Bibliography

[Abney, 1989]
Steven P. Abney. A computational model of human parsing. *Journal of Psycholinguistic Research*, 18:129–144, 1989.

[Abney, 1991]
Steven P. Abney. Parsing by chunks. In Robert C. Berwick, Steven P. Abney, and Carol Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, volume 44 of *Studies in Linguistics and Philosophy*. Kluwer, 1991.

[Abney, 1996a]
Steven Abney. Part-of-speech tagging and partial parsing. In Ken Church, Steve Young, and Gerrit Bloothooft, editors, *Corpus-Based Methods in Language and Speech*. Kluwer Academic Publishers, Dordrecht, 1996.

[Abney, 1996b]
Steven Abney. Statistical methods and linguistics. In Judith Klavans and Philip Resnik, editors, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. The MIT Press, 1996.

[Abney, 2008]
Steven Abney. *Semisupervised Learning for Computational Linguistics*. Chapman and Hall, 2008.

[Agirre and Edmonds, 2007]
Eneko Agirre and Philip Edmonds. *Word Sense Disambiguation: Algorithms and Applications*. Springer, 2007.

[Alpaydin, 2004]
Ethem Alpaydin. *Introduction to Machine Learning*. MIT Press, 2004.

[Ananiadou and McNaught, 2006]
Sophia Ananiadou and John McNaught, editors. *Text Mining for Biology and Biomedicine*. Artech House, 2006.

[Androutsopoulos et al., 1995]
Ion Androutsopoulos, Graeme Ritchie, and Peter Thanisch. Natural language interfaces to databases — an introduction. *Journal of Natural Language Engineering*, 1:29–81, 1995.

[Artstein and Poesio, 2008]
Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. *Computational Linguistics*, pages 555–596, 2008.

[Baayen, 2008]
Harald Baayen. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press, 2008.

[Bachenko and Fitzpatrick, 1990]
J. Bachenko and E. Fitzpatrick. A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16:155–170, 1990.

[Baldwin and Kim, 2010]
Timothy Baldwin and Su Nam Kim. Multiword expressions. In *Handbook of Natural Language Processing*. CRC Press, Taylor and Francis Group, 2 edition, 2010.

[Beazley, 2006]
David M. Beazley. *Python Essential Reference*. Developer's Library. Sams Publishing, third edition edition, 2006.

[Biber et al., 1998]
Douglas Biber, Susan Conrad, and Randi Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press, 1998.

[Bird and Liberman, 2001a]
Steven Bird and Mark Liberman. [A formal framework for linguistic annotation](). *Speech Communication*, 33:23–60, 2001.

[Bird and Liberman, 2001b]
Steven Bird and Mark Liberman. [A formal framework for linguistic annotation](). *Speech Communication*, 33:23–60, 2001.

[Bird and Simons, 2003]
Steven Bird and Gary Simons. Seven dimensions of portability for language documentation and description. *Language*, 79:557–82, 2003.

[Bird, 1999]
Steven Bird. Multidimensional exploration of online linguistic field data. In Pius Tamanji, Masako Hirotani, and Nancy Hall, editors, *Proceedings of the 29th Annual Meeting of the Northeast Linguistics Society*, pages 33–47. GLSA, University of Massachussetts at Amherst, 1999.

[Blackburn and Bos, 2005]
Patrick Blackburn and Johan Bos. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. CSLI Publications, Stanford, Ca, 2005.

[BNC, 1999]
BNC. British National Corpus, 1999. [http://info.ox.ac.uk/bnc/].

[Brent and Cartwright, 1995]

Michael Brent and Timothy Cartwright. Distributional regularity and phonotactic constraints are useful for segmentation. In Michael Brent, editor, *Computational Approaches to Language Acquisition*. MIT Press, 1995.

[Bresnan and Hay, 2006]
Joan Bresnan and Jennifer Hay. Gradient grammar: An effect of animacy on the syntax of *give* in varieties of English. unpublished ms., 2006.

[Budanitsky and Hirst, 2006]
Alexander Budanitsky and Graeme Hirst. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32:13–48, 2006.

[Burton-Roberts, 1997]
Noel Burton-Roberts. *Analysing Sentences*. Longman, 1997.

[Buseman et al., 1996]
Alan Buseman, Karen Buseman, and Rod Early. *The Linguist's Shoebox: Integrated Data Management and Analysis for the Field Lingui st*. Waxhaw NC: SIL, 1996.

[Carpenter, 1992]
Bob Carpenter. *The Logic of Typed Feature Structures*. Cambridge University Press, Cambridge, England, 1992.

[Carpenter, 1997]
Bob Carpenter. *Type-Logical Semantics*. The MIT Press, 1997.

[Chierchia and McConnell-Ginet, 1990]
Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar: An Introduction to Meaning*. The MIT Press, Cambridge, Mass, 1990.

[Chomsky and Halle, 1968]
Noam Chomsky and Morris Halle. *The Sound Pattern of English*. New York: Harper and Row, 1968.

[Chomsky, 1965]
Noam Chomsky. *Aspects of the Theory of Syntax*. The MIT Press, Cambridge, MA, 1965.

[Chomsky, 1970]
Noam Chomsky. Remarks on nominalization. In R. Jacobs and P. Rosenbaum, editors, *Readings in English Transformational Grammar*. Blaisdell, Waltham, MA, 1970.

[Church and Patil, 1982]

Kenneth Church and Ramesh Patil. Coping with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics*, 8:139–149, 1982.

[Cohen and Hunter, 2004]
K. Bretonnel Cohen and Lawrence Hunter. Natural language processing and systems biology. In Werner Dubitzky and Francisco Azuaje, editors, *Artificial Intelligence Methods and Tools for Systems Biology*, page ??? Springer Verlag, 2004.

[Cole, 1997]
Ronald Cole, editor. *Survey of the State of the Art in Human Language Technology*. Studies in Natural Language Processing. Cambridge University Press, 1997.

[Copestake, 2002]
Ann Copestake. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA, 2002.

[Corbett, 2006]
Greville G Corbett. *Agreement*. Cambridge University Press, Cambridge, UK, 2006.

[Croft et al., 2009]
Bruce Croft, Donald Metzler, and Trevor Strohman. *Search Engines: Information Retrieval in Practice*. Addison Wesley, 2009.

[Daelemans and van den Bosch, 2005]
Walter Daelemans and Antal van den Bosch. *Memory-Based Language Processing*. Cambridge University Press, 2005.

[Dagan et al., 2006]
Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In J. Quinonero-Candela, I. Dagan, B. Magnini, and F. d'Alché Buc, editors, *Machine Learning Challenges*, volume 3944 of *Lecture Notes in Computer Science*, pages 177–190. Springer, 2006.

[Dale et al., 2000]
Robert Dale, Hermann Moisl, and Harold Somers, editors. *Handbook of Natural Language Processing*. Marcel Dekker, 2000.

[Dalrymple et al., 1999]
Mary Dalrymple, V. Gupta, John Lamping, and V. Saraswat. Relating resource-based semantics to categorial semantics. In Mary Dalrymple, editor, *Semantics and syntax in Lexical Functional Grammar: the resource logic approach*, pages 261–280. MIT Press, Cambridge, MA, 1999.

[Dalrymple, 2001]
Mary Dalrymple. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. Academic Press, New York, 2001.

[Dowty et al., 1981]

David R. Dowty, Robert E. Wall, and Stanley Peters. *Introduction to Montague Semantics*. Kluwer Academic Publishers, 1981.

[Earley, 1970]
Jay Earley. An efficient context-free parsing algorithm. *Communications of the Association for Computing Machinery*, 13:94–102, 1970.

[Emele and Zajac, 1990]
Martin C. Emele and Rémi Zajac. Typed unification grammars. In *Proceedings of the 13th Conference on Computational linguistics*, pages 293–298, Morristown, NJ, USA, 1990. Association for Computational Linguistics.

[Farghaly, 2003]
Ali Farghaly, editor. *Handbook for Language Engineers*. CSLI Publications, 2003.

[Feldman and Sanger, 2007]
Ronen Feldman and James Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[Fellbaum, 1998]
Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998. http://wordnet.princeton.edu/.

[Finegan, 2007]
Edward Finegan. *Language: Its Structure and Use*. Wadsworth, 5 edition, 2007.

[Forsyth and Martell, 2007]
Eric N. Forsyth and Craig H. Martell. Lexical and discourse analysis of online chat dialog. In *Proceedings of the First IEEE International Conference on Semantic Computing*, pages 19–26, 2007.

[Friedl, 2002]
Jeffrey E. F. Friedl. *Mastering Regular Expressions*. O'Reilly, second edition edition, 2002.

[Gamut, 1991a]
L. T. F. Gamut. *Intensional Logic and Logical Grammar*, volume Volume 2 of *Logic, Language and Meaning*. University of Chicago Press, Chicago, 1991.

[Gamut, 1991b]
L. T. F. Gamut. *Introduction to Logic*, volume Volume 1 of *Logic, Language and Meaning*. University of Chicago Press, Chicago, 1991.

[Garofolo et al., 1986]

John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathon G. Fiscus, David S. Pallett, and Nancy L. Dahlgren. *The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM*. NIST, 1986.

[Gazdar et al., 1985]
Gerald Gazdar, Ewan Klein, Geoffrey Pullum, and Ivan Sag (1985). *Generalized Phrase Structure Grammar*. Basil Blackwell, 1985.

[Gomes et al., 2006]
Bruce Gomes, William Hayes, and Raf Podowski. Text mining. In Darryl Leon and Scott Markel, editors, *In Silico Technologies in Drug Target Identification and Validation*, chapter Chapter 6, page ??? Taylor & Francis Ltd, 2006.

[Gries, 2009]
Stefan Gries. *Quantitative Corpus Linguistics with R: A Practical Introduction*. Routledge, 2009.

[Guzdial, 2005]
Mark Guzdial. *Introduction to Computing and Programming in Python: A Multimedia Approach*. Prentice Hall, 2005.

[Harel, 2004]
David Harel. *Algorithmics: The Spirit of Computing*. Addison Wesley, 2004.

[Hastie et al., 2009]
Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2 edition, 2009.

[Hearst, 1992]
Marti Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics (COLING)*, pages 539–545, 1992.

[Heim and Kratzer, 1998]
Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, 1998.

[Hirschman et al., 2005]
Lynette Hirschman, Alexander Yeh, Christian Blaschke, and Alfonso Valencia. Overview of biocreative: critical assessment of information extraction for biology. *BMC Bioinformatics*, 6, May 2005. Supplement 1.

[Hodges, 1977]
Wilfred Hodges. *Logic*. Penguin Books, Harmondsworth, 1977.

[Huddleston and Pullum, 2002]

Rodney D. Huddleston and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002.

[Hunt and Thomas, 2000]
Andrew Hunt and David Thomas. *The Pragmatic Programmer: From Journeyman to Master*. Addison Wesley, 2000.

[Indurkhya and Damerau, 2010]
Nitin Indurkhya and Fred Damerau, editors. *Handbook of Natural Language Processing*. CRC Press, Taylor and Francis Group, 2 edition, 2010.

[Jackendoff, 1977]
Ray Jackendoff. *X-Syntax: a Study of Phrase Strucure*. Number 2 in Linguistic Inquiry Monograph. The MIT Press, Cambridge, MA, 1977.

[Johnson, 1988]
Mark Johnson. *Attribute Value Logic and Theory of Grammar*. CSLI Lecture Notes Series. Chicago University Press, 1988.

[Jurafsky and Martin, 2008]
Daniel Jurafsky and James H. Martin. *Speech and Langugage Processing*. Prentice Hall, New Jersey, second edition edition, 2008.

[Kamp and Reyle, 1993]
Hans Kamp and Uwe Reyle. *From Discourse to the Lexicon: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers, 1993.

[Kaplan and Bresnan, 1982]
Ronald Kaplan and Joan Bresnan. Lexical-functional grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 173–281. The MIT Press, Cambridge, Mass., 1982.

[Kaplan, 1989]
Ronald Kaplan. The formal architecture of Lexical-Functional Grammar. In Chu-Ren Huang and Keh-Jiann Chen, editors, *Proceedings of ROCLING II*, pages 1–18. CSLI, 1989. Reprinted in Dalrymple, Kaplan, Maxwell, and Zaenen (eds), Formal Issues in Lexical-Functional Grammar, 7-27. Stanford: Center for the Study of Language and Information. 1995.

[Kasper and Rounds, 1986]
Robert T. Kasper and William C. Rounds. A logical semantics for feature structures. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, pages 257–266, Morristown, NJ, USA, 1986. Association for Computational Linguistics.

[Kathol, 1999]
Andreas Kathol. Agreement and the syntax-morphology interface in HPSG. In Robert D. Levine and Georgia M. Green, editors, *Studies in Contemporary Phrase Structure Grammar*, pages 223–274. Cambridge University Press, Cambridge, 1999.

[Kay, 1985]
Martin Kay. Unification in grammar. In Verónica Dahl and Patrick Saint-Dizier, editors, *Natural Language Understanding and Logic Programming,* pages 233–240. North-Holland, 1985. Proceedings of the First International Workshop on Natural Language Understanding and Logic Programming.

[Kiss and Strunk, 2006]
Tibor Kiss and Jan Strunk. Unsupervised multilingual sentence boundary detection. In *Computational Linguistics*, volume 32, pages 485–525, 2006.

[Kiusalaas, 2005]
Jaan Kiusalaas. *Numerical Methods in Engineering with Python*. Cambridge University Press, 2005.

[Klein and Manning, 2003]
Dan Klein and Christopher D. Manning. A* parsing: Fast exact viterbi parse selection. In *Proceedings of HLT-NAACL 03*, 2003.

[Knuth, 2006]
Donald E. Knuth. *The Art of Computer Programming, Volume 4: Generating All Trees*. Addison Wesley, 2006.

[Lappin, 1996]
Shalom Lappin, editor. *The Handbook of Contemporary Semantic Theory*. Blackwell Publishers, Oxford, 1996.

[Larson and Segal, 1995]
Richard Larson and Gabriel Segal. *Knowledge of Meaning: An Introduction to Semantic Theory*. The MIT Press, Cambridge, Mass, 1995.

[Levin, 1993]
Beth Levin. *English Verb Classes and Alternations*. Chicago, 1993.

[Levitin, 2004]
Anany Levitin. *The Design and Analysis of Algorithms*. Addison Wesley, 2004.

[Lutz and Ascher, 2003]
Mark Lutz and David Ascher. *Learning Python*. O'Reilly, second edition edition, 2003.

[MacWhinney, 1995]
Brian MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum., second edition, 1995. [childes.psy.cmu.edu/].

[Madnani, 2007]
Nitin Madnani. Getting started on natural language processing with python. *ACM Crossroads*, 13(4), 2007.

[Manning and Schutze, 1999]
Christopher Manning and Hinrich Schutze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, 1999.

[Manning et al., 2008]
Christopher Manning, Prabhakar Raghavan, and Hinrich Schutze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[Manning, 2003]
Christopher Manning. Probabilistic syntax. In *Probabilistic Linguistics*, pages 289–341. MIT Press, Cambridge, MA, 2003.

[McCawley, 1998]
James McCawley. *The Syntactic Phenomena of English*. Chicago University Press, 1998.

[McConnell, 2004]
Steve McConnell. *Code Complete: A Practical Handbook of Software Construction*. Microsoft Press, 2004.

[McCune, 2008]
William McCune. Prover9: Automated theorem prover for first-order and equational logic, 2008.

[McEnery, 2006]
Anthony McEnery. *Corpus-Based Language Studies: An Advanced Resource Book*. Routledge, 2006.

[Melamed, 2001]
Dan Melamed. *Empirical Methods for Exploiting Parallel Texts*. MIT Press, 2001.

[Mertz, 2003]
David Mertz. *Text Processing in Python*. Addison-Wesley, Boston, 2003.

[Meyer, 2002]
Charles Meyer. *English Corpus Linguistics: An Introduction*. Cambridge University Press, 2002.

[Miller and Charles, 1998]
George Miller and Walter Charles. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6:1–28, 1998.

[Mitkov, 2002a]
Ruslan Mitkov. *Anaphora Resolution*. Longman, 2002.

[Mitkov, 2002b]
Ruslan Mitkov, editor. *Oxford Handbook of Computational Linguistics*. Oxford University Press, 2002.

[Müller, 2002]
Stefan Müller. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Number 13 in Studies in Constraint-Based Lexicalism. Center for the Study of Language and Information, Stanford, 2002. http://www.dfki.de/ stefan/Pub/complex.html.

[Nerbonne et al., 1994]
John Nerbonne, Klaus Netter, and Carl Pollard. *German in Head-Driven Phrase Structure Grammar*. CSLI, Stanford, CA, 1994.

[Nespor and Vogel, 1986]
Marina Nespor and Irene Vogel. *Prosodic Phonology*. Number 28 in Studies in Generative Grammar. Foris Publications, Dordrecht, 1986.

[Nivre et al., 2006]
J. Nivre, J. Hall, and J. Nilsson. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of LREC*, pages 2216–2219, 2006.

[Niyogi, 2006]
Partha Niyogi. *The Computational Nature of Language Learning and Evolution*. MIT Press, 2006.

[O'Grady et al., 2004]
William O'Grady, John Archibald, Mark Aronoff, and Janie Rees-Miller. *Contemporary Linguistics: An Introduction*. St. Martin's Press, 5 edition, 2004.

[OSU, 2007]
OSU, editor. *Language Files: Materials for an Introduction to Language and Linguistics*. Ohio State University Press, 10 edition, 2007.

[Partee, 1995]
Barbara Partee. Lexical semantics and compositionality. In L. R. Gleitman and M. Liberman, editors, *An Invitation to Cognitive Science: Language*, volume Volume 1, pages 311–360. The MIT Press, 1995.

[Pasca, 2003]
Marius Pasca. *Open-Domain Question Answering from Large Text Collections*. CSLI, 2003.

[Pevzner and Hearst, 2002]
L. Pevzner and M. Hearst. A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28:19–36, 2002.

[Pullum, 2005]
Geoffrey K. Pullum. Fossilized prejudices about "however", 2005.

[Radford, 1988]
Andrew Radford. *Transformational Grammar: An Introduction*. Cambridge University Press, Cambridge, 1988.

[Ramshaw and Marcus, 1995]
Lance A. Ramshaw and Mitchell P. Marcus. Text chunking using transformation-based learning. In *Proceedings of the Third ACL Workshop on Very Large Corpora*, pages 82–94, 1995.

[Reppen et al., 2005]

Randi Reppen, Nancy Ide, and Keith Suderman. *American National Corpus*. Linguistic Data Consortium, 2005.

[Robinson et al., 2007]
Stuart Robinson, Greg Aumann, and Steven Bird. Managing fieldwork data with toolbox and the natural language toolkit. *Language Documentation and Conservation*, 1:44–57, 2007.

[Sag and Wasow, 1999]
Ivan A. Sag and Thomas Wasow. *Syntactic Theory: A Formal Introduction*. CSLI Publications, 1999.

[Sampson and McCarthy, 2005]
Geoffrey Sampson and Diana McCarthy. *Corpus Linguistics: Readings in a Widening Discipline*. Continuum, 2005.

[Scott and Tribble, 2006]
Mike Scott and Christopher Tribble. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. John Benjamins, 2006.

[Segaran, 2007]
Toby Segaran. *Collective Intelligence*. O'Reilly Media, 2007.

[Shatkay and Feldman, 2004]
Hagit Shatkay and Ronan Feldman. Mining the biomedical literature in the genomic era: An overview. *Journal of Computational Biology*, 10:821–855, 2004.

[Shieber et al., 1983]
Stuart Shieber, Hans Uszkoreit, Fernando Pereira, Jane Robinson, and Mabry Tyson. The formalism and implementation of PATR-II. In Barbara J. Grosz and Mark Stickel, editors, *Research on Interactive Acquisition and Use of Knowledge*, pages 39–79. SRI International, Menlo Park, CA, 1983. (PDF)

[Shieber, 1986]
Stuart M. Shieber. *An Introduction to Unification-Based Approaches to Grammar*, volume 4 of *CSLI Lecture Notes Series*. Center for the Study of Language and Information, Stanford, CA, 1986.

[Simons and Bird, 2003]
Gary Simons and Steven Bird. The Open Language Archives Community: An infrastructure for distributed archiving of language resources. *Literary and Linguistic Computing*, 18:117–128, 2003.

[Sproat et al., 2001]
Richard Sproat, Alan Black, Stanley Chen, Shankar Kumar, Mari Ostendorf, and Christopher Richards. Normalization of non-standard words. *Computer Speech and Language*, 15:287–333, 2001.

[Strunk and White, 1999]
William Strunk and E. B. White. *The elements of style*. Boston: Allyn and Bacon, 1999.

[Thompson and McKelvie, 1997]
Henry S. Thompson and David McKelvie. Hyperlink semantics for standoff markup of read-only documents. In *SGML Europe '97*, 1997. http://www.ltg.ed.ac.uk/ ht/sgmleu97.html.

[TLG, 1999]
TLG. Thesaurus Linguae Graecae, 1999.

[Turing, 1950]
Alan M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[van Benthem and ter Meulen, 1997]
Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. The MIT Press, Cambridge, Mass, 1997.

[van Rossum and Drake, 2006a]
Guido van Rossum and Fred L. Drake. *An Introduction to Python — The Python Tutorial*. Network Theory Ltd, Bristol, 2006.

[van Rossum and Drake, 2006b]
Guido van Rossum and Fred L. Drake. *The Python Language Reference Manual*. Network Theory Ltd, Bristol, 2006.

[Warren and Pereira, 1982]
David H. D. Warren and Fernando C. N. Pereira. An efficient easily adaptable system for interpreting natural language queries. *AJCL*, 8(3-4):110–122, 1982.

[Wechsler and Zlatic, 2003]
Stephen Mark Wechsler and Larisa Zlatic. *The Many Faces of Agreement*. Stanford Monographs in Linguistics. Center for the Study of Language and Information, Stanford, 2003.

[Weiss et al., 2004]
Sholom Weiss, Nitin Indurkhya, Tong Zhang, and Fred Damerau. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer, 2004.

[Woods et al., 1986]
Anthony Woods, Paul Fletcher, and Arthur Hughes. *Statistics in Language Studies*. Cambridge University Press, 1986.

[Zhao and Zobel, 2007]

Y. Zhao and J. Zobel. Search with style: authorship attribution in classic literature. In *Proceedings of the Thirtieth Australasian Computer Science Conference*. Association for Computing Machinery, 2007.