

Automatisches Textverstehen

Michael Strube

michael.strube at h-its.org

February 1, 2013

Course Description

A text is more than a sequence of sentences. For understanding a text, the reader needs to infer semantic and pragmatic relations between the sentences. In Computational Linguistics methods have been developed for capturing the specific character of text: models of local and global coherence, coreference resolution algorithms, theories describing the rhetorical, temporal, causal, and argumentative structure of text. In the classe, first a few classical papers dealing with these topics will be discussed. Then we will learn about current models, methods and algorithms. However, their benefit over simpler approaches can only be evaluated within NLP applications. Therefore, the class will also cover the evaluation of discourse processing methods in applications such as information extraction, question answering, automatic summarization, sentiment analysis, etc.

Textbooks

- Stede, Manfred (2012). *Discourse Processing*, Morgan & Claypool Publishers.

selected chapters from:

- Jurafsky, Daniel & James H. Martin (2008). *Speech and Language Processing*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall.
- Bird, Steven, Ewan Klein & Edward Loper (2009). *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit*. O'Reilly.

References

- Webber, Bonnie, Markus Egg & Valia Kordoni (2012). Discourse structure and language technology. In *Natural Language Engineering*, 18(4), pp.437-490. (http://journals.cambridge.org/repo_A86UtBSJ)

Grading

- weekly: 2 questions submitted to instructor and participation in discussion (1/3)
- once: paper presentation (Referat, 1/3)
- once: essay (Hausarbeit, at most 15 pages) or programming project (demo/executable code and short (5 pages) description) (1/3) – **Deadline: 2013/03/07**

Schedule

2012/10/25

Introduction, Overview about schedule, assignments, grading ...

(Stede, 2012; Webber et al., 2012; Webber & Joshi, 2012)

2012/11/08

Discourse structure

(Grosz & Sidner, 1986; Morris & Hirst, 1991; Passonneau & Litman, 1997)

Document structure in different genres

(Teufel & Moens, 2002; Power et al., 2003; Sporleder & Lapata, 2004; Filippova & Strube, 2006; Chung, 2009; Teufel et al., 2009; Liakata et al., 2010)

to be read: (Webber & Joshi, 2012)

2012/11/15

Introduction to local coherence analysis: Information Status, Coreference

Presenter: Michael Strube: (Hobbs, 1978)

Presenter: Lyubov Nakryiko: (Lappin & Leass, 1994; Kennedy & Boguraev, 1996)

to be read: (Hobbs, 1978) or (Lappin & Leass, 1994)

2012/11/22

Introduction to centering

Presenter: Michael Strube: (Brennan et al., 1987; Grosz et al., 1995; Poesio et al., 2004)

Introduction to global coherence analysis

Presenter: Mareike Hartmann: (Kehler et al., 2008)

to be read: (Hobbs, 1979) or (Hobbs, 1985) or (Grosz et al., 1995)

optional: (Prince, 1981; Grosz & Sidner, 1986; Mann & Thompson, 1988; Prince, 1992; Knott & Dale, 1994; Strube, 1998; Webber & Joshi, 1998; Strube & Hahn, 1999; Webber et al., 2012)

2012/11/29

Topic segmentation: Distributional approaches

Presenters: Chen Li (Beeferman et al., 1999), **Danny Rehl** (Pevzner & Hearst, 2002)

optional: (Reynar, 1999; Choi, 2000; Pevzner & Hearst, 2002)

to be read: (Hearst, 1997)

2012/12/06

Topic segmentation: Lexical chains

Presenters: Joachim Bingel (Ye & Chua, 2006; Ye et al., 2007), **Martina Lindahl**

(Galley et al., 2003)

optional: (Morris & Hirst, 1991; Barzilay & Elhadad, 1997; Hirst & St-Onge, 1998; Barzilay & Elhadad, 1999; Silber & McCoy, 2002; Stokes et al., 2004; Medelyan, 2007)

to be read: (Barzilay & Elhadad, 1999)

2012/12/13

Probabilistic models for topic segmentation

Presenter: Hans-Martin Ramsel: (Utiyama & Isahara, 2001; Eisenstein & Barzilay, 2008)

optional: (Shafiei & Milios, 2008; Chen et al., 2009)

Hierarchical topic segmentation

Presenter: Sokchea Kor: (Hsueh et al., 2006)

optional: (Grosz & Sidner, 1986; Eisenstein, 2009)

to be read: (Utiyama & Isahara, 2001) or (Hsueh et al., 2006)

2012/12/20

Segmentation applications: Summarization

Presenter: Annika Berger: (Stokes et al., 2004)

optional: (Goldstein et al., 2000; Teufel & Moens, 2002; Narayanan & Harabagiu, 2004)

Local coherence applications: Readability

Presenter: Thomas Bögeler: (Higgins et al., 2004; Heilman et al., 2007)

optional: (Miltasakaki & Kukich, 2000; 2004; Schwarm & Ostendorf, 2005; Miltasakaki & Troutt, 2008)

to be read: (Stokes et al., 2004) or (Higgins et al., 2004) or (Heilman et al., 2007)

2013/01/10

Local coherence: Entity grid with applications to readability, summary coherence rating

Presenter: Christoph Mayer: (Barzilay & Lapata, 2008; Elsner & Charniak, 2011)

optional: (Elsner et al., 2007; Filippova & Strube, 2007)

Local coherence applications: Information ordering

Presenter: Frank Marczewski: (Barzilay & Lapata, 2008; Karamanis et al., 2009)

optional: (Karamanis, 2007)

to be read: (Barzilay & Lapata, 2005) or (Karamanis et al., 2004)

2013/01/17

Local coherence: Information status

Presenter: Sariya Karimova: (Nissim et al., 2004; Nissim, 2006; Rahman & Ng,

2011a)

Local coherence: Fine-grained Information status

Presenter: Yufang Hou: (Markert et al., 2012)

optional: (Cahill & Riester, 2009; Riester et al., 2010; Riester & Baumann, 2011)

to be read: (Nissim, 2006)

2013/01/24

Local coherence: Machine learning for coreference resolution

Presenter: ~~Yulia Pilkevich~~ Michael Strube

(Soon et al., 2001; Ng & Cardie, 2002; Luo et al., 2004; Nicolae & Nicolae, 2006; Yang et al., 2008a; 2008b; Ng, 2008; 2010)

Local coherence: Recent work on coreference resolution

Presenter: Carolin Haas (Raghunathan et al., 2010; Lee et al., 2011)

optional: (Denis & Baldridge, 2007; Klenner, 2007; Finkel & Manning, 2008; Poon & Domingos, 2008; Denis & Baldridge, 2009; Sapena et al., 2010; Cai & Strube, 2010; Rahman & Ng, 2011b)

to be read: (Soon et al., 2001) or (Ng, 2010)

2013/01/31

Local coherence applications: Coreference and summarization, QA

Presenter: Magnus Bodén: (Vicedo & Ferrández, 2006)

optional: (Azzam et al., 1999; Boguraev & Kennedy, 1999; Morton, 2000; Stuckhardt, 2003; Watson et al., 2003; Steinberger et al., 2005; 2007)

Global coherence: Rhetorical Structure Theory (RST)

Presenter: Huiqin Körkel-Qu: (Feng & Hirst, 2012)

optional: (Marcu & Echihiabi, 2002; Carlson et al., 2003; Soricut & Marcu, 2003; Subba & Di Eugenio, 2009; Hernault et al., 2012)

to be read: (Vicedo & Ferrández, 2006) or (Feng & Hirst, 2012)

2013/02/07

Global coherence: Penn Discourse Treebank (PDTB)

Presenter: Angela Schneider/~~Eric Hildebrand~~: Intro to PDTB and (Pitler & Nenkova, 2009b)

optional: (Pitler & Nenkova, 2009a; Ghosh et al., 2012; Lin et al., 2012)

ALL: Discussion about properties of discourse structure (segments, local and global coherence), multilinguality, applications, future directions, ...

to be read: (Prasad et al., 2008) – send me **one** question about this paper and **two** questions about issues/topics from the whole term you want to discuss

Optional:

Global coherence: Applications

(Marcu, 1997; Maslennikov & Chua, 2007; Clarke & Lapata, 2010; Zirn et al., 2011)

Notes

The questions about the material **to be read** will be part of the grading. Please send me your questions via email until 1pm on the day of the class.

Attendance in class will be checked. If you really cannot attend one class, let me know beforehand via email and produce a (doctor's) certificate afterwards.

The *Hausarbeit* should be at most 15 pages including references. It can be written in German or English. Please submit a PDF-file as I don't have Microsoft software and cannot guarantee that I can read and print Microsoft Word generated files correctly. Please don't forget to put your name and Matrikelnummer on the front page.

Literature: Most of the literature cited can be accessed via the ACL anthology (<http://www.aclweb.org/anthology/>), in particular all papers which appeared in the *Computational Linguistics Journal*, the (*E/NA*)*ACL*, *IJCNLP*, and *EMNLP* conferences. Many other journals are available electronically through the university library (UB).

References

- Azzam, Saliha, Kevin Humphreys & Robert Gaizauskas (1999). Using coreference chains for text summarization. In *Proceedings of the Workshop on Coreference and Its Applications*, College Park, Md., 22 June 1999, pp. 77–84.
- Barzilay, Regina & Michael Elhadad (1997). Using lexical chains for text summarization. In *Proceedings of the ACL Workshop on Intelligent and Scalable Text Summarization, Madrid, Spain, July 1997*, pp. 10–17.
- Barzilay, Regina & Michael Elhadad (1999). Using lexical chains for text summarization. In Inderjeet Mani & Mark T. Maybury (Eds.), *Advances in Automatic Text Summarization*, pp. 111–121. Cambridge, Mass.: MIT Press.
- Barzilay, Regina & Mirella Lapata (2005). Modeling local coherence: An entity-based approach. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pp. 141–148.
- Barzilay, Regina & Mirella Lapata (2008). Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.
- Beeferman, Doug, Adam L. Berger & John Lafferty (1999). Statistical models for text segmentation. *Machine Learning*, 34:177–210.
- Bird, Steven, Ewan Klein & Edward Loper (2009). *Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit*. O'Reilly.
- Boguraev, Branimir & Christopher Kennedy (1999). Salience-based content characterisation of text documents. In Inderjeet Mani & Mark T. Maybury (Eds.), *Advances in Automatic Text Summarization*, pp. 99–110. Cambridge, Mass.: MIT Press.
- Brennan, Susan E., Marilyn W. Friedman & Carl J. Pollard (1987). A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Stanford, Cal., 6–9 July 1987, pp. 155–162.

- Cahill, Aoife & Arndt Riester (2009). Incorporating information status into generation ranking. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing*, Singapore, 2–7 August 2009, pp. 817–825.
- Cai, Jie & Michael Strube (2010). End-to-end coreference resolution via hypergraph partitioning. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China, 23–27 August 2010, pp. 143–151.
- Carlson, Lynn, Daniel Marcu & Mary Ellen Okurowski (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In J. van Kuppevelt & R. Smith (Eds.), *Current Directions in Discourse and Dialogue*, pp. 85–112. Dordrecht, The Netherlands: Kluwer.
- Chen, Harr, S.R.K. Branavan, Regina Barzilay & David R. Karger (2009). Content modeling using latent permutations. *Journal of Artificial Intelligence Research*, 36:129–163.
- Choi, Freddy Y. Y. (2000). Advances in domain independent linear text segmentation. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, Wash., 29 April – 3 May 2000, pp. 26–33.
- Chung, Grace (2009). Sentence retrieval for abstracts of randomized trials. *BMC Medical Informatics and Decision Making*, 9(10). 13 pages.
- Clarke, James & Mirella Lapata (2010). Discourse constraints for document compression. *Computational Linguistics*, 36(3):411–441.
- Denis, Pascal & Jason Baldridge (2007). Joint determination of anaphoricity and coreference resolution using integer programming. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, Rochester, N.Y., 22–27 April 2007, pp. 236–243.
- Denis, Pascal & Jason Baldridge (2009). Global joint models for coreference resolution and named entity classification. *Procesamiento del Lenguaje Natural*, 42:87–96.
- Eisenstein, Jacob (2009). Hierarchical text segmentation from multi-scale lexical cohesion. In *Proceedings of Human Language Technologies 2009: The Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, Col., 31 May – 5 June 2009, pp. 353–361.
- Eisenstein, Jacob & Regina Barzilay (2008). Bayesian unsupervised topic segmentation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25–27 October 2008, pp. 334–343.
- Elsner, Micha, Joseph Austerweil & Eugene Charniak (2007). A unified local and global model for discourse coherence. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, Rochester, N.Y., 22–27 April 2007, pp. 436–443. Read this version: <http://www.cs.brown.edu/~melsner/order.pdf>.
- Elsner, Micha & Eugene Charniak (2011). Extending the entity grid with entity-specific features. In *Proceedings of the ACL 2011 Conference Short Papers*, Portland, Oreg., USA, 19–24 June 2011, pp. 125–129.
- Feng, Vanessa Wei & Graeme Hirst (2012). Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju Island, Korea, 8–14 July 2012, pp. 60–68.
- Filippova, Katja & Michael Strube (2006). Using linguistically motivated features for paragraph segmentation. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, 22–23 July 2006, pp. 267–274.
- Filippova, Katja & Michael Strube (2007). Extending the entity-grid coherence model to semantically related entities. In *Proceedings of the 11th European Workshop on Natural Language Generation*, Schloss Dagstuhl, Germany, 17–20 June 2007, pp. 139–142.
- Finkel, Jenny Rose & Christopher Manning (2008). Enforcing transitivity in coreference resolution. In *Companion Volume to the Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, 15–20 June 2008, pp. 45–48.
- Galley, Michel, Kathleen R. McKeown, Eric Fosler-Lussier & Hongyan Jing (2003). Discourse

- segmentation of multi-party conversation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 562–569.
- Ghosh, Sucheta, Giuseppe Riccardi & Richard Johansson (2012). Global features for shallow discourse parsing. In *Proceedings of the SIGdial 2012 Conference: The 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Seoul, Korea, 5–6 July 2012, pp. 150–159.
- Goldstein, Jade, Vibhu Mittal, Jaime Carbonell & Mark Kantrowitz (2000). Multi-document summarization by sentence extraction. In *Proceedings of the Workshop on Automatic Summarization at ANLP/NAACL 2000*, Seattle, Wash., 30 April 2000, pp. 40–48.
- Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Grosz, Barbara J. & Candace L. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Hearst, Marti A. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.
- Heilman, Michael J., Kevyn Collins-Thompson, Jamie Callan & Maxine Eskenazi (2007). Combining lexical and grammatical features to improve readability measures for first and second language texts. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, Rochester, N.Y., 22–27 April 2007, pp. 460–467.
- Hernault, Hugo, Helmut Prendinger, David A. duVerle & Mitsuru Ishizuka (2012). HILDA: a discourse parser using support vector classification. *Dialogue and Discourse*, 1(3):1–33.
- Higgins, Derrick, Jill Burstein, Daniel Marcu & Claudia Centile (2004). Evaluating multiple aspects of coherence in student essays. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Boston, Mass., 2–7 May 2004, pp. 185–192.
- Hirst, Graeme & David St-Onge (1998). Lexical chains as representations of context for the detection and correction of malapropisms. In Christiane Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*, pp. 305–332. Cambridge, Mass.: MIT Press.
- Hobbs, Jerry R. (1978). Resolving pronominal references. *Lingua*, 44:311–338.
- Hobbs, Jerry R. (1979). Coherence and coreference. *Cognitive Science*, 3:67–90.
- Hobbs, Jerry R. (1985). *On the coherence and structure of discourse*. Technical Report CSLI-85-37: Stanford: SRI/CSLI.
- Hsueh, Pei-Yun, Johanna D. Moore & Steve Renals (2006). Automatic segmentation of multiparty dialogue. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 3–7 April 2006, pp. 273–280.
- Jurafsky, Daniel & James H. Martin (2008). *Speech and Language Processing* (2nd. ed.). Upper Saddle River, N.J.: Prentice Hall.
- Karamanis, Nikiforos (2007). Supplementing entity coherence with local rhetorical relations for information ordering. *Journal of Logic, Language and Information*, 16(4):445–464.
- Karamanis, Nikiforos, Chris Mellish, Massimo Poesio & Jon Oberlander (2009). Evaluating centering for information ordering using corpora. *Computational Linguistics*, 35(1):29–46.
- Karamanis, Nikiforos, Massimo Poesio, Chris Mellish & Jon Oberlander (2004). Evaluating centering-based metrics of coherence for text structuring using a reliably annotated corpus. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 21–26 July 2004, pp. 392–393.
- Kehler, Andrew, Laura Kertz, Hannah Rohde & Jeffrey R. Elman (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44.
- Kennedy, Christopher & Branimir Boguraev (1996). Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, 5–9 August 1996, Vol. 1, pp. 113–118.
- Klenner, Manfred (2007). Enforcing consistency on coreference sets. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Borovets, Bulgaria, 27–29 September 2007, pp. 323–328.

- Knott, Alistair & Robert Dale (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1):35–62.
- Lappin, Shalom & Herbert J. Leass (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561.
- Lee, Heeyoung, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu & Dan Jurafsky (2011). Stanford’s multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In *Proceedings of the Shared Task of the 15th Conference on Computational Natural Language Learning*, Portland, Oreg., 23–24 June 2011, pp. 28–34.
- Liakata, Maria, Simone Teufel, Advait Siddharthan & Colin Batchelor (2010). Corpora for the conceptualisation and zoning of scientific papers. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, La Valetta, Malta, 17–23 May 2010.
- Lin, Ziheng, Hwee Tou Ng & Min-Yen Kan (2012). A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, ??(??):?? To appear.
- Luo, Xiaoqiang, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla & Salim Roukos (2004). A mention-synchronous coreference resolution algorithm based on the Bell Tree. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 21–26 July 2004, pp. 136–143.
- Mann, William C. & Sandra A. Thompson (1988). Rhetorical structure theory. Toward a functional theory of text organization. *Text*, 8(3):243–281.
- Marcu, Daniel (1997). The rhetorical parsing of natural language texts. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and of the 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, 7–12 July 1997, pp. 365–372.
- Marcu, Daniel & Abdessamad Echihabi (2002). An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Penn., 7–12 July 2002, pp. 368–375.
- Markert, Katja, Yufang Hou & Michael Strube (2012). Collective classification for fine-grained information status. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju Island, Korea, 8–14 July 2012, pp. 795–804.
- Maslennikov, Mstislav & Tat-Seng Chua (2007). A multi-resolution framework for information extraction from free text. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, 23–30 June 2007, pp. 592–599.
- Medelyan, Olena (2007). Computing lexical chains with graph clustering. In *Proceedings of the ACL 2007 Student Research Workshop*, Prague, Czech Republic, 25–26 June 2007, pp. 85–90.
- Miltsakaki, Eleni & Karen Kukich (2000). The role of centering theory’s rough-shifts in the teaching and evaluation of writing skills. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, China, 1–8 August 2000, pp. 408–415.
- Miltsakaki, Eleni & Karen Kukich (2004). Evaluation of text coherence for electronic essay scoring systems. *Natural Language Engineering*, 10(1):25–55.
- Miltsakaki, Eleni & Audrey Troutt (2008). Real-time web text classification and analysis of reading difficulty. In *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications at ACL’08*, Columbus, Ohio, 19 June 2008, pp. 89–97.
- Morris, Jane & Graeme Hirst (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1):21–48.
- Morton, Thomas S. (2000). Coreference for NLP applications. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, China, 1–8 August 2000, pp. 173–180.
- Narayanan, Srinivas & Sanda A. Harabagiu (2004). Answering questions using advanced semantics and probabilistic inference. In *Proceedings of the Pragmatics of Question Answering at HLT-NAACL 2004*, Boston, Mass., USA, 6–7 May 2004, pp. 10–16.
- Ng, Vincent (2008). Unsupervised models for coreference resolution. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25–27 October 2008, pp. 640–649.

- Ng, Vincent (2010). Supervised noun phrase coreference research: The first fifteen years. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, 11–16 July 2010, pp. 1396–1411.
- Ng, Vincent & Claire Cardie (2002). Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Penn., 7–12 July 2002, pp. 104–111.
- Nicolae, Cristina & Gabriel Nicolae (2006). BestCut: A graph algorithm for coreference resolution. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, 22–23 July 2006, pp. 275–283.
- Nissim, Malvina (2006). Learning information status of discourse entities. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, 22–23 July 2006, pp. 94–102.
- Nissim, Malvina, Shipara Dingare, Jean Carletta & Mark Steedman (2004). An annotation scheme for information status in dialogue. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 26–28 May 2004, pp. 1023–1026.
- Passonneau, Rebecca & Diane J. Litman (1997). Discourse segmentation by human and automated means. *Computational Linguistics*, 23(1):103–139.
- Pevzner, Lev & Marti Hearst (2002). A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28(1):19–36.
- Pitler, Emily & Ani Nenkova (2009a). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing*, Singapore, 2–7 August 2009, pp. 683–691.
- Pitler, Emily & Ani Nenkova (2009b). Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Singapore, 2–7 August 2009, pp. 13–16.
- Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio & Janet Hitzeman (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3). 309–363.
- Poon, Hoifung & Pedro Domingos (2008). Joint unsupervised coreference resolution with Markov Logic. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25–27 October 2008, pp. 650–659.
- Power, Richard, Donia Scott & Nadjat Bouayad-Agha (2003). Document structure. *Computational Linguistics*, 29(2):211–260.
- Prasad, Rashmi, Nikhil Dineh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi & Bonnie Webber (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, 26 May – 1 June 2008.
- Prince, Ellen F. (1981). Towards a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics*, pp. 223–255. New York, N.Y.: Academic Press.
- Prince, Ellen F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W.C. Mann & S.A. Thompson (Eds.), *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising Text*, pp. 295–325. Amsterdam: John Benjamins.
- Raghunathan, Karthik, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky & Christopher Manning (2010). A multi-pass sieve for coreference resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, Mass., 9–11 October 2010, pp. 492–501.
- Rahman, Altaf & Vincent Ng (2011a). Learning the information status of noun phrases in spoken dialogues. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, U.K., 27–29 July 2011, pp. 1069–1080.
- Rahman, Altaf & Vincent Ng (2011b). Narrowing the modeling gap: A cluster-ranking approach to coreference resolution. *Journal of Artificial Intelligence Research*, 40:469–521.
- Reynar, Jeffrey C. (1999). Statistical models for topic segmentation. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Md., 20–26 June 1999, pp. 357–364.

- Riester, Arndt & Stefan Baumann (2011). Information structure annotation and secondary accents. In S. Dipper & H. Zinsmeister (Eds.), *Corpus-based Investigations of Pragmatic and Discourse Phenomena*, Vol. 3, Bochumer Linguistische Arbeitsberichte, pp. 111–127. University of Bochum, Bochum, Germany.
- Riester, Arndt, David Lorenz & Nina Seemann (2010). A recursive annotation scheme for referential information status. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, La Valetta, Malta, 17–23 May 2010, pp. 717–722.
- Sapena, Emili, Lluís Padró & Jordi Turmo (2010). A global relaxation labeling approach to coreference resolution. In *Proceedings of Coling 2010: Poster Volume*, Beijing, China, 23–27 August 2010, pp. 1086–1094.
- Schwarm, Sarah E. & Mari Ostendorf (2005). Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pp. 523–530.
- Shafiei, M. Mahdi & Evangelos E. Milios (2008). A statistical model for topic segmentation and clustering. In *Proceedings of the 21st Canadian Conference on Artificial Intelligence*, Windsor, Ontario, Canada, 28–30 May 2008, pp. 283–295.
- Silber, Gregory H. & Kathleen F. McCoy (2002). Efficiently computed lexical chains as an intermediate representation for automatic text summarization. *Computational Linguistics*, 28(4):487–496.
- Soon, Wee Meng, Hwee Tou Ng & Daniel Chung Yong Lim (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.
- Soricut, Radu & Daniel Marcu (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Edmonton, Alberta, Canada, 27 May–1 June 2003, pp. 149–156.
- Sporleder, Caroline & Mirella Lapata (2004). Automatic paragraph identification: A study across languages and domains. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, 25–26 July 2004, pp. 72–79.
- Stede, Manfred (2012). *Discourse Processing*. Morgan & Claypool Publishers.
- Steinberger, Josef, Mijail A. Kabadjov & Massimo Poesio (2005). Improving LSA-based summarization with anaphora resolution. In *Proceedings of the Human Language Technology Conference and the 2005 Conference on Empirical Methods in Natural Language Processing*, Vancouver, B.C., Canada, 6–8 October 2005, pp. 1–8.
- Steinberger, Josef, Massimo Poesio, Mijail A. Kabadjov & Karel Ježek (2007). Two uses of anaphora resolution in summarization. *Information Processing and Management*, 43(6):1663–1680.
- Stokes, Nicola, Joe Carthy & Alan F. Smeaton (2004). SeLeCT: a lexical cohesion based news story segmentation system. *AI Communications*, 17(1):3–12.
- Strube, Michael (1998). Never look back: An alternative to centering. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Québec, Canada, 10–14 August 1998, Vol. 2, pp. 1251–1257.
- Strube, Michael & Udo Hahn (1999). Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.
- Stuckhardt, Roland (2003). Coreference-based summarization and question answering: A case for high precision anaphor resolution. In *Proceedings of the 2003 International Symposium on Reference Resolution and Its Applications to Question Answering and Summarization*, Venice, Italy, 23–24 June 2003, pp. 33–42.
- Subba, Rajen & Barbara Di Eugenio (2009). An effective discourse parser that uses rich linguistic information. In *Proceedings of Human Language Technologies 2009: The Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, Col., 31 May – 5 June 2009, pp. 566–574.
- Teufel, Simone & Marc Moens (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409–445.

- Teufel, Simone, Advaith Siddharthan & Colin Batchelor (2009). Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapore, 6–7 August 2009, pp. 1493–1502.
- Utiyama, Masao & Hitoshi Isahara (2001). A statistical model for domain-independent text segmentation. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, pp. 499–506.
- Vicedo, Jose L. & Antonio Ferrández (2006). Coreference in q&a. In T. Strzalkowski & S. Harabagiu (Eds.), *Advances in Open Domain Question Answering*, pp. 71–96. Dordrecht, The Netherlands: Springer.
- Watson, Rebecca, Judita Preiss & Ted Briscoe (2003). The contribution of domain-independent robust pronominal anaphora resolution to open-domain question-answering. In *Proceedings of the 2003 International Symposium on Reference Resolution and Its Applications to Question Answering and Summarization*, Venice, Italy, 23–24 June 2003, pp. 75–82.
- Webber, Bonnie, Markus Egg & Valia Kordoni (2012). Discourse structure and language technology. *Natural Language Engineering*, 18(4):437–490.
- Webber, Bonnie & Aravind Joshi (2012). Discourse structure and computation: Past, present and future. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*, Jeju Island, Korea, 10 July 2012, pp. 42–54.
- Webber, Bonnie L. & Aravind K. Joshi (1998). Anchoring a lexicalized Tree-Adjoining Grammar for discourse. In *COLING-ACL '98 Workshop on Discourse Relations and Discourse Markers*, Montréal, Québec, Canada, 15 August, 1998, pp. 86–92.
- Yang, Xiaofeng, Jian Su, Jun Lang, Chew Lim Tan, Ting Liu & Sheng Li (2008a). An entity-mention model for coreference resolution with Inductive Logic Programming. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Columbus, Ohio, 15–20 June 2008, pp. 843–851.
- Yang, Xiaofeng, Jian Su & Chew Lim Tan (2008b). A twin-candidate model for learning-based anaphora resolution. *Computational Linguistics*, 34(3):327–356.
- Ye, Shiren & Tat-Seng Chua (2006). NUS at DUC 2006: Document concept lattice for summarization. In *Proceedings of the 2006 Document Understanding Conference held at the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, New York, N.Y., 8–9 June 2006.
- Ye, Shiren, Tat-Seng Chua, Min-Yen Kan & Long Qiu (2007). Document concept lattice for text understanding and summarization. *Information Processing and Management*, 43:1643–1663.
- Zirn, Căcilia, Matthias Niepert, Heiner Stuckenschmidt & Michael Strube (2011). Fine-grained sentiment analysis with structural features. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, Chiang Mai, Thailand, 8–13 November 2011, pp. 336–344.