\in\mathbb{Z}}\varphi^i(\mathbf{z})$ be a shift space generated by the fixed point $\mathbf{z}$ and $x \in X$ be a bi-infinite word of the shift space.

Let $x, y \in X$ be such that $\varphi(x) = \varphi(y)$. From Theorem 3.5.6, the morphism is two-sided recognizable for a constant $L > 0$ (with respect to asymptotic injectivity). For the admissible fixed point $\mathbf{z}$, and for $i \in \mathbb{Z}$, there exist $k, \ell \in \mathbb{Z}$ such that $\mathbf{z}_{[k-L,k+L]} = x_{[i-L,i+L]}$ and $\mathbf{z}_{[\ell-L,\ell+L]} = y_{[i-L,i+L]}$. Then, by asymptotic injectivity, follows that $\mathbf{z}_\ell = \mathbf{z}_k$ and thus $x_i = x_j$. □

3.6 Synchronizing delay and Circular morphisms

3.6.1 Synchronizing delay

Let us now return to the case of right infinite fixed points. The following definition was given by F. Mignosi and P. Séébold in [79].

Definition 3.6.1. [79] Let $\varphi : A^* \to A^*$ be a morphism with a fixed point $\mathbf{u}$. We say that φ has *synchronizing delay* $D > 0$ if the following property holds: for all words $u \in \mathcal{L}(\mathbf{u})$, if u admits two distinct internal interpretations, say

$$(p, v, s) \text{ and } (p', v', s')$$

where $v = v_1 \dots v_n$, $v' = v'_1 \dots v'_m$ with $v_i, v'_i \in A$ and $v, v' \in \mathcal{L}(\mathbf{u})$, then whenever $|p\varphi(v_1) \dots \varphi(v_{i-1})| > D$ and $|\varphi(v_{i+1}) \dots \varphi(v_n)s| > D$, there exists a number $j \in \mathbb{N}$ such that $p\varphi(v_1) \dots \varphi(v_i) = p'\varphi(v'_1) \dots \varphi(v'_j)$ and $v_i = v'_j$.
That being the case, the two decompositions are synchronized at distance D from the borders. See Figure 3.16 for a graphical representation of the above definition.

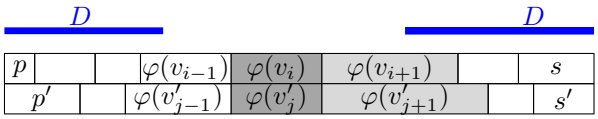


Figure 3.16: A graphical representation of Definition 3.6.1.

Remark 3.6.2. In order to check if a morphism has a synchronizing delay D, it is enough to check all words of length smaller than $2D+2||\varphi||+2$. The $2||\varphi||$ is needed since every factor of that length has at least two cutting points (i.e. there are two positions in $f(\mathbb{N})$), while $2D+2$ is needed so that the length of what is before and what is after the fixed points will be longer than D. To be more precise, it is enough to check the words with lenght n that is,

$$2D + |\varphi| + 2 \leq n \leq 2D + 2||\varphi|| + 2.$$

The reason why there is a lower bound is that if a word is shorter than that bound, then the conditions of the definition cannot be applied.

The following proposition gives some relation between the notion of synchronizing delay and the one of strongly circularity.

Proposition 3.6.3. Any bifix morphism with finite synchronizing delay is strongly circular.

Proof. If a morphism φ has synchronizing delay D then we know that any word $v \in \mathcal{L}(\mathbf{u})$ with length more than $2D + 2||\varphi|| + 2$ has a synchronization point. Also, since it is bifix, the decoding of v before and after the synchronization point has to be unique and therefore there is unique internal interpretation for v. Hence, φ is strongly circular. □

3.6.2 Circular morphisms

Definition 3.6.4. Let $\varphi : A^* \to A^*$ be a morphism and $\mathbf{u} \in A^{\mathbb{N}}$ its fixed point. We say that two external interpretations (p, v, s) and (p', v', s') of a word $u \in \mathcal{L}(\mathbf{u})$ are *synchronized at position* k if there exist indices i and j such that

$$\varphi(v_0 \dots v_i) = pu_0 \dots u_k \text{ and } \varphi(v'_0 \dots v'_j) = p'u_0 \dots u_k$$

with $v = v_0 \dots v_n$, $v' = v'_0 \dots v'_m$, $0 \leq i \leq n$ and $0 \leq j \leq m$.

Example 3.6.5. Let $\varphi : \{0,1\}^* \to \{0,1\}^*$ with $\varphi(0) = 001$, $\varphi(1) = 01$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. The word $u = 010010 \in \mathcal{L}(\mathbf{u})$ has three different external interpretations, $(\varepsilon, 100, 01)$, $(\varepsilon, 101, 1)$, $(0, 001, 1)$ and all of them are synchronized at positions 2 and 5.

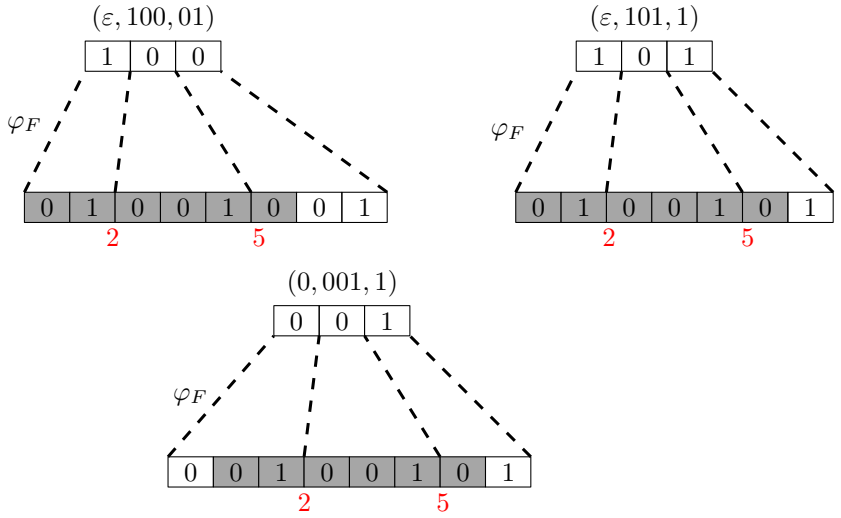


Figure 3.17: External interpretations of $u = 010010$ synchronized at positions 2 and 5 of u.

Remark 3.6.6. If all external interpretations of a word $u \in \mathcal{L}(\mathbf{u})$ are synchronized at position k, with $0 \leq k \leq |u|$, then the position k is a synchronizing point of u.

The inverse is also true. More precisely, if a word has a synchronizing point, then all its external interpretations, by definition, pass from this point and therefore they are synchronized at the position of the synchronizing point.

Definition 3.6.7. [65] Let $\varphi : A \to A^*$ be a morphism and $\mathbf{u} \in A^{\mathbb{N}}$ a fixed point of φ. The morphism φ is called *weakly circular* if there is a constant $D > 0$ such that any v in the language $\mathcal{L}(\mathbf{u})$ longer than D has a synchronization point.

Example 3.6.8. The Fibonacci morphism φ_F (Example 2.2.2) is weakly circular for $D = 1$, since all words with length at least 2 have as factor the letter 1 at least once or they have two consecutive 0's. In the first case, the position after the appearance of 1 is a synchronization point. In the second case, the position in between the two 0's is a synchronization point. Hence, all external interpretations are pairwise synchronized at the synchronization points.

It can be proved that the definition of two-sided recognizability is stronger than the one of weakly circularity, as follows later from Theorem 3.6.9.

Theorem 3.6.9. *Any two-sided recognizable morphism is weakly circular.*

Proof. Let us suppose that φ is two-sided recognizable for a constant $L > 0$. In order to obtain weakly circularity we have to ensure the existence of a synchronizing point. Let us set $D = L + ||\varphi||$ and we will prove that weakly circularity holds for this constant. For any $i > D$ position in $\mathbf{u}$ such that $\mathbf{u}_{[i-D,i+D]} \in \mathcal{L}(\mathbf{u})$ it can be seen that the factor $\mathbf{u}_{[i-||\varphi||,i+||\varphi||]}$ of $\mathbf{u}_{[i-D,i+D]}$ has a position $k \in E_{\mathbf{u}}$ with $i - ||\varphi|| \leq k \leq i + ||\varphi||$. Also, since φ is two-sided recognizable with constant L, it holds that if there is an integer j such that $\mathbf{u}_{[j-L,j+L]} = \mathbf{u}_{[k-L,k+L]}$, then $j \in E_{\mathbf{u}}$. It follows that the position k of the word $u_{[k-L,k+L]}$ is a synchronization point and since $u_{[k-L,k+L]}$ is factor of $u_{[i-D,i+D]}$, it is also a synchronization point of his. Hence, for $D = L+||\varphi||$, every word $\mathbf{u}_{[i-D,i+D]}$ has at least a synchronization point. □

Proposition 3.6.10. An injective morphism is weakly circular if and only if is two-sided recognizable.

Proof. Let φ be an injective weakly circular morphism for $D > 0$ with a fixed point $\mathbf{u}$. That being the case, every $w = \mathbf{u}_{[i-D,i+D]} \in \mathcal{L}(\mathbf{u})$ has at least two synchronizing points at positions k and l, $i - D \leq k < i$ and $i < l \leq i + D$. The factor $\mathbf{u}_{[k,l)}$ of w has unique preimage because of the injectivity and therefore for any

position in the interval $[k, l)$, including the position i, if $i \in E_{\mathbf{u}}$ then it is a synchronization point. More precisely, if there is $j \in \mathbb{N}$ such that $\mathbf{u}_{[j-D,j+D]} = \mathbf{u}_{[i-D,i+D]}$ with the above reasoning there are positions k', l' with $k - (i - D) = k' - (j - D)$ and $i + D - l = j + D - l'$ synchronization points such that $\mathbf{u}_{[k',l')} = \mathbf{u}_{[k,l)}$ and because of injectivity the preimage is unique and thus if $i \in E_{\mathbf{u}}$ then $j \in E_{\mathbf{u}}$. Hence, the morphism φ is two-sided recognizable for a constant $L = D$. The other side follows from Theorem 3.6.9. □

The above proposition is not true for the non-injective case as it is demonstrated in Example 3.6.11.

Example 3.6.11. Consider the morphism $\varphi : 0 \to 012, 1 \to 201, 2 \to 2$ with fixed point $\mathbf{u} = \varphi^{\omega}(0)$. The morphism is non-injective since $\varphi(20) = \varphi(12)$. The fixed point is periodic with period $p = 4$. This morphism is weakly circular for $D = 4$, since every word $v \in \mathcal{L}(\mathbf{u})$ longer than D has as factor the word 22 which has a synchronizing point between the two 2 (2 is always either the beginning or the end of an image of a letter and there is no letter $a \in A$ such that 22 would have been a factor of $\varphi(a)$). The above can be verified from the words of length 5. However, the morphism φ is not two-sided recognizable. Indeed, for $L = 3$ we can find a word $v = 1220122 \in \mathcal{L}(\mathbf{u})$ that has two different internal interpretations $\mathbf{u}_{[i-3,i+3]} = (1, 220, 2)$ and $\mathbf{u}_{[j-3,j+3]} = (12, 122, \varepsilon)$. For the first one, $\mathbf{u}_i = 0$ with $i \in E_{\mathbf{u}}$, while for the second one $\mathbf{u}_j = 0$ with $j \notin E_{\mathbf{u}}$. We have that $|v| = 7$ and since the word starts with the letter 1, the only possible extension on the left by a letter is 0. Since $|0v| = 8 = 2p$, it follows that $(0v)^* \in \mathcal{L}(\mathbf{u})$. Similarly, all words of the form $v(0v)^{3n-1} \in \mathcal{L}(\mathbf{u})$, for $n \in \mathbb{N}$ we have two internal interpretations $(1, (2201)^{n-1}220, 2)$, and $(12, (1220)^{n-1}122, \varepsilon)$, where $\mathbf{u}_i = 0$ with $i \in E_{\mathbf{u}}$ but $\mathbf{u}_j = 0$ with $j \notin E_{\mathbf{u}}$. It has just been proved that all words of the previous form (these words have length $7n$, $n \in \mathbb{N}$) do not satisfy the definition of two-sided recognizability and therefore $L \neq 3n$ for all $n \in \mathbb{N}$. Thus, the morphism is not two-side recognizable since there can be found arbitrary long words $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ for which the positions i and j are not simultaneously in $E_{\mathbf{u}}$ or not in $E_{\mathbf{u}}$.

In the definition of weakly circularity, any word long enough has at least a synchronization point. In the following definition of circularity, any long enough word has to verify the two following conditions:

- it should have at least a synchronization point,
- it has a unique preimage, except of some prefix and/or suffix.

Let us now present the definition of circularity,

Definition 3.6.12. [65] Let $\varphi : A^* \to A^*$ be a morphism with fixed point $\mathbf{u} \in A^{\mathbb{N}}$. Let (p, v, s) and (p', v', s') be two external interpretations of a non-empty word $u \in \mathcal{L}(\mathbf{u})$ with $v = v_0 v_1 \dots v_n$ and $v' = v'_0 v'_1 \dots v'_m$.
We say that φ is *circular* with delay $D > 0$ if whenever we have

$$|\varphi(v_0 \dots v_i)| - |p| > D \text{ and } |\varphi(v_{i+1} \dots v_n)| - |s| > D$$

for some $0 \leq i \leq n$, then there is $0 \leq j \leq m$ such that

$$|\varphi(v_0 \dots v_{i-1})| - |p| = |\varphi(v'_0 \dots v'_{j-1})| - |p'|$$

and $v_i = v'_j$.

(See Figure 3.18)

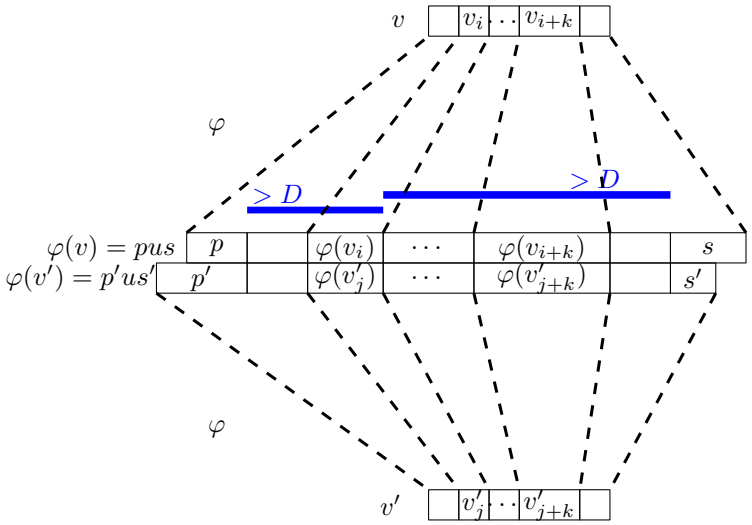


Figure 3.18: A graphical illustration of Definition 3.6.12.

In other words, a long enough word has unique φ-preimage except for some prefix and suffix shorter than a constant $D + 1$. Indeed, it

can be seen that for a long enough word u in the language $\mathcal{L}(\mathbf{u})$ of a circular morphism, we may be able to apply Definition 3.6.12 to all positions $i+k \in \mathbb{N}$ for $k > 0$, $k \in \mathbb{N}$ that satisfy the following inequalities,

$$|\varphi(v_0 \dots v_{i+k})| - |p| > D \text{ and } |\varphi(v_{i+k+1} \dots v_n)| - |s| > D.$$

Let $K \in \mathbb{N}$ be the largest such integer k. It follows that $v_i = v'_j$, $v_{i+1} = v'_{j+1}, \dots, v_{i+K} = v'_{j+K}$. Thus, the word u can be decomposed as follows,

$$u = u_0 \dots u_{f(i)-1}\varphi(v_i)\varphi(v_{i+1}) \dots \varphi(v_{i+K})u_{f(i+K+1)} \dots u_{|u|-1}$$

where,

$$|u_0 \dots u_{f(i)-1}| < D \text{ and } |u_{f(i+K+1)} \dots u_{|u|-1}| < D.$$

The word $v_i v_{i+1} \dots v_{i+K}$ is factor of any preimage of the word u.

Remark 3.6.13. The definition of circularity applies to words $u \in \mathcal{L}(\mathbf{u})$ longer than $2D+1$. Any word u longer than $2D + ||\varphi||$ of the language of a circular morphism with constant D have at least a synchronizing point, since $|u_{[D,|u|-D]}| \geq ||\varphi||$ and for that factor we will have unique preimage and therefore at least a synchronization point. The existence of the synchronization point implies weak circularity for the morphism.

Observation 3.6.14. Circularity implies weak circularity but the inverse is not always true.

Example 3.6.15. Let us work with the non-injective morphism $\varphi : \{0,1,2\}^* \to \{0,1,2\}^*$ with $\varphi(0) = 0120$, $\varphi(1) = 12$, $\varphi(2) = 12$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. The language $\mathcal{L}(\mathbf{u})$ is not circular since for every $m \in \mathbb{N}$ the word $(12)^{2m}$ has two different preimages $(12)^m$ and $(21)^m$. However, the corresponding external interpretations have synchronizing points for $m > 1$ at the positions $2k$ for $0 \leq k \leq m$. One can easily check that the language $\mathcal{L}(\mathbf{u})$ is weakly circular since every $v \in \mathcal{L}(\mathbf{u})$ with length more than $D = 5$ has a synchronization point. Indeed, $D \neq 3$ since the word $1201 \in \mathcal{L}(\mathbf{u})$ has two distinct external interpretations $(\varepsilon, 20, 20)$ and $(0, 01, 2)$ and no synchronization point. Similarly, $D \neq 4$ since the word 12012 has two distinct interpretations $(0, 01, \varepsilon)$ and $(\varepsilon, 20, 0)$ and no synchronization point. It can be verified that $D = 5$ by checking that all

words of length 6 have synchronization point, and thus the morphism is weakly circular but not circular.

The delay of circularity and that of weak circularity are not necessarily equal. If a morphism has circular delay D, it is not implied that every word u of length at least $D+1$ has a synchronizing point. However, for the words longer than $2D+1$, if there is a cutting point in between the positions D and $|u|-D$, then this position is a synchronization point because of circularity. The example 3.6.16 demonstrates the case of a morphism that is circular and weakly circular with the two delays not being equal.

Example 3.6.16. Let φ_{TM} be the Thue-Morse morphism (Example 3.1.8). It can be verified that φ_{TM} is circular for a delay $D = 1$. For $D = 1$ it is enough to check the words of length 5 (since the definition holds for length $|u| > 2 \times 1 + ||\varphi_{TM}|| = 4$). Since circularity implies weak circularity, the morphism φ_{TM} is also weakly circular. It is easy to see that weak circularity does not hold for $D' = 1$, since 01 does not have any synchronization point. Also, $D' \neq 2$ since the word 010, with length more than $D' = 2$ does not have a synchronizing point. It can be verified that the definition of weak circularity holds for $D' = 3$, since all words longer than 3 have at least a synchronization point. Indeed, the words 0101 and 1010 that belong in the language are the only words with length at least 3 that do not have as factor two consecutive 0's or 1's. Both of them have unique external interpretation, $(\varepsilon, 00, \varepsilon)$ and $(\varepsilon, 11, \varepsilon)$ respectively. Therefore, the positions where the interpretations pass are synchronization points. All the other words longer that 4 have at least one factor 00 or 11, in which case we know that the position in between the same letters is a synchronizing point.

The following proposition proves that weak circularity implies circularity in the case of injective substitutions.

Proposition 3.6.17. An injective weakly circular morphism with delay D is circular for a certain delay D'.

Proof. Let the injective morphism $\varphi : A^* \to A^*$ with fixed point $\mathbf{u}$ be weakly circular for a delay $D > 0$. Now let us take a word $w \in \mathcal{L}(\mathbf{u})$ such that $|w| > 2D+1$. We recall that $f(i) = |\varphi(\mathbf{u}_{[0,i)})|$. Because of the chosen length of w we have that it has at least two distinct synchronization points in positions $f(i), f(j)$ of w, with $0 \leq$

$f(i) < D+1 \leq f(j) \leq 2D+1$. Since the morphism is injective, the factor $w_{[f(i),f(j)-1]} = \varphi(x)$ has unique preimage $x \in \mathcal{L}(\mathbf{u})$. Now let (p, v, s) and (p', v', s') be two external interpretations of the word w. If there is an integer κ such that,

$$|\varphi(v_0 \dots v_\kappa)| - |p| > D \wedge |\varphi(v_{\kappa+1} \dots v_n)| - |s| > D$$

and for the position $f(\kappa)$ holds that $f(i) \leq f(\kappa) \leq f(j)$, then the letter v_κ is unique as it is factor of the unique preimage x. Hence, for any other interpretation (p', v', s') of w there is an integer λ such that,

$$|\varphi(v_0 \dots v_{\kappa-1})| - |p| = |\varphi(v'_0 \dots v'_{\lambda-1})| - |p'|$$

and $v_\kappa = v'_\lambda$. It follows then that the morphism is circular for D, but it is possible that there are smallest constants for which circularity is satisfied ($D' \leq D$). □

The injectivity is an important condition in the above proposition. As it is indicated by Example 3.6.15, the above is not true for the non-injective case.

The notion of circularity seems similar to the one of synchronizing delay. The equivalence between the two is given by the following lemma.

Lemma 3.6.18. A primitive morphism $\varphi : A^* \to A^*$ has finite synchronizing delay if and only if it is circular.

Proof. Let φ be a primitive morphism and $\mathcal{L}(\mathbf{u})$ be the language of the morphism with a fixed point $\mathbf{u}$. Let φ be circular morphism for a delay $D > 0$ and u a word in $\mathcal{L}(\mathbf{u})$ with external interpretations (p_1, v, s_1) and (p_2, v', s_2), with $v = (v_i)_{0 \leq i \leq n}$ and $v' = (v'_i)_{0 \leq i \leq m}$. It follows that there are internal interpretations $(p'_1, v_1 \dots v_{n-1}, s'_1)$ and $(p'_2, v'_1 \dots v'_{m-1}, s'_2)$, with words p'_1, p'_2, s'_1, s'_2, such that $\varphi(v_0) = p_1 p'_1$, $\varphi(v_n) = s'_1 s_1$, $\varphi(v'_0) = p'_2 p_2$ and $\varphi(v'_n) = s'_2 s_2$. Let $i \in \mathbb{N}$ be an integer that satisfies the following inequalities,

$$\begin{aligned} |p'_1 \varphi(v_1) \dots \varphi(v_{i-1})| &= |\varphi(v_0) \dots \varphi(v_{i-1})| - |p_1| > D \\ |\varphi(v_{i+1}) \dots \varphi(v_{n-1}) s'_1| &= |\varphi(v_{i+1}) \dots \varphi(v_n)| - |s_1| > D. \end{aligned}$$

Because of circularity, there is an integer j such that $0 \leq j \leq m$ such that,

$$|p'_1 \varphi(v_1 \dots v_{i-1})| = |p'_2 \varphi(v'_1 \dots v'_{j-1})| \text{ and } v_i = v'_j.$$

Hence,

$$|p'_1\varphi(v_1)\dots\varphi(v_{i-1})\varphi(v_i)| = |p'_2\varphi(v'_1)\dots\varphi(v'_j)|$$

and since the words $p'_1\varphi(v_1\dots v_i)$ and $p'_2\varphi(v'_1\dots v'_j)$ are prefixes of u, it follows that D is a synchronizing delay.

For the other direction, we suppose that there is word $u \in \mathcal{L}(\mathbf{u})$ such that $u = p\varphi(v_1)\dots\varphi(v_{n-1})s = p'\varphi(v'_1)\dots\varphi(v'_{m-1})s'$. Because of primitivity, the morphism is uniformly recurrent and consequently every word in the language is extendable in both sides. That implies that there are $v, v' \in \mathcal{L}(\mathbf{u})$ such that $\varphi(v) = p_1p\varphi(v_1)\dots\varphi(v_{n-1})ss_1$ and $\varphi(v') = p'_1p'\varphi(v'_1)\dots\varphi(v'_{m-1})s's'_1$. Thus, there exist two external interpretations $(p_1, v, s_1), (p'_1, v', s'_1)$ of the word u. From the definition of synchronizing delay we have that if $|p\varphi(v_1)\dots\varphi(v_{i-1})| > D$ and $|\varphi(v_{i+1})\dots\varphi(v_{n-1})s| > D$ then there is j integer such that $p\varphi(v_1)\dots\varphi(v_i) = p'\varphi(v'_i)\dots\varphi(v'_j)$ and $v_i = v_j$. We have also that

$$|p\varphi(v_1)\dots\varphi(v_{i-1})| > D \Rightarrow |p\varphi(v_1)\dots\varphi(v_{i-1})\varphi(v_i)| > D$$

and $|\varphi(v_{i+1})\dots\varphi(v_{n-1})s| > D$. Thus, if

$$|p\varphi(v_1\dots v_i)| > D + ||\varphi||$$

and

$$|\varphi(v_{i+1}\dots v_{n_1}s)| > D + ||\varphi||$$

there is a position j of v' for which holds that,

$$\begin{aligned} |\varphi(v_0v_1\dots v_i)| - |p_1| &= |p\varphi(v_1\dots v_i)| = \\ &= |p'\varphi(v'_1\dots v'_j)| = |\varphi(v'_0\dots v'_j)| - |p'_1| \end{aligned}$$

and $v_i = v'_j$, from which we conclude that the language $\mathcal{L}(\mathbf{u})$ is circular. □

Observation 3.6.19. Primitivity is needed for the proof of Theorem 3.6.18 since if primitivity does not hold, we cannot convert the internal interpretations to external interpretations.

Observation 3.6.20. It follows from the proof of Lemma 3.6.18 that if a morphism is circular with delay D then the same constant satisfies the definition of synchronizing delay.

In the work of K. Klouda and S. Starosta [65] we can find the following corollary that guarantees the circularity for the case of primitive injective morphisms with aperiodic fixed point.

Corollary 3.6.21. [65] If $\mathbf{u}$ is an aperiodic fixed point of a primitive injective morphism, then the morphism is circular.

The bound on the value of the delay D is hard to find, however a bound is given in [64] for the simple case of a circular k-uniform morphism on a two letter alphabet, with $k \geq 2$.

Theorem 3.6.22. *[64] If a morphism $\varphi : \{a, b\}^* \to \{a, b\}^*$ is circular with fixed point $\mathbf{u} = \varphi^\omega(a)$ and k-uniform, then the minimum value of its delay, denoted by D_{min}, is bounded as follows:*

1. $D_{min} \leq 8$ *if* $k = 2$,
2. $D_{min} \leq k^2 + 3k - 4$ *if k is an odd prime number,*
3. $D_{min} \leq k^2(\frac{k}{d} - 1) + 5k - 4$ *otherwise,*

where d is the least divisor of k greater than 1.

Observation 3.6.23. If a language $\mathcal{L}(\mathbf{u})$ contains arbitrary long words with two different φ-preimages it cannot be circular.

Example 3.6.24. Let φ be morphism on $A = \{0, 1, 2\}$ with $\varphi(0) = 0120$, $\varphi(1) = 121$, $\varphi(2) = 212$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. It is easy to verify that $(12)^n \in \mathcal{L}(\mathbf{u})$ for all $n > 0$. For all integers $D > 0$ there is word $v \in (12)^* \cap \mathcal{L}(\mathbf{u})$ of length $|v| = 2D$. The image of the word v, let $u = \varphi(v) \in (12)^*$, is also in $\mathcal{L}(\mathbf{u})$ and has length $|u| = 3(2D) = 6D$ and it is of the form $u = 1212 \cdots 2$. Every such word u has two different external interpretations, $(\varepsilon, (12)^D, \varepsilon)$ and $(p, v, s) = (2, (21)^D 2, 12)$. For the interpretation $(\varepsilon, (12)^D, \varepsilon)$ and the position $i = D - 1$ of the word v it holds that,

$$|\varphi(v_0 \dots v_i)| - |\varepsilon| = 3|v_0 \dots v_i| - 0 = 3D > D$$

and

$$|\varphi(v_{i+1} \dots v_{2D-1})| - |\varepsilon| = 3(2D) = 6D > D.$$

However, for the external interpretation $(p', v', s') = (2, (21)^D 2, 12)$, there is no integer j such that $|\varphi(v'_0 \dots v'_j)| - |p'| = 3D$. Indeed,

$$|\varphi(v'_0 \dots v'_j)| - |p'| = 3(j - 1) - 1 = 3j - 2 \neq 3D$$

for any $D \in \mathbb{N}$. Thus, the definition of circularity is not verified for any constant $D > 0$ which proves that the morphism is not circular.

The following theorem gives the equivalence between circularity and strongly recognizability.

Theorem 3.6.25. *A morphism is circular if and only if it is strongly two-sided recognizable.*

Proof. Let φ be a circular morphism with delay $D > 0$, with fixed point $\mathbf{u}$. Let $u \in \mathcal{L}(\mathbf{u})$ be longer than $2D + 2||\varphi||$, with external interpretation (p, v, s), where $v = (v_i)_{0 \leq i \leq n}$. Because of the length of u the factor $u_{[D,|u|-D]}$ of u is longer than $2||\varphi||$. Therefore, there are at least two distinct positions in $E_{\mathbf{u}}$, $f(i), f(j)$ with $0 \leq f(i) < D + ||\varphi|| + 1$ and $|u| - (D + ||\varphi|| + 1) \leq f(j) < |u|$. The integers i, j are positions in v such that the following inequalities hold,

$$|\varphi(v_0 \dots v_i)| - |p| > D \quad \text{and} \quad |\varphi(v_{i+1} \dots v_n)| - |s| > D$$
$$|\varphi(v_0 \dots v_j)| - |p| > D \quad \text{and} \quad |\varphi(v_{j+1} \dots v_n)| - |s| > D.$$

Without any loss of generality i can be the smallest such integer and j the largest. Because of circularity, the factor $v_i \dots v_j$ of v is "fixed" and is factor of any other external interpretation. Hence, we have that φ is strongly two-sided recognizable for constant $\ell = D + ||\varphi|| + 1$.

In order to prove the other way we suppose that φ is strongly two-sided recognizable with a constant $\ell > 0$. The definition of circularity holds for $D = \ell$. Indeed, if (p, v, s) and (p', v', s') are external interpretations of u and $|\varphi(v_0 \dots v_i)| - |p| > \ell$ and $|\varphi(v_{i+1} \dots v_n)| - |s| > \ell$, then there exist positive integers α, β, with $0 \leq \alpha < i$ and $i + 1 \leq \beta < n$ such that $u_{[f(\alpha),f(\beta))} = \varphi(v_\alpha \dots v_{\beta-1})$, where $v_\alpha \dots v_{\beta-1}$ is "fixed" because of strongly recognizability ("fixed" in the sense that is always factor of any v' for which (p, v', s) external interpretation of u and that its image corresponds always to the same position in u). Thus, since the letter v_i is factor of $v_\alpha \dots v_{\beta-1}$, the letter v_i is also "fixed", which implies circularity for $D = \ell$. (See Figure 3.19) □

The following result is a stronger than Corollary 3.6.21, given by K. Klouda and S. Starosta in [65], since it is not required injectivity.

Theorem 3.6.26. *Any primitive aperiodic morphism is circular.*

Proof. Let us suppose that φ is primitive aperiodic morphism. Then from Mossé's Theorem 3.4.1 follows that the morphism φ is strongly recognizable. From Theorem 3.6.25 it follows that since φ is strongly two-sided recognizable, it is circular. □

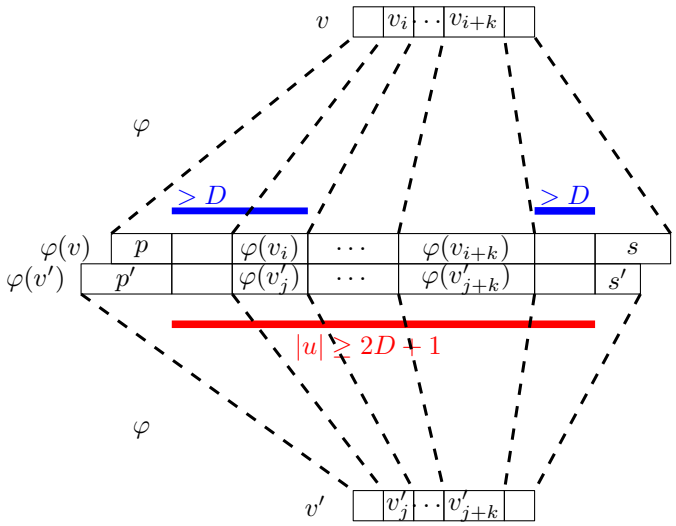


Figure 3.19: A graphical explanation of the proof of Theorem 3.6.25.

3.7 Summary and graphical illustration of the chapter

In this section we will summarize the results presented in the chapter using a graphical representation of the relations that exist between the different definitions.

The following graph (Figure 3.23) will make clear which definitions are stronger and which are equivalent under certain conditions in the general case of a substitutive language $\mathcal{L}(\mathbf{u})$ of a morphism φ with an aperiodic fixed point $\mathbf{u}$.

The set diagram presented in Figure 3.20 shows all possible relations between the finite collection of different sets that satisfy the definitions that are mentioned in this chapter.

It is important to mention that the graph in Figure 3.20 is simplified in the case of primitive morphisms as it can be seen in the following remark and in Figure 3.22.

Remark 3.7.1. Let $\varphi : A \to A^*$ be an aperiodic primitive morphism with φ-fixed point $\mathbf{u}$. Then, the morphism φ is:

1. two-sided recognizable,

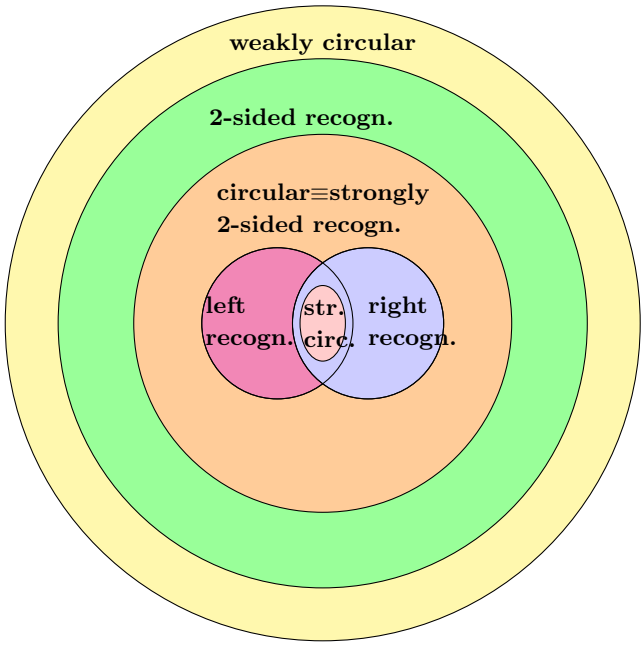


Figure 3.20: Relations between the definitions of recognizability and circularity.

2. strongly two-sided recognizable,
3. circular,
4. weakly circular.

Remark 3.7.1 is the result of propositions and theorems presented earlier in this chapter. More precisely, strongly two-sided recognizability implies two-sided recognizability from Observation 3.3.28, while the other direction follows from Mossé's Theorem 3.4.1, since every primitive morphism is strongly two-sided recognizable and therefore every primitive two-sided recognizable morphism is strongly two-sided recognizable. The equivalence between strongly two-sided recognizability and circularity holds from Theorem 3.6.25. Lastly, circularity implies weakly circularity, as stated in the Observation 3.6.14, while the other direction follows in the primitive aperiodic case from Mossé's Theorem 3.4.1 and Theorem 3.6.25, since every primitive aperiodic morphism is strongly two-sided recognizable and therefore circular, which means that adding the condition that the morphism is weakly circular does not change the previous result. A graphical illustration of the equivalences described above is the Figure 3.21.

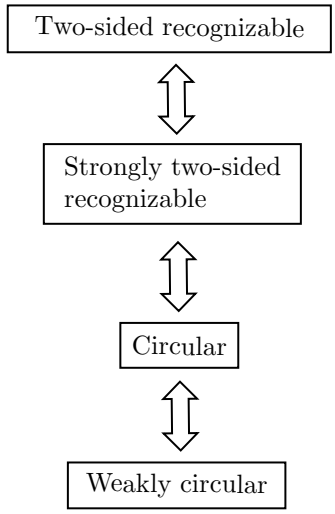


Figure 3.21: Relations between the definitions of recognizability and circularity for the case of primitive and aperiodic morphisms.

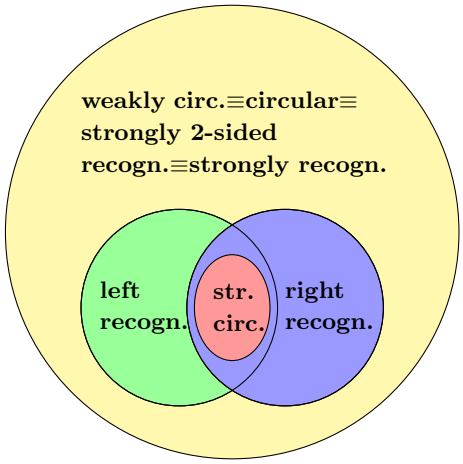


Figure 3.22: Relations between the definitions of recognizability and circularity for the primitive case.

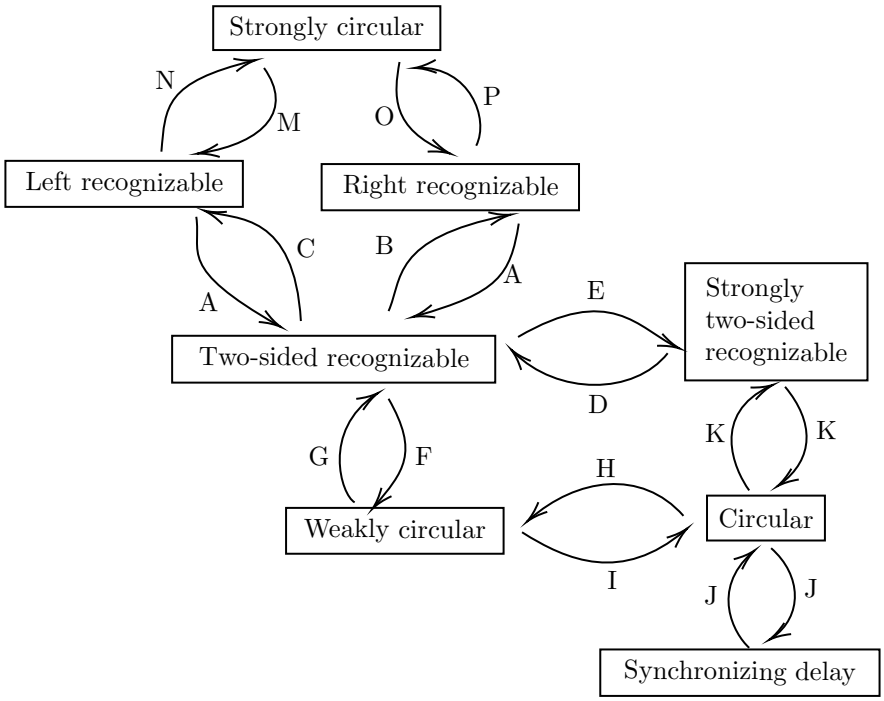


Figure 3.23: Relations between the definitions for an aperiodic morphism.

A: Proposition 3.3.21 (holds always),

B: Proposition 3.3.23 (under the condition of suffix morphism),

C: similar to Proposition 3.3.25 (under the condition of prefix morphism),

D: Observation 3.3.28 (holds always),

E: Lemma 3.3.36 (under the condition of suffix or prefix morphism),

F: Theorem 3.6.9 (holds always),

G: Proposition 3.6.10 (under the condition of injective morphism),

H: Observation 3.6.14 (holds always),

I: Proposition 3.6.17 (under the condition of injective morphism),

J: Lemma 3.6.18 (under the condition of primitive morphism),

K: Theorem 3.6.25 (holds always),

M: Proposition 3.3.15 (holds always),

N: Remark 3.3.17 (hunder the condition of suffix morphism),

O: Proposition 3.3.8 (holds always),

P: Remark 3.3.10 (under the condition of prefix morphism).

4

Decidable properties of extension graphs

Given a set of words S, one can associate with every word $u \in S$ a graph called *extension graph*. This graph carries the information of how the word u can be extended on the left and on the right in S. The reason why extension graphs are worth studying is that they allow the characterization of a set as acyclic, dendric set, neutral set, etc.
Normally, in order to be able to decide some of the above properties we should construct the extension graph for all the words of the set S, and check if the property holds for all of these graphs. However, since we may work on an infinite set S, such as substitutive languages, the above method cannot be applied. In order to be able to decide if some of the above properties holds, we use the notion of strong two-sided recognizability.

4.1 Extension graphs

4.1.1 Extension graph

Let S be a bi-extendable set of words. For $w \in S$, we consider the undirected bipartite graph $\mathcal{E}_S(w)$ with set of vertices the disjoint union of $L_S(w)$ and $R_S(w)$ and edges the pairs $(a, b) \in B_S(w)$, where $L_S(S), R_S(S), B_S(S)$ are the ones defined in chapter Preliminaries. This graph is called the *extension graph* of w. When it is clear from the context we omit the S. We note that, since $\mathcal{E}_S(w)$ has $\ell(w) + r(w)$ vertices and $b(w)$ edges, the number $1 - m(w)$ is the

Euler characteristic of the graph $\mathcal{E}_S(w)$ (see [33]).

If the extension graph $\mathcal{E}_S(w)$ is acyclic, then $m(w) \leq 0$. Thus w is weak or neutral. More precisely, one has in this case that $c = 1 - m(w)$ is the number of connected components of the graph $\mathcal{E}_S(w)$.

A graph is called *acyclic* if there is no way to start at any vertex and follow a consistently-directed sequence of edges that eventually loops back to that vertex again. A bi-extendable set S is called *acyclic* if for every $w \in S$ the graph $\mathcal{E}_S(w)$ is acyclic. A graph is *connected* when there is a path between every pair of vertices. The set S is *connected* if for every $w \in S$ the graph $\mathcal{E}_S(w)$ is connected.

A graph is *dendric* if it is connected and acyclic. A bi-extendable set S is called a *dendric set* of characteristic c if for any nonempty $w \in S$, the graph $\mathcal{E}_S(w)$ is a dendric and if $\mathcal{E}_S(\varepsilon)$ is a union of c dendric graphs (the definition of dendric set in [19] corresponds to a dendric set of characteristic 1). It is clear that a dendric set is always a connected and acyclic set. Note that a dendric set of characteristic c is a neutral set of characteristic c.

Example 4.1.1. Let φ_F be the Fibonacci morphism (Example 2.2.2) with fixed point $\mathbf{u}$ the Fibonacci word. The extension graphs of the empty word ε and of the two letters are shown in Figure 4.1. The extension graphs $\mathcal{E}(\varepsilon)$, $\mathcal{E}(0)$ and $\mathcal{E}(1)$ are connected and acyclic.

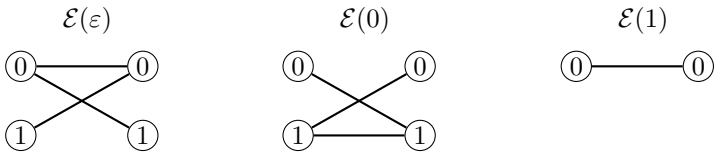


Figure 4.1: The graphs $\mathcal{E}(\varepsilon)$, $\mathcal{E}(0)$ and $\mathcal{E}(1)$ for the language $\mathcal{L}(\mathbf{u})$.

Example 4.1.2. Let φ_{TM} be the Thue-Morse morphism (Example 3.1.8) with fixed point $\mathbf{u} = \varphi_{TM}^{\omega}(0)$. The extension graphs of the empty word and of the two letters are shown in Figure 4.2. All of the graphs are connected while the graph $\mathcal{E}(\varepsilon)$ has a cycle and the graphs $\mathcal{E}(0)$ and $\mathcal{E}(1)$ are acyclic.

Remark 4.1.3. Any non bispecial word has extension graph that is dendric.

If a word is not bispecial, then its extension graph has one of the forms as seen in the Figure 4.3. The graph $\mathcal{E}_1$ corresponds to

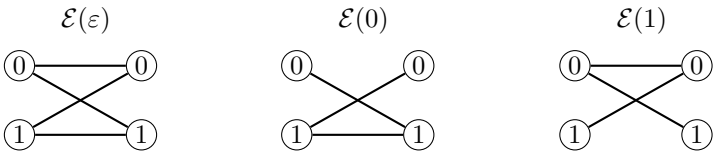


Figure 4.2: The graphs $\mathcal{E}(\varepsilon)$, $\mathcal{E}(0)$ and $\mathcal{E}(1)$ for the language $\mathcal{L}(\mathbf{u})$.

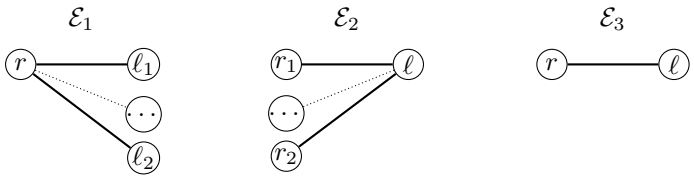


Figure 4.3: Extension graphs of non bispecial words.

the words that are right special but not left special, the graph $\mathcal{E}_2$ corresponds to the words that are left special but not right special, while the graph $\mathcal{E}_3$ corresponds to the words that are neither right special nor left special.

A *planar dendric set* of characteristic c with respect to two orders $<_L$ and $<_R$ on the alphabet A is a dendric set of characteristic c compatible with the two orders (see [20]), i.e., for any $w \in S$ one has

$$a <_L c \implies b \leq_R d \qquad \text{for any } (a,b),(c,d) \in B_S(w).$$

As explained in [20] we can extend $<_L$ and $<_R$ to two orders over A^* by taking the induced anti-lexicographic order for $<_L$ and induced lexicographic order for $<_R$.

Thus, if the extension graph $\mathcal{E}(w)$ is a planar dendric graph and it has the vertices of $L_S(w)$ placed ordered by $<_L$ on a line and those of $R_S(w)$ placed ordered by $\leq_R$ on a parallel line, then the edges of the graph will be straight non-crossing segments.

4.1.2 Generalized extension graph

For two sets of words X, Y and a word $w \in S$, we denote

$$\begin{aligned} L_S^X(w) &= \{x \in X \mid xw \in S\}, \\ R_S^Y(w) &= \{y \in Y \mid wy \in S\}, \\ B_S^{X,Y}(w) &= \{(x,y) \in X \times Y \mid xwy \in S\}. \end{aligned}$$

We omit the subscript S when it is clear from the context.

As for $\mathcal{E}(w)$, we consider $\mathcal{E}^{X,Y}(w)$ as an undirected bipartite graph on the set of vertices which is the disjoint union of $L^X(w)$ and $R^Y(w)$. Such a graph is called a *generalized extension graph*. When $X = A^n$, and S is understood, we write $L^{(n)}(w)$ instead of $L^{A^n}(w)$. Similarly we define $R^{(m)}(w)$, $B^{(n,m)}$ and $\mathcal{E}^{(n,m)}(w)$. Moreover, when $n = m$, we set $B^{(n)} = B^{(n,n)}$ and we call the graph $\mathcal{E}^{(n)}(w) = \mathcal{E}^{(n,n)}(w)$ the *uniform generalized extension graph* (of degree n) of w. Note that $\mathcal{E}(w) = \mathcal{E}^{(1,1)}(w)$.

Observation 4.1.4. Note that identifying in a generalized extension graph $\mathcal{E}^{(n,m)}(u)$ of a word u in a substitutive language all vertices on the left having the same last letter and all vertices on the right having the same first letter, we find the extension graph $\mathcal{E}^{(1,1)}(u)$.

Indeed, if there is an edge (u_1a, bu_2) in the graph $\mathcal{E}^{(n,m)}(u)$, then $u_1aubu_2 \in \mathcal{L}(\mathbf{u})$ and therefore there is edge $(a, b) \in \mathcal{E}(u)$. Let us now suppose that there is an edge $(c, d) \in \mathcal{E}(u)$ that does not correspond to any edge in $\mathcal{E}^{(n,m)}(u)$. That means that there are no words $u_3, u_4 \in \mathcal{L}(\mathbf{u})$, with $|u_3| = n - 1$ and $|u_4| = m - 1$, such that $u_3cudu_4 \in \mathcal{L}(\mathbf{u})$, which implies that there is a word in $\mathcal{L}(\mathbf{u})$ that is not extendable, which leads to a contradiction.

Example 4.1.5. Let φ_F be the Fibonacci morphism (Example 2.2.2) with fixed point $\mathbf{u}$ the Fibonacci word.

The extension graph $\mathcal{E}(\varepsilon)$ and the generalized uniform extension graph $\mathcal{E}^{(3)}(\varepsilon)$ of the empty word are represented in Figure 4.4. It can be seen that as it was mention in the above Observation, identifying in $\mathcal{E}^{(3)}(\varepsilon)$ all vertices on the left having the same last letter and all vertices on the right having the same first letter, we find the graph $\mathcal{E}(\varepsilon)$.

Figure 4.4: The graphs $\mathcal{E}(\varepsilon)$ (on the left) and $\mathcal{E}^3(\varepsilon)$ (on the right). Both of the graphs are connected and acyclic, hence they are dendric.

The Fibonacci set is a planar dendric set with respect to the two orders $1 <_L 0$ and $0 <_R 1$ (see [20]).

4.2 Antecedent

We will define in this section the notion of the *antecedent* of a word in a substitutive language and we will correlate this notion with the one of strongly two-sided recognizability, that was thoroughly presented in the previous chapter.

Let us first give the following examples as motivation.

Example 4.2.1. Let φ_F be the Fibonacci morphism with fixed point $\mathbf{u}$ the Fibonacci word (Example 2.2.2). The morphism φ_F is right recognizable as it has already seen in Example 3.3.3 and it is suffix. It can be verified that any word $u \in \mathcal{L}(\mathbf{u}) \cap A^*1$ has a unique internal interpretation and therefore it can be uniquely written as $u = s\varphi_F(v)p$, where $v \in \mathcal{L}(\mathbf{u})$ and s a proper suffix of $\varphi(0)$ or $\varphi(1)$, i.e $s \in \{\varepsilon, 1\}$ and $p = \varepsilon$. The word v is considered as the preimage of u or, as we will call it later, its *antecedent*.

Example 4.2.2. Let ψ be the following morphism,

$$\psi : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 01 \end{cases},$$

with fixed point $\mathbf{u} = \psi^\omega(0) = 0101010101\cdots$ and therefore language $\mathcal{L}(\mathbf{u}) = (\varepsilon + 1)(01)^*(\varepsilon + 0)$. Any word $u \in \mathcal{L}(\mathbf{u})$ is of the form

$$u = s(01)^k p$$

for some words $p \in \{\varepsilon, 0\}, s \in \{\varepsilon, 1\}$. We thus get two possible factorizations

$$s\psi((10)^{\lfloor k/2 \rfloor})p = u = s\psi((01)^{\lfloor k/2 \rfloor})p,$$

which prevents us from defining *the* antecedent of u.

The next example shows that antecedents may be undefined even if the considered morphism has an aperiodic fixed point.

Example 4.2.3. Consider the following morphism,

$$\rho : \begin{cases} 0 \mapsto 0120 \\ 1 \mapsto 121 \\ 2 \mapsto 212 \end{cases}$$

with fixed point $\mathbf{u} = \rho^\omega(0) = 01201212120120\cdots \in A^{\mathbb{N}}$. The fixed point $\mathbf{u}$ is not ultimately periodic, since for all integers $n \in \mathbb{N}$ the word $0(12)^{3^n}0$ occurs in the language $\mathcal{L}(\mathbf{u})$. We are unable to define the antecedent of the words of the form $(12)^k$. For example, the word $u = (12)^3 = 12121$ has internal interpretations $(\varepsilon, 1, 21)$, $(1, 2, 1)$ and $(12, 1, \varepsilon)$ and therefore we cannot define the antecedent of u.

Motivated by the previous examples, we will now formally define the notion of the antecedent.

Definition 4.2.4 (Antecedent)**.** Let $\varphi : A^* \to A^*$ be a morphism with fixed point $\mathbf{u}$ and $u \in \mathcal{L}(\mathbf{u})$. The *antecedent* of u in $\mathcal{L}(\mathbf{u})$ (if it exists) is the longest non-empty word $w \in \mathcal{L}(\mathbf{u})$ such that

1. $u = x\varphi(w)y$ for some words $x, y \in A^*$ and
2. for any internal interpretation (s, v, p) of u, there exists $i, j \in \{0, \ldots, |v| - 1\}$, $i \leq j$, such that $w = v_{[i,j]}$, $x = s\varphi(v_{[0,i)})$ and $y = \varphi(v_{(j,|v|-1]})p$.

(See Figure 4.5)

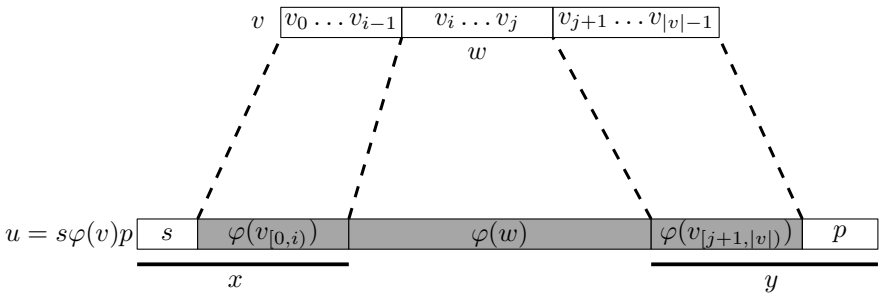


Figure 4.5: A graphical explanation of the definition of the antecedent.

In other words, the antecedent is the longest common factor from whom pass all the internal interpretations.

Remark 4.2.5. If the antecedent of a word in a substitutive language exists, then it is unique.

The uniqueness of the antecedent follows directly from the second condition in the definition of the antecedent and the fact that the antecedent is the longest such word.

Remark 4.2.6. If a word $u \in \mathcal{L}(\mathbf{u})$ admits an antecedent w, then any word $u' \in \mathcal{L}(\mathbf{u})$ that has the word u as factor admits an antecedent w'. More precisely, the antecedent w' of u' has as factor the antecedent w of u.

Example 4.2.7. Let φ_F be the Fibonacci morphism (Example 2.2.2) with fixed point $\mathbf{u}$ the Fibonacci word. The word $u = 10010 \in \mathcal{L}(\mathbf{u})$ has two distinct internal interpretations $(1, 101, \varepsilon)$ and $(1, 10, 0)$. It can be verified that the antecedent of u is the word $w = 10 \in \mathcal{L}(\mathbf{u})$, that is the longest common factor of the words 101 and 10.

The existence of the antecedent for a long enough word in a substitutive language is guaranteed in the case that strong two-sided recognizability holds for the language, as we will further explain it in the following part of this section.

Let us recall the definition of a *strongly two-sided recognizable* morphism on a fixed point (after Proposition 3.3.37), as the morphism $\varphi : A^* \to A^*$ with a fixed point $\mathbf{u}$ for which there is an integer $L > 0$ such that for every $i > L$ if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_{\mathbf{u}}$, then $j \in E_{\mathbf{u}}$ and $\mathbf{u}_{i'} = \mathbf{u}_{j'}$, where $f(i') = i$ and $f(j') = j$.

The next proposition states that strong two-sided recognizability ensures the existence of the antecedent for all words long enough in the language.

Proposition 4.2.8. If φ is a strongly two-sided recognizable morphism on a fixed point $\mathbf{u}$ for a constant $L > 0$, then for all words $u \in \mathcal{L}(\mathbf{u})$ of length at least $2L + ||\varphi||$, u has an antecedent w. Furthermore, if x, y are as in Definition 4.2.4, then we have $|x| < L + ||\varphi||$ and $|y| < L + 1$.

Proof. We will first prove the existence of the antecedent for a long enough word u. Let $u \in \mathcal{L}(\mathbf{u})$ with $|u| \geq 2L + ||\varphi||$. In the factor $v = u_{[L,|u|-1-L]}$ of $u = u_0 \dots u_{|u|-1}$, there is at least one cutting point, because of its length ($|v| \geq ||\varphi||$). Let $i, j \in E_{\mathbf{u}}$ be such that $L \leq i \leq L + ||\varphi||$ and $i \leq j \leq |u| - 1 - L$, be the smallest and the largest position respectively in the interval $[L, |u| - 1 - L]$ that belong in $E_{\mathbf{u}}$. Let us note that $\mathrm{Card}([L, |u| - 1 - L] \cap E_{\mathbf{u}}) > 0$. For any position $k \in [L, |u| - 1 - L] \cap E_{\mathbf{u}}$ there is a factor $u_{[k-L,k+L]}$ of the word u, where we can apply the definition of strong two-sided recognizability. Thus, the position k is a strong synchronization point (from Proposition 3.3.38) and the letter $\mathbf{u}_{k'}$ with $f(k') = k$ is

unique. Therefore, by concatenation on the positions in $[L, |u|-1-L] \cap E_{\mathbf{u}}$ the word $v' = \mathbf{u}_{[i',j']}$ with $f(i') = i$ and $f(j') = j$ is unique and thus it is a common factor of all the internal interpretations of u.

Let $((s_m, v_m, p_m))_{m \leq M}$ be the set of all internal interpretations of the word u and $M > 0$ be the number of distinct internal interpretations that correspond to u. Note that M is finite since every finite word has finite number of distinct internal interpretations. Since the word $v' = \mathbf{u}_{[i',j']}$ is a factor of any v_m, there are words $r_m, t_m \in \mathcal{L}(\mathbf{u})$ such that $v_m = r_m v' t_m$. It is then enough to consider the word $w = rv't$, where r (resp. t) is the longest common suffix (resp. prefix) of all $(r_m)_{m \leq M}$ (resp. $(t_m)_{m \leq M}$). It then follows that the word w is the antecedent of u.

Consider now an internal interpretation (s, v, p) of the word u and let w be the antecedent of u. Let $u = x\varphi(w)y$ for $x, y \in \mathcal{L}(\mathbf{u})$ and let $w = v_{[\ell,h]}$, for $0 \leq \ell, h \leq |v| - 1$. Since

$$u = x\varphi(w)y = x\varphi(v_{[\ell,h]})y = s\varphi(v)p$$

it follows that,

$$|x| \leq |s\varphi(v_{[0,\ell)})| \leq |s\varphi(v_{[0,\ell-1)})| + ||\varphi|| < L + ||\varphi||$$

and

$$|y| \leq |\varphi(v_{[h,|v|-1)}p)| \leq L.$$

□

Corollary 4.2.9. Let φ be a strongly two-sided recognizable morphism on a fixed point $\mathbf{u}$ for a constant $L > 0$. Let $x, y \in \mathcal{L}(\mathbf{u})$ be as in Definition 4.2.4 with $y \neq \varepsilon$ (resp. $x \neq \varepsilon$). If φ is suffix (resp. prefix) then there is a letter a such that y is a proper prefix (resp. x is a proper suffix) of $\varphi(a)$.

Proof. Let us assume that the morphism φ is prefix (resp. suffix) and strongly two-sided recognizable on $\mathcal{L}(\mathbf{u})$, where $\mathbf{u}$ is a fixed point of the morphism. Let x, y, w be defined as in the general case and let (s, v, p) be an internal interpretation of the word $u \in \mathcal{L}(\mathbf{u})$, with $u = s\varphi(v)p$. Let i, j with $i \leq j$ be positions in the word v such that $w = v_{[i,j]}$, where w is the antecedent of u. Thus, $y = \varphi(v_{(j,|v|-1]})p$. Considering another interpretation (s', v', p') of u, we also have $y = \varphi(v'_{(j',|v|-1]})p'$. Since w is the longest common

factor of the preimages and since the morphism φ is prefix, one has $\varphi(v_{(j,|v|-1]}) = \varphi(v'_{(j',|v'|-1]}) = \varepsilon$. Thus $y = s = s'$. □

We say that a morphism φ is *everywhere growing* if $|\varphi^n|$ goes to infinity when n increases. A primitive morphism is everywhere growing.

Corollary 4.2.10. Let φ be an everywhere growing morphism with a fixed point $\mathbf{u}$. If φ is strongly two-sided recognizable for some constant $L > 0$ on $\mathbf{u}$, then for all $u \in \mathcal{L}(\mathbf{u})$ with length at least $2L + ||\varphi||$, there exists a unique finite sequence $(u_1, u_2, \ldots, u_k)$ of words in $\mathcal{L}(\mathbf{u})$ such that

i) u_1 does not have any antecedent (in particular, $|u_1| < 2L+||\varphi||$);

ii) u_i is the antecedent of u_{i+1};

iii) $u_k = u$.

The previous sequence will be called *the sequence of antecedents of* u in $\mathcal{L}(\mathbf{u})$.

The above corollary is the result of applying recursively Proposition 4.2.8 on the words that are longer that $2L + ||\varphi||$, as it is demonstrated in the following example.

Example 4.2.11. The Fibonacci morphism φ_F with fixed point the Fibonacci word (Example 2.2.2) is everywhere growing and strongly two-sided recognizable for a constant $L = 1$. Let us consider the word $u_3 = 0010010100$. It can be verified that the antecedent of u_3 is the word $u_2 = 01001$, while the antecedent of the former is $u_1 = 010$ (see Figure 4.6). One has $|u_3| = 10 \geq 2 \cdot 1 + 2$, $|u_2| = 5 \geq 2 \cdot 1 + 2$ and $|u_1| = 3 < 2 \cdot 1 + 2$. Thus, for the word u_3 there is a unique finite sequence $(u_1, u_2, u_3) = (010, 01001, 0010010100)$, that satisfies Corollary 4.2.10 and it is the sequence of antecedents of u_3 in $\mathcal{L}(\mathbf{u})$.

Since strong two-sided recognizability can guarantee the existence of the antecedent of a long enough word in a substitutive language, it is important to be able to decide whether a morphism is strongly two-sided recognizable or not. The following theorems that appear in [44] and [79] lead to the result that under certain conditions it is decidable whether the morphism is strongly two-sided recognizable. Let us first give the following definitions.

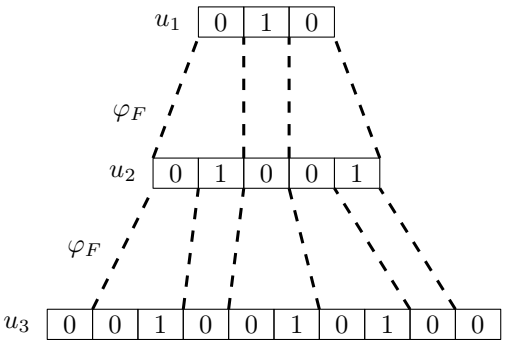


Figure 4.6: The sequence of antecedents of the word $u_3 = 0010010100$ in the Fibonacci language.

A word $u \in \mathcal{L}(\mathbf{u})$ is said to be *bounded* if $(|\varphi^n(u)|)_{n\geq 0}$ is a bounded sequence. A morphism φ is said to be *pushy* if there are arbitrarily long bounded words in $\mathcal{L}(\mathbf{u})$, otherwise it is *non-pushy*.

Let us also now mention that a word is called *k-power-free* if it does not contain any k-power as a factor, i.e., it has no factor of the form v^k, for some integer $k \geq 2$. A morphism φ is called *k-power-free* for some $k \geq 2$ if for every k-power-free word u in $\mathcal{L}(\varphi)$, its image $\varphi(u)$ is also k-power-free.

Theorem 4.2.12. *[44] It is decidable whether a morphism is pushy or not. Moreover, it is decidable whether a morphism is k-power-free for some $k \geq 2$.*

Theorem 4.2.13. *[79] Let φ be non-erasing. If φ is k-power-free for some $k \geq 2$, then it is strongly two-sided recognizable. If φ is non-pushy and strongly two-sided recognizable, then it is k-power-free for some $k \geq 2$.*

Corollary 4.2.14. If φ is non-erasing and non-pushy, then it is decidable whether or not it is strongly two-sided recognizable.

Note that a pushy morphism can be either strongly two-sided recognizable or not.

Example 4.2.15. The Fibonacci morphism φ_F (Example 2.2.2) with fixed point $\mathbf{u}$ the Fibonacci word is not pushy, since it is everywhere growing and thus no word in $\mathcal{L}(\mathbf{u})$ is bounded. On the other

hand, the morphism $\zeta : \{0,1\}^* \to \{0,1\}^*$,

$$\begin{aligned} \zeta : 0 &\mapsto 001, \\ 1 &\mapsto 1 \end{aligned}$$

with fixed point $\mathbf{u}_\zeta = \zeta^\omega(0)$ and the morphism $\eta : \{0,1,2,3\}^* \to \{0,1,2,3\}^*$,

$$\begin{aligned} \eta : 0 &\mapsto 01203, \\ 1 &\mapsto 121, \\ 2 &\mapsto 212, \\ 3 &\mapsto 3 \end{aligned}$$

with fixed point $\mathbf{u}_\eta = \eta^\omega(0)$ are pushy. Indeed, $1^m \in \mathcal{L}(\mathbf{u}_\zeta)$ and $3^m \in \mathcal{L}(\mathbf{u}_\eta)$ for all for all $m \geq 0$ (and $|\zeta^n(1^m)| = |\eta^n(3^m)| = m$ for all $n \geq 1$). While it can be checked that ζ is strongly two-sided recognizable with constant of recognizability 1, we can prove that the morphism η is not strongly two-sided recognizable. Indeed all words of the form $(12)^n \in \mathcal{L}(\mathbf{u}_\eta)$ for $n \in \mathbb{N}$ have at least two distinct internal interpretations that pass from totally different positions. Therefore, in the words of the previous form there is no synchronization point, let alone a strong synchronization point.

Lastly, let us recall Mossé's Theorem 3.4.1 that states that any primitive aperiodic morphism is strongly two-sided recognizable.

4.3 Evolution of the extensions of bispecial words

In what follows we will suppose that φ is a non-erasing non-pushy strongly two-sided recognizable morphism for a constant $L > 0$ on a fixed point $\mathbf{u}$ of φ.

We also assume that the substitutive language $\mathcal{L}(\mathbf{u})$ is *bi-extendable*, i.e., for all $u \in \mathcal{L}(\mathbf{u})$, there exist $a, b \in A$ such that $aub \in \mathcal{L}(\mathbf{u})$. We say that the morphism φ is *left-marked* (resp. *right-marked*) if for all distinct letters $a, b \in A$, $\varphi(a)$ and $\varphi(b)$ have distinct first letter (resp. distinct last letter).

Remark 4.3.1. If φ is a primitive morphism with fixed point $\mathbf{u}$, then the language $\mathcal{L}(\mathbf{u})$ is bi-extendable.

Example 4.3.2. Let φ_F be the Fibonacci morphism (Example 2.2.2) with fixed point $\mathbf{u}$ the Fibonacci word. Since the Fibonacci morphism is primitive, the substitutive language $\mathcal{L}(\mathbf{u})$ is bi-extendable. Also, since $\varphi(0), \varphi(1) \in 0A^*$ the morphism is not left-marked but it is right-marked since the two images end in different letters.

If $u \in \mathcal{L}(\mathbf{u})$ admits an antecedent w with $u = x\varphi(w)y$, then by definition there exist some words ℓ, r such that $\ell wr \in \mathcal{L}(\mathbf{u})$, x is a suffix of $\varphi(\ell)$ and y is a prefix of $\varphi(r)$. By Proposition 4.2.8, x and y have length smaller than $L + ||\varphi||$ and L respectively. Since φ is non-erasing, the same property holds for ℓ and r. For all integers $m, n > 0$, we set

$$B_{x,y}^{(m,n)}(w) = \{(\ell, r) \in B^{(m,n)}(w) \mid \varphi(\ell) \in A^*x \text{ and } \varphi(r) \in yA^*\}.$$

We say that a pair $((\ell, r), (\ell', r')) \in B_{x,y}^{(m,n)}(w) \times B_{x,y}^{(m,n)}(w)$ is a *left special pair* for w (resp. a *right special pair*) if ℓ, ℓ' (resp. r, r') have distinct last letter (resp. first letter). A pair $((\ell, r), (\ell', r')) \in B_{x,y}^{(m,n)}(w) \times B_{x,y}^{(m,n)}(w)$ is a *bispecial pair* if it is right-special and left-special.

Remark 4.3.3. Every right special (resp. left special) word has a right-special (resp. left-special) pair.

Proposition 4.3.4. If $w \in \mathcal{L}(\mathbf{u})$ is bispecial then there is a bispecial pair in $B_{x,y}^{(m,n)}(w) \times B_{x,y}^{(m,n)}(w)$.

Proof. A bispecial word $w \in \mathcal{L}(\mathbf{u})$ is right special and left special by definition. Thus, by Remark 4.3.3 there is a left-special pair

$$((\ell_1, r_1), (\ell_2, r_2)) \in B_{x,y}^{(m,n)}(w) \times B_{x,y}^{(m,n)}(w)$$

and right-special pair

$$((\ell_3, r_3), (\ell_4, r_4)) \in B_{x,y}^{(m,n)}(w) \times B_{x,y}^{(m,n)}(w).$$

Thus ℓ_1, ℓ_2 end in different letters and r_3, r_4 start with different letters. If ℓ_3, ℓ_4 end with different letters, then $((\ell_3, r_3), (\ell_4, r_4))$ is a bispecial pair. Similarly, if r_1, r_2 start with different letters, then $((\ell_1, r_1), (\ell_2, r_2))$ is a bispecial pair.

Let us now examine the case where $\ell_3, \ell_4 \in A^*a$ for some letter $a \in A$. Then at least one of the words ℓ_1, ℓ_2 does not end with the letter a. Without loss of generality, let $\ell_1 \in A^*b$ for some b distinct

from a. Similarly, the right extension $r_1 \in cA^*$ should have different first letter with at least one of the right extensions r_3, r_4. Without any loss of generality, let $r_3 \in dA^*$ for a letter d distinct from c. Thus, there is a bispecial pair

$$((\ell_1, r_1), (\ell_3, r_3)) \in B_{x,y}^{(m,n)}(w) \times B_{x,y}^{(m,n)}(w).$$

□

Let us consider words $u, v \in A^+$. We let $f_L(u, v)$ (resp. $f_R(u, v)$) denote the longest common suffix (resp. longest common prefix) of the words $\varphi(u)$ and $\varphi(v)$.

As it has been already mentioned in the previous chapter, strong two-sided recognizability implies that any long enough word in the language has a unique preimage except of some prefix and suffix. The length of that prefix and suffix in the case of a strongly two-sided recognizable language for a constant $L > 0$, depends on the constant L. More precisely, after Proposition 4.2.8 and Corollary 4.2.9 we obtain the upper bounds for the length of the prefix and suffix that are denoted as C_L and C_R respectively and they are the following ones,

$$C_L = \begin{cases} 1 & \text{if } \varphi \text{ is right-marked,} \\ ||\varphi|| & \text{if } \varphi \text{ is suffix, but not right-marked,} \\ L + ||\varphi|| & \text{otherwise;} \end{cases}$$

$$C_R = \begin{cases} 1 & \text{if } \varphi \text{ is left-marked,} \\ \min\{||\varphi||, L+1\} & \text{if } \varphi \text{ is prefix, but not left-marked,} \\ L+1 & \text{otherwise.} \end{cases}$$

Example 4.3.5. Let φ be the Fibonacci morphism with fixed point $\mathbf{u}$ the Fibonacci word (Example 2.2.2). Let us remind that the morphism is strongly two-sided recognizable for a constant $L = 1$. Since $\varphi(0) = 01 \in A^*1$ and $\varphi(1) = 0 \in A^*0$ we have that φ is right-marked and thus $C_L = 1$. However since $\varphi(0), \varphi(1) \in 0A^*$ the morphism φ is not left-marked and also since $\varphi(0) = 01 = \varphi(1)0$ it is also not prefix. Thus, $C_R = L + 1 = 1 + 1 = 2$.

Lemma 4.3.6. Let φ be a strongly two-sided recognizable morphism with fixed point $\mathbf{u}$. Let $u \in \mathcal{L}(\mathbf{u})$ and let w be its antecedent with $u = x\varphi(w)y$ and $x, y \in A^*$. If u is left special then so is

his antecedent w. More precisely, there is a left special pair in $B_{x,y}^{(|x|+1,|y|+1)}(w) \times B_{x,y}^{(|x|+1,|y|+1)}(w)$.

Proof. We proceed by contradiction, assuming that there is a letter $a \in A$ such that for all $(\ell, r) \in B_{x,y}^{(|x|+1,|y|+1)}(w)$, the last letter of ℓ is a. Let (s, v, p) be an internal interpretation of u with s being a proper suffix of the image of a letter $\alpha \in A$ and p a proper prefix of a letter $\beta \in A$ with $\alpha v \beta \in \mathcal{L}(\mathbf{u})$. Then, by definition of the antecedent, we have that $v = twq$ for some words t, q such that $x = s\varphi(t)$ and $y = \varphi(q)p$. By our assumption on the left extensions of w either $t = \varepsilon$ (and thus $\alpha \in \{\varepsilon, a\}$) or $t \in A^*a$. The second possibility cannot occur for all interpretations, otherwise w, who is a factor of aw would not be the antecedent of u (See Figure 4.7).

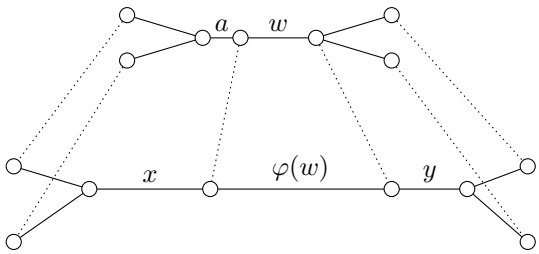


Figure 4.7: A graphical explanation of Lemma 4.3.6.

Thus, there exists at least one interpretation of the form (s, v, p) with $\alpha \in \{\varepsilon, a\}$ and $v \in wA^*$. In particular, we have $x = s$, where x is a suffix of $\varphi(a)$ distinct from $\varphi(a)$. This actually implies that all interpretations are of this form. Indeed, if (s', v', p') is an interpretation of u with $v' = t'wq'$ and $t' = t''a$, we must have $x = s = s'\varphi(t') = s'\varphi(t'')\varphi(a)$, which is impossible, since x is a suffix of $\varphi(a)$ distinct from $\varphi(a)$. We finally deduce that u has a unique left extending letter which is the letter b such that bx is suffix of $\varphi(a)$. Hence, u is not left special, which is a contradiction.

More precisely, the previous prove the existence of a left special pair in $B_{x,y}^{(|x|+1,|y|+1)}(w) \times B_{x,y}^{(|x|+1,|y|+1)}(w)$. □

The following lemma is similar to Lemma 4.3.6, considering the right special case instead of the left special one.

Lemma 4.3.7. Let φ be a strongly two-sided recognizable morphism with fixed point $\mathbf{u}$. Let $u \in \mathcal{L}(\mathbf{u})$ and let w be its antecedent

with $u = x\varphi(w)y$ and $x, y \in A^*$. If u is right special then so is his antecedent w. More precisely, there is a right special pair in $B_{x,y}^{(|x|+1,|y|+1)}(w) \times B_{x,y}^{(|x|+1,|y|+1)}(w)$.

The proof is analogous to the proof of Lemma 4.3.6.

The following lemma gives more precise information on the relation of a word and its antecedent.

Lemma 4.3.8. Let φ be a strongly two-sided recognizable morphism with fixed point $\mathbf{u}$. Let $u \in \mathcal{L}(\mathbf{u})$ and let w be its antecedent with $u = x\varphi(w)y$, with $x, y \in A^*$.

1. If u is left special, then so is w and there exists a left special pair
$$((\ell, r), (\ell', r')) \in B_{x,y}^{(C_L,C_R)}(w) \times B_{x,y}^{(C_L,C_R)}(w)$$
such that $x = f_L(\ell, \ell')$. In particular $|x| < C_L$.

2. If u is right special, then so is w and there exists a right special pair
$$((\ell, r), (\ell', r')) \in B_{x,y}^{(C_L,C_R)}(w) \times B_{x,y}^{(C_L,C_R)}(w)$$
such that $y = f_R(r, r')$. In particular, $|y| < C_R$.

Proof. Let us prove the first item and the second one can be proved in a symmetrical way.

The existence of a left special pair in $B_{x,y}^{(|x|+1,|y|+1)}(w) \times B_{x,y}^{(|x|+1,|y|+1)}(w)$ is given by Lemma 4.3.6.

We will now show that such a left special pair can be chosen to satisfy $x = f_L(\ell, \ell')$. By definition of $B_{x,y}^{(|x|+1,|y|+1)}(w)$, for all $(\ell, r) \in B_{x,y}^{(|x|+1,|y|+1)}(w)$, x is a suffix of $\varphi(\ell)$ and $x \neq \varphi(\ell)$ (because φ is non-erasing). If for all $(\ell, r), (\ell', r') \in B_{x,y}^{(|x|+1,|y|+1)}(w)$, we have $|f_L(\ell, \ell')| > |x|$, then u is not left special, which is a contradiction. Thus, there exists pairs $(h, s), (h', s') \in B_{x,y}^{(|x|+1,|y|+1)}(w)$ satisfying $x = f_L(h, h')$ (and in particular $x \notin \{\varphi(h), \varphi(h')\}$). If $f_L(\ell, \ell') = x$ or if $((h, s), (h', s'))$ is a left special pair, we are done. Let us show by contradiction that the converse cannot happen. We assume that $f_L(\ell, \ell') = x'$ with $|x'| > |x|$ (and x suffix of x') and that $a \in A$ is such that $h, h' \in A^*a$. Since $((\ell, r), (\ell', r'))$ is a left special pair, there exists $b, c \in A$ such that $\ell \in A^*b$ and $\ell' \in A^*c$, with $b \neq c$. We assume without loss of generality that $a \neq b$. Writing $z = f_L(\ell, h)$ and $z' = f_L(\ell, h')$, we have x suffix of z and of z'. If both z and z' are

longer than x, then since x' is also longer than x, this implies that x is a strict suffix of $f_L(h, h')$, which is a contradiction. Therefore, either $((\ell, r), (h, s))$ is a left special pair satisfying $x = f_L(\ell, h)$, or $((\ell, r), (h', s'))$ is a left special pair satisfying $x = f_L(\ell, h')$.

Let us finally show that we can replace $B_{x,y}^{(|x|+1,|y|+1)}(w)$ by $B_{x,y}^{(C_L,|y|+1)}(w)$. We consider a left special pair $((\ell, r), (\ell', r')) \in B_{x,y}^{(|x|+1,|y|+1)}(w) \times B_{x,y}^{(|x|+1,|y|+1)}(w)$ such that $x = f_L(\ell, \ell')$. In particular, we have $x \notin \{\varphi(\ell), \varphi(\ell')\}$. If $|x| + 1 = C_L$, we are done. If $|x| + 1 < C_L$, then since $\mathcal{L}(\mathbf{u})$ is bi-extendable, there exists h, h' such that $(h, r), (h', r') \in B_{x,y}^{(C_L,|y|+1)}(w)$ with $h \in A^*\ell$ and $h' \in A^*\ell'$. As $x \notin \{\varphi(\ell), \varphi(\ell')\}$, we have $f_L(h, h') = f_L(\ell, \ell') = x$. The last possibility $|x| + 1 > C_L$ cannot happen. Indeed, if φ is right-marked, then $x = \varepsilon$ and $C_L = 1$. Otherwise, Proposition 4.2.8 ensures that $|x| < C_L$.

Lastly, since we are interested in the left special pairs and since we are working on bi-extendable languages, we can change without any problem the length of the right extensions to C_R. □

Proposition 4.3.9. The antecedent w of a bispecial word u, with $u = x\varphi(w)y$, is bispecial and also there is a bispecial pair $((\ell_1, r_1), (\ell_2, r_2)) \in B_{x,y}^{(C_L,C_R)}(w) \times B_{x,y}^{(C_L,C_R)}(w)$, with $x = f_L(\ell_1, \ell_2)$ and $y = f_R(r_1, r_2)$. In particular, $|y| < C_R$ and $|x| < C_L$.

Proof. A word is bispecial when it is right special and left special. From Lemma 4.3.6 and Lemma 4.3.7 we have that the antecedent w of u is right special pair and left special respectively and thus we have that it is bispecial. Furthermore, from the previous Lemma, there is at least one left special pair and one right special pair in $B_{x,y}^{(C_L,C_R)}(w) \times B_{x,y}^{(C_L,C_R)}(w)$ and thus, following the same procedure as in Proposition 4.3.4, it can be verified that there is bispecial pair in $B_{x,y}^{(C_L,C_R)}(w) \times B_{x,y}^{(C_L,C_R)}(w)$.

Let us now suppose that for none of these pairs holds that $x = f_L(\ell_1, \ell_2)$ and $y = f_R(r_1, r_2)$. Let us study the case where $x = f_L(\ell_1, \ell_2)$ but $y \neq f_R(r_1, r_2)$. If $y \neq f_R(r_1, r_2)$ for all bispecial pairs in $B_{x,y}^{(C_L,C_R)}(w) \times B_{x,y}^{(C_L,C_R)}(w)$, then $|f_R(r_1, r_2)| > |y|$ for all bispecial pairs. Thus, there is a word $v \in A^+$ for each pair such that $f_R(r_1, r_2) = yv$. All those words (v) have to start with the same letter, because if not there can be constructed a pair such that $f_R = |y|$ that is not possible because of our assumptions. Therefore, the word

$x\varphi(w)y$ is left special but not right special, which leads to a contradiction.

The cases where $x \neq f_L(\ell_1, \ell_2)$ and $y \neq f_R(r_1, r_2)$ or $x \neq f_L(\ell_1, \ell_2)$ and $y \neq f_R(r_1, r_2)$ are similar. □

Remark 4.3.10. There exist two computable constants D_L and D_R such that for all $\ell \in A^{D_L}$ and all $r \in A^{D_R}$, $|\varphi(\ell)| \geq D_L + C_L - 1$ and $|\varphi(r)| \geq D_R + C_R - 1$.

Indeed, the existence of the constants D_L and D_R follows from the fact the morphism φ is assumed to be non-pushy and non-erasing and moreover, they are trivially computable.

Example 4.3.11. Let φ be the Fibonacci morphism with fixed point the Fibonacci word (Example 2.2.2). We have already seen in Example 4.3.5 that $C_L = 1$ and $C_R = 2$. The constant $D_L = 1$, since $|\varphi(0)| = |01| = 2 \geq D_L + C_L - 1 = 1 + 1 - 1 = 1$ and $|\varphi(1)| = |0| = 1 \geq D_L + C_L - 1 = 1$. The constant $D_R = 2$. Indeed, $D_R \neq 1$ since $|\varphi(1)| = |0| = 1 < D_R + C_R - 1 = 1 + 2 - 1 = 2$. By verifying all the words of length 2 in the substitutive language we have that $|\varphi(01)| = |010| = 3 \geq D_R + C_R - 1 = 2 + 2 - 1 = 3$, $|\varphi(00)| = |0101| = 4 \geq D_R + C_R - 1 = 3$ and $|\varphi(10)| = |001| = 3 \geq D_R + C_R - 1 = 3$.

Lemma 4.3.12. Let $u \in \mathcal{L}(\mathbf{u})$ bispecial with antecedent $w \in \mathcal{L}(\mathbf{u})$ such that $u = x\varphi(w)y$ and constants D_L, D_R as described in Remark 4.3.10. Then,

$$\mathrm{Card}(B^{(D_L,D_R)}(u)) \leq \mathrm{Card}(B_{x,y}^{(D_L,D_R)}(w)).$$

Note that the $\mathrm{Card}(B^{(D_L,D_R)}(u))$ is equal to the number of edges in the graph $\mathcal{E}^{(D_L,D_R)}(u)$. Thus, the above lemma proves that the extension graph $\mathcal{E}^{(D_L,D_R)}(w)$ of the antecedent has at least as many edges as the extension graph of the decedent.

Proof. Let φ be a morphism with a fixed point $\mathbf{u}$ and $u \in \mathcal{L}(\mathbf{u})$ with antecedent $w \in \mathcal{L}(\mathbf{u})$ such that $u = x\varphi(w)y$ and from Proposition 4.3.9 we have that $|x| < C_L$ and $|y| < C_R$.
The proof is done in two steps. In the first step we will prove that for every pair $(\ell, r) \in B^{(D_L,D_R)}(u)$ there is a pair $(\ell', r') \in B_{x,y}^{(D_L,D_R)}(w)$ such that $\varphi(\ell') \in A^*\ell x$ and $\varphi(r') \in yrA^*$. In the second step we

will prove that two distinct pairs in $B^{(D_L,D_R)}(u)$ cannot be related to the same pair in $B^{(D_L,D_R)}(w)$.

For the first part of the proof, let $(\ell, r) \in B^{(D_L,D_R)}(u)$, i.e., $\ell ur \in \mathcal{L}(\mathbf{u})$ with $|\ell| = D_L$ and $|r| = D_R$. Then,

$$\ell ur = \ell x\varphi(w)yr \in \mathcal{L}(\mathbf{u})$$

and since the language is bi-extendable there are words ℓ', r' such that $\ell' wr' \in \mathcal{L}(\mathbf{u})$, $|\ell'| = D_L$, $|r'| = D_R$. Let us remind that because of Remark 4.3.10 $|\varphi(\ell')| \geq D_L + C_L - 1$ and $|\varphi(r')| \geq D_R + C_R - 1$. Thus,

$$\varphi(\ell' wr') = \varphi(\ell')\varphi(w)\varphi(r') = u_1\ell x\varphi(w)yru_2 \in \mathcal{L}(\mathbf{u})$$

for some words $u_1, u_2 \in A^*$. The last follows from the fact that $|x| \leq C_L$ and $|\ell| = D_L$, which implies that

$$|x| + |\ell| < C_L + D_L \leq C_L + D_L - 1 = |\varphi(\ell')|.$$

Thus, there is a pair $(\ell', r') \in B_{x,y}^{(D_L,D_R)}(w)$. with $\varphi(\ell') \in A^*\ell x$ and $\varphi(r') \in yrA^*$.

We now move to the second part of the proof. We proceed by contradiction. Let the extension pairs $(\ell_1, r_1), (\ell_2, r_2) \in B^{(D_L,D_R)}(u)$ be distinct and let us also suppose that both of them result from the pair $(\ell', r') \in B_{x,y}^{(D_L,D_R)}(w)$. Since $\ell' wr' \in \mathcal{L}(\mathbf{u})$ with $|\ell'| = D_L$ and $|r'| = D_R$, then $\varphi(\ell')\varphi(w)\varphi(r') \in \mathcal{L}(\mathbf{u})$ with $|\varphi(\ell')| \geq D_L + C_L - 1$ and $|\varphi(r')| \geq D_R + C_R - 1$ (Remark 4.3.10). Thus, there are words words u_1, u_2 such that

$$u_1x\varphi(w)yu_2 \in \mathcal{L}(\mathbf{u})$$

with

$$|u_1| = |\varphi(\ell')| - |x| \geq (D_L + C_L) - C_L \geq D_L$$

and

$$|u_2| = |\varphi(r')| - |y| \geq (D_R + C_R) - C_R \geq D_R.$$

Therefore, the words ℓ_1, ℓ_2 are suffixes of length D_L of u_1 and the words r_1, r_2 are prefixes of length D_R of the word u_2, which implies that $\ell_1 = \ell_2$ and $r_1 = r_2$ respectively, which leads to a contradiction.

Lastly, combining the two parts we can conclude that

$$\mathrm{Card}(B^{(D_L,D_R)}(u)) \leq \mathrm{Card}(B_{x,y}^{(D_L,D_R)}(w)).$$

□

The following corollary is a direct consequence of the proof of Lemma 4.3.12.

Corollary 4.3.13. Let $u \in \mathcal{L}(\mathbf{u})$ with antecedent $w \in \mathcal{L}(\mathbf{u})$ such that $u = x\varphi(w)y$ and constants D_L, D_R as described in Remark 4.3.10. Then,

$$B^{(D_L,D_R)}(u) = \left\{ (\ell, r) \mid \exists\, (\ell', r') \in B_{x,y}^{(D_L,D_R)}(w) : \begin{array}{l} \varphi(\ell') \in A^*\ell x \\ \varphi(r') \in yrA^* \end{array} \right\}.$$

Corollary 4.3.14. Let $u \in \mathcal{L}(\mathbf{u})$ with antecedent $w \in \mathcal{L}(\mathbf{u})$ and constants D_L, D_R as described in Remark 4.3.10. Then,

$$\mathrm{Card}(B^{(D_L,D_R)}(u)) \leq \mathrm{Card}(B^{(D_L,D_R)}(w)).$$

The above corollary follows directly from Lemma 4.3.12 and the fact that $\mathrm{Card}(B_{x,y}^{(D_L,D_R)}(w)) \leq \mathrm{Card}(B^{(D_L,D_R)}(w))$ for all words $w \in \mathcal{L}(\mathbf{u})$. The last inequality holds since the word w may have right extensions $\ell \in L^{(D_L)}$ with $\varphi(\ell) \notin A^*x$ and right extensions $r \in R^{D_R}$ with $\varphi(r) \notin yA^*$.

Lemma 4.3.15. Let φ be a morphism with a fixed point $\mathbf{u}$. Let $w \in \mathcal{L}(\mathbf{u})$ be a bispecial factor such that $\varphi(w)$ admits an antecedent. If $((\ell, r), (\ell', r')) \in B^{(D_L,D_R)}(w) \times B^{(D_L,D_R)}(w)$ is a left special pair (resp. right special pair), then $|f_L(\ell, \ell')| < C_L$ (resp. $|f_R(r, r')| < C_R$).

Proof. Let us first prove that for a left special pair $((\ell, r), (\ell', r'))$, we have $|f_L(\ell, \ell')| < C_L$. The proof is symmetric for a right special pair.

The result is trivial if φ is right marked, since in this case $f_L(\ell, \ell') = \varepsilon$ always, so let us assume it is not. Let us consider the word $W = f_L(\ell, \ell')\varphi(w)$ and let us write $\ell = \ell_1 \cdots \ell_{D_L}$ and $\ell' = \ell'_1 \cdots \ell'_{D_L}$, with $\ell_i, \ell'_j \in A$ for all i, j. We also fix $\ell_0 = \ell'_0 = \varepsilon$. By definition of f_L, there exist $m, n \in \{0, \ldots, D_L\}$ such that $f_L(\ell, \ell') = \alpha\varphi(\ell_{(m,D_L]}) = \beta\varphi(\ell'_{(n,D_L]})$ with α, β respectively suffix of $\varphi(\ell_m)$ and $\varphi(\ell'_n)$. We furthermore assume without loss of generality that $\alpha \neq \varphi(\ell_m)$ whenever $m \neq 0$ and that $\beta \neq \varphi(\ell_n)$ whenever $n \neq 0$. We thus get the two following external interpretations of W:

$$(\ell_m, \ell_{(m,D_L]}w, \varepsilon) \text{ and } (\ell_n, \ell_{(n,D_L]}w, \varepsilon).$$

Since $\varphi(w)$ admits an antecedent and it is a factor of W, by Remark 4.2.6. Thus, the word W admits an antecedent w' such

that $W = x\varphi(w')y$ with $|x| < C_L$ (Proposition 4.3.8). More precisely, for any internal interpretation (s, v, p) of W, there exist $i, j \in \{0, \ldots, |v| - 1\}$, $i \leq j$, such that $w' = v_{[i,j]}$, $x = s\varphi(v_{[0,i)})$ and $y = \varphi(v_{(j,|v|-1]})p$. Considering the two previous internal interpretations of W, if $|f_L(\ell, \ell')| \geq C_L$, there exist $m' \in \{m, \ldots, D_L\}, n' \in \{n, \ldots, D_L\}$ and $j \in \{1, \ldots, |w|\}$ such that $x = \alpha\varphi(\ell_{(m,m')}) = \beta\varphi(\ell'_{(n,n')})$ and $w' = \ell_{[m',D_L]}w_{[1,j]} = \ell'_{[n',D_L]}w_{[1,j]}$. This implies that $\ell_{D_L} = \ell'_{D_L}$, contradicting the fact that $((\ell, r), (\ell', r'))$ is a left special pair. We conclude that $|f_L(\ell, \ell')| < C_L$. □

Corollary 4.3.16. Let φ be a morphism with a fixed point $\mathbf{u}$. Let $w \in \mathcal{L}(\mathbf{u})$ be a bispecial word such that $\varphi(w)$ admits an antecedent. If $((\ell_1, r_1), (\ell_2, r_2)) \in B^{(D_L,D_R)}(w) \times B^{(D_L,D_R)}(w)$ is a bispecial pair, then

$$|f_L(\ell_1, \ell_2)| < C_L \text{ and } |f_R(r_1, r_2)| < C_R.$$

Furthermore, the word $u = f_L(\ell_1, \ell_2)\varphi(w)f_R(r_1, r_2) \in \mathcal{L}(\mathbf{u})$ is bispecial and

$$B^{(D_L,D_R)}(u) \supset \left\{ (\ell, r) \mid \exists\, (\ell_1, r_1) \in B^{(D_L,D_R)}(w) : \begin{array}{c} \varphi(\ell_1) \in A^*\ell f_L(\ell_1, \ell_2) \\ \varphi(r_1) \in f_R(r_1, r_2)rA^* \end{array} \right\}. \tag{4.3.1}$$

Proof. Let φ be a morphism with a fixed point $\mathbf{u}$. Let $w \in \mathcal{L}(\mathbf{u})$ be a bispecial word, then by Proposition 4.3.4 there is a bispecial pair in $B^{(D_L,D_R)}(w) \times B^{(D_L,D_R)}(w)$. Since this pair is by definition left-special and right-special, it follows by Lemma 4.3.15 that $|f_L(\ell_1, \ell_2)| < C_L$ and $|f_R(r_1, r_2)| < C_R$ respectively.

Let now $((\ell_1, r_1), (\ell_2, r_2)) \in B^{(D_L,D_R)}(w) \times B^{(D_L,D_R)}(w)$ be a bispecial pair. We will prove that the word $u = f_L(\ell_1, \ell_2)\varphi(w)f_R(r_1, r_2) \in \mathcal{L}(\mathbf{u})$ is bispecial. Since $\ell_i w r_i \in \mathcal{L}(\mathbf{u})$ for $i \in \{1, 2\}$, then $\varphi(\ell_i)\varphi(w)\varphi(r_i) \in \mathcal{L}(\mathbf{u})$ for all $i \in \{1, 2\}$. Also, since $|\ell_1| = |\ell_2| = D_L$, because of the definition of D_L (Remark 4.3.10), we have that $|\varphi(\ell_i)| \geq D_L + C_L - 1$. Similarly, since $|r_1| = |r_2| = D_R$, we have that $|\varphi(r_i)| \geq D_R + C_R - 1$ for all $i \in \{1, 2\}$. We conclude that there are words $h_i, s_i \in \mathcal{L}(\mathbf{u})$ for $i \in \{1, 2\}$ such that,

$$\varphi(\ell_i) = h_i f_L(\ell_1, \ell_2) \text{ and } \varphi(r_i) = f_L(r_1, r_2)s_i.$$

It can be seen that the words h_i and s_i have length at most D_L and D_R respectively. Indeed,

$$|h_i| = |\varphi(\ell_i)| - |f_L(\ell_1, \ell_2)| > (D_L + C_L - 1) - C_L = D_L - 1 \geq D_L$$

and

$$|s_i| = |\varphi(r_i)| - |f_R(r_1, r_2)| > (D_R + C_R - 1) - C_R =\geq D_R.$$

Furthermore, h_1, h_2 have distinct last letter and s_3, s_4 have distinct first letter, showing that u is bispecial. Equation (4.3.1) is immediate. □

Definition 4.3.17. If w and u are as in Corollary 4.3.16, we say that u is a *bispecial extended image* of w.

Observe that while any long enough word has a unique antecedent (by Proposition 4.2.8), a given bispecial factor may have several bispecial extended images. In the case of a *bifix morphism* (i.e. a prefix and suffix morphism), any long enough bispecial word v has a unique bispecial extended image u, and the antecedent of u is exactly v (see [48]).

4.4 Graph of extension graphs

Let us remind that $\varphi : A^* \to A^*$ is a non-erasing non-pushy strongly two-sided recognizable morphism on the substitutive language $\mathcal{L}(\mathbf{u})$ for a constant $L > 0$. In this section we will introduce a finite graph whose set of vertices is the set of extension graphs of all bispecial words in the substitutive language $\mathcal{L}(\mathbf{u})$.

Let $\mathcal{BE} = \left\{\mathcal{E}^{(D_L,D_R)}(u) \mid u \in \mathcal{L}(\mathbf{u}), u \text{ bispecial}\right\}$. Any generalized extension graph contains at most $\mathrm{Card}(A^{D_L}) + \mathrm{Card}(A^{D_R})$ vertices, so this set is finite. Our aim is to give an algorithm that describe, among all possible graphs with vertices in $A^{D_L} \cup A^{D_R}$, which ones belong to $\mathcal{BE}$ and that also describe the relation between antecedents and decedents.

We first consider the subset of *initial generalized extension graphs* $\mathcal{IE} \subset \mathcal{BE}$ defined as $\mathcal{IE} = \left\{\mathcal{E}^{(D_L,D_R)}(v) \in \mathcal{BE} \mid |v| \leq 2L + ||\varphi||\right\}$. From Proposition 4.2.8, we cannot guarantee the existence of the antecedent for the set of words $\mathcal{L}(\mathbf{u}) \cap A^{<2L+||\varphi||}$.

Example 4.4.1. The Fibonacci morphism φ has constants $D_L = 1$ and $D_R = 2$, as it has been explained in Example 4.3.11. The set $\mathcal{IE}$ contains the generalized extension graphs of the bispecial words of length smaller than $2L + ||\varphi|| = 4$, i.e., of $\varepsilon, 0$ and 010 (Figure 4.8).

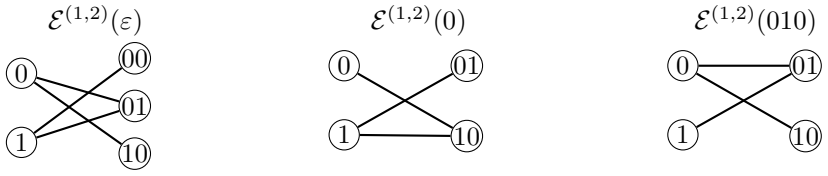


Figure 4.8: The graphs $\mathcal{E}^{(1,2)}(\varepsilon)$, $\mathcal{E}^{(1,2)}(0)$ and $\mathcal{E}^{(1,2)}(010)$.

Because of strong two-sided recognizability, any word longer than $2L + ||\varphi||$ can be uniquely factorized as $u = x\varphi(w)y$ with $w \in \mathcal{L}(\mathbf{u})$ being the antecedent of u. Moreover, Corollary 4.3.13 implies that the sets of (left-, right- and bi-) extensions of a bispecial word u depend only on the set of extensions of its antecedent w and not on the word w itself.

We define a relation $\mathcal{R}$ on $\mathcal{BE} \times \mathcal{BE}$ as follows:

$$\left(\mathcal{E}^{(D_L,D_R)}(u), \mathcal{E}^{D_L,D_R}(u')\right) \in \mathcal{R} \iff \begin{cases} |u|, |u'| \geq 2L + ||\varphi|| \\ \mathcal{E}^{(D_L,D_R)}(u) = \mathcal{E}^{(D_L,D_R)}(u') \end{cases}$$

Let $[\mathcal{E}^{(D_L,D_R)}(u)]$ be the corresponding class of the graph $\mathcal{E}^{(D_L,D_R)}(u)$.

Example 4.4.2. Let us consider the Fibonacci morphism φ and the two bispecial words $u = 010010$ and $v = 01001010010$ with length more than 4 (Example 4.4.1). The generalized extension graphs of the classes $G = [\mathcal{E}^{(1,2)}(u)]$ and $H = [\mathcal{E}^{(1,2)}(v)]$ are shown in Figure 4.9.

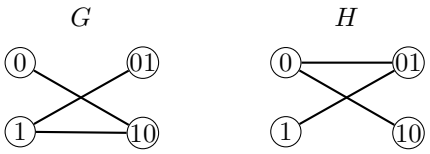


Figure 4.9: The graphs of the classes $G = [\mathcal{E}^{(1,2)}(u)]$ (on the left) and $H = [\mathcal{E}^{(1,2)}(v)]$ (on the right).

In the previous section we defined the bispecial extended image of a bispecial word in a substitutive language. A bispecial word v may have more than one bispecial extended images. This number is nevertheless finite (a trivial bound is given by $\binom{\mathrm{Card}(B^{(D_L,D_R)}(v))}{2}$. On the other hand, any long enough bispecial word u in a strongly two-sided recognizable language has a unique antecedent w and, in this case, $\mathrm{Card}(B^{(D_L,D_R)}(u)) \leq \mathrm{Card}(B^{(D_L,D_R)}(w))$.

Let us now show that given another bispecial word v such that $u = s\varphi(v)p$ is a bispecial extended image of v for some words s, p, one has $\mathrm{Card}(B_{x,y}^{(D_L,D_R)}(v)) \leq \mathrm{Card}(B_{x,y}^{(D_L,D_R}(w))$. Let us consider a pair $(\ell, r) \in B_{x,y}^{(D_L,D_R)}(v)$. Since w is factor of any preimage of u one has $v = twq$ for some words t, q. It follows that there is a pair $(\ell', r') \in B_{x,y}^{(D_L,D_R)}(w)$ with $\ell' = \ell''t$ and $r' = qr''$, for some words ℓ'', r''. This shows the claim.

Let us call *adequate* a substitutive language in which if a bispecial word w_2 has extension graph in the same class of graphs as another word w_1, i.e., $[\mathcal{E}^{(D_L,D_R)}(w_1)] = [\mathcal{E}^{(D_L,D_R)}(w_2)]$, then if u_1 is a bispecial extended image of w_1 that has w_1 as antecedent and it can be written as $u_1 = x\varphi(w_1)y$, then there is a bispecial extended image u_2 of w_2 that has the word w_2 as antecedent which is written as $u_2 = x\varphi(w_2)y$ and $[\mathcal{E}^{(D_L,D_R)}(u_1)] = [\mathcal{E}^{(D_L,D_R)}(u_2)]$.

We say that a morphism with a fixed point $\mathbf{u}$ is *adequate* if the language $\mathcal{L}(\mathbf{u})$ is adequate.

A substitutive language $\mathcal{L}(\mathbf{u})$ will be called *ultimately adequate* if there is an integers $N > 0$ such that the previous definition holds for all words longer than N.

A trivial category of morphisms that are adequate is the class of the morphisms that are *stable*.

Definition 4.4.3. A morphism φ is *stable* if there are words $u, v \in A^*$ such that for any two distinct letters $a, b \in A$ holds that $f_L(a, b) = u$ and $f_R(a, b) = v$.

Example 4.4.4. Let $\zeta : \{0,1,2\}^* \to \{0,1,2\}^*$ be the following morphism,

$$\begin{aligned} \zeta : 0 &\mapsto 020, \\ 1 &\mapsto 00, \\ 2 &\mapsto 01210. \end{aligned}$$

It can be verified that ζ is stable since $f_L(0,1) = f_L(0,2) = f_L(1,2) = 0$ and $f_R(0,1) = f_R(0,2) = f_R(1,2) = 0$. However, the morphism $\eta : \{0,1,2\}^* \to \{0,1,2\}^*$ defined as follows,

$$\begin{aligned} \eta : 0 &\mapsto 020, \\ 1 &\mapsto 00, \\ 2 &\mapsto 0210, \end{aligned}$$

is not stable since $f_R(0,1) = 0 \neq f_R(0,2) = 02$.

Remark 4.4.5. If a morphism is right-marked and left-marked then it is stable.

Indeed, if a morphism is right-marked and left-marked then for any two letter $a, b \in A$ holds that $f_L(a,b) = f_R(a,b) = \varepsilon$ and thus the morphism is stable.

Proposition 4.4.6. Let φ be a morphism on an alphabet A satisfying the usual conditions. If the morphism φ is stable, then φ is adequate.

Proof. Let φ be a morphism with fixed point $\mathbf{u}$ and let words $u, v \in A^*$ such that for any two distinct letters $a, b \in A$, $f_L(a,b) = u$ and $f_R(a,b) = v$. Let w be a bispecial word in $\mathcal{L}(\mathbf{u})$. Since for any two distinct right extensions $r_1, r_2 \in R_{\mathcal{L}(\mathbf{u})}(w)$ it holds that $f_R(r_1, r_2) = v$ and since for any two distinct left extensions $\ell_1, \ell_2 \in L_{\mathcal{L}(\mathbf{u})}(w)$ it holds that $f_L(\ell_1, \ell_2) = u$, then the only possible bispecial extended image of w is $w' = u\varphi(w)v$. In order to prove that w is the antecedent of w' we should suppose that $|w'| \geq 2L + ||\varphi||$, since otherwise we cannot assure the existence of the antecedent. Let us now suppose that a word z distinct to the word w is the antecedent of w'. Thus, since the antecedent is the longest common factor of all the preimages it holds that the word z must be a factor of w. Let $\alpha, \beta \in A^*$, with at least one of them the non-empty word, be such that $w = \alpha z \beta$. Since the bispecial word z is the antecedent of u, there are $x = f_L(\ell_1', \ell_2') = u$ and $y = f_R(r_1', r_2') = v$, with $\ell_1', \ell_2' \in L^{D_L}_{\mathcal{L}(\mathbf{u})}(z)$ and $r_1', r_2' \in R^{D_R}_{\mathcal{L}(\mathbf{u})}(z)$, such that

$$w' = u\varphi(z)v.$$

We also have that $w' = u\varphi(w)v$, from which follows that,

$$u\varphi(w)v = u\varphi(z)v$$

$$\Rightarrow \varphi(z) = \varphi(x) = \varphi(\alpha z \beta)$$

that leads to a contradiction since at least one of the words α, β is non-empty and the morphism is non-erasing. Thus, every word w has a unique decedent that is the word $w' = u\varphi(w)v$, and this word has w as antecedent. The graph $\mathcal{E}^{(D_L, D_R)}(u\varphi(w)v)$ depends only on the extension pairs of the antecedent w, or equivalently it

depends only on graph $\mathcal{E}^{(D_L,D_R)}(w)$. Every word w'' with extension graph in the class $[\mathcal{E}(w)]$ has exactly the same extension pairs and therefore its bispecial extended image $u\varphi(w'')v$ has the word w'' as the antecedent and $[\mathcal{E}^{(D_L,D_R)}(u\varphi(w'')v] = [\mathcal{E}^{(D_L,D_R)}(u\varphi(w)v)]$, which proves that the morphism φ is adequate. □

Let us now prove that the bifix morphisms are also adequate.

Proposition 4.4.7. Let φ be a morphism on an alphabet A satisfying the usual conditions. If the morphism φ is bifix, then φ is adequate.

Proof. Let w be a bispecial word in the substitutive language $\mathcal{L}(\mathbf{u})$, where $\mathbf{u}$ is a fixed point of the morphism φ. Let the word $u = x\varphi(w)y$ be a bispecial extended image of w. We suppose that the length of u is more than $2L + ||\varphi||$ and we firstly prove that w is the antecedent of u. Let us prove it by contradiction, supposing that the antecedent of u is a word w' distinct to w. Therefore the word w' is a proper factor of w. Let $w = \alpha w'\beta$, where at least one of the $\alpha, \beta \in A^*$ is the non-empty word. Thus,

$$\begin{aligned} u &= x\varphi(w)y \\ &= x\varphi(\alpha w'\beta)y \\ &= (x\varphi(\alpha))\varphi(w')(\varphi(\beta)y) \\ &= x'\varphi(w')y' \end{aligned}$$

with $x' = x\varphi(\alpha)$, $y = \varphi(\beta)y$ and with at least one of the words α, β a non-empty word. Let us suppose that $\alpha \neq \varepsilon$ (resp. $\beta \neq \varepsilon$). We have that $x' = f_L(\ell_1, \ell_2)$, for ℓ_1, ℓ_2 left extensions in a left special pair $((\ell_1, r_1), (\ell_2, r_2)) \in B^{(D_L,D_R)}_{x',y'}(w')$. Because of the definition of the antecedent we have that $(x', \varphi(w')y')$ is a strong synchronization point of u. We also have that $x' = x\varphi(\alpha) = x\varphi(\alpha_{[0,|\alpha|-1)}\varphi(\alpha_{|\alpha|-1}$ where $\alpha_i \in A$, for $i \in \{0, \ldots, |\alpha| - 1\}$. Since there is a decomposition of u as $u = x\varphi(\alpha_{[0,|\alpha|-1)}\varphi(\alpha_{|\alpha|-1}\varphi(w')y'$ where $(x', \varphi(w')y')$ is a synchronization point and since the morphism is bifix, it follows that $(x\varphi(\alpha_{[0,|\alpha|-1)}, \varphi(\alpha_{|\alpha|-1}\varphi(w')y')$ is a strong synchronization point too. The previous implies that there is a word $\alpha_{|\alpha|-1}w' \neq w'$ that satisfies the definition of the antecedent which leads to a contradiction.

Let us now recall that words with extension graphs in the same class have the same number of bispecial extended images because they depend only on the bispecial pairs, or equivalently they depend

on the edges of the graph. Since all bispecial extended images of a word w in a class of graphs $[\mathcal{E}^{(D_L,D_R)}(w)]$ have the word w as the antecedent, then their extension graph depends only on the graph of w. Therefore the same holds for any other word v such that $[\mathcal{E}^{(D_L,D_R)}(v)] = [\mathcal{E}^{(D_L,D_R)}(w)]$, which implies that for each bispecial extended image $u = x\varphi(w)y$ of w there is a bispecial extended image $u' = x\varphi(v)y$ for which holds that $[\mathcal{E}^{(D_L,D_R)}(u)] = [\mathcal{E}^{(D_L,D_R)}(u')]$. Thus, the bifix morphism φ is adequate. □

Definition 4.4.8. Let us consider the labeled directed graph $\mathcal{K}(\mathcal{L}(\mathbf{u}))$, having set of vertices $\mathcal{IE} \sqcup (\mathcal{BE}/\mathcal{R})$ and set of edges defined as follows: there is an edge in $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ going from G to H, labeled $(f_L(\ell_1, \ell_2), f_R(r_1, r_2))$, if

1. $G = \mathcal{E}^{(D_L,D_R)}(u) \in \mathcal{IE}$ or $G = [\mathcal{E}^{(D_L,D_R)}(u)] \in (\mathcal{BE}/\mathcal{R})$,
2. $H = [\mathcal{E}^{(D_L,D_R)}(v)] \in (\mathcal{BE}/\mathcal{R})$,
3. there exist a bispecial pair $((\ell_1, r_1), (\ell_2, r_2))$ in $\mathcal{E}^{(D_L,D_R)}(u)$ such that
   (a) $L^{(D_L)}(v) = \left\{t \mid \exists\, s \in L^{(D_L)}(u) : \varphi(s) \in A^* t f_L(\ell_1, \ell_2)\right\}$,
   (b) $R^{(D_R)}(v) = \left\{q \mid \exists\, p \in R^{(D_R)}(u) : \varphi(p) \in f_R(r_1, r_2) q A^*\right\}$,
   (c) $B^{(D_L,D_R)}(v) = \left\{(t, q) \mid \exists\, (s, p) \in B^{(D_L,D_R)}(u) : \begin{array}{l} \varphi(s) \in A^* t f_L(\ell_1, \ell_2) \\ \varphi(p) \in f_R(r_1, r_2) q A^* \end{array}\right\}$,
4. the word u is the antecedent of the word $v = f_L(\ell_1, \ell_2)\varphi(u)f_R(r_1, r_2)$ in $\mathcal{L}(\mathbf{u})$.

Theorem 4.4.9. *Let φ be a strongly two-sided recognizable morphism on the adequate substitutive language of a fixed point $\mathbf{u}$. The graph $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ is finite and computable.*

Proof. The subgraph of $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ involving vertices in $\mathcal{IE}$ is computable since it only involves words of bounded length and this bound is computable by Theorem 3.4.5. For any integer n, the number of possible graphs such that $\mathrm{Card}(B^{(D_L,D_R)}(w)) = n$ for a certain word w is finite. Moreover, for any edge from u to v in the graph we have $\mathrm{Card}(B^{(D_L,D_R)}(v)) \leq \mathrm{Card}(B^{(D_L,D_R)}(u))$, since u is the antecedent of v (Corollary 4.3.14). It follows from Corollary 4.2.10 that the maximal value of $\mathrm{Card}(B^{(D_L,D_R)}(w))$, for $w \in \mathcal{L}(\mathbf{u})$, is attained in the initial part of the graph (i.e. for a w of length less then $2L + ||\varphi||$). Thus, the set $(\mathcal{BE}/R)$ is also finite and computable. □

The following theorem proves that all possible extension graphs of the bispecial words in an adequate strongly two-sided recognizable substitutive language appear in the graph $\mathcal{K}$.

Theorem 4.4.10. *Let φ be a strongly two-sided recognizable morphism on the adequate substitutive language of a fixed point $\mathbf{u}$. If v is a bispecial word in $\mathcal{L}(\mathbf{u})$, then $[\mathcal{E}^{(D_L,D_R)}(v)] \in \mathcal{K}(\mathcal{L}(\mathbf{u}))$.*

Proof. Let the word $v \in \mathcal{L}(\mathbf{u})$ be bispecial. We will prove by recurrence on the length of the word v that the algorithm for the construction of the graph of extension graphs $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ includes the construction of the graph $\mathcal{E}^{(D_L,D_R)}(v)$. If $|v| \leq 2L+||\varphi||$, then $\mathcal{E}^{(D_L,D_R)}(v) \in \mathcal{IE}$ and therefore $\mathcal{E}^{(D_L,D_R)}(v) \in \mathcal{K}(\mathcal{L}(\mathbf{u}))$. If $|v| > 2L+||\varphi||$, then because of strong two-sided recognizability, it holds that the word v has an antecedent. Let u be the antecedent of v, with $|u| < |v|$. By the recurrence hypothesis we have that $[\mathcal{E}^{(D_L,D_R)}(u)] \in \mathcal{K}(\mathcal{L}(\mathbf{u}))$. Since the word v is bispecial we have that v is bispecial extended image of its antecedent u, and it is of the form $v = f_L(\ell_1, \ell_2)\varphi(u)f_R(r_1, r_2)$, with $((\ell_1, r_1), (\ell_2, r_2)) \in B^{(D_L,D_R)}(u) \times B^{(D_L,D_R)}(u)$ a bispecial pair of u. Regarding the constructions of the graph $\mathcal{E}^{(D_L,D_R)}(v)$ there are the two possible cases. The first case is that there is a word $z \in \mathcal{L}(\mathbf{u})$ such that $\mathcal{E}^{(D_L,D_R)}(z) \in \mathcal{K}(\mathcal{L}(\mathbf{u}))$ and $[\mathcal{E}^{(D_L,D_R)}(v)] = [\mathcal{E}^{(D_L,D_R)}(z)]$. In this case, since the morphism is adequate, it is enough to consider the graph of the word z and the bispecial extended images only for z and not those of v. The second case is that the graph $\mathcal{E}^{(D_L,D_R)}(v)$ is not in the same class with any of the graphs that are so far constructed in $\mathcal{K}(\mathcal{L}(\mathbf{u}))$, in the later case the class $[\mathcal{E}^{(D_L,D_R)}(v)]$ should be added in $\mathcal{K}(\mathcal{L}(\mathbf{u}))$. In both cases we draw an edge from the class $[\mathcal{E}^{(D_L,D_R)}(u)]$ to the class $[\mathcal{E}^{(D_L,D_R)}(v)]$. Thus, it has been proved that $[\mathcal{E}^{(D_L,D_R)}(v)] \in \mathcal{K}(\mathcal{L}(\mathbf{u}))$. □

Example 4.4.11. Let $\mathcal{L}(\mathbf{u})$ be the substitutive language of the Fibonacci morphism φ_F, that is strongly two-sided recognizable for a constant $L = 1$ (Example 3.3.29). Let us also remind that for the Fibonacci morphism we have that $D_L = 1$ and $D_R = 2$ (Example 4.3.11). It can be verified that the Fibonacci morphism is stable since $\varphi_L(0, 1) = \varepsilon$ and $\varphi_R(0, 1) = 0$. Thus, the morphism φ_F is adequate and it is enough to construct the graph $\mathcal{K}$.

The initial generalized extension graphs for the Fibonacci morphism φ_F are given in Example 4.4.1. Let us remind that the length of the words that we consider in order to construct the subgraph

$\mathcal{IE}$ is less than 4, since $4 = 2L + ||\varphi||$. Let us now construct the rest of the graph $\mathcal{K}(\mathcal{L}(\mathbf{u}))$, as seen in the Figure 4.10. We start by the word ε. It can be verified that ε has only one extended bispecial image, that is the word 0. Indeed, for all bispecial pairs $((\ell_1, r_1), (\ell_2, r_2))$ in $B^{(1,2)}(\varepsilon) \times B^{(1,2)}(\varepsilon)$ it holds that $f_L(\ell_1, \ell_2) = \varepsilon$ and $f_R(r_1, r_2) = 0$, which implies that the only bispecial extended image of ε is the word $f_L(\ell_1, \ell_2)\varphi_F(\varepsilon)f_R(r_1 r_2) = \varepsilon\varepsilon 0$. However the word 0 is shorter than 4 and therefore $\mathcal{E}^{(1,2)}(0) \in \mathcal{IE}$ and thus we do not draw an edge. We follow now the same procedure for the word 0, which for the same reasoning has also one bispecial extended image, that is the word 010. Since $|010| = 3 \leq 4$ and the graph $\mathcal{E}^{(1,2)}(010)$ is in $\mathcal{IE}$ we also do not draw an edge. We move now to the word 010 which has for bispecial extended image the word $u = 010010$, with length $|u| = 6 > 4$. Before drawing the graph we should verify that the word 010 is the antecedent of u. The word u has unique internal interpretation $(s, v, p) = (\varepsilon, 010, 0)$ and therefore the antecedent is the word $v = 010$. Thus, we should add the graph $\mathcal{E}^{(1,2)}(u)$ in $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ and we should design an edge from $\mathcal{E}^{(1,2)}(010)$ to $\mathcal{E}^{(1,2)}(u)$ labeled by $(\varepsilon, 0)$. Similarly, we can verify that there is only one extended image of u and that it is the word $w = \varepsilon\varphi_F(u)0 = 01001010010$, and that the antecedent of w is u. Since there is no graph in $\mathcal{K}(\mathcal{L}(\mathbf{u})) \setminus \mathcal{IE}$ that is in the same class as $\mathcal{E}^{(1,2)}(w)$, we design the graph $\mathcal{E}^{(1,2)}(w)$ and an edge from $\mathcal{E}^{(1,2)}(u)$ to $\mathcal{E}^{(1,2)}(w)$ labeled by $(\varepsilon, 0)$. Then we compute the extended image z of w and we verify that w is the antecedent of z. We have that $z = \varepsilon\varphi_F(w)0 = \varphi_F(01001010010)0 = 0100101001001010010$. The graph $\mathcal{E}^{(1,2)}(z)$ is identical to the graph $\mathcal{E}^{(1,2)}(u)$ and therefore they are in the same class $[\mathcal{E}^{(1,2)}(u)]$. Thus, we draw an edge from the graph $\mathcal{E}^{(1,2)}(w)$ to the graph $\mathcal{E}^{(1,2)}(u)$, labeled $(\varepsilon, 0)$. We do not have any other graphs to calculate and thus we have the whole graph $\mathcal{K}(\mathcal{L}(\mathbf{u}))$.

Let us define the function $\Phi : \mathbb{N} \to \mathbb{N}$ such that $\Phi(n) = n||\varphi|| + C_L + C_R$. We also use the notation $\Phi^k(n) = \Phi^{k-1}(\Phi(n))$ for all $k > 1$.

Proposition 4.4.12. Let φ be strongly two-sided recognizable morphism with the fixed point $\mathbf{u}$ for a constant $L > 0$. Let u be a bispecial factor of length n. Then all bispecial extended images of u have length at most $\Phi(n)$.

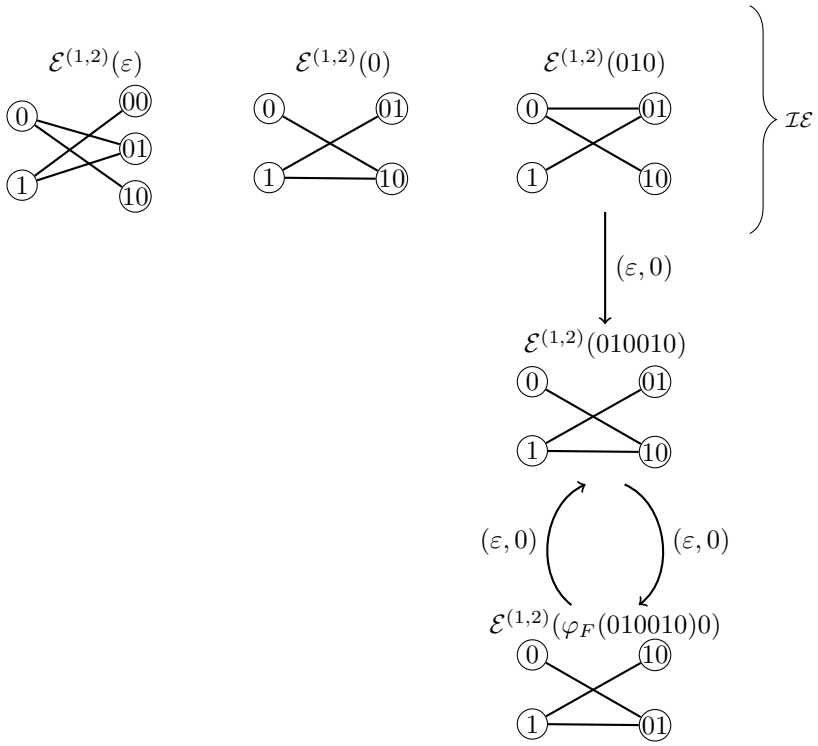


Figure 4.10: The graphs $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ for the Fibonacci word $\mathbf{u}$.

Proof. Since u is bispecial, there is a bispecial pair $((\ell_1, r_1), (\ell_2, r_2))$ in $B^{(D_L, D_R)}(u)$ which leads to a bispecial image $v = f_L(\ell_1, \ell_2)\varphi(u)f_R(r_1, r_2)$. By Lemma 4.3.8, $|f_L(\ell_1, \ell_2)| < C_L$ and $|f_R(r_1, r_2)| < C_R$. Hence,

$$|f_L(\ell_1, \ell_2)\varphi(u)f_R(r_1, r_2)| < C_L + |\varphi(u)| + C_R \leq C_L + n||\varphi|| + C_R = \Phi(n).$$

□

The following theorem gives an upper bound on the length of the words that we considered in order to construct the graph $\mathcal{K}$, but let us first define the constant δ as follows.

Let $\delta = \delta(\mathcal{K}(\mathcal{L}(\mathbf{u})))$ be the length of the longest simple path in the graph $\mathcal{K}(\mathcal{L}(\mathbf{u}))$

Theorem 4.4.13. *Let φ be a strongly two-sided recognizable morphism for a constant $L > 0$ on the adequate substitutive language $\mathcal{L}(\mathbf{u})$. For any long enough bispecial word $v \in \mathcal{L}(\mathbf{u})$ there is a*

bispecial word $z \in \mathcal{L}(\mathbf{u})$ *with* $|z| \leq \Phi^{\delta}(2L + ||\varphi|| - 1)$ *such that* $[\mathcal{E}^{(D_L,D_R)}(v)] = [\mathcal{E}^{(D_L,D_R)}(z)]$.

Proof. It follows from the definition of δ that every possible extension graph of a bispecial word in the language $\mathcal{L}(\mathbf{u})$ must appear either in the initial graphs $\mathcal{IE}$ if the word is shorter that $2L + ||\varphi||$, where $L > 0$ is the constant of recognizability, or it will appear in a path of length at most δ starting from an initial graph $G \in \mathcal{IE}$ (and not passing through $\mathcal{IE}$ again). From Proposition 4.4.12 it follows that for any bispecial word v with length at least $2L + ||\varphi||$ it is possible to choose a word z of length at most $\Phi^{\delta}(2L + ||\varphi|| - 1)$ that has the exact same extension graph, i.e., $\mathcal{E}^{(D_L,D_R)}(v) \in [\mathcal{E}^{(D_L,D_R)}(z)]$. □

Example 4.4.14. Let φ_F be the Fibonacci morphism with fixed point the Fibonacci word. The only possible extension graphs of a bispecial word in $\mathcal{L}(\mathbf{u})$ are showed in Figure 4.10. It follows from the graph $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ that the length of the longest simple path is $\delta = 2$. Thus, for a bispecial word $u \in \mathcal{L}(\mathbf{u})$ there exists a bispecial word $v \in \mathcal{L}(\mathbf{u})$ of length $|v| \leq \Phi^2(3) = 21$, such that $\mathcal{E}^{(1,2)}(u) \in [\mathcal{E}^{(1,2)}(v)]$.

Corollary 4.4.15. Let φ be an adequate strongly two-sided recognizable morphism on a fixed point $\mathbf{u}$. Then it is decidable whether the substitutive language $\mathcal{L}(\mathbf{u})$ is a dendric set (resp. acyclic, connected, neutral).

We can actually decide any property of the substitutive language $\mathcal{L}(\mathbf{u})$ that depends only on the shape of the extension graphs of words in $\mathcal{L}(\mathbf{u})$. Given a particular extension graph, we can also decide whether there exists some word in $\mathcal{L}(\mathbf{u})$ that has that graph as extension graph and we can even describe all words in $\mathcal{L}(\mathbf{u})$ for which this is the case, as it can be seen in the following example.

Example 4.4.16. Let us work with the Fibonacci morphism $\varphi = \varphi_F$ with graph of extension graphs $\mathcal{K}(\mathcal{L}(\mathbf{u}))$ as seen in the Figure 4.10. Let $w = 010010$ for which the extension graph is already calculated. Since the class $[\mathcal{E}^{(1,2)}(w)]$ is in a loop of length 2, the words in the language $\mathcal{L}(\mathbf{u})$ whose graph is in the same class can be obtained calculating the 2^k-time bispecial extended image, for $k > 0$. More precisely, the word $w' = \varepsilon\varphi(\varepsilon\varphi(w)0)0$ is bispecial with extension graph in the class $[\mathcal{E}^{(1,2)}(w)]$. Similarly, the word $w'' = \varphi(\varepsilon\varphi(w')0)0$ is bispecial with graph in the class $[\mathcal{E}^{(1,2)}(w)]$.

All bispecial words with graphs in the class $[\mathcal{E}^{(1,2)}(w)]$ can be obtained in the same way.

The previous definition of the graph of extension graphs $\mathcal{K}$ can be generalized to the case of ultimately adequate morphisms as follows.

Definition 4.4.17. Let us consider the labeled directed graph $\tilde{\mathcal{K}}(\mathcal{L}(\mathbf{u}))$ of an ultimately adequate substitutive language of a fixed point $\mathbf{u}$ of a morphism φ. If $N \leq 2L + ||\varphi||$ then $\tilde{\mathcal{K}} = \mathcal{K}$. If $N > 2L + ||\varphi||$ we consider the labeled directed graph $\tilde{\mathcal{K}}(\mathcal{L}(\mathbf{u}))$ having set of vertices $\mathcal{IE} \sqcup \mathcal{IEE} \sqcup (\mathcal{BE}/\mathcal{R})$. The set $\mathcal{IE}$ contains the extension graphs of all bispecial words of length less that $2L+||\varphi||$, the set $\mathcal{IEE}$ is the set of extension graphs of all bispecial words with length at least $2L+||\varphi||$ and at most N, where N is the constant that satisfies the ultimate adequate propriety and $(\mathcal{BE}/\mathcal{R})$ is the class of graphs defined as in the previous case. The set of edges is defined as follows: there is an edge in $\tilde{\mathcal{K}}(\mathcal{L}(\mathbf{u}))$ going from G to H, labeled $(f_L(\ell_1, \ell_2), f_R(r_1, r_2))$, if

1. $G = \mathcal{E}^{(D_L,D_R)}(u) \in \mathcal{IE}$ or $G = \mathcal{E}^{(D_L,D_R)}(u) \in \mathcal{IEE}$ and $H = \mathcal{E}^{(D_L,D_R)}(v) \in \mathcal{IEE}$ or $H = [\mathcal{E}^{(D_L,D_R)}(v)] \in (\mathcal{BE}/\mathcal{R})$,
2. $G = [\mathcal{E}^{(D_L,D_R)}(u)] \in (\mathcal{BE}/\mathcal{R})$ and $H = [\mathcal{E}^{(D_L,D_R)}(v)] \in (\mathcal{BE}/\mathcal{R})$,
3. there exist a bispecial pair $((\ell_1, r_1), (\ell_2, r_2))$ in $\mathcal{E}^{(D_L,D_R)}(u)$ such that
   (a) $L^{(D_L)}(v) = \left\{t \mid \exists\, s \in L^{(D_L)}(u) : \varphi(s) \in A^* t f_L(\ell_1, \ell_2)\right\}$;
   (b) $R^{(D_R)}(v) = \left\{q \mid \exists\, p \in R^{(D_R)}(u) : \varphi(p) \in f_R(r_1, r_2) q A^*\right\}$;
   (c) $B^{(D_L,D_R)}(v) = \left\{(t, q) \mid \exists\, (s, p) \in B^{(D_L,D_R)}(u) : \begin{array}{l} \varphi(s) \in A^* t f_L(\ell_1, \ell_2) \\ \varphi(p) \in f_R(r_1, r_2) q A^* \end{array}\right\}$,
4. the word u is the antecedent of the word v in $\mathcal{L}(\mathbf{u})$.

Remark 4.4.18. Let us point out that in the sets $\mathcal{IE}$ and $\mathcal{IEE}$ there are no classes of graphs. In particular, even if two words u, v have exactly the same graph, we have to draw two different graphs $\mathcal{E}^{(D_R,D_L)}(v)$ and $\mathcal{E}^{(D_R,D_L)}(u)$.

Theorem 4.4.19. *Let φ be a strongly two-sided recognizable ultimately adequate morphism on the substitutive language of a fixed point $\mathbf{u}$. The graph $\tilde{K}(\mathcal{L}(\mathbf{u}))$ is finite and it contains all the classes of the graphs for all the bispecial words in $\mathcal{L}(\mathbf{u})$.*

Proof. The proof is the same as those of Theorem 4.4.9 and Theorem 4.4.10 except of the fact that it is added the subgraph $\mathcal{IEE}$. The subgraph $\mathcal{IEE}$ is also finite since there are finitely many bispecial words with length smaller than $N > 0$, where N is the smallest constant that satisfies the definition of ultimate adequety. □

Similarly with the $\mathcal{K}$ graph, in the case of the graph $\tilde{\mathcal{K}}$ it holds that for any long enough bispecial word $u \in \mathcal{L}(\mathbf{u})$ there is a bispecial word $v \in \mathcal{L}(\mathbf{u})$ such that $[\mathcal{E}^{(D_L,D_R)}(u)] = [\mathcal{E}^{(D_L,D_R)}(v)]$. For this, it is enough to define $\tilde{\delta} = \tilde{\delta}(\tilde{\mathcal{K}}(\mathcal{L}(\mathbf{u})))$ to be the length of the longest simple path in the graph $\tilde{\mathcal{K}}(\mathcal{L}(\mathbf{u}))$. It can be proved that $|v| \leq \Phi^{\tilde{\delta}}(2L + ||\varphi|| - 1)$.

Conjecture 4.4.20. Any strongly two-sided recognizable morphism on substitutive language is adequate.

Notice that if the above conjecture is true, then Theorem 4.4.13 and Corollary 4.4.15 hold for all strongly two-sided recognizable morphisms.

5

Profinite Semigroups

This chapter is based on a survey on profinite semigroups [69] and most of the results that are presented by Jorge Almeida and Alfredo Costa (see [7] and [8]). The motivation for the following introduction to profinite semigroups is to present some relation between the profinite semigroups and the recognizability, which is the principal notion of the thesis.

5.1 Profinite semigroups and symbolic dynamics

As an introduction to the profinite topology we consider the example of the Fibonacci morphism φ_F, defined in Example 2.2.2. The Fibonacci sequence can be extended to a converging sequence in the profinite topology. A subsequence of the Fibonacci sequence that contains words converging both in the suffix and the prefix topology, but not in the profinite topology, is the following subsequence,

$$\begin{aligned} \varphi_F^2(0) &= 010 \\ \varphi_F^4(0) &= 01001010 \\ \varphi_F^6(0) &= 010010100100101001010 \\ &\cdots \end{aligned}$$

defined by $w_{2n+2} = \varphi_F^{2n}(0)$, with $n \geq 1$. Indeed, in this sequence, each term is both a prefix and a suffix of the next one, but since

the length of the words is alternately odd and even, the sequence is not a Cauchy sequence with respect to the profinite topology that we will see how it is defined later.

We will see that the subsequence

$$\begin{aligned}\varphi_F^2(0) &= 010\\ \varphi_F^6(0) &= 010010100100101001010\\ \varphi_F^{24}(0) &= 01001010\cdots01001010\\ &\cdots\end{aligned}$$

defined by $w_{n!+2} = \varphi_F^{n!}(0)$, with $n \geq 1$, converges in the profinite topology to a pseudoword denoted $\varphi_F^\omega(0)$ which, in a certain sense, begins by the Fibonacci infinite word with prefixes $(\varphi_F^n(0))_{n\geq 0}$ and ends with the left infinite word with suffixes $(\varphi_F^{2n}(0))_{n\geq 0}$. The pseudoword $\varphi^\omega(0)$, as an element of the free profinite monoid, is the limit of the sequence $w_{n!+2} = \varphi_F^{n!}(0)$ and its length is a profinite natural number, the limit of the sequence $F_{n!+2}$ in $\hat{\mathbb{N}}$. If we consider the sequence of Fibonacci words as a sequence of group words, then their lengths should be seen as elements of $\hat{\mathbb{Z}}$. From this point of view, the length of $\varphi_F^\omega(0)$ as a *group pseudoword* (that is, considering the projection of $\varphi_F^\omega(0)$ in the free profinite group generated by A), is the limit of $F_{n!+2}$ in the ring of profinite integers, which is $F_2 = 1$, as we have seen in the previous section. Hence, $\varphi_F^\omega(0)$, as a group pseudoword, can be supposed to be a letter.

Before moving to the notions of profinite groups and free profinite groups, we will make a quick reminder of what a free group is.

Let A be a set. Then the *free group* $FG(A)$ or otherwise denoted as FG if A is clear from the context, is the set that consists of expressions that can be build of all elements of the set A, considering expressions different, unless they are equivalent because of the group axioms. We call *generators* of the free group $FG(A)$ the elements of A. A group G is called *free* if it is isomorphic to $FG(A)$ for some subset A of G (i.e. every element of G has a unique decomposition in elements of $A \cup A^{-1}$).

5.1.1 Projective limits

We define profinite semigroups as projective limit of finite semigroups. What is interesting about that limit is that it preserves the properties that hold true for all finite semigroups. Let us now define

the notion of projective limit, since it is necessary for the definition of profinite semigroups.

It is important to recall that the morphisms between topological semigroups are the continuous morphisms of semigroups.

A *projective system* (or *inverse system*) of topological semigroups is given by

(i) a directed set I, that is a partially ordered set (poset) in which any two elements have a common upper bound,

(ii) a topological semigroup S_i for each i in I,

(iii) a *connecting morphism* $\psi_{i,j} : S_i \to S_j$ for each pair $(i,j) \in I$ with $i \geq j$, such that $\psi_{i,i}$ is the identity on S_i and for $i \geq j \geq k$, $\psi_{i,k} = \psi_{j,k} \circ \psi_{i,j}$ (i.e. the diagram in Figure 5.1 commute).

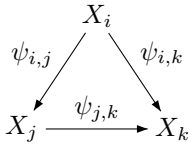


Figure 5.1: Projective system.

The pair $((S_i)_{i\in I}, (\psi_{i,j})_{i\geq j})$ is an inverse system, and the morphisms $\psi_{i,j}$ are the transition morphisms of the system.

Example 5.1.1. The set $\mathbb{Z}^+$ of positive integers, ordered by divisibility (that is, $n \geq m$ if and only if $m|n$), is a directed set. The family of cyclic groups $(\mathbb{Z}/n\mathbb{Z})_{n\in\mathbb{Z}^+}$ forms a projective system for the morphisms $\psi_{n,m}$ defined by $\psi_{n,m}(x) = x \bmod m$.

In the same way, the family of cyclic groups $(\mathbb{Z}/n!\mathbb{Z})_{n\in\mathbb{Z}^+}$, indexed by the set $\mathbb{Z}^+$ with the natural total order on integers, is a projective system.

More precisely, let $n = 4$ and $m = 12$, where $4|12$. We have $S_4 = \mathbb{Z}/4\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}\}$ and $S_{12} = \mathbb{Z}/12\mathbb{Z} = \{\bar{0}, \ldots, \bar{11}\}$. There is a connecting morphism $\psi_{12,4} : S_{12} \to S_4$, such that $\psi_{12,4}(x) = x \bmod 4$, for example $\psi_{12,4}(\bar{7}) = \bar{7} \bmod \bar{4} = \bar{3}$.

Example 5.1.2. Let $M_{i,p}$ for $i \geq 0$, $p \geq 1$ be the family of monoids $M_{i,p} = \{1, \ldots, a^{i+p-1}\}$ with $a^{i+p} = a^i$ (Figure 5.2). This family is a profinite system for the connecting morphisms $M_{i,p} \to M_{j,q}$ for $j \geq i$ and $q|p$, sending the generator a of $M_{i,p}$ on the generator a of $M_{j,q}$.

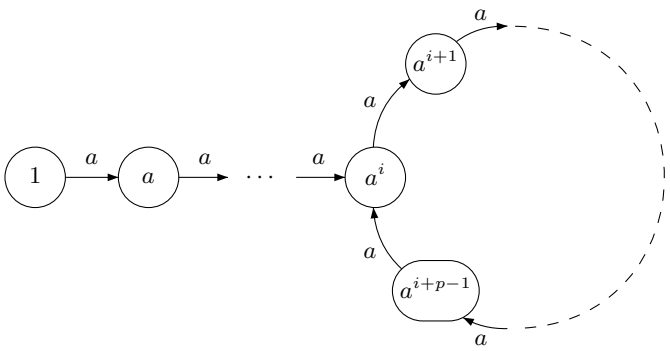


Figure 5.2: The monoid $\in M_{i,p}$.

The *projective limit* (or *inverse limit*) of the projective system, that has been defined above, is a topological semigroup S together with the morphisms $\Phi_i : S \to S_i$ such that the following conditions hold (see Figure 5.3),

1. for all $i, j \in I$ with $i \geq j$, one has $\psi_{i,j} \circ \Phi_i = \Phi_j$;

2. for any topological semigroup T and any family of morphisms $\Psi_i : T \to S_i$ such that, for all $i, j \in I$ with $i \geq j$, the equality $\psi_{i,j} \circ \Psi_i = \Psi_j$ holds, there exists a unique morphism $\theta : T \to S$ such that $\Phi_i \circ \theta = \Psi_i$ for all $i \in I$.

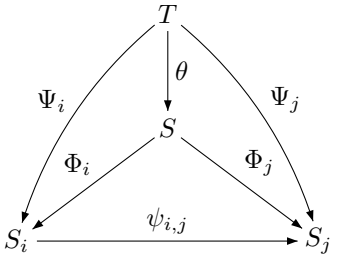


Figure 5.3: The projective limit.

The projective limit of a projective system always exists and it is unique. In order to prove the existence of the projective limit we have to take the subsemigroup S of the product $\prod_{i \in I} S_i$ consisting

of all $(s_i)_{i\in I}$ such that, for all $i, j \in I$ with $i \geq j$,

$$\psi_{i,j}(s_i) = s_j,$$

endowed with the product topology. The maps $\Phi_i : S \to S_i$ are the projections, that is, if $s = (s_i)_{i\in I}$, then $\Phi_i(s) = s_i$.
We will use the notation $\varprojlim_{i\in I} S_i$ for the projective limit and if the projective system is known we can use the notation $\varprojlim S_i$, for S.

Remark 5.1.3. A sequence $(s_n)_n$ converges in $\varprojlim_{i\in I} S_i$ if and only if, for every $i \in I$, the sequence $(\Phi_i(s_n))_n$ converges in S_i.

Let us mention that it is possible that the projective limit may be empty, as it happens in the following example,

Example 5.1.4. Let S_n be the set of positive integers $\mathbb{Z}^+$ under addition, with $\psi_{n,m}(x) = 2^{n-m}x$ being the connecting morphisms. In this case the projective limit is empty. That follows from the definition of the projective limit as a semigroup $S \subseteq \prod_{i\in I} S_i$, consisting of sequences $(s_i)_{i\in I}$ such that $\Psi_{i,j}(s_i) = s_j$, for all $i \geq j$. In this specific example we want $s_i = 2s_{i+1}$, that is true only if $s_i = 0$ for each i in I, but since $0 \notin S_n$ we conclude that $S = \{\emptyset\}$.

However, that is never the case under the conditions mentioned in the following proposition,

Proposition 5.1.5. [89, Lemma 3.1.26] Consider a projective limit $S = \varprojlim S_i$ of compact semigroups. Then S is a nonempty compact semigroup. If, moreover, the connecting morphisms $\psi_{i,j} : S_i \to S_j$ are onto, then each component projection $\Phi_i : S \to S_i$ is onto.

In particular the result holds when the semigroups S_i are finite.

From now on we will suppose that the semigroups in the projective system are compact and the connecting morphisms are onto.

The projective system and the projective limit of topological monoids or groups are defined in exactly the same way. For a projective system of topological monoids, one has to take all morphisms as continuous monoid morphisms, and similarly for groups[1] and for every abstract algebra.

Example 5.1.6. The projective limit of the family of cyclic groups in Example 5.1.1 is the group of profinite integers $\widehat{\mathbb{Z}} = \varprojlim_{n\geq 1} \mathbb{Z}/n\mathbb{Z}$.

[1] Actually, a monoid morphism between groups is already a group morphism.

In $\widehat{\mathbb{Z}}$ one has $\lim n! = 0$. As in any group there is a unique idempotent, which is 0.

Example 5.1.7. The projective limit of the family $M_{i,p}$ is the monoid $\widehat{\mathbb{N}}$ of profinite natural integers. In $\widehat{\mathbb{N}}$ there are two idempotents: 0 and $\omega = \lim n!$.

5.1.2 Free profinite monoids

In order to better explain the idea of profinite monoids we present the specific case of a finite alphabet A. Let u and v be two distinct words in A^*. We can consider a finite monoid M and some homomorphism $\varphi\colon A^* \to M$ such that the images of the two distinct words are distinct, $\varphi(u) \neq \varphi(v)$. We say that such a monoid *separates* the words u and v of A^*.

Example 5.1.8. Let us consider the morphism $\varphi : \{a, b\}^* \to \mathbb{Z}/2\mathbb{Z}$, such that $\varphi(x) = |x| \mod 2$. It can be seen that φ separates the words $aba, ab \in A^*$, since $|aba| \mod 2 = 1 \neq 0 = |ab| \mod 2$.

Let us now define the *natural metric* as the function $d\colon A^* \times A^* \to \mathbb{R}^+$ such that

$$d(u,v) = \begin{cases} 2^{-r(u,v)} & \text{if } u \neq v \\ 0 & \text{if } u = v \end{cases}$$

where $r(u, v)$ is the minimal cardinality of a monoid separating u and v. Because of the definition of natural metric, we have that two words are closer the larger the monoid that separates them is. The natural metric is an *ultrametric* since it satisfies the following conditions,

- $d(u, v) \leq \min(d(u, w), d(w, u))$,
- $d(u, v) = d(v, u)$,
- $d(ux, vy) \leq \max(d(u, v), d(x, y))$.

That metric space has a completion $\widehat{A^*}$, which is a compact metric space. The multiplication on A^* is uniformly continuous with respect to the metric d, and so the completion $\widehat{A^*}$ is a topological monoid whose multiplication is the unique continuous extension of the multiplication on A^*. More precisely, the metric space A^* is a

dense submonoid of $\widehat{A^*}$. An easy way to think of the elements of $\widehat{A^*}$ is as generalizations of words over the alphabet A. The elements of A^* are isolated in the topological space $\widehat{A^*}$. The elements of $\widehat{A^*}$ are called *pseudowords* over the alphabet A, the elements of $\widehat{A^*} \setminus A^*$ are called *infinite pseudowords* and the ones in A^* are called *finite pseudowords*.

An intuitively way to define the *profinite monoid* is as compact Hausdorff monoid M that is if u and $v \in M$ are distinct, then there is a continuous homomorphism $\varphi\colon M \to N$ with $\varphi(u) \neq \varphi(v)$, where N is a finite monoid endowed with the discrete topology.

Remark 5.1.9. The finite monoids are profinite, if endowed with the discrete topology.

Coming back to the special case of $\widehat{A^*}$, it can be characterized as *free profinite monoid generated by* A in the sense that if $\varphi\colon A \to N$ is a mapping into a profinite monoid N, then there is a unique extension of the morphism φ to a continuous homomorphism $\widehat{\varphi} : \widehat{A^*} \to N$.

A more formal way to define the free profinite monoid on a finite alphabet A is as the projective limit of this family, which does also satisfy the universal property, as seen in the following proposition.

Proposition 5.1.10. The natural mapping $\iota : A \to \widehat{A^*}$ is such that for any map $\varphi : A \to M$ into a profinite monoid there exists a unique continuous morphism $\widehat{\varphi} : \widehat{A^*} \to M$ such that $\widehat{\varphi} \circ \iota = \varphi$.

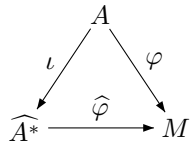


Figure 5.4: The universal property of $\widehat{A^*}$.

Example 5.1.11. The free profinite monoid on the one-letter alphabet $A = \{a\}$ is the set $\widehat{A^*}$ identified with $\widehat{\mathbb{N}}$ (A^* is identified with $\mathbb{N}$), with

$$\widehat{A^*} = \{a^x | x \in \widehat{\mathbb{N}}\}.$$

5.2 Free profinite groups

The notion of profinite monoids can be generalized to the notions of *profinite semigroups* and *profinite groups*. The formal way to define the *profinite group* is as a projective limit of a projective system of finite groups, while the *free profinite group* $\widehat{FG}(A)$ is as projective limit of the projective system formed by the isomorpism classes of A-generated finite groups.

Let us present the specific case of the profinite semigroup generated by the alphabet A, denoted as $\widehat{A^+}$. The way to construct this set is exactly the same as the one that is presented above for the construction of $\widehat{A^*}$. The profinite semigroup $\widehat{A^+}$ is basically the set $\widehat{A^*} \setminus \{\varepsilon\}$.

Finally, we use the notation $FG(A)$ and $\widehat{FG}(A)$ for the *free group* generated by A and the *free profinite group* generated by A respectively. The second one has $FG(A)$ as a dense subgroup. The *rank* of $\widehat{FG}(A)$ is the cardinality of the alphabet A.

The free group $FG(A)$ can be considered a profinite monoid, in which case, the canonical projection from $\widehat{A^*}$ onto $FG(A)$ is the unique continuous homomorphism $p_G \colon \widehat{A^*} \to FG(A)$ fixing the elements of A.

Example 5.2.1. The free group $FG(\{a\})$ identified with $\widehat{\mathbb{Z}}$.

Theorem 5.2.2. *[86] For each $x \in A^*$, the sequence $(x^{n!})_{n\geq 0}$ converges in $\widehat{A^*}$ to an idempotent denoted by x^ω.*

The way that x^ω is defined is as the limit

$$x^\omega = \lim_{n\to\infty} x^{n!} \tag{5.2.1}$$

of the sequence $(x^{n!})$ in $\widehat{A^*}$.

Proof. In order to prove that the sequence $(x^{n!})$ converges in $\widehat{A^*}$, we will prove that it is a Cauchy sequence. Let $\varepsilon > 0$, we would like to find integer $n > 0$ such that for all integers $q, p \geq n$, $d(x^{q!}, x^{p!}) < \varepsilon$. Because of the way that the natural metric is defined, it is enough to show that there is no monoid M with size $|M| \leq n$ that separates the elements $x^{q!}$ and $x^{p!}$. Let $\varphi : A^* \to M$ be such that $\mathrm{Card}(M) \leq n$ and let us consider $s = \varphi(x)$. Since the monoid M is finite, there

is an integer $r \leq n$ such that $s^r = 1$. Also, since $p, q \geq n$, one has that $n|p!$ and $n|q!$. Hence, $x^{q!} = x^{q!} = 1$ which implies that $d(s^{p!}, s^{q!}) = 0$ and thus there is no such monoid M that separates the two elements.

Now, it remains to see that x^ω is an idempotent. That follows from the fact that for a large enough integer n one has that $\varphi(x^{n!})\varphi(x^{n!}) = s^{n!}s^{n!} = 1 \cdot 1 = 1 = \varphi(x^{n!})$. □

Remark 5.2.3. The limit x^ω is a notation and it should not be considered an infinite word.

As it has been seen in Example 5.1.1, for $\widehat{\mathbb{Z}}$ we have $\lim n! = 0$.

5.3 Green's relations

Before giving to the definition of Green's relations let us present the notion of subshift. A *two-sided infinite word* (or *bi-infinite word*) on an alphabet A is an element $x = (x_n)_{n\in\mathbb{Z}}$ of $A^{\mathbb{Z}}$. Given a factorial biextendable subset F of A^*, the set of two-sided infinite words with all their factors in F is denoted $\mathcal{X}(F)$. It is closed for the product topology of $A^{\mathbb{Z}}$. The set $\mathcal{X}(F)$ is also invariant by the shift mapping $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$, defined by $y = \sigma(x)$ if $y_n = x_{n+1}$ for any $n \in \mathbb{Z}$. A nonempty closed and shift-invariant subset of $A^{\mathbb{Z}}$ is called a *subshift* (also called *shift space*) of $A^{\mathbb{Z}}$. The subshifts of $A^{\mathbb{Z}}$ are precisely the subsets of the form $\mathcal{X}(F)$, with F an infinite factorial biextendable subset of A^*. Moreover, whenever F and F' are infinite factorial biextendable sets, one has $\mathcal{X}(F) \subset \mathcal{X}(F')$ if and only if $F \subset F'$, and in particular each subshift is completely determined by a unique factorial biextendable set.

Let φ be a primitive substitution over a finite alphabet A. It is known that each such substitution can be associated with a minimal subshift $\mathcal{X}_\varphi$.

The *Green's relations* are five equivalence relations that characterize the elements of a semigroup in terms of the principal ideals that they generate. The three of the Green's relations are $\mathcal{J}$, $\mathcal{R}$ and $\mathcal{L}$, which are defined as follows,

- $u \,\mathcal{J}\, v \Leftrightarrow MuM = MvM$,
- $u \,\mathcal{R}\, v \Leftrightarrow uM = vM$,

- $u \mathcal{L} v \Leftrightarrow Mu = Mv$.

Therefore, we say that the relation $u\mathcal{R}v$ holds if and only if there exists elements $m, m' \in M$ such that $u = vm$ and $v = um'$ and similarly for all the other relations. Because of the above definitions we have that $\mathcal{R} \subset \mathcal{J}$ and $\mathcal{L} \subset \mathcal{J}$. The other two Green's relations are defined as follows,

- $\mathcal{H} = \mathcal{R} \cap \mathcal{L}$,
- $\mathcal{D} = \mathcal{R} \vee \mathcal{L}$.

The latest relation $\mathcal{D}$ is otherwise defined as the equivalence $\mathcal{RL} = \mathcal{LR}$. Two elements s, t are said to be $\mathcal{D}$-equivalent if and only if the relation $s\mathcal{L} \circ \mathcal{R}t$ holds (i.e. there exists element $m \in M$ such that $s\mathcal{R}m$ and $m\mathcal{L}t$ or there exists element $m' \in M$ such that $s\mathcal{L}m'$ and $m'\mathcal{R}t$).

It can be seen easily that the relations $\mathcal{R}$ and $\mathcal{L}$ imply the relation $\mathcal{D}$, while the relation $\mathcal{H}$ implies both of the relations $\mathcal{R}$ and $\mathcal{L}$. Let us use the notation H_a, R_a, L_a, D_a for the classes $\mathcal{H}, \mathcal{R}, \mathcal{L}, \mathcal{D}$ respectively that contain the element $a \in M$.

Remark 5.3.1. Over finite monoids one has $\mathcal{D} = \mathcal{J}$.

A $\mathcal{J}$-class decomposes into $\mathcal{R}$-classes and $\mathcal{L}$-classes, and those classes decompose into $\mathcal{H}$-classes. Moreover, as mentioned above, in the case of a finite monoid holds that $\mathcal{J} = \mathcal{D} = \mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$, which implies that the intersection of every such $\mathcal{L}$-class and a $\mathcal{R}$-class is non-empty. The following proposition can be used in order to prove the above.

Proposition 5.3.2. For any finite monoid, if $s\mathcal{J}t$ then $\mathcal{L}(s) \cap \mathcal{R}(t) \neq \emptyset$.

Proof. $s\mathcal{J}t \Rightarrow s\mathcal{D}t \Rightarrow \exists\ x \in M$ s.t. $s\mathcal{L}x\mathcal{R}t \Rightarrow \mathcal{L}(s) \cap \mathcal{R}(t) \neq \emptyset$ □

A $\mathcal{D}$-class can be presented by a graph called *"egg-box"*, that is illustrated in Figure 5.5. In this graph, the rows correspond to the $\mathcal{R}$-classes, while the columns correspond to the $\mathcal{L}$-classes. Every sub-box of the "egg-box" is an intersection of an $\mathcal{R}$-class and an $\mathcal{L}$-class, thus it is a non empty $\mathcal{H}$-class. According to Green's theorem any two $\mathcal{D}$-equivalent $\mathcal{H}$-classes have the same cardinality. The $\mathcal{D}$ relation between two elements $m, n \in M$ can be represented using

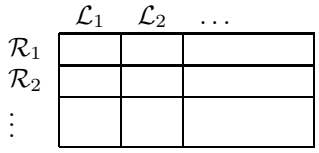


Figure 5.5: $\mathcal{D}$-class representation as an "egg-box".

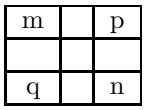


Figure 5.6: $\mathcal{D}$-relation between two elements $m, n \in M$.

the "egg-box" as seem in Figure 5.6, which illustrates the existence of elements $q, p \in M$ such that $m\mathcal{L}q\mathcal{R}n$ and $m\mathcal{R}p\mathcal{L}n$. This implies that there is a permutability of $\mathcal{R}$ and $\mathcal{L}$, since if $\mathcal{R}(a) \cap \mathcal{L}(b) \neq \emptyset$, then $\mathcal{R}(b) \cap \mathcal{L}(a) \neq \emptyset$. Also, the family of $\mathcal{J}$-classes (resp. $\mathcal{R}$, $\mathcal{L}$-classes) is ordered by $u \leq_{\mathcal{J}} v$ if and only if $MuM \subset MvM$.

Example 5.3.3. Let $M = \{\varepsilon, a, b\}$ be the monoid defined by the automaton presented in Figure 5.7. It is easy to see that the following equations hold,

$$ab = b,\ bb = b,\ ba = a,\ aa = a.$$

Those lead to the result that $a\mathcal{R}b$, while a and b are not $\mathcal{L}$-equivalent as it can be seen from the "egg-box" in Figure 5.7.

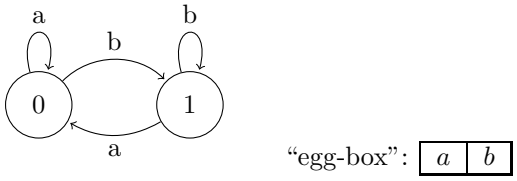


Figure 5.7: Automaton that generates the monoid $M = \{\varepsilon, a, b\}$.

Example 5.3.4. Let $M = \{1, a, b, ab, ba, 0\}$ be the monoid defined by the automaton presented in Figure 5.8 for which the "egg-box" presented in the same figure illusterates the relations between the classes.

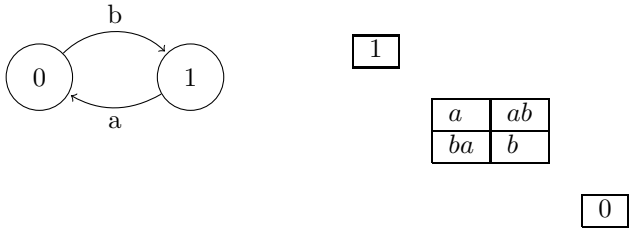


Figure 5.8: Automaton that generates the monoid $M = \{\varepsilon, a, b\}$.

Definition 5.3.5. A $\mathcal{D}$-class (or an $\mathcal{H}$-class or $\mathcal{R}$-class or $\mathcal{L}$-class) is *regular* if it contains an idempotent.

Lemma 5.3.6. (Location Lemma) Let M be a finite monoid and $s, t \in M$ such that $s\mathcal{D}t$. Then $st\mathcal{D}t$ (resp. $st\mathcal{R}s$, $st\mathcal{L}t$) if and only if the $\mathcal{H}$-class $\mathcal{L}(s) \cap \mathcal{R}(t)$ contains an idempotent.

An element u of a monoid M is characterized as *regular* if $u \in uMu$ (i.e. there is $x \in M$ such that $uxu = u$).

Lemma 5.3.7. (Von Neumann) The following three propositions concerning an element a of a monoid M are equivalent,

1. a is regular,
2. L_a contains an idempotent,
3. R_a contains an idempotent.

In a compact monoid, a $\mathcal{J}$-class contains a regular element if and only if all its elements are regular, if and only if it contains an idempotent, and for that reason, if $\mathcal{K}$ is one of the relations $\mathcal{J}, \mathcal{R}, \mathcal{L}$, $\mathcal{H}$, one says that a $\mathcal{K}$-class is regular when some of (equivalently, all of) its elements are regular.

Lemma 5.3.8. Let M be a finite monoid. A $\mathcal{D}$-class is regular if and only if every element in the class is regular if and only if contains a regular element.

Proof. Let us suppose that there is a regular element s in a $\mathcal{D}$-class D. Then there is element $t \in D$ such that $s = sts$, which implies that $st = stst$ and hence st is an idempotent. Furthermore, the equality $s = sts$ implies that $st\mathcal{D}s$ and hence D is regular.

In order to prove the next part of the lemma we will suppose that D is regular, which implies that every $\mathcal{R}$-class contains an idempotent, let us call the idempotent e. For all elements $s \in D$ there are elements $e, t \in D$ such that $st = e$ from which follows that $sts = s$, and hence every element is regular. □

The regular $\mathcal{D}$-classes are the ones that contain idempotents. In the monoid M, the $\mathcal{H}$-classes of idempotents are the maximal subgroups of M, with respect to the inclusion relation (a *subgroup* of M is a subsemigroup of M which is a group). The maximal subgroups contained in the same regular $\mathcal{D}$-class are isomorphic. If M is profinite, then $\mathcal{J} = \mathcal{D}$, and the maximal subgroups of the same regular $\mathcal{D}$-class are isomorphic as profinite groups.

A profinite monoid M also satisfies the *stability property*, $u \mathcal{J} ux$ if and only if $u \mathcal{R} ux$, and dually for the relation $\mathcal{L}$, we have that $u \mathcal{J} xu$ if and only if $u \mathcal{L} xu$, for every $u, x \in M$.

Let H be an $\mathcal{H}$-class of M. Set $T(H) = \{x \in M \mid Hx = H\}$. Each $x \in T(H)$ defines a map $\rho_x : H \to H$ by $\rho_x(h) = hx$. The set of the translations ρ_x for $x \in T(H)$ is a topological group acting by permutations on H, denoted $\Gamma(H)$. The groups corresponding to different $\mathcal{H}$-classes contained in the same $\mathcal{J}$-class J are continuously isomorphic and their equivalence class, called the *Schützenberger group* of J, is denoted $G(J)$. Every $\mathcal{H}$-class has the same cardinality as its Schützenberger group $G(J)$. If the $\mathcal{D}$-class is regular, then there is at least one group in every column and every raw and as said above, all of them are isomorphic one to each other.

If J is a regular $\mathcal{J}$-class, then every $\mathcal{H}$-class of J which is a group is isomorphic to $G(J)$. Indeed, $H \subset T(H)$ and the restriction to H of the mapping $\rho : x \in T(H) \to \rho_x \in \Gamma(H)$ is an isomorphism (see [70] for a more detailed presentation).

A subshift $\mathcal{X}$ is *minimal* if it does not contain subshifts different that $\mathcal{X}$. It can be proved that if a subshift $\mathcal{X}$ is minimal, then $\overline{\mathcal{L}(\mathcal{X})} \backslash A^+$ is contained in a regular $\mathcal{J}$-class, denoted as $J(\mathcal{X})$. Thus, $J(\mathcal{X})$ is by definition the set of infinite pseudowords such that all their finite factors are in $\mathcal{L}(\mathcal{X})$.

Proposition 5.3.9. $J(\mathcal{X})$ is a J-class of $\widehat{A^*}$.

More precisely, $J(\mathcal{X})$ is a maximal not trivial $\mathcal{J}$-class.

Before moving to the next lemma let us use the notation $\overrightarrow{v}$ for the right infinite word whose finite prefixes are those of v, while

the notation $\overleftarrow{v}$ is the notation used for the left infinite word whose finite suffixes are those of v.

Lemma 5.3.10. [5] For a minimal subshift $\mathcal{X}$, two elements $u, v \in J(\mathcal{X})$ are $\mathcal{R}$-equivalent (resp. $\mathcal{L}$-equivalent) if and only if $\overrightarrow{u} = \overrightarrow{v}$ (resp. $\overleftarrow{u} = \overleftarrow{v}$).

It is known from [14] that $w \in J(\mathcal{X})$ lies in a subgroup if and only if the two-sided infinite word $\overleftarrow{w}.\overrightarrow{w} \in \mathcal{X}$. It follows that the maximal subgroups $H \subset J(\mathcal{X})$ are in bijection with the elements of $\mathcal{X}$, with any element $w \in H$ being sent to the two-sided infinite words $\overleftarrow{w}.\overrightarrow{w}$. It is denoted by H_x the maximal subgroup corresponding to x.

5.4 Limit return sets

In order to define limit return sets, we have to present the notion of *return words*.

Let F be a factorial set. A *right return word*, or otherwise called just *return word*, to $x \in F$ is a nonempty word $w \in F$ such that xw begins and ends with x but has no internal factor equal to x. We denote by $\mathcal{R}_F(x)$ (or simply $R(x)$ if F is clear) the set of return words to x. Note that $\mathcal{R}_F(x)$ is a prefix code, since if it was not, that would imply that there is a word in $\mathcal{R}_F(x)$ that has x as internal factor and that contradicts the definition.

For $x \in F$, we denote

$$\Gamma_F(x) = \{w \in F \mid xw \in F \cap A^*x\}.$$

Thus $\mathcal{R}_F(x)$ is the set of nonempty words in $\Gamma_F(x)$ without any proper prefix in $\Gamma_F(x)$. Similarly, a *left return word* to $x \in F$ is a nonempty word $w \in F$ such that wx begins and ends with x but has no internal factor equal to x. The set of the left return words to x is denoted by $\mathcal{R}'_F(x)$ and the relation $\mathcal{R}'_F(x) = x\mathcal{R}_F(x)x^{-1}$ holds always.

There is a similar definition presented in [39] of the sets $\mathcal{R}_{u.v}$, called the *n-delayed return words of* uv, where $|u| = n$. This is the set of all the factors $x_{[i,j)}$ such that i, j are two consecutive occurrences of the words uv in x. It can be easily seen that,

$$\mathcal{R}\,(uv) = u^{-1}\,\mathcal{R}_{u.v}\,v \quad \text{and} \quad \mathcal{R}\,(u) = \mathcal{R}_{\varepsilon.u}\,.$$

Example 5.4.1. Let F be the Fibonacci set. The sets of right and left return words to a are $\mathcal{R}_F(a) = \{a, ba\}$ and $\mathcal{R}'_F(a) = \{a, ab\}$.

Let us recall that a word is called *primitive* if it has only a trivial representation as a power of other words. Non-primitive words are the periodic words.

Example 5.4.2. Let F be a periodic set. Let w be a primitive word of length n such that $F = F(w^*)$. Then, for any word $x \in F$ of length at least n, the set $\mathcal{R}_F(x)$ is reduced to one word of length n. It is true that there is a word of length n in $\mathcal{R}_F(x)$, and that is because n is the period of the fixed point. Also, if there was a longer one then it would have internal factor x because of the period of the fixed point. Lastly, there cannot be a word shorter than n, since if there was such a word, that would mean that there is a word z such that $w = z^k$ that is not possible since w is chosen to be primitive. More precisely, let $x = wu$ and $v \in R_F(x)$, where $|v| < n$. If $|uv| < n$, then we can write w as $w = zuv$ for some $z \in F$. However if such the case, it can be easily seen that z is also a suffix of w, hence $w = zz'z$, for some $z' \in F$. Following the same procedure we can conclude that $w^* = z^*$ that leads to a contradiction, since w is chosen to be primitive.

The following result appears in [22].

Proposition 5.4.3. [22] Let F be a factorial set. For any $x \in F$, one has $\Gamma_F(x) = \mathcal{R}_F(x)^* \cap x^{-1}F$.

Proof. If a nonempty word w is in $\Gamma_F(x)$ and not in $\mathcal{R}_F(x)$, then w has an internal factor x. Thus, setting v as the factor after the first appearance of x in w until the end of it, we have $w = uv$ with $u \in \Gamma_F(x)$ and v nonempty. Since $\Gamma_F(x)$ is right unitary, we have $v \in \Gamma_F(x)$, from which we conclude $w \in \mathcal{R}_F(x)^*$ by induction on the length of w. Moreover, $xw \in F$ and thus $w \in x^{-1}F$.

Conversely, assume that w is a nonempty word in $\mathcal{R}_F(x)^* \cap x^{-1}F$. Set $w = uv$ with $u \in \mathcal{R}_F(x)$ and $v \in \mathcal{R}_F(x)^*$. Then $xw = xuv \in A^*xv \subset A^*x$, since xv ends with x ($v \in R_F(x)$) and $xw \in F$. Thus, $w \in \Gamma_F(x)$. □

Note that for any words $x, y \in F$ such that $xy \in F$, it holds the following,

$$\mathcal{R}_F(xy) \subset \mathcal{R}_F(y)^*. \tag{5.4.1}$$

Indeed, let $w \in R_F(xy)$, which implies that $xyw \in F$ has the word xy as suffix and prefix but has no other appearance of it. It follows that $yw \in x^{-1}F$ has y as suffix and prefix and it possible has other appearances of the factor y, which implies that $w \in \Gamma_F(y)$. From Proposition 5.4.3 we have that $w \in R_F(y)^* \cap y^{-1}F \subset R_F(y^*)$. Thus, $R_F(xy) \subset R_F(y)^*$.

A similar relation stands for the left return words,

$$\mathcal{R}'_F(xy) \subset \mathcal{R}'_F(y)^*. \tag{5.4.2}$$

Example 5.4.4. Let $\varphi : a \mapsto ab, b \mapsto a^3b$ and F be the set of finite factors of the fixed point $\varphi^\omega(a)$. One has $\mathcal{R}_F(a) = \{a, ba\}$ and $\mathcal{R}_F(aa) = \{a, babaa, bababaa\}$, thus $\mathcal{R}_F(aa) \subset \mathcal{R}_F(a)^*$ as it was expected by Equation (5.4.1).

The first theorems that indicate some relation between the characterization of a set as dendric or connected and the free groups are the following ones presented in [19, Theorem 4.5] and [19, Theorem 4.7].

Theorem 5.4.5. *[19] Let F be a uniformly recurrent dendric set. For any $x \in F$, the set of return words $\mathcal{R}_F(x)$ is a basis of $FG(A)$.*

The above theorem shows that in a dendric set a property much stronger than Equation (5.4.1) holds, and in order to be proved it uses Equation (5.4.1) and the following theorem.

Theorem 5.4.6. *[19] Let F be a uniformly recurrent connected set. For any $w \in F$, the set $\mathcal{R}_F(w)$ generates the free group $FG(A)$.*

Example 5.4.7. Let F be the Tribonacci set on $A = \{a, b, c\}$, defined by the morphism $\tau : a \mapsto ab$, $b \mapsto ac$, $c \mapsto a$ with fixed point $\tau^\omega(a)$. Then the set of return words is $\mathcal{R}_F(a) = \{a, ba, ca\}$, which is a basis of $FG(A)$.

5.5 Groups of fixed points of morphisms

Let $\varphi : A^* \to A^*$ be a primitive substitution and let $\mathcal{L}(\varphi)$ be the uniformly recurrent set of factors of a fixed point of φ, noted as $\mathbf{u}$. Let $J(\varphi)$ be the $\mathcal{J}$-class $J(F(\varphi))$, and $G(\varphi)$ the Schützenberger

group of $J(\varphi)$. In that case, it holds that $\varphi(F(\varphi)) \subset F(\varphi)$, which, in view of $J(\varphi) = \overline{F(\varphi)} \setminus A^*$, justifies the following observation.

Let φ be a substitution over A. Then φ extends to a unique continuous endomorphism of $\widehat{A^*}$, which we still denote by φ. We can prove that the topological monoid $\mathrm{End}(\widehat{A^*})$ is profinite. Thus, the morphism φ^ω is well defined as the unique idempotent in the closure of the subsemigroup of $\mathrm{End}(\widehat{A^*})$ generated by φ. By definition of the pointwise topology, one has $\varphi^\omega(u) = \lim \varphi^{n!}(u)$, for every $u \in \widehat{A^*}$.

Remark 5.5.1. The inclusions $\varphi^\omega(F(\varphi)) \subseteq J(\varphi)$ and $\varphi(J(\varphi)) \subseteq J(\varphi)$ hold.

Recall that a *connection* for a substitution φ is a word $ba \in \mathcal{L}(\varphi)$ with $b, a \in A$, such that the first letter of $\varphi^\omega(a)$ is a, and the last letter of $\varphi^\omega(b)$ is b. It has been proved by Almeida [15, Lemma 4.1], that for every primitive substitution there is a connection. In terms of subshifts $\mathcal{X}_\varphi$ of a morphism φ with fixed point x, a connection can be seen as a word $x_{-1}x_0 \in \mathcal{L}(x)$. A *connective power* of φ is a finite power of φ, noted as $\tilde{\varphi}$, that is such that the first letter of $\tilde{\varphi}(a)$ is a and the last letter of $\tilde{\varphi}(b)$ is b. The intersection of the $\mathcal{R}$-class containing $\varphi^\omega(a)$ with the $\mathcal{L}$-class containing $\varphi^\omega(b)$ is a maximal subgroup of $J(\varphi)$ and is noted as H_{ba}.

The following proposition proves the existence of a connection for all primitive substitutions.

Proposition 5.5.2. [12] Every primitive substitution has a connection.

Proof. Let $a, b \in A$ be such that $ba \in \mathcal{L}(\varphi)$. Since the alphabet A is chosen to be finite, there are integers $i, j \geq 0$ and $p, q \geq 1$ such that $\varphi^i(a)$ and $\varphi^{i+p}(a)$ start with the same letter and the words $\varphi^j(b)$ and $\varphi^j(b)$ and $\varphi^{j+q}(b)$ end with the same letter. Let us set the integer $k = \max(i, j)$ and let us also set c be the first letter of $\varphi^k(a)$ and d be the last letter of $\varphi^k(b)$. Since $ba \in \mathcal{L}(\varphi)$ so does its k-image $\varphi^k(ba)$, and since dc is a factor of $\varphi^k(ba)$ we have that $dc \in \mathcal{L}(\varphi)$. Also, it can be seen that φ since $\varphi^p(c) \in A^*$, then $\varphi^\omega(c) \in c\widehat{A^*}$, and similarly $\varphi^\omega(d) \in \widehat{A^*}d$. Thus, it is implied that dc is a connection of the primitive substitution. □

Example 5.5.3. Let $\tau : a \mapsto ab, b \mapsto ba$ be the Thue-Morse morphism. The word aa is a connection for τ and $\tilde{\tau} = \tau^2$ is a connective power of τ, since $\tau^\omega(a) \in aA^*a$.

The following lemma presented in [7] indicates that natural actions are induced on certain subsets of the J-class of a primitive substitution.

Lemma 5.5.4. [7] Let φ be a primitive substitution and let ba be a connection for φ. If $\tilde{\varphi}$ is a connective power of φ, then $\tilde{\varphi}(H_{ba}) \subseteq H_{ba}$.

Definition 5.5.5. A sequence x on a finite alphabet is *linearly recurrent* for a constant $K \in \mathbb{N}$ if it is recurrent and if, for every factor $u \in \mathcal{L}(x)$, the difference between two successive occurrences of u in x is less than $K|u|$.

Remark 5.5.6. While in the case of linearly recurrent sequence the constant K does not depend on the word u, in the uniformly recurrent case the constant may depends on the word u.

The following theorem bounds the cardinality of the set of return words to a word u and it has been proved by F.Durand, B. Host and C.Skau in [39].

Theorem 5.5.7. *Let* $\mathbf{u}$ *be a linearly recurrent aperiodic sequence with constant* K*. Then, for all* $u \in \mathcal{L}(\mathbf{u})$*,* $card(R_F(u)) \leq K(K+1)^2$*.*

Also, in the work of F.Durand [37] it has been proved the following proposition.

Proposition 5.5.8. [37] All substitutive subshifts are linearly recurrent.

Thus, the following corollary easily follows,

Corollary 5.5.9. Let φ be a non-periodic primitive substitution with biinfinite fixed point $\mathbf{u}$. Then, the sequence $|R(n, \mathbf{u}_{[-n,n]})|$ is bounded.

Remark 5.5.10. The set H_{ba} is a maximal subgroup of $J(\varphi)$.

The fact that the $\mathcal{H}$-class H_{ba} is a group follows immediately from basic properties of Green's relations and from $\varphi^\omega(a)$, $\varphi^\omega(b)$ and $\varphi^\omega(ba)$ being all elements of $J(\varphi)$, where the latter is true because $ba \in \mathcal{L}(\varphi)$ (cf. Remark 5.5.1).

Proposition 5.5.11. [7] Let $\mathcal{X}$ be a minimal aperiodic subshift of $A^{\mathbb{Z}}$ and $x \in \mathcal{X}$. Suppose there are strictly increasing sequences of positive integers $(p_n)_n$ and $(q_n)_n$ such that $R(p_n, x_{[-p_n,p_n]}) =$

$\{r_{n_1}, r_{n_2} \dots r_{n_M}\}$ for all $n > 0$. Let $(r_1, r_2 \dots r_M)$ be an arbitrary accumulation point of the sequence $(r_{n_1}, r_{n_2} \dots r_{n_M})_n$ in the $\widehat{A^+}^M$. Then $\overline{< r_1, r_2, \dots, r_M >}$ is the maximal subgroup H_x of $J(\mathcal{X})$.

The following result by J.Almeida and A.Costa [7] uses the notion of two-sided recognizability in order to prove that there is related to the relation between the set H_{ba} of a connection ba of a morphism φ and image of any connective power $\tilde{\varphi}$ of φ.

Proposition 5.5.12. [7] Let φ be an aperiodic primitive substitution and let ba be a connection of φ. If $\tilde{\varphi}$ is a connective power of φ, then $H_{ba} \subseteq Im(\tilde{\varphi})$.

Proof. Because of the definition of H_x as a maximal subgroup corresponding to x we have that there is a unique element x of the subshift $\mathcal{X}_\varphi$ such that $H_x = H_{ba}$. Because of Lemma 5.5.4, the word x can also be written as

$$\dots \tilde{\varphi}(x_{-2})\tilde{\varphi}(x_{-1}) \cdot \tilde{\varphi}(x_0)\tilde{\varphi}(x_1)\tilde{\varphi}(x_2) \dots$$

which implies that x is a fixed point of the connective power $\tilde{\varphi}$. Hence, the connective power $\tilde{\varphi}$, because of Theorem 3.4.1, is two-sided recognizable with a constant of recognizability $l > 0$.

Now, by Corollary 5.5.9 the sequence $|R(n, \mathbf{u}_{[-n,n]})|$ is bounded. Thus, we can find a strictly increasing sequence of integers (p_n) such that $|R(n, \mathbf{u}_{[-n,n]})| = M$ for $M \in \mathbb{N}$ and $p_n > l$. Since the cardinality of the set of return words is M, we can denote $R(n, \mathbf{u}_{[-p_n,p_n]}) = \{r_{n_1}, r_{n_2} \dots r_{n_M}\}$. As it has been proved above, the connective power $\tilde{\varphi}$ is two-sided recognizable and since p_n is longer than $2l + 1$ and the position $0 \in E_{\mathbf{u}}$, we have that any other appearance of $x_{[-p_n,p_n]}$ in the fixed point x, such as $x_{[i-p_n,i+p_n]}$, $i \in E_{\mathbf{u}}$. Also, since x is is linearly recurrent, there are infinite appearances of $x_{[-p_n,p_n]}$ in bounded distances. Hence, there are integers $i, j > 0$ with $j > i$ such that $x_{[i-p_n,i+p_n]} = x_{[j-p_n,j+p_n]} = x_{[-p_n,p_n]}$ and $i, j \in E_{\mathbf{u}}$. That means that the factor $x_{[i,j-1]} \in Im(\tilde{\varphi})$.

Let us denote by $(r_1, r_2, \dots, r_M)$ the accumulation point of the sequence $(r_{n_1}, r_{n_2} \dots r_{n_M})_n$, which belongs to $Im(\tilde{\varphi})^M$. Because of Proposition 5.5.11 we can conclude that $H_{ba} \subseteq Im(\tilde{\varphi})$. □

The above Theorem 5.5.12 that uses the notion of two-sided recognizability is used for the proof of the following theorem that was also presented in [7].

Theorem 5.5.13. *Let φ be a non-periodic primitive substitution. Consider a connection ba for φ and a connective power $\tilde{\varphi}$. Then $H_{ba} = \tilde{\varphi}(H_{ba}) = \varphi^{\omega}(H_{ba})$.*

Proof. Since $\tilde{\varphi}$ is a connective power of φ, so it is any of its powers $\tilde{\varphi}^k$. By Proposition 5.5.12 we have that $H_{ba} \subseteq Im(\tilde{\varphi}^k)$. Hence, for any element $g \in H_{ab}$ there is $u_k \in \widehat{A^+}$ such that $g = (\widehat{\varphi})^{k!}(u_k)$. Since the mapping on continuous endomorphisms of finitely generated profinite semigroups is continuous, there is word u such that $g = \tilde{\varphi}^{\omega}(u) = \varphi^{\omega}(u)$. Thus, we have that $g = \varphi^{\omega}(g)$ which proves the equality $H_{ba} = \varphi^{\omega}(H_{ba})$.

The relation $\tilde{\varphi}(H_{ba}) \subseteq H_{ba}$ holds because of Lemma 5.5.4. As it has been mentioned above, any k power of a connective power $\tilde{\varphi}$ is also a connective power for ba. Hence, $\tilde{\varphi}^{k!-1}(H_{ba}) \subseteq H_{ba}$. Also, it follows from the definition of R,L-classes that all such classes are closed, and thus so it is the H-class (as an intersection of closed classes). Since H_{ba} is closed we can take the limit for k and obtain $\tilde{\varphi}^{\omega-1}(H_{ba}) \subseteq H_{ba}$. We have,

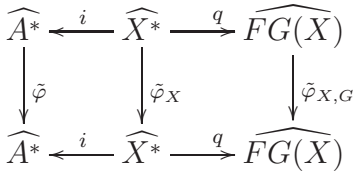
$$H_{ba} = \varphi^{\omega}(H_{ba}) = \tilde{\varphi}^{\omega}(H_{ba}) = \tilde{\varphi}(\varphi^{\omega-1}(H_{ba})) \subseteq \tilde{\varphi}(H_{ba})$$

which proves the equality $\tilde{\varphi}(H_{ba}) = H_{ba}$. □

In order to prove the main theorem of this chapter we should introduce some more notions first. Let φ be a morphism, $\mathcal{X}_{\varphi}$ be its subshift and $ba \in \mathcal{L}(\mathcal{X}_{\varphi})$ be a connection and $\tilde{\varphi}$ be a connective power. We will use the notation i for the unique homomorphism from the semigroup freely generated by $X = b^{-1}R(ba)b$ into the semigroup freely generated by the alphabet A that is such that $i(x) = x$ for all $x \in X$. As it has been mentioned earlier X is a code, which implies that the morphism i is injective. The homomorphism i can be extended uniquely to a continuous homomorphism $i : \widehat{X^*} \to \widehat{A^*}$ called the *encoding associated with the connection ba*.It has been proven in [75] that the last mapping i is also injective.

Let us now define the unique continuous homomorphism $q : \widehat{X^*} \to \widehat{FG(X)}$ that is the identity on the generators. There are also the continuous endomorphisms $\tilde{\varphi}_X$ and $\tilde{\varphi}_{X,G}$ such that the fol-

lowing diagram commutes:



More precisely the endomorphism $\tilde{\varphi}_X$ and $\tilde{\varphi}_{X,G}$ are such that $\tilde{\varphi}_X(x) = w_X$ and $\tilde{\varphi}_{X,G}(x) = w_X$ for $x \in X$, where w_X is a semigroup words and a group word respectively.

Let us also note that if a semigroup admits the following presentation,

$$\langle X \mid R\rangle_S$$

for a relation R, then it also admits the presentation,

$$\langle X \mid (q \times q)(R)\rangle_G.$$

It has been proven in [7] that a profinite subgroup associated with a primitive substitution φ and a connection ba admits a semigroup presentation, as it follows from the next proposition.

Proposition 5.5.14. [7] Let φ be a primitive substitution over the alphabet A, ba be a connection for φ, and $\tilde{\varphi}$ be a connective power of φ. Put $X = X_\varphi(a, b)$ and $H = Im(\varphi^\omega \circ i)$, where i is the encoding associated with ba. Then $Ker(\varphi^\omega \circ i) \subseteq \tilde{\varphi_X^\omega}$ and so H admits the presentation

$$\langle X | \tilde{\varphi_X^\omega}(x) = x \ (x \in X)\rangle_S. \tag{5.5.1}$$

In the proof of Proposition 5.5.14 presented by J.Almeida and A.Costa [7] one can obtain the following commutative diagram,

Lemma 5.5.15. [15] If u is a non-empty factor of the $\mathcal{J}$-class $J(\varphi)$, then $\varphi^\omega(u)$ belongs in $J(\varphi)$.

Proof. Let $\varphi^\omega(v) \in J(\varphi)$, such that u is one of its factors. Then there are words $x, y \in \widehat{A^*}$ such that $\varphi^\omega(v) = xuy$. Applying the homomorphism φ^ω we have $\varphi^\omega(\varphi^\omega(v)) = \varphi^\omega(xuy)$. Since φ^ω is an idempotent homomorphism it holds that $\varphi^\omega(v) = \varphi^\omega(x)\varphi^\omega(u)\varphi^\omega(y)$, which implies that the infinite word $\varphi^\omega(u)$ is a factor of $\varphi^\omega(v)$. It is known that $J(\varphi)$ is a J-maximal $\mathcal{J}$-class, which implies that $\varphi^\omega(u) \in J(\varphi)$. □

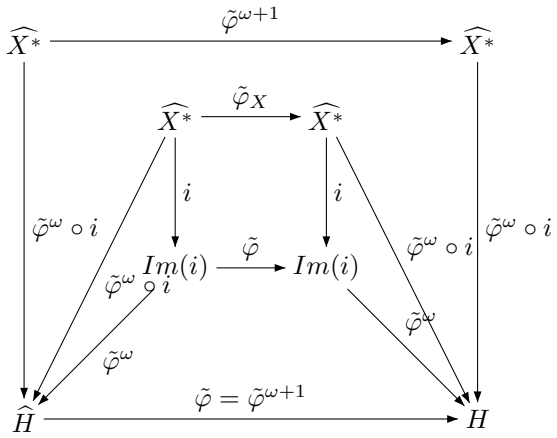


Figure 5.9: The universal property of $\widehat{A^*}$.

Lemma 5.5.16. [7] Let $\mathcal{X}$ be a minimal subshift of $A^{\mathbb{Z}}$ and $v \in \widehat{A^*}$ such that it belongs to maximal subgroup of $J(\mathcal{X})$. If the words $u_1, u_1 \in A^*$ are such that they are suffix and prefix of the word v respectively, then $v \in \langle \overline{R(u_1, u_2)} \rangle$.

Proof. Let $v \in J(\mathcal{X})$ that satisfies the above conditions, which implies that $v^3 \in J(\mathcal{X})$. Hence, $u_1 v u_2 \in J(\mathcal{X})$. There must exist sequence of words (w_n) ,such that all elements w_n belong in $\mathcal{L}(\mathcal{X})$, that converges to $u_1 v u_2 \in u_1 u_2 \widehat{A^*} \cap \widehat{A^*} u_1 u_2$. It can be assumed that all w_n have the factor $u_1 u_2$ as suffix and prefix. It is implied that $w_n (u_1 u_2)^{-1} \in \Gamma(u_1 u_2)$ (i.e. it is a product of words in $R(u_1, u_2)$), and we conclude that $u_1^{-1} w_n u_2^{-1} \in \langle \overline{R(u_1, u_2)} \rangle$, and hence so does the limit. □

Proposition 5.5.17. Let φ be a primitive substitution over an alphabet A, ba be a connection of the morphism φ, and $\tilde{\varphi}$ be a connective power of φ. Let $X = X_\varphi(a, b)$ and $H = (\varphi^\omega \circ i)$, where i is the encoding associated with ba. Then, $H = H_{ba}$.

Proof. Firstly we will prove that H is contained in H_{ba}. We have that $a, b \in A$ are such that ba is a factor of $J(\varphi)$. The words $\varphi^\omega(a)$, $\varphi^\omega(b)$ and all other elements of $\varphi^\omega(X_\varphi(a, b))$ belong to $J(\varphi)$, by Lemma 5.5.15. Also, by definition we have that if $v \in \varphi^\omega(X_\varphi(a, b))$ that means that $v \in \varphi^\omega(a)\widehat{A^*} \cap \widehat{A^*}\varphi^\omega(b)$. Hence, $\varphi^\omega(X_\varphi(a, b)) \in \mathcal{R}(\varphi^\omega(a)) \cap \mathcal{L}(\varphi^\omega(b)) = H_{ba}$, from which we conclude that $H =$

$\varphi^\omega(i(X)) \subseteq H_{ba}$. Let us now prove the other direction. Let $v \in H_{ba}$, which implies that it has suffix b and prefix a. From Lemma 5.5.16 we have that $v \in \overline{X^*}$, and because of definition of i, $v \in Im(i)$ that implies $H_{ba} \subseteq Im(i)$ and applying the morphism φ^ω we have $\varphi(H_{ba}) \subseteq Im(\varphi^\omega \circ i)$. Since $H_{ba} = \varphi^\omega(H_{ba})$, by Theorem 5.5.13, it holds that $H_{ba} \subseteq H$. □

The following is the main result in [7].

Theorem 5.5.18. *Let φ be a aperiodic primitive substitution over the alphabet A. Let ba be a connection of the morphism φ and let $X_\varphi = aR_F(ba)a^{-1}$. Then $G(\varphi)$ admits the presentation*

$$\langle X \mid \tilde{\varphi}^\omega_{X,G}(x) = x, x \in X\rangle_G. \tag{5.5.2}$$

Proof. By Proposition 5.5.17 we have that $H = H_b a$ that implies that H is a Schützenberger group $G(\varphi)$. It follows then from Proposition 5.5.14 that H admits the presentation,

$$\langle X \mid \tilde{\varphi}^\omega_X(x) = x, x \in X\rangle_S.$$

It is known that there is a canonical projection $q : \widehat{X^*} \to \widehat{FG(X)}$ such that $\tilde{\varphi}_X = \tilde{\varphi}_{X,G} \circ q$ and taking that into consideration we can write the above presentation as follows,

$$\langle X \mid q(\tilde{\varphi}^\omega_X(x)) = q(x), x \in X\rangle_G.$$

Finally, from the diagram in Figure 5.9 we conclude that H is generated as follows,

$$\langle X \mid \tilde{\varphi}^\omega_{X,G}(x) = x, x \in X\rangle_G.$$

□

Example 5.5.19. Let $\varphi \mapsto ab, b \mapsto ba$ be the Thue-Morse morphism (defined in Example 3.1.8). The word aa is a connection for φ and $\tilde{\varphi} = \varphi^2$ is a connective power of the morphism φ. The set $R_F(aa)$ has four elements and thus so does the set $X = aR_{\mathcal{L}(\varphi)}(aa)a^{-1}$. Specifically, X has the following four elements: $x = abba$, $y = ababba$, $z = abbaba$ and $t = ababbaba$. One has $\tilde{\varphi}_X(x) = zxy$, $\tilde{\varphi}_X(y) = ztxy$, $\tilde{\varphi}_X(z) = zxty$ and $\tilde{\varphi}_X(t) = ztxty$. By Theorem 5.5.18, the group $G(\varphi)$ is the profinite group generated by X with the relations $\varphi^\omega_{X,G}(u) = u$ for $u \in X$. Actually, since $\varphi^\omega(y)\varphi^{\omega-1}(x)\varphi^\omega(z) = \varphi^\omega_{TM}(t)$, the relation $xy^{-1}z = t$ is a consequence of the relations above and thus $G(\varphi)$ is generated by x, y, z.

6

Conclusions

The current thesis focused on the topic of combinatorics on words and symbolic dynamical systems. The main subject of the thesis was the recognizability of words generated by primitive morphisms. There were presented various definitions that are supposed to be equivalent in the bibliography and it was provided a study proves their equivalence or indicates the differences among them.

Some of the many applications of the notion of recognizability were presented in Chapter Decidable properties of extension graphs and in Chapter Profinite Semigroups.

In the first one the notion of recognizability alongside a technique used in [65] were used in order to prove the decidability of different properties of extension graphs (defined in [19]) of elements of a language such as acyclic sets, dendric sets, neutral sets, etc. More precisely, the use the recognizability provided decidability of extension graphs and all the properties of the language that depend on the graphs themselves. The results presented in this thesis regarding this topic are for substitutive language of right infinite fixed points generated by a morphism. However, the results can be generalized to the case of the two-sided infinite fixed points or even to the case of the language of a morphism.

Lastly, in Chapter Profinite Semigroups the notion of recognizability was used in the subject of Profinite Semigroups. Specifically, it was presented the link with Profinite extensions of a morphism.

There are many other circumstances in which the property of recognizability is used. One of the many is the representation of

symbolic minimal dynamical systems by sequences of partitions in towers. This representation (due to Richard H. Herman, Ian F. Putnam, and Christian F. Sk [56]) is used to compute the dimension group of a minimal system (for further details see the work of Fabien Durand, Bernard Host, Dominique Perrin [38]).

Bibliography

[1] Jorge Almeida. *Finite semigroups and universal algebra*, volume 3 of *Series in Algebra*. World Scientific Publishing Co., Inc., River Edge, NJ, 1994. Translated from the 1992 Portuguese original and revised by the author.

[2] Jorge Almeida. Dynamics of implicit operations and tameness of pseudovarieties of groups. *Trans. Amer. Math. Soc.*, 354(1):387–411, 2002.

[3] Jorge Almeida. Finite semigroups: an introduction to a unified theory of pseudovarieties. In G. M. S. Gomes, J.-E. Pin, and P. V. Silva, editors, *Semigroups, Algorithms, Automata and Languages*, pages 3–64, Singapore, 2002. World Scientific.

[4] Jorge Almeida. Profinite semigroups and applications. In *Structural theory of automata, semigroups, and universal algebra*, volume 207 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 1–45. Springer, Dordrecht, 2005. Notes taken by Alfredo Costa.

[5] Jorge Almeida and Alfredo Costa. Infinite-vertex free profinite semigroupoids and symbolic dynamics. *J. Pure Appl. Algebra*, 213(5):605–631, 2009.

[6] Jorge Almeida and Alfredo Costa. On the transition semigroups of centrally labeled Rauzy graphs. *Internat. J. Algebra Comput.*, 22(2):1250018, 25, 2012.

[7] Jorge Almeida and Alfredo Costa. Presentations of Schützenberger groups of minimal subshifts. *Israel J. Math.*, 196(1):1–31, 2013.

[8] Jorge Almeida and Alfredo Costa. A geometric interpretation of the Schützenberger group of a minimal subshift. *Ark. Mat.*, 54(2):243–275, 2016.

[9] Jorge Almeida and Alfredo Costa. Equidivisible pseudovarieties of semigroups. *Publ. Math. Debrecen*, 90:435–453, 2017.

[10] Jorge Almeida, Alfredo Costa, José Carlos Costa, and Marc Zeitoun. The linear nature of pseudowords. Technical report, 2017. arXiv:1702.08083 [cs.FL].

[11] Jorge Almeida, Alfredo Costa, Revekka Kyriakoglou, and Dominique Perrin. On the group of a rational maximal bifix code. *17e journées montoises d'informatique théorique*, 2018.

[12] Jorge Almeida, Alfredo Costa, Revekka Kyriakoglou, and Dominique Perrin. Profinite semigroups and symbolic dynamics. in preparation.

[13] Jorge Almeida and Benjamin Steinberg. Rational codes and free profinite monoids. *J. Lond. Math. Soc. (2)*, 79(2):465–477, 2009.

[14] Jorge Almeida and Mikhail Volkov. Subword complexity of profinite words and subgroups of free profinite semigroups. *Internat. J. Algebra Comput.*, 16:221–258, 2006.

[15] Zh. Almeĭda. Profinite groups associated with weakly primitive substitutions. *Fundam. Prikl. Mat.*, 11(3):13–48, 2005.

[16] Lubomíra Balková, Edita Pelantová, and Wolfgang Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4):251–263, 2008.

[17] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.

[18] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.

[19] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatshefte für Mathematik*, 176(4):521–550, 2015.

[20] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and interval exchanges. *Journal of Pure and Applied Algebra*, 219(7):2781 – 2798, 2015.

[21] 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.

[22] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Dicrete Math.*, 338:725–742, 2015.

[23] S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra*. Number 78 in Grad. Texts in Math. Springer, Berlin, 1981.

[24] Julien Cassaigne. An algorithm to test if a given circular HD0L-language avoids a pattern. In *Information processing '94, Vol. I (Hamburg, 1994)*, IFIP Trans. A Comput. Sci. Tech., A-51, pages 459–464. North-Holland, Amsterdam, 1994.

[25] Julien Cassaigne. Complexity and special factors. (complexit et facteurs spciaux.). *Bulletin of the Belgian Mathematical Society - Simon Stevin*, 4(1):67–88, 1997.

[26] Christian Choffrut and Juhani Karhumäki. Combinatorics of words. In *Handbook of formal languages, Vol. 1*, pages 329–438. Springer, Berlin, 1997.

[27] Alfredo Costa. Conjugacy invariants of subshifts: an approach from profinite semigroup theory. *IJAC*, 16:629–655, 2006.

[28] Alfredo Costa and Benjamin Steinberg. Profinite groups associated to sofic shifts are free. *Proc. Lond. Math. Soc. (3)*, 102(2):341–369, 2011.

[29] José Carlos Costa. Free profinite locally idempotent and locally commutative semigroups. *JPAA*, 163:19–47, 2001.

[30] Thierry Coulbois, Mark Sapir, and Pascal Weil. A note on the continuous extensions of injective morphisms between free groups to relatively free profinite groups. *Publ. Mat.*, 47(2):477–487, 2003.

[31] Andreas Distler and Tom Kelsey. The semigroups of order 9 and their automorphism groups. *Semigroup Forum*, 88(1):93–112, 2014.

[32] Andreas Distler and James D. Mitchell. The number of nilpotent semigroups of degree 3. *Electron. J. Combin.*, 19(2):51–19, 2012.

[33] Francesco Dolce and Dominique Perrin. Neutral and tree sets of arbitrary characteristic. *Theoretical Computer Science*, pages –, 2016.

[34] Francesco Dolce and Perrin Perrin. Eventually dendric shifts. In *in: CSR COMPLETER*, 2019.

[35] Francesco Donce, Revekka Kyriakoglou, and Julien Leroy. Decidable properties of extension graphs for substitutive languages. *15èmes Journées Montoises dinformatique théorique*, 2016.

[36] 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.

[37] Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Mat.*, 179:89–101, 1998.

[38] Fabien Durand, Bernard Host, and Dominique Perrin. Dimension groups. 2019.

[39] Fabien Durand, Bernard Host, and Christian Skau. Substitutional dynamical systems, bratteli diagrams and dimension groups. *Ergodic Theory and Dynamical Systems*, 19(4):953–993, 1999.

[40] Fabien Durand and Julien Leroy. The constant of recognizability is computable for primitive morphisms. *Preprint*.

[41] Fabien Durand, Julien Leroy, and Gwenaël Richomme. Do the properties of an S-adic representation determine factor complexity? *J. Integer Seq.*, 16(2):Article 13.2.6, 30, 2013.

[42] Andrzej Ehrenfeucht and Grzegorz Rozenberg. Elementary homomorphisms and a solution of the d0l sequence equivalence problem. *Theoretical Computer Science*, 7(2):169–183, 1978.

[43] Andrzej Ehrenfeucht and Grzegorz Rozenberg. Simplifications of homomorphisms. *Inform. and Control*, 38(3):298–309, 1978.

[44] Andrzej Ehrenfeucht and Grzegorz Rozenberg. Repetition of subwords in dol languages. *Information and Control*, 59(1):13 – 35, 1983.

[45] Samuel Eilenberg. *Automata, Languages and Machines*, volume A. Academic Press, New York, 1974.

[46] Samuel Eilenberg. *Automata, Languages and Machines*, volume B. Academic Press, New York, 1976.

[47] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[48] Anna E. Frid and Sergey V. Avgustinovich. On bispecial words and subword complexity of D0L sequences. In C. Ding, T. Helleseth, and H. Niederreiter, editors, *Sequences and their Applications*, pages 191–204. Springer London, 1999.

[49] Michael D. Fried and Moshe Jarden. *Field arithmetic*, volume 11 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, third edition, 2008. Revised by Jarden.

[50] Amy Glen and Jacques Justin. Episturmian words: a survey. *Theor. Inform. Appl.*, 43:403–442, 2009.

[51] Samuel J. van Gool and Benjamin Steinberg. Pro-aperiodic monoids via saturated models. Technical report, 2016. arXiv:1609.07736.

[52] Marshall Hall, Jr. Coset representations in free groups. *Trans. Amer. Math. Soc.*, 67:421–432, 1949.

[53] Marshall Hall, Jr. A topology for free groups and related groups. *Ann. of Math. (2)*, 52:127–139, 1950.

[54] Tero Harju and Matti Linna. On the periodicity of morphisms on free monoids. *RAIRO*, 20:47–54, 1986.

[55] Wolfgang Herfort. Introduction to profinite groups.

[56] Richard H Herman, Ian F Putnam, and Christian F Skau. Ordered bratteli diagrams, dimension groups and topological dynamics. *International Journal of Mathematics*, 3(06):827–864, 1992.

[57] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[58] Bernard Host. Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable. *Ergodic Theory Dynam. Systems*, 6(4):529–540, 1986.

[59] Shunji Ito and Yoichihiro Takahashi. Markov subshifts and realization of β-expansions. *J. Math. Soc. Japan*, 26:33–55, 1974.

[60] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.

[61] Ilya Kapovich and Alexei Myasnikov. Stallings foldings and subgroups of free groups. *J. Algebra*, 248:608–668, 2002.

[62] Lila Kari, Grzegorz Rozenberg, and Arto Salomaa. L systems. In *Handbook of formal languages, Vol. 1*, pages 253–328. Springer, Berlin, 1997.

[63] Karel Klouda. Bispecial factors in circular non-pushy d0l languages. *Theor. Comput. Sci.*, 445:63–74, 2012.

[64] Karel Klouda and Kateina Medková. Synchronizing delay for binary uniform morphisms. *Theoretical Computer Science*, 615:12 – 22, 2016.

[65] Karel Klouda and Štěpán Starosta. Characterization of circular d0l-systems. *Theoretical Computer Science*, 2019.

[66] Donald E. Knuth. *The art of computer programming. Vol. 2.* Addison-Wesley, Reading, MA, 1998. Seminumerical algorithms, Third edition [of MR0286318].

[67] Neal Koblitz. *p-adic numbers, p-adic analysis, and zeta-functions*, volume 58 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1984.

[68] Petr Kůrka. *Topological and symbolic dynamics*, volume 11 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, 2003.

[69] Revekka Kyriakoglou and Dominique Perrin. Profinite semigroups. *arXiv preprint arXiv:1703.10088*, 2017.

[70] Gérard Lallement. *Semigroups and combinatorial applications*. John Wiley & Sons, New York-Chichester-Brisbane, 1979. Pure and Applied Mathematics, A Wiley-Interscience Publication.

[71] Hendrick Lenstra. Profinite Fibonacci numbers. *Nieuw Archief voor Wiskunde*, 6:297–300, 2005.

[72] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

[73] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge, UK, 2002.

[74] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1977 edition.

[75] S. Margolis, Mark Sapir, and Pascal Weil. Irreducibility of certain pseudovarieties. *Comm. Algebra*, 26(3):779–792, 1998.

[76] Stuart W. Margolis, Mark Sapir, and Pascal Weil. Closed subgroups in pro-$\mathbf{V}$ topologies and the extension problem for inverse automata. *Internat. J. Algebra Comput.*, 11(4):405–445, 2001.

[77] John C. Martin. Minimal flows arising from substitutions of non-constant length. *Math. Systems Theory*, 7:72–82, 1973.

[78] J. D. McKnight, Jr. and A. J. Storey. Equidivisible semigroups. *J. Algebra*, 12:24–48, 1969.

[79] Filippo Mignosi and Patrice Séébold. If a DOL language is k-power free then it is circular. In *Automata, languages and programming (Lund, 1993)*, volume 700 of *Lecture Notes in Comput. Sci.*, pages 507–518. Springer, Berlin, 1993.

[80] Brigitte Mossé. Puissances de mots et reconnaissabilité des points fixes d'une substitution. *Theoret. Comput. Sci.*, 99(2):327–334, 1992.

[81] Brigitte Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France*, 124(2):329–346, 1996.

[82] Jean-Jacques Pansiot. Decidability of periodicity for infinite words. *RAIRO*, 20:43–46, 1986.

[83] Dominique Perrin. Codes and automata in minimal sets. In *Combinatorics on Words - 10th International Conference, WORDS 2015, Kiel, Germany, September 14-17, 2015, Proceedings*, pages 35–46, 2015.

[84] Dominique Perrin and Jean-Éricric Pin, editors. *Infinite WordsAutomata, Semigroups, Logic and Games*, volume 141 of *Pure and Applied Mathematics*. Elsevier, 2004.

[85] Jean-Éric Pin. *Varieties of Formal Languages*. Plenum, London, 1986. English translation.

[86] Jean-Éric Pin. Profinite methods in automata theory. In *26th International Symposium on Theoretical Aspects of Computer Science STACS 2009*, pages 31–50. IBFI Schloss Dagstuhl, 2009.

[87] Martine Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2010. first edition, 1987.

[88] Christophe Reutenauer. Une topologie du monoïde libre. *Semigroup Forum*, 18(1):33–49, 1979.

[89] John Rhodes and Benjamin Steinberg. *The q-theory of finite semigroups*. Springer Monographs in Mathematics. Springer, 2009.

[90] Luis Ribes and Pavel Zalesskii. *Profinite groups*, volume 40 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, second edition, 2010.

[91] S. Satoh, K. Yama, and M. Tokizawa. Semigroups of order 8. *SF*, 49:7–30, 1994.

[92] Marcel-Paul Schtzenberger. On finite monoids having only trivial subgroups. *Inform. and Control*, 8:190–194, 1965.

[93] John R. Stallings. Topology of finite graphs. *Inventiones Mathematicae*, 71:551–565, 1983.

[94] Stephen Willard. *General topology*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1970 original [Addison-Wesley, Reading, MA; MR0264581].

[95] Hisatishi Yuasa. A review of Mossé's recognizability theorem. unpublished, 2014.