Discrete Mathematics 16 (1976) 299-307.

RUN LANGUAGES*

J.A. BRZOZOWSKI
Computer Science Department, University of Waterloo, Waterloo, Ont., Canada

Received 28 July 1975

The family $\mathscr{R}$ of run languages is a generalization of the family of finite/cofinite languages. For an alphabet A, the family $\mathscr{W}_{\oplus}$ of "$\oplus$words" consists of finite products of factors each of which is of the form $\{a\}$ or a^+, $a \in A$, and $\mathscr{R}$ is the Boolean closure of $\mathscr{W}_{\oplus}$. It is shown that each run language or its complement is a finite union of $\oplus$words, and that $\mathscr{R}$ is also the Boolean closure of $\mathscr{W}_{+}$, the set of "words" over $\{a^+: a \in A\}$. Lastly we prove that $\mathscr{R}$ is contained in the family γ_1 of languages whose syntactic monoids are $\mathscr{J}$-trivial, and that for a two-letter alphabet $\mathscr{R} = \gamma_1$.

1. Introduction

The initial goal of this work is to provide a better understanding of a family γ_1 of languages whose syntactic monoids are $\mathscr{J}$-trivial. This family plays a key role in the study of star-free (or aperiodic) languages [2, 3]. In case the cardinality of the alphabet is ≤ 2, γ_1 coincides with the family $\mathscr{R}$ of run languages introduced here. Otherwise, $\mathscr{R}$ is properly contained in γ_1. Run languages are interesting in their own right since they are a generalization of finite/cofinite languages.

We use the following notation. The cardinality of a set S is denoted card S, and $S \cup S'$, $S \cap S'$, $S - S'$ and $\bar{S}$ denote union, intersection, difference and complement. The empty set is $\emptyset$. If A is a finite non-empty alphabet, A^+ (respectively A^*) is the free semigroup (respectively free monoid) generated by A. The empty word is denoted by 1. Any subset L or A^* is a language over A. The product or concatenation of two languages is denoted by $L \cdot L'$ or LL'. The subsemigroup and submonoid of A^* generated by $L \subset A^*$ are L^+ and L^* respectively. The length of a word $w \in A^*$ is $|w|$.

The family of all languages over A is a monoid under concatenation and a Boolean algebra under finite union, finite intersection and complement. For any family $\mathscr{X}$ of languages over A, $\mathscr{X}M$ and $\mathscr{X}B$ denote the monoid and Boolean algebra (respectively) generated by $\mathscr{X}$.

* Part of this work was done while the author was at the University of Paris VI and VII, under the scientific exchange program between Canada and France. The financial assistance provided by this program is gratefully acknowledged.

2. Run languages

We shall use the following finite families over a given alphabet A:

$$\mathscr{L} = \{\{a\}: a \in A\}, \quad L_+ = \{a^+: a \in A\}, \quad \mathscr{L}_* = \{a^*: a \in A\}, \quad L_\oplus = \mathscr{L} \cup \mathscr{L}_+.$$

In this section we consider $L_\oplus$ as an extended alphabet of $\oplus$*letters*, and $\mathscr{W}_\oplus = \mathscr{L}_\oplus M$ as the set of words over this alphabet, i.e. $\oplus$*words*. Our main interest is in the elements of $\mathscr{W}_\oplus$ as languages over A. To simplify notation we often write w instead of $\{w\}$ for $w \in A^*$, when there is no risk of ambiguity.

Any language $u \in \mathscr{W}_\oplus$, $u \neq 1$, can be written in the form $u = u_1 \ldots u_p$, where $p \geq 1$ and

(a) $u_i \subset a_i^+$, for some $a_i \in A$, $1 \leq i \leq p$;

(b) $a_{i+1} \neq a_i$, $1 \leq i \leq p-1$.

(Except for condition (b), the a_i are not necessarily distinct.) Furthermore, since $aa^+ = a^+a = a^+a^+$ for all $a \in A$, we can assume:

(c) either $u_i = a_i^n$ or $u_i = (a_i^+)^n$, $n \geq 1$.

We say that $u \in \mathscr{W}_\oplus$ is in *run form* iff either $u = 1$ or u satisfies (a), (b) and (c) above. Define the *run length* $\|u\|$ of u as follows. Let $\|1\| = 0$ and for $u \neq 1$, $\|u\| = p$, the number of factors in the run form of u.

Note that, if $u \neq 1$ is in run form $u = u_1 \ldots u_p$ and $w \in A^*$, then $w \in u$ iff $w = a_1^{n_1} \ldots a_p^{n_p}$, where $a_i^{n_i} \in u_i$, $1 \leq i \leq p$. This factorization of w into p factors is unique; in fact this is the run form of w considered as an element of $\mathscr{W}_\oplus$.

It is convenient to associate with each $u \in \mathscr{W}_\oplus$ an element $u\rho$ of $\mathscr{L}_+M = \mathscr{W}_+$. Let $1\rho = 1$ and, for $u \neq 1$, let $u\rho = a_1^+ \ldots a_p^+$. Note that $w \in u$ implies $w\rho = u\rho$.

The family $\mathscr{R}$ of run languages is now defined as $\mathscr{R} = \mathscr{W}_\oplus B = L_\oplus MB$.

We illustrate these notions by some examples. The $\oplus$word $u' = aa^+baac^+$ is not in run form, but can be rewritten as $u = (a^+)^2ba^2c^+ = u_1u_2u_3u_4$, $u_1 = (a^+)^2$, $u_2 = b$, $u_3 = a^2$, $u_4 = c^+$, which is in run form. If $w \in u$ then w has the form $a^nba^2c^m$, where $n \geq 2$, $m \geq 1$. Finally $u\rho = w\rho = a^+b^+a^+c^+$.

3. Finite/cofinite languages over $L_\oplus$

A language $L \subset A^*$ is said to be $\oplus$*finite* iff it is a finite union of $\oplus$words. Let $\mathscr{F}_\oplus$ be the family of $\oplus$finite languages, and let $\mathscr{C}_\oplus = \{L \subset A^*: \bar{L} \in \mathscr{F}_\oplus\}$ be the family of $\oplus$*cofinite* languages.

Lemma 3.1. *The intersection of any two $\oplus$finite languages is $\oplus$finite.*

Proof. The problem easily reduces to showing that $u \cap u'$ is $\oplus$finite for any two $\oplus$words u and u'. First note that $u \cap u' \neq \emptyset$ implies $u\rho = u'\rho$. Hence we can assume that the run forms of u and u' are $u = u_1 \ldots u_p$, $u' = u'_1 \ldots u'_p$ when $u \cap u' \neq \emptyset$. Under this condition we claim that

$$u \cap u' = (u_1 \cap u_1') \cdot \ldots \cdot (u_p \cap u_p'). \tag{1}$$

The containment $\supset$ holds for any languages $u_1, \ldots, u_p$, $u_1', \ldots, u_p'$ and u, u' such that $u = u_1 \ldots u_p$ and $u' = u_1' \ldots u_p'$. Conversely, suppose $w \in u \cap u'$. Then $w = w_1 \ldots w_p$, where $w_1 \in u_1, \ldots, w_p \in u_p$, and $w = w_1' \ldots w_p'$, where $w_1' \in u_1', \ldots, w_p' \in u_p'$. But both these factorizations of w correspond to the unique run form of w; hence they are identical and (1) follows.

Next we show for $\|u\| = \|u'\| = 1$, i.e. for the one-letter alphabet, that $u \cap u'$ is $\oplus$finite. The following listing covers all cases for $A = \{a\}$, n, $m \geq 1$:

$$a^m \cap a^m = a^m,$$

$$a^m \cap a^n = \emptyset, \quad \text{if } m \neq n,$$

$$a^m \cap (a^+)^n = a^m, \quad \text{if } m \geq n,$$

$$a^m \cap (a^+)^n = \emptyset, \quad \text{if } m < n,$$

$$(a^+)^m \cap (a^+)^n = (a^+)^{\max(m,n)}.$$

In each case the intersection is either empty or a $\oplus$word. By (1), $u \cap u'$ is a product of such intersections; hence $u \cap u'$ is either empty or a $\oplus$word. In any case it is $\oplus$finite.

Lemma 3.2. *The difference of any two $\oplus$finite languages is $\oplus$finite.*

Proof. As in Lemma 3.1, the problem reduces to showing that $u - u'$ is $\oplus$finite for any two $\oplus$words u and u'. Note that $u\rho \neq u'\rho$ implies $u \cap u' = \emptyset$ and $u - u' = u$, which is $\oplus$finite. Hence assume $u\rho = u'\rho$ and $u = u_1 \ldots u_p$, $u' = u_1' \ldots u_p'$ are in run form. Also let $\bar{u} = u_2 \ldots u_p$, $\bar{u}' = u_2' \ldots u_p'$. (The case $u = 1$ is trivial so we can assume $p \geq 1$; $\bar{u}$ and $\bar{u}'$ may be the empty word.) Under these conditions we claim that

$$u - u' = (u_1 - u_1') \cdot \bar{u} \cup u_1 \cdot (\bar{u} - \bar{u}'). \tag{2}$$

For suppose $w \in u - u'$. Then $w = w_1\bar{w}$, $w_1 \in u_1$, $\bar{w} \in \bar{u}$ and $w \notin u'$. Thus either $w_1 \notin u_1'$ or $\bar{w} \notin \bar{u}'$, i.e. $w \in (u_1 - u_1')\bar{u}$ or $w \in u_1(\bar{u} - \bar{u}')$. Conversely, suppose $w \in (u_1 - u_1')\bar{u}$. Then $w = w_1\bar{w}$, $w_1 \in u_1$, $w_1 \notin u_1'$ and $\bar{w} \in \bar{u}$. Clearly $w \in u$. If also $w \in u'$ then $w_1 \in u_1'$ since the run form of w is unique. This is a contradiction and $(u_1 - u_1') \cdot \bar{u} \subset u - u'$. The case $w \in u_1(\bar{u} - \bar{u}')$ follows similarly and (2) holds.

We now prove the lemma for the one-letter alphabet by the following cases, where $A = \{a\}$ and n, $m \geq 1$:

$$a^m - a^n = a^m, \quad \text{if } m \neq n,$$

$$a^m - a^m = \emptyset,$$

$$a^m - (a^+)^n = a^m, \quad \text{if } m < n,$$

$$a^m - (a^+)^n = \emptyset, \quad \text{if } m \geq n,$$

$$
\begin{aligned}
(a^+)^n - a^m &= a^n \cup a^{n+1} \cup \ldots \cup a^{m-1} \cup (a^+)^{m+1}, && \text{if } m > n,\\
&= (a^+)^{m+1}, && \text{if } m = n,\\
&= (a^+)^n, && \text{if } m < n,\\
(a^+)^n - (a^+)^m &= a^n \cup a^{n+1} \cup \ldots \cup a^{m-1}, && \text{if } m > n,\\
&= \emptyset, && \text{if } m \leqslant n.
\end{aligned}
$$

In all cases the difference is ⊕finite. Now proceed by induction on p. Assume that for all u such that $\|u\| = p-1$, the difference $u - u'$ is ⊕finite. The induction step follows from (2).

We now show that each run language is either ⊕finite or ⊕cofinite.

Theorem 3.3. $\mathcal{R} = \mathcal{F}_\oplus \cup \mathcal{C}_\oplus$.

Proof. Evidently, $\mathcal{R} = \mathcal{L}_\oplus MB \supset \mathcal{F}_\oplus \cup \mathcal{C}_\oplus$. Since $\mathcal{L}_\oplus M \subset \mathcal{F}_\oplus$, it is sufficient to verify that $\mathcal{F}_\oplus \cup \mathcal{C}_\oplus$ is a Boolean algebra in order to show that $\mathcal{L}_\oplus MB \subset \mathcal{F}_\oplus \cup \mathcal{C}_\oplus$. Closure under complementation is obvious. Also, for L, $L' \in \mathcal{F}_\oplus$,

$$L \cup L' \in \mathcal{F}_\oplus,$$

$$L \cup \bar{L}' = \overline{L' - L} \in \mathcal{C}_\oplus, \quad \text{since } L' - L \in \mathcal{F}_\oplus \text{ by Lemma 3.2},$$

$$\bar{L} \cup \bar{L}' = \overline{L \cap L'} \in \mathcal{C}_\oplus, \quad \text{since } L \cap L' \in \mathcal{F}_\oplus \text{ by Lemma 3.1}.$$

Hence $\mathcal{F}_\oplus \cup \mathcal{C}_\oplus$ is also closed under union.

4. The hierarchy generated by $\mathcal{L}_\oplus$

The family $\mathfrak{A}$ of aperiodic languages can be generated by $\mathcal{L}$ in the sense that

$$\mathfrak{A} = \bigcup_{n \geqslant 0} \mathcal{L}(MB)^n = \bigcup_{n \geqslant 0} \mathcal{L}(BM)^n.$$

Furthermore it is unimportant whether $\mathcal{L}$ is first closed under Boolean operations or under concatenation [1], since $\mathcal{L}(MB)^2 = \mathcal{L}BMB$, i.e. the two sequences $[\mathcal{L}(MB)(MB)^n]$ and $[\mathcal{L}B(MB)^n]$ are identical for $n \geqslant 1$. An analogous situation exists if $\mathcal{L}_\oplus$ is used instead of $\mathcal{L}$.

Theorem 4.1. $\mathcal{L}_\oplus MBMB = \mathcal{L}_\oplus BMB$.

Proof. Obviously, $\mathcal{L}_\oplus MBMB \supset \mathcal{L}_\oplus BMB$. Conversely, it suffices to show that $\mathcal{L}_\oplus MBM \subset \mathcal{L}_\oplus BMB$. By Theorem 3.3, $\mathcal{L}_\oplus MB = \mathcal{F}_\oplus \cup \mathcal{C}_\oplus$. Hence we need to show that every product of factors from $\mathcal{F}_\oplus \cup \mathcal{C}_\oplus$ is in $\mathcal{L}_\oplus BMB$.

Let L be $\oplus$cofinite and let $p-1$ be the run length of the run-longest $\oplus$word in $\bar{L}$. Clearly L contains the union U_p of all the $\oplus$words of run length $\geq p$. We have

$$U_p = \Big[\bigcup a_1^+ \dots a_p^+ \Big] A^*,$$

where the union is over all $(a_1, \dots, a_p)$ such that $a_i \in A$, $1 \leq i \leq p$, and $a_i \neq a_{i+1}$, $1 \leq i \leq p-1$. For a finite alphabet A there is only a finite number of such $\oplus$words $(a_1^+ \dots a_p^+)$. Hence for every $\oplus$cofinite L we can write

$$L = L_1 \cup L_2 A^* = L_1 \cup A^* L_2,$$

where L_1 and L_2 are $\oplus$finite.

It now follows that every $L \in \mathscr{L}_\oplus MBM = (\mathscr{F}_\oplus \cup \mathscr{C}_\oplus)M$ is a finite union of products of $\oplus$words and A^*. Now $A^* \cup \mathscr{L}_\oplus \subset \mathscr{L}_\oplus B$ and $(A \cup \mathscr{L}_\oplus)M = (A^* \cup \mathscr{W}_\oplus)M \subset \mathscr{L}_\oplus BM$. Thus every product of $\oplus$words and A^* is in $\mathscr{L}_\oplus BM$ and $\mathscr{L}_\oplus MBMB = \mathscr{L}_\oplus BMB$.

5. Run languages over $\mathscr{L}_+$

The family $\mathscr{R}$ of run languages can be obtained from a reduced basis, namely $\mathscr{L}_+$, but at the expense of the finite/cofinite analogy.

Theorem 5.1. $\mathscr{R} = \mathscr{L}_+ MB$.

Proof. Since $\mathscr{L}_\oplus \supset \mathscr{L}_+$, $\mathscr{R} \supset \mathscr{L}_+ MB$. Conversely we show $\mathscr{W}_\oplus \subset \mathscr{L}_+ MB$. Clearly $1 \in \mathscr{L}_+ MB$. For $\|u\| = 1$, either $u = (a^+)^n$ (and $u \in \mathscr{L}_+ M$) or $u = a^n = (a^+)^n - (a^+)^{n+1}$ and $u \in \mathscr{L}_+ MB$. In general, let $u = u_1 \dots u_p$ be in run form. If u has no finite factors (i.e. factors of the form a^n) then $u \in \mathscr{W}_+$. Otherwise suppose $u_i = a^n$ and let $u' = u_1 \dots u_{i-1}$ and $u'' = u_{i+1} \dots u_p$. Then one easily verifies that

$$u = u'(a^+)^n u'' - u'(a^+)^{n+1} u''.$$

The proof now follows by induction on the number of finite factors in u.

From the proof of Lemma 3.1 it follows that the intersection of two $+$finite languages (i.e. unions of $+$words) is $+$finite. However, the statement corresponding to Lemma 3.2 is false. For let $A = \{a\}$. Then $\{a\} = \overline{\{1\} \cup (a^+)^2}$ is $+$cofinite. However $\{a^2\} = (a^+)^2 - (a^+)^3$ is clearly not $+$finite. Also $\overline{a^2} = \{1\} \cup \{a\} \cup (a^+)^3$ is not $+$finite since $\{a\}$ is $+$cofinite. Therefore $\mathscr{F}_+ \cup \mathscr{C}_+$ is not a Boolean algebra.

6. Containment in γ_1

For $w = a_1 \dots a_p \in A^*$ define $[w$ to be

$$[w = [(a_1 \dots a_p) = A^* a_1 A^* a_2 \dots A^* a_p A^*.$$

Thus $[w$ is the set of all words that contain w as a subword in the above sense. Similarly, let $[\mathscr{W} = \{[w : w \in A^*\}$. The family γ_1 (see [2, 3]) is $\gamma_1 = ([\mathscr{W})B$. We now show that $\mathscr{R} \subset \gamma_1$.

Let ω be the homomorphism, $\omega : \mathscr{W}_+ \to A^*$, which changes + words into words, i.e. $1\omega = 1$ and $a^+\omega = a$ for all $a \in A$. In fact ω is an isomorphism between $\mathscr{W}_+$ and A^*.

Lemma 6.1. *Let $v \in \mathscr{W}_+$. Then $v = (v\rho) \cap [(v\omega)$.*

Proof. Note that $v \subset v\rho$ since ρ replaces $(a^+)^n$ by a^+. Also $v\omega$ is the shortest word in v; hence $v \subset [(v\omega)$. This proves the containment $\subset$. If $v = (a_1^+)^{n_1} \dots (a_p^+)^{n_p}$ is in run form, then $v\rho = a_1^+ \dots a_p^+$ and $v\omega = a_1^{n_1} \dots a_p^{n_p}$. Hence $w \in v\rho$ implies $w = a_1^{m_1} \dots a_p^{m_p}$ and $w \in [(v\omega)$ means $m_i \geqslant n_i$, $1 \leqslant i \leqslant p$, because of the uniqueness of the run form. Hence $v \supset (v\rho) \cap [(v\omega)$, and the lemma follows.

It is now more convenient to consider $\mathscr{W}_* = \mathscr{L}_*M$ rather than $\mathscr{W}_+$. Since $a^* = a^+ \cup 1$ for all $a \in A$, it follows that $\mathscr{L}_*MB \subset \mathscr{L}_+MB$. The converse containment is false since $a^+a^+ \notin \mathscr{L}_*MB$. To prove this let $A = \{a\}$. Then

$$\mathscr{L}_*MB = \{a^*\}MB = \{\{1\}, a^*\}\ B = \{\emptyset, \{1\}, a^+, a^*\},$$

and $(a^+)^2 \notin \mathscr{L}_*MB$. It is however true that an important subset of $\mathscr{L}_+MB$ is contained in $\mathscr{L}_*MB$.

Lemma 6.2. $\mathscr{L}_*MB = (\mathscr{L}_+M)\rho B$.

Proof. Suppose $v \in (\mathscr{L}_+M)\rho = \mathscr{W}_+\rho$. If $v = 1$ then $v \in \mathscr{W}_*$. Hence assume $\|v\| \geqslant 1$, i.e. $v = a_1^+ \dots a_p^+$, $p \geqslant 1$. Let $v_* = a_1^* \dots a_p^*$ and, for each i, $1 \leqslant i \leqslant p$, let $v_*^{-i} = a_1^* \dots a_{i-1}^* a_{i+1}^* \dots a_p^*$, where $(a_1^*)^{-1} = 1$. We now claim that

$$v \doteq v_* - \bigcup_{i=1}^{p} v_*^{-i}. \tag{4}$$

For $w \in v$ implies $w = a_1^{n_1} \dots a_p^{n_p}$, $n_i > 0$, $1 \leqslant i \leqslant p$. Hence $w \in v_*$ and $w \notin v_*^{-i}$ for all i. Conversely, if $w \in v_*$ then $w = w_1 \dots w_p$, $w_i \in a_i^*$, $1 \leqslant i \leqslant p$. If $w \notin v$, at least one of the w_i must be 1. Hence $w \in v_*^{-i}$. Hence $w \in v_* - \bigcup_{i=1}^{p} v_*^{-i}$ implies $w \in v$ and $(\mathscr{L}_+M)\rho B \subset \mathscr{L}_*MB$.

The converse containment is shown as follows. Let $x \in \mathscr{L}_*M$, $x = a_1^* \dots a_p^*$, $a_i \neq a_{i+1}$, $1 \leqslant i \leqslant p-1$. Then $x = a_2^* \dots a_p^* \cup a_1^+ a_2^* \dots a_p^*$. If $a_1 \neq a_3$, then $a_1^+ a_2^* \dots a_p^* = a_1^+ a_3^* \dots a_p^* \cup a_1^+ a_2^+ \dots a_p^*$. If $a_1 = a_3$, then $a_1^+ a_2^* \dots a_p^* = a_1^+ a_2^+ a_3^* \dots a_p^* \cup a_1^+ a_4^* \dots a_p^*$, and $a_1 = a_3 \neq a_4$. Proceeding in this manner we can write any $*$word as a finite union of $+$words in $\mathscr{W}_+\rho$. Hence $\mathscr{L}_*MB \subset \mathscr{W}_+\rho B$.

Now let $n = \operatorname{card} A$ and let $\mathscr{X}_1$ be the free monoid generated by n idempotents $a_1^*, \dots, a_n^*$. For example, for $A = \{a, b\}$,

$$\mathscr{X}_1 = \{1, a^*, b^*, a^*b^*, b^*a^*, a^*b^*a^*, \dots\}.$$

The monoid $\mathscr{X}_1$ is partially ordered by inclusion. Note that $x \subset x'$ implies that x is a sub$*$word of x'. Conversely let x be a sub$*$word of x' and let x'' be the element of $\mathscr{X}_1$ corresponding to x, after x has been reduced to the form where no two consecutive $*$letters are equal. Then $x'' \subset x'$. For example, $a^*b^*a^*c^*$ contains $a^*b^*a^*$, $a^*b^*c^*$, a^*c^* (obtained by reducing $a^*a^*c^*$), $b^*a^*c^*$, a^*b^*, b^*a^*, b^*c^*, a^*, b^*, c^* and 1.

For a given element $y \in \mathscr{X}_1$ define $Z_y = \{x : x \not\subset y\}$. Clearly Z_y is an ideal of $\mathscr{X}_1$. Furthermore, the quotient monoid $M_y = \mathscr{X}_1/Z_y$ is finite because all the $*$words in $\mathscr{X}_1$ of length exceeding the length of y are in Z_y. Let $\varphi_y : \mathscr{X}_1 \to M_y$ be the surjective morphism corresponding to the quotient of $\mathscr{X}_1$ by Z_y, i.e. $Z_y\varphi_y = 0$ in M_y. Call $x \in \mathscr{X}_1$ *reduced in* M_y iff there is no x' in $\mathscr{X}_1$ such that $x' \subset x$ and $x\varphi_y = x'\varphi_y$. For example, let $A = \{a, b, c\}$ and $y = a^*b^*$. The reduced $*$words are 1, a^*, b^*, a^*b^*, c^* and b^*a^*. Any other $*$word x satisfies $x\varphi_y = 0$ and contains either c^* or b^*a^*.

Note that, if the length of $y \in \mathscr{X}_1$ (as a $*$word) is k, all $*$words of length $k+1$ are in Z_y. No $*$word x of length $> k+1$ can be reduced in M_y since it has a prefix x' of length $k+1$ in Z_y and $x' \subset x$. Therefore the number of reduced words is finite.

Let $\omega : \mathscr{L}_*M \to A^*$ be the isomorphism that changes $*$words into words, i.e. $1\omega = 1$, $a^*\omega = a$, for all $a \in A$.

Lemma 6.3. *Let* $y \in \mathscr{X}_1$. *Then* $y = \bigcap_{x \in R} \overline{[(x\omega)}$, *where R is the finite set of reduced words x such that* $x\varphi_y = 0$.

Proof. We have

$$1 = \bigcap_{a \in A} \overline{[a} = \bigcap_{a \in A} \overline{[(a^*\omega)}.$$

Henceforth assume that $y = a_1^* \ldots a_p^*$, $p \geq 1$. Suppose $w = b_1^{n_1} \ldots b_q^{n_q}$ is in y, where w is in run form. Then, if $w_* = b_1^* \ldots b_q^*$, we have $w_* \subset y$. If also $w \in [(x\omega)$ for some $x \in R$, then $x\omega$ is a subword of w. This in turn implies that x is a sub$*$word of w_*, i.e. $x \subset w_*$. Thus $x \subset w_* \subset y$, contradicting the fact that $x \not\subset y$ since $x \in Z_y$. Therefore $y \subset \bigcap_{x \in R} \overline{[(x\omega)}$. Conversely suppose $w \notin [(x\omega)$ for any reduced x. Then $w \notin [(x'\omega)$ for any x' such that $x'\varphi_y = 0$. Hence $w_*\varphi_y \neq 0$, $w_* \subset y$, and $w \in y$. Hence also $y \supset \bigcap_{x \in R} \overline{[(x\omega)}$.

Theorem 6.4. $\mathscr{R} \subset \gamma_1$.

Proof. It is sufficient to show that any $v \in \mathscr{L}_+M$ is in γ_1, in view of Theorem 5.1. By Lemma 6.1, it is sufficient to prove this for vp. By Lemma 6.2, we reduce this to showing that $x \in \mathscr{L}_*M$ implies $x \in \gamma_1$. This last result follows from Lemma 6.3.

For card $A \geq 3$, the containment in Theorem 6.4 is proper. For, let $L = \{a, b\}^*$. Clearly L is not $\oplus$finite. For $A = \{a, b, c\}$, $\bar{L} = A^*cA^*$ is not $\oplus$finite either. Hence $L \notin \mathscr{R}$, but $L = \overline{[c}$ is in γ_1. Note, however, that L *is* a run language for $A = \{a, b\}$, since $L = \bar{\emptyset}$ is $\oplus$cofinite. Thus the run property is alphabet dependent.

7. The two-letter case

We will now show that for card $A \leqslant 2$, $\mathscr{R} = \gamma_1$. Consider first $A = \{a\}$. We have for $n \geqslant 1$,

$$[a^n = \overline{1 \cup a \cup \ldots \cup a^{n-1}}.$$

Hence $[\mathscr{W} \subset \mathscr{R}$ and also $\gamma_1 \subset \mathscr{R}$. To prove the two-letter case, we will use the following lemma which holds for any alphabet. For any $a_i \in A$ let $A_i = A - a_i$.

Lemma 7.1. (a) $\overline{[a_i} = A_i^*$.
(b) $\overline{[a_i^{n+1}} = (A_i^* \cup A_i^* a_i A_i^*)^n$, *for* $n \geqslant 1$.
(c) *For all* $a_1^{n_1} \ldots a_p^{n_p}$ *in run form*:

$$\overline{[(a_1^{n_1} \ldots a_p^{n_p})} = \overline{[a_1^{n_1}} \cdot \ldots \cdot \overline{[a_p^{n_p}}.$$

Proof. (a) and (b) are easily verified. For (c) note that $w \in [(xy)$ iff there exist w_1, $w_2 \in A^*$ such that $w = w_1 w_2$, $w_1 \in [x$ and $w_2 \in [y$. To prove the lemma it suffices to show that for all $a, b \in A$, $a \neq b$, $i, j > 0$, $x \in A^*$,

$$\overline{[(a^i b^j x)} = \overline{[a^i} \cdot \overline{[b^j x}. \qquad (5)$$

If $w \in \overline{[a^i} \cdot \overline{[(b^j x)}$, then $w = w_1 w_2$, $w_1 \notin [a^i$, $w_2 \notin [(b^j x)$. For any decomposition $w = w_1' w_2'$ with $|w_1'| < |w_1|$, we have $w_1' \notin [a^i$. Similarly if $|w_1'| > |w_1|$, then $w_2' \notin [(b^j x)$. Hence there is no decomposition $w = w_1' w_2'$ such that $w_1' \in [a^i$ and $w_2' \in [(b^j x)$. Thus $w \notin [a^i[(b^j x) = [(a^i b^j x)$ and the containment $\supset$ for (5) follows.

Conversely, suppose $w \in \overline{[(a^i b^j x)}$. Suppose also $w \in [a^i$ and let $w = w_1 a w_2$, where $w_1 a$ is the shortest prefix of w such that $w_1 a \in [a^i$, implying $w_1 \in \overline{[a^i}$. If $w_2 \in [(b^j x)$, then $w = (w_1 a) w_2 \in [(a^i b^j x)$, a contradiction. Hence $w_2 \in \overline{[(b^j x)}$, $a w_2 \in \overline{[(b^j x)}$ and $w = w_1(a w_2) \in \overline{[a^i} \cdot \overline{[(b^j x)}$. Finally if $w \notin [a^i$, then $w = w \cdot 1 \in \overline{[a_i} \cdot \overline{[(b^j x)}$. Hence the containment $\subset$ for (5) also holds.

Theorem 7.2. *For* card $A \leqslant 2$, $\mathscr{R} = \gamma_1$.

Proof. The case card $A = 1$ was already shown. For $A = \{a, b\}$, $\overline{[a} = b^*$ and $\overline{[a^{n+1}} = (b^* \cup b^* a b^*)^n$ for $n \geqslant 1$, by Lemma 7.1 (a) and (b). By Lemma 7.1 (c), $\overline{[w}$ is a finite union of products of elements from $\mathscr{L}_* \cup \mathscr{L}$. Since $a^* = 1 \cup a^+$, every such product is in $\mathscr{L}_\oplus MB$. Hence $\gamma_1 \subset \mathscr{R}$. By Theorem 6.4, $\mathscr{R} \subset \gamma_1$.

8. Automata accepting run languages

Let $\mathscr{A} = \langle S, s_0, F, f \rangle$ be a reduced finite automaton over alphabet A, with S as the set of states, $s_0 \in S$ the initial state, $F \subset S$ the set accepting states, and $f : S \times A \to S$ the transition function.

One verifies that if L satisfies any one of the following 3 properties then it cannot be $\oplus$finite:

(a) $L \supset xy^*z$, $|y| \geq 2$, $y \in [a \cap [b, a, b \in A$;

(b) $L \supset x(a^n)^*z$, $n \geq 2, a \in A$;

(c) $L \supset x(a, b)^*z, a, b \in A$.

If $\mathscr{A}$ accepts a $\oplus$finite language, then it must have a "dead" state $s_\emptyset$ such that $f(s_\emptyset, a) = s_\emptyset$, for all $a \in A$. Let $\mathscr{A}'$ be the incompletely specified automaton obtained from $\mathscr{A}$ by removing $s_\emptyset$ and all transitions into $s_\emptyset$. Then $\mathscr{A}'$ must satisfy, for all a, $b \in A$, $x \in A^*$, $s \in S - s_\emptyset$,

(1) $f(s, x) = s$ implies $|x| = 1$,

(2) $f(s, a) = f(s, b) = s$ implies $a = b$.

Otherwise one of the conditions (a)–(c) will be true for L. Thus (1) and (2) are necessary in order that $\mathscr{A}$ accept a $\oplus$finite language. One easily verifies that these conditions are also sufficient. Hence one can easily recognize whether a given reduced automaton $\mathscr{A}$ accepts a $\oplus$finite language. It now follows that one can easily test whether $\mathscr{A}$ accepts a run language.

On the other hand, run languages cannot be characterized by their syntactic monoids or semigroups since for $A = \{a, b, c\}$, $L = \{a, b\} \notin \mathscr{R}$, $L' = a^* \in \mathscr{R}$ and the monoids of L and L' are isomorphic, as are their semigroups.

Acknowledgment

The author wishes to thank S. Eilenberg and M.P. Schützenberger for their constructive comments on this work.

References

[1] R.S. Cohen and J.A. Brzozowski, Dot depth of star-free events, J. Comput. System Sci. (February 1971) 1–16.

[2] I. Simon, Hierarchies of events with dot-depth one, Ph. D. Thesis, Computer Science Department, University of Waterloo, Waterloo, Ont. (August 1972).

[3] I. Simon, Piecewise testable events, 2nd GI-Professional Conference on Automata Theory and Formal Languages, Kaiserslautern (May 1975), Lecture Notes in Computer Science (Springer, Berlin) to appear.