ONE-SIDED RANDOM CONTEXT GRAMMARS: ESTABLISHED RESULTS AND OPEN PROBLEMS

Petr Zemek
Doctoral Degree Programme (3), FIT BUT
E-mail: xzemek02@stud.fit.vutbr.cz

Supervised by: Alexander Meduna
E-mail: meduna@fit.vutbr.cz

Abstract: In this paper, we give an overview of the established results concerning one-sided random context grammars, and point out open problems. Additionally, we propose several areas suggested as topics of future investigation related to the study of one-sided random context grammars.

Keywords: formal languages, one-sided random context grammars, open problems

1 INTRODUCTION

A *one-sided random context grammar* (see [4]) represents a variant of a random context grammar (see [2]). In this variant, a set of *permitting symbols* and a set of *forbidding symbols* are attached to every rule, and its set of rules is divided into the set of *left random context rules* and the set of *right random context rules*. A left random context rule can rewrite a nonterminal if each of its permitting symbols occurs to the left of the rewritten symbol in the current sentential form while each of its forbidding symbols is absent therein. A right random context rule is applied analogically except that the symbols are examined to the right of the rewritten symbol.

During the last several years, these grammars have been intensively studied (see [1, 3–8, 11]). To summarize the body of knowledge concerning these grammars, this paper gives an overview of the established results, and points out many open problems. The motivation behind the present paper is to give prospective scientists interested in these grammars a base where to start a promising research, and provide a summary of open questions and areas to be investigated.

This paper is organized as follows. Section 2 defines one-sided random context grammars and their variants, and illustrates the definitions by an example. Then, Section 3 gives an overview of the established results concerning these grammars and mentions open problems that are related to them. Section 4 concludes this paper by proposing several areas suggested as topics of future investigation.

2 PRELIMINARIES, DEFINITIONS, AND EXAMPLES

We assume that the reader is familiar with formal language theory (see [10]). For a set Q, $\mathrm{card}(Q)$ denotes the cardinality of Q, and 2^Q denotes the power set of Q. For an alphabet V, V^* represents the free monoid generated by V. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$. For $x \in V^*$, $|x|$ denotes the length of x, and $\mathrm{alph}(x)$ denotes the set of symbols occurring in x.

Definition 1 (see [4]). A *one-sided random context grammar* is a quintuple

$$G = \big(N, T, P_L, P_R, S\big)$$

where N and T are two disjoint alphabets, $S \in N$, and $P_L, P_R \subseteq N \times \big(N \cup T\big)^* \times 2^N \times 2^N$ are two finite relations. The components N, T, P_L, P_R, and S are the alphabet of *nonterminals*, the alphabet of *terminals*, the set of *left random context rules*, the set of *right random context rules*, and the *start symbol*, respectively. Each $(A,x,U,W) \in P_L \cup P_R$ is written as $\lfloor A \to x,U,W \rfloor$ throughout this paper. For $\lfloor A \to x,U,W \rfloor \in P_L$, U and W are called the *left permitting context* and the *left forbidding context*, respectively. For $\lfloor A \to x,U,W \rfloor \in P_R$, U and W are called the *right permitting context* and the *right forbidding context*, respectively. Let $V = N \cup T$ be the *total alphabet*. The *direct derivation relation* over V^* is denoted by $\Rightarrow$ and defined as follows. Let $u,v \in V^*$ and $\lfloor A \to x,U,W \rfloor \in P_L \cup P_R$. Then, $uAv \Rightarrow uxv$ in G if and only if

$$\lfloor A \to x,U,W \rfloor \in P_L, U \subseteq \mathrm{alph}(u), \text{ and } W \cap \mathrm{alph}(u) = \emptyset$$

or

$$\lfloor A \to x,U,W \rfloor \in P_R, U \subseteq \mathrm{alph}(v), \text{ and } W \cap \mathrm{alph}(v) = \emptyset$$

Let $\Rightarrow^n$ and $\Rightarrow^*$ denote the nth power of $\Rightarrow$, for some $n \geq 0$, and the reflexive-transitive closure of $\Rightarrow$, respectively. The *language of* G is denoted by $L(G)$ and defined as $L(G) = \{w \in T^* \mid S \Rightarrow^* w\}$. □

Next, we illustrate the previous definition by an example.

Example 1. Consider the one-sided random context grammar

$$G = \big(\{S,A,B,\bar{A},\bar{B}\},\{a,b,c\},P_L,P_R,S\big)$$

where P_L contains the following four rules:

$$\lfloor S \to AB,\emptyset,\emptyset \rfloor \qquad \lfloor \bar{B} \to B,\{A\},\emptyset \rfloor$$
$$\lfloor B \to b\bar{B}c,\{\bar{A}\},\emptyset \rfloor \qquad \lfloor B \to \varepsilon,\emptyset,\{A,\bar{A}\} \rfloor$$

and P_R contains the following three rules:

$$\lfloor A \to a\bar{A},\{B\},\emptyset \rfloor \qquad \lfloor \bar{A} \to A,\{\bar{B}\},\emptyset \rfloor \qquad \lfloor A \to \varepsilon,\{B\},\emptyset \rfloor$$

It is rather easy to see that every derivation that generates a non-empty string of $L(G)$ is of the form

$$\begin{array}{rcl} S & \Rightarrow & AB \\ & \Rightarrow & a\bar{A}B \\ & \Rightarrow & a\bar{A}b\bar{B}c \\ & \Rightarrow & aAb\bar{B}c \\ & \Rightarrow & aAbBc \\ & \Rightarrow^* & a^nAb^nBc^n \\ & \Rightarrow & a^nb^nBc^n \\ & \Rightarrow & a^nb^nc^n \end{array}$$

where $n \geq 1$. The empty string is generated by $S \Rightarrow AB \Rightarrow B \Rightarrow \varepsilon$. Based on the previous observations, we see that G generates the non-context-free language $\{a^nb^nc^n \mid n \geq 0\}$. □

Moreover, the following variants of one-sided random context grammars have been introduced.

Definition 2 (see [4])**.** Let $G = (N,\, T,\, P_L,\, P_R,\, S)$ be a one-sided random context grammar, and set $P = P_L \cup P_R$. If $\lfloor A \to x,U,W \rfloor \in P$ implies that $|x| \geq 1$, then G is a *propagating one-sided random context grammar*. If $\lfloor A \to x,U,W \rfloor \in P$ implies that $W = \emptyset$, then G is a *one-sided permitting grammar*. If $\lfloor A \to x,U,W \rfloor \in P$ implies that $U = \emptyset$, then G is a *one-sided forbidding grammar*. □

Definition 3 (see [4])**.** Let $G = (N,\, T,\, P_L,\, P_R,\, S)$ be a one-sided random context grammar. If $P_R = \emptyset$, then instead of one-sided, we use the term *left*. By analogy with one-sided random context grammars, we define a *left random context grammar*, a *left forbidding grammar*, a *left permitting grammar*, and their propagating variants. □

3 ESTABLISHED RESULTS AND OPEN PROBLEMS

In this section, we give an overview of the established results concerning one-sided random context grammars, and point out open problems that are related to them.

3.1 GENERATIVE POWER

In [4], it has been proved that one-sided random context grammars characterize the family of recursively enumerable languages while their propagating variants define the family of context-sensitive languages. Therefore, we should move our attention on the generative power of their variants.

One-sided forbidding grammars have been shown to be equivalent to so-called selective substitution grammars (see [6]). Nevertheless, their generative power is not known. Can they generate all recursively enumerable languages? By answering this question, we would also establish the generative power of selective substitution grammars, which is unknown (see [6]). Interestingly, left forbidding grammars generate only the family of context-free languages (see [3]).

Both left and one-sided permitting grammars are strictly stronger than context-free grammars, but at most as strong as so-called scattered context grammars (see [1, 4]). An analogous result holds for their propagating variants. The question is: Are these grammars, in fact, equivalent to scattered context grammars?

The generative power of left random context grammars remains unresolved, and we suggest it to be studied.

3.2 DESCRIPTIONAL COMPLEXITY

In [5], it has been proved that any recursively enumerable language can be generated by a one-sided random context grammar with no more than 10 nonterminals. Can this limit be improved? Furthermore, the aforementioned paper has introduced the notion of a *right random context nonterminal*, defined as a nonterminal that appears on the left-hand side of a right random context rule. The paper has demonstrated how to convert any one-sided random context grammar to an equivalent one-sided random context grammar with 2 right random context nonterminals. This result has been proved also for propagating one-sided random context grammars. What is the generative power of one-sided random context grammars having only a single right random context nonterminal?

Apart from the reduction of the overall number of nonterminals and right random context nonterminals, we should also investigate the reduction of the number of right random context rules. Recall that a right random context rule is a rule that checks the presence and absence of symbols to the right of the rewritten nonterminal. Theorem 1 in [4] implies the following corollary. Can we improve it?

Corollary 1. *Every one-sided random context grammar G can be converted into a one-sided random context grammar H having 4 right random context rules such that $L(G) = L(H)$.* □

3.3 NORMAL FORMS

Four normal forms of one-sided random context grammars have been established in [11]. More specifically, the first of them has the set of left random context rules coinciding with the set of right random context rules. The second normal form, in effect, consists in demonstrating how to turn any one-sided random context grammar to an equivalent one-sided random context grammar with the sets of left and right random context rules being disjoint. The third normal form resembles the Chomsky normal form for context-free grammars. In the fourth normal form, each rule has its permitting or forbidding context empty.

It is an open problem whether a one-sided random context grammar $G = (N, T, P_L, P_R, S)$ can be turned into an equivalent one-sided random context grammar in any of the following four normal forms:

1. either $P_L = \emptyset$ or $P_R = \emptyset$;
2. $\lfloor A \to x, U, W \rfloor \in P_L \cup P_R$ implies that $\mathrm{card}(U) + \mathrm{card}(W) \leq 1$;
3. $P_L = \emptyset$ and $\lfloor A \to x, U, W \rfloor \in P_R$ implies that $W = \emptyset$;
4. $P_R = \emptyset$ and $\lfloor A \to x, U, W \rfloor \in P_L$ implies that $W = \emptyset$.

3.4 LEFTMOST DERIVATIONS

By analogy with the three well-known types of leftmost derivations in regulated grammars (see [2]), three types of leftmost derivation restrictions placed upon one-sided random context grammars have been defined and studied in [7]. In the *type*-1 *derivation restriction*, during every derivation step, the leftmost occurrence of a nonterminal has to be rewritten. In the *type*-2 *derivation restriction*, during every derivation step, the leftmost occurrence of a nonterminal which can be rewritten has to be rewritten. In the *type*-3 *derivation restriction*, during every derivation step, a rule is chosen, and the leftmost occurrence of its left-hand side is rewritten. The following three results have been demonstrated: (I) One-sided random context grammars with type-1 leftmost derivations characterize the family of context-free languages. (II) One-sided random context grammars with type-2 and type-3 leftmost derivations characterize the family of recursively enumerable languages. (III) Propagating one-sided random context grammars with type-2 and type-3 leftmost derivations characterize the family of context-sensitive languages.

What is the generative power of the variants of one-sided random context grammars restricted to perform their derivations in the aforementioned three ways?

4 NEW INVESTIGATION AREAS

In this concluding section, we propose several areas suggested as topics of future investigation related to the study of one-sided random context grammars.

4.1 APPLICATIONS

As can be seen from the previous section, one-sided random context grammars have been studied from a rather strictly theoretical point of view. Nevertheless, we should study the applicability of these grammars in practice. Specifically, we believe that one-sided random context grammars can formally and elegantly simulate processing information in molecular genetics, including information concerning macromolecules, such as DNA, RNA, and polypeptides. For instance, consider an organism consisting of DNA molecules made by enzymes. It is a common phenomenon that a molecule m made by a specific enzyme can be modified unless molecules made by some other enzymes occur either to the left or to the right of m in the organism. Consider a word w that formalizes this organism so every molecule is represented by a symbol. As obvious, to simulate a change of the symbol a that represents m requires random context occurrences of some symbols that either precede or follow a in w. As obvious, one-sided random context grammars can provide a string-changing formalism that can capture this random context requirement in a very succinct and elegant way. To put it more generally, one-sided random context grammars can simulate the behavior of molecular organisms in a rigorous and uniform way. Application-oriented topics like this obviously represent a future investigation area concerning one-sided random context grammars.

4.2 GENERALIZED VERSIONS OF ONE-SIDED RANDOM CONTEXT GRAMMARS

We may generalize the concept of one-sided context from symbols to strings. Obviously, as one-sided random context grammars already characterize the family of recursively languages, such a generalization cannot increase their strength. However, a generalization likes this makes sense in terms of variants of one-sided random context grammars. In [8], one-sided forbidding grammars that can forbid strings instead of single symbols are studied, and it has been proved that they are computationally complete, even if all strings are formed by at most two symbols. What about other variants of one-sided random context grammars?

4.3 OTHER FORMAL MODELS EQUIPPED WITH ONE-SIDED RANDOM CONTEXT

One-sided random context grammars are based upon context-free grammars. It is only natural to consider other types of grammars and equip them with one-sided random context. Some preliminary results in this direction have been achieved in [9], where ET0L grammars and their variants enhanced with left random context are studied. Study this topic in terms of other types of grammars.

ACKNOWLEDGEMENTS

This work was supported by the following grants: BUT FIT-S-11-2, MŠMT ED1.1.00/02.0070, TAČR TE01010415, and CEZ MŠMT MSM0021630528.

REFERENCES

[1] E. Csuhaj-Varjú, T. Masopust, and G. Vaszil. Cooperating distributed grammar systems with permitting grammars as components. *Romanian Journal of Information Science and Technology*, 12(2):175–189, 2009.

[2] J. Dassow and G. Păun. *Regulated Rewriting in Formal Language Theory*. Springer, 1989.

[3] F. Goldefus, T. Masopust, and A. Meduna. Left-forbidding cooperating distributed grammar systems. *Theoretical Computer Science*, 20(3):1–11, 2010.

[4] A. Meduna and P. Zemek. One-sided random context grammars. *Acta Informatica*, 48(3):149–163, 2011.

[5] A. Meduna and P. Zemek. Nonterminal complexity of one-sided random context grammars. *Acta Informatica*, 49(2):55–68, 2012.

[6] A. Meduna and P. Zemek. One-sided forbidding grammars and selective substitution grammars. *International Journal of Computer Mathematics*, 89(5):586–596, 2012.

[7] A. Meduna and P. Zemek. One-sided random context grammars with leftmost derivations. In *LNCS Festschrift Series: Languages Alive*, volume 7300, pages 160–173. Springer Verlag, 2012.

[8] A. Meduna and P. Zemek. Generalized one-sided forbidding grammars. *International Journal of Computer Mathematics*, 90(2):127–182, 2013.

[9] A. Meduna and P. Zemek. Left random context ET0L grammars. *Fundamenta Informaticae*, to appear in 2013.

[10] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages, Volumes 1 through 3*. Springer, 1997.

[11] P. Zemek. Normal forms of one-sided random context grammars. In *Proceedings of the 18th Conference STUDENT EEICT 2012*, pages 430–434, Brno University of Technology, CZ, 2012.