RAIRO. INFORMATIQUE THÉORIQUE

ROBERT KNAST

A semigroup characterization of dot-depth one languages

RAIRO. Informatique théorique, tome 17, nº 4 (1983), p. 321-330

<http://www.numdam.org/item?id=ITA_1983__17_4_321_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 17, n° 4, 1983, p. 321 à 330)

A SEMIGROUP CHARACTERIZATION OF DOT-DEPTH ONE LANGUAGES (*)

by Robert KNAST ([1])

Communicated by J.-F. PERROT

Abstract. — *It is shown that one can decide whether a langage has dot-depth one in the dot-depth hierarchy introduced by Brzozowski. The decision procedure is based on an algebraic characterization of the syntactic semigroup of a langage of dot-depth* 0 *or* 1.

Résumé. — *On démontre que l'on peut décider si un langage est de hauteur* 1 *dans la hiérarchie de concaténation introduite par Brzozowski. L'algorithme de décision est basé sur une condition algèbrique qui caractérise les semigroupes syntactiques des langages de hauteur inférieure ou égale à* 1.

1. INTRODUCTION

Let A be a non-empty finite set, called alphabet. A^+ (respectively A^*) is the free semigroup (respectively free monoid) generated by A. Elements of A^* are called words. The empty word in A^* is denoted by λ (the identity of A^*). The concatenation of two words x, y is denoted by xy. The length of a word x is denoted by $|x|$.

Any subset of A^* is called a language. If L_1 and L_2 are languages, then $L_1 \cup L_2$ is their union, $L_1 \cap L_2$ is their intersection, and $\overline{L}_1 = A^* - L_1$ is the complement of L_1 with respect to A^*. Also $L_1 L_2 = \{ w \in A^* \mid w = xy,\ x \in L_1,\ y \in L_2 \}$ is the concatenation of L_1 and L_2.

Let $\sim$ be an equivalence relation on A^*. For $x \in A^*$ we denote by $[x]_\sim$ the equivalence class of $\sim$ containing x. An equivalence relation $\sim$ on A^* is a congruence iff for all $x, y \in A^*$, $x \sim y$ implies $uxv \sim uyv$ for any $u, v \in A^*$.

The syntactic congruence of a language L is defined as follows: for $x, y \in A^*$, $x \equiv_L y$ iff for all $u, v \in A^*$ ($uxv \in L$ iff $uyv \in L$). The syntactic semigroup of L is the quotient semigroup $A^+ / \equiv_L$.

Let η be any family of languages. Then ηM (ηB) will denote the smallest family of languages containing η and closed under concatenation (finite union and complementation respectively).

(*) Received February 1981, revised May 1983.
([1]) Institute of Mathematics, Polish Academy of Sciences, 61-725 Poznan, Poland.

Let $\varepsilon = \{\{\lambda\}, \{a\}; a \in A\}$ be the family of elementary languages. Then define:

$$\mathscr{B}_0 = \varepsilon B,$$
$$\mathscr{B}_k = \mathscr{B}_{k-1} MB \quad \text{for} \quad k \geqq 1.$$

This sequence $(\mathscr{B}_0, \mathscr{B}_1, \ldots, \mathscr{B}_k, \ldots)$ is called the dot-depth hierarchy. A langage L is of dot-depth at most k if $L \in \mathscr{B}_k$.

The dot-depth hierarchy was introduced in [3]. It was proved in [2] that it is infinite if the alphabet has two or more letters. In [4] it was shown that $(\mathscr{B}_0, \mathscr{B}_1, \ldots)$ forms a hierarchy of $+$ $-$varieties of languages. Therefore, in the rest of the paper we consider languages as subsets of A^+. For an excellent and general presentation of problems related to this paper the reader is referred to Brzozowski's survey paper [1] or the above mentioned monograph of Eilenberg [4].

In [6] Simon conjectured that a language L is in $\mathscr{B}_1$ iff its syntactic semigroup S_L is finite and there exists an integer $n > 0$ such that for each idempotent e in S_L, and any elements $a, b \in S_L$:

$$(eaeb)^n eae = (eaeb)^n e = ebe(aebe)^n.$$

Simon also proved that $L \in \mathscr{B}_1$ implies this condition. By an example we show that this conjecture fails. We present a necessary and sufficient condition for a syntactic semigroup to be the syntactic semigroup of a language of dot-depth at most one. The main result is as follows: Let L be a language and let S_L be its syntactic semigroup. Then $L \in \mathscr{B}_1$ iff S_L is finite and there exists an integer $n > 0$ such that for all idempotents e_1, e_2 in S_L and any elements $a, b, c, d \in S_L$:

$$(e_1 ae_2 b)^n e_1 ae_2 de_1 (ce_2 de_1)^n = (e_1 ae_2 b)^n e_1 (ce_2 de_1)^n.$$

We will refer to this as the "dot-depth one" condition. This semigroup characterization gives a decision procedure for testing whether or not a regular language is in $\mathscr{B}_1$.

In the proof of this characterization we use a theorem on graphs from [5].

We will say that a language $L \subset A^+$ is a $\sim$ language, if L is a union of congruence classes of $\sim$. Let L be a language and let S_L be its syntactic semigroup. The class $[x] \equiv_L$, as an element of S_L, will be also denoted by $\underline{x}$, where $x \in A^+$. Then $x \equiv_L y$ iff $\underline{x} = \underline{y}$ in S_L.

2. BASIC CONGRUENCE$_m \sim_k$ [6]

Let k, m be integers, $k \geqq 1$, $m \geqq 0$. Let $v = (w_1, w_2, \ldots, w_m)$ be an m-tuple of words w_i of length k, i. e. $|w_i| = k$, $w_i \in A^*$ $i = 1, 2, \ldots, m$. We say that v occurs in

x, $x \in A^*$ (we write $\nu \underline{\in} x$), if $x = u_i w_i v_i$, for some $u_i, v_i \in A^*$ $(i = 1, 2, \ldots, m)$ such that $|u_j| < |u_{j+1}|$, $j = 1, 2, \ldots, m-1$.

Let us set:

$$\tau_{m,k}(x) = \{\nu \mid \nu \in (A^k)^m \text{ and } \nu \underline{\in} x\}.$$

By convention $\tau_{0,k} x = \emptyset$.

For $x \in A^*$ and $n \geqq 0$ define $f_n(x)$ as follows: if $|x| \leqq n$, then $f_n(x) = x$; otherwise $f_n(x)$ is the prefix of x of length n. Similarly, $t_n(x) = x$ if $|x| \leqq n$, and $t_n(x)$ is the suffix of length n of x otherwise.

Now, for $x, y \in A^*$ and $k \geqq 0$, $m \geqq 0$ we define:

$$\begin{aligned} x\,{}_m\!\sim_k y \text{ iff } & x = y \text{ if } |x| \leqq m + k - 1 \\ & \text{or } f_k(x) = f_k(y),\ t_k(x) = t_k(y) \\ & \text{and } \tau_{m,k+1}(x) = \tau_{m,k+1}(y) \text{ otherwise.} \end{aligned}$$

In the case $k = 0$ we write τ_m instead $\tau_{m,0}$ and ${}_m\!\sim$ instead ${}_m\!\sim_0$. If $m = 1$, we also write τ instead τ_1.

PROPOSITION 1: *(a) ${}_m\!\sim_k$ is a congruence of finite index on A^*; (b) $x\,{}_m\!\sim_k y$ implies $x\,{}_{m-1}\!\sim_k y$, for $m \geqq 1$ and all $x, y \in A^*$; (c) $w(xw)^m\,{}_m\!\sim_k w(xw)^{m+1}$, for $w, x \in A^*$ and $|w| = k$; (d) $(w_1 x w_2 y)^m \quad w_1 x w_2 v w_1 (u w_2 v w_1)^m\,{}_m\!\sim_k (w_1 x w_2 y)^m$ $w_1 (u w_2 v w_1)^m$, for $w_1, w_2, x, y, u, v \in A^*$ and $|w_1| = |w_2| = k$.*

Proof: The verification of (a), (b) and (c) is straightforward.
(d) By (b):

$$\tau_{m,k+1}(x) = \tau_{m,k+1}(y)$$

implies:

$$\tau_{j,k+1}(x) = \tau_{j,k+1}(y),$$

for all $x, y \in A^*$ and $j \in \{0, 1, \ldots, m\}$. If

$$\nu_1 = (w_1, \ldots, w_i) \in (A^{k+1})^i$$

and

$$\nu_2 = (v_1, \ldots, v_j) \in (A^{k+1})^j,$$

we denote by (ν_1, ν_2) the $i+j$-tuple $(w_1, \ldots, w_i, v_1, \ldots, v_j) \in (A^{k+1})^{i+j}$. Evidently:

$$\tau_{m,\,k+1}((w_1\,x\,w_2\,y)^m\,w_1)\subseteq\tau_{m,\,k+1}((w_1\,x\,w_2\,y)^m\,w_1\,xw_2)$$
$$\subseteq\tau_{m,\,k+1}((w_1\,xw_2\,y)^{m+1}\,w_1).$$

Using (c), we have:

$$\tau_{m,\,k+1}((w_1\,xw_2\,y)^m\,w_1\,xw_2)=\tau_{m,\,k+1}((w_1\,xw_2\,y)^m\,w_1).$$

Similarly:

$$\tau_{m,\,k+1}(w_2\,vw_1\,(uw_2\,vw_1)^m)=\tau_{m,\,k+1}(w_1\,(uw_2\,vw_1)^m).$$

Since $|w_1|=|w_2|=k$, by the above conclusions from (b) and (c):

$$\tau_{m,\,k+1}((w_1\,xw_2\,y)^m\,w_1\,xw_2\,vw_1\,(uw_2\,vw_1)^m)=\bigcup_{\substack{i+j=m\\ m\geqq i,\,j\geqq 0}}\{(\nu_1,\,\nu_2)\,|\,\nu_1$$
$$\in\tau_{i,\,k+1}((w_1\,x\,w_2\,y)^m\,w_1\,xw_2),\ \nu_2\in\tau_{j,\,k+1}(w_2\,vw_1\,(uw_2\,vw_1)^m)\}$$
$$=\bigcup_{\substack{i+j=m\\ m\geqq i,\,j\geqq 0}}\{(\nu_1,\,\nu_2)\,|\,\nu_1\in\tau_{i,\,k+1}((w_1 x\,w_2\,y)^m\,w_1),\ \nu_2\in\tau_{j,\,k+1}(w_1\,(uw_2\,vw_1)^m)\}$$
$$=\tau_{m,\,k+1}((w_1 x\,w_2\,y)^m\,w_1\,(uw_2\,vw_1)^m).\quad\square$$

THEOREM 2 (Simon [6]): *A language L is of dot-depth at most one, $L\in\mathscr{B}_1$, iff L is a ${}_m\sim_k$ language for some $m,\,k\geqq 0$.*

3. GRAPHS AND THE INDUCED SYNTACTIC GRAPH CONGRUENCE

First we briefly recall Eilenberg's terminology for graphs [4]. A directed graph G consists of two sets, an alphabet A and a set of vertices V, along with two functions: $\alpha,\,\omega : A\to V$. Elements of A are also called edges in this case.

Two letters (or edges) $a,\,b\in A$ are called consecutive if $a\,\omega=b\,\alpha$. Let $D\subset A^2$ be the set of all words ab such that a and b are non-consecutive. Then the set of all paths of G is:

$$P=A^+-A^*\,DA^*.$$

Functions α, ω can be extended to $\alpha,\,\omega : P\to V$ in the following way: if $p=a_1\,a_2\ldots a_n\in P,\,a_1,a_2,\ldots,a_n\in A$, then $p\,\alpha=a_1\,\alpha, p\,\omega=a_n\,\omega$. For each vertex v we adjoint to P a trivial path 1_v where $1_v\,\alpha=1_v\,\omega=v$. If $p=a_1\,a_2\ldots a_n\in P$, then the length of p, $|p|=n$.

A path p is called a loop if $p\,\alpha=p\,\omega$. We say that two paths p_1 and p_2 are consecutive if $p_1\,\omega=p_2\,\alpha$. In this case the concatenation $p_1\,p_2$ is again a path. Two paths p_1 and p_2 are coterminal if $p_1\,\alpha=p_2\,\alpha$ and $p_1\,\omega=p_2\,\omega$.

An equivalence relation $\sim$ on P is called a graph congruence if it satisfies the following conditions:

(i) if $p_1 \sim p_2$, then p_1 and p_2 are coterminal;

(ii) if $p_1 \sim p_2$ and $p_3 \sim p_4$ and p_1, p_3 are consecutive, then $p_1 p_3 \sim p_2 p_4$.

For trivial paths, by convention we set $\tau_m(1_v) = \emptyset$. Thus the relation ${}_m\sim$ (${}_m\sim_1$) is also defined on P. In [5] the following theorem is proved:

THEOREM 3: *Let $\sim$ be a graph congruence of finite index on P satisfying the condition:*

$$\text{(A)} \qquad (p_1 p_2)^n p_1 p_4 (p_3 p_4)^n \sim (p_1 p_2)^n (p_3 p_4)^n,$$

for some $n \geqq 1$ and $p_1, p_2, p_3, p_4 \in P$. (Note that $p_1 p_2$ and $p_3 p_4$ must be loops about the same vertex).

Then there exists an integer $m \geqq 1$ such that for any two coterminal paths x and y, $x\,{}_m\!\sim y$ implies $x \sim y$.

We will use this theorem in proving the semigroup characterization of languages of dot-depth at most one ($\mathscr{B}_1$).

Let A be a finite alphabet. Define a graph $G_k = (V, E, \alpha, \omega)$ for $k \geqq 0$ as follows:

$$V = \{ w \mid w \in A^* \text{ and } |w| = k \} \text{ is the set of vertices,}$$

$$E = \{ (w_1, \sigma, w_2) \mid \sigma \in A,\ w_1,\ w_2 \in V \text{ and } t_k(w_1 \sigma) = w_2 \},$$

is the set of edges (letters)

$$\alpha, \omega : E \to V, (w_1, \sigma, w_2)\alpha = w_1, (w_1, \sigma, w_2)\omega = w_2.$$

Let P be the set of all paths in G_k, including the empty path over each vertex from V. Now, let us define the mapping:

$$^{-} : A^k A^* \to P,$$

recursively as follows:

$$\bar{x} = 1_x \quad \text{if} \quad x \in A^k,$$

$$\overline{x\sigma} = \bar{x}(t_k(x), \sigma, t_k(x\sigma)).$$

For $k = 0$, by convention $A^0 = \{\lambda\}$. One can verify that the mapping $^{-}$ is bijective. It follows from the definition that $|x| = k + h$, $h \geqq 0$ iff $|\bar{x}| = h$.

If ρ is a congruence relation on A^*, then by $\bar{\rho}$ we will denote the induced congruence on P defined in the following way: for $\bar{x}, \bar{y} \in P$, $x, y \in A^k A^*$, $\bar{x} \bar{\rho} \bar{y}$ if $\bar{x}, \bar{y}$ are coterminal paths and $x \rho y$. One can verify that $\bar{\rho}$ is a graph congruence on P.

PROPOSITION 4: *Let G_k be a graph for $k \geqq 1$ and P be the set of all paths of G_k. Let $x \in A^k A^*$. If $x = x_1 x_2$, then $\bar{x} = \bar{x}_1 \overline{t_k(x_1) x_2}$, for $|x_1| \geqq k$.*

Proof: If $|x| = k$, then the only decomposition possible is $x = x\lambda$. But $\bar{x} = 1_x = 1_x 1_x = \bar{x}\bar{\lambda} = \bar{x}\,\overline{t_k(x)\lambda}$. Induction assumption: the proposition is true for x such that $|x| = k+h$, $h \geqq 0$. Suppose $x = x_1 x_2 \sigma$, where $|x_1 x_2| = k+h$ and $|x_1| \geqq k$. By definition:

$$\bar{x} = \overline{x_1 x_2}\,(t_k(x_1 x_2),\ \sigma,\ t_k(x_1 x_2 \sigma)).$$

By the induction assumption:

$$\overline{x_1 x_2} = x_1\, \overline{t_k(x_1) x_2}.$$

Hence:

$$\bar{x} = \bar{x}_1\, \overline{t_k(x_1) x_2}\,(t_k(x_1 x_2),\ \sigma,\ t_k(x_1 x_2 \sigma)).$$

Again by definition:

$$\overline{t_k(x_1) x_2 \sigma} = \overline{t_k(x_1) x_2}\,(t_k(t_k(x_1) x_2),\ \sigma,\ t_k(t_k(x_1) x_2 \sigma)).$$

Thus $\bar{x} = \bar{x}_1\, \overline{t_k(x_1) x_2 \sigma}$ because $t_k(x_1 x_2) = t_k(t_k(x_1) x_2)$. Thus the induction step holds. □

LEMMA 5: *Let $x \in A^k A^*$ and $\bar{x} = a_1 a_2 \ldots a_n$, $a_j \in E$, $j = 1, 2, \ldots, n$. Then for $i \in \{1, 2, \ldots, n\}$ $a_i = (w, \sigma, t_k(w\sigma))$ iff $x = x_1 w \sigma x_2$ for some $x_1, x_2 \in A^*$ and $|x_1 w \sigma| = k+i$.*

Proof: Suppose $f_{k+i}(x) = x_1 w \sigma$. By Proposition 3 $\bar{x} = \overline{x_1 \dot{w}}\, \overline{w \sigma x_2}$. By the definition of $\bar{\ }$ it follows from Proposition 3 that $\overline{\dot{w} \sigma x_2} = (w, \sigma, t_k(\dot{w}\sigma))\, \overline{t_k(\dot{w}\sigma) x_2}$. Also by the definition of $\bar{\ }$ $|\overline{x_1 \dot{w}}| = i-1$, because $|x_1 w| = k+i-1$. Hence $a_i = (\dot{w}, \sigma, t_k(\dot{w}\sigma))$.

The converse follows in the similar way. □

PROPOSITION 6: *For any $x, y \in A^k A^*$:*

$$x\ {}_m\!\sim_k y \quad \textit{implies} \quad \bar{x}\ {}_m\!\sim \bar{y},$$

where $\bar{x}, \bar{y} \in P$ of G_k.

Proof: If $|x| \leqq m+k$, then $x = y$ and consequently, $\bar{x}\ {}_m\!\sim \bar{y}$. Otherwise, let $\tau_{m,k+1}(x) = \tau_{m,k+1}(y) \neq \emptyset$. It follows from Lemma 5 that $((\dot{w}_1, \sigma_1, v_1), \ldots, (\dot{w}_m, \sigma_m, v_m)) \in \tau_m(\bar{x})$ implies $(\dot{w}_1 \sigma_1, \ldots, \dot{w}_m \sigma_m) \in \tau_{m,k+1}(x) = \tau_{m,k+1}(y)$. Hence, again by Lemma 4 $((\bar{w}_1, \sigma_1, v_1), \ldots, (\bar{w}_m, \sigma_m, v_m)) \in \tau_m(\bar{y})$. Thus, $\tau_m(\bar{x}) \subseteq \tau_m(\bar{y})$. By symmetry, $\tau_m(\bar{y}) \subseteq \tau_m(\bar{x})$.

Since $f_k(x)=f_k(y)$ and $t_k(x)=t_k(y)$, then $\bar{x}$ and $\bar{y}$ are coterminal.

Consequently, $\bar{x}_m \sim \bar{y}$. □

PROPOSITION 7: *Let $L \subseteq A^+$ and let S_L be the finite syntactic semigroup of L, satisfying the condition: there exists m, $m>0$, such that for all idempotents e_1, e_2 in S_L and any elements $a, b, c, d \in S_L$:*

$$(e_1 a e_2 b)^m e_1 a e_2 d e_1 (c e_2 d e_1)^m = (e_1 a e_2 b)^m e_1 (c e_2 d e_1)^m.$$

Then the congruence $\overline{\equiv_L}$ on P of G_K for $k = \text{card } S_L + 1$, induced by the syntactic congruence $\equiv_L$ satisfies condition (A) *of Theorem* 2 *and is of finite index on P.*

Proof: Since G_k is finite and $\equiv_L$ *is of finite index on A^+*, then $\overline{\equiv_L}$ is of finite index on P.

We have to show that there is an integer n, $n>0$ *such that:*

(A) $$(p_1 p_2)^n p_1 p_4 (p_3 p_4)^n \overline{\equiv_L} (p_1 p_2)^n (p_3 p_4)^n,$$

for $p_1, p_2, p_3, p_4 \in P$.

Since $p_1 p_2$ and $p_3 p_4$ are loops about the same vertex and since paths p_1 and p_4 are consecutive by (A), then $p_1 \alpha = p_2 \omega = p_3 \alpha = p_4 \omega = w$, and $p_1 \omega = p_2 \alpha = p_3 \omega = p_4 \alpha = v$ for some $w, v \in A^k$. Therefore we may assume that $p_1 = \overline{wu_1}$, $p_2 = \overline{vu_2}$, $p_3 = \overline{wu_3}$, $p_4 = \overline{vu_4}$ for some $u_1, u_2, u_3, u_4 \in A^*$ such that $t_k(wu_1) = t_k(wu_3) = v$, $t_k(vu_2) = t_k(vu_4) = w$. Consequently:

$$(p_1 p_2)^n p_1 p_4 (p_3 p_4)^n = \overline{w (u_1 u_2)^n u_1 u_4 (u_3 u_4)^n}.$$

Similarly:

$$(p_1 p_2)^n (p_3 p_4)^n = \overline{w (u_1 u_2)^n (u_3 u_4)^n}.$$

By the definition of $\overline{\equiv_L}$ it is sufficient to show that there exists n, $n>0$, such that:

$$w (u_1 u_2)^n u_1 u_4 (u_3 u_4)^n \equiv_L w (u_1 u_2)^n (u_3 u_4)^n,$$

i. e.:

(1) $$\underline{w} (\underline{u}_1 \underline{u}_2)^n \underline{u}_1 \underline{u}_4 (\underline{u}_3 \underline{u}_4)^n = \underline{w} (\underline{u}_1 \underline{u}_2)^n (\underline{u}_3 \underline{u}_4)^n.$$

Let $s \in S_L$. Since S_L is finite, then s^r is an idempotent for some $r \geqq 1$ ([4], Proposition 4.2, p. 68). Now, since S_L satisfies the dot-depth one condition, there is $m \geqq 1$ such that:

$$s^r (s s^r)^m = s^r (s s^r)^{m+1}$$

i. e. $s^r s^m = s^r s^m s$. It follows that there exists an integer q such that for any $s \in S_L$ $s^q = s^{q+1}$ i. e. S_L is aperiodic.

We claim that (1) holds for $n > m, q$. First we will show that if $|u_1 u_2| > 0$ ($|u_3 u_4| > 0$) then we may consider u_1, u_2 (u_3, u_4 respectively) such that $|u_1|, |u_2| \geqq k$ ($|u_3|, |u_4| > k$ *respectively*). Since $n > q$, then by the aperiodicity of S_L:

$$\underline{w} (\underline{u_1} \underline{u_2})^n = \underline{w} (\underline{u_1} \underline{u_2})^{n(2k+1)}.$$

Let us define:

$$\tilde{u}_1 = (u_1 u_2)^k u_1, \; \tilde{u}_2 = u_2 (u_1 u_2)^k.$$

Evidently:

$$|\tilde{u}_1|, |\tilde{u}_2| \geqq k, \qquad t_k(w \tilde{u}_1) = v, \qquad t_k(v \tilde{u}_2) = w$$

and:

$$\underline{w} (\underline{u_1} \underline{u_2}) = w (\tilde{u}_1 \tilde{u}_2)^n.$$

Similarly, we may proceed for u_3 and u_4.

Now, we consider the full case if $|u_1 u_2|, |u_3 u_4| > 0$. The other cases if $|u_1 u_2| = 0$ or $|u_3 u_4| = 0$ follow in the same way. By the above, instead of proving (1) it is sufficient to show that:

$$\text{(2)} \qquad \underline{w} (\underline{u_1} \underline{v} \; \underline{u_2} \underline{w})^n \underline{u_1} \underline{v u_4} \underline{w} (\underline{u_3} \underline{v u_4} \underline{w})^n = \underline{w} (\underline{u_1} \underline{v u_2} \underline{w})^n (\underline{u_3} \underline{v u_4} \underline{w})^n,$$

holds.

Now, since $|w| = |v| = k > \text{card } S_L + 1$, then $w = w_1 w_2 w_3$ and $v = v_1 v_2 v_3$ for $w_1, w_3, v_1, v_3 \in A$, $w_2, v_2 \in A^+$ such that $\underline{w}_1 = \underline{w}_1 \underline{w}_2^i$, $\underline{v}_1 = \underline{v}_1 \underline{v}_2^i$ for any $i \geqq 0$. So as before, we can choose i such that $\underline{w}_2^i$ and $\underline{v}_2^i$ are idempotents in S_L. Thus (2) can be rewritten in a form:

$$\underline{w}_1 e_1 (a e_1 b e_1)^n a e_2 d e_1 (c e_2 d e_1)^n \underline{w}_3 = \underline{w}_1 e_1 (a e_2 b e_1)^n (c e_2 d e_1)^n \underline{w}_3,$$

where:

$$e_1 = \underline{w}_2^i, \qquad e_2 = \underline{v}_2^i, \qquad a = \underline{w}_3 \underline{u}_1 \underline{v}_1,$$

$$b = \underline{v}_3 \underline{u}_2 \underline{w}_1, \qquad c = \underline{w}_3 \underline{u}_3 \underline{v}_1$$

and $d=\underline{v}_3\underline{u}_4\underline{w}_1$. Thus by the dot-depth one condition, (2) holds. □

4. SEMIGROUP CHARACTERIZATION OF $\mathscr{B}_1$

Now we are in a position to prove our main result.

THEOREM 8: *Let L be a language, $L \subseteq A^+$ and let S_L be its syntactic semigroup. Then the following are equivalent:*

(i) $L \in \mathscr{B}_1$;

(ii) *L is a ${}_m\sim_k$ language for some m, $k \geqq 1$;*

(iii) *S_L is finite and there is an integer $n>0$ such that for all idempotents e_1, e_2 in S_L and any elements a, b, c, d in S_L:*

$$(e_1 ae_2 b)^n e_1 ae_2 de_1 (ce_2 de_1)^n = (e_1 ae_2 b)^n e_1 (ce_2 de_1)^n.$$

Proof: (i) ⇔ (ii) by Theorem 2;

(ii) ⇒ (iii) : by (*a*) of Proposition 1 S_L is finite.

Now, let $e_1=\underline{z}_1, e_2=\underline{z}_2, a=\underline{x}, b=\underline{y}, c=\underline{u}, d=\underline{v}$ for some $z_1, z_2, x, y, u, v \in A^+$. Define $w_1=z_1^h$, $w_2=z_2^h$ for h such that $|w_1|, |w_2| \geqq k$. Consequently, $e_1=\underline{w}_1$, $e_2=\underline{w}_2$. By (*d*) of Proposition 1 for ${}_m\sim_k$:

$$(\underline{w}_1\,\underline{xw_2}\,\underline{y})^m\,\underline{w}_1\,\underline{xw_2}\,\underline{vw_1}\,(\underline{uw_2}\,\underline{vw_1})^m = (\underline{w}_1\,\underline{xw_2}\,\underline{y})^m\,\underline{w}_1\,(\underline{uw_2}\,\underline{vw_1})^m.$$

Thus S_L satisfies the dot-depth one condition with $n=m$.

(iii) ⇒ (ii): suppose S_L satisfies the dot-depth one condition with n. Let $k=\text{card}\, S+1$. By Proposition 7 the induced syntactic congruence $\overline{\equiv}_L$ on P of G_k, satisfies the condition (A) of the theorem on graphs with some $n_1>n$, q, and is of finite index on P. Hence by Theorem 3 there exists m such that for any two coterminal paths x, y.

$$\bar{x}\,{}_m\sim\bar{y} \quad \text{implies} \quad \bar{x}\,\overline{\equiv}_L\,\bar{y}.$$

Now, consider $x, y \in A^k A^*$, and the congruence ${}_m\sim_k$. We have that $x\,{}_m\sim_k\,y$ implies $\bar{x}\,{}_m\sim\bar{y}$ and that $\bar{x}, \bar{y}$ are coterminal. Hence, $x\,{}_m\sim_k\,y$ implies $\bar{x}\,\overline{\equiv}_L\,\bar{y}$ and consequently, $x \equiv_L y$. If $|x| \leqq k$, then $x\,{}_m\sim_k\,y$ implies $x=y$ and consequently, $x \equiv_L y$. Thus L is a ${}_m\sim_k$ language. □

It is easy to see that if a syntactic semigroup satisfies the dot-depth one condition, then it also satisfies the condition: there exists an integer $n>0$ such that for any idempotent e in S_L and any elements a, b S_L:

$$(eaeb)^n eae = (eaeb)^n e = ebe(aebe)^n.$$

The following example shows that the converse is not true.

Let $A=\{0, 1, 2, 3\}$ and let $L=(01^{+} \cup 02^{+})^{*} 01^{+} 3(2^{+} 3 \cup 1^{+} 3)^{*}$. The syntactic semigroups S_L of L satisfies the above condition, but it fails the dot-depth one condition. By Theorem 8 $L \notin \mathscr{B}_1$. On the other hand one can verify that $L \notin \mathscr{B}_1$, apart from Theorem 8, using (*d*) of Proposition 1 and proving that for any m, k L cannot be a ${}_m\sim_k$ language.

REFERENCES

1. J. A. BRZOZOWSKI, *Hierarchies of a Periodic Languages*, R.A.I.R.O., Informatique Théorique, Vol. 10, No. 8, 1976, pp. 33-49.
2. J. A. BRZOZOWSKI a,d R. KNAST, *The Dot Depth Hierarchy of Star-Free Languages is Infinité*, J. Computer and System Sc., Vol. 16, No. 1, 1978, pp. 37-55.
3. R. S. COHEN and J. A. BRZOZOWSKI, *Dot-Depth of Star-Free Events*, J. Computer and System Sc., Vol. 5, 1971, pp. 1-16.
4. S. EILENBERG, *Automata, Languages and Machines*, Vol. B, Academic Press, New York, 1976.
5. R. KNAST, *Some Theorems on Graph Congruences*, R.A.I.R.O., Informatique Théorique, Vol. 17, No. 4, pp. 331-342.
6. I. SIMON, *Hierarchies of Events with Dot-Depth One*, Dissertation, University of Waterloo, Canada, 1972.