Acta Cybernetica **18** (2008) 783–793.

Two Power-Decreasing Derivation Restrictions in Generalized Scattered Context Grammars*

Tomáš Masopust†, Alexander Meduna†, and Jiří Šimáček†

Abstract

The present paper introduces and discusses generalized scattered context grammars that are based upon sequences of productions whose left-hand sides are formed by nonterminal strings, not just single nonterminals. It places two restrictions on the derivations in these grammars. More specifically, let k be a positive integer. The first restriction requires that all rewritten symbols occur within the first k symbols of the first continuous block of nonterminals in the sentential form during every derivation step. The other restriction defines derivations over sentential forms containing no more than k occurrences of nonterminals. As its main result, the paper demonstrates that both restrictions decrease the generative power of these grammars to the power of context-free grammars.

Keywords: scattered context grammar; grammatical generalization; derivation restriction; generative power.

1 Introduction

Scattered context grammars are based upon finite sets of sequences of context-free productions having a single nonterminal on the left-hand side of every production (see [5]). According to a sequence of n context-free productions, these grammars simultaneously rewrites n nonterminals in the current sentential form according to the n productions in the order corresponding to the appearance of these productions in the sequence. It is well-known that they characterize the family of recursively enumerable languages (see [8]).

In this paper, we generalize these grammars so that the left-hand side of every production may consist of a string of several nonterminals rather than a single nonterminal. Specifically, we discuss two derivation restrictions in scattered context grammars generalized in this way. To explain these restrictions, let k be a constant. The first restriction requires that all simultaneously rewritten symbols

*This work was supported by the Czech Ministry of Education under the Research Plan No. MSM 0021630528 and the Czech Grant Agency project No. 201/07/0005.

†Faculty of Information Technology, Brno University of Technology, Božetěchova 2, Brno 61266, Czech Republic, E-mail: {masopust,meduna}@fit.vutbr.cz, xsimac00@stud.fit.vutbr.cz

occur within the first k symbols of the first continuous block of nonterminals in the current sentential form during every derivation step. The other restriction defines the grammatical derivations over sentential forms containing no more than k occurrences of nonterminals. As the main result, this paper demonstrates that both restrictions decrease the generative power of generalized scattered context grammars to the generative power of context-free grammars. As ordinary scattered context grammars represent special cases of their generalized versions, they also characterize only the family of context-free languages if they are restricted in this way.

This result concerning the derivation restrictions is of some interest when compared to analogical restrictions in terms of other grammars working in a context-sensitive way. Over its history, formal language theory has studied many restrictions placed on the way grammars derive sentential forms and on the forms of productions. In [6], Matthews studied derivations of grammars in the strictly leftmost (rightmost) way—that is, rewritten symbols are preceded (succeeded) only by terminals in the sentential form during the derivation. Later, in [7], he combined both approaches—leftmost and rightmost derivations—so that any sentential form during the derivation is of the form xWy, where x and y are terminal strings, W is a nonterminal string, and a production is applicable only to a leftmost or rightmost substring of W. In both cases, these restrictions result into decreasing the generative power of type-0 grammars to the power of context-free grammars.

Whereas Matthews studied restrictions placed on the forms of derivations, other authors studied the forms of productions. In [2], Book proved that if the left-hand side of any non-context-free production contains besides exactly one nonterminal only terminals, then the generative power of type-0 grammars decreases to the power of context-free grammars. He also proved that if the left-hand side of any non-context-free production has as its left context a terminal string and the left context is at least as long as the right context, then the generative power of type-0 grammars decreases to the power of context-free grammars, too. In [4], Ginsburg and Greibach proved that if the left-hand side of any production is a nonterminal string and the right-hand side contains at least one terminal, then the generated language is context-free. Finally, in [1], Baker proved a stronger result. This result says that if any left-hand side of a production either has, besides terminals, only one nonterminal, or there is a terminal substring, β, on the right-hand side of the production such that the length of β is greater than the length of any terminal substring of the left-hand side of the production, then the generative power of type-0 grammars decreases to the power of context-free grammars. For more details, see page 198 in [9] and the literature cited there.

2 Preliminaries

In this paper, we assume that the reader is familiar with formal language theory (see [10]). For a set Q, $|Q|$ denotes the cardinality of Q. For an alphabet (finite nonempty set) V, V^* represents the free monoid generated by V. The identity of

V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$. For $w \in V^*$, $|w|$ and w^R denote the length and the mirror image of w, respectively, and $sub(w)$ denotes the set of all substrings of w. For $W \subseteq V$, $occur(w, W)$ denotes the number of occurrences of symbols from W in w.

A *pushdown automaton* is a septuple $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, where Q is a finite set of states, Σ is an input alphabet, $q_0 \in Q$ is the initial state, Γ is a pushdown alphabet, δ is a finite set of rules of the form $Zqa \to \gamma p$, where $p, q \in Q$, $Z \in \Gamma \cup \{\varepsilon\}$, $a \in \Sigma \cup \{\varepsilon\}$, $\gamma \in \Gamma^*$, F is a set of final states, and Z_0 is the initial pushdown symbol. Let ψ denote a bijection from δ to Ψ (Ψ is an alphabet of rule labels). We write $r.Zqa \to \gamma p$ instead of $\psi(Zqa \to \gamma p) = r$.

A configuration of M is any word from $\Gamma^* Q \Sigma^*$. For any configuration $xAqay$, where $x \in \Gamma^*$, $y \in \Sigma^*$, $q \in Q$, and any $r.Aqa \to \gamma p \in \delta$, M makes a move from $xAqay$ to $x\gamma py$ according to r, written as $xAqay \Rightarrow x\gamma py\,[r]$, or, simply, $xAqay \Rightarrow x\gamma py$. If $x, y \in \Gamma^* Q \Sigma^*$ and $m > 0$, then $x \Rightarrow^m y$ if and only if there exists a sequence $x_0 \Rightarrow x_1 \Rightarrow \cdots \Rightarrow x_m$, where $x_0 = x$ and $x_m = y$. Then, we say $x \Rightarrow^+ y$ if and only if there exists $m > 0$ such that $x \Rightarrow^m y$, and $x \Rightarrow^* y$ if and only if $x = y$ or $x \Rightarrow^+ y$. The language of M is defined as $\mathcal{L}(M) = \{w \in \Sigma^* : Z_0 q_0 w \Rightarrow^* f,\ f \in F\}$.

A *phrase-structure grammar* or a *grammar* is a quadruple $G = (V, T, P, S)$, where V is a total alphabet, $T \subseteq V$ is an alphabet of terminals, $S \in V - T$ is the start symbol, and P is a finite relation over V^*. Set $N = V - T$. Instead of $(u, v) \in P$, we write $u \to v \in P$ throughout. We call $u \to v$ a production; accordingly, P is G's set of productions. If $u \to v \in P$, $x, y \in V^*$, then G makes a derivation step from xuy to xvy, symbolically written as $xuy \Rightarrow xvy$. If $x, y \in V^*$ and $m > 0$, then $x \Rightarrow^m y$ if and only if there exists a sequence $x_0 \Rightarrow x_1 \Rightarrow \cdots \Rightarrow x_m$, where $x_0 = x$ and $x_m = y$. We write $x \Rightarrow^+ y$ if and only if there exists $m > 0$ such that $x \Rightarrow^m y$, and $x \Rightarrow^* y$ if and only if $x = y$ or $x \Rightarrow^+ y$. The language of G is defined as $\mathcal{L}(G) = \{w \in T^* : S \Rightarrow^* w\}$.

3 Definitions

This section defines a new notion of generalized scattered context grammars. In addition, it formalizes two derivation restrictions studied in this paper.

A *generalized scattered context grammar*, a **SCG** for short, is a quadruple $G = (V, T, P, S)$, where V is a total alphabet, $T \subseteq V$ is an alphabet of terminals, $S \in N$ ($N = V - T$) is the start symbol, and P is a finite set of productions such that each production p has the form $(\alpha_1, \ldots, \alpha_n) \to (\beta_1, \ldots, \beta_n)$, for some $n \geq 1$, where $\alpha_i \in N^+$, $\beta_i \in V^*$, for all $1 \leq i \leq n$. If each production p of the above form satisfies $|\alpha_i| = 1$, for all $1 \leq i \leq n$, then G is an ordinary *scattered context grammar*. Set $\pi(p) = n$. If $\pi(p) \geq 2$, then p is said to be a *context-sensitive* production. If $\pi(p) = 1$, then p is said to be *context-free*. If $(\alpha_1, \ldots, \alpha_n) \to (\beta_1, \ldots, \beta_n) \in P$, $u = x_0\alpha_1 x_1 \ldots \alpha_n x_n$, and $v = x_0 \beta_1 x_1 \ldots \beta_n x_n$, where $x_i \in V^*$, $1 \leq i \leq n$, then $u \Rightarrow v\ [(\alpha_1, \ldots, \alpha_n) \to (\beta_1, \ldots, \beta_n)]$ in G or, simply, $u \Rightarrow v$. Let $\Rightarrow^+$ and $\Rightarrow^*$ denote the transitive and the reflexive and transitive closure of $\Rightarrow$, respectively. The language of G is defined as $\mathcal{L}(G) = \{w \in T^* : S \Rightarrow^* w\}$.

For an alphabet $T = \{a_1, \dots, a_n\}$, there is an *extended Post correspondence problem*, E, defined as

$$E = (\{(u_1, v_1), \dots, (u_r, v_r)\}, (z_{a_1}, \dots, z_{a_n}))\,,$$

where $u_i, v_i, z_{a_j} \in \{0, 1\}^*$, for each $1 \le i \le r$, $1 \le j \le n$. The language represented by E is the set

$$\mathcal{L}(E) = \{b_1 \dots b_k \in T^* : \text{ exists } s_1, \dots, s_l \in \{1, \dots, r\}, l \ge 1, \\ v_{s_1} \dots v_{s_l} = u_{s_1} \dots u_{s_l} z_{b_1} \dots z_{b_k} \text{ for some } k \ge 0\}.$$

It is well known that for each recursively enumerable language, L, there is an extended Post correspondence problem, E, such that $\mathcal{L}(E) = L$ (see Theorem 1 in [3]).

Next, we define two derivation restrictions discussed in this paper.

Let $k \ge 1$. If there is $(\alpha_1, \dots, \alpha_n) \to (\beta_1, \dots, \beta_n) \in P$, $u = x_0\alpha_1 x_1 \dots \alpha_n x_n$, and $v = x_0\beta_1 x_1 \dots \beta_n x_n$, where

1. $x_0 \in T^* N^*$,
2. $x_i \in N^*$, for all $0 < i < n$,
3. $x_n \in V^*$, and
4. $\mathrm{occur}(x_0\alpha_1 x_1 \dots \alpha_n, N) \le k$,

then $u \; {}_k\!\Leftrightarrow v$ $[r]$ in G or, simply, $u \; {}_k\!\Leftrightarrow v$. Let ${}_k\!\Leftrightarrow^n$ denote the n-fold product of ${}_k\!\Leftrightarrow$, where $n \ge 0$. Furthermore, let ${}_k\!\Leftrightarrow^*$ denote the reflexive and transitive closure of ${}_k\!\Leftrightarrow$. Set ${}_{k-left}\mathcal{L}(G) = \{w \in T^* : S \; {}_k\!\Leftrightarrow^* w\}$.

Let $m, h \ge 1$. $W(m)$ denotes the set of all strings $x \in V^*$ satisfying 1 given next. $W(m, h)$ denotes the set of all strings $x \in V^*$ satisfying 1 and 2 given next.

1. $x \in (T^* N^*)^m T^*$;
2. $(y \in \mathrm{sub}(x)$ and $|y| > h)$ implies $\mathrm{alph}(y) \cap T \ne \emptyset$.

If there is $(\alpha_1, \dots, \alpha_n) \to (\beta_1, \dots, \beta_n) \in P$, $u = x_0\alpha_1 x_1 \dots \alpha_n x_n$, and $v = x_0\beta_1 x_1 \dots \beta_n x_n$, where

1. $x_0 \in V^*$,
2. $x_i \in N^*$, for all $0 < i < n$, and
3. $x_n \in V^*$,

then $u \Leftrightarrow v$ $[r]$ in G or, simply, $u \Leftrightarrow v$. Let $\Leftrightarrow^n$ denote n-fold product of $\Leftrightarrow$, where $n \ge 0$. Furthermore, let $\Leftrightarrow^*$ denote the reflexive and transitive closure of $\Leftrightarrow$.

Let $u, v \in V^*$, and $u \Leftrightarrow v$.

$$u \; {}_m^h\!\Leftrightarrow v$$

if and only if $u, v \in W(m, h)$, and

$$u \; {}_m\!\Leftrightarrow v$$

if and only if $u, v \in W(m)$. Set ${}_{nonter}\mathcal{L}(G, m, h) = \{w \in T^* : S \; {}_m^h\!\Leftrightarrow^* w\}$ and ${}_{nonter}\mathcal{L}(G, m) = \{w \in T^* : S \; {}_m\!\Leftrightarrow^* w\}$.

3.1 Language Families

Let **SCG**s denote the family of generalized scattered context grammars. Define these language families:

$$\begin{aligned} {}_{nonter}SC(m,h) &= \{L : L = {}_{nonter}\mathcal{L}(G,m,h), G \in \mathbf{SCG}\text{s}\} \text{ for all } m,h \geq 1 \\ {}_{nonter}SC(m) &= \{L : L = {}_{nonter}\mathcal{L}(G,m), G \in \mathbf{SCG}\text{s}\} \text{ for all } m \geq 1 \\ {}_{k-left}SC &= \{L : L = {}_{k-left}\mathcal{L}(G), G \in \mathbf{SCG}\text{s}\} \text{ for all } k \geq 0 \end{aligned}$$

Let **CF**, **CS**, and **RE** denote the families of context-free, context-sensitive, and recursively enumerable languages, respectively. For all $k \geq 0$, ${}_kCF$ denote the family of languages generated by context-free grammars of index k.

4 Results

This section presents the main results of this paper. First, it demonstrates that, for every $k \geq 1$, $CF = {}_{k-left}SC$, then that $RE = {}_{nonter}SC(1)$, and, finally, that for every $m, h \geq 1$, ${}_mCF = {}_{nonter}SC(m,h)$.

Theorem 1. *Let k be a positive integer. Then, $CF = {}_{k-left}SC$.*

Proof. Let $G = (V, T, P, S)$ be a generalized scattered context grammar. Consider the following pushdown automaton

$$M = (\{q,r,f\} \cup \{[\gamma,s] : \gamma \in N^*, |\gamma| \leq k, s \in \{q,r\}\}, T, V \cup \{Z\}, \delta, [S,q], Z, \{f\}),$$

where $Z \notin V$, and δ contains rules of the following forms:

1. $[\beta_0 A_1 \beta_1 \dots A_n \beta_n, q] \to (\beta_0 \alpha_1 \beta_1 \dots \alpha_n \beta_n)^R [\varepsilon, r]$
   if $(A_1, \dots, A_n) \to (\alpha_1, \dots, \alpha_n) \in P$; $\beta_i \in N^*$, $0 \leq i \leq n$;
2. $A[A_1 \dots A_n, r] \to [A_1 \dots A_n A, r]$ if $n < k$, $A \in N$;
3. $[A_1 \dots A_k, r] \to [A_1 \dots A_k, q]$;
4. $a[A_1 \dots A_n, r] \to a[A_1 \dots A_n, q]$ if $n < k$, $a \in T$;
5. $Z[A_1 \dots A_n, r] \to Z[A_1 \dots A_n, q]$ if $n < k$;
6. $a[\varepsilon, r]a \to [\varepsilon, r]$ if $a \in T$;
7. $Z[\varepsilon, r] \to f$.

We prove that $\mathcal{L}(M) = {}_{k-left}\mathcal{L}(G)$.

($\subseteq$:) By induction on the number of rules constructed in 1 used in a sequence of moves, we prove the following claim.

Claim 1. *If $Z\alpha^R[\beta_0 A_1 \beta_1 \dots A_n \beta_n, q]w \Rightarrow^* f$, then $\beta_0 A_1 \beta_1 \dots A_n \beta_n \alpha \;{}_k\!\Leftrightarrow^* w$.*

Proof. Basis: Only one rule constructed in 1 is used. Then,

$$Z\alpha^R[\beta_0 A_1 \beta_1 \dots A_n \beta_n, q]uw \Rightarrow Z(\beta_0 \alpha_1 \beta_1 \dots \alpha_n \beta_n \alpha)^R [\varepsilon, r]uw \Rightarrow^* f,$$

where $(A_1, \dots, A_n) \rightarrow (\alpha_1, \dots, \alpha_n) \in P$, $n \leq k$, and $\beta_0\alpha_1\beta_1 \dots \alpha_n\beta_n\alpha \in T^*$. Therefore, $\beta_0 = \dots = \beta_n = \varepsilon$, and $\alpha_1 \dots \alpha_n\alpha = uw$. Then,

$$A_1 \dots A_n w \ {}_k\!\Leftrightarrow\!\Rightarrow uw.$$

Induction hypothesis: Suppose that the claim holds for all sequences of moves containing no more than i rules constructed in 1.

Induction step: Consider a sequence of moves containing $i+1$ rules constructed in 1. Then,

$$\begin{array}{lll} & Z\alpha^R[\beta_0 A_1\beta_1 \dots A_l\beta_l, q]w & \\ \Rightarrow & Z\alpha^R(\beta_0\alpha_1\beta_1 \dots \alpha_l\beta_l)^R[\varepsilon, r]w & \text{(by a rule constructed in 1)} \\ \Rightarrow^* & Z\alpha'[\varepsilon, r]w' & \text{(by rule constructed in 6)} \\ \Rightarrow^* & Z\alpha''[\beta'_0 B_1\beta'_1 \dots B_m\beta'_m, r]w' & \text{(by rule constructed in 2)} \\ \Rightarrow & Z\alpha''[\beta'_0 B_1\beta'_1 \dots B_m\beta'_m, q]w' & \text{(by a rule constructed in 3, 4, or 5)} \\ \Rightarrow^* & f & \end{array}$$

where $\alpha' \in V^*N \cup \{\varepsilon\}$, $v \in T^*$, $\alpha' v^R = \alpha^R(\beta_0\alpha_1\beta_1 \dots \alpha_l\beta_l)^R$, and $vw' = w$. Then, by the production $(A_1, \dots, A_l) \rightarrow (\alpha_1, \dots, \alpha_l)$,

$$\beta_0 A_1\beta_1 \dots A_l\beta_l\alpha \ {}_k\!\Leftrightarrow\!\Rightarrow \beta_0\alpha_1\beta_1 \dots \alpha_l\beta_l\alpha,$$

where $|\beta_0 A_1\beta_1 \dots A_l\beta_l| \leq k$,

$$\beta_0\alpha_1\beta_1 \dots \alpha_l\beta_l\alpha = v(\alpha')^R = v\beta'_0 B_1\beta'_1 \dots B_m\beta'_m(\alpha'')^R,$$

and, by the induction hypothesis,

$$v\beta'_0 B_1\beta'_1 \dots B_m\beta'_m(\alpha'')^R \ {}_k\!\Leftrightarrow\!\Rightarrow^* vw'.$$

Hence, the inclusion holds. △

($\supseteq$:) First, we prove the following claim.

Claim 2. *If $\beta \ {}_k\!\Leftrightarrow\!\Rightarrow^* w$, where $\beta \in NV^*$, then $Z\beta^R[\varepsilon, r]w \Rightarrow^* f$.*

Proof. By induction on the length of derivations.

Basis: Let $A_1 \dots A_n w \ {}_k\!\Leftrightarrow\!\Rightarrow \alpha_1 \dots \alpha_n w$ $(\alpha_1 \dots \alpha_n = \alpha)$, where $\alpha w \in {}_{k-left}\mathcal{L}(G)$, and $(A_1, \dots, A_n) \rightarrow (\alpha_1, \dots, \alpha_n) \in P$, $1 \leq n \leq k$. M simulates this derivation step as follows.

$$\begin{array}{lll} & Zw^R A_n \dots A_1[\varepsilon, r]\alpha w & \\ \Rightarrow^n & Zw^R[A_1 \dots A_n, r]\alpha w & \text{(by rule constructed in 2)} \\ \Rightarrow & Zw^R[A_1 \dots A_n, q]\alpha w & \text{(by a rule constructed in 4 or 5)} \\ \Rightarrow & Zw^R\alpha^R[\varepsilon, r]\alpha w & \text{(by a rule constructed in 1)} \\ \Rightarrow^{|\alpha w|} & Z[\varepsilon, r] & \text{(by rule constructed in 6)} \\ \Rightarrow & f & \text{(by the rule constructed in 7)} \end{array}$$

Induction hypothesis: Suppose that the claim holds for all derivations of length i or less.

Induction step: Consider a derivation of length $i+1$. Let

$$\beta_0 B_1 \beta_1 \dots B_l \beta_l \gamma \ {}_k\!\Leftrightarrow \beta_0 \alpha_1 \beta_1 \dots \alpha_l \beta_l \gamma \ {}_k\!\Leftrightarrow^i \varphi w,$$

where $\varphi w \in {}_{k-left}\mathcal{L}(G)$, $\beta_0 B_1 \beta_1 \dots B_l \beta_l \in N^+$, and either $|\beta_0 B_1 \beta_1 \dots B_l \beta_l| = k$, or $|\beta_0 B_1 \dots B_l \beta_l| < k$, $\beta_0 \alpha_1 \beta_1 \dots \alpha_l \beta_l \gamma = \varphi \psi$, where $\varphi \in T^*$, $\psi \in NV^* \cup \{\varepsilon\}$, and $\gamma \in TV^* \cup \{\varepsilon\}$. Then,

	$Z(\beta_0 B_1 \beta_1 \dots B_l \beta_l \gamma)^R [\varepsilon, r] \varphi w$	
$\Rightarrow^*$	$Z \gamma^R [\beta_0 B_1 \beta_1 \dots B_l \beta_l, r] \varphi w$	(by rule constructed in 2)
$\Rightarrow$	$Z \gamma^R [\beta_0 B_1 \beta_1 \dots B_l \beta_l, q] \varphi w$	(by a rule constructed in 3 or 4)
$\Rightarrow$	$Z(\varphi \psi \gamma)^R [\varepsilon, r] \varphi w$	(by a rule constructed in 1)
$\Rightarrow^*$	$Z(\psi \gamma)^R [\varepsilon, r] w$	(by a rule constructed in 6)
$\Rightarrow^*$	f	(by the induction hypothesis)

Hence, the claim holds. △

Now, if $S \Rightarrow u\alpha \Rightarrow^* uw$, where $u \in T^*$ and $\alpha \in NV^*$, then $Z[S,q]uw \Rightarrow Z(u\alpha)^R[\varepsilon, r]uw \Rightarrow^* Z\alpha^R[\varepsilon, r]w \Rightarrow^* f$, by rules constructed in 1 and 6 and the previous claim. For $\alpha = \varepsilon$, $Z[S,q]u \Rightarrow Zu^R[\varepsilon, r]u \Rightarrow^* f$. Hence, the other inclusion holds. □

Theorem 2. $RE = {}_{nonter}SC(1)$.

Proof. Let $L \subseteq \{a_1, \dots, a_n\}^*$ be a recursively enumerable language. There is an extended Post correspondence problem,

$$E = (\{(u_1, v_1), \dots, (u_r, v_r)\}, (z_{a_1}, \dots, z_{a_n})),$$

where $u_i, v_i, z_{a_j} \in \{0,1\}^*$, for each $1 \le i \le r$, $1 \le j \le n$, such that $\mathcal{L}(E) = L$; that is, $w = b_1 \dots b_k \in L$ if and only if $w \in \mathcal{L}(E)$. Set $V = \{S, A, 0, 1, \$\} \cup T$. Define the **SCG** $G = (V, T, P, S)$ with P constructed as follows:

1. For every $a \in T$, add
   a) $(S) \to ((z_a)^R S a)$, and
   b) $(S) \to ((z_a)^R A a)$ to P;
2. a) For every $(u_i, v_i) \in E$, $1 \le i \le r$, add $(A) \to ((u_i)^R A v_i)$ to P;
   b) Add $(A) \to (\$\$)$ to P;
3. Add
   a) $(0, \$, \$, 0) \to (\$, \varepsilon, \varepsilon, \$)$,
   b) $(1, \$, \$, 1) \to (\$, \varepsilon, \varepsilon, \$)$, and
   c) $(\$) \to (\varepsilon)$ to P.

Claim 3. *Let $w_1, w_2 \in \{0,1\}^*$. Then, $w_1\$\$w_2 \Rightarrow_G^* \varepsilon$ if and only if $w_1 = (w_2)^R$.*

Proof. If: Let $w_1 = (w_2)^R = b_1 \dots b_k$, for some $k \geq 0$. By productions (3a) and (3b) followed by two applications of (3c), we obtain

$$\begin{aligned} b_k \dots b_2 b_1 \$\$ b_1 b_2 \dots b_k &\Rightarrow b_k \dots b_2 \$\$ b_2 \dots b_k \\ &\Rightarrow^* b_k \$\$ b_k \\ &\Rightarrow \$\$ \Rightarrow \$ \Rightarrow \varepsilon. \end{aligned}$$

Therefore the if-part of the claim holds.

Only if: Suppose that $|w_1| \leq |w_2|$. We demonstrate that

$$w_1\$\$w_2 \Rightarrow_G^* \varepsilon \text{ implies } w_1 = (w_2)^R$$

by induction on $k = |w_1|$.

Basis: Let $k = 0$. Then, $w_1 = \varepsilon$ and the only possible derivation is

$$\$\$w_2 \Rightarrow \$w_2\ [(3c)] \Rightarrow w_2\ [(3c)].$$

Hence, we can derive ε only if $w_1 = (w_2)^R = \varepsilon$.

Induction Hypothesis: Suppose that the claim holds for all w_1 satisfying $|w_1| < k$ for some $k \geq 0$.

Induction Step: Consider $w_1 a\$\$ b w_2$ with $a \neq b$, $a, b \in \{0,1\}$. If $w_1 = w_{11} b w_{12}$, $w_{11}, w_{12} \in \{0,1\}^*$, then either (3a) or (3b) can be used. In either case, we obtain

$$w_1 a\$\$ b w_2 \Rightarrow w_{11}\$ w_{12} a w_{21} \$ w_{22},$$

where $bw_2 = w_{21} b w_{22}$, $w_{21}, w_{22} \in \{0,1\}^*$, and $w_{12} a w_{21} \in N^+$ cannot be removed by any production from the sentential form. The same is true when $w_2 = w'_{21} a w'_{22}$, $w'_{21}, w'_{22} \in \{0,1\}^*$. Therefore, the derivation proceeds successfully only if $a = b$. Thus,

$$w_1 a\$\$ b w_2 \Rightarrow w_1\$\$w_2 \Rightarrow^* \varepsilon,$$

and from the induction hypothesis,

$$w_1 = (w_2)^R.$$

Analogously, the same result can be proved for $|w_1| \geq |w_2|$, which implies that the only-if part of the claim holds.

Therefore, the claim holds. △

Examine the introduced productions to see that G always generates $b_1 \dots b_k \in \mathcal{L}(E)$ by a derivation of this form:

$$\begin{array}{lll} S & \Rightarrow & (z_{b_k})^R S b_k \\ & \Rightarrow & (z_{b_k})^R (z_{b_{k-1}})^R S b_{k-1} b_k \\ & \Rightarrow^* & (z_{b_k})^R \dots (z_{b_2})^R S b_2 \dots b_k \\ & \Rightarrow & (z_{b_k})^R \dots (z_{b_2})^R (z_{b_1})^R A b_1 b_2 \dots b_k \\ & \Rightarrow & (z_{b_k})^R \dots (z_{b_1})^R (u_{s_l})^R A v_{s_l} b_1 \dots b_k \\ & \Rightarrow^* & (z_{b_k})^R \dots (z_{b_1})^R (u_{s_l})^R \dots (u_{s_1})^R A v_{s_1} \dots v_{s_l} b_1 \dots b_k \\ & \Rightarrow & (z_{b_k})^R \dots (z_{b_1})^R (u_{s_l})^R \dots (u_{s_1})^R \$\$ v_{s_1} \dots v_{s_l} b_1 \dots b_k \\ & & = (u_{s_1} \dots u_{s_l} z_{b_1} \dots z_{b_k})^R \$\$ v_{s_1} \dots v_{s_l} b_1 \dots b_k \\ & \Rightarrow^* & b_1 \dots b_k. \end{array}$$

Productions introduced in steps 1 and 2 of the construction find nondeterministically the solution of the extended Post correspondence problem which is subsequently verified by productions from step 3. Therefore, $w \in L$ if and only if $w \in \mathcal{L}(G)$ and the theorem holds. □

Theorem 3. *Let m and h be positive integers. Then, ${}_m CF = {}_{nonter}SC(m, h)$.*

Proof. Obviously, ${}_m CF \subseteq {}_{nonter}SC(m, h)$.

We prove that ${}_{nonter}SC(m, h) \subseteq {}_m CF$. Let $\alpha = x_0 y_1 x_1 \dots y_n x_n$, where $x_i \in T^*$, $y_i \in N^+$, for $0 \le i \le n$, and for all $0 < i < n$, $x_i \neq \varepsilon$. Define $f(\alpha) = x_0 \langle y_1 \rangle x_1 \dots \langle y_n \rangle x_n$, where $\langle y_i \rangle$ is a new nonterminal, for all $0 \le i \le n$. Let $G_{SC} = (V, T, P, S)$ be a generalized scattered context grammar. Introduce a context-free grammar $G_{CF} = (V', T, P', \langle S \rangle)$, where $V' = \{\langle \gamma \rangle : \gamma \in N^*, 1 \le |\gamma| \le h\} \cup T$ and P' is constructed as follows:

1. for each $\gamma = x_0 \alpha_1 x_1 \dots \alpha_n x_n$, where $x_i \in N^*$, $\alpha_i \in N^+$, $1 \le |\gamma| \le h$, and $(\alpha_1, \dots, \alpha_n) \to (\beta_1, \dots, \beta_n) \in P$, add $\langle \gamma \rangle \to f(x_0 \beta_1 x_1 \dots \beta_n x_n)$ to P'.

Claim 4. *Let $S\ {}_m^h\!\circ\!\!\Rightarrow^k \omega$ in G_{SC}, where $\omega \in V^*$, $k \ge 0$. Then, $\langle S \rangle\ {}_m\!\Rightarrow^k f(\omega)$ in G_{CF}.*

Proof. By induction on $k = 0, 1, \dots$.

Basis: Let $k = 0$, thus $S\ {}_m^h\!\circ\!\!\Rightarrow^0 S$ in G_{SC}. Then, $\langle S \rangle\ {}_m\!\Rightarrow^0 \langle S \rangle$ in G_{CF}. As $f(S) = \langle S \rangle$, the basis holds.

Induction hypothesis: Suppose that the claim holds for all $0 \le m \le k$, where k is a non-negative integer.

Induction step: Let $S\ {}_m^h\!\circ\!\!\Rightarrow^k \phi\gamma\psi\ {}_m^h\!\circ\!\!\Rightarrow \phi\gamma'\psi$ in G_{SC}, and the last production applied during the derivation is $(\alpha_1, \dots, \alpha_n) \to (\beta_1, \dots, \beta_n)$, where $\phi \in V^*T \cup \{\varepsilon\}$, $\gamma = x_0 \alpha_1 x_1 \dots \alpha_n x_n$, $\psi \in TV^* \cup \{\varepsilon\}$, $\gamma' = x_0 \beta_1 x_1 \dots \beta_n x_n$, $\alpha_i, x_i \in N^*$, and $\beta_i \in V^*$. By the induction hypothesis,

$$\langle S \rangle\ {}_m\!\Rightarrow^k f(\phi\gamma\psi).$$

By the definition of f, ϕ, and ψ, $f(\phi\gamma\psi) = f(\phi)\langle\gamma\rangle f(\psi)$. Hence, we can use the production $\langle\gamma\rangle \to f(\gamma') \in P'$ introduced in 1 in the construction to obtain

$$f(\phi)\langle\gamma\rangle f(\psi) \ {}_m\!\Rightarrow f(\phi)f(\gamma')f(\psi).$$

By the definition of f, ϕ, and ψ, $f(\phi)f(\gamma')f(\psi) = f(\phi\gamma'\psi)$. As a result,

$$\langle S\rangle \ {}_m\!\Rightarrow^k f(\phi)\langle\gamma\rangle f(\psi) \ {}_m\!\Rightarrow f(\phi\gamma'\psi)$$

and, therefore, $\langle S\rangle \ {}_m\!\Rightarrow^{k+1} f(\phi\gamma'\psi)$ and the claim holds for $k+1$. △

Claim 5. *Let $\langle S\rangle \ {}_m\!\Rightarrow^k \omega$ in G_{CF}, where $\omega \in V'^*$, $k \geq 0$. Then, $S \ {}_m^h\!\circ\!\!\Rightarrow^k f^{-1}(\omega)$ in G_{SC}.*

Proof. By induction on $k = 0, 1, \ldots$.

Basis: Let $k = 0$, thus $\langle S\rangle \ {}_m\!\Rightarrow^0 \langle S\rangle$ in G_{CF}. Then $S \ {}_m^h\!\circ\!\!\Rightarrow^0 S$ in G_{SC}. As $f^{-1}(\langle S\rangle) = S$, the basis holds.

Induction hypothesis: Suppose that the claim holds for all $0 \leq m \leq k$, where k is a non-negative integer.

Induction step: Let $\langle S\rangle \ {}_m\!\Rightarrow^k \phi\langle\gamma\rangle\psi \ {}_m\!\Rightarrow \phi\gamma'\psi$ in G_{CF}, and the last production applied during the derivation is $\langle\gamma\rangle \to \gamma'$, where $\phi \in V^*T \cup \{\varepsilon\}$, $\gamma = x_0\alpha_1x_1\ldots\alpha_nx_n$, $\psi \in TV^* \cup \{\varepsilon\}$, $\gamma' = f(x_0\beta_1x_1\ldots\beta_nx_n)$, $\alpha_i, x_i \in N^*$, and $\beta_i \in V^*$. By the induction hypothesis,

$$S \ {}_m^h\!\circ\!\!\Rightarrow^k f^{-1}(\phi\langle\gamma\rangle\psi).$$

By the definition of f, ϕ, and ψ, $f^{-1}(\phi\langle\gamma\rangle\psi) = f^{-1}(\phi)\gamma f^{-1}(\psi)$. There exists $(\alpha_1,\ldots,\alpha_n) \to (\beta_1,\ldots,\beta_n) \in P$ by 1, thus

$$f^{-1}(\phi)\gamma f^{-1}(\psi) \ {}_m^h\!\circ\!\!\Rightarrow f^{-1}(\phi)f^{-1}(\gamma')f^{-1}(\psi).$$

By the definition of f, ϕ, and ψ, $f^{-1}(\phi)f^{-1}(\gamma')f^{-1}(\psi) = f^{-1}(\phi\gamma'\psi)$. As a result

$$S \ {}_m^h\!\circ\!\!\Rightarrow^k f^{-1}(\phi)\gamma f^{-1}(\psi) \ {}_m^h\!\circ\!\!\Rightarrow f^{-1}(\phi\gamma'\psi)$$

and, therefore, $S \ {}_m^h\!\circ\!\!\Rightarrow^{k+1} f^{-1}(\phi\gamma'\psi)$ and the claim holds for $k+1$. △

Hence, the theorem holds. □

References

[1] Baker, B. S. Context-sesitive grammars generating context-free languages. In Nivat, M., editor, *Automata, Languages and Programming*, pages 501–506. North-Holland, Amsterdam, 1972.

[2] Book, R. V. Terminal context in context-sensitive grammars. *SIAM Journal of Computing*, 1:20–30, 1972.

[3] Geffert, V. Context-free-like forms for the phrase-structure grammars. In Chytil, M., Janiga, L., and Koubek, V., editors, *Mathematical Foundations of Computer Science*, volume 324 of *Lecture Notes in Computer Science*, pages 309–317. Springer-Verlag, 1988.

[4] Ginsburg, S. and Greibach, S. Mappings which preserve context-sensitive languages. *Information and Control*, 9:563–582, 1966.

[5] Greibach, S. and Hopcroft, J. Scattered context grammars. *Journal of Computer and System Sciences*, 3:233–247, 1969.

[6] Matthews, G. A note on symmetry in phrase structure grammars. *Information and Control*, 7:360–365, 1964.

[7] Matthews, G. Two-way languages. *Information and Control*, 10:111–119, 1967.

[8] Meduna, A. A trivial method of characterizing the family of recursively enumerable languages by scattered context grammars. *EATCS Bulletin*, pages 104–106, 1995.

[9] Rozenberg, G. and Salomaa, A., editors. *Handbook of Formal Languages*, volume 1. Springer-Verlag, Berlin, 1997.

[10] Salomaa, A. *Formal languages.* Academic Press, New York, 1973.

Received 18th July 2007