Logic and Rational Languages of Words Indexed by Linear Orderings

Nicolas Bedon[1], Alexis Bès[2], Olivier Carton[3], and Chloé Rispal[1]

[1] 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
[2] Université Paris-Est, LACL
Email: bes@univ-paris12.fr,
[3] 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 ω 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, 27]) 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 [27] 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 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 recognizing 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 proven 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 contained 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, rational expressions for words over linear orderings, automata and MSO. In Section 3 we show that rational languages are MSO-definable. Section 4 deals with the converse problem. We conclude the paper with some open questions.

2 Preliminaries

2.1 Linear Orderings

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 ω 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 $j < k < j'$. An ordering is *scattered* if it contains no dense sub-ordering.

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 $j \in J$. For an ordering J, we denote by $\hat{J}^*$ the set $\hat{J} \setminus \{(\varnothing, J), (J, \varnothing)\}$ where $(\varnothing, J)$ and $(J, \varnothing)$ are the first and last cut. 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 $K \neq \varnothing$, $L \neq \varnothing$, K has no last element and L has no first element. An ordering J is *complete* if it has no gap. **For example, the linear ordering of the real numbers $\mathbb{R}$ is complete, whereas the linear ordering of the rational numbers $\mathbb{Q}$ is not.**

We respectively denote by $\mathcal{N}$, $\mathcal{O}$ and $\mathcal{L}$ the class of finite orderings, the class of all ordinals and the class of all linear orderings.

2.2 Words and rational expressions

Given a finite alphabet A and a linear ordering J, 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* ε is indexed by

the empty linear ordering $J = \varnothing$. Usual finite words are the words indexed by finite orderings $J = \{1, 2, \ldots, n\}$, $n \geq 0$. A word of length $J = \omega$ is usually called an ω-word or an infinite word. A word of length $\zeta = -\omega + \omega$ is a sequence $\ldots a_{-2}a_{-1}a_0a_1a_2 \ldots$ 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 $a^\zeta = a^{-\omega}a^\omega$ of length ζ is the product of the two words $a^{-\omega}$ and a^ω of length $-\omega$ and ω respectively.

We now recall the notion of rational languages 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 ω usually used to construct ω-words and the ordinal iteration $\sharp$ introduced by Wojciechowski [29] 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. Given two classes X and Y of words, define

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

We denote by $A^\diamond$ the class 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 $L_1, \ldots, L_n \subseteq A^\diamond$, we define

$$\mathrm{sh}(L_1, \ldots, L_n)$$

as the class 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 $(J_1, \ldots, J_n)$ 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 $\{\varnothing\} \cup A$ with the symbols denoting the rational operations as function symbols. Each rational expression denotes a class of words which is inductively defined by the above definitions of the rational operations. A class of words is *rational* if it can be denoted by a rational expression. As usual, the dot denoting concatenation is omitted in rational expressions.

Example 1. Consider the word $w = (w_r)_{r\in\mathbb{R}}$ of length $\mathbb{R}$ over the alphabet $A = \{a, b\}$, defined by $w_r = a$ if $r \in \mathbb{Q}$, and $w_r = b$ otherwise. Then it is not difficult to check that $w \in \mathrm{sh}(a, b)$. Consider now the word $w' = (w'_q)_{q\in\mathbb{Q}}$ of length $\mathbb{Q}$ over the alphabet A, defined by $w'_q = a$ if $q \in \{m/2^n \mid m \in \mathbb{Z}, n \in \mathbb{N}\}$, and $w'_q = b$ otherwise. Here $w' \notin \mathrm{sh}(a, b)$ because $\mathbb{Q}$ is not complete, but it can be checked that $w' \in \mathrm{sh}(a, b, \varepsilon)$ (the use of ε in $\mathrm{sh}(a, b, \varepsilon)$ allows to complete $\mathbb{Q}$).

Example 2. The rational expression $a^*(\varepsilon+\mathrm{sh}(a^*, \varepsilon))a^*$ denotes the class of words (over the unary alphabet $\{a\}$) whose length is an ordering containing no infinite sequence of consecutive elements. It is clear that the length of any word denoted by this expression cannot contain an infinite sequence of consecutive elements. Conversely, let J be such an ordering. Define the equivalence relation $\sim$ on J by $x \sim y$ if and only if there are finitely many elements between x and y. The classes of $\sim$ are then finite intervals. Furthermore the ordering of these intervals must be a dense ordering with possibly a first and a last element. This completes the converse.

Example 3. The rational expression $(\varepsilon + \mathrm{sh}(a)) \diamond a$ denotes the class of words (over the unary alphabet $\{a\}$) whose length is a complete ordering. The shuffle operator is defined using complete orderings J, and the ordering $\hat{J}$ is always complete. It follows from these two facts that the length of any word denoted by this expression is complete. Conversely, let J be a complete ordering. Define the equivalence relation $\sim$ on J by $x \sim y$ if and only if there is an open dense interval containing both x and y. Each class of $\sim$ is either a singleton or an open dense interval. Let K be the ordering of the singleton classes and let L_0 be the ordering of the dense classes. Let L_1 be the ordering of pairs of consecutive elements in K and let L be $L_0 \cup L_1$ equipped with the natural ordering. It can be shown that $K = \hat{L}$. This gives the expression $(\varepsilon + \mathrm{sh}(a)) \diamond a$ where ε is due to L_1, $\mathrm{sh}(a)$ to L_0 and a to K. This completes the converse.

2.3 Automata

We recall the definition given in [4] for automata accepting words on linear orderings. As already noted in [4], this definition is actually suitable for all linear orderings.

Automata accepting words on linear orderings are classical finite automata equipped with limit transitions. They are defined as $\mathcal{A} = (Q, A, E, I, F)$, where Q denotes the finite set of states, A is a finite alphabet, and I, F denote the set of initial and final states, respectively. The set E consists in three types of transitions: the usual *successor* transitions in $Q \times A \times Q$, the *left limit* transitions which belong to $2^Q \times Q$ and the *right limit* transitions which belong to $Q \times 2^Q$. A left (respectively right) limit transition $(P, q) \in 2^Q \times Q$ (respectively, $(q, P) \in Q \times 2^Q$) will usually be denoted by $P \to q$ (respectively $q \to P$).

We sometimes write that an automaton $\mathcal{A}$ has transitions $P_1, \ldots, P_m \to q_1, \ldots, q_n$ when $\mathcal{A}$ has all left limit transitions $P_i \to q_j$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. Analogously we shall use the notation $q_1, \ldots, q_n \to P_1, \ldots, P_m$ for right limit transitions.

We now turn to the definition of a path in an automaton. Let J be a linear ordering. Observe that the ordering $\hat{J}$ always has a first element and a last element, namely the cuts $c_{\min} = (\varnothing, J)$ and $c_{\max} = (J, \varnothing)$. For any cut $c \in \hat{J}$, define the sets $\lim_{c^-} \gamma$ and $\lim_{c^+} \gamma$ as follows:

$$\lim_{c^-} \gamma = \{q \in Q \mid \forall c' < c \ \ \exists k \quad c' < k < c \text{ and } q = q_k\},$$
$$\lim_{c^+} \gamma = \{q \in Q \mid \forall c < c' \ \ \exists k \quad c < k < c' \text{ and } q = q_k\}.$$

A sequence $\gamma = (q_c)_{c \in \hat{J}}$ of states is a path for (or *labeled* by) the word $x = (a_j)_{j \in J}$ if the following conditions are fulfilled. For any pair (c_j^-, c_j^+) of consecutive cuts of J, the automaton must have the successor transition $q_{c_j^-} \xrightarrow{a_j} q_{c_j^+}$. For any cut $c \neq c_{\min}$ which has no predecessor in $\hat{J}$, $\lim_{c^-} \gamma \ \to \ q_c$ must be a left limit transition. For any cut $c \neq c_{\max}$ in $\hat{J}$ which has no successor, $q_c \to \lim_{c^+} \gamma$ must be a right limit transition.

The *content* of a path γ is the set of states which occur inside γ. It does not take account the first and the last state of the path. We denote by $q \xrightarrow{w} q'$ any path labeled by w and whose first (resp. last) element is q (resp. q').

A path is *successful* if its first state $q_{c_{\min}}$ is initial and its last state $q_{c_{\max}}$ is final. A word is *accepted* by $\mathcal{A}$ if it is the label of a successful path. The *language* $L(\mathcal{A})$ of the automaton $\mathcal{A}$ is the class of the words it accepts. A class of words is *regular* if it is the language of some automaton.

Example 4. Let $A = \{a, b\}$. The automaton $\mathcal{A}$ pictured in Fig. 1 has two successor transitions, three left limit transitions and three right limit transitions. State 0 is the only initial state, and state 5 is the only final state.

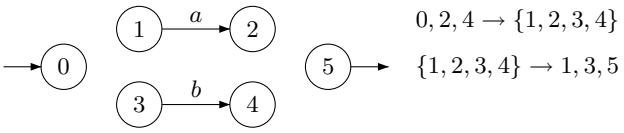

Fig. 1. An automaton $\mathcal{A}$ with $L(\mathcal{A}) = \mathrm{sh}(a, b)$.

Let us show that $L(\mathcal{A}) = \mathrm{sh}(a, b)$. Consider indeed a word $w = (w_j)_{j \in J}$, and assume first that w is accepted by $\mathcal{A}$. Let $\gamma = (q_c)_{c \in \hat{J}}$ be a successful path labeled by w. The ordering J must be dense since there are no consecutive successor transitions in $\mathcal{A}$. It must also be complete since there is no state with incoming left limit transitions and leaving right limit transitions. Occurrences of both a and b must be dense in J since all limit transitions involve the four states $\{1, 2, 3, 4\}$. Finally, J cannot have a first or a last element. Indeed, the only transition leaving state 0 is a limit one, and similarly for the only transition entering state 5.

Conversely, let $w = (w_j)_{j \in J}$ be a word indexed by a complete ordering J without first and last element, and such that occurrences of both a and b are

dense in J. Since J is complete, any cut of J (apart from $c_{\min}$ and $c_{\max}$) is either preceded or followed by a letter. Then the sequence $\gamma = (q_c)_{c \in \hat{J}}$ defined as follows is a successful path labeled by w.

- $q_{c_{\min}} = 0$,
- $q_{c_{\max}} = 5$,
- $q_c = 1$ if c is followed by an a and $q_c = 2$ if c is preceded by an a,
- $q_c = 3$ if c is followed by a b and $q_c = 4$ if c is preceded by a b.

Example 5. Let $A = \{a, b\}$. The automaton $\mathcal{A}$ pictured in Fig. 2 has two successor transitions, two left and two right limit transitions. State 0 is the only initial state, and state 2 is the only final state. Then $L(\mathcal{A}) = a^\zeta \diamond b$. Note that

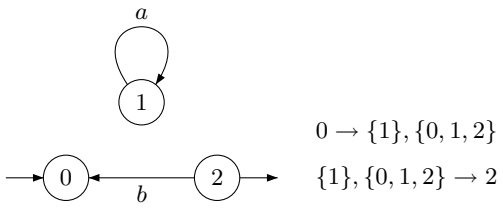

Fig. 2. An automaton $\mathcal{A}$ with $L(\mathcal{A}) = a^\zeta \diamond b$.

the automaton without the transitions $0 \to \{0,1,2\}$, $\{0,1,2\} \to 2$ and $2 \xrightarrow{b} 0$ accepts only a^ζ. Considering also the transition $2 \xrightarrow{b} 0$ it accepts $a^\zeta(ba^\zeta)^*$. Now consider the entire automaton. An accepted word w has neither a first nor a last letter. Each occurence of a^ζ in w which is not the last (resp. first) is followed (resp. preceded) by a b. Two b can not be consecutive. Moreover, if J is the linear ordering indexing the occurences of a^ζ in w, then the occurences of b are indexed by $\hat{J}^*$ and they are ordered as in $J \cup \hat{J}^*$. Reciprocally, checking that any word of $a^\zeta \diamond b$ is accepted by the automaton is just pure verification.

The following theorem was proven in [4] for the restricted case of countable scattered linear orderings; the general case is proved in [3].

Theorem 1. *A class of words over linear orderings is rational if and only if it is regular.*

2.4 Monadic Second-Order Logic

Let us 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 [28].

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 $x, y, \ldots$ interpreted as elements of the domain of the structure, monadic second-order variable symbols $X, Y, \ldots$ 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. 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 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 an MSO-formula over $\mathcal{L}$, say $\varphi(x_1, \ldots, x_m, X_1, \ldots, X_n)$, which is true in M if and only if $(x_1, \ldots, x_m, X_1, \ldots, X_n)$ is interpreted by an $(m+n)-$tuple of R.

Given a 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_ε which has an empty domain, we will allow structures to be empty. Given an MSO sentence φ over the signature $\mathcal{L}_A$, we define the language L_φ as the class 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 an MSO-sentence φ over the signature $\mathcal{L}_A$ such that $L = L_\varphi$.

Example 6. In this example we assume the axiom of choice. Let $A = \{a, b\}$ and L be the class of words w over A such that w contains a sub-sequence of a indexed by ω. Then L is the language of the following MSO-sentence φ:

$$\varphi \equiv \exists X((\exists x\ x \in X) \wedge \forall x(x \in X \Rightarrow (P_a(x) \wedge \exists y\ (y \in X \wedge x < y))))$$

If $w \in L$ then choose X to be the positions of the letters of the sub-sequence. Then X is not empty, each element of X is the index of a letter a and as X is isomorphic to ω it has no last element. Thus $M_w \models \varphi$. Conversely assume that w is such that $M_w \models \varphi$. Then w contains a non-empty ordered set X of elements all labeled by a and that contains no last element. As a consequence of the axiom of choice, an ordered sub-sequence of type ω can be extracted from X.

3 Rational languages are MSO-definable

3.1 Introduction and examples

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 ω, 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, 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 $L \subseteq A^\diamond$, 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 an 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 φ when X is interpreted as an interval of 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 a rational expression denoting L; this approach is not new, see e.g. [15] where it is used in the case of finite words. 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 interval X can be partitioned in some way in intervals. Consider for instance the case of the ω-power operation. Assume that L is definable by the MSO formula $\varphi(X)$. Then L^ω could be defined by an MSO formula which expresses 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, maximal for inclusion, which consists in elements of X_1 only, or elements of X_2 only (see Figure 3). 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 ω. This allows to find an MSO formula which defines L^ω. 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 7. Let L_1 be the class of words $w = (a_j)_{j\in J}$ (with $a_j \in A$) such that J has a first 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 $aA^\diamond bA^\diamond$. It is also MSO-definable by the sentence

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

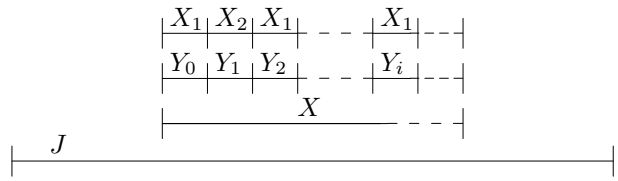

Fig. 3. Let L be a language (assume $\epsilon \notin L$), w be a word of length J, and X be an interval of J. Denote by $w[X]$ the factor of w defined by X. Then $w[X] \in L^\omega$ iff there exists a partition of X into X_1, X_2 such that $w[I] \in L$ for every maximal interval I of X_1 and of X_2, and an order isomorphic to ω is built by picking one element into such maximal intervals.

Example 8. Let L_2 be the class 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 $\mathrm{sh}(a, b, \varepsilon)$. It is MSO-definable by the sentence

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

Example 9. The language $L_3 = a^\omega a^{-\omega}$ is definable in MSO by the sentence

$$\begin{aligned} \forall x\ P_a(x) \wedge \exists X_1 \exists X_2 (\forall x\ (x \in X_1 \Leftrightarrow x \notin X_2) \\ \wedge \forall x \forall y ((x \in X_1 \wedge y \in X_2) \Rightarrow x < y) \\ \wedge \mathrm{Omega}(X_1) \wedge \mathrm{MinusOmega}(X_2)) \end{aligned}$$

where $\mathrm{Omega}(X_1)$ (respectively $\mathrm{MinusOmega}(X_2)$) expresses that the order type of X_1 is ω (respectively $-\omega$). One can show that the predicates Omega and MinusOmega are MSO-definable (see Lemma 1).

Example 10. The language L_4 of words whose length is a complete ordering can be represented by the rational expression $(\varepsilon + \mathrm{sh}(a + b)) \diamond (a + b)$. Let φ_{lub} be the following sentence, where $\varphi(x, Y)$ is an abbreviation for $\forall y \in Y\ y \leq x$.

$$\varphi_{\mathrm{lub}} \equiv \forall Y ((\exists x \varphi(x, Y)) \Longrightarrow (\exists x (\varphi(x, Y) \wedge \forall z (\varphi(z, Y) \Longrightarrow x \leq z))))$$

Then, for any word $(w_j)_{j \in \mathcal{J}}$, $M_w \models \varphi_{\mathrm{lub}}$ if and only if every non-empty subordering Y of J which is bounded above has a least upper bound in J. Similarly to φ_{lub}, a sentence φ_{glb} can be written to express that every non-empty subordering Y of J which is bounded below has a greatest lower bound in J. Thus $M_w \models \varphi_{\mathrm{glb}} \wedge \varphi_{\mathrm{lub}}$ if and only if the length of w is a complete ordering.

Example 11. 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 $\mathrm{sh}(A^\diamond(a+b)A^\diamond, \varepsilon)$. The language L_5 can also be defined by the following MSO formula

$$\exists X(\exists x_1 \in X\ \exists x_2 \in X\ x_1 < x_2 \\ \wedge \forall y_1 \in X\ \forall y_2 \in X(y_1 < y_2 \Longrightarrow \exists z\ z \in X \wedge (y_1 < z \wedge z < y_2))).$$

3.2 Proof of Proposition 1

We shall prove that for every rational language L there exists an MSO formula $\varphi(X)$ in the language $\mathcal{L}_A$ such that for every word $w = (w_j)_{j\in J}$ where $w_j \in A$, then $w \in L$ if and only if M_w satisfies φ when X is interpreted by J.

This will yield 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 a rational expression denoting L.

The following lemma provides auxiliary predicates which will be useful later.

Lemma 1. *Let $L = \{<\}$ be a language where $<$ is interpreted as a linear ordering over the domain of the structure. The following relations are MSO-definable in the language $\mathcal{L}_A$.*

- *The relations $x = y$, $x \leq y$, $x \in [y; z]$, $X = \varnothing$, $X \subseteq Y$, $x \in Y \cap Z$, $x \in Y \setminus Z$;*
- *"x, y are consecutive elements of Z", denoted by* $\mathrm{Consec}(x, y, Z)$
- *"X is an interval", denoted by* $\mathrm{Interval}(X)$
- *"X is maximal among the intervals contained in Y", denoted by* $X \subseteq_{max} Y$
- *"X is an ordinal", denoted by* $\mathrm{Ord}(X)$
- *"X is an ordinal less than or equal to ω", denoted by* $\mathrm{Ord}_{\leq\omega}(X)$
- *"X is a finite ordinal", denoted by* $\mathrm{Ord}_{fin}(X)$
- *"X equals ω", denoted by* $\mathrm{Omega}(X)$
- *"T is contained in X, and every maximal interval of X contains exactly one element of T", denoted by* $\mathrm{Trace}(X, T)$
- *"X is complete", denoted by* $\mathrm{Complete}(X)$
- *Given $n \geq 1$, the relation "$X_1, \ldots, X_n$ form a partition of X", denoted by* $\mathrm{Partition}(X, X_1, \ldots, X_n)$. *We shall "overload" the predicate symbol* Partition *and use it with several values of n.*
- *"X is the sum of Y and Z" denoted by $X = Y + Z$.*

Proof. We give below the formal definitions for each relation, except for the relations of the first item.

- $x \in [y; z] :\ y \leq x \leq z \vee z \leq x \leq y$
- $\mathrm{Consec}(x, y, Z) :\ x \in Z \wedge y \in Z \wedge x < y \wedge \neg\exists z \in Z(x < z \wedge z < y)$
- $\mathrm{Interval}(X) :\ \forall x, x' \in X(\forall y\ x \leq y \leq x' \Longrightarrow y \in X)$
- $X \subseteq_{max} Y :$

$$\mathrm{Interval}(X) \wedge X \subseteq Y \wedge X \neq \varnothing \wedge \forall y \in (Y \setminus X)\ \forall x \in X\ \exists z \notin Y\ z \in [x; y]$$

- $\mathrm{Ord}(X) :\ \forall Y \subseteq X\ (Y \neq \varnothing \Longrightarrow \exists y \in Y\ \forall y' \in Y\ y \leq y')$

– $\mathrm{Ord}_{\leq\omega}(X)$:

$\mathrm{Ord}(X) \wedge \forall x \in X((\forall y \in X\ x \leq y) \vee \exists z \in X\ (z < x \wedge \forall y \in X\ (y < x \Longrightarrow y \leq z))$

(i.e., every $x \in X$ is either the first element of X or has a predecessor in X)
– $\mathrm{Ord}_{fin}(X)$: $\mathrm{Ord}_{\leq\omega}(X) \wedge \exists y \in X \forall z \in X(z \leq y)$
– $\mathrm{Omega}(X)$: $\mathrm{Ord}_{\leq\omega}(X) \wedge \neg\, \mathrm{Ord}_{fin}(X)$
– $\mathrm{Trace}(X, T)$: $T \subseteq X \wedge (\forall x \in X\ \exists! t \in T\ \forall x' \in [x; t]\ x' \in X)$
– $\mathrm{Complete}(X)$:

$\forall Y \subseteq X((\exists x \in X\ \varphi(x, Y)) \Longrightarrow (\exists x \in X(\varphi(x, Y) \wedge \forall z \in X(\varphi(z, Y) \Longrightarrow x \leq z))))$

where $\varphi(x, Y)$ is an abbreviation for $\forall y \in Y\ y \leq x$
– $\mathrm{Partition}(X, X_1, \ldots, X_n)$:

$$\bigwedge_{1 \leq i \leq n} X_i \subseteq X \wedge \bigwedge_{1 \leq i < j \leq n} X_i \cap X_j = \varnothing \wedge \forall x \in X \bigvee_{1 \leq i \leq n} x \in X_i$$

– $X = Y + Z$: $\mathrm{Partition}(X, Y, Z) \wedge \forall y \forall z((y \in Y \wedge z \in Z) \Rightarrow y < z)$.

We now start the proof of Proposition 1, by induction on a rational expression denoting L. Note that $L = \varnothing$ is obviously MSO-definable.

Empty word. The language $L = \{\varepsilon\}$ is defined by the formula

$$\varphi(X) : \neg \exists x \in X$$

Singletons. For every $a \in A$, the language $\{a\}$ is defined by the formula

$$\varphi(X) : \exists x \in X(P_a(x) \wedge \forall x' \in X\ x' = x)$$

Product operation. Assume that L_1, L_2 are languages defined by the MSO formulas $\varphi_1(X), \varphi_2(X)$, respectively. Then the language $L_1 \cdot L_2$ is defined by the formula

$$\varphi(X) : \exists X_1 \exists X_2 (X = X_1 + X_2 \wedge \varphi_1(X_1) \wedge \varphi_2(X_2)).$$

Star operation. Assume that L_1 is defined by the MSO formula $\psi(X)$. Let us find an MSO formula which defines the language $L = L_1^*$.

Consider first the case where $\varepsilon \notin L_1$ and $L_1 \neq \varnothing$. Then by definition a word w belongs to L if and only if it can be written as $w = \prod_{i \in I} w_i$ where I is finite and all w_i's are non-empty words which belong to L_1. The MSO-formula which defines L expresses this property as the existence a partition of X into two subsets X_1, X_2 such that each interval I of X which is maximal among subsets of X_1 (resp. X_2) corresponds to one of the subwords w_i of w. Moreover, in order to express that there are finitely many such maximal intervals, the formula states the existence of two subsets T_1, T_2 of X such that every maximal interval $I \subseteq X_1$ (resp. $I \subseteq X_2$) contains exactly one element of T_1 (resp. T_2), and

such that $T_1 \cup T_2$ is finite. It is easy to check that these properties are equivalent to $w \in L_1^*$.

This leads to define the language L by the formula

$$\begin{aligned}\varphi(X):\ & \exists X_1 \exists X_2 \exists T_1 \exists T_2 \\ & (\mathrm{Partition}(X, X_1, X_2) \wedge \mathrm{Trace}(X_1, T_1) \wedge \mathrm{Trace}(X_2, T_2) \wedge \mathrm{Ord}_{fin}(T_1 \cup T_2) \\ & \wedge \forall U((U \subseteq_{max} X_1 \vee U \subseteq_{max} X_2) \Rightarrow \psi(U)).\end{aligned}$$

The case $L = \varnothing$ is trivial, and the case where $\varepsilon \in L$ can be handled by considering the formula $\varphi'(X):\ \varphi(X) \vee \neg \exists x \in X$.

Power operations. The language L^ω (respectively $L^\sharp$) is defined using the same ideas as above, except that the formula $\mathrm{Ord}_{fin}(T_1 \cup T_2)$ has to be replaced by $\mathrm{Omega}(T_1 \cup T_2)$ (respectively $\mathrm{Ord}(T_1 \cup T_2)$). The cases of $L^{-\omega}$ and $L^{-\sharp}$ are similar.

Diamond operation. Assume that L_1, L_2 are languages defined by the MSO formulas $\varphi_1(X), \varphi_2(X)$, respectively. We shall find a formula that defines the language $L = L_1 \diamond L_2$. We have to consider several cases depending on whether ε belongs to L_1 and L_2.

First case: $\varepsilon \notin L_1 \cup L_2$. In this case we use the following lemma from [3], which gives necessary and sufficient conditions which ensure that a partition (J, J') of an ordering K satisfies $J' = \hat{J}^*$. We recall the proof for the convenience of the reader.

Lemma 2. *Let K be a complete linear ordering, and let (J, J') be a partition of K. Assume that if K has a first (resp. last) element then it belongs to J. Suppose that any non-first and non-last element of J has a predecessor and successor in J', that the first (resp. last) element of J, if exists, has a successor (resp. predecessor) in J', and that there is at least one element of J between two elements of J'. Then J' equals $\hat{J}^*$, that is $K = J \cup \hat{J}^*$.*

Note that one checks easily that the converse of Lemma 2 holds.

Proof. We define a function f from K into $J \cup \hat{J}^*$ as follows. For any $k \in K$, define

$$f(k) = \begin{cases} k & \text{if } k \in J \\ \big(\{j \in J \mid j < k\}, \{j \in J \mid k < j\}\big) & \text{if } k \in J'. \end{cases}$$

Since $J \cap J' = \varnothing$ and $K = J \cup J'$, the function f is well defined. The restriction of f to J is the identity. The image of an element of J' is a cut of J. Therefore f is a function from K into $J \cup \hat{J}^*$.

We claim that the function f is one-to-one. We first show that $k \neq k'$ implies $f(k) \neq f(k')$. If $k \in J$ or $k' \in J$, the result is trivial. Suppose then that $k, k' \in J'$ and that $k < k'$. By our second hypothesis there exists $j' \in J$ such that $k < j' < k'$, which implies $\{j \in J \mid j < k\} \neq (\{j \in J \mid j < k'\}$, thus $f(k) \neq f(k')$.

We now prove that the function f is onto. It is clear that $J \subseteq f(K)$. Let $(L, M) \in \hat{J}^*$. We claim that there is $k \in J'$ such that $(L, M) = f(k)$. Define the two elements l and m of K by $l = \sup(L)$ and $m = \inf(M)$. If l belongs to L, it has a successor k in J' and one has $(L, M) = f(k)$. If m belongs to M, it has a predecessor k in J' and one has $(L, M) = f(k)$. If l and m do not belong to L and M, they belong to J' and their image by f is the cut (L, M). Since f is one-to-one, l and m are equal.

Lemma 2 allows to define the language $L_1 \diamond L_2$ in MSO logic, by expressing that there exists a partition of X into two subsets X_1, X_2 such that the set of maximal intervals of X_1 or X_2 has an underlying ordering of the form $J \cup \hat{J}^*$, where J (resp. $\hat{J}^*$) is the ordering of maximal intervals of X_1 (resp. X_2), and each maximal interval of X_1 (resp. X_2) corresponds to a subword of w which belongs to L_1 (resp. L_2).

Let us give a formal definition:

$$\varphi(X) : \ \exists X_1 \exists X_2 \exists T_1 \exists T_2 \ (\text{Complete}(T_1 \cup T_2) \tag{1}$$
$$\wedge \text{Partition}(X, X_1, X_2) \wedge \text{Trace}(X_1, T_1) \wedge \text{Trace}(X_2, T_2) \tag{2}$$
$$\wedge \forall U \subseteq_{max} X_1 \ \ \varphi_1(U) \tag{3}$$
$$\wedge \forall U \subseteq_{max} X_2 \ \ \varphi_2(U) \tag{4}$$
$$\wedge \forall t \in T_1 \ (\exists u \in T_1 \cup T_2 \ u < t) \Rightarrow (\exists u \in T_2 \ \ \text{Consec}(u, t, T_1 \cup T_2)) \tag{5}$$
$$\wedge \forall t \in T_1 \ (\exists u \in T_1 \cup T_2 \ t < u) \Rightarrow (\exists u \in T_2 \ \ \text{Consec}(t, u, T_1 \cup T_2)) \tag{6}$$
$$\wedge \forall u_1 \in T_2 \ \forall u_2 \in T_2 \ (u_1 < u_2 \Rightarrow \exists t \in T_1 \ (u_1 < t < u_2))) \tag{7}$$
$$\wedge \forall t \in (T_1 \cup T_2)((\forall u \in (T_1 \cup T_2) \ t \leq u) \Rightarrow t \in T_1) \tag{8}$$
$$\wedge \forall t \in (T_1 \cup T_2)((\forall u \in (T_1 \cup T_2) \ u \leq t) \Rightarrow t \in T_1)) \tag{9}$$

Lines (5), (6) and (7) express the conditions of Lemma 2 for the orderings T_1 and T_2, while line (8) (respectively (9)) expresses that if $T_1 \cup T_2$ has a first (resp. last) element then it belongs to T_1; this allows to take into account the fact that we deal with an ordering of the form $J \cup \hat{J}^*$ and not $J \cup \hat{J}$.

Second case: $\varepsilon \in L_1$ and $\varepsilon \notin L_2$. This case can be handled by proving a variant of Lemma 2. Indeed we have $w \in L_1 \diamond L_2$ if and only if w satisfies the following condition, which we denote by (C_1): w can be written as $w = \prod_{k \in K} w_k$ where K is a complete ordering and there exists a partition of K in two subsets J_1, J_2 such that

- $w_k \neq \varepsilon$ for every $k \in K$;
- $w_k \in L_1$ if $k \in J_1$, and $w_k \in L_2$ if $k \in J_2$;
- every element of J_1 which is not the first (resp. last) element of K admits a predecessor (resp. successor) which belongs to J_2;
- K does not contain any dense interval consisting only in elements of J_2.

Indeed assume first that $w \in L_1 \diamond L_2$. By definition w can be written as $w = \prod_{j \in J \cup \hat{J}^*} w_j$; where $w_j \in L_1$ if $j \in J$ and $w_j \in L_2$ if $j \in \hat{J}^*$. If one removes from $J \cup \hat{J}^*$ all elements $j \in J$ such that $w_j = \varepsilon$, one can re-write w as

$w = \prod_{j \in J_1 \cup J_2} w_j$ where $J_2 = \hat{J}^*$, and J_1 corresponds to the remaining elements of J. Let us prove that the set $K = J_1 \cup J_2$ and the partition (J_1, J_2) satisfy (C_1). It is easy to check that K is complete, and that every element of K which belongs to J_1 has a predecessor and a successor which belong to J_2. Moreover let us show that every dense interval I of K contains at least an element of J_1. Let x, y be two elements of I such that $x < y$. If x or y belong to J_1 then the result follows. Now if both x, y belong to $J_2 = \hat{J}^*$, then x and y correspond to different cuts of J, which implies that there exists in $J \cup \hat{J}^*$ an element $z \in J$ between x and y. Let us prove that $z \in J_1$, which will yield the result since $z \in I$. Assume for a contradiction that $z \notin J_1$. The element z admits in $J \cup \hat{J}^*$ a predecessor x' and a successor y' which both belong to $\hat{J}^*$, that is to J_2. In this case x' and y', seen as elements of $K = J_1 \cup J_2$, become consecutive elements, which contradict the fact that I is a dense interval of K.

Conversely assume that $w = \prod_{k \in K} w_k$, where K is a complete ordering and there exists a partition of K in two subsets J_1, J_2 satisfying condition (C_1). Let us add to K a new element of J_1 between every pair of elements of J_2 which are consecutive in K, and also (if necessary) elements of J_1 at both extremities of K. This gives rise to a set K' and a partition (J_1', J_2') of K' such that J_1' corresponds to the union of J_1 and the new elements, and $J_2' = J_2$. If we associate to every new element the empty word, we can re-write w as $w = \prod_{k \in K'} w_k$ where K' is complete, $w_k \in L_1$ if $k \in J_1'$, and $w_k \in L_2$ if $k \in J_2'$. Let us prove that $J_2' = \hat{J'}_1^*$. By Lemma 2 it suffices to prove, on one hand, that every element of J_1', apart from the first (resp. last) element of K if it exists, admits a predecessor (resp. successor) in J_2', and on the other hand that between two elements of J_2' there exists at least an element of J_1'. The first fact follows easily from the construction of J_1'. For the second fact, assume for a contradiction that there exist two elements $x, y \in J_2'$ with $x < y$ such that the interval $[x, y]$ does not contain any element of J_1'. By our hypothesis on $J_2 = J_2'$, the interval $[x, y]$, seen as an interval of K, cannot be dense, thus it contains two consecutive elements $x' < y'$ which both belong to J_2. This implies by definition of J_1' that there exists (in K') an element of J_1' between x' and y', that is between x and y, which leads to a contradiction.

We proved that $w \in L_1 \diamond L_2$ if and only if w satisfies (C_1). It remains to prove that (C_1) can be expressed with a MSO-formula, which can be done easily using ideas similar to the first case.

Third case: $\varepsilon \notin L_1$ and $\varepsilon \in L_2$. This case is similar to the previous one. In this case we have $w \in L_1 \diamond L_2$ if and only if w satisfies the following conditions, which we denote by (C_2): w can be written as $w = \prod_{k \in K} w_k$ where K is any ordering and there exists a partition of K in two subsets J_1, J_2 such that

- $w_k \neq \varepsilon$ for every $k \in K$;
- $w_k \in L_1$ if $k \in J_1$, and $w_k \in L_2$ if $k \in J_2$;
- between two elements of J_2 there exists at least an element of J_1;
- if K has a first (resp. last) element then it belongs to J_1.

Indeed assume first that $w \in L_1 \diamond L_2$. By definition w can be written as $w = \prod_{j \in J \cup \hat{J}^*} w_j$ where $w_j \in L_1$ if $j \in J$ and $w_j \in L_2$ if $j \in \hat{J}^*$. If one

removes from $J \cup \hat{J}^*$ all elements $j \in \hat{J}^*$ such that $w_j = \varepsilon$, one can re-write w as $w = \prod_{j \in J_1 \cup J_2} w_j$ where $J_1 = J$ and J_2 corresponds to the remaining elements of $\hat{J}^*$. It is easy to check that J_1 and J_2 satisfy the properties required in (C_2).

Conversely assume that $w = \prod_{k \in K} w_k$, where K is a complete ordering and there exists a partition of K into two subsets J_1, J_2 satisfying condition (C_2). We shall add elements to J_2 in order to fill gaps in $J_1 \cup J_2$. For each cut (K_1, K_2) of $J_1 \cup J_2$ such that K_1 does not admit a last element in J_2 and K_2 does not admit a first element in J_2, we add to K a new element of J_2 between K_1 and K_2. This gives rise to a set K' and a partition (J_1', J_2') of K' such that $J_1' = J_1$, and J_2' corresponds to the union of J_2 and the new elements. If we associate to every new element the empty word, we can re-write w as $w = \prod_{k \in K'} w_k$ where K' is complete, $w_k \in L_1$ if $k \in J_1'$, and $w_k \in L_2$ if $k \in J_2'$. It follows from the construction of J_2' that $J_2' = \hat{J'}_1^*$. This proves that $w \in L_1 \diamond L_2$.

We proved that $w \in L_1 \diamond L_2$ if and only if w satisfies (C_2). It remains to prove that (C_2) can be expressed with a MSO-formula, which again can be done easily using ideas similar to the first case.

Fourth case: $\varepsilon \in L_1 \cap L_2$. In this case we have $w \in L_1 \diamond L_2$ if and only if w satisfies the following conditions, which we denote by (C_3): w can be written as $w = \prod_{k \in K} w_k$ where K is any ordering and w_k is a non empty word which belongs to $L_1 \cup L_2$ for every $k \in K$.

As in the previous cases, it is easy to check that if $w \in L_1 \diamond L_2$ then w satisfies (C_3). Conversely assume that $w = \prod_{k \in K} w_k$ for some ordering K. Let (J_1, J_2) be the partition of K such that $w_k \in J_1$ if and only if $k \in L_1$. We proceed in a similar way as in the two previous cases. First, as in the second case, we add a new element of J_1 between every pair of elements of J_2 which are consecutive in K, and also (if necessary) elements of J_1 at both extremities of K. Then, as in the third case, for each cut (K_1, K_2) of $J_1 \cup J_2$ such that K_1 does not admit a last element in J_2 and K_2 does not admit a first element in J_2, we add a new element of J_2 between K_1 and K_2. These two steps give rise to a set K' and a partition (J_1', J_2') of K' such that J_1' corresponds to the union of J_1 and the new elements added during the first step, and J_2' corresponds to the union of J_2 and the new elements added during the second step. If we associate to all new elements the empty word, we can re-write w as $w = \prod_{k \in K'} w_k$ where K' is complete, $w_k \in L_1$ if $k \in J_1'$, and $w_k \in L_2$ if $k \in J_2'$. Using ideas from the previous cases one can check that $J_2' = \hat{J'}_1^*$. This proves that $w \in L_1 \diamond L_2$.

It is easy to express condition (C_3) in MSO.

This completes the proof that $L_1 \diamond L_2$ is definable in MSO.

Shuffle operation. Let $n \geq 1$, and assume that $L_1, \ldots, L_n$ are languages defined by the MSO formulas $\varphi_1(X), \ldots, \varphi_n(X)$, respectively. We shall find a formula that defines the language $L = \mathrm{sh}(L_1, \ldots, L_n)$. We have to consider again several cases depending on whether ε belongs to some of the L_i's.

First case: $\varepsilon \notin \bigcup_i L_i$. In this case the language L can be defined by a formula which expresses in a direct way the definition of the shuffle operation. Here is

such a formula:

$$\varphi(X):\ \exists X_1 \dots \exists X_n \exists T_1 \dots \exists T_n \exists T$$
$$(\mathrm{Partition}(X, X_1, \dots, X_n) \wedge \bigwedge_{1\leq i\leq n} \mathrm{Trace}(X_i, T_i)\ \wedge T = \bigcup_{1\leq i\leq n} T_i \wedge \mathrm{Complete}(T)$$
$$\wedge\ \forall x \in T\ (\exists y_1 \in T\ y_1 < x \wedge \exists y_2 \in T\ x < y_2)$$
$$\wedge\ \forall x, y \in T\ (x < y \Longrightarrow \bigwedge_{1\leq i\leq n} (\exists z \in T_i\ x < z < y))$$
$$\wedge \bigwedge_{1\leq i\leq n} (\forall U \subseteq_{max} X_i\ \ \varphi_i(U))\ )$$

Second case: there exists i such that $\varepsilon \in L_i$. In the sequel assume without loss of generality that there exists k such that for every i we have $\varepsilon \in L_i$ if and only if $i \leq k$.

We need the following lemma (see e.g. [24]). We denote by $|Z|$ the cardinality of Z.

Lemma 3. *(Assuming the Axiom of Choice) Let X be a non-empty dense set and let $n \geq 1$ be a natural number. There exists a partition of X into n subsets $X_1, \dots, X_n$ which are dense in X.*

Proof. It suffices to prove the case $n = 2$.

Consider the binary relation $\equiv$ defined on X as follows: given $x, y \in X$, $x \equiv y$ if and only if

- $x = y$, or
- $x \neq y$ and there exist $x', y' \in X$ such that $x' < [x; y] < y'$ and for every $a, b \in]x'; y'[$ with $a < b$ we have $|[a; b]| = |[x; y]|$.

It is easy to check that $\equiv$ is a condensation, i.e. an equivalence relation whose equivalence classes are intervals of X. We shall choose, in each equivalence class of $\equiv$, which elements belong to X_1 (respectively X_2).

Given an equivalence class $x/\equiv$ which contains more than one element, all intervals $[a, b]$ such that $a < b$ and a, b belong to $x/\equiv$ have (by definition of $\equiv$) the same cardinality, say λ. The cardinal λ is infinite since X is dense; moreover we have $\lambda \leq |x/\equiv|$.

Assume first that $\lambda = |x/\equiv|$. Consider an enumeration $\{(a_i, b_i) : i < \lambda\}$ of pairs of elements of $x/\equiv$ such that $a_i < b_i$. We define by induction on i two sequences $(x_i^1)_{i<\lambda}$ and $(x_i^2)_{i<\lambda}$ as follows: set $x_0^1 = a_0$ and $x_0^2 = b_0$. Then assuming that x_i^1 and x_i^2 are defined for every $i < j$, choose for x_j^1 any element of the set $[a_j, b_j] \setminus (\{x_i^1 : i < j\} \cup \{x_i^2 : i < j\})$, and for x_j^2 any element of the set $[a_j, b_j] \setminus (\{x_i^1 : i \leq j\} \cup \{x_i^2 : i < j\})$. This is always possible since $|[a_j, b_j]| = \lambda$ by the very definition of $\equiv$, and λ is infinite. Finally we add elements of the sequence $(x_i^1)_{i<\lambda}$ (resp. $(x_i^2)_{i<\lambda}$) to the set X_1 (resp. X_2). It is clear that $X_1 \cap x/\equiv$ and $X_2 \cap x/\equiv$ are dense in $x/\equiv$.

Consider now the case $\lambda < |x/\equiv|$. Then it is not difficult to prove that there exists a partition of $|x/\equiv|$ into (λ^+) intervals of cardinal λ, and we can apply again the above construction to each such interval.

Finally we add to X_1 all elements of X which were not included into $X_1 \cup X_2$ up to now (in particular, all $x \in X$ such that $|x/\equiv| = 1$).

Let us show that X_1 and X_2 are dense in X. Consider $a, b \in X$ such that $a < b$. Let γ be the least cardinal of an interval $[a', b'] \subseteq [a, b]$ with $a' < b'$. The fact that X is dense ensures that such an interval exists and that γ is infinite. Any non-singleton interval included in $[a', b']$ has cardinality γ, which implies that $a' \equiv b'$, and the previous construction ensures that $[a', b']$ has a non-empty intersection with X_1 and X_2.

The following lemma gives a necessary and sufficient condition which ensures that $w \in \mathrm{sh}(L_1, \ldots, L_n)$ in case at least one of the L_i's contain the empty word. We define the *completion* of an ordering Z, and denote by $\overline{Z}$, the minimal complete ordering which contains Z.

Lemma 4. *For every word $w \in A^\diamond$, we have $w \in \mathrm{sh}(L_1, \ldots, L_n)$ if and only if there exists an ordering J and a partition $J_1, \ldots, J_n$ of J such that w can be written as $w = \prod_{j \in J} w_j$ where*

- *J has neither first nor last element;*
- *for every $j \in J$, $j \in J_i$ if and only if $w_j \in L_i$;*
- *for every $i > k$, the set J_i is dense in J;*
- *for every interval I of J such that $I \cap J_i = \varnothing$ for some $i \leq k$, the set of gaps of I is dense in the completion of I.*

Proof. Assume first that w satisfies the hypotheses of Lemma 4. We add elements to J in such a way that the resulting ordering satisfies the conditions given in the definition of the shuffle. More precisely consider the completion $J' = \overline{J}$. We prove that there exist a partition $J'_1, \ldots, J'_n$ of J such that

- $w = \prod_{j \in J'} w'_j$ where $w'_j = w_j$ for every $j \in J$, and $w'_j = \varepsilon$ for every $j \in J' \setminus J$;
- $J_i \subseteq J'_i$ for every $i \leq k$, and $J'_i = J_i$ for every $i > k$;
- J'_i is dense in J' for every i;
- $w'_j \in L_i$ if and only if $j \in J'_i$.

We start by setting $J'_i = J_i$ for every i. We shall add each element of $J' \setminus J$ to one of the sets $J'_1, \ldots, J'_k$. Consider the binary relation $\sim$ defined on J as follows: given $x, y \in J$, $x \sim y$ if and only if

- $x = y$, or
- $x \neq y$ and
  1. either every J_i is dense in the interval $[x; y]$
  2. or for every interval $I \subseteq [x; y]$ there exists i such that $J_i \cap I$ is not dense in I

One checks easily that $\sim$ is a condensation. Let $I \subseteq J$ be an equivalence class of $\sim$ with more than one element. If I arises from case (1) above, then we choose to add every gap x of I to the subset J'_1, and set $w'_x = \varepsilon$. If I arises from case (2) above, then by definition of $\sim$ and the hypotheses of Lemma 4, the set of gaps of I is dense in its completion. Thus by Lemma 3 there exists a partition of $(\overline{I} \setminus I)$ into n dense subsets $X_1, \ldots, X_n$. We choose to add every gap x of I to the subset J'_i such that $x \in X_i$, and we set $w'_x = \varepsilon$. One checks that the sets $J'_1, \ldots, J'_n$ satisfy the required properties, which ensures that $w \in \mathrm{sh}(L_1, \ldots, L_n)$.

Conversely if $w \in \mathrm{sh}(L_1, \ldots, L_n)$ then w can be written as $w = \prod_{j \in J} w_j$ where J satisfies the conditions required in the definition of the shuffle operation. By removing from J all elements j such that $w_j = \varepsilon$, one can re-write w as $w = \prod_{j \in J'} w_j$ where $J' \subseteq J$ and $w_j \neq \varepsilon$. Now if we set $J'_i = J_i \cap J'$ for every i, then one can check that the set J' and the partition $(J'_1, \ldots, J'_n)$ of J' satisfy the hypotheses of the lemma.

In order to complete the proof that $\mathrm{sh}(L_1, \ldots, L_n)$ is MSO-definable, it suffices to prove that there exists a MSO-formula which expresses the properties required in the statement of the above lemma. Since this formula is a variant of the one given in the first case, we leave this task to the reader.

This concludes the proof of Proposition 1.

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.*

Proof. Assume that $A = \{a_1, \ldots, a_n\}$, and let L_1, L_2 be two rational languages of words over A indexed by countable orderings. By Proposition 1, one can construct effectively from L_1 and L_2 two MSO-formulas $\varphi_1(X)$ and $\varphi_2(X)$ which respectively define L_1 and L_2.

Consider the MSO-sentence ψ defined as

$$\forall X \forall X_1 \ldots \forall X_n((\mathrm{Partition}(X, X_1, \ldots, X_n) \wedge \varphi'_1(X)) \Rightarrow \varphi'_2(X))$$

where $\varphi'_1(X)$ (resp. $\varphi'_2(X)$) is obtained from $\varphi_1(X)$ (resp. $\varphi_2(X)$) by replacing all atomic formulas of the form $P_{a_i}(x)$ by $x \in X_i$.

It is easy to check that $L_1 \subseteq L_2$ if and only if ψ is true in the monadic second order theory of countable linear orderings. Now by Rabin [19] this theory is decidable, from which the result follows.

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

4 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.*

Proof. We give a sketch of the proof. The "only if" part comes from Proposition 1, and the "if" part is a direct adaptation of Büchi's proof [8], which goes by induction on a MSO-formula (in prenex form) defining L. The crucial argument here is that by [21] the class of rational languages of words on countable scattered linear orderings is closed under complementation. Let us recall quickly the main arguments of the proof. We refer e.g. to [25] for more explanation.

First of all, we have to extend our notion of definable language to the case of MSO formulas with free variables. Assume that $\varphi(X_1, \ldots, X_m, x_1, \ldots, x_n)$ is an MSO formula whose free variables are $X_1, \ldots, X_m, x_1, \ldots, x_n$ (the case where the free variables are only first-order variables, or only second-order variables, are handled similarly).

Then we associate with φ the language L_φ defined as the class of words w over the alphabet $\{0,1\}^{m+n} \times A$ such that:

- for every $j \in [m+1, n]$, there is exactly one symbol from w such that its $j-$th component equals 1;
- $M_w \models \varphi(X_1, \ldots, X_m, x_1, \ldots, x_n)$ where
  - every X_i is interpreted as the set of positions in w carrying a letter $a \in \{0,1\}^{m+n} \times A$ whose $i-$th component equals 1.
  - every x_j is interpreted as the only position in w carrying a letter whose $j-$th component equals 1.
  - $P_a(x)$ is interpreted as "the position x carries a symbol whose last component equals a"

With this definition, we can prove that every MSO formula φ defines a rational language by induction on the construction of φ, which we can assume to be in prenex form. It is easy to check that atomic formulas $P_a(x)$, $x < y$, $x \in X$ define rational languages. The case $\varphi = \varphi_1 \vee \varphi_2$ can be solved using the fact that the class of rational languages is closed under union and cylindrification. The case $\varphi = \neg\varphi'$ is handled thanks to the fact that by [21] the class of rational languages of words on countable scattered linear orderings is closed under complementation. Finally the closure of rational languages under projection (respectively projection and intersection) allows to deal with the case $\varphi = \exists X\ \varphi'$ (resp. $\varphi = \exists x\ \varphi'$).

The effectiveness of the previous 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, i.e. the complement of the language L_5 of Example 11. Since L_5 is definable in MSO, the same holds for S_A. However the following holds.

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

Proof. By Theorem 1 it suffices to prove that S_A is not regular.

Let us introduce some useful vocabulary. An automaton $\mathcal{A}$ is said to be *trim* if and only if every state and every transition of $\mathcal{A}$ appears in at least one successful path. If an automaton is not trim, it can easily be trimmed by removing any state and any transition which does not appear in a successful path. The class of words recognized by the automaton is of course not changed by this operation.

By a slight abuse of language, a word is called *scattered* if its length is a scattered ordering.

The proof follows directly from the following two lemmas.

Lemma 5. *Let $\mathcal{A}$ be a trim automaton such that there exist a left limit transition $P \to p$ and a right limit transition $q \to P$ with $p, q \in P$. Then the automaton accepts a non scattered word.*

Proof. Since the automaton is trim, there is a path $i \xrightarrow{u} q$ from an initial state i to q. There is also a path $p \xrightarrow{w} f$ from p to a final state. There is also a path $p \xrightarrow{v} q$ whose content is exactly P. Then it is clear that the word $x = uv^{\mathbb{R}}w$ is accepted by the automaton.

Example 12. The automaton of Example 5 is trim. It accepts $(a^{\zeta}b)^{\mathbb{R}}$, which is not scattered.

Let $(x_n)_{n\geq 0}$ be the sequence of words defined by induction by $x_0 = a$ and $x_{n+1} = x_n^{\zeta}$. The length of the word x_n is the ordering ζ^n. Note that each word x_n is scattered.

Lemma 6. *If an automaton with m states accepts x_{2m+3}, then it also accepts a non scattered word.*

Proof. By the previous lemma, it suffices to prove that there are two limit transitions $P \to p$ and $q \to P$ with $p, q \in P$.

The ordering ζ^n is the ordering $J_n = \mathbb{Z}^n$ of all the n-tuples $(i_1, \ldots, i_n)$ of relative integers with the lexicographic ordering. This means that $(i_1, \ldots, i_n) < (j_1, \ldots, j_n)$ if $i_k < j_k$ where k is the least integer such that $i_k \neq j_k$. We say that an r-tuple $(i_1, \ldots, i_r)$ is a *prefix* of a s-tuple $(j_1, \ldots, j_s)$ if $r \leq s$ and $i_k = j_k$ for any $1 \leq k \leq r$.

We first give an explicit description of the ordering $\hat{J}_n^*$ of non-trivial cuts of the ordering J. Let us denote by $\mathbb{Z} + \frac{1}{2}$ the set $\{n + \frac{1}{2} \mid n \in \mathbb{Z}\}$ and let K_n be defined by

$$K_n = \{(i_1, \ldots, i_r) \mid 1 \leq r \leq n,\ \ i_1, \ldots, i_{r-1} \in \mathbb{Z} \text{ and } i_r \in \mathbb{Z} + \frac{1}{2}\}.$$

The fact that the last element of a tuple in K_n is in $\mathbb{Z} + \frac{1}{2}$ makes K_n disjoint from J_n and prevents two tuples of K_n from being prefix of each other. The set K_n is endowed with the lexicographic ordering. The relation $(i_1, \ldots, i_r) < (j_1, \ldots, j_s)$ holds if $i_k < j_k$ where k is the least integer such that $i_k \neq j_k$. Note that this integer k always exists since $i_r, j_s \in \mathbb{Z} + \frac{1}{2}$. The orderings of J_n and K_n can be extended to an ordering of $J_n \cup K_n$.

We claim that K_n with this ordering is isomorphic to the ordering $\hat{J}_n^*$. It is clear that each element k of K_n defines the cut (K, L) of J_n where $K = \{j \in J_n \mid j < k\}$ and $L = \{j \in J_n \mid k < j\}$. It is easy to see that any cut of J_n is of this form.

Let $\mathcal{A}$ be an automaton with m states which accepts the word x_{2m+3}. We set $n = 2m+3$. Let γ be an accepting path labeled by x_n. This path γ is a function from K_n to the state set Q of $\mathcal{A}$. Let L_n be the set

$$L_n = \{(i_1, \ldots, i_r) \mid 1 \leq r \leq n, \text{ and } i_1, \ldots, i_r \in \mathbb{Z}\}.$$

We define a function Γ from L_n to the power set 2^Q of Q as follows.

$$\Gamma(i_1, \ldots, i_r) = \{\gamma(j_1, \ldots, j_s) \mid (i_1, \ldots, i_r) \text{ is a prefix of } (j_1, \ldots, j_s)\}.$$

It follows from the definition that if $(i_1, \ldots, i_r)$ is a prefix of $(j_1, \ldots, j_s)$, then $\Gamma(i_1, \ldots, i_r) \supseteq \Gamma(j_1, \ldots, j_s)$. Since $n = 2m + 3$, there is an element $(i_1, \ldots, i_r)$ of L_n such that for any $j, j' \in \mathbb{Z}$

$$\Gamma(i_1, \ldots, i_r, j, j') = \Gamma(i_1, \ldots, i_r).$$

Otherwise, we may find a strictly decreasing sequence of subsets of Q of length $m+2$ which is impossible. Let P be the set $\Gamma(i_1, \ldots, i_r)$ and let p and q be the states $\gamma(i_1, \ldots, i_r, -\frac{1}{2})$ and $\gamma(i_1, \ldots, i_r, \frac{1}{2})$. By definition of Γ, both states q and p belong to P. Furthermore, since $\Gamma(i_1, \ldots, i_r, 0, j)$ is equal to P for each $j \in \mathbb{Z}$, $p \to P$ and $P \to q$ are two limit transitions of $\mathcal{A}$. By the previous lemma, this completes the proof of the lemma.

This concludes the proof of Proposition 3.

Since the language L_5 was shown to be rational (see Example 11), 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 contained in the class of MSO-definable languages.*

5 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 φ 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 ω (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^ω 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 [26] 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.

Acknowledgements

The authors wish to thank the anonymous referees for useful suggestions.

References

1. N. Bedon. Logic over words on denumerable ordinals. *J. Comput. System Sci.*, 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 $< N, +, V_k >$. (the k-recognizable sets are definable in $< N, +, V_k >$). *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. H. Straubing. *Finite automata, formal logic and circuit complexity*. Birkhäuser, 1994.
26. W. Thomas. Star free regular sets of ω-sequences. *Inform. Control*, 42:148–156, 1979.
27. 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.
28. 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.
29. J. Wojciechowski. Finite automata on transfinite sequences and regular expressions. *Fundamenta informaticæ*, 8(3-4):379–396, 1985.