

CONTEXT-SENSITIVE LANGUAGES, RATIONAL GRAPHS AND DETERMINISM

ARNAUD CARAYOL AND ANTOINE MEYER

IRISA – CAMPUS DE BEAULIEU – 35042 RENNES CEDEX – FRANCE

E-mail address: `Arnaud.Carayol@irisa.fr`

LIAFA – UNIVERSITÉ DE PARIS 7 – 2 PLACE JUSSIEU, CASE 7014, 75251 PARIS CEDEX 05 – FRANCE

E-mail address: `Antoine.Meyer@liafa.jussieu.fr`

ABSTRACT. We investigate families of infinite automata for context-sensitive languages. An infinite automaton is an infinite labeled graph with two sets of initial and final vertices. Its language is the set of all words labelling a path from an initial vertex to a final vertex. In 2001, Morvan and Stirling proved that rational graphs accept the context-sensitive languages between rational sets of initial and final vertices. This result was later extended to sub-families of rational graphs defined by more restricted classes of transducers.

We present a clear map of which characteristics of transducers and of the graphs they define (out-degree, set of initial vertices) are needed to accept the context-sensitive languages. Our main contribution is to provide syntactical and self-contained proofs, whereas the constructions presented up to now rely on a non-trivial normal form of context-sensitive grammars. We use these new constructions both to simplify existing proofs, and to present original results for sub-families of rational graphs and context-sensitive languages.

1. INTRODUCTION

One of the cornerstones of formal language theory is the well-known hierarchy introduced by Chomsky in [Cho59]. It consists of the regular, context-free, context-sensitive and recursively enumerable languages. This hierarchy was originally defined by imposing syntactical restrictions on the rules of grammars generating the languages. These four families of languages as well as some of their sub-families have been extensively studied. In particular, they were given alternative characterizations in terms of finite acceptors. They are respectively accepted by finite automata, pushdown automata, linearly bounded automata and Turing machines. Recently, these families of languages have been characterised by families of infinite automata. An infinite automaton is a labelled countable graph together with a set of initial and a set of final vertices. The language it accepts (or simply its *language*) is the set of all words labelling a path from an initial vertex to a final vertex. In [CK02], a summary of four families of graphs corresponding to the four families in the Chomsky hierarchy was given: they are respectively the finite graphs, *prefix-recognisable*

Key words and phrases. language theory, infinite graphs, automata, determinism.

graphs [Cau96, Cau03a], *rational* graphs [Mor00] and transition graphs of Turing machines [Cau03b] (for a survey, see for instance [Tho01]).

This work specifically deals with a family of infinite automata for context-sensitive languages. The first result on this topic is due to Morvan and Stirling [MS01], who showed that the languages accepted by rational graphs, whose vertices are words and whose edges are defined by rational transducers, taken between rational or finite sets of vertices, are precisely the context-sensitive languages. This result was later extended by Rispal [Ris02] to the more restricted families of *synchronized* rational graphs, and even to *synchronous* graphs. A summary can be found in [MR04]. All proofs provided in these works use context-sensitive grammars in Penttonen normal form [Pen74] to characterize context-sensitive languages, which has two main drawbacks. First, this normal form is far from being obvious, and the proofs and constructions provided in [Pen74] are known to be difficult. Second, and more importantly, there is no grammar-based characterization of deterministic context-sensitive languages, which forbids one to adapt these results to the deterministic case.

Our main contributions are a new syntactical proof of the theorem by Stirling and Morvan based on the tight correspondance between tiling systems and synchronized graphs and an in depth study of the trade off between the structure of the rational graphs (number of initial vertices and out-degree), the transducers defining them, and the family of languages they accept, as summarized in Table 1.

Each row of the table concerns a family or sub-family of rational graphs, and each column corresponds to a structural restriction of that family with respect to sets of initial vertices and degree. The first case is that of rational (infinite) sets of initial vertices, while the second case only considers the fixed rational initial set $\{a\}^*$ over a single letter a . The two remaining cases concern graphs with a unique initial vertex, with respectively arbitrary and finite out-degree. A cell containing an equality symbol indicates that the languages accepted by the considered family of graphs (row) from the considered set of initial vertices (column) are the context-sensitive languages. An inclusion symbol indicates that their languages are strictly included in context-sensitive languages. A question mark denotes a conjecture. When relevant, we give a reference to the proposition, theorem or remark which states each result.

Family of graphs	Set of initial vertices			
	Rational set	Set $\{a\}^*$	Unique vertex	Unique vertex (finite degree)
Rational [MS01]	=	=	=	= [4.6]
Synchronized [Ris02]	=	=	= [4.1]	\subset (?) [4.10]
Synchronous [Ris02]	=	= [3.9]	\subset [4.2]	\subset
Sequential synchronous	= [3.11]	\subset (?) [5.3]	\subset	\subset

Table 1: Families of rational graphs and their languages.

Finally, we investigate the case of deterministic languages. A long-standing open problem in language theory is the equivalence between deterministic and non-deterministic (or even unambiguous) context-sensitive languages [Kur64]. Thanks to our constructions, we

characterize two syntactical sub-families of rational graphs respectively accepting the unambiguous and deterministic context-sensitive languages.

Outline. Our presentation is structured along the following lines. The definitions of rational graphs and context-sensitive languages are given in Section 2. The results concerning languages accepted by rational and synchronous rational graphs are given in Section 3. In Section 4, we investigate rational graphs under structural constraints, and finally Section 5 is devoted to deterministic context-sensitive languages.

2. DEFINITIONS

2.1. Notations

Before all, we fix notations for words, languages and automata, as well as directed graphs and the *languages* they accept. For a more thorough introduction to monoids and rationality, the interested reader is referred to [Ber79, Sak03].

2.1.1. Languages and Automata

We consider finite sets of symbols, or *letters*, called *alphabets*. In the following, Σ and Γ always denote finite alphabets. Tuples of letters are called *words*, and sets of words *languages*. The word u corresponding to the tuple (u_1, \dots, u_n) is written $u_1 \dots u_n$. Its i -th letter is denoted by $u(i) = u_i$. The set of all words over Σ is written Σ^* . The number of letter occurrences of u is its length, written $|u| = n$. The unique word of length 0 is written ε . The concatenation of two words $u = u_1 \dots u_n$ and $v = v_1 \dots v_m$ is the word $uv = u_1 \dots u_n v_1 \dots v_m$. This operation extends to sets of words: for all $A, B \subseteq \Sigma^*$, AB stands for the set $\{uv \mid u \in A \text{ and } v \in B\}$. By a slight abuse of notation, we will usually denote by u both the word u and the singleton $\{u\}$.

A *monoid* is composed of a set M together with an associative internal binary law on M called *product*, with a neutral element in M . The product of two elements x and y of M is written $x \cdot y$. An automaton over M is a tuple $A = (L, Q, q_0, F, \delta)$ where $L \subseteq M$ is a finite set of *labels*, Q a finite set of *control states*, $q_0 \in Q$ is the *initial state*, $F \subseteq Q$ is the set of *final states* and $\delta \subseteq Q \times L \times Q$ is the *transition relation* of A .

A *run* of A is a sequence of transitions $(q_0, l_1, q_1) \dots (q_{n-1}, l_n, q_n)$. It is associated to the element $m = l_1 \cdot \dots \cdot l_n \in M$. If q_n belongs to F , the run is *accepting* (or *successful*), and m is accepted, or recognized, by A . The set of elements accepted by A is written $L(A)$. A is *unambiguous* if there is only one accepting run for each element in $L(A)$.

The star of a set $X \subseteq M$ is defined as $X^* := \bigcup_{k \geq 0} X^k$ with $X^0 = \{\varepsilon\}$ and $X^{k+1} = X \cdot X^k$. Similarly, we write $X^+ := \bigcup_{k \geq 1} X^k$. The set of *rational* subsets of a monoid is the smallest set containing all finite subsets and closed under union, product and star. The set of all words over Σ together with the concatenation operation forms the so-called *free monoid* whose neutral element is the empty word ε . Finite automata over the free monoid Σ^* are known to accept the rational subsets of Σ^* , also called rational languages.

2.1.2. Graphs

A labeled, directed and simple *graph* is a set $G \subseteq V \times \Gamma \times V$ where Γ is a finite set of labels and V a countable set of *vertices*. An element (s, a, t) of G is an *edge* of source s , target t and label a , and is written $s \xrightarrow[a]{a} t$ or simply $s \xrightarrow{a} t$ if G is understood. The set of all sources and targets of a graph form its *support* V_G . A sequence of edges

$s_1 \xrightarrow{a_1} t_1, \dots, s_k \xrightarrow{a_k} t_k$ with $\forall i \in [2, k], s_i = t_{i-1}$ is called a *path*. It is written $s_1 \xrightarrow{u} t_k$, where $u = a_1 \dots a_k$ is the corresponding *path label*. A graph is *deterministic* if it contains no pair of edges having the same source and label. The path language of a graph G between two sets of vertices I and F is the set

$$L(G, I, F) := \{ w \mid s \xrightarrow[G]{w} t, s \in I, t \in F \}.$$

If two infinite automata recognize the same language, we say they are *trace-equivalent*. In this paper, we consider infinite automata: infinite graphs together with sets of initial and final vertices. We will no longer distinguish the notion of graph with initial and final vertices from the notion of automaton. However, as we will see in Section 3, with no restriction on the set of initial vertices and on the structure of the graph this might not provide a reasonable extension of finite automata.

2.2. Word transducers

Automata can be used to accept more than languages. In particular, when the edges of an automaton are labelled with pairs of letters (with an appropriate product operation), its language is a set of pairs of words, which can be seen as a binary relation on words. Such automata are called finite automata with output, or *transducers*, and they recognize rational relations. We will now recall their definition as well as some of their important properties. For a detailed presentation of transducers, see for instance [Ber79, Pri00, Sak03].

Consider the monoid whose elements are the pairs of words (u, v) in Σ^* , and whose composition law is defined by $(u_1, v_1) \cdot (u_2, v_2) = (u_1 u_2, v_1 v_2)$, generally called the product monoid and written $\Sigma^* \times \Sigma^*$. A transducer T over a finite alphabet Σ is a finite automaton over $\Sigma^* \times \Sigma^*$ with labels in $(\Sigma \cup \{\varepsilon\}) \times (\Sigma \cup \{\varepsilon\})$. Finite transducers accept the rational subsets of $\Sigma^* \times \Sigma^*$. We do not distinguish a transducer from the relation it accepts and write $(w, w') \in T$ if (w, w') is accepted by T . The *domain* $\text{Dom}(T)$ (resp. *range* $\text{Ran}(T)$) of a transducer T is the set $\{w \mid (w, w') \in T\}$ (resp. $\{w' \mid (w, w') \in T\}$). We also write $T(L)$ the set of all vertices v such that $(u, v) \in T$ for some $u \in L$. A transducer accepting a function is called *functional*.

In general, there is no bound on the size difference between input and output in a transducer. Interesting subclasses are obtained by enforcing some form of synchronization. For instance, *length-preserving* rational relations are recognized by transducers with labels in $\Sigma \times \Sigma$, called *synchronous* transducers. Such relations only pair words of the same size.

A more relaxed form of synchronization was introduced by Elgot and Mezei [EM65]: a transducer over Σ with initial state q_0 is *left-synchronized* if for every path

$$q_0 \xrightarrow{x_0/y_0} q_1 \dots q_{n-1} \xrightarrow{x_n/y_n} q_n,$$

there exists $k \in [0, n]$ such that for all $i \in [0, k]$, x_i and y_i belong to Σ and either $x_j = \varepsilon$ for all $j > k$ or $y_j = \varepsilon$ for all $j > k$. In other terms, a left-synchronized relation is a finite union of relations of the form $S \cdot F$ where S is a synchronous relation and F is either equal to $\{\varepsilon\} \times R$ or $R \times \{\varepsilon\}$ where R is a rational language. *Right-synchronized* transducers are defined similarly. In the following, unless otherwise stated, we will refer to left-synchronized transducers simply as *synchronized transducers*.

The standard notion of determinism for automata does not have much meaning in the case of transducers because it does not rely only on the input but on both the input and the output. A more refined notion is that of *sequentiality*: a transducer T with states Q is

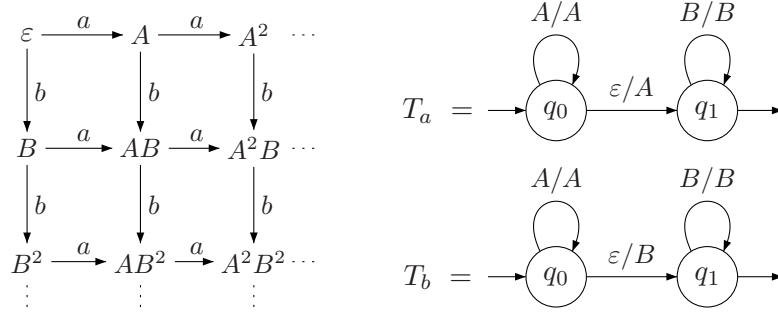


Figure 1: The grid and its associated transducers.

sequential if for all q, q' and q'' in Q , if $q \xrightarrow{x/y} q'$ and $q \xrightarrow{x'/y'} q''$ then either $x = x', y = y'$ and $q' = q''$, or $x \neq \varepsilon, x' \neq \varepsilon$ and $x \neq x'$.

Remark 2.1. The standard determinization procedure applied to a synchronous transducer yields an equivalent unambiguous synchronous transducer (i.e for every pair of words (u, v) accepted by the transducer there is exactly one accepting run of the transducer labelled by u/v). This remains true for synchronized transducers.

It is well-known that there is a close relationship between rational languages and rational transductions. In particular, rational relations have rational domains and ranges, and are closed under restriction to a rational domain or range. Moreover, the restriction of a sequential (resp. synchronous) transducer to a rational domain is still sequential (resp. synchronous) (see for instance [Ber79]).

2.3. Rational graphs

The Chomsky-like hierarchy of graphs presented in [CK02] uses words to represent vertices. Each of these graphs is thus a finite union of binary relations on words, each relation corresponding to a given edge label. In particular, the family of rational graphs owes its name to the fact that their sets of edges are given by rational relations on words, i.e. relations recognized by word transducers.

Definition 2.2 ([Mor00]). A rational graph labelled by Σ with vertices in Γ^* is given by a tuple of transducers $(T_a)_{a \in \Sigma}$ over Γ . For all $a \in \Sigma$, G has an edge labelled by a between vertices u and $v \in \Gamma^*$ if and only if $(u, v) \in T_a$. For $w \in \Sigma^+$ and $a \in \Sigma$, we write $T_{wa} = T_w \circ T_a$, and $u \xrightarrow{w} v$ if and only if $(u, v) \in T_w$.

Note that $T_w \circ T_a$ stands for the set of all pairs (u, v) such that $(u, x) \in T_w$ and $(x, v) \in T_a$ for some x . Figure 1 shows an example of rational graph, the infinite *grid*, with the rational transducers which define its edges. By the properties of rational relations, the support of a rational graph is a rational subset of Γ^* . The rational graphs with synchronized transducers were already defined by Blumensath and Grädel in [BG00] under the name *automatic* graphs and by Rispal in [Ris02] under the name *synchronized* rational graphs.

It follows from the definitions (see Section 2.2) that sequential synchronous, synchronous, synchronized and rational graphs form an increasing hierarchy. This hierarchy is strict (up to isomorphism): first, sequential synchronous graphs are deterministic graphs whereas synchronous graphs can be non-deterministic. Second, synchronous graphs have a finite degree whereas synchronized graphs can have an infinite degree. Finally, to separate

synchronized graphs from rational graphs, we can use the following properties on the growth rate of the out-degree in the case of graphs of finite out-degree.

Proposition 2.3. [Mor01] *For any rational graph G of finite out-degree and any vertex x , there exists $c \in \mathbb{N}$, such that the out-degree of vertices at distance n of x is at most c^{c^n} .*

This upper bound can be reached: consider the unlabeled rational graph $G_0 = \{T\}$ where T is the transducer over $\Gamma = \{A, B\}$ with one state q_0 which is both initial and final and a transition $q_0 \xrightarrow{X/YZ} q_0$ for all X, Y and $Z \in \Gamma$. It has an out-degree of $2^{2^{n+1}}$ at distance n of A . In the case of synchronized graphs of finite out-degree, the bound on the out-degree is simply exponential.

Proposition 2.4. [Ris02] *For any synchronized graph G of finite out-degree and vertex x , there exists $c \in \mathbb{N}$ such that the out-degree of vertices at distance $n > 0$ of x is at most c^n .*

It follows from the above proposition that G_0 is rational but not synchronized. Hence, the synchronized graphs form a strict sub-family of rational graphs.

2.4. Context sensitive languages

In this work, we are concerned with the family of *context-sensitive languages*¹. Several finite formalisms are known to accept this family of languages, the most common being linearly bounded machines (LBM), which are Turing machines working in linear space. Less well-known acceptors for these languages are bounded tiling systems, which are not traditionally studied as language recognizers. However, one can show that these formalisms are equivalent, and that syntactical translations exist between them. Since they are at the heart of our proof techniques, we now give a detailed definition of tiling systems. For more information about linearly bounded machines the reader is referred to [HU79].

Tiling systems were originally defined to recognize or specify *picture languages*, i.e. two-dimensional words on finite alphabets [GR96]. They can be seen as a normalized form of dominos systems [LS97b]. Such sets of pictures are called *local picture languages*. However, by only looking at the words contained in the first row of each picture of a local picture language, one obtains a context-sensitive language [LS97a].

A (n, m) -picture p over an alphabet Γ is a two dimensional array of letters in Γ with n rows and m columns. We denote by $p(i, j)$ the letter occurring in the i th row and j th column starting from the top-left corner, by $\Gamma^{n, m}$ the set of (n, m) -pictures and by Γ^{**} the set of all pictures². Given a (n, m) -picture p over Γ and a letter $\# \notin \Gamma$, we denote by $p_{\#}$ the $(n + 2, m + 2)$ -picture over $\Gamma \cup \{\#\}$ defined by:

- $p_{\#}(i, 1) = p_{\#}(i, m + 2) = \#$ for $i \in [1, n + 2]$,
- $p_{\#}(1, j) = p_{\#}(n + 2, j) = \#$ for $j \in [1, m + 2]$,
- $p_{\#}(i + 1, j + 1) = p(i, j)$ for $i \in [1, n]$ and $j \in [1, m]$.

For any $n, m \geq 2$ and any (n, m) -picture p , $T(p)$ is the set of $(2, 2)$ -pictures appearing in p . A $(2, 2)$ -picture is also called a *tile*. A picture language $K \subseteq \Gamma^{**}$ is *local* if there exists a symbol $\# \notin \Gamma$ and a finite set of tiles Δ such that $K = \{p \in \Gamma^{**} \mid T(p_{\#}) \subseteq \Delta\}$. To any set of pictures over Γ , we can associate a language of words by looking at the frontiers of the pictures. The frontier of a (n, m) -picture p is the word $\text{fr}(p) = p(1, 1) \dots p(1, m)$ corresponding to the first row of the picture.

¹In order to simplify our presentation, we only consider context-sensitive languages that do not contain the empty word ε (this is a standard restriction).

²We do not consider the empty picture.

#	#	#	#	#	#	#	#
#	a	a	a	a	b	b	b
#	a	a	a	a	b	b	b
#	a	a	a	a	⊥	⊥	⊥
#	a	a	a	a	⊥	⊥	⊥
#	⊥	a	a	a	⊥	⊥	⊥
#	⊥	a	⊥	⊥	⊥	b	⊥
#	#	⊥	⊥	⊥	#	⊥	⊥
#	#	⊥	⊥	#	#	⊥	#

Figure 2: A tiling system accepting $a^n b^n$ (Cf. Example 2.7).

Definition 2.5. A tiling system S is a tuple $(\Gamma, \Sigma, \#, \Delta)$ where Γ is a finite alphabet, $\Sigma \subset \Gamma$ is the input alphabet, $\# \notin \Gamma$ is a frame symbol and Δ is a finite set of tiles over $\Gamma \cup \{\#\}$. It recognizes the local picture language $P(S) = \{p \in \Gamma^{**} \mid T(p_{\#}) \subseteq \Delta\}$ and the word language $L(S) = \text{fr}(P(S)) \cap \Sigma^*$.

A tiling system S recognizes a language $L \subseteq \Sigma^+$ in height $f(n)$ for some mapping $f : \mathbb{N} \mapsto \mathbb{N}$ if for all $w \in L(S)$ there exists a (n, m) -picture p in $P(S)$ such that $w = \text{fr}(p)$ and $n \leq f(m)$. We can now precisely state the following well-known equivalence result.

Theorem 2.6. *The following simulations link linearly bounded machines and tiling systems:*

- (1) *A linearly bounded Turing machine T working in $f(n)$ reversals can be simulated by a tiling system of height $f(n) + 2$.*
- (2) *A tiling system of height $f(n)$ can be simulated by a linearly bounded Turing machine working in $f(n)$ reversals.*

Example 2.7. Figure 2 shows the set of tiles Δ of a tiling system S over $\Gamma = \{a, b, \perp\}$, $\Sigma = \{a, b\}$ and the border symbol $\#$. The language $L(S)$ is exactly the set $\{a^n b^n \mid n \geq 1\}$.

A context-sensitive language is called *deterministic* if it can be accepted by a deterministic LBM or tiling system, where a tiling system is deterministic if one can infer from each row in a picture a single possible next row.

3. THE LANGUAGES OF RATIONAL GRAPHS

In this section, we consider the languages accepted by rational graphs and their sub-families from and to a rational set of vertices. We give a simplified presentation of the result by Morvan and Stirling [MS01] stating that the family of rational graphs accepts the context-sensitive languages. This is done in several steps. First, Proposition 3.2 states that the rational graphs are trace-equivalent to the *synchronous* rational graphs. Then, Proposition 3.3 and Proposition 3.7 establish a very tight relationship between synchronous graphs and tiling systems. It follows that the languages of synchronous rational graphs are also the context-sensitive languages (Theorem 3.9). The original result is given as Corollary 3.10. Finally, Proposition 3.11 establishes that even the smallest sub-family we consider,

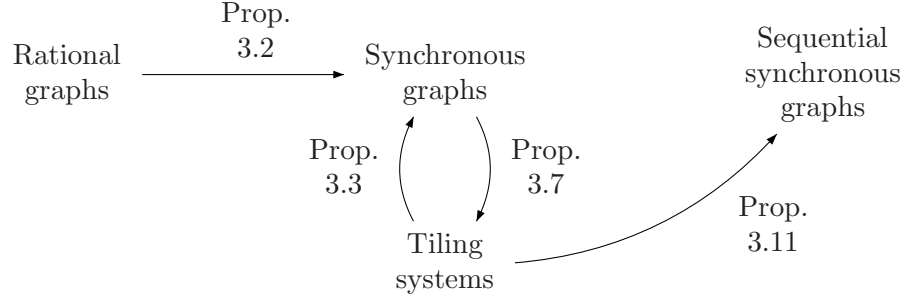


Figure 3: Each edge represents an effective transformation preserving languages.

the family of sequential synchronous rational graphs, accepts all context-sensitive languages. The various transformations presented in this section are summarized in Figure 3.

3.1. From rational graphs to synchronous graphs

We present an effective construction that transforms a rational graph G with two rational sets I and F of initial and final vertices into a synchronous graph G' trace-equivalent between two rational sets I' and F' . The construction is based on replacing the symbol ε in the transitions of the transducers defining G by a fresh symbol $\#$.

Let $(T_a)_{a \in \Sigma}$ be the set of transducers over Γ characterizing G and let $\#$ be a symbol not in Γ . For all a , we define \bar{a} to be equal to a if $a \in \Gamma$, and to ε if $a = \#$. We extend this to a projection from $(\Gamma \cup \#)^*$ to Γ^* in the standard way. We define G' as the rational graph defined by the set of transducers $(T'_a)_{a \in \Sigma}$ where T'_a has the same set of control states Q_a as T_a and a set of transitions given by

$$\{ p \xrightarrow{a/b} q \mid p \xrightarrow{\bar{a}/\bar{b}} q \in T_a \} \cup \{ p \xrightarrow{\#/\#} p \mid p \in Q_a \}.$$

By definition of each T'_a , G' is a synchronous rational graph. Let I' and F' be the two rational sets such that $I' = \{u \mid \bar{u} \in I\}$ and $F' = \{v \mid \bar{v} \in F\}$ (the automaton accepting I' (resp. F') is obtained from the automaton accepting I (resp. F) by adding a loop labeled by $\#$ on each control state). We claim that G' accepts between I' and F' the same language as G between I and F . For example, Figure 4 illustrates the previous construction applied to the graph of Figure 1. Only one connected component of the obtained graph is shown.

Before we prove the correctness of this construction, we need to establish a couple of technical lemmas. Let \mathcal{B} be the set of all mappings from \mathbb{N} to \mathbb{N} . To any mapping $\delta \in \mathcal{B}$, we associate a mapping from $(\Gamma \cup \{\#\})^*$ to $(\Gamma \cup \{\#\})^*$ defined as follows: for all $w = \#^{i_0} a_1 \#^{i_1} \dots a_n \#^{i_n}$ with $a_1, \dots, a_n \in \Gamma$, let $\delta w = \#^{i_0 + \delta(0)} a_1 \#^{i_1 + \delta(1)} \dots a_n \#^{i_n + \delta(n)}$. Before proceeding, we state two properties of these mappings with respect to the sets of transducers (T_a) and (T'_a) .

Lemma 3.1. *We have the following properties:*

$$\forall u, v \in \Gamma^*, (u, v) \in T_a \iff \exists \delta_u, \delta_v \in \mathcal{B}, (\delta_u u, \delta_v v) \in T'_a, \quad (3.1)$$

$$\forall (u, v) \in T'_a, \forall \delta \in \mathcal{B}, \exists \delta' \in \mathcal{B}, (\delta u, \delta' v) \in T'_a, \quad (3.2)$$

$$\text{and dually } \forall (u, v) \in T'_a, \forall \delta \in \mathcal{B}, \exists \delta' \in \mathcal{B}, (\delta' u, \delta v) \in T'_a. \quad (3.3)$$

We can now prove the correctness of the construction: a word w is accepted by G between I and F if and only if it is accepted by G' between I' and F' .

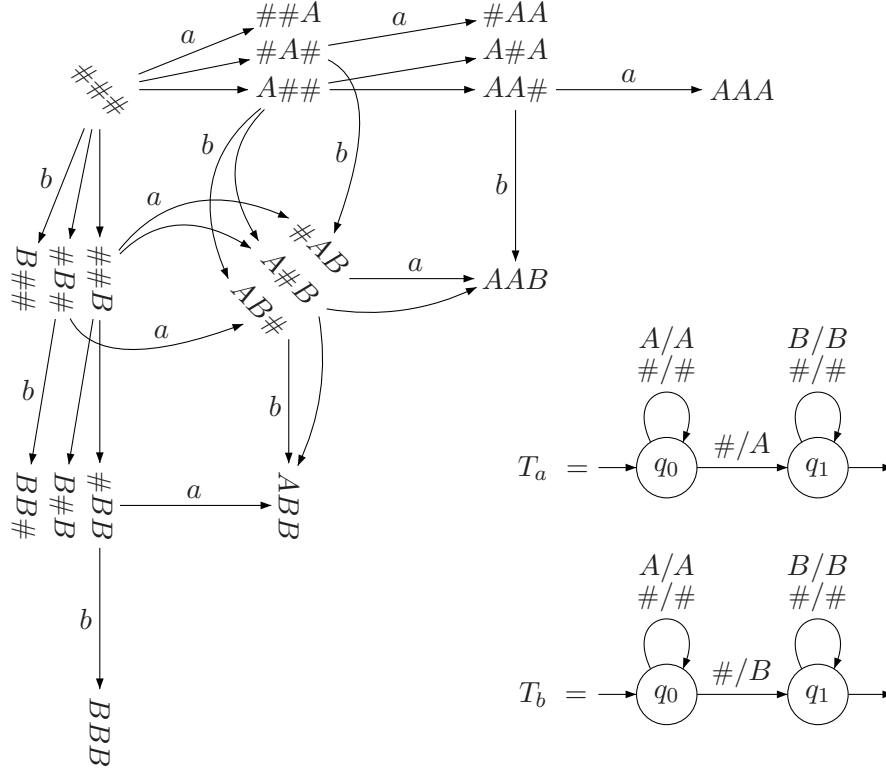


Figure 4: Synchronous graph trace-equivalent to the grid (1 connected component).

Proposition 3.2. *For every rational graph G and rational sets of vertices I and F , there is a synchronous graph G' and two rational sets I' and F' such that $L(G, I, F) = L(G', I', F')$.*

Proof. We show by induction on n that for all $u_0, \dots, u_n \in \Gamma^*$, if there is a path

$$u_0 \xrightarrow[G]{w(1)} u_1 \dots u_{n-1} \xrightarrow[G]{w(n)} u_n,$$

then there exist words $u'_0, \dots, u'_n \in (\Gamma \cup \{\#\})^*$ such that for all i , $\bar{u}'_i = u_i$, and

$$u'_0 \xrightarrow[G']{w(1)} u'_1 \dots u'_{n-1} \xrightarrow[G']{w(n)} u'_n.$$

The case where $n = 0$ is trivial. Suppose the property is true for all paths of length at most n , and consider a path

$$u_0 \xrightarrow[G]{w(1)} \dots \xrightarrow[G]{w(n)} u_n \xrightarrow[G]{w(n+1)} u_{n+1}.$$

By induction hypothesis, one can find mappings $\delta_0, \dots, \delta_n$ such that

$$\delta_0 u_0 \xrightarrow[G']{w(1)} \dots \xrightarrow[G']{w(n)} \delta_n u_n.$$

We now use the properties of mappings stated in Lemma 3.1. By (3.1), there exist δ'_n and δ'_{n+1} such that $\delta'_n u_n \xrightarrow{w(n+1)} \delta'_{n+1} u_{n+1} \in G'$. Let γ_n and γ'_n be two elements of \mathcal{B} such that $\delta'_n \circ \gamma'_n = \delta_n \circ \gamma_n$. By Lemma (3.2) and (3.3), we can find mappings γ'_{n+1} and γ_0 to γ_{n-1} such that:

$$\gamma_0 \delta_0 u_0 \xrightarrow[G']{w(1)} \dots \xrightarrow[G']{w(n)} \gamma_n \delta_n u_n = \gamma'_n \delta'_n u_n \xrightarrow[G']{w(n+1)} \gamma'_{n+1} \delta'_{n+1} u_{n+1}$$

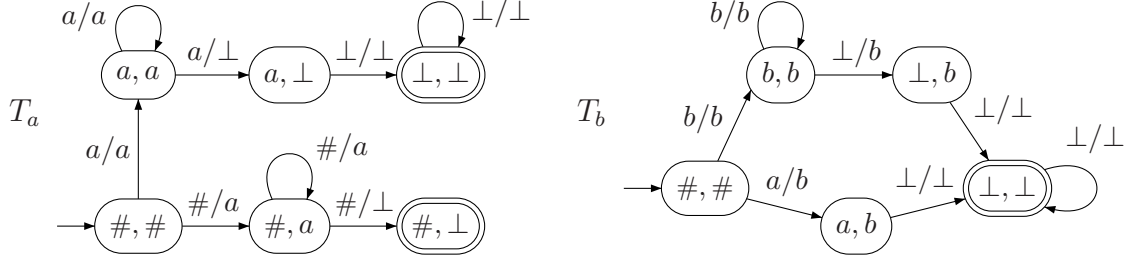


Figure 5: Transducers of a synchronous graph accepting $\{a^n b^n \mid n \geq 1\}$.

which concludes the proof by induction. If we suppose that $u_0 \in I$ and $u_n \in F$, then necessarily $u'_0 \in I'$ and $u'_n \in F'$. It follows that for every path in G between I and F , there is a path in G' between I' and F' with the same path label.

Conversely, by (3.1), for any such path in G' , erasing the occurrences of $\#$ from its vertices yields a valid path in G between I and F . Hence $L(G, I, F) = L(G', I', F')$. \square

3.2. Equivalence between synchronized graphs and tiling systems

The following propositions establish the tight relationship between tiling systems and synchronous rational graphs. Proposition 3.3 presents an effective transformation of a tiling system into a synchronous rational graph.

Proposition 3.3. *Given a tiling system $S = (\Gamma, \Sigma, \#, \Delta)$, there exists a synchronous rational graph G and two rational sets I and F such that $L(G, I, F) = L(S)$.*

Proof. Consider the finite automaton A on Γ with a set of states $Q = \Gamma \cup \{\#\}$, initial state $\#$, a set of final states F and a set of transitions δ given by:

$$\begin{aligned}
 F : a \quad \text{such that} \quad & \begin{array}{|c|c|} \hline a & \# \\ \hline \# & \# \\ \hline \end{array} \in \Delta \\
 \delta : \# \xrightarrow[A]{a} a, a \xrightarrow[A]{b} b \quad \text{for all} \quad & \begin{array}{|c|c|} \hline \# & \# \\ \hline a & \# \\ \hline \end{array}, \begin{array}{|c|c|} \hline a & \# \\ \hline b & \# \\ \hline \end{array} \in \Delta \text{ (respectively).}
 \end{aligned}$$

Call M the language recognized by A , M represents the set of possible last columns of pictures of $P(S)$. Note that this does not imply that each word of M actually *is* the last column of a picture in $P(S)$, only that it is compatible with the right border tiles of Δ .

Let us build a synchronous rational graph G and two rational sets I and F such that $L(G, I, F) = L(S)$. The transitions of the set of transducers $(T_e)_{e \in \Sigma}$ of G are:

$$\begin{aligned}
 (\#, \#) \xrightarrow[T_d]{c/d} (c, d) \quad \text{for all} \quad & \begin{array}{|c|c|} \hline \# & \# \\ \hline c & d \\ \hline \end{array} \in \Delta, d \neq \# \\
 (a, b) \xrightarrow[T_e]{c/d} (c, d) \quad \text{for all} \quad & \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array} \in \Delta, b, d \neq \#, e \in \Sigma
 \end{aligned}$$

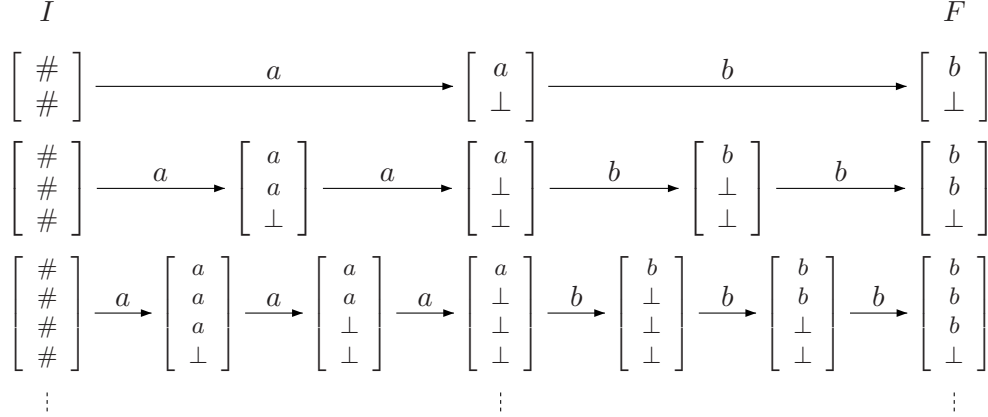


Figure 6: The synchronous rational graph associated to the tiling system of Figure 2 whose transducers are presented in Figure 5.

where $(\#, \#)$ is the unique initial state of each transducer and the set of final states F of each transducer is given by:

$$F : (a, b) \in (\Gamma \cup \{\#\}) \times \Gamma \text{ such that } \begin{array}{|c|c|} \hline a & b \\ \hline \# & \# \\ \hline \end{array} \in \Delta.$$

A pair of words (s, t) is accepted by the transducer T_e if and only if e is the first letter of t , and either s and t are two adjacent columns of a picture in $P(S)$ or $s \in \#^*$ and t is the first column of a picture in $P(S)$. As a consequence, $L(S) = L(G, \#^*, M)$. \square

Example 3.4. Figure 5 shows the transducers obtained using the previous construction on the tiling system of Figure 2. They define a rational graph whose path language between $\#^*$ and $b^+ \perp$ is $\{a^n b^n \mid n \geq 1\}$. Figure 6 presents the corresponding synchronous graph whose vertices are the rational set of words $\#^{\geq 2} \cup a^+ \perp^+ \cup b^+ \perp^+$, the set of initial vertices is $\#^{\geq 2}$ and the set of final vertices is $b^+ \perp$. Remark that in this example, the set of vertices accessible from the initial vertices is rational: this is not true in the general case.

Remark 3.5. The correspondence between a tiling system S and the synchronous graph G constructed from S in Proposition 3.3 is tight: each picture p with frontier w can be mapped to a unique accepting path for w in G (and conversely).

Conversely, Proposition 3.7 states that the languages accepted by synchronous rational graphs between rational sets of vertices can be accepted by a tiling system. To make the construction simpler, we first prove that the sets of initial and final vertices can be chosen over a one-letter alphabet without loss of generality.

Lemma 3.6. *For every synchronous rational graph G with vertices in Γ^* and rational sets I and F , one can find a synchronous rational graph H and two symbols i and $f \notin \Gamma$ such that $L(G, I, F) = L(H, i^*, f^*)$.*

Proof. Let $G = (K_a)_{a \in \Sigma}$ be a synchronous rational graph with vertices in Γ^* . For i, f two new distinct symbols, we define a new synchronous rational graph H characterized by the set of transductions $(T_a = (T_I \circ K_a) \cup K_a \cup (K_a \circ T_F))_{a \in \Sigma}$ where $T_I = \{(i^n, u) \mid n \geq 0, u \in I, |u| = n\}$ and $T_F = \{(v, f^n) \mid n \geq 0, v \in F, |v| = n\}$. For all vertices $u \in I, v \in F$ we have $u \xrightarrow[G]{w} v$ if and only if $i^{|u|} \xrightarrow[H]{w} f^{|u|}$, i.e. $L(G, I, F) = L(H, i^*, f^*)$. \square

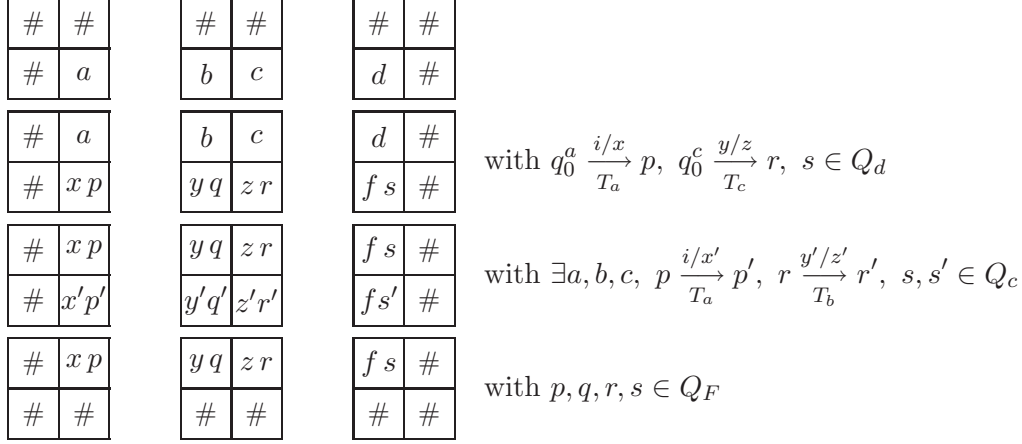


Figure 7: Tiling system accepting the language of a synchronous graph.

We are now able to establish the converse of Proposition 3.3, which states that all the languages accepted by synchronous rational graphs between rational sets of vertices can be accepted by a tiling system.

Proposition 3.7. *Given a synchronous rational graph G and two rational sets I and F , there exists a tiling system S such that $L(S) = L(G, I, F)$.*

Proof. Let $G = (T_a)_{a \in \Sigma}$ be a synchronous rational graph with vertices in Γ^* (with $\Sigma \subseteq \Gamma$). By Lemma 3.6, we can consider without loss of generality that $I = i^*$ and $F = f^*$ for some distinct letters i and f , and that neither i nor f occurs in any vertex which is not in I or F . Furthermore by Remark 2.1, we can assume that T_a is non-ambiguous for all $a \in \Sigma$.

We write Q_a the set of control states of T_a . We suppose that all control state sets are disjoint, and designate by $q_0^a \in Q_a$ the unique initial state of each transducer T_a , and by Q_F the set of final states of all T_a .

Let $a, b, c, d \in \Sigma$, $x, x', y, y', z, z' \in \Gamma$, and $p, p', q, q', r, r', s, s' \in \bigcup_{a \in \Sigma} Q_a$. We define a tiling system $S = (\Gamma, \Sigma, \#, \Delta)$, where Δ is the set of tiles from Figure 7. By construction, $P(S)$ is in exact bijection with the set of accepting paths in G with respect to I and F . Let ϕ be the function associating to a picture $p \in P(S)$ with columns $a_1 w_1, \dots, a_n w_n$, the path $i^{|w_1|} \xrightarrow{a_1} \widetilde{w}_1 \dots \xrightarrow{a_n} \widetilde{w}_n$ where \widetilde{w} is obtained by removing the control states from w . By construction of S , the function ϕ is well defined. It is easy to check that ϕ is an onto function. As the transducers defining G are non-ambiguous, two distinct pictures have distinct images by ϕ and therefore ϕ is an injection.

Hence, the tiling system $(\Gamma, \Sigma, \#, \Delta)$ exactly recognizes $L(G, I, F)$. \square

Remark 3.8. As in Remark 3.5, the set of paths in G from I to F and the set of pictures $P(S)$ accepted by S are in bijection, and the length of the vertices along the path is equal to the height of the corresponding picture.

Putting together Propositions 3.3 and 3.7 and Theorem 2.6, we obtain the following result concerning the path languages of synchronous rational graphs.

Theorem 3.9 ([Ris02]). *The languages accepted by synchronous rational graphs between rational sets of initial and final vertices are the context-sensitive languages.*

Note that this formulation of the theorem could be made a bit more precise by recalling that initial and final sets of vertices only of the form x^* , where x is a letter, are sufficient to accept all context-sensitive languages, as stated in Lemma 3.6. By Proposition 3.2, this implies as a corollary the original result by Morvan and Stirling [MS01].

Corollary 3.10. *The languages accepted by rational graphs between rational sets of initial and final vertices are the context-sensitive languages.*

If we transform a rational graph into a Turing machine by applying successively the construction of Proposition 3.2, Proposition 3.7 and Theorem 2.6, we obtain the same Turing machine as in [MS01].

3.3. Sequential synchronous graphs are enough

Theorem 3.9 shows that when considering rational sets of initial and final vertices, synchronous graphs are enough to accept all context-sensitive languages. Interestingly, when considering rational sets of initial and final vertices, the even more restricted class of *sequential* synchronous transducers are sufficient.

Proposition 3.11. *The languages accepted by sequential synchronous rational graphs between rational sets of initial and final vertices are the context-sensitive languages.*

Proof. Thanks to Proposition 3.7, it suffices to prove that any context sensitive language $L \subseteq \Sigma^*$ is accepted by a synchronous sequential rational graph. By Theorem 2.6, we know that there exists a tiling system $S = (\Gamma, \Sigma, \#, \Delta)$ such that $L(S) = L$.

Let $\Lambda = \Gamma \cup \{\#\}$ and $[$ and $]$ be two symbols that do not belong to Λ . We associate to each picture $p \in \Lambda^{**}$ with rows l_1, \dots, l_n the word $[l_1] \dots [l_n]$. We are going to define a set of sequential synchronous transducers that, when iterated, recognize the words corresponding to pictures in $P(S)$.

First, for any finite set of tiles Δ , we construct a transducer T_Δ which checks that a word in $([\Lambda^{\geq 3}])^{\geq 2}$ represents a picture with tiles in Δ . The checking is done column by column, and we introduce marked letters to keep track of the column being checked. Let $\tilde{\Lambda}$ be a finite alphabet in bijection with but disjoint from Λ . For all $x \in \Lambda$ we write $\tilde{x} \in \tilde{\Lambda}$ the marked version of x . For every word $w = u\tilde{x}v \in \Lambda^*\tilde{\Lambda}\Lambda^*$, we write $\pi(w)$ the word $uxv \in \Lambda^*$ and $\rho(w) = |u| + 1$ designates the position of the marked letter in the word.

We consider words in $(\Lambda^*\tilde{\Lambda}\Lambda^*)^{\geq 2}$. Let Shift be the relation that shifts all marks in a word one letter to the right. More precisely, Shift satisfies $Dom(\text{Shift}) = ([\Lambda^*\tilde{\Lambda}\Lambda^+])^{\geq 2}$, and $\text{Shift}([w_1] \dots [w_n]) = [w'_1] \dots [w'_n]$ with $\pi(w'_i) = \pi(w_i)$ and $\rho(w'_i) = \rho(w_i) + 1$ for all $i \in [1, n]$. The rational relation Shift can be realized by a synchronous sequential transducer T_{Sh} . Consider the following rational language:

$$R_\Delta = \left\{ [w_1 x_1 \tilde{y}_1 w'_1] \dots [w_n x_n \tilde{y}_n w'_n] \mid n \geq 2 \text{ and } \forall i \in [2, n], \begin{array}{|c|c|} \hline x_{i-1} & y_{i-1} \\ \hline x_i & y_i \\ \hline \end{array} \in \Delta \right\}.$$

The transducer T_Δ obtained by restricting T_{Sh} to the domain R_Δ is both synchronous and sequential. For all $w = [w_1] \dots [w_n] \in ([\Lambda\tilde{\Lambda}\Lambda^*])^{\geq 2}$, if $w' = T_\Delta^N(w)$ then $w' = [w'_1] \dots [w'_n]$ with $\pi(w_i) = \pi(w'_i)$ and $\rho(w'_i) = N + 2$ for all $i \in [1, n]$. Let r_i be the word containing the $N + 1$ first letters of w'_i , a straightforward induction on N proves that the picture p formed of the rows r_1, \dots, r_n only has tiles in Δ . In particular, $T_\Delta^N(w)$ belongs to $([\Lambda^*\tilde{\Lambda}])^* \cap R_\Delta$ if and only if $\pi(w)$ represent a picture p of width $N + 2$ such that $T(p) \subseteq \Delta$.

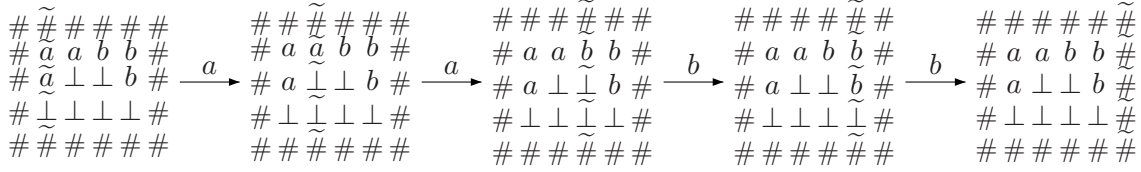


Figure 8: Connected component of a sequential synchronous graph accepting $\{a^n b^n \mid n \geq 1\}$.

We now define more precisely the sequential rational graph $G = (T_a)_{a \in \Sigma}$ accepting L . For all $a \in \Sigma$, the transducer T_a is obtained by restricting the domain of T_Δ to the set of words representing pictures whose marked symbol on the second row is a , i.e. to the set $[(\Lambda \cup \tilde{\Lambda})^*][(\Lambda \tilde{a} \Lambda^*)][(\Lambda \cup \tilde{\Lambda})^*]$. T_a can be chosen synchronous and sequential. The set of initial vertices I is $[\#\#\#^*](\#\tilde{\Gamma}\Gamma^*\#)^*[\#\#\#^*]$ and the set of final vertices F is $[\#\#\#^*](\#\tilde{\Gamma}\Gamma^*\#)^*[\#\#\#^*]$. \square

Example 3.12. Figure 8 shows a part of the result of the previous construction when applied to the language $\{a^n b^n \mid n \geq 1\}$ as recognized by the tiling system of Figure 2. Each vertex is represented by the corresponding picture, instead of the word coding for it. Also, only one connected component of the graph is shown. The other connected components all have the same linear structure: the degree of the graph is bounded by 1. The leftmost vertex belongs to the set I , and the rightmost to the set F , hence the word $a^2 b^2$ is accepted.

Remark 3.13. In the case of synchronized transducers, it has been shown in Lemma 3.6 that I could be taken over a one letter alphabet without loss of generality. This does not seem to hold for sequential transducers as the proof we present relies on the expressiveness of the initial set of vertices. In fact, as shown in Proposition 5.3, the languages recognized by sequential synchronous graph from i^* are deterministic context-sensitive languages.

4. RATIONAL GRAPHS SEEN AS AUTOMATA

The structure of the graphs obtained in the previous section (propositions 3.3 and 3.11) is very poor. Synchronous graphs are by definition composed of a possibly infinite set of *finite* connected components. In the case of Proposition 3.11, we obtain an even more restricted family of graphs since both their in-degree and out-degree is bounded by 1. However, when considering accepted languages from a possibly infinite rational set of vertices, even this extremely restricted family accepts the same languages as the most general rational graphs, namely all context-sensitive languages. This is why, in order to compare the expressiveness of the different sub-families of rational graphs and to obtain graphs with richer structures, we need to impose structural restrictions.

We first consider graphs with a single initial vertex, but this restriction alone is not enough. In fact, both synchronized and rational graphs with a rational set of initial vertices accept the same languages as their counterparts with a single initial vertex.

Lemma 4.1. *For every rational graph (resp. synchronized graph) G and for every pair of rational sets I and F , there exists a rational graph (resp. a synchronized graph) G' , a vertex i and a rational set F' such that $L(G, I, F) = L(G', \{i\}, F')$.*

Proof. Let $G = (T_a)_{a \in \Sigma}$ be a rational graph with vertices in Γ^* and let i be a symbol which does not belong to Γ and $\Gamma' = \Gamma \cup \{i\}$. For all $a \in \Sigma$, let T'_a be a transducer recognizing the rational relation $T_a \cup \{(i, w) \mid w \in T_a(I)\}$. Remark that if T_a is synchronized then T'_a can also be chosen synchronized. If $\varepsilon \notin L(G, I, F)$ then $F' = F$ else $F' = F \cup \{i\}$. It is straightforward to show that $L(G, I, F) = L(G', \{i\}, F')$. \square

It follows from Proposition 3.3 and Lemma 4.1 that the synchronized rational graphs with one initial vertex accept the context-sensitive languages [Ris02].

Remark 4.2. It is fairly obvious that this result does not hold for synchronous graphs: indeed, the restriction of a synchronous rational graph to the vertices reachable from a single vertex is finite. Hence, the languages of synchronous graphs from a single vertex are rational. Similarly, as any rational language is accepted by a deterministic finite graph, it can also be accepted by a sequential synchronous graph with a single initial vertex.

Note that the construction of Lemma 4.1 relies on infinite out-degree to transform a synchronous graph with a rational set of initial vertices into a rational one with a single initial vertex. In order to obtain more satisfactory notions of infinite automata, we now restrict our attention to graphs of finite out-degree with a single initial vertex.

4.1. Rational graphs of finite out-degree with one initial vertex.

We present a syntactical transformation of a synchronous rational graph with a rational set of initial vertices into a rational graph of finite out-degree with a unique initial vertex accepting the same language.

The construction relies on the fact that for a synchronous graphs to recognize a word of length $n > 0$, it is only necessary to consider vertices whose length is smaller than c^n (where c is a constant depending only on the graph). We first establish a similar result for tiling systems and conclude using the close correspondence between synchronous graphs and tiling systems established in Proposition 3.7.

Lemma 4.3. *For any tiling system $S = (\Gamma, \Sigma, \#, \Delta)$, if $p \in P(S)$ then there exists a (n, m) -picture p' such that $\text{fr}(p) = \text{fr}(p')$ and $n \leq |\Gamma|^m$.*

Proof. Let p' be a (n, m) -picture with $n > |\Gamma|^m$, and suppose that p' is the smallest picture in $P(S)$ with frontier $\text{fr}(p)$. Let l_1, \dots, l_n be the rows of p' . As $n > |\Gamma|^m$ then there exists $j > i \geq 1$ such that $l_i = l_j$. Let p'' be the picture with rows $l_1, \dots, l_i, l_{j+1}, \dots, l_n$. It is easy to check that $T(p''_{\#}) \subset T(p'_{\#})$, we have that $p'' \in P(S)$ and as p'' has a smaller height than p' but the same frontier, we obtain a contradiction. \square

We know from Remark 3.8 that for every synchronous rational graph $G = (T_a)_{a \in \Sigma}$ and two rational sets I and F , there exists a tiling system S such that $i \xrightarrow[G]{w} f$ with $i \in I$ and $f \in F$ if and only if there exists $p \in K$ such that $\text{fr}(p) = w$ and p has height $|i| = |f|$. Hence, as a direct consequence of Lemma 4.3, one gets:

Lemma 4.4. *For every synchronous rational graph G and rational sets I and F , there exists $k \geq 1$ such that:*

$$\forall w \in L(G, I, F), \exists i \in I, f \in F \text{ such that } i \xrightarrow[G]{w} f \text{ and } |i| = |f| \leq k^{|w|}.$$

We can now construct of a rational graph of finite out-degree accepting from a single vertex the same language as a synchronous graph with a rational set of initial vertices.

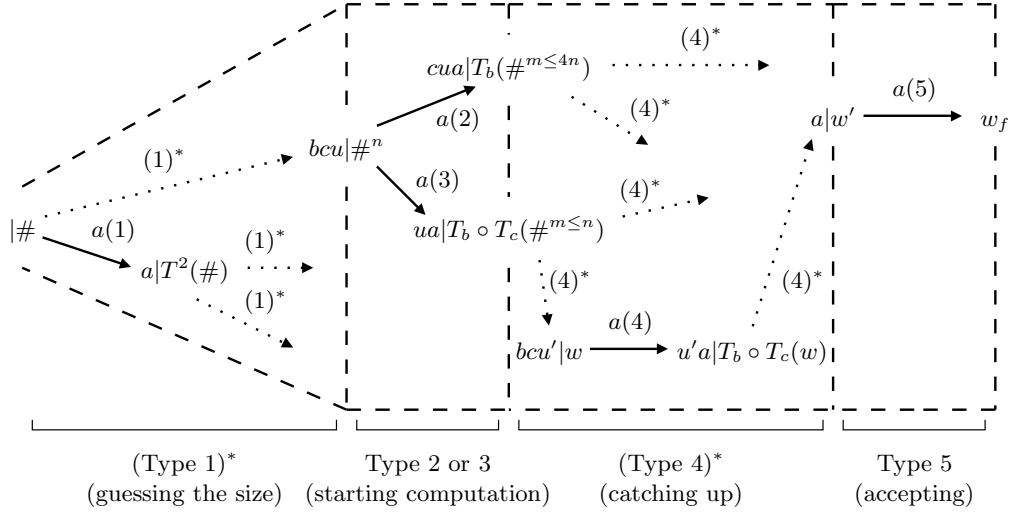


Figure 9: Schema of the construction in Proposition 4.5.

Proposition 4.5. *For every synchronous rational graph G and rational sets I and F such that $I \cap F = \emptyset$, there is a rational graph H of finite out-degree and a vertex i such that $L(G, I, F) = L(H, \{i\}, F)$.*

Proof. According to Lemma 3.6, there exists a synchronous rational graph R described by a set of transducers $(T_a)_{a \in \Sigma}$ over Γ^* such that $L(G, I, F) = L(R, \#^*, F)$. Note that for all $w \in \#^*$ and $w' \in \Gamma^*$, if $w \xrightarrow{R} w'$ then w' does not contain $\#$. We define a graph H such that $L(G, I, F) = L(H, \{i\}, F)$ for some vertex i of H . Let k be the constant involved in Lemma 4.4, T and T' two transducers realizing the rational relations $\{(\#^n, \#^{kn}) \mid n \in \mathbb{N}\}$ and $\{\#^n, \#^m \mid m \in [1, n]\}$ respectively. For all $a, b, c \in \Sigma$ and $u \in \Sigma^*$, H has edges:

$$\begin{array}{llll}
 \forall n \in \mathbb{N}, & u|\#^n & \xrightarrow{a} & ua|T \circ T'(\#^n) & \text{(Type 1)} \\
 \forall n \in \mathbb{N}, & bu|\#^n & \xrightarrow{a} & ua|T \circ T' \circ T_b(\#^n) & \text{(Type 2)} \\
 \forall n \in \mathbb{N}, & bcu|\#^n & \xrightarrow{a} & ua|T' \circ T_b \circ T_c(\#^n) & \text{(Type 3)} \\
 \forall w \in (\Gamma \setminus \{\#\})^*, & bcu|w & \xrightarrow{a} & ua|T_b \circ T_c(w) & \text{(Type 4)} \\
 \forall w \in (\Gamma \setminus \{\#\})^*, & b|w & \xrightarrow{a} & T_b \circ T_a(w) & \text{(Type 5)} \\
 & | \# & \xrightarrow{a} & T \circ T' \circ T_a(\#) & \text{(Type 6)}
 \end{array}$$

The graph H is clearly rational and of finite out-degree. We take $i = | \#$ as initial vertex. Remark that in H an edge of type 2 or 3 cannot be followed by edges of type 1, 2 or 3, and at most one edge of type 2 or 3 and of type 5 or 6 can be applied. Moreover, an edge of type 1 increases the length of the left part of the word by one, and an edge of type 4 decreases it by one. Also, in any accepting path, the last edge is of type 5 or 6. Figure 9 illustrates the structure of the obtained graph. It is technical but straightforward to show a correspondence between accepting paths in H and R , and to conclude that $L(R, \#^*, F) = L(H, \{i\}, F)$. \square

From Proposition 3.2 and Proposition 4.5, we deduce that the rational graphs of finite out-degree with one initial vertex accept all context-sensitive languages. This result was proved in [MS01] using the Penttonen normal form of context-sensitive grammars [Pen74].

Theorem 4.6. *The path languages of rational graphs of finite out-degree from a unique initial vertex to a rational set of final vertices are the context-sensitive languages.*

4.2. Synchronized graphs of finite out-degree with one initial vertex

We now consider the languages of synchronized graphs of finite out-degree with one initial vertex. First, we characterize them as the languages recognized by tiling systems with square pictures (i.e. for which there exists $c \in \mathbb{N}$ such that for every word $w \in L(S)$, there exists a (n, m) -picture in $P(S)$ with $n \leq cm$ and with frontier w). A slight adaptation of the construction of Proposition 4.5 gives the first inclusion.

Proposition 4.7. *Let $S = (\Gamma, \Sigma, \#, \Delta)$ be a tiling system with square pictures. There exists a synchronized rational graph of finite degree accepting $L(S)$ from one initial vertex.*

Proof. Let $G = (T_a)_{a \in \Sigma}$ be the synchronized graph obtained from S in Proposition 3.3. In the construction from the proof of Proposition 4.5, if we replace the transducer T by a transducer S realizing the synchronized relation $\{(\#^n, \#^{n+c}) \mid n \in \mathbb{N}\}$, we obtain a synchronized graph H , a vertex i and a set F such that $L(H, i, F) = L(S)$. \square

Before proceeding with the converse, we state a result similar to Lemma 4.4 for synchronized graphs of finite out-degree that states that when recognizing a word w from a unique initial vertex i , the vertices involved have a length at most linear in the size of w .

Lemma 4.8. *For any synchronized rational graph G of finite out-degree with vertices in Γ^* and for every vertex i , there exists a constant k such that for all w in $L(G, \{i\}, F)$, there exists a path from i to some $f \in F$, labeled by w , and with vertices of size at most $k \cdot |w|$.*

Proof. It follows from the definition of synchronized transducers that for every synchronized transducer of finite out-degree there exists $c \in \mathbb{N}$ such that $(x, y) \in T$ implies that $|x| \leq |y| + c$ (see [Sak03] for a proof of this result). We take k to be the maximum over the set of transducers defining G of these constants. The result follows by a straightforward induction on the size of w . \square

The converse inclusion is obtained by remarking that the composition of the construction of Proposition 3.2 and Proposition 3.7 gives a tiling system with square pictures when applied to a synchronized graph of finite out-degree.

Proposition 4.9. *Let $G = (T_a)_{a \in \Sigma}$ be a synchronized graph of finite out-degree. For every initial vertex i and set of final vertices F , there exists a tiling system S with square pictures such that $L(S) = L(G, \{i\}, F)$.*

Proof. Let G' , I' and F' be the synchronous graph and the rational set of initial and final vertices obtained by applying the constructions of Proposition 3.2 to G , $\{i\}$ and F . It is easy to show that for every word $w \in L(G', I', F')$, there exists $i' \in I'$ and $f' \in F'$ such that $i' \xRightarrow{w} f'$ with $|i'| = |f'| \leq k|w|$ where k is the constant of Lemma 4.8 for G . We conclude by Proposition 3.7, that states the existence of a tiling system S such that $L(S) = L(G', I', F')$. By Remark 3.5, S is a tiling system with square pictures. \square

Putting together Proposition 4.9 and Proposition 4.7 and with the use of the simulation result from Theorem 2.6, we obtain the following theorem.

Theorem 4.10. *The languages accepted by synchronized graphs of finite out-degree from a unique vertex to a rational set of vertices are the context-sensitive languages recognized by non-deterministic linearly bounded machines with a linear number of head reversals.*

We conjecture that this class is strictly contained in the context-sensitive languages. However, few separation results exist for complexity classes defined by time and space restrictions (see for example [vM04]). In particular, the diagonalization techniques (see [For00]) used to prove that the polynomial time hierarchy (with no space restriction) is strict do not apply for lack of a suitable notion of *universal* LBM.

4.3. Bounding the out-degree

It is natural to wonder if the rational graphs still accept the context-sensitive languages when considering bounded out-degree. This is a difficult question, to which we only provide here a partial answer concerning the synchronized graphs of bounded out-degree.

It follows from Lemma 4.8 that the vertices used to accept a word w in a synchronized rational graph have a length at most linear in the length of w and therefore, can be stored on the tape of a LBM. Moreover if the graph is deterministic, we can construct a deterministic LBM accepting its language.

Proposition 4.11. *The language accepted by a deterministic synchronized graph from a unique initial vertex is deterministic context-sensitive.*

Proof. Let $G = (T_a)_{a \in \Sigma}$ be a deterministic synchronized graph over Γ , i a vertex and F a rational set of vertices. We define a deterministic LBM M accepting $L(G, \{i\}, F)$. When accepting $w = a_1 \dots a_{|w|}$, M starts by writing i on its tape. It successively applies $T_{a_1}, \dots, T_{a_{n-1}}$ and T_{a_n} to i . If the image of the current tape content by one of these transducers is not defined, the machine rejects. Otherwise, it checks whether the last tape content represents a vertex which belongs to F .

We now detail how the machine M can apply one of the transducers T of G to a word x in a deterministic manner. As T has a finite image, we can assume without loss of generality that $T = (\Gamma, Q, i, F, \delta)$ is in real-time normal form: $\delta \subset Q \times \Gamma \times \Gamma^* \times Q$ (see for instance [Ber79] for a presentation of this result). The machine enumerates all paths in T of length less than $c|x|$ in the lexicographic order where c is the constant associated to G in Lemma 4.8. For each such path ρ , it checks if it is an accepting path for input x , and in that case replaces x by the output of ρ .

The space used by M when starting with a word w is bounded by $(2c+1)|w|$. Moreover if M accepts w , then there exists a path from i to a vertex F in G labeled by w . Conversely, if w belongs to $L(G, \{i\}, F)$ then by Lemma 4.8, there exists a path in G from i to F with vertices of length at most $c|w|$ and by construction M accepts w . Hence, M is a deterministic linearly bounded Turing machine accepting $L(G, \{i\}, F)$. \square

Remark 4.12. The result of Proposition 4.11 extends to any deterministic rational graph satisfying the property expressed by Lemma 4.8.

The previous result can be extended to synchronized graphs of bounded out-degree thanks to a uniformization result by Weber. First observe that a rational graph is of out-degree bounded by some constant k if and only if it is defined by transducers which associate

at most k distinct images to any input word. The relations realized by these transducers are called k -valued rational relations.

Proposition 4.13 ([Web96]). *For any k -valued rational relation R , there exist k functional rational relations F_1, \dots, F_k such that $R = \bigcup_{i \in [1, k]} F_i$.*

Note that even if R is a synchronized relation, the F_i 's are not necessarily synchronized. However, they still satisfy the inequality $|y| \leq |x| + c$ for all $(x, y) \in F_i$.

To any synchronized graph G with an out-degree bounded by k defined by a set of transducers $(T_a)_{a \in \Sigma}$, we associate the deterministic rational graph H defined by $(F_{a_i})_{a \in \Sigma, i \in [1, k]}$ where for all $a \in \Sigma$, $(F_{a_i})_{i \in [1, k]}$ is the set of rational functions associated to T_a by Proposition 4.13. According to Proposition 4.11 and to Remark 4.12, $L(H, \{i\}, F)$ is a deterministic context-sensitive language. Let π be the alphabetical projection defined by $\pi(a_i) = a$ for all $a \in \Sigma$ and $i \in [1, k]$, it is straightforward to establish that $\pi(L(H, \{i\}, F)) = L(G, \{i\}, F)$. As deterministic context-sensitive languages are closed under alphabetical projections, $L(G, \{i\}, F)$ is a deterministic context-sensitive language.

Theorem 4.14. *The language accepted by a synchronized graph of bounded out-degree from a unique initial vertex is deterministic context-sensitive.*

The converse result is not clear, for reasons similar to those presented in the previous section for synchronized graphs of finite degree. A precise characterization of the family of languages accepted by synchronized rational graphs of bounded degree would be interesting.

5. NOTIONS OF DETERMINISM

In this last part of the section on rational graphs, we investigate families of graphs which accept the deterministic context-sensitive languages. First of all, we examine the family yielded by the previous constructions when applied to deterministic languages. Then, we propose a global property over sets of transducers characterizing a sub-family of rational graphs whose languages are precisely the deterministic context-sensitive languages.

5.1. Unambiguous context-sensitive languages

When applying the construction of Proposition 3.3 to a deterministic tiling system S , one obtains a synchronous rational graph G (which is non-deterministic in general) and two rational sets of vertices I and F such that $L(G, I, F) = L(S)$, with the particularity that for every word w in L , there is *exactly one* path labeled by w leading from some vertex in I to a vertex in F : G is *unambiguous* with respect to I and F . However, the converse is not granted: given a graph G and two rational sets I and F such that G is unambiguous with respect to I and F , we cannot ensure that $L(G, I, F)$ is a deterministic context-sensitive language. Rather, the obtained languages can be accepted by unambiguous linearly bounded machines. This class of languages is called $\text{USPACE}(n)$, and it is not known whether it coincides with either the context-sensitive or deterministic context-sensitive languages.

Theorem 5.1. *Let L be a language, the following properties are equivalent:*

- (1) *L is an unambiguous context-sensitive language.*
- (2) *There exist a rational graph G with unambiguous transducers and two rational sets I and F with respect to which G is unambiguous, such that $L = L(G, I, F)$.*

This result only holds if one considers unambiguous transducers, i.e. transducers in which there is at most one accepting path per pair of words. The reason is that ambiguity in the transducers would induce ambiguity in the machine. However, since synchronized transducers can be made unambiguous (Cf. Remark 2.1), we can drop this requirement in the case of synchronized graphs. Note that the unambiguity of a rational or synchronized graphs with respect to rational sets of vertices is undecidable. However, since any rational function can be realized by an unambiguous transducer [Kob69, Sak03], the language of any deterministic rational graph is, by to Theorem 5.1, unambiguous.

Corollary 5.2. *The languages of deterministic rational graphs from an initial vertex i to a rational set F of vertices are unambiguous context-sensitive languages.*

5.2. Globally deterministic sets of transducers

We just saw an attempt at characterizing natural families of graphs whose languages are the deterministic context-sensitive languages, which was based on a restriction of previous constructions to the deterministic case, but failed to meet its objective because of a slight nuance between the notions of determinism and unambiguity for tiling systems.

First, we naturally consider the class of sequential synchronous automata with an initial set of the form $\{a\}^*$, where a is a letter of the vertex alphabet (in other words, a given initial vertex does not code for any information besides its length). It is easy to check that when applying the construction of Proposition 3.7 to one of these automata, we obtain a deterministic tiling system.

Proposition 5.3. *The languages of sequential synchronous graphs from $\{a\}^*$ are deterministic context-sensitive languages.*

The converse result seems difficult to prove due to the local nature of the determinism involved in this class. Hence, we consider a *global* property of the set of transducers characterizing a rational graph, so as to ensure that each accepting path corresponds to the run of a deterministic linearly bounded machine on the corresponding input, or equivalently that each accepting path corresponds to a picture recognized by a deterministic tiling system and whose upper frontier is the path label under consideration.

For any rational language L , we write T_L the minimal synchronous transducer recognizing the identity relation over L .

Definition 5.4. Let T be a set of synchronous transducers over Γ . We say T is *globally deterministic* with respect to two rational languages I and $F \subseteq \Gamma^*$ if all transducers in T are deterministic³ and for every pair of transducers $T_1 \in T \cup \{T_I\}$ and $T_2 \in T \cup \{T_F\}$, and every pair of control states $q_1 \in Q_{T_1}$ and $q_2 \in Q_{T_2}$, there is at most one b such that

$$q_1 \xrightarrow[T_1]{a/b} q'_1 \wedge q_2 \xrightarrow[T_2]{b/c} q'_2 \text{ for some } a, c \in \Gamma, \ q'_1 \in Q_{T_1}, \ q'_2 \in Q_{T_2}.$$

Intuitively, this condition states that, whenever a part of the output of one transducer can be read as input by a second transducer, there is only one way to add a letter to this word such that it is still compatible with both transducers. This property of sets of transducers is trivially decidable, since it is sufficient to check the above condition for every pair of control states of transducers in $(T \cup \{T_I\}) \times (T \cup \{T_F\})$. This allows us to capture a sub-family of rational graphs whose languages are the deterministic context-sensitive languages.

³i.e. whenever $q \xrightarrow{a/b} q'$ and $q \xrightarrow{c/d} q''$ with $q' \neq q''$, it implies $(a, b) \neq (c, d)$.

Theorem 5.5. *Let L be a language, the following two properties are equivalent:*

- (1) *L is a deterministic context-sensitive language.*
- (2) *There is a synchronous rational graph G and two rational sets I and F such that $L = L(G, I, F)$ and G is globally deterministic between I and F .*

Proof. Let $G = (T_a)_{a \in \Sigma}$ be a synchronous rational graph which is globally deterministic between I and F . The graph $H = (T'_a)_{a \in \Sigma}$ obtained by applying Lemma 3.6 to G is such that $L(H, i^*, f^*) = L(G, I, F)$. Moreover, H is globally deterministic between i^* and f^* .

We will show that the construction of Proposition 3.7, when applied to a rational graph H between i^* and f^* yields a deterministic tiling system. Suppose that this is not the case. Then, by definition of a non-deterministic tiling system, there must be words u , v_1 and v_2 with $v_1 \neq v_2$ such that the two-rows pictures p_1 and p_2 with first row $\#u\#$ and second row $\#v_1\#$ and $\#v_2\#$ respectively only have tiles in Δ . Since $v_1 \neq v_2$, let i be the smallest index such that $v_1(i) \neq v_2(i)$. Let $v_1(i) = xp$, $v_2(i) = x'p'$.

By the construction of Prop. 3.7, there are two transducers T_a and T_b such that

$$q_a \xrightarrow[T_a]{y/x} p \wedge q_b \xrightarrow[T_b]{x/z} q'_b \wedge q_a \xrightarrow[T_a]{y/x'} p' \wedge q_b \xrightarrow[T_b]{x'/z'} q''_b$$

for some symbols $y, y', z, z' \in \Gamma$ and control states q_a, q_b, q'_b and q''_b . As T_a is deterministic, if x is equal to x' , then $p = p'$ and $v_1(i) = v_2(i)$. Hence $x \neq x'$, and the above relations contradicts the global determinacy of H .

To prove the converse, we introduce yet another family of acceptors for context-sensitive languages, namely cellular automata. A cellular automaton is a tuple $(\Gamma, \Sigma, F, [,], \delta)$ where Γ and $\Sigma \subseteq \Gamma$ are the work and input alphabets, $F \subseteq \Gamma$, $[,] \notin \Gamma$ and δ is a set of 4-tuples over Γ called transition rules. These rules induce a transition relation over words of the form $[u] \in [\Gamma^*]$: $c' = [v]$ is a successor of $c = [u]$ if $|c| = |c'| = n$ and for all $i \in [2, n-1]$, $(c(i-1), c(i), c(i+1), c'(i)) \in \delta$. A word w is accepted if, starting from $[w]$ one can derive a word $[u]$ with $u \in F^*$. An cellular automaton is deterministic if for all A, B, C there is at most one D such that $(A, B, C, D) \in \delta$. Equivalence of (deterministic) cellular automata with (deterministic) LBMs or tiling systems is common knowledge.

Let L be any deterministic context-sensitive language, there exists a deterministic cellular automaton $C = (\Gamma, \Sigma, \perp, \delta, [,],)$ recognizing L . One can easily build two rational languages I and F and a set of transducers T globally deterministic with respect to I and F such that $L(G, I, F) = L$ where G is the rational graph defined by T . The work alphabet of T is $\Gamma' = \Sigma \cup \{[,]\} \cup \delta$. The set of control states of transducer $T_a \in T$ is $\{q_0^a\} \cup \{q_{AB}^a \mid A, B \in \Gamma \cup \{[,]\}\}$, where q_0^a is the unique initial state. Its transitions are:

$$\forall a, b \in \Sigma, \quad q_0^a \xrightarrow{[a]} q_a^a \quad \text{and} \quad q_0^a \xrightarrow{b/a} q_{ba}^a$$

$$\forall d_1 = ([, A, B, A') \in \delta, \quad q_{[A}^a \xrightarrow{[d_1]} q_{[A'}^a$$

$$\forall d_1 = (A, B, C, B'), d_2 = (B, C, D, C') \in \delta, \quad q_{BC}^a \xrightarrow{d_1/d_2} q_{B'C'}^a$$

The terminal states of T_a are q_{\perp} and $q_{\perp\perp}$. Now let $I = ([])^*$ and $F = \Sigma R^*$ where $R = \{(a, b, [, b') \in \delta \mid a, b, b' \in \Gamma\}$. By construction and since C is deterministic, T is globally deterministic with respect to I and F . One can easily verify that $L(G, I, F) = L$. \square

6. CONCLUSION

This work is a summary of new and existing results concerning rational graphs and their relation to context-sensitive languages. Its main contributions are, first, to show the language equivalence between rational graphs and synchronous rational graphs, and second to establish a tight connection between synchronous rational graphs and finite tiling systems. Since tiling systems accept precisely the context-sensitive languages, this yields a new and simpler proof that the languages of rational graphs coincide with this family. Thanks to this, we studied the impact of structural restrictions on the obtained family of languages, in particular when considering finite or bounded degree and a single initial vertex.

This approach also enables us to consider the case of deterministic languages. We show how one can define sub-families of rational graphs whose languages are precisely the unambiguous or deterministic context-sensitive languages. However, due to their syntactical nature, these results brings little new insight as to the difficult question of the strictness of inclusions between deterministic, unambiguous and general context-sensitive languages.

This presentation gives rise to a few interesting open questions. A thorough study of graphs of bounded degree seems necessary, albeit difficult. More generally, the question of knowing whether any “tractable” family of graphs accepting the context-sensitive languages exists remains. We saw that synchronous graphs are not a good option since they lose all their expressive power when only a finite number of initial vertices are considered. Synchronized graphs form an interesting class, especially since their first order theory is decidable, but it seems reasonable to believe that they require infinite out-degree to accept all context-sensitive languages.

Another question is to compare the rational graphs with the transition graphs of linearly bounded machines [KP99, Pay00]. This last point is addressed to some extent in [CM05], where it is shown that all bounded degree rational graphs are isomorphic to transition graphs of linearly bounded machines.

ACKNOWLEDGEMENT.

The authors would like to thank Kamal Lodaya for his comments, and for pointing out the interest of using tiling systems, and Didier Caucal for his general advice and support.

REFERENCES

- [Ber79] J. Berstel. *Transductions and Context-Free Languages*. Teubner Verlag, 1979.
- [BG00] A. Blumensath and E. Grädel. Automatic structures. In *Proceedings of the 15th IEEE Symposium on Logic in Computer Science (LICS 2000)*, pages 51–62. IEEE, 2000.
- [Cau96] D. Caucal. On infinite transition graphs having a decidable monadic theory. In *ICALP*, pages 194–205, 1996.
- [Cau03a] D. Caucal. On infinite transition graphs having a decidable monadic theory. *Theoretical Computer Science*, 290:79–115, 2003.
- [Cau03b] D. Caucal. On the transition graphs of Turing machines. *Theoretical Computer Science*, 296:195–223, 2003.
- [Cho59] N. Chomsky. On certain formal properties of grammars. *Information and Control*, 2:137–167, 1959.
- [CK02] D. Caucal and T. Knapik. A Chomsky-like hierarchy of infinite graphs. In *Mathematical Foundations of Computer Science 2002, 27th International Symposium (MFCS 2002)*, volume 2420 of *Lecture Notes in Computer Science*, pages 177–187, 2002.
- [CM05] A. Carayol and A. Meyer. Linearly bounded infinite graphs. In *Mathematical Foundations of Computer Science 2005, 30th International Symposium (MFCS 2005)*, volume 3618 of *Lecture*

- Notes in Computer Science*, pages 180–191. Springer Verlag, 2005. Long version to appear in *Acta Informatica*.
- [EM65] C. Elgot and J. Mezei. On relations defined by finite automata. *IBM Journal of Research and Development*, 9:47–68, 1965.
 - [For00] L. Fortnow. Diagonalization. *Bulletin of the European Association for Theoretical Computer Science*, 71:102–112, 2000.
 - [GR96] D. Giammarresi and A. Restivo. *Handbook of Formal Languages*, volume 3, chapter Two-dimensional languages. Springer Verlag, 1996.
 - [HU79] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 1979.
 - [Kob69] K. Kobayashi. Classification of formal languages by functional binary transductions. *Information and Control*, 15:95–109, 1969.
 - [KP99] T. Knapik and É. Payet. Synchronized product of linear bounded machines. In *Fundamentals of Computation Theory, 12th International Symposium (FCT 1999)*, volume 1684 of *Lecture Notes in Computer Science*, pages 362–373, 1999.
 - [Kur64] S. Kuroda. Classes of languages and linear-bounded automata. *Information and Control*, 7(2):207–223, 1964.
 - [LS97a] M. Latteux and D. Simplot. Context-sensitive string languages and recognizable picture languages. *Information and Computation*, 138(2):160–169, 1997.
 - [LS97b] M. Latteux and D. Simplot. Recognizable picture languages and domino tiling. *Theoretical Computer Science*, 178(1-2):275–283, 1997.
 - [Mor00] C. Morvan. On rational graphs. In *Foundations of Software Science and Computation Structures, Third International Conference (FoSSaCS 2000)*, volume 1784 of *Lecture Notes in Computer Science*, pages 252–266, 2000.
 - [Mor01] C. Morvan. *Les graphes rationnels*. PhD thesis, IFSIC, Université de Rennes 1, 2001.
 - [MR04] C. Morvan and C. Rispal. Families of automata characterizing context-sensitive languages. *to appear in Acta Informatica*, 2004.
 - [MS01] C. Morvan and C. Stirling. Rational graphs trace context-sensitive languages. In *Mathematical Foundations of Computer Science 2001, 26th International Symposium (MFCS 2001)*, volume 2136 of *Lecture Notes in Computer Science*, pages 548–559, 2001.
 - [Pay00] E. Payet. *Thue Specifications, Infinite Graphs and Synchronized Product*. PhD thesis, Université de la Réunion, 2000.
 - [Pen74] M. Penttonen. One-sided and two-sided context in formal grammars. *Information and Control*, 25(4):371–392, 1974.
 - [Pri00] Ch. Prieur. *Fonctions rationnelles de mots infinis et continuité*. PhD thesis, Université de Paris 7, 2000.
 - [Ris02] C. Rispal. The synchronized graphs trace the context-sensitive languages. In *4th International Workshop on Verification of Infinite-State Systems (Infinity 2002)*, volume 68 of *Electronic Notes in Theoretical Computer Science*, 2002.
 - [Sak03] J. Sakarovitch. *Éléments de théorie des automates*. Éditions Vuibert, 2003.
 - [Tho01] W. Thomas. A short introduction to infinite automata. In *Developments in Language Theory, 5th International Conference (DLT 2001)*, volume 2295 of *Lecture Notes in Computer Science*, pages 130–144, 2001.
 - [vM04] D. van Melkebeek. Time-space lower bounds for NP-complete problems. *Current Trends in Theoretical Computer Science*, pages 265–291, 2004.
 - [Web96] A. Weber. Decomposing a k -valued transducer into k unambiguous ones. *Informatique Théorique et Applications*, 30(5):379–413, 1996.

IRISA – CAMPUS DE BEAULIEU – 35042 RENNES CEDEX – FRANCE
E-mail address: Arnaud.Carayol@irisa.fr

LIAFA – UNIVERSITÉ DE PARIS 7 – 2 PLACE JUSSIEU, CASE 7014, 75251 PARIS CEDEX 05 – FRANCE
E-mail address: Antoine.Meyer@liafa.jussieu.fr