

# Complementation of Rational Sets on Scattered Linear Orderings of Finite Rank

Olivier Carton

LIAFA, Université Paris 7,

<http://www.liafa.jussieu.fr/~carton/>

Chloé Rispal

IGM, Université de Marne-la-Vallée,

<http://www-igm.univ-mlv.fr/~rispal/>

April 17, 2006

## Abstract

In a preceding paper [6], 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 show that deterministic automata do not have the same expressive power. Despite this negative result, we prove that rational sets of words of finite ranks are closed under complementation.

## 1 Introduction

Automata were first introduced by Kleene who showed that they have the same expressive power as rational expressions [13]. Since then, many extensions of this deep result have been proved. Different kinds of structures have been considered such as infinite words [8, 14], bi-infinite words [2, 15] and transfinite words [10, 11], finite and infinite trees [18], finite and infinite traces and pictures.

In [6, 7], automata that accept linearly-ordered structures were introduced. These automata are a simple and natural generalization of usual automata with additional limit transitions of the form  $P \rightarrow q$  and  $q \rightarrow P$  where  $P$  is a subset of states. They allow finite words, infinite words, bi-infinite words and transfinite words to be treated in the same framework. These automata have been shown to be equivalent to certain rational expressions when the orderings are restricted to scattered orderings [6]. Recall that scattered orderings are those orderings which do not contain a dense sub-ordering like  $\mathbb{Q}$ . Scattered orderings include the ordinals and their mirrors.

One main property of rational sets is the closure under complementation. This property means that for any automaton  $\mathcal{A}$ , there exists another automaton  $\mathcal{B}$  which accepts exactly the structures not accepted by  $\mathcal{A}$ . This property holds for almost all structures: finite and infinite words, finite and infinite trees and even for transfinite words on ordinals.

This property is important both from the practical and the theoretical point of view. The effective closure under complementation means that the class of rational sets forms an effective boolean algebra. This closure property is used whenever some logic is translated into automata. For instance, in the proof of the decidability of the monadic second-order theory of the integers by Büchi [9], and also in the proof of the decidability of the monadic second-order theory of the infinite binary tree by Rabin [18], the closure under complementation of automata is the key property. It is well known that automata have the same expressive power as the monadic second order theory on many structures such as finite, infinite, and transfinite words and trees. A nice result would be to extend this equivalence to linear orderings. Proving closure under complementation is one step towards this result.

In [7], the question of closure under complementation was left open. In this paper, we address and solve this problem for a subclass of scattered linear orderings. Namely, we prove that rational sets of words on scattered orderings of finite ranks are closed under complementation. Recall that Hausdorff's result [12] states that scattered orderings can be obtained from finite orderings by repetitive applications of  $\omega$ -sums and  $-\omega$ -sums (see Theorem 1). The rank of a scattered linear ordering is the number of nested  $\omega$ -sums and  $-\omega$ -sums needed to obtain it. The ranks of all countable scattered linear orderings range over all countable ordinals. Therefore the rank of a scattered linear ordering can be seen as a measure of its complexity. For instance,  $\omega$  and  $\zeta$  are scattered orderings of rank 1. Our result for scattered linear orderings of finite rank generalizes both the complementation of infinite and bi-infinite words. The class of scattered orderings of finite rank includes all the ordinals smaller than  $\omega^\omega$ . Therefore, our result holds for the sets of transfinite words studied by Choueka [11].

The classical method for constructing an automaton for the complement of a set of finite words accepted by an automaton  $\mathcal{A}$  is through determinization [1]. Another method uses algebraic objects such as semigroups [17]. The determinization method can still be used for infinite words but it becomes more involved [21, 3]. This method has been pushed further by Büchi for countable transfinite words but it is then very complex [10]. The algebraic method can also be extended to ordinals [4, 5]. In our case, this method can not be applied since automata can not be made deterministic. In Section 4, we give an example of a rational set of words that cannot be accepted by a deterministic automaton. Therefore, to prove the closure under complementation, we use another method which was introduced by Büchi for infinite words. This method is based on an equivalence relation on words whose classes are shown to be rational.

The paper is organized as follows. In Section 2, we introduce words indexed by linear orderings and recall the Hausdorff characterization of countable scattered linear orderings. Then rational sets of words are defined in terms of

rational operators and automata in Section 3. We finally prove in Section 4 that rational sets of words indexed by countable scattered linear orderings of finite ranks are closed under complementation.

## 2 Words on linear orderings

In this section, we recall some definitions and operations on linear orderings but we refer the reader to [20] for a complete introduction to linear orderings. We give Hausdorff's characterization of countable scattered linear orderings and introduce words indexed by linear orderings.

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  $k < j$ . A linear ordering  $J$  containing at least two elements is *dense* if for every  $j$  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  $\omega$  of natural integers and the ordering  $\zeta$  of relative integers are scattered. More generally, ordinals are scattered orderings.

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  $\omega$  is the length of right-infinite words  $a_0a_1\dots$  and  $\zeta$  is the length of bi-infinite words  $\dots a_{-1}a_0a_1\dots$ .

In order to define the rank of scattered linear orderings, we recall the following operations.

### 2.1 Operations on linear orderings

For any linear ordering  $J$ , we denote by  $-J$  the backward linear ordering that is the set  $J$  equipped with the reverse 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 original ordering  $<$  on the sets  $J$  and  $K$ , extended to a total ordering by setting  $j < k$  for every  $j \in J$  and  $k \in K$ . For instance,  $\zeta = -\omega + \omega$ . Formally, 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  $K_{j_1})$ .

We now give a definition of products of words. The sum operator for linear orderings defined above will help us define the length of a product 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  $L = \sum_{j \in J} K_j$ . We say that a product indexed by the ordering  $J$  is a *J-product*. For instance, the  $\omega$ -product of the word  $a^\omega$  is the word  $(a^\omega)^\omega$  of length  $\sum_\omega \omega$ . The sequence  $(x_j)_{j \in J}$  of words is a *J-factorization* of the word  $x = \prod_{j \in J} x_j$ . Let  $(x_i)_{i \in \omega}$  be an  $\omega$ -factorization of a

word  $x = \prod_{i \in \omega} x_i$ . A *superfactorization*  $x = (y_i)_{i \in \omega}$  is a factorization obtained from an original factorization  $x = (x_i)_{i \in \omega}$  by grouping the factors; i.e., if there exists a strictly increasing sequence  $(k_i)_{i \in \omega}$  such that  $y_0 = x_0 \dots x_{k_0}$  and  $y_i = x_{k_{i-1}+1} \dots x_{k_i}$  for all  $i \geq 1$ , then  $x = (y_i)_{i \in \omega}$  is a superfactorization.

$$x = \underbrace{(x_0 x_1 \dots x_{k_0})}_{y_0} \underbrace{(x_{k_0+1} x_{k_0+2} \dots x_{k_1})}_{y_1} \dots$$

## 2.2 Construction of countable scattered linear orderings

Countable scattered linear orderings are defined in terms of a forbidden property, namely that they may not contain a dense sub-ordering. Hausdorff's theorem states that countable scattered linear orderings can be constructed from finite orderings.

We denote by  $\mathcal{N}$  the subclass of finite linear orderings,  $\mathcal{O}$  the class of countable ordinals, and  $\mathcal{S}$  the class of countable scattered linear orderings.

**Theorem 1.** [12] *A countable linear ordering  $J$  is scattered if and only if  $J$  belongs to  $\bigcup_{\alpha \in \mathcal{O}} V_\alpha$ , where the classes  $V_\alpha$  are inductively defined by:*

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

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

Intuitively, the rank of a linear ordering is the maximum number of nested  $\omega$  and  $-\omega$ . It is linked to its Hausdorff class. For instance the orderings  $\omega$  of rank 1 and  $\omega^2$  of rank 2 belong respectively to  $V_1$  and  $V_2$ . Nevertheless, the class  $V_\alpha$  is not exactly the set of orderings of rank  $\alpha$ . For instance, the ordering  $\omega + \omega$  is of rank 1, and belongs to  $V_2$ . Therefore, we work on slightly different inductive classes. For any  $\alpha \in \mathcal{O}$ , we define the class  $W_\alpha$  by :

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

Those classes are strictly intermediate to the Hausdorff's ones: the inclusions  $V_\alpha \subset W_\alpha \subset V_{\alpha+1}$  hold for any ordinal  $\alpha$ . For instance, the ordering  $\omega^\alpha + \omega^\alpha$  belongs to  $W_\alpha$  but does not belong to  $V_\alpha$ , and the ordering  $\omega^{\alpha+1}$  belongs to  $V_{\alpha+1}$ , but does not belong to  $W_\alpha$ . Formally, the *rank* of a linear ordering  $J$  is the smallest ordinal  $\alpha$  such that  $J \in W_\alpha$ . For instance, the orderings of rank 0 are the finite ones. In this paper, we restrict to the set  $\bigcup_{n < \omega} W_n = \bigcup_{n < \omega} V_n$  of linear orderings of finite ranks.

By extension, the rank of a word is the rank of its length and the rank of a set of words is the upper bound of the ranks of its elements.

We denote by  $A^\diamond$  the set of all words indexed by countable scattered linear orderings and we also denote by  $A^{W_r}$  (respectively  $A^{V_r}$ ) the set of words whose length is an ordering in  $W_r$  (respectively  $V_r$ ) for some integer  $r$ . Thus the words of  $A^{W_r}$  have a rank lower than or equal to  $r$ .

### 2.3 Ramseyan factorization

Let  $\varphi : A^* \rightarrow E$  be a map from  $A^*$  into a finite set  $E$ . An  $\omega$ -factorization  $(x_j)_{j \in \omega}$  of a right-infinite word  $x \in A^\omega$  is said to be *Ramseyan* for  $\varphi$  if there exists an element  $e$  of  $E$  such that for any strictly positive integer  $j$ , we have  $\varphi(x_j) = e$ . The elements of  $E$  are usually called *colors*. The following theorem, due to Ramsey, states that every infinite word admits an  $\omega$ -factorization whose factors are all of the same color, except perhaps the first.

**Theorem 2. [Ramsey]** *Let  $\varphi : A^* \rightarrow E$  be a map from  $A^*$  into a finite set  $E$ . Every infinite word of  $A^\omega$  admits a ramseyan factorization for  $\varphi$ .*

Theorem 2 is a particular case of the Ramsey theorem [19]. Different versions of this theorem can be found in [16] with well detailed proofs.

In the case of words indexed by linear orderings, the previous theorem can be applied with a map from  $A^\diamond$  into a finite set when words admit  $\omega$ -factorizations.

**Corollary 3.** *Let  $\varphi : A^\diamond \rightarrow E$  be a map from  $A^\diamond$  into a finite set  $E$ . Let  $x \in A^\diamond$  be a word indexed by a scattered linear ordering. If  $x$  admits an  $\omega$ -factorization, then  $x$  admits a ramseyan superfactorization for  $\varphi$ .*

Note that Corollary 3 can be used symmetrically for words admitting an  $-\omega$ -factorization.

## 3 Rational sets of words on linear orderings

Bruyère and Carton [6] introduced rational expressions and automata for words indexed by countable scattered linear orderings. They proved that a set of words is rational if and only if it is recognizable, extending Kleene's theorem. More precisely, they defined an entire hierarchy of rational sets [7]. For each subset of rational operations, they consider the class of corresponding rational languages, and define transition functions for automata capturing the same languages. In this paper, we only consider rational sets of words of finite rank.

### 3.1 Rational expressions

The rational sets of finite rank can be obtained from finite sets of finite words using the union  $+$ , the concatenation  $\cdot$ , the star  $*$ , the omega iteration  $\omega$ , and the backwards omega iteration  $-\omega$ . Let  $X$  and  $Y$  be two sets of words. We define:

$$\begin{aligned}
X + Y &= \{z \mid z \in X \cup Y\} \\
X \cdot Y &= \{x \cdot y \mid x \in X, y \in Y\} \\
X^* &= \left\{ \prod_{j=1}^n x_j \mid n \in \mathcal{N}, x_j \in X \right\} \\
X^\omega &= \left\{ \prod_{j \in \omega} x_j \mid x_j \in X \right\} \\
X^{-\omega} &= \left\{ \prod_{j \in -\omega} x_j \mid x_j \in X \right\}
\end{aligned}$$

To define rational sets of words indexed by all linear orderings, three more operations are needed : the ordinal iteration  $\#$ , the backwards ordinal iteration  $-\#$  and the iteration for all linear countable scattered orderings  $\diamond$ .

$$\begin{aligned}
X^\# &= \left\{ \prod_{j \in J} x_j \mid J \in \mathcal{O}, x_j \in X \right\} \\
X^{-\#} &= \left\{ \prod_{j \in -J} x_j \mid J \in \mathcal{O}, x_j \in X \right\} \\
X \diamond Y &= \left\{ \prod_{j \in J \cup \hat{J}^*} z_j \mid J \in \mathcal{S} \setminus \emptyset, z_j \in X \text{ if } j \in J \text{ and } z_j \in Y \text{ if } j \in \hat{J}^* \right\}
\end{aligned}$$

In this paper, we are only interested in languages which are defined using  $+$ ,  $\cdot$ ,  $\omega$  and  $-\omega$ . We refer the reader to [6] for a precise definition of other rational operations. A set of words on linear orderings of finite ranks is *rational* if it is obtained from finite sets of finite words using  $+$ ,  $\cdot$ ,  $\omega$  and  $-\omega$ .

### 3.2 Automata on linear orderings

Let  $(Q, A, E, I, F)$  be a classical automaton on finite words with the usual notation. As the set  $E$  of transitions is a subset of  $Q \times A \times Q$ , the paths of such an automaton are finite. In Büchi automata, a word is accepted if it is the label of a path going infinitely many times through a given set of states. The problem is that this accepting condition does not allow the concatenation of infinite words to be recognized. To cope with this difficulty, a set of limit transitions included in  $\mathcal{P}(Q) \times Q$  is introduced. In this way, if an infinite path goes infinitely many times through the states of a set  $P$ , and the transition  $(P, q)$  exists, then the next state of the path may be  $q$ .

*Example 4.* Consider the automaton  $\mathcal{A} = (Q, A, E, I, F)$  of Figure 1 where  $Q = \{1, 2, 3\}$ ,  $A = \{a, b\}$ ,  $I = \{1\}$ ,  $F = \{3\}$ .

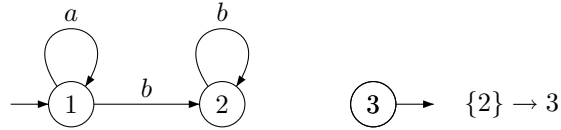


Figure 1: Automaton recognizing  $a^*b^\omega$

A limit transition  $\{2\} \rightarrow 3$  is added to  $E$ . Intuitively, an infinite path going through state 2 infinitely many times leads to state 3, and a path in  $\mathcal{A}$  leading

from state 2 to state 3 is labelled  $b^\omega$ . Finally, this automaton recognizes the language  $a^*b^\omega$ .

The previous limit transitions, called *left limit transitions*, allow an automaton to recognize sets of words indexed by countable ordinals. In order to get words indexed by linear scattered orderings, we also need *right limit transitions*.

**Definition 5.** An automaton  $\mathcal{A}$  on linear orderings is defined by a tuple  $\mathcal{A} = (Q, A, E, I, F)$  where  $Q$  is a finite set of states,  $A$  is a finite alphabet,  $E \subseteq (Q \times A \times Q) \cup (\mathcal{P}(Q) \times Q) \cup (Q \times \mathcal{P}(Q))$  is the set of transitions, and  $I \subseteq Q$  and  $F \subseteq Q$  are respectively the sets of initial and final states.

Right limit transitions are used symmetrically when a path has a limit length on the left. In order to use nested limit transitions, it is necessary to define the left (respectively right) limit sets of states in a given point of the path.

Consider a finite path  $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_n} q_n$  labelled  $x = a_1 \dots a_n$ . Note that a state is inserted between any two consecutive letters of  $x$ . In other words, a state  $q_k$  is associated with every two-factorization  $x = (a_1 \dots a_k)(a_{k+1} \dots a_n)$  of  $x$ . This definition of paths is generalized to automata on linear orderings in the following way: Let  $x$  be a word indexed by a linear scattered ordering  $J$ . To any two-factorization  $x = yz$  of  $x$ , one can associate a partition of  $J$  into two intervals  $(K, L)$  such that  $|y| = K$ , and  $|z| = L$ . Then a path labelled  $x$  is a function from the set  $\hat{J} = \{(K, L) \mid K \cup L = J \wedge \forall k \in K, \forall l \in L, k < l\}$  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  $\hat{J}$ . An element of  $\hat{J}$  is called a *cut*.

Let  $\gamma = (q_c)_{c \in \hat{J}}$  be a word of length  $\hat{J}$  over  $Q$ . We are now able to define the limit sets of states of  $\gamma$  in a given cut  $c$  of  $\hat{J}$ :

$$\lim_{c^-} \gamma = \{q \in Q \mid \forall c' < c, \exists c' < c'' < c \text{ such that } q = q_{c''}\}$$

$$\lim_{c^+} \gamma = \{q \in Q \mid \forall c' > c, \exists c < c'' < c' \text{ such that } q = q_{c''}\}$$

For instance, in Example 4, the word  $\gamma = (q_c)_{c \in \hat{\omega}}$  defined by  $q_{(\emptyset, \omega)} = 1$ ,  $q_{(\{0, 1, \dots, n\}, \{n+1, \dots\})} = 2$  for any positive integer  $n$ , and  $q_{(\omega, \emptyset)} = 3$  has the following nonempty limit  $\lim_{(\omega, \emptyset)^-} \gamma = \{2\}$ .

Finally, a path has to be compatible with the automata transitions:

**Definition 6.** 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  $\gamma$  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 every  $(K, L) \in \hat{J}$ :

- If there exists  $l \in L$  such that  $(K \cup \{l\}, L \setminus \{l\}) \in \hat{J}$ , then  $q_{(K, L)} \xrightarrow{a_l} q_{(K \cup \{l\}, L \setminus \{l\})} \in E$  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 is built on 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. It is said to be *successful* if it leads from an initial state to a final state. A word is *recognized* by an automata if it is the label of a successful path.

We denote by  $p \xrightarrow{x} q$  the existence of a path leading from state  $p$  to  $q$  of label  $x$ . The *content* of a path is the set of states occurring in the path (including the first and last states). A path leading from  $p$  to  $q$  of label  $x$  and of content  $P$  is denoted by  $p \xrightarrow[x]{P} q$ .

### 3.3 Generalizations of Kleene's theorem

Bruyère and Carton have generalized Kleene's theorem on words indexed by countable scattered linear orderings of finite ranks.

**Theorem 7.** [7] *A set of words of finite rank is rational if and only if it is recognized by an automata on linear orderings where limit transitions  $P \rightarrow q$  or  $q \rightarrow P$  verify  $q \notin P$ .*

This Theorem is illustrated by the Example of Figure 2.

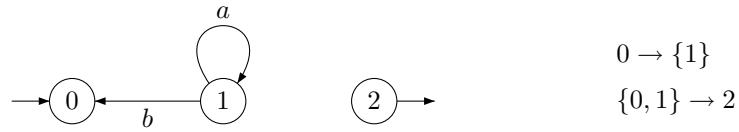


Figure 2: Automaton on linear orderings recognizing  $(a^{-\omega}b)^{\omega}$

## 4 Complement of a rational set of finite rank

A rational set of finite rank is thus both defined by a rational expression using operators  $+$ ,  $*$ ,  $\cdot$ ,  $\omega$  and  $-\omega$  and by automata on linear orderings where limit transitions  $P \rightarrow q$  or  $q \rightarrow P$  verify  $q \notin P$ . In this section, we prove that this class of rational sets is closed under complementation. In the case of finite words, it is known that rational sets are closed under complementation. Given an automaton on finite words recognizing a language  $L$ , the construction of an automaton recognizing the complement  $A^* \setminus L$  is based on the property that any finite automaton on finite words can be determinized. Büchi has generalized



this result for sets of words indexed by countable ordinals of finite ranks [10]. However, this property does not hold for automata on linear orderings. An automaton on linear orderings  $\mathcal{A} = (Q, A, E, I, F)$  is *deterministic* if for any state  $q \in Q$  and any word  $u \in A^\omega$ , there exists at most one path labelled  $u$  starting from  $q$ .

**Proposition 8.** *The language  $(a^{-\omega})^{-\omega}$  can not be accepted by a deterministic automaton.*

*Proof.* By contradiction, let us suppose the existence of a deterministic automaton  $\mathcal{A}$  recognizing  $(a^{-\omega})^{-\omega}$ . Denote the path labelled  $(a^{-\omega})^{-\omega}$  in the following way:

$$q \dots \dots \dots q_{-1,-2} \xrightarrow{a} q_{-1,-1} \xrightarrow{a} q_{-1,0} \dots q_{0,-2} \xrightarrow{a} q_{0,-1} \xrightarrow{a} q_{0,0}$$

Remove the last  $a$  and the state  $q_{0,0}$ . The word  $(a^{-\omega})^{-\omega}$  remains unchanged. As the automaton is deterministic, there exists a unique path starting from  $q$  and labelled  $(a^{-\omega})^{-\omega}$ . Thus we have  $q_{0,-1} = q_{0,0}$ . By induction, we obtain  $q_{0,m} = q_{0,0}$  for all  $m < 0$ . Using the same reasoning, one can remove the last  $a^{-\omega}$  and show that, except maybe the first state  $q$ , every state of the path is equal to  $q_{0,0}$ . Thus the automaton  $\mathcal{A}$  has the transition  $q_{0,0} \rightarrow \{q_{0,0}\}$ , which means that it accepts the set  $a^{-\#}$ .  $\square$

To cope with this difficulty of determinism, we use a different method based on equivalence classes to prove the closure of rational sets under complementation. Up to now, we are only able to prove this result in the case of rational sets of words of finite ranks.

**Theorem 9.** *Let  $L$  be a rational set of words on linear orderings, and let  $r$  be a finite integer. The complement  $A^{Wr} \setminus L$  is rational.*

In the case of finite words, Büchi has given a proof of the closure under complement of rational sets which does not work by constructing automata. It does not need the property of determinizability and it is based on the following equivalence relation defined for any finite automaton  $\mathcal{A} = (Q, A, E, I, F)$  on finite words:

$$u \sim v \text{ if and only if } \forall p \in Q, \forall q \in Q, p \xrightarrow{u} q \iff p \xrightarrow{v} q$$

Note that if a word  $u$  is the label of a successful path in  $\mathcal{A}$ , it holds for any equivalent word. So any equivalence class is either contained in the language  $L$  recognized by  $\mathcal{A}$ , or disjoint from  $L$ . Moreover, equivalence classes are rational thus the complement of  $L$  is rational as a finite union of equivalence classes. We extend this proof to automata on linear orderings of finite ranks. Let  $\mathcal{A} = (Q, A, E, I, F)$  be an automaton on linear orderings recognizing  $L$ . Recall that a path from  $p$  to  $q$  with label  $u$  and content  $P$  is denoted by  $p \xrightarrow[u]{P} q$ . As

the contents of paths are needed in limit transitions, we define the equivalence relation  $\sim$  by:

$$u \sim v \text{ if and only if } \forall p \in Q, \forall q \in Q, \forall P \subseteq Q, p \xrightarrow[u]{P} q \iff p \xrightarrow[v]{P} q$$

Note first that the equivalence relation has finitely many classes. Indeed the class of a word  $u$  depends on whether there is a path from  $p$  to  $q$  with content  $P$  for each triple  $(p, q, P)$ . If  $n$  denotes the size of  $Q$ , there are  $n^2 2^n$  such triples, and the relation  $\sim$  has at most  $2^{n^2 2^n}$  equivalence classes. We denote by  $\mathcal{C}$  the set of all equivalence classes of  $\sim$ . For any integer  $r$  and for any equivalence class  $C \in \mathcal{C}$ , the set  $C \cap A^{W_r}$  is called an *equivalence class of rank  $r$* , even if it contains all words of  $C$  of rank at most  $r$ . For each integer  $r$ , we denote by  $\mathcal{C}_r = \{C \cap A^{W_r} \mid C \in \mathcal{C}\}$  the set of equivalence classes of rank  $r$ . The cardinality of  $\mathcal{C}_r$  is at most the cardinality of  $\mathcal{C}$ . As in the case of finite words, each class  $C$  is either contained in  $L$  or disjoint from  $L$ . Therefore we have both equalities

$$L = \bigcup_{C \in \mathcal{C}, C \cap L \neq \emptyset} C \text{ and } \bar{L} = A^\diamond \setminus L = \bigcup_{C \in \mathcal{C}, C \cap L = \emptyset} C$$

The same holds for words of rank at most  $r$ .

$$L \cap A^{W_r} = \bigcup_{C \in \mathcal{C}_r, C \cap L \neq \emptyset} C \text{ and } A^{W_r} \setminus L = \bigcup_{C \in \mathcal{C}_r, C \cap L = \emptyset} C.$$

For each integer  $r$ , the family  $\mathcal{C}_r$  contains finitely many classes. To prove that  $A^{W_r} \setminus L$  is rational, it suffices to prove that each  $C \in \mathcal{C}_r$  is rational. We prove that claim by induction on  $r$ . The result holds obviously for  $r = 0$ , and the induction step is based on the following idea. Suppose that  $\mathcal{C}_r$  contains the classes  $\{C_1, \dots, C_m\}$ . We define rational expressions using the  $C_i$  as letters. An *elementary expression* is an expression of the form  $C_i, C_i^\omega$  or  $C_i^{-\omega}$  where  $C_i$  is a class of  $\mathcal{C}_r$ . We denote by  $B$  the set of elementary expressions. We consider the set  $B^*$  of all expressions obtained by concatenation of elementary expressions. Suppose for instance that  $\mathcal{C}_r = \{C_1, C_2\}$ . The set of elementary expressions is  $B = \{C_1, C_1^\omega, C_1^{-\omega}, C_2, C_2^\omega, C_2^{-\omega}\}$ . A typical example of an element of  $B^*$  is  $C_2^\omega C_1 C_2^{-\omega} C_1 C_2^{-\omega}$ . We consider each element of  $B^*$  as a rational expression over the letters  $C_i$ . Each expression of  $B^*$  denotes a set of words of rank at most  $r+1$ . By a slight abuse of language, we say that a word belongs to an expression  $R$  in  $B^*$  if it actually belongs to the set denoted by  $R$ . The two following lemmas are needed in the proof of Proposition 12. In Lemma 10, we first prove that each word of rank at most  $r+1$  belongs to at least one expression in  $B^*$ .

**Lemma 10.**  $A^{W_{r+1}} = \bigcup_{R \in B^*} R.$

*Proof.* Since equivalence classes of rank  $r$  contain words of rank at most  $r$ , each elementary expression of  $B$  contains words of rank at most  $r+1$ . Moreover, the rank is not modified by finite product, thus each  $R \in B^*$  is included in  $A^{W_{r+1}}$ .

Conversely, let  $x \in A^{W_{r+1}}$ . Since the length  $J$  of  $x$  belongs to the class  $W_{r+1}$ , it is a finite sum of linear orderings of class  $V_{r+1}$  defined by Hausdorff's

characterization:  $J = \sum_{i=1}^n K_i$  where  $\forall 1 \leq i \leq n, K_i \in V_{r+1}$ . Consider the

associated factorization:  $x = \prod_{i=1}^n x_i$  where  $\forall 1 \leq i \leq n, |x_i| = K_i$ .

We claim that each  $x_i$  belongs to an expression  $R_i$  of  $B^*$ . Lemma 10 follows from this claim since  $x$  belongs then to the expression  $R_1 \dots R_n$  of  $B^*$ .

We show that any factor  $x_i$  of  $x$  belongs to an expression  $R_i$  of  $B^*$ :

- Case 1: The rank of  $x_i$  is at most  $r$ ,  $|x_i| \in W_r$ . Thus  $x_i$  belongs to an equivalence class  $C$  of rank  $r$ . Thus we set  $R_i = C$ .

- Case 2:  $x_i$  is an  $\omega$ -product of words of rank at most  $r$  :

The ordering  $K_i$  can be decomposed  $K_i = \sum_{j \geq 0} K_{i,j}$  where  $K_{i,j} \in W_r$  for all  $j \geq 0$ . Thus  $x_i = \prod_{j \geq 0} x_{i,j}$  with  $x_{i,j} \in A^{W_r}$ . The number of equivalence classes of  $\mathcal{C}_r$  is finite. Therefore, by Corollary 3 there exists a Ramseyan superfactorization  $x_i = \prod_{j \geq 0} x'_{i,j}$  where all  $x'_{i,j}$  are words of rank at most  $r$  as a finite product of  $x_{i,j}$  and belong to the same equivalence class  $C$  of  $\mathcal{C}_r$ . Denote by  $C'$  the equivalence class of  $x_{i,0}$ . We set  $R_i = C' C^\omega$ .

- Case 3:  $x_i$  is a  $-\omega$ -product of words of rank at most  $r$  :

Similarly to case 2, we prove that  $x_i$  belongs to an expression  $C^{-\omega} C'$  where  $C, C' \in \mathcal{C}_r$  and we set  $R_i = C^{-\omega} C'$ .

- Case 4:  $x_i$  is a  $\zeta$ -product of words of rank at most  $r$  : There exists two words  $y_1$  and  $y_2$  such that  $x = y_1 \cdot y_2$  where  $y_1$  is a  $-\omega$ -product, and  $y_2$  is an  $\omega$ -product of words of rank at most  $r$ . Thus  $x_i$  belongs to an expression of the form  $C_1^{-\omega} C'_1 C'_2 C_2^\omega$ , where  $C_1, C'_1, C_2, C'_2 \in \mathcal{C}_r$  and we set  $R_i = C_1^{-\omega} C'_1 C'_2 C_2^\omega$ .

□

In Lemma 11, we prove that two words belonging to the same expression of  $B^*$  are  $\sim$ -equivalent. This means that each set denoted by an expression of  $B^*$  is included in a single  $\sim$ -class.

**Lemma 11.** *If two words  $x, y$  of rank at most  $r+1$  belong to the same expression  $R$  of  $B^*$ , then they satisfy  $x \sim y$ .*

*Proof.* Let  $R \in B^*$ . Let  $x$  and  $y$  be two words verifying  $x \in R$  and  $y \in R$ . We first show that  $x \sim y$  when  $R$  is an elementary expression of  $B$ .

- Case 1:  $R = C$  for some class  $C$  of  $\mathcal{C}_r$ .

As  $x$  and  $y$  belong to  $C$ , we get that  $x \sim y$ .

- Case 2:  $R = C^\omega$  for some class  $C$  of  $\mathcal{C}_r$ .

Let  $x = \prod_{i \geq 0} x_i$  and  $y = \prod_{i \geq 0} y_i$  be the factorizations satisfying  $x_i, y_i \in C$ ,  $\forall i \geq 0$ .

Let  $\gamma : p \xrightarrow{P} q$  be a path in  $\mathcal{A}$  labelled  $x$ . The path  $\gamma$  can be decomposed according to the factorization of  $x$  as a path

$$p = p_0 \xrightarrow{P_0} p_1 \xrightarrow{P_1} p_2 \dots q$$

ending with a limit transition  $P' \rightarrow q$ .

For each  $i \geq 0$ , we have  $x_i \sim y_i$  thus  $p_i \xrightarrow{P_i} p_{i+1}$ . Finally,  $p \xrightarrow{P} q$  is a path of  $\mathcal{A}$  which concludes our proof that  $x \sim y$ .

- Case 3:  $R = C^{-\omega}$  for some class  $C$  of  $\mathcal{C}_r$ . Similarly to case 2, we get that  $x \sim y$ .

Now we suppose that  $R$  is a finite product of elementary expressions of  $B$  :  
 $R = \prod_{i \geq 0}^n R_i$ .

There exist two factorizations  $x = \prod_{i=1}^n x_i$  and  $y = \prod_{i=1}^n y_i$  such that for all  $1 \leq i \leq n$ ,  $x_i \in R_i$  and  $y_i \in R_i$ .

Let us suppose the existence of a path  $p \xrightarrow{P} q$  in  $\mathcal{A}$ . Decompose it according to the previous factorization of  $x$ :

$$p = q_0 \xrightarrow{P_1} q_1 \xrightarrow{P_2} q_2 \dots \xrightarrow{P_n} q_n = q$$

Since the  $R_i$  are elementary expressions, we have  $x_i \sim y_i$  for all  $1 \leq i \leq n$ . Thus  $q_{i-1} \xrightarrow{P_i} q_i$  for all  $1 \leq i \leq n$  and it follows that  $p \xrightarrow{P} q$ .

We conclude that  $x \sim y$ . □

It follows from Lemmas 10 and 11 that each class  $C$  in  $\mathcal{C}_{r+1}$  satisfies

$$C = \bigcup_{R \in B^*, C \cap R \neq \emptyset} R$$

However, this is not a rational expression for  $C$  because there are infinitely many such expressions  $R$  included in  $C$ . In the following proposition, we show that the set of rational expressions included in some class  $C$  can be described by a rational expression over elementary expressions.

**Proposition 12.** *Each equivalence class in  $\mathcal{C}_r$  is rational.*

*Proof.* The proof is by induction on the rank  $r$ .

- $r = 0$ :

We consider classes of finite words. Given  $p \in Q$ ,  $q \in Q$  and  $P \subseteq Q$ , we denote by  $L_{p,q,P}$  the set of finite labels of paths in  $\mathcal{A}$  leading from  $p$  to  $q$  and of content  $P$ . One can easily prove that  $L_{p,q,P}$  is rational. Thus any  $\sim$ -class of finite words is rational as a boolean combination of the sets  $L_{p,q,P}$ .

- $r \geq 0$ :

Let  $\mathcal{C}_r = \{C_1, C_2, \dots, C_m\}$  be the set of equivalence classes of rank  $r$ . By inductive hypothesis, each class of rank  $r$  is rational:

$$\forall C \in \mathcal{C}_r, C \in \text{Rat}(A^{W_r})$$

Let  $B = \{C_1, C_1^\omega, C_1^{-\omega}, \dots, C_m, C_m^\omega, C_m^{-\omega}\}$ . From  $\mathcal{A}$ , we construct an automaton  $\mathcal{B}$  which links any expression  $R$  of  $B^*$  with the words of  $A^{W_{r+1}}$  belonging to  $R$ . Then the classes of  $\mathcal{C}_{r+1}$  will be defined as rational sets of elements of  $B^*$ .

Define the automaton  $\mathcal{B} = (Q \times \mathcal{P}(Q), B, E', I', F')$  by:

$$E' = \{(p, S) \xrightarrow{b} (p', S \cup T) \mid \exists x \in b \text{ such that } p \xrightarrow{x}_T p' \text{ in } \mathcal{A}\}$$

$$I' = \{(p, \emptyset) \mid p \in Q\} \text{ and } F' = Q \times \mathcal{P}(Q)$$

Note that a label of a path in  $\mathcal{B}$  is an element of  $B^*$ . Thus it is an expression denoting a set of words of  $A^{W_{r+1}}$ . Let  $L_{(p,\emptyset),(q,P)}$  be the set of labels of paths leading from  $(p, \emptyset)$  to  $(q, P)$  in  $\mathcal{B}$ , and let  $L_{p,q,P}^{r+1}$  be the set of labels of paths in  $\mathcal{A}$  leading from  $p$  to  $q$  of content  $P$  and of rank at most  $r+1$ . We claim that a word  $x$  in  $A^{W_{r+1}}$  belongs to  $L_{p,q,P}^{r+1}$  if and only if  $x$  belongs to an expression  $R$  of  $L_{(p,\emptyset),(q,P)}$ . Thus the set  $L_{p,q,P}^{r+1}$  is defined as the set of words belonging to a label of  $L_{(p,\emptyset),(q,P)}$ . The result follows as a class  $C$  of  $\mathcal{C}_{r+1}$  is defined as a boolean combination of the rational sets  $L_{(p,\emptyset),(q,P)}$  of finite words.

We now prove the claim. We show that for all  $p \in Q$ ,  $q \in Q$ ,  $P \subseteq Q$  and  $R \in B^*$ ,  $(p, \emptyset) \xrightarrow{R} (q, P)$  in  $\mathcal{B}$  if and only if there exists a word  $x$  in  $R$  such that  $p \xrightarrow{x}_P q$  in  $\mathcal{A}$ .

Let  $x$  be a word of rank at most  $r+1$ . By Lemma 10, there exists  $R \in B^*$  such that  $x \in R$ . Denote by  $R = b_1 b_2 \dots b_n$  with  $n \in \mathbb{N}$  and  $b_i \in B$  for any  $1 \leq i \leq n$ . We prove by induction on  $n \geq 1$  that for all  $p \in Q$ ,  $q \in Q$ , and  $P \subseteq Q$ ,  $p \xrightarrow{x}_P q$  in  $\mathcal{A}$  implies  $(p, \emptyset) \xrightarrow{R} (q, P)$  in  $\mathcal{B}$ . If  $n = 1$ ,  $R \in B$ , and the result is obtained by definition of  $\mathcal{B}$ . If  $n > 1$ ,  $x$  can be written  $x = x_1 x_2 \dots x_n$  with  $x_i \in b_i$  for any  $1 \leq i \leq n$ . Suppose that  $p \xrightarrow{x}_P q$  in  $\mathcal{A}$ . There exists  $q' \in Q$ ,  $P_1, P_2 \subseteq Q$  such that  $p \xrightarrow{x_1 \dots x_{n-1}}_{P_1} q'$  and  $q' \xrightarrow{x_n}_{P_2} q$  with

$P_1 \cup P_2 = P$ . By the inductive hypothesis,  $(p, \emptyset) \xrightarrow{b_1 \dots b_{n-1}} (q', P_1)$  in  $\mathcal{B}$  and by definition of  $\mathcal{B}$ ,  $(q', P_1) \xrightarrow{b_n} (q, P_1 \cup P_2)$ , thus  $(p, \emptyset) \xrightarrow{R} (q, P)$ .

Conversely, let  $R \in B^*$  and let  $x \in R$ . Suppose that  $(p, \emptyset) \xrightarrow{R} (q, P)$ . By construction of  $\mathcal{B}$ , there exists  $y \in R$  such that  $p \xrightarrow{y} q$  in  $\mathcal{A}$ . Moreover, from Lemma 11 we have  $x \sim y$ , thus  $p \xrightarrow{x} q$  in  $\mathcal{A}$ .

□

We come back to the proof of Theorem 9.

*Proof.* Let  $\mathcal{A}$  be an finite automaton on linear orderings recognizing  $L$  and let  $r$  be a finite rank. Let  $\mathcal{C}_r$  be the set of equivalence classes of rank  $r$  according to  $\mathcal{A}$ . From Proposition 12, we have that each class of  $\mathcal{C}_r$  is rational. Moreover, considering the definition of  $\sim$ , we note that if a word  $u$  is the label of a successful path in  $\mathcal{A}$ , it holds for any equivalent word. So an equivalence class is either contained in  $L$  or disjoint of  $L$ . We deduce a rational expression of  $A^{W_r} \setminus L$  as a finite union of the classes of  $\mathcal{C}_r$ :

$$A^{W_r} \setminus L = \bigcup_{C \in \mathcal{C}_r, C \cap L = \emptyset} C$$

□

The following example illustrates the construction of rational expressions of equivalence classes:

*Example 13.* Let  $\mathcal{A}$  be the automata of Figure 3 recognizing the set  $L = (a^\zeta)^*$ . We look for a rational expression of the equivalence classes of rank 1.

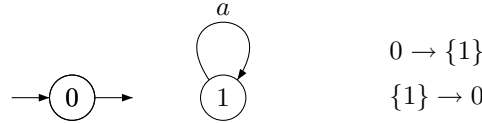


Figure 3: Automaton on linear orderings recognizing  $(a^\zeta)^*$

Let  $\mathcal{C}_0$  be the set of equivalence classes of finite words. The automata has two equivalence classes:  $\mathcal{C}_0 = \{a^+, \epsilon\}$ . Define the set  $B = \{a^+, a^\omega, a^{-\omega}, \epsilon\}$ .

The corresponding finite automaton  $\mathcal{B}$  is given in Figure 4. Note that the letter  $\epsilon$  has not been written in the automaton, and that terminal states have not been marked.

We deduce the following  $\sim$ -classes of rank 1:  $\mathcal{C}_1 = \{C_0, C_1, C_2, C_3, C_4, C_5, C_6, \}$ .

$$\begin{aligned} C_0 &= \epsilon, & C_1 &= a^+, & C_2 &= (a^\zeta)^+ a^{-\omega}, \\ C_3 &= a^\omega (a^\zeta)^*, & C_4 &= (a^\zeta)^+, & C_5 &= a^\omega (a^\zeta)^* a^{-\omega} \\ \text{and } C_6 &= (a + a^\omega + a^{-\omega})^* \cdot (a^\omega \cdot a + a \cdot a^{-\omega}) \cdot (a + a^\omega + a^{-\omega})^* \end{aligned}$$

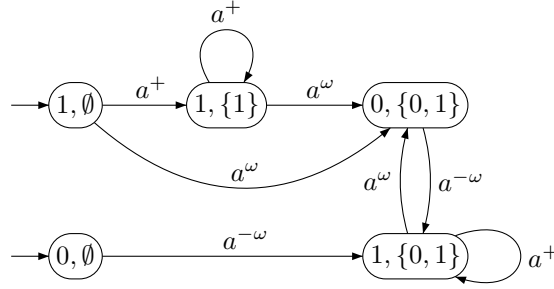


Figure 4: Automaton  $\mathcal{B}$

where the last class contains words which are not labels of a path in  $\mathcal{A}$ . Consider the automaton  $\mathcal{B}$  having all states final. Since it is an automaton on finite words,  $C_6$  is obtained from its complement.

Except for the classes  $C_0$  and  $C_4$  which are included in  $L$ , any other class of  $\mathcal{C}_1$  is disjoint of  $L$ . We get the following rational expression of  $L$  and its complement in the set of words of rank at most 1:

$$L \cap A^{W_1} = C_0 + C_4$$

$$A^{W_1} \setminus L = C_1 + C_2 + C_3 + C_5 + C_6$$

As a conclusion, we mention a question that is left open by this paper. A generalization of our result is that the class of rational sets of countable scattered linear orderings is closed under complementation.

## References

- [1] A. V. Aho, J. E. Hopcroft, J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, London, 1974.
- [2] D. Beauquier, Bilimites de langages reconnaissables, Theoret. Comput. Sci. 33 (2-3) (1984) 335-342.
- [3] N. Bedon, Finite automata and ordinals, Theoret. Comput. Sci. 156 (1996) 119-144.
- [4] N. Bedon, Automata, semigroups and recognizability of words on ordinals, Int. J. Alg. Comput. 8 (1998) 1-21.
- [5] N. Bedon, O. Carton, An Eilenberg theorem for words on countable ordinals, in: C. L. Lucchesi, A. V. Moura (Eds.), Latin'98: Theoretical Informatics, Vol. 1380 of Lect. Notes in Comput. Sci., Springer-Verlag, 1998, pp. 53-64.
- [6] V. Bruyère, O. Carton, Automata on linear orderings, in: J. Sgall, A. Pultr, P. Kolman (Eds.), MFCS'2001, Vol. 2136 of Lect. Notes in Comput. Sci., 2001, pp. 236-247, iGM report 2001-12.

- [7] V. Bruyère, O. Carton, Hierarchy among automata on linear orderings, in: IFIP TCS'2002, 2002, pp. 107–118.
- [8] J. R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik und grundl. Math.* 6 (1960) 66–92.
- [9] 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, Stanford University Press, 1962, pp. 1–11.
- [10] J. R. Büchi, Transfinite automata recursions and weak second order theory of ordinals, in: *Proc. Int. Congress Logic, Methodology, and Philosophy of Science*, Jerusalem 1964, North Holland, 1965, pp. 2–23.
- [11] Y. Choueka, Finite automata, definable sets, and regular expressions over  $\omega^n$ -tapes, *J. Comput. System Sci.* 17 (1) (1978) 81–97.
- [12] F. Hausdorff, *Set theory*, Chelsea Publishing Co., New York, 1962.
- [13] S. C. Kleene, Representation of events in nerve nets and finite automata, in: C. Shannon (Ed.), *Automata studies*, Princeton university Press, Princeton, 1956, pp. 3–41.
- [14] D. Muller, Infinite sequences and finite machines, in: *P. of Fourth Annual IEEE Symp. (Ed.), Switching Theory and Logical Design*, 1963, pp. 3–16.
- [15] M. Nivat, D. Perrin, Ensembles reconnaissables de mots bi-infinis, in: *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, 1982, pp. 47–59.
- [16] D. Perrin, J.-E. Pin, *Infinite words*, Vol. 141 of *Pure and applied mathematics*, Elsevier, 2004.
- [17] J.-E. Pin, Syntactic semigroups, in: *Handbook of formal languages*, Vol. 1, Springer, Berlin, 1997, pp. 679–746.
- [18] M. O. Rabin, Decidability of second-order theories and automata on infinite trees, *Trans. Amer. Math. Soc.* 141 (1969) 1–35.
- [19] F. D. Ramsey, On a problem of formal logic, *Proc. of the London math. soc.* 30 (1929) 338–384.
- [20] J. G. Rosenstein, *Linear ordering*, Academic Press, New York, 1982.
- [21] S. Safra, On the complexity of  $\omega$ -automata, in: *29th Annual Symposium on Foundations of computer sciences*, 1988, pp. 24–29.