Hierarchy among Automata on Linear Orderings

Véronique Bruyère
Institut d'Informatique*
Université de Mons-Hainaut

Olivier Carton
LIAFA†
Université Paris 7

Abstract

In a preceding paper, automata and rational expressions have been introduced for words indexed by linear orderings, together with a Kleene-like theorem. We here pursue this work by proposing a hierarchy among the rational sets. Each class of the hierarchy is defined by a subset of the rational operations that can be used. We then characterize any class by an appropriate class of automata, leading to a Kleene theorem inside the class. A characterization by particular classes of orderings is also given.

1 Introduction

The first result in automata theory and formal languages is the Kleene theorem which establishes the equivalence between sets of words accepted by automata and sets of words described by rational expressions. Since the seminal paper of Kleene [15], this equivalence has been extended to many kinds of structures: infinite words, bi-infinite words, finite and infinite trees, finite and infinite traces, pictures, *etc.*

In [5, 6], we have considered linear structures in a general framework, *i.e.*, words indexed by a linear ordering. This approach allows us to treat in the same way finite words, left- and right-infinite words, bi-infinite words, ordinal words which are studied separately in the literature. We have introduced a new notion of automaton accepting words on linear orderings, which is simple, natural and includes previously defined automata. We have also defined rational expressions for such words. We have proved the related Kleene-like theorem when the orderings are restricted to countable scattered linear orderings. This result extends Kleene's theorem for finite words [15], infinite words [9, 18], bi-infinite words [14, 19] and ordinal words [10, 13, 24].

Another jewel of formal languages is the characterization of star-free languages by first-order logic [16] or by group-free semigroups [22]. A set of finite

*Université de Mons-Hainaut, Le Pentagone, 6 avenue du Champ de Mars, B-7000 Mons, Belgium, Email: Veronique.Bruyere@umh.ac.be

†Université Paris 7, 5 place Jussieu, F-75251 Paris Cedex 05 France, Email: Olivier.Carton@liafa.jussieu.fr

words is star-free if it can be described by a rational expression using concatenation, union and complementation only. The class of star-free sets is thus obtained by restricting the rational operations. The star iteration is replaced by complementation which is weaker when union and concatenation are already allowed.

In this paper, we propose a hierarchy among rational sets of words on linear orderings. As for star-free sets, this hierarchy is obtained by restricting the rational operations that can be used. Each class contains the rational sets that can be described by a given subset of the rational operations.

The rational operations introduced in [6] include the usual Kleene operations: union, concatenation and star iteration. They also include the omega iteration usually used to construct infinite words and the ordinal iteration introduced by Wojciechowski [24] for ordinal words. Three new operations are added: the backwards omega iteration, the backwards ordinal iteration and a last operation which is a kind of iteration for all countable scattered linear orderings. The lowest class of the hierarchy contains sets that can be described by rational expressions using union, concatenation and star iteration. This is of course the class of rational sets of finite words. The greatest class contains sets that can be described by rational expressions using all rational expressions introduced in [6]. It contains all rational sets of words on scattered linear orderings. Some other classes of words already studied in the literature appear naturally in our framework. Sets of words on ordinals introduced by Büchi [10] or sets of words on ordinals smaller than ω^ω studied by Choueka [13] form two classes of our hierarchy.

We give a characterization of each class of the hierarchy by a corresponding class of automata. A set of words belongs to the given class if and only if it is recognized by an automaton of the corresponding class. Each of these characterizations is thus a Kleene's theorem which holds for that class. For well-known classes, these Kleene's theorems were already proved by Wojciechowski [24] for words on ordinals or by Choueka [13] for words on ordinals smaller than ω^ω. In each case, the corresponding class of automata is obtained naturally by restricting the kind of transitions that can be used. For instance, the automata for words on ordinals do have left limit transitions but no right limit transitions as there are defined by Büchi [10].

The last rational operation defined in [6] works like an iteration for all countable scattered linear orderings. It is binary. In this paper, we consider a simpler definition of this iteration as a unary operation. This simplified definition seems to be more natural but it turns out to be weaker. The results of this paper show that the binary operation is really needed to obtain the Kleene's theorem of [6]. This question was actually the original motivation of our work.

We also give a characterization of each class of the hierarchy (except one) by a corresponding class of orderings. A set of words belongs to the given class if and only if the length of each of its words belongs to the corresponding class of orderings. For some classes as the class of sets of words on ordinals, this characterization is straightforward. However some other classes need that suitable classes of orderings are defined. These definitions are inspired by the characterization of countable scattered orderings due to Hausdorff.

To summarize, the results of the paper establish a hierarchy among rational sets of words on linear orderings, with connections between natural classes of orderings, rational operations and the types of transitions in automata.

The paper is organized as follows. In Sections 2, 3 and 4, we briefly recall the new notions introduced in [6]: words on linear orderings, automata and rational expressions. We refer the reader to [21] for a complete introduction to linear orderings. The different classes of the hierarchy are described in Section 5. This hierarchy is summarized on Figure 7 and illustrated by some examples. The characterization by appropriate orderings is studied in Section 6. The characterization by appropriate automata of the new classes of the hierarchy is studied in Section 8. Some technical proofs on orderings are put in Section 7.

The results of this paper were first presented in [7].

2 Orderings and Words

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

In this article, linear orderings are thus used to index sequences. Therefore we are only interested in orderings up to isomorphism. We freely say that two orderings are equal if they are actually isomorphic.

Given a linear ordering J, we denote by $-J$ the *backwards* linear ordering obtained by reversing the ordering relation. For instance, $-\omega$ is the backwards linear ordering of ω which is used to index the so-called left-infinite words. For a class $\mathcal{V}$ of linear orderings, we denote by $-\mathcal{V}$ the class $\{-J \mid J \in \mathcal{V}\}$.

Given two linear orderings J and K, the linear ordering $J + K$ is obtained by juxtaposition of J and K, *i.e.*, it is the linear ordering on the disjoint union $J \cup K$ extended with $j < k$ for any $j \in J$ and any $k \in K$. For instance, the linear ordering ζ can be obtained as the sum $-\omega + \omega$. More generally, let J and K_j for $j \in J$, be linear orderings. The linear ordering $\sum_{j \in J} K_j$ is obtained by juxtaposition of the orderings K_j with respect to J. More formally, the *sum* $\sum_{j \in J} K_j$ is the set L of all pairs (k, j) such that $k \in K_j$. The relation $(k_1, j_1) < (k_2, j_2)$ holds iff $j_1 < j_2$ or else $j_1 = j_2$ and $k_1 < k_2$ in K_{j_1}.

The sum operation on linear orderings leads to a notion of product of words as follows. Let J and K_j for $j \in J$, be linear orderings. Let $x_j = (a_{k,j})_{k \in K_j}$ be a word of length K_j, for any $j \in J$. The *product* $\prod_{j \in J} x_j$ is the word of length $L = \sum_{j \in J} K_j$ equal to $(a_{k,j})_{(k,j) \in L}$. For instance, the word $a^{-\omega} \cdot a^{\omega}$ of length

$\zeta = -\omega + \omega$ is the product of the two words $a^{-\omega}$ and a^{ω} of length $-\omega$ and ω respectively.

An ordering J is *dense* if for any $i < k$ in J, there is $j \in J$ such that $i < j < k$. An ordering is *scattered* if it has no dense subordering. In this paper as in [6], we only consider linear orderings which are *countable* and *scattered*. This class is denoted by $\mathcal{S}$ and its elements are shortly called *orderings*. We use the notation $\mathcal{N}$ for the subclass of $\mathcal{S}$ of finite linear orderings and $\mathcal{O}$ for the subclass of countable ordinals. Recall that an *ordinal* is a linear ordering which is well-ordered, that is, it does not contain the subordering $-\omega$.

The following characterization of the class $\mathcal{S}$ is due to Hausdorff [21]. The notation 1 is used for the finite ordering with one element and the notation $\beta < \alpha$ denotes the usual ordering on ordinals.

Theorem 1 (Hausdorff) *$\mathcal{S} = \bigcup_{\alpha \in \mathcal{O}} U_\alpha$ where the classes U_α are inductively defined by*

1. $U_0 = \{\varnothing, 1\}$;
2. $U_\alpha = \{\sum_{j \in J} K_j \mid J \in \mathcal{N} \cup \{\omega, -\omega, \zeta\}$ *and* $K_j \in \bigcup_{\beta < \alpha} U_\beta\}$.

Example 2 The ordinal ω belongs to U_1 because $\omega = \sum_{j \in \omega} K_j$ with $K_j = 1 \in U_0$. More generally one can check that the ordinal ω^n belongs to U_n, for any $n \geq 1$. Finally, ω^ω belongs to U_ω since it equals $\sum_{j \in \omega} K_j$ with $K_j = \omega^j \in U_j$.

In this article, we propose two new families of classes to characterize $\mathcal{S}$.

Definition 3 *For any $\alpha \in \mathcal{O}$,*

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

One easily checks that for any $\alpha \in \mathcal{O}$, the class $-V_\alpha$ can be defined by $\{\sum_{j \in J} K_j \mid J \in -\mathcal{O} \cup \{\omega, \zeta\}$ and $K_j \in \bigcup_{\beta < \alpha} -V_\beta\}$.

Definition 4 *For any $\alpha \in \mathcal{O}$,*

1. $W_0 = \{\varnothing, 1\}$;
2. $W_\alpha = \{\sum_{j \in J} K_j \mid J \in \mathcal{O} \cup -\mathcal{O} \cup \{\zeta\}$ *and* $K_j \in \bigcup_{\beta < \alpha} W_\beta\}$.

Example 5 The ordinal ω^ω belongs to V_1, to $-V_\omega$ and to W_1. More generally, $\mathcal{O} \subseteq V_1$ and $\mathcal{O} \cup -\mathcal{O} \subseteq W_1$. It can be proved that the ordering ζ^ω belongs to W_ω. The ordering ζ^α is defined inductively on $\alpha \in \mathcal{O}$ as follows (see [21, p. 90]):

$$\begin{aligned}
\zeta^0 &= 1 \\
\zeta^{\alpha+1} &= -(\zeta^\alpha \cdot \omega) + \zeta^\alpha + \zeta^\alpha \cdot \omega = \zeta^\alpha \cdot \zeta \\
\zeta^\alpha &= -\sum \{\zeta^\beta \mid \beta < \alpha\} + 1 + \sum \{\zeta^\beta \mid \beta < \alpha\} \quad \text{if } \alpha \text{ is a limit ordinal.}
\end{aligned}$$

It is easy to see that $U_\alpha \subseteq V_\alpha$ and $V_\alpha \subseteq W_\alpha$ and that the classes V_α and W_α contain only scattered linear orderings. It follows from Theorem 1 that $\mathcal{S}$ is equal to the union of the classes V_α and to the union of the classes W_α as stated in the following theorem.

Theorem 6 *The following equalities hold.*

$$\mathcal{S} = \bigcup_{\alpha \in \mathcal{O}} U_\alpha = \bigcup_{\alpha \in \mathcal{O}} V_\alpha = \bigcup_{\alpha \in \mathcal{O}} W_\alpha.$$

3 Automata

In this section, we present the definition proposed in [6] for automata accepting words on linear orderings. Before defining them in detail, we recall how usual automata accept finite words. Consider an automaton $\mathcal{A} = (Q, A, E, I, F)$ with Q the set of states, A the alphabet, $E \subseteq Q \times A \times Q$ the set of edges, $I \subseteq Q$ the set of initial states and $F \subseteq Q$ the set of final states. A word $x = a_1 a_2 \cdots a_n \in A^*$ is accepted by $\mathcal{A}$ if there exists a sequence $\gamma = q_0 q_1 \cdots q_n$ of states that can be *inserted* between the letters of w such that $q_0 \in I$, $q_n \in F$ and $q_{j-1} \xrightarrow{a_j} q_j \in E$ for all $1 \leq j \leq n$ (see Figure 1). In terms of orderings, the ordering $\{0, 1, \ldots, n\}$ used for the sequence γ corresponds to the set of *cuts* of the ordering $\{1, \ldots, n\}$ used for the word x.

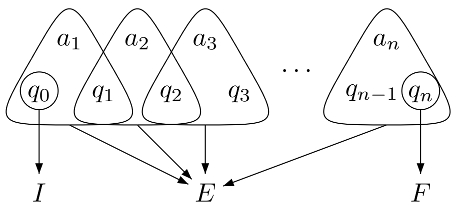


Figure 1: Successful run in a usual automaton

We recall that a *cut* of an ordering J is a pair (K, L) of intervals such that $J = K \cup L$ and $k < l$ for any $k \in K$, $l \in L$. The two intervals K and L must be disjoint and they form a partition of the set J. The set of all cuts of the ordering J is denoted by $\hat{J}$. The cuts can be linearly ordered as follows. Let $c_1 = (K_1, L_1)$ and $c_2 = (K_2, L_2)$ be two cuts of J. Define $c_1 < c_2$ whenever $K_1 \subsetneq K_2$. Note that $\hat{J}$ has always a least cut $(\varnothing, J)$ denoted $c_{\min}$ and a greatest cut $(J, \varnothing)$ denoted $c_{\max}$.

The set $J \cup \hat{J}$ is naturally endowed with a linear ordering such that J and $\hat{J}$ are two of its suborderings as follows. For $j \in J$ and a cut $c = (K, L)$, the relations $j < c$ and $c < j$ are respectively defined by $j \in K$ and $j \in L$. For any element $j \in J$, there are two consecutive cuts denoted by $c_j^- \in \hat{J}$ and $c_j^+ \in \hat{J}$ such that $c_j^- < j < c_j^+$. These cuts are given by $c_j^- = (K, \{j\} \cup L)$ and $c_j^+ = (K \cup \{j\}, L)$ with $K = \{k \mid k < j\}$ and $L = \{k \mid j < k\}$. See Figure 2 where each element of J is represented by a bullet, and each cut by a vertical bar.

$$\Big| \cdots | \bullet | \bullet | \bullet | \cdots \Big| \cdots | \bullet | \bullet | \bullet | \cdots \Big|$$

Figure 2: Ordering $J \cup \hat{J}$ for $J = \zeta + \zeta$

In our previous situation, $J = \{1, \ldots, n\}$ and $\hat{J} = \{0, 1, \ldots, n\}$. The least and greatest cuts of $\hat{J}$ are indices of an initial and a final state respectively. Furthermore, $q_{c_j^-} \xrightarrow{a_j} q_{c_j^+}$ is a transition of the automaton for any $j \in J$.

Automata accepting words on linear orderings are a natural extension of finite automata. As above, they are defined as $\mathcal{A} = (Q, A, E, I, F)$. The set E is composed of three types of transitions: the usual *successor* transitions in $Q \times A \times Q$, the *left limit* transitions which belong to $\mathcal{P}(Q) \times Q$ and the *right limit* transitions which belong to $Q \times \mathcal{P}(Q)$.

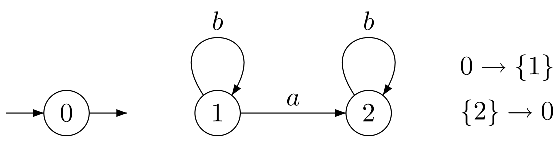


Figure 3: An automaton on linear orderings

Example 7 The automaton depicted in Figure 3 has one left limit transition $\{2\} \to 0$ and one right limit transition $0 \to \{1\}$.

A word $x = (a_j)_{j \in J}$ of length J is accepted by $\mathcal{A}$ if it is the label of a successful path. A *path* γ is a sequence of states $\gamma = (q_c)_{c \in \hat{J}}$ of length $\hat{J}$ verifying the following intuitive conditions. For two consecutive states in γ, there must be a successor transition labeled by the letter in between. For a state q in γ which has no predecessor in γ, there must be a left limit transition $P \to q$ where P is the limit set of γ to the left of q. Right limit transitions are used similarly when q has no successor in γ. A path is *successful* if its first state $q_{c_{\min}}$ is initial and its last state $q_{c_{\max}}$ is final.

More precisely, for any cut $c \in \hat{J}$, define the sets $\lim_{c^-} \gamma$ and $\lim_{c^+} \gamma$ as follows:

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

For any consecutive cuts c_j^- and c_j^+ of $\hat{J}$, $q_{c_j^-} \xrightarrow{a_j} q_{c_j^+}$ must be a successor transition. For any cut $c \neq c_{\min}$ in $\hat{J}$ which has no predecessor, $\lim_{c^-} \gamma \ \rightarrow \ q_c$ must be a left limit transition. For any cut $c \neq c_{\max}$ in $\hat{J}$ which has no successor, $q_c \rightarrow \lim_{c^+} \gamma$ must be a right limit transition.

Finally, a set of words is said to be *recognizable* if it is the set of words accepted by some automaton.

$$\begin{array}{ccccccccccccccccccccccc} | & \cdots b & | & b & | & a & | & b & | & b \cdots & | & \cdots b & | & b & | & a & | & b & | & b \cdots & | \\ 0 & \{1\} & 1 & & 1 & & 2 & & 2 & \{2\} & 0 & \{1\} & 1 & & 1 & & 2 & & 2 & \{2\} & 0 \end{array}$$

Figure 4: The word $(b^{-\omega}ab^{\omega})^2$ is accepted

Example 8 The word $(b^{-\omega}ab^{\omega})^2$ of length $J = \zeta + \zeta$ is accepted by the automaton of Figure 3. Indeed, there exists a sequence of states $\gamma = 01^{-\omega}2^{\omega}01^{-\omega}2^{\omega}0$ of length $\hat{J}$ such that the first state is initial, the last state is final and the successor, left limit and right limit transitions are respected. See Figure 4 where the sets $\lim_{c^-} \gamma$ and $\lim_{c^+} \gamma$ are also indicated. Note that the word $(b^{-\omega}ab^{\omega})^{\omega}$ cannot be accepted by the automaton. Indeed the only way to find an appropriate sequence of states is to consider the sequence $\gamma = (01^{-\omega}2^{\omega})^{\omega} q_{c_{\max}}$. There is no possible choice for $q_{c_{\max}}$ because the cut $c_{\max}$ has no predecessor, $\lim_{c_{\max}^-} \gamma$ is equal to $\{0, 1, 2\}$ and there exists no left limit transition $\{0, 1, 2\} \rightarrow q$ in the automaton.

The notion of path γ we have introduced for words on orderings coincides with the usual notion of path considered in the literature for finite words [20], ω-words [23] and ordinal words [2]. For a Muller automaton accepting ω-words, a left limit set P is computed at the end of the path. It is nothing else than the states appearing infinitely often along the path. In our context, the path then ends with an additional left limit transition from P to a state q which is final.

An automaton is said to be *trim* iff any transition is used in a successful path. If an automaton is not trim, it can easily be trimmed by removing any state and any transition which does not appear in a successful path. The set of words recognized by the automaton is of course not changed by this operation.

4 Rational Expressions

We now recall the notion of rational set of words on linear orderings as defined in [6]. The rational operations include of course the usual Kleene operations for finite words which are the union $+$, the concatenation $\cdot$ and the star operation $*$. They also include the omega iteration ω usually used to construct ω-words and the ordinal iteration $\sharp$ introduced by Wojciechowski [24] for ordinal words. Three new operations are also needed: the backwards omega iteration $-\omega$, the backwards ordinal iteration $-\sharp$ and a last binary operation denoted $\diamond$ which is a kind of iteration for all orderings.

Given two sets X and Y of words, we define

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

The last operation needs some explanation. The notation $\hat{J}^{*}$ is used for the set $\hat{J} \setminus \{c_{\min}, c_{\max}\}$. A word x belongs to $X \diamond Y$ iff there is a nonempty ordering J such that x is the product indexed by the ordering $J \cup \hat{J}^{*}$ of words where each word indexed by an element of J belongs to X and each word indexed by a cut in $\hat{J}^{*}$ belongs to Y (see Figure 5). We use the notation $X^{\diamond}$ for the set $(X \diamond \varepsilon) + \varepsilon$. Hence

$$X^{\diamond} = \{\prod_{j \in J} x_j \mid J \in \mathcal{S}, x_j \in X\}.$$

Note that contrarily to $X \diamond Y$, the empty ordering $J = \varnothing$ is allowed in $X^{\diamond}$ in a way to construct the empty word $\varepsilon \in X^{\diamond}$. When operation $\diamond$ is used as a unary operation, we also use the notation $\diamond_1$. When it is used as a binary operation, we use the notation $\diamond_2$.

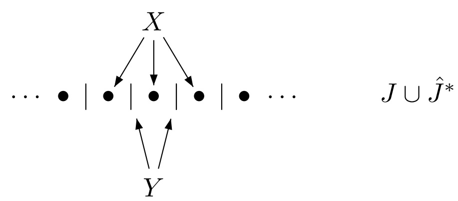


Figure 5: The operation $X \diamond Y$

Note that the definitions for the operations $*$, $\sharp$ and $\diamond_1$ are similar to each other. The only difference is that the products are over any $J \in \mathcal{N}$ for operation $*$ whereas they are over any $J \in \mathcal{O}$ for operation $\sharp$, and over any $J \in \mathcal{S}$ for operation $\diamond_1$.

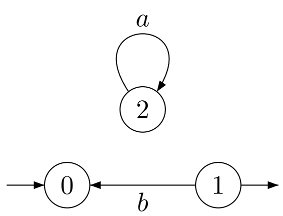
An abstract *rational expression* is a well-formed term of the free algebra over $\{\varnothing\} \cup A$ with the symbols denoting the rational operations as function symbols. Each rational expression denotes a set of words which is inductively defined by the above definitions of the rational operations. A set of words is said to be *rational* if it can be denoted by some rational expression.

Example 9 The set of all words over the alphabet A is the rational set $A^\diamond$. The set of words accepted by the automaton of Figure 3 is the rational set $(b^{-\omega}ab^{\omega})^*$.

It is stated in [6] that a set of words on countable scattered linear orderings is accepted by an automaton iff it can be described by a rational expression. This result extends the well-known Kleene's theorem on finite words, its extension to ω-words [9] and to ordinal words [24].

Theorem 10 ([6]) *Over countable scattered linear orderings, $X \subseteq A^\diamond$ is recognizable iff it is rational.*

The proof that any rational set of words is recognizable is by induction on the rational expression denoting the set by giving the corresponding construction for the automaton. The constructions for the union, the concatenation and the star iteration are very similar to the classical ones for automata on finite words [20]. The proof that any set of words accepted by an automaton is rational is a generalization of the McNaughton and Yamada algorithm [17]. It is based on an induction on the number of states of the automaton and the type of limit transitions that are used in the path. The base of the induction is Kleene's theorem on finite and ω-words. This generalization of the McNaughton and Yamada algorithm is actually the most difficult part of the proof.



$0 \to \{2\}$

$0 \to \{0, 1, 2\}$

$\{2\} \to 1$

$\{0, 1, 2\} \to 1$

Figure 6: Automaton accepting the set $(a^{-\omega}a^{\omega}) \diamond b$

Example 11 The automaton pictured in Figure 6 accepts the set denoted by the rational expression $(a^{-\omega}a^{\omega}) \diamond b$. The part of the automaton given by state 2 and the two limit transitons $0 \to \{2\}$ and $\{2\} \to 1$ accepts the word $a^{-\omega}a^{\omega}$ whereas the part given by the successor transition from state 1 to state 0 accepts the word b. Any occurrence of $a^{-\omega}a^{\omega}$ is preceded and followed by an occurrence of b in the automaton. Thanks to the limit transitions $0 \to \{0, 1, 2\}$ and $\{0, 1, 2\} \to 1$, the occurrences of $a^{-\omega}a^{\omega}$ are indexed by an ordering J, the occurrences of b are indexed by the ordering $\hat{J}^*$ and they are interleaved according to the ordering $J \cup \hat{J}^*$.

5 Hierarchy

We now come to the main result of this paper. We introduce a hierarchy among rational sets of words on countable scattered linear orderings. This hierarchy is obtained by restricting the rational operations which can be used in the rational expressions. Each class contains the rational sets which can be denoted by rational expressions using only a given subset of the rational operations. We do not consider all possible subsets of rational operations. The smallest considered subset is the subset $\{+, \cdot, *\}$ of the three usual Kleene operations. The corresponding class is of course the class of rational sets of finite words. Then, the other operations are added to these three operations by increasing power. This means that $\sharp$ is added after ω, $-\sharp$ after $-\omega$, $\diamond_1$ after $\sharp$ and $-\sharp$, and finally $\diamond_2$ after $\diamond_1$. This leads to 11 different subsets of rational operations and 11 corresponding classes of rational sets. The 11 classes of the hierarchy are summarized in Figure 7.

The *mirror* of a word x of length J is the word $-x$ of length $-J$ (the backwards linear ordering) defined by $(-x)_j = x_j$ for any $j \in J$. Similarly, the *mirror* of a set X of words is the set $-X = \{-x \mid x \in X\}$ and the *mirror* of a class $\mathcal{C}$ of sets of words is the class $\{-X \mid X \in \mathcal{C}\}$. Note that if the set X of words is recognized by an automaton $\mathcal{A}$, the set $-X$ is recognized by the automaton obtained from $\mathcal{A}$ by reversing the successor transitions, changing the left limit transitions into right limit transitions and vice-versa, and exchanging the sets of initial and final states. In this hierarchy, some classes are the mirror of another class. Classes $1'$, $3'$ and $4'$ are the respective mirrors of Classes 1, 3 and 4 and any statement concerning these classes can be easily deduced from the statement concerning 1, 3 and 4. In the sequel, we omit the statement concerning these classes.

The most interesting fact about this hierarchy is that each of its classes of rational sets can be characterized by a class of automata of a special form. Each characterization states that a rational set belongs to a given class of the hierarchy iff it can be recognized by an automaton of the corresponding class. Furthermore, for all classes but one, we have obtained a stronger result which states that if a rational set belongs to a given class, then *any* automaton recognizing this set belongs to the corresponding class. The characterization of each class is actually a Kleene-like theorem since it states the equivalence between some subset of rational operations and a class of automata of a special form. The corresponding classes of automata are defined by restricting the limit transitions of the automata. Therefore, these Kleene-like theorems enlighten the connections between the form of the limit transitions used in the automata and the rational operations needed to denote the accepted sets.

Finally, all classes but one can be characterized by a particular class of orderings. Each characterization states that a rational set belongs to a given class of the hierarchy iff the lengths of all its words belong to the corresponding class of orderings. The classes of orderings measure the structural complexity of the orderings. Therefore, these characterizations establish connections between the rational operations used in expressions and the complexity of the words denoted by the expression.

The hierarchy is pictured in Figure 7. Each class is described in a box in the following way. The first line lists the rational operations which are allowed in rational expressions. The second line describes the form of the limit transitions of the corresponding automata. The third line gives the corresponding class of orderings. For instance, Class 1 is the class of all rational sets denoted by expressions using $+$, $\cdot$, $*$ and ω. The corresponding automata may have any left limit transitions $P \to q$ with $q \notin P$ but they have no right limit transitions. The corresponding class of orderings is the class of all ordinals smaller than ω^ω.

Let us give a precise description of each class.

Class 0: This class corresponds to Kleene's theorem [15] on rational sets of finite words. The rational operations are the usual Kleene operations: union, concatenation and star iteration. The automata are the usual automata on finite words with no limit transitions. The orderings are the finite ones.

Class 1: This class corresponds to Choueka's theorem [13] on rational sets of words of length an ordinal smaller than ω^ω. The rational operations are the Kleene ones and the omega iteration. Automata considered by Choueka are a special kind of automata on ordinals introduced by Büchi. It is shown in [1] that these automata are equivalent to automata with no right limit transitions and such that the left limit transitions are of the form $P \to q$ with $q \notin P$. The corresponding orderings are the ordinals smaller than ω^ω.

Class 2: The rational operations of Class 2 are the Kleene ones, the omega iteration and the backwards omega iteration. The automata have left limit transitions $P \to q$ with $q \notin P$ and right limit transitions $q' \to P'$ with $q' \notin P'$. The associated orderings are the orderings belonging to $\bigcup_{n<\omega} U_n$ where the classes U_n have been defined in Theorem 1.

Class 3: This class corresponds to Wojciechowski's theorem [24] on rational sets of words on countable ordinals. The rational operations are the Kleene ones, the omega iteration and the ordinal iteration introduced by Wojciechowski. The automata are those on ordinals introduced by Büchi [10]. They have any left limit transitions of the form $P \to q$ but they have no right limit transitions. The related orderings are exactly those of the class $\mathcal{O}$ of all countable orderings.

Class 4: The rational operations of Class 4 are the Kleene one, the omega iteration, the backwards omega iteration and the ordinal iteration. The automata have any left limit transitions $P \to q$ but right limit transitions $q' \to P'$ limited by the condition $q' \notin P'$. The associated orderings are the orderings belonging to $\bigcup_{n<\omega} V_n$ where the classes V_n are described in Definition 3.

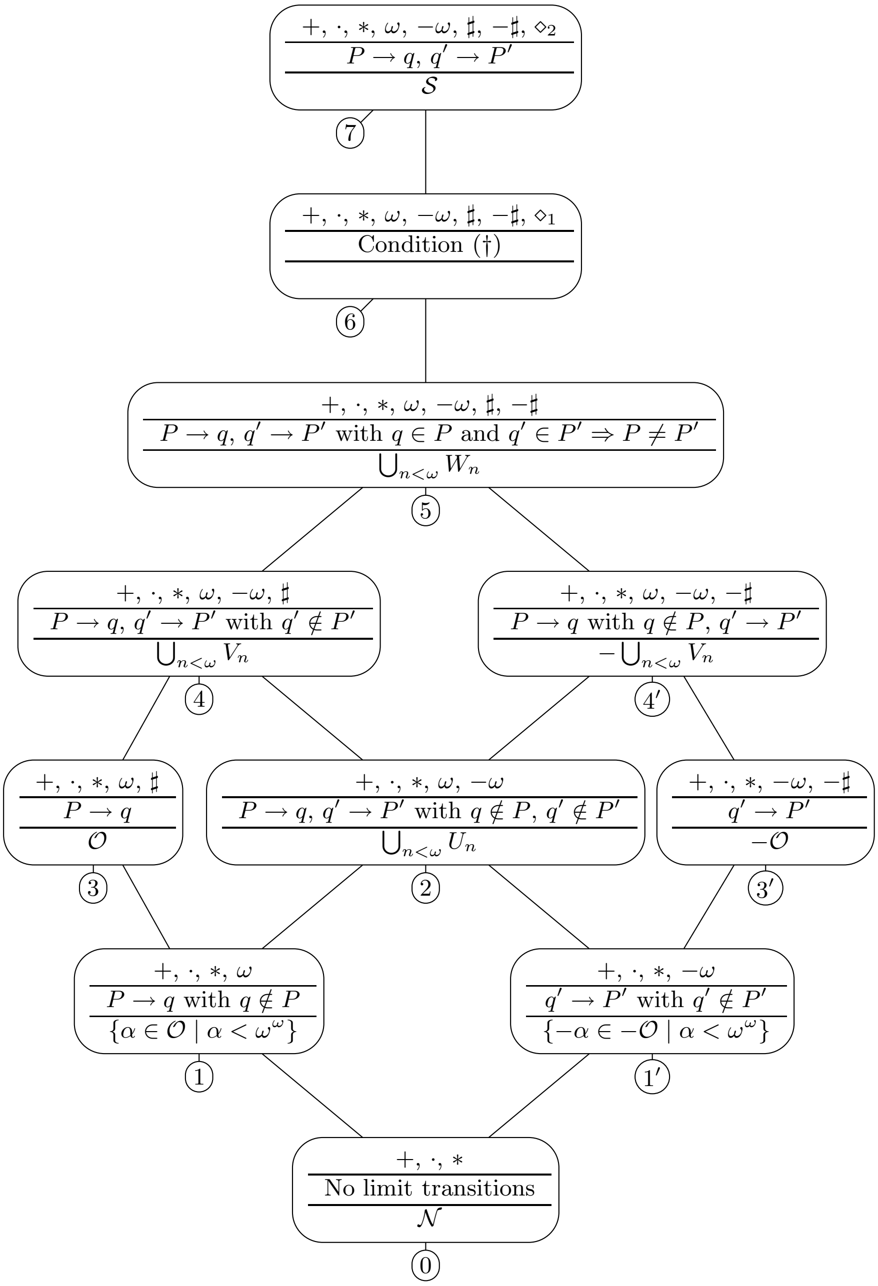


Figure 7: Classes of the hierarchy

Class 5: In this class, all rational operations but the operation $\diamond$ are allowed. The automata may have left and right limit transitions but these transitions must satisfy the following condition. If $P \to q$ and $q' \to P'$ are two transitions and if furthermore $q \in P$ and $q' \in P'$, then one has necessarily $P \neq P'$. The related class of orderings is equal to $\bigcup_{n<\omega} W_n$ (see Definition 4).

Class 6: In this class, all rational operations are allowed but the operation $\diamond$ can only be used as a unary operation, that is the operation $X \diamond Y$ must be restricted to the case $Y = \{\varepsilon\}$. The automata may have left and right limit transitions but these transitions satisfy the following condition (†).

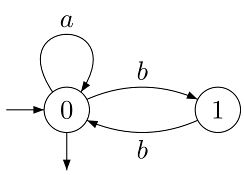
Condition (†): let $P \to q$ be a left limit transition and $q' \to P'$ be a right limit transition. If $q \in P$, $q' \in P'$ and $P = P'$, then $q = q'$ and for any $R \subseteq P$ with $q \in R$, the left and right transitions $R \to q$ and $q \to R$ must appear among the transitions of the automaton.

We do not know a characterization by a particular class of orderings.

Class 7: This class corresponds to the Kleene theorem of [6] for all countable scattered orderings. The operation $\diamond$ is here used as a binary operation.

The new Classes 2, 4, 5 and 6 will be studied in detail in Sections 6 and 8. Let us illustrate by examples some classes of the hierarchy.

Example 12 The set $(b^{-\omega}ab^{\omega})^*$ of Examples 7 and 9 belong to Class 2. The related orderings are $K_n = \sum_{j\in\{1,\ldots,n\}} \zeta$ which belong to U_2. Since the linear ordering ζ is neither an ordinal nor a backwards ordinal, the set $(b^{-\omega}ab^{\omega})^*$ cannot belong to a lower class.



$0 \to \{0\}$

$0 \to \{0,1\}$

$\{0\} \to 0$

$\{0,1\} \to 0$

Figure 8: Automaton accepting the set $(a + bb)^{\diamond}$

Example 13 The automaton of Figure 8 accepts the rational set $(a + bb)^{\diamond}$. The operation $\diamond$ is unary and the automaton satisfies Condition (†). Hence it is an example of Class 6. Note that it does not belong to Class 5. Indeed, the condition on the limit transitions of the automaton is not respected. Moreover, the set $(a + bb)^{\diamond}$ contains the word $\prod_{j\in J} a$ of length $J = \zeta^{\omega}$ and we will see in Lemma 16 that J belongs to W_ω but not to $\bigcup_{n<\omega} W_n$.

The following example shows that Class 6 of the hierarchy is strictly included in Class 7. In other words, the operation $\diamond$ turns out to be weaker when it is considered as a unary instead of binary operation. This question was actually the original motivation of the article.

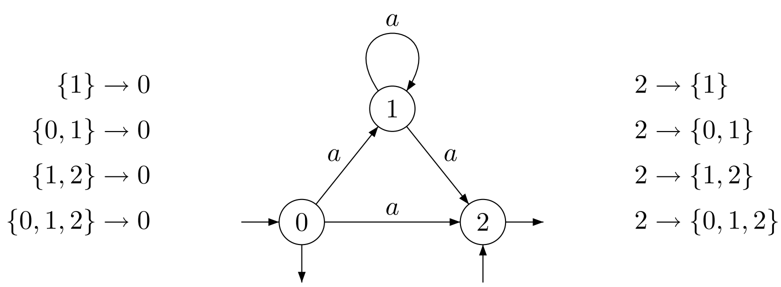


Figure 9: Automaton accepting the set $\varepsilon \diamond a$

Example 14 The automaton pictured in Figure 9 accepts the set $\varepsilon \diamond a$. This set is the set of words over the alphabet $\{a\}$ whose length is a complete ordering. This will be explained below.

Recall that a linear ordering is *complete* if any nonempty subset which is upper bounded has a least upper bound (or equivalently if any nonempty subset which is lower bounded has a greatest lower bound).

This automaton does not satisfy Condition (†). Indeed there exist left and right limit transitions $P \to q$ and $q' \to P'$ such that $q \in P$, $q' \in P'$ and $P = P'$ but $q \neq q'$. Take $P = P' = \{0, 1, 2\}$, $q = 0$ and $q' = 2$. It can be shown that any rational expression with a unary $\diamond$ denotes a set which contains words with a non complete length. Therefore the set $\varepsilon \diamond a$ does not belong to Class 6.

Let us now explain that (i) the rational expression $\varepsilon \diamond a$ denotes the set of words the length of which is a complete ordering, (ii) the automaton pictured in Figure 9 accepts this set.

(i) Coming back to the definition of the operation $\diamond$, the set $\varepsilon \diamond a$ is equal to $\{\prod_{j \in \hat{J}^*} a_j \mid J \in \mathcal{S} \setminus \{\varnothing\} \text{ and } a_j = a\}$. It can be shown that a scattered linear ordering K is complete iff there is a scattered linear ordering J such that $K = \hat{J}^*$. Therefore, a word x belongs to $\varepsilon \diamond a$ iff its length is a complete scattered linear ordering.

(ii) Let $x = (a_j)_{j \in J}$ with J a complete ordering, labeling a path $(q_c)_{c \in \hat{J}}$. As J is complete, any cut $c \in \hat{J}$ has a successor or a predecessor. So there are three types of cuts : type 0 if c has a successor and no predecessor, type 1 if c has a successor and a predecessor, type 2 if c has a predecessor and no successor. These three types correspond to the three states of the automaton of Figure 9. We now explain the transitions of the automaton. It is easy to check its successor transitions by considering all the possible pairs of consecutive cuts. The limit transitions are any $P \to 0$ or $2 \to P$, for any $P \subseteq \{0, 1, 2\}$. However some are never used inside paths. Indeed any scattered ordering with at least three elements has always three consecutive elements. It follows that the limit transitions can be restricted to the transitions of the form $P \to 0$ or $2 \to P$ with P containing state 1.

Thus the automaton of Figure 9 accepts the set $\varepsilon \diamond a$.

The proofs of the different characterizations of the classes are presented in the following three sections. Recall that the characterizations of Classes 0, 1 and 3 are just another formulation of the respective results of Kleene [15], Choueka [13] and Wojciechowski [24]. The characterization of Class 7 is just a reformulation of our previous result in [6]. Note also that the characterizations of classes $1'$, $3'$ and $4'$ follow directly from the characterizations of classes 1, 3 and 4. The remaining classes are Classes 2, 4, 5 and 6.

Classes 2, 4 and 5 are characterized by suitable rational expressions, automata and linear orderings. In order to avoid unnecessary proofs, it could be proved that the characterization by rational expressions implies the characterization by orderings which in turn implies the characterization by automata and that finally the characterization by automata implies the characterization by rational expressions. This outline of proof cannot be followed for Class 6 since there is no corresponding class of orderings. Furthermore, much of the material which is needed to present the proofs for Class 6 can also be used for the other classes. Therefore, we have preferred to split the proofs into two parts. In the next section, we first prove that the characterizations of Classes 2, 4 and 5 by rational expressions and by orderings are equivalent. The proof of an order-theoretic lemma (Lemma 16) is delayed to Section 7. In Section 8, we finally prove that the characterizations of Classes 2, 4, 5 and 6 by rational expressions and by automata are equivalent.

6 Connections with Classes of Orderings

In this section, we state and prove the characterization of Classes 2, 4 and 5 by appropriate classes of orderings. These characterizations are interesting by themselves since they establish connections between the rational operations and the structural complexity of the lengths of the words.

We recall that Hausdorff has characterized the class $\mathcal{S}$ as the union $\bigcup_{\alpha \in \mathcal{O}} U_\alpha$ (see Theorem 1). We have proposed two new families of classes to characterize $\mathcal{S}$ (see Definitions 3 and 4). The class $\mathcal{S}$ is then also equal to $\bigcup_{\alpha \in \mathcal{O}} V_\alpha$ and to $\bigcup_{\alpha \in \mathcal{O}} W_\alpha$ (see Theorem 6).

The next theorem gives a characterization of Classes 2, 4 and 5 of the hierarchy thanks to the classes of orderings U_n, V_n and W_n. We use the notation $\mathrm{Rat}_A(+, \cdot, *, \omega, -\omega)$ to express the alphabet and the rational operations which are allowed in the rational expressions.

Theorem 15 *Let $X \subseteq A^\diamond$ be a rational set. Then*

$$\begin{array}{lllll} 1. & X \in \mathrm{Rat}_A(+,\cdot,*,\omega,-\omega) & \iff & \exists n \geq 0 \ \forall x \in X & |x| \in U_n \\ 2. & X \in \mathrm{Rat}_A(+,\cdot,*,\omega,-\omega,\sharp) & \iff & \exists n \geq 0 \ \forall x \in X & |x| \in V_n \\ 3. & X \in \mathrm{Rat}_A(+,\cdot,*,\omega,-\omega,\sharp,-\sharp) & \iff & \exists n \geq 0 \ \forall x \in X & |x| \in W_n \end{array}$$

The proof of the theorem uses the following three lemmas. The first lemma is a pure ordering theoretic lemma. It provides three orderings which do not belong respectively to the classes U_n, V_n and W_n. These orderings will be used to show that, if the lengths of the words of a set X belong to a class U_n, V_n or W_n, then some rational operation cannot appear in a rational expression denoting X.

Lemma 16 *One has the following memberships.*

$$\omega^\omega \in U_\omega \setminus \bigcup_{n<\omega} U_n \ , \qquad -\omega^\omega \in V_\omega \setminus \bigcup_{n<\omega} V_n \quad \textit{and} \quad \zeta^\omega \in W_\omega \setminus \bigcup_{n<\omega} W_n$$

The proof that the ordering ω^ω belongs to U_ω but to no U_n with $n < \omega$ can be found in Chapter 5 of [21]. In that chapter, a notion of rank $r(J)$ is introduced for any ordering J in $\mathcal{S}$. The definition relies on the concept of iterated condensations. It is then proved that given an ordering $J \in \mathcal{S}$, the value α of its rank gives the smallest class U_α to which J belongs. For example, the rank of the ordering ω^ω can be easily computed. It is equal to ω showing that ω^ω belongs to U_ω but not to U_n for any $n < \omega$. We have developed the same approach for classes V_α and W_α, with $\alpha \in \mathcal{O}$, by defining the adequate notions of rank and iterated condensations. The related ranks of the orderings $-\omega^\omega$ and ζ^ω can be easily computed, showing that $-\omega^\omega \in V_\omega \setminus \bigcup_{n<\omega} V_n$ and $\zeta^\omega \in W_\omega \setminus \bigcup_{n<\omega} W_n$. All these developments are provided in Section 7, together with a proof of Lemma 16. The main difficulty was here to define correctly the new classes V_α and W_α, in a way to obtain Theorem 15. The reader who is willing to believe Lemma 16 could skip Section 7.

The following lemma states that the classes U_α, V_α and W_α are closed under taking subordering.

Lemma 17 *Let K be a subordering of an ordering L. If L belongs to U_α (V_α, W_α respectively) for some $\alpha \in \mathcal{O}$, then K also belongs to U_α (V_α, W_α respectively).*

Proof This property is easily proved, by an induction on α, using the definition of the classes U_α, V_α and W_α. It is based on the following fact. Suppose that L is equal to $\sum_{j \in J} L_j$. As a subordering of L, K is thus equal to $\sum_{j \in J} K_j$ where each K_j is a subordering of L_j. □

The proof of Theorem 15 finally uses the next lemma. Its proof is omitted since it can be easily established by induction on the rational expression E.

Lemma 18 *Let X be a rational set denoted by a rational expression E. Let F be a subexpression of E and Y the rational set denoted by F. Then, for any $y \in Y$, there exists $x \in X$ such that $|y|$ is a subordering of $|x|$.*

We now come to the proof of Theorem 15.

Proof of Theorem 15

($\Rightarrow$) The proof is by induction on the rational expression denoting X. If $X = \varnothing$, $X = \varepsilon$ or $X = a$ with $a \in A$, then for any $x \in X$, we have $|x| \in U_0 = V_0 = W_0$. The case $X = X_1 + X_2$ is trivial and the case $X = X_1X_2$ follows from the following fact. If two orderings J_1 and J_2 belong to U_{n_1} (V_{n_1}, W_{n_1} respectively) and U_{n_2} (V_{n_2}, W_{n_2} respectively), then the ordering $J_1 + J_2$ belongs to U_n (V_n, W_n respectively) where $n = 1 + \max(n_1, n_2)$. Suppose now that $X = Y^*$, $X = Y^\omega$ or $X = Y^{-\omega}$, with Y a rational set. Then any $x \in X$ is equal to $x = \prod_{j \in J} y_j$ with $y_j \in Y$ and $J \in \mathcal{N} \cup \{\omega, -\omega\}$. Suppose by induction hypothesis that

$$\exists n \geq 0 \; \forall y \in Y \quad |y| \in U_n \quad (V_n,\, W_n \text{ respectively}).$$

Since $J \in \mathcal{N} \cup \{\omega, -\omega\}$, it follows that $|x| \in U_{n+1}$ (V_{n+1}, W_{n+1} respectively). The remaining cases $X = Y^\sharp$ and $X = Y^{-\sharp}$ are solved similarly using orderings $J \in \mathcal{O} \cup -\mathcal{O}$.

($\Leftarrow$) Each of the three statements is proved in the same way.

1. Suppose that $F^\sharp$ is a subexpression of the rational expression denoting X. Let Z be the set denoted by F. We can suppose that $Z \neq \varnothing$ and $Z \neq \varepsilon$, otherwise the expression $F^\sharp$ can be replaced by ε. Take $z \neq \varepsilon$ in Z. Then the word $y = z^{\omega^\omega}$ belongs to $Z^\sharp$ and ω^ω is a subordering of its length. By Lemma 18, $|y|$ is a subordering of $|x|$ for some $x \in X$. Thus ω^ω is also a subordering of $|x|$. This is impossible by Lemmas 16 and 17.

If $F^{-\sharp}$ is a subexpression of the rational expression denoting X, then the same argument leads to the existence of a word x in X such that $-\omega^\omega$ is a subordering of $|x|$. By Lemmas 16 and 17, $|x|$ belongs to no V_n and hence to no U_n since $U_n \subseteq V_n$. This states the contradiction.

The argument is similar if $F_1 \diamond F_2$ is a subexpression of the rational expression denoting X. Let Z_1 and Z_2 be the nonempty sets denoted by F_1 and F_2. We can choose z_1 in Z_1 and z_2 in Z_2 such that $z_1z_2 \neq \varepsilon$. Let J be equal to the ordering ω^ω. One verifies that $\hat{J}^*$ is isomorphic to J. Consider the word y of $Z_1 \diamond Z_2$ equal to $\prod_{j \in J \cup \hat{J}^*} t_j$ with $t_j = z_1$ if $j \in J$ and $t_j = z_2$ if $j \in \hat{J}^*$. Then ω^ω is a subordering of its length and the contradiction holds as before.

2. If $F^{-\sharp}$ is a subexpression of the rational expression denoting X, we get a contradiction as before. If $F_1 \diamond F_2$ is a subexpression, then the proof is exactly the same as before but with the ordering $-\omega^\omega$, and we get again a contradiction.

3. Suppose that $F_1 \diamond F_2$ is a subexpression of the rational expression denoting X. The proof is now repeated with the ordering $J = \zeta^\omega$. One checks that J is a subordering of $\hat{J}^*$. The contradiction holds again. $\square$

7 An Order-Theoretic Excursion

This section is completely devoted to the proof of Lemma 16. The proof that the ordering ω^ω belongs to $U_\omega \setminus \bigcup_{n<\omega} U_n$ can be found in Chapter 5 of [21]. We first recall the notions of condensation and rank introduced in this reference which are necessary to obtain this result. We then show how to adapt these notions in a way to prove the other results of Lemma 16. The reader not interested in this order-theoretic excursion is free to skip the section.

For two orderings K and L, we write $K \simeq L$ to denote that they are isomorphic. We recall that the notation 1 is used for the ordering with one element. Let L be a linear ordering. A *condensation* of L is an equivalence relation on L such that each of its classes is an interval. If c is a condensation of L, the class of an element $x \in L$ is denoted $c(x)$. The quotient of L by c, that is the set of classes, is denoted by $c[L]$. Since disjoint intervals are naturally ordered, the quotient $c[L]$ can be endowed with a linear ordering induced by the ordering of L. If c is a condensation of L and if c' is a condensation of $c[L]$, the composition of c and c' is the condensation c^* of L such that the class $c^*(x)$ is $c'(c(x))$.

In the sequel, we consider condensations which are defined on all orderings. Such a condensation c is thus defined on any ordering L but also on the quotient $c[L]$. Therefore, the condensation can be applied again to $c[L]$ to get the ordering $c^2[L]$. More generally, such a condensation can be *iterated* as follows using an induction on the ordinal $\alpha \in \mathcal{O}$ as follows.

$$\begin{aligned} c^0(x) &= \{x\} \\ c^{\alpha+1}(x) &= \{y \mid c(c^\alpha(x)) = c(c^\alpha(y))\} \\ c^\alpha(x) &= \bigcup_{\beta<\alpha} c^\beta(x) && \text{if } \alpha \text{ is a limit ordinal.} \end{aligned}$$

A particular condensation is much studied in [21, p. 71]:

$$c_{\mathcal{N}}(x) = \{y \mid [x,y] \in \mathcal{N}\}$$

The interval notation $[x,y]$ is used even when $y < x$. In that case $[x,y]$ denotes the interval $[y,x]$. Thus, x and y are in the same interval with respect to condensation $c_{\mathcal{N}}$ iff the ordering $[x,y]$ is finite.

Note that $c_{\mathcal{N}}(x)$ may not be finite. It can be an infinite interval whose ordering is ω, $-\omega$ or ζ. Hence $c_{\mathcal{N}}[L] \simeq 1$ for any ordering L in $\mathcal{N} \cup \{\omega, -\omega, \zeta\}$. If $L = \omega + (-\omega)$ then $c_{\mathcal{N}}[L]$ is the finite ordering with two elements. Examples of iterated condensations are $c^n_{\mathcal{N}}[\omega^\omega] \simeq \omega^\omega$ for any $n < \omega$ and $c^\omega_{\mathcal{N}}[\omega^\omega] \simeq 1$.

A notion of *rank* $r_{\mathcal{N}}$ in relation with condensation $c_{\mathcal{N}}$ is proposed in [21, p. 82] as follows. If L is any ordering of $\mathcal{S}$, then there exists $\alpha \in \mathcal{O}$ such that $c^{\alpha}_{\mathcal{N}}[L] \simeq 1$. The smallest ordinal α satisfying that equality is called the rank $r_{\mathcal{N}}(L)$ of L. For instance, $r_{\mathcal{N}}(\omega) = 1$, $r_{\mathcal{N}}(\omega + (-\omega)) = 2$ and $r_{\mathcal{N}}(\omega^{\omega}) = \omega$.

Theorem 1 and the next theorem are proved thanks to the notions of iterated condensation and rank (see [21, p. 84]). The following theorem states that the rank of an ordering L is the ordinal α such that L belongs to U_{α} but L does not belong to any U_{β} for $\beta < \alpha$.

Theorem 19 *For any ordering $L \in \mathcal{S}$, one has*

$$r_{\mathcal{N}}(L) = \min\{\alpha \mid L \in U_{\alpha}\}.$$

As a corollary, we get the first result of Lemma 16 that the ordering ω^{ω} belongs to $U_{\omega} \setminus \bigcup_{n<\omega} U_n$.

For a class $\mathcal{C}$ of orderings, we denote by $c_{\mathcal{C}}$ the condensation defined by

$$c_{\mathcal{C}}(x) = \{y \mid [x, y] \in \mathcal{C}\}.$$

In what follows, we consider the three condensations for $\mathcal{C}$ equal to the class $\mathcal{N}$ of finite orderings, the class $\mathcal{O}$ of ordinals, and a new class $\mathcal{A}$ that we now introduce. The class $\mathcal{A}$ is defined by

$$\mathcal{A} = \{\sum_{j \in J} K_j \mid J \in \mathcal{N} \text{ and } K_j \in \mathcal{O} \cup -\mathcal{O}\}.$$

An ordering K belongs then to the class $\mathcal{A}$ if K is a finite sum of ordinals and reverse of ordinals. The condensation $c_{\mathcal{O}}$ has been already considered in [21, p. 72].

For $\mathcal{C}$ equal to $\mathcal{N}$, $\mathcal{O}$ or $\mathcal{A}$, we denote by $\widehat{\mathcal{C}}$ the class of orderings to which any $c_{\mathcal{C}}(x)$ belongs. For instance, we have seen before that $\widehat{\mathcal{N}} = \mathcal{N} \cup \{\omega, -\omega, \zeta\}$. It is proved in [21, p. 72] that $\widehat{\mathcal{O}} = \{\sum_{j \in -\omega} K_j \mid K_j \in \mathcal{O}\}$. It can also be analogously proved that $\widehat{\mathcal{A}} = \{\sum_{j \in \zeta} K_j \mid K_j \in \mathcal{O} \cup -\mathcal{O}\}$. It should be noted that any of the classes $\widehat{\mathcal{N}}$, $\widehat{\mathcal{O}}$ or $\widehat{\mathcal{A}}$ is closed under taking suborderings.

We now study iterated condensations $c^{\alpha}_{\mathcal{C}}$, with $\alpha \in \mathcal{O}$, for any class $\mathcal{C}$ equal to $\mathcal{N}$, $\mathcal{O}$ or $\mathcal{A}$.

Lemma 20 *Let x be an element of an ordering L. The following inclusions hold for any $\alpha \in \mathcal{O}$.*

$$c^{\alpha}_{\mathcal{N}}(x) \subseteq c^{\alpha}_{\mathcal{O}}(x) \subseteq c^{\alpha}_{\mathcal{A}}(x)$$

Proof We only give the proof of the first inclusion. The second one can be proved similarly using the properties that any class $\mathcal{C}$ is closed under taking suborderings and that $\mathcal{O} \subseteq \mathcal{A}$.

The proof is by induction on α. If $\alpha = 0$, then $c^{\alpha}_{\mathcal{C}}(x) = \{x\}$ for any class $\mathcal{C}$ and the inclusion holds trivially.

If α is a limit ordering, the proof is easily obtained by induction since $c^{\alpha}_{\mathcal{C}}(x) = \bigcup_{\beta<\alpha} c^{\beta}_{\mathcal{C}}(x)$.

Suppose that $\alpha = \beta+1$. Let us show that the inclusion $c^{\alpha}_{\mathcal{N}}(x) \subseteq c^{\alpha}_{\mathcal{O}}(x)$ holds. If $c_{\mathcal{N}}(c^{\beta}_{\mathcal{N}}(x)) = c_{\mathcal{N}}(c^{\beta}_{\mathcal{N}}(y))$, then by definition of $c_{\mathcal{N}}$, we have $[c^{\beta}_{\mathcal{N}}(x), c^{\beta}_{\mathcal{N}}(y)] \in \mathcal{N}$ where any $c^{\beta}_{\mathcal{N}}(z)$ is an element of the ordering $c^{\beta}_{\mathcal{N}}[L]$. By the induction hypothesis, the intervals $c^{\beta}_{\mathcal{N}}(z)$ and $c^{\beta}_{\mathcal{O}}(z)$ verify $c^{\beta}_{\mathcal{N}}(z) \subseteq c^{\beta}_{\mathcal{O}}(z)$ for any $z \in L$, and thus the ordering of $[c^{\beta}_{\mathcal{O}}(x), c^{\beta}_{\mathcal{O}}(y)]$ is a subordering of $[c^{\beta}_{\mathcal{N}}(x), c^{\beta}_{\mathcal{N}}(y)]$. Therefore $[c^{\beta}_{\mathcal{O}}(x), c^{\beta}_{\mathcal{O}}(y)] \in \mathcal{N} \subseteq \mathcal{O}$. It follows that $c_{\mathcal{O}}(c^{\beta}_{\mathcal{O}}(x)) = c_{\mathcal{O}}(c^{\beta}_{\mathcal{O}}(y))$. Hence $c^{\alpha}_{\mathcal{N}}(x) \subseteq c^{\alpha}_{\mathcal{O}}(x)$. □

For any class $\mathcal{C}$ equal to $\mathcal{N}$, $\mathcal{O}$ or $\mathcal{A}$ and for any ordering $L \in \mathcal{S}$, we define the rank $r_{\mathcal{C}}(L)$ to be the smallest ordinal α such that $c^{\alpha}_{\mathcal{C}}[L] \simeq 1$. This notion is well-defined thanks to properties of rank $r_{\mathcal{N}}$ and to the next corollary which follows directly from the previous lemma.

Corollary 21 *The following inequalities hold for any ordering L.*

$$r_{\mathcal{A}}(L) \leq r_{\mathcal{O}}(L) \leq r_{\mathcal{N}}(L).$$

Let us compute the rank of the orderings considered in Lemma 16.

Example 22 We have seen before that $r_{\mathcal{N}}(\omega^{\omega}) = \omega$. In the same way, we have $r_{\mathcal{N}}(-\omega^{\omega}) = \omega$. As the condensations $c_{\mathcal{N}}$ and $c_{\mathcal{O}}$ act similarly on the ordering $-\omega^{\omega}$, it follows that $r_{\mathcal{O}}(-\omega^{\omega}) = \omega$. One also checks that $r_{\mathcal{A}}(\zeta^{2n}) = n$ and $r_{\mathcal{A}}(\zeta^{\omega}) = \omega$.

The next proposition characterizes orderings of finite rank $r_{\mathcal{C}}$ for any class $\mathcal{C}$.

Proposition 23 *Let L be a linear ordering. Then*

$$\begin{array}{llll} 1. & r_{\mathcal{N}}(L) < \omega & \Longleftrightarrow & L \in \bigcup_{n<\omega} U_n \\ 2. & r_{\mathcal{O}}(L) < \omega & \Longleftrightarrow & L \in \bigcup_{n<\omega} V_n \\ 3. & r_{\mathcal{A}}(L) < \omega & \Longleftrightarrow & L \in \bigcup_{n<\omega} W_n \end{array}$$

The proof of this proposition is based on the following lemma. This lemma is stated in [21, p. 84] for the particular case $\mathcal{C} = \mathcal{N}$.

Lemma 24 *Let L be a linear ordering such that $L = \sum_{j\in J} K_j$ with $J \in \widehat{\mathcal{C}}$. If there exists $\alpha \in \mathcal{O}$ such that $r_{\mathcal{C}}(K_j) \leq \alpha$ for all $j \in J$, then $r_{\mathcal{C}}(L) \leq \alpha + 1$.*

Proof Let us denote by c_L (respectively c_{K_j}) the condensation $c_{\mathcal{C}}$ applied to L (respectively K_j).

We first show by induction on $\alpha \in \mathcal{O}$ that if K is an interval of L, then for any element x of K, the following equality holds.

$$c^{\alpha}_{K}(x) = c^{\alpha}_{L}(x) \cap K.$$

This equality will be applied later to the particular intervals K_j.

The cases $\alpha = 0$ and α a limit ordinal are immediate using the definition of $c_K^\alpha(x)$.

Suppose that $\alpha = \beta + 1$. Consider the intervals $c_L^\beta(z)$ partitioning the ordering L. The interval K divides at most two of them. Hence by induction hypothesis, the ordering of the $c_K^\beta(z)$ is the same as the ordering of the $c_L^\beta(z)$ restricted to the elements z of K. Thus applying condensation once more leads to $c_K^\alpha(x) = c_L^\alpha(x) \cap K$.

Let us now suppose that $r_{\mathcal{C}}(K_j) \leq \alpha$ for all $j \in J$. In other words, $c_{K_j}^\alpha(x) = K_j$ for all elements x of K_j. By the first part of the proof we have also $c_{K_j}^\alpha(x) = c_L^\alpha(x) \cap K_j$. Thus $K_j \subseteq c_L^\alpha(x)$. It follows that the ordering of the intervals $c_L^\alpha(x)$, $x \in L$, is a subordering of $J \in \widehat{\mathcal{C}}$. If one recalls that any class $\widehat{\mathcal{C}}$ is closed under taking suborderings, the application of the condensation c_L to these intervals $c_L^\alpha(x)$ leads to one interval equal to L. Therefore $c_L^{\alpha+1}(x) = L$ and $r_{\mathcal{C}}(L) \leq \alpha+1$. $\square$

Proof of Proposition 23 We first prove by induction on n, that any ordering L in U_n (respectively V_n and W_n) has a finite rank $r_{\mathcal{N}}(L)$ (respectively $r_{\mathcal{O}}(L)$ and $r_{\mathcal{A}}(L)$). The case $n = 0$ is trivial. The induction step follows directly from Lemma 24.

We now suppose that the rank $r_{\mathcal{N}}(L)$ (respectively $r_{\mathcal{O}}(L)$ and $r_{\mathcal{A}}(L)$) is equal to an integer n and we prove by induction on n that L belongs to U_m (respectively V_m and W_m) for some integer m. The case $n = 0$ is again trivial. By definition of the rank, one has $c_{\mathcal{C}}^n[L] \simeq 1$. It follows that $c_{\mathcal{C}}^{n-1}[L] \in \widehat{\mathcal{C}}$. The ordering L is then equal to $\sum_{j \in J} L_j$ where $J \in \widehat{\mathcal{C}}$ and $r_{\mathcal{C}}(L_j) = n - 1$. By the induction hypothesis the orderings L_j belong to U_m (respectively V_m and W_m) for some m. Since the ordering $J \in \widehat{\mathcal{C}}$ belongs to U_1 (respectively V_2 and W_2), it follows that L belongs to U_{m+1} (respectively V_{m+2} and W_{m+2}). $\square$

We conclude this section with the proof of Lemma 16. It is obtained as a direct consequence of Examples 5 and 22, and of Proposition 23.

8 Connections with Classes of Automata

In this section, we study the characterizations by automata of Classes 2, 4, 5 and 6 of the hierarchy. For each of these classes, we define a corresponding class of automata. These classes of automata are obtained by restricting the limit transitions which can occur in an automaton. We have the following theorem.

Theorem 25 *Let $X \subseteq A^\diamond$ be a rational set. Then*

1. *X belongs to* $\mathrm{Rat}_A(+, \cdot, *, \omega, -\omega)$ *if and only if the limit transitions of any trim automaton recognizing X satisfy*

$$P \to q \in E \implies q \notin P \quad \textit{and} \quad q' \to P' \in E \implies q' \notin P'$$

2. *X belongs to* $\mathrm{Rat}_A(+, \cdot, *, \omega, -\omega, \sharp)$ *if and only if the limit transitions of any trim automaton recognizing* X *satisfy*

$$q' \to P' \in E \quad \Longrightarrow \quad q' \notin P'$$

3. *X belongs to* $\mathrm{Rat}_A(+, \cdot, *, \omega, -\omega, \sharp, -\sharp)$ *if and only if the limit transitions of any trim automaton recognizing* X *satisfy*

$$\left.\begin{array}{l} P \to q \ \in E \quad \text{with } q \in P \\ q' \to P' \in E \quad \text{with } q' \in P' \end{array}\right\} \quad \Longrightarrow \quad P \neq P'$$

4. *X belongs to* $\mathrm{Rat}_A(+, \cdot, *, \omega, -\omega, \sharp, -\sharp, \diamond_1)$ *if and only if* X *is recognized by an automaton satisfying the following Condition* (†).

$$\left.\begin{array}{l} P \to q \in E \text{ with } q \in P \\ q' \to P' \in E \text{ with } q' \in P' \\ P = P' \end{array}\right\} \quad \Longrightarrow \quad \left\{\begin{array}{l} q = q' \\ \text{for all } R \subseteq P \text{ with } q \in R: \\ \qquad R \to q, q \to R \in E \end{array}\right.$$

It must be pointed out that the last statement of the theorem is different from the first three ones. It only states that for any set X in Class 6, there is *an* automaton which satisfies Condition (†) whereas the first three ones state that *any* automaton recognizing a set in Classes 2, 4 or 5 satisfies the given condition. This difference comes from the lack of a corresponding class of orderings for Class 6. Indeed the characterization of Classes 2, 4 and 5 by adequate orderings is used in the proof of Theorem 25.

It is required in the first three statements that the automaton is trim. This condition is of course necessary since transitions which cannot occur in a successful path can have any form. This assumption however is harmless since an automaton can easily be trimmed by removing useless states and transitions.

The proof of Theorem 25 follows the steps developed in [5] for the proof of Theorem 10. We recall the tools used in that proof in the next two subsections.

8.1 From Rational Expressions to Automata

The proof given in [5] that any rational set is recognized by an automaton is by induction on the rational expression denoting the set. For each rational operation, let us recall the corresponding construction for the automaton. These constructions apply to normalized automata defined as follows. An automaton is *normalized* if it has a unique initial state i and a unique final state $f \neq i$ and if it has no transition which enters i or leaves f. In particular, we can suppose that a normalized automaton has no transition of the form $P \to q$ or $q \to P$ where P contains i or f.

A normalized automaton never accepts the empty word. The next lemma shows that the empty word can be added or removed without changing recognizability (see [5] for a proof).

Lemma 26 *A set $X \subseteq A^\diamond$ is recognizable iff $X \setminus \{\varepsilon\}$ is recognizable. Furthermore if $\varepsilon \notin X$, then X can be recognized by a normalized automaton.*

Suppose that the sets X_1 and X_2 do not contain the empty word and that they are recognized by the normalized automata $\mathcal{A}_1 = (Q_1, E_1, I_1, F_1)$ and $\mathcal{A}_2 = (Q_2, E_2, I_2, F_2)$, such that Q_1 and Q_2 are disjoint. We recall the constructions of the automata for the sets $X_1 + X_2$, X_1X_2, X_1^*, X_1^ω, $X_1^{-\omega}$, $X_1^\sharp$, $X_1^{-\sharp}$, $X_1^\diamond$ and $X_1 \diamond X_2$. The situations where $\varepsilon \in X_1$ or $\varepsilon \in X_2$ will be considered later.

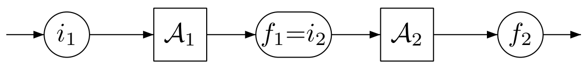


Figure 10: Automaton for X_1X_2

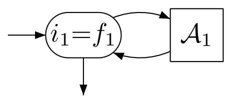


Figure 11: Automaton for X_1^*

The case of the union is easy, the constructions for the concatenation and the star operation are pictured in Figures 10 and 11.

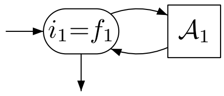
Figure 12: Automaton for X_1^ω

Let us consider the omega iteration (see Figure 12). The automaton has a new final state f and additional left limit transitions $P \to f$ for all P containing i_1. The construction for $X_1^{-\omega}$ is symmetrical and is thus omitted.

The construction for the ordinal iteration is pictured in Figure 13. The new automaton has additional left limit transitions $P \to i_1$ for all P containing i_1. The construction for $X_1^{-\sharp}$ is symmetrical and is thus omitted.

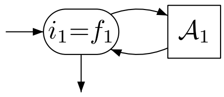
The construction of an automaton for $X_1^\diamond$ is the same as for $X_1^\sharp$ except that transitions $i_1 \to P$ for all P containing i_1 are also added (see Figure 14). The resulting automaton satisfies Condition (†).

When the $\diamond$ operation is binary, the automaton is obtained by merging the initial state i_1 of $\mathcal{A}_1$ and the final state f_2 of $\mathcal{A}_2$, by merging the initial state i_2 of $\mathcal{A}_2$ and the final state f_1 of $\mathcal{A}_1$, and by adding the left limit transition $P \to f_1$ and the right limit transition $i_1 \to P$ for all P containing both i_1 and f_1 (see Figure 15).



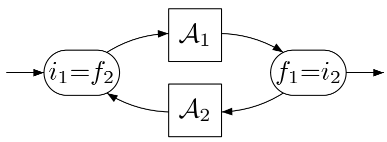
$$\{i_1, \ldots\} \to i_1$$

Figure 13: Automaton for $X_1^\sharp$



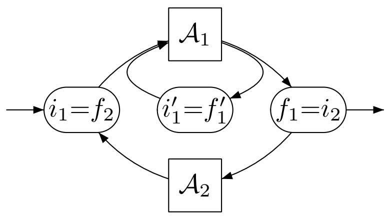
$$\{i_1, \ldots\} \to i_1$$
$$i_1 \to \{i_1, \ldots\}$$

Figure 14: Automaton for $X_1^\diamond$



$$\{i_1, f_1, \ldots\} \to f_1$$
$$i_1 \to \{i_1, f_1 \ldots\}$$

Figure 15: Automaton for $X_1 \diamond X_2$



$$\{i_1, f_1 \ldots\} \to f_1, i'_1$$
$$\{i'_1, \ldots\} \to f_1, i'_1$$
$$i_1, i'_1 \to \{i_1, f_1 \ldots\}$$
$$i_1, i'_1 \to \{i'_1, \ldots\}$$

Figure 16: Automaton for $X_1 \diamond X_2$ when $\varepsilon \notin X_1, \varepsilon \in X_2$

We conclude this list of constructions by treating the empty word. Any set X is decomposed as $X = X' + \varepsilon(X)$ where X' is $X \setminus \{\varepsilon\}$ and where $\varepsilon(X)$ is either ε if $\varepsilon \in X$ or $\varnothing$ otherwise. Recall that if X is recognizable, then X' can be recognized by a normalized automaton (Lemma 26). We have the following equalities $X_1X_2 = X_1'X_2' + \varepsilon(X_1)X_2 + X_1\varepsilon(X_2) + \varepsilon(X_1)\varepsilon(X_2)$, $X_1^* = {X_1'}^*$, $X_1^\omega = \varepsilon(X_1){X_1}^* + {X_1'}^\omega$, $X_1^\sharp = {X_1'}^\sharp$ and $X_1^\diamond = {X_1'}^\diamond$, showing that the previous constructions are sufficient. Similar equalities do not exist for $X_1 \diamond X_2$ which means that particular constructions have to be proposed. The case $\varepsilon \notin X_1$ and $\varepsilon \in X_2$ is described in Figure 16 where $\mathcal{A}_2$ is a normalized automaton accepting X_2'. The construction of Figure 15 must be slightly adapted but it remains essentially the same. The main difference is that the initial state i_1 and the final state f_1 of $\mathcal{A}_1$ are duplicated, as well as the related transitions (see the new states i_1' and f_1' in Figure 16). The other cases $\varepsilon \in X_1$ and $\varepsilon \notin X_2$, $\varepsilon \in X_1$ and $\varepsilon \in X_2$, are similar and left to the reader.

8.2 From Automata to Rational Expressions

We now recall the main steps of the proof given in [5] showing that any set of words recognized by an automaton is rational. It is a generalization of McNaughton and Yamada's algorithm.

Let $\mathcal{A} = (Q, A, E, I, F)$ be a fixed automaton. The *content* $\mathrm{C}(\gamma)$ of a path γ is the set of states which occur inside γ. It does not take into account the first and the last state of the path. A path γ from a state p to a state p' which is of content S and labeled by x is denoted by $\gamma : p \xrightarrow[S]{x} p'$. When $x \neq \varepsilon$, the path γ uses a first transition σ which leaves p and a last transition σ' which enters p'. The path is then denoted $\gamma : \sigma \xrightarrow[S]{x} \sigma'$.

Let S be a subset of states and let σ and σ' be two transitions of $\mathcal{A}$. The sets of words $\Pi^S_{\sigma,\sigma'}$, $\nabla^S_{\sigma,\sigma'}$, $\Delta^S_{\sigma,\sigma'}$, and $\Gamma^S_{\sigma,\sigma'}$ are defined as follows.

$$\begin{aligned}
\Pi^S_{\sigma,\sigma'} &= \{x \mid \sigma \xrightarrow[S]{x} \sigma'\} \\
\nabla^S_{\sigma,\sigma'} &= \{x \mid \sigma \xrightarrow[S]{x} \sigma' \text{ without any transition } S \to s\} \\
\Delta^S_{\sigma,\sigma'} &= \{x \mid \sigma \xrightarrow[S]{x} \sigma' \text{ without any transition } s \to S\} \\
\Gamma^S_{\sigma,\sigma'} &= \{x \mid \sigma \xrightarrow[S]{x} \sigma' \text{ without any transition } s \to S \text{ or } S \to s\}.
\end{aligned}$$

Note that *without any transition* $S \to s$ means that the path γ does not use any left limit transition of the form $S \to s$ for any $s \in \mathrm{C}(\gamma)$, except perhaps for the last transition if σ' is a left limit transition of this form. Similarly *without any transition* $s \to S$ means that γ does not use any right limit transition of this form except perhaps for the first transition.

Both sets $\nabla^S_{\sigma,\sigma'}$ and $\Delta^S_{\sigma,\sigma'}$ are subsets of $\Pi^S_{\sigma,\sigma'}$ and the set $\Gamma^S_{\sigma,\sigma'}$ is equal to the intersection $\nabla^S_{\sigma,\sigma'} \cap \Delta^S_{\sigma,\sigma'}$.

The set $X \setminus \{\varepsilon\}$ of nonempty words recognized by the automaton $\mathcal{A}$ is equal to the finite union

$$X \setminus \{\varepsilon\} \quad = \bigcup_{S \subseteq Q, \sigma \in \mathrm{Out}(I), \sigma' \in \mathrm{In}(F)} \Pi^S_{\sigma,\sigma'} \tag{1}$$

where $\mathrm{Out}(q)$ (respectively $\mathrm{In}(q)$) denotes the set of transitions leaving (respectively entering) state q. In [5], to show that X is rational, it is proved that $\Pi^S_{\sigma,\sigma'}$ is rational by induction on the cardinality of S. The case $S = \varnothing$ is easily solved since all four sets $\Pi^S_{\sigma,\sigma'}$, $\nabla^S_{\sigma,\sigma'}$, $\Delta^S_{\sigma,\sigma'}$ and $\Gamma^S_{\sigma,\sigma'}$ are equal and they are included in the alphabet A. The case $S \neq \varnothing$ is treated in four steps assuming that sets $\Pi^R_{\tau,\tau'}$, $\nabla^R_{\tau,\tau'}$, $\Delta^R_{\tau,\tau'}$ and $\Gamma^R_{\tau,\tau'}$ are rational for any $R \subsetneq S$ and any transitions τ, τ'. We briefly outline these four steps:

1. It is first proved that $\Gamma^S_{\sigma,\sigma'}$ is rational and this proof needs the classical Kleene's theorem for finite words as well as Kleene's theorem for ω-words. The set $\Gamma^S_{\sigma,\sigma'}$ is expressed as a rational expression with union, product, star iteration, omega iteration and backwards omega iteration over the alphabet $B = \{\Pi^R_{\tau,\tau'} \mid R \subsetneq S, \tau, \tau' \in E\}$:

$$\Gamma^S_{\sigma,\sigma'} \in \mathrm{Rat}_B(+, \cdot, *, \omega, -\omega). \tag{2}$$

2. It is secondly proved that the set $\Delta^S_{\sigma,\sigma'}$ is rational thanks to the induction hypothesis and the previous step. The rational expression which is obtained is over the alphabet $C = B \cup \{\Gamma^S_{\tau,\tau'} \mid \tau, \tau' \in E\}$ and it uses the additional rational operation $\sharp$.

$$\Delta^S_{\sigma,\sigma'} \in \mathrm{Rat}_C(+, \cdot, *, \omega, -\omega, \sharp). \tag{3}$$

3. The rationality of the set $\nabla^S_{\sigma,\sigma'}$ is proved similarly. The additional rational operation is here $-\sharp$.

$$\nabla^S_{\sigma,\sigma'} \in \mathrm{Rat}_C(+, \cdot, *, \omega, -\omega, -\sharp). \tag{4}$$

4. The last step showing that $\Pi^S_{\sigma,\sigma'}$ is rational is the most difficult. It is based on the induction hypothesis and the rationality of sets $\Gamma^S_{\tau,\tau'}$, $\Delta^S_{\tau,\tau'}$ and $\nabla^S_{\tau,\tau'}$. The operation $\diamond_2$ appears in this step.

$$\Pi^S_{\sigma,\sigma'} \in \mathrm{Rat}_D(+, \cdot, *, \omega, -\omega, \sharp, -\sharp, \diamond_2) \tag{5}$$

   with $D = C \cup \{\Delta^S_{\tau,\tau'} \mid \tau, \tau' \in E\} \cup \{\nabla^S_{\tau,\tau'} \mid \tau, \tau' \in E\}$.

8.3 Proof

In this subsection, we give the proof of Theorem 25.

Proof of Theorem 25

Proof of the implication ($\Rightarrow$) for each statement

1. Assume that the automaton recognizing X has a left limit transition $P \to q$ with $q \in P$. We can suppose that this transition occurs in a successful path. There are then a path from an initial state to q, a path from q to q with content P and a path from q to a final state. Let u, $v \neq \epsilon$ and w be the labels of these three paths. Each word of the set $uv^\sharp w$ is then accepted by the automaton. The word v^{ω^ω} belongs to the set $v^\sharp$. The ordering ω^ω is thus a subordering of the length of an accepted word. By Lemmas 16 and 17, there is a contradiction with the first statement of Theorem 15. This proves that the automaton cannot have transitions of the form $P \to q$ with $q \in P$. By symmetry, it cannot have transitions of the form $q' \to P'$ with $q' \in P'$.

2. Mimicking the proof of the previous case, we can show that if the automaton has a right limit transition $q' \to P'$ with $q' \in P'$, it accepts a subset $uv^{-\sharp}v$ for three words u, $v \neq \epsilon$ and w. The ordering $-\omega^\omega$ is thus a subordering of the length of an accepted word. This leads to a contradiction with the second statement of Theorem 15.

3. Assume that the automaton has a left limit transition $P \to q$ with $q \in P$ and a right limit transition $q' \to P$ with $q' \in P$. We can suppose that each of them occurs in a successful path. Thanks to the transition $P \to q$ and $q, q' \in P$, there is a path from an initial state to q', a path from q' to q with content P and a path from q to a final state. Let u, $v \neq \epsilon$ and w be the labels of these three paths. In the same way with the transition $q' \to P$ and $q, q' \in P$, there is a path from an initial state to q, a path from q to q' with content P and a path from q' to a final state. Their respective labels are u', $v' \neq \epsilon$ and w'. Then any word of the set $u(v \diamond v')w$ is accepted by the automaton. Consider the word $y \in v \diamond v'$ equal to $\prod_{j \in J \cup \hat{J}^*} t_j$ with $J = \zeta^\omega$, $t_j = v$ if $j \in J$ and $t_j = v'$ if $j \in \hat{J}^*$. Then the ordering ζ^ω is a subordering of the length of the word uyw accepted by the automaton. By Lemmas 16 and 17, there is a contradiction with the third statement of Theorem 15. Therefore, the automaton cannot have a left limit transition $P \to q$ with $q \in P$ and a right limit transition $q' \to P$ with $q' \in P$.

4. This last case is treated differently. Given a set X in $\mathrm{Rat}_A(+, \cdot, *, \omega, -\omega, \sharp, -\sharp, \diamond_1)$, the constructions described before in Subsection 8.1 yield an automaton accepting X that satisfies Condition (†).

Proof of the implication ($\Leftarrow$) for each statement

1. Consider an automaton $\mathcal{A}$ accepting a set X, such that $q \notin P$ and $q' \notin P'$ for any transition $P \to q$ or $q' \to P'$. Let us show that the sets $\Pi^S_{\sigma,\sigma'}$ and $\Gamma^S_{\sigma,\sigma'}$ are equal. With Equations (1) and (2), it then follows that $X \in \mathrm{Rat}_A(+, \cdot, *, \omega, -\omega)$. Recall that $\Gamma^S_{\sigma,\sigma'} \subseteq \Pi^S_{\sigma,\sigma'}$. Let us prove that the other inclusion also holds. Let γ be a path labeled by a word x belonging to $\Pi^S_{\sigma,\sigma'}$. Assume that a limit transition $S \to s$ or $s \to S$ is used inside this path, *i.e.*,

$s \in \mathrm{C}(\gamma)$. Then $s \in S$ since $S = \mathrm{C}(\gamma)$, which is impossible. Therefore, x belongs to $\Gamma^S_{\sigma,\sigma'}$ and the sets $\Pi^S_{\sigma,\sigma'}$ and $\Gamma^S_{\sigma,\sigma'}$ are equal.

2. The second case is solved in the same way as for Case 1. The sets $\Pi^S_{\sigma,\sigma'}$ and $\Delta^S_{\sigma,\sigma'}$ are equal and $X \in \mathrm{Rat}_A(+,\cdot,*,\omega,-\omega,\sharp)$ by Equations (1) and (3).

3. The same arguments are again used to show that the set $\Pi^S_{\sigma,\sigma'}$ is equal to $\Delta^S_{\sigma,\sigma'} \cup \nabla^S_{\sigma,\sigma'}$. Assume that a path γ labeled by a word $x \in \Pi^S_{\sigma,\sigma'}$ uses a left limit transition $S \to s$ and a right limit transition $s' \to S$ with $s, s' \in \mathrm{C}(\gamma)$. Then s and s' belong to S and we get a contradiction with the hypothesis.

4. This last case needs more attention because the operation $\diamond_1$ has not been studied in [5]. Let $\mathcal{A}$ be an automaton accepting a set X and suppose that it satisfies Condition (†). We are going to show an equation similar to (5) but with operation $\diamond_1$:

$$\Pi^S_{\sigma,\sigma'} \in \mathrm{Rat}_D(+,\cdot,*,\omega,-\omega,\sharp,-\sharp,\diamond_1). \tag{6}$$

Let $\gamma = (q_c)_{c\in\hat{J}}$ be a path labeled by a word $x = (a_j)_{j\in J}$ in $\Pi^S_{\sigma,\sigma'}$, with content S, first transition σ and last transition σ'. Let p and p' be the first and the last state of this path.

If γ has only left limit transitions $S \to s$, or has only right limit transitions $s \to S$, then

$$x \in \Delta^S_{\sigma,\sigma'} \cup \nabla^S_{\sigma,\sigma'}. \tag{7}$$

Otherwise, there exist a left limit transition $S \to s$ and a right limit transition $s' \to S$ inside the path γ, with $s, s' \in \mathrm{C}(\gamma) = S$. Note that for any such transitions, the states s, s' are equal, say to state q, by Condition (†). Consider the subordering K of $\hat{J}$ defined by the occurrences of limit transitions $S \to q$ and $q \to S$ inside γ, that is

$$K = \{c \in \hat{J} \mid \lim_{c^-} \gamma = S \text{ or } \lim_{c^+} \gamma = S\}.$$

One can verify that K has a least and a greatest element respectively denoted by $\min(K)$ and $\max(K)$.

If K has cardinality 2, take $c \in \hat{J}$ such that $\min(K) < c < \max(K)$. Hence the path γ decomposes as the path $p \to q_c$ followed by the path $q_c \to p'$ showing that

$$x \in \Delta^S_{\sigma,\tau'} \nabla^S_{\tau,\sigma'} \cup \nabla^S_{\sigma,\tau'} \Delta^S_{\tau,\sigma'} \tag{8}$$

with $\tau \in \mathrm{Out}(q_c)$, $\tau \neq q_c \to S$ and $\tau' \in \mathrm{In}(q_c)$, $\tau' \neq S \to q_c$.

Let us now suppose that K has more than two elements. Let $k < k'$ be two consecutive elements of K. Let $\tau_k \in \mathrm{Out}(q)$ be the transition leaving $q_k = q$ and $\tau'_{k'} \in \mathrm{In}(q)$ the transition entering $q_{k'} = q$. Then the label of the path $q_k \to q_{k'}$ belongs to $\Gamma^S_{\tau_k,\tau'_{k'}}$ or $\Pi^R_{\tau_k,\tau'_{k'}}$ for some $R \subsetneq S$.

Let us study the set

$$Y = T^{\diamond} \quad \text{where} \quad T = \bigcup_{\tau \in \mathrm{Out}(q), \tau' \in \mathrm{In}(q)} \left(\Gamma^S_{\tau,\tau'} \cup \bigcup_{R \subsetneq S} \Pi^R_{\tau,\tau'} \right). \tag{9}$$

Given an ordering $L \in \mathcal{S}$ and paths δ_l for $l \in L$ with respective labels $y_l \in T$, there exists a path δ with label $y = \prod_{l \in L} y_l$ obtained by concatenation of the paths δ_l. Indeed, any path δ_l is from state q to state q. Thus two such consecutive paths can be concatenated. More generally, let us show that any number of such consecutive paths can be concatenated. Suppose that this implies the presence of a left limit or a right limit. Such a limit is equal to some $R \subseteq S$ with $q \in R$. By Condition (†), the limit transitions $R \to q$ and $q \to R$ exist and can be used to concatenate the paths. Note that the content of δ is included in S.

Let us come back to the path γ. Inside γ, the path $\gamma' : q_{\min(K)} \to q_{\max(K)}$ has label x' which belongs to the set Y previously defined and its content is equal to S. This last statement is not imposed by the definition of Y. To impose a content equal to S, we are going to show that the word x' belongs to a set Y' defined by:

$$Y' = Y \left(\bigcup_{\tau \in \mathrm{Out}(q), \tau' \in \mathrm{In}(q)} \Gamma^S_{\tau,\tau'} \right) Y. \tag{10}$$

This is possible as follows. As K has at least three elements and it is a scattered ordering, it follows that there exist in K three consecutive elements $k < k' < k''$. Let δ and δ' be the paths $q_k \to q_{k'}$ and $q_{k'} \to q_{k''}$ inside the path γ. At least one, say δ, has its content equal to S by definition of K. Moreover, its label belongs to the set $\Gamma^S_{\tau,\tau'}$, for $\tau \in \mathrm{Out}(q_k)$ and $\tau' \in \mathrm{In}(q_{k'})$. Thus the word x' belongs to the set Y' as defined by (10).

We now treat different cases according to the form of the first and the last transitions σ and σ' of the path γ.

- **$\sigma \neq p \to S$ and $\sigma' \neq S \to p$.** As $\sigma \neq p \to S$, we have $\min(\hat{J}) < \min(K)$. Similarly with σ', we have $\max(K) < \max(\hat{J})$. The path γ decomposes as a path $\delta : p \to q_{\min(K)}$ followed by the path γ' and then by a path $\delta' : q_{\max(K)} \to p'$. The label of δ belongs to the set $Z = \bigcup_{\tau' \in \mathrm{In}(q)} (\Gamma^S_{\sigma,\tau'} \cup \bigcup_{R \subsetneq S} \Pi^R_{\sigma,\tau'})$. The label of δ' belongs to the set $Z' = \bigcup_{\tau \in \mathrm{Out}(q)} (\Gamma^S_{\tau,\sigma'} \cup \bigcup_{R \subsetneq S} \Pi^R_{\tau,\sigma'})$. Hence

$$x \in ZY'Z'. \tag{11}$$

Therefore, for this first case, $\Pi^S_{\sigma,\sigma'}$ is the union of the sets given in (7), (8) and (11). It thus belongs to $\mathrm{Rat}_D(+, \cdot, *, \omega, -\omega, \sharp, -\sharp, \diamond_1)$.

- **$\sigma = p \to S$ and $\sigma' \neq S \to p$.** As $\sigma = p \to S$, we now have $\min(\hat{J}) = \min(K)$. We first assume that the minimum of the set $K' = K \setminus \{\min(K)\}$

exists. We change the previous definition of the path δ to $\delta : p \to q_{\min(K')}$. Hence the definition of Z changes to $Z = \bigcup_{\tau' \in \mathrm{In}(q)} \Gamma^S_{\sigma,\tau'}$. We get $x \in ZY'Z'$ as in (11). We secondly assume that $\min(K')$ does not exist. It means that there exists an infinite decreasing sequence

$$\cdots < k_n < \cdots < k_1 < k_0$$

in K' whose infimum is equal to $\min(\hat{J})$. Without loss of generality we can suppose that for any $i \geq 1$, the content of the path $q_{k_i} \to q_{k_{i-1}}$ is equal to S. Hence the label of these paths belongs to Y' (see the definitions (9) and (10) of Y and Y'). It follows that

$$x \in Y'^{-\omega} Z'. \tag{12}$$

With (7), (8), (11) and (12), we have $\Pi^S_{\sigma,\sigma'} \in \mathrm{Rat}_D(+, \cdot, *, \omega, -\omega, \sharp, -\sharp, \diamond_1)$.

The case where $\sigma \neq p \to S$ and $\sigma' = S \to p$ is symmetrical and omitted.

• $\sigma = p \to S$ **and** $\sigma' = S \to p$**.** In this last case, there are a state s inside γ, and two transitions $\tau' \in \mathrm{In}(s)$, $\tau \in \mathrm{Out}(s)$ that are not limit transitions like $S \to s$ and $s \to S$. Hence, the path γ decomposes as a path $p \to s$ followed by a path $s \to p'$ and the word x belongs to a product $\Pi^S_{\sigma,\tau'}\Pi^S_{\tau,\sigma'}$ of subsets studied in the previous case. This proves that $\Pi^S_{\sigma,\sigma'} \in \mathrm{Rat}_D(+, \cdot, *, \omega, -\omega, \sharp, -\sharp, \diamond_1)$. This completes the proof. □

9 Conclusion

In this article, we investigate automata accepting sets of words indexed by linear orderings and rational expressions denoting such sets of words. In [5, 6] a Kleene-type equivalence is established between sets of words defined by automata vs. rational expressions. By properly restricting the classes of automata (of rational operations respectively), we here obtain several such equivalence results. Furthermore, except for one class (Class 6), we give close connections to the type of the underlying linear ordering. On the other hand, we show in this paper that the unary version of $\diamond$ is strictly weaker than its binary version.

Let us mention recent results and open problems in the framework of automata on linear orderings. It is proved in [11] that the emptiness problem for automata on linear orderings is decidable with a polynomial time complexity. The equivalence problem for automata on linear orderings is proved to be decidable in [8]. Other results related to combinatorics on words or equational theories can be found in [3, 4].

Since the work of Büchi, automata and logics have been shown to have strong connections. Connections between automata on linear orderings and adequate logics should be investigated. The first step of this study relies on the closure of the class of recognizable sets under the boolean operations. We recently know that this class is closed under complementation [12].

Acknowledgments

We would like to thank the referees for their numerous valuable comments on an earlier version of this article.

References

[1] N. Bedon. *Langages reconnaissables de mots indexés par des ordinaux.* Thèse de doctorat, Université de Marne-la-Vallée, 1998.

[2] N. Bedon and O. Carton. An Eilenberg theorem for words on countable ordinals. In Cláudio L. Lucchesi and Arnaldo V. Moura, editors, *Latin'98: Theoretical Informatics*, volume 1380 of *Lect. Notes in Comput. Sci.*, pages 53–64. Springer-Verlag, 1998.

[3] S. L. Bloom and Ch. Choffrut. Long words: the theory of concatenation and omega-power. *Theoret. Comput. Sci.*, 259:533–548, 2001.

[4] S. L. Bloom and Z. Esik. Axiomatizing omega and omega-op powers of words. *RAIRO Theoretical informatics*, 38:3–17, 2004.

[5] V. Bruyère and O. Carton. Automata on linear orderings. Technical Report 2000-12, Institut Gaspard Monge, 2000. Submitted.

[6] V. Bruyère and O. Carton. Automata on linear orderings. In J. Sgall, A. Pultr, and P. Kolman, editors, *MFCS'2001*, volume 2136 of *Lect. Notes in Comput. Sci.*, pages 236–247, 2001.

[7] V. Bruyère and O. Carton. Hierarchy among automata on linear orderings. In R. Baeza-Yates, U. Montanari, and N. Santoro, editors, *Foundation of Information technology in the era of network and mobile computing*, pages 107–118. Kluwer Academic Publishers, 2002. TCS'2002/IFIP'2002.

[8] V. Bruyère, O. Carton, and G. Sénizergues. Tree automata and automata on linear orderings. In Tero Harju and Juhani Karhumäki, editors, *WORDS'2003*, volume 27 of *TUCS General Publication*, pages 222–231, 2003.

[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. Transfinite automata recursions and weak second order theory of ordinals. In *Proc. Int. Congress Logic, Methodology, and Philosophy of Science, Jerusalem 1964*, pages 2–23. North Holland, 1965.

[11] O. Carton. Accessibility in automata on scattered linear orderings. In K. Diks and W. Rytter, editors, *MFCS'2002*, volume 2420 of *Lect. Notes in Comput. Sci.*, pages 155–164, 2002.

[12] O. Carton and C. Rispal. Complementation of rational sets on scattered linear orderings of finite rank. In *LATIN'2004*, volume 2976 of *Lect. Notes in Comput. Sci.*, pages 292–301, 2004.

[13] Y. Choueka. Finite automata, definable sets, and regular expressions over ω^n-tapes. *J. Comput. System Sci.*, 17(1):81–97, 1978.

[14] D. Girault-Beauquier. Bilimites de langages reconnaissables. *Theoret. Comput. Sci.*, 33(2–3):335–342, 1984.

[15] S. C. Kleene. Representation of events in nerve nets and finite automata. In C.E. Shannon, editor, *Automata studies*, pages 3–41. Princeton university Press, Princeton, 1956.

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

[17] R. McNaughton and H. Yamada. Regular expressions and state graphs for automata. *IEEE Trans. on Electronic Computers*, 9:39–47, 1960.

[18] D. Muller. Infinite sequences and finite machines. In *Switching Theory and Logical Design, Proc. Fourth Annual IEEE Symp.*, pages 3–16, 1963.

[19] 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.

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

[21] J. G. Rosenstein. *Linear ordering.* Academic Press, New York, 1982.

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

[23] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 4, pages 133–191. Elsevier, 1990.

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