Acta Informatica 31, 539–557 (1994)

Well quasi-orders and regular languages

Aldo de Luca [1] **and Stefano Varricchio** [2]

[1] Dipartimento di Matematica, Università di Roma "La Sapienza", Piazzale A. Moro, 2, I-00185 Roma, Italy, and Istituto di Cibernetica del CNR, Arco Felice, Napoli, Italy
[2] Dipartimento di Matematica, Università di Catania, Catania, Italy

Received May 20, 1993 / November 29, 1993

Abstract. An extension of Myhill's theorem of automata theory, due to Ehrenfeucht et al. [4] shows that a subset X of a semigroup S is recognizable if and only if X is closed with respect to a monotone well quasi-order on S. In this paper we prove that a similar extension of Nerode's theorem is not possible by showing that there exist non-regular languages on a binary alphabet which are closed with respect to a right-monotone well quasi-order. We give then some additional conditions under which a set $X \subseteq S$ closed with respect to a right-monotone well quasi-order becomes recognizable. We prove the following main proposition: A subset X of S is recognizable if and only if X is closed with respect to two well quasi-orders $\leqq_1$ and $\leqq_2$ which are right-monotone and left-monotone, respectively. Some corollaries and applications are given. Moreover, we consider the family $\mathscr{F}$ of all languages which are closed with respect to a right-monotone well quasi-order on a finitely generated free monoid. We prove that $\mathscr{F}$ is closed under rational operations, intersection, inverse morphisms and direct non-erasing morphisms. This implies that $\mathscr{F}$ is closed under faithful rational transductions. Finally we prove that the languages in $\mathscr{F}$ satisfy a suitable 'pumping' lemma and that $\mathscr{F}$ contains languages which are not recursively enumerable.

0. Introduction

Regular languages play a central role in theoretical computer science, since they can be recognized by *finite automata*. It is well known that regular languages admit several characterizations based on different concepts such as: *finite automata, right-grammars, syntactic semigroups, logical formulas, rational expressions, iteration properties, permutation properties, well quasi-orders* (cf. [5], [4], [10], [11], [12] and references therein).

In this paper we shall refer to well quasi-orders. There exist various characterizations of this concept, which was often rediscovered by many different authors (cf. [7]). Nowadays there exists a large literature on this subject. Well quasi-orders are important in mathematics and in many areas of theoretical computer science as well. Basic theorems such as the Higman theorem [6] and the Simon theorem [13] give a deep insight into combinatorics on words and language theory. As shown by Ehrenfeucht et al. [4] *regular languages can be characterized as the closed sets of well quasi-orders* in the set of all the words over a finite alphabet; moreover, well quasi-orders can be naturally associated with the derivation relations of suitable rewriting (or semi-Thue) systems. Ehrenfeucht et al. [4], for instance, characterized the unitary rewriting systems whose derivation relations are well quasi-orders.

In [11] we proved that the derivation relations associated to two suitable classes of rewriting systems are well quasi-orders. As a consequence we obtained some noteworthy regularity conditions extending a result of Restivo and Reutenauer [12] and a theorem of Latteux and Rozenberg [8]. In this paper we consider the recognizable parts of an arbitrary semigroup S. A subset X of S is recognizable if and only if it is a closed set of a monotone well quasi-order in S (generalized Myhill theorem). Moreover, the class of all monotone quasi-orders with respect to which X is a closed set, admits a maximum element w.r.t. the inclusion, that we call the Myhill quasi-order relative to X.

A problem that naturally arises is whether one can generalize also the Nerode theorem of automata theory. In other words let X be a closed set with respect to a well quasi-order monotone on the right. Is then X recognizable? One can easily prove that the class of all quasi-orders monotone on the right, w.r.t. which X is closed has a maximum element called the Nerode quasi-order relative to X. However, as we will prove in detail in Sect. 5, the answer to the preceding problem is negative. For instance the non-regular language $L=\{a^n b^m \mid n \geqq m \geqq 0\}$ is such that the Nerode quasi-order of L is a well quasi-order. In Sect. 2 we consider subsets X of S which are closed with respect to a well quasi-order monotone on the right; the problem that we face is to find some further conditions under which X becomes recognizable. In particular we prove that X is recognizable if and only if X is a closed set of a well quasi-order $\leqq_1$ monotone on the right and of a well quasi-order $\leqq_2$ monotone on the left. A stronger result is given by the weaker requirement (cf. Corollary 2.10) that $\leqq_1$ and $\leqq_2$ are well quasi-orders, respectively, on the sets $P(X)$ of the left-factors and $S(X)$ of the right-factors of the elements of X.

In Sect. 3 we shall refer to finitely generated free monoids and consider the family $\mathscr{F}$ of all languages which are closed with respect to a well quasi-order monotone on the right. We prove some remarkable closure properties of $\mathscr{F}$. The family $\mathscr{F}$ is rationally closed; moreover, $\mathscr{F}$ is closed under intersection, inverse morphisms and direct non-erasing morphisms. Since $\mathscr{F}$ contains Rat, it follows that $\mathscr{F}$ is an AFL, so that it is closed under faithful rational transductions. We stress that the proofs of these closure properties use techniques based on well-quasi orders; they can be utilized to show in a non-standard way the existence of these closure properties for the family Rec = Rat.

In Sect. 4 we prove that the languages of $\mathscr{F}$ satisfy a *pumping* lemma. However, our lemma differs from the classical formulation for regular languages in that the *pump* is *positive*, and located 'near' the initial parts of the words of the language.

In Sect. 5 we prove that $\mathscr{F}$ is a large family of languages such to contain languages which are not recursively enumerable. In conclusion some auxiliary results are given and some open problems are formulated.

1. Well quasi-orders

A binary relation $\leqq$ on a set S is a *quasi-order* (qo) if $\leqq$ is reflexive and transitive. If for all $s, t \in S$, $s \leqq t \leqq s$ implies $s = t$ then $\leqq$ is a partial order (po). If $s \leqq t$ implies $t \leqq s$ then $\leqq$ is an equivalence relation. The meet $\leqq \cap \leqq^{-1}$ is an equivalence relation $\sim$ and the quotient of S by $\sim$ is a poset (partially ordered set). It is clear that any qo generates a po over the equivalence classes.

An elements $s \in X \subseteq S$ is *minimal* (resp. *maximal*) in X with respect to $\leqq$ if, for every $x \in X$, $x \leqq s$ (resp. $s \leqq x$) implies that $x \sim s$. For $s, t \in S$ if $s \leqq t$ and s is not equivalent to t modulo $\sim$, then we set $s < t$. A subset X of S is upper-closed, or simply *closed*, with respect to $\leqq$ (or X is a closed set of $\leqq$), if the following condition is satisfied:

$$[x \in X \text{ and } x \leqq y] \Rightarrow y \in X.$$

A quasi-order on S is called a *well quasi-order* (*wqo*) *if every non-empty subset X of S has at least one minimal element but no more than a finite number of* (*non-equivalent*) *minimal elements.* It is clear that a well ordered set is well quasi-ordered. There exist several conditions which characterize the concept of well quasi-order and that can be assumed as equivalent definitions (cf. [6]).

Theorem 1.1 *Let S be a set quasi-ordered set by $\leqq$. The following conditions are equivalent*:
(i) $\leqq$ is a well quasi-order,
(ii) the ascending chain condition holds for the closed subsets of S (ordered by inclusion),
(iii) every infinite sequence of elements of S has an infinite ascending subsequence,
(iv) if $s_1, s_2, \ldots$ is an infinite sequence of elements of S, then there exist integers i, j such that $0 < i < j$ and $s_i \leqq s_j$,
(v) there exists neither an infinite strictly descending sequence in S (i.e. $\leqq$ is well founded), nor an infinity of mutually incomparable elements of S,
(vi) S has the "finite basis property", i.e. for each subset X of S there exists a finite subset F_X of X such that for every $x \in X$ there exists a $y \in F_X$ such that $y \leqq x$.

If ρ and σ are two relations on the sets S and T respectively, then the direct product $\rho \times \sigma$ is the relation on $S \times T$ defined as: $(a, b)\ \rho \times \sigma\ (c, d)$ if and only if $a \rho c$ and $b \sigma d$. From condition (iii) of Theorem 1.1 one derives immediately the following:

Proposition 1.2 *Every subset of a* wqo *set is* wqo. *If S and T are well quasi-ordered by the relations $\leqq_1$ and $\leqq_2$, respectively, then $S \times T$ is well quasi-ordered by $\leqq_1 \times \leqq_2$. If $\leqq_1$ is a* wqo *on S, then any other* wqo *$\leqq_2$ on S such that $\leqq_1 \subseteq \leqq_2$ is a* wqo.

As a consequence of the fact that $\mathbb{N}$ is well ordered one easily derives the following famous lemma, due to Dickson (cf. [9, Chapt. 6, Problem 6.1.2]):

Lemma 1.3 *Let $q > 0$ and consider on $\mathbb{N}^q$ the relation $\leqq$ defined as: For $(r_1, \ldots, r_q)$, $(s_1, \ldots, s_q) \in \mathbb{N}^q$*

$$(r_1, \ldots, r_q) \leqq (s_1, \ldots, s_q) \quad \textit{if and only if } \forall i \in [1, q], \quad r_i \leqq s_i.$$

Then $\leqq$ is a well quasi-order on $\mathbb{N}^q$.

Lemma 1.4 *Let* $\leqq$ *be a* wqo *on* S *and* $\sim$ *the equivalence relation* $\sim\,=\,\leqq\cap\leqq^{-1}$. *The following conditions are equivalent*:
i) $\sim$ *is an equivalence of finite index,*
ii) $\leqq^{-1}$ *is a* wqo,
iii) $\leqq^{-1}$ *is well founded.*

Proof. i)$\Rightarrow$*ii)*. An equivalence of finite index is trivially a wqo. Since $\sim\,\subseteq\,\leqq^{-1}$ it follows from Proposition 1.2 that $\leqq^{-1}$ is a wqo. The implication ii)$\Rightarrow$iii) is a consequence of Theorem 1.1 (cf. condition *v*). *iii)*$\Rightarrow$i) Suppose, by contradiction, that $\sim$ has not a finite index. There will exist elements $s_1, s_2, \ldots, s_n, \ldots$ such that for all positive integers i, j, $i \neq j$, s_i is not equivalent to s_j, modulo $\sim$. Since $\leqq$ is a wqo there exists an infinite subsequence $s_{r_1}, s_{r_2}, \ldots, s_{r_n}, \ldots$ such that $r_1 < r_2 \ldots < r_n < \ldots$ and

$$s_{r_1} < s_{r_2} < \ldots < s_{r_n} < \ldots,$$

which contradicts the hypothesis that $\leqq^{-1}$ is well founded. Q.E.D.

Let $\leqq$ be a quasi-order defined on a set S. We can consider the set Σ of all infinite sequences of elements of S, $\mathbf{y}\colon \mathbb{N} \to S$,

$$\mathbf{y} = y_0 y_1 \cdots y_n \cdots,$$

where $y_i = \mathbf{y}(i)$ for all $i \geqq 0$. The quasi-order $\leqq$ on S naturally induces a quasi-order $\leqq$ on Σ as follows: Let $\mathbf{y}$, $\mathbf{z} \in \Sigma$. If for all $i \geqq 0$, $y_i \sim z_i$, then we set $\mathbf{y} \sim \mathbf{z}$, otherwise we define $i = \min\,\{j \in \mathbb{N} \mid y_j$ is not equivalent to z_j, modulo $\sim\}$ and set $\mathbf{y} < \mathbf{z}$ if $y_i < z_i$. One easily verifies that the relation $\mathbf{y} \leqq \mathbf{z}$ if and only if $\mathbf{y} \sim \mathbf{z}$ or $\mathbf{y} < \mathbf{z}$, is a quasi-order on Σ.

Let $\leqq$ be a quasi-order on S and $\mathbf{x}\colon \mathbb{N} \to S$ be an infinite sequence of elements of S. We call $\mathbf{x}$ 'bad' if for all i, j such that $0 \leqq i < j$, one has $x_i \nleqq x_j$. Many proofs on well quasi-orders, as well as the proof of the Higman theorem, are based on the following proposition essentially due to Nash-Williams (cf. [9], [7]):

Proposition 1.5 *Let* $\leqq$ *be a well-founded quasi-order on S. Let* $\leqq'$ *be a quasi-order on S which is not a well quasi-order. Then there exists a 'bad' sequence* (w.r.t. $\leqq'$) *which is minimal with respect to the order* $\leqq$.

Proof. The order $\leqq'$ is not a wqo so that the negation of statement (*v*) in Theorem 1.1 has to be true. There exists at least one infinite sequence $\mathbf{y}\colon \mathbb{N} \to S$, which is 'bad'. Let Σ_0 be the set of all such sequences. We construct then a sequence $\mathbf{x}\colon \mathbb{N} \to S$ inductively, as follows: let x_0 be a minimal element w.r.t. to $\leqq$ in the set $\{y_0 \mid \mathbf{y} \in \Sigma_0\}$. For all $i \geqq 0$ one defines $\Sigma_{i+1} = \{\mathbf{y} \in \Sigma_i \mid y_i = x_i\}$ and x_{i+1} is any minimal element, w.r.t. $\leqq$, in the set $\{y_{i+1} \mid \mathbf{y} \in \Sigma_{i+1}\}$. Let us prove that $\mathbf{x} \in \Sigma_0$, i.e. for all i, j, $0 \leqq i < j$, one has $x_i \nleqq' x_j$. Suppose in fact, by contradiction, that there exist i, j, $0 \leqq i < j$, for which $x_i \leqq' x_j$. This is absurd since $x_0, x_1, \ldots, x_j$ are the first $j+1$ terms of a sequence of Σ_0. By construction the sequence $\mathbf{x}$ is minimal in Σ_0 w.r.t. the order $\leqq$. Q.E.D

If we take $\leqq\,=\,\leqq'$ one obtains the following:

Corollary 1.6 *Let* $\leqq$ *be a well founded quasi-order on S that is not a well quasi-order. Then there exists a minimal bad sequence with respect to* $\leqq$.

Let $\leqq$ be a wqo on a set S. We denote by $\mathrm{Fin}(S)$ the family of all finite subsets of S. We can quasi-order $\mathrm{Fin}(S)$ by $\leqq'$ defined as: For $A, B \in \mathrm{Fin}(S)$

$$A \leqq' B \Leftrightarrow \forall s \in A \exists t \in B \quad \text{such that } s \leqq t.$$

The following lemma has some useful applications in the following sections:

Lemma 1.7 *If* $\leqq$ *is a* wqo *on* S, *then* $\leqq'$ *is a* wqo *on* $\mathrm{Fin}(S)$.

Proof. We can order the finite subsets of S by their cardinalities. We denote this order by $\lhd$. By Proposition 1.4 if $\leqq'$ is not a well quasi-order, then there will exist a sequence:

$$(1.1) \qquad A_1, A_2, \ldots, A_n, \ldots,$$

which is a 'bad' sequence minimal w.r.t. $\lhd$. The sets A_i, $i>0$ have to be nonempty since, otherwise, one would contradict the fact that the sequence is 'bad'. Let us then take in each A_i, $i>0$, one element $a_i \in A_i$, and consider the infinite sequence:

$$a_1, a_2, \ldots, a_n, \ldots.$$

Since $\leqq$ is a wqo there exists an infinite ascending subsequence:

$$a_{j_1} \leqq a_{j_2} \leqq \ldots \leqq a_{j_n} \leqq \ldots.$$

Let us observe that for all $s>0$, $a_{j_s} \in A_{j_s}$ and $\mathrm{Card}(A_{j_s})>1$. Indeed, if there exists an integer $s>0$ such that $A_{j_s}=\{a_{j_s}\}$ then one would derive $A_{j_s} \leqq' A_{j_{s+1}}$ which is absurd.

Let us set $j=j_1$ and consider then the sequence:

$$A_1, A_2, \ldots, A_{j-1}, B_j, B_{j+1}, \ldots, B_{j+n}, \ldots.$$

where for every $s \geqq 0$, $B_{j+s}=A_{j_{s+1}} \setminus \{a_{j_{s+1}}\}$. This sequence is still a 'bad' sequence. We prove that $A_s \not\leqq' B_{j+n} (s=1, \ldots, j-1; n \geqq 0)$ and $B_{j+r} \not\leqq' B_{j+s}$ for $0 \leqq r<s$. Indeed, if $A_s \leqq' B_{j+n}$ since $B_{j+n} \leqq' A_{j_{n+1}}$ one would derive $A_s \leqq' A_{j_{n+1}}$ which is a contradiction. Let us now suppose that $B_{j+r} \leqq' B_{j+s}$ with $0 \leqq r<s$; since $a_{j_{r+1}} \leqq a_{j_{s+1}}$ one derives $A_{j_{r+1}} \leqq' A_{j_{s+1}}$ which is again a contradiction. Hence the above sequence is 'bad' and this contradicts the minimality of the sequence (1.1) w.r.t. $\lhd$ because $\mathrm{Card}(B_j)<\mathrm{Card}(A_j)$. Q.E.D.

Let us now suppose that the set S is a semigroup. A quasi-order $\leqq$ on a semigroup S is *monotone on the right* (resp. *on the left*) if for all $x_1, x_2, y \in S$

$$x_1 \leqq x_2 \quad \text{implies} \quad x_1 y \leqq x_2 y \quad (\text{resp. } y x_1 \leqq y x_2).$$

A quasi-order is *monotone* if and only if it is monotone on the right and on the left. One has in particular that a monotone equivalence is a congruence in S.

2. Recognizable parts of a semigroup. The generalized Myhill and Nerode Theorems

Let S be a semigroup. A subset X of a semigroup S is *recognizable* if it is a closed set of a congruence in S of finite index defined in it (cf. [5]).

As is well known, a characterization of the recognizable subsets of S can be given in terms of the *syntactic semigroup* $S(X)$ of X. More precisely there exists a maximal congruence $\equiv_X$ w.r.t. which X is closed (i.e. X is union of classes). This is called the *syntactic congruence* of X and it is defined as follows: for any $s \in S$ let $\mathrm{Cont}_X(s) = \{(u, v) \in S^1 \times S^1 \mid usv \in X\}$; one sets for $s, t \in S$, $s \equiv_X t$ if and only if $\mathrm{Cont}_X(s) = \mathrm{Cont}_X(t)$. The syntactic semigroup is then defined as $S(X) = S/\equiv_X$. One easily derives that *a subset X of S is recognizable if and only if $S(X)$ is a finite semigroup* (Myhill's theorem). In the following the class of recognizable subsets of S will be denoted by $\mathrm{Rec}(S)$.

In a semigroup S one can introduce the so-called *rational operations* of union ($\cup$), product (.) and ($+$). For any subset X of S, X^+ is the subsemigroup of S generated by X. The class $\mathrm{Rat}(S)$ of the rational subsets of S is then defined as the smallest family of subsets of S containing the finite subsets and closed under the rational operations. If S is a *finitely generated free semigroup* then the recognizable subsets of S coincide also with the class of subsets definable in terms of *finite automata*. Moreover, in this case a fundamental theorem due to Kleene (cf. [5]) states that $\mathrm{Rec}(S) = \mathrm{Rat}(S)$.

A characterization of recognizable subsets of a semigroup can be obtained in terms of well quasi-orders. In fact the following theorem holds. (When S is a finitely generated free semigroup this is usually called the *generalized Myhill theorem* [4]).

Theorem 2.1 *Let X be a subset of a semigroup S. $X \in \mathrm{Rec}(S)$ if and only if X is closed with respect to a monotone well quasi-order on S.*

A more general version of Theorem 2.1 will be proved in the sequel. The link between congruences and well quasi-orders is given by the following:

Proposition 2.2 *A congruence in a semigroup S has a finite index if and only if it is a* wqo.

Proof. A congruence θ in S is a monotone quasi-order. Since $\theta^{-1} = \theta$ by Lemma 1.4 the result follows. Q.E.D.

Let $X \subseteq S$; we introduce the following relation $\leqq_X$ on S defined as: For all $s, t \in S$,

$$s \leqq_X t \quad \text{if and only if} \quad \mathrm{Cont}_X(s) \subseteq \mathrm{Cont}_X(t).$$

We call $\leqq_X$ the *Myhill quasi-order relation relative to X*. One easily verifies the following facts: 1. $\leqq_X$ is a monotone quasi-order; 2. X is closed w.r.t. $\leqq_X$; 3. the equivalence relation $\leqq_X \cap (\leqq_X)^{-1}$ coincides with the syntactic congruence of X; 4. $\leqq_X$ is maximum (w.r.t. inclusion) in the set of all monotone quasi-orders with respect to which X is closed.

Proposition 2.3 *Let X be a subset of a semigroup S. $X \in \mathrm{Rec}(S)$ if and only if $\leqq_X$ is a well quasi-order.*

Proof. ($\Leftarrow$) It is trivial by the fact that X is closed w.r.t. the monotone well quasi-order $\leqq_X$ so that the result follows from the generalized Myhill theorem.

($\Rightarrow$) If $X \in \mathrm{Rec}(S)$ then $\equiv_X$ is of finite index. Since $\equiv_X \subseteq \leqq_X$ then by Propositions 1.2 and 2.2 one has that $\leqq_X$ is a wqo. Q.E.D.

Recognizable subsets of a semigroup can also be described by equivalence relations monotone on the right or on left, called *right* and *left congruences*, respectively. It is well known that *a subset X of a semigroup S is recognizable if and only if it is a closed set of a right (or left) congruence of finite index* (Nerode's theorem). For any subset X of S one can consider the Nerode right congruence $\mathfrak{N}_X^r$ of X defined in the following way. For any $s \in S$ set

$$s^{-1}X = \{t \in S \mid st \in X\} \quad \text{and} \quad Xs^{-1} = \{t \in S \mid ts \in X\}.$$

One has then

$$s\,\mathfrak{N}_X^r\,t \quad \text{if and only if } s^{-1}X = t^{-1}X.$$

In a symmetric way the Nerode left congruence $\mathfrak{N}_X^l$ of X is defined as:

$$s\,\mathfrak{N}_X^l\,t \quad \text{if and only if } Xs^{-1} = Xt^{-1}.$$

In terms of Nerode's congruences the property of being recognizable becomes: X is recognizable if and only if $\mathfrak{N}_X^r$ or $\mathfrak{N}_X^l$ is of finite index.

For any subset X of S we can then introduce the quasi-order $\leqq_X^r$ that we call the (*right*) *Nerode quasi-order relative to* X, defined as: For $s, t \in S$

$$s \leqq_X^r t \quad \text{if and only if } s^{-1}X \subseteq t^{-1}X.$$

One easily verifies the following facts: 1. $\leqq_X^r$ is a quasi-order monotone on the right; 2. X is closed w.r.t. $\leqq_X^r$; 3. the equivalence $\leqq_X^r \cap (\leqq_X^r)^{-1}$ is the Nerode equivalence $\mathfrak{N}_X^r$; 4. $\leqq_X^r$ is maximum in the set of all quasi-orders monotone on the right w.r.t. which X is closed; 5. $\leqq_X \subseteq \leqq_X^r$.

In a symmetric way the quasi-order $\leqq_X^l$ is defined as: for $s, t \in S$:

$$s \leqq_X^l t \quad \text{if and only if } Xs^{-1} \subseteq Xt^{-1}.$$

The relation $\leqq_X^l$ satisfies properties prefectly symmetric to those of $\leqq_X^r$. Since $\leqq_X \subseteq \leqq_X^r$ one has that if X is recognizable then $\leqq_X^r$, as well as $\leqq_X^l$, is a well quasi-order. One can ask the question whether, in analogy with Nerode's theorem, the fact that $\leqq_X^r$ is a well quasi-order assures that X is a recognizable subset of S. The answer is negative as will be shown in Sect. 5. In fact, when $S = A^*$ is a finitely generated free monoid, we prove that there exists a very large class of languages L over A which are not regular and such that $\leqq_L^r$ is a wqo. For instance, when $A = \{a, b\}$ the language $L = \{a^n b^m \mid n \geqq m \geqq 0\}$ is not regular and is such that $\leqq_L^r$ is a wqo (cf. Proposition 5.1).

A partial generalization of Nerode's theorem and of Theorem 2.1 as well, is given by the following:

Theorem 2.4 *Let X be a subset of S. The following conditions are equivalent*
i. X is recognizable
ii. X is a closed set of a wqo $\leqq_1$ *monotone on the right and of a* wqo $\leqq_2$ *monotone on the left.*

iii. The quasi-orders $\leqq_X^r$ and $\leqq_X^l$ are wqo.

Proof. The implication $i \Rightarrow ii$ is trivial since if X is recognizable then it is a closed set of a congruence in S of finite index. The implication $ii \Rightarrow iii$ follows from the fact that $\leqq_1 \subseteq \leqq_X^r$ and $\leqq_2 \subseteq \leqq_X^l$. Let us then prove the implication $iii \Rightarrow i$. If X is not recognizable, then there exists an infinite sequence:

$$x_1, x_2, \ldots, x_n, \ldots$$

of elements of S such that for all integers $i, j, i \neq j$, one has that x_i is not equivalent to x_j, modulo $\mathfrak{N}_X^r$. Since $\leqq_X^r$ is a wqo there exists an infinite subsequence $y_1, y_2, \ldots, y_n, \ldots$ which is strictly increasing with respect to $\leqq_X^r$:

$$y_1 <_X^r y_2 <_X^r \ldots <_X^r y_n <_X^r \ldots.$$

This implies that for all $i \geqq 1$, $y_i^{-1} X \subset y_{i+1}^{-1} X$. Now, as one easily verifies, any set $y_i^{-1} X$ is a closed set of $\leqq_X^l$ so that we have an infinite strictly increasing chain of closed sets of $\leqq_X^l$. By Theorem 4.1 this contradicts the fact that $\leqq_X^l$ is a wqo. Q.E.D.

We now give an application of Theorem 2.4. Let S be a semigroup. We say that S satisfies the maximal condition $M_R(M_L)$ on right (left) ideals if any chain of right (left) ideals, ordered by inclusion, terminates. We recall that S can be quasi-ordered by the relations $\leqq_{\mathscr{R}}(\leqq_{\mathscr{L}})$ defined as: For $s, t \in S$,

$$s \leqq_{\mathscr{R}} t \quad \text{if and only if } sS^1 \subseteq tS^1 \quad \text{and} \quad s \leqq_{\mathscr{L}} t \quad \text{if and only if } S^1 s \subseteq S^1 t.$$

One has that $\leqq_{\mathscr{R}}(\leqq_{\mathscr{L}})$ is monotone on the left (right); moreover, $\leqq_{\mathscr{R}} \cap \leqq_{\mathscr{R}}^{-1} = \mathscr{R}$ and $\leqq_{\mathscr{L}} \cap \leqq_{\mathscr{L}}^{-1} = \mathscr{L}$, where $\mathscr{R}(\mathscr{L})$ is the $\mathscr{R}$-relation ($\mathscr{L}$-relation) of Green (cf. [3]).

Lemma 2.5 *In a semigroup S the following two conditions are equivalent*
i. S satisfies $M_R(M_L)$,
ii. the quasi order $\leqq_{\mathscr{R}}^{-1}(\leqq_{\mathscr{L}}^{-1})$ is a wqo.

Proof. Let us first prove that a subset R of S is a closed set of $\leqq_{\mathscr{R}}^{-1}$ if and only if R is a right ideal of S. Indeed, let R be a closed set of $\leqq_{\mathscr{R}}^{-1}$, $r \in R$ and $s \in S$. Since $rsS^1 \subseteq rS^1$ one has $r \leqq_{\mathscr{R}}^{-1} rs$, so that $rs \in R$. Conversely, let R be a right ideal of S and $r \in R$ be such that $r \leqq_{\mathscr{R}}^{-1} s$. One has then $sS^1 \subseteq rS^1$, so that $s = rv$ with $v \in S^1$, which implies $s \in R$. Hence R is a closed set of $\leqq_{\mathscr{R}}^{-1}$. Let us now suppose that S satisfies M_R. If $R_1 \subseteq R_2 \subseteq \ldots \subseteq R_n \subseteq \ldots$ is an ascending chain of closed sets of $\leqq_{\mathscr{R}}^{-1}$, then by M_R the chain has to terminate so that by Theorem 1.1, $\leqq_{\mathscr{R}}^{-1}$ is a wqo. Conversely, suppose that $\leqq_{\mathscr{R}}^{-1}$ is a wqo and be $R_1 \subseteq R_2 \subseteq \ldots \subseteq R_n \subseteq \ldots$ an ascending chain of right ideals. Since the sets R_i, $i \geqq 0$, are closed w.r.t. $\leqq_{\mathscr{R}}^{-1}$ if follows by Theorem 1.1 that the chain has to terminate, so that the condition M_R has to be satisfied. A symmetric argument shows that S satisfies M_L if and only if $\leqq_{\mathscr{L}}^{-1}$ is a wqo. Q.E.D.

Corollary 2.6 *Let S be a semigroup satisfying M_R and M_L. Then for any two-sided ideal J, $J \in \mathrm{Rec}(S)$.*

Proof. By the preceding lemma $\leqq_{\mathscr{R}}^{-1}$ and $\leqq_{\mathscr{L}}^{-1}$ are wqo. Let J be a two-sided ideal of S. As we have seen in the proof of Lemma 2.5, J is a closed set of

$\leqq_{\mathscr{R}}^{-1}$ and of $\leqq_{\mathscr{L}}^{-1}$. Since $\leqq_{\mathscr{R}}^{-1}$ ($\leqq_{\mathscr{L}}^{-1}$) is monotone on the left (on the right) then by Theorem 2.4, the result follows. Q.E.D.

For any $X \subseteq S$ we denote by X^c the complement of the set X.

Proposition 2.7 *If* $\leqq_X^r$ *is a* wqo *and* $\leqq_{X^c}^r$ *is well founded, then* $X \in \mathrm{Rec}(S)$.

Proof. Let us observe that $\leqq_{X^c}^r = (\leqq_X^r)^{-1}$ so that $\leqq_X^r \cap \leqq_{X^c}^r = \mathfrak{N}_X^r$. By Lemma 1.4, $\mathfrak{N}_X^r$ is of finite index, so that X is a recognizable set. Q.E.D.

Corollary 2.8 *Let* $X \in \mathrm{Rec}(S)$. *If* Y *is closed with respect to* $\leqq_X^r$, *then* $Y \in \mathrm{Rec}(S)$.

Proof. One has that

$$\mathfrak{N}_X^r \subseteq \leqq_X^r \subseteq \leqq_Y^r \quad \text{and} \quad \mathfrak{N}_X^r = \mathfrak{N}_{X^c}^r \subseteq \leqq_{X^c}^r = (\leqq_X^r)^{-1} \subseteq (\leqq_Y^r)^{-1} = \leqq_{Y^c}^r.$$

Since $\mathfrak{N}_X^r$ is of finite index, then $\leqq_Y^r$ and $\leqq_{Y^c}^r$ will be wqo so that from Proposition 2.7, $Y \in \mathrm{Rec}(S)$. Q.E.D.

We make the following conjecture:

Conjecture. Let $\leqq_X^r$ be a wqo and $\leqq_X^l$ be well founded. Then $X \in \mathrm{Rec}(S)$.

Let us remark that in the case of the non-regular language $L = \{a^n b^m \mid n \geqq m \geqq 0\}$ over the alphabet $\{a, b\}$, which is such that $\leqq_L^r$ is a wqo, the order $\leqq_L^l$ is not well founded. Indeed, as one easily verifies, for all integers i, j such that $0 < i < j$, one has that $b^j <_L^l b^i$.

Let S be a semigroup and $s, t \in S$. We say that s is a *factor* of t if $t \in S^1 s S^1$, i.e. there exist $x, y \in S^1$ such that $t = xsy$. If $x = 1$ ($y = 1$), then s is called a *left-factor* (*right-factor*) of t. For any $t \in S$ we denote by $F(t)(P(t), S(t))$ the set of all factors (left-factors, right-factors) of t. For any $X \subseteq S$ we set $F(X) = \bigcup_{x \in X} F(x)(P(X) = \bigcup_{x \in X} P(x), S(X) = \bigcup_{x \in X} S(x))$.

Let $\leqq$ be a quasi-order on S and X be a subset of S. We say that $\leqq$ is a wqo on X if X is well quasi-ordered by $\leqq$.

Proposition 2.9 *Let X be a closed set of a right-monotone quasi-order $\leqq$ on S. If $\leqq$ is a* wqo *on $P(X)(S(X))$ then X is a closed set of a right-monotone (left-monotone) well quasi-order on S.*

Proof. Let us first observe that since X is closed w.r.t. $\leqq$ so will be $P(X)$. Indeed, let $x \in P(X)$ and be $x \leqq y$. There exists $u \in S^1$ such that $xu \in X$. Since $\leqq$ is right-monotone one has $yu \in X$ and then $y \in P(X)$. We introduce now on S a quasi-order $\leqq^e$ which is right-monotone and such that X is a closed w.r.t. $\leqq^e$. For all $u, v \in S$ we set:

$$u \leqq^e v \quad \text{if and only if either } u \notin P(X) \quad \text{or} \quad u \leqq v.$$

Let us verify that $\leqq^e$ is a quasi-order on S. Trivially $\leqq^e$ satisfies the reflexive property. Let us then verify that $\leqq^e$ satisfies the transitive property. Let $u, v, w \in S$ be such that $u \leqq^e v$ and $v \leqq^e w$. If $u \notin P(X)$, then, trivially, $u \leqq^e w$. Let us then suppose that $u \in P(X)$. One has then $u \leqq v$ and $v \in P(X)$. Since $v \leqq^e w$ it follows also $v \leqq w$ and $w \in P(X)$. Hence since $u \leqq w$, one has that $u \leqq^e w$.

We now show that $\leqq^e$ is right-monotone. Indeed, let $x \leqq^e y$ and $\lambda \in S^1$. If $x \notin P(X)$, then $x\lambda \notin P(X)$ so that $x\lambda \leqq^e y\lambda$. If $x \in P(x)$, then $x \leqq y$. Since $\leqq$ is right-monotone one has $x\lambda \leqq y\lambda$. If $x\lambda \notin P(X)$ then one has $x\lambda \leqq^e y\lambda$. If $x\lambda \in P(x)$, then, since $x\lambda \leqq y\lambda$, it follows $x\lambda \leqq^e y\lambda$.

X is closed w.r.t. $\leqq^e$. In fact, let $x \in X$ and be $x \leqq^e y$. Since $x \in P(X)$ then $x \leqq y$. Since X is closed w.r.t. $\leqq$, one derives $y \in X$.

Let us now prove that $\leqq^e$ is a wqo. Indeed, take an infinite sequence $x_1, \ldots, x_n \ldots$ of elements of S. If there exists an integer $i \geqq 1$ such that $x_i \notin P(X)$, then $x_i \leqq^e x_j$ for any $j \geqq i$, so that the result is trivially achieved. We can then suppose that all the elements of the preceding sequence belong to $P(X)$. Since $\leqq$ is a wqo on $P(X)$ then there will exist integers $i, j, 0 < i < j$, such that $x_i \leqq x_j$ and then $x_i \leqq^e x_j$.

An entirely symmetric argument can be given when $\leqq$ is a left-monotone wqo on $S(X)$. In this case X will be closed w.r.t. a left-monotone well quasi-order $\leqq^f$ on S. Q.E.D.

Corollary 2.10 *Let X be a closed set of the quasi-orders $\leqq_1$ and $\leqq_2$ in S monotone on the right and on the left, respectively. If $\leqq_1$ is a* wqo *on $P(X)$ and $\leqq_2$ is a* wqo *on $S(X)$, then $X \in \mathrm{Rec}(S)$.*

Proof. By the preceding proposition X is closed w.r.t. a right-monotone wqo $\leqq^e$ on S, and closed w.r.t. a left-monotone wqo $\leqq^f$ on S. By Theorem 2.4 it follows that $X \in \mathrm{Rec}(S)$. Q.E.D.

Corollary 2.11 *Let X be closed with respect to a monotone quasi-order $\leqq$ on S. If $\leqq$ is a* wqo *on $F(X)$, then $X \in \mathrm{Rec}(S)$.*

Proof. Since $P(X)$, $S(X)$ are subsets of $F(X)$, then $\leqq$ is a wqo on $P(X)$ and on $S(X)$. In view of the fact that $\leqq$ is monotone, from Corollary 2.10 the result follows. Q.E.D.

Let us remark that by Theorem 2.4 the statements of Corollaries 2.10 and 2.11 are not only sufficient, but also necessary conditions in order that $X \in \mathrm{Rec}(S)$. When $\leqq$ is an equivalence relation θ, then Corollary 2.11 gives rise to the following statement about recognizable sets which can be also, directly proved in a different way:

Corollary 2.12 *Let X be closed part of a congruence relation θ in S. If the index of θ in $F(X)$ is finite, then $X \in \mathrm{Rec}(S)$.*

Proof. It is sufficient to observe that the condition that θ is a wqo on $F(X)$ is equivalent to the condition that $F(X)$ is union of a finite number of classes of the congruence θ. Q.E.D.

3. The case of free monoids

In this section we shall consider the case when S is a finitely generated free monoid. More precisely let A be a finite non-empty set, or *alphabet*, and $A^*(A^+)$ the *free monoid* (*free semigroup*) over A. The elements of A are called *letters* and those of A^* *words*. The identity element of A^* is called the *empty word* and denoted by Λ. For any word w, $|w|$ denotes its *length*. A factor (left-factor, right-factor) of a word w is usually called *subword* (*prefix*, *suffix*) of w. Any subset L of A^* is called *language*. A rational language is also called *regular*. We recall that the *reversal operation* ($^\sim$) in A^* is defined as: $\Lambda^\sim = \Lambda$ and for any $u \in A^*$ and $a \in A$, $(ua)^\sim = au^\sim$; the reversal operation is an involutory anti-automorphism of A^* since for any $u, v \in A^*$, $(uv)^\sim = v^\sim u^\sim$ and $(u^\sim)^\sim = u$. For any $L \subseteq A^*$, $L^\sim$ will denote the language $L^\sim = \{w^\sim \in A^* | w \in L\}$.

We denote by $\mathscr{F}_A$ the class of languages $L \subseteq A^*$ which are closed sets of a well quasi-order on A^* monotone on the right. In a symmetric way $\mathscr{G}_A$ will be the class of languages $L \subseteq A^*$ which are closed sets of a well quasi-order on A^* monotone on the left. As we have seen in the preceding section $L \in \mathscr{F}_A (L \in \mathscr{G}_A)$ if and only if $\leqq_L^r (\leqq_L^l)$ is a wqo. By Theorem 2.4, one has that $\mathscr{F}_A \cap \mathscr{G}_A = \mathrm{Rat}(A^*)$. We shall prove now some remarkable closure properties of $\mathscr{F}_A$.

Proposition 3.1 *$\mathscr{F}_A$ is closed under union and intersection.*

Proof. Let L, $M \in \mathscr{F}_A$. By hypothesis $\leqq_L^r$ and $\leqq_M^r$ are wqo. Let us first prove that the relation $\leqq_L^r \cap \leqq_M^r$ is a wqo. Indeed, let

$$u_1, u_2, \ldots, u_n, \ldots$$

be an infinite sequence of words. We then consider the infinite sequence of elements of $A^* \times A^*$:

$$(u_1, u_1), \ldots, (u_n, u_n) \ldots .$$

Since from Proposition 1.2, $\leqq_L^r \times \leqq_M^r$ is a wqo on $A^* \times A^*$, then from Theorem 1.1 there exist integers i, j, $0 < i < j$, such that

$$(u_i, u_i)(\leqq_L^r \times \leqq_M^r)(u_j, u_j).$$

This implies $u_i \leqq_L^r u_j$ and $u_i \leqq_M^r u_j$, i.e. $u_i(\leqq_L^r \cap \leqq_M^r) u_j$. Then by Theorem 1.1 it follows that $\leqq_L^r \cap \leqq_M^r$ is a wqo.

Let now u, $v \in A^*$ be such that $u(\leqq_L^r \cap \leqq_M^r)v$. If $u \in L$ then since $u \leqq_L^r v$ it follows that $v \in L$. If $u \in M$, then since $u \leqq_M^r v$ it follows that $v \in M$. Thus if $u \in L \cup M$, then $v \in L \cup M$ and if $u \in L \cap M$ then $v \in L \cap M$. This shows that $L \cup M$ and $L \cap M$ are closed w.r.t. the well quasi-order monotone on the right $\leqq_L^r \cap \leqq_M^r$, so that $L \cup M$, $L \cap M \in \mathscr{F}_A$. Q.E.D.

Proposition 3.2 *$\mathscr{F}_A$ is not closed under complementation.*

Proof. Consider a language $L \in \mathscr{F}_A$. If its complement L^c belongs to $\mathscr{F}_A$ then $\leqq_{L^c}^r$ is a wqo. This implies by Proposition 2.7 that L is regular. Since $\mathscr{F}_A$ contains non-regular languages (c.f. Corollary 5.3) the result follows. Q.E.D.

Lemma 3.3 *If $\leqq_L^r$ is a* wqo, *then $\leqq_{L^\sim}^l$ is a* wqo.

Proof. Let $u_1, u_2, \ldots, u_n, \ldots$ be an infinite sequence of words. We consider then the infinite sequence $u_1^\sim, u_2^\sim, \ldots, u_n^\sim, \ldots$. Since $\leqq_L^r$ is a wqo there exist integers i, j, $0 < i < j$, such that $u_i^\sim \leqq_L^r u_j^\sim$, i.e. for any $\zeta \in A^*$, $u_i^\sim \zeta \in L \Rightarrow u_j^\sim \zeta \in L$. This implies $\zeta^\sim u_i \in L^\sim \Rightarrow \zeta^\sim u_j \in L^\sim$. Thus $u_i \leqq_{L^\sim}^l u_j$ so that $\leqq_{L^\sim}^l$ is a wqo. Q.E.D.

Proposition 3.4 *$\mathscr{F}_A$ is not closed under the reversal operation.*

Proof. Consider a language $L \in \mathscr{F}_A$. If $L^\sim \in \mathscr{F}_A$, then $\leqq_{L^\sim}^r$ is a wqo. Moreover, by Lemma 3.3 the fact that $\leqq_L^r$ is a wqo implies that $\leqq_{L^\sim}^l$ is a wqo. From this it follows that $L^\sim$, as well as L, is a regular language. Q.E.D.

Proposition 3.5 *$\mathscr{F}_A$ is closed under right-quotients.*

Proof. Let $L \in \mathscr{F}_A$ and $X \subseteq A^*$. We prove that LX^{-1} is closed w.r.t. $\leqq_L^r$ that implies $LX^{-1} \in \mathscr{F}_A$. Indeed, let $w \in LX^{-1}$ and $w \leqq_L^r u$. Since there exists $x \in X$ such that $wx \in L$ one has $wx \leqq_L^r ux$ and then $ux \in L$, i.e. $u \in LX^{-1}$. Q.E.D.

Proposition 3.6 *Let $\phi: A^* \to B^*$ be a morphism, where A and B are finite alphabets. If $L \in \mathscr{F}_B$, then $\phi^{-1}(L) \in \mathscr{F}_A$.*

Proof. If $L \in \mathscr{F}_B$, then $\leqq_L^r$ is a wqo on B^*. Let us then define in A^* the quasi-order $\leqq$ as: For $u, v \in A^*$,

$$u \leqq v \Leftrightarrow \phi(u) \leqq_L^r \phi(v).$$

The quasi-order $\leqq$ is monotone on the right. Let $u \leqq v$ and $h \in A^*$. One has $\phi(u)\phi(h) \leqq_L^r \phi(v)\phi(h)$ i.e. $\phi(uh) \leqq_L^r \phi(vh)$ that implies $uh \leqq vh$.

The relation $\leqq$ is a wqo. Indeed, let $u_1, u_2, \ldots, u_n, \ldots$ be an infinite sequence of words of A^* and consider the sequence:

$$\phi(u_1), \phi(u_2), \ldots, \phi(u_n), \ldots.$$

Since $\leqq_L^r$ is a wqo on B^* there exist integers i, j, $0 < i < j$, such that $\phi(u_i) \leqq_L^r \phi(u_j)$. This implies $u_i \leqq u_j$. We prove that $\phi^{-1}(L)$ is closed w.r.t. $\leqq$. Let $u \in \phi^{-1}(L)$ and $u \leqq v$. One has $\phi(u) \leqq_L^r \phi(v)$. Since $\phi(u) \in L$ then $\phi(v) \in L$ so that $v \in \phi^{-1}(L)$. Q.E.D.

Proposition 3.7 *$\mathscr{F}_A$ is closed under product.*

Proof. Let $L, M \in \mathscr{F}_A$ so that $\leqq_L^r$ and $\leqq_M^r$ are wqo. Let us define the map $\sigma_{L,M}$: $A^* \to \mathrm{Fin}(A^*)$, that we denote simply by σ, as: For any $w \in A^*$,

$$\sigma(w) = L^{-1}w \cap M(A^*)^{-1}.$$

In other terms a word $\lambda \in \sigma(w)$ if and only if λ is a prefix of a word of M and, moreover, there exists a word $v \in L$ such that $w = v\lambda$. It is clear that $\sigma(w)$ is a finite, possibly empty, set. We want to prove that $\leqq_{LM}^r$ is a wqo.

We recall that by Lemma 1.7, the relation $(\leqq_M^r)'$ is a wqo in $\mathrm{Fin}(A^*)$. Thus the set $A^* \times \mathrm{Fin}(A^*)$ is well quasi-ordered by the relation $\leqq_L^r \times (\leqq_M^r)'$. Consider now an arbitrary infinite sequence of words: $w_1, w_2, \ldots, w_n, \ldots$ and the infinite sequence of elements of $A^* \times \mathrm{Fin}(A^*)$:

$$(w_1, \sigma(w_1)), (w_2, \sigma(w_2)), \ldots, (w_n, \sigma(w_n)), \ldots.$$

There will exist integers i and j, $0 < i < j$, for which

$$(w_i, \sigma(w_i))(\leqq_L^r \times (\leqq_M^r)')(w_j, \sigma(w_j))$$

that implies $w_i \leqq_L^r w_j$ and $\sigma(w_i)(\leqq_M^r)'\sigma(w_j)$.

Let now x be a word such that $w_i x \in LM$. There exist words $v_1 \in L$ and $v_2 \in M$ such that $w_i x = v_1 v_2$.

Suppose first that $|v_1| > |w_i|$. In this case we have $v_1 = w_i y \in L$ and $x = y v_2$. Since $w_i \leqq_L^r w_j$ it follows $v_1 \leqq_L^r w_j y$ so that $w_j y \in L$. This implies $w_j y v_2 = w_j x \in LM$.

Let us now suppose that $|v_1| \leqq |w_i|$. One has $w_i = v_1\lambda$, $v_2 = \lambda x$ with $\lambda \in \sigma(w_i)$. Since $\sigma(w_i)(\leqq^r_M)'\sigma(w_j)$ there exists $\lambda' \in \sigma(w_j)$ such that $\lambda \leqq^r_M \lambda'$. Since $\lambda x \in M$ then $\lambda' x \in M$. By the definition of $\sigma(w_j)$ there exists $v \in L$ for which $w_j = v\lambda'$. Thus

$$w_j x = v\lambda' x \in LM.$$

Hence again $w_i x \in LM \Rightarrow w_j x \in LM$ so that in conclusion $w_i \leqq^r_{LM} w_j$. This shows that $\leqq^r_{LM}$ is a wqo. Q.E.D.

Proposition 3.8 *$\mathscr{F}_A$ is closed under the star operation.*

Proof. Let $L \in \mathscr{F}_A$. We define the map $\sigma_L : A^* \to \mathrm{Fin}(A^*)$, that we simply denote by σ, defined as: for any $w \in A^*$

$$\sigma(w) = (L^*)^{-1} w,$$

i.e. $\lambda \in \sigma(w)$ if and only if $w \in L^*\lambda$. Since any such λ is a suffix of w it is clear that $\sigma(w)$ is a finite set. Let us now prove that $\leqq^r_{L^*}$ is a wqo.

The relation $\leqq^r_L$ is a wqo on A^* so that, by Lemma 1.7, $\mathrm{Fin}(A^*)$ is well quasi-ordered by the relation $(\leqq^r_L)'$. Consider now an arbitrary infinite sequence of words: $w_1, w_2, \ldots, w_n, \ldots$. We can always suppose that all the words in the sequence are not empty. Let us now consider the infinite sequence of elements of $\mathrm{Fin}(A^*)$:

$$\sigma(w_1), \sigma(w_2), \ldots, \sigma(w_n), \ldots.$$

There will exist integers i and j, $0 < i < j$, for which

$$\sigma(w_i)(\leqq^r_L)'\sigma(w_j).$$

Let now x be a word such that $w_i x \in L^*$. We can factor $w_i x$ like as:

$$w_i x = u\lambda\mu\xi, \quad \text{where} \quad w_i = u\lambda, u \in L^*, \xi \in L^* \quad \text{and} \quad \lambda\mu \in \mathrm{L}.$$

Since $\lambda \in \sigma(w_i)$ there exists $\lambda' \in \sigma(w_j)$ such that $\lambda \leqq^r_L \lambda'$. One has then $\lambda\mu \leqq^r_L \lambda'\mu$, so that $\lambda'\mu \in L$. In view of the fact that $\lambda' \in \sigma(w_j)$ we can write:

$$w_j = v\lambda', \quad \text{with } v \in L^*.$$

Since $x = \mu\xi$, one derives $w_j x = v\lambda'\mu\xi \in L^*$. This shows that $w_i \leqq^r_{L^*} w_j$ so that $\leqq^r_{L^*}$ is a wqo. Q.E.D.

Proposition 3.9 *Let $\phi : A^* \to B^*$ be a non-erasing morphism, where A and B are finite alphabets. If $L \in \mathscr{F}_A$, then $\phi(L) \in \mathscr{F}_B$.*

Proof. Let $\phi : A^* \to B^*$ be a morphism such that ϕ is non-erasing, i.e. $\phi(A) \subseteq B^+$. We want to prove that if $L \in \mathscr{F}_A$, then $\phi(L) \in \mathscr{F}_B$. Let us denote by $P(\phi(A))$ the finite set of prefixes of the words of $\phi(A)$. For each word $u \in B^*$ and $\delta \in P(\phi(A))$ we consider the (possibly empty) set,

$$X(u, \delta) = \{\xi \in A^* \mid \phi(\xi)\delta = u\}.$$

We observe that $X(u, \delta)$ is a finite set. Indeed, if $\xi \in \mathrm{X}(\mathrm{u}, \delta)$, then, since the morphism ϕ is not erasing, one has $|\xi| \leqq |\phi(\xi)| \leqq |u|$.

By hypothesis $L \in \mathscr{F}_A$ so that $\leqq_L^r$ is a wqo. This implies by Lemma 1.7 that $(\leqq_L^r)'$ is a wqo in $\mathrm{Fin}(A^*)$. Moreover, for each k, the set V of k-tuples $(F_1, \ldots, F_k)$, with $F_i \in \mathrm{Fin}(A^*)$, $i=1, \ldots, k$, is also well quasi-ordered by the natural extension to V (cf. Proposition 1.2) of $(\leqq_L^r)'$. We want now prove that $\leqq_{L'}^r$, where $L' = \phi(L)$, is a wqo on B^*.

Let $u_1, u_2, \ldots, u_n, \ldots$ be an infinite sequence of words of B^*. We set $k = \mathrm{Card}(P(\phi(A)))$ and enumerate as $\delta_1, \delta_2, \ldots, \delta_k$ the elements of $P(\phi(A))$. We consider then the infinite sequence of k-vectors:

$$(X(u_j, \delta_1), \ldots, X(u_j, \delta_k)), \quad j > 0.$$

Since V is well quasi-ordered there exist integers i, j, $0 < i < j$, such that

$$X(u_i, \delta)(\leqq_L^r)' X(u_j, \delta), \quad \text{for each } \delta \in P(\phi(A)) \tag{3.1}$$

We want to prove that $u_i \leqq_{L'}^r u_j$. Indeed, suppose that $u_i \zeta \in L' = \phi(L)$; this means that there exists a word $w = a_1 \ldots a_n \in L$, $a_i \in A$, $1 \leqq i \leqq n$, such that $u_i \zeta = \phi(a_1) \ldots \phi(a_n)$. Hence there exists an integer r, $0 \leqq r \leqq n-1$, and $\delta, \eta \in A^*$ such that:

$$u_i = \phi(a_1 \ldots a_r)\delta, \quad \zeta = \eta\, \phi(a_{r+2} \ldots a_n), \quad \delta\eta = \phi(a_{r+1}),$$

where, conventionally, when $r=0$ (resp. $r=n-1$) the word $a_1 \ldots a_r$ (resp. $a_{r+2} \ldots a_n$) is equal to the empty word of A^*.

Let us set $\xi = a_1 \ldots a_r$. One has $u_i = \phi(\xi)\delta$, with $\delta \in P(\phi(A))$. Hence $\xi \in X(u_i, \delta)$. By Eq. (3.1) there exists $\xi' \in X(u_j, \delta)$ such that $\xi \leqq_L^r \xi'$. Since $\xi a_{r+1} \ldots a_n \in L$ it follows that $\xi' a_{r+1} \ldots a_n \in L$. This implies:

$$\phi(\xi')\delta\eta\,\phi(a_{r+2}) \ldots \phi(a_n) = \phi(\xi')\delta\zeta = u_j \zeta \in \phi(L).$$

Hence $\leqq_{L'}^r$ is a wqo and $\phi(L) \in \mathscr{F}_B$. Q.E.D.

Two words $x, y \in A^*$ are *conjugate* if there exist words $u, v \in A^*$ such that $x = uv$ and $y = vu$. For any $x \in A^*$ and $L \subseteq A^*$, denote by $c(x)$ the set of all words of A^* which are conjugate to x and by $c(L)$ the *conjugation closure* of L, i.e. $c(L) = \bigcup_{x \in L} c(x)$. A language is closed under conjugation if $L = c(L)$. Note that a commutative language is certainly closed under conjugation.

Proposition 3.10 *A language $L \in \mathscr{F}_A$ closed under conjugation is regular.*

Proof. Let us prove that $\leqq_L^r = \leqq_L^l$. Indeed, let $u, v \in A^*$ be such that $u \leqq_L^r v$. For any $\zeta \in A^*$ one has $\zeta u \in L \Rightarrow u\zeta \in L \Rightarrow v\zeta \in L \Rightarrow \zeta v \in L$, i.e. $u \leqq_L^l v$. By Theorem 2.4 the result follows. Q.E.D.

Lemma 3.11 *If $L \in \mathscr{F}_A$ and $A \subseteq B$, then $L \in \mathscr{F}_B$.*

Proof. We denote by $\leqq_{L,A}^r$ and $\leqq_{L,B}^r$ the Nerode quasi-orders of L on A^* and on B^*, respectively. Since $L \in \mathscr{F}_A$, then $\leqq_{L,A}^r$ is a wqo on A^*. We want to prove that $\leqq_{L,B}^r$ is a wqo on B^*. Let us first observe that if $u \in A^*$ then $\{w \in B^* \mid uw \in L\} = \{w \in A^* \mid uw \in L\}$. Hence if $u, v \in A^*$ and $u \leqq_{L,A}^r v$ then $u \leqq_{L,B}^r v$. Let $u_1, u_2, \ldots, u_n, \ldots$ be an infinite sequence of words of B^*. If there exists a word $u_i \notin A^*$, then $\{w \in B^* \mid u_i w \in L\} = \emptyset$ so that $u_i \leqq_{L,B}^r u_j$ for any $j > i$. Suppose then that $u_i \in A^*$, for any $i \geq 1$. Since $\leq^r_{L,A}$ is a wqo, there exist integers i, j such that $0 < i < j$ and $u_i \leq^r_{L,A} u_j$. This implies $u_i \leq^r_{L,B} u_j$. Q.E.D.

Let us denote by $\mathscr{F}$ the class of all languages of $\mathscr{F}_A$ where A is any finite subset of a given fixed infinite alphabet. $\mathscr{F}$ is a *family* of languages (cf. [1, 2]). Indeed, one can easily derive (for instance by Proposition 3.9) that if $\phi: A^* \to B^*$ is an isomorphism and $L \in \mathscr{F}_A$, then $\phi(L) \in \mathscr{F}_B$.

Proposition 3.12 *The family $\mathscr{F}$ is closed under the following operations: union, intersection, product, star, inverse morphisms, direct non-erasing morphisms, intersection with regular sets.*

Proof. The closure of $\mathscr{F}$ under star, inverse morphisms and direct non-erasing morphisms is given by Propositions 3.6, 3.8 and 3.9. The closure under union, intersection and product is obtained as follows: let $L \in \mathscr{F}_A$ and $M \in \mathscr{F}_B$; one considers an alphabet C such that $L, M \subseteq C^*$. By Lemma 3.11 it follows that $L, M \in \mathscr{F}_C$. Hence by Propositions 3.1 and 3.7, $L \cup M, L \cap M, LM \in \mathscr{F}_C$ and the result follows. Let $L \in \mathscr{F}$ and $M \in \mathrm{Rat}$. Since $\mathrm{Rat} \subseteq \mathscr{F}$, then, by the closure under intersection, one has $L \cap M \in \mathscr{F}$. Q.E.D.

Let us recall that a *rational transduction* $\mathbf{t}: A^* \to B^*$ is a rational subset of $A^* \times B^*$. A transduction $\mathbf{t}$ is called *continuous* (or Λ-free) if $\mathbf{t}(A^+) \subseteq B^+$; $\mathbf{t}$ is called *faithful* if for any $f \in B^+$ one has $\mathbf{t}^{-1}(f)$ is a finite subset of A^*. We recall (cf. [1, 2]) that a *cone* (*quasi-cone*) is a family of languages which is closed under rational transductions (faithful rational transductions). A cone (quasi-cone) which is rationally closed is called a "full AFL" (AFL).

Corollary 3.13 *The family $\mathscr{F}$ is an* AFL.

Proof. From Propositions 3.1, 3.7 and 3.8 one has that $\mathscr{F}$ is a rationally closed family of languages. Moreover, from Propositions 3.6 and 3.9 one has that $\mathscr{F}$ is closed under inverse and direct non-erasing morphisms. Since $\mathscr{F}$ is closed under intersection and contains the regular languages, one has that $\mathscr{F}$ is closed under intersection with rational languages. This implies that $\mathscr{F}$ is closed under continuous and faithful rational transductions. Moreover, since $\mathscr{F}$ is rationally closed, then it contains languages having the empty string. Thus from [1, pag. 68], it follows that $\mathscr{F}$ is closed under faithful rational transductions. This implies that $\mathscr{F}$ is an AFL. Q.E.D.

4. A pumping lemma

Let L be a language over A and $P(L)$ the set of the prefixes of the words of L. If $\mathbf{x}: \mathbb{N} \to A$ is an infinite word, then $P(\mathbf{x})$ will denote the set of all prefixes of $\mathbf{x}$ of finite length. The following lemma (whose proof is part of the folklore (cf. [9]) holds:

Lemma 4.1 *Let L be an infinite language over the finite alphabet A. There exists an infinite word $\mathbf{x}: \mathbb{N} \to A$ such that $P(\mathbf{x}) \subseteq P(L)$.*

Let w be a finite or infinite word, we denote by $w[i]$ the prefix of w of length i; when w is a finite word then $w[i]$ is defined only for $0 \leq i \leq |w|$.

Proposition 4.2 *Let $L \in \mathscr{F}$. There exists an integer N_L such that if $w \in A^*$ and $|w| \geqq N_L$, then there exist integers i, j, $0 < i < j < |w|$, and such that*

$$w[i] \leqq_L^r w[j].$$

Proof. Suppose by contradiction, that there exist infinitely many words:

$$w_1, w_2, \ldots, w_n, \ldots$$

such that for any $n > 0$ and integers i, j, $0 < i < j \leqq |w_n|$, one has

$$w_n[i] \nleqq_L^r w_n[j].$$

Let us denote by W the language $W = \bigcup_{i \in \mathbb{N}} \{w_i\}$. By Lemma 4.1 there exists an infinite word $\mathbf{x} : \mathbb{N} \to A$ such that $P(\mathbf{x}) \subseteq P(W)$. Let

$$\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[n], \ldots$$

be the sequence of all prefixes of $\mathbf{x}$ of positive length. Since $\leqq_L^r$ is a well quasi-order there exist integers i, j, $0 < i < j$, such that $\mathbf{x}[i] \leqq_L^r \mathbf{x}[j]$. Now $\mathbf{x}[i]$ is a prefix of $\mathbf{x}[j]$ and $\mathbf{x}[j]$ is a prefix of a word $w \in W$. Hence $w[i] \leqq_L^r w[j]$ and we reach a contradiction. Q.E.D.

Proposition 4.3 *Let $L \in \mathscr{F}$. There exists an integer N_L such that if $w \in A^*$ and $|w| \geqq N_L$, then there exist integers i, j, $0 < i < j \leqq N_L$, such that $w = p\zeta s$ with $|\zeta| = j - i$, $|p| < N_L$, $s \in A^*$, and for all $v \in A^*$*

$$pv \in L \Rightarrow p\zeta v \in L \quad \text{and} \quad p\zeta v \in L \Rightarrow p\zeta^n v \in L, \quad \text{for all } n > 0.$$

Proof. Let us write $w = w'h$ with $|w'| = N_L$ and $h \in A^*$. From the preceding proposition there exist integers i, j, $0 < i < j \leqq |w'|$, for which

$$w'[i] \leqq_L^r w'[j].$$

Let us set $p = w'[i]$, $w'[j] = p\zeta$ with $|p| = |w'[i]| < N_L$ and $|\zeta| = j - i$. We can then write $w = p\zeta s$, $s \in A^*$. Moreover, since $\leqq_L^r$ is right-monotone, one has that $p \leqq_L^r p\zeta$ and $p\zeta \leqq_L^r p\zeta^n$ for all $n > 0$. This implies then Eq. (3.2). Q.E.D.

Corollary 4.4 *Let $L \in \mathscr{F}$. There exists an integer N_L such that if $w \in L$ and $|w| \geqq N_L$, then w can be factorized as $w = uvt$, with $v \neq A$, $|u| < N_L$ and*

$$uv^+ t \subseteq L.$$

Proof. Obvious from Proposition 4.3. Q.E.D.

Let us observe that in the statement of the preceding pumping lemma the "pump" is positive (i.e. it starts with $n = 1$), and, moreover, differently from the 'classical' formulation for regular languages, it is located 'near' the initial parts of the words of the language (and not in any sufficiently large factor). For instance in the case of the language $L = \{a^n b^m \mid n \geqq m \geqq 0\}$ if we take a sufficiently large word $a^n b^m$ with $n \geqq m$, then the pump can be located only in the prefix a^n. As an application of the pumping lemma we derive that the language

$L=\{a^n b^n | n \geqq 0\}$ does not belong to $\mathscr{F}$. In fact, if one takes $n \geqq N_L$, then there would exist integers i and j such that $0<i<j \leqq n$ and $a^i \leqq_L^r a^j$. This implies $a^i b^i \leqq_L^r a^j b^i$. Since $a^i b^i \in L$ we reach the contradiction $a^j b^i \in L$.

5. Non-recursively enumerable languages of $\mathscr{F}$

In this section we prove that $\mathscr{F}$ contains languages which are not recursively enumerable. Let f be any non-decreasing function $f: \mathbb{N} \to \mathbb{N}$; we associate to it the language over the alphabet $A=\{a, b\}$:

$$L_f=\{a^n b^m | f(n) \geqq m \geqq 0\}.$$

Proposition 5.1 *For any non-decreasing function f: $\mathbb{N} \to \mathbb{N}$, $L_f \in \mathscr{F}$.*

Proof. Let us simply denote L_f by L. We prove that the relation $\leqq_L^r$ is a wqo. Let σ be an infinite sequence of words of $\{a, b\}^*$:

$$x_1, \ldots, x_n, \ldots.$$

We show that there exist integers $i, j, 0<i<j$, such that $x_i \leqq_L^r x_j$. Let us consider the complement L^c of the set L

$$L^c=b^+ A^* \cup \{a^n b^m | f(n)<m\} \cup a^+ b^+ a A^*;$$

for any pair $y_1, y_2 \in L^c$, $y_1^{-1} L=y_2^{-1} L=\emptyset$ so that $y_1 \mathfrak{R}_L^r y_2$. If in σ there are two elements of the set L^c, then the result is, trivially, achieved. We suppose then that all the elements of σ belong to L. We consider first the case when in σ there are infinitely many elements which belong to the set $a^* \subseteq L$. One has that there exist at least two elements $x_i=a^p$ and $x_j=a^q$ such that $0<i<j$ and $p<q$. We prove that $x_i \leqq_L^r x_j$. Indeed, if $a^p \zeta \in L$, then $\zeta=a^s b^r$ and $f(p+s) \geqq r$. Let us then consider $a^q \zeta=a^{q+s} b^r$; since f is non-decreasing $f(q+s) \geqq f(p+s) \geqq r$, so that $a^q \zeta \in L$. Thus in this case the result is achieved.

We have now to consider the case that for a sufficiently large n all the elements of σ belong to $L \backslash a^*$. One may suppose then, without loss of generality, that all the elements of σ belong to $L \backslash a^*$. To each word $x=a^i b^j \in L$, $f(i) \geqq j>0$, we can associate the non-negative integer $d(x)=f(i)-j$. Let us then consider the sequence of integers:

$$d(x_1), d(x_2), \ldots, d(x_n), \ldots;$$

since $\mathbb{N}$ is well-ordered w.r.t. $\leqq$ there exists always a pair of integers m, n such that $0<m<n$ and $d(x_m) \leqq d(x_n)$. Let $x_m=a^i b^j$ and $x_n=a^p b^q$; one has then $f(i)-j \leqq f(p)-q$. Let $\zeta \in A^*$ be such that $x_m \zeta \in L_f$. This implies $\zeta=b^r$ for a suitable $r \geqq 0$ and $a^i b^j b^r \in L_f$. One has then $f(i) \geqq j+r$. Since $f(p)-q \geqq f(i)-j \geqq r$ it follows that $a^p b^q b^r \in L$. Hence $x_i \leqq_L^r x_j$. Q.E.D.

If we take the map f equal to the identity (i.e. $f(n)=n$, for all $n \geqq 0$), then we obtain that the non-regular context-free language $L=\{a^n b^m | n \geqq m \geqq 0\}$ is such that $\leqq_L^r$ is a wqo.

Let us denote by $F(\mathbb{N})$ the set of all non-decreasing maps from $\mathbb{N}$ to $\mathbb{N}$.

Lemma 5.2 *Let $f, g \in F(\mathbb{N})$ and $f \neq g$ then $L_f \neq L_g$.*

Proof. If $f \neq g$ then there exists an integer i such that $f(i) \neq g(i)$. Suppose that $g(i) > f(i)$. We have that $a^i b^{g(i)} \in L_g$. However, $a^i b^{g(i)} \notin L_f$. Indeed, otherwise, $f(i) \geqq g(i)$ which is a contradiction. Q.E.D.

Corollary 5.2 *There are languages in the family $\mathscr{F}_A$, $A = \{a, b\}$, which are not recursively enumerable.*

Proof. By the Cantor theorem the set $F(\mathbb{N})$ is not enumerable. Hence by Lemma 5.2 the class $\mathscr{C}$ of languages L_f with $f \in F(\mathbb{N})$ is not enumerable. By Proposition 5.1, $\mathscr{C} \subseteq \mathscr{F}_A$, so that $\mathscr{F}_A$ is also not enumerable. Since, as it is well known, the class of recursively enumerable languages is enumerable (by enumerating the Turing machines), we obtain that $\mathscr{F}_A$ contains languages which are not recursively enumerable. Q.E.D.

6. Concluding remarks

In the previous section we saw that the non-regular context-free language $L = \{a^n b^m \mid n \geqq m \geqq 0\}$ is in the class $\mathscr{F}$; now we consider a larger class of context-free and non-regular languages contained in $\mathscr{F}$. For each $n > 0$ let $Z_n = A_n \cup A'_n$, where $A_n = \{a_1, \ldots, a_n\}$ and $A'_n = \{a'_1, \ldots, a'_n\}$. We consider the language $L_n \subseteq Z_n^*$, where

$$L_n = \{w \in Z_n^* \mid \text{for each prefix } u \text{ of } w \text{ and } 1 \leqq i \leqq n \text{ one has } |u|_{a_i} \geqq |u|_{a'_i}\}$$

Using an argument similar to that of the proof of Proposition 5.1 and Lemma 1.3, one can prove the following:

Proposition 6.1 *For each positive integer n, L_n is a language of the class $\mathscr{F}$.*

Let us consider in Z_n^* the congruence δ'_n generated by the relations

$$a_k a'_k \sim 1, \qquad (k = 1, \ldots, n).$$

We recall that the restricted Dyck language $D_n'^*$ is defined as

$$D_n'^* = \{w \in Z_n^* \mid w \equiv 1, \text{ modulo } \delta'_n\}.$$

Moreover $D_n'^*$ can also be defined as follows (cf. [2], pag. 48, Corollary 4.2): let $Z_1 = \{a, a'\}$, then

$$D_1'^* = \{w \in Z_1^* \mid |w|_a = |w|_{a'} \text{ and for each prefix } u \text{ of } w, |u|_a \geqq |u|_{a'}\}.$$

Hence the set of prefixes $D_1'^*$ coincides with L_1. In other words $P(D_1'^*) \in \mathscr{F}$.

The cone and the quasi-cone generated by $D_1'^*$ are the families of one counter languages (OCL) and restricted one counter languages (ROCL), respectively (cf. [2]). This lead us to formulate the following:

Conjecture. Let P(ROCL) the family of all the languages of the kind $P(L)$ with L in ROCL. Then $P(\text{ROCL}) \subseteq \mathscr{F}$.

We remark that, if one could prove that P(ROCL) is the quasi-cone generated by $P(D_1'^*)$, as it seems reasonable, then the conjecture would follow from the closure of $\mathscr{F}$, with respect to the faithful rational transductions.

References

1. Autebert, J.M., Boasson, L.: Transductions rationnelles. Applications aux languages algébriques. Paris: Masson 1988
2. Berstel, J.: Transductions and context-free languages. Stuttgart: Teubner 1979
3. Clifford, A.H., Preston, G.B.: The algebraic theory of semigroups, vol. 1. (Math. Surveys Monogr., No 7) Providence, RI: American Methematical Society 1961
4. Ehrenfeucht, A., Haussler, D., Rozenberg, G.: On regularity of context-free languages. Theor. Comput. Sci. 27, 311–332 (1983)
5. Eilenberg, S.: Automata, languages and machines, vol. A. New York: Academic Press 1974
6. Higman, G.H.: Ordering by divisibility in abstract algebras. Proc. London Math. Soc. **3**, 326–336 (1952)
7. Kruskal, J.: The theory of well-quasi-ordering: a frequently discovered concept. J. Comb. Theory Ser. A **13**, 297–305 (1972)
8. Latteux, M., Rozenberg, G.: Commutative one-counter languages are regular. J. Comput. Syst. Sci. **29**, 54–57 (1984)
9. Lothaire, M.: Combinatorics on words. Reading, Mass: Addison Wesley 1983
10. Luca, A. de, Varricchio, S.: Finiteness and iteration conditions for semigroups. Theor. Comput. Sci. **87**, 315–327 (1991)
11. Luca, A. de, Varricchio, S.: Some regularity conditions based on well quasi-orders (Lect. Notes Comput. Sci., vol. 583, pp. 356–371) Berlin, Heidelberg, New York: Springer 1992
12. Restivo, A., Reutenauer, C.: Some applications of a theorem of Shirshov to language theory. Inf. Control **57**, 205–213 (1983)
13. Simon, I.: Piecewise testable events (Lect. Notes Comput. Sci., vol. 33, pp. 214–222) Berlin, Heidelberg, New York: Springer 1975