ON THE MEMBERSHIP QUESTION IN SOME LINDENMAYER-SYSTEMS

BY

P. G. DOUCET

(Communicated by Prof. A. van Wijngaarden at the meeting of June 26, 1971)

Summary

Lindenmayer-systems are a family of string-generating systems, and several types can be distinguished, of different strength. The subject of this article is the membership question in the languages generated by determinate, no-input Lindenmayer-systems (or, shortly, D0L-languages). For propagating D0L-systems (which contain no production rules yielding the empty string) the question is easily answered. For D0L-systems in general it can be solved in two steps, involving the construction of an intermediate "backbone" system, which is propagating.

Notation; description of D0L-systems

The notation conforms to common usage in language theory, and I shall only state a few conventions.

Any sequence of elements of an alphabet Σ is called a *word* over Σ; Λ is the empty word.

If $\sigma \in \Sigma$, then $\sigma^2 = \sigma\sigma$, etc.; $\sigma^0 = \Lambda$.

Σ^* is the Kleenean closure of Σ.

$\# \Sigma$ denotes the number of elements of Σ.

$|w|$ denotes the length of a word w.

If δ is a mapping of a set into itself, then $\delta^2(x) = \delta(\delta(x))$, etc.; $\delta^0(x) = x$.

Elements (or *letters*) of an alphabet shall usually be denoted by $\sigma, \tau, \ldots$, words by $u, v, w, x, \ldots$.

L-systems were first proposed by A. Lindenmayer in 1968 [3], as a model for the development of filamentous organisms. D0L-systems, which form a subclass of the L-systems, can be formally described as follows:

Definition. A *D0L-system* is an ordered triple $G = \langle \Sigma, \delta, w_0 \rangle$, where Σ (the *alphabet*) is a finite, nonempty set, δ (the set of *productions*) is a total mapping from Σ into Σ^*, and w_0 (the *axiom*) is a non-empty word over Σ.

The domain of δ is extended from Σ to Σ^* in a natural way, by defining

1. $\delta(\Lambda) = \Lambda$.
2. $\delta(\sigma w) = \delta(\sigma)\ \delta(w)$ for all $\sigma \in \Sigma$ and $w \in \Sigma^*$.

The relation $\Rightarrow$ in D0L-systems (and in L-systems in general) is not the same as in the usual grammars:

DEFINITION. Let $G=\langle \Sigma, \delta, w_0 \rangle$ be a D0L-system.

1. $w_1 \Rightarrow w_2$ if and only if $\delta(w_1)=w_2$.
2. $w_1 \overset{*}{\Rightarrow} w_2$ if and only if $\delta^m(w_1)=w_2$ for some $m \geqslant 0$.
3. The *language* of G is defined by $L(G)=\{w: w_0 \overset{*}{\Rightarrow} w\}$.

Note that

(a) if $w_1 \Rightarrow w_2$, then w_2 is the result of a δ-mapping applied to *each* letter of w_1.

(b) there is no distinction between terminal and non-terminal letters, and every word derived from w_0 is in $L(G)$.

These properties are characteristic for L-systems; they account for many of the differences with other grammars.

If $\delta(\sigma) \neq \Lambda$ for all $\sigma \in \Sigma$, then G is called a *propagating* D0L-system (or, with biological connotations, *without cell-death*), abbreviated as PD0L-system.

EXAMPLE 1. $G=\langle \{0, 1\}, \delta, 011 \rangle$ with δ as follows:
$0 \rightarrow 1001$, $1 \rightarrow \Lambda$.
From 011, words are produced as follows:
$011 \Rightarrow 1001 \Rightarrow 10011001 \Rightarrow (1001)^4 \Rightarrow (1001)^8$ etc.
So $L(G)=\{011\} \cup \{(1001)^{2n}: n \geqslant 0\}$.
G is a D0L-system; it is not propagating.

EXAMPLE 2. $H=\langle \{0, 1\}, \delta, 10 \rangle$, with δ: $0 \rightarrow 00$, $1 \rightarrow 0$, $1 \rightarrow 1$.
H is propagating; but it is no D0L-system, the two-valuedness of $\delta(1)$ violating the condition that δ be a mapping from Σ into Σ^*.

PD0L-SYSTEMS

LEMMA 1. Let $G=\langle \Sigma, \delta, w_0 \rangle$ be a PD0L-system.
If $L(G)$ is infinite, then it contains at most $\# \Sigma$ words of a given length; formally:

$$(\forall k)(\# \{w \in L(G): |w|=k\} \leqslant \# \Sigma).$$

Proof: Because $L(G)$ is infinite, δ contains some strictly increasing production rules (i.e., with $|\delta(\sigma)| \geqslant 2$).

Because G is propagating, word length cannot decrease, so all words of the same length follow immediately upon each other.

Let w_1 and w_z be the first and last words of length k. Then w_z contains at least one letter, say σ_z, with $|\delta(\sigma_z)| \geqslant 2$.

Let $\sigma_1 \Rightarrow \sigma_2 \Rightarrow \ldots \Rightarrow \sigma_{z-1} \Rightarrow \sigma_z$ (with each σ_i in w_i). Then $|\delta(\sigma_i)|=1$ for all σ_i except σ_z.

Now all $\sigma_1, \ldots, \sigma_z$ have to be different (for, if two of them were equal, σ_z would also be equal to some σ_y; but $|\delta(\sigma_y)|=1$ while $|\delta(\sigma_z)| \geqslant 2$, and in a determinate system this contradicts $\sigma_y=\sigma_z$), and this implies $z \leqslant \# \Sigma$.

COROLLARY: In PD0L-systems the membership question is effectively solvable.

The decision procedure is to simply start producing words from w_0 on, and stop as soon as either some previous word reappears (in which case $L(G)$ is finite), or a word longer than w (the word in question) is produced. Lemma 1 assures us that, if $L(G)$ is infinite, this will happen in at most $(|w|+1-|w_0|) \cdot \# \Sigma$ production steps.

LEMMA 2. Let $G=\langle \Sigma, \delta, w_0 \rangle$ be a D0L-system. If there exist different m_1 and m_2 such that $\delta^{m_1}(w_0)=\delta^{m_2}(w_0)$, then $L(G)$ is finite.

Proof: Let $m_2=m_1+q$ with $q>0$.
For any $z>m_2$, $z=m_1+\lambda q+\mu$ for some $\lambda \geqslant 1$ and $0 \leqslant \mu < q$.
Then $\delta^z(w_0)=\delta^{m_1+\lambda q+\mu}(w_0)=\delta^{m_1+\mu}(w_0)$.
Consequently, no new words are produced after w_{m_2}. So $L(G)$ is finite.

The converse of lemma 2 also holds.

LEMMA 3. Let $G=\langle \Sigma, \delta, w_0 \rangle$ be a PD0L-system with $\# \Sigma = n$.
1. It is effectively decidable whether $L(G)$ is finite.
2. If $L(G)$ is finite, then $\# L(G) \leqslant n^n+n$.

Proof: I shall use the following conventions:
- $\sigma \in \Sigma$ will be called *ultimately periodic* if

$$(\exists a \geqslant 0, \pi \geqslant 1)(\delta^{a+\pi}(\sigma)=\delta^a(\sigma)).$$

Note that the smallest among possible a's is always smaller than n. The smallest among different π's is called the *period* of σ. Note that this period is never larger than n.
- Σ can be partitioned as follows:

$$\Sigma_1=\{\sigma \in \Sigma: |\delta^n(\sigma)|=1\}.$$
$$\Sigma_2=\Sigma-\Sigma_1.$$

1. Obviously,

$$\text{if } \sigma \in \Sigma_1, \text{ then } |\delta^p(\sigma)|=1 \text{ for all } p \qquad \text{(i)}$$

and

$$\text{if } \sigma \in \Sigma_1, \text{ then } \sigma \text{ is ultimately periodic.} \qquad \text{(ii)}$$

One can determine $\delta^n(\sigma)$ for each $\sigma \in \Sigma$, and then see whether or not $\delta^n(\sigma) \in \Sigma_1{}^*$.

a. If $\delta^n(\sigma) \in \Sigma_1{}^*$, then σ is ultimately periodic; for, if $\delta^n(\sigma)=\sigma_1 \ldots \sigma_m$ and each σ_i is ultimately periodic with period π_i, then

$$\delta^{n+\pi_1 \cdot \pi_2 \ldots \pi_m}(\sigma)=\delta^n(\sigma).$$

b. If $\delta^n(\sigma) \notin \Sigma_1{}^*$, then $\delta^n(\sigma)= \ldots \tau_n \ldots$ for some $\tau_n \in \Sigma_2$.
Consider τ_n with its ancestors: $\tau_0(=\sigma), \tau_1, \ldots, \tau_{n-1}, \tau_n$.
According to (i), $\tau_i \in \Sigma_2$ for each τ_i.

As $\# \Sigma_2 \leqslant \# \Sigma = n$, at least two of the τ_i's must be equal, say, τ_k and τ_{k+j}.

Then $\delta^j(\tau_k) = w_1 \tau_{k+j} w_2 = w_1 \tau_k w_2$.

$\tau_k \in \Sigma_2$, so w_1 and w_2 cannot both be empty. If one considers $\delta^{2j}(\tau_k)$, $\delta^{3j}(\tau_k)$ etc., one observes that the offspring of τ_k (and of σ) strictly increases in length. So σ is not ultimately periodic.

The above argument shows that, for every $\sigma \in \Sigma$, it can be determined whether σ is ultimately periodic. Now, if all letters in w_0 are ultimately periodic, the same holds for w_0 itself; and then, by lemma 2, $L(G)$ is finite. If, on the other hand, not all letters in w_0 are ultimately periodic, then obviously $L(G)$ is infinite.

2. Consider some σ_i in w_0. Since $L(G)$ is finite, σ_i is ultimately periodic. As shown in 1^a (and in the same notation),

$$\delta^{n+\pi_1 \pi_2 \ldots \pi_m}(\sigma_i) = \delta^n(\sigma_i).$$

By simple concatenation one obtains a similar relation for w_0 instead of σ_i. Taking into account that only different π_i's have to be included in the left-hand exponent, each of them no larger than n, it is clear that this exponent can always be made smaller than $n^n + n$. This sets an upper bound to $\# L(G)$.

Propagating sub-systems of D0L-systems

From any D0L-system G a simpler system $\tilde{G}$ can be constructed, which is propagating. The construction is as follows:

Let $G = \langle \Sigma, \delta, w_0 \rangle$ be a D0L-system with $\# \Sigma = n$.

Define an erasing homomorphism χ on $\Sigma \cup \{\Lambda\}$:

$$\chi(\sigma) = \begin{cases} \sigma & \text{if } \delta^n(\sigma) \neq \Lambda \\ \Lambda & \text{if } \delta^n(\sigma) = \Lambda. \end{cases}$$

χ is extended to words over Σ^* by

$$\chi(\sigma * w) = \chi(\sigma) * \chi(w).$$

Now define

$$\Sigma = \{\sigma \in \Sigma : \chi(\sigma) = \sigma\}$$

$$\tilde{\delta} = \chi \circ \delta, \text{ defined on } \tilde{\Sigma}$$

$$\tilde{w} = \chi(w) \text{ for every } w \in \Sigma^*$$

$$\tilde{G} = \langle \tilde{\Sigma}, \tilde{\delta}, \tilde{w} \rangle.$$

The following example illustrates the construction of $\tilde{G}$ from G.

Let $G = \langle \{1, 2, 3, 4\}, \delta, 44231 \rangle$ with

$$\delta(1) = 42 \quad \delta(2) = 3 \quad \delta(3) = \Lambda \quad \delta(4) = 124$$

Determine $\delta^4(\sigma)$ for $\sigma = 1, 2, 3, 4$: $\delta^4(1) \neq \Lambda$, $\delta^4(2) = \Lambda$, $\delta^4(3) = \Lambda$, $\delta^4(4) \neq \Lambda$.

So $\tilde{\Sigma}=\{1, 4\}$, and $\tilde{G}=\langle\{1, 4\}, \tilde{\delta}, 441\rangle$ with

$$\tilde{\delta}(1)=4 \quad \tilde{\delta}(4)=14.$$

Remark: In larger D0L-systems, the computation of $\delta^n(\sigma)$ for all $\sigma \in \Sigma$ can be tedious. An easier way of obtaining $\tilde{G}$ is the construction of intermediate reduced systems. By omitting from Σ those σ for which $\delta(\sigma)=\Lambda$, one obtains Σ' and the corresponding $G'=\langle\Sigma', \delta', w_0'\rangle$. From G', G'' is constructed, and so on until $G^{(k+1)}=G^{(k)}$ for some k. $G^{(k)}$ is then the desired $\tilde{G}$.

I omit the proof that the method is equivalent to the "direct" one. In the example, the procedure goes on as follows:

$G' =\langle\{1, 2, 4\}, \delta', 4421\rangle$ with $\delta'(1)=42$, $\delta'(2)=\Lambda$, $\delta'(4)=124$.
$G'' =\langle\{1, 4\}, \delta'', 441\rangle$ with $\delta''(1)=4$, $\delta''(4)=14$.
$G'''=G''$, so $\tilde{G}=G''$.

The construction of $\tilde{G}$, the propagating "backbone" of G, is essential in the decision procedure for the membership question "$w \in L(G)$?" (it still must be proved, of course, that $\tilde{G}$ is propagating). The procedure is, roughly speaking, the following: it is first determined whether $\tilde{w} \in L(\tilde{G})$. One cannot find a positive answer to that question without obtaining a production tree, and that tree is used to decide whether $w \in L(G)$. As I shall show, the procedure is not uniform, but depends on an intermediate result, namely, the finiteness of $L(\tilde{G})$.

LEMMA 4. $\tilde{G}$ is propagating.

Proof: Let $\sigma \in \tilde{\Sigma}$. By construction of $\tilde{G}$, $\delta^n(\sigma)$ is not empty. Consider an arbitrary letter, say σ', in the word $\delta^n(\sigma)$; this word can then be written as $w'\sigma'w''$ for some $w', w'' \in \Sigma^*$. Consider the line of predecessors of σ': $\sigma=\tau_0, \tau_1, \ldots, \tau_{n-1}, \tau_n=\sigma'$. This line is unique, and $\delta(\tau_i)=a_{i+1}\tau_{i+1}b_{i+1}$ (for some a_{i+1} and $b_{i+1} \in \Sigma^*$) for each i.

As $\# \Sigma=n$, at least two of these τ's are equal, say $\tau_{k+j}=\tau_k$. Then also $\tau_n=\tau_{n-j}$. Hence $\delta(\tau_n)=\delta(\tau_{n-j})\neq\Lambda$.

Now

$$\delta^n \circ \delta(\tau_0)=\delta \circ \delta^n(\tau_0)=\delta(w'\tau_n w'')=\delta(w')\delta(\tau_n)\delta(w'')\neq\Lambda,$$

so $\chi \circ \delta(\sigma)\neq\Lambda$; in other words, $\tilde{\delta}(\sigma)\neq\Lambda$.

This holds for all $\sigma \in \tilde{\Sigma}$; so $\tilde{G}$ is propagating.

LEMMA 5. $\tilde{\delta}(\tilde{w})=\widetilde{\delta(w)}$.

Proof: Let $w=\sigma_1 \ldots \sigma_m$.

$$\begin{aligned}\tilde{\delta}(\tilde{w}) &=\chi \circ \delta(\chi(\sigma_1) \ldots \chi(\sigma_m))\\ &=\chi \circ \delta \circ \chi(\sigma_1) \ldots \chi \circ \delta \circ \chi(\sigma_m).\end{aligned}$$

If $\sigma_i \in \tilde{\Sigma}$, then $\chi \circ \delta \circ \chi(\sigma_i) = \chi \circ \delta(\sigma_i)$.

If $\sigma_i \notin \tilde{\Sigma}$, then $\chi(\sigma_i) = \Lambda \Rightarrow \delta^n(\sigma_i) = \Lambda \Rightarrow \delta^n \circ \delta(\sigma_i) = \Lambda \Rightarrow \chi \circ \delta(\sigma_i) = \Lambda$. On the other hand, $\chi \circ \delta \circ \chi(\sigma_i) = \Lambda$.

So, in both cases, $\chi \circ \delta \circ \chi(\sigma_i) = \chi \circ \delta(\sigma_i)$.

Hence $\tilde{\delta}(\tilde{w}) = \chi \circ \delta(\sigma_1) \ldots \chi \circ \delta(\sigma_m) = \chi \circ \delta(w) = \widetilde{\delta(w)}$.

COROLLARY: If $\widetilde{L(G)}$ is defined as $\{\tilde{w} | w \in L(G)\}$, then $L(\tilde{G}) = \widetilde{L(G)}$.

LEMMA 6. Let $G = \langle \Sigma, \delta, w_0 \rangle$ be a D0L-system, and $L = L(G)$. Then L is finite if and only if $\tilde{L}$ is finite.

Proof: $\Rightarrow$: trivial.

$\Leftarrow$: Let $\# \Sigma = N$; $\max_{\sigma \in \Sigma} |\delta(\sigma)| = K$; $\max_{w \in \tilde{L}} |w| = M$.

Vital and *mortal* letters are letters from $\tilde{\Sigma}$ and $\Sigma - \tilde{\Sigma}$, respectively.

In each word of L, every mortal letter stems from a *first mortal predecessor* (FMP), which is unique.

Since G is determinate, the words of L are produced in a fixed order, so w_q $(= \delta^q(w_0))$ refers to only one word.

If σ is mortal, then, by definition,

$$\delta^N(\sigma) = \Lambda. \tag{i}$$

It follows from (i) that a word stemming from a mortal letter can not be longer than K^{N-1}:

$$(\forall p)(\forall \sigma \in \Sigma - \tilde{\Sigma})(|\delta^p(\sigma)| \leqslant K^{N-1}). \tag{ii}$$

Now consider an arbitrary word $w_q \in L$ and an arbitrary mortal letter in it. According to (i), the FMP of this letter can only belong to one of the $N-1$ words preceding w_q. By its definition, each FMP is the immediate successor of a vital letter. Since a word contains no more than M vital letters, the next word contains no more than $M \cdot K$ FMPs (of letters in w_q). All w_q's FMPs belong to the $N-1$ words preceding w_q, so the set of w_q's FMPs contains no more than $(N-1)MK$ letters. According to (ii), each FMP can only produce strings of length K^{N-1} or less, therefore the number of mortal letters in w_q is certainly not larger than $(N-1)MK \cdot K^{N-1}$.

By addition,

$$|w_q| \leqslant (N-1)MK^N + M; \tag{iii}$$

so L is finite.

Now it becomes possible to make a rough estimate of $\# L$ if L is finite. Again, put $\# \Sigma = N$, $\max_{\tilde{w} \in \tilde{L}} |\tilde{w}| = M$, and $\max_{\sigma \in \Sigma} |\delta(\sigma)| = K$.

By lemma 3, $\# \tilde{L} \leqslant N^N + N$.

If one puts $M = |\tilde{w}_0| \cdot K^{N^N+N}$, then

$$\tilde{w} \leqslant M \text{ for all } \tilde{w} \in \tilde{L}.$$

By (iii) in the proof of lemma 6,

$$|w_q| \leqslant (N-1)MK^N + M.$$

Hence

$$\# L \leqslant N^{(N-1)MK^N+M}.$$

This means that someone writing out a finite $L(G)$ has the advantage of knowing an upper bound to the amount of work he has to do.

THEOREM: In D0L-systems the membership question is effectively decidable.

Proof: Let $G = \langle \Sigma, \delta, w_0 \rangle$ be a D0L-system, with $\# \Sigma = n$.

Given $w \in \Sigma^*$, the problem is: $w \in L(G)$?

Construct $\tilde{G}$ as indicated before. By lemma 4, $\tilde{G}$ is a PD0L-system.

By lemma 1 (corollary), it can be determined whether $\tilde{w} \in \tilde{L}$.

(i) If $\tilde{w} \notin \tilde{L}$, then $w \notin L$ (lemma 5, corollary).

(ii) If $\tilde{w} \in \tilde{L}$, then together with this answer an m was found for which $\tilde{\delta}^m(\tilde{w}_0) = \tilde{w}$.

Now determine whether $\tilde{L}$ is finite, with the procedure given in lemma 3.

– If $\tilde{L}$ is finite, then, by lemma 6, L is finite, and the question can be decided by writing out the whole of L.

– If $\tilde{L}$ is infinite, determine whether $\delta^m(w_0) = w$.

If so, then of course $w \in L$.

If not, then $w \notin L$. For, if there existed a $p \neq m$ for which $\delta^p(w_0) = w$, then also $\tilde{\delta}^p(\tilde{w}_0) = \tilde{w}$ (lemma 5).

Together with $\tilde{\delta}^m(\tilde{w}_0) = \tilde{w}$, this would imply that $\tilde{L}$ was finite (lemma 2). But, by assumption, $\tilde{L}$ is infinite.

The following example shows the decision procedure as well as the reason why a distinction has to be made between finite and infinite $\tilde{L}$.

EXAMPLE. $G = \langle \{1, 2, 3, 4, 5\}, \delta, 3142 \rangle$ with

$$\delta(1) = 24 \quad \delta(2) = 3 \quad \delta(3) = 15 \quad \delta(4) = 5 \quad \delta(5) = \Lambda.$$

Question: is $w = 152435 \in L(G)$?

Solution: Construct $\tilde{G}$.

$$\delta^5(1) \neq \Lambda,\ \delta^5(2) \neq \Lambda,\ \delta^5(3) \neq \Lambda,\ \delta^5(4) = \Lambda,\ \delta^5(5) = \Lambda;$$

so $\tilde{G} = \langle \{1, 2, 3\}, \tilde{\delta}, 321 \rangle$ with $\tilde{\delta}(1) = 2,\ \tilde{\delta}(2) = 3,\ \tilde{\delta}(3) = 1.$

$$\tilde{w} \in \tilde{L}, \text{ for } \tilde{\delta}(w_0) = 123 = \tilde{w}.$$

$\tilde{L}$ is finite; $\tilde{w}_0 = 312 \underset{\tilde{G}}{\Rightarrow} 123 \underset{\tilde{G}}{\Rightarrow} 231 \underset{\tilde{G}}{\Rightarrow} 312$ etc.

Now if $\tilde{L}$ were infinite, it would be sufficient to take the equality $\tilde{\delta}(\tilde{w}_0)=\tilde{w}$ (which established the relation $\tilde{w} \in \tilde{L}$) and determine whether the corresponding $\delta(w_0)=w$ also holds. The result is negative, and for an infinite $\tilde{L}$ the conclusion $w \notin L$ would be valid. But here it is not, as is shown by $\delta^4(w_0)=152435=w$.

Conclusion

While reading the manuscript of this paper, dr. G. Rozenberg put forward the conjecture that a D0L-language could be accepted by a linear bounded automaton. This proved to be right for D0L-languages and also for the larger class of 0L-languages. As a consequence, 0L-languages are context-sensitive, and the main theorem of this paper is thereby reduced to a corollary of the new result [5]. Still, this paper gives a direct decision procedure which is unusual in not being uniform, and the reduction of a D0L-system to a PD0L-system may also be of some interest.

Acknowledgement

I wish to thank Professor A. Lindenmayer and Dr. D. van Dalen for their stimulating help and useful criticism.

Filosofisch Instituut
Rijksuniversiteit, Utrecht

REFERENCES

1. Dalen, D. van, A note on some systems of Lindenmayer *Mathematical Systems Theory*, 5, 128–140 (1971).
2. Herman, G. T., The computing ability of a developmental model for filamentous organisms, *Journal of Theoretical Biology*, 25, 421–435 (1969).
3. Lindenmayer, A., Mathematical models for cellular interactions in development, I, II, *Journal of Theoretical Biology*, 18, 280–315 (1968).
4. ———, Developmental systems without cellular interactions, their languages and grammars J. Theor. Biol. (1971) 30, 455–484 (1970).
5. Rozenberg, G. and P. G. Doucet, On 0L-languages. *Information and Control* (in press) (1971).