

Lecture Notes in Computer Science

Edited by G. Goos and J. Hartmanis

93

Anton Nijholt

Context-Free Grammars:
Covers, Normal Forms,
and Parsing



Springer-Verlag
Berlin Heidelberg New York 1980

Editorial Board

W. Brauer P. Brinch Hansen D. Gries C. Moler G. Seegmüller
J. Stoer N. Wirth

Author

Anton Nijholt
Vrije Universiteit
Wiskundig Seminarium
De Boelelaan 1081
Postbus 7161
1007 MC Amsterdam
The Netherlands

AMS Subject Classifications (1979): 68F05, 68F25

CR Subject Classifications (1974): 4.12, 5.23

ISBN 3-540-10245-0 Springer-Verlag Berlin Heidelberg New York

ISBN 0-387-10245-0 Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Data. Nijholt, Anton, 1946-
Context-free grammars. (Lecture notes in computer science; 93) Bibliography: p.
Includes index. I. Formal languages. I. Title. II. Series.
QA267.3.N54. 511.3. 80-21378

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machine or similar means, and storage in data banks. Under § 54 of the German Copyright Law where copies are made for other than private use, a fee is payable to the publisher, the amount of the fee to be determined by agreement with the publisher.

© by Springer-Verlag Berlin Heidelberg 1980
Printed in Germany

Printing and binding: Beltz Offsetdruck, Hemsbach/Bergstr.
2145/3140-543210

PREFACE

This monograph develops a theory of grammatical covers, normal forms and parsing. Covers, formally defined in 1969, describe a relation between the sets of parses of two context-free grammars. If this relation exists then in a formal model of parsing it is possible to have, except for the output, for both grammars the same parser.

Questions concerning the possibility to cover a certain grammar with grammars that conform to some requirements on the productions or the derivations will be raised and answered. Answers to these cover problems will be obtained by introducing algorithms that describe a transformation of an input grammar into an output grammar which satisfies the requirements.

The main emphasis in this monograph is on transformations of context-free grammars to context-free grammars in some normal form. However, not only transformations of this kind will be discussed, but also transformations which yield grammars which have useful parsing properties.

Organization of the monograph

This monograph can be viewed as consisting of four parts.

The first part, Chapters 1 through 3, introduces the cover concept, the motivation of our research, the problems and, moreover, it reviews previous research.

The second part, Chapters 4 through 7, provides cover results for normal form transformations of context-free and regular grammars.

The third part, Chapters 8 through 10, is devoted to cover results for three classes of deterministically parsable grammars, viz. $LL(k)$, strict deterministic and $LR(k)$ grammars. In this part, a discussion of some syntactic aspects of compiler writing systems is included.

The fourth and final part of this monograph consists of Chapters 11 and 12. Chapter 11 contains a detailed discussion on simple chain grammars. Chapter 12 surveys parsing strategies for context-free grammars. In this chapter cover properties of transformations to $LL(k)$ and some other classes of grammars are considered.

A Bibliography and an Index appear at the end of the monograph.

A few sections and notes in this monograph are marked with a star. These starred sections and notes can be skipped without loss of continuity. Some of these starred sections and notes deal with syntax categories and grammar functors. Others deal with technical arguments on parsing at a moment that a reader who is not acquainted with some less conventional ideas of parsing will not grasp their significance.

The sections and notes on syntax categories are included to give the interested reader and the reader who is familiar with these concepts a notion of the differences and the similarities between these concepts and the grammar cover concept.

Moreover, it will become clear that in our grammar cover framework of Chapter 2 we have borrowed from ideas of the grammar functor approach.

We have tried to give full and formal proofs for most of the results which appear in this monograph. Only in those cases that proofs are available in publications elsewhere or in cases that *we* had the idea that a certain result should be clear because of its simplicity or because of what has been proven in the foregoing parts of the monograph, we have omitted a proof or formal detail.

Acknowledgements

Several people have helped me prepare this monograph. I should like to mention particularly Michael A. Harrison of the University of California at Berkeley and Jaco W. de Bakker of the Vrije Universiteit and the Mathematical Centre in Amsterdam for providing time, confidence and for their comments on a first handwritten version of the manuscript. Although not all their suggestions have been incorporated many improvements are due to their comments.

Other people, maybe sometimes unknowingly, did encourage me. Especially I want to mention Derick Wood of McMaster's University at Hamilton.

This monograph was prepared during my stay with the Vakgroep Informatica of the Department of Mathematics of the Vrije Universiteit in Amsterdam. I want to express my gratitude to Marja H., Marja V., Betty and Carla for being there and helping me. Carla Reuvecamp did an excellent job of typing the lengthy manuscript.

Anton Nijholt

April 1980.

CONTENTS

1. INTRODUCTION AND PRELIMINARIES	1
1.1. Introduction	1
1.2. Overview of the contents	4
1.3. Preliminaries	5
1.3.1. Grammars, automata and transducers	5
1.3.2.* Syntax categories	12
2. GRAMMAR COVERS AND RELATED CONCEPTS	14
2.1. Grammar covers	14
2.2. Restrictions on parse relations	20
2.2.1.* Some notes on parsing	22
2.2.2. Production directed parses	24
2.3.* Grammar functors	28
2.4. Related concepts	29
3. COVERS, PARSING AND NORMAL FORMS	32
3.1. Covers and parsing	32
3.2. Covers and normal forms: Historical notes	34
3.3. Covers and normal forms: An introduction	35

4. PROPERTIES OF COVERS AND PRELIMINARY TRANSFORMATIONS	38
4.1. Properties of covers	38
4.2. Preliminary transformations	41
5. NORMAL FORM COVERS FOR CONTEXT-FREE GRAMMARS	48
5.1. From proper grammars to non-left-recursive grammars	48
5.2. From non-left-recursive to Greibach normal form grammars	51
5.2.1. The 'substitution' transformation	51
5.2.2. The left part transformation	55
5.3. Transformations on Greibach normal form grammars	76
6. THE COVER-TABLE FOR CONTEXT-FREE GRAMMARS	80
7. NORMAL FORM COVERS FOR REGULAR GRAMMARS	85
8. DETERMINISTICALLY PARSABLE GRAMMARS	98
8.1. Introduction	98
8.2. Preliminaries	105
9. COVERS AND DETERMINISTICALLY PARSABLE GRAMMARS	111
9.1. Deterministically parsable grammars	111
9.2. On the covering of deterministically parsable grammars	117
10. NORMAL FORM COVERS FOR DETERMINISTICALLY PARSABLE GRAMMARS	127
10.1. Normal form covers for $LL(k)$ grammars	127
10.2. Normal form covers for strict deterministic grammars	141
10.3. Normal form covers for $LR(k)$ grammars	164

VII

11. COVER PROPERTIES OF SIMPLE CHAIN GRAMMARS	171
11.1. Simple chain grammars	172
11.2. Relationships between simple chain grammars and other classes of grammars	180
11.3. Simple chain languages	184
11.4. A left part theorem for simple chain grammars	189
11.5. Left part parsing and covering of simple chain grammars	196
12. TRANSFORMATIONS AND PARSING STRATEGIES: A CONCRETE APPROACH	205
12.1. Introduction	205
12.2. From LL(k) to LR(k) grammars: Parsing strategies	207
12.3. Transformations to LL(k) grammars	210
12.4. Parsing strategies revisited: A survey of recent research	230
BIBLIOGRAPHY	239
INDEX	248