

The State of the Art in Computational Linguistics: How to get at  
information encoded in natural language

Amazon Developer Conference

January 17th, 2007

Emily M. Bender

Assistant Professor & Faculty Director

Professional Master's in Computational Linguistics

University of Washington

ebender@u.washington.edu

## References

- Baker, Collin F., and Beau Fillmore, Charles J. Cronin. 2003. The structure of the Framenet database. *International Journal of Lexicography* 16:281–296.
- Brill, Eric. 1995. Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics* 21:543–565.
- Briscoe, Ted, John Carroll, and Rebecca Watson. 2006. The second release of the RASP system. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, 77–80, Sydney, Australia. Association for Computational Linguistics.
- Collins, Michael. 1999. *Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania.
- Fellbaum, Christiane (Ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Giozzo, Alfio, and Carlo Strapparava. 2006. Exploiting comparable corpora and bilingual dictionaries for cross-language text categorization. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 553–560, Sydney, Australia, July. Association for Computational Linguistics.
- Joachims, Thorsten. 2002. *Learning to Classify Text Using Support Vector Machines*. Kluwer.
- Koike, A., and T. Takagi. 2004. Gene/protein/family name recognition in biomedical literature. In *Proc. Biolink 2004*, 9–16.
- Lindberg, D.A., B.L. Humphreys, and A. T. McCray. 1993. The unified medical language system. *Methods in Inf. Med.* 32:281–291.
- Lv, Yuanhua, Le Sun, Junlin Zhang, Jian-Yun Nie, Wan Chen, and Wei Zhang. 2006. An iterative implicit feedback approach to personalized search. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the*

- Association for Computational Linguistics*, 585–592, Sydney, Australia, July. Association for Computational Linguistics.
- Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19:313–330.
- Miyao, Yusuke, Tomoko Ohta, Katsuya Masuda, Yoshimasa Tsuruoka, Kazuhiro Yoshida, Takashi Ninomiya, and Jun'ichi Tsujii. 2006. Semantic retrieval for the accurate identification of relational concepts in massive textbases. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 1017–1024, Sydney, Australia, July. Association for Computational Linguistics.
- Miyao, Yusuke, and Jun'ichi Tsujii. 2005. Probabilistic disambiguation models for wide-coverage HPSG parsing. In *Proc. 43rd ACL*, 83–90.
- Oepen, Stephan, Kristina Toutanova, Stuart Shieber, Chris Manning, Dan Flickinger, and Thorsten Brants. 2002. The LinGO Redwoods treebank. Motivation and preliminary applications. In *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan.
- Pianta, Emanuele, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, Mysore, India.
- Porter, Martin H. 1980. An algorithm for suffix stripping. *Program* 14:130–137.
- Tateisi, Y., A. Yakushiji, T. Ohta, and J. Tsujii. 2005. Syntax annotation for the GENIA corpus. In *Proc. IJCNLP 2005, Companion volume*, 222–227.
- Tsuruoka, Y., and J. Tsujii. 2004. Improving the performance of dictionary-based approaches in protein name recognition. *Journal of Biomedical Informatics* 37:461–470.