GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS
Volume 17

Section: Algebra
P. M. Cohn and Roger Lyndon, *Section Editors*

Combinatorics on Words

M. Lothaire

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS
Volume 17

Section: Algebra
P. M. Cohn and Roger Lyndon, *Section Editors*

Combinatorics on Words

M. Lothaire

Foreword by

Roger Lyndon
University of Michigan

1983

Addison-Wesley Publishing Company
Advanced Book Program/World Science Division
Reading, Massachusetts

London • Amsterdam • Don Mills, Ontario • Sydney • Tokyo

Library of Congress Cataloging in Publication Data

Lothaire, M.
Combinatorics on words.

(Encyclopedia of mathematics and its applications; v. 17)
Bibliography: p.
Includes index.
1. Combinatorial analysis. 2. Word problems (Mathematics) I. Title. II. Series.
QA164.L67 511'.6 82-6716
ISBN 0-201-13516-7 AACR2

American Mathematical Society (MOS) Subject Classification Scheme (1980): 05-01, 05A99, 20MO5, 68-01, 68E99.

Copyright © 1983 by Addison-Wesley Publishing Company, Inc.
Published simultaneously in Canada.

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, Addison-Wesley Publishing Company, Inc., Advanced Book Program/World Science Division, Reading, Massachusetts 01867, U.S.A.

ABCDEFGHIJ-HA-898765432

Manufactured in the United States of America

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Volume		Section
1	LUIS A. SANTALO **Integral Geometry and Geometric Probability,** 1976 (2nd printing, with revisions, 1979)	Probability
2	GEORGE E. ANDREWS **The Theory of Partitions,** 1976 (2nd printing, 1981)	Number Theory
3	ROBERT J. McELIECE **The Theory of Information and Coding** A Mathematical Framework for Communication, 1977 (2nd printing, with revisions, 1979)	Probability
4	WILLARD MILLER, Jr. **Symmetry and Separation of Variables,** 1977	Special Functions
5	DAVID RUELLE **Thermodynamic Formalism** The Mathematical Structures of Classical Equilibrium Statistical Mechanics, 1978	Statistical Mechanics
6	HENRYK MINC **Permanents,** 1978	Linear Algebra
7	FRED S. ROBERTS **Measurement Theory** with Applications to Decisionmaking, Utility, and the Social Sciences, 1979	Mathematics and the Social Sciences
8	L. C. BIEDENHARN and J. D. LOUCK **Angular Momentum in Quantum Physics:** Theory and Application, 1981	Mathematics of Physics
9	L. C. BIEDENHARN and J. D. LOUCK **The Racah-Wigner Algebra in Quantum Theory,** 1981	Mathematics of Physics

GIAN-CARLO ROTA, *Editor*

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Volume		Section
10	JOHN D. DOLLARD and CHARLES N. FRIEDMAN **Product Integration** with Application to Differential Equations, 1979	Analysis
11	WILLIAM B. JONES and W. J. THRON **Continued Fractions:** Analytic Theory and Applications, 1980	Analysis
12	NATHANIEL F. G. MARTIN and JAMES W. ENGLAND **Mathematical Theory of Entropy,** 1981	Real Variables
13	GEORGE A. BAKER, Jr. and PETER GRAVES-MORRIS **Padé Approximants, Part I** **Basic Theory,** 1981	Mathematics of Physics
14	GEORGE A. BAKER, Jr. and PETER R. GRAVES-MORRIS **Padé Approximants, Part II:** **Extensions and Applications,** 1981	Mathematics of Physics
15	E. C. BELTRAMETTI and G. CASSINELLI **The Logic of Quantum** **Mechanics,** 1981	Mathematics of Physics
16	G. D. JAMES and A. KERBER **The Representation Theory** **of the Symmetric Group,** 1981	Algebra
17	M. LOTHAIRE **Combinatorics on Words,** 1982	Algebra

试读结束：需要全本请在线购买： www.ertongbook.com

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Volume		Section
18	HECTOR FATTORINI **The Abstract Cauchy Problem,** 1983	Analysis
19	G. G. LORENTZ, K. JETTER, and S. D. RIEMENSCHNEIDER **Birkhoff Interpolation,** 1983	Combinatorics
	Other volumes in preparation	

ENCYCLOPEDIA OF MATHEMATICS and Its Applications

GIAN-CARLO ROTA, Editor
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Editorial Board

Janos D. Aczel, *Waterloo*
George E. Andrews, *Penn State*
Richard Askey, *Madison*
Michael F. Atiyah, *Oxford*
Donald Babbitt, *U.C.L.A.*
Edwin F. Beckenbach, *U.C.L.A.*
Lipman Bers, *Columbia*
Garrett Birkhoff, *Harvard*
Salomon Bochner, *Rice*
Raoul Bott, *Harvard*
James K. Brooks, *Gainesville*
Felix E. Browder, *Chicago*
A. P. Calderón, *Buenos Aires*
Peter A. Carruthers, *Los Alamos*
S. Chandrasekhar, *Chicago*
S. S. Chern, *Berkeley*
Hermann Chernoff, *M.I.T.*
P. M. Cohn, *Bedford College, London*
H. S. MacDonald Coxeter, *Toronto*
George B. Dantzig, *Stanford*
Nelson Dunford, *Sarasota, Florida*
F. J. Dyson, *Inst. for Advanced Study*
Harold M. Edwards, *Courant*
Harvey Friedman, *Ohio State*
Giovanni Gallavotti, *Rome*
Andrew M. Gleason, *Harvard*
James Glimm, *Rockefeller*
A. González Domínguez, *Buenos Aires*
M. Gordon, *Essex*
Peter Henrici, *ETH, Zurich*
Nathan Jacobson, *Yale*
Mark Kac, *U.S.C.*
Shizuò Kakutani, *Yale*
Samuel Karlin, *Stanford*
J. F. C. Kingman, *Oxford*
Donald E. Knuth, *Stanford*
Joshua Lederberg, *Rockefeller*
André Lichnerowicz, *Collège de France*
M. J. Lighthill, *London*
Chia-Chiao Lin, *M.I.T.*
Jacques-Louis Lions, *Paris*
G. G. Lorentz, *Austin*
Roger Lyndon, *Ann Arbor*
Robert J. McEliece, *Urbana*
Henry McKean, *Courant*
Marvin Marcus, *Santa Barbara*
N. Metropolis, *Los Alamos*
Jan Mycielski, *Boulder*
L. Nachbin, *Rio de Janeiro* and *Rochester*
Steven A. Orszag, *M.I.T.*
Alexander Ostrowski, *Basle*
Roger Penrose, *Oxford*
Carlo Pucci, *Florence*
C. R. Rao, *Indian Statistical Institute*
Fred S. Roberts, *Rutgers*
Abdus Salam, *Trieste*
M. P. Schützenberger, *Paris*
Jacob T. Schwartz, *Courant*
Irving Segal, *M.I.T.*
Oved Shisha, *Technion*
I. M. Singer, *Berkeley*
Olga Taussky, *Caltech*
René Thom, *Bures-sur-Yvette*
John Todd, *Caltech*
John W. Tukey, *Princeton*
Stanislaw Ulam, *Colorado*
Veeravalli S. Varadarajan, *U.C.L.A.*
Antoni Zygmund, *Chicago*

Editor's Statement

A large body of mathematics consists of facts that can be presented and described much like any other natural phenomenon. These facts, at times explicitly brought out as theorems, at other times concealed within a proof, make up most of the applications of mathematics, and are the most likely to survive change of style and of interest.

This ENCYCLOPEDIA will attempt to present the factual body of all mathematics. Clarity of exposition, accessibility to the non-specialist, and a thorough bibliography are required of each author. Volumes will appear in no particular order, but will be organized into sections, each one comprising a recognizable branch of present-day mathematics. Numbers of volumes and sections will be reconsidered as times and needs change.

It is hoped that this enterprise will make mathematics more widely used where it is needed, and more accessible in fields in which it can be applied but where it has not yet penetrated because of insufficient information.

GIAN-CARLO ROTA

Section Editor's Foreword

This is the first book devoted to broad study of the combinatorics of *words*, that is to say, of sequences of symbols called *letters*. This subject is in fact very ancient and has cropped up repeatedly in a wide variety of contexts. Even in the most elegant parts of abstract pure mathematics, the proof of a beautiful theorem surprisingly often reduces to some very down to earth combinatorial lemma concerning linear arrays of symbols. In applied mathematics, that is in the subjects to which mathematics can be applied, such problems are even more to be expected. This is true especially in those areas of contemporary applied mathematics that deal with the discrete and non-commutative aspects of the world about us, notably the theory of automata, information theory, and formal linguistics.

The systematic study of words seems to have been initiated by Axel Thue in three papers [Norske Vid. Selsk. Skr. I Mat. Nat. Kl., Christiania, 1906, 1–22; 1912, 1–67; 1914, 1–34.]. Even more than for his theorems, we owe him a great debt for delineating this subject. Both before and after his time, a multitude of fragmentary results have accumulated in the most diverse contexts, and a substantial but not very widely known lore was beginning to crystallize to the point where a systematic treatment of the subject was badly needed and long over due.

This need is splendidly fulfilled by the present volume. It provides a clear and easily accessible introduction to the area, treating in some depth a representative selection of the most important branches of the subject. In particular, connections with free Lie algebras and algebras with polynomial identities are treated in full. The Preface by Dominique Perrin gives a lucid account of this book, and we need not say more on that matter. However, we want to amplify his remarks on the origins of this undertaking.

First, Marcel P. Schützenberger should be acknowledged as the "grandfather" of the book. It was Marco who initiated the systematic combinatorial and algebraic study of monoids, the natural habitat of words, and of the connections and applications of this subject to such classical areas as group representation theory, infinite groups, Lie algebras, probability theory, as well as to more recent "applied" subjects, notably computer science and mathematical linguistics. In addition to his own important and seminal work in these subjects, it was he who founded the most important school dealing with these and related subjects (now the Laboratoire Informatique Théorique et Programmation at Paris). Most of the contributors to this book are his former students, or students of theirs, and all are disciples of his teachings. Today, when he is not promulgating the virtues of monoids in hazardous foreign climes, he is walking the corridors of Paris VII stimulating the workers there and instigating new lines of research.

Second, this book is the result of a friendly collaboration of the group of authors that have realized it. This collective enterprise was initiated by Jean François Perrot and led to its conclusion by Dominique Perrin, who played the role of editor of the volume. He can be considered to be the "biological father" of this book, and I have been privileged to see him proving theorems, testing conjectures, as well as phoning long distance to obtain copy, editing copy, and carrying it to the post office. It is to this team that we are indebted for both the existence and the high quality of the present work.

It is a pleasure to witness such an auspicious official inauguration of a newly recognized mathematical subject, one which carries with it certain promise of continued increasingly broad development and application.

ROGER LYNDON

Preface

Combinatorics on words is a field that has grown separately within several branches of mathematics, such as group theory or probabilities, and appears frequently in problems of computer science dealing with automata and formal languages. It may now be considered as an independent theory because of both the number of results that it contains and the variety of possible applications.

This book is the first attempt to present a unified treatment of the theory of combinatorics on words. It covers the main results and methods in an elementary presentation and can be used as a textbook in mathematics or computer science at undergraduate or graduate level. It will also help researchers in these fields by putting together a lot of results scattered in the literature.

The idea of writing this book arose a few years ago among the group of people who have collectively realized it. The starting point was a mimeographed text of lectures given by M. P. Schützenberger at the University of Paris in 1966 and written down by J. F. Perrot. The title of this text was "Quelques Problèmes combinatoires de la théorie des automates." It was widely circulated and served many people (including most of the authors of this book) as an introduction to this field. It was J. F. Perrot's idea to make a book out of these notes, whose title varied from *The Little Red Book* in the sixties (by the color of its cover, later selected for this series) to *The Apocryphal*, a name that could dilute the responsibility for mistakes, if any, in the text.

Let us put aside for now the domestic history of this book and turn to its subject. The objects considered by people who study combinatorics on words are words, that is to say, sequences of elements taken from a set. Typical phenomena that can be observed in a word are certain kinds of repetitions, decompositions into words of a special sort, and the results of rearrangement of the letters. The type of results obtained is perhaps reminiscent of the beginnings of number theory.

The first significant works on the subject go back to the start of this century, appearing in A. Thue's papers on square-free words and MacMahon's treatise on combinatory analysis. Apart from pure combinatorics this kind of problem has also been studied by scholars dedicated to probabilities, especially to fluctuations of random variables. In pure algebra, problems on words appear in a number of situations, including free algebras, free groups, and free semigroups. More recently, the theory of au-

试读结束：需要全本请在线购买： www.ertongbook.com

tomata and formal languages was developed, taking inspiration from problems of computer science. For these problems, words are the natural object, because any computation gives rise to a sequence of elementary operations and also because a sequential machine can process an object only if it has a linear structure, which again is a word. The same observations obviously apply to natural languages. In fact, the beginnings of automata theory and of modern linguistics are interconnected.

The main originality of this book is that it gathers for the first time the pieces of the jigsaw puzzle just described. It is dedicated neither to group or semigroup theory nor to automata theory but only to words (although the origin or the consequences of the methods and results presented are explicitly mentioned all along). The subtlest difference between the subject of this book and any of the aforementioned theories is perhaps with automata theory. It can be roughly said that automata theory (and formal language theory) deals with sets of words whereas combinatorics on words considers properties of *one* word. This distinction is sometimes rather artificial, however, and the situation is the same as for determining what the work "combinatorial" exactly means. We have put aside another subject that is rather wide and also closely related to this one, the theory of codes, which will be the subject of another book. Let me now briefly present the contents of this book.

Chapter 1 contains the main definitions, together with some elementary properties of words frequently used in the sequel. The following three chapters deal with *unavoidable regularities*, which are properties of words that become true when their length tends to infinity. It is therefore a study of asymptotic properties of words.

Chapters 5–7 may be considered another block. They deal with properties of words related with classical noncommutative algebra. The first, Chapter 5, treats Lie algebras, the second is linked with nilpotent groups, and the third algebras with polynomial identities. The rest of the book, Chapters 8–11, deals with specific aspects of words, each worthy of a complete volume.

This book is written at a completely elementary level so as to be accessible to anyone with a standard mathematical background. The authorship of each chapter is different, but the notation is uniform throughout and the architecture of the book (including use of results from one chapter in another) is the result of the joint conception of the coauthors.

Each chapter ends with a series of problems. Either they comment upon particular cases of the results of the chapter, or, more often, they mention some additional results. Difficult ones are indicated with an asterisk or double asterisk.

It is a pleasure to express the thanks of the authors of this book for the collaboration that we received during its preparation. Dorothée Reutenauer translated the paper of Shirshov on which Chapter 7 relies. Howard

Straubing read several chapters and formulated useful comments. Volker Strehl carefully read Chapter 10 and helped with suggestions. We are also indebted to Georges Hansel, Gérard Lallement, André Lentin, Roger Lyndon, and Maurice Nivat and, of course, to Jean-François Perrot whose work was the starting point for this enterprise. The help of Claudine Beaujean, Maryse Brochot, Martine Chaumard, Monique Claverie, Arlette Dupont and Sylvie Lutzing, in particular for typing, is gratefully acknowledged.

To close this preface, I should like to mention the belief my coauthors and I share that this book will serve as an incentive for further developments of this beautiful theory. It might be the case that a good part of it will be superseded within a few years, and this is exactly what we hope.

DOMINIQUE PERRIN

Contents

CHAPTER 1

Words

1.0. Introduction

This chapter contains the main definitions used in the rest of the book. It also presents some basic results about words that are of constant use in the sequel. In the first section are defined words, free monoids, and some terms about words, such as length and factors.

Section 1.2 is devoted to submonoids and to morphism of free monoids, one of the basic tools for words. Many of the proofs of properties of words involve a substitution from the alphabet into words over another alphabet, which is just the definition of a morphism of free monoids. A nontrivial result called the *defect theorem* is proved. The theorem asserts that if a relation exists among words in a set, those words can be written on a smaller alphabet. This is a weak counterpart for free monoids of the Nielsen–Schreier theorem for subgroups of a free group.

In Section 1.3 the definition of conjugate words is given, together with some equivalent characterizations. Also defined are *primitive words*, or words that are not a repetition of another word. A very useful result, due to Fine and Wilf, is proved that concerns the possibility of multiple repetitions. The last section introduces the notation of formal series that deal with linear combinations of words, which will be used in Chapters 5–7 and 11.

A list of problems, some of them difficult, is collected at the end. Two of them (1.1.2 and 1.2.1) deal with free groups; their object is to point out the existence of a combinatorial theory of words in free groups, although the theory is not developed in the present book (see Lyndon and Schupp 1977). Two others (1.1.3 and 1.3.5) deal with the analysis of algorithms on words.

1.1. Free Monoids and Words

Let A be a set that we shall call an *alphabet*. Its elements will be called *letters*. (In the development of this book, it will often be necessary to suppose that the alphabet A is finite. Because this assumption is not always necessary, however, it will be mentioned explicitly whenever it is used.)

试读结束：需要全本请在线购买： www.ertongbook.com