Complementation of rational sets on countable scattered linear orderings

Chloé Rispal
IGM, Université de Marne-la-Vallée,
http://www-igm.univ-mlv.fr/~rispal/

Olivier Carton
LIAFA, Université Paris 7,
http://www.liafa.jussieu.fr/~carton/

April 17, 2006

Abstract

In a preceding paper (Bruyère and Carton, automata on linear orderings, MFCS'01), automata have been introduced for words indexed by linear orderings. These automata are a generalization of automata for finite, infinite, bi-infinite and even transfinite words studied by Büchi. Kleene's theorem has been generalized to these words. We prove that rational sets of words on countable scattered linear orderings are closed under complementation using an algebraic approach.

1 Introduction

In his seminal paper [14], Kleene showed that automata on finite words and regular expressions have the same expressive power. Since then, this result has been extended to many classes of structures like infinite words [7, 17], bi-infinite words [12, 18], transfinite words [9, 1], traces, trees, pictures...

In [5], automata accepting linear-ordered structures have been introduced with corresponding rational expressions. These linear structures include finite words, infinite, transfinite words and their mirrors. These automata are usual automata on finite words, extended with limit transitions. A Kleene-like theorem was proved for words on countable scattered linear orderings. Recall that an ordering is scattered if it does not contain a dense sub-ordering.

For many structures, the class of rational sets is closed under many operations like substitutions, inverse substitutions and boolean operations. As for boolean operations, the closure under union and intersection are almost always easy to get. The closure under complementation is often much more difficult to prove. This property is important both from the practical and the theoretical point of view. It means that the class of rational sets forms an effective boolean algebra. It is used whenever some logic is translated into automata. For instance, in both proofs of the decidability of the monadic second-order theory of the integers by Büchi [8] and the decidability of the monadic second-order theory of the infinite binary tree by Rabin [21], the closure under complementation of automata is the key property.

In [5], the closure under complementation was left as an open problem. In this paper, we solve that problem in a positive way. We show that the complement of a rational set of words on countable scattered linear orderings is also rational.

The classical method to get an automaton for the complement of a set of finite words accepted by an automaton $\mathcal{A}$ is through determinization. It is already non-trivial that the complement of a rational set of infinite words is also rational. The determinization method cannot be easily extended to infinite words. In his seminal paper [8], Büchi used another approach based on a congruence on finite words and Ramsey's theorem. This method is somehow related to our algebraic approach. McNaughton extended the determinization method to infinite words [15] proving that any Büchi automaton is equivalent to a deterministic Muller automaton. Büchi pushed further this method and extended it to transfinite words [9]. It is then very complex. In [4], the algebraic approach was used to give another proof of the closure under complementation for transfinite words. In [11], we have already proved the result for words on countable scattered linear orderings of finite ranks. The determinization method cannot be applied because any automaton is not equivalent to a deterministic one. In that paper, we extended the method used by Büchi in [8] using an additional induction on the rank. Since ranks of countable scattered linear orderings range over all countable ordinals, this approach is not suitable for words on all these orderings. In this paper, we prove the whole result for all countable scattered linear orderings using an algebraic approach. We define a generalization of semigroups, called $\diamond$-semigroups. We show that, when finite, these $\diamond$-semigroups are equivalent to automata. We also show that, by analogy with the case of finite words, a canonical $\diamond$-semigroup, called the syntactic $\diamond$-semigroup, can be associated with any rational set X. It has the property of being the smallest $\diamond$-semigroup recognizing X. A continuation of this paper would be to extend the equivalence between star-free sets, first order logic and aperiodic semigroups [24, 16, 3] and also between rational sets and the monadic second order theory.

Both hypotheses that the orderings are scattered and countable are really necessary. Büchi already pointed out that rational sets of transfinite words of length greater that ω_1 (the least non-countable ordinal) are not closed under complement. It can be proved that the set of all words on scattered linear orderings, viewed as a subset of words on all linear orderings, is not rational although its complement is rational.

Our proof of the complementation closure is effective. Given an automaton $\mathcal{A}$, it gives another automaton $\mathcal{B}$ that accepts words that are not accepted by $\mathcal{A}$. It gives another proof of the decidability of the equivalence of these automata [6].

This paper is organized as follows. Definitions concerning linear orderings and rational sets are first recalled in Sections 2 and 3. Then, Section 4 introduces the algebraic structure of $\diamond$-semigroup. Section 5 sketches the proof that any set accepted by a finte automaton is recognizable and the full proof of the converse is given in Section 6. Since rational sets are exactly those accepted by automata [5], this proves that rational sets are exactly recognizable sets. This algebraic caracterisation proves trivially the closure under complementation of rational sets. Finally, the syntactic $\diamond$-semigroup corresponding to a rational set is defined in Section 7.

2 Words on linear orderings

This section recalls basic definitions on linear orderings but the reader is referred to [23] for a complete introduction. Hausdorff's characterization of countable scattered linear orderings is given and words indexed by linear orderings are introduced.

Let J be a set equipped with an order $<$. The ordering J is *linear* if for any j and k in J such that $j \neq k$, either $j < k$ or $k < j$. Let A be a finite alphabet. A *word* $x = (a_j)_{j \in J}$ indexed by a linear ordering J is a function from J to A. J is called the *length* of x. For instance ω is the length of right-infinite words $a_0 a_1 \cdots$ and ζ is the length of bi-infinite words $\cdots a_{-1} a_0 a_1 \cdots$.

2.1 Product of words indexed by linear orderings

For any linear ordering J, we denote by $-J$ the opposite linear ordering that is the set J equipped with the opposite ordering. For instance, $-\omega$ is the linear ordering of negative integers.

The sum $J + K$ of two linear orderings is the set $J \cup K$ equipped with the ordering $<$ extending the orderings of J and K by setting $j < k$ for any $j \in J$ and $k \in K$. Next, the *sum* $\sum_{j \in J} K_j$ is the set of all pairs (k, j) such that $k \in K_j$ equipped with the ordering defined by $(k_1, j_1) < (k_2, j_2)$ if and only if $j_1 < j_2$ or ($j_1 = j_2$ and $k_1 < k_2$ in K_{j_1}).

The sum of linear orderings helps to define the products of words. Let J be a linear ordering and let $(x_j)_{j \in J}$ be words of respective length K_j for any $j \in J$. The word $x = \prod_{j \in J} x_j$ obtained by concatenation of the words x_j with respect to the ordering on J is of length $L = \sum_{j \in J} K_j$. For instance, if for any $j \in \omega$, we set $x_j = a^{\omega^j}$, then $x = \prod_{j \in \omega} x_j$ is the word $x = a^{\omega^\omega}$ of length $\sum_{j \in \omega} \omega^j = \omega^\omega$. The sequence $(x_j)_{j \in J}$ of words is called a *J-factorization* of the word $x = \prod_{j \in J} x_j$.

2.2 Scattered linear orderings

A linear ordering J is *dense* if for any j and k in J such that $j < k$, there exists an element i of J such that $j < i < k$. It is *scattered* if it contains no dense sub-ordering. The ordering ω of natural integers and the ordering ζ of relative integers are scattered. More generally, ordinals are scattered orderings. We denote by $\mathcal{N}$ the subclass of finite linear orderings, $\mathcal{O}$ the class of ordinals and $\mathcal{S}$ the class of countable scattered linear orderings. The following characterization of scattered linear orderings is due to Hausdorff.

Theorem 1. *[Hausdorff [13]] A countable linear ordering J is scattered if and only if J belongs to $\bigcup_{\alpha \in \mathcal{O}} V_\alpha$ where the classes V_α are inductively defined by:*

1. $V_0 = \{\mathbf{0}, \mathbf{1}\}$
2. $V_\alpha = \{\sum_{j \in J} K_j \mid J \in \mathcal{N} \cup \{\omega, -\omega, \zeta\}$ *and* $K_j \in \bigcup_{\beta < \alpha} V_\beta\}$.

where $\mathbf{0}$ *and* $\mathbf{1}$ *are respectively the orderings with zero and one element.*

For instance, ω^2 and $\sum_{\omega} -\omega$ belong to V_2 and for any integer n, ω^n belongs to V_n. Each scattered linear ordering is defined from the ordering of one element using finite sums, ω-sums and $-\omega$-sums. Intuitively, the rank of a linear ordering is the maximum number of nested ω-sums and $-\omega$-sums. For any ordinal α, the class V_α does not contain exactly the orderings of rank α since the finite product does not modify the rank. For instance, the ordering $\omega + \omega$ of rank 1 belongs to the class V_2. Thus, we use slightly different inductive classes : For any ordinal α, the class W_α is defined by :

$$W_\alpha = \left\{ \sum_{j \in J} K_j \mid J \in \mathcal{N} \text{ et } K_j \in V_\alpha \right\}.$$

Those classes are strictly intermediate to the previous ones : the inclusions $V_\alpha \subset W_\alpha \subset V_{\alpha+1}$ are satisfied for any ordinal α. Formally, the *rank* of a linear ordering J is the smallest ordinal α such that $J \in W_\alpha$. The orderings of rank 0 are the finite ones. In order to simplify the proofs, we also use the classes U_α defined for any ordinal α by

$$U_\alpha = \left\{ \sum_{j \in J} K_j \mid J \in \mathcal{N} \cup \{\omega, -\omega\} \text{ et } K_j \in \bigcup_{\beta < \alpha} W_\beta \right\}.$$

By Theorem 1, the ranks of scattered linear orderings range over all ordinals : $\mathcal{S} = \bigcup_{\alpha \in \mathcal{O}} W_\alpha$. To prove that a property holds for all scattered linear orderings, we often use an induction on the rank. We prove that the property holds for the set of finite orderings W_0. Then, for any ordinal α, we suppose that the property holds for any ordering of rank $\beta < \alpha$ and prove that it is stable under

ω-sums and $-\omega$-sums. This shows that the property is verified for all linear orderings of U_α. Finally, the case of linear orderings of W_α is checked with the finite sum.

$$W_\alpha = \left\{ \sum_{j \in J} K_j \mid J \in \mathcal{N} \text{ et } K_j \in U_\alpha \right\}.$$

We denote by $A^\diamond$ the set of all words over A indexed by countable scattered linear orderings and for any ordinal α, we denote by A^{W_α} (respectively A^{U_α}) the set of words of rank lower than or equal to α (respectively the set of words whose length belongs to U_α).

3 Rational sets of words on linear orderings

Bruyère and Carton have introduced rational expressions and automata for words indexed by countable scattered linear orderings. They have proved that a set of words is rational if and only if it is accepted by a finite automaton. This result is an extention of Kleene's theorem. This section shortly recalls definitions of rational operations and automata but the reader is referred to [5] for more details.

3.1 Rational expressions

Let A be a finite alphabet. The set $Rat(A^\diamond)$ of rational sets of words over A indexed by countable scattered linear orderings is the smallest set containing $\{a\}$ for any $a \in A$ and closed under the following rational operations defined for any subsets X and Y of $A^\diamond$ by :

$$\begin{array}{lcllcl}
X + Y & = & \{z \mid z \in X \cup Y\} & & & \\
X \cdot Y & = & \{x \cdot y \mid x \in X, y \in Y\} & X^* & = & \{\prod_{j=1}^{n} x_j \mid n \in \mathcal{N}, x_j \in X\} \\
X^\omega & = & \{\prod_{j \in \omega} x_j \mid x_j \in X\} & X^{-\omega} & = & \{\prod_{j \in -\omega} x_j \mid x_j \in X\} \\
X^\# & = & \{\prod_{j \in \alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\} & X^{-\#} & = & \{\prod_{j \in -\alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\}
\end{array}$$

$$X \diamond Y = \{\prod_{j \in J \cup \hat{J}^*} z_j \mid J \in \mathcal{S} \setminus \emptyset, z_j \in X \text{ if } j \in J \text{ and } z_j \in Y \text{ if } j \in \hat{J}^*\} \text{ where}$$

$$\hat{J}^* = \hat{J} \setminus \{(\emptyset, J), (J, \emptyset)\}.$$

The notation $\hat{J}$ is defined in the next section.

3.2 Automata on linear orderings

An automaton on linear orderings is a classical finite automaton with additional limit transitions of the form $P \longrightarrow q$ or $q \longrightarrow P$ where P is a set of states.

Definition 2. *An* automaton $\mathcal{A} = (Q, A, E, I, F)$ on linear orderings *is defined by a finite set of states Q, a finite alphabet A, a set of transitions $E \subseteq (Q \times$*

$A \times Q) \cup (\mathcal{P}(Q) \times Q) \cup (Q \times \mathcal{P}(Q))$ and sets of initial and final states $I \subseteq Q$ and $F \subseteq Q$.

The definition of paths is based on the notion of cut that we explain now. Let $J \in \mathcal{S}$ be a scattered linear ordering. A *cut* of J is a a partition of J into two intervals (K, L) such that $\forall k \in K, \forall \ell \in L, k < \ell$. The set $\hat{J}$ denotes the set of all cuts of the ordering J: $\hat{J} = \{(K, L) | K \cup L = J \wedge \forall k \in K, \forall \ell \in L, k < \ell\}$. Then, a path labelled by a word x of lenght J is a function from the set $\hat{J}$ into the set of states. As the set $\hat{J}$ is naturally equipped with the ordering $(K_1, L_1) < (K_2, L_2)$ if and only if $K_1 \subset K_2$, a path labelled by a word of length J is a word over Q of length $\hat{J}$.

Let $\gamma = (q_c)_{c \in \hat{J}}$ be a word of length $\hat{J}$ over Q, the limit sets of states of γ at a given cut c of $\hat{J}$ are defined by:

$$\lim_{c^-} \gamma = \{q \in Q | \ \forall c^{'} < c, \ \exists c^{''} \quad c^{'} < c^{''} < c \text{ and } q = q_{c^{''}}\}$$

$$\lim_{c^+} \gamma = \{q \in Q | \ \forall c^{'} > c, \ \exists c^{''} \quad c < c^{''} < c^{'} \text{ and } q = q_{c^{''}}\}$$

Definition 3. *Let $\mathcal{A} = (Q, A, E, I, F)$ be an automaton on linear orderings and let $x = (a_j)_{j \in J}$ be a word of length J on A. A* path *γ of label x in $\mathcal{A}$ is a word $\gamma = (q_c)_{c \in \hat{J}}$ of length $\hat{J}$ over Q such that for any $(K, L) \in \hat{J}$:*

- *If there exists $\ell \in L$ such that $(K \cup \{\ell\}, L \setminus \{\ell\}) \in \hat{J}$ then $q_{(K,L)} \xrightarrow{a_\ell} q_{(K \cup \{\ell\}, L \setminus \{\ell\})} \in E$ else $q_{(K,L)} \rightarrow \lim_{(K,L)^+} \gamma \in E$*
- *If there exists $k \in K$ such that $(K \setminus \{k\}, L \cup \{k\}) \in \hat{J}$ then $q_{(K \setminus \{k\}, L \cup \{k\})} \xrightarrow{a_k} q_{(K,L)} \in E$ else $\lim_{(K,L)^-} \gamma \rightarrow q_{(K,L)} \in E$.*

Thus, if a cut has a predecessor or a successor, usual transitions are used, otherwise the path uses limit transitions.

As $\hat{J}$ has the least element $(\emptyset, J)$ and the greatest element $(J, \emptyset)$ for any linear ordering J, a path has always a first and a last state. A word is *accepted* by an automata if it is the label of a path leading from an initial state to a final state. We denote by $p \stackrel{x}{\Longrightarrow} q$ the existence of a path leading from the state p to the state q of label x.

Example 4. For instance, the word $(a^{-\omega}b)^\omega$ of lenght $\sum_\omega -\omega$ is accepted by the automaton of Figure 1. It is the label of the path $(01^{-\omega})^\omega 0$. For any cut of the form $((-\omega) * n, \sum_\omega -\omega)$ where n is a natural integer, this path has $\{1\}$ as right limit set and uses the limit transition $0 \rightarrow \{1\}$. At the last cut $(\sum_\omega -\omega, \emptyset)$, the left limit set is $\{0, 1\}$ and the path ends with the limit transition $\{0, 1\} \rightarrow 0$.

It has been proved in [5] that automata and rational expressions have the same expressive power.

Theorem 5. *[5] A set of words indexed by countable scattered linear orderings is rational if and only if it is accepted by a finite automata.*

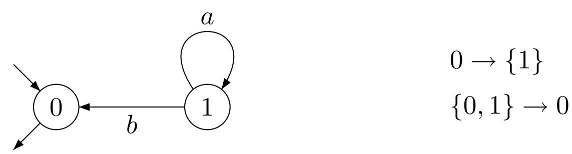

Figure 1: Automaton on linear orderings accepting the set $(a^{-\omega}b)^{\#}$.

4 Algebraic characterization of rational sets

A semigroup is a set S equipped with an associative binary product. The semigroup S in which had been added a neutral element is denoted by S^1. An element $e \in S$ is an *idempotent* if $e^2 = e$ and the set of idempotents of S is denoted by $E(S)$. A pair $(s, e) \in S \times S$ is *right linked* (respectively left linked) if $e \in E(S)$ and $se = s$ (respectively $es = s$). Two right linked pairs (s_1, e_1) and (s_2, e_2) are *conjugated* if there exists a, $b \in S^1$ such that $e_1 = ab$, $e_2 = ba$, $s_1 a = s_2$ and $s_2 b = s_1$. The conjugacy relation is an equivalence relation on right linked pairs [19]. Recall that the Green's relations are defined from the following preorders:

$$\begin{array}{ccc} s \leq_{\mathcal{R}} t & \Longleftrightarrow & \exists a \in S^1,\ s = ta \\ s \leq_{\mathcal{L}} t & \Longleftrightarrow & \exists a \in S^1,\ s = at \\ s \leq_{\mathcal{J}} t & \Longleftrightarrow & \exists a, b \in S^1,\ s = atb \end{array}$$

For any $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}\}$, $s\mathcal{K}t$ if and only if $s \leq_{\mathcal{K}} t$ and $t \leq_{\mathcal{K}} s$. We also denote by $s <_{\mathcal{K}} t$ iff $s \leq_{\mathcal{K}} t$ and not $t \leq_{\mathcal{K}} s$. Recall that the equivalence relation $\mathcal{D} = \mathcal{R}\mathcal{L} = \mathcal{L}\mathcal{R}$ is equal to $\mathcal{J}$ when S is finite.

4.1 $\diamond$-semigroups

The product of semigroups is generalized to recognize sets of words indexed by countable scattered linear orderings. A $\diamond$-semigroup is a generalization of a usual semigroup. The product of a sequence indexed by any scattered ordering is defined. For any set S, recall that $S^\diamond$ denotes the set of words over S indexed by scattered linear orderings.

Definition 6. *A* $\diamond$-semigroup *is a set S equipped with product $\pi : S^\diamond \longrightarrow S$ which maps any word of $S^\diamond$ to an element of S such that*

- *for any element s of S, $\pi(s) = s$.*

- *for any word x of $S^\diamond$ and for any factorization $x = \prod_{j \in J} x_j$ where $J \in \mathcal{S}$,*

$$\pi(x) = \pi(\prod_{j \in J} \pi(x_j))$$

The latter condition is a generalization of associativity.

For instance, the set $A^\diamond$ equipped with the concatenation is a $\diamond$-semigroup.

Example 7. The set $S = \{0, 1\}$ equipped with the product π defined for any $u \in S^\diamond$ by $\pi(u) = 0$ if u has at least one occurrence of the letter 0 and $\pi(u) = 1$ otherwise is a $\diamond$-semigroup.

For any two elements s and t of a $\diamond$-semigroup (S, π), the finite product $\pi(st)$ is merely denoted by st.

A *sub-$\diamond$-semigroup* T of a $\diamond$-semigroup S is a subset of S closed under product. A *morphism of $\diamond$-semigroup* is an application which preserves the product. A *congruence* of $\diamond$-semigroup is an equivalence relation $\sim$ stable under product: If $s_j \sim t_j$ for any $j \in J$, then $\pi(\prod_{j \in J} s_j) \sim \pi(\prod_{j \in J} t_j)$. The set $S/\sim$ is a $\diamond$-semigroup. A $\diamond$-semigroup T is a *quotient* of a $\diamond$-semigroup S if there exists an onto morphism from S to T. A $\diamond$-semigroup T *divides* S if T is the quotient of a sub-$\diamond$-semigroup of S.

4.2 Finite $\diamond$-semigroups

A $\diamond$-semigroup (S, π) is said to be finite if S is finite. Even when S is finite, the function π is not easy to describe because the product of any sequence has to be given. It turns out that the function π can be described using a semigroup structure on S with two additional functions (called τ and $-\tau$) from S to S. This gives a finite description of the function π. The functions τ and $-\tau$ are the counterpart of limit transitions of automata. This finite description is based on the next Lemma which follows directly from Ramsey's Theorem [22].

Let $x = \prod_{i \in \omega} x_i$ an ω-factorization. Another factorization $x = \prod_{i \in \omega} y_i$ is called a *superfactorization* if there is a sequence $(k_i)_{i \in \omega}$ of integers such that $y_0 = x_0 \ldots x_{k_0}$ and $y_i = x_{k_{i-1}+1} \ldots x_{k_i}$ for all $i \geq 1$.

Lemma 8. *Let $\varphi : A^\diamond \longrightarrow S$ be a morphism into a finite $\diamond$-semigroup. For any factorization $x = \prod_{i \in \omega} x_i$, there exists a superfactorization $x = \prod_{i \in \omega} y_i$ and a right linked pair $(s, e) \in S \times E(S)$ such that $\varphi(y_0) = s$ and $\varphi(y_i) = e$ for any $i > 0$.*

Such a factorization is called a *ramseyan factorization*, see Theorem 3.2 in [20].

Definition 9. *Let S be a semigroup. A function $\tau : S \longrightarrow S$ (respectively $-\tau : S \longrightarrow S$) is* compatible to the right with *S (respectively* to the left*) if and only if for any s, t in S and any integer n the following properties hold: $s(ts)^\tau = (st)^\tau$ and $(s^n)^\tau = s^\tau$ (respectively $(st)^{-\tau} s = (ts)^{-\tau}$ and $(s^n)^{-\tau} = s^{-\tau}$).*

The product of a finite $\diamond$-semigroup S can be finitely described by functions compatible to the right and to the left with S.

Theorem 10. *Let (S, π) be a finite $\diamond$-semigroup. The binary product defined for any s, t in S by $s \cdot t = \pi(st)$ naturally endows a structure of semigroup and*

the functions τ and $-\tau$ respectively defined by $s^\tau = \pi(s^\omega)$ and $s^{-\tau} = \pi(s^{-\omega})$ are respectively compatible to the right and to the left with S.
Conversely, let S be a finite semigroup and let τ and $-\tau$ be functions respectively compatible to the right and to the left with S. Then S can be uniquely endowed with a structure of $\diamond$-semigroup (S,π) such that $s^\tau = \pi(s^\omega)$ and $s^{-\tau} = \pi(s^{-\omega})$.

Proof. The first part of the theorem follows directly from the associativity of the product π. Conversely, let S be a finite semigroup and let τ and $-\tau$ be functions respectively compatible to the right and to the left with S. The product of a word $x = (s_j)_{j\in J}$ over S of length $J \in \mathcal{S}$ is defined by induction on $\alpha \in \mathcal{O}$ for any $J \in W_\alpha$ by the following way:

Let $J \in W_0$ and let $x \in S^J$. There exists an integer m and $s_1, \ldots, s_m$ in S such that $x = s_1 \ldots s_m$. We set $\pi(x) = s_1 \cdot s_2 \ldots s_m$.

Let $J \in U_\alpha$ where $\alpha > 1$ and let $x \in S^J$. The linear ordering J can be decomposed as a sum $J = \sum_{i\in I} K_i$ where $I \in \mathcal{N} \cup \{\omega, -\omega\}$ and for all $i \in I$, $K_i \in \bigcup_{\beta<\alpha} W_\beta$. There exists a factorization $x = \prod_{i\in I} x_i$ such that for all $i \in I$, $|x_i| = K_i$.

- $J = \{1, \ldots, m\} \in \mathcal{N}$: we set $\pi(x) = \pi(x_1) \ldots \pi(x_m)$.
- $J = \omega$: There exists a superfactorization $x = \prod_{i\in\omega} y_i$ and a right linked pair $(s, e) \in S \times E(S)$ such that $\varphi(y_0) = s$ and $\varphi(y_i) = e$ for any $i > 0$. We set $\pi(x) = se^\tau$.
- $J = -\omega$: Symmetrically to the previous case, we set $\pi(x) = e^{-\tau}s$.

Finally, let $J \in W_\alpha$. The linear ordering J can be decomposed as a finite sum $J = K_1 + K_2 + \ldots K_m$ where for all $1 \le i \le m$, $K_i \in U_\alpha$. There exists a factorization $x = x_1x_2 \ldots x_m$ such that for all $1 \le i \le m$, $|x_i| = K_i$ and we set $\pi(x) = \pi(x_1) \ldots \pi(x_m)$.

Since two linked pairs associated with two factorizations of a word are conjugated [20], it can be proved by induction on α that π is uniquely defined and associative on $S^\diamond$.

Example 11. The $\diamond$-semigroup $S = \{0, 1\}$ of Example 7 is defined by the finite product $00 = 01 = 10 = 0$ and $11 = 1$ and by the compatible functions τ and $-\tau$ defined by $0^\tau = 0^{-\tau} = 0$ and $1^\tau = 1^\tau = 1$.

4.3 Recognizability

It is well known that rational sets of finite words are exactly those recognized by finite semigroups. This result is generalized for words indexed by countable scattered linear orderings.

Definition 12. *Let S and T be two $\diamond$-semigroups. The $\diamond$-semigroup T* recognizes *a subset X of S if and only if there exists a morphism $\varphi : S \longrightarrow T$ and a* *subset $P \subseteq T$ such that $X = \varphi^{-1}(P)$. A set $X \subseteq A^\diamond$ is* recognizable *if and only if there exists a* ***finite*** *$\diamond$-semigroup recognizing it.*

Example 13. The set $S = \{0,1\}$ equipped with the product π defined for any $u \in S^\diamond$ by $\pi(u) = 1$ if $u \in 1^\#$ and $\pi(u) = 0$ otherwise is a $\diamond$-semigroup. It is also defined by the finite product $00 = 01 = 10 = 0$ et $11 = 1$ and by the compatible functions τ and $-\tau$ defined by $0^\tau = 0^{-\tau} = 1^{-\tau} = 0$ and $1^\tau = 1$. Define the morphism of $\diamond$-semigroup $\varphi : A^\diamond \longrightarrow S$ by $\varphi(a) = 1$ for any $a \in A$. The set $A^\#$ is recognizable since $A^\# = \varphi^{-1}(\{1\})$.

For any finite alphabet A, $Rec(A^\diamond)$ denotes the set of subsets of $A^\diamond$ recognizable by a finite $\diamond$-semigroup.

Theorem 14. *A set of words indexed by countable scattered linear orderings is rational iff it is recognizable.*

Example 15. The set $X = (ab)^\diamond$ is recognized by the $\diamond$-semigroup $S = \{s,t,e,f,0\}$ whose product is defined by $st = e$, $ts = f$, $ee = e$, $ff = f$, $es = s$, $ft = t$, $sf = s$, $te = t$, $e^\tau = e$, $e^{-\tau} = e$, $f^\tau = t$, $f^{-\tau} = s$ where any other product is equal to 0. Defining the morphism $\varphi : A^\diamond \to S$ by $\varphi(a) = s$ and $\varphi(b) = t$, we get $X = \varphi^{-1}(e)$.

If X is recognized by a morphism $\varphi : S \longrightarrow T$, the set $A^\diamond \setminus X$ is also recognized by φ since $A^\diamond \setminus X = \varphi^{-1}(S \setminus P)$. Therefore, we obtain following theorem.

Theorem 16. *Rational sets of words on countable scattered linear orderings are closed under complementation.*

Example 17. The set $X = A^*$ is recognized by the $\diamond$-semigroup $S = \{0,1\}$ whose product is defined by $11 = 1$, $01 = 10 = 00 = 0$ and by the compatible functions $0^\tau = 0^{-\tau} = 1^\tau = 1^{-\tau} = 0$. Define the morphism $\varphi : A^\diamond \to S$ by $\varphi(a) = 1$ for any $a \in A$. One gets $X = \varphi^{-1}(1)$ and the complement $A^\diamond\ X = (A^\diamond)^\omega A^\diamond + A^\diamond (A^\diamond)^{-\omega} = \varphi^{-1}(0)$.

The two next sections are devoted the proof of Theorem 14.

5 From automata to $\diamond$-semigroups

This proof is just an adaptation of a proof of [4].

Proposition 18. *Any rational set of $A^\diamond$ is recognizable.*

Proof. Let $\mathcal{A} = (Q, A, E, I, F)$ be an automaton on linear orderings accepting a set $X \subseteq A^\diamond$. The *content* of a path is the set of states occurring in the path and $p \overset{x}{\underset{P}{\Longrightarrow}} q$ denotes a path leading from p to q of label x and of content P. Let $T = \mathcal{P}(Q)$ be the set of all subsets of Q and $K = \mathcal{P}(T)$ be the set of subsets of T. The set K is equipped with the following product and union:

$$kk' = \{t \cup t' \mid t \in k,\, t' \in k'\} \text{ and } k + k' = k \cup k'$$

Let S be the set of all $Q \times Q$ matrices whose entries are in K with product defined by:

$$(m \cdot m')_{q,q'} = \bigcup_{p \in Q} m_{q,p} \cdot m'_{p,q'} = \{t \cup t' \mid \exists p \in Q,\, t \in m_{q,p},\, t' \in m'_{p,q'}\}$$

The semigroup S is finite and by Theorem 10, it suffices to define compatible functions to endow a structure of $\diamond$-semigroup. Define the function τ by :

$$m^{\tau}_{q,q'} = \{t \cup \{q'\} \mid \exists t' \subset t,\, \exists p \in Q,\, t \in m^{\pi}_{q,p},\, t' \in m^{\pi}_{p,p} \text{ and } t' \longrightarrow q' \in E\}$$

where π is the smallest integer such that m^{π} is an idempotent matrix. The function $-\tau$ is defined symmetrically and it can be proved that τ and $-\tau$ are functions respectively compatible on the right and left with S. It remains to define a morphism $\varphi : A^{\diamond} \longrightarrow S$ recognizing X. For each letter a of A, we define the matrix $m_a = \varphi(a)$ corresponding to the edges of $\mathcal{A}$ labelled by a: The entry (q, q') of m_a is equal to $\{\{q, q'\}\}$ if $q \xrightarrow{a} q' \in E$ or $\emptyset$ otherwise. An induction on the rank would show that for all word $x \in A^{\diamond}$, $\varphi(x) = m$ where the matrix m memorizes the contents of paths labelled by x:

$$m_{q,q'} = \{l \mid q \underset{l}{\overset{x}{\Longrightarrow}} q'\}$$

A word $x \in A^{\diamond}$ belongs to X iff $\varphi(x)$ has a (i, f)non-empty entry where i and f are respectively initial and final states. Thus X is recognized by S.

6 From $\diamond$-semigroups to automata

In this section, we prove that a recognizable set is rational. The proof is by induction on the structure of the $\diamond$-semigroup (S, π). The problem is first reduced in Lemma 19. After two technical lemmas 20 and 21, an automaton accepting words over S whose product belongs to a given $\mathcal{D}$-class is defined. Lemmas 22 to 26 are devoted to the proof that the defined automaton is sound. Finally, the result is proved in Proposition 27.

Let (S, π) be a finite $\diamond$-semigroup. By Theorem 10, the product π is defined by compatible functions τ and $-\tau$. Let X be a subset of $A^{\diamond}$ recognized by S. There exists a morphism of $\diamond$-semigroup $\varphi : A^{\diamond} \longrightarrow S$ and a subset P of S such that $X = \varphi^{-1}(P)$. Since rational sets are closed under finite union, one may suppose that P is a single element $\{p\}$. The following Lemma shows that it suffices to prove that for any element s of S, the set $\pi^{-1}(s)$ of words over S whose product is s, is rational.

Lemma 19. *Let (S, π) be a finite $\diamond$-semigroup. Let $\varphi : A^{\diamond} \longrightarrow S$ be a morphism of $\diamond$-semigroup and let $p \in S$. If $\pi^{-1}(p) \in Rat(S^{\diamond})$ then $\varphi^{-1}(p) \in Rat(A^{\diamond})$.*

Proof. Let h be the finite substitution which associates to each element s of S the set $\varphi^{-1}(s) \cap A$. We prove that the following equality holds:

$$\varphi^{-1}(p) = h(\pi^{-1}(p) \cap \varphi(A)^{\diamond})$$

Let $x \in \varphi^{-1}(p)$. Denote by $x = \prod_{j\in J} a_j$ where $J \in \mathcal{S}$, $a_j \in A$ for any $j \in J$ and $\varphi(x) = \pi(\prod_{j\in J} \varphi(a_j)) = p$. We have

$$x \in \prod_{j\in J} (\varphi^{-1}(\varphi(a_j)) \cap A) \subseteq \prod_{j\in J} h(\varphi(a_j)) = h(\prod_{j\in J} \varphi(a_j)) \subseteq h(\pi^{-1}(p) \cap \varphi(A)^{\diamond})$$

Conversely, let $x \in h(\pi^{-1}(p) \cap \varphi(A)^{\diamond})$. There exists $u \in \pi^{-1}(p) \cap \varphi(A)^{\diamond}$ such that $x \in h(u)$. Denote by $u = \prod_{j\in J} s_j$ where $J \in \mathcal{S}$, $s_j \in S$ for any $j \in J$. Since $x = \prod_{j\in J} h(s_j)$,

$$\varphi(x) = \pi(\prod_{j\in J} \varphi(h(s_j))) = \pi(\prod_{j\in J} \varphi(\varphi^{-1}(s_i) \cap A)) = \pi(\prod_{j\in J} s_j) = \pi(u) = p.$$

For any element s of S, it suffices to construct an automaton on linear orderings accepting the set $\pi^{-1}(p)$. In the case of finite words, when S is a semigroup, it is natural to define the automaton having S^1 as set of states and $\{s \xrightarrow{t} st \mid s \in S^1, t \in S\}$ as set of transitions. In the case of words on linear orderings, we also need to define limit transitions. The difficulty comes from the fact that two words u and v can be labels of paths having the same cofinal set of states even if $\pi(u)$ is not egal to $\pi(v)$. Thus, it is necessary to define conditions on limit transitions in order to guarantee that the labels of paths admitting the same cofinal set of states P admits ramseyan factorizations associated to conjugated linked pairs. In this case, $\pi(u) = \pi(v)$:

Lemma 20. *Let S be a finite $\diamond$-semigroup, let τ be a function compatible to the right with S and let $(s, e) \in S \times E(S)$ and $(t, f) \in S \times E(S)$ be two right linked pairs. If the two pairs are conjugated, then $se^{\tau} = tf^{\tau}$. Symmetrically, if τ is a function compatible on the left with S and if the pairs $(e, s) \in E(S) \times S$ and $(f, t) \in E(S)\times$ are conjugated to the left, then $e^{-\tau}s = f^{-\tau}t$.*

Proof. Let $(s, e) \in S \times E(S)$ and $(t, f) \in S \times E(S)$ be two right linked pairs. If the pairs are conjugated, there exists $a, b \in S^1$ such that $e = ab$, $f = ba$, $sa = t$ and $tb = s$. Since τ is compatible to the right, we get

$$se^{\tau} = s(ab)^{\tau} = (sa)(ba)^{\tau} = tf^{\tau}$$

The case of left linked pairs is symmetrical.

The construction of the automaton is based on an induction on the $\mathcal{D}$-class structure of S. It uses the following property of linked pairs of a same $\mathcal{D}$-class.

Lemma 21. *[[10], Lemma 65] Let D be a $\mathcal{D}$-class of a finite semigroup S and let (s_1, e_1) and (s_2, e_2) be two right linked pairs (respectively left linked pairs) such that s_1, s_2, e_1, $e_2 \in D$. The linked pairs are conjugated if and only if $s_1 \mathcal{R} s_2$ (respectively $s_1 \mathcal{L} s_2$).*

For any $\mathcal{D}$-class D of S, denote by:

$$S_D = \{s \in S \mid \forall p \in D, s \geq_{\mathcal{J}} p\} \text{ and } T_D = \{s \in S \mid \forall p \in D, s >_{\mathcal{J}} p\}$$

We define an automaton on linear orderings accepting words over S_D and computing the product π of its paths labels in both directions.

Let $\mathcal{A}_D = (Q_D, S_D, E_D)$ be the automaton defined by:

$$\begin{array}{rcl} Q_D & = & S_D^1 \times S_D^1 \times \mathbb{B} \text{ is the set of states where } \mathbb{B} = \{0,1\} \\ E_D & = & \{(s, rt, b) \xrightarrow{r} (sr, t, b') \mid b \in \mathbb{B}, b' = (r \in D)\} \\ & \cup & \{\{(s_i, t_i, b_i)\}_{1 \leq i \leq m} \longrightarrow (s, t, b) \mid b \in \mathbb{B}, \exists 1 \leq i \leq m\ ,\ b_i = 1\ , \\ & & \exists 1 \leq k \leq m\ ,\ \exists e \in E(D)\ ,\ s_k e = s_k\ ,\ e t_k = t_k\ ,\ s = s_k e^{\tau} \text{ and } t_k = e^{\tau} t\} \\ & \cup & \{(s, t, b) \longrightarrow \{(s_i, t_i, b_i)\}_{1 \leq i \leq m} \mid b \in \mathbb{B}, \exists 1 \leq i \leq m\ ,\ b_i = 1\ , \\ & & \exists 1 \leq k \leq m\ ,\ \exists e \in E(D)\ ,\ s_k e = s_k\ ,\ e t_k = t_k\ ,\ t = e^{-\tau} t_k \text{ and } s_k = s e^{-\tau}\} \end{array}$$

The boolean component of Q_D allows limit transitions only if the label of the path admits a ramseyan factorization associated to an idempotent of D. Using Lemma 21, we prove that the automaton $\mathcal{A}_D$ computes properly the product π.

Lemma 22. *For any word u of countable scattered linear length over S and for any states $q = (s, t, b)$ and $q' = (s', t', b')$ of Q_D, if there exists a path in $\mathcal{A}_{\mathcal{D}}$ labelled u leading from q to q' then $s' = s\pi(u)$ and $t = \pi(u)t'$.*

Proof. We prove by induction on $\alpha \in \mathcal{O}$ that the Lemma holds for any word u of rank α.

When $\alpha = 0$, the result follows from the definition of $\mathcal{A}_D$.

Let $\alpha > 0$. Suppose the existence of a path γ in $\mathcal{A}_{\mathcal{D}}$ leading from a state $q = (s, t, b)$ to a state $q' = (s', t', b')$ of label $u \in S^{W_\alpha}$. Denote by $I = |u|$ the length of u.

Suppose that $I \in U_\alpha$. The linear ordering I can be written as a sum $I = \sum_{j \in J} K_j$ where for any $j \in J$, K_j is of rank strictly lower than α: $K_j \in \bigcup_{\beta < \alpha} W_\beta$ and where $J \in \mathcal{N} \cup \{\omega, -\omega\}$.

In the case where $J \in \mathcal{N}$, we use the inductive hypothesis since I is of rank strictly lower than α.

Suppose that $J = \omega$ (the case $J = -\omega$ is symmetrical). There exists an ω-factorization $u = \prod_{j \in \omega} u_j$ and a right linked pair (r, e) such that $\pi(u_0) = r$ and $\pi(u_j) = e$ for any $j > 0$. Each factor u_j is of rank strictly lower than α. By definition of the product π, $\pi(u) = re^{\tau}$. We want to prove that $s' = sre^{\tau}$ and $t = re^{\tau}t'$. The path γ of label u ends with a left limit transition $P \longrightarrow q'$. The cofinal set of states $P = cof_{(I,\emptyset)^-}\gamma$ is denoted by $P = \{(s_1, t_1, b_1), \ldots, (s_m, t_m, b_m)\}$ and the path γ is represented by the following way:

$$\gamma : q = q_0 \overset{u_0}{\Longrightarrow} q_1 \overset{u_1}{\Longrightarrow} q_2 \overset{u_2}{\Longrightarrow} q_3 \ldots\ P \longrightarrow q'$$

By inductive hypothesis, for any $j > 0$, the first component of the state q_j, denoted by s_{q_j}, satisfies $s_{q_j} = sr$, $s_{q_j}e = s_{q_j}$ and the second component t_{q_j} verifies $t = rt_{q_j}$ and $et_{q_j} = t_{q_j}$. Thus, there exists a state of P satisfying those equalities. Let $1 \leq k \leq m$ such that the state $p_k \in P$ verifies $s_k = sr$, $s_k e = s_k$, $t = rt_k$ and $et_k = t_k$. Since $P \longrightarrow q' \in E_D$, there exists at least one state of P whose boolean component is equal to 1. By construction, it means that $e \in D$: for any element $i \in I$, there exists $i' > i$ such that the letter of u indexed by i' belongs D. Up to a factorization, one can suppose that for any $j > 0$, the factor u_j contains a letter $d_j \in D$. Then we get $\pi(u_j) = e \leq_J d_j$. Moreover, $e \geq_J sre$ and by the inductive hypothesis $sre = s_{q_j} \in S_D$ thus $e \in D$.

On the other hand, since $P \longrightarrow q'$ belongs to the set of transitions E_D, there exists an idempotent $f \in E(D)$ and $1 \leq l \leq m$ such that the state $p_l \in P$ verifies $s_l = s_l f$, $t_l = ft_l$, $s' = s_l f^\tau$ and $t_l = f^\tau t'$.

We first show that $s' = s\pi(u)$. Since $e \in D$, $f \in D$ and that the pairs (s_k, e) and (s_l, f) are right linked, we know that $s_k \in D$ and $s_l \in D$. Since $p_k \in P$ and $p_l \in P$, we get that $s_k \mathcal{R} s_l$ by construction. From Lemma 21, those pairs are conjugated. Finally, Lemma 20 gives $s_l f^\tau = s_k e^\tau$ i.e. $s' = sre^\tau = s\pi(u)$.

Symmetrically, (e, t_k) and (f, t_l) are left linked pairs of D and $t_k \mathcal{L} t_l$. From lemma 21, those pairs are conjugated on the left thus there exists $a, b \in S$ such that $e = ab$, $f = ba$, $t_l = bt_k$ and $t_k = at_l$. Then, we get $t = rt_k = rat_l = raf^\tau t' = ra(ba)^\tau t' = re^\tau t' = \pi(u)t'$ which concludes the case where $|u| \in U_\alpha$.

Now suppose that $I \in W_\alpha$. The length of u is a finite sum $I = \sum_{j=0}^{n} K_j$ where for all $0 \leq j \leq n$, $K_j \in U_\alpha$. Let $u = \prod_{j=0}^{n} u_j$ be the associated factorization. The path γ is denoted by $q = q_0 \stackrel{u_0}{\Longrightarrow} q_1 \stackrel{u_1}{\Longrightarrow} q_2 \ldots \stackrel{u_n}{\Longrightarrow} q_{n+1} = q'$. From the preceding case, for all $0 \leq j \leq n$, $s_{q_j}\pi(u_j) = s_{q_{j+1}}$ and $t_{q_j} = \pi(u_j)t_{q_{j+1}}$. Thus $s\pi(u) = s'$ and $t = \pi(u)t'$.

For any $s \in S_D$, denote by $\mathcal{A}_s$ the automaton $\mathcal{A}_D$ with initial state $\{(1, s, 0)\}$ and final set of states $\{(s, 1, b) \mid b \in \mathbb{B}\}$. We also denote by L_s the set of words accepted by $\mathcal{A}_s$. The preceding Lemma shows that for any $s \in S_D$, $L_s \subseteq \pi^{-1}(s)$. This inclusion is strict. Because of the restrictions on limit transitions, a word $u \in S_D^\diamond$ whose product $\pi(u)$ belongs to D is not always the label of a path in $\mathcal{A}_D$. In order to describe the set L_s of accepted words, we first give some properties on the paths of $\mathcal{A}_D$.

Lemma 23. *If a word $u \in S^\diamond$ is the label of a path $\gamma = ((s_c, t_c, b_c))_{c \in |\hat{u}|}$ in $\mathcal{A}_D$, then for any elements s and t of S_D, the path $((ss_c, t_c t, b_c))_{c \in |\hat{u}|}$ is also a path in $\mathcal{A}_D$ of label u.*

Proof. The Lemma is proved for any word $u \in A^\diamond$ by induction on the rank $\alpha \in \mathcal{O}$ of u.

If u is a finite word ($\alpha = 0$), the result follows from the definition of $\mathcal{A}_D$.

Let $\alpha > 0$, $I \in W_\alpha$ and let $u \in S^I$. Suppose the existence of a path $\gamma = ((s_c, t_c, b_c))_{c \in \hat{I}}$ in $\mathcal{A}_\mathcal{D}$ of label u and let $s, t \in S_D$.

We first suppose that $I \in U_\alpha$. The ordering I can be decomposed as a sum $I = \sum_{j \in J} K_j$ where for any $j \in J$, K_j is of rank strictly lower than α and where $I \in \mathcal{N} \cup \{\omega, -\omega\}$.

When $J \in \mathcal{N}$, the result follows from the inductive hypothesis.

Suppose that $J = \omega$ (the case $J = -\omega$ is treated symmetrically). Let $u = \prod_{j \in \omega} u_j$ be the factorization such that $|u_j| = K_j$ for any $j \in \omega$. We denote by c_0 the first cut $(\emptyset, I)$ and for all $j > 0$, we denote by $c_j = (\bigcup_{0 \leq i < j} K_i, \bigcup_{j \leq i} K_i)$ the cut of $\hat{I}$ corresponding to the factorization $u = (u_0 \ldots u_{j-1})(u_j \ldots)$. By the inductive hypothesis, for any $j \in \omega$, $((ss_c, t_c t, b_c))_{c_j \leq c \leq c_{j+1}}$ is a path in $\mathcal{A}_D$ of label u_j. It remains to show that the path $\gamma' = ((ss_c, t_c t, b_c))_{c \in \hat{I} \setminus \{(I, \emptyset)\}}$ is followed by a limit transition. Let $P = cof_{(I,\emptyset)^-} \gamma$ and $P' = cof_{(I,\emptyset)^-} \gamma'$. By construction, if $P = \{(s_i, t_i, b_i)\}_{1 \leq i \leq m}$ then $P' = \{(ss_i, t_i t, b_i)\}_{1 \leq i \leq m}$. Since $P \longrightarrow (s_{(J,\emptyset)}, t_{(J,\emptyset)}, b_{(J,\emptyset)}) \in E_D$, there exists $1 \leq i \leq m$ such that $b_i = 1$. There exists also $1 \leq k \leq m$ and $e \in E(D)$ such that $s_k e = s_k$, $e t_k = t_k$, $s_{(J,\emptyset)} = s_k e^\tau$ and $t_k = e^\tau t_{(J,\emptyset)}$. Then we have $ss_k e = ss_k$, $e t_k t = t_k t$, $ss_{(J,\emptyset)} = ss_k e^\tau$ and $t_k t = e^\tau t_{(J,\emptyset)} t$ which proves that $P' \longrightarrow (ss_{(J,\emptyset)}, t_{(J,\emptyset)} t, b_{(J,\emptyset)}) \in E_D$.

Now suppose that $I \in W_\alpha$. The ordering I is a finite sum $I = \sum_{j=0}^{n} K_j$ where for all $0 \leq j \leq n$, $K_j \in U_\alpha$. Keeping the same notations and using the preceding case, $((ss_c, t_c t, b_c))_{c_j \leq c \leq c_{j+1}}$ is a path in $\mathcal{A}_D$ of label u_j for any $0 \leq j \leq n$.

Note also that, by definition of $\mathcal{A}_D$, the boolean component of the first state of a path is not important.

Lemma 24. *If a word $u \in S^\diamond$ of length I is the label of a path $\gamma = ((s_c, t_c, b_c))_{c \in \hat{I}}$ in $\mathcal{A}_D$, then, for any boolean $b \in \mathbb{B}$, the path*

$$(s_{(\emptyset,I)}, t_{(\emptyset,I)}, b)((s_c, t_c, b_c))_{c \in \hat{I} \setminus \{(\emptyset,I)}$$

is also a path in $\mathcal{A}_D$ of label u.

Using Lemmas 23 and 24, it is possible to concatenate the paths of $\mathcal{A}_D$, a finite number of times first.

Lemma 25. *$\forall s \in S_D$, $\forall t \in S_D$, $L_s L_t \subseteq L_{st}$.*

Proof. Let $s, t \in S_J$. Suppose the existence of paths

$$(1, s, 0) \stackrel{u}{\Longrightarrow} (s, 1, b) \text{ and } (1, t, 0) \stackrel{v}{\Longrightarrow} (t, 1, b')$$

of labels $u \in L_s$ and $v \in L_t$. From Lemmas 23 and 24,

$$(1, st, 0) \stackrel{u}{\Longrightarrow} (s, t, b) \text{ and } (s, t, b) \stackrel{v}{\Longrightarrow} (st, 1, b')$$

are also paths of $\mathcal{A}_D$ of labels u and v. Thus $(1, st, 0) \stackrel{uv}{\Longrightarrow} (st, 1, b')$ i.e. $uv \in L_{st}$.

Suppose that D contains an idempotent e and let $u \in L_e^\omega$. Using the preceding Lemma, one can construct a path of left limit length whose label coincides with u. If the sequence of letters of u belonging to D is cofinal, then the cofinal set of states contains a positive boolean component and the path ends with a limit transition. Otherwise, the limit transition is forbidden since all boolean components of the cofinal set are equal to 0. In this case, the word u admits a suffix over the alphabet T_D. Let E_d be the set of finite words over the alphabet T_D of product d:

$$E_d = \{s_1 s_2 \dots s_m \in T_D^+ \mid s_1 \dots s_m = d\}$$

Lemma 26. *Let $e \in E(D)$ such that $e^\tau \in D$ (respectively $e^{-\tau} \in D$). Then ${L_e}^\omega \subseteq L_{e^\tau} \cup L_e {E_e}^\omega$ (respectively ${L_e}^{-\omega} \subseteq L_{e^{-\tau}} \cup {E_e}^{-\omega} L_e$).*

Proof. Let $e \in E(D)$ such that $e^\tau \in D$ and let $u \in {L_e}^\omega$ a word of length $I \in \mathcal{S}$. Let $u = \prod_{j \in \omega} u_j$ be the factorization such that for all $j \in \omega$, $u_j \in L_e$. For any $j \in \omega$, there exists a path γ_j of label u_j:

$$\gamma_j : (1, e, 0) \overset{u_j}{\Longrightarrow} (e, 1, b_j).$$

From Lemmas 23 and 24, the paths

$$\gamma'_0 : (1, e^\tau, 0) \overset{u_0}{\Longrightarrow} (e, e^\tau, b_0) \text{ and } \gamma'_j : (e, e^\tau, b_{j-1}) \overset{u_j}{\Longrightarrow} (e, e^\tau, b_j)$$

are also labelled u_j for all $j > 0$. Thus, we have constructed a path γ' of left limit length whose label coincides with u. Let $P = cof_{(I,\emptyset)^-}\gamma'$. The set $P = \{(s_i, t_i, b_i)\}_{1 \le i \le m}$ contains at least one state of the form (e, e^τ, b) for some boolean b. Thus, there exists $1 \le k \le m$ such that $s_k = e$, $t_k = e^\tau$, $s_k e = e = s_k$ and $e t_k = e^\tau = t_k$. Two cases have to be considered depending whether P contains at least one boolean component equals to 1 or not:

- There exists $1 \le l \le m$ such that $b_l = 1$. Then $P \longrightarrow (e^\tau, 1, 0) \in E_D$ and $u \in L_{e^\tau}$.
- For any $1 \le l \le m$, $b_l = 0$. Denote by $u = \prod_{i \in I} a_i$ with $a_i \in S$ for all $i \in I$. By construction, there exists $i_0 \in I$ such that for all $i > i_0$, $a_j \in T_D$. Thus, $u \in L_e(L_e \cap T_D^\diamond)^\omega \subseteq L_e(L_e \cap T_D^+)^\omega \subseteq L_e {E_e}^\omega$.

The inclusion ${L_e}^{-\omega} \subseteq L_{e^{-\tau}} \cup {E_e}^{-\omega} L_e$ is symmetrical.

For every $d \in D$, the language L_d accepted by $\mathcal{A}_d$ is strictly included in the set $\pi^{-1}(d)$. The preceding Lemma shows that $\pi^{-1}(d)$ is not included in L_d : the words over the alphabet T_D which are not finite are not labels of any path in $\mathcal{A}_D$. We solve this problem using an induction on the $\mathcal{D}$-class structure of the $\diamond$-semigroup S. A word of $\pi^{-1}(d)$ can be obtained by a finite product of words of $\mathcal{D}$-classes strictly $\mathcal{J}$-greater than D. For every $d \in D$, we define

$$F_d = \bigcup_{\substack{s_1, \dots, s_m >_{\mathcal{J}} d, \\ s_1 \dots s_m = d}} \pi^{-1}(s_1) \dots \pi^{-1}(s_m).$$

If there exists an idempotent e strictly $\mathcal{J}$-greater than D such that e^τ or $e^{-\tau}$ falls in D, the words of $\pi^{-1}(d)$ can also be obtained from an ω-product or from an $-\omega$-product. For every $d \in D$, we define

$$G_d = \bigcup_{\substack{t,e >_\mathcal{J} d,\\ te^\tau = d}} \pi^{-1}(t)\pi^{-1}(e)^\omega \cup \bigcup_{\substack{t,e >_\mathcal{J} d,\\ e^{-\tau} t = d}} \pi^{-1}(e)^{-\omega}\pi^{-1}(t).$$

For all $p \in D$, we substitute in L_p any letter $d \in D$ with the sets F_d et G_d. Then we obtain a rational expression of $\pi^{-1}(p)$.

Proposition 27. *Let (S, π) be a finite $\diamond$-semigroup. For any $p \in S$, the set $\pi^{-1}(p)$ is rational.*

Proof. Let $p \in S$. We prove that $\pi^{-1}(p)$ is rational by induction on the $\mathcal{D}$-class D of p.

Suppose that D is $\mathcal{J}$-maximal i.e. $\forall s \in S$, if $p \leq_\mathcal{J} s$ then $s \in D$. In this case, we prove that $\pi^{-1}(p) = L_p$. From Lemma 22, $L_p \subseteq \pi^{-1}(p)$. Conversely, we prove that for any $u \in \pi^{-1}(p)$, $u \in L_p$ by induction on the rank of u. By definition of the automaton $\mathcal{A}_D$, the inclusion is true for all finite words. In the other hand, when D is $\mathcal{J}$-maximal, the set E_d is empty for all $d \in D$. If D contains an idempotent e such that $e^\tau \in D$ (respectively $e^{-\tau} \in D$), Lemma 26 gives ${L_e}^\omega \subseteq L_{e^\tau}$ (respectively ${L_e}^{-\omega} \subseteq L_{e^{-\tau}}$). It follows that $L_p = \pi^{-1}(p)$ which concludes the basic case of the induction on the $\mathcal{D}$-class.

By the inductive hypothesis, for every $s \in T_D$, $\pi^{-1}(s)$ is rational. Note that for any $d \in D$, F_d is also rational since the set E_d of finite words over T_D of product d is accepted by the finite automaton $({T_D}^1, T_D, \{s \xrightarrow{r} sr\})$. The set G_d is rational too as a finite union of rational sets. We define the rational substitution f by

$$\begin{array}{rccl} f: & S_D & \longrightarrow & Rat({S_D}^\diamond) \\ & s & \longrightarrow & \begin{cases} \pi^{-1}(s) & \text{if } s \in T_D \\ \{s\} \cup F_s \cup G_s & \text{if } s \in D \end{cases} \end{array}$$

and we prove that $f(L_p) = \pi^{-1}(p)$.

By Lemma 22, $L_p \subseteq \pi^{-1}(p)$. By definition, f preserves the product : for any $u \in {S_D}^\diamond$, $\pi(f(u)) = \pi(u)$. Thus $f(L_p) \subseteq \pi^{-1}(p)$. Conversely, we prove that for any $u \in \pi^{-1}(p)$, $u \in f(L_p)$ by induction on the rank of u. Let u be a finite word such that $\pi(u) = p$. By definition of f, $u \in f(u)$ and by definition of $\mathcal{A}_D$, $u \in L_p$ thus $u \in f(L_p)$. This concludes the case of rank 0.

Let $\alpha > 0$ and let $u \in S^{W_\alpha} \cap \pi^{-1}(p)$. Let $I = |u|$ be the length of u.

Suppose first that $I \in U_\alpha$. The linear ordering I can be decomposed as a sum $I = \sum_{J \in J} K_J$ such that for all $j \in J$, $K_j \in \bigcup_{\beta<\alpha} W_\beta$ and $J \in \mathcal{N} \cup \{\omega, -\omega\}$. In the case where $J \in \mathcal{N}$, $I \in \bigcup_{\beta<\alpha} W_\beta$ and the result is obtained by inductive hypothesis. Suppose that $J = \omega$. There exists a right linked pair $(s, e) \in S_D \times E(S_D)$ and a factorization $u = \prod_{j\in\omega} u_j$ such that $\pi(u_0) = s$ and for all

$j > 0$, $\pi(u_j) = e$ with $|u_j| \in \bigcup_{\beta<\alpha} W_\beta$ for every $j \in \omega$. Note that $s \leq_{\mathcal{J}} e$ since (s, e) is a right linked pair and that $e^\tau \in S_D$ since $\pi(u) \in D$. The following cases are distinguished:

- $s \in T_D$ and $e \in T_D$:
  then $u \in \pi^{-1}(s)\pi^{-1}(e)^\omega \subseteq G_p \subseteq f(p) \subseteq f(L_p)$.
- $s \in D$ and $e \in T_D$:
  by the inductive hypothesis on the rank, $u \in f(L_s)\pi^{-1}(e)^\omega$. Let $t = e^\tau$.
  - If $t \in T_D$, then $u \in f(L_s)\pi^{-1}(t)$.
  - If $t \in D$, then $u \in f(L_s)G_t$.

  In both cases $u \in f(L_s)f(t) \subseteq f(L_s)f(L_t) \subseteq f(L_p)$.
- $s \in D$ and $e \in D$:
  By the inductive hypothesis on the rank, $u \in f(L_s)f(L_e)^\omega = f(L_s L_e^\omega)$. By Lemma 26, $L_s L_e^\omega \subseteq L_s L_{e^\tau} \cup L_s E_e^\omega$.
  - If $u \in f(L_s L_{e^\tau})$ then $u \in f(L_p)$.
  - Otherwise $u \in f(L_s)f(E_e)^\omega = f(L_s)F_e^\omega \subseteq f(L_s)f(e)^\omega = f(L_s)f(e^\omega) \subseteq f(L_s)f(L_{e^\tau}) \subseteq f(L_p)$.

The case $J = -\omega$ is symmetrical which concludes the case where $I \in U_\alpha$.

Suppose that $I \in W_\alpha$. There exists a finite factorization $u = \prod_{j=1}^{n} u_j$ such that for all $1 \leq j \leq n$, $|u_j| \in U_\alpha$ thus $u_j \in f(L_{s_j})$ where $s_j = \pi(u_j)$. Thus $u \in \prod_{j=1}^{n} f(L_{s_j}) = f(L_{s_1} \ldots L_{s_n}) \underset{Lemma\ 26}{\subseteq} f(L_p)$.

The construction is illustrated by the following example.

Example 28. Let $S = \{0, a, b, s, e, f\}$ be the finite $\diamond$-semigroup whose product π is defined by the finite product

$$aa = a \ , \ ee = sb = e \ , \ ff = bs = fa = af = f,$$

$$es = sf = sa = s \ , \ be = fb = ab = b,$$

the others finite products being equal to 0. The compatible functions τ and $-\tau$ are defined by

$$a^\tau = a \ , \ a^{-\tau} = s \ , \ e^\tau = e \ , \ f^\tau = b,$$

the others values being equal to 0. Let $\varphi : A^\diamond \to S$ be the morphism defined by $\varphi(a) = a$ and $\varphi(b) = b$. We look for a rational expression of the set $\varphi^{-1}(e)$. The $\mathcal{D}$-class structure of the $\diamond$-semigroup S is represented on the following way:

a^*

s	e^*
f^*	b

0^*

The automaton $\mathcal{A}_a$ of the $\mathcal{J}$-maximal $\mathcal{D}$-class of S is the following:

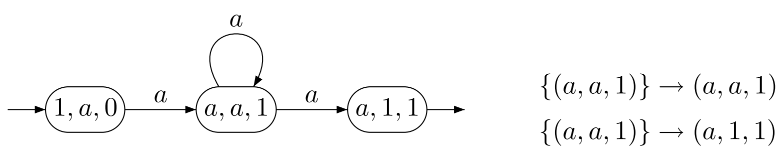

Figure 2: Automaton $\mathcal{A}_a$ accepting the set $L_a = a^{\#}$.

The automaton $\mathcal{A}_e$ is given in Figure 3.

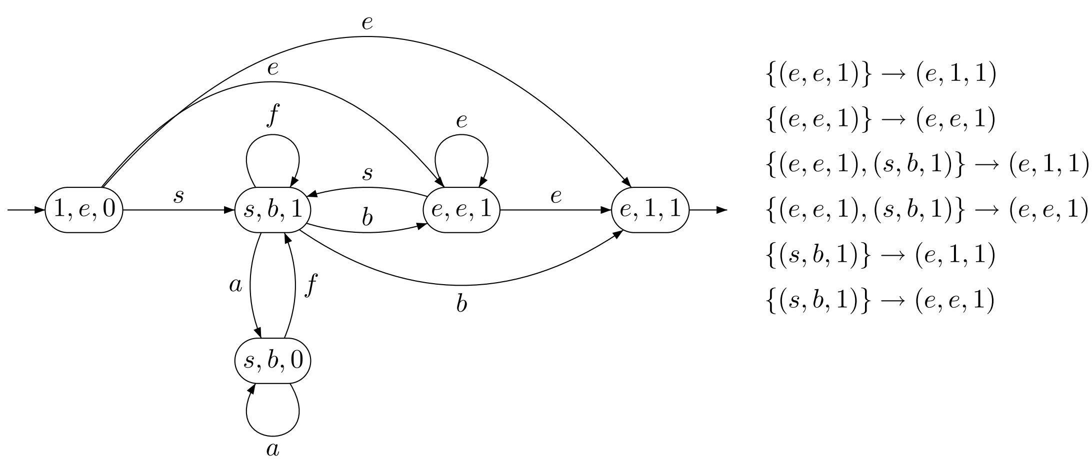

Figure 3: Automaton $\mathcal{A}_e$ accepting the set $L_e = (e + s(a^* f)^* b)^{\#}$.

We look for a rational expression of the set $\pi^{-1}(e)$. For any $d \in D = \{s, e, b, f\}$ the sets E_d and F_d are empty. The only non-empty set G_d is $G_s = (\pi^{-1}(a))^{-\omega} = (a^{\#})^{-\omega}$. We deduce that $\pi^{-1}(e) = f(L_e) = (e+(s+(a^{\#})^{-\omega})(a^{\#} f)^* b)^{\#}$. Using Lemma 19, we get $\varphi^{-1}(e) = ((a^{\#})^{-\omega} b)^{\#}$.

We have proved that any recognizable language of $A^\diamond$ is rational. The converse, given by the preceding section, concludes the proof of Theorem 14.

7 Syntactic $\diamond$-semigroup

Let X be a recognizable subset of $A^\diamond$. Among all $\diamond$-semigroups recognizing X, there exists one which is minimal in the sense of division. It is called the syntactic $\diamond$-semigroup of X and is the first canonical object associated to rational sets on linear orderings. Because of a lack of space, the proofs of this section are omitted but we refer the reader to [2] whose proofs are very similar. For any $\diamond$-semigroup (S, π) and any set $P \subseteq S$, the equivalence relation $\sim_P$ is defined for any s,t in S by $s \sim_P t$ iff for any integer m:

$$\forall s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_m \in S^1 \ , \forall \theta_1, \theta_2, \ldots, \theta_{m-1} \in \{\omega, -\omega\} \cup \mathcal{N},$$

$$\pi(s_m(\ldots(s_2(s_1 s t_1)^{\theta_1} t_2)^{\theta_2} \ldots)^{\theta_{m-1}} t_m) \in P$$
$$\iff \pi(s_m(\ldots(s_2(s_1 t t_1)^{\theta_1} t_2)^{\theta_2} \ldots)^{\theta_{m-1}} t_m) \in P$$

The equivalence relation $\sim_P$ is a congruence of $\diamond$-semigroup. If S finite, then and the quotient $S/{\sim_P}$ is an effective $\diamond$-semigroup. If X is a recognizable subset of $A^\diamond$, then the quotient $A^\diamond/{\sim_X}$ is finite and recognizes X.

Proposition 29. *Let X be a rational subset of $A^\diamond$. The relation $\sim_X$ is a congruence of $\diamond$-semigroup of finite index.*

Proof. The proof is similar to Lemmas 37 and 39 of [2] for ordinal words. If $X = \varphi^{-1}(P)$, where $\varphi : A^\diamond \to S$ is a morphism of $\diamond$-semigroup, it can be proved that for any words $u, v \in A^\diamond$, $u \sim_X v$ if and only if $\varphi(u) \sim_P \varphi(v)$. Note that when S is a finite $\diamond$-semigroup, the integer m in the definition of $\sim_P$ is bounded.

For any recognizable subset X of $A^\diamond$, the $\diamond$-semigroup $A^\diamond/{\sim_X}$ is called the *syntactic semigroup* of X and is denoted by $S(X)$. It is the smallest $\diamond$-semigroup recognizing X in the sense of division.

Proposition 30. *Let X be a recognizable set of $A^\diamond$ and let T be a $\diamond$-semigroup. Then T recognizes X if and only if $S(X)$ divides T.*

Proof. The proof is the same of the cases of finite words ([20] Corollary 8.10) and ordinal words ([4] Theorem 43). In particular, for any recognizable set X, the relation $\sim_X$ is the coarsest congruence such that the quotient $A^\diamond/{\sim_X}$ recognizes X. From Theorem 14 and Proposition 30, it follows that the syntactic $\diamond$-semigroup of a rational set is finite.

Theorem 31. *The syntactic $\diamond$-semigroup of a rational set of words indexed by countable scattered linear orderings is finite.*

References

[1] N. Bedon. Finite automata and ordinals. *Theoret. Comput. Sci.*, 156:119–144, 1996.

[2] N. Bedon. Automata, semigroups and recognizability of words on ordinals. *Int. J. Alg. Comput.*, 8(1):1–21, 1998.

[3] N. Bedon. Star-free sets of words on ordinals. *Inform. Comput.*, 166:93–111, 2001.

[4] N. Bedon and O. Carton. An Eilenberg theorem for words on countable ordinals. In Cláudio L. Lucchesi and Arnaldo V. Moura, editors, *Latin'98: Theoretical Informatics*, volume 1380 of *Lect. Notes in Comput. Sci.*, pages 53–64. Springer-Verlag, 1998.

[5] V. Bruyère and O. Carton. Automata on linear orderings. In Jiří Sgall, Aleš Pultr, and Petr Kolman, editors, *MFCS'2001*, volume 2136 of *Lect. Notes in Comput. Sci.*, pages 236–247, 2001.

[6] V. Bruyère, O. Carton, and G. Sénizergues. Tree automata and automata on linear orderings. In Tero Harju and Juhani Karhumäki, editors, *WORDS'2003*, pages 222–231. Turku Center for Computer Science, 2003.

[7] J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik und grundl. Math.*, 6:66–92, 1960.

[8] J. R. Büchi. On a decision method in the restricted second-order arithmetic. In *Proc. Int. Congress Logic, Methodology and Philosophy of Science, Berkeley 1960*, pages 1–11. Stanford University Press, 1962.

[9] J. R. Büchi. Decision methods in the theory of ordinals. *Bull. Am. Math. Soc.*, 71:767–770, 1965.

[10] O. Carton and M. Michel. Unambiguous Büchi automata. *Theoret. Comput. Sci.*, 297:37–81, 2003. IGM report 2001–06.

[11] O. Carton and C. Rispal. Complementation of rational sets on scattered linear orderings of finite rank. In *LATIN'2004*, volume 2976 of *Lect. Notes in Comput. Sci.*, pages 292–301, 2004.

[12] D. Girault-Beauquier. Bilimites de langages reconnaissables. *Theoret. Comput. Sci.*, 33(2–3):335–342, 1984.

[13] F. Hausdorff. *Set Theory.* Chelsea, New York, 1957.

[14] S. C. Kleene. Representation of events in nerve nets and finite automata. In C.E. Shannon, editor, *Automata studies*, pages 3–41. Princeton university Press, Princeton, 1956.

[15] R. McNaughton. Testing and generating infinite sequences by a finite automaton. *Inform. Control*, 9:521–530, 1966.

[16] R. McNaughton and S. Papert. *Counter free automata.* MIT Press, Cambridge, MA, 1971.

[17] D. Muller. Infinite sequences and finite machines. In Proc. of Fourth Annual IEEE Symp., editor, *Switching Theory and Logical Design*, pages 3–16, 1963.

[18] M. Nivat and D. Perrin. Ensembles reconnaissables de mots bi-infinis. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, pages 47–59, 1982.

[19] J.-P. Pécuchet. Variétés de semigroupes et mots infinis. In B. Monien and G. Vidal-Naquet, editors, *STACS'86*, volume 210 of *Lect. Notes in Comput. Sci.*, pages 180–191, 1986.

[20] D. Perrin and J.-É. Pin. *Infinite Words.* Elsevier, 2004.

[21] M. O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.

[22] F. P. Ramsey. On a problem on formal logic. *Proc. London Math. Soc.*, 30(2):264–286, 1929.

[23] J. G. Rosenstein. *Linear Orderings.* Academic Press, 1982.

[24] M.-P. Schützenberger. On finite monoids having only trivial subgroups. *Inform. Control*, 8:190–194, 1965.