

References

- [AL03] Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, 2003.
- [ASM10] Amal Al-Saif and Katja Markert. The Leeds Arabic Discourse Treebank: Annotating discourse connectives for Arabic. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, page 8 pages, Valletta, Malta, 2010.
- [ASM11] Amal Al-Saif and Katja Markert. Modelling discourse relations for Arabic. In *EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011.
- [BE97] Regina Barzilay and Michael Elhadad. Using lexical chains for text summarization. In *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pages 10–17, 1997.
- [Bes06] Yves Bestgen. Improving text segmentation using Latent Semantic Analysis: A reanalysis of Choi, Wiemer-Hastings, and Moore (2001). *Computational Linguistics*, 32(1):5–12, 2006.
- [BKrKeM09] Matthias Buch-Kromann, Iørn Korzen, and Henrik Høeg Müller. Uncovering the ‘lost’ structure of translations with parallel treebanks. In Fabio Alves, Susanne Göpferich, and Inger Mees, editors, *Copenhagen Studies of Language: Methodology, Technology and Innovation in Translation Process Research*, Copenhagen Studies of Language, vol. 38, pages 199–224. Copenhagen Business School, 2009.
- [BL04] Regina Barzilay and Lillian Lee. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proceedings of the 2nd Human Language Technology Conference and Annual Meeting of the North American Chapter*, pages 113–120, 2004.
- [BL05] Jason Baldridge and Alex Lascarides. Probabilistic Head-driven Parsing for Discourse Structure. In *Proceedings of CoNLL'05*, 2005.
- [BMAC01] Jill Burstein, Daniel Marcu, Slava Andreyev, and Martin Chodorow. Towards automatic classification of discourse elements in essays. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 98–105, Toulouse, France, 2001. Association for Computational Linguistics.
- [BMK03] Jill Burstein, Daniel Marcu, and Kevin Knight. Finding the WRITE stuff: Automatic identification of discourse structure in student essays. *IEEE Intelligent Systems: Special Issue on Advances in Natural Language Processing*, 18:32–39, 2003.
- [CB08] Chris Callison-Birch. Syntactic constraints on paraphrases extracted from parallel corpora. In *Proceedings of EMNLP*, 2008.
- [CBBK09] Harr Chen, S. R. K. Branavan, Regina Barzilay, and David Karger. Global models of document structure using latent permutations. In *NAACL '09: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 371–379, 2009.
- [Chu09] Grace Chung. Sentence retrieval for abstracts of randomized controlled trials. *BMC Medical Informatics and Decision Making*, 10(9), February 2009.
- [CL10] James Clarke and Mirella Lapata. Discourse constraints for document compression. *Computational Linguistics*, 36:411–441, 2010.

- [CMO03] Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In J. van Kuppevelt & R. Smith, editor, *Current Directions in Discourse and Dialogue*. Kluwer, New York, 2003.
- [COCO98] Simon Corston-Oliver and Simon H. Corston-Oliver. Beyond string matching and cue phrases: Improving efficiency and coverage in discourse analysis. In *The AAAI Spring Symposium on Intelligent Text Summarization*, pages 9–15, 1998.
- [CW97] Dan Cristea and Bonnie Webber. Expectations in incremental discourse processing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL97/EACL97)*, pages 88–95, Madrid, Spain, 1997.
- [CWHM01] Freddy Choi, Peter Wiemer-Hastings, and Johanna Moore. Latent Semantic Analysis for text segmentation. In *EMNLP '01: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 109–117, 2001.
- [Dal92] Robert Dale. *Generating Referring Expressions*. MIT Press, Cambridge MA, 1992.
- [DM02] Hal Daume III and Daniel Marcu. A noisy-channel model for document compression. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 449–456, Philadelphia PA, 2002.
- [EB08a] Jacob Eisenstein and Regina Barzilay. Bayesian unsupervised topic segmentation. In *EMNLP '08: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 334–343, 2008.
- [EB08b] Robert Elwell and Jason Baldridge. Discourse connective argument identification with connective specific rankers. In *Proceedings of the IEEE Conference on Semantic Computing (ICSC-08)*, Santa Clara CA, 2008.
- [FIK10] George Foster, Pierre Isabelle, and Roland Kuhn. Translating structured documents. In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas*, Denver CO, 2010.
- [FKL98] Peter Foltz, Walter Kintsch, and Thomas Landauer. The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, 25:285–307, 1998.
- [GBC01] Hatem Ghorbel, Afzal Ballim, and Giovanni Coray. Rosetta: Rhetorical and semantic environment for text alignment. In *Proceedings of Corpus Linguistics 2001*, page 224–233, Lancaster, 2001.
- [GJW95] Barbara Grosz, Aravind Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.
- [GKL⁺10] Yufan Guo, Anna Korhonen, Maria Liakata, Ilona Silins, Lin Sun, and Ulla Stenius. Identifying the information structure of scientific abstracts: An investigation of three different schemes. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pages 99–107, Uppsala, Sweden, 2010.
- [GMFLJ03] Michel Galley, Kathleen McKeown, Eric Fosler-Lussier, and Hongyan Jing. Discourse segmentation of multi-party conversation. In *Proceedings of the 41st Annual Conference of the Association for Computational Linguistics*, 2003.
- [GS86] Barbara Grosz and Candy Sidner. Attention, intention and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

- [GS90] Barbara Grosz and Candace Sidner. Plans for discourse. In Philip Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*, pages 417–444. MIT Press, Cambridge MA, 1990.
- [HBMG04] Derrick Higgins, Jill Burstein, Daniel Marcu, and Claudia Gentile. Evaluating multiple aspects of coherence in student essays. In *Proceedings of HLT-NAACL*, pages 185–192, Boston MA, 2004. Association for Computational Linguistics.
- [HE10] Daniel Hardt and Jakob Elming. Incremental re-training for post-editing SMT. In *Proceedings of AMTA*, 2010.
- [Hea97] Marti Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64, 1997.
- [HF10] Christian Hardmeier and Marcello Federico. Modelling pronominal anaphora in Statistical Machine Translation. In *Proceedings of the 7th International Workshop on Spoken Language Translation*, Paris, 2010.
- [HMP⁺06] Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. OntoNotes: The 90% solution. In *Proceedings of HLT-NAACL*, pages 57–60, New York, 2006. Association for Computational Linguistics.
- [HOAI08] Kenji Hirohata, Naoki Okazaki, Sophia Ananiadou, and Mitsuru Ishizuka. Identifying sections in scientific abstracts using conditional random fields. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing*, pages 381–388, 2008.
- [HSAM93] Jerry R. Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142, 1993.
- [Joa02] Thorsten Joachims. *Learning to classify text using Support Vector Machines – methods, theory, and algorithms*. Kluwer/Springer, 2002.
- [KO11] Moshe Koppel and Noam Ordan. Translationese and its dialects. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1318–1326, Portland, Oregon, 2011.
- [KOOM01] Alistair Knott, Jon Oberlander, Michael O’Donnell, and Chris Mellish. Beyond elaboration: The interaction of relations and focus in coherent text. In Ted Sanders, Joost Schilperoord, and Wilbert Spooren, editors, *Text representation: linguistic and psycholinguistic aspects*, pages 181–196. Benjamins, Amsterdam, 2001.
- [KR93] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer Academic Publishers, 1993.
- [LKDFK06] Jimmy Lin, Damianos Karakos, Dina Demner-Fushman, and Sanjeev Khudanpur. Generative content models for structural analysis of medical abstracts. In *Proceedings of the HLT-NAACL Workshop on BioNLP*, pages 65–72, 2006.
- [LNK10] Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. A PDTB-styled end-to-end discourse parser. Technical report, Department of Computing, National University of Singapore, November 2010. <http://arxiv.org/abs/1011.0835>.
- [LTSB10] Maria Liakata, Simone Teufel, Advait Siddharthan, and Colin Batchelor. Corpora for the conceptualisation and zoning of scientific papers. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, 2010.

- [Mar97] Daniel Marcu. The Rhetorical Parsing of Natural Language Texts. In *Proceedings of ACL/EACL 1997*, 1997.
- [Mar98] Daniel Marcu. To build text summaries of high quality, nuclearity is not sufficient. In *Proceedings of AAAI-98 Spring Symposium on Intelligent Text Summarization*, 1998.
- [Mar99] Daniel Marcu. A decision-based approach to rhetorical parsing. In *Proceedings of ACL'99*, pages 365–372, 1999.
- [Mar00a] Daniel Marcu. The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*, 26:395–448, 2000.
- [Mar00b] Daniel Marcu. *The theory and practice of discourse parsing and summarization*. MIT Press, 2000.
- [Mar00c] James Martin. Beyond exchange: Appraisal systems in English. In S. Hunston and G. Thompson, editors, *Evaluation in Text: Authorial Distance and the Construction of Discourse*, pages 142–175. Oxford University Press, Oxford, 2000.
- [MB06] Igor Malioutov and Regina Barzilay. Minimum cut model for spoken lecture segmentation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 25–32, 2006.
- [MC07] Mstislav Maslennikov and Tat-Seng Chua. A multi-resolution framework for information extraction from free text. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 592–599, Prague, Czech Republic, 2007. Association for Computational Linguistics.
- [MCW00] Daniel Marcu, Lynn Carlson, and Maki Watanabe. The automatic translation of discourse structures. In *Proceedings of the 1st Conference of the North American Chapter of the ACL*, pages 9–17, 2000.
- [ME02] Daniel Marcu and Abdessamad Echihabi. An unsupervised approach to recognizing discourse relations. In *Proceedings of ACL'02*, 2002.
- [Mey11] Thomas Meyer. Disambiguating temporal-contrastive connectives for machine translation. In *Proceedings of the ACL 2011 Student Session*, pages 46–51, Portland, OR, June 2011.
- [MKMC06] Yoko Mizuta, Anna Korhonen, Tony Mullen, and Nigel Collier. Zone analysis in biology articles as a basis for information extraction. *International Journal of Medical Informatics*, 75:468–487, 2006.
- [Moe01] Marie-Francine Moens. Innovative techniques for legal text retrieval. *Artificial Intelligence and Law*, 9:29–57, 2001.
- [MS03] Larry McKnight and Padmini Srinivasan. Categorization of sentence types in medical abstracts. In *Proceedings of the AMIA Annual Symposium*, pages 440–444, 2003.
- [MT87] William C. Mann and Sandra A. Thompson. Antithesis: A study in clause combining and discourse structure. In *Language topics: Essays in Honour of Michael Halliday*, pages 359–384, Amsterdam, 1987. John Benjamins.
- [MT88] William Mann and Sandra Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

- [MUD99] Marie-Francine Moens, Caroline Uyttendaele, and Jos Dumortier. Information extraction from legal texts: the potential of discourse analysis. *International Journal of Human-Computer Studies*, 51:1155–1171, 1999.
- [MUD00] Marie-Francine Moens, Caroline Uyttendaele, and Jos Dumortier. Intelligent information extraction from legal texts. *Information & Communications Technology Law*, 9:17–26, 2000.
- [NC02] Vincent Ng and Claire Cardie. Identifying anaphoric and non-anaphoric noun phrases to improve coreference resolution. In *Proceedings of COLING 2002*, 2002.
- [NK10] Ronan Le Nagard and Philipp Koehn. Aiding pronoun translation with co-reference resolution. In *Proceedings of the Joint Workshop on Statistical Machine Translation and MetricsMATR*, Uppsala, Sweden, 2010.
- [NM01] Tadashi Nomoto and Yuji Matsumoto. A new approach to unsupervised text summarization. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 26–34, New York, NY, USA, 2001. ACM.
- [OPK⁺09] Umangi Oza, Rashmi Prasad, Sudheer Kolachina, Dipti Misra Sharma, and Aravind Joshi. The hindi discourse relation bank. In *Proc. 3rd ACL Language Annotation Workshop (LAW III)*, Singapore, 2009.
- [OSM94] Kenji Ono, Kazuo Sumita, and Seiji Miike. Abstract generation based on rhetorical structure extraction. In *COLING*, pages 344–348, 1994.
- [PDL⁺07] Rashmi Prasad, Nikhil Dinesh, Alan Lee, Aravind Joshi, and Bonnie Webber. Attribution and its annotation in the Penn Discourse TreeBank. *TAL (Traitement Automatique des Langues)*, 42(2), 2007.
- [PDL⁺08] Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, and et al. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, 2008.
- [PGKT06] Matthew Purver, Tom Griffiths, K.P. Körding, and Joshua Tenenbaum. Unsupervised topic modelling for multi-party spoken discourse. In *Proceedings of the 21st International Conf. on Computational Linguistics (COLING) and the 44th annual meeting of the Association for Computational Linguistics*, pages 17–24, 2006.
- [P JW10a] Rashmi Prasad, Aravind Joshi, and Bonnie Webber. Exploiting scope for shallow discourse parsing. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, page 8 pages, Valletta, Malta, 2010.
- [P JW10b] Rashmi Prasad, Aravind Joshi, and Bonnie Webber. Realization of discourse relations by other means: Alternative lexicalizations. In *Proceedings of COLING*, page 9 pages, Beijing, China, 2010.
- [PL05] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the ACL*, pages 115–124, 2005.
- [PLV02] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using Machine Learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, Philadelphia, July 2002, 2002. Association for Computational Linguistics.

- [PN09] Emily Pitler and Ani Nenkova. Using syntax to disambiguate explicit discourse connectives in text. In *ACL-IJCNLP '09: Proceedings of the 47th Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing*, 2009.
- [PPB11] Volha Petukhova, Laurent Prévot, and Harry Bunt. Multi-level discourse relations between dialogue units. In *Sixth Joint ACL - ISO Workshop on Interoperable Semantic Annotation (isa-6)*, pages 18–27, 2011.
- [PR07] S. Patwardhan and E. Riloff. Effective information extraction with semantic affinity patterns and relevant regions. In *Proceedings of the 2007 Conference on Empirical Methods in Natural Language Processing (EMNLP-07)*, 2007.
- [PSEH04] Massimo Poesio, Rosemary Stevenson, Barbara Di Eugenio, and Janet Hitzeman. Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30:309–363, 2004.
- [Pur11] Matthew Purver. Topic segmentation. In Gokhan Tur and Renato de Mori, editors, *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*. Wiley, Hoboken NJ, 2011.
- [PZ04] Livia Polanyi and Annie Zaenen. Contextual valence shifters. In *Proceedings of AAAI Spring Symposium on Attitude*, page 10 pages, 2004.
- [RBC⁺07] Patrick Ruch, Celia Boyer, Christine Chichester, Imad Tbahriti, Antoine Geissbühler, Paul Fabry, and et al. Using argumentation to extract key sentences from biomedical abstracts. *International Journal of Medical Informatics*, 76(2–3):195–200, 2007.
- [Sch02] Frank Schilder. Robust discourse parsing via discourse markers, topicality and position. *Natural Language Engineering*, 8(3):235–255, 2002.
- [Sib92] Penni Sibun. Generating text without trees. *Computational Intelligence*, 8(1):102–122, 1992.
- [SM03] Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of NAACL 2003*, 2003.
- [Ste04] Manfred Stede. The Potsdam Commentary Corpus. In *ACL Workshop on Discourse Annotation*, Barcelona, Spain, 2004.
- [Ste08a] Manfred Stede. Disambiguating rhetorical structure. *Research on Language and Computation*, 6:311–332, 2008.
- [Ste08b] Manfred Stede. RST revisited: Disentangling nuclearity. In *Subordination versus Coordination in Sentence and Text*. John Benjamins, Amsterdam, 2008.
- [TBS09] Maite Taboada, J. Brooke, and Manfred Stede. Genre-based paragraph classification for sentiment analysis. In *Proceedings of the 10th Annual SIGDIAL Conference on Discourse and Dialogue*, pages 62–70, London, 2009.
- [Teu10] Simone Teufel, editor. *The Structure of Scientific Articles*. CSLI Publications, 2010.
- [TM02] Simone Teufel and Marc Moens. Summarizing scientific articles - experiments with relevance and rhetorical status. *Computational Linguistics*, 28:409–445, 2002.
- [TMM04] Kristina Toutanova, Penka Markova, and Christopher D. Manning. The leaf path projection view of parse trees: Exploring string kernels for HPSG parse selection. In *EMNLP'04*, pages 166–173, 2004.

- [Tur02] Peter Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424, Philadelphia, July 2002, 2002. Association for Computational Linguistics.
- [TvdBPC04] G. Thione, M. van den Berg, L. Polanyi, and C. Culy. Hybrid text summarization: combining external relevance measures with structural analysis. In *Proceedings of the ACL 2004 Workshop Text Summarization Branches Out*, 2004.
- [UPN10] Vinícius Rodrigues Uzêda, Thiago Alexandre Salgueiro Pardo, and Maria Das Graças Volpe Nunes. A comprehensive comparative evaluation of RST-based summarization methods. *ACM Transactions on Speech and Language Processing*, 6:1–20, 2010.
- [vdVBB⁺11] Nynke van der Vliet, Ildikó Berzlánovich, Gosse Bouma, Markus Egg, and Gisela Redeker. Building a discourse-annotated Dutch text corpus. In *Beyond Semantics*, page 9 pages, Gottingen, Germany, February 2011.
- [VT07] Kimberly Voll and Maite Taboada. Not all words are created equal: Extracting semantic orientation as a function of adjective relevance. In *Proceedings of the 20th Australian Joint Conference on Artificial Intelligence*, pages 337–346, Gold Coast, Australia, 2007.
- [WG05] Florian Wolf and Edward Gibson. Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31:249–287, 2005.
- [WP07] Ben Wellner and James Pustejovsky. Automatically identifying the arguments of discourse connectives. In *Proceedings of the 2007 Conference on Empirical Methods in Natural Language Processing (EMNLP-07)*, 2007.
- [WSMP07] Marilyn Walker, Amanda Stent, Francois Mairesse, and Rashmi Prasad. Individual and domain adaptation in sentence planning for dialogue. *Journal of Artificial Intelligence Research*, 30:413–456, 2007.
- [ZDSc⁺10] Deniz Zeyrek, Işın Demirşahin, Ayıışı Sevdik-Çallı, Hale Ögel Balaban, İhsan Yalçınkaya, and Ümut Deniz Turan. The annotation scheme of the Turkish Discourse Bank and an evaluation of inconsistent annotations. In *Proceedings of the 4th Linguistic Annotation Workshop (LAW III)*, Uppsala, Sweden, 2010.
- [ZTB⁺09] Deniz Zeyrek, Ümut Deniz Turan, Cem Bozsahin, Ruket Cakici, and et al. Annotating Subordinators in the Turkish Discourse Bank. In *Proceedings of the 3rd Linguistic Annotation Workshop (LAW III)*, Singapore, 2009.
- [ZW08] Deniz Zeyrek and Bonnie Webber. A discourse resource for Turkish: Annotating discourse connectives in the METU corpus. In *Proceedings of the 6th Workshop on Asian Language Resources (ALR6)*, Hyderabad, India, 2008.