# Algebraic characterization of FO for scattered linear orderings*

**Alexis Bès[1] and Olivier Carton[2]**

1 **LACL, Université Paris-Est, Créteil, France**
bes@u-pec.fr

2 **LIAFA, Université Paris Diderot Paris, France**
Olivier.Carton@liafa.jussieu.fr

**Abstract**

We prove that for the class of sets of words indexed by countable scattered linear orderings, there is an equivalence between definability in first-order logic, star-free expressions with marked product, and recognizability by finite aperiodic semigroups which satisfy the equation $x^\omega x^{-\omega} = x^\pi$.

**1998 ACM Subject Classification** F.4.1 Mathematical Logic; F.4.3 Formal Languages

**Keywords and phrases** Linear orderings, first-order logic, semigroups, rational sets

**Digital Object Identifier** 10.4230/LIPIcs.CSL.2011.67

## 1 Introduction

One of the fundamental results in formal language theory is the equivalence between automata on finite words, rational expressions, recognizability by finite semigroups, and definability in monadic second-order logic (see e.g. [26]). This has been specified by Schützenberger [25], McNaughton and Papert [13], which show equivalence between counter-free automata, star-free expressions, recognizability by finite aperiodic semigroups, and definability in first-order logic.

These results have been extended to many classes of structures like infinite words [6, 14], bi-infinite words [10, 15], transfinite words [7, 1], traces, trees, pictures...

In [5], Bruyère and Carton introduce automata and rational expressions for words indexed by linear orderings. These notions unify naturally previously defined notions for finite words, left- and right-infinite words, bi-infinite words, and ordinal words. The question to know whether the above equivalence results hold in this setting has been addressed in several papers since then. Up to now, most results hold when one restricts to sets of words indexed by countable scattered orderings; recall that a linear ordering is scattered if it does not contain any dense sub-ordering. For this class of sets, the paper [5] already proves that a Kleene-like theorem holds. The works [23, 22] introduce a notion of ⋄-semigroup and show equivalence between recognizability by finite ⋄-semigroups and rational expressions. Finally [2] shows equivalence between rational expressions and monadic second-order logic.

Let us now consider the extension of Schützenberger-McNaughton-Papert results for sets of words indexed by countable scattered orderings. Bedon and Rispal [3] prove equivalence between star-free expressions and recognizability by finite aperiodic ⋄-semigroups. From

* Supported by ANR Frec


Computer Science Logic 2011 (CSL'11).
Editor: Marc Bezem; pp. 67–81
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

their work it is possible to obtain a characterization of star-free expressions in terms of FO definability in structures where one can quantify over elements *and cuts* of the underlying ordering. However it seems natural to consider the more classical logical framework (used e.g. in [2]) where the domain of the structure consists only of elements of the ordering, and ask for a characterization of FO-definable sets. For many classes of words (such as words indexed by $\omega$, $\mathbb{Z}$, an ordinal, or $\mathbb{R}$ [20]), the equivalence between star-free expressions with FO logic is a relatively simple generalization of McNaughton-Papert proof. Let us try to explain why this is not the case here. A crucial point in the proof that star-free sets are FO-definable is the possibility to define in FO the product $L = L_1 \cdot L_2$ of two FO-definable sets $L_1$ and $L_2$. Intuitively, this can be done by expressing (with a FO sentence) that $w \in L$ iff there exists some position $x$ in $w$ such that the prefix of $w$ which corresponds to positions before $x$ belongs to $L_1$, and the remaining suffix of $w$ belongs to $L_2$. The existence of $x$ is ensured by the fact that the underlying ordering is complete. We cannot use this idea when considering any countable scattered ordering.

In this paper we characterize FO-definable sets in terms of rational expressions and recognizability by semigroups. For rational expressions, we consider the class of star-free marked sets, which is a variant of the class of star-free sets where one uses the *marked* product instead of the classical product. The operation of marked product has already been studied extensively, in particular in connection with the hierarchy of concatenation [19, Sect 7.1]. For the algebraic side, this corresponds to the class of sets which can be recognized by a $\diamond$-semigroup which is finite and aperiodic, and satisfies the additional equation $x^\omega x^{-\omega} = x^\pi$. As an immediate corollary of this characterization, we prove that it is decidable whether a rational set of words indexed by countable scattered orderings is star-free marked or, equivalently, FO-definable. We obtain as a byproduct that marked products are less expressive than usual products for linear ordering whereas they have the same expressive power for finite words.

Let us mention partial results for the class of words indexed by any linear ordering. The paper [4] introduces a new rational operation of shuffle of sets which allows to deal with dense orderings, and extends the Kleene-like theorem proved in [5] to sets of words indexed by all linear orderings. The work [2] shows that rational sets are definable in MSO logic, but not the converse (for instance, it is shown that the class of scattered orderings is MSO-definable but not rational). The decidability of FO can also be obtained with automata through linear temporal logic which is equivalent on linear orderings [9].

The paper is organized as follows. Definitions and useful results concerning linear orderings, rational sets, logic, and semigroups are recalled respectively in Sections 2, 3, 4 and 5. Section 6 states and gives a sketch of the proof of the main result.

## 2 Words on scattered linear orderings

### 2.1 Scattered linear orderings

This section recalls basic definitions on linear orderings but the reader is referred to [24] for a complete introduction. Hausdorff's characterization of countable scattered linear orderings is given and words indexed by linear orderings are introduced.

A *linear ordering* $(J, <)$ is a total 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 $j \in J$.

A *gap* of an ordering $J$ is a cut $(K, L)$ such that $K \neq \emptyset$, $L \neq \emptyset$, $K$ 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.

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

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

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

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

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

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

The *rank* $r(J)$ of a countable scattered ordering $J$ is defined as the least ordinal $\alpha$ such that $J \in V_\alpha$.

## 2.2 Words

Let $A$ be a finite alphabet. A *word* $w = (a_j)_{j \in J}$ indexed by a linear ordering $J$ is a function from $J$ to $A$. $J$ is called the *length* of $w$. For instance $\omega$ is the length of right-infinite words $a_0 a_1 \cdots$ and $\zeta$ is the length of bi-infinite words $\cdots a_{-1} a_0 a_1 \cdots$.

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

We denote by $A^\diamond$ the set of all words over $A$ indexed by countable scattered linear orderings. The *rank* $r(w)$ of a word $w \in A^\diamond$, is, by definition the rank of its length $J$.

# 3 Rational sets of words on linear orderings

Bruyère and Carton have introduced rational expressions and automata for words indexed by countable scattered linear orderings [5]. They have proved that a set of words is described by a rational expression if and only if it is accepted by some finite automaton. Such a set is called *rational* in this paper. This result is an extension of the classical Kleene theorem on finite words. This section briefly recalls definitions of rational operations. In this paper, we will mainly use union and concatenation but the other operations are often useful to denote sets of words.

Let $A$ be a fixed finite alphabet. We consider the rational operations defined for any subsets $X$ and $Y$ of $A^\diamond$ by :

$$X + Y = \{z \mid z \in X \cup Y\} \qquad X \cdot Y = \{x \cdot y \mid x \in X, y \in Y\},$$
$$X^* = \{\textstyle\prod_{j\in\{1,\ldots,n\}} x_j \mid n \in \mathcal{N}, x_j \in X\}, \qquad X^\diamond = \{\textstyle\prod_{j\in J} x_j \mid J \in \mathcal{S}, x_j \in X\},$$
$$X^\omega = \{\textstyle\prod_{j\in\omega} x_j \mid x_j \in X\}, \qquad X^{-\omega} = \{\textstyle\prod_{j\in-\omega} x_j \mid x_j \in X\},$$
$$X^\sharp = \{\textstyle\prod_{j\in\alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\}, \qquad X^{-\sharp} = \{\textstyle\prod_{j\in-\alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\}.$$

As usual, the dot denoting concatenation is omitted in rational expressions. A *marked product* of $X$ and $Y$ is a product of the form $XaY$ for some letter $a \in A$.

The operator $\diamond$ that we have defined above is actually a special case of a more general binary operator defined in [5]. This binary operator is really needed to capture all rational sets but it is not used in this paper.

Let us define the main classes of sets that we use in this paper.

- The class $\mathrm{Rat}(A^\diamond)$ of all rational sets over $A$ indexed by countable scattered linear orderings is the smallest set containing $\{a\}$ for any $a \in A$, the empty set, and closed under all rational operations. It is proved in [23] that this class is closed under complementation and thus under all boolean operations.
- The set $\mathrm{SF}(A^\diamond)$ of star-free sets over $A$ indexed by countable scattered linear orderings is the smallest set containing $\{a\}$ for any $a \in A$, the empty set, and closed under product and all boolean operations.
- The set $\mathrm{SFM}(A^\diamond)$ of star-free marked sets over $A$ indexed by countable scattered linear orderings is the smallest set containing $\{a\}$ for any $a \in A$, the empty set, and closed under *marked* product and all boolean operations.

It is interesting to note that, for finite words, the corresponding classes $\mathrm{SF}(A^*)$ and $\mathrm{SFM}(A^*)$ coincide. Any product $KL$ is indeed equal to a finite union $K\varepsilon(L) + \sum_{a\in A} KaL_a$ where $L_a = a^{-1}L$. They do not coincide in our case as it is shown by Example 4. Let us illustrate these definitions by some examples.

▶ **Example 2.** Consider the set $X_1 \subseteq A^\diamond$ of words $w$ over $A = \{a, b\}$ such that every position in $w$ (apart from the last position, if any) admits a next position, and every position (apart from the first position, if any) admits a previous position. We have

$$X_1 = A^\diamond \setminus [(A^\diamond AA^\diamond)^\omega AA^\diamond + A^\diamond A(A^\diamond AA^\diamond)^{-\omega}].$$

Moreover $(A^\diamond AA^\diamond)^\omega = A^\diamond AA^\diamond \setminus A^\diamond A$ and $(A^\diamond AA^\diamond)^{-\omega} = A^\diamond AA^\diamond \setminus AA^\diamond$, thus $X_1$ is a star-free marked set.

▶ **Example 3.** Consider the set $X_2 = (a^\diamond aa^\diamond)^\omega(b^\diamond bb^\diamond)^{-\omega}$ of words $w$ over $A = \{a, b\}$ which can be written as $w = w_1w_2$ where $w_1$ is non-empty, contains only $a$ and has no last $a$ and $w_2$ is non-empty, contains only $b$ and has no first $b$. There is then a gap between $w_1$ and $w_2$. A star-free marked expression for $X_2$ is

$$A^\diamond aA^\diamond bA^\diamond \setminus (A^\diamond bA^\diamond aA^\diamond \cup a^\diamond ab^\diamond \cup a^\diamond bb^\diamond).$$

Observe that $a^\diamond = A^\diamond \setminus A^\diamond bA^\diamond$ and $b^\diamond = A^\diamond \setminus A^\diamond aA^\diamond$.

The following example gives a set $X_3$ which seems very close to the set $X_2$ of the previous example. This set will be star-free but our characterization of $SFM(A^\diamond)$ will allow us to prove that $X_3$ is not a star-free marked set (see Example 17).

▶ **Example 4.** Consider the set $X_3 = (A^\diamond AA^\diamond)^\omega (A^\diamond AA^\diamond)^{-\omega}$ of words $w$ over $A = \{a, b\}$ such that the underlying ordering of $w$ contains at least one gap. We have $X_3 \in SF(A^\diamond)$.

## 4 Logic

Let us recall useful elements of logic, and settle some notations. For more details we refer e.g. to Thomas' survey paper [26].

We consider first-order (shortly: FO) logic over relational signatures. As usual, we will often identify logical symbols with their interpretation. We call FO sentence every FO formula without free variable.

For every finite alphabet $A = \{a_1, \ldots, a_n\}$ we consider the relational signature $\mathcal{L}_A = \{<, P_{a_1}, \ldots, P_{a_n}\}$ where $<$ denotes a binary predicate symbol and every $P_{a_i}$ denotes a unary predicate symbol. 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 an FO sentence $\varphi$ over the signature $\mathcal{L}_A$, we define the set $L_\varphi$ as the set of words $w \in A^\diamond$ such that $M_w \models \varphi$. This definition extends to the case of FO formulas with free variables. For every word $w = (a_j)_{j\in J}$ over $A$ and every $n$-tuple $b_1, \ldots, b_n$ of elements of $J$, we define $w(b_1, \ldots, b_n)$ as the word $w' = (a'_j)_{j\in J}$ over the alphabet $\{0, 1\}^n \times A$ such that for every $j \in J$, the last component of $a'_j$ equals $a_j$, and for every $i \in \{1, \ldots, n\}$, the $i$-th component of $a'_j$ equals 1 if and only if $j = b_i$. Now, given a FO formula $\varphi(x_1, \ldots, x_n)$ with free variables $x_1, \ldots, x_n$, we define $L_\varphi$ as the class of words of the form $w(b_1, \ldots, b_n)$ over the alphabet $\{0, 1\}^n \times A$ such that $M_w \models \varphi(b_1, \ldots, b_n)$.

We will say that a set $X \subseteq A^\diamond$ is *FO-definable* if there exists an FO-formula $\varphi$ over the signature $\mathcal{L}_A$ such that $X = L_\varphi$.

▶ **Example 5.** The set $X_1$ of Example 2 is FO definable. We first define the auxiliary predicate $suc(x, y)$ as $x < y \wedge \neg\exists z(x < z \wedge z < y)$. Then $X_1$ is definable by the formula

$$\forall x[(\exists y\ x < y \longrightarrow \exists y\ suc(x, y)) \wedge (\exists y\ y < x \longrightarrow \exists y\ suc(y, x))].$$

▶ **Example 6.** The set $X_2$ of Example 3 can be defined by the FO-formula

$$\exists x\ P_a(x) \wedge \exists y\ P_b(y) \wedge \neg\exists x\exists y(x < y \wedge P_b(x) \wedge P_a(y)) \quad (1)$$
$$\wedge\forall x(P_a(x) \rightarrow \exists y(y > x \wedge P_a(y))) \quad (2)$$
$$\wedge\forall y(P_b(y) \rightarrow \exists x(x < y \wedge P_b(x))) \quad (3)$$

The sub-formula (1) expresses that the word contains some $a$ and some $b$, and that no $a$ occurs after some $b$. The sub-formula (2) (resp. (3)) ensures that there is no last $a$ (resp. no first $b$).

## 5 Algebraic characterization of rational sets

The algebraic objets that we use to characterize FO over countable scattered linear orderings are semigroups enriched with operations that make them suitable for linear orderings. We start with the definition of a semigroup.

A semigroup is a set $S$ equipped with an associative binary product. Since the product is associative, the product $s_1 \cdots s_n$ of $n$ elements $s_1, \ldots, s_n$ is well-defined. The semigroup $S^1$ is $S$ if $S$ has already a neutral element and it is the semigroup obtained by adding a neutral element otherwise. An *idempotent* $e$ of a semigroup is an element such that $e^2 = e$.

## 5.1 $\diamond$-semigroups

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

▶ **Definition 7.** A $\diamond$-*semigroup* is a set $S$ equipped with product $\pi : S^\diamond \to S$ which maps any word of $S^\diamond$ to an element of $S$ such that

- for any element $s$ of $S$, $\pi(s) = s$;
- for any word $x$ of $S^\diamond$ and for any factorization $x = \prod_{j\in J} x_j$ where $J \in S$,

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

The latter condition is a generalization of associativity. It states that for any factorization $x = \prod_{j\in J} x_j$ of some word $x \in S^\diamond$, the product of $x$ can be obtained by first computing the product $\pi(x_j)$ of each word $x_j$ to get a word $y = \prod_{j\in J} \pi(x_j)$ of length $J$ and then computing the product $\pi(y)$ of that word $y$.

Note that a $\diamond$-semigroup $(S, \pi)$ is already a semigroup. For any two elements $s$ and $t$ of $S$, the finite product $\pi(st)$ (more precisely, the product of the sequence $st$ of length 2) is defined. It is merely denoted by $st$. The generalization of associativity ensures that $r(st) = \pi(r\pi(st)) = \pi(rst) = \pi(\pi(rs)t) = (rs)t$ for any $r, s, t \in S$.

The set $A^\diamond$ equipped with the concatenation is a $\diamond$-semigroup. All $\diamond$-semigroups considered in this paper are either of the form $A^\diamond$ for some alphabet $A$ or they are finite. The following example is a $\diamond$-semigroup where the underlying set $S$ is finite (these $\diamond$-semigroups will be studied in Section 5.2).

▶ **Example 8.** The set $S = \{0, 1\}$ equipped with the product $\pi$ defined for any $u \in S^\diamond$ by $\pi(u) = 0$ if $u$ has at least one occurrence of the letter 0, and $\pi(u) = 1$ otherwise, is a $\diamond$-semigroup.

A *sub-$\diamond$-semigroup* $T$ of a $\diamond$-semigroup $S$ is a subset of $S$ closed under product. A *morphism of $\diamond$-semigroup* is an application which preserves the product. A function $\varphi : S \to T$ is a morphism from $(S, \pi_S)$ to $(T, \pi_T)$ if for any word $x = (s_j)_{j\in J}$, one has $\pi_T(\varphi(x)) = \varphi(\pi_S(x))$ where $\varphi(x) = (\varphi(s_j))_{j\in J}$. A $\diamond$-semigroup $T$ is a *quotient* of a $\diamond$-semigroup $S$ if there exists an onto morphism from $S$ to $T$. A $\diamond$-semigroup $T$ *divides* $S$ if $T$ is the quotient of a sub-$\diamond$-semigroup of $S$.

## 5.2 Finite $\diamond$-semigroups

A $\diamond$-semigroup $(S, \pi)$, of course, is said to be finite if $S$ is finite. Even when $S$ is finite, the function $\pi$ is not easy to describe because the product of any sequence has to be given. It turns out that the function $\pi$ can be described using a semigroup structure on $S$ with two additional functions from $S$ to $S$. This gives a finite description of the product $\pi$.

It has already been noted that a $\diamond$-semigroup $(S, \pi)$ has already a structure of semigroup since $\pi$ is defined on words of length 2. Let us define two functions $\tau : S \to S$ and $-\tau : S \to S$.

The images of these functions are denoted using exponentiation : $\tau : s \mapsto s^\tau$ and $-\tau : s \mapsto s^{-\tau}$. For any $s \in S$, $s^\tau$ and $s^{-\tau}$ are respectively equal to $\pi(s^\omega)$ and $\pi(s^{-\omega})$ where $s^\omega = sss\cdots$ and $s^{-\omega} = \cdots sss$ are the two words of length $\omega$ and $-\omega$ in which $s$ occurs at all positions.

The functions $\tau$ and $-\tau$ satisfy the following equations. For any $s, t \in S$ and for any integer $n$, one has $s(ts)^\tau = (st)^\tau$, $(s^n)^\tau = s^\tau$, $(st)^{-\tau}s = (ts)^{-\tau}$ and $(s^n)^{-\tau} = s^{-\tau}$. Equations for $\tau$ follow from the equality between the $\omega$-words $(s^n)^\omega$ and $s^\omega$ and from the equality between the $\omega$-words $(st)^\omega$ and $s(ts)^\omega$. Equations for $-\tau$ follow from similar relations for words of lengths $-\omega$. Functions satisfying these equations are respectively called *compatible to the right* and *compatible to the left* with $S$.

Note that these two functions $\tau$ and $-\tau$ can be defined even when the $\diamond$-semigroup is not finite. When the $\diamond$-semigroup $(S, \pi)$ is finite, the semigroup structure of $S$ and the two functions $\tau$ and $-\tau$ completely describe its product $\pi$. This is stated in the following theorem.

▶ **Theorem 9** ([23, 22]). *Let $S$ be a finite semigroup and let $\tau$ and $-\tau$ be functions respectively compatible to the right and to the left with $S$. Then $S$ can be uniquely endowed with a structure of $\diamond$-semigroup $(S, \pi)$ such that $s^\tau = \pi(s^\omega)$ and $s^{-\tau} = \pi(s^{-\omega})$.*

The previous theorem means that a finite $\diamond$-semigroup has a finite description. It suffices to give a semigroup product and two compatible functions to fully characterize the product.

We briefly explain how the product $\pi$ can be recovered from the semigroup structure and the compatible functions. The construction of $\pi$ is based on the next Lemma which follows directly from Ramsey's Theorem [21].

Let $S$ be a semigroup. We denote by $\varphi$ the natural morphism from $S^*$ to $S$ which maps any finite sequence of elements to their product. Let $x = (s_j)_{j\in\omega}$ be an $\omega$-word over $S$. A $\omega$-factorization of $x$ is $\omega$-sequence $(x_j)_{j\in\omega}$ of finite words such that $x = \prod_{j\in\omega} x_j$. A *right linked pair* of a semigroup $S$ is a pair $(s, e)$ such that $se = s$ and $e^2 = e$.

▶ **Lemma 10.** *For any $\omega$-word $x$ over a semigroup $S$, there is an $\omega$-factorization $x = \prod_{j\in\omega} x_j$ and a right linked pair $(s, e)$ such that $\varphi(x_0) = s$ and $\varphi(x_j) = e$ for any $j \geq 1$.*

Such a factorization is called a *ramseyan factorization*, see Theorem 3.2 in [17]. If $x = \prod_{j\in\omega} x_j$ is a ramseyan factorization of $x$, then $\pi(x)$ must be equal to $se^\tau$ since $\pi$ satisfies a generalized associativity. The product $\pi$ is then defined on all words in $S^\diamond$ by induction on their rank. A word of rank $\alpha$ is, indeed, either a finite, or an $\omega$ , or a $-\omega$-product of words of ranks smaller than $\alpha$.

Note that a given $\omega$-word over $S$ may have several ramseyan factorizations related to different right linked pairs $(s_1, e_1)$ and $(s_2, e_2)$. It turns out that these linked pairs are then conjugated. There exist elements $x, y \in S^1$ such that $s_1x = s_2$, $e_1 = xy$ and $e_2 = yx$. Since the function $\tau$ is compatible, one has $s_1e_1^\tau = s_1(xy)^\tau = s_1x(yx)^\tau = s_2e_2^\tau$.

The functions $\tau$ and $-\tau$ are usually denoted $\omega$ and $-\omega$. This may cause a small confusion since $s^\omega$ is either an $\omega$-word over $S$ or its product in $S$ but it is always clear from the context which one is meant.

▶ **Example 11.** Consider again the $\diamond$-semigroup $S = \{0, 1\}$ of Example 8. Its semigroup structure is $\{0, 1\}$ with the usual multiplication ($11 = 1$ and $00 = 01 = 10 = 0$). The compatible functions $\omega$ and $-\omega$ are defined by $0^\omega = 0^{-\omega} = 0$ and $1^\omega = 1^{-\omega} = 1$.

▶ **Example 12.** The set $S = \{0, 1\}$ equipped with the product $\pi$ defined for any $u \in S^\diamond$ by $\pi(u) = 1$ if only 1 occurs in $u$ and if the length of $u$ is an ordinal and $\pi(u) = 0$ otherwise, is a $\diamond$-semigroup. Its semigroup structure is again $\{0, 1\}$ with the usual multiplication but the compatible functions $\omega$ and $-\omega$ defined by $0^\omega = 0^{-\omega} = 1^{-\omega} = 0$ and $1^\omega = 1$.

Since any element $s$ of a finite semigroup has a power $s^n$ which is an idempotent and since $(s^n)^\omega = s^\omega$, it suffices to give the values of $e^\omega$ when $e$ is an idempotent to completely describe the function $\omega$. The same applies to the function $-\omega$.

## 5.3 Recognizability

It is well known that rational sets of finite words are exactly those recognized by finite semigroups (see e.g. [18]). This result can be generalized for words indexed by countable scattered linear orderings.

Let $S$ and $T$ be two $\diamond$-semigroups. The $\diamond$-semigroup $T$ *recognizes* a subset $X$ of $S$ if and only if there exists a morphism $\varphi : S \to T$ and a subset $P \subseteq T$ such that $X = \varphi^{-1}(P)$.

The following theorem states that finite $\diamond$-semigroups are equivalent to rational expressions and automata for words indexed by countable scattered orderings.

▶ **Theorem 13** ([23]). *A set $X \subseteq A^\diamond$ is rational if and only if it is recognized by a finite $\diamond$-semigroup.*

It is also proved in [23] that any rational set $X$ of words over countable scattered linear orderings has a *syntactic $\diamond$-semigroup*. This is a smallest $\diamond$-semigroup recognizing $X$ in a strong way. Not only it is the smallest in cardinality but it also divides any other $\diamond$-semigroup recognizing $X$. As in the case of finite words, this syntactic $\diamond$-semigroup can be obtained by quotienting any $\diamond$-semigroup recognizing $X$ by the relation that identifies elements which cannot be distinguished by contexts (intuitively, contexts are terms which involve $\omega$ and $-\omega-$products, and with a hole in it).

The following example shows how the $\diamond$-semigroup introduced in Example 12 can be used to recognize the set of words of ordinal length.

▶ **Example 14.** Consider the $\diamond$-semigroup $S$ defined in Example 12 and the morphism $\varphi : A^\diamond \to S$ defined by $\varphi(a) = 1$ for any $a \in A$. The set of words with countable ordinal length is recognizable since it is equal to $\varphi^{-1}(\{1\})$. It will be shown after Theorem 21 that this set is not FO-definable. This is a variant of Tarski's result (see [24, Theorem 13.13]) that the class of well-orderings is not elementary.

We give below some examples of morphisms from $A^\diamond$ into finite $\diamond$-semigroups that recognize subsets of $A^\diamond$ that have been already encountered in Examples 2, 3 and 4. It can be checked that, in each example, the given $\diamond$-semigroup is actually the syntactic $\diamond$-semigroup of the set it recognizes. For each $\diamond$-semigroup, we give the $\mathcal{D}$-classes structure.

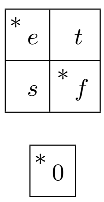


**Figure 1** $\mathcal{D}$-classes structure of $\diamond$-semigroup of Example 15

▶ **Example 15.** The set $X_1$ of Example 2 is recognized by the $\diamond$-semigroup $S_1 = \{0, e, t, s, f\}$ whose product is defined by $ts = e^2 = e$, $et = tf = t$, $se = fs = s$, $st = f^2 = f$, $e^\omega = t$, $e^{-\omega} = s$, $f^\omega = f^{-\omega} = f$ and any other product is equal to 0. Define the morphism

$\varphi_1 : A^\diamond \to S_1$ by $\varphi_1(a) = \varphi_1(b) = e$. Observe that $\varphi_1^{-1}(e)$ is the set of words which have a first and a last element, and $\varphi_1^{-1}(f)$ is the set of words which have neither a first nor a last element. We have $X_1 = \varphi_1^{-1}(S_1 \setminus \{0\})$.

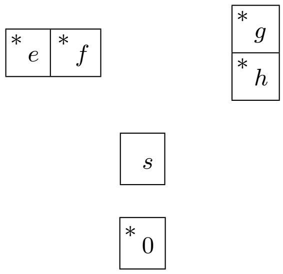


**Figure 2** $\mathcal{D}$-classes structure of $\diamond$-semigroup of Example 16

▶ **Example 16.** The set $X_2$ of Example 3 is recognized by the $\diamond$-semigroup $S_2 = \{0, e, f, g, h, s\}$ whose product is defined by $e^2 = fe = e^{-\omega} = e$, $f^2 = ef = e^\omega = f^\omega = f^{-\omega} = f$, $g^2 = gh = g^\omega = g$, $h^2 = hg = h^\omega = g^{-\omega} = h^{-\omega} = h$, $fh = es = fs = sg = sh = s$, and any other product is equal to 0. Define the morphism $\varphi_2 : A^\diamond \to S_2$ by $\varphi_2(a) = e$ and $\varphi_2(b) = g$. We have $X_2 = \varphi_2^{-1}(s)$. It will be seen after Theorem 21 that $X_2$ is FO-definable.

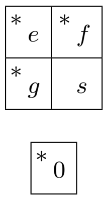


**Figure 3** $\mathcal{D}$-classes structure of $\diamond$-semigroup of Example 17

▶ **Example 17.** The set $X_3$ of Example 4 is recognized by the $\diamond$-semigroup $S_3 = \{0, e, f, g, s\}$ whose product is defined by $e^2 = fe = eg = e$, $f^2 = ef = es = e^\omega = f^\omega = f$, $g^2 = ge = se = e^{-\omega} = g^{-\omega} = g$, $gs = sf = gf = g^\omega = f^{-\omega} = s$ and any other product is equal to 0. Define the morphism $\varphi_3 : A^\diamond \to S_3$ by $\varphi_3(a) = \varphi_3(b) = e$. We have $X_3 = \varphi_3^{-1}(0)$. Our characterization of FO-definability (Theorem 21) will imply that $X_3$ is not FO-definable.

## 5.4 Equivalence between rational sets, $\diamond$-semigroups, and logic

The equivalence between rational expressions, $\diamond$-semigroups and logic was proved in [23, 22, 2]. The logical side involves Monadic Second-Order (shortly: MSO) logic. Recall that MSO logic is an extension of first-order logic that allows to quantify over elements as well as subsets of the domain of the structure. The notion of MSO-definable set extends in a natural way the one of FO-definable set. For more details about MSO logic we refer e.g. to [11, 26].

The following theorem is an extension of the classical theorem of Büchi [6] which states that, for finite words, MSO exactly defines rational sets of words.

▶ **Theorem 18** ([23, 22, 2]). *Let $A$ be a finite alphabet, and let $X \subseteq A^\diamond$ be a set of words indexed by countable scattered linear orderings. Then the following properties are equivalent:*

1. *$X$ is rational;*
2. *$X$ is recognizable by a finite $\diamond$-semigroup;*
3. *$X$ is MSO-definable.*

Bedon and Rispal [3] extended Schützenberger' theorem [25] to the class of sets of words indexed by countable scattered linear orderings. In order to state their result, we recall the definitions of an aperiodic semigroup and of an aperiodic $\diamond$-semigroup.

▶ Proposition 19. ([17, Annex A,Prop. 2.9]) Let $S$ be a finite semigroup. The following conditions are equivalent.
1. There exists an integer $n$ such that $s^{n+1} = s^n$ for any $s \in S$;
2. every group in $S$ is trivial;
3. Each $\mathcal{H}$-class is trivial.

A semigroup $S$ satisfying these conditions is called *aperiodic*.

Note that if $s^{n+1} = s^n$, then $s^{m+1} = s^m$ for any $m \geq n$. If a finite semigroup $S$ is aperiodic, then $s^{n+1} = s^n$ for any $s \in S$ and for any large enough integer $n$. Such an integer is traditionally denoted by $\omega$ in semigroup theory but we will denote it $\pi$. This symbol $\pi$ is also used for the product of the $\diamond$-semigroup but this will not lead to ambiguous interpretations in the sequel. A $\diamond$-semigroup $S$ is said to be aperiodic whenever its semigroup structure is an aperiodic semigroup. This definition allows us to state the characterization of star-free sets due to Bedon and Rispal.

▶ **Theorem 20** ([3]). *Let $A$ be a finite alphabet, and let $X \subseteq A^\diamond$ be a set of words indexed by countable scattered linear orderings. Then the following properties are equivalent:*
1. *$X$ is star-free;*
2. *$X$ is recognizable by a $\diamond$-semigroup which is finite and aperiodic;*
3. *$X$ can be defined by FO over the cuts.*

In the previous theorem, "defined by FO over the cuts" means that $X$ is defined by a first order formula with quantification over positions of the words, (that is, elements of its length) but also over cuts of its length. This is not, strictly speaking, FO since the cuts are not part of the structure $\mathcal{M}_w$ of a word $w$. The last statement is not given in [3] but the equivalence between star-freeness and FO over the cuts is not difficult. In this paper, we give the equivalence between marked star-freeness and FO (without the cuts). The proof carries easily over star-freeness and the cuts.

For sets of finite words, McNaughton-Papert theorem [13] states that star-free sets coincide with FO-definable sets. For sets of words indexed by countable scattered linear orderings, one can prove that FO-definable sets are star-free (see Proposition 23), but the converse does not hold anymore. For instance, it is easy to check that the set $X = \{a^\omega\}$ is star-free, but it can be shown that $X$ is not FO-definable (this comes from the fact that the ordering $\omega$ is undistinguishable from any ordering of the form $\omega + \zeta \times \alpha$ in FO logic, see e.g. [24, Proposition 6.12]). In the next section we provide a characterization of FO-definable sets.

## 6 Main result

We finally come to the main result of the paper, characterization of FO for words over countable scattered linear orderings

▶ **Theorem 21.** *Let $A$ be a finite alphabet, and let $X \subseteq A^\diamond$ be a set of words indexed by countable scattered linear orderings. Then the following properties are equivalent:*
1. *$X$ is a star-free marked set;*

2. *$X$ is FO-definable;*
3. *$X$ is recognizable by a finite aperiodic $\diamond$-semigroup satisfying the equation $x^\omega x^{-\omega} = x^\pi$;*
4. *the syntactic $\diamond$-semigroup of $X$ is finite, aperiodic and satisfies the equation $x^\omega x^{-\omega} = x^\pi$.*

In the sequel, a finite aperiodic $\diamond$-semigroup satisfying the equation $x^\omega x^{-\omega} = x^\pi$ is called an *FO-semigroup*. The theorem is illustrated by the following examples.

The set $X_1$ of Example 2 is star-free marked, FO-definable (see Example 5) and the $\diamond$-semigroup provided in Example 15 is a FO-semigroup. Similarly, the set $X_2$ of Example 3 is star-free marked, FO-definable (see Example 6) and the $\diamond$-semigroup provided in Example 16 is a FO-semigroup. On the other hand, the set $X_3$ of Example 4 is star-free but its syntactic $\diamond$-semigroup given in Example 17 is not a FO-semigroup. Indeed we have $e^\omega e^{-\omega} = fg = 0 \neq e^\pi$ since $e^\pi = e$. Thus $X_3$ is not a star-free marked set, and is not FO-definable.

Theorem 21 yields an effective procedure to test whether a rational set $X \subseteq A^\diamond$ is star-free marked. Indeed Theorem 18 allows to compute effectively the finite syntactic $\diamond-$semigroup of $X$, from which one can decide whether $S$ is aperiodic and satisfies the equation $x^\omega x^{-\omega} = x^\pi$.

▶ **Corollary 22.** *Let $X \subseteq A^\diamond$ be a rational set of words indexed by countable scattered linear orderings. Then it is decidable whether $X$ is FO-definable.*

The proof of Theorem 21 is organized as follows. Section 6.1 proves the equivalence between (1) and (2). This is a straightforward generalization of the case of sets of finite words. In Proposition 24 we prove that (1) implies (3); it is again a rather easy extension of the case of finite words. The most difficult part is to prove that (3) implies (1), namely that sets recognizable by FO-semigroups are star-free marked sets. The proof is long and technically involved. It relies on the study of the structure of $\mathcal{D}$-classes of a FO-semigroup which recognizes the set. In Section 6.2.2 we give the general structure of the proof, but details are omitted.

## 6.1 First-order logic vs star-free marked sets

In this section we state equivalence between star-free marked expressions and FO-definability.

▶ Proposition 23. Let $A$ be a finite alphabet, and let $X \subseteq A^\diamond$ be a set of words indexed by countable scattered linear orderings. The set $X$ is a star-free marked set if and only if $X$ is definable in first-order logic.

**Proof.** (sketch) The proof is a straightforward generalization of the proof of McNaughton-Papert Theorem given in [26, Theorem 4.4].

The "only if" part is proved by induction on a star-free marked expression denoting $X$.

For the converse, consider a first-order formula $\varphi(x_1, \ldots, x_n)$ with quantifier-depth $m$, and assume (without loss of generality) that $\varphi$ holds only if $x_1 < x_2 \cdots < x_n$. Then one can prove by induction on $m$ (using Ehrenfeucht-Fraïssé games) that $\varphi$ is equivalent to a disjunction of formulas of the form

$$\psi_0 \wedge P_{a_1}(x_1) \wedge \psi_1 \wedge \cdots \wedge P_{a_n}(x_n) \wedge \psi_n$$

where all formulas $\psi_i$ have quantifier depth $m$, and for every $1 \leq i \leq n-1$ (respectively $i = 0$, $i = n$), $\psi_i$ is a formula where all quantifiers are relativized to the interval $(x_i, x_{i+1})$, except for $\psi_0$ (respectively $\psi_n$) for which all quantifiers are relativized to elements less than $x_0$ (respectively greater than $x_n$).

Now assume that $X$ is FO-definable by a sentence $\psi$. Assume first that $\psi$ is of the form $\exists x \varphi(x)$. Then using the above result, $\psi$ is equivalent to a disjunction of formulas of the form

$\exists x(\psi_0 \wedge P_{a_i}(x) \wedge \psi_1)$. Each such formula defines a star-free marked set, thus $X \in SFM(A^\diamond)$. In case $\psi$ has the form $\forall x \varphi(x)$, we use the equivalence $\psi \equiv \neg\exists x \neg\varphi(x)$. ◀

## 6.2 FO-semigroups vs star-free marked sets

### 6.2.1 From star-free marked sets to FO-semigroups

▶ Proposition 24. Let $A$ be a finite alphabet, and let $X \subseteq A^\diamond$ be a set of words indexed by countable scattered linear orderings. If $X$ is a star-free marked set then $X$ is recognizable by a FO-semigroup.

The proof is very close to the one given in [16] for the case of sets of finite words. It goes by induction on a star-free marked expression denoting the set $X$. Here we have to show, in addition, that a $\diamond$-semigroup $S$ which recognizes $X$ satisfies the equation $x^\omega x^{-\omega} = x^\pi$.

For every set $X \subseteq A^\diamond$ recognizable by a finite aperiodic $\diamond$-semigroup $S$, we define $i(X)$ as the least integer $n$ such that for every $x \in A^\diamond$ and every context $C$, we have $C(x^{n+1}) \in X \Leftrightarrow C(x^n) \in X$ (let us recall that a context is, intuitively, a term with a hole in it).

**Proof.** The proof goes by induction on a star-free marked expression denoting $X$.

The cases when $X = \emptyset$, and $X = \{a\}$ with $a \in A$, are easy.

Assume now that $X_1$ and $X_2$ are star-free marked sets which are recognized by the FO-semigroup $S_1$ and $S_2$, respectively. The proof that the sets $Y_1 = X_1 + X_2$ and $Y_2 = A^\diamond \setminus X_1$ are recognizable by a FO-semigroup is easy. They are both recognized by the $\diamond$-semigroup $S_1 \times S_2$ with the component-wise product. This $\diamond$-semigroup is obviously a FO-semigroup.

Let us prove that every set $X \subseteq A^\diamond$ of the form $X = X_1 a X_2$ with $a \in A$, is recognizable by a FO-semigroup.

Let $S$ be a finite $\diamond$-semigroup which recognizes $X$, and let $\varphi : A^\diamond \to S$ be the associated morphism. Let us show that $S$ is aperiodic with $i(X) \leq i(X_1) + i(X_2) + 1$. This amounts to show that for all words $u, v, w \in A^\diamond$ and every integer $n \geq i(X_1) + i(X_2) + 1$, one has $uv^n w \in X$ if and only if $uv^{n+1} w \in X$. It is actually sufficient to prove that if $uv^n w \in X$, then $uv^{n+1} w \in X$ since the finiteness of $S$ implies that there always exists an integer $p \geq 1$ such that for $n$ large enough, one has $uv^n w \in X$ if and only if $uv^{n+p} w \in X$. Assume first that $uv^n w \in X$. By definition of $X$ there exists $z_1 \in X_1$ and $z_2 \in X_2$ such that $uv^n w = z_1 a z_2$. We consider several cases:

- if $uv^n w' = z_1$ for some prefix $w'$ of $w$, then it follows from our hypothesis on $X_1$ and the fact that $n \geq i(X_1)$ that $uv^{n+1} w' \in X_1$, thus $uv^{n+1} w \in X$.
- The case when $u' v^n w = z_2$ for some suffix $u'$ of $u$ is similar to the previous case.
- Assume now that $z_1 = uv^{n_1} v_1$ and $z_2 = v_2 v^{n_2} w$ with $v_1 a v_2 = v$ and $n_1 + n_2 + 1 = n$. By hypothesis we have $n_1 + n_2 + 1 \geq i(X_1) + i(X_2) + 1$, thus either $n_1 \geq i(X_1)$, or $n_2 \geq i(X_2)$. If $n_1 \geq i(X_1)$ then it follows from our hypothesis on $X_1$ that $uv^{n_1+1} v_1 \in X_1$, which yields $uv^{n+1} w \in X$. The case when $n_2 \geq i(X_2)$ is similar.

Let us now prove that $x^\omega x^{-\omega} = x^\pi$ for every $x \in S$. This amounts to show that for all words $u, v, w \in A^\diamond$, there exists an integer $n > i(X)$ such that $uv^\omega v^{-\omega} w \in X$ if and only if $uv^n w \in X$. The case $v = \varepsilon$ is trivial. We suppose now that $v \neq \varepsilon$.

Assume first that $uv^\omega v^{-\omega} w \in X$. By definition of $X$ there exists $z_1 \in X_1$ and $z_2 \in X_2$ such that $uv^\omega v^{-\omega} w = z_1 a z_2$. We consider several cases:

- if $uv^\omega$ is a prefix of $z_1$, then it is a strict prefix since $az_2$ cannot be equal to $v^{-\omega} w$. Thus $z_1 = uv^\omega v^{-\omega} w'$ for some prefix $w'$ of $w$. It follows from our hypothesis on $X_1$ that for every $n \geq i(X_1)$ we have $uv^n w' \in X_1$, thus $uv^n w \in X$.

- the case when $z_1$ is a prefix of $uv^\omega$ is similar.

Note that we have really used here that this is a marked product.

Conversely assume that $uv^n w \in X$ for some integer $n > i(X)$, and let $uv^n w = z_1 a z_2$ with $z_1 \in X_1$ and $z_2 \in X_2$. By definition of $i(X)$ we can assume that $n \geq i(X_1) + i(X_2) + 1$. Let $y = uv^n w$.

If $y$ can be written as $y = uv^{n_1} w_1 w_2$ with $n_1 \geq i(X_1)$ and $uv^{n_1} w_1 \in X_1$, then the induction hypothesis implies that $uv^\omega v^{-\omega} w_1 \in X_1$, hence $uv^\omega v^{-\omega} w \in X$.

The case where $y$ can be written as $y = u_1 u_2 v^{n_2} w$ with $n_2 \geq i(X_2)$ and $u_2 v^{n_2} w \in X_2$ is similar. ◀

### 6.2.2 From FO-semigroups to star-free marked sets

In this section we state the following result, and discuss about its proof.

▶ Proposition 25. Let $A$ be a finite alphabet, and let $X \subseteq A^\diamond$ be a set of words indexed by countable scattered linear orderings. If $X$ is recognizable by a FO-semigroup then $X$ is a star-free marked set.

Let us explain the main ingredients of the proof of Proposition 25. The structure of the proof is similar to the one of the proof of Schützenberger's theorem given in [16]. The proof goes by induction on the $\mathcal{D}$-classes of the FO-semigroup recognizing $X$.

Let us give some details. Assume that $X$ is recognized by the morphism $\varphi : A^\diamond \to S$ into a FO-semigroup $S$. There exists a subset $P$ of $S$ such that $X = \varphi^{-1}(P)$. Since $X = \bigcup_{s \in P} \varphi^{-1}(s)$ and since star-free marked sets are closed under union, it is sufficient to prove that $\varphi^{-1}(s)$ is a star-free marked set for each $s \in S$.

For every subset $P$ of $S$, let $X_P = \varphi^{-1}(P)$. In case $P$ is reduced to a singleton set $\{s\}$, we simply denote $X_P$ as $X_s$. We shall prove that for every $s \in S$, the set $X_s$ is marked star-free. This is proved by induction on the integer $h(s) = |S| - |S^1 s S^1|$ where $|P|$ denotes the cardinality of $P$.

The following definition is frequently used in the proof. Let $X$ and $Y$ be two subsets of $A^\diamond$ and let $D$ be a $\mathcal{D}$-class of $S$. We say that $Y$ is a *$D$-approximation of* $X$ if

$$X \subseteq Y \subseteq X \cup \varphi^{-1}(\{s \mid s <_{\mathcal{J}} D\})$$

We often use this definition when $X = \varphi^{-1}(P)$ for some subset $P$ of $D$.

The main difference with the proof of Schützenberger's theorem in [16], is that sets of the form $X_s X_t$ for $s, t \in S$ appear in some rational expressions. Since these products are not marked, it is necessary to prove that these sets are also star-free marked. To cope with this problem, we actually prove by induction on $k \geq 0$ the following two statements.

($P_1$) for every $s \in S$, if $h(s) \leq k$ then $X_s$ is a star-free marked set.

($P_2$) for all $s, t \in S$, if $h(s) \leq k$, $h(t) \leq k$ and $h(st) > k$ then $X_s X_t$ is a star-free marked set.

Note that $X_s X_t$ is contained in $X_{st}$. If $h(st) \leq k$, the set $X_s X_t$ can always be replaced in expressions by $X_{st}$, which is already star-free marked by ($P_1$).

We do not discuss the case $k = 0$ which is easy. Assume now that $k > 0$. Observe first that the set $\{s \mid h(s) = k\}$ is a union of $\mathcal{D}$-classes. Indeed for all $s, t \in S$, the relation $s \mathrel{\mathcal{D}} t$ implies $h(s) = h(t)$. Let $D$ be one of these $\mathcal{D}$-classes and let $s_0$ be an element of $D$. Let $R$ and $L$ be the $\mathcal{R}$-class and the $\mathcal{L}$-class of $s_0$, respectively. The main steps of the proof of ($P_1$) are the following.

1. We show that there exist two star-free marked sets $Y_R$ and $Y_L$ which are $D$-approximations of $X_R$ and $X_L$. Since $S$ is aperiodic, we have $\{s_0\} = R \cap L$, and $Y_R \cap Y_L$ is a $D$-approximation of $X_{s_0}$.

2. We show that the set $Z = \varphi^{-1}(\{s \mid s <_{\mathcal{J}} D\})$ is star-free marked set. This gives the equality $X_{s_0} = (Y_R \cap Y_L) \setminus Z$ which shows that $X_{s_0}$ is star-free marked.

## 7 Conclusion

Let us mention a few problems that are raised by our work.

We proved that for countable scattered orderings, FO logic captures the class of star-free marked sets. Which extension of FO does capture the class of star-free sets ? By [2], we know that this logic is a strict fragment of MSO. It can be shown that the existential fragment of MSO is not convenient, since for instance even the set $\{a^\omega\}$ is not definable in this fragment.

In the case of finite words, some subclasses of FO have already been algebraically characterized. Let us mention, for instance, that FO with two variables, usually called $FO^2$, correspond to a class of semigroups called DA. It would be interesting to know whether this is still true for linear orderings.

A lot of results concerning FO over linear orderings are obtained with Ehrenfeucht-Fraïssé games [24]. Some of them may deserve to be reconsidered using an algebraic approach.

Another interesting question is to remove the hypothesis *countable* or *scattered*. Very recently, the second author, together with Colcombet and Puppis [8], have extended the algebraic framework, and also the equivalence with MSO logic, to the case of all countable orderings. Does the characterization of FO still hold in that framework ?

**Acknowledgements** We thank the anonymous referees for useful suggestions.

### References

1 N. Bedon. Finite automata and ordinals. *Theoret. Comput. Sci.*, 156:119–144, 1996.
2 N. Bedon, A. Bès, O. Carton, and C. Rispal. Logic and rational languages of words indexed by linear orderings. *Theory of Computing Systems*, 46(4):737–760, 2010.
3 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.
4 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.
5 V. Bruyère and O. Carton. Automata on linear orderings. *J. Comput. System Sci.*, 73(1):1–24, 2007.
6 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.
7 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.
8 O. Carton, T. Colcombet, and G. Puppis. Regular languages of words over countable linear orderings. 2011. submitted.
9 J. Cristau. Automata and temporal logic over arbitrary linear time. *CoRR*, abs/1101.1731, 2011.
10 D. Girault-Beauquier. Bilimites de langages reconnaissables. *Theoret. Comput. Sci.*, 33(2–3):335–342, 1984.
11 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.

**12** F. Hausdorff. *Set Theory.* Chelsea, New York, 1957.

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

**14** D. E. Muller. Infinite sequences and finite machines. In *Switching Circuit Theory and Logical Design: Proc. Fourth Annual Symp.*, pages 3–16. I.E.E.E., New York, 1963.

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

**16** D. Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 1, pages 1–57. Elsevier, 1990.

**17** D. Perrin and J.-É. Pin. *Infinite Words.* Elsevier, 2004.

**18** J.-É. Pin. *Variétés de Langages Formels.* Masson, Paris, 1984.

**19** J.-É. Pin. Syntactic semigroups. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 679–746. Springer-Verlag, 1997.

**20** A. Rabinovich. Star free expressions over the reals. *Theoret. Comput. Sci.*, 233(1–2):233–245, 2000.

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

**22** C. Rispal. *Automates sur les ordres linéaires: complémentation.* PhD thesis, University of Marne-la-Vallée, France, 2004.

**23** 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.

**24** J. G. Rosenstein. *Linear orderings.* Academic Press, New York, 1982.

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

**26** 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.