A Note on the Descriptional Complexity of Semi-Conditional Grammars

Tomáš Masopust

Dept. of Information Systems, Faculty of Information Technology,
Brno University of Technology, Božetěchova 2,
612 66 Brno, Czech Republic
masopust@fit.vutbr.cz

Abstract. The descriptional complexity of semi-conditional grammars is studied. A proof that every recursively enumerable language is generated by a semi-conditional grammar of degree $(2, 1)$ with no more than seven conditional productions and eight nonterminals is given.

Keywords: formal languages, descriptional complexity, semi-conditional grammars

1 Introduction

This paper studies the descriptional complexity of semi-conditional grammars (see [4, 7–9] for more details) with respect to the number of conditional productions and nonterminals.

Semi-conditional grammars are modified context-free grammars, where a permitting and a forbidding context is associated with each production. This means that a production is applicable if its permitting context is contained in the current sentential form and its forbidding context is not. As a special case of semi-conditional grammars, we obtain simple semi-conditional grammars introduced in [3], where one of the contexts is required to be a special symbol 0, i.e., either a permitting or a forbidding context is associated with each production.

Whereas the descriptional complexity of simple semi-conditional grammars has been studied carefully (see [5, 7, 8, 10]), the descriptional complexity of semi-conditional grammars has not been studied at all, and all results concerning the descriptional complexity of semi-conditional grammars are consequences of results concerning the descriptional complexity of simple semi-conditional grammars. Specifically, in [8], a proof that every recursively enumerable language is generated by a (simple) semi-conditional grammar of degree $(2, 1)$ with no more than twelve conditional productions and thirteen nonterminals was given. Later, in [10], this result was improved and a proof that every recursively enumerable language is generated by a (simple) semi-conditional grammar of degree $(2, 1)$ with no more than ten conditional productions and twelve nonterminals was given. Finally, the result from [10] was improved in [5], where a proof that every recursively enumerable language is generated by a (simple) semi-conditional grammar of degree $(2, 1)$ with no more than nine conditional productions and ten nonterminals was given. However, a better result can be achieved for semi-conditional grammars than for simple semi-conditional grammars. In this paper, a proof that every recursively enumerable language is generated by a semi-conditional grammar of degree $(2, 1)$ with no more than seven conditional productions and eight nonterminals is given.

2 Preliminaries and Definitions

This paper assumes that the reader is familiar with the theory of formal languages (see [1, 6]). For an alphabet V, V^* represents the free monoid generated by V. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$. Set $sub(w) = \{u : u$ is a substring of $w\}$.

In [2], it was shown that every recursively enumerable language is generated by a grammar

$$G = (\{S, A, B, C\}, T, P \cup \{ABC \to \varepsilon\}, S)$$

in the *Geffert normal form*, where P contains context-free productions of the form

$$\begin{array}{ll} S \to uSa, & \text{where } u \in \{A, AB\}^*, \ a \in T, \\ S \to uSv, & \text{where } u \in \{A, AB\}^*, \ v \in \{BC, C\}^*, \\ S \to uv, & \text{where } u \in \{A, AB\}^*, \ v \in \{BC, C\}^*. \end{array}$$

In addition, any terminal derivation is of the form

$$S \Rightarrow^* w_1 w_2 w$$

by productions from P, where $w_1 \in \{A, B\}^*$, $w_2 \in \{B, C\}^*$, $w \in T^*$, and

$$w_1 w_2 w \Rightarrow^* w$$

by $ABC \to \varepsilon$.

A *semi-conditional grammar*, G, is a quadruple

$$G = (N, T, P, S),$$

where

- N is a nonterminal alphabet,
- T is a terminal alphabet such that $N \cap T = \emptyset$,
- $S \in N$ is the start symbol, and
- P is a finite set of productions of the form

$$(X \to \alpha, u, v)$$

with $X \in N$, $\alpha \in (N \cup T)^*$, and $u, v \in (N \cup T)^+ \cup \{0\}$, where $0 \notin N \cup T$ is a special symbol.

If $u \neq 0$ or $v \neq 0$, then the production $(X \to \alpha, u, v) \in P$ is said to be *conditional.* G has *degree* (i, j) if for all productions $(X \to \alpha, u, v) \in P$, $u \neq 0$ implies $|u| \leq i$ and $v \neq 0$ implies $|v| \leq j$. For $x \in (N \cup T)^+$ and $y \in (N \cup T)^*$, x *directly derives* y according to the production $(X \to \alpha, u, v) \in P$, denoted by

$$x \Rightarrow y$$

if $x = x_1 X x_2$, $y = x_1 \alpha x_2$, for some $x_1, x_2 \in (N \cup T)^*$, and $u \neq 0$ implies that $u \in sub(x)$ and $v \neq 0$ implies that $v \notin sub(x)$. As usual, $\Rightarrow$ is extended to $\Rightarrow^i$, for $i \geq 0$, $\Rightarrow^+$, and $\Rightarrow^*$. The language generated by a semi-conditional grammar, G, is defined as

$$\mathscr{L}(G) = \{w \in T^* : S \Rightarrow^* w\}.$$

Let $G = (N, T, P, S)$ be a semi-conditional grammar. If $(X \to \alpha, u, v) \in P$ implies that $0 \in \{u, v\}$, then G is said to be a *simple semi-conditional grammar.*

3 Main Result

This section presents the main result concerning the descriptional complexity of semi-conditional grammars.

Theorem 1. *Every recursively enumerable language is generated by a semi-conditional grammar of degree* $(2, 1)$ *with no more than 7 conditional productions and 8 nonterminals.*

Proof idea.
The main idea of the proof is to simulate a terminal derivation of a grammar, G, in the Geffert normal form.

To do this, we first apply all context-free productions as applied in the G's derivation, and then we simulate the production $ABC \to \varepsilon$ so that we mark with $'$ only one occurrence of A, one of B, and one of C and check that these marked symbols form a substring $A'B'C'$ of the current sentential form. If so, the marked symbols can be removed, which completes the simulation of the production $ABC \to \varepsilon$ in G; otherwise, the derivation must be blocked.

The formal proof follows.

Proof. Let L be a recursively enumerable language. There is a grammar

$$G = (\{S, A, B, C\}, T, P \cup \{ABC \to \varepsilon\}, S)$$

in the Geffert normal form such that $L = \mathscr{L}(G)$. Construct the grammar

$$G' = (\{S, A, B, C, A', B', C', \$\}, T, P' \cup P'', S),$$

where

$$P' = \{(X \to \alpha, 0, 0) : X \to \alpha \in P\},$$

and P'' contains following seven conditional productions:

1. $(A \to \$A', 0, \$)$,
2. $(B \to B', A', B')$,
3. $(C \to C'\$, A'B', C')$,
4. $(B' \to \varepsilon, B'C', 0)$,
5. $(C' \to \varepsilon, A'C', 0)$,
6. $(A' \to \varepsilon, A'\$, 0)$,
7. $(\$ \to \varepsilon, 0, A')$.

To prove that $\mathscr{L}(G) \subseteq \mathscr{L}(G')$, consider a derivation

$$S \Rightarrow^* wABCw'v \Rightarrow ww'v$$

in G by productions from P with only one application of the production $ABC \to \varepsilon$, where $w, w' \in \{A, B, C\}^*$ and $v \in T^*$. Then,

$$S \Rightarrow^* wABCw'v$$

in G' by productions from P'. Moreover, by productions 1, 2, 3, 4, 5, 6, 7, 7, we get

$$\begin{aligned} wABCw'v &\Rightarrow w\$A'BCw'v \\ &\Rightarrow w\$A'B'Cw'v \\ &\Rightarrow w\$A'B'C'\$w'v \\ &\Rightarrow w\$A'C'\$w'v \\ &\Rightarrow w\$A'\$w'v \\ &\Rightarrow w\$\$w'v \\ &\Rightarrow w\$w'v \\ &\Rightarrow ww'v. \end{aligned}$$

The inclusion follows by induction.

To prove that $\mathscr{L}(G) \supseteq \mathscr{L}(G')$, consider a terminal derivation. Let $X \in \{A, B, C\}$ be in a sentential form of this derivation. To eliminate X, there are following three possibilities:

1. If $X = A$, then there must be C and B (by productions 6 and 3) in the derivation;
2. If $X = B$, then there must be C and A (by productions 4 and 3) in the derivation;
3. If $X = C$, then there must be A and B (by productions 5 and 3) in the derivation.

In all above cases, there are A, B, and C in the derivation. By productions 1, 2, 3, and 7, there cannot be more than one A', B', and C' in any sentential form of this terminal derivation. Moreover, by productions 3 and 4, $A'B'C'$ is a substring of a sentential form of this terminal derivation, and there is no terminal symbol between any two nonterminals; otherwise, there will be a situation in which (at least) one of productions 3 and 4 will not be applicable. Thus, any first part of a terminal derivation in G' is of the form

$$S \Rightarrow^* w_1 ABC w_2 w \Rightarrow^3 w_1 \$A'B'C'\$ w_2 w \tag{1}$$

by productions from P' and productions 1, 2, and 3, where $w_1 \in \{A, B\}^*$, $w_2 \in \{B, C\}^*$, and $w \in T^*$. Next, only production 4 is applicable. Thus,

$$\Rightarrow w_1 \$A'C'\$ w_2 w\,.$$

Besides a possible application of production 2, only production 5 is applicable. Thus,

$$\Rightarrow^+ w_1' \$A'\$ w_2' w$$

where $w_1' \in \{A, B, B'\}^*$, $w_2' \in \{B, B', C\}^*$. Besides a possible application of production 2, only production 6 is applicable. Thus,

$$\Rightarrow^+ w_1'' \$\$ w_2'' w$$

where $w_1'' \in \{A, B, B'\}^*$, $w_2'' \in \{B, B', C\}^*$. Finally, only production 7 is applicable, i.e.,

$$\Rightarrow^2 w_1'' w_2'' w\,.$$

Thus, by productions 1, 2, 3, or 1, 3, if production 2 has already been applied, we get

$$\Rightarrow^* uvw\,.$$

Here,

$$uvw \in \{u_1 \$A'B'C'\$ u_2 w : u_1 \in \{A, B\}^*, u_2 \in \{B, C\}^*\}$$

or $uv = \varepsilon$.

Thus, the substring ABC and only this substring was eliminated during the previous derivation. By induction (see (1)), the inclusion holds. This derivation can be performed in G with an application of the production $ABC \to \varepsilon$, too. □

This work has been supported by the Grant Agency of the Czech Republic within the project No. 102/05/H050, FRVŠ grant No. FR762/2007/G1, and the Czech Ministry of Education under the Research Plan No. MSM 0021630528.

References

1. J. Dassow and Gh. Păun. *Regulated Rewriting in Formal Language Theory.* Springer-Verlag, Berlin, 1989.
2. V. Geffert. Context-free-like forms for the phrase-structure grammars. In M. Chytil, L. Janiga, and V. Koubek, editors, *MFCS*, volume 324 of *Lecture Notes in Computer Science*, pages 309–317. Springer, 1988.

3. A. Gopalaratnam and A. Meduna. On semi-conditional grammars with productions having either forbidding or permitting conditions. *Acta Cybernetica*, 11(4):307–324, 1994.
4. J. Kelemen. Conditional grammars: Motivations, definitions, and some properties. In *Proc. Conf. Automata, Languages and Mathematical Sciences*, pages 110–123, Salgótarján, 1984.
5. T. Masopust. An improvement of the descriptional complexity of grammars regulated by context conditions. In *Second Doctoral Workshop on Mathematical and Engineering Methods in Computer Science (MEMICS 2006)*, pages 105–112, Mikulov, 2006.
6. A. Meduna. *Automata and Languages: Theory and Applications*. Springer-Verlag, London, 2000.
7. A. Meduna and M. Švec. Reduction of simple semi-conditional grammars with respect to the number of conditional productions. *Acta Cybernetica*, 15:353–360, 2002.
8. A. Meduna and M. Švec. *Grammars with Context Conditions and Their Applications*. John Wiley & Sons, New York, 2005.
9. Gh. Păun. A variant of random context grammars: Semi-conditional grammars. *Theoretical Computer Science*, 41:1–17, 1985.
10. Gy. Vaszil. On the descriptional complexity of some rewriting mechanisms regulated by context conditions. *Theoretical Computer Science*, 330:361–373, 2005.