



**HAL**  
open science

## On languages of one-dimensional overlapping tiles

David Janin

► **To cite this version:**

| David Janin. On languages of one-dimensional overlapping tiles. 2012. hal-00659202v2

**HAL Id: hal-00659202**

**<https://hal.science/hal-00659202v2>**

Submitted on 14 Jan 2012 (v2), last revised 11 Jul 2012 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



LaBRI, CNRS UMR 5800  
Laboratoire Bordelais de Recherche en Informatique

Rapport de recherche RR-1457-12

On languages of one-dimensional overlapping tiles

January 14, 2012

David Janin,  
LaBRI, IPB, Université de Bordeaux

## Abstract

In the perspective of defining a modeling language for computational music theory [9], this paper deals with languages of one-dimensional discrete overlapping tiles, i.e. triples of words of the form  $(u, v, w)$  where  $v$  is the root of tile and  $u$  and  $w$  define compatibility conditions for a (partial) product.

This leads to define a monoid called the monoid of positive tiles. Completed with negative tiles, it is shown to be isomorphic with the McAlister (inverse) monoid [16, 14].

Several properties of this monoid are studied. We prove in particular that some canonical left and right residual operations are defined on positive tiles and that they encode within positive tiles product with negative tiles in McAlister monoid.

Then we define particular classes of languages of tiles: languages recognizable by finite monoids, rational languages of tiles, and languages definable by monadic second order (MSO) formulae.

In contrast with former studies of word languages recognizable by inverse monoids [15, 21], recognizable languages of tiles are shown to be even weaker : they are at most characterized by finite sets of bi-infinite periodic words. Rational languages of tiles are shown to be MSO-definable.

For the most general class of MSO-definable languages, we provide a simple though indirect Myhill–Nerode like characterization. It induces a simple characterization of MSO definable languages of positive tiles by means of rational languages of words.

This shows that the class of MSO definable languages of tiles is a class of languages as robust and as simple the class of rational languages of words.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Monoid of positive tiles</b>	<b>5</b>
<b>3</b>	<b>Natural order and residuals</b>	<b>8</b>
<b>4</b>	<b>McAlister monoid of arbitrary tiles</b>	<b>11</b>
<b>5</b>	<b>Recognizable languages of tiles</b>	<b>14</b>
<b>6</b>	<b>MSO-definable languages of tiles</b>	<b>19</b>
<b>7</b>	<b>Discussion on open perspectives</b>	<b>22</b>

# 1 Introduction

In this paper, we study languages of one dimensional discrete overlapping tiles : tiles that can be used to cover words with arbitrary overlaps between two consecutive tiles.

There are several ways to define one dimensional overlapping tiles (just called tiles in the sequel) on alphabet  $A$ . In fact, tiles convey such a huge amount of intuition and mathematical richness that they have been defined many times in many fields. For instance, they appear more or less implicitly in the theory of zig-zag codes [3, 22], or analyzing two-way automata [19], or, explicitly, studying the structure of tilings of the  $d$ -dimensional euclidian space  $\mathbb{R}^d$  [10, 11, 8, 1].

In fact, tiles can be combined sequentially, say from left to right, forming bigger and bigger tiles. With caution, we obtain a monoid of tiles with a curious partial associative product. Evenmore, equipping this monoid with reverse tiles : tiles that goes from right to left, we define a monoid known for years as a particularly interesting inverse monoid : the monoid of McAlister [14, 16].

An algebraic definition of this inverse monoid is given by Lawson [14]. More precisely, following Munn's representation theorem of elements of the free inverse monoid [17], tiles can be defined as the birooted word trees that are both unidirectional and linear [18].

Lawson and Munn each provide an explicit definition of this monoid [14, 18]. We give a third one that put emphasis on the quasi-inverse submonoid induced by positive (or left to right) tiles which is the main subject of our study.

Oddly enough, our interest in (re)defining this monoid came from an application perspective in computational music theory [9].

There, every musical pattern is modeled as a triple  $(u, v, w)$  with introduction  $u$ , development  $v$  and conclusion  $w$ . Combining two musical patterns by synchronizing their developments one after the other, one obtains overlaps between developments, introductions and conclusions that are commonly encountered in music.

Of course, these overlaps may need to satisfy some coherence constraints. Abstracting a little, say from melody to harmony, this constraints just become letter by letter equality. And there comes tile product.

Obviously, these overlapping tiles and tile product can be used in many other application context. For instance, in concurrent process behavior modeling, positive tiles product could also be a way to define some notion of directed communication with explicit and unbounded information transfer.

Indeed, if each process behavior is defined as a tile, the sequential product of two tiles can be seen as a directed rendez-vous of processes, and overlaps define the knowledge transferred from the first process to the second.

In other words, languages of tiles can be used in many modeling context. But what are the properties of these languages ? Is there a class of such languages - as rational languages of words - robust enough to be closed under sequential product or iterated product but also simple enough to be decidable ? Is there an automata oriented language theory (or even better an algebraic language theory) associated with this class ? Which specification and synthesis techniques, known for rational languages of words, can be lifted up to languages of tiles ?

Languages of tiles convey such a high modeling potential that it is worth looking for precise answers to these questions. As far as we know, despite McAlister monoid has been known and studied for long, it seems that no systematic study of languages of tiles have been conducted so far.

In this paper, we start such a study. We prove that the class of MSO definable languages of tiles is that expected class of languages : both simple and robust.

## Main results and structure of the paper

We first define positive tiles and the associated product. This leads us to define the monoid of positive tiles which basic properties are studied.

A natural order (analogous to the natural order in inverse monoid [13]) is defined over positive tiles. The notion of left and right residuals induced by that order are considered. Although there can be several residuals, we show that there always exist a canonical least one.

Equipped with residuals, the monoid of positive tiles can be seen as a completion of the free monoid by some *recorder* of word cancelation operators.

Extending the monoid of positive tiles with negative tiles we obtain McAlister's inverse monoid [13]. Canonical left and right residuals and shift operators, defined for positive tiles, are then characterized by means of product and inverse operator in arbitrary tiles.

We then consider recognizable languages, i.e. languages definable as inverse images by morphism of finite sets. The class of such languages turns out to be fairly small. In contrast with languages of words recognizable by finite inverse monoid [15, 21], recognizable languages of tiles are shown to be even weaker : they are at most characterized by finite sets of bi-infinite periodic words. A non trivial example of such a recognizable language of positive tiles is shown to exist.

Last, we consider the class of MSO-definable languages of tiles, i.e. languages definable by formulae of monadic second order logic. It is shown that it contains rational languages of tiles. This result is proved by showing that the class of MSO-definable languages is not only closed under boolean

operators and projection, but also closed under sequential product, iterated product, residuals (in positive tiles) or inverses (in arbitrary tiles).

Despite this robustness, we still obtain a simple Myhill-Nerode like characterization of that class by means of a *word* congruence associated to every language of tiles. This leads to an even simpler characterization of these tile languages.

*Despite the high modeling potential and robustness of these languages of tiles, MSO-definable languages of tiles are shown essentially as simple as rational languages of words.*

We conclude this paper discussing several perspectives open by our study.

## 2 Monoid of positive tiles

We define in this section the monoid of positive tiles and show how it can be extended with left and right shift operators in order to be finitely generated.

### Some notations

Given a finite alphabet  $A$ , let  $A^*$  be the free monoid generated by  $A$  with neutral element denoted by  $1$  and let  $A^\omega$  be the set of infinite words on the alphabet  $A$ .

Let  $\leq_p$  stand for the prefix order relations over words, i.e. for any two words  $x$  and  $y \in A^*$ ,  $x \leq_p y$  when there is a (unique) word denoted by  $x^{-1}(y)$  such that  $xx^{-1}(y) = y$ . Likewise, let  $\leq_s$  stand for the suffix order relations over words, i.e. for any two words  $x$  and  $y \in A^*$ ,  $x \leq_s y$  when there is some word denoted by  $(y)x^{-1}$  such that  $y = (y)x^{-1}x$ .

Extending  $A^*$  with  $0$ , with  $0u = u0 = 0$  for arbitrary  $u \in A^*$ , the resulting monoid can be seen as a complete bi-lattice with  $\vee_p$  and  $\wedge_p$  the joint and meet operators for the prefix order and  $\vee_s$  and  $\wedge_s$  the joint and meet operators for the suffix order.

Given  $\bar{A}$  a disjoint copy of  $A$ , we write also  $u \mapsto \bar{u}$  for the mapping from  $(A + \bar{A})^*$  to itself inductively defined by  $\bar{1} = 1$ , for every letter  $a \in A$ ,  $\bar{a}$  is the copy of  $a$  in  $\bar{A}$ ,  $\bar{\bar{a}} = a$  and, for every word  $u \in (A + \bar{A})^*$ ,  $\overline{\bar{u}} = u$ .

The free group  $FG(A)$  generated by  $A$  is defined to be  $(A + \bar{A})^*$  quotiented by the least congruence over  $(A + \bar{A})^*$  such that, for every letter  $a \in A$ ,  $a\bar{a} = 1$  and  $\bar{a}a = 1$ .

For every  $w \in (A + \bar{A})^*$  there is a unique word  $red(w)$  equivalent to  $w$ , such that there is no occurrence of  $a\bar{a}$  nor  $\bar{a}a$  in  $red(w)$ . In the remainder of the text, we shall identify elements of  $FG(A)$  to their reductions.

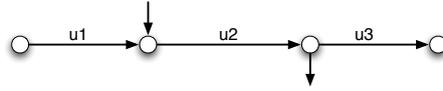
With that assumption, for any two words  $u$  and  $v$  in  $A^*$ , if  $u \leq_p v$  then  $u^{-1}(v) = \bar{u}v$  and if  $u \leq_s v$  then  $(v)u^{-1} = v\bar{u}$ .

## Positive tiles

A positive (or left to right) tile over the alphabet  $A$  is define to be a triple of words  $u = (u_1, u_2, u_3) \in A^* \times A^* \times A^*$ .

In such a tile  $u = (u_1, u_2, u_3)$ ,  $u_1$  is called the *prefix*,  $u_2$  is called the *root* and  $u_3$  is called the *suffix* of tile  $u$ . The word  $u_1u_2u_3$  is called the *support* or the *domain* of tile  $u$ .

Such a tile is conveniently drawn as a (linear, unidirectional and left to right) Munn's birooted word tree [17].

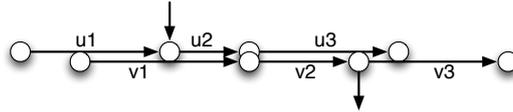


The beginning of the root is marked with a dangling input arrow and the end of the root is marked with a dangling output arrow.

For technical reason, we define an extra tile, written 0 and called the *undefined tile*.

## Tiles product

The sequential product of two non-zero tiles  $u = (u_1, u_2, u_3)$  and  $v = (v_1, v_2, v_3)$  is intuitively defined as the tile (if it exists) resulting of the superposition of the two tiles positioned in such a way that the end of the root of the first tile is *synchronized* with the beginning of the root of the second tile.



A *pattern-matching constraint*, to the left and to the right of the synchronisation point, forces superposed letters to be equal.

The root of the resulting product tile is then defined as the concatenation of the roots of its terms.

Formally, for any two positive tiles  $u = (u_1, u_2, u_3)$  and  $v = (v_1, v_2, v_3)$ , the *sequential product of u and v* is defined by

$$u.v = ((u_1.u_2 \vee_s v_1)u_2^{-1}, u_2v_2, v_2^{-1}(u_3 \vee_p v_2v_3))$$

when the *left matching constraint*  $u_1.u_2 \vee_s v_1 \neq 0$ , i.e.  $u_1u_2$  and  $v_1$  are suffix one of the other, and the *right matching constraint*  $u_3 \vee_p v_2v_3 \neq 0$ , i.e.  $u_3$  and  $v_2v_3$  are prefix one of the other, are both satisfied.

This partial product is completed by  $u.v = 0$  when the matching constraint are not satisfied and by  $u.0 = 0.u = 0$ .

*Example:* Let  $a, b, c$  and  $d \in A$  be four distinct letters. Then  $(a, b, c).(b, c, d) = (a, bc, d)$  while  $(a, b, c).(a, c, d) = 0$ . In the latter case, the left matching constraint is violated because  $a \neq b$ .

One can observe that the domain of a non-zero product contains the domains of its components. In fact, it is defined as the (well positioned and compatible) superposition of their domains.

Observe also that the domain of the product can arise from both tiles involved in the product. There is no limit nor order in the (matching) overlaps that are allowed in sequential product of tiles. For instance,  $(a, b, cc).(b, c, 1) = (a, bc, c)$  with a resulting domain that entirely comes from the first tile.

## The monoid of positive tiles

The set  $T_A$  of positive tiles is defined as the set  $A^* \times A^* \times A^* + 0$ . denoting 1 the tile  $1 = (1, 1, 1)$  we can state that:

**Theorem 1** *The set  $T_A$  of positive tiles equipped with the sequential product of tiles is a monoid with neutral element 1 and absorbant element 0.*

*Proof.* The fact that 1 is neutral is immediate. Associativity of the sequential product comes from the fact that, as already noticed, the domain of non-zero a product always contains the domains of its components.  $\square$

**Lemma 2** *The mapping  $i : A^* \rightarrow T_A$ , defined, by  $i(u) = (1, u, 1)$  for every word  $u \in A^*$  is an embedding of  $A^*$  in  $T_A$ , i.e. it is a one to one monoid morphism.*

*Proof.* Straightforward.  $\square$

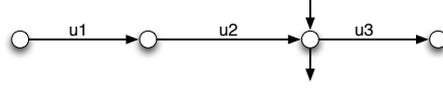
One may ask what could be a set of generators of the monoid of positive tiles on the alphabet  $A$ . Are some irreducible tiles behaving in  $T_A$  much in the same way elements of  $A$  generate the free monoid  $A^*$  ?

It occurs that, in sequential products of tiles, although roots are indeed combined sequentially, both prefixes and suffixes decrease. More precisely, arbitrary roots can be generated, via tiles product, from the (canonical images of) letters of  $A$ , but, the monoid  $T_A$  itself is not finitely generated, i.e. in some sense, monoid  $T_A$  is incomplete. We need some other operators on tiles to generate all tiles from a finite subset of tiles.

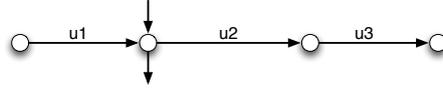
The left and right shifts defined below give a radical solution for generating arbitrary tiles from (canonical images of) words of  $A^*$ . The left and right residuals defined in the next section give an alternative and more subtle way to complete the monoid of positive tiles.

### Left and right shift of a tile

For every tiles  $u \in T_A$  we define the *left shift*  $\sigma_L(u)$  of tile  $u$  to be the tile 0 when  $u = 0$  or the tile  $\sigma_L(u) = (u_1 u_2, 1, u_3)$  when  $u = (u_1, u_2, u_3)$ .



Symmetrically, for every tiles  $u \in T_A$  we define the *right shift*  $\sigma_R(u)$  of tile  $u$  to be the tile 0 when  $u = 0$  or the tile  $\sigma_R(u) = (u_1, 1, u_2 u_3)$  when  $u = (u_1, u_2, u_3)$ .



In the sequel, tiles of the form  $(u, 1, w) \in T_A$  are called *context tiles*.

**Lemma 3** *The set  $C_A$  of context tiles extended with zero is a (sub)monoid of  $T_A$  which elements are commuting idempotents.*

*Proof.* Straightforward. □

**Theorem 4** *The monoid of positive tiles  $T_A$  is finitely generated from (canonical images of) letters of  $A$ , product and left and right shift operators.*

*Moreover, for every word  $u \in A^*$  still denoting  $u$  for the tile  $(1, u, 1)$ , and denoting  $u_L$  for the tiles  $(u, 1, 1) = \sigma_L(u)$  and  $u_R$  for the tile  $(1, 1, u) = \sigma_R(u)$ , one has  $(u, v, w) = u_L v w_R$  for every non zero tiles  $(u, v, w) \in T_A$ .*

*Proof.* As (canonical images of) any words are generated from letters and products this is immediately follows from the second statement that is immediate. □

## 3 Natural order and residuals

Although the monoid of positive tiles is not an inverse monoid [13], it happens that positive tiles can be naturally ordered in a way similar to the way inverse monoids are naturally ordered.

### The natural order

We say that a tile  $(u_1, u_2, u_3)$  is smaller than a tile  $(v_1, v_2, v_3)$ , which we write  $(u_1, u_2, u_3) \leq (v_1, v_2, v_3)$ , when  $v_1 \leq_s u_1$ ,  $v_2 = u_2$  and  $v_3 \leq_p u_3$ . We extend this relation with  $0 \leq u$  for every tile  $u \in T_A$ .

One can easily check that this relation is an order. It is called the *natural order* over tiles. Its properties are described in the following three lemmas.

**Lemma 5 (The submonoid of roots)** *Maximal elements of  $T_A$  under the natural order form a monoid isomorphic with the free monoid  $A^*$ .*

*Proof.* Maximal elements are all elements of the form  $i(u) = (1, u, 1)$  with  $u \in A^*$ .  $\square$

**Lemma 6 (The meet semi-lattice of contexts)** *For any context tiles  $u$  and  $v \in C_A$ :*

- (1)  $u \leq v$  if and only if  $uv = u$ ,
- (2)  $u \wedge v = uv$ .

*Proof.* Elements of  $C_A$  are commuting idempotent elements hence it is a classical result that  $C_A$  is a meet semi-lattice with respect to the order relation defined by  $x \preceq y$  when  $xy = x$ . In that case,  $x \wedge y = xy$ . One can then easily check that this order  $\preceq$  and the natural order  $\leq$  coincide.  $\square$

**Lemma 7 (Natural order and shift operators)** *For every tiles  $u$  and  $v \in T_A$ ,  $u \leq v$  if and only if  $u = v \cdot \sigma_L(u)$  if and only if  $u = \sigma_R(u) \cdot v$ .*

*Proof.* Straightforward.  $\square$

The natural order defined here is actually the restriction to  $T_A$  of the natural order (in the sense of inverse monoid) of the McAlister monoid defined in Section 4.

### Prefix, suffix and residuals

Prefix and suffix preorder relations can be defined as for words. Formally, for every  $u$  and  $v \in T_A$ ,  $u \leq_p v$  (resp.  $u \leq_s v$ ) when there exists some  $w \in T_A$  such that  $u \cdot w = v$  (resp.  $w \cdot u = v$ ).

The prefix and suffix relations over non zero tiles are characterized in  $T_A$  by the following technical lemma.

**Lemma 8** *For every  $u = (u_1, u_2, u_3)$ ,  $v = (v_1, v_2, v_3) \in T_A$  and  $w = (w_1, w_2, w_3) \in T_A$ :*

- (1)  $u \leq_p w$  if and only if  $u_1 \leq_s w_1$ ,  $u_2 \leq_p w_2$  and  $u_2 u_3 \leq_p w_2 w_3$ ,
- (2)  $v \leq_s w$  if and only if  $v_1 v_2 \leq_s w_1 w_2$ ,  $v_2 \leq_s w_2$  and  $v_3 \leq_p w_3$ .

*In particular, both relations  $\leq_p$  and  $\leq_s$  in  $T_A$  are partial order relations in  $T_A$  with 1 as least element and 0 as greatest elements.*

*Proof.* Straightforward from the product definition since, by definition,  $w = uv$  means that

$$(w_1, w_2, w_3) = ((u_1.u_2 \vee_s v_1)u_2^{-1}, u_2v_2, v_2^{-1}(u_3 \vee_p v_2v_3)) \quad \square$$

For every  $u, v$  and  $w \in T_A$ , we say that  $u$  is a *left residual of  $w$  by  $v$*  and that  $v$  is a *right residual of  $w$  by  $u$*  when  $w \leq uv$ .

Over words, with natural order defined as identity, left and right residuals are uniquely defined when they exists and coincide with string cancellation. Over positive tiles, the situation is different as there might be several left or right residuals.

The following lemma gives some necessary and sufficient conditions for  $u$  (resp.  $v$ ) to be a left (resp. right) residual of  $w$  by  $v$  (resp.  $w$  by  $u$ ).

**Lemma 9** *For all positive tiles  $u = (u_1, u_2, u_3)$ ,  $v = (v_1, v_2, v_3)$  and  $w = (w_1, w_2, w_3)$ , the following propositions are equivalent:*

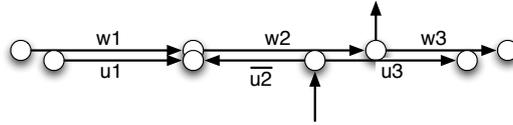
- (1)  $w \leq uv$
- (2)  $v \leq_s w$  with  $u_1 \leq_s w_1$ ,  $u_2 = (w_2)v_2^{-1}$  (henceforth  $u_2 \leq_p w_2$ ) and  $u_3 \leq_p v_2w_3$ ,
- (3)  $u \leq_p w$  with  $v_1 \leq_s w_1u_2$ ,  $v_2 = u_2^{-1}(w_2)$  (henceforth  $v_2 \leq_s w_2$ ) and  $v_3 \leq_p w_3$ .

*In particular, given  $w$  and  $u$  (resp.  $w$  and  $v$ ) there always exists a least  $v$  (resp. a least  $u$ ) such that  $w = uv$ .*

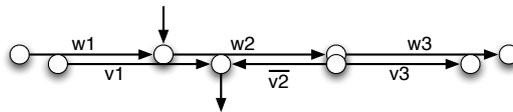
*Proof.* Straightforward. □

### Canonical residuals

For all tiles  $u = (u_1, u_2, u_3)$  and  $w = (w_1, w_2, w_3) \in T_A$  the *canonical left residual*  $u^{-1}(w)$  of  $w$  by  $u$  is defined by  $u^{-1}(w) = (w_1u_2, u_2^{-1}(w_2), w_3)$  when  $u \leq_p w$  and  $u^{-1}(w) = 0$  otherwise. This is illustrated as follows:



Symmetrically, for all tiles  $v = (v_1, v_2, v_3)$  and  $w = (w_1, w_2, w_3) \in T_A$  the *canonical right residual*  $(w)v^{-1}$  of  $w$  by  $v$  is defined by  $(w)v^{-1} = (w_1, (w_2)v_2^{-1}, v_2w_3)$  when  $v \leq_s w$  and  $(w)v^{-1} = 0$  otherwise. This is illustrated as follows:



Left and right residuals are related with left and right shifts as follows.

**Lemma 10** For all  $u \in T_A$ ,  $\sigma_L(u) = u^{-1}(u)$  and  $\sigma_R(u) = (u)u^{-1}$ .

*Proof.* Straightforward from the definitions.  $\square$

**Theorem 11** The monoid  $T_A$  of positive tiles over the alphabet  $A$  is generated by the (images of) letters of  $A$ , product, and (canonical) left and right residuals.

More precisely, denoting  $u$  the canonical image of  $u \in A^*$  in  $T_A$ , for all  $(u, v, w) \in T_A$  we have  $(u, v, w) = (u^{-1}(uvw))w^{-1} = u^{-1}((uvw)w^{-1})$ .

*Proof.* Immediate consequence of the previous definitions.  $\square$

*Remark:* This says that any tile  $(w, v, w)$  can be seen in some sense as a representation of the word residual expression  $(u^{-1}(u)).v.((w)w^{-1})$  defined in  $A^*$  where the canceled strings  $u$  and  $w$  have been recorded into the structure of the corresponding tile in  $T_A$ .

## 4 McAlister monoid of arbitrary tiles

In the definition of the canonical left (resp. right) residual of  $w$  by  $u$  (resp.  $w$  by  $v$ ), everything *looks* as if this residual is computed by means of a generalized sequential product of some  $u^{-1}$  by  $w$  (resp.  $w$  by some  $v^{-1}$ ) where  $u^{-1}$  is obtained from  $u$  (resp.  $v^{-1}$  is obtained from  $v$ ) by flipping the input and output vertices.

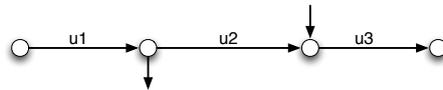
This observation leads to the definition of negative tiles. This *completion* of the monoid of positive tiles by negative tiles define an (inverse) monoid  $M_A$  known as the monoid of McAlister [14].

One shall keep in mind however that residuals and product by inverse, though tightly related, *are* distinct operations as shown by Theorem 13 below.

### Negative tiles

A *negative (or right to left) tile* over the alphabet  $A$  is define to be a triple of word of the form  $v = (u_1u_2, \bar{u}_2, u_2u_3) \in A^* \times \bar{A}^* \times A^*$  with  $u_1, u_2$  and  $u_3 \in A^*$ .

Such a tile is conveniently drawn as a Munn's birooted word tree



which is obtained from the birooted word tree representing the tile  $(u_1, u_2, u_3)$  just by flipping the input and output arrows.

We write  $M_A$  for the set of positive and negative tiles extended with 0. The product of tiles in  $M_A$  is then defined accordingly, the product of two words in  $A^* + \bar{A}^*$  being interpreted in the free group  $FG(A)$  generated by alphabet  $A$ .

## Generalized tiles product

For any two non zero (positive or negative) tiles  $u = (u_1, u_2, u_3)$  and  $v = (v_1, v_2, v_3) \in M_A$ , the sequential product of  $u$  and  $v$  is defined by

$$uv = ((u_1u_2 \vee_s v_1) \cdot \bar{u}_2, u_2v_2, \bar{v}_2(u_3 \vee_s v_2v_3))$$

when both  $u_1u_2 \vee_s v_1 \neq 0$  and  $u_3 \vee_s v_2v_3 \neq 0$ , and  $uv = 0$  otherwise.

**Theorem 12** *The set  $M_A$  equipped with the generalized sequential product is a monoid with neutral element 1 and absorbant element 0. The set  $T_A \subseteq M_A$  of positive tiles is a submonoid of  $M_A$ .*

*Proof.* The generalized product is well defined, i.e. in defined cases we have  $(u_1u_2 \vee_s v_1) \cdot \bar{u}_2 \in A^*$ ,  $\bar{v}_2(u_3 \vee_s v_2v_3) \in A^*$  and, in the case  $u_2v_2 \in \bar{A}^*$  (with elements of the free group  $FG(A)$  always reduced),  $\overline{u_2v_2} \leq_s (u_1u_2 \vee_s v_1) \cdot \bar{u}_2$  and  $\overline{u_2v_2} \leq_p \bar{v}_2(u_3 \vee_s v_2v_3)$ .

Moreover, when both  $u$  and  $v$  are positive tiles, this definition coincides with the sequential product previously defined (Section 2).  $\square$

Observe that all non zero tiles of the form  $(u_1, 1, u_3)$  are both positive and negative tiles.

## Inverse tiles

For every non-zero tile  $u = (u_1, u_2, u_3) \in M_A$ , let  $u^{-1} \in M_A$  be the inverse of tile  $u$  defined by  $u^{-1} = (u_1u_2, \bar{u}_2, u_2u_3)$  and let  $0^{-1} = 0$ .

Left and right shift operators are extended to  $M_A$  just with the same definition, i.e.  $\sigma_L(u_1, u_2, u_3) = (u_1u_2, 1, u_2)$  and  $\sigma_R(u_1, u_2, u_3) = (u_1, 1, u_2u_3)$ .

**Theorem 13** *For every positive tile  $u \in M_A$ ,  $u \cdot u^{-1} = \sigma_L(u)$  and  $u^{-1} \cdot u = \sigma_R(u)$ .*

*Moreover, for all positive tiles  $u$  and  $v \in T_A$ , if  $u^{-1}(v) \neq 0$  (resp.  $(v)u^{-1} \neq 0$ ) in  $T_A$  then  $u^{-1}(v) = u^{-1} \cdot v$  (resp.  $(v)u^{-1} = v \cdot u^{-1}$ ) in  $M_A$ .*

*Proof.* Straightforward.  $\square$

Observe that a left or right residual in monoid  $T_A$  can be zero even though the corresponding left or right multiplication by inverse is positive. In fact, residual  $u^{-1}(v)$  (resp.  $(v)u^{-1}$ ) is non-zero only when  $u \leq_p v$  (resp.  $u \leq_s v$ ).

## McAlister monoid

**Theorem 14** *The monoid  $M_A$  is an inverse monoid isomorphic to the monoid of McAlister.*

*Proof.* (sketch of) We use a representation of McAlister monoid as the quotient of the free inverse monoid  $FIM(A)$  by the ideal  $\perp$  of non unidirectional and non linear tiles where, following Munn's result, elements of  $FIM(A)$  are seen as birooted word trees

The definition of  $FIM(A)$  and the monoid of McAlister  $FIM(A)/\perp$  are first recalled below. Then we explicit an isomorphism from  $M_A$  to  $FIM(A)/\perp$ .

In [17], Munn give the following description of the free inverse monoid on alphabet  $A$ .

A birooted word tree is (conveniently described as) a pair  $(P, u)$  where  $P$  is a non empty finite and prefix-closed subset of (reduced elements of) the free group  $FG(A)$  generated by  $A$ , with  $u \in P$ .

The product of two birooted trees  $(P, u)$  and  $(Q, v)$  is defined by

$$(P, u).(Q, v) = (P \cup uQ, uv)$$

It is a consequence of Munn's representation theorem [17] that the resulting monoid is (isomorphic to)  $FIM(A)$ , the free inverse monoid generated by alphabet  $A$  which can also be defined as the quotient of the free monoid  $(A + \bar{A})^*$  by the least congruence such that, for all words  $u$  and  $v \in (A + \bar{A})^*$ ,  $u\bar{u}u \simeq u$  and  $u\bar{u}v\bar{v} \simeq v\bar{v}u\bar{u}$ .

The monoid of McAlister can in turn be defined as follows. A birooted word tree  $(P, u)$  is said *unidirectional* when  $P \subseteq A^* + \bar{A}^*$ , and *linear* when both  $P \cap A^*$  and  $P \cap \bar{A}^*$  are totally ordered by the prefix order.

It is straightforward that the set  $\perp$  of non unidirectional or non linear birooted trees is an ideal. Following Lawson or Munn, we define thus the monoid of McAlister as the Rees quotient  $FIM(A)/\perp$ .

In that monoid, given two linear and unidirectional birooted word trees  $(P, u)$  and  $(Q, v)$ , the product of these two tiles is defined to be  $(P \cup uQ, uv)$  as in  $FIM(A)$  when the resulting birooted tree is linear and unidirectional, and  $\perp$  (from now on written 0) otherwise.

We prove now that our monoid  $M_A$  is just the McAlister monoid.

For every tile  $u = (u_1, u_2, u_3)$  let

$$P_u = \{x \in \bar{A}^* : x \leq_p \bar{u}_1\} \cup \{x \in A^* : x \leq_p u_2u_3\}$$

and let  $t_u = (P_u, u_2)$  be the resulting birooted tree.

Observe that  $t_u$  is indeed a well defined unidirectional and linear birooted tree. In fact, when  $u$  is a positive tile, we have  $u_2 \leq u_2u_3$  hence  $u_2 \in P_u$ .

When  $u$  is a negative tile, i.e. with  $u_2 \in \bar{A}^*$  we have both  $u_1u_2$  and  $u_2u_3 \in A^*$  and thus  $u_2 \leq_p \bar{u}_1$  hence  $u_2 \in P_u$ .

We claim that the mapping  $\varphi : M_A \rightarrow FIM(A)/\perp$  defined by  $\varphi(0) = 0$  and for any non-zero tile  $u = (u_1, u_2, u_3) \in M_A$ ,  $\varphi(u) = (P_u, u_2)$  is an isomorphism.

First, it is easy to check that it is a bijection. In fact, given a linear and unidirectional tile  $(P, u)$  one can define  $\varphi^{-1}((P, u))$  to be the tile  $(u_1, u_2, u_3)$  such that  $u_2 = u$ ,  $u_1 = \bigvee_s \overline{P \cap \bar{A}^*}$  and  $u_3 = \bar{u}_2 \bigvee_p P \cap A^*$ .

It remains to show that it is indeed a monoid morphism. For this, it is enough to check that for any two non zero tiles  $u = (u_1, u_2, u_3)$  and  $v = (v_1, v_2, v_3) \in M_A$ , one indeed has  $t_{u.v} = (P_u \cup u_2P_v, u_2v_2)$  which is straightforward.

As a consequence, as McAlister is an inverse monoid, so is the monoid  $M_A$ .  $\square$

In particular, the natural order defined in  $T_A$  extends to  $M_A$  and it coincides with the natural order defined in any inverse monoid. We refer the interested reader to the book of Lawson on inverse monoids [13], chapter 9.4, for a description of the numerous additional properties of McAlister monoid.

## 5 Recognizable languages of tiles

In this section, we consider languages of arbitrary or positives tiles that are recognizable in the algebraic sense.

Since the monoid of positive tiles is not an inverse monoid or since we consider recognizable subsets of McAlister inverse monoid, the result we obtain rather differs from the studies of languages of words recognized by finite inverse monoids (see e.g. [15, 21]).

### Preliminaries on recognizable languages

Following the standard definition we say that a language  $L \subseteq T_A$  (resp.  $L \subseteq M_A$ ) is a recognizable language of positive tiles (resp. arbitrary tiles) when there is a (surjective) monoid morphism  $\varphi : T_A \rightarrow S$  (resp.  $\varphi : M_A \rightarrow S$ ), with  $S$  a finite monoid such that  $L = \varphi^{-1}(\varphi(L))$ .

Given such  $L$ , we define the syntactic congruence  $\simeq_L$  over  $L$  by : for all  $u$  and  $v \in T_A$  (resp.  $u$  and  $v \in M_A$ ),  $u \simeq_L v$  when for all  $x$  and  $y \in T_A$ ,  $xuy \in L \Leftrightarrow xvy \in L$ . As well known, such a language is recognizable if and only if its syntactic congruence is of finite index.

The following Lemma explain why we need to restrict slightly the definition of recognizability in  $T_A$ . Otherwise there might be some far too complex languages (even non computable ones) that would be considered recognizable.

**Lemma 15** *Let  $w \in A^*$  be an arbitrary infinite word on the alphabet  $A$ . The positive tiles language  $L_w = \{(1, 1, v) \in T_A : 1 <_p v \leq_p w\}$  is recognizable.*

*In particular, if  $w$  is not computable, so is  $L_w$  while it remains recognizable.*

*Proof.* Let  $S = \{0, 1, s\}$  be the monoid defined by usual product rules for 0 and 1 and  $ss = s$ , and let  $\varphi_w : T_A \rightarrow S$  be the mapping defined for every  $q \in T_A$  by  $\varphi_w(q) = 1$  when  $q = 1$ ,  $\varphi_w(q) = s$  when  $q \in L_w$  and  $\varphi_w(q) = 0$  in the remaining cases.

We claim that  $\varphi_w$  is a morphism monoid. In fact, for every  $q_1$  and  $q_2 \in L_w$  one indeed has  $q_1q_2 = q_1$  or  $q_1q_2 = q_2$  hence  $\varphi(q_1q_2) = \varphi(q_1)\varphi(q_2)$ . It follows that, indeed,  $L_w = \varphi_w^{-1}(\varphi_w(s))$ .  $\square$

This undesirable property results from the fact that  $T_A$  is not finitely generated. It follows that there might some element  $u \in T_A$  such that  $\varphi(\sigma_R(u)) \neq 0$  (or  $\varphi(\sigma_L(u)) \neq 0$ ) while  $\varphi(u) = 0$ .

Observe that in McAlister monoid, finitely generated, this cannot happen since for every  $u \in M_A$ ,  $\sigma_L(u) = u^{-1}u$  and  $\sigma_R(u) = uu^{-1}$ .

In the sequel, whenever considering recognizable languages of positive tiles, we shall restrict ourselves to *safely recognizable* languages in the sense that the morphism  $\varphi : T_A \rightarrow S$  is such that for every  $u \in T_A$ , if  $\varphi(\sigma_L(u)) \neq 0$  or  $\varphi(\sigma_R(u)) \neq 0$  then  $\varphi(u) \neq 0$ .

The next result, negative, tells us that rather simple languages of tiles are not recognizable.

**Lemma 16** *Assume  $A$  has at least two distinct letters  $a$  and  $b$ . Language  $L = \{(ba^m, a^n, 1) \in T_A : m, n \in \mathbb{N}\}$  i.e.  $L = (ba^*)^{-1}(ba^*) - 0$  within positive tiles, is not recognizable.*

*Proof.* We prove that the syntactical congruence  $\simeq_L$  associated to  $L$  in  $T_A$  (or in  $M_A$ ) is of infinite index.

Consider for every  $m \in \mathbb{N}$ , the tile  $u_m = (ba^m, 1, 1)$ . It is an easy observation that for any two  $m$  and  $n \in \mathbb{N}$ ,  $u_m \simeq_L u_n$  if and only if  $m = n$  hence the claim.

Indeed, for every  $k \in \mathbb{N}$  let  $v_k = (a^k, 1, 1)$ . We have for any given  $m \in \mathbb{N}$ ,  $u_m v_k \in L$  if and only if  $k \leq m$ . Now if for some  $m$  and  $n \in \mathbb{N}$  one has  $u_m \simeq_L u_n$  then for every  $k \in \mathbb{N}$ ,  $u_m v_k \in L$  if and only if  $u_n v_k \in L$ . It follows that, for every  $k$ ,  $k \leq m$  if and only if  $k \leq n$ , hence  $m = n$ .  $\square$

We are now ready to study (safely) recognizable subsets of the monoid of positive or arbitrary tiles on a given alphabet  $A$ . We assume that  $A$  has two distinct letters  $a$  and  $b$  at least.

From now on, let  $\varphi : T_A \rightarrow S$  be a safe morphism (or  $\varphi : M_A \rightarrow S$ ). For every word  $u \in A^*$ , let  $u_L = (u, 1, 1)$ ,  $u_C = (1, u, 1)$  and  $u_R = (1, 1, u)$

and let  $\varphi_L$ ,  $\varphi_C$  and  $\varphi_R$  be the induced morphisms from  $A^*$  to  $S$  respectively defined by  $\varphi_L(u) = \varphi(u_L)$ ,  $\varphi_C(u) = \varphi(u_C)$  and  $\varphi_R(u) = \varphi((1, 1, u))$ . In McAlister monoid  $M_A$ , let also  $u_I = (u, \bar{u}, u)$  and  $\varphi_I(u) = \varphi(u_I)$ .

For all non-zero positive tiles  $(u_1, u_2, u_3) \in T_A$  one has  $\varphi((u_1, u_2, u_3)) = \varphi_L(u_1)\varphi_C(u_2)\varphi_R(u_3)$ . It follows that morphism  $\varphi$  on  $T_A$  is completely determined by  $\varphi_L$ ,  $\varphi_C$  and  $\varphi_R$ . Similarly, in  $M_A$ ,  $\varphi$  is completely determined by  $\varphi_L$ ,  $\varphi_C$ ,  $\varphi_R$  and  $\varphi_I$  since for all negative tiles  $(u_1u_2, \bar{u}_2, u_2u_3) \in M_A$  one has  $\varphi((u_1u_2, \bar{u}_2, u_2u_3)) = \varphi_L(u_1)\varphi_I(u_2)\varphi_R(u_3)$ .

## On recognizable tiles' root languages

**Theorem 17** *For all non-zero  $s \in S$ , language  $L_s = \varphi_C^{-1}(s)$  is totally ordered by both prefix and suffix order.*

*Moreover,  $L_s$  is either empty, or a singleton or else  $L_s = u(vu)^m(vu)^*$  for some integer  $m$  and some  $u \in A^*$  and  $v \in A^+$ .*

*Proof.* Let  $s \in S$  be a non zero element of  $S$ .

Let us first prove that  $L_s$  is totally ordered by prefix and suffix order. Let  $u$  and  $v \in L_s$ . We need to prove that both  $u \vee_s v$  and  $u \vee_p v$  are non zero. By symmetry, it suffices to prove that  $u \vee_s v \neq 0$ .

Assume  $u \vee_s v = 0$ . This implies that  $u_C u_L = u_C$  while  $v_C u_L = 0$ . Now, since  $\varphi(u_C) = \varphi(v_C)$  this means

$$\varphi(u_C u_L) = \varphi(u_C)\varphi(u_L) = \varphi(v_C)\varphi(u_L) = \varphi(v_C u_L)$$

henceforth  $s = 0$  : contradiction.

Assume  $L_s$  is non empty with two distinct words  $u$  and  $v$  such that  $|u| < |v|$ . Without loss of generality, assume  $u$  and  $v$  are the shortest distinct words of  $L_s$ . Since  $L_s$  is totally ordered by prefix and suffix we know these two words are uniquely determined. Moreover, there are two non empty words  $x$  and  $y \in A^*$  such that  $v = ux = yu$  hence  $\varphi(u) = \varphi(ux) = \varphi(u)\varphi(x) = \varphi(u)\varphi(x)^k$  for every integer  $k$ , hence  $ux^* = y^*u \subseteq L_s$ .

We now claim that  $L_s = ux^* = y^*u$ . Indeed, let  $w \in L_s$ . Since  $L_s$  is totally ordered by prefix there is a unique  $k$  such that  $w = ux^k w'$  with  $1 \leq w' <_p x$ . But this means that  $\varphi_C(u)\varphi_C(x^k)\varphi_C(w') = \varphi_C(u)\varphi_C(w') = \varphi_C(u)$  hence  $uw' \in L_s$  with  $u \leq_p uw' <_p ux$ . Since we have chosen  $u$  and  $v = ux$  minimal in length, this means  $w' = 1$ .

In other words,  $L_s = ux^* = y^*u$ .

Let us prove now that  $L_s$  is of the form stated above.

If  $u \leq_s x$  (equivalently  $u \leq_p y$ ) then  $v = uwu$  for some  $w \in A^*$  and  $L_s = u(wu)^*$  and we are done renaming  $w$  in  $v$ .

Otherwise, let  $m$  be the greatest integer such that  $x^m \leq_s u$  (equivalently  $y^m \leq_p u$ ). We have  $u = y^m u_y = u_x x^m$  for some  $u_x$  and  $u_y \in A^*$  with  $|u_y| = |u_x| < |x| = |y|$  and, for every  $k$ , we also have  $u_x x^m x^k = y^k y^m u_y$ .

This means that  $x = u'_y u_y$  and  $y = u_x u'_x$  for some  $u'_y$  and  $u'_x \in A^*$  and, thus for every  $k$ ,  $u_x (u'_y u_y)^{m+k} = (u_x u'_x)^{m+k} u_y$  hence  $u'_y = u'_x$  (from now on written  $w$ ) and  $u_x = u_y$  (from now on written  $z$ ) hence  $x = wz$  and  $y = zw$  and  $u = z(wz)^m$  and  $L_s = z(wz)^m (wz)^*$  which terminates the proof (renaming  $z$  in  $u$  and  $w$  in  $v$ ).  $\square$

*Remark:* By symmetry, it is clear that in  $M_A$  a similar statement holds for the inverse image by  $\varphi_I$  of any non zero element.

Observe also that in the above proof, we never assume that  $S$  is finite !

### On recognizable tiles' context languages

**Theorem 18** *For every non zero  $s \in S$ , language  $L_s = \varphi_L^{-1}(s)$  (resp.  $L_s = \varphi_R^{-1}(s)$ ) is either empty or  $s$  is idempotent and  $L_s$  is totally ordered by the suffix order (resp. prefix order).*

*Moreover, either  $L_s$  is finite or there exists two words  $u$  and  $v \in A^*$  with  $v \neq 1$  such that  $L_s$  is a co-finite subset of the set of all suffixes (resp. prefixes) of words of the form  $u(vu)^*$ .*

*Proof.* Observe first that the set  $\varphi_L(A^*) + 0$  (resp.  $\varphi_R(A^*) + 0$ ) is a submonoid of  $S$  of commuting idempotents elements partially ordered by the semi-lattice order  $\preceq$  defined for every  $x$  and  $y \in S_L$  (resp.  $S_R$ ) by  $x \preceq y$  when  $xy = x$ . This is an immediate consequence of the fact that both left (resp. right) context tiles in  $T_A$  form a submonoid with the above properties (see Lemma 3).

By symmetry, we only consider the case of  $\varphi_R(s)$ . Let then  $s \in S$  be a non zero element such that  $L_s = \varphi_R^{-1}(s)$  is non empty.

We claim that  $L_s$  is totally ordered by prefix. Indeed, let  $u$  and  $v \in L_s$ . By morphism definition,  $\varphi(u_R v_R) = \varphi_R(u) \varphi_R(v)$ , hence, by idempotence of  $s$ ,  $\varphi(u_R v_R) = s \neq 0$ , hence  $u \vee_p v \neq 0$ .

Assuming  $L_s$  is infinite, let  $w \in A^*$  be the unique infinite word such that all words of  $L_s$  are prefixes of  $w$  and let  $L_w$  be the set of all finite prefixes of  $w$ .

Observe that for every  $u$  and  $v \in L_w$ , if  $u \leq_p v$  then  $\varphi_R(u) \succeq \varphi_R(v)$ . This immediately follows from  $u_R v_R = v_R$ .

But this also implies that  $L_w - L_s$  is finite. Indeed, infinitely many prefixes of  $w$  belong to  $L_s$  hence, for every word  $u \in L_w$  greater (by prefix order) than the least element  $v_1$  of  $L_s$  there is a word  $v_2 \in L_s$  such that  $v_1 \leq_p u \leq_p v_2$ , thus  $s = \varphi_R(v_1) \succeq \varphi_R(u) \succeq \varphi_R(v_2) = s$ , thus  $\varphi_R(u) = s$  since  $\succeq$  is an order in  $\varphi_R(A^*) + 0$ .

Observe also that  $\varphi_R(L_w)$  does not contain 0. Indeed, every elements are greater than or equal to  $s$ . By the safety assumption, this implies that the

language of tiles' context  $\varphi_C(L_w)$  does not contain 0. Then, by arguments similar to those of the proof of Theorem 17, there exists two words  $u \in A^*$  and  $v \in A^+$  such that  $w = u(vu)^\omega$ .  $\square$

## A non trivial recognizable language of positive tiles

Our point in this section, is to demonstrate that algebraic recognizability somehow collapses for languages of tiles. This is already fairly well illustrated by the necessary conditions stated in Theorems 17 and 18 above.

Still, does there exist a non trivial recognizable language that satisfies these conditions? We provide here such an example by means of a surjective morphism from  $M_A$  to a non trivial finite (inverse) monoid  $S$ .

Let  $a$  and  $b$  be two distinct letters of  $A$ , and let

$$S = \{0, 1, (a, 1, b), (b, 1, a), (b, a, b), (a, b, a)\} \subseteq T_A$$

with product  $\odot$  define as expected for 0 and 1 and defined according to the following product table:

$\odot$	$(a, 1, b)$	$(b, 1, a)$	$(b, a, b)$	$(a, b, a)$
$(a, 1, b)$	$(a, 1, a)$	0	0	$(a, b, a)$
$(b, 1, a)$	0	$(b, 1, a)$	$(b, a, b)$	0
$(b, a, b)$	$(b, a, b)$	0	0	$(b, 1, a)$
$(a, b, a)$	0	$(a, b, a)$	$(a, 1, b)$	0

**Lemma 19**  $(S, \odot)$  is an inverse monoid.

*Proof.* One can check that product  $\odot$  is associative hence  $S$  is a monoid. It remains to show [13] that the set  $E(S)$  of idempotent of  $S$  is a commutative submonoid and that for every non-idempotent element  $x \in S$  there is  $x^{-1} \in S$  such that  $x \odot x^{-1} \odot x = x$  and  $x^{-1} \odot x \odot x^{-1} = x^{-1}$ .

Since  $E(S) = \{0, 1, (a, 1, b), (b, 1, a)\}$  commutation of idempotent immediately follows from the fact that  $(a, 1, b) \odot (b, 1, a) = (b, 1, a) \odot (a, 1, b) = 0$ .

One can then check that  $(a, b, a) \odot (b, a, b) \odot (a, b, a) = (a, b, a)$  and  $(b, a, b) \odot (a, b, a) \odot (b, a, b) = (b, a, b)$  hence we have  $(a, b, a)^{-1} = (b, a, b)$  and  $(b, a, b)^{-1} = (a, b, a)$ .  $\square$

From now on, this monoid is just denoted by  $S$ .

Let then  $F = (ab)^* + (ba)^* + a(ba)^* + b(ab)^*$ , i.e.  $F$  is the set of factors of words of  $(ab)^*$ , and let  $\varphi : M_A \rightarrow S$  be the mapping defined by  $\varphi(0) = 0$ ,  $\varphi(1) = 1$  and for all  $(u, v, w) \in M_A$  such that  $uvw \neq 1$ :

- (1)  $\varphi(u, v, w) = 0$  when  $uvw \notin F$ ,

- (2)  $\varphi(u, v, w) = (a, 1, b)$  or, resp.,  $(b, 1, a)$  when  $v = 1$  with  $a \leq_s u$  or  $b \leq_p w$  or, resp.,  $b \leq_s u$  or  $a \leq_p w$ , i.e. when  $(u, v, w)$  is idempotent,
- (3)  $\varphi(u, v, w) = (b, 1, a), (a, 1, b), (a, b, a)$  or resp.  $(b, a, b)$  when  $v \in (ab)^+, v \in (ba)^+, v \in b(ab)^*$  or, resp.  $v \in a(ba)^*$ , i.e. when  $(u, v, w)$  is strictly positive,
- (4)  $\varphi(u, v, w) = \varphi(uv, \bar{v}, vw)^{-1}$  when  $v \in \bar{A}^+$ , i.e. when  $(u, v, w)$  is strictly negative.

**Theorem 20** *The mapping  $\varphi : M_A \rightarrow S$  is a (surjective) morphism.*

*Proof.* This follows from the fact that, for all  $u$  and  $v \in M_A$ ,

$$\varphi(u) \odot \varphi(v) = \varphi(uv) = \varphi(\varphi(u)\varphi(v))$$

□

*Remark:* In other words, any  $\varphi^{-1}(X)$  for  $X \subseteq S$  is a recognizable languages of arbitrary tiles. Restricting  $\varphi$  to  $T_A$  gives languages of positive tiles.

This result suggest that their might be some link between our study of recognizable languages of tiles and the study of inverse monoids induced by one-dimensional periodic tilings made in [8, 1].

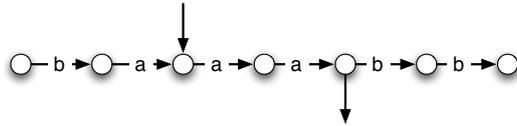
## 6 MSO-definable languages of tiles

We consider in this section the languages of tiles definable by means of monadic second order formulae.

### MSO definability

First, we need a FO-structure description of tiles. For this, we uses a typical encoding of words into FO-structures that amount to encode each letter  $a \in A$  as a relation between elements of the domain. This way, there is no need of end markers and the empty word is simply modeled as the structure with singleton domains and empty relations. We raise models of words to models of tiles just by marking (as for birooted trees) the entry point and exit point of each tile.

For instance, the triple  $(ba, aa, bb)$  is modeled as indicated by the following picture



where, as before, the entry point is marked by a dangling input arrow, and the exit point is marked by a dangling output arrow.

In the sequel, the model of a tile  $t$  is still written  $t$ . The associated domain of its underlying FO-structure is written  $dom(t)$ , the entry point is written  $in(t)$  and the exit point is written  $out(t)$ .

A language  $L \subseteq M_A$  is MSO definable ( $MSO(T_A)$ ) when there is a MSO formula of the form  $\varphi_L(U, x, y)$  where  $U$  is a set variable and  $x$  and  $y$  are two FO-variables such that, for every  $t \in T_A$ ,  $t \in L$  if and only if  $t \models \varphi_L(dom(t), in(t), out(t))$ .

## Closure properties

**Theorem 21** *For any  $L$  and  $M \subseteq T_A$  (resp.  $L$  and  $M \subseteq M_A$ ), if  $L$  and  $M$  are MSO definable then so are  $T_A - L$  (resp.  $M_A - L$ ),  $L \cup M$ ,  $L \cap M$ ,  $L.M$ ,  $L^*$ ,  $L^{-1}(M)$  and  $(M)L^{-1}$  (resp.  $L^{-1}$  in  $M_A$ ).*

*Proof.* Let  $\varphi_L(U, x, y)$  and  $\varphi_M(U, x, y)$  be two formulae defining respectively the language of tiles  $L$  and  $M$ . We assume that these formulae also check that both  $x$  and  $y$  belongs to  $U$  (with  $x \leq y$  when dealing with languages of positive tiles) and that  $U$  is connected.

Case of boolean combination of  $L$  and  $M$  : straightforward.

Case of  $LM$  : take  $\psi(U, x, y)$  stating that there exist two sets  $X$  and  $Y$  such that  $U = X \cup Y$  and there is  $z$  such that both  $\varphi_L(X, x, z)$  and  $\varphi_M(Y, z, y)$  hold.

Case of  $L^*$  : in order to define  $\varphi(U, x, y)$ , the main idea is to consider the reflexive transitive closure  $R^+(x, y)$  of the binary relation  $R(x_1, x_2)$  defined by  $\exists X \varphi_L(X, x_1, x_2)$ ; one must take care, however, that set  $U$  is completely covered by (sub)tiles' domains; this is equivalent to the fact, as domains necessarily overlap, that each extremity (left most or right most element) of the domain  $U$  belongs to one of these sets  $X$  at least. This is easily encoded by a disjunction of the three possible cases : extremities are reached in a single intermediate tile, left extremity is reached first or right extremity is reached first.

Case of  $L^{-1}(M)$  (within positive tiles) : take  $\psi(U, x, y)$  stating that there exist a set  $X \subseteq U$  and there is  $z$  such that  $z \leq x$  and  $z \leq y$  and both  $\varphi_L(X, z, x)$  and  $\varphi_M(U, z, y)$  hold.

Case of  $(M)L^{-1}$  (within positive tiles) : take  $\psi(U, x, y)$  stating that there exist a set  $X \subseteq U$  and there is  $z$  such that  $x \leq z$  and  $y \leq z$  and both  $\varphi_M(U, x, z)$  and  $\varphi_L(X, y, z)$  hold.

Case of  $L^{-1}$  (within McAlister monoid) : take  $\psi(U, x, y)$  stating that  $\varphi_L(U, y, x)$  hold.

□

A language  $L \subseteq T_A$  is rational when it can be defined by means of a finite expression built over letters of  $A$ , sum, product, star iteration of the

product and left and right residuals. Rational languages if  $M_A$  are defined as usual.

**Theorem 22** *Rational languages of positive or arbitrary tiles are MSO definable.*

*Proof.* Immediate consequence of the closure properties above.  $\square$

### A canonical word congruence for languages of tiles

We aim at achieving a simple characterization of MSO definable language of tiles. For this purpose, we first define a notion of congruence relation over  $A^*$  which is defined for every language of tiles. It occurs that this congruence is of finite index if and only if the language of tile is definable in MSO.

Given a language  $L \subseteq T_A - 0$ , we define three relation over words of  $A^*$ , the prefix relation  $\simeq_{PL}$ , the root relation  $\simeq_{RL}$  and the suffix relation  $\simeq_{SL}$  as follows: for every two words  $u$  and  $v \in A^*$  :

- (1)  $u \simeq_{PL} v$  when, for all words  $w_1, w_2, w_3, w_4 \in A^*$ ,  $(w_1uw_2, w_3, w_4) \in L \Leftrightarrow (w_1vw_2, w_3, w_4) \in L$ ,
- (2)  $u \simeq_{RL} v$  when, for all words  $w_1, w_2, w_3, w_4 \in A^*$ ,  $(w_1, w_2uw_3, w_4) \in L \Leftrightarrow (w_1, w_2vw_3, w_4) \in L$ ,
- (3)  $u \simeq_{SL} v$  when, for all words  $w_1, w_2, w_3, w_4 \in A^*$ ,  $(w_1, w_2, w_3uw_4) \in L \Leftrightarrow (w_1, w_2, w_3vw_4) \in L$ ,

Last, we define the relation  $\simeq_L$  over  $A^*$  as follows : for every two words  $u$  and  $v \in A^*$ ,  $u \simeq_L v$  when the three relations  $u \simeq_{PL} v$ ,  $u \simeq_{RL} v$  and  $u \simeq_{SL} v$  hold, i.e.  $\simeq_L$  is the intersection of the prefix, root and suffix relations associated to  $L$ .

### A Myhill-Nerode like theorem

**Theorem 23** *Let  $L \subseteq T_A - 0$ . All relations  $\simeq_{PL}$ ,  $\simeq_{RL}$ ,  $\simeq_{SL}$  and  $\simeq_L$  are congruence relations over  $A^*$ .*

*Moreover,  $L$  is MSO-definable if and only if all these relations are of finite index.*

*Proof.* Let  $L \subseteq T_A - 0$  be a language of tiles. The fact that these relation are congruences is immediate from the definition.

Now,  $L = \Sigma_{(u,v,w) \in L} [u]_L \times [v]_L \times [w]_L$ . Indeed, this means that whenever  $(u, v, w) \in L$  for some  $u, v$  and  $w \in A^*$  then if  $u \simeq u'$ ,  $v \simeq v'$  and  $w \simeq w'$  for some  $u', v'$  and  $w' \in A^*$ , we also have  $(u', v', w') \in L$  whenever . This property just follows from the definition of the word congruence  $\simeq_L$ .

Moreover, if  $\simeq_L$  is of finite index, any language of the form  $[w] \subseteq A^*$  with  $w \in A^*$  is rational henceforth MSO definable, and the above sum is actually finite, henceforth there exists an MSO formula  $\varphi(U, x, y)$  over tiles that define  $L$ .

Conversely, observe that language  $L \subseteq T_A$  can be encoded into a language of words  $M \subseteq A_P^* A_R^* A_S^*$  where  $A_P$ ,  $A_R$  and  $A_S$  are three disjoint copies of the alphabet  $A$  for encoding prefixes, roots and suffixes of tiles.

If  $L$  is MSO definable then, clearly, so is  $M$ . Thus, by Büchi theorem, this means that  $M$  is also rational hence  $M$  is recognizable and thus its syntactic congruence  $\simeq_M$  the word language  $M$  is of finite index hence so are  $\simeq_{PL}$ ,  $\simeq_{RL}$  and  $\simeq_{SM}$ . Indeed, for every word  $u$  and  $v \in A^*$ , we have, for  $X$  denoting  $P$ ,  $R$  or  $S$ ,  $u \simeq_{XL} v$  if and only if  $u_X \simeq_M v_X$  where we write  $w_X$  for the renaming of any word  $w \in A^*$  in the copy alphabet  $A_X$ .  $\square$

## Languages of positive tiles vs languages of words

A reformulation of Theorem 23 shows that despite its high modeling potential, despite its closure under tiles product and residuals, this class is somehow as simple as the class of rational languages of words.

**Corollary 24** *A language of non zero positive tiles  $L \subseteq T_A$  is MSO definable if and only if there are finitely many rational languages  $P_i$ ,  $R_i$  and  $S_i \subseteq A^*$  with  $i \in I$  such that  $L = \Sigma_{i \in I} P_i \times R_i \times S_i$ .*

*Proof.* Immediate consequence of the proof of Theorem 23.  $\square$

## 7 Discussion on open perspectives

### Automata theory for tile languages

Our application perspectives in music call for efficient automata theory adapted to languages of tiles.

There is a (possibly tight) connection between languages of tiles and two-way automata on words. In fact, our last characterization of MSO definable languages of tiles implicitly define such automata. A run over a tile could be, first a back and forth run on its prefix, then a forth run on its root and, last, a run forth and back on its suffix.

From the opposite direction, studying two-way automata, it occurs that Pécuchet in the mid 80's already implicitly defined tiles as partial runs of two-way automata [19]. Following Pécuchet terminology, tiles are actually nothing but pairs of (as called) sections of their domains and his notion of mode 2 automata runs is obviously close to what could be defined as a run over a tile.

However, the languages of tiles are not explicitly defined in [19]. Interested in languages of words, Pécuchet observes the robustness of two-way automata on words described by the equivalence of mode 1 and mode 2 induced notion of recognizability. Our last characterization of MSO definable languages could perhaps explain it ?

### **Algebraic theory for tile languages**

In his algebraic proof of Shepherdson's reduction of two-way automata to one-way automata [20, 23], Birget [6] almost provides an algebraic characterization of two-way automata that has been used till recently [12]. But, as he mentioned, there is still no defined class of algebraic structures that characterizes two-way automata in the same way that the class of monoids characterizes one way automata.

Of course, theorems 17 and 18 tell us that the notion of morphism is far too strong for that. Even for positive tiles, recognizability collapses. An alternative would consist in considering instead (sort of) prehomomorphisms (see [13]) defined as monotonic mappings  $\varphi$  between (naturally) ordered monoid such that neutral and zero elements are mapped to neutral and zero elements and such that  $\varphi(xy) \leq \varphi(x)\varphi(y)$  for all  $x$  and  $y$ .

Over arbitrary tiles or, perhaps more efficiently, over positive tiles only, could the notion of prehomomorphism leads to define an interesting class of (quasi-)algebraic definable languages of tiles ?

### **Rational timed expressions for timed tile languages**

Another potentially interesting development, especially in our perspective of music modeling, is to consider tiles over a timed alphabet : pairs of letters with positive durations. Together with timed constraints - as in timed regular expressions [4] - over tiles domain, one can define a class of rational timed tile languages worth being studied.

In fact, in a product of (timed) tiles, a given timed symbol may occur in many different tiles and thus may be involved in several timed constraints; even in the absence of intersection or projection operators on languages. It follows that rational expression of timed tile languages just extended with time constraints could defined a class of timed tile languages considerably richer than the class of rational timed languages of words which, in absence of projection and intersection, equals the class of one-clock timed automata languages [4].

Will timed tiles overlaps capture, via rational timed expressions of tile languages, the notion of balanced timed expressions [5, 7] ? A positive answer to this question could lead to a fairly simple specification formalism that would capture the class of all recognizable timed languages [2].

## Acknowledgment

This research could not have been achieved without the patient ears of Sylvain Salvati and Marc Zeitoun in Bordeaux whom the author wish to thank.

A special thank is also due to Géraud Sénizergues and Pascal Weil who both suggested to have a look at the theory of inverse monoids in general, and Lawson's book in particular, where the author eventually found a definition of McAlister monoid.

The author feels especially in debt to Anne Dicky for noticing, thanks to her attentive reading, several mistakes in a previous version of this work.

## References

- [1] Filipa Soares de Almeida. *Algebraic Aspects of Tiling Semigroups*. PhD thesis, Universidade de Lisboa, Faculdade de Ciências Departamento de Matemática, 2010.
- [2] Rajeev Alur and David L. Dill. A theory of timed automata. *Theor. Comput. Sci.*, 126(2):183–235, 1994.
- [3] Marcella Anselmo. Automates et code zigzag. *ITA*, 25:49–66, 1991.
- [4] Eugene Asarin, Paul Caspi, and Oded Maler. Timed regular expressions. *J. ACM*, 49(2):172–206, 2002.
- [5] Eugene Asarin and Catalin Dima. Balanced timed regular expressions. *Electr. Notes Theor. Comput. Sci.*, 68(5):16–33, 2002.
- [6] Jean-Camille Birget. Concatenation of inputs in a two-way automation. *Theoretical Computer Science*, 63(2):141 – 156, 1989.
- [7] Catalin Dima. Regular expressions with timed dominoes. In *DMTCS*, volume 2731 of *LNCS*, pages 141–154. DMTCS, 2003.
- [8] E. R. Dombi and N. D. Gilbert. The tiling semigroups of one-dimensional periodic tilings. *Journal of the Australian Mathematical Society*, 87(02):153–160, 2009.
- [9] David Janin. Modélisation compositionnelle des structures rythmiques : une exploration didactique. Technical Report RR-1455-11, LaBRI, Université de Bordeaux, 2011.
- [10] J. Kellendonk. The local structure of tilings and their integer group of coinvariants. *Comm. Math. Phys.*, 187:115–157, 1997.
- [11] Johannes Kellendonk and Mark V. Lawson. Universal groups for point-sets and tilings. *Journal of Algebra*, 276:462–492, 2004.

- [12] Michal Kunc and Alexander Okhotin. Describing periodicity in two-way deterministic finite automata using transformation semigroups. In *Developments in Language Theory*, volume 6795 of *Lecture Notes in Computer Science*, pages 324–336. Springer, 2011.
- [13] Mark V. Lawson. *Inverse Semigroups : The theory of partial symmetries*. World Scientific, 1998.
- [14] Mark V. Lawson. McAlister semigroups. *Journal of Algebra*, 202(1):276–294, 1998.
- [15] Stuart W. Margolis and Jean-Eric Pin. Languages and inverse semigroups. In *ICALP*, volume 172 of *Lecture Notes in Computer Science*, pages 337–346. Springer, 1984.
- [16] D.B. McAlister. Inverse semigroups which are separated over a sub-semigroups. *Trans. Amer. Math. Soc.*, 182:85–117, 1973.
- [17] W. D. Munn. Free inverse semigroups. *Proceedings of the London Mathematical Society*, 29(3):385–404, 1974.
- [18] W. D. Munn. On McAlister’s monoid and its contracted algebra. In *Proceedings of the Conference on Semigroups and Languages*, Lisbon, 2005.
- [19] Jean-Pierre Pécuchet. Automates boustrophedon, semi-groupe de birget et monoïde inversif libre. *ITA*, 19(1):71–100, 1985.
- [20] J. C. Shepherdson. The reduction of two-way automata to one-way automata. *IBM J. Res. Dev.*, 3:198–200, April 1959.
- [21] Pedro V. Silva. On free inverse monoid languages. *ITA*, 30(4):349–378, 1996.
- [22] Do Long Van, Bertrand LeSaëc, and Igor Litovsky. On coding morphisms for zigzag codes. *ITA*, 26:565–, 1992.
- [23] Moshe Y. Vardi. A note on the reduction of two-way automata to one-way automata. *Information Processing Letters*, 30:261–264, 1989.