A Note on the Cooperation in Rewriting Systems with Context-Dependency Checking

Zbyněk Křivka and Tomáš Masopust

Faculty of Information Technology, Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic
krivka@fit.vutbr.cz, masopust@fit.vutbr.cz

Extended Abstract

Rewriting systems based on simple forms of productions play an important role in formal language theory. Therefore, it is no surprise that context-free grammars and their modifications are one of the most studied models. However, many of the systems describing practically interesting applications, such as the parsers of natural and programming languages, require additional mechanisms that allow these systems to check for some kind of context dependencies. From that point of view, context-free grammars are not convenient for those applications because they are too simple to describe such context dependencies.

Naturally, a method how to increase the generative power of rewriting systems is to compose them of several simple components, and to define how these components cooperate to generate the common sentential form. In the literature, such devices are called *cooperating distributed (CD) grammar systems*. Specifically, CD grammar systems [1] are rewriting devices composed of several components represented by grammars or other rewriting systems, and of a protocol describing the mutual cooperation of components in the generative process. These protocols describe (roughly speaking) the number of steps the component has to make to allow another component to work. For instance, the most interesting protocol is so-called *terminal mode* making the component work until it is not able to make another derivation step.

It is well-known that the cooperation (as specified above) has a significant effect on context-free grammars—nontrivial cooperation protocols make context-free CD grammar systems more powerful than context-free grammars (the reader is referred to [2]).

As mentioned above, simple rewriting devices that are able to check for some context dependencies are of interest. One of such simple rewriting systems, so-called *random context grammars* (see [3]), is a natural generalization of context-free grammars, where a mechanism checking for context dependencies is added. Specifically, two finite sets of nonterminal symbols are attached to each context-free production—a *permitting* and a *forbidding* set—and such a production is applicable only if all permitting symbols appear in the current sentential form, while no forbidding symbol does. It is known that the family of random context languages contains the whole family of context-free languages and that is properly included in the family of context sensitive languages. In addition, considering only those random context grammars that have all permitting (forbidding) sets empty results to the introduction of *forbidding* (*permitting*) *grammars*, which are less powerful than random context grammars (the reader is referred to [4, 5]).

Recently, it has been shown that the t-mode cooperation protocol has a significant effect on both permitting and forbidding grammars. More specifically, it is demonstrated in [6] and [7], respectively, that they are as powerful as random context grammars. Particularly, it is surprising that this cooperation protocol is able to compensate the absence of permitting sets in case of CD grammar systems with forbidding components (see [6]), and, even more surprisingly, that it is able to compensate the absence of forbidding sets in case of CD grammar systems with permitting components (see [7]). This is primarily interesting in comparison with the open problem of whether the families of permitting and forbidding languages are comparable.

This paper studies the cooperation in terms of CD grammar systems with respect to all usual cooperation protocols and its effect on the generative power of random context grammars. In comparison with the results achieved for context-free, permitting, and forbidding grammars, this paper demonstrates that the cooperation has no effect on the generative power of random context grammars. Specifically, using any of the cooperation protocols, CD grammar systems with random context grammars as their components generate only random context languages. Therefore, as far as the authors know, the family of random context grammars is the smallest family of rewriting devices more powerful than context-free grammars, where the cooperation of several components has no effect on the generative power.

1 Definitions

For an alphabet V, V^* represents the free monoid generated by V, the unit is denoted by λ, and $V^+ = V^* - \{\lambda\}$. For $w \in V^*$, let $|w|$ denote the length of w and $alph(w)$ denote the set of all symbols occurring in w.

A *random context grammar* (see [3]) is a quadruple $G = (N, T, P, S)$, where N and T are alphabets of nonterminals and terminals, respectively, such that $N \cap T = \emptyset$, $V = N \cup T$, $S \in N$ is the start symbol, and P is a finite set of productions of the form $(A \to x, Per, For)$, where $A \in N$ and $x \in V^*$, and $Per, For \subseteq N$. If for each $(A \to x, Per, For) \in P$, $Per = \emptyset$, then G is said to be a *forbidding* random context grammar. If for each $(A \to x, Per, For) \in P$, $For = \emptyset$, then G is said to be a *permitting* random context grammar. For $u, v \in V^*$ and $(A \to x, Per, For) \in P$, $uAv \Rightarrow uxv$ holds provided that $Per \subseteq alph(uv)$ and $alph(uv) \cap For = \emptyset$. Extend $\Rightarrow$ to $\Rightarrow^n$, for $n \geq 0$, $\Rightarrow^+$, and $\Rightarrow^*$. A random context grammar is *nonerasing* if no production has λ on its right-hand side. The language generated by G is defined as $L(G) = \{w \in T^* : S \Rightarrow^* w\}$.

In the literature, the relation of the direct derivation step is also defined so that for $u, v \in V^*$ and $(A \to x, Per, For) \in P$, $uAv \Rightarrow uxv$ holds provided that $Per \subseteq alph(uAv)$ and $alph(uAv) \cap For = \emptyset$. It is not hard to prove that these definitions are equivalent for (permitting, forbidding) random context grammars.

Lemma 1. *Definitions* (1) *and* (2) *are equivalent for (permitting, forbidding) random context grammars.*

Proof. (1) $\Rightarrow$ (2): Let $G = (N, T, P, S)$ be a (permitting, forbidding) random context grammar using definition (1). Construct the grammar $G' = (N \cup N', T, P', S)$ of the same type using definition (2), where $N' = \{A' : A \in N\}$ is such that $N \cap N' = \emptyset$, and P' is defined as follows.

1. If G is a forbidding random context grammar or a random context grammar, then $P' = \{(A \to A', \emptyset, N'), (A' \to x, \mathit{Per}, \mathit{For}) : (A \to x, \mathit{Per}, \mathit{For}) \in P\}$.
2. On the other hand, if G is a permitting random context grammar, then $P' = \{(A \to A', \emptyset, \emptyset), (A' \to x, \mathit{Per}, \emptyset) : (A \to x, \mathit{Per}, \emptyset) \in P\}$.

It is not hard to see that $L(G) = L(G')$.

(2) $\Rightarrow$ (1): Let G be a (permitting, forbidding) random context grammar using definition (2). Construct the grammar G' of the same type using definition (1) as follows. For each production $p = (A \to x, \mathit{Per}, \mathit{For})$ of G with $A \notin \mathit{For}$, add $p' = (A \to x, \mathit{Per} - \{A\}, \mathit{For})$ to the set of productions of G'. Clearly, p is applicable in G if and only if p' is applicable in G'. $\square$

In what follows, definition (1) is used. However, using definition (2), does not change the results.

A *CD grammar system* is a construct $\Gamma = (N, T, P_1, P_2, \ldots, P_n, S)$, for some $n \geq 2$, where N, T, S, and V have the same meaning as in random context grammars and for each $i = 1, 2, \ldots, n$, P_i is a finite set of productions. For $u, v \in V^*$ and $1 \leq k \leq n$, let $u \Rightarrow_k v$ denote a derivation step made by a production from P_k. As usual, extend the relation $\Rightarrow_k$ to $\Rightarrow_k^m$, for $n \geq 1$, $\Rightarrow_k^+$, and $\Rightarrow_k^*$. In addition, define the relation $u \Rightarrow_k^t v$ so that $u \Rightarrow_k^+ v$ and there is no $w \in V^*$ such that $v \Rightarrow_k w$. Now, we can define the languages generated by Γ working in the f-mode, denoted by $L_f(\Gamma)$, as follows:

f-mode: $L_f(\Gamma) = \{w \in T^* :$ there exists $\ell \geq 1$ such that $\alpha_i \Rightarrow_{k_i}^f \alpha_{i+1}$, $1 \leq k_i \leq n$, $i = 1, \ldots, \ell - 1$, $\alpha_1 = S$, and $\alpha_\ell = w\}$, $f \in \{t, *\} \cup \{=m : m \geq 1\}$.

$\leq m$-mode: $L_{\leq m}(\Gamma) = \{w \in T^* :$ there is $\ell \geq 1$ such that $\alpha_i \Rightarrow_{k_i}^{j_i} \alpha_{i+1}$, $1 \leq k_i \leq n$, $1 \leq j_i \leq m$, $i = 1, \ldots, \ell - 1$, $\alpha_1 = S$, and $\alpha_\ell = w\}$, for $m \geq 1$.

$\geq m$-mode: $L_{\geq m}(\Gamma) = \{w \in T^* :$ there is $\ell \geq 1$ such that $\alpha_i \Rightarrow_{k_i}^{j_i} \alpha_{i+1}$, $1 \leq k_i \leq n$, $j_i \geq m$, $i = 1, \ldots, \ell - 1$, $\alpha_1 = S$, and $\alpha_\ell = w\}$, for $m \geq 1$.

The families of languages generated by CD grammar systems with n context-free components working in the f-mode, where $f \in \{*, t\} \cup \{\leq k, =k, \geq k : k \geq 1\}$, are denoted by $\mathbf{CD}_f(\mathbf{CF}_\lambda, n)$, or $\mathbf{CD}_f(\mathbf{CF}, n)$ if the components are nonerasing. Let **CF**, **ET0L**, and **RE** denote the families of context-free, ET0L, and recursively enumerable languages, respectively, and let **MAT** and $\mathbf{MAT}_\lambda$ denote the families of languages generated by matrix grammars without and with erasing productions, respectively. It is well-known (see [2]) that

1. $\mathbf{CD}_f(\mathbf{CF}_\lambda, n) = \mathbf{CD}_f(\mathbf{CF}, n) = \mathbf{CF}$, $n \geq 1$, $f \in \{=1, \geq 1, *\} \cup \{\leq k : k \geq 1\}$,
2. $\mathbf{CF} \subset \mathbf{CD}_f(\mathbf{CF}, 2) \subseteq \mathbf{CD}_f(\mathbf{CF}, r) \subseteq \mathbf{MAT}$, $f \in \{=k, \geq k : k \geq 2\}$, $r \geq 3$,

3. $\mathbf{CF} \subset \mathbf{CD}_f(\mathbf{CF}_\lambda, 2) \subseteq \mathbf{CD}_f(\mathbf{CF}_\lambda, r) \subseteq \mathbf{MAT}_\lambda$, $f \in \{=k, \geq k : k \geq 2\}$, $r \geq 3$,
4. $\mathbf{CD}_t(\mathbf{CF}_\lambda, 2) = \mathbf{CD}_t(\mathbf{CF}, 2) = \mathbf{CF}$, and
5. $\mathbf{CD}_t(\mathbf{CF}_\lambda, n) = \mathbf{CD}_t(\mathbf{CF}, n) = \mathbf{ET0L}$, for $n \geq 3$.

The family of languages generated by CD grammar systems with n random context components working in the f-mode, for $f \in \{*, t\} \cup \{\leq k, =k, \geq k : k \geq 1\}$, is denoted by $\mathbf{CD}_f(\mathbf{RC}_\lambda, n)$, or $\mathbf{CD}_f(\mathbf{RC}, n)$ if the components are nonerasing.

2 Results

This section presents the main results of this paper. We only describe the construction parts of the proofs and basic ideas, leaving the complete demonstrations to the reader.

First, however, as it is well-known that $\mathbf{RC}_\lambda = \mathbf{RE}$ (see [8]), the following theorem holds.

Theorem 1. $\mathbf{RC}_\lambda = \mathbf{CD}_f(\mathbf{RC}_\lambda, n)$*,* $n \geq 2$*,* $f \in \{*, t\} \cup \{\leq k, =k, \geq k : k \geq 1\}$*.*

Given a random context grammar, if this grammar is considered to be the only nonempty component (with productions of the form $(A \to A, \emptyset, \emptyset)$ added if needed), we have the following result.

Theorem 2. $\mathbf{RC} \subseteq \mathbf{CD}_f(\mathbf{RC}, n)$*, for* $n \geq 2$*,* $f \in \{*, t\} \cup \{\leq k, =k, \geq k : k \geq 1\}$*.*

On the other hand, it is not hard to see that $\mathbf{CD}_f(\mathbf{RC}, n) \subseteq \mathbf{RC}$, for all $f \in \{=1, \geq 1, *\} \cup \{\leq k : k \geq 1\}$. In what follows, we prove that the analogous inclusion holds for the other derivation modes as well.

Lemma 2. $\mathbf{CD}_t(\mathbf{RC}, n) \subseteq \mathbf{RC}$*, for* $n \geq 2$*.*

Proof. Let $\Gamma = (N, T, P_1, P_2, \ldots, P_n, S)$ be a CD system with n random context components. Construct the random context grammar $G = (N' \cup \{\bar{S}\}, T \cup \{c\}, P', \bar{S})$, where c, $\bar{S}$ are new symbols, $c, \bar{S} \notin T \cup N'$,

$$\begin{aligned} N' = N &\cup \{X' : X \in N\} \\ &\cup \{[Q_i], \langle p, Q_i \rangle, [p, Q_i], [i] : Q_i \subseteq P_i,\, p \in Q_i,\, 1 \leq i \leq n\}\,, \end{aligned}$$

and P' is constructed as follows:

1. For each $(A \to x, Per, For) \in P_i$, add $(A \to x, Per \cup \{[i]\}, For)$ to P'.
2. For $1 \leq i, \ell \leq n$, add to P'
   (a) $(\bar{S} \to S[i], \emptyset, \emptyset)$,
   (b) $([i] \to [P_i], \emptyset, \emptyset)$,
   (c) $([\emptyset] \to [\ell], \emptyset, \emptyset)$,
   (d) $([i] \to c, \emptyset, \emptyset)$.
3. For $Q_i \subseteq P_i$ and $p = (A \to x, Per, For) \in Q_i$, $1 \leq i \leq n$, add also to P'
   (e) $([Q_i] \to [Q_i - \{p\}], \emptyset, \{A\})$,
   (f) $([Q_i] \to \langle p, Q_i \rangle, \{A\}, \emptyset)$,

(g) $(A \to A', \{\langle p, Q_i \rangle\}, \{A'\})$,
(h) $(\langle p, Q_i \rangle \to [p, Q_i], \{A'\}, \{X\})$, where $X \in \mathit{Per}$,
(i) $(\langle p, Q_i \rangle \to [p, Q_i], \{A', X\}, \emptyset)$, where $X \in \mathit{For}$,
(j) $(A' \to A, \emptyset, \emptyset)$,
(k) $([p, Q_i] \to [Q_i - \{p\}], \emptyset, \{A'\})$.

From the construction follows that $L(G) = L_t(\Gamma)c$. As it is known that nonerasing random context grammars are closed under restricted homomorphisms (see Lemma 1.3.3 in [8]), there is a random context grammar H such that $L_t(\Gamma) = L(H)$.

Basic Idea. Every derivation step made by a production from P_i is simulated by one or more rewriting steps in G. Specifically, at the end of each sentential form, there is a special nonterminal (of the form $[i]$, or $[Q_i]$, or $[p, Q_i]$, or $\langle p, Q_i \rangle$, for some $1 \leq i \leq n$) with a single occurrence (see step (2a) and the consequent steps), which determines the simulated component of Γ. To check the same conditions required by Γ in the t-mode when the current component is changed from, say, P_i to P_ℓ, G checks that there is no applicable production in P_i (see steps (2b), (3), and (2c)). More specifically, if $[Q_i]$ occurs in the current sentential form and $p \in Q_i$ is a production then the steps (3e) through (3k) verify that p is not applicable to the corresponding sentential form in Γ. Finally, the last nonterminal, $[i]$, is rewritten to c (see (2d)).

The following lemma discusses the effect of the remaining derivation modes on the generative power of CD grammar systems with random context components.

Lemma 3. $\mathbf{CD}_f(\mathbf{RC}, n) \subseteq \mathbf{RC}$, *for* $n \geq 2$ *and* $f \in \{=k, \geq k : k \geq 2\}$.

Proof. Let $\Gamma = (N, T, P_1, P_2, \ldots, P_n, S)$ be a CD grammar system with n random context components working in the $\geq k$-mode, for some $k \geq 2$. Construct the random context grammar $G = (N' \cup \{S'\}, T \cup \{c\}, P', S')$, where c and S' are new symbols, $c, S' \notin T \cup N'$,

$$N' = N \cup \bar{N} \cup \{[i, m], \langle i, m \rangle : 1 \leq i \leq n, 0 \leq m \leq k\},$$

$\bar{N} = \{\langle x \rangle : (A \to x, \mathit{Per}, \mathit{For}) \in P_i, 1 \leq i \leq n\}$, and P' is constructed as follows.

For each component P_i, $1 \leq i \leq n$, and each production $(A \to x, \mathit{Per}, \mathit{For}) \in P_i$, add the following productions to P':

1. $(S' \to S[i, k], \emptyset, \emptyset)$,
2. $(A \to \langle x \rangle, \mathit{Per} \cup \{[i, m] : 1 \leq m \leq k\}, \mathit{For} \cup \bar{N})$,
3. $(\langle x \rangle \to x, \{\langle i, m \rangle : 0 \leq m \leq k\}, \emptyset)$,
4. $([i, k] \to \langle i, k \rangle, \bar{N}, \emptyset)$,
5. $([i, m] \to \langle i, m - 1 \rangle, \bar{N}, \emptyset)$, where $1 \leq m \leq k$,
6. $(\langle i, m \rangle \to [i, m], \emptyset, \bar{N})$, where $0 \leq m \leq k$,
7. $([i, 0] \to [j, k], \emptyset, \emptyset)$, where $1 \leq j \leq n$, and
8. $([i, 0] \to c, \emptyset, \emptyset)$.

From the above construction, we get that $L(G) = L_{\geq k}(\Gamma)c$ and, again, by [8, Lemma 1.3.3], it follows that there exists a random context grammar H such that $L_{\geq k}(\Gamma) = L(H)$.

Finally, it is not hard to see that by omitting all productions constructed in (4), G simulates exactly k applications of productions of the given component. Thus, it also proves the statement for $=k$-mode, where $k \geq 2$.

Basic Idea. Each nonterminal of the form $[i, m]$ or $\langle i, m\rangle$ consists of two components—(i) the index, i, of the simulated component of Γ, and (ii) the counter, m, of the number of productions from P_i to be applied by Γ to allow another component to work. By analogy with the proof of Lemma 2, the simulation uses nonterminals of these forms, with a single occurrence in every sentential form, to distinguish several phases:

(a) the grammar does not decrease the counter after the application of a production from P_i, see production (4);
(b) the grammar does decrease the counter after the application of a production from P_i, see production (5);
(c) the grammar sets the counter to the initial value (k) and changes the simulated component after at least k applications of productions from P_i, see production (7).

Thus, the derivation of G can be described by the following regular expression

$$1\,(\,(2436)^*\ (2536)^k\ 7)^+\,8\,,$$

where the numbers denote the sets of productions from construction steps (1) through (8).

Thus, we can summarize these results in the following theorem.

Theorem 3. $\mathbf{CD}_f(\mathbf{RC}, n) = \mathbf{RC}$*, for* $n \geq 2$*,* $f \in \{*, t\} \cup \{\leq k, =k, \geq k : k \geq 1\}$*.*

Proof. It follows from Theorem 2, the note below it, and Lemmas 2 and 3. □

3 Conclusion and discussion

Recall that CD grammar systems with forbidding (permitting) components are as powerful as random context grammars. In case of the generative power of forbidding components, definition (2) was used in [6]. However, the constructions can be modified so that the results hold using definition (1) as well. In addition, all the results concerning CD grammar systems with random context components proved in this paper hold for both definitions, too. Clearly, productions constructed in (3g) and (3j) are removed from the construction of the proof of Lemma 2 and symbol A' is replaced with A in the productions constructed in (3) of that proof. Obviously, Lemma 3 holds for both definitions.

On the other hand, to achieve the results concerning the generative power of permitting components proved in [7], definition (1) is used. In comparison with forbidding components, we do not know whether the same results can also be achieved using definition (2). It is not hard to see that definition (1) allow us to check for at least two occurrences of a given symbol in the sentential form—the rewritten one and the one occurring on the left or on the right of the rewritten symbol—while definition (2) seems to be too weak to check for that property.

Acknowledgments

This work was supported by the Czech Ministry of Education under the Research Plan No. MSM 0021630528 and the Grant No. 2C06008 (Virtual Laboratory of Microprocessor Technology Application).

References

1. Csuhaj-Varjú, E., Dassow, J.: On cooperating/distributed grammar systems. Journal of Information Processing and Cybernetics (EIK) **26**(1-2) (1990) 49–63
2. Csuhaj-Varjú, E., Dassow, J., Kelemen, J., Păun, G.: Grammar Systems: A Grammatical Approach to Distribution and Cooperation. Gordon and Breach Science Publishers, Topics in Computer Mathematics 5, Yverdon (1994)
3. van der Walt, A.P.J.: Random context grammars. In: Proceedings of the Symposium on Formal Languages. (1970) 163–165
4. van der Walt, A.P.J., Ewert, S.: A shrinking lemma for random forbidding context languages. Theoretical Computer Science **237**(1-2) (2000) 149–158
5. Ewert, S., van der Walt, A.P.J.: A pumping lemma for random permitting context languages. Theoretical Computer Science **270**(1–2) (2002) 959–967
6. Masopust, T.: On the terminating derivation mode in cooperating distributed grammar systems with forbidding components. International Journal of Foundations of Computer Science **20**(2) (2009) 331–340
7. Csuhaj-Varjú, E., Masopust, T., Vaszil, G.: Cooperating distributed grammar systems with permitting grammars as components. Romanian Journal of Information Science and Technology (ROMJIST). To appear.
8. Dassow, J., Păun, G.: Regulated Rewriting in Formal Language Theory. Springer-Verlag, Berlin (1989)