

On Word Rewriting Systems Having a Rational Derivation

Didier Caucal

IRISA–CNRS, Campus de Beaulieu,
35042 Rennes, France
caucal@irisa.fr

Abstract. We define four families of word-rewriting systems: the prefix/suffix systems and the left/right systems. The rewriting of prefix systems generalizes the prefix rewriting of systems: a system is prefix (suffix) if a left hand side and a right hand side are overlapping only by prefix (suffix). The rewriting of right systems generalizes the mechanism of transducers: a system is right (left) if a left hand side overlaps a right hand side only on the right (left).

We show that these systems have a rational derivation even if they are not only finite but recognizable. Besides these four families, we give simple systems having a non rational derivation.

1 Introduction

A general approach to verify properties for systems is to decide whether formulas are verified by their transition graphs: systems with isomorphic transition graphs have the same properties. These graphs are in general infinite but we have a hierarchy of graph families: finite graphs, regular graphs, prefix-recognizable graphs, rational graphs. A family of infinite graphs has been defined in [MS 85]: the connected regular graphs of finite degree meaning that they have a finite number of non isomorphic connected components by decomposition by distance from any vertex. The regular graphs of finite degree are the transition graphs of pushdown automata (restricted to a rational configuration set) and are also the prefix transition graphs of finite word-rewriting systems [Ca 90]: finite unions of elementary graphs of the form $(u \xrightarrow{a} v).W = \{uw \xrightarrow{a} vw \mid w \in W\}$ where u, v are words and W is a rational language. This family has been extended in [Co 90] to all the regular graphs (or equational graphs): the graphs generated by the deterministic graph grammars. A larger family is composed of the prefix-recognizable graphs [Ca 96] which are the prefix transition graphs of the recognizable word-rewriting systems: finite union of elementary graphs of the form $(U \xrightarrow{a} V).W$ where U, V, W are rational languages. Finally, an even larger family of graphs is the set of rational graphs studied in [Mo 00]: the graphs recognized by transducers with labelled outputs. Clearly, all these representations are heterogeneous and a central question is to find a simple and uniform specification for all these graphs.

A solution has been proposed in [CK 98] which considers the ‘Cayley graph’ of any word-rewriting system: the set of transitions $u \xrightarrow{a} v$ if u, v are irreducible words, a is a letter and ua derives into v . To represent as Cayley graphs the regular graphs and the prefix-recognizable graphs, we translate the prefix (resp. suffix) rewriting of systems [Bü 64] into the rewriting of particular systems, called prefix (resp. suffix) systems. A system is called prefix (resp. suffix) if a left hand side and a right hand side are overlapping only by prefix (resp. suffix). To represent as Cayley graphs the rational graphs, we translate the mechanism of transducers into the rewriting of particular systems, called right systems. A system is called right (resp. left) if a left hand side overlaps a right hand side only on the right (resp. left). These systems yield a uniform characterization of all the previous families of graphs.

In this paper, we show that these systems have a rational derivation: derivation relation itself (the reflexive and transitive closure by composition of the rewriting) is recognizable by a transducer (a finite automaton where each label is a couple of words), and we can construct such a transducer in polynomial time. Such a result is general: a rational relation preserves rational and context-free languages, and the composition of rational relations remains rational. Many others properties are well known [Be 79], [AB 88]. Furthermore the derivation is rational when the systems (left, right, prefix, suffix) are not only finite but recognizable (and false for rational systems). Finally, it appears that we can have a non rational derivation for the remaining families of rewriting systems, defined by overlapping between the left hand sides and the right hand sides.

2 Rational and Recognizable Relations

We present notations and basic properties for rational relations and recognizable relations.

For any set E , we denote by $\#E$ its cardinal and by 2^E its powerset. Let \mathbb{N} be the set of nonnegative integers and for any $n \in \mathbb{N}$, let $[n] = \{1, \dots, n\}$ with $[0] = \emptyset$.

A binary (total) *operation* \cdot on a set E is a mapping from $E \times E$ into E and we write $a \cdot b$ instead of $\cdot(a, b)$. A set M with a binary operation \cdot on M is a *monoid* if \cdot is *associative*: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for every $a, b, c \in M$, and has a (unique) neutral element 1 : $a \cdot 1 = 1 \cdot a = a$ for every $a \in M$. The powerset 2^M of M is a monoid for operation \cdot extended by union to subsets: $P \cdot Q = \{a \cdot b \mid a \in P \wedge b \in Q\}$ for every $P, Q \subseteq M$; $\{1\}$ is the neutral element. A subset P of a monoid M is a *submonoid* of M if P is a monoid for \cdot of M : $P \cdot P \subseteq P$ and $1 \in P$. The smallest (for inclusion) submonoid of M containing P and called the *submonoid generated* by P , is the following subset $P^* = \bigcup_{n \geq 0} P^n$ with $P^0 = \{1\}$ and $P^{n+1} = P^n \cdot P$ for every n . The subset P^* is also called the *reflexive and transitive closure* by \cdot of P . Note that $(P^*)^* = P^*$ and $\emptyset^* = \{1\}$.

We say that M is *finitely generated* if $M = P^*$ for some finite P . We say that M is *free* if $M = P^*$ for some *code* P : there is no two factorizations in P^* of a same element *i.e.* if $a_1 \dots a_m = b_1 \dots b_n$ for $a_1, \dots, a_m, b_1, \dots, b_n \in P$ then $m = n$ and $u_i = v_i$ for all i .

We say that M is free finitely generated if $M = P^*$ for some finite code P .

The set $Rat(M)$ of the *rational subsets* of M is the smallest subset of 2^M containing the finite subsets of M and closed by the three operations $\cup, \cdot, *$.

We can also recognize the rational subsets by finite automata.

Let P be a subset of M . A (simple oriented labelled) P -graph G is a subset of $V \times P \times V$ where V is an arbitrary set. Any (s, a, t) of G is a *labelled arc* of *source* s , of *target* t , with *label* a , and is identified with the *labelled transition* $s \xrightarrow[a]{a} t$ or directly $s \xrightarrow{a} t$ if G is understood. We denote by $V_G :=$

$\{ s \mid \exists a \exists t, s \xrightarrow{a} t \vee t \xrightarrow{a} s \}$ the set of *vertices* of G . A graph is *deterministic* if \xrightarrow{a} is a function for every $a \in P$ *i.e.* distinct arcs with the same source have distinct labels: $r \xrightarrow{a} s \wedge r \xrightarrow{a} t \implies s = t$. The set $2^{V \times P^* \times V}$ of P^* -graphs with vertices in V is a monoid for the *composition*: $G \circ H := \{ r \xrightarrow{a \cdot b} t \mid \exists s, r \xrightarrow[a]{a} s \wedge s \xrightarrow[b]{b} t \}$ for any $G, H \subseteq V \times P^* \times V$;

its neutral element is $\{ s \xrightarrow{1} s \mid s \in V \}$ (in fact $2^{V \times P^* \times V}$ is the powerset monoid of the partial semigroup $V \times P^* \times V$ with the partial operation $(r, a, s) \circ (s, b, t) = (r, a \cdot b, t)$).

The relation $\xrightarrow[G^*]{u}$ denoted by $\xrightarrow[G]{u}$ or simply by \xRightarrow{u} if G is understood, is the existence of a *path* in G labelled $u \in P^*$. The labels $L(G, E, F)$ of paths from a set E to a set F is the following subset of P^* : $L(G, E, F) := \{ u \in M \mid \exists s \in E, \exists t \in F, s \xrightarrow[G]{u} t \}$; in particular $1 \in L(G, E, F)$ when $E \cap F \neq \emptyset$.

A P -automaton A is a P -graph G whose vertices are called *states*, with a subset I of *initial states* and a subset F of *final states*; the automaton recognizes the subset $L(A) = L(G, I, F)$ of P^* . An automaton is finite if its graph is finite. An automaton is deterministic if its graph is deterministic and there is a unique initial state. This permits to express a standard result on rational subsets:

Given a subset P of a monoid M , $Rat(P^*)$ is equivalently

- the smallest subset of 2^M containing \emptyset and $\{a\}$ for each $a \in P$, and closed by $\cup, \cdot, *$
- the set of subsets recognized by the finite P -automata
- the set of subsets recognized by the finite deterministic P -automata.

Given monoids M and N , the *cartesian product* $M \times N = \{ (m, n) \mid m \in M \wedge n \in N \}$ is a monoid for the operation defined by $(m, n) \cdot (m', n') = (m \cdot_M m', n \cdot_N n')$ for every $m, m' \in M$ and every $n, n' \in N$. A *relation* R from M into N is a subset of $M \times N$ and we write also $u R v$ for $(u, v) \in R$. In particular R is a *rational relation* if R belongs to $Rat(M \times N)$ *i.e.* R is recognized by a finite (and deterministic) $M \times N$ -automaton: any transition $\xrightarrow{(u,v)}$ is written simply $\xrightarrow{u/v}$. For any relations R, S from M into N , we have $R \cdot S = \{ (m \cdot m', n \cdot n') \mid m R m' \wedge n S n' \}$ and R^* is the reflexive and transitive closure by \cdot of R . As usual, we denote by $R^{-1} = \{ (v, u) \mid u R v \}$ the *inverse* of R

and $R(P) = \{ v \mid \exists u \in P, u R v \}$ is the *image* by R of $P \subseteq M$. In particular $Dom(R) = R^{-1}(N)$ is the *domain* of R and $Im(R) = R(M)$ is the *image* of R . Note that for $R \in Rat(M \times N)$, we have $R^{-1} \in Rat(N \times M)$, $Dom(R) \in Rat(M)$ and $Im(R) \in Rat(N)$.

Note that the family $2^{M \times M}$ of binary relations on M coincides with the set $2^{M \times \{1\} \times M}$ of (unlabelled) \emptyset^* -graphs: (u, v) coincides with the transition $u \xrightarrow{1} v$. So $2^{M \times M}$ is a monoid for the *relational composition* $R \circ S = \{ (u, w) \mid \exists v, u R v \wedge v S w \}$ for every $R, S \subseteq M \times M$ with the neutral element $Id_M = \{ (u, u) \mid u \in M \}$. Furthermore for every $R \subseteq M \times M$, $R^* = \bigcup_{n \geq 0} R^{(n)}$ is the reflexive and transitive closure of R for \circ : $R^{(0)} = Id_M$ and $R^{(n+1)} = R^{(n)} \circ R$.

Another family of subsets of a monoid M are defined by inverse morphism. A mapping h from M into a monoid N is a (monoid) *morphism* if $h(1) = 1$ and $h(a \cdot b) = h(a) \cdot h(b)$ for every $a, b \in M$. A subset P of M is *recognizable* if there exists a morphism h from M into a finite monoid N such that $P = h^{-1}(h(P))$; we denote by $Rec(M)$ the family of recognizable subsets of M .

Recognizable subsets are also recognizable by automata. We say that a P -graph G is (source) *complete* if for every $a \in P$, every vertex $s \in V_G$ is source of an arc labelled a : $\exists t, s \xrightarrow{a} t$. We say also that G is *path-deterministic* if G^* is deterministic: \xrightarrow{u} is a function for every $u \in P^*$ i.e. if $r \xrightarrow{u} s$ and $r \xrightarrow{u} t$ then $s = t$. Given a subset P of a monoid M , $Rec(P^*)$ is the set of subsets recognized by the path-deterministic and complete P -automata having a finite set of states. Another way to characterize a recognizable subset is by residual:

$$P \in Rec(M) \iff \{ u^{-1}P \mid u \in M \} \text{ is finite}$$

where the set $Q^{-1}P = \{ v \mid \exists u \in Q, u \cdot v \in P \}$ is the *left residual* of P by $Q \subseteq M$. We denote also by $PQ^{-1} = \{ u \mid \exists v \in Q, u \cdot v \in P \}$ the *right residual* of P by Q . The characterizations of the rational and recognizable subsets by automata permit to deduce usual facts:

- $Rec(M)$ is a boolean algebra
- $P \cap Q \in Rec(M)$ for every $P \in Rat(M)$ and $Q \in Rec(M)$
- $R(P) \in Rec(N)$ for every $R \in Rat(M \times N)$ and $P \in Rec(M)$
- $Rec(M) \subseteq Rat(M)$ if M is finitely generated (McKnight theorem)
- $Rec(M) = Rat(M)$ if M is free finitely generated (Kleene theorem)
- $R \in Rec(M \times N) \iff R = \bigcup_{i \in I} P_i \times Q_i$ for I finite
with $P_i \in Rec(M), Q_i \in Rec(N)$ (Mezei theorem)

We restrict now to rational and recognizable relations on words. Henceforth N is an *alphabet* i.e. a finite set of symbols called *letters*.

The set $N^* = \{ (a_1, \dots, a_n) \mid n \geq 0 \wedge a_1, \dots, a_n \in N \}$ is a monoid for the *concatenation* operator: $(a_1, \dots, a_m) \cdot (b_1, \dots, b_n) = (a_1, \dots, a_m, b_1, \dots, b_n)$. Any element (a_1, \dots, a_n) is written simply $a_1 \dots a_n$ and called a *word*, and the neutral element $()$ is denoted by ε and called the *empty word*. Note that a word u over N of length $|u| \in \mathbb{N}$ is a mapping from $[|u|]$ into N represented by $u(1) \dots u(|u|) = u$. The *mirror* of any word u is the word $\tilde{u} = u(|u|) \dots u(1)$. A *language* L is a set of words: $L \subseteq N^*$. Let $\tilde{L} = \{ \tilde{u} \mid u \in L \}$ the mirror of any language L , and let $\tilde{R} = \{ (\tilde{u}, \tilde{v}) \mid u R v \}$ the mirror of any binary

relation R on N^* . When L and R are finite, we denote by $|L| = \sum_{u \in L} |u|$ the length of L and by $|R| = \sum_{(u,v) \in R} |u| + |v|$ the length of R ; in particular $|Dom(R)| + |Im(R)| \leq |R|$. Furthermore we denote by $N_L = \{ u(i) \mid u \in L \wedge i \in [|u|] \}$ the alphabet of letters in L , and by $N_R = N_{Dom(R)} \cup N_{Im(R)}$ the alphabet of R . As N^* is the free monoid generated by N , $Rec(N^*) = Rat(N^*)$ is the set of languages recognized by the (deterministic and/or complete) N -automata. A rational relation on N^* , *i.e.* an element of $Rat(N^* \times N^*)$, is a relation recognized by a finite $N^* \times N^*$ -automaton called a *transducer*. Furthermore $R \in Rec(N^* \times N^*)$ if and only if $R = \bigcup_{i \in I} P_i \times Q_i$ for some finite I with $P_i, Q_i \in Rat(N^*)$; in particular $Id_{N^*} \in Rat(N^* \times N^*) - Rec(N^* \times N^*)$. Another remark is that $R(P) \in Rat(N^*)$ for every $R \in Rat(N^* \times N^*)$ and $P \in Rat(N^*)$. A crucial property which is not true for any monoid product is the Elgot-Mezei theorem: $Rat(N^* \times N^*)$ is closed by composition. The family $Rec(N^* \times N^*)$ is also closed by composition, and more generally $R \circ S, S \circ R \in Rec(N^* \times N^*)$ for every $R \in Rat(N^* \times N^*)$ and $S \in Rec(N^* \times N^*)$. Obviously $Rec(N^* \times N^*)$ is closed by mirror, and $Rat(N^* \times N^*)$ is also closed by mirror: for any $N^* \times N^*$ -graph G , we have $L(G, E, F)^\sim = L(\tilde{G}, F, E)$ with $\tilde{G} = \{ q \xrightarrow{\tilde{u}/\tilde{v}} p \mid p \xrightarrow{u/v}_G q \}$.

3 Rational Derivation

We consider the word-rewriting systems (see for instance the survey [DJ 90] and [BO 93]) associated with a language of admissible words such that any derivation between admissible words contains only admissible words. Like in [Sé 93], we define several subclasses of rewriting systems by considering the overlappings between the left hand sides (of the rules) and the right hand sides, inside of the admissible words. We extract two families of systems, the right systems and the prefix systems, having a rational derivation even if the systems are recognizable (Theorems 3.8 and 3.11). By mirror, we obtain two others families of systems, the left systems and the suffix systems. Besides these four families, we give simple systems having a non rational derivation.

A (word) *rewriting system* (R, C) is a binary relation R on N^* and a language $C \subseteq N^*$ of *configurations* (or *admissible words*). A system (R, C) is respectively *finite*, *recognizable*, *rational* if C is rational, and if R is respectively finite, recognizable, rational. The *rewriting* $\xrightarrow{R, C}$ according to any system (R, C) is

$$\xrightarrow{R, C} := \{ (xyy, xvy) \in C \times C \mid u R v \wedge x, y \in N^* \}$$

the application of R under any left and right contexts, but restricted to configurations. Furthermore the *derivation* $\xrightarrow{R, C}^*$ according to (R, C) is

$$\xrightarrow{R, C}^* := \{ (u_0, u_n) \in C \times C \mid n \geq 0 \wedge \exists u_1, \dots, u_{n-1}, u_0 \xrightarrow{R, C} u_1 \dots u_{n-1} \xrightarrow{R, C} u_n \}$$

the reflexive (restricted to C) and transitive closure of $\xrightarrow{R, C}$ by composition *i.e.*

$$\xrightarrow{R, C}^* = \bigcup_n \xrightarrow{R, C}^n \quad \text{where} \quad \xrightarrow{R, C}^0 = Id_C \quad \text{and} \quad \xrightarrow{R, C}^{n+1} = \xrightarrow{R, C}^n \circ \xrightarrow{R, C} \quad \forall n \geq 0.$$

Note that for any system (R, C) ,

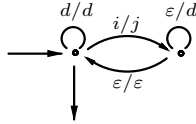
$$\begin{aligned} \xrightarrow{\tilde{R}, \tilde{C}} &= \left(\xrightarrow{R, C}\right)^{\sim} \quad \text{and} \quad \xrightarrow{R^{-1}, C} = \left(\xrightarrow{R, C}\right)^{-1} \\ \text{hence } \xrightarrow{\tilde{R}, \tilde{C}}^* &= \left(\xrightarrow{R, C}^*\right)^{\sim} \quad \text{and} \quad \xrightarrow{R^{-1}, C}^* = \left(\xrightarrow{R, C}^*\right)^{-1}. \end{aligned}$$

When the configuration set $C = N^*$, it can be omitted: we usually say that the relation R is a rewriting system and we denote by \xrightarrow{R} its rewriting (instead of $\xrightarrow{R, N^*}$) and by \xrightarrow{R}^* its derivation. Note that $\xrightarrow{R, C} = \xrightarrow{R} \cap C \times C$.

Even if the configuration set C is rational, it is possible by restriction to C to control the rewriting of simple finite relations in order to get a non rational derivation. This is shown in the following example.

Example 3.1 Consider the finite relation $R = \{(a, bd), (b, c), (c, a)\}$ and the rational configuration set $C = \bigcup_{p \neq q} \in \{a, b, c\} pd^*qd^*$. This finite system has a non rational derivation $\xrightarrow{R, C}^*$ because the language $\xrightarrow{R, C}^*(ab) \cap ad^*bd^* = \{ad^nbd^n \mid n \geq 0\}$ is not rational.

However such a relation R is prefix and is left (\tilde{R} is right) as defined below (cf. Theorems 3.8 and 3.11), and in particular its derivation \xrightarrow{R}^* is rational; it is recognized by the following transducer:



with $(i, j) \in \{(a, a), (a, bd), (a, cd), (b, a), (b, b), (b, c), (c, a), (c, bd), (c, c)\}$.

Thus we introduce a general condition on the systems to be study. A system (R, C) is *stable* if it satisfies the following condition:

$$s \xrightarrow{R}^* r \xrightarrow{R}^* t \wedge s, t \in C \implies r \in C$$

Such a general condition is undecidable but there exists decidable sufficient conditions like the closure of C by rewriting: $\xrightarrow{R}(C) \subseteq C$. In particular any relation R (on N^*) is stable. A basic property of any stable system is that its derivation is the restriction to the configurations of the derivation of its relation.

Lemma 3.2 For any stable system (R, C) , $\xrightarrow{R, C}^* = \xrightarrow{R}^* \cap C \times C$.

For C rational, $C \times C$ is a recognizable relation, and Lemma 3.2 implies that if \xrightarrow{R}^* is a rational (resp. recognizable) relation then $\xrightarrow{R, C}^*$ is a rational (resp. recognizable) relation. However we will give general families of systems (R, C) such that $\xrightarrow{R, C}^*$ is rational, but not containing (R, N^*) in general.

To study the rationality of derivation $\xrightarrow[\text{R}, \text{C}]{*}$ we consider the composition $\xrightarrow[\text{R}, \text{C}]{} \circ \xrightarrow[\text{R}, \text{C}]{} \xrightarrow[\text{R}, \text{C}]{*}$ of two rewritings, and we examine the possible overlappings between the right hand side of the rule applied in the first rewriting, with the left hand side of the rule applied in the second rewriting. We extricate families of systems having a rational derivation by discarding the undesirable overlapping rules.

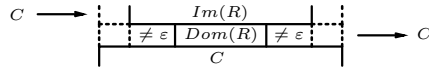
It is easy to find simple finite relations having a non rational derivation.

Example 3.3 For $R = \{(ab, aabb)\}$, $\xrightarrow[\text{R}]{*}(ab) = \{a^n b^n \mid n \geq 1\} \notin \text{Rat}(\{a, b\}^*)$ hence $\xrightarrow[\text{R}]{*}$ is not rational. Similarly the derivation of $R^{-1} = \{(aabb, ab)\}$ is not rational because $\xrightarrow[\text{R}^{-1}]{*} = (\xrightarrow[\text{R}]{*})^{-1}$. These relations are strict-internals as defined below.

We say that a system (R, C) is *domain-strict-internal* if

$$\exists s, t \in N^* \quad \exists (w, xuy), (u, v) \in R, \quad x, y \neq \varepsilon \wedge swt, sxuyt, sxvyt \in C$$

meaning that the following representation is allowed:



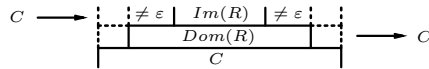
which is decidable for (R, C) rational: $\exists s, t \in N^*$,

$$(R \cap s^{-1}Ct^{-1} \times s^{-1}Ct^{-1}) \circ (Id_{N^+} . R . Id_{N^+} \cap s^{-1}Ct^{-1} \times s^{-1}Ct^{-1}) \neq \emptyset$$

Let us illustrate the significance of the configuration set C for this definition.

Example 3.4 The relation $R = \{(\varepsilon, ab)\}$ is domain-strict-internal and its derivation is not rational because the language $\xrightarrow[\text{R}]{*}(\varepsilon) \cap a^*b^* = \{a^n b^n \mid n \geq 0\}$ is not rational. On the other hand $(R, (ab)^*)$ is not domain-strict-internal (and is stable) and $\xrightarrow[\text{R}, (ab)^*]{*} = Id_{(ab)^*} . (\{\varepsilon\} \times (ab)^*)$ is rational.

Similarly a system (R, C) is *image-strict-internal* if (R^{-1}, C) is domain-strict-internal, meaning that the following representation is allowed:

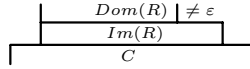


Finally a system is *strict-internal* if it is domain-strict-internal or image-strict-internal.

Another notions of internal systems can be obtained by prefixity and suffixity. A system (R, C) is *domain-prefix-internal* if

$$\exists s, t \in N^* \quad \exists (w, uy), (u, v) \in R, \quad y \neq \varepsilon \wedge swt, suyt, svyt \in C$$

meaning that the following representation is allowed:



which is decidable for (R, C) rational:

$$\exists s, t \in N^*, (R \cap s^{-1}Ct^{-1} \times s^{-1}Ct^{-1}) \circ (R.Id_{N^+} \cap s^{-1}Ct^{-1} \times s^{-1}Ct^{-1}) \neq \emptyset$$

Similarly we say that (R, C) is

- image-prefix-internal* if (R^{-1}, C) is domain-prefix-internal,
- domain-suffix-internal* if (\tilde{R}, \tilde{C}) is domain-prefix-internal,
- image-suffix-internal* if $(\tilde{R}^{-1}, \tilde{C})$ is domain-prefix-internal.

Finally a system is *prefix-internal* if it is domain-prefix-internal or image-prefix-internal. Similarly a system is *suffix-internal* if it is domain-suffix-internal or image-suffix-internal. Furthermore a system is *domain-internal* if it is domain-strict-internal or domain-prefix-internal or domain-suffix-internal. Similarly a system is *image-internal* if it is image-strict-internal or image-prefix-internal or image-suffix-internal.

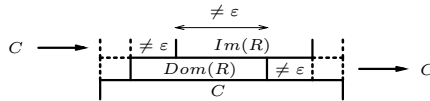
Note that it is again easy to find non-internal relations having a non rational derivation.

Example 3.5 For $R = \{(ba, ab)\}$, the language $\xrightarrow{*}_R((ab)^*) \cap a^*b^*$ is equal to $\{a^n b^n \mid n \geq 0\}$ hence $\xrightarrow{*}_R$ is not rational. Such a relation is together left-overlapping and right-overlapping as defined below.

We say that a system (R, C) is *left-overlapping* if

$$\exists s, t \in N^* \quad \exists (u, yz), (xy, v) \in R, x, y, z \neq \varepsilon \wedge sxut, sxyzt, svzt \in C$$

meaning that the following representation is allowed:



Let us verify that we can decide whether a rational system (R, C) is left-overlapping. Let $\$$ be a new symbol (not in N). So

$$C_{\$} = \{u\$v \mid uv \in C\} = Id_{N^*} \cdot \{(\varepsilon, \$)\} \cdot Id_{N^*}(C)$$

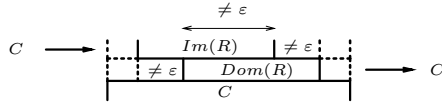
$$\text{and } R_{\$} = \{(u\$v, w) \mid uv R w \wedge v \neq \varepsilon\} = Id_{N^*} \cdot \{(\$, \varepsilon)\} \cdot Id_{N^+} \circ R$$

are rational: $C_{\$} \in Rat((N \cup \{\$\})^*)$ and $R_{\$} \in Rat((N \cup \{\$\})^* \times N^*)$.

Then (R, C) is left-overlapping if and only if it satisfies the following decidable property: $\exists s, t \in N^*$,

$$(Id_{N^+} \cdot (\varepsilon, \$) \cdot R \cap s^{-1}Ct^{-1} \times s^{-1}C_{\$}t^{-1}) \circ (R_{\$} \cdot Id_{N^+} \cap s^{-1}C_{\$}t^{-1} \times s^{-1}Ct^{-1}) \neq \emptyset$$

Similarly a system (R, C) is *right-overlapping* if (R^{-1}, C) is left-overlapping (which is equivalent to (\tilde{R}, \tilde{C}) is left-overlapping), meaning that the following representation is allowed:



Finally a system is *overlapping* if it is left-overlapping or right-overlapping.

We are ready to give stable systems such that the derivation can be done increasingly.

Precisely, we denote by $\xrightarrow{R, C}_n$ the rewriting of (R, C) at letter position $n + 1$:

$$xuy \xrightarrow{R, C}_n xvy \quad \text{for every } u R v \text{ and } xuy, xvy \in C \text{ with } |x| = n.$$

This permits to define the following *increasing derivation* :

$$\xrightarrow{R, C}^* = \bigcup_{n \geq 0} \xrightarrow{R, C}_n$$

where $\xrightarrow{R, C}_0 = Id_C$

and $\xrightarrow{R, C}_n = \bigcup_{\ell_1 \leq \dots \leq \ell_n} \xrightarrow{R, C}_{\ell_1} \circ \dots \circ \xrightarrow{R, C}_{\ell_n}$ for every $n > 0$

with $l_{i-1} = l_i \implies \xrightarrow{R, C}_{\ell_i}$ is only according to $R - \{\varepsilon\} \times N^*$.

This last condition means that the following derivation:

$$xuy \xrightarrow{R}_{|x|} xvy \xrightarrow{R}_{|x|} xwvy \quad \text{with } u R v \text{ and } \varepsilon R w$$

is not increasing. In fact in this derivation, the rule $\varepsilon \rightarrow w$ is on the ‘left’ of the rule $u \rightarrow v$ and must be applied before to give the following increasing derivation:

$$xuy \xrightarrow{R}_{|x|} xwuy \xrightarrow{R}_{|xw|} xwvy \quad (\text{assuming that } w \neq \varepsilon).$$

Lemma 3.2 remains true for increasing derivations.

Lemma 3.6 For any stable system (R, C) , $\xrightarrow{R, C}^* = \xrightarrow{R}^* \cap C \times C$.

The increasing derivation coincides with the derivation for stable systems having no overlapping configurations where the domain begins before the image.

Lemma 3.7 For any stable system (R, C) which is not left-overlapping, not image-strict-internal and not image-suffix-internal, we have $\xrightarrow{R, C}^* = \xrightarrow{R, C}^*$.

Proof.

By definition, we have $\xrightarrow{R, C}^* \subseteq \xrightarrow{R, C}^*$.

Let us prove the converse. As $\xrightarrow{R, C}^* = \xrightarrow{R - \{\varepsilon, \varepsilon\}, C}^*$, we may assume that $(\varepsilon, \varepsilon) \notin R$.

We show the following four inclusions:

$$\begin{aligned} \xrightarrow{u, v, C} \geq n \circ \xrightarrow{\varepsilon, v', C} n &\subseteq \xrightarrow{\varepsilon, v', C} n \circ \xrightarrow{u, v, C} > n \\ \xrightarrow{u, v, C} > n \circ \xrightarrow{u', v', C} n &\subseteq \xrightarrow{u', v', C} n \circ \xrightarrow{u, v, C} \geq n \quad \text{with } u, u' \neq \varepsilon \\ \xrightarrow{\varepsilon, v, C} > n \circ \xrightarrow{u', v', C} n &\subseteq \xrightarrow{u', v', C} n \circ \xrightarrow{\varepsilon, v, C} > n \quad \text{with } u', v' \neq \varepsilon \\ \xrightarrow{\varepsilon, v, C} > n \circ \xrightarrow{u', \varepsilon, C} n &\subseteq \xrightarrow{u', \varepsilon, C} n \circ \xrightarrow{\varepsilon, v, C} > n \cup \xrightarrow{\varepsilon, v, C} n \circ \xrightarrow{u', \varepsilon, C} > n \end{aligned}$$

where $\xrightarrow{u, v, C} P = \bigcup_{n \in P} \xrightarrow{\{(u, v)\}, C} n$ for any integer subset P and with $(u, v) \in R$.

Using these inclusions, we sort increasingly any derivation by applying the bubble sort. \square

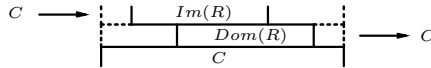
A first class of rewriting systems with a decidable rational derivation is obtained by generalizing the mechanism of a transducer. Let (G, E, F) be a transducer: G is a finite $N^* \times N^*$ -automaton and we assume that its vertex set V_G is disjoint of N . We convert G into the following relation:

$$R_G = \{ (pu, vq) \mid p \xrightarrow[G]{u/v} q \}$$

in such a way that the language recognized by the transducer is obtained by derivation of R_G as follows:

$$L(G, E, F) = \{ (u, v) \mid pu \xrightarrow[R_G]{*} vq \wedge p \in E \wedge q \in F \}$$

Such a system $(R_G, N^*V_GN^*)$ is *right* meaning that it is not strict-internal, not domain-prefix-internal, not image-suffix-internal, and not left-overlapping. So a right system (R, C) is a system where the overlapping configurations have only the following form:



It is important to remark that relation R_G (on $(N \cup V_G)^*$) is not right: we may have the overlapping configuration $puwvq$ with $puw \in \text{Dom}(R_G)$ and $wvq \in \text{Im}(R_G)$.

The derivation $\xrightarrow[R, C]{*}$ of any finite right stable system (R, C) can be always recognized by a transducer that we can construct from (R, C) , and this can be generalized to any recognizable right stable system.

Theorem 3.8 *For any recognizable right stable system (R, C) , the derivation $\xrightarrow[R, C]{*}$ is an effective rational relation.*

Proof.

i) Let us reduce the proof of this theorem to $C = N^*$.

Let (R, C) be a recognizable system: R is recognizable and C is rational.

Furthermore we assume that (R, C) is stable and is right: it is not left-overlapping, not strict-internal, not domain-prefix-internal and not image-suffix-internal.

Let $\$, \#$ be two new symbols: $\$, \# \notin N$. We consider the following system:

$$S = \{ (x\#y, \$v) \mid xy R v \}$$

So $S = \{(\varepsilon, \$)\} \cdot (Id_{N^*} \cdot \{(\#, \varepsilon)\}) \cdot Id_{N^*} \circ R$ is recognizable.

As $S \subseteq N^* \# N^* \times \N^* , S is a right relation (on $(N \cup \{\$, \#\})^*$).

We verify that

$$\xrightarrow[\substack{R, C}{*}] = \{ (h_{\#}(s), h_{\$}(t)) \in C \times C \mid s \xrightarrow[S]{*} t \}$$

with for $\& \in \{\#, \$\}$, $h_{\&}$ is the morphism from $(N \cup \{\&\})^*$ to N^* erasing $\&$:

$$h(\&) = \varepsilon \text{ and } h(a) = a \text{ for every } a \in N.$$

Thus

$$\xrightarrow[\substack{R, C}{*}] = \left((\{(\varepsilon, \#)\} \cup Id_N)^* \circ \xrightarrow[S]{*} \circ (\{(\$, \varepsilon)\} \cup Id_N)^* \right) \cap C \times C$$

implying that $\xrightarrow[\substack{R, C}{*}]$ is rational if $\xrightarrow[S]{*}$ is rational for the recognizable right relation S .

ii) Let R be a finite right relation.

Let us construct from R a transducer to recognize $\xrightarrow[R]{*}$.

Its finite set of states Q is

$$Q = \{ w \mid \exists x, y \in N^*, xw \in Im(R) \wedge wy \in Dom(R) \}$$

which contains ε . Its finite graph G is

$$G = H \cup I$$

$$\text{where } H = \{ w \xrightarrow{x/y} z \mid w, z \in Q \wedge wx R yz \wedge |x|, |y| \text{ minimal} \}$$

$$\text{and } I = \{ w \xrightarrow{x/\varepsilon} wx \mid w, wx \in Q \wedge |x| \geq 1 \text{ minimal} \}$$

$$\cup \{ yz \xrightarrow{\varepsilon/y} z \mid z, yz \in Q \wedge |y| \geq 1 \text{ minimal} \}$$

$$\cup \{ \varepsilon \xrightarrow{a/a} \varepsilon \mid a \in N \}$$

We take ε as the initial state and as the unique final state. We show that the transducer $(G, \{\varepsilon\}, \{\varepsilon\})$ recognizes $\xrightarrow[R]{*}$.

To implement I and as $\#N$ can be large, we use a new symbol \bullet to designate any letter in N and the label \bullet/\bullet means any couple a/a for $a \in N$.

The minimality of $|x|$ and $|y|$ in H and I is useless but it permits to construct a graph isomorphic to G with a (worst case) complexity $O(|R|)$ in space and $O(|R|^2 \#N_R)$ in time *i.e.* linear in space and quadratic in time if we assume that the number of letters is a constant.

iii) Let us extend (ii) to any recognizable right relation R . As R is recognizable,

$$R = \bigcup_{i=1}^p L(G_i, r_i, F_i) \times L(G'_i, r'_i, F'_i)$$

where $(G_i, r_i, F_i)_{1 \leq i \leq p}$ and $(G'_i, r'_i, F'_i)_{1 \leq i \leq p}$ are finite automata such that their vertex sets $V_{G_1}, V_{G'_1}, \dots, V_{G_p}, V_{G'_p}$ are pairwise disjoint.

Let us construct from R a transducer to recognize \xrightarrow{R}^* .

Let r be a new symbol. We define the following finite graph:

$$G = H \cup I$$

$$\text{with } I = \{ r \xrightarrow{a/a} r \mid a \in N \} \cup$$

$$\{ r \xrightarrow{\varepsilon/\varepsilon} r_i \mid i \in [p] \} \cup \{ s' \xrightarrow{\varepsilon/\varepsilon} r \mid \exists i \in [p], s' \in F'_i \} \cup$$

$$\{ s \xrightarrow{a/\varepsilon} t \mid \exists i \in [p], s \xrightarrow{a} t \}_{G_i} \cup \{ s' \xrightarrow{\varepsilon/a} t' \mid \exists i \in [p], s' \xrightarrow{a} t' \}_{G'_i}$$

$$\text{and } H = \{ s \xrightarrow{\varepsilon/\varepsilon} r'_i \mid i \in [p] \wedge s \in F_i \} \cup$$

$$\{ t' \xrightarrow{\varepsilon/\varepsilon} s \mid \exists i, j \in [p], \exists u \neq \varepsilon, u \in L(G'_i, t', F'_i) \cap L(G_j, r_j, s) \}$$

We take r as the initial state and as the unique final state. We show that the transducer (G, r, r) recognizes \xrightarrow{R}^* .

□

Similarly a system (R, C) is *left* if $(\widetilde{R}, \widetilde{C})$ is right: (R, C) is not strict-internal, not domain-suffix-internal, not image-prefix-internal, and not right-overlapping. Note that (R, C) is left if and only if (R^{-1}, C) is right. By Theorem 3.8 and as the rational relations are preserved by mirror (or by inverse), the derivation of any recognizable left stable system is also an effective rational relation. Note that the condition for a right system to be not domain-prefix-internal is necessary to have a rational derivation.

Example 3.9 For $R = \{ (\$, a\&), (\&, \$b) \}$, $\xrightarrow{R}^*(\$) \cap \{ a, b, \$ \}^* = \{ a^n \$ b^n \mid n \geq 0 \}$ hence \xrightarrow{R}^* is not rational. This relation R is not overlapping (not left-overlapping and not right-overlapping), is not strict-internal (not domain-strict-internal and not image-strict-internal), and is not image-internal (not image-strict-internal, not image-prefix-internal and not image-suffix-internal). In particular, its derivation is increasing but the system is not right because it is domain-prefix-internal (and domain-suffix-internal).

A second class of rewriting systems follows from the relation of prefix rewriting. The *prefix rewriting* \xrightarrow{R} of a system R is the restriction of the rewriting \xrightarrow{R} obtained by applying the rules only by prefix:

$$uw \xrightarrow{R} vw \quad \text{for every } u R v \quad \text{and } w \in N^*$$

meaning that the prefix rewriting \xrightarrow{R} is the relation $R.Id_{N^*}$. The *prefix*

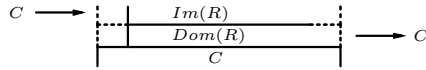
derivation \xrightarrow{R}^* is the reflexive and transitive closure for the composition of the prefix rewriting. Büchi has shown that the prefix derivation $\xrightarrow{R}^*(u)$ of any finite

relation from any word u is a rational language which can be constructed in exponential time [Bü 64]. Boasson and Nivat have extended this result: the prefix derivation $\xrightarrow[\text{R}]{*}$ of any recognizable relation R is a rational relation [BN 84] and a transducer can be constructed in polynomial time [Ca 90]. To adapt this result for rewriting systems, let us remark that for any system R , we have

$$x \xrightarrow[\text{R}]{*} y \iff \$x \xrightarrow[\text{\$R}]{*} \$y$$

where $\text{\$R} = \{ (\$u, \$v) \mid u R v \}$ with $\text{\$}$ a new symbol.

We say that a rewriting system (R, C) is *prefix* if (R, C) is not overlapping (not left-overlapping and not right-overlapping), not strict-internal (not domain-strict-internal and not image-strict-internal), and not suffix-internal (it is not domain-suffix-internal and not image-suffix-internal). So a prefix system is a system where the overlapping configurations have only the following form:



Note that (R, C) is prefix is equivalent to (R^{-1}, C) is prefix.

A usual finite prefix system is a (unlabelled) *pushdown automaton* i.e. a finite $R \subset Q.P \times Q.P^*$ with the language $C = Q.P^*$ of configurations, where $P+Q = N$ (N is partitionned into the stack alphabet P and the state alphabet Q); note that such a R (with $C = N^*$) is also a prefix relation.

The derivation of any prefix stable system is the restriction to the admissible configurations of the concatenation closure of the prefix derivation of its relation.

Proposition 3.10 *For any prefix stable system (R, C) , we have*

$$\xrightarrow[\text{R,C}]{*} = (\xrightarrow[\text{R}]{*})^* \cap C \times C.$$

Proposition 3.10 permits to extend the rationality of the prefix derivation of any recognizable relation to the derivation of any recognizable prefix stable system.

Theorem 3.11 *For any recognizable prefix stable system (R, C) , the derivation $\xrightarrow[\text{R,C}]{*}$ is an effective rational relation.*

Proof.

By Proposition 3.10, it remains to show that $\xrightarrow[\text{R}]{*}$ is a rational relation for any recognizable relation R . This has been proved in [BN 84]. A short proof is due to J.-M. Autebert and follows from Corollary 3.3 of [Ca 96] :

$$\xrightarrow[\text{R}]{*} = \xrightarrow[\text{S}]{*} = S \cdot Id_{N^*} \quad \text{for a recognizable relation } S = \bigcup_{i=1}^q U_i \times V_i$$

such that for \overline{N} a new alphabet in bijection to N , we have

$$\bigcup_{i=1}^q \overline{U}_i \widetilde{V}_i = \xrightarrow[\text{P}]{*}(\{ \overline{u}\widetilde{v} \mid u R v \}^*) \cap \overline{N}^* N^*$$

where $P = \{ (x\overline{x}, \varepsilon) \mid x \in N \}$.

Note that we can specify directly S from R :

$$S = \{(\varepsilon, \varepsilon)\} \cup \bigcup \{ \xrightarrow[R^{-1}]{*}(U) \times \xrightarrow[R]{*}(V) \mid (U, V) \in R \}$$

The construction of S hence of $S.Id_{N^*}$ can be done in polynomial time. But we do not give a precise majoration of the order like for Theorem 3.11 (ii).

□

Similarly a system (R, C) is *suffix* if (\tilde{R}, \tilde{C}) is prefix: (R, C) is not overlapping, not strict-internal and not prefix-internal. By Theorem 3.11 and as the rational relations are preserved by mirror, the derivation of any suffix recognizable stable system is also an effective rational relation.

Example 3.9 shows that we have non rational derivations for prefix finite systems which can be domain-suffix-internal or image-suffix-internal (systems which are not overlapping, not strict-internal, and not domain-suffix-internal or not image-suffix-internal). This includes the basic systems [Sé 93] even if they are not strict-internal, where a *basic* system is a not overlapping and not domain-internal system (the inverse of the system defined in Example 3.9 is basic and not strict-internal).

Furthermore we cannot combine our two theorems as shown below by modifying slightly Example 3.9.

Example 3.12 The overlapping configurations of $R = \{(\$, a\&), (\&b, \$bb)\}$ are only prefix $(N^*\$bbN^*)$ and right $(N^*a\&bN^*)$ but $\xrightarrow[R]{*}$ is not rational because the language $\xrightarrow[R]{*}(\$b) \cap \{a, b, \&\}^* = \{a^n \& b^n \mid n \geq 1\}$ is not rational.

Finally Theorems 3.8 and 3.11 cannot be extended to respectively any rational right stable system and any rational prefix stable system as shown in the following example.

Example 3.13 The relation $R = \{(\$xus, \$usx) \mid u \in \{a, b\}^* \wedge x \in \{a, b\}\}$ is rational and taking $C = \$\{a, b\}^*\$\{a, b\}^*$, the rational system (R, C) has only domain-prefix-internal overlapping configurations. So (R, C) is prefix and is left but

$$\xrightarrow[R, C]{*} \cap (\$\{a, b\}^*\$) \times (\$\$\{a, b\}^*) = \{(\$us, \$\tilde{u}) \mid u \in \{a, b\}^*\}$$

is not a rational relation, hence $\xrightarrow[R, C]{*}$ is not rational.

Acknowledgements

Many thanks to Teodor Knapik for the definition of a right relation and the fact that it preserves the rationality, and for the notion of increasing derivation.

References

- [AB 88] J.-M. AUTEBERT and L. BOASSON *Transductions rationnelles*, Ed. Masson, pp 1-133, 1988.
- [Be 79] J. BERSTEL *Transductions and context-free languages*, Ed. Teubner, pp. 1–278, 1979.
- [BN 84] L. BOASSON and M. NIVAT *Centers of context-free languages*, LITP report 84-44, 1984.
- [BO 93] R. BOOK and F. OTTO *String-rewriting systems*, Text and monographs in computer science, Springer Verlag, 189 pages, 1993.
- [Bü 64] R. BÜCHI *Regular canonical systems*, Archiv für Mathematische Logik und Grundlagenforschung 6, pp. 91–111, 1964
or in *The collected works of J. Richard Büchi*, edited by S. Mac Lane and D. Siefkes, Springer-Verlag, New York, pp. 317–337, 1990.
- [CK 98] H. CALBRIX and T. KNAPIK *A string-rewriting characterization of Muller and Schupp's context-free graphs*, FSTTCS 98, LNCS 1530, pp. 331–342, 1998.
- [Ca 90] D. CAUCAL *On the regular structure of prefix rewriting*, CAAP 90, LNCS 431, pp. 87–102, 1990, selected in TCS 106, pp. 61–86, 1992.
- [Ca 96] D. CAUCAL *On infinite transition graphs having a decidable monadic theory*, ICALP 96, LNCS 1099, pp. 194–205, 1996.
- [Co 90] B. COURCELLE *Graph rewriting: an algebraic and logic approach*, Handbook of TCS, Vol. B, Elsevier, pp. 193–242, 1990.
- [DJ 90] N. DERSHOWITZ and J.-P. JOUANNAUD *Rewrite systems*, Handbook of TCS, Vol. B, Elsevier, pp. 243–320, 1990.
- [Mo 00] C. MORVAN *On rational graphs*, FOSSACS 2000, in this volume.
- [MS 85] D. MULLER and P. SCHUPP *The theory of ends, pushdown automata, and second-order logic*, TCS 37, pp. 51–75, 1985.
- [Sé 93] G. SÉNIZERGUES *Formal languages and word-rewriting*, LNCS 909, pp. 75–94, 1993.