

References

- Abney, S (1997) Part-of-speech tagging and partial parsing. In: Young, S and Bloothoof, G (eds.) (1997) *Corpus-based methods in language and speech processing*. Dordrecht: Kluwer Academic Publishers.
- Anttila, A (1995) How to recognise subjects in English. In: Karlsson et al. (1995).
- Armstrong, S, Church, K, Isabelle, P, Manzi, S, Tzoukermann, E and Yarowsky, D (eds.) (1999) *Natural language processing using very large corpora*. Dordrecht: Kluwer Academic Publishers.
- Atwell, E (1987) Constituent-likelihood grammar. In: Garside, Leech and Sampson (1987).
- Atwell, E, Demetriou, G, Hughes, J, Schiffrin, A, Souter, C, and Wilcock, S (2000) A comparative evaluation of modern English corpus grammatical annotation schemes. In: *ICAME Journal*, 24: 7-23.
- Bahl, LR and Mercer, RL (1976) Part of speech assignment by a statistical decision algorithm. In: *IEEE International Symposium on Information Theory*, 88-89. Ronneby.
- Bailey, TG, ed. by Firth, JR and Harley, AH (1956) *Teach Yourself Urdu*. New York: David McKay Company.

- Baker, P, Burnard, L, McEnery, T, and Sebba, M (1999) Locating minority language speaking communities in the UK. MILLE Working Paper 3. Department of Linguistics and Modern English Language, Lancaster University.
- Barz, RK (1977) *An Introduction to Hindi and Urdu*. Canberra: Australian National University Press.
- Benello, J, Mackie, AW and Anderson, JA (1989) Syntactic category disambiguation with neural networks. In: *Computer Speech and Language*, 3: 203-217.
- Bharati, A, Chaitanya, V, Kulkarni, AP, Sharma, DP, and Sangal, R (2001). LERIL: Collaborative effort for creating lexical resources. Report submitted to the LESAL workshop at NCST, Mumbai, April 2001. Available on the internet at <http://www.emille.lancs.ac.uk/lesal.htm>
- Bhatia, TK (1987) *A History of the Hindi Grammatical Tradition*. Leifden: E. J. Brill.
- Bhatia, TK and Koul, A (2000) *Colloquial Urdu*. London: Routledge.
- Black, E, Jelinek, F, Lafferty, J, Mercer, R and Roukos, S (1992) Decision tree models applied to the labeling of text with parts-of-speech. In: *Darpa Workshop on Speech and Natural Language*. Harriman, NY.

- Blackwell, S (1987) Syntax versus orthography: Problems in the automatic parsing of idioms. In: Garside, Leech and Sampson (1987).
- Brill, E (1992) A simple rule-based part of speech tagger. In: *Proceedings of the Third Conference on Applied Natural Language Processing (ANLP'92)*. Trento.
- Brill, E (1994) Some advances in transformation-based part-of-speech tagger. In: *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI'94)*, Seattle, Washington.
- Brill, E (1995) Transformation-based error-driven learning and Natural Language Processing: a case study in part-of-speech tagging. In: *Computational Linguistics*, 21 (4): 543-565.
- Brill, E (1999) Corpus-based rules. In: van Halteren (1999a).
- Brill, E and Pop, M (1999) Unsupervised learning of disambiguation rules for part of speech tagging. In: Armstrong et al. (1999).
- Butt, M (1995) *The Structure of Complex Predicates in Urdu*. Stanford, California: CSLI Publications.
- Calzolari, N, McNaught, J, and Zampolli, A (1996) *Eagles Editor's Introduction*. Available on the internet at <http://www.ilc.pi.cnr.it/EAGLES96/edintro/edintro.html>

- Calzolari, N and Monachini, M (1996) *Multext: common specifications and notation for lexicon encoding*. Available on the internet at <http://www.lpl.univ-aix.fr/projects/multext/LEX/LEX1.html>
- Cardie, C (1996) Embedded machine learning systems for natural language processing: a general framework. In: Wermter, S Riloff, E, and Scheler, G (eds.) (1996) *Connectionist, statistical and symbolic approaches to learning for natural language processing*, lecture notes in artificial intelligence. Berlin: Springer.
- Cardona, G (1976) *Pāṇini: a survey of research*. The Hague: Mouton.
- Chae, Y-S and Choi, K-S (2000) Introduction of KIBS (Korean Information Base System) Project. In: Gavrilidou et al. (2000): vol. 3, 1731-1735.
- Chanod, J-P and Tapanainen, P (1995a) Tagging French – comparing a statistical and a constraint-based method. In: *Proceedings of the Seventh Conference of the European Chapter of the ACL*. Dublin: Association for Computational Linguistics.
- Chanod, J-P and Tapanainen, P (1995b) Creating a tagset, lexicon and guesser for a French tagger. In: *Proceedings of the European Chapter of the Association for Computational Linguistics EACL-SIGDAT Workshop*. Dublin.

- Charniak, E, Hendrickson, C, Jacobson, N and Perkowitz, M (1993) Equations for part of speech tagging. In: *Proceedings of the Eleventh National Conference on Artificial Intelligence*. Menlo Park: AAAI Press/MIT Press.
- Chomsky, N (1957) *Syntactic structures*. Mouton.
- Church, K (1988) A stochastic parts program and noun phrase parser for unrestricted text. In: *Proceedings of the second conference on Applied Natural Language Processing, ACL*.
- Church, K and Mercer, RL (1993) Introduction to a special issue on computational linguistics using large corpora. In: *Computational Linguistics* 19 (1): 1-24.
- Cloeren, J (1999) Tagsets. In: van Halteren (1999a).
- Comrie, B (ed.) (1990) *The Major Languages of South Asia, the Middle East and Africa*. London: Routledge.
- CRATER (1996) Mapping of the French tagset to the EAGLES guidelines. CRATER Deliverable Report 24. Internal report: Department of Linguistics and Modern English Language, University of Lancaster.
- Cutting, D, Kupiec, J, Pederson, J, and Sibun, P (1992) A practical part-of-speech tagger. In: *Proceedings of the third conference on Applied Natural Language Processing, ACL*.

- Daelemans, W (1999) Machine learning approaches. In: van Halteren (1999a).
- Daelemans, W, Zavrel, J, Berck, P and Gillis, S (1996) MBT: a memory-based part of speech tagger-generator. In Ejerhed, E and Dagan, I (eds.) (1996) *Proceedings of the Fourth Workshop on Very Large Corpora, Copenhagen*.
- de Marcken, CG (1990) Parsing the LOB corpus. In: *Proceedings of the 1990 Conference of the Association for Computational Linguistics*, 243-251.
- DeRose, SJ (1988) Grammatical category disambiguation by statistical optimization. In: *Computational Linguistics*, 14(1): 31-39.
- Dixon, RMW (1979) Ergativity. In: *Language* 55 (1): 59-138.
- Dixon, RMW (1994) *Ergativity*. Cambridge: Cambridge University Press.
- Eklund, R (1993) A probabilistic tagging module based on surface pattern matching. In: Eklund, R (ed.) (1993) *NODALIDA '93 - Proceedings of 9:e Nordiska Datalingvistikdagarna*. Stockholm: Stockholm University.
- El-Beze, M and Merialdo, B (1999) Hidden Markov models. In: van Halteren (1999a).

- Ellegård, A (1978) *The syntactic structure of English texts : a computer-based study of four kinds of texts in the Brown University Corpus*. Gothenburg Studies in English, 43. Gothenburg: Gothenburg University.
- Francis, WN and Kučera, H (1982) *Frequency analysis of English usage*. Boston: Houghton Mifflin.
- Garside, R (1987) The CLAWS word-tagging system. In: Garside, Leech and Sampson (1987).
- Garside, R, Leech, G and McEnery A (eds.) (1997) *Corpus annotation: linguistic information from computer text corpora*. London: Longman.
- Garside, R, Leech, G and Sampson, G (eds.) (1987) *The computational analysis of English*. London: Longman.
- Garside, R and Smith, N (1997) A hybrid grammatical tagger: CLAWS4. In: Garside, Leech and McEnery (1997).
- Gavrilidou, M, Carayannis, G, Markantonatou, S, Piperidis, S and Stainhaouer, G (eds.) (2000) *Second International Conference on Language Resources and Evaluation: Proceedings*. 3 volumes. Athens: European Language Resources Association.

Greenbaum, S (ed.) (1996) *Comparing English Worldwide: the International Corpus of English*. Oxford: Clarendon Press.

Greenbaum, S and Yibin, N (1996) About the ICE Tagset. In: Greenbaum (1996).

Greene, BB and Rubin, GM (1971) *Automatic grammatical tagging of English*. Providence, Rhode Island: Brown University Department of Linguistics.

van Halteren, H (ed.) (1999a) *Syntactic wordclass tagging*. Dordrecht: Kluwer Academic Publishers.

van Halteren, H (1999b) Performance of taggers. In: van Halteren (1999a).

van Halteren, H and Oostdijk, N (1993) The TOSCA analysis system. In: Aarts, J, de Haan, P and Oostdijk, N (1993) *English language corpora: design, analysis and exploitation. Papers from the thirteenth International Conference on English Language Research on Computerised Corpora, Nijmegen 1992*. Amsterdam: Rodopi.

van Halteren, H and Voutilainen, A (1999) Automatic taggers: an introduction. In: van Halteren (1999a).

Haq, A (2001) *English-Urdu/Urdu-English combined dictionary*. Previously published in two volumes. New Delhi: Star.

- Harris, ZS (1962) *String analysis of sentence structure*. The Hague: Mouton.
- Heikkilä, J (1995) A TWOL-based lexicon and feature system for English. In: Karlsson et al. (1995).
- Hetzron, R (1990) Semitic languages. In: Comrie (1990).
- Hindle, D (1989) Acquiring disambiguation rules from text. In: *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL-89), Vancouver*.
- Ide, N and Véronis, J (1994) MULTTEXT: multilingual test tools and corpora. In: *Proceedings of the 14th International Conference on Corpus Linguistics, COLING '94, Kyoto, Japan, volume 1, 588-592*.
- Jelinek, F (1985) Markov source modeling of text generation. In: Skwirzinski, JK (ed.) (1983) *Impact of Processing Techniques on Communication: Proceedings of the NATO Advanced Study Institute 1983*. Dordrecht: Nijhoff.
- Joshi, AK and Hopely, P (1997) A parser from antiquity. In: *Natural Language Engineering*, 2 (4): 291-294.
- Kachru, Y (1990) Hindi-Urdu. In: Comrie (1990).

- Kahrel, P, Barnett, R and Leech, G (1997) Towards cross-linguistic standards or guidelines for the annotation of corpora. In: Garside, Leech and McEnery (1997).
- Karlsson, F (1995a) Designing a parser for unrestricted text. In: Karlsson et al. (1995).
- Karlsson, F (1995b) The formalism and environment of Constraint Grammar Parsing. In: Karlsson et al. (1995).
- Karlsson, F, Voutilainen, A, Heikkilä, J and Anttila, A (eds.) (1995) *Constraint Grammar: a language-independent system for parsing unrestricted text*. Berlin: Mouton de Gruyter.
- Katamba, F (1993) *Morphology*. London: Macmillan.
- Kaye, AS (1990) Arabic. In: Comrie (1990).
- Kellogg, SH (1875) *A grammar of the Hindí language, in which are treated the High Hindí, Braj, and the eastern Hindí of the Rámáyan of Tulsí Dás*. (Reprinted 1965.) London: Routledge and Kegan Paul.
- Khoja, S, Garside, R, and Knowles, G (2001) A tagset for the morphosyntactic tagging of Arabic. Paper given at the Corpus Linguistics 2001 conference, Lancaster.

- Klein, S and Simmons, RF (1963) A computational approach to grammatical coding of English words. In: *Journal of the Association for Computing Machinery*, 10: 334-347.
- Koskenniemi, K (1990) Finite-state parsing and disambiguation. In: Karlgren, H (ed) (1990) *COLING-90*. Helsinki University.
- Kupiec, J 1992 Robust part of speech tagging using a hidden Markov model. *Journal of Computer Speech and Language*, 6 (3): 225-242.
- Leech, G (1987) General introduction. In: Garside, Leech and Sampson (1987).
- Leech, G (1991) The state of the art in corpus linguistics. In: Aijmer, K and Altenberg, B (eds.) (1991) *English corpus linguistics: studies in honour of Jan Svartvik*. London: Longman.
- Leech, G (1997a) Introducing corpus annotation. In: Garside, Leech and McEnery (1997).
- Leech, G (1997b) Grammatical tagging. In: Garside, Leech and McEnery (1997).
- Leech, G, Garside, R and Bryant, M (1994) CLAWS4: The tagging of the British National Corpus. In: *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)* Kyoto, Japan: 622-628.

Leech, G and Smith, N (1999) The use of tagging. In: van Halteren (1999a).

Leech, G and Wilson, A (1999) Standards for tagsets. In: van Halteren (1999a).

(Edited version of *EAGLES Recommendations for the Morphosyntactic Annotation of Corpora* (1996): available on the internet at <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html> .)

Magerman, D (1995) Statistical decision tree models for parsing. In: *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, Cambridge, MA*.

Marcus, M, Santorini, B and Marcinkiewicz, MA (1993) Building a large annotated corpus of English: the Penn Treebank. In: *Computational Linguistics*, 19(2): 313-330.

Marshall, I (1987) Tag selection using probabilistic methods. In: Garside, Leech and Sampson (1987).

Masica, CP (1991) *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

McEnery, T (1995) *Computational pragmatics – probability, deeming and uncertain references*. Doctoral thesis: Department of Linguistics, Lancaster University.

McEnery, T (1996) A mapping of the CLAWS7 tagset onto the obligatory features of the EAGLES morphosyntactic annotation group guidelines. CRATER project Deliverable Report 25. Internal report: Department of Linguistics and Modern English Language, University of Lancaster.

McEnery, T, Baker, P, and Burnard, L (2000) Corpus resources and minority language engineering. In: Gavrilidou et al. (2000): vol. 2, 801-806.

McEnery, T, Baker, P and Hardie, A (2000) Swearing and abuse in modern British English. In: Lewandowska-Tomaszczyk, B and Melia, PJ (eds.) (2000) *PALC'99: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang.

McEnery, T and Wilson, A (1996) *Corpus linguistics*. Edinburgh: Edinburgh University Press.

McEnery, T, Wilson, A, Sánchez-León, F and Nieto-Serrano, A (1997) Multilingual resources for European languages: Contributions of the CRATER project. In: *Literary and Linguistic Computing* 12 (4): 219-226.

McGregor, RS (1972) *Outline of Hindi grammar*. Oxford: Clarendon Press.

Meriello, B (1994) Tagging English text with a probabilistic model. In: *Computational Linguistics*, 20 (2): 155-171.

- Misra, VN (1966) *The descriptive technique of Pāṇini*. The Hague: Mouton.
- Nakamura, M, Maruyama, K, Kawabata, T and Shikano, K (1990) Neural network approach to word category prediction for English texts. In: Karlgren, H (ed.) (1990) *Proceedings of COLING 90, Helsinki*.
- Oriental Book Society (date unknown). *New edition popular Oxford combined dictionary*. Lahore: Oriental Book Society.
- Payne, JR (1990) Persian. In: Comrie (1990).
- Piao, S (2000) *Sentence and word alignment between Chinese and English*. PhD thesis: Department of Linguistics and Modern English Language, University of Lancaster.
- Piao, S and McEnery, T (2003) A tool for text comparison. In: Archer, D, Rayson, P, Wilson, A and McEnery, T (eds.) (2003) *Proceedings of the Corpus Linguistics 2003 conference*.
- Platts, JT (1884) *A Dictionary of Urdu, Classical Hindi, and English*. Oxford: Oxford University Press (reprinted 1965).
- Rajman, M, Lecomte, J and Paroubek, P (1997) *Format de description lexicale pour le français, partie 2: Description morpho-syntaxique*. GRACE GTR-3-2.1. Available on the internet at <http://www.limsi.fr/TLP/grace/doc/GTR-3-2.1.ps>

Roche, E and Schabes, Y (1995) Deterministic part-of-speech tagging with finite-state transducers. In: *Computational Linguistics*, 21 (2): 227-253.

Ruhlen, M (1987) *A guide to the world's languages. Volume 1: classification*.
London: Edward Arnold.

Sampson, G (1985). *Writing Systems*. London: Hutchinson.

Sampson, G (1987) Alternative grammatical coding systems. In: Garside, Leech and Sampson (1987).

Sampson, G (1995) *English for the computer: the SUSANNE Corpus and Analytic Scheme*. Oxford: Clarendon Press.

Sánchez León, F and Nieto Serrano, AF (1997) Retargeting a tagger. In: Garside, Leech and McEnery (1997).

Schmid, H. (1994) Part-of-speech tagging with neural networks. In *Proceedings of COLING-94, Kyoto*.

Schmidt, RL (1999) *Urdu: an essential grammar*. London: Routledge.

- Schütze, H and Singer, Y (1994) Part of speech tagging using a variable memory Markov model. In *Proceedings of the Association for Computational Linguistics*, Las Cruces, NM.
- Short, M, Semino, E and Culpeper, J (1996) Using a corpus for stylistics research: speech presentation. In: Thomas, J and Short, M (eds.) (1996) *Using corpora for language research*. London: Longman.
- Smith, N (1997) Improving a tagger. In: Garside, Leech and McEnery (1997).
- Smith, N and McEnery, T (2000) Can we improve part-of-speech tagging by inducing probabilistic part-of-speech annotated lexicons from large corpora? In: Nicolov, N and Mitkov, R (eds.) (2000) *Recent advances in natural language processing II: Selected papers from RANLP'97*. Amsterdam: John Benjamins.
- Stolz, WS, Tannenbaum, PH and Carstensen, FV (1965) A stochastic approach to the grammatical coding of English. In: *Communications of the ACM*, 8 (6): 399-405.
- Svartvik, J (1990) Tagging and parsing on the TESS project. In: Svartvik, J (ed.) (1990) *The London-Lund Corpus of Spoken English: description and research*. Lund: Lund University Press.

- Tapanainen, P and Voutilainen, A (1994) Tagging accurately – don't guess if you know. In: *Proceedings of the Fourth Conference on Applied Natural Language Processing*. Stuttgart.
- Véronis, J and Khouri, L (1995) *Etiquetage grammatical multilingue: modèle*. Draft document: available on the internet at <http://www.lpl.univ-aix.fr/projects/multext/LEX/LEX2.html>
- Voutilainen, A (1995) Morphological disambiguation. In: Karlsson et al. (1995).
- Voutilainen, A (1999a) Orientation. In: van Halteren (1999a).
- Voutilainen, A (1999b) A short history of tagging. In: van Halteren (1999a).
- Voutilainen, A (1999c) Hand-crafted rules. In: van Halteren (1999a).
- Weischedel, R, Meteor, M, Schwartz, R, Ramshaw, L and Palmuzzi, J (1993) Coping with ambiguity and unknown words through probabilistic models. In: *Computational Linguistics*, 19 (2): 359-382.
- Whaley, LJ (1997) *Introduction to typology*. Thousand Oaks, CA: Sage.