Journal of Computer and System Sciences **63**, 394–431 (2001)
doi:10.1006/jcss.2001.1782, available online at http://www.idealibrary.com on IDEAL®

Logic over Words on Denumerable Ordinals

Nicolas Bedon

Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France
E-mail: Nicolas.Bedon@univ-mlv.fr
http://www-igm.univ.mlv.fr/~ bedon/

Received March 20, 2000; revised April 10, 2001

The main result of this paper is the extension of the theorem of Schützenberger, McNaughton, and Papert on star-free sets of finite words to languages of words of countable length. We also give another proof of the theorem of Büchi which establishes the equivalence between automata and monadic second-order sentences for defining sets of words of denumerable length.

Büchi [Büc60] was the first to use certain formulæ of logic, known as (monadic) second order formulæ, in order to define sets of finite words. Monadic second-order formulæ are built from (first-order) variables $x, y, \ldots$ representing positions in words, (second-order) variables $X, Y, \ldots$ representing sets of positions, an ordering relation $<$ between positions, a unary relation $X(x)$ which allows us to test whether the value of a first-order variable belongs to a set of positions or not, and, for every letter a of the alphabet, a unary relation $R_a(x)$ whose signification is to test if the letter at position x is an a. The goal of Büchi was to prove the decidability of such a logic. The main argument of the proof of this result consists in representing the formula with a Kleene automaton. Büchi showed a stronger result: logic formulæ are in fact equivalent to automata, in the sense that a set of words described with one formalism can also be described with the other. The restriction of this logic to first-order logic, that is, the same logic without second-order variables, was first investigated by McNaughton and Papert [MP71]. They proved that the class of languages obtained in this way is exactly the class of languages obtained from the letters by finite boolean operations and product. Such languages are called *star-free*. An algebraic formalism, the finite monoids, is also equivalent to second-order formulæ and automata to define sets of words. Schützenberger [Sch65] proved a strong result of characterization of star-free languages; the star-free languages are exactly those defined by a finite group-free monoid. Thus, first-order formulæ, finite aperiodic monoids and star-free expressions, both define the same class of languages of finite words.

0022-0000/01 $35.00

Finite automata on ω-words, that is, words whose letters are indexed by all nonnegative integers, were first introduced by Büchi [Büc62] to extend his result on finite words to infinite words. The logic formulæ remain the same, but they are interpreted using ω-words instead of finite words. Büchi's automata are like Kleene's, but with an accepting condition adapted to the recognition of infinite words. Those automata have been widely studied since their introduction by Büchi. A first attempt in the direction of the algebraic approach to the theory of ω-words was made by Pécuchet [Péc86a, Péc86b], but a more satisfying one is due to Wilke [Wil91] and Perrin and Pin [PP97] with the introduction of ω-semigroups. The result on star-free sets on finite words was extended to ω-words by Ladner [Lad77], Thomas [Tho79], and Perrin [Per84]: the star-free languages of ω-words are exactly those recognized by finite groups-free ω-semigroups or equivalently those defined by first-order formulæ.

Büchi, in [Büc64], generalized his idea of automata recognizing ω-words to transfinite words, i.e., words whose letters are indexed by ordinals. He defined, among others, classes of automata recognizing words of length less than ω^n, where n is a given integer. We proved [Bed98b, Bed98a] that those automata are equivalent to a generalization of ω-semigroups that are finite algebraic structures called ω^n-semigroups. We also extended the star-free results on finite and ω-words to languages of words of length less than ω^n: again, the star-free languages of ω^n-words are exactly those recognized by finite group-free ω^n-semigroups or equivalently those defined by first-order formulæ.

Since logicians are not only interested in small ordinals Büchi also worked on automata recognizing words whose length is a countable ordinal. He proved again the equivalence between his automata and second-order formulæ to define sets of words. We introduced [Bed98b, BC98] an algebraic structure, the ω_1-semigroups, adapted to study of the languages recognized by these automata.

In this paper we extend the star-free results on finite words, ω-words, and words of length less than ω^n to languages of words of countable length: again, the star-free languages of words of countable length are exactly those recognized by finite group-free ω_1-semigroups or equivalently those defined by first-order formulæ. We also give another proof of the result of Büchi which establishes the equivalence between automata and second-order formulæ to define sets of words of countable length. The proof we give is an extension of an elegant one from Straubing [Str94] for finite word case.

Some knowledge of ordinals is required to read the paper. Although we tried to obtain a self-contained paper, some previous knowledge of automata and semigroups is also welcome.

1. NOTATION AND DEFINITIONS

For the theory of ordinals we refer to [Sie65] or [Ros82]. We note by *Succ* the class of successor ordinals, *Lim* the class of limit ordinals and $Ord = Succ \cup Lim \cup \{0\}$. As usual we identify the linear order on ordinals with the membership. An ordinal α is then identified with the set of all ordinals smaller than α. If

$\omega^{\alpha_1} \cdot n_1 + \omega^{\alpha_2} \cdot n_2 + \cdots + \omega^{\alpha_k} \cdot n_k$ is the Cantor normal form of an ordinal α the *end* of α, noted by *end*(α), is ω^{α_k}. An increasing sequence $(\alpha_\beta)_{\beta<\gamma}$ of ordinals less than ξ is *cofinal* with an ordinal ξ if, for any $\delta < \xi$, there exists $\beta < \gamma$ such that $\delta < \alpha_\beta < \xi$. Observe that this implies that $\xi \in Lim$. The first ordinal of uncountable cardinality is noted ω_1. We recall the following well-known theorem on limit ordinals less than ω_1:

THEOREM 1.1. *Let $\xi < \omega_1$ be a limit ordinal. There exists an increasing sequence $(\alpha_i)_{i<\omega}$ of ordinals less than ξ which is cofinal with ξ. Furthermore, there does not exist such a sequence for ω_1, which is the smallest limit ordinal having this property.*

Let α be an ordinal and A a finite set. The set A is usually called an *alphabet*. Each element of an alphabet is a *letter*. A *word* u of *length* α on A is a function $u: \alpha \to A$, which associates a letter to any position in the word. A position in the word is any ordinal less than α. A word u of length α can also be seen as sequence $u = (u_\beta)_{\beta<\alpha}$ of α letters (or α-*sequence*) of A. For this reason we sometimes allow ourselves to take one for the other in the remainder of the paper. The word of length 0 is the *empty word* and is noted by λ.

EXAMPLE 1.1. Let $A = \{a, b, c\}$. The word u of length 2 on A defined by $u(0) = a$ and $u(1) = b$ (or equally $u_0 = a$ and $u_1 = b$) is the only word of length 2 whose first letter in an "a" and second letter is a "b." For practical reasons u is also noted by mere concatenation : $u = ab$.

EXAMPLE 1.2. Let $A = \{a, b\}$. The word u of length ω defined by $u_{2k} = a$ and $u_{2k+1} = b$ for any integer k is the only word in which the indexes of the letters are exactly all the integers and formed by infinite (ω) repetition of ab: "a" appears at even positions, "b" at odd positions.

EXAMPLE 1.3. Let $A = \{a, b\}$. The word u of length $\omega+2$ defined by $u_{2\alpha} = a$, with $\alpha \leqslant \omega$, and whose other letters are a "b" is the only word of length $\omega+2$ formed by infinite ($\omega+1$) repetition of ab.

Let u be a word of length α on a finite set A_u and v be a word of length β on a finite set A_v. The *product* of u and v, noted by $u \cdot v$, or uv for short, is the word u of length $\alpha + \beta$ on $A_u \cup A_v$ such that:

$$w_\gamma = \begin{cases} u_\gamma & \text{if } \ 0 \leqslant \gamma < \alpha \\ u_{\gamma-\alpha} & \text{if } \ \alpha \leqslant \gamma < \alpha + \beta. \end{cases}$$

Clearly the product of words is an associative operation. If $(u_\beta)_{\beta<\alpha}$ is a sequence of words over A then $\prod_{\beta<\alpha} u_\beta$ is the word v of length $\sum_{\beta<\alpha} |u_\beta|$ defined by, for any $\gamma < |v|$,

$$u(\gamma) = u_\delta(\gamma - \varepsilon),$$

where δ is the greatest possible ordinal such that $\sum_{\beta<\delta} |u_\beta| \leqslant \gamma$ and $\varepsilon = \sum_{\beta<\delta} |u_\beta|$.

EXAMPLE 1.4. Let u be the word of Example 1.1 and v the word of Example 1.2. The product of v and u is the word of Example 1.3. Observe that the product of words is not a commutative operation, since in this example $uv = v \neq vu$.

If $w = xyz$ then x, y, and z are called *factors* of w, x a left factor (or *prefix*) of w and z a right factor (or *suffix*) of w. A factor of w is *proper* if it is different from w. Let α and β be ordinals with $\alpha < \beta$ and u a word such that $|u| \geqslant \beta$. By $u[\alpha, \beta[$ we note the word of length $\beta - \alpha$ such that $u[\alpha, \beta[(\gamma) = u(\alpha + \gamma)$ for any $0 \leqslant \gamma < \beta - \alpha$. A decomposition of a word u into a product of factors $\prod_{\beta < \alpha}(v_\beta)$ is called a *factorization* of u. The factorization is *cofinal* with $|u|$ if $(\sum_{\beta < \gamma} |v_\beta|)_{\gamma < \alpha}$ is cofinal with $|u|$. Let A be an alphabet, α, β ordinals such that $\beta < \alpha$ and n an integer. We note by A^α the set of all words on A of length α, $A^{<\alpha}$ is the set of all words on A of length less than α, and $A^{[\beta, \alpha[}$ the set of all words on A of length γ such that $\beta \leqslant \gamma < \alpha$. A set of words is also called a *language*. Let L be a language and α an ordinal. Then

$$L^\alpha = \left\{ u \colon u = \prod_{\beta < \alpha} v_\beta \text{ with } v_\beta \in L \text{ for any } \beta < \alpha \right\} \qquad \text{and} \qquad L^{<\alpha} = \bigcup_{\beta < \alpha} L^\beta.$$

If L_1 and L_2 are two languages, $L_1 \cup L_2$ is sometimes noted by $L_1 + L_2$. For practical reasons, a language composed of only one word u is sometimes simply noted u.

The powerset of a set S is noted by $\mathscr{P}(S)$, $\mathscr{P}(S) - \{\varnothing\}$ by $[S]$, $[S] \cup S$ by $[S]_0^1$, and the cardinal of S by $|S|$.

1.1. Automata

In this section we give the definition of automata used in this paper.

Büchi automata [Büc65] on transfinite words are a generalization of usual (Kleene) automata on finite words, with a second transition function for limit ordinals. States reached at limit points depend only on states reached before.

DEFINITION 1.1. An automaton $\mathscr{A}$ is a 5-tuple (Q, A, E, I, F) where Q is the finite set of states, A a finite alphabet, $E \subseteq ([Q]_0^1 \times A \times Q)$ the set of transitions, $I \subseteq Q$ the set of initial states, and $F \subseteq [Q]_0^1$ the set of final elements.

We now explain how these automata are used to define languages. In order to define the notion of path we need the following definition:

DEFINITION 1.2. Let E be a nonempty set, α a countable ordinal, and $e = (e_\beta)_{\beta < \alpha}$ a sequence of elements of $[E]_0^1$. The sequence e is *continuous* if

- $e_\beta \in E$ for every nonlimit ordinal β less than α,
- $e_\xi \in [E]$ for every limit ordinal ξ less than α,
- if ξ is a limit ordinal less than α and $q \in E$ then $q \in e_\xi$ iff there exists an increasing ω-sequence $(\alpha_i)_{i < \omega}$ of nonlimit ordinals less than ξ which is cofinal with ξ such that $q = e_{\alpha_i}$ for every integer i.

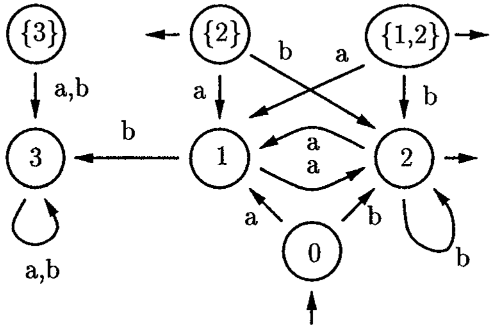

FIG. 1. An automaton recognizing $(aa+b)^{<\omega_1}-\lambda$.

Observe that the values of a continuous sequence at limit points are entirely determined by the ones indexed by smaller successor ordinals.

We are now ready to define paths in an automaton:

DEFINITION 1.3. Let $\mathscr{A}=(Q, A, E, I, F)$ be an automaton, $p \in Q$, and $p' \in [Q]_0^1$. A *path* c of length α from p to p' in $\mathscr{A}$ is a countable continuous $(\alpha+1)$-sequence of elements of $[Q]_0^1$ such that

- $c_0=p$ and $c_\alpha=p'$,
- for any $\beta<\alpha$ there exists $a_\beta \in A$ such that $(c_\beta, a_\beta, c_{\beta+1})$ is a transition of $\mathscr{A}$.

The word $u=(a_\gamma)_{\gamma<\alpha}$ is called a *label* of c. The path is *successful* iff $p \in I$ and $p' \in F$. We note by $\mathscr{L}(\mathscr{A})$ the class of labels of successful paths. A word is *accepted* (or *recognized*) by $\mathscr{A}$ iff it belongs to $\mathscr{L}(\mathscr{A})$. We say that a language L is *accepted* (or *recognized*) by $\mathscr{A}$ if $L=\mathscr{L}(\mathscr{A})$.

An automaton (Q, A, E, I, F) can be pictured as a graph in which the nodes represent the elements of $[Q]_0^1$. The edges are labeled and represent the transitions. Initial states have a small ingoing arrow, and the elements of F a small outgoing arrow.

EXAMPLE 1.5. Let $A=\{a, b\}$. The automaton $\mathscr{A}$ in Fig. 1 recognizes $(aa+b)^{<\omega_1}-\lambda$.

EXAMPLE 1.6. Let $A=\{a, b\}$. The automaton $\mathscr{A}$ drawn in Fig. 2 recognizes $(ab)^{<\omega_1}-\lambda$.

The reader familiar with automata on ω-words should have noticed that this definition of automata extends the definition of usual Muller automata [Mul63] on ω-words. We refer the reader interested in rational expressions equivalent to this kind of automata to [Woj84, Woj85].

This theorem will be needed in the remainder of the paper.

THEOREM 1.2 (Büchi). *Let A be an alphabet and $\mathscr{A}$ and $\mathscr{B}$ be two automata recognizing languages on A. Automata recognizing $\mathscr{L}(\mathscr{A}) \cup \mathscr{L}(\mathscr{B})$, $\mathscr{L}(\mathscr{A}) \cap \mathscr{L}(\mathscr{B})$, and $A^{<\omega_1}-\mathscr{L}(\mathscr{A})$ can effectively be built from $\mathscr{A}$ and $\mathscr{B}$.*

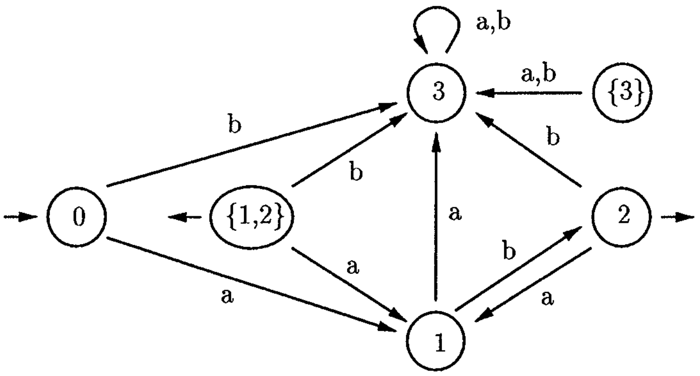

FIG. 2. An automaton recognizing $(ab)^{<\omega_1} - \lambda$.

1.2. Semigroups

A *semigroup* S is a set equipped with an internal associative function written in multiplicative form; for short we write xy instead of $x \cdot y$. An element s of S is the *zero* of S if $xs = sx = s$ for any $x \in S$. It is called the *neutral element* of S if $xs = sx = x$ for any $x \in S$. A zero is usually noted by 0 and a neutral element by 1. An element e of a semigroup is called *idempotent* if $e^2 = e$. The set of all idempotents of a semigroup S is noted $E(S)$. A pair (s, e) of elements of S is *linked* if $se = s$ and e is an idempotent. It is well known that each element of a finite semigroup S has an idempotent power (that is, for every $s \in S$, there exists an integer n_s such that $(s^{n_s})^2 = s^{n_s}$). The least common multiple of all such n_s is called the *exponent* of S and is usually noted by π. A semigroup S is *aperiodic* if there exists an integer (called the *index* of S) n such that for any $s \in S$, $s^n = s^{n+1}$. A *monoid* is a semigroup with an identity, usually noted 1. Let S be a semigroup. A *subsemigroup* S' of S is a subset of S such that S' is a semigroup. We note by S^1 the monoid $S \cup \{1\}$ if S is not a monoid, S otherwise. A subset I of a semigroup S is an *ideal* of S iff $S^1IS^1 = I$. A *morphism* between two algebraic structures of the same kind is a function preserving operations. For example, if S and T are two semigroups and φ is a morphism from S to T, then for all x, y in S, $\varphi(x \cdot y) = \varphi(x) \cdot \varphi(y)$. A semigroup T is a *quotient* of a semigroup S if there exists a surjective morphism $\varphi: S \to T$. A *congruence* is an equivalence relation preserving operations, usually noted $\sim$. For example, a semigroup congruence $\sim$ verifies $x \sim y \Rightarrow uxv \sim uyv$. This condition ensures that the set of equivalence class $S/\sim$ can naturally be equipped with an associative product and that the mapping which associates to an element its equivalence class is a (surjective) semigroup morphism. This remark is also true for algebras more complex than semigroups. If $\sim_1$ and $\sim_2$ are two congruences on an algebraic structure S we say that $\sim_1$ is a *refinement* of $\sim_2$ if and only if, for every $x, y \in S$, $x \sim_1 y \Rightarrow x \sim_2 y$. It is well known that finite semigroups are equivalent to usual automata on finite words to define sets of words and that to any rational language one can attach a canonical finite semigroup. A similar result holds in the theory of ω-words.

The relations we introduce now enable the study of the multiplicative structure of finite semigroups. The four preorders (reflexive and transitive relations) $\leqslant_{\mathscr{D}}$, $\leqslant_{\mathscr{R}}$, $\leqslant_{\mathscr{L}}$, $\leqslant_{\mathscr{H}}$ are defined by:

$$s_1 \leqslant_{\mathscr{D}} s_2 \Leftrightarrow S^1 s_1 S^1 \subseteq S^1 s_2 S^1$$

$$s_1 \leqslant_{\mathscr{R}} s_2 \Leftrightarrow s_1 S^1 \subseteq s_2 S^1$$

$$s_1 \leqslant_{\mathscr{L}} s_2 \Leftrightarrow S^1 s_1 \subseteq S^1 s_2$$

$$s_1 \leqslant_{\mathscr{H}} s_2 \Leftrightarrow s_1 \leqslant_{\mathscr{R}} s_2 \text{ et } s_1 \leqslant_{\mathscr{L}} s_2.$$

In other words, $s_1 \leqslant_{\mathscr{D}} s_2$ iff there exist $t_1, t_2 \in S^1$ such that $s_1 = t_1 s_2 t_2$, $s_1 \leqslant_{\mathscr{R}} s_2$ iff there exist $t \in S^1$ such that $s_1 = s_2 t$, and $s_1 \leqslant_{\mathscr{L}} s_2$ iff there exist $t \in S^1$ such that $s_1 = t s_2$.

We deduce from these four preorders the following four equivalence relations, known as *Green's relations*:

$$\begin{array}{lcl} s_1 \mathscr{D} s_2 & \Leftrightarrow & s_1 \leqslant_{\mathscr{D}} s_2 \text{ and } s_2 \leqslant_{\mathscr{D}} s_1 \\ s_1 \mathscr{R} s_2 & \Leftrightarrow & s_1 \leqslant_{\mathscr{R}} s_2 \text{ and } s_2 \leqslant_{\mathscr{R}} s_1 \\ s_1 \mathscr{L} s_2 & \Leftrightarrow & s_1 \leqslant_{\mathscr{L}} s_2 \text{ and } s_2 \leqslant_{\mathscr{L}} s_1 \\ s_1 \mathscr{H} s_2 & \Leftrightarrow & s_1 \leqslant_{\mathscr{H}} s_2 \text{ and } s_2 \leqslant_{\mathscr{H}} s_1. \end{array}$$

In particular, $s_1 \mathscr{H} s_2$ iff $s_1 \mathscr{R} s_2$ and $s_1 \mathscr{L} s_2$.

If $\mathscr{K} \in \{\mathscr{D}, \mathscr{R}, \mathscr{L}, \mathscr{H}\}$ we write $s_1 <_{\mathscr{K}} s_2$ if $s_1 \leqslant_{\mathscr{K}} s_2$ and $s_2 \not\leqslant_{\mathscr{K}} s_1$. An equivalence class for $\mathscr{K}$ is called a $\mathscr{K}$*-class*. The $\mathscr{K}$-class of $s \in S$ is noted by $\mathscr{K}(s)$.

We will use the following propositions on elements of semigroups. We refer to [Pin84] for proofs.

PROPOSITION 1.1. *Let S be a semigroup and e, e' two idempotents of S. Then $e \mathscr{D} e'$ iff there exist $x, y \in S$ such that $e = xy$ and $e' = yx$.*

PROPOSITION 1.2. *Let S be a finite semigroup and a and b be two elements of S. If $a \mathscr{D} b$ and $a \leqslant_{\mathscr{R}} b$ (resp. $a \leqslant_{\mathscr{L}} b$) then $a \mathscr{R} b$ (resp. $a \mathscr{L} b$).*

We also have this property on $\mathscr{H}$-classes:

PROPOSITION 1.3. *Let S be a finite semigroup. An $\mathscr{H}$-class of S containing an idempotent e is a group whose neutral element is e.*

We now describe algebraic structures adapted to the study of words whose length is a denumerable ordinal. Those structure where introduced in [Bed98b, BC98]. The following theorem, whose proof uses Ramsey-type arguments, lays the foundations for extending finite semigroups in order to deal with words of infinite length:

THEOREM 1.3. *Let S be a finite semigroup and $x = (x_i)_{i<\omega}$ be an ω-sequence over S. There exist a cofinal factorization $\prod_{i<\omega} x[k_i, k_{i+1}[$ of x and a linked pair (s, e) of S such that $x_{k_0} \ldots x_{k_1 - 1} = s$ and $x_{k_j} \ldots x_{k_{j+1}-1} = e$ for $0 < j < \omega$.*

We now define the notion of an ω_1-semigroup. Roughly speaking, an ω_1-semigroup is a set S equipped with a product which maps any sequence of countable length over S to an element of S. This notion generalizes the usual notion of a semigroup where the product is defined on finite sequences of elements. Semigroups adapted to ordinals, in particular ω_1-semigroups, were introduced in [Bed98b].

DEFINITION 1.4. An ω_1-*semigroup* is a set S equipped with a function φ: $S^{[1, \omega_1[} \to S$, called *product*, which satisfies the following properties

1. For any element $s \in S$, $\varphi(s) = s$.
2. For any word x of countable length over S, and any factorization

$\prod_{\gamma<\beta} x[\alpha_\gamma, \alpha_{\gamma+1}[$ of x, then

$$\varphi(x) = \varphi\left(\prod_{\gamma<\beta} \varphi(x[\alpha_\gamma, \alpha_{\gamma+1}[)\right).$$

The second condition on φ ensures that it verifies an extension of the associativity of the product semigroups are endowed with.

The following example corresponds to the free semigroup A^+ over a finite alphabet A.

EXAMPLE 1.7. Let A be an alphabet and let $A^{<\omega_1}$ be the set of words over A of countable length. The concatenation maps any sequence of words of $A^{<\omega_1}$ to a word of $A^{<\omega_1}$. It can easily be verified that $A^{<\omega_1}$ equipped with the concatenation as the product is an ω_1-semigroup. This ω_1-semigroup is actually the free ω_1-semigroup on A.

Since the description of products of elements of infinite sequences is infinite, ω_1-semigroups are not very interesting objects, even if the number of elements is finite. Wilke [Wil91] proved that when a semigroup S is finite, then the product of elements of ω-sequences of elements of S is entirely determined by the infinite products of the form $s^\omega = ssss\ldots$. In other words, the product of elements of an ω-sequence is entirely determined by the products of elements of ω-sequences composed of the same element. As a consequence, since every countable limit ordinal is the limit of an increasing ω-sequence, products of elements of α-sequences, with α a countable ordinal, are entirely determined by the product of elements of finite sequence and by the products of elements of ω-sequences composed of the same element. Thus, ω_1-semigroups really become finite objects when they have a finite number of elements (i.e., finite ω_1-semigroups are objects of finite signature).

Definition 1.5. An ω_1-*Wilke algebra* S is a finite semigroup equipped with an internal unary operation $\omega: S \to S$ (noted in a postfix form) verifying, for any $s, t \in S$

- $s(ts)^\omega = (st)^\omega$,
- $(s^n)^\omega = s^\omega$ for any positive integer n.

We refer to [Bed98b] for the proof of the following theorem:

Theorem 1.4. *Let S be a finite set equipped with an associative binary product $\cdot$ and with a unary map $\omega: S \to S$ verifying the properties of the ω operator of Definition* 1.5. *Then S can be endowed in an unique manner of a product φ verifying:*

- $\varphi((x_i)_{i<2}) = x_0 \cdot x_1$ *for any* $x_0, x_1 \in S$,
- $\varphi(t) = s^\omega$ *for any* $s \in S$, *if t is the ω-sequence whose elements are all equal to s,*
- *conditions on the φ application of Definition* 1.4.

Conversely, if S is a finite set equipped with an application φ that makes S a finite ω_1-semigroup, then S can be endowed in a unique manner of an associative binary product $\cdot$ and a unary map $\omega: S \to S$ such that

- $\varphi((x_i)_{i<2}) = x_0 \cdot x_1$ *for any* $x_0, x_1 \in S$,
- $\varphi(t) = s^\omega$ *for any* $s \in S$, *if t is the ω-sequence whose elements are all equal to s,*
- *ω verifies conditions on the ω application of Definition* 1.5.

The previous theorem justifies that from now on we shall not distinguish between finite ω_1-semigroups and ω_1-Wilke algebras. We also shall not differenciate between s^ω and the product of the elements of an ω-sequence whose elements are all equal to s.

Even if the notion of an ω_1-semigroup does not really fit into the general framework of a universal algebra, the following notions are self-understanding: morphism of ω_1-semigroups, quotient of ω_1-semigroups, sub-ω_1-semigroup, congruence of ω_1-semigroups. For an ω_1-semigroup S, we note by S^1 the ω_1-semigroup obtained by adding a neutral element to S. We say that a morphism of ω_1-semigroups $\varphi: S \to T$ *recognizes* a subset X of S if $\varphi^{-1}\varphi(X) = X$. This subset X is *recognizable* if there exist a finite ω_1-semigroup T' and a morphism $\varphi': S \to T'$ of ω_1-semigroups such that φ' recognizes X. We also say that T *recognizes* X if there exists $\varphi'': S \to T$ such that $\varphi''^{-1}\varphi''(X) = X$.

By extension to the finite words case, with any recognizable subset X of an ω_1-semigroup S, one can effectively associate a canonical finite ω_1-semigroup $\mathrm{synt}(X)$ which divides any ω_1-semigroup recognizing X. This syntactic ω_1-semigroup is the quotient of S by a syntactic congruence $\sim_X$ we now define. This syntactic congruence is actually the counterpart of Arnold's congruence [Arn85] for recognizable languages of ω-words.

DEFINITION 1.6. Let S be an ω_1-semigroup and X a subset of S. For any $x, y \in S$, we say that $x \sim_X y$ iff for any positive integer m and any elements $s_0, \ldots, s_m \in S^1$,

$$s_0(\ldots(((xs_1)^\omega s_2)^\omega s_3)^\omega \ldots)^\omega s_m \in X \Leftrightarrow s_0(\ldots(((ys_1)^\omega s_2)^\omega s_3)^\omega \ldots)^\omega s_m \in X.$$

For $m=1$, the expression $s_0(\ldots(((xs_1)^\omega s_2)^\omega s_3)^\omega \ldots)^\omega s_m$ should be understood as $s_0 x s_1$.

THEOREM 1.5. *Let X be a recognizable subset of an ω_1-semigroup S. The relation $\sim_X$ is a congruence of an ω_1-semigroup of finite index and the quotient $S/\sim_X$ of S by $\sim_X$ divides any ω_1-semigroup recognizing X. The ω_1-semigroup $S/\sim_X$ is called the syntactic ω_1-semigroup* of X and is noted by synt(X).

In particular, synt(X) is finite and is smaller than any ω_1-semigroup recognizing X.

EXAMPLE 1.8. Let $A=\{a, b\}$ be an alphabet and $S=\{a, b, 0, 1, ab, ba, aba, a^\omega, a^\omega a\}$ the ω_1-semigroup whose $\mathscr{D}$-classes structure is[1]

* 1, a

* b	ba
ab	* aba
* a^ω	* $a^\omega a$

* 0

and such that 0 is a zero for S, 1 the neutral element of S for the product only, $a^2=1$, $bab=0$, $ba^\omega=b$, $a^\omega b=a^\omega$, $a^\omega ab=0$, $b^\omega=b$, and $(a^\omega)^\omega=a^\omega$. Let $\varphi\colon A^{[1,\omega_1[} \to S$ be the morphism of ω_1-semigroups defined by $\varphi(a)=a$ and $\varphi(b)=b$. Then S recognizes $L=(aa+b)^{<\omega_1}-\lambda$ since $L=\varphi^{-1}(\{1, b, a^\omega\})$. Furthermore, S is the syntactic ω_1-semigroup of L.

The following theorem [Bed98b, BC98] establishes the link between finite ω_1-semigroups and automata. Its proof (see [Bed98b]) uses effective constructions to obtain an object of one of the two formalisms from the other.

THEOREM 1.6. *Let A be a finite alphabet and L a subset of $A^{<\omega_1}$. There exists an automaton $\mathscr{A}$ such that $\mathscr{L}(\mathscr{A})=L$ iff* synt(L) *is finite.*

EXAMPLE 1.9. Let $A=\{a, b\}$ and $L=(aa+b)^{<\omega_1}$. The language L is recognized by the automaton of Example 1.5 and its finite syntactic ω_1-semigroup is given in Example 1.8.

We now give a few definitions and propositions needed in the last sections of the paper.

[1] $\mathscr{H}$-classes with a star inside contain an idempotent, while others do not.

If X and Y are sets of words we note by $\overrightarrow{X \cdot Y}$ the set of words u that verify for every $0 < x < |u|$ there exist $x \leqslant y < |u|$ and $y < z < |u|$ such that $u[0, y[\in X$ and $u[y, z[\in Y$.

PROPOSITION 1.4. *Let A be an alphabet, S be a finite ω_1-semigroup, (s, e) a linked pair of S, and φ: $A^{[1, \omega_1[} \to S$ a morphism of ω_1-semigroups. Then*

$$\varphi^{-1}(s)\, \varphi^{-1}(e)^{\omega} \subseteq \overrightarrow{\varphi^{-1}(s) \cdot \varphi^{-1}(e)} \subseteq \bigcup_{f \in P_{s,e}} \varphi^{-1}(s)\, \varphi^{-1}(f)^{\omega},$$

where $P_{s,e} = \{f \in S : sf = s, ef = f, \text{ and } f^2 = f\}$.

Proof. The left inclusion is immediate. Let us turn to the other one. Assume $u \in \overrightarrow{\varphi^{-1}(s) \cdot \varphi^{-1}(e)}$. Now let $(x_j y_j)_{j<\omega}$ be an ω-sequence of prefixes of u such that $x_j \in \varphi^{-1}(s)$, $y_j \in \varphi^{-1}(e)$, $|x_j| > |x_{j-1} y_{j-1}|$ for every integer $j > 0$ and $(|x_i y_i|)_{i<\omega}$ is cofinal with $|u|$. Let $(z_j)_{j<\omega}$ be the ω-sequence of words such that $x_{j+1} = x_j z_j$ for any integer j. As a consequence of Theorem 1.3, $u = x_0 \prod_{i<\omega} z_i$ has a factorization $u = (x_0 z_0 \cdots z_{n_0-1}) \prod_{i<\omega} (z_{n_i} \cdots z_{n_{i+1}-1})$ such that $\varphi(x_0 z_0 \cdots z_{n_0-1}) = r$ and $\varphi(z_{n_j} \cdots z_{n_{j+1}-1}) = f$ for a linked pair (r, f) of S. Since $\varphi(x_0 z_0 \cdots z_{n_0-1}) = \varphi(x_j)$ for some j it follows that $r = s$. Since $\varphi(z_{n_0} \cdots z_{n_1-1}) = f$, y_{n_0} is a prefix of z_{n_0} and $\varphi(y_{n_0}) = e$ it follows that $f = eg$ for some $g \in S$, so $ef = eeg = eg = f$, which ends the proof of the right inclusion. ∎

COROLLARY 1.1. $\varphi^{-1}(e)^{\omega} = \overrightarrow{\varphi^{-1}(s) \cdot \varphi^{-1}(e)}$.

Proof. It suffices to use the previous proposition with $s = e$. Since $ef = e$ and $ef = f$ then $e = f$. ∎

We will use the following propositions on idempotents of finite ω_1-semigroups:

PROPOSITION 1.5. *Let S be a finite ω_1-semigroup and e, e' two idempotents of S. If $e \mathscr{D} e'$ then $e^{\omega} \mathscr{L} e'^{\omega}$.*

Proof. According to Proposition 1.1 there exist $x, y \in S$ such that $e = xy$ and $e' = yx$. So $e^{\omega} = (xy)^{\omega} = x(yx)^{\omega} = xe'^{\omega}$ and $e'^{\omega} = (yx)^{\omega} = y(xy)^{\omega} = ye^{\omega}$ which proves that $e^{\omega} \mathscr{L} e'^{\omega}$. ∎

PROPOSITION 1.6. *Let S be a finite ω_1-semigroup and e, e' two idempotents of S. If $e \mathscr{R} e'$ then $e^{\omega} = e'^{\omega}$.*

Proof. Since $e \mathscr{R} e'$ there exists $x \in S^1$ such that $e'x = e$. One can easily show that $e \mathscr{R} e' \mathscr{R} ee' \mathscr{R} e'e$. This implies in particular that $e' \mathscr{D} ee'$. Since $ee' \leqslant_{\mathscr{L}} e'$ and $e' \mathscr{D} ee'$ it follows from Proposition 1.2 that $ee' \mathscr{H} e'$; i.e., $e'xe' \mathscr{H} e'$. Let $h = e'xe'$. Since $\mathscr{H}(e')$ is a group of neutral element e' according to Proposition 1.3, there exists a positive integer n such that $h^n = e'$. Let now $y = h^{n-1} e'x$. We have $e'y = e'(e'xe')^{n-1}\, e'x = (e'x)^n = e^n = e$ and $ye' = (e'xe')^{n-1}\, e'xe' = (e'xe')^n = e'$. So $e^{\omega} = (e'y)^{\omega} = e'(ye')^{\omega} = e'e'^{\omega} = e'^{\omega}$. ∎

We say that an ω_1-semigroup S is *aperiodic* if S viewed as a simple semigroup is aperiodic, or, equivalently, if there is an integer n such that $s^n = s^{n+1}$ for any $s \in S$, i.e., S contains only tivial subgroups (S is group-free).

The following results are trivial adaptations of the analog on semigroups.

PROPOSITION 1.7. *Let $\sim_1$ and $\sim_2$ be two congruences on an ω_1-semigroup S. Then $\sim_1$ is a refinement of $\sim_2$ iff there exists a surjective morphism from $S/\sim_1$ into $S/\sim_2$.*

PROPOSITION 1.8. *Let A be an alphabet and X a recognizable subset of $A^{[1,\omega_1[}$. Then X is recognizable by an aperiodic ω_1-semigroup iff $A^{[1,\omega_1[}/\sim_X$ is aperiodic.*

PROPOSITION 1.9 (Cancellation proposition). *Let $p, q,$ and r be elements of an aperiodic ω_1-semigroup S. If $p = qpr$ then $p = qp = pr$.*

Proof. If S is aperiodic there exists an integer m such that $q^m = q^{m+1}$, so $p = qpr = q^m pr^m = q^{m+1} pr^m = qp$. The proof of $p = qpr \Rightarrow p = pr$ is similar. ▮

PROPOSITION 1.10. *Let p be an element of an aperiodic ω_1-semigroup S. Then $\mathscr{H}(p) = \{p\}$.*

Proof. By definition $\mathscr{H}(p) = \mathscr{R}(p) \cap \mathscr{L}(p)$. Let $x \in \mathscr{H}(p)$. There exist $a, b \in S^1$ such that $x = pa$ and $p = bx$. Using the cancellation proposition it follows that $x = pa = bxa = bx = p$. ▮

PROPOSITION 1.11. *Let x and y be two elements of an aperiodic ω_1-semigroup S. If $xy \mathscr{D} x$ (resp. $yx \mathscr{D} x$) then $xy \mathscr{R} x$ (resp. $yx \mathscr{L} x$).*

Proof. If $xy \mathscr{D} x$ there exist $a, b \in S^1$ such that $x = axyb$, and by the cancellation proposition $x = xyb$, so $x \leqslant_{\mathscr{R}} xy$. Since obviously $xy \leqslant_{\mathscr{R}} x$ then $xy \mathscr{R} x$. The proof that $yx \mathscr{D} x$ implies $yx \mathscr{L} x$ is similar. ▮

As an immediate corollary:

COROLLARY 1.2. *Let s be an element of an aperiodic ω_1-semigroup S. If $s \mathscr{D} s^\omega$ then $s \mathscr{R} s^\omega$.*

Proof. It suffices to replace y by x^ω in the statement of Proposition 1.11. The result holds because $xx^\omega = x^\omega$. ▮

PROPOSITION 1.12. *Let r, s, t be elements of an aperiodic ω_1-semigroup S. Then $rs \mathscr{D} st \mathscr{D} s$ iff $rst \mathscr{D} s$.*

Proof. We first prove that $rs \mathscr{D} st \mathscr{D} s$ implies $rst \mathscr{D} s$. If $rs \mathscr{D} st \mathscr{D} s$ there exist $a, b, c, d \in S^1$ such that $s = arsb$ and $s = cstd$, so by the cancellation proposition $s = ars$ and $s = std$, which implies $s = arstd$ and $s \leqslant_{\mathscr{D}} rst$. Since obviously $rst \leqslant_{\mathscr{D}} s$ we have $rst \mathscr{D} s$.

Let us show now that $rst \mathscr{D} s$ implies $rs \mathscr{D} st \mathscr{D} s$. Obviously $st \leqslant_{\mathscr{R}} s$. By hypothesis there exist $a, b \in S^1$ such that $s = arstb$, so $s = stb$ by the cancellation proposition, so $st \mathscr{R} s$. The proof that $rs \mathscr{L} s$ uses the same argument. ▮

1.3. Logic

We now define sets of words by sentences of formal logic, that is, by logical properties of words; this is based on the sequential calculus of Büchi.

1.3.1. Syntax

Let A be an alphabet. Our *first-order formulæ* are inductively built from a set of element variables usually noted by $x, y, z, x_1, y_1, z_1, \ldots$, a unary predicate R_a for each $a \in A$, a binary relation symbol $<$, an existential quantifier $\exists$ on variables, a binary logical connector $\vee$, and a unary one $\neg$:

- If x is a variable and $a \in A$, then $R_a(x)$ is a formula.
- If x and y are variables, then $x < y$ is a formula.
- If ϕ is a formula, then so is $\neg\phi$.
- If ϕ and ψ are formulæ, then so is $\phi \vee \psi$.
- If x is a variable and ϕ a formula, then $\exists x\phi$ is a formula.

We shall add parentheses for clarity. For convenience, we define the abbreviations $\forall x\phi$ for $\neg\exists x\neg\phi$, $\phi \to \psi$ for $\neg\phi \vee \psi$, $\phi \leftrightarrow \psi$ for $(\phi \to \psi) \wedge (\psi \to \phi)$, $\phi \wedge \psi$ for $\neg(\neg\phi \vee \neg\psi)$, $x = y$ for $(\neg(x < y)) \wedge (\neg(y < x))$, $x \leqslant y$ for $(x = y) \vee (x < y)$, $x \neq y$ for $\neg(x = y)$, $x = y + 1$ for $y < x \wedge \neg(\exists z\, z < x \wedge y < z)$, $\forall_z^y x\, \psi$ for $\forall x((z \leqslant x \wedge x < y) \to \psi)$, $\exists_z^y x\, \psi$ for $\exists x(z \leqslant x \wedge x < y \wedge \psi)$, and $Lim(x)$ for $(\neg\exists y\, x = y+1) \wedge (\exists y\, y < x)$. When one of those abbreviations introduces a new name of variable we suppose that this name is new in the formula in which the abbreviation is used.

If x and y are variables and a a letter, the formulæ $R_a(x)$ and $x < y$ are called *atomic formulæ*.

DEFINITION 1.7. Let ϕ be a first-order formula and x a first-order variable. The *quantifier height* of ϕ, noted by $hq(\phi)$, is inductively defined on the structure of ϕ:

- $hq(x < y) = hq(R_a(x)) = 0$
- $hq(\neg\phi) = hq(\phi)$
- $hq(\phi \vee \psi) = max(hq(\phi), hq(\psi))$
- $hq(\exists x\phi) = hq(\phi) + 1$.

For every formula ϕ we define by induction the set $FV(\phi)$ of *free variables* of ϕ:

- $FV(R_a(x)) = \{x\}$
- $FV(x < y) = \{x, y\}$
- $FV(\neg\phi) = FV(\phi)$.
- $FV(\phi \vee \psi) = FV(\phi) \cup FV(\psi)$
- $FV(\exists x\phi) = FV(\phi) - \{x\}$

An occurence of a variable x in a formula ϕ is *free* if there does not exist any subformula ψ of ϕ such that the occurrence is in ψ and $x \notin FV(\psi)$. A nonfree occurrence of a variable in a formula is said to be *bounded*. A *sentence* is a formula ϕ such that $FV(\phi) = \emptyset$. For simplicity, we assume that if x is a variable, $\exists x$ appears at most one time in a formula and that if ϕ is a formula and $x \in FV(\phi)$, then $\exists x \psi$ is not a subformula of ϕ. In other words, we assume that in formula all identical names of variables refer to the same variable.

Our *monadic second-order formulæ* (or second-order formulæ for short) are first-order formulæ in which variables of sets, also called (monadic) second-order variables, are allowed. We make a difference between second-order and first-order vatiables by noting the former using upper-cases letters and the latter with lower case letters. Formally, we build second-order formulæ by adding five items to the rules of construction of first-order formulæ:

- Any first-order formula is considered as a second-order formula,
- If x and X are respectively first- and second-order variables, then $X(x)$ is a (atomic) second-order formula.
- If X is a second-order variable and ϕ a second-order formula, then $\exists X \phi$ is a second-order formula.
- If ϕ and ψ are both second-order formulæ then so are $\phi \vee \psi$ and $\neg \psi$,
- If x is a first-order variable and ϕ a second-order formula then so is $\exists x \phi$.

Free variables are defined for second-order formulæ as for first-order formulæ. The set of free first-order variables of a formula ϕ will be noted by $FV1(\phi)$, the set of its free second-order variables by $FV2(\phi)$. The set of free variables of a second-order formula is the union of its free first-order variables and second-order variables ($FV(\phi) = FV1(\phi) \cup FV2(\phi)$).

1.3.2. Semantics

We now explain the meaning of formulæ. We define $\mathscr{L}(\phi)$, the set of words verifying properties described by the formula ϕ as [PP86] (see also [Str94]):

Definition 1.8. Let V, V' be respectively finite sets of first-order and second-order variables, A an alphabet, and α a countable ordinal. A *(V, V')-marked word* of length α over A is a word $\prod_{\beta<\alpha}(a_\beta, V_\beta, V'_\beta)$ over $A \times \mathscr{P}(V) \times \mathscr{P}(V')$ such that $V_\beta \cap V_\gamma = \emptyset$ if $\beta \neq \gamma$ and $\bigcup_{\beta<\alpha} V_\beta = V$.

Definition 1.9. Let ϕ be a formula, V, V' two finite sets such that $FV1(\phi) \subseteq V$ (resp. $FV2(\phi) \subseteq V'$) and there is no $x \in V$ (resp. $X \in V'$) that appears bounded in ϕ, and $w = \prod_{\beta<\alpha}(a_\beta, V_\beta, V'_\beta)$ a (V, V')-marked word of countable length α over an alphabet A. We say that w *satisfies* ϕ, and note $w \models \phi$, iff

- in case ϕ has the form $\neg \psi$, not $w \models \psi$,
- in case ϕ has the form $\psi \vee \chi$, $w \models \psi$, or $w \models \chi$,
- in case ϕ has the form $x < y$, $x \in V_\beta$, $y \in V_\gamma$, and $\beta < \gamma$ (the symbol $<$ represents the usual linear ordering on ordinals),

- in case ϕ has the form $R_a(x)$, $(a_\beta, V_\beta, V'_\beta)$ is a letter of w with $a_\beta = a$ and $x \in V_\beta$,
- in case ϕ has the form $X(x)$, $(a_\beta, V_\beta, V'_\beta)$ is a letter of w with $x \in V_\beta$ and $X \in V'_\beta$,
- in case ϕ has the form $\exists x\psi$, there exists $\beta < \alpha$ such that $(a_0, V_0, V'_0)\cdots(a_\beta, V_\beta \cup \{x\}, V'_\beta)\cdots$ is a $(V \cup \{x\}, V')$-marked word w' over A and $w' \models \psi$ for some $\beta < \alpha$,
- in case ϕ has the form $\exists X\psi$, there exists a set (eventually empty) B of ordinals smaller than α such that w in which each letter $(a_\beta, V_\beta, V'_\beta)$ of index $\beta \in B$ is replaced by $(a_\beta, V_\beta, V'_\beta \cup \{X\})$ satisfies ϕ.

If $w = \prod_{\beta<\alpha} a_\beta$ is a word over A and ϕ a sentence, then $w \models \phi$ iff $\prod_{\beta<\alpha} (a_\beta, \emptyset) \models \phi$.

Let ϕ be a sentence. We say that a word $w \in \mathscr{L}(\phi)$ iff $w \models \phi$.

Example 1.10. The sets of words of successor length containing an “a” letter is defined by the sentence:

$$\exists x R_a(x) \wedge \exists y\, \forall z (z \leqslant y).$$

Let ϕ and ψ be two first-order formulæ. We say that ϕ and ψ are (logically) equivalent, and write $\phi \equiv \psi$, if $\mathscr{L}(\phi) = \mathscr{L}(\psi)$. If α is an ordinal, A an alphabet, and ϕ a first-order formula then $\mathscr{L}^{<\alpha}(\phi)$ denotes $\mathscr{L}(\phi) \cap A^{<\alpha}$, $\mathscr{L}^{[1,\alpha[}(\phi)$ denotes $\mathscr{L}(\phi) \cap A^{[1,\alpha[}$, and $\mathscr{L}^{\alpha}(\phi)$ denotes $\mathscr{L}(\phi) \cap A^{\alpha}$.

This is a well-known result on formulæ:

Definition 1.10. A first-order formula ϕ is in *disjunctive normal form* if

- $hq(\phi) = 0$ and ϕ is

$$\bigvee_{i=1}^{m} \bigwedge_{j=1}^{p_i} \phi_{(i,j)},$$

where each $\phi_{(i,j)}$ is an atomic formula or a negation of atomic formula and there does not exist any repetition of a conjuct or a disjunct,

- $hq(\phi) = n+1$ and ϕ is

$$\bigvee_{i=1}^{m} \bigwedge_{j=1}^{p_i} \phi_{(i,j)},$$

where each $\phi_{(i,j)}$ is one of $\exists x\varphi$, $\neg\exists x\varphi$, φ with φ a first-order formula in disjunctive normal form, $hq(\varphi) \leqslant n$, and there does not exist any repetition of a conjuct or a disjunct.

Proposition 1.13. *Every first-order formula is logically equivalent to a first-order formula in disjunctive normal form of the same quantifier height.*

COROLLARY 1.3. *Let V be a finite set of variables and n an integer. There exist only a finite number of first-order formulæ ϕ such that $hq(\phi) \leqslant n$, modulo the logical equivalence, with variables in V.*

PROPOSITION 1.14. *For every first-order formula ϕ there exists a first-order formula*

$$Q_1 x_1 \cdots Q_n x_n \psi$$

which is logically equivalent to ϕ, where $Q_1 \cdots Q_n$ are $\exists$ or $\forall$, $x_1 \ldots x_n$ first-order variables and ψ a first-order formula without any quantifier.

1.4. Ehrenfeucht–Fraïssé Games

Ehrenfeucht–Fraïssé games are a play tool from model theory. We use first-order Ehrenfeucht–Fraïssé games for proving that two words satisfy exactly the same logic formulæ. Since they are not the subject of this paper, we only introduce here the material needed in the remainder of the paper. We refer to [Str94] for a very clear and more exhaustive presentation of the subject.

Let u, v be two $\{\,\}$-marked words and n an integer. The Ehrenfeucht–Fraïssé games are two players games. Let $\mathfrak{U}$ and $\mathfrak{V}$ note these two players. $\mathfrak{U}$ tries to prove that u and v do not satisfy the same atomic formulæ, while $\mathfrak{V}$ tries to displease his or her opponent. Each player has n pebbles, labeled $z_1, \ldots, z_n$. $\mathfrak{U}$ plays first: he or she chooses between u and v (say u for example) and places the pebble z_1 on a position of u, thus building a $\{z_1\}$-marked word. $\mathfrak{V}$ plays his or her pebble z_1 on the other marked word, and so on. The game ends when the two players have no more pebbles. $\mathfrak{U}$ has won the game if there exists an atomic formula with free variables in $\{z_1, \ldots, z_n\}$ that is satisfied by one of the two obtained $\{z_1, \ldots, z_n\}$-marked words but not the other, otherwise $\mathfrak{V}$ has won. We say that a player has a *winning strategy* if he or she wins the game, independently of what his or her opponent plays.

For a proof of the following well-known results on games on words, see [Ehr61, Lad77, Str94].

PROPOSITION 1.15. *Let n be an integer and u and v two $\{\,\}$-marked words. One of the two players has a winning strategy on the game on (u, v) with n pebbles.*

We write $u \sim_n v$ iff $\mathfrak{V}$ has a winning strategy on (u, v) using n pebbles, or $u \nsim_n v$ otherwise.

PROPOSITION 1.16. *$u \sim_n v$ iff u and v satisfy exactly the same first-order sentences of quantifier height at most n.*

Clearly, $\sim_n$ is an equivalence relation.

PROPOSITION 1.17. *Let n be an integer. Then $\sim_n$ has a finite number of equivalence classes.*

PROPOSITION 1.18. *Let x_1, x_2, y_1, and y_2 be $\{\,\}$-marked words and n an integer. If $x_1 \sim_n y_1$ and $x_2 \sim_n y_2$ then $x_1 x_2 \sim_n y_1 y_2$.*

Proof. The winning strategy of $\mathfrak{B}$ consists in partitioning the game in two parts: pebbles played on (x_1, y_1) and pebbles played on (x_2, y_2). He or she just applies his or her winning strategies on each of the two parts. To prove that this strategy suffices for $\mathfrak{B}$ to win the game, assume he or she loses; i.e., $x_1 x_2 \not\sim_n y_1 y_2$. An atomic formula is verified in one marked word (the marked-word build from $x_1 x_2$, for example) and not in the other. Assume first this atomic formula is $x < y$. If pebbles labeled x and y were both played in x_1 then the other pebbles labeled x and y were played in y_1, according to the strategy of $\mathfrak{B}$. Then $\mathfrak{U}$ has a winning strategy for the game (x_1, y_1) using n pebbles: it suffices to play exactly like he or she did in the game $(x_1 x_2, y_1 y_2)$ without playing the pebbles he or she played on x_2 or y_2. So $x_1 \not\sim_n y_1$, which is a contradiction. The rest of the proof uses similar arguments. ▮

This result can be easily generalized:

Proposition 1.19. *Let $(x_\beta)_{\beta<\alpha}$ and $(y_\beta)_{\beta<\alpha}$ be two sequences of $\{\}$-marked words and n an integer. If $x_\beta \sim_n y_\beta$ for every $\beta < \alpha$ then $\prod_{\beta<\alpha} x_\beta \sim_n \prod_{\beta<\alpha} y_\beta$.*

Proof. As in the previous position. ▮

The ordinal number α can be thought as a word of length α on an alphabet containing only one letter. The following is a well-known result of Ehrenfeucht–Fraïssé games on ordinals. For proofs, see for example [Ros82].

Proposition 1.20. *Let n be an integer. For every $k \geqslant 2^n - 1$, $k \sim_n k+1$.*

2. EQUIVALENCE BETWEEN MONADIC SECOND-ORDER FORMULAE AND AUTOMATA

In this section we give another proof of this well-known result of Büchi:

Theorem 2.1 (Büchi). *Let A be a finite alphabet. A subset L of $A^{<\omega_1}$ is recognizable by an automaton iff there exists a monadic second-order formula ϕ such that $L = \mathscr{L}(\phi)$.*

We emphasize that the constructions we will give in this proof are effective. These constructions will be used in the section on star-free sets. They are an adaptation of constructions from Straubing [Str94], who gave a proof of a similar theorem, but restricted to finite words.

Starting from an automaton $\mathscr{A} = (Q, A, E, I, F)$, we build a monadic second-order sentence ϕ such that $\mathscr{L}(\phi) = \mathscr{L}(\mathscr{A})$ by coding the states of $\mathscr{A}$ by second-order variables and using sentences meaning that a word u belongs to $\mathscr{L}(\mathscr{A})$ iff it is the label of a successful path in $\mathscr{A}$, i.e., iff there exist sets $(X_s)_{s \in [Q]_0^1}$ of ordinals less than $|u|$ such that (we assume first that $\lambda \notin \mathscr{L}(\mathscr{A})$)

1. $0 \in \bigcup_{i \in I} X_i$,
2. both of the following conditions are true:

   (i) if $|u| \in Succ$ and $|u| - 1 \in X_s$ there exists $f \in F$ such that $(s, u_{|u|-1}, f) \in E$,

   (ii) if $|u| \in Lim$ there exists $\{s_1, ..., s_p\} \in F$ such that the X_{s_i} are exactly those containing a sequence of ordinals cofinal with $|u|$,

3. $\alpha \in X_s$ for $s \in [Q]$ iff $\alpha \in Lim$,
4. if $\alpha \in X_s$ and $\alpha + 1 \in X_t$ then $(s, u_\alpha, t) \in E$,
5. $\alpha \in X_{\{s_1, \ldots, s_p\}}$ iff X_{s_i} are exactly those containing a sequence of ordinals cofinal with α,
6. each ordinal less than $|u|$ belongs to a X_s,
7. all the X_s are disjoints.

These properties are coded by the following second-order sentences:

1. $\psi_{1.} \equiv \forall x \neg \exists y \; y < x \to \bigvee_{i \in I} X_i(x)$,
2. $\psi_{2.} \equiv \psi_{(i)(i)} \wedge \psi_{(ii)(ii)}$ with

(i)

$$\psi_{(i)(i)} \equiv (\exists x(\forall y \; y \leqslant x)) \to \exists z \left((\forall y \; y \leqslant z) \bigwedge_{s \in [Q]_0^1} \left(X_s(z) \to \bigvee_{\substack{(s', a, f) \in E \\ s' = s \\ f \in F}} R_a(z) \right) \right),$$

(ii)

$$\psi_{(ii)(ii)} \equiv (\neg \exists x(\forall y \; y \leqslant x))$$
$$\to \bigvee_{\{s_1, \ldots, s_p\} \in F} \left(\left(\bigwedge_{s_i \in \{s_1, \ldots, s_p\}} \forall y \, \exists z \; y < z \wedge X_{s_i}(z) \right) \bigwedge_{t \in Q - \{s_1, \ldots, s_p\}} \neg \forall y \, \exists z \; y < z \wedge X_t(z) \right),$$

3.

$$\psi_{3.} \equiv \forall x((Lim(x) \wedge \exists y \; y < x) \leftrightarrow \bigvee_{s \in [Q]} X_s(x)),$$

4.

$$\psi_{4.} \equiv \forall x \, \forall y \left(y = x + 1 \to \bigwedge_{(s, t) \in [Q]_0^1 \times Q} (X_s(x) \wedge X_t(y) \to \bigvee_{\substack{(s', a, t') \in E \\ s' = s \\ t' = t}} R_a(x) \right),$$

5.

$$\psi_{5.} \equiv \bigwedge_{\{s_1, \ldots, s_p\} \in [Q]} \forall x (X_{\{s_1, \ldots, s_p\}}(x)$$
$$\leftrightarrow \left(\left(\bigwedge_{s_i \in \{s_1, \ldots, s_p\}} \forall_0^x y \; \exists_y^x z \; X_{s_i}(z) \right) \bigwedge_{t \in Q - \{s_1, \ldots, s_p\}} \neg \forall_0^x y \; \exists_y^x z \; X_t(z) \right) \right),$$

6.

$$\psi_{6.} \equiv \forall x \bigvee_{s \in [Q]_0^1} X_s(x),$$

7.

$$\psi_{7.} \equiv \bigwedge_{s \neq t,\, (s,\, t) \in ([Q]_0^1)^2} \neg \exists x (X_s(x) \wedge X_t(x)).$$

If $[Q]_0^1 = \{s_1, ..., s_p\}$ and $\lambda \notin \mathscr{L}(\mathscr{A})$ we have

$$\mathscr{L}(\mathscr{A}) = \mathscr{L}\left(\exists X_{s_1} \cdots \exists X_{s_p} \bigwedge_{1 \leqslant i \leqslant 7} \psi_i\right).$$

If $[Q]_0^1 = \{s_1, ..., s_p\}$ and $\lambda \in \mathscr{L}(\mathscr{A})$ we have

$$\mathscr{L}(\mathscr{A}) = \mathscr{L}\left((\forall x \neg x = x) \vee \left(\exists X_{s_1} \cdots \exists X_{s_p} \bigwedge_{1 \leqslant i \leqslant 7} \psi_i\right)\right).$$

We now prove the converse of the theorem. Let A be an alphabet and ϕ a second-order formula. We build an automaton $\mathscr{A}$ such that $\mathscr{L}(\mathscr{A}) = \mathscr{L}(\phi)$ by induction on the structure of ϕ. Let V_1 and V_2 be respectively the sets of first- and second-order variables appearing in ϕ, and let L be the set of (V_1, V_2)-marked words of countable length on A. If $a \in A$ and $x \in V_1$ an automaton accepting only the (V_1, V_2)-marked words on L with both x in the second component and a as the first component in a position can easily be built. In other words, this automaton recognizes only the (V_1, V_2)-marked words $u \in L$ such that $u \models R_A(x)$. Such an automaton can be built for each atomic formula. Since Theorem 1.2 shows that the class of languages accepted by automata is closed under the boolean operations, it just remains to prove that if ϕ looks like $\exists x \psi$ (or $\exists X \psi$), then we can build from an automaton recognizing $\mathscr{L}(\psi)$ an automaton $\mathscr{A}$ such that $\mathscr{L}(\mathscr{A}) = \mathscr{L}(\phi)$. By the induction hypothesis we can build an automaton $\mathscr{B} = \langle Q, A \times V'_1 \times V'_2, E, I, F\rangle$ such that $\mathscr{L}(\mathscr{B}) = \mathscr{L}(\psi)$. Let $\mathscr{A} = \langle Q', A \times (V'_1 - \{x\}) \times V'_2, E', I', F'\rangle$ the automaton defined by:

- $Q' = Q \times \{0, 1\}$,
-

$$E' = \bigcup \left\{ \begin{array}{l} \{(q, (a, W_1, W_2), p) \in [Q']_0^1 \times (A \times V'_1 \times V'_2) \times Q' : \\ \qquad (\mathscr{P}_1(q), (a, W_1, W_2), \mathscr{P}_1(p)) \in E \text{ and } \mathscr{P}_2(p) = \mathscr{P}_2(q) \text{ and } x \notin W_1\} \\ \{(q, (a, W_1 - \{x\}, W_2), p) \in [Q']_0^1 \times (A \times V'_1 \times V'_2) \times Q' : \\ \qquad (\mathscr{P}_1(q), (a, W_1, W_2), \mathscr{P}_1(p)) \in E \text{ and } \mathscr{P}_2(p) = 1, \mathscr{P}_2(q) = 0 \text{ and } x \in W_1\} \end{array} \right.$$

- $I' = \{(p, 0) \colon p \in I\}$,
- $F' = \{p \in [Q']_0^1 : \mathscr{P}_1(p) \in F \text{ and } \mathscr{P}_2(p) = 1\}$,

where $\mathscr{P}_1$ is the function from $[Q']_0^1$ to $[Q]_0^1$ defined by

- $\mathscr{P}_1(\{p_1, ..., p_k\}) = \{\mathscr{P}_1(p_1), ..., \mathscr{P}_1(p_k)\}$
- $\mathscr{P}_1((p, q)) = p$

and $\mathscr{P}_2$ is the function from $[Q']_0^1$ to $\{0, 1\}$ defined by

- $\mathscr{P}_2(\{p_1, ..., p_k\}) = 1$ iff there exists $i \in 1 \dots k$ such that $\mathscr{P}_2(p_i) = 1$
- $\mathscr{P}_2((p, q)) = q$.

The intuition is that the second component of the elements of Q' is a boolean marking the passage in $\mathscr{B}$ by a transition labeled by a letter in which x belongs to the second component. One can esaily verify that $\mathscr{L}(\mathscr{A}) = \mathscr{L}(\exists x\psi)$. The construction for $\exists X\psi$ uses similar arguments. This ends the proof of Theorem 2.1.

EXAMPLE 2.1. Let L be the language recognized by the automaton of Example 1.5 and ϕ be the following second-order sentence:

$$\begin{aligned}
&\exists X\, \forall x\, \forall y(R_b(x) \wedge y = x+1 \wedge R_a(y)) \rightarrow X(y)\\
&\quad \wedge \forall x(X(x) \rightarrow (R_a(x) \wedge \exists y(y = x+1 \wedge R_a(y) \wedge \neg X(y))))\\
&\quad \wedge \forall x((\neg X(x) \wedge R_a(x)) \rightarrow (\forall y(y = x+1 \wedge R_a(y)) \rightarrow X(y)))\\
&\quad \wedge \forall x((R_a(x) \wedge \neg \exists y\, x = y+1) \rightarrow X(x))\\
&\quad \wedge \exists x(R_a(x) \vee R_b(x)).
\end{aligned}$$

Then $\mathscr{L}(\phi) = L$. Each word of L can be factorized as a product of "aa" and "b." The idea of the sentence is that each position of an "a" is either the first or the second letter of a factor of such a factorization. The positions of the first "a" letter of the factors are memorized in a set X. The first line of the sentence declares the existence of X. The second line says that if a "b" in the word is followed by an "a" then this "a" is the first letter of a factor "aa," so the position just after "b" belongs to X. The third line signifies that each position in X is the position of an "a" and is followed by an "a" whose position is not in X. The fourth line says that if an "a" whose position is not in X is followed by another "a," then the position of the latter is in X. The fifth line signifies that if an "a" occurs at a nonsuccessor position, then this position must belong to X since it begins an occurrence of a factor "aa." Finally, the last line excludes the empty word from $\mathscr{L}(\phi)$.

EXAMPLE 2.2. Let L be the language recognized by the automaton of Example 1.6 and ϕ be the following second-order sentence:

$$\begin{aligned}
&\exists X\, \exists Y\, \forall x(\neg \exists y\, x = y+1) \rightarrow X(x)\\
&\quad \wedge \forall x_1\, X(x_1) \rightarrow (R_a(x_1) \wedge \exists y_1\, y_1 = x_1 + 1 \wedge Y(y_1))\\
&\quad \wedge \forall x_2\, Y(x_2) \rightarrow (R_b(x_2) \wedge \forall y_2\, y_2 = x_2 + 1 \rightarrow X(y_2))\\
&\quad \wedge \exists x(R_a(x) \vee R_b(x)).
\end{aligned}$$

Then $\mathscr{L}(\phi) = L$. The idea of the sentence is that positions of the "a" are kept in a set X, and positions of the "b" are kept in a set Y. The first line declares the existence of such sets and says that the positions in the word that are not successors belong to X. The second line signifies that every position in X is the position of an "a" and that every "a" is followed by a position in Y. The third line says that every position in Y is the position of a "b" and that if this is not the last position, the following position belongs to X. Finally, the last line excludes the empty word from $\mathscr{L}(\phi)$.

3. STAR-FREE SETS

This section is devoted to the citation of the different star-free theorems by increasing the lengths of words considered.

The first result was obtained on finite words:

DEFINITION 3.1. Let A be an alphabet. The class $SF(A, <\omega)$ of *star-free sets of finite words* on A is the smallest set containing all $\{a\}$ for $a \in A$ and closed under finite union, complement with respect to $A^{<\omega}$ and product.

THEOREM 3.1 [MP71, Sch65]. *Let A be an alphabet and X a recognizable subset of $A^{<\omega}$. The following conditions are equivalent:*

- $X \in SF(A, <\omega)$,
- $A^{<\omega}/\sim_X$ *is aperiodic,*
- $X = \mathscr{L}^{<\omega}(\phi)$ *for a first-order sentence* ϕ.

A similar result holds for sets of ω-words:

DEFINITION 3.2. Let A be an alphabet. The class $SF(A, \omega)$ of *star-free sets of ω-words* on A is the smallest containing $\varnothing$ closed under finite union, complement with respect to A^{ω}, and product on the left only by an element of $SF(A, <\omega)$.

THEOREM 3.2 [Lad77, Tho79, Per84]. *Let A be an alphabet and X a recognizable subset of A^{ω}. The following conditions are equivalent:*

- $X \in SF(A, \omega)$,
- $A^{[1, \omega^2[}/\sim_X$ *is aperiodic,*
- $X = \mathscr{L}^{\omega}(\phi)$ *for a first-order sentence* ϕ.

A similar result also holds for sets of words of length less than ω^{n+1}, where n is a fixed integer. It was obtained using algebras adapted to the study of such words, called ω^n-semigroups, which are a special case of ω_1-semigroups.

DEFINITION 3.3. Let A be an alphabet and n an integer. The class $SF(A, [1, \omega^{n+1}[)$ of *star-free sets of transfinite words* of length less than ω^{n+1} on A is the smallest set containing all $\{a\}$ for $a \in A$ and closed under finite union, complement with respect to $A^{[1, \omega^{n+1}[}$ and product.

THEOREM 3.3 [Bed98b, Bed]. *Let A be an alphabet, n an integer, and X a recognizable subset of $A^{[1,\omega^{n+1}[}$. The following conditions are equivalent:*

- $X \in SF(A, [1, \omega^{n+1}[)$,
- $A^{[1,\omega^{n+1}[}/\sim_X$ *is aperiodic,*
- $X = \mathscr{L}^{[1,\omega^{n+1}[}(\phi)$ *for a first-order sentence* ϕ.

And finally for sets of words of countable length:

DEFINITION 3.4. Let A be an alphabet. The class $SF(A, [1, \omega_1[)$ of *star-free sets of transfinite words* of countable length on A is the smallest set containing all $\{a\}$ for $a \in A$ and closed under finite union, complement with respect to $A^{[1,\omega_1[}$ and product.

THEOREM 3.4. *Let A be an alphabet and X a recognizable subset of $A^{[1,\omega_1[}$. The following conditions are equivalent:*

- $X \in SF(A, [1, \omega_1[)$,
- $A^{[1,\omega_1[}/\sim_X$ *is aperiodic,*
- $X = \mathscr{L}^{[1,\omega_1[}(\phi)$ *for a first-order sentence* ϕ.

Observe that the only difference between the definitions of star-free sets is the length of words in the complementation operation, except for Definition 3.2 which is a little bit more technical, because the words considered there have a fixed length ω.

The (effective) proof of the previous theorem occupies the remainder of this paper.

COROLLARY 3.1. *Let A be an alphabet. It is decidable whether a recognizable subset X of $A^{[1,\omega_1[}$ is star-free.*

COROLLARY 3.2. *Let A be an alphabet and ϕ a second-order formulæ. It is decidable if there exists a first-order formulæ ψ such that $\mathscr{L}(\psi) = \mathscr{L}(\phi)$. Furthermore, ψ can effectively be built from ϕ.*

4. FROM STAR-FREE SETS TO SENTENCES

Let $E \in SF(A, [1, \omega_1[)$ and $u = a_0 a_1 \cdots \in A^{[1,\omega_1[}$. We first prove that there exists a first-order formula ϕ_E which has exactly two free variables x and y such that

$$(a_0, \varnothing) \cdots (a_\alpha, \{x\}) \cdots (a_\beta, \{y\}) \cdots (\$, \varnothing) \vDash \phi_E \qquad \text{iff} \quad u[\alpha, \beta[\in E,$$

where \$ is a new letter which is not in A appearing only at the last position of the marked word (i.e., the index of $(\$, \varnothing)$ is $|u|$ in the left member of the equivalence above). The method is very similar to the one usually used for the finite word case. If r is a free variable of a formula ϕ the formula $\phi\{r \leftarrow s\}$ is ϕ in which the name r has been replaced by s.

If $E = \varnothing$ then $\phi_E \equiv (x = y) \wedge (x \neq y)$. If $E = \{a\}$ where $a \in A$ then $\phi_E \equiv y = x + 1 \wedge R_a(x)$. Assume now the existence of ϕ_L and ϕ_M for two star-free sets L and M. Then $\phi_{LM} \equiv \exists r(\phi_L\{y \leftarrow r\} \wedge (\phi_M\{x \leftarrow r\}))$ and $\phi_{L \cup M} \equiv \phi_L \vee \phi_M$. Let us turn finally to the complement operation. We have $\phi_{\neg E} \equiv x < y \wedge \neg \phi_E$.

Thus, we have inductively built ϕ_E from a star-free set E. It remains to get rid of the two free variables x and y. Let $\phi'_E \equiv \exists z[(\forall x\, z \leqslant x) \wedge (\phi_E\{x \leftarrow z\})]$, where z is a name that does not appear in ϕ_E. The only free variable of ϕ'_E is y. Let ϕ''_E be the sentence obtained from ϕ'_E substituting the subformulæ of the form $r < y$ by $r = r$, and $y < r$ or $y = r$ or $r = y$ (if r is not y) by $r \neq r$, where r is any variable of ϕ'_E.

It is not difficult to verify that if E is a star-free set then

$$u \in E \qquad \text{iff} \qquad u \vDash \phi''_E.$$

5. FROM SENTENCES TO FINITE APERIODIC ω_1-SEMIGROUPS

Let A be an alphabet and ϕ a first-order sentence. In this section we use games on words to prove that $\mathscr{L}(\phi)$ is recognizable by a finite aperiodic ω_1-semigroup.

Propositions 1.17 and 1.19 show that $A^{[1,\omega_1[}/\sim_{hq(\phi)}$ is a finite ω_1-semigroup, and Proposition 1.16 shows that $A^{[1,\omega_1[}/\sim_{hq(\phi)}$ recognizes $\mathscr{L}(\phi)$.

The proof that $A^{[1,\omega_1[}/\sim_{hq(\phi)}$ is aperiodic directly follows from this proposition:

PROPOSITION 5.1. *Let $n \in \mathbb{N}$ and $k = 2^n - 1$. For every word $y \in A^{[1,\omega_1[}$ then $y^{k+1} \sim_n y^k$.*

Proof. As an immediate corollary of Proposition 1.20 we have $a^{k+1} \sim_n a^k$ for $a \in A$. Let $y^{k+1} = y_1 y_2 \cdots y_{k+1}$ and $y^k = y'_1 y'_2 \cdots y'_k$ where $y_i = y'_i = y$ for every $1 \leqslant i \leqslant k$ and $y_{k+1} = y$. We consider that $\mathfrak{U}$ and $\mathfrak{B}$ play simultaneously two different games on n turns: the first one on a^{k+1} and a^k and the second one on y^{k+1} and y^k. $\mathfrak{U}$ plays first on the second game. If he or she plays in y^{k+1} (the other case is similar) on y_i at relative position α then he or she also plays on the first game on a^{k+1} at position i. $\mathfrak{B}$ applies his winning strategy in the first game: he or she plays on a^k at position j. His winning strategy in the second game is to play on y'_j at relative position α. ∎

Thus, $A^{[1,\omega_1[}/\sim_{\mathscr{L}(\phi)}$ is a finite aperiodic ω_1-semigroup.

We emphasize that the construction of $A^{[1,\omega_1[}/\sim_{\mathscr{L}(\phi)}$ is effective. Indeed, we showed in Section 2 that the construction of an automaton from a second-order sentence (in particular, from a first-order sentence) is effective. Furthermore, the proof of Theorem 1.6 uses effective constructions to prove the equivalence between automata and finite ω_1-semigroups. Because of its length and technical complexity the proof of this theorem is not given in this paper. We refer to the reader interested in the proof to [Bed98b].

6. FROM FIRST-ORDER SENTENCES TO STAR-FREE SETS

Let ϕ be a first-order sentence and A be an alphabet. In the previous section we showed that the set of words $u \in A^{[1,\omega_1[}$ such that $u \vDash \phi$ is a finite union of equivalence classes for $\sim_{hq(\phi)}$. We now prove that each such class is in $SF(A, [1, \omega_1[)$.

Since the star-free sets are closed under finite union, it follows that the set of words $u \in A^{[1,\omega_1[}$ such that $u \models \phi$ is in $SF(A, [1, \omega_1[)$.

If $x \in A^{[1,\omega_1[}$ we note by $\langle x \rangle_n$ the equivalence class of x for Ehrenfeucht–Fraïssé games in n turns. The statement of the following proposition is from Ladner (personal communication).

PROPOSITION 6.1. *Let m, n be two integers and x a word such that $0 < |x| < \omega_1$. Then*

$$\langle x \rangle_n = \left(\bigcap_{(u,a,v) \in P} \langle u \rangle_{n-1}\, a \langle v \rangle_{n-1} \right) \Big\backslash \left(\bigcup_{(u,a,v) \in Q} \langle u \rangle_{n-1}\, a \langle v \rangle_{n-1} \right),$$

where $P = \{(u, a, v) \in A^{<\omega_1} \times A \times A^{<\omega_1} \mid uav = x\}$ and $Q = \{(u, a, v) \in A^{<\omega_1} \times A \times A^{<\omega_1}$ such that for any factorization $x = u'a'v'$ then $u \nsim_{n-1} u'$ or $a \neq a'$ or $v \nsim_{n-1} v'\}$.

This lemma will be useful in the proof of the proposition.

LEMMA 6.1. *Let x and y be two words such that $x \nsim_n y$. If x_1, x_2, y_1, y_2 are four words and a and b two letters determined by the first turn of the game such that $x_1 a x_2 = x$ and $y_1 b y_2 = y$, either $x_1 \nsim_{n-1} y_1$ or $x_2 \nsim_{n-1} y_2$ or $a \neq b$.*

Proof. We note by x^i and y^i the index of letters of x and y played at turn i. In his or her winning strategy, $\mathfrak{A}$ plays his or her first pebble and $\mathfrak{B}$ answers, defining the factorizations of x and y of the statement of the lemma. If $\mathfrak{B}$ could not play on the same letter as $\mathfrak{A}$ in the other word, we have $a \neq b$. Assume $\mathfrak{B}$ could. Since $\mathfrak{A}$ wins, there exist two integers $i, j \leqslant n$ such that one of the two following conditions is true:

1. $R_c(x^i)$, $R_d(y^i)$ and $c \neq d$
2. $x^i < x^j$ and not $y^i < y^j$.

Since playing two times at the same position is not the advantage of $\mathfrak{A}$, and since $\mathfrak{B}$ can always do the same, we can assume that all his or her moves are different. Assume 1 is true and that $\mathfrak{A}$ has played at turn i on the left of the first move (the other case is similar). Since $\mathfrak{B}$ could not find the good letter at turn i on the left of the first move on the other word, and since pebbles played on the right of the first move are not useful for the winning strategy of $\mathfrak{A}$, $\mathfrak{A}$ has a winning strategy on $x[0, x^1[$ and $y[0, y^1[$ in $n-1$ turns. The second case is similar. ∎

We can now prove the proposition:

Proof. Let $y \in \langle x \rangle_n$. We start by proving that for any factorization $x = uav$ of w, where u and v are words and a a letter, there exist two words u' and v' such that $y = u'av'$ with $u' \sim_{n-1} u$ and $v' \sim_{n-1} v$. Assume that it is false, that is to say that for every u' and v' we have $u' \nsim_{n-1} u$ or $v' \nsim_{n-1} v$. It follows that $\mathfrak{A}$ has a winning strategy on the words x and y in n turns: he or she put his or her first pebble on a on x, and $\mathfrak{B}$ answers on y. If he cannot play on a letter a, he or she loses in only one turn. Otherwise, he factorizes y in $u'av'$, and since either $u' \nsim_{n-1} u$ or $v' \nsim_{n-1} v$ $\mathfrak{A}$ has just to apply his or her winning strategy in $n-1$ turns either on the left or on the right of the first turn. We now show that there does exist u, a, and v such that for any factorization $x = u'av'$ we have $y \in \langle u \rangle_{n-1}\, a \langle v \rangle_{n-1}$ and $u \not\sim_{n-1} u'$ or $v \not\sim_{n-1} v'$ or $a \neq a'$. Assume that such u, a, and v exist, and let $uav = z$. The winning strategy of $\mathfrak{U}$ consists in playing a on y, determinizing a factorization $y = u''av''$. $\mathfrak{B}$ answers in x determinizing a factorization $x = u'a'v'$. If $a' \neq a$, $\mathfrak{U}$ wins in only one turn. Otherwise, since $u'' \sim_{n-1} u \not\sim_{n-1} u'$ or $v'' \sim_{n-1} v \not\sim_{n-1} v'$, $\mathfrak{U}$ applies his or her winning strategy either on u'' and u' or on v'' and v'. We thus have obtained the contradiction $x \not\sim_n y$.

Now let y be a word of the right member of the equality of the statement of the proposition. We show that $\mathfrak{B}$ wins the game between x and y in n turns. Assume (for a contradiction) that $x \not\sim_n y$. $\mathfrak{U}$ plays his or her first pebble following his or her winning strategy; $\mathfrak{B}$ answers. If $\mathfrak{U}$ played on x, he or she chooses a factorization of $x = uav$ such that he or she wins for any factorization of $y = u'a'v'$ determined by the first play of $\mathfrak{B}$. If $a \neq a'$, $\mathfrak{U}$ wins in a single turn. Otherwise, according to the preceding lemma, either $u \not\sim_{n-1} u'$ or $v \not\sim_{n-1} v'$, that is to say, there does not exist a factorization $y = u'a'v'$ such that $u \sim_{n-1} u'$ and $v \sim_{n-1} v'$ and $a = a'$, which implies that y does not belong to the intersection of the right member of the equality, which is a contradiction. If $\mathfrak{U}$ played on y, he or she factorized it such that for any factorization $x = u'a'v'$ determined by the first pebble of $\mathfrak{B}$ we have either $a \neq a'$ or $u \not\sim_{n-1} u'$ or $v \not\sim_{n-1} v'$, and thus y belongs to the union of the right member of the equality, which contradicts the fact that y is in the right member of the equality. ∎

7. FROM FINITE APERIODIC ω_1-SEMIGROUPS TO STAR-FREE SETS

Let A be an alphabet and S a finite aperiodic ω_1-semigroup. In this section we prove that a language X recognized by a morphism $\varphi\colon A^{[1,\omega_1[} \to S$ of ω_1-semigroups is in $SF(A, [1, \omega_1[)$.

Let $P = \varphi(X) = \{p_1, \ldots, p_x\}$. Since $X = \varphi^{-1}(P) = \bigcup_{i=1\ldots x} \varphi^{-1}(p_i)$ and $SF(A, [1, \omega_1[)$ is closed under finite union it suffices to prove that $\varphi^{-1}(p_i) \in SF(A, [1, \omega_1[)$ for any $i \in 1\ldots x$, so we can assume that P contains only one element $p \in S$.

The proof is by induction on the structure in $\mathscr{D}$-classes of S: we will assume that $\varphi^{-1}(x) \in SF(A, [1, \omega_1[)$ for every $x \in S$ such that $p <_{\mathscr{D}} x$. This is the technique used in the original proof [Sch65] of Theorem 3.1, but the proof we present now is an adaptation of the proof of Theorem 3.1 from [Per90]. The advantage of the latter is that the semigroup does not change during the proof, which is not the case in the former.

Before starting the proof, which is long and technical, we give a (very) short informal description about it, in order to give some insight to the reader experienced with semigroup theory. The $\mathscr{D}$-classes of a finite ω_1-semigroup S can be preordered by $D \leqslant_{\mathscr{D}} D'$ iff $d \leqslant_{\mathscr{D}} d'$ for any $(d, d') \in D \times D'$. As S is aperiodic, any $\mathscr{H}$-class is a singleton, and thus $\varphi^{-1}(p) = \varphi^{-1}(\mathscr{L}(p)) \cap \varphi^{-1}(\mathscr{R}(p))$. In order to prove that $\varphi^{-1}(p)$ is star-free, it suffices to show that $\varphi^{-1}(\mathscr{R}(p))$ and $\varphi^{-1}(\mathscr{L}(p))$ are star-free. We show that $\varphi^{-1}(\mathscr{R}(p)) = \varphi^{-1}(\mathscr{R}(p))\, A^{<\omega_1} - \bigcup_{p \not\leqslant_{\mathscr{D}} r} \varphi^{-1}(r)$ (a similar equality holds for $\varphi^{-1}(\mathscr{L}(p))$) and that each element of the right side of the equality is star-free. The proof of the last fact is the long part of the proof and proceeds by induction on the $\leqslant_{\mathscr{D}}$ preorder. Here is the main idea, explained with finite words. Considering words indexed by countable ordinals uses the same arguments, but with many technical difficulties added because a nonempty factor of such a word may not have a last letter and the word cannot be read from the right to the left as it can be read from the left to the right: any nonlast letter has a successor letter, but there exists nonfirst letters without predecessor letters: the letters at limit positions. The proof that $\varphi^{-1}(\mathscr{R}(p))\, A^{<\omega_1}$ for finite words is star-free is by induction on the structure in $\mathscr{D}$-classes of S. The induction hypothesis is "$\varphi^{-1}(x)$ is star-free if $p <_{\mathscr{D}} x$." Now $\varphi^{-1}(\mathscr{R}(p))\, A^{<\omega_1} = (\bigcup_{\substack{s \in S^1 \\ a \in A \\ s\varphi(a)\, \mathscr{R}\, p \\ s \not\mathscr{R}\, p}} \varphi^{-1}(s)\, aA^{<\omega_1})$. If $sx \,\mathscr{R}\, p$ but $s \not\mathscr{R}\, p$ then $p <_{\mathscr{D}} s$. We then obtain that $\varphi^{-1}(\mathscr{R}(p))\, A^{<\omega_1}$ is star-free using the induction hypothesis. The remainder of the proof uses similar arguments.

We now start the proof.

If S does not possess a neutral element we add it: since $1^2 = 1$ this does not change the aperiodicity of S nor $\varphi^{-1}(s)$ for every $s \in S$. Thus, $1x = x1 = 1 = 1^{\omega}$ for every x of S.

We start by showing that

$$\varphi^{-1}(p) = \left(\varphi^{-1}(\mathscr{R}(p))\, A^{\langle\omega_1} \cap A^{<\omega_1}\varphi^{-1}(\mathscr{L}(p))\right) - \bigcup_{p \not\leqslant_{\mathscr{D}} r} \varphi^{-1}(r). \qquad (1)$$

First, if $u \in \varphi^{-1}(p)$, then obviously u has a prefix v such that $\varphi(v) \in \mathscr{R}(p)$ and a suffix w such that $\varphi(w) \in \mathscr{L}(p)$. Furthermore, u cannot be in $\bigcup_{p \not\leqslant_{\mathscr{D}} r} \varphi^{-1}(r)$.

Now let u be in the right member of the equality. Then u has a prefix v and a suffix w such that $\varphi(v) \in \mathscr{R}(p)$, $\varphi(w) \in \mathscr{L}(p)$, and of course $\varphi(u) \leqslant_{\mathscr{R}} \varphi(v)$ and $\varphi(u) \leqslant_{\mathscr{L}} \varphi(w)$. Since $u \notin \bigcup_{p \not\leqslant_{\mathscr{D}} r} \varphi^{-1}(r)$ then $p \leqslant_{\mathscr{D}} \varphi(u)$. Since $\varphi(u) \leqslant_{\mathscr{R}} \varphi(v)$ (resp. $\varphi(u) \leqslant_{\mathscr{L}} \varphi(w)$) implies $\varphi(u) \leqslant_{\mathscr{D}} \varphi(v)$ (resp. $\varphi(u) \leqslant_{\mathscr{D}} \varphi(w)$) and $\varphi(v) \in \mathscr{R}(p)$ (resp. $\varphi(w) \in \mathscr{L}(p)$) implies $\varphi(v) \in \mathscr{D}(p)$ (resp. $\varphi(w) \in \mathscr{D}(p)$) then $\varphi(u)\, \mathscr{D}\, p\, \mathscr{D}\, \varphi(v)\, \mathscr{D}\, \varphi(w)$. Using Proposition 1.11 it follows that $\varphi(u) \in \mathscr{R}(p) \cap \mathscr{L}(p)$, so $\varphi(u) = p$ by usage of Proposition 1.10.

It remains to show that each component of the right member of equality 1 belongs to $SF(A, [1, \omega_1[)$. It will prove that $\varphi^{-1}(p) \in SF(A, [1, \omega_1[)$ because $SF(A, [1, \omega_1[)$ is closed under the operations used in the right member of equality 1.

The following lemmas will be needed:

LEMMA 7.1. *Let A be an alphabet, n an integer, and $(L_i)_{i<n}$ a finite family of sets such that $L_i \in SF(A, [1, \omega_1[)$ for every $i < n$. Then $(\bigcup_{i<n} L_i A^{<\omega_1})^{\omega} \in SF(A, [1, \omega_1[)$.*

Proof. If every $L_i \in SF(A, [1, \omega_1[)$, according to Section 4 there exists a first-order sentence ϕ^i such that $L_i = \mathscr{L}(\phi^i)$. A first-order formula $\phi^i_{x,y}$ with exactly two free variables x and y (two names that do not already appear in ϕ^i) and such that

$$(a_0, \varnothing) \cdots (a_\alpha, \{x\}) \cdots (a_\beta, \{y\}) \cdots \models \phi^i_{x,y} \qquad \text{iff} \quad \prod_{\alpha \leqslant \gamma < \beta} a_\gamma \in L_i$$

can easily be built from ϕ^i. Now let ϕ be the first-order sentence defined by

$$\phi \equiv \exists x' \, \exists z (\forall y \; z \leqslant y) \wedge \left(\bigvee_{i<n} \phi^i_{x,y}\{x \leftarrow z, y \leftarrow x'\} \right)$$
$$\wedge \forall x' \, \exists y' \, \exists z' \; x' < y' \wedge y' < z' \wedge \left(\bigvee_{i<n} \phi^i_{x,y}\{x \leftarrow y', y \leftarrow z'\} \right).$$

Informally speaking, the first line says that a model of ϕ has a proper prefix in $\bigcup_{i<n} L_i$ and the second line that after any position there exists a proper factor of the model, which is not a suffix, and that belongs to $\bigcup_{i<n} L_i$. We let the reader check that $\mathscr{L}(\phi) = (\bigcup_{i<n} L_i A^{<\omega_1})^\omega$. According to Section 6, $(\bigcup_{i<n} L_i A^{<\omega_1})^\omega \in SF(A, [1, \omega_1[)$. ▮

Lemma 7.2. *If $X, Y \in SF(A, [1, \omega_1[)$ then $\overrightarrow{X \cdot Y} \in SF(A, [1, \omega_1[)$.*

Proof. If $X, Y \in SF(A, [1, \omega_1[)$ there exist two first-order sentences ϕ^X and ϕ^Y such that $\mathscr{L}(\phi^X) = X$ and $\mathscr{L}(\phi^Y) = Y$. Again, a first-order formula $\phi^X_{x,y}$ with exactly two free variables x and y (two names that do not already appear in ϕ^X) and such that

$$(a_0, \varnothing) \cdots (a_\alpha, \{x\}) \cdots (a_\beta, \{y\}) \cdots \models \phi^X_{x,y} \text{ iff } \prod_{\alpha \leqslant \gamma < \beta} a_\gamma \in X$$

can easily be built from ϕ^X. The same holds for ϕ^Y. Now let $\phi^{\overrightarrow{X \cdot Y}}$ be the first-order sentence defined as

$$\phi^{\overrightarrow{X \cdot Y}} \equiv \exists z ((\forall x \; z \leqslant x) \wedge (\forall x' \; z < x' \rightarrow \exists y' \, \exists z' \; x' \leqslant y' \wedge y' < z'$$
$$\wedge \phi^X_{x,y}\{x \leftarrow z, y \leftarrow y'\} \wedge \phi^Y_{x,y}\{x \leftarrow y', y \leftarrow z'\})) .$$

Then $\mathscr{L}(\phi^{\overrightarrow{X \cdot Y}}) = \overrightarrow{X \cdot Y}$. According to Section 6, $\overrightarrow{X \cdot Y} \in SF(A[1, \omega_1[)$. ▮

As a consequence of Lemma 7.2 and Corollary 1.1:

Lemma 7.3. *Let e be an idempotent of a finite ω_1-semigroup such that $\varphi^{-1}(e) \in SF(A, [1, \omega_1[)$. Then $\varphi^{-1}(e)^\omega \in SF(A, [1, \omega_1[)$.*

Lemma 7.4. *Let S be a finite ω_1-semigroup and p be an element of S such that $\varphi^{-1}(s) \in SF(A, [1, \omega_1[)$ for any $p <_{\mathscr{D}} s$. Then*

$$\left(\bigcup_{\substack{s \in S^1 \\ a \in A \\ s\varphi(a) \,\mathscr{R}\, p \\ s \,\not\mathscr{R}\, p}} \varphi^{-1}(s) \, aA^{<\omega_1} \right) \cup \left(\bigcup_{\substack{(s', e') \in S \times E(S) \\ s'e'^\omega \,\mathscr{R}\, p \\ s' \,\not\mathscr{R}\, p \\ s'e' \,\not\mathscr{R}\, p}} \varphi^{-1}(s') \, \varphi^{-1}(e')^\omega A^{<\omega_1} \right) \in SF(A, [1, \omega_1[).$$

Proof. Since S is finite the number of unions involved is finite. So it suffices to prove that $\varphi^{-1}(s) \in SF(A[1, \omega_1[)$ for each s in the first union and that $\varphi^{-1}(s') \, \varphi^{-1}(e')^\omega \in SF(A, [1, \omega_1[)$ for each pair (s', e') of the second union. If $p \not<_{\mathscr{D}} s$ then $s\varphi(a) \,\mathscr{R}\, p$ is impossible. If $s \,\mathscr{D}\, p$ then $s \,\mathscr{R}\, p$ because $s\varphi(a) \,\mathscr{R}\, p$ and by

Proposition 1.2. So $p <_{\mathscr{D}} s$, which proves that $\varphi^{-1}(s) \in SF(A[1, \omega_1[)$. The proof that $\varphi^{-1}(s') \in SF(A, [1, \omega_1[)$ is identical. Now if $p \nless_{\mathscr{D}} e'$ then $s'e'^{\omega} \mathscr{R}\, p$ is impossible. If $e' \mathscr{D}\, p$ then $s'e' \mathscr{D}\, p$ because $s'e'^{\omega} \mathscr{D}\, p$. Because $s'e'^{\omega} \mathscr{R}\, p$ and by Proposition 1.2 it follows that $s'e' \mathscr{R}\, p$, which is a contradiction. So $p <_{\mathscr{D}} e'$, which proves $\varphi^{-1}(e') \in SF(A, [1, \omega_1[)$. Finally, Lemma 7.3 proves that $\varphi^{-1}(e')^{\omega} \in SF(A, [1, \omega_1[)$. ∎

LEMMA 7.5. *Let e be an idempotent of a finite ω_1-semigroup S such that $\varphi^{-1}(x) \in SF(A, [1, \omega_1[)$ for every $e <_{\mathscr{D}} x$. Then*

$$\Bigg(\Bigg(\bigcup_{\substack{s \in S^1 \\ a \in A \\ s\varphi(a)\, \mathscr{R}\, e \\ s \not\mathscr{R}\, e}} \varphi^{-1}(s)\, aA^{<\omega_1}\Bigg) \cup \Bigg(\bigcup_{\substack{(s', e') \in S \times E(S) \\ s'e'^{\omega}\, \mathscr{R}\, e \\ s' \not\mathscr{R}\, e \\ s'e' \not\mathscr{R}\, e}} \varphi^{-1}(s')\, \varphi^{-1}(e')^{\omega}\, A^{<\omega_1}\Bigg)\Bigg)^{\omega} \in SF(A[1, \omega_1[).$$

Proof. Recall that S is finite, so the number of unions involved is finite. By Lemma 7.1 it suffices to prove that $\varphi^{-1}(s) \in SF(A, [1, \omega_1[)$ for each s in the first union and that $\varphi^{-1}(s')\, \varphi^{-1}(e')^{\omega} \in SF(A, [1, \omega_1[)$ for each pair (s', e') of the second union. The proof exactly follows the proof of Lemma 7.4. ∎

We now prove that

LEMMA 7.6. *If S is a finite aperiodic ω_1-semigroup and p an element of S such that $\varphi^{-1}(s) \in SF(A, [1, \omega_1[)$ for every $s \in S$ such that $p <_{\mathscr{D}} s$, then $\bigcup_{p \nleq_{\mathscr{D}} r} \varphi^{-1}(r) \in SF(A, [1, \omega_1[)$.*

Proof. In order to do it we first show that

$$\begin{aligned}\bigcup_{p \nleq_{\mathscr{D}} r} \varphi^{-1}(r) = \bigcup_{\substack{a \in A \\ p \nleq_{\mathscr{D}} \varphi(a)}} & A^{<\omega_1} a A^{<\omega_1} \bigcup_{\substack{e \in E(S) \\ p \leqslant_{\mathscr{D}} e \\ p \nleq_{\mathscr{D}} e^{\omega}}} A^{\omega_1} \varphi^{-1}(e)^{\omega}\, A^{<\omega_1} \\ & \bigcup_{\substack{a, b \in A \\ s \in S^1 \\ p \leqslant_{\mathscr{D}} \varphi(a)\, s \\ p \leqslant_{\mathscr{D}} s\varphi(b) \\ p \nleq_{\mathscr{D}} \varphi(a)\, s\varphi(b)}} A^{<\omega_1} a \varphi^{-1}(s)\, bA^{<\omega_1} \bigcup_{\substack{a \in A \\ (s, e) \in S^1 \times E(S) \\ p \leqslant_{\mathscr{D}} e^{\omega} s \\ p \leqslant_{\mathscr{D}} s\varphi(a) \\ p \nleq_{\mathscr{D}} e^{\omega} s\varphi(a)}} A^{<\omega_1} \varphi^{-1}(e)^{\omega}\, \varphi^{-1}(s)\, aA^{<\omega_1} \\ & \bigcup_{\substack{(s, e_1, e_2) \in S^1 \times E(S) \times E(S) \\ p \leqslant_{\mathscr{D}} e_1^{\omega} s \\ p \leqslant_{\mathscr{D}} s e_2^{\omega} \\ p \nleq_{\mathscr{D}} e_1^{\omega} s e_2^{\omega}}} A^{<\omega_1} \varphi^{-1}(e_1)^{\omega}\, \varphi^{-1}(s)\, \varphi^{-1}(e_2)^{\omega} A^{<\omega_1}.\end{aligned} \tag{2}$$

Clearly a word in the right member of the equality belongs to the left member too. Assume now that u is a word of the left member. If u has a letter a such that $p \nleq_{\mathscr{D}} \varphi(a)$ then u belongs to the first union of the right member of the equality. Assume u does not have such a letter, and let w be a smallest factor of u such that $p \nleq_{\mathscr{D}} \varphi(w)$. Write $|w|$ is normal form $|w| = \omega^{\alpha_0} + \cdots \omega^{\alpha_n}$, where n is an integer. Assume first there is only one term in this sum; i.e., $n = 0$. Then $\alpha_0 > 0$ since u does not have a letter a such that $p \nleq_{\mathscr{D}} \varphi(a)$. So u belongs to the second union of the right member of the equality. Suppose now that $n > 0$. We have two cases to examine: either $|w| \in Lim$ or $|w| \in Succ$. Assume first that $|w| \in Lim$. Then w has a factorization $w = w_1 w_2 w_3$ with $|w_1| = \omega^{\alpha_0}$ and $|w_3| = \omega^{\alpha_n}$. Both of these lengths are limit ordinals, so w belongs to the last union of the right member of the equality. Assume now that $|w| \in Succ$. Then w has a last letter and if $\alpha_0 \in Lim$, w belongs to the third union of the right member of the equality or to the fourth union of the right member of the equality otherwise. Thus the equality is proved; it remains to show that each union of its right member belongs to $SF(A, [1, \omega_1[)$. It is trivial for the first one. Let us see the second one. We obviously have

$$\bigcup_{\substack{e \in E(S) \\ p \leqslant_{\mathscr{D}} e \\ p \nleqslant_{\mathscr{D}} e^\omega}} A^{<\omega_1} \varphi^{-1}(e)^\omega A^{<\omega_1} = \bigcup_{\substack{e \in E(S) \\ p <_{\mathscr{D}} e \\ p \nleqslant_{\mathscr{D}} e^\omega}} A^{<\omega_1} \varphi^{-1}(e)^\omega \bigcup_{\substack{e \in E(S) \\ p \mathscr{D} e \\ p \nleqslant_{\mathscr{D}} e^\omega}} A^{<\omega_1} \varphi^{-1}(e)^\omega A^{<\omega_1}.$$

So we have to prove that the right member of the equality is in $SF(A, [1, \omega_1[)$. The first union belongs to $SF(A, [1, \omega_1[)$ because of the hypothesis and Lemma 7.3. Furthermore, it can easily be checked that

$$\bigcup_{\substack{e \in E(S) \\ p \mathscr{D} e \\ p \nleqslant_{\mathscr{D}} e^\omega}} A^{<\omega_1} \varphi^{-1}(e)^\omega A^{<\omega_1}$$

$$\subseteq \bigcup_{\substack{e \in E(S) \\ p \mathscr{D} e \\ p \nleqslant_{\mathscr{D}} e^\omega}} A^{<\omega_1} \Bigg(\bigcup_{\substack{s' \in S^1 \\ a \in A \\ s'\varphi(a) \mathscr{R} e \\ s' \not\mathscr{R} e}} \varphi^{-1}(s')\, a A^{<\omega_1} \bigcup_{\substack{(s', e') \in S \times E(S) \\ s'e'^\omega \mathscr{R} e \\ s' \not\mathscr{R} e \\ s'e' \not\mathscr{R} e}} \varphi^{-1}(s')\, \varphi^{-1}(e')^\omega A^{<\omega_1} \Bigg)^\omega A^{<\omega_1} \tag{3}$$

and that the right member of the inclusion is in $SF(A, [1, \omega_1[)$ because of Lemma 7.5. Furthermore, the right member of the inclusion it itself included in $\bigcup_{p \nleqslant_{\mathscr{D}} r} \varphi^{-1}(r)$: if u is a word that belongs to the right member of the inclusion then it has a factor w such that w has a factorization $\prod_{i<\omega} v_i$ verifying either $\varphi(v_i) \mathscr{R} e$ or $\varphi(v_i) <_{\mathscr{D}} e$ for each integer i (by usage of Proposition 1.11). If $\varphi(v_i) <_{\mathscr{D}} e$ for some i then $\varphi(u) <_{\mathscr{D}} e$, so $p \nleqslant_{\mathscr{D}} \varphi(u)$. If this case never happens then according to Theorem 1.3 we can suppose the existence of a linked pair (s', e') such that $\varphi(v_0) = s'$ and $\varphi(v_i) = e'$ for any positive integer i. So $\varphi(w) = s'e'^\omega$. Since $e' \mathscr{R} e$, by Proposition 1.6 we have $e'^\omega = e^\omega$, so $p \nleqslant_{\mathscr{D}} e'^\omega$ because $p \nleqslant_{\mathscr{D}} e^\omega$, and $p \nleqslant_{\mathscr{D}} \varphi(u)$. The ideas for the two last unions of the right member of equality (2) are the same. For the fourth union we have

$$\bigcup_{\substack{a \in A \\ (s, e) \in S^1 \times E(S) \\ p \leqslant_{\mathscr{D}} e^\omega s \\ p \leqslant_{\mathscr{D}} s\varphi(a) \\ p \nleqslant_{\mathscr{D}} e^\omega s\varphi(a)}} A^{<\omega_1} \varphi^{-1}(e)^\omega \varphi^{-1}(s)\, a A^{<\omega_1}$$

$$= \bigcup_{\substack{a \in A \\ (s, e) \in S^1 \times E(S) \\ p <_{\mathscr{D}} e \\ p \leqslant_{\mathscr{D}} e^\omega s \\ p \leqslant_{\mathscr{D}} s\varphi(a) \\ p \nleqslant_{\mathscr{D}} e^\omega s\varphi(a)}} A^{<\omega_1} \varphi^{-1}(e)^\omega \varphi^{-1}(s)\, a A^{<\omega_1} \bigcup_{\substack{a \in A \\ (s, e) \in S^1 \times E(S) \\ p \mathscr{D} e \\ p \leqslant_{\mathscr{D}} e^\omega s \\ p \leqslant_{\mathscr{D}} s\varphi(a) \\ p \nleqslant_{\mathscr{D}} e^\omega s\varphi(a)}} A^{<\omega_1} \varphi^{-1}(e)^\omega \varphi^{-1}(s)\, a A^{<\omega_1}$$

and

$$\bigcup_{\substack{a \in A \\ (s,e) \in S^1 \times E(S) \\ p \mathscr{D} e \\ p \leqslant_{\mathscr{D}} e^{\omega} s \\ p \leqslant_{\mathscr{D}} s\varphi(a) \\ p \not\leqslant_{\mathscr{D}} e^{\omega} s\varphi(a)}} A^{<\omega_1} \varphi^{-1}(e)^{\omega} \varphi^{-1}(s)\, aA^{<\omega_1}$$

$$\subseteq \bigcup_{\substack{a \in A \\ (s,e) \in S^1 \times E(S) \\ p \mathscr{D} e \\ p \leqslant_{\mathscr{D}} e^{\omega} s \\ p \leqslant_{\mathscr{D}} s\varphi(a) \\ p \not\leqslant_{\mathscr{D}} e^{\omega} s\varphi(a)}} A^{<\omega_1} \Bigg(\bigcup_{\substack{s' \in S^1 \\ b \in A \\ s'\varphi(b) \mathscr{R} e \\ s' \not\mathscr{R} e}} \varphi^{-1}(s')\, bA^{<\omega_1} \bigcup_{\substack{(s',e') \in S \times E(S) \\ s'e'^{\omega} \mathscr{R} e \\ s' \not\mathscr{R} e \\ s'e' \not\mathscr{R} e}} \varphi^{-1}(s')\, \varphi^{-1}(e')^{\omega} A^{<\omega_1} \Bigg)^{\omega}$$

$$\times \varphi^{-1}(s)\, aA^{\omega_1} \tag{4}$$

and for the last one (for short we write $ABBREV(A, e_1, s, e_2)$ for $A^{<\omega_1}\varphi^{-1}(e_1)^{\omega}$ $\varphi^{-1}(s)\, \varphi^{-1}(e_2)^{\omega} A^{<\omega_1}$)

$$\bigcup_{\substack{s \in S^1 \\ e_1 \in E(S) \\ e_2 \in E(S) \\ p \leqslant_{\mathscr{D}} e_1^{\omega} s \\ p \leqslant_{\mathscr{D}} s e_2^{\omega} \\ p \not\leqslant_{\mathscr{D}} e_1^{\omega} s e_2^{\omega}}} ABBREV(A, e_1, s, e_2)$$

$$= \bigcup_{\substack{s \in S^1 \\ e_1 \in E(S) \\ e_2 \in E(S) \\ p <_{\mathscr{D}} e_1 \\ p <_{\mathscr{D}} e_2 \\ p \leqslant_{\mathscr{D}} e_1^{\omega} s \\ p \leqslant_{\mathscr{D}} s e_2^{\omega} \\ p \not\leqslant_{\mathscr{D}} e_1^{\omega} s e_2^{\omega}}} ABBREV(A, e_1, s, e_2)$$

$$\bigcup_{\substack{s \in S^1 \\ e_1 \in E(S) \\ e_2 \in E(S) \\ p \mathscr{D} e_1 \\ p <_{\mathscr{D}} e_2 \\ p \leqslant_{\mathscr{D}} e_1^{\omega} s \\ p \leqslant_{\mathscr{D}} s e_2^{\omega} \\ p \not\leqslant_{\mathscr{D}} e_1^{\omega} s e_2^{\omega}}} ABBREV(A, e_1, s, e_2) \bigcup_{\substack{s \in S^1 \\ e_1 \in E(S) \\ e_2 \in E(S) \\ p <_{\mathscr{D}} e_1 \\ p \mathscr{D} e_2 \\ p \leqslant_{\mathscr{D}} e_1^{\omega} s \\ p \leqslant_{\mathscr{D}} s e_2^{\omega} \\ p \not\leqslant_{\mathscr{D}} e_1^{\omega} s e_2^{\omega}}} ABBREV(A, e_1, s, e_2)$$

$$\bigcup_{\substack{(s, e_1, e_2) \in S^1 \times E(S) \times E(S) \\ p \mathscr{D} e_1 \\ p \mathscr{D} e_2 \\ p \leqslant_{\mathscr{D}} e_1^{\omega} s \\ p \leqslant_{\mathscr{D}} s e_2^{\omega} \\ p \not\leqslant_{\mathscr{D}} e_1^{\omega} s e_2^{\omega}}} ABBREV(A, e_1, s, e_2)$$

and

$$\bigcup_{\substack{(s,e_1,e_2)\in S^1\times E(S)\times E(S)\\ p\,\mathscr{D}\,e_1\\ p\,\mathscr{D}\,e_2\\ p\leqslant_{\mathscr{D}} e_1^\omega s\\ p\leqslant_{\mathscr{D}} se_2^\omega\\ p\nleqslant_{\mathscr{D}} e_1^\omega se_2^\omega}} A^{<\omega_1}\varphi^{-1}(e_1)^\omega\,\varphi^{-1}(s)\,\varphi^{-1}(e_2)^\omega\,A^{<\omega_1}$$

$$\subseteq \bigcup_{\substack{s\in S^1\\ e_1\in E(S)\\ e_2\in E(S)\\ p\,\mathscr{D}\,e_1\\ p\,\mathscr{D}\,e_2\\ p\leqslant_{\mathscr{D}} e_1^\omega s\\ p\leqslant_{\mathscr{D}} se_2^\omega\\ p\nleqslant_{\mathscr{D}} e_1^\omega se_2^\omega}} A^{<\omega_1}\Bigg(\bigcup_{\substack{s'\in S^1\\ a\in A\\ s'\varphi(a)\,\mathscr{R}\,e_1\\ s'\,\not\mathscr{R}\,e_1}}\varphi^{-1}(s')\,aA^{<\omega_1}\bigcup_{\substack{s'\in S\\ e'\in E(S)\\ s'e'^\omega\,\mathscr{R}\,e_1\\ s'\,\not\mathscr{R}\,e_1\\ s'e'\,\not\mathscr{R}\,e_1}}\varphi^{-1}(s')\,\varphi^{-1}(e')^\omega\,A^{<\omega_1}\Bigg)^\omega\varphi^{-1}(s)$$

$$\times\Bigg(\bigcup_{\substack{s'\in S^1\\ a\in A\\ s'\varphi(a)\,\mathscr{R}\,e_2\\ s'\,\not\mathscr{R}\,e_2}}\varphi^{-1}(s')\,aA^{<\omega_1}\bigcup_{\substack{(s',e')\in S\times E(S)\\ s'e'^\omega\,\mathscr{R}\,e_2\\ s'\,\not\mathscr{R}\,e_2\\ s'e'\,\not\mathscr{R}\,e_2}}\varphi^{-1}(s')\,\varphi^{-1}(e')^\omega\,A^{<\omega_1}\Bigg)^\omega A^{<\omega_1} \tag{5}$$

and so on.

It remains to prove that $\varphi^{-1}(s)\in SF(A,[1,\omega_1[)$ in the right members of inclusions 4, 5 in the third union of the right member of equality (2) to show that $\bigcup_{p\nleqslant_{\mathscr{D}} r}\varphi^{-1}(r)\in SF(A,[1,\omega_1[)$. In order to do it we will show that if there exist two elements x, y of S such that $p\leqslant_{\mathscr{D}} xs$, $p\leqslant_{\mathscr{D}} sy$, and $p\nleqslant_{\mathscr{D}} xsy$ then $p<_{\mathscr{D}} s$. It follows from the hypothesis that $\varphi^{-1}(s)\in SF(A,[1,\omega_1[)$.

We can deduce from $p\leqslant_{\mathscr{D}} xs$ that $p\leqslant_{\mathscr{D}} s$. Assume $p\,\mathscr{D}\,s$. Since $p\leqslant_{\mathscr{D}} xs$ and $p\leqslant_{\mathscr{D}} sy$ then $p\,\mathscr{D}\,xs\,\mathscr{D}\,sy\,\mathscr{D}\,s$. Applying Proposition 1.12 it follows that $p\,\mathscr{D}\,xsy$, which is a contradiction. So $p<_{\mathscr{D}} s$ and $\varphi^{-1}(s)\in SF(A,[1,\omega_1[)$. This ends the proof of the lemma. ∎

We now prove that $\varphi^{-1}(\mathscr{R}(p))\,A^{<\omega_1}\in SF(A,[1,\omega_1[)$ in equality (1):

LEMMA 7.7. *If S is a finite ω_1-semigroup and p an element of S such that $\varphi^{-1}(s)\in SF(A,[1,\omega_1[)$ for every $s\in S$ such that $p<_{\mathscr{D}} s$, then $\varphi^{-1}(\mathscr{R}(p))\,A^{<\omega_1}\in SF(A,[1,\omega_1[)$.*

Proof. It can easily be checked that

$$\varphi^{-1}(\mathscr{R}(p))\,A^{<\omega_1}=\Bigg(\bigcup_{\substack{s\in S^1\\ a\in A\\ s\varphi(a)\,\mathscr{R}\,p\\ s\,\not\mathscr{R}\,p}}\varphi^{-1}(s)\,aA^{<\omega_1}\Bigg)\cup\Bigg(\bigcup_{\substack{(s',e')\in S\times E(S)\\ s'e'^\omega\,\mathscr{R}\,p\\ s'\,\not\mathscr{R}\,p\\ s'e'\,\not\mathscr{R}\,p}}\varphi^{-1}(s')\,\varphi^{-1}(e')^\omega\,A^{<\omega_1}\Bigg)$$

which belongs to $SF(A,[1,\omega_1[)$ by Lemma 7.4. ∎

We can prove a stronger result:

LEMMA 7.8. *If S is a finite ω_1-semigroup and p an element of S such that $\varphi^{-1}(s) \in SF(A, [1, \omega_1[)$ for every $s \in S$ such that $p <_{\mathscr{D}} s$, then $\varphi^{-1}(\mathscr{R}(p)) \in SF(A, [1, \omega_1[)$.*

Proof. We first show the equality

$$\varphi^{-1}(\mathscr{R}(p)) = \varphi^{-1}(\mathscr{R}(p))\, A^{<\omega_1} - \bigcup_{p \nleqslant_{\mathscr{D}} r} \varphi^{-1}(r).$$

The inclusion from left to right is trivial. Let us see the other one. Let u be a word of the right member of the equality. Then u has a prefix v such that $\varphi(v) \in \mathscr{R}(p)$. Since $u \notin \bigcup_{p \nleqslant_{\mathscr{D}} r} \varphi^{-1}(r)$ then $p \leqslant_{\mathscr{D}} \varphi(u)$, so $\varphi(u)\ \mathscr{D}\ p$ because $\varphi(v) \in \mathscr{R}(p)$. By Proposition 1.2 it follows that $\varphi(u) \in \mathscr{R}(p)$. The lemma follows from Lemmas 7.7 and 7.6. ∎

We next show that $A^{<\omega_1}\varphi^{-1}(\mathscr{L}(p)) \in SF(A, [1, \omega_1[)$ in equality (1):

LEMMA 7.9. *If S is a finite ω_1-semigroup and p an element of S such that $\varphi^{-1}(s) \in SF(A, [1, \omega_1[)$ for every $s \in S$ such that $p <_{\mathscr{D}} s$, then $A^{<\omega_1}\varphi^{-1}(\mathscr{L}(p)) \in SF(A, [1, \omega_1[)$.*

Proof. We show that

$$A^{<\omega_1}\varphi^{-1}(\mathscr{L}(p)) = \Bigg(\bigcup_{\substack{s \in S^1 \\ a \in A \\ \varphi(a)\, s \mathscr{L} p \\ s \not\mathscr{L}\, p}} A^{<\omega_1} a \varphi^{-1}(s) \Bigg) \cup \Bigg(\bigcup_{\substack{(s', e') \in S^1 \times S \\ e'^{\omega} s' \mathscr{L} p \\ s' \not\mathscr{L}\, p}} A^{<\omega_1}\, \overrightarrow{\varphi^{-1}(\mathscr{R}(e')) \cdot \varphi^{-1}(\mathscr{R}(e'))}\, \varphi^{-1}(s') \Bigg).$$

First the inclusion from right to left. The first union of the right member of the equality is clearly included in the left member of the equality. Assume now that u belongs to the second union. Then u has a suffix in $\overrightarrow{\varphi^{-1}(\mathscr{R}(e')) \cdot \varphi^{-1}(\mathscr{R}(e'))}\, \varphi^{-1}(s')$. Let w be a prefix of this suffix that belongs to $\overrightarrow{\varphi^{-1}(\mathscr{R}(e')) \cdot \varphi^{-1}(\mathscr{R}(e'))}$ such that $u = xwv$ with $v \in \varphi^{-1}(s')$ and $x \in A^{<\omega_1}$. Let $(x_j y_j)_{j<\omega}$ be an ω-sequence of prefixes of w such that $x_j, y_j \in \varphi^{-1}(\mathscr{R}(e'))$, $|x_j| > |x_{j-1} y_{j-1}|$ for every integer $j > 0$ and $(|x_i y_i|)_{i<\omega}$ is cofinal with $|w|$. Let $(z_j)_{j<\omega}$ be the ω-sequence of words such that $x_{j+1} = x_j z_j$ for any integer j. Since $\varphi(x_i)\ \mathscr{R}\ e'$ and $(|x_i|)_{i<\omega}$ is cofinal with $|w|$, there does not exist a factor w' of w such that $e' \nleqslant_{\mathscr{D}} \varphi(w')$. Since $z_i \in y_i A^{<\omega_1}$ and $\varphi(y_i)\ \mathscr{R}\ e'$ for any integer i, then $\varphi(z_i)\ \mathscr{R}\ e'$ by Proposition 1.2. But $w = x_0 \prod_{i<\omega} z_i$, so $w \in \varphi_i^{-1}(\mathscr{R}(e'))^{\omega}$. As a consequence of Theorem 1.3 there exist a strictly increasing $(n_i)_{i<\omega}$ of integers and a linked pair (s'', e'') of elements of S such that $\varphi(x_0 z_0 \cdots z_{n_0}) = s''$ and $\varphi(z_{n_i+1} \cdots z_{n_i+1}) = e''$ for any integer i. Since in $x_0 z_0 \cdots z_{n_0}$ (resp. $z_{n_i+1} \cdots z_{n_i+1}$ for any integer i) $\varphi(x_0)\ \mathscr{R}\ e'$ (resp. $(z_{n_i+1}\ \mathscr{R}\ e')$ then $\varphi(x_0 z_0 \cdots z_{n_0})\ \mathscr{R}\ e'$ (resp. $\varphi(z_{n_i+1} \cdots z_{n_i+1})\ \mathscr{R}\ e'$) according to Proposition 1.2. So $u \in A^{<\omega_1}\varphi^{-1}(s'')\, \varphi^{-1}(e'')^{\omega}\, \varphi^{-1}(s)$ with $s''\ \mathscr{R}\ e'$ and $e''\ \mathscr{R}\ e'$. But $e''^{\omega} = e'^{\omega}$ according to Proposition 1.6. So $u \in A^{<\omega_1}(e'^{\omega}s)$ and u belongs to the left member of the equality since $e'^{\omega}s \mathrel{\mathscr{L}} p$.

Let us turn now to the inclusion from left to right. Let u be a word in the left member of the equality. Then u has a suffix v such that $\varphi(v) \mathrel{\mathscr{L}} p$. Take v as small as possible (observe that v is not unique). Let x be the smallest prefix of v such that $v = xy$ and $\varphi(y) \not\mathrel{\mathscr{L}} p$. If $|x| \in Succ$ let a be the last letter of x. Then $\varphi(ay) \mathrel{\mathscr{L}} p$ so u belongs to the first union of the right member of the equality. If $|x| \in Lim$ write $|x|$ in Cantor normal form, $|x| = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n}$ (with $\alpha_n > 0$), and let w be a suffix of x of length ω^{α_n}. Since $|w|$ is a countable limit ordinal as a consequence of Theorem 1.3 there exists a linked pair (s', e') of S such that $w \in \varphi^{-1}(s')\,\varphi^{-1}(e')^{\omega}$. Let w_1 and w_2 be such that $w = w_1 w_2$, $w_1 \in \varphi^{-1}(s')$, and $w_2 \in \varphi^{-1}(e')^{\omega} \subseteq \overrightarrow{\varphi^{-1}(e') \cdot \varphi^{-1}(e')} \subseteq \overrightarrow{\varphi^{-1}(\mathscr{R}(e')) \cdot \varphi^{-1}(\mathscr{R}(e'))}$. Then u has $w_2 y$ as a suffix, $\varphi(w_2 y) \mathrel{\mathscr{L}} p$, $\varphi(y) \not\mathrel{\mathscr{L}} p$, and $e'^{\omega}\varphi(y) \mathrel{\mathscr{L}} p$ which proves that u belongs to the second union of the right member of the equality.

It remains to show that the right member of the equality belongs to $SF(A, [1, \omega_1[)$. Since S is finite the number of unions involved are finite, and since $SF(A, [1, \omega_1[)$ is closed under finite unions, products, and $\vec{\ }$ according to Lemma 7.2 it suffices to show that every $\varphi^{-1}(x)$ of the right member of the equality belongs to $SF(A, [1, \omega_1[)$. Let us begin by $\varphi^{-1}(s)$ in the first union. If $p \not\leqslant_{\mathscr{D}} s$ then $\varphi(a)\, s \mathrel{\mathscr{L}} p$ is impossible. So $p \leqslant_{\mathscr{D}} s$. If $p <_{\mathscr{D}} s$ then $\varphi^{-1}(s) \in SF(A, [1, \omega_1[)$ by the hypothesis of the lemma. Otherwise, $p \mathrel{\mathscr{D}} s$. If $p \mathrel{\mathscr{D}} s$ then since $\varphi(a)\, s \mathrel{\mathscr{L}} p$ implies $\varphi(a)\, s \mathrel{\mathscr{D}} p$, using Proposition 1.11 we have $p \mathrel{\mathscr{L}} s$, which is impossible. Now we show that $\varphi^{-1}(\mathscr{R}(e')) \in SF(A, [1, \omega_1[)$ in the second union. Since $e'^{\omega}s' \mathrel{\mathscr{D}} p$ then either $p <_{\mathscr{D}} e'$ or $p \mathrel{\mathscr{D}} e'$. In the first case $\varphi^{-1}(\mathscr{R}(e')) \in SF(A, [1, \omega_1[)$ follows from the induction hypothesis and the finiteness of S, in the second case from Lemma 7.8. Finally we check that $\varphi^{-1}(s') \in SF(A, [1, \omega_1[)$ in the second union using the same arguments as for $\varphi^{-1}(s)$ in the first union. This ends the proof of the lemma. ∎

It follows from Lemmas 7.6, 7.7, and 7.9 that every component of the right member of equality (1) belongs to $SF(A, [1, \omega_1[)$. We thus gave an effective algorithm to compute a star-free expression from any finite ω_1-semigroup.

8. EXAMPLES

We now give two examples of characterizations of star-free languages using their syntactic ω_1-semigroup. We first show an example of a language which is not star-free and an example of a star-free language.

EXAMPLE 8.1. Let $A = \{a, b\}$ and $L = (aa + b)^{<\omega_1} - \lambda$. The syntactic ω_1-semigroup S of L is given in Example 1.8. Since $a \mathrel{\mathscr{H}} 1$ then S contains a nontrivial group $\{1, a\}$; i.e., S is not aperiodic, so L is not star-free and not definable by a first-order sentence. An automaton recognizing L is given in Example 1.5 and a second-order sentence ϕ such that $\mathscr{L}(\phi) = L$ in Example 2.1.

EXAMPLE 8.2. Let $A = \{a, b\}$ and $S = \{a, b, 0, ab, ba\}$ be the ω_1-semigroup whose $\mathscr{D}$-classes structure is

a	$^* ab$
$^* ba$	b

$^* 0$

and such that 0 is a zero S, $a^2 = b^2 = 0$, $aba = a$, $bab = b$, and $(ab)^\omega = ab$. Let φ: $A^{[1, \omega_1[} \to S$ be the morphism of ω_1-semigroups defined by $\varphi(a) = a$ and $\varphi(b) = b$. Then S recognizes $L = (ab)^{<\omega_1} - \lambda$ since $L = \varphi^{-1}(\{ab\})$. Furthermore, S is the syntactic ω_1-semigroup of L and is aperiodic. So we have $L \in SF(A, [1, \omega_1[)$:

$$L = A^{[1, \omega_1[} - (L'bA^{<\omega_1} \cup A^{<\omega_1}a \cup A^{<\omega_1}aaA^{<\omega_1} \cup A^{<\omega_1}bbA^{<\omega_1}),$$

where $L' = A^{<\omega_1} - A^{<\omega_1}A$ is the set of words of limit or zero length.

A first-order sentence ϕ such that $\mathscr{L}(\phi) = L$ is

$$\begin{aligned}\phi = {} & (\forall x(\neg \exists y\; x = y+1 \to R_a(x))) \wedge (\forall x R_a(x) \to (\exists y\; y = x+1 \wedge R_b(y))) \\ & \wedge (\forall x(\exists y\; x < y \wedge R_b(x)) \to \exists z\; z = x+1 \wedge R_a(z)) \wedge (\exists x\; R_a(x) \vee R_b(x)).\end{aligned}$$

The first term of the conjunction expresses that every letter without predecessor is an "a," the second that every "a" is followed by a "b," the third one that every "b" which is not the last letter of the word is followed by an "a," and the last excludes the empty word from $\mathscr{L}(\phi)$. An automaton recognizing L is given in Example 1.6 and a second-order sentence logically equivalent to ϕ in Example 2.2.

9. VARIETIES

The links between varieties and star-free sets of words of countable length are discussed in this section.

The following definitions and results are issued from the theory of universal algebra (see [Alm94]). Let S and T be two ω_1-semigroups whose products are respectively denoted by φ_S and φ_T. We say that S *divises* T if there exists a surjective morphism from a sub-ω_1-semigroup of T of S. The *product* of S and T is the ω_1-semigroup composed of the elements of $S \times T$ and verifying, for any countable ordinal α and sequences $(s_\beta)_{\beta<\alpha}$ and $(t_\beta)_{\beta<\alpha}$ of respectively elements of S and T:

$$\varphi_{S\times T}\left(\prod_{\beta<\alpha} (s_\beta, t_\beta)\right) = \left(\varphi_S\left(\prod_{\beta<\alpha} s_\beta\right), \varphi_T\left(\prod_{\beta<\alpha} t_\beta\right)\right).$$

DEFINITION 9.1. A *pseudo-variety* of finite ω_1-semigroups is a class of finite ω_1-semigroups closed under product and division.

PROPOSITION 9.1. *Let A be an alphabet and $\{(u_i, v_i)\colon 0 < i < \omega\} \subseteq A^{[1,\omega[} \times A^{[1,\omega[}$. The class of finite ω_1-semigroups that satisfy all or all but finitely many of the identities $u_i = v_i$ is a pseudo-variety of finite ω_1-semigroups.*

As an example, the class of finite aperiodic ω_1-semigroups is a pseudo-variety of finite ω_1-semigroups since any element x of any finite ω_1-semigroup of this class verifies all but finitely many of the identities $(x^i = x^{i+1})_{i<\omega}$.

We now turn to the definition of a variety of languages. Let S be an ω_1-semigroup, $X \subseteq S$, $s \in S$, and

- $s^{-1}X = \{t \in S : st \in X\}$
- $Xs^{-\omega} = \{t \in S : (ts)^\omega \in X\}$
- $Xs^{-1} = \{t \in S : ts \in X\}$.

DEFINITION 9.2. A *variety of* ω_1-languages $\mathscr{V}$ is a function which associates to any alphabet A a class $A^{[1,\omega_1[}\mathscr{V}$ of rational ω_1-languages of $A^{[1,\omega_1[}$ such that:

- for any alphabet A, $A^{[1,\omega_1[}\mathscr{V}$ is a boolean algebra;
- for any alphabet A, if $X \in A^{[1,\omega_1[}\mathscr{V}$ and $x \in A^{[1,\omega_1[}$, $x^{-1}X \in A^{[1,\omega_1[}\mathscr{V}$, $Xx^{-1} \in A^{[1,\omega_1[}\mathscr{V}$, and $Xx^{-\omega} \in A^{[1,\omega_1[}\mathscr{V}$;
- if $\varphi : A^{[1,\omega_1[}\ B^{[1,\omega_1[}$ is a morphism of free ω_1-semigroups and $X \in B^{[1,\omega_1[}\mathscr{V}$ then $\varphi^{-1}(X) \in A^{[1,\omega_1[}\mathscr{V}$.

We refer to [Bed98b] for a proof of the following theorem:

THEOREM 9.1 [Bed98b, BC98]. *The map $\mathbf{V} \to \mathscr{V}$ is a bijection between pseudo-varieties of finite ω_1-semigroups and varieties of ω_1-languages.*

In particular, the star-free languages of words of countable length are a variety of languages.

10. CONCLUSION

This paper extends to languages of words of any countable length the theorem of Schützenberger, McNaughton, and Papert that establishes that any language of finite words expressed with one of the three formalisms, the finite aperiodic semigroups, first-order sentences, and star-free expressions, can also be expressed in the two others. All constructions used in the proof are effective.

The star-free languages of finite words were the first class of languages characterized by algebraic properties. A lot of other subclasses of recognizable sets have also been characterized by their algebraic properties since that time. Such results could be extended to languages of words of countable length.

Two kinds of logic formulæ were used in this paper: first-order formulæ and second-order formulæ. Another kind of logics is often used to define sets of finite or ω words: the temporal logics. Temporal logics are used in practice to represent the behavior of processes. They do not use quantifiers. Particular temporal logic formulæ are built from atomic formulæ R_a where a is a letter of the alphabet and by induction using the usual boolean connectors and three temporal connectors

N, $\Diamond$, and $\mathbb{U}$: if ϕ and ψ are both temporal logic formulæ, then so are $\varphi \vee \psi$, $\neg \varphi$, $\mathbf{N}\varphi$, $\Diamond\varphi$, and $\varphi \mathbb{U} \psi$. The semantics of those formulæ is defined by:

- $u \models R_a$ if the first letter of u is an a,
- $u \models \varphi \vee \psi$ if $u \models \varphi$ or $u \models \psi$,
- $u \models \neg \varphi$ if $u \not\models \varphi$,
- $u \models \mathbf{N}\varphi$ if $u[1, |u|[\models \varphi$,
- $u \models \Diamond\varphi$ if there exists $\beta < |u|$ such that $u[\beta, |u|[\models \varphi$,
- $u \models \varphi \mathbb{U} \psi$ if there exists $\beta < |u|$ such that $u[\beta, |u|[\models \psi$ and $u[\alpha, |u|[\models \varphi$ for every $\alpha < \beta$.

There exist many proofs [Kam68, GPSS80, CPP93, CC91] of the equivalence between this kind of temporal logic and the first-order logic used in this paper, if the length of words considered is finite or at most ω. Rohde [Roh97] recently showed that this temporal logic is *stricly* included in the first order logic we used as soon as models of length greater than ω are considered. The algebraic characterization of languages described by formulæ of this kind of temporal logic is still an open problem. In fact we do not know if the class of such languages is even a variety of languages.

Finally, the empty word was excluded from the languages studied in this paper. Taking it into account can easily be done by replacing ω_1-semigroups with ω_1-monoids (ω_1-semigroups with a neutral element).

ACKNOWLEDGMENTS

The author expresses his thanks to Dominique Perrin for directing the research and to Richard E. Ladner for Proposition 6.1. The anonymous referee had a difficult task in the careful reading of a previous version of this paper. His comments helped improve the quality of this final version.

REFERENCES

[Alm94] J. Almeida, "Finite Semigroups and Universal Algebra," Series in Algebra, Vol. 3, World Scientific, Singapore, 1994.

[Arn85] A. Arnold, A syntactic congruence for rational ω-languages, *Theoret. Comput. Sci.* **39** (1985), 333–335.

[BC98] N. Bedon and O. Carton, An Eilenberg theorem for words on countable ordinals, *in* "Latin'98: Theoretical Informatics, Campinas, Brazil, April 1998," Vol. 1380, pp. 53–64, (C. L. Lucchesi and A. V. Moura, Eds.), Springer-Verlag, Berlin/New York.

[Bed] N. Bedon, "Star-Free Sets of Words on Ordinals," *Inform. and Comput.*, in press.

[Bed98a] N. Bedon, Automata, semigroups and recognizability of words on ordinals, *Internat. J. Algebra Comput.* **8** (1998), 1–21.

[Bed98b] N. Bedon, "Langages Reconnaissables de mots indexés par des ordinaux," Ph.D. thesis, Université de Marne-la-Vallée, Cité Descartes, 5, Boulevard Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée, Cedex 2, France, January 1998. Available at http://www-igm.univ-mlv.fr/~ bedon/Recherche/These/these.shtml.

[Büc60] J. R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik. Grund. Math.* **6** (1960), 66–92.

[Büc62] J. R. Büchi, On a decision method in the restricted second-order arithmetic, *in* "Logic, Methodology and Philosophy of Science: Proc. Intern. Congr.," pp. 1–11, Stanford University Press, Stanford, CA, 1962.

[Büc64] J. R. Büchi, Transfinite automata recursions and weak second order theory of ordinals, *in* "Proc. Int. Congress Logic, Methodology, and Philosophy of Science," Jerusalem, pp. 2–23, 1964, North-Holland, Amsterdam.

[Büc65] J. R. Büchi, Decision methods in the theory of ordinals, *Bull. Amer. Math. Soc.* **71** (1965), 767–770.

[CC91] J. Cohen-Chesnot, On the expressive power of temporal logic for infinite words, *Theoret. Comput. Sci.* **83** (1991), 301–312.

[CPP93] J. Cohen, D. Perrin, and J.-E. Pin, On the expressive power of temporal logic, *J. Comput. System Sci.* **46** (1993), 271–294.

[Ehr61] A. Ehrenfeucht, An application of games to the completeness problem for formalized theories, *Fund. Math.* **49** (1961), 129–141.

[GPSS80] D. M. Gabbay, A. Pnueli, S. Shelah, and J. Stavi, On the temporal basis of fairness, *in* "Conference Record of the Seventh Annual ACM Symposium on Principles of Programming Languages, Las Vegas, Nevada, January 1980," pp. 163–173.

[Kam68] J. Kamp, "Tense Logic and the Theory of Linear Order," Ph.D. thesis, University of California, Los Angeles, 1968.

[Lad77] R. E. Ladner, Application of model theoretic games to discrete linear orders and finite automata, *Inform. and Control* **33** (1977), 281–303.

[MP71] R. McNaughton and S. Papert, "Counter Free Automata," MIT Press, Cambridge, MA, 1971.

[Mul63] D. Muller, Infinite sequences and finite machines, *in* "Switching Circuit Theory and Logical Design : Proc. Fourth Annual Symp. IEEE," pp. 3–16, 1963.

[Péc86a] J.-P. Pécuchet, "Etude syntaxique des parties reconnaissables de mots infinis," Lecture Notes in Computer Science, Vol. 226, pp. 294–303, Springer-Verlag, Berlin/New York 1986.

[Péc86b] J.-P. Pécuchet, "Variétés de semigroupes et mots infinis," Lecture Notes in Computer Science, Vol. 210, pp. 180–191, Springer-Verlag, Berlin/New-York, 1986.

[Per84] D. Perrin, Recent results on automata and infinite words, *in* "Mathematical Foundations of Computer Science" (M. P. Chytil and V. Koubek, Eds.), Lecture Notes in Computer Science, Vol. 176, pp. 134–148, Springer-Verlag, Berlin/New York, 1984.

[Per90] D. Perrin, Finite automata, *in* "Handbook of Theoretical Computer Science," Vol. B, pp. 1–53. Elsevier, Amsterdam, 1990.

[Pin84] J.-E. Pin, "Variétés de langages formels," Masson, Paris, 1984. [English version: "Varieties of Formal Languages," Plenum Press, New York, 1986]

[PP86] D. Perrin and J.-E. Pin, First order logic and star-free sets, *J. Comput. System Sci.* **32** (1986), 393–406.

[PP97] D. Perrin and J.-E. Pin, Mots infinis (Infinite words), LITP report 97–04, 1997, http://www.liafa.jussieu.fr/~jep/Resumes/InfiniteWords.html. [English version "Infinite words"].

[Roh97] S. Rohde, "Alternating Automata and the Temporal Logic of Ordinals," Ph.D. thesis, University of Illinois, Urbana-Champaign, IL, 1997.

[Ros82] J. G. Rosenstein, "Linear Ordering," Academic Press, New York 1982.

[Sch65] M.-P. Schützenberger, On finite monoids having only trivial subgroups, *Inform. and Control* **8** (1965), 190–194.

[Sie65] W. Sierpiński, "Cardinals and Ordinal Numbers," Polish Scientific, Varsovie, 1965.

[Str94] H. Straubing, "Finite Automata, Formal Logic and Circuit Complexity," Birkhäuser, Basel, 1994.

[Tho79] W. Thomas, Star free regular sets of ω-sequences, *Inform. and Control* **42** (1979), 148–156.

[Wil91] T. Wilke, An Eilenberg theorem for ∞-languages, *in* "Automata, Languages and Programming: Proc. of 18th ICALP Conference," pp. 588–599, Springer-Verlag, Berlin/New York, 1991.

[Woj84] J. Wojciechowski, Classes of transfinite sequences accepted by finite automata, *Fund. Inform.* **7** (1984), 191–223.

[Woj85] J. Wojciechowski, Finite automata on transfinite sequences and regular expressions, *Fund. Inform.* **8** (1985), 379–396.