

# **K-LIMITED ERASING PERFORMED BY SCATTERED CONTEXT GRAMMARS**

**Jiří Techet**

Doctoral Degree Programme (2), FIT BUT

E-mail: [techet@fit.vutbr.cz](mailto:techet@fit.vutbr.cz)

Supervised by: Alexander Meduna

E-mail: [meduna@fit.vutbr.cz](mailto:meduna@fit.vutbr.cz)

## **ABSTRACT**

A scattered context grammar,  $G$ , erases nonterminals in a  $k$ -limited way, where  $k \geq 1$ , if in every sentential form of any derivation, between every two symbols from which  $G$  derives non-empty strings, there occur no more than  $k$  nonterminals from which  $G$  derives empty words. This paper demonstrates that any scattered context grammar that erases nonterminals in this way can be converted to an equivalent scattered context grammar without any erasing productions while in general, this conversion is impossible.

## **1 INTRODUCTION**

This paper discusses scattered context grammars, which represent an important type of semi-parallel grammars (see [1, 2, 3, 4, 6, 8, 9]). It concentrates its investigation on the role of erasing productions and the way they are applied in these grammars. While scattered context grammars with erasing productions characterize the family of recursively enumerable languages, the same grammars without erasing productions cannot generate any non-context-sensitive language (see [3, 4]). As a result, in general, we cannot convert any scattered context grammar with erasing productions to an equivalent scattered context grammar without these productions. In this paper, we demonstrate that this condition is always possible if the original grammar *erases its nonterminals in a  $k$ -limited way*, where  $k$  is a positive integer; in every sentential form of any derivation, between any two symbols from which the grammar derives non-empty strings, there occur no more than  $k$  nonterminals from which the grammar derives empty strings later in the derivation. Consequently, the scattered grammars that have erasing productions but apply them in a  $k$ -limited way are equivalent to the same grammars that do not have erasing productions at all.

In [3] it was demonstrated that a language generated by propagating scattered context grammars is closed under restricted homomorphism. Note that our definition of  $k$ -limited erasing differs significantly from the definition of restricted homomorphism. While in case of restricted homomorphism a language can be generated by a propagated scattered context grammar in case that at most  $k$  symbols are deleted between every two terminals in a sentence, in case of a scattered context grammar which erases its nonterminals in a  $k$ -limited way virtually unlimited number of symbols can be deleted between every two terminals in a sentence in case that during the

derivation process there are always at most  $k$  erasable symbols between two non-erasable symbols. Therefore, the result presented in this paper represents a generalization of the previously published result.

## 2 PRELIMINARIES

We assume that the reader is familiar with the language theory (see [5, 7]).  $V^*$  represents the free monoid generated by  $V$  under the operation of concatenation. The unit of  $V^*$  is denoted by  $\epsilon$ . Set  $V^+ = V^* - \{\epsilon\}$ . For  $w \in V^*$ ,  $|w|$  and  $\text{alph}(w)$  denote the length of  $w$  and the set of symbols occurring in  $w$ , respectively. For  $L \subseteq V^*$ ,  $\text{alph}(L) = \{a : a \in \text{alph}(w), w \in L\}$ . Let  $\text{pos}(a_1 \dots a_i \dots a_n, i) = a_i$  for  $1 \leq i \leq n$ ,  $a_1 \dots a_n \in V^*$ .

A *context-free grammar* (see [5]), a *CFG* for short, is a quadruple,  $G = (V, T, P, S)$ , where  $V$  is an alphabet,  $T \subseteq V$ ,  $S \in V - T$ , and  $P$  is a finite set of productions such that each production has the form  $A \rightarrow x$ , where  $A \in V - T$ ,  $x \in V^*$ . Let  $\text{lhs}(A \rightarrow x)$  and  $\text{rhs}(A \rightarrow x)$  denote  $A$  and  $x$ , respectively. If  $A \rightarrow x \in P$ ,  $u = rAs$ , and  $v = rxs$ , where  $r, s \in V^*$ , then  $G$  makes a *derivation step* from  $u$  to  $v$  according to  $A \rightarrow x$ , symbolically written as  $u \Rightarrow v [A \rightarrow x]$  in  $G$  or, simply,  $u \Rightarrow v$ . Let  $\Rightarrow^+$  and  $\Rightarrow^*$  denote the transitive closure of  $\Rightarrow$  and the transitive-reflexive closure of  $\Rightarrow$ , respectively. The *language of  $G$*  is denoted by  $L(G)$  and defined as  $L(G) = \{x : x \in T^*, S \Rightarrow^* x\}$ .

## 3 DEFINITIONS AND EXAMPLES

A *scattered context grammar* (see [1, 2, 3, 4, 6, 8, 9]), a *SCG* for short, is a quadruple,  $G = (V, T, P, S)$ , where  $V$  is an alphabet,  $T \subseteq V$ ,  $S \in V - T$ , and  $P$  is a finite set of productions such that each production has the form  $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)$ , for some  $n \geq 1$ , where  $A_i \in V - T$ ,  $x_i \in V^*$ , for  $1 \leq i \leq n$ . If every production  $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n) \in P$  satisfies  $x_i \in V^+$  for all  $1 \leq i \leq n$ ,  $G$  is a *propagating scattered context grammar*, a *PSCG* for short. If  $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n) \in P$ ,  $u = u_1 A_1 u_2 \dots u_n A_n u_{n+1}$ , and  $v = u_1 x_1 u_2 \dots u_n x_n u_{n+1}$ , where  $u_i \in V^*$ ,  $1 \leq i \leq n$ , then  $G$  makes a *derivation step* from  $u$  to  $v$  according to  $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)$ , symbolically written as  $u \Rightarrow v [(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)]$  in  $G$  or, simply,  $u \Rightarrow v$ . Set  $\pi((A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)) = |A_1 \dots A_n| = n$  and  $\rho((A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)) = \{A_1 \rightarrow x_1, \dots, A_n \rightarrow x_n\}$ . Let  $\Rightarrow^+$  and  $\Rightarrow^*$  denote the transitive closure of  $\Rightarrow$  and the transitive-reflexive closure of  $\Rightarrow$ , respectively. The *language of  $G$*  is denoted by  $L(G)$  and defined as  $L(G) = \{x : x \in T^*, S \Rightarrow^* x\}$ .

The *core grammar underlying a scattered context grammar*,  $G = (V, T, P, S)$ , is denoted by  $\text{core}(G)$  and defined as the context-free grammar  $\text{core}(G) = (V, T, \text{cf}(P), S)$  with  $\text{cf}(P) = \{B \rightarrow y : B \rightarrow y \in \rho(p) \text{ for some } p \in P\}$ . Let  $v = u_1 A_1 u_2 A_2 \dots u_n A_n u_{n+1} \Rightarrow u_1 x_1 u_2 x_2 \dots u_n x_n u_{n+1} = w [(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)]$  in  $G$ . The *partial  $m$ -step context-free simulation* of this step by  $\text{core}(G)$  is denoted by  $\text{pcf}_m(v \Rightarrow w)$  and defined as  $\text{core}(G)$ 's  $m$ -step derivation of the form  $u_1 A_1 u_2 A_2 \dots u_n A_n u_{n+1} \Rightarrow u_1 x_1 u_2 A_2 \dots u_n A_n u_{n+1} \Rightarrow \dots \Rightarrow u_1 x_1 u_2 x_2 \dots u_m x_m u_{m+1} A_{m+1} \dots u_n A_n u_{n+1}$  where  $m \leq n$ . The *context-free simulation* is a special case of the partial  $m$ -step context-free simulation for  $m = n$ , denoted by  $\text{cf}(v \Rightarrow w)$ . Let  $v = v_1 \Rightarrow^* v_n = w$  be a derivation in  $G$  of the form  $v_1 \Rightarrow v_2 \Rightarrow v_3 \Rightarrow \dots \Rightarrow v_n$ . The context-free simulation of  $v \Rightarrow^* w$  by  $\text{core}(G)$  is denoted as  $\text{cf}(v \Rightarrow^* w)$  and defined as  $v_1 \Rightarrow^* v_2 \Rightarrow^* v_3 \Rightarrow^* \dots \Rightarrow^* v_n$  such that for all  $1 \leq i \leq n - 1$ ,  $v_i \Rightarrow^* v_{i+1}$  in  $\text{core}(G)$  is the context-free simulation of  $v_i \Rightarrow v_{i+1}$  in  $G$ . Let  $S \Rightarrow^* x$  in  $G$  be of the form  $S \Rightarrow^* uAv \Rightarrow^* x$ . Let  $\text{cf}(S \Rightarrow^* x)$  in  $\text{core}(G)$  be the context-free simulation of  $S \Rightarrow^*$

$x$  in  $G$ . Let  $t$  be the derivation tree corresponding to  $S \Rightarrow^* x$  in  $\text{core}(G)$  (regarding derivation trees and related notions, we use the terminology of [5]). Consider a subtree rooted at  $A$  in  $t$ . If the frontier of this subtree is  $\varepsilon$ , then  $G$  erases  $A$  in  $S \Rightarrow^* uAv \Rightarrow^* x$ , symbolically written as  $\check{A}$ , and if this frontier differs from  $\varepsilon$ , then  $G$  does not erase  $A$  during this derivation, symbolically written as  $\hat{A}$ . If  $w = \hat{A}_1 \dots \hat{A}_n$  or  $w = \check{A}_1 \dots \check{A}_n$ , we write  $\hat{w}$  or  $\check{w}$ , respectively. Let  $G = (V, T, P, S)$  be a SCG, and let  $k \geq 0$ .  $G$  erases its nonterminals in a  $k$ -limited way if for every  $y \in L(G)$  there exists a derivation  $S \Rightarrow^* y$  such that every sentential form  $x$  of the derivation satisfies the following two properties:

1. Every  $x = uAvBw, \hat{A}, \hat{B}, \check{v}$ , satisfies  $|v| \leq k$ .
2. Every  $x = uAw, \hat{A}$ , satisfies: if  $\check{u}$  or  $\check{w}$ , then  $|u| \leq k$  or  $|w| \leq k$ , respectively.

### Examples

1. Observe that the grammar  $G_1 = (\{S, A, B, C, A', B', C', a, b, c\}, \{a, b, c\}, \{(S) \rightarrow (ABC), (A) \rightarrow (aAA'), (B) \rightarrow (bBB'), (C) \rightarrow (cCC'), (A, B, C) \rightarrow (y, y, y), (A', B', C') \rightarrow (y, y, y)\}, S)$  generates the language  $L(G_1) = \{a^n y^{n+1} b^n y^{n+1} c^n y^{n+1} : n \geq 0\}$ . Therefore, there does not exist any restricted homomorphism  $h$  such that  $h(L(G_1)) = \{a^n b^n c^n : n \geq 0\}$ . However, as demonstrated by the following example, there exists a scattered context grammar which erases its nonterminals in a  $k$ -limited way.
2. Observe that the grammar  $G_2 = (\{S, A, B, C, A', B', C', a, b, c\}, \{a, b, c\}, \{(S) \rightarrow (ABC), (A) \rightarrow (aAA'), (B) \rightarrow (bBB'), (C) \rightarrow (cCC'), (A, B, C) \rightarrow (\varepsilon, \varepsilon, \varepsilon), (A', B', C') \rightarrow (\varepsilon, \varepsilon, \varepsilon)\}, S)$  generates the language  $L(G_2) = \{a^n b^n c^n : n \geq 0\}$ . As the derivation of any string  $aa \dots aabb \dots bbcc \dots cc \in L(G_2)$  may be of the form

$$\begin{aligned} S &\Rightarrow ABC \Rightarrow^* aAA' bBB' cCC' \Rightarrow aAbBcC \\ &\Rightarrow^* aaAA' bbBB' ccCC' \Rightarrow aaAbbBccC \\ &\Rightarrow^* aa \dots aaAA' bb \dots bbBB' cc \dots ccCC' \\ &\Rightarrow aa \dots aaAbb \dots bbBcc \dots ccC \Rightarrow aa \dots aabb \dots bbcc \dots cc, \end{aligned}$$

the grammar erases its nonterminals in a 2-limited way.

3. Consider the grammar  $G_3 = (\{S, A, B, A', B', a, b, c\}, \{a, b, c\}, \{(S) \rightarrow (AA), (A, A) \rightarrow (aA, A'A), (A, A) \rightarrow (B, B), (B, B) \rightarrow (bBc, B'B), (B, B) \rightarrow (\varepsilon, \varepsilon), (A', B') \rightarrow (\varepsilon, \varepsilon)\}, S)$ . Observe that  $L(G_3) = L(G_2)$ . However, because the first part of every derivation has the form

$$S \Rightarrow AA \Rightarrow aAA'A \Rightarrow aaAA'A'A \Rightarrow^* aa \dots aaA'A' \dots A'A$$

and all  $A'$ 's are deleted in the second part of the derivation, there does not exist any  $k$  such that  $G_3$  erases its nonterminals in a  $k$ -limited way.

## 4 RESULTS

The main result of this paper follows next.

**Theorem 1.** *For every SCG,  $G$ , which erases its nonterminals in a  $k$ -limited way there exists a PSCG,  $\bar{G}$ , such that  $L(G) = L(\bar{G})$ .*

*Proof.* Let  $G = (V, T, P, S)$  be a SCG which erases its nonterminals in a  $k$ -limited way. For every  $p = (A_1, \dots, A_i, \dots, A_n) \rightarrow (x_1, \dots, x_i, \dots, x_n) \in P$  let  $\lfloor p, i \rfloor$  denote  $A_i \rightarrow x_i$  for all  $1 \leq i \leq n$ . Let  $\Psi = \{\lfloor p, i \rfloor : p \in P, 1 \leq i \leq \pi(p)\}$  and  $\Psi' = \{\lfloor p, i \rfloor' : \lfloor p, i \rfloor \in \Psi\}$ . Set  $\bar{N}_1 = \{\langle x \rangle : x \in (V - T)^* \cup (V - T)^* T (V - T)^*, |x| \leq 2k + 1\}$ . For every  $\langle x \rangle \in \bar{N}_1$  and  $\lfloor p, i \rfloor \in \Psi$ , define

$$\text{lhs-replace}(\langle x \rangle, \lfloor p, i \rfloor) = \{\langle x_1 \lfloor p, i \rfloor x_2 \rangle : x_1, x_2 \in V^*, x_1 \text{ lhs}(\lfloor p, i \rfloor) x_2 = x\}.$$

Set  $\bar{N}_2 = \{\langle x \rangle : \langle x \rangle = \text{lhs-replace}(\langle y \rangle, \lfloor p, i \rfloor), \langle y \rangle \in \bar{N}_1, \lfloor p, i \rfloor \in \Psi\}$ . For every  $\langle x \rangle \in \bar{N}_1$  and  $\lfloor p, i \rfloor' \in \Psi'$ , define

$$\text{insert}(\langle x \rangle, \lfloor p, i \rfloor') = \{\langle x_1 \lfloor p, i \rfloor' x_2 \rangle : x_1, x_2 \in V^*, x_1 x_2 = x\}.$$

Set  $\bar{N}_2' = \{\langle x \rangle : \langle x \rangle = \text{insert}(\langle y \rangle, \lfloor p, i \rfloor'), \langle y \rangle \in \bar{N}_1, \lfloor p, i \rfloor' \in \Psi'\}$ . For every  $x = \langle x_1 \rangle \langle x_2 \rangle \dots \langle x_n \rangle \in (\bar{N}_1 \cup \bar{N}_2 \cup \bar{N}_2')^*$  for some  $n \geq 1$ , define

$$\text{join}(x) = x_1 x_2 \dots x_n.$$

For every  $x \in \bar{N}_1 \cup \bar{N}_2 \cup \bar{N}_2'$ , define

$$\text{split}(x) = \{y : x = \text{join}(y)\}.$$

Set  $\bar{V} = T \cup \bar{N}_1 \cup \bar{N}_2 \cup \bar{N}_2' \cup \{\bar{S}\}$ . Define the PSCG,

$$\bar{G} = (\bar{V}, T, \bar{P}, \bar{S}),$$

with  $\bar{P}$  constructed as follows:

1. For every  $p = (S) \rightarrow (x) \in P$ , add  $(\bar{S}) \rightarrow (\langle \lfloor p, 1 \rfloor \rangle)$  to  $\bar{P}$ ;
2. For every  $\langle x \rangle \in \bar{N}_1$ , every  $X \in \text{insert}(\langle x \rangle, \lfloor p, n \rfloor')$ , where  $p \in P$ ,  $\pi(p) = n$ , every  $\langle y \rangle \in \bar{N}_1$ , and every  $Y \in \text{lhs-replace}(\langle y \rangle, \lfloor q, 1 \rfloor)$ , where  $q \in P$ , add
  - (a)  $(X, \langle y \rangle) \rightarrow (\langle x \rangle, Y)$ , and
  - (b)  $(\langle y \rangle, X) \rightarrow (Y, \langle x \rangle)$  to  $\bar{P}$ ;
  - (c) if  $\langle x \rangle = \langle y \rangle$ , add  $(X) \rightarrow (Y)$  to  $\bar{P}$ ;
  - (d)  $(X) \rightarrow (\langle x \rangle)$  to  $\bar{P}$ ;
3. For every  $\langle x \rangle \in \bar{N}_1$ , every  $X \in \text{insert}(\langle x \rangle, \lfloor p, i \rfloor')$ , where  $p \in P$ ,  $i < \pi(p)$ , every  $\langle y \rangle \in \bar{N}_1$ , and every  $Y \in \text{lhs-replace}(\langle y \rangle, \lfloor p, i + 1 \rfloor)$ , where  $q \in P$ , add
  - (a)  $(X, \langle y \rangle) \rightarrow (\langle x \rangle, Y)$  to  $\bar{P}$ ;
  - (b) if  $\langle x \rangle = \langle y \rangle$  and  $\text{pos}(X, l) = \lfloor p, i \rfloor'$ ,  $\text{pos}(Y, m) = \lfloor p, i + 1 \rfloor'$ ,  $l < m$ , add  $(X) \rightarrow (Y)$  to  $\bar{P}$ ;
4. For every  $\langle x_1 \lfloor p, i \rfloor x_2 \rangle \in \text{lhs-replace}(\langle x \rangle, \lfloor p, i \rfloor)$ ,  $\langle x \rangle \in \bar{N}_1$ ,  $\lfloor p, i \rfloor \in \Psi$ ,  $x_1, x_2 \in V^*$ , and every  $Y \in \text{split}(x_1 \text{ rhs}(\lfloor p, i \rfloor) \lfloor p, i \rfloor' x_2)$ , add  $(\langle x_1 \lfloor p, i \rfloor x_2 \rangle) \rightarrow (Y)$  to  $\bar{P}$ ;
5. For every  $a \in T$ , add  $(\langle a \rangle) \rightarrow (a)$  to  $\bar{P}$ .

Denote the set of productions introduced in step  $i$  of the construction by  ${}_i P$ , for  $1 \leq i \leq 5$ .

**Basic Idea**  $\bar{G}$  simulates  $G$  by using nonterminals of the form  $\langle \dots \rangle$ . In each nonterminal of this form, during every simulated derivation step,  $\bar{G}$  records a substring of the corresponding current sentential form of  $G$ .

The rule constructed in (1) only initializes the simulation process. By rules introduced in (2) through (4),  $\bar{G}$  simulates the application of a scattered context rule  $p$  from  $P$  in a left-to-right way. In greater detail, by using a rule of (2),  $\bar{G}$  nondeterministically selects a scattered context rule  $p$  from  $P$ . Suppose that  $p$  consists of context-free rules  $r_1, \dots, r_{i-1}, r_i, \dots, r_n$ . By using rules of (3) and (4),  $\bar{G}$  simulates the application of  $r_1$  through  $r_n$  one by one. To explain this in greater detail, suppose that  $\bar{G}$  has just completed the simulation of  $r_{i-1}$ . Then, to the right of this simulation,  $\bar{G}$  selects  $\text{lhs}(r_i)$  by using a rule of (3). That is, this selection is made inside of  $\bar{G}$ 's nonterminal in which the simulation of  $r_{i-1}$  has been performed or in one of the nonterminals appearing to the right of this nonterminal. After this selection, by using a rule of (4),  $\bar{G}$  performs the replacement of the selected symbol  $\text{lhs}(r_i)$  with  $\text{rhs}(r_i)$ .

If a terminal occurs inside of a nonterminal of  $\bar{G}$ , then a rule of (5) allows  $\bar{G}$  to change this nonterminal to the terminal string contained in it.

**Formal Proof** Due to the requirements imposed on the length of this paper, the formal proof is omitted. ■

## ACKNOWLEDGEMENTS

This work was supported by GAČR grant 102/05/H050 and FRVŠ grant FR762/2007/G1.

## REFERENCES

- [1] Fernau, H.: Scattered context grammars with regulation, *Annals of Bucharest Univ., Math.-Informatics Series*, 45(1), 41–49, 1996
- [2] Gonczarowski, J., Warmuth, M.: Scattered versus context-sensitive rewriting, *Acta Informatica*, 27, 81–95, 1989
- [3] Greibach, S., Hopcroft, J.: Scattered context grammars, *Journal of Computer and System Sciences*, 3, 233–247, 1969
- [4] Meduna, A.: A trivial method of characterizing the family of recursively enumerable languages by scattered context grammars, *EATCS Bulletin*, 56, 104–106, 1995
- [5] Meduna, A.: *Automata and Languages: Theory and Applications*, Springer, London, 2000
- [6] Meduna, A., Techet, J.: Generation of sentences with their parses: the case of propagating scattered context grammars, *Acta Cybernetica*, 17, 11–20, 2005
- [7] Salomaa, A.: *Formal Languages*, Academic Press, London, 1973
- [8] Vaszil, G.: On the descriptive complexity of some rewriting mechanisms regulated by context conditions, *Theoretical Computer Science*, 330, 361–373, 2005
- [9] Virkkunen, V.: On scattered context grammars, *Acta Universitatis Ouluensis, Series A, Mathematica* 6, 75–82, 1973