RAIRO-Theor. Inf. Appl. 56 (2022) 8
https://doi.org/10.1051/ita/2022008

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

ON RATIONALLY CONTROLLED ONE-RULE INSERTION SYSTEMS

Michel Latteux and Yves Roos*

Abstract. Rationally controlled one-rule insertion systems are one-rule string rewriting systems for which the rule, that is to insert a given word, may be applied in a word only behind a prefix that must belong to a given rational language called the control language. As for general string rewriting systems, these controlled insertion systems induce a transformation over languages: from a starting word, one can associate all its descendants. In this paper, we investigate the behavior of these systems in terms of preserving the classes of languages: finite, rational and context-free languages. We show that, even for very simple such systems, the images of finite or rational languages need not be context-free. In the case when the control language is in the form u^* for some word u, we characterize one-rule insertion systems that induce a rational transduction and we prove that for these systems, the image of any context-free language is always context-free.

Mathematics Subject Classification. 68Q45, 68Q42, 68R15.

Received April 30, 2020. Accepted September 6, 2022.

1. Introduction

String rewriting systems are of primordial interest for computational problems. Mainly, the problems that are investigated for rewriting systems are the accessibility problem, the common descendant problem, the confluence problem, the termination and uniform termination problem. For several years they have been intensively studied and several deep results have been obtained. However some intriguing decidability problems remain open even for very simple rewriting systems. The most known among these problems is certainly the decidability of the termination of one-rule rewriting systems, a question that remains open for more than thirty years.

Other problems consider rewriting systems as transformation operations on languages: given a rewriting system S and a word w, $S^*(w)$ is the set of all the descendants of w in the rewriting system S. Thus S induces a transformation relation S^* over languages and, from there, one can wonder how these transformations on languages can interact with the classical families of formal languages [17, 18]. In particular, a natural question is the following: given two classes of languages $\mathcal{C}_1$ and $\mathcal{C}_2$, and a family of rewriting systems $\mathcal{F}$, is it true that, for every system S in $\mathcal{F}$ and for every language L in $\mathcal{C}_1$, it holds that $S^*(L)$ is in $\mathcal{C}_2$. If the property is satisfied then rewriting systems in $\mathcal{F}$ are said to be $\mathcal{C}_1/\mathcal{C}_2$ and, in the case when $\mathcal{C}_1 = \mathcal{C}_2$, we rather say that rewriting systems in $\mathcal{F}$ *preserve* $\mathcal{C}_1$.

In this context, some families of rewriting systems have been identified as preserving rational languages ($\mathcal{C}_1 = \mathcal{C}_2 = \mathsf{RAT}$) like *k-period expanding* systems [12], *deleting systems* [9] and *match-bounded* systems [7] or

Keywords and phrases: String rewriting, context-free languages.

CRIStAL, University of Lille, France.

* Corresponding author: yves.roos@univ-lille.fr

preserving context-free languages ($\mathcal{C}_1 = \mathcal{C}_2 = \mathsf{CF}$) like systems *with inhibitor* [14] and *inverse match-bounded* systems [8].

Even for very simple rewriting systems, the question whether these systems are $\mathcal{C}_1/\mathcal{C}_2$ for some given classes of languages $\mathcal{C}_1$ and $\mathcal{C}_2$ is not always so easy to answer. One-rule rewriting systems are among the simplest rewriting systems since they are defined by only two words u, v over an alphabet A. Clearly, one-rule rewriting systems do not always preserve regular languages: the simplest example of such a one-rule (length-preserving) rewriting system is the system $S = \{ba \mapsto ab\}$ but it has been proved RAT/CF in [5], in the context of a particular class of rewriting systems called semi-commutations. From this, one could think that one-rule rewriting systems are at least FIN/CF where FIN is the class of finite languages, but, rather surprisingly, it is not the case: for the one-rule rewriting system $S = \{ba \mapsto a^2b^2\}$ it has been proved that $S(b^2a^2)$ is not a context-free language [10]. Since then, one-rule *grid* rewriting systems, introduced in [6] have been proved FIN/CF in [10].

In [11], we have considered *prefixal* one-rule rewriting systems that are systems in the form $S = \{u \mapsto uf\}$ for some word u and some nonempty word f; in particular, we have proved that theses systems are FIN/CF. One can observe that a prefixal one-rule rewriting systems $S = \{u \mapsto uf\}$ may be seen as a *controlled rewriting system* [3, 4, 17]: indeed, a controlled rewriting system is a rewriting system S equipped with a given language L (the *control language*). In such a system, a rule $l \mapsto r$ may be applied on a word w only behind a prefix of w that must belong to the control language L. When L is the singleton that only contains the empty word, this corresponds to prefix rewriting as defined in [2] and a prefixal system $S = \{u \mapsto uf\}$ corresponds to the controlled system defined by the insertion rule $\{\varepsilon \mapsto f\}$ where ε is the empty word with the control language $L = A^*u$ where A is the used alphabet. Since such a system is completely defined by the word f to insert and the control language L, we denote by $I_{L|f}$ the corresponding system and $I^*_{L|f}$ its associated transformation on languages. In this paper, we focus on one-rule controlled insertion systems $I_{L|f}$ when the control language L is rational.

After some preliminaries given in Section 2, we give some basic properties of controlled one rule insertion systems in Section 3. In particular we characterize when such systems are deterministic, codeterministic or unambiguous. Then we give some sufficient conditions on controlled one rule insertion systems in order to preserve rational or context-free languages in the case when the control language is rational. We finish Section 3 by introducing the notion of *maximal control of a word for insertion* for some alphabet A, denoted by $\mathcal{C}_{\max}(f)$, whose definition is that for all languages L it holds that $I^*_{L|f}(\varepsilon) = I^*_{A^*|f}(\varepsilon)$ if and only if $\mathcal{C}_{\max}(f) \subseteq L$. We show that such a maximal control language effectively exists in the case when the root r of the word f is *unbordered*, that is if no proper suffix of r is a prefix of r; moreover in this case $I^*_{\mathcal{C}_{\max}(f)|f}$ is codeterministic. On the other hand, we prove that when the root of f is bordered, there does not exist any language K such that $I^*_{K|f}(\varepsilon) = I^*_{A^*|f}(\varepsilon)$ with $I^*_{K|f}$ being codeterministic.

In Section 4 we show that, even in the case of a rational control language and even in the case of a single insertion rule, it is possible to define a system $I_{L|f}$ such that $I^*_{L|f}$ is not (FIN/CF). More precisely, we define a system such that for all words w, $I^*_{L|f}(w)$ is not context-free. Moreover, we prove that as soon as a word f contains at least two distinct letters, there exists a rational language R_f such that for all words w, $I^*_{R_f|f}(w)$ is not context-free.

This result motivates Section 5 where we consider rationally one-rule insertion systems $I_{R|f}$ with $f \in a^+$ for some letter a. We prove that such systems are $\mathsf{FIN}/\mathsf{RAT}$ but, rather surprisingly, we also prove that these systems are not RAT/CF by giving two examples where the inserted word f is reduced to a single letter a. For one of these examples, we exhibit a rational language K such that $I^*_{K|a}(b^*)$ is not context-free.

Section 6 is devoted to rationally controlled one-rule insertion systems in the case when the rational control language R is defined as $R = u^*$ for some word u with no constraint on the word to insert f. In particular we characterize when such a system $I_{u^*|f}$ leads to a transformation $I^*_{u^*|f}$ that corresponds to a rational transduction and we prove that these systems preserve context-free languages. To take into account the relations between the words u and f, we study in particular, for a one-rule insertion system $I_{R|f}$ over an alphabet A the following language $\mathcal{R}_{R|f} = \{w \in A^* \mid I^*_{R|f}(w) \cap R \neq \emptyset\}$ that plays a central role in most of the results of this section.

We conclude in Section 7 by some open questions and some perpectives that deserve to be studied.

2. Preliminaries

2.1. Notations

Let A be a finite alphabet, A^* is the free monoid over A and ε is the empty word in A^*. For a word $w \in A^*$, the length of the word w is denoted by $|w|$ and, for any letter $a \in A$, the number of occurrences of the letter a in w is denoted by $|w|_a$.

A word w' is a *factor* of a word w if there exist two words w_1 and w_2 such that $w = w_1 w' w_2$. We denote by $\mathsf{RF}(w)$ (respectively $\mathsf{LF}(w)$) the set of *right factors* (respectively *left factors*) of the word w, that is:

$$\mathsf{RF}(w) = \{w' \in A^* \mid \exists w'' \in A^*, w = w''w'\},$$

$$\mathsf{LF}(w) = \{w' \in A^* \mid \exists w'' \in A^*,\ w = w'w''\}.$$

A word $u \in A^*$ is said to be *primitive* if it is not a proper power of a shorter word, that is $u = v^n$ with $v \in A^*$ implies $n = 1$ and $v = u$. The *root* of a word $u \in A^+$ is the unique primitive word ρ such that $u = \rho^n$ for some natural number n. Observe that, from the definition, the empty word ε is not primitive. A nonempty set $L \subseteq A^*$ is called a *code* if every equation $u_1 u_2 \cdots u_m = v_1 v_2 \cdots v_n$ with $u_i, v_j \in L$ for all i and j implies $n = m$ and $u_i = v_i$ for all i. A language L is *prefix* if it satisfies the condition: for all words $w, w' \in L$, if $w = w'w''$ for some word w'' then $w'' = \varepsilon$. Clearly, a prefix set is a code.

We denote by FIN the family of finite languages, by RAT the family of rational languages and by CF the family of context-free languages. Abusing notations we identify a rational language with the rational expressions that describe it for instance $(a+b)^*$ with $\{a,b\}^*$ or $\varepsilon + ab$ with $\{\varepsilon, ab\}$.

We denote by ${D'_1}^*$ the Dyck language over $\{a,b\}$, that is the language ${D'_1}^* = \{w \in (a+b)^* \mid |w|_a = |w_b| \wedge \forall x \in \mathsf{LF}(w), |x|_a \geq |x|_b\}$; it is the set of well balanced words of $(a+b)^*$ where a is seen as an open parenthesis and b is the corresponding closing one. We also denote by D_1^* the language $D_1^* = \{w \in (a+b)^* \mid |w|_a = |w|_b\}$.

A *rational relation* from A^* to B^* for some alphabets A and B is a rational subset of $A^* \times B^*$. A *rational transduction* $\tau : A^* \mapsto B^*$, is a mapping from A^* to 2^{B^*} such that its graph is a rational relation from A^* to B^*. A well know result of [15] states that any rational transduction is equivalent to the composition of an inverse morphism, an intersection with a rational language and a morphism. The reader can refer to [1] for detailed informations on rational transductions.

A language L *rationally dominates* a language L', denoted by $L \rightsquigarrow_{\text{rat}} L'$, if there exists a rational transduction τ such that $L' = \tau(L)$. When $L \rightsquigarrow_{\text{rat}} L'$ and $L' \rightsquigarrow_{\text{rat}} L$ the two languages are said to be *rationally equivalent*, that is denoted by $L \underset{\text{rat}}{\equiv} L'$.

A rewriting system over an alphabet A is a subset $S \subseteq A^* \times A^*$. Members of S are denoted by $u \mapsto v$. One-step derivation, denoted by $\rightarrow$, is the binary relation over words defined by : for all words w, w' in $A^*, w \rightarrow w'$ iff there exists $u \mapsto v \in S$ and $\alpha, \beta \in A^*$ such that $w = \alpha u \beta$ and $w' = \alpha v \beta$. The relation $\xrightarrow{*}$, called *derivation relation*, is the reflexive and transitive closure of the relation $\rightarrow$ and we denote by $\xrightarrow{+}$ the transitive closure of the relation $\rightarrow$. For a derivation $w = w_0 \rightarrow w_1 \cdots \rightarrow w_n = w'$, n is called the length of the derivation. Observe that we consider in this paper that a derivation $w = w_0 \rightarrow w_1 \ldots \rightarrow w_n = w'$ is completely characterized by the list of words $[w_0, \ldots, w_n]$ independently from the indexes where the rule is applied at each step of rewriting.

A rewriting system S induces a transformation over languages: for every word $w \in A^*$, we shall denote by $S^*(w)$ the set $S^*(w) = \{w' \in A^* \mid w \xrightarrow{*} w'\}$ and $S^+(w)$ the set $S^+(w) = \{w' \in A^* \mid w \xrightarrow{+} w'\}$; then, for every language $L \subseteq A^*$, $S^*(L) = \bigcup_{w \in L} S^*(w)$ and $S^+(L) = \bigcup_{w \in L} S^+(w)$. We say that a language L is *closed by* a rewriting system S if for all words $w \in L$ it holds that $S^*(w) \subseteq L$.

A rewriting system S over an alphabet A is *confluent* if for all words $w \in A^*$, $u, v \in S^*(w)$ it holds that $S^*(u) \cap S^*(v) \neq \emptyset$.

Given two classes of languages $\mathcal{C}_1$ and $\mathcal{C}_2$, and a family of rewriting systems $\mathcal{F}$, one can wonder whether it holds that $S^*(L)$ is in $\mathcal{C}_2$ for every system S in $\mathcal{F}$ and for every language L in $\mathcal{C}_1$. If the property is satisfied then the rewriting systems in $\mathcal{F}$ are said to be $\mathcal{C}_1/\mathcal{C}_2$ and, in the case when $\mathcal{C}_1 = \mathcal{C}_2$, we rather say that the rewriting systems in $\mathcal{F}$ *preserve* $\mathcal{C}_1$.

We say that a string rewriting system S is *rational* if the relation $\{(w, w') \mid w' \in S^*(w)\}$ is a rational relation. In this case, S^* is a rational transduction and preserves RAT and CF.

3. Controlled insertion systems

3.1. Definitions

To all pair $\langle L, f\rangle$ where $f \in A^+$ is a nonempty word and $L \subseteq A^*$ is a nonempty language called the *control language* is associated a *controlled one-rule insertion system* that is a binary relation over A^*, denoted by $I_{L|f}$ and defined by $(w, w') \in I_{L|f}$ if $w = w_1w_2$, $w' = w_1fw_2$ for some words w_1, w_2 with $w_1 \in L$. In this paper, we shall identify a controlled one-rule insertion system with its associated relation $I_{L|f}$ and, as said before, we shall consider the reflexive and transitive closure $I^*_{L|f}$ of the relation $I_{L|f}$ as a transformation over languages: for all words $w \in A^*, I^*_{L|f}(w) = \{w' \mid (w, w') \in I^*_{L|f}\}$ and for all languages $K \subseteq A^*, I^*_{L|f}(K) = \bigcup_{w\in K} I^*_{L|f}(w)$. So, abusing notations, $I^*_{L|f}$ will represent both this transformation on languages and the set of couple of words that are in relation. We shall often use the following *string rewriting like* alternative notation: $w \xrightarrow[I_{L|f}]{} w'$ for $(w, w') \in I_{L|f}$, $w \xrightarrow[I_{L|f}]{*} w'$ for $w' \in I^*_{L|f}(w)$ and $w \xrightarrow[I_{L|f}]{n} w'$ when the derivation has length n. When there is no ambiguity on the insertion system that is used, we will simply denote by $w \to w'$, $w \xrightarrow{*} w'$ and $w \xrightarrow{n} w'$.

Example 3.1. Let $R = \varepsilon + aba$ and $f = ab$ then using the fact that $(aba, abaab) \in I_{R|ab}$ and $(aba, ababa) \in I_{R|ab}$, it can be proved that $I^*_{R|ab}(\varepsilon) = \varepsilon + ab + aba(ab + ba)^*b$.

Example 3.2. Let $R = a^*$ and $f = ab$, then $I^*_{a^*|ab}(\varepsilon) = {D'_1}^*$: indeed, $I^*_{a^*|ab}(\varepsilon)$ is clearly included in $I^*_{(a+b)^*|ab}(\varepsilon) = {D'_1}^*$. Conversely, let $w \in {D'_1}^*$, we can prove by induction on $|w|$, the length of w, that $w \in I^*_{a^*|ab}(\varepsilon)$: indeed, it is clearly true for $w = \varepsilon$. If $w \neq \varepsilon$, then $w = a^iabw'$ for some natural number i and some word w' with $a^iw' \in {D'_1}^*$. From the inductive hypothesis, we get $a^iw' \in I^*_{a^*|ab}(\varepsilon)$ which implies $w = a^iabw' \in I^*_{a^*|ab}(\varepsilon)$.

We observe that for all words w in $I^*_{a^*|ab}(\varepsilon)$ there exists a unique derivation from ε to w. We shall name below this property *unambiguity* and this statement will be generalized in Proposition 6.14 to $I^*_{u^*|f}$ for all words u and all word $f \neq \varepsilon$.

3.2. Basic properties

Some of the following properties, stated in Proposition 3.4, that are satisfied by controlled insertion systems are clear and do not need a proof, nevertheless they deserve to be mentioned. In these statements, L and L' are any nonempty languages; f and f' are any nonempty words and w, w' are any words all defined over an alphabet A. We shall use the following properties which are consequences of the Schützenberger-Lyndon Theorem [13]:

Lemma 3.3. *Let u and v be words of A^*.*

- *if $uv = vu$ then there exists some (primitive) word r such that $u \in r^*$ and $v \in r^*$.*
- *If r is a primitive word and $urv \in r^*$ then $u \in r^*$ and $v \in r^*$.*

Proposition 3.4.

1. $I_{L|f} = I^*_{L|f} \cap \{(w, w') \mid |w'| = |w| + |f|\}$.
2. *If* $I^*_{L|f} = I^*_{L'|f'}$ *then* $f = f'$.
3. *If* $L \subseteq L'$ *then* $I^*_{L|f} \subseteq I^*_{L'|f}$.

4. $I^*_{L|f} = I^*_{L'|f}$ *if and only if* $L_0 \subseteq L' \subseteq L_0 r^*$ *where* r *is the root of* f *and* $L_0 = L \setminus Lr^+$.
5. $L \subsetneq L'$ *does not imply* $I^*_{L|f} \subsetneq I^*_{L'|f'}$.
6. *If* $(w, w') \in I^*_{L|f}$ *then for all words* α *it holds that* $(w\alpha, w'\alpha) \in I^*_{L|f}$.
7. $(w\alpha, w'\alpha) \in I^*_{L|f}$ *for some nonempty word* α *does not imply* $(w, w') \in I^*_{L|f}$.
8. $(w, w') \in I^*_{L|f}$ *does not imply that for all words* α, $(\alpha w, \alpha w') \in I^*_{L|f}$.
9. *If* $(w, w') \in I^*_{L^*|f}$ *then for all words* $\alpha \in L^*$, *it holds that* $(\alpha w, \alpha w') \in I^*_{L^*|f}$.
10. *If* $LA^* \cap L'A^* = \emptyset$ *then* $I^*_{L\cup L'|f} = I^*_{L|f} \cup I^*_{L'|f}$.
11. *For all* $g \in I^*_{L^*|f}(f)$ *it holds that* $I^*_{L^*|g} \subseteq I^*_{L^*|f}$

Proof.

1. Obviously true.
2. Let w be the shortest word in $L \cup L'$ and assume $w \in L$. We get $(w, wf) \in I^*_{L|f} = I^*_{L'|f'}$. Since w is the shortest word in $L \cup L'$, there is no proper left factor of w in L', so we get $w \in L'$ and $(w, wf') \in I^*_{L'|f'}$. If we assume $|f| \neq |f'|$, this would lead to the contradiction $I^*_{L|f}(w) \neq I^*_{L'|f'}(w)$ so $f' = f$.
3. Obviously true.
4. The condition is necessary: assume $I^*_{L|f} = I^*_{L'|f}$ and let $w \in L'$. Then $(w, wf) \in I^*_{L'|f} = I^*_{L|f}$ and from this follows $w = w_1 w_2$ for some words w_1 and w_2 with $w_1 \in L$ and $w_1 f w_2 = w_1 w_2 f$ which implies $w_2 \in r^*$ from Lemma 3.3. Since $w_1 \in L$, $w_1 r^* \subseteq L_0 r^*$ so $w = w_1 w_2 \in L_0 r^*$. Let now $w \in L_0$ then $(w, wf) \in I^*_{L|f} = I^*_{L'|f}$ and we get $w = v_1 v_2$ for some words v_1 and v_2 with $v_1 \in L'$ and $v_2 \in r^*$. Since $L' \subseteq L_0 r^*$, we get $v_1 = v_1' v_1''$ with $v_1' \in L_0$ and $v_1'' \in r^*$. Now, $w = v_1' v_1'' v_2 \in L_0$ implies $v_1'' = v_2 = \varepsilon$ so $w = v_1 \in L'$.
   The condition is sufficient: assume $L_0 \subseteq L' \subseteq L_0 r^*$, then $I^*_{L_0|f} \subseteq I^*_{L'|f} \subseteq I^*_{L_0 r^*|f}$ from 3. We shall prove $I^*_{L_0 r^*|f} \subseteq I^*_{L_0|f}$ that will imply $I^*_{L_0 r^*|f} = I^*_{L_0|f} = I^*_{L'|f} = I^*_{L|f}$; clearly, it is sufficient to prove $I_{L_0 r^*|f} \subseteq I_{L_0|f}$. Let $(w, w') \in I_{L_0 r^*|f}$ then $w = w_1 r^i w_2$ for some $w_1 \in L_0$ and $w' = w_1 r^i f w_2$. From this follows $w' = w_1 f r^i w_2$ so $(w, w') \in I_{L_0|f}$.
5. It is in fact a consequence of 4 in the case when $L_0 \subsetneq L_0 r^*$: for instance, let $A = \{a\}$, $L = \varepsilon$ and $f = a$, then $L_0 = \varepsilon$, $L_0 r^* = a^*$ and $I^*_{\varepsilon|a} = I^*_{a^*|a}$.
6. Obviously true.
7. Let $A = \{a\}$, $L = a$ and $f = a$. We get $(a, aa) \in I^*_{a|a}$ but $(\varepsilon, a) \notin I^*_{a|a}$.
8. Let $A = \{a, b\}$, $L = \varepsilon$ and $f = ab$. We get $(\varepsilon, ab) \in I^*_{\varepsilon|ab}$ but $(b, bab) \notin I^*_{\varepsilon|ab}$.
9. Since $L^* = L^* L^*$, we directly get that if $(w, w') \in I_{L^*|f}$ then for all words $\alpha \in L^*$, it holds that $(\alpha w, \alpha w') \in I_{L^*|f}$ which easily implies 9 by induction.
10. Obviously true.
11. Since $(\varepsilon, g) \in I^*_{L^*|f}$, it follows from 9 that for all words $w \in L^*$, it holds that $(w, wg) \in I^*_{L^*|f}$. This clearly implies $I_{L^*|g} \subseteq I^*_{L^*|f}$.

□

3.3. Determinism and ambiguity

In the following, L is any nonempty language and f is any nonempty word both defined over an alphabet A.

Definition 3.5. $I^*_{L|f}$ is *deterministic* if for all words $w \in A^*$ there exists at most one word w' such that $(w, w') \in I_{L|f}$.

And we can state:

Lemma 3.6. $I^*_{L|f}$ *is* deterministic *if and only if* $L^{-1}L \subseteq r^*$ *where* r *is the root of* f.

Proof. There exist $(w, w_1) \in I_{L|f}$ and $(w, w_2) \in I_{L|f}$ for some distinct words w_1 and w_2 if and only if $w = u_1 v_1 = u_2 v_2$ with $u_1, u_2 \in L$ and $u_1 f v_1 \neq u_2 f v_2$ (so $u_1 \neq u_2$). Assume $|u_2| < |u_1|$; from this follows $u_1 = u_2 \alpha$ and $\alpha v_1 = v_2$ for some word $\alpha \neq \varepsilon$. If we assume $\alpha f = f \alpha$, then we get $u_1 f v_1 = u_2 \alpha f v_1 = u_2 f \alpha v_1 = u_2 f v_2$, a contradiction. So these words u_1 and u_2 exist if and only if there exists some $\alpha \in L^{-1}L$ with $\alpha \notin r^*$ since $\alpha f \neq f\alpha$. □

Definition 3.7. $I^*_{L|f}$ is *codeterministic* if for all words $w' \in A^*$ there exists at most one word w such that $(w, w') \in I_{L|f}$.

With a similar proof to the proof of Lemma 3.6, we can state:

Lemma 3.8. $I^*_{L|f}$ *is* codeterministic *if and only if* $(Lf)^{-1}(Lf) \subseteq r^*$ *where* r *is the root of* f.

Definition 3.9. $I^*_{L|f}$ is *unambiguous* if for all $(w, w') \in I^*_{L|f}$ there exists a unique derivation $w = w_0 \xrightarrow[I_{L|f}]{} w_1 \xrightarrow[I_{L|f}]{} \dots \xrightarrow[I_{L|f}]{} w_n = w'$.

Proposition 3.10.

1. *If L is prefix then $I^*_{L|f}$ is deterministic; the converse does not hold.*
2. *If $I^*_{L|f}$ is deterministic then it is codeterministic; the converse does not hold.*
3. *If $I^*_{L|f}$ is codeterministic then it is unambiguous; the converse does not hold.*
4. *$I^*_{L|f}$ is unambiguous and confluent if and only if it is deterministic.*

Proof.
1. The implication L prefix then $I^*_{L|f}$ deterministic is obvious. On the other hand $I^*_{a^*|a}$ is deterministic but a^* is not a prefix language.
2. If $I^*_{L|f}$ is deterministic then $L^{-1}L \subseteq r^*$ from Lemma 3.6. That implies $(Lf)^{-1}(Lf) \subseteq r^*$: indeed let $w \in (Lf)^{-1}(Lf)$ then there exist $w_1, w_2 \in L$ such that $w_1 f = w_2 f w$. Since $L^{-1}L \subseteq r^*$, we get $w_1 = w_2 r^i$ for some natural number i. That implies $w_2 r^i f = w_2 f w$ so $fw = r^i f$ and we get $w = r^i$ in r^*. On the other hand, we have seen before that $I^*_{a^*|ab}$ is codeterministic and it is clearly nondeterministic.
3. The implication codeterministic implies unambiguous is clear. Conversely let $L = \varepsilon + ab$. Then $I^*_{L|a}$ is not codeterministic since $(ba, aba) \in I_{L|a}$ and $(ab, aba) \in I_{L|a}$ but it is unambiguous: indeed let w and w' be two words such that $(w, abw') \in I^*_{L|a}$ and $(w, baw') \in I^*_{L|a}$ which could lead to a situation of ambiguity for $(w, abaw')$. But $(w, baw') \in I^*_{L|a}$ implies $w = baw'$ that leads to a contradiction because $abw' \notin I^*_{L|a}(baw')$.
4. If $I^*_{L|f}$ is deterministic then it is confluent, and it is unambiguous from 2 and 3. Conversely, assume that there exist some words w and $w_1 \neq w_2$ such that $w_1, w_2 \in I_{L|f}(w)$. Since $I^*_{L|f}$ is confluent, $I_{L|f}(w_1) \cap I_{L|f}(w_2) \neq \emptyset$ so $I^*_{L|f}$ is ambiguous.

□

3.4. Rationally controlled insertion systems

When L is a rational language, the one-rule insertion system $I_{L|f}$ is said to be rationally controlled.
In the following proposition, f is any nonempty word.

Proposition 3.11.

1. *If F is a finite language then $I^*_{F|f}$ is rational.*
2. *If R is a rational language and $I^*_{R|f}$ is deterministic then $I^*_{R|f}$ is rational.*
3. *If R is a rational language that satisfies $R = RA^*$ then $I^*_{R|f}$ preserves the context-free languages.*
4. *If $R = A^*u$ for some word u then $I^*_{R|f}$ is* FIN/CF.

Proof.

1. If F is finite then $I_{F|f}$ is equivalent to the prefix rewriting system associated with the set of rules $\{u \mapsto uf \mid u \in F\}$. A prefix rewriting system S is a rewriting system where the rewriting rules can only be applied on left factors of the words: $w \xrightarrow[S]{} w'$ if $w = u\alpha$ and $w' = v\alpha$ for some rule $u \mapsto v \in S$ and some word α. The relations associated with prefix rewriting systems have been proved rational in [2].

2. Observe first that, for all languages L and K, it holds that $I^*_{L|f}(K) = (K \setminus LA^*) \cup I^*_{L|f}(K \cap LA^*)$, so it is sufficient to prove that $I^*_{R|f} \cap (RA^* \times A^*)$ is rational. If $I^*_{R|f}$ is deterministic then $R^{-1}R \subseteq r^*$ where r is the root of f from Lemma 3.6. That implies that for all words $w \in RA^*$, it holds that $I^*_{R|f}(w) = w_1 I^*_{\varepsilon|f}(w_2) = w_1 f^* w_2$ where $w = w_1 w_2$ with w_1 being the shortest left factor of w that belongs to R. Indeed, if $w_2 = w_2' w_2''$ with $w_1 w_2' \in R$, we get $w_2' \in r^*$ which implies $w_2' f = f w_2'$. So $I^*_{R|f}(w)$ can be obtained as $s \circ g(w)$ with $g(w) = w_1 \# w_2$ where $\#$ is a fresh letter, *i.e.* a new letter that does not belong to A, and $s : A \cup \{\#\} \mapsto A^*$ is the rational substitution defined by $s(a) = a$ for all $a \in A$ and $s(\#) = f^*$. Since g is a rational function and s is a rational substitution, we get that $I^*_{R|f} = s \circ g$ is a rational transduction.
3. With the same observations as in the proof of 2, we focus on words in RA^*: for all words $w \in RA^*$, it holds that $I^*_{R|f}(w) = w_1 I^*_{A^*|f}(w_2)$ where $w = w_1 w_2$ with w_1 being the shortest left factor of w that belongs to R. Moreover, $I^*_{A^*|f}(w_2) = L_f x_1 L_f \cdots L_f x_n L_f$ where $w_2 = x_1 \cdots x_n, x_i \in A$ and $L_f = I^*_{A^*|f}(\varepsilon)$. From this follows $I^*_{R|f}(w)$ can be obtained as $I^*_{R|f}(w) = s' \circ g'(w)$ where $g'(w) = w_1 \# g''(w_2)$ with $g'' : A^* \mapsto (A \cup \{\#\})^*$ being the morphism defined as $g''(a) = a\#$ for all $a \in A$, and $s' : A \cup \{\#\} \mapsto A^*$ being the substitution defined by $s'(a) = a$ for all $a \in A$ and $s'(\#) = L_f$. Since g'' is a morphism, we get that g' is a rational function. Moreover L_f is a context-free language since it can clearly be generated by a context-free grammar. That implies that s' is a context-free substitution so $I^*_{R|f} = s' \circ g'$ preserves the context-free languages.
4. This statement has been proved in [11] (Prop. 11).

□

We observe that if a rational language R is prefix then for all words f, $I^*_{R|f}$ is deterministic hence is rational from Item 2 of Proposition 3.11.

In the remainder of this section, we shall address the problem to know, given a control language L and a word f, whether the language $I^*_{L|f}(\varepsilon)$ is in fact an *uncontrolled* language, that is whether $I^*_{L|f}(\varepsilon) = I^*_{A^*|f}(\varepsilon)$. To make the notation less cluttered, we will use in the following: $L_{K|f} = I^*_{K|f}(\varepsilon)$ for all languages $K \neq A^*$ and $L_f = I^*_{A^*|f}(\varepsilon)$. Let us also denote $\mathsf{LF}(f) \setminus r^*$ by F where r is the root of f. Observe that, though F^* is not always closed by left factor, *i.e.* generally, $F^* \neq \mathsf{LF}(F^*)$, it holds that $\mathsf{LF}(F^*) \setminus A^*r \subseteq F^*$: indeed we can first observe that $F = \mathsf{LF}(f) \setminus A^*r = \mathsf{LF}(F) \setminus A^*r$ since $f \in r^*$. On the other hand, it holds that $\mathsf{LF}(F^*) \setminus A^*r = (F^*\mathsf{LF}(F)) \setminus A^*r$; moreover r is a primitive word so we get $(F^*\mathsf{LF}(F)) \setminus A^*r \subseteq F^*(\mathsf{LF}(F) \setminus A^*r) \subseteq F^*$.

We can also state the following equality:

Proposition 3.12. $L_f = L_{F^*|f}$.

Proof. $L_{F^*|f} \subseteq L_f$ from Item 3 of Proposition 3.4. Conversely, we prove that for all words $w \in L_f$ it holds that $w \in L_{F^*|f}$ by induction on $|w|$, the length of the word w. If $w = \varepsilon$ or $w = f$, it is clearly true. Consider now $w = w_1 f w_2$ with w_1, w_2 in A^* and $w_1 w_2 \in L_f \cap A^+$. From the inductive hypothesis, $w_1 w_2 = \alpha f \beta$ with $\alpha \in F^*$ and $\alpha\beta \in L_{F^*|f}$. Let us consider two cases:

1. $|w_1| < |\alpha f|$. Set $w_1 = w_1' r^k$ such that $w_1' \notin A^*r$. From this follows $w_1' \in F^*$ so $w_1 f w_2 = w_1' r^k f w_2 = w_1' f r^k w_2 \in L_{F^*|f}$.
2. $|w_1| \geq |\alpha f|$. Then $w_1 = \alpha f \beta'$ and $\beta = \beta' w_2$ for some word β'. Since $\alpha\beta' w_2 = \alpha\beta \in L_{F^*|f}$, we get $\alpha\beta' f w_2 \in L_f$. Moreover $|\alpha\beta' f w_2| = |w_1 w_2|$, so we may apply the inductive hypothesis and we get $\alpha\beta' f w_2 \in L_{F^*|f}$ which implies $\alpha f \beta' f w_2 = w \in L_{F^*|f}$.

□

Observe that the property $I^*_{a^*|ab} = I^*_{(a+b)^*|ab} = {D_1'}^*$ that has been proved in Example 3.2 is, as a matter of fact, a consequence of Proposition 3.12.

This property states that it is always possible to get some word of L_f from ε with insertions of the word f only behind some left factor in F^*. More generally, we introduce the following notion of maximal control of a word f for insertion:

Definition 3.13. A word f *possesses a maximal control for insertion*, denoted by $\mathcal{C}_{\mathsf{max}}(f)$ if for all languages K it holds that $L_{K|f} = L_f$ if and only if $\mathcal{C}_{\mathsf{max}}(f) \subseteq K$.

It is an open problem to know whether every word possesses a maximal control for insertion. If it is not the case, is it possible to decide whether a given word possesses a maximal control for insertion? When a word f possesses a maximal control for insertion $\mathcal{C}_{\mathsf{max}}(f)$, is $\mathcal{C}_{\mathsf{max}}(f)$ always a rational language? We do not know the answers of these questions. However, we shall prove in Proposition 3.17 the existence of $\mathcal{C}_{\mathsf{max}}(f)$ when there are no overlaps between distinct occurrences of the root of f, that is when the root of f is unbordered.

Definition 3.14. A word $w \in A^*$ is called *bordered* if $(\mathsf{LF}(w) \cap \mathsf{RF}(w)) \setminus (\varepsilon + w) \neq \emptyset$ else it is *unbordered*.

In other words w is unbordered if no proper right factor of w is a left factor of w.

We can first state:

Lemma 3.15. *Let r be the root of f. If r is bordered then $F^* = \mathsf{LF}(f)^*$ else $F^* = \mathsf{LF}(f)^* \setminus (A^*fA^* \cup A^*r)$.*

Proof. Assume r to be bordered. We can factorize r as $r = r'mr'$ for some words r' and m with $mr' \neq r'm$ since r is a bordered primitive word. From this follows $r'm \in F$ and $r' \in F$ so $r \in F^*$. That implies $\mathsf{LF}(f) \subseteq F^*$ and we get $F^* = \mathsf{LF}(f)^*$.

Assume now r to be unbordered and suppose $r \in \mathsf{RF}(F)F^*$. From r primitive follows $r \notin \mathsf{RF}(F)$: indeed if $r \in \mathsf{RF}(F)$ then $f = r'rr''$ for some words r' and r'' and this implies $r'r \in r^*$ from Lemma 3.3, a contradiction. From $r \notin \mathsf{RF}(F)$ follows that there exists some word $z \in F \cap \mathsf{RF}(r)$. Moreover $z \in \mathsf{LF}(r)$ since $z \in \mathsf{LF}(f)$ and $|z| < |r|$. This leads to the contradiction that r is bordered so $F^* \cap A^*r = \emptyset$. From this, we can deduce $F^* \cap A^*fA^* = \emptyset$. Indeed assume $f \in \mathsf{RF}(F)F^*\mathsf{LF}(F)$ and set $f = xy$ with $x \in \mathsf{RF}(F)F^*$ and $y \in \mathsf{LF}(F)$. Observe that $x \neq \varepsilon$ since $|y| < |f|$. From $F^* \cap A^*r = \emptyset$ follows $y \neq \varepsilon$ and $y \notin F$ so $y \in r^+$ but that implies $x \in r^+$ that also contradicts $F^* \cap A^*r = \emptyset$. The two properties $F^* \cap A^*r = \emptyset$ and $F^* \cap A^*fA^* = \emptyset$ directly imply $F^* \subseteq \mathsf{LF}(f)^* \setminus (A^*fA^* \cup A^*r)$.

Conversely, let $w \in \mathsf{LF}(f)^* \setminus (A^*fA^* \cup A^*r)$. We prove that $w \in F^*$ by induction on $|w|$, the length of the word w. If $w = \varepsilon$ then $w \in F^*$ else set $w = \alpha\beta$ such that β is the longest word in $\mathsf{LF}(f) \cap \mathsf{RF}(w)$. Observe that $\alpha \in \mathsf{LF}(f)^*$ and $\beta \neq \varepsilon$. That implies $\beta \in F$ else it would follow $\beta \in r^+$ that contradicts $w \notin A^*r$. On the other hand, $\alpha \notin A^*fA^*$ and, if we assume $\alpha = \alpha'r$ for some word α', it would follow $f \in \mathsf{LF}(r\beta)$ from the definition of β that contradicts $w \notin A^*fA^*$. From the inductive hypothesis, we get $\alpha \in F^*$ so $w = \alpha\beta \in F^*$. □

Lemma 3.16. *Let r be the root of f then L_f is closed by the rewriting system $S = \{f \mapsto \varepsilon\}$ if and only if r is unbordered.*

Proof. Assume that r is unbordered, we shall prove that for all words $w = w_1fw_2 \in L_f$ it holds that $w_1w_2 \in L_f$ by induction on $|w|$, the length of the word w. If $w = f$ then $w_1w_2 = \varepsilon \in L_f$. Else, since the occurrence of f that is highlighted in the factorisation $w = w_1fw_2$ need not be obtained by insertion, we consider a factorisation $w = \alpha f\beta$ for some words α and β with $\alpha\beta \neq \varepsilon$ and $\alpha\beta \in L_f$. Let us distinguish four cases:

1. $w_1 = \alpha f\beta'$ and $\beta = \beta' fw_2$ for some word β': then $\alpha\beta = \alpha\beta' fw_2 \in L_f$. From the inductive hypothesis follows $\alpha\beta' w_2 \in L_f$ which implies $w_1w_2 = \alpha f\beta' w_2 \in L_f$.
2. $\alpha = w_1fw'_2$ and $w_2 = w'_2f\beta$ for some word w'_1: then $\alpha\beta = w_1fw'_2\beta$. From the inductive hypothesis follows $w_1w'_2\beta \in L_f$ which implies $w_1w_2 = w_1w'_2f\beta \in L_f$.
3. $w_1 = \alpha f_1$ and $\beta = f_3w_2$ with $f = f_1f_2 = f_2f_3$. Since r is unborded, we get that f_1, f_2 and f_3 are in r^* so $f_1 = f_3$ which implies $w_1w_2 = \alpha f_1w_2 = \alpha f_3w_2 = \alpha\beta \in L_f$.
4. $\alpha = w_1f_1$ and $w_2 = f_3\beta$ with $f = f_1f_2 = f_2f_3$. Since r is unborded, we get $f_1 = f_3$ which implies $w_1w_2 = w_1f_3\beta = w_1f_1\beta = \alpha\beta \in L_f$.

Conversely, if r is bordered, then $r = r_1r_2 = r_3r_1$ for some non empty words r_1, r_2 and r_3. In particular $r_1 \notin r^*$. This implies that $f = r_1f_1 = f_2r_1$ for some words f_1 and f_2. Observe first that $f_1r_1 \neq f$: indeed if we assume $f_1r_1 = r_1f_1 = f$ we get $r_1 \in r^*$ by Lemma 3.3, a contradiction. From this follows $f_1r_1 \notin L_f$. We get $f_2r_1f_1r_1 \in L_f$ but $f_2r_1f_1r_1 \xrightarrow[f\mapsto\varepsilon]{} f_1r_1$ with $f_1r_1 \notin L_f$. ☐

Thanks to Lemmas 3.15 and 3.16, we can state the existence of $\mathcal{C}_{\max}(f)$ when r, the root of f, is unbordered.

Proposition 3.17. *Let $f \in A^+$ and r be the root of f. Then $\mathcal{C}_{\max}(f) = F^*$ if and only if r is unbordered. Moreover if r is unbordered then $I^*_{F^*|f}$ is codeterministic.*

Proof. We first prove the *only if* part: assume r to be bordered. From $r \in (\mathsf{LF}(r) \setminus \{r\})^*$ follows $f \in F^*$. Let $K = F^* \setminus f^+$. We get $L_{K|f} = L_f$ but $K \subsetneq F^*$.

We prove now the *if* part: assume $F^* \subseteq K$ for some language K. From Item 3 of Proposition 3.4 follows $L_{F^*|f} \subseteq L_{K|f} \subseteq L_f$ so we get $L_{K|f} = L_f$ from Proposition 3.12.

Conversely, assume $L_{K|f} = L_f$. Let $w \in F^*$, we shall prove $w \in K$ by induction on $|w|$, the length of the word w. Since $f \in L_{K|f}$, we get $\varepsilon \in K$. If $w \in F^+$, we can consider w', the shortest word such that $ww' \in L_f$ since $w \in \mathsf{LF}(L_f)^*$. From Lemma 3.16, $w' \notin A^*fA^*$. Moreover $w' \in rA^*$ implies $w \in A^*r$ so $w' \notin rA^*$. Let us consider now the word wfw' that belongs to L_f. Since $L_{K|f} = L_f$, we get $wfw' = xfy$ for some word $x \in K$ and some word $y \in A^*$. If $|x| < |w|$, we get $w \in A^*r$ and if $|x| > |w|$, we get $w' \in rA^*$ so $w = x \in K$ which implies $\mathcal{C}_{\max}(f) = F^*$.

In order to prove that $I^*_{F^*|f}$ is codeterministic, we shall prove that F^*f is a prefix set. Assume $w_0f = w_0'fw_1'$ for some distinct words w_0 and w_0' in F^*. Since $F^* \cap A^*fA^* = \emptyset$, we get $|w_0'f| > |w_0|$ which implies $f = f_1f_2 = f_2f_3$ for some $f_1 \neq \varepsilon$ with $w_0 = w_0'f_1$ and $w_0f = w_0'ff_3$. Since r is unbordered, we get in particular $f_1 \in r^+$ so $w_0 \in A^*r$, a contradiction with Lemma 3.15. Hence $I^*_{F^*|f}$ is codeterministic.

☐

Proposition 3.17 does not hold anymore when the root of f is bordered: let $f = aba$, a primitive bordered word. In this case, it can be proved that $\mathcal{C}_{\max}(aba) \subsetneq F^* = (a + ab)^*$. More precisely, we prove that $\mathcal{C}_{\max}(aba) = K$ where $K = a^* + (ab)^*$ (Lems. 3.18 and 3.19) and that $I^*_{K|aba}$ is unambiguous (Lem. 3.22).

Lemma 3.18. *For all words $w \in L_{aba} \setminus \{\varepsilon\}$ it holds that $w = \gamma^i abaw'$ for some natural number i and some word w' with $\gamma \in \{a, ab\}$ and $\gamma^i w' \in L_{aba}$. So $L_{K|aba} = L_{aba}$.*

Proof. First we can prove that for all words w_1 and w_2 it holds that if $w_1aabw_2 \in L_{aba}$ then $w_1w_2 \in L_{aba}$. The proof is an induction on $|w_1w_2| = 3n$ for some natural number $n > 0$. If $n = 1$ then $w_1aabw_2 = aababa$ or $w_1w_2 = abaaba$ so the property is satisfied. If $n > 1$ then $w_1aabw_2 = \alpha aba\beta$ for some words α and β with $\alpha\beta \in L_{aba}$. Let us consider four cases:

1. $|\alpha aba| \leq |w_1|$. Then $w_1 = \alpha abaw_1'$ for some word w_1' and $\beta = w_1'aabw_2$. From this follows $\alpha w_1'aabw_2 \in L_{aba}$ and from the inductive hypothesis, we get $\alpha w_1'w_2 \in L_{aba}$ which implies $\alpha abaw_1'w_2 = w_1w_2 \in L_{aba}$.
2. $w_1 = \alpha ab$ and $\beta = abw_2$. Then $w_1w_2 = \alpha abw_2 = \alpha\beta \in L_{aba}$.
3. $\alpha = w_1a$ and $w_2 = a\beta$. Then $w_1w_2 = w_1a\beta = \alpha\beta \in L_{aba}$.
4. $|aba\beta| \leq |w_2|$. Then $w_2 = w_2'aba\beta$ for some word w_2' and $\alpha = w_1aabw_2'$. From this follows $w_1aabw_2'\beta \in L_{aba}$ and from the inductive hypothesis, we get $w_1w_2'\beta \in L_{aba}$ so $w_1w_2'aba\beta = w_1w_2 \in L_{aba}$.

From this property, we can directly deduce the property $(\mathcal{P})$: L_{aba} is closed by the rewriting system defined by the rule $aaba \mapsto a$. Symmetrically, we can prove the property $(\mathcal{Q})$: L_{aba} is closed by the rewriting system defined by the rule $abaa \mapsto a$ as a consequence of the property: for all words w_1 and w_2 it holds that if $w_1baaw_2 \in L_{aba}$ then $w_1w_2 \in L_{aba}$.

Let us now consider a word $w \in L_{aba} \setminus (\varepsilon + aba)$. We have $|w|_a = 2|w|_b > |w|_b + 1$ and $w \notin (bA^* \cup A^*bbA^*)$ so $w = (ab)^i aaw'$ for some natural number i and some word w'. Let us consider two cases.

1. $i = 0$. Then $w = aaw'$ for some word $w' \in a^*baA^*$. From this follows $w = a^k aabaw''$ for some natural number k and some word w'' with $a^k aw'' \in L_{aba}$ from Property $(\mathcal{P})$:
2. $i > 0$. Then $w = (ab)^{i-1}abaaw'$ for some word w' and $(ab)^{i-1}aw' \in L_{aba}$ from Property $(\mathcal{Q})$.

$\square$

Lemma 3.19. *Let $R \subseteq A^*$.*

1. *If $a^i \notin R$ then $a^i aba(ba)^i \in L_{aba} \setminus L_{R|aba}$.*
2. *If $(ab)^i \notin R$ then $(ab)^i abaa^i \in L_{aba} \setminus L_{R|aba}$.*

Proof. If another occurrence of aba than the occurence that is just after the left factor a^i (or the left factor $(ab)^i$ in the second case) is erased then one obtain a word that belongs to $bA^* \cup A^*bbA^*$, a contradiction since $L_{aba} \cap (bA^* \cup A^*bbA^*) = \emptyset$. $\square$

From Lemmas 3.18 and 3.19, we get $\mathcal{C}_{\max}(aba) = K$: indeed, let R be a language such that $K \subseteq R \subseteq (a+b)^*$ then $L_{R|aba} = L_{aba}$ from Lemma 3.18 and Item 3 of Proposition 3.4. Conversely, if a language R satisfies $L_{R|aba} = L_{aba}$ then $K \subseteq R$ from Lemma 3.19.

The following lemma will be used in the proof of Lemma 3.21. It is more general that needed in the proof because it will also be used in Section 6.

Lemma 3.20. *If $u \in A^*bd$ and $f \in A^*ad$ for some word d and some distincts letters a and b, then for all words $\alpha \in A^*$, $I^*_{u^*|f}(A^*a\alpha) \cap A^*b\alpha = \emptyset$.*

Proof. We prove this property by induction on $|\alpha|$. If $\alpha = \varepsilon$ the existence of a derivation $wa \xrightarrow[I_{u^*|f}]{*} w'b$ for some words w and w' would imply $u \in A^*a$ and $f \in A^*b$, a contradiction. Assume now $\alpha \neq \varepsilon$ and $wa\alpha \xrightarrow[I_{u^*|f}]{*} w'b\alpha$. From Lemma 6.2 follows $wa\alpha = w_1w_2 \xrightarrow[I_{u^*|f}]{*} u^iw_2 \xrightarrow[I_{u^*|f}]{} u^ifw_2 \xrightarrow[I_{u^*|f}]{*} w_1'w_2 = w'b\alpha$ with, in particular, $\alpha = \alpha'w_2$. If $w_2 \neq \varepsilon$, it would follow $wa\alpha' = w_1 \xrightarrow[I_{u^*|f}]{*} u^i \xrightarrow[I_{u^*|f}]{} u^if \xrightarrow[I_{u^*|f}]{*} w_1' = w'b\alpha'$ with $|\alpha'| < |\alpha|$, in contradiction with the inductive hypothesis. So we can assume $wa\alpha \xrightarrow[I_{u^*|f}]{*} u^i \xrightarrow[I_{u^*|f}]{} u^if \xrightarrow[I_{u^*|f}]{*} w'b\alpha$. Since $I^*_{u^*|f}(A^*ad) = A^*ad$, we get $wa\alpha \in A^*bd$ and $w'b\alpha \in A^*ad$, a contradiction. $\square$

Lemma 3.21.

1. *For all words $w_1 \notin a^*$ and for all words w_2, it holds that $I^*_{K|aba}(w_1aaw_2) = I^*_{K|aba}(w_1)aaw_2$.*
2. *For all integers $i > 1$ and for all words w, it holds that $I^*_{K|aba}(a^ibw) = I^*_{K|aba}(a^i)bw$*
3. *For all words $\beta \in (a+b)^*$, $I^*_{K|aba}((ab)^+a\beta) \cap A^*b\beta = \emptyset$.*

Proof.

1. Clearly, $I^*_{K|aba}(w_1)aaw_2 \subseteq I^*_{K|aba}(w_1aaw_2)$. Conversely, it is sufficient to prove that for all words $w \in I_{K|aba}(w_1aaw_2)$, it holds that $w = w_1'aaw_2$ with $w_1' \in I_{K|aba}(w_1)$ that is clearly satisfied since $\mathsf{LF}(w_1aaw_2) \cap w_1aA^* \cap K = \emptyset$.
2. Clearly, $I^*_{K|aba}(a^i)bw \subseteq I^*_{K|aba}(a^ibw)$. Conversely, let $w' \in I_{K|aba}(a^ibw)$ then $w' = abaa^ibw$ or $w' = a^jabaa^kbw$ with $j > 0$ and $i = j+k$. From 1,

$$I^*_{K|aba}(abaa^ibw) = I^*_{K|aba}(ab)aa^ibw \subseteq I^*_{K|aba}(a^i)bw$$

and on the other hand we get

$$I^*_{K|aba}(a^jabaa^kbw) = I^*_{K|aba}(a^{j+1})baa^kbw \subseteq I^*_{K|aba}(a^i)bw$$

by induction on the length of derivation.

3. Assume $\alpha b\beta \in I^*_{K|aba}((ab)^i a\beta)$ for some word α and some natural number $i > 0$. From Lemma 3.20, $\alpha b\beta \notin I^*_{(ab)^*|aba}((ab)^i a\beta)$ so there exists some word w such that $abw \in I^*_{(ab)^*|aba}((ab)^i a\beta)$ and $\alpha b\beta \in I^*_{K|aba}(aababw)$. From 2, $I^*_{K|aba}(aababw) = I^*_{K|aba}(aa)babw$.
Moreover $|w| \geq |a\beta|$ so $I^*_{K|aba}(aababw) \subseteq A^*a\beta$ so $\alpha b\beta$ cannot belong to $I^*_{K|aba}(aababw)$ and this implies $\alpha b\beta \notin I^*_{K|aba}((ab)^i a\beta)$.

□

Lemma 3.22. *$I^*_{K|aba}$ is unambiguous.*

Proof. Assume that $I^*_{K|aba}$ is ambiguous. Then there exists words $w, \alpha_1, \alpha_2, \beta$ such that

- $w = \alpha_1\alpha_2\beta$
- $\alpha_1 \in K$
- $\alpha_1\alpha_2 \in K$
- $\alpha_2 \neq \varepsilon$
- $I^*_{K|aba}(w_1) \cap I^*_{K|aba}(w_2) \neq \emptyset$ where $w_1 = \alpha_1 aba\alpha_2\beta$ and $w_2 = \alpha_1\alpha_2 aba\beta$.

Let us consider two cases:

1. $\alpha_1\alpha_2 \in a^*$. In this case, $w_1 = a^i abaa^t\beta$ with $t > 0$ and $w_2 = a^{i+t}aba\beta$. From Item 1 of Lemma 3.21 follows $I^*_{K|aba}(w_1) = I^*_{K|aba}(a^{i+1}b)a^{t+1}\beta \subseteq A^*aa\beta$ and from Item 2 of Lemma 3.21 follows
$$I^*_{K|aba}(w_2) = I^*_{K|aba}(a^{i+t+1})ba\beta \subseteq A^*ba\beta$$
so $I^*_{K|aba}(w_1) \cap I^*_{K|aba}(w_2) = \emptyset$, a contradiction.
2. $\alpha_1\alpha_2 \in (ab)^+$. We have to consider two sub-cases:
   (a) $\alpha_1 \in (ab)^*$. In this case, $w_1 = (ab)^i aba(ab)^t\beta$ and $w_2 = (ab)^{i+t}aba\beta$ for some natural numbers i and t with $t > 0$. From Item 1 of Lemma 3.21 follows $I^*_{K|aba}(w_1) = I^*_{K|aba}((ab)^i ab)a(ab)^t\beta \subseteq A^*b\beta$. Now, since $w_2 \in (ab)^+a\beta$, we get $I^*_{K|aba}(w_1) \cap I^*_{K|aba}(w_2) = \emptyset$ from Item 3 of Lemma 3.21.
   (b) $\alpha_1 = a$. In this case, $w_1 = aabab(ab)^i\beta$ and $w_2 = (ab)^{i+1}aba\beta$ for some natural number i. From Item 2 of Lemma 3.21, $I^*_{K|aba}(w_1) = I^*_{K|aba}(aa)b(ab)^{i+1}\beta \subseteq A^*b\beta$. Since, again, $w_2 \in (ab)^+a\beta$, we get $I^*_{K|aba}(w_1) \cap I^*_{K|aba}(w_2) = \emptyset$ from Item 3 of Lemma 3.21.

In all cases, we get $I^*_{K|aba}(w_1) \cap I^*_{K|aba}(w_2) = \emptyset$, a contradiction. So $I^*_{K|aba}$ is unambiguous. □

Moreover, $\mathcal{C}_{\max}(aba)$ is also different from $R = \mathsf{LF}(aba)^* \setminus (a+b)^*aba(a+b)^*$: if we consider any language Z satisfying $L_{Z|aba} = L_{aba}$, then $abab$ must belong to Z. Indeed $w = abababaaa \in L_{aba}$ so there exist some words α and β such that $\alpha \in Z$, $\alpha\beta \in L_{aba}$ and $w = \alpha aba\beta$. Since, clearly, $L_{aba} \cap bA^* = L_{aba} \cap A^*bbA^* = \emptyset$, we get $babaaa \notin L_{aba}$ and $abbaaa \notin L_{aba}$. Hence $\alpha = abab \in Z$ but $abab \notin R$ so $w = abababaaa \in L_{aba} \setminus L_{R|aba}$. Moreover $I^*_{Z|aba}$ is not codeterministic: indeed, $(abab, abababa) \in I_{Z|aba}$ and, since clearly ε must belong to Z, $(baba, abababa) \in I_{Z|aba}$.

More generally we get Proposition 3.25 that needs the following lemmas:

Lemma 3.23. *Let $\alpha f\beta = xfy$ with $\alpha\beta = xy$ for some words f, α, β, x, y. If $|x| < |\alpha|$ then $\alpha \in xr^+$ and $y \in r^+\beta$ where r is the root of f. Moreover, if $\alpha \in \mathsf{LF}(r^*)$ then $\alpha, \beta, x, y \in r^*$.*

Proof. Since $|x| < |\alpha|$, there exist some nonempty words α', y' such that $\alpha = x\alpha'$ and $y = y'\beta$. Then $\alpha\beta = x\alpha'\beta = xy = xy'\beta$ which implies $\alpha' = y'$. From this follows $x\alpha' f\beta = xf\alpha'\beta$ so $\alpha' \in r^+$ from Lemma 3.3. Now, if $\alpha \in \mathsf{LF}(r^*) \cap A^*r$ then $\alpha \in r^*$ since r is a primitive word and we get $\beta, x, y \in r^*$. □

As a consequence of Lemma 3.23, we get:

Corollary 3.24. *If $L_{K|f} = L_f$ for some language K then $F \subseteq K$.*

Proof. Assume $L_{K|f} = L_f$ and let $f_1 \in F$. Then $f_1 f f_2 \in L_f = L_{K|f}$ for some word f_2. If $f_1 \notin K$ then there exist some words $x \in K$ and $y \in A^*$ such that $xy = f_1 f_2 = f$ and $xfy = f_1 f f_2$. From Lemma 3.23, it would follow $f_1 \in xr^+$ or $x \in f_1 r^+$ with $f_1 \in r^*$ in both cases that contradicts $f_1 \in F$. □

Now we can prove:

Proposition 3.25. *If r, the root of f, is bordered then there does not exist any language K such that $L_{K|f} = L_f$ with $I^*_{K|f}$ codeterministic.*

Proof. If r is bordered then there exists some word $f_2 \in \mathsf{LF}(r) \setminus \{\varepsilon, r\}$ such that $f = f_1 f_2 = f_2 f_3$ for some words f_1, f_3. Moreover, since $f_2 \notin r^*$, we get $f_2 f_3 \neq f_3 f_2$ else r would not be primitive. From that also follows $f_1 \in F$. Since $F \subseteq K$ from Corollary 3.24, we get $f_1 \in K$. On the other hand, clearly $\varepsilon \in K$ since $f \in L_{K|f}$. From this follows $(f_1 f_2, f_1 f f_2) \in I_{K|f}$ and $(f_3 f_2, f f_3 f_2) \in I_{K|f}$. Since $f_1 f f_2 = f_1 f_2 f_3 f_2 = f f_3 f_2$ and $f_1 f_2 \neq f_3 f_2$, we get that $I^*_{K|f}$ is not codeterministic. □

4. A non-(FIN/CF) rationally controlled insertion system

The following language $K_0 = {D'_1}^* \cap K$ where $K = \{w \in (a+b)^* \mid \forall x \in \mathsf{LF}(w), |x|_a \leq 2|x|_b + 1\}$ will play a central role in this section. First, observe that this language is not context-free: indeed $K_0 \cap (ab)^* a^+ b^+ = \{(ab)^p a^n b^n \mid p \geq n-1 \geq 0\}$ that is not a context-free language. On the other hand, K_0 enjoys some easy but useful properties:

Lemma 4.1.

1. $K_0 = \mathsf{LF}(K_0) \cap D_1^*$,
2. $K_0 K_0^{-1} = K_0 = K_0^*$,
3. $K_0 \subsetneq K_0^{-1} K_0 = {D'_1}^*$,
4. $\mathsf{RF}(K_0) = \mathsf{RF}({D'_1}^*) = (\mathsf{RF}(K_0))^*$.

Proof. 1 and 2 are consequences of the two equalities: ${D'_1}^* = \mathsf{LF}({D'_1}^*) \cap D_1^*$ and $K = \mathsf{LF}(K)$. For 3, $K_0 \subsetneq {D'_1}^*$ since $aabb \notin K_0$; it only remains to prove $K_0^{-1} K_0 = {D'_1}^*$. First, $K_0^{-1} K_0 \subseteq ({D'_1}^*)^{-1} {D'_1}^* = {D'_1}^*$. Conversely, let $w \in {D'_1}^*$ with $|w|_a = n$; we get $(ab)^n w \in K_0$ which implies $w \in K_0^{-1} K_0$. At last, for 4, $\mathsf{RF}({D'_1}^*) \subseteq \mathsf{RF}(K_0)$ follows from 3 so $\mathsf{RF}(K_0) = \mathsf{RF}({D'_1}^*)$ and $\mathsf{RF}(K_0) = \mathsf{RF}(K_0)^*$ since $\mathsf{RF}({D'_1}^*) = \mathsf{RF}({D'_1}^*)^*$. □

Let us denote by $R = \{w \in A^* \mid |w|_a = 2n, n \geq 0\}$ and $I_0 = I_{R|ab}$. We have:

Lemma 4.2. *$I_0^*(\varepsilon) = K_0$, thus $I_0^*(\varepsilon)$ is not a context-free language.*

Proof.

- $I_0^*(\varepsilon) \subseteq K_0$: the inclusion $I_0^*(\varepsilon) \subseteq D'_1$ is clear. The proof of the inclusion $I_0^*(\varepsilon) \subseteq K$ is an induction on the length of the derivation from ε: assume $\varepsilon \xrightarrow{*} xy \to xaby$ for some words x and y with $x \in R$. From the inductive hypothesis, $|x|_a \leq 2|x|_b + 1$. Moreover, since $x \in R$, we get that $|x|_a$ is even and this implies $|x|_a < 2|x|_b + 1$. From this follows $|xa|_a \leq 2|xa|_b + 1$ and $\mathsf{LF}(xaby) \subseteq K$.
- $K_0 \subseteq I_0^*(\varepsilon)$: the proof is by induction on the length of $w \in K_0$; clearly $\varepsilon \in I_0^*(\varepsilon)$. If $w \neq \varepsilon$, we get $w = abw'$ for some word $w' \in {D'_1}^*$. Observe that w' need not belong to K_0 like for instance $w' = aabb$, so we consider two cases:
  1. $w' \in K_0$. In this case, $w = abw' \in I_0^*(\varepsilon)$ from the inductive hypothesis.
  2. $w' \notin K_0$. In this case, $w' \notin K$ because $w' \in {D'_1}^*$. Let w'_1 be the shortest word such that $w = w'_1 w'_2$ for some word w'_2 and $|w'_1|_a - 2|w'_1|_b > 1$. We get $w'_1 = w''_1 a$ and $|w''_1|_a - 2|w''_1|_b = 1$ for some word w''_1. This implies $w'_2 = bw''_2$ for some word w''_2 because $abw'_1 a \notin K$ and $K = \mathsf{LF}(K)$. Observe that $|abw''_1 ab|_a - 2|abw''_1 ab|_b = 0$ so $|w''_2|_a - 2|w''_2|_b = |w|_a - 2|w|_b \leq 1$. We get $w''_2 \in K$ so $abw''_1 w''_2 \in K$ since $K = \mathsf{LF}(K) = K^*$. On the other hand, $abw''_1 w''_2$ clearly belongs to ${D'_1}^*$, so $abw''_1 w''_2 \in I_0^*(\varepsilon)$ from the

inductive hypothesis. Moreover $|w_1''|_a$ is odd since $|w_1''|_a - 2|w_1''|_b = 1$ so $abw_1'' \in R$ and $w = abw_1''abw_2'' \in I_0^*(\varepsilon)$.

□

As a direct consequence, for every word $w \in A^*$, it holds that there exists some rational language $R_w = wR$ such that $I^*_{R_w|ab}(w)$ is not a context-free language. One can state a stronger result where the rational control is independent from the word w.

Observe that K_0 is closed by I_0^*: indeed $I_0^*(K_0) = I_0^*(I_0^*(\varepsilon)) = I_0^*(\varepsilon) = K_0$. Like K_0, both $\mathsf{LF}(K_0)$ and $\mathsf{RF}(K_0)$ are closed by I_0^*. The equality $\mathsf{LF}(K_0) = I_0^*(\mathsf{LF}(K_0))$ is a particular case of the following property: for all nonempty languages L and for all nonempty words f, it clearly holds that for all languages K, $\mathsf{LF}(K) = I^*_{L|f}(\mathsf{LF}(K))$. For $\mathsf{RF}(K_0)$, we can prove:

Lemma 4.3. $\mathsf{RF}(K_0) = I_0^*(\mathsf{RF}(K_0))$.

Proof. We only have to prove $I_0^*(\mathsf{RF}(K_0)) \subseteq \mathsf{RF}(K_0)$. To prove this inclusion, it is sufficient to consider a single step of derivation: let $w = xyz \in K_0$ for some words x, y and z with $y \in R$. If $xy \in R$ then we get $w \to xyabz$ so $yabz \in \mathsf{RF}(K_0)$, else $abxy \in R$ and $abw \to abxyabz$. Since $abK_0 \subseteq K_0^* = K_0$, we get $yabz \in \mathsf{RF}(K_0)$. □

For every word $w \in (a+b)^*$, $\mathsf{l}(w)$ is the longest left factor of w that belongs to $\mathsf{LF}(K_0)$ and $\mathsf{r}(w)$ is the longest right factor of w that belongs to $\mathsf{RF}(K_0)$. First, we get from Lemma 4.3:

Lemma 4.4. *Let $w \in (a+b)^*$. For all $w' \in I_0^*(w)$, it holds that:*

1. *either $|\mathsf{l}(w')| > |\mathsf{l}(w)|$ or $\mathsf{l}(w') = \mathsf{l}(w)$,*
2. *either $|\mathsf{r}(w')| > |\mathsf{r}(w)|$ or $\mathsf{r}(w') = \mathsf{r}(w)$.*

Proof. The first property is clearly true since $\mathsf{LF}(K_0) = I_0^*(\mathsf{LF}(K_0))$. For the second one, it is sufficient to prove the property for $(w, w') \in I_0$. Let $w = w_1\mathsf{r}(w)$; if $w' = w_1'abw_1''\mathsf{r}(w)$ for some word $w_1'' \neq \varepsilon$ and $w_1' \in R$ then $\mathsf{r}(w') = \mathsf{r}(w)$. Else $w' = w_1w_2'abw_2''$ with $w_2'w_2'' = \mathsf{r}(w)$ and we shall prove $w_2'abw_2'' \in \mathsf{RF}(K_0)$ by considering two cases:

- $|w_2'|_a$ is even. In this case, $w_2'abw_2'' \in I_0^*(\mathsf{RF}(K_0))$ and we get $w_2'abw_2'' \in \mathsf{RF}(K_0)$ from Lemma 4.3.
- $|w_2'|_a$ is odd. Then $abw_2'abw_2'' \in I_0^*(\mathsf{RF}(K_0))$ and we get $abw_2'abw_2'' \in \mathsf{RF}(K_0)$ from Lemma 4.3 so $w_2'abw_2'' \in \mathsf{RF}(K_0)$.

As a consequence, we get $|\mathsf{r}(w')| \geq |w_2'abw_2''| > |\mathsf{r}(w)|$. □

We also obtain these technical results:

Lemma 4.5. *Let $w \notin \mathsf{LF}(K_0)$ and α be the shortest left factor of w that is not in $\mathsf{LF}(K_0)$. Then*

1. *If $|\alpha|_a$ is odd then $\alpha \in K_0 b$,*
2. *if $|\alpha|_a$ is even then $I_0^*(\alpha^{-1}w) = \alpha^{-1}I_0^*(w)$.*

Proof.

1. Assume $\alpha = \alpha' a$ for some word α'. From the choice of α follows $\alpha' \in \mathsf{LF}(K_0)$; moreover, since $|\alpha|_a$ is odd, we get that $|\alpha'|_a$ is even so $\alpha' ab \in I_0^*(\mathsf{LF}(K_0)) = \mathsf{LF}(K_0)$ which implies that $\alpha = \alpha' a \in \mathsf{LF}(K_0)$, a contradiction, hence $\alpha = \alpha' b$ for some word α'. Now, since $\alpha' \in \mathsf{LF}(K_0)$ and $\alpha' b \notin \mathsf{LF}(K_0)$, we get that $\alpha' \in K$, $\alpha' \in \mathsf{LF}(D_1'^*)$ and $\alpha' b \notin \mathsf{LF}(D_1'^*)$ so $\alpha' \in D_1'^*$ which implies $\alpha' \in K_0$.
2. Set $w = \alpha\beta$. The inclusion $I_0^*(\beta) \subseteq \alpha^{-1}(I_0^*(w))$ is clear and, conversely, we shall prove by induction on the length of a derivation $\alpha\beta \xrightarrow{*} \alpha\beta'$ that $\beta \xrightarrow{*} \beta'$. If the length of the derivation is 0 then $\beta = \beta'$, else let us consider the first step of the derivation: $\alpha\beta = uv \to uabv \xrightarrow{*} \alpha\beta'$ for some words u and v. Observe that, from Lemma 4.4, $\mathsf{l}(uabv) = \mathsf{l}(\alpha)$ since $\alpha \notin \mathsf{LF}(K_0)$. Let $\alpha = \alpha' x$ with $x \in A$. If we assume $|u| \leq |\alpha'|$ we get $\alpha' = uu'$ for some word u'. From $\alpha' \in \mathsf{LF}(K_0)$ follows $uabu' \in \mathsf{LF}(K_0)$ and this implies $|\mathsf{l}(uabv)| > |\mathsf{l}(\alpha)|$, a contradiction. Hence $|u| > |\alpha'|$ and $u = \alpha\beta''$ for some word $\beta'' \in A^*$. We get $\alpha\beta = \alpha\beta'' v \to \alpha\beta'' abv \xrightarrow{*} \alpha\beta'$.

From the inductive hypothesis, we get $\beta''abv \xrightarrow{*} \beta'$; moreover, since $|\alpha|_a$ is even, we get that $|\beta''|_a$ is even so $\beta = \beta''v \to \beta''abv$ which implies $\beta \xrightarrow{*} \beta'$.

□

Symmetrically one can prove:

Lemma 4.6. *Let $w \notin \mathsf{RF}(K_0)$ and β be the shortest right factor of w that is not in $\mathsf{RF}(K_0)$. Then $I_0^*(w\,\beta^{-1}) = I_0^*(w)\beta^{-1}$.*

Proof. Set $w = \alpha\beta$. The inclusion $I_0^*(\alpha) \subseteq I_0^*(w)\beta^{-1}$ is immediate. Conversely, assume $w'\beta \in I_0^*(\alpha\beta)$ for some word w'. If $w' = \alpha$, we directly get $w' \in I_0^*(w)\beta^{-1}$. Else, $\alpha\beta = uv \to uabv \xrightarrow{*} w'\beta$. One can consider two cases:

- $|u| > |\alpha|$. In this case, $u = \alpha x\beta'$ with $x \in \{a, b\}$, $\beta' \in (a + b)^*$ and $\beta = x\beta'v$. We get $\alpha x\beta'v \to \alpha x\beta'abv \xrightarrow{*} w'\beta$. Moreover, from the choice of β follows $\beta'v \in \mathsf{RF}(K_0)$. That implies $\beta'abv \in \mathsf{RF}(D_1'^*) = \mathsf{RF}(K_0)$, a contradiction with Lemma 4.4 since $|\beta'abv| = |\beta| + 1$.
- $|u| \le |\alpha|$. In this case, $\alpha = uw_1'$ for some word w_1' and we have $uw_1'\beta \to uabw_1'\beta \xrightarrow{*} w'\beta$. By induction over the length of the derivations, we get $w' \in I_0^*(uabw_1')$ which implies $w' \in I_0^*(\alpha)$.

□

We can now state that for all words w, $I_0^*(w)$ rationally dominates K_0:

Lemma 4.7. *For every word $w \in (a + b)^*$, $I_0^*(w) \rightsquigarrow_{\text{rat}} K_0$, thus $I_0^*(w)$ is not a context-free language.*

Proof. Clearly, it is sufficient to prove that for all words $w \in A^+$ there exists some word w' with $|w'| < |w|$ and $I_0^*(w) \rightsquigarrow_{\text{rat}} I_0^*(w')$. Since $K_0 = \mathsf{LF}(K_0) \cap \mathsf{RF}(K_0)$, we consider three cases:

1. $w \in K_0$: in this case, we directly prove $I_0^*(\varepsilon) = I_0^*(w)w^{-1}$. The inclusion $K_0 = I_0^*(\varepsilon) \subseteq I_0^*(w)w^{-1}$ is clear. Conversely, let w' such that $w'w \in I_0^*(w)$. Since $w \in K_0$, we get $w'w \in K_0$ and, from Item 2 of Lemma 4.1, we get $w' \in K_0$.
2. $w \notin \mathsf{RF}(K_0)$: then, from Lemma 4.6, $I_0^*(\alpha) = I_0^*(w)\beta^{-1}$ where $w = \alpha\beta$ with β the shortest right factor of w that is not in $\mathsf{RF}(K_0)$. Since $\beta \ne \varepsilon$, $|\alpha| < |w|$.
3. $w \notin \mathsf{LF}(K_0)$: Set $w = \alpha\beta$ where α is the shortest left factor of w that is not in $\mathsf{LF}(K_0)$ and let us consider two cases:
   - $|\alpha|_a$ is even. In this case, $I_0^*(\beta) = \alpha^{-1}(I_0^*(w))$ from Item 2 of Lemma 4.5 and $|\beta| < |w|$.
   - $|\alpha|_a$ is odd. From Item 1 of Lemma 4.5, $\alpha = \alpha'b$ for some word $\alpha' \in K_0$; we get that $|\alpha'|_a$ is odd. Let us consider two subcases:
     (a) $|\beta|_a > 0$. Set $\beta = b^ia\beta'$; we get $w = \alpha'b^{i+1}a\beta'$ and, in this case, we prove $I_0^*(\beta') = (\alpha'b^{i+1}a)^{-1}I_0^*(w)$. Since $|\alpha'b^{i+1}a|_a$ is even, the inclusion $I_0^*(\beta') \subseteq (\alpha'b^{i+1}a)^{-1}I_0^*(w)$ is clear. The converse inclusion follows from Item 1 of Lemma 4.4: the property implies that no insertion can be done inside the left factor $\alpha'b^{i+1}a$.
     (b) $|\beta|_a = 0$. Set $\beta = b^ia$; we get $w = \alpha'b^{i+1}$ and, in this case, we shall prove $I_0^*(\varepsilon) = (abw)^{-1}I_0^*(w)$. The inclusion $I_0^*(\varepsilon) \subseteq (abw)^{-1}I_0^*(w)$ is clear. Conversely, let $w = \alpha\beta \xrightarrow{*} abw\gamma$ for some word γ and let us consider the first step of this derivation. Since $|\alpha'|_a$ is odd, we get: $w = \alpha'b^{i+1} = \alpha_1'\alpha_2'b^{i+1} \to \alpha_1'ab\alpha_2'b^{i+1} \xrightarrow{*} ab\alpha_1'\alpha_2'b^{i+1}\gamma$ for some words α_1' and α_2' with $|\alpha_1'|_a$ even and $|\alpha_2'|_a$ odd. From Item 1 of Lemma 4.4 follows $\alpha_1'ab\alpha_2'b^{i+1} = ab\alpha_1'\alpha_2'b^{i+1}$ and also no insertion in the derivation $\alpha_1'ab\alpha_2'b^{i+1} \xrightarrow{*} ab\alpha_1'\alpha_2'b^{i+1}\gamma$ can appear inside the left factor $\alpha_1'ab\alpha_2'b^{i+1}$.

□

Our aim is now to generalize Lemma 4.7 to cases when the inserted word f of the system is some word such that $|f|_a > 0$ and $|f|_b > 0$ for some distinct letters a and b. We need the following lemma involving prefix morphisms that will also be very useful in Section 6. A morphism $h : A^* \mapsto B^*$ is prefix if $h(A)$ is a prefix language.

Lemma 4.8. *Let A and B be two alphabets. Let $f \in A^*$, h be a prefix morphism from A^* to B^* and $L \subseteq A^*$ be a language. Then for all words $w \in A^*$, $I^*_{L'|f'}(w') = h(I^*_{L|f}(w))$ and $I^*_{L|f}(w) = h^{-1}(I^*_{L'|f'}(w'))$ where $f' = h(f)$, $L' = h(L)$ and $w' = h(w)$.*

Proof. Let $w \in A^*$, assume $w = xy$ for some words x, y with $x \in L$ and consider $w'' = xfy$ obtained by a single insertion step in $I^*_{L|f}$. Since $h(x) \in h(L)$, we get $h(x)f'h(y) = h(w'') \in I^*_{L'|f'}(w')$. Moreover, since $h(A)$ is a prefix code, we get $h^{-1}(h(xfy)) = xfy = w''$ so $w'' \in h^{-1}(I^*_{L'|f'}(w'))$. Conversely, let $h(w) = x'y'$ for some words $x', y' \in B^*$ with $x' \in h(L)$ and consider $w''' = x'f'y'$, obtained by a single insertion step in $I^*_{L'|f'}$. Since $x' \in h(L)$ and h is prefix (so injective), we get $w = xy$ where $h^{-1}(x') = \{x\}$ with $x \in L$, $h^{-1}(y') = y$ and $h^{-1}(w''') = \{xfy\}$. Since $x \in L$, we get $xfy \in I^*_{L|f}(w)$ so $w''' \in h(I^*_{L|f}(w))$ and $h^{-1}(w''') \subseteq I^*_{L|f}(w)$.

Finally, by induction on the length of the derivations in $I^*_{L|f}$ and $I^*_{L'|f'}$, we get the equalities $I^*_{L'|f'}(w') = h(I^*_{L|f}(w))$ and $I^*_{L|f}(w) = h^{-1}(I^*_{h(L)|h(f)}(h(w)))$. □

We observe that the property of being prefix is crucial for morphisms in Lemma 4.8; the result does not hold in general for an injective morphism. As a consequence of this lemma, we can state the following general proposition:

Proposition 4.9. *Let X be an alphabet and f be some word of X^* such that $|f|_a > 0$ and $|f|_b > 0$ for some distinct letters a and b. Then there exists a rational language R' such that for every word $w \in X^*$, $I^*_{R'|f}(w)$ is not a context-free language.*

Proof. We may assume $f = a^i bf'$ for some strictly positive integer i and some word f'. Let $A = \{a, b\}$ and $h : A^* \mapsto X^*$ be the prefix morphism defined by $h(a) = a^i$ and $h(b) = bf'$. Let $R = \{w \in A^* \mid |w|_a = 2n, n \geq 0\}$ and $R' = h(R)$.

Let $w' \in X^*$, it can be factorized as $w' = w'_1 w'_2$ with w'_1 being the longest left factor of w' that is in $h(A)^*$. Since h is prefix, there exists a unique word $w_1 \in A^*$ such that $h(w_1) = w'_1$. We claim that $I^*_{R'|f}(w') = I^*_{R'|f}(w'_1)w'_2$. Indeed, the inclusion $I^*_{R'|f}(w'_1)w'_2 \subseteq I^*_{R'|f}(w')$ is clear and conversely, the inclusion $I^*_{R'|f}(w') \subseteq I^*_{R'|f}(w'_1)w'_2$ can be proved by induction over the length of the derivation by considering a single insertion step: $w' = xy \to x\ h(f)\ y$ with $x \in R'$. From this follows $x \in h(A)^*$ so $w'_1 = xx'$ with $x' \in h(A)^*$. That implies $x\ h(f)\ y = x\ h(f)\ x'w'_2$ and $x\ h(f)\ x'$ is the longest left factor of $x\ h(f)\ y$ that is in $h(A)^*$.

Now, from Lemma 4.8, we get $I^*_0(w_1) = h^{-1}(I^*_{R'|f}(w'_1))$ that is not context-free from Lemma 4.7. That implies that $I^*_{R'|f}(w'_1)$ cannot be context-free so $I^*_{R'|f}(w') = I^*_{R'|f}(w'_1)w'_2$ is not context-free. □

5. Single letter controlled insertion system

In the previous section, Proposition 4.9 needs that $|f|_a > 0$ and $|f|_b > 0$ for some distinct letters a and b in order to build a rational language R' such that for every word $w \in X^*$, $I^*_{R'|f}(w)$ is not a context-free language. Since this proposition does not hold when $f \in a^+$ for some letter a, a natural question is to wonder whether such controlled insertion systems are FIN/CF or RAT/CF. While the answer is positive for the FIN/CF property,[1] it is, rather surprisingly, not the case for the RAT/CF property. We start this section by giving two examples of such non-RAT/CF controlled insertion systems in the case when the inserted word consists in a single occurrence of a single letter a. The proof and the rational control of the first example are quite simple while for the second example it is the starting rational language which is as simple as possible.

Lemma 5.1. *Let $R = R_0 \cup R_1 \cup R_2$ with $R_0 = a^*$, $R_1 = \{a^i(ba)^j b \mid i + j \text{ is odd}\}$ and $R_2 = \{a^i(ba)^j b^p c(ba)^k b \mid j + k \text{ is odd}\}$, then $I^*_{R|a}(b^*cb^*)$ is not a context-free language.*

[1] In fact, we shall even see that the answer is positive for the FIN/RAT property.

Proof. Let $L_0 = a^*(ba)^*c(ba)^*$ and $L_1 = \{a^i(ba)^jc(ba)^k \mid i \geq j \geq k\}$. Observe that L_1 is not a context-free language. We shall prove that $I^*_{R|a}(b^*cb^*) \cap L_0 = L_1$ that implies the non-context-freeness of $I^*_{R|a}(b^*cb^*)$.

- $L_1 \subseteq I^*_{R|a}(b^*cb^*) \cap L_0$: let $i \geq j \geq k$ be three natural numbers. We get the derivation:

$$b^jcb^k \xrightarrow[I_{R_0|a}]{} ab^jcb^k \xrightarrow[I_{R_1|a}]{} a(ba)b^{j-1}cb^k \xrightarrow[I_{R_2|a}]{} a(ba)b^{j-1}c(ba)b^{k-1}$$

$$\xrightarrow[I_{R|a}]{*} a^k(ba)^kb^{j-k}c(ba)^k \xrightarrow[I_{R_0\cup R_1|a}]{*} a^j(ba)^jc(ba)^k \xrightarrow[I_{R_0|a}]{*} a^i(ba)^jc(ba)^k.$$

- $I^*_{R|a}(b^*cb^*) \cap L_0 \subseteq L_1$: let $H = A^*baaA^*$ with $A = \{a, b, c\}$. Clearly $I^*_{R|a}(H) = H$ and $H \cap L_0 = \emptyset$. We shall prove that , for all words $w \in L_2 = \{a^i(ba)^jb^pc(ba)^kb^q \mid i \geq j \geq k\}$, if $w \xrightarrow[I_{R|a}]{} w' \xrightarrow[I_{R|a}]{*} w'' \in L_0$ then $w' \in L_2$.
  Assume $w = a^i(ba)^jb^pc(ba)^kb^q$ with $i \geq j \geq k$, we can distinguish three cases:
  1. $(w, w') \in I_{R_0|a}$: in this case, $w' = a^{i+1}(ba)^jb^pc(ba)^kb^q$ is clearly in L_2.
  2. $(w, w') \in I_{R_1|a}$: since $w' \notin H$ (else $w'' \notin L_0$), we ge't $w' = a^ic(ba)^{j+1}b^{p-1}c(ba)^kb^q$. Moreover $i + j$ is odd so $i > j$ and we get $i \geq j + 1 > k$ so $w' \in L_2$.
  3. $(w, w') \in I_{R_2|a}$: since $w' \notin H$, we get $w' = a^i(ba)^jb^pc(ba)^{k+1}b^{q-1}$ with $j + k$ odd. That implies $j > k$ so $i \geq j \geq k + 1$ so $w' \in L_2$.

  Finally, by induction over the length of the derivation, we get that for all $w \in L_2$, if $w \xrightarrow[I_{R|a}]{*} w'' \in L_0$ then $w'' \in L_2 \cap L_0 = L_1$. Now, since $b^*cb^* \subseteq L_2$ we get $I^*_{R|a}(b^*cb^*) \cap L_0 \subseteq L_1$.

□

We observe that a similar construction would allow to get an example of a rational control R such that $I^*_{R|a}(b^*a^2b^*)$ is not a context-free language, nevertheless, in the following second example, the starting rational language is as simple as possible: indeed, we shall see later that single letter controlled insertion systems are FIN/RAT.

Lemma 5.2. *Let* $R = R_0 \cup R_1 \cup R_2 \cup R_3$ *where* $R_0 = (b^6a)^*b^6$, $R_1 = a^*$, $R_2 = \{a^i(b^2a)^jb^2 \mid i + j \text{ is odd}\}$ *and* $R_3 = \{a^i(b^2a)^jb^4a(b^6a)^p(b^5a)^qb^5 \mid j + q \text{ is odd}\}$, *then* $I^*_{R|a}(b^*)$ *is not a context-free language.*

Proof. Let $L_0 = a^+(b^2a)^+(b^5a)^+$ and $L_1 = \{a^i(b^2a)^{3k}(b^5a)^q \mid i > k \geq q > 0\}$, we shall prove that $I^*_{R|a}(b^*) \cap L_0 = L_1$, a non-context-free language.

- $L_1 \subseteq I^*_{R|a}(b^*) \cap L_0$: let $i > k \geq q > 0$ be three integers.
  We first prove that for all integers t such that $0 \leq t \leq q$, it holds that

$$b^{6k+5q} \xrightarrow[I_{R|a}]{*} a^{t+1}(b^2a)^{3t}(b^6a)^{k-t}(b^5a)^t(b^5)^{q-t}$$

  by induction on t.
  - For the case $t = 0$, we get $b^{6k+5q} \xrightarrow[I_{R_0|a}]{*} (b^6a)^kb^{5q} \xrightarrow[I_{R_1|a}]{*} a(b^6a)^kb^{5q}$.
  - For $0 \leq t < q$, we get $a^{t+1}(b^2a)^{3t}(b^6a)^{k-t}(b^5a)^t(b^5)^{q-t}$
    $\xrightarrow[I_{R_2|a}]{*} a^{t+1}(b^2a)^{3t+1}b^4a(b^6a)^{k-t-1}(b^5a)^t(b^5)^{q-t}$
    $\xrightarrow[I_{R_3|a}]{*} a^{t+1}(b^2a)^{3t+1}b^4a(b^6a)^{k-t-1}(b^5a)^{t+1}(b^5)^{q-t-1}$

$\xrightarrow[I_{R_2|a}]{*} a^{t+1}(b^2a)^{3t+3}(b^6a)^{k-t-1}(b^5a)^{t+1}(b^5)^{q-t-1}$
$\xrightarrow[I_{R_1|a}]{*} a^{t+2}(b^2a)^{3t+3}(b^6a)^{k-t-1}(b^5a)^{t+1}(b^5)^{q-t-1}$.

Setting now $t = q$ in $a^{t+1}(b^2a)^{3t}(b^6a)^{k-t}(b^5a)^t(b^5)^{q-t}$, we get
$b^{6k+5q} \xrightarrow[I_{R|a}]{*} a^{q+1}(b^2a)^{3q}(b^6a)^{k-q}(b^5a)^q \xrightarrow[I_{R_2|a}]{*} a^{q+1}(b^2a)^{3k}(b^5a)^q \xrightarrow[I_{R_1|a}]{*} a^i(b^2a)^{3k}(b^5a)^q$.

- $I^*_{R|a}(b^*) \cap L_0 \subseteq L_1$: let $H = A^*bA^*a(\varepsilon + b + b^3)aA^* + a^+(b^2a)^*b^*$. It is easily seen that $H = I^*_{R|a}(H)$ and $H \cap L_0 = \emptyset$. From this follows that for every words w and $w' \in I^*_{R|a}(w)$, if w' is in L_0 then w is not in H. Let $L_2 = \{a^i(b^2a)^j(b^4a)^t(b^6a)^p(b^5a)^qb^r \mid (t = 0 \vee t = 1) \wedge (\exists k, j + 2t = 3k \wedge i + t > k \geq q)\}$. Observe that, in order to reach some word in L_0 from some word of b^* by $I^*_{R|a}$, it is necessary to begin with $I^*_{R_0|a}$ then to continue with at most one step with $I^*_{R_1|a}$ before using $I^*_{R_2|a}$ and $I^*_{R_3|a}$. Moreover, it is easily seen that for every word $w \in I^*_{R|a}(a^+b^*)$, if $w \notin H$ then it is necessary to first reach a word in the form $a^i(b^6a)^jb^r$ with $i > 0$ and $j > 0$, a word that is in L_2.
  We shall now prove that for any word $w = a^i(b^2a)^j(b^4a)^t(b^6a)^p(b^5a)^qb^r$ in L_2, for any word w' that can be obtained from w by a single step with $I^*_{R|a}$, it holds that if $w' \notin H$ then $w' \in L_2$. Since $I^*_{R_0|a}$ cannot be used anymore on a word of L_2, we can have $(w, w') \in I_{R_1|a}$, $(w, w') \in I_{R_2|a}$ or $(w, w') \in I_{R_3|a}$. If $(w, w') \in I_{R_1|a}$, we clearly get $w' \in L_2$ so we consider the two other cases:
  1. $(w, w') \in I_{R_2|a}$. we have again to consider two cases:
     (a) if $t = 1$: in this case $w = a^i(b^2a)^jb^4a(b^6a)^p(b^5a)^qb^r$ with $j + 2 = 3k$ and $i + 1 > k \geq q$. Observe that, since $w' \notin H$, we cannot insert an a with $I^*_{R_2|a}$ in the factor $(b^2a)^j$ of w. From this follows $w' = a^i(b^2a)^{j+2}(b^6a)^p(b^5a)^qb^r$ and $i + j$ is odd. Hence $i + j + 2t = i + 3k$ is odd so $i + k$ is odd and this implies $i \neq k$ so $i > k$ and $w' \in L_2$.
     (b) if $t = 0$: in this case $w = a^i(b^2a)^j(b^6a)^p(b^5a)^qb^r$ with $j = 3k$ and $i > k \geq q$. Observe that we cannot have $p = 0$ else $q > 0$ since $w \notin H$ and this leads to insert an a into the factor $(b^5a)^q$ that would give a word that belongs to H. Hence $w' = a^i(b^2a)^{j+1}(b^4a)(b^6a)^{p-1}(b^5a)^qb^r$ with $j + 1 + 2 = 3k + 3$ and $i + 1 > k + 1 \geq q$.
  2. $(w, w') \in I_{R_3|a}$. In this case $w = a^i(b^2a)^jb^4a(b^6a)^p(b^5a)^qb^r$ with $j + 2 = 3k$ and $i + 1 > k \geq q$. Observe that we cannot insert an a neither into the factor $(b^6a)^p$ nor into the factor $(b^5a)^q$ since it would give a word of H. From this follows $w' = a^i(b^2a)^jb^4a(b^6a)^p(b^5a)^{q+1}b^{r-5}$ with $j + 2 = 3k$ and $i + 1 > k$. Moreover, from $j + q$ odd follows $j + 2 + q = 3k + q$ odd then $k + q$ odd. That implies $k \neq q$ so $k \geq q + 1$.

  We have proved that for every word w in $I^*_{R|a}(b^*)$, it holds that if $I^*_{R|a}(w) \cap L_0 \neq \emptyset$ then $w \in L_2$, that is $w = a^i(b^2a)^j(b^4a)^t(b^6a)^p(b^5a)^qb^r \mid (t = 0 \vee t = 1) \wedge (\exists k, j + 2t = 3k \wedge i + t > k \geq q$. Now if $w \in L_0$ we get $t = p = r = 0$ so $j = 3k$ and $i > k \geq q$ which implies $w \in L_1$.

□

Lemma 5.3. *For all languages $L \subseteq A^*$ and for all words $w \in A^*$, $\mathsf{LF}(I^*_{L|a}(w)) = I^*_{L|a}(\mathsf{LF}(w))$.*

Proof. Clearly, if $w = w_1w_2$ and $w'_1 \in I^*_{L|a}(w_1)$ then $w'_1w_2 \in I^*_{L|a}(w)$ and we get $I^*_{L|a}(\mathsf{LF}(w)) \subseteq \mathsf{LF}(I^*_{L|a}(w))$. Conversely, let $w = xy$ for some words x and y with $x \in L$. Assume $xy \to xay = w'_1w'_2$; we can consider two cases:

- $x = w'_1x'$ for some word x'. In this case, $w'_1 \in I^*_{L|a}(\mathsf{LF}(w))$.
- $w'_1 = xay'$ for some word y' with $y = y'w'_2$. From this follows $xy' \to xay' = w'_1$ so $w'_1 \in I^*_{L|a}(\mathsf{LF}(w))$. Now, by induction on the length of the derivations, we get $\mathsf{LF}(I^*_{L|a}(w)) \subseteq I^*_{L|a}(\mathsf{LF}(w))$.

□

As said before, it has been proved in [2], through a more general result, that $I^*_{R|a}$ is a rational transduction in the case when R is finite: indeed this case corresponds to the prefix rewriting system associated with the set of rules $\{u \mapsto ua \mid u \in R\}$. A prefix rewriting system S is a rewriting system where the rewriting rules can only be applied on left factors of the words: $w \underset{S}{\rightarrow} w'$ if $w = u\alpha$ and $w' = v\alpha$ for some rule $u \mapsto v \in S$ and some word α. We prove here the following more general result that is somehow *optimal* as stated forward in Proposition 5.7.

In the following, A is an alphabet with $a \in A$ and $B = A \setminus \{a\}$. The projection of a word $w \in A^*$ over the alphabet B is the morphism from A^* to B^*, denoted by Π_B and defined by for all letters $x \in B, \Pi_B(x) = x$ and $\Pi_B(a) = \varepsilon$.

Proposition 5.4. *For all rational languages $R \subseteq A^*$ such that $\Pi_B(R)$ is finite, for all words $f \in a^*$, $I^*_{R|f}$ is a rational transduction.*

Proof. The proof is an induction over $k = \mathsf{max}(\{|w| \mid w \in \Pi_B(R)\})$.

- if $k = 0$, $R \subseteq a^*$. For all words $w \in A^*$, if $R \cap \mathsf{LF}(w) = \emptyset$, we get $I^*_{R|f}(w) = \{w\}$ else $I^*_{R|f}(w) = f^*w$ so we can set $I^*_{R|f}(w) = \{w\} \cup \{f^*(\{w\} \cap RA^*)\}$.
- if $k > 0$, let $R_k = \{w \in R \mid |\Pi_B(w)| = k\}$ and $R_{<k} = \{w \in R \mid |\Pi_B(w)| < k\}$. Observe that $I^*_{R_{<k}|f}$ is a rational transduction from the inductive hypothesis. On the other hand, for all words w, $I^*_{R_k|f}(w) = \{w\} \cup \{w'f^*w'' \mid w = w'w'' \wedge w'$ is the shortest leftfactor of$w$ that belongs to $R_k)$ so $I^*_{R_k|f}$ is a rational transduction. We claim that $I^*_{R|f} = I^*_{R_{<k}|f} \circ I^*_{R_k|f} \circ I^*_{R_{<k}|f}$, so $I^*_{R_k|f}$ is a rational transduction: indeed, for all words w, we clearly have $I^*_{R_{<k}|f} \circ I^*_{R_k|f} \circ I^*_{R_{<k}|f}(w) \subseteq I^*_{R|f}(w)$. Conversely, let $w = w_1 \xrightarrow[I_{R|f}]{} w_2 \ldots \xrightarrow[I_{R|f}]{} w_n = w'$ for some $n > 1$.
  If for all $0 \leq i < n$, $w_i \notin R_kA^*$ then $w' \in I^*_{R_{<k}|f}(w)$. Else let i be the smallest index such that $w_i \in R_k$, we get $w_i = w'_i\alpha$ and $w_1 = w'_1\alpha$ with $w'_1, \alpha \in A^*$, $w'_i \in R_k \cap I^*_{R_{<k}|f}(w'_1)$. Moreover, for all words $v \in R_kA^*$ and $v' \in I^*_{R|f}(v)$, it holds that $v' \in I^*_{R_{<k}|f} \circ I^*_{R_k|f}(v)$ so $w' = w_n \in I^*_{R_{<k}|f} \circ I^*_{R_k|f}(w_i)$. Since $w_i \in I^*_{R_{<k}|f}(w)$, we get $w' \in I^*_{R_{<k}|f} \circ I^*_{R_k|f} \circ I^*_{R_{<k}|f}(w)$.

□

Corollary 5.5. *Let $R \subseteq A^*$ and $K \subseteq A^*$ be two rational languages and $f \in a^*$. If $\Pi_B(K)$ is finite, then $I^*_{R|f}(K) \in \mathsf{RAT}$.*

Proof. Let $k = \mathsf{max}(\{\Pi_B(w) \mid w \in K\})$ and $R' = \{w \in R \mid |\Pi_B(w)| \leq k\}$. Clearly $I^*_{R|a}(K) = I^*_{R'|a}(K)$ and, since $\Pi_B(R')$ is finite, we get from Proposition 5.4 that $I^*_{R'|a}(K)$ is a rational language. □

Conversely, using Lemma 4.8 and Lemma 5.2, we get:

Corollary 5.6. *If $K \subseteq B^*$ is infinite then there exists a rational language R such that $I^*_{R|a}(K)$ is not a context-free language.*

Proof. Let $h : A^* \mapsto (a+b)^*$ be the morphism defined by $h(x) = b$ for all $x \in B$ and $h(a) = a$. For any rational language R, it holds that $I^*_{R|a}(h(w)) = h(I^*_{h^{-1}(R)|a}(w))$ from Lemma 4.8 so it remains to prove that for all infinite languages $K \subseteq b^*$, there exists $R \in \mathsf{RAT}$ such that $I^*_{R|a}(K) \notin \mathsf{CF}$. We shall first prove that, for all words $w \in b^*$ and for all languages L, it holds that $I^*_{L|a}(\mathsf{LF}(w)) = \mathsf{LF}(I^*_{L|a}(w))$. The inclusion $I^*_{L|a}(\mathsf{LF}(w)) \subseteq \mathsf{LF}(I^*_{L|a}(w))$ is clear. Conversely, let $w' \in \mathsf{LF}(I^*_{L|a}(w))$; there exists some word w'' such that $w \xrightarrow[I_{L|a}]{n} w'w''$ and we shall prove by induction on n that $w' \in I^*_{L|a}(\mathsf{LF}(w))$.

If $n = 0$, $w' \in \mathsf{LF}(w)$, else $w = \alpha\beta \xrightarrow[I_{L|a}]{} \alpha a\beta \xrightarrow[I_{L|a}]{n-1} w'w''$ for some words α and β with $\alpha \in L$. From the inductive hypothesis, there exist two words α' and β' such that $\alpha a\beta = \alpha'\beta'$ with $\alpha' \xrightarrow[I_{L|a}]{*} w'$. Let us consider two cases:

- $|\alpha| \geq |\alpha'|$: we directly get $w' \in I^*_{L|a}(\mathsf{LF}(w))$.

- $|\alpha| < |\alpha'|$: in this case, $\alpha' = \alpha a \beta'$ and $\beta = \beta'\beta''$ for some words β' and β''. From this follows $w = \alpha\beta'\beta''$ and we get $\alpha\beta' \xrightarrow[I_{L|a}]{} \alpha a\beta' = \alpha' \xrightarrow[I_{L|a}]{*} w'$ so $w' \in I^*_{L|a}(\mathsf{LF}(w))$.

Now, let $R = R_0 \cup R_1 \cup R_2 \cup R_3$ where $R_0 = (b^6a)^*b^6$, $R_1 = a^*$, $R_2 = \{a^i(b^2a)^jb^2 \mid i+j \text{ is odd}\}$ and $R_3 = \{a^i(b^2a)^jb^4a(b^6a)^p(b^5a)^qb^5 \mid j+q \text{ is odd}\}$. Let $K \subset b^*$ be an infinite language. We get $\mathsf{LF}(I^*_{R|a}(K)) = I^*_{R|a}(\mathsf{LF}(K)) = I^*_{R|a}(b^*)$ that is not a context-free language from Lemma 5.2; this implies $I^*_{R|a}(K) \notin \mathsf{CF}$. ☐

As a direct consequence of Corollary 5.5 and Corollary 5.6, we get:

Proposition 5.7. *For all languages $K \subseteq B^*$, the three following statements are equivalent:*

1. *There exists some rational language R such that $I^*_{R|a}(K)$ is not a context-free language.*
2. *There exists some rational language R such that $I^*_{R|a}(K)$ is not a rational language.*
3. *K is infinite.*

6. Simple rational control

This section is devoted to rationally controlled one-rule insertion systems in the case when the rational control language R is defined as $R = u^*$ for some word u. In particular we shall characterize when such systems correspond to a rational transduction and we shall prove that these systems preserve context-free languages. We begin this section with the study of the particular case $u \in a^*$ and $f \in a^+b^*$ for some distinct letters a and b.

Proposition 6.1. *For all natural numbers t, i and j,*

- *if $1 \le t \le i$ and $1 \le j$ then $I^*_{(a^t)^*|a^ib^j}(\varepsilon) \underset{\text{rat}}{\equiv} {D'_1}^*$,*
- *else $I^*_{(a^t)^*|a^ib^j}(\varepsilon) = (a^ib^j)^*$.*

Proof. Clearly, if $t > i$ or if $t = 0$ then $I^*_{(a^t)^*|a^ib^j}(\varepsilon) = I^*_{\varepsilon|a^ib^j}(\varepsilon) = (a^ib^j)^*$ and if $j = 0$ then $I^*_{(a^t)^*|a^i}(\varepsilon) = (a^i)^*$.

If $1 \le t \le i$ and $1 \le j$ then $i = st + r$ for some $s > 0$ and some $r < t$. Let $h : (a+b)^* \mapsto (a+b)^*$ be the morphism defined by $h(a) = a^t$ and $h(b) = a^rb^j$. Since $r < t$, the morphism h is prefix and we get $h(I^*_{a^*|a^sb}(\varepsilon)) = I^*_{(a^t)^*|a^{st}a^rb^j}(\varepsilon)$ from Lemma 4.8. It remains to prove that for all $i > 0$ it holds that $I^*_{a^*|a^ib}(\varepsilon) \underset{\text{rat}}{\equiv} {D'_1}^*$.

Thanks again to Lemma 4.8, respectively using the morphisms $g_1 : (a+b)^* \mapsto (a+b)^*$ defined by $g_1(a) = a$ and $g_1(b) = b^i$ and $g_2 : (a+b)^* \mapsto (a+b)^*$ defined by $g_2(a) = a^i$ and $g_2(b) = b^i$ we get the two following properties:

$$I^*_{a^*|a^ib}(\varepsilon) \underset{\text{rat}}{\equiv} I^*_{a^*|a^ib^i}(\varepsilon) \tag{6.1}$$

$$I^*_{(a^i)^*|a^ib^i}(\varepsilon) \underset{\text{rat}}{\equiv} I^*_{a^*|ab}(\varepsilon) \tag{6.2}$$

Moreover, we claim:

$$I^*_{a^*|a^ib^i}(\varepsilon) = I^*_{a^*|ab}(\varepsilon) \cap (a+b^i)^* \tag{6.3}$$

Indeed, the inclusion $I^*_{a^*|a^ib^i}(\varepsilon) \subseteq I^*_{a^*|ab}(\varepsilon) \cap (a+b^i)^*$ is clear and we can prove the converse inclusion by induction on the length of a derivation $\varepsilon \xrightarrow[I_{a^*|ab}]{*} w$ from ε to a word $w \in (a+b^i)^*$: if $w = \varepsilon$ then $w \in I^*_{a^*|a^ib^i}(\varepsilon)$, else $w = a^kb^iw'$ for some word w' and $k \ge i > 0$. From this follows $a^{k-i}w' \in I^*_{a^*|ab}(\varepsilon) \cap (a+b^i)^*$ and, from the inductive hypothesis, we get $a^{k-i}w' \in I^*_{a^*|a^ib^i}(\varepsilon)$ which implies $w \in I^*_{a^*|a^ib^i}(\varepsilon)$.

We also have:

$$I^*_{(a^i)^*|a^ib^i}(\varepsilon) = I^*_{a^*|a^ib^i}(\varepsilon) \cap (a^i + b)^* \tag{6.4}$$

Indeed, $I^*_{(a^i)^*|a^ib^i}(\varepsilon)$ is clearly included into $I^*_{a^*|a^ib^i}(\varepsilon) \cap (a^i + b)^*$. We prove the converse inclusion by induction on the length of a derivation $\varepsilon \xrightarrow[I_{a^*|a^ib^i}]{*} w$ from ε to a word $w \in (a^i + b)^*$: if $w = \varepsilon$ then $w \in I^*_{(a^i)a^*|a^ib^i}(\varepsilon)$, else

$$\varepsilon \xrightarrow[I_{a^*|a^ib^i}]{*} a^k w' \xrightarrow[I_{a^*|a^ib^i}]{} w = a^k a^i b^i w'$$

for some word w' and some natural number k. Moreover, since $w \in (a^i + b)^*$, we get $a^k \in (a^i)^*$ so $w \in I^*_{(a^i)^*|a^ib^i}(a^k w')$ and $a^k w' \in (a^i + b)^*$. From the inductive hypothesis, we get $a^k w' \in I^*_{(a^i)^*|a^ib^i}(\varepsilon)$ which implies $w = a^k a^i b^i w' \in I^*_{(a^i)^*|a^ib^i}(\varepsilon)$. Now, we get ${D'_1}^* = I^*_{a^*|ab}(\varepsilon)$, as seen in Example 3.2, and $I^*_{a^*|ab}(\varepsilon) \leadsto_{\text{rat}} I^*_{a^*|a^ib^i}(\varepsilon)$ from (6.3) so ${D'_1}^* \leadsto_{\text{rat}} I^*_{a^*|a^ib^i}(\varepsilon)$.

Conversely, $I^*_{a^*|a^ib^i}(\varepsilon) \leadsto_{\text{rat}} I^*_{(a^i)^*|a^ib^i}$ from (6.4) and $I^*_{(a^i)^*|a^ib^i} \underset{\text{rat}}{\equiv} {D'_1}^*$ from (6.2). That implies $I^*_{a^*|a^ib^i}(\varepsilon) \leadsto_{\text{rat}} {D'_1}^*$ so $I^*_{a^*|a^ib^i}(\varepsilon) \underset{\text{rat}}{\equiv} {D'_1}^*$. Finally, since $I^*_{a^*|a^ib^i}(\varepsilon) \underset{\text{rat}}{\equiv} I^*_{a^*|a^ib}(\varepsilon)$ from (6.1), it holds that ${D'_1}^* \underset{\text{rat}}{\equiv} I^*_{a^*|a^ib}(\varepsilon)$ that ends the proof of the proposition.

□

Given a one-rule insertion system $I_{R|f}$ over an alphabet A for some regular language R and some word f, the following language $\mathcal{R}_{R|f} = \{w \in A^* \mid I^*_{R|f}(w) \cap R \neq \emptyset\}$ will be very useful in order to study the properties of the system $I_{R|f}$. This is shown by the following elementary lemmas. The first one highlights the longest right factor that is never used in a derivation step.

Lemma 6.2. *Let $I_{R|f}$ be a one-rule insertion system over an alphabet A for some regular language R and some word f. For all derivations $w_0 \xrightarrow[I_{R|f}]{} w_1 \dots \xrightarrow[I_{R|f}]{} w_n$ with $n > 0$, there exists a word β such that for all integers $j \in [0, n], w_j = \alpha_j \beta$ for some word α_j with for all integers $k \in [0, n[, \alpha_k \xrightarrow[I_{R|f}]{} \alpha_{k+1}$. Moreover there exists some integer $i \in [0, n[$ with $\alpha_i \in R \wedge \alpha_{i+1} = \alpha_i f$ and $\alpha_0 \in \mathcal{R}_{R|f}$.*

Proof. Clearly, $\alpha_0 \in \mathcal{R}_{R|f}$ is a consequence of the other properties. If $w_0 \xrightarrow[I_{R|f}]{} w_1 \dots \xrightarrow[I_{R|f}]{} w_n$ with $n > 0$, then there exist words $\alpha'_0, \dots, \alpha'_n$, $\beta'_0, \dots, \beta'_n$ such that $w_0 = \alpha'_0\beta'_0$, $w_n = \alpha'_n\beta'_n$ and for all $0 \leq j < n$, $\alpha'_j \in R$ and $\alpha'_j f \beta'_j = \alpha'_{j+1}\beta'_{j+1}$. These factorisations correspond to each application of the rewrite rule in the derivation step. Let $i \in [0 \dots n[$ such that β'_i is the shortest word in $\{\beta'_0, \dots, \beta'_{n-1}\}$ then we can take $\beta = \beta'_i$ and for all integer $j \in [0, n], \alpha_j = w_j\beta^{-1}$. □

A simpler statement of this property is the following:

Corollary 6.3. *For all words $w \in A^*$ and $w' \in I^*_{R|f}(w)$, there exist some words $w_1, w'_1, w_2 \in A^*$ and $w''_1 \in R$ such that $w = w_1 w_2$, $w' = w'_1 w_2$, $w''_1 \in I^*_{R|f}(w_1)$ and $w'_1 \in I^*_{R|f}(w''_1)$.*

We also get as a corollary of Lemma 6.2:

Corollary 6.4. *For all words $w \in A^*$, $I^*_{R|f}(w) = I^*_{R|f}(w_1)w_2$ with $w = w_1 w_2$ where w_1 is the longest left factor of w that belongs to $\mathcal{R}_{R|f}$.*

When $R = u^*$ for some word u, we observe that $\mathcal{R}_{u^*|f} = \mathcal{R}^*_{u^*|f}$ since $\varepsilon \in u^*$. We will now prove, in this case $R = u^*$, that it is possible to build a code C such that $\mathcal{R}_{u^*|f} = C^*$: let $I_{R|f}$ be a one-rule insertion system over an alphabet A for some regular language R and some word f. Let $E_1 = \{e \in A^* \mid |e| < |u| \wedge fe \in u^*\}$ and, for every $i > 0$, $E_{i+1} = E_i \cup \{e \in \mathsf{RF}(u) \mid fe \in (E_i + u)^* e', (e' \in E_i + u) \wedge (|e'| > |e|)\}$. We define $E = \cup_{i>0} E_i$

and $E_u = E \cup \{u\}$. It is easily seen that, for every $i > 0$, $E_i \subseteq E_{i+1} \subseteq \mathsf{RF}(u)$ so there exists $k > 0$ such that $E = E_k = E_{k+1}$. From the inductive construction of E we get:

Lemma 6.5.

1. $E_u^* \subseteq \mathcal{R}_{u^*|f}$,
2. $E = \{e \in A^* \mid \exists e' \in E_u, |e| < |e'| \wedge fe \in E_u^* e'\}$,
3. $fw \in E_u^*$ *if and only if* $w \in EE_u^*$.

Proof.

1. Since $\mathcal{R}_{u^*|f} = \mathcal{R}_{u^*|f}^*$, it is sufficient to prove $E_u \subseteq \mathcal{R}_{u^*|f}$ that is to prove $E \subseteq \mathcal{R}_{u^*|f}$ and we easily get by induction that for all $i > 0$ it holds that $E_i \subseteq \mathcal{R}_{u^*|f}$.
2. Let $L = \{e \in A^* \mid \exists e' \in E_u, |e| < |e'| \wedge fe \in E_u^* e'\}$. Clearly, for all $i > 0$, $E_i \subseteq L$. Conversely, let $e \in L$, then there exists some $e' \in E_u$ with $|e'| > |e|$ and $fe \in E_u^* e'$. If $fe \in u^*$ then $e \in E_1 \subseteq E$. Else $fe = e_1 \cdots e_n$ such that for all $i \in [1..n]$, it holds that $e_i = u$ or $e_i \in E_{k_i}$ for some k_i and $e_n = e'$. Let t be the biggest index among the $k_i's$, we can assume that t exists else $fe \in u^*$. From this follows $e' \in E_t$ so $e \in E_{t+1} \subseteq E$.
3. From 2, we only have to prove that $fw \in E_u^*$ implies $w \in EE_u^*$. If $fw \in E_u^*$ then $f = f'f''$ and $w = w'w''$ with $f' \in E_u^*$, $f'' \neq \varepsilon$, $w'' \in E_u^*$ and $f''w' \in E_u$. From this follows $fw' \in E_u^*(f''w')$ and, since $w' \in \mathsf{RF}(u)$, we get $w' \in E_u$. Moreover $|w'| < |u|$ since $|f''| > 0$ so $w' \in E$ and $w \in EE_u^*$.

□

We shall now prove the converse inclusion $\mathcal{R}_{u^*|f} \subseteq E_u^*$. The following lemma highlights the different cases that must be considered.

Lemma 6.6. *Let* $u, f \in A^+$. *Then*

1. *either* $uf = fu$,
2. *or* $f = u^s z$ *with* $\{u, z\}$ *prefix for some integer* $s > 0$,
3. *or* $u = f^i z$ *with* $\{z, f\}$ *prefix for some integer* $i \geq 0$,
4. *or* $\{u, f\}$ *is suffix.*

Proof. Let $u \in A^+$ and $f \in A^+$ such that $uf \neq fu$. Assume first $u \notin \mathsf{LF}(f^*)$; we get $u = f^i z$ with $\{z, f\}$ prefix for some integer $i \geq 0$. Otherwise, if $u \in \mathsf{LF}(f^*)$, we can first observe that $|u| \neq |f|$ else this implies $uf = fu$, a contradiction. Hence, we consider two cases:

- $|u| > |f|$. In this case, $f = f'f''$ for some words f' and f'' and $u = f^i f'$ with $i > 0$. Since $uf \neq fu$ we get $f' \neq \varepsilon$ and $f'f'' \neq f''f'$ so $f \notin \mathsf{RF}(ff')$ which implies $\{u, f\}$ suffix.
- $|f| > |u|$. In this case $f = u^s z$ for some word z with $u \notin \mathsf{LF}(z)$ and some natural number s. Observe that $s > 0$ since $u \in \mathsf{LF}(f^*)$ and $z \neq \varepsilon$ else $uf = fu$. Assume that $\{u, z\}$ is not prefix. We get $z \in \mathsf{LF}(u)$ and $u = zz'$ for some word z'. From $uf \neq fu$ follows $zz' \neq z'z$ so $u \notin \mathsf{RF}(uz)$ which implies $\{u, f\}$ suffix.

□

We can observe that the different cases that appear in the previous lemma are not necessarily exclusive to each other.

Lemma 6.7.

If $u \in A^*bd$ and $f \in A^*ad$ for some word d and some distinct letters a and b then:

1. $dE_u \subseteq A^*bd$,
2. $adE \subseteq \mathsf{RF}(du)$.

Proof. From $fE_1 \subseteq u^*$ and $E_1 \subseteq \mathsf{RF}(u)$ follows $adE_1 \subseteq \mathsf{RF}(du)$ which implies $dE_1 \subseteq A^*bd$. By induction on the construction of E, we get $adE \subseteq \mathsf{RF}(du)$ and $dE \subseteq A^*bd$. □

We can now prove the crucial following lemma:

Lemma 6.8. *Let $u, f \in A^+$. For every word $w \in A^*$, if $uw \in E_u^*$ then $w \in E_u^*$.*

Proof. From Lemma 6.6 we can consider four cases.

1. $uf = fu$. In this case, $u^p = f^q$ for some strictly positive integers p and q. If $uw \in E_u^*$ then $u^p w \in E_u^*$ which implies $f^q w \in E_u^*$ and we get $w \in E_u^*$ from Item 3 of Lemma 6.5.
2. $f = u^s z$ with $\{u, z\}$ prefix for some integer $s \geq 0$. The property is clearly true since in this case $E_1 = \emptyset$ so $E_u^* = u^*$.
3. $u = f^i z$ with $\{z, f\}$ prefix for some integer $i \geq 0$. It is easily seen that in this case, $E = \cup_{0 \leq j < i}\{f^j z\}$ so $E_u = \cup_{0 \leq j \leq i}\{f^j z\}$, a prefix code and this implies the property.
4. $\{u, f\}$ is suffix. In this case $u \in A^* bd$ and $f \in A^* ad$ for some word d and some distinct letters a and b. Let $v = du$ and $D = \{\alpha \mid ad\alpha \in \mathsf{RF}(v)\}$. Observe that $E \subseteq D$ from Lemma 6.7; on the other hand it has been proved in [11] (Prop. 4) that $vD^* \cap A^+ v A^* = \emptyset$. Now, assume $uw \in E_u^*$ and $w \notin E_u^*$. From this follows $uw \in EE_u^*$ which implies $uw' \in E^*$ for some word w': indeed if $uw \in EE_u^*$, then there exist words w'', α, β and γ such that $uw = u\alpha w''$ with $u = \beta\gamma$, $\beta \in E^*$ and $\gamma\alpha \in E_u$. If $\gamma\alpha \in E$ then we can take $w' = \alpha$ to get $uw' \in E^*$, else $u\alpha = \beta u$ which implies $u\alpha \in \mathsf{LF}(\beta^*) \subseteq \mathsf{LF}(E^*)$ so there exists a word α' such that $u\alpha\alpha' \in E^*$ and we can take $w' = \alpha\alpha'$ to get $uw' \in E^*$. Finally, $uw' \in E^*$ leads to a contradiction: $vuw' \in vD^*$ and $vuw' \in A^+ v A^*$ since $vu = duu \in A^* bduA^* = A^* bvA^*$.

□

As a consequence, we get:

Proposition 6.9. *It holds that $\mathcal{R}_{u^*|f} = E_u^*$. Moreover*

1. *if $uf \neq fu$ then E_u is a code and $f \notin E_u^*$,*
2. *else $E_u^* = x^*$ where x is the longest word such that u and f are in x^*.*

Proof. From Lemma 6.5 we only have to prove $\mathcal{R}_{u^*|f} \subseteq E_u^*$. The proof is an induction on the length of the derivation from a word in $\mathcal{R}_{u^*|f}$ to a word in u^*. If this length is null then $w \in u^* \subseteq E_u^*$. Else $w = u^i\alpha \xrightarrow[I_{u^*|f}]{} u^i f\alpha \xrightarrow[I_{u^*|f}]{*} w'$ for some natural number i and some $w' \in u^*$. From the inductive hypothesis follows $u^i f\alpha \in E_u^*$. This implies $f\alpha \in E_u^*$ from Lemma 6.8 and $\alpha \in E_u^*$ from Item 3 of Lemma 6.5. Finally we get $w = u^i\alpha \in E_u^*$.

- To prove 1, from Lemma 6.6 we can consider three cases:
  - $f = u^s z$ with $\{u, z\}$ prefix for some integer $s \geq 0$. In this case $E_1 = \emptyset$ so $E_u = \{u\}$.
  - $u = f^i z$ with $\{z, f\}$ prefix for some integer $i \geq 0$. In this case, $E = \cup_{0 \leq j < i}\{f^j z\}$ so $E_u = \cup_{0 \leq j \leq i}\{f^j z\}$, a prefix code that does not contain f.
  - $\{u, f\}$ is suffix. In this case $u \in A^* bd$ and $f \in A^* ad$ for some word d and some distinct letters a and b. Let $w \in E_u^+$ and assume $w = \alpha e = \beta e'$ for some $e, e' \in E_u$ with $|e'| < |e|$. From this follows $e' \in E$ and there exists a word $e'' \in E_u$ such that $w \in A^* e'' e'$. From Lemma 6.7, we get $de'' \in A^* bd$ so $bde' \in \mathsf{RF}(de) \subseteq \mathsf{RF}(du)$. This leads to a contradiction since $ade' \in \mathsf{RF}(du)$ from Item 2 of Lemma 6.7.
- To prove 2, let x be the longest word such that u and f belong to x^*. From the definition of E follows $E_u \subseteq x^*$. Conversely, let $y = x^k$ for some $k > 0$ be the shortest word in $E_u \setminus \{\varepsilon\}$. Necessarily, $f \in y^*$, else $f = y^i x^{k'}$ for some integer i and some integer k' with $0 < k' < k$ which implies $x^{k'}$ in E_u, a contradiction. Assume $E_u \not\subseteq y^*$ and let e be the shortest word in E_u such that $e \notin y^*$. There are two cases:
  1. $|e| < |f|$: in this case, $f = ey^i x^{k'}$ with $0 < k' < k$ that leads again to a contradiction.
  2. $|e| > |f|$: in this case, $e = fe'$ which implies $e' \in E_u$ from Item 2 of Lemma 6.5. Moreover $e' \notin y^*$ and is shorter than e, a contradiction.

  So $E_u \subseteq y^*$. In particular, $u \in y^*$. That implies $k = 1$: indeed, $u = x^p$ and $f = x^q$ for some integers p and q that are prime between them else x is not the longest word such that $u \in x^*$ and $f \in x^*$. Finally, $x = y$ so $x \in E_u$.

□

As a consequence, we get:

Lemma 6.10. *If $uf \neq fu$ then*

1. $u^*fuA^* \cap \mathcal{R}_{u^*|f} = \emptyset$,
2. $u^*f^+ \cap \mathcal{R}_{u^*|f} = \emptyset$,
3. *for all natural numbers p and for all words w,* $I^*_{u^*|f}(u^pfuw) = I^*_{u^*|f}(u^pfu)w$,
4. *for all words $w \in A^*$,* $I^*_{u^*|f}(uw) = I^*_{u^*|f}(u)I^*_{u^*|f}(w)$.

Proof.

1. Assume $u^ifuw \in \mathcal{R}_{u^*|f} = E^*_u$ for some natural number i and some word w. By induction on i, we get $fuw \in E^*_u$ from Lemma 6.8. Moreover, from Item 3 of Lemma 6.5, we get $uw \in EE^*_u$. That implies $uw \in E^*_u$ so $w \in E^*_u$ from Lemma 6.8. We finally get $uw \in EE^*_u \cap uE^*_u$ that contradicts Item 1 of Proposition 6.9.
2. Assume $u^if^j \in \mathcal{R}_{u^*|f} = E^*_u$ for some natural number i and some strictly positive integer j. By induction on i, we get $f^j \in E^*_u$ from Lemma 6.8 and by induction on j we get $f \in E^*_u$ from Item 3 of Lemma 6.5 that contradicts Item 1 of Proposition 6.9.
3. The inclusion $I^*_{u^*|f}(u^pfu)w \subseteq I^*_{u^*|f}(u^pfuw)$ is clear. Conversely, from Corollary 6.4 follows $I^*_{R|f}(u^pfuw) = I^*_{R|f}(w_1)w_2$ with $u^pfuw = w_1w_2$ where w_1 is the longest left factor of u^pfuw that belongs to $\mathcal{R}_{R|f}$. From Item 1, $|w_1| < |u^pfu|$ so $I^*_{u^*|f}(u^pfuw) \subseteq I^*_{u^*|f}(u^pfu)w$.
4. $I^*_{u^*|f}(u)I^*_{u^*|f}(w)$ is clearly included into $I^*_{u^*|f}(uw)$. Conversely, the proof is an induction on the length of a derivation from uw to some word w'. By considering the first step of the derivation, one can distinguish two cases:
   - $uw \xrightarrow[I_{u^*|f}]{u} w'' \xrightarrow[I_{u^*|f}]{*} w'$ with $w \xrightarrow[I_{u^*|f}]{} w''$. From the inductive hypothesis, we get $w'' \in I^*_{u^*|f}(u)I^*_{u^*|f}(w'') \subseteq I^*_{u^*|f}(u)I^*_{u^*|f}(w)$.
   - $uw \xrightarrow[I_{u^*|f}]{} fuw \xrightarrow[I_{u^*|f}]{*} w'$. We get from Item 3 that $w' = w''w$ with $w'' \in I^*_{u^*|f}(fu) \subseteq I^*_{u^*|f}(u)$ so $w' \in I^*_{u^*|f}(u)I^*_{u^*|f}(w)$.

□

We can observe that Item 3 and Item 4 of Lemma 6.10 remain true in the case $uf = fu$ unlike items 1 and 2. To finish the preliminary results of this section, we shall now prove a stronger version of Lemma 6.2 in the case when $R = u^*$ and $uf \neq fu$. In this case, the word β of Lemma 6.2 only depends on the words w_0 and w_n; in particular that implies by induction on the length of derivations the unicity of the derivation from a word w to any word of $I^*_{u^*|f}(w)$. We need first:

Lemma 6.11. *Let u and f be two words with $uf \neq fu$. For all words $w \in E^*_u$ there exists a unique natural number n such that $u^n \in I^*_{u^*|f}(w)$.*

Proof. The proof is an induction on the length of a shortest derivation from $w \in E^*_u$ to some word of u^*. If $w = u^n$ for some n we get from Item 1 of Lemma 6.10 that for all $p \neq n$, $u^p \notin I^*_{u^*|f}(w)$. Else $w \xrightarrow[I_{u^*|f}]{} w_1 \xrightarrow[I_{u^*|f}]{*} u^n$ with $w = u^k\alpha$ and $w_1 = u^kf\alpha$. Since $w_1 \in E^*_u$ we get from Item 1 of Lemma 6.10 that $\alpha \notin uA^*$. That implies that w_1 is uniquely defined and, from the inductive hypothesis, n is unique. □

In the following, thanks to this lemma, when u and f satisfy $uf \neq fu$, for all words $w \in E^*_u$, we shall denote by $\varphi(w)$ this unique word that belongs to $I^*_{u^*|f}(w) \cap u^*$. We observe that, for all words $w_1, w_2 \in E^*_u$, $\varphi(w_1w_2) = \varphi(w_1)\varphi(w_2)$.

We also need the following lemma that states that for all prefix sets P and for all distincts words x, y in P, to give two words $w \in P^*x\beta$ and $w' \in P^*y\beta$ uniquely defines the word β. More precisely:

Lemma 6.12. *Let $P \subseteq A^*$ be a prefix set and x and y be two distinct words in P. For all words $w, w' \in A^*$, there exists at most one word β such that $w\beta^{-1} \in P^*x$ and $w'\beta^{-1} \in P^*y$.*

Proof. Assume there exists two distinct words β_1 and β_2 such that $w \in P^*x\beta_1 \cap P^*x\beta_2$ and $w' \in P^*y\beta_1 \cap P^*y\beta_2$ and assume $|\beta_2| < |\beta_1|$. Then $x\beta_1 \in P^+x\beta_2$ and, since $x \in P$ which is a prefix code, we get $\beta_1 \in P^*x\beta_2$. Similarly, we get $\beta_1 \in P^*y\beta_2$, a contradiction since $P^*x \cap P^*y = \emptyset$. ☐

We can now state:

Lemma 6.13. *For all words u, f, w and $w' \neq w \in I^*_{u^*|f}(w)$, there exist words w''_1 and w''_2 such that, for all derivations $w = w_0 \xrightarrow[I_{u^*|f}]{} w_1 \dots \xrightarrow[I_{u^*|f}]{} w_n = w'$, there exists some index $i \in [0 \dots n[$ such that $w_i = w''_1$ and $w_{i+1} = w''_2$.*

Proof. Observe first that, if $uf = fu$, we can take $w''_1 = w$ and $w''_2 = u^k f w_2$ where $w = u^k w_2$ with $w_2 \notin uA^*$ so we assume $uf \neq fu$. From Lemma 6.2, there exists a word β with:

- For all integer $j \in [0, n], w_j = \alpha_j\beta$ for some word α_j with for all integer $k \in [0, n[, \alpha_k \xrightarrow[I_{u^*|f}]{} \alpha_{k+1}$,
- there exists some integer $i \in [0, n[$ with $\alpha_i \in u^* \wedge \alpha_{i+1} = \alpha_i f$,
- $\alpha_0 \in E^*_u$.

We first show that β only depends on w and w' and not on the derivation itself. From Lemma 6.6, we can consider three cases:

1. $u \in A^*bd$ and $f \in A^*ad$ for some word d and some distinct letters a and b: from $\alpha_0 \in E^*_u$ and thanks to Item 1 of Lemma 6.7, we get $d\alpha_0 \in A^*bd$ which implies $dw_0 \in A^*bd\beta$. Moreover, $\alpha_n \in I^*_{u^*|f}(\alpha_i f) \subseteq I^*_{u^*|f}(A^*ad) = A^*ad$ so $w_n \in A^*ad\beta$. Hence β is uniquely defined from these two properties : $dw \in A^*bd\beta$ and $w' \in A^*ad\beta$.
2. $u = f^i z$ with $P = \{f, z\}$ prefix. In this case, $E^*_u = \{z, fz, \dots f^i z\}^*$ so $z\alpha_0 \in P^*z$ and $zw \in P^*z\beta$. Moreover, $\alpha_n \in I^*_{u^*|f}(u^*f) \subseteq I^*_{u^*|f}(P^*f) = P^*f$ so $w' \in P^*f\beta$. From Lemma 6.12 we get that β is unique.
3. $f = u^s z$ with $P = \{u, z\}$ prefix. In this case, $E^*_u = u^*$ so $u\alpha_0 \in P^*u$ which implies $uw \in P^*u\beta$. Moreover, $\alpha_n \in I^*_{u^*|f}(u^*f) \subseteq I^*_{u^*|f}(P^*z) = P^*z$ so $w' \in P^*z\beta$ and, from Lemma 6.12, we get that β is unique.

Finally, from the uniqueness of β, we can define $w''_1 = \varphi(w\beta^{-1})\beta$ and $w''_2 = \varphi(w\beta^{-1})f\beta$. ☐

By induction, we can deduce from this lemma:

Proposition 6.14. *For all words w and $w' \in I^*_{u^*|f}(w)$, there exists a unique derivation from w to w'.*

We can now address the main results of this section that is to characterize, given two words u and f, when $I^*_{u^*|f}$ is rational and to prove that these systems preserve context-free languages.

Lemma 6.15. *If $uf = fu$ then for all words w, $I^*_{u^*|f}(w) = f^*w$.*

Proof. Clearly, in this case, for all words $w, w' \in A^*$, if $w \xrightarrow[I_{u^*|f}]{} w'$ then $w' = fw$ and, by induction we get that $I^*_{u^*|f}(w) = f^*w$. ☐

So, when $uf = fu$, $I^*_{u^*|f}$ is rational. It is also the case when $u \notin \mathsf{LF}(f^*)$:

Lemma 6.16. *If $u \notin \mathsf{LF}(f^*)$ then $I^*_{u^*|f}$ is rational and $L_{u^*|f} = f^*$.*

Proof. If $u \notin \mathsf{LF}(f^*)$ then $u = f^i z$ for some $i \geq 0$ and some word z with $\{f, z\}$ being a prefix set. In this case, $E^*_u = (\cup_{0 \leq j \leq i}\{f^j z\})^*$. Let B be the alphabet $B = \cup_{0 \leq j \leq i}\{b_j\}$ and $h : (B \cup A)^* \mapsto A^*$ be the morphism defined by $h(b_j) = f^j z$ for all $b_j \in B$ and $h(a) = a$ for all $a \in A$. Let $s : (B \cup A)^* \mapsto A^*$ be the rational substitution defined by $s(b_j) = f^*uf^*$ for all $b_j \in B$ and $s(a) = a$ for all $a \in A$.

For all $w \in A^*$, it holds that $I^*_{u^*|f}(w) = f^*s(h^{-1}(w) \cap B^*A^*)$: indeed we clearly have $f^*s(h^{-1}(w) \cap B^*A^*) \subseteq I^*_{u^*|f}(w)$; conversely, from Lemma 6.2, if $w' \in I^*_{u^*|f}(w)$, there exists $\alpha \in E^*_u$, $\alpha', \beta \in A^*$ such that $w = \alpha\beta$, $w' =$

$\alpha'\beta$ and $\alpha' \in I^*_{u^*|f}(\varphi(\alpha))$. Since $E^*_u = (\cup_{0\le j\le i}\{f^j z\})^*$, we get $\alpha' \in (h^{-1}(\alpha) \cap B^*)$ so $w' \in f^* s(h^{-1}(w) \cap B^* A^*)$. In particular, $L_{u^*|f} = f^*$. ☐

We observe that if $f \in a^*$ for some letter a then $u \in \mathsf{LF}(f^*)$ implies $u \in a^*$ so $uf = fu$ and we get as a corollary:

Corollary 6.17. *If $f \in a^*$ for some letter a then $I^*_{u^*|f}$ is rational.*

This result does not hold when $u \in a^*$: indeed we have seen before that $I^*_{a^*|ab}(\varepsilon) = {D'_1}^*$. More generally, for all $s \ge 0$, let g_s be the morphism from $(a+b)^*$ to $(a+b)^*$ defined by $g(a) = a^s$ and $g(b) = b$ then it holds that $g^{-1}(I^*_{a^*|a^s b}(\varepsilon)) = {D'_1}^*$. As a matter of fact, we can state:

Lemma 6.18. *If $uf \ne fu$ and $u \in \mathsf{LF}(f^*)$ then $L_{u^*|f} \notin \mathsf{RAT}$.*

Proof. From Lemma 6.6, we can consider two cases: indeed the case $u = f^i z$ with $\{f, z\}$ prefix is not possible since $u \in \mathsf{LF}(f^*)$.

1. $f = u^s z$ for some integer $s > 0$ and some word z with $\{u, z\}$ prefix. Let $h : (a+b)^* \mapsto A^*$ be the (prefix) morphism defined by $h(a) = u$ and $h(b) = z$. From Lemma 4.8 we get $I^*_{a^*|a^s b}(\varepsilon) = h^{-1}(L_{u^*|f})$ and from $I^*_{a^*|a^s b}(\varepsilon) \notin \mathsf{RAT}$ follows $L_{u^*|f} \notin \mathsf{RAT}$.
2. $\{u, f\}$ is a suffix set. In this case, $u \in A^*bd$ and $f \in A^*ad$ for some word d and some distinct letters a and b. Let us consider the factorization $f = \alpha z$ where α is the longest left factor of f that belongs to E^*_u and let $\varphi(\alpha) = u^n$. Observe that from $u \in \mathsf{LF}(f^*)$ follows $n > 0$ so $f \ne z$ and $\varphi(\alpha) \ne \varepsilon$. Let L be the nonregular language $L = \{u^{tn} z^t \mid t \ge 0\}$; we shall prove $L_{u^*|f} \cap u^* z^* = L$ which implies $L_{u^*|f} \notin \mathsf{RAT}$.
   Clearly, $L \subseteq L_{u^*|f} \cap u^* z^*$. Conversely, we shall prove by induction over the length of the derivation that $\varepsilon \xrightarrow[I_{u^*|f}]{*} wz^t$ for some word $w \in E^*_u$ implies $\varphi(w) z^t \in L$. It is clearly true if $wz^t = \varepsilon$ else $\varepsilon \xrightarrow[I_{u^*|f}]{*} u^s w' \xrightarrow[I_{u^*|f}]{} u^s f w'$ for some natural number s and some word w' with $u^s f w' \in E^*_u z^*$ so $u^s f w' = w z^t$ with $w \in E^*_u$.
   If $w' = z^k$ for some natural number k, we get $u^s w' = u^s z^k$ and from the inductive hypothesis follows $s = kn$. Now $u^s f w' = u^s \alpha z^{k+1}$ with $\varphi(u^s \alpha) = u^{s+n} = u^{(k+1)n}$ which implies $u^s f w' \in L$. So we assume in the following $w' \notin z^*$ and we consider two cases:
   (a) $|z^t| < |w'|$. In this case, $w = u^s f w''$ for some word $w'' \ne \varepsilon$ since $w' \notin z^*$. We get $f w'' \in E^*_u$ from Lemma 6.8 so $w'' \in E E^*_u$ from Lemma 6.5. Moreover, since $w'' \notin uA^*$, we get $\varphi(f w'') = \varphi(w'')$ so $\varphi(u^s f w'') z^t = \varphi(u^s w'') z^t \in L$.
   (b) $|z^t| > |w'|$. One can distinguish two sub-cases:
      i. $|zw'| > |z^t|$. In this case, $w = u^s \alpha z_1$ with $z_1 \ne \varepsilon$ and αz_1 a left factor of f. Since $w \in E^*_u$, from Lemma 6.8 follows $\alpha z_1 \in E^*_u$, a contradiction with the definition of α.
      ii. $|zw'| < |z^t|$. In this case, $z = z_1 z_2$ for some nonempty suffix z_1 of z and some nonempty left factor z_2 of z and $u^s \alpha = \beta z_1$ for some word $\beta \in E^*_u z^*$. Moreover, since $\alpha \in E^*_u$, we get from Item 1 of Lemma 6.7 that $d u^s \alpha \in A^* bd$.
      On the other hand, let r be the root of z. We get $z_1 = r^i$ and $z_2 = r^j$ for some positive integers i and j. Since $dz = dr^{i+j} \in A^* ad$, we get $d \in \mathsf{RF}(r^*)$. That implies $dr^* \subseteq \mathsf{RF}(r^*)$ and we get $dr \in \mathsf{RF}(dr^{i+j}) \subseteq \mathsf{RF}(A^* ad)$ so $dz_1 \in A^* ad$. This leads to a contradiction: $d u^s \alpha \in A^* bd$ and $d\beta z_1 \in A^* ad$ but $d u^s \alpha = d \beta z_1$.

☐

We shall now prove that for all words w, $I^*_{u^*|f}(w) \rightsquigarrow_{\text{rat}} L_{u^*|f}$. We will need the two following lemmas that hold in the case when $\{u, f\}$ is a suffix set.

Lemma 6.19. *Assume $u \in A^* bd$ and $f \in A^* ad$ for some word d and some distincts letters a and b, then for all words $w \in A^*$ and $\alpha, \alpha' \in (A^* d)^*$, if $\alpha' w \in I^*_{u^*|f}(\alpha w)$ then $\alpha' \in I^*_{u^*|f}(\alpha)$.*

Proof. The proof is an induction on the length of the derivation $\alpha w \xrightarrow[I_{u^*|f}]{*} \alpha' w$. If this length is 0 then $\alpha = \alpha'$ else $\alpha w = u^i w' \xrightarrow[I_{u^*|f}]{} u^i f w' \xrightarrow[I_{u^*|f}]{*} \alpha' w$ and $\alpha' \neq \varepsilon$. If $i = 0$ then $u^i f w' = f\alpha w$ and, from the inductive hypothesis, we get $f\alpha \xrightarrow[I_{u^*|f}]{*} \alpha'$ hence $\alpha \xrightarrow[I_{u^*|f}]{} f\alpha \xrightarrow[I_{u^*|f}]{*} \alpha'$. Else we have to consider two cases:

- $|u^i| < |\alpha|$. In this case, $\alpha = u^i\alpha''$ and $w' = \alpha'' w$ for some word α'' that satisfies $d\alpha'' \in A^*d$. Then we get $u^i\alpha'' w \xrightarrow[I_{u^*|f}]{} u^i f\alpha'' w \xrightarrow[I_{u^*|f}]{*} \alpha' w$. From the inductive hypothesis, we get $u^i f \alpha'' \xrightarrow[I_{u^*|f}]{*} \alpha'$ so $\alpha \xrightarrow[I_{u^*|f}]{} u^i f\alpha'' \xrightarrow[I_{u^*|f}]{*} \alpha'$.
- $|u^i| > |\alpha|$. In this case, $u^i = \alpha w''$ and $w = w'' w'$ for some word $w'' \neq \varepsilon$. Since $\alpha \in (A^*d)^*$ we get $dw'' \in A^*bd$ and from $\alpha' \in A^*d$ follows $\alpha' w'' \in A^*bd$. From the derivation $\alpha w \xrightarrow[I_{u^*|f}]{} u^i f w' \xrightarrow[I_{u^*|f}]{*} \alpha' w'' w'$ we get from the inductive hypothesis $u^i f \xrightarrow[I_{u^*|f}]{*} \alpha' w''$ and this leads to a contradiction with Lemma 3.20 since $u^i f \in A^*ad$ and $\alpha' w'' \in A^*bd$.

□

In the general case, for all words w, it holds that $L_{u^*|f} \subseteq I^*_{u^*|f}(w)w^{-1}$ but the converse is not always true: for instance if $u = b$ and $f = ab$ for some distinct letters a and b, we get $bab \in I^*_{u^*|f}(b)$ but $ba \notin L_{u^*|f}$. Nevertheless, we can state:

Proposition 6.20. *For all u and f, there exists a rational language K such that for all words w, $L_{u^*|f} = I^*_{u^*|f}(w)w^{-1} \cap K$.*

Proof. Thanks to Lemma 6.6, Lemma 6.15 and Lemma 6.16, we can consider three cases:

1. $uf = fu$ or $u \notin \mathsf{LF}(f^*)$. In these cases, $L_{u^*|f} = f^*$ so we can take $K = f^*$: indeed, for all words w, it holds that $L_{u^*|f} \subseteq I^*_{u^*|f}(w)w^{-1}$ so $L_{u^*|f} \subseteq I^*_{u^*|f}(w)w^{-1} \cap f^*$ and conversely $I^*_{u^*|f}(w)w^{-1} \cap f^* \subseteq f^* = L_{u^*|f}$.
2. $f = u^s z$ for some $s > 0$ with $P = \{u, z\}$ prefix. In this case, we can take $K = P^*$: indeed, clearly $I^*_{u^*|f}(P^*) = P^*$ and $I^*_{u^*|f}(P^*z) = P^*z$. From this follows $L_{u^*|f} \subseteq P^*$ so $L_{u^*|f} \subseteq I^*_{u^*|f}(w)w^{-1} \cap P^*$. Conversely, observe first that, clearly, for all $x \in P^*z$ and for all $y \in A^*$, it holds that $I^*_{u^*|f}(xy) = I^*_{u^*|f}(x)y$. Now, if we consider a derivation $w \xrightarrow[I_{u^*|f}]{*} \alpha w$ for some $\alpha \in P^*$, either $\alpha = \varepsilon \in L_{u^*|f}$ or $w = u^i w' \xrightarrow[I_{u^*|f}]{} u^i f w' \xrightarrow[I_{u^*|f}]{*} \alpha w = \alpha u^i w'$. If $i > 0$ then $\alpha w \in P^*zw' \cap P^*uw'$, a contradiction so $i = 0$ and $\alpha \in I^*_{u^*|f}(f) \subseteq L_{u^*|f}$.
3. $\{u, f\}$ is suffix. In this case, $f \in A^*ad$ and $u \in A^*bd$ for some word d and some distinct letters a and b and let $K = (A^*d)^*$.
   - $L_{u^*|f} \subseteq (I^*_{u^*|f}(w)w^{-1}) \cap K$: the inclusion $L_{u^*|f} \subseteq (I^*_{u^*|f}(w)w^{-1})$ is always satisfied and, clearly, $L_{u^*|f} \setminus \{\varepsilon\} = I^*_{u^*|f}(f) \subseteq A^*d$.
   - $(I^*_{u^*|f}(w)w^{-1}) \cap K \subseteq L_{u^*|f}$: if $w \xrightarrow[I_{u^*|f}]{*} \alpha' w$ with $\alpha' \in K$, then either $\alpha' = \varepsilon \in L_{u^*|f}$ or $\alpha' \in A^*d$ and $\alpha' \in L_{u^*|f}$ from Lemma 6.19 by taking $\alpha = \varepsilon$.

□

By combining the previous results of this section, we can now precisely characterize when $I^*_{u^*|f}*$ is rational:

Proposition 6.21. *The following statements are equivalent:*

1. $L_{u^*|f} \in \mathsf{RAT}$.
2. $\exists w \in A^* \mid I^*_{u^*|f}(w) \in \mathsf{RAT}$.

3. $u \notin \mathsf{LF}(f^*) \vee uf = fu$.
4. $L_{u^*|f} = f^*$.
5. $I^*_{u^*|f}$ *is rational.*

Proof. Observe that implications between some of these statements are clear: 4 implies 1, 5 implies 1, 1 implies 2. Moreover, Lemma 6.15 and Lemma 6.16 prove that statement 3 implies both statement 4 and statement 5. Hence it remains to prove that statement 2 implies statement 3 to complete the proof. That is indeed the case since Proposition 6.20 proves that statement 2 implies statement 1 and Lemma 6.18 proves that statement 1 implies statement 3. □

The rest of the section is devoted to the proof of the fact that $I^*_{u^*|f}$ preserves the context-free languages, in other words $I^*_{u^*|f}$ is CF/CF. We first state:

Proposition 6.22. *$I^*_{u^*|f}(u)$ and $L_{u^*|f}$ are two context-free languages.*

Proof. We define the following set $\Delta = \Delta_1 \cup \Delta_2 \cup \Delta_3 \cup \Delta_4$ of context-free rules where A is the terminal alphabet and where S_ε and S_u are variables:

$$\begin{array}{lcl} \Delta_1 & = & \{S_\varepsilon \mapsto \varepsilon\} \\ \Delta_2 & = & \{S_\varepsilon \mapsto S_\varepsilon (S_u)^t f'' \mid f = f'f'', f' \in E_u^*, \varphi(f') = u^t\} \\ \Delta_3 & = & \{S_u \mapsto S_\varepsilon S_u S_\varepsilon\} \\ \Delta_4 & = & \{S_u \mapsto (S_u)^k u'' \mid u = u'u'', u' \in E_u^*, u'' \neq \varepsilon, \varphi(u') = u^k\} \end{array}$$

From this definition, we can observe the following:

- in Δ_2, the words f'' are nonempty since $f \notin E_u^*$,
- in Δ_2, if we take $f' = \varepsilon$, we get the rule $S_\varepsilon \mapsto S_\varepsilon f$,
- in Δ_4, the words f'' is assumed to be nonempty in order to avoid the rule $S_u \mapsto S_u$,
- in Δ_2, if we take $u' = \varepsilon$, we get the rule $S_u \mapsto u$.

We shall prove that for all words $w \in A^*$, for all $x \in \varepsilon + u$, $S_x \xrightarrow[\Delta]{*} w$ if and only if $w \in I^*_{u^*|f}(x)$. Since all rules in Δ are context-free which implies that $I^*_{u^*|f}(u)$ and $L_{u^*|f}$ are both context-free. The proof is an induction on $p = |w| - |x|$.

1. $x \xrightarrow[I_{u^*|f}]{*} w$ implies $S_x \xrightarrow[\Delta]{*} w$: if $p = 0$ then $x = w = \varepsilon$ or $x = w = u$ and we have the rules $S_x \mapsto x$. If $p > 0$, assume first $x = \varepsilon$ and $\varepsilon \xrightarrow[I_{u^*|f}]{} f \xrightarrow[I_{u^*|f}]{*} w$. From Lemma 6.2 we get $f = f'f''$ and $w = w'f''$ with $f' \xrightarrow[I_{u^*|f}]{*} \varphi(f') = u^k \xrightarrow[I_{u^*|f}]{*} w'$. If $f' = \varepsilon$, we have $w = f$ and the property is true thanks to the derivation $S_\varepsilon \xrightarrow[\Delta]{} S_\varepsilon f \xrightarrow[\Delta]{} f$. Else, from Item 4 of Lemma 6.10, $w' = w_1 \cdots w_k$ with $u \xrightarrow[I_{u^*|f}]{*} w_i$ for all $1 \leq i \leq k$. Observe that for all $1 \leq i \leq k$, $|w_i| - |u| < |w|$, so from the inductive hypothesis follows $S_u \xrightarrow[\Delta]{*} w_i$ for all $1 \leq i \leq k$. On the other hand, Δ_2 contains the rule $S_\varepsilon \mapsto S_\varepsilon (S_u)^k f''$ so we get the derivation $S_\varepsilon \xrightarrow[\Delta]{} S_\varepsilon (S_u)^k f'' \xrightarrow[\Delta]{} (S_u)^k f'' \xrightarrow[\Delta]{*} w_1 \cdots w_k f'' = w$.
Assume now $x = u$. We have to consider two cases:
   (a) $u \xrightarrow[I_{u^*|f}]{} uf \xrightarrow[I_{u^*|f}]{*} w$: from Item 4 of Lemma 6.10 follows $w = w_1 w_2$ for some words w_1, w_2 with $u \xrightarrow[I_{u^*|f}]{*} w_1$ and $f \xrightarrow[I_{u^*|f}]{*} w_2$. Since $|w_1| - |u| < |w| - |u| = p$, we get from the inductive hypothesis $S_u \xrightarrow[\Delta]{*} w_1$. Moreover, from $|w_1| \geq |u|$ follows $|w_2| \leq |w| - |u| = p$ so we can apply the inductive hypothesis on the derivation $\varepsilon \xrightarrow[I_{u^*|f}]{} f \xrightarrow[I_{u^*|f}]{*} w_2$ to get $S_\varepsilon \xrightarrow[\Delta]{*} w_2$. Finally we have $S_u \xrightarrow[\Delta]{} S_\varepsilon S_u S_\varepsilon \xrightarrow[\Delta]{} S_u S_\varepsilon \xrightarrow[\Delta]{*} w_1 w_2 = w$.

(b) $u \xrightarrow[I_{u^*|f}]{} fu \xrightarrow[I_{u^*|f}]{*} w$: if $w = fu$, we have the derivation $S_u \xrightarrow[\Delta]{} S_\varepsilon S_u S_\varepsilon \xrightarrow[\Delta]{} S_\varepsilon S_u \xrightarrow[\Delta]{} S_\varepsilon u \xrightarrow[\Delta]{} S_\varepsilon fu \xrightarrow[\Delta]{} fu$. Else, from Lemma 6.2 there exist words α, β, w' such that $fu = \alpha\beta$, $w = w'\beta$ and $\alpha \xrightarrow[I_{u^*|f}]{*} u^t \xrightarrow[I_{u^*|f}]{*} w'$ for some t. Observe that $\alpha\beta = uf \notin E_u^*$ from Item 1 of Lemma 6.10. That implies $\beta \neq \varepsilon$ so $|w'| < |w|$. Moreover, from Item 4 of Lemma 6.10, $w' = w'_1 \cdots w'_t$ with $u \xrightarrow[I_{u^*|f}]{*} w'_i$ and $|w'_i| < |w|$ for all $1 \leq i \leq t$.

From the inductive hypothesis, we get $S_u \xrightarrow[\Delta]{*} w'_i$ for all $1 \leq i \leq t$ so $(S_u)^t\beta \xrightarrow[\Delta]{*} w$. We have to consider again two cases by comparing $|f|$ and $|\alpha|$:

- $f = \alpha f''$ and $\beta = f''u$ for some word f''. In this case, we easily get the derivation: $S_u \xrightarrow[\Delta]{} S_\varepsilon S_u S_\varepsilon \xrightarrow[\Delta]{} S_\varepsilon u \xrightarrow[\Delta]{} S_\varepsilon (S_u)^t f''u \xrightarrow[\Delta]{} (S_u)^t f''u = (S_u)^t\beta \xrightarrow[\Delta]{*} w$.
- $\alpha = fu'$ and $u = u'\beta$ for some word u'. From the derivation $u' \xrightarrow[I_{u^*|f}]{} fu' \xrightarrow[I_{u^*|f}]{*} u^t$ follows $u' \in E_u^*$ with $\varphi(u') = u^t$. Moreover $\beta \neq \varepsilon$ so the rule $S_u \mapsto (S_u)^t\beta$ belongs to Δ_4 which implies $S_u \xrightarrow[\Delta]{*} w$.

2. $S_x \xrightarrow[\Delta]{*} w$ implies $x \xrightarrow[I_{u^*|f}]{*} w$: if $p = 0$ then $w = x$ so $x \xrightarrow[I_{u^*|f}]{*} w$. If $p > 0$ and $x = \varepsilon$ then $S_\varepsilon \xrightarrow[\Delta]{} S_\varepsilon (S_u)^k f'' \xrightarrow[\Delta]{*} w$ with $f = f'f''$ and $\varphi(f') = u^k$. From this follows $w = w_0 w_1 \cdots w_k f''$ and, since $f'' \neq \varepsilon$, $|w_i| < |w|$ for all $0 \leq i \leq k$. We can apply the inductive hypothesis on the derivations $S_\varepsilon \xrightarrow[\Delta]{*} w_0$ and $S_u \xrightarrow[\Delta]{*} w_i$ for all $1 \leq i \leq k$ so we get $\varepsilon \xrightarrow[I_{u^*|f}]{*} u^k f'' \xrightarrow[I_{u^*|f}]{*} w_1 \cdots w_k f'' \xrightarrow[I_{u^*|f}]{*} w_0 w_1 \cdots w_k f'' = w$.

Assume now $p > 0$ and $x = u$ and consider a shortest derivation $S_u \xrightarrow[\Delta]{*} w$, we have two cases:

(a) $S_u \xrightarrow[\Delta]{} (S_u)^k u'' \xrightarrow[\Delta]{*} w$ with $u = u'u''$, $u'' \neq \varepsilon$ and $\varphi(u') = u^k$. From this follows $w = w_1 \cdots w_k u''$ with $|w_i| < |w|$ for all $1 \leq i \leq k$ since $u'' \neq \varepsilon$. From the inductive hypothesis, for all $1 \leq i \leq k$, $u \xrightarrow[I_{u^*|f}]{*} w_i$ which implies $u = u'u'' \xrightarrow[I_{u^*|f}]{*} u^k u'' \xrightarrow[I_{u^*|f}]{*} w$.

(b) $S_u \xrightarrow[\Delta]{} S_\varepsilon S_u S_\varepsilon \xrightarrow[\Delta]{*} w$: then $w = w_1 w_2 w_3$ for some words w_1, w_2, w_3 with $S_\varepsilon \xrightarrow[\Delta]{*} w_1$, $S_u \xrightarrow[\Delta]{*} w_2$ and $S_\varepsilon \xrightarrow[\Delta]{*} w_3$. Moreover, since we consider a shortest derivation, $w_1 w_3 \neq \varepsilon$. That implies $|w_2| < |w|$ so, from the inductive hypothesis, we get $u \xrightarrow[I_{u^*|f}]{*} w_2$. Now, since $|w_2| \geq |u|$ we get $|w_1| \leq |w| - |u|$ and $|w_3| \leq |w| - |u|$ so, from the inductive hypothesis follows $\varepsilon \xrightarrow[I_{u^*|f}]{*} w_1$ and $\varepsilon \xrightarrow[I_{u^*|f}]{*} w_3$ which implies $u \xrightarrow[I_{u^*|f}]{*} uw_3 \xrightarrow[I_{u^*|f}]{*} w_2 w_3 \xrightarrow[I_{u^*|f}]{*} w_1 w_2 w_3 = w$.

□

Thanks to Proposition 6.22, we can now state that $I^*_{u^*|f}$ preserves the context-free languages:

Proposition 6.23. *For every context-free language L, for all u and for all f, $I^*_{u^*|f}(L) \in \mathsf{CF}$.*

Proof. We shall prove that $I^*_{u^*|f} = s \circ \tau$ where τ is a rational transduction and s is a context-free substitution that will prove the property. Let $B = \{b_e \mid e \in E_u\}$ be an alphabet in bijection with the set E_u, we define the morphism $h : (A \cup B)^* \mapsto A^*$ by $h(a) = a$ for all $a \in A$ and $h(b_e) = e$ for all $b_e \in B$. We also define the morphism $g : (A \cup B)^* \mapsto (A \cup B)^*$ by $g(a) = a$ for all $a \in A$ and $g(b_e) = (b_u)^i$ for all $b_e \in B$ with $\varphi(b_e) = u^i$. We can now define $\tau : A^* \mapsto (A \cup B \cup \{\#\})^*$ where $\#$ is a fresh letter by $\tau(w) = \#g(h^{-1}(w) \cap B^*A^*)$ for all $w \in A^*$and $s : (A \cup B \cup \{\#\})^* \mapsto A^*$ by $s(a) = a$ for all $a \in A$, $s(\#) = L_{u^*|f}$ and $s(b_u) = I^*_{u^*|f}(u)$.

We observe that for all words w,

$$\tau(w) = \{\#(b_u)^t w'' \mid w = w'w'', w' \in E_u^*, \varphi(w') = u^t\}.$$

This equality shows in particular that the fresh letter # is only useful when $t = 0$. We also observe that $w \in s \circ \tau(w)$ for all $w \in A^*$.

We shall now prove that for all words $w, w' \in A^*$, it holds that $w' \in I^*_{u^*|f}(w)$ if and only if $w' \in s \circ \tau(w)$.

1. if $w' \in I^*_{u^*|f}(w)$ then $w' \in s \circ \tau(w)$: the proof is an induction on $|w'|$, the length of the word w'. The base case $|w'| = |w|$ is clear since $w \in s \circ \tau(w)$. Else, if $w = w_1 w_2$ and $w' = w'_1 w_2$ for some words w_1, w_2, w'_1 with $w_2 \neq \varepsilon$ and $w'_1 \in I^*_{u^*|f}(w_1)$, from the inductive hypothesis follows $w'_1 \in s \circ \tau(w_1)$ so $w'_1 w_2 \in s \circ \tau(w_1 w_2)$. The last case is a derivation $w \xrightarrow[I_{u^*|f}]{*} u^k \xrightarrow[I_{u^*|f}]{*} w'$ for some natural number k. In this case, $\#(b_u)^k \in \tau(w)$ so $w' \in I^*_{u^*|f}(u^k)$ which implies $w' \in L_{u^*|f}(I^*_{u^*|f}(u))^k \subseteq s(\#)s((b_u)^k) \subseteq s(\#(b_u)^k) \subseteq s \circ \tau(w)$.
2. if $w' \in s \circ \tau(w)$ then $w' \in I^*_{u^*|f}(w)$: if $w' \in s \circ \tau(w)$ then there exists $\#(b_u)^k w'' \in \tau(w)$ for some word $w'' \in A^*$ such that $w' \in s(\#(b_u)^k)w''$. From the definition of τ, $w = \alpha w''$ with $\alpha \xrightarrow[I_{u^*|f}]{*} u^k$. Moreover, $w' = \beta w''$ for some word β with $\beta \in s(\#(b_u)^k = I^*_{u^*|f}(u^k)$. Finally $w = \alpha w'' \xrightarrow[I_{u^*|f}]{*} u^k w'' \xrightarrow[I_{u^*|f}]{*} \beta w'' = w'$.

□

7. Conclusion and perspectives

In this paper we have shown that, even in the case of a rational control language and even in the case of a single insertion rule, it is possible to define a system $I_{L|f}$ such that $I^*_{L|f}$ is not (FIN/CF). In particular, we have proved that as soon as a word f contains at least two distinct letters, there exists a rational language R_f such that for all words w, $I^*_{R_f|f}(w)$ is not context-free. On the other hand, it is easily seen that, for all R and all f, $I^*_{R_f|f}(w)$ is recursive: for all words, by erasing, iteratively and nondeterministically, occurrences of f it is possible to check if w is reached with this procedure that clearly stops since the length of the input strictly decreases at each step. Moreover this last remark shows that for all context-sensitive languages L, $I^*_{R_f|f}(L)$ is in fact context-sensitive since it can be recognized by a linear bounded automaton.

Among the different questions that arose on controlled insertions systems, the following ones deserve to be studied as a complement of Section 3.3: given a rational language L and a word f, is it possible to decide whether or not

- $I^*_{L|f}$ is confluent?
- $I^*_{L|f}$ is unambiguous?

In Section 3.4, we have defined the maximal control language of a word f for insertion, denoted by $\mathcal{C}_{\max}(f)$ when it exists. We have proved its existence for all words f such that r, the root of f, is unbordered: in this case, $\mathcal{C}_{\max}(f) = \mathsf{LF}(f)^* \setminus (A^* f A^* \cup A^* r)$ and $I^*_{\mathcal{C}_{\max}(f)|f}$ is codeterministic. We have seen that this result does not hold anymore when f is bordered. Nevertheless this proposition does not prove that there is no maximal control language in this case. For instance, if $f = aba$ it can be proved that $\mathcal{C}_{\max}(aba) = a^* + (ab)^*$. We can observe that $\mathcal{C}_{\max}(aba)$ and $R = \mathsf{LF}(aba)^* \setminus A^* aba A^*$ are not comparable with respect to inclusion: $abab \in \mathcal{C}_{\max}(aba) \setminus R$ and $aab \in R \setminus \mathcal{C}_{\max}(aba)$. On the other hand, it can also be proved that $I^*_{\mathcal{C}_{\max}(aba)|aba}$ is unambiguous, so, in the light of this example, we can summarize the questions about the maximal control language as follows:

- Does every word possesses a maximal control language for insertion?
- If the answer of the previous question is negative, given a word f, is it decidable to know whether or not $\mathcal{C}_{\max}(f)$ exists?
- When $\mathcal{C}_{\max}(f)$ exists, is it always rational?
- When $\mathcal{C}_{\max}(f)$ exists, does it always hold that $I^*_{\mathcal{C}_{\max}(f)|f}$ is unambiguous?

In Section 6 we have defined the language $\mathcal{R}_{R|f} = \{w \in A^* \mid I^*_{R|f}(w) \cap R \neq \emptyset\}$. We have proved that in the case when $R = u^*$ for some word u, $\mathcal{R}_{R|f}$ is a rational language and we have given an algorithm to compute it. We think that $\mathcal{R}_{R|f}$ is also rational when $I^*_{R|f}$ is codeterministic. More precisely, we conjecture the following:

If $I^*_{R|f}$ is codeterministic then

1. $I^*_{R|f}$ preserves the context-free languages,
2. $\mathcal{R}_{R|f}$ is a rational language.

More generally, a natural question would be to know whether $\mathcal{R}_{R|f}$ is a recursive set or not when R belongs to different classes of languages.

In the same section, we have characterized in Proposition 6.21 when such a system $I_{u^*|f}$ leads to a transformation $I^*_{u^*|f}$ that corresponds to a rational transduction and we have proved that these systems preserve context-free languages in Proposition 6.23. More generally, given a rational language R and a word f, is it possible to decide whether $I^*_{R|f}$ is rational and is it possible to decide whether $I^*_{R|f}$ preserves context-free languages? When R is a finite language, the controlled rewriting corresponds to prefix rewriting as defined in [2] where it is proved that the corresponding transformation is a rational transduction. Clearly, it is also the case when R is not necessarily finite but is a prefix set. It is worth studying whether there are some other families of rational languages that satisfy these properties. In particular, can the results established in Section 6 be extended to the case when R is a bounded rational language?

At last, we state here one conjecture and two questions when $R = u^*$ for some word u:

- First, in order to characterize when $I_{S|U}$ is a rational transduction, we have proved in Proposition 6.20 that for all words w, i$I^*_{u^*|f}(w) \rightsquigarrow_{\text{rat}} L_{u^*|f}$. In fact we conjecture that we also have $L_{u^*|f} \rightsquigarrow_{\text{rat}} I^*_{u^*|f}(w)$ *i.e.* $I^*_{u^*|f}(w)$ and $L_{u^*|f}$ are rationally equivalent.
- Second, to prove that $I^*_{u^*|f}$ preserves context-free languages, we have proved in Proposition 6.22 that for all u and f, $L_{u^*|f}$ is a context-free language. The following questions would precise this property:
  - Does that hold that ${D'_1}^* \rightsquigarrow_{\text{rat}} L_{u^*|f}$?
  - Conversely, does that hold that either $L_{u^*|f}$ is a rational language or $L_{u^*|f} \rightsquigarrow_{\text{rat}} {D'_1}^*$?

Acknowledgements. The authors would like to thank the anonymous referees for their careful reading and their helpful comments that improved the quality of the manuscript.

Appendix I

Some of the notations that are used in this article are listed below.

- $\rightsquigarrow_{\text{rat}}$: a language L rationally dominates a language L', denoted by $L \rightsquigarrow_{\text{rat}} L'$, if there exists a rational transduction τ such that $L' = \tau(L)$.
- $\underset{\text{rat}}{\equiv}$: two languages L and L' are rationally equivaeant, denoted by $L \underset{\text{rat}}{\equiv} L'$, if $L \rightsquigarrow_{\text{rat}} L'$ and $L' \rightsquigarrow_{\text{rat}} L$.
- $I_{L|f}$: the binary relation on words associated with the controlled one-rule insertion system with L as control language and f as word to be inserted. It can be seen as a one-rule rewriting system where the unique rule $\varepsilon \mapsto f$ can only be applied in a position defining a left context that is in L.
- $I^*_{L|f}$: the reflexive and transitive closure of $I_{L|f}$. It is also the corresponding transformation over languages: for all words $w \in A^*, I^*_{L|f}(w) = \{w' \mid (w, w') \in I^*_{L|f}\}$ and for all languages $K \subseteq A^*, I^*_{L|f}(K) = \bigcup_{w \in K} I^*_{L|f}(w)$.
- $\xrightarrow[S]{}$: a single step of rewriting using a rule of the rewriting system S.
- $\xrightarrow[S]{*}$: the derivation relation that is the reflexive and transitive closure of $\xrightarrow[S]{}$. In this paper, a derivation $w = w_0 \underset{S}{\rightarrow} w_1 \cdots \underset{S}{\rightarrow} w_n = w'$ is completely characterized by the list of words $[w_0, \ldots, w_n]$ independently from the indexes where the rule is applied in the left-hand side of each step of rewriting.
- $\xrightarrow{*}$: the derivation relation in the case when the rewriting system is implicitly defined.
- L_f : the language $I^*_{A^*|f}(\varepsilon)$ for some word $f \in A^*$.
- $L_{K|f}$: the language $I^*_{K|f}(\varepsilon)$ for some control language K and some word f.

$\mathcal{C}_{\max}(f)$: the maximal control for insertion of the word f. It satisfies that for all languages K, $L_{K|f} = L_f$ if and only if $\mathcal{C}_{\max}(f) \subseteq K$. The fact that for all f, $\mathcal{C}_{\max}(f)$ exists is an open question.

K_0 : the language $K_0 = {D'_1}^* \cap K$ where $K = \{w \in (a+b)^* \mid \forall x \in \mathsf{LF}(w), |x|_a \leq 2|x|_b + 1\}$.

I_0 : the insertion system $I_0 = I_{R|ab}$ where $R = \{w \in A^* \mid |w|_a = 2n, n \geq 0\}$.

$\mathcal{R}_{L|f}$: the language $\mathcal{R}_{L|f} = \{w \in A^* \mid I^*_{L|f}(w) \cap L \neq \emptyset\}$.

E^*_u : a language obtained by an algorithmic construction and proved to be equal to $\mathcal{R}_{u^*|f}$.

$\varphi(w)$: defined only for a word $w \in \mathcal{R}_{u^*|f}$ when u and f satisfy $uf \neq fu$. The word $\varphi(w)$ is the unique word of $I^*_{u^*|f}(w) \cap u^*$.

References

[1] J. Berstel, Transductions and context-free languages, vol. 38 of *Teubner Studienbücher : Informatik.* Teubner (1979).

[2] D. Caucal, On the regular structure of prefix rewriting, *Theor. Comput. Sci.* **106** (1992) 61–86.

[3] L. Chottin, Strict deterministic languages and controlled rewriting systems, in Automata, Languages and Programming, 6th Colloquium, Graz, Austria, July 16–20, 1979, Proceedings, *Lecture Notes in Computer Science*, edited by H.A. Maurer. vol. 71, Springer (1979) 104–117.

[4] L. Chottin, Langages Algébriques et Systèmes de Réécriture Rationels. *ITA* **16** (1982) 93–112.

[5] M. Clerbout and Y. Roos, Semi-commutations and algebraic languages, in C. Choffrut and T. Lengauer (editors), STACS 90, *Lecture Notes in Computer Science.* Springer Berlin/Heidelberg (1990), vol. 415, 82–94.

[6] A. Geser, Decidability of termination of grid string rewriting rules. *SIAM J. Comput.* **31** (2002) 1156–1168.

[7] A. Geser, D. Hofbauer and J. Waldmann, Match-bounded string rewriting systems. *Appl. Algebra Eng. Commun. Comput.* **15** (2004) 149–171.

[8] A. Geser, D. Hofbauer and J. Waldmann, Termination proofs for string rewriting systems via inverse match-bounds. *J. Autom. Reason.* **34** (2005) 365–385.

[9] D. Hofbauer and J. Waldmann, Deleting string rewriting systems preserve regularity, *Theor. Comput. Sci.* **327** (2004) 301–317.

[10] M. Latteux and Y. Roos, On one-rule grid semi-thue systems. *Fundam. Inform.* **116** (2012) 189–204.

[11] M. Latteux and Y. Roos, On prefixal one-rule string rewrite systems. *Theor. Comput. Sci.* **795** (2019) 240–256.

[12] P. Leupold, On regularity-preservation by string-rewriting systems, in C. Martín-Vide, F. Otto and H. Fernau (editors), Language and Automata Theory and Applications, Second International Conference, LATA 2008, Tarragona, Spain, March 13–19, 2008. Revised Papers, *Lecture Notes in Computer Science*, vol. 5196, Springer (2008) 345–356.

[13] R.C. Lyndon and M.P. Schützenberger, The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.* **9** (1962) 289–298.

[14] R. McNaughton, Semi-Thue systems with an inhibitor. *J. Autom. Reason.* **26** (2001) 409–431.

[15] M. Nivat, Transductions des langages de Chomsky. *Ann. l'Inst. Fourier* **18** (1968) 339–455.

[16] J. Sakarovitch, Elements of Automata Theory, Cambridge University Press, New York, NY, USA (2009).

[17] G. Sénizergues, Some decision problems about controlled rewriting systems. *Theor. Comput. Sci.* **71** (1990) 281–346.

[18] G. Sénizergues, Formal languages & word-rewriting, in H. Comon and J.-P. Jounnaud (editors), Term Rewriting: French Spring School of Theoretical Computer Science Font Romeux, France, May 17–21, 1993 Advanced Course, Springer Berlin Heidelberg, Berlin, Heidelberg (1995) 75–94.