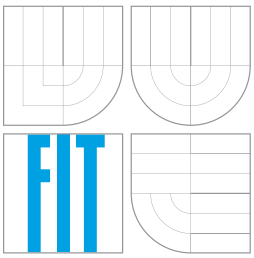
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INFORMAČNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INFORMATION SYSTEMS

ONE-SIDED RANDOM CONTEXT GRAMMARS

JEDNOSTRANNÉ GRAMATIKY S NAHODILÝM KONTEXTEM

ROZŠÍŘENÝ ABSTRAKT DISERTAČNÍ PRÁCE
EXTENDED ABSTRACT OF PHD THESIS

AUTOR PRÁCE — Ing. PETR ZEMEK
AUTHOR

VEDOUCÍ PRÁCE — prof. RNDr. ALEXANDER MEDUNA, CSc.
SUPERVISOR

BRNO 2014

Abstract

The 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 some of 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, generalized and parsing-related 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.

Keywords

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

Bibliographic Citation

Zemek, P.: One-sided random context grammars, Ph.D. thesis, Faculty of Information Technology, Brno University of Technology, Brno, CZ (2014)

Declaration

I hereby declare that the thesis is my own work that has been created under the supervision of prof. RNDr. Alexander Meduna, CSc. It is based on the following two books, one book chapter, and nine papers that I have written jointly with my supervisor: [12–20, 22, 23]. Furthermore, Chapter 9 is based on [11], which is a paper written together with Lukáš Vrábel. Where other sources of information have been used, they have been duly acknowledged.

Petr Zemek
March 1, 2014

Acknowledgements

The thesis was supported by several grants—namely, BUT FIT grants FIT-S-11-2 and FIT-S-14-2299, research plan CEZ MŠMT MSM0021630528, European Regional Development Fund in the IT4Innovations Centre of Excellence (MŠMT CZ1.1.00/02.0070), and Visual Computing Competence Center (TE01010415).

I wish to thank prof. RNDr. Alexander Meduna, CSc. for his support during his supervision of this work, for valuable and inspiring consultations, and for his advice and recommendations from which I have benefited greatly. I would also like to thank many colleagues from the university for fruitful discussions about formal languages. My special thanks go to Lukáš Vrábel, Zbyněk Křivka, Jiří Koutný, Ondřej Soukup, and Martin Čermák. Finally, I thank my family for their enthusiastic encouragement; most importantly, I deeply appreciate the great patience and constant support of my girlfriend Daniela.

Contents

Chapter 1
Introduction

Formal Languages and Regulated Grammars

Formal languages, such as programming languages, are applied in a great number of scientific disciplines, ranging from biology through linguistics up to informatics (see [21]). 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 this 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 regulated grammars (see [3, 13, 20], Chapter 13 of [7], and Chapter 3 of the second volume of [21] 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 present 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. Specifically, 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 [6]). 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 summarize, 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 4.2.2). 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 4.3.2). In the present thesis, 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 propagating one-sided permitting grammars and propagating one-sided random context grammars are equivalent, then so are propagating scattered context grammars and context-sensitive grammars (see Theorem 4.3.1), so the PSC = CS problem would be solved.

Organization

The text is divided into ten chapters. After this introductory Chapter 1, Chapter 2 briefly reviews formal language theory. It covers all the notions that are necessary to follow the rest of the thesis.

Chapters 3 through 9 represent the heart of the thesis. They introduce one-sided random context grammars and study them from many points of view. In a greater detail, Chapter 3 defines one-sided random context grammars and illustrates them by examples. Chapter 4 studies the generative power of these grammars. Chapter 5 establishes four normal forms of one-sided random context grammars. Chapter 6 investigates their descriptional complexity. Chapter 7 introduces three types of left-most derivation restrictions placed upon one-sided random context grammars, and studies their effect to the generative power of these grammars. Chapter 8 introduces and investigates generalized versions of one-sided random context grammars. Chapter 9 introduces and investigates parsing-related variants of one-sided random context grammars, which may be applied in practice.

Chapter 10 closes the thesis 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 references, and open problem areas.

Chapter 2
Rudiments of Formal Language Theory

The present chapter briefly reviews formal language theory. It consists of three sections. Section 2.1 gives the used mathematical notation. Section 2.2 covers strings and languages. Section 2.3 concerns grammars and language families.

2.1 Mathematical Notation

For a set Q, $\mathrm{card}(Q)$ denotes the cardinality of Q, and 2^Q denotes the power set of Q. For two sets P and Q, $P \subseteq Q$ denotes that P is a subset of Q; $P \subset Q$ denotes that $A \subseteq B$ and $A \neq B$. Set difference is denoted by $-$. The empty set is denoted by $\emptyset$.

2.2 Strings and Languages

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.

2.3 Grammars and Language Families

RE, **CS**, and **CF** denote the families of recursively enumerable languages, context-sensitive languages, and context-free languages, respectively. **RC**, **Per**, **For**, **S**, and **SC** denotes the families of languages generated by random context grammars, permitting grammars, forbidding grammars, selective substitution grammars, and scattered context grammars. To indicate that only propagating grammars—that is, grammars having no erasing rules—are considered, we use the upper index $-\varepsilon$. For example, $\mathbf{RC}^{-\varepsilon}$ denote the family of languages generated by propagating random context grammars.

Chapter 3
Definitions and Examples

This three-section chapter defines one-sided random context grammars and their variants, and illustrates them by examples. More specifically, Section 3.1 gives formal definitions of these grammars, Section 3.2 illustrates them by two examples, and Section 3.3 presents a denotation of language families generated by these grammars.

3.1 Definitions

Without further ado, let us define one-sided random context grammars formally.

Definition 3.1.1. A *one-sided random context grammar* is a quintuple

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

where N and T are two disjoint alphabets, $S \in N$, and

$$P_L, P_R \subseteq N \times \left(N \cup T\right)^* \times 2^N \times 2^N$$

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 3.1.2. 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 $\Rightarrow_G^n$ and $\Rightarrow_G^*$ denote the nth power of $\Rightarrow_G$, for some $n \geq 0$, and the reflexive-transitive closure of $\Rightarrow_G$, respectively. □

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 3.1.3. 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 3.1.4. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. 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 3.1.5. 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. □

3.2 Examples

Next, we illustrate the above definitions by two examples.

Example 3.2.1. 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

$$\begin{array}{ll} (S \to AB,\emptyset,\emptyset) & (\bar{B} \to B,\{A\},\emptyset) \\ (B \to b\bar{B}c,\{\bar{A}\},\emptyset) & (B \to \varepsilon,\emptyset,\{A,\bar{A}\}) \end{array}$$

and P_R contains the following three rules

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

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

$$\big\{a^nb^nc^n \mid n \geq 0\big\} \qquad \square$$

Example 3.2.2. 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

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

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, it is rather easy to see that every derivation that generates $w \in L(G)$ is of the form

$$\begin{aligned} S &\Rightarrow_G AX \\ &\Rightarrow^*_G a^u AX \\ &\Rightarrow_G a^{u+1} BX \\ &\Rightarrow_G a^{u+1} BbYc \\ &\Rightarrow_G a^{u+1} AbYc \\ &\Rightarrow^*_G a^{u+1+v} AbYc \\ &\Rightarrow_G a^{u+1+v} AbXc \\ &\quad\vdots \\ &\Rightarrow^*_G a^{n-1} Ab^{m-1} Xc^{m-1} \\ &\Rightarrow^2_G a^n b^m c^m = w \end{aligned}$$

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

3.3 Denotation of Language Families

Throughout the rest of the thesis, the language families under discussion are denoted in the following way. **ORC**, **OPer**, and **OFor** denote the language families generated by one-sided random context grammars, one-sided permitting grammars, and one-sided forbidding grammars, respectively. **LRC**, **LPer**, and **LFor** 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.

Chapter 4
Generative Power

In this chapter, consisting of Sections 4.1 through 4.3, we establish relations between the language families defined in the previous chapter and some well-known language families from Chapter 2.

4.1 One-Sided Random Context Grammars

First, we consider one-sided random context grammars.

Theorem 4.1.1. $\mathbf{ORC}^{-\varepsilon} = \mathbf{CS}$ *and* $\mathbf{ORC} = \mathbf{RE}$ □

4.2 One-Sided Forbidding Grammars

Next, we consider one-sided forbidding grammars.

Theorem 4.2.1. $\mathbf{OFor}^{-\varepsilon} = \mathbf{S}^{-\varepsilon}$ *and* $\mathbf{OFor} = \mathbf{S}$ □

Theorem 4.2.2. $\mathbf{LFor}^{-\varepsilon} = \mathbf{LFor} \subset \mathbf{For}^{-\varepsilon} \subseteq \mathbf{OFor}^{-\varepsilon} \subseteq \mathbf{OFor}$ □

Theorem 4.2.3. *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 $P_L = P_R$.* □

4.3 One-Sided Permitting Grammars

Finally, we consider one-sided permitting grammars and their generative power.

Theorem 4.3.1. $\mathbf{CF} \subset \mathbf{OPer}^{-\varepsilon} \subseteq \mathbf{SC}^{-\varepsilon} \subseteq \mathbf{CS} = \mathbf{ORC}^{-\varepsilon}$ □

Theorem 4.3.2. $\mathbf{CF} \subset \mathbf{LPer}^{-\varepsilon} \subseteq \mathbf{SC}^{-\varepsilon} \subseteq \mathbf{CS} = \mathbf{ORC}^{-\varepsilon}$ □

Theorem 4.3.3. $\mathbf{RC}^{-\varepsilon} \subset \mathbf{ORC}^{-\varepsilon} \subset \mathbf{RC} = \mathbf{ORC}$ □

Chapter 5
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 (see [9]). 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.

The present chapter establishes four normal forms for one-sided random context grammars. 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, mentioned above. In the fourth normal form, each rule has its permitting or forbidding context empty.

This chapter is divided into Sections 5.1 through 5.4. Each section establishes one of the above-mentioned normal forms of one-sided random context grammars.

5.1 First Normal Form

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

Theorem 5.1.1. *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 $P'_L = P'_R$.* □

Theorem 5.1.1 also holds if we restrict ourselves only to propagating one-sided random context grammars.

Theorem 5.1.2. *Let $G = (N, T, P_L, P_R, S)$ be a propagating one-sided random context grammar. Then, there is a propagating one-sided random context grammar, $H = (N', T, P'_L, P'_R, S)$, such that $L(H) = L(G)$ and $P'_L = P'_R$.* □

5.2 Second Normal Form

The second normal form represents a dual normal form to that in Theorems 5.1.1 and 5.1.2. 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 5.2.1. *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, then so is H.* □

5.3 Third Normal Form

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 5.3.1. *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, then so is H.* □

5.4 Fourth Normal Form

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

Theorem 5.4.1. *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, then so is H.* □

Chapter 6
Reduction

Recall that one-sided random context grammars characterize the family of recursively enumerable languages (see Theorem 4.1.1). 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 chapter, consisting of three sections, gives an affirmative answer to this question.

More specifically, in Section 6.1, we show that every recursively enumerable language can be generated by a one-sided random context grammar with no more than ten nonterminals. In addition, we show that an analogous result holds for thirteen nonterminals in terms of these grammars with the set of left random context rules coinciding with the set of right random context rules.

Then, in Section 6.2, we approach the discussion concerning the reduction of these grammars with respect to the number of nonterminals in a finer way. Indeed, we introduce 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, and demonstrate how to convert any one-sided random context grammar G to an equivalent one-sided random context grammar H with two right random context nonterminals. We also explain how to achieve an analogous conversion in terms of propagating versions of these grammars (recall that they characterize the family of context-sensitive languages, see Theorem 4.1.1). Similarly, we introduce the notion of a *left random context nonterminal* and show how to convert any one-sided random context grammar G to an equivalent one-sided random context grammar H with two left random context nonterminals. We explain how to achieve an analogous conversion in terms of propagating versions of these grammars, too.

Apart from reducing the number of nonterminals, we reduce the number of rules. More specifically, in Section 6.3, we show that any recursively enumerable language can be generated by a one-sided random context grammar having no more than two right random context rules. As a motivation behind limiting the number of right random context rules in these grammars, consider left random context gram-

mars, which are one-sided random context grammars with no right random context rules (see Section 3). Recall that it is an open question whether these grammars are equally powerful to one-sided random context grammars. To give an affirmative answer to this question, it is sufficient to show that in one-sided random context grammars, no right random context rules are needed. From this viewpoint, the above-mentioned result may fulfill a useful role during the solution of this problem in the future.

6.1 Total Number of Nonterminals

First, we investigate a reduction of the total number of nonterminals.

Theorem 6.1.1. *Let K be a recursively enumerable language. Then, there is a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$ and* $\mathrm{card}(N) = 10$. □

Theorem 6.1.2. *Let K be a recursively enumerable language. Then, there is a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$, $P_L = P_R$, and* $\mathrm{card}(N) = 13$. □

6.2 Number of Left and Right Random Context Nonterminals

In this section, we approach the discussion concerning the reduction of one-sided random context grammars with respect to the number of nonterminals in a finer way. Indeed, we introduce 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, and demonstrate how to convert any one-sided random context grammar G to an equivalent one-sided random context grammar H with two right random context nonterminals. We also explain how to achieve an analogous conversion in terms of propagating versions of these grammars (recall that they characterize the family of context-sensitive languages, see Theorem 4.1.1). Similarly, we introduce the notion of a *left random context nonterminal* and show how to convert any one-sided random context grammar G to an equivalent one-sided random context grammar H with two left random context nonterminals.

First, we define these two new measures formally.

Definition 6.2.1. 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

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

Left random context nonterminals and their number in a one-sided random context grammar are defined analogously.

Definition 6.2.2. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. If $(A \to x, U, W) \in P_L$, then A is a *left random context nonterminal*. The *number of left random context nonterminals* of G is denoted by $\mathrm{nlrcn}(G)$ and defined as

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

Theorem 6.2.3. *Let K be a recursively enumerable language. Then, there is a one-sided random context grammar, $H = (N, T, P_L, P_R, S)$, such that $L(H) = K$, $\mathrm{nrrcn}(H) = 4$, and $\mathrm{nlrcn}(H) = 6$.* $\square$

Theorem 6.2.4. *For every recursively enumerable language K, there exists a one-sided random context grammar H such that $L(H) = K$ and $\mathrm{nrrcn}(H) = 2$.* $\square$

Theorem 6.2.5. *For every recursively enumerable language K, there exists a one-sided random context grammar H such that $L(H) = K$ and $\mathrm{nlrcn}(H) = 2$.* $\square$

Theorem 6.2.6. *For every context-sensitive language K, there exists a propagating one-sided random context grammar H such that $L(H) = K$ and $\mathrm{nrrcn}(H) = 2$.* $\square$

Theorem 6.2.7. *For every context-sensitive language K, there exists a propagating one-sided random context grammar H such that $L(H) = K$ and $\mathrm{nlrcn}(H) = 2$.* $\square$

6.3 Number of Right Random Context Rules

In this section, we show that any recursively enumerable language can be generated by a one-sided random context grammar having no more than two right random context rules.

Theorem 6.3.1. *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 $\mathrm{card}(P_R) = 2$.* $\square$

Theorem 6.3.2. *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 $\mathrm{card}(P_R) = 2$.* $\square$

Chapter 7
Leftmost Derivations

The investigation of grammars that perform leftmost derivations is central to formal language theory as a whole. Indeed, from a practical viewpoint, leftmost derivations fulfill a crucial role in parsing, which represents a key application area of formal grammars (see [1]). From a theoretical viewpoint, an effect of leftmost derivation restrictions to the power of grammars restricted in this way represents an intensively investigated area of this theory as clearly indicated by many studies on the subject.

Considering the significance of leftmost derivations, it comes as no surprise that the present chapter pays a special attention to them. Indeed, it introduces three types of leftmost derivation restrictions placed upon one-sided random context grammars. In the *type*-1 *derivation restriction*, discussed in Section 7.1, during every derivation step, the leftmost occurrence of a nonterminal has to be rewritten. In the *type*-2 *derivation restriction*, covered in Section 7.2, during every derivation step, the leftmost occurrence of a nonterminal which can be rewritten has to be rewritten. In the *type*-3 *derivation restriction*, studied in Section 7.2, during every derivation step, a rule is chosen, and the leftmost occurrence of its left-hand side is rewritten.

In this chapter, we place the three above-mentioned leftmost derivation restrictions on one-sided random context grammars, and study their effect to the generative power of one-sided random context grammars.

7.1 Type-1 Leftmost Derivations

In the first derivation restriction type, during every derivation step, the leftmost occurrence of a nonterminal has to be rewritten. This type of leftmost derivations corresponds to the well-known leftmost derivations in context-free grammars.

Definition 7.1.1. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. The *type*-1 *direct leftmost derivation relation* over V^*, symbolically denoted by ${}_{\text{lm}}^{1}\Rightarrow_G$, is defined as follows. Let $u \in T^*$, $A \in N$ and $x, v \in V^*$. Then,

$$uAv \,{}_{\text{lm}}^{1}\!\Rightarrow_G uxv$$

if and only if

$$uAv \Rightarrow_G uxv$$

Let ${}_{\text{lm}}^{1}\!\Rightarrow_G^{n}$ and ${}_{\text{lm}}^{1}\!\Rightarrow_G^{*}$ denote the nth power of ${}_{\text{lm}}^{1}\!\Rightarrow_G$, for some $n \geq 0$, and the reflexive-transitive closure of ${}_{\text{lm}}^{1}\!\Rightarrow_G$, respectively. The ${}_{\text{lm}}^{1}$-*language* of G is denoted by $L(G, {}_{\text{lm}}^{1}\!\Rightarrow)$ and defined as

$$L\big(G, {}_{\text{lm}}^{1}\!\Rightarrow\big) = \big\{w \in T^* \mid S \,{}_{\text{lm}}^{1}\!\Rightarrow_G^{*} w\big\} \qquad \square$$

Notice that if the leftmost occurrence of a nonterminal cannot be rewritten by any rule, then the derivation is blocked.

The language families generated by one-sided random context grammars with type-1 leftmost derivations and propagating one-sided random context grammars with type-1 leftmost derivations are denoted by $\mathbf{ORC}({}_{\text{lm}}^{1}\!\Rightarrow)$ and $\mathbf{ORC}^{-\varepsilon}({}_{\text{lm}}^{1}\!\Rightarrow)$, respectively.

Theorem 7.1.2. $\mathbf{ORC}^{-\varepsilon}({}_{\text{lm}}^{1}\!\Rightarrow) = \mathbf{ORC}({}_{\text{lm}}^{1}\!\Rightarrow) = \mathbf{CF}$ $\square$

7.2 Type-2 Leftmost Derivations

In the second derivation restriction type, during every derivation step, the leftmost occurrence of a nonterminal that can be rewritten has to be rewritten.

Definition 7.2.1. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. The *type*-2 *direct leftmost derivation relation* over V^*, symbolically denoted by ${}_{\text{lm}}^{2}\!\Rightarrow_G$, is defined as follows. Let $u, x, v \in V^*$ and $A \in N$. Then,

$$uAv \,{}_{\text{lm}}^{2}\!\Rightarrow_G uxv$$

if and only if $uAv \Rightarrow_G uxv$ and there is no $B \in N$ and $y \in V^*$ such that $u = u_1Bu_2$ and $u_1Bu_2Av \Rightarrow_G u_1yu_2Av$.

Let ${}_{\text{lm}}^{2}\!\Rightarrow_G^{n}$ and ${}_{\text{lm}}^{2}\!\Rightarrow_G^{*}$ denote the nth power of ${}_{\text{lm}}^{2}\!\Rightarrow_G$, for some $n \geq 0$, and the reflexive-transitive closure of ${}_{\text{lm}}^{2}\!\Rightarrow_G$, respectively. The ${}_{\text{lm}}^{2}$-*language* of G is denoted by $L(G, {}_{\text{lm}}^{2}\!\Rightarrow)$ and defined as

$$L\big(G, {}_{\text{lm}}^{2}\!\Rightarrow\big) = \big\{w \in T^* \mid S \,{}_{\text{lm}}^{2}\!\Rightarrow_G^{*} w\big\} \qquad \square$$

The language families generated by one-sided random context grammars with type-2 leftmost derivations and propagating one-sided random context grammars with type-2 leftmost derivations are denoted by $\mathbf{ORC}({}_{\text{lm}}^{2}\!\Rightarrow)$ and $\mathbf{ORC}^{-\varepsilon}({}_{\text{lm}}^{2}\!\Rightarrow)$, respectively.

Theorem 7.2.2. $\mathbf{ORC}^{-\varepsilon}({}_{\text{lm}}^{2}\!\Rightarrow) = \mathbf{CS}$ *and* $\mathbf{ORC}({}_{\text{lm}}^{2}\!\Rightarrow) = \mathbf{RE}$ $\square$

7.3 Type-3 Leftmost Derivations

In the third derivation restriction type, during every derivation step, a rule is chosen, and the leftmost occurrence of its left-hand side is rewritten.

Definition 7.3.1. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. The *type*-3 *direct leftmost derivation relation* over V^*, symbolically denoted by ${}_{\mathrm{lm}}{}^{3}\Rightarrow_G$, is defined as follows. Let $u, x, v \in V^*$ and $A \in N$. Then,

$$uAv \; {}_{\mathrm{lm}}{}^{3}\Rightarrow_G \; uxv$$

if and only if $uAv \Rightarrow_G uxv$ and $\mathrm{alph}(u) \cap \{A\} = \emptyset$.

Let ${}_{\mathrm{lm}}{}^{3}\Rightarrow_G^n$ and ${}_{\mathrm{lm}}{}^{3}\Rightarrow_G^*$ denote the nth power of ${}_{\mathrm{lm}}{}^{3}\Rightarrow_G$, for some $n \geq 0$, and the reflexive-transitive closure of ${}_{\mathrm{lm}}{}^{3}\Rightarrow_G$, respectively. The ${}_{\mathrm{lm}}{}^{3}$-*language* of G is denoted by $L(G, {}_{\mathrm{lm}}{}^{3}\Rightarrow)$ and defined as

$$L\big(G, {}_{\mathrm{lm}}{}^{3}\Rightarrow\big) = \big\{ w \in T^* \mid S \; {}_{\mathrm{lm}}{}^{3}\Rightarrow_G^* \; w \big\} \qquad \square$$

Notice the following difference between the second and the third type. In the former, the leftmost occurrence of a rewritable nonterminal is chosen first, and then, a choice of a rule with this nonterminal on its let-hand side is made. In the latter, a rule is chosen first, and then, the leftmost occurrence of its left-hand side is rewritten.

The language families generated by one-sided random context grammars with type-3 leftmost derivations and propagating one-sided random context grammars with type-3 leftmost derivations are denoted by $\mathbf{ORC}({}_{\mathrm{lm}}{}^{3}\Rightarrow)$ and $\mathbf{ORC}^{-\varepsilon}({}_{\mathrm{lm}}{}^{3}\Rightarrow)$, respectively.

Theorem 7.3.2. $\mathbf{ORC}^{-\varepsilon}({}_{\mathrm{lm}}{}^{3}\Rightarrow) = \mathbf{CS}$ *and* $\mathbf{ORC}({}_{\mathrm{lm}}{}^{3}\Rightarrow) = \mathbf{RE}$ $\square$

Chapter 8
Generalized One-Sided Forbidding Grammars

In [8], so-called *generalized forbidding grammars* that are based upon context-free rules, each of which may be associated with finitely many *forbidding strings*, were introduced and investigated. A rule like this can rewrite a nonterminal provided that none of its forbidding strings occur in the current sentential form; apart from this, these grammars work just like context-free grammars. As opposed to context-free grammars, however, they are computationally complete—that is, they generate the family of recursively enumerable languages (see Theorem 1 in [8]), and this property obviously represents their crucially important advantage over ordinary context-free and forbidding grammars.

Taking a closer look at the rewriting process in generalized forbidding grammars, we see that they always verify the absence of forbidding strings within their entire sentential forms. To simplify and accelerate their rewriting process, it is obviously more than desirable to modify these grammars so they make this verification only within some prescribed portions of the rewritten sentential forms while remaining computationally complete. *Generalized one-sided forbidding grammars*, which are defined and studied in the present chapter, represent a modification satisfying these properties.

More precisely, in a generalized one-sided forbidding grammar, the set of rules is divided into the set of *left forbidding rules* and the set of *right forbidding rules*. When applying a left forbidding rule, the grammar checks the absence of its forbidding strings only in the prefix to the left of the rewritten nonterminal in the current sentential form. Similarly, when applying a right forbidding rule, it performs an analogous check to the right. Apart from this, it works like any generalized forbidding grammar.

This chapter is divided into two sections. First, Section 8.1 defines generalized one-sided forbidding grammars and illustrate them by an example. Then, Section 8.2 establishes their generative power.

8.1 Definitions and Examples

Without further ado, let us define generalized one-sided forbidding grammars and illustrate them by an example. For an alphabet N and a string $x \in N^*$, $\mathrm{sub}(x)$ denotes the set of all substrings of x, and $\mathrm{fin}(N)$ denotes the set of all finite languages over N.

Definition 8.1.1. A *generalized one-sided forbidding 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

$$P_L, P_R \subseteq N \times \big(N \cup T\big)^* \times \mathrm{fin}(N)$$

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 forbidding rules*, the set of *right forbidding rules*, and the *start symbol*, respectively. Each $(A, x, F) \in P_L \cup P_R$ is written as $(A \to x, F)$ throughout this chapter. For $(A \to x, F) \in P_L$, F is called the *left forbidding context*. Analogously, for $(A \to x, F) \in P_R$, F is called the *right forbidding context*. The *direct derivation relation* over V^*, symbolically denoted by $\Rightarrow_G$, is defined as follows. Let $u, v \in V^*$ and $(A \to x, F) \in P_L \cup P_R$. Then,

$$uAv \Rightarrow_G uxv$$

if and only if

$$(A \to x, F) \in P_L \text{ and } F \cap \mathrm{sub}(u) = \emptyset$$

or

$$(A \to x, F) \in P_R \text{ and } F \cap \mathrm{sub}(v) = \emptyset$$

Let $\Rightarrow_G^n$ and $\Rightarrow_G^*$ denote the nth power of $\Rightarrow_G$, for some $n \geq 0$, and the reflexive-transitive closure of $\Rightarrow_G$, respectively. The *language* of G is denoted by $L(G)$ and defined as

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

Next, we introduce the notion of a degree of G. Informally, it is the length of the longest string in the forbidding contexts of the rules of G. Let N be an alphabet. For $L \in \mathrm{fin}(N)$, $\text{max-len}(L)$ denotes the length of the longest string in L. We set $\text{max-len}(\emptyset) = 0$.

Definition 8.1.2. Let $G = (N, T, P_L, P_R, S)$ be a generalized one-sided forbidding grammar. G is of *degree* (m, n), where $m, n \geq 0$, if $(A \to x, F) \in P_L$ implies that $\text{max-len}(F) \leq m$ and $(A \to x, F) \in P_R$ implies that $\text{max-len}(F) \leq n$. □

Next, we illustrate the previous definitions by an example.

Example 8.1.3. Consider the generalized 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

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

and P_R contains the following four rules

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

Since the length of the longest string in the forbidding contexts of rules from P_L and P_R is 1, G is of degree $(1,1)$. It can be seen that G generates the non-context-free language

$$\big\{a^n b^n c^n \mid n \geq 0\big\}$$ □

The language family generated by generalized one-sided forbidding grammars of degree (m,n) is denoted by $\mathbf{GOF}(m,n)$. Furthermore, set

$$\mathbf{GOF} = \bigcup_{m,n\geq 0} \mathbf{GOF}(m,n)$$

8.2 Generative Power

In this section, we establish the generative power of generalized one-sided forbidding grammars.

Theorem 8.2.1. $\mathbf{GOF}(n,0) = \mathbf{GOF}(0,n) = \mathbf{CF}$ *for every* $n \geq 0$. □

Theorem 8.2.2. $\mathbf{CF} \subset \mathbf{GOF}(1,1) = \mathbf{OFor}$ □

Theorem 8.2.3. $\mathbf{GOF}(1,2) = \mathbf{GOF}(2,1) = \mathbf{RE}$ □

Theorem 8.2.4. *A language K is context-free if and only if there is a generalized one-sided forbidding grammar, $G = (N, T, P_L, P_R, S)$, satisfying $K = L(G)$ and $P_L = P_R$.* □

Chapter 9
LL One-Sided Random Context Grammars

In the previous chapters, 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—the subject of the present chapter.

LL one-sided random context grammars, introduced in this chapter, 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 $k \geq 1$.

Recall that one-sided random context grammars characterize the family of recursively enumerable languages (see Theorem 4.1.1). Of course, it is natural to ask whether LL one-sided random context grammars generate the family of LL context-free languages or whether they are more powerful. As its main result, this chapter shows that the families of LL one-sided random context languages and LL context-free languages coincide.

In fact, we take a closer look at the generation of languages by both versions of LL grammars. That is, we demonstrate an advantage of LL one-sided random context grammars over LL context-free grammars. More precisely, for every $k \geq 1$, we present a specific LL(k) one-sided random context grammar G and show that every equivalent LL(k) context-free grammar has necessarily more nonterminals or rules than G. Thus, to rephrase this result more broadly and pragmatically, we actually show that LL(k) one-sided random context grammars can possibly allow us

to specify LL(k) languages more succinctly and economically than LL(k) context-free grammars do.

This chapter is divided into three sections. First, Section 9.1 defines LL one-sided random context grammars. Then, Section 9.2 gives a motivational example. After that, Section 9.3 shows the main result sketched above, and formulates three open problems.

9.1 Definitions

In this section, we define LL one-sided random context grammars. Since we pay a principal attention to context-free and one-sided random context grammars working in the leftmost way, in what follows, by a context-free and one-sided random context grammar, respectively, we always mean a context-free and one-sided random context grammar working in the leftmost way, respectively. In terms of one-sided random context grammars, by this leftmost way, we mean the type-1 leftmost derivations (see Section 7.1).

Definition 9.1.1. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar and $\$ \notin N \cup T$ be a symbol. For every $r = (A \to x, U, W) \in P_L \cup P_R$ and $k \geq 1$, define

$$\text{Predict}_k(r) \subseteq T^*\{\$\}^*$$

as follows: $\gamma \in \text{Predict}_k(r)$ if and only if $|\gamma| = k$ and

$$S\$^k \overset{1}{\underset{\text{lm}}{\Rightarrow}}{}^*_G uAv\$^k \overset{1}{\underset{\text{lm}}{\Rightarrow}}_G uxv\$^k \overset{1}{\underset{\text{lm}}{\Rightarrow}}{}^*_G u\gamma w$$

where $u \in T^*$, $v, x \in V^*$, $w \in V^*\{\$\}^*$, and r is leftmost-applicable to uAv. □

Making use of the above definition, we next define LL one-sided random context grammars.

Definition 9.1.2. Let $G = (N, T, P_L, P_R, S)$ be a one-sided random context grammar. G is an *LL(k) one-sided random context grammar*, where $k \geq 1$, if it satisfies the following condition: for any $p = (A \to x, U, W), r = (A \to x', U', W') \in P_L \cup P_R$ such that $p \neq r$, if $\text{Predict}_k(p) \cap \text{Predict}_k(r) \neq \emptyset$, then there is no $w \in V^*$ such that $S \overset{1}{\underset{\text{lm}}{\Rightarrow}}{}^*_G w$ with both p and r being leftmost-applicable to w.

If there exists $k \geq 1$ such that G is an LL(k) one-sided random context grammar, then G is an *LL one-sided random context grammar*. □

9.2 A Motivational Example

In this short section, we give an example of an LL(k) one-sided random context grammar, for every $k \geq 1$. In this example, we argue that LL(k) one-sided random context grammars can describe some languages more succinctly than LL(k) context-free grammars.

Example 9.2.1. Let k be a positive integer and $G = (N, T, \emptyset, P_R, S)$ be a one-sided random context grammar, where $N = \{S\}$, $T = \{a, b, c, d\}$, and

$$P_R = \big\{(S \to d^{k-1}c, \emptyset, \emptyset), (S \to d^{k-1}aSS, \emptyset, \{S\}), (S \to d^{k-1}bS, \{S\}, \emptyset)\big\}$$

Notice that G is an LL(k) one-sided random context grammar. Observe that the second rule can be applied only to a sentential form containing exactly one occurrence of S, while the third rule can be applied only to a sentential form containing at least two occurrences of S. The generated language $L(G)$ can be described by the following expression

$$\big(d^{k-1}a(d^{k-1}b)^*d^{k-1}c\big)^*d^{k-1}c$$

In the thesis, we argue that $L(G)$ cannot be generated by any LL(k) context-free grammar having a single nonterminal and at most three rules. This shows us that for some languages, LL(k) one-sided random context grammars need fewer rules or nonterminals than LL(k) context-free grammars do to describe them. □

9.3 Generative Power

In this section, we show that LL one-sided random context grammars characterize the family of LL context-free languages. For every $k \geq 1$, let $\mathbf{LL\text{-}CF}(k)$ and $\mathbf{LL\text{-}ORC}(k)$ denote the families of languages generated by LL(k) context-free grammars and LL(k) one-sided random context grammars, respectively.

Theorem 9.3.1. $\mathbf{LL\text{-}ORC}(k) = \mathbf{LL\text{-}CF}(k)$ *for* $k \geq 1$. □

Define the language families $\mathbf{LL\text{-}CF}$ and $\mathbf{LL\text{-}ORC}$ as

$$\mathbf{LL\text{-}CF} = \bigcup_{k \geq 1} \mathbf{LL\text{-}CF}(k)$$

$$\mathbf{LL\text{-}ORC} = \bigcup_{k \geq 1} \mathbf{LL\text{-}ORC}(k)$$

Theorem 9.3.2. $\mathbf{LL\text{-}ORC} = \mathbf{LL\text{-}CF}$ □

Chapter 10
Concluding Remarks

This concluding chapter 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 10.1). Then, it chronologically summarizes the concepts and results achieved in most significant studies on the subject of the present thesis (Section 10.2). Finally, this chapter lists the most important open problems resulting from the study of the thesis (Section 10.3).

10.1 Application Perspectives

As already stated in Chapter 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 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 4.1.1), Rice's theorem (see Section 9.3.3 in [5]) 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.

In Chapter 9, we have studied 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 (see Theorem 9.3.2), they still may be useful since for some languages, they can describe languages in a more economical way (see Section 9.2).

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 (see Chapter 8), 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.

10.2 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 in [14]. 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 [16]. The nonterminal complexity of one-sided random context grammars was investigated in [15]. A reduction of the number of right random context rules was the topic of [19]. Several normal forms of these grammars were established in [22]. Leftmost derivations were studied in [17]. The generalized version of one-sided forbidding grammars was introduced and investigated in [18]. A list of open problems concerning these grammars appears in [23]. Finally, the LL versions of one-sided random context grammars are based on [11] and appear in the thesis for the first time.

10.3 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 they are equally powerful as selective substitution grammars (see Theorem 4.2.1). Thus, by establishing the generative power of one-sided forbidding grammars, we would establish the power of selective substitution grammars, too.

(III) By Theorem 6.1.1, 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 6.1.1 be improved?

(IV) Recall that propagating one-sided random context grammars characterize the family of context-sensitive languages (see Theorem 4.1.1). Can we also limit

the overall number of nonterminals in terms of this propagating version like in Theorem 6.1.1?

(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 one-sided permitting grammars and propagating one-sided random context grammars are equivalent, then so are propagating scattered context grammars and context-sensitive grammars (see Theorem 4.3.1), so the PSC = CS problem would be solved.

(VI) By Theorem 6.2.4, 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 6.3.1, 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.

References

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

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

[3] Dassow, J., Păun, G.: Regulated Rewriting in Formal Language Theory. Springer, New York (1989)

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

[5] Hopcroft, J.E., Motwani, R., Ullman, J.D.: Introduction to Automata Theory, Languages, and Computation, 3rd edn. Addison-Wesley, Boston (2006)

[6] Kleijn, H.C.M.: Selective substitution grammars based on context-free productions. Ph.D. thesis, Leiden University, Netherlands (1983)

[7] Martín-Vide, C., Mitrana, V., Păun, G. (eds.): Formal Languages and Applications. Springer, Berlin (2004)

[8] Meduna, A.: Generalized forbidding grammars. International Journal of Computer Mathematics **36**(1-2), 31–38 (1990)

[9] Meduna, A.: Automata and Languages: Theory and Applications. Springer, London (2000)

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

[11] Meduna, A., Vrábel, L., Zemek, P.: LL one-sided random context grammars. Unpublished manuscript

[12] Meduna, A., Zemek, P.: One-sided random context grammars: A survey. Unpublished manuscript

[13] Meduna, A., Zemek, P.: Regulated Grammars and Their Transformations. Faculty of Information Technology, Brno University of Technology, Brno, CZ (2010)

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

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

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

[17] Meduna, A., Zemek, P.: One-sided random context grammars with leftmost derivations. In: LNCS Festschrift Series: Languages Alive, vol. 7300, pp. 160–173. Springer Verlag (2012)

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

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

[20] Meduna, A., Zemek, P.: Regulated Grammars and Automata. Springer, New York (2014)

[21] Rozenberg, G., Salomaa, A. (eds.): Handbook of Formal Languages, Volumes 1 through 3. Springer, New York (1997)

[22] Zemek, P.: Normal forms of one-sided random context grammars. In: Proceedings of the 18th Conference STUDENT EEICT 2012, vol. 3, pp. 430–434. Brno University of Technology, Brno, CZ (2012)

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

Curriculum Vitae

Ing. Petr Zemek

Contact Information	Božetěchova 99 612 00 Brno Czech Republic	Phone: +420 608 453 199 E-mail: petr.zemek@volny.cz Web: http://www.fit.vutbr.cz/~izemek/
Personal Information	Date of Birth: December 13, 1985 Place of Birth: Opava	Citizenship: Czech Republic Gender: Male

Education

Brno University of Technology, Brno

Doctoral degree Ph.D., Faculty of Information Technology **2010 – ***

- Study specialization: Computer Science and Engineering
- Dissertation topic: *One-Sided Random Context Grammars*
- Expected finish: 2014

Master's degree Ing., Faculty of Information Technology **2008 – 2010**

- Study specialization: Information Systems
- Thesis topic: *On Erasing Rules in Regulated Grammars*
- Final state examination, red diploma

Bachelor's degree Bc., Faculty of Information Technology **2005 – 2008**

- Study specialization: Information Technology
- Thesis topic: *Canonical Derivations in Programmed Grammars*
- Final state examination, red diploma

Mendel Grammar School, Opava **2001 – 2005**

Graduation exam

Publications

- 2 books
- 1 book chapter
- 12 international journal papers
- 10 international conference papers
- 3 international conference posters/presentations
- 5 student competition contributions

Honors and Awards

- Rector's award for excellent results obtained during Ph.D. study, Brno, 2013
- Joseph Fourier Prize, Prague, 2013, 5th place (student competition)
- Student EEICT 2013, Brno, 1st place (student competition)
- Student EEICT 2012, Brno, 1st place (student competition)
- Student EEICT 2011, Brno, 1st place (student competition)
- Dean's award for an excellent master's thesis, Brno, 2010
- Student EEICT 2010, Brno, 3rd place (student competition)
- Dean's award for an excellent bachelor's thesis, Brno, 2008
- Student EEICT 2008, Brno, 2nd place (student competition)
- International Young Physicists' Tournament, Leoben, Austria, 2004, 3rd place (physics, team competition)
- Becario Challenge, Prague, 2003, 3rd place (physics, team competition)

Professional Carreer

AVG Technologies, Brno

Developer — **4/2011 – ***

Faculty of Information Technology, BUT, Brno

Teacher and researcher — **11/2010 – ***

Private Secondary School of Business, Opava

Web and information system maintainer — **1/2003 – ***

Calitko, An extensible software P2P framework (open-source project)

Developer — **5/2007 – 10/2007**

Language Skills

English fluent both written and spoken

German basic knowledge of the language

Czech native speaker

Research and Other Interests

theoretical computer science (formal languages), programming languages and practices, software development, reverse engeneering (decompilation), open-source software, and operating systems (Linux)

Hobbies

technical literature, movies, music, mountain bike, nutrition