Kybernetika

Jiří Pik

A note on connections between 0L-languages and context-free languages

Kybernetika, Vol. 15 (1979), No. 2, (136)--140

Persistent URL: http://dml.cz/dmlcz/124479

KYBERNETIKA — VOLUME 15 (1979), NUMBER 2

A Note on Connections between 0L-Languages and Context-free Languages

JIŘÍ PIK

Connections between the families of 0L-languages, context-free languages, and languages of sentential forms are considered; a possible generalization of the concept of 0L-systems in relation to context-free languages is presented.

1. INTRODUCTION

L-systems were introduced by A. Lindenmayer for the purposes of theoretical biology in 1968. Now, theory of these systems, often called developmental or Lindenmayer systems, forms an important part of the theory of formal languages. The main feature of L-systems consists in the simultaneous application of productions to all symbols of the processed word. Context-free languages and their theory represent a "classical" part of formal languages theory, they belong to the Chomsky hierarchy and they are also called languages of type 2. A language of sentential forms is a set of all words generated by a context-free grammar, i.e., a set of words over the union of terminals and nonterminals of this grammar obtained from the initial symbol by applications of production rules.

This paper contains a remark on the comparison of the family of 0L-languages with the family of context-free languages, and with the family of languages of sentential forms. The concept of 0L-systems can be a subject of some generalizations, a possible one in relation to the family of context-free languages is also considered.

We assume the reader is familiar with the elements of formal languages theory including the theory of 0L-systems, e.g., see Salomaa [4]; here we give only some definitions and theorems relevant for this paper.

An alphabet is an arbitrary finite nonempty set of symbols. If Σ is an alphabet, then Σ^* denotes the set of all words over Σ including the word λ consisting of no symbols. A language is any set of words over an alphabet.

Let A and B be two sets, then $A \subseteq B$ denotes inclusion of A in B, $A \subset B$ denotes strict inclusion of A in B, and $A \nsubseteq B$ denotes the negation of $A \subseteq B$. Two sets A and B are incomparable if $A \nsubseteq B$ and $B \nsubseteq A$.

For $a \in \Sigma$, let $\tau(a)$ denote a set of words over Σ, $\tau(a) \subset \Sigma^*$. Let τ be defined on Σ^* by $\tau(\lambda) = \{\lambda\}$ and $\tau(x) = \tau(x_1), \ldots, \tau(x_k)$ for $x = x_1, \ldots, x_k$, $x \in \Sigma^*$, $x_1, \ldots \ldots, x_k \in \Sigma$ for each $k \geqq 1$. Further, τ is extended to languages over Σ by defining $\tau(A) = \bigcup_{x \in A} \tau(x)$ for all $A \subseteq \Sigma^*$. Then τ is a substitution on Σ^*. A substitution τ is called context-free if $\tau(a)$ is a context-free language for each a in Σ.

Definition. Let Σ be a fixed alphabet. The family of 0L-languages $\mathcal{O}_\Sigma$ is the set

$$\mathcal{O}_\Sigma = \{M : M \subseteq \Sigma^* \text{ and there exists a 0L-system } H \text{ such that } L(H) = M\} .$$

Definition. Let Σ be a fixed alphabet. The family of context-free languages $\mathcal{L}_\Sigma^{CF}$ is the set

$$\mathcal{L}_\Sigma^{CF} = \{M : M \subseteq \Sigma^* \text{ and there exists a context-free grammar } G \text{ such that } L(G) = M\} .$$

Theorem 2.1. (Rozenberg and Doucet [3]) For every alphabet Σ, $\mathcal{O}_\Sigma$ and $\mathcal{L}_\Sigma^{CF}$ are incomparable, but not disjoint.

Theorem 2.2. (Rozenberg and Doucet [3]) If $G = (V, \Sigma, P, S)$ is a context-free grammar, then there exists a 0L-system H such that $L(H) \cap \Sigma^* = L(G)$.

Theorem 2.3. (Rozenberg and Doucet [3]) Let $H = (\Sigma, P, \sigma)$ be a 0L-system with the property that (a, a) is in P for every $a \in \Sigma$. Then $L(H)$ is a context-free language.

Theorem 2.4. (Král [2]) Let $A \subseteq \Sigma^*$ be a context-free language and let τ be a context-free substitution on Σ^*. Let $\tau(A) \subseteq \Sigma^*$ and $a \in \tau(a)$ for each $a \in \Sigma$. Set $\tau^1(A) = \tau(A)$ and for $n \geqq 1$, let $\tau^{n+1}(A) = \tau(\tau^n(A))$. Then $\lim_{n \to \infty} \tau^n(A)$ is a context-free language. (The limit is taken in the sense of set theory. The inclusion $\tau^{n+1}(A) \supseteq \tau^n(A)$ holds for every $n \geqq 1$, and $\lim_{n \to \infty} \tau^n(A) = \bigcup_n \tau^n(A)$.)

For any context-free grammar G a word over the union of the terminal and nonterminal alphabets derived from the initial symbol by applications of productions of this grammar G is called a sentential form of G. The language of sentential forms of G, $L_{SF}(G)$, contains all sentential forms of G. A language L is called an SF-language if and only if there exists a context-free grammar G such that $L = L_{SF}(G)$.

Definition. Let Σ be a fixed alphabet. The family of SF-languages $\mathscr{L}_\Sigma^{SF}$ is the set

$$\mathscr{L}_\Sigma^{SF} = \{M : M \subseteq \Sigma^* \text{ and } M \text{ is an SF-language}\} .$$

Theorem 2.5. (Salomaa [5]) Every SF-language is generated by a 0L-system.

3. 0L-LANGUAGES AND CONTEXT-FREE LANGUAGES

Rozenberg and Doucet in [3] considered a generalization of the 0L-system, they called it a 0L-system with context-free root.

Definition. A system $H = (\Sigma, P, \Theta)$ will be called a 0L-system with context-free root if Σ and P are defined as in 0L-systems, and Θ is a context-free language over Σ, serving as a set of axioms. The relations $\Rightarrow_H$ and $\Rightarrow_H^*$ are defined as in 0L-systems, and the language generated by H is

$$L(H) = \{x : x \in \Sigma^* \text{ and there exists } \sigma, \sigma \in \Theta, \text{ such that } \sigma \Rightarrow_H^* x\} .$$

The concept of the 0L-system can be generalized also in another way than by the change of the set of axioms, namely by the change of the productions of this system. We can give the following definition.

Definition. A system $H = (\Sigma, Q, \Theta)$ will be called a 0L-system with context-free root and context-free productions if Σ is defined as in 0L-systems, Θ is a context-free language over Σ, serving as a set of axioms, and the set of productions Q is a finite subset of $\Sigma \times \mathscr{L}_\Sigma^{CF}$, such that for every $a \in \Sigma$ there exists $\alpha \in \mathscr{L}_\Sigma^{CF}$ with the property $(a, \alpha) \in Q$.

The relation $\Rightarrow_H$ for 0L-systems with context-free root and context-free productions is defined as follows.

Definition. Let $H = (\Sigma, Q, \Theta)$ be a 0L-system with context-free root and context-free productions, let $x \in \Sigma^+$, $x = x_1, \ldots, x_m$ with $m \geqq 1$ and $x_j \in \Sigma$ for $j = 1, \ldots, m$, let $y \in \Sigma^*$. Then $x \Rightarrow_H y$ if and only if there exist $p_1, \ldots, p_m$ in Q such that $p_j = (x_j, \alpha_j)$ and $y = y_1, \ldots, y_m$, where $y_j \in \alpha_j$ for every $j = 1, \ldots, m$.

Now, the relation $\Rightarrow_H^*$ and the language of this 0L-system are defined in the usual way as in 0L-systems. The language, generated by this 0L-system H is

$$L(H) = \{x : x \in \Sigma^* \text{ and there exists } \sigma, \sigma \in \Theta, \text{ such that } \sigma \Rightarrow_H^* x\} .$$

The next theorem holds for these 0L-systems.

Theorem 3.1. Let $H = (\Sigma, Q, \Theta)$ be a 0L-system with context-free root and context-free productions. If for every $a \in \Sigma$ there exists α such that $(a, \alpha) \in Q$ and $a \in \alpha$, then $L(H)$ is a context-free language.

Proof. The set Θ is a context-free language as well as the set α, $\alpha \subseteq \Sigma^*$, $(a, \alpha) \in Q$, for every $a \in \Sigma$. In fact, the application of productions in 0L-systems is a repeated context-free substitution. The conditions of Theorem 2.4 are satisfied and the language generated by a 0L-system with context-free root and context-free productions is therefore context-free.

The families of 0L-languages and context-free languages are incomparable, but not disjoint. The family of SF-languages is a proper subset of the family of context-free languages. For the family of SF-languages we can formulate the following corollary which deals with its connection with the family of 0L-languages.

We shall use the symbol $\mathcal{O}_\Sigma^{CF}$ for the intersection of $\mathcal{O}_\Sigma$ and $\mathcal{L}_\Sigma^{CF}$, $\mathcal{O}_\Sigma^{CF} = \mathcal{O}_\Sigma \cap \mathcal{L}_\Sigma^{CF}$.

Corollary 3.1. For every alphabet Σ, $\mathcal{O}_\Sigma^{CF} \supset \mathcal{L}_\Sigma^{SF}$.

Proof. The inclusion $\mathcal{O}_\Sigma^{CF} \supseteq \mathcal{L}_\Sigma^{SF}$ follows from Theorem 2.5. (The construction of the required 0L-system H for an SF-language generated by a grammar G is very simple. The alphabet Σ of this 0L-system H contains all terminal and nonterminal symbols of the grammar G, as the axiom of H we take the initial symbol of G, and to obtain the productions of this system H we add the productions (a, a), for every $a \in \Sigma$, to those of the grammar G.)

To prove the strict inclusion let us consider an unary 0L-language which is not deterministic and not propagating, e.g., $L(H) = \{a^{2n} : n \geqq 0\}$. This language is evidently context-free (it is even regular), but it is not an SF-language. (We cannot find a context-free grammar, where the union of the sets of terminals and nonterminals is $\{a\}$ and the generated language is $L(H)$.) It follows that $\mathcal{O}_\Sigma^{CF} \neq \mathcal{L}_\Sigma^{SF}$, which completes the proof.

Clearly, Theorem 2.2 is a consequence of Theorem 2.5. Indeed, if $G = (V, \Sigma, P, S)$ is a context-free grammar, then $L(G) = L_{SF}(G) \cap \Sigma^*$. By Theorem 2.5, there exists a 0L-system H such that $L_{SF}(G) = L(H)$. This implies Theorem 2.2.

Some special subclasses of the class $\mathcal{O}_\Sigma^{CF}$ are included in $\mathcal{L}_\Sigma^{SF}$. The class of languages generated by 0L-systems (Σ, P, σ) with $(a, a) \in P$ for every $a \in \Sigma$ has this property.

(Received December 19, 1977.)

REFERENCES

[1] G. T. Herman, G. Rozenberg: Developmental Systems and Languages. North-Holland Publishing Company, Amsterdam — Oxford, American Elsevier Publishing Company, Inc., New York 1975.

[2] J. Král: A Modification of a Substitution Theorem and Some Necessary and Sufficient Conditions for Sets to Be Context-Free. Mathematical Systems Theory *4* (1970), 2, 129—139.

[3] G. Rozenberg, P. G. Doucet: On 0L-Languages. Informations and Control *19* (1971), 4, 302—318.

[4] A. Salomaa: Formal Languages. Academic Press, New York—London 1973.

[5] A. Salomaa: On Sentential Forms of Context-Free Grammars. Acta Informatica *2* (1973), 1, 40—49.

Ing. Jiří Pik, Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.