References

Adamczewski, B. (2003). Balances for fixed points of primitive substitutions, *Theoret. Comput. Sci.*, *307*, 47–75.

Adams, W. W. (1969). Simultaneous asymptotic diophantine approximations to a basis of a real cubic number field, *J. Number Theory*, *1*, 179–194.

Adebiyi, E. F., Jiang, T., and Kaufmann, M. (2001). An efficient algorithm for finding short approximate non-tandem repeats, In *Intelligent Systems for Molecular Biology (ISMB)*, pp. 5–12.

Adebiyi, E. F. and Kaufmann, M. (2002). Extracting common motifs under the Levenshtein measure: Theory and experimentation, In Guigó, R. and Gusfield, D. (Eds.), *Algorithms in Bioinformatics (WABI 2002)*, Vol. 2452 of *Lect. Notes Comp. Sci.*, pp. 140–156. Springer-Verlag.

Aho, A. V., Hopcroft, J. E., and Ullman, J. D. (1975). *The Design and Analysis of Computer Algorithms*. Addison-Wesley.

Aho, A. V., Hopcroft, J. E., and Ullman, J. D. (1983). *Data Structures and Algorithms*. Addison-Wesley.

Aho, A. V., Sethi, R., and Ullman, J. D. (1986). *Compilers: Principles, Techniques and Tools*. Addison-Wesley.

Akiyama, S. (1999). Self affine tiling and Pisot numeration system, In *Number Theory and its Applications (Kyoto, 1997)*, Vol. 2 of *Dev. Math.*, pp. 7–17. Kluwer.

Akiyama, S. and Pethö, A. (2002). On canonical number systems, *Theoret. Comput. Sci.*, *70*, 921–933.

Allauzen, C. and Mohri, M. (2003). Efficient algorithms for testing the twins property, *J. Autom. Lang. Comb.*, *8*(2), 117–144.

Allauzen, C., Mohri, M., and Riley, M. (2004). Statistical modeling for unit selection in speech synthesis, In 42*nd Meeting Assoc. Computational Linguistics (ACL 2004)*, Barcelona, Spain.

Allouche, J.-P. (1990). Sur la transcendance de la série formelle π, *Sém. Théor. Nombres Bordeaux*, *2*, 103–117.

Allouche, J.-P. (1997). Note on the transcendence of a generating function, In Laurincikas, A., Manstavicius, E., and Stakenas, V. (Eds.), *Proc. Palanga Conference*, Vol. 4 of *New Trends in Probability and Statistics*, pp. 461–465. VSP Science.

Allouche, J.-P. and Shallit, J. (1999). The ubiquitous Prouhet-Thue-Morse sequence, In Ding, C., Helleseth, T., and Niederreiter, H. (Eds.), *Sequences and Their Applications, Proceedings of SETA '98*, pp. 1–16. Springer-Verlag.

Allouche, J.-P. and Shallit, J. (2003). *Automatic Sequences*. Cambridge University Press.

Allouche, J.-P. and Zamboni, L. Q. (1998). Algebraic irrational binary numbers cannot be fixed points of non-trivial constant-length or primitive morphisms, *J. Number Theory*, *69*, 119–124.

Alon, N. and Spencer, J. (1992). *The Probabilistic Method.* J. Wiley and Sons.

Alonso, L., Rémy, J. L., and Schott, R. (1997). A linear-time algorithm for the generation of trees, *Algorithmica*, *17*(2), 162–182.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool, *J. Mol. Biol.*, *215*, 403–410.

Ambjørn, J., Durhuus, B., and Jonsson, T. (1997). *Quantum Geometry.* Cambridge Monographs on Mathematical Physics. Cambridge University Press.

Angluin, D. (1982). Inference of reversible languages, *J. Assoc. Comput. Mach.*, *29*(3), 741–765.

Angluin, D. (1987). Learning regular sets from queries and counterexamples, *Inform. Comput.*, *75*(2), 87–106.

Apostolico, A. (1985). The myriad virtues of suffix trees, In Apostolico, A. and Galil, Z. (Eds.), *Combinatorial Algorithms on Words*, Vol. 12 of *NATO Advanced Science Institutes, Series F*, pp. 85–96. Springer-Verlag.

Apostolico, A., Bock, M., and Xuyan, X. (1998). Annotated statistical indices for sequence analysis, In *Compression and Complexity of Sequences 97*, pp. 215–229. IEEE Computer Society Press.

Apostolico, A. and Crochemore, M. (1991). Optimal canonization of all substrings of a string, *Inform. Comput.*, *95*(1), 76–95.

Apostolico, A. and Preparata, F. (1983). Optimal off-line detection of repetitions in a string, *Theoret. Comput. Sci.*, *22*(3), 297–315.

Apostolico, A. and Szpankowski, W. (1992). Self-alignments in words and their applications, *J. Algorithms*, *13*, 446–467.

Arnoux, P. (1988). Un exemple de semi-conjugaison entre un échange d'intervalles et une translation sur le tore, *Bull. Soc. Math. France*, *116*(4), 489–500.

Arnoux, P. (1989). Some remarks about Fibonacci multiplication, *Appl. Math. Lett.*, *2*(4), 319–320.

Arnoux, P. and Ito, S. (2001). Pisot substitutions and Rauzy fractals, *Bull. Belg. Math. Soc. Simon Stevin*, *8*(2), 181–207.

Arnoux, P. and Rauzy, G. (1991). Représentation géométrique de suites de complexité $2n+1$, *Bull. Soc. Math. France*, *119*(2), 199–215.

Arnoux, P. and Yoccoz, J.-C. (1981). Construction de difféomorphismes pseudo-Anosov, *C. R. Acad. Sci. Paris Sér. I Math.*, *292*(1), 75–78.

Arratia, R., Bollobas, B., Coppersmith, D., and Sorkin, G. (2000). Euler circuits and DNA sequencing by hybridization, *Discr. Appl. Math.*, *104*, 63–69.

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations : the Chen-Stein method, *Ann. Probab.*, *17*, 9–25.

Arratia, R., Goldstein, L., and Gordon, L. (1990). Poisson approximation and the Chen-Stein method, *Statist. Sci.*, *5*, 403–434.

Arratia, R., Martin, D., Reinert, G., and Waterman, M. S. (1996). Poisson approximation for long repeats in a random sequence with application to sequencing by hybridization, *J. Comput. Biol.*, *3*, 425–463.

Arratia, R., Morris, P., and Waterman, M. S. (1988). Stochastic Scrabble: large deviations for sequences with scores, *J. Appl. Probab.*, *25*, 106–119.

Arratia, R. and Waterman, M. (1994). A phase transition for the score in matching random sequences allowing deletions, *Ann. Appl. Probab.*, *4*, 200–225.

Arratia, R. and Waterman, M. S. (1989). The Erdös-Rényi strong law for pattern matching with a given proportion of mismatches, *Ann. Probab.*, *17*, 1152–1169.

Ash, R. B. (1990). *Information Theory.* Dover. Corrected reprint of the 1965 original edition.

Atallah, M., Jacquet, P., and Szpankowski, W. (1993). A probabilistic analysis of a pattern matching problem, *Random Structures & Algorithms*, *4*, 191–213.

Baake, M. and Moody, R. V. (Eds.) (2000). *Directions in Mathematical Quasicrystals*, Vol. 13 of *CRM Monograph Series.* Amer. Math. Soc.

Baeza-Yates, R. and Ribero-Neto, B. (1999). *Modern Information Retrieval.* Addison-Wesley.

Bahl, L. R., Jelinek, F., and Mercer, R. (1983). A maximum likelihood approach to continuous speech recognition, *IEEE Trans. Pattern Anal. Machine Intell.*, *5*(2), 179–190.

Baldi, P. and Brunak, S. (1998). *Bioinformatics. The Machine Learning Approach.* The MIT Press.

Banderier, C. and Flajolet, P. (2002). Basic analytic combinatorics of directed lattice paths, *Theoret. Comput. Sci.*, *281*(1-2), 37–80.

Barbour, A. D., Chen, L., and Loh, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method, *Ann. Probab.*, *20*, 1843–1866.

Barbour, A. D. and Chryssaphinou, O. (2001). Compound Poisson approximation: a user's guide, *Ann. in Appl. Probab.*, *11*, 964–1002.

Barbour, A. D., Chryssaphinou, O., and Vaggelatou, E. (2001). Applications of compound Poisson approximation, In Charalambides, C. A., Koutras, M. V., and Balakrishnan, N. (Eds.), *Probability and Statistical Models with Applications*, pp. 41–62. Chapman and Hall.

Barbour, A. D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.

Barbour, A. D. and Mansson, M. (2002). Compound Poisson process approximation, *Ann. Probab.*, *30*, 1492–1537.

Barbour, A. D. and Utev, S. (1998). Solving the Stein equation in compound Poisson approximation, *Adv. in Appl. Probab.*, *30*, 449–475.

Barbour, A. D. and Xia, A. (1999). Poisson perturbations, *ESAIM Probab. Statist.*, *3*, 131–150.

Barcucci, E., Pergola, E., Pinzani, R., and Rinaldi, S. (2001). A bijection for some paths on the slit plane, *Adv. in Appl. Math.*, *26*(2), 89–96.

Barcucci, E., Pinzani, R., and Sprugnoli, R. (1995). The random generation of underdiagonal walks, *Discrete Math.*, *139*(1-3), 3–18.

Béal, M.-P. and Carton, O. (2001). Computing the prefix of an automaton, *Theoret. Inform. Appl.*, *34*(6), 503–515.

Béal, M.-P. and Carton, O. (2002). Determinization of transducers over finite and infinite words, *Theoret. Comput. Sci.*, *289*(1), 225–251.

Bedford, T. (1986). Generating special Markov partitions for hyperbolic toral automorphisms using fractals, *Ergodic Theory Dynam. Systems*, *6*(3), 325–333.

Bell, T. C., Cleary, J. G., and Witten, I. H. (1990). *Text Compression.* Prentice Hall.

Bender, E. (1973). Central and local limit theorems applied to asymptotic enumeration, *J. Combin. Th. A*, *15*, 91–111.

Bender, E. and Kochman, F. (1993). The distribution of subword counts is usually normal, *European J. Combin.*, *14*, 265–275.

Bergeron, F., Labelle, G., and Leroux, P. (1998). *Combinatorial Species and Tree-like Structures*, Vol. 67 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press.

Berstel, J. (1979). *Transductions and Context-Free Languages.* Teubner.

Berstel, J. and Reutenauer, C. (1984). *Rational Series and Their Languages.* Springer-Verlag.

Bertrand, A. (1977). Développements en base de Pisot et répartition modulo 1, *C. R. Acad. Sci. Paris, Sér. A*, *285*, 419–421.

Bétréma, J. and Penaud, J.-G. (1993). Modèles avec particules dures, animaux dirigés et séries en variables partiellement commutatives, available at `arXiv:math.CO/0106210`.

Biaudet, V., El Karoui, M., and Gruss, A. (1998). Codon usage can explain GT-rich islands surrounding Chi sites on the *Escherichia coli* genome, *Mol. Microbiol.*, *29*, 666–669.

Bieganski, P., Riedl, J., Carlis, J. V., and Retzel, E. (1994). Generalized suffix trees for biological sequence data: applications and implementations, In *27th Hawai Int. Conf. on Systems Science*, pp. 35–44. IEEE Computer Society Press.

Biggins, J. and Cannings, C. (1987). Markov renewal processes, counters and repeated sequences in Markov chains, *Ann. in Appl. Probab.*, *19*, 521–545.

Billingsley, P. (1961). Statistical methods in Markov chains, *Ann. Math. Statist.*, *2*, 12–40.

Blanchette, M. (2001). Algorithms for phylogenetic footprinting, In *5th Research in Computational Molecular Biology (RECOMB)*, pp. 49–58. ACM Press.

Blanchette, M., Schwikowski, B., and Tompa, M. (2000). An exact algorithm to identify motifs in orthologous sequences from multiple species, In *Intelligent Systems for Molecular Biology (ISMB)*, pp. 37–45.

Blanchette, M., Schwikowski, B., and Tompa, M. (2002). Algorithms for phylogenetic footprinting, *J. Comput. Biol.*, *9*(2), 211–223.

Blanchette, M. and Tompa, M. (2002). Discovery of regulatory elements by a computational method for phylogenetic footprinting, *Genome Research*, *12*, 739–748.

Blom, G. and Thorburn, D. (1982). How many random digits are required until given sequences are obtained?, *J. Appl. Probab.*, *19*, 518–531.

Blumer, A., Blumer, J., Ehrenfeucht, A., Haussler, D., and McConnel, R. (1983). Linear size finite automata for the set of all subwords of a word: an outline of results, *Bull. Eur. Assoc. Theoret. Comput. Sci.*, *21*, 12–20.

Blumer, A., Blumer, J., Haussler, D., Ehrenfeucht, A., Chen, M. T., and Seiferas, J. (1985). The smallest automaton recognizing the subwords of a text, *Theoret. Comput. Sci.*, *40*(1), 31–55.

Blumer, A., Ehrenfeucht, A., and Haussler, D. (1989). Average size of suffix trees and DAWGS, *Discr. Appl. Math.*, *24*, 37–45.

Booth, K. S. (1980). Lexicographically least circular substrings, *Inform. Proc. Letters*, *10*(4-5), 240–242.

Bourdon, J. and Vallée, B. (2002). Generalized pattern matching statistics, In *Mathematics and Computer Science II (Versailles, 2002)*, Trends. Math., pp. 249–265. Birkhäuser.

Bousquet-Mélou, M. (1996). *Habilitation à diriger des recherches.* Université Bordeaux I, available at `http://www.labri.fr/~bousquet/`.

Bousquet-Mélou, M. (2001). Walks on the slit plane: other approaches, *Adv. in Appl. Math.*, *27*(2-3), 243–288.

Bousquet-Mélou, M. (2002). Counting walks in the quarter plane, In *Mathematics and Computer Science II (Versailles, 2002)*, Trends Math., pp. 49–67. Birkhäuser.

Bousquet-Mélou, M. and Guttmann, A. J. (1997). Three-dimensional self-avoiding convex polygons, *Phys. Rev. E (3)*, *55*(6, part A), R6323–R6326.

Bousquet-Mélou, M. and Schaeffer, G. (2002). Walks on the slit plane, *Probab. Theory Related Fields*, *124*(3), 305–344.

Bouttier, J., Di Francesco, P., and Guitter, E. (2002). Census of planar maps: from the one-matrix model solution to a combinatorial proof, *Nuclear Phys. B*, *645*(3), 477–499.

Brazma, A., Jonassen, I., Eidhammer, I., and Gilbert, D. (1998a). Approaches to the automatic discovery of patterns in biosequences, *J. Comput. Biol.*, *5*(2), 277–304.

Brazma, A., Jonassen, I., Vilo, J., and Ukkonen, E. (1998b). Pattern discovery in biosequences, In Honavar, V. and Slutzki, G. (Eds.), *Grammatical Inference (ICGI-98)*, Vol. 1433 of *Lect. Notes Comp. Sci.*, pp. 255–270. Springer-Verlag.

Brazma, A., Jonassen, I., Vilo, J., and Ukkonen, E. (1998c). Predicting gene regulatory elements *in silico* on a genomic scale, *Genome Research*, *8*(11), 1202–1215.

Breen, S., Waterman, M. S., and Zhang, N. (1985). Renewal theory for several patterns, *J. Appl. Probab.*, *22*, 228–234.

Brendel, V., Beckmann, J. S., and Trifonov, E. N. (1986). Linguistics of nucleotide sequences: Morphology and comparison of vocabularies, *J. Biomol. Struct. Dynamics*, *4*, 11–21.

Breslauer, D. (1998). The suffix tree of a tree and minimizing sequential transducers, *Theoret. Comput. Sci.*, *191*(1-2), 131–144.

Brodal, G., Lyngsø, R., Pedersen, C., and Stoye, J. (2000). Finding maximal pairs with bounded gap, *J. Discrete Algorithms*, *1*(1), 77–104.

Brown, T. A. (1999). *Genomes.* BIOS Scientific Publishers.

Bucklew, J. A. (1990). *Large Deviation Techniques in Decision, Simulation, and Estimation.* J. Wiley and Sons.

Buhler, J. and Tompa, M. (2001). Finding motifs using random projections, In *5th Research in Computational Molecular Biology (RECOMB)*, pp. 69–76. ACM Press.

Burdík, Č., Frougny, C., Gazeau, J. P., and Krejcar, R. (1998). Beta-integers as natural counting systems for quasicrystals, *J. Phys. A*, *31*(30), 6449–6472.

Burdík, Č., Frougny, C., Gazeau, J.-P., and Krejcar, R. (2000). Beta-integers as a group, In Gambaudo, J.-M., Hubert, P., Tisseur, P., and Vaienti, S. (Eds.), *Dynamical Systems (Luminy-Marseille, 1998)*, pp. 125–136. World Scientific.

Canterini, V. and Siegel, A. (2001a). Automate des préfixes-suffixes associé à une substitution primitive, *J. Théor. Nombres Bordeaux*, *13*(2), 353–369.

Canterini, V. and Siegel, A. (2001b). Geometric representation of substitutions of Pisot type, *Trans. Amer. Math. Soc.*, *353*(12), 5121–5144.

Cartier, P. and Foata, D. (1969). *Problèmes combinatoires de commutation et réarrangements*, Vol. 85 of *Lect. Notes Math.* Springer-Verlag.

Cassaigne, J., Ferenczi, S., and Zamboni, L. Q. (2000). Imbalances in Arnoux-Rauzy sequences, *Ann. Inst. Fourier (Grenoble)*, *50*(4), 1265–1276.

Cassels, J. W. S. (1957). *An Introduction to Diophantine Approximation.* Cambridge University Press.

Castelli, M. G., Mignosi, F., and Restivo, A. (1999). Fine and Wilf's theorem for three periods and a genereralization of Sturmian words, *Theoret. Comput. Sci.*, *218*, 83–94.

Chedin, F., Noirot, P., Biaudet, V., and Ehrlich, S. (1998). A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of *Bacillus subtilis*, *Mol. Microbiol.*, *31*, 1369–1377.

Chekhova, N. (2000). Covering numbers of rotations, *Theoret. Comput. Sci.*, *230*(1-2), 97–116.

Chekhova, N., Hubert, P., and Messaoudi, A. (2001). Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci, *J. Théor. Nombres Bordeaux*, *13*(2), 371–394.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials, *Ann. Probab.*, *3*, 534–545.

Chen, L. H. Y. and Xia, A. (to appear). Stein's method, Palm theory and Poisson process approximation, *Ann. Probab.*

Choffrut, C. (1977). Une caractérisation des fonctions séquentielles et des fonctions sous-séquentielles en tant que relations rationnelles, *Theoret. Comput. Sci.*, *5*(3), 325–337.

Choffrut, C. (1979). A generalization of Ginsburg and Rose's characterization of G-S-M mappings, In Maurer, H. A. (Ed.), *6th Automata, Languages and Programming (ICALP)*, Vol. 71 of *Lect. Notes Comp. Sci.*, pp. 88–103. Springer-Verlag.

Choffrut, C. (2003). Minimizing subsequential transducers: a survey, *Theoret. Comput. Sci.*, *292*(1), 131–143.

Choffrut, C. and Karhumäki, J. (1997). Combinatorics of words, In Rozenberg, G. and Salomaa, A. (Eds.), *Handbook of Formal Languages*, Vol. I, pp. 329–438. Springer-Verlag.

Chomsky, N. (1956). Three models for the description of language, *IRE Trans. Inform. Theory*, *2*(3), 113–124.

Chomsky, N. (1957). *Syntactic Structures*. Mouton, The Hague.

Chomsky, N. and Schützenberger, M. P. (1963). The algebraic theory of context-free languages, In Braffort, P. and Hirschberg, D. (Eds.), *Computer Programming and Formal Systems*, pp. 118–161. North Holland, Amsterdam.

Chottin, L. and Cori, R. (1982). Une preuve combinatoire de la rationalité d'une série génératrice associée aux arbres, *RAIRO Inform. théor.*, *16*(2), 113–128.

Christol, G. (1979). Ensembles presque périodiques k-reconnaissables, *Theoret. Comput. Sci.*, *9*, 141–145.

Christol, G., Kamae, T., Mendès France, M., and Rauzy, G. (1980). Suites algébriques, automates et substitutions, *Bull. Soc. Math. France*, *108*, 401–419.

Chryssaphinou, O. and Papastavridis, S. (1988a). A limit theorem for the number of non-overlapping occurrences of a pattern in a sequence of independent trials, *J. Appl. Probab.*, *25*, 428–431.

Chryssaphinou, O. and Papastavridis, S. (1988b). A limit theorem on the number of overlapping appearances of a pattern in a sequence of independent trials, *Probab. Theory Related Fields*, *79*, 129–143.

Chryssaphinou, O., Papastavridis, S., and Vaggelatou, E. (2001). Poisson approximation for the non-overlapping appearances of several words in Markov chains, *Combin. Probab. Comput.*, *10*, 293–308.

Chung, K. L. (1974). *A Course in Probability Theory*. Academic Press. 2nd edition.

Churchill, G. A. (1989). Stochastic models for heterogeneous DNA sequences, *Bull. Math. Biol.*, *51*, 79–94.

Churchill, G. A., Daniels, D. L., and Waterman, M. S. (1990). The distribution of restriction enzyme sites in *Escheria coli*, *Nucl. Acids Res.*, *18*, 589–597.

Clement, J., Flajolet, P., and Vallée, B. (2001). Dynamic sources in information theory: a general analysis of trie structures, *Algorithmica*, *29*, 307–369.

Cobham, A. (1972). Uniform tag sequences, *Math. Systems Theory*, *6*, 164–192.

Cole, R. and Hariharan, R. (1998). Approximate String Matching: A Simpler Faster Algorithm, In *9th SIAM Symposium on Discrete Algorithms (SODA)*, pp. 463–472.

Cori, R. (1975). *Un code pour les graphes planaires et ses applications*. Société Mathématique de France. Astérisque, No. 27.

Cori, R. and Machì, A. (1992). Maps, hypermaps and their automorphisms: a survey. I, II, III, *Exposition. Math.*, *10*(5), 403–427, 429–447, 449–467.

Cori, R. and Vauquelin, B. (1981). Planar maps are well labeled trees, *Canad. J. Math.*, *33*(5), 1023–1042.

Cowan, R. (1991). Expected frequencies of DNA patterns using Whittle's formula, *J. Appl. Probab.*, *28*, 886–892.

Crochemore, M. (1981). An optimal algorithm for computing the repetitions in a word, *Inform. Proc. Letters*, *12*, 244–250.

Crochemore, M. (1983). Recherche linéaire d'un carré dans un mot, *Comptes Rendus Acad. Sci. Paris Sér. I Math.*, *296*, 781–784.

Crochemore, M. (1986). Transducers and repetitions, *Theoret. Comput. Sci.*, *45*(1), 63–86.

Crochemore, M. (1987). Longest common factor of two words, In Ehrig, H., Kowalski, R., Levi, G., and Montanari, U. (Eds.), *2nd Theory and Practice of Software Development (TAPSOFT), Pisa, Italy*, Vol. 249 of *Lect. Notes Comp. Sci.*, pp. 26–36. Springer-Verlag.

Crochemore, M., Hancart, C., and Lecroq, T. (2001). *Algorithmique du texte*. Vuibert.

Crochemore, M., Mignosi, F., Restivo, A., and Salemi, S. (2000). Data compression using antidictonaries, *Proceedings of the IEEE*, *88*(11), 1756–1768. Special issue *Lossless data compression* edited by J. Storer.

Crochemore, M. and Rytter, W. (1994). *Text Algorithms*. The Clarendon Press Oxford University Press.

Crochemore, M. and Rytter, W. (1995). Squares, cubes, and time-space efficient string searching, *Algorithmica*, *13*, 405–425.

Crochemore, M. and Vérin, R. (1997). On compact directed acyclic word graphs, In Mycielski, J., Rozenberg, G., and Salomaa, A. (Eds.), *Structures in Logic and Computer Science*, Vol. 1261 of *Lect. Notes Comp. Sci.*, pp. 192–211. Springer-Verlag.

Culik II, K. and Kari, J. (1997). Digital Images and Formal Languages, In Rozenberg, G. and Salomaa, A. (Eds.), *Handbook of Formal Languages*, Vol. 3, pp. 599–616. Springer-Verlag.

Dacunha-Castelle, D. and Duflo, M. (1983). *Probabilités et statistiques 2.Problèmes à temps mobile*. Masson.

De Luca, A. (1981). A combinatorial property of the Fibonacci words, *Inform. Proc. Letters*, *12*(4), 193–195.

Deheuvels, P., Devroye, L., and Lynch, J. (1986). Exact convergence rate in the limit theorems of Erdös-Rényi and Shepp, *Ann. Probab.*, *14*, 209–223.

Dekking, F. M. (1982). Recurrent sets, *Adv. in Math.*, *44*(1), 78–104.

Del Lungo, A., Mirolli, M., Pinzani, R., and Rinaldi, S. (2001). A bijection for directed-convex polyominoes, In Cori, R., Mazoyer, J., Morvan, M., and Mosseri, R. (Eds.), *Discrete Models: Combinatorics, Computation, and Geometry, DM-CCG 2001*, Vol. AA of *DMTCS Proceedings*, pp. 133–144.

Dembo, A. and Karlin, S. (1992). Poisson approximations for r-scan processes, *Ann. Appl. Probab.*, *2*, 329–357.

Denise, A. and Régnier, M. (2004). Rare events and conditional events on random strings, *Discrete Math. Theor. Comput. Sci.*, *6*, 191–214.

Denise, A., Régnier, M., and Vandenbogaert, M. (2001). Assessing the statistical significance of overrepresented oligonucleotides, In Gascuel, O. and Moret, B. M. E. (Eds.), *Algorithms in Bioinformatics (WABI 2001)*, Vol. 2149 of *Lect. Notes Comp. Sci.*, pp. 85–97. Springer-Verlag.

Dershowitz, N. and Zaks, S. (1990). The cycle lemma and some applications, *European J. Combin.*, *11*, 35–40.

Devroye, L., Szpankowski, W., and Rais, B. (1992). A note of the height of suffix trees, *SIAM J. Comput.*, *21*, 48–53.

Di Francesco, P. (2001). Matrix model combinatorics: applications to folding and coloring, In *Random Matrix Models and Their Applications*, Vol. 40 of *Math. Sci. Res. Inst. Publ.*, pp. 111–170. Cambridge University Press.

Duchon, P., Flajolet, P., Louchard, G., and Schaeffer, G. (2002). Random sampling from Boltzmann principles, In Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., and Conejo, R. (Eds.), *29th Automata, Languages and Programming (ICALP '02)*, Vol. 2380 of *Lect. Notes Comp. Sci.*, pp. 501–513. Springer-Verlag.

Dumont, J.-M. and Thomas, A. (1989). Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoret. Comput. Sci.*, *65*(2), 153–169.

Dumont, J.-M. and Thomas, A. (1993). Digital sum moments and substitutions, *Acta Arith.*, *64*, 205–225.

Durand, F. (1998). A characterization of substitutive sequences using return words, *Discrete Math.*, *179*(1-3), 89–101.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). *Biological Sequence Analysis. Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press.

Duval, J.-P. (1983). Factorizing words over an ordered alphabet, *J. Algorithms*, *4*(4), 363–381.

Duval, J.-P. (1998). Périodes locales et propagation de périodes dans un mot, *Theoret. Comput. Sci.*, *204*(1-2), 87–98.

Duval, J.-P., Kolpakov, R., Kucherov, G., Lecroq, T., and Lefebvre, A. (2003). Linear-time computation of local periods, In Rovan, B. and Vojtas, P. (Eds.), *28th Mathematical Foundations of Computer Science (MFCS 2003)*, Vol. 2747 of *Lect. Notes Comp. Sci.*, pp. 388–397. Springer-Verlag.

Duval, J.-P., Mignosi, F., and Restivo, A. (2001). Recurrence and periodicity in infinite words from local periods, *Theoret. Comput. Sci.*, *262*(1), 269–284.

Dvoretzky, A. and Motzkin, T. (1947). A problem of arrangements, *Duke Math. J.*, *14*, 305–313.

Dyck, W. (1882). Gruppentheoretische Studien, *Math. Ann.*, *20*, 1–44.

Eichelsbacher, P. and Roos, M. (1999). Compound Poisson approximation for dissociated random variables via Stein's method, *Combin. Probab. Comput.*, *8*, 335–346.

Eilenberg, S. (1974). *Automata, Languages, and Machines. Vol A.* Academic Press.

El Karoui, M., Biaudet, V., Schbath, S., and Gruss, A. (1999). Characteristics of Chi distribution on several bacterial genomes, *Res. Microbiol.*, *150*, 579–587.

Erdős, P. and Rényi, A. (1970). On a new law of large numbers, *J. Anal. Math.*, *23*, 103–111.

Erhardsson, T. (1997). *Compound Poisson approximation for Markov chains.* Ph.D. thesis, Royal Institute of Technology, Stockholm.

Erhardsson, T. (1999). Compound Poisson approximation for Markov chains using Stein's method, *Ann. Probab.*, *27*, 565–596.

Erhardsson, T. (2000). Compound Poisson approximation for counts of rare patterns in Markov chains and extreme sojourns in birth-death chains, *Ann. in Appl. Probab.*, *10*, 573–591.

Eskin, E., Gelfand, M. S., and Pevzner, P. A. (2003). Genome-wide analysis of bacterial promoter regions, In *Pacific Symposium on Biocomputing (PSB)*, pp. 29–40.

Eskin, E. and Pevzner, P. A. (2002). Finding composite regulatory patterns in DNA sequences, In *Intelligent Systems for Molecular Biology (ISMB)*, pp. 354–363.

Farach, M. (1997). Optimal suffix tree construction with large alphabets, In *38th Foundations of Computer Science (FOCS)*, pp. 137–143, Miami Beach, FL.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I. J. Wiley and Sons. Third edition.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II. J. Wiley and Sons. Second edition.

Fellows, M. R., Gramm, J., and Niedermeier, R. (2002). On the parameterized intractability of CLOSEST SUBSTRING and related problems, In Alt, H. and Ferreira, A. (Eds.), *Theoretical Aspects of Computer Science (STACS '02)*, Vol. 2285 of *Lect. Notes Comp. Sci.*, pp. 262–273. Springer-Verlag.

Ferenczi, S. (1992). Bounded remainder sets, *Acta Arith.*, *61*, 319–326.

Ferenczi, S. and Mauduit, C. (1997). Transcendence of numbers with a low complexity expansion, *J. Number Theory*, *67*, 146–161.

Fitch, W. (1975). Toward defining the course of evolution: minimum change for a specified tree topology, *Systematic Zoology*, *20*, 406–416.

Flajolet, P. (1987). Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.*, *49*(2-3), 283–309.

Flajolet, P., Gourdon, X., and Dumas, P. (1995). Mellin transforms and asymptotics: harmonic sums, *Theoret. Comput. Sci.*, *144*, 3–58.

Flajolet, P., Guivarc'h, Y., Szpankowski, W., and Vallée, B. (2001). Hidden pattern statistics, In Oreijas, F., Spirakis, P., and Leeuwen, J. van (Eds.), *Automata, Languages and Programming (ICALP 2001)*, Vol. 2076 of *Lect. Notes Comp. Sci.*, pp. 152–165. Springer-Verlag.

Flajolet, P. and Sedgewick, R. (2002). *Analytic Combinatorics – Symbolic Combinatorics*. Technical report, INRIA.

Flajolet, P., Zimmerman, P., and Van Cutsem, B. (1994). A calculus for the random generation of labelled combinatorial structures, *Theoret. Comput. Sci.*, *132*(1-2), 1–35.

Fogg, N. P. (2002). *Substitutions in Dynamics, Arithmetics and Combinatorics*, Vol. 1794 of *Lect. Notes Math.* Springer-Verlag. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

Fraenkel, A. and Simpson, J. (1995). How many squares must a binary sequence contain?, *Electronic J. Comb.*, *2*(R2), 9pp.

Fraenkel, A. and Simpson, J. (1998). How many squares can a string contain?, *J. Combin. Th. A*, *82*, 112–120.

Fraenkel, A. and Simpson, J. (1999). The exact number of squares in Fibonacci words, *Theoret. Comput. Sci.*, *218*(1), 83–94.

Fredricksen, H. and Maiorana, J. (1978). Necklaces of beads in k colors and k-ary de Bruijn sequences, *Discrete Math.*, *23*(3), 207–210.

Frougny, C. and Solomyak, B. (1992). Finite beta-expansions, *Ergodic Theory Dynam. Systems*, *12*(4), 713–723.

Fu, J. C. (1993). Poisson convergence in reliability of a large linearly connected system as related to coin tossing, *Statist. Sinica*, *3*, 261–275.

Galil, Z. and Seiferas, J. (1983). Time-space optimal string matching, *J. Comput. System Sci.*, *26*(3), 280–294.

Gambaudo, J.-M., Hubert, P., Tisseur, P., and Vaienti, S. (Eds.) (2000). *Dynamical Systems (Luminy-Marseille, 1998)*. World Scientific.

Gantmacher, F. R. (1959). *The Theory of Matrices* Vols 1, 2. Chelsea. Translated from the Russian original.

Gardner, M. (1966). *New Mathematical Diversions from Scientific American*. Simon and Schuster.

Gentleman, J. (1994). The distribution of the frequency of subsequences in alphabetic sequences, as exemplified by deoxyribonucleic acid, *Appl. Statist.*, *43*, 404–414.

Gentleman, J. and Mullin, R. (1989). The distribution of the frequency of occurrence of nucleotide subsequences, based on their overlap capability, *Biometrics*, *45*, 35–52.

Geske, M. X., Godbole, A. P., Schaffner, A. A., Skolnick, A. M., and Wallstrom, G. L. (1995). Compound Poisson approximations for word patterns under Markovian hypotheses, *J. Appl. Probab.*, *32*, 877–892.

Gessel, I. M. (2000). On the number of convex polyominoes, *Ann. Sci. Math. Québec*, *24*(1), 63–66.

Giegerich, R., Kurtz, S., and Stoye, J. (1999). Efficient implementation of lazy suffix trees, In *3rd Workshop on Algorithmic Engineering (WAE99)*, Vol. 1668 of *Lect. Notes Comp. Sci.*, pp. 30–42. Springer-Verlag.

Glaz, J., Naus, J., and Wallenstein, S. (2001). *Scan Statistics*. Springer-Verlag.

Godbole, A. P. (1991). Poisson approximations for runs and patterns of rare events, *Adv. in Appl. Probab.*, *23*, 851–865.

Godbole, A. P. and Schaffner, A. A. (1993). Improved Poisson approximations for word patterns, *Adv. in Appl. Probab.*, *25*, 334–347.

Gold, E. M. (1967). Language identification in the limit, *Inform. and Control*, *10*(5), 447–474.

Good, I. (1953). The population frequencies of species and the estimation of population parameters, *Biometrica*, *40*, 237–264.

Gordon, L., Schilling, M., and Waterman, M. (1986). An extreme value theory for long head runs, *Probab. Theory Related Fields*, *72*, 279–287.

Goulden, I. P. and Jackson, D. M. (1983). *Combinatorial Enumeration*. J. Wiley and Sons.

Gouyou-Beauchamps, D. and Viennot, G. (1988). Equivalence of the two-dimensional directed animal problem to a one-dimensional path problem, *Adv. in Appl. Math.*, *9*(3), 334–357.

Gross, M. (1975). *Méthodes en syntaxe*. Hermann, Paris.

Gross, M. (1979). On the failure of generative grammar, *Language*, *55*(4), 859–885.

Gross, M. (1989). The use of finite automata in the lexical representation of natural language, In Gross, M. and Perrin, D. (Eds.), *Electronic Dictionaries and Automata in Computational Linguistics*, Vol. 377 of *Lect. Notes Comp. Sci.*, pp. 34–50. Springer-Verlag.

Gross, M. (1995). Representation of finite utterances and the automatic parsing of texts, *Language Research*, *31*(2), 291–307. Seoul National University.

Gross, M. and Lentin, A. (1967). *Notions sur les grammaires formelles*. Gauthier-Villars, Paris. 2nd edition: 1970.

Guibas, L. and Odlyzko, A. (1980). Long repetitive patterns in random sequences, *Z. Wahrsch. Verw. Gebiete*, *53*, 241–262.

Guibas, L. and Odlyzko, A. (1981a). Periods in strings, *J. Combin. Th. A*, *30*, 19–43.

Guibas, L. and Odlyzko, A. (1981b). String overlaps, pattern matching, and nontransitive games, *J. Combin. Th. A*, *30*, 183–208.

Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press.

Gusto, G. (2000). Approximation par une loi de Poisson composée de la loi du comptage d'un mot rare dans une chaîne de Markov, Master's thesis, DEA Modélisation Stochastique et Statistique, Université Paris XI, Orsay.

Guthrie, D. and Youssef, M. (1970). Empirical evaluation of some chi-square tests for the order of a Markov chain, *J. Amer. Statist. Soc.*, *65*, 631–634.

Guy, R. K., Krattenthaler, C., and Sagan, B. E. (1992). Lattice paths, reflections, & dimension-changing bijections, *Ars Combin.*, *34*, 3–15.

Gwadera, R., Atallah, M., and Szpankowski, W. (2003). Reliable detection of episodes in event sequences, In *3rd IEEE Conf. on Data Mining*, pp. 67–74. IEEE Computer Soc.

Harris, Z. S. (1952). Discourse analysis, *Language*, *28*, 1–30.

Harris, Z. S. (1970). *Papers in Structural and Transformational Linguistics*. D. Reidel, Dordrecht/Holland.

Hirano, K. and Aki, S. (1993). On number of occurrences of success runs of specified length in a two-state Markov chain, *Statist. Sinica*, *3*, 313–320.

Hirschberg, D. S. (1977). Algorithms for the longest common subsequence problem, *J. Assoc. Comput. Mach.*, *24*(4), 664–675.

Hochstättler, W., Loebl, M., and Moll, C. (1996). Generating convex polyominoes at random, *Discrete Math.*, *153*, 165–176.

Holton, C. and Zamboni, L. Q. (1998). Geometric realizations of substitutions, *Bull. Soc. Math. France*, *126*(2), 149–179.

Holton, C. and Zamboni, L. Q. (1999). Descendants of primitive substitutions, *Theory Comput. Syst.*, *32*(2), 133–157.

Hopcroft, J. (1971). An $n \log n$ algorithm for minimizing states in a finite automaton, In *Theory of Machines and Computations*, pp. 189–196. Academic Press.

Huang, H. (2002). Error bounds on multivariate normal approximations for word count statistics, *Adv. in Appl. Probab.*, *34*, 559–587.

Hui, L. (1992). Color set size problem with applications to string matching, In *Combinatorial Pattern Matching*, Vol. 644 of *Lect. Notes Comp. Sci.*, pp. 230–243. Springer-Verlag.

Hunt, J. W. and Szymanski, T. G. (1977). A fast algorithm for computing longest common subsequences, *Comm. Assoc. Comput. Mach.*, *20*, 350–353.

Hwang, H.-K. (1996). Large deviations for combinatorial distributions I: Central limit theorems, *Ann. in Appl. Probab.*, *6*, 297–319.

Iliopoulos, C., Moore, D., and Smyth, W. (1997). A characterization of the squares in a Fibonacci string, *Theoret. Comput. Sci.*, *172*, 281–291.

Inenaga, S., Hoshino, H., Shinohara, A., Takeda, M., Arikawa, S., Mauri, G., and Pavesi, G. (2001). On-line construction of compact directed acyclic word graphs, In Amir, A. and Landau, G. M. (Eds.), *12th Combinatorial Pattern Matching,* Jerusalem, Vol. 2089 of *Lect. Notes Comp. Sci.*, pp. 169–180. Springer-Verlag.

Ito, S. (1996). Simultaneous approximations and dynamical systems (on the simultaneous approximation of (α, α^2) satisfying $\alpha^3 + k\alpha - 1 = 0$), *Sūrikaisekikenkyūsho Kōkyūroku*, *958*, 59–61. Analytic number theory (Japanese) (Kyoto, 1994).

Ito, S. (2000). On periodic expansions of cubic numbers and Rauzy fractals, In Gambaudo, J.-M., Hubert, P., Tisseur, P., and Vaienti, S. (Eds.), *Dynamical Systems (Luminy-Marseille, 1998)*, pp. 144–164. World Scientific.

Ito, S., Fujii, J.and Higashino, H., and Yasutomi, S.-i. (2003). On simultaneous approximation to (α, α^2) with $\alpha^3 + k\alpha - 1 = 0$, *J. Number Theory*, *99*(2), 255–283.

Ito, S. and Kimura, M. (1991). On Rauzy fractal, *Japan J. Indust. Appl. Math.*, *8*(3), 461–486.

Ito, S. and Ohtsuki, M. (1993). Modified Jacobi-Perron algorithm and generating Markov partitions for special hyperbolic toral automorphisms, *Tokyo J. Math.*, *16*(2), 441–472.

Ito, S. and Sano, Y. (2001). On periodic β-expansions of Pisot numbers and Rauzy fractals, *Osaka J. Math.*, *38*(2), 349–368.

Jacquet, P. and Régnier, M. (1986). Trie partitioning process: limiting distributions, In *CAAP '86 (Nice, 1986)*, Vol. 214 of *Lect. Notes Comp. Sci.*, pp. 196–210. Springer-Verlag.

Jacquet, P. and Szpankowski, W. (1991). Analysis of digital tries with Markovian dependency, *IEEE Trans. Inform. Theory*, *37*, 1470–1475.

Jacquet, P. and Szpankowski, W. (1994). Autocorrelation on words and its applications: analysis of suffix trees by string-ruler approach, *J. Combin. Th. A*, *66*, 237–269.

Jacquet, P. and Szpankowski, W. (1998). Analytical de-Poissonization and its applications, *Theoret. Comput. Sci.*, *201*, 1–62.

Jacquet, P., Szpankowski, W., and Apostol, I. (2002). A universal predictor based on pattern matching, *IEEE Trans. Inform. Theory*, *48*, 1462–1472.

Janson, S. (1997). *Gaussian Hilbert Spaces.* Cambridge University Press.

Janson, S. (to appear). Large deviations for sums of partially dependent random variables, *Random Structures & Algorithms*.

Johnson, C. D. (1972). *Formal Aspects of Phonological Description*. Mouton, The Hague.

Jonassen, I., Collins, J. F., and Higgins, D. (1995). Finding flexible patterns in unaligned protein sequences, *Protein Science*, *4*(8), 1587–1595.

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall.

Justin, J. (2000). On a paper by M. Castelli, F. Mignosi, A. Restivo, *Theor. Inform. Appl.*, *34*(5), 373–377.

Justin, J. and Pirillo, G. (2002a). Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.*, *276*(1-2), 281–313.

Justin, J. and Pirillo, G. (2002b). On a characteristic property of Arnoux-Rauzy sequences, *Theor. Inform. Appl.*, *36*(4), 385–388 (2003).

Karhumäki, J. (1983). On cube-free ω-words generated by binary morphisms, *Discr. Appl. Math.*, *5*(3), 279–297.

Kärkkäinen, J. and Sanders, P. (2003). Simple linear work suffix array constructio, In Baeten, J. C. M., Lenstra, J. K., Parrow, J., and Woeginger, G. J. (Eds.), *30th Automata, Languages and Programming (ICALP '03)*, Vol. 2719 of *Lect. Notes Comp. Sci.*, pp. 943–955. Springer-Verlag.

Karlin, S., Burge, C., and Campbell, A. M. (1992). Statistical analyses of counts and distributions of restriction sites in DNA sequences, *Nucl. Acids Res.*, *20*, 1363–1370.

Karlin, S. and Dembo, A. (1992). Limit distributions of maximal segmental score among Markov dependent partial sums, *Adv. in Appl. Probab.*, *24*, 113–140.

Karlin, S. and Macken, C. (1991). Assessment of inhomogeneities in an *E. coli* physical map, *Nucl. Acids Res.*, *19*, 4241–4246.

Karlin, S. and Ost, F. (1987). Counts of long aligned word matches among random letter sequences, *Adv. in Appl. Probab.*, *19*, 293–351.

Karlin, S. and Taylor, H. (1975). *A First Course in Stochastic Processes* (Second edition). Academic Press.

Kato, T. (1980). *Perturbation Theory for Linear Operators*. Springer-Verlag.

Katz, S. (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer, *IEEE Trans. Acoust. Speech Signal Process.*, *35*, 400–401.

Kim, D. K., Sim, J. S., Park, H., and Park, K. (2003). Linear-time construction of suffix arrays, In Baeza-Yates, R., Chávez, E., and Crochemore, M. (Eds.), *Combinatorial Pattern Matching (CPM 2003)*, Vol. 2676 of *Lect. Notes Comp. Sci.*, pp. 186–199. Springer-Verlag.

Klarner, D. A. (1997). Polyominoes, In *Handbook of Discrete and Computational Geometry*, CRC Press Ser. Discrete Math. Appl., pp. 225–240. CRC.

Kleffe, J. and Borodovsky, M. (1992). First and second moment of counts of words in random texts generated by Markov chains, *Comput. Appl. Biosci.*, *8*, 433–441.

Kleffe, J. and Langbecker, U. (1990). Exact computation of pattern probabilities in random sequences generated by Markov chains, *Comput. Appl. Biosci.*, *6*, 347–353.

Knuth, D. E. (1965). On the translation of languages from left to right, *Inform. and Control*, *8*, 607–639.

Knuth, D. E. (1988). Fibonacci multiplication, *Appl. Math. Lett.*, *1*(1), 57–60.

Knuth, D. E., Morris, J. H., and Pratt, V. R. (1977). Fast pattern matching in strings, *SIAM J. Comput.*, *6*, 323–350.

Ko, P. and Aluru, S. (2003). Space-efficient linear-time construction of suffix arrays, In Baeza-Yates, R., Chávez, E., and Crochemore, M. (Eds.), *Combinatorial Pattern Matching (CPM 2003)*, Vol. 2676 of *Lect. Notes Comp. Sci.*, pp. 200–210. Springer-Verlag.

Kolpakov, R., Bana, G., and Kucherov, G. (2003). `mreps`: efficient and flexible detection of tandem repeats in DNA, *Nucl. Acids Res.*, *31*(13), 3672–3678.

Kolpakov, R. and Kucherov, G. (1999). Finding maximal repetitions in a word in linear time, In *40th Symp. Foundations of Computer Science (FOCS)*, pp. 596–604. IEEE Computer Society Press.

Kolpakov, R. and Kucherov, G. (2000a). Finding repeats with fixed gap, In *7th Symp. String Processing and Information Retrieval (SPIRE)*, pp. 162–168. IEEE Computer Society Press.

Kolpakov, R. and Kucherov, G. (2000b). On maximal repetitions in words, *J. Discrete Algorithms*, *1*(1), 159–186.

Kolpakov, R. and Kucherov, G. (2003). Finding approximate repetitions under Hamming distance, *Theoret. Comput. Sci.*, *33*(1), 135–156.

Kosaraju, S. (1994). Computation of squares in string, In Crochemore, M. and Gusfield, D. (Eds.), *5th Combinatorial Pattern Matching*, Vol. 807 of *Lect. Notes Comp. Sci.*, pp. 146–150. Springer-Verlag.

Koskenniemi, K. (1983). *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*. University of Helsinki, Department of General Linguistics. Publications, no. 11.

Kucherov, G., Ochem, P., and Rao, M. (2003). How many square occurrences must a binary sequence contain, *Electronic J. Comb.*, *10*(1), 11pp.

Kucherov, G. and Rusinowitch, M. (1997). Matching a set of strings with variable length don't cares, *Theoret. Comput. Sci.*, *178*, 129–154.

Kuich, W. and Salomaa, A. (1986). *Semirings, Automata, Languages*. Springer-Verlag.

Landau, G. and Schmidt, J. (1993). An algorithm for approximate tandem repeats, In Apostolico, A., Crochemore, M., Galil, Z., and Manber, U. (Eds.), *4th Combinatorial Pattern Matching (Padova)*, Vol. 684 of *Lect. Notes Comp. Sci.*, pp. 120–133. Springer-Verlag, Padova, Italy.

Laporte, E. (1997). Rational transductions for phonetic conversion and phonology, In Roche, E. and Schabès, Y. (Eds.), *Finite-state Language Processing*, Language, Speech, and Communication Series, chap. 14, pp. 407–428. MIT Press.

Lee, K.-F. (1990). Context dependent phonetic hidden Markov models for continuous speech recognition, *IEEE Trans. Acoust. Speech Signal Process.*, *38*(4), 599–609.

Lempel, A. and Ziv, J. (1976). On the complexity of finite sequences, *IEEE Trans. Inform. Theory*, *IT-22*, 75–81.

Leung, M. Y., Marsh, G. M., and Speed, T. P. (1996). Over and underrepresentation of short DNA words in Herpesvirus genomes, *J. Comput. Biol.*, *3*, 345–360.

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics Dokl.*, *10*, 707–710.

Li, S.-Y. (1980). A martingale approach to the study of occurrence of sequence patterns in repeated experiments, *Ann. Probab.*, *8*, 1171–1176.

Lind, D. and Marcus, B. (1996). *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press.

Lint, J. H. van and Wilson, R. M. (1992). *A Course in Combinatorics*. Cambridge University Press.

Lonardi, S. (2001). *Global detectors of unusual words design, implementation and applications to pattern discovery in biosequences*. Ph.D. thesis, Purdue University, West Lafayette, Indiana.

Lothaire, M. (1997). *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press. Reprint with corrections of the 1983 original edition.

Lothaire, M. (2002). *Algebraic Combinatorics on Words*, Vol. 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press.

Lundstrom, R. (1990). *Stochastic models and statistical methods for DNA sequence data*. Ph.D. thesis, University of Utah.

Mahmoud, H. (1992). *Evolution of Random Search Trees*. J. Wiley and Sons.

Main, M. (1989). Detecting leftmost maximal periodicities, *Discr. Appl. Math.*, *25*, 145–153.

Main, M. and Lorentz, R. (1984). An $O(n \log n)$ algorithm for finding all repetitions in a string, *J. Algorithms*, *5*(3), 422–432.

Main, M. and Lorentz, R. (1985). Linear time recognition of square free strings, In Apostolico, A. and Galil, Z. (Eds.), *Combinatorial Algorithms on Words*, Vol. 12 of *NATO Advanced Science Institutes, Series F*, pp. 272–278. Springer-Verlag.

Manber, U. and Myers, G. (1993). Suffix arrays: a new method for on-line string searches, *SIAM J. Comput.*, *22*(5), 935–948.

Manzini, G. (2001). An analysis of the Burrows-Wheeler transform, *J. Assoc. Comput. Mach.*, *48*(3), 407–430.

Marsan, L. and Sagot, M.-F. (2000a). Algorithms for extracting structured motifs using a suffix tree with an application to promoter and regulatory site consensus identification, *J. Comput. Biol.*, *7*, 345–362.

Marsan, L. and Sagot, M.-F. (2000b). Extracting structured motifs using a suffix tree. Algorithms and application to consensus identification, In *4th Research in Computational Molecular Biology (RECOMB)*, pp. 210–219. ACM Press.

Marsan, L. and Sagot, M.-F. (2001). Algorithms for extracting structured motifs using a suffix tree with application to promoter and regulatory consensus identification, *J. Comput. Biol.*, *7*, 345–360.

McCluer, C. R. (2000). The many proofs and applications of Perron's theorem, *SIAM Review*, *42*, 487–498.

McCreight, E. M. (1976). A space-economical suffix tree construction algorithm, *J. Assoc. Comput. Mach.*, *23*(2), 262–272.

McEliece, R. J. (2002). *The Theory of Information and Coding* (2nd edition)., Vol. 86 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press.

Mehldau, G. and Myers, E. (1993). A system for pattern matching applications on biosequences, *Computer Applied Bioscience (CABIOS)*, *9*(3), 299–314.

Messaoudi, A. (1998). Propriétés arithmétiques et dynamiques du fractal de Rauzy, *J. Théor. Nombres Bordeaux*, *10*(1), 135–162.

Messaoudi, A. (2000a). Frontière du fractal de Rauzy et système de numération complexe, *Acta Arith.*, *95*(3), 195–224.

Messaoudi, A. (2000b). Généralisation de la multiplication de Fibonacci, *Math. Slovaca*, *50*(2), 135–148.

Messaoudi, A. (2002). Tribonacci multiplication, *Appl. Math. Lett.*, *15*(8), 981–985.

Mignosi, F. and Pirillo, G. (1992). Repetitions in the Fibonacci infinite word, *Theoret. Inform. Appl.*, *26*(3), 199–204.

Mignosi, F., Restivo, A., and Salemi, S. (1995). A periodicity theorem on words and applications, In *20th Mathematical Foundations of Computer Science (MFCS)*, Vol. 969 of *Lect. Notes Comp. Sci.*, pp. 337–348. Springer-Verlag.

Mignosi, F. and Zamboni, L. Q. (2002). On the number of Arnoux-Rauzy words, *Acta Arith.*, *101*(2), 121–129.

Miller, G. A. and Chomsky, N. (1963). Finitary models of language users, In Luce, R. D., Bush, R. R., and Galanter, E. (Eds.), *Handbook of Mathematical Psychology*, Vol. 2, chap. 13, pp. 419–491. Wiley, New York.

Mohri, M. (1994). Minimization of sequential transducers, In Crochemore, M. and Gusfield, D. (Eds.), *Combinatorial Pattern Matching 94*, Vol. 807 of *Lect. Notes Comp. Sci.*, pp. 151–163. Springer-Verlag.

Mohri, M. (1997). Finite-state transducers in language and speech processing, *Comput. Linguistics*, *23*(2), 269–311.

Mohri, M. (2002). Semiring frameworks and algorithms for shortest-distance problems, *J. Autom. Lang. Comb.*, *7*(3), 321–350.

Mohri, M., Pereira, F., and Riley, M. (2000). The design principles of a weighted finite-state transducer library, *Theoret. Comput. Sci.*, *231*(1), 17–32.

Mohri, M., Pereira, F. C. N., and Riley, M. (1996). Weighted automata in text and speech processing, In *12th European Conf. Artificial Intelligence (ECAI 1996), Workshop on Extended finite state models of language, Budapest, Hungary*. J. Wiley and Sons.

Mohri, M., Pereira, F. C. N., and Riley, M. (2002). Weighted finite-state transducers in speech recognition, *Computer Speech and Language*, *16*(1), 69–88. Available at `http://www.research.att.com/~mohri/postscript/csl01.ps`.

Moody, R. V. (1997). Meyer sets and their duals, In Moody, R. V. (Ed.), *The Mathematics of Long-Range Aperiodic Order*, Vol. 13 of *CRM Monograph Series*, pp. 403–441. Kluwer.

Moore, E. F. (1956). Gedanken-experiments on sequential machines, In *Automata Studies*, Vol. 34 of *Annals of Mathematics Studies*, pp. 129–153. Princeton University Press.

Muri, F. (1998). Modelling Bacterial Genomes using Hidden Markov Models, In Payne, R. and Green, P. (Eds.), *Compstat'98*, pp. 89–100. Physica-Verlag.

Nicodème, P., Salvy, B., and Flajolet, P. (2002). Motif statistics, *Theoret. Comput. Sci.*, *287*(2), 593–617. Algorithms (Prague, 1999).

Nijenhuis, A. and Wilf, H. S. (1978). *Combinatorial Algorithms* (2nd edition). Academic Press.

Nuel, G. (2001). *Grandes déviations et chaînes de Markov pour l'étude des mots exceptionnels dans les séquences biologiques*. Ph.D. thesis, Université d'Evry Val d'Essonne.

Parida, L., Rigoutsos, I., Floratos, A., Platt, D., and Gao, Y. (2000). Pattern discovery on character sets and real-valued data: Linear bound on irredundant motifs and efficient polynomial time algorithm, In *11th SIAM Symposium on Discrete Algorithms (SODA)*, pp. 297–308. ACM Press.

Parida, L., Rigoutsos, I., and Platt, D. (2001). An output-sensitive flexible pattern discovery algorithm, In Amir, A. and Landau, G. (Eds.), *12th Combinatorial Pattern Matching*, Vol. 2089 of *Lect. Notes Comp. Sci.*, pp. 131–142. Springer-Verlag.

Paumier, S. (2001). Some remarks on the application of a lexicon-grammar, *Lingvisticae Investigationes*, *24*(2), 245–256.

Pavesi, G., Mauri, G., and Pesole, G. (2001a). An algorithm for finding signals of unknown length in DNA sequences, *Bioinformatics*, *17*(Suppl. 1), 207–214.

Pavesi, G., Mauri, G., and Pesole, G. (2001b). Methods for pattern discovery in unaligned biological sequences, *Briefings in Bioinformatics*, *2*(4), 417–430.

Paz, A. (1971). *Introduction to Probabilistic Automata*. Academic Press.

Pereira, F. C. N. and Riley, M. D. (1997). Speech recognition by composition of weighted finite automata, In *Finite-State Language Processing*, pp. 431–453. MIT Press.

Petersen, H. (1994). The ambiguity of primitive words, In Enjalbert, P., Mayr, E. W., and Wagner, K. W. (Eds.), *Theoretical Aspects of Computer Science (STACS '94)*, Vol. 775 of *Lect. Notes Comp. Sci.*, pp. 679–690. Springer-Verlag.

Petersen, H. (1996). On the language of primitive words, *Theoret. Comput. Sci.*, *161*, 141–156.

Pevzner, P. A. (1989). l-tuple DNA sequencing: computer analysis, *J. Biomol. Struct. Dynamics*, *7*, 63–73.

Pevzner, P. A. (1995). DNA physical mapping and alternating Eulerian cycles in colored graphs, *Algorithmica*, *13*, 77–105.

Pevzner, P. A. (2000). *Computational Molecular Biology. An Algorithmic Approach*. The MIT Press.

Pevzner, P. A. and Sze, S.-H. (2000). Combinatorial approaches to finding subtle signals in DNA sequences, In *Intelligent Systems for Molecular Biology (ISMB)*, pp. 269–278.

Pin, J.-E. (1986). *Varieties of Formal Languages*. North Oxford.

Pirillo, G. (1997). Fibonacci numbers and words, *Discrete Math.*, *173*(1-3), 197–207.

Pisanti, N., Crochemore, M., Grossi, R., and Sagot, M.-F. (2003). A basis of tiling motifs for generating repeated patterns and its complexity for higher quorum, In Rovan, B. and Vojtás, P. (Eds.), *Mathematical Foundations of Computer Science (MFCS 2003)*, Vol. 2747 of *Lect. Notes Comp. Sci.*, pp. 622–631. Springer-Verlag.

Pitman, J. (1998). Enumerations of trees and forests related to branching processes and random walks, In Aldous, D. and Propp, J. (Eds.), *Microsurveys in Discrete Probability*, Vol. 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* Amer. Math. Soc.

Poulalhon, D. and Schaeffer, G. (2003). Optimal coding and sampling of triangulations, In Baeten, J. C. M., Lenstra, J. K., Parrow, J., and Woeginger, G. J. (Eds.), *30th Automata, Languages and Programming (ICALP '03)*, Vol. 2719 of *Lect. Notes Comp. Sci.*, pp. 1080–1094. Springer-Verlag.

Praggastis, B. (1999). Numeration systems and Markov partitions from self-similar tilings, *Trans. Amer. Math. Soc.*, *351*(8), 3315–3349.

Prum, B., Rodolphe, F., and Turckheim, É. (1995). Finding words with unexpected frequencies in deoxyribonucleic acid sequences, *J. Roy. Statist. Soc. Ser. B*, *57*, 205–220.

Queffélec, M. (1987). *Substitution Dynamical Systems–Spectral Analysis*, Vol. 1294 of *Lect. Notes Math.* Springer-Verlag.

Rabiner, L. (1989). A tutorial on hidden Markov models, *Proceedings of the IEEE*, *77*(2), 257–286.

Raffinot, M. (1997). Asymptotic estimation of the average number of terminal states in DAWGs, In Baeza-Yates, R. (Ed.), *4th South American Workshop on String Processing*, pp. 140–148. Carleton University Press, Valparaiso, Chile.

Rajarshi, M. B. (1974). Success runs in a two-state Markov chain, *J. Appl. Probab.*, *11*, 190–192.

Raney, G. N. (1960). Functional composition patterns and power series reversion, *Trans. Amer. Math. Soc.*, *94*, 441–451.

Rauzy, G. (1982). Nombres algébriques et substitutions, *Bull. Soc. Math. France*, *110*(2), 147–178.

Rauzy, G. (1988). Rotations sur les groupes, nombres algébriques, et substitutions, In *Sém. Théor. Nombres (Talence, 1987–1988)*. Univ. Bordeaux I. Exp. No. 21.

Rauzy, G. (1990). Sequences defined by iterated morphisms, In *Sequences (Naples/Positano, 1988)*, pp. 275–286. Springer, New York.

Régnier, M. (2000). A unified approach to word occurrence probabilities, *Discr. Appl. Math.*, *104*, 259–280.

Régnier, M. and Szpankowski, W. (1998a). On pattern frequency occurrences in a Markovian sequence, *Algorithmica*, *22*, 631–649.

Régnier, M. and Szpankowski, W. (1998b). On the approximate pattern occurrences in a text, In *Compression and Complexity of Sequences 97*, pp. 253–264. IEEE Computer Society Press.

Reinert, G. and Schbath, S. (1998). Compound Poisson and Poisson process approximations for occurrences of multiple words in Markov chains, *J. Comput. Biol.*, *5*, 223–253.

Reinert, G., Schbath, S., and Waterman, M. (2000). Probabilistic and statistical properties of words: an overview, *J. Comput. Biol.*, *7*, 1–46.

Rensburg, E. J. J. van (2000). *The Statistical Mechanics of Interacting Walks, Polygons, Animals and Vesicles*, Vol. 18 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press.

Reutenauer, C. (1990). Subsequential functions: characterizations, minimization, examples, In Dassow, J. and Kelemen, J. (Eds.), *6th International Meeting of Young Computer Scientists*, Vol. 464 of *Lect. Notes Comp. Sci.*, pp. 62–79. Springer-Verlag.

Revuz, D. (1992). Minimisation of acyclic deterministic automata in linear time, *Theoret. Comput. Sci.*, *92*(1), 181–189.

Rice, J. (1995). *Mathematical Statistics and Data Analysis*. Duxbury Press.

Ridout, D. (1957). Rational approximations to algebraic numbers, *Mathematika*, *4*, 125–131.

Rinott, Y. and Rotar, V. (1996). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics, *J. Multivariate Anal.*, *56*, 333–350.

Risley, R. N. and Zamboni, L. Q. (2000). A generalization of Sturmian sequences: combinatorial structure and transcendence, *Acta Arith.*, *95*(2), 167–184.

Robin, S. (2002). A compound Poisson model for words occurrences in DNA sequences, *J. Roy. Statist. Soc. Ser. C*, *51*, 437–451.

Robin, S. and Daudin, J.-J. (1999). Exact distribution of word occurrences in a random sequence of letters, *J. Appl. Probab.*, *36*, 179–193.

Robin, S. and Daudin, J.-J. (2001). Exact distribution of the distances between any occurrences of a set of words, *Ann. Inst. Statist. Math.*, *36*(4), 895–905.

Robin, S., Daudin, J.-J., Richard, H., Sagot, M.-F., and Schbath, S. (2002). Occurrence probability of structured motifs in random sequences, *J. Comput. Biol.*, *9*, 761–773.

Robin, S. and Schbath, S. (2001). Numerical comparison of several approximations of the word count distribution in random sequences, *J. Comput. Biol.*, *8*, 349–359.

Rocha, E., Viari, A., and Danchin, A. (1998). Oligonucleotide bias in *Bacillus subtilis*: general trends and taxonomic comparisons, *Nucl. Acids Res.*, *26*, 2971–2980.

Roche, E. (1997). Compact factorization of finite-state transducers and finite-state automata, *Nordic J. Comput.*, *4*(2), 187–216.

Rodeh, M., Pratt, V. R., and Even, S. (1981). Linear algorithm for data compression via string matching, *J. Assoc. Comput. Mach.*, *28*(1), 16–24.

Roos, M. (1993). *Stein-Chen method for compound Poisson approximation*. Ph.D. thesis, University of Zurich.

Rudander, J. (1996). *On the first occurrence of a given pattern in a semi-Markov process*. Ph.D. thesis, Uppsala, Sweden.

Sagot, M.-F. (1996). *Ressemblance lexicale et structurale entre macromolécules – Formalisation et approches combinatoires*. Ph.D. thesis, Université de Marne-la-Vallée, Noisy le Grand, France. Thèse de doctorat.

Sagot, M.-F. (1998). Spelling approximate repeated or common motifs using a suffix tree, In Lucchesi, C. and Moura, A. (Eds.), *LATIN'98: Theoretical Informatics: Third Latin American Symposium*, Vol. 1380 of *Lect. Notes Comp. Sci.*, pp. 111–127. Springer-Verlag.

Sagot, M.-F. and Myers, E. (1998). Identifying satellites and periodic repetitions in biological sequences, *J. Comput. Biol.*, *5*(3), 539–554.

Sagot, M.-F., Soldano, H., and Viari, A. (1995). A distance-based block searching algorithm, In *Intelligent Systems for Molecular Biology (ISMB)*, pp. 322–331.

Sagot, M.-F. and Viari, A. (1996). A double combinatorial approach to discovering patterns in biological sequences, In Hirschberg, D. and Myers, G. (Eds.), *7th Combinatorial Pattern Matching*, Vol. 1075 of *Lect. Notes Comp. Sci.*, pp. 186–208. Springer-Verlag.

Sagot, M.-F., Viari, A., Pothier, J., and Soldano, H. (1995). Finding flexible patterns in a text - an application to 3D molecular matching, *Computer Applied Bioscience (CABIOS)*, *11*(1), 59–70.

Sagot, M.-F., Viari, A., and Soldano, H. (1997). Multiple sequence comparison - A peptide matching approach, *Theoret. Comput. Sci.*, *180*(1-2), 115–137.

Sakarovich, J. (2004). *Éléments de théorie des automates*. Vuibert.

Salomaa, A. and Soittola, M. (1978). *Automata-Theoretic Aspects of Formal Power Series*. Springer-Verlag.

Salton, G. (1989). *Automatic Text Processing*. Addison-Wesley.

Sankoff, D. and Kruskal, J. B. (Eds.) (1983). *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison*. Addison-Wesley.

Sano, Y. (2002). On purely periodic β-expansions of Pisot numbers, *Nagoya Math. J.*, *166*, 183–207.

Sano, Y., Arnoux, P., and Ito, S. (2001). Higher dimensional extensions of substitutions and their dual maps, *J. Anal. Math.*, *83*, 183–206.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Harcourt, Brace and World.

Schaeffer, G. (1997). Bijective census and random generation of Eulerian planar maps with prescribed vertex degrees, *Electronic J. Comb.*, *4*(1), Research Paper 20, 14 pp.

Schbath, S. (1995a). Compound Poisson approximation of word counts in DNA sequences, *ESAIM Probab. Statist.*, *1*, 1–16. (`http://www.emath.fr/ps/`).

Schbath, S. (1995b). *Étude asymptotique du nombre d'occurrences d'un mot dans une chaîne de Markov et application à la recherche de mots de fréquence exceptionnelle dans les séquences d'ADN*. Ph.D. thesis, Université René Descartes, Paris V.

Schmidt, K. (1980). On periodic expansions of Pisot numbers and Salem numbers, *Bull. London Math. Soc.*, *12*, 269–278.

Schützenberger, M.-P. (1961). A remark on finite transducers, *Inform. and Control*, *4*, 185–196.

Schützenberger, M.-P. (1964). On the synchronizing properties of certain prefix codes, *Inform. and Control*, *7*, 23–36.

Schützenberger, M.-P. (1977). Sur une variante des fonctions séquentielles, *Theoret. Comput. Sci.*, *4*(1), 47–57.

Schützenberger, M.-P. and Chomsky, N. (1963). The algebraic theory of context-free languages, In *Computer Programming and Formal Systems*, pp. 118–161. North Holland.

Sedgewick, R. (1983). *Algorithms*. Addison-Wesley.

Sedgewick, R. and Flajolet, P. (1995). *An Introduction to the Analysis of Algorithms*. Addison-Wesley.

Senellart, J. (1998). Reconnaissance automatique des entrées du lexique-grammaire des phrases figées, *Travaux de linguistique*, *37*, 109–125.

Senellart, J. (1999). *Outils de reconnaissance d'expressions linguistiques complexes dans de grands corpus*. Ph.D. thesis, LADL, University of Paris 7.

Senoussi, R. (1990). Statistique asymptotique presque-sûre de modèles statistiques convexes, *Ann. Inst. Henri Poincaré*, *26*, 19–44.

Shamir, R. and Tsur, D. (2001). Large scale sequencing by hybridization, In *5th Research in Computational Molecular Biology (RECOMB)*, pp. 269–277. ACM Press.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Tech. J.*, *27*, 379–423, 623–656.

Shannon, C. E. (1951). Prediction and entropy of printed English, *Bell System Tech. J.*, *30*, 50–64.

Shapiro, H. S. (1952). Extremal problems for polynomials and power series, Master's thesis, MIT.

Shields, P. C. (1969). *The Ergodic Theory of Discrete Sample Paths*. Amer. Math. Soc.

Shiloach, Y. (1981). Fast canonization of circular strings, *J. Algorithms*, *2*(2), 107–121.

Silberztein, M. (1994). INTEX: a corpus processing system, In *15th Conf. Comput. Linguistics (COLING'94)*, Vol. 1, pp. 579–583, Kyoto, Japan. Available at http://acl.ldc.upenn.edu/C/C94/C94-1095.pdf.

Sirvent, V. F. (2000a). The common dynamics of the Tribonacci substitutions, *Bull. Belg. Math. Soc. Simon Stevin*, *7*(4), 571–582.

Sirvent, V. F. (2000b). Geodesic laminations as geometric realizations of Pisot substitutions, *Ergodic Theory Dynam. Systems*, *20*(4), 1253–1266.

Sirvent, V. F. and Wang, Y. (2002). Self-affine tiling via substitution dynamical systems and Rauzy fractals, *Pacific J. Math.*, *206*(2), 465–485.

Slisenko, A. (1983). Detection of periodicities and string matching in real time, *J. Soviet Math.*, *22*, 1316–1386. translation from the Russian ordiginal.

Soldano, H., Viari, A., and Champesme, M. (1995). Searching for flexible repeated patterns using a non transitive similarity relation, *Pattern Recognition Letters*, *16*(3), 233–246.

Sourice, S., Biaudet, V., El Karoui, M., Ehrlich, S., and Gruss, A. (1998). Identification of the Chi site of *Haemophilus influenzae* as several sequences related to the *Escherichia coli* Chi site, *Mol. Microbiol.*, *27*, 1021–1029.

Sproat, R. (1997). Multilingual text analysis for text-to-speech synthesis, *J. Natural Language Engineering*, *2*(4), 369–380.

Stanley, R. P. (1999). *Enumerative Combinatorics. Vol. II*, Vol. 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press.

Stefanov, V. (2003). The intersite distances between pattern occurrences in strings generated by general discrete - and continuous- time models: an algorithmic approach, *J. Appl. Probab.*, *40*.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, Vol. 2, pp. 583–602. Univ. California Press, Berkeley.

Storer, J. A. (1988). *Data Compression: Methods and Theory*. Computer Science Press.

Stoye, J. and Gusfield, D. (1998). Simple and flexible detection of contiguous repeats using a suffix tree, In Farach-Colton, M. (Ed.), *9th Combinatorial Pattern Matching*, No. 1448 in Lect. Notes Comp. Sci., pp. 140–152. Springer-Verlag.

Szpankowski, W. (1993a). Asymptotic properties of data compression and suffix trees, *IEEE Trans. Inform. Theory*, *39*, 1647–1659.

Szpankowski, W. (1993b). A generalized suffix tree and its (un)expected asymptotic behaviors, *SIAM J. Comput.*, *22*, 1176–1198.

Szpankowski, W. (2001). *Average Case Analysis of Algorithms on Sequences*. J. Wiley and Sons.

Tanushev, M. S. (1996). Central limit theorem for several patterns in a Markov chain sequence of letters,. Preprint.

Tanushev, M. S. and Arratia, R. (1997). Central limit theorem for renewal theory for several patterns, *J. Comput. Biol.*, *4*, 35–44.

Thurston, W. P. (1989). Groups, tilings and finite state automata, Lectures notes distributed in conjunction with the Colloquium Series, in *AMS Colloquium lectures*.

Tutte, W. T. (1962). A census of planar triangulations, *Canad. J. Math.*, *14*, 21–38.

Ukkonen, E. (1992). Approximate string-matching with q-grams and maximal matches, *Theoret. Comput. Sci.*, *92*, 191–211.

Ukkonen, E. (1995). On-line construction of suffix trees, *Algorithmica*, *14*(3), 249–260.

Valiant, L. G. (1984). A theory of the learnable, *Comm. Assoc. Comput. Mach.*, *27*(11), 1134–1142.

Vallée, B. (2001). Dynamical sources in information theory: fundamental intervals and word prefixes, *Algorithmica*, *29*, 262–306.

Vanet, A., Marsan, L., and Sagot, M.-F. (1999). Promoter sequences and algorithmical methods for identifying them, *Res. Microbiol.*, *150*(1), 1–21.

Verger Gaugry, J.-L. and Gazeau, J.-P. (2004). Geometric study of the beta-integers for a Perron number and mathematical quasicrystals, to appear in *J. Théor. Nombres Bordeaux*.

Vidal, J. and Mosseri, R. (2000). Generalized Rauzy tilings: construction and electronic properties, *Materials Science and Engineering A*, *294–296*, 572–575.

Vidal, J. and Mosseri, R. (2001). Generalized quasiperiodic Rauzy tilings, *J. Phys. A*, *34*(18), 3927–3938.

Viennot, G. X. (1986). Heaps of pieces. I. Basic definitions and combinatorial lemmas, In *Combinatoire énumérative (Montréal, 1985)*, Vol. 1234 of *Lect. Notes Math.*, pp. 321–350. Springer-Verlag.

Wade, L. I. (1941). Certain quantities transcendental over $GF(p^n, x)$, *Duke Math. J.*, *8*, 701–720.

Waterman, M. S. (1995). *Introduction to Computational Biology. Maps, Sequences and Genomes*. Chapman and Hall.

Weber, A. and Klemm, R. (1995). Economy of description for single-valued transducers, *Inform. and Comput.*, *118*(2), 327–340.

Weiner, P. (1973). Linear pattern matching algorithm, In *14th IEEE Symposium on Switching and Automata Theory (SWAT)*, pp. 1–11, Washington, DC.

Welsh, D. (1988). *Codes and Cryptography*. The Clarendon Press Oxford University Press.

Wharton, R. M. (1974). Approximate language identification, *Inform. and Control*, *26*, 236–255.

Wozny, N. and Zamboni, L. Q. (2001). Frequencies of factors in Arnoux-Rauzy sequences, *Acta Arith.*, *96*(3), 261–278.

Wu, S. and Manber, U. (1995). Fast text searching allowing errors, *Comm. Assoc. Comput. Mach.*, *35*, 983–991.

Wyner, A. and Ziv, J. (1989). Some asymptotic properties of the entropy of a stationary ergodic dat a source with applications to data compression, *IEEE Trans. Inform. Theory*, *35*, 1250–1258.

Wyner, A. J. (1997). The redundancy and distribution of the phrase lengths of the fixed-da tabase Lempel-Ziv algorithm, *IEEE Trans. Inform. Theory*, *43*, 1439–1465.

Yang, E. and Kieffer, J. (1998). On the performance of data compression algorithms based upon string matching, *IEEE Trans. Inform. Theory*, *44*, 47–65.

Zamboni, L. Q. (1998). Une généralisation du théorème de Lagrange sur le développement en fraction continue, *C. R. Acad. Sci. Paris, Série I*, *327*, 527–530.

Zipf, G. K. (1935). *Psycho-Biology of Languages.* Houghton-Mifflin. Republished by MIT Press, 1965.

Ziv, J. and Lempel, A. (1977). A universal algorithm for sequential data compression, *IEEE Trans. Inform. Theory*, *IT-23*, 337–343.

Ziv, J. and Lempel, A. (1978). Compression of individual sequences via variable-rate coding, *IEEE Trans. Inform. Theory*, *24*, 530–536.

Ziv, J. and Merhav, N. (1993). A measure of relative entropy between individual sequences with application to universal classification, *IEEE Trans. Inform. Theory*, *39*, 1270–1279.

General Index

T