TM-738/053/00

Chains of Full AFLs

Sheila A. Greibach

SCIENTIFIC REPORT
NO. 26

12 May 1969

Prepared for
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH, USAF
BEDFORD, MASSACHUSETTS
and
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH, USAF
ARLINGTON, VIRGINIA

Contract Monitor Thomas V. Griffiths Data Sciences Laboratory
Monitored by: Major Russell Ives, SRIR (AFOSR) Supported by:
SDC's Independent Research Program:
Contract F1962867C0008, Programming (Algorithmic) Languages,
Project No. 5632, Task No. 563205; and Grant No. AF/AFOSR/1203-67A

Work Unit No. 56320501 29

AFCRL-69-0293

TM-738/053/00

SDC TM SERIES

The work reported herein was supported by SDC and Contract F1962867C0008, Programming (Algorithmic) Languages; Project No. 5632, Task No. 563205; Work Unit No. 56320501; and Grant No. AF-AFOSR-1203-67; and by Harvard University and Contract CF19628C0029.

SCIENTIFIC REPORT NO. 26

12 May 1969

CHAINS OF FULL AFLs

by
Sheila A. Greibach*

Monitored by: Contract Monitor
Thomas V. Griffiths
Data Sciences Laboratory
Major Russell Ives, USAF
SRIR (AFOSR)

Prepared for:

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH, USAF
BEDFORD, MASSACHUSETTS 01730
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH, USAF
ARLINGTON, VIRGINIA

SYSTEM
DEVELOPMENT
CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

*Harvard University, Cambridge, Massachusetts and Consultant, System Development Corporation

ABSTRACT

If a full AFL $\mathcal{L}$ is not closed under substitution, then $\mathcal{L} \hat{\sigma} \mathcal{L}$, the result of substituting members of $\mathcal{L}$ into $\mathcal{L}$, is not substitution closed and hence $\mathcal{L}$ generates an infinite hierarchy of full AFLs. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are two incomparable full AFLs, then the least full AFL containing $\mathcal{L}_1$ and $\mathcal{L}_2$ is not substitution closed. In particular, the substitution closure of any full AFL properly contained in the context-free languages is itself properly contained in the context-free languages. If any set of languages generates the context-free languages, one of its members must do so. The substitution closure of the one-way stack languages is properly contained in the nested stack languages. For each n, there is a class of full context-free AFLs whose partial ordering under inclusion is isomorphic to the natural partial ordering on n-tuples of positive integers.

CHAINS OF FULL AFLs*

1. Introduction

Recently there have been several investigations of the closure under substitution of various families of languages such as the linear context-free languages [10], [12], [5], counter languages [6] and stack languages [7], [11]. In some of these cases a chain of full AFL's, $\mathcal{L}_1, \ldots, \mathcal{L}_n, \ldots$ was exhibited such that each $\mathcal{L}_n$ is properly contained in $\mathcal{L}_{n+1}$, obtained from $\mathcal{L}_n$ by application of a substitution operator. Hence the infinite union $\bigcup_n \mathcal{L}_n$ is a full AFL which cannot be principal [4]. In this paper we shall show that for any full AFL $\mathcal{L}$, if $\mathcal{L}$ is not substitution closed then we can exhibit such a chain and hence the substitution closure of $\mathcal{L}$ is not principal. To make this precise we need a few definitions.

<u>Definition 1.1</u> A <u>full semi-AFL</u> is a family of languages containing at least one nonempty set and closed under union, homomorphism, inverse homomorphism and intersection with regular sets. A <u>full AFL</u> is a full semi-AFL closed under concatenation and Kleene closure.

<u>Definition 1.2</u> If $\mathfrak{D}$ is a family of languages, $\hat{\mathcal{F}}(\mathfrak{D})$ is the least full AFL containing $\mathfrak{D}$. If $\mathcal{L} = \hat{\mathcal{F}}(\mathfrak{D})$, then $\mathfrak{D}$ is a <u>core</u> of $\mathcal{L}$. If for each L in $\mathfrak{D}$, $\hat{\mathcal{F}}(\mathfrak{D}) \neq \hat{\mathcal{F}}(\mathfrak{D} - \{L\})$, then $\mathfrak{D}$ is <u>independent</u>. If $\mathfrak{D} = \{L\}$, then we write $\hat{\mathcal{F}}(L)$ for $\hat{\mathcal{F}}(\mathfrak{D})$ and $\mathcal{L} = \hat{\mathcal{F}}(L)$ is <u>full principal</u> and L is a <u>generator</u> of $\mathcal{L}$.

---

* Research sponsored in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, USAF, under contracts F-19628-67-C-0008 and F-19628-68-C-0029, and by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under AFOSR Grant No. AF-AFOSR-1203-67A and the Division of Engineering and Applied Physics of Harvard University.

<u>Definition 1.3</u> Let Σ_1 be finite and for each a in Σ_1 let $\tau(a)$ be a language. Let $\tau(e) = \{e\}$+. Let $\tau(a_1 \ldots a_n) = \tau(a_1) \ldots \tau(a_n)$, $a_i \in \Sigma_1$, and for $L \subseteq \Sigma_1^*$, let $\tau(L) = \bigcup_{w \in L} \tau(w)$. Then τ is a <u>substitution</u> on L. If each $\tau(a)$ is in $\mathcal{L}$ for $a \in \Sigma_1$, then τ is an $\mathcal{L}$-<u>substitution</u>. If each $\tau(a)$ is regular, then τ is a <u>regular substitution</u>. We let $\mathcal{R}$ be the family of regular sets.

<u>Definition 1.4</u> Let $\mathcal{L}_1 \hat{\sigma} \mathcal{L}_2 = \{\tau(L) \mid L \in \mathcal{L}_1, \tau \text{ an } \mathcal{L}_2\text{-substitution}\}$. $\mathcal{L}$ is <u>substitution closed</u> if $\mathcal{L} \hat{\sigma} \mathcal{L} \subseteq \mathcal{L}$. Let $\hat{\mathcal{S}}_0(\mathcal{L}) = \mathcal{R}$, $\hat{\mathcal{S}}_1(\mathcal{L}) = \mathcal{L}$, and $\hat{\mathcal{S}}_{n+1}(\mathcal{L}) = \hat{\mathcal{S}}_n(\mathcal{L}) \hat{\sigma} \mathcal{L}$. Let $\hat{\mathcal{S}}(\mathcal{L})$ be the least substitution closed full AFL containing $\mathcal{L}$.

Our general result will be that if $\mathcal{L}$ is a full semi-AFL that is not substitution closed, then $\mathcal{L} \hat{\sigma} \mathcal{L}$ is not substitution closed and $\hat{\mathcal{S}}(\mathcal{L})$ is not a full principal AFL. Since the context-free languages are a full principal AFL, if $\mathcal{L}$ is a full semi-AFL contained in the context-free languages but not substitution closed, then $\hat{\mathcal{S}}(\mathcal{L})$ is properly contained in the context-free languages. This provides an immediate proof of the result of Yntema [12] and Nivat [10] that the standard matching choice languages (the least substitution closed full AFL containing the linear context-free languages) are properly contained in the context-free languages, and yields the further result that the substitution closure of the linear context-free and the one counter languages is likewise properly contained in the context-free languages. Similar reasoning establishes that the one-way stack languages are properly contained in the nested stack languages [1].

+ We let A* be the monoid generated by A with identity e. A <u>language</u> is any subset of a finitely generated free monoid.

In order to prove that the one-way stack languages are not substitution closed, the author of this paper examined a special class of substitutions [7]. In the next section we prove a lemma about these substitutions which is used to establish the main result in Section 3. A variant of the basic lemma is used in Section 4 to show the companion result that if $\mathcal{L}_1$ and $\mathcal{L}_2$ are incomparable full semi-AFLs, then the least full AFL containing $\mathcal{L}_1$ and $\mathcal{L}_2$ is not substitution closed. This result yields many further chains of AFLs and shows that there is no way to factor the context-free languages into a finite number of subAFLs which jointly generate the whole family; the argument shows that any core of a substitution closed full principal AFL contains a generator and that a substitution closed full AFL that is not principal has no independent core. In Section 5 we extend the results of the previous sections to exhibit for each n a class of context-free full principal AFLs whose partial ordering under inclusion is isomorphic to the natural partial ordering on n-tuples of positive integers. The corresponding lattice is not a sublattice of the lattice of all context-free full AFLs with

l.u.b. $(\mathcal{L}_1, \mathcal{L}_2) = \hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$.

2. The τ_L^Σ Lemmas

In this section we establish a series of lemmas concerning particular types of substitutions. These lemmas will be used to yield the results in the following sections. In this and subsequent sections we shall freely use the results established in [8] that $\hat{\sigma}$ is associative (among full semi-AFLs), that $\mathcal{L}$ is substitution closed if and only if $\mathcal{L} = \mathcal{L} \hat{\sigma} \mathcal{L}$, and that $\hat{\mathcal{J}}(\mathcal{L}) = \bigcup_n \hat{\mathcal{J}}_n(\mathcal{L})$. Also, if $\mathcal{L}$ is a full semi-AFL, $\mathcal{L} = \mathcal{L} \hat{\sigma} \mathcal{R}$ and $\hat{\mathcal{F}}(\mathcal{L}) = \mathcal{R} \hat{\sigma} \mathcal{L}$ and if $\mathcal{L}$ is a full AFL, $\mathcal{L} = \mathcal{L} \hat{\sigma} \mathcal{R} = \mathcal{R} \hat{\sigma} \mathcal{L}$ [3].

Definition 2.1 Let $\Sigma_1 \cap \Sigma_2 = \phi$ and let $L_2 \subseteq \Sigma_2^*$. Let $\tau_{L_2}^{\Sigma_1}$ be the substitution defined by $\tau_{L_2}^{\Sigma_1}(a) = aL_2$ for each $a \in \Sigma_1$.

Lemma 2.1 Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be full semi-AFLs. Let $L_1 \subseteq \Sigma_1\Sigma_1^*$, $L_2 \subseteq \Sigma_2\Sigma_2^*$, $\Sigma_1 \cap \Sigma_2 = \phi$. If $\tau_{L_2}^{\Sigma_1}(L_1)$ is in $\mathcal{L}_1 \hat{\sigma} \mathcal{L}_2$, then either L_1 is in $\mathcal{L}_1$ or L_2 is in $\mathcal{L}_2$.

Proof

Let $L = \tau_{L_2}^{\Sigma_1}(L_1)$. If L is in $\mathcal{L}_1 \hat{\sigma} \mathcal{L}_2$, then $L = \tau(L_3)$ for some $L_3 \in \mathcal{L}_1$, where $L_3 \subseteq \Sigma_3^*$ and $\tau(a) \in \mathcal{L}_2$ for each $a \in \Sigma_3$. Since $\mathcal{L}_1$ is a full semi-AFL we can obviously assume that $\Sigma_3 \cap (\Sigma_1 \cup \Sigma_2) = \phi$ and that $\tau(a) \neq \phi$ for each $a \in \Sigma_3$. We shall think of $\tau(L_3)$ and $\tau_{L_2}^{\Sigma_1}(L_1)$ as representing alternative parsings of words in L. The idea of our proof is that the form of L forces such constraints on the $\tau(L_3)$ parsing that either L_1 can be recovered from L_3 by semi-AFL operations or else L_2 can be obtained from the $\tau(a)$ by semi-AFL operations.

For any w in L a factorization of w is an expression $a_1y_1 \ldots a_ny_n$, where $a_1 \ldots a_n \in L_3$, $y_i \in \tau(a_i)$ for each i and $w = y_1 \ldots y_n$. A factorization of w is a parsing of w as a member of $\tau(L_3)$; by its very form w is already uniquely parsed as a member of $\tau_{L_2}^{\Sigma_1}(L_1)$. Note that in this factorization, $n \geq 1$ since L does not contain e. Further, $w \in \tau(a_1 \ldots a_n)$ and each w in L has at least one factorization. For each $b_1 \ldots b_m$ in L_1, $m \geq 1$, b_i in Σ_1 and each z in L_2, let $\mu(b_1 \ldots b_m, z) = b_1z \ldots b_mz$. By definition of L, $\mu(b_1 \ldots b_m, z)$ is in L for each $b_1 \ldots b_m$ in L_1 and z in L_2.

We shall restrict attention to factorizations of the $\mu(b_1 \ldots b_m, z)$. We say that such a factorization $a_1y_1 \ldots a_ny_n$ <u>splits</u> if each y_i contains at most one b_k, that is, $y_i \in \Sigma_2^* \Sigma_1 \Sigma_2^* \cup \Sigma_2^*$. Let h_1 be the homomorphism defined by $h_1(b) = b$ for $b \in \Sigma_1$ and $h_1(c) = e$ elsewhere. Then the substitution defined by $\bar{\tau}(a) = h_1(\tau(a) \cap [\Sigma_2^* \Sigma_1 \Sigma_2^* \cup \Sigma_2^*])$ is regular since each $\bar{\tau}(a)$ is finite. Let $L_1' = \bar{\tau}(L_3)$. Since every full semi-AFL is closed under regular substitution [3], L_1' is in $_1$. Now observe that if the factorization $a_1y_1 \ldots a_ny_n$ of $\tau(b_1 \ldots b_m, z)$ splits, then $h_1(y_i) \in \bar{\tau}(a_i)$ for each i; thus $b_1 \ldots b_m \in \bar{\tau}(a_1 \ldots a_n) \subseteq \bar{\tau}(L_3) = L_1'$. On the other hand, $L_1' = \bar{\tau}(L_3) \subseteq h_1[\tau(L_3)] = L_1$. Thus L_1' is contained in L_1 and contains each $b_1 \ldots b_m$ in L_1 for which there exists a z in L_2 such that $\mu(b_1 \ldots b_m, z)$ has a splitting factorization.

Consider the following condition:

(A) For <u>every</u> $b_1 \ldots b_m$ in L_1, there is some z in L_2 such that $\mu(b_1 \ldots b_m, z)$ has a factorization that splits.

If (A) holds, then $L_1 = L_1'$ and L_1 is in $\mathcal{L}_1$. Since every $\mu(b_1 \ldots b_m, z)$ has at least one factorization, if (A) does not hold then surely the following must be true:

(B) There is a $b_1 \ldots b_m$ in L_1 such that for all z in L_2, $\mu(b_1 \ldots b_m, z)$ has a factorization that does not split.

Now suppose that $\mu(b_1 \ldots b_m, z)$ has a factorization $a_1y_1 \ldots a_ny_n$ that does not split. Thus for some i, y_i = xbucx' for some x in Σ_2^*, u in Σ_2^*, x' in $(\Sigma_1 \cup \Sigma_2)^*$ and b, c in Σ_1. But $\mu(b_1 \ldots b_m, z)$ is in L so that the only members of Σ_1 in

any y_i are the b_i and only z can appear between consecutive b_i. Hence u = z and for some k, $1 \leq k < n$, $b = b_k$ and $c = b_{k+1}$. Now if we let M be the gsm mapping which erases all symbols up to and including the first member of Σ_1, then acts as the identity until another member of Σ_1 is encountered whereupon it erases that symbol and all subsequent symbols, then clearly $z = M(xb_k z b_{k+1} x') = M(y_i)$. Now since all $\tau(a) \neq \phi$, if a appears in any word of L_3 and y is any member of $\tau(a)$, then y must appear in some word of L. Let Σ_3' contain all and only members of Σ_3 appearing in at least one member of L_3 and for each a, let $\tau'(a) = \tau(a) \cap [\Sigma_2^* \Sigma_1 \Sigma_2^* \Sigma_1 (\Sigma_1 \cup \Sigma_2)^*]$. The previous argument shows that for a in Σ_3' and y in $\tau'(a)$, M(y) is in L_2, and $\tau'(a)$ is in $\mathcal{L}_2$. Hence, if we let $L_2' = \bigcup_{a \in \Sigma_3'} M(\tau'(a))$, then $L_2' \subseteq L_2$, and L_2' is in $\mathcal{L}_2$ since every full semi-AFL is closed under gsm mapping [3].
Furthermore, if $\mu(b_1 \ldots b_m, z)$ has a factorization that does not split, then z is in L_2'. Therefore, if (B) holds, $L_2' = L_2$, so that L_2 is in $\mathcal{L}_2$.

Thus either (A) holds and L_1 is in $\mathcal{L}_1$, or (B) holds and L_2 is in $\mathcal{L}_2$.

Lemma 2.1 is a syntactic lemma regarding possible factorizations or parsings of certain languages. We shall find it more convenient to express it in the following form as a lemma regarding families of languages.

<u>Lemma 2.2</u> Let $n \geq 1$, and for each i, $1 \leq i \leq n$, let $\mathcal{L}_i$ and $\mathcal{L}_i'$ be full semi-AFLs such that $\mathcal{L}_i - \mathcal{L}_i' \neq \phi$. Then
$\mathcal{L}_1 \hat{\sigma} \mathcal{L}_2 \hat{\sigma} \ldots \hat{\sigma} \mathcal{L}_n - \mathcal{L}_1' \hat{\sigma} \mathcal{L}_2' \hat{\sigma} \ldots \hat{\sigma} \mathcal{L}_n' \neq \phi$.

Proof

We establish the result for $n = 2$; the general result then follows by an obvious induction on n. Assume $L_1 \in \mathcal{L}_1 - \mathcal{L}'_1$ and $L_2 \in \mathcal{L}_2 - \mathcal{L}'_2$. Since we are dealing with full semi-AFLs we can assume that $L_1 \subseteq \Sigma_1 \Sigma_1^*$, $L_2 \subseteq \Sigma_2 \Sigma_2^*$ with $\Sigma_1 \cap \Sigma_2 = \emptyset$. Let $L = \tau_{L_2}^{\Sigma_1}(L_1)$. Then by definition L is in $\mathcal{L}_1 \hat{\sigma} \mathcal{L}_2$. But by Lemma 2.1, L cannot be in $\mathcal{L}'_1 \hat{\sigma} \mathcal{L}'_2$, since L_1 is not in $\mathcal{L}'_1$ and L_2 is not in $\mathcal{L}'_2$.

For Section 4 we need a variant of Lemma 2.1, for which we give the following definitions.

Definition 2.2 If $\mathcal{L}_1$ and $\mathcal{L}_2$ are families of languages, let

$\mathcal{L}_1 \cup \mathcal{L}_2 = \{L \mid L \text{ in } \mathcal{L}_1 \text{ or } L \text{ in } \mathcal{L}_2\}$, and

$\mathcal{L}_1 \vee \mathcal{L}_2 = \{L_1 \cup L_2 \mid L_1 \text{ in } \mathcal{L}_1, L_2 \text{ in } \mathcal{L}_2\}$.

Lemma 2.3 Let $\Sigma_1 \cap \Sigma_2 = \emptyset$, $L_1 \subseteq \Sigma_1 \Sigma_1^*$, $L_2 \subseteq \Sigma_2 \Sigma_2^*$, and let $\mathcal{L}_1$, $\mathcal{L}_2$ be full semi-AFLs. If $\tau_{L_2}^{\Sigma_1}(L_1)$ is in $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$, then either L_1 is in $\hat{\mathcal{F}}(\mathcal{L}_1)$ or L_2 is in $\mathcal{L}_2$.

Proof

The proof is similar to the proof of Lemma 1. Note that $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) = \mathcal{R}\, \hat{\sigma}\, (\mathcal{L}_1 \cup \mathcal{L}_2)$ [3]. Again, we assume that $L = \tau_{L_2}^{\Sigma_1}(L_1) = \tau(R)$, $R \subseteq \Sigma_3^*$, $\Sigma_3 \cap (\Sigma_1 \cup \Sigma_2) = \emptyset$, $\tau(a) \neq \emptyset$ for a in Σ_3, where this time R is regular and for each a in Σ_3, either $\tau(a)$ is in $\mathcal{L}_1$ or $\tau(a)$ is in $\mathcal{L}_2$.

We define $\mu(b_1 \ldots b_m, z)$ and factorizations as before. This time we say that a factorization $a_1 y_1 \ldots a_n y_n$ of $\mu(b_1 \ldots b_m, z)$ <u>splits with respect to</u> $\mathcal{L}_1$

if for each i either $y_i \varepsilon \Sigma_2^* \ \Sigma_1 \ \Sigma_2^* \cup \Sigma_2^*$ or else $\tau(a_i)$ is in $\mathcal{L}_1$. We define h_1 as before ($h_1(b) = b$, b in Σ_1, $h_1(c) = e$, c in Σ_2) and let

$$\bar{\tau}(a) = \begin{cases} h_1(\tau(a)), & \tau(a) \text{ in } \mathcal{L}_1 \\ h_1(\tau(a) \cap (\Sigma_2^* \ \Sigma_1 \ \Sigma_2^* \cup \Sigma_2^*)), & \text{otherwise} \end{cases}$$

and $L_1' = \bar{\tau}(R)$. As before we note that $L_1' \subseteq h_1(\tau(R)) = L_1$ and that $b_1 \ldots b_m \ \varepsilon \ L_1'$ if any factorization of $\mu(b_1 \ldots b_m, z)$ splits with respect to $\mathcal{L}_1$. Furthermore, a full AFL is closed under substitution into regular sets and contains all regular sets [3]; thus each $\bar{\tau}(a)$ is in $\mathcal{L}_1$ and hence L_1' is in $\hat{\mathcal{F}}(\mathcal{L}_1)$. Hence, if we have:

(A) For every $b_1 \ldots b_m$ in L_1 there is a z in L_2 such that $\mu(b_1 \ldots b_m, z)$ splits with respect to $\mathcal{L}_1$

then $L_1 = L_1'$ and so $L_1 \ \varepsilon \ \hat{\mathcal{F}}(\mathcal{L}_1)$. If (A) does not hold, we surely must have:

(B) There is a $b_1 \ldots b_m$ in L_1 such that for every z in L_2 no factorization of $\mu(b_1 \ldots b_m, z)$ splits with respect to $\mathcal{L}_1$.

If the factorization $a_1y_1 \ldots a_ny_n$ of $\mu(b_1 \ldots b_m, z)$ does not split with respect to $\mathcal{L}_1$, then for some i, $\tau(a_i) \ \varepsilon \ \mathcal{L}_2$ and $y_i \ \varepsilon \ \Sigma_2^* \ \Sigma_1 \ \Sigma_2^* \ \Sigma_1 \ (\Sigma_1 \cup \Sigma_2)^*$, so that $y_i = xb_kzb_{k+1}x'$ as before. Let Σ_3' be the set of all a in Σ_3 such that a appears in some member of R and $\tau(a)$ is in $\mathcal{L}_2$. Define the gsm mapping M as before. Let $L_2' = \bigcup_{a \varepsilon \Sigma_3'} M(\tau(a) \cap [\Sigma_2^* \ \Sigma_1 \ \Sigma_2^* \ \Sigma_1 \ (\Sigma_1 \cup \Sigma_2)^*])$. As before,

L_2' is in $\mathcal{L}_2$ and $L_2' \subseteq L_2$. Also, if any factorization of $\mu(b_1 \dots b_m, z)$ does not split with respect to $\mathcal{L}_1$, then z is in L_2'. Hence, if (B) holds, $L_2 = L_2'$ so L_2 is in $\mathcal{L}_2$. But either (A) or (B) must hold.

We note the following result for full semi-AFLs.

<u>Lemma 2.4</u> Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be incomparable full semi-AFLs. Then $\mathcal{L}_1 \cup \mathcal{L}_2 \subsetneq \mathcal{L}_1 \vee \mathcal{L}_2 \subsetneq \hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$.

<u>Proof</u>

Let L_1 be in $\mathcal{L}_1 - \mathcal{L}_2$ and L_2 in $\mathcal{L}_2 - \mathcal{L}_1$. As before, we can assume $L_1 \subseteq \Sigma_1 \Sigma_1^*$ and $L_2 \subseteq \Sigma_2 \Sigma_2^*$ with $\Sigma_1 \cap \Sigma_2 = \phi$. Let c be a new symbol. Clearly, $L_1 \cup cL_2$ is in $\mathcal{L}_1 \vee \mathcal{L}_2 - \mathcal{L}_1 \cup \mathcal{L}_2$. Let $L = L_1cL_2$. Obviously, L is in $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$. We claim that L is not in $\mathcal{L}_1 \vee \mathcal{L}_2$. For suppose $L = L_3 \cup L_4$, L_3 in $\mathcal{L}_1$, L_4 in $\mathcal{L}_2$. Let $h_1(a) = a$, a in Σ_1, $h_2(a) = a$, a in Σ_2, $h_1(c) = h_2(c) = e$, $h_1(a) = e$, a in Σ_2, $h_2(a) = e$, a in Σ_1. Either there is a w in L_1 such that $wcL_2 \subseteq L_3$, in which case $L_2 = h_2(L_3)$ or for each w in L_1 there is a z in L_2 with wcz in L_4 in which case $L_1 = h_1(L_4)$. Hence, if L is in $\mathcal{L}_1 \vee \mathcal{L}_2$, either L_1 is in $\mathcal{L}_2$ or L_2 is in $\mathcal{L}_1$. Thus L is not in $\mathcal{L}_1 \vee \mathcal{L}_2$.

3. <u>Chains of Full AFLs</u>

We are now ready to establish the main results concerning substitution and AFL chains: namely that a full AFL that is not substitution closed generates under substitution an infinite ascending chain of full AFLs that are not substitution closed.

Theorem 3.1 Let $\mathcal{L}$ be a full semi-AFL. Then $\mathcal{L}$ is substitution closed if and only if $\mathcal{L} \hat{\sigma} \mathcal{L}$ is substitution closed.

Proof

Clearly $\mathcal{L} = \mathcal{L} \hat{\sigma} \mathcal{L}$ if $\mathcal{L}$ is substitution closed. Now if $\mathcal{L}$ is not substitution closed, $\mathcal{L} \hat{\sigma} \mathcal{L} - \mathcal{L} \neq \phi$ and hence by Lemma 2.2, $((\mathcal{L} \hat{\sigma} \mathcal{L}) \hat{\sigma} (\mathcal{L} \hat{\sigma} \mathcal{L})) - (\mathcal{L} \hat{\sigma} \mathcal{L}) \neq \phi$. Therefore, $\mathcal{L} \hat{\sigma} \mathcal{L}$ is not substitution closed.

Corollary 1 A full semi-AFL $\mathcal{L}$ is substitution closed if and only if $\hat{\mathcal{F}}(\mathcal{L})$ is substitution closed.

Proof

Since $\hat{\mathcal{F}}(\mathcal{L}) = \mathcal{R} \hat{\sigma} \mathcal{L}$, and every full AFL contains all regular sets [3], if $\mathcal{L}$ is substitution closed then $\hat{\mathcal{F}}(\mathcal{L}) = \mathcal{L}$. By Theorem 3.1, if $\mathcal{L}$ is not substitution closed, neither is $\mathcal{L} \hat{\sigma} \mathcal{L}$. But $\hat{\mathcal{F}}(\mathcal{L}) = \mathcal{R} \hat{\sigma} \mathcal{L} \subseteq \mathcal{L} \hat{\sigma} \mathcal{L}$, and if $\hat{\mathcal{F}}(\mathcal{L})$ is substitution closed, $\mathcal{L} \hat{\sigma} \mathcal{L} \subseteq \hat{\mathcal{F}}(\mathcal{L})$, so $\hat{\mathcal{F}}(\mathcal{L}) = \mathcal{L} \hat{\sigma} \mathcal{L}$. Hence if $\mathcal{L}$ is not substitution closed, neither is $\hat{\mathcal{F}}(\mathcal{L})$.

Corollary 2 If a full semi-AFL $\mathcal{L}$ is not a full AFL, then $\hat{\mathcal{F}}(\mathcal{L})$ is not substitution closed.

Corollary 2 to Theorem 3.1 tells us for example that in order to show that the least full AFL containing the linear context-free languages is not substitution closed it suffices to show that the linear languages are not closed under concatenation. In the corollaries to the next theorem and in the rest of the paper we assume the reader to be familiar with the definitions of "context-free," "linear context-free" [2], "stack automaton" and "nested stack automaton" [1]. We shall call a family of languages "context-free" if all of its members are context-free.

<u>Theorem 3.2</u> Let $\mathcal{L}$ be a full semi-AFL that is not substitution closed. Then for each $n \geq 0$, $\hat{\mathcal{J}}_n(\mathcal{L}) \subsetneq \hat{\mathcal{J}}_{n+1}(\mathcal{L})$ and $\hat{\mathcal{J}}(\mathcal{L})$ is not full principal.

<u>Proof</u>

Since the regular sets are substitution closed, if $\mathcal{L}$ is not substitution closed, then $\hat{\mathcal{J}}_0(\mathcal{L}) = \mathcal{R} \subsetneq \mathcal{L} = \hat{\mathcal{J}}_1(\mathcal{L})$. Now if $\hat{\mathcal{J}}_n(\mathcal{L})$ is substitution closed for any $n \geq 1$, then clearly $\hat{\mathcal{J}}(\mathcal{L}) = \hat{\mathcal{J}}(\hat{\mathcal{J}}_n(\mathcal{L})) = \hat{\mathcal{J}}_n(\mathcal{L})$, so that $\hat{\mathcal{J}}_n(\mathcal{L}) = \hat{\mathcal{J}}_{n+k}(\mathcal{L})$ and $\hat{\mathcal{J}}_{n+k}(\mathcal{L})$ is substitution closed for all $k \geq 0$. Now let n be the smallest positive integer such that $\hat{\mathcal{J}}_n(\mathcal{L})$ is substitution closed. By hypothesis, $n > 1$. Then $\hat{\mathcal{J}}_{n-1}(\mathcal{L})$ is not substitution closed. By Theorem 3.1, $\hat{\mathcal{J}}_{2n-2}(\mathcal{L}) = \hat{\mathcal{J}}_{n-1}(\mathcal{L}) \,\sigma\, \hat{\mathcal{J}}_{n-1}(\mathcal{L})$ is not substitution closed. But $2n-2 \geq n$, so by the previous argument $\hat{\mathcal{J}}_n(\mathcal{L})$ is not substitution closed. Hence, for all n, $\hat{\mathcal{J}}_n(\mathcal{L})$ is not substitution closed. Now if $\hat{\mathcal{J}}_n(\mathcal{L}) = \hat{\mathcal{J}}_{n+1}(\mathcal{L})$, then by induction $\hat{\mathcal{J}}_{2n}(\mathcal{L}) = \hat{\mathcal{J}}_n(\mathcal{L})$, so $\hat{\mathcal{J}}_n(\mathcal{L})$ is substitution closed. Therefore, $\hat{\mathcal{J}}_n(\mathcal{L}) \subsetneq \hat{\mathcal{J}}_{n+1}(\mathcal{L})$ for all $n \geq 0$. Since $\hat{\mathcal{J}}_0(\mathcal{L}) \subsetneq \hat{\mathcal{J}}_1(\mathcal{L}) \subsetneq \hat{\mathcal{J}}_2(\mathcal{L}) \ldots$

is an infinite strictly increasing chain of full AFLs, $\hat{\mathcal{S}}(\mathcal{L})$ is not full principal [4].

Corollary 1 If a context-free full semi-AFL is not substitution closed, then its substitution closure is properly contained in the context-free languages.

Proof

The context-free languages form a full principal AFL closed under substitution [2].

Corollary 2 The substitution closure of the linear context-free languages is properly contained in the context-free languages.

Corollary 2 only says that some context-free language is not contained in the substitution closure $\mathcal{S}$ of the linear languages. More delicate methods are needed to establish the precise result of Nivat [10] and Yntema [12] that there is a one counter language not contained in $\mathcal{S}$; we shall use that result in Section 5 to establish infinite hierarchies of context-free AFLs.

Corollary 3 The substitution closure of the one-way stack languages is properly contained in the nested stack languages.

Proof

The one-way stack languages are not substitution closed [7], [11]. The nested stack languages [1] form a full principal AFL [4].

In the next section we shall use Lemmas 2.3 and 2.4 to establish results on incomparable full AFLs analagous to Theorem 3.1 and to examine the question of independent cores for substitution closed full AFLs.

4. Substitution and Cores of Full AFLs

In this section we use Lemmas 2.2 and 2.4 to show that if $\mathcal{L}_1$ and $\mathcal{L}_2$ are incomparable full semi-AFLs then $\hat{\mathcal{J}}(\mathcal{L}_1 \cup \mathcal{L}_2)$ is not substitution closed. This yields some interesting results on the cores of substitution closed full AFLs as well as further chains of full AFLs.

Theorem 4.1 Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be incomparable full semi-AFLs. Then (1) $\mathcal{L}_1 \hat{\sigma} \mathcal{L}_2$ and $\mathcal{L}_2 \hat{\sigma} \mathcal{L}_1$ are incomparable, (2) $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$ is not substitution closed, and (3) $\hat{\mathcal{J}}(\mathcal{L}_1 \cup \mathcal{L}_2)$ is not full principal.

Proof

Part (1) is a direct application of Lemma 2.2, since $\mathcal{L}_1 - \mathcal{L}_2 \neq \emptyset$ and $\mathcal{L}_2 - \mathcal{L}_1 \neq \emptyset$. Now $\hat{\mathcal{J}}(\mathcal{L}_1 \cup \mathcal{L}_2) = \hat{\mathcal{J}}(\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2))$, so that statement (3) follows directly from (2) by Theorem 3.2. Since $\mathcal{L}_1$ and $\mathcal{L}_2$ are incomparable, each must contain nonregular sets, so $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) - \mathcal{R} \neq \emptyset$. By Lemma 2.4, $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) - (\mathcal{L}_1 \vee \mathcal{L}_2) \neq \emptyset$. Now $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) = \hat{\mathcal{F}}(\mathcal{L}_1 \vee \mathcal{L}_2) = \mathcal{R} \hat{\sigma} (\mathcal{L}_1 \vee \mathcal{L}_2)$, and $\mathcal{L}_1 \vee \mathcal{L}_2$ is a full semi-AFL. Hence by Lemma 2.2, $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) \hat{\sigma} \hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) - \hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2) \neq \emptyset$, so that $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$ is not substitution closed.

Corollary 1 Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be incomparable context-free full semi-AFLs. Then $\hat{\mathcal{J}}(\mathcal{L}_1 \cup \mathcal{L}_2)$ is properly contained in the context-free languages.

Corollary 2 The substitution closure of the linear context-free languages and the one counter languages is properly contained in the context-free languages.

If we are dealing with full AFLs, Lemma 2.3 yields a further condition we shall use in the next section.

Theorem 4.2 Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be incomparable full AFLs. Then $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$ is properly contained in $\mathcal{L}_2 \hat{\sigma} \mathcal{L}_1$.

Proof

Let L_1 be in $\mathcal{L}_1 - \mathcal{L}_2$ and L_2 in $\mathcal{L}_2 - \mathcal{L}_1$. We can assume $L_1 \subseteq \Sigma_1 \Sigma_1^*$ and $L_2 \subseteq \Sigma_2 \Sigma_2^*$ and $\Sigma_1 \cap \Sigma_2 = \emptyset$. Let $L = \tau_{L_1}^{\Sigma 2}(L_2)$. By definition L is in $\mathcal{L}_2 \hat{\sigma} \mathcal{L}_1$. If L is in $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$, then by Lemma 2.3, either L_2 is in $\mathcal{L}_1$ or L_1 is in $\mathcal{L}_2$, contradicting the definition of L_1 and L_2. Hence L is not in $\hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$.

We use Theorem 4.1 to show that if $\mathcal{L}$ is a substitution closed full AFL, either $\mathcal{L}$ has no independent core or else every core contains a generator.

Theorem 4.3 Let $\mathcal{L}$ be a substitution closed full AFL and let $\mathcal{D}$ be any core of $\mathcal{L}$. If $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, either $\mathcal{D}_1$ or $\mathcal{D}_2$ is a core of $\mathcal{L}$.

Proof

Let $\mathcal{L}_1 = \hat{\mathcal{F}}(\mathcal{D}_1)$ and $\mathcal{L}_2 = \hat{\mathcal{F}}(\mathcal{D}_2)$. Then $\mathcal{L} = \hat{\mathcal{F}}(\mathcal{D}_1 \cup \mathcal{D}_2) = \hat{\mathcal{F}}(\mathcal{L}_1 \cup \mathcal{L}_2)$. Since $\mathcal{L}$ is substitution closed, $\mathcal{L}_1$ and $\mathcal{L}_2$ cannot be incomparable. Hence either $\mathcal{L}_1 \subseteq \mathcal{L}_2 = \mathcal{L}$ or $\mathcal{L}_2 \subseteq \mathcal{L}_1 = \mathcal{L}$, and either $\mathcal{D}_1$ or $\mathcal{D}_2$ is a core of $\mathcal{L}$.

<u>Corollary 1</u> If $\mathcal{L}$ is a substitution closed full AFL that is not full principal, then $\mathcal{L}$ has no independent core.

<u>Proof</u>

If $\mathcal{D}$ is a core of $\mathcal{L}$, and $L \in \mathcal{D}$, either $\{L\}$ or $\mathcal{D} - \{L\}$ must be a core. But since $\mathcal{L}$ is not full principal, $\{L\}$ cannot be a core.

<u>Corollary 2</u> If $\mathcal{L}$ is a substitution closed full principal AFL, any core of $\mathcal{L}$ must contain a generator of $\mathcal{L}$.

<u>Proof</u>

Let $\mathcal{D}$ be a core of $\mathcal{L}$. Since $\mathcal{L}$ is full principal, $\mathcal{L} = \hat{\mathcal{F}}(L)$ for some L in $\mathcal{L}$. Hence for some finite $\mathcal{D}' \subseteq \mathcal{D}$, $L \in \hat{\mathcal{F}}(\mathcal{D}')$, so $\mathcal{D}'$ is a core. Let $\mathcal{D}_1$ be a finite independent core contained in $\mathcal{D}'$. By Theorem 4.3, $\mathcal{D}_1$ must be of size one, that is, $\mathcal{D}_1 = \{L_1\}$ where L_1 is in $\mathcal{D}$ and is a generator of $\mathcal{L}$.

<u>Corollary 3</u> If $\mathcal{L}$ is a substitution closed full AFL, either $\mathcal{L}$ has no independent core or else every core contains a generator.

Theorem 4.3 and its corollaries yield examples of full AFLs (such as the context-free languages) with no independent core of size larger than one. These could be called essentially full principal since every core contains a single generator. The regular sets, the context-free languages, the

nested stack languages and the recursively enumerable languages all form essentially full AFLs. For example, if K_2 is the Dyck set on two letters and R is any regular set, then either $K_2 \cap R$ or $K_2 - (K_2 \cap R)$ must generate the context-free languages. Another consequence of this is that if L is any generator of the context-free languages then the least full semi-AFL containing L is a full AFL, that is, closed under concatenation and Kleene closure. This is not always true of AFLs that are not substitution closed; for example, the Dyck set on one letter, K_1, generates the one counter languages [6], but the least full semi-AFL containing K_1 is not a full AFL.

The results of Section 5 yield context-free independent cores of size n for each finite n. It is an open question if there is any full AFL with an infinite independent core. We conjecture that no context-free full AFL can have an infinite independent core and indeed that there is no infinite class of mutually incomparable context-free full AFLs. If $L_k = \{a^{n^k} \mid n \geq 1\}$, we conjecture that $\mathfrak{D} = \{L_k \mid k \geq 1\}$ is an independent core of $\hat{\mathcal{F}}(\mathfrak{D})$.

We conclude this section by showing that a substitution closed full principal AFL always contains a unique maximal subAFL, unless it is equal to the regular sets.

<u>Theorem 4.4</u> Let $\mathcal{L}$ be a substitution closed full principal AFL properly containing the regular sets. Then there is a unique maximal full AFL properly contained in $\mathcal{L}$.

Proof

Let $\mathcal{L}' = \{L \text{ in } \mathcal{L} \mid \hat{\mathcal{F}}(L) \neq \mathcal{L}\}$. Clearly $\mathcal{L}'$ contains all full AFLs properly contained in $\mathcal{L}$. Since $\mathcal{L}$ is full principal it properly contains $\mathcal{L}'$. Thus, if $\mathcal{L}'$ is a full AFL it will be the unique maximal full AFL properly contained in $\mathcal{L}$. Since $\mathcal{R} \neq \mathcal{L}$, $\mathcal{L}'$ is nonempty. Consider $\hat{\mathcal{F}}(\mathcal{L}')$. If $\hat{\mathcal{F}}(\mathcal{L}')$ is not contained in $\mathcal{L}'$, then it contains a generator of $\mathcal{L}$, so that $\hat{\mathcal{F}}(\mathcal{L}') = \mathcal{L}$. Then $\mathcal{L}'$ is a core of $\mathcal{L}$. By Corollary 2 of Theorem 4.3, $\mathcal{L}'$ must already contain a generator of $\mathcal{L}$. This contradicts the definition of $\mathcal{L}'$. Hence $\hat{\mathcal{F}}(\mathcal{L}')$ must be contained in $\mathcal{L}'$. So $\mathcal{L}'$ is a full AFL.

Remarks

Theorem 4.4 above shows that there is a maximal full AFL properly contained in the context-free languages. We conjecture that this full AFL is the full AFL described in Corollary 2 to Theorem 4.1, namely the substitution closure of the linear context-free languages and the one counter languages, and that this is also the intersection of the context-free languages with the one-way nonerasing stack languages; its members are all one-way nonerasing stack languages, in fact, checking automata languages [7].

Other conditions besides substitution closure imply that $\mathcal{L}'$ is a full AFL and hence the unique maximal full AFL properly contained in $\mathcal{L}$. Families having maximal subAFLs include the one-way stack languages and the least full AFL containing the linear context-free languages. It is an interesting question to identify some of these subAFLs. We can also turn the question around and ask if, for example, $\mathcal{R} = \mathcal{L}'$ for any $\mathcal{L}$. Equivalently, is there any full AFL covering the regular sets, in the sense that it properly contains no full AFL properly containing the regular sets?

5. Lattices of Full AFLs.

In this section we generalize the results of [6] by showing that for each n there is a class of context-free full principal AFLs whose partial ordering under inclusion is isomorphic to the natural partial ordering of n-tuples of positive integers. To do so, we need the existence of two full context-free AFLs which might be called "strongly independent;" not only they but also their substitution closures are incomparable.

Lemma 5.1 There are full context-free AFLs $\mathcal{L}_1$ and $\mathcal{L}_2$ such that

(1) $\mathcal{L}_1 - \mathcal{L}_1 \hat{\sigma} \mathcal{L}_1 \neq \phi \neq \mathcal{L}_2 - \mathcal{L}_2 \hat{\sigma} \mathcal{L}_2$, and

(2) $\mathcal{L}_1 - \hat{\mathcal{J}}(\mathcal{L}_2) \neq \phi \neq \mathcal{L}_2 - \hat{\mathcal{J}}(\mathcal{L}_1)$.

Proof

Let $\mathcal{L}_1$ be the one counter languages and $\mathcal{L}_2$ the least full AFL containing the linear context-free languages. Neither $\mathcal{L}_1$ nor $\mathcal{L}_2$ is substitution closed [6]. The Dyck set on one letter is in $\mathcal{L}_1 - \hat{\mathcal{J}}(\mathcal{L}_2)$ [10], [12], and the set of palindromes on two letters is in $\mathcal{L}_2 - \hat{\mathcal{J}}(\mathcal{L}_1)$ [6].

In the sequel we assume that $\mathcal{L}_1$ and $\mathcal{L}_2$ are full AFLs with the properties listed in Lemma 5.1. For convenience, we introduce the following notation, used in the rest of this section. We let $p(n) = 1$ if n is odd and $p(n) = 2$

if n is even. Then we define $\mathcal{L}_{<m>} = \hat{\mathcal{J}}_{3m}(\mathcal{L}_1)$ and $\mathcal{L}_{<m_1, \ldots, m_n, t>} = \mathcal{L}_{<m_1, \ldots, m_n>} \hat{\sigma} \hat{\mathcal{J}}_{3t}(\mathcal{L}_{p(n+1)})$. We define $(m_1, \ldots, m_n) \leq (k_1, \ldots, k_n)$ if and only if $m_i \leq k_i$ for $1 \leq i \leq n$. This is the natural partial order on n-tuples of positive integers. We shall show that $\mathcal{L}_{<m_1, \ldots, m_n>} \subseteq \mathcal{L}_{<k_1, \ldots, k_n>}$ if and only if $(m_1, \ldots, m_n) \leq (k_1, \ldots, k_n)$. First we establish two preliminary lemmas. We shall use the fact that $\hat{\mathcal{J}}_0(\mathcal{L}) = \mathcal{L}$ and for a full AFL $\mathcal{L}$, $\mathcal{R} \hat{\sigma} \mathcal{L} = \mathcal{L}$ as well as the associativity of $\hat{\sigma}$. Thus, for example, $\mathcal{L}_{<0,m,n>} = \hat{\mathcal{J}}_{3m}(\mathcal{L}_2) \hat{\sigma} \mathcal{L}_{<n>}$, $\mathcal{L}_{<0,0,m,n>} = \mathcal{L}_{<m,n>}$, and $\mathcal{L}_{<m,n,r,s>} = \mathcal{L}_{<m,n>} \hat{\sigma} \mathcal{L}_{<r,s>}$.

<u>Lemma 5.2</u> Let $n \geq 1$, and $m_i \geq 1$ for $1 \leq i \leq n$. Then for any nonnegative integers $k_1, \ldots, k_n$,

(1) $\mathcal{L}_{<m_1, \ldots, m_n>} \hat{\sigma} \mathcal{L}_{p(n+1)} - \mathcal{L}_{<k_1, \ldots, k_n>} \neq \emptyset$, and

(2) $\mathcal{L}_2 \hat{\sigma} \mathcal{L}_{<m_1, \ldots, m_n>} - \mathcal{L}_{<k_1, \ldots, k_n>} \neq \emptyset$, and
$\mathcal{L}_1 \hat{\sigma} \mathcal{L}_{<0, m_1, \ldots, m_n>} - \mathcal{L}_{<0, k_1, \ldots, k_n>} \neq \emptyset$.

<u>Proof</u>

We shall show (1) by induction on n; the proof of (2) is similar, despite the asymmetry caused by the notation. For n = 1, observe that $\mathcal{L}_{<m_1>} - \mathcal{R} \neq \emptyset$ since $\mathcal{L}_1$ and thus $\mathcal{L}_{<m_1>}$ contain nonregular sets. We defined p(2) = 2. Since $\mathcal{L}_{<k_1>} \subseteq \hat{\mathcal{J}}(\mathcal{L}_1)$ and $\mathcal{L}_2 - \hat{\mathcal{J}}(\mathcal{L}_1) \neq \emptyset$ by hypothesis, then $\mathcal{L}_{p(2)} - \mathcal{L}_{<k_1>} \neq \emptyset$. Hence by Lemma 2.2, $\mathcal{L}_{<m_1>} \hat{\sigma} \mathcal{L}_{p(2)} - \mathcal{L}_{<k_1>} \neq \emptyset$, since $\mathcal{L}_{<k_1>} = \mathcal{R} \hat{\sigma} \mathcal{L}_{<k_1>}$.

Suppose we have shown (1) for some $n \geq 1$. We have the following inclusions:

(3) $\mathcal{L}_{<m_1, \ldots, m_n>} \hat{\sigma} \mathcal{L}_{p(n+1)} \subseteq \mathcal{L}_{<m_1, \ldots, m_{n+1}>} = \mathcal{L}_{<m_1, \ldots, m_n>} \hat{\sigma} \hat{\mathcal{J}}_{3m_{n+1}}(\mathcal{L}_{p(n+1)})$.

and

(4) $\mathcal{L}_{<k_1, \ldots, k_{n+1}>} \subseteq \mathcal{L}_{<k_1, \ldots, k_n>} \hat{\sigma} \hat{\mathcal{J}}(\mathcal{L}_{p(n+1)})$.

Combining the induction hypothesis and (3) yields:

(5) $\mathcal{L}_{<m_1, \ldots, m_{n+1}>} - \mathcal{L}_{<k_1, \ldots, k_n>} \neq \phi$.

By assumption $\mathcal{L}_{p(n+2)} - \hat{\mathcal{J}}(\mathcal{L}_{p(n+1)}) \neq \phi$. Hence by Lemma 2.2:

(6) $\mathcal{L}_{<m_1, \ldots, m_{n+1}>} \hat{\sigma} \mathcal{L}_{p(n+2)} - \mathcal{L}_{<k_1, \ldots, k_n>} \hat{\sigma} \hat{\mathcal{J}}(\mathcal{L}_{p(n+1)}) \neq \phi$.

Combining (6) and (4) yields (1) for $n + 1$, namely:

$$\mathcal{L}_{<m_1, \ldots, m_{n+1}>} \hat{\sigma} \mathcal{L}_{p(n+2)} - \mathcal{L}_{<k_1, \ldots, k_{n+1}>} \neq \phi.$$

<u>Lemma 5.3</u> Let $n \geq 1$, $m_j \geq 1$, $k_j \geq 1$ for $1 \leq j \leq n$ and let $m_t = k_t + r$ for some $r \geq 1$ and some t, $1 \leq t \leq n$. Then $\mathcal{L}_{<m_1, \ldots, m_n>} - \mathcal{L}_{<k_1, \ldots, k_n>} \neq \phi$.

<u>Proof</u>

If $n = 1$, Theorem 3.2 and the fact that $\mathcal{L}_1$ is not substitution closed yields the desired result. Now let $n > 1$. We define $\mathcal{F}_1 = \mathcal{L}_{<m_1, \ldots, m_{t-1}>} \hat{\sigma} \mathcal{L}_{p(t)}$ and $\mathcal{F}_2 = \mathcal{L}_{p(t)} \hat{\sigma} \mathcal{L}_{<0, \ldots, 0, m_{t+1}, \ldots, m_n>}$, where $\mathcal{L}_{<m_1, \ldots, m_{t-1}>}$ is taken as $\mathcal{R}$ if $t = 1$ and similarly $\mathcal{L}_{<0, \ldots, 0, m_{t+1}, \ldots, m_n>}$ is taken as $\mathcal{R}$ if $t = r$.

By Lemma 5.2, $\mathcal{F}_1 - \mathcal{L}_{<k_1, \ldots, k_{t-1}>} \neq \phi \neq \mathcal{F}_2 - \mathcal{L}_{<0, \ldots, 0, k_{t+1}, \ldots, k_n>}$.

Since $\mathcal{L}_{p(t)}$ is not substitution closed and $k_t = m_t - r$, we have by Theorem 3.2 $\hat{\mathcal{J}}_{3k_t}(\mathcal{L}_{p(t)}) = \hat{\mathcal{J}}_{3m_t-3r}(\mathcal{L}_{p(t)}) \subsetneq \hat{\mathcal{J}}_{3m_t-2}(\mathcal{L}_{p(t)})$. Now

$\mathcal{L}_{<k_1, \ldots, k_n>} = \mathcal{L}_{<k_1, \ldots, k_{t-1}>} \hat{\sigma} \hat{\mathcal{J}}_{3k_t}(\mathcal{L}_{p(t)}) \hat{\sigma} \mathcal{L}_{<0, \ldots, 0, k_{t+1}, \ldots, k_n>}$.

Similarly, $\mathcal{L}_{<m_1, \ldots, m_n>} = \mathcal{F}_1 \hat{\sigma} \hat{\mathcal{J}}_{3m_t-2}(\mathcal{L}_{p(t)}) \hat{\sigma} \mathcal{F}_2$. Hence by Lemma 2.2,

$\mathcal{L}_{<m_1, \ldots, m_n>} - \mathcal{L}_{<k_1, \ldots, k_n>} \neq \phi$.

<u>Lemma 5.4</u> Let $n \geq 1$, $m_j, k_j \geq 1$ for $1 \leq j \leq n$. Then $\mathcal{L}_{<m_1, \ldots, m_n>} \subseteq \mathcal{L}_{<k_1, \ldots, k_n>}$ if and only if $(m_1, \ldots, m_n) \leq (k_1, \ldots, k_n)$.

<u>Proof</u>

Since $\mathcal{L}_i \subseteq \mathcal{L}_i \hat{\sigma} \mathcal{L}_i$ for $i = 1, 2$ and $\hat{\sigma}$ is associative, it is clear that if $(m_1, \ldots, m_n) \leq (k_1, \ldots, k_n)$, then $\mathcal{L}_{<m_1, \ldots, m_n>} \subseteq \mathcal{L}_{<k_1, \ldots, k_n>}$. Now if we do not have $(m_1, \ldots, m_n) \leq (k_1, \ldots, k_n)$, then for some t, $m_t \geq k_t + 1$. Then by Lemma 5.3, $\mathcal{L}_{<m_1, \ldots, m_n>} - \mathcal{L}_{<k_1, \ldots, k_n>} \neq \phi$.

The four lemmas give us the appropriate lattices.

<u>Theorem 5.1</u> For each integer $n \geq 1$, there is a class of context-free full principal AFLs whose partial ordering under inclusion is isomorphic to the natural partial ordering of n-tuples of positive integers.

Proof

Let $\mathcal{L}_1$ be the one counter languages and $\mathcal{L}_2$ the least full AFL containing the linear context-free languages. Let the $\mathcal{L}_{<m_1, \ldots, m_n>}$ be defined as above. Then Lemma 5.4 shows that this class of families of languages has the desired partial ordering under inclusion. Now $\mathcal{L}_1$ and $\mathcal{L}_2$ are full principal context-free AFLs and hence so are the $\mathcal{L}_{<m_1, \ldots, m_n>}$. [2], [4], [6].

Corollary For each n, there is a full principal context-free AFL with an independent core of size n.

For each n, the $\mathcal{L}_{<m_1, \ldots, m_n>}$ form a lattice with

l.u.b. $(\mathcal{L}_{<m_1, \ldots, m_n>}, \mathcal{L}_{<k_1, \ldots, k_n>}) = \mathcal{L}_{<\mathrm{Max}(m_1, k_1), \ldots, \mathrm{Max}(m_n, k_n)>}$

and the g.l.b. similarly defined. Theorem 4.2 says that this lattice is not a sublattice of the lattice of all context-free full AFLs (with l.u.b. $(\mathcal{F}_1, \mathcal{F}_2) = \hat{\mathcal{F}}(\mathcal{F}_1 \cup \mathcal{F}_2)$) which in turn is not a sublattice of the lattice of all families of context-free languages (with l.u.b. $(\mathcal{F}_1, \mathcal{F}_2) = \mathcal{F}_1 \cup \mathcal{F}_2$), since for example,

$$\mathcal{L}_{<1, 2>} \cup \mathcal{L}_{<2, 1>} \subsetneqq \mathcal{L}_{<1, 2>} \vee \mathcal{L}_{<2, 1>} \subsetneqq \hat{\mathcal{F}}(\mathcal{L}_{<1, 2>} \cup \mathcal{L}_{<2, 1>}) \subsetneqq \mathcal{L}_{<2, 2>}.$$

We conjecture that this also holds for the g.l.b., that for example,

$$\mathcal{L}_{<1, 1>} \subsetneqq \mathcal{L}_{<1, 2>} \cap \mathcal{L}_{<2, 1>}.$$

Lemma 5.1 demonstrates the existence of two incomparable substitution closed full AFLs. We conjecture that Lemma 5.1 cannot be extended to three AFLs, that is, there are no three full context-free AFLs $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ such that no $\mathcal{L}_i$ is substitution closed, and for $i \neq j$, $1 \leq i, j \geq 3$, $\mathcal{L}_i$ and $\hat{\mathcal{J}}(\mathcal{L}_j)$ are incomparable. This conjecture is obviously related to the one in Section 4 regarding the maximal full AFL properly contained in the context-free languages.

References

1. Aho, A. V., "Indexed Grammars, an Extension of Context Free Grammars," Journal of the Association for Computing Machinery, Vol. 15, No. 4 (1968), pp. 647-671.

2. Chomsky, N. and M. P. Schutzenberger. "The Algebraic Theory of Context Free Languages," in P. Braffort and D. Hirschberg (eds.), Computer Programming and Formal Systems. Amsterdam: North Holland Publishing Co., 1963, pp. 118-161.

3. Ginsburg, S. and S. Greibach. "Abstract Families of Languages," to appear in a Memoir of the American Mathematical Society.

4. ________ "Principal AFL" (in preparation).

5. Ginsburg, S. and E. Spanier. "Derivation-Bounded Languages," IEEE Ninth Annual Symposium on Switching and Automata Theory, October 1968, pp. 306-314.

6. Greibach, S. "An Infinite Hierarchy of Context-Free Languages," Journal of the Association for Computing Machinery, Vol. 16, No. 1 (1969), pp. 91-106.

7. ________ "Checking Automata and One-Way Stack Languages," Journal of Computer and Systems Sciences, April 1, 1968, pp. 196-217.

8. Greibach, S. and S. Ginsburg. "Multitape AFA" (in preparation).

9. Greibach, S. and J. Hopcroft. "Independence of AFL Operations," to appear in a Memoir of the American Mathematical Society.

10. Nivat, M. "Transductions des languages de Chomsky," Grenoble University Thesis, 1967.

11. Schkolnick, M. "Two-Type Bracketed Grammars," IEEE Ninth Annual Symposium on Switching and Automata Theory, October 1968, pp. 315-326.

12. Yntema, M. K. "Inclusion Relations Among Families of Context-Free Languages," Information and Control, Vol. 10 (1967), pp. 572-597.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation Santa Monica, California 90406	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

CHAINS OF FULL AFLS.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

"Scientific Interim."

5. AUTHOR(S) (First name, middle initial, last name)

Sheila A. Greibach

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
12 May 1969	26	12

8a. CONTRACT OR GRANT NO. F1962867C0008, Grant No. AF-AFOSR-1203-67A b. PROJECT NO. 5632 Task No. 563205 c. Work Unit No. 56320501 DoD Element: 6144501F d. DoD Subelement: 681305	9a. ORIGINATOR'S REPORT NUMBER(S) TM-738/053/00 Scientific Report No. 26 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFCRL-69-0293

10. DISTRIBUTION STATEMENT

1- Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Air Force Office of Scientific Research Office of Aerospace Research, USAF Arlington, Virginia	Air Force Cambridge Research Laboratories (CRB) L.G. Hanscom Field Bedford, Massachusetts 01730

13. ABSTRACT

If a full ~~AFL~~ $\mathcal{L}$ is not closed under substitution, then $\cancel{\mathcal{L} \hat{\sigma} \mathcal{L}}$, the result of substituting members of $\mathcal{L}$ into $\mathcal{L}$, is not substitution closed and hence $\mathcal{L}$ generates an infinite hierarchy of full AFLs. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are two incomparable full AFLs, then the least full AFL containing $\mathcal{L}_1$ and $\mathcal{L}_2$ is not substitution closed. In particular, the substitution closure of any full AFL properly contained in the context-free languages is itself properly contained in the context-free languages. If any set of languages generates the context-free languages, one of its members must do so. The substitution closure of the one-way stack languages is properly contained in the nested stack languages. For each n, there is a class of full context-free AFLs whose partial ordering under inclusion is isomorphic to the natural partial ordering on n-tuples of positive integers.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
AFL Abstract families of languages Context free languages Stack languages						

Security Classification