ABSTRACT FAMILIES OF DETERMINISTIC LANGUAGES

W. J. Chandler

University of Southern California
Los Angeles, California

System Development Corporation
Santa Monica, California

Abstract

Abstract families of (one-way) deterministic acceptors are formalized and shown to be characterized by families of languages closed under marked union, marked *, and inverse marked gsm mapping. The independence of these closure operations is also proved.

Introduction

In recent years, automata theorists have worked towards developing generalized models of automata. Families of one-way nondeterministic acceptors have been shown to be characterized by a set of closure properties.[2] Balloon automata were formulated to model general classes of automata.[6] In this paper, we formalize the notion of an abstract family of (one-way) deterministic acceptors, and characterize the related family of languages by closure properties. This and related results are presented in three sections.

In the first section, we recall several definitions and concepts from formal language theory. Then we define a marked generalized sequential machine (abbreviated mgsm), which is basically a generalized sequential machine with a right endmarker. We then define an abstract family of deterministic languages (abbreviated AFDL) as a family of languages which is closed under marked union, marked * (marked Kleene closure), and inverse mgsm mapping.

In the second section, an abstract family of (one-way) deterministic acceptors (abbreviated AFDA) is introduced. The main result of the paper is then proved, namely that a family of languages is an AFDL if and only if there exists an AFDA which accepts exactly that family. Finally, we consider quasi-realtime and memory bounded AFDA, and show that the family of languages accepted by each is an AFDL.

In the final section, by exhibiting three counterexamples, we prove that the three operations in the definition of an AFDL are mutually independent. This implies that an arbitrary AFDL cannot be defined by only two of these operations.

Section 1. Preliminaries

In this section, we introduce the families of languages with which we shall be concerned, namely the abstract families of deterministic languages (AFDL). We shall also present some of their elementary closure properties.

First, we shall recall some well-known concepts from formal language theory.

<u>Definition</u>. A <u>family of languages</u> is an ordered pair $(\Sigma, \mathcal{L})$, or $\mathcal{L}$ when Σ is understood, where:

(1) Σ is an infinite set of symbols,

(2) $\mathcal{L}$ is a family of subsets of Σ^*, (1)

(3) $L \neq \emptyset$ for some L in $\mathcal{L}$, and

(4) for each L in $\mathcal{L}$, there exists a finite set $\Sigma_1 \subset \Sigma$ such that $L \subseteq \Sigma_1^*$.

Condition (3) of the previous definition is used only for avoiding the trival family of languages (i.e., $\mathcal{L}$ only contains the empty set). In the sequel, Σ will denote an infinite set (not necessarily countable) of symbols, and Σ subscripted will denote a finite subset of Σ.

We now define some marked operations as well as set derivatives.

<u>Definition</u>. Let $L_1 \subseteq \Sigma_1^*$, $L_2 \subseteq \Sigma_2^*$, and $\{a,b\} \subset \Sigma - (\Sigma_1 \cup \Sigma_2)$. The sets $L_1 a L_2$, $(aL_1) \cup (bL_2)$, $(aL_1)^*$, and $(aL_1)+$ are called a <u>marked product of L_1 and L_2</u>, a <u>marked union of L_1 and L_2</u>, a <u>marked * of L_1</u>, and a <u>marked + of L_1</u>, respectively.(2)

<u>Definition</u>. Given $L \subset \Sigma^*$ and $w \in \Sigma_1^*$, let $L/w = \{w_1 \mid w_1 w \in L\}$ and $w\backslash L = \{w_1 \mid ww_1 \in L\}$. The sets L/w and $w\backslash L$ are called <u>right</u> and <u>left derivatives</u>, respectively.

(1) For each set of symbols A, let A^0 be the set consisting of the empty word (denoted by ε). For each $i \geq 1$, let A^i be the set consisting of all words of length i over A, and $A^* = \bigcup_{i=0}^{\infty} A^i$.

(2) If A is an alphabet, then $A+ = \bigcup_{i=1}^{\infty} A^i$. Thus, $A+ = AA^*$.

Finally, we recall the definition of a generalized sequential machine.

Definition. A generalized sequential machine (gsm) is a 6-tuple $G = (K_1, \Sigma_1, \Sigma_2, \delta, \lambda, p_1)$ where:

(1) K_1 is a finite, non-empty set (states),

(2) Σ_1 is a finite set (input alphabet),

(3) Σ_2 is a finite set (output alphabet),

(4) δ is a mapping from $K_1 \times \Sigma_1$ into K_1 (next state function),

(5) λ is a mapping from $K_1 \times \Sigma_1$ into Σ_2^* (output function), and

(6) p_1 is an element of K_1 (start state).

The functions δ and λ are extended to mappings from $K_1 \times \Sigma_1^*$ by defining $\delta(p,\epsilon) = p$, $\lambda(p,\epsilon) = \epsilon$, $\delta(p,wa) = \delta[\delta(p,w), a]$, and $\lambda(p,wa) = \lambda(p,w)\lambda[\delta(p,w), a]$ for all $p \in K_1$, $w \in \Sigma_1^*$, and $a \in \Sigma_1$.

Definition. For each gsm $G = (K_1, \Sigma_1, \Sigma_2, \delta, \lambda, p_1)$, the function from $2^{\Sigma_1^*}$ into $2^{\Sigma_2^*}$ defined by $G(X) = \{y | \lambda(p_1,x) = y \text{ for some } x \in X\}$ is called a gsm mapping. The function from $2^{\Sigma_2^*}$ into $2^{\Sigma_1^*}$ defined by $G^{-1}(Y) = \{x | \lambda(p_1,x) \in Y\}$ is called an inverse gsm mapping.

In order to treat families of deterministic languages, we need to extend the notion of a gsm to that of a machine with a right endmarker. We therefore introduce the concept of a "marked gsm."

Definition. Let $ be an abstract symbol. A marked generalized sequential machine (mgsm) is a 6-tuple $G = (K_1, \Sigma_1, \Sigma_2, \delta, \lambda, p_1)$ where:

(1) K_1, Σ_1, Σ_2, and p_1 are as defined for a gsm,

(2) $\Sigma_1 \cup \Sigma_2$ does not contain $,

(3) δ is mapping from $K_1 \times (\Sigma_1 \cup \{\$\})$ into K_1 (next state function), and

(4) λ is a mapping from $K_1 \times (\Sigma_1 \{\$\})$ into $\Sigma_2^* \cup \Sigma_2^*\$$ such that $\lambda(K_1, \Sigma_1) \subset \Sigma_2^*$ and $\lambda(K_1, \$) \subset \Sigma_2^*\$$ (output function).

The functions δ and λ are extended to mappings from $K_1 \times (\Sigma_1 \cup \{\$\})^*$ as for a gsm.

Definition. For each mgsm $G = (K_1, \Sigma_1, \Sigma_2, \delta, \lambda, p_1)$ the function from $2^{\Sigma_1^*}$ into $2^{\Sigma_2^*}$ defined by $G(X) = \{y | \lambda(p_1, x\$) = y\$ \text{ for some } x \in X\}$ is called a mgsm mapping. The function from $2^{\Sigma_2^*}$ into $2^{\Sigma_1^*}$ defined by $G^{-1}(Y) = \{x | \lambda(p_1, x\$) \in Y\$\}$ is called an inverse mgsm mapping.

We are now ready for the families of languages with which we shall be concerned.

Definition. An abstract family of deterministic languages (abbreviated AFDL) is a family of languages closed under (1) marked union, (2) marked *, and (3) inverse mgsm mapping.

The adjective "deterministic" refers to the fact, as will be seen in Section 2, that AFDL coincide with the families of languages recognized by families of one-way deterministic acceptors.

Examples of AFDL are the family of regular sets, the family of context free languages, the family of sets recognized by deterministic pushdown acceptors, and the family of recursive sets. The family of linear languages is not an AFDL because it is not closed under marked *.

Due to closure under marked * (and inverse mgsm mapping), each AFDL contains $\{\epsilon\}$. This inclusion of $\{\epsilon\}$ excludes many families such as the context sensitive languages and the ϵ-free regular sets. However, $\{\epsilon\}$ can be appended to most ϵ-free families without significantly altering the properties or structure of that family. In particular, the family $\{L, L \cup \{\epsilon\} | L \text{ is context sensitive}\}$ is an AFDL.

Now, we turn to some elementary closure properties of AFDL. First though, we present a lemma on some closure properties implied by inverse mgsm mapping. This lemma clearly illustrates the additional "power" obtained by adding a right endmarker to a gsm.

Lemma 1.1 Let $\mathcal{L}$ be a family of languages closed under inverse mgsm mapping, and $L \subseteq \Sigma_1^*$, a language in $\mathcal{L}$. Let R be a regular set, F a finite set, w a word in Σ_1^*, and G a gsm. Then (a) $G^{-1}(L)$, (b) $L \cap R$, (c) $L - R$, (d) R, (e) $L \cup F$, (f) L/w, (g) $w\backslash L$, (h) wL, and (i) Lw are in $\mathcal{L}$.

We may restate (d) as a corollary.

Corollary 1 Every family of languages closed under inverse mgsm mapping contains the family of regular sets.

Corollary 2 Let $\mathcal{L}$ be a family of languages closed under inverse mgsm mapping. Let $L = L_1cL_2$ be in $\mathcal{L}$ where $L_1 \subseteq \Sigma_1^*$, $L_2 \subseteq \Sigma_1^*$, $c \in \Sigma - \Sigma_1$, and $L \neq \emptyset$. Then L_1 and L_2 are in $\mathcal{L}$.

We now present some closure properties of AFDL.

Theorem 1.1 Let be an AFDL. Let L be in $\mathcal{L}$, R a regular set, F a finite set, w a word, and G a gsm. Then $G^{-1}(L)$, $L \cap R$, $L - R$, R, $L \cup F$, L/w, $w\backslash L$, wL, and Lw are in $\mathcal{L}$. Also, $\mathcal{L}$ is closed under marked product.

Proof Since $\mathcal{L}$ is closed under inverse mgsm mapping, it follows from Lemma 1.1 that $G^{-1}(L)$, $L \cap R$, $L - R$, R, $L \cup F$, L/w, $w\backslash L$, wL, and Lw are in $\mathcal{L}$. Consider marked product. Let $L_1 \subseteq \Sigma_1^*$ and $L_2 \subseteq \Sigma_1^*$ be in $\mathcal{L}$. Let a, b, and c be three symbols in $\Sigma - \Sigma_1$. Let $L_3 = [c(aL_1 \cup bL_2)]^*$ and $L_4 = L_3 \cap (ca\Sigma_1^*cb\Sigma_1^*)$. Then $L_4 = caL_1cbL_2$. Clearly L_3, and hence L_4, is in $\mathcal{L}$. Let G_2 be a mgsm such that $G_2(w_1) = caw_2cbw_3$ if $w_1 = w_2cw_3$, and $G_2(w_1) = caw_1$ otherwise. Then $L_1cL_2 = G_2^{-1}(L_4)$, and thus L_1cL_2 is in $\mathcal{L}$.

Corollary Every AFDL contains the family of regular sets.

In the following theorem, we consider an alternate formulation of AFDL which will be useful in later proofs. First, we introduce a lemma.

Lemma 1.2 A family of languages $\mathcal{L}$ is closed under inverse mgsm mapping if and only if $\mathcal{L}$ is closed under inverse gsm mapping and right derivative by a single symbol.

Corollary The family of regular sets is closed under inverse mgsm mapping.

Thus, the family of regular sets is the smallest family closed under inverse mgsm mapping. In view of Lemma 1.2, we have the next theorem directly.

Theorem 1.2 A family of languages $\mathcal{L}$ is an AFDL if and only if $\mathcal{L}$ is closed under marked union, marked *, inverse gsm mapping, and right derivative by a single symbol.

Corollary The family of regular sets is an AFDL.

From this corollary and the corollary to Theorem 1.1, we have that the family of regular sets is the smallest AFDL.

Section 2. AFDA

In this section, an abstract family of (one-way) deterministic acceptors (AFDA) is discussed. The main result of the paper is then proved, namely that a family of languages is an AFDL if and only if there exists an AFDA which accepts exactly that family. Finally, we consider memory-bounded and time-bounded AFDA.

We now give an intuitive description of AFDA. Basically, it is a family of one-way, deterministic acceptors which contain an input tape, a finite control, and an auxiliary memory. Its input tape has a unique symbol \$ as its rightmost symbol. The finite control consists of a finite number of states. One of the states is distinguished and is called the start state. A subset of these states is distinguished and is called the set of accepting states. The auxiliary memory (which is potentially infinite) consists of a word over a set of memory symbols. A write function is used to change the state of the auxiliary memory, and a read function is used to determine the "effective state" of the memory. Each acceptor operates in the following manner:

(1) The finite control and the auxiliary memory are initialized to the start state and the empty word, respectively.

(2) Both the next state and a memory instruction are then determined by either (a) the current state, the "effective state" of the auxiliary memory, and the leftmost symbol of the input tape, or (b) the current state and the "effective state" of the auxiliary memory (the latter is called an ϵ-move).

(3) Simultaneous with the state change, the auxiliary memory is altered as a function of the selected memory instruction and the contents of the memory.

(4) The input tape is scanned from left to right until the acceptor attempts to scan a symbol to the right of \$. At that time, the acceptor blocks, and the input tape is accepted if the finite control is in an accepting state and the auxiliary memory is empty.

We now formalize the notion of an AFDA.

Definition Let \$ be an abstract symbol. An abstract family of (one-way) deterministic acceptors (abbreviated AFDA) is an ordered pair $(\Omega, \mathfrak{D})$, or $\mathfrak{D}$ when Ω is understood, with the following properties:

(1) Ω is a 6-tuple $(K, \Sigma, \Gamma, I, f, g)$, where

(a) K and Σ are infinite abstract sets (states and input alphabet, respectively), with \$ not in Σ.

(b) Γ and I are non-empty abstract sets (possibly infinite).

(c) f is a mapping from $\Gamma^* \times I$ into $\Gamma^* \cup \emptyset$ (f is called the write function).

(d) g is a mapping from Γ^* into $\Gamma^* \cup \emptyset$ such that $g(\gamma) = \epsilon$ if and only if $\gamma = \epsilon$ (g is called the read function).

(e) for each $\dot{\gamma}$ in $g(\Gamma^*)$, there exists an element $1_{\dot{\gamma}}$ in I such that $f(\gamma', 1_{\dot{\gamma}}) = \gamma$ for all γ such that $\dot{\gamma} = g(\gamma')$.

(f) for each u in I, there exists a finite set $\Gamma_u \subseteq \Gamma$ with the following property: If $\Gamma_1 \subseteq \Gamma$, γ is in Γ_1^*, and $f(\gamma, u) \neq \emptyset$, then $f(\gamma, u)$ is in $(\Gamma_1 \cup \Gamma_u)^*$.

(2) $\mathfrak{D}$ is a family of elements (called acceptors) $D = (K_1, \Sigma_1, \delta, p_o, F)$, where

(a) K_1 and Σ_1 are finite subsets of K and Σ, respectively.

(b) p_o is an element of K_1 (start state).

(c) F is a subset of K_1 (set of accepting states).

(d) δ is a mapping from $K_1 \times (\Sigma_1 \cup \{\epsilon\} \cup \{\$\}) \times g(\Gamma^*)$ into $(K_1 \times I) \cup \{\emptyset\}$ such that $\{\gamma \mid \delta(p, x, \gamma) \neq \emptyset$ for some $p \in K_1$ and $x \in (\Sigma_1 \cup \{\epsilon\} \cup \{\$\})\}$ is a finite set.

(e) for all $p \in K_1$ and $\gamma \in \Gamma^*$, if $\delta(p,x,\gamma) \neq \emptyset$ for some $x \in (\Sigma_1 \cup \{\$\})$, then $\delta(p,\epsilon,\gamma) = \emptyset$; and if $\delta(p,\epsilon,\gamma) \neq \emptyset$, then $\delta(p,x,\gamma) = \emptyset$ for all $x \in (\Sigma_1 \cup \{\$\})$.

We now give a brief intuitive discussion of AFDA. The set K is the universe of allowable states, Σ is the set of all allowable input symbols (hence each input word is in $\Sigma^*\$$), Γ is the set of allowable auxiliary memory symbols (thus a memory configuration is a word in Γ^*), and I is a set of memory instructions. The read and write functions are allowed to be partially defined functions because all words in Γ^* might not be permissible memory configurations. Condition 1e is used to permit state changes in the finite control while preserving the auxiliary memory. Conditions 1f, 2a, and 2d require that each acceptor in $\mathfrak{D}$ operate with words in a finite subset of Γ. Conditions 2a and 2d dictate that each acceptor is finitely described.

We now present two examples of AFDA.

Example 1 Deterministic Pushdown Acceptors.[3] Let k and Σ be infinite abstract sets. Let $\Gamma = \Sigma$, and $I = \{z \in \Sigma^* \mid |z| \leq 2\}$. Let g be defined as $g(\gamma z) = z$ and $g(\epsilon) = \epsilon$ for all $\gamma \in \Gamma^*$ and $z \in \Gamma$. Let f be defined as $f(\gamma z, z') = \gamma z'$ and $f(\epsilon, z') = z'$ for all $\gamma \in \Gamma^*$, $z \in \Gamma$, and $z' \in I$.

Example 2 One-way Nonerasing Deterministic Stack Automata. This device differs from the original version[5] in that a special memory instruction is included to empty the memory (independent of its contents). While a proof is not given here, it can be shown that this example is equivalent to the original formulation. Let K and Σ be infinite abstract sets. Let the set $\{-1,0,+1,c,\uparrow\}$ be disjoint from Σ. Let $\Gamma = \Sigma \cup \{\uparrow\}$, and let $I = \Sigma \cup \Sigma^2 \cup \{-1,0,+1,c\}$. Let g be defined by $g(\epsilon) = \epsilon$ and $g(\gamma z \uparrow \gamma') = z$ for all $\gamma \in \Sigma^*$, $\gamma' \in \Sigma^*$, and $z \in \Sigma$. Let f be defined as follows:

(1) $f(\epsilon, i) = \epsilon$ for $i \in \{-1,0,+1\}$.

(2) $f(\epsilon, z) = z\uparrow$ for all $z \in \Sigma$.

(3) $f(\gamma, c) = \epsilon$ for all $\gamma \in \Gamma^*$. (This rule permits emptying the memory.)

(4) $f(\gamma, 0) = \gamma$ for all $\gamma \in \Gamma^*$.

(5) $f(\gamma \uparrow z \gamma', +1) = \gamma z \uparrow \gamma'$ for all $\gamma \in \Sigma^+$, $z \in \Sigma$, and $\gamma' \in \Sigma^*$.

(6) $f(\gamma z \uparrow \gamma', -1) = \gamma \uparrow z \gamma'$ for all $\gamma \in \Sigma^+$, $z \in \Sigma$, and $\gamma' \in \Sigma^*$.

(7) $f(\gamma z \uparrow, z') = \gamma z' \uparrow$ for all $z \in \Sigma$, $\gamma \in \Sigma^*$, and $z' \in (\Sigma \cup \Sigma^2)$.

(8) $f(z \uparrow \gamma, -1) = z \uparrow \gamma$ for all $z \in \Sigma$ and $\gamma \in \Sigma^*$.

(This rule prevents the stack marker from leaving the memory on the left.)

(9) $f(\gamma \uparrow, +1) = \gamma \uparrow$ for all $\gamma \in \Sigma^+$. (This rule prevents the stack marker from leaving the memory on the right.)

To conveniently describe the action of an acceptor, we adopt the following notation.

Notation Let $D = (K_1, \Sigma_1, \delta, p_1, F)$ be in $\mathfrak{D}$. For $k \geq 0$, $\vdash_D^k$ is a relation on $K_1 \times (\Sigma_1 \cup \{\$\})^* \times \Gamma^*$ defined as follows:

(1) $(p,w,\gamma) \vdash_D^0 (p,w,\gamma)$ for all $p \in K_1$, $\gamma \in \Gamma^*$, and $w \in (\Sigma_1 \cup \{\$\})^*$, and

(2) For $k \geq 0$, w and w" in $(\Sigma_1 \cup \{\$\})^*$, γ, γ', and γ'' in Γ^*, p, p', and p'' in K_1, and x in $(\Sigma_1 \cup \{\epsilon\} \cup \{\$\})$, $(p,w,\gamma) \vdash_D^{k+1} (p'',w'',\gamma'')$ if $(p,w,\gamma) \vdash_D^k (p',xw'',\gamma')$, $\delta[p',x,g(\gamma')] = (p'',u)$, and $f(\gamma',u) = \gamma''$.

Write $(p,w,\gamma) \vdash_D^* (p',w',\gamma')$ if $(p,w,\gamma) \vdash_D^k (p',w',\gamma')$ for some $k \geq 0$.

When D is understood, $\vdash_D^*$, $\vdash_D^k$ and $\vdash_D^1$ are written as $\vdash^*$, $\vdash^k$, and $\vdash$, respectively.

We now formally define the notion of "acceptance". In essence, the auxiliary memory must be emptied and the device must end in an accepting state in order to accept an input word. Note that \$ is not a symbol in the accepted words.

Definition Let $D = (K_1, \Sigma_1, \delta_1, p_o, F_1)$ be in $\mathfrak{D}$. A word $w \in \Sigma_1^*$ is accepted by D if $(p_o, w\$, \epsilon) \vdash_D^* (p, \epsilon, \epsilon)$ for some $p \in F_1$. The set of words accepted by D is defined to be the set $L(D) = \{w \in \Sigma_1^* \mid (p_o, w\$, \epsilon) \vdash_D^* (p, \epsilon, \epsilon)$ for some $p \in F_1\}$. The family of sets accepted by elements of $\mathfrak{D}$ is defined to be $\mathcal{L}(\mathfrak{D}) = \{L(D) \mid D \text{ is in } \mathfrak{D}\}$.

Now we show that the family of languages accepted by each AFDA is an AFDL. First though, we present an intermediate lemma.

Lemma 2.1 If $\mathfrak{D}$ is an AFDA, then $\mathcal{L}(\mathfrak{D})$ is closed under marked union and marked *.

We now have the first major result of this section.

Theorem 2.1 Let $\mathfrak{D}$ be an AFDA. Then $\mathcal{L}(\mathfrak{D})$ is an AFDL.

Proof By Lemma 2.1, $\mathcal{L}(\mathfrak{D})$ is closed under marked union and marked *. Hence it suffices to show that $\mathcal{L}(\mathfrak{D})$ is closed under inverse mgsm mapping. To see this, let $D_1 = (K_1, \Sigma_1, \delta_1, p_1, F_1)$ be in $\mathfrak{D}$. Let $G = (K_2, \Sigma_2, \Sigma_1, \delta_2, \lambda_2, p_2)$ be a mgsm, and let

$m = \text{Max}\ \{|\lambda_2(p,x)| \mid p \in K_2,\ x \in (\Sigma_2 \cup \{\$\})\}$(3)
Let $D_3 = (K_3, \Sigma_2, \delta_3, p_3, F_3)$, where $K_3 = K_2 \times K_1 \times \{w \mid w \in (\Sigma_1^* \cup \Sigma_1^*\$) \text{ and } |w| \leq m\}$, $F_3 = K_2 \times F_1 \times \{\epsilon\}$, and δ_3 is defined as follows:

(1) $\delta_3[(p,q,\epsilon),x,\gamma] = [(p',q,w),1_\gamma]$ if $x \in (\Sigma_2 \cup \{\$\})$, $\delta_2(p,x) = p'$, $\lambda_2(p,x) = w$, and $\delta_1(p,\epsilon,\gamma) = \emptyset$. (This rule simulates a move in G and encodes the output of G into the next state of D_3)

(2) $\delta_3[(p,q,xw),\epsilon,\gamma] = [(p,q',w),u]$ if $w \in (\Sigma_1^* \cup \Sigma_1^*\$)$, $|w| \leq m$, $x \in \Sigma_1 \cup \{\epsilon\} \cup \{\$\})$, and $\delta_1(q,x,\gamma) = (q',u)$. (This rule simulates a move in D_1 under the output of G which is encoded in the state of D_3.)

Then $L(D_3) = G^{-1}[L(D_1)]$, so that $\mathcal{L}(\mathfrak{D})$ is closed under inverse mgsm mapping. Therefore, $\mathcal{L}(\mathfrak{D})$ is an AFDL.

We now turn to the main result, namely that $\mathcal{L}$ is an AFDL if and only if there exists an AFDA $\mathfrak{D}$ such that $\mathcal{L} = \mathcal{L}(\mathfrak{D})$. First though, we present several definitions.

Definition Let $(\Sigma, \mathcal{L})$ be a family of languages where $\mathcal{I} \neq \emptyset$ is an index set disjoint from Σ and $\mathcal{L} = \{L_i \mid i \in \mathcal{I}\}$. Then $(\Omega, \mathfrak{D}(\mathcal{L}))$ denotes an AFDA with the following properties:

(1) $\Omega = (K, \Sigma, \Gamma, I, f, g)$,

(2) α, β, and 1_ϵ are symbols not in $\Sigma \cup \mathcal{I}$,

(3) $\Gamma = \Sigma \cup \mathcal{I} \cup \{\alpha, \beta\}$,

(4) $I = \Sigma \cup \mathcal{I} \cup \{\beta, 1_\epsilon\}$,

(5) g is defined by $g(\epsilon) = \epsilon$ and $g(\gamma) = \alpha$ for all $\gamma \in \Gamma^+$, and

(6) f is defined as follows:

(i) $f(\gamma, 1_\epsilon) = \gamma$ for all $\gamma \in \Gamma^*$,

(ii) $f(\gamma, z) = \gamma z$ for all $\gamma \in \Gamma^*$ and $z \in (\Sigma \cup \mathcal{I})$,

(iii) $f(\gamma, \beta) = \epsilon$ if $\gamma = i\gamma'$ for some $i \in \mathcal{I}$ and $\gamma' \in L_i$, and

(iv) $f(\gamma, \beta) = \gamma\beta$ if $\gamma \neq i\gamma'$ for any $i \in \mathcal{I}$ and $\gamma' \in L_i$.

We now define notationally the smallest AFDL containing a family of languages. It exists because it is the intersection of all AFDL that contain the family.

(3) If w is a word, then $|w|$ is its length. That is, if $w = x_1, x_2, \ldots x_n$ is in Σ_1^* with each x_i in Σ_1, then $|w| = n$. Also, $|\epsilon| = 0$.

Definition For each family of languages $\mathcal{L}$, let $\mathcal{F}(\mathcal{L})$ be the smallest AFDL containing $\mathcal{L}$.

Before proceeding to the main result, we present several lemmas.

Lemma 2.2 $\mathcal{F}(\mathcal{L}) \subseteq \mathcal{L}[\mathfrak{D}(\mathcal{L})]$.

We now define a type of acceptor in $\mathfrak{D}(\mathcal{L})$ with a specified structure. Intuitively, this acceptor D operates independent of the contents of the auxiliary memory. The existence of such a D is possible because the read has a finite range.

Definition For each $D = (K_1, \Sigma_1, \delta, p_1, F)$ in $\mathfrak{D}(\mathcal{L})$, let

$\underline{C_0} = \{p \mid (p_1, w\$, \epsilon) \overset{*}{\vdash}_D (p, w'\$, \gamma)$ for some $w \in \Sigma_1^*$, $w' \in \Sigma_1^*$, and $\gamma \in \Gamma^*(\Gamma - \{\beta\})\}$,

$\underline{C_1} = \{p \mid (p_1, w\$, \epsilon) \overset{*}{\vdash}_D (p, w'\$, \gamma)$ for some $w \in \Sigma_1^*$, $w' \in \Sigma_1^*$, and $\gamma \in (\Gamma^*\beta)^*\}$,

$\underline{C_2} = \{p \mid (p_1, w\$, \epsilon) \overset{*}{\vdash}_D (p, \epsilon, \gamma)$ for some $w \in \Sigma_1^*$, and $\gamma \in \Gamma^*(\Gamma - \{\beta\})\}$, and

$\underline{C_3} = \{p \mid (p_1, w\$, \epsilon) \overset{*}{\vdash}_D (p, \epsilon, \gamma)$ for some $w \in \Sigma_1^*$ and $\gamma \in (\Gamma^*\beta)^*\}$.

Then D is said to be a disjoint acceptor if:

(1) $K_1 = C_0 \cup C_1 \cup C_2 \cup C_3$,

(2) C_0, C_1, C_2, and C_3 are pairwise disjoint sets,

(3) $\delta(p,x,\epsilon) = \delta(p,x,\alpha)$ for all $p \in K_1$ and $x \in (\Sigma_1 \cup \{\$\} \cup \{\epsilon\})$,

(4) $F \subseteq C_3$,

(5) $\delta(p,x,\epsilon) = \emptyset$ for all $p \in (C_2 \cup C_3)$ and $x \in (\Sigma_1 \cup \{\$\})$, and

(6) $\delta(p,\epsilon,\epsilon) = \emptyset$ for all $p \in F$.

We now present a lemma concerning disjoint acceptors.

Lemma 2.3 Let D_1 be in $\mathfrak{D}(\mathcal{L})$. Then there exists a disjoint acceptor D_2 in $\mathfrak{D}(\mathcal{L})$ such that $L(D_1) = L(D_2)$.

We now define three sets relative to an acceptor in $\mathfrak{D}(\mathcal{L})$.

Definition Let $(\Sigma, \mathcal{L})$ be a family of languages, where $\mathcal{I} \neq \emptyset$, $\mathcal{I} \cap \Sigma = \emptyset$, and $\mathcal{L} = \{L_i \mid i \text{ in } \mathcal{I}\}$. For each $i \in \mathcal{I}$, let $L_i \subseteq \Sigma_i^*$. Then for each $D = (K_1, \Sigma_1, \delta, p_1, F)$ in $\mathfrak{D}(\mathcal{L})$, let

$\underline{\mathcal{I}_D} = \{i \in \mathcal{I} \mid \delta(p,x,\gamma) = (p',i)$ for some p, p', x, and $\gamma\}$

$$U_D = \{w \in \Sigma_1^* \mid (p_1, w\$, \epsilon) \vdash_D^* (p', w', \gamma) \text{ for some } p' \in K_1,\ w' \in (\Sigma_1 \cup \{\$\})^*, \text{ and } \gamma \in \Gamma^* \rho\, \Gamma^* \}, \text{ and}$$

$$M_D = \Big[\bigcup_{i \in \mathcal{A}_D} c_i L_i c \Big]^*, \text{ where } \{c, c_i \mid i \in \mathcal{A}_D\} \subseteq \Big[\Sigma - \Big(\bigcup_{i \in \mathcal{A}_D} \Sigma_i\Big)\Big].$$

In the next two lemmas, we show the existence of a fsa and a mgsm which are vital to the proof of the main result.

Lemma 2.4 Let D be in $\mathfrak{D}(\mathcal{L})$. Then there exists a fsa A such that $L(D) \subseteq L(A)$ and $L(A) - L(D) \subseteq U_D$.

Proof Because of Lemma 2.3, we may assume that D is a disjoint acceptor. Let $D = (K_1, \Sigma_1, \delta, p_1, F)$. Let F_A be the set of all $p_o' \in K_1$ with the following property: There exists a finite sequence of states $p_o', p_1', \ldots, p_n', q_o', q_1', \ldots, q_m'$ such that $n \geq 0$, $m \geq 0$, $\delta(p_i', \epsilon, \epsilon) = (p_{i+1}', u_{i+1}')$ for $0 \leq i \leq n-1$, $\delta(p_n', \$, \epsilon) = (q_o', u_o)$, $\delta(q_i', \epsilon, \epsilon) = (q_{i+1}', u_{i+1})$ for $0 \leq i \leq m-1$, $\delta(q_m', \epsilon, \epsilon) = \emptyset$, and $q_m' \in F$. Let q_o be a symbol not in K_1. Let $A = (K_1 \cup \{q_o\}, \Sigma_1, \delta_A, p_1, F_A)$ be a fsa where δ_A is defined as follows:

(1) $\delta_A(p_1', x) = p_n'$ if there exists a finite sequence of states $p_1', p_2', \ldots, p_n'$ in K_1 such that $n \geq 2$, $\delta(p_1', \epsilon, \epsilon) = (p_{i+1}', u_{i+1})$ for $1 \leq i \leq n-2$, and $\delta(p_{n-1}', x, \epsilon) = (p_n', u_n')$ for some $x \in \Sigma_1$.

(2) $\delta_A(p, x) = q_o$ for all $p \in (K_1 \cup \{q_o\})$ and $x \in \Sigma_1$ such that (1) is not satisfied.

Hence $(p_1, w\$, \epsilon) \vdash_D^* (p', \epsilon, \epsilon)$ for $w \in \Sigma_1^*$ and $p' \in F$ implies that $\delta_A(p_1, w) \in F_A$. Thus $L(D) \subseteq L(A)$. Let w be in $L(A) - L(D)$. Then $(p_1, w\$, \epsilon) \vdash_D^* (q, \epsilon, \gamma)$ for some $q \in F$ and $\gamma \in \Gamma^+$. Now $\gamma \neq \epsilon$ and, by definition of disjoint acceptor, $F \subseteq \{p \mid (p_1, w'\$, \epsilon) \vdash_D^* (p, \epsilon, \gamma')$ for some $w' \in \Sigma_1^*$ and $\gamma' \in (\Gamma^* \rho)^*\}$. Therefore, $\gamma \in \Gamma^* \rho$. Hence w is in U_D.

Lemma 2.5 Let D be in $\mathfrak{D}(\mathcal{L})$. Then there exists a mgsm G such that $L(D) \subseteq G^{-1}(M_D)$ and $G^{-1}(M_D) \cap U_D = \emptyset$.

Proof By lemma 2.3, we may assume that D is a disjoint acceptor. Let $D = (K_1, \Sigma_1, \delta, p_1, F)$. For each $i \in \mathcal{A}$, let $L_i \subseteq \Sigma_i^*$. Let $\{c, c_i \mid i \in \mathcal{A}_D\} \subset [\Sigma - (\bigcup_{i \in \mathcal{A}_D} \Sigma_i)]$. By definition, $M_D = [\bigcup_{i \in \mathcal{A}_D} c_i L_i c]^*$. Let $\Delta_1 = \{u \mid \delta(p, x, \gamma) = (p', u)$ for some p, p', x, and $\gamma\}$. Let h be the homomorphism from $(\Delta_1 \cup \{1_\epsilon, \rho\})^*$ into Σ^* defined by $h(1_\epsilon) = \epsilon$, $h(\rho) = c$, $h(i) = c_i$ for all $i \in \mathcal{A}_D$, and $h(x) = x$ otherwise. Let q_o be a symbol not in K_1, let $\Delta = \{h(u) \mid u \in \Delta_1\}$, and let d be in $\Sigma - \Delta$. Let $G = (K_1 \cup \{q_o\}, \Sigma_1, \Delta \cup \{d\}, \delta_1, \lambda_1, p_1)$, where δ_1 and λ_1 are defined as follows for $x \in \Sigma_1$:

(1) $\delta_1(p_o', x) = p_n'$ and $\lambda_1(p_o', x) = h(u_1 u_2 \ldots u_n)$ if there exists a finite sequence of states $p_o', p_1', \ldots, p_n'$ in K_1 such that $n \geq 1$, $\delta(p_i', \epsilon, \epsilon) = (p_{i+1}', u_{i+1})$ for $0 \leq i \leq n-2$, and $\delta(p_{n-1}', x, \epsilon) = (p_n', u_n)$.

(2) $\delta_1(p_o', \$) = q_m'$ and $\lambda_1(p_o', \$) = h(u_1 u_2 \ldots u_n u_o' u_1' \ldots u_m')\$$ if there exists a finite sequence of states $p_o', p_1', \ldots, p_n', q_o', q_1', \ldots, q_m'$ such that $n \geq 0$, $m \geq 0$, $\delta(p_i', \epsilon, \epsilon) = (p_{i+1}', u_{i+1})$ for $0 \leq i \leq n-1$, $\delta(p_n', \$, \epsilon) = (q_o', u_o')$, $\delta(q_i', \epsilon, \epsilon) = (q_{i+1}', u_{i+1}')$ for $0 \leq i \leq m-1$, $\delta(q_m', \epsilon, \epsilon) = \emptyset$, and $q_m' \in F$.

(3) $\delta_1(p, x) = q_o$ and $\lambda_1(p, x) = d$ if $p \in (K_1 \cup \{q_o\})$ and (1) does not hold.

(4) $\delta_1(p, \$) = q_o$ and $\lambda_1(p, \$) = d\$$ if $p \in (K_1 \cup \{q_o\})$ and (2) does not hold.

For all $p' \in K_1$, $w \in L(D)$, $w' \in (\Sigma_1^* \$)^*$, and $\gamma \in \Gamma^*$, such that $(p_1, w\$, \epsilon) \vdash_D^* (p', w', \gamma)$, $\lambda_1(p_1, w\$/w') = h(\gamma'\gamma)$ for some $\gamma' \in \Gamma^*$. Hence $L(D) \subseteq G^{-1}(M_D)$. If $(p_1, w\$, \epsilon) \vdash_D^* (p', w', \gamma)$ for some $\gamma \in \Gamma^* \rho \Gamma^*$, then $\lambda_1(p_1, w\$) \notin M_D$. Thus $G^{-1}(M_D) \cap U_D = \emptyset$.

Lemma 2.6 $\mathcal{F}(\mathcal{L}) = \mathcal{L}[\mathfrak{D}(\mathcal{L})]$ for every family of languages $\mathcal{L}$.

We now have the main result of the paper.

Theorem 2.2 A family of languages $\mathcal{L}$ is an AFDL if and only if there exists an AFDA $\mathfrak{D}$ such that $\mathcal{L} = \mathcal{L}(\mathfrak{D})$.

Proof The "if" is proved in Theorem 2.1. Consider the "only if". Let $\mathcal{L}$ be an AFDL, and $\mathfrak{D} = \mathfrak{D}(\mathcal{L})$. By Lemma 2.6, $\mathcal{L}[\mathfrak{D}(\mathcal{L})] = \mathcal{F}(\mathcal{L}) = \mathcal{L}$.

The significant of this theorem becomes clear when examples of AFDL are recalled: context free languages, context sensitive languages (if ϵ is added), recursive sets, etc. A stronger observation is that the family of languages accepted by each closed class of balloon automata[6] is an AFDL, and thus there exists an equivalent

AFDA. This is independent of whether the closed class is one-way, two-way, deterministic, or nondeterministic. Hence almost all known acceptors are equivalent to some AFDA. One outstanding counterexample is the ultralinear languages[4] which are not closed under marked $*$, but which are accepted by finite turn pushdown acceptors.

We now turn to subsets of AFDA which have restrictions placed on the number of ϵ-moves.

Definition Let k be a non-negative integer, and $\mathfrak{D}$ an AFDA. Let $\mathfrak{D}_k^t$ be the set of all D in $\mathfrak{D}$ such that if $(p,w,\gamma) \vdash_D^m (p',w,\gamma')$ then $m \leq k$. Let $\mathcal{L}^t(\mathfrak{D}) = \{L(D) \mid D \text{ is in } \mathfrak{D}_k^t \text{ for some } k \geq 0\}$. Each L in $\mathcal{L}^t(\mathfrak{D})$ is called quasi-real-time.

The family $\mathcal{L}(\mathfrak{D}_0^t)$ is usually called the real-time languages for the AFDA $\mathfrak{D}$. For AFDA which possess a "speed-up" property, $\mathcal{L}^t(\mathfrak{D})$ is identical to $\mathcal{L}(\mathfrak{D}_0^t)$. As is well known, the family of languages accepted by real-time Turing Acceptors (which have a "speed-up" property) is a proper subfamily of the family accepted by Turing Acceptors. Thus, in general, $\mathcal{L}^t(\mathfrak{D}) \neq \mathcal{L}(\mathfrak{D})$.

We now present a lemma showing a relation between acceptors in $\mathfrak{D}_k^t$ for $k \geq 0$.

Lemma 2.7 Let $\mathfrak{D}$ be an AFDA, and $k \geq 0$. If D_1 is in $\mathfrak{D}_k^t$, then there exists a D_2 in $\mathfrak{D}_0^t$ and a mgsm G such that $L(D_1) = G^{-1}[L(D_2)]$.

We now have a result concerning $\mathcal{L}^t(\mathfrak{D})$.

Theorem 2.3 Let $\mathfrak{D}$ be an AFDA, and $k \geq 0$. Then $\mathcal{L}^t(\mathfrak{D})$ is the smallest AFDL containing $\mathcal{L}(\mathfrak{D}_k^t)$.

Proof First, we show that $\mathcal{L}^t(\mathfrak{D})$ is an AFDL. Clearly $\mathcal{L}^t(\mathfrak{D})$ is closed under marked union and marked $*$. Consider inverse mgsm mapping. Let $G = (K_1, \Sigma_1, \Sigma_2, \delta, \lambda, p_1)$ be a mgsm, and let $m = \text{Max}\{|\lambda(p,x)| \mid p \in K_1 \text{ and } x \in (\Sigma_1 \cup \{\$\})\}$. Using a construction identical to that in the proof of Theorem 2.1, it is clear that if L is in $\mathcal{L}(\mathfrak{D}_l^t)$ for some $l \geq 0$, then $G^{-1}(L)$ is in $\mathcal{L}(\mathfrak{D}_i^t)$ where $i = (1+l)(m+1)$. Therefore, $\mathcal{L}^t(\mathfrak{D})$ is closed under mgsm mapping, and thus $\mathcal{L}^t(\mathfrak{D})$ is an AFDL. To complete the proof, it suffices to show that every AFDL containing $\mathcal{L}(\mathfrak{D}_k^t)$ also contains $\mathcal{L}^t(\mathfrak{D})$. Let L_1 be in $\mathcal{L}^t(\mathfrak{D})$. Then there exists an integer j and D_1 in $\mathfrak{D}_j^t$ such that $L_1 = L(D_1)$. By Lemma 2.7, there exists a D_2 in $\mathfrak{D}_0^t$ and a mgsm G_1 such that $L(D_1) = G^{-1}[L(D_2)]$. Since $G^{-1}[L(D_2)]$ is in every AFDL $\mathcal{L}$ containing $\mathcal{L}(\mathfrak{D}_k^t)$ and $\mathcal{L}(\mathfrak{D}_k^t)$ contains $\mathcal{L}(\mathfrak{D}_0^t)$, $\mathcal{L}^t(\mathfrak{D})$ is in $\mathcal{L}$. Thus $\mathcal{L}^t(\mathfrak{D})$ is the smallest AFDL containing $\mathcal{L}(\mathfrak{D}_k^t)$.

Corollary Let $\mathfrak{D}$ be an AFDA. Then $\mathcal{L}(\mathfrak{D}_k^t) = \mathcal{L}^t(\mathfrak{D})$ for every $k \geq 0$ if and only if $\mathcal{L}(\mathfrak{D}_0^t)$ is an AFDL.

Next, we consider subsets of AFDA with restrictions on the length of the auxiliary memory.

Definition Let $T(n)$ be a total function from the non-negative integers into the non-negative integers. Then $T(n)$ is a tape function if $T(n+1) \geq T(n)$ for all $n \geq 0$. For each AFDA $\mathfrak{D}$, let $\mathfrak{D}_{T(n)}$ be the family of all $D = (K_1, \Sigma_1, \delta, p_1, F)$ in $\mathfrak{D}$ such that if $(p_1, w\$, \epsilon) \vdash_D^* (p', w', \gamma)$, then $|\gamma| \leq T(|w|)$. For $k \geq 0$, let $T_k(n) = T(kn)$. Let $\mathcal{L}_{T(n)}(\mathfrak{D}) = \bigcup_{k=0}^{\infty} \mathcal{L}(\mathfrak{D}_{T_k(n)}) = \{L(D) \mid D \text{ is in } \mathfrak{D}_{T_k(n)} \text{ for some } k \geq 0\}$.

First, we present a technical lemma before proceeding to the main result.

Lemma 2.8 Let $\mathfrak{D}$ be an AFDA, $T(n)$ a tape function, and k a non-negative integer. If D_1 is in $\mathfrak{D}_{T_k(n)}$, then there exists a mgsm G and D_2 in $\mathfrak{D}_{T(n)}$ such that $L(D_1) = G^{-1}[L(D_2)]$.

We now have a result concerning $\mathcal{L}_{T(n)}(\mathfrak{D})$.

Theorem 2.4 Let $\mathfrak{D}$ be an AFDA, $T(n)$ a tape function, and k a positive integer. Then $\mathcal{L}_{T(n)}(\mathfrak{D})$ is the smallest AFDA containing $\mathcal{L}(\mathfrak{D}_{T_k(n)})$.

Proof First we show that $\mathcal{L}_{T(n)}(\mathfrak{D})$ is an AFDL. Clearly $\mathcal{L}(\mathfrak{D}_{T_k(n)})$ is closed under marked union and marked $*$. Consider inverse mgsm marking. Let $G = (K_1, \Sigma_1, \Sigma_2, \delta, \lambda, p_1)$ be a mgsm, and let $m = \text{Max}\{|\lambda(p,x)| \mid p \in K \text{ and } x \in (\Sigma_1 \cup \{\$\})\}$. Using the construction in the proof of Theorem 2.1, it is clear that if L is in $\mathcal{L}(\mathfrak{D}_{T_l(n)})$ for some $l \geq 0$, then $G^{-1}(l) - \{\epsilon\}$ is in $\mathcal{L}(\mathfrak{D}_{T'(n)})$ where $T'(n) = T[lm(n+1)]$. Since $n > 0$, we have that $G^{-1}(L) - \{\epsilon\}$ is in $\mathcal{L}(\mathfrak{D}_{T_i(n)})$ where $i = 2lm$. Therefore, $\mathcal{L}_{T(n)}(\mathfrak{D})$ is closed under inverse mgsm mapping, and thus $\mathcal{L}_{T(n)}(\mathfrak{D})$ is an AFDL.

To complete the proof, it suffices to show that every AFDL containing $\mathcal{L}(\mathfrak{D}_{T_k(n)})$ also contains $\mathcal{L}_{T(n)}(\mathfrak{D})$. Let L_1 be in $\mathcal{L}_{T(n)}(\mathfrak{D})$. Then there exists an integer j and D_1 in $\mathfrak{D}_{T_j(n)}$ such that $L_1 = L(D_1)$. By Lemma 2.8, there exists a D_2 in $\mathfrak{D}_{T(n)}$ and a mgsm G_1 such that $L(D_1) = G_1^{-1}[L(D_2)]$. For $k > 0$, $\mathfrak{D}_{T(n)}$ is a subset of $\mathfrak{D}_{T_k(n)}$. Then

Then $G_1^{-1}[L(D_2)]$ is in every AFDL $\mathcal{L}$ containing $\mathcal{L}(\mathfrak{D}_{T_k(n)})$, and hence $\mathcal{L}_{T(n)}(\mathfrak{D})$ is in $\mathcal{L}$. Thus $\mathcal{L}_{T(n)}(\mathfrak{D})$ is the smallest AFDL containing $\mathcal{L}(\mathfrak{D}_{T_k(n)})$. This completes the proof of the theorem.

<u>Corollary</u> Let $\mathfrak{D}$ be an AFDA, and $T(n)$ a tape function. Then $\mathcal{L}_{T(n)}(\mathfrak{D}) = \mathcal{L}(\mathfrak{D}_{T_k(n)})$ for every $k > 0$ if and only if $\mathcal{L}(\mathfrak{D}_{T(n)})$ is an AFDL.

Now, we present a theorem which links $\mathcal{L}_n(\mathfrak{D})$ and $\mathcal{L}^{t}(\mathfrak{D})$.

<u>Theorem 2.5</u> Let $(\Omega, \mathfrak{D})$ be an AFDA where $\Omega = (K, \Sigma, \Gamma, I, f, g)$ and f satisfies the following condition: For each u in I, there exists an integer M_u such that $|f(\gamma,u)| \le M_u + |\gamma|$ for all $\gamma \in \Gamma^*$ having $f(\gamma,u) \ne \emptyset$. Let $T(n) = n$ for all $n \ge 0$. Then $\mathcal{L}_{T(n)}(\mathfrak{D}) \supseteq \mathcal{L}^{t}(\mathfrak{D})$.

<u>Proof</u> Let $D = (K_1, \Sigma_1, \delta, p_1, F)$ be in $\mathfrak{D}_k^{t}$ for some $k \ge 0$. Let $m_1 = \text{Max}\{M_u \mid \delta(p,x,\gamma) = (p',u)$ for some p, x, γ, and $p'\}$. Since both $\mathcal{L}^{t}(\mathfrak{D})$ and $\mathcal{L}_{T(n)}(\mathfrak{D})$ are AFDL, $L(D)$ is in $\mathcal{L}_{T(n)}(\mathfrak{D})$ if and only if $L(D) - \{\epsilon\}$ is in $\mathcal{L}_{T(n)}(\mathfrak{D})$. Thus we need consider only $w \in \Sigma_1^+$. If $(p_1, w\$, \epsilon) \overset{l}{\underset{D}{\vdash}} (p', w', \gamma)$, then $l \le k|w| + 2k + |w| + 1$ and $|\gamma| \le m_1(k|w| + 2k + |w| + 1)$. Since $|w| > 0$, we have that $|\gamma| \le m_1(3k + 2)|w|$ and $|\gamma| \le T_i(|w|)$ where $i = 3km_1 + 2m_1$. Thus $L(D) - \{\epsilon\}$ is in $\mathcal{L}(\mathfrak{D}_{T_i(n)})$. Hence $\mathcal{L}^{t}(\mathfrak{D}) \subseteq \mathcal{L}_{T(n)}(\mathfrak{D})$.

<u>Section 3.</u>
<u>Independence of AFDL Operations</u>

In this section, we prove that the three operations in the definition of an AFDL are independent of each other. (This implies that an arbitrary AFDL cannot be defined by only two of these operations.) This result is obtained by exhibiting three counter-examples.

Before proceeding to the main result, we establish some basic mgsm properties in the following two lemmas.

<u>Lemma 3.1</u> Let $L \subseteq \Sigma_1^*$ be a language. Let $\mathcal{L}$ be the smallest family containing L and closed under marked union and inverse mgsm mapping. Then for each L_1 in $\mathcal{L}$, there exists a mgsm G such that $L_1 = G^{-1}(L)$.

<u>Lemma 3.2</u> Let $L \subseteq \Sigma_1^*$ be a language, and c be in $\Sigma - \Sigma_1$. Let $\mathcal{L}$ be the smallest family containing L and closed under inverse mgsm mapping and marked $*$. Then for each L_1 in $\mathcal{L}$, there exists a mgsm G such that $L_1 = G^{-1}[(cL)^*]$.

<u>Corollary</u> Let L_1, L_2, ..., and L_n be languages for $n \ge 1$. Let $\mathcal{L}$ be the smallest family containing L_1, L_2, ..., and L_n and closed under inverse mgsm mapping and marked $*$. Then for each L in $\mathcal{L}$, there exists an integer j, $1 \le j \le n$, and a mgsm G such that $L = G^{-1}[(cL_j)^*]$, where $(cL_j)^*$ is a marked $*$ of L_j.

We are now ready for the main result of this section.

<u>Theorem 3.1</u> Closure under marked $*$, marked union, and inverse mgsm mapping are independent of each other.

<u>Proof</u> This proof will be in three parts, labelled (a), (b), and (c).

(a) Closure under inverse mgsm mapping is independent of closure under marked union and marked $*$. To see this, let $L_0 = \{ww^R \mid w \in \{a,b\}+\}$.[4] Let $\mathcal{L}$ be the smallest family containing L_0 and closed under marked union and marked $*$. Note that every L in $\mathcal{L}$ contains a word $w = w_1aw_2$ for some w_1 and w_2. Therefore, the set $\{ww^R \mid w \in \{c,d\}+\}$ is not in $\mathcal{L}$. Hence $\mathcal{L}$ is not closed under inverse mgsm mapping.

(b) Closure under marked $*$ is independent of closure under marked union and inverse mgsm mapping. To see this, let $L_1 = \{a^ib^i \mid i \ge 1\}$, and let $\mathcal{L}_1$ be the smallest family containing L_1 and closed under marked union and inverse mgsm mapping. It sufficies to show that $\mathcal{L}_1$ is not closed under marked $*$. Suppose $(cL_1)^*$ is in $\mathcal{L}_1$. In view of Lemma 3.1, there exists a mgsm $G = (K_1, \{a,b,c\}, \Sigma_1, \delta, \lambda, p_1)$ such that $\{a,b\} \subseteq \Sigma_1$ and $(cL_1)^* = G^{-1}(L_1)$. Let n_1 be a positive integer. Suppose that $\lambda(p_1, ca^{n_1}b^{n_1}) = a^{m_1}b^{m_2}$ for some $m_1 > 0$ and $m_2 > 0$ with $m_2 \le m_1$. Hence for all $w \in L_1$, $|\lambda(p_1, ca^{n_1}b^{n_1}cw)| \le 2m_1$. Then there exists integers $n_3 > 0$ and $n_4 > 0$ and $q \in K_1$ such that $\delta(p_1, ca^{n_1}b^{n_1}ca^{n_3}) = q$, $\delta(q, a^{n_4}) = q$, and $\lambda(q, a^{n_4}) = \epsilon$. Thus $G(ca^{n_1}b^{n_1}ca^{n_3+kn_4}b^{n_3+n_4})$ is in L_1 for all $k \ge 0$. This is a contradiction. Therefore, $\lambda(p_1, cL_1) \subseteq a^*$. Let $k_1 = \text{Max}\{|\lambda(p,x)| \mid p \in K_1, x \in (\Sigma_1 \cup \{\$\})\}$. Since for each $w \in L_1$, $\lambda(p_1, cw\$)$ is in L_1 and $|\lambda(p,\$)| \le k_1$ for all $p \in K_1$, it follows that $|\lambda(p_1, cw\$)| \le 2k_1$. Then there exists integers $n_5 > 0$ and $n_6 > 0$ and $q' \in K_1$

---

(4) If w is a word in Σ_1^*, then w^R is the reverse of w. That is, if $w = x_1x_2\ldots x_n$ with x_1, x_2, ..., and x_n in Σ_1, then $w^R = x_nx_{n-1}\ldots x_2x_1$. Note that $\epsilon^R = \epsilon$.

such that $\delta(p_1, ca^{n_5}) = q'$, $\delta(q', a^{n_6}) = q'$, and $\lambda(q, a^{n_6}) = \epsilon$. Thus $G(ca^{n_5+kn_6} b^{n_5+n_6})$ is in L_1 for all $k \geq 0$. This is a contradiction. Therefore, G cannot exist, and hence $\mathcal{L}_1$ is not closed under marked *.

(c) Closure under marked union is independent of closure under marked * and inverse mgsm mapping. To see this, let $L_2 = \{zz^R \mid z \in \{a,b\}^*\}$ and $L_3 = \{a^n b^n c^n \mid n \geq 1\}$. Let $\mathcal{L}_2$ be the smallest family containing L_2 and L_3 and closed under marked * and inverse mgsm mapping. It suffices to show that $\mathcal{L}_2$ is not closed under marked union. Let L be a marked union of L_2 and L_3, denoted by $(eL_2) \cup (fL_3)$. Suppose L is in $\mathcal{L}_2$. In view of the Corollary to Lemma 3.2, there exists a mgsm G_1 such that either $L = G_1^{-1}[(dL_2)^*]$ or $L = G_1^{-1}[(dL_3)^*]$, where d is a new symbol. By (b) and (g) of Lemma 1.1, there exists mgsm G_2 and G_3 such that either:

(1) $L_2 = G_2^{-1}[(dL_2)^*]$ and $L_3 = G_3^{-1}[(dL_2)^*]$

or

(2) $L_2 = G_3^{-1}[(dL_3)^*]$ and $L_3 = G_2^{-1}[(dL_3)^*]$.

We now show that it is contradictory for either to occur.

Suppose (1) occurs. Then $L_3 = G_3^{-1}[(dL_2)^*]$. Since L_2 is a context free language (CFL) [1] and the family of CFL is a full AFL, it follows that $G_3^{-1}(dL_2)^*$ is a CFL. But, as is well known, L_3 is not a CFL. [1] Hence (1) cannot occur.

Suppose (2) occurs. Then $L_2 = G_3^{-1}[(dL_3)^*]$. Let $G_3 = (K_3, \{a,b\}, \Sigma_3, \delta_3, \lambda_3, p_3)$, where $\{a,b,c,d\} \subseteq \Sigma_3$. Then there exists integers n_1 and n_2 and $q \in K_3$ such that $n_1 > 0$, $n_2 > 0$, $\delta_3(p_3, a^{n_1}) = q$, and $\delta_3(q, a^{n_2}) = q$. Since $G_3(a^{n_1} a^{2kn_2} a^{n_1})$ is in $(dL_3)^*$ for each $k \geq 0$, there exists u_1, u_2, and u_3 such that $\lambda_3(q, a^{n_2}) = u_1 u_2 u_3$, $u_2 \in (dL_3)^*$, $u_3 u_1 \in (dL_3 \cup \{\epsilon\})$, and $\lambda_3(p_3, a^{n_1}) u_1 \in (dL_3)^*$. Five cases arise:

(i) If $\lambda_3(p_3, a^{n_1}) = \epsilon$, then $\lambda_3(q, a^{n_2}) \in (dL_3)^*$.

(ii) If $\lambda_3(p_3, a^{n_1}) \in (dL_3)^* d$, then $\lambda_3(q, a^{n_2}) \in (L_3 d)^*$.

(iii) If $\lambda_3(p_3, a^{n_1}) \in (dL_3)^* da^{m_1}$ for some $m_1 > 0$, then $\lambda_3(q, a^{n_2}) \in (a^{m_2-m_1} b^{m_2} c^{m_2} (dL_3)^* da^{m_1}) \cup \{\epsilon\}$ for some $m_2 \geq m_1$.

(iv) If $\lambda_3(p_3, a^{n_1}) \in (dL_3)^* da^{m_1} b^{m_2}$ for some $m_1 > 0$ and $m_1 \geq m_2 > 0$, then $\lambda_3(q, a^{n_2}) \in (b^{m_1-m_2} c^{m_1} (dL_3)^* da^{m_1} b^{m_2}) \cup \{\epsilon\}$.

(v) If $\lambda_3(p_3, a^{n_1}) \in (dL_3)^* da^{m_1} b^{m_1} c^{m_2}$ for some $m_1 > 0$ and $m_1 \geq m_2 > 0$, then $\lambda_3(q, a^{n_2}) \in (a^{m_1-m_2} (dL_3)^* da^{m_3} b^{m_3} c^{m_3+m_2-m_1}) \cup \{\epsilon\}$ for some $m_3 > m_1 - m_2$.

Since $G(a^{n_1+n_2} b^2 a^{n_1+n_2})$ is in $(dL_3)^*$, cases (i) thru (v) imply that $G(a^{n_1+kn_2} b^2 a^{n_1+n_2})$ is in $(dL_3)^*$ for all $k \geq 1$. Thus G_3 cannot exist. Hence $\mathcal{L}_2$ is not closed under marked union. This completes the proof of the theorem.

In the proof of (c) of the previous theorem, it was necessary to use a family of languages generated by two distinct languages. A natural question is the existence of a counterexample which is generated by a single language. In other words, does there exist a family of languages generated by one language which is closed under marked * and inverse mgsm mapping and not closed under marked union. We now show that the answer is negative. Let $L \subseteq \Sigma_1^*$ be a language, and $\mathcal{L}$ be the smallest family containing L and closed under marked * and inverse mgsm mapping. In view of Lemma 3.2, if L_0 is in $\mathcal{L}$, then there exists a mgsm G such that $L_0 = G^{-1}[(cL)^*]$, where $(cL)^*$ is a marked * of L. Let L_1 and L_2 be in $\mathcal{L}$, and G_1 and G_2 be mgsm such that $L_1 = G_1^{-1}[(cL)^*]$ and $L_2 = G_2^{-1}[(cL)^*]$. As shown in the proof of Lemma 3.1, there exists a mgsm G_3 such that $G_3^{-1}[(cL)^*] = aL_1 \cup bL_2$, where $aL_1 \cup bL_2$ is a marked union of L_1 and L_2. Hence $\mathcal{L}$ is closed under marked union.

References

1. Ginsburg, S., "The Mathematical Theory of Context Free Languages," McGraw-Hill, New York, 1966.

2. Ginsburg, S. and Greibach, S. A. Abstract families of languages. IEEE Conference Record on Switching and Automata Theory, 1967, pp. 128-139.

3. Ginsburg, S. and Greibach, S. A. Deterministic context free languages. Information and Control 9 (December 1966), pp. 620-648.

4. Ginsburg, S. and Spanier, E. H. Finite-turn pushdown automata. SIAM J. Control 4 (1967), pp. 172-201.

5. Ginsburg, S., Greibach, S. A., and Harrison, M. A. One-way stack automata. J. ACM 14 (April 1967), pp. 389-418.

6. Hopcroft, J. E. and Ullman, J. D. An approach to a unified theory of automata. Bell System Technical J. 46 (1967), pp. 1793-1829.

Acknowledgement

The author thanks Professor Seymour Ginsburg for suggestions and Caryl Johnston for typing this paper. This research sponsored in part by, but does not necessarily constitute the opinion of, the Air Force Cambridge Research Laboratories, Office of Aerospace Research USAF, under Contract F1962867C0008, and by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under AFOSR Grant No. AF-AFOSR-1203-67, and in part by the University of Southern California.