Proceedings of Grammar Systems Week 2004, edited by E. Csuhaj-Varjú and Gy. Vaszil, MTA SZTAKI, Budapest, pages 50 - 65, 2004.

Active Symbols in Pure Systems

Suna Bensch, Henning Bordihn
Universität Potsdam, Institut für Informatik
August-Bebel-Str. 89, 14482 Potsdam, Germany
{aydin, henning}@cs.uni-potsdam.de

Abstract

In this paper, we consider the number of (statically measured) active symbols for Lindenmayer systems without interaction and some variants thereof as well as for pure CD grammar systems, where no distinction between terminal and nonterminal symbols is made. This measure of descriptional complexity gives rise to infinite hierarchies in all cases considered here. Moreover, all the devices under consideration are compared with respect to their generative power when the number of active symbols is bounded. Finally, some closure and non-closure properties of the corresponding language families with a fixed number of active symbols are proved.

1 Introduction

In the last years the concept of active symbols was studied in several papers [5, 8, 10, 14, 15] within the framework of extended tabled Lindenmayer systems without interaction (ET0L systems). A symbols is said to be active if and only if it can be non-identically rewritten. ¿From the biological point of view active symbols can be interpreted as the maximum number of cells which are simultaneously contributing to the growing of the organism.

The authors in [2] investigate the concept of active symbols for deterministic ET0L systems (EDT0L systems) as well as for cooperating distributed grammar systems (CD grammar systems for short) working in the t-mode of derivation. Furthermore, they introduce the notion of *dynamically active symbols*.

In this paper we will only consider the *statically measured active symbols* which we will refer to as *active symbols* throughout the paper.

CD grammar systems have been introduced in [3] and have further been investigated in [4] as models of distributed problem solving. CD grammar system with context-free productions can be viewed as a generalization of context-free grammars in which the set of rules is divided into parts which are called components of the system. These components work on a common sentential form in turns according to some cooperation protocol, which determines when a component is allowed to start and to stop rewriting the sentential form. For example, in the so-called t-mode of derivation a component, once started, has to remain active as long as possible, that is, until none of its productions can be applied to the current sentential form. In what follows, we always consider context-free CD grammar systems in the t-mode of derivation without further mentioning.

Context-free CD grammar system working in the t-mode of derivation can be considered as the sequential counterparts of ET0L systems, just having components instead of tables. Analogously, pure CD (pCD) grammar systems, where no distinction between terminal and nonterminal symbols is made, may be viewed as sequential counterparts of T0L systems. For a more detailed discussion see [1].

In this paper we investigate the number of active symbols in pure grammar formalisms, that is, we will study T0L systems and languages as well as their sequential counterparts, namely the pCD grammar systems and languages, within the framework of active symbols.

One reason why such pure grammars and systems are of interest is that there is no distinction between a sentential form and a word in the language generated. Thus, all information about the derivation process is stored in the language. This may be useful for purposes of syntax analysis. Moreover it may help to improve the understanding of the relationship between parallel and sequential rewriting mechanisms.

The paper is organized as follows. Section 2 provides the necessary definitions of the language generating devices under consideration. In section 3 we will define the notion of active symbols and show that various kinds of Lindenmayer systems without interaction as well as pCD grammar systems lead to infinite hierarchies induced by the measure of active symbols. In section 4 we will reprove and partly extend known hierarchies of some basic families of languages defined by Lindenmayer systems and show that these families of languages build the same kind of hierarchies when the number of active symbols is regarded and disregarded. Some closure and non-closure properties for the respective language families are investigated in section 5. In the conclusions we will summarize our results and state some unsolved problems.

2 Definitions and Preliminaries

We assume the reader to be familiar with basic notions in the theory of formal languages. With our notation we mainly follow [6]. In general, we have the following conventions: $\subseteq$ denotes set inclusion, while $\subset$ denotes strict set inclusion. Set difference will be denoted by $\setminus$. The set of positive integers is denoted by $\mathbb{N}$ and the cardinality of a set M is denoted by $\#M$. Let V be some alphabet, that is, a finite and non-empty set; by V^+ we denote the set of all nonempty words over V; if the empty word λ is included, then we use the notation V^*. For a word $x \in V^*$, its length is denoted by $|x|$. For any set $W \subseteq V$, $|x|_W$ is the number of occurrences of letters of W in the word x. Frequently, for singletons $\{a\}$ we simply write a. We consider two families $\mathcal{L}_1$ and $\mathcal{L}_2$ of languages to be equal if they distinguish from each other at most by the empty set, that is, if $\mathcal{L}_1 \setminus \{\emptyset\} = \mathcal{L}_2 \setminus \{\emptyset\}$. The families of regular, context-free and context-sensitive languages are denoted by $\mathcal{L}(\mathrm{REG})$, $\mathcal{L}(\mathrm{CF})$ and $\mathcal{L}(\mathrm{CS})$, respectively.

A *pure CD grammar system* (pCD grammar system for short) of degree n is an

$(n+2)$-tuple $G = (V, P_1, P_2, \ldots, P_n, S)$, where V is some alphabet, $\omega \in V^+$ is the axiom and, for $1 \le i \le n$, $P_i \subseteq V \times V^*$ is a finite set of pure context-free productions, which are called the components of G. A production sets P_i is called λ-free, if there are no productions of the form $A \to \lambda$ in P_i.

The t-derivation step according to component P_i is defined as follows: for $x, y \in V^*$, we write $x \overset{t}{\underset{i}{\Longrightarrow}} y$ if and only if one of the following conditions hold:

(i) there exist strings $x_0, x_1, \ldots, x_k$, $k \ge 0$, such that $x_0 = x$, $x_k = y$, $x_j \underset{i}{\Longrightarrow} x_{j+1}$, $0 \le j \le k-1$, and there is no z such that $y \underset{i}{\Longrightarrow} z$, or

(ii) $y = x$.

Here $x_j \underset{i}{\Longrightarrow} x_{j+1}$ denotes a direct derivation step in which the component P_i is applied, that is, $x_j = z_1 a z_2$, $x_{j+1} = z_1 v z_2$, for some $a \to v \in P_i$ and $z_1, z_2 \in V^*$. The set $\mathrm{SF}(x \overset{t}{\underset{i}{\Longrightarrow}} y)$ of the sentential forms of the t-derivation step according to P_i is (i) the set $\mathrm{SF}(x \overset{t}{\underset{i}{\Longrightarrow}} y) = \{x_0, x_1, \ldots, x_k\}$ or (ii) the set $\mathrm{SF}(x \overset{t}{\underset{i}{\Longrightarrow}} y) = \{x\}$, respectively. Note that the second case (ii) in this definition is not just a special case of (i) with $k = 0$, since there may be derivations $x \overset{*}{\underset{i}{\Longrightarrow}} y$, $y \ne x$, but all these derivations will not terminate.

A t-derivation in a pCD grammar system is a sequence of t-derivations according to arbitrary components of the system: for $x, y \in V^*$, we write $x \overset{t*}{\Longrightarrow} y$ if and only if there are strings $x_0, x_1, \ldots, x_k$, $k \ge 0$, such that $x_0 = x$, $x_k = y$, and $x_j \overset{t}{\underset{i_j}{\Longrightarrow}} x_{j+1}$, for $1 \le i_j \le n$, $0 \le j \le k-1$. The set $\mathrm{SF}(x \overset{t*}{\Longrightarrow} y)$ of its sentential forms is defined to be the union of the sets $\mathrm{SF}(x_j \overset{t}{\underset{i_j}{\Longrightarrow}} x_{j+1})$. The language $L(G)$ generated by a pCD grammar system G is the set of all sentential forms in a t-derivation in G starting with the axiom ω:

$$L(G) = \{\, w \in V^* \mid w \in \mathrm{SF}(\omega \overset{t*}{\Longrightarrow} y) \text{ for some } y \in V^* \,\}.$$

Note that the language consists of all words generated by iterated t-derivation steps and all the intermediate words appearing along these derivations. Sentential forms of derivations where the active component will not terminate remain excluded.

The family of languages generated by pCD grammar systems in the t-mode of derivation is denoted by $\mathcal{L}(\mathrm{pCD})$.

In order to clarify this definition, we repeat an example given in [1], characterizing all pCD languages over a one-letter alphabet.

Example 2.1 *(see [1, Example 2.2]) Every language over a one-letter alphabet in* $\mathcal{L}(\mathrm{pCD})$ *is either of the form* $L_{0-n} = \{a^n, a^{n-1}, a^{n-2}, \ldots, a, \lambda\}$ *or it contains exactly one nonempty word.*

A TF0L system is a triple $G = (\Sigma, H, \Omega)$, where Σ is the alphabet, H is a finite set of finite substitutions from Σ into Σ^*, and Ω is a finite, non-empty subset of Σ^+, called the *set of axioms of* G. A substitution h in H is called a table of G. If H

only contains homomorphisms or non-erasing substitutions, the TF0L system is said to be *deterministic* or *propagating* and is referred to as DTFOL or PTF0L system, respectively. If for some TF0L system $\#H = 1$, then the TF0L system is called an F0L system and if for some TF0L system $\#\Omega = 1$, then the TF0L system is called a T0L system.

If $v \in h(a)$, $a \in \Sigma$, the we say that $a \to v$ is a production in the table h. For x and y in Σ^*, we write $x \underset{h}{\Longrightarrow} y$ for some $h \in H$ if and only if $y \in h(x)$. Hence, subscript h refers to the table which is used. Let $x \overset{*}{\underset{h}{\Longrightarrow}} y$ denote the reflexive and transitive closure of the relation $x \underset{h}{\Longrightarrow} y$.

The language generated by G is defined as

$$L(G) = \{\, w \in \Sigma^* \mid \exists \omega \in \Omega, \omega \underset{h_{i_1}}{\Longrightarrow} w_1 \underset{h_{i_2}}{\Longrightarrow} \cdots \underset{h_{i_m}}{\Longrightarrow} w_m = w \text{ for some } m \geq 0 \text{ and } h_{i_j} \in H \text{ with } 1 \leq j \leq m \,\}.$$

Any combination of the denotations D, P, T and F leads to various classes of languages. In what follows, we will consider all the families of languages $\mathcal{L}(X)$ with $X \in \{\mathrm{P}, \mathrm{D}, \mathrm{PD}, \lambda\}\{\mathrm{T}, \lambda\}\{\mathrm{F}, \lambda\}\{0\mathrm{L}\}$. The set $\{\mathrm{P}, \mathrm{D}, \mathrm{PD}, \lambda\}\{\mathrm{T}, \lambda\}\{\mathrm{F}, \lambda\}\{0\mathrm{L}\}$ will be denoted by $\mathcal{M}$ for better readability.

Figure 1 and Figure 2 show some known hierarchies of these language families (for proofs see [7, 9, 11, 12, 13]). In these figures we write X instead of $\mathcal{L}(X)$ in order to obtain a better appearance of the figures; moreover an arrow denotes strict inclusion of the lower language family in the upper one, and if two families are not connected, then they are incomparable.

In [1], the language family $\mathcal{L}(\mathrm{pCD})$ is located in a part of this hierarchy, proving the strict inclusion $\mathcal{L}(\mathrm{pCD}) \subset \mathcal{L}(\mathrm{CS})$ and showing the incomparability of $\mathcal{L}(\mathrm{pCD})$ with $\mathcal{L}(\mathrm{CF})$ as well as with each $\mathcal{L}(X)$, $X \in \{\mathrm{P,D,PD},\lambda\}\{\mathrm{T},\lambda\}\{0\mathrm{L}\}$.

Moreover, by definition and [9, 12] the hierarchy in Figure 2 holds.

3 Active Symbols as Connected Measure of Syntactical Complexity

First, we provide the formal definition of the number of active symbols for TF0L systems and languages.

Definition 3.1 Let $G = (\Sigma, H, \omega)$ be a TF0L system. We define the number of active symbols in a table $h \in H$ by

$$as(h) = \#\{\, a \mid a \to w \in h \text{ with } a \neq w \,\}.$$

For G we set

$$as(G) = \max\{\, as(h) \mid h \in H \,\},$$

and for a language L in $\mathcal{L}(\mathrm{TF0L})$, we define

$$as_{\mathrm{T0L}}(L) = \min\{\, as(G) \mid G \text{ is a TF0L system and } L = L(G) \,\}.$$

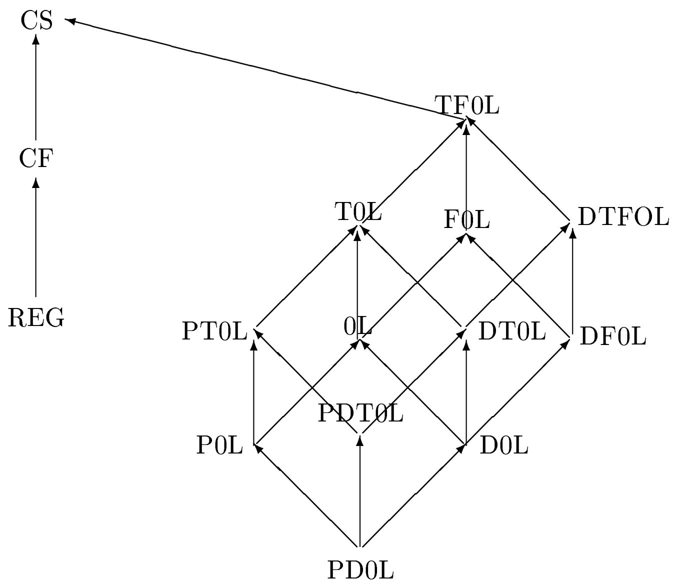

Figure 1: Preliminary hierarchy results (1)

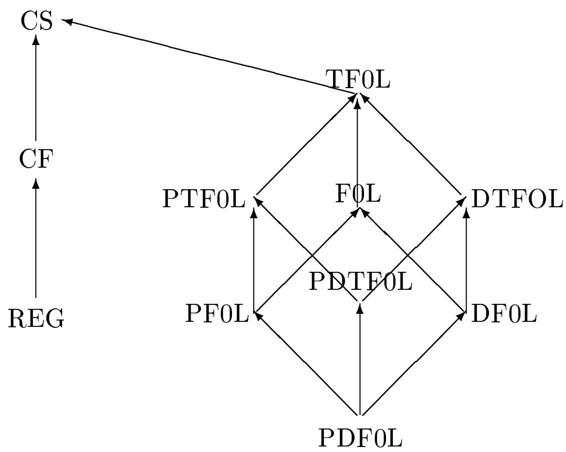

Figure 2: Preliminary hierarchy results (2)

For $X \in \mathcal{M}$ the notion $as_X(L)$ for a language L in $\mathcal{L}(X)$ is defined analogously to Definition 3.1.

In the case of pCD grammar system or CD grammar system, where components are used instead of tables, the number of the active symbols is defined analogously.

For $n \geq 0$ and $X \in \mathcal{M} \cup \{\text{pCD}\}$, let

$$\mathcal{L}(X, n) = \{ L \in \mathcal{L}(X) \mid as_X(L) \leq n \}$$

be the family of languages which can be generated by some X system with at most n active symbols. By definition, $\mathcal{L}(X, n) \subseteq \text{Ł}(X, n+1)$ holds.

Moreover, the following lemmata hold by definition.

Lemma 3.1 *(i) Let $X \in \{\text{P,D,PD},\lambda\}\{T, \lambda\}\{\text{0L}\} \cup \{\text{pCD}\}$. For any language L, $L \in \mathcal{L}(X)$, we have*

$$as_X(L) = 0 \text{ iff } \#(L) \leq 1.$$

(ii) Let $X \in \{\text{P,D,PD},\lambda\}\{T, \lambda\}\{\text{F0L}\}$. For any language L, $L \in \mathcal{L}(X)$ we have

$$as_X(L) = 0 \text{ iff } L \text{ is finite (or empty).}$$

Lemma 3.2 *Let L be a language over a single-letter alphabet. If $L \in \mathcal{L}(X)$ with $X \in \mathcal{M}$, then $L \in \mathcal{L}(X, 1)$.*

Next, we are going to show that the number of active symbols is a connected measure of syntactical complexity with respect to all systems under consideration, that is, this measure induces infinite hierarchies of language families for all the system types considered here. More precisely, the following two theorems hold.

Theorem 3.3 *Let $X \in \mathcal{M}$. For every $n \geq 0$, there exists a language L in $\mathcal{L}(X)$ such that $as_X(L) = n$.*

Proof. The statement has been proved for $n = 0$ by Lemma 3.1. Let $n \geq 1$ and set $\Sigma_n = \{a_1, a_2, \ldots, a_n\}$.

The language

$$L_n = \{ a_1^{2^i} a_2^{2^i} \ldots a_n^{2^i} \mid i \geq 0 \}$$

is generated by the PD0L system with n active symbols

$$G_n = (\Sigma_n, \{ a_i \to a_i^2 \mid 1 \leq i \leq n \}, a_1 a_2 \ldots a_n).$$

Hence, $as_X(L_n) \leq n$, for all $X \in \mathcal{M}$.

Let us assume that there is an TF0L system $G'_n = (\Sigma_n, h_1, h_2, \ldots, h_k, \omega_n)$ with $L(G'_n) = L_n$ and $as(G'_n) < n$. Then for each i, $1 \leq i \leq k$, there is at least one symbol $a_j \in \Sigma_n$ such that $h_i(a_j) = \{a_j\}$. Due to the structure of the words in L_n, any derivation step according to G'_n is of the form $u \underset{h_i}{\Longrightarrow} v$, for some $u = a_1^{2^l} a_2^{2^l} \ldots a_n^{2^l}$, $v = a_1^{2^m} a_2^{2^m} \ldots a_n^{2^m}$. Since a_j is inactive in h_i, $m = l$ has to hold. Therefore, $L(G'_n) = \{\omega_n\}$, and this contradicts $L(G'_n) = L_n$.

In conclusion, $as_X(L_n) = n$, for all $X \in \mathcal{M}$. □

Theorem 3.4 *For every $n \geq 0$, there exists a language L in $\mathcal{L}(\text{pCD})$ such that $as_{\text{pCD}}(L) = n$.*

Proof. For $n = 0$ the statement has been proved by Lemma 3.1. Consider the pCD grammar system of degree 2,

$$G_n = (\Sigma_n \cup \Delta_n, P_{n,1}, P_{n,2}, \omega_n), \text{ where}$$

$\Sigma_n = \{a_1, a_2, \ldots, a_n\}$, $\Delta_n = \{b_1, b_2, \ldots, b_n\}$, $\omega_n = a_1^3 a_2^3 \ldots a_n^3$, and

$$P_{n,1} = \{\, a_i \to b_i^3 \mid 1 \leq i \leq n \,\},$$
$$P_{n,2} = \{\, b_i \to a_i \mid 1 \leq i \leq n \,\}.$$

Starting off with a word $a_1^{3^i} a_2^{3^i} \ldots a_n^{3^i}$, $i \geq 1$, only $P_{n,1}$ is applicable, yielding the word $b_1^{3^{i+1}} b_2^{3^{i+1}} \ldots b_n^{3^{i+1}}$ in the t-mode of derivation.

Since the a_i's are sequentially replaced in arbitrary order, all the sentential forms in $K_{n,i} = \{u_1 u_2 \ldots u_n \mid u_j \in \{a_j, b_j^3\}^{3^i}\}$ can be obtained during all possible t-mode derivation steps starting off with $a_1^{3^i} a_2^{3^i} \ldots a_n^{3^i}$.

Next, only $P_{n,2}$ can be applied to $b_1^{3^{i+1}} b_2^{3^{i+1}} \ldots b_n^{3^{i+1}}$, where the intermediate sentential forms of all possible t-mode derivation steps build the set

$$M_{n,i} = \{\, v_1 v_2 \ldots v_n \mid v_j \in \{a_j, b_j\,\}^{3^{i+1}} \}.$$

Hence, $L_n = L_t(G_n) = \bigcup\limits_{i \geq 0} (K_{n,i} \cup M_{n,i})$.

Since G_n has n active symbols, we have $as_{\text{pCD}}(L_n) \leq n$.

Let $G'_n = (\Sigma_n \cup \Delta_n, P_{n,1}, P_{n,2}, \ldots, P_{n,k}, \omega'_n)$ be a pCD grammar system generating L_n in t-mode of derivation with $as(G'_n) < n$. Since for any two words w and w' of L with $|w| \neq |w'|$, we have $||w| - |w'|| \geq 2$, G' is λ-free.

Thus, the shortest word in L is the axiom, that is, $\omega'_n = a_1^3 a_2^3 \ldots a_n^3$. Let $P_{n,l}$ be a component which can successfully be applied to the axiom, that is, there is some u, $u \neq \omega'_n$, such that a derivation $\omega'_n \underset{n,l}{\overset{t}{\Longrightarrow}} u$ is possible according to G'_n.

If $a_i \to y \in P_{n,l}$, for some $1 \leq i \leq n$ and $y \in (\Sigma_n \cup \Delta_n)^*$, then this production can be applied to each occurrence of a_i in ω'_n. Therefore, $y \in \{a_i, b_i^3\}$ has to hold. Analogously, if $b_i \to y \in P_{n,l}$, then $y \in \{a_i, b_i\}$ has to hold. Since $as(P_{n,l}) < n$, there is at least one j, $1 \leq j \leq n$, such that a_j is not active, i.e., there is no production replacing a_j in $P_{n,l}$. Note that $\omega'_n \underset{n,l}{\overset{t}{\Longrightarrow}} u$ is a t-mode derivation step, where the presence of the production $a_j \to a_j$ would be blocking this derivation. In conclusion, $u = u_1 u_2 \ldots u_n$ with $u_i = \{a_i, b_i^3\}^3$, for $1 \leq i \leq n$, where $u_j = a_j^3$ holds.

Therefore, the word $x = a_1^2 b_1^3 a_2^2 b_2^3 \ldots a_n^2 b_n^3$ does not appear as intermediate sentential form during $\omega'_n \underset{n,l}{\overset{t}{\Longrightarrow}} u$.

On the other hand, because of $u \neq \omega'_n$, there is at least one j, $1 \leq j \leq n$, such that $u_j = b_j^9$.

Therefore the word x cannot be obtained during further derivations, since G'_n is λ-free.

Hence, $x \notin L_t(G'_n)$, contradicting $L_t(G'_n) = L_n$. $\square$

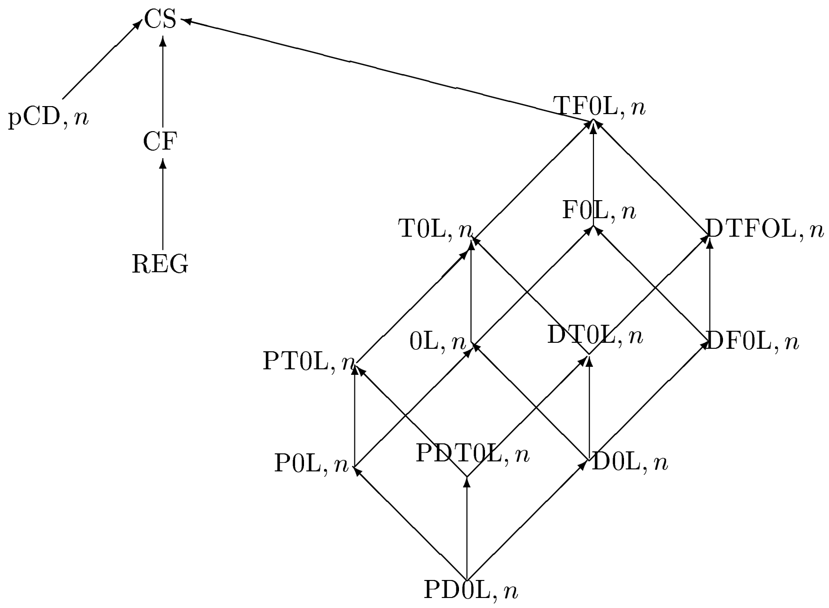

Figure 3: Hierarchy of language families with bounded number of active symbols (1)

4 Hierarchies Induced by a Bounded Number of Active Symbols

In this section we will extend the known hierarchies presented in section 2 involving the family $\mathcal{L}(\mathrm{pCD})$. We compare the generative power of the mechanisms when the number of active symbols is taken into consideration. We are going to prove that the same hierarchical relationships are obtained both when regarding and disregarding this syntactical measure. More precisely Theorems 4.1 and 4.2 hold.

In Figure 3 and Figure 4 which are given below we write X, n instead of $\mathcal{L}(X, n)$, for $X \in \mathcal{M} \cup \{\mathrm{pCD}\}$ (or we write REG, CF, CS instead of $\mathcal{L}(\mathrm{REG})$, $\mathcal{L}(\mathrm{CF})$, $\mathcal{L}(\mathrm{CS})$, respectively) in order to obtain a better appearance of the figures; moreover an arrow denotes strict inclusion of the lower language family in the upper one, and if two families are not connected, then they are incomparable.

Theorem 4.1 *For any integer $n \geq 1$, the hierarchy presented in Figure 3 holds.*

Proof. Let $n \geq 1$. The strict inclusion $\mathcal{L}(\mathrm{pCD}, n) \subset$ Ł(CS) follows from the fact $\mathcal{L}(\mathrm{pCD}) \subset \mathcal{L}(\mathrm{ET0L})$ which has been shown in [1, Theorem 3.6]. The reader may readily verify that it is sufficient to show the following 10 facts.
i) Since every finite language is in $\mathcal{L}(\mathrm{DF0L}, 0)$, the language

$$L_1 = \{a^2, ab, ba, b^2, ab^2, b^2a, b^4\}$$

is contained in $\mathcal{L}(\mathrm{DF0L}, n)$, but L_1 is not a T0L language (see [1, Theorem 3.5]). Hence, $L_1 \in \mathcal{L}(\mathrm{DF0L}, n) \setminus \mathcal{L}(\mathrm{T0L}, n)$ holds.
ii) Consider the P0L language

$$L_2 = \{a\}^+$$

which can be generated by a P0L system with one active symbol, but L_2 is not a DTF0L language (see [12]). Hence, $L_2 \in \mathcal{L}(\mathrm{P0L}, n) \setminus \mathcal{L}(\mathrm{DTF0L}, n)$ holds.
iii) The PDT0L language

$$L_3 = \{\,\$w\$w\$w \mid w \in \{a, b\}^*\,\}$$

can be generated by the PDT0L system with one active symbol

$$G = (\{a, b, \$\}, \{\{\$ \to \$a, a \to a, b \to b\}, \{\$ \to \$b, a \to a, b \to b\}\}, \$\$\$)\,,$$

but L_3 is not an E0L language (see [13, Exercise IV.1.2]). Since every F0L language is an E0L language [7], $L_3 \in \mathcal{L}(\mathrm{PDT0L}, n) \setminus \mathcal{L}(\mathrm{F0L}, n)$ holds.
iv) The D0L language

$$L_4 = \{a, ab\}$$

can be generated by the D0L system $(\{a, b\}, \{a \to a, b \to \lambda\}, ab)$ having only one active symbol, but L_4 is not a PT0L language. Assume the contrary, that is, there exists a PT0L system generating L_4. Since the shortest word in L_4 is the axiom, $\omega = a$ has to hold. Then there exists at least one table h with $a \to ab \in h$ and a derivation $a \underset{h}{\Longrightarrow} ab \underset{h}{\Longrightarrow} ab\beta$, where $\beta \neq \lambda$. Therefore $|ab\beta| \geq 3$ has to hold, and $ab\beta \notin L_4$. Hence, $L_4 \in \mathcal{L}(\mathrm{D0L}, n) \setminus \mathcal{L}(\mathrm{PT0L}, n)$ holds.
v) The PD0L language

$$L_5 = \{\, a^{2^i} \mid i \geq 0\,\}$$

can be generated by a PD0L grammar system with one active symbol, but L_5 is not a pCD language due to Example 2.1. Hence, $L_5 \in \mathcal{L}(\mathrm{PD0L}, n) \setminus \mathcal{L}(\mathrm{pCD}, n)$ holds.
vi) Conversely, consider the pCD grammar with one active symbol

$$G = (\{a, b, c, d\}, \{a \to b\}, \{a \to b^2\}, \{a \to a, c \to dc, c \to d\}, da^2c)\,.$$

Since the components $\{a \to b\}$ and $\{a \to b^2\}$ are applicable to the axiom da^2c, the words b^2 and b^4 are obtained by the respective t-mode derivations, such that the set $\{\, cxd \mid x \in \{a^2, ab, ba, b^2, ab^2, b^2a, b^4\}\}$ is a subset of the language $L_6 = L(G)$. A derivation using the component $\{a \to a, c \to dc, c \to d\}$ can not terminate if an a occurs in the sentential form; thus it is applicable only to b^2 and b^4. Therefore, G generates the language

$$\begin{aligned} L_6 \;&=\; \{da^2c, dabc, dbac, db^2c, dab^2c, db^2ac, db^4c\} \\ &\quad\cup \{\, db^2d^ic,\ db^4d^ic,\ db^2d^i,\ db^4d^i \mid i \geq 1\,\} \end{aligned}$$

Hence, $L_6 \in \mathcal{L}(\mathrm{pCD}, 1)$, but L_6 is not a TF0L language.

Suppose there exists a TF0L system $G = (\{a, b, c, d\}, H, \Omega)$ generating L_6. For all $h \in H$, the following properties hold.

(1) If $d \to \delta$ in h, then $|\delta|_{\{a,b,c\}} = 0$ since h can be applied to db^4d^4 and no words in L_6 contain more than two a, more than two c, or more than four b. Furthermore, since all words in L_6 have a prefix da or db, $|\delta| = |\delta|_d \leq 1$, hence $\delta \in \{\lambda, d\}$ has to hold.

(2) If $b \to \beta$ is a rule in h, then $|\beta|_{\{a,c\}} = 0$, since $x\beta^4 y \in h(db^4c)$ for some $x, y \in \{a, b, c, d\}^*$, but there are no words in L_6 containing at least four occurrences of a or four occurrences of c. Analogously, $|\beta|_b \leq 1$ has to hold. If $|\beta|_b = 1$ and $|\beta|_d > 0$, then $x\beta^4 y \in h(db^4c)$ contains more than one subwords db, hence it is not in L_7. Thus, $\beta \in \{b\} \cup \{\, d^i \mid i \geq 0\,\}$ has to hold. By item (1), $h(d) \in \{\lambda, d\}$. Hence, if $\beta \in d^+$, then a word with prefix d^4 is contained in $h(db^4d)$, a contradiction to the structure of the words in L_6. Therefore, $\beta \in \{\lambda, b\}$.

(3) Because of items (1) and (2) we can argue as follows: since $\delta\beta^2\delta \in h(db^2d)$, neither β nor δ can be the empty word. Hence, $b \to b$ and $d \to d$ are the only rules for b and d, respectively.

(4) Let $a \to \alpha$ be a rule in h, then $d\alpha b^2$ is prefix of all words in $h(dab^2c)$. Therefore, $\alpha \in \{\lambda, a, b^2\}$ has to hold. Since h is applicable to db^4c, $h(c)$ must be taken from d^*c or from d^*. Thus, $\alpha = \lambda$ implies $h(da^2c) \subseteq \{c, d\}^+$, a contradiction. On the other hand, $\alpha = b^2$ implies $|x|_b = 3$ for all $x \in h(dabc)$, which is a contradiction, too. Hence, $\alpha = a$.

(5) According to the above arguments, all words $h(da^2c)$ have the prefix da^2. Hence, $c \to c$ is the only rule for c in h.

In conclusion, each table h in H is the identity, and $L(G)$ is finite. This contradicts the assumption $L(G) = L_6$.

In conclusion, $L_6 \in \mathcal{L}(\text{pCD}, n) \setminus \mathcal{L}(\text{TF0L}, n)$ holds.

vii) The regular language

$$L_7 = \{\, a^i b \mid i \geq 1 \,\} \cup \{\, ab^i \mid i \geq 1 \,\}$$

is not a TF0L language (see [12, Lemma 5]). Hence, $L_7 \in \mathcal{L}(\text{REG}) \setminus \mathcal{L}(\text{TF0L}, n)$ holds.

viii) The regular language $L_8 = \{a, a^2\}$ is no pCD language due to Example 2.1, hence $L_8 \in \mathcal{L}(\text{REG}) \setminus \mathcal{L}(\text{pCD}, n)$.

ix) The language $L_5 = \{\, a^{2^i} \mid i \geq 0 \,\} \in \mathcal{L}(\text{PD0L}, 1)$ is not a CF language. Hence $L_5 \in \mathcal{L}(\text{PD0L}, n) \setminus \mathcal{L}(\text{CF})$ holds.

x) The language $L_t(G_1)$ of the proof of Theorem 3.4 is contained in $\mathcal{L}(\text{pCD}, 1)$. Assume that it is context-free. Then, its intersection with the regular set $\{a\}^+$ is context-free as well, but $\{\, a^{3^i} \mid i \geq 0 \,\}$ is not context-free, a contradiction. In conclusion, $L_t(G_1) \in \mathcal{L}(\text{pCD}, 1) \setminus \mathcal{L}(\text{CF})$. □

Theorem 4.2 *For any integer $n \geq 3$, the hierarchy presented in Figure 4 holds.*

Proof. The statement follows from Theorem 4.1 and the following facts. Let $n \geq 3$.

i) Consider the DF0L language

$$L_9 = \{\, bc(ab^2)^{2^i} \mid i \geq 0 \,\}$$

which can be generated by the DF0L system with 3 active symbols

$$G = (\{a, b, c\}, \{a \to ab^2ab^2, b \to \lambda, c \to bc\}, bcab^2)\,,$$

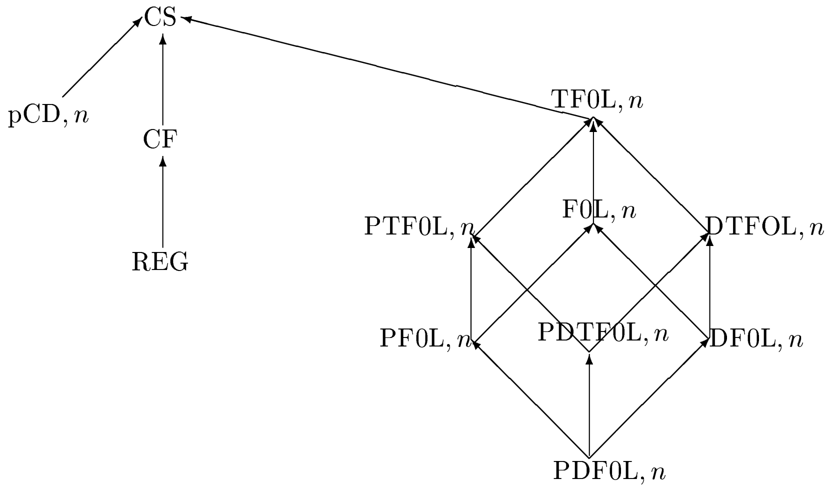

Figure 4: Hierarchy of language families with bounded number of active symbols (2)

but L_9 is not a PTF0L language (see [9, Theorem 12.1]).

Hence, $L_9 \in \mathcal{L}(\mathrm{DF0L}, n) \setminus \mathcal{L}(\mathrm{PTF0L}, n)$ holds.

ii) The language $L_2 = \{a\}^+ \in \mathcal{L}(\mathrm{PF0L}, 1)$ is not a DTFOL language (see [12]). Hence, $L_2 \in \mathcal{L}(\mathrm{PF0L}, n) \setminus \mathcal{L}(\mathrm{DTFOL}, n)$ holds.

iii) The language $L_3 = \{\ \$w\$w\$w \mid w \in \{a, b\}^*\ \} \in \mathcal{L}(\mathrm{PDTF0L}, 1)$ is not an F0L language (see [13, Exercise IV.1.2]). Hence, $L_3 \in \mathcal{L}(\mathrm{PDTF0L}, n) \setminus \mathcal{L}(\mathrm{F0L}, n)$ holds.

□

5 Closure Properties

In this section, we investigate whether or not certain operations on languages lead out of the families of languages generated by some pure system with a bounded number of active symbols. For the definitions of the considered operations we refer to any standard text book on formal languages or to [6].

Due to Lemma 3.1, we immediately obtain the following result.

Theorem 5.1 *(i) For $X \in$ {P,D,PD,λ}{T,λ}{0L} ∪ {pCD}, each of the families of languages $\mathcal{L}(X, 0)$ is closed with respect to product, intersection, homomorphism, and intersection with a regular set, but it is not closed with respect to union, complement, Kleene star, Kleene plus, and inverse homomorphism.*

(ii) For $X \in$ {P,D,PD,λ}{T,λ}{F0L}, each of the families of languages $\mathcal{L}(X, 0)$ is closed with respect to union, intersection, product, homomorphism, and intersection with a regular set, but it is not closed with respect to complement, Kleene star, Kleene plus, inverse homomorphism.

Apart from this, we obtain only non-closure properties.

Theorem 5.2 *For any integer $n \geq 1$, the family of languages $\mathcal{L}(\mathrm{pCD}, n)$ is not closed with respect to union, intersection, complement, product, Kleene star, Kleene plus, intersection with a regular set, homomorphism, inverse homomorphism.*

Proof. Let $n \geq 1$.

Union.
Consider the languages $\{a\}$ and $\{a^3\}$ which are both in $\mathcal{L}(\mathrm{pCD}, n)$, but their union $\{a, a^3\}$ is not in $\mathcal{L}(\mathrm{pCD})$, thus it is not in $\mathcal{L}(\mathrm{pCD}, n)$.

Complement. The language $\{a\}^* \setminus \{a\}$ is not contained in $\mathcal{L}(\mathrm{pCD})$, hence not in $\mathcal{L}(\mathrm{pCD}, n)$.

Kleene star and Kleene plus.
The languages $\{a\}^*$ and $\{a\}^+$ are not contained in $\mathcal{L}(\mathrm{pCD}, n)$.

Intersection with a regular set.
The pCD grammar system $G = (\{a, b\}, \{a \to a^2, a \to b\}, a)$ generates the language $\{a, b\}^+$. The intersection $\{a, b\}^+ \cap \{a\}^+ = \{a\}^+$ is not in $\mathcal{L}(\mathrm{pCD}, n)$.

Intersection.
The languages $\{a, c\}^+$ and $\{a, b\}^+$ are both in $\mathcal{L}(\mathrm{pCD}, 1)$, but their intersection $\{a, c\}^+ \cap \{a, b\}^+ = \{a\}^+$ is not contained in $\mathcal{L}(\mathrm{pCD}, n)$.

Product.
Consider the languages $\{\lambda, a\}$ and $\{a^2\}$ which are both in $\mathcal{L}(\mathrm{pCD}, 1)$ (see Example 2.1). Their product $\{\lambda, a\}\{a^2\} = \{a^2, a^3\}$ is not in $\mathcal{L}(\mathrm{pCD})$, hence not in $\mathcal{L}(\mathrm{pCD}, n)$.

Homomorphism.
The pCD grammar system $G = (\{a, b\}, \{a \to b\}, a)$, generates the language $\{a, b\}$, hence it is contained in $\mathcal{L}(\mathrm{pCD}, 1)$.

Let $\phi : \{a, b\}^* \to \{a, b\}^*$ be a homomorphism such that $\phi(a) = a$ and $\phi(b) = a^3$, then $\phi\{a, b\} = \{a, a^3\}$, which is not contained in $\mathcal{L}(\mathrm{pCD}, n)$ (see Example 2.1).

Inverse homomorphism.
Consider the language $\{\lambda, a\} \in \mathcal{L}(\mathrm{pCD}, 1)$ and let $\phi : \{a\}^* \to \{a\}^*$ a homomorphism such that $\phi(a) = \lambda$. Then $h^{-1}(L) = \{a\}^*$ which is not in $\mathcal{L}(\mathrm{pCD}, n)$. □

Theorem 5.3 *(i) For any integer $n \geq 1$, the following statements hold.*

- *(a) Each of the families $\mathcal{L}(X, n)$ with $X \in \mathcal{M}$ is not closed with respect to union, complement, product, homomorphism, and intersection with a regular set.*
- *(b) Each of the families $\mathcal{L}(X, n)$ with $X \in \mathcal{M} \setminus \{\mathrm{PD0L}, \mathrm{D0L}\}$ is not closed with respect to intersection.*
- *(c) Each of the families $\mathcal{L}(X, n)$ with $X \in \mathcal{M}$ is not closed with respect to Kleene star and Kleene plus.*
- *(d) Each of the families of languages $\mathcal{L}(X, n)$ with $X \in \mathcal{M} \setminus \{\mathrm{PF0L}, \mathrm{F0L}, \mathrm{PTF0L}, \mathrm{TF0L}\}$ is not closed with respect to inverse homomorphisms.*

(ii) For any integer $n \geq 2$, the following statements hold.

(a) *The families* $\mathcal{L}(\mathrm{D0L})$ *and* $\mathcal{L}(\mathrm{PD0L})$ *are not closed with respect to intersection.*

(b) *Each of the families* $\mathcal{L}(X,n)$ *with* $X \in \{\mathrm{P},\lambda\}\{\mathrm{T},\lambda\}\{\mathrm{F},\lambda\}\{\mathrm{0L}\}$ *is not closed with respect to Kleene star and Kleene plus.*

(c) *Each of the families* $\mathcal{L}(X,n)$ *with* $X \in \{\mathrm{PF0L}, \mathrm{F0L}, \mathrm{PTF0L}, \mathrm{TF0L}\}$ *is not closed with respect to inverse homomorphisms.*

Proof.
(i) Let $n \geq 1$.

Union.
Consider the languages $\{a^3\}$ and $\{\, a^{2^i} \mid i \geq 1 \,\}$ which are both in $\mathcal{L}(\mathrm{PD0L}, n)$, but their union $\{a^3\} \cup \{\, a^{2^i} \mid i \geq 1 \,\}$ is not in $\mathcal{L}(\mathrm{TF0L})$ (see [12, Theorem 2]), thus it is not in $\mathcal{L}(\mathrm{TF0L}, n)$.

Complement.
The language $\{\, a^{2^i} \mid i \geq 1 \,\}$ is in $\mathcal{L}(\mathrm{PD0L}, n)$, but $\{a\}^* \setminus \{\, a^{2^i} \mid i \geq 1 \,\}$ is not in $\mathcal{L}(\mathrm{TF0L})$ (see [12, Theorem 2]) and therefore not contained in $\mathcal{L}(\mathrm{TF0L}, n)$.

Intersection.
Consider the P0L systems $G_1 = (\{a,b,c\}, \{a \to a, b \to b, c \to acb, c \to ab\}, c)$ and $G_2 = (\{a,b,d\}, \{a \to a, b \to b, d \to adb, d \to ab\}, d)$.

Then, $L(G_1) \cap L(G_2) = \{\, a^i b^i \mid i \geq 1 \,\}$.

Assume that this language is generated by some TF0L system G. If $a \to \alpha$ is a rule in some table of G, then $\alpha \in a^*$ has to hold, and for any rule $b \to \beta$, $\beta \in b^*$ has to hold. If there are two productions of these forms in one table with $|\alpha| \neq |\beta|$, then a string of the form $a^j b^k$ with $j \neq k$ can be obtained from ab. In conclusion, G is deterministic. Since $L(G)$ is infinite, $|\alpha| = |\beta| > 0$ in each table of G. But this leads to an exponential progression, a contradiction to our assumption. Hence, $\{\, a^i b^i \mid i \geq 1 \,\} \notin \mathcal{L}(\mathrm{TF0L})$. This shows the non-closure for all propagating classes.

Clearly, $L(G_1)$ and $L(G_2)$ are in $\mathcal{L}(\mathrm{DT0L}, 1)$, as well (just divide the productions appropriately into two tables). Thus, the non-closure for all tabled classes is proved.

Intersection with a regular set.
Use the language $L(G_1)$ of the proof for the non-closure under intersection (see *iii)*), and the regular set a^*b^*. This proves the non-closure for all propagating or tabled classes.

In order to complete the argument, consider the language $\{\, a^{2^i} \mid i \geq 0 \,\} \in \mathcal{L}(\mathrm{PD0L}, 1)$, again. Its intersection with the regular set $\{a^2, a^4\}$ is no T0L language, see [11, Theorem 2].

Product.
Consider the languages $\{a\}$ and $\{\, a^{2^i} \mid i \geq 0 \,\}$. Their product $\{a\}\{\, a^{2^i} \mid i \geq 0 \,\}$ is not in $\mathcal{L}(\mathrm{TF0L})$ (see [12, Theorem 2]) and hence not in $\mathcal{L}(\mathrm{TF0L}, n)$.

Kleene star and Kleene plus.
The claim is shown using $\{a\}$ and the fact that neither $\{a\}^*$ nor $\{a\}^+$ is a DTFOL language [12].

Homomorphism.
First, consider the PD0L language $\{a^2, b^4\}$ which is generated by the PD0L system $G = (\{a,b\}, \{a \to b^2, b \to b\}, a^2)$. Hence, the language $\{a^2, b^4\}$ is in $\mathcal{L}(\mathrm{PD0L}, 1)$.

Let $\phi : \{a,b\}^* \to \{a,b\}^*$ be a homomorphism such that $\phi(a) = \phi(b) = a$, then $\phi(\{a^2,b^4\}) = \{a^2,a^4\}$ which is not a T0L language (see [11, Theorem 2]) and therefore also not in $\mathcal{L}(\text{T0L},n)$.

Now, consider the PDF0L system $G = (\{a,b\}\{a \to a^2, b \to b\}, \{a,b^3\})$ that is generating the language $\{b^3\} \cup \{\, a^{2^i} \mid i \geq 1\,\}$ which is therefore in $\mathcal{L}(\text{PDF0L},n)$. Let $\phi : \{a,b\}^* \to \{a,b\}^*$ be a homomorphism such that $\phi(a) = \phi(b) = a$, then

$$\phi(\{b^3\} \cup \{\, a^{2^i} \mid i \geq 1\,\}) = \{a^3\} \cup \{\, a^{2^i} \mid i \geq 1\,\},$$

which is not a TF0L language (see [12, Theorem 2]) and therefore also not contained in $\mathcal{L}(\text{TF0L},n)$.

Inverse homomorphism.
At first, $\{a^2,b^3\} = \phi^{-1}(\{c^6\})$, if ϕ is the homomorphism $\phi : \{a,b\}^* \to \{c\}^*$ with $\phi(a) = c^3$ and $\phi(b) = c^2$. On the other hand, it is proved in [11, Theorem 2] that $\{a^2,b^3\}$ is no T0L language.

Furthermore, $\{a\}^* = h^{-1}(\{a,\lambda\})$, if h is the homomorphism defined by $h(a) = \lambda$, but $\{a\}^* \notin \mathcal{L}(\text{DTFOL})$.

(ii) Now, let $n \geq 2$.

Intersection. The D0L systems $G_1 = (\{a,b,c\}, \{a \to a^2b, b \to a^2b\}, a)$ and $G_2 = (\{a,b\}, \{a \to \lambda, b \to a\}, a^2b)$ generate the languages $\{a\} \cup \{\, (a^2b)^{3^i} \mid i \geq 0\,\}$ and $\{\lambda, a, a^2b\}$, respectively, each with two active symbols. The intersection $L(G_1) \cap L(G_2) = \{a, a^2b\}$ is known to be no D0L language (see [11, Theorem 2]).

The PD0L systems $G_1 = (\{a,b,c\}, \{a \to ab, b \to c, c \to c\}, a)$ and $G_2 = (\{a,b,d\}, \{a \to ab, b \to d, d \to d\}, a)$ generate the languages $\{\, a, abc^n \mid n \geq 0\,\}$ and $\{\, a, abd^n \mid n \geq 0\,\}$, respectively. The intersection $L(G_1) \cap L(G_2) = \{a, ab\}$ is no PT0L language (see iv) in the proof of Theorem 4.1).

Kleene star and Kleene plus.
The language $\{\, a^{2^i}b^{2^i} \mid i \geq 0\,\}$ is in $\mathcal{L}(\text{PD0L},n)$, but $\{\, a^{2^i}b^{2^i} \mid i \geq 0\,\}^*$ is not in $\mathcal{L}(\text{TF0L})$ (see [12, Theorem 2]) and hence not in $\mathcal{L}(\text{TF0L},n)$. The non-closure under Kleene plus is shown analogously.

Inverse homomorphism. Consider the PF0L system

$$G = (\{a,b,c,d\}, \{a \to a, b \to b, c \to ac, c \to b, d \to bd, d \to a\}, \{ac, ad\}).$$

Let $\phi : \{a,b,c,d\}^* \to \{a,b,c,d\}^*$ be the homomorphism defined by $\phi(a) = a$, $\phi(b) = b$, $\phi(c) = \phi(d) = cd$. Then, we have

$$\phi^{-1}(L(G)) = \{\, a^ib \mid i \geq 1\,\} \cup \{\, ab^i \mid i \geq 1\,\},$$

which is no TF0L language [12, Lemma 5]. □

6 Concluding Remarks

The number of (statically measured) active symbols, which has extensively been investigated for ET0L , EDT0L , and CD grammar systems, is considered in the present paper for the pure versions of these systems and several variants thereof. It is shown that this measure of syntactical complexity is connected for all cases under consideration. Next, the known hierarchies of the underlying families of languages are proved to be valid also if one bounds the number of active symbols by any constant $n \geq 1$ (or $n \geq 3$ in a few cases). One should remark that the condition $n \geq 3$ is needed only in one of the numerous constructions (which is only used for proving certain incomparability results, not affecting the strictness of any inclusion). Therefore, in principle a better result than that stated in Theorem 4.2 has been shown. Finally, some closure and, mainly, non-closure properties of the considered families of languages with bounded number of active symbols are proved, where except from a few marginal cases, optimal results could be achieved.

The problems which have been left open are:

1. Is there a language in $\mathcal{L}(\mathrm{DF0L}, 1) \setminus \mathcal{L}(\mathrm{PTF0L}, 2)$?

2. Which of $\mathcal{L}(\mathrm{D0L}, 1)$ and $\mathcal{L}(\mathrm{PD0L}, 1)$ is closed under intersection?

3. Is $\mathcal{L}(X, 1)$ for nondeterministic types X of systems closed under Kleene star or plus?

4. Is $\mathcal{L}(X, 1)$ with $X \in \{\mathrm{PF0L}, \mathrm{F0L}, \mathrm{PTF0L}, \mathrm{TF0L}\}$ closed under inverse homomorphism?

In [1] also deterministic pCD grammar systems and languages have been considered. Due to the proof of Theorem 3.4 the number of active symbols induces an infinite hierarchy also in this deterministic case, more precisely

> *For every $n \geq 0$, there exists a language L generated by a deterministic* pCD *grammar system such that $as_{\mathrm{pCD}}(L) = n$.*

Unfortunately, some constructions used in the subsequent proofs make use of nondeterministic pCD grammar systems. This leads to a lot of open problems.

Moreover, the computability of the number of active symbols for given pure languages is of interest. Finally, also the dynamic interpretation of the number of active symbols could be treated in the framework of pure systems.

References

[1] H. Bordihn, E. Csuhaj-Varjú, and J. Dassow. CD grammars systems versus L systems. In G. Păun and A. Salomaa, editors, *Grammatical Models of Multi-Agent Systems*, pages 18–32. Gordon and Breach Publishers, 1999.

[2] H. Bordihn and M. Holzer. On the number of active symbols in L and CD grammar systems. *Journal of Automata, Languages and Combinatorics*, 6:411–426, 2001.

[3] E. Csuhaj-Varjú and J. Dassow. On cooperating/distributed grammar systems. *Journal of Information Processing and Cybernetics (EIK)*, 26:49–63, 1990.

[4] E. Csuhaj-Varjú, J. Dassow, J. Kelemen, and Gh. Păun. *Grammar Systems. A Grammatical Approach to Distribution and Cooperation.* Gordon and Breach, London, 1994.

[5] J. Dassow. $L(ETOL_{[6]}) = L(ET0L)$. Unpublished note, 1981.

[6] J. Dassow and Gh. Păun. *Regulated Rewriting in Formal Language Theory.* Akademie-Verlag, Berlin, 1989.

[7] L. Kari, G. Rozenberg, and A. Salomaa. L systems. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages, Vol. 1*, pages 253–328. Springer-Verlag, 1997.

[8] H.C.M. Kleijn and G. Rozenberg. A study in parallel rewriting systems. *Information and Control*, 44:134–163, 1980.

[9] M. Nielson, G. Rozenberg, A. Salomaa, and S. Skyum. Nonterminals, homomorphisms and codings in different variations of 0L systems, ii. nondeterministic systems. *Acta Informatica*, 3:357–364, 1974.

[10] Gh. Păun and J. Dassow. $L(ETOL_{[5]}) = L(ET0L)$. *Elektronische Informationsverarbeitung und Kybernetik (J. Inf. Process. Cybern. EIK)*, 18:345–348, 1982.

[11] G. Rozenberg. T0L systems and languages. *Information and Control*, 23:357–381, 1973.

[12] G. Rozenberg and K. P. Lee. Development systems with finite axiom sets, part 1: Systems without interactions. *Intern. J. Computer Math.*, 4:43–68, 1974.

[13] G. Rozenberg and A. Salomaa. *The Mathematical Theory of* L *Systems.* Academic Press, New York, 1980.

[14] T. Yokomori, D. Wood, and K.-J. Lange. A three-restricted normal form for ET0L languages. *Information Processing Letters*, 14:97–100, 1982.

[15] T. Yokomori, D. Wood, and K.-J. Lange. Erratum at a three-restricted normal form for ET0L languages. *Information Processing Letters*, 21:53, 1985.