MULTIGENERATIVE GRAMMAR SYSTEMS

Ing. Roman LUKÁŠ, Doctoral Degree Programme (1)
Dept. of Information Systems, FIT, BUT
E-mail: lukas@fit.vutbr.cz

Supervised by: Dr. 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 N-MULTIGENERATIVE NONTERMINAL-SYNCHRONIZED GRAMMAR SYSTEM

1.1 BASIC DEFINITION

An *n-multigenerative nonterminal-synchronized grammar system* (MGN) is n+1 tuple

$$\Gamma = (G_1, G_2, \ldots, G_n, Q), \text{ where:}$$

- $G_i = (N_i, T_i, P_i, S_i)$ is a context-free grammar for each $i = 1, \ldots, n$,
- 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 $i = 1, \ldots, n$.

1.2 SENTENTIAL N-FORM

Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be a MGN. Then a *sentential n-form of 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 $i = 1, \ldots, n$.

1.3 DIRECT DERIVATION STEP

Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be a MGN. Let $\chi = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$ and $\chi' = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$ 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 $A_i \rightarrow x_i \in P_i$ for all all $i = 1, \ldots, n$ and $(A_1, A_2, \ldots, A_n) \in Q$. Then χ directly derives χ' in Γ, denoted by $\chi \Rightarrow \chi'$.

1.4 SEQUENCE OF DERIVATION STEPS, PART 1

Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be a MGN.

- Let χ by any sentential n-form of Γ. Γ makes a *zero-step* derivation from χ to χ, which is written as $\chi \Rightarrow^0 \chi$.
- Let there exists a sequence of sentential n-forms $\chi_0, \chi_1, \ldots, \chi_k$ for some $k \geq 1$ such that $\chi_{i-1} \Rightarrow \chi_i$ for all $i = 1, \ldots, k$. Then, Γ makes *n-step derivation* from χ_0 to χ_k, which is written as $\chi_0 \Rightarrow^n \chi_k$.

1.5 SEQUENCE OF DERIVATION STEPS, PART 2

Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be a MGN, let χ and χ' be two sentential n-forms of Γ.

- If there exists $k \geq 1$ so $\chi \Rightarrow^k \chi'$ in Γ, then $\chi \Rightarrow^+ \chi'$,
- If there exists $k \geq 0$ so $\chi \Rightarrow^k \chi'$ in Γ, then $\chi \Rightarrow^* \chi'$.

1.6 N-LANGUAGE

Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be a MGN. The n-language of Γ, $n\text{-}L(\Gamma)$, is defined as:

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

1.7 THREE TYPES OF GENERATED LANGUAGES

- 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 leftmost mode*, $L_{lm}(\Gamma)$, is defined as:

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

1.8 EXAMPLE

$\Gamma = (G_1, G_2, Q)$, where:

- $G_1 = (\{S_1, A_1\}, \{a, b, c\}, \{S_1 \rightarrow aS_1, S_1 \rightarrow aA_1, A_1 \rightarrow bA_1c, A_1 \rightarrow bc\}, S_1)$,
- $G_2 = (\{S_2, A_2\}, \{d\}, \{S_2 \rightarrow S_2A_2, S_2 \rightarrow A_2, A_2 \rightarrow d\}, S_2)$,
- $Q = \{(S_1, A_1), (S_2, A_2)\}$

is a 2-multigenerative nonterminal-synchronized grammar system.

Notice that this system generates following languages in the different modes:

- $L_{union}(\Gamma) = \{a^nb^nc^n: n \geq 1\} \cup \{d^n: n \geq 1\}$,
- $L_{conc}(\Gamma) = \{a^nb^nc^nd^n: n \geq 1\}$,
- $L_{lm}(\Gamma) = \{a^nb^nc^n: n \geq 1\}$.

2 N-MULTIGENERATIVE RULE-SYNCHRONIZED GRAMMAR SYSTEM

2.1 BASIC DEFINITION

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

$$\Gamma = (G_1, G_2, \ldots, G_n, Q), \text{ where:}$$

- $G_i = (N_i, T_i, P_i, S_i)$ is a context-free grammar for each $i = 1, \ldots, n$,
- 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 $i = 1, \ldots, n$.

2.2 SENTENTIAL N-FORM

A sentential n-form for MGR is defined analogically as the sentential n-form for a MGN.

2.3 DIRECT DERIVATION STEP

Let $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ be a MGR. Let $\chi = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$ and $\chi' = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$ 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: $A_i \rightarrow x_i \in P_i$ for all all $i = 1, \ldots, n$ and $(p_1, p_2, \ldots, p_n) \in Q$. Then χ directly derives χ' in Γ, denoted by $\chi \Rightarrow \chi'$.

2.4 SEQUENCE OF DERIVATION STEPS

A sequence of derivation steps for MGR is defined analogically as the sequence of derivation steps for a MGN.

2.5 N-LANGUAGE

An n-language for MGR is defined analogically as the n-language for a MGN.

2.6 THREE TYPES OF GENERATED LANGUAGES

A language generated by MGN in the X mode, for each $X \in \{union, conc, lm\}$, is defined analogically as the language generated by MGR in the X mode.

2.7 EXAMPLE

$\Gamma = (G_1, G_2, Q)$, where:

- $G_1 = (\{S_1, A_1\}, \{a, b, c\}, \{\mathbf{1:}\ S_1 \rightarrow aS_1, \mathbf{2:}\ S_1 \rightarrow aA_1, \mathbf{3:}\ A_1 \rightarrow bA_1c, \mathbf{4:}\ A_1 \rightarrow bc\}, S_1)$,
- $G_2 = (\{S_2\}, \{d\}, \{\mathbf{1:}\ S_2 \rightarrow S_2S_2, \mathbf{2:}\ S_2 \rightarrow S_2, \mathbf{3:}\ S_2 \rightarrow d\}, S_2)$,
- $Q = \{(\mathbf{1}, \mathbf{1}), (\mathbf{2}, \mathbf{2}), (\mathbf{3}, \mathbf{3}), (\mathbf{4}, \mathbf{3})\}$.

is 2-multigenerative rule-synchronized grammar system.

Notice that this system generates following languages in the different modes:

- $L_{union}(\Gamma) = \{a^n b^n c^n: n \geq 1\} \cup \{d^n: n \geq 1\}$,
- $L_{conc}(\Gamma) = \{a^n b^n c^n d^n: n \geq 1\}$,
- $L_{lm}(\Gamma) = \{a^n b^n c^n: n \geq 1\}$.

3 CONVERSIONS BETWEEN MGN AND MGR

3.1 ALGORITHM 1: CONVERSION FROM MGN TO MGR

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

OUTPUT: MGR $\Gamma' = (G_1, G_2, \ldots, G_n, Q')$; $L_X(\Gamma) = L_X(\Gamma')$,
for each $X \in \{union, conc, lm\}$

METHOD:

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

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

3.2 ALGORITHM 2: CONVERSION FROM MGR TO MGN

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

OUTPUT: MGN $\Gamma' = (G'_1, G'_2, \ldots, G'_n, Q')$; $L_X(\Gamma) = L_X(\Gamma')$,
where $X \in \{union, conc, lm\}$

METHOD:

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

- $G'_i = (N'_i, T_i, P'_i, S_i)$ for all $i = 1, \ldots, n$, where:
  - $N'_i := \{<A, x>: A \rightarrow x \in P_i\} \cup \{S_i\}$,
  - $P'_i := \{<A, x> \rightarrow y: A \rightarrow x \in P_i, y \in \tau_i(x)\} \cup \{S_i \rightarrow y: y \in \tau_i(S_i)\}$,

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

    $\tau_i(a) = \{a\}$ for all $a \in T_i$; $\tau_i(A) = \{<A, x>: A \rightarrow x \in P_i\}$ for all $A \in N_i$.
  - $Q' := \{(<A_1, x_1>, <A_2, x_2>, \ldots, <A_n, x_n>: (A_1 \rightarrow x_1, A_2 \rightarrow x_2, \ldots, A_n \rightarrow x_n) \in Q\} \cup \{(S_1, S_2, \ldots, S_n)\}$

3.3 COROLLARY

The class of languages generated by MGNs in the X mode, where $X \in \{union, conc, lm\}$ is equivalent with the class of language generated by MGRs in the X mode.

Proof:

This corollary follows from Algorithm 1 and Algorithm 2.

4 GENERATIVE POWER OF MGN AND MGR

4.1 CLAIM

For every recursive enumerable language L over an alphabet T there exist a MGR,

$$\Gamma = ((N'_1, T, P'_1, S_1), (N'_2, T, P'_2, S_2), Q), \text{ such that:}$$

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

4.2 THEOREM 1:

For every recursive enumerable language L over an alphabet T there exist a MGR,

$$\Gamma = (G_1, G_2, Q), \text{ such that: } L_{union}(\Gamma) = L.$$

4.3 THEOREM 2:

For every recursive enumerable language L over an alphabet T there exist a MGR,

$$\Gamma = (G_1, G_2, Q), \text{ such that: } L_{lm}(\Gamma) = L.$$

4.4 THEOREM 3:

For every recursive enumerable language L over an alphabet T there exist a MGR,

$$\Gamma = (G_1, G_2, Q), \text{ such that: } L_{conc}(\Gamma) = L.$$

5 CONCLUSION

Let $L(\text{MGN}_X)$ and $L(\text{MGR}_X)$ denote the language families defined by MGN in the X mode and MGR in the X mode, respectively, where $X \in \{union, conc, lm\}$. From the previous results, we obtain $L(\text{RE}) = L(\text{MGN}_X) = L(\text{MGR}_X)$.

REFERENCES

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

[2] Salomaa, A: Formal Languages. Academic Press, 1973