JOURNAL OF COMPUTER AND SYSTEM SCIENCES: 5, 41–53 (1971)

Classification of Noncounting Events*

J. A. BRZOZOWSKI, K. ČULÍK II, AND A. GABRIELIAN

Department of Applied Analysis and Computer Science,
University of Waterloo, Waterloo, Ontario, Canada

Received August 12, 1970

An event E is a subset of the free monoid A^* generated by the finite alphabet A. E is noncounting if and only if there exists an integer $k \geqslant 0$, called the order of E, such that for any $x, y, z \in A^*$, $xy^kz \in E$ if and only if $xy^{k+1}z \in E$. From semigroup theory it follows that the number of noncounting events of order $\leqslant 1$ is finite. Each such event is regular and the finite automata accepting such events over a fixed alphabet are homomorphic images of a universal automaton. Star-free regular expressions for such events are easily obtainable. It is next shown that the number of distinct noncounting events of order $\geqslant 2$ over any alphabet with two or more letters is infinite. Furthermore, there exist noncounting events which are of any "arbitrary degree of complexity," e.g. not recursively enumerable.

1. INTRODUCTION

An *event* E, i.e., a subset of the free monoid A^* generated by the finite *alphabet* A, is *noncounting* if and only if there exists an integer $k \geqslant 0$, called the *order* of the noncounting event, such that for arbitrary $x, y, z \in A^*$, $xy^kz \in E$ if and only if $xy^{k+1}z \in E$. Schutzenberger [8] and Papert and McNaughton [6] showed that an event that is regular (acceptable by a finite automaton) is noncounting if and only if there exists a star-free regular expression representing it. Cohen and Brzozowski [1] and Meyer [5] also proved this equivalence by invoking the Krohn–Rhodes decomposition theorem for finite automata. Using some results from semigroup theory, we show in Sections 3 and 4 of this paper that there exist only a finite number of noncounting events of order $\leqslant 1$ and that they are all regular. Furthermore, we show that for each finite alphabet A there exists a universal finite automaton whose homomorphic images constitute precisely the set of all automata that accept noncounting events of order $\leqslant 1$ over the alphabet A. We also prove directly that each such event can be represented by a star-free regular expression. In Section 5 we consider noncounting events of order $\geqslant 2$ and we show that for each order $\geqslant 2$ there exist an infinite number of

* This work was supported by the National Research Council of Canada, under Grants A-1617 and A-7403.

distinct regular noncounting events over an alphabet with two or more letters. In fact, there exist noncounting events that are, in a sense, of any arbitrary degree of complexity, i.e., they are not regular, context-free or even recursively enumerable.

2. Preliminaries

Definition 2.1. For u, v in A^* write $u \leftrightarrow_k v$ (or $u \leftrightarrow v$, if k is understood) if and only if one of the following three conditions holds:

(i) $u = v$;

(ii) $u = xy^kz$ and $v = xy^{k+1}z$, for some x, $z \in A^*$, $y \in A^+$, where for any $E \subseteq A^*$, $E^+ = E^* - \{\lambda\}$, and λ is the empty word;

(iii) $u = xy^{k+1}z$ and $v = xy^kz$, for some x, $z \in A^*$, $y \in A^+$.

The transitive closure of the relation $\leftrightarrow_k$ is denoted by $\sim_k$ (or $\sim$).

For each k, $\sim_k$ is obviously a congruence relation on A^* with respect to concatenation. The congruence class containing $u \in A^*$ will be denoted by $[u]_k$ (or $[u]$).

Definition 2.2. Let E be an arbitrary subset of A^*. E is called *noncounting* of order k ($k \geqslant 0$) if and only if the following holds: For u, v, $w \in A^*$, $uw^kv \in E$ if and only if $uw^{k+1}v \in E$. In other words, E is noncounting if and only if it is a union of congruence classes of $\sim_k$ for some k.

Definition 2.3. For u, $v \in A^*$, u is said to be a *prefix* (*suffix*) of v if there exists a $w \in A^*$ such that $uw = v$ ($wu = v$). $u \not\leftrightarrow v$ means that $u \leftrightarrow v$ is false, similarly for $\not\sim$. ϕ denotes the empty set, and $I = A^* = \bar{\phi}$, where the bar stands for complementation. F_k is the family of all noncounting events that are of order k but not of any lower order, and B_k is the family of all noncounting events of order $\leqslant k$.

Lemma 2.1. *If the congruence $\sim_k$ on A^* is of finite index, then all noncounting events of order k on the alphabet A are regular.*

Proof. This follows from the theorem of Myhill that an event is regular if and only if it is a union of some equivalence classes of a congruence relation over A^* of finite index. See Rabin and Scott [7].

Lemma 2.2. $F_0 = B_0 = \{\phi, I\}$.

Proof. It is clear that ϕ is noncounting of every order. Let E be noncounting of order zero. Suppose $x \in E$. If $x = \lambda$, then $\lambda = \lambda y^0 \lambda \in E$ implying that $\lambda y^1 \lambda = y \in E$ by Definition 2.2. Therefore, $y \in E$ for all $y \in A^*$, i.e., $E = I$. If $x \neq \lambda$, $x = \lambda x^1 \lambda \in E$ implies $\lambda x^0 \lambda = \lambda \in E$. So this reduces to the first case.

LEMMA 2.3. *For each k, B_k is a Boolean algebra.*

Proof. It is easy to verify (see Meyer [5]) that if $E = P \cup Q$, $P \in F_{k_P}$, $Q \in F_{k_Q}$, then $E \in B_{\max[k_P, k_Q]}$. Also $E \in F_k$ if and only if $\bar{E} \in F_k$.

3. NONCOUNTING EVENTS OF ORDER $\leqslant 1$ OVER A TWO-LETTER ALPHABET

Let $A = \{0, 1\}$ and let $S_2 = A^*/\sim_1$ be the monoid of congruence classes, called the free idempotent monoid generated by A, because S_2 satisfies the relation $x = x^2$ for each $x \in S_2$, but is otherwise free. One can easily verify that S_2 contains exactly seven elements, namely, $[\lambda]$, $[0]$, $[1]$, $[01]$, $[10]$, $[010]$ and $[101]$. Any other sequence of 0's and 1's will contain a repeated subsequence and so is reducible to one of the above. Also it is easily shown that all seven elements are distinct. General methods of characterizing the equivalence classes are described in Section 4.

Let $\mathscr{A}_2 = \langle Q, M, [\lambda], -\rangle$ be the finite automaton over the input alphabet A, where $Q = \{[x] \mid [x] \in S_2\}$ is the set of states, for $a \in A$, $[x] \in Q$ the transition function is $M([x], a) = [xa]$, $[\lambda]$ is the initial state, and the set of final states is not specified. The state graph of $\mathscr{A}_2$ is shown in Fig. 1. Let $\mathscr{A}_{2[x]} = \langle Q, M, [\lambda], \{[x]\}\rangle$ and let R_x be the set of words accepted by $\mathscr{A}_{2[x]}$, which is clearly the set of all words that take $\mathscr{A}_{2[x]}$ from $[\lambda]$ to $[x]$.

One can verify that

$$R_\lambda = \lambda$$

$$R_0 = 00^* = \overline{I1I} - \lambda = \overline{\bar{\phi}1\bar{\phi}} - \lambda \qquad R_1 = 11^* = \overline{\bar{\phi}0\bar{\phi}} - \lambda$$
$$R_{01} = 0I1 = 0\bar{\phi}1 \qquad R_{10} = 1I0 = 1\bar{\phi}0$$
$$R_{010} = 00^*1I0 = 0I1I0 = 0\bar{\phi}1\bar{\phi}0 \qquad R_{101} = 1I0I1 = 1\bar{\phi}0\bar{\phi}1$$

Note that $\mathscr{A}_{2[x]}$ for $x \in S_2$ is not a reduced automaton. For example, in $\mathscr{A}_{2[01]}$, the states corresponding to $[1]$, $[10]$ and $[101]$ are all equivalent as are $[0]$ and $[010]$. The reduced version of $\mathscr{A}_{2[01]}$ is shown in Fig. 2.

By the construction of $\mathscr{A}_2$ it is clear that for any $y \in A^*$, $M([\lambda], y) = [x]$ iff $y \sim_1 x$. The regular expression R_x above denotes precisely the set of all words equivalent to x, i.e. $R_x = [x]$, where $[x]$ is interpreted as a set of words. Thus each equivalence class is a star-free event. As arbitrary noncounting events of order $\leqslant 1$ are unions of these equivalence classes they must be star-free also. Recall that B_1, the set of all noncounting events of order $\leqslant 1$, is a Boolean algebra. The events $[x]$ denoted by R_x are clearly the atoms of this algebra and B_1 has 2^7 elements, corresponding to the subsets of the set of seven atoms.

The automaton $\mathscr{A}_2$ can be considered as a *universal* automaton for B_1 in the sense that, if $E \in B_1$ and $E = \bigcup_{x \in X} [x]$, $X \subseteq S_2$, then $\mathscr{A}_{2X} = \langle Q, M, [\lambda], \{[x] \mid x \in X\}\rangle$ accepts precisely E. The reduced version of $\mathscr{A}_{2X}$ must then be a homomorphic image of $\mathscr{A}_2$.

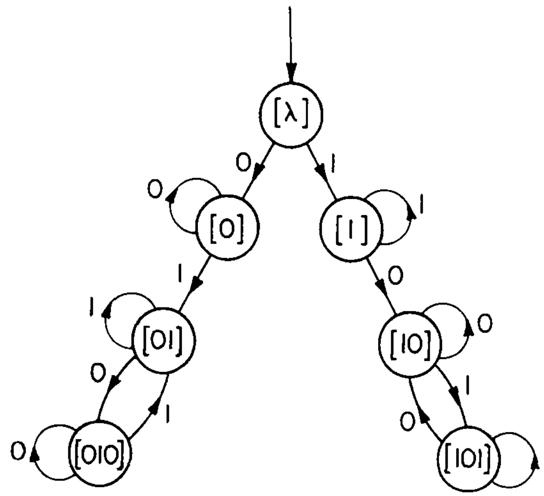

FIG. 1. Automaton $\mathscr{A}_2$.

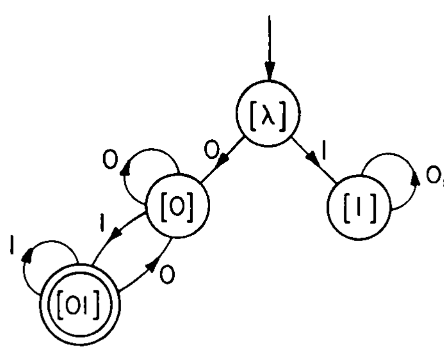

FIG. 2. Reduced version of $\mathscr{A}_{2[01]}$.

4. Noncounting Events of Order $\leqslant 1$ over Arbitrary Finite Alphabets

Using the results of Green and Rees [3] the characterization of Section 3 can be generalized to the case of an arbitrary finite alphabet. The first part of this section through Corollary 4.5 is based entirely on their work.

Let $S_n = A^*/\sim_1$ be the free idempotent monoid generated by A (i.e. $x = x^2$, for all $x \in S_n$), where A has n letters.

LEMMA 4.1. *Two words in S_n are equivalent only if they contain the same letters.*

Proof. Replacement of xy^2z by xyz, or vice versa, in S_n cannot eliminate any existing a_i, $i = 1,..., n$, nor can it introduce new ones.

DEFINITION 4.2. C_k is the set of all elements of S_n that contain each of the letters $a_1,..., a_k$. The cardinality of C_k will be denoted by c_k and that of S_n by s_n.

LEMMA 4.2. $s_n = \sum_{k=0}^{n} \binom{n}{k} c_k$, *where* $c_0 = 1$.

Proof. Obviously, the cardinality of C_k is independent of any permutation on the set $A = \{a_1,..., a_n\}$. For each $a_{i_1},..., a_{i_k}$, with $i_s \neq i_t$ for $s \neq t$, and $i_s \in \{1,..., n\}$, for $s = 1,..., k$, there exist c_k words in S_n containing $a_{i_1},..., a_{i_k}$.

DEFINITION 4.2. The *initial mark* of a word w in S_n is that letter "a" whose earliest appearance (from the left) in w is farthest to the right. Thus $w = w_1aw_2$ and w_1 contains all the letters in w except the letter "a." The word w_1 appearing to the left of this "a" is called the *initial* of w. Similarly, the *terminal mark* of w is the letter "a" whose last appearance in w is farthest to the left. The *terminal* of w is the word to the right of this "a."

LEMMA 4.3. *Two words w and w' in S_n are equivalent ($w \sim_1 w'$) if and only if they have the same initial marks, the same terminal marks, equivalent initials and equivalent terminals.*

Proof. By Lemma 4.1 it is sufficient to consider $w, w' \in C_n$. By Lemma 2 of [3], if $w \sim_1 w'$, then w and w' have the same initial marks, the same terminal marks, equivalent initials, and equivalent terminals. By Lemma 5 of [3] the set G of all elements of C_n which have initial u, initial mark a, terminal v, and terminal mark b is a group. $G = \{w \mid w = uax = ybv, \text{ for } x, y \in S_n\}$ is idempotent since it is a subgroup of S_n. Therefore it has only one element (i.e., $w \in G$ implies $w \sim uabv$). Now if $w \sim uabv$ and $w' \sim u'abv'$, with $u \sim u'$ and $v \sim v'$, then $w \sim w'$.

THEOREM 4.4. *c_n and s_n are finite for all finite n. Furthermore,* $c_n = n^2c_{n-1}^2 = n^2(n-1)^4(n-2)^8 \cdots$.

COROLLARY 4.5. $s_1 = 2$, $s_2 = 7$, $s_3 = 160$, $s_1 = 332381$, *etc.*

THEOREM 4.6. (*McLean* [4]). $s_n - 1 = \sum_{r=1}^{n} \binom{n}{r} \prod_{i=1}^{r} (r - i + 1)^{2^i}$.

A simplified characterization of idempotent monoids can be achieved by noting the symmetry between possible initials and terminals of words in S_n.

COROLLARY 4.7. *The idempotent monoid S_n has isomorphic proper endomorphs* *(i.e., homomorphic images which are subsets of* S_n*)* S_n' *and* S_n'' *such that* $S_n = S_n' \cdot S_n'' = \{xy \mid x \in S_n' \text{ and } y \in S_n''\}$*, for* $n = 2, 3,...$.

Proof. Let S_n' be the quotient of S_n under the congruence $w_1 \sim' w_2$ if and only if w_1 and w_2 have the same initial marks and equivalent initials. Similarly, let S_n'' be the quotient of S_n under the congruence $w_1 \sim'' w_2$ if and only if w_1 and w_2 have the same terminal marks and equivalent terminals. Clearly S_n' and S_n'' are isomorphic and by Lemma 4.3 $S_n = S_n' \cdot S_n''$. S_n will be equal to S_n' and S_n'' only for $n = 1$.

THEOREM 4.8. *Given an alphabet* $A = \{a_1 ,..., a_n\}$*, one can construct a universal finite automaton* $\mathscr{A}_n$ *such that any reduced automaton accepting a noncounting event of order* $\leqslant 1$ *over* A *is a reduced version of* $\mathscr{A}_n$.

Proof. As in the case of $A = \{0, 1\}$, we construct the automaton $\mathscr{A}_n = \{Q, M, [\lambda], -\}$, where $Q = \{[x] \mid x \in S_n\}$, $[\lambda]$ is the initial state and $M([x], a) = [xa]$, $x \in S_n$. As S_n is finite, $\mathscr{A}_n$ is finite and any arbitrary noncounting event of order $\leqslant 1$ over A can be considered as the set of words in A^* that take $\mathscr{A}_n$ from $[\lambda]$ to some set of final states $Q_F \subseteq Q$. The reduced automaton corresponding to this event will simply be a reduced version of $\mathscr{A}_n$.

COROLLARY 4.9. *There exist* 2^{s_n} *regular noncounting events of order* $\leqslant 1$ *over an alphabet of n letters.*

Proof. The number of possible sets of final states in $\mathscr{A}_n$ is 2^{s_n}.

Remark. Recall that by Lemma 4.3 two words w_1 and w_2 take $\mathscr{A}_n$ to the same state if and only if w_1 and w_2 have the same initial marks, the same terminal marks, and equivalent initials and terminals, respectively. The s_n distinct regular languages, each of them having a particular state of $\mathscr{A}_n$ as the final state, form the atoms of the finite Boolean algebra B_1 over the alphabet $A = \{a_1 ,..., a_n\}$. The regular language L corresponding to a particular initial u, initial mark a_k , terminal mark a_p , and terminal v can be expressed by a star-free expression as follows: $L = L_u a_k A^* \cap A^* a_p L_v = L_u a_k I \cap I a_p L_v$, where L_u and L_v represent the sets of words in S_n that are equivalent to u and v, respectively. L_u and L_v can in turn be replaced by expressions similar to L. Note, however, that u and v can contain at most $(n - 1)$ letters and instead of I one must use $(A - a_k)^* = \overline{Ia_kI}$ for L_u and $\overline{Ia_pI}$ for L_v . The recursive application of this process will result in a star-free expression for L in at most $(n - 1)$ steps.

Our final result in this section sets an upper bound on the length of representatives of minimal length for the equivalence classes of A^* under the relation $x^2 = x$, for all $x \in A^*$.

THEOREM 4.10. *A word* $w \in S_n$ *can be represented as a product* $a_{i_1} \cdots a_{i_m}$, $i_k \in \{1,..., n\}$*,* $k = 1,..., m$*, with* $m \leqslant 5.2^{n-2} - 2$*, for* $n = 2, 3,...$. *For* $n = 1$, $m \leqslant 1$.

Proof. Let M_n be the desired bound for S_n. By the proof of Lemma 4.3, if $w \in S_{n+1}$ has initial u, initial mark a, terminal mark b and terminal v, then $w \sim uabv$. Therefore, $M_{n+1} \leqslant 2M_n + 2$, as u and v are equivalent, respectively, to words no longer than M_n. Consider the difference equation $M_{n+1} = 2M_n + 2$. Its solution is of the form $M_n = C2^n + D$. $C2^{n+1} + D = 2(C2^n + D) + 2$ implies that $D = -2$. We also have the initial condition $M_2 = 3$ as $a_1 a_2 a_1$ is irreducible in S_2. Therefore, $3 = C2^1 - 2$ and $C = 5/4$. The reason M_1 is treated separately is that $M_2 = 2M_1 + 1$ so that a better bound is obtained if M_2 is used as the initial condition.

The bound obtained above cannot be improved. The proof is by induction on n, with $n = 2$ as the basis. Let $w_n = a_{i_1} \cdots a_{i_m}$ be a word of length m not equivalent to a shorter word over the alphabet $\{a_1, \ldots, a_n\}$, where we can assume that $a_{i_m} = a_1$. Let $w_n' = a_{j_1} \cdots a_{j_m}$, where for each k, $j_k = i_k + 1$, and let w_n^T be the reverse of w_n, $w_n^T = a_{i_m} \cdots a_{i_1}$. Then $w_{n+1} = w_n a_{n+1} a_1 (w_1')^T$ is of length $2m + 2$, and one verifies that w_{n+1} is not equivalent to a shorter word.

5. Noncounting Events of Order ⩾2 over an Alphabet of Two or More Letters

We first establish some preliminary results necessary to show that the index of the congruence $\sim_k$ for $k \geqslant 2$ over an alphabet of two or more letters is infinite. It is sufficient to prove this for $k = 2$ and a two letter alphabet; in the following $A = \{0, 1\}$ and by $\leftrightarrow$, $\sim$ the relations $\leftrightarrow_2$, $\sim_2$ are understood.

Definition 5.1. Let $u \in A^*$; u can be written uniquely in the form $0^{i_1} 1^{i_2} \cdots 0^{i_{2m-1}} 1^{i_{2m}} 0^{i_{2m+1}}$, where $m \geqslant 0$, $i_1 \geqslant 0$, $i_{2m+1} \geqslant 0$, $i_k \geqslant 1$ for $2 \leqslant k \leqslant 2m$. Denote $r_1(u) = 0^{j_1} 1^{j_2} \cdots 1^{j_{2m}} 0^{j_{2m+1}}$, where $j_k = \min(i_k, 2)$. Obviously $u \sim r_1(u)$ for each $u \in A^*$.

Lemma 5.1. *If $u \leftrightarrow v$ then $r_1(u) \leftrightarrow r_1(v)$, for any u, v in A^*.*

Proof. In the definition of $\leftrightarrow$ there are three cases,

(i) $u = v$; trivially also $r_1(u) = r_1(v)$.

(ii) $u = xy^2z$, $v = xy^3z$;

we have the following subcases:

(a) $y = 0^k$ or $y = 1^k$, $k \geqslant 1$, then obviously $r_1(u) = r_1(v)$.

(b) $y \neq 0^k$ and $y \neq 1^k$.

We can then assert the following: for any s, t in A^* there is a prefix s' of $r_1(s)$ such that

$r_1(st) = s'r_1(t)$ and there is a suffix t' of t such that $r_1(st) = r_1(s)\, t'$. Using this we can write:

$$r_1(yz) = r_1(y)\, z', \qquad \text{where } z' \text{ is a suffix of } r_1(z);$$
$$r_1(y^2z) = y'r_1(y)\, z', \qquad \text{where } y' \text{ is a prefix of } r_1(y);$$
$$r_1(y^3z) = y'y'r_1(y)\, z' \qquad \text{(we get the same prefix of } y);$$
$$r_1(u) = r_1(xy^2z) = x'r_1(y^2z) = x'y'r_1(y)\, z', \qquad \text{where } x' \text{ is a prefix of } r_1(x);$$
$$r_1(v) = r_1(xy^3z) = x'y'y'r_1(y)\, z'.$$

Since y' is a prefix of $r_1(y)$ we have shown that $r_1(u) \leftrightarrow r_1(v)$ holds.

(iii) The case $u = xy^3z$, $v = xy^2z$ is symmetric with (ii).

DEFINITION 5.2. Let $u \in A^*$. If u does not have 01 as a substring, denote $r_2(u) = u$. Otherwise, u can be written uniquely in the form $u_0(01)^{i_1}u_1(01)^{i_2} \cdots u_{m-1}(01)^{i_m}u_m$, where $m \geqslant 1$, u_i is in A^* and does not have 01 as a substring for $0 \leqslant i \leqslant m$, $u_i \neq \lambda$ for $1 \leqslant i \leqslant m-1$, and $i_j \geqslant 1$ for $1 \leqslant j \leqslant m$. Then denote $r_2(u) = u_0(01)^{j_1}u_1(01)^{j_2} \cdots u_{m-1}(01)^{j_m}u_m$, where $j_k = \min(i_k, 2)$ for $1 \leqslant k \leqslant m$.

DEFINITION 5.3. A string u in A^* is called *simple* if $u = x(10)^k y$, $k \geqslant 2$, implies $x = x'0$, $y = 1y'$, for some x' and y' in A^*.

LEMMA 5.2. *Let $u, v \in A^*$, with u simple $u \leftrightarrow v$. Then v is simple and $r_2(u) \leftrightarrow r_2(v)$.*

Proof. The first part of lemma can be easily verified. To prove the second part we consider the three cases in the definition of $\leftrightarrow$;

(i) $u = v$, trivially also $r_2(u) = r_2(v)$;

(ii) $u = xy^2z$, $v = xy^3z$;

we have the following subcases:

(a) $y = (01)^k$ or $y = (10)^k$, then obviously $r_2(u) = r_2(v)$;

(b) $y \neq (01)^k$ and $y \neq (10)^k$.

The following two assertions can be easily verified:

Assertion A. If $s \neq s_1 0$ or $t \neq 10101t_1$ for all $s_1, t_1 \in A^*$, then there exists a prefix s' of $r_2(s)$ such that $r_2(st) = s'r_2(t)$.

Assertion B. If $s \neq t_1 01010$ or $t \neq 1t_1$ for all $s_1, t_1 \in A^*$, then there exists a suffix t' of $r_2(t)$ such that $r_2(st) = r_2(s)\, t'$.

We consider now three subcases of (b);

(b1) Let $y = 10101y'$ for some y'. Then, since u is simple, the last symbol of both x and y must be 0 and we can "move the squared and cubed substrings one symbol to the right," i.e. there exist a, b, c in A^* such that $u = ab^2c$, $v = ab^3c$, $x = a0$, $0y = b0$ and $0z = c$. Assertion A holds for pairs of strings (a, b) and (b, b); since the first symbol of c is 0, Assertion B is fulfilled for the pairs of strings (b, c).

(b2) Let $y = y'01010$ for some y'. Then since u is simple, the first symbol of both y and z must be 1 and there exist a, b, c in A^* such that $u = ab^2c$, $v = ab^3c$, $x1 = a$, $y1 = 1b$ and $z = 1c$. Assertion B is fulfilled for the pair (b, c); since the last symbol of a is 1 also Assertion A is fulfilled for pairs (a, b) and (b, b).

(b3) If y is neither of the form $10101y'$ nor of the form $y'01010$ let $a = x$, $b = y$, $c = z$. Therefore, in all subcases of (b) we have:

By Assertion B

$$r_2(bc) = r_2(b)\, c', \qquad \text{where } c' \text{ is a suffix of } r_2(c);$$

by Assertion A

$$r_2(b^2c) = b'r_2(bc) = b'r_2(b)\, c', \qquad \text{where } b' \text{ is a prefix of } r_2(b);$$

by Assertion A

$$r_2(u) = r_2(ab^2c) = a'r_2(b^2c) = a'b'r_2(b)\, c', \qquad \text{where } a' \text{ is a prefix of } r_2(a);$$

and similarly

$$r_2(v) = r_2(ab^3c) = a'b'b'r_2(b)\, c'.$$

Since b' is a prefix of $r_2(b)$ we have shown that $r_2(u) \leftrightarrow r_2(v)$ holds.

(iii) The case $u = xy^3z$, $v = xy^2z$, is symmetric to (ii).

Note. The assumption of simplicity in Lemma 5.2 is essential as shown by following example:

$$u = 0(10101)(10101), \qquad v = 0(10101)(10101)(10101)$$
$$r_2(u) = 010110101, \qquad r_2(v) = 01011010110101.$$

Obviously, $u \leftrightarrow v$ and $r_2(u) \sim r_2(v)$, but $r_2(u) \not\leftrightarrow r_2(v)$.

DEFINITION 5.4. For w in A^* the *reduced form* of w is defined by $r(w) = r_2(r_1(w))$.

COROLLARY 5.3. *If $u, v \in A^*$, u is simple and $u \leftrightarrow v$ then v is simple and $r(u) \leftrightarrow r(v)$.*

Proof. Follows from Lemma 5.1, Lemma 5.2, and the fact that $r_1(u)$ is simple for simple u.

DEFINITION 5.5. Let h be the homomorphism from A^* into A^* defined by $h(\lambda) = \lambda$, $h(0) = 001$, $h(1) = 011$, $h(xy) = h(x)h(y)$.

LEMMA 5.4. *Let $x \in A^*$ and $h(x)$ be of the form uw^pv, where $u, v, w \in A^*$, $| w | \geqslant 3$, $p \geqslant 2$. ($| w |$ is the length of w.) Then there exist $a, b, c \in A^*$ such that $x = ab^pc$ and $h(ab^kc) = uw^kv$ for each $k \geqslant 2$.*

Note. We have $h(x) = h(ab^pc) = h(a)(h(b))^p h(c)$ but not necessarily $h(a) = u$, $h(b) = w$, $h(c) = v$.

Proof. Since w^2 is a subword of $h(x)$, w must be of one of the following forms for some $t \in A^*$.

(a) $001t$,
(b) $011t$,
(c) $01t0$,
(d) $11t0$,
(e) $1t00$,
(f) $1t01$.

Cases (a) *and* (b). Since w^2 is a substring of $h(x)$ of the form $001t001t$ or $011t011t$ the length of w has to be a multiple of three and thus there exist a, b, c in A^* such that $u = h(a)$, $w = h(b)$, $v = h(c)$ and a, b, c fulfil the stated requirements.

Case (c). Since w^2 is a substring of $h(x)$ of the form $01t001t0$ either $w = 010$ if $t = \lambda$ or $w = 010t0$, $t \in A^+$. In both cases the last symbol of u must be 0 and we can write $h(x) = u_1w_1{}^pv_1$, where $u_10 = u$, $w_10 = 0w$ and $v_1 = 0v$. Since the lengths of u_1, w_1 and v_1 are multiples of three there exist $a, b, c \in A^*$ such that $h(a) = u_1$, $h(b) = w$, $h(c) = v_1$, $x = ab^pc$. Obviously, $h(ab^kc) = h(a)(h(b))^k h(c) = u_1w_1{}^kv_1$ for each $k \geqslant 2$. Since $w_1{}^k0 = 0w^k$ for each k, also $h(ab^kc) = uw^kv$ holds for each $k \geqslant 2$.

Case (d). Since w is a substring of $h(x)$ of the form $11t011t0$ the last symbol of u must be 0; the rest is the same as in the case (c).

By interchaning zeros and ones and reversing the strings, case (e) becomes case (d) and case (f) becomes case (c).

LEMMA 5.5. *$x \sim y$ if and only if $h(x) \sim h(y)$.*

Proof. 1. Let

$$x = x_1 \leftrightarrow x_2 \leftrightarrow \cdots \leftrightarrow x_n = y,$$

then obviously

$$h(x) = h(x_1) \leftrightarrow h(x_2) \leftrightarrow \cdots \leftrightarrow h(x_n) = h(y).$$

2. Let

$$h(x) = t_1 \leftrightarrow t_2 \leftrightarrow \cdots \leftrightarrow t_n = h(y).$$

Denote $z_i = r(t_i)$ for $1 \leqslant i \leqslant n$. For every s in A^* $h(s)$ has no substrings of the form w^3, where $1 \leqslant |w| \leqslant 2$; thus $r(h(s)) = h(s)$ for every s. Furthermore, it is clear that $h(s)$ is simple. Hence, by the repeated use of Corollary 5.3

$$h(x) = z_1 \leftrightarrow z_2 \leftrightarrow \cdots \leftrightarrow z_n = h(y).$$

We now construct by induction a sequence of strings $x_1, x_2, ..., x_n$ in A^* such that $h(x_i) = z_i$ for $1 \leqslant i \leqslant n$ and

$$x = x_1 \leftrightarrow x_2 \leftrightarrow \cdots \leftrightarrow x_n = y.$$

Define $x_1 = x$. Suppose now that $x_1, x_2, ..., x_k$ have already been constructed for some k, $1 \leqslant k < n$. Since $z_k \leftrightarrow z_{k+1}$, we have three possibilities by the definition of $\leftrightarrow$;

(i) $z_k = z_{k+1}$; define $x_{k+1} = x_k$, then $h(x_{k+1}) = h(x_k) = z_{k+1} = z_k$;

(ii) $z_k = uw^2v$ and $z_{k+1} = uw^3v$ for some $u, v \in A^*$, $w \in A^+$.

By the induction assumption $z_k = h(x_k)$ for some x_k and therefore z_k has no substring of the form $(10)^j$, $j \geqslant 2$. Since z_{k+1} is reduced it does not contain any substring of the form s^m, $m \geqslant 3$ for $s \in \{0, 1, 00, 11, 01\}$. Since w^2 is a substring of z_k and w^3 is a substring of z_{k+1} it follows that $|w| \geqslant 3$. Thus Lemma 5.4 is applicable and there exist a, b, c in A^* such that $x_k = ab^2c$ and $h(ab^3c) = uw^3v = z_{k+1}$. Define $x_{k+1} = ab^3c$; obviously $h(x_{k+1}) = z_{k+1}$ and $x_k \leftrightarrow x_{k+1}$.

(iii) $z_k = uw^3v$ and $z_{k+1} = uw^2v$ for some $u, v \in A^*$, $w \in A^+$.

In this case w^3 is a substring of z_k which is both reduced and by the induction assumption it is without any substring of the form $(10)^j$, $j \geqslant 2$. Thus again $|w| \geqslant 3$ and by Lemma 5.4 there exist a, b, c in A^* such that $x_k = ab^3c$ and $h(ab^2c) = uw^2v = z_{k+1}$. Define $x_{k+1} = ab^2c$; obviously $h(x_{k+1}) = z_{k+1}$ and $x_k \leftrightarrow x_{k+1}$.

THEOREM 5.6. *Let $i, j \geqslant 1$ then* $h^i(0) \sim h^j(0)$ if and only if $i = j$.

Proof. 1. If $i = j$ then $h^i(0) = h^j(0)$.

2. There is at least one occurrence of the symbol 1 in $h^k(0)$ for each $k \geqslant 1$, thus $h^k(0) \not\sim 0$ for $k \geqslant 1$. Assume $h^i(0) \sim h^j(0)$ for some $i > j \geqslant 1$. Then $h^{i-j}(0) \sim 0$ by the repeated use of Lemma 5.4. This is a contradiction.

COROLLARY 5.7. *The index of the congruence relation $\sim_2$ is infinite.*

In [9] it was shown that there exists an infinite subset X of $\{0, 1, 2\}^*$ such that no string in X has the form xy^2z with y in $\{0, 1, 2\}^+$. It was also shown in [9] that there exists an infinite subset Y of $\{0, 1\}^*$ such that no string in Y is of the form xy^3z, with y in $\{0, 1\}^+$.

LEMMA 5.8. *Let $E_1 \subset X$ and $E_2 \subset Y$ then E_1 and E_2 are noncounting events of order 2 and of order 3, respectively.*

Proof. Obvious.

From the *uvwxy*-Theorem [2] it follows that X and Y are not context-free languages. Because of Lemma 5.8 we can construct by diagonalisation a noncounting event of order 3 over an alphabet with at least two letters or a noncounting event of order 2 over an alphabet with at least three letters which is not context-sensitive or even recursively enumerable. Such events exist also for order 2 and a 2-letter alphabet. In this case we can diagonalise over an infinite set of mutually nonequivalent strings which was proven to exist in Theorem 5.6 and then consider not the direct result of the diagonalisation but the union of corresponding congruence classes.

In Lemma 2.3 we have shown that the family of all noncounting events of order $\leqslant k$ over an finite alphabet is a Boolean algebra. This Boolean algebra is atomic, the atoms are the congruence classes of order k. Theorem 4.4 shows that B_k is finite for $k \leqslant 1$ and Corollary 5.7 shows that B_k is infinite for $k \geqslant 2$ if the alphabet has two or more letters.

Let R_k be the family of all regular noncounting events of order k over an alphabet A. R_k is also a Boolean algebra, a subalgebra (proper for $k \geqslant 2$ and an alphabet with at least two letters) of B_k. For $k \leqslant 1$ the equivalence classes are regular, thus B_k and R_k are identical. For order $\geqslant 2$ and an alphabet with at least three letters or for order $\geqslant 3$ and an alphabet with at least two letters B_k is infinite since in these cases there exist an infinite number of congruence classes containing exactly one element which are of course regular. These singletons are elements of X, Y, respectively.

We now show that R_2 is also infinite for a 2-letter alphabet. In the following $A = \{0, 1\}$, $k = 2$, and the maps h and r are as defined above.

LEMMA 5.9. *Let $x, z \in A^*$, with $h(x) \sim r(z)$. Then there exists a $y \in A^*$ such that $h(y) = r(z)$.*

Proof. Follows from the proof of Lemma 5.5.

LEMMA 5.10. $[h^{i+1}(q)] = r^{-1}(h([h^i(q)]))$ *for each $q \in A^*$ and $i \geqslant 0$.*

Proof. Let $w \in [h^{i+1}(q)]$, i.e. $w \sim h^{i+1}(q)$. Since $r(w) \sim w$ we have $r(w) \sim h^{i+1}(q)$. By Lemma 5.9 there exists a y such that $h(y) = r(w)$. By Lemma 5.5 $y \sim h^i(q)$, thus $r(w) \in h([h^i(q)])$. Hence $r^{-1}(r(w)) \subset r^{-1}(h([h^i(q)]))$. Since $w \in r^{-1}(r(w))$ and w is an arbitrary element of $[h^{i+1}(q)]$ we have $[h^{i+1}(q)] \subset r^{-1}(h([h^i(q)]))$. The reverse inclusion is obvious.

THEOREM 5.11. $[h^i(0)]$ *is regular for each $i \geqslant 0$.*

Proof. In [2] a device called sequential transducer is defined and it is shown that regular sets are preserved by sequential transducer mappings. The map r^{-1} is realized by the sequential transducer $S = (K, A, A, H, s_\lambda)$, where

$$K = \{s_\lambda, s_0, s_1, s_{00}, s_{01}, s_{11}, s_{0101}\}$$

and

$$\begin{aligned} H = {} & \{(q, 0, 0, s_0) : q \in (K - \{s_0, s_{00}\})\} \cup \{(q, 1, 1, s_1) : q \in \{s_\lambda, s_0, s_{00}\}\} \\ & \cup \{(q, 1, 1, s_{11}) : q \in \{s_1, s_{01}, s_{0101}\}\} \\ & \cup \{(s_0, 1, 1, s_{01}), (s_{00}, 1, 1, s_{01}), (s_0, 0, 0, s_{00}), (s_{01}, 01, 01, s_{0101}), \\ & (s_{00}, \lambda, 0, s_{00}), (s_{11}, \lambda, 1, s_{11}), (s_{0101}, \lambda, 01, s_{0101})\}. \end{aligned}$$

Since regular sets are closed under homomorphisms, $r^{-1}(h([h^i(0)]))$ is regular if $[h^i(0)]$ is regular. Since $[0]$ is regular the proof can be completed by induction using Lemma 5.10.

COROLLARY 5.12. *For order $k \geqslant 2$ if A has at least two letters the Boolean algebra R_k is infinite.*

Proof. We can obviously restrict ourselves to $k = 2$ and $A = \{0, 1\}$. By Theorem 5.11 $\{[h^i(0)] : i = 0, 1, ...\} \subset R_2$. By Theorem 5.6, $[h^i(0)] \neq [h^j(0)]$ for $i \neq j$.

REFERENCES

1. R. S. COHEN AND J. A. BRZOZOWSKI, On star-free events, *in* "Proceedings of the Hawaii International Conference on System Sciences," pp. 1–4, University of Hawaii Press, Honolulu, Hawaii, 1968.
2. S. GINSBURG, "The Mathematical Theory of Context-Free Languages," McGraw-Hill, New York, 1966.
3. J. A. GREEN AND D. REES, On semigroups in which $x^r = x$, *Proc. Cambridge Philos. Soc.* **48** (1952), 35–40.
4. D. MCLEAN, Idempotent semigroups, *Amer. Math. Monthly* **61** (1954), 110–113.
5. A. R. MEYER, A note on star-free events, *J. Amer. Comput. Mach.* **16** (1969), 220–225.
6. S. PAPERT AND R. MCNAUGHTON, On topological events, *in* "Theory of Automata," University of Michigan Engineering Summer Conference, Ann Arbor, Mich., 1966.
7. M. O. RABIN AND D. SCOTT, Finite Automata and their decision problems, *IBM J. Res. Develop.* **3** (1959), 114–125.
8. M. P. SCHÜTZENBERGER, On a family of sets related to McNaughtons L-Language, *in* "Automata Theory," (E. R. Caianiello, Ed.), pp. 320–324, Academic Press, New York, 1966.
9. A. THUE, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Videnskabsselskabets Skrifter, I Mat.-Nat. Kl.*, Christiania, 1912.