One-Sided Random Context Grammars

Petr Zemek*
Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 1/2, 612 66 Brno, Czech Republic
izemek@fit.vutbr.cz

Abstract

This extended abstract of the doctoral thesis introduces the notion of a *one-sided random context grammar* as a context-free-based regulated grammar, in which 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 does not occur there. A right random context rule is applied analogically except that the symbols are examined to the right of the rewritten symbol.

The thesis is divided into three parts. The first part gives a motivation behind introducing one-sided random context grammars and places all the covered material into the scientific context. Then, it gives an overview of formal language theory and some of its lesser-known areas that are needed to fully grasp the upcoming topics.

The second part forms the heart of the thesis. It formally defines one-sided random context grammars and studies them from many points of view. Generative power, relations to other types of grammars, reduction, normal forms, leftmost derivations, parsing-related and generalized versions all belong between the studied topics.

The final part of the thesis closes its discussion by adding remarks regarding its coverage. More specifically, these remarks concern application perspectives, bibliography, and open problem areas.

Categories and Subject Descriptors

F.4.3 [**Mathematical Logic and Formal Languages**]: Formal Languages—*Classes defined by grammars or automata, Operations on languages*

*Recommended by thesis supervisor: Prof. Alexander Meduna. Defended at Faculty of Information Technology, Brno University of Technology on TBA, 2014.

Keywords

formal language theory, regulated grammars, random context grammars, one-sided random context grammars, permitting grammars, forbidding grammars, reduction, generative power, normal forms, leftmost derivations, generalized versions, LL versions

1. Introduction

Formal languages, such as programming languages, are applied in a great number of scientific disciplines, ranging from biology through linguistics up to informatics (see [20]). As obvious, to use them properly, they have to be precisely specified in the first place. Most often, they are defined by mathematical models with finitely many rules by which they rewrite sequences of symbols, called strings.

Over its history, formal language theory has introduced a great variety of these language-defining models. Despite their diversity, they can be classified into two basic categories—generative and recognition language models. Generative models, better known as *grammars*, define strings of their language so their rewriting process generates them from a special start symbol. On the other hand, recognition models, better known as *automata*, define strings of their language by a rewriting process that starts from these strings and ends in a special set of strings, usually called final configurations.

Concerning grammars, the classical theory of formal languages has often classified all grammars into two fundamental categories—*context-free grammars* and *non-context-free grammars*. As their name suggests, context-free grammars are based upon context-free rules, by which these grammars rewrite symbols regardless of the context surrounding them. As opposed to them, non-context-free grammars rewrite symbols according to context-dependent rules, whose application usually depends on rather strict conditions placed upon the context surrounding the rewritten symbols, and such a way of context-dependent rewriting often makes them clumsy and inapplicable in practice. From this point of view, we obviously always prefer using context-free grammars, but they have their drawbacks, too. Perhaps most importantly, context-free grammars are significantly less powerful than non-context-free grammars. Considering all these pros and cons, it comes as no surprise that modern formal language theory has intensively and systematically struggled to come with new types of grammars that are underlined by context-free rules, but which are more powerful than ordinary context-free grammars. Regulated versions of context-free grammars, briefly referred to as *regulated grammars* in the thesis, represent perhaps the most successful and significant achievement in this direction. They are based upon context-free grammars extended by additional regulating mechanisms by which they control the way the language generation is performed.

Over the last four decades, formal language theory has introduced an investigated many types of regulated grammars (see [3, 11, 18], Chapter 13 of [9], and Chapter 3 of the second volume of [20] for an overview of the most important results). Arguably, one of the most studied type of regulated grammars are random context grammars, which are central to the thesis.

Random Context Grammars

In essence, *random context grammars* (see Section 1.1 in [3]) regulate the language generation process so they require the presence of some prescribed symbols and, simultaneously, the absence of some others in the rewritten sentential forms. More precisely, random context grammars are based upon context-free rules, each of which may be extended by finitely many *permitting* and *forbidding nonterminal symbols*. A rule like this can rewrite the current sentential form provided that all its permitting symbols occur in the sentential form while all its forbidding symbols do not occur there.

Random context grammars are significantly stronger than ordinary context-free grammars. In fact, they characterize the family of recursively enumerable languages (see Theorem 1.2.5 in [3]), and this computational completeness obviously represents their indisputable advantage. Also, *propagating random context grammars*, which do not have any erasing rules—that is, rules with the empty string on their right-hand sides—are stronger than context-free grammars. However, they are strictly less powerful than context-sensitive grammars. Indeed, they generate a language family that is strictly included in the family of context sensitive languages (see Theorem 1.2.4 in [3]).

From a pragmatical standpoint, however, random context grammars have a drawback consisting in the necessity of scanning the current sentential form in its entirety during every single derivation step. From this viewpoint, it is highly desirable to modify these grammars so they scan only a part of the sentential form, yet they keep their computational completeness. *One-sided random context grammars*—the topic of the thesis—represent a modification like this.

One-Sided Random Context Grammars

Specifically, in every one-sided random context grammar, the set of rules is divided into the set of *left random context rules* and the set of *right random context rules*. When applying a left random context rule, the grammar checks the existence and absence of its permitting and forbidding symbols, respectively, only in the prefix to the left of the rewritten nonterminal in the current sentential form. Analogously, when applying a right random context rule, it checks the existence and absence of its permitting and forbidding symbols, respectively, only in the suffix to the right of the rewritten nonterminal. Otherwise, it works just like any ordinary random context grammar.

As the main result of the thesis, we demonstrate that propagating versions of one-sided random context grammars, which possess no erasing rules, characterize the family of context-sensitive languages, and with erasing rules, they characterize the family of recursively enumerable languages.

Furthermore, we discuss the generative power of several special cases of one-sided random context grammars. In a greater detail, we prove that *one-sided permitting grammars*, which have only permitting rules, are more powerful than context-free grammars; on the other hand, they are no more powerful than so-called scattered context grammars (see [10]). *One-sided forbidding grammars*, which have only forbidding rules, are equivalent to so-called selective substitution grammars (see [7]). Finally, *left forbidding grammars*, which have only left-sided forbidding rules, are only as powerful as context-free grammars.

Apart from the generative power of one-sided random context grammars and their special cases, we investigate the following aspects of these grammars. First, we establish four normal forms of one-sided random context grammars, in which all rules satisfy some prescribed properties or format. Then, we study a reduction of one-sided random context grammars with respect to the number of nonterminals and rules. After that, we place three leftmost derivation restrictions on one-sided random context grammars and investigate their generative power. We also study generalized versions of one-sided random context grammars, in which strings of symbols rather than single symbols can be required or forbidden. Finally, we study one-sided random context grammars from a more practical viewpoint by investigating their parsing-related variants.

To give a summary, the thesis is primarily and principally meant as a theoretical treatment of one-sided random context grammars, which represent a modification of random context grammars. Apart from this theoretical treatment, however, we also cover some application perspectives to give the reader ideas about their applicability in practice.

Motivation

Taking into account the definition of one-sided random context grammars and all the results sketched above, we see that these grammars may fulfill an important role in the language theory and its applications for the following four reasons.

(I) From a practical viewpoint, one-sided random context grammars examine the existence of permitting symbols and the absence of forbidding symbols only within a portion of the current sentential form while ordinary random context grammars examine the entire current sentential form. As a result, the one-sided versions of these grammars work in a more economical and, therefore, efficient way than the ordinary versions. Moreover, one-sided random context grammars provide a finer control over the regulation process. Indeed, the designer of the grammar may select whether the presence or absence of symbols is examined to the left or to the right. In the case of ordinary random context grammars, this selection cannot be done since they scan the sentential forms in their entirety.

(II) The one-sided versions of propagating random context grammars are stronger than ordinary propagating random context grammars. Indeed, the language family defined by propagating random context grammars is properly included in the family of context-sensitive languages (see Theorem 1.2.4 in [3]). One-sided random context grammars are as powerful as ordinary random context grammars. These results come as a surprise because one-sided random context grammars examine only parts of sentential forms as pointed out in (I) above.

(III) Left forbidding grammars were introduced in [4], which also demonstrated that these grammars only define the family of context-free languages (see Theorem 1 in [4]). It is more than natural to generalize left forbidding grammars to one-sided forbidding grammars, which are stronger than left forbidding grammars (see Theorem 8 in this paper). As a matter of fact, even *propagating left permitting grammars*, introduced in [2], are stronger than left forbidding grammars because they define a proper superfamily of the family of context-free languages (see Theorem 10 in this paper). We also generalize left permitting grammars to one-sided permitting grammars and study their properties.

(IV) In the future, one might find results achieved in the thesis useful when attempting to solve some well-known open problems. Specifically, recall that every propagating scattered context grammar can be turned to an equivalent context-sensitive grammar (see Theorem 3.21 in [10]), but it is a longstanding open problem whether these two kinds of grammars are actually equivalent—the *PSC = CS problem* (see [10]). If in the future one proves that the propagating versions of one-sided permitting grammars and one-sided random context grammars are equivalent, then so are propagating scattered context grammars and context-sensitive grammars (see Theorem 9 in this paper), so the PSC = CS problem would be solved.

Organization

This extended abstract of the thesis is divided into five sections. After this introductory Section 1, Section 2 briefly reviews the needed notions from formal language theory. Then, Section 3 defines one-sided random context grammars and illustrates them by examples.

Section 4 represents the heart of this extended abstract. Indeed, it gives an overview of the established results and investigated topics concerning one-sided random context grammars. Generative power, relations to other types of grammars, reduction, normal forms, leftmost derivations, parsing-related and generalized versions all belong between the studied topics.

Section 5 closes the text by making several final remarks concerning the covered material with a special focus on its future developments. It concerns application perspectives of one-sided random context grammars, bibliographic comments, and open problem areas.

2. Preliminaries

We assume that the reader is familiar with formal language theory (see [20]). For a set Q, $\mathrm{card}(Q)$ denotes the cardinality of Q, and 2^Q denotes the power set of Q. For an alphabet (finite nonempty set) V, V^* represents the free monoid generated by V under the operation of concatenation. The unit of V^* is denoted by ε. For $x \in V^*$, $|x|$ denotes the length of x and $\mathrm{alph}(x)$ denotes the set of symbols occurring in x.

A *random context grammar* (see Section 1.1 in [3]) is a quadruple, $G = (N, T, P, S)$, where N and T are two disjoint alphabets of *nonterminals* and *terminals*, respectively, $S \in N$ is the *start symbol*, and $P \subseteq N \times (N \cup T)^* \times 2^N \times 2^N$ is a finite relation, called the set of *rules*. Set $V = N \cup T$. Each rule $(A, x, U, W) \in P$ is written as $(A \rightarrow x,\ U,\ W)$. The *direct derivation relation* over V^*, symbolically denoted by $\Rightarrow_G$, is defined as follows: if $u, v \in V^*$, $(A \rightarrow x,\ U,\ W) \in P$, $U \subseteq \mathrm{alph}(uAv)$, and $W \cap \mathrm{alph}(uAv) = \emptyset$, then $uAv \Rightarrow_G uxv$. U is called the *permitting context* and W is called the *forbidding context*. Let $\Rightarrow_G^*$ denote the reflexive-transitive closure of $\Rightarrow_G$. The *language of* G is denoted by $L(G)$ and defined as $L(G) = \{w \in T^* \mid S \Rightarrow_G^* w\}$.

Let $G = (N,\ T,\ P,\ S)$ be a random context grammar. Rules of the form $(A \rightarrow \varepsilon,\ U,\ W)$ are called *erasing rules*. If $(A \rightarrow x,\ U,\ W) \in P$ implies that $|x| \geq 1$, then G is a *propagating random context grammar*. If $(A \rightarrow x,\ U,\ W) \in P$ implies that $W = \emptyset$, then G is a *permitting grammar*. If $(A \rightarrow x,\ U,\ W) \in P$ implies that $U = \emptyset$, then G is a *forbidding grammar*. By analogy with propagating random context grammars, we define a *propagating permitting grammar* and a *propagating forbidding grammar*, respectively.

Denotation of Language Families

Throughout the rest of this paper, the language families under discussion are denoted in the following way. **RC**, **P**, and **F** denote the language families generated by random context grammars, permitting grammars, and forbidding grammars, respectively. The notation with the upper index $-\varepsilon$ stands for the corresponding propagating family. For example, $\mathbf{RC}^{-\varepsilon}$ denotes the family of languages generated by propagating random context grammars. **CF**, **CS**, and **RE** denote the families of context-free languages, context-sensitive languages, and recursively enumerable languages, respectively. $\mathbf{SC}^{-\varepsilon}$, **S**, and $\mathbf{S}^{-\varepsilon}$ denote the language families generated by propagating scattered context grammars (see [5]), selective substitution grammars (see [8, 19]), and propagating selective substitution grammars—that is, selective substitution grammars without erasing rules—, respectively.

3. Definitions and Examples

Next, we formally define one-sided random context grammars and their variants. In addition, we illustrate them by examples.

Definition 1. 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. Set $V = N \cup T$. The components V, N, T, P_L, P_R, and S are called the *total alphabet*, 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

$$(A \to x, U, W)$$

For $(A \to x,\ U,\ W) \in P_L$, U and W are called the *left permitting context* and the *left forbidding context*, respectively. For $(A \to x,\ U,\ W) \in P_R$, U and W are called the *right permitting context* and the *right forbidding context*, respectively. □

When applying a left random context rule, the grammar checks the existence and absence of its permitting and forbidding symbols, respectively, only in the prefix to the left of the rewritten nonterminal in the current sentential form. Analogously, when applying a right random context rule, it checks the existence and absence of its permitting and forbidding symbols, respectively, only in the suffix to the right of the rewritten nonterminal. The following definition states this formally.

Definition 2. Let $G = (N,\ T,\ P_L,\ P_R,\ S)$ be a one-sided random context grammar. The *direct derivation relation* over V^* is denoted by $\Rightarrow_G$ and defined as follows. Let $u, v \in V^*$ and $(A \to x,\ U,\ W) \in P_L \cup P_R$. Then,

$$uAv \Rightarrow_G uxv$$

if and only if

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

or

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

Let $\Rightarrow_G^*$ denote the reflexive-transitive closure of $\Rightarrow_G$. □

The language generated by a one-sided random context grammar is defined as usual—that is, it consists of strings over the terminal alphabet that can be generated from the start symbol.

Definition 3. Let $G = (N,\ T,\ P_L,\ P_R,\ S)$ be a one-sided random context grammar. The *language of* G is denoted by $L(G)$ and defined as

$$L(G) = \{w \in T^* \mid S \Rightarrow_G^* w\}$$ □

Next, we define several special variants of one-sided random context grammars.

Definition 4. Let $G = (N,\ T,\ P_L,\ P_R,\ S)$ be a one-sided random context grammar. Rules of the form $(A \to \varepsilon,\ U,\ W)$ are called *erasing rules*. If $(A \to x,\ U,\ W) \in P_L \cup P_R$ implies that $|x| \geq 1$, then G is a *propagating one-sided random context grammar*. If $(A \to x,\ U,\ W) \in P_L \cup P_R$ implies that $W = \emptyset$, then G is a *one-sided permitting grammar*. If $(A \to x, U, W) \in P_L \cup P_R$ implies that $U = \emptyset$, then G is a *one-sided forbidding grammar*. By analogy with propagating one-sided random context grammars, we define a *propagating one-sided permitting grammar* and a *propagating one-sided forbidding grammar*, respectively. □

Definition 5. Let $G = (N,\ T,\ P_L,\ P_R,\ S)$ be a one-sided random context grammar. If $P_R = \emptyset$, then G is a *left random context grammar*. If $P_R = \emptyset$ and $(A \to x,\ U,\ W) \in P_L$ implies that $W = \emptyset$, then G is a *left permitting grammar* (see [2]). If $P_R = \emptyset$ and $(A \to x,\ U,\ W) \in P_L$ implies that $U = \emptyset$, then G is a *left forbidding grammar* (see [4]). Their propagating versions are defined analogously as the propagating version of one-sided random context grammars. □

Next, we illustrate the above definitions by three examples.

Example 1. Consider the following 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 next four rules

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

and P_R contains the next three rules

$$(A \to a\bar{A}, \{B\}, \emptyset) \qquad (A \to \varepsilon, \{B\}, \emptyset)$$
$$(\bar{A} \to A, \{\bar{B}\}, \emptyset)$$

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

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

where $n \geq 1$. The empty string is generated by

$$S \Rightarrow_G AB \Rightarrow_G B \Rightarrow_G \varepsilon$$

Based on the previous observations, we see that G generates the non-context-free language $\{a^nb^nc^n \mid n \geq 0\}$. □

Example 2. Consider $K = \{a^nb^mc^m \mid 1 \leq m \leq n\}$. This non-context-free language is generated by the one-sided permitting grammar

$$G = \big(\{S, A, B, X, Y\}, \{a, b, c\}, P_L, \emptyset, S\big)$$

with P_L containing the following seven rules

$$(S \to AX,\ \emptyset,\ \emptyset) \qquad (X \to bc,\ \emptyset,\ \emptyset)$$
$$(A \to a,\ \emptyset,\ \emptyset) \qquad (X \to bYc,\ \{B\},\ \emptyset)$$
$$(A \to aB,\ \emptyset,\ \emptyset) \qquad (Y \to X,\ \{A\},\ \emptyset)$$
$$(B \to A,\ \emptyset,\ \emptyset)$$

Notice that G is, in fact, a propagating left permitting grammar. Observe that $(X \to bYc,\ \{B\},\ \emptyset)$ is applicable if B, produced by $(A \to aB,\ \emptyset,\ \emptyset)$, occurs to the left of X in the current sentential form. Similarly, $(Y \to X,\ \{A\},\ \emptyset)$ is applicable if A, produced by $(B \to A,\ \emptyset,\ \emptyset)$, occurs to the left of Y in the current sentential form. Consequently, we see that every derivation that generates $w \in L(G)$ is of the form

$$\begin{aligned} S &\Rightarrow_G AX \\ &\Rightarrow_G^* a^u AX \\ &\Rightarrow_G a^{u+1}BX \\ &\Rightarrow_G a^{u+1}BbYc \\ &\Rightarrow_G a^{u+1}AbYc \\ &\Rightarrow_G^* a^{u+1+v}AbYc \\ &\Rightarrow_G a^{u+1+v}AbXc \\ &\vdots \\ &\Rightarrow_G^* a^{n-1}Ab^{m-1}Xc^{m-1} \\ &\Rightarrow_G ab^{m-1}Xc^{m-1} \\ &\Rightarrow_G a^n b^m c^m = w \end{aligned}$$

where $u, v \geq 0$, $1 \leq m \leq n$. Hence, $L(G) = K$. □

Example 3. Consider the one-sided forbidding grammar

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

where P_L contains the following five rules

$$(S \to AB, \emptyset, \emptyset) \qquad (B' \to B, \emptyset, \{A'\})$$
$$(B \to bB'c, \emptyset, \{A, \bar{A}\}) \qquad (\bar{B} \to \varepsilon, \emptyset, \{\bar{A}\})$$
$$(B \to \bar{B}, \emptyset, \{A, A'\})$$

and P_R contains the following four rules

$$(A \to aA', \emptyset, \{B'\}) \qquad (A' \to A, \emptyset, \{B\})$$
$$(A \to \bar{A}, \emptyset, \{B'\}) \qquad (\bar{A} \to \varepsilon, \emptyset, \{B\})$$

Notice that every derivation that generates a nonempty string of $L(G)$ is of the form

$$\begin{aligned} S &\Rightarrow_G AB \\ &\Rightarrow_G aA'B \\ &\Rightarrow_G aA'bB'c \\ &\Rightarrow_G aAbB'c \\ &\Rightarrow_G aAbBc \\ &\Rightarrow_G^* a^n Ab^n Bc^n \\ &\Rightarrow_G a^n \bar{A}b^n Bc^n \\ &\Rightarrow_G a^n \bar{A}b^n \bar{B}c^n \\ &\Rightarrow_G a^n b^n \bar{B}c^n \\ &\Rightarrow_G a^n b^n c^n \end{aligned}$$

where $n \geq 1$. The empty string is generated by

$$S \Rightarrow_G AB \Rightarrow_G \bar{A}B \Rightarrow_G \bar{A}\bar{B} \Rightarrow_G \bar{B} \Rightarrow_G \varepsilon$$

Based on the previous observations, we see that G generates the non-context-free language $\{a^n b^n c^n \mid n \geq 0\}$. □

Denotation of Language Families

Throughout the rest of this paper, the language families under discussion are denoted in the following way. **ORC**, **OP**, and **OF** denote the language families generated by one-sided random context grammars, one-sided permitting grammars, and one-sided forbidding grammars, respectively. **LRC**, **LP**, and **LF** denote the language families generated by left random context grammars, left permitting grammars, and left forbidding grammars, respectively.

The notation with the upper index $-\varepsilon$ stands for the corresponding propagating family. For example, $\mathbf{ORC}^{-\varepsilon}$ denotes the family of languages generated by propagating one-sided random context grammars.

4. Results

In this section, we give an overview of the established results and studied topics concerning one-sided random context grammars.

The present section is divided into four subsections. Section 4.1 studies the generative power of one-sided random context grammars. Section 4.2 establishes four normal forms of these grammars. Section 4.3 investigates their descriptional complexity. Finally, Section 4.4 briefly mentions other topics related to one-sided random context grammars that have been investigated in the thesis.

4.1 Generative Power

First, we have investigated the generative power of one-sided random context grammars. In the thesis, it is proved that one-sided random context grammars characterize the family of recursively enumerable languages, and that their propagating versions characterize the family of context-sensitive languages.

Theorem 1. $\mathbf{ORC}^{-\varepsilon} = \mathbf{CS}$ *and* $\mathbf{ORC} = \mathbf{RE}$

Since $\mathbf{RC}^{-\varepsilon} \subset \mathbf{CS}$ and $\mathbf{RC} = \mathbf{RE}$ (see [3]), we have that one-sided random context grammars are equally powerful as random context grammars, while propagating one-sided random context grammars are more powerful than propagating random context grammars.

Theorem 2. $\mathbf{RC}^{-\varepsilon} \subset \mathbf{ORC}^{-\varepsilon} \subset \mathbf{RC} = \mathbf{ORC}$

Next, we consider one-sided forbidding grammars. In the thesis, it is proved that they have the same power as selective substitution grammars (see [8, 19]).

Theorem 3. $\mathbf{OF}^{-\varepsilon} = \mathbf{S}^{-\varepsilon}$ *and* $\mathbf{OF} = \mathbf{S}$

It is not known whether one-sided forbidding grammars or selective substitution grammars characterize the family of recursively enumerable languages. Also, it is not known whether these grammars without erasing rules characterize the family of context-sensitive languages.

Moreover, the thesis proves the following two results concerning the generative power of one-sided forbidding grammars, where the set of left random context rules coincides with the set of right random context rules.

Theorem 4. *A language K is context-free if and only if there is a one-sided forbidding grammar, $G = (N, T, P_L, P_R, S)$, satisfying $K = L(G)$ and $P_L = P_R$.*

Theorem 5. *Let $G = (N, T, P_L, P_R, S)$ be a one-sided forbidding grammar satisfying $P_L = P_R$. Then, there is a propagating one-sided forbidding grammar H such that $L(H) = L(G) - \{\varepsilon\}$.*

One-sided forbidding grammars are at least as powerful as forbidding grammars. This is stated in the next theorem.

Theorem 6. $\mathbf{F}^{-\varepsilon} \subseteq \mathbf{OF}^{-\varepsilon}$ *and* $\mathbf{F} \subseteq \mathbf{OF}$

In terms of left forbidding grammars and their power, [4] proves that they are no more powerful than context-free grammars.

Theorem 7. $\mathbf{LF}^{-\varepsilon} = \mathbf{LF} = \mathbf{CF}$

From Theorems 6 and 7 above and from the fact that $\mathbf{CF} \subset \mathbf{F}^{-\varepsilon}$ (see [3]), we obtain the following theorem, which relates the language families generated by left forbidding grammars, one-sided forbidding grammars, and forbidding grammars.

Theorem 8. $\mathbf{LF}^{-\varepsilon} = \mathbf{LF} \subset \mathbf{F}^{-\varepsilon} \subseteq \mathbf{OF}^{-\varepsilon} \subseteq \mathbf{OF}$

Finally, the following two theorems relate the language families generated by propagating one-sided permitting grammars and propagating left permitting grammars to other families of languages.

Theorem 9. $\mathbf{CF} \subset \mathbf{OP}^{-\varepsilon} \subseteq \mathbf{SC}^{-\varepsilon} \subseteq \mathbf{CS} = \mathbf{ORC}^{-\varepsilon}$

Theorem 10. $\mathbf{CF} \subset \mathbf{LP}^{-\varepsilon} \subseteq \mathbf{SC}^{-\varepsilon} \subseteq \mathbf{CS} = \mathbf{ORC}^{-\varepsilon}$

Recall that it is not known whether propagating scattered context grammars characterize the family of context-sensitive languages—that is, whether the inclusion $\mathbf{SC}^{-\varepsilon} \subseteq \mathbf{CS}$ above is, in fact, an identity (see [10]).

4.2 Normal Forms

Formal language theory has always struggled to turn grammars into *normal forms*, in which grammatical rules satisfy some prescribed properties or format because they are easier to handle from a theoretical as well as practical standpoint. Concerning context-free grammars, there exist two famous normal forms—the Chomsky and Greibach normal forms. In the former, every grammatical rule has on its right-hand side either a terminal or two nonterminals. In the latter, every grammatical rule has on its right-hand side a terminal followed by zero or more nonterminals. Similarly, there exist normal forms for general grammars, such as the Kuroda, Penttonen, and Geffert normal forms. In this section, we present four normal forms for one-sided random context grammars.

In the first normal form, the set of left random context rules coincides with the set of right random context rules.

Theorem 11. *Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. Then, there is a one-sided random context grammar, $H = (N', T, P'_L, P'_R, S)$, such that $L(H) = L(G)$ and $P'_L = P'_R$. Furthermore, if G is propagating, so is H.*

The second normal form represents a dual normal form to that in Theorem 11. Indeed, every one-sided random context grammar can be turned into an equivalent one-sided random context grammar with the sets of left and right random context rules being disjoint.

Theorem 12. *Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. Then, there is a one-sided random context grammar, $H = (N', T, P'_L, P'_R, S)$, such that $L(H) = L(G)$ and $P'_L \cap P'_R = \emptyset$. Furthermore, if G is propagating, so is H.*

The third normal form represents an analogy of the well-known Chomsky normal form for context-free grammars. However, since one-sided random context grammars with erasing rules are more powerful than their propagating versions, we allow the presence of erasing rules in the transformed grammar.

Theorem 13. *Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. Then, there is a one-sided random context grammar, $H = (N', T, P'_L, P'_R, S)$, such that $L(H) = L(G)$ and $(A \to x, U, W) \in P'_L \cup P'_R$ implies that $x \in N'N' \cup T \cup \{\varepsilon\}$. Furthermore, if G is propagating, so is H.*

In the fourth normal form, every rule has its permitting or forbidding context empty.

Theorem 14. *Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. Then, there is a one-sided random context grammar, $H = (N', T, P'_L, P'_R, S)$, such that $L(H) = L(G)$ and $(A \to x, U, W) \in P'_L \cup P'_R$ implies that $U = \emptyset$ or $W = \emptyset$. Furthermore, if G is propagating, so is H.*

4.3 Reduction

Recall that one-sided random context grammars characterize the family of recursively enumerable languages (see Theorem 1). Of course, it is more than natural to ask whether the family of recursively enumerable languages is characterized by one-sided random context grammars with a limited number of nonterminals or rules. The present section gives an affirmative answer to this question.

The next theorem states that ten nonterminals suffice to generate any recursively enumerable language by a one-sided random context grammar.

Theorem 15. *For every recursively enumerable language K, there exists a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$ and* $\mathrm{card}(N) = 10$.

The number of nonterminals can be also limited in terms of one-sided random context grammars satisfying the normal form from Theorem 11.

Theorem 16. *For every recursively enumerable language K, there exists a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$, $P_L = P_R$, and* $\mathrm{card}(N) = 13$.

To approach the reduction of the number of nonterminals in a finer way, the notion of a *right random context nonterminal* is introduced in the thesis. It is defined as a nonterminal that appears on the left-hand side of a right random context rule. The thesis has demonstrated how to convert any one-sided random context grammar to an equivalent one-sided random context grammar with two right random context nonterminals. This result has been proved also for propagating one-sided random context grammars.

Let us first define the above-mentioned measure formally.

Definition 6. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. If $(A \to x, U, W) \in P_R$, then A is a *right random context nonterminal.* The *number of right random context nonterminals* of G is denoted by $\mathrm{nrrcn}(G)$ and defined as

$$\mathrm{nrrcn}(G) = \mathrm{card}\big(\{A \mid (A \to x, U, W) \in P_R\}\big) \qquad \square$$

The next two theorems state that two right random context nonterminals suffice to keep the power of one-sided random context grammars unchanged.

Theorem 17. *For every recursively enumerable language K, there is a one-sided random context grammar H such that $L(H) = K$ and* $\mathrm{nrrcn}(H) = 2$.

Theorem 18. *For every context-sensitive language J, there is a propagating one-sided random context grammar H such that $L(H) = J$ and* $\mathrm{nrrcn}(H) = 2$.

By analogy with Definition 6, we may define a *left random context nonterminal* and their number in one-sided random context grammars. Then, in the thesis, it is shown that Theorems 17 and 18 can be reformulated in terms of left random context nonterminals and their number. The thesis also proves that we may limit both the total number of right and left random context nonterminals at the same time.

Finally, apart from reducing the overall number of nonterminals and right random context nonterminals, a reduction of the number of right random context rules has been investigated. 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 (see Definition 1).

Theorem 19. *For every recursively enumerable language K, there exists a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$ and* $\mathrm{card}(P_R) = 2$.

That is, we know that two right random context rules suffice to keep the generative power of one-sided random context grammars unchanged.

The next theorem says that it is possible to simultaneously reduce both the number of nonterminals and the number of right random context rules.

Theorem 20. *For every recursively enumerable language K, there exists a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$,* $\mathrm{card}(N) = 13$, $\mathrm{nrrcn}(H) = 2$, *and* $\mathrm{card}(P_R) = 2$.

4.4 Other Topics of Investigation

We conclude Section 4 by briefly mentioning other topics related to one-sided random context grammars that have been investigated in the thesis.

Leftmost Derivations

By analogy with the three well-known types of leftmost derivations in regulated grammars (see [3]), three types of leftmost derivation restrictions placed upon one-sided random context grammars have been defined and studied in the thesis. 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. In the thesis, the following three results are 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.

Generalized 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 power. However, a generalization likes this makes sense in terms of variants of one-sided random context grammars. In the thesis, 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.

LL One-Sided Random Context Grammars

In the previous sections, have introduced and studied one-sided random context grammars from a purely theoretical viewpoint. From a more practical viewpoint, however, it is also desirable to make use of them in such grammar-based application-oriented fields as syntax analysis (see [1]). An effort like this obviously gives rise to introducing and investigating their parsing-related variants, such as LL versions.

LL one-sided random context grammars, introduced in the thesis, represent ordinary one-sided random context grammars restricted by analogy with LL requirements placed upon LL context-free grammars. That is, for every positive integer k, (1) LL(k) one-sided random context grammars always rewrite the leftmost nonterminal in the current sentential form during every derivation step, and (2) if there are two or more applicable rules with the same nonterminal on their left-hand sides, then the sets of all terminal strings of length k that can begin a string obtained by a derivation started by using these rules are disjoint. The class of LL grammars is the union of all LL(k) grammars, for every $k \geq 1$.

We have introduced and investigated LL versions of one-sided random context grammars. We have proved that they generate the family of LL context-free languages. Taking a finer look at this generation, we have also demonstrated that the generation of languages by LL one-sided random context grammars can be more succinct than that by LL context-free grammars.

5. Concluding Remarks

This concluding section makes several final remarks concerning the material covered in the thesis with a special focus on its future developments. First, it suggests application perspectives of one-sided random context grammars (Section 5.1). Then, it chronologically summarizes the concepts and results achieved in most significant studies on the subject of the thesis (Section 5.2). Finally, this section lists the most important open problems resulting from the study of the thesis (Section 5.3).

5.1 Application Perspectives

As already stated in Section 1, the thesis is primarily and principally meant as a theoretical treatment of one-sided random context grammars. Nevertheless, to demonstrate their possible practical importance, we make some general remarks regarding their applications in the present section.

Taking the definition and properties of one-sided random context grammars into account, we see that they are suitable to underly information processing based on the existence or absence of some information parts. Therefore, in what follows, we pay major attention to this application area.

Molecular Genetics

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 string 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.

Computer Science

Considering the fact that one-sided random context grammars have a greater power than context-free grammars, we may immediately think of applying them in terms of syntax analysis of complicated non-context-free structures during language translation. However, as one-sided random context grammars are computationally complete (see Theorem 1), Rice's theorem (see Section 9.3.3 in [6]) implies that we cannot use them to parse all recursively enumerable languages. Therefore, we should focus on variants of one-sided random context grammars that are not computationally complete, such as propagating one-sided random context grammars.

At the end of Section 4.4, we have mentioned LL versions of one-sided random context grammars, which may be suitable for syntax analysis. Even though they are equally powerful as context-free grammars, they still may be useful since for some languages, they can describe languages in a more economical way. See the thesis for more details.

Linguistics

In terms of linguistics, one-sided random context grammars may be used for generating or verifying that the given texts contain no forbidding passages, such as vulgarisms or classified information. More specifically, generalized one-sided forbidding grammars, which are one-sided forbidding grammars that can forbid the occurrences of strings, are suitable to formally capture such applications.

Another application area of one-sided random context grammars may be syntax-oriented linguistics. Observe that many common English sentences contain expressions and words that mutually depend on each other although they are not adjacent to each other in the sentences. For example, consider the following sentence: *He sometimes goes to bed very late.* The subject (*he*) and the predicator (*goes*) are related. Therefore, we cannot rewrite *goes* to *go* because of the subject. One-sided random context grammars form a suitable formalism to capture and verify such dependencies.

Application-oriented topics like the ones outlined in this section obviously represent a future investigation area concerning one-sided random context grammars.

5.2 Bibliographical and Historical Remarks

This section gives an overview of the crucially important studies published on the subject of the thesis from a historical perspective.

One-sided random context grammars were introduced and investigated in [12]. Their special variants, left permitting and left forbidding grammars, were originally introduced in [2] and [4], respectively. The generative power of one-sided forbidding grammars and their relation to selective substitution grammars were studied in [14]. The nonterminal complexity of one-sided random context grammars was investigated in [13]. A reduction of the number of right random context rules was the topic of [17]. Several normal forms of these grammars were established in [21]. Leftmost derivations were studied in [15]. The generalized version of one-sided forbidding grammars was introduced and investigated in [16]. A list of open problems concerning these grammars appears in [22]. Finally, the LL versions of one-sided random context grammars appear in the thesis for the first time.

5.3 Open Problem Areas

We finish the thesis by summarizing open problems concerning one-sided random context grammars.

(I) What is the generative power of left random context grammars? What is the role of erasing rules in this left variant? That is, are left random context grammars more powerful than propagating left random context grammars?

(II) What is the generative power of one-sided forbidding grammars? We only know that these grammars are equally powerful as selective substitution grammars (see Theorem 3). Thus, by establishing the generative power of one-sided forbidding grammars, we would establish the power of selective substitution grammars, too.

(III) By Theorem 15, ten nonterminals suffice to generate any recursively enumerable language by a one-sided random context grammar. Is this limit optimal? In other words, can Theorem 15 be improved?

(IV) Recall that propagating one-sided random context grammars characterize the family of context-sensitive languages (see Theorem 1). Can we also limit the overall number of nonterminals in terms of this propagating version like in Theorem 15?

(V) What is the generative power of one-sided forbidding grammars and one-sided permitting grammars?

Moreover, what is the power of left permitting grammars? Recall that every propagating scattered context grammar can be turned to an equivalent context-sensitive grammar (see Theorem 3.21 in [10]), but it is a longstanding open problem whether these two kinds of grammars are actually equivalent—the *PSC = CS problem*. If in the future one proves that propagating versions of one-sided permitting grammars and one-sided random context grammars are equivalent, then so are propagating scattered context grammars and context-sensitive grammars (see Theorem 9), so the PSC = CS problem would be solved.

(VI) By Theorem 17, any recursively enumerable language is generated by a one-sided random context grammar having no more than two right random context nonterminals. Does this result hold with one or even zero right random context nonterminals? Notice that by proving that no right random context nonterminals are needed, we would establish the generative power of left random context grammars.

(VII) By Theorem 19, any recursively enumerable language is generated by a one-sided random context grammar having no more than two right random context rules. Does this result hold with one or even zero right random context rules? Again, notice that by proving that no right random context rules are needed, we would establish the generative power of left random context grammars.

Acknowledgements. This work was supported by the following grants: BUT FIT FIT-S-14-2299, European Regional Development Fund in the IT4Innovations Centre of Excellence (MŠMT CZ1.1.00/02.0070), and Visual Computing Competence Center (TE01010415).

References

[1] A. V. Aho, M. S. Lam, R. Sethi, and J. D. Ullman. *Compilers: Principles, Techniques, and Tools*. Addison-Wesley, Boston, 2nd edition, 2006.

[2] 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.

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

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

[5] S. A. Greibach and J. E. Hopcroft. Scattered context grammars. *Journal of Computer and System Sciences*, 3(3):233–247, 1969.

[6] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Boston, 3rd edition, 2006.

[7] H. C. M. Kleijn. *Selective Substitution Grammars Based on Context-Free Productions*. PhD thesis, Leiden University, Netherlands, 1983.

[8] H. C. M. Kleijn. Basic ideas of selective substitution grammars. In *Trends, Techniques, and Problems in Theoretical Computer Science*, volume 281 of *Lecture Notes in Computer Science*, pages 75–95. Springer, 1987.

[9] C. Martín-Vide, V. Mitrana, and G. Păun, editors. *Formal Languages and Applications*. Springer, Berlin, 2004.

[10] A. Meduna and J. Techet. *Scattered Context Grammars and their Applications*. WIT Press, Southampton, 2010.

[11] A. Meduna and P. Zemek. *Regulated Grammars and Their Transformations*. Faculty of Information Technology, Brno University of Technology, Brno, CZ, 2010.

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

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

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

[15] 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.

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

[17] A. Meduna and P. Zemek. One-sided random context grammars with a limited number of right random context rules. *Theoretical Computer Science*, 516(1):127–132, 2014.

[18] A. Meduna and P. Zemek. *Regulated Grammars and Automata*. Springer, New York, 2014.

[19] G. Rozenberg. Selective substitution grammars (towards a framework for rewriting systems). Part 1: Definitions and examples. *Elektronische Informationsverarbeitung und Kybernetik*, 13(9):455–463, 1977.

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

[21] P. Zemek. Normal forms of one-sided random context grammars. In *Proceedings of the 18th Conference STUDENT EEICT 2012*, volume 3, pages 430–434, Brno, CZ, 2012. Brno University of Technology.

[22] P. Zemek. One-sided random context grammars: Established results and open problems. In *Proceedings of the 19th Conference STUDENT EEICT 2013*, volume 3, pages 222–226, Brno, CZ, 2013. Brno University of Technology.

Selected Publications by the Author

A. Meduna, P. Zemek. *Regulated Grammars and Automata.* Springer, New York, 2014.

A. Meduna, P. Zemek. One-sided random context grammars with a limited number of right random context rules. *Theoretical Computer Science*, 516(1): 127–132, 2014.

A. Meduna, P. Zemek. Left random context ET0L grammars. *Fundamenta Informaticae*, 123(3): 289–304, 2013.

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

P. Zemek. One-sided random context grammars: Established results and open problems. In *Proceedings of the 19th Conference STUDENT EEICT 2013, Volume 3*, pages 222–226. Brno University of Technology, Brno, CZ, 2013.

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

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

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

P. Zemek. Normal forms of one-sided random context grammars. In *Proceedings of the 18th Conference STUDENT EEICT 2012, Volume 3*, pages 430–434. Brno University of Technology, Brno, CZ, 2012.

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

A. Meduna, P. Zemek. *Regulated Grammars and Their Transformations*. Faculty of Information Technology, Brno University of Technology, Brno, CZ, 2010.