Text Processing in NLP and its Applications

Michael Strube

`michael.strube@h-its.org`

May 5, 2012

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. After giving a short introduction to Natural Language Processing (NLP), these models, methods and algorithms will be discussed. However, their benefit over simpler approaches can only be evaluated within NLP applications. Therefore, the class will also cover applications processing texts such as information extraction, question answering, automatic summarization, sentiment analysis, etc.

Objectives

Students should understand which NLP application requires which discourse processing component. Students should be able to extend applications by such components and evaluate them.

Textbooks

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, to appear. (`http://journals.cambridge.org/repo_A84ql5gR`)

Evaluation

- project and project presentation – maybe in small groups (30%)

- paper presentation (20%)

- questions submitted to instructor and/or exercises solved (30%)

- participation in discussion (20%)

Schedule

2012/03/02
Overview about schedule, assignments, grading . . .

to be read for lecture on 2012/03/09: (Jurafsky & Martin, 2008a, Chapter 1)

2012/03/09
Introduction to NLP I: Preprocessing, Tokenization, POS Tagging

homework for lecture on 2012/03/16:
`http://michael.kimstrube.de/tnlpa/homework120309.pdf`

2012/03/16
Introduction to NLP II: Syntax, Parsing

homework for lecture on 2012/03/23:
`http://michael.kimstrube.de/tnlpa/homework120316.pdf`

2012/03/23
Introduction to NLP III: Computational Semantics

homework for lecture on 2012/03/30:
`http://michael.kimstrube.de/tnlpa/homework120323.pdf`

2012/03/30
Introduction to NLP IV: Lexical Semantics

homework for lecture on 2012/04/06:
`http://michael.kimstrube.de/tnlpa/homework120330.pdf`

2012/04/06
a. Text structure: local and global coherence (Halliday & Hasan, 1976; Kehler, 2000; Moore & Wiemer-Hastings, 2003; Jurafsky & Martin, 2008b; Webber et al., 2012)
b. Model: Centering (Poesio et al., 2004)
homework for lecture on 2012/0f/13:
`http://michael.kimstrube.de/tnlpa/homework120406.pdf`

2012/04/13
a. Application: Anaphora resolution with Centering (Brennan et al., 1987; Kehler, 1997; Strube, 1998; Strube & Hahn, 1999; Tetreault, 2001)
b. Application:Evaluating readability with Centering (Miltsakaki & Kukich, 2000; 2004; Higgins et al., 2004; Heilman et al., 2007; Miltsakaki & Troutt, 2008)

2012/04/20
no class

2012/04/27
a. Application: Information ordering with Centering (Karamanis et al., 2004; Karamanis, 2007; Karamanis et al., 2009)
b. Application: Information ordering with language models, LSA, machine learning (Foltz et al., 1998; Althaus et al., 2004; Lapata, 2006)
c. Method Information structure/information status (Prince, 1981; Lambrecht, 1994; Kruijff-Korbayová & Steedman, 2003; Nissim et al., 2004; Nissim, 2006; Rahman & Ng, 2011b; Markert et al., 2012)

2012/05/04
a. Model: Lexical cohesion, lexical chains (Morris & Hirst, 1991; Galley et al., 2003)
b. Application: Automatic summarization using lexical chains (Barzilay & Elhadad, 1999; Silber & McCoy, 2002)

2012/05/11
Presenter: Youngsam Kim
a. Model: Entity grid (Barzilay & Lapata, 2008; Elsner & Charniak, 2011)

Presenter: Michael Strube
b. Application: . . . using the entity grid (Elsner et al., 2007; Barzilay & Lapata, 2008; Filippova & Strube, 2007)

2012/05/18
Presenter: Carlos Alfonso Torres Fernandez
b. Application: Discourse structure and question answering (Verberne et al., 2007; Verberne, 2007a; 2007b)

Presenter: Michael Strube
a. Model: Discourse structure (RST, DLTAG) (Mann & Thompson, 1987; 1988; Webber & Joshi, 1998; Webber, 2004; Prasad & Joshi, 2008)
c. Application: Discourse structure and sentiment analysis (Zirn et al., 2011)
d. Application: Discourse structure and automatic summarization (Marcu, 1997; 1999; 2000)

2012/05/25
Presenter: Michael Strube
a. Model: Introduction to coreference resolution (task, linguistic issues, corpora, evaluation) (Strube, 2007; Ng, 2010)

Presenter: Sang Woo Lee
b. Model: Machine learning for coreference resolution (Soon et al., 2001; Ng & Cardie, 2002; Luo et al., 2004; Ponzetto & Strube, 2006; Yang et al., 2008a; 2008b; Bengtson & Roth, 2008; Stoyanov et al., 2009; Rahman & Ng, 2011a)

2012/06/01
Presenter: Ji-Seob Kim
a. Model: Graph-based approaches to coreference resolution (Nicolae & Nicolae, 2006; Cai & Strube, 2010; Sapena et al., 2010)

Presenter: Michael Strube
b. Application: Coreference resolution for automatic summarization; coreference resolution for question answering (Azzam et al., 1999; Boguraev & Kennedy, 1999; Morton, 2000; Stuckhardt, 2003; Watson et al., 2003; Steinberger et al., 2005; 2007)

2012/06/08
Presenter: Eun-Sol Kim
a. Model: Text segmentation (Hearst, 1997; Passonneau & Litman, 1997; Beeferman et al., 1999; Reynar, 1999; Choi, 2000; Pevzner & Hearst, 2002; Galley et al., 2003; Chen et al., 2009)

Presenter: Michael Strube
b. Application: Segmentation for automatic summarization (Goldstein et al., 2000; Teufel & Moens, 2002; Stokes et al., 2004; Narayanan & Harabagiu, 2004)

2012/06/15
a. Project presentations
b. Future Directions
c. Wrap-up

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 Thursday before each 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.

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.

References

Althaus, Ernst, Nikiforos Karamanis & Alexander Koller (2004). Computing locally coherent discourses. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics,* Barcelona, Spain, 21–26 July 2004, pp. 400–407.

Azzam, Saliha, Kevin Humphreys & Robert Gaizauskas (1999). Using coreference chains for text summarization. In *Proceedings of the Workshop on Coreference and It's Applications,* College Park, Md., 22 June 1999, pp. 77–84.

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 (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.

Bengtson, Eric & Dan Roth (2008). Understanding the value of features for coreference resolution. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing,* Waikiki, Honolulu, Hawaii, 25–27 October 2008, pp. 294–303.

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.

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.

Chen, Harr, S.R.K. Branavan, Regina Barzilay & David R. Karger (2009). Global models of document structure using latent permutations. 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. 371–379.

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.

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, Michae & 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.

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.

Foltz, Peter, Walter Kintsch & Thomas Landauer (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, 25(2):285–307.

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.

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.

Halliday, M. A. K. & Ruqaiya Hasan (1976). *Cohesion in English.* London, U.K.: Longman.

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.

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.

Jurafsky, Daniel & James H. Martin (2008a). *Speech and Language Processing* (2nd. ed.)., Chp. 1. Introduction. Upper Saddle River, N.J.: Prentice Hall.

Jurafsky, Daniel & James H. Martin (2008b). *Speech and Language Processing* (2nd. ed.)., Chp. 21. Computational Discourse, pp. 681–723. 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 (1997). Current theories of centering for pronoun interpretation: A critical evaluation. *Computational Linguistics*, 23(3):467–475.

Kehler, Andrew (2000). Discourse. In Daniel Jurafsky & James H. Martin (Eds.), *Speech and Language Processing*, pp. 669–718. Upper Saddle River, N.J.: Prentice Hall.

Kruijff-Korbayová, Ivana & Mark Steedman (2003). Discourse and information structure. *Journal of Logic, Language and Information. Special Issue on Discource and Information Structure*, 12(3):149–259.

Lambrecht, Knud (1994). *Information Structure and Sentence Form.* Cambridge, U.K.: Cambridge University Press.

Lapata, Mirella (2006). Automatic evaluation of information ordering: Kendall's tau. *Computational Linguistics*, 32(4):471–484.

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 (1987). *A theory of text organisation.* Technical Report ISI/RS-87-190: Information Sciences Institute at the University of Southern California, Marina del Rey, Cal. Available at: http://www.sfu.ca/rst/pdfs/Mann_Thompson_1987.pdf.

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 (1999). Discourse trees are good indicators of importance in text. In I. Mani & M.T. Maybury (Eds.), *Advances in Automatic Text Summarization*, pp. 123–136. Cambridge, Mass.: MIT Press.

Marcu, Daniel (2000). *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, Mass.: The MIT Press.

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. To appear.

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.

Moore, Johanna D. & Peter Wiemer-Hastings (2003). Discourse in Computational Linguistics and Artificial Intelligence. In A. Graesser, M. Gernsbacher & S. Goldman (Eds.), *Handbook of Discourse Processes*, pp. 439–486. Mahwah, N.J.: Lawrence Erlbaum.

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 (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. 94–012.

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.

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio & Janet Hitzeman (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3). 309-363.

Ponzetto, Simone Paolo & Michael Strube (2006). Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, New York, N.Y., 4–9 June 2006, pp. 192–199.

Prasad, Rashmi & Aravind Joshi (2008). A discourse-based approach to generating why-questions from texts. In *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*, Arlington, Va., September 2008, p. ?? download at: http://www.cis.upenn.edu/ rjprasad/papers/prasad-joshi-qg08.pdf.

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.

Rahman, Altaf & Vincent Ng (2011a). Coreference resolution with world knowledge. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,* Portland, Oreg., USA, 19–24 June 2011, pp. 814–824.

Rahman, Altaf & Vincent Ng (2011b). 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.

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.

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.

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.

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). A lexical cohesion based news story segmentation system. *AI Communications*, 17(1):3–12.

Stoyanov, Veselin, Nathan Gilbert, Claire Cardie & Ellen Riloff (2009). Conundrums in noun phrase coreference resolution: Making sense of the state-of-the-art. 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. 656–664.

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 (2007). Corpus-based and machine learning approaches to coreference resolution. In M. Schwarz-Friesel, M. Consten & M. Knees (Eds.), *Anaphors in Text. Cognitive, formal and applied approaches to anaphoric reference*, pp. 207–222. Amsterdam, The Netherlands: John Benjamins.

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 prcision 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.

Tetreault, Joel R. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27(4):507–520.

Teufel, Simone & Marc Moens (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409–445.

Verberne, Suzan (2007a). Evaluating answer extraction for Why-QA using RST-annotated Wikipedia texts. In *Proceedings of the Twelfth ESSLLI Student Session,* Dublin, Ireland, 6–17 August 2007, pp. 255–266.

Verberne, Suzan (2007b). Paragraph retrieval for why-question answering. Exploiting discourse structure for intelligent paragraph retrieval for why-QA. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Amsterdam, The Netherlands, 23–27 July 2007, pp. 922–926.

Verberne, Suzan, Lou Boves, Nelleke Oostdijk & Peter-Arno Coppen (2007). Discourse-based answering of why-questions. *Traitement Automatique des Langues*, 47(2):21–41.

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 (2004). D-LTAG: Extending lexicalized TAG to discourse. *Cognitive Science*, 28(5):751–779.

Webber, Bonnie, Markus Egg & Valia Kordoni (2012). Discourse structure and language technology. *Natural Language Engineering*. To appear.

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.

Zirn, Cäcilia, 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.