

 Open access • Journal Article • DOI:10.1007/S00224-009-9222-6

## Logic and Rational Languages of Words Indexed by Linear Orderings — [Source link](#)

Nicolas Bedon, Alexis Bès, Olivier Carton, Chloé Rispal

**Institutions:** Institut Gaspard Monge, University of Paris

**Published on:** 01 May 2010 - Theory of Computing Systems ∨ Mathematical Systems Theory (Springer-Verlag)

**Topics:** Indexed language, Countable set, Decidability, Formal language and Regular language

Related papers:

- [Automata on linear orderings](#)
- [Decidability of second-order theories and automata on infinite trees.](#)
- [Complementation of rational sets on countable scattered linear orderings](#)
- [Transfinite Automata Recursions and Weak Second Order Theory of Ordinals](#)
- [Regular languages of words over countable linear orderings](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/logic-and-rational-languages-of-words-indexed-by-linear-jejqe9ddm6>

# Logic and Rational Languages of Words Indexed by Linear Orderings

Nicolas Bedon<sup>1</sup>, Alexis Bès<sup>2</sup>, Olivier Carton<sup>3</sup>, and Chloé Rispal<sup>1</sup>

<sup>1</sup> Université Paris-Est and CNRS

Laboratoire d'informatique de l'Institut Gaspard Monge, CNRS UMR 8049

Email: `Nicolas.Bedon@univ-mlv.fr`, `Chloe.Rispal@univ-mlv.fr`

<sup>2</sup> Université Paris-Est, LACL

Email: `bes@univ-paris12.fr`,

<sup>3</sup> Université Paris 7 and CNRS

LIAFA, CNRS UMR 7089

Email: `Olivier.Carton@liafa.jussieu.fr`

**Abstract.** We prove that every rational language of words indexed by linear orderings is definable in monadic second-order logic. We also show that the converse is true for the class of languages indexed by countable scattered linear orderings, but false in the general case. As a corollary we prove that the inclusion problem for rational languages of words indexed by countable linear orderings is decidable.

## 1 Introduction

In [4, 6], Bruyère and Carton introduce automata and rational expressions for words on linear orderings. These notions unify naturally previously defined notions for finite words, left- and right-infinite words, bi-infinite words, and ordinal words. They also prove that a Kleene-like theorem holds when the orderings are restricted to countable scattered linear orderings; recall that a linear ordering is scattered if it does not contain any dense sub-ordering. Since [4], the study of automata on linear orderings was carried on in several papers. The emptiness problem and the inclusion problem for rational languages is addressed in [7, 11]. The papers [5, 2] provide a classification of rational languages with respect to the rational operations needed to describe them. Algebraic characterizations of rational languages are presented in [2, 21, 20]. The paper [3] introduces a new rational operation of shuffle of languages which allows to deal with dense orderings, and extends the Kleene-like theorem proved in [4] to languages of words indexed by all linear orderings.

In this paper we are interested in connections between rational languages and languages definable in a logical formalism. The main motivations are, on one hand, to extend the classical results to the case of linear orderings, and on the other hand to get a better understanding of monadic second order (shortly: MSO) theories of linear orderings. Let us recall the state-of-the-art. In his seminal paper [8], Büchi proved that rational languages of finite words coincide with languages definable in the weak MSO theory of  $(\omega, <)$ , which allowed him to

prove decidability of this theory. In [9] he proved that a similar equivalence holds between rational languages of infinite words of length  $\omega$  and languages definable in the MSO theory of  $(\omega, <)$ . The result was then extended to languages of words indexed by a countable ordinal [10]. What can be said about MSO theories for linear orderings beyond ordinals? Using the automata technique, Rabin proved decidability of the MSO theory of the binary tree [19], from which he deduces decidability of the MSO theory of  $\mathbb{Q}$ , which in turn implies decidability of the MSO theory of countable linear orderings. Shelah [24] (see also [12, 26]) improved model-theoretical techniques that allow him to reprove almost all known decidability results about MSO theories, as well as new decidability results for the case of linear orderings. He proved in particular that the MSO theory of  $\mathbb{R}$  is undecidable. Shelah's decidability method is model-theoretical, and up to now no corresponding automata techniques are known. This led Thomas to ask [26] whether there is an appropriate notion of automata for words indexed by linear orderings beyond the ordinals. As mentioned in [4], this question was an important motivation for the introduction of automata over words indexed by linear orderings.

In this paper we study rational languages in terms of definability in MSO logic. Our main result is that, assuming the axiom of choice, every rational language of words indexed by linear orderings is definable in MSO logic. The proof does not rely on the classical encoding of an accepting run of an automaton accepting the language, but on an induction on the rational expression denoting the language. As a corollary we prove that the inclusion problem for rational languages of countable linear orderings is decidable, which extends [7] where the result was proved for countable scattered linear orderings. We also study the converse problem, i.e. whether every MSO-definable language of words indexed by linear orderings is rational. A key argument in order to prove this kind of results is the closure of the class of rational languages under complementation. Carton and Rispal [21] proved (using semigroup theory) that the class of rational languages of words indexed by countable scattered orderings is closed under complementation; building on this, we prove that every MSO-definable language of words indexed by countable scattered linear orderings is rational, giving thus the equivalence between rational expressions and MSO logic in this case. On the other hand we show that for every finite alphabet  $A$  the language of words over  $A$  indexed by scattered orderings is not rational, while its complement is. This proves that the class of rational languages of words over linear orderings is not closed under complementation, and as a corollary of the previous results that this class is strictly included in the class of MSO-definable languages.

The paper is organized as follows: we recall in Section 2 some useful definitions related to linear orderings. Section 3 introduces rational expressions for words over linear orderings. Section 4 recalls useful notions related to MSO. In Section 5 we show that rational languages are MSO-definable. Section 6 deals with the converse problem. We conclude the paper with some open questions.

## 2 Linear Orderings

In this section we recall useful definitions and results about linear orderings. A good reference on the subject is Rosenstein's book [22].

A *linear ordering*  $J$  is an ordering  $<$  which is total, that is, for any  $j \neq k$  in  $J$ , either  $j < k$  or  $k < j$  holds. Given a linear ordering  $J$ , we denote by  $-J$  the *backwards* linear ordering obtained by reversing the ordering relation. For instance,  $-\omega$  is the backwards linear ordering of  $\omega$  which is used to index the so-called left-infinite words.

The sum of orderings is concatenation. Let  $J$  and  $K_j$  for  $j \in J$ , be linear orderings. The linear ordering  $\sum_{j \in J} K_j$  is obtained by concatenation of the orderings  $K_j$  with respect to  $J$ . More formally, the *sum*  $\sum_{j \in J} K_j$  is the set  $L$  of all pairs  $(k, j)$  such that  $k \in K_j$ . The relation  $(k_1, j_1) < (k_2, j_2)$  holds 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 two orderings  $K_1$  and  $K_2$  is denoted  $K_1 + K_2$ .

Given two elements  $j, k$  of a linear ordering  $J$ , we denote by  $[j; k]$  the interval  $[\min(j, k), \max(j, k)]$ . The elements  $j$  and  $k$  are called *consecutive* if  $j < k$  and if there is no element  $i \in J$  such that  $j < i < k$ . An ordering is *dense* if it contains no pair of consecutive elements. More generally, a subset  $K \subset J$  is *dense* in  $J$  if for any  $j, j' \in J$  such that  $j < j'$ , there is  $k \in K$  such that  $j < k < j'$ .

A *cut* of a linear ordering  $J$  is a pair  $(K, L)$  of intervals such that  $J = K \cup L$  and such that for any  $k \in K$  and  $l \in L$ ,  $k < l$ . The set of all cuts of the ordering  $J$  is denoted by  $\hat{J}$ . This set  $\hat{J}$  can be linearly ordered by the relation defined by  $c_1 < c_2$  if and only if  $K_1 \subsetneq K_2$  for any cuts  $c_1 = (K_1, L_1)$  and  $c_2 = (K_2, L_2)$ . This linear ordering can be extended to  $J \cup \hat{J}$  by setting  $j < c_1$  whenever  $j \in K_1$  for any  $j \in J$ .

The consecutive elements of  $\hat{J}$  deserve some attention. For any element  $j$  of  $J$ , define two cuts  $c_j^-$  and  $c_j^+$  by  $c_j^- = (K, \{j\} \cup L)$  and  $c_j^+ = (K \cup \{j\}, L)$  where  $K = \{k \mid k < j\}$  and  $L = \{k \mid j < k\}$ . It can be easily checked that the pairs of consecutive elements of  $\hat{J}$  are the pairs of the form  $(c_j^-, c_j^+)$ .

A *gap* of an ordering  $J$  is a cut  $(K, L)$  such that  $K \neq \emptyset$ ,  $L \neq \emptyset$ ,  $K$  has no greatest element and  $L$  has no least element. An ordering  $J$  is *complete* if it has no gap.

## 3 Words and rational expressions

Given a finite alphabet  $A$ , a *word*  $(a_j)_{j \in J}$  is a function from  $J$  to  $A$  which maps any element  $j$  of  $J$  to a letter  $a_j$  of  $A$ . We say that  $J$  is the *length*  $|x|$  of the word  $x$ . For instance, the *empty word*  $\varepsilon$  is indexed by the empty linear ordering  $J = \emptyset$ . Usual finite words are the words indexed by finite orderings  $J = \{1, 2, \dots, n\}$ ,  $n \geq 0$ . A word of length  $J = \omega$  is usually called an  $\omega$ -word or an infinite word. A word of length  $\zeta = -\omega + \omega$  is a sequence  $\dots a_{-2}a_{-1}a_0a_1a_2\dots$  of letters which is usually called a bi-infinite word.

The sum operation on linear orderings leads to a notion of product of words as follows. Let  $J$  and  $K_j$  for  $j \in J$ , be linear orderings. Let  $x_j = (a_{k,j})_{k \in K_j}$  be a

word of length  $K_j$ , for any  $j \in J$ . The *product*  $\prod_{j \in J} x_j$  is the word  $z$  of length  $L = \sum_{j \in J} K_j$  equal to  $(a_{k,j})_{(k,j) \in L}$ . For instance, the word  $a^\zeta = b^{-\omega} a^\omega$  of length  $\zeta$  is the product of the two words  $b^{-\omega}$  and  $a^\omega$  of length  $-\omega$  and  $\omega$  respectively.

We now recall the notion of rational sets of words indexed by linear orderings as defined in [4, 3]. The rational operations include of course the usual Kleene operations for finite words which are the union  $+$ , the concatenation  $\cdot$  and the star operation  $*$ . They also include the omega iteration  $\omega$  usually used to construct  $\omega$ -words and the ordinal iteration  $\sharp$  introduced by Wojciechowski [28] for ordinal words. Four new operations are also needed: the backwards omega iteration  $-\omega$ , the backwards ordinal iteration  $-\sharp$ , a binary operation denoted  $\diamond$  which is a kind of iteration for all orderings, and finally a shuffle operation which allows to deal with dense linear orderings.

We respectively denote by  $\mathcal{N}$ ,  $\mathcal{O}$  and  $\mathcal{L}$  the classes of finite orderings, the class of all ordinals and the class of all linear orderings. For an ordering  $J$ , we denote by  $\hat{J}^*$  the set  $\hat{J} \setminus \{(\emptyset, J), (J, \emptyset)\}$  where  $(\emptyset, J)$  and  $(J, \emptyset)$  are the first and last cut. Given two sets  $X$  and  $Y$  of words, define

$$\begin{aligned} X + Y &= \{z \mid z \in X \cup Y\}, \\ X \cdot Y &= \{x \cdot y \mid x \in X, y \in Y\}, \\ X^* &= \{\prod_{j \in \{1, \dots, n\}} x_j \mid n \in \mathcal{N}, x_j \in X\}, \\ X^\omega &= \{\prod_{j \in \omega} x_j \mid x_j \in X\}, \\ X^{-\omega} &= \{\prod_{j \in -\omega} x_j \mid x_j \in X\}, \\ X^\sharp &= \{\prod_{j \in \alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\}, \\ X^{-\sharp} &= \{\prod_{j \in -\alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\}, \\ X \diamond Y &= \{\prod_{j \in J \cup \hat{J}^*} z_j \mid J \in \mathcal{L}, z_j \in X \text{ if } j \in J \text{ and } z_j \in Y \text{ if } j \in \hat{J}^*\}. \end{aligned}$$

We denote by  $A^\diamond$  the set of words over  $A$  indexed by linear orderings. Note that we have  $A^\diamond = (A \diamond \varepsilon) + \varepsilon$ .

For every finite alphabet  $A$ , every  $n \geq 1$ , and all languages  $L_1, \dots, L_n \subseteq A^\diamond$ , we define  $\text{sh}(L_1, \dots, L_n)$  as the set of words  $w \in A^\diamond$  that can be written as  $w = \prod_{j \in J} w_j$ , where  $J$  is a complete linear ordering without first and last element, and there exists a partition  $(J_1, \dots, J_n)$  of  $J$  such that all  $J_i$ 's are dense in  $J$ , and for every  $j \in J$ , if  $j \in J_k$  then  $w_j \in L_k$ .

An abstract *rational expression* is a well-formed term of the free algebra over  $\{\emptyset\} \cup A$  with the symbols denoting the rational operations as function symbols. Each rational expression denotes a set of words which is inductively defined by the above definitions of the rational operations. A set of words is *rational* if it can be denoted by a rational expression. As usual, the dot denoting concatenation is omitted in rational expressions.

Automata which recognize languages of words indexed by linear orderings were introduced in [4]. In the latter paper a Kleene-like theorem was also shown for the special case of languages of words indexed by countable scattered linear orderings. Recall that a linear ordering is scattered if it does not contain any dense sub-ordering. The general case of words indexed by all linear orderings was proven in [3]. We refer e.g. to these papers for more details about automata; in this paper we shall deal only with rational expressions.

## 4 Monadic Second-Order Logic

In this section we recall useful elements of monadic second-order logic, and settle some notations. For more details about MSO logic we refer e.g. to Thomas' survey paper [27]. Monadic second-order logic is an extension of first-order logic that allows to quantify over elements as well as subsets of the domain of the structure.

Given a signature  $\mathcal{L}$ , one can define the set of *MSO-formulas* over  $\mathcal{L}$  as well-formed formulas that can use first-order variable symbols  $x, y, \dots$  interpreted as elements of the domain of the structure, monadic second-order variable symbols  $X, Y, \dots$  interpreted as subsets of the domain, symbols from  $\mathcal{L}$ , and a new binary predicate  $x \in X$  interpreted as “ $x$  belongs to  $X$ ”. We call *MSO sentence* any MSO formula without free variable.

Given a signature  $\mathcal{L}$  and an  $\mathcal{L}$ -structure  $M$  with domain  $D$ , we say that a relation  $R \subseteq D^m \times (2^D)^n$  is *MSO-definable* in  $M$  if and only if there exists a MSO-formula over  $\mathcal{L}$ , say  $\varphi(x_1, \dots, x_m, X_1, \dots, X_n)$  which is true in  $M$  if and only if  $(x_1, \dots, x_m, X_1, \dots, X_n)$  is interpreted by an  $(m+n)$ -tuple of  $R$ .

As usual, we will often confuse logical symbols with their interpretation. Moreover we will use freely abbreviations such as  $\exists x \in X \varphi$ ,  $\forall X \subseteq Y \varphi$ ,  $\exists! t \varphi$ , and so on.

Given any finite alphabet  $A$ , let us consider the signature  $\mathcal{L}_A = \{<, (P_a)_{a \in A}\}$  where  $<$  is a binary relation symbol and the  $P_a$ 's are unary predicates (over first-order variables). One can associate to every word  $w = (a_j)_{j \in J}$  over  $A$  (where  $a_j \in A$  for every  $j$ ) the  $\mathcal{L}_A$ -structure  $M_w = (J; <; (P_a)_{a \in A})$  where  $<$  is interpreted as the ordering over  $J$ , and  $P_a(x)$  holds if and only if  $a_x = a$ . In order to take into account the case  $w = \varepsilon$ , which leads to the structure  $M_\varepsilon$  which has an empty domain, we will allow structures to be empty.

Given a MSO sentence  $\varphi$  over the signature  $\mathcal{L}_A$ , we define the language  $L_\varphi$  as the set of words  $w$  over  $A$  such that  $M_w \models \varphi$ . We will say that a language  $L$  over  $A$  is definable in MSO logic (or *MSO-definable*) if and only if there exists a MSO-sentence  $\varphi$  over the signature  $\mathcal{L}_A$  such that  $L = L_\varphi$ .

## 5 Rational languages are MSO-definable

Büchi's proof [8] that every rational language  $L$  of finite words is definable in MSO logic relies on the encoding of an accepting run of an automaton  $\mathcal{A}$  recognizing  $L$ . Given a word  $w$ , one expresses the existence of a successful path in  $\mathcal{A}$  labeled by  $w$ , by encoding each state of the path on a position of  $w$ , which is possible because - up to a finite number of elements - the underlying ordering of the path is the same as the one of the word. This property still holds when one considers infinite words of length  $\omega$ , and more generally of any ordinal length. However it does not hold anymore for words indexed by all linear orderings, since for a word of length  $J$ , the path of the automaton is defined on the set  $\hat{J}$  of cuts of  $J$  (see [4]), and in general  $\hat{J}$  can be quite different from  $J$  - consider e.g. the case  $J = \mathbb{Q}$  for which  $J$  is countable while  $\hat{J}$  is not. Thus in our situation

there seems to be no natural extension of the classical Büchi's encoding technique. In order to overcome this issue, we use a proof by induction over rational expressions.

**Proposition 1.** *(Assuming the Axiom of Choice) For every finite alphabet  $A$  and every language  $L \subseteq A^\omega$ , if  $L$  is rational then it is definable in monadic second-order logic.*

Let us give a quick outline of the proof. One proves that for every rational language  $L$  there exists a MSO formula  $\varphi(X)$  over the signature  $\mathcal{L}_A$  such that for every word  $w$  over  $A$  indexed by some linear ordering  $J$ , we have  $w \in L$  if and only if  $M_w$  satisfies  $\varphi$  when  $X$  is interpreted by  $J$ . This yields Proposition 1 since every rational language  $L$  can then be defined by the MSO sentence  $\exists X(\varphi(X) \wedge \forall x x \in X)$ . The proof proceeds by induction on the rational expression denoting  $L$ ; this approach is not new, see e.g. [15]. The case of the empty word, as well as the one of singletons, union and product operations, are easy. For the other rational operations one has to find a way to express that the set  $X$  can be partitioned in some way in intervals. Consider for instance the case of the  $\omega$ -power operation. Assume that  $L$  is definable by the MSO formula  $\varphi(X)$ . Then  $L^\omega$  could be defined by a MSO formula which express the existence of a partition of  $X$  in a sequence  $(Y_i)_{i \in \omega}$  of intervals  $Y_i$  such that  $\varphi(Y_i)$  holds for every  $i$ . Since the existence of such a partition cannot be expressed directly in MSO, one reformulates this property as the existence of a partition of  $X$  in two subsets  $X_1, X_2$  such that every  $Y_i$  corresponds to an interval which consists in elements of  $X_1$  only, or elements of  $X_2$  only, and which is maximal for inclusion among such intervals. These maximal intervals are definable in MSO in terms of  $X, X_1$  and  $X_2$ , and moreover one can express that the order type of the sequence of these maximal intervals is  $\omega$ . This allows to find a MSO formula which defines  $L^\omega$ . The idea of interleaving finitely many subsets in order to encode some partition of  $X$  in intervals is also used for the other rational operations.

We illustrate Proposition 1 with several examples, over the alphabet  $A = \{a, b\}$ .

*Example 1.* Let  $L_1$  be the set of words  $w = (a_j)_{j \in J}$  (with  $a_j \in A$ ) such that  $J$  has a least element  $j_0$  with  $a_{j_0} = a$ , and  $a_j = b$  for some  $j \in J$ . This language can be represented by the rational expression  $aA^\omega bA^\omega$ . It is also MSO-definable by the sentence

$$\exists x \exists y (P_a(x) \wedge P_b(y) \wedge \neg \exists z z < x).$$

*Example 2.* Let  $L_2$  be the set of words indexed by a linear ordering  $J$  such that the set of positions  $j \in J$  for which  $w_j = a$  (respectively  $w_j = b$ ) is dense in  $J$ . This language can be represented by the rational expression  $\text{sh}(a, b, \varepsilon)$ . It is MSO-definable by the sentence

$$\forall x \forall y (x < y \implies \exists z \exists t (x < z < y \wedge P_a(z) \wedge x < t < y \wedge P_b(t))).$$

*Example 3.* The language  $L_3 = a^\omega a^{-\omega}$  is definable in MSO by the formula

$$\begin{aligned} & \forall x P_a(x) \wedge \exists X_1 \exists X_2 (\forall x (x \in X_1 \leftrightarrow x \notin X_2) \\ & \wedge \forall x \forall y ((x \in X_1 \wedge y \in X_2) \rightarrow x < y) \\ & \wedge \text{Omega}(X_1) \wedge \text{MinusOmega}(X_2)) \end{aligned}$$

where  $\text{Omega}(X_1)$  (respectively  $\text{MinusOmega}(X_2)$ ) expresses that the order type of  $X_1$  is  $\omega$  (respectively  $-\omega$ ). One can show that the predicates  $\text{Omega}$  and  $\text{MinusOmega}$  are MSO-definable.

*Example 4.* The language  $L_4$  of words whose length is a complete ordering can be represented by the rational expression  $(\varepsilon + \text{sh}(a + b)) \diamond (a + b)$ . It is also MSO-definable by the sentence

$$\forall Y ((\exists x \varphi(x, Y)) \implies (\exists x (\varphi(x, Y) \wedge \forall z (\varphi(z, Y) \implies x \leq z))))$$

where  $\varphi(x, Y)$  is an abbreviation for  $\forall y \in Y y \leq x$ .

*Example 5.* Consider the language  $L_5$  of words over  $A$  whose length is a non-scattered ordering. It follows from [22, chap. 4] that  $L_5$  consists in words  $w$  which can be written as  $w = \prod_{k \in K} w_k$  where  $K$  is a dense ordering, and  $w_k \neq \varepsilon$  for every  $k \in K$ . From this decomposition one can deduce that a convenient rational expression for  $L_5$  is  $\text{sh}(A^\diamond(a + b)A^\diamond, \varepsilon)$ . The language  $L_5$  can also be defined by the following MSO formula

$$\begin{aligned} & \exists X (\exists x_1 \in X \exists x_2 \in X x_1 < x_2 \\ & \wedge \forall y_1 \in X \forall y_2 \in X (y_1 < y_2 \implies \exists z \in X (y_1 < z \wedge z < y_2))). \end{aligned}$$

Combining Proposition 1 and Rabin's result [19] about the decidability of the MSO theory of countable linear orderings, yields the following result.

**Corollary 1.** *The inclusion problem for rational languages of words over countable linear orderings is decidable.*

This improves [7] where the authors prove the result for words over *scattered* countable linear orderings.

## 6 MSO-definable languages vs rational languages

In this section we consider the problem whether MSO-definable languages are rational. The answer is positive if we consider words indexed by countable scattered linear orderings. Indeed we can prove the following result.

**Proposition 2.** *For every finite alphabet  $A$  and every language  $L$  of words over  $A$  indexed by countable scattered linear orderings,  $L$  is rational if and only if it is MSO-definable.*



As for the finite words case, the proof that  $L = L_\phi$  for some MSO sentence  $\phi$  implies that  $L$  is rational relies on the construction of an automaton accepting  $L$ , by induction on the structure of  $\phi$ . The effectiveness of this construction, together with the decidability of the emptiness problem for automata on words indexed by countable scattered linear orderings [11], yield the following corollary.

**Corollary 2.** *The monadic second order theory of countable scattered linear orderings is decidable.*

Note that the latter result is also a direct consequence of Rabin’s result [19] about the decidability of the MSO theory of countable linear orderings (the property “to be scattered” is expressible in the latter theory).

Proposition 2 does not hold anymore if we consider languages of words indexed by all linear orderings. Indeed consider, for every finite alphabet  $A$ , the language  $S_A$  of words over  $A$  indexed by scattered linear orderings. The language  $S_A$  is the complement of the language  $L_5$  of Example 5. Since  $L_5$  is definable in MSO, the same holds for  $S_A$ . However one can show by a pumping argument that no automaton can recognize  $S_A$ . This fact together with the equivalence between rational languages and languages recognizable by automata [3] yield the following result.

**Proposition 3.** *For every finite alphabet  $A$ , the language  $S_A$  of words over  $A$  indexed by scattered linear orderings is not rational.*

On the other hand the language  $L_5$  was shown to be rational. Thus we can deduce the following result from Propositions 1 and 3.

**Corollary 3.** *For every finite alphabet  $A$ , the class of rational languages over  $A$  is not closed under complementation, and is strictly included in the class of MSO-definable languages.*

## 7 Open questions

Let us mention some related problems. It would be interesting to determine which syntactic fragment of the monadic second-order theory captures rational languages. The proof of Proposition 1, which uses an induction on the rational expression, gives rise to defining formulas where the alternation of (second-order) quantifiers is unbounded. However if we consider the special form of formulas used in the proof, together with classical techniques of re-using variables we can show that every rational language can be defined by MSO formulas of the form  $\forall X_1 \dots \forall X_m \exists Y_1 \dots \exists Y_n \forall Z_1 \dots \forall Z_p \varphi$ , where  $\varphi$  has no monadic second-order quantifier. We already know that the  $\forall\exists\forall$ -fragment of MSO contains non-rational languages, since by Proposition 3 the language of words indexed by scattered orderings, which can be defined by a  $\forall$ -formula, is not rational. Thus it would be interesting to know the expressive power of smaller syntactic fragments with respect to rational languages, and in particular the existential fragment.

Recall that for the MSO theory of  $\omega$  (and more generally any countable ordinal) the existential fragment is equivalent in terms of expressive power to the full theory. This comes from the fact that the formula encoding a successful run of an automaton is existential (for second-order variables). In our context the existential fragment does not capture all rational languages, as one can prove e.g. that the language  $a^\omega$  is not existentially definable. We conjecture that the class of languages definable by existential formulas is strictly included in the class of rational languages.

Another related problem is the expressive power of first-order logic. For finite words the McNaughton-Papert Theorem [14] shows that sets of finite words defined by first-order sentences coincide with star-free languages. Schützenberger gave another characterization of star-free sets, based on the equivalence of automata and an algebraic formalism, the finite monoids, for the definition of sets of finite words. He proved that the star-free sets are exactly those definable by a finite group-free monoid [23]. This double equivalence of Schützenberger, McNaughton and Papert was already extended to the infinite words by Ladner [13], Thomas [25] and Perrin [17], to words whose letters are indexed by all the relative integers by Perrin and Pin [17, 16, 18], and to the countable ordinals case by Bedon [1]. We already know [2] that a language of countable scattered linear orderings is star-free if and only if its syntactic  $\diamond$ -semigroup is finite and aperiodic. However, one can show that first-order definable languages of countable scattered linear orderings do not coincide any more with star-free and aperiodic ones [2, 22]. It would be interesting to characterize languages which are first-order definable.

## References

1. N. Bedon. Logic over words on denumerable ordinals. *Journal of Computer and System Science*, 63(3):394–431, 2001.
2. N. Bedon and C. Rispal. Schützenberger and Eilenberg theorems for words on linear orderings. In C. De Felice and A. Restivo, editors, *DLT'2005*, volume 3572 of *Lect. Notes in Comput. Sci.*, pages 134–145. Springer-Verlag, 2005.
3. A. Bès and O. Carton. A Kleene theorem for languages of words indexed by linear orderings. *Int. J. Found. Comput. Sci.*, 17(3):519–542, 2006.
4. V. Bruyère and O. Carton. Automata on linear orderings. In J. Sgall, A. Pultr, and P. Kolman, editors, *MFCS'2001*, volume 2136 of *Lect. Notes in Comput. Sci.*, pages 236–247, 2001.
5. V. Bruyère and O. Carton. Hierarchy among automata on linear orderings. In R. Baeza-Yate, U. Montanari, and N. Santoro, editors, *Foundation of Information technology in the era of network and mobile computing*, pages 107–118. Kluwer Academic Publishers, 2002.
6. V. Bruyère and O. Carton. Automata on linear orderings. *J. Comput. System Sci.*, 73(1):1–24, 2007.
7. V. Bruyère, O. Carton, and G. Sénizergues. Tree automata and automata on linear orderings. In T. Harju and J. Karhumäki, editors, *WORDS'2003*, pages 222–231. Turku Center for Computer Science, 2003.

8. J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik und grundl. Math.*, 6:66–92, 1960.
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*, pages 1–11. Stanford University Press, 1962.
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*, pages 2–23. North Holland, 1965.
11. O. Carton. Accessibility in automata on scattered linear orderings. In K. Diks and W. Rytter, editors, *MFCS'2002*, volume 2420 of *Lect. Notes in Comput. Sci.*, pages 155–164, 2002.
12. Y. Gurevich. Monadic second-order theories. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, pages 479–506. Springer-Verlag, Perspectives in Mathematical Logic, 1985.
13. R. E. Ladner. Application of model theoretic games to discrete linear orders and finite automata. *Inform. Control*, 33, 1977.
14. R. McNaughton and S. Papert. *Counter free automata*. MIT Press, Cambridge, MA, 1971.
15. C. Michaux and F. Point. Les ensembles  $k$ -reconnaissables sont définissables dans  $\langle N, +, V_k \rangle$ . (the  $k$ -recognizable sets are definable in  $\langle N, +, V_k \rangle$ ). *C. R. Acad. Sci. Paris, Sér. I*(303):939–942, 1986.
16. D. Perrin. An introduction to automata on infinite words. In M. Nivat, editor, *Automata on infinite words*, volume 192 of *Lect. Notes in Comput. Sci.*, pages 2–17. Springer, 1984.
17. D. Perrin. Recent results on automata and infinite words. In M. P. Chytil and V. Koubek, editors, *Mathematical foundations of computer science*, volume 176 of *Lect. Notes in Comput. Sci.*, pages 134–148, Berlin, 1984. Springer.
18. D. Perrin and J. E. Pin. First order logic and star-free sets. *J. Comput. System Sci.*, 32:393–406, 1986.
19. M.O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.
20. C. Rispal. *Automates sur les ordres linéaires: complémentation*. PhD thesis, University of Marne-la-Vallée, France, 2004.
21. C. Rispal and O. Carton. Complementation of rational sets on countable scattered linear orderings. In C. S. Calude, E. Calude, and M. J. Dinneen, editors, *DLT'2004*, volume 3340 of *Lect. Notes in Comput. Sci.*, pages 381–392, 2004.
22. J. G. Rosenstein. *Linear orderings*. Academic Press, New York, 1982.
23. M. P. Schützenberger. On finite monoids having only trivial subgroups. *Inform. Control*, 8:190–194, 1965.
24. S. Shelah. The monadic theory of order. *Annals of Mathematics*, 102:379–419, 1975.
25. W. Thomas. Star free regular sets of  $\omega$ -sequences. *Inform. Control*, 42:148–156, 1979.
26. W. Thomas. Ehrenfeucht games, the composition method, and the monadic theory of ordinal words. In *Structures in Logic and Computer Science, A Selection of Essays in Honor of A. Ehrenfeucht*, number 1261 in *Lect. Notes in Comput. Sci.*, pages 118–143. Springer-Verlag, 1997.
27. W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume III, pages 389–455. Springer-Verlag, 1997.

28. J. Wojciechowski. Finite automata on transfinite sequences and regular expressions. *Fundamenta informaticæ*, 8(3-4):379–396, 1985.