Scattered Context Grammars and their Applications

WIT PRESS

WIT Press publishes leading books in Science and Technology.
Visit our website for the current list of titles.
www.witpress.com

WIT eLibrary

Home of the Transactions of the Wessex Institute, the WIT electronic-library provides the international scientific community with immediate and permanent access to individual papers presented at WIT conferences. Visit the WIT eLibrary at
http://library.witpress.com

Scattered Context Grammars and their Applications

A. Meduna & J. Techet

Brno University of Technology, Czech Republic

WIT PRESS Southampton, Boston

A. Meduna & J. Techet
Brno University of Technology, Czech Republic

Published by

WIT Press
Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK
Tel: 44 (0) 238 029 3223; Fax: 44 (0) 238 029 2853
E-Mail: witpress@witpress.com
http://www.witpress.com

For USA, Canada and Mexico

WIT Press
25 Bridge Street, Billerica, MA 01821, USA
Tel: 978 667 5841; Fax: 978 667 7582
E-Mail: infousa@witpress.com
http://www.witpress.com

British Library Cataloguing-in-Publication Data

A Catalogue record for this book is available
from the British Library

ISBN: 978-1-84564-426-0

Library of Congress Catalog Card Number: 2009931446

The texts of the papers in this volume were set individually by the authors or under their supervision.

No responsibility is assumed by the Publisher, the Editors and Authors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. The Publisher does not necessarily endorse the ideas held, or views expressed by the Editors or Authors of the material contained in its publications.

© WIT Press 2010

Printed in Great Britain by Athenaeum Press Ltd.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Publisher.

Contents

Preface ix

PART I: INTRODUCTION

CHAPTER 1
Motivation **3**

CHAPTER 2
Definitions **7**

2.1 Mathematical background 7
2.2 Basics of formal language theory 9
2.3 Scattered context grammars 22

PART II: THEORY

CHAPTER 3
Basic properties **29**

3.1 Normal forms 29
3.2 Closure properties 32
3.3 Generative power 42

CHAPTER 4
Further properties **49**

4.1 Terminating left-hand sides 49
4.2 Generalized *k*-limited erasing 52

CHAPTER 5
Restrictions and extensions **69**

5.1 *n*-Limited derivations 69
5.2 Leftmost derivations 83
5.3 Maximal and minimal derivations 89
5.4 Unordered scattered context grammars 98
5.5 Linear scattered context grammars 101
5.6 Extended propagating scattered context grammars 106

CHAPTER 6
Reduction and economy 113

6.1 Reduction 113
6.2 Economical transformations 126

CHAPTER 7
Parses and their generators 135

7.1 Terminology 136
7.2 General generators 139
7.3 Canonical generators 145
7.4 Reduced generators 152

PART III: APPLICATIONS AND CONCLUSIONS

CHAPTER 8
Applications in linguistics 161

8.1 Syntax and related linguistic terminology 162
8.2 Transformational scattered context grammars 165
8.3 Scattered context in English syntax 169

CHAPTER 9
Concluding remarks 179

Bibliography 185

Language family index 191

Symbol index 195

Subject index 197

Preface

Objective

Many modern information technologies cope with processing broadly scattered pieces of mutually related information throughout various computer systems. As a result, a mathematical exploration of scattered information processing based upon adequate formal models represents an important investigation area of today's informatics. In formal language theory, which often provides informatics with formal models, information processing of this kind is formalized by *scattered context grammars*. As these grammars formalize scattered information processing in a very natural and rigorous way, it comes as no surprise that formal language theory has devoted to their investigation a great number of studies since their introduction four decades ago. Although these studies have achieved many remarkable results, published in many international computer science journals and conferences, formal language theory lacks any monograph that systematically and compactly summarizes them. To fill this gap, the principal goal of the present book is to provide informatics with a monograph like this. This book gives an in-depth exploration of scattered context grammars and their languages from a theoretical point of view. In addition, from a practical point of view, it sketches how to process scattered information by these grammars in linguistics.

Approach

This book is primarily a theoretically oriented treatment of scattered context grammars. As a result, it introduces all formalisms with enough rigor to make all results clear and valid. Every complex concept defined in the text is immediately illustrated by some examples and every complicated mathematical passage is followed by its intuitive explanation, so the reader should easily follow every proof in the book. By describing their applications in linguistics and sketching several other application areas, the text also demonstrates their use in practice.

Organization

The text consists of the following three parts.

Part I consists of two chapters. Chapter 1 gives a rudimentary introduction to scattered context processing and its formalization by scattered context grammars. It explains why this formalization and its investigation is important in today's informatics. Chapter 2 reviews basic mathematical concepts, including fundamental notions of formal language theory. Most importantly, it rigorously defines scattered context grammars and their languages.

Part II, consisting of Chapters 3 through 7, presents the theory of scattered context grammars and their languages, and as such, it represents the core of this monograph. Chapters 3 and 4 establish crucial properties concerning the languages generated by scattered context grammars. Chapter 5 discusses several restricted and extended versions of scattered context grammars. Compared to the first three chapters of Part II, which are mostly mathematically oriented, Chapters 6 and 7 represent theoretical treatments with significant consequences regarding the use of scattered context grammars in practice. Chapter 6 explains how to describe these grammars in an economical way. In particular, it studies their reduction with respect to some of their components, such as the number of symbols or productions. As obvious, compared to their general versions, the resulting grammars described in this succinct way are easier to use in practice. Chapter 7 studies parsing based upon scattered context grammars. Apart from its interesting theoretical properties, parsing plays an important role in several pragmatically oriented areas of computer science and informatics, such as computational linguistics and compiler writing. As a result, this final chapter of Part II brings its theoretical discussion to the application-oriented discussion of the third part.

Part III consists of Chapters 8 and 9. To demonstrate the use of scattered context grammars, Chapter 8 describes their applications in linguistics. Finally, Chapter 9 closes the monograph by making several remarks, suggesting further theoretically oriented investigation areas concerning scattered context grammars, and bringing the reader's attention to some attractive application fields of these grammars, such as compiler writing and biology.

Prerequisites

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. Nonetheless, some background in today's formal language theory is definitely helpful to grasp the material of this monograph easily, and for this purpose, we recommend *Handbook of Formal Languages* (see [1]), which gives both an introduction to this theory and an overview of its recent trends.

Use

Principally, this monograph is meant as a general source of information about scattered context grammars, and therefore, it is useful to everybody interested in theoretical fundamentals of scattered information processing. Secondarily, it can be used as a textbook for a graduate-level course in formal language theory. Although it contains enough material for a full semester, we recommend to use it as a complementary text for a part of such a course, covering some selected essential topics of scattered context grammars. The text allows the flexibility needed to make a selection of this kind. Further backup materials related to this book, such as suggestions for further reading, are available at

`http://www.fit.vutbr.cz/~meduna/books/scg/`

Acknowledgments

We are indebted to many people for their assistance in various aspects of creating this book. We have greatly benefited from conversations with many of our colleagues at Brno University of Technology. Our special thanks go to Zbyněk Křivka and Tomáš Masopust, whose help was enormous. We are grateful to Elizabeth Cherry and Brian Privett at WIT Press, who were always helpful and encouraging during the preparation of this book. We gladly acknowledge the support of the GACR 102/05/H050, GACR 201/07/0005, and MSM 0021630528 grants. Most importantly, we thank our families for their constant support during our work on this book.

Alexander Meduna & Jiří Techet