

This excerpt from

Foundations of Statistical Natural Language Processing.
Christopher D. Manning and Hinrich Schütze.
© 1999 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact cognetadmin@cognet.mit.edu.

Bibliography

The following conference abbreviations are used in this bibliography:

ACL *n* Proceedings of the *n*th Annual Meeting of the Association for Computational Linguistics

ANLP *n* Proceedings of the *n*th conference on Applied Natural Language Processing

COLING *n* Proceedings of the *n*th International Conference on Computational Linguistics (COLING-*year*)

EACL *n* Proceedings of the *n*th Conference of the European Chapter of the Association for Computational Linguistics

EMNLP *n* Proceedings of the *n*th Conference on Empirical Methods in Natural Language Processing

WVLC *n* Proceedings of the *n*th Workshop on Very Large Corpora

These conference proceedings are all available from the Association for Computational Linguistics, P.O. Box 6090, Somerset NJ 08875, USA, acl@aclweb.org, <http://www.aclweb.org>.

SIGIR 'y Proceedings of the (y - 77)th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval. Available from the Association for Computing Machinery, acmhelp@acm.org, <http://www.acm.org>.

Many papers are also available from the Computation and Language subject area of the Computing Research Repository e-print archive, a part of the xxx.lanl.gov e-print archive on the World Wide Web.

Abney, Steven. 1991. Parsing by chunks. In Robert C. Berwick, Steven P. Abney, and Carol Tenny (eds.), *Principle-Based Parsing*, pp. 257-278. Dordrecht: Kluwer Academic.

- Abney, Steven. 1996a. Part-of-speech tagging and partial parsing. In Steve Young and Gerrit Bloothoof (eds.), *Corpus-Based Methods in Language and Speech Processing*, pp. 118-136. Dordrecht: Kluwer Academic.
- Abney, Steven. 1996b. Statistical methods and linguistics. In Judith L. Klavans and Philip Resnik (eds.), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, pp. 1-26. Cambridge, MA: MIT Press.
- Abney, Steven P. 1997. Stochastic attribute-value grammars. *Computational Linguistics* 23:597-618.
- Ackley, D. H., G. E. Hinton, and T. J. Sejnowski. 1985. A learning algorithm for Boltzmann machines. *Cognitive Science* 9:147-169.
- Aho, Alfred V., Ravi Sethi, and Jeffrey D. Ullman. 1986. *Compilers: Principles, Techniques, and Tools*. Reading, MA: Addison-Wesley.
- Allen, James. 1995. *Natural Language Understanding*. Redwood City, CA: Benjamin Cummings.
- Alshawi, Hiyan, Adam L. Buchsbaum, and Fei Xia. 1997. A comparison of head transducers and transfer for a limited domain translation application. In *ACL 35/EACL 8*, pp. 360-365.
- Alshawi, Hiyan, and David Carter. 1994. Training and scaling preference functions for disambiguation. *Computational Linguistics* 20:635-648.
- Anderson, John R. 1983. *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, John R. 1990. *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.
- Aone, Chinatsu, and Douglas McKee. 1995. Acquiring predicate-argument mapping information from multilingual texts. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 175-190. Cambridge, MA: MIT Press.
- Appelt, D. E., J. R. Hobbs, J. Bear, D. Israel, and M. Tyson. 1993. Fastus: A finite-state processor for information extraction from real-world text. In *Proc. of the 13th IJCAI*, pp. 1172-1178, Chambéry, France.
- Apresjan, Jurij D. 1974. Regular polysemy. *Linguistics* 142:5-32.
- Apté, Chidanand, Fred Damerau, and Sholom M. Weiss. 1994. Automated learning of decision rules for text categorization. *ACM Transactions on Information Systems* 12:233-251.
- Argamon, Shlomo, Ido Dagan, and Yuval Krymolowski. 1998. A memory-based approach to learning shallow natural language patterns. In *ACL 36/COLING 17*, pp. 67-73.

- Atwell, Eric. 1987. Constituent-likelihood grammar. In Roger Garside, Geoffrey Leech, and Geoffrey Sampson (eds.), *The Computational Analysis of English: A Corpus-Based Approach*. London: Longman.
- Baayen, Harald, and Richard Sproat. 1996. Estimating lexical priors for low-frequency morphologically ambiguous forms. *Computational Linguistics* 22: 155-166.
- Bahl, Lalit R., Frederick Jelinek, and Robert L. Mercer. 1983. A maximum likelihood approach to continuous speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-5:179-190. Reprinted in (Waibel and Lee 1990), pp. 308-319.
- Bahl, Lalit R., and Robert L. Mercer. 1976. Part-of-speech assignment by a statistical decision algorithm. In *International Symposium on Information Theory*, Ronneby, Sweden.
- Baker, James K. 1975. Stochastic modeling for automatic speech understanding. In D. Raj Reddy (ed.), *Speech Recognition: Invited papers presented at the 1974 IEEE symposium*, pp. 521-541. New York: Academic Press. Reprinted in (Waibel and Lee 1990), pp. 297-307.
- Baker, James K. 1979. Trainable grammars for speech recognition. In D. H. Klatt and J. J. Wolf (eds.), *Speech Communication Papers for the 97th Meeting of the Acoustical Society of America*, pp. 547-550.
- Baldi, Pierre, and Søren Brunak. 1998. *Bioinformatics: The Machine Learning Approach*. Cambridge, MA: MIT Press.
- Barnbrook, Geoff. 1996. *Language and computers: a practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Basili, Roberto, Maria Teresa Pazienza, and Paola Velardi. 1996. Integrating general-purpose and corpus-based verb classification. *Computational Linguistics* 22:559-568.
- Basili, Roberto, Gianluca De Rossi, and Maria Teresa Pazienza. 1997. Inducing terminology for lexical acquisition. In *EMNLP 2*, pp. 125-133.
- Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* 41:164-171.
- Beeferman, Doug, Adam Berger, and John Lafferty. 1997. Text segmentation using exponential models. In *EMNLP 2*, pp. 35-46.
- Bell, Timothy C., John G. Cleary, and Ian H. Witten. 1990. *Text Compression*. Englewood Cliffs, NJ: Prentice Hall.
- Benello, Julian, Andrew W. Mackie, and James A. Anderson. 1989. Syntactic category disambiguation with neural networks. *Computer Speech and Language* 3:203-217.

- Benson, Morton. 1989. The structure of the collocational dictionary. *International Journal of Lexicography* 2:1-14.
- Benson, Morton, Evelyn Benson, and Robert Ilson. 1993. *The BBI combinatory dictionary of English*. Amsterdam: John Benjamins.
- Berber Sardinha, A. P. 1997. *Automatic Identification of Segments in Written Texts*. PhD thesis, University of Liverpool.
- Berger, Adam L., Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22:39-71.
- Berry, Michael W. 1992. Large-scale sparse singular value computations. *The International Journal of Supercomputer Applications* 6:13-49.
- Berry, Michael W., Susan T. Dumais, and Gavin W. O'Brien. 1995. Using linear algebra for intelligent information retrieval. *SIAM Review* 37:573-595.
- Berry, Michael W., and Paul G. Young. 1995. Using latent semantic indexing for multilanguage information retrieval. *Computers and the Humanities* 29: 413-429.
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In J. R. Hayes (ed.), *Cognition and the development of language*. New York: Wiley.
- Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8:243-257.
- Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Black, Ezra. 1988. An experiment in computational discrimination of English word senses. *IBM Journal of Research and Development* 32:185-194.
- Black, E., S. Abney, D. Flickinger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini, and T. Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proceedings, Speech and Natural Language Workshop*, pp. 306-311, Pacific Grove, CA. DARPA.
- Black, Ezra, Fred Jelinek, John Lafferty, David M. Magerman, Robert Mercer, and Salim Roukos. 1993. Towards history-based grammars: Using richer models for probabilistic parsing. In *ACL 31*, pp. 31-37. Also appears in the Proceedings of the DARPA Speech and Natural Language Workshop, Feb. 1992, pp. 134-139.
- Bod, Rens. 1995. *Enriching Linguistics with Statistics: Performance Models of Natural Language*. PhD thesis, University of Amsterdam.

- Bod, Rens. 1996. Data-oriented language processing: An overview. Technical Report LP-96-13, Institute for Logic, Language and Computation, University of Amsterdam.
- Bod, Rens. 1998. *Beyond Grammar: An experience-based theory of language*. Stanford, CA: CSLI Publications.
- Bod, Rens, and Ronald Kaplan. 1998. A probabilistic corpus-driven model for lexical-functional analysis. In *ACL 36/COLING 17*, pp. 145–151.
- Bod, Rens, Ron Kaplan, Remko Scha, and Khalil Sima'an. 1996. A data-oriented approach to lexical-functional grammar. In *Computational Linguistics in the Netherlands 1996*, Eindhoven, The Netherlands.
- Boguraev, Bran, and Ted Briscoe. 1989. *Computational Lexicography for Natural Language Processing*. London: Longman.
- Boguraev, Branimir, and James Pustejovsky. 1995. Issues in text-based lexicon acquisition. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 3–17. Cambridge MA: MIT Press.
- Boguraev, Branimir K. 1993. The contribution of computational lexicography. In Madeleine Bates and Ralph M. Weischedel (eds.), *Challenges in natural language processing*, pp. 99–132. Cambridge: Cambridge University Press.
- Bonnema, R. 1996. Data-oriented semantics. Master's thesis, Department of Computational Linguistics, University of Amsterdam.
- Bonnema, Remko, Rens Bod, and Remko Scha. 1997. A DOP model for semantic interpretation. In *ACL 35/EACL 8*, pp. 159–167.
- Bonzi, Susan, and Elizabeth D. Liddy. 1988. The use of anaphoric resolution for document description in information retrieval. In *SIGIR '88*, pp. 53–66.
- Bookstein, Abraham, and Don R. Swanson. 1975. A decision theoretic foundation for indexing. *Journal of the American Society for Information Science* 26:45–50.
- Booth, Taylor L. 1969. Probabilistic representation of formal languages. In *Tenth Annual IEEE Symposium on Switching and Automata Theory*, pp. 74–81.
- Booth, Taylor L., and Richard A. Thomson. 1973. Applying probability measures to abstract languages. *IEEE Transactions on Computers* C-22:442–450.
- Borthwick, Andrew, John Sterling, Eugene Agichtein, and Ralph Grishman. 1998. Exploiting diverse knowledge sources via maximum entropy in named entity recognition. In *WVLC 6*, pp. 152–160.
- Bourigault, Didier. 1993. An endogeneous corpus-based method for structural noun phrase disambiguation. In *EACL 6*, pp. 81–86.
- Box, George E. P., and George C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

- Brants, Thorsten. 1998. Estimating Hidden Markov Model Topologies. In Jonathan Ginzburg, Zurab Khasidashvili, Carl Vogel, Jean-Jacques Lévy, and Enric Vallduví (eds.), *The Tbilisi Symposium on Logic, Language and Computation: Selected Papers*, pp. 163–176. Stanford, CA: CSLI Publications.
- Brants, Thorsten, and Wojciech Skut. 1998. Automation of treebank annotation. In *Proceedings of NeMLaP-98*, Sydney, Australia.
- Breiman, Leo. 1994. Bagging predictors. Technical Report 421, Department of Statistics, University of California at Berkeley.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.
- Brent, Michael R. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics* 19:243–262.
- Brew, Chris. 1995. Stochastic HPSG. In *EACL 7*, pp. 83–89.
- Brill, Eric. 1993a. Automatic grammar induction and parsing free text: A transformation-based approach. In *ACL 31*, pp. 259–265.
- Brill, Eric. 1993b. *A Corpus-Based Approach to Language Learning*. PhD thesis, University of Pennsylvania.
- Brill, Eric. 1993c. Transformation-based error-driven parsing. In *Proceedings Third International Workshop on Parsing Technologies*, Tilburg/Durbuy, The Netherlands/Belgium.
- Brill, Eric. 1995a. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics* 21:543–565.
- Brill, Eric. 1995b. Unsupervised learning of disambiguation rules for part of speech tagging. In *WVLC 3*, pp. 1–13.
- Brill, Eric, David Magerman, Mitch Marcus, and Beatrice Santorini. 1990. Deducing linguistic structure from the statistics of large corpora. In *Proceedings of the DARPA Speech and Natural Language Workshop*, pp. 275–282, San Mateo CA. Morgan Kaufmann.
- Brill, Eric, and Philip Resnik. 1994. A transformation-based approach to prepositional phrase attachment disambiguation. In *COLING 15*, pp. 1198–1204.
- Briscoe, Ted, and John Carroll. 1993. Generalized probabilistic LR parsing of natural language (corpora) with unification-based methods. *Computational Linguistics* 19:25–59.
- Britton, J. L. (ed.). 1992. *Collected Works of A. M. Turing: Pure Mathematics*. Amsterdam: North-Holland.

- Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics* 16: 79-85.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, John D. Lafferty, and Robert L. Mercer. 1992a. Analysis, statistical transfer, and synthesis in machine translation. In *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, pp. 83-100.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, Jennifer C. Lai, and Robert L. Mercer. 1992b. An estimate of an upper bound for the entropy of English. *Computational Linguistics* 18:31-40.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1991a. A statistical approach to sense disambiguation in machine translation. In *Proceedings of the DARPA Workshop on Speech and Natural Language Workshop*, pp. 146-151.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1991b. Word-sense disambiguation using statistical methods. In *ACL 29*, pp. 264-270.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics* 19:263-311.
- Brown, Peter F., Vincent J. Della Pietra, Peter V. deSouza, Jennifer C. Lai, and Robert L. Mercer. 1992c. Class-based n -gram models of natural language. *Computational Linguistics* 18:467-479.
- Brown, Peter F., Jennifer C. Lai, and Robert L. Mercer. 1991c. Aligning sentences in parallel corpora. In *ACL 29*, pp. 169-176.
- Bruce, Rebecca, and Janyce Wiebe. 1994. Word-sense disambiguation using decomposable models. In *ACL 32*, pp. 139-145.
- Bruce, Rebecca F., and Janyce M. Wiebe. 1999. Decomposable modeling in natural language processing. *Computational Linguistics*. to appear.
- Brundage, Jennifer, Maren Kresse, Ulrike Schwall, and Angelika Storrer. 1992. Multiword lexemes: A monolingual and contrastive typology for natural language processing and machine translation. Technical Report 232, Institut fuer Wissensbasierte Systeme, IBM Deutschland GmbH, Heidelberg.
- Buckley, Chris, Amit Singhal, Mandar Mitra, and Gerard Salton. 1996. New retrieval approaches using SMART: TREC 4. In D. K. Harman (ed.), *The Second Text REtrieval Conference (TREC-2)*, pp. 25-48.
- Buitelaar, Paul. 1998. *CoreLex: Systematic Polysemy and Underspecification*. PhD thesis, Brandeis University.

- Burgess, Curt, and Kevin Lund. 1997. Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes* 12:177-210.
- Burke, Robin, Kristian Hammond, Vladimir Kulyukin, Steven Lytinen, Noriko Tomuro, and Scott Schoenberg. 1997. Question answering from frequently asked question files. *AI Magazine* 18:57-66.
- Caraballo, Sharon A., and Eugene Charniak. 1998. New figures of merit for best-first probabilistic chart parsing. *Computational Linguistics* 24:275-298.
- Cardie, Claire. 1997. Empirical methods in information extraction. *AI Magazine* 18:65-79.
- Carletta, Jean. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics* 22:249-254.
- Carrasco, Rafael C., and Jose Oncina (eds.). 1994. *Grammatical inference and applications: second international colloquium, ICGI-94*. Berlin: Springer-Verlag.
- Carroll, G., and E. Charniak. 1992. Two experiments on learning probabilistic dependency grammars from corpora. In Carl Weir, Stephen Abney, Ralph Grishman, and Ralph Weischedel (eds.), *Working Notes of the Workshop Statistically-Based NLP Techniques*, pp. 1-13. Menlo Park, CA: AAAI Press.
- Carroll, John. 1994. Relating complexity to practical performance in parsing with wide-coverage unification grammars. In *ACL 32*, pp. 287-294.
- Chang, Jason S., and Mathis H. Chen. 1997. An alignment method for noisy parallel corpora based on image processing techniques. In *ACL 35/EACL 8*, pp. 297-304.
- Chanod, Jean-Pierre, and Pasi Tapanainen. 1995. Tagging French - comparing a statistical and a constraint-based method. In *EACL 7*, pp. 149-156.
- Charniak, Eugene. 1993. *Statistical Language Learning*. Cambridge, MA: MIT Press.
- Charniak, Eugene. 1996. Tree-bank grammars. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI '96)*, pp. 1031-1036.
- Charniak, Eugene. 1997a. Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI '97)*, pp. 598-603.
- Charniak, Eugene. 1997b. Statistical techniques for natural language parsing. *AI Magazine* pp. 33-43.
- Charniak, Eugene, Curtis Hendrickson, Neil Jacobson, and Mike Perkowitz. 1993. Equations for part-of-speech tagging. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pp. 784-789, Menlo Park, CA.

- Cheeseman, Peter, James Kelly, Matthew Self, John Stutz, Will Taylor, and Don Freeman. 1988. AutoClass: A Bayesian classification system. In *Proceedings of the Fifth International Conference on Machine Learning*, pp. 54-64, San Francisco, CA: Morgan Kaufmann.
- Chelba, Ciprian, and Frederick Jelinek. 1998. Exploiting syntactic structure for language modeling. In *ACL 36/COLING 17*, pp. 225-231.
- Chen, Jen Nan, and Jason S. Chang. 1998. Topical clustering of MRD senses based on information retrieval techniques. *Computational Linguistics* 24:61-95.
- Chen, Stanley F. 1993. Aligning sentences in bilingual corpora using lexical information. In *ACL 31*, pp. 9-16.
- Chen, Stanley F. 1995. Bayesian grammar induction for language modeling. In *ACL 33*, pp. 228-235.
- Chen, Stanley F., and Joshua Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *ACL 34*, pp. 310-318.
- Chen, Stanley F., and Joshua Goodman. 1998. An empirical study of smoothing techniques for language modeling. Technical Report TR-10-98, Center for Research in Computing Technology, Harvard University.
- Chi, Zhiyi, and Stuart Geman. 1998. Estimation of probabilistic context-free grammars. *Computational linguistics* 24:299-305.
- Chitrao, Mahesh V., and Ralph Grishman. 1990. Statistical parsing of messages. In *Proceedings of the DARPA Speech and Natural Language Workshop, Hidden Valley, PA*, pp. 263-266. Morgan Kaufmann.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.
- Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Prager.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Choueika, Yaacov. 1988. Looking for needles in a haystack or locating interesting collocational expressions in large textual databases. In *Proceedings of the RIAO*, pp. 43-38.
- Choueika, Yaacov, and Serge Lusignan. 1985. Disambiguation by short contexts. *Computers and the Humanities* 19:147-158.
- Church, Kenneth, William Gale, Patrick Hanks, and Donald Hindle. 1991. Using statistics in lexical analysis. In Uri Zernik (ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pp. 115-164. Hillsdale, NJ: Lawrence Erlbaum.

- Church, Kenneth, and Ramesh Patil. 1982. Coping with syntactic ambiguity or how to put the block in the box on the table. *Computational Linguistics* 8: 139-149.
- Church, Kenneth W. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *ANLP* 2, pp. 136-143.
- Church, Kenneth Ward. 1993. Char_align: A program for aligning parallel texts at the character level. In *ACL* 31, pp. 1-8.
- Church, Kenneth Ward. 1995. One term or two? In *SIGIR '95*, pp. 310-318.
- Church, Kenneth W., and William A. Gale. 1991a. A comparison of the enhanced Good-Turing and deleted estimation methods for estimating probabilities of English bigrams. *Computer Speech and Language* 5:19-54.
- Church, Kenneth W., and William A. Gale. 1991b. Concordances for parallel text. In *Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research*, pp. 40-62, Oxford.
- Church, Kenneth W., and William A. Gale. 1995. Poisson mixtures. *Natural Language Engineering* 1:163-190.
- Church, Kenneth Ward, and Patrick Hanks. 1989. Word association norms, mutual information and lexicography. In *ACL* 27, pp. 76-83.
- Church, Kenneth Ward, and Mark Y. Liberman. 1991. A status report on the ACL/DCI. In *Proceedings of the 7th Annual Conference of the UW Centre for New OED and Text Research: Using Corpora*, pp. 84-91.
- Church, Kenneth W., and Robert L. Mercer. 1993. Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics* 19:1-24.
- Clark, Eve, and Herbert Clark. 1979. When nouns surface as verbs. *Language* 55: 767-811.
- Cleverdon, Cyril W., and J. Mills. 1963. The testing of index language devices. *Aslib Proceedings* 15:106-130. Reprinted in (Sparck Jones and Willett 1998).
- Coates-Stephens, Sam. 1993. The analysis and acquisition of proper names for the understanding of free text. *Computers and the Humanities* 26:441-456.
- Collins, Michael John. 1996. A new statistical parser based on bigram lexical dependencies. In *ACL* 34, pp. 184-191.
- Collins, Michael John. 1997. Three generative, lexicalised models for statistical parsing. In *ACL 35/EACL* 8, pp. 16-23.
- Collins, Michael John, and James Brooks. 1995. Prepositional phrase attachment through a backed-off model. In *WVLC* 3, pp. 27-38.
- Copestake, Ann, and Ted Briscoe. 1995. Semi-productive polysemy and sense extension. *Journal of Semantics* 12:15-68.

- Cormen, Thomas H., Charles E. Leiserson, and Ronald L. Rivest. 1990. *Introduction to Algorithms*. Cambridge, MA: MIT Press.
- Cottrell, Garrison W. 1989. *A Connectionist Approach to Word Sense Disambiguation*. London: Pitman.
- Cover, Thomas M., and Joy A. Thomas. 1991. *Elements of Information Theory*. New York: John Wiley & Sons.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- Croft, W. B., and D. J. Harper. 1979. Using probabilistic models of document retrieval without relevance information. *Journal of Documentation* 35:285-295.
- Crowley, Terry, John Lynch, Jeff Siegel, and Julie Piau. 1995. *The Design of Language: An introduction to descriptive linguistics*. Auckland: Longman Paul.
- Crystal, David. 1987. *The Cambridge Encyclopedia of Language*. Cambridge, England: Cambridge University Press.
- Cutting, Doug, Julian Kupiec, Jan Pedersen, and Penelope Sibun. 1991. A practical part-of-speech tagger. In *ANLP 3*, pp. 133-140.
- Cutting, Douglas R., David R. Karger, and Jan O. Pedersen. 1993. Constant interaction-time scatter/gather browsing of very large document collections. In *SIGIR '93*, pp. 126-134.
- Cutting, Douglas R., Jan O. Pedersen, David Karger, and John W. Tukey. 1992. Scatter/gather: A cluster-based approach to browsing large document collections. In *SIGIR '92*, pp. 318-329.
- Daelemans, Walter, and Antal van den Bosch. 1996. Language-independent data-oriented grapheme-to-phoneme conversion. In J. Van Santen, R. Sproat, J. Olive, and J. Hirschberg (eds.), *Progress in Speech Synthesis*, pp. 77-90. New York: Springer Verlag.
- Daelemans, Walter, Jakub Zavrel, Peter Berck, and Steven Gillis. 1996. MBT: A memory-based part of speech tagger generator. In *WVLC 4*, pp. 14-27.
- Dagan, Ido, Kenneth Church, and William Gale. 1993. Robust bilingual word alignment for machine aided translation. In *WVLC 1*, pp. 1-8.
- Dagan, Ido, and Alon Itai. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics* 20:563-596.
- Dagan, Ido, Alon Itai, and Ulrike Schwall. 1991. Two languages are more informative than one. In *ACL 29*, pp. 130-137.
- Dagan, Ido, Yael Karov, and Dan Roth. 1997a. Mistake-driven learning in text categorization. In *EMNLP 2*, pp. 55-63.

- Dagan, Ido, Lillian Lee, and Fernando Pereira. 1997b. Similarity-based methods for word sense disambiguation. In *ACL 35/EACL 8*, pp. 56-63.
- Dagan, Ido, Fernando Pereira, and Lillian Lee. 1994. Similarity-based estimation of word cooccurrence probabilities. In *ACL 32*, pp. 272-278.
- Damerau, Fred J. 1993. Generating and evaluating domain-oriented multi-word terms from texts. *Information Processing & Management* 29:433-447.
- Darroch, J. N., and D. Ratcliff. 1972. Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics* 43:1470-1480.
- de Saussure, Ferdinand. 1962. *Cours de linguistique générale*. Paris: Payot.
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41:391-407.
- DeGroot, Morris H. 1975. *Probability and Statistics*. Reading, MA: Addison-Wesley.
- Della Pietra, Stephen, Vincent Della Pietra, and John Lafferty. 1997. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.
- Demers, A.J. 1977. Generalized left corner parsing. In *Proceedings of the Fourth Annual ACM Symposium on Principles of Programming Languages*, pp. 170-181.
- Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society Series B* 39: 1-38.
- Dermatas, Evangelos, and George Kokkinakis. 1995. Automatic stochastic tagging of natural language texts. *Computational Linguistics* 21:137-164.
- DeRose, Steven J. 1988. Grammatical category disambiguation by statistical optimization. *Computational Linguistics* 14:31-39.
- Derouault, Anne-Marie, and Bernard Merialdo. 1986. Natural language modeling for phoneme-to-text transcription. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8:742-649.
- Dietterich, Thomas G. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation* 10:1895-1924.
- Dini, Luca, Vittorio Di Tomaso, and Frédérique Segond. 1998. Error-driven word sense disambiguation. In *ACL 36/COLING 17*, pp. 320-324.
- Dolan, William B. 1994. Word sense ambiguity: Clustering related senses. In *COLING 15*, pp. 712-716.

- Dolin, Ron. 1998. *Pharos: A Scalable Distributed Architecture for Locating Heterogeneous Information Sources*. PhD thesis, University of California at Santa Barbara.
- Domingos, Pedro, and Michael Pazzani. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29:103-130.
- Doran, Christy, Dania Egedi, Beth Ann Hockey, B. Srinivas, and Martin Zaidel. 1994. XTAG system - a wide coverage grammar for English. In *COLING 15*, pp. 922-928.
- Dorr, Bonnie J., and Mari Broman Olsen. 1997. Deriving verbal and compositional lexical aspect for nlp applications. In *ACL 35/EACL 8*, pp. 151-158.
- Dras, Mark, and Mike Johnson. 1996. Death and lightness: Using a demographic model to find support verbs. In *Proceedings of the 5th International Conference on the Cognitive Science of Natural Language Processing*, Dublin.
- Duda, Richard O., and Peter E. Hart. 1973. *Pattern classification and scene analysis*. New York: Wiley.
- Dumais, Susan T. 1995. Latent semantic indexing (LSI): TREC-3 report. In *The Third Text REtrieval Conference (TREC 3)*, pp. 219-230.
- Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19:61-74.
- Dunning, Ted. 1994. Statistical identification of language. Technical report, Computing Research Laboratory, New Mexico State University.
- Durbin, Richard, Sean Eddy, Anders Krogh, and Graeme Mitchison. 1998. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge: Cambridge University Press.
- Eeg-Olofsson, Mats. 1985. A probability model for computer-aided word class determination. *Literary and Linguistic Computing* 5:25-30.
- Egan, Dennis E., Joel R. Remde, Louis M. Gomez, Thomas K. Landauer, Jennifer Eberhardt, and Carol C. Lochbaum. 1989. Formative design-evaluation of superbook. *ACM Transactions on Information Systems* 7:30-57.
- Eisner, Jason. 1996. Three new probabilistic models for dependency parsing: An exploration. In *COLING 16*, pp. 340-345.
- Ellis, C. A. 1969. *Probabilistic Languages and Automata*. PhD thesis, University of Illinois. Report No. 355, Department of Computer Science.
- Elman, Jeffrey L. 1990. Finding structure in time. *Cognitive Science* 14:179-211.
- Elworthy, David. 1994. Does Baum-Welch re-estimation help taggers? In *ANLP 4*, pp. 53-58.
- Estoup, J. B. 1916. *Gammes Sténographiques*, 4th edition. Paris.

- Evans, David A., Kimberly Ginther-Webster, Mary Hart, Robert G. Lefferts, and Ira A. Monarch. 1991. Automatic indexing using selective NLP and first-order thesauri. In *Proceedings of the RIAO*, volume 2, pp. 624-643.
- Evans, David A., and Chengxiang Zhai. 1996. Noun-phrase analysis in unrestricted text for information retrieval. In *ACL 34*, pp. 17-24.
- Fagan, Joel L. 1987. Automatic phrase indexing for document retrieval: An examination of syntactic and non-syntactic methods. In *SIGIR '87*, pp. 91-101.
- Fagan, Joel L. 1989. The effectiveness of a nonsyntactic approach to automatic phrase indexing for document retrieval. *Journal of the American Society for Information Science* 40:115-132.
- Fano, Robert M. 1961. *Transmission of information; a statistical theory of communications*. New York: MIT Press.
- Fillmore, Charles J., and B. T. S. Atkins. 1994. Starting where the dictionaries stop: The challenge of corpus lexicography. In B.T.S. Atkins and A. Zampolli (eds.), *Computational Approaches to the Lexicon*, pp. 349-393. Oxford: Oxford University Press.
- Finch, Steven, and Nick Chater. 1994. Distributional bootstrapping: From word class to proto-sentence. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 301-306, Hillsdale, NJ: Lawrence Erlbaum.
- Finch, Steven Paul. 1993. *Finding Structure in Language*. PhD thesis, University of Edinburgh.
- Firth, J. R. 1957. A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32. Oxford: Philological Society. Reprinted in F. R. Palmer (ed), *Selected Papers of J. R. Firth 1952-1959*, London: Longman, 1968.
- Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society* 222:309-368.
- Fontenelle, Thierry, Walter Bröls, Luc Thomas, Tom Vanallemeersch, and Jacques Jansen. 1994. DECIDE, MLAP-Project 93-19, deliverable D-1a: survey of collocation extraction tools. Technical report, University of Liege, Liege, Belgium.
- Ford, Marilyn, Joan Bresnan, and Ronald M. Kaplan. 1982. A competence-based theory of syntactic closure. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pp. 727-796. Cambridge, MA: MIT Press.
- Foster, G. F. 1991. Statistical lexical disambiguation. Master's thesis, School of Computer Science, McGill University.
- Frakes, William B., and Ricardo Baeza-Yates (eds.). 1992. *Information Retrieval*. Englewood Cliffs, NJ: Prentice Hall.
- Francis, W. Nelson, and Henry Kučera. 1964. *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Providence, RI: Dept of Linguistics, Brown University.

- Francis, W. Nelson, and Henry Kučera. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston, MA: Houghton Mifflin.
- Franz, Alexander. 1996. *Automatic Ambiguity Resolution in Natural Language Processing*, volume 1171 of *Lecture Notes in Artificial Intelligence*. Berlin: Springer Verlag.
- Franz, Alexander. 1997. Independence assumptions considered harmful. In *ACL 35/EACL 8*, pp. 182-189.
- Franz, Alexander Mark. 1995. *A Statistical Approach to Syntactic Ambiguity Resolution*. PhD thesis, CMU.
- Frazier, Lyn. 1978. *On Comprehending Sentences: Syntactic Parsing Strategies*. PhD thesis, University of Connecticut.
- Freedman, David, Robert Pisani, and Roger Purves. 1998. *Statistics*. New York: W. W. Norton. 3rd ed.
- Friedl, Jeffrey E. F. 1997. *Mastering Regular Expressions*. Sebastopol, CA: O'Reilly & Associates.
- Fu, King-Sun. 1974. *Syntactic Methods in Pattern Recognition*. London: Academic Press.
- Fung, Pascale, and Kenneth W. Church. 1994. K-vec: A new approach for aligning parallel texts. In *COLING 15*, pp. 1096-1102.
- Fung, Pascale, and Kathleen McKeown. 1994. Aligning noisy parallel corpora across language groups: Word pair feature matching by dynamic time warping. In *Proceedings of the Association for Machine Translation in the Americas (AMTA-94)*, pp. 81-88.
- Gale, William A., and Kenneth W. Church. 1990a. Estimation procedures for language context: Poor estimates of context are worse than none. In *Proceedings in Computational Statistics (COMPSTAT 9)*, pp. 69-74.
- Gale, William A., and Kenneth W. Church. 1990b. Poor estimates of context are worse than none. In *Proceedings of the June 1990 DARPA Speech and Natural Language Workshop*, pp. 283-287, Hidden Valley, PA.
- Gale, William A., and Kenneth W. Church. 1991. A program for aligning sentences in bilingual corpora. In *ACL 29*, pp. 177-184.
- Gale, William A., and Kenneth W. Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics* 19:75-102.
- Gale, William A., and Kenneth W. Church. 1994. What's wrong with adding one? In Nelleke Oostdijk and Pieter de Haan (eds.), *Corpus-Based Research into Language: in honour of Jan Aarts*. Amsterdam: Rodopi.
- Gale, William A., Kenneth W. Church, and David Yarowsky. 1992a. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *ACL 30*, pp. 249-256.

- Gale, William A., Kenneth W. Church, and David Yarowsky. 1992b. A method for disambiguating word senses in a large corpus. *Computers and the Humanities* 26:415-439.
- Gale, William A., Kenneth W. Church, and David Yarowsky. 1992c. A method for disambiguating word senses in a large corpus. Technical report, AT&T Bell Laboratories, Murray Hill, NJ.
- Gale, William A., Kenneth W. Church, and David Yarowsky. 1992d. Using bilingual materials to develop word sense disambiguation methods. In *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, pp. 101-112.
- Gale, William A., Kenneth W. Church, and David Yarowsky. 1992e. Work on statistical methods for word sense disambiguation. In Robert Goldman, Peter Norvig, Eugene Charniak, and Bill Gale (eds.), *Working Notes of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, pp. 54-60, Menlo Park, CA. AAAI Press.
- Gale, William A., and Geoffrey Sampson. 1995. Good-Turing frequency estimation without tears. *Journal of Quantitative Linguistics* 2:217-237.
- Gallager, Robert G. 1968. *Information theory and reliable communication*. New York: Wiley.
- Garside, Roger. 1995. Grammatical tagging of the spoken part of the British National Corpus: a progress report. In Geoffrey N. Leech, Greg Myers, and Jenny Thomas (eds.), *Spoken English on computer: transcription, mark-up, and application*. Harlow, Essex: Longman.
- Garside, Roger, and Fanny Leech. 1987. The UCREL probabilistic parsing system. In Roger Garside, Geoffrey Leech, and Geoffrey Sampson (eds.), *The Computational Analysis of English: A Corpus-Based Approach*, pp. 66-81. London: Longman.
- Garside, Roger, Geoffrey Sampson, and Geoffrey Leech (eds.). 1987. *The Computational analysis of English: a corpus-based approach*. London: Longman.
- Gaussier, Éric. 1998. Flow network models for word alignment and terminology extraction from bilingual corpora. In *ACL 36/COLING 17*, pp. 444-450.
- Ge, Niyu, John Hale, and Eugene Charniak. 1998. A statistical approach to anaphora resolution. In *WVLC 6*, pp. 161-170.
- Ghahramani, Zoubin. 1994. Solving inverse problems using an EM approach to density estimation. In Michael C. Mozer, Paul Smolensky, David S. Touretzky, and Andreas S. Weigend (eds.), *Proceedings of the 1993 Connectionist Models Summer School*, Hillsdale, NJ. Erlbaum Associates.
- Gibson, Edward, and Neal J. Pearlmutter. 1994. A corpus-based analysis of psycholinguistic constraints on prepositional-phrase attachment. In Charles

- Clifton, Jr., Lyn Frazier, and Keith Rayner (eds.), *Perspectives on Sentence Processing*, pp. 181–198. Hillsdale, NJ: Lawrence Erlbaum.
- Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10:447–474.
- Goldszmidt, Moises, and Mehran Sahami. 1998. A probabilistic approach to full-text document clustering. Technical Report SIDL-WP-1998-0091, Stanford Digital Library Project, Stanford, CA.
- Golub, Gene H., and Charles F. van Loan. 1989. *Matrix Computations*. Baltimore: The Johns Hopkins University Press.
- Good, I. J. 1953. The population frequencies of species and the estimation of population parameters. *Biometrika* 40:237–264.
- Good, I. J. 1979. Studies in the history of probability and statistics. XXXVII: A. M. Turing's statistical work in World War II. *Biometrika* 66:393–396.
- Goodman, Joshua. 1996. Parsing algorithms and metrics. In *ACL 34*, pp. 177–183.
- Greenbaum, Sidney. 1993. The tagset for the International Corpus of English. In Eric Atwell and Clive Souter (eds.), *Corpus-based Computational Linguistics*, pp. 11–24. Amsterdam: Rodopi.
- Greene, Barbara B., and Gerald M. Rubin. 1971. Automatic grammatical tagging of English. Technical report, Brown University, Providence, RI.
- Grefenstette, Gregory. 1992a. Finding semantic similarity in raw text: the deese antonyms. In Robert Goldman, Peter Norvig, Eugene Charniak, and Bill Gale (eds.), *Working Notes of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, pp. 61–65, Menlo Park, CA. AAAI Press.
- Grefenstette, Gregory. 1992b. Use of syntactic context to produce term association lists for text retrieval. In *SIGIR '92*, pp. 89–97.
- Grefenstette, Gregory. 1994. *Explorations in Automatic Thesaurus Discovery*. Boston: Kluwer Academic Press.
- Grefenstette, Gregory. 1996. Evaluation techniques for automatic semantic extraction: Comparing syntactic and window-based approaches. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, chapter 11, pp. 205–216. Cambridge, MA: MIT Press.
- Grefenstette, Gregory (ed.). 1998. *Cross-language information retrieval*. Boston, MA: Kluwer Academic Publishers.
- Grefenstette, Gregory, and Pasi Tapanainen. 1994. What is a word, what is a sentence? Problems of tokenization. In *Proceedings of the Third International Conference on Computational Lexicography (COMPLEX '94)*, pp. 79–87, Budapest. Available as Rank Xerox Research Centre technical report MLTT-004.

- Grenander, Ulf. 1967. Syntax-controlled probabilities. Technical report, Division of Applied Mathematics, Brown University.
- Günter, R., L. B. Levin, B. Shapiro, and P. Wagner. 1996. Zipf's law and the effect of ranking on probability distributions. *International Journal of Theoretical Physics* 35:395-417.
- Guthrie, Joe A., Louise Guthrie, Yorick Wilks, and Homa Aidinejad. 1991. Subject-dependent co-occurrence and word sense disambiguation. In *ACL 29*, pp. 146-152.
- Guthrie, Louise, James Pustejovsky, Yorick Wilks, and Brian M. Slator. 1996. The role of lexicons in natural language processing. *Communications of the ACM* 39:63-72.
- Halliday, M. A. K. 1966. Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.), *In memory of J. R. Firth*, pp. 148-162. London: Longmans.
- Halliday, M. A. K. 1994. *An introduction to functional grammar*, 2nd edition. London: Edward Arnold.
- Harman, D.K. (ed.). 1996. *The Third Text REtrieval Conference (TREC-4)*. Washington DC: U.S. Department of Commerce.
- Harman, D. K. (ed.). 1994. *The Second Text REtrieval Conference (TREC-2)*. Washington DC: U.S. Department of Commerce. NIST Special Publication 500-215.
- Harnad, Stevan (ed.). 1987. *Categorical perception: the groundwork of cognition*. Cambridge: Cambridge University Press.
- Harris, B. 1988. Bi-text, a new concept in translation theory. *Language Monthly* 54.
- Harris, T. E. 1963. *The Theory of Branching Processes*. Berlin: Springer.
- Harris, Zellig. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
- Harrison, Philip, Steven Abney, Ezra Black, Dan Flickinger, Ralph Grishman, Claudia Gdaniec, Donald Hindle, Robert Ingria, Mitch Marcus, Beatrice Santorini, and Tomek Strzalkowski. 1991. Natural Language Processing Systems Evaluation Workshop, Technical Report RL-TR-91-362. In Jeannette G. Neal and Sharon M. Walter (eds.), *Evaluating Syntax Performance of Parser/Grammars of English*, Rome Laboratory, Air Force Systems Command, Griffis Air Force Base, NY 13441-5700.
- Harter, Steve. 1975. A probabilistic approach to automatic keyword indexing: Part II. an algorithm for probabilistic indexing. *Journal of the American Society for Information Science* 26:280-289.

- Haruno, Masahiko, and Takefumi Yamazaki. 1996. High-performance bilingual text alignment using statistical and dictionary information. In *ACL 34*, pp. 131-138.
- Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. 1993. Towards the automatic identification of adjectival scales: clustering adjectives according to meaning. In *ACL 31*, pp. 172-182.
- Hawthorne, Mark. 1994. The computer in literary analysis: Using TACT with students. *Computers and the Humanities* 28:19-27.
- Hearst, Marti, and Christian Plaunt. 1993. Subtopic structuring for full-length document access. In *SIGIR '93*, pp. 59-68.
- Hearst, Marti A. 1991. Noun homograph disambiguation using local context in large text corpora. In *Seventh Annual Conference of the UW Centre for the New OED and Text Research*, pp. 1-22, Oxford.
- Hearst, Marti A. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING 14*, pp. 539-545.
- Hearst, Marti A. 1994. *Context and Structure in Automated Full-Text Information Access*. PhD thesis, University of California at Berkeley.
- Hearst, Marti A. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23:33-64.
- Hearst, Marti A., and Hinrich Schütze. 1995. Customizing a lexicon to better suit a computational task. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 77-96. Cambridge, MA: MIT Press.
- Henderson, James, and Peter Lane. 1998. A connectionist architecture for learning to parse. In *ACL 36/COLING 17*, pp. 531-537.
- Hermjakob, Ulf, and Raymond J. Mooney. 1997. Learning parse and translation decisions from examples with rich context. In *ACL 35/EACL 8*, pp. 482-489.
- Hertz, John A., Richard G. Palmer, and Anders S. Krogh. 1991. *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.
- Herwijnen, Eric van. 1994. *Practical SGML*, 2nd edition. Dordrecht: Kluwer Academic.
- Hickey, Raymond. 1993. Lexa: Corpus processing software. Technical report, The Norwegian Computing Centre for the Humanities, Bergen.
- Hindle, Donald. 1990. Noun classification from predicate argument structures. In *ACL 28*, pp. 268-275.
- Hindle, Donald. 1994. A parser for text corpora. In B.T.S. Atkins and A. Zampolli (eds.), *Computational Approaches to the Lexicon*, pp. 103-151. Oxford: Oxford University Press.

- Hindle, Donald, and Mats Rooth. 1993. Structural ambiguity and lexical relations. *Computational Linguistics* 19:103-120.
- Hirst, Graeme. 1987. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge: Cambridge University Press.
- Hodges, Julia, Shiyun Yie, Ray Reighart, and Lois Boggess. 1996. An automated system that assists in the generation of document indexes. *Natural Language Engineering* 2:137-160.
- Holmes, V. M., L. Stowe, and L. Cupples. 1989. Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language* 28:668-689.
- Honavar, Vasant, and Giora Slutzki (eds.). 1998. *Grammatical inference: 4th international colloquium, ICGI-98*. Berlin: Springer.
- Hopcroft, John E., and Jeffrey D. Ullman. 1979. *Introduction to automata theory, languages, and computation*. Reading, MA: Addison-Wesley.
- Hopper, Paul J., and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hornby, A. S. 1974. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. Third Edition.
- Horning, James Jay. 1969. *A study of grammatical inference*. PhD thesis, Stanford.
- Huang, T., and King Sun Fu. 1971. On stochastic context-free languages. *Information Sciences* 3:201-224.
- Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Hull, David. 1996. Stemming algorithms - A case study for detailed evaluation. *Journal of the American Society for Information Science* 47:70-84.
- Hull, David. 1998. A practical approach to terminology alignment. In Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme (eds.), *Proceedings of Computerm '98*, pp. 1-7, Montreal, Canada.
- Hull, David, and Doug Oard (eds.). 1997. *AAAI Symposium on Cross-Language Text and Speech Retrieval*. Stanford, CA: AAAI Press.
- Hull, David A., and Gregory Grefenstette. 1998. Querying across languages: A dictionary-based approach to multilingual information retrieval. In Karen Sparck Jones and Peter Willett (eds.), *Readings in Information Retrieval*. San Francisco: Morgan Kaufmann.
- Hull, David A., Jan O. Pedersen, and Hinrich Schütze. 1996. Method combination for document filtering. In *SIGIR '96*, pp. 279-287.
- Hutchins, S. E. 1970. *Stochastic Sources for Context-free Languages*. PhD thesis, University of California, San Diego.

- Ide, Nancy, and Jean Véronis (eds.). 1995. *The Text Encoding Initiative: Background and Context*. Dordrecht: Kluwer Academic. Reprinted from *Computers and the Humanities* 29(1-3), 1995.
- Ide, Nancy, and Jean Véronis. 1998. Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics* 24:1-40.
- Ide, Nancy, and Donald Walker. 1992. Introduction: Common methodologies in humanities computing and computational linguistics. *Computers and the Humanities* 26:327-330.
- Inui, K., V. Sornlertlamvanich, H. Tanaka, and T. Tokunaga. 1997. A new formalization of probabilistic GLR parsing. In *Proceedings of the Fifth International Workshop on Parsing Technologies (IWPT-97)*, pp. 123-134, MIT.
- Isabelle, Pierre. 1987. Machine translation at the TAUM group. In Margaret King (ed.), *Machine Translation Today: The State of the Art*, pp. 247-277. Edinburgh: Edinburgh University Press.
- Jacquemin, Christian. 1994. FASTR: A unification-based front-end to automatic indexing. In *Proceedings of RIAO*, pp. 34-47, Rockefeller University, New York.
- Jacquemin, Christian, Judith L. Klavans, and Evelyne Tzoukermann. 1997. Expansion of multi-word terms for indexing and retrieval using morphology and syntax. In *ACL 35/EACL 8*, pp. 24-31.
- Jain, Anil K., and Richard C. Dubes. 1988. *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall.
- Jeffreys, Harold. 1948. *Theory of Probability*. Oxford: Clarendon Press.
- Jelinek, Frederick. 1969. Fast sequential decoding algorithm using a stack. *IBM Journal of Research and Development* pp. 675-685.
- Jelinek, Frederick. 1976. Continuous speech recognition by statistical methods. *IEEE* 64:532-556.
- Jelinek, Frederick. 1985. Markov source modeling of text generation. In J. K. Skwirzynski (ed.), *The Impact of Processing Techniques on Communications*, volume E91 of *NATO ASI series*, pp. 569-598. Dordrecht: M. Nijhoff.
- Jelinek, Fred. 1990. Self-organized language modeling for speech recognition. Printed in (Waibel and Lee 1990), pp. 450-506.
- Jelinek, Frederick. 1997. *Statistical Methods for Speech Recognition*. Cambridge, MA: MIT Press.
- Jelinek, Frederick, Lalit R. Bahl, and Robert L. Mercer. 1975. Design of a linguistic statistical decoder for the recognition of continuous speech. *IEEE Transactions on Information Theory* 21:250-256.
- Jelinek, F., J. Lafferty, D. Magerman, R. Mercer, A. Ratnaparkhi, and S. Roukos. 1994. Decision tree parsing using a hidden derivation model. In *Proceedings of the 1994 Human Language Technology Workshop*, pp. 272-277. DARPA.

- Jelinek, Fred, and John D. Lafferty. 1991. Computation of the probability of initial substring generation by stochastic context-free grammars. *Computational Linguistics* 17:315-324.
- Jelinek, F., J. D. Lafferty, and R. L. Mercer. 1990. Basic methods of probabilistic context free grammars. Technical Report RC 16374 (#72684), IBM T. J. Watson Research Center.
- Jelinek, F., J. D. Lafferty, and R. L. Mercer. 1992a. Basic methods of probabilistic context free grammars. In P. Laface and R. De Mori (eds.), *Speech Recognition and Understanding: Recent Advances, Trends, and Applications*, volume 75 of *Series F: Computer and Systems Sciences*. Springer Verlag.
- Jelinek, Fred, and Robert Mercer. 1985. Probability distribution estimation from sparse data. *IBM Technical Disclosure Bulletin* 28:2591-2594.
- Jelinek, Frederick, Robert L. Mercer, and Salim Roukos. 1992b. Principles of lexical language modeling for speech recognition. In Sadaoki Furui and M. Mohan Sondhi (eds.), *Advances in Speech Signal Processing*, pp. 651-699. New York: Marcel Dekker.
- Jensen, Karen, George E. Heidorn, and Stephen D. Richardson (eds.). 1993. *Natural language processing: The PLNLP approach*. Boston: Kluwer Academic Publishers.
- Johansson, Stig, G. N. Leech, and H. Goodluck. 1978. *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers*. Oslo: Dept of English, University of Oslo.
- Johnson, Mark. 1998. The effect of alternative tree representations on tree bank grammars. In *Proceedings of Joint Conference on New Methods in Language Processing and Computational Natural Language Learning (NeM-LaP3/CoNLL98)*, pp. 39-48, Macquarie University.
- Johnson, W. E. 1932. Probability: deductive and inductive problems. *Mind* 41: 421-423.
- Joos, Martin. 1936. Review of *The Psycho-Biology of Language*. *Language* 12: 196-210.
- Jorgensen, Julia. 1990. The psychological reality of word senses. *Journal of Psycholinguistic Research* 19:167-190.
- Joshi, Aravind K. 1993. Tree-adjointing grammars. In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press.
- Justeson, John S., and Slava M. Katz. 1991. Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics* 17:1-19.
- Justeson, John S., and Slava M. Katz. 1995a. Principled disambiguation: Discriminating adjective senses with modified nouns. *Computational Linguistics* 24: 1-28.

- Justeson, John S., and Slava M. Katz. 1995b. Technical terminology: some linguistic properties and an algorithm for identification in text. *Natural Language Engineering* 1:9-27.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky (eds.). 1982. *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.
- Kan, Min-Yen, Judith L. Klavans, and Kathleen R. McKeown. 1998. Linear segmentation and segment significance. In *WVLC 6*, pp. 197-205.
- Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pp. 173-281. Cambridge, MA: MIT Press.
- Karlsson, Fred, Atro Voutilainen, Juha Heikkilä, and Arto Anttila. 1995. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Berlin: Mouton de Gruyter.
- Karov, Yael, and Shimon Edelman. 1998. Similarity-based word sense disambiguation. *Computational Linguistics* 24:41-59.
- Karttunen, Lauri. 1986. Radical lexicalism. Technical Report 86-68, Center for the Study of Language and Information, Stanford CA.
- Katz, Slava M. 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-35:400-401.
- Katz, Slava M. 1996. Distribution of content words and phrases in text and language modelling. *Natural Language Engineering* 2:15-59.
- Kaufman, Leonard, and Peter J. Rousseeuw. 1990. *Finding groups in data*. New York: Wiley.
- Kaufmann, Stefan. 1998. Second-order cohesion: Using wordspace in text segmentation. Department of Linguistics, Stanford University.
- Kay, Martin, and Martin Röscheisen. 1993. Text-translation alignment. *Computational Linguistics* 19:121-142.
- Kehler, Andrew. 1997. Probabilistic coreference in information extraction. In *EMNLP 2*, pp. 163-173.
- Kelly, Edward, and Phillip Stone. 1975. *Computer Recognition of English Word Senses*. Amsterdam: North-Holland.
- Kempe, André. 1997. Finite state transducers approximating hidden markov models. In *ACL 35/EACL 8*, pp. 460-467.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.

- Kent, Roland G. 1930. Review of *Relative Frequency as a Determinant of Phonetic Change*. *Language* 6:86–88.
- Kilgarriff, Adam. 1993. Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities* 26:365–387.
- Kilgarriff, Adam. 1997. “i don’t believe in word senses”. *Computers and the Humanities* 31:91–113.
- Kilgarriff, Adam, and Tony Rose. 1998. Metrics for corpus similarity and homogeneity. Manuscript, ITRI, University of Brighton.
- Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science* 220:671–680.
- Klavans, Judith, and Min-Yen Kan. 1998. Role of verbs in document analysis. In *ACL 36/COLING 17*, pp. 680–686.
- Klavans, Judith L., and Evelyne Tzoukermann. 1995. Dictionaries and corpora: Combining corpus and machine-readable dictionary data for building bilingual lexicons. *Journal of Machine Translation* 10.
- Klein, Sheldon, and Robert F. Simmons. 1963. A computational approach to grammatical coding of English words. *Journal of the Association for Computing Machinery* 10:334–347.
- Kneser, Reinhard, and Hermann Ney. 1995. Improved backing-off for *m*-gram language modeling. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, volume 1, pp. 181–184.
- Knight, Kevin. 1997. Automating knowledge acquisition for machine translation. *AI Magazine* 18:81–96.
- Knight, Kevin, Ishwar Chander, Matthew Haines, Vasileios Hatzivassiloglou, Edward Hovy, Masayo Iida, Steve Luk, Richard Whitney, and Kenji Yamada. 1995. Filling knowledge gaps in a broad-coverage MT system. In *Proceedings of IJCAI-95*.
- Knight, Kevin, and Jonathan Graehl. 1997. Machine transliteration. In *ACL 35/EACL 8*, pp. 128–135.
- Knight, Kevin, and Vasileios Hatzivassiloglou. 1995. Two-level, many-paths generation. In *ACL 33*, pp. 252–260.
- Knill, Kate M., and Steve Young. 1997. Hidden markov models in speech and language processing. In Steve Young and Gerrit Bloothoof (eds.), *Corpus-Based Methods in Language and Speech Processing*, pp. 27–68. Dordrecht: Kluwer Academic.
- Kohonen, Teuvo. 1997. *Self-Organizing Maps*. Berlin, Heidelberg, New York: Springer Verlag. Second Extended Edition.
- Korfhage, Robert R. 1997. *Information Storage and Retrieval*. Berlin: John Wiley.

- Krenn, Brigitte, and Christer Samuelsson. 1997. The linguist's guide to statistics. manuscript, University of Saarbrücken.
- Krovetz, Robert. 1991. Lexical acquisition and information retrieval. In Uri Zernik (ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pp. 45-64. Hillsdale, NJ: Lawrence Erlbaum.
- Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1-27.
- Kruskal, J. B. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115-129.
- Kučera, Henry, and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Kupiec, Julian. 1991. A trellis-based algorithm for estimating the parameters of a hidden stochastic context-free grammar. In *Proceedings of the Speech and Natural Language Workshop*, pp. 241-246. DARPA.
- Kupiec, Julian. 1992a. An algorithm for estimating the parameters of unrestricted hidden stochastic context-free grammars. In *COLING 14*, pp. 387-393.
- Kupiec, Julian. 1992b. Robust part-of-speech tagging using a Hidden Markov Model. *Computer Speech and Language* 6:225-242.
- Kupiec, Julian. 1993a. An algorithm for finding noun phrase correspondences in bilingual corpora. In *ACL 31*, pp. 17-22.
- Kupiec, Julian. 1993b. MURAX: A robust linguistic approach for question answering using an on-line encyclopedia. In *SIGIR '93*, pp. 181-190.
- Kupiec, Julian, Jan Pedersen, and Francine Chen. 1995. A trainable document summarizer. In *SIGIR '95*, pp. 68-73.
- Kwok, K. L., and M. Chan. 1998. Improving two-stage ad-hoc retrieval for short queries. In *SIGIR '98*, pp. 250-256.
- Lafferty, John, Daniel Sleator, and Davy Temperley. 1992. Grammatical trigrams: A probabilistic model of link grammar. In *Proceedings of the 1992 AAAI Fall Symposium on Probabilistic Approaches to Natural Language*.
- Lakoff, George. 1987. *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.
- Landauer, Thomas K., and Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review* 104:211-240.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, volume 1. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar*, volume 2. Stanford, CA: Stanford University Press.

- Laplace, Pierre Simon marquis de. 1814. *Essai philosophique sur les probabilités*. Paris: Mme. Ve. Courcier.
- Laplace, Pierre Simon marquis de. 1995. *Philosophical Essay On Probabilities*. New York: Springer-Verlag.
- Lari, K., and S. J. Young. 1990. The estimation of stochastic context-free grammars using the inside-outside algorithm. *Computer Speech and Language* 4: 35-56.
- Lari, K., and S. J. Young. 1991. Application of stochastic context free grammar using the inside-outside algorithm. *Computer Speech and Language* 5:237-257.
- Lau, Raymond. 1994. Adaptive statistical language modelling. Master's thesis, Massachusetts Institute of Technology.
- Lau, Ray, Ronald Rosenfeld, and Salim Roukos. 1993. Adaptive language modeling using the maximum entropy principle. In *Proceedings of the Human Language Technology Workshop*, pp. 108-113. ARPA.
- Lauer, Mark. 1995a. Corpus statistics meet the noun compound: Some empirical results. In *ACL 33*, pp. 47-54.
- Lauer, Mark. 1995b. *Designing Statistical Language Learners: Experiments on Noun Compounds*. PhD thesis, Macquarie University, Sydney, Australia.
- Leacock, Claudia, Martin Chodorow, and George A. Miller. 1998. Using corpus statistics and Wordnet relations for sense identification. *Computational Linguistics* 24:147-165.
- Lesk, Michael. 1986. Automatic sense disambiguation: How to tell a pine cone from an ice cream cone. In *Proceedings of the 1986 SIGDOC Conference*, pp. 24-26, New York. Association for Computing Machinery.
- Lesk, M. E. 1969. Word-word association in document retrieval systems. *American Documentation* 20:27-38.
- Levin, Beth. 1993. *English Verb Classes and Alternations*. Chicago: The University of Chicago Press.
- Levine, John R., Tony Mason, and Doug Brown. 1992. *Lex & Yacc*, 2nd edition. Sebastopol, CA: O'Reilly & Associates.
- Levinson, S. E., L. R. Rabiner, and M. M. Sondhi. 1983. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Technical Journal* 62:1035-1074.
- Lewis, David D. 1992. An evaluation of phrasal and clustered representations on a text categorization task. In *SIGIR '92*, pp. 37-50.
- Lewis, David D., and Karen Sparck Jones. 1996. Natural language processing for information retrieval. *Communications of the ACM* 39:92-101.

- Lewis, David D., and Marc Ringuette. 1994. A comparison of two learning algorithms for text categorization. In *Proc. SDAIR 94*, pp. 81-93, Las Vegas, NV.
- Lewis, David D., Robert E. Schapire, James P. Callan, and Ron Papka. 1996. Training algorithms for linear text classifiers. In *SIGIR '96*, pp. 298-306.
- Li, Hang, and Naoki Abe. 1995. Generalizing case frames using a thesaurus and the mdl principle. In *Proceedings of Recent Advances in Natural Language Processing*, pp. 239-248, Tzgov Chark, Bulgaria.
- Li, Hang, and Naoki Abe. 1996. Learning dependencies between case frame slots. In *COLING 16*, pp. 10-15.
- Li, Hang, and Naoki Abe. 1998. Word clustering and disambiguation based on co-occurrence data. In *ACL 36/COLING 17*, pp. 749-755.
- Li, Wentian. 1992. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory* 38:1842-1845.
- Lidstone, G. J. 1920. Note on the general case of the Bayes-Laplace formula for inductive or *a priori* probabilities. *Transactions of the Faculty of Actuaries* 8: 182-192.
- Light, Marc. 1996. Morphological cues for lexical semantics. In *ACL 34*, pp. 25-31.
- Littlestone, Nick. 1995. Comparing several linear-threshold learning algorithms on tasks involving superfluous attributes. In A. Prieditis (ed.), *Proceedings of the 12th International Conference on Machine Learning*, pp. 353-361, San Francisco, CA. Morgan Kaufmann.
- Littman, Michael L., Susan T. Dumais, and Thomas K. Landauer. 1998a. Automatic cross-language information retrieval using latent semantic indexing. In Gregory Grefenstette (ed.), *Cross Language Information Retrieval*. Kluwer.
- Littman, Michael L., Fan Jiang, and Greg A. Keim. 1998b. Learning a language-independent representation for terms from a partially aligned corpus. In Jude Shavlik (ed.), *Proceedings of the Fifteenth International Conference on Machine Learning*, pp. 314-322. Morgan Kaufmann.
- Losee, Robert M. (ed.). 1998. *Text Retrieval and Filtering*. Boston, MA: Kluwer Academic Publishers.
- Lovins, Julie Beth. 1968. Development of a stemming algorithm. *Translation and Computational Linguistics* 11:22-31.
- Luhn, H. P. 1960. Keyword-in-context index for technical literature (KWIC index). *American Documentation* 11:288-295.
- Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

- MacDonald, M. A., N. J. Pearlmutter, and M. S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676-703.
- MacKay, David J. C., and Linda C. Peto. 1990. Speech recognition using hidden Markov models. *The Lincoln Laboratory Journal* 3:41-62.
- Magerman, David M. 1994. *Natural language parsing as statistical pattern recognition*. PhD thesis, Stanford University.
- Magerman, David M. 1995. Statistical decision-tree models for parsing. In *ACL 33*, pp. 276-283.
- Magerman, David M., and Mitchell P. Marcus. 1991. Pearl: A probabilistic chart parser. In *EACL 4*. Also published in the Proceedings of the 2nd International Workshop for Parsing Technologies.
- Magerman, David M., and Carl Weir. 1992. Efficiency, robustness, and accuracy in Picky chart parsing. In *ACL 30*, pp. 40-47.
- Mandelbrot, Benoit. 1954. Structure formelle des textes et communication. *Word* 10:1-27.
- Mandelbrot, Benoit B. 1983. *The Fractal Geometry of Nature*. New York: W. H. Freeman.
- Mani, Inderjeet, and T. Richard MacMillan. 1995. Identifying unknown proper names in newswire text. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 41-59. Cambridge, MA: MIT Press.
- Manning, Christopher D. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. In *ACL 31*, pp. 235-242.
- Manning, Christopher D., and Bob Carpenter. 1997. Probabilistic parsing using left corner language models. In *Proceedings of the Fifth International Workshop on Parsing Technologies (IWPT-97)*, pp. 147-158, MIT.
- Marchand, Hans. 1969. *Categories and types of present-day English word-formation*. München: Beck.
- Marcus, Mitchell, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *ARPA Human Language Technology Workshop*, pp. 110-115.
- Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics* 19:313-330.
- Markov, Andrei A. 1913. An example of statistical investigation in the text of 'Eugene Onyegin' illustrating coupling of 'tests' in chains. In *Proceedings of the Academy of Sciences, St. Petersburg*, volume 7 of VI, pp. 153-162.

- Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman.
- Marshall, Ian. 1987. Tag selection using probabilistic methods. In Roger Garside, Geoffrey Sampson, and Geoffrey Leech (eds.), *The Computational analysis of English: a corpus-based approach*, pp. 42–65. London: Longman.
- Martin, James. 1991. Representing and acquiring metaphor-based polysemy. In Uri Zernik (ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pp. 389–415. Hillsdale, NJ: Lawrence Erlbaum.
- Martin, W. A., K. W. Church, and R. S. Patil. 1987. Preliminary analysis of a breadth-first parser algorithm: Theoretical and experimental results. In Leonard Bolc (ed.), *Natural Language Parsing Systems*. Berlin: Springer Verlag. Also MIT LCS technical report TR-261.
- Masand, Brij, Gordon Linoff, and David Waltz. 1992. Classifying news stories using memory based reasoning. In *SIGIR '92*, pp. 59–65.
- Maxwell, III, John T. 1992. The problem with mutual information. Manuscript, Xerox Palo Alto Research Center, September 15, 1992.
- McClelland, James L., David E. Rumelhart, and the PDP Research Group (eds.). 1986. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge, MA: The MIT Press.
- McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*, 2nd edition, chapter 4, pp. 101–123. Chapman and Hall.
- McDonald, David D. 1995. Internal and external evidence in the identification and semantic categorization of proper names. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 21–39. Cambridge MA: MIT Press.
- McEnery, Tony, and Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McGrath, Sean. 1997. *PARSEME.1ST: SGML for Software Developers*. Upper Saddle River, NJ: Prentice Hall PTR.
- McMahon, John G., and Francis J. Smith. 1996. Improving statistical language model performance with automatically generated word hierarchies. *Computational Linguistics* 22:217–247.
- McQueen, C.M. Sperberg, and Lou Burnard (eds.). 1994. *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. Chicago, IL: ACH/ACL/ALLC (Association for Computers and the Humanities, Association for Computational Linguistics, Association for Literary and Linguistic Computing).
- McRoy, Susan W. 1992. Using multiple knowledge sources for word sense disambiguation. *Computational Linguistics* 18:1–30.

- Melamed, I. Dan. 1997a. A portable algorithm for mapping bitext correspondence. In *ACL 35/EACL 8*, pp. 305–312.
- Melamed, I. Dan. 1997b. A word-to-word model of translational equivalence. In *ACL 35/EACL 8*, pp. 490–497.
- Mel'čuk, Igor Aleksandrovich. 1988. *Dependency Syntax: theory and practice*. Albany: State University of New York.
- Mercer, Robert L. 1993. Inflectional morphology needs to be authenticated by hand. In *Working Notes of the AAAI Spring Symposium on Building Lexicons for Machine Translation*, pp. 99–99, Stanford, CA. AAAI Press.
- Merialdo, Bernard. 1994. Tagging English text with a probabilistic model. *Computational Linguistics* 20:155–171.
- Miclet, Laurent, and Colin de la Higuera (eds.). 1996. *Grammatical inference: learning syntax from sentences: Third International Colloquium, ICGI-96*. Berlin: Springer.
- Miikkulainen, Risto (ed.). 1993. *Subsymbolic Natural Language Processing*. Cambridge MA: MIT Press.
- Mikheev, Andrei. 1998. Feature lattices for maximum entropy modelling. In *ACL 36*, pp. 848–854.
- Miller, George A., and Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6:1–28.
- Miller, Scott, David Stallard, Robert Bobrow, and Richard Schwartz. 1996. A fully statistical approach to natural language interfaces. In *ACL 34*, pp. 55–61.
- Minsky, Marvin Lee, and Seymour Papert (eds.). 1969. *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press. Partly reprinted in (Shavlik and Dietterich 1990).
- Minsky, Marvin Lee, and Seymour Papert (eds.). 1988. *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press. Expanded edition.
- Mitchell, Tom M. 1980. The need for biases in learning generalizations. Technical Report Department of Computer Science. CBM-TR-117, Rutgers University. Reprinted in (Shavlik and Dietterich 1990), pp. 184–191.
- Mitchell, Tom M. (ed.). 1997. *Machine Learning*. New York: McGraw-Hill.
- Mitra, Mandar, Chris Buckley, Amit Singhal, and Claire Cardie. 1997. An analysis of statistical and syntactic phrases. In *Proceedings of RIAO*.
- Moffat, Alistair, and Justin Zobel. 1998. Exploring the similarity space. *ACM SIGIR Forum* 32.
- Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes. 1974. *Introduction to the theory of statistics*. New York: McGraw-Hill. 3rd edition.

- Mooney, Raymond J. 1996. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. In *EMNLP 1*, pp. 82–91.
- Moore, David S., and George P. McCabe. 1989. *Introduction to the practice of statistics*. New York: Freeman.
- Morris, Jane, and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics* 17: 21–48.
- Mosteller, Frederick, and David L. Wallace. 1984. *Applied Bayesian and Classical Inference - The Case of The Federalist Papers*. Springer Series in Statistics. New York: Springer-Verlag.
- Nagao, Makoto. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. In Alick Elithorn and Ranan B. Banerji (eds.), *Artificial and Human Intelligence*, pp. 173–180. Edinburgh: North-Holland.
- Neff, Mary S., Brigitte Bläser, Jean-Marc Langé, Hubert Lehmann, and Isabel Zapata Dominguez. 1993. Get it where you can: Acquiring and maintaining bilingual lexicons for machine translation. In *Working Notes of the AAAI Spring Symposium on Building Lexicons for Machine Translation*, pp. 98–98, Stanford, CA. AAAI Press.
- Nevill-Manning, Craig G., Ian H. Witten, and Gordon W. Paynter. 1997. Browsing in digital libraries: a phrase-based approach. In *Proceedings of ACM Digital Libraries*, pp. 230–236, Philadelphia, PA. Association for Computing Machinery.
- Newmeyer, Frederick J. 1988. *Linguistics: The Cambridge Survey*. Cambridge, England: Cambridge University Press.
- Ney, Hermann, and Ute Essen. 1993. Estimating 'small' probabilities by leaving-one-out. In *Eurospeech '93*, volume 3, pp. 2239–2242. ESCA.
- Ney, Hermann, Ute Essen, and Reinhard Kneser. 1994. On structuring probabilistic dependencies in stochastic language modeling. *Computer Speech and Language* 8:1–28.
- Ney, Hermann, Sven Martin, and Frank Wessel. 1997. Statistical language modeling using leaving-one-out. In Steve Young and Gerrit Bloothoof (eds.), *Corpus-Based Methods in Language and Speech Processing*, pp. 174–207. Dordrecht: Kluwer Academic.
- Ng, Hwee Tou, and John Zelle. 1997. Corpus-based approaches to semantic interpretation in natural language processing. *AI Magazine* 18:45–64.
- Ng, Hwee Tou, and Hian Beng Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *ACL 34*, pp. 40–47.

- Nie, Jian-Yun, Pierre Isabelle, Pierre Plamondon, and George Foster. 1998. Using a probabilistic translation model for cross-language information retrieval. In *WVLC 6*, pp. 18–27.
- Nießen, S., S. Vogel, H. Ney, and C. Tillmann. 1998. A DP based search algorithm for statistical machine translation. In *ACL 36/COLING 17*, pp. 960–967.
- Nunberg, Geoffrey. 1990. *The Linguistics of Punctuation*. Stanford, CA: CSLI Publications.
- Nunberg, Geoff, and Annie Zaenen. 1992. Systematic polysemy in lexicology and lexicography. In *Proceedings of Euralex II*, Tampere, Finland.
- Oaksford, M., and N. Chater. 1998. *Rational Models of Cognition*. Oxford, England: Oxford University Press.
- Oard, Douglas W., and Nicholas DeClaris. 1996. Cognitive models for text filtering. Manuscript, University of Maryland, College Park.
- Ostler, Nicholas, and B. T. S. Atkins. 1992. Predictable meaning shift: Some linguistic properties of lexical implication rules. In James Pustejovsky and Sabine Bergler (eds.), *Lexical Semantics and Knowledge Representation: Proceedings of the 1st SIGLEX Workshop*, pp. 76–87. Berlin: Springer Verlag.
- Paik, Woojin, Elizabeth D. Liddy, Edmund Yu, and Mary McKenna. 1995. Categorizing and standardizing proper nouns for efficient information retrieval. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 61–73. Cambridge MA: MIT Press.
- Palmer, David D., and Marti A. Hearst. 1994. Adaptive sentence boundary disambiguation. In *ANLP 4*, pp. 78–83.
- Palmer, David D., and Marti A. Hearst. 1997. Adaptive multilingual sentence boundary disambiguation. *Computational Linguistics* 23:241–267.
- Paul, Douglas B. 1990. Speech recognition using hidden markov models. *The Lincoln Laboratory Journal* 3:41–62.
- Pearlmutter, N., and M. MacDonald. 1992. Plausibility and syntactic ambiguity resolution. In *Proceedings of the 14th Annual Conference of the Cognitive Society*.
- Pedersen, Ted. 1996. Fishing for exactness. In *Proceedings of the South-Central SAS Users Group Conference*, Austin TX.
- Pedersen, Ted, and Rebecca Bruce. 1997. Distinguishing word senses in untagged text. In *EMNLP 2*, pp. 197–207.
- Pereira, Fernando, and Yves Schabes. 1992. Inside-outside reestimation from partially bracketed corpora. In *ACL 30*, pp. 128–135.
- Pereira, Fernando, Naftali Tishby, and Lillian Lee. 1993. Distributional clustering of English words. In *ACL 31*, pp. 183–190.

- Pinker, Steven. 1994. *The Language Instinct*. New York: William Morrow.
- Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.
- Pook, Stuart L., and Jason Catlett. 1988. Making sense out of searching. In *Information Online* 88, pp. 148-157, Sydney. The Information Science Section of the Library Association of Australia.
- Porter, M. F. 1980. An algorithm for suffix stripping. *Program* 14:130-137.
- Poznański, Victor, and Antonio Sanfilippo. 1995. Detecting dependencies between semantic verb subclasses and subcategorization frames in text corpora. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 175-190. Cambridge, MA: MIT Press.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1988. *Numerical Recipes in C*. Cambridge: Cambridge University Press.
- Procter, P. (ed.). 1978. *Longman dictionary of contemporary English*. Harlow, England: Longman Group.
- Prokosch, E. 1933. Review of selected studies of the principle of relative frequency in language. *Language* 9:89-92.
- Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics* 17: 409-441.
- Pustejovsky, James, Sabine Bergler, and Peter Anick. 1993. Lexical semantic techniques for corpus analysis. *Computational Linguistics* 19:331-358.
- Qiu, Yonggang, and H.P. Frei. 1993. Concept based query expansion. In *SIGIR '93*, pp. 160-169.
- Quinlan, J. R. 1986. Induction of decision trees. *Machine Learning* 1:81-106. Reprinted in (Shavlik and Dietterich 1990).
- Quinlan, John Ross. 1993. *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann Publishers.
- Quinlan, J. R. 1996. Bagging, boosting, and C4.5. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI '96)*, pp. 725-730.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rabiner, Lawrence, and Biing-Hwang Juang. 1993. *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: PTR Prentice-Hall.
- Rabiner, Lawrence R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of IEEE* 77:257-286. Reprinted in (Waibel and Lee 1990), pp. 267-296.

- Ramsey, Fred L., and Daniel W. Schafer. 1997. *The statistical sleuth: a course in methods of data analysis*. Belmont, CA: Duxbury Press.
- Ramshaw, Lance A., and Mitchell P. Marcus. 1994. Exploring the statistical derivation of transformational rule sequences for part-of-speech tagging. In *The Balancing Act. Proceedings of the Workshop*, pp. 86-95, Morristown NJ. Association of Computational Linguistics.
- Rasmussen, Edie. 1992. Clustering algorithms. In William B. Frakes and Ricardo Baeza-Yates (eds.), *Information Retrieval*, pp. 419-442. Englewood Cliffs, NJ: Prentice Hall.
- Ratnaparkhi, Adwait. 1996. A maximum entropy model for part-of-speech tagging. In *EMNLP 1*, pp. 133-142.
- Ratnaparkhi, Adwait. 1997a. A linear observed time statistical parser based on maximum entropy models. In *EMNLP 2*, pp. 1-10.
- Ratnaparkhi, Adwait. 1997b. A simple introduction to maximum entropy models for natural language processing. Technical Report IRCS Report 97-08, Institute for Research in Cognitive Science, Philadelphia, PA.
- Ratnaparkhi, Adwait. 1998. Unsupervised statistical models for prepositional phrase attachment. In *ACL 36/COLING 17*, pp. 1079-1085.
- Ratnaparkhi, Adwait, Jeff Reynar, and Salim Roukos. 1994. A maximum entropy model for prepositional phrase attachment. In *Proceedings of the ARPA Workshop on Human Language Technology*, pp. 250-255, Plainsboro, NJ.
- Read, Timothy R. C., and Noel A. C. Cressie. 1988. *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer Verlag.
- Resnik, Philip. 1992. Probabilistic tree-adjoining grammar as a framework for statistical natural language processing. In *COLING 14*, pp. 418-425.
- Resnik, Philip. 1996. Selectional constraints: an information-theoretic model and its computational realization. *Cognition* 61:127-159.
- Resnik, Philip, and Marti Hearst. 1993. Structural ambiguity and conceptual relations. In *WVLC 1*, pp. 58-64.
- Resnik, Philip, and David Yarowsky. 1998. A perspective on word sense disambiguation methods and their evaluation. In *Proceedings of the SIGLEX workshop Tagging Text with Lexical Semantics*, pp. 79-86, Washington, DC.
- Resnik, Philip Stuart. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. PhD thesis, University of Pennsylvania.
- Reynar, Jeffrey C., and Adwait Ratnaparkhi. 1997. A maximum entropy approach to identifying sentence boundaries. In *ANLP 5*, pp. 16-19.
- Riley, Michael D. 1989. Some applications of tree-based modeling to speech and language indexing. In *Proceedings of the DARPA Speech and Natural Language Workshop*, pp. 339-352. Morgan Kaufmann.

- Riloff, Ellen, and Jessica Shepherd. 1997. A corpus-based approach for building semantic lexicons. In *EMNLP 2*, pp. 117-124.
- Ristad, Eric Sven. 1995. A natural law of succession. Technical Report CS-TR-495-95, Princeton University.
- Ristad, Eric Sven. 1996. Maximum entropy modeling toolkit. Manuscript, Princeton University.
- Ristad, Eric Sven, and Robert G. Thomas. 1997. Hierarchical non-emitting Markov models. In *ACL 35/EACL 8*, pp. 381-385.
- Roark, Brian, and Eugene Charniak. 1998. Noun-phrase co-occurrence statistics for semi-automatic semantic lexicon construction. In *ACL 36/COLING 17*, pp. 1110-1116.
- Robertson, S.E., and K. Sparck Jones. 1976. Relevance weighting of search terms. *Journal of the American Society for Information Science* 27:129-146.
- Rocchio, J. J. 1971. Relevance feedback in information retrieval. In Gerard Salton (ed.), *The Smart Retrieval System - Experiments in Automatic Document Processing*, pp. 313-323. Englewood Cliffs, NJ: Prentice-Hall.
- Roche, Emmanuel, and Yves Schabes. 1995. Deterministic part-of-speech tagging with finite-state transducers. *Computational Linguistics* 21:227-253.
- Roche, Emmanuel, and Yves Schabes. 1997. *Finite-State Language Processing*. Boston, MA: MIT Press.
- Roget, P. M. 1946. *Roget's International Thesaurus*. New York: Thomas Y. Crowell.
- Rosenblatt, Frank (ed.). 1962. *Principles of neurodynamics; perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan Books.
- Rosenfeld, Ronald. 1994. *Adaptive Statistical Language Modeling: A Maximum Entropy Approach*. PhD thesis, CMU. Technical report CMU-CS-94-138.
- Rosenfeld, Roni. 1996. A maximum entropy approach to adaptive statistical language modelling. *Computer Speech and Language* 10:187-228.
- Rosenfeld, Ronald, and Xuedong Huang. 1992. Improvements in stochastic language modeling. In *Proceedings of the DARPA Speech and Natural Language Workshop*, pp. 107-111. Morgan Kaufmann.
- Rosenkrantz, Stanley J., and Philip M. Lewis, II. 1970. Deterministic left corner parser. In *IEEE Conference Record of the 11th Annual Symposium on Switching and Automata*, pp. 139-152.
- Ross, Ian C., and John W. Tukey. 1975. Introduction to these volumes. In John Wilder Tukey (ed.), *Index to Statistics and Probability*, pp. iv-x. Los Altos, CA: R & D Press.

- Roth, Dan. 1998. Learning to resolve natural language ambiguities: A unified approach. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, Menlo Park CA. AAAI Press.
- Rumelhart, D. E., and J. L. McClelland. 1986. On learning the past tenses of English verbs. In James L. McClelland, David E. Rumelhart, and the PDP Research Group (eds.), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*, pp. 216-271. Cambridge, MA: The MIT Press.
- Rumelhart, David E., James L. McClelland, and the PDP research group (eds.). 1986. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: The MIT Press.
- Rumelhart, David E., and David Zipser. 1985. Feature discovery by competitive learning. *Cognitive Science* 9:75-112.
- Russell, Stuart J., and Peter Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.
- Sakakibara, Y., M. Brown, R. Hughey, I. S. Mian, K. Sjölander, R. C. Underwood, and D. Haussler. 1994. Stochastic context-free grammars for tRNA modeling. *Nucleic Acids Research* 22:5112-5120.
- Salton, Gerard. 1971a. Experiments in automatic thesaurus construction for information retrieval. In *Proceedings IFIP Congress*, pp. 43-49.
- Salton, Gerard (ed.). 1971b. *The Smart Retrieval System - Experiments in Automatic Document Processing*. Englewood Cliffs, NJ: Prentice-Hall.
- Salton, Gerard. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA: Addison Wesley.
- Salton, G., J. Allan, C. Buckley, and A. Singhal. 1994. Automatic analysis, theme generation and summarization of machine-readable texts. *Science* 264:1421-1426.
- Salton, Gerard, and James Allen. 1993. Selective text utilization and text traversal. In *Proceedings of ACM Hypertext 93*, New York. Association for Computing Machinery.
- Salton, Gerard, and Chris Buckley. 1991. Global text matching for information retrieval. *Science* 253:1012-1015.
- Salton, Gerard, Edward A. Fox, and Harry Wu. 1983. Extended boolean information retrieval. *Communications of the ACM* 26:1022-1036.
- Salton, Gerard, and Michael J. McGill. 1983. *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Salton, Gerard, and R. W. Thorpe. 1962. An approach to the segmentation problem in speech analysis and language translation. In *Proceedings of the 1961*

- International Conference on Machine Translation of Languages and Applied Language Analysis*, volume 2, pp. 703-724, London. Her Majesty's Stationery Office.
- Sampson, Geoffrey. 1989. How fully does a machine-usable dictionary cover English text? *Literary and Linguistic Computing* 4:29-35.
- Sampson, Geoffrey. 1995. *English for the Computer*. New York: Oxford University Press.
- Sampson, Geoffrey. 1997. *Educating Eve*. London: Cassell.
- Samuel, Ken, Sandra Carberry, and K. Vijay-Shanker. 1998. Dialogue act tagging with transformation-based learning. In *ACL 36/COLING 17*, pp. 1150-1156.
- Samuelsson, Christer. 1993. Morphological tagging based entirely on bayesian inference. In *9th Nordic Conference on Computational Linguistics*, Stockholm University, Stockholm, Sweden.
- Samuelsson, Christer. 1996. Handling sparse data by successive abstraction. In *COLING 16*, pp. 895-900.
- Samuelsson, Christer, and Atro Voutilainen. 1997. Comparing a linguistic and a stochastic tagger. In *ACL 35/EACL 8*, pp. 246-253.
- Sanderson, Mark, and C. J. van Rijsbergen. 1998. The impact on retrieval effectiveness of the skewed frequency distribution of a word's senses. *ACM Transactions on Information Systems*. To appear.
- Sankoff, D. 1971. Branching processes with terminal types: applications to context-free grammars. *Journal of Applied Probability* 8:233-240.
- Santorini, Beatrice. 1990. Part-of-speech tagging guidelines for the Penn treebank project. 3rd Revision, 2nd printing, Feb. 1995. University of Pennsylvania.
- Sapir, Edward. 1921. *Language: an introduction to the study of speech*. New York: Harcourt Brace.
- Sato, Satoshi. 1992. CTM: An example-based translation aid system. In *COLING 14*, pp. 1259-1263.
- Saund, Eric. 1994. Unsupervised learning of mixtures of multiple causes in binary data. In J. Cowan, G. Tesauro, and J. Alspector (eds.), *Advances in Neural Information Processing Systems 6*. San Mateo, CA: Morgan Kaufmann Publishers.
- Schabes, Yves. 1992. Stochastic lexicalized tree-adjoining grammars. In *COLING 14*, pp. 426-432.
- Schabes, Yves, Anne Abeillé, and Aravind Joshi. 1988. Parsing strategies with lexicalized grammars: Tree adjoining grammars. In *COLING 12*, pp. 578-583.
- Schabes, Yves, Michal Roth, and Randy Osborne. 1993. Parsing the Wall Street Journal with the Inside-Outside algorithm. In *EACL 6*, pp. 341-347.

- Schapire, Robert E., Yoram Singer, and Amit Singhal. 1998. Boosting and Rocchio applied to text filtering. In *SIGIR '98*, pp. 215-223.
- Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*, pp. 44-49, Manchester, England.
- Schütze, Carson T. 1996. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.
- Schütze, Hinrich. 1992a. Context space. In Robert Goldman, Peter Norvig, Eugene Charniak, and Bill Gale (eds.), *Working Notes of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, pp. 113-120, Menlo Park, CA. AAAI Press.
- Schütze, Hinrich. 1992b. Dimensions of meaning. In *Proceedings of Supercomputing '92*, pp. 787-796, Los Alamitos, CA. IEEE Computer Society Press.
- Schütze, Hinrich. 1993. Part-of-speech induction from scratch. In *ACL 31*, pp. 251-258.
- Schütze, Hinrich. 1995. Distributional part-of-speech tagging. In *EACL 7*, pp. 141-148.
- Schütze, Hinrich. 1997. *Ambiguity Resolution in Language Learning*. Stanford, CA: CSLI Publications.
- Schütze, Hinrich. 1998. Automatic word sense discrimination. *Computational Linguistics* 24:97-124.
- Schütze, Hinrich, David A. Hull, and Jan O. Pedersen. 1995. A comparison of classifiers and document representations for the routing problem. In *SIGIR '95*, pp. 229-237.
- Schütze, Hinrich, and Jan O. Pedersen. 1995. Information retrieval based on word senses. In *Fourth Annual Symposium on Document Analysis and Information Retrieval*, pp. 161-175, Las Vegas, NV.
- Schütze, Hinrich, and Jan O. Pedersen. 1997. A cooccurrence-based thesaurus and two applications to information retrieval. *Information Processing & Management* 33:307-318.
- Schütze, Hinrich, and Yoram Singer. 1994. Part-of-speech tagging using a variable memory Markov model. In *ACL 32*, pp. 181-187.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379-423, 623-656.
- Shannon, Claude E. 1951. Prediction and entropy of printed English. *Bell System Technical Journal* 30:50-64.
- Shavlik, Jude W., and Thomas G. Dietterich (eds.). 1990. *Readings in Machine Learning*. San Mateo, CA: Morgan Kaufmann.

- Shemtov, Hadar. 1993. Text alignment in a tool for translating revised documents. In *EACL 6*, pp. 449-453.
- Sheridan, Paraic, and Alan F. Smeaton. 1992. The application of morpho-syntactic language processing to effective phrase matching. *Information Processing & Management* 28:349-370.
- Sheridan, Paraic, Martin Wechsler, and Peter Schäuble. 1997. Cross language speech retrieval: Establishing a baseline performance. In *SIGIR '97*, pp. 99-108.
- Shimohata, Sayori, Toshiyuko Sugio, and Junji Nagata. 1997. Retrieving collocations by co-occurrences and word order constraints. In *ACL 35/EACL 8*, pp. 476-481.
- Siegel, Sidney, and N. John Castellan, Jr. 1988. *Nonparametric Statistics for the Behavioral Sciences*, 2nd edition. New York: McGraw Hill.
- Silverstein, Craig, and Jan O. Pedersen. 1997. Almost-constant-time clustering of arbitrary corpus subsets. In *SIGIR '97*, pp. 60-66.
- Sima'an, Khalil. 1996. Computational complexity of probabilistic disambiguation by means of tree-grammars. In *COLING 16*, pp. 1175-1180.
- Sima'an, Khalil, Rens Bod, S. Krauwer, and Remko Scha. 1994. Efficient disambiguation by means of stochastic tree substitution grammars. In *Proceedings International Conference on New Methods in Language Processing*.
- Simard, Michel, G. F. Foster, and P. Isabelle. 1992. Using cognates to align sentences in bilingual corpora. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, pp. 67-81.
- Simard, Michel, and Pierre Plamondon. 1996. Bilingual sentence alignment: Balancing robustness and accuracy. In *Proceedings of the First Conference of the Association for Machine Translation in the Americas (AMTA-96)*, pp. 135-144.
- Sinclair, John (ed.). 1995. *Collins COBUILD English dictionary*. London: Harper Collins. New edition, completely revised.
- Singhal, Amit, Gerard Salton, and Chris Buckley. 1996. Length normalization in degraded text collections. In *Fifth Annual Symposium on Document Analysis and Information Retrieval*, pp. 149-162, Las Vegas, NV.
- Sipser, Michael. 1996. *Introduction to the theory of computation*. Boston, MA: PWS Publishing Company.
- Siskind, Jeffrey Mark. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61:39-91.
- Skut, Wojciech, and Thorsten Brants. 1998. A maximum-entropy partial parser for unrestricted text. In *WVLC 6*, pp. 143-151.

- Smadja, Frank. 1993. Retrieving collocations from text: Xtract. *Computational Linguistics* 19:143-177.
- Smadja, Frank, Kathleen R. McKeown, and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics* 22:1-38.
- Smadja, Frank A., and Kathleen R. McKeown. 1990. Automatically extracting and representing collocations for language generation. In *ACL* 28, pp. 252-259.
- Smeaton, Alan F. 1992. Progress in the application of natural language processing to information retrieval tasks. *The Computer Journal* 35:268-278.
- Smith, Tony C., and John G. Cleary. 1997. Probabilistic unification grammars. In *1997 Australasian Natural Language Processing Summer Workshop*, pp. 25-32, Macquarie University.
- Snedecor, George Waddel, and William G. Cochran. 1989. *Statistical methods*. Ames: Iowa State University Press. 8th edition.
- Sparck Jones, Karen. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28:11-21.
- Sparck Jones, Karen, and Peter Willett (eds.). 1998. *Readings in Information Retrieval*. San Francisco: Morgan Kaufmann.
- Sproat, Richard William. 1992. *Morphology and computation*. Cambridge, MA: MIT Press.
- Sproat, Richard W., Chilin Shih, William Gale, and Nancy Chang. 1996. A stochastic finite-state word-segmentation algorithm for Chinese. *Computational Linguistics* 22:377-404.
- St. Laurent, Simon. 1998. *XML: A Primer*. Foster City, CA: MIS Press/IDG Books.
- Stanfill, Craig, and David Waltz. 1986. Toward memory-based reasoning. *Communications of the ACM* 29:1213-1228.
- Steier, Amy M., and Richard K. Belew. 1993. Exporting phrases: A statistical analysis of topical language. In R. Casey and B. Croft (eds.), *Second Annual Symposium on Document Analysis and Information Retrieval*, pp. 179-190, Las Vegas, NV.
- Stolcke, Andreas. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics* 21:165-202.
- Stolcke, Andreas, and Stephen M. Omohundro. 1993. Hidden Markov model induction by Bayesian model merging. In S. J. Hanson, J. D. Cowan, and C. Lee Giles (eds.), *Advances in Neural Information Processing Systems* 5, pp. 11-18, San Mateo, CA. Morgan Kaufmann.
- Stolcke, Andreas, and Stephen M. Omohundro. 1994a. Best-first model merging for hidden Markov model induction. Technical Report TR-94-003, International Computer Science Institute, University of California at Berkeley.

- Stolcke, Andreas, and Stephen M. Omohundro. 1994b. Inducing probabilistic grammars by Bayesian model merging. In *Grammatical Inference and Applications: Proceedings of the Second International Colloquium on Grammatical Inference*. Springer Verlag.
- Stolcke, A., E. Shriberg, R. Bates, N. Coccaro, D. Jurafsky, R. Martin, M. Meteer, K. Ries, P. Taylor, and C. Van Ess-Dykema. 1998. Dialog act modeling for conversational speech. In *Applying Machine Learning to Discourse Processing*, pp. 98–105, Menlo Park, CA. AAAI Press.
- Stolz, Walter S., Percy H. Tannenbaum, and Frederick V. Carstensen. 1965. A stochastic approach to the grammatical coding of English. *Communications of the ACM* 8:399–405.
- Strang, Gilbert. 1988. *Linear algebra and its applications*, 3rd edition. San Diego: Harcourt, Brace, Jovanovich.
- Strzalkowski, Tomek. 1995. Natural language information retrieval. *Information Processing & Management* 31:397–417.
- Stubbs, Michael. 1996. *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell.
- Suppes, Patrick. 1970. Probabilistic grammars for natural languages. *Synthese* 22:95–116.
- Suppes, Patrick. 1984. *Probabilistic Metaphysics*. Oxford: Blackwell.
- Suppes, Patrick, Michael Böttner, and Lin Liang. 1996. Machine learning comprehension grammars for ten languages. *Computational Linguistics* 22:329–350.
- Tabor, Whitney. 1994. *Syntactic Innovation: A Connectionist Model*. PhD thesis, Stanford.
- Tague-Sutcliffe, Jean. 1992. The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management* 28:467–490. Reprinted in (Sparck Jones and Willett 1998).
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical form. In Timothy Shopen (ed.), *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*, pp. 57–149. Cambridge, MA: Cambridge University Press.
- Tanenhaus, M. K., and J. C. Trueswell. 1995. Sentence comprehension. In J. Miller and P. Eimas (eds.), *Handbook of Perception and Cognition*, volume 11, pp. 217–262. San Diego: Academic Press.
- Tesnière, Lucien. 1959. *Éléments de Syntaxe Structurale*. Paris: Librairie C. Klincksieck.
- Tomita, Masaru (ed.). 1991. *Generalized LR parsing*. Boston: Kluwer Academic.

- Towell, Geoffrey, and Ellen M. Voorhees. 1998. Disambiguating highly ambiguous words. *Computational Linguistics* 24:125-146.
- Trask, Robert Lawrence. 1993. *A dictionary of grammatical terms in linguistics*. London: Routledge.
- van Halteren, Hans, Jakub Zavrel, and Walter Daelemans. 1998. Improving data driven wordclass tagging by system combination. In *ACL 36/COLING 17*, pp. 491-497.
- van Riemsdijk, Henk, and Edwin Williams. 1986. *Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.
- van Rijsbergen, C. J. 1979. *Information Retrieval*. London: Butterworths. Second Edition.
- Velardi, Paola, and Maria Teresa Pazienza. 1989. Computer aided interpretation of lexical cooccurrences. In *ACL 27*, pp. 185-192.
- Viegas, Evelyne, Boyan Onyshkevych, Victor Raskin, and Sergei Nirenburg. 1996. From submit to submitted via submission: On lexical rules in large-scale lexicon acquisition. In *ACL 34*, pp. 32-39.
- Viterbi, A. J. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* IT-13: 1260-269.
- Vogel, Stephan, Hermann Ney, and Christoph Tillmann. 1996. HMM-based word alignment in statistical translation. In *COLING 16*, pp. 836-841.
- Voutilainen, A. 1995. A syntax-based part of speech analyser. In *EACL 7*, pp. 157-164.
- Waibel, Alex, and Kai-Fu Lee (eds.). 1990. *Readings in Speech Recognition*. San Mateo, CA: Morgan Kaufmann.
- Walker, Donald E. 1987. Knowledge resource tools for accessing large text files. In Sergei Nirenburg (ed.), *Machine Translation: Theoretical and methodological issues*, pp. 247-261. Cambridge: Cambridge University Press.
- Walker, Donald E., and Robert A. Amsler. 1986. The use of machine-readable dictionaries in sublanguage analysis. In Ralph Grishman and Richard Kittredge (eds.), *Analyzing language in restricted domains: sublanguage description and processing*, pp. 69-84. Hillsdale, NJ: Lawrence Erlbaum.
- Walker, Marilyn A., Jeanne C. Fromer, and Shrikanth Narayanan. 1998. Learning optimal dialogue strategies: A case study of a spoken dialogue agent for email. In *ACL 36/COLING 17*, pp. 1345-1351.
- Walker, Marilyn A., and Johanna D. Moore. 1997. Empirical studies in discourse. *Computational Linguistics* 23:1-12.

- Wang, Ye-Yi, and Alex Waibel. 1997. Decoding algorithm in statistical machine translation. In *ACL 35/EACL 8*, pp. 366-372.
- Wang, Ye-Yi, and Alex Waibel. 1998. Modeling with structures in statistical machine translation. In *ACL 36/COLING 17*, pp. 1357-1363.
- Waterman, Scott A. 1995. Distinguished usage. In Branimir Boguraev and James Pustejovsky (eds.), *Corpus Processing for Lexical Acquisition*, pp. 143-172. Cambridge, MA: MIT Press.
- Weaver, Warren. 1955. Translation. In William N. Locke and A. Donald Booth (eds.), *Machine Translation of Languages: Fourteen Essays*, pp. 15-23. New York: John Wiley & Sons.
- Webster, Mort, and Mitch Marcus. 1989. Automatic acquisition of the lexical semantics of verbs from sentence frames. In *ACL 27*, pp. 177-184.
- Weinberg, Sharon L., and Kenneth P. Goldberg. 1990. *Statistics for the behavioral sciences*. Cambridge: Cambridge University Press.
- Weischedel, Ralph, Marie Meteer, Richard Schwartz, Lance Ramshaw, and Jeff Palmucci. 1993. Coping with ambiguity and unknown words through probabilistic models. *Computational Linguistics* 19:359-382.
- Wiener, Erich, Jan Pedersen, and Andreas Weigend. 1995. A neural network approach to topic spotting. In *Proc. SDAIR 95*, pp. 317-332, Las Vegas, NV.
- Wilks, Yorick, and Mark Stevenson. 1998. Word sense disambiguation using optimized combination of knowledge sources. In *ACL 36/COLING 17*, pp. 1398-1402.
- Willett, Peter. 1988. Recent trends in hierarchic document clustering: A critical review. *Information Processing & Management* 24:577-597.
- Willett, P., and V. Winterman. 1986. A comparison of some measures for the determination of inter-molecular structural similarity. *Quantitative Structure-Activity Relationships* 5:18-25.
- Witten, Ian H., and Timothy C. Bell. 1991. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory* 37:1085-1094.
- Wittgenstein, Ludwig. 1968. *Philosophical Investigations [Philosophische Untersuchungen]*, 3rd edition. Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.
- Wong, S. K. M., and Y. Y. Yao. 1992. An information-theoretic measure of term specificity. *Journal of the American Society for Information Science* 43:54-61.
- Wood, Mary McGee. 1993. *Categorial Grammars*. London: Routledge.
- Wolf, Henry Bosley (ed.). 1973. *Webster's new collegiate dictionary*. Springfield, MA: G. & C. Merriam Co.

- Wu, Dekai. 1994. Aligning a parallel English-Chinese corpus statistically with lexical criteria. In *ACL 32*, pp. 80-87.
- Wu, Dekai. 1995. Grammarless extraction of phrasal examples from parallel texts. In *Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*.
- Wu, Dekai. 1996. A polynomial-time algorithm for statistical machine translation. In *ACL 34*, pp. 152-158.
- Wu, Dekai, and Hongsing Wong. 1998. Machine translation with a stochastic grammatical channel. In *ACL 36/COLING 17*, pp. 1408-1415.
- Yamamoto, Mikio, and Kenneth W. Church. 1998. Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus. In *WVLC 6*, pp. 28-37.
- Yang, Yiming. 1994. Expert network: Effective and efficient learning from human decisions in text categorization and retrieval. In *SIGIR '94*, pp. 13-22.
- Yang, Yiming. 1995. Noise reduction in a statistical approach to text categorization. In *SIGIR '95*, pp. 256-263.
- Yang, Yiming. 1998. An evaluation of statistical approaches to text categorization. *Information Retrieval*.
- Yarowsky, David. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *COLING 14*, pp. 454-460.
- Yarowsky, David. 1994. Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In *ACL 32*, pp. 88-95.
- Yarowsky, David. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *ACL 33*, pp. 189-196.
- Youmans, Gilbert. 1991. A new tool for discourse analysis: The vocabulary-management profile. *Language* 67:763-789.
- Younger, D. H. 1967. Recognition and parsing of context free languages in time n^3 . *Information and Control* 10:198-208.
- Zavrel, Jakub, and Walter Daelemans. 1997. Memory-based learning: Using similarity for smoothing. In *ACL 35/EACL 8*, pp. 436-443.
- Zavrel, Jakub, Walter Daelemans, and Jorn Veenstra. 1997. Resolving PP attachment ambiguities with memory-based learning. In *Proceedings of the Workshop on Computational Natural Language Learning*, pp. 136-144, Somerset, NJ. Association for Computational Linguistics.
- Zernik, Uri. 1991a. Introduction. In *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pp. 1-26. Hillsdale, NJ: Lawrence Erlbaum.

- Zernik, Uri. 1991b. Train1 vs. train2: Tagging word sense in corpus. In Uri Zernik (ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pp. 91-112. Hillsdale, NJ: Lawrence Erlbaum.
- Zipf, George Kingsley. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40:1-95.
- Zipf, George Kingsley. 1935. *The Psycho-Biology of Language*. Boston, MA: Houghton Mifflin.
- Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

This excerpt from

Foundations of Statistical Natural Language Processing.
Christopher D. Manning and Hinrich Schütze.
© 1999 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact cognetadmin@cognet.mit.edu.