Factorization forests for infinite words and applications to countable scattered linear orderings

Thomas Colcombet[a]

[a] *Laboratoire* Liafa
Université Paris Diderot - Paris 7 and Cnrs
Case 7014
75205 Paris Cedex 13, France

Abstract

The theorem of *factorization forests* of Imre Simon shows the existence of nested factorizations — à la Ramsey — for finite words. This theorem has important applications in semigroup theory, and beyond.

We provide two improvements to the standard result. First we improve on all previously known bounds. Second, we extend it to 'every linear ordering'.

We use this last variant in a simplified proof of the translation of recognisable languages over countable scattered linear orderings to languages accepted by automata.

Key words: Formal languages, semigroups, infinite words, automata, factorization trees

1. Introduction

Factorization forests were introduced by Simon [32]. The associated theorem — which we call the theorem of factorization forests below — states that for every semigroup morphism from words to a finite semigroup S, every word has a Ramseyan factorization tree of height linearly bounded by $|S|$ (see below). An alternative presentation states that for every semigroup morphism φ from A^+ to some finite semigroup S, there exists a regular expression evaluating to A^+ in which the Kleene star L^* is allowed only when $\varphi(L) = \{e\}$ for some idempostent e in S; i.e., the Kleene star is allowed only if it produces a Ramseyan factorization of the word.

The theorem of factorization forests provides a very deep insight on the structure of finite semigroups, and has therefore many applications. Let us cite some of them. Distance automata are non-deterministic finite automata mapping words to non-negative integers. An important question concerning them is the limitedness problem: decide whether the range of this mapping is bounded or not. It has been shown decidable by Simon using the theorem of factorization forests [32] (another proof was known before from Haschigushi [17]). An extension of this proof to a more general form of automata has been done in [1]. This theorem also allows a constructive proof of Brown's lemma on locally finite semigroups [6]. It is also used in the characterization of subfamilies of

URL: thomas.colcombet@liafa.jussieu.fr (Thomas Colcombet)

the regular languages, for instance the polynomial closure of varieties in [28] and more recently the polynomial closure of lattices of regular languages [27]. In the context of languages of infinite words indexed by ω, it has also been used in a complementation procedure [5] extending Büchi's proof [10].

The present paper aims first at advertising the theorem of factorization forest which, though already used in many papers, is in fact known only to a quite limited community. The reason for this is that its proofs rely on the use of Green's relations: Green's relations form an important tool in semigroup theory, but are rather technical to work with. The merit of the factorization forest theorem is that it is usable without any significant knowledge of semigroup theory, while it encapsulates nontrivial parts of this theory. Furthermore, as briefly mentioned above, this theorem has natural applications in automata theory.

This paper contains three contributions. First, we provide a new proof of the original theorem. Our proof improves on previously known bounds in [32] and [12], and yields a bound comparable to the recent one obtained by Kufleitner independently [19, 20]. Our result also improves on other works, since it does not apply to finite words only, but also to words indexed by ordinals. Second, we extend the result to the general case, i.e., to words that can be infinite, even if not indexed with an ordinal (though we use a different presentation). However, the bound is not as good in this situation. Third, we use this last extension in a simplified proof of complement for automata on countable scattered linear orderings, a result known from Carton and Rispal [11].

The content of the paper is organized as follows. Section 2 is dedicated to definitions. Section 3 presents the original theorem of factorization forests as well as a variant in terms of Ramseyan splits. In Section 4, the result is extended to infinite linear orderings. In Section 5 we apply this last extension to the complementation of automata over countable scattered linear orderings.

2. Definitions

In this section, we successively present linear orderings, words indexed by them, semigroups and additive labelings.

2.1. Linear orderings

A *linear ordering* $\alpha = (L, <)$ is a set L equipped with a total ordering relation $<$; i.e., an irreflexive, antisymmetric and transitive relation such that for every distinct elements x, y in L, either $x < y$ or $y < x$. Two linear orderings $\alpha = (L, <)$ and $\beta = (L', <')$ have same *order type* if there exists a bijection f from L onto L' such that for every x, y in L, $x < y$ iff $f(x) <' f(y)$. We denote by $\omega, -\omega, \zeta$ the order types of respectively $(\mathbb{N}, <)$, $(-\mathbb{N}, <)$ and $(\mathbb{Z}, <)$. Below, we do not distinguish between a linear ordering and its order type unless necessary. This is safe since all the constructions we perform are defined up to similar order type.

A *subordering* of a linear ordering $\alpha = (L, <)$ is a linear ordering of the form $\beta = (L', <')$ in which $L' \subseteq L$ and $<' = < \cap (L')^2$. We write $\beta \subseteq \alpha$. A subset $S \subseteq \alpha$ is *convex* if for all $x, y \in S$ and $x < z < y$, $z \in S$. We use the notations $[x, y]$, $[x, y[$, $]x, y]$, $]x, y[$, $]-\infty, y]$, $]-\infty, y[$, $[x, +\infty[$ and $]x, +\infty[$ for denoting the usual *intervals*. Intervals are convex, but the converse does not hold in general if α is not complete (see below). Given two subsets X, Y of a linear ordering, $X < Y$ holds if for all $x \in X$ and $y \in Y$, $x < y$.

The *sum* of two linear orderings $\alpha_1 = (L_1, <_1)$ and $\alpha_2 = (L_2, <_2)$ (up to renaming, assume L_1 and L_2 disjoint), denoted $\alpha_1 + \alpha_2$, is the linear ordering $(L_1 \cup L_2, <)$ with $<$ coinciding with $<_1$ on L_1, with $<_2$ on L_2 and such that $L_1 < L_2$. More generally, given a linear ordering $\alpha = (L, <)$ and for each $x \in L$ a linear ordering $\beta_x = (K_x, <_x)$ (the K_x are assumed disjoint), we denote by $\sum_{x\in\alpha} \beta_x$ the linear ordering $(\cup_{x\in L} K_x, <')$ with $x' <' y'$ if $x < y$ or ($x = y$ and $x' <_x y'$), where $x' \in K_x$ and $y' \in K_y$.

A linear ordering α is *complete* if every nonempty subset of α with an upper bound has a least upper bound in α, and every nonempty subset of α with a lower bound has a greatest lower bound in α.

A (Dedekind) *cut* of a linear ordering $\alpha = (L, <)$ is a couple (E, F) where $\{E, F\}$ is a partition of L, and $E < F$. Cuts are totally ordered by $(E, F) < (E', F')$ if $E \subsetneq E'$. This order has a minimal element $\bot = (\emptyset, L)$ and a maximal element $\top = (L, \emptyset)$. We denote by $\overline{\alpha}$ the set of cuts of α. It is classical that $(\overline{\alpha}, <)$ is a complete linear ordering. We also abbreviate by $\overline{\alpha}^{[]}, \overline{\alpha}^{[[}, \overline{\alpha}^{]]}, \overline{\alpha}^{][}$ the sets $\overline{\alpha}$, $\overline{\alpha} \setminus \{\top\}$, $\overline{\alpha} \setminus \{\bot\}$, $\overline{\alpha} \setminus \{\bot, \top\}$ respectively. Cuts can be thought as new elements located between the elements of α: given $x \in \alpha$, $x^- = (]-\infty, x[, [x, +\infty[)$ represents the cut placed just before x, while $x^+ = (]-\infty, x],]x, +\infty[)$ is the cut placed just after x. We say in this case that x^+ is *the successor of* x^- *through* x.

Each element x in a linear ordering α can be of three forms depending on nature of the interval $]-\infty, x[$; (a) if it is empty then x is the minimal element of α, (b) if it has a maximal element, then x is called a *successor*, and finally (c) if it is nonempty but has no maximal element, then x is called a *limit from the left*. Remark that this definition of a successor is consistent with the one introduced just above in the context of Dedekind cuts. By symmetry, the same separation into maximal element, *predecessors* and *limits from the right* is used. Below, we will use this in the case of countable linear orderings, or of the Dedekind cuts of countable orderings. In those two cases, x being a limit from the left is equivalent to the existence of a sequence $x_1 < x_2 < \dots$ of order type ω (an *ω-sequence* for short) of supremum x. This does not hold in the general case for which sequences indexed by higher ordinals are necessary.

A linear ordering α is *dense* if for every $x < y$ in α, there exists z in $]x, y[$. A linear ordering is *scattered* if it is not dense on any subordering. For instance $(\mathbb{Q}, <)$ and $(\mathbb{R}, <)$ are dense, while $(\mathbb{N}, <)$ and $(\mathbb{Z}, <)$ are scattered. Being scattered is preserved under taking a subordering. A scattered sum of scattered linear orderings also yields a scattered linear ordering. Every ordinal is scattered. Furthermore, if α is scattered, then $\overline{\alpha}$ is scattered. And if α is countable and scattered, then $\overline{\alpha}$ is also countable and scattered.

Additional material on linear orderings can be found in [31].

2.2. Words, languages

We use a generalized version of words: words indexed by a linear ordering. Given a linear ordering $\alpha = (L, <)$ and a finite alphabet A, an *α-word u over the alphabet A* is a mapping from L to A. We also say that α is the *domain* of the word u, or that u is a word *indexed* by α. Below, we always consider words up to isomorphism of their domain (unless a specific presentation of the domain is required). Standard finite words are simply the words indexed by finite linear orderings. The set of finite words over an alphabet A is denoted A^*. The set of words indexed by countable scattered linear orderings is denoted $A^\diamond$. Words in $A^\diamond$ are also called *$\diamond$-words*.

Given a word u of domain α and $\beta \subseteq \alpha$, we denote by $u|_\beta$ the word u restricted to its positions in β. Given an α-word u and a β-word v, uv represents the $(\alpha+\beta)$-word defined by $(uv)(x)$ is $u(x)$ if x belongs to α and $v(x)$ if x belongs to β. The product is extended to languages of words in a natural way. The product of words is naturally generalized to the infinite product $\prod_{i\in\alpha} u_i$, where α is an order type and each u_i for $i \in \alpha$ is a β_i-word; the result being a $(\sum_{i\in\alpha} \beta_i)$-word. For a language W and a linear ordering α, one defines W^α to be the language containing all the words $\prod_{i\in\alpha} u_i$, where $u_i \in W$ for all $i \in \alpha$.

2.3. Semigroups and additive labelings

For a thorough introduction to semigroups, we refer the reader to [21, 26]. A *semigroup* $(S, \cdot)$ is a set S equipped with an associative binary operator written multiplicatively. Groups and monoids are particular instances of semigroups. The set of nonempty finite words A^+ over an alphabet A is a semigroup – it is the semigroup freely generated by A. A *morphism of semigroups* from a semigroup $(S, \cdot)$ to a semigroup $(S', \cdot')$ is a mapping φ from S to S' such that for all x, y in S, $\varphi(x \cdot y) = \varphi(x) \cdot' \varphi(y)$. An *idempotent* in a semigroup is an element e such that $e^2 = e$.

Let α be a linear ordering and $(S, \cdot)$ be a semigroup. A mapping σ from ordered pairs $(x, y) \in \alpha^2$ with $x < y$ to S is called an *additive labeling* if for every $x < y < z$ in α, $\sigma(x, y)\sigma(y, z) = \sigma(x, z)$.

Given a semigroup morphism φ from $(A^\diamond, \cdot)$ to some semigroup $(S, \cdot)$ and a word u in $A^\diamond$ of domain α, there is a natural way to construct an additive labeling φ_u from $\overline{\alpha}$ to $(S, \cdot)$: for every two cuts $x < y$ in $\overline{\alpha}$, set $\varphi_u(x, y)$ to be $\varphi(u_{x,y})$, where $u_{x,y}$ is the word u restricted to its positions between x and y; i.e., $u_{x,y} = u|_{F\cap E'}$ for $x = (E, F)$ and $y = (E', F')$.

2.4. Standard results on finite semigroups

In this section, we recall some basic definitions and gather results concerning finite semigroups. The reader can refer to [21, 26] for more details on the subject.

Given a semigroup S, S^1 denotes the monoid S itself if S is a monoid, or the semigroup S augmented with a new neutral element 1 otherwise, thus making S a monoid. The Green's relations are defined by:

$$\begin{array}{llll} a \leq_{\mathcal{L}} b & \text{if} & a = cb \text{ for some } c \text{ in } S^1 & a \mathcal{L} b \quad \text{if} \quad a \leq_{\mathcal{L}} b \text{ and } b \leq_{\mathcal{L}} a \\ a \leq_{\mathcal{R}} b & \text{if} & a = bc \text{ for some } c \text{ in } S^1 & a \mathcal{R} b \quad \text{if} \quad a \leq_{\mathcal{R}} b \text{ and } b \leq_{\mathcal{R}} a \\ a \leq_{\mathcal{J}} b & \text{if} & a = cbc' \text{ for some } c, c' \text{ in } S^1 & a \mathcal{J} b \quad \text{if} \quad a \leq_{\mathcal{J}} y \text{ and } b \leq_{\mathcal{J}} a \\ a \leq_{\mathcal{H}} b & \text{if} & a \leq_{\mathcal{L}} b \text{ and } a \leq_{\mathcal{R}} b & a \mathcal{H} b \quad \text{if} \quad a \mathcal{L} b \text{ and } a \mathcal{R} b \end{array}$$

Fact 1. *Let a, b, c be in S. If $a \mathcal{L} b$ then $ac \mathcal{L} bc$. If $a \mathcal{R} b$ then $ca \mathcal{R} cb$. For every a, b in S, $a \mathcal{L} c \mathcal{R} b$ for some c iff $a \mathcal{R} c' \mathcal{L} b$ for some c'.*

As a consequence of the last equivalence, one defines the last of Green's relations:

$$\begin{array}{lll} a \mathcal{D} b & \text{iff} & a \mathcal{L} c \mathcal{R} b \text{ for some } c \text{ in } S\,, \\ & \text{iff} & a \mathcal{R} c' \mathcal{L} b \text{ for some } c' \text{ in } S\,. \end{array}$$

From now we assume that the semigroup $(S, \cdot)$ is *finite*. This assumption is mandatory for the correctness of what follows.

Fact 2. $\mathcal{D}{=}\mathcal{J}$.

For this reason, we refer from now on only to $\mathcal{D}$ and not $\mathcal{J}$. However, we will use the preorder $\leq_{\mathcal{J}}$ (which is an order over the $\mathcal{D}$-classes).

An element a in S is called *regular* if $asa = a$ for some s in S. A $\mathcal{D}$-class is *regular* if all its elements are regular.

Fact 3. *A $\mathcal{D}$-class D is regular, iff it contains an idempotent, iff every $\mathcal{L}$-class in D contains an idempotent, iff every $\mathcal{R}$-class in D contains an idempotent, iff there exists a, b in D such that $ab \in D$.*

Fact 4. *For every a, b in a $\mathcal{D}$-class D such that $ab \in D$, then $a \mathcal{R}\ ab\ \mathcal{L}\ b$. Furthermore, there is an idempotent e in D such that $a \mathcal{L}\ e\ \mathcal{R}\ b$.*

Fact 5. *For all a, b, c in S such that ab, b, and bc belong to the same $\mathcal{D}$-class D, then $abc \in D$.*

Fact 6 (from Green's lemma). *All $\mathcal{H}$-classes in a $\mathcal{D}$-class have the same size.*

Fact 7. *Let H be an $\mathcal{H}$-class in S. Either for all a, b in H, $ab \notin H$; or for all a, b in H, $ab \in H$, and furthermore $(H, \cdot)$ is a group.*

3. Simon's factorization forest theorem: a new proof

In this section, we first give the original statement of the theorem of factorization forest of Simon (Section 3.1). In Section 3.2, we introduce the notion of a split and use it in a different presentation of the result yielding better bounds. In Section 3.3, we establish the result. Some lemmas in the proof are used again in the sequel of the paper.

3.1. Factorization forest theorem

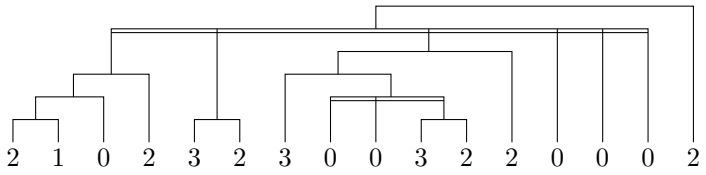


Figure 1: A factorization tree

Fix an alphabet A and a semigroup morphism φ from A^+ to a finite semigroup $(S, \cdot)$. A *factorization tree* is an ordered unranked tree in which each node is either a leaf labeled by a letter, or an internal node. The *value* of a node is the word obtained by reading the leaves below from left to right. A *factorization tree* of a word $u \in A^+$ is a factorization tree of value u. The *height* of the tree is defined as usual, with the convention that the height of a single leaf is 0. A factorization tree is *Ramseyan* (for φ) if every node 1) is a leaf, or 2) has two children, or, 3) the values of its children are all mapped by φ to the same idempotent of S.

Example 8. *We provide here an example of a semigroup which is in fact a group, though this case cannot be considered as representative of the problem in its generality. Fix* $A = \{0, 1, 2, 3, 4\}$, $(S, \cdot) = (\mathbb{Z}/5\mathbb{Z}, +)$ *and* φ *to be the only semigroup morphism from* A^+ *to* $(S, \cdot)$ *mapping each letter to its value. Figure 1 presents a Ramseyan factorization tree for the word*

$$u = 2102323003220002 \ .$$

In this drawing, internal nodes appear as horizontal lines. Double lines correspond to case 3 in the definition of Ramseyness.

The theorem of factorization forests is then the following.

Theorem 1 (factorisation forests [32]). *For every alphabet* A*, finite semigroup* $(S, \cdot)$*, semigroup morphism* φ *from* A^+ *to* S *and word* u *in* A^+*,* u *has a Ramseyan factorization tree of height at most* $9|S|$*.*

The original theorem of Simon [32], gives the bound of $9|S|$. An improved bound of $7|S|$ has been established by Chalopin and Leung [12]. A value of $3|S|$ is a byproduct of the present work (see Remark 10 in Section 3.2). Another proof of the $3|S|$ bound has been independently obtained by Kufleitner [19]. This line of improvements reaches its end in a paper of Kufleitner to appear: a new proof of Simon's result is given with the improved upper bound of $3|S| - 1$, together with a proof of optimality of this bound:

Theorem 2 ([20]). *For every alphabet* A*, finite semigroup* $(S, \cdot)$*, semigroup morphism* φ *from* A^+ *to* S *and word* u *in* A^+*,* u *has a Ramseyan factorization tree of height at most* $3|S| - 1$*. Furthermore, this bound is tight when* $(S, \cdot)$ *is a group; i.e., there exists a word* u *such that every Ramseyan factorization tree for* u *has height at least* $3|S| - 1$*.*

Let us finally mention that in this paper the specific case of aperiodic semigroups is also treated. In this more restricted situation, Ramseyan factorization trees of height at most $2|S|$ always exist, and this bound is tight.

3.2. A variant via Ramseyan splits (for ordinals)

The variant of the factorization forest theorem presented here (Theorem 3) uses the notion of splits. We use this formalism in the sequel of the paper. Our result is slightly more general than Theorem 1 above since it can be applied not only to finite words, but more generally to words indexed by ordinals (though the presentation is not given in terms of words).

The reason for introducing Ramseyan splits in replacement of Ramseyan factorization trees is that this is an object much easier to manipulate by a word automaton. We advocate the use of splits in this context. It simplifies also slightly the proof of Simon's theorem by avoiding to work with the structure of trees. Finally, the extension to the general case (in which words are not necessary finite) is natural with splits, while the original presentation of the result would require more care.

For N a non-negative integer, a *split of height* N of a linear ordering α is a mapping s from α to $[1, N]$ (N can be null, and in this case α has to be empty). Given a split, two elements x and y in α such that $s(x) = s(y) = k$ are *k-neighbors* if $s(z) \geq k$

for all $z \in [x, y]$. k-neighborhood is an equivalence relation over $s^{-1}(k)$. A class of k-neighborhood is also called a *k-class*.

Fix an *additive labeling* σ from α to some finite semigroup S. A split of α is *Ramseyan* for σ — we also say a *Ramseyan split for* (α, σ) — if for every k-class X, there exists an idempotent e such that $\sigma(x, y) = e$ for all $x < y$ in X.

Example 9. *Let S be $\mathbb{Z}/5\mathbb{Z}$ equipped with the addition $+$. Consider the linear ordering of 17 elements and the additive labeling σ defined by:*

| 2 | 1 | 0 | 2 | 3 | 2 | 3 | 0 | 0 | 3 | 2 | 2 | 0 | 0 | 0 | 2 |

Each symbol '|' represents an element, the elements being ordered from left to right. Between two consecutive elements x and y is represented the value of $\sigma(x, y) \in S$. In this situation, the value of $\sigma(x, y)$ for every $x < y$ is uniquely defined according to the additivity of σ: it is obtained by summing all the values between x and y modulo 5.

A split s of height 3 is the following, where we have written above each element x the value of $s(x)$:

1	3	2	2	1	2	1	2	2	2	3	2	1	1	1	1	2
|	2	1	0	2	3	2	3	0	0	3	2	2	0	0	0	2	|

In particular, if you choose $x < y$ such that $s(x) = s(y) = 1$, then the sum of elements between them is 0 modulo 5. If you choose $x < y$ such that $s(x) = s(y) = 2$ but there is no element z in between with $s(z) = 1$ — i.e., x and y are 2-neighbors — the sum of values separating them is also 0 modulo 5. Finally, it is impossible to find two distinct 3-neighbors in our example.

Our variant of Theorem 1 is phrased as follows (this time for every ordinal).

Theorem 3. *For every ordinal α, every finite semigroup $(S, \cdot)$ and additive labeling σ from α to S, there exists a Ramseyan split for (α, σ) of height at most $|S|$. Furthermore this bound is tight for $(S, \cdot)$ a group and α a finite linear ordering.*

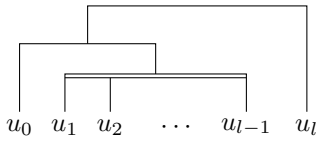
The proof of this theorem is the subject of Section 3.3. The link with Ramseyan factorization forests, as well as the optimality of the bound, are considered in the following remark.

Remark 10. *Fix an alphabet A, a semigroup S, a morphism φ from A^+ to S and a word $u \in A^+$ of finite domain α. The Ramseyan factorization trees and the Ramseyan splits are linked as follows:*

- *every Ramseyan factorization tree of height k of u can be turned into a Ramseyan split of height at most k for $(\overline{\alpha}^{]\![}, \varphi_u)$,*
- *every Ramseyan split of height k for $(\overline{\alpha}^{]\![}, \varphi_u)$ can be turned into a factorization tree of height at most $3k$ of u.*

For the first item, we set the value of the split for $x \in \overline{\alpha}^{]\![}$, say for x the cut between letter i and letter $i + 1$ in u, to be the maximal depth of a node that has the ith and the $(i + 1)$th letter below it. It is not difficult to see that this defines a split of height at most k, and that it is Ramseyan for $(\overline{\alpha}^{]\![}, \varphi_u)$.

For the second item, one remarks that the only 1*-class (we assume that there is one) factorizes the word* u *into* $u = u_0u_1 \dots u_l$ *in such a way that* $\phi(u_1) = \cdots = \phi(u_{l-1})$ *is an idempotent. Hence we construct the prefix of a tree as:*



and then proceed inductively with the subwords $u_0, \dots, u_l$*. We get at the end a Ramseyan factorization tree, and its height is at most* $3k$*.*

Using the second item together with Theorem 3, we obtain a bound of $3|S|$ *for Theorem 1.*

The optimality part in Theorem 3 is obtained by the same argument. Indeed, assume the bound of $|S|-1$ *was possible in Theorem 3, this would mean that Ramseyan factorization trees of height at most* $3(|S| - 1)$ *would be possible. This contradicts the optimality result of Kufleitner (Theorem 2).*

3.3. Proof of the result in the ordinal case

In this section, we establish Theorem 3. Some of the intermediate results are also used in the proof of Theorem 4.

Hence, let σ denote an additive labeling from some linear ordering α to some finite semigroup $(S, \cdot)$. We denote by β a subordering of α. We slightly abuse the notation, and write (β, σ) for $(\beta, \sigma|_\beta)$ in which $\sigma|_\beta$ is the additive labeling obtained by restricting σ to β. We also denote by $\sigma(\beta)$ the set $\{\sigma(x, y) \ : \ x < y, \ x, y \in \beta\}$.

Aiming at their use in the proof of Theorem 4, the splits constructed by Lemmas 11 and 12 have an extra property of being 1-right. This extra constraint is irrelevant in the ordinal case. A split s of a linear ordering α is called 1-*right* if for every $x \in \alpha$, there exists $y \geq x$ in α with $s(y) = 1$. Hence a split is 1-right if either α has a maximal element and this element has split value 1, or if the elements of split value 1 'reach at limit' the right side of α. The notion of being 1-*left* is obtained by symmetry.

The proof works by studying successively different cases according to Green's relations. The first one is the case of a single regular $\mathcal{H}$-class.

Lemma 11. *Let* H *be an* $\mathcal{H}$*-class in* S *such that* $(H, \cdot)$ *is a group, and* β *be such that* $\sigma(\beta) \subseteq H$*. Then there exists a Ramseyan split of height at most* $|H|$ *of* (β, σ)*.*

Furthermore, the split can be chosen to be 1*-right.*

Proof. Since $(H, \cdot)$ is a group, it is natural to extend the definition of σ over β in the following way. For every x in β, let $\sigma(x, x)$ be 1_H, the neutral element of the group $(H, \cdot)$, and for every $y < x$ in β, let $\sigma(x, y)$ be $\sigma(y, x)^{-1}$, the inverse of $\sigma(x, y)$ in H. As expected, this extended version of σ satisfies for every x, y, z in β, $\sigma(x, z) = \sigma(x, y)\sigma(y, z)$. Let n be a mapping numbering the elements of H from 1 to $|H|$. Fix an element x_0 in β. Let s be defined for all x by $s(x) = n(\sigma(x_0, x))$.

Let us show that s defined this way is indeed a Ramseyan split for σ. Let $x < y$ be such that $s(x) = s(y)$, then $\sigma(x_0, x) = \sigma(x_0, y)$ since n is a injection from H onto $[1, |H|]$. Hence $\sigma(x, y) = \sigma(x, x_0)\sigma(x_0, y) = \sigma(x_0, x)^{-1}\sigma(x_0, y) = 1_H$. Hence, given

$x < y$ and $x' < y'$ such that x, y, x' and z' are k-neighbors, we have $\sigma(x, y) = 1_H = \sigma(x', y') = 1_H^2$.

In order to choose the split to be 1-right, it is sufficient to start from the split described so far, to select a split value l such that for all $x \in \alpha$, there exists $y \geq x$ in α with $s(y) = 1$, and swap the role of the split values 1 and l (this amounts to number differently the elements of H). The split obtained this way is still Ramseyan, and it is furthermore 1-right. □

The second case corresponds to a single regular $\mathcal{D}$-class.

Lemma 12. *Let D be a regular $\mathcal{D}$-class in S, and β be such that $\sigma(\beta) \subseteq D$. Then there exists a Ramseyan split of height at most $|D|$ of (β, σ).*

Furthermore, the split can be chosen to be 1-right.

Proof. For every $x \in \beta$ nonmaximal, set $r(x)$ to be the $\mathcal{R}$-class of $\sigma(x, z)$ for some $z > x$; this value is independent of the choice of z according to Fact 4. Similarly, for every x in β nonminimal, set $l(x)$ to be the $\mathcal{L}$-class of $\sigma(y, x)$ for some $y < x$. If β has a maximal element M, choose $r(M)$ to an $\mathcal{R}$-class be such that $l(M) \cap r(M)$ is a subgroup of S; this is possible according to Fact 3. Similarly if β has a minimal element m, choose $l(m)$ to be an $\mathcal{L}$-class such that $l(m) \cap r(m)$ is a subgroup of S. Set for all x in β, $h(x) = l(x) \cap r(x)$.

We claim that for every x in β, $h(x)$ is a subgroup of S. Indeed, if x is either the minimal or the maximal element of β, this follows from the definition of $r(M)$ and $l(m)$ and Fact 7. Otherwise, there exists y, z such that $y < x < z$. Let a be $\sigma(y, x) \in l(x)$ and b be $\sigma(x, z) \in r(x)$. By Fact 4, since $ab = \sigma(y, z) \in D$, there exists an idempotent e in D such that $a \mathcal{L} e$ and $b \mathcal{R} e$; i.e., $e \in h(x)$. And by Fact 7, $h(x)$ is a subgroup of S. The claim holds.

According to Fact 6, there is a positive integer N such that all $\mathcal{H}$-classes included in D have size N. Let $H_0, \ldots, H_{d-1}$ be the $\mathcal{H}$-classes included in D which are subgroups of S. For $k = 0, \ldots, d-1$, set β_k to be the linear ordering $\{x \in \beta \ : \ h(x) = H_k\}$. By Fact 4, $\sigma(\beta_k) \subseteq H_k$. By Lemma 11, there exists a Ramseyan split s_k for (β_k, σ) of height at most $|H_k| = N$.

We set now for all x in β, $s(x) = kN + s_k(x)$ where k is such that $x \in \beta_k$. Let us establish that s is a Ramseyan split for (β, σ). Let $x < y$ and $x' < y'$ be such that $s(x) = s(y) = s(x') = s(y')$. By definition of s, x, y, x', y' belong to the same β_k. Furthermore, since $s(x) = s(y) = s(x') = s(y')$, we have $s_k(x) = s_k(y) = s_k(x') = s_k(y')$. Hence, by Ramseyness of s_k over (β_k, σ), $\sigma(x, y) = \sigma(x', y') = \sigma(x, y)^2$. We conclude that the mapping s is a Ramseyan split for (β, σ), and its height is at most $dN \leq |D|$.

In order to construct a 1-right split, it is sufficient to choose the correct enumeration of $H_0, \ldots, H_{k-1}$. Indeed, there in an $\mathcal{H}$-class H such that for every $x \in \alpha$, there exists $y \geq x$ in α with $h(x) = H$. We apply the construction as above, but choosing to number the H_i in a way such that $H_0 = H$. Furthermore, we require de split s_0 to be 1-right (this is possible according to Lemma 11). The resulting split is Ramseyan as above. Now it is also 1-right. □

From now, we assume that β is an ordinal. We say that $E \subseteq S$ is $\mathcal{D}$-closed if $x \in E$ and $x \mathcal{D} y$ implies y in E. Given a nonempty ordinal β, one denotes by $\dot{\beta}$ the linear ordering $\beta \setminus \{0_\beta\}$, where 0_β is the minimal element of β. This technique of removing 0_β is a trick for not getting into trouble with non-regular $\mathcal{D}$-classes. Without this precision, the lemma below would not hold for instance if E is a non-regular $\mathcal{D}$-class of size 1 and β has size 2.

Lemma 13. *Let $E \subseteq S$ be a $\mathcal{D}$-closed subset of S and $\beta \subseteq \alpha$ be such that $\sigma(\beta) \subseteq E$. Then there exists a Ramseyan split of height at most $|E|$ for $(\dot{\beta}, \sigma)$.*

Proof. The proof is done by induction on the size of E. If E is empty, then β contains at most one element. Hence $\dot{\beta}$ is empty. We can give a split of height 0 over the empty linear ordering.

Otherwise, let D be a minimal $\mathcal{D}$-class in E (for the $\leq_{\mathcal{J}}$-order). Let $\gamma \subseteq \beta$ be the least set satisfying:

- $0_\beta \in \gamma$,
- if $x \in \gamma$ then $\min\{y > x \ : \ \sigma(x, y) \in D\} \in \gamma$.

It is not difficult to check that the following fact holds.

Fact 14. *For every x, y in β, if $]x, y] \cap \gamma$ is empty, then $\sigma(x, y) \notin D$. If $[x, y] \cap \gamma$ contains at least two elements, then $\sigma(x, y) \in D$.*

Define $\sim$ to be the least equivalence relation over β such that for all $x < y$, if $]x, y] \cap \gamma = \emptyset$ then $x \sim y$. Let η be an equivalence class for $\sim$. By Fact 14, $\sigma(\eta) \cap D = \emptyset$. Hence, one can apply the induction hypothesis and obtain a Ramseyan split $s_{\dot{\eta}}$ for $(\dot{\eta}, \sigma)$ of height at most $|E| - |D|$. Remark that $\dot{\eta} = \eta \setminus \gamma$.

At this point, two cases may happen depending on the cardinal of γ. If γ contains exactly two elements (the case of γ being a singleton is similar), we define $s_{\dot{\beta}}$ over $\dot{\beta}$ by $s(x) = 1$ for $x \in \gamma$, else $s(x) = s_{\dot{\eta}}(x) + 1$ for η the equivalence class of x. This split is Ramseyan since the value 1 is used at most once (in $\dot{\gamma}$), and the Ramseyness is inherited from the induction hypothesis elsewhere. By induction hypothesis, this split has height at most $|E| - |D| + 1 \leq |E|$. Remark that in this case, constructing the split for β instead of $\dot{\beta}$ would require to use another split value for 0_β; thus transforming the $+1$ into a $+2$. This would provide an upper bound for the height of the split of $2|E|$ instead of $|E|$.

Otherwise, γ contains at least three elements, say $x < y < z$. Since $\sigma(x, y)$, $\sigma(y, z)$ and $\sigma(x, y)\sigma(y, z) = \sigma(x, z)$ all belong to D, we deduce by Fact 3 that D is regular. Hence, by Lemma 12, we obtain a Ramseyan split s_γ of height at most $|D|$ for (γ, σ). Then define s over $\dot{\beta}$ by $s(x) = s_\gamma(x)$ for $x \in \gamma$, else $s(x) = |D| + s_\eta(x)$ for η the equivalence class of x. It follows from the definition that s is a Ramseyan split of $(\dot{\beta}, \sigma)$ of height at most $|E| - |D| + |D| = |E|$. □

We can now conclude the proof of Theorem 3. Applying the previous lemma to α would give us a split for $\dot{\alpha}$. What remains to be done is to remove the dot.

Proof. Given an ordinal α, and an additive labeling σ from α to S. Fix a value a_0 in S, construct the linear ordering $\alpha' = 1 + \alpha$, where 1 is a linear ordering containing the single element 0. Set $\sigma'(x, y)$ for $x < y$ in α to be $\sigma(x, y)$, set $\sigma'(0, 0_\alpha) = a_0$, and propagate accordingly for making σ' an additive labeling. By Lemma 13, there exists a Ramseyan split s for $(\dot{\alpha}', \sigma')$ of height at most $|S|$. By construction of α' and σ', s is also a Ramseyan split for (α, σ). □

4. Ramseyan splits, the general case

We generalize Theorem 3 to infinite linear orderings as follows[1].

Theorem 4. *For every linear ordering α, every finite semigroup $(S, \cdot)$ and additive labeling σ from α to S, there exists a Ramseyan split for (α, σ) of height at most $2|S|$.*

Compared to Theorem 3, we trade the ordinal assumption for a bound of $2|S|$ which replaces a bound of $|S|$. We do not know if this bound is tight.

Proof. The proof is by induction on the size of $|S|$. Let D be a maximal $\mathcal{D}$-class in S.

A *D-set* is a subset $X \subseteq \alpha$ such that for all $x < y$ in X, $\sigma(x, y) \in D$. We are interested in *maximal D-sets*: the one maximal for the inclusion. The reason for that is that our construction of the Ramseyan split heavily relies on the structure of the D-sets, and the way these sets cover α.

We first start by establishing some results on the D-sets. Those simple facts are stated in the items below.

1. The convex hull of a D-set is a D-set, and as a consequence, maximal D-sets are convex. Indeed, let $x < y$ belong to Y. By definition of Y, there exists $x' \leq x$ in X and $y' \geq y$ in X. Let us assume that $x' < x$ and $y < y'$, the three other cases being similar. We have $\sigma(x, y) \geq_{\mathcal{J}} \sigma(x', x)\sigma(x, y)\sigma(y, y') = \sigma(x', y') \in D$ by definition of $\leq_{\mathcal{J}}$ and the assumption that X is a D-set. Thus by maximality of D, $\sigma(x, y) \in D$. By consequence, Y is a D-set as claimed.
2. The maximal D-sets can be *totally ordered* by $\prec$, where $X \prec Y$ holds if there exists $x \in X$ with $x < y$ for all $y \in Y$.
3. Every D-set is contained in a maximal D-set. One proof is by a simple use of Zorn's lemma. A theoreticaly simpler argument is the following: consider a nonempty D-set X, and Y, Z defined as follows:

$$\begin{aligned} Y &= \{y \ : \ \exists x{\in}X.\ y \leq x \text{ and } \forall x.\ (x \in X \wedge y < x) \to \sigma(y, x) \in D\} \\ \text{and } Z &= \{z \ : \ \exists y{\in}Y.\ y \leq z \text{ and } \forall y.\ (y \in Y \wedge y < z) \to \sigma(y, z) \in D\}\ . \end{aligned}$$

   By construction and by use of Item 1, it is not difficult to see that Y and Z are convex D-sets. Remark also that $X \subseteq Y \subseteq Z$. For the sake of contradiction, assume now that there exists $z \notin Z$ such that $Z \cup \{z\}$ is a D-set. Two cases can happen, either $z < Z$ or $z > Z$. Let us assume first that $z < Z$. We have $z < X$ since $X \subseteq Z$. We also have that for all $x \in X$, $\sigma(x, z) \in D$. By consequence, we should have $z \in Y \subseteq Z$; a contradiction. The same arguments can be used for $z > Z$. Hence Z is a maximal D-set containing X.
4. The maximal D-sets cover α. Indeed, every singleton satisfy that for all $x < y$ in X, $\sigma(x, y) \in D$, i.e, every singleton is a D-set. Using Item 3, every element belongs to a maximal D-set.

[1] In the conference version of this work [14], a weaker variant of the result is announced, in which the linear ordering is assumed to be complete, and the upper bound on the height is only $3|S|$.

5. Two maximal D-sets having at least two elements in their intersection are equal. What we prove is that given two convex D-sets X, Y that have at least two elements in common, then $X \cup Y$ is also a D-set. Indeed, assume $x < y$ with $x, y \in X \cap Y$ where X, Y are two distinct convex D-sets. Let $x' < y'$ belong to $X \cup Y$. If both x', y' belong to either X or Y, then we have $\sigma(x', y') \in D$ by assumption on X or Y. Otherwise, let us assume wlog. that $x' \in X \setminus Y$ and $y' \in Y \setminus X$. By convexity of Y, this means that $x' < x$, and similarly, $y > y'$. Using Fact 5 on $\sigma(x', x), \sigma(x, y)$ and $\sigma(y, y')$ we get $\sigma(x', y') \in D$. Hence $X \cup Y$ is also a D-set.
6. Each element $x \in \alpha$ belongs to at most two distinct maximal D-sets. Indeed, assume that three distinct convex subsets share one element, then at least two among them must share at least two elements. By Item 5, both can't be maximal simultaneously. Hence, if three convex subsets share one element, then one of them is not maximal.

We now turn ourselves to the construction of the split itself. Let $\beta \subseteq \alpha$ be maximal such that for all D-set X, $|\beta \cap X| \leq 1$. Such a set exists by Zorn's lemma. For every $x < y$ in β, by construction of β, x and y cannot be in the same D-set. Thus $\sigma(x, y) \notin D$, and we obtain $\sigma(\beta) \subseteq S \setminus D$. By induction hypothesis applied to β, we get a Ramseyan split s' for (β, σ), of height at most $2(|S| - |D|)$.

Consider now a maximal (nonempty) convex subset $Z \subseteq \alpha$ which does not intersect β. We aim at constructing a Ramseyan split s_Z for (Z, σ) of height at most $2|D|$ (in fact $|D| + 1$). For this, we would like to use Lemma 12. Unfortunately, Z is not a D-set in general, and hence Lemma 12 cannot be applied directly. For this reason, we start by an analysis of the set Z and show that it has to be contained in the union of two convex D-sets (claim $\star$ below).

Call β^- (resp. β^+) the linear ordering β restricted to elements smaller than Z (resp. greater than Z). By Items 1 and 6, there is at most one maximal D-set that intersects both β^- and Z (resp. β^+ and Z). Let X^+ (resp.X^-) be this maximal D-set if it exists, or else be $\emptyset$. We claim that $Z \subseteq X^- \cup X^+$ ($\star$). For the sake of contradiction consider some $y \in Z \setminus (X^+ \cup X^-)$. Let Y be a maximal D-set containing y (it exists by Item 4). Assume it intersects β^-, then by Item 5, we should have $Y = X^-$: a contradiction since $y \in Y \setminus X^-$. Similarly, it cannot intersect β^+. Hence, $Y \subseteq Z$. We obtain that for all D-set X, $|(\beta \cup \{y\}) \cap X| \leq 1$. This contradicts the maximality of β. This concludes the proof of Claim $\star$.

If Z contains exactly two elements, say y, z (the case of Z being a singleton is not different), then s_Z defined by $s_Z(y) = 1$ and $s_Z(z) = 2$ is a Ramseyan split for (Z, σ) of height at most $2 \leq |D| + 1$.

Otherwise Z contains at least three elements. By Claim $\star$, $Z \subseteq X^- \cup X^+$. Thus either $X^- \cap Z$ or $X^+ \cap Z$ has size at least 2. Wlog., let us assume this is X^-. Since furthermore X^- intersects by construction β^- while Z doesn't, X^- has cardinality at least 3. Let $x < y < z$ be three elements in X^-, one has $\sigma(x, y) \in D$, $\sigma(y, z) \in D$ and $\sigma(x, y) \cdot \sigma(y, z) = \sigma(x, z) \in D$. Hence by Fact 3, D is a regular. This menas that one can use Lemma 12 and obtain a 1-right Ramseyan split s^- for (X^-, σ) of height at most $|D|$. Symetricaly, by Lemma 12, we obtain a 1-left Ramseyan split s^+ for $(X^+ \setminus X^-, \sigma)$ of

height at most $|D|$. We set s_Z to be:

$$s_Z(x) = \begin{cases} s^-(x) & \text{for } x \text{ in } Z \cap X^- \\ s^+(x) + 1 & \text{for } x \text{ in } Z \cap X^+ \setminus X^- \end{cases}$$

By convexity of Z, X^-, and X^+, and using the fact that s^- is 1-right and s^+ is 1-left, we obtain that s_Z is a Ramseyan split for (Z, σ) of height at most $|D| + 1$.

We can now construct the split s by:

$$s(x) = \begin{cases} s'(x) & \text{if } x \in \beta, \\ s_Z(x) + 2(|S| - |D|) & \text{otherwise, where } Z \text{ is the maximal convex} \\ & \text{non-intersecting } \beta \text{ and containing } x. \end{cases}$$

By construction, this split is Ramseyan, and has height at most $2|S|$. □

5. Application to countable scattered linear orderings

The subject of this section is to give a new simplified proof of the equivalence between the acceptance by automata and the recognizability by finite $\diamond$-semigroups of languages of words indexed by countable scattered linear orderings. We first give an overview of this field.

5.1. Overview of the field

This line of research pursue the sequence of works establishing the equivalence for a language of finite words L between the following items:

1. L is described by a regular expression,
2. L is accepted by a deterministic automaton,
3. L is recognized by a morphism of monoid/semigroup,
4. L is accepted by a non-deterministic automaton,
5. L is definable in monadic second-order logic.

Those equivalences are well known in formal language theory. The equivalence between Item 1 and 2 is the famous theorem of Kleene [18]. Myhill has established the equivalence between Items 2 and 3 (more precisely, Rabin and Scott give credit to Myhill for that). Rabin and Scott have introduced in their seminal paper the notion of non-deterministic automaton and presented the modern view on the equivalences of all items up to 4 [30].

The languages satisfying the condition above are called regular. From either Item 1 or Item 4, one easily gets that regular languages are closed under union, intersection and projection (i.e., image under a letter to letter morphism). From either Item 2 or 3, one obtains easily the closure under union, intersection and complement.

The monadic second-order logic (MSO for short) is the extension of first-order logic with set-quantification. Each formula in this logic describes the set of words that satisfies it. Item 5 states the equivalence between this form of description and regular languages. One direction can be established using the closure properties of regular languages under union, intersection, complement and projection. The other direction consists in encoding

the semantic of automata in MSO, and is natural. This equivalence was used by Büchi in his proof of the decidability of the satisfaction problem for MSO over finite words [9].

Those equivalences have been extended to various other settings. We are interested here only in the extension to transfinite words: languages of words that are indexed by possibly infinite linear orderings.

The starting point in this direction is the work of Büchi [10]. In this work, Büchi extends non-deterministic automata to words of length ω. A new acceptance condition is required: a run is accepting if it visits infinitely often some fixed set of states (the so-called Büchi acceptance condition). The languages accepted by such automata happen to be closed under union, intersection, projection, and complement (the difficult part). This is sufficient for establishing the link with MSO, i.e., the implication from Item 4 to Item 5 (the other direction of the implication being once more simple). The equivalence with a suitable form of regular expression is also natural, yielding an equivalence between Item 1 and Item 4. Automata using the Büchi acceptance condition cannot be determinized in general. Müller introduced independently a more expressive form of acceptance conditions [23] (a Müller condition consists of a set of sets of states F, and a run is accepting if the set of states visited infinitely often belongs to F). McNaughton showed that automata using a Büchi acceptance condition can be transformed into deterministic automata using a Müller acceptance condition [22], thus establishing the equivalence between Items 2 and 4. The corresponding algebraic notions are Wilke algebras [33] and ω-semigroups [24]. Those two objects are of different nature: Wilke algebras provide a finite model for defining regular sets of infinite words, while ω-semigroups more precisely reflect the algebraic structure of words of length ω (the set of infinite words of length ω is the free ω-semigroup, while the free Wilke algebra is smaller and contains only ultimately periodic infinite words). The link between those two notions is that every ω-semigroup induces a Wilke algebra, while every finite Wilke algebra can be uniquely prolongated into an ω-semigroup. Those works complete the task of extending the equivalences above to words of length ω. The interested reader can refer to [25] for more detailed information.

The next step consists in extending the model to ordinals beyond ω. The first work in this direction is also due to Büchi who introduced the notion of deterministic automata running on words of countable ordinal length [8]. Those automata used, apart from standard transitions of finite word automata of the form $Q \times A \to Q$ (in which Q is the set of states and A the alphabet), limit transitions of the form $\mathcal{P}(Q) \to Q$. Those transitions are fired when encountering a limit ordinal, based on the set of states that appears infinitely close to the limit (i.e., the set of states reached cofinaly in the past). Büchi establishes that the languages accepted by such automata are closed under union, intersection, complement and projection, and thus are at least as expressive as MSO (more precisely, the two formalisms are equi-expressive). Since a non-deterministic automaton can be seen as accepting the projection of the language of its accepting runs (which is accepted by a deterministic automaton), this work establishes in fact the link between Item 2 and 4. From the closure properties we get the implication from Item 5 to Item 2. The corresponding regular expressions have been studied by Wojciechowski [34] who established their equivalence with automata. The algebraic view of regularity over countable ordinals was first studied in [3] through the use of ω^n-semigroups, which recognize languages of words indexed by ordinals less than ω^{n+1}. Those correspond to the expressive power of Choueka automata [13]. Finally, the algebraic structure of ω_1-semigroups introduced by Bedon and Carton gives a notion of recognizability suitable

for languages of words indexed by countable ordinals, and is equivalent to the automata introduced by Büchi for ordinals [4].

Though automata running over uncountable ordinals have been considered – e.g. by Büchi –, the properties of such automata are not as good: the closure under complement is lost, and as a consequence the equivalence with MSO does not remain either.

Beyond ordinals are scattered linear orderings: linear orderings that are nowhere dense. Bruyère and Carton introduced regular expressions for words indexed by countable scattered linear orderings [7], which happen to be equivalent to a natural extension of non-deterministic ordinal automata by limit transitions from the right. Hence the equivalence holds between 1 and 4. In the context of scattered linear orderings, it is not clear what a deterministic automaton does mean. For this reason, we do not consider the equivalence with item 2. The algebraic variant is due to Carton and Rispal and goes through the use of $\diamond$-semigroups [11]. From those equivalences, every language definable in MSO over countable scattered linear orderings is accepted by an automaton. The converse also holds [2], and is more involved than in the finite case.

Automata running over words indexed by linear orderings beyond countable scattered ones have been studied. For instance, the equivalence between Items 1 and 4 has been extended to all linear orderings in [2]. However, the family of automata/expressions described in the corresponding models are not closed under complement, i.e., this defines classes of languages for which there is no hope to have an equivalence with Items 3 and 5.

From the equivalence with Item 5 over countable scattered linear orderings, we get the decidability of the satisfiability of MSO over countable scattered linear orderings. However, there is another way to obtain this decidability result, namely by an elementary reduction to MSO over the rationals. Indeed every countable linear ordering is isomorphic to a subset of $(\mathbb{Q}, <)$. Furthermore, among those subsets, the one that are scattered are definable in MSO. Hence every satisfaction problem for MSO over countable scattered linear orderings is reducible to the question whether $(\mathbb{Q}, <)$ models some MSO formula. This latter problem is known to be decidable from the famous theorem of Rabin [29], and also using the compositional method developed by Gurevich and Shelah in this context [15, 16].

For this reason the interest of this study is not the decidability of the MSO theory of countable scattered linear orderings itself, but rather the development of an automata-theoretic comprehension of it. In particular the proof using the theorem of Rabin makes use of tree automata, which is a much more involved object. Using automata running directly over linear orderings is thus interesting by itself.

The contribution of this section is a new simplified proof of the translation from $\diamond$-semigroups to automata over countable scattered linear orderings. Our new proof relies on the use of Theorem 4. Le us remark that our result is in fact slightly stronger than the one in [11]. The difference is that our model of automata is *a priori* weaker than the automaton used in the original proof. In [11], the limit transitions follow a Müller acceptance condition, i.e., limit transitions are fired based on the set of states appearing infinitely close to the limit. In the present work, priority conditions are used: each state has a priority, i.e., a nonnegative integer, and limit transitions are fired based on the highest priority appearing infinitely close to the limit. This makes it more ressemblant to the parity condition used in games and tree automata.

The remaining of this section is organized as follows. We first introduce automata over countable scattered linear orderings (Section 5.2), then we present $\diamond$-semigroups

(Section 5.3). In Section 5.4 we present a useful lemma concerning scattered linear orderings. We finally establish the equivalence between automata and $\diamond$-semigroups in Section 5.5.

5.2. Automata over countable scattered linear orderings

In this section, we define priority automata and show how they accept words indexed by countable scattered linear orderings. Similar automata were introduced in [7], but with a different definition for limit transitions. While the automata in [7] use limit transitions that resemble the Müller acceptance condition, our model uses transitions that are similar to the parity accepting condition. The notion of priority automaton defined just below is new to this respect.

Definition 15. *A* priority automaton $\mathcal{A} = (Q, A, I, F, p, \delta)$ *consists of a finite set of* states Q*, a finite alphabet* A*, a set of* initial states I*, a set of final states* F*, a* priority mapping $p : Q \mapsto [1, N]$ *(*N *being a nonnegative integer) and a* transition relation $\delta \subseteq (Q \times A \times Q) \cup ([1, N] \times Q) \cup (Q \times [1, N])$.

A *run* of the automaton $\mathcal{A}$ over an α-word u is a mapping ρ from $\overline{\alpha}$ to Q such that for all cuts c, c':

- if c' is the successor of c through x, then $(\rho(c), u(x), \rho(c')) \in \delta$,
- if c is a limit from the left, then $(k, \rho(c)) \in \delta$ for $k = \max \bigcap_{c'<c} p(\rho(]c', c[))$,
- if c is a limit from the right, then $(\rho(c), k) \in \delta$ for $k = \max \bigcap_{c'>c} p(\rho(]c, c'[))$.

The first case corresponds to standard automata on finite words: a transition links one state to another while reading a single letter in the word. The second case checks that the highest priority appearing infinitely close to the left of c is allowed by the transition relation. The third case is symmetric. An α-word u is *accepted* by $\mathcal{A}$ if there is a run ρ of $\mathcal{A}$ over u such that $\rho(\bot) \in I$ and $\rho(\top) \in F$.

Example 16. *Consider the priority automaton with states* q, r*, both of priority* 0*, alphabet* $\{a\}$*, initial states* $\{q, r\}$*, final state* q*, and transitions* $\{(q, a, q), (q, a, r), (0, q), (r, 0)\}$*. It accepts those words in* $\{a\}^\diamond$ *which have a complete domain. For this, note that a linear ordering is complete iff no cut is a limit simultaneously from the left and from the right.*

Consider a word $u \in \{a\}^\diamond$ *which has a complete domain* α*. For* $c \in \overline{\alpha}$*, set* $\rho(c)$ *to be* q *if* c *is* $\top$ *or if* c *has a successor, else* $\rho(c)$ *is* r*. Under the hypothesis of completeness, it is simple to verify that* ρ *is a run witnessing the acceptance of the word. Conversely, assume that there is a run* ρ *over the* α*-word* u *with* α *not complete. There is a cut* $c \in \overline{\alpha}$ *which is both a limit from the left and from the right. If* $\rho(c)$ *is* r*, then, as* c *is a limit from the left and there is no corresponding transition from its left, there is a contradiction; else* $\rho(c)$ *is* q*, and the same argument can be applied from the right of* c*. In both cases there is a contradiction.*

It is easy to prove that the languages of $\diamond$-words accepted by priority automata are closed under union, and projection. The closure under intersection is simple in the model used in [11], while it is not easy in our model. It is also easy to establish the decidability of the emptiness problem. Below, after introducing the notion of $\diamond$-semigroup, we show the more difficult closure under complement.

5.3. On $\diamond$-semigroups

We present in this section $\diamond$-semigroups and the corresponding recognizable languages.

Formally, a *$\diamond$-semigroup* (S, π) is a set S equipped with an operator π mapping $S^\diamond$ to S, and which satisfies:

- for all $s \in S$, $\pi(s) = s$, and,
- for all countable scattered linear orderings α and families $(u_i)_{i\in\alpha}$ of words in $S^\diamond$,

$$\pi\left(\prod_{i\in\alpha} \pi(u_i)\right) = \pi\left(\prod_{i\in\alpha} u_i\right) .$$

Those properties express the fact that π is a generalized product operator: more precisely, the rules correspond to a generalized form of associativity. For instance, for every u, v, w in S, $\pi(u\pi(vw)) = \pi(uvw) = \pi(\pi(uv)w)$. In this sense, every $\diamond$-semigroup can be seen as a semigroup with the product defined by $u \cdot v = \pi(uv)$. The free $\diamond$-semigroup generated from a finite alphabet A is $(A^\diamond, \prod)$.

Given two $\diamond$-semigroups (S, π) and (S', π'), a mapping φ from S to S' is a *morphism of $\diamond$-semigroups* if for every scattered linear ordering α, and every $(x_i)_{i\in\alpha}$ in S, $\varphi(\pi(\prod_{l\in\alpha} x_l)) = \pi'(\prod_{l\in\alpha} \varphi(x_l))$. A language $K \subseteq A^\diamond$ is *$\diamond$-recognizable* if there exists a morphism of $\diamond$-semigroups from $A^\diamond$ to a finite $\diamond$-semigroup saturating K; i.e., such that $\varphi^{-1}(\varphi(K)) = K$. As usual with recognizability, $\diamond$-recognizable languages are closed under union, intersection and complement.

From now, we denote $\pi(uv)$ simply by uv. More generally, given a word u in $S^\diamond$, we do not distinguish between u and $\pi(u)$. Similarly, we abbreviate $\pi(\prod_{i\in(\mathbb{N},<)} u)$ by u^ω and $\pi(\prod_{i\in(-\mathbb{N},<)} u)$ by $u^{-\omega}$. We also denote by u^ζ the value $u^{-\omega}u^\omega$.

The following theorem establishes that a finite $\diamond$-semigroup is entirely described by the semigroup it induces together with the exponent ω and $-\omega$. This result was announced by Zoltán Ézik during his invited talk at DLT 02. The proof is due to Carton and Rispal [11].

Theorem 5 (Theorem 10 in [11]). *Given a finite semigroup $(S, \cdot)$, and mappings $r : S \to S$, and $l : S \to S$ satisfying for all $a, b \in S$ and positive integer n:*

$$r(a \cdot b) = a \cdot r(b \cdot a) , \qquad l(a \cdot b) = l(b \cdot a) \cdot b ,$$
$$r(a^n) = r(a) , \qquad \text{and} \quad l(a^n) = l(a) ,$$

there exists a unique $\diamond$-semigroup (S, π) which coincides with $S, \cdot$ as a semigroup, and such that $s^\omega = r(s)$ and $s^{-\omega} = l(s)$ for all $s \in S$.

The proof of this result is technical and tedious since both the existence and the uniqueness of the $\diamond$-semigroup have to be established. This theorem is in fact an extension to the case of scattered linear orderings of Wilke's result sating that every Wilke algebra can be uniquely extended into an ω-semigroup.

Example 17. *Consider the set* $S = (\{0,1\} \times \{0,1\}) \cup \{\perp\}$. *Define the product* . *and the exponent mappings* ω *and* $-\omega$ *by, for every* x *in* S *and* a, b, a', b' *in* $\{0,1\}$,

$$\perp x = x \perp = \perp \qquad (a,b)(a',b') = \begin{cases} \perp & \text{if } b = a' = 1 \\ (a,b') & \text{otherwise} \end{cases}$$

$$\perp^{\omega} = (1,1)^{\omega} = \perp \qquad (a,b)^{\omega} = \begin{cases} \perp & \text{if } a = b = 1 \\ (a,1) & \text{otherwise} \end{cases}$$

$$\perp^{-\omega} = (1,1)^{-\omega} = \perp \qquad (a,b)^{-\omega} = \begin{cases} \perp & \text{if } a = b = 1 \\ (1,b) & \text{otherwise.} \end{cases}$$

By Theorem 5, this $(S, \cdot)$ *together with the mappings* ω *and* $-\omega$ *defines uniquely a* $\diamond$*-semigroup* (S, π).

Let u *be in* $\{a\}^{\diamond}$ *of domain* α. *Set* $\varphi(u)$ *to be* $\perp$ *if* α *is not complete. If* α *is complete, set* $\varphi(u)$ *to be* (a,b) *where* $a = 0$ *if* α *has a minimal element, else* $a = 1$, *and* $b = 0$ *if* α *has a maximal element, else* $b = 1$. *This* φ *is a morphism from* $(\{a\}^{\diamond}, \prod)$ *to* (S, π). *It follows that the set of words in* $\{a\}^{\diamond}$ *of complete domain is* $\diamond$*-recognizable: it is equal to* $\varphi^{-1}(\{0,1\} \times \{0,1\})$.

5.4. A lemma for scattered linear orderings

The subject of this section is to establish Lemma 18 which is a convenient way to prove results on scattered linear orderings.

Let X, Y be two nonempty subsets of a linear ordering α with $X < Y$; one says that X and Y are *contiguous* if there is no z such that $X < z < Y$. The following lemma is used for proving the correctness of the construction in Section 5.

Lemma 18. *Given a scattered linear ordering* α *and an equivalence relation* R *over* α *satisfying:*

for all $X < Y$ *contiguous subsets of* α,

$$X^2 \subseteq R \text{ and } Y^2 \subseteq R \quad \text{implies} \quad (X \cup Y)^2 \subseteq R;$$

Then $R = \alpha^2$.

Proof. Consider the set S of equivalence relations included in R such that every equivalence class is convex. It is nonempty since the equality relation over α belongs to S. Order S by inclusion. Given a chain in S, the union of all relations in the chain is itself an element of S: the chain has an upper bound in S. Then, according to Zorn's lemma, there is a maximal element $\sim$ in S. Since α is scattered and $\sim \in S$, $\alpha/_{\sim}$ is itself a scattered linear ordering. Assume that it has at least two distinct equivalence classes. Since $\alpha/_{\sim}$ is scattered, there exist two contiguous equivalence classes $X < Y$. Applying the hypothesis leads to $(X \cup Y)^2 \subseteq R$, and consequently $\sim \subsetneq (\sim \cup (X \cup Y)^2) \in S$. It contradicts the maximality of $\sim$. □

5.5. Equivalence between automata and recognizability

We now aim at:

Theorem 6 ([11]). *A language of $\diamond$-words is $\diamond$-recognizable iff it is accepted by a priority automaton.*

Let us remark once more that the proof here is a bit more general since our automata are parity-like, while the original proof uses Müller-like automata. Since recognizability entails directly closure under complement, we obtain as a corollary.

Corollary 19. *The languages of $\diamond$-words accepted by priority automata are closed under complement.*

The proof from right to left of Theorem 6 is standard (see [11]), we do not establish it here. Let us establish the other direction. We fix ourselves a $\diamond$-semigroup (S, π), a morphism of $\diamond$-semigroups φ from $(A^\diamond, \prod)$ to (S, π) and a language $L = \varphi^{-1}(\varphi(L))$. Let k be a non-negative integer, a be in S, and ι be among $[], [[,]],]$[, set $S_k^\iota(a)$ to be the set of $\diamond$-words u such that $\varphi(u) = a$ and $(\overline{\alpha}^\iota, \varphi_u)$ admits a Ramseyan split of height k (by convention, ε does not belong to $S_k^{]][}(a)$). We claim the following.

Claim: For every non-negative integer k, a in S and $\iota = [], [[,]],]$[, $S_k^\iota(a)$ is accepted by a priority automaton.

Assuming the claim, since by Theorem 4, $\varphi^{-1}(a) = S_{2|S|}^{[]}(a)$, we deduce that $\varphi^{-1}(a)$ is accepted by a priority automaton. Since furthermore L is a finite union of such languages and priority automata are closed under union, this concludes the proof.

The claim itself is obtained by an induction on k. For the base case $k = 0$ one remarks that the following languages are accepted by priority automata (recall that $S_0^{]][}(a)$ does not contain ε by convention):

$$S_0^{]][}(a) = \varphi^{-1}(a) \cap A\,, \qquad S_0^{[[}(a) = S_0^{]]}(a) = \varphi^{-1}(a) \cap \{\varepsilon\}\,, \qquad \text{and} \quad S_0^{[]}(a) = \emptyset\,.$$

Let now $k \geq 1$. Given an idempotent e in S, set $C_{e,k}$ to be the set of $\diamond$-words u of domain α such that $\varphi(u) = e$, and there exists a Ramseyan split s of height k of $\overline{\alpha}$ such that $s(\bot) = s(\top) = 1$.

Our first step is to show how to construct an automaton accepting $C_{e,k+1}$ from automata accepting the languages $S_k^\iota(a)$. For this, consider the following languages:

$$M_{e,k} = S_k^{]][}(e)\,, \qquad M_{e,k}^{\leftarrow} = \sum_{ae^{-\omega}=e} S_k^{]]}(a),$$

$$M_{e,k}^{\rightarrow} = \sum_{e^\omega a=e} S_k^{[[}(a), \qquad \text{and } M_{e,k}^{\rightarrow\leftarrow} = \sum_{e^\omega ae^{-\omega}=e} S_k^{[]}(a).$$

By induction hypothesis, those languages are accepted by priority automata, namely $\mathcal{M}_{e,k}, \mathcal{M}_{e,k}^{\leftarrow}, \mathcal{M}_{e,k}^{\rightarrow}$ and $\mathcal{M}_{e,k}^{\rightarrow\leftarrow}$. Wlog., we choose them to use distinct priorities, and we set $n-1$ to be the maximal priority involved in those automata. We use them in the construction of the automaton $\mathcal{A}_{e,k+1}$ depicted Fig. 2.

This construction makes use of ε-transitions. Those transitions are here just for clarifying the presentation, and can be removed using standard techniques. The automaton

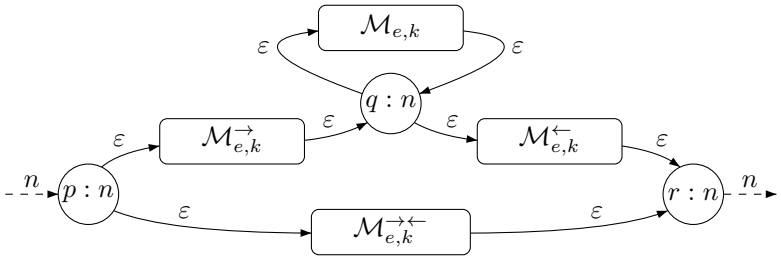


Figure 2: The automaton $\mathcal{A}_{e,k+1}$

itself is made of disjoint copies of the automata accepting $\mathcal{M}_{e,k}, \mathcal{M}^{\rightarrow\leftarrow}_{e,k}, \mathcal{M}^{\rightarrow}_{e,k}$, and $\mathcal{M}^{\leftarrow}_{e,k}$, together with three new states p, q, r. Each ε-transition entering one of the subautomata represents in fact all possible ε-transitions with as destination an initial state of this automaton; similarly, every ε-transition exiting a subautomaton represents all possible ε-transitions with as origin any of the final states of the automaton. The priority of the new state q is n, a priority unused elsewhere by construction. One chooses also p and r to have priority n (this is not of real importance since it is impossible to see infinitely often p or r in a run without seeing infinitely often q: the priority of q only matters). The two dashed arrows represent the two limit transitions (n, p) and (r, n).

Let $L_{e,k+1}[q_1, q_2]$ be the language accepted by this automaton with initial state q_1 and final state q_2 for q_1, q_2 among p, q, r.

The core of the proof is embedded in the following lemma.

Lemma 20. *For every idempotent e, $L_{e,k+1}[q, q] = C_{e,k+1}$.*

Proof. We first prove $L_{e,k+1}[q, q] \subseteq C_{e,k+1}$.

Let u be in $L_{e,k+1}[q, q]$, we have to construct a Ramseyan split s of height $k+1$ of $\varphi^{[]}_u$ with $s(\bot) = s(\top) = 1$. Since $u \in L_{e,k+1}[q, q]$, there exists a corresponding run ρ of the automaton $\mathcal{A}_{e,k+1}$ from state q to state q. Let I be the set of cuts c such that $\rho(c) = q$.

Set $s(c) = 1$ for all c in I. Let now $J \subseteq \overline{\alpha}$ be a maximal interval not intersecting I. We aim at defining s over J. Let

$$x = \sup\{z < J \ : \ z \in I\} \qquad \text{and,} \quad y = \inf\{z > J \ : \ z \in I\} \ .$$

Then J is either $[x, y]$, $[x, y[$, $]x, y]$ or $]x, y[$. We only treat the case $J = [x, y[$ which is representative of the other cases. In this case, since $y \notin J$, $y \in I$ and hence $\rho(y) = q$. Furthermore, since $x \in J$, there exists an infinite sequence $x_1 < x_2 < \dots$ of length ω and limit x in I. As the priority of $\rho(x_i) = q$ is the maximal one, namely n, the only possible state for $\rho(x)$ compatible with limit transitions is p. Furthermore the state q is never visited by ρ in $[x, y[$ (by definition of J). By inspecting the automaton, we conclude that the only possibility is that ρ restricted to $[x, y]$ is in fact a run of the subautomaton $\mathcal{M}^{\rightarrow}_{e,k}$. By induction hypothesis, since $M^{\rightarrow}_{e,k}$ is a union of languages $S^{[[}_k$, this means that there exists a split s_J of height k of J, Ramseyan for σ. We set $s(z) = s_J(z) + 1$ for all $z \in J$. For the other possibilities for J, runs of the automata $\mathcal{M}^{\rightarrow}_{e,k}$, $\mathcal{M}^{\leftarrow}_{e,k}$ and $\mathcal{M}^{\rightarrow\leftarrow}_{e,k}$ are involved in a similar way.

We claim now: For every $x < y$ in I, $\varphi_u(x, y) = e$.

Let R be the equivalence relation over I defined by xRy if $x = y$ or $x < y$ and $\sigma(x, y) = e$ or $y < x$ and $\sigma(y, x) = e$. Let $X < Y$ be nonempty contiguous subsets of I with $X^2 \subseteq R$ and $Y^2 \subseteq R$. Let J be

$$\bigcap_{\substack{x \in X \\ y \in Y}}]x, y[.$$

By construction, J is an interval. Furthermore, since X and Y are contiguous *in* I, J does not intersect I. It is also very easy to show that J is a maximal interval nonintersecting I. Let

$$x = \sup\{z < J \ : \ z \in I\} \qquad \text{and,} \quad y = \inf\{z > J \ : \ z \in I\} \ .$$

We perform a case distinction on whether J is $[x, y]$, $[x, y[$, $]x, y]$ or $]x, y[$. Let us treat the case $J = [x, y[$. As mentioned in the construction of s, this means that there is an accepting run of $\mathcal{M}^{\rightarrow}_{e,k}$ over $[x, y]$. Hence, $\varphi_u(x, y) = a$ is such that $e^\omega a = e$ (definition of $M^{\rightarrow}_{e,k}$). Let $x' \in X$. By definition of J, $y \in Y$. Let us show $\varphi_u(x', y) = e$. For this, remark that there exists an infinite sequence $x' = x_0 < x_1 < \dots$ in X of length ω and limit x. Since all the x_i's belong to X, $\varphi_u(x_i, x_{i+1}) = e$ for all $i \in \omega$. We conclude that $\varphi_u(x', x) = e^\omega$. Hence

$$\varphi_u(x', y) = \varphi_u(x', x)\varphi_u(x, y) = e^\omega a = e \ .$$

Let now y' be in Y. If $y' = y$ then $\varphi_u(x', y') = \varphi_u(x', y) = e$. Otherwise, $y' > y$, and $\varphi_u(y, y') = e$ since $y, y' \in Y$. And we have $\varphi_u(x', y') = \varphi_u(x', y)\varphi_u(y, y') = ee = e$. We conclude that XRY. The same argument is used for the other possibilities for J.

Hence the hypothesis of Lemma 18 holds, and by application of the lemma, we deduce that $R = I^2$. This concludes the proof of the claim.

Let us prove that s is Ramseyan. Let $x < y$ and $x' < y'$ lie all four in the same class of n-neighborhood. If $n \geq 2$, according to the n-neighborhood relation, x, y, x' and y' lie in a common interval J nonintersecting I. Wlog., let us choose J maximal. According to the definition of s, s equals $s_J + 1$ over J. This means that x, y, x', y' were $(n-1)$-neighbors in J. Hence, $\varphi_u(x, y) = \varphi_u(x', y') = \varphi_u(x, y)^2$ since s_J is Ramseyan for (J, φ_u). If $n = 1$, all x, y, x', y' lie in I. In this case $\varphi_u(x, y) = \varphi_u(x', y') = e = e^2$ according to the claim above.

Overall, $L_{e,k+1}[q, q] \subseteq C_{e,k+1}$

Let us turn ourselves to the other inclusion: $C_{e,k+1} \subseteq L_{e,k+1}[q, q]$. Let u be a $\diamond$-word indexed by α in $C_{e,k+1}$. By definition of $C_{e,k+1}$, there exists a Ramseyan split s of $\overline{\alpha}$ of height $k + 1$ such that $s(\bot) = s(\top) = 1$. Let I be $s^{-1}(1)$.

We construct a run $\rho \in Q^{\overline{\alpha}}$ in the following way (Q is the set of states of $A_{e,k+1}$). Set $\rho(x) = q$ for all x in I. We define ρ elsewhere by copying runs of the automata $\mathcal{M}_{e,k}$, $\mathcal{M}^{\leftarrow}_{e,k}$, $\mathcal{M}^{\rightarrow}_{e,k}$ and $\mathcal{M}^{\rightarrow\leftarrow}_{e,k}$. More precisely, consider a maximal interval $J \subseteq \overline{\alpha}$ nonintersecting I. Let us define ρ over J and let

$$x = \sup\{z < J \ : \ z \in I\} \qquad \text{and,} \quad y = \inf\{z > J \ : \ z \in I\} \ .$$

Four cases happen depending on whether J is $[x, y], [x, y[,]x, y]$ or $]x, y[$. We treat the case of $[x, y[$, the others being similar.

If $J = [x, y[$, this means that $x \notin I$, but $y \in I$. As a consequence, there is a sequence $x_1 < x_2 < \dots$ in I of length ω and limit x. Since s is Ramseyan, $\varphi_u(x_i, x_{i+1}) = e$ for all i. This means that $\varphi_u(x_1, x) = e^\omega$. Furthermore, still by Ramseyness, $\varphi_u(x_1, y) = e$. We deduce that $e = \varphi_u(x_1, y) = \varphi_u(x_1, x)\varphi_u(x, y) = e^\omega \varphi_u(x, y)$. By definition of $M^{\rightarrow}_{e,k}$ we obtain that $u|_{x,y}$ is accepted by $\mathcal{M}^{\rightarrow}_{e,k}$. We define ρ to replicate the corresponding run over J using the instance of $\mathcal{M}^{\rightarrow}_{e,k}$ it contains. We have to prove that this choice indeed produces a run. Over $]x, y[$ this is a correct run since the original run was itself correct. It remains to show the correctness of the run to the left of x. But, we already know that the maximal priority reaching x from the left is n since the sequence of the x_i's tends to x and by construction correspond to a priority n which is maximal. We conclude that there is a corresponding transition in $A_{e,k+1}$. □

We can derive from the last lemma the following.

Corollary 21. *$L_{e,k+1}[q, p] = C^{\omega}_{e,k+1}$, $L_{e,k+1}[r, q] = C^{-\omega}_{e,k+1}$, and $L_{e,k+1}[r, p] = C^{\zeta}_{e,k+1}$.*

Finally, for $\xi, \xi' \in \{[,]\}$ and $a \in S$, $S^{\xi\xi'}_{k+1}(a)$ is proved accepted by a priority automaton using the following equation:

$$
\begin{aligned}
S^{\xi\xi'}_{k+1}(a) = {} & S^{\xi\xi'}_{k} && + \sum_{bc=a} S^{\xi[}_{k}(b) S^{]\xi'}_{k}(c) \\
& + \sum_{\substack{bec = a \\ e^2 = e}} S^{\xi[}_{k}(b)\, C_{e,k+1}\, S^{]\xi'}_{k}(c) && + \sum_{\substack{be^\omega c = a \\ e^2 = e}} S^{\xi[}_{k}(b)\, C^{\omega}_{e,k+1}\, S^{[\xi'}_{k}(c) \\
& + \sum_{\substack{be^{-\omega} c = a \\ e^2 = e}} S^{\xi]}_{k}(b)\, C^{-\omega}_{e,k+1}\, S^{]\xi'}_{k}(c) && + \sum_{\substack{be^\zeta c = a \\ e^2 = e}} S^{\xi]}_{k}(b)\, C^{\zeta}_{e,k+1}\, S^{[\xi'}_{k}(c)
\end{aligned}
$$

This equation is obtained by a case analysis on the split s witnessing $u \in S^{\xi\xi'}_{k+1}(a)$. The first case is when $s^{-1}(1)$ is empty, the second when it is a singleton. The following are when $s^{-1}(1)$ contains at least two elements; the four cases corresponding to the four possibilities of presence or absence of a minimal and/or a maximal element in $s^{-1}(1)$.

Acknowledgment

I thank Olivier Carton, Manfred Kufleinter, Gabriele Puppis, Sasha Rubin, and Thomas Weidner for their comments and discussions on this work. I am also grateful to the two anonymous referees for their numerous comments which greatly improved the quality of this work.

References

[1] P. A. Abdulla, P. Krcál, and W. Yi. R-automata. In *CONCUR 2008*, volume 5201, pages 67–81, 2008.

[2] N. Bedon, A. Bès, O. Carton, and C. Rispal. Logic and rational languages of words indexed by linear orderings. In *CSR*, pages 76–85, 2008.

[3] N. Bedon. Automata, semigroups and recognizability of words on ordinals. *Int. J. of Algebra and Computation*, 8:1–21, 1998.

[4] N. Bedon and O. Carton. An Eilenberg theorem for words on countable ordinals. 1380:53–64, 1998.

[5] M. Bojańczyk and T. Colcombet. Bounds in omega-regularity. In *LICS'06*, pages 285–296, 2006.
[6] T. C. Brown. An interesting combinatorial method in the theory of locally finite semigroups. *Pacific J. Math.*, 36:285–289, 1971.
[7] V. Bruyère and O. Carton. Automata on linear orderings. In *MFCS*, volume 2136, pages 236–247, 2001.
[8] J. R. Büchi. Transfinite automata recursions and weak secondo order theory of ordinals. In *Proceedings of the International Congress on Logic, Methodology and Philosophy of Science*, pages 2–23. Stanford University press, 1965.
[9] J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik und Grundl. Math.*, 6:66–92, 1960.
[10] J. R. Büchi. On a decision method in restricted second-order arithmetic. pages 1–11, 1962.
[11] O. Carton and C. Rispal. Complementation of rational sets on countable scattered linear orderings. *Int. J. Found. Comput. Sci.*, 16(4):767–786, 2005.
[12] J. Chalopin and H. Leung. On factorization forests of finite height. *Theoretical Computer Science*, 310(1–3):489–499, jan 2004.
[13] Y. Choueka. Finite automata, definable sets, and regular expressions over $mega^n$-tapes. *J. Comput. System Sci.*, 17:81–97, 1978.
[14] T. Colcombet. Factorisation forests for infinite words. In *FCT'07*, pages 226–237. Springer, 2007.
[15] Y. Gurevich. Modest theory of short chains. i. *J. Symb. Log.*, 44(4):481–490, 1979.
[16] Y. Gurevich and S. Shelah. Modest theory of short chains. ii. *J. Symb. Log.*, 44(4):491–502, 1979.
[17] K. Hashiguchi. Algorithms for determining relative star height and star height. *Inf. Comput.*, 78(2):124–169, 1988.
[18] S. C. Kleene. Representation of events in nerve nets and finite automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 3–42. Princeton University Press, Princeton, New Jersey, 1956.
[19] M. Kufleitner. A proof of the factorization forest theorem. Technical Report 2007/05, october 2007.
[20] M. Kufleitner. The height of factorization forests. In *MFCS'08*, volume 5162, pages 443–454. Springer, 2008.
[21] G. Lallement. *Semigroups and Combinatorial Applications*. Wiley, 1979.
[22] R. McNaughton. Testing and generating infinite sequences by a finite automaton. *Information and Control*, 9:521–530, 1966.
[23] Infinite sequences and finite machines. *FOCS* 1963: 3-16
[24] D. Perrin and J-É. Pin. Semigroups and automata on infinite words. In J. Fountain, editor, *NATO Advanced Study Institute Semigroups, Formal Languages and Groups*, pages 49–72. Kluwer academic publishers, 1995.
[25] D. Perrin and J-É. Pin. Infinite words: Automata, semigroups, logic and games. *Pure and Applied Mathematics*, 141, 2004.
[26] J-É. Pin. *Varieties of formal languages*. North Oxford, London and Plenum, New-York, 1986.
[27] J-É. Pin and M. J. J. Branco. Equations defining the polynomial closure of a lattice of regular languages. In *ICALP'09*, LNCS. Springer, 2009.
[28] J-É. Pin and P. Weil. Polynominal closure and unambiguous product. *Theory Comput. Syst.*, 30(4):383–422, 1997.
[29] M. O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, July 1969.
[30] M. O. Rabin and D. Scott. Finite automata and their decision problems. Technical report, 1964.
[31] J. G. Rosenstein. *Linear Orderings*. Academic Press, 1982.
[32] I. Simon. Factorization forests of finite height. *Theoret. Comput. Sci.*, 72:65–94, 1990.
[33] T. Wilke. An Eilenberg theorem for ∞-languages. In *Automata, Languages and Programming*, volume 510, pages 588–599. Springer Verlag, Berlin, Heidelberg, New York, 1991.
[34] J. Wojciechowski. Finite automata on transfinite sequences and regular expressions. *Fundamenta Informaticae*, 8:379–396, 1985.