INFORMATION AND COMPUTATION **73**, 59–74 (1987)

Semi-commutations

M. CLERBOUT AND M. LATTEUX

CNRS, LA 369, University of Lille 1, Lille, France

Nous étendons la notion de commutation partielle en introduisant celle de semi-commutation qui en est la version non symétrique. Une fonction de semi-commutation f est associée à un système de réécriture $\langle X, P\rangle$ où toutes les productions sont de la forme $yx \to xy$ avec $x, y \in X$. Nous étudions cette opération en liaison avec les opérations rationnelles. Nous prouvons que si $f(R)$ et $f(R')$ sont des langages rationnels, $f(RR')$ est aussi un langage rationnel et nous donnons une condition suffisante pour que $f(R^*)$ soit un langage rationnel. © 1987 Academic Press, Inc.

We extend the notion of partial commutation by introducing that of semi-commutation which is its non-symetrical version. A semi-commutation function f is associated to a semi-Thue system $\langle X, P\rangle$ where the rules are of the form $yx \to xy$ with $x, y \in X$. We study this operation in connection with rational operations. We prove that if $f(R)$ and $f(R')$ are regular languages then $f(RR')$ is a regular language and we give a sufficient condition which ensures that $f(R^*)$ is a regular language. © 1987 Academic Press, Inc.

INTRODUCTION

The study of free partially commutative monoïds has been initiated by Cartier and Foata (1969). Trace languages, which are subsets of a free partially commutative monoïd have been proposed by Mazurkiewicz (1977) as a tool for describing the behavior of concurrent program schemes. To further that proposition, there were a lot of papers devoted to the study of trace languages and free partially commutative monoïds. A survey concerning the theory of traces can be found an Aalbersberg and Rozenberg (1985). A free partially commutative monoïd can be viewed as a Thue system $S = \langle X, P\rangle$, where X is a finite alphabet and P is a set of symmetrical productions of the form $xy \leftrightarrow yx$ with $x, y \in X$. The operation which links to a word $u \in X^*$ the set of the words which can be obtained from u in the system S is called a partial commutation. In Clerbout and Latteux (1984), partial commutations are examined in relation to other classical operations on languages and some representation theorems for well-known language families are derived. Recently, several papers gave sufficient conditions which ensure that the image by a partial commutation

0890-5401/87 $3.00

of a regular language remains regular (Fle and Roucairol, 1982; Cori and Perrin, 1985; and Cori and Metivier, 1985).

In this paper, we extend the notion of partial commutation by introducing a non-symmetrical version of that operation. A semi-commutation is a semi-Thue system $S = \langle X, P \rangle$, where X is a finite alphabet and P a set of productions of the form $yx \to xy$ with $x, y \in X$. The operation which links to a work $u \in X^*$ the set of the words which can be derived from u in the system S, is called a semi-commutation. Note that semi-commutations can be of interest for describing synchronization processes since the semi-Dyck language which often appears for such problems is, quite naturally, the image of the regular language $(a\bar{a})^*$ by the semi-commutation associated to the system $\langle X, P \rangle$, where $X = \{a, \bar{a}\}$ and $P = \{\bar{a}a \to a\bar{a}\}$.

First, we prove several easy basic results concerning semi-commutations and we deduce that if f is a semi-commutation and $f(R_1)$, $f(R_2)$ are regular languages, then $f(R_1 R_2)$ is a regular language. Then, we present the main result, that is, a sufficient condition which ensures that if $f(R)$ is a regular language, then $f(R^*)$ is also a regular language. At last, note that some other properties of semi-commutations are established in Clerbout (1984).

I. Definitions and Notations

First, we recall some classical definitions. For a finite alphabet X, X^* will denote the free monoïd generated by X and the empty word will be represented by ε. For $u \in X^*$, $\mathrm{alph}(u) = \{x \in X / |u|_x \neq 0\}$, where $|u|_x$ denotes the number of occurrences of the letter x in the word u.

A mapping $\tau: X^* \to 2^{Y^*}$ is

— *alphabetic* if $\forall x \in X$, $\tau(x) \subset Y \cup \{\varepsilon\}$,

— *strictly alphabetic* if $\forall x \in X$, $\tau(x) \subset Y$,

— *a substitution* if $\tau(\varepsilon) = \{\varepsilon\}$ and $\forall u_1, u_2 \in X^*$,

$$\tau(u_1 u_2) = \tau(u_1)\, \tau(u_2)$$

— *a projection from X to Y*, denoted $\mathrm{proj}(X, Y)$ if $Y \subset X$ and τ is a substitution defined on X by: $\forall x \in X$, $\tau(x) = \{x\}$ if $x \in Y$, $\tau(x) = \{\varepsilon\}$ otherwise.

The *inverse* of τ, denoted $\tau^{-1}: Y^* \to 2^{X^*}$ is defined by: $\forall v \in Y^*$, $\tau^{-1}(v) = \{u \in X^* / v \in \tau(u)\}$.

We have, now, to introduce a new definition.

Definition. A *semi-commutation system* $S = \langle X, P \rangle$ is a semi-Thue system, where X is a finite alphabet and P is a set of rules of the form $yx \to xy$ with $x, y \in X$ and $x \neq y$.

For $w, w' \in X^*$, we shall write $w \xrightarrow{*}_S w'$ if there exists, in the system S, a derivation from w to w'. Then, clearly, w and w' have the same commutative image.

The *semi-commutation function* $f\colon X^* \to 2^{X^*}$ associated to the system S is defined by: $\forall w \in X^*, f(w) = \{w'/w \xrightarrow{*}_S w'^*\}$. We shall use the obvious fact that f^{-1}, the inverse of f is also a semi-commutation function. Indeed, f^{-1} is associated to $S^{-1} = \langle X, P^{-1}\rangle$, where $P^{-1} = \{v \to u/u \to v \in P\}$.

II. Semi-commutation and Product

Now, we shall present some basic results concerning semi-commutations in relation with alphabetic substitutions. First, we define the image of a semi-commutation system under an alphabetic substitution.

Definition. Let $S = \langle X, P\rangle$ be a semi-commutation system and $\theta\colon X^* \to 2^{Y^*}$ be an alphabetic substitution. Then the image of S under θ is $S_\theta = \langle Y, P_\theta\rangle$, where $P_\theta = \{y_1 y_2 \to y_2 y_1 \mid y_1, y_2 \in Y, y_1 \neq y_2, \exists u \to v \in P$ s.t. $y_1 y_2 \in \theta(u)\}$.

The relations between the semi-commutations f and f_θ associated respectively to S and S_θ are given in the following lemma.

Lemma 1. *The two following properties are satisfied:*

Ⓐ $\forall w \in X^*$, $\theta \circ f(w) \subset f_\theta \circ \theta(w)$ *and* $\forall w' \in Y^*$, $f \circ \theta^{-1}(w') \subset \theta^{-1} \circ f_\theta(w')$.

Ⓑ *Moreover, if* $\theta = g^{-1}$, *where g is strictly alphabetic morphism, then* $\theta \circ f = f_\theta \circ \theta$ *and* $g \circ f_\theta = f \circ g$.

Proof. Let $x_1 x_2 \to x_2 x_1$ be a production from P and v be a word in $\theta(x_2 x_1)$. By construction of P_θ, either v is in $\theta(x_1 x_2)$ or these exists a u in $\theta(x_1 x_2)$ such that $u \to v \in P_\theta$. Thus, if $w_1 \to_S w_2$, $w'_1 \in \theta(w_1)$, and $w'_2 \in \theta(w_2)$, then $w'_2 \in f_\theta(w'_1)$. Hence, $w'_2 \in f_\theta \circ \theta(w_1)$, $w_2 \in \theta^{-1} \circ f_\theta(w'_1)$, and we get property A by induction on the length of the derivation in the system S.

For property B, let us take $w_1 \in X^*$, $w'_1 \in \theta(w_1)$, and $w'_1 \to_{S_\theta} w'_2$. Then $w'_1 = u' y_1 y_2 v'$, $w'_2 = u' y_2 y_1 v'$ with $y_1 y_2 \to y_2 y_1 \in P_\theta$. Then $w_1 = u x_1 x_2 v$ with $u = g(u')$, $x_1 = g(y_1)$, $x_2 = g(y_2)$, and $v = g(v')$. By construction of P_θ, $x_1 x_2 \to x_2 x_1 \in P$, hence $w'_2 \in \theta \circ f(w_1)$. By induction on the length of the derivation in the system S_θ, we obtain: $\forall w \in X^*, f_\theta \circ \theta(w) \subset \theta \circ f(w)$, which implies together with property A, $\theta \circ f = f_\theta \circ \theta$.

Now, $g \circ \theta$ is the identity function on X^* and $\forall w' \in Y^*$, $w' \in \theta \circ g(w')$. Thus, for every $w' \in Y^*$, we can write $g \circ f_\theta(w') \subset g \circ f_\theta \circ \theta \circ g(w') =$

$g \circ \theta \circ f \circ g(w') = f \circ g(w')$. Since $g = \theta^{-1}$, we can deduce from property A that $g \circ f_\theta = f \circ g$. ▮

An interesting particular case is when $\forall x \in X$, $\theta(x) \subset \{x, \varepsilon\}$. Then, $Y = X \cap \theta(X) \subset X$ and clearly, $\forall w' \in Y^*$, $\theta(w') = w'$. Thus, $\theta \circ \theta = \theta$, $f_\theta \circ \theta = f \circ \theta$, and $\forall w \in X^*$, $f \circ \theta(w) \subset \theta \circ f \circ \theta(w)$. Conversely, from property A of Lemma 1, $\forall w \in X^*$, $\theta \circ f \circ \theta(w) \subset f_\theta \circ \theta \circ \theta(w) = f \circ \theta(w)$ and we get

LEMMA 2. *If for every $x \in X$, $\theta(x) \subset \{x, \varepsilon\}$, then $f \circ \theta = \theta \circ f \circ \theta$.*

Another result can be useful.

LEMMA 3. *If $wa \in f(uav)$ with $u, v, w \in X^*$, $a \in X$, and $a \notin \mathrm{alph}(v)$, then:*

(i) $\forall x \in \mathrm{alph}(v)$, $ax \to xa \in P$,

(ii) $w \in f(uv)$.

Proof. Property (i) is obvious.

For property (ii), let us consider $Y = X \cup \{\bar{a}\}$, θ the substitution defined on X by $\theta(x) = x$ if $x \neq a$, $\theta(a) = \{a, \bar{a}\}$ and h the morphism defined on X by $h(x) = x$ if $x \in X$, $h(\bar{a}) = \varepsilon$. From Lemma 1, we have $g \circ f_\theta(u\bar{a}v) = f \circ g(u\bar{a}v) = f(uav)$ with $g = \theta^{-1}$. Thus $wa \in g \circ f_\theta(u\bar{a}v)$, which implies that there exists $w'' \in f_\theta(u\bar{a}v)$ such that $g(w'') = wa$. Since $a\bar{a} \to \bar{a}a \notin P_\theta$, necessarily $w'' = w\bar{a}$. By construction, $(f_\theta)_h = f$ and from Lemma 1, we get $w = h(w\bar{a}) \in h \circ f_\theta(u\bar{a}v) \subset (f_\theta)_h \circ h(u\bar{a}v) = f(uv)$. ▮

LEMMA 4. *Let w_1 be in X_1^*, w_2 be in X_2^*, where X_1, X_2 are disjoint subsets of X. Then, $\forall w \in f(w_1 w_2)$, $w \in f(\pi_1(w)\,\pi_2(w))$, where for $i \in [1, 2]$, $\pi_i = \mathrm{proj}(X, X_i)$.*

Proof. We shall reason by induction on the length of w. If $w = \varepsilon$, we are done. Let us take $w = w'a$ with $a \in X_1 \cup X_2$. By hypothesis there exists $uav \in X_1^* X_2^*$ with $a \notin \mathrm{alph}(v)$ such that $w'a \in f(uav)$. From the previous lemma, $\forall x \in \mathrm{alph}(v)$, $ax \to xa \in P$, and $w' \in f(uv)$ with $uv \in X_1^* X_2^*$. By inductive hypothesis, $w' \in f(\pi_1(w')\,\pi_2(w'))$. If $a \in X_2$, then $w = w'a \in f(\pi_1(w')\,\pi_2(w')a) = f(\pi_1(w)\,\pi_2(w))$. If $a \in X_1$, $\mathrm{alph}(\pi_2(w')) \subset \mathrm{alph}(v)$ and $a\pi_2(w') \in f(\pi_2(w')a)$ and we get $w = w'a \in f(\pi_1(w')\,\pi_2(w')a) \subset f(\pi_1(w')\,a\pi_2(w')) = f(\pi_1(w)\,\pi_2(w))$. ▮

Now, if L_1 and L_2 are languages over disjoint alphabets X_1 and X_2, we can express $f(L_1 L_2)$ by means of $f(L_1)$, $f(L_2)$, and $f(X_1^* X_2^*)$.

PROPOSITION 5. *Let X_1, X_2 be disjoint subsets of X and for $i \in [1, 2]$, π_i be equal to $\mathrm{proj}(X, X_i)$. Then, for $L_1 \subset X_1^*$ and $L_2 \subset X_2^*$, $f(L_1 L_2) = f(X_1^* X_2^*) \cap \pi_1^{-1}(f(L_1)) \cap \pi_2^{-1}(f(L_2))$.*

Proof. Let us take $w \in f(w_1 w_2)$ with $w_1 \in L_1$ and $w_2 \in L_2$. Then, $w \in f(X_1^* X_2^*)$ and from Lemma 1, $\pi_1(w) \in \pi_1 \circ f(w_1 w_2) \subset f_{\pi_1} \circ \pi_1(w_1 w_2) = f(w_1) \subset f(L_1)$, which implies $w \in \pi_1^{-1}(f(L_1))$. Similarly, we get $w \in \pi_2^{-1}(f(L_2))$, hence a first inclusion.

Conversely, let us take $w \in f(X_1^* X_2^*)$ with $\pi_1(w) \in L_1$ and $\pi_2(w) \in L_2$. From Lemma 4, $w \in f(\pi_1(w)\, \pi_2(w)) \subset f(L_1 L_2)$. ∎

Let us consider, now, $X_1, X_2 \subset X$ and $L = f(X_1^* X_2^*)$. In order to prove that L is a regular language, one can construct, directly, an finite automaton which recognizes L. Another way is to prove that L is closed under subword, since such languages are regular (see Lothaire, 1983). For that let SW be the substitution defined on X by: $\forall x \in X$, $SW(x) = \{x, \varepsilon\}$. Since $SW(X_1^* X_2^*) = X_1^* X_2^*$, we get, from Lemma 2, $SW(L) = SW \circ f(X_1^* X_2^*) = SW \circ f \circ SW(X_1^* X_2^*) = f \circ SW(X_1^* X_2^*) = L$, hence:

Lemma 6. *$f(X_1^* X_2^*)$ is a regular for every X_1, X_2 included in X.*

We can, now, extend this result.

Proposition 7. *Let $f(L_1)$ and $f(L_2)$ be regular languages over the alphabet X. Then $f(L_1 L_2)$ is also a regular language.*

Proof. Let us define $\bar{X} = \{\bar{x}/x \in X\}$, $Y = X \cup \bar{X}$, θ the substitution by: $\forall x \in X$, $\theta(x) = \{x, \bar{x}\}$, $\bar{L}_2 = \theta(L_2) \cap \bar{X}^*$, $g = \theta^{-1}$. From Lemma 1, $f(L_1 L_2) = f \circ g(L_1 \bar{L}_2) = g \circ f_\theta(L_1 \bar{L}_2)$ and it suffices to prove that $f_\theta(L_1 \bar{L}_2)$ is a regular language. From Proposition 5, $f_\theta(L_1 \bar{L}_2) = f_\theta(X^* \bar{X}^*) \cap \pi^{-1}[f_\theta(L_1)] \cap \pi'^{-1}[f_\theta(\bar{L}_2)]$, where $\pi = \mathrm{proj}(Y, X)$ and $\pi' = \mathrm{proj}(Y, \bar{X})$. Clearly, $f_\theta(L_1) = f(L_1)$and $f_\theta(\bar{L}_2) = \theta(f(L_2)) \cap \bar{X}^*$ are regular languages. Hence, $f_\theta(L_1 \bar{L}_2)$ which is the intersection of three regular languages is a regular language. ∎

Note that our construction permits us to also prove that if $f(L_1)$ is a context-free language and $f(L_2)$ is a regular language, then $f(L_1 L_2)$ is context-free language.

III. Semi-commutation and Star Operation

We shall, now, study the same problem for the star operation. Clearly, a similar result does not hold since there are regular languages $f(R)$ such that $f(R^*)$ is not regular. For example, let f be the semi-commutation function associated to the system $\{X, P\}$ with $X = \{a, b\}$ and $P = \{ba \to ab\}$. For $R = f(R) = \{ab\}$, it is clear that $f(R^*)$ is the semi-Dyck language, hence $f(R^*)$ is not regular.

Thus, we are just going to give a sufficient condition in order to obtain a positive answer. For that, let us introduce the non-commutation graph $\Gamma(S)$ associated to a semi-commutation system $S = \langle X, P \rangle$. The directed graph $\Gamma(S)$ is defined by its vertices, X, and by its edges: (x, y) is an edge if $x \neq y$ and $xy \to yx \notin P$. Note that this notion has been introduced for partial commutations in Cori and Metivier (1984).

The restriction of $\Gamma(S)$ to $X' \subset X$ will be denoted by $(\Gamma(S), X')$. At last, a graph is strongly connected if for every couple x, y of vertices with $x \neq y$, there is a path from x to y. We can, now, give our main result.

PROPOSITION 8. *Let f be the semi-commutation associated to the system $S = \langle X, P \rangle$ and $R = f(R) \subset X^*$ be a regular language. If, for every word $w \in R$, the graph $(\Gamma(S), \mathrm{alph}(w))$ is strongly connected, then $f(R^*)$ is a regular language.*

Proof. From the finite deterministic automaton $M = (X, Q, q_0, *, F)$ which accepts R, we are going to construct a right linear grammar $G = (X, N, D, (\varepsilon))$ which generates the language $f(R^*)$.

The set of non-terminals N is defined by

$N = \{((X_1, q_1)(X_2, q_2) \cdots (X_k, q_k))/k \geqslant 0,\ \forall i \in [1, k],\ X_i \subset X,$ $q_i \in Q$, and there is at most one index j such that $X_j \subset \bigcup_{t>j} X_t\}$ (Note that N is a finite set), where $(\varepsilon) \in N$ is the axiom.

At last, the productions of D are of five kinds:

① $(\varepsilon) \to \varepsilon$

② $((X_1, q_1)(X_2, q_2) \cdots (X_k, q_k)) \to x((X_1, q_1) \cdots (\{x\}, q_0 * x)(X_i, q_i) \cdots (X_k, q_k))$ if

(i) $((X_1, q_1) \cdots (\{x\}, q_0 * x) \cdots (X_k, q_k))$ is in N

(ii) $\forall t \in [i, k], \forall y \in X_t, xy \to yx \in P$

③ $((X_1, q_1) \cdots (X_k, q_k)) \to x((X_1, q_1) \cdots (X_i \cup \{x\}, q_i * x) \cdots (X_k, q_k))$ if

(i) $((X_1, q_1) \cdots (X_i \cup \{x\}, q_i * x) \cdots (X_k, q_k))$ is in N

(ii) $\forall t \in [i+1, k]$, $\forall y \in X_t$, $xy \to yx \in P$

④ $((X_1, q_1) \cdots (X_k, q_k)) \to ((X_2, q_2) \cdots (X_k, q_k))$ if q_1 is a terminal state of the automaton M.

⑤ $((X_1, q_1) \cdots (X_k, q_k)) \to ((X_1, q_1) \cdots (X_{i-1}, q_{i-1})(X_{i+1}, q_{i+1}) \cdots (X_k, q_k))$ if $X_i \subset \bigcup_{t>i} X_t$ and q_i is a terminal state.

Intuitively, in the beginning of a word of $f(R^*)$, we can find subwords which are words of R, and subwords which are left parts of words of R.

Every (X_i, q_i) is a trace of one of these left factors: X_i is the alphabet on which it is defined and q_i is the state in the automaton M.

When we read a letter x which is the next letter of the left factor w_i associated to (X_i, q_i), we know that it had to commute with the letters of other left factors which have been read after w_i (rules ② and ③). When a complete word has been recognized, we erase its trace in the non-terminal (rules ④ and ⑤).

This appears in the following example. Let us take $X = \{a, b, c, d, e\}$ and define the semi-commutation system $S = \langle X, P\rangle$ by its non-commutation graph $\Gamma(S)$ (see Fig. 1).

Let us consider the regular language $R = \{de\} \cup \{a^i b / i \geqslant 2\} \cup \{ab^j c / j \geqslant 1\}$ recognized by the automaton of the Fig. 2. Then $f(R) = R$, where f is the semi-commutation associated to the system S. For every $u \in R$, $\text{alph}(u) \in \{X_1, X_2, X_3\}$ with $X_1 = \{a, b\}$, $X_2 = \{a, b, c\}$, and $X_3 = \{d, e\}$. For $i = 1, 2, 3$, the graph $(\Gamma(S, X_i)$ is clearly strongly connected. Note that this property does not hold for all subgraphs of $\Gamma(S)$. For instance, $(\Gamma(S), \{a, b, c, e\})$ is not strongly connected.

Let us consider the word $u = adbacaeb \in f(deabcaab) \subset f(R^*)$. The derivation $(\varepsilon) \stackrel{*}{\Rightarrow} u$ in the grammar defined above will use the following productions:

— $(\varepsilon) \to a(X_0, q_1)$ of type ② with $X_0 = \{a\}$,

— $(X_0, q_1) \to d((X_4, q_2)(X_0, q_1))$ of type ② with $X_4 = \{d\}$, since $da \to ad \in P$,

— $((X_4, q_2)(X_0, q_1)) \to b((X_4, q_2)(X_1, q_4))$ of type ③, since $X_1 = X_0 \cup \{b\}$ and $q_4 = q_1 * b$,

— $((X_4, q_2)(X_1, q_4)) \to a((X_4, q_2)(X_1, q_4)(X_0, q_1))$ of type ②,

— $((X_4, q_2)(X_1, q_4)(X_0, q_1)) \to c((X_4, q_2)(X_2, q_5)(X_0, q_1))$ of type ③, since $X_2 = X_1 \cup \{c\}$, $q_5 = q_4 * c$ and $ca \to ac \in P$,

— $((X_4, q_2)(X_2, q_5)(X_0, q_1)) \to a((X_4, q_2)(X_2, q_5)(X_0, q_2))$ of type ③, since $X_0 = X_0 \cup \{a\}$ and $q_1 * a = q_2$,

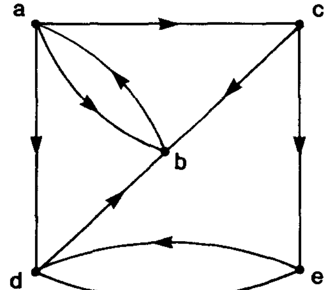

FIGURE 1

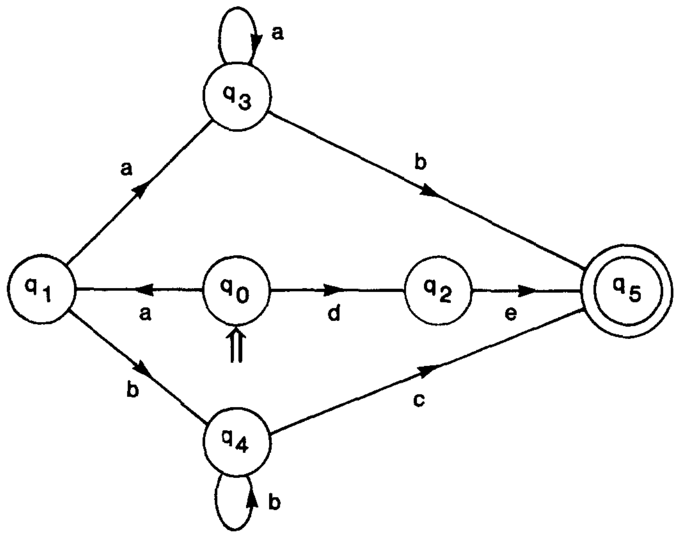

FIGURE 2

— $((X_4, q_2)(X_2, q_5)(X_0, q_2) \to e((X_3, q_5)(X_2, q_5)(X_0, q_2))$ of type ③, since $X_3 = X_4 \cup \{e\}$, $q_5 = q_2 * e$ and $ec \to ce$, $eb \to be$, $ea \to ae \in P$.
— $((X_3, q_5)(X_2, q_5)(X_0, q_2)) \to ((X_2, q_5)(X_0, q_2)) \to (X_0, q_2)$ of type ④, since q_5 is a terminal state.

At last we get the derivation $(X_0, q_2) \Rightarrow_{③} b(X_1, q_5) \Rightarrow_{④} b(\varepsilon) \Rightarrow_{①} b$.

On the contrary, the word $v = adabc$ is not a left factor of a word in $f(R^*)$. In the grammar G, we have, for instance, the derivations:

$(\varepsilon) \xrightarrow{*} adabc((X_0, q_1)(X_4, q_2)(X_3, q_5))$ and $(\varepsilon) \xRightarrow{*} adab((X_4, q_2)(X_2, q_5))$. But one can verify that $w_1, w_2 \in X^*$ do not exist such that $((X_0, q_1)(X_4, q_2)(X_3, q_5)) \xRightarrow{*} w_1$ and $((X_4, q_2)(X_2, q_5)) \xRightarrow{*} cw_2$.

Let us return now to the formal proof of the equality $L(G, (\varepsilon)) = f(R^*)$. We need two lemmas which deal with the links between a derivation in the system S and a derivation according to the grammar G. In these lemmas, we shall use a subset N' of the set of non-terminals N defined by

$$N' = \left\{((X_1, q_1) \cdots (X_k, q_k)) \in N / \forall i \in [1, k],\ X_i \not\subset \bigcup_{t > i} X_t\right\}.$$

LEMMA 9. *Let $S = \langle X, P\rangle$ be a semi-commutation system and w be a word of X^*. If $w \xrightarrow{*}_{S^{-1}} u_1 w_1 u_2 w_2 \cdots u_k w_k$, with the following conditions:*

(C1) $((\mathrm{alph}(w_1), q_0 * w_1)(\mathrm{alph}(w_2), q_0 * w_2) \cdots (\mathrm{alph}(w_k), q_0 * w_k)) \in N'$

(C2) $\forall i \in [1, k],\ u_i \in R^*$.

(C3) $\forall i \in [2, k]$, $\mathrm{alph}(u_i) \subseteq \mathrm{alph}(w_i \cdots w_k)$.

(C4) $\forall i \in [1, k-1]$, there is a subset $Y_i \subset X$ such that, $\forall x \in Y_i$, $\forall y \in \mathrm{alph}(w_{i+1} \cdots w_k)$, $xy \to yx \in P$, and the graph $(\Gamma(S), Y_i \cup \mathrm{alph}(w_i))$ is strongly connected.

then, there exist k words $z_1, z_2, \ldots, z_k$ such that

$$(\varepsilon) \xRightarrow[G]{*} w((\mathrm{alph}(w_1), q_0 * z_1) \cdots (\mathrm{alph}(w_k), q_0 * z_k))$$

$$\text{and} \qquad \forall i \in [1, k], \quad w_i \xrightarrow[S]{*} z_i.$$

Proof. We reason by induction on the length l of the word w:

If $l = 0$, there is no problem.

For $l > 0$, set $w' = wa$, where $a \in X$ and $wa \xrightarrow{*}_{S^{-1}} u_1 w_1 \cdots u_k w_k$ with conditions (C1)–(C4).

Then, we can write $u_1 w_1 \cdots u_k w_k = z'az''$ with $|z''|_a = 0$. From Lemma 3, we deduce the two following properties:

(i) $w \xrightarrow[S^{-1}]{*} z'z''$

(ii) $az'' \xrightarrow[S]{*} z''a$.

We now just have to find a "good" factorization of w (according to the lemma), in order to apply the induction hypothesis. For that, we are led to study two different cases.

Case 1. The letter "a" is appearing in a "w_i", that is, there exists $i \in [1, k]$ such that $w_i = w_i' a w_i''$, with $|w_i''|_a = 0$ ($z'' = w_i'' u_{i+1} \cdots w_k$). So,

$$w \xrightarrow[S^{-1}]{*} u_1 w_1 \cdots u_i w_i' w_i'' u_{i+1} \cdots u_k w_k.$$

($*$) In order to prove that this factorization verifies (C1), we shall prove that $((\mathrm{alph}(w_1), q_0 * w_1) \cdots (\mathrm{alph}(w_i' w_i''), q_0 * w_i' w_i'') \cdots (\mathrm{alph}(w_k), q_0 * w_k))$ is in N'.

— $\forall j \neq i$, $\mathrm{alph}(w_j) \not\subset \mathrm{alph}(w_{j+1} \cdots w_k)$ by hypothesis.

— For $j = i$, there exists Y_i such that the subgraph $(\Gamma(S), \mathrm{alph}(w_i' a w_i'') \cup Y_i)$ is strongly connected and $\forall y \in Y_i$, $\forall z \in \mathrm{alph}(w_{i+1} \cdots w_k)$, $yz \to zy \in P$.

Set $Y_i' = Y_i \cup \{a\}$. Thus, we can write that $(\Gamma(S), \mathrm{alph}(w_i' w_i'') \cup Y_i')$ is strongly connected and $\forall z \in \mathrm{alph}(w_{i+1}, \ldots, w_k)$, $yz \to zy \in P$ (since $az'' \xrightarrow{*}_S z''a$), and we can deduce $\mathrm{alph}(w_i' w_i'') \not\subset \mathrm{alph}(w_{i+1} \cdots w_k)$.

Indeed, otherwise, $\forall y \in Y_i'$, $\forall z \in \mathrm{alph}(w_i' w_i'')$, $yz \to zy \in P$, and the subgraph $(\Gamma(S), Y_i' \cup \mathrm{alph}(w_i' w_i''))$ is not strongly connected, a contradiction.

(∗) It is obvious that this factorization verifies (C2).

(∗) This factorization verifies (C4). Indeed. $\forall t \in [1, k]$, if $t \neq i$, y_t exists by hypothesis. If $t = i$, one can take $Y_i' = Y_i \cup \{a\}$ (see above).

(∗) If this factorization verifies (C3), then we can apply the induction hypothesis to the word w and this factorization. So there exist k words $z_1, ..., z_k$ such that $w_t \xrightarrow{*}_S z_t$ for each $t \neq i$, $w_i' w_i'' \xrightarrow{*}_S z_i$, and

$$(\varepsilon) \underset{G}{\overset{*}{\Longrightarrow}} w((\text{alph}(w_1), q_0 * z_1) \cdots (\text{alph}(w_i' w_i''), q_0 * z_i) \cdots (\text{alph}(w_k), q_0 * z_k)).$$

Since $aw_{i+1} \cdots w_k \xrightarrow{*}_S w_{i+1} \cdots w_k a (ii)$, we have

$$(\varepsilon) \underset{G}{\overset{*}{\Longrightarrow}} wa((\text{alph}(w_1), q_0 * z_1) \cdots \text{alph}(w_i), q_0 * z_i a) \\ \cdots (\text{alph}(w_k), q_0 * z_k)),$$

and $w_i' a w_i'' = w_i \xrightarrow{*}_S w_i' w_i'' a \xrightarrow{*}_S z_i a$.

(∗) If the factorization does not berify (C3): Let j be the greatest index such that $2 \leqslant j \leqslant k$ and $\text{alph}(u_j) \not\subset \text{alph}(w_j \cdots w_k)$.

Necessarily, we have $|u_j|_a \neq 0$ and $j \leqslant i$. Set $u_j = uzatv$ with u, $v \in R^*$, $|v|_a = 0$, $zat \in R$, and $|t|_a = 0$. So, we can write

$$w \xrightarrow[S]{*} u_1 w_1 \cdots w_{j-1} u \,\text{zat}\, v w_j \cdots u_i w_i' w_i'' \cdots u_k w_k.$$

Set $w_j' = zat$. It remains to prove that this factorization verifies conditions (C1)–(C4).

(∗) It verifies (C1), that is $(\text{alph}(w_1), q_0 * w_1) \cdots (\text{alph}(w_j'), q_0 * w_j') \cdots (\text{alph}(w_i' w_i''), q_0 * w_i' w_i'') \cdots (\text{alph}(w_k), q_0 * w_k))$ is in N'. Indeed,

- for $t > i$, there is no problem, since $\text{alph}(w_t) \not\subset \text{alph}(w_{t+1} \cdots w_k)$ by hypothesis;
- for $t = i$, we reason as above;
- for $j \leqslant t < i$, $\text{alph}(w_t) \not\subset \text{alph}(w_j \cdots w_i' w_i'' \cdots w_k)$, by hypothesis;
- since $a \in \text{alph}(w_j')$ and $a \notin \text{alph}(w_j \cdots w_i' w_i'' \cdots w_k)$, we get $\text{alph}(w_j') \not\subset \text{alph}(w_j \cdots w_i' w_i'' \cdots w_k)$;
- for $1 \leqslant t < j$, $\text{alph}(w_j') \subset \text{alph}(u_j) \subset \text{alph}(w_j \cdots w_i \cdots w_k)$, and $\text{alph}(w_j \cdots w_i \cdots w_k) = \text{alph}(w_j \cdots w_i' w_i'' \cdots w_k) \cup \{a\}$.

So, $\text{alph}(w_j' w_j \cdots w_i' w_i'' \cdots w_k) = \text{alph}(w_j \cdots w_i \cdots w_k)$. By hypothesis, $\text{alph}(w_t) \not\subset \text{alph}(w_{t+1} \cdots w_j \cdots w_i \cdots w_k)$, hence we get $\text{alph}(w_t) \not\subset \text{alph}(w_{t+1} \cdots w_j' w_j \cdots w_i' w_i'' \cdots w_k)$.

(∗) It is clear that this factorization verifies (C2).

(∗) Let us prove that it verifies (C3):

— $\forall t \geqslant j+1$, $\mathrm{alph}(u_t) \subset \mathrm{alph}(w_t \cdots w_i' w_i'' \cdots w_k)$ by hypothesis.

— $\mathrm{alph}(v) \subset \mathrm{alph}(u_j) \backslash \{a\} \subset \mathrm{alph}(w_j \cdots w_i' w_i'' \cdots w_k)$ by hypothesis.

— $\mathrm{alph}(u) \subset \mathrm{alph}(u_j) \subset \mathrm{alph}(w_j \cdots w_i' w_i'' \cdots w_k) \cup \{a\} = \mathrm{alph}(w_j' w_j \cdots w_i' w_i'' \cdots w_k)$.

— $\forall t < j$, $\mathrm{alph}(u_t) \subset \mathrm{alph}(w_t \cdots w_j \cdots w_i' w_i'' \cdots w_k) \cup \{a\} = \mathrm{alph}(w_t \cdots w_j' w_j \cdots w_i' w_i'' \cdots w_k)$.

(∗) It verifies (C4) from the following assertions:

— $\forall t > i$, $\forall t < j$, $\forall t \in [j, i-1]$, Y_t exists by hypothesis.

— For $t = i$, $Y_i' = Y_i \cup \{a\}$ is right.

— For w_j', we can take $Y_j' = \varnothing$, since $w_j' \in R$, and so the graph $(\Gamma(S), \mathrm{alph}(w_j'))$ is strongly connected.

We can now apply the induction hypothesis to w and to this factorization. We have $k+1$ words $z_1, \ldots, z_j', z_j, \ldots, z_k$ such that $\forall t \neq i$, $w_t \xrightarrow{*}_S z_t$, $w_i' w_i'' \xrightarrow{*}_S z_i$, $w_j' \xrightarrow{*}_S z_j'$, and

$$(\varepsilon) \xRightarrow{*}_G w((\mathrm{alph}(w_1), q_0 * z_1) \cdots (\mathrm{alph}(w_j', q_0 * z_j') \cdots (\mathrm{alph}(w_i' w_i''), q_0 * z_i) \cdots (\mathrm{alph}(w_k), q_0 * z_k))$$

with (ii), we can write

$$(\varepsilon) \xRightarrow{*}_G wa((\mathrm{alph}(w_1), q_0 * z_1) \cdots (\mathrm{alph}(w_i' w_i'' a), q_0 * z_i a) \cdots (\mathrm{alph}(w_k), q_0 * z_k)).$$

The non-terminal built here is in N. Indeed,

$$\mathrm{alph}(w_j') \subset \mathrm{alph}(u_j) \subset \mathrm{alph}(w_j \cdots w_i \cdots w_k)$$

and

$$\forall t \in [1, k], \qquad \mathrm{alph}(w_t) \not\subset \mathrm{alph}(w_{t+1} \cdots w_k) \qquad \text{by hypothesis.}$$

Furthermore, $w_j' \in R = f(R)$. Thus, $z_j' \in R$, and we can apply a production of type 5 which erases $(\mathrm{alph}(w_j'), q_0 * z_j')$. Thus

$$(\varepsilon) \xRightarrow{*}_G wa((\mathrm{alph}(w_1), q_0 * z_1) \cdots (\mathrm{alph}(w_i), q_0 * z_i' a) \cdots (\mathrm{alph}(w_k), q_0 * z_k))$$

with, for $t \neq i$, $w_t \xrightarrow{*}_S z_t$ and for $t = i$, $w_i \xrightarrow{*}_S w_i' w_i'' a \xrightarrow{*}_S z_i a$.

Case 2. We have now to study the case where the letter "a" appears in a "u_j"; that is, there is an index j such that $u_j = u'_j a v'_j$, with $|v'_j|_a = 0$ and $z'' = v'_j w_j u_{j+1} \cdots w_k$.

If j is greater than 2, $\operatorname{alph}(u_j) \subset \operatorname{alph}(w_j \cdots w_k)$ by hypothesis and we know that $a \notin \operatorname{alph}(w_j \cdots w_k)$. So we deduce that $j = 1$ and $u_1 = uu'au''v$ with $u,\ v \in R^*,\ u'au'' \in R,\ |u''v|_a = 0,\ z'' = u'vw_1u_2 \cdots w_k$, and $w \xrightarrow{*}_{S^{-1}} uu'u''vw_1u_2w_2 \cdots u_k w_k$.

From this factorization, we have to find another which verifies the conditions (C1)–(C4) of the lemma, in order to be allowed to apply the induction.

Set $v = v_1 v_2 \cdots v_n$, with $n \geq 0$ and $\forall i \in [1, n]$, $v_i \in R$. If $\operatorname{alph}(v_n) \subset \operatorname{alph}(w_1 \cdots w_k)$, v_n will be considered as a "u_i," otherwise, it will be considered as a "w_i."

We iterate this process for $v_{n-1}, v_{n-2}, \ldots$, and v_1, and during this iteration we take into account the last results. Thus, we get a decomposition of the word v: $v = u'_1 w'_1 \cdots w'_{p-1} u'_p$ with

— $\forall i \in [1, p]$, $u'_i \in R^*$, and $\operatorname{alph}(u'_i) \subset \operatorname{alph}(w'_i \cdots w'_{p-1} w_1 \cdots w_k)$

— $\forall i \in [1, p-1]$, $w'_i \in R$, and $\operatorname{alph}(w'_i) \not\subset \operatorname{alph}(w'_{i+1} \cdots w'_{p-1} w_1 \cdots w_k)$.

Thus, $w \xrightarrow{*}_{S^{-1}} uu'u''u'_1 w'_1 \cdots w'_{p-1} u'_p w_1 u_2 w_2 \cdots u_k w_k$. We are now going to prove that this factorization satisfies (C1)–(C4).

(∗) It verifies (C1). Indeed,

— $\forall t \in [1, k]$, $\operatorname{alph}(w_t) \not\subset \operatorname{alph}(w_{t+1} \cdots w_k)$ by hypothesis.

— $\forall t \in [1, p-1]$, $\operatorname{alph}(w'_t) \not\subset \operatorname{alph}(w'_{t+1} \cdots w'_{p-1} w_1 \cdots w_k)$ by construction.

— For $u'u''$, $u'au'' \in R$, and $au''vw_1u_2 \cdots w_k \xrightarrow{*}_S u''vw_1u_2 \cdots w_k a$, thus the graph$(\Gamma(S), \operatorname{alph}(u'au''))$ is strongly connected, and we can find a letter x in $\operatorname{alph}(u'u'')$, $x \neq a$, such that $ax \to xa \notin P$.

We deduce that $\operatorname{alph}(u'u'') \not\subset \operatorname{alph}(v_1 w_1 u_2 \cdots w_k)$, hence $\operatorname{alph}(u'u'') \not\subset \operatorname{alph}(w'_1 \cdots w'_{p-1} w_1 \cdots w_k)$.

(∗) It is clear that (C2) is verified, and by hypothesis and by construction, it is easy to verify (C3).

(∗) It remains to examine condition (C4):

— $\forall t \in [1, k-1]$, we associate to w_t the subset Y_t given by the induction hypothesis.

— $\forall t \in [1, p-1]$, we can associate to w'_t the subset $Y'_t = \emptyset$. Indeed, $w'_t \in R$, hence the subgraph $(\Gamma(S), \operatorname{alph}(w'_t))$ is strongly connected.

— for $u'u''$, since $u'au'' \in R$, and $aw'_1 \cdots w'_{p-1} w_1 \cdots w_k \xrightarrow{*}_S w'_1 \cdots w_k a$, the subset $Y = \{a\}$ can be taken.

Then, by the induction hypothesis, we can write:

$$(\varepsilon) \xrightarrow{*}_G w(\text{alph}(u'u''), q_0 * z)(\text{alph}(w'_1), q_0 * z'_1)$$

$$\cdots (\text{alph}(w_1), q_0 * z_1) \cdots (\text{alph}(w_k), q_0 * z_k)),$$

with $u'u'' \xrightarrow{*}_S z$, $\forall t \in [1, p-1]$, $w'_t \xrightarrow{*}_S z'_t$, and $\forall t \in [1, k]$, $w_t \xrightarrow{*}_S z_t$.

Since $aw'_1 \cdots w'_{p-1} w_1 \cdots w_k \xrightarrow{*}_S w'_1 \cdots w'_{p-1} w_k \cdots w_k a$, we have

$$(\varepsilon) \xrightarrow{*}_G wa((\text{alph}(u'u''a), q_0 * za)(\text{alph}(w'_1), q_0 * z'_1)$$

$$\cdots (\text{alph}(w_k), q_0 * z_k)).$$

(It is clear that this non-terminal is in N.)

Furthermore, $u'au'' \in R$ and $u'au'' \xrightarrow{*}_S u'u''a \xrightarrow{*}_S za$, hence $za \in f(R) = R$; Similarly, $\forall t \in [1, p-1]$, $w'_t \in R$ and $w'_t \xrightarrow{*}_S z'_t \in f(R) = R$.

So we are allowed to apply successively p rewriting rules of the kind ④ according to G, thus we get

$$(\varepsilon) \xrightarrow{*}_G wa((\text{alph}(w_1), q_0 * z_1) \cdots (\text{alph}(w_k), q_0 * z_k)),$$

which ends the proof of Lemma 9. ∎

LEMMA 10. *Let* $(\varepsilon) \xrightarrow{*}_G w((X_1, q_1)(X_2, q_2) \cdots (X_k, q_k))$ *be a derivation in* G. *Then, we can find* k *words* $w_1, \dots, w_k$ *such that, for each* i *in* $[1, k]$, $q_i = q_0 * w_i$, *and* $X_i = \text{alph}(w_i)$, *and* k *words* $u_1, \dots, u_k$ *in* R^* *which verify: for each* i *in* $[1, k]$, $\text{alph}(u_i) \subseteq \text{alph}(w_i \cdots w_k)$ *and* $w \xrightarrow{*}_{S^{-1}} u_1 w_1 u_2 w_2 \cdots u_k w_k$.

Proof. The proof is by induction on the length L of the derivation in G. If $L = 0$, the result is obvious.

For $L = l + 1$, with $l \geqslant 0$, we can write

$$(\varepsilon) \xrightarrow[G]{l} vA \xrightarrow[G]{*} w((X_1, q_1) \cdots (X_k, q_k)) \quad \text{with} \quad v \in X^*, A \in N.$$

We now discuss the last derivation. There are two cases:

Case 1. $v = w$, that is, $A \to ((X_1, q_1) \cdots (X_k, q_k))$ is a production of the grammar G. Then the nonterminal A is either equal to

(a) $((Y, q_F)(X_1, q_1) \cdots (X_k, q_k))$ with $q_F \in F$ or

(b) $((X_1, q_1) \cdots (X_i, q_1)(Y, q_F)(X_{i+1}, q_{i+1}) \cdots (X_k, q_k))$, where $i \geqslant 1$, $q_F \in F$, and $Y \subseteq \bigcup_{t > i} X_t$.

If $A = ((Y, q_F)(X_1, q_1) \cdots (X_k, q_k))$ with $q_F \in F$, by applying the induction hypothesis to the derivation $(\varepsilon) \Rightarrow_G^l wA$, we can write: $w \xrightarrow{*}_{S^{-1}} u_1' v_1 u_1'' w_1 u_2 w_2 \cdots u_k w_k$, where

- $u_1', u_1'', u_2, \ldots, u_k \in R^*$
- $\forall i \geqslant 2$, $\text{alph}(u_i) \subset \text{alph}(w_i \cdots w_k)$ and $\text{alph}(u_1'') \subset \text{alph}(w_1 \cdots w_k)$
- $\text{alph}(v_1) = Y$, $q_F = q_0 * v_1$ and, $\forall i \in [1, k]$, $\text{alph}(w_i) = X_i$, $q_0 * w_i = q_i$.

Since $q_F \in F$, v_1 is in R and, if we note $u_1 = u_1' v_1 u_1''$, we have $w \xrightarrow{*}_{S^{-1}} u_1 w_1 u_2 w_2 \cdots u_k w_k$ and this factorization verifies the conditions of the lemma.

Let us consider, now, the subcase 1(b). Then

$$A = ((X_1, q_1) \cdots (X_i, q_i)(Y, q_F)(X_{i+1}, q_{i+1}) \cdots (X_k, q_k)).$$

The induction hypothesis applied to the derivation $(\varepsilon) \Rightarrow_G^l wA$ gives us $w \xrightarrow{*}_{S^{-1}} u_1 w_1 \cdots u_i w_i u_{i+1}' v' u_{i+1}'' w_{i+1} \cdots u_k w_k$ with

- $\text{alph}(v') = Y$, $q_0 * v' = q_F$, and $\forall j \in [1, k]$, $X_j = \text{alph}(w_j)$ and $q_j = q_0 * w_j$.
- $\forall j \in [1, k]$, $j \neq i+1$, $u_j \in R^*$; $u_{i+1}' \in R^*$, $u_{i+1}'' \in R^*$.
- $\forall j \in [2, i]$, $\text{alph}(u_j) \subset \text{alph}(w_j \cdots w_i v' w_{i+1} \cdots w_k)$; $\text{alph}(u_{i+1}') \subset \text{alph}(v' w_{i+1} \cdots w_k)$ and $\text{alph}(u_{i+1}'') \subset \text{alph}(w_{i+1} \cdots w_k)$; $\forall j \in [i+2, k]$, $\text{alph}(u_j) \subset \text{alph}(w_j \cdots w_k)$.

Since $q_F \in F$, $v' \in R$. Thus, if we set $u_{i+1} = u_{i+1}' v' u_{i+1}'' \in R^*$, we get $w \xrightarrow{*}_{S^{-1}} u_1 w_1 \cdots u_i w_i u_{i+1} w_{i+1} \cdots u_k w_k$, and, clearly, this factorization verifies the conditions of the lemma.

Case 2. If $w = va$, with $a \in X$, there are, again, two subcases:

(a) $A = ((X_1, q_1) \cdots (X_i', q_i') \cdots (X_k, q_k))$ with $X_i = X_i' \cup \{a\}$, $q_i = q_i' * a$ and $A \to a((X_1, q_1) \cdots (X_i, q_i) \cdots (X_k, q_k))$ is a production of G. Then, $\forall t \in [i+1, k]$, $\forall y \in X_t$, $ay \to ya \in P$.

From the induction hypothesis applied to the derivation $(\varepsilon) \Rightarrow_G^l vA$, we get $v \xrightarrow{*}_{S^{-1}} u_1 w_1 \cdots u_i w_i' u_{i+1} \cdots u_k w_k$, where $X_i' = \text{alph}(w_i')$, $q_i' = q_0 * w_i'$, and

- $\forall j \in [1, k]$, $u_j \in R^*$, and if $j > 2$, $\text{alph}(u_j) \subset \text{alph}(w_j \cdots w_i' \cdots w_k)$
- $\forall j \in [1, k]$, with $j \neq i$, $q_j = q_0 * w_j$, and $X_j = \text{alph}(w_j)$.

Then, $w = va \xrightarrow{*}_{S^{-1}} u_1 w_1 \cdots u_i w_i' u_{i+1} \cdots u_k w_k a$.

Since $\forall y \in \text{alph}(w_{i+1} \cdots w_k)$, $ay \to ya \in P$, and $\forall t \in [i+2, k]$, $\text{alph}(u_t) \subset \text{alph}(w_t \cdots w_k)$, we can deduce that $au_{i+1} w_{i+1} \cdots u_k w_k \xrightarrow{*}_S u_{i+1} w_{i+1} \cdots u_k w_k a$. So, $w \xrightarrow{*}_{S^{-1}} u_1 w_1 \cdots u_i w_i' a u_{i+1} \cdots u_k w_k$.

If we note $w_i = w_i' a$, it is easy to see that $\forall j \in [1, k]$, $u_j \in R^*$, $q_j = q_0 * w_j$, $\text{alph}(w_j) = X_j$, and if $j > 2$, $\text{alph}(u_j) \subset \text{alph}(w_j \cdots w_k)$; hence the result.

(b) The second possibility is to have $A = ((X_1, q_1) \cdots (X_{i-1}, q_{i-1})(X_{i+1}, q_{i+1}) \cdots (X_k, q_k))$, where $i \geqslant 1$ and $A \to a((X_1, q_1) \cdots (X_k, q_k))$ is a production of G, that is, $X_i = \{a\}$, $q_i = q_0 * a$, and $\forall y \in \bigcup_{t \geqslant i+1} X_t$, $ay \to ya \in P$.

Since $(\varepsilon) \Rightarrow_G^l vA$, we can write $v \xrightarrow[S^{-1}]{*} u_1 w_1 \cdots u_{i-1} w_{i-1} u_{i+1} w_{i+1} \cdots u_k w_k$, where $\forall j \in [1, k]$, with $j \neq i$, $u_j \in R^*$, $q_j = q_0 * w_j$, $X_j = \text{alph}(w_j)$, and, if $j > 2$, $\text{alph}(u_j) \subset \text{alph}(w_j \cdots w_{i-1} w_{i+1} \cdots w_k)$. Then

$$\begin{aligned} w = va &\xrightarrow[S^{-1}]{*} u_1 w_1 \cdots u_{i-1} w_{i-1} u_{i+1} w_{i+1} \cdots u_k w_k a \\ &\xrightarrow[S^{-1}]{*} u_1 w_1 \cdots u_{i-1} w_{i-1} a u_{i+1} w_{i+1} \cdots u_k w_k, \end{aligned}$$

since $\bigcup_{t \geqslant i+1} X_t = \text{alph}(w_{i+1} \cdots w_k) = \text{alph}(u_{i+1} w_{i+1} \cdots u_k w_k)$.

If we note $u_i = \varepsilon$ and $w_i = a$, it is clear that this derivation verifies the conditions of the lemma, which ends the proof of Lemma 10. ∎

We are now able to finish the proof of Proposition 8. First, we shall prove that $f(R^*) \subseteq L(G, (\varepsilon))$. Let w be a word of $f(R^*)$. We can find a word u of R^* such that $w \xrightarrow{*}_{S^{-1}} u$. It is obvious that this derivation verifies the four conditions of the Lemma 9. Thus, we can deduce from this lemma that $(\varepsilon) \xrightarrow{*}_G w(\varepsilon)$ and, since $(\varepsilon) \Rightarrow \varepsilon$ is a derivation according to G, we get $(\varepsilon) \xrightarrow{*}_G w$, hence w is in $L(G, (\varepsilon))$.

For the reverse inclusion, let w be a word of $L(G, (\varepsilon))$. Then, there is a derivation according to G: $(\varepsilon) \xrightarrow{*}_G w(\varepsilon) \Rightarrow_G w$.

From Lemma 10, we can find a word u of R^* such that $w \xrightarrow{*}_{S^{-1}} u$, hence w is in $f(u) \subset f(R^*)$. ∎

We have to note that the condition of Proposition 8 is not a necessary condition. For example, if we take $R = \{a, b, ab\}$ and f the semi-commutation associated to the system $S = \langle \{a, b\}, \{ba \to ab\} \rangle$, it is clear that $f(R) = R$, and that the subgraph $(\Gamma(S), \text{alph}(ab))$ is not strongly connected. But $f(R^*) = \{a, b\}^*$ is obviously a regular language.

However, in the particular case where $f(R) = f(w)$ for some word w, we get

PROPOSITION 11. *Let f be the semi-commutation fonction associated to the system $S = \langle X, P \rangle$ and let w be a word of X^*. Then, $f(w^*)$ is a regular language if and only if the subgraph $(\Gamma(S)$, alph$(w))$ is strongly connected.*

Proof. Let us assume that the subgraph $(\Gamma(S), \text{alph}(w))$ is not strongly connected. Then, we can find two letters x and y such that no path is directed from x to y.

Set $E_1 = \{z \in \text{alph}(w)/\text{there is a path from } x \text{ to } z\}$ and $E_2 = \{z \in \text{alph}(w)/\text{there is no path from } x \text{ to } z\}$.

It is clear that $\{E_1, E_2\}$ is a partition of alph(w). Furthermore, there is no edge from an element s of E_1 to an element t of E_2. Thus, $\forall x \in E_1$, $\forall y \in E_2$, $xy \to yx \in P$.

For $i \in [1, 2]$, set $\pi_i = \text{proj}(X, E_i)$ and $w_i = \pi_i(w)$. Since $w \xrightarrow{*}_S w_2 w_1$ and $w_1 w_2 \xrightarrow{*}_S w_2 w_1$, $f(w^*) \cap w_2^* w_1^* = \{w_2^n w_1^n / n \geqslant 0\}$, which is a non-regular language since w_1 and w_2 are non-empty words over disjoint alphabets. ∎

It would be interesting to extend the above proposition by giving a decidable necessary and sufficient condition concerning a semi-commutation f and a finite set F which ensures that $f(F^*)$ is a regular language. Note that in the particular case where f is the total commutation, noted com, there is an answer, since it is known that for a finite set F with alph$(F) = X$, com(F^*) is regular if and only if for each $x \in X$, $F \cap x^+ \neq \emptyset$.

RECEIVED February 1985; ACCEPTED July 1986

REFERENCES

AALBERSBERG, I. J., AND ROZENBERG, G. (1985), "Theory of Traces," Rep. No. 85–16, University of Leiden.

CARTIER, P., AND FOATA, D. (1969), "Problèmes combinatoires de commutation et réarrangements," Lecture Notes in Mathematics 85, Springer-Verlag, Berlin.

CLERBOUT, M. (1984), "Commutations partielles et familles de langages," Thesis, Lille.

CLERBOUT, M., AND LATTEUX, M. (1984), Partial commutations and faithful rational transductions, *Theoret. Comput. Sci.* **34**, 241–254.

CORI, R., AND METIVIER, Y. (1985), Recognizable subsets of some partially abelian monoïds, *Theor. Comput. Sci.* **35**, 179–189.

CORI, R., AND PERRIN, D. (1985), Automates et commutations partielles, *RAIRO Inform. Théor.* **19**, 21–32.

FLE, M. P., AND ROUCAIROL, G. (1982), On serializability of iterated transactions, *in* "ACM Sigact–Sigops," pp. 194–200.

LOTHAIRE, M. (1983), "Combinatorics on Words," Addison-Wesley, Reading, MA.

MAZURKIEWICZ, A. (1977), "Concurrent Program Schemes and Their Interpretations," DAIMI, PB 78, Aarhus University.