

Contents of Volume 1

Chapter 1. Formal Languages: an Introduction and a Synopsis

<i>Alexandru Mateescu and Arto Salomaa</i>	1
1. Languages, formal and natural	1
1.1 Historical linguistics	3
1.2 Language and evolution	7
1.3 Language and neural structures	9
2. Glimpses of mathematical language theory	9
2.1 Words and languages	10
2.2 About commuting	12
2.3 About stars	15
2.4 Avoiding scattered subwords	19
2.5 About scattered residuals	24
3. Formal languages: a telegraphic survey	27
3.1 Language and grammar. Chomsky hierarchy	27
3.2 Regular and context-free languages	31
3.3 L Systems	33
3.4 More powerful grammars and grammar systems	35
3.5 Books on formal languages	36
References	38

Chapter 2. Regular Languages

<i>Sheng Yu</i>	41
1. Preliminaries	43
2. Finite automata	45
2.1 Deterministic finite automata	45
2.2 Nondeterministic finite automata	49
2.3 Alternating finite automata	54
2.4 Finite automata with output	66
3. Regular expressions	70
3.1 Regular expressions – the definition	70
3.2 Regular expressions to finite automata	71
3.3 Finite automata to regular expressions	74
3.4 Star height and extended regular expressions	77
3.5 Regular expressions for regular languages of polynomial density	80
4. Properties of regular languages	82
4.1 Four pumping lemmas	82
4.2 Closure properties	88
4.3 Derivatives and the Myhill-Nerode theorem	92

5. Complexity issues 96
 5.1 State complexity issues 96
 5.2 Time and space complexity issues 102
 References 105

Chapter 3. Context-Free Languages and Pushdown Automata

Jean-Michel Autebert, Jean Berstel, and Luc Boasson 111
 1. Introduction 111
 1.1 Grammars 112
 1.2 Examples 113
 2. Systems of equations 114
 2.1 Systems 115
 2.2 Resolution 120
 2.3 Linear systems 121
 2.4 Parikh’s theorem 123
 3. Normal forms 124
 3.1 Chomsky normal form 124
 3.2 Greibach normal forms 125
 3.3 Operator normal form 133
 4. Applications of the Greibach normal form 134
 4.1 Shamir’s theorem 134
 4.2 Chomsky-Schützenberger’s theorem 135
 4.3 The hardest context-free language 136
 4.4 Wechler’s theorem 138
 5. Pushdown machines 138
 5.1 Pushdown automata 138
 5.2 Deterministic pda 144
 5.3 Pushdown store languages 151
 5.4 Pushdown transducers 154
 6. Subfamilies 156
 6.1 Linear languages 156
 6.2 Quasi-rational languages 160
 6.3 Strong quasi-rational languages 163
 6.4 Finite-turn languages 164
 6.5 Counter languages 165
 6.6 Parenthetic languages 167
 6.7 Simple languages 169
 6.8 LL and LR languages 170
 References 172

Chapter 4. Aspects of Classical Language Theory

Alexandru Mateescu and Arto Salomaa 175
 1. Phrase-structure grammars 175
 1.1 Phrase-structure grammars and Turing machines 176
 1.2 Normal forms for phrase-structure grammars 180

1.3	Representations of recursively enumerable languages	180
1.4	Decidability. Recursive languages	184
2.	Context-sensitive grammars	186
2.1	Context-sensitive and monotonous grammars	186
2.2	Normal forms for context-sensitive grammars	190
2.3	Workspace	191
2.4	Linear bounded automata	192
2.5	Closure properties of the family CS	194
2.6	Decidable properties of the family CS	196
2.7	On some restrictions on grammars	197
3.	AFL-theory	199
3.1	Language families. Cones and AFL's	199
3.2	First transductions, then regular operations	204
4.	Wijngaarden (two-level) grammars	210
4.1	Definitions, examples. The generative power	210
4.2	Membership problem and parsing	214
4.3	W-grammars and complexity	218
5.	Attribute grammars	221
5.1	Definitions and terminology	221
5.2	Algorithms for testing the circularity	222
5.3	Restricted attribute grammars	226
5.4	Other results	229
6.	Patterns	230
6.1	Erasing and nonerasing patterns	230
6.2	The equivalence problem	232
6.3	The inclusion problem	234
6.4	The membership problem	238
6.5	Ambiguity	238
6.6	Multi-pattern languages. Term rewriting with patterns	241
7.	Pure and indexed grammars. Derivation languages	242
7.1	Pure grammars	242
7.2	Indexed grammars	243
7.3	The derivation (Szilard) language	245
	References	246

Chapter 5. L Systems

	<i>Lila Kari, Grzegorz Rozenberg, and Arto Salomaa</i>	253
1.	Introduction	253
1.1	Parallel rewriting	253
1.2	Callithamnion roseum, a primordial alga	254
1.3	Life, real and artificial	256
2.	The world of L, an overview	257
2.1	Iterated morphisms and finite substitutions: D0L and 0L	257
2.2	Auxiliary letters and other auxiliary modifications	264
2.3	Tables, interactions, adults, fragmentation	268

3. Sample L techniques: avoiding cell death if possible	275
4. L decisions	281
4.1 General. Sequences versus languages	281
4.2 D0L sequence equivalence problem and variations	285
5. L Growth	288
5.1 Communication and commutativity	288
5.2 Merging and stages of death	294
5.3 Stagnation and malignancy	298
6. L codes, number systems, immigration	301
6.1 Morphisms applied in the way of a fugue	301
6.2 An excursion into number systems	305
6.3 Bounded delay and immigration	308
7. Parallel insertions and deletions	312
8. Scattered views from the L path	322
References	324

Chapter 6. Combinatorics of Words

<i>Christian Choffrut and Juhani Karhumäki</i>	329
1. Introduction	329
2. Preliminaries	331
2.1 Words	332
2.2 Periods in words	334
2.3 Repetitions in words	337
2.4 Morphisms	338
2.5 Finite sets of words	339
3. Selected examples of problems	342
3.1 Injective mappings between F -semigroups	342
3.2 Binary equality sets	347
3.3 Separating words via automata	351
4. Defect effect	354
4.1 Basic definitions	354
4.2 Defect theorems	357
4.3 Defect effect of several relations	360
4.4 Relations without the defect effect	364
4.5 The defect theorem for equations	366
4.6 Properties of the combinatorial rank	367
5. Equations as properties of words	370
5.1 Makanin’s result	370
5.2 The rank of an equation	371
5.3 The existential theory of concatenation	372
5.4 Some rules of thumb for solving equations by “hand”	374
6. Periodicity	376
6.1 Definitions and basic observations	376
6.2 The periodicity theorem of Fine and Wilf	377

6.3	The Critical Factorization Theorem	380
6.4	A characterization of ultimately periodic words	385
7.	Finiteness conditions	389
7.1	Orders and quasi-orderings	390
7.2	Orderings on words	391
7.3	Subwords of a given word	394
7.4	Partial orderings and an unavailability	395
7.5	Basics on the relation quasi-ordering \preceq_r	398
7.6	A compactness property	400
7.7	The size of an equivalent subsystem	404
7.8	A finiteness condition for finite sets of words	405
8.	Avoidability	408
8.1	Prelude	408
8.2	The basic techniques	409
8.3	Repetition-free morphisms	415
8.4	The number of repetition-free words	416
8.5	Characterizations of binary 2^+ -free words	418
8.6	Avoidable patterns	422
9.	Subword complexity	423
9.1	Examples and basic properties	423
9.2	A classification of complexities of fixed points of iterated morphisms	427
	References	431

Chapter 7. **Morphisms**

	<i>Tero Harju and Juhani Karhumäki</i>	439
1.	Introduction	439
2.	Preliminaries	441
2.1	Words and morphisms	441
2.2	Rational transductions	442
2.3	Word problem for finitely presented semigroups	443
2.4	Semi-Thue systems	445
3.	Post correspondence problem: decidable cases	446
3.1	Basic decidable cases	446
3.2	Generalized post correspondence problem	449
3.3	(G)PCP in the binary case	451
4.	Undecidability of PCP with applications	456
4.1	PCP(9) is undecidable	456
4.2	A mixed modification of PCP	460
4.3	Common relations in submonoids	461
4.4	Mortality of matrix monoids	463
4.5	Zeros in upper corners	466
5.	Equality sets	468
5.1	Basic properties	468
5.2	Some restricted cases	470

5.3	On the regularity of equality sets	471
5.4	An effective construction of regular equality sets	474
6.	Ehrenfeucht's Conjecture and systems of equations	477
6.1	Systems of equations	478
6.2	Ehrenfeucht's Conjecture	478
6.3	Equations with constants	483
6.4	On generalizations of Ehrenfeucht's Conjecture	484
6.5	Ehrenfeucht's Conjecture for more general mappings	486
7.	Effective subcases	487
7.1	Finite systems of equations	487
7.2	The effectiveness of test sets	488
7.3	Applications to problems of iterated morphisms	490
8.	Morphic representations of languages	492
8.1	Classical results	492
8.2	Representations of recursively enumerable languages	493
8.3	Representations of regular languages	496
8.4	Representations of rational transductions	498
9.	Equivalence problem on languages	500
9.1	Morphic equivalence on languages	501
9.2	More general mappings	502
10.	Problems	504
	References	505

Chapter 8. **Codes**

	<i>Helmut Jürgensen and Stavros Konstantinidis</i>	511
1.	Introduction	511
2.	Notation and basic notions	516
3.	Channels and codes	522
4.	Error correction, synchronization, decoding	530
5.	Homophonic codes	541
6.	Methods for defining codes	545
7.	A hierarchy of classes of codes	558
8.	The syntactic monoid of a code	567
9.	Deciding properties of codes	575
10.	Maximal codes	582
11.	Solid codes	585
12.	Codes for noisy channels	595
13.	Concluding remarks	600
	References	600

Chapter 9. **Semirings and Formal Power Series**

	<i>Werner Kuich</i>	609
1.	Introduction	609
2.	Semirings, formal power series and matrices	610
3.	Algebraic systems	619

4. Automata and linear systems	626
5. Normal forms for algebraic systems	631
6. Pushdown automata and algebraic systems	637
7. Transductions and abstract families of elements	644
8. The Theorem of Parikh	658
9. Lindenmayerian algebraic power series	662
10. Selected topics and bibliographical remarks	667
References	671

Chapter 10. Syntactic Semigroups

<i>Jean-Eric Pin</i>	679
1. Introduction	679
2. Definitions	681
2.1 Relations	681
2.2 Semigroups	681
2.3 Morphisms	682
2.4 Groups	683
2.5 Free semigroups	684
2.6 Order ideals	684
2.7 Idempotents	685
2.8 Green’s relations	685
2.9 Categories	686
3. Recognizability	687
3.1 Recognition by ordered semigroups	687
3.2 Syntactic order	688
3.3 Recognizable sets	689
3.4 How to compute the syntactic semigroup?	691
4. Varieties	692
4.1 Identities	692
4.2 The variety theorem	695
5. Examples of varieties	697
5.1 Standard examples	698
5.2 Commutative varieties	699
5.3 Varieties defined by local properties	701
5.4 Algorithmic problems	705
6. Some algebraic tools	705
6.1 Relational morphisms	706
6.2 Mal’cev product	708
6.3 Semidirect product	708
6.4 Representable transductions	712
7. The concatenation product	716
7.1 Polynomial closure	716
7.2 Unambiguous and deterministic polynomial closure	717
7.3 Varieties closed under product	718
7.4 The operations $L \rightarrow LaA^*$ and $L \rightarrow A^*aL$	719

- 7.5 Product with counters 720
- 7.6 Varieties closed under product with counter 721
- 8. Concatenation hierarchies 722
 - 8.1 Straubing-Thérien’s hierarchy 723
 - 8.2 Dot-depth hierarchy 727
 - 8.3 The group hierarchy 728
 - 8.4 Subhierarchies 730
 - 8.5 Boolean-polynomial closure 731
- 9. Codes and varieties 733
- 10. Operators on languages and varieties 735
- 11. Further topics 736
- References 738

Chapter 11. Regularity and Finiteness Conditions

- Aldo de Luca and Stefano Varricchio* 747
- 1. Combinatorics on words 749
 - 1.1 Infinite words and unavoidable regularities 749
 - 1.2 The Ramsey theorem 751
 - 1.3 The van der Waerden theorem 753
 - 1.4 The Shirshov theorem 754
 - 1.5 Uniformly recurrent words 756
 - 1.6 Some extensions of the Shirshov theorem 759
- 2. Finiteness conditions for semigroups 764
 - 2.1 The Burnside problem 764
 - 2.2 Permutation property 767
 - 2.3 Chain conditions and J -depth decomposition 771
 - 2.4 Iteration property 775
 - 2.5 Repetitive morphisms and semigroups 779
- 3. Finitely recognizable semigroups 782
 - 3.1 The factor semigroup 783
 - 3.2 On a conjecture of Brzozowski 785
 - 3.3 Problems and results 786
- 4. Non-uniform regularity conditions 789
 - 4.1 Pumping properties 790
 - 4.2 Permutative property 792
- 5. Well quasi-orders 795
 - 5.1 The generalized Myhill-Nerode Theorem 796
 - 5.2 Quasi-orders and rewriting systems 798
 - 5.3 A regularity condition for permutable languages 800
 - 5.4 A regularity condition for almost-commutative languages 803
- References 806

Chapter 12. Families Generated by Grammars and L Systems

<i>Gheorghe Păun and Arto Salomaa</i>	811
1. Grammar forms	811
1.1 Introduction: structural similarity	811
1.2 Interpretations and grammatical families	813
1.3 Closure properties and normal forms	818
1.4 Completeness	821
1.5 Linguistical families and undecidability	825
2. L forms	829
2.1 EOL and ETOL forms	829
2.2 Good and very complete forms	834
2.3 PDOL forms	835
3. Density	838
3.1 Dense hierarchies of grammatical families	838
3.2 Color families of graphs and language interpretations	843
3.3 Maximal dense intervals	847
4. Extensions and variations	851
4.1 Context-dependent grammar and L forms	851
4.2 Matrix forms	853
4.3 Further variants	856
References	859
Index	863

Authors' Addresses

Jean-Michel Autebert
UFR d'Informatique, Université Denis Diderot-Paris 7
2, place Jussieu, F-75005 Paris, Cedex 05, France
autebert@litp.ibp.fr

Jean Berstel
LITP, IBP, Université Pierre et Marie Curie
Laboratoire LITP, Institut Blaise Pascal
4, place Jussieu, F-75252 Paris, Cedex 05, France
jean.berstel@litp.ibp.fr

Luc Boasson
LITP, IBP, Université Denis Diderot
2, place Jussieu, F-75251 Paris Cedex 05, France
luc.boasson@litp.ibp.fr

Christian Choffrut
LITP, Université de Paris VI
4, place Jussieu, F-75252 Paris Cedex 05, France
cc@litp.ibp.fr

Tero Harju
Department of Mathematics, University of Turku
FIN-20014 Turku, Finland
harju@sara.utu.fi

Helmut Jürgensen
Department of Computer Science, The University of Western Ontario
London, Ontario N6A 5B7, Canada
helmut@uwo.ca
and Institut für Informatik, Universität Potsdam
Am Neuen Palais 10, D-14469 Potsdam, Germany
helmut@mpag-inf.uni-potsdam.de

Juhani Karhumäki
Department of Mathematics, University of Turku
FIN-20014 Turku, Finland
karhumak@utu.fi

Lila Kari
Department of Computer Science, University of Western Ontario
London, Ontario N6A 5B7, Canada
lila@csd.uwo.ca

Stavros Konstantinidis
Department of Mathematics and Computer Science, University of Lethbridge
4401 University Drive, Lethbridge, Alberta T1K 3M4 Canada
stavros@csd.uleth.ca

Werner Kuich
Abteilung für Theoretische Informatik, Institut für Algebra
und Diskrete Mathematik, Technische Universität Wien
Wiedner Hauptstrasse 8-10, A-1040 Wien, Austria
kuich@email.tuwien.ac.at

Aldo de Luca
Dipartimento di Matematica, Università di Roma "La Sapienza"
Piazzale Aldo Moro 2, I-00185 Roma, Italy
deluca@mercurio.mat.uniroma1.it

Alexandru Mateescu
Faculty of Mathematics, University of Bucharest
Academiei 14, RO-70109 Bucharest, Romania
and Turku Centre for Computer Science (TUCS)
Lemminkäisenkatu 14 A, FIN-20520 Turku, Finland
mateescu@sara.utu.fi

Gheorghe Păun
Institute of Mathematics of the Romanian Academy
P.O. Box 1-764, RO-70700 Bucharest, Romania
gpaun@imar.ro

Jean-Eric Pin
LITP, IBP, CNRS, Université Paris VI
4, place Jussieu, F-75252 Paris Cedex 05, France
jep@litp.ibp.fr

Grzegorz Rozenberg
Department of Computer Science, Leiden University
P.O. Box 9512, NL-2300 RA Leiden, The Netherlands
and Department of Computer Science
University of Colorado at Boulder, Campus 430
Boulder, CO 80309, U.S.A.
rozenber@wi.leidenuniv.nl

Arto Salomaa
Academy of Finland and Turku Centre for Computer Science (TUCS)
Lemminkäisenkatu 14 A, FIN-20520 Turku, Finland
asalomaa@sara.cc.utu.fi

Stefano Varricchio
Dipartimento di Matematica, Università di L'Aquila
Via Vetoio Loc. Coppito, I-67100 L'Aquila, Italy
varricch@univaq.it

Sheng Yu
Department of Computer Science, University of Western Ontario
London, Ontario N6A 5B7, Canada
syu@csd.uwo.ca