stichting
mathematisch
centrum

AFDELING INFORMATICA IW 40/75 SEPTEMBER

P.M.B. VITÁNYI

DIGRAPHS ASSOCIATED WITH DOL SYSTEMS

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 68A30, 68A25, 94A30, 92A05

ACM -Computing Reviews- categories: 5.22, 5.23

Digraphs associated with DOL Systems *)

by

P.M.B. Vitányi

KEY WORDS & PHRASES: *Formal language theory, Lindenmayer systems, DOL systems, structural complexity, digraphs, growth functions, locally catenative property, regularity and context freeness, decision methods, developmental biology, cell lineage, cell differentiation.*

---

*) This paper is not for review; it is meant for publication elsewhere.

DIGRAPHS ASSOCIATED WITH DOL SYSTEMS$^{*)}$

Paul M.B. Vitányi
Mathematisch Centrum
2e Boerhaavestraat 49
Amsterdam, The Netherlands

ABSTRACT

Directed graphs associated with homomorphisms are used as a tool for reasoning about structure of derivations in DOL systems, thus relating local and global properties of the derived string sequences. Associated digraphs can be used to establish a partial ordering on structural complexity classes of (semi) DOL systems. Applications of the use of associated digraphs are given in the theory of growth functions, for DOL systems possessing the locally catenative property, and for questions on the regularity or context-free-ness of the produced languages. With respect to developmental biology it is argued that associated digraphs contribute to the theoretical framework in that field and may constitute a conceptual help to those engaged in problems of cell lineage, cell differentiation and cell potential.

1. INTRODUCTION

We shall be concerned with relations between homomorphisms on finitely generated free monoids and certain associated digraphs. The interest in this area is mainly due to the parallel rewriting systems introduced by LINDENMAYER [6] to model the growth and development of filamentous organisms. These Lindenmayer systems have been subject of investigation in a large number of papers from the formal language theory point of view and a (not so large) number of papers from the viewpoint of developmental biology, see e.g. [5, 10]. One of the most thoroughly investigated classes of L systems are the so called DOL systems (deterministic context-free Lindenmayer systems) which have been especially fruitful in yielding mathematically and biologically interesting theories such as those about growth functions [14, 12, 16] and locally catenative sequences [9]. The purpose of the present contribution is to make explicit the structural approach the author has used in a number of papers on DOL systems [15, 16, 17], which approach seems to have an obvious interpretation in terms of cell lineage, cell differentiation and cell potential. Related to this is the use of dependence graphs by ROZENBERG & LINDENMAYER in [9].

A DOL system is a string rewriting system, where each letter of a string symbolizes the presence in that position of a cell of a certain type / state and the whole string symbolizes a filament of cells. Time is assumed to be discrete and, between two consecutive moments of time, say between t and t + 1, each letter of the string is rewritten simultaneously as a string (which may be empty). The string at time t + 1 consists of the concatenation of the strings resulting from the rewriting of the individual letters of the string existing at time t. In this way we obtain a sequence of strings symbolizing the developmental history of the modelled filamentous organism.

Although objections may be raised against the adequacy of L systems to model phenomena occurring in actual biological development, and against the usefulness of sophisticated mathematical theorems in developmental biology, it seems to the author that developmental biologists might find conceptual help from the more superficial aspects of the theoretical framework embodied by Lindenmayer's model. Some of the mathematical theorems might be useful to confirm or refute biological hypotheses - but only after careful scrutiny as to whether the assumptions under which the theorem holds are reflected entirely by the biological reality in the case under consideration. For a more extensive discussion along these lines see

*) The work reported in this paper was supported by the Mathematical Center under #IW 40/75.

DOUCET [1,2].

As a reference frame to think about cell lineage, cell differentiation, cell potential and the like, the associated digraphs introduced in section 2 may be useful to developmental biologists. In this respect also the theorems in sections 3-5 about growth functions, locally catenative systems, etc., may prove worthwhile. With this idea in mind we digress in section 6 from mathematics into possible biological interpretations.

In section 2 four associated digraphs are constructed from a (semi) DOL system which form in increasing levels of abstraction a representation of the structure of derivations between letters in the system. These are the associated digraph, the condensed associated digraph, the recursive structure and the unlabelled recursive structure, respectively. In section 3 we investigate the relations between types of growth functions and types of recursive structures. Here we interpret [3] and use results of [16]. Section 4 contains necessary conditions on the recursive structure of a DOL system in order that it can have the locally catenative property, and an order $e^{\sqrt{n \log n}}$ "worst case" lower bound on the minimal depth of a locally catenative formula for a locally catenative DOL system with an n letter alphabet. It is shown that the sequence and language equivalence problems for locally catenative DOL systems are decidable. Furthermore it is shown that deciding whether a DOL system has the locally catenative property is equivalent to deciding whether the monoid generated by the language of a DOL system is finitely generated. Section 5 consists of the application of associated digraphs on recent results by SALOMAA [13] and gives necessary (and sometimes sufficient) conditions on the recursive structure of a DOL system for the produced languages to be regular or context-free. In section 6 we return to developmental biology once more and try to point out possibly relevant aspects of the developed notions and results.

The attitude throughout the paper is that of local versus global properties: local in the sense of the defining parameters of the DOL systems (alphabet, homomorphism, initial string), global in the sense of the overall properties of the produced sequence of strings.

2. TERMINOLOGY AND DEFINITIONS

We shall use the terminology of formal language theory as in e.g. [11] and that of graph theory as in e.g. [4]; "string" and "word" are used interchangeably, $\#Z$ denotes the cardinality of a set Z, $\lg(z)$ denotes the length of a word z, and λ denotes the empty word, i.e. $\lg(\lambda) = 0$.

A *semi DOL system* $S = \langle W,\delta\rangle$ consists of a finite nonempty *alphabet* W and a *homomorphism* δ from W^* into W^*. A *DOL system* $G = \langle W,\delta,w\rangle$ consists of a semi DOL system $\langle W,\delta\rangle$ and an *initial string* $w \in W^+$. The composition of i copies of δ is defined inductively by $\delta^0(v) = v$ and $\delta^i(v) = \delta(\delta^{i-1}(v))$ for $i > 0$ and $v \in W^*$. The *string sequence* produced by G is defined by $S(G) = w,\delta(w),\delta^2(w),\ldots$; the *language* produced by G is $L(G) = \{\delta^i(w) \mid i \geq 0\}$; the *growth function* $f_G : \mathbb{N} \to \mathbb{N}$ associated with G is defined by $f_G(t) = \lg(\delta^t(w))$. According to [12], f_G is a generalized exponential polynomial $f_G(t) = \Sigma_{i=1}^{r} p_i(t)c_i^t$ where the c_i's are distinct (and possibly complex) constants and the p_i's are polynomials in t (with possibly complex coefficients) such that $\Sigma_{i=1}^{r}(\text{degree } p_i+1) \leq \#W$.

A DOL system $G = \langle W,\delta,w\rangle$ has the *locally catenative* property [9] if there exist fixed positive integers $n_0,i_1,i_2,\ldots,i_k$ such that $\delta^n(w) = \delta^{n-i_1}(w)\delta^{n-i_2}(w)\ldots\delta^{n-i_k}(w)$ for all $n \geq n_0$. n_0 is called the *cut* and $\max\{i_1,i_2,\ldots,i_k\}$ the *depth* of the *locally catenative formula* $(n_0,i_1,i_2,\ldots,i_k)$.

Classify the letters of a semi DOL system $S = \langle W,\delta\rangle$ as follows. A letter $a \in W$ is called *mortal* ($a\in M_S$) if $\delta^i(a) = \lambda$ for some i; *vital* ($a\in V_S$) if $a \notin M_S$; *recursive* ($a\in R_S$) if $\delta^i(a) = v_1av_2$ for some $i > 0$ and $v_1,v_2 \in W^*$; *monorecursive* ($a\in MR_S$) if $\delta^i(a) = v_1av_2$ for some $i > 0$ and $v_1,v_2 \in M_S^*$; *expanding* ($a\in E_S$) if

$\delta^i(a) = v_1 a v_2 a v_3$ for some i and $v_1,v_2,v_3 \in W^*$; *accessible* from $v \in W^+$ ($a \in U(v)$) if $\delta^i(v) = v_1 a v_2$ for some $i > 0$ and $v_1,v_2 \in W^*$. We dispense with the subscripts if S is understood. It is not difficult to see [16] that we can determine the above classes of letters by examining the letters occurring in $\delta^i(b)$, $1 \le i \le \#W$, for all $b \in W$.

The global properties of (sequences of) strings produced by a DOL system such as the "patterns" (characteristic substrings) occurring, or the growth of the length of strings, are essentially due to the recursive letters and the derivational relations between them. E.g. a language like $\{a^{2^n} b^{2^n} c^{3^n} \mid n \ge 0\}$ can only be produced by the DOL system

$$G = \langle \{a,b,c\}, \{\delta(a) = a^2, \delta(b) = b^2, \delta(c) = c^3\}, abc \rangle.$$

Hence from the produced patterns it can be readily deduced that the system has to contain 3 expanding recursive letters with no derivational relations between each other at all. We shall see in the sequel that types of growth functions, locally catenativeness, regularity and context-free-ness depends to a very large extent on the recursive letters and the accessibility between them: properties of recursive letters govern the relation between local properties of DOL systems and global properties of the derived sequences.

We define an equivalence relation on R by $a \sim b$ if $a \in U(b)$ and $b \in U(a)$. $\sim$ induces a partition on R in disjoint equivalence classes and

$$R/\sim \; = \{[a] \mid [a] = \{b \in R \mid b \sim a\}\}.$$

For properties of such equivalence classes see [16]. E.g., $\delta^i(a) \in W^*[a]W^*$ for all $a \in R$ and all $i \ge 0$.

We now construct four digraphs from a semi DOL system $S = \langle W,\delta \rangle$ which form in increasing levels of abstraction a representation of the derivational relations between letters.

I. The *associated digraph* of S (*AD(S)*), called the *dependence graph* in [9], is the labelled digraph AD(S) = (W,A) where W is the set of *points* and A the set of *directed arcs* defined by

$$A = \{(a,b) \mid \delta(a) = v_1 b v_2, \quad a,b \in W \quad v_1,v_2 \in W^*\}.$$

Note that we identify points with their *labels* since in all digraphs we discuss there is a one-to-one correspondence between the set of points and the set of labels. We admit digraphs with *loops*, i.e., a point can be connected to itself by an arc.

A digraph is *strong* if every two points are mutually reachable, i.e., if p,q are two points of the digraph then there is a sequence of arcs $(p_1,p_2),(p_2,p_3),\ldots,(p_{n-1},p_n)$ such that $p_1 = p$ and $p_n = q$. (We consider the graph on a single point without arcs to be a strong digraph.) A *strong component* of a digraph is a maximal strong subgraph. Let $D_1,D_2,\ldots,D_n$ be strong components of a digraph D. The *condensation* $D^{(c)}$ of D has the strong components of D as its points, with an arc from D_i to D_j ($i \ne j$) whenever there is at least one arc in D from a point in D_i to a point in D_j. It follows from the maximality of the strong components that the condensation of a digraph has no cycles.

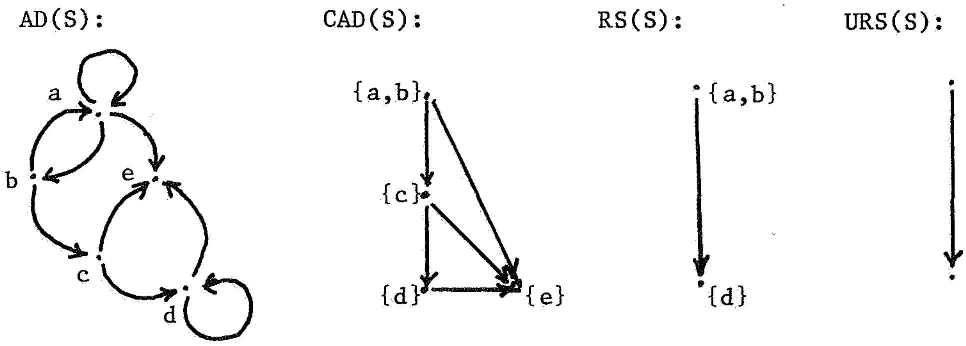
II. The *condensed associated digraph* of S (*CAD(S)*) is the condensation of AD(S). A point in CAD(S) is labelled by the set of letters labelling the points of the corresponding strong component in AD(S).

III. The *recursive structure* of S (*RS(S)*) is obtained from CAD(S) by deleting all points labelled by {a} where a is not a recursive letter. Two points p,q in RS(S) are connected by an arc (p,q) if there is a sequence of arcs $(p_1,p_2),(p_2,p_3),\ldots,(p_{n-1},p_n)$ in CAD(S) such that $p_1 = p$, $p_n = q$ and $p_i \in \{\{a\} \mid a \notin R$ for all i, $1 < i < n\}$.

IV. The *unlabelled recursive structure* of S (*URS(S)*) is obtained from RS(S) by removing the labels.

Example:

Let $S = \langle\{a,b,c,d,e\},\{\delta(a) = abe, \delta(b) = ac, \delta(c) = de, \delta(d) = de, \delta(e) = \lambda\}\rangle$.



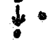
We now tie in the digraph approach with the preceding classification of letters. It is easy to see that $RS(S) = (P_3,A_3)$ is the labelled acyclic digraph such that $P_3 = R/\sim$ and $A_3 \subseteq R/\sim \times R/\sim$ is as defined in III. Similarly, each subset of $W - R$ labelling a point in CAD(S) is a singleton subset of $W - R$ and conversely. A letter $a \in W$ labelling a point in AD(S) with no outgoing arcs is an element of M, etc.

For each unlabelled acyclic digraph D we can find a semi DOL system S such that $URS(S) \cong D$ ("$\cong$" means "is isomorphic with"). Hence the set of all homomorphisms $\delta: W^* \to W^*$, where W is a finite nonempty subset of some infinite alphabet Σ, can be divided into disjoint classes of homomorphisms having isomorphic unlabelled recursive structures. It is natural to assign to a given homomorphism its URS(S) as its complexity (*structural* complexity which should not be confused with *computational* complexity). We define a partial ordering on the thus constructed disjoint complexity classes as a partial ordering according to graph inclusion. It is of interest to see how many different URS's are possible for an alphabet of n letters. If we call the number of unlabelled acyclic digraphs on n points H(n) then this is given by $F(n) = \Sigma_{i=0}^{n} H(n)$. ROBINSON [8] gives a method to compute H(n) for all n; in particular this yields: $F(0) = 1$, $F(1) = 2$, $F(2) = 4$, $F(3) = 10$, $F(4) = 41$, $F(5) = 343$, and $F(6) = 6327$. The partial ordering $\leq$ induced by "being a subgraph of" on the set of unlabelled acyclic digraphs (on i points, $0 \leq i \leq n$) has a 0 element: the empty graph; and a 1 element: the *complete* unlabelled acyclic digraph (on n points), i.e., the unlabelled acyclic digraph with the maximal number of arcs ($\frac{1}{2}n(n-1)$) which is unique up to isomorphism. In a similar way we can define complexity classes of (semi) DOL systems and a partial ordering between them with respect to the levels of abstraction I-III.

A DOL system $G = \langle W,\delta,w\rangle$ is reduced if all leters of W occur in L(G), or equivalently, if the axiom w contains letters from each point which is a maximal element (point without incoming arcs) of $CAD(\langle W,\delta\rangle)$. Considerable attention has been given to the problem which properties are possible for DOL systems with different initial strings and the same semi DOL system $S = \langle W,\delta\rangle$. This problem is reduced to looking at subgraphs of CAD(S) with as maximal elements points labelled by the sets of letters in which occur the letters in the chosen initial string.

We have seen that the set of all (semi) DOL systems is partitioned in disjoint classes having isomorphic characteristic digraphs. We would like to know to what extent this is also the case for the corresponding classes of languages. However, there are DOL systems G_1,G_2 such that $L(G_1) = L(G_2)$ while $URS(G_1) \not\cong URS(G_2)$ as is shown by the example:

$G_1 = \langle\{a,b,c\},\{\delta(a) = a, \delta(b) = ba, \delta(c) = ac\}, bac\rangle$, URS($G_1$) =  .

$G_2 = \langle\{a,b,c\},\{\delta(a) = a, \delta(b) = b, \delta(c) = a^2c\}, bac\rangle$, URS($G_2$) =  .

Yet it is to be expected that DOL systems with different associated digraphs generate different languages. For instance, the previously mentioned language $\{a^{2^n}b^{2^n}c^{3^n} \mid n \geq 0\}$ can only be produced by a DOL system having a totally disconnected URS on three points. For the class of DOL languages such that each language in the class can be produced by exactly one DOL system obviously the URS complexity classes are disjoint. Research in this direction might shed some light on the DOL language equivalence problem.

3. GROWTH FUNCTIONS

First we explore the build up of CAD(S) and RS(S) of a semi DOL S in connection with the distribution of different letter types over the labels in the digraph. Let $S = \langle W,\delta\rangle$ be a semi DOL system. A letter $a \in W$ is of *growth type* 3 (exponential) if $\lim_{t\to\infty} \lg(\delta^t(a)) / x^t > 0$ for some $x > 1$; of *growth type* 2 (polynomial) if there exist polynomials p,q such that $p(t) \leq \lg(\delta^t(a)) \leq q(t)$ for all t; of *growth type* 1 (limited) if there is a constant c such that $1 \leq \lg(\delta^t(a)) \leq c$ for all t; of *growth type* 0 if $\lg(\delta^t(a)) = 0$ for all $t \geq \#W$. Similarly we classify growth types of DOL systems $G = \langle W,\delta,w\rangle$ where we substitute w for a in the definition (note that $\lg(\delta^t(a_1a_2\ldots a_n)) = \lg(\delta^t(a_1)) + \lg(\delta^t(a_2)) + \ldots + \lg(\delta^t(a_n))$). A complete investigation into growth types of letters, DOL systems and semi DOL systems appears in [16]. By definition $GT(i) = \{a \in W \mid a \text{ is of growth type } i\}$, $i = 0,1,2,3$. We say that a point p is *reachable* from a point q in a digraph D if there is a sequence $(p_1,p_2),(p_2,p_3),\ldots,(p_{n-1},p_n)$ in D such that $q = p_1$ and $p = p_n$.

We can distinguish two distinct regions in CAD(S): an exponential region and a polynomial region (and of course a region consisting of points labelled by mortal letters). Clearly no point in the exponential region (labelled by subsets of GT(3)) is reachable from a point in the polynomial region (labelled by subsets of $GT(2) \cup GT(1)$). Both regions have minimal elements; for the exponential region

$$M_E = \{[a] \in R/\!\sim \mid [a] \text{ is the label of a minimal exponential point}\},$$

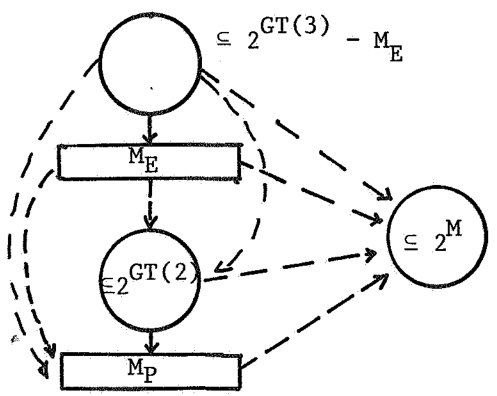
and

$$M_E \subseteq \{[a] \in R/\!\sim \mid [a] \subseteq E\};$$

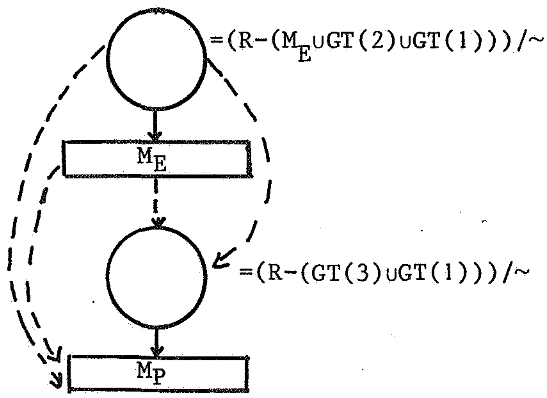
for the polynomial region:

$$M_P = \{[a] \in R/\!\sim \mid [a] \subseteq MR\}.$$

For RS(S) the same is the case except that only labels in $R/\!\sim$ occur. The following pictures hold, where a solid arrow implies that at least one point in the lower set is reachable from each point in the upper set; a square means that distinct points in this set can not be reached from each other; a dotted arrow implies possible reachability.



CAD(S)



RS(S)

If we talk about a digraph associated with a DOL system G we shall assume that G is reduced and we restrict the homomorphism involved accordingly and write CAD(G), RS(G), etc.

THEOREM 1.
(i) *The language generated by a DOL system* $G = \langle W,\delta,w\rangle$ *is finite iff* RS(G) *is totally disconnected and* $\cup R/\sim = MR$ *(i.e., if* RS(G) *consists of the bottom rectangle in the previous figure only).*
(ii) *If* RS(G) *is nonempty, totally disconnected and* $\cup R/\sim \neq MR$ *then* L(G) *is infinite and* f_G *is exponential (i.e., if* RS(G) *contains at least the upper rectangle and at most both rectangles without reachability between them).*

PROOF.
(i) According to [17] L(G) is finite iff R = MR and clearly points in MR/~ can not be reachable from each other.
(ii) Easy and left to the reader. □

Following EHRENFEUCHT & ROZENBERG [3] we define the *rank* of a DOL system. Let $G = \langle W,\delta,w\rangle$ be a DOL system.
(i) If $lg(\delta^t(a)) \leq c$ for some constant c and all t then $\rho_G(a) = 1$.
(ii) Let $W_0 = W$ and $\delta_0 = \delta$. For $j \geq 1$, δ_j denotes the restriction of δ to $W_j = W - \{a \mid \rho_G(a) \leq j\}$. For $j \geq 1$, if $lg(\delta_j^t(a)) \leq c_j$ for some constant c_j and all t then $\rho_G(a) = j + 1$.

$\rho_G(a)$ is called the *rank* of a letter a in G. If each letter $a \in W$ has a rank then G has a rank. The rank of G is the largest of the ranks of all letters accessible from the axiom, equivalently, of all letters in the axiom of G. According to [3] G is a DOL system with rank iff there are polynomials (p,q) of degree (rank G - 1) such that $p(t) \leq f_G(t) \leq q(t)$ for all t.

THEOREM 2. *The rank of* G *is equal to the length of the longest path in* RS(G) *iff* $E = \emptyset$.

PROOF (outline).
"only if". By induction on the length of a path. Remember that all letters in MR are of rank 1.
"if". Trivial from [12, 16]. □

4. THE LOCALLY CATENATIVE PROPERTY

We now turn our attention to locally catenative DOL systems producing infinite languages (finite DOL languages are trivially locally catenative), i.e., k > 1 in the locally catenative formula.

THEOREM 3. *If a DOL system* $G = \langle W,\delta,w\rangle$ *is locally catenative then* RS(G) *is a directed labelled rooted tree with branches of at most length* 1 *such that* $[c] = E$ *labels the root and* $R/\sim -\{[c]\}$ *labels the leaves,* $\cup(R/\sim -\{[c]\}) = MR$.

PROOF. If $G = \langle W,\delta,w\rangle$ is locally catenative there are fixed integers $n_0, i_1, i_2, \ldots, i_k$ such that

$$\delta^n(w) = \delta^{n-i_1}(w)\delta^{n-i_2}(w)\ldots\delta^{n-i_k}(w)$$

for all $n \geq n_0$. Therefore, $L(G) \subseteq \{\delta^i(w) \mid i < n_0\}^*$ and if $\delta^t(w) = v_1 a v_2 b v_3$, $a \sim b$ and $v_2 \in (W-[a])^*$, then

(1) $lg(v_2) < 2\max\{lg(\delta^i(w)) \mid i < n_0\}$.

Assume that a, $b \in R$ and $a \notin U(b)$. Since G is reduced at least one letter from both [a] and [b] occurs in $\delta^{n_0}(w)$. By the locally catenative property there must be an i such that

$$\delta^i(w) \in W^*[a]W^*[b]W^*[a]W^*.$$

Then for all t holds: $\delta^{i+t}(w) = v_1cv_2dv_3$ for some $c, d \in [a]$, $v_2 \in (W-[a])^+$ and $lg(v_2) \geq lg(\delta^t(e))$ for some $e \in [b]$. By (1) it follows that $[b] \subseteq MR$. Since for all $[a], [b] \in R/\sim$, $[a] \neq [b]$, either $a \notin U(b)$ or $b \notin U(a)$ we have: either all $[b] \in R/\sim$ are contained in MR and L(G) is finite or there is exactly one $[c] \in R/\sim$ which is not contained in MR. Since the assumption that there exists a $[b] \neq [c]$ such that $b \notin U(c)$ leads to the contradiction that $[c] \subseteq (R-MR) \cap MR$ we have that

(2) $b \in U(c)$ for all $b \in R - [c]$.

If $[c] \nsubseteq MR$ then L(G) is infinite by Theorem 1(i), and under the assumption that G is locally catenative, $k > 1$ in the locally catenative formula. Then, as we can easily see, f_G is exponential, and by [12] $E \neq \emptyset$. Hence $[c] = E$ and by (2) the theorem follows. □

Note that Theorem 3 gives a necessary but not sufficient condition for a DOL system to possess the locally catenative property. For instance

$G = \langle\{a,b\},\{\delta(a) = b, \delta(b) = ab\},ba\rangle$

with

$S(G) = ba, abb, babab, abbabbab, \ldots$

is easily proven not to be locally catenative but

$G = \langle\{a,b\},\{\delta(a) = b, \delta(b) = ab\},a\rangle$

with

$S(G) = a, b, ab, \ldots$

is locally catenative.

LEMMA 1. *Let* $G = \langle W,\delta,w\rangle$ *be a DOL system such that there exist integers* $n_0, i_1, i_2, \ldots, i_k$ *for which* $\delta^{n_0}(w) = \delta^{n_0-i_1}(w)\delta^{n_0-i_2}(w)\ldots\delta^{n_0-i_k}(w)$ *then G is locally catenative with formula* $(n_0, i_1, i_2, \ldots, i_k)$.

PROOF. By induction on n, $n \geq n_0$, in $\delta^n(w)$. □

Obviously, any locally catenative DOL sequence can be characterized by infinitely many locally catenative formulas. From the above lemma we see that we can assign a unique locally catenative formula to such a sequence. E.g. given a loc. cat. DOL sequence assign to it the first formula in the lexicographical ordering of the set of formulas satisfying the sequence. We call this locally catenative formula the *canonical* locally catenative formula of the DOL system. The following two decision problems suggest themselves immediately.

(i) Decide whether or not a given DOL system is locally catenative.
(ii) Decide whether two locally catenative DOL systems produce the same sequences (languages), i.e., given two locally catenative DOL systems G, G' decide whether or not $S(G) = S(G')$ $(L(G)=L(G'))$.

In view of the preceding remark on canonical locally catenative formulas the second question is settled easily, although the problem for DOL systems in general is still open.

THEOREM 4. *The sequence (language) equivalence is decidable for locally catenative DOL systems.*

PROOF. Let $G_i = \langle W,\delta_i,w\rangle$, $i = 1,2$, be two locally catenative DOL systems. $S(G_1) = S(G_2)$ iff both G_1, G_2 have the same canonical locally catenative formula, say $(n_0, i_1, i_2, \ldots, i_k)$, and $\delta_1^i(w) = \delta_2^i(w)$ for all $i < n_0$. By [7] a decision procedure for the sequence equivalence can be extended to a decision procedure for the language equivalence. □

To decide whether a DOL system is locally catenative is much more difficult but we shall prove some related results.
Define the functions c,d: $\mathbb{N} \to \mathbb{N}$ as follows:

$c(n) = \sup \{n_0 \mid G$ is a loc. cat. DOL system with an n letter alphabet and $n_0 = \inf \{m \mid m$ is the cut of a loc. cat. formula for $G\}\}$;

$d(n) = \sup \{d \mid G$ is a loc. cat. DOL system with an n letter alphabet and $d = \inf \{ m \mid m$ is the depth of a loc. cat. formula for $G\}\}$.

To decide whether or not a given DOL system is locally catenative it suffices to exhibit a total function $k: \mathbb{N} \to \mathbb{N}$ such that $k(n) \geq c(n)$ for all n and show that k is computable. We shall prove that if such a k exists then $\log k(n)$ is asymptotically greater or equal to $\sqrt{n \log n}$. For technical reasons k, c and d assume the value ∞ if they are undefined in some argument. Clearly, $k(n) \geq c(n) \geq d(n)$ for all n. First we prove a stronger result.

THEOREM 5.

$$\liminf_{n\to\infty} \frac{\log d(n)}{\sqrt{n \log n}} \geq 1.$$

PROOF. Let $G_1 = \langle W_1,\delta_1,w_1\rangle$ be a DOL system with $\#W_1 = n - 1$ and $L(G_1)$ is finite. We construct a DOL system $G_2 = \langle W_2,\delta_2,w_2\rangle$ where $W_2 = W_1 \cup \{a\}$, $a \notin W_1$, $\delta_2 = \delta_1 \cup \{\delta_2(a) = aw_1a\}$, $w_2 = aw_1$.

Claim. For all $i > 0$, $\delta_2^i(w_2) = \delta_2^{i-1}(w_2)\delta_2^{i-2}(w_2)\ldots\delta_2^0(w_2)a\delta_2^i(w_1)$.

Proof of Claim. By induction on i.

$i = 1$. $\quad \delta_2(aw_1) = aw_1a\delta_2(w_1) = \delta_2^0(w_2)a\delta_2(w_1)$.

$i > 1$. $\quad$ Suppose the claim is true for all $j \leq i$.

$$\begin{aligned}\delta_2^{i+1}(aw_1) &= \delta_2(\delta_2^i(aw_1)) = \delta_2(\delta_2^{i-1}(w_2)\delta_2^{i-2}(w_2)\ldots\delta_2^0(w_2)a\delta_2^i(w_1))\\ &= \delta_2^i(w_2)\delta_2^{i-1}(w_2)\ldots\delta_2^1(w_2)awa\delta_2^{i+1}(w_1)\\ &= \delta_2^i(w_2)\delta_2^{i-1}(w_2)\ldots\delta_2^0(w_2)a\delta^{i+1}(w_1),\end{aligned}$$

which proves the claim.

Since $L(G_1)$ is finite there are unique integers $t_0, u \in \mathbb{N}$ such that

(3) $\delta_1^t(w_1) = \delta_1^{t'}(w_1)$ for $t,t' \geq t_0$ and $t \equiv t' \bmod u$,

(4) $\delta_1^t(w_1) \neq \delta_1^{t'}(w_1)$ for $t < t_0$ and $t' \geq t_0$ or for $t,t' \geq t_0$ and $t \not\equiv t' \bmod u$.

i.e., $\#L(G_1) = t_0 + u$.

G_2 is locally catenative since for all $t \geq t_0$:

$$\begin{aligned}\delta_2^{t+u}(w_2) &= \delta_2^{t+u-1}(w_2)\delta_2^{t+u-2}(w_2)\ldots\delta_2^0(w_2)a\delta_2^{t+u}(w_1) && \text{(by the claim)}\\ &= \delta_2^{t+u-1}(w_2)\delta_2^{t+u-2}(w_2)\ldots\delta_2^t(w_2)\delta_2^{t-1}(w_2)\ldots\delta_2^0(w_2)a\delta_2^t(w_1) && \text{(by (3))}\\ &= \delta_2^{t+u-1}(w_2)\delta_2^{t+u-2}(w_2)\ldots\delta_2^t(w_2)\delta_2^t(w_2) && \text{(by the claim).}\end{aligned}$$

Since for each i holds $\delta_2^i(w_2) = \ldots a\delta_2^i(w_1)$ we see from the locally catenative formula above and from (4) that if $(n_0,i_1,i_2,\ldots,i_k)$ is a locally catenative formula for G_2 then $i_k \geq u$ and

(5) depth (loc. cat. formula of G_2) $\geq u$.

In [17] the maximum cardinality of a finite DOL language over n letters was studied. Let $u(n)$ be the maximum *period* of a finite DOL language over n letters, i.e.,

$u(n) = \sup \{u \mid G = \langle W,\delta,w\rangle$ with $\#W = n$ is a DOL system generating a finite language with u defined by (3) and (4)$\}$,

then, according to [17],

$u(n) = \sup \{\text{l.c.m.}(k_1,k_2,\ldots,k_q) \mid k_1,k_2,\ldots,k_q$ is a partition of n in $q \le n$ positive integral summands$\}$

and

$$\lim_{n\to\infty} \frac{\log u(n)}{\sqrt{n \log n}} = 1. \qquad \text{Hence also} \qquad \lim_{n\to\infty} \frac{\log u(n-1)}{\sqrt{n \log n}} = 1,$$

and by (5) $d(n) \ge u(n-1)$ for all n. Therefore,

$$\lim_{n\to\infty} \inf \frac{\log d(n)}{\sqrt{n \log n}} \ge 1. \qquad \square$$

COROLLARY 1.

$$\lim_{n\to\infty} \inf \frac{\log k(n)}{\sqrt{n \log n}} \ge \lim_{n\to\infty} \inf \frac{\log c(n)}{\sqrt{n \log n}} \ge \lim_{n\to\infty} \inf \frac{\log d(n)}{\sqrt{n \log n}} \ge 1,$$

where we can substitute $\sqrt{p_n}$ *for* $\sqrt{n \log n}$ *in the formulas by the well-known asymptotic approximation of the n-th prime number* p_n.

Finally, we provide an equivalent form of the locally catenative property, which links this property of the derived sequence with a property of the derived language.

THEOREM 6. *Let* G *be a DOL system. The following two statements are equivalent:*
(i) G *is locally catenative*
(ii) *The monoid* $L(G)^*$ *is finitely generated.*

PROOF.
(i) $\to$ (ii). Let $G = \langle W,\delta,w\rangle$ be a locally catenative DOL system with formula $(n_0,i_1,i_2,\ldots,i_k)$. Then $L(G) \subseteq \{\delta^i(w) \mid i < n_0\}^* \subseteq L(G)^*$ and therefore $L(G)^* = \{\delta^i(w) \mid i < n_0\}^*$.
(ii) $\to$ (i). Suppose $L(G)^* = \{v_1,v_2,\ldots,v_\ell\}^* \subseteq W^*$. Without loss of generality we can assume that $v_i \notin \{v_1,v_2,\ldots,v_{i-1},v_{i+1}\ldots,v_\ell\}^*$ for all i, $1 \le i \le \ell$. Hence $v_i \in L(G)$ for all i, $1 \le i \le \ell$, and there is a j, $i \le j \le \ell$, such that $v_j = \delta^t(w)$ for some t and for no j', $1 \le j' \le \ell$, $v_{j'} = \delta^{t'}(w)$ with $t' > t$. Hence there exist $j_1,j_2,\ldots,j_k$ such that $\delta^{t+1}(w) = v_{j_1}v_{j_2}\ldots v_{j_k}$ and therefore there are $i_1,i_2,\ldots,i_k$ such that $\delta^{t+1}(w) = \delta^{t+1-i_1}(w)\delta^{t+1-i_2}(w)\ldots\delta^{t+1-i_k}(w)$ where $\delta^{t+1-i_h}(w) = v_{j_h}$ for all h, $1 \le h \le k$. By Lemma 1 G is locally catenative. $\square$

5. REGULARITY AND CONTEXT-FREE-NESS

In [13] SALOMAA proves that the regularity and context-free-ness of DOL languages is decidable. Roughly, this is achieved as follows. Given a DOL system G, with at most a linear growth function, we can construct (a decomposition of G in) DOL systems $G_1,G_2,\ldots,G_k$ such that $L(G) = h(L(G_1) \cup L(G_2) \cup\ldots\cup L(G_k))$ where h is a nonerasing homomorphism. $G_1,G_2,\ldots,G_k$ satisfy restrictions like: there are no mortal letters in G_i and every letter from the alphabet of G_i occurs in each word in $L(G_i)$. Salomaa then gives a definition of the degree of a DOL system G satisfying said restrictions and proves:

LEMMA 2 (SALOMAA). *If* G *has degree* ≤ 1 *then* L(G) *is regular. If* G *is of degree* > 4 *then* L(G) *is noncontext-free. If* G *is of degree* 2, L(G) *is context-free and possibly regular. If* G *is of degree* 3 *or* 4, L(G) *is nonregular (but possibly context-free). It is decidable which of the alternatives hold in the last two sentences.*

Since a DOL system can only generate a context-free language if its associated growth function is bounded by a linear polynomial we have the following. If L(G) is context-free then RS(G) contains paths of at most length 1 and $E = \emptyset$. We can improve on Salomaa's results by showing that under a slightly modified definition of degree decomposition of G is not necessary.

For the vital letters of a DOL system $G = \langle W,\delta,w \rangle$ we define the *degree* as follows.

degree (a) = 0 iff $U(a) \cap (R-MR) = \emptyset$; $O = \{a \mid \text{degree}(a) = 0\}$,

degree (a) = 2 iff $U(a) \cap (R-MR) = [a]$ and $\delta^i(a) = v_1 a v_2$ for some $i \leq \#W$ and $v_1, v_2 \in (O \cup M)^* O (O \cup M)^*$,

degree (a) = 1 iff $U(a) \cap (R-MR) = [a]$ and $\delta^i(a) = v_1 a v_2$ or $v_2 a v_1$ for some $i \leq \#W$ and $v_1 \in (O \cup M)^* O (O \cup M)^*$, $v_2 \in M^*$.

The degree of G is found by adding the degrees of all vital letters in $\delta^{\#(W-R)}(w)$ where each letter is counted as many times as it occurs. Note that f_G is linear iff all letters occurring in $\delta^{\#(W-R)}(w)$ have a degree or are mortal.

THEOREM 7. *Under the given definition of the degree of a DOL system Lemma 2 holds for arbitrary DOL systems.*

INDICATION OF PROOF. The degree of a letter is invariant if we substitute δ by δ^k in the definitions, i.e., under decomposition. Furthermore, the degree of a letter is invariant under restriction of δ to the vital letters, or equivalently, if G has degree i then the PDOL G', constructed such that there is a nonerasing homomorphism h such that $hS(G') = S(G)$, has degree i. Therefore each G_i, $1 \leq i \leq k$, in the above decomposition of G in $G_1, G_2, \ldots, G_k$ has the degree of G.

Since each letter in $[a] \in R/\sim$ must have the same degree in G (if f_G is bounded by a linear polynomial) we say degree [a] = degree (a). If degree [a] = 1,2 then $[a] \subseteq R - (MR \cup E)$ and $[b] < [a] \Rightarrow b \subseteq MR$. (N.B. $[b] < [a]$ if $b \in U(a)$ and $a \notin U(b)$.) From sections 3,4,5 we have obtained good criteria to prove that a language does not belong to a given language family.

COROLLARY 2.

- $L(G)$ *is finite iff* $\sum_{[a]\in R/\sim}$ *degree* $[a] = 0$.
- *If* $L(G)$ *is regular then* $\sum_{[a]\in R/\sim}$ *degree* $[a] \leq 2$.
- *If* $L(G)$ *is context-free then* $\sum_{[a]\in R/\sim}$ *degree* $[a] \leq 4$.
- *If* $L(G)$ *is infinite and locally catenative then* $E = [b]$ *for some letter* b *and* $\sum_{[a]\in R/\sim-\{[b]\}}$ *degree* $[a] = 0$.

6. BIOLOGICAL INTERPRETATION

In biology we encounter the phenomenon of cell differentiation as opposed to cell potential. In higher species cells become so specialized (highly differentiated) that they loose their ability to produce cells of other types (low potential). In the embryonic stage, and to a large extent in the vegetatative kingdom this seems not to be the case (low differentiation and high potential). The associated digraphs, as I-IV, form in increasing levels of abstraction a formal representation of cell lineage and cell differentiation of an organism modelled by a DOL system. In I the AD depicts the cell lineage. The CAD in II shows us the stages of cell differentiation where the labels consisting of sets of recursive letters correspond to meta-stable stages of cell differentiation, i.e. the descendancy of such a cell always contains a cell with the same cell potential as the original one and each cell type of a meta-stable stage of differentiation occurs in the descendancy or each other cell type of this stage. The points labelled by singleton sets of vital nonrecursive letters correspond to transitory stages of cells between one meta-stable stage of cell differentiation and a next one. The RS shows us the lineage between the meta-stable stages which is of prime importance and the URS the same structure without labels.

EXAMPLES.

(i) If the CAD consists of the graph on one point the modelled organism is very regenerative: each cell type has the possibility of deriving any other cell type.

(ii) If the CAD consists of a directed tree we observe a type of cell differentiation similar to that in higher organisms. Cells in the leaves of the tree are completely specialized and have no regenerative capacity to produce cells of other types in their progeny, as opposed to the cells at the root which can produce all other cell types.

(iii) To be able to reproduce from a single cell the CAD of the associated DOL system must be such that every two points of the CAD have a common ancestral point while the unique maximal element is labelled by an equivalence class of recursive letters. The rules must be such that at any time the description of the organism (i.e., the produced string) contains a cell in the maximal point of the CAD. All living plants and animals seem always to contain some cells which are capable of division, and through that to give rise to cells from which a new similar organism can be derived.

To interpret some of the results in sections 3-5:

If an organism grows under optimal conditions (and if it can be adequately modelled by a DOL system) it exhibits linear growth iff it has exactly two meta-stable levels of cell differentiation. More generally, if it exhibits polynomial growth of degree n it has exactly n + 1 meta-stable levels of cell differentiation (by this we mean that if we trace the cell lineage from a least differentiated cell to a most differentiated cell there is at least one cell lineage such that we meet n + 1 different meta-stable stages of differentiation).

If an organism has the locally catenative property, i.e., if at a time t the organism is composed from previous stages in its developmental history [9], it contains at most two meta-stable levels of differentiation and it can be grown from cells occurring in a single uppermost meta-stable stage of differentiation. The RS is a tree of at most two levels, with a meta-stable stage of cell differentiation at the top from which all other, completely differentiated cell types are derived without intermediate meta-stable stages of differentiation. Another result shows that if a relatively simple organism, i.e. one having not many different cell types, is locally catenative we might have to wait a very long time to see that it is such.

In general we can think of the URS, or the genealogical relations between meta-stable stages of cell differentiation, as a measure of the complexity of the organism, see e.g. Corrolary 2.

REFERENCES

[1] DOUCET, P.G., *On the applicability of L systems in developmental biology*. In: Abstracts of Conf. on Formal languages, Automata and Development held at Noordwijkerhout, The Netherlands (A. Lindenmayer and G. Rozenberg, eds.) (1975).

[2] DOUCET, P.G., Ph.D. thesis, Univ. of Utrecht (forthcoming).

[3] EHRENFEUCHT, A. & G. ROZENBERG, *DOL systems with rank*. In: [10], 136-141.

[4] HARARY, F., *Graph Theory*. Addison-Wesley (1971).

[5] HERMAN, G.T. & G. ROZENBERG, *Developmental Systems and Languages*. North-Holland / American Elsevier (1975).

[6] LINDENMAYER, A., *Mathematical models for cellular interaction in development I & II*. J. Theoret. Biol. 18 (1968), 280-315.

[7] NIELSEN, M., *On the decidability of some equivalence problems for DOL systems*. Inf. Contr. 25 (1974), 166-193.

[8] ROBINSON, R.W., *Enumeration of acyclic digraphs*. In: Proc. Second Chapel Hill Conf. on Comb. Math. and its Applications, Univ. of North Carolina, Chapel Hill, N.C. (1970), 391-399.

[9] ROZENBERG, G. & A. LINDENMAYER, *Developmental systems with locally catenative formulas*, Acta Informatica 2 (1973), 214-248.

[10] ROZENBERG, G. & A. SALOMAA (eds.), *L systems*. Lecture Notes in Computer Science, vol. 15 (1974), Springer Verlag, Berlin, etc.

[11] SALOMAA, A., *Formal Languages*. Academic Press (1973).

[12] SALOMAA, A., *On exponential growth in Lindenmayer systems*, Indag. Math. 35 (1973), 23-30.

[13] SALOMAA, A., *Comparative decision problems between sequential and parallel rewriting* (extended abstract).

[14] SZILARD, A.L., *Growth functions of Lindenmayer systems*, Univ. of Western Ontario, Comp. Sci. Dept., Tech. Rept. #4 (1971).

[15] VITÁNYI, P.M.B., *DOL languages and a feasible solution for a word problem*. Mathematisch Centrum, Amsterdam, Tech. Rept. MR 138/72 (1972).

[16] VITÁNYI, P.M.B., *Structure of growth in DOL systems*. Indag. Math. 35 (1973), 247-253.

[17] VITÁNYI, P.M.B., *On the size of DOL languages*. In: [10], 78-92.