Star-free sets of words on ordinals*

Nicolas Bedon†
Institut Gaspard Monge
Université de Marne-la-Vallée
5, boulevard Descartes
Champs-sur-Marne
77454 Marne-la-Vallée Cedex 2
France

February 21, 2001

Abstract

Let n be a fixed integer; we extend the theorem of Schützenberger, McNaughton and Papert on star-free sets of finite words to languages of words of length less than ω^n.

Finite automata are a formalism for defining sets of words. They began to be studied in the 1950s. Among the first important results of this theory, Kleene proved [Kle56] that this formalism, when used to define sets of finite words, is equivalent to another one, the rational expressions. The class of rational expressions is the smallest class containing the letters and closed under finite union, product, and Kleene closure. It is also a well-known result that finite automata, monadic second-order logic [Büc60], and finite semigroups are equivalent formalisms for defining sets of finite words. The algebraic approach gives access to powerful tools for the study of properties of such sets. By analogy with the automata theory, one can attach to any rational set of finite words X a canonical semigroup, called the syntactic semigroup of X. Algebraic properties of such semigroups can be used to define subclasses of the rational sets of finite words. In particular, a rational set belongs to the smallest set containing the letters and closed under finite boolean operations and product if and only if its syntactic semigroup is finite and group-free [Sch65]. Such sets, called **star-free**, are also definable by first-order logic formulae, and conversely [MP71].

Finite automata on ω-words were first introduced by Büchi [Büc62] to prove the decidability of the monadic second-order theory of integers. As for the finite word case, finite automata on ω-words are equivalent to rational expressions introduced by McNaughton [McN66], looking like those of Kleene but with an

†Nicolas.Bedon@univ-mlv.fr, http://www-igm.univ-mlv.fr/~bedon/

added unary ω operator standing for the ω repetition of a rational set of finite words. Both formalisms are equivalent to finite semigroups with an adapted structure for the infinite product. A first attempt in the direction of the algebraic approach to the theory of ω-words was made by J.P. Pécuchet [Péc86a, Péc86b], but a more satisfying one is due to T. Wilke [Wil91] and D. Perrin and J.E. Pin [PP97] with the introduction of ω-semigroups. As for the finite word case, one can link to any rational set X of ω-words the syntactic ω-semigroup of X, which is finite and unique. This differs from the automata theory, where we do not know how to attach a canonical "minimal" automaton to any rational set of ω-words. The result on star-free sets on finite words was extended to ω-words by R.E. Ladner [Lad77], W. Thomas [Tho79], and D. Perrin [Per84].

Büchi [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 that those automata are equivalent to a generalization of ω-semigroups, that are finite algebraic structures called ω^n-semigroups [Bed98b, Bed98a]. As for the finite and ω-words cases, there exists for every set of words accepted by a Büchi automata an ω^n-semigroup which is canonical and finite and recognizes the same set.

In this paper we first recall the algebraic definitions on ω^n-semigroups and introduce logic formulae to define sets of words. Then, we extend the theorem on star-free sets of finite and ω-words to sets of words of length less than ω^n for an integer n. In order to obtain effective constructions we extend the ideas of [Lad77] to obtain a decision procedure for the question "$x \models \phi$" for a first-order sentence ϕ, where x belongs to a particular class of words on ordinals.

A reader knowledge of ordinals is assumed. Although we tried to write a self-contained article, previous knowledge of automata and semigroups is also beneficial.

1 Notations and definitions

For the theory of ordinals we refer to [Sie65] or [Ros82]. We denote by Succ the class of successor ordinals, Lim the class of limit ordinals, and $\mathit{Ord} = \mathit{Succ} \cup \mathit{Lim} \cup \{0\}$. 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 $\mathit{end}(\alpha)$, is ω^{α_k}. Let α be an ordinal and A a finite 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 an ordinal smaller 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 use them interchangeably. The **length** of u is denoted by $|u|$. The only word of length 0 is the **empty word**.

Example 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 is an "a" and second letter is a "b". For pratical reasons u is also denoted by mere concatenation: $u = ab$.*

Example 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 and "b" at odd positions.*

Example 3 *Let $A = \{a, b\}$. The word u of length $\omega + 2$ defined by $u_{2\alpha} = a$, with $\alpha \leq \omega$, 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, denoted $u \cdot v$, or uv, is the word w of length $\alpha + \beta$ on $A_u \cup A_v$ such that:

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

Example 4 *Let u be the word of example 1 and v the word of example 2. The product of v and u is the word of example 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* of w, and z a *right factor* of w. Let α and β be ordinals with $\alpha < \beta$ and u a word such that $|u| \geq \beta$. By $u[\alpha, \beta[$ we denote the word such that $u[\alpha, \beta[(\gamma) = u(\alpha + \gamma)$ for any $0 \leq \gamma < \beta - \alpha$. A decomposition of a word into a product of factors is called a **factorization**. Let A be an alphabet, α and β be ordinals such that $\beta < \alpha$. 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 $\beta \leq \gamma < \alpha$. The powerset of a set S is denoted by $\mathcal{P}(S)$ and its cardinal $|S|$.

1.1 Semigroups

A **semigroup** 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 e of a semigroup is called **idempotent** if $e^2 = e$. 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 denoted 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 denoted 1. Let S be a semigroup. A **sub-semigroup** S' of S is a subset of S such that

S' is a semigroup. We denote by S^1 the monoid $S \cup \{1\}$ if S is not a monoid and S otherwise. A subset I of a semigroup S is an **ideal** of S iff $S^1 I S^1 = I$. 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 **quotient** of a semigroup S if there exists a surjective morphism $\varphi : S \to T$. A **congruence** is an equivalence relation preserving operations, usually denoted $\sim$. For example, a semigroup congruence $\sim$ verifies $x \sim y \Rightarrow uxv \sim uyv$. This condition ensures that if S is a semigroup, then the set of equivalence classes $S/\!\sim$ 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 of finite words one can attach a canonical finite semigroup. A similar result holds in the theory of ω-words.

Let us turn to the case of words of length less than ω^n. We refer to [Bed98a, Bed98b] for more details about the basic theory of ω^n-semigroups. 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 5 *Let A be an alphabet, i an integer, u a word over A such that $|u| = \omega^i$, S a finite set, and $\varphi : A^{[1,\omega^i[} \to S$ a function. Let $u = u_0 u_1 u_2 \dots$ be the factorization of u such that $|u_j| = \omega^{i-1}$ for any integer j. There exist an increasing infinite sequence of integers $(k_j)_{j \in \mathbb{N}}$ and $s, t \in S$ such that $\varphi(u_0 \dots u_{k_0}) = s$ and $\varphi(u_{k_j+1} \dots u_{k_{j+1}}) = t$ for any integer j.*

Definition 6 *Let n be an integer. An ω^n**-semigroup** S is a set equipped with a partial function (called the* product *of S) $\psi : \cup_{0<\alpha<\omega^{n+1}} S^\alpha \to S$ such that*

1. *$\psi(s) = s$ for any $s \in S$,*
2. *if $\alpha < \omega^{n+1}$ and $(s_\beta)_{\beta<\alpha}$ is a sequence of elements of S, then for any increasing sequence $(\gamma_\delta)_{\delta<\delta_l}$ such that $\gamma_0 = 0$ and $\delta_l \leq \alpha$,*

$$\psi(s_0, s_1, \dots) = \psi(\psi(s_{\gamma_0}, s_{\gamma_0+1}, \dots), \psi(s_{\gamma_1}, s_{\gamma_1+1}, \dots), \psi(s_{\gamma_2}, s_{\gamma_2+1}, \dots), \dots)$$

3. *S, which is then equipped with a structure of semigroup, is partitioned into $n+1$ sub-semigroups $S_0, S_1, \dots, S_n$,*
4. *$\cup_{i \leq j} S_i$ is a semigroup of ideal S_j for any $j \leq n$,*
5. *if $s = (s_k)_{k<\omega}$ is a sequence of elements of S_i, then $\psi(s) \in S_{i+1}$ if $i < n$, and is not defined otherwise.*

Observe the notation $\psi(s_0, s_1, \dots)$ for $\psi(t)$, where $t = (s_\beta)_{\beta<\alpha}$ is a sequence of elements of S, and that the notation $s_0 s_1 s_2 \dots$ can unambiguously be used for $\psi(s_0, s_1, s_2, \dots)$.

Example 7 *Let A be an alphabet and n an integer. Then the product of words naturally equips $A^{[1,\omega^{n+1}[}$ with a structure of ω^n-semigroup. We thus have $A_i = A^{[\omega^i,\omega^{i+1}[}$ for any $i \leq n$.*

Example 8 *An ω^0-semigroup is an ordinary semigroup.*

Definition 9 *Let n be an integer. An ω^n-**Wilke algebra** S is a finite semigroup partitioned into $n+1$ sub-semigroups $S_0, S_1, \dots, S_n$ such that for every $j \leq n$, S_j is an ideal of $\cup_{i\leq j} S_i$, and equipped with a family of n functions from S_i to S_{i+1} denoted by $s \to s^{\omega_i}$ such that, for all $s, t \in S_i$,*

$$s(ts)^{\omega_i} = (st)^{\omega_i} \tag{1}$$

$$(s^n)^{\omega_i} = s^{\omega_i} \text{ for all } n > 0 \tag{2}$$

For brevity, we shall omit the subscripts of ω's.

The following theorem is a direct consequence of Wilke's theory (see [Wil91]). It shows that finite ω^n-semigroups are equivalent to ω^n-Wilke algebras.

Theorem 10 *Let n be an integer and S a finite semigroup partitioned into $n+1$ sub-semigroups $S_0, \dots, S_n$ such that for every $j \leq n$, S_j is an ideal of $\cup_{i\leq j} S_i$. Assume there exist n unary functions $\omega_i : S_i \to S_{i+1}$ for $0 \leq i < n$ such that, for all $s, t \in S_i$, (1) and (2) of definition 9 are verified. Then S (as a set) can be equipped in a unique way with a structure of ω^n-semigroup such that $s^\omega = \psi((t_k)_{k<\omega})$, where $t_k = s$ for any integer k, and $\psi(s,t) = s \cdot t$ for any integer $s, t \in S$. Conversely, let S be a finite ω^n-semigroup. Then S (as a set) can be equipped in a unique way of a finite associative product $\cdot$ such that $s \cdot t = \psi(s,t)$ for any $s, t \in S$, and with n unary functions $\omega_i : S_i \to S_{i+1}$ for $0 \leq i < n$ such that, for all $s, t \in S_i$, (1) and (2) of definition 9 are verified, and $s^\omega = \psi((t_k)_{k<\omega})$, where $t_k = s$ for any integer k.*

From now on we shall not differentiate between finite ω^n-semigroups and ω^n-Wilke algebras.

Morphisms of ω^n-semigroups are defined like in universal algebra:

Definition 11 *Let n be an integer, S and T two ω^n-semigroups. A **morphism of ω^n-semigroups** $\varphi : S \to T$ is a function verifying, for any sequence $(s_\beta)_{\beta<\alpha}$ of elements of S such that $\psi_S(s_0, s_1, \dots)$ is defined,*

$$\varphi(\psi_S(s_0, s_1, \dots)) = \psi_T(\varphi(s_0), \varphi(s_1), \dots)$$

We say that φ **recognizes** a subset X of S if $\varphi^{-1}\varphi(X) = X$. This subset X is **recognizable** if there exist a finite ω^n-semigroup T' and a morphism $\varphi' : S \to T'$ of ω^n-semigroups such that φ' recognizes X. We also say that T **recognizes** X if there exists $\varphi'' : S \to T$ such that $\varphi''^{-1}\varphi''(X) = X$.

Remark 12 *Let A be an alphabet, n an integer, u a word over A of length less than ω^{n+1}, S an ω^n-semigroup, and $\varphi : A^{[1,\omega^{n+1}[} \to S$ a morphism of ω^n-semigroups. Then $\varphi(u) \in S_i$ iff $|u| = \sum_{j=i}^{0} \omega^j a_j$ where each a_j is an integer and a_i is not null.*

The notation s^ω now stands for the infinite product of ω elements $sss\ldots$.

Proposition 13 *Let A be an alphabet, n be an integer, S be a finite ω^n-semigroup and $\varphi : A^{[1,\omega^{n+1}[} \to S$ be a morphism of ω^n-semigroups. Let $x \in S_i$ and $\sum_{j=i}^{0} \omega^j a_j$ (with $a_i > 0$) be the length of the shortest word u such that $\varphi(u) = x$. Then $\sum_{j=0}^{i} a_j \leq |S_i|$.*

Proof Assume it is false. Let $(s_j)_{1 \leq j \leq \sum_{j=0}^{i} a_j}$ be the sequence of elements of S_i defined by ($t = \sum_{j=i-k}^{i} a_j + l$ with $l < a_{i-k-1}$)

$$s_t = \varphi(u[0, \sum_{j=i}^{i-k} \omega^j a_j + \omega^{i-k-1} l[)$$

If $\sum_{j=0}^{i} a_j > |S_i|$ there exist two integers k and l ($k < l$) less than or equal to $\sum_{j=0}^{i} a_j$ such that $s_k = s_l$. Let (with $l_1 < a_{i-k_1-1}$ and $l_2 < a_{i-k_2-1}$)

$$k = \sum_{j=i-k_1}^{i} a_j + l_1 \qquad \text{and} \qquad l = \sum_{j=i-k_2}^{i} a_j + l_2$$

Let

$$w = u[0, \sum_{j=i}^{i-k_1} \omega^j a_j + \omega^{i-k_1-1} l_1[\qquad \text{and} \qquad v = u[\sum_{j=i}^{i-k_2} \omega^j a_j + \omega^{i-k_2-1} l_2, |u|[$$

We have $\varphi(w) = s_k$. Let $\varphi(v) = y$. We have $\varphi(u) = s_k y = x$. Since

$$|v| = \omega^{i-k_2-1}(a_{i-k_2-1} - l_2) + \sum_{j=i-k_2-2}^{0} \omega^j a_j$$

and since either $k_2 > k_1$, or $k_2 = k_1$ and $l_2 > l_1$ one can verify that $|wv| < |u|$, but $\varphi(wv) = \varphi(w)\varphi(v) = s_k y = x$, which is a contradiction. ∇

Proposition 14 *Let A be an alphabet, n be an integer, S be a finite ω^n-semigroup and $\varphi : A^{[1,\omega^{n+1}[} \to S$ be a morphism of ω^n-semigroups. If $0 < i \leq n$ then for every $m \in S_i$,*

$$\varphi^{-1}(m) \cap A^{\omega^i} = \bigcup_{(s,e) \in P} \varphi^{-1}(s)\varphi^{-1}(e)^\omega$$

with $P = \{(s,e) \in S_{i-1} \times S_{i-1} \mid se = s,\ e^2 = e,\ \text{and } se^\omega = m\}$.

Proof First let $u \in \varphi^{-1}(s)\varphi^{-1}(e)^\omega$ such that $(s,e) \in P$. Then u has a factorization in ω factors $u = u_0u_1\dots$ such that $\varphi(u_0) = s$ and $\varphi(u_j) = e$ for every positive integer. It follows from $\omega^{i-1} \leq |u_j| < \omega^i$ for every integer j that $|u| = \omega^i$. The inclusion from right to left follows since $\varphi(u) = se^\omega = m$. Let us turn to the converse. Assume $u \in \varphi^{-1}(m) \cap A^{\omega^i}$. Using theorem 5, u has a factorization in ω factors $u = u_0u_1\dots$ such that $\varphi(u_0) = s_0$ and $\varphi(u_j) = t$ for every integer $j > 0$, with $|u_j| = \omega^{i-1}k_j$, where $k_j > 0$ is a integer for every $j \geq 0$, so $s_0 \in S_{i-1}$ and $t \in S_{i-1}$. Since every element of a finite semigroup has an idempotent power, there exists an integer k such that $t^k = t^{2k}$, and then a factorization of u in ω factors

$$u = u_0(u_1\dots u_{k+1})(u_{k+2}\dots u_{2(k+1)})\dots(u_{jk+j+1}\dots u_{(j+1)(k+1)})\dots$$

such that $\varphi(u_{jk+j+1}\dots u_{(j+1)(k+1)}) = t^k$ for every integer j. Let $e = t^k$ and $s = s_0e$. We have $s_0ee = se = s_0e = s$, so $u \in \cup_{(s,e)\in P}\varphi^{-1}(s)\varphi^{-1}(e)^\omega$. ∇

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

Proposition 15 *Let A be an alphabet, n be an integer, S be a finite ω^n-semigroup, s and e be elements of S_i such that $se = s$ and $e^2 = e$ and $\varphi : A^{[1,\omega^{n+1}[} \to S$ a morphism of ω^n-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 $P_{s,e} = \{f \in S_i \mid sf = s,\ ef = f \text{ and } f^2 = f\}$.

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 $(x_jy_j)_{j\in\mathbb{N}}$ 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 $(|x_jy_j|)_{j<\omega}$ is cofinal with $|u|$. Let $(z_j)_{j\in\mathbb{N}}$ be the ω-sequence of words such that $x_{j+1} = x_jz_j$ for any integer j. Using the same kind of argument as in the proof of proposition 14, $u = x_0z_0z_1\dots$ has a factorization $u = (x_0z_0\dots z_{n_0-1})(z_{n_0}\dots z_{n_1-1})(z_{n_1}\dots z_{n_2-1})\dots$ such that $\varphi(x_0z_0\dots z_{n_0-1}) = r$ and $\varphi(z_{n_j}\dots z_{n_{j+1}-1}) = f$ for some $r, f \in S_i$ such that $rf = r$ and $f^2 = f$. Since $\varphi(x_0z_0\dots z_{n_0-1}) = \varphi(x_j)$ for some j it follows that $r = s$. Since $\varphi(z_{n_0}\dots z_{n_1-1}) = f$, y_{n_0} is a prefix of z_{n_0} and $\varphi(y_{n_0}) = e$ it follows $f = eg$ for some $g \in \cup_{0\leq j\leq i}S_j$, so $ef = eeg = eg = f$, which ends the proof of the right inclusion. ∇

Corollary 16 $\varphi^{-1}(e)^\omega = \overrightarrow{\varphi^{-1}(e) \cdot \varphi^{-1}(e)}$

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

Theorem 17 *Let n be an integer, S an ω^n-semigroup, and X a recognizable subset of S. Among all congruences of ω^n-semigroups $\sim_X$ such that $S/\sim_X$ recognizes X, there exists an unique one from which all others are refinements. The number of equivalence classes of this congruence, which is minimal, is finite. This congruence of ω^n-semigroup, called* **syntactic congruence of** *X, is defined by the following: for any integer i less than $n+1$ and $x, y \in S_i$, $x \sim_X y$ if, for all $r, t \in S^1$,*

$$rxt \in X \iff ryt \in X \tag{3}$$

and, for any $m \in \mathbb{N}$ and $y_0, y_1, \ldots, y_m \in S^1$ such that

$$y_0(\ldots(((xy_1)^\omega y_2)^\omega y_3)^\omega \ldots)^\omega y_m$$

is defined

$$\begin{aligned} & y_0(\ldots(((xy_1)^\omega y_2)^\omega y_3)^\omega \ldots)^\omega y_m \in X \\ & \qquad \iff y_0(\ldots(((yy_1)^\omega y_2)^\omega y_3)^\omega \ldots)^\omega y_m \in X \end{aligned} \tag{4}$$

The quotient of S under the syntactic congruence of X is called the **syntactic ω^n-semigroup** of X and is usually denoted $S/\sim_X$. The function which associates to every element of S its congruence class in $S/\sim_X$ is a morphism of ω^n-semigroup, called the **syntactic morphism** of X.

We say that S is aperiodic if S viewed as a simple semigroup is aperiodic.

Proposition 18 *Let $\varphi : S \to T$ be a morphism of ω^n-semigroup that recognizes a subset X of S. Let $\sim_\varphi$ be the equivalence relation defined on S by $x \sim_\varphi y$ iff $\varphi(x) = \varphi(y)$. Then:*

1. *$\sim_\varphi$ is a congruence of ω^n-semigroups,*
2. *$S/\sim_\varphi$ recognizes X,*
3. *if T is aperiodic then so is $S/\sim_\varphi$. Furthermore, $S/\sim_\varphi$ is isomorphic to $\varphi(S)$,*
4. *the natural morphism $\varphi' : S \to S/\sim_\varphi$ which associates to any element of S its congruence class for $\sim_\varphi$ is surjective.*

Proposition 19 *Let $\sim_1$ and $\sim_2$ be two congruences on an ω^n-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 20 *Let A be an alphabet, n an integer and X a recognizable subset of $A^{[1,\omega^{n+1}[}$. Then X is recognizable by an aperiodic ω^n-semigroup iff $A^{[1,\omega^{n+1}[}/\sim_X$ is aperiodic.*

Proposition 21 *Let p, q and r be elements of an aperiodic ω^n-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 22 *Let p be an element of an aperiodic ω^n-semigroup S. Then $p = pS^1 \cap S^1 p \backslash \{r/p \notin S^1 r S^1\}$.*

Proof It is clear that p belongs to the right side of the equality. Now let n be in the right side of the equality. There exist x, y, r and s in S^1 such that $n = px = yp$ and $p = rns$. So $n = rnsx$ and it follows that $n = rn$ from proposition 21. We can prove $n = ns$ using the same argument. So $n = rns = p$. ∇

1.2 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.

1.2.1 Syntax

Let A be an alphabet. Our **first-order formulae** are inductively built from a set of (first-order) variables usually denoted by $x, y, z, x_1, y_1, z_1, \ldots$, an unary predicate R_a for each $a \in A$, a binary relation symbol $<$ of linear order, an existential quantifier $\exists$ on variables, a binary logical connector $\vee$, and an 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 formulae, 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 $\neg\phi \vee \psi$, $\phi \wedge \psi$ for $\neg(\neg\phi \vee \neg\psi)$, $x = y$ for $(\neg(x < y)) \wedge (\neg(y < x))$, $x \leq y$ 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 $\forall x((z \leq x \wedge x < y) \to \psi)$ and $\exists_z^y x\ \psi$ for $\exists x(z \leq x \wedge x < y \wedge \psi)$.

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

Definition 23 *Let ϕ be a first-order formula and x a first-order variable. The* ***quantifier height*** *of ϕ, denoted 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:

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

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 sub-formula of ϕ. An occurrence of a variable x in a formula ϕ is said to be **free** if $x \in FV(\psi)$. A non-free occurrence of a variable in a formula is said to be **bounded**. A **sentence** is a formula ϕ such that $FV(\phi) = \emptyset$.

Our **monadic second-order formulae** (or second-order formulae for short) are first-order formulae in which variables of sets, also called (monadic) second-order variables, are allowed. We make a difference between second-order and first-order variables by denoting the former using uppercase letters and the latter using lowercase letters. Formally, we build second-order formulae by adding five items to the rules of construction of first-order formulae:

- 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 monadic second-order formula.
- If X is a monadic second-order variable and ϕ a monadic second-order formula, then $\exists X\phi$ is a monadic second-order formula.
- If ϕ and ψ are both monadic second-order formulae then so are $\phi \vee \psi$ and $\neg\psi$.
- If x is a first-order variable and ϕ a monadic second-order formula then so is $\exists x\phi$.

1.2.2 Semantics

We now explain the meaning of a first-order formula (the semantic of monadic second-order formulae is not needed in the remainder of this work). We define $\mathcal{L}(\phi)$, the set of words verifying properties described by the formula ϕ, as in [PP86] (see also [Str94]):

Definition 24 *Let V be a finite set of variables and A an alphabet. A* ***V-marked word*** *of length α over A is a word $(a_0, V_0)(a_1, V_1)\cdots$ over $A \times \mathcal{P}(V)$ such that $V_\beta \cap V_\gamma = \emptyset$ if $\beta \neq \gamma$ and $\cup_{\beta<\alpha} V_\beta = V$.*

Definition 25 *Let ϕ be a formula, V a finite set such that $FV(\phi) \subseteq V$ and there is no $x \in V$ that appears bounded in ϕ, and $w = (a_0, V_0)\ldots(a_\beta, V_\beta)\ldots$ a V-marked word over an alphabet A. We say that w* ***satisfies*** *ϕ, and note $w \models \phi$, iff*

- *if ϕ has the form $\neg\psi$, not $w \models \psi$,*
- *if ϕ has the form $\psi \vee \chi$, $w \models \psi$ or $w \models \chi$,*
- *if ϕ has the form $x < y$, $x \in V_\beta$, $y \in V_\gamma$ and $\beta < \gamma$,*
- *if ϕ has the form $R_a(x)$, (a, V_β) is a letter of w with $x \in V_\beta$,*
- *if ϕ has the form $\exists x\psi$, $(a_0, V_0)\cdots(a_\beta, V_\beta \cup \{x\})\cdots \models \psi$ for some $\beta < |w|$.*

If $w = a_0 a_1 \ldots$ is a word over A and ϕ a first-order sentence, then $w \models \phi$ iff $(a_0, \emptyset)(a_1, \emptyset)\cdots \models \phi$.

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

Example 26 *The set of words of successor length containing an "a" letter is defined by the sentence*

$$\exists x R_a(x) \wedge \exists y \forall z (z \leq y)$$

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

This is a well-known result on formulae:

Definition 27 *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 exists any repetition of a conjunct 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, $hq(\varphi) \leq n$ and there does not exists any repetition of a conjunct or a disjunct.

Proposition 28 *Every first-order formula is logically equivalent to a first-order formula in disjunctive normal form of the same quantifier height.*

Corollary 29 *Let V be a finite set of first-order variables and n an integer. There exist only a finite number of first-order formulae ϕ such that $hq(\phi) \leq n$, modulo the logical equivalence, with variables in V.*

Proof We prove the result by induction on n. Since V is finite there exist only a finite number, say m, of formulae of the form $x < y$ or $R_a(x)$ or $\neg(x < y)$ or $\neg R_a(x)$, where $a \in A$ and x, y are variables. The number of conjunctions of disjunctions of such formulae is 2^{2^m}. Now let P be the set of first-order formulae of quantifier height less than n, $p = |P|$ and $\phi \in P$. There exist $2p$ formulae of the form $\exists\phi$ or $\neg\exists\phi$, and 2^{2p} conjunctions of such formulae. To each of this conjunction we must add formulae of P: we obtain $(p+1)2^{2p}$ formulae. The total number of disjunctions is $2^{(p+1)2^{2p}}$. ∇

Remark 30 *Observe that the proof gives the number of first-order formulae ϕ such that $hq(\phi) \leq n$, modulo the logical equivalence, with variables in V.*

Proposition 31 *For every first-order formula ϕ there exists a first-order formula*

$$Q_1 x_1 \ldots Q_n x_n \psi$$

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

1.3 Ehrenfeucht-Fraïssé Games

Let u, v be two $\{\}$-marked words and n an integer. The Ehrenfeucht-Fraïssé games are two-player games. Let $\mathfrak{A}$ and $\mathfrak{B}$ denote these two players. $\mathfrak{A}$ tries to prove that u and v do not satisfy the same atomic formulae, while $\mathfrak{B}$ tries to displease his opponent. Each player has n pebbles, labeled $z_1, \ldots, z_n$. $\mathfrak{A}$ plays first: he 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. $\mathfrak{B}$ plays his pebble z_1 on the other marked word, and so on. The game ends when the two players have no more pebbles. $\mathfrak{A}$ has won the game if there exists an atomic formula with free variables in $\{z_1, \ldots, z_n\}$ that satisfies one of the two obtained $\{z_1, \ldots, z_n\}$-marked words but not the other, otherwise $\mathfrak{B}$ has won. We say that a player has a **winning strategy** if he wins the game, independently of what his opponent plays.

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

Proposition 32 *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 $\mathfrak{B}$ has a winning strategy on (u,v) using n pebbles, $u \not\sim_n v$ otherwise.

Proposition 33 *$u \sim_n v$ 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 34 *Let n be an integer. Then $\sim_n$ has a finite number of equivalence classes.*

Proposition 35 *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 $x_1x_2 \sim_n y_1y_2$.*

Proof The winning strategy of $\mathfrak{B}$ consists of partitioning the game in two parts: pebbles played on (x_1, y_1) and pebbles played on (x_2, y_2). $\mathfrak{B}$ just applies his winning strategies on each of the two parts. To prove that this strategy suffices for $\mathfrak{B}$ to win the game, assume he loses, i.e., $x_1x_2 \not\sim_n y_1y_2$. An atomic formula is verified in one marked-word (the marked-word built from x_1x_2, for example) and not in the other. Assume first that this atomic formula is $x < y$. If pebbles labeled x and y were both played in x_1 then the others pebbles labeled x and y were played in y_1, according to the strategy of $\mathfrak{B}$. Then $\mathfrak{A}$ has a winning strategy for the game (x_1, y_1) using n pebbles: it suffices for $\mathfrak{A}$ to play exactly like he did in the game (x_1x_2, y_1y_2) without playing the pebbles he 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 easily be generalized:

Proposition 36 *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 $x_0x_1x_2 \cdots \sim_n y_0xy_1y_2 \ldots$.*

Proof As for the previous proposition. ∇

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

Proposition 37 *Let n be an integer. For every $k \geq 2^n - 1$, $k \sim_n k + 1$.*

Proposition 38 *Let n be an integer. If $\alpha < \omega^{n+1} < \beta$, then*

1. $\alpha \not\sim_{2n+2} \omega^{n+1}$
2. $\alpha \not\sim_{2n+3} \beta$

Proposition 39 *Let n be an integer and α and β two ordinals such that $\alpha < \omega^{n+1} \leq \beta$. Then $\alpha \not\sim_{2n+3} \beta$.*

2 Star-free sets

We recall in this section the different definitions of star-free sets, classified by the length of words considered. We also recall, for each such class of words, the main theorem for star-free sets, which establishes the equivalence between the three formalisms to define sets of words: first-order logic, finite algebras and star-free expressions. The section ends with the formulation of the theorem for star-free sets of words of length less than ω^{n+1}, whose proof is the subject of the paper.

Definition 40 *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 41 ([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*
- $X = \mathcal{L}^{<\omega}(\phi)$ *for a first-order sentence ϕ.*

A similar result holds for sets of ω-words:

Definition 42 *Let A be an alphabet. The class $SF(A, \omega)$ of* ***star-free sets of ω-words*** *on A is the smallest set containing $\emptyset$ closed under finite union, complement with respect to A^{ω} and product on the left only by an element of $SF(A, <\omega)$.*

Theorem 43 ([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*
- $X = \mathcal{L}^{\omega}(\phi)$ *for a first-order sentence ϕ.*

And for sets of words of length less than ω^{n+1}:

Definition 44 *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 45 *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*
- $X = \mathcal{L}^{[1,\omega^{n+1}[}(\phi)$ *for a first-order sentence* ϕ.

The (constructive) proof of this theorem occupies all of the remainder of this paper.

Corollary 46 *Let A be an alphabet and n an integer. It is decidable whether a recognizable subset X of $A^{[1,\omega^{n+1}[}$ is star-free.*

3 From star-free sets to sentences

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

$$(a_0, \emptyset) \ldots (a_\alpha, \{x\}) \ldots (a_\beta, \{y\}) \ldots (\$, \emptyset) \models \phi_E \text{ iff } u[\alpha, \beta[\in E$$

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 $(\$, \emptyset)$ is $|u|$ in the left side 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 $E = \emptyset$ 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. It follows from proposition 38 that ω^{n+1} is definable by a first-order sentence $\phi_{\omega^{n+1}}$, that is to say, $\mathcal{L}(\phi_{\omega^{n+1}})$ is the set of words over A of length ω^{n+1}. From this sentence one can build a first-order formula $\phi'_{\omega^{n+1}}$ which has exactly two free variables x and y such that

$$(a_0, \emptyset) \ldots (a_\alpha, \{x\}) \ldots (a_\beta, \{y\}) \ldots (\$, \emptyset) \models \phi'_{\omega^{n+1}} \text{ iff } u[\alpha, \beta[\models \phi_{\omega^{n+1}}$$

It suffices to replace in $\phi_{\omega^{n+1}}$ each occurrence of $\exists z\psi$ (resp. $\forall z\psi$), where z is a variable and ψ a sub-formula of $\phi_{\omega^{n+1}}$, by $\exists_x^y z\ \psi$ (resp. $\forall_x^y z\ \psi$).

Since the words of length less than ω^{n+1} are those without any factor of length ω^{n+1} we have

$$\phi_{\neg E} \equiv x < y \wedge (\neg\phi_E) \wedge (\neg\exists_x^y z_1\ \exists_x^y z_2\ (\phi'_{\omega^{n+1}}\{x \leftarrow z_1\}\{y \leftarrow z_2\})) \wedge \neg\phi'_{\omega^{n+1}}$$

Thus, we have inductively build ϕ_E from a star-free set E. It remains to get rid of the two free variables x and y. Let $\phi'_E \equiv \exists z[(\forall x\ z \leq x) \wedge (\phi_E\{x \leftarrow z\})]$, 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 sub-formulae of the form $r < y$ by $r = r$, and $y < r$ 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

$$u \in E \text{ iff } u \models \phi''_E$$

4 From sentences to aperiodic ω^n-semigroups

Let n be a positive integer, A an alphabet and ϕ a first-order sentence. In this section we use games on words to prove that $\mathcal{L}^{[1,\omega^{n+1}[}(\phi)$ is recognizable by a finite aperiodic ω^n-semigroup. We will first describe a construction for a finite aperiodic ω^n-semigroup recognizing $\mathcal{L}^{[1,\omega^{n+1}[}(\phi)$. We shall next show that this construction is effective. Throughout the section h is $max(2n+1, hq(\phi))$.

4.1 Construction

Propositions 36 and 34 show that $A^{[1,\omega^{n+1}[}/\sim_h$ is a finite ω^n-semigroup, and proposition 33 showes that $A^{[1,\omega^{n+1}[}/\sim_h$ recognizes $\mathcal{L}^{[1,\omega^{n+1}[}(\phi)$ for any first-order formula of quantifier height at most h.

It remains to prove that $A^{[1,\omega^{n+1}[}/\sim_h$ is aperiodic, which is a direct consequence of the following proposition:

Proposition 47 *Let $n \in \mathbb{N}$ and $k = 2^n - 1$. For every word $y \in A^{[1,\omega^{n+1}[}$ then $y^{k+1} \sim_n y^k$.*

Proof As an immediate corollary of proposition 37 we have $a^{k+1} \sim_n a^k$ for $a \in A$. Let $y^{k+1} = y_1 y_2 \dots y_{k+1}$ and $y^k = y'_1 y'_2 \dots y'_k$ where $y_i = y'_i = y$ for every $1 \leq i \leq k$ and $y_{k+1} = y$. We consider that $\mathfrak{A}$ and $\mathfrak{B}$ plays 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. $\mathfrak{A}$ plays first in the second game. If he plays in y^{k+1} (the other case is similar) on y_i at relative position α then he also plays on the first game on a^{k+1} at position i. $\mathfrak{B}$ applies his winning strategy in the first game: he plays on a^k at position j. His winning strategy in the second game is to play on y'_j at relative position α. ∇

4.2 Effectivity

We now prove that the construction of a finite aperiodic ω^n-semigroup S isomorphic to $A^{[1,\omega^{n+1}[}/\sim_h$ is effective. We first show how to build the semigroup $\cup_{j \leq n} S_j$ by induction on n. We note by $\varphi : A^{[1,\omega^{n+1}[} \to A^{[1,\omega^{n+1}[}/\sim_h$ the natural morphism of ω^n-semigroup which associate to any element of $A^{[1,\omega^{n+1}[}$ its equivalence class in $A^{[1,\omega^{n+1}[}/\sim_h$.

Recall that the cardinal of $A^{[1,\omega^{n+1}[}/\sim_h$ is finite. Let M be an upper bound of this cardinal. According to remark 30 the cardinal of the class C of first-order sentences of quantifier height at most n is effectively known. Since each equivalence class D of $\sim_h$ is characterized by the subclass of C composed of all the sentences of C satisfying every word of D, M is at most $2^{|C|}$.

Let also

$$Y_j = \{x \in A^{[1,\omega^{j+1}[} || x| = \sum_{i=j}^{0} \omega^i a_i, a_j > 0, \text{if } 0 \le i \le j \text{ then } a_i < \omega,$$
$$\text{and } \sum_{i=0}^{j} a_i \le M\}$$

and

$$X_j = \{(x,P) |\ x \in Y_j, P = \{\phi |\ x \models \phi\}, \text{ and there is no } y \text{ such that } (y,P) \in X_j\}$$

In other words, Y_j is the set of words x such that $|x|$ verifies:

- $\omega^j \le |x| < \omega^{j+1}$,
- the Cantor normal form of $|x|$ written as a sum of terms of the form ω^k with $k \le j$ has at most M terms.

Note also that X_j is isomorphic to $Y_j / \sim_h$.

Proposition 13 shows that the set X_j is isomorphic to the set S_j. Informally speaking, in X_j each element s of S_j is represented by a pair (x,P) such that x is a word verifying $\varphi(x) = s$ and P is the set of all sentences of quantifier height less than or equal to h satisfied by every word y such that $\varphi(y) = s$.

If $i = 0$, using proposition 13, each class of $A^{[1,\omega[} / \sim_h$ contains a word of length less than or equal to M. Since the alphabet A is finite, all of those words can effectively be enumerated. Let x be one of these. Let V be a finite set of first-order variables such that $|V| = h$. According to corollary 29 we can enumerate all sentences ϕ such that $hq(\phi) \le h$ with variables names in V, modulo the logical equivalence. Since $|x|$ is finite one can effectively decide if $x \models \phi$. So the construction of X_0 is effective. Furthermore, X_0 can effectively be equipped with an associative product: if (x_1,P_1) and (x_2,P_2) are elements of X_0 then $(x_1,P_1)(x_2,P_2) = (x,P)$ where $(x,P) \in X_0$ and $x_1x_2 \models \phi$ iff $x \models \phi$ for any first-order sentence ϕ.

We now assume that X_j for every $j \le i$ can effectively be obtained, and we compute X_{i+1}. Let $s \in S_{i+1}$. Our first task is to find a word x such that $\varphi(x) = s$. Since there is no empty equivalence class, there exists a word y such that $\varphi(y) = s$. Using proposition 13, we can suppose that $y \in Y_{i+1}$. Let $|y| = \sum_{j=i+1}^{0} \omega^j a_j$, with $a_{i+1} > 0$, and $(y_r)_{0 < r \le \sum_{t=0}^{i+1} a_t}$ be the serie of factors of y defined by

$$y_z = y[\sum_{j=i+1}^{i+1-k} \omega^j a_j + \omega^{i+1-k-1} l, \sum_{j=i+1}^{i+1-k} \omega^j a_j + \omega^{i+1-k-1}(l+1)[$$

where z is a sum of a_j's in decreasing order of indices, and with as many terms as possible, plus a rest l, that is to say, $z = (\sum_{j=i+1-k}^{i+1} a_j) + l + 1$ where $-1 \le$

$k \leq i$, k as great as possible and $l < a_{i+1-k-1}$. According to proposition 14, there exists a word x_z such that $x_z \sim_h y_z$ for every $0 < z \leq \sum_{j=1}^{i+1} a_j$, and $x_z = x_{z,1}x_{z,2}^\omega$, where $x_{z,1}$ and $x_{z,2}$ are words already enumerated by induction hypothesis. There also exist finite words x_z such that $x_z \sim_h y_z$, and x_z is already enumerated too, for $\sum_{j=1}^{i+1} a_j < z \leq \sum_{j=0}^{i+1} a_j$. If $z = \sum_{j=1}^{i+1} a_j$ the word $x_1x_2 \ldots x_z x_{z+1} \ldots x_{z+a_0}$ is equivalent to y, and to x, and can effectively be constructed from words of X_j, where $j \leq i$. We now enumerate all sentences ϕ such that $hq(\phi) \leq h$. We have to decide whether or not $x \models \phi$ for such an x.

Proposition 48 *Let V be a finite set of finite words, and V' the closure of V under finite use of $\cdot$ and ω. We modify the rules of Ehrenfeucht-Fraïssé games on two marked words x and y built from words of V' in order to oblige the players to put their pebbles only on a finite number of finite areas of positions of y that change dynamically over the game. We note by $P(y)$ the set of such positions. If $Y = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a finite set of ordinals and β is an ordinal then $Y \uparrow \beta$ denotes the set $\{\alpha_1 + \beta, \alpha_2 + \beta, \ldots, \alpha_n + \beta\}$. $P(y)$ is defined inductively on the structure of y:*

- *If $y \in V$ then $P(y) = \{1, \ldots, |y|\}$*
- *If $y = y_1y_2$, where $y_1, y_2 \in V'$ then $P(y) = P(y_1) \cup (P(y_2) \uparrow |y_1|)$*
- *If $y = y_1^\omega$, where $y_1 \in V'$ and i is the smallest integer such that all pebbles already played on y are in the prefix y_1^i, then $P(y) = P(y_1^i) \cup \bigcup_{j=0}^{k-1} P(y_1) \uparrow |y_1^{i+j}|$ where k is the index of $\cup_{j=0}^{m} X_j$ if $|y_1| = \sum_{j=m}^{0} \omega^j a_j$.*

We write $x \simeq_n y$ if $\mathfrak{B}$ wins this restricted game on (x, y) with n pebbles. We claim that $y \simeq_n y$ and that if $x \simeq_n y$ then $x \models \phi$ iff $y \models \phi$ for any first-order formula ϕ of quantifier height at most n.

Proof The proof of $y \simeq_n y$ is by induction on the structure of y. In order to avoid heavy notations we say that a marked word $(a_1, V_1)(a_2, V_2) \ldots$ is a prefix (resp. a factor) of a word u if $a_1a_2 \ldots$ is a prefix (resp. a factor) of u.

If y is issued from V then $y \simeq_n y$ iff $y \sim_n y$, so $y \simeq_n y$.

If $y = y_1y_2$ by induction hypothesis $y_1 \simeq_n y_1$ and $y_2 \simeq_n y_2$, and it is easy to prove that $y_1y_2 \simeq_n y_1y_2$ using the same argument as in the proof of proposition 35.

Let $y = y_1^\omega$. We denote by $y|$ the marked word which is in the left side of the $\simeq_n$ sign and $|y$ the one on the right side. We prove that at j^{th} turn $\mathfrak{B}$ can divide the game in m partitions, that is to say, m "sub-games" denoted by $(y|_1, |_1y)$, $(y|_2, |_2y)$, $\ldots$, $(y|_m, |_my)$, such that $m \leq j$, $y|_1y|_2 \ldots y|_m$ is a prefix of $y|$, $|_1y|_2y \ldots |_my$ is a prefix of $|y$ and $y|_i \simeq_n |_iy$ for every $1 \leq i \leq m$. Assume that j turns have been played and that the game is partitioned as explain below. If $\mathfrak{A}$ plays in an existing part $y|_i$ (resp. $|_iy$) then $\mathfrak{B}$ plays his winning strategy on $(y|_i, |_iy)$. If $\mathfrak{A}$ plays on $|y$ on the right of $|_my$, then the obtained marked word can be written $|_1y \ldots |_myy y^k y^\omega$ where $\mathfrak{A}$ played his pebble on the factor y^k at relative position α. The answer of $\mathfrak{B}$ is to play on $y|$ at position $|y|_1 \ldots y|_m| + \alpha$.

We say that $|_{m+1}y$ is the factor y^k of $|y$ in which $\mathfrak{A}$ has just placed his pebble and $y|_{m+1}$ the factor y^k of $y|$ in which $\mathfrak{B}$ answered according to his winning strategy. Assume now that $\mathfrak{A}$ places his pebble on $y|$ on the right of $y|_m$ at relative position α and let l be the smallest integer such that $\alpha \leq |y^{k+l}|$. Since it is not difficult to modify the proof of proposition 47 to show that $y^{k+1} \simeq_n y^k$ then $y^{k+l} \simeq_n y^k$. We say that $y|_{m+1}$ is the factor y^{k+l} of $y|$ in which $\mathfrak{A}$ has just placed his pebble and $|_{m+1}y$ the factor y^k of $|y$ in which $\mathfrak{B}$ answered according to his winning strategy. This ends the global winning strategy of $\mathfrak{B}$.

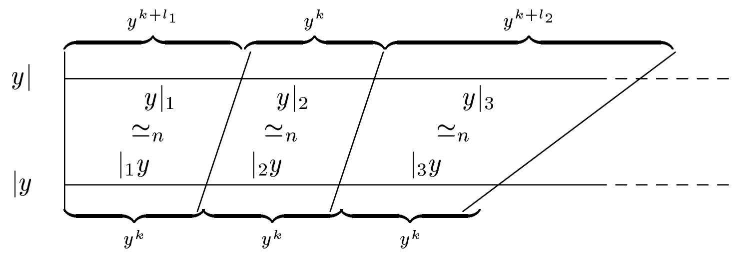

Figure 1: The winning strategy of $\mathfrak{B}$

We now prove the second part of the claim, if $x \simeq_n y$ then $x \models \phi$ iff $y \models \phi$ for any first-order formula ϕ of quantifier height at most n, by induction on n. If $n = 0$ by definition of the game x and y satisfy the same atomic formulae. Assume that the claim is true for $n-1$, that $x \simeq_n y$ and that there exists a first-order formula ϕ such that $hq(\phi) = k \leq n$, $x \models \phi$ and $y \not\models \phi$, that is to say $y \models \neg\phi$. We can suppose that $\phi = \exists z \psi$ (the other case is similar), where ψ is a first-order formula such that $hq(\psi) = k - 1$. We put the pebble z on x such that the obtained marked word x' verifies $x' \models \psi$. Wherever we put a pebble z on y, the obtained marked word y' verifies $y' \models \neg\psi$. Since $x \simeq_n y$, then $x' \simeq_{n-1} y'$, so by induction hypothesis x' and y' satisfy exactly the same first-order formulae of quantifier height at most $n-1$, which is a contradiction. ∇

The previous proposition shows that in order to answer the question "$x \models Q_1x_1 \ldots Q_mx_m\psi$", where for every $1 \leq i \leq m$ Q_i is a quantifier, x_i is a variable, and ψ is a first-order formula such that $hq(\psi) = 0$ and $FV(\psi) = \{x_1, \ldots, x_m\}$, it suffices to enumerate all possible positioning of pebbles $x_1, \ldots, x_m$ in a finite number of finite factors of x which depends only on the structure of x and on an integer k effectively computable by induction hypothesis, and to verify if the obtained $\{x_1, \ldots, x_m\}$-marked word satisfies ψ, which is effective.

This proves that the construction of a semigroup $\cup_{j \leq n} X_j$ isomorphic to $\cup_{j \leq n} S_j$ is effective. We now have to equip this semigroup with ω operators. Let $s \in X_j$ with $j < n$. Assume $s = (x, P)$. Since for any first-order formula ϕ the question "$x^\omega \models \phi$" is decidable, we can effectively find $s' = (y, P') \in X_{j+1}$ such that x^ω and y satisfy exactly the same first-order formulae of quantifier height less or equal than h. The obtained algebraic structure is isomorphic to S.

5 From first-order sentences to star-free sets

Let ϕ be a first-order sentence, A an alphabet and n an integer. In the previous section we showed that the set of words $u \in A^{[1,\omega^{n+1}[}$ such that $u \models \phi$ is a finite union of equivalence classes for $\sim_{max(2n+1,hq(\phi))}$. We now prove that each such class is in $SF(A,[1,\omega^{n+1}[)$. Since the star-free sets are closed under finite union, it follows that the set of words $u \in A^{[1,\omega^{n+1}[}$ such that $u \models \phi$ is in $SF(A,[1,\omega^{n+1}[)$.

If $x \in A^{[1,\omega^{n+1}[}$ we denote by $< x >_n$ the equivalence class of x for Ehrenfeucht-Fraïssé games in n turns. The statement of the following proposition is from Ladner.

Proposition 49 *Let m, n be two integers and x a word such that $0 < |x| < \omega^m$. Then*

$$< x >_n = (\bigcap_{(u,a,v)\in P} < u >_{n-1} a < v >_{n-1}) \setminus (\bigcup_{(u,a,v)\in Q} < u >_{n-1} a < v >_{n-1})$$

where $P = \{(u,a,v) \in A^{<\omega^m} \times A \times A^{<\omega^m} | uav = x\}$ and $Q = \{(u,a,v) \in A^{<\omega^m} \times A \times A^{<\omega^m}$ such that for any factorization $x = u'a'v'$ then $u \not\sim_{n-1} u'$ or $a \neq a'$ or $v \not\sim_{n-1} v'\}$.

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

Lemma 50 *Let x and y be two words such that $x \not\sim_n y$. 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 $x_1 \not\sim_{n-1} y_1$ or $x_2 \not\sim_{n-1} y_2$ or $a \neq b$.*

Proof We denote by x^i and y^i the index of letters of x and y played at turn i. In his winning strategy, $\mathfrak{A}$ plays his first pebble, and $\mathfrak{B}$ answers, defining the factorizations of x and y of the statement of the lemma. If $\mathfrak{B}$ could not have played 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 $i, j \leq n$ 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 to the advantage of $\mathfrak{A}$, since $\mathfrak{B}$ can always do the same, we can assume that all his 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 case of 2 is similar. ∇

We can now prove the proposition:

Proof Let $y \in< x >_n$. We start by proving that for any factorization $x = uav$ of x, 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 $u' \not\sim_{n-1} u$ or $v' \not\sim_{n-1} v$. It follows that $\mathfrak{A}$ has a winning strategy on the words x and y in n turns: he put his first pebble on a on x, and $\mathfrak{B}$ answers on y. If he cannot play on a letter a, he will loose in only one turn. Otherwise, he will factorize y in $u'av'$, and since either $u' \not\sim_{n-1} u$ or $v' \not\sim_{n-1} v$, $\mathfrak{A}$ just has to apply his winning strategy in $n-1$ turns either on the left or on the right of the first turn. We now show that there do not exists u, a, and v such that for any factorization $x = u'av'$ we have $y \in< u >_{n-1} a < v >_{n-1}$ 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 $\mathfrak{A}$ 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 $a' \neq a$, $\mathfrak{A}$ wins in only one turn. Otherwise, since $u'' \sim_{n-1} u \not\sim_{n-1} u'$ or $v'' \sim_{n-1} v \not\sim_{n-1} v'$, $\mathfrak{A}$ applies his 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 side 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 (wrongly) that $x \not\sim_n y$. $\mathfrak{A}$ plays his first pebble following his winning strategy, $\mathfrak{B}$ answers. If $\mathfrak{A}$ played on x, he chose a factorization of $x = uav$ such that he wins for any factorization of $y = u'a'v'$ determined by the first play of $\mathfrak{B}$. If $a \neq a'$, $\mathfrak{A}$ 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 exists 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 side of the equality, which is a contradiction. If $\mathfrak{A}$ played on y, he 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 side of the equality, which contradicts the fact that y is on the right side of the equality. ∇

6 From aperiodic ω^n-semigroups to star-free sets

Let A be an alphabet, n an integer and S a finite aperiodic ω^n-semigroup. In this section we prove that a set X recognized by a morphism $\varphi : A^{[1,\omega^{n+1}[} \to S$ of ω^n-semigroups is in $SF(A,[1,\omega^{n+1}[)$.

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

Our proof is by induction on i. Let us start the induction. If $i=0$ the result directly follows from proposition 20 and theorem 41. We now suppose that the result is true for $0 \leq i \leq n-1$ and we prove it for $i+1$.

Lemma 51 *If* $m \in S_{i+1}$ *then* $\varphi^{-1}(m) \cap A^{\omega^{i+1}} \in SF(A, [1, \omega^{n+1}[)$.

Proof According to proposition 14,

$$\varphi^{-1}(m) \cap A^{\omega^{i+1}} = \bigcup_{(s,e) \in P} \varphi^{-1}(s) \varphi^{-1}(e)^{\omega}$$

with $P = \{(s,e) \in S_i \times S_i \mid se = s,\ e^2 = e \text{ and } se^{\omega} = m\}$. Using corollary 16 we obtain

$$\varphi^{-1}(m) \cap A^{\omega^{i+1}} = \bigcup_{(s,e) \in P} \varphi^{-1}(s) \overrightarrow{\varphi^{-1}(e) \cdot \varphi^{-1}(e)}$$

Using the induction hypothesis, $\varphi^{-1}(s)$ and $\varphi^{-1}(e)$ are both in $SF(A, [1, \omega^{n+1}[)$, and using results of section 3 equivalent to first-order formulae ϕ_s and ϕ_e having exactly two free variables x and y such that (the same holds for ϕ_e):

$$(a_0, \emptyset) \ldots (a_\alpha, \{x\}) \ldots (a_\beta, \{y\}) \ldots (\$, \emptyset) \models \phi_s \text{ iff } u[\alpha, \beta[\in \varphi^{-1}(s)$$

where $a_0 a_1 \cdots = u$ and $\$$ is a new letter which is not in A that has been concatenated to u. One can understand this new letter has a marker to the end of u. The formula

$$\phi \equiv \forall r\ x < r \rightarrow \exists l\ \exists f\ r \leq l \wedge l < f \wedge \phi_e\{y \leftarrow l\} \wedge (\phi_e\{x \leftarrow l\}\{y \leftarrow f\})$$

has only one free variable x, and verifies $(a_0, \emptyset) \ldots (a_\alpha, \{x\}) \cdots \models \phi$ iff $u[\alpha, |u|[\in \overrightarrow{\varphi^{-1}(e) \cdot \varphi^{-1}(e)}$. Using arguments of section 3 one can build a sentence ϕ_m such that for any word $u \in A^{[1, \omega^{n+1}[}$, $u \models \phi_m$ iff $u \in \varphi^{-1}(m) \cap A^{\omega^{i+1}}$. We have $\mathcal{L}(\phi_m) = \mathcal{L}^{[1, \omega^{n+1}[}(\phi_m)$. According to results of section 5 $\mathcal{L}^{[1, \omega^{n+1}[}(\phi_m) \in SF(A, [1, \omega^{n+1}[)$. ∇

We now return to our main proof, adapting the proof of theorem 41 from [Per90]. Assume that $p \in S_{i+1}$. We introduce a new notation: if $s \in S_j$, we denote $\varphi^{-1}(s) \cap A^{\omega^j}$ by $\overline{\varphi^{-1}(s)}$. If S does not possesses a neutral element we add it: since $1^2 = 1$ this does not change the aperiodicity of S nor $\varphi^{-1}(s)$ for every $s \in S$. We start by showing that

$$\varphi^{-1}(p) = (UA^{\omega^{<n+1}} \cap A^{\omega^{<n+1}} V) \backslash (A^{\omega^{<n+1}} W A^{\omega^{<n+1}}) \tag{5}$$

where

$$U = (\bigcup_{\substack{s \in S \\ sS = pS}} \overline{\varphi^{-1}(s)}) \cup (\bigcup_{\substack{r,s \in S \\ rsS = pS \neq rS}} \varphi^{-1}(r) \overline{\varphi^{-1}(s)})$$

$$V = (\bigcup_{\substack{s \in S \\ Ss = Sp}} \overline{\varphi^{-1}(s)}) \cup (\bigcup_{\substack{r,s \in S \\ Ssr = Sp \neq Sr}} \overline{\varphi^{-1}(s)} \varphi^{-1}(r))$$

and

$$W = (\bigcup_{\substack{s\in S\\ p\notin SsS}} \overline{\varphi^{-1}(s)}) \cup (\bigcup_{\substack{s,t\in S\\ p\in SsS\cap StS\\ p\notin SstS}} \overline{\varphi^{-1}(s)\varphi^{-1}(t)})$$

$$\cup(\bigcup_{\substack{r,s,t\in S\\ p\in SrsS\cap SstS\\ p\notin SrstS}} \overline{\varphi^{-1}(r)}\varphi^{-1}(s)\overline{\varphi^{-1}(t)})$$

and we end by showing that $\varphi^{-1}(p) \in SF(A,[1,\omega^{n+1}[)$ by proving that U, V and W belong to $SF(A,[1,\omega^{n+1}[)$ using a decreasing induction on $|SpS|$. The final result directly follows since $SF(A,[1,\omega^{n+1}[)$ is closed under finite boolean operations and product.

We first show the inclusion from left to right of 5. Let $x \in \varphi^{-1}(p)$ and w a left factor of x such that $\varphi(w) \in pS$ and there does not exists a left factor w' of x such that $\varphi(w') \in pS$ and $|w'| < |w|$. If $|w| = \omega^m$ for an integer m then $w \in \overline{\varphi^{-1}(\varphi(w))}$, so $w \in U$, and $x \in UA^{<\omega^{n+1}}$. Else we write $|w|$ in Cantor normal form: $|w| = \omega^{m_1}\cdot n_1 + \omega^{m_2}\cdot n_2 + \cdots + \omega^{m_k}\cdot n_k$ and we factorize w in yz such that $|z| = \omega^{m_k}$. Since $|y| < |w|$ it follows that $\varphi(y) \notin pS$, so $x \in UA^{<\omega^{n+1}}$. The proof that $x \in A^{<\omega^{n+1}}V$ is similar, but this time we force the length of y (instead of z) to be ω^{m_1} (instead of ω^{m_k}). If $x \in A^{<\omega^{n+1}}WA^{<\omega^{n+1}}$ we cannot have $\varphi(x) = p$, so the inclusion from left to right is proved.
Now let x be in the right side of (5). Since $x \in UA^{<\omega^{n+1}}$ and $\varphi(U) \subseteq pS$ then $\varphi(x) \in pS$. We prove similarly that $\varphi(x) \in Sp$. Using proposition 22 it suffices to show that $p \in S\varphi(x)S$ to obtain our inclusion. Assume it is false and let w be a factor of x such that $p \notin S\varphi(w)S$ and there does not exists another factor w' of x that verifies $|w'| < |w|$ and $p \notin S\varphi(w')S$. If $|w| = \omega^m$ for an integer m then $w \in \overline{\varphi^{-1}(\varphi(w))}$ and since $p \notin S\varphi(w)S$ it follows that $w \in W$, which is a contradiction. Else we write $|w|$ in Cantor normal form: $|w| = \omega^{m_1}\cdot n_1 + \omega^{m_2}\cdot n_2 + \cdots + \omega^{m_k}\cdot n_k$ and we factorize w in $w_1w_2w_3$, with $|w_1| = \omega^{m_1}$, $|w_3| = \omega^{m_k}$ and w_2 possibly empty. If w_2 is not the empty word then $w \in \overline{\varphi^{-1}(\varphi(w_1))}\varphi^{-1}(w_2)\overline{\varphi^{-1}(\varphi(w_3))}$, $p \notin S\varphi(w_1)\varphi(w_2)\varphi(w_3)S$, but $p \in S\varphi(w_1)\varphi(w_2)S$ and $p \in S\varphi(w_2)\varphi(w_3)S$, so $w \in W$. We obtain the same contradiction if w_2 is the empty word. This ends the proof of (5).

It remains to prove that $\varphi^{-1}(p) \in SF(A,[1,\omega^{n+1}[)$. First observe that for every $x \in S$, $SxS \subseteq S1S$. Since S is aperiodic, using proposition 21, $1 = xyz = xy1z = 1z = z = x1y = x = y$, so if $x \neq 1$ then $|S1S| > |SxS|$. Using the same kind of argument, and since $1 \in S_0$, it follows that $\varphi^{-1}(1) = A^{[1,\omega[}\backslash A^{<\omega}(\cup_{\varphi(a)\neq 1} a)A^{<\omega} \in SF(A,[1,\omega[)$ where every $a \in A$. Assume now that $p \neq 1$. We begin by showing that $U \in SF(A,[1,\omega^{n+1}[)$ using the induction hypothesis. According to lemma 51, for every $s \in S$, $\overline{\varphi^{-1}(s)} \in SF(A,[1,\omega^{n+1}[)$. Now let r, s and p be elements of S such that $rsS = pS$ and $rS \neq pS$. There exists $x \in S$ such that $p = rsx$, so $SpS \subseteq SrS$ and $pS \subseteq rS$. If $SpS = SrS$ there exist $y, z \in S$ such that $r = ypz = y(rsx)z = pz$ according to proposition 21, so $rS \subseteq pS$ and $rS = pS$, which is a contradiction.

The proof of $V \in SF(A, [1, \omega^{n+1}[)$ is symmetrical. Now let $p \in SrsS \cap SstS$ such that $p \notin SrstS$. There exist $a, b, c, d \in S$ such that $p = arsb = cstd$, so $MpM \subseteq MsM$. If $MpM = MsM$ then $s = xpy$ for some $p, y \in S$, and using proposition 21 $s = xarsby = xars$, so $p = cxarstd$, which is a contradiction, so $W \in SF(A, [1, \omega^{n+1}[)$.

7 Examples

We give here two examples of recognizable sets. The first one is not star-free, and the second is star-free.

Example 52 *Let* $A = \{a, b\}$ *and* $S = (\{a, b, ab, ba, aba, 0, 1\}, \{a^\omega, ab^\omega, ba^\omega, 0'$ $, a^\omega a, ab^\omega a, ba^\omega a\})$ *be the* ω^1*-semigroup with the product defined by*

	a	b	ab	ba	aba	0	1	a^ω	ab^ω	ba^ω	$0'$	$a^\omega a$	$ab^\omega a$	$ba^\omega a$
a	1	ab	b	aba	ba	0	a	a^ω	ba^ω	ab^ω	$0'$	$a^\omega a$	$ba^\omega a$	$ab^\omega a$
b	ba	b	0	ba	0	0	b	ba^ω	$0'$	ba^ω	$0'$	$ba^\omega a$	$0'$	$ba^\omega a$
ab	aba	ab	0	aba	0	0	ab	ab^ω	$0'$	ab^ω	$0'$	$ab^\omega a$	$0'$	$ab^\omega a$
ba	b	0	b	0	ba	0	ba	ba^ω	ba^ω	$0'$	$0'$	$ba^\omega a$	$ba^\omega a$	$0'$
aba	ab	0	ab	0	aba	0	aba	ab^ω	ab^ω	$0'$	$0'$	$ab^\omega a$	$ab^\omega a$	$0'$
0	0	0	0	0	0	0	0	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$
1	a	b	ab	ba	aba	0	1	a^ω	ab^ω	ba^ω	$0'$	$a^\omega a$	$ab^\omega a$	$ba^\omega a$
a^ω	$a^\omega a$	a^ω	$0'$	$a^\omega a$	$0'$	$0'$	a^ω	a^ω	$0'$	a^ω	$0'$	$a^\omega a$	$0'$	$a^\omega a$
ab^ω	$ab^\omega a$	ab^ω	$0'$	$ab^\omega a$	$0'$	$0'$	ab^ω	ab^ω	$0'$	ab^ω	$0'$	$ab^\omega a$	$0'$	$ab^\omega a$
ba^ω	$ba^\omega a$	ba^ω	$0'$	$ba^\omega a$	$0'$	$0'$	ba^ω	ba^ω	$0'$	ba^ω	$0'$	$ba^\omega a$	$0'$	$ba^\omega a$
$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$
$a^\omega a$	a^ω	$0'$	a^ω	$0'$	$a^\omega a$	$0'$	$a^\omega a$	a^ω	a^ω	$0'$	$0'$	$a^\omega a$	$a^\omega a$	$0'$
$ab^\omega a$	ab^ω	$0'$	ab^ω	$0'$	$ab^\omega a$	$0'$	$ab^\omega a$	ab^ω	ab^ω	$0'$	$0'$	$ab^\omega a$	$ab^\omega a$	$0'$
$ba^\omega a$	ba^ω	$0'$	ba^ω	$0'$	$ba^\omega a$	$0'$	$ba^\omega a$	ba^ω	ba^ω	$0'$	$0'$	$ba^\omega a$	$ba^\omega a$	$0'$

and the ω *operator by*

a	b	ab	ba	aba	0	1
a^ω	ba^ω	$0'$	$0'$	ab^ω	$0'$	a^ω

Let $\varphi : A^{[1,\omega^2[} \to S$ *be the morphism of* ω^1*-semigroups defined by* $\varphi(a) = a$ *and* $\varphi(b) = b$*, and* $X = \{1, b, a^\omega, ba^\omega\}$*. Then* S *recognizes* $\varphi^{-1}(X)$*, the set of non-empty words on* A *of length less than* ω^2 *having an even number (that is to say, there exists an ordinal* α *such that this number is equal to* $2 \cdot \alpha$*) of consecutive "a" letter. Furthermore,* S *is the syntactic* ω^1*-semigroup of this set. Since* $a^2 = 1$ *and* $1a = a$ *then* S *is not aperiodic, and* $\varphi^{-1}(X)$ *is not star-free or definable by a first-order sentence.*

Example 53 *Let* $A = \{a, b\}$ *and* $S = (\{a, b, 0, ab, ba\}, \{0', (ab)^\omega, (ba)^\omega, (ab)^\omega a,$

$(ba)^\omega a\})$ be the ω^1-semigroup with the product defined by

	a	b	0	ab	ba	$0'$	$(ab)^\omega$	$(ba)^\omega$	$(ab)^\omega a$	$(ba)^\omega a$
a	0	ab	0	0	a	$0'$	$0'$	$(ab)^\omega$	$0'$	$(ab)^\omega a$
b	ba	0	0	b	0	$0'$	$(ba)^\omega$	$0'$	$(ba)^\omega a$	$0'$
0	0	0	0	0	0	$0'$	$0'$	$0'$	$0'$	$0'$
ab	a	0	0	ab	0	$0'$	$(ab)^\omega$	$0'$	$(ab)^\omega a$	$0'$
ba	0	b	0	0	ba	$0'$	$0'$	$(ba)^\omega$	$0'$	$(ba)^\omega a$
$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$	$0'$
$(ab)^\omega$	$(ab)^\omega a$	$0'$	$0'$	$(ab)^\omega$	$0'$	$0'$	$(ab)^\omega$	$0'$	$(ab)^\omega a$	$0'$
$(ba)^\omega$	$(ba)^\omega a$	$0'$	$0'$	$(ba)^\omega$	$0'$	$0'$	$(ba)^\omega$	$0'$	$(ba)^\omega a$	$0'$
$(ab)^\omega a$	$0'$	$(ab)^\omega$	$0'$	$0'$	$(ab)^\omega a$	$0'$	$0'$	$(ab)^\omega$	$0'$	$(ab)^\omega a$
$(ba)^\omega a$	$0'$	$(ba)^\omega$	$0'$	$0'$	$(ba)^\omega a$	$0'$	$0'$	$(ba)^\omega$	$0'$	$(ba)^\omega a$

and the ω operator by

a	b	0	ab	ba
$0'$	$0'$	$0'$	$(ab)^\omega$	$(ba)^\omega$

Let $\varphi : A^{[1,\omega^2[} \to S$ be the morphism of ω^1-semigroups defined by $\varphi(a) = a$ and $\varphi(b) = b$, and $X = \{ab, (ab)^\omega\}$. Then S recognizes $\varphi^{-1}(X)$, the set of non-empty words on A of length less than ω^2 formed by repetitions of ab. One can verify that S is aperiodic of index 2, and is the syntactic ω^1-semigroup of this set. So we have $\varphi^{-1}(X) \in SF(A, [1, \omega^2[)$:

$$\varphi^{-1}(X) = A^{[1,\omega^2[} \backslash (LbA^{<\omega^2} \cup A^{<\omega^2} a \cup A^{<\omega^2} aaA^{<\omega^2} \cup A^{<\omega^2} bbA^{<\omega^2})$$

where $L = A^{<\omega^2} \backslash (A^{<\omega^2} A)$ is the set of limit words union the singleton which contains only the empty word. A first-order sentence ϕ such that $\mathcal{L}(\phi) = \varphi^{-1}(X)$ is

$$\begin{aligned}
\phi =& \phi_{[1,\omega^2[} \wedge \Big(\forall x \big(\neg \exists y \ x = y + 1 \to R_a(x)\big)\Big) \\
& \wedge \Big(\forall x R_a(x) \to \big(\exists y \ y = x + 1 \wedge R_b(y)\big)\Big) \\
& \wedge \Big(\forall x \big(\exists y \ x < y \wedge R_b(x)\big) \to \exists z \ z = x + 1 \wedge R_a(z)\Big)
\end{aligned}$$

where $\phi_{[1,\omega^2[}$ is a first-order sentence saying that every word is non-empty and of length less than ω^2.

8 Conclusion

Büchi proved that second-order sentences used for defining set of words on ordinal are equivalent to a certain class A of automata. If we restrict the set of words recognized by automata of class A to words of length less than ω^{n+1}, this class is equivalent to another one B, studied by Choueka in [Cho78]. We proved in [Bed98a] that automata of class B are equivalent to finite ω^n-semigroups. The constructions to obtain

- an automaton of class A from a second-order sentence,
- an automaton of class B from an automaton of class A,
- a finite ω^n-semigroup from an automaton of class B,
- a star-free expression from a finite aperiodic ω^n-semigroup,
- a first-order sentence from a star-free expression

are effective. As an immediate consequence:

Corollary 54 *Let ϕ be a second-order sentence and n an integer. It is decidable whether there exists a first-order sentence ψ such that $\mathcal{L}^{[1,\omega^{n+1}[}(\phi)=\mathcal{L}^{[1,\omega^{n+1}[}(\psi)$. Furthermore, if ψ exists, it can effectively be built from ϕ.*

Ideas in subsection 4.2 for the decision procedure of $x \models \phi$, where ϕ is a first-order sentence, are a generalization of those in [Lad77]. The ideas can also be generalized to second-order sentences to obtain another proof of Büchi's theorem for words of length less than ω^n.

Syntactic semigroups and logics over finite and ω-words have been widely studied to obtain hierarchies over rational languages (see for example [Pin96]). This can also be generalized to words of length less than ω^n.

Finally, we introduced in [Bed98b, BC98] an algebraic structure adapted to the study of words of any denumerable length. The main theorem of this paper (theorem 45) is extended to sets of words of denumerable length in [Bed].

Acknowledgments

The author thanks Dominique Perrin for directing the researches and Richard E. Ladner for the statement of proposition 49.

References

[BC98] Nicolas Bedon and Olivier Carton. An Eilenberg theorem for words on countable ordinals. In Cláudio L. Lucchesi and Arnaldo V. Moura, editors, *Latin'98: Theoretical Informatics*, volume 1380 of *Lecture Notes in Computer Science*, pages 53–64, Campinas, Brazil, April 1998. Springer.

[Bed] Nicolas Bedon. Logic over words on denumerable ordinals. To appear.

[Bed96] Nicolas Bedon. Finite automata and ordinals. *Theoretical Computer Science*, 156:119–144, March 1996.

[Bed98a] Nicolas Bedon. Automata, semigroups and recognizability of words on ordinals. *International Journal of Algebra and Computation*, 8(1):1–21, February 1998.

[Bed98b] Nicolas Bedon. *Langages reconnaissables de mots indexés par des ordinaux.* PhD thesis, Université de Marne-la-Vallée, Cité Descartes, 5, boulevard Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France, January 1998. http://www-igm.univ-mlv.fr/~bedon/Recherche/These/these.shtml.

[Büc60] J. Richard Büchi. Weak second-order arithmetic and finite automata. *Zeit. Math. Logik. Grund. Math.*, 6:66–92, 1960.

[Büc62] J. Richard Büchi. On a decision method in the restricted second-order arithmetic. In *Logic, Methodology and Philosophy of science: Proc. Intern. Congr.*, pages 1–11, Stanford, California, 1962. Stanford University Press.

[Büc64] J. Richard Büchi. Transfinite automata recursions and weak second order theory of ordinals. In *Proc. Int. Congress Logic, Methodology, and Philosophy of Science*, pages 2–23, Jerusalem, 1964. North-Holland Publ. Co. Invited address.

[Cho78] Yaacov Choueka. Finite automata, definable sets, and regular expressions over ω^n-tapes. *Journal of Computer and System Sciences*, 17:81–97, 1978.

[Ehr61] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fund. Math.*, 49:129–141, 1961. [MR 23, #3666].

[Kle56] Stephen C. Kleene. Representation of events in nerve nets and finite automata. In Shannon and McCarthy, editors, *Automata studies*, pages 3–42, Princeton, New Jersey, 1956. Princeton University Press.

[Lad77] Richard E. Ladner. Application of model theoretic games to discrete linear orders and finite automata. *Information and Control*, 33:281–303, 1977.

[McN66] Robert McNaughton. Testing and generating infinite sequences by a finite automaton. *Information and Control*, 9:521–530, 1966.

[MP71] Robert McNaughton and Seymour Papert. *Counter free automata.* MIT Press, Cambridge, MA, 1971.

[Péc86a] Jean-Pierre Pécuchet. Etude syntaxique des parties reconnaissables de mots infinis. *Lecture Notes in Computer Science*, 226:294–303, 1986.

[Péc86b] Jean-Pierre Pécuchet. Variétés de semigroupes et mots infinis. *Lecture Notes in Computer Science*, 210:180–191, 1986.

[Per84] Dominique Perrin. Recent results on automata and infinite words. In M. P. Chytil and V. Koubek, editors, *Mathematical foundations of computer science*, volume 176 of *Lecture Notes in Computer Science*, pages 134–148, Berlin, 1984. Springer.

[Per90] Dominique Perrin. *Handbook of theoretical computer science*, volume B, chapter Finite automata, pages 1–53. Elsevier, 1990.

[Pin96] Jean-Eric Pin. Logic, semigroups and automata on words. *Ann. Math. Artificial Intelligence*, 16(1-4):343–384, 1996.

[PP86] Dominique Perrin and Jean-Eric Pin. First order logic and star-free sets. *Journal of Computer and System Sciences*, 32:393–406, 1986.

[PP97] Dominique Perrin and Jean-Eric Pin. *Mots infinis.* 1997. To appear (LITP report 97-04), http://www.liafa.jussieu.fr/~jep/Resumes/InfiniteWords.html. English version: "Infinite words".

[Ros82] J. G. Rosenstein. *Linear ordering.* Academic Press, New York, 1982.

[Sch65] Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.

[Sie65] Waclaw Sierpiński. *Cardinal and ordinal numbers.* Polish Scientific Publishers, Varsovie, 1965.

[Str94] Howard Straubing. *Finite automata, formal logic and circuit complexity.* Birkhäuser, 1994.

[Tho79] Wolfgang Thomas. Star free regular sets of ω-sequences. *Information and Control*, 42:148–156, 1979.

[Wil91] Thomas Wilke. An Eilenberg theorem for ∞-languages. In *Automata, Languages and Programming: Proc. of 18th ICALP Conference*, pages 588–599. Springer, 1991.