

References

- Callaway, Charles B. (2003). Integrating discourse markers into a pipelined natural language generation architecture. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 264–271.
- Carletta, Jean (1996). Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2):249–254.
- Cohen, Jacob (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.
- Elting, Christian, Gregor Möhler, Stefan Rapp & Michael Strube (2003). Architecture and implementation of multimodal plug and play. In *Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI-03)*, Vancouver, B.C., Canada, 5–7 November, 2003. To appear.
- Fleiss, Joseph L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.
- Kowtko, Jacqueline C., Stephen D. Isard & Gwyneth M. Doherty (1992). *Conversational games within dialogue*. Technical Report HCRC/RP-31: Human Communication Research Centre, University of Edinburgh.
- Krippendorff, Klaus (1980). *Content Analysis: An Introduction into its Methodology*. Beverly Hills, Cal.: Sage Publications.
- Litman, Diane & Julia Hirschberg (1990). Disambiguating cue phrases in text and speech. In *Proceedings of the 13th International Conference on Computational Linguistics*, Helsinki, Finland, 20–25 August 1990, Vol. 2, pp. 251–256.
- Müller, Christoph, Stefan Rapp & Michael Strube (2002). Applying Co-Training to reference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Penn., 7–12 July 2002, pp. 352–359.
- Müller, Christoph & Michael Strube (2001a). Annotating anaphoric and bridging relations with MMAX. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg, Denmark, 1-2 September 2001, pp. 90–95.
- Müller, Christoph & Michael Strube (2001b). MMAX: A tool for the annotation of multi-modal corpora. In *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Seattle, Wash., 5 August 2001, pp. 45–50.
- Müller, Christoph & Michael Strube (2002). An API for discourse-level access to XML-encoded corpora. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Canary Islands, Spain, 29-31 May, 2002, pp. 26–30.
- Müller, Christoph & Michael Strube (2003). Multi-level annotation in MMAX. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan, 5-6 July 2003, pp. 198–207.
- Passonneau, Rebecca (1997). *Applying reliability metrics to co-reference annotation*. Technical Report CUCS-017-97: Columbia University, Department of Computer Science. Available at: <http://arxiv.org/ps/cmp-1g/9706011>.
- Quinlan, J. Ross (1993). *C4.5: Programs for Machine Learning*. San Mateo, Cal.: Morgan Kaufman.
- Rapp, Stefan & Michael Strube (2002). An iterative data collection approach for multimodal dialogue systems. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Canary Islands, Spain, 29-31 May, 2002, pp. 661–665.
- Salmon-Alt, Susanne & Renata Vieira (2002). Nominal expressions in multilingual corpora: Definites and demonstratives. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Canary Islands, Spain, 29-31 May, 2002, pp. 1627–1634.
- Siegel, Sidney & N. John Castellan (1988). *Nonparametric Statistics for the Behavioral Sciences* (2nd ed.). New York: McGraw-Hill.
- Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert & Julia Hirschberg (1992). TOBI: A standard for labeling English prosody. In *Proceedings of the International Conference on Speech and Language Processing (ICSLP-92)*, Vol. 2, pp. 867–870.
- Strube, Michael & Christoph Müller (2003). A machine learning approach to pronoun resolution in spoken dialogue. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 168–175.

Walker, Marilyn A., Diane J. Litman, Candace A. Kamm & Alicia Abella (1997). PARADISE: A framework for evaluating spoken dialogue agents. 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. 271–280.