

References

(Statistical Natural Language Processing)

- [Abney 99] Abney, S., Schapire, R.E. and Singer, Y., "Boosting Applied to Tagging and PP-attachment." In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, EMNLP-VLC*, pages 38--45, College Park, MD, 1999.
- [Abney 02] Abney, Steven, "Bootstrapping." *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, ACL-2002*, 2002.
- [Aho 86] Aho, A. V., R. Sethi, and J. D. Ullman, "Compilers: Principles, Techniques, and Tools", Addison-Wesley, USA, 1986.
- [Amari 67] Amari, S., "A Theory of Adaptive Pattern Classifiers", *IEEE Transaction on Electronic Computers*, Vol. EC-16, pp. 299-307, June 1967.
- [Bahl 83] Bahl, L. R., F. Jelinek, and R. Mercer, "A Maximum Likelihood Approach to Continuous Speech Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 2, pp. 179-190, March 1993.
- [Berger 96] Berger, A. L., Della Pietra, S. A., & Della Pietra, V. J., "A Maximum Entropy Approach to Natural Language Processing." *Computational Linguistics*. Vol. 22, No. 1, pp. 39-71. 1996.
- [Bickel 77] Bickel, P. J. and K. A. Doksum, "Mathematical Statistics: Basic Ideas and Selected Topics", Holden-Day Inc., USA, 1977.
- [Blahut 87] Blahut, Richard E., *Principles and Practice of Information Theory*, Addison-Wesley, USA, 1987.
- [Blum 98] Blum, A., & Mitchell, T., "Combining labeled and unlabeled data with co-training." In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT '98)*, pp. 92-100, 1998.
- [Brants 00] Brants, Thorsten, "TnT – A statistical Part-of-Speech tagger." In *Proceedings of the 6th Conference on Applied Natural Language Processing*, pp. 224–231, Seattle, Washington, 29 April – 4 May 2000.
- [Breiman 84] Breiman, L., J. H. Friedman, R.A. Olshen and C. J. Stone, *Classification And Regression Trees*, Wadsworth Inc., CA, USA, 1984.
- [Brill 94] Brill, Eric, "Some Advances in Transformation-Based Part of Speech Tagging", *Proceedings AAAI94*, 1994.
- [Brill 95] Brill, Eric, "Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging", *Computational Linguistics*, vol. 21, no. 4, pp. 543-565, December 1995.
- [Brown 90] Brown, P. et al., "A Statistical Approach to Machine Translation," *Computational Linguistics*, vol. 16, no. 2, pp. 79-85, June 1990.
- [Brown 91] Brown, P. et al., "Word-Sense Disambiguation Using Statistical Methods," *Proceedings of 29th Annual Meeting of the Association for Computational Linguistics*, pp. 264-270, California, USA, June 18-21, 1991.
- [Brown 92] Brown, P. et al., "Class-Based n-gram Models of Natural Language," *Computational Linguistics*, vol. 18, no. 4, pp. 467-479, December, 1992.
- [Burges 98] Burges, Christopher J.C., "A tutorial on support vector machines for pattern recognition." *Data Mining and Knowledge Discovery*, 2(2):121--167, June, 1998.
- [Carletta 96] Carletta, J., "Assessing Agreement on Classification Tasks: The Kappa Statistic," *Computational Linguistics*, vol. 22, no. 2, pp. 249-254, June, 1996.
- [Castelli 95] Castelli, V., and Cover, T. M., "On the exponential value of labeled samples." *Pattern Recognition Letters*, 16 (1), 105-111, 1995.
- [Chang 92] Chang, J.-S., Y.-F. Luo and K.-Y. Su, "GPSM: A Generalized Probabilistic Semantic Model for Ambiguity Resolution," *Proceedings of ACL-92*, pp. 177--184, 30th Annual

Meeting of the Association for Computational Linguistics, University of Delaware, Newark, DE, USA, 28 June--2 July, 1992.

- [Chang 93] Chang, J.-S. and K.-Y. Su, "A Corpus-Based Statistics-Oriented Transfer and Generation Model for Machine Translation," *Proceedings of TMI-93*, pp. 3-14, 1993.
- [Chang 97a] Chang, J.-S., *Automatic Lexicon Acquisition and Precision-Recall Maximization for Untagged Text Corpora*, Ph.D. dissertation, Department of Electrical Engineering, National Tsing-Hua University, Hsinchu, Taiwan, R.O.C., July, 1997.
- [Chang 97b] Chang, J.-S. and Keh-Yih Su, "An Unsupervised Iterative Method for Chinese New Lexicon Extraction," *International Journal of Computational Linguistics & Chinese Language Processing*, vol. 2, no. 2, pp. 97-148, August, 1997.
- [Charniak 97] Charniak, E., "Statistical Parsing with a context-free grammar and word statistics", In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pp. 598-603, AAAI Press/MIT Press, Menlo Park, CA, 1997.
- [Charniak 99] Charniak, E., *A maximum-entropy-inspired parser* (Technical Report CS99-12). Department of Computer Science, Brown University, RI, 1999.
- [Charniak 00] Charniak, E. "A maximum-entropy-inspired parser." In *Proceedings of the 2000 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 132-139, ACL, New Brunswick NJ, 2000.
- [Charniak 01a] Charniak, Eugene, "Unsupervised learning of name structure from coreference data," *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics*, pp 48-54, 2001.
- [Charniak 01b] Charniak, Eugene, "Immediate-Head Parsing for Language Models", *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001.
- [Chen 91] Chen, S.-C., J.-S. Chang, J.-N. Wang and K.-Y. Su, "ArchTran: A Corpus-Based Statistics-Oriented English-Chinese Machine Translation System," *Proceedings of Machine Translation Summit III*, pp. 33--40, Washington, D.C., USA, July 1--4, 1991.
- [Chiang 92a] Chiang, T.-H., Y.-C. Lin and K.-Y. Su, "Syntactic Ambiguity Resolution Using A Discrimination and Robustness Oriented Adaptive Learning Algorithm", *Proceedings of COLING-92*, vol. I, pp. 352--358, 14th Int. Conference on Computational Linguistics, Nantes, France, July 23--28, 1992.
- [Chiang 92b] Tung-Hui Chiang, Jing-Shin Chang, Ming-Yu Lin and Keh-Yih Su, "Statistical Models for Word Segmentation and Unknown Word Resolution," *Proceedings of ROCLING-V*, pp. 123-146, Taipei, Taiwan, R.O.C., 1992.
- [Chiang 95] Chiang, Tung-Hui, Yi-Chung Lin, and Keh-Yih Su, "Robust Learning, Smoothing, and Parameter Tying on Syntactic Ambiguity Resolution", *Computational Linguistics*, Vol. 21, No. 3, 1995, pp.322-349.
- [Chiang 96] Chiang, Tung-Hui, Yi-Chung Lin, and Keh-Yih Su, "On Jointly Learning the Parameters in a Character-Synchronous Integrated Speech and Language Model", *IEEE Transactions on Speech and Audio Processing*, Vol. 4, No. 3, May 1996, pp. 167-189.
- [Church 88] Church, K., "A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text," *ACL Proc. 2nd Conf. on Applied Natural Language Processing*, pp. 136-143, Austin, Texas, USA, 9-12 Feb. 1988.
- [Church 89] Church, K. and P. Hanks, "Word Association Norms, Mutual Information, and Lexicography," *Proc. 27th Annual Meeting of the ACL*, pp. 76-83, University of British Columbia, Vancouver, British Columbia, Canada, 26-29 June 1989.
- [Collins 96] Collins, M., "A New Statistical Parser Based on Bigram Lexical Dependencies," *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 184-191, 1996.
- [Collins 97] Collins, M. J. "Three generative lexicalized models for statistical parsing." In *Proceedings of the 35th Annual Meeting of the ACL*, pp. 16-23, 1997.
- [Collins 99a] Collins, M. J., *Head-Driven Statistical Models for Natural Language Parsing*. University of Pennsylvania, Ph.D. Dissertation, 1999.
- [Collins 99b] Collins, M. and Y. Singer. "Unsupervised models for named entity classification." In *Proceedings of EMNLP/WVLC*, pp. 100-110, 1999.
- [Cristianini 00] Cristianini, Nello and John Shawe-Taylor, *AN INTRODUCTION TO SUPPORT*

VECTOR MACHINES (and other kernel-based learning methods), Cambridge University Press, 2000.

- [Cutting 92] Cutting, D., J. Kupiec, J. Pedersen, and P. Sibun. "A practical part-of-speech tagger." In *Proc. of the 3rd Conference on Applied Natural Language Processing*, pages 133--140, Trento, Italy, 1992.
- [Dagon 91] Dagan, Ido, Alon Itai, and Ulrike Schwall, "Two Languages Are More Informative Than One", *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp. 130-137, UC Berkeley, CA, USA, 18-21 June 1991.
- [Dasgupta 01] Dasgupta, Sanjoy, Michael Littman and David McAllester, "PAC Generalization Bounds for Co-Training", *Neural Information Processing Systems, NIPS 01*, Vancouver, British Columbia, Canada, 2001.
- [Dempster 77] Dempster, A. P., N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the *EM* Algorithm", *Journal of the Royal Statistical Society*, 39 (b), pp. 1-38, 1977.
- [Devijver 82] Devijver, P. A. and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall International, Inc., London, 1982.
- [Duda 73] Duda, Richard O., and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons Inc., USA, 1973.
- [Dudewicz 88] Dudewicz, Edward J. and Satya N. Mishra, *Modern Mathematical Statistics*, John Wiley & Sons Inc., USA, 1988.
- [Fienberg 72] Fienberg, S. E. and P. W. Holland, "On the Choice of Flattening Constants for Estimating Multinomial Probabilities." *Journal of Multivariate Analysis*, Vol. 2, pp. 127-134, 1972.
- [Freund 99] Freund, Yoav and Robert E. Schapire, "A short introduction to boosting." *Journal of the Japanese Society for Artificial Intelligence*, 14(5):771-780, September, 1999. (Appearing in Japanese, translation by Naoki Abe.)
- [Garside 87] Garside, Roger, Geoffrey Leech and Geoffrey Sampson (eds.), *The Computational Analysis of English: A Corpus-Based Approach*, Longman Inc., New York, 1987.
- [Goldberg 89] Goldberg, David E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Company Inc., USA, 1989.
- [Goldman 00] Goldman, Sally, and Yah Zhou. "Enhancing supervised learning with unlabeled data." In *Proceedings of the 17th International Conference on Machine Learning*, pages 327-334, 2000.
- [Good 53] Good, I. J., "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, no. 3, 4, pp. 237--264, 1953.
- [Jelinek 97] F. Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, USA, 1997.
- [Joachim 99] Joachims, Thorsten, "Transductive inference for text classification using support vector machines." In *Proc. 16th International Conf. on Machine Learning*, pages 200—209, Morgan Kaufmann, San Francisco, CA, 1999.
- [Juang 92] Juang, Bing-Hwang, Shigeru Katagiri, "Discriminative Learning for Minimum Error Classification", *IEEE Transactions on Signal Processing*, Vol. 40, No. 12, December 1992, pp. 3043-3054.
- [Jurafsky 00] Jurafsky, Daniel & James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, USA, 2000.
- [Katagiri 91] Katagiri, S., C. H. Lee, and B. H. Juang, "New Discriminative Training Algorithm Based on the Generalized Probabilistic Decent Method," *Proceedings of 1991 IEEE Workshop Neural Networks for Signal Processing*, pp. 299-308, Piscataway, NJ, Aug. 1991.
- [Katz 87] Katz, S. M., "Estimation of Probabilities From Sparse Data for the Language Model Component of a Speech Recognizer," *IEEE Transactions on Acoustics, Speech and Signal Processing*, No. 35, pp. 400--401, 1987.
- [Kudo 00] Kudo Taku, Yuji Matsumoto, "Use of Support Vector Learning for Chunk Identification," *Proc. of CoNLL-2000*, 2000.
- [Kudo 01] Kudo, Taku, Yuji Matsumoto, "Chunking with Support Vector Machines," *Proc. of*

NAACL-2001, 2001.

- [Kuang 02] Kuang, Tz-Liang, Keh-Yih Su, "A Robust Cross-Style Bilingual Sentences Alignment Model", *Proc. of COLING-2002*, Taipei, 2002.
- [Lin 95] Lin, Yi-Chung, Tung-Hui Chiang, and Keh-Yih Su, "The Effects of Learning, Parameter Tying and Model Refinement for Improving Probabilistic Tagging", *Computer Speech and Language*, No. 9, 1995, pp. 37-61.
- [Lin 99] Lin, Yi-Chung and Keh-Yih Su, "A Level-Synchronous Approach to Ill-formed Sentence Parsing and Error Recovery," *International Journal of Computational Linguistics & Chinese Language Processing*, vol. 4, no. 1, pp. 25-58, February, 1999.
- [Liu 90] Liu, C.-L., J.-S. Chang and K.-Y. Su, "The Semantic Score Approach to the Disambiguation of PP Attachment Problem," *Proceedings of ROCLING-III*, pp. 253-270, National Tsing-Hua Univ., Taipei, R.O.C., Sept. 21-23, 1990.
- [Manning 99] Manning, Christopher D. and Hinrich Schutze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999.
- [Marcu 00] Marcu, Daniel, *The Theory and Practice of Discourse Parsing and Summarization*, MIT Press, 2000.
- [Mitchell 97] Mitchell, Tom, *Machine Learning*, McGraw Hill, 1997.
- [Montgomery 01] Montgomery, Douglas C., *Design and Analysis of Experiments*, 5th Edition, John Wiley & Son Inc., 2001.
- [Mueller 02] Mueller, Christoph, Stefan Rapp and Michael Strube, "Applying Co-Training to Reference Resolution", *Proc. of ACL-2002*, pp. 352-359, 2002.
- [Murata 01] Murata, Masaki, Qing Ma, and Hitoshi Isahara, "Part of Speech Tagging in Thai Language Using Support Vector Machine", *Proc. of the Second Workshop on Natural Language Processing and Neural Networks (NLPNN 2001)*, NLPNS-01 post-conference workshop, Tokyo, Japan, Nov. 30, 2001.
- [Muslea 00] Muslea, Ion, Steven Minton, and Craig A. Knoblock. "Selective sampling with redundant views." In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, AAAI-2000, pages 621-626. Menlo Park, CA: AAAI Press, 2000.
- [Nakagawa 01] Nakagawa, Tetsuji, Taku Kudoh and Yuji Matsumoto, "Unknown Word Guessing and Part-of-Speech Tagging Using Support Vector Machines", *Proc. of NLPNS-01*, pp. 325-331, 2001.
- [Nakagawa 02] Nakagawa, Tetsuji, Taku Kudo, Yuji Matsumoto, "Revision Learning and its Application to Part-of-Speech Tagging", *Proc. of ACL-2002*, 2002.
- [Nigam 00a] Nigam, K., and R. Ghani. "Analyzing the effectiveness and applicability of co-training." In *Proceedings of the Ninth International Conference on Information and Knowledge Management*, pp. 86-93, 2000.
- [Nigam 00b] Nigam, K., A. McCallum, S. Thrun, and T. Mitchell. "Text classification from labeled and unlabeled documents using EM." *Machine Learning*, 39(2/3):103-134, 2000.
- [Papoulis 84] Papoulis A., *Probability, Random Variable, and Stochastic Processes*, 2nd Ed. McGraw-Hill, USA. 1984.
- [Papoulis 90] Papoulis A., *Probability & Statistics*, Prentice-Hall, USA. 1990.
- [Pearl 88] Pearl, Judea, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, USA, 1988.
- [Pierce 01] Pierce, David, and Claire Cardie, "Limitations of Co-Training for Natural Language Learning from Large Datasets", *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP-2001)*, pp. 1-9, 2001.
- [Pinker 00] Pinker, Steven, *The Language Instinct: How the Mind Creates Language*, HarperCollins Publishers, 2000.
- [Quinlan 93] Quinlan, J. Ross, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, CA, USA, 1993.
- [Rabiner 86] Rabiner, L. R., J. G. Wilpon, and B. H. Juang, "A Segmental k-Means Training Procedure for Connected Word Recognition," *AT&T Tech. Journal*, Vol. 65, No. 3, pp.21-31, May-June 1986.
- [Rabiner 89] Rabiner, L. R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceeding of IEEE*, pp.1404-1413, Vol. 77, No. 2, Feb. 1989.
- [Rabiner 93] Rabiner, Lawrence and Biing-Hwang Juang, *Fundamentals of Speech*

Recognition, Prentice-Hall, USA, 1993.

- [Ratnaparkhi 96] Ratnaparkhi, Adwait, "A maximum entropy model for part-of-speech tagging." In Eric Brill and Kenneth Church, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 133--142, University of Pennsylvania, Philadelphia, Pa., May 1996.
- [Rennie 01] Rennie, Jason D. M., Ryan Rifkin, "Improving multi-class text classification with the support vector machine," *AI Memo, AIM-2001-026*, MIT, 2001.
- [Resnik 92] Resnik, P., "Probabilistic tree-adjoining grammars as a framework for statistical natural language processing." In *Proc. of COLING '92*, volume 2, pages 418--424, Nantes, France, 1992.
- [Resnik 99] Resnik, Philip, 1999, "Mining the Web for Bilingual Text", *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*, pp. 527-534, University of Maryland, Maryland, USA, June 20-26, 1999.
- [Riloff 99] Riloff, E., and R. Jones. "Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping." In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pp. 474-479, 1999.
- [Rosenfeld 96] Rosenfeld, R., "A maximum entropy approach to adaptive statistical language modeling." *Computer, Speech, and Language*, Vol. 10, No. 3, pp. 187-228, 1996.
- [Sarkar 2001] Sarkar, Anoop. "Applying Co-training Methods to Statistical Parsing." In *Proc. of the 2nd NAACL*, pp. 175--182, Pittsburgh, PA, 2--7 June, 2001.
- [Schabes 92] Schabes, Y., "Stochastic lexicalized tree-adjoining grammars." In *Proc. of COLING '92*, volume 2, pages 426--432, Nantes, France, 1992.
- [Schabes 93] Y. Schabes and R. C. Waters. "Stochastic Lexicalized Context-Free Grammar." In *Proceedings of the 3rd International Workshop on Parsing Technologies (IWPT'93)*, pages 257--266, Tilburg, The Netherlands, 1993.
- [Sowa 99] Sowa, J. F., *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, 1999.
- [Su 88] Su, K.-Y. and J.-S. Chang, "Semantic and Syntactic Aspects of Score Function," *Proc. of COLING-88*, vol. 2, pp. 642--644, 12th Int. Conf. on Computational Linguistics, Budapest, Hungary, August 22--27, 1988.
- [Su 89] Su, K.-Y., M.-H. Su and L.-M. Kuan., "Smoothing Statistic Databases in a Machine Translation System," *Proceedings of ROCLING-II*, pp. 333-347, Academia Sinica, Taipei, Taiwan, R.O.C., Sept. 22-24, 1989.
- [Su 90] Su, K.-Y. and J.-S. Chang, "Some Key Issues in Designing MT Systems," *Machine Translation*, vol. 5, no. 4, pp. 265-300, 1990.
- [Su 91a] Su, Keh-Yih and Chin-Hui Lee, "Robustness and Discrimination Oriented Speech Recognition Using Weighted HMM and Subspace Projection Approaches", *Proceedings of IEEE 1991 International Conference of Acoustic, Speech, and Signal Processing (ICASSP-91)*, vol. 1, pp. 541-544, Toronto, Ontario, Canada, May 14-17, 1991.
- [Su 91b] Su, K.-Y., J.-N. Wang, M.-H. Su and J.-S. Chang, "GLR Parsing with Scoring," In M. Tomita (ed.), *Generalized LR Parsing*, Chapter 7, pp. 93-112, Kluwer Academic Publishers, 1991.
- [Su 92] Su, Keh-Yih and Jing-Shin Chang, "Why Corpus-Based Statistics-Oriented Machine Translation," *Proceedings of TMI-92*, pp. 249-262, 4th International Conference on Theoretical and Methodological Issues in Machine Translation, Montreal, Canada, June 25-27, 1992.
- [Su 94] Su, Keh-Yih and Chin-Hui Lee, "Speech Recognition Using Weighted HMM and Subspace Projection Approaches", *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 1, Part 1, pp. 69-79, January 1994.
- [Su 95] Su, Keh-Yih, Jing-Shin Chang, and Yu-Liang Una Hsu, 1995, "A Corpus-based Two-Way Design for Parameterized MT Systems: Rationale, Architecture and Training Issues", *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95)*, Vol. 2, pp. 334-353, Leuven, Belgium, July 5-7, 1995.
- [Su 99] Su, Keh-Yih and Jing-Shin Chang, 1999, "A Customizable, Self-Learnable Parameterized MT Systems: The Next Generation", *Proceedings of the Machine Translation Summit VII International Conference*, pp. 182-188, Singapore, September 13-17, 1999.
- [Tou 74] Tou, J.T. and R.C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley,

Massachusetts, 1974.

- [Wahlster 00] W. Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, New York, Berlin, 2000.
- [Weston 98] Weston, J., and C. Watkins, "Multi-class support vector machines," *Technical Report, CSD-TR-98-04*, Royal Holloway University of London, May 20, 1998.
- [Winston 92] Winston, P.H, *Artificial Intelligence*, third edition, AddisonWesley, Massachusetts, USA, 1992.
- [Yarowsky 95] Yarowsky, D.. "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods." In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*. Cambridge, MA, pp. 189-196, 1995.