Languages, D0L-Systems, Sets of Curves, and Surface Automorphisms

Luis-Miguel Lopez* **and Philippe Narbel****

(*) Sunshine College, Higashi-Ikebukuro 4-23-4, Toshima-Ku, 170-8434 TOKYO.
(**) LABRI, Université Bordeaux I. 351, Cours de la Libération 33405 Talence, France
e-mails: lopez@sunshine.ac.jp,, narbel@labri.u-bordeaux.fr

(Information and Computation, Academic Press, Vol. 180 (1), pp. 30-52, 2003.)

Abstract. We introduce,extend and apply some relationships between formal language theory and surface theory. First we show how boundaries of languages topologized with the Cantor metric can be mapped to sets of curves on surfaces, namely laminations. Second we present how and when endomorphisms of free monoids, i.e. substitutions, can be mapped to automorphisms of surfaces, so that D0L-systems correspond to iterations of these automorphisms. Third, we apply these ideas to construct sets of non-periodic irreducible automorphisms of surfaces following [43, 35, 14], so that the involved proofs do not use any differential or algebraic tools, but, accordingly, substitution and D0L-system theory of [11, 12, 37, 31, 32] – mainly the decidability for a D0L-language [12] to be strongly repetitive or not. This construction also yields effective symbolic descriptions of the stable sets of curves associated with these automorphisms.

Keywords. Free monoids, substitutions, D0L-systems, repetitiveness, surface theory, surface automorphisms, Nielsen-Thurston's Theorem, laminations, train-tracks.

1 Introduction

This paper introduces, extends and apply several relationships between formal language theory and surface theory:

1. Boundaries of languages [4, 32] are related to specific sets of curves on surfaces, i.e. laminations – see definitions in Sections 2.1 and 2.3.

2. Endomorphisms on words (substitutions) are related to specific transformations of surfaces, i.e. surface automorphisms.

3. Iterating endomorphisms, i.e. using D0L-systems (e.g. [39, 40]) – see exact definition in Section 3.1 – is related to iterating automorphisms of surfaces.

The idea of representing curves by words, and transformations of curves by transformations of words is quite old, probably coming first from Dehn and Morse [8, 10, 29, 30]. It is still in use (see e.g [2, 3]). With respect to this, the above relationships are an attempt to go one step further by showing how specific results of formal language theory – here, substitutions and D0L-system theory – directly apply to such symbolic representations. From the surface theory point-of-view, the stakes are to obtain proofs with a minimum need of geometric, differential or measure theoretic tools, and to obtain full constructive proofs with tractable algorithms. On the other hand, from the formal language theory point-of-view, classical notions could be treated in new ways by using geometric representations.

The first above relationship involves languages and sets of curves on surfaces. These sets are made of non self-intersecting pairwise disjoint infinite curves called *laminations* (see e.g. [43, 7, 19, 27]), which are closely related to *singular foliations*. Laminations can be combinatorially described by using graphs – also called *train-tracks* (see e.g. [19, 2, 36]) – on which curves can be continuously deformed; one then says that curves are *carried*. When the carrying graphs are labelled, it becomes possible to code carried sets of curves into languages : curves are just represented by the labels of the admissible paths on which they can be deformed. By using the classical topology on words, our first results state how and when *boundaries of languages* (see [4, 32]) are lamination codings. Informally, these results can be summed up as follows (see Theorem 2.7 for a full statement):

Theorem: *Let Γ be a directed graph embedded in a surface M. Let L be the coding language of a set of simple finite pairwise disjoint curves carried by Γ. Then, a maximal lamination in M rel. to Γ corresponds to the boundary of L.*

The second above relationship is based on the fact that surface automorphisms, i.e. bijective bicontinuous maps of a surface onto itself, can be conjugated to automorphisms of *free groups* by using *graph maps* (or *train-track maps*, see e.g. [2, 3]). Here we focus on those which can directly be conjugated to morphisms of *free monoids*, in fact endomorphisms on words, i.e. substitutions (an idea already in [22, 25, 23]). With respect to this, the main result we get is an extension of the above one which leads to effective descriptions of laminations (see Theorem 3.3):

Theorem: *Let Γ be a directed graph embedded in a surface M, invariant under an automorphism h of M. Then there is a lamination which is stable under h and which*

corresponds to the boundary of a language obtained by iterating a substitution, i.e. the boundary of a D0L-language.

The third above relationship is based on the following fact: if iterating an automorphism conjugates to iterating a substitution, results in D0L-systems theory (see e.g. [39, 40]) apply. Here we use these relationships to study the group of automorphisms of a compact oriented surface. These automorphisms have been classified a long time ago as periodic, reducible or non-periodic irreducible [33]. More recently, this classification has been made more precise [15, 28, 18, 7, 43]. In particular, non-periodic irreducible automorphisms were proved to enjoy the property of setwise fixing two laminations. Much effort has been dedicated to finding systematic constructions of these automorphisms and their fixed laminations [1, 13, 35, 43, 7, 34, 41], with the recent achievement of obtaining procedures of decision for the three families [26, 3, 17]. Within this framework, we here focus on a generic construction of non-periodic irreducible automophisms indicated by Thurston [43], and generalized by Penner [35, 14]. Based on the above results, the novelty of our proof is that it makes only use of symbolic tools, i.e. results about substitutions and D0L-system theory found in [11, 12, 37, 31, 32], instead of differential, measure theoretic or linear algebraic tools. Another benefit of this approach is to yield an effective description of the associated fixed laminations: individual curves and all subcurves with prescribed finite lengths can be described by computable words. The results we obtain can be then stated as (see Theorems 4.1 and 4.13):

Theorem: *For every compact surface M, semi-groups of non-periodic irreducible automorphisms can be obtained, so that their fixed laminations are described by boundaries of D0L-languages.*

Our proof of this goes as follows: Given a compact surface M, two mutually transverse finite sets of pairwise disjoint simple closed curves $C \cup D$ are embedded in M. These curves can be associated with basic automorphisms called *Dehn twists*. The above theorem is then the consequence of the fact that every positive composition involving at least once each of these twists is a non-periodic and irreducible automorphism of M. The first steps we take to prove this follow the classical ones [35]: We consider the labelled directed graph Γ arising from $C \cup D$. It is shown invariant under the application of the associated set of Dehn twists. As a consequence, Dehn twists can be interpreted as graph maps. The next steps are different as we stick to symbolic tools: given an automorphism h described in terms of the considered Dehn twists, we apply our above relationships to obtain one of the two corresponding fixed laminations. Its coding is accordingly a set of two-way infinite words given by the boundary $Bi(L_h)$ of a D0L language L_h based on the iterations of a substitution θ_h associated with h (see [32]). The properties of h are thus checked against those of θ_h: Since θ_h is strictly expanding (*strongly growing*), h cannot be periodic. Next, deciding irreducibility of h is shown to be the same as deciding whether or not $Bi(L_h)$ contains periodic words. This is equivalent to deciding whether or not the corresponding D0L-language is *strongly repetitive*, so that one can use the decidability result of Ehrenfeucht and Rozenberg [12]. This amounts in our case to checking that θ_h is a *primitive*, *elementary*, and *non-cyclic* substitution. As a final step, we also show how to build the other fixed lamination of h by "inverting" θ_h, an operation made possible since θ_h is also *recognizable* (see [37, 31]).

2 Curves and Words

2.1 Basic Definitions

A **surface** M is a topological space locally modelled after $\mathbb{R}^2$, that is a *2-manifold*. We here consider only closed oriented surfaces with negative Euler-Poincaré characteristic, i.e. those without boundary having a finite triangulation with negative Euler characteristic. Such a surface, a one with two holes is pictured in Fig. 7. These surfaces have metrics for which the universal covering is the hyperbolic plane; from now on we always assume M has been given such a metric.

A **simple curve** on M is a non self-intersecting curve, i.e. an injective continuous map from an interval of the real line or from the circle to M. Surfaces as defined above possess many subsets of pairwise disjoint simple curves (see e.g. [15, 5, 42]). For instance, solutions of differential equations give rise to such sets of curves. Here we shall consider sets of curves which can be constructed in a combinatorial way.

Let $\Gamma = (V, E)$ be a finite directed graph embedded in M with set of vertices V and set of edges E. An **admissible path** in Γ is an indexing map from an interval of $\mathbb{Z}$ towards E, such that the end of an edge is the origin of the following one. It inherits in an obvious way an orientation. According to if the interval is finite, half-infinite or all of $\mathbb{Z}$, a path is respectively said to be **finite**, **one-way infinite** or **two-way infinite**. It is said to be **closed** if it is defined by a two-way infinite periodic indexing map.

Two curves γ, γ' on M are said **homotopic rel. endpoints**, or for short just **homotopic**, if there is a uniformly continuous deformation fixing their endpoints (if any) which sends γ to γ' (see e.g. [38]). Uniformity is here required to allow one to speak about infinite paths and curves: uniform homotopy is equivalent to ask that two homotopic curves have lifts to M's universal covering at bounded distance from each other. A curve γ in M is said to be **carried by** a graph Γ if it is homotopic to an admissible path in Γ. An example is shown in Figure 1. Graphs like Γ used to carry curves are often called

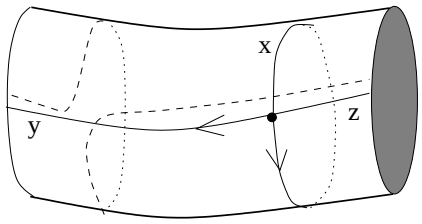

Figure 1: *A cylinder-like piece of a surface M where a piece of a curve (dashed style) is carried by a piece of an embedded graph (normal style) along the path labelled zxy.*

train-tracks (see e.g. [20, 19, 36]). Note however that we consider only globally directed graphs so that all the carried curves inherit that global orientation. From the set of curves viewpoint, this implies no loss of generality up to considering *orientation double coverings* (see e.g. [16]) of the carrying graphs.

A **geodesic lamination** of M is a closed set of complete simple geodesics (see e.g. [7]). Up to homotopy there is another equivalent definition based on embedded graphs (close to the one defined in [19]): A **lamination** is a set $\mathcal{L}$ of pairwise disjoint, pairwise non-

uniformly homotopic, simple infinite curves called **leaves**, carried by some graph Γ such that $\mathcal{L}$ is maximal with respect to inclusion. Given Γ and the lamination, we say that the lamination is **maximal rel.** to Γ.

2.2 Coding Sets of Curves into Languages

Let the edges of the graph Γ be injectively labelled over an alphabet A. The **coding of an admissible path** of Γ is the word obtained by concatenating the edges labels according to the indexing map. An embedded graph Γ is said to be **free** if it does not have distinct admissible paths which are homotopic on M. A necessary and sufficient condition to be free is that no disk or annulus component in $M \setminus \Gamma$ has its boundary made of two admissible paths. Given a free graph, the **coding of a carried leaf** ℓ, denoted by $cod(\ell)$, can be defined as the coding of its unique homotopic admissible path. By extension, the **coding of a carried set of curves** $\mathcal{C}$ is the language of the codings of all the curves of $\mathcal{C}$.

Note that this coding process can be seen as if Γ would be an embedded finite Büchi automaton with all its states being final. Accordingly, *cod* is then a correspondence from a maximal lamination $\mathcal{L}$ rel. to Γ to the whole recognized language. However, even if a lamination is maximal with respect to inclusion, *cod* has no reason to be surjective since leaves must be pairwise disjoint, which is a geometric constraint inducing a strict subset of all the admissible paths.

2.3 Set of Curves and Boundaries of Languages

Let L be a language of finite words over an alphabet A. Its so-called **pointed language** $\widehat{L}$ is obtained by giving every possible origin to each word of L indexing its remaining letters accordingly. These finite words can be assumed to be padded to both infinities with some dummy symbol, say δ, not already in A. The language $\widehat{L}$ becomes then a subset of $(A \cup \{\delta\})^{\mathbb{Z}}$ and inherits the topology of this set with the usual **Cantor metric** defined as follows: let $\dots w_{-1}w_0w_1\dots$ and $\dots v_{-1}v_0v_1\dots$ be two-way infinite words, then their distance is 0 if they are equal, and 2^{-n} if they are not, where $n \geq 0$ is the largest integer for which $w_k = v_k$, $|k| \leq n$.

The **boundary** of $\widehat{L}$ (see [32]) is the set of completion points of $\widehat{L}$ in $(A \cup \{\delta\})^{\mathbb{Z}}$. It is a set of one-way and two-way infinite words over A. The **shift** of a word $w = \dots w(-1)w(0)w(1)\dots$ is defined as $\sigma(w)(i) = w(i+1)$, $i \in \mathbb{Z}$. A pointed language can be "unpointed" by using the equivalence relation: $w \sim w'$ iff there is a power $n \in \mathbb{Z}$ such that $\sigma^n(w) = w'$. We denote by $Bi(L)$ a set of representatives of the quotient by $\sim$ of the boundary of $\widehat{L}$.

The above could seem quite intricate for the single purpose of obtaining asymptotic languages, but there is no simple way of directly defining a meaningful topology on the set of unpointed words. For curves it is easier as a topology already exists for M: one can readily use the Hausdorff topology for closed sets in M. For this topology a sequence of closed sets $\{\gamma_n\}$ in M converges to some set γ iff for each point x of γ there is a sequence of points $x_n \in \gamma_n$ such that x_n converges to x and conversely all such converging sequences of points have their limits in γ. If the γ_n are curves carried by an embedded graph Γ no pointing is needed in these definitions.

The next technical lemma shows the relationship between sets of finite curves carried by a graph Γ, their closures for the Hausdorff topology, and the associated coding words. It can be seen as a combinatorial and topological version of the classical result about geodesic laminations, saying that the closure of any non-empty disjoint union of geodesics is a geodesic lamination (see e.g. [7]).

An embedded graph Γ is said to be **coherent** iff for every vertex of Γ all its incoming edges are adjacent (so that all its outgoing edges are as well). From the set of curves viewpoint, coherence is not a restrictive assumption. Indeed, given a set of curves $\mathcal{C}$ carried by Γ, every vertex not satisfying the coherence condition can be split into n vertices each satisfying this condition, where n is the number of maximal sets of adjacent edges being either incoming or outgoing. These vertices are linked in the natural way so to make an n-gon in $M \setminus \Gamma$ together with some of its diagonals, in such a way that $\mathcal{C}$ is still carried. If Γ is free, every coherent graph obtained from Γ by this operation remains free

Lemma **2.1** *Let $\Gamma = (V, E)$ be a free coherent directed graph labelled over A and embedded in a surface M. Let L be a language of finite words over A, and let $\mathcal{C}_L$ be a set of finite simple curves in M carried by Γ such that:*
- Each word is the coding of a curve in $\mathcal{C}_L$ and vice-versa;
- Up to moving their extremities in an arbitrarily small neighbourhood of Γ's vertices, all the curves in $\mathcal{C}_L$ can be made simultaneously pairwise disjoint on M;
- The boundary of L is not empty.
Then $Bi(L)$ is the coding of a closed set of curves $\mathcal{C}$ which are pairwise disjoint, pairwise non-homotopic and two-way infinite.

Proof.

Step 1 (Construction of a closed regular neighbourhood of Γ in M):

Since Γ is assumed coherent and directed, to each vertex $v \in V$, we associate a closed unit square whose two opposite sides are respectively labelled $(v, origin)$ and (v, end). These sides correspond respectively to the incoming and outgoing edges of Γ at v. Next, to each edge $\eta \in E$ with origin v_1 and end v_2, we associate the symmetry axis of a closed Euclidean isosceles trapezium of height 1, whose base lines are respectively labelled $(v_1, origin)$ and (v_2, end). The lengths of the base lines of the trapezia can be fixed so that the sums of the lengths of the base lines with same common label equal 1 (this is the advantage of using trapezia). Trapezia are then glued with squares along their identical labelled sides following the same adjacencies as Γ's edges. This yields a surface N which is a closed regular neighbourhood of Γ in M (see Figures 3 and 4 for a local appearance of N). Note that N is fibered with arcs parallel to the trapezia bases.

Step 2 (Existence of a closed set of two-way infinite curves $\mathcal{C}$ in M, all carried by Γ, and derived from $\mathcal{C}_L$):

According to the hypothesis, the curves in $\mathcal{C}_L$ can be moved in N so to be pairwise disjoint with extremities fixed at the base lines of the trapezia (recall that a finite curve carried by Γ must have its endpoints on Γ's vertices). This move can also be achieved so that each curve point lies at distance at least $\epsilon > 0$ from the boundary of N in M, and we can also assume that each segment of curve crossing a quadrilateral of N is linear.

We take the closure of $\mathcal{C}_L$ in M with respect to Hausdorff topology for the closed sets in M (each curve in $\mathcal{C}_L$ is compact). We claim that this closure can also be seen as a set of curves: a limit point in the interior of a quadrilateral Q_i of N is the limit of a sequence of points on linear segments in the quadrilateral, hence the limit point itself belongs to a unique linear segment ℓ disjoint from the arcs induced by $\mathcal{C}_L$. Moreover all the points of ℓ lie in the closure. Now, keeping in mind the definition of Hausdorff topology and the relative positions of the curves of $\mathcal{C}_L$ in N, we look at the two adjacent quadrilaterals Q_j, Q_k, containing ℓ's extremities. Only two cases may occur: either there are two other linear limit segments in Q_j and Q_k which fit at ℓ's extremities, or only one in either Q_j or Q_k (this case corresponds to a sequence of curves in $\mathcal{C}_L$ with one extremity lying in the same quadrilateral). Inductively applying this reasoning to the limit segments through adjacent quadrilaterals yields curves: the former above case yields two-way infinite curves and the latter one-way infinite curves. This is so since the boundary of L is assumed not empty: $\mathcal{C}_L$ must be made of infinitely many curves for which there is only a finite number of them which have a smaller length than any specified value.

We consider only the two-way infinite curves and we denote the set they form by $\mathcal{C}$. Again by definition of convergence in the Hausdorff topology a converging sequence of points belonging to two-way infinite curves necessarily converges to a point on a two-way infinite curve. Hence $\mathcal{C}$ is closed as a set of points in M.

Step 3 (The coding of $\mathcal{C}$ can be made to be exactly $Bi(L)$):

By construction, the coding of a curve γ in $\mathcal{C}$ has its subwords in L. Thus, by definition of the boundary of L, and after choosing an adequate origin for γ, its coding belongs to $Bi(L)$.

Conversely, take a pointed two-way infinite word $w \in Bi(L)$. Consider a sequence of words in $\widehat{L}$ converging to w. There is a corresponding sequence s_w of finite curves in $\mathcal{C}_L$ with an origin fixed on them. We must prove that it determines a limit curve belonging to $\mathcal{C}$: Consider the trapezium Q corresponding to the edge labelled with the origin letter of w. We can extract from s_w a sequence whose intersection segments with Q converge to some segment because Q is compact. Next the same is done for the quadrilaterals adjacent to Q and the preceeding extracted sequence and so on, thus obtaining a decreasing sequence (in the setwise sense) of sequences. Using a diagonal argument, the sequence formed by the n-th curve of the n-th sequence gives a sequence of finite curves which converges to a limit curve with origin marked in Q and whose coding is w.

Step 4 (Making $\mathcal{C}$ a set of non-pointed simple pairwise disjoint limit curves):

We have proved that the closed subset $\mathcal{C}$ of the closure of $\mathcal{C}_L$ can be seen as a set of two-way infinite curves. However, these need not be pairwise disjoint, nor even simple. An intersection component of two curves must be also a curve, finite or not, with at least one endpoint on a base line of one of the quadrilaterals making N. Recall also that N can be fibered by arcs parallel to the trapezia bases.

Now, consider the intersection of $\mathcal{C}$ with $|V|$ fibers, one for each unit square of N linking trapezia together. Since $\mathcal{C}$ is closed its intersection with these fibers is closed, hence the complement is open, and so this complement is an at most enumerable disjoint union of open intervals whose closures we call here **transverse segments**. Consider

all of them and follow them by parallel translation towards the two possible directions, subject to the following rules: (1) If in this process a transverse segment hits a trapezium corner, just stop; (2) Else if a transverse segment shrinks to a single point, mark this point as an intersection curve endpoint; (3) Else if the process never stops towards both directions, do nothing. Rule (2) exhausts all the intersection curves endpoints. Hence, there are at most an enumerable number of them. We then cut open along all these curves starting from their marking points (an idea already used in [19]), and we undergo again the process of translating the transverse segments. But this time only the conditions of rules (1) and (3) may occur. The former means that the two curves visiting the endpoints of the translated segment are non-homotopic. The latter means that these two curves are homotopic: performing this homotopy in M makes one curve out of the two. After these transformations, the curves in $\mathcal{C}$ are pairwise disjoint and pairwise non-homotopic. $\diamondsuit$

We now look for the conditions to ensure that $\mathcal{C}$ is a lamination, maximal rel. to the graph Γ which carries it. For this purpose, we introduce two graph transformations. First, the **trimming** of Γ rel. to a set of two-way infinite curves $\mathcal{C}$ consists of the removing of every edge and vertex of Γ not used to carry curves in $\mathcal{C}$. The second transformation is the **cutting** and involves several definitions:
- An **in-vertex cutting** of v (see Figure 2) is defined by a 3-tuple (a_1, a_2, b_1) where a_1,a_2 are two distinct adjacent incoming edges at v, and b_1 is an outgoing edge linking v to another vertex v'. The in-vertex cutting of v consists in splitting v into two vertices v_1, v_2, erasing b_1 and replacing it with two edges b_1 and b_2 linking v_j to v' for $j = 1, 2$, linking a_j to v_j for $j = 1, 2$, and in distributing the remaining incident edges of v so that the former adjacency relation around v is kept from a_j to b_j for $j = 1, 2$.
- A **mid-vertex cutting** of v (see Figure 2) is similarly defined by a 4-tuple (a_1, a_2, b_1, b_2) where a_1, a_2 (resp. b_1,b_2) are two adjacent incoming (resp. outgoing) edges of v. The difference is that after the splitting of v, the edge b_j is linked to v_j for $j = 1, 2$. -

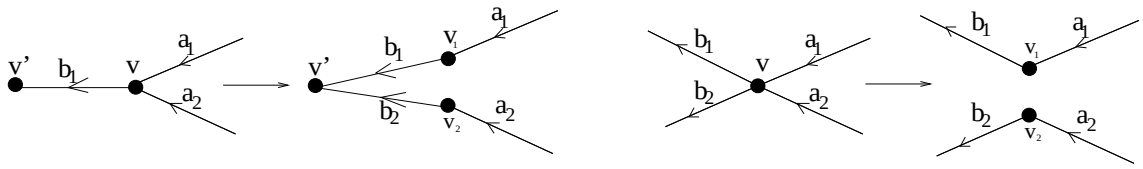

Figure 2: *To the left, an in-vertex cutting, and to the right, an mid-vertex cutting.*

A **cutting curve** ρ is a curve carried by Γ starting at a vertex v where two adjacent incoming edges are marked, and ending at a vertex v' where two adjacent outgoing edges are marked. Cutting along ρ consists then in applying in-vertex cuttings successively to every vertex along the admissible path in Γ associated with ρ, fixing the incoming pair of edges as the two edges resulting from the preceding split edge. This is done until the last vertex, at which a mid-vertex cutting is applied. If $v = v'$, the cutting curve is of length 0 and amounts to a plain mid-vertex cutting.
- A **cutting curve rel. to** $\mathcal{C}$ is a one such that after cutting along it, the curves in $\mathcal{C}$ are all still carried by the resulting graph. Such a move is similar to the so-called *collision* used in the framework of *measured train-tracks* theory (see e.g. p.119 of [36]).

- The **cutting** of a graph Γ rel. to $\mathcal{C}$ amounts to performing all the cuttings along a maximal set of cutting curves rel. to $\mathcal{C}$. If a cutting creates new edges, the labelling of the new graph is extended so that each edge has again a distinct label.

Now, cutting is related to the notion of maximality rel. to a carrying graph as follows:

Lemma **2.2** *Let $\mathcal{C}$ be a closed set of pairwise disjoint, pairwise non-homotopic and two-way infinite curves carried by Γ. If a cutting curve rel. to $\mathcal{C}$ exists, then there exists a carried simple two-way infinite curve γ non-homotopic to any curve of $\mathcal{C}$, and such that $\{\gamma\} \cup \mathcal{C}$ is still a set of pairwise disjoint curves.*

Proof. Let N be a regular neighborhood of Γ as described in the proof of Lemma 2.1. A cutting curve ρ is a finite curve within N carried by Γ, and its extremities can be moved so to lie in corners of N, so to be respectively the intersections of trapezia Q_i, Q_j and Q_k, Q_l (see Figure 3). Assume that following the orientation of Γ, the trapezium

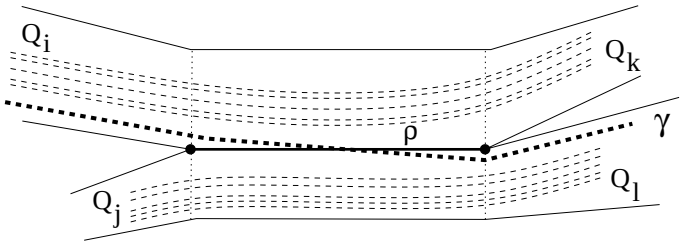

Figure 3: *Inserting a curve γ (dashed bold style) when a cutting curve ρ exists (bold style).*

Q_i is above Q_j and Q_k is above Q_l. Let γ_{min} (respect. γ_{max}) be the minimum (respect. maximum) curve passing through Q_i (respect. Q_l). A curve γ is then constructed as the cutting curve stretched by following γ_{min} in one direction and γ_{max} in the other. This curve has the announced properties. $\diamondsuit$

Lemma **2.3** *Let $\mathcal{C}$ be as in Lemma 2.2. Then there is most a finite number of cutting curves rel. to $\mathcal{C}$ which are pairwise non-homotopic and pairwise disjoint after homotopy in N.*

Proof. Since the cutting curves can be taken pairwise disjoint in N, cutting along anyone of these curves leaves the other cutting curves unchanged. In particular, these remain cutting curves for the resulting graph. Now, each cutting along a cutting curve implies at least one strict decreasing of the incoming (respect. outgoing) degree of its starting (respect. ending) vertex) without increasing the others. Thus, the number of cutting curves is bounded by the number of edges of Γ. $\diamondsuit$

Corollary **2.4** *A cutting rel. to $\mathcal{C}$ can involve only a finite number of cutting curves.*

Lemma **2.5** *Let Γ be a free directed graph embedded in M. After trimming and cutting, the resulting graph is still a free graph (so that the coding of its carried leaves is still well-defined).*

Proof. This is clear for trimming. For cutting, first consider in-vertex cuttings: they do not change the type of any component of $M \setminus \Gamma$. After a mid-vertex cutting, a priori a disk or an annulus could be created, but this is impossible since this would imply the existence of a disk in $M \setminus \Gamma$ with its boundary made of two admissible paths, contradicting that Γ is free. $\diamondsuit$

Lemma **2.6** *Let $\mathcal{C}$ be as in Lemma 2.2. Let the carrying graph Γ be a free and coherent directed graph embedded in M on which trimming and cutting have been applied rel. to $\mathcal{C}$. Then $\mathcal{C}$ is a maximal lamination rel. to Γ.*

Proof. Let γ be a two-way infinite curve carried by Γ, missing $\mathcal{C}$ and not parallel to any of its curves. We put γ in the regular neighbourhood N defined as in the proof of Lemma 2.1.

We claim there is a transverse segment $[x', x'']$ which has a point x of intersection with γ and such that (see next Figure 4) there are two curves γ' and γ'' (not necessarily distinct) in $\mathcal{C}$ going respectively through x' and x''.

Begin with any tranverse segment $[x', x'']$ meeting γ. If both endpoints lie on curves of $\mathcal{C}$ the claim is proved. If not, at least one among x', x'', say x'', is visited by some curve in $\mathcal{C}$. This is because Γ is trimmed, hence there is at least one two-way infinite curve going through every trapezium. In this case the other endpoint of the segment, x', lies on the boundary of N. If translation of $[x', x'']$ can be undergone forever towards one direction, γ'', hence γ will both be homotopic to a boundary component of N, and necessarily the same one: a contradiction. So translation must stop somewhere, and it can stop only when a translate of $[x', x'']$ hits a trapezium corner. A trapezium corner can be hit in two ways: either by the interior of a segment, or by one of its endpoints. In the former case, if c_0 is such a corner, $[c_0, x']$ will define a segment transverse to a whole quadrilateral of N visited by no curve of $\mathcal{C}$, which contradicts Γ being trimmed. So only the latter occurs, and we have a segment of the type $[c_0, x']$ which can be streched from c_0 towards the direction opposite to x' in a quadrilateral Q. Again because there is no trimming left, this streching will hit the lowest leaf γ'' in Q at a point x'' (the lowest leaf exists because as a set of points $\mathcal{C} \cap Q$ is compact). Then $[x', c_0] \cup [c_0, x''] = [x', x'']$ is a transverse segment as promised.

Now we translate the segment $[x', x'']$ of the above claim. But note that this time, if a trapezium corner is hit it must be hit by the interior of the segment translate. If no translate hits any trapezium corner towards any of the two directions, then γ' and γ'' are either distinct homotopic curves in $\mathcal{C}$, which is not possible, or a single curve, but then γ must be homotopic to it, which is a contradiction. So let c be the first trapezium corner hit by the open segment (x', x''). This is in fact the only one: towards the other direction the translates of (x', x'') cannot hit again any other corner. As a matter of fact, if it did, say at some corner c_0, this would allow a finite curve missing $\mathcal{C}$ to be carried by Γ – with endpoints c and c_0 –, in contradiction with the fact that cutting has been applied to Γ.

So let c be the unique corner hit. Assuming that c belongs to the translates of the subsegment (x, x''), we translate the other subsegment (x', x). Again we claim that it cannot hit any other corner. Indeed, suppose there is such a hit at c'. The curve γ must travel under c', through a trapezium Q. Here too since $\mathcal{C}$ is compact, there is a

highest curve in each trapezium of N. Let γ_{max} be the one in Q. By construction it must lie under γ.

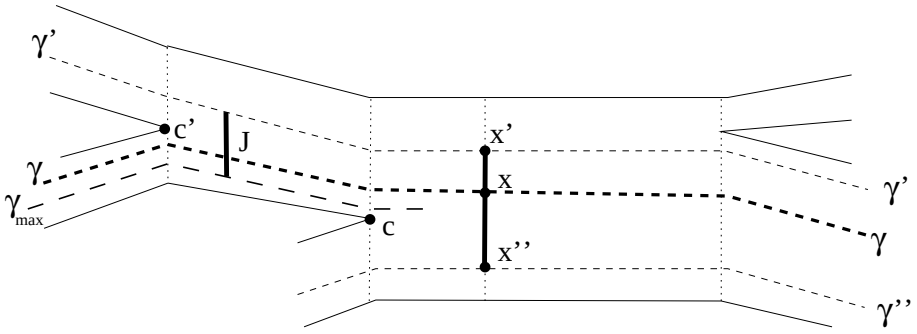

Figure 4: *The attempt of inserting a curve γ (bold dashed style).*

So consider the transverse segment J whose endpoints belong to γ' and γ_{max}, and which hits c'. Its interior must meet γ. But again, since cutting have been applied to Γ, the translates of J cannot hit any other corner than c'. Therefore, it reaches the initial segment (x', x''): this is a contradiction since γ_{max} is then a curve of $\mathcal{C}$ meeting the interior of (x', x''). Hence γ is parallel to γ', contradicting the assumption. So $\mathcal{C}$ is a maximal lamination. $\diamondsuit$

Theorem **2.7** *Let Γ, L and $\mathcal{C}$ be defined as in Lemma 2.1. Let trimming and cutting be applied to Γ rel. to $\mathcal{C}$, so that by Lemma 2.6, $\mathcal{C}$ becomes a maximal lamination. If cutting involves at most mid-vertex cuttings, $Bi(L)$ remains the coding of $\mathcal{C}$. Otherwise, $Bi(L)$ is the image of a bijective projection of the coding of $\mathcal{C}$.*

Proof. Trimming means reducing the alphabet only to the necessary letters: it has no effect on the coding of $\mathcal{C}$. Cutting involving only mid-vertex cuttings generates no new edges, so that the coding is left fixed.

Otherwise by Lemma 2.5, there is a coding $Bi'(L)$ of $\mathcal{C}$ rel. to the transformed graph. The projection from $Bi'(L)$ to $Bi(L)$ is defined by inverting the sequence of cuttings along the cutting curves, and by projecting the labels of the two edges created by each in-vertex cutting towards the label of the initial edge. Injectivity of the projection is ensured since homotopy classes are left stable by cutting or uncutting. Surjectivity is ensured since every curve carried by Γ is still carried after transformation. $\diamondsuit$

3 Automorphisms of Surfaces and D0L-systems

3.1 Basic Definitions

A **homeomorphism** is a bijective bicontinuous map $X \to Y$ between two topological spaces X, Y. When $X = Y$, it is called an **automorphism**. **Dehn twists** are basic automorphisms of an oriented surface M (see [9, 10, 7, 38]) which are built from simple closed curves – called **twist curves** – which are homotopically non-trivial, i.e. not continuously deformable into a single point. A Dehn twist along γ can be described by

three steps: M is cut along γ, one (or more) turn(s) are applied to one of the separated parts, and finally M is pasted back along γ (see Figure 5). More formally, a Dehn twist applies to a parametrized annulus $\{(r, e^{i\alpha}) \mid r \in [r_0, r_1], 0 < r_0 < r_1, \alpha \in [0, 2\pi)\}$ embedded in M and is homotopic to $f_n : (r, e^{i\alpha}) \to (r, e^{2i\pi n \frac{r-r_0}{r_1-r_0}} e^{i\alpha})$ for some $n \in \mathbb{Z}$. According to n's sign, a Dehn twist is said **positive** or **negative**. The annulus is used to describe a neighborhood of a homotopically non trivial simple closed curve γ. This is well-defined since f_n's restriction to the boundary is the identity, and different annuli neighbourhoods yield homotopic maps. Thus Dehn twists can be considered locally on the surface. When performing a Dehn twist on a surface along a curve we shall always assume that this neighbourhood has been fixed.

The effect of a Dehn twist on curves running on M is to drag those which are transverse to the twist curve along itself. An example is shown in Figure 5.

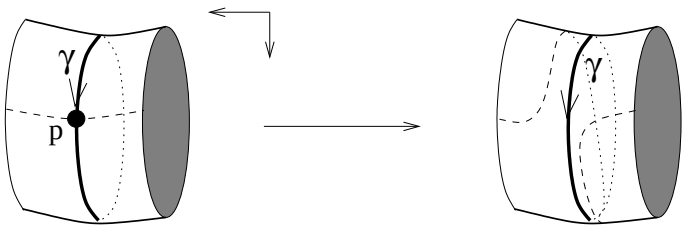

Figure 5: *A cylinder-like piece of a surface M where one can observe the local effect of a twist along γ (bold style) on a transverse curve (dashed style) intersecting γ at p: applying one turn of the twist drags the curve along γ from p.*

Edges of an embedded graph Γ could also be transverse to twist curves, and applying associated Dehn twists induces graph transformations of Γ. This idea can be formalized and generalized as follows: let Γ and Γ' be two directed graphs; then a **directed graph map** sends vertices of Γ to vertices of Γ', and directed edges of Γ to admissible paths of Γ'. An automorphism h of M always induces a graph map up to homotopy from Γ to some Γ'. In case this graph map sends Γ to itself, we say that Γ is **invariant** under h.

A **substitution**[1] θ over an alphabet A is a transformation which sends the letters of A to words over A, and which extends to every one-way and two-way infinite word $...w_i w_{i+1}...$ with $w_i \in A$ by sending it to $...\theta(w_i)\theta(w_{i+1})....$ On finite words, a substitution is simply an endomorphism on the free monoid spanned by A, henceforth denoted by A^*. A **D0L-system** is a 3-tuple (A, θ, w) where θ is a substitution over A and w is a nonempty word of A^* on which θ is iterated (see e.g. [39, 40]).

3.2 Coding Automorphisms into Substitutions

A direct consequence of the last sections is that applying automorphisms to laminations can be studied after their graph representation, i.e. from the induced graph maps. These can be related to substitutions:

[1] Substitutions are here taken in the restricted sense they have in *symbolic dynamics* (see e.g. [37]). In formal language theory, a substitution may be also non-deterministic, and its image words are not necessarily over the origin alphabet (see e.g. [40]).

Lemma 3.1 *Let Γ be a free directed graph labelled over A and embedded in M. Let Γ be invariant under an automorphism h of M. Let $\mathcal{C}$ be the whole set of curves carried by Γ and let cod be the map which assigns to each curve of $\mathcal{C}$ its coding. Then h induces a unique substitution θ_h over A such that $cod(h(\mathcal{C})) = \theta_h(cod(\mathcal{C}))$.*

Proof. Let g_h be the directed graph map induced by h. For each edge η of Γ, we define $\theta_h(cod(\eta))$ as $cod(g_h(\eta))$. This is well-defined because of three properties: the invariance of Γ, the fact that Γ is free, and every edge is assumed to have its own distinct label. Now, consider a finite or infinite admissible path $\gamma = ...\eta_{-1}\eta_0\eta_1...$ in Γ. Then $g_h(\gamma) = ...g_h(\eta_{-1})g_h(\eta_0)g_h(\eta_1)...$, i.e. g_h must preserve concatenation since it is induced by h. On the other hand,

$$\theta_h(cod(\gamma)) = ...\theta_h(cod(\eta_{-1}))\theta_h(cod(\eta_0))\theta_h(cod(\eta_1))...$$

since any simplification would mean that there is a subword coding an admissible path homotopic to the identity, a contradiction with the fact that Γ is free. Thus θ_h is a substitution and the commutation holds. $\diamondsuit$

A generic instance of what happens in Lemma 3.1 is shown in Figure 6.

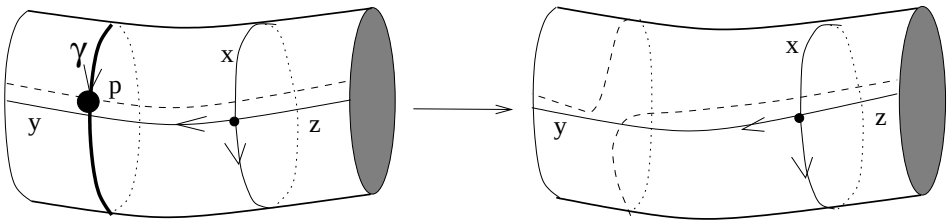

Figure 6: *A cylinder-like piece of a surface M where one can observe the local effect of a twist along γ (bold style) on a transverse curve (dashed style) intersecting γ at p, coded by zy and carried by Γ (normal style): applying one turn of the twist drags the curve along γ from p and results in a curve coded by zxy still carried by Γ. More generally, Γ is invariant by the twist as the image curve of the edge coded by y can be sent to the admissible path coded by xy. This yields a substitution which sends y to xy leaving all the other letters fixed.*

Note that the above lemma is not new in itself. It is already known that automorphisms of surfaces on embedded graph induce automorphisms of free groups (see e.g. [2, 3]). However here we focus on those which can be seen as endomorphisms of free monoids, i.e. substitutions.. Note that substitutions are obtained here since we consider directed graphs: our invariance definition includes admissible paths, preventing any back and forth movements for the carried curves which would induce group relations.

A full example is described in Figure 7 where M is a two-holed torus with an embedded graph Γ and a set of five associated twists. The graph Γ can be checked to be invariant under each of these twists. Substitutions can thus be computed for each of them. For instance, the twist along γ_1 intersects the edge x_1, and applying it drags x_1 along y_1y_2 before it could go along x_1. As a result, the associated substitution

sends x_1 to $y_1y_2x_1$. Hereafter we give the five associated substitutions (θ_γ denotes the substitution associated with the twist along γ, and only letters whose images are not the identity are indicated):

$$\begin{array}{llll} \theta_{\gamma_1}(x_1) & = y_1y_2x_1, & \theta_{\gamma_1}(z_1) & = y_2y_1z_1, \\ \theta_{\gamma_2}(z_2) & = s_2s_1z_2, & \theta_{\gamma_2}(t_1) & = s_1s_2t_1, \\ \theta_{\delta_2}(y_2) & = z_1z_2y_2, & \theta_{\delta_2}(s_2) & = z_2z_1s_2, \\ \theta_{\delta_1}(y_1) & = x_1y_1, & & \\ \theta_{\delta_3}(s_1) & = t_1s_1. & & \end{array} \tag{1}$$

Composing invariant automorphisms leads to invariant automorphisms. For instance,

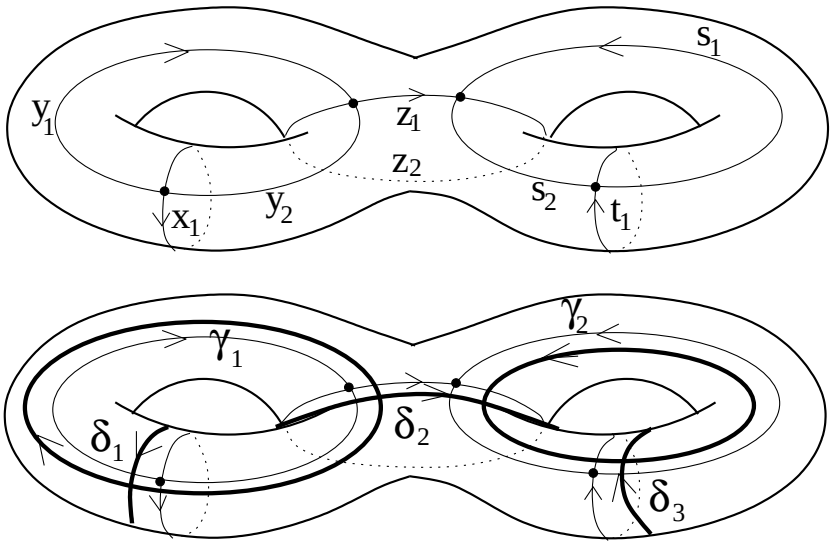

Figure 7: *In the upper part is pictured a two-holed torus with an embedded graph* Γ *with its edges labelled over* $A = \{x_1, y_1, y_2, s_1, s_2, t_1, z_1, z_2\}$. *In the lower part, five twists curves (bold style)* $\{\gamma_1, \gamma_2, \delta_1, \delta_2, \delta_3\}$ *have been added.* Γ *is invariant under each one. Note that* δ_1 *can be readily seen in the same relative position as* γ *in Figures 5 and 6.*

based on the twists of Figure 7, we define:

$$h = h_{\gamma_1}h_{\gamma_2}h_{\delta_1}h_{\delta_2}h_{\delta_3}, \tag{2}$$

where h_γ denotes the twist along γ. The graph Γ is still invariant under h, and the associated substitution θ_h over A is obtained by composing the respective substitutions found in Eq. (1), that is:

$$\begin{array}{l} \theta_h(x_1) = y_1y_2x_1, \\ \theta_h(y_1) = y_1y_2x_1y_1, \\ \theta_h(y_2) = y_2y_1z_1s_2s_1z_2y_2, \\ \theta_h(s_1) = s_1s_2t_1s_1, \\ \theta_h(s_2) = s_2s_1z_2y_2y_1z_1s_2, \\ \theta_h(t_1) = s_1s_2t_1, \\ \theta_h(z_1) = y_2y_1z_1, \\ \theta_h(z_2) = s_2s_1z_2. \end{array} \tag{3}$$

3.3 Stable Sets of Curves and Boundaries of D0L-Languages

The **D0L-language** of a D0L-system (A, θ, w) is defined as $\{v \in A^* \mid v = \theta^n(w), n \in \mathbb{N}\}$. Given a substitution θ, the associated language of θ is the union of the D0L-languages whose starting words are the letters of the alphabet A, i.e. $L_\theta = \{v \in A^* \mid v = \theta^n(s), s \in A, n \in \mathbb{N}\}$.

Lemma **3.2** *Let θ be any substitution over A. Then the set $Bi(L_\theta)$ of the two-way infinite boundary words of L_θ is stable up to shift under θ.*

Proof. Recall that $Bi(L_\theta)$ is defined as a set of representatives of the quotient by the shift of the boundary of the corresponding pointed language $\widehat{L_\theta}$ (cf. Section 2.3). Let w be in $Bi(L_\theta)$ and be a representative of v_w: there is a Cauchy sequence $\{w_n\}_{n\in\mathbb{N}}$ of finite pointed words in $\widehat{L_\theta}$ whose limit is w.

All the shifted versions of a two-way pointed word w are in the boundary of $\widehat{L_\theta}$. So, without loss of generality, we can define the application of θ on a pointed word $\ldots w_{-1}w_0w_1\ldots$ by setting the origin on the first letter of $\theta(w_0)$. So for all n, $\theta(w_n)$ is also in $\widehat{L_\theta}$, and $\{\theta(w_n)\}_{n\in\mathbb{N}}$ is also a Cauchy sequence which converges to a word in $\widehat{L_\theta}$, which is a pointed representative of $\theta(v_w)$. $\diamondsuit$

A substitution θ is said to be **strongly growing** if for every $s \in A$, the length of $\theta^n(s)$ with $n \in \mathbb{N}$ becomes arbitrarily large. This is a sufficient condition so that L_θ is infinite and $Bi(L_\theta)$ is not empty.

Theorem **3.3** *Let Γ be a free coherent directed graph labelled over A and embedded in M, which is invariant under an automorphism h of M. Let θ_h be its associated substitution over A and assume that it is strongly growing. Then the boundary set $Bi(L_{\theta_h})$ is the coding of a maximal lamination invariant under h.*

Proof. Consider a regular neighbourhood N of Γ in M as defined in the proof of Lemma 2.1. Let N' be another closed neighborhood of Γ contained in the interior of N. We assume that N and N' contain respectively two neighbourhoods N_V and N'_V of the set V of $\Gamma's$ vertices, as small as one wants and with the same inclusion property.

Let $\mathcal{C}^{(0)}$ be a set of pairwise disjoint simple finite curves on M with extremities in N'_V, which are contained in the interior of N' and homotopic to exactly one of each edge of Γ up to moving their extremities within N'_V. These are coded by the letters of A, and clearly, they satisfy the first two properties of Lemma 2.1.

Suppose next that we have obtained the set $\mathcal{C}^{(n)}$ corresponding to a set of finite curves on M with the same geometric properties as $\mathcal{C}^{(0)}$ and homotopic to each admissible path corresponding to the language $\{w \in A^* \mid w = \theta^m(s), s \in A, 0 \leq m \leq n\}$, i.e. the words up to the n-th iteration of θ_h on the letters of A.

Since h is a homeomorphism, the set $h(\mathcal{C}^{(n)})$ is also a set of pairwise disjoint simple finite curves. Moreover, by continuity and invariance of Γ, the extremities of $h(\mathcal{C}^{(n)})$ can still be assumed to be in N'_V, and we can push $h(\mathcal{C}^{(n)})$ into the interior of N' preserving the fact that they are still pairwise disjoint after these homotopies.

Now consider a set of curves contained in $N \setminus N'$ and homotopic to $\mathcal{C}^{(0)}$ up to moving their extremities within N_V. By construction, this set has no intersection with $h(\mathcal{C}^{(n)})$. But these can also be pushed inside N' without impairing disjointness. Thus

$\mathcal{C}^{(n+1)} = \mathcal{C}^{(0)} \cup h(\mathcal{C}^{(n)})$ is also a set of curves satisfying the same geometric properties as $\mathcal{C}^{(0)}$, and coded by $\{w \in A^* \mid w = \theta^m(s), s \in A, 0 \leq m \leq n+1\}$.

Using this construction inductively, we can therefore obtain an infinite set of finite curves $\mathcal{C}_{L_\theta}$ satisfying the first two conditions of Lemma 2.1 and coded by the associated language L_θ of θ_h, i.e. $\{w \in A^* \mid \theta^n(s), s \in A, n \in \mathbb{N}\}$. Since θ_h is assumed strongly growing, the third condition of Lemma 2.1 is also satisfied, so that a set of pairwise disjoint, pairwise non-homotopic and two-way infinite curves $\mathcal{C}$ can be obtained.

Now if there is a cutting curve ρ of non-zero length rel. to $\mathcal{C}$, consider $h^n(\rho \cup \mathcal{C}) = h^n(\rho) \cup h^n(\mathcal{C})$. This is a set of curves homotopic to $h^n(\rho) \cup \mathcal{C}$, where $h^n(\rho)$ is disjoint from $\mathcal{C}$ because h^n is a homeomorphism. This implies that for every $n > 0$, $h^n(\rho)$ is also a cutting curve rel. to $\mathcal{C}$. But according to Lemma 2.3, there is at most a finite number of cutting curves which are pairwise disjoint, pairwise non-homotopic. So there exists at least some n_0 and some cutting curve ρ_0 such that up to homotopy, $h^{n_0}(\rho_0) = \rho_0$. Since ρ_0 is carried by Γ, it has a coding word $v = cod(\rho_0)$. According to Lemma 3.1, the coding of $h^{n_0}(\rho_0)$ is the coding of $\theta_h^{n_0}(cod(\rho_0))$ which must be then equal to v. This is impossible since θ_h has been assumed strongly growing.

Thus, according to Lemma 2.6, the set $\mathcal{C}$ is a maximal lamination up to applying trimming and at most mid-vertex cuttings to Γ. So, according to Theorem 2.7, $Bi(L_{\theta_h})$ is the coding of $\mathcal{C}$. $\diamondsuit$

For instance, the automorphism h given in Eq. (2) leaving stable the embedded graph Γ pictured in Figure 7, i.e. $h = h_{\gamma_1} h_{\gamma_2} h_{\delta_1} h_{\delta_2} h_{\delta_3}$, has a substitution θ_h given in Eq. (3) which is clearly strongly growing. Thus the language $Bi(L_{\theta_h})$ describes a maximal lamination on the two-holed torus rel. to Γ.

4 Application: Construction of Non-Periodic Irreducible Automorphisms

4.1 Basic Definitions

An automorphism h of a surface M is called **periodic** when there is some $n > 0$ such that h^n is homotopic to the identity. It is called **reducible** when it is homotopic to an automorphism which fixes a finite set of pairwise disjoint simple closed curves. An automorphism is called **pseudo-Anosov** if it is homotopic to one which fixes a pair of measured transverse laminations without closed leaves, being expansive on the measure of one and shrinking on the measure of the other with mutually inverse coefficients. The relative classification theorem is the following (see [33, 15, 28, 18, 7, 43]):

Theorem (Nielsen-Thurston). *Let M be a closed orientable surface with negative Euler characteristic. Then every automorphism of M is homotopic to one which is of the above three types, i.e. periodic, reducible or pseudo-Anosov.*

In particular this theorem asserts that non-periodic irreducible automorphisms are pseudo-Anosov. A generic way of constructing semigroups of pseudo-Anosov automorphisms up to homotopy comes from Thurston and Penner [43, 35]. It relies on **systems of curves** in a surface M, i.e. unions of pairs of finite sets of pairwise non-homotopic

and disjoint oriented simple closed curves, denoted by C and D, such that the components of $M \setminus C \cup D$ are homeormorphic to disks (i.e. $C \cup D$ **fills** M) with not less than three vertices on their boundaries (i.e. C **hits** D **efficiently**), and such that the orientation given by (c, d) at any point of $c \cap d$, $c \in C$ and $d \in D$, agrees with that of M.

A system of curves $C \cup D$ can be considered as a directed graph Γ embedded in M: its set of vertices V is $C \cap D$ and its set of oriented edges E is the set of segments of $C \cup D$ between consecutive intersection points. This graph is directed and coherent. Accordingly, a system of curves is said to be **free** if its associated graph Γ is. For instance, in the upper part of Figure 7, the considered graph comes from a free system of curves where C consists of the curves coded by y_1y_2 and s_1s_2, and D consists of x_1, t_1 and z_1z_2.

Given a system $C \cup D$, another copy of it can be put in **general position** with respect to M's orientation and the original $C \cup D$, either slightly **over** it or slightly **under** it. This copy is considered as a set of twist curves such that it defines a set of Dehn twists $\mathcal{T}_{C \cup D}$ on M. This is what is illustrated in the lower part of Figure 7 where five Dehn twists along $\{\gamma_1, \gamma_2, \delta_1, \delta_2, \delta_3\}$ are defined after having put $C \cup D$ under Γ.

Now consider the following semigroup of automorphisms of M:

$$H^+(C, D) = \{t_{i_1} \circ ... \circ t_{i_n},\ t_{i_j} \in \mathcal{T}_{C \cup D},\ \text{where each twist of } \mathcal{T}_{C \cup D} \text{ occurs at least once}\}.$$

The sequel of the paper is dedicated to present an alternative proof of the following result (see [35, 14]):

Theorem **4.1** *Let $C \cup D$ be a free system of curves. Then each automorphism in $H^+(C, D)$ is homotopic to one which is non-periodic and irreducible (i.e. pseudo-Anosov).*

4.2 A Proof Based on Formal Language Theory

Lemma **4.2** *Let Γ be an embedded graph defined by a free system of curves $C \cup D$ on M. Then Γ is invariant under every $h \in H^+(C, D)$.*

Proof. Let γ be the twist curve associated with the twist t_γ in $\mathcal{T}_{C \cup D}$. Since $\mathcal{T}_{C \cup D}$ has been defined by a copy of $C \cup D$ in general position, γ hits efficiently a set of edges of Γ. Take one of them, say η. Thus $\eta \cap \gamma$ must be one point p. The effect of performing t_γ on η is to drag it along γ, beginning at p. In case the general position of $\mathcal{T}_{C \cup D}$ rel. to $C \cup D$ is under (respect. over), the resulting curve must be homotopic to the path which first visits the admissible cycle homotopic to γ and next η (respect. first η and next γ). So Γ is invariant under each individual twist in $\mathcal{T}_{C \cup D}$. Since each $h \in H^+(C, D)$ is a composition of these twists, the result follows. $\diamondsuit$

Therefore, according to Lemma 3.1, there is a substitution θ_h associated with each $h \in H^+(C, D)$ over the alphabet A consisting of the labels of Γ.

Lemma **4.3** *Let Γ be as in Lemma 4.2 and let $h \in H^+(C, D)$. Then the associated substitution θ_h is strongly growing.*

Proof. By the above lemma, applying t_γ to η results in a curve carried by an admissible path of length at least 2. Thus the asssociated substitution has an image of length at least 2 on the letter coding η. Now, according to the construction of $\mathcal{T}_{C\cup D}$, each edge of Γ is at least in intersection with one twist of $\mathcal{T}_{C\cup D}$. Since an automorphism h in $H^+(C,D)$ must be such that each twist of $\mathcal{T}_{C\cup D}$ occurs at least once, and since the associated substitution θ_h is the composition of the substitutions associated with the individual twists, each letter of the label alphabet A as an image by θ_h whose length is at least 2. $\diamondsuit$

Proposition **4.4** *Let $C \cup D$ be a free system of curves on M. Every automorphism $h \in H^+(C,D)$ is non-periodic (up to homotopy).*

Proof. If there exists some $n > 0$ such that h^n is homotopic to the identity, its corresponding substitution θ_h must be the identity too. Since θ_h must be strongly growing, this is impossible. $\diamondsuit$

Proposition **4.5** *Let Γ be as in Lemma 4.2 and let $h \in H^+(C,D)$. Let θ_h be its associated substitution. Then the boundary set $Bi(L_{\theta_h})$ is the coding of a lamination $\mathcal{L}_h$ carried by Γ and invariant by h.*

Proof. A graph Γ defined by a free system of curves $C \cup D$ is directed and coherent. By the above lemmas Γ is invariant under h, and θ_h is strongly growing. Thus Theorem 3.3 applies. $\diamondsuit$

Now we study reducibility of the automorphisms in $H^+(C,D)$. We shall use D0L-system theory, and so we need more properties to characterize substitutions. A substitution θ is said to be:
- **primitive** (see e.g. [37]) if there exists a finite **primitivity power** K such that for all $s \in A$, $alph(\theta^K(s)) = A$, where $alph(w)$ is the set of letters that occur in w;
- **strongly closed on** $A' \subseteq A$ if for all $s \in A'$, $alph(\theta(s)) = A'$;
- **cyclic** if there exists a cyclic permutation π of the alphabet A, such that for each $s \in A$, $\theta(s) = s_1 ... s_m$, then $s_{i+1} = \pi(s_i)$, with $1 \leq i \leq m-1$; and for each pair $s, t \in A$ such that $\pi(s) = t$, then $\pi(last(\theta(s))) = first(\theta(t))$, where $first(w)$ extracts the first letter of w, and $last(w)$ its last one;
- **simplifiable** if there exists an alphabet B, such that $|B| < |A|$, where $|.|$ denotes the cardinality, and morphisms $f : A^* \to B^*$ and $g : B^* \to A^*$ such that $\theta = gf$.
- **elementary** if it cannot be simplified;
- **strongly repetitive** if there is a non-empty word u such that for each $k > 1$, $u^k = uuu...u$ is a subword of a word in $L_\theta = \{w \in A^* \mid w = \theta^n(s), s \in A, n \in \mathbb{N}\}$.

We also shall need a known result of lamination theory (see [19], Lemma 2.4 and Proposition 2.5) that we here state relatively to our framework:

Lemma **4.6** *Let h be in $H^+(C,D)$, and $\mathcal{L}_h$ be a fixed lamination as obtained in Proposition 4.5. If $\mathcal{L}_h$ does not contain any closed curve, h is irreducible.*

In the symbolic world, following the definition of the boundary of a language we have that (see [32]):

Lemma **4.7** *There is a periodic word in $Bi(L_{\theta_h})$, i.e. a closed curve in the coded lamination, iff θ_h is strongly repetitive.*

Proof. To a periodic word $w = {}^{\omega}u^{\omega}$ in $Bi(L_{\theta_h})$ corresponds a sequence of finite words in L_{θ_h} which must contain u^k with k larger and larger. Conversely, construct a sequence of pointed words of L_{θ_h} with synchronizing u^k's which grow to the two directions so that the limit is $w = {}^{\omega}u^{\omega}$. $\diamondsuit$

Finally, there is an important result in D0L-system theory coming from Ehrenfeucht and Rozenberg [12] (see also [21]) stating that it is decidable whether a substitution is strongly repetitive. It is based on the following Lemma (see [12], Claim 6.11):

Lemma **4.8** *(Ehrenfeucht-Rozenberg). An elementary strongly growing substitution over A is strongly repetitive iff it is strongly closed and cyclic over a subalphabet $A' \subseteq A$.*

The purpose of the next results is to be able to apply this lemma.

Lemma **4.9** *Let Γ be as in Lemma 4.2 and let $h \in H^+(C, D)$. Then the associated substitution θ_h is primitive.*

Proof. A system of curves on a surface M is such that the components of $M \setminus \Gamma$ are disks, so Γ is certainly connected when seen as a non-directed graph. But since every edge of Γ belongs to a cycle, all the vertices of the boundary of each of these disks must also be connected by a path on the directed graph Γ.

Now, recall that each edge Γ is in intersection with the curve of some twist in $\mathcal{T}_{C\cup D}$. So begin with an edge η, and add to it the edges and vertices corresponding to the twist curve in intersection with it. Do the same thing in the obtained graph with all the added edges one by one, and carry on this process until no more new vertex or edge can be added. If the last graph obtained is a strict subgraph of Γ there is another component in its complement and no two vertices of these two components can be joined by a path, by construction. So the last graph is all of Γ. In view of Lemma 4.2 we can rephrase this by saying that applying to η the twists in the order they appear in the above inductive construction produces a path visiting all the edges of Γ.

By definition of $H^+(C, D)$, every twist of $\mathcal{T}_{C\cup D}$ must appear in the definition of h, so according to the above arguments there is a power $n > 0$ such that $h^n(\eta)$ yields a curve whose admissible path contains every edge of Γ (in fact any $n > |\mathcal{T}_{C\cup D}|$ will do). From the symbolic point-of-view, this means that $\theta_h^n(e)$, e being the label of η, contains every letter of A, which proves the lemma. $\diamondsuit$

Lemma **4.10** *Let Γ be as in Lemma 4.2 and let $h \in H^+(C, D)$. Let $\mathcal{L}_h$ be a lamination as in Proposition 4.5. Then exactly three types of length-2 edge-subpaths of the paths coding $\mathcal{L}_h$ go through each vertex v of Γ.*

Proof. Since θ_h is primitive, the paths coding $\mathcal{L}_h$ visit all of Γ's edges. So, the two only possible length-2 edge-subpaths making a turn at each vertex v of Γ must appear since $\mathcal{L}_h$ is made of two-infinite curves pairwise disjoint. These are indicated in Figure 8 by two pieces of curves γ and γ'. Now, there must be two twists of $\mathcal{T}_{C\cup D}$ near v respectively hitting γ and γ'. By definition of $H^+(C, D)$, these two twists must appear in the decomposition of h. Looking for the last occurence among these two twists determines the twist which drags either γ or γ' along its curve. This also determines a straight length-2 edge-path going through v (also indicated in Figure 8). $\diamondsuit$

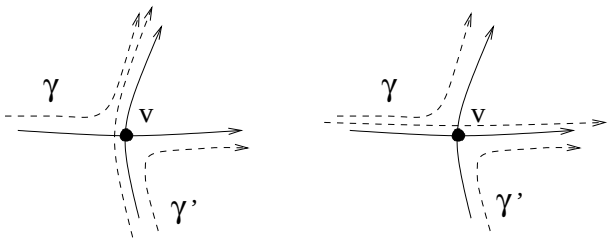

Figure 8: *The two possible 3-tuples length-2 edge-paths going through v.*

Corollary **4.11** *Let Γ be as in Lemma 4.2. No cutting is necessary on Γ rel. to $\mathcal{L}_h$*

Proof. According to the proof of Theorem 3.3, if cutting is necessary, only mid-vertex cuttings can occur. But mid-vertex cuttings depends on the subwords of length-2 centered at each vertex v of Γ: for a graph defined by a system of curves $C \cup D$, at most two subwords among the four possible ones must occur to imply a cutting. According to the above Lemma 4.10, this cannot be the case. $\diamondsuit$

Proposition **4.12** *Let $C \cup D$ be a free system of curves on M. Every automorphism $h \in H^+(C, D)$ is irreducible.*

Proof. To simplify the associated substitution θ_h amounts to being able to factorize the words $\theta_h(A)$ into less subwords than $|A|$ such that each subword is over a subalphabet A' of A and such that the rest can be written over $A \setminus A'$. In terms of edge-paths of the associated graph Γ, this means that some of them can be replaced by single edges. But this implies cutting since all the single edges are originally in Γ followed by a vertex which is a switch. So by Corollary 4.11, θ_h is elementary.

Cyclicity implies that every letter in the image of $\theta_h(s)$, $s \in A$ can be followed by only one distinct letter. But according to Lemma 4.10, this cannot be the case since there are letters (in fact, as many as there are vertices in Γ) which are followed by two different ones.

Finally, according to Lemma 4.9, θ_h is primitive. Hence, there is some finite power $m > 0$ –the primitivity power–, such that θ_h^m is strongly closed on A, and clearly there is no proper subalphabet $A' \subset A$ such that θ_h^m could be strongly closed. Primitivity implies also that $Bi(L_{\theta_h}) = Bi(L_{\theta_h^m})$, i.e. the same lamination is obtained. Moreover, θ_h^m is still elementary and non-cyclic.

So, we can apply the Ehrenfeucht-Rozenberg Lemma 4.8. No two-way infinite periodic word can occur in $Bi(L_{\theta_h})$, i.e. no closed leaf occur in $\mathcal{L}_h$. Lemma 4.6 gives then the result. $\diamondsuit$

Combining the above Proposition 4.12 with Propositions 4.4 and 4.5 gives a proof of Theorem 4.1. In our terms, this theorem becomes :

Theorem **4.13** *For every compact surface M, semi-groups $H^+(C, D)$ of non-periodic irreducible automorphisms can be obtained such that their fixed laminations $\mathcal{L}_h$ are described by $Bi(L_{\theta_h})$, i.e. by boundaries of D0L-languages.*

Thus, for instance, the automorphism h given in Eq. (2) is pseudo-Anosov. Its substitution θ_h given in Eq. (3) leads to a language $Bi(L_{\theta_h})$ which describes a lamination fixed by h.

A consequence of the above proof is that one can either extract codings of individual leaves of $\mathcal{L}_h$ by simply iterating θ_h, or systematically extract all the codings of individual leaves by the procedure described in [32]. Another possibility is given by the following corollary (in [24], we present how to compute the exact number of the involved subwords for every lamination $\mathcal{L}_h$ with $h \in H^+(C,D)$):

Corollary **4.14** *For every $h \in H^+(C,D)$ and every $n > 0$, all the length-n edge-paths homotopic to subcurves of the leaves in $\mathcal{L}_h$ can be described in an effective way.*

Proof. Since the associated substitution θ_h is primitive, each coding word of the individual leaf in $\mathcal{L}_h$ has the same set of subwords. These must also appear recurrently with bounded gaps (see [37]). Moreover those gaps can be effectively bounded [6]. So by finitely iterating θ_h on any letter $s \in A$ gives a word which contains all the possible subwords. $\diamondsuit$

4.3 Remarks about the Proof

A non-periodic and irreducible automorphism setwise fixes two transverse laminations. The lamination $\mathcal{L}_h$ used above is one of them. The other one can also be described through formal language theory as follows. Denote by Γ^{-1} a copy of Γ with its arcs labelled over the same alphabet but with orientations reversed, and in general position rel. to Γ. The graph Γ^{-1} is invariant under h^{-1} for the same reason Γ is under h. Consider the associated lamination $\mathcal{L}_{h^{-1}}$. By construction, it is transverse to $\mathcal{L}_h$. We claim that it is also invariant under h. This can be seen as follows: denote by $\theta_{h^{-1}}$ the substitution associated with h^{-1}; to apply h on $\mathcal{L}_{h^{-1}}$ corresponds on languages to apply the inverse of $\theta_{h^{-1}}$ on $Bi(L_{\theta_{h^{-1}}})$, i.e. to apply the symbolic transformation which consists in slicing the words of $Bi(L_{\theta_{h^{-1}}})$ into the image words $\{\theta_{h^{-1}}(s),\ \ s \in A\}$ of the alphabet A, and then replacing these occurences by the letters whose images they are (see e.g. [37, 31, 32]). Existence and uniqueness of such a slicing for two-way infinite words generated by a non-strongly repetitive and injective substitution on the letters is proved in [31] (Theorem 3.1*bis*). Since elementarity implies injectivity (see [11]) $\theta_{h^{-1}}$ has all the required properties. Finally, by using the same arguments as in Lemma 3.2 $Bi(L_{\theta_{h^{-1}}})$ is stable under this slicing transformation. We therefore have two transverse laminations stable under h.

Measure can be also considered on these two laminations so that h is respectively expanding on one and shrinking on the other one by factors inverse to each other (see [15, 28, 18, 7, 43]). We have been able to avoid these considerations in the whole paper. However, measure is readily at disposal by using the symbolic coding and results in [37]. Indeed, the substitution θ_h has a linear abelianized representation M whose entries $m_{i,j}$ denote the number of letters s_j appearing in $\theta(s_i)$. Primitivity of θ implies that there is some power of M so that all its entries are positive. Perron-Frobenius's Theorem says then that there is a largest real eigenvalue $\lambda > 1$. Using [37], one can conclude that the expanding coefficient of h over $Bi(L_{\theta_h})$ is λ, and the Perron-Frobenius eigenvector gives the relative occurences of the letters of every word in $Bi(L_{\theta_h})$. Similarly to the intersection form on a surface, there is a non-singular bilinear form on the

vector space generated by $C \cup D$'s edges which describes intersections among edges, and it is invariant under the action induced by any homeomorphism $h \in H^+(C, D)$ (via the linear abelianized representation of its associated substitution). This implies that M^{-1} and the transposed of M are conjugate, hence that M and M^{-1} have the same eigenvalues. But the substitution $\theta_{h^{-1}}$ has a matrix conjugate to M^{-1} as linear representation (the conjugacy uses a matrix which inverses some edges orientations), hence the linear representations of θ_h and $\theta_{h^{-1}}$ have the same Perron-Frobenius eigenvalue λ. In particular this implies that λ^{-1} is the shrinking coefficient of the inverse slicing transformation of $\theta_{h^{-1}}$.

Transforming an automorphism of a surface into a substitution has been done here for constructing effective and tractable representations of non-periodic irreducible automorphisms. The originality of the method is to use morphisms on monoids instead of morphisms on groups as in [2, 3] with the benefit that D0L-system theory readily applies. The consequence is the ability of easily describing an infinite set of pseudo-Anosov automorphisms for every closed oriented surface with negative Euler-Poincaré characteristic. However, in view of the classification theorem of Nielsen-Thurston which says that an automorphism is either periodic, irreducible, or pseudo-Anosov, we did not present an answer to the decision problem as in [3, 17]. As a matter of fact it is known that the Thurston-Penner construction, even without the restriction to "free systems of curves", does not yield every possible pseudo-Anosov automorphism of a surface M [35]. Even if the method presented here can be applied to more general invariant orientable embedded graphs with no 0-gons or 1-gons components in their complements (as it is originally done in [35, 14]), this is not sufficient to fill the gap yet. In order to solve the whole classification problem it remains to handle the backtrackings that leaves could make in the graphs when both positive and negative twists occur. As a matter of fact, we could use monoid morphisms because only positive twists were allowed, hence that directed graphs were sufficient. It seems however that backtrackings can be handled, with the stakes that solving the full classification problem of automorphisms in the vein of the present paper would lead to full effective constructions of the objects that characterize each class.

References

[1] P. Arnoux and JC. Yoccoz, *Construction de diffeomorphismes pseudo-Anosov*, C.R. Acad. Sci., Paris, Ser. I **292** (1981), 75–78.

[2] M. Bestvina and M. Handel, *Train tracks and automorphisms of free groups*, Ann. of Math. **135** (1992), 1–51.

[3] ———, *Train-tracks for surface homeomorphisms.*, Topology **34** (1995), no. 1, 109–140.

[4] L. Boasson and N. Nivat, *Adherence of languages*, J. Comp. Syst. Sc. **20** (1980), 285–309.

[5] C. Camacho and A. Lins Neto, *Geometric theory of foliations*, Birkäuser, 1985.

[6] J. Cassaigne, *Recurrence in infinite words*, STACS'2001, 2001, Lecture Notes in Comp. Sci. vol. 2010, pp. 1–11.

[7] A.J. Casson and S. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*, Student Text, no. 9, London Mathematical Society, 1988.

[8] M. Dehn, *Transformation der kurven auf zweiseitigen flächen*, Math. Ann. **72** (1912), 413–421.

[9] ———, *Uber kurvensysteme auf zweiseitigen flächen mit anwendung auf das abbildungsproblem*, 1922, Lecture notes, Math. Kolloq., Breslau.

[10] ———, *Papers on group theory and topology*, Springer Verlag, 1987, (presented by J. Stillwell).

[11] A. Ehrenfeucht and G. Rozenberg, *Simplifications of homomorphisms*, Information and Control **38** (1978), 298–309.

[12] ———, *Repetitions of subwords in D0L languages*, Information and Control **59** (1983), 13–35.

[13] A. Fathi, *Dehn twists and pseudo-Anosov diffeomorphisms*, Invent. Math. **87** (1987), 129–151.

[14] ———, *Démonstration d'un théorème de Penner sur la composition des twists de Dehn*, Bull. Soc. Math. France **120** (1992), 467–484.

[15] A. Fathi, F. Laudenbach, and V. Poenaru (eds.), *Travaux de Thurston sur les surfaces*, Soc. Math. de France, 1979, Astérisque, Volume 66-67.

[16] W.J. Floyd, *Incompressible surfaces in 3-manifolds: the space of boundary curves*, London Math. Soc. Lecture Notes Series (1984), no. 112, 131–144, (D.B.A. Epstein Ed.).

[17] H. Hamidi-Tehrani and Ch. Zong-He, *Surface diffeomorphisms via train-tracks*, Topology and its Appl. **73** (1996), 141–167.

[18] M. Handel and W. P. Thurston, *New proofs of some results of Nielsen*, Adv. in Math. **56** (1985), 173–191.

[19] A.E. Hatcher, *Measured laminations spaces for surfaces from the topological viewpoint*, Topology and its Applications **30** (1988), 63–88.

[20] S.P. Kerckhoff, *Simplicial systems for interval exchange maps and measured foliations*, Ergodic Th. and Dynamical Sys. **5** (1985), 257–271.

[21] Y. Kobayashi and F. Otto, *Repetitiveness of D0L languages is decidable in polynomial time*, Proceedings of MFCS'97, Slovakia, Springer Verlag, 1997, Lecture Notes in Comp. Sci., 1295, pp. 337–346.

[22] L.-M. Lopez and Ph. Narbel, *Generalized Sturmian languages*, Intl. Conf. in Automata Language and Programming'95 (Szeged, Hungary), Lecture Notes in Computer Science 944, Springer-Verlag, 1995, pp. 336–347.

[23] ———, *An algorithm to symbolically describe flows on surfaces*, Tenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, Maryland, USA), 1999, pp.941-942.

[24] ———, *Linear complexity words and surface laminations*, Bull. of Belg. Math. Soc. **8** (2001), 307–323.

[25] ———, *Substitutions and interval exchange transformations of rotation class*, Theoretical Computer Science **1-2** (2001), no. 255, 323–344.

[26] J. E. Los, *Pseudo-Anosov maps and invariant train tracks in the disc: a finite algorithm*, Proc. of London Math. Soc. **66** (1993), no. 2, 400–430.

[27] K. Matsuzaki and M. Taniguchi, *Hyperbolic manifolds and Kleinian groups*, Clarendon Press, Oxford, 1998.

[28] R.T. Miller, *Geodesic laminations from Nielsen's viewpoint*, Adv. in Math. **45** (1982), 189–212.

[29] M. Morse, *A one-to-one representation of geodesics on a surface of negative curvature*, American Journal of Mathematics **43** (1921), 33–51.

[30] ———, *Recurrent geodesics on a surface of negative curvature*, Transactions of Amer. Math. Soc. **22** (1921), 84–100.

[31] B. Mossé, *Puissances de mots et reconnaissabilité des points fixes d'une substitution*, Theoretical Computer Science **99** (1992), 327–334.

[32] Ph. Narbel, *The boundary of iterated morphisms on free semi-groups*, Intern. J. of Algebra and Computation **6** (1996), no. 2, 229–260.

[33] J. Nielsen, *Untersuchungen zur topologie des geschlossenen zweiseitigen flächen*, Acta Math. **50** (1927), 189–358.

[34] A. Papadopoulos and R.C. Penner, *Enumerating pseudo-Anosov foliations*, Pacific Journ. of Math. **142** (1990), no. 1, 159–173.

[35] R.C. Penner, *A construction of pseudo-Anosov homeomorphisms*, Transc. of the Amer. Math. Soc. **310** (1988), no. 1, 179–197.

[36] R.C. Penner and J.L. Harer, *Combinatorics of train tracks*, Princeton University Press, 1993, Annals of Math. Studies, Study 125.

[37] M. Quéffelec, *Substitution dynamical systems. spectral analysis*, Lecture Notes in Mathematics, no. 1294, Springer-Verlag, 1987.

[38] J.G. Ratcliffe, *Foundations of hyperbolic manifolds*, Springer Verlag, 1994.

[39] G. Rozenberg and A. Salomaa, *The mathematical theory of L systems*, Academic press, 1980.

[40] G. Rozenberg and A. Salomaa, *Handbook of formal languages (vol. 1) word languages grammar*, Springer Verlag, Berlin, 1997.

[41] I. Takarajima, *On a construction of pseudo-Anosov diffeomorphism by sequences of train tracks*, Pacific Journ. of Math. **166** (1994), no. 1, 123–191.

[42] I. Tamura, *Topology of foliations: an introduction.*, Translations of Mathematical Monograph, no. 97, American Mathematical Society, 1992.

[43] William P. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Am. Math. Soc., New Ser. **19** (1988), no. 2, 417–431.