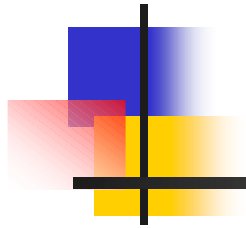


Scattered Context Grammars

Alexander Meduna

**Faculty of Information Technology
Brno University of Technology
Brno, Czech Republic, Europe**





Based on these Papers

- Meduna, A.: Coincidental Extension of Scattered Context Languages, *Acta Informatica* 39, 307-314, 2003
- Meduna, A. and Fernau, H.: On the Degree of Scattered Context-Sensitivity. *Theoretical Computer Science* 290, 2121-2124, 2003
- Meduna, A.: Descriptive Complexity of Scattered Rewriting and Multirewriting: An Overview. *Journal of Automata, Languages and Combinatorics*, 571-579, 2002
- Meduna, A. and Fernau, H.: A Simultaneous Reduction of Several Measures of Descriptive Complexity in Scattered Context Grammars. *Information Processing Letters*, 214-219, 2003



Classification of Parallel Grammars

I. Totally parallel grammars, such as L systems, rewrite **all** symbols of the sentential form during a single derivation step (not discussed in this talk).

II. Partially parallel grammars rewrite **some** symbols while leaving the other symbols unrewritten.

- **Scattered Context Grammars** work in a partially parallel way.
- These grammars are **central to this talk**.



Scattered Context Grammars (SCGs)

Essence

- semi-parallel grammars
- application of several context-free productions during a single derivation step
- stronger than CFGs

Main Topics under Discussion

- reduction of the grammatical size
- new language operations



Concept

Concept

- sequences of context-free productions
- several nonterminals are rewritten in parallel while the rest of the sentential form remains unchanged



Definition

Scattered context grammar :

- $G = (N, T, P, S)$
- $N, T,$ and S as in a CFG
- P is a finite set of productions of the form
 $(A_1, A_2, \dots, A_n) \rightarrow (x_1, x_2, \dots, x_n)$
where $A_i \in N$ and $x_j \in V^*$ with $V = N \cup T$

Direct derivation:

- $u_1 A_1 u_2 A_2 u_3 \dots u_n A_n u_{n+1} \Rightarrow u_1 x_1 u_2 x_2 u_3 \dots u_n x_n u_{n+1}$ if
 $(A_1, A_2, \dots, A_n) \rightarrow (x_1, x_2, \dots, x_n)$

Generated language:

- $L(G) = \{w. S \Rightarrow^* w \text{ and } w \in T^*\}$



Example

Productions:

$$(S) \rightarrow (AA), (A, A) \rightarrow (aA, bAc), (A, A) \rightarrow (\varepsilon, \varepsilon)$$

Derivation:

$$S \Rightarrow AA \Rightarrow aAbAc \Rightarrow aaAbbAcc \Rightarrow aabbcc$$

Generated Language:

$$L(G) = \{a^i b^i c^i : i \geq 0\}$$



Language Families

Language Families

- **CS** - Context Sensitive Languages
- **RE** - Recursively Enumerable Languages
- **SC** = $\{L(G): G \text{ is a SCG}\}$

for every $n \geq 1$,

- **SC(n)** = $\{L(G): G \text{ is a SCG with no more than } n \text{ nonterminals}\}$



Reduction of SCGs

Reduction of SCGs

- **(A) reduction of the number of nonterminals**
- (B) reduction of the number of context (non-context-free) productions
- (C) simultaneous reduction of (A) and (B)



Reduction (A) 1/2

Reduction of the Number of Nonterminals

- **Theorem 1:** $RE = SC(3)$
- **Theorem 2:** $CS \not\subseteq SC(1)$
- **Proof (Sketch):** Let $L = \{a^h : h = 2^n, n \geq 1\}$. Assume that $L = L(G)$, where $G = (\{S\}, \{a\}, P, S)$ is a SCG. In G ,

$$S \Rightarrow^* a^i S a^j \Rightarrow^* a^i a^k a^j$$

for some $i, j \geq 0$ such that $i + j, k \geq 1$. Thus,

$$S \Rightarrow^* a^{in} S a^{jn} \Rightarrow^* a^{in} a^k a^{jn}$$

for every $n \geq 0$. As $a^i a^k a^j \in L$, $|a^i a^k a^j| = i + k + j = 2^m$. Consider $v = a^{2i} a^k a^{2j} \in L$. Then, $2^m < |v| = 2^m + i + j < 2^{m+1}$, so $v \notin L$ —a contradiction.



Reduction (A) 2/2

- **Corollary:** $SC(1) \subset SC(3) = RE$
- **Open Problem:** $RE = SC(2)$?



Reduction (B)

Reduction of SCGs

- (A) reduction of the number of nonterminals
- **(B) reduction of the number of context (non-context-free) productions**
- (C) reduction of (A) and (B)



Reduction (B) 1/5

Reduction of the Number of Context Productions

- A *context production* means a non-context-free production

$$(A_1, A_2, \dots, A_n) \rightarrow (x_1, x_2, \dots, x_n) \text{ with } n \geq 2$$

- **Theorem 4:** Every language in **RE** is generated by a scattered context grammar with only these two context productions:

$$(\$, 0, 0, \$) \rightarrow (\varepsilon, \$, \$, \varepsilon)$$

$$(\$, 1, 1, \$) \rightarrow (\varepsilon, \$, \$, \varepsilon)$$



Reduction (B) 2/5

I. Left-Extended Queue Grammar

$$Q = (V, T, W, F, s, R)$$

R - finite set of productions of the form (a, q, z, r) . Every generation of $h \in L(Q)$ has this form

$$\# a_0 q_0$$

$$\Rightarrow a_0 \# a_1 x_0 q_1$$

$$[(a_0, q_0, z_0, q_1)]$$

$$\Rightarrow a_0 a_1 \# a_2 x_1 q_2$$

$$[(a_1, q_1, z_1, q_2)]$$

$$\Rightarrow a_0 a_1 \dots a_k \# a_{k+1} x_k q_{k+1}$$

$$\Rightarrow a_0 a_1 \dots a_k a_{k+1} \# a_{k+2} x_{k+1} y_1 q_{k+2}$$

$$[(a_{k+1}, q_{k+1}, y_1, q_{k+2})]$$

$$\Rightarrow a_0 a_1 \dots a_k a_{k+1} \dots a_{k+m-1} \# a_{k+m} y_1 \dots y_{m-1} q_{k+m}$$

$$[(a_{k+m-1}, q_{k+m-1}, y_{m-1}, q_{k+m})]$$

$$\Rightarrow a_0 a_1 \dots a_k a_{k+1} \dots a_{k+m} \# y_1 \dots y_m q_{k+m+1}$$

$$[(a_{k+m}, q_{k+m}, y_m, q_{k+m+1})]$$

where $h = y_1 \dots y_m$ with $q_{k+m+1} \in F$



Reduction (B) 3/5

II. Substitutions

g : binary code of symbols from V

h : binary code of states from W

III. Introduction of SCG

$G = (N, T, CF \cup Context, S)$

$Context = \{ (\$, 0, 0, \$) \rightarrow (\varepsilon, \$, \$, \varepsilon),$
 $(\$, 1, 1, \$) \rightarrow (\varepsilon, \$, \$, \varepsilon) \}$

IV. CF used to generate

$\$g(a_0 a_1 \dots a_k a_{k+1} \dots a_{k+m})y_1 \dots y_m h(q_{k+m} \dots q_{k+1} q_k \dots q_1 q_0)\$$



Reduction (B) 4/5

V. Context used to verify

$$g(a_0 a_1 \dots a_k a_{k+1} \dots a_{k+m}) = h(q_0 q_1 \dots q_k q_{k+1} \dots q_{k+m})$$

$$\text{let } g(a_0 a_1 \dots a_k a_{k+1} \dots a_{k+m}) = c_0 c_1 \dots c_{(k+m)2n}$$

$$\text{let } h(q_0 q_1 \dots q_k q_{k+1} \dots q_{k+m}) = d_0 d_1 \dots d_{(k+m)2n}$$

where each $c_i, d_i \in \{0, 1\}$

By using $(\$, 0, 0, \$) \rightarrow (\varepsilon, \$, \$, \varepsilon)$ and

$(\$, 1, 1, \$) \rightarrow (\varepsilon, \$, \$, \varepsilon)$, G makes

$$\$c_0 c_1 c_2 \dots c_{(k+m)2n} y_1 \dots y_m d_{(k+m)2n} \dots d_2 d_1 d_0 \$$$

$$\$c_1 c_2 \dots c_{(k+m)2n} y_1 \dots y_m d_{(k+m)2n} \dots d_2 d_1 \$$$

$$\$c_2 \dots c_{(k+m)2n} y_1 \dots y_m d_{(k+m)2n} \dots d_2 \$$$

$$\$y_1 \dots y_m \$$$

$$y_1 \dots y_m$$



Reduction (B) 5/5

- **Corollary 5:** The *SCGs* with two context productions characterize ***RE***.
- **Open Problem:** What is the power of the *SCGs* with a single context production?



Reduction of SCGs

Reduction of SCGs

- (A) reduction of the number of nonterminals
- (B) reduction of the number of context (non-context-free) productions
- **(C) reduction of (A) and (B)**



Simultaneous Reduction (A) & (B)

Simultaneous Reduction of the Number of Nonterminals and the Number of Context Productions

- **Note:** Next two theorems were proved in cooperation with H. Fernau (Germany).
- **Theorem:** Every type-0 language is generated by a SCG with no more than **seven context productions and** no more than **five nonterminals**
- **Theorem:** Every type-0 language is generated by a SCG with no more than **six context productions and** no more than **six nonterminals**
- **Open Problem:** Can we improve the above theorems?



New Operations

ε -free SCGs

- ε -free SCG: each production $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)$ satisfies $x_j \neq \varepsilon$
- ε -free **SC** = $\{L(G): G \text{ is an } \varepsilon\text{-free SCG}\}$
- ε -free **SC** \subseteq **CS** \subset **SC** = **RE**

- Objective: Increase of ε -free **SC** to **RE** by a simple language operation over ε -free **SC**



Coincidental Extension 1/6

Coincidental Extension

- For a symbol, $\#$, and a string, $x = a_1 a_2 \dots a_{n-1} a_n$, any string of the form $\#^i a_1 \#^i a_2 \#^i \dots \#^i a_{n-1} \#^i a_n \#^i$, where $i \geq 0$, is a *coincidental $\#$ -extension* of x .
- A language, K , is a coincidental $\#$ -extension of L if every string of K represents a coincidental extension of a string in L and the deletion of all $\#$ s in K results in L , symbolically written as $L \# \blacktriangleleft K$
- If $L \# \blacktriangleleft K$ and there are an infinitely many coincidental extensions of x in K for every $x \in L$, we write $L \# \blacktriangleleft_{\infty} K$



Coincidental Extension 2/6

Examples:

For $X = \{ \#^i a \#^i b \#^i : i \geq 5 \} \cup \{ \#^i c^n \#^i d^n \#^i : n, i \geq 0 \}$ and

$Y = \{ ab \} \cup \{ c^n d^n : n \geq 0 \},$

$Y \# \triangleleft_{\infty} X,$ so $Y \# \triangleleft X.$

For $A = \{ \# a \# b \# \} \cup \{ \#^i c^n \#^i d^n \#^i : n, i \geq 0 \},$

$Y \# \triangleleft A$ holds, but $Y \# \triangleleft_{\infty} A$ does not hold.

$B = \{ \#^i a \#^i b \#^i : i \geq 5 \} \cup \{ \#^i c^n \#^i d^n \#^{i+1} : n, i \geq 0 \}$ is not the coincidental $\#$ -extension of any language.



Coincidental Extension 3/6

- **Theorem:** Let $K \in RE$. Then, there exists a ε -free SCG, G , such that $K \# \triangleleft_{\infty} L(G)$.
- **Proof (Sketch):** Let $K \in RE$. There exists a SCG, G , such that $L = L(G)$. Construct a ε -free SCG, $G = (V, P, S, \{\#\} \cup T)$, so that $L \# \triangleleft_{\infty} L(G)$.

Homomorphism h :

$h(A) = A$ for every nonterminal A

$h(a) = a$ for every terminal a

$h(\varepsilon) = Y$



Coincidental Extension 4/6

P constructed by performing the next six steps:

- I.** add $(\bar{Z}) \rightarrow (YS\$)$ to P
- II.** for every $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n) \in P$, add $(A_1, \dots, A_n, \$) \rightarrow (h(x_1), \dots, h(x_n), \$)$ to P
- III.** add $(Y, \$) \rightarrow (YY, \$)$ to P
- IV.** for every $a, b, c \in T$,
add $(\langle a \rangle, \langle b \rangle, \langle c \rangle, \$) \rightarrow (\langle 0a \rangle, \langle 0b \rangle, \langle 0c \rangle, \$)$ to P
- V.** for every $a, b, c, d \in T$, add
 $(Y, \langle 0a \rangle, Y, \langle 0b \rangle, Y, \langle 0c \rangle, \$) \rightarrow (\#, \langle 0a \rangle, X, \langle 0b \rangle, Y, \langle 0c \rangle, \$)$,
 $(\langle 0a \rangle, \langle 0b \rangle, \langle 0c \rangle, \$) \rightarrow (\langle 4a \rangle, \langle 1b \rangle, \langle 2c \rangle, \$)$,
 $(\langle 4a \rangle, X, \langle 1b \rangle, Y, \langle 2c \rangle, \$) \rightarrow (\langle 4a \rangle, \#, \langle 1b \rangle, X, \langle 2c \rangle, \$)$,
 $(\langle 4a \rangle, \langle 1b \rangle, \langle 2c \rangle, \langle d \rangle, \$) \rightarrow (a, \langle 4b \rangle, \langle 1c \rangle, \langle 2d \rangle, \$)$,
 $(\langle 4a \rangle, \langle 1b \rangle, \langle 2c \rangle, \$) \rightarrow (a, \langle 1b \rangle, \langle 3c \rangle, \$)$,
 $(\langle 1a \rangle, X, \langle 3b \rangle, Y, \$) \rightarrow (\langle 1a \rangle, \#, \langle 3b \rangle, \#, \$)$
to P



Coincidental Extension 5/6

VI. for every $a, b \in T$, add

$(\langle 1a \rangle, X, \langle 3b \rangle, \S) \rightarrow (a, \#, b, \#)$ to P.

G generates every $y \in L(G)$ in this way

$$Z \Rightarrow YS\$ \Rightarrow^+ x\$ \Rightarrow v\$ \Rightarrow^+ z\$ \Rightarrow y$$

where $v \in (\mathcal{T}\{Y\}^+)^+\{\$\}$. In addition,

$$v = u_0\langle 0a_1 \rangle u_1\langle 0a_2 \rangle u_2\langle 0a_3 \rangle \dots u_{n-1}\langle a_n \rangle u_n\$$$

if and only if $a_1a_2a_3\dots a_n \in L(G)$

Coincidental Extension 6/6

In greater detail, $v\xi \Rightarrow^+ z\xi \Rightarrow y$ can be expressed as

$$\begin{aligned}
 & Y\langle 0a_1 \rangle Y\langle 0a_2 \rangle Y\langle 0a_3 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow^i & \#^i\langle 0a_1 \rangle X\langle 0a_2 \rangle Y\langle 0a_3 \rangle Y\langle a_4 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow & \#^i\langle 4a_1 \rangle X\langle 1a_2 \rangle Y\langle 2a_3 \rangle Y\langle a_4 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow^i & \#^i\langle 4a_1 \rangle \#^i\langle 1a_2 \rangle X\langle 2a_3 \rangle Y\langle a_4 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow & \#^i a_1 \#^i\langle 4a_2 \rangle X\langle 1a_3 \rangle Y\langle 2a_4 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow^i & \#^i a_1 \#^i\langle 4a_2 \rangle \#^i\langle 1a_3 \rangle X\langle 2a_4 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow & \#^i a_1 \#^i a_2 \#^i\langle 4a_3 \rangle X\langle 1a_4 \rangle Y\langle 2a_5 \rangle \dots Y\langle a_n \rangle Y^{\neq 1}\xi \\
 & \vdots \\
 & \#^i a_1 \#^i a_2 \#^i a_3 \dots \langle 4a_{n-2} \rangle \#^i\langle 1a_{n-1} \rangle X\langle 2a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow & \#^i a_1 \#^i a_2 \#^i a_3 \dots a_{n-2} \#^i\langle 1a_{n-1} \rangle X\langle 3a_n \rangle Y^{\neq 1}\xi \\
 \Rightarrow^{i-1} & \#^i a_1 \#^i a_2 \#^i a_3 \dots \#^i a_{n-2} \#^i\langle 1a_{n-1} \rangle \#^j X\langle 3a_n \rangle \#^{\neq 1}\xi \\
 \Rightarrow & \#^i a_1 \#^i a_2 \#^i a_3 \dots \#^i a_{n-2} \#^i a_{n-1} \#^i a_n \#^i
 \end{aligned}$$

- Corollary:** Let $K \in RE$. Then, there exists a ε -free SCG, G , such that $K \# \blacktriangleleft L(G)$.



Use in Theoretical Computer Science

Use in Theoretical Computer Science

- **Corollary:** For every language $K \in RE$, there exists a homomorphism h and a language $H \in \varepsilon\text{-free } \mathbf{SC}$ such that $K = h(H)$.
- In a complex way, this result was proved on page 245 in [Greibach, S. A. and Hopcroft, J. E.: Scattered Context Grammars. *J. Comput. Syst. Sci.* 3, 232-247 (1969)]



Future Investigation

Future Investigation: *k-limited coincidental extension*

- Let k be a non-negative integer.
- For a symbol, $\#$, and a string, $x = a_1 a_2 \dots a_{n-1} a_n$, any string of the form $\#^i a_1 \#^i a_2 \#^i \dots \#^i a_{n-1} \#^i a_n \#^i$, where $k \geq i \geq 0$, is a *k-limited coincidental #-extension* of x .
- A language, K , is a coincidental a *k-limited #-extension* of L if every string of K represents a *k-limited coincidental extension* of a string in L and the deletion of all $\#$ s in K results in L , symbolically written as $L \stackrel{k \geq \#}{\blacktriangleleft} K$

Example

- For $X = \{\#^i a \#^i b \#^i : 2 \geq i \geq 0\} \cup \{\#^i c_n \#^i d_n \#^i : n \geq 0, 4 \geq i \geq 0\}$ and $Y = \{ab\} \cup \{c_n d_n : n \geq 0\}$,

$$Y \stackrel{4 \geq \#}{\blacktriangleleft} X$$



Very Important Open Problem

Important Open Problem: ε -free **SC** = **CS**?

- Does there exist a non-negative integer k , such that for every $L \in CS$, $L \stackrel{k \geq \#}{\leftarrow} L(H)$ for some ε -free SCG, H ?
- If so, I know how to prove ε -free **SC** = **CS** 😊.

END