

## Referências

- (Afonso, 2006) Susana Afonso. Árvores deitadas: Descrição do formato e das opções de análise na Floresta Sintáctica. 12 de Fevereiro de 2006. <http://www.linguateca.pt/Floresta/ArvoresDeitadas.doc>.
- (Afonso et al., 2002) Susana Afonso, Eckhard Bick, Renato Haber e Diana Santos. Floresta sintá(c)tica: um treebank para o português. Em Anabela Gonçalves e Clara Nunes Correia, editores, *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística*, APL 2001. Lisboa, Portugal. 2-4 de Outubro de 2002. p. 533–545.
- (Agichtein e Gravano, 2000) Eugene Agichtein e Luis Gravano. Snowball: Extracting Relations from Large Plain-Text Collections. Em *Proceedings of the Fifth ACM Conference on Digital Libraries*. San Antonio, TX, EUA. 2-7 de Junho de 2000. p. 85–94.
- (Almeida e Pinto, 1995) José João Almeida e Ulisses Pinto. Jspell – um módulo para análise léxica genérica de linguagem natural. Em *Actas do X Encontro Nacional da Associação Portuguesa de Linguística*. Évora, Portugal. 6-8 de Outubro de 1995. p. 1–15.
- (Almeida e Simões, 2006a) José João Almeida e Alberto Manuel Simões. Publishing multilingual ontologies: a quick way of obtaining feedback. Em Bob Martens e Milena Dobrevá, editores, *Digital spectrum : integrating technology and culture : proceedings of the International Conference on Electronic Publishing*, ELPUB2006. Bansko, Bulgária. Junho de 2006. p. 373–374.
- (Almeida e Simões, 2006b) José João Almeida e Alberto Manuel Simões. T2O - Recycling Thesauri into a Multilingual Ontology. Em Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik e Daniel Tapias, editores, *Proceedings of the 5th International Conference on Language Resources and Evaluation*, LREC'2006. Génova, Itália. 22-28 de Maio de 2006. p. 1466–1471.
- (Alves e Almeida, 2006) Edgar Alves e José João Almeida. Manual de utilizador do RENA. Relatório técnico. Universidade do Minho, Departamento de Informática. Julho de 2006.

- (Amitay et al., 2004) Einat Amitay, Nadav Har'El, Ron Sivan e Aya Soffer. Web-a-Where: Geotagging Web content. Em Mark Sanderson, Kalervo Järvelin, James Allan e Peter Bruza, editores, *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04. Sheffield, Reino Unido. 25-29 de Julho de 2004. p. 273–280.
- (Appelt et al., 1995) Douglas E. Appelt, Jerry R. Hobbs, John Bear, David Israel, Megumi Kameyama, Andy Kehler, David Martin, Karen Myers e Marby Tyson. SRI International FASTUS system MUC-6 test results and analysis. Em *Proceedings of the 6th Message Understanding Conference*, MUC-6. Columbia, MD, EUA. 6-8 de Novembro de 1995. p. 237–248.
- (Arévalo et al., 2002) Montserrat Arévalo, Xavier Carreras, Lluís Màrquez, Toni Martí, Lluís Padró e Maria José Simon. A proposal for wide-coverage Spanish named entity recognition. *Sociedad Española para el Procesamiento del Lenguaje Natural*. 28:63–80. Maio de 2002.
- (Baptista et al., 2006) Jorge Baptista, Fernando Batista, Nuno Mamede e Cristina Mota. Npro: um novo recurso para o processamento computacional do português. Em Joaquim Barbosa e Fátima Oliveira, editores, *Textos seleccionados do XXI Encontro da Associação Portuguesa de Linguística*. 2006.
- (Ben-Kiki et al., 2005) Oren Ben-Kiki, Clark Evans e Brian Ingerson. YAML specification. 2005. <http://yaml.org/spec/>.
- (Ben-Kiki et al., 2006) Oren Ben-Kiki, Clark Evans e Brian Ingerson. YAML cookbook. 2006. <http://yaml4r.sourceforge.net/cookbook/>.
- (Bick, 2000) Eckhard Bick. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Tese de doutoramento. Universidade de Aarhus. Aarhus University Press. Novembro de 2000.
- (Bick, 2003) Eckhard Bick. Multi-level NER for Portuguese in a CG framework. Em Nuno J. Mamede, Jorge Baptista, Isabel Trancoso e Maria das Graças Volpe Nunes, editores, *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003. Faro, Portugal, June 2003*. Springer. Berlin/Heidelberg. 2003. p. 118–125.
- (Bick, 2004) Eckhard Bick. A Named Entity Recognizer for Danish. Em Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa e Raquel Silva, editores, *Proceedings of LREC2004, the 4th International Conference on Language Resources and Evaluation*. Lisboa, Portugal. 26-28 de Maio de 2004. p. 305–308.
- (Bick, 2006a) Eckhard Bick. Functional Aspects in Portuguese NER. Em Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oli-

- veira e Maria Carmelita Dias, editores, *Computational Processing of the Portuguese Language: 7th International Workshop, PROPOR 2006. Itatiaia, Brazil, May 2006*. Springer. Berlin/Heidelberg. 2006. p. 80–89.
- (Bick, 2006b) Eckhard Bick. Functional Aspects on Portuguese NER. Encontro do HAREM. Porto, Portugal. Apresentação. 15 de Julho de 2006. <http://www.linguateca.pt/documentos/HAREM2006Bick.pdf>.
- (Bick et al., 2007) Eckhard Bick, Diana Santos, Susana Afonso e Rachel Marchi. Floresta Sintá(c)tica: Ficção ou realidade? Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. Lisboa, Portugal. 20 de Março de 2007. p. 291–300.
- (Bikel et al., 1997) Daniel M. Bikel, Scott Miller, Richard Schwartz e Ralph Weischedel. Nymble: a high performance learning name-finder. Em *Proceedings of the Fifth Conference on Applied Natural Language Processing, ANLP'97*. Washington DC, DC, EUA. 31 de Março a 3 de Abril de 1997. p. 194–201.
- (Bikel et al., 1999) Daniel M. Bikel, Richard Schwartz e Ralph Weischedel. An algorithm that learns what's in a name. *Machine Learning*. 34(1-3):211–231. Fevereiro de 1999.
- (Black et al., 1998) William J. Black, Fabio Rinaldi e David Mowatt. FACILE: Description of the NE system used for MUC-7. Em *Proceedings of the 7th Message Understanding Conference, MUC-7*. Fairfax, VI, EUA. 29 de Abril a 1 de Maio de 1998.
- (Blume, 2005) Matthias Blume. Automatic Entity Disambiguation: Benefits to NER, Relation Extraction, Link Analysis, and Inference. Em *Proceedings of 2005 International Conference on Intelligence Analysis*. McLean, VA, EUA. 2-4 de Maio de 2005.
- (Bontcheva et al., 2002) Kalina Bontcheva, Hamish Cunningham, Valentin Tablan, Diana Maynard e Oana Hamza. Using GATE as an Environment for Teaching NLP. Em *Proceedings of the ACL'02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*. Filadélfia, PA, EUA. Julho de 2002.
- (Borthwick, 1999) Andrew Borthwick. *A Maximum Entropy Approach to Named Entity Recognition*. Tese de doutoramento. Universidade de Nova Iorque, EUA. Setembro de 1999.
- (Borthwick et al., 1998) Andrew Borthwick, John Sterling, Eugene Agichtein e Ralph Grishman. Exploiting Diverse Knowledge Sources via Maximum Entropy in Named Entity Recognition. Em *Proceedings of the 6th Workshop on Very Large Corpora, WVLC-98*. Montreal, Quebec, Canadá. 15-16 de Agosto de 1998.
- (Brin, 1998) Sergey Brin. Extracting Patterns and Relations from the World Wide Web. Em *Workshop on the Web and Database, WebDB'98*. Valência, Espanha. 27-28 de Março de 1998. p. 172–183.

- (Buckley e Voorhees, 2000) Chris Buckley e Ellen M. Voorhees. Evaluating evaluation measure stability. Em Nicholas J. Belkin, Peter Ingwersen e Mun-Kew Leong, editores, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2000*. Atenas, Grécia. 24-28 de Julho de 2000. p. 33-40.
- (Buitelaar, 1998) Paul Buitelaar. CoreLex: An Ontology of Systematic Polysemous Classes. Em *Proceedings of International Conference on Formal Ontology in Information Systems, FOIS'98*. Trento, Itália. 6-8 de Junho de 1998.
- (Burns, 1991) Linda Claire Burns. *Vagueness: An Investigation into Natural Languages and the Sorites Paradox*. Kluwer Academic Publishers. Dordrecht. 1991.
- (Cardoso, 2006a) Nuno Cardoso. Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas. Tese de mestrado. Faculdade de Engenharia da Universidade do Porto. Outubro de 2006. Republicado como DI/FCUL TR-06-26, Departamento de Informática, Universidade de Lisboa, Novembro 2006.
- (Cardoso, 2006b) Nuno Cardoso. HAREM e MiniHAREM: Uma análise comparativa. Encontro do HAREM. Porto, Portugal. Apresentação. 15 de Julho de 2006. [http://www.linguateca.pt/documentos/encontroHAREM\\_cardoso.pdf](http://www.linguateca.pt/documentos/encontroHAREM_cardoso.pdf).
- (Carreras e Padró, 2002) Xavier Carreras e Lluís Padró. A flexible distributed architecture for natural language analyzers. Em Manuel González Rodrigues e Carmen Paz Suarez Araujo, editores, *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation*. Las Palmas de Gran Canaria, Espanha. 29-31 de Maio de 2002. p. 1813-1817.
- (Carreras et al., 2002) Xavier Carreras, Lluís Màrques e Lluís Padró. Named Entity Extraction using AdaBoost. Em Dan Roth e Antal van den Bosch, editores, *Proceedings of CoNLL-2002, the 6th Conference on Natural Language Learning*. Taipé, Formosa. 31 de Agosto a 1 de Setembro de 2002. p. 167-170.
- (Carreras et al., 2003a) Xavier Carreras, Lluís Màrquez e Lluís Padró. Named entity recognition for Catalan using only Spanish resources and unlabelled data. Em *10th Conference of the European Chapter of the Association for Computational Linguistics, EACL'03*. Budapeste, Hungria. Abril de 2003. p. 43-50.
- (Carreras et al., 2003b) Xavier Carreras, Lluís Màrquez e Lluís Padró. A Simple Named Entity Extractor using AdaBoost. Em Walter Daelemans, Miles Osborne, Walter Daelemans e Miles Osborne, editores, *Proceedings of the Conference on Computational Natural Language Learning, CoNLL-2003*. Edmonton, Canadá. 31 de Maio a 1 de Junho de 2003. p. 152-155.

- (Chaves et al., 2005) Marcirio Silveira Chaves, Mário J. Silva e Bruno Martins. A Geographic Knowledge Base for Semantic Web Applications. Em C. A. Heuser, editor, *Proceedings do 20º Simpósio Brasileiro de Banco de Dados, SBBD*. Uberlândia, MG, Brasil. 3-7 de Outubro de 2005. p. 40–54.
- (Chinchor, 1992) Nancy Chinchor. The Statistical Significance of MUC-4 Results. Em *Proceedings of the 4th Conference on Message Understanding, MUC-4*. McLean, VI, EUA. 16-18 de Junho de 1992. p. 30–50.
- (Chinchor, 1995) Nancy Chinchor. Statistical significance of MUC-6 results. Em *Proceedings of the 6th Message Understanding Conference, MUC-6*. Columbia, MD, EUA. 6-8 de Novembro de 1995. p. 39–43.
- (Chinchor, 1998a) Nancy Chinchor. Statistical Significance of MUC-7 Results. Em *Proceedings of the 7th Message Understanding Conference, MUC-7*. Fairfax, VI, EUA. 29 de Abril a 1 de Maio de 1998.
- (Chinchor e Marsh, 1998) Nancy Chinchor e Elaine Marsh. MUC-7 Named Entity Task Definition (version 3.5). Em *Proceedings of the 7th Message Understanding Conference, MUC-7*. Fairfax, VI, EUA. 29 de Abril a 1 de Maio de 1998.
- (Chinchor, 1998b) Nancy A. Chinchor. Overview of MUC-7/MET-2. Em *Proceedings of the 7th Message Understanding Conference, MUC-7*. Fairfax, VI, EUA. 29 de Abril a 1 de Maio de 1998.
- (Christ et al., 1999) Oliver Christ, Bruno M. Schulze, Anja Hofmann e Esther Koenig. The IMS Corpus Workbench: Corpus Query Processor (CQP): User's Manual (CQP V2.2). Relatório técnico. Universidade de Estugarda. 16 de Agosto de 1999. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPUserManual/HTML/>.
- (Cohen e Sarawagi, 2004) William W. Cohen e Sunita Sarawagi. Exploiting dictionaries in named entity extraction: combining semi-Markov extraction processes and data integration methods. Em *Proceedings of KDD-04, the 10th International Conference on Knowledge Discovery and Data Mining*. Seattle, WA, EUA. 22-25 de Agosto de 2004.
- (Costa et al., 2007) Luís Costa, Paulo Rocha e Diana Santos. Organização e resultados morfolímpicos. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. Lisboa, Portugal. 20 de Março de 2007. p. 15–33.
- (Cruse, 2004) Alan Cruse. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press. Oxford. 2004.

- (Cunningham, 2005) Hamish Cunningham. Information Extraction, Automatic. Em *Encyclopedia of Language and Linguistics*. Elsevier. 2ª edição. 2005. p. 665–677.
- (Cunningham et al., 2002) Hamish Cunningham, Diana Maynard, Kalina Bontcheva e Valentin Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. Em *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, ACL'02*. Filadélfia, PA, EUA. Julho de 2002.
- (Curran e Clark, 2003) James R. Curran e Stephen Clark. Language independent NER using a maximum entropy tagger. Em Walter Daelemans e Miles Osborne, editores, *Proceedings of the Conference on Computational Natural Language Learning, CoNLL-2003*. Edmonton, Canadá. 31 de Maio a 1 de Junho de 2003. p. 164–167.
- (Daelemans et al., 2003) Walter Daelemans, Jakub Zavrel, Ko van der Sloot e Antal van den Bosch. TiMBL: Tilburg Memory-Based Learner. Relatório Técnico ILK 03-10. Universidade de Tilburg. 2003.
- (Day et al., 1997) David Day, John Aberdeen, Lynette Hirschman, Robyn Kozierek e Patricia Robinson. Mixed-Initiative Development of Language Processing Systems. Em *Proceedings of the Fifth Conference on Applied Natural Language Processing, ANLP'97*. Washington DC, DC, EUA. 31 de Março a 3 de Abril de 1997. p. 88–95.
- (Delboni, 2005) Tiago M. Delboni. *Expressões de posicionamento como fonte de contexto geográfico na Web*. Tese de doutoramento. Universidade Federal de Minas Gerais. 2005.
- (Densham e Reid, 2003) Ian Densham e James Reid. A geo-coding service encompassing a geo-parsing tool and integrated digital gazetteer service. Em *Proceedings of the HLT-NAACL 2003 Workshop on the Analysis of Geographic References*. Edmonton, Canadá. 27 de Maio a 1 de Junho de 2003.
- (Dietterich, 2000) Thomas G. Dietterich. Ensemble methods in machine learning. Em J. Kittler e F. Roli, editores, *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21-23, 2000 Proceedings*. Springer. Nova Iorque, NY, EUA. 2000. p. 1–15.
- (Doddington et al., 2004) George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel e Ralph Weischedel. The Automatic Content Extraction (ACE) Program. Tasks, Data and Evaluation. Em Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa e Raquel Silva, editores, *Proceedings of LREC2004, the 4th International Conference on Language Resources and Evaluation*. Lisboa, Portugal. 26-28 de Maio de 2004. p. 837–840.

- (Douthat, 1998) Aaron Douthat. The Message Understanding Conference Scoring Software User's Manual. Em *Proceedings of the 7th Message Understanding Conference, MUC-7*. Fairfax, VI, EUA. 29 de Abril a 1 de Maio de 1998.
- (Efron, 1981) Bradley Efron. Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods. *Biometrika*. 81(3):589–599. Dezembro de 1981.
- (Ellis, 1993) John M. Ellis. *Language, Thought and Logic*. Northwestern University Press. Evanston, IL, EUA. 1993.
- (Evert, 2005) Stefan Evert. The CQP Query Language Tutorial (CWB version 2.2.b90). Relatório técnico. Universidade de Estugarda. 10 de Julho de 2005.
- (Fairon, 1999) Cédric Fairon. Parsing a Web site as a corpus. Em Cédric Fairon, editor, *Analyse lexicale et syntaxique: Le système INTEX*. John Benjamins Publishing. Amsterdão, Países Baixos. 1999. p. 327–340.
- (Ferrández et al., 2005) Óscar Ferrández, Zornitsa Kozareva, Andrés Montoyo e Rafael Muñoz. NERUA: sistema de detección y clasificación de entidades utilizando aprendizaje automático. *Sociedad Española para el Procesamiento del Lenguaje Natural*. 35:37–44. 2005.
- (Ferrández et al., 2006) Óscar Ferrández, Antonio Toral e Rafael Muñoz. Fine tuning features and post-processing rules to improve Named Entity Recognition. Em Christian Kop, Günther Fliedl, Heinrich C. Mayr e Elisabeth Métais, editores, *Processing and Information Systems, 11th International Conference on Applications of Natural Language to Information Systems, NLDB 2006, Klagenfurt, Austria, May 31 - June 2, 2006, Proceeding*. Springer. Berlin/Heidelberg. 2006. p. 176–185.
- (Fillmore, 1968) Charles J. Fillmore. The case for case. Em Emmon Bach e Robert T. Harms, editores, *Universals in Linguistic Theory*. Holt, Rinehart and Winston. Londres. 1968. p. 1–88.
- (Florian et al., 2003) Radu Florian, Abe Ittycheriah, Hongyan Jing e Tong Zhang. Named Entity Recognition through Classifier Combination. Em Walter Daelemans e Miles Osborne, editores, *Proceedings of the Conference on Computational Natural Language Learning, CoNLL-2003*. Edmonton, Canadá. 31 de Maio a 1 de Junho de 2003. p. 168–171.
- (Frankenberg-Garcia e Santos, 2002) Ana Frankenberg-Garcia e Diana Santos. COMPARA, um corpus paralelo de português e de inglês na Web. *Cadernos de Tradução*. IX(1):61–79. 2002.
- (Friburger, 2002) Nathalie Friburger. *Reconnaissance automatique des noms propres. Application à la classification automatique de textes journalistiques*. Tese de doutoramento. Universidade François Rabelais, Tours, França. 2 de Dezembro de 2002.

- (Gale et al., 1992) William A. Gale, Kenneth W. Church e David Yarowsky. One Sense Per Discourse. Em *Proceedings of the 4th DARPA Speech and Natural Language Workshop*. Harriman, NY, EUA. 23-26 de Fevereiro de 1992. p. 233–237.
- (Gey et al., 2006) Frederic Gey, Ray Larson, Mark Sanderson, Hideo Joho e Paul Clough. GeoCLEF: the CLEF 2005 Cross-Language Geographic Information Retrieval Track. Em Carol Peters, Frederic Gey, Julio Gonzalo, Henning Müeller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini e Maarten de Rijke, editores, *Accessing Multilingual information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005. Vienna, Austria, September 2005. Revised Selected papers*. Springer. Berlin/Heidelberg, 2006. p. 908–919.
- (Gey et al., 2007) Fredric Gey, Ray Larson, Mark Sanderson, Kerstin Bischoff, Thomas Mandl, Christa Womser-Hacker, Diana Santos, Paulo Rocha, Giorgio M. Di Nunzio e Nicola Ferro. GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. Em Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de Rijke e Maximilian Stempfhuber, editores, *Evaluation of Multilingual and Multi-modal Information Retrieval - 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September, 2006. Revised Selected papers*. Springer. Berlin / Heidelberg, 2007. p. 852–876.
- (Ginsberg, 1987) Matthew Ginsberg. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann. Los Altos, CA, EUA. 1987.
- (Gomes e Silva, 2006) Daniel Gomes e Mário J. Silva. Modelling Information Persistence on the Web. Em *Proceedings of the 6th International Conference on Web Engineering, ICWE 2006*. Palo Alto, CA, EUA. 11-14 de Julho de 2006. p. 193–200.
- (Good, 2000) Philip Good. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer. Nova Iorque, NY, EUA. 2ª edição. 2000.
- (Grishman e Sundheim, 1995) Ralph Grishman e Beth Sundheim. Design of the MUC-6 Evaluation. Em *Proceedings of the 6th Message Understanding Conference, MUC-6*. Columbia, MD, EUA. 6-8 de Novembro de 1995. p. 413–422.
- (Grishman e Sundheim, 1996) Ralph Grishman e Beth Sundheim. Message Understanding Conference 6: A Brief History. Em *Proceedings of the 16th International Conference on Computational Linguistics, COLING 96*. Copenhaga, Dinamarca. 5-9 de Agosto de 1996. p. 466–471.
- (Gross, 1975) Maurice Gross. *Méthodes en Syntaxe - Régime des constructions complétives*. Hermann. Paris, França. 1975.

- (Guthrie et al., 2004) Louise Guthrie, Roberto Basili, Eva Hajicova e Frederick Jelinek, editores. *Workshop proceedings of LREC 2004: Beyond Entity Recognition - Semantic Labelling for NLP Tasks*. ELRA. Lisboa, Portugal. 25 de Maio de 2004.
- (Harman, 1993) Donna Harman. Overview of the First TREC Conference. Em Robert Korfhage, Edie Rasmussen e Peter Willett, editores, *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 93*. Pittsburgh, PA, EUA. 27 de Junho a 1 de Julho de 1993. p. 36–47.
- (Harpring, 1997) Patricia Harpring. Proper words in proper places: The thesaurus of geographic names. *MDA Information*. 2(3):5–12. 1997.
- (Hill, 2000) Linda L. Hill. Core elements of digital gazetteers: Placenames, categories, and footprints. Em José Luis Borbinha e Thomas Bake, editores, *Research and Advanced Technology for Digital Libraries, 4th European Conference, ECDL 2000, Lisbon, Portugal, September 18-20, 2000, Proceedings*, ECDL-00. Springer. Berlin/Heidelberg. 2000. p. 280–291.
- (Hill et al., 1999) Linda L. Hill, James Frew e Qi Zheng. Geographic names: the implementation of a gazetteer in a georeferenced digital library. *D-Lib Magazine*. 5(1). Janeiro de 1999.
- (Hirschman, 1998) Lynette Hirschman. The evolution of evaluation: Lessons from the Message Understanding Conferences. *Computer Speech and Language*. 12(4):281–305. 1998.
- (Holte et al., 1989) Robert C. Holte, Liane Acker e Bruce W. Porter. Concept Learning and the Problem of Small Disjuncts. Em N. S. Sridharan, editor, *Proceedings of the Eleventh Joint International Conference on Artificial Intelligence, IJCAI*. Detroit, MI, EUA. Agosto de 1989. p. 813–818.
- (Hughes e Cresswell, 1968) George E. Hughes e Maxwell J. Cresswell. *An Introduction to Modal Logic*. Methuen & Co., Ltd. Londres. 1968.
- (Inácio e Santos, 2006) Susana Inácio e Diana Santos. Syntactical Annotation of COMPARA: Workflow and First Results. Em Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Computational Processing of the Portuguese Language: 7th International Workshop, PROPOR 2006, Itatiaia, Brazil, May 2006*. Springer. Berlin/Heidelberg. 2006. p. 256–259.
- (Japkowicz, 2003) Nathalie Japkowicz. Class Imbalances: Are we Focusing on the Right Issue? Em Nitesh Chawla, Nathalie Japkowicz e Aleksander Kolcz, editores, *Proceedings of The Twentieth International Conference on Machine Learning (ICML-2003), Workshop on Learning from Imbalanced Data Sets II*. Washington DC, DC, EUA. Agosto de 2003.

- (Joachims, 1999) Thorsten Joachims. Transductive inference for text classification using support vector machines. Em Saso Dzeroski e Ivan Bratko, editores, *Proceedings of the Sixteenth International Conference on Machine Learning, ICML 1999*. Bled, Eslovénia. Junho de 1999. p. 200–209.
- (Joachims, 2002) Thorsten Joachims. *Learning to Classify Text using Support Vector Machines: Methods Theory and Algorithms*. Kluwer Academic Publishers. Norwell, MA, EUA. Maio de 2002.
- (Johannessen et al., 2005) Janne Bondi Johannessen, Kristin Hagen, Åsne Haaland, Andra Björk Jónsdóttir, Anders Nøklestad, Dimitris Kokkinakis, Paul Meurer, Eckhard Bick e Dorte Haltrup. Named Entity Recognition for the Mainland Scandinavian Languages. *Literary and Linguistic Computing*. 20(1):91–102. 2005.
- (Jones et al., 2004) Cristopher B. Jones, Alia I. Abdelmoty, David Finch, Gaihua Fu e Subodh Vaid. The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. Em Max J. Egenhofer, Christian Freska e Harvey Miller, editores, *Geographic Information Science, Third International Conference, GIScience 2004, Adelphi, MD, USA, October 20-23, 2004, Proceedings*. Springer. Berlin/Heidelberg. 2004. p. 125–139.
- (Jones e Bates, 1977) Karen Sparck Jones e R. G. Bates. Research on Automatic Indexing 1974-1976. Relatório técnico. Computer Laboratory, University of Cambridge. 1977.
- (Kamp e Reyle, 1993) Hans Kamp e Uwe Reyle. *From Discourse to Logic: an Introduction to Model Theoretic Semantics of Natural language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Press. Dordrecht. 1993.
- (Koehn, 2004) Philip Koehn. Statistical significance tests for machine translation evaluation. Em Dekang Lin e Dekai Wu, editores, *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004*. Barcelona, Espanha. Julho de 2004. p. 388–395.
- (Kohler, 2003) Janet Kohler. *Analysing search engine queries for the use of geographic terms*. Tese de doutoramento. Universidade de Sheffield. 2003.
- (Kornai e Sundheim, 2003) Andras Kornai e Beth Sundheim, editores. *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*. Association for Computational Linguistics. Morristown, NJ, EUA. 27 de Maio a 1 de Junho de 2003.
- (Kozareva et al., 2007) Zornitsa Kozareva, Óscar Ferrández, Andrés Montoyo, Rafael Muñoz, Armando Suárez e Jaime Gómez. Combining data-driven systems for improving Named Entity Recognition. *Data & Knowledge Engineering*. 61(3):449–466. 2007.

- (Krupka e Hausman, 1998) George R. Krupka e Kevin Hausman. IsoQuest Inc.: Description of the NetOwITM extractor system as used for MUC-7. Em *Proceedings of the 7th Message Understanding Conference, MUC-7*. Fairfax, VI, EUA. 29 de Abril a 1 de Maio de 1998.
- (Lakoff, 1987) George Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press. Chicago & Londres. 1987.
- (Lakoff e Johnson, 1980) George Lakoff e Mark Johnson. *Metaphors We Live By*. University of Chicago Press. Chicago & Londres. 1980.
- (Lansing, 2001) Jeff Lansing. Geoparser Service Specification 0.71. Relatório Técnico OGC-01-035. Open Geospatial Consortium. Março de 2001.
- (Leidner, 2004) Jochen L. Leidner. Towards a Reference Corpus for Automatic Toponym Resolution Evaluation. Em *Proceedings of the Workshop on Geographic Information Retrieval held at the 27th Annual International ACM SIGIR Conference, GIR 2004*. Sheffield, Reino Unido. 25-29 de Julho de 2004.
- (Leidner et al., 2003) Jochen L. Leidner, Gail Sinclair e Bonnie Webber. Grounding Spatial Named Entities for Information Extraction and Question Answering. Em *Proceedings of the HLT-NAACL 2003 Workshop on the Analysis of Geographic References*. Edmonton, Canadá. 27 de Maio a 1 de Junho de 2003. p. 31-38.
- (Leveling e Hartrumpf, 2006) Johannes Leveling e Sven Hartrumpf. On metonymy recognition for Geographic IR. Em *Proceedings of the Workshop on Geographic Information Retrieval held at the 29th Annual International ACM SIGIR Conference, GIR 2006*. Seattle, WA, EUA. Agosto de 2006.
- (Leveling e Veiel, 2006) Johannes Leveling e Dirk Veiel. University of Hagen at GeoCLEF 2006: Experiments with metonymy recognition in documents. Em Alessandro Nardi, Carol Peters e José Luís Vicedo, editores, *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop*. 2006. s/pp.
- (Li et al., 2002) Huifeng Li, Rohini Srihari, Cheng Niu e Wei Li. Location normalization for information extraction. Em *Proceedings of the 19th International Conference on Computational Linguistics, COLING 2002*. Taipé, Formosa. 24 de Agosto a 1 de Setembro de 2002. p. 1-7.
- (Li, 1992) Wentian Li. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*. 38(6):1842-1845. 1992.
- (Lin e Hauptmann, 2005) Wei-Hao Lin e Alexander Hauptmann. Revisiting the Effect of Topic Set Size on Retrieval Experiment Error. Em Ricardo A. Baeza-Yates, Nivio Ziviani,

- Gary Marchionini, Alistair Moffat e John Tait, editores, *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2005*. Salvador, Brasil. 15-19 de Agosto de 2005. p. 637–638.
- (Ling e Li, 1998) Charles X. Ling e Chenghui Li. Data mining for direct marketing: Problems and solutions. Em Rakesh Agrawal, Paul E. Stolorz e Gregory Piatetsky-Shapiro, editores, *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, KDD-98*. Nova Iorque, NY, EUA. 31 de Agosto de 1998. p. 73–79.
- (Madrigal et al., 2003) Víctor J. Díaz Madrigal, José Troyano e Fernando Enríquez. Aplicación de Modelos de Markov y Maquinas SVM al Reconocimiento de Entidades. Em *Actas de las X Jornadas de la CAEPIA y V de la TTIA, CAEPIA'2003*. San Sebastián, Espanha. 11-14 de Novembro de 2003. p. 55–58.
- (Malouf, 2002) Robert Malouf. Markov models for language-independent named entity recognition. Em Dan Roth e Antal van den Bosch, editores, *Proceedings of CoNLL-2002, the 6th Conference on Natural Language Learning*. Taipé, Formosa. 31 de Agosto a 1 de Setembro de 2002. p. 187–190.
- (Mamede et al., 2003) Nuno J. Mamede, Jorge Baptista, Isabel Trancoso e Maria das Graças Volpe Nunes, editores. *Computational Processing of the Portuguese Language, 6th International Workshop, PROPOR 2003*. Springer. Berlin/Heidelberg. 2003.
- (Mandl et al., 2007) Thomas Mandl, Fredric Gey, Giorgio Di Nunzio, Nicola Ferro, Ray Larson, Mark Sanderson, Diana Santos, Christa Womser-Hacker e Xing Xie. GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview. Em Alessandro Nardi e Carol Peters, editores, *Working Notes for the CLEF 2007 Workshop*. 2007. s/pp.
- (Manov et al., 2003) Dimitar Manov, Atanas Kiryakov, Borislav Popov, Kalina Bontcheva, Diana Maynard e Hamish Cunningham. Experiments with geographic knowledge for information extraction. Em *Proceedings of the HLT-NAACL 2003 Workshop on the Analysis of Geographic References*. Edmonton, Canadá. 27 de Maio a 1 de Junho de 2003.
- (Marcelino, 2005) Isabel Marcelino. Documentação do ELLE. 29 de Setembro de 2005. [http://www.linguateca.pt/Equipa/isabel/Documentacao\\_ELLE.pdf](http://www.linguateca.pt/Equipa/isabel/Documentacao_ELLE.pdf).
- (Markert e Nissim, 2002) Katja Markert e Malvina Nissim. Towards a corpus annotated for metonymies: the case of location names. Em Manuel González Rodríguez, Carmen Paz Suárez Araujo, Manuel González Rodrigues e Carmen Paz Suarez Araujo, editores, *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation*. Las Palmas de Gran Canaria, Espanha. 29-31 de Maio de 2002. p. 1385–1392.

- (Martins et al., 2005) Bruno Martins, Mário J. Silva e Marcirio Silveira Chaves. Challenges and Resources for Evaluating Geographical IR. Em *Proceedings of the 2005 Workshop On Geographic Information Retrieval, GIR 2005, Bremen, Germany, November 4, 2005*. Bremen, Alemanha. Outubro de 2005. p. 31–34.
- (Martins et al., 2006) Bruno Martins, Marcirio Chaves e Mario J. Silva. O sistema CaGE para Reconhecimento de referências geográficas em textos na língua portuguesa. Encontro do HAREM. Porto, Portugal. Apresentação. 15 de Julho de 2006. <http://www.linguateca.pt/documentos/CaGEHAREM.ppt>.
- (Martins et al., 2007) Bruno Martins, Nuno Cardoso, Marcirio Silveira Chaves, Leonardo Andrade e Mário J. Silva. The University of Lisbon at GeoCLEF 2006. Em Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de Rijke e Maximilian Stempfhuber, editores, *Evaluation of Multilingual and Multi-modal Information Retrieval - 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September, 2006. Revised Selected papers*. Springer. Berlin / Heidelberg. 2007. p. 986–994.
- (Maynard et al., 2002) Diana Maynard, Hamish Cunningham, Kalina Bontcheva e Marin Dimitrov. Adapting a Robust Multi-genre NE System for Automatic Content Extraction. Em D. R. Scott, editor, *Artificial Intelligence: Methodology, Systems, and Applications, 10th International Conference, AIMSA 2002, Varna, Bulgaria, September 4-6, 2002, Proceedings*. Springer. Berlin/Heidelberg. 2002. p. 264–273.
- (Maynard et al., 2003a) Diana Maynard, Kalina Bontcheva e Hamish Cunningham. Towards a semantic extraction of named entities. Em *Recent Advances in Natural Language Processing, RANLP*. Borovets, Bulgária. 10-12 de Setembro de 2003.
- (Maynard et al., 2003b) Diana Maynard, Valentin Tablan, Kalina Bontcheva, Hamish Cunningham e Yorick Wilks. MUSE: a MUlti-Source Entity recognition system. Em apreciação pela revista *Computers and the Humanities*. 2003. <http://gate.ac.uk/sale/muse/muse.pdf>.
- (McDonald, 1996) David D. McDonald. Internal and External Evidence in the Identification and Semantic Categorization of Proper Names. Em Branimir Boguraev e James Pustejovsky, editores, *Corpus Processing for Lexical Acquisition*. MIT Press. Cambridge, MA & Londres. 1996. p. 21–39.
- (Merchant et al., 1996) Roberta Merchant, Mary Ellen Okurowski e Nancy Chinchor. The Multilingual Entity Task (MET) Overview. Em *Proceedings of TIPSTER Text Program (Phase II)*. Vienna, VI, EUA. Maio de 1996. p. 449–451.

- (Mikheev et al., 1999) Andrei Mikheev, Marc Moens e Claire Grover. Named Entity Recognition without Gazetteers. Em *Proceedings of EACL'99: Ninth Conference of the European Chapter of the Association for Computational Linguistics*. Bergen, Noruega. 8-12 de Junho de 1999. p. 1-8.
- (Mitsumori et al., 2004) Tomohiro Mitsumori, Sevrani Fation, Masaki Murata, Kouichi Doi e Hirohumi Doi. Boundary correction of protein names adapting heuristic rules. Em Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Fifth International Conference, CICLing 2004*. Springer. Berlin/Heidelberg. 2004. p. 172-175.
- (Moore et al., 2002) David S. Moore, George P. McCabe, William M. Duckworth e Stanley L. Sclove. *The Practice of Business Statistics: Using Data for Decisions*. W. H. Freeman. Novembro de 2002.
- (Morgan, 2006) William Morgan. Statistical Hypothesis Tests for NLP. 16 de Fevereiro de 2006. <http://nlp.stanford.edu/local/talks/sigtest.pdf>.
- (Mota e Moura, 2003) Cristina Mota e Pedro Moura. ANELL: A Web System for Portuguese Corpora Annotation. Em Nuno J. Mamede, Jorge Baptista, Isabel Trancoso e Maria das Graças Volpe Nunes, editores, *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003. Faro, Portugal, June 2003*. Springer. Berlin/Heidelberg. 2003. p. 184-188.
- (Mota et al., 2007) Cristina Mota, Diana Santos e Elisabete Ranchhod. Avaliação de reconhecimento de entidades mencionadas: princípio de AREM. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. Lisboa, Portugal. 20 de Março de 2007. p. 161-176.
- (Nissim et al., 2004) Malvina Nissim, Colin Matheson e James Reid. Recognising Geographical Entities in Scottish Historical Documents. Em *Proceedings of the Workshop on Geographic Information Retrieval held at the 27th Annual International ACM SIGIR Conference, GIR 2004*. Sheffield, Reino Unido. 25-29 de Julho de 2004.
- (Noreen, 1989) Eric W. Noreen. *Computer Intensive Methods for Testing Hypotheses: An Introduction*. John Wiley & Sons. Nova Iorque, NY, EUA. 1989.
- (Oliveira e Ribeiro, 2003) José N. Oliveira e Óscar Ribeiro. Knowledge renovator - Requirements Specification. Relatório técnico. Universidade do Minho, Departamento de Informática. 2003. IKF-P partner - IKF (E!2235).
- (Olligschlaeger e Hauptmann, 1999) Andreas M. Olligschlaeger e Alexander G. Hauptmann. Multimodal Information Systems and GIS: The Informedia Digital Video Library. Em *Proceedings of the 1999 ESRI User Conference*. San Diego, CA, EUA. 26-30 de Julho de 1999.

- (Palmer e Day, 1997) David D. Palmer e David S. Day. A Statistical Profile of the Named Entity Task. Em *Proceedings of the Fifth ACL Conference for Applied Natural Language Processing*, ANLP'97. Washington DC, DC, EUA. Abril de 1997. p. 190–193.
- (Papineni et al., 2001) Kishore Papineni, Salim Roukos, Todd Ward e Wei-Jing Zhuw. BLEU: a Method for Automatic Evaluation of Machine Translation. Relatório Técnico RC22176 (W0109-022). Computer Science IBM Research Division, T.J.Watson Research Center. 17 de Setembro de 2001. <http://domino.watson.ibm.com/library/CyberDig.nsf/Home>. Republicado em ACL'02.
- (Pasca, 2004) Marius Pasca. Acquisition of categorized named entities for web search. Em *Proceedings of the 2004 ACM CIKM International Conference on Information and Knowledge Management*. Washington DC, DC, EUA. 8-13 de Novembro de 2004. p. 137–145.
- (Paumier, 2002) Sébastien Paumier. Manuel d'utilisation du logiciel Unitex. Relatório técnico. Universidade de Marne-la-Vall. Julho de 2002. <http://www-igm.univ-mlv.fr/~unitex/manuelunitex.pdf>.
- (Petasis et al., 2000) Georgios Petasis, Alessandro Cucchiarelli, Paola Velardi, Georgios Paliouras, Vangelis Karkaletsis e Constantine D. Spyropoulos. Automatic adaptation of proper noun dictionaries through cooperation of machine learning and probabilistic methods. Em Nicholas J. Belkin, Peter Ingwersen e Mun-Kew Leong, editores, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR 2000. Atenas, Grécia. 24-28 de Julho de 2000. p. 128–135.
- (Petasis et al., 2004) Georgios Petasis, Vangelis Karkaletsis, Claire Grover, Benjamin Hachey, Maria-Teresa Pazienza, Michele Vindigni e Jose Coch. Adaptive, Multilingual Named Entity Recognition in Web Pages. Em *Proceedings of the 16th European Conference on Artificial Intelligence*, ECAI 2004. Valência, Espanha. 22-27 de Agosto de 2004. p. 1073–1074.
- (Platt, 1999) John C. Platt. Fast training of support vector machines using sequential minimal optimization. Em B. Schölkopf, C.J.C. Burges e A.J. Smola, editores, *Advances in Kernel Methods - Support Vector Learning*. MIT Press. Cambridge, MA, EUA. 1999. p. 185–208.
- (Purves e Jones, 2004) Ross Purves e Christopher B. Jones. Workshop on Geographic Information Retrieval. *SIGIR Forum*. 38(1). 2004.
- (Pustejovsky, 1994) James Pustejovsky. Semantic Typing and Degrees of Polymorphism. Em Carlos Martin-Vide, editor, *Current Issues in Mathematical Linguistics*. Elsevier. Amsterdão, Holanda. 1994. p. 221–238.

- (Pustejovsky, 1995) James Pustejovsky. *The Generative Lexicon*. MIT Press. Cambridge, MA, EUA. 1995.
- (Rauch et al., 2003) Erik Rauch, Michael Bukatin e Kenneth Baker. A confidence-based framework for disambiguating geographic terms. Em *Proceedings of the HLT-NAACL 2003 Workshop on the Analysis of Geographic References*. Edmonton, Canadá. 27 de Maio a 1 de Junho de 2003.
- (Riezler e Maxwell III, 2005) Stefan Riezler e John T. Maxwell III. On Some Pitfalls in Automatic Evaluation and Significance Testing for MT. Em *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and Summarization, MTSE 2005*. Ann Arbor, MI, EUA. Junho de 2005. p. 57–64.
- (Riloff, 1996) Ellen Riloff. Automatically Generating Extraction Patterns from Untagged Text. Em *Proceedings of the Thirteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, AAAI 96, IAAI 96*. Portland, OR, EUA. 4-8 de Agosto de 1996. p. 1044–1049.
- (Rocha e Santos, 2007) Paulo Rocha e Diana Santos. CLEF: Abrindo a porta à participação internacional em avaliação de RI do português. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. Lisboa, Portugal. 20 de Março de 2007. p. 143–158.
- (Rocha e Santos, 2000) Paulo Alexandre Rocha e Diana Santos. CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa. Em Maria das Graças Volpe Nunes, editor, *V Encontro para o processamento computacional da língua portuguesa escrita e falada*, PROPOR 2000. Atibaia, SP, Brasil. 19-22 de Novembro de 2000. p. 131–140.
- (Sakai, 2006) Tetsuya Sakai. Bootstrap-Based Comparisons of IR Metrics for Finding One Relevant Document. Em Hwee Tou Ng, Mun-Kew Leong, Min-Yen Kan e Donghong Ji, editores, *Proceedings of Third Asia Information Retrieval Symposium, AIRS 2006*. Springer. Nova Iorque, NY, EUA. 2006. p. 374–389.
- (Sang, 2002) Erik F. Tjong Kim Sang. Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition. Em Dan Roth e Antal van den Bosch, editores, *Proceedings of CoNLL-2002, the 6th Conference on Natural Language Learning*. Taipé, Formosa. 31 de Agosto a 1 de Setembro de 2002. p. 155–158.
- (Sang e Meulder, 2003) Erik F. Tjong Kim Sang e Fien De Meulder. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. Em Walter Daelemans e Miles Osborne, editores, *Proceedings of the Conference on Computational Natural Language Learning, CoNLL-2003*. Edmonton, Canadá. 31 de Maio a 1 de Junho de 2003. p. 142–147.

- (Santos, 1940) Delfim Santos. *Conhecimento e realidade*. Tese de doutoramento. Universidade de Coimbra. Lisboa, Portugal. 1940.
- (Santos, 2007a) Diana Santos, editor. *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. Lisboa, Portugal. 20 de Março de 2007.
- (Santos, 1997) Diana Santos. The importance of vagueness in translation: Examples from English to Portuguese. *Romansk Forum*. 5:43–69. Junho de 1997. Versão bilingue revista em TradTerm 5.1, Revista do centro interdepartamental de tradução e terminologia, FFLCH - Universidade de São Paulo, 1998, "A relevância da vagueza para a tradução, ilustrada com exemplos de inglês para português", pp.41-70 / "The relevance of vagueness for translation: Examples from English to Portuguese", pp. 71-78.
- (Santos, 1999) Diana Santos. Processamento computacional da língua portuguesa: Documento de trabalho. Versão base de 9 de Fevereiro de 1999; revista a 13 de Abril de 1999. 1999. <http://www.linguateca.pt/branco/index.html>.
- (Santos, 2000) Diana Santos. O projecto Processamento Computacional do Português: Balanço e perspectivas. Em Maria das Graças Volpe Nunes, editor, *V Encontro para o processamento computacional da língua portuguesa escrita e falada*, PROPOR 2000. Atibaia, SP, Brasil. 19-22 de Novembro de 2000. p. 105–113.
- (Santos, 2002) Diana Santos. Um centro de recursos para o processamento computacional do português. *DataGramZero - Revista de Ciência da Informação*. 3(1). Fevereiro de 2002. [http://www.dgz.org.br/fev02/Art\\_02.htm](http://www.dgz.org.br/fev02/Art_02.htm).
- (Santos, 2006a) Diana Santos. HAREM: the first evaluation contest for Named Entity Recognition in Portuguese. Palestra convidada no IST. Lisboa, Portugal. 24 de Fevereiro de 2006. <http://www.linguateca.pt/documentos/SantosISTFev2006.pdf>.
- (Santos, 2006b) Diana Santos. Reconhecimento de entidades mencionadas. Palestra convidada na PUC. Rio de Janeiro, Brasil. 18 de Maio de 2006. <http://www.linguateca.pt/Diana/download/SantosPalestraPUCRio2006.pdf>.
- (Santos, 2006c) Diana Santos. Resumo da actividade da LINGUATECA de 15 de Maio de 2003 a 15 de Dezembro de 2006. Relatório técnico. LINGUATECA. Dezembro de 2006. <http://www.linguateca.pt/documentos/RelatorioLinguateca2003-2006.pdf>. Com a colaboração (por ordem alfabética) de Alberto Simões, Ana Frankenberg-Garcia, Belinda Maia, Luís Costa, Luís Miguel Cabral, Luís Sarmiento, Marcirio Chaves, Mário J. Silva, Nuno Cardoso, Paulo Gomes e Rui Vilela.
- (Santos, 2006d) Diana Santos. What is natural language? Differences compared to artificial languages, and consequences for natural language processing. Palestra con-

- vidada no SBLP2006 e no PROPOR'2006 . Itatiaia, RJ, Brasil. 15 de Maio de 2006. <http://www.linguateca.pt/Diana/download/SantosPalestraSBLPPropor2006.pdf>.
- (Santos, 2007b) Diana Santos. Avaliação conjunta. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. Lisboa, Portugal. 20 de Março de 2007. p. 1–12.
- (Santos e Bick, 2000) Diana Santos e Eckhard Bick. Providing Internet access to Portuguese corpora: the AC/DC project. Em Maria Gavrilidou, George Carayannis, Stella Markantonatou, Stelios Piperidis e Gregory Stainhauer, editores, *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000*. Atenas, Grécia. 31 de Maio a 2 de Junho de 2000. p. 205–210.
- (Santos e Cardoso, 2006) Diana Santos e Nuno Cardoso. A Golden Resource for Named Entity Recognition in Portuguese. Em Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Computational Processing of the Portuguese Language: 7th International Workshop, PROPOR 2006*. Itatiaia, Brazil, May 2006. Springer. Berlin/Heidelberg. 2006. p. 69–79.
- (Santos e Costa, 2005) Diana Santos e Luís Costa. A Linguateca e o projecto 'Processamento Computacional do português'. *Terminómetro - Número especial nº 7 - A terminologia em Portugal e nos países de língua portuguesa em África*. p. 63–69. 2005.
- (Santos e Gasperin, 2002) Diana Santos e Caroline Gasperin. Evaluation of parsed corpora: experiments in user-transparent and user-visible evaluation. Em Manuel González Rodríguez e Carmen Paz Suarez Araujo, editores, *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation*. Las Palmas de Gran Canaria, Espanha. 29-31 de Maio de 2002. p. 597–604.
- (Santos e Inácio, 2006) Diana Santos e Susana Inácio. Annotating COMPARA, a grammar-aware parallel corpus. Em Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik e Daniel Tapias, editores, *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC'2006*. Génova, Itália. 22-28 de Maio de 2006. p. 1216–1221.
- (Santos e Ranchhod, 1999) Diana Santos e Elisabete Ranchhod. Ambientes de processamento de corpora em português: Comparação entre dois sistemas. Em Irene Rodrigues e Paulo Quaresma, editores, *Actas do IV Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada, PROPOR'99*. Évora, Portugal. 20-21 de Setembro de 1999. p. 257–268.
- (Santos e Sarmiento, 2003) Diana Santos e Luís Sarmiento. O projecto AC/DC: acesso a corpora/disponibilização de corpora. Em Amália Mendes e Tiago Freitas, editores, *Ac-*

- tas do XVIII Encontro Nacional da Associação Portuguesa de Linguística*, APL 2002. Porto, Portugal. 2-4 de Outubro de 2003. p. 705–717.
- (Santos et al., 2003) Diana Santos, Luís Costa e Paulo Rocha. Cooperatively evaluating Portuguese morphology. Em Nuno J. Mamede, Jorge Baptista, Isabel Trancoso e Maria das Graças Volpe Nunes, editores, *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003*. Faro, Portugal, June 2003. Springer. Berlin/Heidelberg. 2003. p. 259–266.
- (Santos et al., 2004) Diana Santos, Alberto Simões, Ana Frankenberg-Garcia, Ana Pinto, Anabela Barreiro, Belinda Maia, Cristina Mota, Débora Oliveira, Eckhard Bick, Elisabete Ranchhod, José João Dias de Almeida, Luís Cabral, Luís Costa, Luís Sarmiento, Marcirio Chaves, Nuno Cardoso, Paulo Rocha, Rachel Aires, Rosário Silva, Rui Vilela e Susana Afonso. Linguateca: um centro de recursos distribuído para o processamento computacional da língua portuguesa. Em Guillermo De Ita Luna, Olac Fuentes Chávez e Mauricio Osorio Galindo, editores, *Proceedings of the international workshop "Taller de Herramientas y Recursos Lingüísticos para el Español y el Portugués"*, IX Iberoamerican Conference on Artificial Intelligence, IBERAMIA 2004. Puebla, México. Novembro de 2004. p. 147–154.
- (Santos et al., 2006) Diana Santos, Nuno Seco, Nuno Cardoso e Rui Vilela. HAREM: An Advanced NER Evaluation Contest for Portuguese. Em Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik e Daniel Tapias, editores, *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC'2006*. Génova, Itália. 22-28 de Maio de 2006. p. 1986–1991.
- (Sarmiento, 2006a) Luís Sarmiento. BACO - A large database of text and co-occurrences. Em Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik e Daniel Tapias, editores, *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC'2006*. Génova, Itália. 22-28 de Maio de 2006. p. 1787–1790.
- (Sarmiento, 2006b) Luís Sarmiento. SIEMÊS - A Named Entity Recognizer for Portuguese Relying on Similarity Rules. Em Renata Vieira, Paulo Quaresma, Maria da Graça Volpe Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Computational Processing of the Portuguese Language: 7th International Workshop, PROPOR 2006*. Itatiaia, Brazil, May 2006. Springer. Berlin/Heidelberg. 2006. p. 90–99.
- (Sarmiento e Mota, 2006) Luís Sarmiento e Cristina Mota. HAREM 2.0 Proposta. Encontro do HAREM. Porto, Portugal. Apresentação. 15 de Julho de 2006. [http://www.linguateca.pt/documentos/harem\\_2.0.ppt](http://www.linguateca.pt/documentos/harem_2.0.ppt).

- (Sarmiento et al., 2004) Luís Sarmiento, Belinda Maia e Diana Santos. The Corpógrafo - a Web-based environment for corpora research. Em Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa e Raquel Silva, editores, *Proceedings of LREC2004, the 4th International Conference on Language Resources and Evaluation*. Lisboa, Portugal. 26-28 de Maio de 2004. p. 449–452.
- (Sarmiento et al., 2006) Luís Sarmiento, Ana Sofia Pinto e Luís Cabral. REPENTINO - A Wide-Scope Gazetteer for Entity Recognition in Portuguese. Em Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Computational Processing of the Portuguese Language: 7th International Workshop, PROPOR 2006. Itatiaia, Brazil, May 2006*. Springer. Berlin/Heidelberg. 2006. p. 31–40.
- (Savoy, 1997) Jacques Savoy. Statistical Inference in Retrieval Effectiveness Evaluation. *Information Processing and Management*. 33:495–512. 1997.
- (Schank e Rieger, 1974) Roger Schank e Charles Rieger. Inference and the computer understanding of natural languages. *Artificial Intelligence*. 5(4):373–412. 1974.
- (Schilder et al., 2004) Frank Schilder, Yannick Versley e Christopher Habel. Extracting spatial information: grounding, classifying and linking spatial expressions. Em *Proceedings of the Workshop on Geographic Information Retrieval held at the 27th Annual International ACM SIGIR Conference, GIR 2004*. Sheffield, Reino Unido. 25-29 de Julho de 2004.
- (Schölkopf e Smola, 2002) Bernhard Schölkopf e Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press. Cambridge, MA, EUA. 2002.
- (Schröer, 2002) Ingo Schröer. A Case Study in Part-of-Speech tagging Using the ICOPOST Toolkit. Relatório técnico. Departamento de Informática da Universidade de Hamburgo. 2002.
- (Seco et al., 2006) Nuno Seco, Diana Santos, Rui Vilela e Nuno Cardoso. A Complex Evaluation Architecture for HAREM. Em Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Computational Processing of the Portuguese Language: 7th International Workshop, PROPOR 2006. Itatiaia, Brazil, May 2006*. Springer. Berlin/Heidelberg. 2006. p. 260–263.
- (Sekine et al., 2002) Satoshi Sekine, Kiyoshi Sudo e Chikashi Nobata. Extended named entity hierarchy. Em Manuel González Rodrigues e Carmen Paz Suarez Araujo, editores, *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation*. Las Palmas de Gran Canaria, Espanha. 29-31 de Maio de 2002. p. 1818–1824.

- (Sheskin, 2000) David J. Sheskin. *Handbook of Parametric and Nonparametric Statistical Procedures*. Springer. Nova Iorque, NY, EUA. 2ª edição. 2000.
- (Silberztein, 1993) Max Silberztein. *Dictionnaires électroniques et analyse lexicale du français. Le système INTEX*. Masson. Paris, França. 1993.
- (Silberztein, 2004) Max Silberztein. NooJ: A Cooperative, Object-Oriented Architecture for NLP. Em *INTEX pour la Linguistique et le traitement automatique des langues*. Presses Universitaires de Franche-Comté. Besançon, França. 2004.
- (Silva, 2004) Cândida Gonçalves da Silva. Specification of the knowledge representation standard of IKF-P (E!2235). Relatório técnico. Universidade do Minho, Departamento de Informática. 2004.
- (Simões e Almeida, 2002) Alberto Manuel Simões e José João Almeida. Jspell.pm - um módulo de análise morfológica para uso em Processamento de Linguagem Natural. Em Anabela Gonçalves e Clara Nunes Correia, editores, *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística*, APL 2001. Lisboa, Portugal. 2-4 de Outubro de 2002. p. 485–495.
- (Smith e Crane, 2001) David A. Smith e Gregory Crane. Disambiguating Geographic Names in a Historical Digital Library. Em *Research and Advanced Technology for Digital Libraries, 5th European Conference, ECDL 2001, Darmstadt, Germany, September 4-9, 2001, Proceedings*. Springer. Berlin/Heidelberg. 2001. p. 127–136.
- (Smith e Mann, 2003) David A. Smith e Gideon S. Mann. Bootstrapping toponym classifiers. Em *Proceedings of the HLT-NAACL 2003 Workshop on the Analysis of Geographic References*. Edmonton, Canadá. 27 de Maio a 1 de Junho de 2003.
- (Solorio, 2005) Tamar Solorio. *Improvement of Named Entity Tagging by Machine Learning*. Tese de doutoramento. Instituto Nacional de astrofísica, óptica e electrónica, Puebla, México. 2005.
- (Stitson et al., 1996) Mark O. Stitson, Jason A. E. Weston, Alex Gammerman, Volodya Vovk e Vladimir Vapnik. Theory of support vector machines. Relatório Técnico CSD-TR-96-17. Universidade de Londres, Royal Holloway. Egham, Reino Unido. Dezembro de 1996.
- (Sundheim, 1995) Beth Sundheim. Overview of Results of the MUC-6 Evaluation. Em *Proceedings of the 6th Message Understanding Conference, MUC-6*. Columbia, MD, EUA. 6-8 de Novembro de 1995. p. 13–31.
- (Suárez e Palomar, 2002) Armando Suárez e Manuel Palomar. A Maximum Entropy-based Word Sense Disambiguation System. Em *Proceedings of the 19th International Conference on*

- Computational Linguistics, COLING 2002*. Taipé, Formosa. 24 de Agosto a 1 de Setembro de 2002. p. 960–966.
- (Tettamanzi, 2003) Andrea G. B. Tettamanzi. Approaches to knowledge extraction based on soft computing. Relatório Técnico IKF-I Attività R1, 15.1. Universidade de Milão. 2003.
- (Télliez et al., 2005) Alberto Télliez, Manuel Montes y Gómez e Luis Villaseñor. A machine learning approach to information extraction. Em Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Sixth International Conference, CICLing 2005*. Springer. Berlin/Heidelberg. 2005. p. 539–547.
- (Tong e Koller, 2001) Simon Tong e Daphne Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*. 2:45–66. 2001.
- (Toral, 2005) Antonio Toral. DRAMNERI: a free knowledge based tool to Named Entity Recognition. Em *Proceedings of the 1st Free Software Technologies Conference*. A Coruña, Espanha. 23-25 de Março de 2005. p. 27–32.
- (Toral e Muñoz, 2006) Antonio Toral e Rafael Muñoz. A proposal to automatically build and maintain gazetteers for Named Entity Recognition by using Wikipedia. Em *Proceedings of the workshop on New Text Wikis and blogs and other dynamic text sources, 11th Conference of the European Chapter of the Association for Computational Linguistics*. Trento, Itália. Abril de 2006.
- (Vapnik, 1995) Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer. Nova Iorque, NY, EUA. 1995.
- (Vatant, 2006) Bernard Vatant. The geonames ontology. 2006. <http://www.geonames.org/ontology/>.
- (Voorhees e Buckley, 2002) Ellen M. Voorhees e Chris Buckley. The effect of topic set size on retrieval experiment error. Em Kalervo Järvelin, Micheline Beaulieu, Ricardo Baeza-Yates e Sung Hyon Myaeng, editores, *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2002*. Tampere, Finlândia. 11-15 de Agosto de 2002. p. 316–323.
- (Wakao et al., 1996) Takahiro Wakao, Robert Gaizauskas e Yorick Wilks. Evaluation of an Algorithm for the Recognition and Classification of Proper Names. Em Bente Maegaard, editor, *Proceedings of the 16th International Conference of Computational Linguistics, COLING'96*. Copenhaga, Dinamarca. 5-9 de Agosto de 1996. p. 418–423.

- (Will, 1993) Craig A. Will. Comparing human and machine performance for natural language information extraction: results for English microelectronics from the MUC-5 evaluation. Em *Proceedings of the 5th Message Understanding Conference, MUC-5*. Baltimore, MD, EUA. 25-25 de Agosto de 1993. p. 53–67.
- (Witten e Frank, 1999) Ian H. Witten e Eibe Frank. *Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann. San Francisco, CA, EUA. 1999.
- (Zhang e Mani, 2003) Jianping Zhang e Inderjeet Mani. kNN Approach to Unbalanced Data Distributions: A Case Study involving Information Extraction. Em *Proceedings of The Twentieth International Conference on Machine Learning (ICML-2003), Workshop on Learning from Imbalanced Data Sets II*. Washington DC, DC, EUA. Agosto de 2003.
- (Zhou e Su, 2002) GuoDong Zhou e Jian Su. Named entity recognition using an HMM-based chunk tagger. Em *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, ACL'02*. Filadélfia, PA, EUA. Julho de 2002. p. 473–480.
- (Zipf, 1949) George K. Zipf. *Human Behaviour and the Principle of Least Effort*. Addison-Wesley. Reading, MA. 1949.