Descriptional Complexity of Multi-Parallel Grammars

Tomáš Masopust

Faculty of Information Technology, Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic

Abstract

This paper studies the descriptional complexity of multi-parallel grammars with respect to the number of nonterminals and selectors, and the length of these selectors. As a result, it proves that every recursively enumerable language is generated by a multi-parallel grammar with no more than seven nonterminals and four selectors of length five.

Key words: formal languages, multi-parallel grammars, descriptional complexity
1991 MSC: 68Q19

1 Introduction

The major topic of the descriptional complexity of grammars is to study how to reduce the number of grammatical components, such as the number of productions, selectors, nonterminals, etc., in order to make the grammars small, succinct, and, therefore, easy to use.

The present paper studies the descriptional complexity of multi-parallel grammars originally introduced (as multi grammars) by Kleijn and Rozenberg in 1983 (see [2]). Over its history, however, the original definition of multi-parallel grammars has been simplified. For that reason, the definition used in this paper is based on the definition given by Meduna and Kolář in [5]. Informally, multi-parallel grammars are E0L grammars (see [6, Chapter 2]) having a set of selectors, represented by simple regular languages, which allows or disallows the current sentential form to parallelly rewrite all its symbols. More precisely, any sentential form, say u, is rewritten if (i) either $u = S$, where S is the start

Email address: masopust@fit.vutbr.cz (Tomáš Masopust).

symbol, and there is a production $S \to v$, or (ii) $u \neq S$, $u \in \pi$, for some selector π, and all symbols of u are parallelly rewritten according to the set of productions.

Recall that Meduna and Kolář [5] have proved that every recursively enumerable language is generated by a multi-parallel grammar with no more than eight nonterminals. However, no reduction in the number of selectors and in the length of these selectors has been studied so far. Thus, this gives rise to the open problem whether a similar result can be established for fewer nonterminals and the reduced number and length of selectors. This paper affirmatively answers this question. More specifically, it proves that every recursively enumerable language is generated by a multi-parallel grammar with no more than seven nonterminals and four selectors of length five.

2 Preliminaries and Definitions

This paper assumes that the reader is familiar with formal language theory (see [4]). For an alphabet (finite nonempty set) V, V^* represents the free monoid generated by V. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$. For $w \in V^*$, w^R denotes the mirror image of w.

Let $m \geq 1$ be a positive integer. A *multi-parallel grammar* is a quintuple $G = (N, T, P, S, K)$, where N is a nonterminal alphabet, T is a terminal alphabet, $V = N \cup T$, $S \in N$ is the start symbol, P is a finite set of productions of the form $a \to x$, for $a \in V$ and $x \in V^*$, and $K = \{\pi_1, \dots, \pi_\ell\}$ is a finite set of *selectors*, for some $\ell \geq 0$, where for each $i = 1, \dots, \ell$, π_i is a language of the form $\pi_i = F_1 F_2 \dots F_m$ with $F_j \in \{W^+ : W \subseteq V, W \neq \emptyset\}$, for all $j = 1, \dots, m$.

We say that G has selectors of length m. If G has k nonterminals, then G is said to be a (k, ℓ, m) multi-parallel grammar; i.e., G has k nonterminals and ℓ selectors of length m.

Let $u, v \in V^*$. Then, G directly derives v from u according to a selector $\pi \in K$, symbolically denoted as $u \Rightarrow v\ [\pi]$, provided that

(1) either $u = S$ and $S \to v \in P$, or
(2) there exists $k \geq 1$ so that
   (a) $u = a_1 \dots a_k$ with $a_i \in V$, for all $i = 1, \dots, k$,
   (b) $u \in \pi$,
   (c) $v = x_1 \dots x_k$ with $a_i \to x_i \in P$.

In the standard manner, extend $\Rightarrow$ to $\Rightarrow^n$, for all $n \geq 0$, $\Rightarrow^+$, and $\Rightarrow^*$. The language of G is defined as $L(G) = \{w \in T^* : S \Rightarrow^* w\}$.

3 Result

This section presents the main result of this paper.

Theorem 1. *Every recursively enumerable language is generated by a* $(7, 4, 5)$ *multi-parallel grammar.*

PROOF. Let L be a recursively enumerable language. It is well-known that there is a grammar $G = (\{S, 0, 1, \$\}, T, P \cup \{0\$0 \to \$, 1\$1 \to \$, \$ \to \varepsilon\}, S)$ with $\{S, 0, 1, \$\} \cap T = \emptyset$ and P containing (context-free) productions of the form

$$S \to uSa,$$
$$S \to uSv,$$
$$S \to u\$v,$$

where $u, v \in \{0, 1\}^*$ and $a \in T$, such that $L = L(G)$ and any terminal derivation in G is of the form $S \Rightarrow^* u\$u^R w$ by productions from P, where $u \in \{0, 1\}^*$ and $w \in T^*$, and $u\$u^R w \Rightarrow^* w$ by productions $0\$0 \to \$$, $1\$1 \to \$$, and $\$ \to \varepsilon$ (see [1] and [3]).

Construct the following $(7, 4, 5)$ multi-parallel grammar

$$G' = (N, T, P_1 \cup P_2, S, K),$$

where $N = \{S, 0, 1, 2, 0', 1', \$\}$ such that $N \cap T = \emptyset$,

$$\begin{aligned} P_1 = &\{a \to a : a \in T\} \\ &\cup \{0 \to 0, 1 \to 1, \$ \to \$\} \\ &\cup \{0 \to 20', 0 \to 0'2, 1 \to 21', 1 \to 1'2, \$ \to 2\$2\} \\ &\cup \{0' \to \varepsilon, 1' \to \varepsilon, 2 \to \varepsilon, \$ \to \varepsilon\}, \end{aligned}$$

P_2 contains productions of the form

$$\begin{aligned} &S \to 2u2S2a2 &&\text{if } S \to uSa \in P, \\ &S \to 2u2S2v2 &&\text{if } S \to uSv \in P, \\ &S \to 2u2\$2v2 &&\text{if } S \to u\$v \in P, \end{aligned}$$

and K contains the following four selectors:

(1) $\pi_1 = \{0, 1, 2\}^+\{2\}^+\{S\}^+\{2\}^+(\{0, 1, 2\} \cup T)^+$,
(2) $\pi_2 = \{0, 1, 2\}^+\{2\}^+\{\$\}^+\{2\}^+(\{0, 1, 2\} \cup T)^+$,
(3) $\pi_3 = \{0, 1, 2\}^+\{0'\}^+\{\$\}^+\{0'\}^+(\{0, 1, 2\} \cup T)^+$,

(4) $\pi_4 = \{0,1,2\}^+\{1'\}^+\{\$\}^+\{1'\}^+(\{0,1,2\} \cup T)^+$.

To prove that $L(G) \subseteq L(G')$, consider a derivation in G,

$$S \Rightarrow^* u\$u^R w\,,$$

by productions from P, where $u \in \{0,1\}^*$ and $w \in T^*$, and $u\$u^R w \Rightarrow^* w$ by productions $0\$0 \to \$$, $1\$1 \to \$$, and $\$ \to \varepsilon$. It is not hard to see that (by selector π_1)

$$S \Rightarrow^* u_1 2 u_2 2\$2 u_3 2 u_4 w$$

in G', where $u = u_1u_2 = (u_3u_4)^R$, because during this part of the derivation, exactly four symbols 2 are generated (see the productions in P_2) and all symbols 2 occurring in the sentential form are removed (see the only production having 2 on its left-hand side, $2 \to \varepsilon$). To complete the simulation, assume that $u = u_1u_2 = \overline{u}x$, where $x \in \{0,1,\varepsilon\}$ and $x = \varepsilon$ if and only if $\overline{u}x = \varepsilon$.

If $x = \varepsilon$, then

$$22\$22w \Rightarrow w\ [\pi_2]$$

by productions $2 \to \varepsilon$ and $\$ \to \varepsilon$ (and $a \to a$, a occurs in w; however, we will not mention such productions in the remainder of this proof).

If $x \in \{0,1\}$, then

$$u_1 2 u_2 2\$2 u_3 2 u_4 w \Rightarrow \overline{u}2x'\$x'2\overline{u}^R w\ [\pi_2]$$

by productions $2 \to \varepsilon$, $x \to 2x'$, $x \to x'2$ (for instance, for $x = 0$, $x \to 2x'$ means $0 \to 20'$). If $\overline{u} = \varepsilon$, then

$$\overline{u}2x'\$x'2\overline{u}^R w = 2x'\$x'2w \Rightarrow w\ [\pi_3 \text{ or } \pi_4]$$

by productions $2 \to \varepsilon$, $x' \to \varepsilon$, $\$ \to \varepsilon$. Otherwise,

$$\overline{u}2x'\$x'2\overline{u}^R w \Rightarrow \overline{u}2\$2\overline{u}^R w\ [\pi_3 \text{ or } \pi_4]$$

by productions $2 \to \varepsilon$, $x' \to \varepsilon$, $\$ \to 2\$2$. The proof then follows by induction.

To prove the other inclusion, $L(G) \supseteq L(G')$, consider a derivation in G'. Such a derivation is of the form

$$S \Rightarrow^* w_1 x w_2 2\$2 w_3 y w_4\,,$$

where $w_1, w_2 \in \{0,1,2\}^*$, $w_3, w_4 \in (\{0,1,2\} \cup T)^*$, $x, y \in \{0,1,2\}$, $x = 2$ if and only if $y = 2$ if and only if $w_1 x w_2 = 2$, and the derivation is according to selector π_1. It means that during this part of the derivation, only S is expanded according to productions from P_2 while all other symbols are replaced with themselves (except for 2 that is removed). Thus, during the first part of the derivation, there are exactly four symbols 2 in any sentential form different from the start form (symbol); two on the left-hand side of \$ and two on the right-hand side (see the productions in P_2 and production $2 \to \varepsilon$).

If $x = 2$, then the derivation is of the form

$$S \Rightarrow^* 22\$2w_3 2w_4 \,.$$

Then, the only applicable selector is π_2. Thus,

$$22\$2w_3 2w_4 \Rightarrow w_3 w_4 \,,$$

where $w_3 w_4 \in T^*$, because if \$ is not removed, then the sentential form is of the form $\$w_3w_4$ (by $\$ \to \$$) or $2\$2w_3w_4$ (by $\$ \to 2\$2$), and it is not hard to see that no selector is applicable.

In G, $S \Rightarrow^* \$w_3w_4 \Rightarrow w_3w_4$.

Assume that $x \neq 2$, then next n steps of the derivation, for some $n \geq 1$, continue as follows:

$$\begin{aligned}
w_1 x w_2 2\$2 w_3 y w_4 &\Rightarrow h(w_1)xh(w_2)2\$2h(w_3)yh(w_4)\ [\pi_2]\\
&\Rightarrow h(w_1)xh(w_2)2\$2h(w_3)yh(w_4)\ [\pi_2]\\
&\ \vdots\\
&\Rightarrow h(w_1)xh(w_2)2\$2h(w_3)yh(w_4)\ [\pi_2]\\
&\Rightarrow h(w_1)2x'h(w_2)\$h(w_3)y'2h(w_4)\ [\pi_2]\,,
\end{aligned}$$

where the first $n-1$ steps is by $2 \to \varepsilon$ and $\$ \to 2\$2$, and the last step is by $x \to 2x'$, $y \to y'2$, $2 \to \varepsilon$, and $\$ \to \$$, and where h is a homomorphism on V^* defined as $h(2) = \varepsilon$ and $h(a) = a$, for all $a \in V - \{2\}$.

Therefore, $h(w_2w_3) = \varepsilon$, $x = y$, and a production of the form $x \to 2x'$ is applied on the left-hand side of \$ and of the form $x \to x'2$ on the right-hand side—not reversely; otherwise, no selector is applicable.

Thus, the derivation continues as follows.

If $h(w_1) = \varepsilon$, then

$$2x'\$x'2h(w_4) \Rightarrow h(w_4)\ [\pi_3 \text{ or } \pi_4]$$

by productions $2 \to \varepsilon$, $x' \to \varepsilon$, $\$ \to \varepsilon$, where $h(w_4) \in T^*$; otherwise, no selector is applicable.

If $h(w_1) \neq \varepsilon$, then either

$$h(w_1)2x'\$x'2h(w_4) \Rightarrow h(w_1)2\$2h(w_4)\ [\pi_3 \text{ or } \pi_4]$$

by productions $2 \to \varepsilon$, $x' \to \varepsilon$, $\$ \to 2\$2$, or

$$h(w_1)2x'\$x'2h(w_4) \Rightarrow w_5 2z'\$z'2w_6\ [\pi_3 \text{ or } \pi_4]$$

by productions $2 \to \varepsilon$, $x' \to \varepsilon$, $\$ \to \$$, $z \to 2z'$, $z \to z'2$, where $h(w_1) = w_5 z$ and $h(w_4) = zw_6$, for some $z \in \{0,1\}$ and $w_5, w_6 \in \{0,1\}^*$. The proof then follows by induction.

In G, $S \Rightarrow^* h(w_1)x\$xh(w_4) \Rightarrow h(w_1)\$h(w_4)$ by productions from P and production $x\$x \to \$$. □

4 Conclusion

This paper decreases the known number of nonterminals needed by multi-parallel grammars to generate any recursively enumerable language from eight to seven. In addition, in this case, it studies the number and the length of selectors, and it proves that four selectors of length five are sufficient. However, although this paper improves the result from [5], it does not prove the minimality of the new result. Therefore, the question whether or not the numbers of nonterminals and selectors, and the length of selectors, are minimal remains open.

Acknowledgements

The author thanks the anonymous referee for valuable suggestions.

This work was supported by the Czech Ministry of Education under the Research Plan No. MSM 0021630528.

References

[1] V. Geffert, Normal forms for phrase-structure grammars, Theoretical Informatics and Applications 25 (5) (1991) 473–496.

[2] H. Kleijn, G. Rozenberg, Multi grammars, International Journal of Computer Mathematics 12 (3–4) (1983) 177–201.

[3] T. Masopust, A. Meduna, Descriptional complexity of generalized forbidding grammars, in: V. Geffert, G. Pighizzini (eds.), Proceedings of the 9th International Workshop on Descriptional Complexity of Formal Systems, DCFS'07, High Tatras, Slovakia, 2007.

[4] A. Meduna, Automata and Languages: Theory and Applications, Springer-Verlag, London, 2000.

[5] A. Meduna, D. Kolář, Descriptional complexity of multi-parallel grammars with respect to the number of nonterminals, in: C. Martín-Vide, V. Mitrana (eds.), Grammars and Automata for String Processing, vol. 9 of Topics in Computer Mathematics, Taylor and Francis, 2003.

[6] A. Meduna, M. Švec, Grammars with Context Conditions and Their Applications, John Wiley & Sons, New York, 2005.