Grammars with Context Conditions and Their Applications
Grammars with Context Conditions and Their Applications
Grammars with Context Conditions and Their Applications

Alexander Meduna
Martin Švec

Brno University of Technology
Faculty of Information Technology
Brno, Czech Republic
To Professor Miroslav Novotný
Contents

Preface ix
Acknowledgments ix

1 Introduction 1

2 Preliminaries and Definitions 5
  2.1 Basic Definitions ........................................ 5
  2.2 Grammars .................................................. 8

3 Conditions Placed on Derivation Domains 15
  3.1 Sequential Grammars over Word Monoids ................. 15
  3.2 Parallel Grammars over Word Monoids .................... 24

4 Conditions Placed on the Use of Productions 33
  4.1 Sequential Conditional Grammars ....................... 33
    4.1.1 Context-Conditional Grammars .................... 33
    4.1.2 Random-Context Grammars ....................... 40
    4.1.3 Generalized Forbidding Grammars ................. 44
    4.1.4 Semi-conditional Grammars ..................... 58
    4.1.5 Simple Semi-conditional Grammars ............... 62
  4.2 Parallel Conditional Grammars ....................... 91
    4.2.1 Context-Conditional ET0L Grammars .......... 91
    4.2.2 Forbidding ET0L Grammars ...................... 98
    4.2.3 Simple Semi-conditional ET0L Grammars ....... 118
  4.3 Global Context Conditional Grammars ................. 131

5 Conditions Placed on the Neighborhood of Rewritten Symbols 137
  5.1 Continuous Context ..................................... 137
    5.1.1 Sequential Uniform Rewriting .................. 138
    5.1.2 Parallel Uniform Rewriting .................... 145
  5.2 Scattered Context ..................................... 149
    5.2.1 Scattered Context Grammars and Their Reduction ... 150
    5.2.2 Semi-parallel Uniform Rewriting ............... 162

6 Grammatical Transformations and Derivation Simulations 169
  6.1 Derivation Simulation .................................. 169
  6.2 Grammatical Simulation ................................ 175
  6.3 Simulation of E(0,1)L Grammars ...................... 177
7 Applications and Implementation
  7.1 Applications ........................................ 185
  7.2 Implementation .................................... 192

8 Concluding and Bibliographical Notes 199

Bibliography 201

Denotations of Language Families 213

Subject Index 215
Preface

The investigation of grammars with context conditions represents an exciting trend within the formal language theory. Although this investigation has introduced a number of new grammatical concepts and brought many remarkable results, all these concepts and results are scattered in various journal papers. In effect, the formal language theory lacks any monograph systematically and compactly summarizing this important trend. The primary goal of the present book is to provide such a monograph.

According to the types of context conditions, the present book classifies the grammars into three classes and sums up the crucial results about them. Specifically, this classification comes from the distinction between context conditions placed on (1) the domains of grammatical derivations, (2) the use of grammatical productions, and (3) the neighborhood of the rewritten symbols. In all three cases, the main attention is on establishing the grammatical generative power and important properties. In particular, this book studies how to reduce these grammars with respect to some of their components, such as the number of grammatical symbols or productions, in order to make the grammars small, succinct, and therefore easy to use. To demonstrate this practical use, it also discusses the applications and implementation of grammars with context conditions. Most of the applications are related to microbiology, which definitely belongs to the central application areas of computer science today.

No previous knowledge concerning the subject of this book is assumed on the part of the reader. Indeed, this book is self-contained in the sense that no other sources are needed for understanding all the presented material. Almost every new concept defined in the text is immediately illustrated by some examples to give it grasp. Every complicated mathematical passage is preceded by its intuitive explanation, so the reader should easily follow every proof in the book. All the applications given in the book are explained in a realistic way to clearly demonstrate the strong relation between the theoretical concept and its use in practice. Additional information found at http://www.fit.vutbr.cz/~meduna/books/gwcc.en.

Acknowledgments

We are indebted to many people for their assistance in various aspects of creating this book. We greatly benefited from conversations with our colleagues at the Brno University of Technology. Our special thanks go to Vladimír Čech. We are grateful to our Editor, Val Molière and Editorial Assistant, Emily Simmons at John Wiley & Sons, who were very encouraging and helpful during the preparation of this book. We gladly acknowledge support of GACR grant 201/04/0441. Most important, we thank our families for their constant patience and encouragement.

Alexander Meduna and Martin Švec.
Chapter 1
Introduction

Formal languages fulfill a crucial role in many computer science areas, ranging from compilers through mathematical linguistics to molecular genetics. In dealing with these languages, we face the problem of choosing appropriate models in order to capture their structure elegantly and precisely. By analogy with the specification of natural languages, we often base these models on suitable grammars.

A grammar generates its language by performing derivation steps that change strings, called sentential forms, to other strings according to its grammatical productions. During a derivation step, the grammar rewrites a part of its current sentential form with a string according to one of its productions. If in this way it can make a sequence of derivation steps from its start symbol to a sentential form consisting of terminal symbols—that is, the symbols over which the language is defined—the resulting sentential form is called a sentence and belongs to the generated language. The set of all sentences made in this way is the language generated by the grammar.

In classical formal language theory, we can divide grammatical productions into context-dependent and context-independent productions. Based on this division, we can make a natural distinction between context-dependent grammars, such as phrase-structure grammars, and context-independent grammars, such as context-free grammars. The derivation step by context-dependent productions depends on rather strict conditions, usually placed on the context surrounding the rewritten symbol, while the derivation step by context-independent productions does not have any restrictions. For this reason, we tend to use context-independent grammars. Unfortunately, compared to context-dependent grammars, context-independent grammars are far less powerful; in fact, most of these grammars are incapable of grasping some basic aspects of common programming languages. On the other hand, most context-dependent grammars are as powerful as the Turing machines, and this remarkable power gives them an indisputable advantage.

From a realistic point of view, the classical context-independent and context-dependent grammars have some other disadvantages. Consider, for instance, English. Context-independent grammars are obviously incapable of capturing all the contextual dependencies in this complex language. However, we may find even the classical context-dependent grammars clumsy for this purpose. To illustrate, in an English sentence, the proper form of verb usually depends on the form of the subject. For instance, we write I do it, not I does it, and it is the subject, I, that implies the proper form of do. Of course, there may occur several words, such as
adverbs, between the subject and the verb. We could extend *I do it* to *I often do it*, *I very often do it* and infinitely many other sentences in this way. At this point, however, the classical context-dependent productions, whose conditions are placed on the context surrounding the rewritten symbol, are hardly of any use. The proper form of the verb follows from a subject that does not surround the verb at all; it can occur many words ahead of the verb.

To overcome the difficulties and, at the same time, maintain the advantages described above, modern language theory has introduced some new grammars that simultaneously satisfy these three properties:

- They are based on context-independent productions.
- Their context conditions are significantly more simple and flexible than the strict condition placed on the context surrounding the rewritten symbol in the classical context-dependent grammars.
- They are as powerful as classical context-dependent grammars.

In the present book, we give an overview of the most essential types of these grammars. Their alternative context conditions can be classified into these three categories:

- Context conditions placed on derivation domains.
- Context conditions placed on the use of productions.
- Context conditions placed on the neighborhood of the rewritten symbols.

As already pointed out, we want the context conditions to be as small as possible. For this reason, we pay a lot of attention to the reduction of context conditions in this book. Specifically, we reduce the number of their components, such as the number of nonterminals or productions. We study how to achieve this reduction without any decrease of their generative power, which coincides with the power of the Turing machines. By achieving this reduction, we actually make the grammars with context conditions more succinct and economical, and these properties are obviously highly appreciated both from a practical and theoretical standpoint. Regarding each of the discussed grammars, we introduce and study their parallel and sequential versions, which represent two basic approaches to grammatical generation of languages in today’s formal language theory. To be more specific, during a sequential derivation step, a grammar rewrites a single symbol in the current sentential form whereas during a parallel derivation step, a grammar rewrites all symbols. As context-free and E0L grammars represent perhaps the most fundamental sequential and parallel grammars, respectively, we usually base the discussion of sequential and parallel generation of languages on them.
Chapter 1: Introduction

Organization

The text consists of the following chapters:

Chapter 2 gives an introduction to formal languages and their grammars.

Chapter 3 restricts grammatical derivation domains in a very simple and natural way. Under these restrictions, both sequential and parallel context-independent grammars characterize the family of recursively enumerable languages, which are defined by the Turing machines.

Chapter 4 studies grammars with conditional use of productions. In these grammars, productions may be applied on condition that some symbols occur in the current sentential form and some others do not. We discuss many sequential and parallel versions of these grammars in detail. Most important, new characterizations of some well-known families of L languages, such as the family of ET0L languages, are obtained.

Chapter 5 studies grammars with context conditions placed on the neighborhood of rewritten symbols. We distinguish between scattered and continuous context neighborhood. The latter strictly requires that the neighborhood of the rewritten symbols forms a continuous part of the sentential form while the former drops this requirement of continuity.

Chapter 6 takes a closer look at grammatical transformations, many of which are mentioned in the previous chapters. Specifically, it studies how to transform grammars with context-conditions to some other equivalent grammars so that both the input grammars and the transformed grammars generate their languages in a very similar way.

Chapter 7 demonstrates the use of grammars with context conditions by several applications related to biology.

Chapter 8 summarizes the main results of this book and presents several open problems. It makes historical notes and suggests some general references regarding the theoretical background of grammars with context conditions. In addition, it proposes new directions in the investigation of these grammars.

Approach

This book is theoretically oriented in its treatment of the grammars. It presents the formalism concerning grammars with enough rigor to make all results quite clear and valid. Every complicated mathematical passage is preceded by its intuitive explanation so that even the most complex parts of the book are easy to grasp. As most proofs of the results contain some transformations of grammars, the present book also provides an algorithmic approach to the grammatical models under discussion and shows how they are used in practice. Several worked-out examples and real-world applications give further illustrations of the theoretical notions.