JOURNAL OF COMPUTER AND SYSTEM SCIENCES 7, 583–596 (1973)

# Two Iteration Theorems for Some Families of Languages

L. BOASSON

*Institut de Programmation, Université de Paris, Paris V, France*

Received September 10, 1971; revised March 5, 1973

The "full Abstract Families of Languages," abbreviated full AFL's, were introduced by S. Ginsburg and S. Greibach in [11]. It is well known that the family of context-free languages is a full AFL [9]. It is also a rational cone (according to Eilenberg's terminology) [5]. Let $D_n{}^*$ (respectively, $D_n'^*$) be the Dyck language (respectively, semi-Dyck language) on $2n$ letters, $\sigma_i$, $\bar{\sigma}_i$, $i = 1,..., n$, i.e., the class of 1 in the congruence generated by $\sigma_i\bar{\sigma}_i = \bar{\sigma}_i\sigma_i = 1$ (respectively, $\sigma_i\bar{\sigma}_i = 1$) [9–14]. The Chomsky–Schützenberger theorem [9] implies that, for any $n \geqslant 2$, $D_n{}^*$ (respectively, $D_n'^*$) is a full generator [10–11] of the AFL as well as of the rational cone of the context-free languages. It seems, then, natural to look at the rational cones and the AFL's generated by the Dyck languages $D_1{}^*$ and the semi-Dyck language $D_1'^*$ on two letters.

We proved elsewhere [2] that the rational cone $\mathscr{C}$ generated by $D_1'^*$ can be characterized by the structure of the pushdown automata recognizing the languages in $\mathscr{C}$. The main restriction we impose on them is that they should use a single pushdown symbol. Since we can consider the pushdown store as a counter, we call such an automaton a "one-counter automaton" and we call one-counter languages the elements of $\mathscr{C}$. It is rather important to notice that this family $\mathscr{C}$ is not the family $\mathscr{F}$ studied by Greibach in [13]. However, $\mathscr{F}$ and $\mathscr{C}$ are closely related: $\mathscr{F}$ is the full AFL generated by $D_1'^*$ [13]. Consequently, from a result of [11] restated in [3], $\mathscr{F}$ is the closer of $\mathscr{C}$ under union, product, and star operation.

In this paper, we prove two pumping lemmas (Theorems 3 and 4) which yield corollaries such as:

*Neither $D_1{}^*$, nor $D_1'^*$ can be mapped one on the other by any rational transduction* (= *mapping performed by an a-transducer* [11]).

*The full* AFL*'s they generate are different and strictly included in the* AFL *of context-free languages.*

In the first part of this paper, we recall some notations, terminology, and properties we use about Dyck languages and rational transductions. In the second, after further definitions, we state our two theorems and prove several corollaries. The proof of these two theorems is given in a third part.

The main results of this paper have already been given, without any proof, in [1].

## I. Preliminaries

Using Eilenberg's terminology [5–7], we call *rational languages* the regular languages. Let $\Sigma_n = \{\sigma_i, \bar{\sigma}_i \mid i = 1,..., n\}$ a $2n$-letter alphabet. Let 1 be the empty word of $\Sigma_n{}^*$, the free monoid generated by $\Sigma_n$. We define the congruence $\theta$ (respectively $\theta'$) on $\Sigma_n{}^*$ as the Thue's congruence generated by the relations $\sigma_i\bar{\sigma}_i = \bar{\sigma}_i\sigma_i \equiv 1$ (respectively, $\sigma_i\bar{\sigma}_i \equiv 1$) for $i = 1,..., n$. It is well known that $\Sigma_n{}^*/\theta$ is the free group generated by $\Sigma_n$ where $\sigma_i$ and $\bar{\sigma}_i$ are the inverse of each other. On the other hand, $\Sigma_n{}^*/\theta'$ is the polycyclic monoid generated by $\Sigma_n$ [15], which generalizes the bicyclic monoid [4] $\Sigma_1{}^*/\theta'$. The *Dyck language*, $D_n{}^*$, is the class of 1 mod $\theta$, and the *semi-Dyck language*, $D_n'^*$, is the class of 1 mod $\theta'$. So, what we call the Dyck language is the kernel of the canonical homomorphism of $\Sigma_n{}^*$ to $\Sigma_n{}^*/\theta$. It is well known (see [14], for instance) that $D_n{}^*$ and $D_n'^*$ are both context-free, non-ambiguous, and that they are two free submonoids of $\Sigma_n{}^*$, respectively, generated by $D_n$ and $D_n'$ where $D_n$ (respectively, $D_n'$) is the set of words in $D_n{}^*$ (respectively, $D_n'^*$) which have no proper non-empty left factor in $D_n{}^*$ (respectively, in $D_n'^*$).

An homomorphism is called *alphabetical* iff the image of one letter under it is either a letter or the empty word.

Any application from the free monoid $X^*$ in the parts of the free monoid $Y^*$ will be called a *transduction* of $X^*$ to $Y^*$. Such a transduction $\tau$ is said to be *rational* iff $\hat{\tau} = \{(f, g) \mid f \in X^* \text{ and } g \in \tau f\}$ is a rational part of $X^* \times Y^*$ [5–14]. The rational transductions have been introduced (as binary transductions) by Elgot and Mezei [8] and studied (as $K$-transductions) by Nivat [14]. They are precisely those mappings performed by an $a$-transducer of [11]. These rational transductions can be used because of the following characterization ([14, 11]):

Nivat's Theorem 1. *A transduction $\tau$ from $X^*$ to $Y^*$ is rational iff there exists a (finite) set $Z$, a rational language $R$ on $Z$, and two alphabetical homomorphisms $\varphi$ and $\psi$ from $Z^*$ to $X^*$ and $Y^*$, respectively, such that:*

$$\forall f \in X^*, \qquad \tau f = \psi(\varphi^{-1} f \cap R).$$

Using Eilenberg's terminology [5], we say a family $\mathscr{L}$ of languages is *a rational cone* iff $\mathscr{L}$ is closed under rational transductions. Let us point out that this notion is slightly different from the "*full semi-AFL*" one [13] because such a rational cone may not be closed under union; in the particular case in which the cone is principal, it is a full semi-AFL; at any rate, we shall still call it a rational cone as the union closure will not have much to do there.

Let $\mathscr{L}$ be a family of languages; then Rat $\circ$ $\mathscr{L}$ is the closure of $\mathscr{L}$ under union product, and star operation, and $\mathscr{C}(\mathscr{L})$ is the rational cone, $\hat{\mathscr{F}}(\mathscr{L})$ the full AFL generated by $\mathscr{L}$. As a consequence of the fact that the rational transductions are closed under composition [8],

$$\mathscr{C}(\mathscr{L}) = \{L' \mid L' = \tau L, L \in \mathscr{L} \text{ and } \tau \text{ is a rational transduction}\}.$$

Theorem 1 implies that a family $\mathscr{L}$ is a full AFL iff it is a rational cone closed under union, product, and star operation; therefore, we have the following [11, 10, 3]:

THEOREM 2. *Let $\mathscr{L}$ be any family of languages, then*

$$\hat{\mathscr{F}}(\mathscr{L}) = \mathrm{Rat} \circ \mathscr{C}(\mathscr{L}).$$

Finally, let CFL be the family of context-free languages; we can, then deduce from the Chomsky–Schutzenberger theorem

PROPOSITION 1 [3]. *For any integer $n \geqslant 2$,*

$$\hat{\mathscr{F}}(D_n{}^*) = \hat{\mathscr{F}}(D_n'^*) = \hat{\mathscr{F}}(D_n) = \hat{\mathscr{F}}(D_n')$$
$$= \mathscr{C}(D_n{}^*) = \mathscr{C}(D_n'^*) = \mathscr{C}(D_n) = \mathscr{C}(D_n') = \mathscr{C}(D_2'^*) = \mathrm{CFL}.$$

## II. THE MAIN RESULTS

We need further definitions before stating our two iteration theorems (Theorems 3 and 4).

For convenience, we will call *factor* of the word $f$ in $X^*$ an occurrence of a non-empty factor, i.e., a triplet $(f_1, f_2, f_3)$ in $X^* \times XX^* \times X^*$ such that $f = f_1f_2f_3$. The word $f$ is said to have *two disjoint factors* $(f_1, f_2, f_3)$ and $(f_1', f_2', f_3')$ iff there exists $h$ in $X^*$ such that $f_1' = f_1f_2h$ and $f_3 = hf_2'f_3'$ (and therefore $f = f_1f_2hf_2'f_3'$). We will say that $(f_1, f_2, f_3)$ is the first factor (instead of the leftmost one).

Let $L \subsetneq X^*$ be a language and $f$ a word in $L$. The factor $(f_1, f_2, f_3)$ of $f$ is an *iterative factor* of $f$ in $L$ iff:

$$n \geqslant 0 \to f_1f_2{}^nf_3 \in L.$$

Two disjoint factors $\alpha = (g_1, g_2, g_3)$, the first one, and $\beta = (g_1', g_2', g_3')$, the second one, constitute an *iterative pair* $\pi = (\alpha, \beta)$ of $f$ in $L$, iff:

(1) $f = g_1g_2hg_2'g_3'$,

(2) $n \geqslant 0 \to g_1g_2{}^nhg_2'^ng_3' \in L$.

The iterative pair $\pi$ is *strict* iff, for any integer $n$, the set of integers $m$ such that $g_1g_2{}^nhg_2'^mg_3' \in L$ or $g_1g_2{}^mhg_2'^ng_3' \in L$ is finite.

We define the $n$-th iteration of the pair $\pi$ as the pair $\pi_n = (\alpha_n, \beta_n)$ defined by $\alpha_n = (g_1, g_2^n, hg_2'^n g_3')$ and $\beta_n = (g_1 g_2^n h, g_2'^n, g_3')$. The word $f_n = g_1 g_2^n h g_2'^n g_3'$ is said to be deduced from $f$ by taking the $n$-th iteration of $\pi$.

It is well known that any long enough word $f$ of a context-free language $L$ has either an iterative factor or a (strict) iterative pair in $L$.

Still using the same notations, $\pi = (\alpha, \beta)$ is said to be a *pre-iterative pair* of $f = g_1 g_2 h g_2' g_3'$ in $L$ iff there exist two integers $\lambda$ and $\mu$ such that $\alpha(\lambda) = (g_1, g_2^\lambda, hg_2'^\mu g_3')$ and $\beta(\mu) = (g_2 g_2^\lambda h, g_2'^\mu, g_3')$ constitute an iterative pair $(\alpha(\lambda), \beta(\mu))$ of $g_1 g_2^\lambda h g_2'^\mu g_3'$ in $L$.

The word $f$ in $L$ has two *disjoint iterative pairs* in $L$ (abbreviated two *dips*) $\pi = (\alpha, \beta)$ and $\pi' = (\alpha', \beta')$ iff:

(1) $\pi$ and $\pi'$ are two strict iterative pairs of $f$ in $L$,

(2) the four factors $\alpha$, $\beta$, $\alpha'$, $\beta'$ are pairwise disjoint,

(3) the word you get by taking the $n$-th iteration of $\pi$ and the $m$-th of $\pi'$ is in $L$ for any $n$, $m \geqslant 0$.

We will say $\pi$ is the first of the two dips iff $\alpha$ is the first of the four factors $\alpha$, $\alpha'$, $\beta$, $\beta'$. Then, three cases arise: Let

$$f = g_1 u g_2 v g_3 u' g_4 v' g_5$$

and $\alpha = (g_1, u, g_2 v g_3 u' g_4 v' g_5)$,

$$t_1 = (g_1 u g_2, v, g_3 u' g_4 v' g_5),$$

$$t_2 = (g_1 u g_2 v g_3, u', g_4 v' g_5),$$

$$t_3 = (g_1 u g_2 v g_3 u' g_4, v', g_5).$$

1. $\pi$ and $\pi'$ are of type 1: iff $\beta = t_1$, $\alpha' = t_2$, and $\beta' = t_3$.
2. $\pi$ and $\pi'$ are of type 2: iff $\beta = t_2$, $\alpha' = t_1$, and $\beta' = t_3$.
3. $\pi$ and $\pi'$ are of type 3: iff $\beta = t_3$, $\alpha' = t_1$, and $\beta' = t_2$.

We denote the $f_{n,m}$ the word (in $L$) we get by taking the $n$-th iteration or the first iterative pair $\pi$ and the $m$-th of the second one $\pi'$, i.e.,

if $\pi$ and $\pi'$ are of type 1:

$$f_{n,m} = g_1 u^n g_2 v^n g_3 u'^m g_4 v'^m g_5 ;$$

if $\pi$ and $\pi'$ are of type 2:

$$f_{n,m} = g_1 u^n g_2 v^m g_3 u'^n g_4 v'^m g_5 ;$$

if $\pi$ and $\pi'$ are of type 3:

$$f_{n,m} = g_1 u^n g_2 v^m g_3 u'^m g_4 v'^n g_5 .$$

In any of those tree cases, $f_{n,m}$ admits two canonical dips $\pi_{n,m}$ and $\pi'_{n,m}$ in $L$; in the first case, we have:

$$\pi_{n,m} = (\alpha_{n,m}, \beta_{n,m}), \qquad \pi' = (\alpha'_{n,m}, \beta'_{n,m})$$

where

$$\alpha_{n,m} = (g_1 , u, u^{n-1} g_2 v^n g_3 u'^m g_4 v'^m g_5),$$

$$\beta_{n,m} = (g_1 u^n g_2 , v, v^{n-1} g_3 u'^m g_4 v'^m g_5),$$

$$\alpha'_{n,m} = (g_1 u^n g_2 v^n g_3 , u', u'^{m-1} g_4 v'^m g_5),$$

$$\beta'_{n,m} = (g_1 u^n g_2 v^n g_3 u'^m g_4 , v', v'^{m-1} g_5).$$

(The two other cases are handled in the same obvious way.)

We can now state our two theorems:

THEOREM 3. *Let $L$ be in $\mathscr{C}(D'^*)$. If $f$ is a word in $L$ which has two dips, $\pi = (\alpha, \beta)$ the first pair and $\pi' = (\alpha', \beta')$ the second one, there exists two integers $n, m \geqslant 1$ such that $(\alpha_{n,m} , \beta'_{n,m})$ is a pre-iterative pair of $f_{n,m}$ in $L$. Moreover, if $\pi$ and $\pi'$ are of type 3, $(\alpha'_{n,m} , \beta_{n,m})$ is a pre-iterative pair of $f_{n,m}$ in $L$ too.*

THEOREM 4. *Let $L$ be in $\mathscr{C}(D_1{}^*)$. If $f$ is a word in $L$ which has two dips $\pi = (\alpha, \beta)$ and $\pi' = (\alpha', \beta')$, there exist two integers $n$, $m \geqslant 1$ such that $f_{n,m}$ will admit four pre-iterative pairs in $L$ among the six pairs possible to build out of $\alpha_{n,m}$, $\beta_{n,m}$, $\alpha'_{n,m}$, and $\beta'_{n,m}$.*

*Remark.* Any language in $\mathscr{C}(D_1'^*)$ or in $\mathscr{C}(D_1{}^*)$ is obviously context-free. Bearing in mind a result of S. Ginsburg and E. H. Spanier [12], it may seem strange to consider the case of dips of type 2. However, the following example will show that this case may arise. It will also show that, in the case of such dips, we cannot get a new preiterative pair, different from that announced in Theorem 3. Let

$$L_1 = \{x^n y^m z^p t^q \mid n = n_1 + n_2 \text{ and } 3n_1 + n_2 = m + p + q\},$$
$$L_2 = \{x^n y^m z^p t^q \mid q = q_1 + q_2 \text{ and } 3q_1 + q_2 = n + m + p\},$$
$$L = L \cup L_2 .$$

It is easy to prove that $L \in \mathscr{C}(D_1'^*)$. Then $f = xyzt \in L$ admits two dips of type two: $\pi = ((1, x, yzt), (xy, z, t))$ and $\pi' = ((x, y, zt), (xyz, t, 1))$. Theorem 3 gives a pair of the form $((1, x, yzt), (xyz, t, 1))$. Remark that there is no pre-iterative pair of the form $((x, y, zt), (xy, z, t))$.

The proofs of Theorems 3 and 4 are given in the third part of this paper. Let us look at some corollaries first.

COROLLARY 1. (1) *$D_1{}^*$ is not image of $D_1'^*$ through any rational transduction.*

(2) *$D_1'^*$ is not image of $D_1{}^*$ through any rational transduction.*

(1) has already been proved by S. Eilenberg in a completely different manner [6].

1. Assume $D_1{}^* \in \mathscr{C}(D_1'^*)$. Consider then $f = \sigma\bar{\sigma}\bar{\sigma}\sigma \in D_1{}^*$. It has two dips $\pi = (\alpha, \beta)$ and $\pi' = (\alpha', \beta')$ where

$$\alpha = (1, \sigma, \bar{\sigma}^2\sigma), \qquad \beta = (\sigma, \bar{\sigma}, \bar{\sigma}\sigma),$$

$$\alpha' = (\sigma\bar{\sigma}, \bar{\sigma}, \sigma), \qquad \beta' = (\sigma\bar{\sigma}^2, \sigma, 1),$$

According to Theorem 3, there exist $n$ and $m$ such that $f_{n,m} = \sigma^n\bar{\sigma}^{n+m}\sigma^m$ admits $(\alpha_{n,m}, \beta'_{n,m})$ as a pre-iterative pair where

$$\alpha_{n,m} = (1, \sigma, \sigma^{n-1}\bar{\sigma}^{n+m}\sigma^m),$$

$$\beta'_{n,m} = (\sigma^n\bar{\sigma}^{n+m}, \sigma, \sigma^{m-1}).$$

This is obviously impossible since any word in $D_1{}^*$ must have as many occurrences of $\sigma$ as of $\bar{\sigma}$.

2. Assume now $D_1'^* \in \mathscr{C}(D_1{}^*)$ and consider $f = \sigma\bar{\sigma}\sigma\bar{\sigma} \in D_1'^*$. It has two dips $\pi = (\alpha, \beta)$ and $\pi' = (\alpha', \beta')$ where

$$\alpha = (1, \sigma, \bar{\sigma}\sigma\sigma), \qquad \beta = (\sigma, \bar{\sigma}, \sigma\bar{\sigma}),$$

$$\alpha' = (\sigma\bar{\sigma}, \sigma, \bar{\sigma}), \qquad \beta' = (\sigma\bar{\sigma}\sigma, \bar{\sigma}, 1).$$

According to Theorem 4, there exists $n, m \geqslant 1$ such that $f_{n,m} = \sigma^n\bar{\sigma}^n\sigma^m\bar{\sigma}^m$ has two pre-iterative pairs among the six we can get from $\alpha_{n,m}$, $\alpha'_{n,m}$, $\beta_{,nm}$, and $\beta'_{n,m}$. Obviously, we can build only three of them. Then, it appears clearly that $D_1'^*$ is not in $\mathscr{C}(D_1{}^*)$.

COROLLARY 2 [11]. *Neither $\mathscr{C}(D_1'^*)$ nor $\mathscr{C}(D_1{}^*)$ are full* AFL's.

Consider $L = \{a^nb^n \mid n \geqslant 1\}$ over $X = \{a, b\}$. It is obvious that $L \in \mathscr{C}(D_1{}^*) \cap \mathscr{C}(D_1'^*)$. We will prove that $L^2$ is neither in $\mathscr{C}(D_1{}^*)$ nor in $\mathscr{C}(D_1'^*)$. The corollary will then be established as neither of these two families will be closed under product. Let $f =$

$abab \in L^2$; $f$ has two dips in $L^2$, $\pi = (\alpha, \beta)$ and $\pi' = (\alpha', \beta')$ where

$$\alpha = (1, a, bab), \qquad \beta = (a, b, ab),$$

$$\alpha' = (ab, a, b), \qquad \beta' = (aba, b, 1),$$

Then, let us carry on exactly as in the proof of Corollary 1.

We now want to use Theorems 3 and 4 for comparing (with regard to inclusion) the six families $\mathscr{C}(D_1{}^*)$, $\mathscr{C}(D_1'{}^*)$, $\hat{\mathscr{F}}(D_1{}^*)$, $\hat{\mathscr{F}}(D_1'{}^*)$, the family of rational languages $\mathscr{R}$, and that of context-free languages CFL. Let us point out, first, that

$$\mathscr{R} \subsetneq \mathscr{C}(D_1{}^*) \cap \mathscr{C}(D'{}^*),$$

an obvious consequence of the beginning of the proof of Corollary 2. Let us now prove the

PROPOSITION 2. $\mathscr{C}(D_1'{}^*) = \mathscr{C}(D_1') = \mathscr{C}(D_1)$.

(Remember $D_1$ and $D_1'$ have been defined previously as the free generating sets of $D_1{}^*$ and $D_1'{}^*$, respectively).

The fact that $\mathscr{C}(D_1'{}^*) = \mathscr{C}(D_1'{}^*)$ is an obvious consequence of the fact that $D_1' = \{\sigma f \bar{\sigma} \mid f \in D_1'{}^*\}$. Therefore, we only have to prove that $\mathscr{C}(D_1) = \mathscr{C}(D_1')$:

(1) $\mathscr{C}(D_1) \supseteq \mathscr{C}(D_1')$ as $D_1' = D_1 \cap \Sigma_1{}^*\bar{\sigma}$.

(2) $\mathscr{C}(D_1) \subseteq \mathscr{C}(D_1')$ as $D_1 = D_1' \cup \tilde{D}_1'$ where $\tilde{D}_1' = \{\tilde{f} \mid f \in D_1'\}$.

PROPOSITION 3. $\hat{\mathscr{F}}(D_1{}^*) \subsetneq \hat{\mathscr{F}}(D_1'{}^*)$.

We know (through Proposition 2 ) that $D_1 \in \mathscr{C}(D_1'{}^*)$. Then, obviously $D_1{}^* \in \hat{\mathscr{F}}(D_1'{}^*)$; hence $\hat{\mathscr{F}}(D_1{}^*) \subseteq \hat{\mathscr{F}}(D_1'{}^*)$. In order to show that this inclusion is strict, we will prove that there is a language in $\mathscr{C}(D_1'{}^*)$ which is not in $\hat{\mathscr{F}}(D_1{}^*)$.

Let $X = \{a, b, c, d\}$ be an alphabet and

$$L = \{a^{n+m}b^m c^p d^{p+n} \mid n, m, p \geqslant 1\}$$

a language over $X$. It is easy to check that $L \in \mathscr{C}(D_1'{}^*)$. We want to show, first, that, if $L$ were in $\hat{\mathscr{F}}(D_1{}^*)$, there would be in $\mathscr{C}(D_1{}^*)$ a very similar language: according to Theorem 4, this is impossible.

Assume that $L$ is in this full AFL; according to Theorem 2, $L \in \mathrm{Rat} \circ \mathscr{C}(D_1{}^*)$. This means that $L$ is obtained by substituting some elements of $\mathscr{C}(D_1{}^*)$ to the letters of a rational language. Since $L$ has no iterative factor, this rational language must finite. That can be stated in the following way: $L$ is obtained from elements of $\mathscr{C}(D_1{}^*)$ by a finite number of unions and products. So we may write:

$$L = \bigcup_{i=1}^{r} L_i \text{ where } L_i = L_{i_1} \cdots L_{i_{n_i}} \text{ with } L_{i_j} \in \mathscr{C}(D_1{}^*).$$

Let us point out, first, that, if $x$ is any letter of $X$, $L_{i_j} \subset x^*$ implies that $L_{i_j}$ is finite, since no word in $L$ has any iterative factor.

Assume, then, that $L_{i_j}$ contains some occurrences of $c$; either any word of $L_{i_j}$ contains a bounded number of $c$'s, or $L_{i_j}$ contains some $d$'s. It is a consequence of the fact that, if a word in $L_{i_j}$ contains an arbitrary number of $c$'s without any $d$'s, it would contain an iterative factor, which is impossible. In the same way, if $L_{i_j}$ contains some $b$'s, it contains a bounded number or it contains some $a$'s. So we may now assume that one of the following cases arises:

(1) $L_{i_1} \subset aa^*bb^*$, $L_{i_2} \subset cc^*dd^*$.

(2) $L_{i_1} = L_i = aa^*bb^*cc^*dd^*$, $L_{i_2} = 1$.

In case (1), the difference between the number of $a$'s and $b$'s in words of $L$ cannot be bounded unless it is so in $L_i$. Therefore, there exists $i$ such that this difference is not bounded in $L_{i_1}$. Then, if we choose a fixed word in $L_{i_2}$, we build a word in $L_{i_1}L_{i_2} = L_i$ such that for this difference, arbitrarily large, between the number of $a$'s and $b$'s, the difference between the number of $c$'s and $d$'s is fixed. This word cannot be in $L$!

Hence, there exists an integer $M$ such that, whenever in a word of $L$ the difference between the number of $a$'s and $b$'s is larger than $M$, the word has to belong to an $L_i$ of type (2). The language obtained then is very similar to $L$ and is in $\mathscr{C}(D_1{}^*)$: it has to contain the words of $L$ where the above difference is larger than $M$. Such a language cannot be in $\mathscr{C}(D_1{}^*)$, as is easily proved by Theorem 4.

PROPOSITION 4. $\hat{\mathscr{F}}(D_1'^*) \subsetneq \hat{\mathscr{F}}(D_2'^*) = \mathrm{CFL}$.

The inclusion is a direct consequence of Theorem 2. The fact that it is strict can be proved in the same way as in Proposition 3 considering

$$L = \{a^n b^m c^m d^n \mid m, n \geqslant 1\}.$$

Corollary 1 and Propositions 1, 2, 3 can be summarized in

THEOREM 5.

$$\mathscr{R} \subsetneq \mathscr{C}(D_1{}^*) \cap \mathscr{C}(D_1'^*) \left\{ \begin{matrix} \subsetneq \mathscr{C}(D_1{}^*) \subsetneq \hat{\mathscr{F}}(D_1{}^*) \\ \subsetneq \mathscr{C}(D_1'^*) = \mathscr{C}(D_1') = \mathscr{C}(D_1) \end{matrix} \right\} \subsetneq \hat{\mathscr{F}}(D_1'^*) \subsetneq \mathrm{CFL}.$$

## III. Proofs of Theorems 3 and 4

The proof is rather long but it merely consists in checking the properties stated in the theorems. Letting $L$ in $\mathscr{C}(D_1^*)$ (respectively, in $\mathscr{C}(D_1'^*)$) we know there exists a rational transduction $\tau$ such that $L = \tau D_1^*$ (respectively, $L = \tau D_1'^*$). Now, according to Theorems 1, $L = \psi(\varphi^{-1}D_1^* \cap R)$ (respectively $L = \psi(\varphi^{-1}D_1'^* \cap R)$) where $R$ is a rational language.

We shall first give a characterization of the iterative pairs in $D_1'^*$ (Lemma 1) and in $D_1^*$ (Lemma 1′). Then, we shall show that some special iterative factors appear in $R$ (Lemmas 2, 3, 2′, 3′). Lemma 4 will show that, if $A$ is a context-free language, we can assume that the factors constituting iterative pairs in $A \cap R$ are iterative factors of $R$. Finally, we shall prove Theorems 3 and 4 in a particular case of which we shall deduce the results stated.

Lemma 1. *$\pi = (\alpha, \beta)$ is a strict iterative pair of $f$ in $D_1'^*$ iff $\alpha = (f_1, f_1, f_3)$ and $\beta = (g_1, g_2, g_3)$ are two disjoint factors of $f$ such that*

$$f_2 = \bar{\sigma}^{p}\sigma^{p'} (\text{mod } \theta') \qquad \textit{and} \qquad g_2 = \bar{\sigma}^{q}\sigma^{q'} (\text{mod } \theta')$$

*where $q - q' = p' - p > 0$.*

Lemma 1′. *$\pi = (\alpha, \beta)$ is a strict iterative pair of $f$ in $D_1^*$ iff $\alpha = (f_1, f_2, f_3)$ and $\beta = (g_1, g_2, g_3)$ are two dispoint factors of $f$ such that*

$$f_2 = x^{p} (\text{mod } \theta) \qquad \textit{and} \qquad g_2 = y^{q} (\text{mod } \theta)$$

*where $p = q \neq 0$ and $x, y \in \Sigma_1$, $x \neq y$.*

The proofs of these two lemmas are straightforward and well known.

Lemma 2. *Let $Z$ be a finite alphabet, $\varphi$ an alphabetical homomorphism from $Z^*$ to $\Sigma_1^*$, and $R$ a rational language over $Z^*$ such that $\forall n \in \mathbf{N}$, $\exists f \in R$ such that*

$$\varphi f = \bar{\sigma}^{q}\sigma^{q'} (\text{mod } \theta') \qquad \textit{with} \quad q - q' \geqslant n \ (\textit{respectively, } q' - q \geqslant n).$$

*There exists, then, a word $g$ in $R$ which has an iterative factor $(g_1, g', g_2)$ such that*

$$\varphi g' = \bar{\sigma}^{r}\sigma^{r'} (\text{mod } \theta')$$

*where $r - r' > 0$ (respectively, $r' - r > 0$).*

First, let us point out that $h \in \Sigma_1^*$ and $h = \bar{\sigma}^{q}\sigma^{q'} (\text{mod } \theta')$ implies $|h| \geqslant q + q'$; on the other hand, $h = \varphi f$ implies $|h| \leqslant |f|$ since $\varphi$ is an alphabetical homomorphism.

Let $N(R)$ be the integer associated to $R$ such that $f \in R$ and $|f| \geqslant N(R)$ implies $f$ has an iterative factor in $R$.

Throughout this hypothesis, there exists $f$ in $R$ such that $\varphi f = h \equiv \bar{\sigma}^q \sigma^{q'}$ with $q - q' \geqslant N(R)$ (respectively, $q' - q \geqslant N(R)$). Let $f$ be such a word, the shortest possible. Then $|f| \geqslant |h| \geqslant q + q' \geqslant N(R)$. So that $f$ has an iterative factor in $R$, i.e., $f = f_1 f_2 f_3$ and $f_1 f_2{}^* f_3 \subset R$.

Assume $\varphi f_2 \equiv \bar{\sigma}_1{}^{r_1} \sigma_1^{r'}$ with $r_1 - r_1{}' \leqslant 0$ (respectively, $r_1{}' - r_1 \leqslant 0$). Then $f_1 f_3 \in R$ has the same property as $f$ and is shorter, which is impossible. Therefore, $(f_1, f_2, f_3)$ is an iterative factor which has the announced property. In the same way, we can prove:

LEMMA 2'. *Let $Z$ be a finite alphabet, $\varphi$ an alphabetical homomorphism from $Z^*$ to $\Sigma_1{}^*$, and $R$ a rational language over $Z$ such that $\forall n \in \mathbf{N}$, $\exists f \in R$ such that $\varphi f \equiv x^q (\mathrm{mod}\ \theta)$ $x \in \Sigma_1$ and $q \geqslant n$. Then, there exists a word $g$ in $R$ which has an iterative factor $(g_1, g', g_2)$ such that*

$$\varphi g' \equiv x^r (\mathrm{mod}\ \theta) \qquad r > 0.$$

LEMMA 3. *Let $Z$ be a finite alphabet, $\varphi$ an alphabetical homomorphism from $Z^*$ to $\Sigma_1{}^*$, and $R$, an infinite rational language over $Z$ such that $\exists M \in \mathbf{N}$ such that $f \in R$, $\varphi f \equiv \bar{\sigma}^q \sigma^{q'} (\mathrm{mod}\ \theta') \to |q - q'| \leqslant M$. Then, there exists a word $g$ in $R$ which has an iterative factor $(g_1, g', g_2)$ where*

$$\varphi g' \equiv \bar{\sigma}^p \sigma^p (\mathrm{mod}\ \theta').$$

$R$ being an infinite language, there exists a word $g$ in $R$ which has an iterative factor $(g_1, g', g_2)$. Assume now $\varphi g' \equiv \bar{\sigma}^r \sigma^{r'} (\mathrm{mod}\ \theta')$. Then, we have $\varphi(g'^n) \equiv (\bar{\sigma}^r \sigma^{r'})^n \equiv \bar{\sigma}^{r+(n-1)(r-r')} \sigma^{r'}$ if $r > r'$. Then $(g_1 g'^n g_2) \equiv \varphi g_1 \cdot \bar{\sigma}^{r+(n-1)(r-r')} \sigma^{r'} \cdot \varphi g_2 \equiv \bar{\sigma}^p \sigma^{p'}$ where $p - p' = (n-1)(r - r') + (q - q')$ if we suppose $\varphi g \equiv \bar{\sigma}^q \sigma^{q'}$. This is not compatible with the hypothesis on $R$. Obviously, the case in which $r < r'$ is shown to be impossible in the same way. So, the lemma is proved.

In exactly the same way, we get:

LEMMA 3'. *Let $Z$ be a finite alphabet, $\varphi$ an alphabetical homomorphism from $Z^*$ to $\Sigma_1{}^*$ and $R$ an infinite rational language over $Z$ such that $\exists M \in \mathbf{N}$ such that $f \in R$, $\varphi f \equiv x^q (\mathrm{mod}\ \theta)$ $x \in \Sigma_1 \to q < n$. Then, there exists a word $g$ in $R$ which has an iterative factor $(g_1, g', g_2)$ where*

$$\varphi g' \equiv 1 (\mathrm{mod}\ \theta).$$

LEMMA 4. *Let $L$ be a context-free language and $R$ a rational one over $X$. If the word $f$ in $L \cap R$ has an iterative pair $\pi$ in $L$, there exists $n$, $m \geqslant 1$ such that the iterative pairs $\pi_m$ of $f_n$ in $L \cap R$ is made of two iterative factors in $R$.*

We have $f = f_1uf_2vf_3$ and $f_p = f_1u^pf_2v^pf_3 \in L \cap R$. Let $q$ be the number of states of the minimal automation that recognizes $R$ and consider $f_q$, there exists $p_1, p_2, p_3, p_1', p_2', p_3' \geqslant 0$, such that

(1) $p_1 + p_2 + p_3 = p_1' + p_2' + p_3' = q$,

(2) $g_{\lambda,\mu} = f_1u^{p_1}u^{\lambda p_2}u^{p_3}f_2v^{p_1'}v^{p_2'\mu}v^{p_3'}f_3 \in R$ $\lambda, \mu \geqslant 0$.

Choosing now $\lambda = p_2' + 1$ and $\mu = p_2 + 1$, we get

$$g_{p_2'+1,p_2+1} = f_1u^{q+p_2p_2'}f_2v^{q+p_2'p_2}f_3 = f_{q+p_2p_2'}$$

and we check that

$$\pi_{p_2p_2'} = ((f_1u^q, u^{p_2p_2'}, f_2v^{q+p_2'p_2}f_3), (f_1u^{q+p_2p_2'}f_2v^q, v^{p_2'p_2}, f_3))$$

is an iterative pair of $f_{q+p_2p_2'}$ in $L \cap R$ which is made of two iterative factors of $f_{q+p_2p_2'}$ in $R$.

Let us start the proof of Theorem 3. We do not give it for Theorem 4 since it is exactly the same, apart from the fact that Lemmas 1′, 2′, 3′ must be used instead of Lemmas 1, 2, 3.

First, let us look at a particular language $L'$ in $\mathscr{C}(D_1'^*)$ such that:

(1) $L' \subset x_1x_2{}^*x_3x_4{}^*x_5x_6{}^*x_7x_8{}^*x_9$, $x_i \in X$, $i = 1,\dots, 9$,

(2) any word in $L'$ has no iterative factor in $L'$,

(3) $f = x_1x_2x_3x_4x_5x_6x_7x_8x_9 \in L'$ and has two dips in $L'$.

Since $L' \in \mathscr{C}(D_1'^*)$, we have, according to Theorem 1, a finite alphabet $Z$, a rational language $R$ over $Z$, and two alphabetical homomorphism $\varphi$ and $\psi$ from $Z^*$ in $\Sigma_1{}^*$ and $X^*$, respectively, such that:

$$L' = \psi(\varphi^{-1}D_1'^* \cap R).$$

Let $Z_0 = \{z \in Z \mid \psi z = 1\}$ and, for $i = 1,\dots, 9$, $Z_i = \{z \in Z \mid \psi z = x_i\}$. First notice that we can assume that $z \in Z_0$ implies $\varphi z \neq 1$ [3]. Let $A = \varphi^{-1}(D_1'^*) \cap R$; so $L' = \psi A$. From hypothesis (1), we may assume that

$$R \subseteq Z_0{}^*Z_1(Z_2 \cup Z_0)^*\, Z_3(Z_4 \cup Z_0)^*\, Z_5(Z_6 \cup Z_0)^*\, Z_7(Z_8 \cup Z_0)^*\, Z_9Z_0{}^*.$$

(a) Then, let us show that we may assume

LEMMA 5. *Any word $h$ in $R$ is such that any of its factors in $Z_0{}^*$ is length-bounded.*

According to a result of Ogden [16], there exists $N \geqslant 1$ such that, if $h \in A$ has more than $N$ consecutive letters in $Z_0$, it can be factorized in $h = h_1h_2h_3h_4h_5$ where either

$h_2$ or $h_4$ is in $Z_0{}^*$ and $h_1 h_2{}^n h_3 h_4{}^n h_5 \in A$, for any $n \geqslant 0$. Then $g = \psi h \in L'$, As $\psi h_1 h_2{}^n h_3 h_4{}^n h_5 \in L'$, it comes through hypothesis (2) that $\psi h_2 = \psi h_4 = 1$. Then $g = \psi h_1 h_3 h_5$. And, if we intersect $R$ with the words of $Z^*$ which never have more than $N$ consecutive letters in $Z_0{}^*$, we get a rational language $R'$ such that:

$$L' = \psi(\varphi^{-1} D_1'^* \cap R').$$

The lemma is thus proved.

(b) Let $m \geqslant 1$, and if the two dips are of type 1, 2, or 3, let $A_m$ be, respectively:

$$A_m = A \cap \psi^{-1}(x_1 x_2{}^* x_3 x_4{}^* x_5 x_6{}^m x_7 x_8{}^m x_9),$$

$$A_m = A \cap \psi^{-1}(x_1 x_2{}^* x_3 x_4{}^m x_5 x_6{}^* x_7 x_8{}^m x_9),$$

$$A_m = A \cap \psi^{-1}(x_1 x_2{}^m x_3 x_4{}^* x_5 x_6{}^* x_7 x_8{}^m x_9).$$

Consider now $h_0(m)$, $h'_0(m) \in Z_0{}^*$, for $i = 2, 4, 6, 8$, $h_i(m) \in (Z_i \cup Z_0)^*$ and for $i = 1, 3, 5, 7, 9$, the letters $Z_i(m) \in Z_i$. Let then

$$h(m) = h_0(m)\, z_1(m)\, h_2(m)\, z_3(m)\, h_4(m)\, z_5(m)\, h_6(m)\, z_7(m)\, h_8(m)\, z_9(m)\, h_1'(m)$$

be a word in $A_m$. According to Lemma 5, we may assume $h_0(m)$ and $h_0'(m)$ are length bounded. Since $\varphi(h(m))$ is in $D_1'^*$, $\varphi(h_8(m)\, z_9(m)\, h_0'(m))$ is a right factor of $D_1'^*$; so there exists an integer $q(m)$ such that

$$\varphi(h_8(m)\, z_9(m)\, h'_0(m)) \equiv \bar{\sigma}^{q(m)} (\mathrm{mod}\ \theta').$$

Let us show now that $q(m)$ cannot be bounded when $m$ increases. According to Lemma 2, we shall get from this a special iterative factor in $R$.

Assume that there exists an integer $q$ such that $\forall m_0 \in \mathbf{N}$, $\exists m \geqslant m_0$ such that $q(m) \leqslant q$. Since $\psi(h_8(m)) = x_8{}^m$, the length of $h_8(m)$ increases with $m$. So, for a large enough integer $m$, $h_8(m)$ will admit an iterative factor $h_8'(m)$ in $R$, such that $\varphi(h_8'(m)) \equiv \bar{\sigma}^p \sigma^p$. (This is a consequence of Lemma 3 applied to the factors of $R$ in $(Z_8 \cup Z_0)^*$.) According to Lemma 5, $\psi(h_8'(m)) \neq 1$ is an iterative factor of $\psi(h(m))$ in $L$, which is in contradiction with hypothesis (2). Hence, $\forall q \in \mathbf{N}$, $\exists m_0 \in \mathbf{N}$ such that

$$m \geqslant m_0 \rightarrow q(m) > q.$$

Lemma 2 can be applied in exactly the same conditions as above to get an integer $m$ such that $h_8(m)$ has an iterative factor $h_8'(m)$ in $R$ where

$$\varphi(h_8'(m)) \equiv \bar{\sigma}^r \sigma^{r'} (\mathrm{mod}\ \theta') \qquad \text{with } r - r' > 0.$$

Consider then $A_{m_0}$. Let $h \in A_{m_0}$ be a long enough word. We know [16] that it will have an iterative pair in $A_{m_0}$. Now, according to hypothesis (2), this iterative pair is strict. So, we have $h = h_1'h_2'h_3'h_4'h_5'$ where $h_2'$ and $h_4'$ are the occurrences of the factors constituting the iterative pair of $h$ in $A_{m_0}$. From Lemma 1, we get $\varphi h_2' \equiv \bar{\sigma}^s \sigma^{s'}$ (mod $\theta'$) where $s' - s > 0$. Moreover, Lemma 4 shows that we may assume that $h_2'$ and $h_4'$ are occurrences of iterative factors of $h$ in $R$. Then we, can write $h(m_0) = h = uh_2'vh_8'w$ where $h_8' = h_8'(m_0)$ is the iterative factor of $h(m_0)$ we built before. Then, since we have $\varphi h_2' \equiv \bar{\sigma}^s \sigma^{s'}$ where $s' - s > 0$ and $\varphi h_8' \equiv \bar{\sigma}^r \sigma^{r'}$ where $r - r' > 0$, we have $g = uh_2'^{r-r'} vh_8'^{s'-s} w \in A(m_0)$ and $g \in A$.

It is easy to check that $\varphi h_2'^{r-r'} \equiv \bar{\sigma}^{t_1} \sigma^{t'_1}$ and $\varphi h_8'^{s'-s} \equiv \bar{\sigma}^{t_2} \sigma^{t'_2}$ with $t_1' - t_1 = t_2' - t_2 > 0$. Now, according to Lemma 1, $\varphi h_2'^{r'-r}$ and $\varphi h_8'^{s'-s}$ are two occurrences of factors which constitute an iterative pair of $\varphi g$. Then, in $h$, $h_2'$ and $h_8'$ constitute a pre-iterative pair and it is the same thing in $\psi h$. Since $\psi h_2' \in x_2{}^*$ and $\psi h_8' \in x_8{}^*$, the first part of Theorem 3 is proved for such a language as $L'$.

Now, if $\pi$ and $\pi'$ are of type 3, we have to build up another iterative pair. Let us proceed in the same way with the left factor $h_0(m)\, z_1(m)\, h_2(m)$ of $h(m)$. Again, if $\varphi(h_0(m)\, z_1(m)\, h_2(m)) \equiv \sigma^{q'(m)}$, $q'(m)$ is not bounded when $m$ increases. Thus, we know there is an iterative factor $h_2''(m)$ such that $\varphi(h_2''(m)) \equiv \bar{\sigma}^t \sigma^{t'}$(mod $\theta'$) where $t' - t > 0$. Then we can write $h(m_0) = h = u'h_2''v'h_4'w'$ and the proof goes on.

Let us extend the proof to any $L$ in $\mathscr{C}(D_1'^*)$. Assume $L$ is a language over $Y$ and let $f$ have two dips in $L$. We have $f = f_1 f_2 f_3 f_4 f_5 f_6 f_7 f_8 f_9$ where $f_{2i} \neq 1$ for $i = 1, 2, 3, 4$. Let $\lambda$ be the homomorphism from $X^*$ in $Y^*$ given by $\lambda x_i = f_i$, $i = 1,..., 9$. Let now $L' = \lambda^{-1} L \cap x_1 x_2{}^* x_3 x_4{}^* x_5 x_6{}^* x_7 x_8{}^* x_9$. Thus, we have $L' \in \mathscr{C}(D_1'^*)$ for which we check hypotheses (1), (2), and (3). Besides, we have $f = \lambda g$ implies $f_{n,m} = \lambda g_{n,m}$, and the pre-iterative pairs we want for $f$ are obtained from the one we have just built for $g$. Theorem 3 is thus proved.

The proof of Theorem 4 is the same as the one we have just given for Theorem 3.

## References


1. L. Boasson, An iteration theorem for one-counter languages, Third Annual ACM Symposium on Theory of Computing (1970), 116–120.
2. L. Boasson, Une caractérisation algébrique des languages à compteur (to appear).
3. L. Boasson and M. Nivat, Sur les images par transductions de familles de langages, *Acta Informatica* **2** (1973), 180–188.
4. A. H. Clifford and G. B. Preston, "The Algebraic Theory of Semigroups," Vol. 1, American Mathematical Society, Providence, R. I., 1961.
5. S. Eilenberg, Algebraic aspects of automata theory, *Actes Congr. Intern. Math. Nice 1970* **3**, 265–267 (Gauthier-Villars, Paris, 1971).
6. S. Eilenberg, Private communication.
7. S. Eilenberg and M. P. Schutzenberger, Rational sets in commutative monoids, *J. Algebra* **13** (1969), 173–191.
8. C. C. Elgot and J. E. Mezei, On relations defined by generalized finite automata, *IBM J. Res. Develop.* **9** (1962), 47–48.

9. S. GINSBURG, "The Mathematical Theory of Context-Free Languages," McGraw-Hill, New York, 1966.
10. S. GINSBURG AND S. GREIBACH, Principal AFL, *J. Comput. System Sci.* **4** (1970), 308–338.
11. S. GINSBURG, S. GREIBACH, AND J. HOPCROFT, "Studies in Abstract Families of Languages" (Memoirs of the American Mathematical Society No. 87), American Mathematical Society, Providence, R. I., 1969.
12. S. GINSBURG AND E. H. SPANIER, Bounded algol-like languages, *Trans. Amer. Math. Soc.* **113** (1966), 285–296.
13. S. GREIBACH, An infinite hierachy of context free languages, *J. Assoc. Comput. Mach.* **16** (1969), 91–106.
14. M. NIVAT, Transduction des languages de Chomsky, *Ann. Inst. Fourier (Grenoble)* **18** (1968), 339–455.
15. M. NIVAT AND J. F. PERROT, Une généralisation du monoïde bicyclique, *C. R. Acad. Sci. Paris* **271**, Ser. A (1970), 824–827.
16. W. OGDEN, An helpful result for proving inherent ambiguity, *Math. Systems Theory* **2** (1967), 191–194.