JOURNAL OF COMPUTER AND SYSTEM SCIENCES **35**, 311–320 (1987)

On Commutative Context-Free Languages*

J. BEAUQUIER

Université de Paris-Sud, 91405 Orsay, France

M. BLATTNER†

University of California, Davis, and
Lawrence Livermore National Laboratory, Livermore, California 94550

AND

M. LATTEUX

The University of Lille I, C.N.R.S. U.A. 369, France

Received July 1, 1985; revised June 5, 1986

Let $\Sigma = \{a_1, a_2, ..., a_n\}$ be an alphabet and let $L \subset \Sigma^*$ be the commutative image of FP^* where F and P are finite subsets of Σ^*. If, for any permutation σ of $\{1, 2, ..., n\}$, $L \cap a^*_{\sigma(1)} \cdots a^*_{\sigma(n)}$ is context-free, then L is context-free. This theorem provides a solution to the Fliess conjecture in a restricted case. If the result could be extended to finite unions of the FP^* above, the Fliess conjecture could be solved.

1. INTRODUCTION

One of the goals of formal language theory is to discover properties of families of languages, that is, properties that each member of the family will have. Two examples of such structural properties in the context-free languages are: the semilinearity of languages under the Parikh mapping and the pairwise iteration of substrings in words as described by the pumping lemma. It has been shown that languages that do not have these properties are not context-free.

In this paper we are concerned with structural characteristics of the family of commutative context-free languages. The general problem, known as the Fliess conjecture, has been open since 1970 [1]. The conjecture is: given a commutative language L over the alphabet $\Sigma = \{a_1, a_2, ..., a_n\}$, for every permutation σ of

* This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7505-Eng-48 and the P.R.C. Mathématiques et Informatique.

† Also affiliated with the Department of Biomathematics, M. D. Anderson Cancer Research Hospital, University of Texas Medical Center, Houston, Texas, 77030.

0022-0000/87 $3.00

$1, 2, \ldots, n$, if $L \cap a^*_{\sigma(1)} a^*_{\sigma(2)} \cdots a^*_{\sigma(n)}$ is context-free, then L is context-free. The problem has been solved for alphabets with one, two, or three letters [5, 6, 8]; it remains unresolved for alphabets of four or more letters. Note that the hypothesis implies that L has a semilinear Parihk's image $L = \mathrm{com}(L) = \mathrm{com}(F_1 P_1^* \cup \cdots \cup F_r P_r^*)$, where F_i's and P_i's are finite subsets of Σ^*. In the present paper, we solve the problem for an arbitrary alphabet, with the supplementary assumption that $L = \mathrm{com}(FP^*)$, F and P finite.

2. Preliminaries

We assume the reader to be familiar with the fundamental notions of formal languages theory, as they can be found in [2, 3 or 4]. The notation we shall use is that if Σ is a finite alphabet, Σ^* is the free monoid generated by Σ with the empty word e. We will write a^* instead of $\{a\}^*$, for a in Σ, and Σ^+ is $\Sigma^* - \{e\}$. For a word w in Σ^*, $|w|$ is the *length of* w, and for a in Σ, $|w|_a$ the number of occurrences of a in w, alph(w) is the subset of letters in Σ appearing in w. A language $L \subset \Sigma^*$ is *bounded* if there are words $w_1, w_2, \ldots, w_k$ such that $L \subset w_1^* w_2^* \cdots w_k^*$. Let $a = \langle a_1, a_2, \ldots, a_n \rangle$, the mapping Ψ_a or Ψ, if the alphabet is understood, is a function from Σ^* to N^n is defined by $\Psi(w) = (|w|_{a_1}, \ldots |w|_{a_n})$. Let $\Psi(L) = \{\Psi(w) \mid w \in L\}$. Ψ is called a *Parikh mapping*. Another useful function is f_a, which maps members of N^n to Σ^*, where $f_a(i_1, i_2, \ldots, i_n) = a_1^{i_1} a_2^{i_2} \cdots a_n^{i_n}$.

For w in Σ^* and P contained in Σ^* we will write wP instead of $\{w\}\ P$. Com(w) is the set of all permutations of symbols in w, while com(P) is the *commutative closure* of P; that is, $\mathrm{com}(P) = \{\mathrm{com}(w) \mid w \in P\}$. A set of the form $\{\alpha_0 + n_1\alpha_1 + \cdots + n_m\alpha_m \mid n_j \geqslant 0 \text{ for } 1 \leqslant j \leqslant m\}$, where $\alpha_0, \ldots, \alpha_m$ are elements of N^n, is a *linear* subset of N^n, denoted $L(\alpha_0, \{\alpha_1, \ldots, \alpha_m\})$. A *semilinear set* is a finite union of linear sets. It is known that $\Psi(L)$ is semilinear for any context-free language L (Parikh's theorem) [7]. A subset X of N^n is said to be *stratified* if the following two conditions are satisfied: (i) each element in X has at most two nonzero coordinates; (ii) there are no integers i, j, k, m, and $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$ in X such that $1 \leqslant i < j < k < m \leqslant n$ and $x_i x'_j x_k x'_m \neq 0$. The symbol ‡ is used for the shuffle operator. The notation concerning pushdown automata, their moves and their computations, is quite similar to that in the literature.

In order to prove our results, three lemmas from [2] are required. These are stated below for future reference.

Lemma A. *Let $\Sigma = \{a_1, a_2, \ldots, a_n\}$ and $a = \langle a_1, a_2, \ldots, a_n \rangle$. If $W \subset a_1^* a_2^* \cdots a_n^*$ is a context-free language, then $f_a^{-1}(W)$ is a finite union of linear sets, each with a stratified set of periods.*

Lemma B. *Let $L(c; \{q\})$ and $L(d; S)$ be linear sets of N^n such that $L(c; \{q\}) \cap L(d; S)$ is infinite. Then there exists a positive integer k such that kq in $L(0; S)$.*

Lemma C. *If X and Y are linear subsets of N^n, then $X \cap Y$ is semilinear and effectively calculable from X and Y.*

3. Results

Our problem can be stated more simply using the notion of a B-CF language.

Definition. A language L over the alphabet $\Sigma = \{a_1, a_2, ..., a_n\}$ is said to be a *B-CF language* if and only if for every permutation π of $[1, n]$, $L \cap a_{\pi(1)}^* \cdots a_{\pi(n)}^*$ is a context-free language. Then the Fliess conjecture can be stated as follows:

A commutative language is context-free if and only if it is a B-CF language.

Since the Parikh image of a context-free language is semilinear [7], it is sufficient to consider the commutative languages of the form $\mathrm{com}(w_1 P_1^* \cup \cdots \cup w_k P_k^*)$ where, for i in $[1, k]$, w_i is a word and P_i is a finite language. The aim of this paper is to prove the Fliess conjecture in a nontrivial particular case.

Theorem 1. *Let L be equal to* $\mathrm{com}(FP^*)$, *where F and P are finite languages. Then L is a context-free language if and only if L is a B-CF language.*

Remark. Note that the case considered here is slightly more general then the case of context-free languages having a linear Parikh image, since here the set F is not necessarily reduced to a single word.

In order to establish this result, we will prove three lemmas and introduce some notation. Let $\Sigma = \{a_1, a_2, ..., a_n\}$ and $R = \bigcup_{1 \leq i \leq j \leq n} a_i^* a_j^*$. We may assume that F and P are subsets of $a_1^* a_2^* \cdots a_n^*$ since $P' = \mathrm{com}(P) = \mathrm{com}(P' \cap a_1^* a_2^* \cdots a_n^*)$. Let $Q = P \cap R$ and, for i, j in $[1, n]$, $Q_{i,j} = \mathrm{com}(Q) \cap a_i^* a_j^*$ and $Q'_{i,j} = \mathrm{com}(Q) \cap a_i^+ a_j^+$.

Lemma 2. *Let F and P be finite sets, F nonempty, included in $a_1^* a_2^* \cdots a_n^*$ such that $L = \mathrm{com}(FP^*)$ is a B-CF language. Then, P satisfies the two following properties:*

(A) *For all u in P, $u^+ \cap \mathrm{com}(Q^*) \neq \varnothing$.*

(B) *For all u in $Q'_{i,j}$, for all v in $Q'_{s,t}$, where a_i, a_j, a_s, a_t are distinct letters, $(uv)^+ \cap (\mathrm{com}(Q - Q'_{i,j})^* \cup \mathrm{com}(Q - Q'_{s,t})^*) \neq \varnothing$.*

Proof. First, we will prove Property (A). Since $L \cap a_1^* a_2^* \cdots a_n^*$ is assumed to be context-free, it follows from Lemma A that:

(1) $L = \bigcup_{i=1}^m \mathrm{com}(w_i P_i^*)$, for some integer m, where for all i in $[1, m]$, w_i in $a_1^* a_2^* \cdots a_n^*$ and $P_i \subset R$ is finite.

Let u be a word in P, $u \neq e$. Choose any w in F, from (1) it follows that there exists an integer j such that: $wu^* \cap \mathrm{com}(w_j P_j^*)$ is infinite. Then $\Psi(wu^+ \cap \mathrm{com}(w_j P_j^*))$ is infinite. Set $c = \Psi(w)$, $q = \Psi(u)$, $d = \Psi(w_j)$, and $S = \Psi(P_j)$. Since $\mathrm{com}(w_j P_j^*) = \Psi^{-1}(L(d; S))$, we get $\Psi(wu^* \cap \Psi^{-1}(L(d; S))) = \Psi(wu^*) \cap L(d; S) = L(c; \{q\}) \cap L(d; S)$. From Lemma B, there exists a positive integer t such that tq in

$L(0; S)$. Hence, $\Psi(u' \cap \mathrm{com}(P_j^*)) = \Psi(u') \cap \Psi(P_j^*) = tq \cap L(O; S) \neq \emptyset$, which implies that u' in $\mathrm{com}(P_j^*)$).

Now, let v be any word in P_j. From (1) we know that $w_j v^* \subset w_j P_j \subset L$, then $w_j v^* \cap \mathrm{com}(FP^*)$ is infinite and there exists a positive integer l such that v^l in $\mathrm{com}(P^*)$. But, v in $P_j \subset R$ and v in $a_p^* a_q^*$ for some p, q in $[1, n]$. Thus, v^l in $\mathrm{com}(P^*) \cap \{a_p, a_q\}^* = \mathrm{com}((P \cap a_p^* a_q^*)^*) \subset \mathrm{com}(Q^*)$. Since P_j is finite, that implies that there exists a positive integer h such that for every w in P_j, w^h is in $\mathrm{com}(Q^*)$. Let $P_j' = \{w^h \mid w \in P_j\}$, then, $P_j'^* \subset \mathrm{com}(Q^*)$ and u^{th} in $\mathrm{com}(P_j'^*) \subset \mathrm{com}(Q^*)$, hence $u^+ \cap \mathrm{com}(Q^*) \neq \emptyset$.

Next, we will prove Property B. Let π be a permutation of $[1, n]$ such that $\pi(1) = i$, $\pi(2) = s$, $\pi(3) = j$ and $\pi(4) = t$. By hypothesis, $L' = L \cap a_{\pi(1)}^* a_{\pi(2)}^* \cdots a_{\pi(n)}^*$ is context-free. By the definition of $\mathrm{com}(L')$ and Lemma A, $\mathrm{com}(L') = L = \bigcup_{k=1}^{m} \mathrm{com}(w_k P_k^*)$, where for every k in $[1, m]$, w_k in $a_1^* \cdots a_n^*$ and $P_k \subset R$ is a finite set such that: either $P_k \cap a_i^+ a_j^+ = \emptyset$ or $P_k \cap a_s^+ a_t^+ = \emptyset$. We choose any w in F. Since $w(uv)^* \subset L$, $w(uv)^* \cap \mathrm{com}(w_k P_k^*)$ is infinite for some k in $[1, m]$. Thus $(uv)^l$ is in $\mathrm{com}(P_k^*)$ for some positive integer l. One can assume, for instance, that $P_k \cap a_i^+ a_j^+ = \emptyset$. Now by using Property A, we can prove that there exists a positive integer h such that $P_k' = \{w^h \mid w \in P_k\} \subset \mathrm{com}(Q^*)$. But $P_k \subset R - a_i^+ a_j^+$; hence, $P_k' \subset \mathrm{com}(Q - Q_{i,j}')^*$ and $(uv)^{hl} \subset \mathrm{com}(Q - Q_{i,j}')^*$, which implies Property B. ∎

Lemma 2 may be illustrated by the following examples; let $\Sigma = \{a_1, a_2, a_3, a_4\}$.

EXAMPLE 1. Let $P = Q \cup \{a_1 a_2 a_3\}$ with $Q = \{a_1^2 a_2, a_1 a_4^3, a_2 a_3^2, a_3 a_4^3\}$. Property A of Lemma 2 holds because $(a_1 a_2 a_3)^2$ is in $\mathrm{com}((a_1^2 a_2)(a_2 a_3^2)) \subset \mathrm{com}(Q^*)$. Then $\mathrm{com}(P^*) = \mathrm{com}(FQ^*)$ with $F = \{e, a_1 a_2 a_3\}$. Property B is also satisfied because

$$-(a_1^2 a_2 a_3 a_4^3)^2 \text{ in } \mathrm{com}((a_1^2 a_2)(a_1 a_4^3)^2 (a_2 a_3^2)) \subset \mathrm{com}(Q - Q_{3,4}')^*,$$

$$-(a_1 a_4^3 a_2 a_3^2)^2 \text{ in } \mathrm{com}((a_1^2 a_2)(a_2 a_3^2)(a_3 a_4^3)^2) \subset \mathrm{com}(Q - Q_{1,4}')^*.$$

EXAMPLE 2. Let $P = \{a_1^2 a_2, a_1 a_2^2, a_1 a_3, a_1 a_4^3, a_2 a_3, a_3^2 a_4^3\}$. Since $P = Q$, it remains to be shown that property B is satisfied. But

$$-a_1^2 a_2 a_3^2 a_4^3 \text{ in } \mathrm{com}((a_1 a_3)(a_1 a_4^3)(a_2 a_3)) \subset \mathrm{com}(Q - Q_{1,2}')^*,$$

$$-a_1 a_2^2 a_3^2 a_4^3 \text{ in } \mathrm{com}((a_1 a_4^3)(a_2 a_3)^2) \subset \mathrm{com}(Q - Q_{1,2}')^*,$$

$$-(a_1 a_4^3 a_2 a_3)^2 \text{ in } \mathrm{com}((a_1 a_2^2)(a_1 a_4^3)(a_3^2 a_4^3)) \subset \mathrm{com}(Q - Q_{2,3}')^*.$$

EXAMPLE 3. Given $P = \{a_1 a_2^2, a_1 a_3, a_2 a_3, a_2 a_4, a_3^2 a_4\}$ then $a_1 a_2^2 a_3^2 a_4$ is in $\mathrm{com}((a_1 a_3)(a_2 a_3)(a_2 a_4)) \subset \mathrm{com}(Q - Q_{1,2}')^*$. But, $(a_1 a_3 a_2 a_4)^+ \cap (\mathrm{com}(Q - Q_{1,3}')^* \cup \mathrm{com}(Q - Q_{2,4}')^*) = \emptyset$. This comes from the fact that u in $\mathrm{com}(Q - Q_{1,3}')^*$ implies $|u|_{a_2} \geqslant 2\,|u|_{a_1}$ and v in $\mathrm{com}(Q - Q_{2,4}')^*$ implies $|v|_{a_3} \geqslant 2\,|v|_{a_4}$. From the preceding Lemma, $L = \mathrm{com}(P^*)$ is not a B-CF language. (It can also be verified that $L \cap a_1^* a_3^* a_2^* a_4^*$ is context-free but $L \cap a_1^* a_2^* a_3^* a_4^*$ is not context-free.)

If a finite set $Q \subset R$ satisfies Property B, it will be shown that $L = \mathrm{com}(Q^*)$ is context-free. First, we will show that Property B implies another property that will be crucial for our proof. For that we need a new definition.

DEFINITION. A word w in L is said to be *z-reducible*, for some z in Q, if and only if there exists a word u in L such that uz in $\mathrm{com}(w)$.

LEMMA 3. *Let Q be a finite set where $Q \subset R$ and $L = \mathrm{com}(Q^*)$. If Q satisfies Property B of Lemma 2, then it also satisfies the following property:*

(C) *There exists a positive integer k such that for all u in $Q'_{i,j}$, for all v in $Q'_{s,t}$ where a_i, a_j, a_s, a_t are distinct letters, $(uv)^k$ is z-reducible for some z in $Q_{i,s} \cup Q_{j,s}$.*

Proof. Let us consider u in $Q'_{i,j}$ and v in $Q'_{s,t}$ where a_i, a_j, a_s, a_t are distinct letters. The hypothesis implies that there is a positive integer k such that either:

(1) $(uv)^k$ in $\mathrm{com}(Q^*_{i,j} Q^*_{i,s} Q^*_{i,t} Q^*_{j,s} Q^*_{j,t})$ or

(2) $(uv)^k$ in $\mathrm{com}(Q^*_{i,s} Q^*_{i,t} Q^*_{j,s} Q^*_{j,t} Q^*_{s,t})$.

In case (1) the occurrences of symbols a_s must necessarily be produced by $Q^*_{i,s}$ or $Q^*_{j,s}$. Then, in this case, there exists a word z in $Q_{i,s} \cup Q_{j,s}$ such that $(uv)^k$ is z-reducible.

In case (2), there are two subcases. If $(uv)^k$ is z-reducible, for some z in $Q_{i,s} \cup Q_{j,s}$ (case 2.1) and we are done; otherwise, it is easy to show that $(uv)^k$ is in $\mathrm{com}(Q'^{+}_{i,t} Q'^{+}_{s,t} Q'^{+}_{j,t})$ (case 2.2).

We set $Q'_{s,t} = \{v_1, v_2, ..., v_p\}$ with

$$\frac{|v_1|_{a_t}}{|v_1|_{a_s}} \leqslant \frac{|v_2|_{a_t}}{|v_2|_{a_s}} \leqslant \cdots \leqslant \frac{|v_p|_{a_t}}{|v_p|_{a_s}}.$$

In order to prove the property in case (2.2), we will use an induction on the index q of the v_q's. First, we will prove that, if $v = v_1$, case (2.2) cannot occur. That will require showing that $(uv_1)^k$ is z-reducible for some z in $Q_{i,s} \cup Q_{j,s}$. Assume that $(uv_1)^k$ in $\mathrm{com}(w_1 w_2 w_3)$ with w_1 in $Q'^{+}_{i,t}$, w_2 in $Q'^{+}_{s,t}$ and w_3 in $Q'^{+}_{j,t}$. Since $(uv_1)^k$ is not z-reducible for some z in $Q_{s,s}$, w_2 in $\mathrm{com}(Q'_{s,t} \cup Q_{t,t})^*$ and from the choice of v_1,

$$\frac{|w_2|_{a_t}}{|w_2|_{a_s}} \geqslant \frac{|v_1|_{a_t}}{|v_1|_{a_s}}.$$

Thus, we get a contradiction, since

$$|v_1^k|_{a_t} = |(uv)^k|_{a_t} = |w_1|_{a_t} + |w_2|_{a_t} + |w_3|_{a_t} > |w_2|_{a_t}$$

and

$$|v_1^k|_{a_s} = |(uv_1)^k|_{a_s} = |w_2|_{a_s}.$$

Now, we define inductively, $m_1 = k$ and $m_q = (k+1)^{m_{q-1}}$ for $q > 1$. We make the following induction hypothesis: for $1 \leqslant q < l$, l a given integer in $[2, p]$, $(uv_q)^{m_{l-1}}$ is z-reducible, for some z in $Q_{i,s} \cup Q_{j,s}$. This hypothesis holds for $l = 2$. If the word $(uv_l)^k$ is z-reducible, for some z in $Q_{i,s} \cup Q_{j,s}$, then the same holds for $(uv_l)^{m_l}$ and the induction is extended. Otherwise, the remark made for u_1 can be used again. This time, it leads to the conclusion that $(uv_l)^k$ is v_q-reducible for some q in $[1, l-1]$. Then, $(uv_l)^{k+1}$ is (uv_q)-reducible and $(uv_l)^{m_l}$ is $(uv_q)^{m_{l-1}}$-reducible. Since, from the induction hypothesis, $(uv_q)^{m_{l-1}}$ is itself z-reducible for some z in $Q_{i,s} \cup Q_{j,s}$, the same holds for $(uv_l)^{m_l}$ and the induction is extended.

Since Q is finite, that implies property C. ▮

Let us consider, now a finite set $Q \subset R$ satisfying condition C. We will build a pushdown automaton (pda) that recognizes $L = \mathrm{com}(Q^*)$. This pda works non-deterministically, trying to reduce an input word to the empty word, using the set of congruences $\{w \equiv e \mid w \in Q\}$. Clearly, such a pda cannot work in "realtime" and reading something it must sometimes wait to read something more, before performing a useful reduction (useful in the sense that the reduction not lead to a deadlock preventing us from going further in reducing the input word to the empty word). At this step, Lemma 3 plays a crucial role, since it ensures that it is unnecessary to wait more than a fixed number of symbols of the same type, before performing a useful reduction. This waiting time is simulated in the pda by the finite set of states, each state being considered as a word of length less than or equal to a fixed integer.

LEMMA 4. *Let Q be a finite set included in R satisfying Property* C *of Lemma* 3. *Then $L = \mathrm{com}(Q^*)$ is a context-free language.*

Proof. Let k be the integer given by Lemma 3. We set $N = p(1 + nk)$ where $p = \sum_{u \in Q} |u|$. We define the pda $M = \langle S, \Sigma, \Gamma, s_0, z_0, \delta, F \rangle$ where:

(1) $S = \{u \in \Sigma^* \mid \text{for all } a \in \Sigma,\ |u|_a \leqslant N\}$ is the set of states;

(2) $\Sigma = \{a_1, a_2, ..., a_n\}$;

(3) $\Gamma = \Sigma \cup \{z_0\}$;

(4) $s_0 = e$ is the initial state;

(5) $F = \{s_0\}$ is the set of final states;

(6) z_0 is the bottom-of-store symbol;

(7) δ, the transition function, is defined by the allowed moves:

   (a) Stack, for all u in S, for all a in Σ, for all γ in Σ^*, $(u, a, z_0\gamma) \vdash^s (u, e, z_0\gamma a)$

   (b) Pop, for all γ in Σ^*, for all u in S with $|u|_a < N$, $(u, e, z_0\gamma a) \vdash^p (ua, e, z_0\gamma)$;

   (c) Reduce, for all γ in Σ^*, for all u in S, $(u, e, z_0\gamma) \vdash^r (u', e, z_0\gamma)$ if and only if there exists a z in Q such that $u'z$ in $\mathrm{com}(u)$.

Let $T(M)$ be the language recognized by M by empty store (reinitializations of the store are allowed); that is, $T(M) = \{w \in \Sigma^* \mid (e, w, z_0 \overset{*}{\vdash} (e, e, z_0)\}$. We will prove that $T(M) = L$.

Clearly, one can establish by induction on the length of the computation in M that, if $(e, w, z_0) \overset{*}{\vdash} (u, e, z_0\gamma)$, then there exists a word v in Q^* such that w in com-$(vu\gamma)$. Thus, if w in $T(M)$, $(e, w, z_0) \overset{*}{\vdash} (e, e, z_0)$ and w in com(v) for some v in Q^*; hence w in com$(Q^*) = L$ and so $T(M) \subset L$. In order to prove the other inclusion, we must first prove the following claim:

Claim. Let w_1, w_2 be words in Σ^* such that $w_1 w_2$ is in L. Then there exists a computation in M: $(e, w_1, z_0) \overset{*}{\vdash} (y, e, z_0\gamma)$ that satisfies the following four conditions:

(1) $w = yw_2\gamma$ in L;

(2) for all a_i, a_j in alph(γ), for all u in $Q_{i,j}$, w is not u-reducible;

(3) if $\gamma = \gamma' a_j$ then $|y|_{a_j} = N$;

(4) for all a_i in alph(γ), $|y|_{a_i} \geqslant r_i(w)\, kp$, where $r_i(w) = \text{card}\{ j \in [1, n] \mid$ there exists a u in $Q_{i,j}$ such that w is u-reducible$\}$.

Proof of the claim. By induction on the length of w_1.

Basis. $|w_1| = 0$. Then, w_2 in L and taking $y = \gamma = e$, we get $(e, w_1, z_0) \overset{*}{\vdash} (y, e, z_0)$, $w = yw_2\gamma$ in L and, since alph$(\gamma) = \varnothing$, conditions 2, 3, and 4 are satisfied.

Inductive step. We make the induction hypothesis that the claim holds for any w_1 such that $|w_1| < q$ (q a given positive integer) and let w_1, $|w_1| = q$, be a word such that $w_1 w_2$ in L. Setting $w_1 = w_1' a_k$, a_k in Σ, and $w_2' = a_k w_2$, we have from the induction hypothesis that there is a computation: $(e, w_1', z_0) \overset{*}{\vdash} (y, e, z_0\gamma)$ satisfying conditions 1 to 4 with $w = yw_2'\gamma$.

We will now distinguish between two main cases.

Case 1. $|y|_{a_k} < N$. Then $(e, w_1' a_k, z_0) \overset{*}{\vdash} (y, a_k, z_0\gamma) \vdash^{s} (y, e, z_0\gamma a_k) \vdash^{p} (ya_k, e, z_0\gamma)$. From the inductive hypothesis, it is easy to verify that conditions 1 to 4 hold for that computation.

Case 2. $|y|_{a_k} = N$. At this step. We must distinguish between subcases.

Subcase 1. a_k in alph(γ). Then $(e, w_1' a_k, z_0) \overset{*}{\vdash} (y, a_k, z_0\gamma) \vdash^{s} (y, e, z_0\gamma a_k)$. Since alph$(\gamma a_k) = $ alph(γ), it is easy to verify, from the induction hypothesis, that conditions 1 to 4 hold for that computation.

Subcase 2. a_k is not in alph(γ).

Subcase 2.1. There exists a u in $Q_{k,k}$ such that w is u-reducible. Since $|y|_{a_k} = N \geqslant |u|$, there exists a y_1 such that $y_1 u$ in com(y). Then

$$(e, w_1' a_k, z_0\gamma) \overset{*}{\vdash} (y, a_k, z_0\gamma) \vdash^{s} (y, e, z_0\gamma a_k) \vdash^{r} (y_1, e, z_0\gamma_k) \vdash^{p} (y_1 a_k, e, z_0\gamma).$$

Note that pop is allowed since $|y_1|_{a_k} < |e|_{a_k} = N$. Since $w = yw'_2\gamma$ in L is u-reducible, $y_1 a_k w_2 \gamma = y_1 w'_2 \gamma$ is in L and condition 1 holds. Moreover, it is easy to verify that conditions 2 to 4 are satisfied since a_k is not in alph(γ).

Subcase 2.2. For all u in $Q_{k,k}$, w is not u-reducible.

Subcase 2.2.1. For all a_i in alph(γ) and all u in $Q_{i,k}$, w is not u-reducible. Then, clearly the computation $(e, w'_1 a_k, z_0) \overset{*}{\vdash} (y, a_k, z_0\gamma) \vdash^{s} (y, e, z_0\gamma a_k)$ satisfies conditions 1 to 4.

Subcase 2.2.2. $\gamma = \gamma' a_j$ and there exists a u in $Q_{j,k}$ such that w is u-reducible. From Condition 3, we get $|y|_{a_j} = N$. Now, since $|y|_{a_j} = |y|_{a_k} = N$ there exists a word y_1 such that $y_1 u$ in com(y). Then $(e, w'_1 a_k, z_0) \overset{*}{\vdash} (y, a_k, z_0\gamma) \vdash^{s} (y, e, \gamma a_k) \vdash^{r} (y_1, e, \gamma a_k) \vdash^{p^*} (y_1 a_k a_j, e, \gamma')$. Clearly, condition 1 is satisfied. Since $y_1 a_k a_j w_2 \gamma'$ v-reducible implies $yw'_2\gamma$ v-reducible, condition 2 holds. Since $|y|_{a_j} = N = p(kn+1)$ and $|u|_{a_j} < |u| < p$, clearly, $|y_1|_{a_j} \geqslant pkn$ implying $|y_1 a_k a_j|_{a_j} \geqslant r_j(y_1 a_k a_j w_2 \gamma')\, kp$ and condition 4 is satisfied. Now if condition 3 does not hold, it suffices to repeat the pop operation until condition 3 is satisfied, the other conditions remaining true.

Subcase 2.2.3. $\gamma = \gamma' a_j$, for all u in $Q_{j,k}$, w is not u-reducible and there exists a_i in alph(γ) and v in $Q_{i,k}$ such that w is v-reducible.

First, we will prove that w is not v^k-reducible. Let us suppose the contrary and let w' be a word such that $w'v$ in com(w). Since $|w'|_{a_j} = |w|_{a_j} \geqslant N$ and from the choice of N, there necessarily exists an integer t in $[1, n]$ and a word u in $Q_{j,t}$ such that w' is u^k-reducible. Thus, w is $(uv)^k$-reducible. From the induction hypothesis and the hypothesis of Subcase 2.2.3, v is in $Q'_{i,k}$, t not in $\{i, j, k\}$ and u in $Q'_{j,t}$. From Lemma 3, there exists a word z in $Q_{i,j} \cup Q_{j,k}$ such that w is z-reducible, contradicting our hypothesis.

Consequently, there exists a word u_0 in $Q^*_{i,k}$ and a word w'' in L such that $w''u_0$ in com(w), $|u_0| \leqslant pk$ and for all u in $Q_{i,k}$, w'' is not u-reducible. From the induction hypothesis, $|y|_{a_k} = N \geqslant pk$ and $|y|_{a_i} \geqslant r_i(w)\, kp \geqslant pk$; hence, there exists a word y_1 such that $y_1 u_0$ in com(y). Then $(e, w'_1 a_k, z_0) \overset{*}{\vdash} (y, a_k, z_0\gamma) \vdash^{s} (y, e, z_0\gamma a_k) \vdash^{r} (y_1, e, z_0\gamma a_k) \vdash^{p} (y_1 a_k, e, z_0\gamma)$. Clearly, conditions 1 to 3 are satisfied and for condition 4, $|y_1 a_k|_{a_i} \geqslant |y|_{a_i} - |u_0| \geqslant |y|_{a_i} - pk(r_i(w) - 1) \geqslant pk(r_i(y_1 a_k w_2 \gamma))$, by the choice of u_0, $r_i(y_1 a_k w_2 \gamma') = r_i(w) - 1$.

That ends the proof of the claim.

Now let w be a word in L. From the claim, there is a computation $(e, w, z_0) \overset{*}{\vdash} (y, e, z_0\gamma)$ satisfying conditions 1 to 3. In order to prove that w in $T(M)$, it suffices to prove that $(y, e, z_0\gamma) \vdash^{r^*} (e, e, z_0)$. We will make an induction on the length of γ. If $|\gamma| = 0$, then y in L and there is a computation $(y, e, z_0 \overset{*r}{\vdash} (e, e, z_0)$. If $|\gamma| \geqslant 1$, $\gamma = \gamma' a_j$ for some a_j in Σ and there is a word u in Q with $|u|_{a_j} \neq 0$, such that $y\gamma$ is u-reducible. From the claim, there exists an a_s in Σ-alph(γ) such that u in $Q'_{s,j}$ and $|y|_{a_j}=N>|u|_{a_j}$. Since $|y|_{a_s}=|y\gamma|_{a_s}$, there exists a word y' such that $y'u$ in com(y). Now it follows that: $(y, e, z_0\gamma'a_j) \vdash^r (y', e, z_0\gamma'a_j) \vdash^p (y'a_j, e, z_0\gamma) \vdash^{p^*} (y_1, e, z_0\gamma_1)$, where $(y_1, e, z_0\gamma_1)$ satisfies conditions 1 to 3 and $|\gamma_1| \leqslant |\gamma'| < |\gamma|$. Thus, from the (implicit) induction hypothesis, w in $T(M)$. So we get that $L=T(M)$, hence L is context-free. ▮

We now return to the proof of Theorem 1.

Proof of Theorem 1. Let F and P be finite sets included in Σ^*, where $\Sigma=\{a_1, a_2, ..., a_n\}$, $L=\text{com}(FP^*)$; recall that $R=\bigcup_{1\leqslant i\leqslant j\leqslant n} a_i^*a_j^*$, $Q=P\cap R$ and for i, j in $[1, n]$ $Q_{i,j}=\text{com}(Q)\cap a_i^*a_j^*$ and $Q'_{ij}=Q_{ij}-(a_i^*\cup a_j^*)$. If L is a context-free language, then L is a B-CF language.

Conversely, let us assume that L is a B-CF language. From Lemma 2, P satisfies Properties A and B. Hence from Lemma 3, Q satisfies Property C and from Lemma 4, com(Q^*) is a context-free language. Now, from Property A, one can find a finite set G such that $\text{com}(P^*)=\text{com}(G) \ddagger \text{com}(Q^*)$. Since the family of context-free languages is closed under shuffle with a regular language we get that com(P^*) is a contex-free language. At last, $\text{com}(FP^*)=\text{com}(F) \ddagger \text{com}(P^*)$ is also a context-free language. ▮

From Lemmas 2, 3, and 4 we can deduce immediately the following result, which does not appear in the statement of Theorem 1:

PROPOSITION 5. *Let F and P be finite nonempty subsets of Σ^*. Then the following conditions are equivalent*:

(1) com(P^*) *is a context-free language.*

(2) com(FP^*) *is a context-free language.*

(3) *P satisfies the two following properties*:

A. *For all u in P, $u^+\cap\text{com}(Q^*)\neq\varnothing$.*

B. *For all u in $Q'_{i,j}$, for all v in $Q'_{s,t}$ where a_i, a_j, a_s, a_t are distinct letters, $(uv)^+\cap(\text{com}(Q-Q'_{i,j})^*\cup\text{com}(Q-Q'_{s,t})^*)\neq\varnothing$.*

Remark. If $P=Q$, only Property B must be considered. It is then easy to conclude: com(Q^*) is a context-free language if and only if for all i, j, s, t in $[1, n]$, $\text{com}(Q^*)\cap a_i^*a_j^*a_s^*a_t^*$ is context-free. If there is an i in $[1, n]$ such that $Q-\Sigma^*a_i\Sigma^*=\varnothing$, then com$(Q^*)$ is context-free. In the general case, Properties A and B in the above proposition are decidable by Lemma C. Indeed, $u^+\cap\text{com}(Q^*)\neq\varnothing$ if and only if $\Psi(u^+\cap\text{com}(Q^*))\neq\varnothing$. But $\Psi(u^+\cap\text{com}(Q^*))=\Psi(u^+)\cap\Psi(Q^*)$ is a calculable semilinear set by Lemma C since $\Psi(u^+)$ and $\Psi(Q^*)$ are clearly calculable linear sets. Similarly, $\Psi((uv)^+\cap(\text{com}(Q-Q'_{i,j})^*\cup\text{com}(Q-Q'_{s,t})^*))$ is a calculable semilinear set. Thus we can state:

COROLLARY 6. *One can decide, given a finite language P, whether or not* com(P^*) *is a context-free language.*

EXAMPLE 4. Given $\Sigma = \{a_1, a_2, ..., a_n\}$, consider the language L, where $L = \{u \in (\Sigma^2)^* \mid \text{for all } i \in [1, n],\ |u| \geqslant 2\,|u|_{a_i}\}$. It can be shown that $L = \text{com}(P^*)$ with $P = \{a_i a_j \mid 1 \leqslant i < j \leqslant n\}$. If a_i, a_j, a_s, a_t are four distinct letters in Σ, $a_i a_j a_s a_t$ is in $\text{com}((a_i a_s)(a_j a_t)) \subset \text{com}(Q - Q'_{i,j})^*$. Then, we may conclude from Proposition 5, that L is context-free.

EXAMPLE 5. More generally, take two positive numbers s and t, such that $s < t \leqslant ns$ and the commutative language $L(s, t) = \{u \in \Sigma^* \mid$ there exists a k, $|u| = kt$ and for all $i \in [1, n]$, $|u|_{a_i} \leqslant ks\}$ (the preceding case is similar to that of $s = 1$ and $t = 2$). One is able to prove that $L(s, t) = \text{com}(P^*)$ with $P = L(s, t) \cap \Sigma^t$. There are two cases:

(a) If $t > 2s$, then for all i, j in $[1, n]$, $P \cap a_i^* a_j^* = \varnothing$, hence $Q = \varnothing$. Property 3A is not satisfied, and $L(s, t)$ is not a context-free language.

(b) If $t \leqslant 2s$, then $Q = \{a_i^{t_1} a_j^{t_2} \mid 1 \leqslant i < j \leqslant n,\ 1 \leqslant t_1,\ t_2 \leqslant s, \text{ and } t_1 + t_2 = t.\}$ Then one is able to show that Properties 3A and 3B are satisfied, which implies that the language $L(s, t)$ is context-free if and only if $t \leqslant 2s$.

At last, we are able to show, from Proposition 5, that it is easy to deduce a well-known result on commutative languages over three letter alphabets.

COROLLARY 7 [5, 6, 8]. *Let L be a commutative language included in Σ_3^* with $\Sigma_3 = \{a_1, a_2, a_3\}$. Then, L is a context-free language if and only if $L \cap a_1^* a_2^* a_3^*$ is a context-free language.*

Proof. If L is a context-free language, $L \cap a_1^* a_2^* a_3^*$ is a context-free language. Conversely, if $L \cap a_1^* a_2^* a_3^*$ is a context-free language, there exist $w_1, w_2, ..., w_t$ in Σ^* and finite sets $Q_1, Q_2, ..., Q_t \subset (a_1^* a_2^* \cup a_1^* a_3^* \cup a_2^* a_3^*)$ such that $\text{com}(L) = \text{com}(w_1 Q_1^* \cup \cdots \cup w_t Q_1^*)$. For i in $[1, t]$, Q_i satisfies Properties A and B and from Proposition 5, $L_i = \text{com}(w_i Q_i^*)$ is a context-free language. Thus, $\text{com}(L) = \bigcup_{i=1}^{t} L_i$ is a context-free language. ∎

REFERENCES

1. M. FLIESS, personal communication, 1970.
2. S. GINSBURG, "The Mathematical Theory of Context-Free Languages," McGraw-Hill, New York, 1966.
3. S. GINSBURG, "Algebraic and Automata-Theoretic Properties of Formal Languages," North-Holland, Amsterdam, 1975.
4. M. A. HARRISON, "Introduction to Formal Language Theory," Addison-Wesley, Reading, MA, 1978.
5. H. A. MAURER, The solution to a problem by Ginsburg, *Inform. Process. Lett.* **1** (1971), 7–10.
6. I. OSHIBA, On permuting letters of words in context-free languages, *Inform. and Control* **20** (1972), 405–409.
7. R. J. PARIKH, On context-free languages, *J. Assoc. Comput. Mach.* **13** (1966), 570–581.
8. J. F. PERROT, Sur la fermeture commutative des C-langages, *C.R. Acad. Sci. Paris* **265** (1967), 597–600.