On Some Families of Languages related to Developmental Systems

K. Culik, II

CS-73-12

Department of Applied Analysis & Computer Science

University of Waterloo
Waterloo, Ontario, Canada

Faculty of Mathematics

University of Waterloo

Waterloo, Ontario
Canada

Department of Applied Analysis
&
Computer Science

On Some Families of Languages related to Developmental Systems

K. Culik, II

CS-73-12

Department of Applied Analysis &
Computer Science

University of Waterloo
Waterloo, Ontario, Canada

This work was supported by the National Research Council of Canada, Grant No. A7403.

ABSTRACT

OL systems and TOL systems are the simplest mathematical models for the study of the development of biological organisms with or without a variable environment, respectively. This paper contributes to the study of the properties of the languages generated by these systems and by their generalizations. Macro OL (TOL) languages are languages obtained by substituting languages of a given type in OL (TOL) languages. We study properties of certain families of macro OL (TOL) languages, in particular we show that they are full AFL's.

We observe that OL, TOL systems and many of their generalizations can be viewed as special classes of index grammars.

0. Introduction

In 1968 A. Lindemayer introduced mathematical models for developmental systems [8], now called L-systems. Later, he devoted specific attention to systems without interaction, so called OL systems [9]. In [10] a modification of OL systems, called TOL systems, is described to give a model of the behaviour of organisms in the presence of a variable environment. We ask the reader to see the references for the biological motivation of the considered systems. Formal definitions of OL and TOL systems are given as special cases of extended systems in the preliminaries. Mathematical properties of these systems have been studied quite extensively, e.g. [7, 10, 12], and recently various generalizations of OL systems were introduced which, unlike OL systems, generate languages with nice mathematical properties.

In the preliminaries we give modified definitions of extended table systems and their subclasses [11], and of macro OL (TOL) systems [3]. We will observe that the former are equivalent to certain special cases of the latter. As defined here a language L is an $\mathcal{L}$ macro OL (TOL) language, for some family of languages $\mathcal{L}$, iff L is the result of a substitution of type $\mathcal{L}$ into an OL (TOL) language.

In [3] it was shown that the family of regular macro OL languages (RMOL) is the smallest full AFL containing OL languages (full AFL closure of OL languages) and [11] the family of extended table OL languages was shown to be a full AFL. In the present paper we extend these results and we show that for every (full) AFL $\mathcal{L}$ the families $\mathcal{L}$ MOL and $\mathcal{L}$ MTOL are also (full) AFL's. Moreover, if $\mathcal{L}$ is principal [5] then they are also principal. To show the last result we have to show, first, that RMOL is a full principal AFL. We will do it by explicitly constructing a finite transducer (a-transducer in [4]) generator for the RMOL languages over a fixed number of symbols. For a given alphabet such a generator is a subset of the corresponding Dyck language. It is well known that every context-free language can be expressed in the form t(D) or in the form $h(D \cap R)$ where t is a finite transduction, h is a homomorphism, D is the Dyck language over a suitable alphabet and R is a regular set. We will show that for macro OL languages the situation is different - by 'homomorphic characterization' we obtain a proper subfamily of RMOL languages, namely FMOL languages. On the other hand, the families of finite macro TOL languages and regular macro TOL languages are identical and as for context free languages, there exists a single generator which characterizes them either 'homomorphically' or by finite transduction.

We will also show that the family of RMOL languages is not closed under substitution (of itself) and that the result of this substitution, the family RMOLMOL is again a full principal AFL. If we repeatedly perform RMOL-substitutions we can get an infinite hierarchy of full principal AFL properly between the context free and index languages [1].

In the last section we elaborate the relationship of (generalized) developmental systems to index grammars. We show that all the considered families of languages are included in index languages and, moreover, that all the classes of systems considered in [11], namely OL, TOL, EOL and ETOL systems may be looked upon as special classes of index grammars. In particular, a 'table' of an ETOL (TOL) system becomes exactly a 'flag' of the corresponding index grammar. Consequently, for example, it follows immediately that decidable questions about index grammars (membership and emptiness problem) are also decidable for all systems mentioned above and that the family of languages generated by any of these systems is properly included in the family of context sensitive languages.

1. Preliminaries

We assume knowledge of the basic notions and notation of formal language theory, see e.g. [2, 13].

We start with a slightly modified, but equivalent, definition of extended table OL systems [11].

Definition 1.1: An extended table L-system without interaction (ETOL system) is a 4-tuple $G = (V,T,\mathcal{P},\sigma)$ where

(i) V is a finite nonempty set, the alphabet[1] of G,

(ii) $T \subseteq V$, the terminal alphabet of G,

(iii) $\mathcal{P}$ is a finite set of tables. $\mathcal{P} = \{P_1,\dots,P_n\}$ for some $n \geq 1$, where each $P_i \subseteq V \times V^*$. Element (u,v) of P_i, $1 \leq i \leq n$, is called a production and is usually written in the form $u \to v$. Every P_i, $1 \leq i \leq n$, satisfies the following (completeness) condition: For each $a \in V$ there is $w \in V^*$ so that $(a,w) \in P_i$,

(iv) $\sigma \in V^+$, the axiom of G.

Definition 1.2: An ETOL-system $G = (V,T,\mathcal{P},\sigma)$ is called

(i) a TOL-system if $V = T$;

(ii) an EOL-system if $\mathcal{P} = \{P_1\}$;

(iii) an OL-system if $V = T$ and $\mathcal{P} = \{P_1\}$.

Definition 1.3: Given an ETOL-system $G = (V,T,P,\sigma)$ we write $x \underset{G}{\Rightarrow} y$ if there exist $a_1,\dots,a_k \in V$ and $y_1,\dots,y_k \in V^*$ so that, $x = a_1,\dots,a_k$, $y = y_1,\dots,y_k$ and for some $P_i \in \mathcal{P}$, $a_j \to y_j \in P_i$, $j = 1,\dots,k$.

The transitive and reflexive closure of binary relation $\underset{G}{\Rightarrow}$ is denoted by $\underset{G}{\Rightarrow}^*$.

Definition 1.4: Let $G = (V,T,\wp,\sigma)$ be an ETOL system. The language generated by G is denoted by L(G) and defined as $L(G) = \{w \in T^*: \sigma \underset{G}{\Rightarrow}^* w\}$.

Definition 1.5: A language generated by an XYZ system, for any type XYZ is said to be an XYZ language. The family of all XYZ languages is denoted by XYZ. In particular, F, R_0, and R denote the families of finite sets, ε-free regular sets, and regular sets, respectively.

Note that the requirement of completeness is essential for OL and TOL systems. It is, however, unessential for systems allowing nonterminals. We can easily show it by adding a 'dead' nonterminal similar to a 'dead' state in the well known technique for finite automata (see also Theorem 3 of [11]).

Now we give a quite different but obviously equivalent definition of macro OL languages [3] as a special case of a more general definition. We will define macro OL languages directly instead of defining the associated macro OL systems as in [3].

Notation: Let $\mathcal{L}_1, \mathcal{L}_2$ be families of languages. Then $\mathrm{Sub}(\mathcal{L}_1,\mathcal{L}_2) = \{f(L): L \text{ is in } \mathcal{L}_2 \text{ and } f \text{ is an } \mathcal{L}_1\text{-substitution}^2\}$.

Definition 1.6: Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be families of languages. The L is called an $\mathcal{L}_1$ macro $\mathcal{L}_2$ language iff L is the result of an $\mathcal{L}_1$-substitution[2] into a language in $\mathcal{L}_2$. The family of $\mathcal{L}_1$ macro $\mathcal{L}_2$ languages is denoted by $\mathcal{L}_1 M \mathcal{L}_2$, i.e. $\mathcal{L}_1 M \mathcal{L}_2 = \mathrm{Sub}(\mathcal{L}_1,\mathcal{L}_2)$.

The equivalence of Definition 6.1 in the case that $\mathcal{L}_2$ = OL, and the definition of $\mathcal{L}_1$MOL languages in [3] follows from the note about completeness above and from the fact that in $\mathcal{L}_1$MOL systems any derivation leading to a terminal string produces terminals only in its last step. Consequently we also have the following:

Lemma 1.1: FMOL = EOL.

We need to use FMOL and FMTOL systems in some constructions. To avoid further definitions let us define them as modified ETOL-systems. An FMTOL system is an ETOL system $G = (N \cup T, T, \{P_1,\ldots,P_n\},S)$ where $S \in N$, $P_i \subset N \times (N^+ \cup T^*)$ and P_i does not require completeness for $1 \leq i \leq n$. Moreover if $n = 1$ then the system is called an FMOL system.

The biological motivation of macro OL systems was discussed in [3]. If such models are extended to consider a variable environment formalized by 'tables' [10], we get macro TOL languages whose simplest case, the finite MTOL languages, has already been studied under the name of ETOL languages in [11]. Indeed, in the same way that EOL = FMOL, it is also easy to verify the first equality of the following:

Lemma 1.2: ETOL = FMTOL = RMTOL.

Proof: It remains to show the second equality. By definition RMTOL = Sub(R,TOL). Obviously, Sub(R,TOL) $\subseteq$ Sub(R,ETOL). In [11] it was shown that ETOL is a full AFL. Every full AFL is closed under regular substitution [4]. Thus RMTOL = ETOL. □

2. Properties of Families of Macro OL Languages

The following lemma will be very useful.

Lemma 2.1: Let $\mathcal{L}_1$ be an AFL and $\mathcal{L}_2$ be any family of languages. Then

$$\mathrm{Sub}(\mathcal{L}_1,\mathcal{L}_2) = \mathrm{Sub}(\mathcal{L}_1,R_0M\mathcal{L}_2).$$

(R_0 is defined in Def.1.5 as the family of ε-free regular sets).

Proof: For every AFL $\mathcal{L}_1$ containing ε we have $\mathrm{Sub}(\mathcal{L}_1,R_0) = \mathcal{L}_1$ [6]. Therefore by using the definition of $R_0M\mathcal{L}_2$, and the associativity of substitutions we have $\mathrm{Sub}(\mathcal{L}_1,R_0M\mathcal{L}_2) = \mathrm{Sub}(\mathcal{L}_1,\mathrm{Sub}(R_0,\mathcal{L}_2)) =$ $= \mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1,R_0),\mathcal{L}_2) = \mathrm{Sub}(\mathcal{L}_1,\mathcal{L}_2)$. □

Theorem 2.1: Let $\mathcal{L}$ be an (full) AFL. Then the families $\mathcal{L}$MOL and $\mathcal{L}$MTOL are (full) AFL's.

Proof: By definition and by Lemma 2.1

$$\mathcal{L}\mathrm{MOL} = \mathrm{Sub}(\mathcal{L},\mathrm{OL}) = \mathrm{Sub}(\mathcal{L},R_0\mathrm{MOL}) \text{ and}$$

$$\mathcal{L}\mathrm{MTOL} = \mathrm{Sub}(\mathcal{L},\mathrm{TOL}) = \mathrm{Sub}(\mathcal{L},R_0\mathrm{MTOL}).$$

RMOL was shown to be a full AFL in [3]. By Lemma 2.2 RMTOL = ETOL and ETOL was shown to be a full AFL in [11]. Therefore, clearly, R_0MOL and R_0MTOL are AFL's and $\mathrm{Sub}(\mathcal{L},R_0\mathrm{MOL})$ and $\mathrm{Sub}(\mathcal{L},R_0\mathrm{MTOL})$ are (full) AFL's by [6, Corollary 1]. □

Now we proceed to develop a homomorphic characterization of FMOL languages and consequently to show principality of our AFL's.

First we need to show an auxiliary result which will play a similar role for FMOL languages as the Chomsky Normal Theorem does in the proof of the homomorphic characterization of context-free langguages.

Lemma 2.3: Let $L \in$ FMOL, $L \subseteq T^*$ and b is not in T. Then there exists an FMOL system $G = (N \cup T \cup \{b\}, T \cup \{b\}, P, S)$ such that $P \subset N \times (N^2 \cup T \cup \{b\})$ and $L = \mu(L(G))$ where μ is the homomorphism defined by $\mu(a) = a$ for $a \in T$ and $\mu(b) = \varepsilon$.

Proof: Given an FMOL system $G = (N \cup T,T,P,S)$ let r be the maximum of the lengths of the right sides of the productions in P and 2. Let b not be in $N \cup T$ and let $P_1 = \{a_0 \to a_1 \ldots a_s b^{r-s} : a_0 \to a_1 \ldots a_s \in P\} \cup \{b \to b^r\}$, i.e. we fill each production in P by "blanks" up to length r.

Now construct the FMOL system $G' = (N' \cup T,T,P',S')$ where $N' = (N \cup T \cup \{b\} \cup P_1) \times \{1,\ldots,r-1\}$, $S' = (S,1)$ and P' is given as follows:

(i) If $\pi \in P_1$, $\pi = a_0 \to a_1 \ldots a_r$ for $a_i \in N \cup T \cup \{b\}$, then the following productions are in P':

(a) $(a_0,1) \to (a_1,r-1)(\pi,r-1)$;

(b) $(\pi,k+1) \to (a_{r-k},k)(\pi,k)$ for $2 \le k \le r-2$;

(c) $(\pi,2) \to (a_{r-1},1)(a_r,1)$.

(ii) $(a,k+1) \to (a,k)(b,k)$ is in P' for each $a \in N \cup T \cup \{b\}$;

(iii) $(a,1) \to a$ is in P' for each $a \in T \cup \{b\}$.

Clearly, $\mu(L(G')) = L(G)$ and G' is of the required form. □

We have also proved the following normal form theorem for FMOL which is of interest in itself.

Theorem 2.2: Every language in FMOL can be generated by an FMOL system $G = (N \cup T, T, P, S)$ such that $N \cap T = \phi$ and $P \subseteq N \times (N^2 \cup T \cup \{\varepsilon\})$.

Proof: Replace the production $(b,1) \to b$ in the system constructed in the proof of Lemma 2.3 by the production $(b,1) \to \varepsilon$. □

Now we define subsets of Dyck languages [13] which will be used to characterise the families FMOL and RMOL. We will distinguish two kinds of 'brackets' : a_k, a_k' for $1 \leq k \leq m$ which may occur only as inner-most brackets and c_k, c_k' for $1 \leq k \leq n$ which may occur only as non-inner-most brackets for some $m,n \geq 1$. This distinction is unessential but will make the proof of Theorem 2.3 more transparent. See also the note at the end of the proof of Theorem 2.3.

Definition 2.1: Given $\Sigma_{m,n} = \{a_1,\ldots,a_m,a_1',\ldots,a_m',c_1,\ldots,c_n,c_1',\ldots,c_n'\}$ for $m,n \geq 1$, let $G_{m,n}$ be the FMOL system $(N_{m,n},\Sigma_{m,n},P_{m,n},S)$ where $N_{m,n} = \{S,C_1,\ldots,C_n,C_1',\ldots,C_n'\}$ and $P_{m,n}$ consists of the productions

$$S \to C_k S C_k' S \quad \text{for } 1 \leq k \leq n,$$

$$S \to a_i a_i' \quad \text{for } 1 \leq i \leq m,$$

$$C_k \to c_k \mid C_k \quad \text{for } 1 \leq k \leq n, \text{ and}$$

$$C_k' \to c_k' \mid C_k' \quad \text{for } 1 \leq k \leq n.$$

Denote $L(G_{m,n})$ by $D_{m,n}^{OL}$.

Theorem 2.3: A language L is in FMOL if and only if there exist $m,n \geq 1$, an homomorphism h on $\Sigma^*_{m,n}$ and a regular set $R \subseteq \Sigma^*_{m,n}$ such that $L = h(D^{OL}_{m,n} \cap R)$.

Proof: Let $L \subseteq T^*$ and $L = \mu(L(G))$ for FMOL system $G = (N \cup T \cup \{b\}, T \cup \{b\}, P, S)$ and homomorphism μ from Lemma 2.3. Let $T' = T \cup \{b\} = \{a_1, a_2, \ldots, a_m\}$ and $P = \{\pi_1, \pi_2, \ldots, \pi_n\}$. Note that there is no loss of generality in assuming that T' is a fixed alphabet $\{a_1, \ldots, a_m\}$ for any set T' of cardinality m.

Consider now the alphabet $\Sigma_{m,n} = \{a_1, \ldots, a_m, a_1', \ldots, a_m', c_1, \ldots, c_n, c_1', \ldots, c_n'\}$ from definition 2.1, and construct the right linear grammar $G_R = (N, \Sigma_{m,n}, P_R, S)$ where P_R is defined as follows:

(i) If $A \to a$ is in P for $a \in T'$, then $A \to aa'$ is in P_R.

(ii) If $\pi_i = A \to BC$ for B,C in N then

(a) $A \to c_i B$ is in P_R, and

(b) $D \to aa'c_i'C$ is in P_R for every $D \to a$ in P where $a \in T'$.

Let $R = L(G_R)$ and h_1 be the homomorphism on $\Sigma^*_{m,n}$ defined by $h_1(a) = a$ for $a \in T'$, $h_1(a) = \varepsilon$ for $\Sigma_{m,n} - T'$. Similar to the proof of Theorem 3.7.1 in [13] for context-free grammars we can verify that $h_1(D^{OL}_{m,n} \cap R) = L(G)$. Note that $D^{OL}_{m,n}$ is dependent on language L and therefore we may use a different pair of symbols c_i, c_i' for each production in G.

Finally, let h denote the composition of h_1 and μ, i.e. h is the homomorphism defined by $h(a) = \mu(h_1(a))$ for each $a \in \Sigma_{m,n}$. Then $L = h(D^{OL}_{m,n} \cap R)$ and we have the required characterization for every language in FMOL. We get exactly the languages in FMOL since $D^{OL}_{m,n}$ is in FMOL and the family FMOL is closed under homomorphism and intersection with a regular set [3]. □

Note that we may omit the symbols $a'_1,\ldots,a'_m$ in the definition of $D^{OL}_{m,n}$ and in construction of G_R. We have not chosen this simplification because we wanted the languages $D^{OL}_{m,n}$ to be subsets of Dyck languages.

Corollary 2.1: A language L is in RMOL if and only if there exists $n \geq 1$ and a finite transduction[3] so that $L = t(D^{OL}_n)$.

Proof: In [3] it was shown that RMOL is the closure of OL under finite transductions. Thus the result follows immediately by Theorem 2.2. □

Corollary 2.2: RMOL is a full principal AFL.

Proof: We have shown that the AFL RMOL is generated by the family $\mathcal{L} = \{D^{OL}_{m,n} : m,n \geq 1\}$. However, using well known techniques, all the languages in $\mathcal{L}$ can be encoded by a simple language D^{OL} over a two letter alphabet, say $\{0,1\}$, so that for every m and n there exists a finite transduction $t_{m,n}$ so that $D^{OL}_{m,n} = t_{m,n}(D^{OL})$. The proof is completed by the fact that finite transductions are closed under composition. □

Theorem 2.4: Let $\mathcal{L}$ be a full principal AFL. Then $\mathcal{L}$MOL is a full principal AFL.

Proof: By Lemma 2.1 and because $\mathcal{L}$ is a full AFL $\mathcal{L}$ MOL = Sub($\mathcal{L}$,R_0MOL) = = Sub($\mathcal{L}$,RMOL). By Corollary 2.2 RMOL is full principal. In [5, Corollary 1] it was shown that the result of the substitution of a full principal AFL into a full principal AFL again has this property. □

By Theorem 2.4 RMOLMOL is also a full principal AFL. We will show that it properly contains RMOL, i.e. that RMOL is not closed under substitution. We may repeatedly substitute RMOL and get an infinite hierarchy of full AFL $((\mathrm{RMOL})^n\mathrm{MOL}\ n = 1,2,\ldots)$ between the families of context-free and index languages.

To show RMOL $\subsetneq$ RMOLMOL we need the following

Lemma 2.4: Let L be in RMOL but not in FMOL. Then there is a string x in L so that $x = uvw$, $v \neq \varepsilon$ and $uv^kw \in L$ for any $k \geq 0$.

Proof: Since L is in RMOL there exists a regular substitution f and an OL language L', $L' \subseteq \Sigma^*$, so that $L = f(L')$. Since L is not in FMOL f(a) is not finite for at least one a in Σ such that yaz is in L' for some y,z in Σ^*. Therefore, the proof is completed by applying the 'pumping lemma' (see [2] p.128) to the regular set f(a). □

Theorem 2.5: RMOLMOL is a full principal AFL properly containing RMOL.

Proof: RMOLMOL is a full principal AFL by Theorem 2.4. Obviously, it includes RMOL, therefore it remains to show that the inclusion is proper.

Let $L_1 = \{a^{2^n} : n \geq 1\}$, $L_2 = \{b^{2^n} : n \geq 1\}$ and let the substitution f be defined by $f(a) = \{a\}L_2$. Let $L = f(L_1)$, i.e. L is in RMOLMOL. Let h be the homomorphism defined by $h(a) = a$, $h(b) = \varepsilon$. In [7] it was shown that $h^{-1}(L_1)$ is not in FMOL (= EOL). Consider the finite substitution defined by $f(a) = \{a\}$, $f(b) = \{b,\varepsilon\}$. Clearly, $h^{-1}(L_1) = f(L)$ and therefore L is not in FMOL since FMOL is closed under finite substitution.

Since, obviously, L does not satisfy the requirements of Lemma 2.4 L is not in RMOL and the proof is completed. □

Note: It was stated in [3] that RMOL is properly included in the family of index languages [1]. Inclusion was shown but the proof of proper inclusion was omitted. Now, this follows from Theorem 2.5 since index languages are closed under substitution [1] and therefore include RMOLMOL.

3. Properties of Families of Macro TOL Languages

Most of the results proved in the last section for families of macro OL languages hold for families of macro TOL languages with minor modifications.

We start with the definition of 'generators' of FMTOL languages. To simplify the constructions we use the result from [11] that every language in ETOL (= FMTOL) can be generated by a system with only two tables.

Definition 3.1: Given the alphabet $\Delta_{m,n} = \{a_i, a_i' : 1 \leq i \leq m\} \cup \{c_{k,j}, c_{k,j}' : 1 \leq k \leq n,\ j = 1,2\}$ for $m,n \geq 1$, let $G_{m,n}^T$ be the FMTOL system $(N, \Delta_{m,n}, \{P_1, P_2\}, S)$ where $N = \{S\} \cup \{C_{k,j}, C_{k,j}' : 1 \leq k \leq n,\ j = 1,2\}$ and $P_j = \{S \to C_{k,j} S C_{k,j}' S : 1 \leq k \leq n\} \cup P'$ for $j = 1,2$ and $P' = \{S \to a_i a_i' : 1 \leq i \leq m\} \cup \{C_{k,j} \to C_{k,j} | c_{k,j} : 1 \leq k \leq n,\ j = 1,2\} \cup \{C_{k,j}' \to C_{k,j}' | c_{k,j}' : 1 \leq k \leq n,\ j = 1,2\}$. Denote $L(G_{m,n}^T)$ by $D_{m,n}^T$.

The result from [11] that each language in FMTOL can be generated by a system with no more than two tables can easily be modified as follows.

Lemma 3.1: Every language in FMTOL is generated by an FMTOL system $G = (N \cup T, T, \{P_1, P_2, P_3\}, S)$ where $P_1, P_2 \subseteq N \times N^*$ and $P_3 \subseteq N \times T^*$.

We can now generalize Lemma 2.3:

Lemma 3.2: Let $L \in$ FMTOL, $L \subseteq T^*$ and b not be in T. Then there exists an FMTOL system $G = (N \cup T \cup \{b\}, T \cup \{b\}, \{P_1, P_2, P_3\}, S)$ such that

$P_1, P_2 \subseteq N \times N^2$, $P_3 \subseteq N \times (T \cup \{b\})$ and $L = \mu(L(G))$ where h is the homomorphism defined by $\mu(a) = a$ for $a \in T$ and $\mu(b) = \varepsilon$.

Proof: By Lemma 3.1 and the method of the proof of Lemma 2.3. □

We omit the obvious generalization of Theorem 2.2 and proceed to the main result of this section.

Theorem 3.1: A language L is in FMTOL if and only if there exists $m,n \geq 1$, an homomorphism h on $\Delta^*_{m,n}$ and a regular set $R \subset \Delta^*_{m,n}$ such that $L = h(D^T_{m,n} \cap R)$.

Proof: Let $L \subseteq T^*$ and $L = \mu(L(G))$ for FMTOL system $G = (N \cup T \cup \{b\}, T \cup \{b\}, \{P_1,P_2,P_3\},S)$ and homomorphism μ from Lemma 3.2. Let $T' = T \cup \{b\} = \{a_1,\ldots,a_m\}$, $P_1 = \{\pi^1_1,\ldots,\pi^1_r\}$ and $P_2 = \{\pi^2_1,\ldots,\pi^2_s\}$. Let $n = \max(r,s)$ and $\Delta_{m,n}$ be the alphabet from Definition 3.1.

Construct the right linear grammar $G_R = (N,\Delta_{m,n},P_R,S)$ where P_R is defined as follows:

(i) If $A \to a$ is in P, then $A \to aa'$ is in P_R.

(ii) If $\pi^j_i = A \to BC$ is in P_j for B,C in N, then

(a) $A \to c_{i,j}B$ is in P_R, and

(b) $D \to aa'c'_{i,j}C$ is in P_R for every $D \to a$ in P_3.

The remaining part of the proof is the same as the proof of Theorem 2.3 substituting $D^T_{m,n}$ for $D^{OL}_{m,n}$.

We obtain exactly the languages in FMTOL since this family is closed under homomorphism and intersection with a regular set [11]. □

Unlike the case of macro OL languages in section 2, here the homomorphic and the finite transduction characterizations give the same family FMTOL.

Corollary 3.1: A language L is in FMTOL (= RMTOL = ETOL) if and only if there exist $m,n \geq 1$ and a finite transduction so that $L = t(D^T_{m,n})$.

Proof: In [11] it was shown that ETOL is a full AFL and therefore it is closed under finite transductions [4]. For equality of the three families see Lemma 1.2. □

Corollary 3.2: FMTOL (= RMTOL = ETOL) is a full principal AFL.

Proof: See the proof of Corollary 2.2. □

4. Subsets of Index Grammars

We refer to [1] for the definition of index grammars and index languages. We will show that the families OL, TOL, FMOL (= EOL), FMTOL (= ETOL) are subfamilies of index languages and, moreover, that the corresponding systems closely correspond to certain subclasses of index grammars. In particular, a 'table' of an (E)TOL system is nothing else than a 'flag' of the corresponding index grammar.

Definition 4.1: An index grammar $G = (N,T,F,Q,S)$ is called an ETOL-index grammar iff $F = F_1 \cup F_2$ and $N = N' \cup \{S,A\}$ such that

(i) $F_1 \cap F_2 = \phi$, $N' \cap \{A,S\} = \phi$;

(ii) $Q = \{S \to Ag : g \in F_2\} \cup \{A \to Af : f \in F_1\} \cup \{A \to B\}$ for some $B \in N'$;

(iii) $f \subseteq N' \times N'^*$ for each f in F_1;

(iv) $g \subseteq N' \times T^*$ for each g in F_2.

Definition 4.2: Using the notation of Definition 4.1, an ETOL-index grammar G is called

(1) EOL-index grammar if $|F_1| = |F_2| = 1$.

(2) TOL-index grammar if there is an one-to-one mapping μ from N onto T so that $F_2 = \{g\}$ and $g = \{A \to \mu(A) : A \in N\}$, and if for each f in F_1 and each A in N there is $w \in N^*$ so that $A \to w \in f$.

(3) OL-index grammar if both conditions above hold.

We can easily give additional potentially interesting subclasses of index grammars. For example, it is easy to describe a class of index grammars equivalent to ETOL-systems with regular control sets [14], i.e. ETOL systems in which the manner of generating a string is restricted by allowing only given sequences of productions.

Theorem 4.1: The languages generated by X-index grammars are exactly X languages, for X = OL, TOL, EOL, ETOL.

Proof: We will only prove the result for ETOL (= FMTOL). It is easy to verify it for the other types.

1. Given an FMTOL system $G = (N,T,\{P_1,\ldots,P_n\},Z)$ construct an ETOL-index grammar $G' = (N \cup \{S,A\},F,Q,S)$ where $\{S,A\} \cap N = \phi$, Q is from Definition 4.1 and $F = \{P_j^i : i = 1,2;\ j = 1,\ldots,n\}$ for $P_j^1 = P_j \cap N^*$ and $P_j^2 = P_j \cap T^*$ for $j = 1,\ldots,n$. Clearly, $L(G) = L(G')$.

2. Given an ETOL-index grammar $G = (N,T,F,Q,S)$ with $F = F_1 \cup F_2$ and $N = N' \cup \{S,A\}$ satisfying the conditions of Definition 4.1 construct an FMTOL system $G' = (N',T,F,B)$ where B is the nonterminal for which $A \to B$ and A is not in N'. Clearly, $L(G) = L(G')$. □

We have immediately the following two corollaries. The first one is a refinement of already known results with lengthy proofs in [10,12].

Corollary 4.1: The families OL, TOL, EOL (= FMOL), RMOL, FMTOL (= ETOL) are included in the family of index languages and thus properly included in the family of context sensitive languages. Thus the membership problem is decidable for corresponding systems.

Proof: Inclusion in index languages follows from Theorem 4.1. We do not need the part of Theorem 4.1 whose proof has been omitted since OL, TOL, EOL and RMOL are subfamilies of FMTOL. Proper inclusion of index languages in context sensitive languages was shown in [1]. □

Corollary 4.2: The emptiness problem is recursively decidable for FMOL, RMOL and FMTOL (ETOL) systems.

Proof: The emptiness problem is decidable for index languages [1]. Note that the emptiness problem is trivial for OL and TOL systems, they always generate a nonempty language. □

ACKNOWLEDGEMENT

The author is grateful to D. Wood, J. Opatrny and T.W. Ling for discussions concerning the research reported here.

REFERENCES

[1] A.V. AHO, Indexed Grammars - An Extension of Context-free Grammars, J. Assoc. Comput. Mach. 15 (1968), 647-671.

[2] A.V. AHO and J.D. ULLMAN, The Theory of Parsing, Translation, and Compiling, Prentice-Hall, Englewood Cliffs, New Jersey, 1972.

[3] K. CULIK, II and J. OPATRNY, Macro OL-systems, Research Report CS-73-06, University of Waterloo, Waterloo, Ontario, Canada, 1973.

[4] S. GINSBURG and S. GREIBACH, Abstract Families of Languages, Memoirs Amer. Math. Soc. 87 (1969), 1-32.

[5] S. GINSBURG and S. GREIBACH, Principal AFL, JCSS 4 (1970), pp.308-338.

[6] S. GINSBURG and E.H. SPANIER, Substitutions in Families of Languages, Information Sciences 2 (1970), 83-110.

[7] G.T. HERMAN, Closure Properties of Some Families of Languages Associated with Biological Systems, Int. J. of Comp. Math, to appear. Also in Proceedings of the Fifth Annual Princeton Conference on Information Sciences, Princeton, 1971.

[8] A. LINDENMAYER, Mathematical Models for Cellular Interactions in Development, J. Theoret. Biol. 18 (1968), 280-315.

[9] A. LINDENMAYER, Developmental Systems without Cellular Interactions, their Languages and Grammars, J. Theoret. Biol. 30 (1971), 455-484.

[10] G. ROZENBERG, TOL Systems and Languages, Information and Control, to appear.

[11] G. ROZENBERG, Extension of Tables OL-systems and Languages, R. Report 53-1972, Department of Comp. Science SUNY at Buffalo, Amherst N.Y., U.S.A., 1972.

[12] G. ROZENBERG and P.G. DOUCET, On OL-Languages, Information and Control, 19 (1971), 302-318.

[13] S. GINSBURG, The Mathematical Theory of Context-free Languages, McGraw-Hill, New York, 1966.

[14] A. SALOMAA, Formal Languages, Academic Press, New York, to appear.

FOOTNOTES

1. For an alphabet V, V^* is the free semigroup with identity ε generated by V. An element $a_1 \ldots a_n$ of V^* is called a string. A subset of V^* is called a language over V. $V^*-\{\varepsilon\}$ is denoted by V^+.

2. Let $L \subseteq \Sigma^*$ and for each a in Σ let L_a be a language of type A. Let f be the function defined on Σ^* by $f(\varepsilon) = \{\varepsilon\}$, $f(a) = L_a$ for each a in Σ, and $f(a_1 \ldots a_k) = f(a_1) \ldots f(a_k)$ for each $n \geq 1$ and a_i in Σ. Then f is called an A-substitution. f is extended to languages over Σ by defining $f(X) = \cup_{x \text{ in } X} f(x)$ for all $X \subseteq \Sigma^*$.

3. A finite transducer is a 6-tuple $M = (K,\Sigma,\Delta,H,q_0,F)$ where K is a finite set of states, Σ and Δ are input and output alphabets, H is a finite subset of $K \times \Sigma^* \times \Delta^* \times K$, $q_0 \in K$ and $F \subseteq K$. Let $\vdash$ be the relation on $K \times \Sigma^* \times \Delta^*$ defined as follows: Let $(p,xw,z_1) \vdash (q,w,z_2)$ if (p,x,y,q) is in H and $z_2 = z_1 y$. Let $\vdash^*$ be the reflexive and transitive closure of $\vdash$.

M defines a mapping t from Σ^* into 2^{Δ^*}, called a finite transduction and is defined by $t(w) = \{z:(q_0,w,\varepsilon) \vdash^* (p,\varepsilon,z) \text{ for some p in } F\}$. The mapping t is extended for every $L \subseteq \Sigma^*$ by $t(L) = \cup_{w \text{ in } L} t(w)$.