ON TOTAL REGULATORS GENERATED

BY DERIVATION RELATIONS

by

W. Bucher*, A. Ehrenfeucht and D. Haussler**

CU-CS-301-85 September, 1984

All correspondence to W. Bucher.

*Institutes for Information Processing, Technical University of Graz, A-8010, Graz, Austria.

University of Colorado, Department of Computer Science, Boulder, Colorado.

**Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208.

On Total Regulators Generated by Derivation Relations [1]

W. Bucher, A. Ehrenfeucht* and D. Haussler**

Institutes for Information Processing, Technical University of Graz, A-8010 Graz, Austria.

*Department of Computer Science, University of Colorado, Boulder, Colorado 80302, USA.

**Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208, USA.

All correspondence to W. Bucher.

Abstract A derivation relation is a total regulator on Σ^* if for every language $L \subseteq \Sigma^*$, the set of all words derivable from L is a regular language. We show that for a wide class of derivation relations $\overset{*}{\underset{P}{\Rightarrow}}$, $\overset{*}{\underset{P}{\Rightarrow}}$ is a total regulator on Σ^* if and only if it is a well-quasi-order (wqo) on Σ^*. Using wqo theory, we give a characterization of all non-erasing pure context-free (OS) derivation relations which are total regulators.

Keywords: formal languages, regular languages, context-free languages, well-quasi-orders, unavoidable sets, derivation.

[1] The authors gratefully acknowledge the support of NSF grants IST-8317918 and MCS-8305245, and the Austrian Bundesministerium fuer Wissenschaft und Forschung. Part of this work was conducted while the third author visited the Institutes for Information Processing at the Technical University of Graz and the other part while the first author visited the University of Denver. We would like to thank our respective host institutions for these generous invitations.

Introduction.

While most results on finite automata and regular languages are constructive in the sense that the machines and expressions involved can be effectively given, occasionally one comes across a completely non-constructive result. An example is the following result of Haines ([*HAI*]). We say that a word y is a supersequence of a word x if the sequence of letters of y contains the sequence of letters of x as a subsequence. For any language L, consider the language of all words (over a fixed alphabet) which are supersequences of words in L. This language is always regular. Thus, using J. H. Conway's terminology ([*CON*]), the operation of closing a language by adding all words which are supersequences of words in the language is a *total regulator*, since it converts *any* language L into a regular language. For an arbitrary recursive language L this construction cannot be effective, since this would allow us to solve the emptiness problem for recursive languages ([*LEE*]).

In this paper we look further into Conway's notion by investigating total regulators generated by closure under the more common types of derivation relations in Formal Language Theory. For any particular derivation relation $\overset{*}{\underset{P}{=}}>$ defined on words over an alphabet Σ, we will say that $\overset{*}{\underset{P}{=}}>$ is a total regulator on Σ^* if for any $L \subseteq \Sigma^*$, the language of words derived from words in L by $\overset{*}{\underset{P}{=}}>$ is a regular language. Haines' result can be easily cast in this form. For example, if $\Sigma = \{a, b\}$, $P = \{a \to aa \mid ab \mid ba, b \to bb \mid ba \mid ab\}$ is a pure context-free production system (OS scheme) then for any $x, y \in \Sigma^+$, y is a supersequence of x if and only if $x \overset{*}{\underset{P}{=}}> y$.

Haines' result can be derived from earlier results in the theory of well-quasi-orders, given in Higman's seminal paper ([*HIG*]). In [*EHR*], a more general connection between regularity and well-quasi-orders is exhibited, and a generalized version of the Haines/Higman result is given in terms of derivation by repeated insertion of words chosen from a fixed, unavoidable set. Here we carry these results further be showing that for a wide class of derivation relations, including those generated by propagating (non-erasing) OS schemes, $\overset{*}{\underset{P}{=}}>$ is a total regulator on Σ^* if and only if $\overset{*}{\underset{P}{=}}>$ is a well-quasi-order on Σ^* (Theorem 1.1). We then characterize the OS schemes which generate well-quasi-orders on Σ^* using the notion of unavoidability as defined in [*EHR*] (Theorem 2.1). The generalized Haines/Higman result from [*EHR*] is easily obtained as a corollary of this characterization. Another combinatorial result that follows from Theorem

2.1 is given at the end of Section 2. In Section 3 we give some preliminary results toward a more algebraic characterization of OS total regulators.

Several applications of the theory of well-quasi-orders have recently appeared in the literature ([*RUO*], [*LAT*], [*DER*]). It is hoped that the basic results on well-quasi-orders given here will lead to further applications of the theory in these and other areas. In this context, we note that [*LAT*] uses the Haines/Higman result, which is a special case of our characterization theorem, and the main regularity result from [*RUO*] can be derived from the fact that for the OS scheme with $\Sigma = \{a, b\}$ and $P = \{a \to aa \mid aba, b \to bb \mid bab\}$, $\overset{*}{\underset{P}{\Longrightarrow}}$ is a total regulator, which also follows directly from this theorem.

Several immediate directions for further research remain. These are discussed in detail in Section 4. The primary open problem is whether or not the characterization of propagating OS total regulators given by Theorems 1.1 and 2.1 is effective (i.e. is the criterion given in Theorem 2.1 a decidable property of propagating OS schemes). In addition, even if we can establish that $\overset{*}{\underset{P}{\Longrightarrow}}$ is a total regulator by showing that P satisfies the criterion of Theorem 2.1, the regular languages generated by applying this total regulator can not always be effectively given, as mentioned above. J. van Leeuwen ([*LEE*]) has explored the extent to which the Haines/Higman total regulator is effective, and demonstrated that the closure of any context-free language under this total regulator is an effectively given regular language. We have no similar results for an arbitrary propagating OS total regulator $\overset{*}{\underset{P}{\Longrightarrow}}$. In fact, even when R is the regular language language derived from a single letter $a \in \Sigma$ under $\overset{*}{\underset{P}{\Longrightarrow}}$, we cannot give any recursive bounds on the size of the smallest automaton for R in terms of the size of P. It remains to be seen if the non-constructiveness in our results is merely an artifact of our choice of methods or whether it indicates some deeper intractability of the problem.

Notation

For basic definitions in Formal Language Theory we refer the reader to [*HAR*]. Our conventions are as follows. For a finite alphabet Σ, Σ^* denotes the set of words over Σ, λ denotes the empty word and $\Sigma^+ = \Sigma^* - \{\lambda\}$. For $w \in \Sigma^*$, $|w|$ denotes the length of w and $\#_a(w)$ the number of a's in w for any $a \in \Sigma$. A *production system* is a pair (Σ,P) where P is a finite set of productions $P = \{u_1 \to v_1, \ldots, u_k \to v_k\}$ where $u_i \in \Sigma^+$, $v_i \in \Sigma^*$ for $1 \le i \le k$. If for all i, $1 \le i \le k$, $|u_i| \le |v_i|$, (Σ,P) is *propagating* (*length–increasing*); if $|u_i| < |v_i|$ then (Σ,P) is *strictly propagating*; if $|u_i| = 1$ then (Σ,P) is an *OS scheme*.

$u \to v_1 \mid v_2 \mid \cdots \mid v_k$ is shorthand for $u \to v_1, u \to v_2, \ldots, u \to v_k$. $RHS_P(u) = \{v : u \to v \in P\}$. $RHS_P = \{v : u \to v \in P$ for some $u\}$. If $x = x_1ux_2$ and $y = x_1vx_2$, where $x_1, x_2 \in \Sigma^*$ and $u \to v \in P$, then $x \overset{=}{\underset{P}{=}}> y$. $\overset{*}{\underset{P}{=}}>$ denotes the reflexive and transitive closure of $\overset{=}{\underset{P}{=}}>$.

Section 1. Well-quasi-orders and total regulators

We begin by defining the notion of a total regulator, and characterizing this class of relations using the theory of well-quasi-orders (see below). We will restrict ourselves to relations of the following type, which includes many of the common types of derivation relations in Formal Language Theory.

Definition 1.1. A *quasi-order* is a reflexive and transitive relation. A quasi-order $\leq$ on Σ^* is *multiplicative* if for all $x_1, x_2, y_1, y_2 \in \Sigma^*$, $x_1 \leq x_2$ and $y_1 \leq y_2$ implies that $x_1y_1 \leq x_2y_2$. The quasi-order $\leq$ is *length-increasing* if $x \leq y$ implies that $|x| \leq |y|$.

Example 1.1. Let (Σ,P), with $P = \{u_1 \to v_1, \ldots, u_k \to v_k\}$, be a finite production system. Then $\overset{*}{\underset{P}{=}}>$ is a multiplicative quasi-order on Σ^*. If P is length-increasing then $\overset{*}{\underset{P}{=}}>$ is length-increasing.

Definition 1.2. For a quasi-order $\leq$ on Σ^*, $w \in \Sigma^*$ and $L \subseteq \Sigma^*$, let $cl_{\leq}(w) = \{x \in \Sigma^* : w \leq x\}$, $cl_{\leq}(L) = \bigcup_{y \in L} cl_{\leq}(y)$. If $\leq$ is the derivation relation $\overset{*}{\underset{P}{=}}>$ defined by some OS scheme (Σ,P), we write $cl_P(w)$ for $cl_{\leq}(w)$, similarly $cl_P(L)$ for $cl_{\leq}(L)$. The quasi-order $\leq$ is a *regulator* (on Σ^*) if $cl_{\leq}(L)$ is regular for all regular $L \subseteq \Sigma^*$, $\leq$ is a *total regulator* (on Σ^*) if $cl_{\leq}(L)$ is regular for any $L \in \Sigma^*$. A (total) regulator of the form $\overset{*}{\underset{P}{=}}>$, where (Σ,P) is an OS scheme, is also called an *OS* (total) regulator. It is a *propagating* OS (total) regulator if the OS scheme is propagating.

By the results of Haines ([*HAI*]), the supersequence relation given in the Introduction is one example of a propagating OS total regulator, but much simpler examples can be given.

Example 1.2. Let $\Sigma = \{a, b\}$ and let $P = \{a \to b, b \to a \mid bb\}$. Then for any $x, y \in \Sigma^+$, $x \overset{*}{\underset{P}{=}}> y$ if and only if $|x| \leq |y|$. Thus for nonempty $L \subseteq \Sigma^+$, $cl_P(L) = T = \{x \in \Sigma^* : |x| \geq k\}$ where k is the length of the shortest word of L. For $L' = L \cup \{\lambda\}$, $cl_P(L') = T \cup \{\lambda\}$. Hence $\overset{*}{\underset{P}{=}}>$ is a propagating OS total regulator.

An OS total regulator which is not propagating was also given by Haines.

Example 1.3. Let $\Sigma = \{a, b\}$ and let $P = \{a \to \lambda, b \to \lambda\}$. Then for any $x, y \in \Sigma^+$, $x \overset{*}{\underset{P}{=}}> y$ if and only if x is a supersequence of y. This quasi-order is the inverse of the supersequence total regulator discussed in the Introduction, and is also a total regulator by the results of Haines ([*HAI*]). In fact, Haines' argument generalizes to show that the inverse of any total regulator is also a total regulator.

Haines' results can easily be derived from the more general theory of well-quasi-orders, introduced by Higman ([*HIG*]). We give only the basic definitions and results from this theory which will be needed in what follows. For a more complete treatment, the reader is refered to [*KRU*].

Definition 1.3. A quasi-order $\leq$ on a set S is a *well−quasi−order* (*wqo*) on S if and only if for each infinite sequence $\{x_i\}_{i \geq 1}$ of elements in S, there exist $i < j$ such that $x_i \leq x_j$.

Proposition 1.1. ([*HIG*]) Let $\leq$ be a wqo on a set S and let $\leq^E$ be the quasi-order on the set $F(S)$ of finite sequences of elements from S, defined by $<s_1,\ldots,s_k> \; \leq^E \; <t_1,\ldots,t_l>$ if and only if there exists a subsequence $<t_{i_1},\ldots,t_{i_k}>$ of $<t_1,\ldots,t_l>$ such that $s_j \leq t_{i_j}$ for $1 \leq j \leq k$. Then $\leq^E$ is a wqo on $F(S)$.

Proposition 1.2. ([*EHR*]) Let $\leq$ be a quasi-order on Σ^* which is wqo on $L_1, L_2 \subseteq \Sigma^*$. Then $\leq$ is a wqo on $L_1 \cup L_2$ and if $\leq$ is multiplicative, then $\leq$ is a wqo on L_1L_2.

Proposition 1.3. Let $\leq_1$ be a wqo on a set S. If $\leq_2$ is a quasi-order on S such that $x \leq_1 y$ implies that $x \leq_2 y$, then $\leq_2$ is a wqo on S.

Proof. This follows directly from the definition. ▪

Proposition 1.4. Let $\leq$ be a multiplicative quasi-order on Σ^*, and $x_1,\ldots,x_k, y_1,\ldots,y_k$ be words in Σ^+ such that $x_i \leq y_i$ holds for $1 \leq i \leq k$. If $\overset{*}{\underset{P}{=}}>$ is a wqo on Σ^* for the production system (Σ, P), where $P = \{x_1 \to y_1, \cdots, x_k \to y_k\}$, then $\leq$ is a wqo on Σ^*.

Proof. This follows easily from Proposition 1.3. ▪

In [*EHR*], a generalized Myhill/Nerode theorem for regular languages is given wherein the usual notion of a finite congruence on Σ^* is replaced by that of a multiplicative wqo on Σ^* (here our terminology varies slightly from that of [*EHR*]). A consequence of this result is the following.

Proposition 1.5. For any multiplicative wqo $\leq$ on Σ^*, $\leq$ is a total regulator on Σ^*.

For a wide class of derivation relations, this result actually provides a characterization of the total regulators, as is shown in the following.

Theorem 1.1. If $\leq$ is a length-increasing, multiplicative, decidable quasi-order on Σ^*, then $\leq$ is a total regulator on Σ^* if and only if $\leq$ is a wqo on Σ^*.

Proof. The "if" part follows from Proposition 1.5. For the "only if" part, assume that $\leq$ is a total regulator, but not a wqo on Σ^*. Since $\leq$ is not a wqo on Σ^*, there exists an infinite sequence $\{x_i\}_{i \geq 1}$ of words over Σ such that for no pair i, j of numbers, where $1 \leq i < j$, $x_i \leq x_j$ holds. Let $L = \{x : x = x_i$ for some $i \geq 1\}$ and let $X = \{|x| : x \in L\}$. By considering a subsequence of $\{x_i\}_{i \geq 1}$, if necessary, we can assume that $|x_i| < |x_j|$ whenever $i < j$ and that X is not a recursive set of natural numbers.

Since $\leq$ is a total regulator on Σ^*, $cl_{\leq}(L)$ is a regular set and hence it is decidable for any word x if $x \in cl_{\leq}(L)$. Let $Y = \{n \in N :$ there exists $w \in cl_{\leq}(L)$ with $|w| = n$ and for no $y \in cl_{\leq}(L)$ with $|y| < n$ the relation $y \leq w$ holds$\}$. Since $\leq$ is decidable and $cl_{\leq}(L)$ is recursive, Y is recursive. We claim that $X = Y$ and this contradiction establishes the theorem.

The claim is established as follows. If $w \in cl_{\leq}(L)$ then $x_j \leq w$ for some $j \geq 1$. If in addition there is no $y \in cl_{\leq}(L)$ such that $|y| < |w|$ and $y \leq w$, then there is no x_i such that $|x_i| < |w|$ and $x_i \leq w$. Since $\leq$ is length increasing, this implies that $|w| = |x_j|$, hence $|w| \in X$. On the other hand, for any x_j, $j \geq 1$, $x_j \in cl_{\leq}(L)$. Furthermore, there is no $y \in cl_{\leq}(L)$ such that $|y| < |x_j|$ and $y \leq x_j$, because this would imply that $x_i \leq y \leq x_j$ for some $i < j$, which is impossible by our assumption on $\{x_i\}_{i \geq 1}$. ■

In fact, since in the proof of the preceding theorem the regularity of $cl_{\leq}(L)$ is only needed to show that $cl_{\leq}(L)$ is recursive, the proof shows that the following stronger statement holds.

Theorem 1.2. If $\leq$ is a length-increasing, multiplicative, decidable quasi-order on Σ^*, then the following three properties are equivalent.

(i) $\leq$ is a wqo on Σ^*.

(ii) $\leq$ is a total regulator on Σ^*.

(iii) $cl_{\leq}(L)$ is recursive for every subset L of Σ^*.

Section 2. The Main Theorem

We now restrict our attention to derivation relations generated by OS schemes. Since, for propagating OS schemes, these relations fall into the general category of relations covered by Theorem 1.1, we know that a derivation relation of this type is a total regulator if and only if it is a wqo on Σ^*. Therefore we investigate the circumstances under which an OS scheme generates a wqo on Σ^*. In the case of propagating schemes, this leads to a characterization of the total regulators. We need the following concepts.

Definition 2.1. A subset L of Σ^+ is *unavoidable*, if there exists a number $k_0 \in \mathbf{N}$ such that for all $w \in \Sigma^*$, $|w| > k_0$, w has a subword in L, i.e. $w = w_1 x w_2$ for some $w_1, w_2 \in \Sigma^*$, $x \in L$. The smallest such number k_0 is called the *avoidance bound* for L.

It is clear from the definition that if L is unavoidable with avoidance bound k_0, then $\{x \in L : |x| \le k_0\}$ is also unavoidable with avoidance bound k_0. Hence any infinite unavoidable language contains a finite unavoidable subset.

Definition 2.2. Let (Σ, P) be an OS scheme. Then, for $a \in \Sigma$,

$LEFT_P(a) = \{ax : x \in \Sigma^+, \text{ and } a \overset{*}{\underset{P}{\Rightarrow}} ax\}$,

$RIGHT_P(a) = \{xa : x \in \Sigma^+, \text{ and } a \overset{*}{\underset{P}{\Rightarrow}} xa\}$,

$DUAL_P(a) = LEFT_P(a) \cap RIGHT_P(a) = \{axa : x \in \Sigma^*, a \overset{*}{\underset{P}{\Rightarrow}} axa\}$,

$MIXED_P(a) = LEFT_P(a) \cup RIGHT_P(a)$,

$LEFT_P = \bigcup_{a \in \Sigma} LEFT_P(a)$ and $RIGHT_P$, $DUAL_P$, and $MIXED_P$ are defined similarly.

Theorem 2.1. Let (Σ, P) be an OS scheme. Then the following properties are equivalent.

(i) $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo on Σ^*.

(ii) $DUAL_P$ is unavoidable on Σ^*.

(iii) $MIXED_P$ is unavoidable on Σ^*.

The proof of Theorem 2.1 is somewhat involved, and is presented as a sequence of lemmas. The first few lemmas culminate with Lemma 2.3, which formalizes the following observation. If (Σ, P) is a strictly propagating OS scheme such that RHS_P is unavoidable with avoidance bound k_0, then any word in Σ^* can be parsed by repeatedly replacing the leftmost occurrence of a subword in RHS_P with a letter that derives it in such a way that all replacements occur within the first k_0+1 letters of the word and the final result is a word of at most k_0 letters. This "leftmost shift-reduce" parse of an arbitrary word yields a k_0-depth bounded "derivation" for any word in terms of the regular substitution S_P

described below.

Definition 2.3. Let (Σ, P) be a propagating OS scheme and for each letter $a \in \Sigma$, let Z_a be a variable. Let $Z = \{Z_a : a \in \Sigma\}$ and let P' be the set of left linear productions defined by $P' = \{Z_a \to Z_b w,\ Z_a \to bw : a \to bw \in P\}$. Then $S_P(a)$ denotes the regular substitution on Σ^* defined by $S_P(a) = L(G_a) \cup \{a\}$, where $G_a = (Z \cup \Sigma, \Sigma, P', Z_a)$.

$S_P(a)$ is the set of all strings obtained from a by repeatedly replacing leftmost symbols by right hand sides of corresponding rules in P. The subscript P will be omitted when the production system P is clear from the context. Note that for $bx \in S(a)$, $y \in S(b)$ the relation $yx \in S(a)$ holds.

For the next two lemmas let (Σ, P) be a fixed propagating OS scheme.

Lemma 2.1. Let $a \in \Sigma$, $u, w \in \Sigma^+$, $x, y \in \Sigma^*$. If $ax \in S(w)$, $u \in S^{k+1}(a)$, and $y \in S^k(x)$, where $k \geq 0$ is an arbitrary natural number , then $uy \in S^{k+1}(w)$.

Proof. Let $w = bw'$ where $b \in \Sigma$, $w' \in \Sigma^*$. Since $ax \in S(w)$ and P is propagating there are strings $x' \in \Sigma^*$ and $x'' \in \Sigma^*$, such that $x = x'x''$, $ax' \in S(b)$, and $x'' \in S(w')$. Since $u \in S^{k+1}(a)$, there is a word $u' \in S(a)$ such that $u \in S^k(u')$. But then $u'x' \in S(b)$, and consequently $u'x'x'' = u'x \in S(w)$. This implies $uy \in S^{k+1}(w)$. ▪

Lemma 2.2. Let $w \in \Sigma^+$, $y_1, y_2 \in \Sigma^*$ with $|y_1| < k$, and $a \in \Sigma$. If $y_1 a y_2 \in S^k(w)$ and $x \in S(a)$, then $y_1 x y_2 \in S^k(w)$.

Proof. We use induction on k. If $k = 1$, then $y_1 = \lambda$ and $ay_2 \in S(w)$. It follows from Lemma 2.1 that $xy_2 \in S(w)$. Assume that the statement holds for all numbers less than or equal some k. Consider $y = y_1 a y_2 \in S^{k+1}(w)$, where $|y_1| < k+1$, and let $x \in S(a)$. Let $w_1, w_2 \in \Sigma^*$, $b \in \Sigma$, $z_1, z_2, z_1', z_2' \in \Sigma^*$ be such that $w = w_1 b w_2$, $y_1 = z_1 z_1'$, $y_2 = z_2' z_2$, $z_1 \in S^{k+1}(w_1)$, $z_1' a z_2' \in S^{k+1}(b)$, $z_2 \in S^{k+1}(w_2)$ holds. (See Fig. 1). Let $u, u', u'', v, v', v'' \in \Sigma^*$, $c \in \Sigma$ be such that $ucv \in S(b)$, $z_1' = u'u''$, $z_2' = v''v'$, $u' \in S^k(u)$, $u''av'' \in S^k(c)$, $v' \in S^k(v)$ holds. If $z_1 u' \neq \lambda$, then $|u''| < k$, and consequently by the induction hypothesis $u''xv'' \in S^k(c)$, which implies that $y_1 x y_2 \in S^{k+1}(w)$. If $z_1 u' = \lambda$, then $w_1 = \lambda$, $u = \lambda$. Since $x \in S(a)$, we have $u''xv'' \in S^{k+1}(c)$, and therefore by Lemma 2.1, $u''xv''v' \in S^{k+1}(b)$. Consequently, also in this case $y_1 x y_2 \in S^{k+1}(w)$. ▪

Lemma 2.3. Let (Σ, P) be a strictly propagating OS scheme such that RHS_P is unavoidable with avoidance bound k_0. Let $F = \{w \in \Sigma^* : |w| \leq k_0\}$. Then $\Sigma^* = S^{k_0}(F)$.

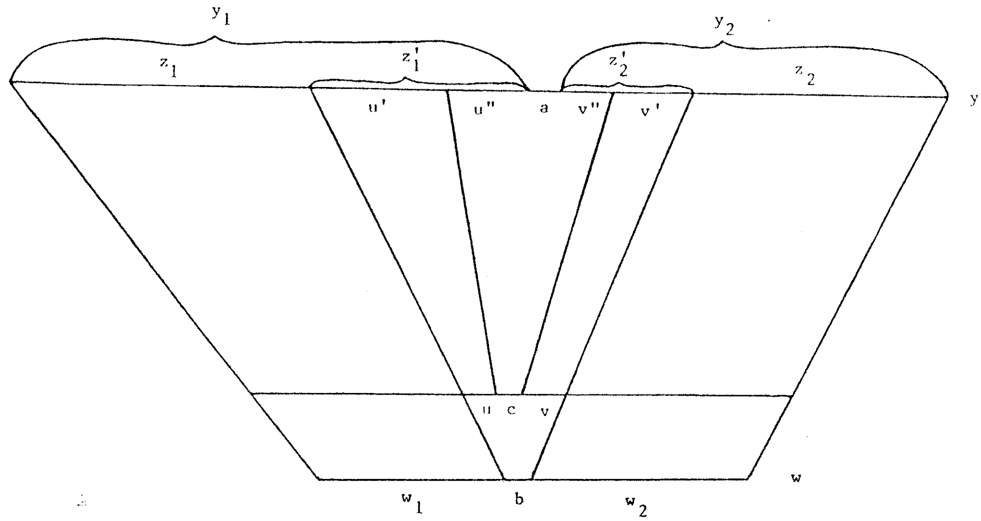

Figure 1.

Proof. Assume to the contrary that $\Sigma^* - S^{k_0}(F) \neq \phi$. Let w be a word in $\Sigma^* - S^{k_0}(F)$ of minimal length. Since $F \subseteq S^{k_0}(F)$, $|w| > k_0$. Since k_0 is the avoidance bound of RHS_P and (Σ,P) is strictly propagating, there are strings $w_1,w_2 \in \Sigma^*$ with $|w_1| < k_0$, and a rule $a \to x \in P$ such that $w = w_1xw_2$. Since w_1aw_2 is shorter than w, $w_1aw_2 \in S^{k_0}(z)$ for some $z \in F$. But then $w = w_1xw_2 \in S^{k_0}(z)$ by Lemma 2.2. Hence $w \in S^{k_0}(F)$, contrary to hypothesis. ∎

We make the following definitions in analogy with those in Definition 2.2.

Definition 2.4. A production $a \to x$ is *left bordered (right bordered)* if $x \in a\Sigma^+$ ($x \in \Sigma^+a$). An OS scheme (Σ,P) is *left (right)* bordered if each production in P is left (right) bordered. (Σ,P) is *dual bordered* if each production in P is both left and right bordered. (Σ,P) is *mixed bordered* if each production in P is either left or right bordered.

The essence of the argument that whenever $DUAL_P$ is unavoidable, $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo (i.e. (ii) implies (i) in Theorem 2.1) is contained in the following result.

Lemma 2.4. If (Σ,P) is a dual bordered OS scheme, then $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo on $S^k(F)$ for every $k \geq 0$ and finite set $F \subseteq \Sigma^*$.

Proof. We use induction on k. If $k = 0$, then $S^k(F) = F$ and the result is trivial. Assume that the result holds for all finite sets F and all numbers less than or equal some k. For $a \in \Sigma$, let $X_a = \{x : a \to axa \in P\}$. Note that since (Σ,P) is dual bordered, $S^{l+1}(a) = (aS^l(X_a))^*a$ holds for all numbers $l \geq 0$.

Fix some $a \in \Sigma$ and consider a sequence $\{w_i\}_{i \geq 1}$ of strings in $S^{k+1}(a)$. Each string w_i can be written in the form $w_i = ay_{i,1}a \cdots ay_{i,n(i)}a$ where $y_{i,l} \in S^k(X_a)$, $1 \leq l \leq n(i)$. Since $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on $S^k(X_a)$ by the induction hypothesis, by Proposition 1.1 there are numbers i and j, with $i < j$, such that for some subsequence $(j_1, \ldots, j_{n(i)})$ of $(1, \ldots, n(j))$, $y_{i,r} \overset{*}{\underset{P}{=\!=\!>}} y_{j,j_r}$ holds, $r = 1, \ldots, n(i)$. Since $a \overset{*}{\underset{P}{=\!=\!>}} ay_{t,l}a$ holds for all numbers t and l, it follows that $w_i \overset{*}{\underset{P}{=\!=\!>}} w_j$. Hence $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on $S^{k+1}(a)$, and consequently, by Proposition 1.2, $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on $S^{k+1}(F)$ for every finite set $F \subseteq \Sigma^*$. ▪

To complete our preparation for the proof of Theorem 2.1 we look now at the relationship between the unavoidability of $MIXED_P$ and that of $DUAL_P$. It is obvious that whenever $DUAL_P$ is unavoidable then $MIXED_P$ is unavoidable, since $DUAL_P \subseteq MIXED_P$. The other direction requires some work. We begin with a simple observation concerning mixed bordered schemes.

Lemma 2.5. Let (Σ,P) be a mixed bordered OS scheme. If $a \overset{*}{\underset{P}{=\!=\!>}} x$, then there are strings $x_1, x_2 \in \Sigma^*$ such that $x = x_1 a x_2$ and $a \overset{*}{\underset{P}{=\!=\!>}} x_1 a$, $a \overset{*}{\underset{P}{=\!=\!>}} a x_2$.

Proof. We use induction on the number of derivation steps in $a \overset{*}{\underset{P}{=\!=\!>}} x$. If $a \overset{0}{\underset{P}{=\!=\!>}} x$, then $x = a$ and the result is trivial. Assume that the claim holds for all derivations $a \overset{k}{\underset{P}{=\!=\!>}} y$, and let $a \overset{k+1}{\underset{P}{=\!=\!>}} x$ be a derivation of length $k+1$. Consequently, there is a word $y \in \Sigma^+$ such that $a \overset{k}{\underset{P}{=\!=\!>}} y$ and $y \underset{P}{=\!=\!>} x$. By the induction hypothesis, $y = y_1 a y_2$ for some words $y_1, y_2 \in \Sigma^*$ such that $a \overset{*}{\underset{P}{=\!=\!>}} y_1 a$ and $a \overset{*}{\underset{P}{=\!=\!>}} a y_2$. Now, if $x = y_1' a y_2$ with $y_1 \underset{P}{=\!=\!>} y_1'$ ($x = y_1 a y_2'$ with $y_2 \underset{P}{=\!=\!>} y_2'$, resp.), the result holds with $x_1 = y_1'$, $x_2 = y_2$ ($x_1 = y_1$, $x_2 = y_2'$, resp.). If $x = y_1 z y_2$, where $a \to z \in P$, then $z = az'$ or $z = z'a$ for some string $z \in \Sigma^*$ since (Σ,P) is mixed bordered. In the first case $x_1 = y_1$, $x_2 = z'y_2$, in the second case $x_1 = y_1 z'$, $x_2 = y_2$ satisfy the claim. ▪

Lemma 2.6. Let (Σ,P) be a mixed bordered OS scheme such that RHS_P is unavoidable with avoidance bound k_0. Then $LEFT_P$ is unavoidable with avoidance bound less than or equal $k_1 = k_0((k_0-1)|\Sigma|+1)^{k_0}-1$.

Proof. Since RHS_P has avoidance bound k_0, we can assume that $a \to x \in P$ implies $|x| \leq k_0$. Let F be chosen as in Lemma 2.3 and let x be a word of length at least $k_1+1 = k_0((k_0-1)|\Sigma|+1)^{k_0}$. By Lemma 2.3, $\Sigma^* = S^{k_0}(F)$, consequently there are words $x_0 \in F$, $x_1, \ldots, x_{k_0} \in \Sigma^+$ such that $x = x_{k_0}$ and $x_i \in S(x_{i-1})$, $i=1, \ldots, k_0$. Since $|x_0| \leq k_0$, there is an index j, $1 \leq j \leq k_0$, such that

$|x_j| \geq |x_{j-1}|((k_0-1)|\Sigma|+1)$ holds. This implies that there is a symbol a in x_{j-1} which contributes at least $(k_0-1)|\Sigma|+1$ symbols to x_j, to be precise: There are strings x_{j-1}', x_{j-1}'', x_j', $x_j'' \in \Sigma^*$, $z \in \Sigma^+$, $a \in \Sigma$ such that $x_{j-1} = x_{j-1}'ax_{j-1}''$, $x_j = x_j'zx_j''$, $x_j' \in S(x_{j-1}')$, $z \in S(a)$, $x_j'' \in S(x_{j-1}'')$ and $|z| \geq (k_0-1)|\Sigma|+1$.

By definition of the substitution S, there are symbols $a_0 = a, a_1, \ldots, a_m \in \Sigma$, and strings $y_1, \ldots, y_m \in \Sigma^+$ such that

$$a_0 \underset{P}{\Rightarrow} a_1y_1 \underset{P}{\Rightarrow} a_2y_2y_1 \underset{P}{\Rightarrow} \cdots \underset{P}{\Rightarrow} a_my_my_{m-1}\cdots y_1 = z$$

is a derivation of z (see Fig. 2), where in each derivation step the leftmost symbol, a_{l-1}, is replaced by a_ly_l according to a production $a_{l-1} \to a_ly_l$ $(1 \leq l \leq m)$.

By assumption, $|y_l| \leq k_0-1$ for $l = 1, \ldots, m$. Since $|z| \geq (k_0-1)|\Sigma|+1$, $m \geq |\Sigma|$. Consequently, there are numbers r and s, $0 \leq r < s \leq m$, such that $a_r = a_s$. Let $z_1 = a_my_m \cdots y_{s+1}$, $z_2 = y_s \cdots y_{r+1}$, $z_3 = y_r \cdots y_1$. Note that $z_2 \neq \lambda$. Since $x_j = x_j'z_1z_2z_3x_j''$ and $x = x_{k_0} \in S^{k_0-j}(x_j)$, there are strings z_1', z_2', z_3' such that for $1 \leq i \leq 3$, $z_i \overset{*}{\underset{P}{\Rightarrow}} z_i'$, and $z_1'z_2'z_3'$ is a substring of x. It follows that $a_r \overset{*}{\underset{P}{\Rightarrow}} a_rz_2 \overset{*}{\underset{P}{\Rightarrow}} a_rz_2'$ and $a_r \overset{*}{\underset{P}{\Rightarrow}} z_1 \overset{*}{\underset{P}{\Rightarrow}} z_1'$.

By Lemma 2.5, there are strings z_1'', z_1''' with $z_1' = z_1''a_rz_1'''$ and $a_r \overset{*}{\underset{P}{\Rightarrow}} a_rz_1'''$. But then $a_r \overset{*}{\underset{P}{\Rightarrow}} a_rz_1'''z_2'$, where $a_rz_1'''z_2'$ is a substring of x and is in $LEFT_P$. This shows that $LEFT_P$ is unavoidable with avoidance bound at most k_1. ∎

Lemma 2.7. Let (Σ,P) be a left bordered OS scheme such that RHS_P is unavoidable with avoidance bound k_0. Then $DUAL_P$ is unavoidable with avoidance bound at most $k_1 = k_0((k_0-1)|\Sigma|+1)^{k_0}-1$.

Proof. For a string $x \in \Sigma^*$ we denote by $x\sim$ the mirror image of x and for a language $L \subseteq \Sigma^*$ by $L\sim$ the mirror image of L, $L\sim = \{x\sim : x \in L\}$. Let $P\sim = \{a \to x\sim : a \to x \in P\}$. Clearly $(\Sigma,P\sim)$ is a right bordered system and $RHS_{P\sim} = (RHS_P)\sim$ is unavoidable with avoidance bound k_0. By Lemma 2.6 $LEFT_{P\sim}$ is unavoidable with avoidance bound at most k_1. The claim follows by observing that $(LEFT_{P\sim})\sim$ is unavoidable and $(LEFT_{P\sim})\sim = DUAL_P$. ∎

We are finally in position to prove the main theorem of this paper.

Proof of Theorem 2.1. $(i) \to (iii)$: Suppose $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo on Σ^*. We show that $LEFT_P$ is unavoidable which shows that $MIXED_P$ is unavoidable, since $MIXED_P$ contains $LEFT_P$. Assume to the contrary that $LEFT_P$ is avoidable. By using Koenig's lemma, there is an infinite string $w = a_1a_2a_3\ldots$ over Σ, $a_i \in \Sigma$ for $i \geq 1$, such that no finite substring of w has a subword in $LEFT_P$. Let $\{w_i\}_{i \geq 1}$ be

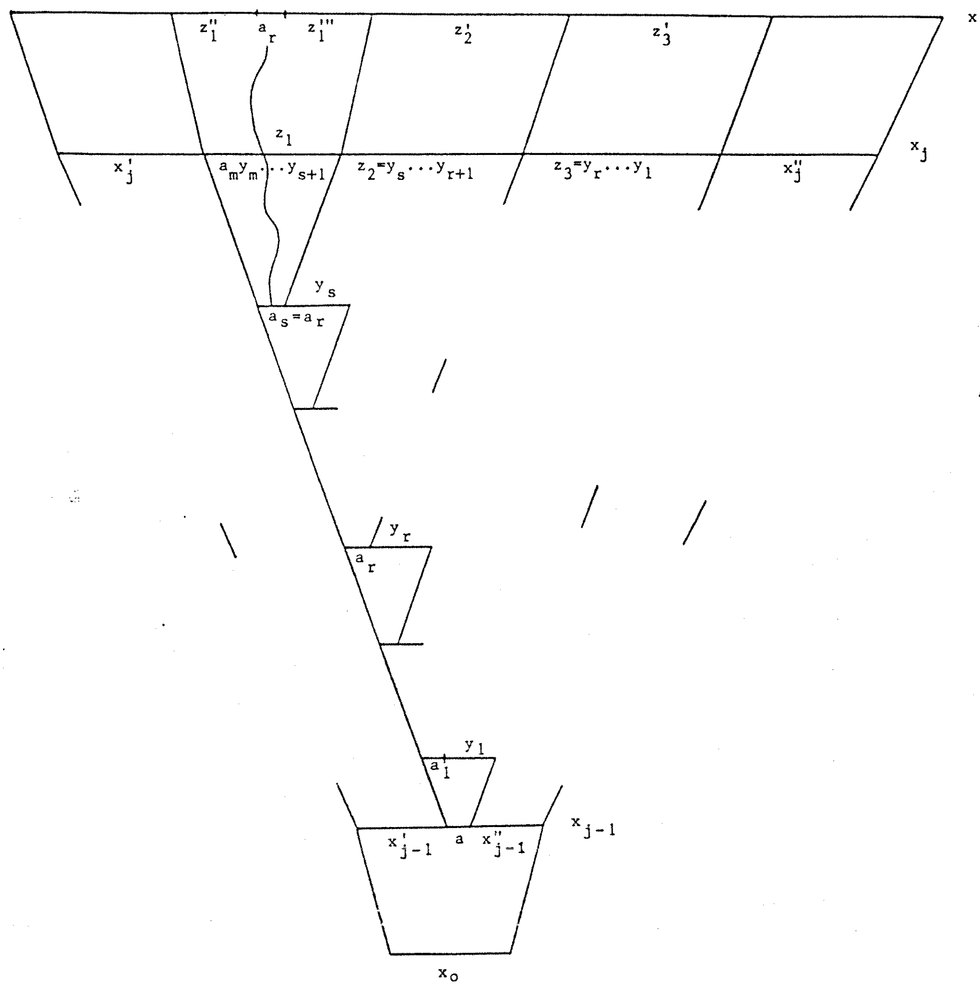

Figure 2.

the sequence of prefixes of w, i.e. $w_i = a_1 a_2 \cdots a_i$, $i = 1, 2, \ldots$. Since $\overset{*}{\underset{P}{=}}>$ is a wqo on Σ^*, there exist numbers i and j, $i < j$, such that $w_i \overset{*}{\underset{P}{=}}> w_j$. Call a letter a of w_i *active*, if a contributes at least two symbols to w_j. Let a_k be the left-most active letter of w_i. Consequently, for some number $n \geq 1$, $a_k \overset{*}{\underset{P}{=}}> a_k \cdots a_{k+n}$. Hence w_j contains a subword in $LEFT_P$, contrary to assumption.

$(iii) \rightarrow (ii)$: Suppose $MIXED_P$ is unavoidable with avoidance bound k_0. If (Σ, P') is the mixed bordered OS scheme defined by $P' = \{a \rightarrow x : a \in \Sigma, x \in MIXED_P(a), |x| \leq k_0\}$, then $RHS_{P'}$, is unavoidable with avoidance bound k_0; consequently, by Lemma 2.6, $LEFT_{P'}$, is unavoidable which shows that $LEFT_P$ is unavoidable, since $LEFT_{P'} \subseteq LEFT_P$. In a similar way we conclude, using Lemma 2.7, that $DUAL_P$ is unavoidable.

$(ii) \rightarrow (i)$: Suppose $DUAL_P$ is unavoidable with avoidance bound k_0. If (Σ, P') is the dual bordered OS scheme defined by $P' = \{a \rightarrow x : a \in \Sigma,\ |x| \leq k_0,\ x \in DUAL_P(a)\}$, then $RHS_{P'}$ is unavoidable with avoidance bound k_0. Consequently, by Lemma 2.4 and Lemma 2.3, $\overset{*}{\underset{P'}{=>}}$ is a wqo on Σ^*. Using Proposition 1.3 we conclude that $\overset{*}{\underset{P}{=>}}$ is a wqo on Σ^*. ∎

For mixed bordered OS schemes, Theorem 2.1 gives a very simple (and easily decidable) characterization of those schemes which generate wqo's.

Corollary 2.1. If (Σ, P) is a mixed bordered OS scheme, then $\overset{*}{\underset{P}{=>}}$ is a wqo on Σ^* if and only if RHS_P is unavoidable.

Proof. If RHS_P is unavoidable, then $MIXED_P$ is unavoidable, since $RHS_P \subseteq MIXED_P$. Hence $\overset{*}{\underset{P}{=>}}$ is a wqo on Σ^* by Theorem 2.1. On the other hand, if $\overset{*}{\underset{P}{=>}}$ is a wqo, then RHS_P must be unavoidable, since otherwise we could find an infinite sequence of strings not derivable from any other string, and hence for no pair x, y of strings in this sequence would $x \overset{*}{\underset{P}{=>}} y$ hold. ∎

For any mixed bordered OS scheme (Σ, P), if $x \overset{*}{\underset{P}{=>}} y$ then y is a supersequence of x. Hence all of the wqos generated by mixed bordered schemes under the conditions of Corollary 2.1 are refinements of the supersequence wqo discussed in the Introduction. One might conjecture that a characterization as in Corollary 2.1 could be given for a larger class of OS schemes which enjoy this property, e.g. for the class of embedding schemes, where an OS scheme (Σ, P) is called *embedding*, if for each production $a \rightarrow x \in P$, x can be written in the form $x = x_1 a x_2$, with $x_1, x_2 \in \Sigma^*$. However, such a generalization of Corollary 2.1 is impossible, as shown by the following example.

Example 2.1. Let $\Sigma = \{a, b, c\}$ and let P be given by the productions $a \rightarrow aa \mid aba \mid acba$, $b \rightarrow bb \mid bab$, $c \rightarrow cc \mid aca \mid bcb \mid bca$. It is readily verified that RHS_P is unavoidable on Σ^*. However, $\overset{*}{\underset{P}{=>}}$ is not a wqo since it can be shown that for no numbers m, n, $m > n$, the relation $(abc)^n \overset{*}{\underset{P}{=>}} (abc)^m$ holds.

On the other hand, Corollary 2.1 does generalize previous results on wqo refinements of the supersequence relation generated by repeated insertion of words from a fixed unavoidable set ([*EHR*]).

Definition 2.4. An OS scheme (Σ, P) is an *insertion system* if there exists a finite set $X \subseteq \Sigma^+$ such that $P = \{a \rightarrow ax \mid xa : a \in \Sigma,\ x \in X\}$. In this case (Σ, P) is the *insertion system generated by X*.

Insertion systems were originally introduced in [*EHR*] in a slightly different way, but it is easy to see how their definition relates to ours.

Corollary 2.2. ([*EHR*]) For a finite set $X \subseteq \Sigma^+$, if (Σ,P) is the insertion system generated by X, then $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on Σ^* if and only if X is unavoidable.

Proof. Clearly, RHS_P is unavoidable if and only if X is unavoidable. Consequently, Corollary 2.2 follows from Corollary 2.1. ▪

As a final example of the use of the wqos given by Corollary 1.1, consider the following proof that "history always repeats itself in ever more elaborate ways".

Definition 2.5. Let Σ be a finite alphabet of "events" and let $\lesssim$ be a total order which ranks the events in Σ. A sequence y of events is an *elaboration* of a sequence x, if $x = a_1 \cdots a_k$ for some $a_1, \ldots, a_k \in \Sigma$, and $y = y_1 \cdots y_k$ for some $y_1, \ldots, y_k \in \Sigma^+$, where for each i either $y_i = a_i$ or $y_i = a_i b_1 \cdots b_n a_i$ for some $n \geq 0$, where $b_j \in \Sigma$ and $a_i \lesssim b_j$, $1 \leq j \leq n$.

Thus we obtain an elaboration of x by replacing each event a of x by a series of events which begins and ends with a, such that no intermediate event has a smaller rank than a.

Corollary 2.3. If Σ is an alphabet and $\lesssim$ is a total order on Σ, then every infinite sequence $\{x_i\}_{i \geq 1}$ of strings in Σ^+ contains strings x_i, x_j, with $i < j$, such that x_j is an elaboration of x_i.

Proof. Let $\leq$ be the quasi-order on Σ^* defined by $x \leq y$ iff y is an elaboration of x, for strings $x, y \in \Sigma^+$. Clearly $\leq$ is multiplicative. For each $a \in \Sigma$ let $L_a = \{x \in \Sigma^+ : a \leq x, a \neq x\}$. Let $L = \cup_{a \in \Sigma} L_a$. It is easily verified by induction on $|\Sigma|$ that L is unavoidable. Let L' be a finite unavoidable subset of L and let $P = \{a \to x : x \in L', a \leq x\}$. The OS scheme (Σ,P) is dual bordered and since $RHS_P = L'$ is unavoidable, $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on Σ^* by Corollary 2.1. The result follows now from Proposition 1.4. ▪

3. Monoid-representations

While in Section 2 total regulators generated by propagating OS schemes were characterized by an unavoidability criterion, in this section an attempt is made to describe such total regulators in a more algebraic way, corresponding to the well known characterization of regular languages in terms of congruences of finite index (finite monoids, resp.). The first result in this section can be seen as the natural extension of this characterization to regulators defined by OS schemes (see Def. 1.2).

Theorem 3.1. For an OS scheme (Σ,P), $\overset{*}{\underset{P}{=}}>$ is a regulator on Σ^* if and only if there is a finite monoid M, a morphism $h : \Sigma^* \to M$, and a multiplicative quasi-order $\leq$ on M such that for all $a \in \Sigma$ and $x \in \Sigma^*$, $a \overset{*}{\underset{P}{=}}> x$ iff $h(a) \leq h(x)$.

Proof. (if part) Let M, h and $\leq$ be as in the statement of the theorem, and for $a \in \Sigma$, let $M_a = \{m \in M : h(a) \leq m\}$. Consequently, $cl_P(a) = h^{-1}(M_a)$, which shows that $cl_P(a)$ is regular for all $a \in \Sigma$. If we define a regular substitution σ on Σ^* by $\sigma(a) = cl_P(a)$, for $a \in \Sigma$, then for every subset L of Σ^* $cl_P(L) = \sigma(L)$. Therefore $cl_P(L)$ is regular for every regular subset L of Σ^*.

(only if part) Assume that $\overset{*}{\underset{P}{=}}>$ is a regulator on Σ^*. For $a \in \Sigma^*$, let $M(a)$ be the syntactic monoid of (the regular language) $cl_P(a)$, let $\pi_a : \Sigma^* \to M(a)$ be the canonical morphism mapping each string of Σ^* to its class modulo the syntactic congruence of $cl_P(a)$, and let $\leq_a$ be the syntactic partial order on $M(a)$, i.e. $\pi_a(x) \leq_a \pi_a(y)$ if and only if for all $u,v \in \Sigma^*$, $uxv \in cl_P(a)$ implies $uyv \in cl_P(a)$. Let $M' = \underset{a \in \Sigma}{X} M(a)$ be the Cartesian product of the monoids $M(a)$, endowed with componentwise multiplication, and let $h : \Sigma^* \to M'$ be the morphism defined by $h(x)=(\pi_a(x))_{a \in \Sigma}$, $x \in \Sigma^*$. Let $M = h(\Sigma^*)$ and define a multiplicative partial order $\leq$ on M by $h(x) \leq h(y)$ if and only if for all $a \in \Sigma$, $\pi_a(x) \leq_a \pi_a(y)$. We will show that M,h and $\leq$ satisfy the claim of the theorem. Indeed, if $h(a) \leq h(x)$, then $\pi_a(a) \leq_a \pi_a(x)$. Since $a \in cl_P(a)$, this implies $x \in cl_P(a)$, i.e. $a \overset{*}{\underset{P}{=}}> x$. If, on the other hand, $a \overset{*}{\underset{P}{=}}> x$, then for all $b \in \Sigma$, $u, v \in \Sigma^*$, $uav \in cl_P(b)$ implies $uxv \in cl_P(b)$. Consequently, $\pi_b(a) \leq_b \pi_b(x)$ for all $b \in \Sigma$ which implies $h(a) \leq h(x)$. This proves the only-if part. ▪

The above theorem suggests the following definition.

Definition 3.1. Let (Σ,P) be an OS scheme, let M be a monoid, $h : \Sigma^* \to M$ a morphism, and $\leq$ a multiplicative quasi-order on M. The triple $(M, h, \leq)$ is called a *(monoid-)representation* of (Σ,P) if for all $a \in \Sigma$ and $x \in \Sigma^*$, $a \overset{*}{\underset{P}{=}}> x$ holds iff $h(a) \leq h(x)$. $(M, h, \leq)$ is a *finite (monoid-)representation* of (Σ,P) if M is finite.

Theorem 3.1 can now be restated as follows: For an OS scheme (Σ,P), $\overset{*}{\underset{P}{=}}>$ is a regulator if and only if (Σ,P) has a finite monoid-representation.

It seems natural to try to characterize wqo's (total regulators) defined by OS schemes in terms of monoid-representations. So far we only have partial answers to this problem, and we restrict ourselves to presenting the following sufficient condition on M to guarantee that $\overset{*}{\underset{P}{=\!=\!>}}$ is a total regulator.

Theorem 3.2. Let (Σ,P) be an OS scheme and let $(G, h, \leq)$ be a monoid-representation of (Σ,P) where G is a finite group. Then $\overset{*}{\underset{P}{=\!=\!>}}$ is a total regulator on Σ^*.

Proof. Let $|G| = n$, and let $x = a_0 a_1 \cdots a_n \in \Sigma^*$, where $a_0, \ldots, a_n \in \Sigma$. Consequently, there are numbers i, j with $0 \leq i < j \leq n$, such that $h(a_0 a_1 \cdots a_i) = h(a_0 a_1 \cdots a_j)$. Since G is a group, $h(a_{i+1} \cdots a_j) = 1$, where 1 is the identity element of G. But then $h(a_i) = h(a_i a_{i+1} \cdots a_j)$, and therefore $a_i \overset{*}{\underset{P}{=\!=\!>}} a_i a_{i+1} \cdots a_j$. This shows that $LEFT_P$ is unavoidable, and thus $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on Σ^* by Theorem 2.1. Hence, $\overset{*}{\underset{P}{=\!=\!>}}$ is a total regulator on Σ^* by Proposition 1.5. ■

The proof of Theorem 3.2 shows that in order to guarantee that $\overset{*}{\underset{P}{=\!=\!>}}$ is a wqo on Σ^*, the following weaker condition on $(G, h, \leq)$ for finite group G is sufficient: For all $a \in \Sigma$, $x \in \Sigma^*$, if $h(a) = h(x)$, then $a \overset{*}{\underset{P}{=\!=\!>}} x$.

There are wqo's (and hence total regulators) defined by OS schemes which cannot be represented in a finite group in the sense of Definition 3.1. For example, the OS scheme (Σ,P) with $\Sigma = \{a,b\}$, $P = \{a \to aa \mid aba,\ b \to bb\}$ defines a wqo $\overset{*}{\underset{P}{=\!=\!>}}$ by Corollary 2.1. On the other hand, if $(G, h, \leq)$ is a representation of (Σ,P) with G a finite group containing n elements, then $h(a) = h(ab^n)$, and consequently $a \overset{*}{\underset{P}{=\!=\!>}} ab^n$, which is a contradiction. This shows that (Σ,P) cannot be represented in a finite group.

In the rest of this section we briefly discuss the question which triples $(M, h, \leq)$ - where M is a finite monoid, $h : \Sigma^* \to M$ is a morphism, and $\leq$ is a multiplicative quasi-order on M - are monoid-representations of some OS scheme. To this end, let for such a triple and for $a \in \Sigma$

$$L_a' = \{x \in \Sigma^* : h(a) \leq h(x),\ a \neq x\},$$

and let $\bar{\Sigma} = \{\bar{a} : a \in \Sigma\}$ be a barred copy of Σ, $\Sigma \cap \bar{\Sigma} = \phi$. Define a substitution σ on Σ^* by $\sigma(a) = \{a, \bar{a}\}$, $a \in \Sigma$, and a substitution ρ on $(\Sigma \cup \bar{\Sigma})^*$ by $\rho(a) = \{a\}$, $\rho(\bar{a}) = L_a'$. For $a \in \Sigma$, let

$$L_a = L_a' - \rho(\sigma(L_a') \cap \Sigma^* \bar{\Sigma} \Sigma^*).$$

L_a is the set of words in L_a' which cannot be obtained from other words in L_a' by substituting words from some sets L_b', $b \in \Sigma$. By construction, L_a is effectively regular for each $a \in \Sigma$.

Lemma 3.1. Let M be a finite monoid, let $h : \Sigma^* \to M$ be a morphism and let $\leq$ be a multiplicative quasi-order on M.

(i) Let (Σ,P) be an OS scheme not containing any rule of the form $a \to a$. If $(M, h, \leq)$ is a representation of (Σ,P) then $\cup_{a \in \Sigma} L_a$ is finite and (*) $\cup_{a \in \Sigma}\{a \to x : x \in L_a\} \subseteq P \subseteq \cup_{a \in \Sigma}\{a \to x : x \in L_a'\}$.

(ii) Conversely, if $\cup_{a \in \Sigma} L_a$ is finite, then for any finite set P of productions satisfying (*), the triple $(M, h, \leq)$ is a representation of (Σ,P).

Proof. (i) For $a \to x \in P$ the relation $a \overset{*}{\underset{P}{=}}> x$ and consequently $x \in L_a'$ holds. On the other hand, assume that $x \in L_a$, but $a \to x \notin P$. Since $(M, h, \leq)$ is a representation of (Σ,P), $a \overset{*}{\underset{P}{=}}> x$. It follows that there are $b \in \Sigma$ and $x_1, x_2, x_3 \in \Sigma^*$ such that $x_1 b x_3 \neq a$, $x_2 \neq b$, $x = x_1 x_2 x_3$, and $a \overset{*}{\underset{P}{=}}> x_1 b x_3$, $b \overset{*}{\underset{P}{=}}> x_2$. This implies $x_1 b x_2 \in L_a'$, $x_2 \in L_b'$, contrary to the assumption $x \in L_a$. Part (ii) is straightforward by definition of the sets L_a, L_a' and the fact that $\leq$ is multiplicative. ∎

As a consequence of Lemma 3.1, it is decidable whether a triple $(M, h, \leq)$ is monoid-representation of some OS scheme (Σ,P) : it suffices to test whether the regular sets L_a are finite. Moreover, there is essentially a unique OS scheme represented by $(M, h, \leq)$, namely $(\Sigma, \cup_{a \in \Sigma}\{a \to x : x \in L_a\})$. This decision problem is not trivial, since there are triples $(M, h, \leq)$ for finite M and multiplicative $\leq$ which are not representations of any OS scheme.

Example 3.1. Let $\Sigma = \{a,b\}$. Let M be the syntactic monoid of $L = (ab^2b^*)^*a$, let $h : \Sigma^* \to M$ be the canonical morphism mapping $x \in \Sigma^*$ to its class modulo L and let the quasi-order $\leq$ on M be the equality relation. A simple computation shows that $L_b' = \phi$, $L_a' = L - \{a\}$, and $L_a = ab^2b^*a$. Consequently, $(M, h, \leq)$ is not monoid-representation of any OS scheme.

However, if M is a finite group, then $(M, h, \leq)$ is a representation of some OS scheme (Σ,P). It should be noted that because of Theorem 3.2, $\overset{*}{\underset{P}{=}}>$ is then a total regulator.

Theorem 3.3. If G is a finite group, $h : \Sigma^* \to G$ a morphism and $\leq$ a multiplicative quasi-order on G, then there is an OS scheme (Σ,P) with representation $(G, h, \leq)$.

Proof. Because of Lemma 3.1 it suffices to show that the sets L_a are finite. More precisely, we prove that $x \in L_a$ implies $|x| \leq |G|+1$. Let $|G| = n$ and assume to the contrary that there is a string $x = a_0 a_1 \cdots a_{n+1}x' \in L_a$, where $a_0, a_1, \ldots, a_{n+1} \in \Sigma$, $x' \in \Sigma^*$. Consequently, there are numbers i and j, $0 \leq i < j \leq n$, such that $h(a_0 \cdots a_i) = h(a_0 \cdots a_j)$. Since G is a group, this implies $h(a_i) = h(a_i a_{i+1} \cdots a_j)$ and $h(a_0 \cdots a_i a_{j+1} \cdots a_{n+1}x') = h(x)$, where $|a_i a_{i+1} \cdots a_j| \geq 2$ and $|a_0 \cdots a_i a_{j+1} \cdots a_{n+1}x'| \geq 2$. We conclude that $a_i a_{i+1} \cdots a_j \in L_{a_i}'$ and $a_0 \cdots a_i a_{j+1} \cdots a_{n+1}x' \in L_a'$. This is a contradiction to $x \in L_a$. ∎

The construction of the proof of Theorem 3.3 gives a tool to construct OS total regulators, however as pointed out above, not every OS total regulator can be obtained in this way. It remains an open problem to characterize those OS total regulators which have a representation in a group.

Example 3.2. Let C_3 be the (additively written) cyclic group with elements 0, 1, 2, let $\Sigma = \{a, b\}$, let $h : \Sigma^* \to C_3$ be defined by $h(a) = 1$, $h(b) = 2$, and $\leq$ by $i \leq j$ iff $i = j$. A straight forward computation shows that $L_a = \{bb, aba\}$, $L_b = \{aa, bab\}$, i.e. $(C_3, h, \leq)$ is a representation of the OS scheme (Σ, P) with $P = \{a \to bb \mid aba,\ b \to aa \mid bab\}$. This result could also be established directly by observing that $a \overset{*}{\underset{P}{=>}} x$ (resp. $b \overset{*}{\underset{P}{=>}} x$) holds if and only if $\#_a(x) - \#_b(x) \equiv 1 \bmod 3$ (resp. $\#_a(x) - \#_b(x) \equiv 2 \bmod 3$).

Section 4. Open Problems

The primary open problem remaining is to show that it is decidable whether or not a propagating OS scheme generates a total regulator. While Theorem 2.1 gives a characterization of such systems, we have been unable to show that this characterization is effective. One approach to this problem would be to investigate the pumping properties of OS total regulators, hoping to find one which is both necessary and sufficient, and effective.

Let (Σ, P) be a propagating OS scheme, and consider the following "pumping" properties:

a) For all $w \in \Sigma^+$ there exist k, l with $k < l$ such that $w^k \overset{*}{\underset{P}{=>}} w^l$

b) For all $w \in \Sigma^+$ there exists $k > 1$ such that $w \overset{*}{\underset{P}{=>}} w^k$

c) For all $w \in \Sigma^+$ there exist $a \in \Sigma$, $w_1, w_2 \in \Sigma^*$ and $k \geq 1$ such that $w = w_1 a w_2$ and $a \overset{*}{\underset{P}{=>}} (a w_2 w_1)^k a$.

While it appears that each of these pumping properties is stronger than the previous one, it can be shown that in fact they are all equivalent for propagating OS schemes. Thus since (a) is obviously implied whenever $\overset{*}{\underset{P}{=>}}$ is a wqo on Σ^*, they are all necessary pumping properties of propagating OS total regulators.

Are they sufficient? We have no counterexample.

While these pumping properties are not effective as given, if it can be shown that, for example, (b) implies that $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo on Σ^* then this, combined with Theorem 2.1, would provide an effective characterization of propagating OS total regulators. The effectiveness follows by considering two semi-algorithms: one which tests if $w^2w^* \cap cl_P(w) = \phi$ for larger and larger w, and the other which checks if F is unavoidable in Σ^* for larger and larger finite subsets F of $DUAL_P$ (or $MIXED_P$).

One appealing aspect of this approach is that property (c) already comes close to implying that $DUAL_P$ is unavoidable in Σ^*. In fact, (c) implies for any word $w \in \Sigma^+$, that w^* contains a word with a subword in $DUAL_P$. Hence we might say that if property (c) holds, then $DUAL_P$ is "periodically unavoidable". Choffrut and Culik II ([*CC*]) have shown that for any regular language $R \subseteq \Sigma^+$, R is unavoidable if and only if it is periodically unavoidable in the above sense. We know that this property does not hold for all languages; $L = \{ww : w \in \Sigma^+\}$, where Σ has at least three letters, is an example of a language which is periodically unavoidable but not unavoidable. However, if it holds for all context-free languages, then (c) would imply that $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo on Σ^*, since $DUAL_P$ is context-free. Hence we would like to know the status of the following:

Conjecture A. For any context-free language $L \subseteq \Sigma^+$, L is unavoidable in Σ^* if and only if it is periodically unavoidable, i.e. if and only if for all $w \in \Sigma^+$, w^* contains a word with a subword in L.

It should be noted that Conjecture A would follow from the stronger conjecture that whenever the syntactic congruence of a context-free language is periodic, then the language is regular (see [*AUT*]); however a counterexample to this conjecture has recently been given by M. Main ([*MAI*]).

Another open problem is to generalize the characterization theorem (Theorem 2.1) to arbitrary length-increasing production systems (i.e. word replacement systems). In addition, it would be nice to know what role such systems play within the class of all length increasing wqo's. By Proposition 1.4, whenever a length-increasing multiplicative quasi-order contains a wqo generated by a finite production system, then it is a wqo. At present we have no counterexample to the following "converse" of this statement:

Conjecture B. For any length-increasing multiplicative wqo $\leq$ on Σ^* there exists a finite production system (Σ, P), with $P = \{u_1 \to v_1, \ldots, u_k \to v_k\}$, $u_i, v_i \in \Sigma^+$ and $u_i \leq v_i$ for all i, $1 \leq i \leq k$, such that $\overset{*}{\underset{P}{\Rightarrow}}$ is a wqo on Σ^*.

References

[AUT] J.M. Autebert, J. Beauquier, L. Boasson, M. Nivat, Quelques problems ouverts en theorie des languages algebriques, *RAIRO Inf. Theor.* **13** (4) (1979) 363-378.

[CC] C. Choffrut, K. Culik II, On extendibility of unavoidable sets, *Dis. Appl. Math.*, **9** (1984) 125-137.

[CON] J. H. Conway, *Regular Algebra and Finite Machines* (Chapman & Hall, London, 1971) pp. 63-64.

[DER] N. Dershowitz, Orderings for term-rewriting systems, *Theor. Comp. Sci.* **17** (1982) 279-301.

[EHR] A. Ehrenfeucht, D. Haussler, G. Rozenberg, On regularity of context-free languages, *Theor. Comp. Sci.* **27** (1983) 311-332.

[HAI] L. H. Haines, On free monoids partially ordered by embedding, *J. Combin. Th.* **6** (1969) 94-98.

[HAR] M. A. Harrison, *Introduction to Formal Language Theory* (Addison-Wesley, Reading, MA. 1978).

[HIG] G. Higman, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* **3** (2) (1952) 326-336.

[KRU] J.B. Kruskal, The theory of well-quasi ordering: A frequently discovered concept, *J. Combin. Theory (Ser. A)* **13** (1972) 297-305.

[LAT] M. Latteux, G. Rozenberg, Commutative one-counter languages are regular, *J. Comp. Sys. Sci.* **29** (1984) 54-57.

[LEE] J. van Leeuwen, Effective constructions in well-partially-ordered monoids, *Discrete Math.*, **21** (1978) 237-252.

[MAI] M. Main, An infinite squarefree co-CFL, *Inf. Proc. Let.*, to appear.

[RUO] K. Ruohonen, A note on off-line machines with 'Brownian' input heads, *Disc. App. Math.* **9** (1984) 69-75.