ma
the
ma
tisch

cen
trum

AFDELING INFORMATICA IW 11/73 NOVEMBER

PAUL M.B. VITÁNYI

A NOTE ON NONRECURSIVE AND DETERMINISTIC
LINDENMAYER LANGUAGES

Prepublication

amsterdam 1973

stichting
mathematisch
centrum

MC

AFDELING INFORMATICA IW 11/73 NOVEMBER

PAUL M.B. VITÁNYI

A NOTE ON NONRECURSIVE AND DETERMINISTIC LINDENMAYER LANGUAGES

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, 94A30
ACM - Computing Reviews - category: 5.22, 5.23

A NOTE ON NONRECURSIVE AND DETERMINISTIC LINDENMAYER LANGUAGES *)

Paul M.B. Vitányi

1. INTRODUCTION

While it is known that there are recursively enumerable Lindenmayer languages which are not recursive, the usual proof of such a fact is relative to an infinite class of Turing machines or an infinite class of inputs on a single Turing machine. Here we consider the construction of a specific nonrecursive Lindenmayer language. Furthermore, we determine completely the place in the Chomsky hierarchy of the families of the deterministic Lindenmayer languages and the deterministic Lindenmayer languages without erasing.

2. L SYSTEMS AND TURING MACHINES

Lindenmayer systems, L Systems for short, are a class of parallel rewriting systems without nonterminals introduced by Lindenmayer as a model for developmental growth in filamentous organisms [3]. The present note claims no relevance to these biological origins. An L System consists of an initial filament symbolized by a string of letters and the subsequent stages of development are obtained by rewriting all letters in a string simultaneously at each time step. When the rewriting of a letter depends on the m left and n right letters, we talk about an (m,n)L System. More precisely, a *deterministic* (m,n)L *System*, D(m,n)L, is a 4 tuple $G = \langle W,\delta,w,g\rangle$ where W is a finite nonempty *alphabet*,

*) This paper is not for review; it is meant for publication in a journal.

$w \in W^*\backslash\{\lambda\}$ is the *axiom* (λ is the *empty word*), $g \notin W$ is the *environmental letter* and δ is a total mapping from $\{g\}^i W^j \{g\}^k$, $i+j+k = m+n+1$, $0 \leq i \leq m$, $0 \leq k \leq n$, into W^* (we consider δ to rewrite the (m+1)th letter from its argument). δ is extended to $\{g\}^m W^* \{g\}^n$ by defining $\delta(g^{m+n}) = \lambda$ and $\delta(b_1 \ldots b_{m+n+h}) = \delta(b_1 \ldots b_{m+n+1}) \ldots \delta(b_h \ldots b_{m+n+h})$, $h > 0$, where $b_1 \ldots b_{m+n+h} \in \{g\}^m W^* \{g\}^n$. We define δ^i inductively by $\delta^0(g^m v g^n) = v$ and $\delta^i(g^m v g^n) = \delta(g^m \delta^{i-1}(g^m v g^n) g^n)$ for $i > 0$. The L *language* produced by G is $L(G) = \{\delta^i(g^m w g^n) \mid i \geq 0\}$.

It was shown by van Dalen [1], and is indeed a nice exercise, that for a suitable standard formulation of Turing machines, e.g. the quintuple version, for every Turing machine T with symbol set S and state set Q we can effectively construct a D(1,1)L $G = \langle W,\delta,w,g \rangle$, $W = Q \cup S$, which simulates it in real time, i.e. the t th *instantaneous description* of T is equal to $\delta^t(gwg)$ (see Minsky [4] for terminology and results on Turing machines). If we do away with the excess blank symbols on the ends of the T.M. tape by letting them derive the empty word λ in the case of the L System, we see that the canonical extensions of the D(1,1)L languages are precisely the recursively enumerable languages, i.e. the languages $h_1(L(G) \cap \Delta^* Q \Delta^*)$, $\Delta \subseteq S$ and h_1 is a homomorphism from $\Delta^* Q \Delta^*$ into Δ^* defined by $h_1(q) = \lambda$ and $h_1(s) = s$ for all $q \in Q$ and all $s \in \Delta$, are exactly the recursively enumerable languages over Δ. We call a homomorphism h *λ-limited* on a set A if there exists an integer $k \geq 0$ such that for all $w \in A$, if $w = xyz$ and $h(y) = \lambda$ then $|y| \leq k$. (For further details concerning operations on languages and closure under these operations see Ginsburg et. al. [2].)

The recursive languages are closed under intersection with regular languages and λ-limited homomorphism. Since $\Delta^*Q\Delta^*$ is regular and h_1 is λ-limited on $\Delta^*Q\Delta^*$ there exist L(G)s which are not recursive. The usual proofs that there are recursively enumerable languages that are not recursive rest on arguments relative to an infinite class of T.M.s or an infinite class of inputs on a single T.M. By application of a result due to Rabin and Wang [5] we can exhibit a specific nonrecursive L language. Let T be a T.M. with S = {b,1,a} where b is the blank. Let the *word* at any moment t in the history of a T.M. be the string consisting of the contents of the minimum block on the tape at t that includes all the marked squares and the square scanned at the initial moment (the origin).

Theorem 1. (Rabin and Wang). For any fixed (finite) word at the initial moment we can find a T.M. T such that the set of words P in its subsequent history is not recursive.

Proof. Take a nonrecursive set $A \subseteq \{1\}^*$ enumerated by a one-one recursive function f; we can recover n from f(n) by f^{-1}. We can now give a T.M. T which first erases the finitely many marks on the initial tape and returns to the origin, puts down the representation of 0 on the tape and calculates the value of f(0). Subsequently, T erases everything else except the representation of f(0), retrieves the representation of 0 from f(0) by f^{-1}, adds one to this representation and computes f(1), and so on.

In particular we can do it in such a way that the symbol a is used, after the initial tape contents is erased, only to mark f(0), f(1),....; it is erased before we evaluate f(n+1) from f(n). Moreover, the string consisting of a followed by the representation of f(n) always begins at the origin. Let h_2 be a homomorphism from $\{a\}\{1\}^*$ into $\{1\}^*$ defined by $h_2(a) = \lambda$ and $h_2(1) = 1$. Now $h_2(P \cap \{a\}\{1\}^*) = A$ where h_2 is λ-limited on $\{a\}\{1\}^*$ and $\{a\}\{1\}^*$ is regular. Since A is nonrecursive P must be nonrecursive by the closure of the recursive languages under λ-limited homomorphism and intersection with regular sets □.

3. SOME NONRECURSIVE L LANGUAGES

Theorem 2. Let G be a D(1,1)L which simulates T in the sense explained above. Then L(G) is not recursive.

Proof. Let h_3 be a homomorphism on L(G) defined by $h_3(s) = s$ and $h_3(q) = \lambda$ for all $s \in S$ and all $q \in Q$. Since $L(G) \subseteq S^*QS^*$, h_3 is λ-limited on L(G). Now $h_3(L(G)) = P$ and since P is not recursive L(G) is not recursive □.

We use G to construct a nonrecursive D(0,1)L language. Let $G = \langle W,\delta,w,g\rangle$ be the D(1,1)L above. Define a D(0,1)L $G' = \langle W',\delta',w',g\rangle$ as follows:

$$W' = W \cup \{¢\} \cup W \times (W \cup \{g\}) \text{ where } ¢ \notin W;$$

$$w' = ¢w;$$

$$\delta'(ab) = (a,b)$$

$$\delta'(ag) = (a,g)$$

$$\delta'(¢a) = ¢$$

$$\delta'(¢g) = ¢$$

$$\delta'((a,b)(b,c)) = \delta(abc)$$

$$\delta'(¢(a,b)) = ¢\delta(gab)$$

$$\delta'(¢(a,g)) = ¢\delta(gag)$$

$$\delta'((a,b)(b,g)) = \delta(abg)$$

$$\delta'((a,g)g) = \lambda, \qquad \text{for all } a,b,c \in W.$$

(The arguments for which δ' is not defined will not occur in our operation of G').

For all words $v = a_1 \ldots a_n \in W^*$ holds:

$$\delta'^2(\not{c}a_1\ldots a_n g) = \delta'(\not{c}(a_1,a_2)(a_2,a_3)\ldots(a_n,g)g) =$$

$$= \not{c}\delta(ga_1a_2)\delta(a_1a_2a_3)\ldots\delta(a_{n-1}a_ng) = \not{c}\delta(gvg), \quad n > 1;$$

$$\delta'^2(\not{c}a_1g) = \delta'(\not{c}(a_1,g)g) = \not{c}\delta(ga_1g) = \not{c}\delta(gvg), \quad n = 1;$$

$$\delta'^2(\not{c}g) = \delta'(\not{c}g) = \not{c}\delta(gg) = \not{c}, \quad n = 0.$$

Therefore $\delta'^{2t}(\not{c}w) = \not{c}\delta^t(w)$ for all t.

Define a homomorphism h_4 from $\{\not{c}\}W^*$ into W^* by $h_4(\not{c}) = \lambda$ and $h_4(a) = a$ for all $a \in W$. Since $h_4(L(G')\cap\{\not{c}\}W^*) = L(G)$, where h_4 is λ-limited on $\{\not{c}\}W^*$ and L(G) is not recursive, L(G') cannot be recursive. Hence we have:

Theorem 3. The D(0,1)L language L(G') is nonrecursive.

Of course, P,L(G) and L(G') are recursively enumerable.

4. DETERMINISTIC L LANGUAGES AND THE CHOMSKY HIERARCHY

From the working space theorem, Salomaa [6] (see also [1]), it follows that the *propagating* D(m,n)Ls ($\delta(.) \neq \lambda$ for all $. \neq g^{m+n}$) produce only context sensitive languages which illustrates the role of erasing productions in L Systems. That the containment of the families of D(m,n)L languages and propagating D(m,n)L (PD(m,n)L) languages in the recursively enumerable and the context sensitive languages is proper follows from the next lemma.

Lemma. There are (non trivial) regular languages over a one letter alphabet which are not D(m,n)L languages.

Proof. $L = (aaa)^*(a \cup aa)$ is such a language. To prove this we make use of:

Claim. If $G = \langle W,\delta,w,g,\rangle$ is a *unary* (i.e. $\#W=1$) D(m,n)L which generates an infinite language then there exist positive integers t_0, p and x such that for all $t \geq t_0$

(1) $|\delta^{t+1}(g^m w g^n)| = p(|\delta^t(g^m w g^n)|-m-n) + x$ where $|v|$ denotes the length of a word v.

Proof of Claim. Let $\delta(a^m a a^n) = a^p$ and

$x = \sum_{i=1}^{m} |\delta(g^i a^{m+n+1-i})| + \sum_{j=1}^{n} |\delta(a^{m+n+1-j} g^j)|$. If L(G) is infinite then there exists a t_0 such that $|\delta^{t_0}(g^m w g^n)| \geq 2(m+n) + x + 1$.

Case 1. $p = 0$. $|\delta^t(g^m w g^n)| \leq (m+n)y$ for all $t > 0$ where $y = \max\{|\delta(v)| \mid |v| = m+n+1\}$: contrary to the assumption.

Case 2. $p > 0$. Clearly, (1) holds.

By observing that $L = \{a^i \mid i \not\equiv 0 \bmod 3\}$ we see that for every positive integer k such that $k \equiv 0 \bmod 3$: $a^{k-1}, a^{k+1}, a^{k+2} \in L$ and $a^k \notin L$. Hence if $L(G) = L$ it follows that $p = 1$ in (1). But then the lengths of the subsequent words of L(G), ordered by increasing length, differ by a constant amount x-m-n and hence $L(G) \neq L$ □.

We are now in the position to determine completely the place in the Chomsky hierarchy of the families of the D(m,n)L and PD(m,n)L languages. By the family of *strictly* X languages we mean the languages which belong to the difference of the type i (i.c. X) and type i+1 languages

$i \in \{0,1,2\}$.

Theorem 4.

(i) For all $m,n \geq 0$ the family of PD(m,n)L languages has nonempty intersections with the regular, strictly context free and strictly context sensitive languages; it is strictly included in the family of context sensitive languages; there are regular, strictly context free and strictly context sensitive languages which are not PD(m,n)L languages.

(ii) For all $m,n \geq 0$ $m+n > 0$ the family of D(m,n)L languages has nonempty intersections with the regular, strictly context free, strictly context sensitive and strictly recursively enumerable languages; it is strictly included in the family of recursively enumerable languages; there are regular, strictly context free, strictly context sensitive and strictly recursively enumerable languages which are not D(m,n)L languages.

(iii) The family of PD(m,n)L languages is strictly included in the family of D(m,n)L languages.

Proof. (i) and (ii). Let G_1, G_2 and G_3 be PD(0,0)Ls defined by

$$G_1 = \langle \{a\}, \{\delta(a)=a\}, a, g \rangle$$

$$G_2 = \langle \{a,b,c\}, \{\delta(a)=a, \delta(b)=b, \delta(c)=acb\}, c, g \rangle$$

$$G_3 = \langle \{a\}, \{\delta(a)=aa\}, a, g \rangle$$

Now $L(G_1) = \{a\}$ is regular; $L(G_2) = \{a^n cb^n | n \geq 0\}$ is a well known strictly context free language; $L(G_3) = \{a^{2^n} | n \geq 0\}$ is context sensitive by the working space theorem and not context free by the uvwxy lemma (see e.g. [6]). Together with the strictly recursively enumerable language

L(G') of theorem 3 this proves the above statements about nonempty intersections.

The language L of the lemma is regular but not a D(m,n)L language. $L \cup L(G_2)$ is strictly context free and it is easy to see that $L \cup L(G_2)$ is not a D(m,n)L language. $L' = \{a^{2^{2^t}} \mid t \geq 0\}$ is strictly context sensitive by the working space theorem and the uvwxy lemma and cannot be produced by a D(m,n)L in view of equation (1). A in theorem 1 is a strictly recursively enumerable language over a one letter alphabet and cannot be produced by a D(m,n)L in view of equation (1). This proves the statements above about languages which are not D(m,n)L languages and hence not PD(m,n)L languages. The statements about strict inclusion were previously established. (iii) follows from (i) and (ii) and the definitions of PD(m,n)L- and D(m,n)L languages □.

REFERENCES

[1] D. van Dalen, *A note on some systems of Lindenmayer*, Math. Systems Theory 5, 128-140 (1971).

[2] S. Ginsburg, S. Greibach, J. Hopcroft. Studies in Abstract Families of Languages. Memoirs of the American Mathematical Society, No 87 (1969).

[3] A. Lindenmayer, *Mathematical models for cellular interactions in development* I *and* II, J. Theoret. Biol. 18, 280-315 (1968).

[4] M. Minsky, Computation: Finite and Infinite Machines. Prentice Hall, Englewood Cliffs (1967).

[5] M.O. Rabin and H. Wang, *Words in the history of a Turing machine with a fixed input*, J. ACM 10, 526-527 (1963).

[6] A. Salomaa. Formal Languages. Academic Press (1973).