

Volume 27

J CRM MONOGRAPH SERIES

Centre de Recherches Mathématiques  
Montréal

# Combinatorics on Words

## Christoffel Words and Repetitions in Words

Jean Berstel  
Aaron Lauve  
Christophe Reutenauer  
Franco V. Saliola



American Mathematical Society

# Combinatorics on Words

Christoffel Words and  
Repetitions in Words





# CRM MONOGRAPH SERIES

---

Centre de Recherches Mathématiques  
Montréal

## Combinatorics on Words Christoffel Words and Repetitions in Words

Jean Berstel  
Aaron Lauve  
Christophe Reutenauer  
Franco V. Saliola

The Centre de Recherches Mathématiques (CRM) of the Université de Montréal was created in 1968 to promote research in pure and applied mathematics and related disciplines. Among its activities are special theme years, summer schools, workshops, postdoctoral programs, and publishing. The CRM is supported by the Université de Montréal, the Province of Québec (FQRNT), and the Natural Sciences and Engineering Research Council of Canada. It is affiliated with the Institut des Sciences Mathématiques (ISM) of Montréal, whose constituent members are Concordia University, McGill University, the Université de Montréal, the Université du Québec à Montréal, and the École Polytechnique. The CRM may be reached on the Web at [www.crm.umontreal.ca](http://www.crm.umontreal.ca).



**American Mathematical Society**  
Providence, Rhode Island USA

The production of this volume was supported in part by the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) and the Natural Sciences and Engineering Research Council of Canada (NSERC).

2000 *Mathematics Subject Classification*. Primary 68R15;  
Secondary 37B10, 11J70, 68W40.

---

For additional information and updates on this book, visit  
[www.ams.org/bookpages/crmm-27](http://www.ams.org/bookpages/crmm-27)

---

**Library of Congress Cataloging-in-Publication Data**

Combinatorics on words : Christoffel words and repetitions in words / Jean Berstel . . . [et al.].  
p. cm. — (CRM monograph series ; v. 27)  
Includes bibliographical references and index.  
ISBN 978-0-8218-4480-9 (alk. paper)  
1. Combinatorial analysis. 2. Word problems (Mathematics) I. Berstel, Jean, 1941—

QA164.C6638 2008  
511'.6—dc22

2008036669

---

**Copying and reprinting.** Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for such permission should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to [reprint-permission@ams.org](mailto:reprint-permission@ams.org).

© 2009 by the American Mathematical Society. All rights reserved.  
The American Mathematical Society retains all rights  
except those granted to the United States Government.  
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines  
established to ensure permanence and durability.

This volume was submitted to the American Mathematical Society  
in camera ready form by the Centre de Recherches Mathématiques.

Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1      14 13 12 11 10 09

Ce livre est dédié à la mémoire de Pierre Leroux (1942–2008)



# Contents

Preface	ix
Notation	xi
<b>Part I. Christoffel Words</b>	<b>1</b>
Chapter 1. Christoffel Words	5
1. Geometric definition	5
2. Cayley graph definition	7
Chapter 2. Christoffel Morphisms	9
1. Christoffel morphisms	9
2. Generators	14
Chapter 3. Standard Factorization	17
1. The standard factorization	17
2. The Christoffel tree	20
Chapter 4. Palindromization	23
1. Christoffel words and palindromes	23
2. Palindromic closures	24
3. Palindromic characterization	29
Chapter 5. Primitive Elements in the Free Group $F_2$	33
1. Positive primitive elements of the free group	33
2. Positive primitive characterization	34
Chapter 6. Characterizations	39
1. The Burrows – Wheeler transform	39
2. Balanced <sub>1</sub> Lyndon words	41
3. Balanced <sub>2</sub> Lyndon words	41
4. Circular words	42
5. Periodic phenomena	44
Chapter 7. Continued Fractions	47
1. Continued fractions	47
2. Continued fractions and Christoffel words	48
3. The Stern – Brocot tree	52
Chapter 8. The Theory of Markoff Numbers	55
1. Minima of binary quadratic forms	55
2. Markoff numbers	56

3. Markoff's condition	58
4. Proof of Markoff's theorem	61
<b>Part II. Repetitions in Words</b>	<b>65</b>
Chapter 1. The Thue–Morse Word	69
1. The Thue–Morse word	69
2. The Thue–Morse morphism	70
3. The Tarry–Escott problem	71
4. Magic squares	74
Chapter 2. Combinatorics of the Thue–Morse Word	77
1. Automatic sequences	77
2. Generating series	79
3. Overlaps	81
4. Complexity	82
5. Formal languages	85
6. The Tower of Hanoi	90
Chapter 3. Square-Free Words	97
1. One example, three constructions	97
2. Square-free morphisms and codes	100
3. A 3-square-free test for square-freeness	102
4. A 2-square-free test for square-freeness	105
Chapter 4. Squares in Words	107
1. Counting squares	107
2. Centered squares	111
3. Prefix arrays	112
4. Crochemore factorization	114
5. Suffix trees	116
Chapter 5. Repetitions and Patterns	125
1. Maximal repetitions	125
2. Repetition thresholds	126
3. Patterns	127
4. Zimin patterns	131
5. Bi-ideal sequences	133
6. Repetitions in Sturmian words	134
Bibliography	135
Index	143

## Preface

This book grew out of two series of five two-hour lectures, given by Jean Berstel and Christophe Reutenauer in March 2007. The lectures were delivered during the school on “Combinatorics on Words” organized by Srećko Brlek, Christophe Reutenauer and Bruce Sagan that took part within the theme semester on *Recent Advances in Combinatorics on Words* at the Centre de recherches mathématiques (CRM), Montréal, Canada.

Notes for the lectures were written down by Aaron Lauve and Franco Saliola. They have augmented their notes with several topics and have added more than 100 exercises. There has been a lot of work in adding bibliographic references and a detailed index.

The text is divided into two parts. Part I, based on the lectures given by Christophe Reutenauer, is a comprehensive and self-contained presentation of the current state of the art in Christoffel words. These are finitary versions of Sturmian sequences. It presents relationships between Christoffel words and topics in discrete geometry, group theory, and number theory. Part I concludes with a new exposition of the theory of Markoff numbers.

Part II, based on the lectures by Jean Berstel, starts with a systematic exposition of the numerous properties, applications, and interpretations of the famous Thue–Morse word. It then presents work related to Thue’s construction of a square-free word, followed by a detailed exposition of a linear-time algorithm for finding squares in words. This part concludes with a brief glimpse of several additional problems with origins in the work of Thue.

**Acknowledgements.** We gratefully acknowledge the generosity of Amy Glen and Gwénaël Richomme, who agreed to read a preliminary version of this text. Implementation of their numerous comments improved the quality of the text tremendously. We also thank Anouk Bergeron-Brlek for lending us a third set of notes and Lise Tourigny through whom all things are possible. Finally, we would like to thank the CRM and François Bergeron, the principal organizer of the CRM theme semester, for providing an excellent scientific program and working environment during the semester as well as support throughout the preparation of this text.

**Typesetting.** The book was typeset with the L<sup>A</sup>T<sub>E</sub>X document preparation system together with the following L<sup>A</sup>T<sub>E</sub>X packages:

algorithm2e	color	multicol	stmaryrd
amsrefs	ednotes	paralist	subfigure
amssymb	gastex	pstricks	upref
array	graphicx	pst-poly	xy
bbm	mathtools	pst-tree	

and Will Robertson's L<sup>A</sup>T<sub>E</sub>X code for typesetting magic squares [Rob2005].

## Notation

We gather in one place the notational conventions shared by the two parts. The reader may also consult the subject index to locate the major occurrences within the text of most of the symbols and bold words below.

Let  $\mathbb{N}$  denote the set of nonnegative integers. If  $a, b$  and  $n$  are integers, then the notation  $a \equiv b \pmod{n}$  shall mean that  $a - b$  is divisible by  $n$ . Equivalently,  $a \equiv b \pmod{n}$  if and only if  $a$  and  $b$  have the same remainder upon division by  $n$ .

Let  $A$  denote a finite set of symbols. The elements of  $A$  are called **letters** and the set  $A$  is called an **alphabet**. A **word** over an alphabet  $A$  is an element of the free monoid  $A^*$  generated by  $A$ . The identity element  $\epsilon$  of  $A^*$  is called the **empty word**. Given a word  $w \in A^*$ , the **square** of  $w$  is the monoid product  $w^2 = ww$  in  $A^*$ . Higher powers of  $w$  are defined analogously. We frequently take  $A$  to be a subset of the nonnegative integers  $\mathbb{N}$ . The reader is cautioned to read 101 not as “one hundred and one” but as “ $1 \cdot 0 \cdot 1$ ,” an element of  $\{0, 1\}^3$ .

If  $w \in A^*$ , then there exists a unique integer  $r \geq 0$  and unique letters  $a_1, a_2, \dots, a_r \in A$  such that  $w = a_1 a_2 \cdots a_r$ ; the number  $r$  is called the **length** of  $w$  and denoted by  $|w|$ . A positive integer  $p$  is a **period** of  $w$  if  $a_i = a_{i+p}$  for all  $1 \leq i \leq |w| - p$ . (Note that if  $p \geq |w|$ , then  $p$  is a period of  $w$ .) If  $w \in A^*$  and  $a \in A$ , then  $|w|_a$  denotes the number of occurrences of the letter  $a$  in the word  $w$  so that

$$|w| = \sum_{a \in A} |w|_a.$$

If  $w = a_1 a_2 \cdots a_r$ , where  $a_1, a_2, \dots, a_r \in A$ , then the **reversal** of  $w$  is the word

$$\tilde{w} = a_r \cdots a_2 a_1.$$

We say  $w$  is a **palindrome** if  $w = \tilde{w}$ .

An **infinite word** is a map from  $\mathbb{N}$  to  $A$ , typically written in bold or as a sequence such as  $\mathbf{w} = w(0)w(1)w(2)\cdots$  or  $\mathbf{w} = w_0 w_1 w_2 \cdots$  (we freely pass between the two notations  $w_n$  and  $w(n)$  in what follows). Any finite word  $m$  gives rise to a periodic infinite word denoted  $\mathbf{m}^\infty$ , namely

$$m^\infty = m m m \cdots.$$

A **factorization** of a finite word  $w$  over  $A$  is a sequence  $(w_1, w_2, \dots, w_r)$  of words over  $A$  such that the relation  $w = w_1 w_2 \cdots w_r$  holds in the monoid  $A^*$ . We sometimes write  $w = (w_1, w_2, \dots, w_r)$  to emphasize a particular factorization of  $w$ . Factorizations of infinite words are similarly defined (with  $w_r$  necessarily the only infinite word in the sequence). If  $w$  is a finite or infinite word over  $A$  and  $w = uv$  for some (possibly empty) words  $u$  and  $v$ , then  $u$  is called a **prefix** of  $w$  and  $v$  is

a **suffix** of  $w$ . Conversely, a **factor** of a finite or infinite word  $w$  is a finite word  $v$  such that  $w = uvu'$  for some words  $u, u'$ ; we say  $v$  is a **proper factor** if  $v \neq \epsilon$  and  $uu' \neq \epsilon$ . Given two words  $w, w' \in A^*$ , we say that  $w$  is a **conjugate** of  $w'$  if there exists  $u, v \in A^*$  such that  $w = uv$  and  $w' = vu$ .

Let  $w$  be a finite or infinite word over an alphabet  $A$  and write  $w = a_0a_1a_2\cdots$ , where  $a_0, a_1, a_2, \dots \in A$ . If  $v$  is a factor of  $w$ , then

$$v = a_i a_{i+1} \cdots a_j \quad \text{for some } 0 \leq i < j,$$

and  $a_i a_{i+1} \cdots a_j$  is said to be an **occurrence** of  $v$  in  $w$ . (Specifically, an occurrence of  $v$  in  $w$  also includes information about where it appears in  $w$ ; for the factor above, we say the **starting index** is  $i$ .) If  $u$  and  $v$  are words, then  $u$  is said to **contain**  $v$  if there is an occurrence of  $v$  in  $u$ .

Given two alphabets  $A, B$ , a **morphism** from  $A^*$  to  $B^*$  shall always mean a “morphism of monoids.” That is, a set mapping  $f: A^* \rightarrow B^*$  satisfying

$$f(uv) = f(u)f(v) \quad \text{for all } u, v \in A^*.$$

In particular,  $f(\epsilon_{A^*}) = \epsilon_{B^*}$  since the empty word  $\epsilon$  is the only element in a free monoid satisfying  $w^2 = w$ . The **identity morphism** on  $A^*$  is the morphism sending each  $w \in A^*$  to itself. The **trivial morphism** from  $A^*$  to  $B^*$  is the morphism sending each  $w \in A^*$  to  $\epsilon_{B^*}$ .

## Bibliography

For the convenience of the reader, each entry below is followed by ↑ and a list of page numbers indicating the location within the book where the citation occurs.

- [ABG2007] A. Aberkane, S. Brlek, and A. Glen, *Sequences having the Thue–Morse complexity* (2007), preprint. ↑84
- [AB2005] B. Adamczewski and Y. Bugeaud, *On the decimal expansion of algebraic numbers*, Fiz. Mat. Fak. Moksl. Semin. Darb. **8** (2005), 5–13. ↑134
- [AL1977] A. Adler and S.-Y. R. Li, *Magic cubes and Prouhet sequences*, Amer. Math. Monthly **84** (1977), no. 8, 618–627. ↑74
- [AAB<sup>+</sup>1995] J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockusch, S. Plouffe, and B. E. Sagan, *A relative of the Thue–Morse sequence*, Discrete Math. **139** (1995), no. 1-3, 455–461. ↑71, 95
- [AARS1994] J.-P. Allouche, D. Astororian, J. Randall, and J. Shallit, *Morphisms, squarefree strings, and the Tower of Hanoi puzzle*, Amer. Math. Monthly **101** (1994), no. 7, 651–658. ↑90, 95, 99
- [AB1992] J.-P. Allouche and R. Bacher, *Toeplitz sequences, paperfolding, Towers of Hanoi and progression-free sequences of integers*, Enseign. Math. (2) **38** (1992), no. 3-4, 315–327. ↑95
- [AS1999] J.-P. Allouche and J. Shallit, *The ubiquitous Prouhet–Thue–Morse sequence*, Sequences and Their Applications (Singapore, 1998) (C. Ding, T. Helleseth, and H. Niederreiter, eds.), Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999, pp. 1–16. ↑69
- [AS2003] ———, *Automatic sequences: Theory, applications, generalizations*, Cambridge Univ. Press, Cambridge, 2003. ↑6, 71, 78, 79, 80, 81
- [AT2007] A. Alpers and R. Tijdeman, *The two-dimensional Prouhet–Tarry–Escott problem*, J. Number Theory **123** (2007), no. 2, 403–412. ↑73
- [Aus2006] P. Auster, *The Brooklyn follies*, Henry Holt and Co., New York, NY, 2006. ↑45
- [BEM1979] D. R. Bean, A. Ehrenfeucht, and G. F. McNulty, *Avoidable patterns in strings of symbols*, Pacific J. Math. **85** (1979), no. 2, 261–294. ↑102, 127
- [Ber1771] J. Bernoulli, *Sur une nouvelle espèce de calcul*, Recueil pour les astronomes, Vol. 1, Berlin, 1771, pp. 255–284. ↑3, 55
- [Ber1990] J. Berstel, *Tracé de droites, fractions continues et morphismes itérés*, Mots, Lang. Raison. Calc., Hermès, Paris, 1990, pp. 298–309. ↑5
- [Ber1995] ———, *Axel Thue’s papers on repetitions in words: A translation*, Publications du LaCIM, vol. 20, UQÀM, Montréal, QC, 1995. ↑81
- [BdL1997] J. Berstel and A. de Luca, *Sturmian words, Lyndon words and trees*, Theoret. Comput. Sci. **178** (1997), no. 1-2, 171–203. ↑20, 28, 41, 47
- [BP1985] J. Berstel and D. Perrin, *Theory of codes*, Pure Appl. Math., vol. 117, Academic Press, Orlando, FL, 1985. ↑105
- [BS2006] J. Berstel and A. Savelli, *Crochemore factorization of Sturmian and other infinite words*, Mathematical Foundations of Computer Science 2006 (Stará Lesná, 2006) (R. Královič and P. Urzyczyn, eds.), Lecture Notes in Comput. Sci., vol. 4162, Springer, Berlin, 2006, pp. 157–166. ↑116

- [BS1993] J. Berstel and P. Séébold, *A characterization of overlap-free morphisms*, Discrete Appl. Math. **46** (1993), no. 3, 275–281. ↑81
- [BdLR2008] V. Berthé, A. de Luca, and C. Reutenauer, *On an involution of Christoffel words and Sturmian morphisms*, European J. Combin. **29** (2008), no. 2, 535–553. ↑14, 15, 25
- [BBGL2007] A. Blondin-Massé, S. Brlek, A. Glen, and S. Labbé, *On the critical exponent of generalized Thue–Morse words*, Discrete Math. Theor. Comput. Sci. **9** (2007), no. 1, 293–304. ↑126
- [BL1993] J.-P. Borel and F. Laubie, *Quelques mots sur la droite projective réelle*, J. Théor. Nombres Bordeaux **5** (1993), no. 1, 23–51. ↑5, 17, 18, 20, 25, 47
- [BR2006] J.-P. Borel and C. Reutenauer, *On Christoffel classes*, Theor. Inform. Appl. **40** (2006), no. 1, 15–27. ↑27, 42, 44
- [BI1994] P. Borwein and C. Ingalls, *The Prouhet–Tarry–Escott problem revisited*, Enseign. Math. (2) **40** (1994), no. 1-2, 3–27. ↑73
- [BLP2003] P. Borwein, P. Lisoněk, and C. Percival, *Computational investigations of the Prouhet–Tarry–Escott problem*, Math. Comp. **72** (2003), no. 244, 2063–2070. ↑73
- [Brl1989] S. Brlek, *Enumeration of factors in the Thue–Morse word*, Discrete Appl. Math. **24** (1989), no. 1-3, 83–96. ↑81, 82
- [BW1994] M. Burrows and D. J. Wheeler, *A block-sorting lossless data compression algorithm*, Technical Report 124, Digital, Palo Alto, CA, 1994. ↑39
- [Car1921] F. Carlson, *Über Potenzreihen mit ganzzahligen Koeffizienten*, Math. Z. **9** (1921), no. 1-2, 1–13. ↑89
- [Car2006] A. Carpi, *On the repetition threshold for large alphabets*, Mathematical Foundations of Computer Science 2006 (Stará Lesná, 2006) (R. Královič and P. Urzyczyn, eds.), Lecture Notes in Comput. Sci., vol. 4162, Springer, Berlin, 2006, pp. 226–237. ↑126
- [Car2007] ———, *On Dejean’s conjecture over large alphabets*, Theoret. Comput. Sci. **385** (2007), no. 1-3, 137–151. ↑126
- [Cas1957] J. W. S. Cassels, *An introduction to Diophantine approximation*, Cambridge Tracts in Mathematics and Mathematical Physics, vol. 45, Cambridge Univ. Press, New York, NY, 1957. ↑56
- [CFL1958] K.-T. Chen, R. H. Fox, and R. C. Lyndon, *Free differential calculus. IV: The quotient groups of the lower central series*, Ann. of Math. (2) **68** (1958), 81–95. ↑41
- [CS1963] N. Chomsky and M.-P. Schützenberger, *The algebraic theory of context-free languages*, Computer Programming and Formal Systems, North-Holland, Amsterdam, 1963, pp. 118–161. ↑89
- [Chr1875] E. B. Christoffel, *Observatio arithmetica*, Ann. Mat. Pura Appl. (2) **6** (1875), 148–152. ↑5, 7, 44
- [Chr1887] ———, *Lehrsätze über arithmetische Eigenschaften der Irrationalzahlen*, Ann. Mat. Pura Appl. (2) **15** (1887–1888), 253–276. ↑5
- [CKMR1980] G. Christol, T. Kamae, M. Mendès France, and G. Rauzy, *Suites algébriques, automates et substitutions*, Bull. Soc. Math. France **108** (1980), no. 4, 401–419. ↑80
- [Cob1972] A. Cobham, *Uniform tag sequences*, Math. Systems Theory **6** (1972), 164–192. ↑78
- [Coh1972] H. Cohn, *Markoff forms and primitive words*, Math. Ann. **196** (1972), 8–22. ↑11
- [CS1966] M. Coudrain and M.-P. Schützenberger, *Une condition de finitude des monoïdes finiment engendrés*, C. R. Acad. Sci. Paris Sér. A-B **262** (1966), A1149–A1151. ↑133
- [CH1973] E. M. Coven and G. A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory **7** (1973), 138–153. ↑43, 85
- [Cro1981] M. Crochemore, *An optimal algorithm for computing the repetitions in a word*, Inform. Process. Lett. **12** (1981), no. 5, 244–250. ↑109
- [Cro1982] ———, *Sharp characterizations of squarefree morphisms*, Theoret. Comput. Sci. **18** (1982), no. 2, 221–226. ↑100, 101
- [Cro1983a] ———, *Recherche linéaire d’un carré dans un mot*, C. R. Acad. Sci. Paris Sér. I Math. **296** (1983), no. 18, 781–784. ↑114, 115
- [Cro1983b] ———, *Régularités évitables*, Thèse d’État, Université de Rouen, 1983. ↑100
- [Cro1986] ———, *Transducers and repetitions*, Theoret. Comput. Sci. **45** (1986), no. 1, 63–86. ↑122

- [CDP2005] M. Crochemore, J. Désarménien, and D. Perrin, *A note on the Burrows–Wheeler transformation*, Theoret. Comput. Sci. **332** (2005), no. 1-3, 567–572. ↑39, 40
- [CHL2001] M. Crochemore, C. Hancart, and T. Lecroq, *Algorithmique du texte*, Vuibert, Paris, 2001; English transl. in *Algorithms on strings*, Cambridge Univ. Press, Cambridge, 2007. ↑111
- [CHL2007] ———, *Algorithms on strings*, Cambridge Univ. Press, Cambridge, 2007. ↑111, 114, 116, 119, 123
- [CI2007] M. Crochemore and L. Ilie, *Analysis of maximal repetitions in strings*, Mathematical Foundations of Computer Science 2007 (Ceský Krumlov, 2007) (L. Kucera and A. Kucera, eds.), Lecture Notes in Comput. Sci., vol. 4708, Springer, Berlin, 2007, pp. 465–476. ↑125, 126
- [CR1995] M. Crochemore and W. Rytter, *Squares, cubes, and time-space efficient string searching*, Algorithmica **13** (1995), no. 5, 405–425. ↑107, 109
- [CR2002] ———, *Jewels of stringology*, World Sci. Publ., River Edge, NJ, 2002. ↑117, 124
- [Cur1993] J. D. Currie, *Unsolved problems: Open problems in pattern avoidance*, Amer. Math. Monthly **100** (1993), no. 8, 790–793. ↑131
- [CM2007] J. D. Currie and M. Mohammad-Noori, *Dejean’s conjecture and Sturmian words*, European J. Combin. **28** (2007), no. 3, 876–890. ↑126
- [CV2007] J. D. Currie and T. I. Visentin, *On abelian 2-avoidable binary patterns*, Acta Inform. **43** (2007), no. 8, 521–533. ↑130
- [CF1989] T. W. Cusick and M. E. Flahive, *The Markoff and Lagrange spectra*, Math. Surveys Monogr., vol. 30, Amer. Math. Soc., Providence, RI, 1989. ↑56, 64
- [Dav1992] H. Davenport, *The higher arithmetic: An introduction to the theory of numbers*, 6th ed., Cambridge Univ. Press, Cambridge, 1992. ↑47
- [Dej1972] F. Dejean, *Sur un théorème de Thue*, J. Combinatorial Theory Ser. A **13** (1972), 90–99. ↑126
- [Dek1979] F. M. Dekking, *Strongly nonrepetitive sequences and progression-free sets*, J. Combin. Theory Ser. A **27** (1979), no. 2, 181–185. ↑128, 130
- [Dic1930] L. E. Dickson, *Studies in the theory of numbers*, Univ. Chicago Press, Chicago, IL, 1930; reprinted as *Introduction to the theory of numbers*, Dover Publ., New York, NY, 1957. ↑55, 56
- [DB1937] H. L. Dorwart and O. E. Brown, *The Tarry–Escott problem*, Amer. Math. Monthly **44** (1937), no. 10, 613–626. ↑71
- [DG1990] S. Dulucq and D. Gouyou-Beauchamps, *Sur les facteurs des suites de Sturm*, Theoret. Comput. Sci. **71** (1990), no. 3, 381–400. ↑41
- [Duv1983] J.-P. Duval, *Factorizing words over an ordered alphabet*, J. Algorithms **4** (1983), no. 4, 363–381. ↑41
- [Erd1961] P. Erdős, *Some unsolved problems*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **6** (1961), 221–254. ↑127
- [Evd1968] A. A. Evdokimov, *Strongly asymmetric sequences generated by a finite number of symbols*, Dokl. Akad. Nauk SSSR **179** (1968), 1268–1271; English transl., Soviet Math. Dokl. **9** (1968), 536–539. ↑128
- [FM1997] S. Ferenczi and C. Mauduit, *Transcendence of numbers with a low complexity expansion*, J. Number Theory **67** (1997), no. 2, 146–161. ↑134
- [FS1998] A. S. Fraenkel and J. Simpson, *How many squares can a string contain?*, J. Combin. Theory Ser. A **82** (1998), no. 1, 112–120. ↑109, 110
- [Fro1913] F. G. Frobenius, *Über die Markoffschen Zahlen*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin (1913), 458–487; reprinted in J.-P. Serre (ed.), *Gesammelte Abhandlungen*, Vol. 3, Springer, Berlin, 1968. ↑56, 58
- [Fur1967] H. Furstenberg, *Algebraic functions over finite fields*, J. Algebra **7** (1967), 271–277. ↑80
- [Gau1986] C. F. Gauss, *Disquisitiones arithmeticae*, Springer, New York, NY, 1986. ↑55
- [GR1993] I. M. Gessel and C. Reutenauer, *Counting permutations with given cycle structure and descent set*, J. Combin. Theory Ser. A **64** (1993), no. 2, 189–215. ↑39
- [GLS2008] A. Glen, A. Lauve, and F. V. Saliola, *A note on the Markoff condition and central words*, Inform. Process. Lett. **105** (2008), no. 6, 241–244. ↑61, 64
- [GV1991] P. Goralčík and T. Vanicek, *Binary patterns in binary words*, Internat. J. Algebra Comput. **1** (1991), no. 3, 387–391. ↑105

- [GKP1994] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics: A foundation for computer science*, 2nd ed., Addison-Wesley, Reading, MA, 1994. ↑47, 53
- [Gus1997] D. Gusfield, *Algorithms on strings, trees, and sequences: Computer science and computational biology*, Cambridge Univ. Press, Cambridge, 1997. ↑117, 119, 122
- [Hal1964] M. Hall Jr., *Generators and relations in groups: The Burnside problem*, Lectures on Modern Mathematics, Vol. II, Wiley, New York, NY, 1964, pp. 42–92. ↑98, 99
- [HYY2003] H. K. Hsiao, Y. T. Yeh, and S. S. Yu, *Square-free-preserving and primitive-preserving homomorphisms*, Acta Math. Hungar. **101** (2003), no. 1-2, 113–130. ↑101
- [Ili2007] L. Ilie, *A note on the number of squares in a word*, Theoret. Comput. Sci. **380** (2007), no. 3, 373–376. ↑110
- [Jac1956] N. Jacobson, *Structure of rings*, Amer. Math. Soc., Colloq. Publ., vol. 37, Amer. Math. Soc., Providence, RI, 1956. ↑133
- [JZ2004] O. Jenkinson and L. Q. Zamboni, *Characterisations of balanced words via orderings*, Theoret. Comput. Sci. **310** (2004), no. 1-3, 247–271. ↑43
- [Jus2005] J. Justin, *Episturmian morphisms and a Galois theorem on continued fractions*, Theor. Inform. Appl. **39** (2005), no. 1, 207–215. ↑25, 29
- [KR2007] C. Kassel and C. Reutenauer, *Sturmian morphisms, the braid group  $B_4$ , Christoffel words and bases of  $F_2$* , Ann. Mat. Pura Appl. (4) **186** (2007), no. 2, 317–339. ↑34
- [Ker1986] V. Keränen, *On the  $k$ -freeness of morphisms on free monoids*, Ann. Acad. Sci. Fenn. Math. Diss. (1986), no. 61, 55. ↑105
- [Ker1987] ———, *On the  $k$ -freeness of morphisms on free monoids*, STACS 87 (Passau, 1987) (F.-J. Brandenburg, G. Vidal-Naquet, and M. Wirsing, eds.), Lecture Notes in Comput. Sci., vol. 247, Springer, Berlin, 1987, pp. 180–188. ↑105
- [Ker1992] ———, *Abelian squares are avoidable on 4 letters*, Automata, Languages and Programming (Vienna, 1992) (W. Kuich, ed.), Lecture Notes in Comput. Sci., vol. 623, Springer, Berlin, 1992, pp. 41–52. ↑128
- [KK1999a] R. Kolpakov and G. Kucherov, *Finding maximal repetitions in a word in linear time*, 40th Annual Symposium on Foundations of Computer Science (New York, NY, 1999), IEEE Computer Soc., Los Alamitos, CA, 1999, pp. 596–604. ↑125
- [KK1999b] ———, *On maximal repetitions in words*, Fundamentals of Computation Theory (Iași, 1999) (G. Ciobanu and G. Păun, eds.), Lecture Notes in Comput. Sci., vol. 1684, Springer, Berlin, 1999, pp. 374–385. ↑125
- [KZ1873] A. Korkine and G. Zolotareff, *Sur les formes quadratiques*, Math. Ann. **6** (1873), no. 3, 366–389. ↑55
- [LS1993] G. M. Landau and J. P. Schmidt, *An algorithm for approximate tandem repeats*, Combinatorial Pattern Matching (Padova, 1993) (A. Apostolico, M. Crochemore, Z. Galil, and U. Manber, eds.), Lecture Notes in Comput. Sci., vol. 684, Springer, Berlin, 1993, pp. 120–133. ↑123
- [LV1986] G. M. Landau and U. Vishkin, *Introducing efficient parallelism into approximate string matching and a new serial algorithm*, Proc. 18th Annual ACM Symposium on Theory of Computing (Berkeley, CA, 1986), ACM Press, New York, NY, 1986, pp. 220–230. ↑123
- [Lec1985] M. Leconte, *A characterization of power-free morphisms*, Theoret. Comput. Sci. **38** (1985), no. 1, 117–122. ↑105
- [Lem1970] A. Lempel, *On a homomorphism of the de Bruijn graph and its applications to the design of feedback shift registers*, IEEE Trans. Computers **C-19** (1970), 1204–1209. ↑78
- [LZ1976] A. Lempel and J. Ziv, *On the complexity of finite sequences*, IEEE Trans. Information Theory **IT-22** (1976), no. 1, 75–81. ↑114
- [Lot1997] M. Lothaire, *Combinatorics on words*, Cambridge Math. Lib., Cambridge Univ. Press, Cambridge, 1997. ↑41, 99, 125
- [Lot2002] ———, *Algebraic combinatorics on words*, Encyclopedia Math. Appl., vol. 90, Cambridge Univ. Press, Cambridge, 2002. ↑6, 9, 14, 41, 43, 125, 130, 131, 133
- [Lot2005] ———, *Applied combinatorics on words*, Encyclopedia Math. Appl., vol. 105, Cambridge Univ. Press, Cambridge, 2005. ↑110
- [dL1997] A. de Luca, *Sturmian words: Structure, combinatorics, and their arithmetics*, Theoret. Comput. Sci. **183** (1997), no. 1, 45–82. ↑24, 25, 27

- [dLM1994] A. de Luca and F. Mignosi, *Some combinatorial properties of Sturmian words*, Theoret. Comput. Sci. **136** (1994), no. 2, 361–385. ↑11, 31, 41
- [dLV1989] A. de Luca and S. Varricchio, *Some combinatorial properties of the Thue–Morse sequence and a problem in semigroups*, Theoret. Comput. Sci. **63** (1989), no. 3, 333–348. ↑82
- [dLV1999] ———, *Finiteness and regularity in semigroups and formal languages*, Monogr. Theoret. Comput. Sci. EATCS Ser., Springer, Berlin, 1999. ↑133
- [Luc1893] É. Lucas, *Récréations mathématiques*, Vol. 3, Gauthier-Villars, Paris, 1893; reprinted by A. Blanchard, Paris, 1979. ↑91
- [LS2001] R. C. Lyndon and P. E. Schupp, *Combinatorial group theory*, Classics Math., Springer, Berlin, 2001. ↑35
- [MKS2004] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial group theory: Presentations of groups in terms of generators and relations*, 2nd ed., Dover Publ., Mineola, NY, 2004. ↑35
- [MRS2003] S. Mantaci, A. Restivo, and M. Sciortino, *Burrows–Wheeler transform and Sturmian words*, Inform. Process. Lett. **86** (2003), no. 5, 241–246. ↑40
- [Man2001] G. Manzini, *An analysis of the Burrows–Wheeler transform*, J. ACM **48** (2001), no. 3, 407–430. ↑39
- [Mar1879] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Math. Ann. **15** (1879), no. 3–4, 381–406. ↑5, 55, 56, 57
- [Mar1880] ———, *Sur les formes quadratiques binaires indéfinies*. 2nd mémoire, Math. Ann. **17** (1880), no. 3, 379–399. ↑5, 55, 56, 57
- [Mar1881] ———, *Sur une question de Jean Bernoulli*, Math. Ann. **19** (1881), no. 1, 27–36. ↑5, 55
- [McC1976] E. M. McCreight, *A space-economical suffix tree construction algorithm*, J. Assoc. Comput. Mach. **23** (1976), no. 2, 262–272. ↑119
- [Mel1999] G. Melançon, *Visualisation de graphes et combinatoire des mots*, thèse d’habilitation à diriger les recherches, Université Bordeaux I, 1999. ↑42
- [Men2003] F. Mendivil, *Fractals, graphs, and fields*, Amer. Math. Monthly **110** (2003), no. 6, 503–515. ↑78
- [Mig1991] F. Mignosi, *On the number of factors of Sturmian words*, Theoret. Comput. Sci. **82** (1991), no. 1, 71–84. ↑134
- [MP1992] F. Mignosi and G. Pirillo, *Repetitions in the Fibonacci infinite word*, RAIRO Inform. Théor. Appl. **26** (1992), no. 3, 199–204. ↑134
- [MS1993] F. Mignosi and P. Séébold, *Morphismes sturmiens et règles de Rauzy*, J. Théor. Nombres Bordeaux **5** (1993), no. 2, 221–233. ↑14
- [Mor2004] E. Moreno, *On the theorem of Fredricksen and Maiorana about de Bruijn sequences*, Adv. in Appl. Math. **33** (2004), no. 2, 413–415. ↑78
- [MH1940] M. Morse and G. A. Hedlund, *Symbolic dynamics. II: Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42. ↑43, 44
- [MH1944] ———, *Unending chess, symbolic dynamics and a problem in semigroups*, Duke Math. J. **11** (1944), 1–7. ↑97, 98, 99
- [Mou1992] J. Moulin-Ollagnier, *Proof of Dejean’s conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters*, Theoret. Comput. Sci. **95** (1992), no. 2, 187–205. ↑126
- [Mye1986] E. W. Myers, *An O(ND) difference algorithm and its variations*, Algorithmica **1** (1986), no. 1, 251–266. ↑123
- [NBT1963] H. Noland, C. H. Braunholtz, and T. Tamura, *Advanced problems and solutions: Solutions: 5030*, Amer. Math. Monthly **70** (1963), no. 6, 675–676. ↑97
- [OZ1981] R. P. Osborne and H. Zieschang, *Primitives in the free group on two generators*, Invent. Math. **63** (1981), no. 1, 17–24. ↑5, 34
- [Pan1984] J.-J. Pansiot, *À propos d’une conjecture de F. Dejean sur les répétitions dans les mots*, Discrete Appl. Math. **7** (1984), no. 3, 297–311. ↑126
- [Pir1999] G. Pirillo, *A new characteristic property of the palindrome prefixes of a standard Sturmian word*, Sémin. Lothar. Combin. **43** (1999), B43f. ↑39
- [Ple1970] P. A. B. Pleasants, *Non-repetitive sequences*, Proc. Cambridge Philos. Soc. **68** (1970), 267–274. ↑128
- [Pro1851] E. Prouhet, *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sci. Paris Sér. I Math. **33** (1851), 225. ↑71, 72

- [Pyt2002] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics* (V. Berthé, S. Ferenczi, C. Mauduit, and A. Siegel, eds.), Lecture Notes in Math., vol. 1794, Springer, Berlin, 2002. ↑6
- [Ram2007] N. Rampersad, *On the context-freeness of the set of words containing overlaps*, Inform. Process. Lett. **102** (2007), no. 2-3, 74–78. ↑89, 90
- [Ran1973] G. N. Raney, *On continued fractions and finite automata*, Math. Ann. **206** (1973), 265–283. ↑28
- [Rau1984] G. Rauzy, *Des mots en arithmétique*, Journées d’Avignon. Théorie des langages et complexité des algorithmes (Avignon, 1983) (B. Rozoy, ed.), Publ. Dép. Math. Nouvelle Sér. B, vol. 84-6, Univ. Claude-Bernard, Lyon, 1984, pp. 103–113. ↑20
- [Reu2006a] C. Reutenauer, *Mots de Lyndon généralisés*, Sém. Lothar. Combin. **54** (2006), B54h. ↑56
- [Reu2006b] ———, *On Markoff’s property and Sturmian words*, Math. Ann. **336** (2006), no. 1, 1–12. ↑56, 60, 61, 62, 64
- [Ric2003] G. Richomme, *Some non-finitely generated monoids of repetition-free endomorphisms*, Inform. Process. Lett. **85** (2003), no. 2, 61–66. ↑100
- [RW2002] G. Richomme and F. Wlazinski, *Some results on k-power-free morphisms*, Theoret. Comput. Sci. **273** (2002), no. 1-2, 119–142. ↑100
- [RW2004] ———, *Overlap-free morphisms and finite test-sets*, Discrete Appl. Math. **143** (2004), no. 1-3, 92–109. ↑101
- [RW2007] ———, *Existence of finite test-sets for k-power-freeness of uniform morphisms*, Discrete Appl. Math. **155** (2007), no. 15, 2001–2016. ↑101
- [Rob2005] W. Robertson, *Square cells: An array cooking lesson*, The PracTEX Journal (2005), no. 2, available at <http://tug.org/pracjourn/2005-2/robertson/>. ↑x
- [Ryt2006] W. Rytter, *The number of runs in a string: Improved analysis of the linear upper bound*, STACS 2006 (Marseille, 2006) (B. Durand and W. Thomas, eds.), Lecture Notes in Comput. Sci., vol. 3884, Springer, Berlin, 2006, pp. 184–195. ↑125
- [Ryt2007] ———, *The number of runs in a string*, Inform. and Comput. **205** (2007), no. 9, 1459–1469. ↑125
- [Sch1961] M.-P. Schützenberger, *On a special class of recurrent events*, Ann. Math. Statist. **32** (1961), 1201–1213. ↑133
- [Séé1982] P. Séébold, *Morphismes itérés, mot de Morse et mot de Fibonacci*, C. R. Acad. Sci. Paris Sér. I Math. **295** (1982), no. 6, 439–441. ↑81
- [Ser1985] C. Series, *The geometry of Markoff numbers*, Math. Intelligencer **7** (1985), no. 3, 20–29. ↑45
- [Sim2004] J. Simpson, *Disjoint Beatty sequences*, Integers **4** (2004), A12. ↑8
- [Šir1957] A. I. Širšov, *On some non-associative null-rings and algebraic algebras*, Mat. Sb. (N.S.) **41(83)** (1957), 381–394 (Russian). ↑133
- [Šir1962] ———, *On the bases of free Lie algebras*, Algebra i Logika Sem. **1** (1962), no. 1, 14–19 (Russian). ↑42
- [Smi1876] H. J. S. Smith, *Note on continued fractions*, Messenger of Mathematics **6** (1876), 1–14. ↑5, 44, 48, 49, 116
- [Sta1997] R. P. Stanley, *Enumerative combinatorics*, Vol. 1, Cambridge Stud. Adv. Math., vol. 49, Cambridge Univ. Press, Cambridge, 1997. ↑117
- [Sta1999] ———, *Enumerative combinatorics*, Vol. 2, Cambridge Stud. Adv. Math., vol. 62, Cambridge Univ. Press, Cambridge, 1999. ↑79
- [SG2002] J. Stoye and D. Gusfield, *Simple and flexible detection of contiguous repeats using a suffix tree*, Theoret. Comput. Sci. **270** (2002), no. 1-2, 843–856. ↑123
- [Thu1906] A. Thue, *Über unendliche Zeichenreihen*, Christiania Vidensk. Selsk. Skr. (1906), no. 7, 1–22; reprinted in T. Nagell, S. Selberg, and K. Thalberg (eds.), *Selected mathematical papers*, Universitetsforlaget, Oslo, 1977. ↑67, 97, 127
- [Thu1912] ———, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Christiania Vidensk. Selsk. Skr. (1912), no. 1, 1–67; reprinted in T. Nagell, S. Selberg, and K. Thalberg (eds.), *Selected mathematical papers*, Universitetsforlaget, Oslo, 1977. ↑67, 81
- [Thu1977] ———, *Selected mathematical papers* (T. Nagell, S. Selberg, and K. Thalberg, eds.), Universitetsforlaget, Oslo, 1977. ↑67

- [TS1995] J. Tromp and J. Shallit, *Subword complexity of a generalized Thue–Morse word*, Inform. Process. Lett. **54** (1995), no. 6, 313–316. ↑82
- [Ukk1995] E. Ukkonen, *On-line construction of suffix trees*, Algorithmica **14** (1995), no. 3, 249–260. See also <http://www.allisons.org/ll/AlgDS/Tree/Suffix/>, where it is really online. ↑119
- [Vie1978] X. G. Viennot, *Algèbres de Lie libres et monoïdes libres*, Lecture Notes in Math., vol. 691, Springer, Berlin, 1978. ↑42
- [Wei1973] P. Weiner, *Linear pattern matching algorithms*, 14th Annual IEEE Symposium on Switching and Automata Theory (Iowa City, IA, 1973), IEEE Comput. Soc., Northridge, CA, 1973, pp. 1–11. ↑119
- [Wel1984] T. A. Welch, *A technique for high-performance data compression*, Computer **17** (1984), 8–19. ↑114
- [WW1994] Z. X. Wen and Z. Y. Wen, *Local isomorphisms of invertible substitutions*, C. R. Acad. Sci. Paris Sér. I Math. **318** (1994), no. 4, 299–304. ↑14, 37
- [Wri1959] E. M. Wright, *Prouhet’s 1851 solution of the Tarry–Escott problem of 1910*, Amer. Math. Monthly **66** (1959), 199–201. ↑71, 72
- [Zim1982] A. I. Zimin, *Blocking sets of terms*, Mat. Sb. (N.S.) **119(161)** (1982), no. 3, 363–375 (Russian). ↑127, 131, 132



# Index

Key words and phrases from the text appear in lexicographic order, as usual. Mathematical symbols appear in the order in which they occurred within the text. Mixtures of symbols and text, such as *k-avoidable pattern*, are likely to be found among the key words (omitting the symbols), e.g., under *avoidable pattern*.

- $\mathbb{N}$ , xi
- $|w|_a$ , xi
- $m^\infty$ , xi
- $a \perp b$ , 5
- $C(a, b)$ , 6
- $xxyxxyxxyxy$ , 6, 17, 20, 24, 25, 30, 44, 48, 49
- $\mathbf{G}$ , 9, 20, 25, 62
- $\mathbf{D}$ , 9, 25
- $\tilde{\mathbf{G}}$ , 9
- $\tilde{\mathbf{D}}$ , 9, 20, 62
- $\mathbf{E}$ , 9
- $\mathcal{G}_j$ , 9, 21
- $\mathcal{D}$ , 10
- $\mathrm{SL}_2(\mathbb{Z})$ , 19
- $w^+$ , 24
- Pal, 24
- $\mathbb{N}^{2 \times 2}$ , 31
- $\mathbb{P}$ , 56
- $\lambda_t(A)$ , 56
- $\mathbb{R}_{>0}$ , 56
- $M(A)$ , 56
- $\mathbf{t}$ , 69
- $\bar{s}$ , 70
- $\varphi^\infty$ , 70
- $\mathrm{bin}(n)$ , 77
- $\mathbb{F}_q$ , 80
- $c_{\mathbf{t}}(n)$ , 82
- $\mathrm{Han}(n, i, j)$ , 92, 95
- $C(h)$ , 102
- $O(n)$ , 109
- $w(i, j)$ , 112
- $w_{(i)}$ , 112
- $w^{(i)}$ , 112
- $\wedge$ , 112, 117
- $\wedge$ , 112
- $\pi_w(u)$ , 115, 117
- $\mathrm{Suff}(w)$ , 117
- $\mathrm{RT}(n)$ , 126
- $\doteq$ , 131
- Abelian *k*-avoidable, 127
- Abelian instance, 127
- accepted by an automaton, 86
- algebraic series, 79, 89
- alphabet, xi
- alternating lexicographic order, 58, 61
- anagram, 127
- aperiodic sequence, 43
- k*-automatic sequence, 78
- automatic sequences, 94
- automaton, 77, *see also* finite deterministic, 86
- pushdown, 88
- k*-avoidable pattern, 127
- k*-avoidable set, 131
- Bézout's Lemma, 20, 35, 53
- balanced<sub>1</sub>, 41, 44, 64
- balanced<sub>2</sub>, 42
- basis, 33, *see also* free group
- bi-ideal sequence, 133
- bifix code, 102
- big-*O*, 109
- binary quadratic form, 55
- discriminant of a, 55
- equivalent, 55
- minimum of a, 55
- reduced, 55
- binary word, 69
- bissective code, 105
- block, 58, 71
- Burrows–Wheeler transform, 39
- Cayley graph, 7, 30, 39
- centered square, 111
- left-, 111
- right-, 111
- Chomsky hierarchy, 85

- Christoffel morphism, 9, 22, 37  
**G**, **D**,  $\tilde{\mathbf{G}}$ ,  $\tilde{\mathbf{D}}$  and **E**, 9  
 Christoffel path, 5, 49  
   closest point for a, 17  
   lower, 5  
   slope, 5  
   upper, 5  
 Christoffel tree, 35, 47, 50, 53  
 Christoffel word, 6  
   lower, 6, 8  
   nontrivial, 6, 23  
   of slope 4/7, 6, 17, 20, 24, 25, 30, 44, 48,  
     49  
   slope, 6  
   standard factorization, 17, 25, 35  
   trivial, 6  
   upper, 6, 23  
 circular word, 42, 79  
 closest point, 17  
 code, 101  
   bifix, 102  
   bissective, 105  
   comma-free, 102  
   faithful, 105  
   infix, 102  
   left synchronizing, 105  
   prefix, 102  
   prefix-suffix, 105  
   ps-code, 105  
   right synchronizing, 105  
   strongly synchronizing, 105  
   suffix, 102  
   synchronizing, 105  
   uniform, 102  
 comma-free code, 102  
 compact suffix tree, 118  
 complexity function, 82  
 conjugate  
   group-theoretic, 13, 33  
   monoid-theoretic, xii, 13, 33  
   root, 7, 51, 63  
 contain, xii  
 context-free grammar, 87  
   derivation, 87  
   leftmost derivation, 89  
 context-free language, 85, 88  
   pumping lemma for a, 90  
   unambiguous, 89  
 continuant, 48  
 continued fraction, 47, 58  
   continuant, 48  
   representation, 47  
 cover relation, 117  
 covers, 117  
 critical exponent, 126, 134  
 Crochemore factorization, 114  
 cutting sequence, 7  
 degree, 71, *see also* Tarry–Escott  
 derivation, 87  
 derived, 87  
 de Bruijn word, 78, 79  
 diagonal, 80, *see also* generating series  
 discretization, 5, 134  
 discriminant, 55  
*n*-division, 133  
 empty word, xi  
 erasing morphism, 14  
 exponent, 126  
   critical, 126, 134  
 extendible repetition, 125  
 extreme point, 49  
 factor, xii  
   left special, 82, 84  
   proper, xii  
   right special, 82  
 Factor Problem, 117  
 factorization, xi  
   Crochemore, 114  
   *n*-division, 133  
   left Lyndon, 42  
   right Lyndon, 42  
   standard, 17, 25  
 faithful code, 105  
 Fibonacci, 117  
   number, 51, 79, 109, 111  
   word, 7, 57, 99, 100, 109, 116, 134  
   words, 109, 116  
 final states, 77, *see also* finite deterministic automaton  
 Fine–Wilf Theorem, 32, 40, 110  
 finite deterministic automaton, 77, 79, 86,  
   87, 91, 94, 100, 119  
   final states, 77  
   initial state, 77  
   next state function, 77  
   states, 77  
 first occurrence, 115, 117  
    $\pi_w(u)$ , 115, 117  
 formal language theory, 85  
 forumlae of Justin, 25–27  
 fractal rendering, 78  
 fractional power, 126  
 free group, 33  
   basis, 33  
   inner automorphism, 34  
   positive element, 33  
   primitive element, 33  
 generalized Thue–Morse word, 72  
 generating series, 79  
   algebraic, 79  
   diagonal, 80  
   Fibonacci numbers, 79  
   rational, 79

- transcendental, 79
- golden ratio, 7, 51, 109
- grammar, 87, *see also* context-free
- Hanoi word, 92
  - and Thue–Morse word, 94
  - automatic sequence, 94
- ideal solution, *see also* Tarry–Escott
- identity morphism, xii
- implicit suffix tree, 119
- index
  - recurrence, 84
  - starting, 115, 117
- infix code, 102
- initial state, 77, *see also* finite deterministic automaton
- inner automorphism, 34
- instance
  - Abelian, 127
  - of a pattern, 127
- iterate of an endomorphism, 70
- iterated palindromic closure, 24
- König’s Lemma, 130
- label, 7, 17
- language, 85, 119
  - context-free, 85, 88
  - regular, 85, 86
- leaf node, 117
- left factorization, 42
- left special factor, 82, 84
- left synchronizing code, 105
- left-centered square, 111
- leftmost derivation, 89
- length
  - in the free group, 33
  - in the free monoid, xi
- letter-doubling morphism, 84
- letters, xi
- Levi’s Lemma, 110
- lexicographic order, 39
  - alternating, 58, 61
- longest common prefix, 112
  - $\wedge$ , 112
- longest common suffix, 112
  - $\wedge$ , 112
- lower Christoffel word, 8
- Lyndon word, 39, 41, 78, 133
  - left factorization, 42
  - right factorization, 42
- magic square, 74
  - La Sagrada Família*, 74
  - Melencolia I*, 74
  - from the Thue–Morse word, 74
  - of order  $2^m$ , 74
  - order, 74
- Markoff numbers, 57
- Markoff spectrum, 64
- maximal repetition, 125
- mediant, 52
- minimal period, 125
- $k$ -mismatch, 123
- morphism, xii
  - Christoffel, 9, 37
  - erasing, 14
  - exchange, 70
  - fixed point of a, 70
  - Hall, 98, 100
  - identity, xii
  - iterate, 70
  - letter-doubling, 84
  - nonerasing, 14, 101
  - period-doubling, 93
  - positive, 37
  - square-free, 100
  - $k$ -square-free, 100
  - Thue–Morse, 70
  - trivial, xii, 100
  - uniform, 101
  - $k$ -uniform, 70, 78
- morphism of Hall, 98, 100
- next state function, 77, *see also* finite deterministic automaton
- nonerasing morphism, 14
- nonerasingmorphism, 101
- nonextendible repetition, 125
- nontrivial Christoffel word, 23
- occurrence, 125
  - extendible repetition, 125
  - nonextendible repetition, 125
  - number of, xi, 26, 107
  - of a factor, xii
  - starting index, xii, 82, 115
- Open Question
  - Thue–Morse generating series, 80
  - context-free language of non-factors of  $t$ , 89
  - Markoff numbers, 58
- order, 74, *see also* magic square
- overlap, 81, 110
- overlap-free word, 81
- palindrome, xi, 13, 23, 61
- palindromic closure, 24
- palindromic prefix, 24, 27
- palindromic suffix, 24
- PATH, 118
- pattern, 127
  - Abelian  $k$ -avoidable, 127
  - Abelian instance of a, 127
  - $k$ -avoidable, 127
  - $k$ -avoidable set, 131
  - instance of a, 127

- k*-unavoidable, 131
- k*-unavoidable set, 131
- period, 25, 27
  - minimal, 125
- period-doubling morphism, 93
- periodic phenomena, 44
- periodic sequences, 44
- Pick's Theorem, 19, 20
- poset
  - cover relation, 117
  - covers, 117
- POSITION, 118
- positive element, 33
- positive morphism, 37
- k*th-power-free word, 97
- prefix, xi
  - longest common, 112
  - palindromic, 24, 27
  - proper, xii
- prefix array, 112
- prefix code, 102
- prefix poset, 107, 117
- prefix-suffix code, 105
- primitive element, 33
- primitive word, 6, 33, 107
- productions, 87
- proper factor, xii
- ps-code, 105
- pumping lemma
  - for context-free languages, 90
  - for regular languages, 86
- pumping length
  - for regular languages, 86
- pushdown automata, 88
- quadratic form
  - binary, 55
  - equivalent, 55
  - minumum of a, 55
  - reduced, 55
- quadratic number, 63
- rational series, 79, 89
- recurrence index, 84
- recurrent, 84
  - uniformly recurrent, 84
- recurrent word, 133
- reduced word, 33
- regular language, 85, 86
  - pumping lemma for a, 86
- rejected by an automaton, 86
- relatively prime, 5
- repetition, 125
  - extendible, 125
  - maximal, 125
  - nonextendible, 125
- repetition threshold, 126
- reversal, xi, 13, 23
- right factorization, 42
- right special factor, 82
- right synchronizing code, 105
- right-centered square, 111
- root node, 117
- run, 125
- sequence
  - aperiodic, 43
  - bi-ideal, 133
  - cutting, 7
  - periodic, 44
  - ultimately periodic, 43, 87, 90
- sesquipowers, 133
- shift registers, 78
- size, 71, *see also* Tarry–Escott
- skew-words, 44
- square
  - centered, 111
  - left-centered, 111
  - right-centered, 111
- square-free morphism, 100
- k*-square-free morphism, 100
- square-free word, 97
- standard factorization, 17, 22, 35, 53
- starting index, xii, 82, 115, 117
  - $\pi_w(u)$ , 115, 117
- starting position, *see also* starting index
- states, 77, *see also* finite deterministic automaton
- Stern–Brocot tree, 47, 52
- strongly synchronizing code, 105
- Sturmian word, 6, 41, 43, 44, 126
  - characteristic, 6, 134
- suffix, xii
  - longest common, 112
  - palindromic, 24
  - proper, xii
  - $\text{Suff}(w)$ , 117
- SUFFIX, 118
- suffix array, 112
- suffix code, 102
- suffix link, 122
- suffix node, 117
- suffix tree, 117
  - $\mathcal{T}(w)$ , 117
  - compact, 118
  - implicit, 119
  - leaf node, 117
  - PATH, 118
  - POSITION, 118
  - root node, 117
  - SUFFIX, 118
  - suffix node, 117
  - unadorned, 117
- synchronizing code, 105
  - left, 105
  - right, 105
  - strongly, 105

- Tarry–Escott problem, 71  
 Thue–Morse word, 72  
 degree of a, 71  
 ideal solution, 73  
 Prouhet’s solution, 72  
 size of a, 71  
 terminal letters, 87  
 test set for square-freeness, 101  
 Three Squares Lemma, 107  
 Thue–Morse morphism, 70, 78  
 Thue–Morse word, 69, 77, 79, 81, 85, 86, 88, 94  
     and Hanoi word, 94  
     automaton, 77  
     complexity function, 82  
     generalized, 72  
     generating series, 80  
     Toeplitz sequence, 95  
     Toeplitz word, 95  
     Tower of Hanoi, 90  
     transcendental series, 79, 89  
     trivial morphism, xii, 100  
     ultimately periodic, 43, 87, 90  
     unadorned suffix tree, 117  
     unambiguous context-free language, 89  
      $k$ -unavoidable pattern, 131  
      $k$ -unavoidable set, 131  
     uniform code, 102  
     uniform morphism, 101  
      $k$ -uniform morphism, 70, 78  
     uniformly recurrent, 84  
     upper Christoffel word, 23  
     variables, 87  
     word, xi  
         accepted, 86  
         anagram of a, 127  
         as a fixed point of a morphism, 70  
         balanced<sub>1</sub>, 41  
         balanced<sub>2</sub> Lyndon, 42  
         binary, 69  
         characteristic Sturmian, 6, 134  
         Christoffel, 6  
         circular, 42, 79  
         conjugate, 6  
         contains, xii  
         critical exponent, 126, 134  
         derived, 87  
         de Bruijn, 78  
         exponent, 126  
         factor of, xii  
         Fibonacci, 100, 116, 134  
         fractional power, 126  
         Hanoi, 92  
         infinite, xi  
         Lyndon, 39, 41, 78, 133  
         minimal period, 125  
         occurrence, xii, 125  
         overlap, 81, 110  
         overlap-free, 81  
         pattern, 127  
          $k$ th-power-free, 97  
         primitive, 6, 33, 107  
         recurrence index, 84  
         recurrent, 84, 133  
         reduced, 33  
         rejected, 86  
         repetition, 125  
         reversal, 23  
         square-free, 97  
         Sturmian, 43, 126  
         Thue–Morse, 69  
         Toeplitz, 95  
         ultimately periodic, 87  
         uniformly recurrent, 84  
         Zimin, 131  
     Zimin word, 131



The two parts of this text are based on two series of lectures delivered by Jean Berstel and Christophe Reutenauer in March 2007 at the Centre de Recherches Mathématiques, Montréal, Canada. Part I represents the first modern and comprehensive exposition of the theory of Christoffel words. Part II presents numerous combinatorial and algorithmic aspects of repetition-free words stemming from the work of Axel Thue—a pioneer in the theory of combinatorics on words.

A beginner to the theory of combinatorics on words will be motivated by the numerous examples, and the large variety of exercises, which make the book unique at this level of exposition. The clean and streamlined exposition and the extensive bibliography will also be appreciated. After reading this book, beginners should be ready to read modern research papers in this rapidly growing field and contribute their own research to its development.

Experienced readers will be interested in the finitary approach to Sturmian words that Christoffel words offer, as well as the novel geometric and algebraic approach chosen for their exposition. They will also appreciate the historical presentation of the Thue–Morse word and its applications, and the novel results on Abelian repetition-free words.

ISBN 978-0-8218-4480-9



9 780821 844809

**CRMM/27**



For additional information  
and updates on this book, visit  
[www.ams.org/bookpages/crmm-27](http://www.ams.org/bookpages/crmm-27)

**AMS on the Web**  
[www.ams.org](http://www.ams.org)