RAIRO. INFORMATIQUE THÉORIQUE

H. J. SHYR
G. THIERRIN

Codes, languages and MOL schemes

RAIRO. Informatique théorique, tome 11, nº 4 (1977), p. 293-301
<http://www.numdam.org/item?id=ITA_1977__11_4_293_0>

R.A.I.R.O. Informatique théorique/Theoretical Computer Science
(vol. 11, n° 4, 1977, p. 293 à 301)

CODES, LANGUAGES AND MOL SCHEMES (*) (1)

H. J. SHYR and G. THIERRIN (1)

Communicated by J.-F. PERROT

Abstract. — *The aim of this paper is to introduce and study a new class of DOL schemes, called MOL schemes. These are characterized by means of the OL languages they generate and by their preservation properties. Several special cases are investigated.*

1. INTRODUCTION.

Let X be an *alphabet* (a non-empty *finite* set) and let X^* be the *free monoid* generated by X. Let $X^+ = X^* - \{ 1 \}$, where 1 is the empty word and let $lg(w)$ denote the *length* of the word $w \in X^*$. Any subset of X^* is called a *language*. For any languages $A, B \subseteq X^*$, let $AB = \{ xy \mid x \in A, y \in B \}$, $A^* = \bigcup_{i=0}^{\infty} A^i$ and $A^+ = \bigcup_{i=1}^{\infty} A^i$ (1-free iteration).

A OL scheme (*see* [1]) is an ordered pair (X, P), where X is an alphabet and P (the set of productions) is a finite non-empty subset of $X \times X^*$ such that for any $a \in X$, there exists at least one $x \in X^*$ such that $(a, x) \in P$. Sometimes the notation $a \to x \in P$ will be used instead of $(a, x) \in P$. A OL scheme is *deterministic* if for every $a \in X$, the element $x \in X^*$ such that $a \to x \in P$ is unique and it is *propagating* if for every $a \to x \in P$, $x \neq 1$. The words DOL and PDOL will be used to represent the deterministic OL schemes and the propagating deterministic OL schemes respectively. If (X, P) is a OL scheme and if $x = a_1 a_2 \ldots a_m$, $m \geqq 0$, $a_i \in X$, $i = 1, 2, \ldots, m$ and $y \in X^*$, then x is said to *directly generate* or *derive* y in (X, P), denoted by $x \Rightarrow y$, if and only if there exist $y_1, y_2, \ldots, y_m$ such that $\{ a_i \to y_i \mid i = 1, 2, \ldots, m \}$ and $y = y_1 y_2 \ldots y_m$. By this definition 1 directly derives y if and only if $y = 1$. The transitive and reflexive closure of the relation $\Rightarrow$ is denoted by $\overset{*}{\Rightarrow}$. When $x \overset{*}{\Rightarrow} y$ then x is said to generate y in (X, P). A OL *system* is a triple (X, P, w), where (X, P) is a OL scheme and $w \in X^*$, called the *axiom* of (X, P, w); (X, P) is called the *scheme* of (X, P, w). The language $L(X, P, w) = \{ y \in X^* \mid w \overset{*}{\Rightarrow} y \}$ is called the OL *language* generated by

(*) Received March 1977.
(1) This research has been supported by Grant A 7877 of the National Research Council of Canada.
(2) Department of Mathematics, University of Western Ontario, London, Ontario, Canada.

(X, P, w); the notation $L(w)$ will also be used when there is no ambiguity concerning the scheme (X, P). A language A is said to be a OL *language* if there exists a OL system (X, P, w) such that A is generated by (X, P, w).

A mapping h of X^* into X^* such that $h(xy) = h(x)h(y)$ for all $x, y \in X^*$ is said to be a *homomorphism* of X^* into X^* or an *endomorphism* of X^*. If furthermore h is injective, i. e., if $h(x) = h(y)$ implies $x = y$, then h is said to be a *monomorphism*. If (X, P) is a DOL scheme, then the mapping h defined by $h(a_i) = x_i$, where $a_i \rightarrow x_i \in P$ determines a homomorphism of X^* into X^*. Conversely, every homomorphism h of X^* into X^* defines a DOL scheme (X, P) where $a_i \rightarrow x_i \in P$ if and only if $h(a_i) = x_i$. *It follows that a* DOL *scheme can be defined either by* (X, P) *or* (X, h). In this paper we will use mainly the second definition. If (X, P, w) is a DOL system, then with the notation (X, h, w), the DOL language $L(w)$ generated by the system is given by $L(w) = \{ h^n(w) \mid n \geqq 0 \}$. If $\mathscr{F}$ is a family of languages over X and if $h(A) \in \mathscr{F}$ for every $A \in \mathscr{F}$, then we say that the DOL scheme (X, h) *preserves* $\mathscr{F}$ or that (X, h) is $\mathscr{F}$-preserving.

A MOL scheme (X, h) is a DOL scheme such that h is a *monomorphism*. It is immediate that a MOL scheme is always a PDOL scheme, but the converse is not true. Let us remark that a DOL scheme (X, h) such that $|X| = 1$ is always a MOL scheme, unless $h(X) = \{ 1 \}$. A DOL system (X, h, w) such that (X, h) is a MOL scheme is called a MOL *system* and the language $L(X, h, w)$ is called a MOL *language*. The purpose of this paper is to establish some properties of the MOL schemes. In section 2, we characterize MOL schemes by using the properties of the OL languages generated by their associated OL systems and we give a biological interpretation of some of these results. In section 3, the characterization of MOL schemes is done by considering some classes of languages which they preserve, and the last section is concerned mainly with the study of particular classes of MOL schemes.

2. MOL SCHEMES AND LANGUAGES

PROPOSITION 1: *Let* (X, h) *be a MOL scheme. If* $L(X, h, w_1) \cap L(X, h, w_2) \neq \emptyset$, *then either* $L(X, h, w_1) \subseteq L(X, h, w_2)$ *or* vice versa.

Proof: There exist $m, n \geqq 0$ such that $h^m(w_1) = h^n(w_2)$. If $m = n$, then $w_1 = w_2$ and $L(w_1) = L(w_2)$. Let $m < n$, $n = m+k$, $k \geqq 1$. Then $h^m(w_1) = h^{m+k}(w_2)$ and $h^m(w_1) = h^m(h^k(w_2))$. Hence $w_1 = h^k(w_2)$ and $L(w_1) \subseteq L(w_2)$. #

Let us remark that if (X, h) is a MOL scheme, then $L(X, h, w_1) \subseteq L(X, h, w_2)$ if and only if $w_1 = h^k(w_2)$ for some $k \geqq 0$.

PROPOSITION 2: *A PDOL scheme* (X, h) *is a MOL scheme if and only if* $L(X, h, w_1) \cap L(X, h, w_2) \neq \emptyset$, $w_1, w_2 \in X^+$, *implies either* $L(X, h, w_1) \subseteq L(X, h, w_2)$ *or* vice versa.

Proof: Necessity. This is Proposition 1. *Sufficiency.* Suppose h is not injective. Then there exist $v, w \in X^+$, $v \neq w$, such that $h(v) = h(w)$. It follows then that $L(v) \cap L(w) \neq \emptyset$ and hence either $L(v) \subseteq L(w)$ or $L(w) \subseteq L(v)$. Let us suppose $L(w) \subseteq L(v)$. Then $h^k(v) = w$ for some $k \geqq 1$ and $h^{k+1}(v) = h(w) = h(v)$. Let $X(w) = \{ x \mid x \in X, x \text{ is a subword of } w \}$. Since $h^{k+1}(v) = h(v)$, then $h^k(w) = w$ and $lg(h(x)) = 1$ for every $x \in X(w) \subseteq X$.

We claim that if $x, y \in X(w)$ and $h(x) = h(y)$, then $x = y$. Suppose on the contrary $x \neq y$ and $h(x) = h(y) = a \in X$. Then $h(xyx) = a^3 = h(yxy)$ and $L(xyx) \cap L(yxy) \neq \emptyset$. Hence $L(xyx) \subseteq L(yxy)$ or *vice versa.* Suppose the first case: then since $xyx \neq yxy$, we have $xyx = h^k(yxy)$ for some $k \geqq 1$. Therefore, $h^k(yxy) = u^3$ for some u and $xyx = u^3$, a contradiction. The second case is also impossible.

Now if $X(v) \subseteq X(w)$, then $h(v) = h(w)$ implies $v = w$, a contradiction. Hence $X(v) \nsubseteq X(w)$ and there exists $z \in X$ such that $z \in X(v)$, $z \notin X(w)$. Therefore $v = v_1 z v_2$ and $h(v) = h(v_1) h(z) h(v_2)$. Since $h(x) \in X$ for $x \in X(w)$ and since $h(v) = h(w)$, it follows then that w can be written in the form $w = y_1 y y_2$ where $h(y) = h(z) = d$. We have $h(zyz) = d^3 = h(yzy)$ and $L(zyz) \cap L(yzy) \neq \emptyset$. Hence $L(zyz) \subseteq L(yzy)$ or *vice versa.* Suppose the first case: then $zyz = h^k(yzy)$, for some $k \geqq 1$ and $zyz = t^3$ for some $t \in X^+$. Since $z \notin X(w)$, then $z \notin X(y)$ and the equality $zyz = t^3$ is impossible. By the same argument we can show that the second case is also impossible. #

The following biological interpretation can be given of the preceding proposition. Let us suppose that we have two organisms which are developing according to the same DOL scheme (X, h). Then the scheme (X, h) is a MOL scheme if and only if either of these two organisms have a completely different development or one of them can be considered as the descendant of the other.

Let (X, h) be a DOL scheme. Define on X^* the *relation* H by $x H y \Leftrightarrow h^m(x) = h^n(y)$ for some $m, n \geqq 0$. This relation is clearly an *equivalence* relation. Let us denote by $H(x)$ the class of x. Every OL language with scheme (X, h) is contained in a class of H.

If (X, h) is a MOL scheme, then $h^m(w_1) = h^n(w_2)$, $m \leqq n$, implies $w_1 = h^{n-m}(w_2)$. Therefore $v H w$ if and only if there exists $n \geqq 0$ such that either $v = h^n(w)$ or $w = h^n(v)$.

PROPOSITION 3: *A PDOL scheme (X, h) is a MOL scheme if and only if every class of H is a OL language.*

Proof: Necessity. Let A be a class of H. If $1 \in A$, then $A = \{ 1 \}$ and $A = L(X, h, 1)$. Let $1 \notin A$ and let B be the set of the words of minimal length in A. For every pair $w_1, w_2 \in B$, then either $w_1 = h^n(w_2)$ or $w_2 = h^n(w_1)$ for some $n \geqq 0$. Since B is finite, there exists $v \in B$ such that, for any $w \in B$, $w = h^n(v)$ for some $n \geqq 0$. Let $u \in A$, $u \notin B$; then $h^m(u) = h^n(v)$ for some $m, n \geqq 0$. Since (X, h) is propagating, then $m \leqq n$ and $u = h^{n-m}(v)$. Therefore $A = L(X, h, v)$.

Sufficiency. Suppose $L(X, h, w_1) \cap L(X, h, w_2) \neq \emptyset$ with $w_1, w_2 \in X^+$. Then, since each OL language with scheme (X, h) is contained in a class of H, $L(X, h, w_1)$ and $L(X, h, w_2)$ are contained in the same class A of H. But $A = L(X, h, v)$ for some $v \in X^+$. Hence $w_1 = h^m(v)$, $w_2 = h^n(v)$ for sorme $m, n \geqq 0$. Suppose $n = m+k$, $k \geqq 0$. Then $w_2 = h^{m+k}(v) = h^k(w_1)$. Therefore $L(X, h, w_2) \subseteq L(X, h, w_1)$ and (X, h) is a MOL scheme by Proposition 2. #

A OL language L with DOL scheme (X, h) is said to be *maximal* if the inclusion $L \subseteq L'$, where L' is a OL language with the same scheme (X, h), implies $L = L'$.

If (X, h) is a PDOL scheme, it is easy to see that every OL language with scheme (X, h) is *contained* in at *least* a *maximal* one. The following example shows that in general there can be *several distinct maximal* OL languages containing the same OL language.

Let $X = \{a, b\}$, $h(a) = ab$, $h(b) = ab$. Then $L(X, h, a)$ and $L(X, h, b)$ are distinct maximal OL languages containing the OL language $L(X, h, ab)$ with the PDOL scheme (X, h).

PROPOSITION 4: *A PDOL scheme (X, h) is a MOL scheme if and only if every OL language L with scheme (X, h) is contained in a unique maximal OL language with the same scheme.*

Proof: Necessity. Since (X, h) is a PDOL scheme, L is contained in at least one maximal OL language. Let M_1 and M_2 be two maximal OL languages containing L. Then $L \subseteq M_1 \cap M_2$, and by Proposition 2, $M_1 \subseteq M_2$ or $M_2 \subseteq M_1$. Hence $M_1 = M_2$.

Sufficiency. Let A be a class of H, $A \neq \{1\}$ and let $v \in A$. Then $L(X, h, v) \subseteq A$ and there is a unique maximal OL language M such that $L(X, h, v) \subseteq M$. It is immediate that $M \subseteq A$. Suppose $M \neq A$. Then there exists $w \in A$, $w \notin M$. Since $v\,H\,w$, then $h^m(v) = h^n(w) = u$ for some $n, m \geqq 0$. Therefore $u \in L(X, h, v) \subseteq M$ and $u \in L(X, h, w) \nsubseteq M$. Let M' be the unique maximal OL language containing $L(X, h, w)$. Since $u \in L(X, h, w)$, we have $L(X, h, u) \subseteq M$ and $L(X, h, u) \subseteq M'$, a contradiction. Hence $M = A$ and every class of H is a OL language. By Proposition 3, it follows then that (X, h) is a MOL scheme. #

3. CODES AND MOL SCHEMES

A non-empty language $A \subseteq X^+$ is said to be a *code* if $a_1 a_2 \ldots a_n = b_1 b_2 \ldots b_m$, $m \geqq 1, n \geqq 1$ and $a_i, b_j \in A$ implies $n = m$ and $a_i = b_i$, $i = 1, 2, \ldots, n$. A code A is called a *prefix code* if $A \cap AX^+ = \emptyset$. (*see* [4]). The relation ρ_c defined on X^* by $x\,\rho_c\,y$ if and only if $y = xu = ux$ for some $u \in X^*$ is a partial order and a non-empty language $A \subseteq X^+$ is called ρ_c*-independent* if for any x, $y \in A$, $x \, \rho_c \, y$ implies $x = y$ (*see* [8]).

PROPOSITION 5: *Every DOL scheme* (X, h) *that is code preserving is propagating.*

Proof: For any $a \in X$, $h(a) \neq 1$, because $\{ a \}$ is a code but $\{ 1 \}$ is not. #

PROPOSITION 6: *A DOL scheme* (X, h) *is a code preserving scheme if and only if* (X, h) *is a MOL scheme.*

Proof: Suppose first that (X, h) is code preserving. Then $h(X)$ is a code. Moreover, if a_i, $a_j \in X$, $a_i \neq a_j$, then $h(a_i) \neq h(a_j)$. Indeed, if $h(a_i) = h(a_j) = c$, then $A = \{ a_i, a_j^2 \}$ is a code but not $h(A) = \{ c, c^2 \}$, a contradiction. Now if h is not injective, then there exist $x \neq y$, x, $y \in X^+$ such that $h(x) = h(y)$. Let

$$x = x_1 \ldots x_m, \quad y = y_1 y_2 \ldots y_n, \quad m \geqq 1, \quad n \geqq 1 \quad \text{and} \quad x_i, y_j \in X;$$

then

$$h(x_1) \ldots h(x_m) = h(x) = h(y) = h(y_1) \ldots h(y_n).$$

Since $h(X)$ is a code, we have $m = n$ and $h(x_i) = h(y_i)$, $i = 1, 2, \ldots, n$. This implies that $x_i = y_i$, $i = 1, 2, \ldots, n$ and $x = y$ holds, a contradiction.

Suppose now that (X, h) is a MOL scheme and that (X, h) is not code preserving. Then there exists a code A over X such that $h(A)$ is not a code, and therefore $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m \in A$, $x_1 \neq y_1$ such that

$$h(x_1) \ldots h(x_n) = h(y_1) \ldots h(y_m).$$

This implies that

$$h(x_1 \ldots x_n) = h(y_1 \ldots y_m).$$

Since h is injective, we have $x_1 \ldots x_n = y_1 \ldots y_m$. It follows then that $x_1 = y_1$ and since A is, by assumption, a code, a contradiction. #

PROPOSITION 7: *A DOL scheme* (X, h) *is a MOL scheme if and only if* $h(X)$ *is a code and* $| h(X) | = | X |$.

Proof: Necessity. This follows immediately from Proposition 6.

Sufficiency. Suppose that h is not injective. Since $h(X)$ is a code, then $1 \notin h(X)$ and there exist x, $y \in X^+$, $x \neq y$, such that $h(x) = h(y)$. Let

$$x = x_1 x_2 \ldots x_m, \quad y = y_1 y_2 \ldots y_n, \quad x_i, y_j \in X, \quad m \geqq 1, \quad n \geqq 1.$$

Then

$$h(x_1 x_2 \ldots x_m) = h(y_1 y_2 \ldots y_n)$$

and

$$h(x_1) h(x_2) \ldots h(x_m) = h(y_1) h(y_2) \ldots h(y_n).$$

Since $h(X)$ is a code by assumption, we have $n = m$ and

$$h(x_i) = h(y_i) \quad \text{for all } i = 1, 2, \ldots, n.$$

Since $|h(X)| = |X|$, and X is finite we have $x_i = y_i$ for all $i = 1, 2, \ldots, n$. Thus $x = y$, a contradiction. Hence h is injective and (X, h) is a MOL scheme. #

If $A \subseteq X^+$, $A \neq \emptyset$, then A is ρ_c-independent if and only if every pair of two distinct elements from A form a code (*see* [8]). We note that for any $x, y \in X^+$, $\{x, y\}$ is a code if and only if $xy \neq yx$.

PROPOSITION 8: *A DOL scheme (X, h) is a MOL scheme if and only if (X, h) preserve the ρ_c-independent languages.*

Proof: Necessity. Let $A \subseteq X^+$ be a ρ_c-independent language. Suppose $h(A)$ is not ρ_c-independant. Then there exist $x, y \in A$, $x \neq y$ such that $\{h(x), h(y)\}$ is not a code.

This implies that $h(x)h(y) = h(y)h(x)$ and $h(xy) = h(yx)$ hods. Since h is injective by assumption, we have $xy = yx$. This contradicts the fact that A is a ρ_c-independent language.

Sufficiency. Suppose that h is not injective. Then there exist $x, y \in X^+$, $x \neq y$, such that

$$h(x) = h(y) = z, \qquad z \neq 1, \qquad \text{and} \qquad h(xy) = h(yx) = z^2.$$

Now if $xy = yx$, then

$$x = p^n, \qquad y = p^m \quad \text{for some } p \in X^+, \qquad \text{and} \qquad m \geqq 1, \quad n \geqq 1.$$

Since $[h(p)]^n = h(x) = h(y) = [h(p)]^m \neq 1$, we have $n = m$, a contradiction. On the other hand, if $xy \neq yx$, then $\{x, y\}$ is a code. The set $A = \{x, xy\}$ is then a code, but $h(x) = z$, $h(xy) = z^2$ and so $\{h(x), h(xy)\}$ is not a code, again a contradiction. #

4. SPECIAL CLASSES OF MOL SCHEMES

In this section, we consider MOL schemes which preserve special classes of languages.

Let us recall that a language A over X is said to be a *right power-bounded language* if there exists a positive integer n such that $yx^m \in A$, $x \neq 1$ implies that $m \leqq n$ (*see*, [9]).

PROPOSITION 9: *Let (X, h) be a DOL scheme such that $h(X) \neq \{1\}$. If (X, h) is a scheme which preserves the regular right power-bounded languages, then (X, h) is a MOL scheme.*

Proof: First we show that for any $a \in X$, $h(a) \neq 1$. Suppose $h(a) = 1$; then there exists $b \in X$ such that $h(b) \neq 1$, since $h(X) \neq \{1\}$ by assumption. The language $A = \{ b^n a \mid n \geqq 1 \}$ is a regular right power-bounded language, but $h(A) = \{ h(b))^n \mid n \geqq 1 \}$ is not a right power-bounded language, a contradiction. Thus $h(a) \neq 1$, for all $a \in X$.

Now suppose h is not injective. Then $h(x) = h(y)$, $x \neq y$, for some $x, y \in X^+$. We can choose x and y such that $x = az_1$, $y = bz_2$, $a \neq b$, $a, b \in X$ and $z_1, z_2 \in X^*$.

Then $h(x) = h(a)\, h(z_1) = h(y) = h(b)\, h(z_2)$. We may assume $lg(h(a)) \leqq lg(h(b))$. Let $h(a) = v$, $h(b) = w$. Then $w = vu$ for some $u = X^*$. The language $A = \{ b^n a \mid n \geqq 1 \}$ is a regular right power-bounded language, but $h(A) = \{ (vu)^n v \mid n \geqq 1 \} = \{ v(uv)^n \mid n \geqq 1 \}$ is not a right power-bounded language, a contradiction. #

The converse of this Proposition is false. For example, let (X, h) be a DOL scheme such that $X = \{ a, b \}$, $h(a) = ba$, $h(b) = b$. Let $A = \{ a^n b \mid n \geqq 1 \}$. Then $| h(X) | = | X |$ and $h(X)$ is a code. Hence by Proposition 7, (X, h) is a MOL scheme. But $h(A) = \{ ((ba)^n b \mid n \geqq 1 \} = \{ b(ab)^n \mid n \geqq 1 \}$ is not a right power-bounded language while A is.

PROPOSITION 10: *Let (X, h) be a DOL scheme such that $| h(X) | = | X |$. Then $h(X)$ is a prefix code if and only if (X, h) is a scheme which preserves the prefix codes.*

Proof: Sufficiency. Trivial.

Necessity. Let A be a prefix code. We have to show that $h(A)$ is also a prefix code. The case $| h(A) | = 1$ is trivial. Suppose that $| h(A) | \geqq 2$ and let $p \neq q \in h(A)$. Then there exist $u, v \in A$ such that $h(u) = p$, $h(v) = q$. Let $u = u_1 u_2 \ldots u_n$, $v = v_1 v_2 \ldots v_m$, $u_i, v_j \in X$. Since $u_1 u_2 \ldots u_n = u \neq v = v_1 v_2 \ldots v_m$, there exists $k \geqq 1$ such that $u_k \neq v_k$ and $u_i = v_i$ for all $i < k$. Since

$$p = h(u) = h(u_1) \ldots h(u_{k-1})\, h(u_k)\, h(u_{k+1} \ldots u_n),$$
$$q = h(v) = h(v_1) \ldots h(v_{k-1})\, h(v_k)\, h(v_{k+1} \ldots v_m),\ h(u_i) = h(v_i)$$

for all $i < k$ and $\{ h(u_k), h(v_k) \}$ a prefix code by assumption, then $\{ p, q \}$ is a prefix code. Therefore $h(A)$ is a prefix code. #

A word $w \in X^+$ is called a *primitive word* if $w = p^n$, $p \in X^+$, implies $n = 1$. It is well known that for any $x \in X^+$, there exists a unique primitive word p and $n \geqq 1$ such that $x = p^n$. Let $Q = \{ p \in X^+ \mid p \text{ is a primitive word} \}$, $Q^{(1)} = Q \cup \{ 1 \}$ and $Q^{(i)} = \{ p^i \mid p \in Q \}$, $i \geqq 2$. Then $X^* = \bigcup_{i=1}^{\infty} Q^{(i)}$ and $Q^{(i)} \cap Q^{(j)} = \emptyset$ if $i \neq j$ (*see* [3]). If $x = p^n$, $p \in Q$, then $\sqrt{x} = p$ is called *the root of* x. In particular $\sqrt{1} = 1$. A language $A \subseteq X^*$ is called *pure* if for any $x \in A^*$, $\sqrt{x} \in A^*$.

A language $A \subseteq X^*$ is called *noncounting* (*left-noncounting*) if there exists $k \geqq 1$ such that $ux^k v \in A$ if and only if $ux^{k+1} v \in A$, ($x^k v \in A$ if and only if

$x^{k+1} v \in A$) for all u, x, $v \in X^*$. A language $A \subseteq X^*$ is said to be a *power-separating language* if there exists $k \geqq 1$, called the order of A, such that for any $x \in X^*$ either $x^k x^* \subseteq A$ or $x^k x^* \cap A = \emptyset$. Every noncounting language is left-noncounting and every left-noncounting language is power-separating, but the converse is not true (*see* [6], [7]).

In [5], Restivo has shown that a finite code $A \subseteq X^*$ is pure if and only if A^* is a noncounting language. In order to extend this result, let us recall that a language $A \subseteq X^*$ is a code if and only if $f \in X^*$, $f A^* \cap A^* \cap A^* f \neq \emptyset$ implies $f \in A^*$. From this, it follows that if A is a code, then x^n and $x^{n+r} \in A^*$ imply $x^r \in A^*$.

PROPOSITION 11: *Let $A \subseteq X^*$ be a finite code. Then the following are equivalent:*

(1) *A is pure;*

(2) *A^* is a power-separating language;*

(3) *A^* is a left-noncounting language.*

Proof: (1) implies (3). Suppose A is pure. Then A^* is a noncounting language (*see* [5]) and hence a left-noncounting language.

(3) implies (2). Immediate.

(2) implies (1). Suppose that A is not pure. Then there exists a word $x \in A^*$ such that $x = p^k$, $k > 1$ and $p \notin A^*$. Thus $p^n \in A^*$ for all $n = kr$, $r \geqq 1$. Since A is a code by assumption and since $p \notin A^*$, then $p^{n+1} \notin A^*$. This implies that A^* is not a power-separating language. #

A DOL scheme (X, h) is said to be a scheme *preserving* the *primitive words*, if for any primitive word $p \in X^+$, $h(p)$ is a primitive word. i. e., if $h(Q) \subseteq Q$.

PROPOSITION 12: *Every MOL scheme (X, h) such that $h(X)$ is a pure code, preserves the primitive words.*

Proof: Let $g \in Q$. Then $h(g) = p^n \in [h(X)]^* \subseteq X^*$, where $p \in Q$. Since $h(X)$ is pure by assumption, we have $p \in [h(X)]^*$. It follows then that for some $x \in X^*$, $h(x) = p$ and $h(x^n) = p^n = h(g)$. Since (X, h) is a MOL scheme, then h is injective and $g = x^n$. Since $g \in Q$, we have $n = 1$. Thus $h(g)$ is a primitive word. #

The MOL scheme (X, h), where $X = \{a, b\}$ and $h(a) = ab$, $h(b) = ba$, is an example of a MOL scheme preserving the primitive words.

PROPOSITION 13: *Every MOL scheme (X, h) such that $h(X)$ is a pure code, preserves the pure languages.*

Proof: Let A be a pure language and let $p^n \in [h(A)]^*$, $p \in Q$. Then there exists $x \in A^*$ such that $h(x) = p^n$ and $x = q^m$, $q \in Q$. This implies that $p^n = h(q^m) = [h(q)]^m$. Since $h(X)$ is a pure code, then by Proposition 12, $h(q)$ is a primitive word. Hence $n = m$ and $p = h(q)$. Since A is pure, then

$x = q^m \in A^*$ implies that $q \in A^*$ and $p = h(q) \in [h(A)]^*$. Therefore $h(A)$ is pure. #

PROPOSITION 14: *Every MOL scheme, such that $h(X)$ is a pure code, preserves the power-separating languages.*

Proof: Since $h(X)$ is a pure code, then by Proposition 11, $[h(X)]^* = h(X^*)$ is a power-separating language, say of order m. Then, by definition, for any $x \in X^*$, either $x^m x^* \subseteq h(X^*)$ or $x^m x^* \cap h(X^*) = \emptyset$. Now let A be any power-separating language of order n. We will show that $h(A)$ is a power-separating language of order nm. Let $x \in X^*$, $x \neq 1$. If $x^m x^* \cap h(X^*) = \emptyset$, then $x^{nm} x^* \cap h(A) = \emptyset$. Now suppose that $x^m x^* \subseteq h(X^*)$. Then there exists $y \in X^*$ such that $h(y) = x^m$. Let $y = p^r$, $x = q^s$, $r, s \geqq 1$, $p, q \in Q$. Then $[h(p)]^r = h(y) = x^m = q^{sm}$. Since $h(X)$ is pure, then by Proposition 12, $h(p)$ is primitive and $h(p) = q$, $r = sm$.

If $p^n p^* \subseteq A$, then

$$p^{nms} p^* \subseteq A \quad \text{and} \quad h(p^{nms} p^*) = [h(p)]^{nms} [h(p)]^* = q^{nms} q^* = x^{nm} q^* \subseteq h(A).$$

This implies that $x^{nm} x^* \subseteq h(A)$, because $x^* \subseteq q^*$.

If $p^n p^* \cap A = \emptyset$, then $p^n p^* \subseteq \bar{A} = X^* - A$ and $\bar{A}$ is also a power-separating language of order n. By using the same argument as above, it can be shown that $x^{nm} x^* \subseteq h(\bar{A})$. Since h is injective, then $h(A) \cap h(\bar{A}) = \emptyset$, and therefore $x^{nm} x^* \cap h(A) = \emptyset$.

It follows then that $h(A)$ is a power-separating language of order nm. #

REFERENCES

1. G. T. HERMAN and G. ROZENBERG, *Developmental Systems and Languages*, Amsterdam, North Holland, 1975.
2. A. LENTIN and M. P. SCHÜTZENBERGER, *A Combinatorial Problem in Theory of Free Monoids*, in *Combinatorial Mathematics and its Applications*, Proceedings of the conference held at University of North Carolina, 1967, pp. 128-144.
3. R. C. LYNDON and M. P. SCHÜTZENBERGER, *On the Equation $a^M = b^N c^P$ in a Free Group*, Michigan Math. J. Vol. 9, 1962, pp. 289-298.
4. M. NIVAT, *Éléments de la théorie générale des codes*, Automata Theory, E. R. CAIANIELLO, Ed., Academic Press, New York, 1966, pp. 278-294.
5. A. RESTIVO, *On a Question of McNaughton and Papert*, Information and Control, Vol. 25, 1974, pp. 93-101.
6. H. J. SHYR and G. THIERRIN, *Power-separating Regular Languages*, Math. Systems Theory, Vol. 8, No. 1, 1974, pp. 90-95.
7. H. J. SHYR and G. THIERRIN, *Left-noncounting languages*, Inter. J. of Computer and Information Sciences, Vol. 4, No. 1, March 1975, pp. 95-102.
8. H. J. SHYR and G. THIERRIN, *Codes and binary relations*, Springer Lecture Notes in Mathematics, Vol. 586, 1977, pp. 180-188.
9. G. THIERRIN, *Regular Prefix Codes and Right Power-bounded Languages*, Semigroup Forum, Vol. 13, 1976, pp. 77-83.