RAIRO-Theor. Inf. Appl. 43 (2009) 221–237
DOI: 10.1051/ita:2008022 Available online at:
www.rairo-ita.org

# MINIMAL NFA AND BIRFSA LANGUAGES *

Michel Latteux[1], Yves Roos[1] and Alain Terlutte[2]

**Abstract.** In this paper, we define the notion of biRFSA which is a residual finate state automaton (RFSA) whose the reverse is also an RFSA. The languages recognized by such automata are called biRFSA languages. We prove that the canonical RFSA of a biRFSA language is a minimal NFA for this language and that each minimal NFA for this language is a sub-automaton of the canonical RFSA. This leads to a characterization of the family of biRFSA languages. In the second part of this paper, we define the family of biseparable automata. We prove that every biseparable NFA is uniquely minimal among all NFAs recognizing a same language, improving the result of H. Tamm and E. Ukkonen for bideterministic automata.

**Mathematics Subject Classification.** 68Q45.

## 1. Introduction

Finite automata constitute a useful model in several domains of computer science like in lexical analysis, coding theory, formal verification... The study of finite automata began in the 50's and led to a very rich and fecund theory (see for instance [7,12]). More recently, finite automata were also used in bioinformatics or in machine learning where, in this last domain, they are used as a representation of regular languages which are the targets of learning algorithms. In this case, it is important to get small automata with respect to the number of states, but it is also important to get a canonical representation of regular languages.

The structure of deterministic finite automata (DFA) recognizing a given recognizable set L is simple since these DFAs have a common model, the unique minimal

*Keywords and phrases.* Residual finite state automata; minimal NFA.

* *This research was partially supported by Inria (MOSTRARE team).*

[1] Laboratoire d'Informatique Fondamentale de Lille, UMR CNRS 8022, Université de Lille 1, 59655 Villeneuve d'Ascq Cedex, France; michel.latteuxyves.roos@lifl.fr

[2] Équipe Grappa–EA 3588, Université de Lille 3, Domaine universitaire du "Pont de bois", BP 149, 59653 Villeneuve d'Ascq Cedex, France; alain.terlutte@univ-lille3.fr

DFA of L. This nice property does not hold for the set of nondeterministic finite automata (NFA) recognizing L. Indeed, it may happen that L is recognizable by two completely different NFAs that are both minimal with respect to the number of states (see [2,3]). Recently the notion of residual finite state automaton (RFSA), a natural extension of DFAs, has been introduced (see [5]). This new family of automata is particularly interesting since the above property of DFAs has been preserved: there exists a unique RFSA called the canonical RFSA of L that is minimal among the RFSAs recognizing $L$. For some recognizable set L, this canonical RFSA is also minimal among the NFAs recognizing L and has much less states than the minimal DFA of L. Unfortunately, this is not true in general and it seems natural to wonder what kind of recognizable sets have this property. In order to get a family of such recognizable sets, we extend the notion of 0-reversible automaton, called also bideterministic automaton, introduced by Angluin [1] for learning questions. So we define biRFSAs that are RFSAs whose the reverse is also an RFSA. The languages recognized by biRFSAs are called biRFSA languages. We prove that the canonical RFSA of a biRFSA language $L$ is a minimal NFA for L. More precisely, each minimal NFA for L is isomorphic to a sub-automaton of the canonical RFSA of L. Canonical RFSAs are full NFAs, that is, it is not possible to add a transition or to transform a non initial (resp. non final) state into an initial (resp. final) one without changing the associated language. As each minimal NFA, they are also thin, that is, it is not possible to remove a state without changing the associated language. With these two properties, one can establish a characterization of the family of biRFSA languages: a recognizable language $L$ is a biRFSA language if and only if there is a unique full and thin NFA recognizing it.

It can be easily seen that the unicity of the minimal NFA does not hold for biRFSA languages. Indeed, the canonical RFSA of a biRFSA L can contain incomparable (minimal) NFAs recognizing L. The second part of this paper concerns recognizable languages having a unique minimal NFA. In [15], it has been proved that bideterministic automata are uniquely minimal, that is they are the unique minimal NFA for the associated language. We improve this result by showing that this property does hold for a larger family of finite automata, called biseparable NFAs. The proof of this result uses the above characterization of the family of biRFSA languages since it is seen that biseparable NFAs are special canonical RFSAs. We also prove the following characterization result: a recognizable language $L$ is recognized by a biseparable NFA if and only if no prime residual of L is included in the union of the other prime residuals of $L$.

## 2. Preliminaries

Let us recall the definition of residuals of a language: let $\Sigma$ be an alphabet and $L \subseteq \Sigma^*$ be a language. A language $L' \subseteq \Sigma^*$ is a *residual* of $L$ if there exists a word $u \in \Sigma^*$ such that $L' = \{v \in \Sigma^* \mid uv \in L\}$, that is denoted $L' = u^{-1}L$. Symmetrically is defined the notion of *left* residual: a language $L'$ is a left residual of $L$ if there exists a word $v \in \Sigma^*$ such that $L' = \{u \in \Sigma^* | uv \in L\}$, that is denoted $L' = Lv^{-1}$. To avoid ambiguity, any residual of a language $L$ will also be called a *right* residual of $L$.

It is well known that a language is recognizable if and only if it has a finite number of residuals. In order to precise the link between residuals of a recognizable language and the states of automata which recognize it, let us introduce the following notation: let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a finite non deterministic automaton (NFA), where $\Sigma$ is the input alphabet, $Q$ is the set of states, $I \subseteq Q$ is the set of initial states, $F \subseteq Q$ is the set of final states, and $\delta$ is the transition function from $Q \times \Sigma \to 2^Q$. We extend, as usually, the transition function $\delta$ in a function from $2^Q \times \Sigma \to 2^Q$ by: $\forall S \subseteq Q, \forall x \in \Sigma, \delta(S, x) = \cup_{q \in S} \delta(q, x)$, and also in a function from $2^Q \times \Sigma^* \to 2^Q$ that is inductively defined by: $\forall S \subseteq Q, \delta(S, \varepsilon) = S$ and $\forall S \subseteq Q, \forall u \in \Sigma^*, \forall x \in \Sigma, \delta(S, xu) = \delta(\delta(S, x), u)$. We sometimes consider $\delta$ as the *set* of transitions of automaton $\mathcal{A}$ that is the subset of $Q \times \Sigma \times Q$ defined by $\delta = \{(q, x, q') \mid q' \in \delta(q, x)\}$. Also, in this paper, we shall always consider equality between automata up to isomorphism, and we say that two automata are equivalent if they recognize a same language.

For any state $q \in Q$, we define the language $\mathsf{post}_{\mathcal{A},q}$ by $\mathsf{post}_{\mathcal{A},q} = \{u \in \Sigma^* \mid \delta(q, u) \cap F \neq \emptyset\}$, and we define the language $\mathsf{pre}_{\mathcal{A},q}$ by $\mathsf{pre}_{\mathcal{A},q} = \{u \in \Sigma^* \mid q \in \delta(I, u)\}$. Notice that $\mathsf{post}_{\mathcal{A},q}$ (resp. $\mathsf{pre}_{\mathcal{A},q}$) is the language recognized by the automaton $\langle \Sigma, Q, \{q\}, F, \delta \rangle$ (resp. $\langle \Sigma, Q, I, \{q\}, \delta \rangle$). When there is no ambiguity on the used automaton, we shall just write $\mathsf{post}_q$ for $\mathsf{post}_{\mathcal{A},q}$ and $\mathsf{pre}_q$ for $\mathsf{pre}_{\mathcal{A},q}$.

If we consider any trim deterministic automaton $\mathcal{A} = \langle \Sigma, Q, \{q_0\}, F, \delta \rangle$, it is clear that, for any state $q$ in $Q$, the language $\mathsf{post}_q$ is a residual of the language recognized by $\mathcal{A}$. Moreover it is well known that the set of states of the minimal deterministic automaton of any recognizable language $L$ is isomorphic to the set of the residuals of $L$. This fine property is not satisfied by non deterministic automata: if $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ is a non deterministic automaton, then for any state $q$ in $Q$, the language $\mathsf{post}_q$ is included in a residual of the language recognized by $\mathcal{A}$, but not always equal to it. This is the reason why the following notion has been introduced in [5]:

**Definition 2.1.** A (non deterministic) automaton $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ is a residual finite state automaton (RFSA for short) if for every state $q \in Q$, the language $\mathsf{post}_q$ is a residual of the language recognized by $\mathcal{A}$.

The notion of minimal deterministic automaton is essential, unfortunately there does not exist a similar notion for NFA. Nevertheless, such a canonical representation exists for the class of RFSA. Indeed it has been proved in [5] that every recognizable language can be recognized by a unique non deterministic *reduced* RFSA, called the canonical RFSA of the language. In order to give its definition, let us first introduce the notion of *prime* residual of a language.

**Definition 2.2.** Let $L$ be a language. A residual of $L$ is *prime* if it is non empty and if it cannot be obtained as the union of other residuals of $L$. The set of all prime residuals of $L$ is denoted by $\mathsf{prime}(L)$.

In a similar way, one can define the notion of prime left residual.

**Definition 2.3.** Let $\Sigma$ be an alphabet and $L \subseteq \Sigma^*$ be a recognizable language. The canonical RFSA $\mathcal{A}$ of $L$ is the automaton $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ where

(1) $\Sigma$ is the alphabet of $L$;
(2) $Q = \mathsf{prime}(L)$, is the set of prime residuals of $L$;
(3) $I \subseteq \mathsf{prime}(L)$ is the set of prime residuals of $L$ which are included in $L$;
(4) $F \subseteq \mathsf{prime}(L)$ is the set of prime residuals of $L$ containing the empty word;
(5) $\forall S \in \mathsf{prime}(L), \forall x \in \Sigma, \delta(S, x) = \{S' \in \mathsf{prime}(L) \mid xS' \subseteq S\}$.

The following result has been proved in [5].

**Lemma 2.4.** *Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a canonical RFSA recognizing a language $L$. Then for any state $q \in Q$, there exists a word $u_q \in \mathsf{pre}_q$ such that $\mathsf{post}_q = u_q^{-1}L$. Such a word $u_q$ is called an* (incoming) characteristic *word of state $q$.*

Let us now recall the definition of the reverse of a language and the reverse of an automaton: let $\Sigma$ be an alphabet. The *reverse* of a word $u \in \Sigma^*$ is denoted $u^R$ and is defined inductively by: $\varepsilon^R = \varepsilon$, and $\forall v \in \Sigma^*, \forall x \in \Sigma, (vx)^R = x(v^R)$. Then this definition is extended to languages: if $L$ is a language, then $L^R = \cup_{u \in L} u^R$.

Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be an automaton. Then the reverse of $\mathcal{A}$ is the automaton $\mathcal{A}^R = \langle \Sigma, Q, F, I, \delta^R \rangle$ where $\delta^R = \{(q, x, q') \mid (q', x, q) \in \delta\}$. It is well known that an automaton $\mathcal{A}$ recognizes a language $L$ if and only if its reverse, $\mathcal{A}^R$, recognizes $L^R$, the reverse of $L$.

The case when the reverse of a deterministic automaton is still deterministic leads to the class of 0-reversible languages (see[1]) or bideterministic languages (see [13,15]) which have been studied in the context of machine learning, or in terms of minimal representation of recognizable languages. When NFA are considered, we define the notion of biRFSA:

**Definition 2.5.**

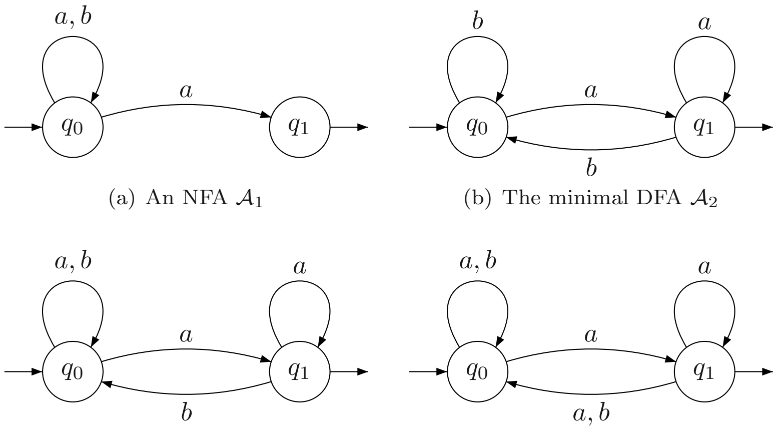
- An automaton $\mathcal{A}$ is a biRFSA if $\mathcal{A}$ is an RFSA and the reverse of $\mathcal{A}$ is also an RFSA.
- A language is a biRFSA language if there exists a biRFSA which recognizes it.

Note that, as an equivalent definition, we can say that $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$, recognizing a language $L$, is a biRFSA if, for any state $q \in Q$, $\mathsf{post}_q$ is a (right) residual of $L$ and $\mathsf{pre}_q$ is a left residual of $L$.

**Example 2.6.** Let us consider the four automata of Figure 1, each of them recognizing the same language $(a + b)^*a$.

The right residuals of $(a+b)^*a$ are the two languages $(a+b)^*a$ and $(a+b)^*a+\varepsilon$. The non empty left residuals of $(a+b)^*a$ are the two languages $(a+b)^*a$ and $(a+b)^*$.

Automaton $\mathcal{A}_1$ is not an RFSA because $\mathsf{post}_{\mathcal{A}_1, q_1} = \{\varepsilon\}$ which is not a right residual of $(a + b)^*a$. Automaton $\mathcal{A}_2$ is an RFSA since it is the minimal DFA of $(a + b)^*a$ but it is not a biRFSA since $\mathsf{pre}_{\mathcal{A}_2, q_0} = (a + b)^*b + \varepsilon$ which is not a left



(a) An NFA $\mathcal{A}_1$ (b) The minimal DFA $\mathcal{A}_2$

(c) A biRFSA $\mathcal{A}_3$ (d) The canonical RFSA $\mathcal{A}_4$

FIGURE 1. Some automata for $(a+b)^*a$.

residual of $(a+b)^*a$. Automaton $\mathcal{A}_3$ is a biRFSA since now $\mathsf{pre}_{\mathcal{A}_3,q_0} = (a+b)^*$, but it is not the canonical RFSA of $(a+b)^*a$. Indeed, $a(\mathsf{post}_{\mathcal{A}_3,q_0}) \subseteq \mathsf{post}_{\mathcal{A}_3,q_1}$ but there is no transition $(q_1, a, q_0)$ in automaton $\mathcal{A}_3$. Finally, automaton $\mathcal{A}_4$ is a biRFSA which is the canonical RFSA of $(a+b)^*a$.

The family of biRFSA languages is strictly included in the family of recognizable languages; even over a one-letter alphabet, there exist finite languages which are not biRFSA languages.

**Example 2.7.** The language $L = a+a^2$ is not a biRFSA language: let us suppose that there exists a biRFSA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ recognizing $L$; the set of non empty right residuals of $L$ is $\{a+a^2, \varepsilon+a, \varepsilon\}$ and it is equal to the set of non empty left residuals of $L$; all these residuals are prime. Since $a^2 \in L$, there exists a state $q \in Q$ such that $a \in \mathsf{pre}_q$ and $a \in \mathsf{post}_q$. Since $a^3 \notin L$, it follows that $\mathsf{pre}_q = \mathsf{post}_q = \varepsilon + a$ then $\mathcal{A}$ would recognize the empty word which is not in $L$. This example also shows that the family of biRFSA languages is not closed by union: indeed, any singleton language is clearly a biRFSA language. It is also easy to verify, as presented in [10], that the family of biRFSA languages is not closed by intersection, complementation, concatenation or quotient. Notice also that the language $L' = \varepsilon + a + a^2$ is a biRFSA language since it is recognized by the biRFSA given in Figure 2.

Despite its bad closure properties, the family of biRFSA languages is an interesting family in terms of identification by learning algorithms from positive examples. Indeed, the task of identifying a language from a set of its words is not an easy one. For instance, it is not feasible to identify regular languages from positive examples in the general case. Therefore, it is interesting to look for subclasses of regular languages that can be identified in this framework. One of the most classical identifiable classes is the class of reversible languages, introduced

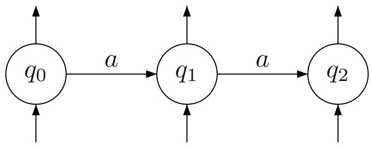



FIGURE 2. A biRFSA for $\varepsilon + a + a^2$.

in [1] and also studied in [13]. Concerning biRFSA languages, the subfamily of *ordered* biRFSA, introduced in [9], strictly contains the family of reversible languages and has been proved identifiable from positive examples. Moreover, the algorithm works in polynomial time from a set of positive examples whose size is polynomially bounded on the size of the target of the learning algorithm.

## 3. A decidable characterization of biRFSA languages

The aim of this section is to give a way for deciding whether a given recognizable language is a biRFSA language. We shall use the following property, satisfied by canonical RFSA:

**Lemma 3.1.** *Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a canonical RFSA. Then for any states $q, q' \in Q$, $\mathsf{post}_q \subseteq \mathsf{post}_{q'}$ if and only if $\mathsf{pre}_{q'} \subseteq \mathsf{pre}_q$.*

*Proof.* Let $u$ be a word which belongs to $\mathsf{pre}_{q'}$; we will prove by induction on $|u|$, the length of $u$, that $u$ belongs to $\mathsf{pre}_q$. If $|u| = 0$, then $u = \varepsilon$ and it follows that $q' \in I$ and $\mathsf{post}_{q'} \subseteq L$, where $L$ denotes the language recognized by automaton $\mathcal{A}$. Since $\mathsf{post}_q \subseteq \mathsf{post}_{q'} \subseteq L$, it follows from item 3 of definition 2.3 that state $q$ is also an initial state and $u = \varepsilon \in \mathsf{pre}_q$. Let us suppose now that $u = u'x$ with $u' \in \Sigma^*$ and $x \in \Sigma$. If $u'x \in \mathsf{pre}_{q'}$, there exists a state $q'' \in Q$ such that $q'' \in \delta(I, u')$ and $q' \in \delta(q'', x)$. Since $\mathcal{A}$ is a canonical RFSA, it follows that $x\mathsf{post}_{q'} \subseteq \mathsf{post}_{q''}$. Now, since $\mathsf{post}_q \subseteq \mathsf{post}_{q'}$, we obtain from item 5 of Definition 2.3 that $x\mathsf{post}_q \subseteq \mathsf{post}_{q''}$ and $q \in \delta(q'', x)$. Finally $q \in \delta(I, u'x)$ that is $u = u'x \in \mathsf{pre}_q$. $\square$

Then we can state:

**Proposition 3.2.** *A recognizable language is a biRFSA language if and only if its canonical RFSA is a biRFSA.*

*Proof.* Let $\mathcal{A} = \langle \Sigma, Q_\mathcal{A}, I_\mathcal{A}, F_\mathcal{A}, \delta_\mathcal{A} \rangle$ be a biRFSA recognizing a language $L$ and $\mathcal{B} = \langle \Sigma, Q_\mathcal{B}, I_\mathcal{B}, F_\mathcal{B}, \delta_\mathcal{B} \rangle$ the canonical RFSA of $L$. Let $q$ be an arbitrary state of $Q_\mathcal{B}$ and $u$ a word such that $\mathsf{post}_{\mathcal{B},q} = u^{-1}L$. We have to prove that $\mathsf{pre}_{\mathcal{B},q}$ is a left residual of $L$.

Set $S = \{s \in Q_\mathcal{A} \mid u \in \mathsf{pre}_{\mathcal{A},s}\}$, then $u^{-1}L = \cup_{s \in S} \mathsf{post}_{\mathcal{A},s}$. Since $u^{-1}L$ is a prime right residual and the $\mathsf{post}_{\mathcal{A},s}$ are right residuals, there exists $p \in S$ such that $u \in \mathsf{pre}_{\mathcal{A},p}$ and $\mathsf{post}_{\mathcal{A},p} = u^{-1}L$. Now, since $\mathsf{pre}_{\mathcal{A},p}$ is a left residual, $\mathsf{pre}_{\mathcal{A},p} = Lv^{-1}$ for some word $v \in \Sigma^*$. Then $uv \in L$ and $v \in u^{-1}L = \mathsf{post}_{\mathcal{B},q}$. It remains to prove that $\mathsf{pre}_{\mathcal{B},q} = Lv^{-1}$.

Clearly, $\mathsf{pre}_{\mathcal{B},q} \subseteq Lv^{-1}$. For the reverse inclusion, let us take $w \in Lv^{-1}$. Then $wv \in L$ and there is a state $p'$ such that $w \in \mathsf{pre}_{\mathcal{B},p'}$, and $v \in \mathsf{post}_{\mathcal{B},p'}$. As $\mathsf{post}_{\mathcal{B},p'}$ is a right residual, $\mathsf{post}_{\mathcal{B},p'} = z^{-1}L$ for some $z \in \Sigma^*$. Then $zv \in L$ and $z \in Lv^{-1} = \mathsf{pre}_{\mathcal{A},p}$. Hence, $z\mathsf{post}_{\mathcal{A},p} \subseteq L$ and $\mathsf{post}_{\mathcal{A},p} = u^{-1}L = \mathsf{post}_{\mathcal{B},q} \subseteq z^{-1}L = \mathsf{post}_{\mathcal{B},p'}$. Since $\mathcal{B}$ is a canonical RFSA, it comes from Lemma 3.1 that $\mathsf{pre}_{\mathcal{B},p'} \subseteq \mathsf{pre}_{\mathcal{B},q}$ and $w \in \mathsf{pre}_{\mathcal{B},q}$. □

The previous result is equivalent to:

**Corollary 3.3.** *A recognizable language $L$ is a biRFSA language if and only if the reverse of its canonical RFSA is the canonical RFSA of the reverse of $L$.*

*Proof.* From Proposition 3.2, it is sufficient to prove that if $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ is the canonical RFSA of a biRFSA language $L$, then for each state $q$ in $Q$, the language $\mathsf{pre}_q$ is prime, that is, there does not exist a set of states $S \subseteq Q$, such that every state $s \in S$ satisfies $\mathsf{pre}_s \subsetneq \mathsf{pre}_q$ and $\mathsf{pre}_q = \cup_{s \in S} \mathsf{pre}_s$. Let us suppose that there exists such a set $S$ and let us consider $u_q$ the characteristic word of $q$, i.e. such that $\mathsf{post}_q = u_q^{-1}L$. We know from Lemma 2.4 that such a $u_q$ exists. Now, since $\mathsf{pre}_q = \cup_{s \in S} \mathsf{pre}_s$, there exists $p \in S$ such that $u_q \in \mathsf{pre}_p$. It follows that $\mathsf{post}_p \subseteq u^{-1}L = \mathsf{post}_q$. From Lemma 3.1, we obtain $\mathsf{pre}_q \subseteq \mathsf{pre}_p$ which leads to a contradiction since every state $s \in S$ satisfies $\mathsf{pre}_s \subsetneq \mathsf{pre}_q$. □

In order to decide whether a given NFA $\mathcal{A}$ recognizes a biRFSA language, Proposition 3.2 and Corollary 3.3 lead to decision algorithms in which the canonical RFSA of the language $L(\mathcal{A})$ recognized by $\mathcal{A}$ must be computed. Unfortunately, it is shown in [5] that the canonical RFSA of the language $L(\mathcal{A})$ may be exponentially larger than the automaton $\mathcal{A}$. In next section, another characterization of biRFSA languages in terms of minimal NFA will lead to a PSPACE-algorithm and we shall prove that the problem to decide whether a given NFA $\mathcal{A}$ recognizes a biRFSA language is PSPACE-complete.

## 4. biRFSA languages and minimal NFA

In general, the canonical RFSA of a recognizable language is not always a minimal NFA. However, we shall prove that it becomes true when we are dealing with biRFSA languages. For biRFSA languages, the notion of canonical biRFSA can be easily defined, thanks to Proposition 3.2: a biRFSA is canonical if it is the canonical RFSA of a biRFSA language.

Clearly, the following is a direct consequence of Lemma 2.4:

**Lemma 4.1.** *Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a canonical biRFSA recognizing a language $L$. Then for any state $q \in Q$, there exist words $u_q, v_q \in \Sigma^*$ such that $\mathsf{post}_q = u_q^{-1}L$ and $\mathsf{pre}_q = Lv_q^{-1}$. The word $u_q$ (resp. $v_q$) is called an* incoming *(resp.* outgoing*)* characteristic *word of state $q$.*

Then we can state that the canonical biRFSA is a minimal NFA which recognizes a biRFSA language:

**Proposition 4.2.** *The canonical RFSA of a biRFSA language $L$ is a minimal NFA for $L$.*

*Proof.* For any canonical biRFSA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$, recognizing a language $L$, and any NFA $\mathcal{A}' = \langle \Sigma, Q', I', F', \delta' \rangle$ recognizing the same language $L$, let us consider now a mapping $h : Q \longrightarrow Q'$ which maps each state $q \in Q$ to a state $h(q) = q' \in Q'$ such that $u_q \in \mathsf{pre}_{\mathcal{A}',q'}$ and $v_q \in \mathsf{post}_{\mathcal{A}',q'}$ where $u_q$ and $v_q$ are respectively incoming and outgoing characteristic words of $q$. For any two distinct states $p, q \in Q$, if $u_p v_q \in L$ then $u_p \in L v_q^{-1}$ and $u_p \mathsf{post}_q \subseteq L$ which implies $\mathsf{post}_q \subseteq u_p^{-1} L = \mathsf{post}_p$. Similarly, if $u_q v_p \in L$ then $\mathsf{post}_p \subseteq \mathsf{post}_q$. Since $p$ and $q$ are distinct, $\mathsf{post}_p \neq \mathsf{post}_q$ and $u_p v_q \notin L$ or $u_q v_p \notin L$. It follows that $h$ is injective. $\square$

The converse is not true: there exist non biRFSA languages such that there does not exist a smaller automaton than their canonical RFSA. For instance the language $a + a^2$ of the example 2.7 is not a biRFSA language, but its canonical RFSA has three states and it is not possible to recognize $a + a^2$ with less states.

We shall see now that a canonical biRFSA of a biRFSA language $L$, satisfies another property than being a minimal NFA recognizing $L$: it also *contains* as sub-automata all the minimal NFAs which recognize $L$. This follows from the fact that a canonical RFSA is *full*, *i.e.* it is not possible to add transitions to it without changing the language that it recognizes. Let us give precisely the definition of sub-automaton; :

**Definition 4.3.** An NFA $\mathcal{B}$ is a sub-automaton of an NFA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ if $\mathcal{B} = \langle \Sigma, Q', I', F', \delta' \rangle$ with $Q' \subseteq Q$, $I' \subseteq I$, $F' \subseteq F$ and $\delta' \subseteq \delta$.

Then we can give the definition of full automata:

**Definition 4.4.** An NFA $\mathcal{A}$ is *full* if for any automaton $\mathcal{B}$ having the same set of states than $\mathcal{A}$ and recognizing the same language, $\mathcal{A}$ is a sub-automaton of $\mathcal{B}$ implies $\mathcal{A} = \mathcal{B}$.

In other words, an NFA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$, is full if adding a non initial state of $Q$ in $I$, or adding a non final state of $Q$ in $F$ or adding a new transition in $\delta$ changes the language recognized by $\mathcal{A}$.

Clearly, every automaton is included in a (non necessarily unique) full automaton recognizing the same language. When an automaton is a canonical RFSA, we have:

**Lemma 4.5.** *Any canonical RFSA is full.*

*Proof.* By definition of a canonical RFSA, a non initial state $q$ of a canonical RFSA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$, recognizing a language $L$, satisfies $\mathsf{post}_q \not\subseteq L$, then it is not possible to add such a $q$ in $I$. Symmetrically, let $p$ be a non final state of $\mathcal{A}$ and let us consider $u_p$, an incoming characteristic word of $p$. Since $p$ is not final, $\varepsilon \notin \mathsf{post}_p = u_p^{-1} L$, then $u_p \notin L$, and it is not possible to add $p$ in $F$. At last, let us consider two states $p, q \in Q$ and a letter $x \in \Sigma$ such that $(p, x, q) \notin \delta$. It follows from item 5 of Definition 2.3 that $x\mathsf{post}_q \not\subseteq \mathsf{post}_p$. Let $v_q \in \mathsf{post}_q$ such that $xv_q \notin \mathsf{post}_p$, then if $u_p$ is an incoming characteristic word of $p$ we get that $u_p x v_q \notin L$ since $\mathsf{post}_q = u^{-1}L$. Then it is impossible to add $(p, x, q)$ in $\delta$ else $\mathcal{A}$ would recognize the word $u_p x v_q$. □

Since we are interested in state minimality, a natural notion is the following:

**Definition 4.6.** An NFA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ is *thin* if, for any proper subset $Q'$ of $Q$, the language recognized by the automaton $\langle \Sigma, Q', I \cap Q', F \cap Q', \delta \cap (Q' \times \Sigma \times Q') \rangle$ is strictly included in the language recognized by $\mathcal{A}$.

In other words, an NFA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$, is thin if deleting any state in $\mathcal{A}$ (and its associated transitions) changes the language recognized by the automaton.

Clearly, every NFA contains a (non necessarily unique) thin sub-automaton recognizing the same language and every minimal NFA is thin. But the converse does not hold.

**Lemma 4.7.** *Any canonical RFSA is thin.*

*Proof.* Let us consider a canonical RFSA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$, recognizing a language $L$ and let $q \in Q$ be a state of $\mathcal{A}$. Let $u \in \Sigma^*$ be an incoming characteristic word of state $q$, then $\mathsf{post}_q = u^{-1}L$. We shall prove that there exists a word $v$ in $\mathsf{post}_q$ such that $uv$ were not recognized anymore if the state $q$ is deleted. Let us consider the language $K = \cup_{p \in \delta(I,u) \setminus \{q\}} \mathsf{post}_p$. The language $K$ is included in $\mathsf{post}_q$ and, since $\mathcal{A}$ is a canonical RFSA, $\mathsf{post}_q$ is a prime residual, this inclusion is strict. It follows that there exists a word $v \in (\mathsf{post}_q \setminus K)$. Then $uv \in L$ and $uv$ were no more recognized if state $q$ is deleted. □

Then we can state:

**Proposition 4.8.** *If a language $L$ is recognized by a unique thin and full NFA then $L$ is a biRFSA language.*

*Proof.* Let $L$ be a recognizable language. Let $\mathcal{A}$ be the canonical RFSA of $L$ and $\mathcal{B}$ be the canonical RFSA of $L^R$, the reverse of $L$. From Lemmata 4.5 and 4.7 , these two automata are full and thin. Moreover, it is clear that the reverse of a full and thin automaton is full and thin, then automaton $\mathcal{B}^R$, the reverse of $\mathcal{B}$, is full and thin. Hence, automata $\mathcal{A}$ and $\mathcal{B}^R$ are full and thin NFA which recognize $L$. Then $\mathcal{A} = \mathcal{B}^R$ and, by Corollary 3.3, $L$ is a biRFSA language. □

In order to obtain the converse of Proposition 4.8, we shall use the following lemmata.

**Lemma 4.9.** *Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a canonical biRFSA. Then for any states $p, q \in Q$ and for any non empty word $u \in \Sigma^+$, if $u\mathsf{post}_q \subseteq \mathsf{post}_p$, then $q \in \delta(p, u)$.*

*Proof.* Let $L$ be the language recognized by automaton $\mathcal{A}$. Let $u_p \in \Sigma^*$ be an incoming characteristic word of state $p$ and $v_q \in \Sigma^*$ be an outgoing characteristic word of state $q$. Then $u_p u v_q \in L$, and $u_p u \in L v_q^{-1} = \mathsf{pre}_q$. Let $u = xu'$ with $x \in \Sigma$. Then, there exists states $s, t \in Q$ such that $u_p \in \mathsf{pre}_s$, $t \in \delta(s, x)$ and $q \in \delta(t, u')$ which implies $x\mathsf{post}_t \subseteq \mathsf{post}_s$. Now, since $u_p \mathsf{post}_s \subseteq L$, it follows $\mathsf{post}_s \subseteq \mathsf{post}_p$, and $x(\mathsf{post}_t) \subseteq \mathsf{post}_p$. Finally, from definition of canonical biRFSA, $t \in \delta(p, x)$ then $q \in \delta(p, u)$. □

Let us consider again the mapping $h$ defined in order to prove Proposition 4.2. Recall that $h : Q \longrightarrow Q'$ maps each state $q \in Q$ to a state $h(q) = q' \in Q'$ such that $u_q \in \mathsf{pre}_{\mathcal{A}',q'}$ and $v_q \in \mathsf{post}_{\mathcal{A}',q'}$ where $u_q$ and $v_q$ are respectively incoming and outgoing characteristic words of $q$. Remark that $\mathsf{post}_{\mathcal{A}',q'} \subseteq u_q^{-1}L = \mathsf{post}_{\mathcal{A},q}$ and $\mathsf{pre}_{\mathcal{A}',q'} \subseteq L v_q^{-1} = \mathsf{pre}_{\mathcal{A},q}$.

**Lemma 4.10.** *Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a canonical biRFSA recognizing a language $L$ and $\mathcal{A}' = \langle \Sigma, Q', I', F', \delta' \rangle$ be an NFA which recognizes $L$. For any states $p, q \in Q$ and for any word $u \in \Sigma^*$, if $h(q) \in \delta'(h(p), u)$ then $q \in \delta(p, u)$.*

*Proof.* If $u = \varepsilon$ then $h(p) = h(q)$ and $p = q$. Else, $u_p u v_q \in L$ and it follows $u_p u \in \mathsf{pre}_{\mathcal{A},q}$ and $u_p u \mathsf{post}_{\mathcal{A},q} \subseteq L$. Then $u\mathsf{post}_{\mathcal{A},q} \subseteq \mathsf{post}_{\mathcal{A},p}$ and, from Lemma 4.9, $q \in \delta(p, u)$. □

We are now able to state:

**Proposition 4.11.** *If $L$ is a biRFSA language, the canonical biRFSA of $L$ is a sub-automaton of any full NFA recognizing $L$.*

*Proof.* Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be the canonical biRFSA of a language $L$ and let $\mathcal{A}' = \langle \Sigma, Q', I', F', \delta' \rangle$ be a full NFA which recognizes $L$. Let us consider automaton $\mathcal{A}'' = \langle \Sigma, h(Q), I' \cap h(Q), F' \cap h(Q), \delta' \cap h(Q) \times \Sigma \times h(Q) \rangle$. We shall prove that $\mathcal{A}$ and $\mathcal{A}''$ are isomorphic.

Let us prove first that, $\forall p, q \in Q, \forall x \in \Sigma, (h(q) \in \delta'(h(p), x)) \iff q \in \delta(p, x)$.

(1) If $h(q) \in \delta'(h(p), x)$ then $u_p x v_q \in L$, where $u_p$ is an incoming characteristic word of $p$ and $v_q$ is an outgoing characteristic word of $q$. It follows $u_p x \in \mathsf{pre}_{\mathcal{A},q}$ and $u_p x \mathsf{post}_{\mathcal{A},q} \subseteq L$. Since $x\mathsf{post}_{\mathcal{A},q} \subseteq \mathsf{post}_{\mathcal{A},p}$, we get $q \in \delta(p, x)$.

(2) If $q \in \delta(p, x)$, let us suppose that $h(q) \notin \delta'(h(p), x)$. Let $\mathcal{B}$ be the automaton obtained from automaton $\mathcal{A}'$, adding the transition $(h(p), x, h(q))$. We shall prove that $\mathcal{B}$ recognizes language $L$ which leads to a contradiction since $\mathcal{A}'$ is full.

Let $w$ be a word recognized by automaton $\mathcal{B}$. If there exists a path labelled by $w$ from $I'$ to $F'$ which does not use transition $(h(p), x, h(q))$ then $w$ is recognized by $\mathcal{A}'$ and $w \in L$. Else, $w = w_0 x w_1 \ldots x w_{k+1}$ with $w_0 \in \mathsf{pre}_{\mathcal{A}',h(p)}$, $w_{k+1} \in \mathsf{post}_{\mathcal{A}',h(q)}$ and, $\forall 1 \le i \le k, h(p) \in \delta'(h(q), w_i)$. From Lemma 4.10, $\forall 1 \le i \le k, p \in \delta(q, w_i)$. Now, since $w_0 \in \mathsf{pre}_{\mathcal{A}',h(p)} \subseteq \mathsf{pre}_{\mathcal{A},p}$ and $w_{k+1} \in \mathsf{post}_{\mathcal{A}',h(q)} \subseteq \mathsf{post}_{\mathcal{A},q}$, we get that $w$ is recognized by $\mathcal{A}$ and $w \in L$.

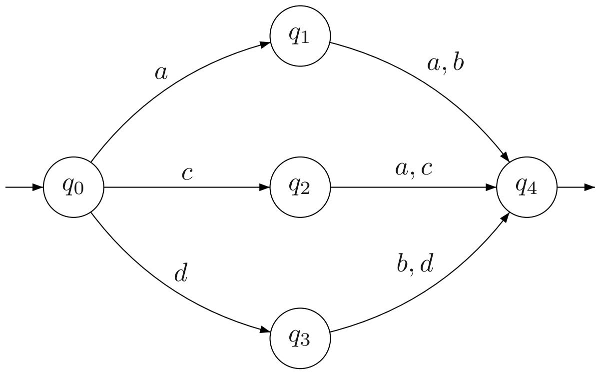



FIGURE 3. The canonical RFSA of $aa + ab + ca + cc + db + dd$.

Finally, since for any state $s$, $\mathsf{post}_{\mathcal{A}',h(s)} \subseteq \mathsf{post}_{\mathcal{A},s}$ and $\mathsf{pre}_{\mathcal{A}',h(s)} \subseteq \mathsf{pre}_{\mathcal{A},s}$, we get $h(I) \subseteq I'$ and $h(F) \subseteq F'$. □

Finally, we are able to give the following characterization of biRFSA languages, in term of minimal automata.

**Theorem 4.12.** *The following properties are equivalent:*

(1) *$L$ is a biRFSA language.*
(2) *The canonical RFSA of $L$ is a sub-automaton of any full NFA recognizing $L$.*
(3) *There exists a unique thin and full NFA recognizing $L$.*

*Proof.* From Proposition 4.11, we get that 1 implies 2, and from Proposition 4.8, we get that 3 implies 1. At last, if the canonical RFSA of a language $L$ is a sub-automaton of any full NFA recognizing $L$, it is a sub-automaton of every thin and full NFA recognizing language $L$, and since it is thin and full, it is isomorphic to every thin and full NFA recognizing language $L$. □

An easy corollary of Theorem 4.12 is the following:

**Corollary 4.13.** *If $L$ is a biRFSA language then every minimal NFA recognizing $L$ is a sub-automaton of the canonical RFSA of $L$.*

The converse of this corollary is false: the canonical RFSA given Figure 3 contains as sub-automata all minimal NFA recognizing the same language, since it is uniquely minimal, but it is not a biRFSA.

Let us finish this section with the study of the complexity of the problem to decide whether a language is a biRFSA language.

**Proposition 4.14.** *The problem to decide whether a given NFA recognizes a biRFSA language is PSPACE-complete.*

*Proof.* Let $\mathcal{A}$ be an NFA, and let us consider the following algorithm:

(1) Compute a minimal NFA $\mathcal{B}$, equivalent to $\mathcal{A}$.

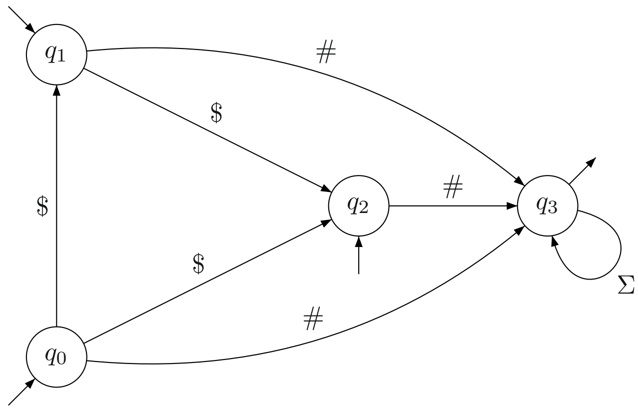



FIGURE 4. The canonical RFSA of $(\varepsilon + \$ + \$^2)\#\Sigma^*$.

(2) Compute a full and thin NFA $\mathcal{C}$ equivalent to $\mathcal{B}$.
(3) Check if $\mathcal{C}$ is a biRFSA.

If $\mathcal{C}$ is a biRFSA, then $L(\mathcal{A}) = L(\mathcal{C})$ is a biRFSA language and if $L(\mathcal{A})$ is a biRFSA language, it follows from the theorem 4.12 that $\mathcal{C}$ is the unique full and thin NFA equivalent to $\mathcal{A}$ and it is the canonical biRFSA of $L(\mathcal{A})$. According to [6], step 1 can be done in PSPACE. For step 2, since $\mathcal{B}$ is minimal, it is thin and it is sufficient to try all possibilities of adding transitions, or transforming states in final of initial ones without changing the recognized language: this can be done in PSPACE since it has been proved in [14] that the equivalence of two NFAs is a PSPACE-complete problem and we have only to add a number of transitions which is polynomially bounded on the number of states of $\mathcal{B}$. Finally it has been proved in [5] that the problem to decide whether a given NFA is an RFSA is a PSPACE-problem, then step 3 can be done in PSPACE testing whether $\mathcal{C}$ is an RFSA and whether the reverse of $\mathcal{C}$ is also an RFSA.

Conversely, in order to prove that the problem is PSPACE-hard, we shall reduce the problem of the universality of NFAs which has been proved to be PSPACE-complete in [8]. Let us consider a regular language $K$ defined over an alphabet $\Sigma$, and two fresh letters \$ and # which do not belong to $\Sigma$. We shall prove that the language $R = (\$ + \$^2)\#\Sigma^* + \#K$ is a biRFSA language if and only if $K = \Sigma^*$. First, it is easy to verify that the language $(\varepsilon + \$ + \$^2)\#\Sigma^*$ is a biRFSA language, since its canonical RFSA, given Figure 4 is clearly a biRFSA. Conversely, let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be the canonical RFSA of $R$ and let us suppose that $\mathcal{A}$ is a biRFSA. If $K \subsetneq \Sigma^*$, then for any word $u \in \Sigma^* \setminus K$ we get $R(\$\#u)^{-1} = \varepsilon + \$$. Moreover, it is easily seen that the language \$ is not a left residual of $R$, then $\varepsilon + \$$ is a prime left residual of $R$ and there exists a state $q \in Q$ such that $\mathsf{pre}_q = \varepsilon + \$$. It follows that $\mathsf{post}_q = \varepsilon^{-1}R = R$ or $\mathsf{post}_q = \$^{-1}R = (\varepsilon + \$)\#\Sigma^*$. In both cases $\mathcal{A}$ would recognize a word that is not in $R$: if $\mathsf{post}_q = R$ then $\$^3\# \in \mathsf{pre}_q.\mathsf{post}_q$ and if $\mathsf{post}_q = (\varepsilon + \$)\#\Sigma^*$ then $\#u \in \mathsf{pre}_q.\mathsf{post}_q$. Thus, if $K \subsetneq \Sigma^\star$, $R$ is not a biRFSA language.

It follows that the problem to decide whether a given NFA recognizes a biRFSA language is PSPACE-hard and then PSPACE-complete. □

## 5. Biseparable automata

In [15], Tamm and Ukkonen prove that the minimal DFA recognizing a bideterministic language $L$, that is a DFA whose reverse is deterministic, is the unique minimal NFA among all NFAs recognizing $L$. Proposition 4.2 states that the canonical RFSA of a biRFSA language is minimal. In this section, we study the family of languages for which the canonical RFSA is the unique minimal NFA.

**Definition 5.1.** A trim NFA $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ is called *separable* if: $\forall q \in Q, \exists u \in \Sigma^*, \delta(I, u) = \{q\}$.

Clearly, any separable automaton is always an RFSA, but the converse is false since the biRFSA given Figure 5 , recognizing the language $a^+$ is not separable:

**Definition 5.2.** An NFA $\mathcal{A}$ is called *biseparable* if both $\mathcal{A}$ and its reverse are separable. A language $L$ is called biseparable if it is recognized by some biseparable NFA.

Any biseparable NFA is clearly a biRFSA; we shall prove in Proposition 5.4 a stronger result. First, let us state:

**Lemma 5.3.** *Any biseparable NFA is full and thin.*

*Proof.* Let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a biseparable automaton for a language $L$. For any state $q \in Q$, there exists a word $u_q$ such that $u_q \in \mathsf{pre}_q$ and $\forall q' \neq q, u_q \notin \mathsf{pre}_{q'}$ and there exists a word $v_q$ such that $v_q \in \mathsf{post}_q$ and $\forall q' \neq q, v_q \notin \mathsf{post}_{q'}$. Clearly, $\mathcal{A}$ is thin, since if we remove a state $q$, the word $u_q v_q$ were not recognized any more. Moreover, if $q \notin I$, it follows that $v_q \notin L$ and it is not possible to add $q$ in $I$ without changing the language recognized by the automaton. Similarly, if $q \notin F$ then $u_q \notin L$ and it is not possible to add $q$ in $F$ without changing the language recognized by the automaton. Let us consider now $(p, x, q) \in (Q \times \Sigma \times Q) \setminus \delta$. Then $x v_q \notin \mathsf{post}_p$, else it should exists a state $q'$ such that $(p, x, q') \in \delta$ and $v_q \in \mathsf{post}'_q$. It follows that $u_p x v_q \notin L$ and it is not possible to add transition $(p, x, q)$. □

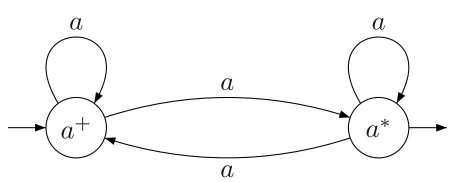



Figure 5. The canonical RFSA of $a^+$.

Clearly, any biseparable NFA is a biRFSA, so we obtain, from Theorem 4.12 and Lemma 5.3, the two following propositions.

**Proposition 5.4.** *Any biseparable NFA is a canonical biRFSA.*

**Proposition 5.5.** *A language is biseparable if and only if its canonical RFSA is biseparable.*

Proposition 5.5 leads to a decision procedure to know whether a regular language is biseparable: for instance, since the canonical RFSA given Figure 5 is not biseparable, it follows that $a^+$ is not a biseparable language. More precisely, we can give the complexity of the problem to decide whether a language is a biRFSA language:

**Proposition 5.6.** *The problem to decide whether a given NFA recognizes a biseparable language is PSPACE-complete.*

*Proof.* Let $\mathcal{A}$ be an NFA, and let us consider the following algorithm:

(1) Compute a minimal NFA $\mathcal{B}$, equivalent to $\mathcal{A}$.
(2) Compute a full and thin NFA $\mathcal{C}$ equivalent to $\mathcal{B}$.
(3) Check if $\mathcal{C}$ is a biRFSA.
(4) Check that it is not possible to remove any transition in $\mathcal{C}$ without changing the language which is recognized.

Steps 1, 2 and 3 are the same as in the PSPACE algorithm given in the proof of Proposition 4.14, and it is easily seen that step 4 can be done in PSACE, then the problem is in PSACE.

Conversely, we shall reduce again the problem of the universality of NFAs. For any regular language $K$ defined over the alphabet $\{x, y\}$, let us consider $L = (a+b)K + a(x+y)^*$. Then $L$ is biseparable if and only if $K = (x+y)^*$ and the problem is PSPACE-hard, then PSPACE complete. □

We are now able to state:

**Proposition 5.7.** *Any biseparable NFA is uniquely minimal.*

*Proof.* Let $\mathcal{B} = \langle \Sigma, Q', I', F', \delta' \rangle$ be a biseparable NFA for a language $L$ and let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be a minimal NFA for $L$. From Proposition 5.4, $\mathcal{B}$ is the canonical RFSA of $L$ and it follows from Corollary 4.13 that $\mathcal{A}$ is a sub-automaton of $\mathcal{B}$. For any state $q$, let us denote by $u_q$ a word such that $\delta'(I', u_q) = \{q\}$ and by $v_q$ a word such that $\delta'^R(F', v_q) = \{q\}$. Since $\mathcal{A}$ is a sub-automaton of $\mathcal{B}$, $\delta(I, u_q) = \{q\}$ and $\delta^R(F, v_q) = \{q\}$. It follows that $\mathcal{A}$ is biseparable, then it is full and thin from Lemma 5.3, and it is equal to $\mathcal{B}$ from Theorem 4.12. □

Remark that the above proposition improves the result of Tamm and Ukkonen since the family of bideterministic automata is strictly included in the family of biseparable automata, even over a one letter alphabet: let $n$ be an integer strictly greater than 1, then the canonical RFSA of language $L_n = (a^n(a^{n-1})^*)^*$ is not bideterministic but is biseparable. For example, the automaton given Figure 6 is the canonical RFSA of language $L_3$.

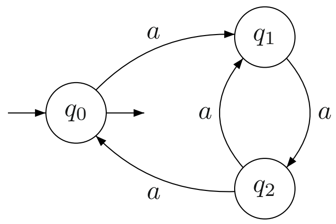



FIGURE 6. The canonical RFSA of $(a^3(a^2)^*)^*$.

Let us finish this section by the following characterization of the family of languages that are recognized by biseparable automata:

**Proposition 5.8.** *For any recognizable language $L$, the three following properties are equivalent:*

(1) *$L$ is a biseparable language;*
(2) *$L$ is a biRFSA language which is recognized by a unique minimal NFA;*
(3) *No prime residual of $L$ is included in the union of the other prime residuals of $L$.*

*Proof.* Implication 1 $\Longrightarrow$ 2 is stated in Proposition 5.7.

Let us prove 2 $\Longrightarrow$ 3: let $\mathcal{A} = \langle \Sigma, Q, I, F, \delta \rangle$ be the canonical RFSA of a biRFSA language $L$, by hypothesis, and from Proposition 4.2, $\mathcal{A}$ is uniquely minimal. For any state $q \in Q$, let $u_q$ and $v_q$ be respectively an incoming and an outgoing characteristic word of state $q$. Let us suppose that there exists a state $p$ such that $\mathsf{post}_p \subsetneq \cup_{q \in Q \setminus \{p\}} \mathsf{post}_q$. Then it exists $p' \neq p$ such that $v_p \in \mathsf{post}_{p'}$. It follows that $\mathsf{pre}_{p'} \subseteq \mathsf{pre}_p$ then $\mathsf{post}_p \subseteq \mathsf{post}_{p'}$. If $p'$ is an initial state then, $p$ is an initial state too since $\mathcal{A}$ is full, but the automaton $\mathcal{A}' = \langle \Sigma, Q, I \setminus \{p\}, F, \delta \rangle$ recognizes $L$ and is different from $\mathcal{A}$. If $p'$ is not an initial state, then there exist a state $s$ and a letter $x \in \Sigma$ such that $(s, x, p') \in \delta$. Then $(s, x, p) \in \delta$ since $\mathcal{A}$ is full, but the automaton $\mathcal{A}' = \langle \Sigma, Q, I, F, \delta \setminus \{(s, x, p)\} \rangle$ recognizes $L$ and is different from $\mathcal{A}$.

Let us now prove implication 3 $\Longrightarrow$ 1: let $L$ be a language such that every prime residual of $L$ is not included in the union of others prime residuals of $L$, let us consider $\mathcal{B} = \langle \Sigma, Q', I', F', \delta' \rangle$ the canonical RFSA of $L$. Let $q \in Q'$, then there exists a word $v_q \in \mathsf{post}_{\mathcal{B},q}$ such that $v_q \notin \cup_{q' \in Q' \setminus \{q\}} \mathsf{post}_{\mathcal{B},q'}$ and the reverse of $\mathcal{B}$ is separable. Let $u_q$ be an incoming characteristic word for state $q$ in $\mathcal{B}$, then for any state $q' \neq q$, it is not possible to have $q' \in \delta'(I, u_q)$, else $\mathsf{post}_{\mathcal{B},q'}$ which is a prime residual of $L$ were included in $\mathsf{post}_{\mathcal{B},q}$ which is another prime residual. It follows that $\mathcal{B}$ is separable, hence biseparable. $\square$

Notice that the unicity of a minimal NFA for a language $L$ is not a sufficient condition for $L$ to be a biseparable language as it is shown in Example 2.7

## 6. Conclusion

We have defined biRFSAs that are RFSAs whose the reverse is also an RFSA and biRFSA languages which are those languages recognized by biRFSAs. We proved that the canonical RFSA of a biRFSA language $L$ is a minimal NFA for $L$ and that each minimal NFA for $L$ is isomorphic to a sub-automaton of the canonical RFSA of $L$.

We have established a characterization of the family of biRFSA languages in term of minimal NFA: a recognizable language $L$ is a biRFSA language if and only if there is a unique full and thin NFA recognizing it.

We have introduced the family of biseparable NFAs which strictly contains the family of bideterministic automata. We have shown that biseparable NFAs are uniquely minimal, improving the result of [15].

Further works on the family of biRFSA languages will concern identification of these languages by learning algorithms. The good properties of the family of biRFSA languages in term of minimal NFAs allow to represent biRFSA languages by *small* canonical (non deterministic) automata that can be the targets of the learning algorithms. Moreover, most of the learning algorithms defined for the already known family of regular languages which are learnable from positive examples exploit the determinism of the target automaton. With the family of biRFSA languages, one can define some efficient new learning algorithms, based on the notion of residuals languages, which infer non deterministic automata.

*Acknowledgements*. We are grateful to the referee for comments which helped improve the presentation of this paper.

## References

[1] Dana Angluin. Inference of reversible languages. *J. ACM* **29** (1982) 741–765.

[2] André Arnold, Anne Dicky, and Maurice Nivat. A note about minimal non deterministic finite automata. *Bull. EATCS* **47** (1992) 166–169.

[3] Christian Carrez. On the minimalization of non-deterministic automaton. Technical report, Laboratoire de Calcul de la Faculté des Sciences de Lille (1970).

[4] Jean-Marc Champarnaud and Fabien Coulon. NFA reduction algorithms by means of regular inequalities. *Theoretical Computer Science* **327** (2004) 241–253.

[5] François Denis, Aurélien Lemay, and Alain Terlutte. Residual finite state automata. In *Proceedings of STACS 2001* **2010**. Springer-Verlag, Dresden (2001) 144–157.

[6] Michael R. Garey and David S. Johnson. *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York (1979).

[7] JE Hopcroft and JD Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, Massachusetts (1979).

[8] Harry B. Hunt III, Daniel J. Rosenkrantz, and Thomas G. Szymanski. On the equivalence, containment, and covering problems for the regular and context-free languages. *Journal of Computer and System Sciences* **12** (1976) 222–268.

[9] Michel Latteux, Aurélien Lemay, Yves Roos, and Alain Terlutte. Identification of biRFSA languages. *Theoretical Computer Science* **356** (2006) 212–223.

[10] Michel Latteux, Yves Roos, and Alain Terlutte. BiRFSA languages and minimal NFAs. Technical Report GRAPPA-0205, GRAppA, (2006).

[11] Oliver Matz and Andreas Potthoff. Computing small nondeterministic automata. In U.H. Engberg, K.G. Larsen, and A. Skou, Eds., *Workshop on Tools and Algorithms for the Construction and Analysis of Systems* (1995).

[12] Dominique Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science, Volume B: Formal Models and Sematics (B)*. Elsevier (1990) 1–57.

[13] Jean-Eric Pin. On reversible automata. In *Proceedings of the first LATIN conference, Saõ-Paulo*. Lecture Notes in Computer Science **583**. Springer Verlag (1992) 401–416.

[14] L.J. Stockmeyer and A.R. Meyer. Word problems requiring exponential time(preliminary report). In *STOC '73: Proceedings of the fifth annual ACM symposium on Theory of computing*. ACM Press, NY, USA (1973) 1–9.

[15] Hellis Tamm and Esko Ukkonen. Bideterministic automata and minimal representations of regular languages. *Theoretical Computer Science* **328** (2004) 135–149.

Communicated by J.-É. Pin.
Received June 1st, 2006. Accepted February 25, 2008.