ELSEVIER

Theoretical Computer Science 246 (2000) 279–284

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

Generative power of three-nonterminal scattered context grammars

Alexander Meduna

Department of Computer Science and Engineering, Technical University of Brno, Božetěchova 2, Brno 61266, Czech Republic

Received 15 September 1999; revised 11 January 2000; accepted 28 February 2000

Abstract

This paper discusses the descriptional complexity of scattered context grammars with respect to the number of nonterminals. It proves that the three-nonterminal scattered context grammars characterize the family of recursively enumerable languages. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

Recently, the formal language theory has intensively investigated the descriptional complexity of grammars with respect to the number of nonterminals (see [6, 7]). This investigation has achieved several characterizations of the family of recursively enumerable languages by various grammars with a reduced number of nonterminals. Specifically, this family was characterized by four-nonterminal scattered context grammars (see [3]). The present paper improves this result by demonstrating that even the three-nonterminal scattered context grammars characterize the family of recursively enumerable languages.

2. Definitions

We assume that the reader is familiar with the language theory (see [1, 5]).

Let V be an alphabet. The cardinality of V is denoted by $card(V)$. V^* represents the free monoid generated by V under the operation of concatenation. The unit of V^* is denoted by ε. We set $V^+ = V^* - \{\varepsilon\}$; algebraically, V^+ is thus the free semigroup

E-mail address: meduna@dcse.fee.vutbr.cz (A. Meduna).

PII: S0304-3975(00)00153-5

generated by V under the operation of concatenation. For $w \in V^*$, $|w|$ denotes the length of w. For $a \in V$ and $w \in V^*$, $occur(a, w)$ denotes the number of occurrences of a in w.

A *scattered context grammar* is a quadruple, $G = (V, P, S, T)$, where V is an alphabet, $T \subseteq V$, $S \in V - T$, and P is a finite set of productions of the form $(A_1, A_2, \ldots, A_n) \to (x_1, x_2, \ldots, x_n)$, where n is a positive integer, and $A_i \in V - T$, $x_i \in V^*$, for $i = 1, 2, \ldots, n$. Let $p \in P$ be a production of the above form; then, $left(p)$ and $right(p)$ denote $A_1 A_2 \ldots A_n$ and $x_1 x_2 \ldots x_n$, respectively. If $p \in P$ is of the form $(A_1, A_2, \ldots, A_n) \to (x_1, x_2, \ldots, x_n)$, $u = u_1 A_1 u_2 A_2 \ldots u_n A_n u_{n+1}$, $v = u_1 x_1 u_2 x_2 \ldots u_n x_n u_{n+1}$, where $u_i \in V^*$, for $i = 1, 2, \ldots, n$, then u directly derives v according to p, denoted by $u \Rightarrow v[p]$ or, simply, $u \Rightarrow v$. In a standard manner, we extend $\Rightarrow$ to $\Rightarrow^n$, where $n \geqslant 0$, and based on $\Rightarrow^n$, we define $\Rightarrow^*$. Let $S \Rightarrow^* x$ with $x \in T^*$, then $S \Rightarrow^* x$ is called a successful derivation. The language of $G, L(G)$, is defined as $L(G) = \{x\colon\ S \Rightarrow^* x \text{ with } x \in T^*\}$.

A *queue grammar* (see [2]) is a sixtuple, $Q = (V, T, W, F, R, g)$, where V and W are two disjoint alphabets, $T \subseteq V$, $F \subseteq W$, $R \in (V - T)(W - F)$, and $g \subseteq (V \times (W - F)) \times (V^* \times W)$ is a finite relation such that for every $a \in V$, there exists an element $(a, b, x, c) \in g$. If there exist $u, v \in V^* W$, $a \in V$, $r, z \in V^*$, and $b, c \in W$ such that $(a, b, z, c) \in g$, $u = arb$, and $v = rzc$, then u directly derives v according to (a, b, z, c), denoted by $u \Rightarrow v$ $[(a, b, z, c)]$ or, simply, $u \Rightarrow v$. In the standard manner, we extend $\Rightarrow$ to $\Rightarrow^n$ and $\Rightarrow^*$ Let $R \Rightarrow^* xq$ in Q with $x \in T^*$ and $q \in F$, then $R \Rightarrow^* xq$ is called a successful derivation in Q. The language of Q, $L(Q)$, is defined as $L(Q) = \{x\colon\ S \Rightarrow^* xq$ with $x \in T^*$ and $q \in F\}$.

Let n be a positive integer. Set $\mathbf{SC}_n = \{L\colon\ L = L(G)$, where $G = (V, P, S, T)$ is a scattered context grammar such that $card(V - T) \leqslant n\}$. Let $\mathbf{RE}$ denote the family of recursively enumerable language.

3. Results

This section demonstrates that $\mathbf{RE} = \mathbf{SC}_3$.

Lemma 1. *For any queue grammar, Q', there exists an equivalent queue grammar, $Q = (V, T, W, F, R, g)$, such that Q generates every $z \in L(Q)$ by the derivation of the form $R \Rightarrow^i u \Rightarrow v \Rightarrow^k w \Rightarrow z$, where $i, k \geqslant 1$, and the derivation satisfies the following properties* 1–4:

1. *each derivation step in $R \Rightarrow^i u$ has the form $a'y'b' \Rightarrow a'y'x'b'$ $[(a', b', x', c')]$, where $a' \in V - T$, $b', c' \in Q - F$, $x', y' \in (V - T)^*$;*
2. *in greater detail, the derivation step $u \Rightarrow v$ has this form $a''y''b'' \Rightarrow a''y''h''x''b''$ $[(a'', b'', h''x'', c'')]$, where $a' \in V - T$, b', $c' \in Q - F$, $h'', y'' \in (V - T)^*$, $x'' \in T^*$;*
3. *each derivation step in $v \Rightarrow^k w$ has the form $a'''y'''h'''b''' \Rightarrow a'''y'''h'''x'''b'''$ $[(a''', b''', x''', c''')]$, where $a''' \in V - T$, $b''', c''' \in Q - F$, $y''' \in (V - T)^*$, $x''', h''' \in T^*$;*

4. *in greater detail, the derivation step* $w \Rightarrow z$ *has the form* $a''''y''''b'''' \Rightarrow y''''x''''c''''$ $[(a'''', b'''', x'''', c'''')]$, *where* $a'''' \in V - T$, $b'''' \in Q - F$, $y'''', x'''' \in T^*$, $w = a''''y''''b''''$, $z = y''''x''''$.

Proof. Let $Q' = (V', T', W', F', R', g')$ be any queue grammar. Introduce these four pairwise disjoint alphabets U, X, Y, and $\{@, \$, \#, \perp\}$ so that $card(U) = card(V')$ and $card(X) = card(Y) = card(W')$. Introduce any bijection, α, from $(V' \cup W')$ onto $(U \cup X)$. Furthermore, introduce another bijection, β, from W' to Y. Set $V = U \cup T' \cup \{@, \#\}$, $T = T'$, $W = X \cup Y \cup \{\$, \perp\}$, $F = \{\perp\}$, and $R = @\$$. Define the queue grammar $Q = (V, T, W, F, R, g)$ with g constructed in the following five-step way:

I. if $R = ab$ with $a \in V - T$ and $b \in W - F$, then add $(@, \$, a, b)$ to g;
II. for every $(a, b, x, c) \in g$ with $a \in V$, $x \in V^*$, and $b, c \in W$, add $(\alpha(a), \alpha(b), \alpha(x), \alpha(c))$ to g;
III. for every $(a, b, xy, c) \in g$ with $a \in V$, $x \in V^*$, $y \in T^*$, and $b, c \in W$, add $(\alpha(a), \alpha(b), \alpha(x)\ \#y, \beta(c))$ to g;
IV. for every $(a, b, y, c) \in g$ with $a \in V$, $y \in T^*$, and $b, c \in W$, add $(\alpha(a), \beta(b), y, \beta(c))$ to g;
V. for every $c \in F$, add $(\#, \beta(b), \varepsilon, \perp)$ to g.

A formal proof that Q satisfies the properties required by lemma is left to the reader. □

Lemma 2. *Let L be a recursively enumerable language. Then, there exists a three-nonterminal scattered context grammar,* $G = (T \cup \{0, 1, 2\}, P, 2, T)$, *satisfying* $L = L(G)$.

Proof. Let L be a recursively enumerable language. By Theorem 2.1 in [2], there exists a queue grammar, $Q = (V, T, W, F, R, g)$, such that $L(Q) = L$. Without any loss of generality, assume that Q satisfies the properties described in Lemma 1. The next construction produces a three-nonterminal scattered context grammar, G, satisfying $L(G) = L(Q)$.

Set $n = card(V \cup W) + 2$. Introduce a bijection, β, from $(V \cup W)$ to $(\{1\}^+\{0\}\{1\}^+ \cap \{0, 1\}^n)$. In a standard manner extend the domain of β to $(V \cup W)^*$. Without any loss of generality assume that $(V \cup W) \cap \{0, 1, 2\} = \emptyset$. Define the scattered context grammar, $G = (T \cup \{0, 1, 2\}, P, 2, T)$, where P is constructed in the following six-step way:

I. if $R = ab$ with $a \in V - T$ and $b \in W - F$, then add $(2) \rightarrow (01^{n-1}\beta(b)22\beta(a)20)$ to P;
II. for every $(a, b, x, c) \in g$ with $a \in V - T$, $x \in (V - T)^*$, and $b, c \in W - F$, add $(d_1, \ldots, d_n, b_1, \ldots, b_n, 2, a_1, \ldots, a_{n-1}, a_n, 2, 2) \rightarrow (d_1, \ldots, d_n, c_1, \ldots, c_n, e_1, e_2, \ldots, e_n, 2, 2, \beta(x)2)$ to P, where $d_1 \ldots d_n = 01^{n-1}$ (that is, $d_1 = 0$ and $d_h = 1$ for $h = 2, \ldots, n$), $b_1 \ldots b_n = \beta(b)$, $a_1 \ldots a_n = \beta(a)$, $c_1 \ldots c_n = \beta(c)$, $e_i = \varepsilon$ for $i = 1, \ldots, n$;
III. for every $(a, b, xy, c) \in g$ with $a \in V - T$, $x \in (V - T)^*$, $y \in T^*$, and $b, c \in W - F$, add $(d_1, \ldots, d_n, b_1, \ldots, b_n, 2, a_1, \ldots, a_{n-1}, a_n, 2, 2) \rightarrow (f_1, \ldots, f_n, c_1, \ldots, c_n, e_1, e_2, \ldots, e_n, 2, 2, \beta(x)y2)$ to P, where $d_1 \ldots d_n = 01^{n-1}$, (that is, $d_1 = 0$ and $d_h = 1$ for

$h=2,\ldots,n), f_1 \ldots f_n = 1^{n-1}0$ (that is, $f_n = 0$ and $f_h = 1$ for $h = 1,\ldots,n-1$), $b_1 \ldots b_n = \beta(b)$, $a_1 \ldots a_n = \beta(a)$, $c_1 \ldots c_n = \beta(c)$, $e_i = \varepsilon$ for $i = 1,\ldots,n$;

IV. for every $(a,b,y,c) \in g$ with $a \in V - T$, $y \in T^*$, and $b,c \in W - F$, add $(f_1,\ldots,f_n, b_1,\ldots,b_n, 2, a_1,\ldots,a_{n-1}, a_n, 2, 2) \rightarrow (f_1,\ldots,f_n, c_1,\ldots,c_n, e_1, e_2,\ldots,e_n, 2, 2, y2)$ to P, where $f_1 \ldots f_n = 1^{n-1}0$ (that is, $f_n = 0$ and $f_h = 1$ for $h = 1,\ldots,n-1$), $b_1 \ldots b_n = \beta(b), a_1 \ldots a_n = \beta(a)$, $c_1 \ldots c_n = \beta(c)$, $e_i = \varepsilon$ for $i = 1,\ldots,n$;

V. for every $(a,b,y,c) \in g$ with $a \in V - T$, $y \in T^*$, $b \in W - F$, and $c \in F$, add $(f_1, \ldots, f_n, b_1,\ldots,b_n, 2, a_1,\ldots,a_{n-1}, a_n, 2, 2) \rightarrow (e_1,\ldots,e_n, e_{n+1},\ldots,e_{2n}, e_{2n+1}, e_{2n+2},\ldots, e_{3n}, \varepsilon, \varepsilon, y)$ to P, where $f_1 \ldots f_n = 1^{n-1}0$ (that is, $f_n = 0$ and $f_h = 1$ for $h = 1,\ldots,n-1$), $b_1 \ldots b_n = \beta(b)$, $a_1 \ldots a_n = \beta(a)$, $e_i = \varepsilon$ for $i = 1,\ldots,3n$;

VI. add $(2,2,a,2) \rightarrow (2,\varepsilon,a2,2)$ to P, where $a \in \{0,1\}$.

To keep this proof readable omit some obvious details from the rest of this proof whose completion is left to the reader.

Claim 1. *Let $2 \Rightarrow^* x$ in G be a derivation in G during which G uses the production introduced in step* I *i times, for some $i \geqslant 1$. Then, $occur(2,x) = (1+2i) - 3j$, $occur(1,x) = (n-1)k$, and $occur(0,x) = k + i - j$, where k is a non-negative integer and j is the number of applications of a production introduced in step* V *during $2 \Rightarrow^* x$ such that $j \geqslant 1$ and $(1+2i) \geqslant 3j$.*

Claim 2. *Let $2 \Rightarrow^* x$ in G be a derivation in G during which G uses the production introduced in step* I *two or more times. Then, $x \notin T^*$.*

Proof 2. Let $2 \Rightarrow^* x$ in G be a derivation in G. If G uses the production introduced in step I two or more times during $2 \Rightarrow^* x$, then the previous claim implies that x contains some occurrences of 0. Thus, $x \notin T^*$ because 0 is a nonterminal. □

Claim 3. *G generates every $w \in L(G)$ as $2 \Rightarrow u\,[p] \Rightarrow^* v \Rightarrow w\,[q]$, where p is the production introduced in I, q is a production introduced in V, during $u \Rightarrow^* v$, G makes every derivation step by a production introduced in* II–IV, *or* VI.

Proof 3. Let $w \in L(G)$. Then, $2 \Rightarrow^* w$ in G and $w \in T^*$. By Claim 1, as $w \in T^*$, G uses the production introduced in I once. Because $2 \Rightarrow^* w$ begins from 2, we can express $2 \Rightarrow^* w$ as $2 \Rightarrow u\,[p] \Rightarrow^* w$, where p is the production introduced in I, and during $u \Rightarrow^* w$, G never uses the production introduced in I. Observe that every production, r, introduced in II–IV, and VI satisfies $occur(left(r),2) = 3$ and $occur(right(r),2) = 3$. Furthermore, notice that every production, q, introduced in V, satisfies $occur(left(q),2) = 3$ and $occur(right(q),2) = 0$. These observations imply $2 \Rightarrow u\,[p] \Rightarrow^* v \Rightarrow w\,[q]$ in G, where p is the production introduced in I, q is a production introduced in V, during $u \Rightarrow^* v$, G makes every step by a production introduced in II–IV, or VI. □

Before describing the form of every successful derivation in G in greater detail, we make some observations about the use of productions introduced in VI.

During any successful derivation in G, a production introduced in step VI is always applied after using a production introduced in steps I–IV (the use of these productions is described below). More precisely, to continue the derivation after applying a production introduced in I–IV, G has to shift the second appearance of 2 right in the current sentential form. G makes this shift by using productions introduced in VI to generate a sentential form having precisely n appearances of d $(d \in \{0,1\})$ between the first appearance of 2 and the second appearance of 2. Indeed, the sentential form has to contain exactly n appearances of d between the first appearance of 2 and the second appearance of 2; otherwise, the successfulness of the derivation is contradicted by arguments O.1 and O.2, which follow next.

O.1. If there exist fewer than n d's between the first appearance of 2 and the second appearance of 2, no rule introduced in I–V can be used, so the derivation ends. If the last sentential form contains nonterminals and if the derivation is not successful, it is a contradiction.

O.2. Assume that there exist more than n d's between the first appearance of 2 and the second appearance of 2. Then, after the next application of a rule introduced in I–V, more than $3n$ d's $(d \in \{0,1\})$ appear before the first appearance of 2. Return to the construction of productions in G to make observations O.2.1–O.2.3:

O.2.1. The production introduced in step I is always used only in the first step of a successful derivation (see Claim 3).

O.2.2. All productions introduced in steps II–IV rewrite $3n$ nonterminals preceding the first appearance of 2 with other $3n$ nonterminals.

O.2.3. Recall that a production introduced in step V is always used in the last derivation step (see Claim 3); furthermore, observe that this production erase precisely $3n$ non-terminals preceding the first appearance of 2.

By observations O.2.1–O.2.3, the occurrence of more than $3n$ d's between the first and second appearance of 2 gives rise to a contradiction of the successfulness of the derivation.

By arguments O.1 and O.2, we see that the sentential form has to contain precisely n appearances of d between the first and second appearance of 2.

Except for the use of productions introduced in step VI (this use is explained above), every successful derivation in G is made as $2 \Rightarrow rhs(p_1)\ [p_1] \Rightarrow^i u \Rightarrow v\,[p_3] \Rightarrow^k w \Rightarrow z$ $[p_5]$, where $i,k \geqslant 1$, and the derivation satisfies the following properties A–D:

A. Each derivation step in $rhs(p_1) \Rightarrow^i u$ has this form $01^{n-1}\beta(b')2\beta(a')2\beta(y')20 \Rightarrow 01^{n-1}\beta(c')22\beta(y'x')20\ [p_2]$, where p_2 is a production introduced in II, $(a',b',x',c') \in g$, $y' \in (V-T)^*$;

B. In greater detail, the derivation step $u \Rightarrow v\,[p_3]$ has this form $01^{n-1}\beta(b'')2\beta(a'')2\beta(h'')20 \Rightarrow 1^{n-1}0\beta(c'')22\beta(h''y'')x''20\ [p_3]$, where $u = 01^{n-1}\beta(b'')2\beta(a'')2\beta(y')20$, $v =$

$1^{n-1}0\beta(c'')22\beta(h''y'')x''20$, p_3 is a production introduced in III, $(a'', b'', y''x'', c'') \in g$, $h'', y'' \in (V-T)^*$, $x'' \in T^*$;

C. Each derivation step in $v \Rightarrow^k w$ has this form $1^{n-1}0\beta(b''')2\beta(a''')2\beta(y''')t'''20 \Rightarrow 1^{n-1}0\beta(c')22\beta(y''')t'''x'''20$ $[p_4]$, where p_4 is a production introduced in IV, $(a''', b''', x''', c''') \in g$, $y''' \in (V-T)^*$, $t''', x''' \in T^*$;

D. In greater detail, the derivation step $w \Rightarrow z$ $[p_5]$ has this form $1^{n-1}0\beta(b'''')2\beta(a'''')2t''''20 \in t''''x''''[p_5]$, where $w = 1^{n-1}0\beta(b'''')2\beta(a'''')2t''''20$, $z = t''''x''''$, p_5 is a production introduced in $V, (a'''', b'''', x'''', c'''') \in g$ with $c'''' \in F$.

Let $2 \Rightarrow rhs(p_1)$ $[p_1] \Rightarrow^i u \Rightarrow v\,[p_3] \Rightarrow^k w \Rightarrow z\,[p_5]$ be any successful derivation in G such that this derivation satisfies the above properties. Observe that at this point $R \Rightarrow^i a''y''b'' \Rightarrow y''x''b''' \Rightarrow^k a''''t''''b'''' \Rightarrow z$ in Q, so $z \in L(Q)$. Consequently, $L(G) \subseteq L(Q)$.

A proof demonstrating that $L(Q) \subseteq L(G)$ is left to the reader. Since $L(Q) = L(G)$ and G has only three nonterminals, 0, 1, and 2, Lemma 2 holds. □

Theorem 3. $\mathbf{RE} = \mathbf{SC}_3$

Proof. Obviously, $\mathbf{SC}_3 \subseteq \mathbf{RE}$. By Lemma 2, we also have $\mathbf{RE} \subseteq \mathbf{SC}_3$. Thus, $\mathbf{SC}_3 = \mathbf{RE}$, and the theorem holds. □

Recall that $\mathbf{SC}_1 \subset \mathbf{RE}$; in fact, the one-nonterminal scattered context grammars cannot even generate some context-sensitive languages (see [4]). However, this paper proves $\mathbf{SC}_3 = \mathbf{RE}$ (see Theorem 3). What is the generative power of two-nonterminal scattered context grammars?

References

[1] J. Dassow, G. Paun, Regulated Rewriting in Formal Language Theory, Springer, New York, 1989.

[2] H.C.M. Kleijn, G. Rozenberg, On the generative power of regular pattern grammars, Acta Inform. 20 (1983) 391–411.

[3] A. Meduna, Four-nonterminal scattered context grammars characterize the family of recursively enumerable languages, Int. J. Comput. Math. 51 (1996) 465–474.

[4] A. Meduna, Terminating left-hand sides of scattered context productions, Theoret. Comput. Sci. 237 (2000) 423–427.

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

[6] Proceedings of the International Workshop on Descriptional Complexity of Automata, Grammars and Related Structures, July 20–23, 1999, Ed. by J. Dassow and D. Wotschke, Otto-von-Guericke University, Magdeburg, Germany, 1999.

[7] Proceedings of the Second Workshop on Descriptional Complexity of Automata, Grammars and Related Structures, July 27–29, 2000, London, Canada (in press).