RONALD V. BOOK
SHEILA A. GREIBACH

The independence of certain operations on semiAFLs


<http://www.numdam.org/item?id=ITA_1978__12_4_369_0>
THE INDEPENDENCE OF CERTAIN OPERATIONS ON SEMIAFLS (*)

by Ronald V. Book (1) and Sheila A. Greibach (2)

Communicé par J. Berstel

Abstract. — It is shown that the operations of homomorphic replication, intersection, and Kleene* are independent for full semiAFLs. Most of the results depend on known properties of certain families of languages; however, some new families are provided via a diagonal construction.

SECTION 1

In a number of recent papers [2, 5, 8, 21, 22, 24, 25], the operation of homomorphic replication (first introduced in [13]) has proved to be useful in characterizing a variety of classes of languages arising naturally in different situations—machines, grammars, string relations, complexity classes, etc. When combined with other operations on languages, simple representations of various classes have been obtained. For example, the class of recursively enumerable sets is the smallest class of languages containing the regular sets and closed under intersection and homomorphic replication, while NP is the smallest class of languages containing the regular sets and closed under intersection and polynomial bounded homomorphic replication [2]. In some of the cases considered, the classes are characterized in terms of some specific operations and then it is shown that they are automatically closed under other operations. Here we concentrate on the operations of homomorphic replication, intersection and Kleene* and the full semiAFL operations (union, homomorphism, inverse homomorphism and intersection with regular sets).

(*) This research was supported in part by the National Science Foundation under Grants MCS 77-11360, DCR 74-15091, and MCS 78-04725.


(1) Department of Mathematics, University of California, Santa Barbara, California, 93106, U.S.A.

(2) Department of System Science, University of California, Los Angeles, California, 90024, U.S.A.
We first study the relationships between these three operations in the context of full semiAFLs. We show that these operations are completely independent in this context. There is a full semiAFL that is closed under none of these operations and one which is closed under all three of these operations and, for every choice of one (two) of these operations, there is a full semiAFL that is closed under that one (two) operations but is not closed under the other two (one) operations. In addition, we consider these operations in the context of semiAFLs that are closed under linear erasing but are not full and of semiAFLs that are not closed under linear erasing. For the most part, the examples and counter-examples presented are classes of languages that have arisen naturally in various different circumstances, and whose properties are already documented in the literature. However, three of the examples are given by diagonal type existence proofs similar to those introduced in [10] and [14] and, since one of the semiAFLs involved must contain languages which are not recursively enumerable, possibly no simple example exists in that case.

SECTION 2

It is assumed that the reader is familiar with the basic concepts from the theories of automata, computability and formal languages. Some of the concepts that are most important for this paper are reviewed here and notation is established.

**Notation:** For a string $w$, $|w|$ denotes the length of $w$. For a finite set $S$, $\# S$ denotes the number of members of $S$. The reversal $w^R$ of a string $w$ is the string obtained by writing $w$ in reverse order. Let $w^1 = w$, $w^{n+1} = ww^n$.

Kleene* is the operation which takes a language $L$ into

$$L^+ = \{ w_1 \ldots w_n \mid n \geq 1, \text{each } w_i \in L \}.$$  

We use $e$ for the empty string, Kleene* is the operation taking $L$ into $L^* = L^+ \cup \{ e \}$. Inverse homomorphism is the operation determined by a homomorphism $h$ taking $L$ into $h^{-1}(L) = \{ w \mid h(w) \in L \}$. By homomorphism, we mean monoid homomorphism, i.e., a function $h: \Sigma^* \rightarrow \Delta^*$ such that for all $x, y \in \Sigma^*$, $h(xy) = h(x)h(y)$.

We shall be concerned with special types of homomorphisms:

**Definition:** A homomorphism $h$ is nonerasing if $h(w) \neq e$ for $w \neq e$. A homomorphism $h$ is linear erasing on a language $L$ if there is a $k > 0$ such that for all $w$ in $L$ with $|w| \geq k$, $|w| \leq k |h(w)|$. A class $\mathcal{L}$ of languages is closed under (nonerasing, linear erasing) homomorphism if for every language $L$
and any homomorphism $h$ (that is nonerasing, linear erasing on $L$), $h(L) = \{ h(w) | w \in L \}$ is in $\mathcal{L}$.

Now we give the definitions and notation used for discussing semiAFLs.

**Definition:** A *semiAFL* is a family of languages containing at least one nonempty set and closed under union, nonerasing homomorphism, inverse homomorphism, and intersection with regular sets. A *full semiAFL* is a semiAFL closed under arbitrary homomorphisms. An *AFL* (full *AFL*) is a semiAFL (full semiAFL) closed under concatenation and Kleene*.

For a family $\mathcal{L}$, we use the notation $\mathcal{M}(\mathcal{L})$ (respectively, $\hat{\mathcal{M}}(\mathcal{L})$, $\mathcal{F}(\mathcal{L})$, $\mathcal{F}^{\text{lin}}(\mathcal{L})$) for the least semiAFL (respectively full semiAFL, AFL, full AFL) containing $\mathcal{L}$. If $\mathcal{L} = \{ L \}$, we write $\mathcal{M}(L)$ etc., and call it a *principal* semiAFL. We add the subscript "\(\cap\)" to require closure under intersection; thus, $\mathcal{M}_{\cap}(L)$ is the least intersection closed semiAFL containing the language $L$. We add the superscript "\(\text{lin}\)" to require closure under linear erasing homomorphism; thus, $\mathcal{F}^{\text{lin}}(\mathcal{L})$ is the least AFL containing $\mathcal{L}$ and closed under linear erasing homomorphism, while $\mathcal{M}^{\text{lin}}_{\cap}(L)$ is the least intersection closed semiAFL containing $L$ and closed under linear erasing homomorphism.

Certain special languages and families of languages occur often enough to deserve special names. We shall reserve for the mirror image language on two letters with center letter the name

$$\text{PAL} = \{ wcw^R | w \in \{ a, b \}^* \}$$

and let $\text{PAL}_e = \text{PAL} \cup \{ e \}$. We let $\mathcal{L}_{\text{BNP}} = \mathcal{M}_{\cap}(\text{PAL}_e)$ (since this family was first discussed in [6]). We use REGL for the family of regular sets, RE for the family of recursive enumerable languages, and RECURSIVE for the family of recursive languages.

The last definitions and rotation of this section involve homomorphic replication.

**Definition:** Let $\rho$ be a function from $\{ 1, \ldots, n \}$ into $\{ 1, R \}$ and for $1 \leq i \leq n$, let $h_i$ be a homomorphism. The operation on languages defined by

$$\langle \rho, h_1, \ldots, h_n \rangle(L) = \{ h_1(w)^{\rho(1)} \ldots h_n(w)^{\rho(n)} | w \in L \}$$

is a *homomorphic replication*. It is *nonerasing* if each $h_i$ is nonerasing and is *linear erasing* on $L$ if each $h_i$ is linear erasing on $L$.

We shall add the subscript $r$ to specify a family closed under homomorphic replication of the appropriate type. Thus, $\mathcal{M}_r(\mathcal{L})$ (or $\hat{\mathcal{M}}_r(\mathcal{L})$) is the least semiAFL (full semiAFL) containing $\mathcal{L}$ and closed under nonerasing homomorphic replication, while $\mathcal{M}^{\text{lin}}_r(\mathcal{L})$ is the least semiAFL containing $\mathcal{L}$
and closed under linear erasing homomorphic replication. Clearly, a class of languages closed under (nonerasing, linear erasing) homomorphic replication is closed under (nonerasing, linear erasing) homomorphism, while a class of languages closed under nonerasing homomorphic replication and under linear erasing (arbitrary) homomorphism is also closed under linear erasing (arbitrary) homomorphic replication.

We conclude with a few useful facts on linear context-free languages, recursively enumerable languages, and homomorphic replication. Further discussion can be found in [1, 2, 5, 6, 16, 20].

**Lemma 2.1:** The class of linear context-free languages is the least semiAFL containing \( \text{PAL} \). Any semiAFL containing \( \{ e \} \) and closed under nonerasing homomorphic replication contains all linear context-free languages.

**Lemma 2.2:** An intersection closed semiAFL containing \( \{ e \} \) is closed under nonerasing homomorphic replication if and only if it contains PAL.

**Lemma 2.3:** An intersection closed full semiAFL contains all recursively enumerable languages if either (1) it is closed under homomorphic replication, or (2) it contains \( \{ a^n b^n \mid n \geq 1 \}^+ \).

**Lemma 2.4:** The class of recursive sets is an intersection closed semiAFL closed under linear erasing homomorphic replication.

**Section 3**

In this section, we demonstrate the independence of the three operations of intersection, Kleene\(^+\) and homomorphic replication, even in the presence of the full semiAFL operations. We establish by example the following independence theorem. All but one of the eight examples are well known and documented in the literature.

**Theorem 3.1:** The operations of homomorphic replication, intersection and Kleene\(^+\) are independent in the presence of the full semiAFL operations. Specifically, there are eight full semiAFLs \( \mathcal{L}_{i,j,k} \), \( i, j, k \in \{ 0, 1 \} \) such that \( \mathcal{L}_{i,j,k} \) is closed under homomorphic replication (respectively, intersection, Kleene\(^+\)) if and only if \( i = 1 \) (respectively, \( j = 1 \), \( k = 1 \)).

First we consider the cases where we have either all three operations or none of them. The family of linear context-free languages, \( \mathcal{N} \) (PAL), serves as \( \mathcal{L}_{0,0,0} \) [1, 13, 15, 16]. The family of recursively enumerable languages is \( \mathcal{L}_{1,1,1} \) since it is the least full semiAFL closed under intersection and homomorphic replication [1, 2].

Now let us consider the other cases in “numerical” order from 001 to 110. The family of context-free languages is \( \mathcal{L}_{0,0,1} \), since it is well known to be a
full AFL closed under none of the other operations, the family of languages accepted by on-line reversal bounded multicounter machines described in [1, 22] is closed under intersection (and is in fact $\hat{\mathcal{M}}_n (\{ a^n b^n | n \geq 0 \})$ but is not closed under homomorphic replication nor under Kleene$^+$ (does not contain $\{ a^n b^n | n \geq 0 \}$) since it is a proper subfamily of RECURSIVE [1]. The family $\hat{\mathcal{M}}_r$ (REGL), the closure of the regular languages under homomorphic replication, is a full semiAFL closed under homomorphic replication but not under intersection since it is also the family of languages accepted by finite reversal on-line nondeterministic checking automata and so properly contained in RECURSIVE [12, 21]. This family is not closed under Kleene$^+$ as a result of a general phenomenon: if $\hat{\mathcal{M}}_r (\mathcal{L})$ is not closed under homomorphic replication, then $\hat{\mathcal{M}}_r (\mathcal{L}^+)$ cannot be closed under Kleene$^+$ [18, 20, 21]. Thus, $\hat{\mathcal{M}}_r$ (REGL) is $\mathcal{L}_{1,0,0}$.

Now we turn to the families closed under two but not three of these operations. The family of regular sets is closed under intersection and Kleene$^+$ but not under replication, and so is $\mathcal{L}_{0,1,1}$. Another candidate for $\mathcal{L}_{0,1,1}$ is given by a diagonal argument in [10]: there is a nonregular language $L \subseteq a^*$ such that $\mathcal{F}_n (L) \cap \text{RE} = \text{REGL}$. Thus $\mathcal{F}_n (L)$ is by definition closed under intersection and Kleene$^+$ but cannot be closed under homomorphic replication since it does not contain RE. We shall use a similar diagonal argument to obtain $\mathcal{L}_{1,1,0}$.

For $\mathcal{L}_{1,0,1}$, we take the family of languages accepted by one-way nondeterministic finite visit checking automata [21] which can also be described as the family of languages generated by absolutely parallel grammars [27] or as the family of languages obtained by taking two-way deterministic finite state transductions of regular sets [28]. This family is closed under homomorphic replication and Kleene$^+$ [21] but cannot be closed under intersection since it contains only recursive sets.

For $\mathcal{L}_{1,1,0}$, we have only an existence proof, given in Theorem 4.2 in the next section. A full semiAFL closed under intersection and homomorphic replication must contain RE; nonclosure under Kleene$^+$ requires it to contain languages which are not recursively enumerable. The families of languages in the arithmetic hierarchy (see [2] for definitions) are all closed under Kleene$^+$ at each stage. Thus, a "natural" candidate for $\mathcal{L}_{1,1,0}$ is not apparent.

What happens if we do not have a full semiAFL? Since the trivial case $\langle \rho, h_1 \rangle (L) = h_1 (L)$ is considered to be a homomorphic replication, closure under homomorphic replication implies closure under all homomorphisms.
Thus, we must consider limitations on the amount of erasing permitted in a homomorphic replication.

If we try to split our above eight cases (i.e., closure or nonclosure under $\cap$, Kleene$^+$ and homomorphic replication) into 24 cases by starting with semi-AFLs and subdividing by closure under homomorphic replication, or under just linear bounded homomorphic replication or nonerasing homomorphic replication, we find that not all cases can occur. In particular, one consequence of Theorem 2.2 of [5] can be stated as follows.

**Proposition 3.2:** If $L$ is a semiAFL closed under intersection and non-erasing homomorphic replication, then $L$ is closed under linear erasing homomorphic replication.

Thus, if we consider intersection closed semiAFLs, it suffices to consider the case of closure under linear erasing but not arbitrary homomorphic replication.

An example of a semiAFL closed under intersection, Kleene$^+$, and linear erasing but not arbitrary homomorphic replication is $\text{NTIME}(n)$, the family of languages accepted in realtime by nondeterministic multitape Turing machines. This family is the least intersection closed semiAFL containing the context-free languages [3] and thus is closed under Kleene$^+$ and linear erasing homomorphic replication (since $M \cap (L)$ is an AFL if $L$ is an AFL [23] and is closed under linear erasing homomorphic replication if $L$ contains PAL [5]). However, $\text{NTIME}(n)$ is not a full semiAFL and indeed is not closed under homomorphisms which erase "more than linearly" (e.g., if $f$ is a time constructible function such that $\lim_{n \to \infty} n + 1/f(n) \leq 0$, there is a language $L \in \text{NTIME}(n)$ and a homomorphism $h$ such that $|w| \leq f(|h(w)|)$ for all $w \in L$ but $h(L) \notin \text{NTIME}(n)$; cf. [7, 29] for further explanation).

Consider semiAFLs closed under intersection but not under nonerasing homomorphic replication. (Note that the full semiAFL cases were discussed earlier.) Let $\text{COUNT}(n)$ ($\text{COUNT}(\text{lin})$) be the family of languages accepted in real time (resp., in linear time) by on-line nondeterministic multicounter machines. Both $\text{COUNT}(n)$ and $\text{COUNT}(\text{lin})$ are semiAFLs closed under intersection and Kleene$^+$, and neither family is closed under nonerasing homomorphic replication since neither contains PAL. The class $\text{COUNT}(\text{lin})$ is closed under linear erasing homomorphism but $\text{COUNT}(n)$ is not [10, 20].

As an example of a semiAFL that is closed under intersection but not under linear erasing homomorphism nor nonerasing homomorphic replication nor Kleene$^+$, consider the family $\text{PBLIND}(n)$ of languages accepted in real time by on-line nondeterministic partially blind multicounter machines.

R.A.I.R.O. Informatique théorique/Theoretical Computer Science
This family can also be described as the family of computation sequence sets or Petri net languages [26] and is the least intersection closed semiAFL containing the Dyck set on one letter [22]. The facts that PBLIND (n) is not closed under linear erasing homomorphism and does not contain PAL (hence, is not closed under nonerasing homomorphic replication) are established in [22]. The claimed decidability of the reachability problem for vector addition systems [30] implies that PBLIND (n) cannot be closed under Kleene+ [22].

We do not have a “natural” example of an intersection closed semiAFL closed under linear erasing but not arbitrary homomorphism and not closed under either Kleene+ or nonerasing homomorphic replication. A complicated diagonal argument (proof of Theorem 4.6 below) shows that such a semiAFL must exist (and can be taken as a subfamily of RECURSIVE). We conjecture that the family PBLIND (lin) of languages accepted in linear time by on-line nondeterministic partially blind multicounter machines has the required properties; all have been shown except nonclosure under arbitrary homomorphism [22].

Let us consider briefly semiAFLs not closed under either nonerasing homomorphic replication or intersection. Six examples (i.e., closed or not closed under Kleene+ and closed under arbitrary homomorphism or closed under linear erasing but not arbitrary homomorphism or not closed under linear erasing homomorphism) can be found by considering principal semiAFLs generated by context-free languages; details can be found in [17].

Now we turn to semiAFLs closed under nonerasing homomorphic replication but not both intersection and Kleene+*. We demonstrate in the next section (Lemma 4.3 and Theorem 4.4), by a diagonal argument, the existence of an intersection closed semiAFL closed under linear erasing but not arbitrary homomorphic replication and not closed under Kleene+. We conjecture that \( \mathcal{L}_{\text{BNP}} \), the family of languages accepted in real time by nondeterministic finite reversal multitape Turing machines, is such a family (see [5, 6]).

When we look for semiAFLs closed under nonerasing homomorphic replication but not under intersection, we encounter difficulty in locating ones which are not closed under arbitrary homomorphic replication and in separating nonerasing versus linear erasing homomorphic replication. All of the “natural” and well-studied cases are full semiAFLs, in part due to the following result.

**Theorem 3.3:** If \( \mathcal{L} \) is a full semiAFL, then the closure of \( \mathcal{L} \) under linear erasing homomorphic replication is exactly the closure of \( \mathcal{L} \) under arbitrary homomorphic replication; that is, \( \hat{\mathcal{M}}_r (\mathcal{L}) = \mathcal{M}^{\text{lin}}_r (\mathcal{L}) \).
Proof: Let $L_1 = \langle p, h_1, \ldots, h_n \rangle (L)$, for $L \in \mathcal{L}$. Let $c$ be a new symbol. Define homomorphisms $h_i, g$, and $h_i'$, $i = 1, \ldots, n$ by $h_i'(a) = c$ if $h_i(a) = e$, $h_i'(a) = h_i(a)$ if $h_i(a) \neq e$, $g(c) = e$, $g(a) = a$ for $a \neq c$, $h(a) = e$ if $h_i(a) = e$ for all $i$, and $h(a) = a$ if $h_i(a) \neq e$ for some $i$. Thus $f = \langle p, h_1', \ldots, h_n' \rangle$ is a nonerasing homomorphic replication and $g$ is linear erasing on $f(h(L))$. Since $\mathcal{L}$ is a full semiAFL, $h(L)$ is in $\mathcal{L}$ and thus the closure of $\mathcal{L}$ under linear erasing homomorphic replication contains $L_1 = f(g(h(L))$. □

In the discussion above, there are four cases that have not been covered: a semiAFL closed under nonerasing but not arbitrary homomorphic replication and not closed under intersection, with or without closure under Kleene* and with or without closure under linear erasing homomorphism. We conjecture that examples of each of these four cases can be found by studying time bounded one-way nondeterministic finite visit or finite reversal checking automata [21], possibly with some variations on the machines. For example, the family of languages accepted in real time by one-way nondeterministic finite reversal checking stack automata is a semiAFL closed under nonerasing homomorphic replication but not under intersection or Kleene* [21] and we conjecture that it is not closed under linear erasing homomorphism; however, if it is so closed, then it is equal to $\hat{\mathcal{A}}_r(REGL)$ by means of Theorem 3.3.

SECTION 4

Now we establish the existence of the “new” families of languages, as described in Section 3. We begin by obtaining $\mathcal{L}_{1,1,0}$ described in the proof of Theorem 3.1.

We adopt the following terminology from [14].

**Definition:** An $n$-ary language operator $f$ is monotone if $L_i \subseteq L_i$, $i = 1, \ldots, n$ always implies that $f(L_1, \ldots, L_n) \subseteq f(L_1, \ldots, L_n)$. It is local if it is monotone and for any $L_1, \ldots, L_n$ and $y \in f(L_1, \ldots, L_n)$ there are finite sets $F_i \subseteq L_i$ such that $y \in f(F_1, \ldots, F_n)$. It is uniformly local with uniform bound $k$ for a positive integer $k$ if it is monotone and for any $L_1, \ldots, L_n$ and $y \in f(L_1, \ldots, L_n)$ there are finite sets $W_i \subseteq L_i$, each with $\# W_i \leq k$, $i = 1, \ldots, n$, such that $y \in f(W_1, \ldots, W_n)$.

An operator is local if we always have

$$f(L_1, \ldots, L_n) = \bigcup_{F_i \subseteq L_i \text{ finite}} f(F_1, \ldots, F_n).$$

For $f$ to be uniformly local, there must be an integer $k$ such that the expression above holds with the $F_i$ restricted not just to be finite but also to contain...
at most \( k \) elements (words). The full semiAFL operations are clearly uniformly local as are homomorphic replication and intersection (see [14] for similar arguments). However, Kleene\(^+\) is local but not uniformly local (if \( L \subseteq S^+ \), \( S \not\in \Sigma^+ \) and \( \# L > k \), we cannot obtain \( L^+ \) as the union of sets \( W^+ \) where \( W \) contains at most \( k \) words). This observation underlies the proof of Lemma 4.1 below. The class of uniformly local operations is closed under composition.

**NOTATION:** For a countable collection \( \mathcal{G} \) of language operators, let \( \mathcal{G} (L) \) be the class of all \( f(L, \ldots, L) \) with \( f \in \mathcal{G} \).

**LEMMA 4.1:** Let \( \mathcal{G} \) be a countably infinite class of uniformly local operations. There exists a nonempty language \( L_0 \subseteq a^+ \) such that \( \mathcal{G} (L_0) \) does not contain \( (L_0 c)^+ \).

**Proof:** Let the operations in \( \mathcal{G} \) be indexed as \( f_k \) for \( k \geq 1 \). For each \( k \geq 1 \), we shall describe finite sets \( U_k \) and \( V_k \) such that \( U_k \subseteq U_{k+1} \), \( V_k \subseteq V_{k+1} \), and \( U_i \cap V_j = \emptyset \) for all \( i, j \). The desired language will be \( L_0 = \bigcup_{k \geq 1} U_k \).

Having obtained \( U_k \), we shall select a word \( z_k \) and a finite set \( W_k \) in such a way that, if we let \( U_{k+1} = U_k \cup W_k \), either \( z_k \in (W_k c)^+ - f_k (L_0, \ldots, L_0) \) or \( z_k \in (W_k c)^+ - f_k (L_0, \ldots, L_0) \). In either case, the monotonicity of \( f_k \) ensures that \( (L_0 c)^+ \neq f_k (L_0, \ldots, L_0) \). Let \( U_1 = \{ a \} \) and \( V_1 = \emptyset \). For \( k \geq 1 \), let \( A_k = a^+ - (U_k \cup V_k) \) and \( B_k = f_k ((A_k \cup U_k), \ldots, (A_k \cup U_k)) \).

Note that \( A_k \) is infinite if \( U_k \cup V_k \) is finite and that, for any \( L \subseteq a^+ \) with \( L \cap V_k = \emptyset \), the monotonicity of \( f_k \) implies that \( f_k (L, \ldots, L) \subseteq B_k \). In particular, since \( L_0 = \bigcup_{k \geq 1} U_k \subseteq a^+ \), for any \( k \geq 1 \) we have \( L_0 \subseteq A_k \cup U_k \) and so \( f_k (L_0, \ldots, L_0) \subseteq B_k \).

Since each \( f_k \) is (by hypothesis) uniformly local, \( f_k \) has a uniform bound, say \( p_k \). Let \( n_k \) be the arity of \( f_k \). Let \( s_k = 1 + n_k p_k + \# U_k \), let \( y_1, \ldots, y_{s_k} \) be any \( s_k \) distinct members of \( A_k \), and let \( z_k = y_1 c \ldots y_{s_k} c \). We consider two cases, either \( z_k \not\in B_k \) or \( z_k \in B_k \).

**CASE 1:** If \( z_k \not\in B_k \), then \( z_k \not\in f_k (L_0, \ldots, L_0) \). In this case, let \( W_k = \{ y_1, \ldots, y_{s_k} \}, \ U_{k+1} = U_k \cup W_k \), and \( V_{k+1} = V_k \). Thus, \( z_k \in (L_0 c)^+ - f_k (L_0, \ldots, L_0) \).

**CASE 2:** If \( z_k \in B_k \), then there are finite sets \( Y_1, \ldots, Y_{n_k} \) each \( Y_i \subseteq A_i \cup U_k \) and each \( \# Y_i \leq p_k \), such that \( z_k \in f_k (Y_1, \ldots, Y_{n_k}) \). Let \( W_k = Y_1 \cup \ldots \cup Y_{n_k} \) and \( U_{k+1} = U_k \cup W_k \). Then we have \( z_k \in f_k (U_{k+1}, \ldots, U_{k+1}) \subseteq f_k (L_0, \ldots, L_0) \).

Since \( s_k = 1 + n_k p_k + \# U_k > \# U_{k+1} \), there is some \( y_{i_0} \) such that \( y_{i_0} \not\in U_{k+1} \). Let \( V_{k+1} = V_k \cup \{ y_{i_0} \} \). Thus, keeping \( L_0 \) and the \( V_i \) disjoint ensures that \( y_{i_0} \not\in L_0 \) and so \( z_k \not\in (L_0 c)^+ \).
Setting \( L_0 = \bigcup_{k \geq 1} U_k \), it is clear from the construction that, for each \( k \), \( U_k \subseteq U_{k+1} \), \( V_k \subseteq V_{k+1} \), and \( L_0 \cap V_k = \emptyset \). For each \( k \), the construction ensured that \( (L_0 c)^+ \neq f_k(L_0, \ldots, L_0) \) and so \\
\( (L_0 c)^+ \notin \mathcal{G}(L_0) = \{ f_k(L_0, \ldots, L_0) \mid k \geq 1 \} \). \( \square \)

Lemma 4.1 could be made somewhat more general: instead of \( (L_0 c)^+ \), we could use \( g(L_0) \) where \( g \) is any monotone but not uniformly local operation with the property that \( g(A) \subseteq g(B) \) if and only if \( A \subseteq B \).

Now we draw the desired conclusion.

**Theorem 4.2:** There is a language \( L_0 \subseteq a^+ \) such that the least full semiAFL containing \( L_0 \) and closed under intersection and homomorphic replication is not closed under Kleene*.

**Proof:** In Lemma 4.1, we take as \( \mathcal{G} \) the closure under composition of the full semiAFL operations and intersection and all homomorphic replications. Thus, \( \mathcal{G}(L_0) = \hat{\mathcal{M}} \cap (\hat{\mathcal{M}}_r(L)) \) contains \( L_0 c \) but not \( (L_0 c)^+ \) and so is not closed under Kleene*.

The language \( L_0 \) of Theorem 4.2 cannot be made recursively enumerable, but can be placed in \( \hat{\mathcal{M}} \) (Co-RE) (where Co-RE is the family of complements of recursively enumerable sets) if we notice that \( L_0 \) could be designed to be accepted by an oracle Turing machine with an oracle for \( \{ (x, y) \mid x \text{ is accepted by the } y\text{-th Turing machine} \} \).

The proof of Lemma 4.1 can be modified to obtain an example expressible as \( \hat{\mathcal{M}} \cap (L_0) \).

**Corollary:** There is a language \( L_0 \subseteq \{ a, b, c \}^* \) such that \( \hat{\mathcal{M}} \cap (L_0) \) contains all recursively enumerable languages and is closed under homomorphic replication but not under Kleene*.

**Proof:** We modify the proof of Lemma 4.1 to start with \( U_1 = \text{PAL} \), and define \( A_k \) as before (so that we only add to \( U_k \) or \( V_k \) words in \( a^+ \) and so not in \( U_1 \)) and let \( s_k = 1 + n_k p_k + \#(U_k \cap a^+). \) We can still conclude that \( (L_0 c)^+ \) is not in \( \mathcal{G}(L_0) \). If we take \( \mathcal{G} \) as the closure under composition of intersection and the full semiAFL operations, then \( \mathcal{G}(L_0) = \hat{\mathcal{M}} \cap (L_0) \) so \( \hat{\mathcal{M}} \cap (L_0) \) is not closed under Kleene*. However, \( \hat{\mathcal{M}} \cap (L_0) \) contains

\[ \text{PAL} = L_0 \cap \{ a, b \}^* c \{ a, b \}^*, \]

so \( \hat{\mathcal{M}} \cap (L_0) \) is closed under homomorphic replication. \( \square \)

It is not known whether we could take the language \( L_0 \) in the Corollary as a one-letter language; the proof in [10] that various one-letter languages \( A \) have the property that RE \( \subseteq \mathcal{F}(A) \) uses the Kleene* operation heavily.
Now we turn to the diagonal constructions needed in the cases where the semiAFL is not full. The first lemma combines the proof of Lemma 4.1 with a modification of the constructions in [17] to obtain a semiAFL that is not closed under arbitrary homomorphism or under Kleene$^*$. To exclude homomorphisms, we need to restrict further our uniformly local operations. We shall now deal with uniformly local operations with uniform bound 1; that is, $f(L_1, \ldots, L_n) = \bigcup_{w_i \in L_i} f(\{w_1\}, \ldots, \{w_n\})$.

**Definition:** An operation $f$ which is uniformly local with uniform bound 1 is linear bounded if there is a $k$ such that if $y \in f(\{w_1\}, \ldots, \{w_n\})$, then $|w_i| \leq k \max (1, |y|)$, $i = 1, \ldots, n$.

**Lemma 4.3:** Let $\mathcal{F}$ be a countably infinite collection of operations on languages, all linear bounded and uniformly local with bound 1. For any language $A$, let $\sigma(A) = \bigcup_{m \geq 1} b^m(A)^m = \{b^m y_1 c \ldots y_m c \mid y_i \in A, m \geq 1\}$. There is a non-empty language $L \subseteq a^+$ such that $(Lc)^+$ is not in $\mathcal{F}(\sigma(L))$.

**Proof:** The proof strategy is similar to that employed in Lemma 4.1 except that, since we are dealing with $\mathcal{F}(\sigma(L))$ instead of $\mathcal{F}(L)$, we must use the linear bounded property to prove that we have essentially the same two cases we had previously. Index $\mathcal{F}$ as in the proof of Lemma 4.1.

Let $U_1 = \{a\}$ and $V_1 = \emptyset$. For $k \geq 1$, let $A_k = a^+ - (U_k \cup V_k)$ and $B_k = f_k(\sigma(A_k \cup U_k), \ldots, \sigma(A_k \cup U_k))$. Let $f_k$ be $n_k$-ary and have linear bound $p_k$. Let $m_k = \min \{r \mid a \in A_k\}$ and $s_k = p_k + n_k + m_k$. Define $\pi$ on $b^*(a^+ c)^+$ by

$$
\pi(b^* a^1 c \ldots a^s c) = \{a^1, \ldots, a^s\}.
$$

Let $z_k = a^{m_k} c a^{m_k+1} c \ldots a^{m_k + s_k n_k - 1} c$.

There are two cases. If $z_k \notin B_k$, let $V_{k+1} = V_k$ and $U_{k+1} = U_k \cup \pi(z_k) \cup \{a^{k+1} \mid a^{k+1} \notin V_k\}$.

Since $\sigma$ is monotone (as are all the operations in $\mathcal{F}$), we have ensured that $z_k$ is in $(Lc)^+$ but not in $f_k(\sigma(L))$.

If $z_k \in B_k$, let $z_k \in f_k(\{y_1\}, \ldots, \{y_{n_k}\})$, each $y_i \in \sigma(A_k \cup U_k)$. Let $W_k = \bigcup_{i=1}^{n_k} \pi(y_i)$. Suppose $\pi(z_k) \subseteq W_k$. Since $\# \pi(z_k) = s_k n_k$, for at least one $y_i$, $\# \pi(y_i) \subseteq s_k$. By the definition of $\sigma$, this means that $|y_i| > s_k = (p_k + n_k + m_k)^5 \geq p_k(n_k s_k)(m_k + s_k n_k) \geq p_k|z_k|$, contradicting the linear boundedness of $f_k$. Hence $\pi(z_k) \subset W_k \neq \emptyset$. Let $V_{k+1} = V_k \cup (\pi(z_k) - W_k)$ and $U_{k+1} = U_k \cup W_k \cup \{a^{k+1} \mid a^{k+1} \notin V_{k+1}\}$.

vol. 12, no 4, 1978
This ensures that $z_k \notin f_k (\sigma (L)), \ldots, \sigma (L))$, but $z_k \in (Lc)^+$. Let $L = \bigcup_{k \geq 1} U_k$.

The arguments used in the proof of Lemma 4.1 show that $(Lc)^+ \notin \mathcal{G} (\sigma (L))$. □

**Theorem 4.4:** There is a semiAFL $\mathcal{L}$ such that:

1) $\mathcal{L}$ is closed under intersection and linear erasing homomorphic replication;
2) $\mathcal{L}$ is not closed under Kleene $^*$ nor under arbitrary homomorphism; and
3) $\mathcal{L}$ is properly contained in the family of recursive sets.

**Proof:** Let the family $\mathcal{G}$ of operations be the closure under composition of the operations of nonerasing homomorphism, inverse homomorphism, intersection with regular sets, intersection, and nonerasing homomorphic replication. These operations are all uniformly local with uniform bound 1 and linear bounded (only inverse homomorphism introduces a linear bound other than 1). Lemma 4.3 yields the existence of $L \subseteq a^+$ such that $\mathcal{G} (\sigma (L))$ does not contain $(Lc)^+$. Arguments similar to those in [11, 23] show that $\mathcal{L} = \mathcal{G} (\sigma (L))$ is closed under union and hence a semiAFL. Proposition 3.2 shows that $\mathcal{L}$ is closed under linear erasing homomorphic replication, so (1) holds for $\mathcal{L}$. If $h$ is the homomorphism $h (b) = e$, $h (a) = a$, $h (c) = c$, then $h (\sigma (L)) = (Lc)^+$, so that $\mathcal{L}$ is not closed under homomorphism. Now $\sigma (L) \cap ba^+ c = bLc$, so $bLc$ and $Lc$ are in $\mathcal{L}$. Hence $\mathcal{L}$ is not closed under Kleene $^*$, so (2) holds.

To show (3), it suffices to show that $L$ can be constructed to be recursive. Notice that each $f_k$ in $\mathcal{G}$ is total recursive as a function from $n_k$-tuples of unit sets into regular sets and is linear bounded, so that for each $z$, the set $g_k (z) = \{ (y_1, \ldots, y_k) | z \in f_k (\{ y_1 \}, \ldots, \{ y_n_k \}) \}$ is finite and $g_k (z)$ is total recursive. We can enumerate $\mathcal{G}$ in such a way that $p_k \leq k$, $n_k \leq k$ and the function $g (k, z) = g_k (z)$ is total recursive as a function of $k$ and $z$. The bound $m_k$ can be computed from $U_k$ and $V_k$. At each step, given $m_k$, we can compute $z_k$ (taking $k$ for $n_k$ and $p_k$) and thus $g (k, z_k)$. For $(y_1, \ldots, y_n_k)$ in $g (k, z_k)$, we can easily test whether $y_i$ is of the form $b^m x_1 c \ldots x_m c$ and if so, whether each $x_j$ is in $V_k$ or in $A_k \cup U_k = a^+ - V_k$. Thus, we can determine whether $g (k, z_k) \cap [\sigma (A_k \cup U_k)]^{m_k}$ is or is not empty and hence the appropriate case can be determined and $V_{k+1}$ and $U_{k+1}$ computed. Hence

$h(r, k) = \begin{cases} 1 & a^r \in U_k, \\ 0 & a^r \notin U_k \end{cases}$

is total recursive. The construction ensures that $a^k$ is in $U_k \cup V_k$, so $a^k$ is in $L$ if and only if $a^k$ is in $U_k$ if and only if $h(k, k) = 1$. Thus, $L$ is a recursive language. □
Remark: By carefully considering the recurrence equation for $s_k$, we can show that $L$ is elementary (i.e., in the Grzegorczyk class $\mathcal{E}_2^0$).

For our final lemma, we must require still stronger conditions on our operations. We rule out closure under homomorphic replication by excluding PAL, and so we must be sure that this language cannot be produced from $\sigma(L)$ by any operation in $\mathcal{G}$.

**Lemma 4.5:** Let $\mathcal{G}$ be a countably infinite collection of operations, each of which is uniformly bounded with uniform bound 1 and linear bounded. Let $\sigma$ be an operation defined by

$$\sigma(A) = \bigcup_{m \geq 1} b^m (Ac)^m = \{ b^m y_1 \ldots y_m | y_1 \in A, m \geq 1 \}.$$  

Let $\mathcal{L}_1$ be a family of languages such that $\mathcal{G}(L_1) \subseteq \mathcal{L}_1$ and if $A \subseteq a^+$ is regular, then $\sigma(A)$ is in $\mathcal{L}_1$. Let $L_1$ be a language not in $\mathcal{L}_1$. Then there is a language $L \subseteq a^+$ such that $\mathcal{G}(\sigma(L))$ contains neither $L_1$ nor $(Lc)^+$.

**Proof:** We proceed as in the proof of Lemma 4.3, but in two stages. The first stage is the same as before, except that we call the sets defined from $U_k$, $V_k$, and $f_k$, $U'_k + 1$ and $V'_k + 1$. We will have $U'_{k+1} \subseteq U_{k+1}$ and $V'_{k+1} \subseteq V_{k+1}$, so the first stage ensures as before that $(Lc)^+ \neq f_k(\sigma(L), \ldots, \sigma(L))$. For the second stage, let $A'_k = a^+ (U'_{k+1} \cup V'_{k+1})$ and $B'_k = f_k(\sigma(A'_k \cup U'_{k+1}), \ldots, \sigma(A'_k \cup U'_{k+1}))$.

Since $A'_k \cup V'_{k+1}$ is cofinite and thus regular, $B'_k$ is in $\mathcal{L}_1$ by hypothesis. Thus $L_1 \neq B'_k$. If $L_1 \neq B'_k \neq \emptyset$, then the monotonicity of $\sigma$ and $f_k$ tells us that $L_1 \neq f_k(\sigma(L), \ldots, \sigma(L))$, so we let $U_{k+1} = U'_{k+1}$ and $V_{k+1} = V'_{k+1}$. Otherwise, there are $x_1, \ldots, x_n$ in $\sigma(A'_k \cup U'_{k+1})$ such that $f_k(\{x_1\}, \ldots, \{x_n\})$ is not contained in $L_1$. Thus, let $U_{k+1} = U'_{k+1} \cup \bigcup_{i=1}^n \pi(x_i)$, and $V_{k+1} = V'_{k+1}$; this ensures that $L_1 \neq f_k(\sigma(L), \ldots, \sigma(L))$. Complete the proof as before. □

**Theorem 4.6:** There is a semiAFL $\mathcal{L}$ such that:

1) $\mathcal{L}$ is closed under intersection and linear erasing homomorphism;
2) $\mathcal{L}$ is not closed under Kleene* or homomorphism;
3) $\mathcal{L}$ does not contain PAL and hence is not closed under nonerasing homomorphic replication;
4) $\mathcal{L}$ does contain $L_2 = \{ w \in \{a, b\}^* | w$ contains the same number of $a$'s as $b$'s $\}$ and
5) $\mathcal{L}$ is properly contained in the family of recursive sets.

**Proof:** In the proof of Lemma 4.5, take $\mathcal{L}_1 =$ COUNTER (lin) and $L_1 =$ PAL, so that $L_1$ is not in $\mathcal{L}_1$, and $\mathcal{L}_1$ is a semiAFL closed under inter-

vol. 12, n° 4, 1978
section and linear erasing homomorphism. A deterministic one-way multi-counter accepter can certainly decide in linear time whether or not \( s = r \), given \( s \) and \( r \) in unary, so \( \sigma (A) \) is in COUNTER (lin) if \( A \) is a regular subset of \( a^+ \). We must exercise some care in the selection of \( \mathcal{G} \). As before, \( \mathcal{G} \) is closed under composition and contains the operations of intersection, nonerasing homomorphism, inverse homomorphism and intersection with regular sets. However, including linear erasing homomorphism requires some care, since the operations in \( \mathcal{G} \) must be everywhere linear bounded. We place \( L_2 \) in \( \mathcal{G} (\sigma (L)) \) by including in \( \mathcal{G} \) the operation of intersection with \( L_2 \); since \( L_2 \) is in \( \mathcal{L}_1 \), \( \mathcal{L}_1 \) is closed under this operation. For any homomorphism \( h \) and positive integer \( t \), define \([h, t]\) on unit sets by \([h, t]\)({\( w \)}) = \( \{ h(w) \} \) if \( |w| \leq t \text{ Max}(1, |h(w)|) \) and \([h, t]\)({\( w \)}) = \( \{ w \} \) elsewhere, and define \([h, t]\) on arbitrary sets by \([h, t](A) = \bigcup_{w \in A} [h, t]\)({\( w \})\). We place \([h, t]\) in \( \mathcal{G} \); clearly, \([h, t]\) is uniformly bounded with uniform bound 1 and linear bounded. We must check that we have not inadvertently put “too much” into \( \mathcal{G} \). However, \([h, t]\)(\( A \)) can be obtained in a uniform manner from \( A \) and \( L_2 \) using the semiAFL operations and intersection and linear erasing homomorphism (i.e., \([h, t]\)(\( A \)) = \( h' (A') \) for \( A' \) in \( \mathcal{M}_\cap (\{ A, L_2 \}) \)) and \( h' \) linear erasing on \( A' \). On the other hand, \( h \) is linear erasing on \( A \), \( h(A) = [h, t]\)(\( A \)) for an appropriate \( t \). Thus, \( \mathcal{G}(\sigma (L)) \) will be \( \mathcal{M}_\cap^{\text{lin}} (\{ \sigma (L), A_2 \}) \), that is, the least intersection closed semiAFL containing \( \sigma (L) \) and \( A_2 \) and closed under linear erasing homomorphism.

Hence, application of Lemma 4.5 yields a language \( L \subseteq a^+ \) such that (1)-(4) hold for \( \mathcal{L} = \mathcal{G}(\sigma (L)) \). It remains to show that \( L \) is recursive. The argument in the proof of Theorem 4.4 will go through if we can show that the construction of \( U_{k+1} \) from \( U'_{k+1} \) can be made algorithmic. We know that either \( L_1 - B'_k \neq \varnothing \) or \( B_k - L \neq \varnothing \) (or both) must hold, so we simple dovetail testing the two conditions. We test \( L_1 - B'_k \neq \varnothing \) by testing successive members of \( L_1 \) for membership in \( B'_k \); if a negative answer turns up, we let \( U_{k+1} = U'_{k+1} \) and stop. We test \( B'_k - L \neq \varnothing \) by enumerating \((x_1, \ldots, x_{n_k})\) in \([\sigma (A_k \cup U'_{k+1})]^n_k \) and testing words in \( f_k (\{ x_1 \}, \ldots, \{ x_{n_k} \}) \) for membership in \( L_1 \); if a negative answer appears, we let \( U_{k+1} = U'_{k+1} \cup \bigcup_{i=1}^{n_k} \pi (x_i) \)

and halt. Thus, \( L \) is recursive. \( \square \)

A homomorphism \( h \) is polynomial erasing on a language \( L \) if there exists \( k \geq 1 \) such that, for all \( w \in L \), \( |w| \leq |h(w)|^k \). In Theorems 4.4 and 4.6, the classes \( \mathcal{L} \) are not closed under polynomial erasing homomorphism.

R.A.I.R.O. Informatique théorique/Theoretical Computer Science
SECTION 5

We conclude by briefly discussing homomorphic duplication, a special case of homomorphic replication studied more thoroughly in a forthcoming paper [4].

**Definition:** A homomorphic replication \( \langle p, h_1, \ldots, h_n \rangle \) is a homomorphic duplication if \( p(i) = 1 \) for \( i = 1, \ldots, n \). For a family of languages \( \mathcal{L} \), let \( \hat{\mathcal{M}}_d(\mathcal{L}) \) be the least full semiAFL containing \( \mathcal{L} \) and closed under homomorphic duplication.

Obviously, closure under homomorphic replication implies closure under homomorphic duplication. The converse is not true. \( \hat{\mathcal{M}}_d(\text{REGL}) \) (which is the family of equal matrix languages of [31]) is not closed under homomorphic replication [24, 25]. Hence we have the following.

**Theorem 5.1:** The operation of homomorphic replication is independent of the operations of homomorphic duplication and the full semiAFL operations.

We conjecture that the same holds if we add Kleene\(^+\) and that the least full AFL closed under homomorphic duplication does not contain PAL.

Homomorphic replication is not independent of duplication under all circumstances. A semiAFL closed under linear erasing homomorphic duplication and containing \( \text{PAL}_e \) must be closed under linear erasing homomorphic replication. An intersection closed semiAFL closed under nonerasing homomorphic duplication and \( n^2 \) bounded erasing (i.e., it contains \( h(L) \) if it contains \( L \) and there is a \( k \) such that \( |w| \leq k \) Max (1, \( |h(w)|e \)) for all \( w \in L \) and so be closed under nonerasing homomorphic replication [4]. It is not known whether this relationship holds if we eliminate the condition of closure under \( n^2 \) erasing. In particular, let

\[
\mathcal{L}_{\text{DUP}} = \mathcal{M}_\cap \{ \text{wcw} \mid w \in \{ a, b \}^* \} \cup \{ e \}
\]

and recall that \( \mathcal{L}_{\text{BNP}} = \mathcal{M}_\cap (\text{PAL}_e) \). It is not known whether \( \mathcal{L}_{\text{DUP}} = \mathcal{L}_{\text{BNP}} \), \( \mathcal{L}_{\text{DUP}} \subseteq \mathcal{L}_{\text{BNP}} \) [4], but \( \mathcal{L}_{\text{BNP}} \subseteq \mathcal{L}_{\text{DUP}} \) if and only if \( \text{PAL}_e \in \mathcal{L}_{\text{DUP}} \). It is known [4] that \( \mathcal{L}_{\text{DUP}} \) is closed under Kleene\(^+\) but in [5] it is conjectured that \( \mathcal{L}_{\text{BNP}} \) is not closed under Kleene\(^+\).

All of the results in Section 3 remain true if we substitute "homomorphic duplication" everywhere, although some comment is necessary regarding the analog of Proposition 3.2. It can be shown that \( \mathcal{L}_{\text{DUP}} \) contains for each finite alphabet \( \Sigma \), symbol \( c \notin \Sigma \) and positive integer \( k \), the language

\[
L_{\Sigma, c, k} = \{(a_1, b_1) \ldots (a_n, b_n) \mid n \geq 1, a_i, b_i \in \Sigma \cup \{ c \}, a_1 \ldots a_n \in (\Sigma^+ c^k)^+, h_c(a_1 \ldots a_n) = h_c(b_1 \ldots b_n)\}
\]

(where \( h_c \) is the homomorphism which erases \( c \) and is the identity elsewhere),

vol. 12, n° 4, 1978
and it is clear that any intersection-closed semiAFL containing each $L_{\Sigma,c,k} \cup \{ e \}$ (or even just $L_{(a,b),c,1} \cup \{ e \}$) is closed under linear erasing homomorphism [4]. (In fact, $L_{\text{DUP}} = \{ L_{(a,b),c,1} \cup \{ e \} \}$.)

REFERENCES