POWER OF MULTIGENERATIVE GRAMMAR SYSTEMS

Ing. Roman LUKÁŠ, Doctoral Degree Programme (3)
Dept. of Information Systems, FIT, BUT
E-mail: lukas@fit.vutbr.cz

Supervised by: Prof. Alexander Meduna

ABSTRACT

This paper presents new models for all recursive enumerable languages. These models are based on multigenerative grammar systems that simultaneously generate several strings in a parallel way. The components of these models are context-free grammars, working in a leftmost way. The rewritten nonterminals are determined by a finite set of nonterminal sequences.

1 INTRODUCTION

The formal language theory has recently intensively investigated various grammar systems (see [1], [2], [8]), which consist of several cooperating components, usually represented by grammars. Although this variaty is extremely broad, all these grammar systems always make a derivation that generates a single string. In this paper, however, we introduce grammar systems that simultaneously generate several strings, which are subsequently composed in a single string by some common string operation, such as concatenation.

More precisely, for a positive integer n, an n-multigenerative grammar system discussed in this paper works with n context-free grammatical components in a leftmost way—that is, in every derivation step, each of these components rewrites the leftmost nonterminal occurring in its current sentential form. These n leftmost derivations are controled n-tuples of nonterminals or rules. Under a control like this, the grammar system generates n strings, out of which the strings that belong to the generated language are made by some basic operations. Specifically, these operations include union, concatenation and a selection of the string generated by the first component.

In this paper, we prove that all the multigenerative grammar systems under discussion characterize the family of recursively enumerable languages. Besides this fundamental result, we give several transformation algorithms of these multigenerative grammar systems.

2 PRELIMINARIES

This paper assumes that the reader is familiar with the formal language theory (see [4]).

For a set, Q, $card(Q)$ denotes the cardinality of Q. For an alphabet, V, V^* represents the free monoid generated by V under the operation of concatenation. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$; algebraically, V^+ is thus the free semigroup generated by V under the operation of concatenation. For every $w \in V^*$, $|w|$ denotes the length of w.

A *context-free grammar* is a quadruple, $G = (N, T, P, S)$, where N and T are two disjoint alphabets. Symbols in N and T are referred to as *nonterminals* and *terminals*, respectively, and $S \in N$ is the *start symbol* of G. P is a finite set of *rules* of the form $A \rightarrow x$, where $A \in N$ and $x \in (N \cup T)^*$. To declare that a label r denotes the rule, this is written as $r: A \rightarrow x$. For every $A \rightarrow x \in P$ and $u, v \in (N \cup T)^*$, write $uAv \Rightarrow uxv$. Let $\Rightarrow^*$ denote the transitive-reflexive closure of $\Rightarrow$. The *language of* G, $L(G)$, is defined as $L(G) = \{w: S \Rightarrow^* w \text{ in } G, \text{ for some } w \in T^*\}$.

3 DEFINITIONS

***Definition 3.1*:** An *n-multigenerative nonterminal-synchronized grammar system* (n-MGN) is an n+1 tuple,

$$\Gamma = (G_1, G_2, \ldots, G_n, Q),$$

where $G_i = (N_i, T_i, P_i, S_i)$ is a context-free grammar for each $i = 1, \ldots, n$, and Q is a finite set of n-tuples of the form $(A_1, A_2, \ldots, A_n)$, where $A_i \in N_i$ for all $i = 1, \ldots, n$. Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be an n-MGN. Then, a *sentential n-form* of n-MGN is an n-tuple of the form $\chi = (x_1, x_2, \ldots, x_n)$, where $x_i \in (N_i \cup T_i)^*$ for all $i = 1, \ldots, n$. Let $\chi = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$ and $\bar{\chi} = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$ be two sentential n-form, where $A_i \in N_i$, $u_i \in T_i^*$, and $v_i, x_i \in (N_i \cup T_i)^*$ for all $i = 1, \ldots, n$. Let $A_i \rightarrow x_i \in P_i$ for all $i = 1, \ldots, n$ and $(A_1, A_2, \ldots, A_n) \in Q$. Then χ directly derives $\bar{\chi}$ in Γ, denoted by $\chi \Rightarrow \bar{\chi}$. In the standard way, we generalize $\Rightarrow$ to $\Rightarrow^k$, $k \geq 0$, $\Rightarrow^+$, and $\Rightarrow^*$. The *n-language of* Γ, n-$L(\Gamma)$, is defined as

$$n\text{-}L(\Gamma) = \{(w_1, w_2, \ldots, w_n): (S_1, S_2, \ldots, S_n) \Rightarrow^* (w_1, w_2, \ldots, w_n), w_i \in T_i^* \text{ for all } i = 1, \ldots, n\}.$$

The *language generated by* Γ *in the union mode*, $L_{union}(\Gamma)$, is defined as

$$L_{union}(\Gamma) = \{w: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma), w \in \{w_i: i = 1, \ldots, n\}\}.$$

The *language generated by* Γ *in the concatenation mode*, $L_{conc}(\Gamma)$, is defined as

$$L_{conc}(\Gamma) = \{w_1w_2\ldots w_n: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\}$$

The *language generated by* Γ *in the first mode*, $L_{first}(\Gamma)$, is defined as

$$L_{first}(\Gamma) = \{w_1: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\}$$

***Example*:** $\Gamma = (G_1, G_2, Q)$, where $G_1 = (\{S_1, A_1\}, \{a, b, c\}, \{S_1 \rightarrow aS_1, S_1 \rightarrow aA_1, A_1 \rightarrow bA_1c, A_1 \rightarrow bc\}, S_1)$, $G_2 = (\{S_2, A_2\}, \{d\}, \{S_2 \rightarrow S_2A_2, S_2 \rightarrow A_2, A_2 \rightarrow d\}, S_2)$, $Q = \{(S_1, S_2), (A_1, A_2)\}$ is a 2-multigenerative nonterminal-synchronized grammar system. Notice that $2\text{-}L(\Gamma) = \{(a^nb^nc^n, d^n): n \geq 1\}$, $L_{union}(\Gamma) = \{a^nb^nc^n: n \geq 1\} \cup \{d^n: n \geq 1\}$, $L_{conc}(\Gamma) = \{a^nb^nc^nd^n: n \geq 1\}$, and $L_{first}(\Gamma) = \{a^nb^nc^n: n \geq 1\}$.

***Definition 3.2*:** An *n-multigenerative rule-synchronized grammar system* (n-MGR) is n+1 tuple

$$\Gamma = (G_1, G_2, \ldots, G_n, Q),$$

where $G_i = (N_i, T_i, P_i, S_i)$ is a context-free grammar for each $i = 1, \ldots, n$, and Q is a finite set of n-tuples of the form $(p_1, p_2, \ldots, p_n)$, where $p_i \in P_i$ for all $i = 1, \ldots, n$. A sentential n-form for n-MGR is defined as the sentential n-form for an n-MGN. Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be an n-MGR. Let $\chi = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$ and $\bar{\chi} = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$ are two sentential n-form, where $A_i \in N_i$, $u_i \in T_i^*$, and $v_i, x_i \in (N_i \cup T_i)^*$ for all $i = 1, \ldots, n$. Let p_i: $A_i \to x_i \in P_i$ for all $i = 1, \ldots, n$ and $(p_1, p_2, \ldots, p_n) \in Q$. Then χ directly derives $\bar{\chi}$ in Γ, denoted by $\chi \Rightarrow \bar{\chi}$. An n-language for any n-MGR is defined as the n-language for any n-MGN, and a language generated by n-MGN in the X mode, for each $X \in \{union, conc, first\}$, is defined as the language generated by n-MGR in the X mode.

***Example*:** $\Gamma = (G_1, G_2, Q)$, where $G_1 = (\{S_1, A_1\}, \{a, b, c\}, \{\mathbf{1}: S_1 \to aS_1, \mathbf{2}: S_1 \to aA_1, \mathbf{3}: A_1 \to bA_1c, \mathbf{4}: A_1 \to bc\}, S_1)$, $G_2 = (\{S_2\}, \{d\}, \{\mathbf{1}: S_2 \to S_2S_2, \mathbf{2}: S_2 \to S_2, \mathbf{3}: S_2 \to d\}, S_2)$, $Q = \{(\mathbf{1}, \mathbf{1}), (\mathbf{2}, \mathbf{2}), (\mathbf{3}, \mathbf{3}), (\mathbf{4}, \mathbf{3})\}$, is 2-multigenerative rule-synchronized grammar system. Notice that $2\text{-}L(\Gamma) = \{(a^nb^nc^n, d^n): n \geq 1\}$, $L_{union}(\Gamma) = \{a^nb^nc^n: n \geq 1\} \cup \{d^n: n \geq 1\}$, $L_{conc}(\Gamma) = \{a^nb^nc^nd^n: n \geq 1\}$, and $L_{first}(\Gamma) = \{a^nb^nc^n: n \geq 1\}$.

4 RESULTS

***Algorithm 4.1*:** Conversion of n-MGN to n-MGR

- ***Input:*** n-MGN $\Gamma = (G_1, G_2, \ldots, G_n, Q)$
- ***Output:*** n-MGR $\bar{\Gamma} = (G_1, G_2, \ldots, G_n, \bar{Q})$; $n\text{-}L(\Gamma) = n\text{-}L(\bar{\Gamma})$
- ***Method:***

  Let $G_i = (N_i, T_i, P_i, S_i)$ for all $i = 1, \ldots, n$, then:

  $\bar{Q} := \{(A_1 \to x_1, A_2 \to x_2, \ldots, A_n \to x_n): A_i \to x_i \in P_i$ for all $i = 1, \ldots, n$, and $(A_1, A_2, \ldots, A_n) \in Q\}$

***Algorithm 4.2*:** Conversion of n-MGR to n-MGN

- ***Input:*** n-MGR $\Gamma = (G_1, G_2, \ldots, G_n, Q)$
- ***Output:*** n-MGN $\bar{\Gamma} = (\bar{G}_1, \bar{G}_2, \ldots, \bar{G}_n, \bar{Q})$; $n\text{-}L(\Gamma) = n\text{-}L(\bar{\Gamma})$
- ***Method:***

  Let $G_i = (N_i, T_i, P_i, S_i)$ for all $i = 1, \ldots, n$, then:

  $\bar{G}_i = (\bar{N}_i, T_i, \bar{P}_i, S_i)$ for all $i = 1, \ldots, n$, where:

  $\bar{N}_i := \{<A, x>: A \to x \in P_i\} \cup \{S_i\}$,

  $\bar{P}_i := \{<A, x> \to y: A \to x \in P_i, y \in \tau_i(x)\} \cup \{S_i \to y: y \in \tau_i(S_i)\}$,

  where τ_i is a substitution from $N_i \cup T_i$ to $\bar{N}_i \cup T_i$ defined as:

  $\tau_i(a) = \{a\}$ for all $a \in T_i$; $\tau_i(A) = \{<A, x>: A \to x \in P_i\}$ for all $A \in N_i$.

  $\bar{Q} := \{(<A_1, x_1>, <A_2, x_2>, \ldots, <A_n, x_n>): (A_1 \to x_1, A_2 \to x_2, \ldots, A_n \to x_n) \in Q\} \cup \{(S_1, S_2, \ldots, S_n)\}$

***Claim 4.3*:** Let Γ be any n-MGN, let $\overline{\Gamma}$ be any n-MGR and let $n\text{-}L(\Gamma) = n\text{-}L(\overline{\Gamma})$. Then, $L_X(\Gamma) = L_X(\overline{\Gamma})$, for each $X \in \{union, conc, first\}$.

Proof:

I. We prove that $L_{union}(\Gamma) = L_{union}(\overline{\Gamma})$: $L_{union}(\Gamma) = \{w: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma), w \in \{w_i: i = 1, \ldots, n\}\} = \{w: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\overline{\Gamma}), w \in \{w_i: i = 1, \ldots, n\}\} = L_{union}(\overline{\Gamma})$.

II. We prove that $L_{conc}(\Gamma) = L_{conc}(\overline{\Gamma})$: $L_{conc}(\Gamma) = \{w_1w_2\ldots w_n: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\} = \{w_1w_2\ldots w_n: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\overline{\Gamma})\} = L_{conc}(\overline{\Gamma})$.

III. We prove that $L_{first}(\Gamma) = L_{first}(\overline{\Gamma})$: $L_{first}(\Gamma) = \{w_1: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\} = \{w_1: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\overline{\Gamma})\} = L_{first}(\overline{\Gamma})$.

***Corollary 4.4*:** The class of languages generated by n-MGN in the X mode, where $X \in \{union, conc, first\}$ is equivalent with the class of language generated by n-MGR in the X mode.

***Proof*:** This corollary follows from Algorithm 4.1, Algorithm 4.2 and Claim 4.3.

***Theorem 4.5*:** For every recursive enumerable language L over an alphabet T there exists a 2-MGR, $\Gamma = ((\overline{N}_1, T, \overline{P}_1, S_1), (\overline{N}_2, T, \overline{P}_2, S_2), Q)$ such that:

1) $\{w: (S_1, S_2) \Rightarrow^* (w, w)\} = L$,
2) $\{w_1w_2: (S_1, S_2) \Rightarrow^* (w_1, w_2), w_1, w_2 \in T^*, w_1 \neq w_2\} = \varnothing$.

Proof: Recall that for every recursive enumerable language L over an alphabet T there exist two context-free grammars $G_1 = (N_1, \overline{T}, P_1, S_1)$, $G_2 = (N_2, \overline{T}, P_2, S_2)$ and homomorphism h: from $\overline{T}$ to T^* such that $L = \{h(x) : x \in L(G_1) \cap L(G_2)\}$. (see Theorem 10.3.1 in [3]). Furthermore, for every context-free grammar, there exists an equivalent context-free grammar in Greibach normal form (see Section 5.1.4.2 in [4]). Hence, without lost of generality, we can assume that G_1 and G_2 are in Greibach normal form. Construct a 2-MGR $\Gamma = (G_1, G_2, Q)$, where:

$G_1 = (\overline{N}_1, T, \overline{P}_1, S_1)$, where $\overline{N}_1 = N_1 \cup \{\bar{a} : a \in \overline{T}\}$, $\overline{P}_1 = \{A \to \bar{a}x : A \to ax \in P_1, a \in \overline{T}, x \in N_1^*\} \cup \{\bar{a} \to h(a): a \in \overline{T}\}$

$G_2 = (\overline{N}_2, T, \overline{P}_2, S_2)$, where $\overline{N}_2 = N_2 \cup \{\bar{a} : a \in \overline{T}\}$, $\overline{P}_2 = \{A \to \bar{a}x : A \to ax \in P_2, a \in \overline{T}, x \in N_2^*\} \cup \{\bar{a} \to h(a): a \in \overline{T}\}$

$Q = \{(A_1 \to \bar{a}x_1, A_2 \to \bar{a}x_2): A_1 \to \bar{a}x_1 \in \overline{P}_1, A_2 \to \bar{a}x_2 \in \overline{P}_2, a \in \overline{T}\} \cup \{(\bar{a} \to h(a), \bar{a} \to h(a)): a \in \overline{T}\}$

***Theorem 4.6*:** For every recursive enumerable language L over an alphabet T there exists a 2-MGR, $\Gamma = (G_1, G_2, Q)$ such that: $L_{union}(\Gamma) = L$.

***Proof*:** Let $\Gamma = \overline{\Gamma}$, where $\overline{\Gamma} = ((N_1, T, P_1, S_1), (N_2, T, P_2, S_2), Q)$ is a MGR from Theorem 4.5. Then, $L_{union}(\Gamma) = \{w: (S_1, S_2) \Rightarrow^* (w_1, w_2), w_i \in T^*$ for $i = 1, 2, w \in \{w_i: i = 1, 2\}\} = \{w: (S_1, S_2) \Rightarrow^* (w, w), w \in T^*\} \cup \{w: (S_1, S_2) \Rightarrow^* (w_1, w_2), w_i \in T^*$ for $i = 1, 2, w \in \{w_i: i = 1, 2\}, w_1 \neq w_2\} = \{w: (S_1, S_2) \Rightarrow^* (w, w), w \in T^*\} \cup \varnothing = \{w: (S_1, S_2) \Rightarrow^* (w, w), w \in T^*\} = L$

***Theorem 4.7*:** For every recursive enumerable language L over an alphabet T there exists a 2-MGR, $\Gamma = (G_1, G_2, Q)$ such that: $L_{first}(\Gamma) = L$.

***Proof*:** Let $\Gamma = \overline{\Gamma}$, where $\overline{\Gamma} = ((N_1, T, P_1, S_1), (N_2, T, P_2, S_2), Q)$ is a MGR from Theorem 4.5.

Then, $L_{first}(\Gamma) = \{w_1: (S_1, S_2) \Rightarrow^* (w_1, w_2), w_i \in T^* \text{ for } i = 1, 2\} = \{w: (S_1, S_2) \Rightarrow^* (w, w), w \in T^*\} \cup \{w_1: (S_1, S_2) \Rightarrow^* (w_1, w_2), w_i \in T^* \text{ for } i = 1, 2, w_1 \neq w_2\} = \{w: (S_1, S_2) \Rightarrow^* (w, w), w \in T^*\} \cup \emptyset = \{w: (S_1, S_2) \Rightarrow^* (w, w), w \in T^*\} = L$.

***Theorem 4.8*:** For every recursive enumerable language L over an alphabet T there exists a 2-MGR, $\Gamma = (G_1, G_2, Q)$ such that: $L_{conc}(\Gamma) = L$.

***Proof*:** Let $\overline{\Gamma} = ((N_1, T, P_1, S_1), (N_2, T, P_2, S_2), Q)$ be a MGR from Theorem 4.5. Let $G_1 = (N_1, T, P_1, S_1)$, $G_2 = (N_2, \emptyset, \overline{P_2}, S_2)$, where $\overline{P_2} = \{A \rightarrow g(x): A \rightarrow x \in P_2\}$, where g is a homomorphism from $(N_2 \cup T)$ to N_2 defined as: For all $X \in N_2$: $g(X) = X$, for all $X \in T$: $g(X) = \varepsilon$. We prove that $L_{conc}(\Gamma) = L$.

I. We prove that $L \subseteq L_{conc}(\Gamma)$: Let $w \in L$. Then, there exists a sequence of derivation in $\overline{\Gamma}$ $(S_1, S_2) \Rightarrow^* (w, w)$, thus, there exist a sequence of derivations in Γ $(S_1, S_2) \Rightarrow^* (w, g(w))$. Because $g(a) = \varepsilon$ for all $a \in T$, then $g(w) = \varepsilon$ for all $w \in T^*$. Thus, there exists a sequence of derivations $(S_1, S_2) \Rightarrow^* (w, \varepsilon)$ in Γ. Hence, $w\varepsilon = w \in L_{conc}(\Gamma)$.

II. We prove that $L_{conc}(\Gamma) \subseteq L$: Let $w \in L_{conc}(\Gamma)$. Then, there exist a sequence of derivations $(S_1, S_2) \Rightarrow^* (w, \varepsilon)$ in Γ, because G_2 derives only empty string. $g(x) = \varepsilon$ for all $x \in T^*$, so there exists a sequence of derivation in $\overline{\Gamma}$ of the form: $(S_1, S_2) \Rightarrow^* (w, x)$, where x is any string. Theorem 4.5 implies that $x = w$, thus: $(S_1, S_2) \Rightarrow^* (w, w)$. Thus, $w \in L$.

5 CONCLUSION

Let L(2-MGN$_X$) and L(2-MGR$_X$) denote the language families defined by 2-MGN in the X mode and 2-MGR in the X mode, respectively, where $X \in \{union, conc, first\}$, let L(RE) denote the family of recursive enumerable languages. From the previous results, we obtain L(RE) = L(MGN$_X$) = L(MGR$_X$).

REFERENCES

[1] Csuhaj-Varju, E., Dassow, J., Kelemen, J., Paun, Gh.: Grammar Systems: A Grammatical Approach to Distribution and Cooperation, Gordon and Breach, London, 1994

[2] Dassow, J., Paun, Gh., Rozenberg, G.: Grammar Systems, In Handbook of Formal Languages, Rozenberg, G. and Salomaa, A. (eds.), Volumes 2, Springer, Berlin, 1997.

[3] Harrison, Michael A.: Introduction to Formal Language Theory. Addison-Wesley, London, 1978.

[4] Meduna, A.: Automata and Languages: Theory and Applications, Springer, London, 2000

[5] Meduna, A.: Two-Way Metalinear PC Grammar Systems and Their Descriptional Complexity, Acta Cybernetica, 2003

[6] Paun, Gh., Salomaa, A. and S. Vicolov, S.: On the generative capacity of parallel communicating grammar systems. International Journal of Computer Mathematics 45, 45-59, 1992.

[7] Salomaa, A.: Formal Languages, Academic Press, New York, 1973.

[8] Vaszil, G.: On simulating Non-returning PC grammar systems with returning systems, Theoretical Computer Science (209) 1-2, 319-329, 1998.