# Ted Pedersen

Department of Computer Science
University of Minnesota
Duluth, MN 55812–3036
tpederse@d.umn.edu
http://umn.edu/home/tpederse

January 5, 2026

## Academic Appointments

- **UNIVERSITY OF MINNESOTA, DULUTH**
  Department of Computer Science
  Professor (with tenure), August 2009 – present
  Associate Professor (with tenure), August 2003 – August 2009
  Assistant Professor (tenure–track), August 1999 – August 2003

- **MINNESOTA SUPERCOMPUTING INSTITUTE, MINNEAPOLIS**
  Principal Investigator, January 2003 – December 2023

- **CALIFORNIA POLYTECHNIC STATE UNIVERSITY, SAN LUIS OBISPO**
  Department of Computer Science
  Assistant Professor (tenure–track), September 1998 – June 1999

## Education

- **SOUTHERN METHODIST UNIVERSITY, DALLAS, TEXAS**
  Ph.D., Computer Science, May 1998
  Dissertation Title: *Learning Probabilistic Models of Word Sense Disambiguation*
  Adviser: Rebecca Bruce

- **UNIVERSITY OF ARKANSAS, FAYETTEVILLE**
  M.S., Computer Science

- **DRAKE UNIVERSITY, DES MOINES, IOWA**
  B.A., Computer Science

## Research Interests

My areas of expertise are in Natural Language Processing and Computational Linguistics. The long term goal of my research is to develop methods and systems that automatically discover the meaning of written text and understand its content. To that end, I have developed techniques that (1) cluster short text snippets based on the similarity of their content, (2) discover word meanings in large corpora of text, (3) assign meanings from a dictionary to words in running text, and (4) measure the semantic similarity and relatedness between concepts. More recently I have become interested in assessing the social impacts of Large Language Models and Artificial Intelligence in general via audits and other systematic approaches. As of January 2026 my publications have been cited 13,260 times, and my h-index is 44 (according to Google Scholar).

## Honors and Awards

- National Science Foundation Early Faculty CAREER Development Award, 2001–2007.

## External Funding

**[2]** Semantic Relatedness for Active Medication Safety and Outcomes Surveillance (co-PI with Serguei Pakhomov (PI), Department of Pharmaceutical Care & Health Systems, University of Minnesota, Twin Cities). Funded by the National Institutes of Health, National Library of Medicine (R01 LM009623-01A2). $935,000 from September 2008 through September 2012.

**[1]** Automatic Resolution of Semantic Ambiguity in Natural Language. Funded by the National Science Foundation Faculty Early Development (CAREER) program (IIS #0092784). $343,000 from March 2001 through February 2007.

## Book Chapters

**[1]** Pedersen, T. (2006) Unsupervised Corpus Based Methods for WSD. In Agirre, E. & Edmonds, P. (Editors), *Word Sense Disambiguation: Algorithms and Applications.* pp. 133–166. Springer.

## Journal Articles

**[8]** Mennes, J. & Pedersen, T. & Lefever, E. (2019) Approaching Terminological Ambiguity in Cross-Disciplinary Communication as a Word Sense Induction Task. A Pilot Study, *Language Resources and Evaluation*, 53, 889-917, April, Springer. [Journal Citation Reports Impact Factor 2018: 1.029]

**[7]** McInnes, B. & Pedersen, T. (2015) Evaluating Semantic Similarity and Relatedness over the Semantic Grouping of Clinical Term Pairs, *Journal of Biomedical Informatics*, 54, 329-336, April, Elsevier. [Journal Citation Reports Impact Factor 2015: 2.447, ranked 2,806 of 11,997 journals (top 23%)]

**[6]** McInnes, B. & Pedersen, T. (2013) Evaluating Measures of Semantic Similarity and Relatedness to Disambiguate Terms in Biomedical Text, *Journal of Biomedical Informatics*, 46(6), 1116 - 1124, December, Elsevier. [Journal Citation Reports Impact Factor 2013: 2.482, ranked 2,243 of 8,539 journals (top 26%)]

**[5]** Pedersen, T. (2012) Rule-based and Lightly Supervised Methods to Predict Emotions in Suicide Notes, *Biomedical Informatics Insights*, 2012:5 (Suppl. 1), pp. 185 – 193, January, Libertas Academica.

**[4]** Pakhomov, S. & Pedersen, T. & McInnes, B. & Melton, G. & Ruggieri, A. & Chute, C. (2011) Towards a Framework for Developing Semantic Relatedness Reference Standards, *Journal of Biomedical Informatics*, 44(2), pp. 251–265, April, Elsevier. [Journal Citation Reports Impact Factor 2010: 1.724, ranked 3,053 of 8,037 journals (top 38%)]

**[3]** Pedersen, T. (2008) Empiricism is Not a Matter of Faith, *Computational Linguistics*, 34(3), pp. 465–470, September, MIT Press. [Journal Citation Reports Impact Factor 2007: 2.367, ranked 1,427 of 6,417 journals (top 22%)]

**[2]** Kulkarni, A. & Pedersen, T. (2008) Name Discrimination and E-mail Clustering Using Unsupervised Clustering of Similar Concepts, *Journal of Intelligent Systems* (Special Issue : Recent Advances in Knowledge-Based Systems and Their Applications), 17(1-3), pp. 37–50, Freund.

**[1]** Pedersen, T. & Pakhomov, S. & Patwardhan, S. & Chute, C. (2007) Measures of Semantic Similarity and Relatedness in the Biomedical Domain, *Journal of Biomedical Informatics*, 40(3), pp. 288–299, June, Elsevier. [Journal Citation Reports Impact Factor 2006: 2.346, ranked 1,347 of 6,164 journals (top 22%)]

## Refereed Conference Papers

**[40]** Shrestha, P. & Rey-Villamizar, N. & Sadeque, F. & Pedersen, T. & Bethard, S. & Solorio, T. (2016) Age and Gender Prediction on Health Forum Data. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*, pp. 3394-3401, Portoroz, Slovenia. [acceptance rate 60%]

**[39]** Evans, R. & Gelbukh, A. & Grefenstette, G. & Hanks, P. & Jakubicek, M. & McCarthy, D. & Palmer, M. & Pedersen, T. & Rundell, M. & Rychly, P. & Sharoff, S. & Tugwell, D. (2016) Adam Kilgarriff's Legacy to Computational Linguistics and Beyond In *Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics*, Konya, Turkey. [invited paper]

[38] McInnes, B. & Pedersen, T. & Liu, Y. & Melton, G. & Pakhomov, S. (2014) U-path : An undirected path-based measure of semantic similarity. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 882-891, Washington, DC. [acceptance rate 35%]

[37] Fokkens, A. & van Erp, M. & Postma, M. & Pedersen, T. & Vossen, P. & Freire, N. (2013) Offspring from Reproduction Problems: What Replication Failure Teaches Us. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 1691-1701, Sofia, Bulgaria. [acceptance rate 26%, nominated for best paper award]

[36] Liu, Y. & Bill, R. & Fiszman, M. & Rindflesch, T. & Pedersen, T. & Melton, G. & Pakhomov, S. (2012) Using SemRep to Label Semantic Relations Extracted from Clinical Text. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 587 - 595, Chicago, IL. [acceptance rate 44%]

[35] Bill, R. & Liu, Y. & McInnes, B. & Melton, G. & Pedersen, T. & Pakhomov, S. (2012) Evaluating Semantic Relatedness and Similarity Measures with Standardized MedDRA Queries. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 43 - 50, Chicago, IL. [acceptance rate 44%]

[34] Liu, Y. & McInnes, B. & Pedersen, T. & Melton-Meaux, G. & Pakhomov, S. (2012) Semantic Relatedness Study Using Second Order Co-occurrence Vectors Computed from Biomedical Corpora, UMLS and WordNet. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pp. 363 – 371, Miami, FL. [acceptance rate 18%]

[33] McInnes, B. & Pedersen, T. & Liu, Y. & Melton, G. & Pakhomov, S. (2011) Knowledge-based Method for Determining the Meaning of Ambiguous Biomedical Terms Using Information Content Measures of Similarity. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 895 – 904, Washington, DC. [acceptance rate 51%]

[32] McInnes, B. & Pedersen, T. & Liu, Y. & Pakhomov, S. & Melton, G. (2011) Using Second-order Vectors in a Knowledge-based Method for Acronym Disambiguation. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, pp. 145 – 153, Portland, OR. [acceptance rate 35%]

[31] Pakhomov, S. & McInnes, B. & Adam, T. & Liu, Y. & Pedersen, T. & Melton, G. (2010) Semantic Similarity and Relatedness between Clinical Terms : An Experimental Study. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 572 – 576, Washington, DC. [acceptance rate 50%]

[30] Pedersen, T. (2010) The Effect of Different Context Representations on Word Sense Discrimination in Biomedical Texts. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pp. 56 – 65, Arlington, VA. [acceptance rate 17%]

[29] Pedersen, T. (2010) Information Content Measures of Semantic Similarity Perform Better Without Sense-Tagged Text. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 329–332, Los Angeles, CA. [acceptance rate 35%]

[28] McInnes, B. & Pedersen, T. & Pakhomov, S. (2009). UMLS-Interface and UMLS-Similarity : Open Source Software for Measuring Paths and Semantic Similarity. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 431–435, San Francisco, CA. [acceptance rate 49%]

[27] Pedersen, T. (2009) Improved Unsupervised Name Discrimination with Very Wide Bigrams and Automatic Cluster Stopping. In *Proceedings of the Tenth International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 294–305, Mexico City. Springer–Verlag. [acceptance rate 26%]

[26] McInnes, B. & Pedersen, T. & Carlis, J. (2007). Using UMLS Concept Unique Identifiers (CUIs) for Word Sense Disambiguation in the Biomedical Domain. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 533–537, Chicago, IL. [acceptance rate 45%]

[25] Pedersen, T. & Kulkarni, A. (2007) Unsupervised Discrimination of Person Names in Web Contexts. In *Proceedings of the Eighth International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 299–310, Mexico City. Springer–Verlag. [acceptance rate 29%]

[24] Joshi, M. & Pakhomov, S. & Pedersen, T. & Chute, C. (2006) A Comparative Study of Supervised Learning as Applied to Acronym Expansion in Clinical Reports. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 399–403, Washington, DC. [acceptance rate 41%]

[23] Pedersen, T. & Kulkarni, A. (2006) Selecting the "Right" Number of Senses Based on Clustering Criterion Functions. In *Proceedings of the Posters and Demo Program of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics*, pp. 111–114, Trento, Italy. [acceptance rate 40%]

[22] Pedersen, T. & Kulkarni, A. & Angheluta, R. & Kozareva, Z. & Solorio, T. (2006) An Unsupervised Language Independent Method of Name Discrimination Using Second Order Co-occurrence Features. In *Proceedings of the Seventh International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 208–222, Mexico City. Springer–Verlag. [acceptance rate 30%]

[21] Joshi, M. & Pedersen, T. & Maclin, R. (2005) A Comparative Study of Support Vectors Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain. In *Proceedings of the Second Indian International Conference on Artificial Intelligence*, pp. 3449–3468, Pune, India. [acceptance rate 35%]

[20] Kulkarni, A. & Pedersen, T. (2005). Name Discrimination and Email Clustering using Unsupervised Clustering and Labeling of Similar Contexts. In *Proceedings of the Second Indian International Conference on Artificial Intelligence*, pp. 703–722, Pune, India. [acceptance rate 35%]

[19] Pakhomov, S. & Pedersen, T. & Chute, C. (2005). Abbreviation and Acronym Disambiguation in Clinical Discourse. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, pp. 589–593, Washington, DC. [acceptance rate 37%]

[18] Pedersen, T. & Purandare, A. & Kulkarni, A. (2005) Name Discrimination by Clustering Similar Contexts. In *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 220–231, Mexico City. Springer–Verlag. [acceptance rate 35%]

[17] Purandare, A. & Pedersen, T. (2004) Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces. In *Proceedings of the Conference on Computational Natural Language Learning*, pp. 41–48, Boston, MA. [acceptance rate 48%]

[16] Mohammad, S. & Pedersen, T. (2004) Combining Lexical and Syntactic Features for Supervised Word Sense Disambiguation. In *Proceedings of the Conference on Computational Natural Language Learning*, pp. 25–32, Boston, MA. [acceptance rate 48%]

[15] Banerjee, S. & Pedersen, T. (2003) Extended Gloss Overlaps as a Measure of Semantic Relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pp. 805–810, Acapulco. [acceptance rate 21%]

[14] Patwardhan, S. & Banerjee, S. & Pedersen, T. (2003) Using Measures of Semantic Relatedness for Word Sense Disambiguation. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 241–257, Mexico City. Springer–Verlag. [acceptance rate 46%]

[13] Banerjee, S. & Pedersen, T. (2003) The Design, Implementation and Use of the Ngram Statistics Package. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 370–381, Mexico City. Springer–Verlag. [acceptance rate 46%]

[12] Mohammad S. & Pedersen, T. (2003) Guaranteed Pre-Tagging with the Brill Tagger. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 148–157, Mexico City. Springer–Verlag. [acceptance rate 46%]

[11] Pedersen, T. (2002) A Baseline Methodology for Word Sense Disambiguation. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 126–135, Mexico City. Springer–Verlag. [acceptance rate 52%]

[10] Banerjee, S. & Pedersen, T. (2002) An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 136–145, Mexico City. Springer–Verlag. [acceptance rate 52%]

[9] Pedersen, T. (2001) A Decision Tree of Bigrams is an Accurate Predictor of Word Sense. In *Proceedings of the Second Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 79–86, Pittsburgh. [acceptance rate 28%]

**[8]** Pedersen, T. (2001) Lexical Semantic Ambiguity Resolution with Bigram–based Decision Trees. In *Proceedings of the Second International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 157–168, Mexico City. Springer–Verlag. [acceptance rate 57%]

**[7]** Pedersen, T. (2000) A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 63–69, Seattle. [acceptance rate 26%]

**[6]** Pedersen, T. & Bruce, R. (1998) Knowledge Lean Word Sense Disambiguation. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pp. 800–805, Madison, WI. AAAI Press/MIT Press. [acceptance rate 30%]

**[5]** Pedersen, T. & Bruce, R. (1997) Distinguishing Word Senses in Untagged Text. In *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*, pp. 197–207, Providence, RI. [acceptance rate 35%]

**[4]** Pedersen, T. & Bruce, R. (1997) A New Supervised Learning Algorithm for Word Sense Disambiguation. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pp. 604–609, Providence, RI. AAAI Press / MIT Press. [acceptance rate 36%]

**[3]** Pedersen, T. & Bruce, R. & Wiebe, J. (1997) Sequential Model Selection for Word Sense Disambiguation. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, pp. 388–395, Washington, DC. [acceptance rate 32%]

**[2]** Pedersen, T. & Kayaalp, M. & Bruce, R. (1996) Significant Lexical Relationships. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pp. 455–460, Portland, OR. AAAI Press/MIT Press. [acceptance rate 30%]

**[1]** Bruce, R. & Wiebe, J. & Pedersen, T. (1996) The Measure of a Model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 101–112, Philadelphia, PA. [acceptance rate 30%]

## Workshop and Other Publications

Nearly all of these papers were peer–reviewed, but acceptance rates were relatively high (greater than 75%), except when otherwise noted.

**[55]** Syed, S. & Pedersen, T. (2025) DuluthNLP at SemEval-2025 Task 7 : TF-IDF with Optimized Vector Dimensions for Multilingual Fact-Checked Claim Retrieval In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval 2025)*, pp. 712-717, Vienna.

**[54]** Akrah, S. & Pedersen, T. (2023) DuluthNLP at SemEval-2023 Task 12 : AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages using Twitter Dataset In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval 2023)*, pp. 1697-1701, Toronto.

**[53]** Martin, A. & Pedersen, T. & D'Souza, J. (2022) NLPSharedTasks: A Corpus of Shared Task Overview Papers in Natural Language Processing Domains In *Proceedings of the First Workshop on Information Extraction from Scientific Publications (WIESP)*, pp. 105-120, virtual.

**[52]** Akrah, S. & Pedersen, T. (2022) DuluthNLP at SemEval-2022 Task 7: Classifying Plausible Alternatives with Pre-trained ELECTRA In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval 2022)*, pp. 1062-1066, Seattle (virtual).

**[51]** D'Souza, J. & Auer, S. & Pedersen, T. (2021) Task 11 at SemEval-2021: NLPCONTRIBUTIONGRAPH - Structuring Scholarly NLP Contributions for a Research Knowledge Graph In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval 2020)*, pp. 364-376, Bangkok (virtual).

**[50]** Martin, A. & Pedersen, T. (2021) Duluth at SemEval-2021 Task 11: Applying DeBERTa to Contributing Sentence Selection and Dependency Parsing for Entity Extraction, In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval 2020)*, pp. 490-501, Bangkok (virtual).

**[49]** Jin, S. & Yin, Y. & Tang, X. & Pedersen, T. (2020) Duluth at SemEval-2020 Task 7: Asking RoBERTa to Judge the Funniness of News Headlines. In *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval 2020)*, pp. 986-994, Barcelona (virtual).

[48] Pedersen, T. (2020) Duluth at SemEval-2020 Task 12: Offensive Tweet Identification in English with Logistic Regression. In *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval 2020)*, pp. 1938-1946, Barcelona (virtual).

[47] Pedersen, T. (2019) Duluth at SemEval-2019 Task 6: Lexical Approaches to Identify and Categorize Offensive Tweets. In *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval 2019)*, pp. 593-599, Minneapolis, MN.

[46] Sengupta, S. & Pedersen, T. (2019) Duluth at SemEval-2019 Task 4: The Pioquinto Manterola Hyperpartisan News Detector. In *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval 2019)*, pp. 949-953, Minneapolis, MN.

[45] Jin, S. & Pedersen, T. (2018) Duluth UROP at SemEval-2018 Task 2: Multilingual Emoji Prediction with Ensemble Learning and Oversampling. In *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval 2018)*, pp. 482-485, New Orleans, LA.

[44] Wang, Z. & Pedersen, T. (2018) UMDSub at SemEval-2018 Task 2: Multilingual Emoji Prediction Multi-channel Convolutional Neural Network on Subword Embedding. In *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval 2018)*, pp. 395-399, New Orleans, LA.

[43] Swanberg, K. & Mirza, M. & Pedersen, T. and Wang, Z. (2018) ALANIS at SemEval-2018 Task 3: A Feature Engineering Approach to Irony Detection in English Tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval 2018)*, pp. 507-511, New Orleans, LA.

[42] Hassan, A. & Vallabhajosyula, M. & Pedersen, T. (2018) UMDuluth-CS8761 at SemEval-2018 Task 9: Hypernym Discovery using Hearst Patterns, Co-occurrence frequencies and Word Embeddings. In *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval 2018)*, pp. 914-918, New Orleans, LA.

[41] McInnes, B. & Pedersen, T. (2017) Improving Correlation with Human Judgments by Integrating Semantic Similarity with Second–Order Vectors. In *Proceedings of the 16th Workshop on Biomedical Natural Language Processing*, pp. 107-116, Vancouver, BC. [acceptance rate 20%]

[40] Yan, X. & Pedersen, T. (2017) Duluth at SemEval-2017 Task 6: Language Models in Humor Detection. In *Proceedings of the 11th International Workshop on Semantic Evaluation*, pp. 376-380, Vancouver, BC.

[39] Pedersen, T. (2017) Duluth at SemEval-2017 Task 7: Puns Upon a Midnight Dreary, Lexical Semantics for the Weak and Weary. In *Proceedings of the 11th International Workshop on Semantic Evaluation*, pp. 407-411, Vancouver, BC.

[38] Rey-Villamizar, N. & Shrestha, P. & Sadeque, F. & Bethard, S. & Pedersen, T. & Mukherjee, A. & Solorio, T. (2016) Analysis of Anxious Word Usage on Online Health Forums. In *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis*, pp. 37-42, Austin, TX.

[37] Sadeque, F. & Pedersen, T. & Solorio, T. & Shrestha, P. & Rey-Villamizar, N. & Bethard, S. (2016) Why Do They Leave: Modeling Participation in Online Depression Forums. In *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, pp. 14-19, Austin, TX.

[36] Rusert, J. & Pedersen, T. (2016) UMNDuluth at SemEval-2016 Task 14: WordNet's Missing Lemmas. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, pp. 1346-1350, San Diego, CA.

[35] Pedersen, T. (2016) Duluth at SemEval 2016 Task 14: Extending Gloss Overlaps to Enrich Semantic Taxonomies. In *Proceedings of the 10th International Workshop on Semantic Evaluation*,pp. 1328-1331, San Diego, CA.

[34] Rey-Villamizar, N. & Shrestha, P. & Solorio, T. & Sadeque, F. & Bethard, S. & Pedersen, T. (2015) Semi-supervised CLPsych 2016 Shared Task System Submission. In *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality*,pp. 171-175, San Diego, CA.

[33] Sadeque, F. & Solorio, T. & Pedersen, T. & Shrestha, P. & Bethard, S. (2015) Predicting Continued Participation in Online Health Forums. In *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, pp. 12-20, September 2015, Lisbon, Portugal.

[32] Pedersen, T. (2015) Screening Twitter Users for Depression and PTSD using Lexical Decision Lists. In *Proceedings of the 2nd Computational Linguistics and Clinical Psychology Workshop - From Linguistic Signal to Clinical Reality*, pp. 46-53, Denver, CO.

[31] Pedersen, T. (2015) Duluth : Word Sense Discrimination in the Service of Lexicography In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pp. 282 – 286, Denver, CO.

[30] Pedersen, T. (2014) Duluth: Measuring Cross-Level Semantic Similarity with First and Second Order Dictionary Overlaps. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), in conjunction with the 25th International Conference on Computational Linguistics*, pp. 247 – 251, Dublin, Ireland.

[29] Pedersen, T. (2013) Duluth: Word Sense Induction Applied to Web Page Clustering. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), in conjunction with the Second Joint Conference on Lexical and Computational Semantics*, pp. 202 – 206, Atlanta, Georgia.

[28] Pedersen, T. (2012) Duluth: Measuring Degrees of Relational Similarity with the Gloss Vector Measure of Semantic Relatedness. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012), in conjunction with the First Joint Conference on Lexical and Computational Semantics*, pp. 497 – 501, Montreal, Canada.

[27] Pedersen, T. (2011) Identifying Collocations to Measure Compositionality : Shared Task System Description. In *Proceedings of Distributional Semantics and Compositionality (DiSCo 2011), an ACL HLT 2011 Workshop*, pp. 33 – 37, Portland, OR.

[26] Pedersen, T. (2010) Duluth-WSI: SenseClusters Applied to the Sense Induction Task of SemEval-2. In *Proceedings of the SemEval 2010 Workshop : the 5th International Workshop on Semantic Evaluations*, pp. 363–366, Uppsala, Sweden.

[25] Pedersen, T. (2008) Learning High Precision Rules to Make Predictions of Morbidities in Discharge Summaries. In *Working Notes of the Second i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, Washington, DC.

[24] McInnes, B. & Pedersen, T. & Pakhomov, S. (2007) Determining the Syntactic Structure of Medical Terms in Clinical Notes. In *Proceedings of the ACL Workshop BioNLP 2007: Biological, translational, and clinical language processing*, pp. 9–16, Prague, Czech Republic. [acceptance rate 29%]

[23] Patwardhan, S. & Banerjee, S. & Pedersen, T. (2007) UMND1 : Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness. In *Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations*, pp. 390–393, Prague, Czech Republic.

[22] Pedersen, T. (2007) UMND2 : SenseClusters Applied to the Sense Induction Task of Senseval-4. In *Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations*, pp. 394–397, Prague, Czech Republic.

[21] Pedersen, T. & Kulkarni, A. (2007) Discovering Identities in Web Contexts with Unsupervised Clustering. In *Proceedings of the IJCAI Workshop on Analytics for Noisy Unstructured Data*, pp. 23–30, Hyderabad, India.

[20] Pedersen, T. (2006) Determining Smoker Status using Supervised and Unsupervised Learning with Lexical Features. In *Working Notes of the i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, Washington, DC.

[19] Patwardhan, S. & Pedersen, T. (2006) Using WordNet Based Context Vectors to Estimate the Semantic Relatedness of Concepts. In *Proceedings of the EACL Workshop Making Sense of Sense - Bringing Computational Linguistics and Psycholinguistics Together*, pp. 1–8, Trento, Italy.

[18] Pedersen, T. & Kulkarni, A. & Angheluta, R. & Kozareva, Z. & Solorio, T. (2006) Improving Name Discrimination : A Language Salad Approach In *Proceedings of the EACL Workshop on Cross-Language Knowledge Induction*, pp. 25–32, Trento, Italy.

[17] Martin, J. & Mihalcea, R. & Pedersen, T. (2005) Word Alignment for Languages with Scarce Resources. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, pp. 65–74, Ann Arbor, MI.

[16] Purandare, A. & Pedersen, T. (2004) Improving Word Sense Discrimination with Gloss Augmented Feature Vectors. In *Proceedings of the Iberamia Workshop on Lexical Resources for the Web and Word Sense Disambiguation*, pp. 123–130, Puebla, Mexico.

[15] Pedersen, T. (2004) The Duluth Lexical Sample Systems in Senseval-3. In *Proceedings of the ACL Workshop on Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pp. 203–208, Barcelona, Spain.

[14] Mohammad, S. & Pedersen, T. (2004) Complementarity of Lexical and Simple Syntactic Features: The SyntaLex Approach to Senseval-3. In *Proceedings of the ACL Workshop on Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pp. 159–162, Barcelona, Spain.

[13] Chklovski, T. & Mihalcea, R. & Pedersen, T. & Purandare, A. (2004) The Senseval-3 Multilingual English-Hindi Lexical Sample Task. In *Proceedings of the ACL Workshop on Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pp. 5–8, Barcelona, Spain.

[12] Mihalcea, R. & Pedersen, T. (2003) An Evaluation Exercise for Word Alignment. In *Proceedings of the NAACL Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pp. 1–10, Edmonton, Canada.

[11] Thomson-McInnes, B. & Pedersen, T. (2003) The Duluth Word Alignment System. In *Proceedings of the NAACL Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pp. 40–43, Edmonton, Canada.

[10] Pedersen, T. (2002) Assessing System Agreement and Instance Difficulty in the Lexical Sample Tasks of SENSEVAL-2. In *Proceedings of ACL Workshop Word Sense Disambiguation: Recent Successes and Future Directions*, pp. 40–46, Philadelphia, PA.

[9] Pedersen, T. (2002) Evaluating the Effectiveness of Ensembles of Decision Trees in Disambiguating Senseval Lexical Samples. In *Proceedings of ACL Workshop Word Sense Disambiguation: Recent Successes and Future Directions*, pp. 81–87, Philadelphia, PA.

[8] Pedersen, T. (2001) Machine Learning and Lexical Features: The Duluth Approach to SENSEVAL-2. In *Proceedings of SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pp. 139–144, Toulouse, France.

[7] Pedersen, T. (2000) An Ensemble Approach to Corpus Based Word Sense Disambiguation. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics*, pp. 205–218, Mexico City.

[6] Pedersen, T. (1999) Search Techniques for Learning Probabilistic Models of Word Sense Disambiguation. In *Working Notes of the AAAI Spring Symposium on Search Techniques for Problem Solving under Uncertainty and Incomplete Information*, pp. 107–112, Palo Alto, CA.

[5] Pedersen, T. (1999) Integrating Natural Language Subtasks with Bayesian Belief Networks. In *Proceedings of the Pacific Asian Conference on Expert Systems*, pp. 1–6, Los Angeles, CA.

[4] Pedersen, T. (1998) Naive Bayes as a Satisficing Model. In *Working Notes of the AAAI Spring Symposium on Satisficing Models*, pp. 60–67, Palo Alto, CA.

[3] Kaayalp, M. & Pedersen, T. & Bruce, R. (1997) A Statistical Decision Making Method: A Case Study in Prepositional Phrase Attachment. In *Proceedings of the Computational Natural Language Learning Workshop*, pp. 33–42, Madrid, Spain.

[2] Pedersen, T. (1996) Fishing for Exactness. In *Proceedings of the 1996 South–Central Regional SAS Conference*, pp. 188–200, Austin, TX.

[1] Pedersen, T. & Chen, W. (1995) Lexical Acquisition via Constraint Solving. In *Working Notes of the AAAI Spring Symposium on Representation and Acquisition of Lexical Knowledge*, pp. 118–122, Palo Alto, CA.

## System Descriptions

These papers document implemented systems that were demonstrated at major international conferences. The submissions are generally not reviewed for technical content, but are instead selected based on the functionality of the system and its potential appeal to a broadly based audience. However, when an acceptance rate is shown then the system description was fully refereed.

[**13**] McInnes, B. & Liu, Y. & Pedersen, T. & Melton, G. & Pakhomov, S. (2013) UMLS::Similarity: Measuring the Relatedness and Similarity of Biomedical Concepts. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 28–31, Atlanta, Georgia. [acceptance rate 53%]

[**12**] Pedersen, T. & Banerjee, S. & McInnes, B. & Kohli, S. & Joshi, M. & Liu, Y. (2011). The Ngram Statistics Package (Text::NSP) - A Flexible Tool for Identifying Ngrams, Collocations, and Word Associations. In *Proceedings of Multiword Expressions : from Parsing and generation to the Real World (MWE 2011), an ACL HLT 2011 Workshop*, pp 131–122. Portland, OR.

[**11**] Pedersen, T. & Kolhatkar, V. (2009) WordNet::SenseRelate::AllWords - A Broad Coverage Word Sense Tagger that Maximizes Semantic Relatedness. In *Proceedings of the Demonstration Session of the Human Language Technology Conference and the Tenth Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 17–20. Boulder, Colorado.

[**10**] Joshi, M. & Pedersen, T. & Maclin, R. & Chute, C. (2006) An End-to-End Supervised Target-Word Sense Disambiguation System. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pp. 1941–1942. Boston, MA. AAAI Press / MIT Press.

[**9**] Pedersen, T. & Kulkarni, A. (2006) Automatic Cluster Stopping with Criterion Functions and the Gap Statistic. In *Proceedings of the Demonstration Session of the Human Language Technology Conference and the Sixth Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 276–279. New York City.

[**8**] Pedersen, T. and Kulkarni, A. (2005) Identifying Similar Words and Contexts in Natural Language with SenseClusters. In *Proceedings of the Twentieth National Conference on Artificial Intelligence*. pp. 1694–1695. Pittsburgh, PA.

[**7**] Patwardhan, S. & Banerjee, S. & Pedersen, T. (2005) SenseRelate::TargetWord – A Generalized Framework for Word Sense Disambiguation. In *Proceedings of the Twentieth National Conference on Artificial Intelligence*. pp. 1692–1693. Pittsburgh, PA.

[**6**] Kulkarni, A. and Pedersen, T. (2005) SenseClusters: Unsupervised Clustering and Labeling of Similar In *Proceedings of the Demonstration and Interactive Poster Session of the 43rd Annual Meeting of the Association for Computational Linguistics*. pp. 105–108. Ann Arbor, MI. [acceptance rate 55%]

[**5**] Patwardhan, S. & Banerjee, S. & Pedersen, T. (2005) SenseRelate::TargetWord - A Generalized Framework for Word Sense Disambiguation. In *Proceedings of the Demonstration and Interactive Poster Session of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 73–76. Ann Arbor, MI. [acceptance rate 55%]

[**4**] Purandare, A. & Pedersen, T. (2004) SenseClusters - Finding Clusters that Represent Word Senses. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence*. pp. 1030–1031. San Jose, CA. AAAI Press / MIT Press.

[**3**] Pedersen, T. & Patwardhan, S. & Michelizzi, J. (2004) WordNet::Similarity - Measuring the Relatedness of Concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence*. pp. 1024–1025. San Jose, CA. AAAI Press / MIT Press.

[**2**] Purandare, A. & Pedersen, T. (2004) SenseClusters - Finding Clusters that Represent Word Senses. In *Proceedings of the Fifth Annual Meeting of the North American Chapter of the Association for Computational Linguistics*. pp. 26–29. Boston, MA.

[**1**] Pedersen, T. & Patwardhan, S. & Michelizzi, J. (2004) WordNet::Similarity - Measuring the Relatedness of Concepts. In *Proceedings of the Fifth Annual Meeting of the North American Chapter of the Association for Computational Linguistics*. pp. 38–41. Boston, MA.

## Published Abstracts of Poster Presentations

These are short written summaries of work in progress that were the subject of a more comprehensive poster presentation. Nearly all of these abstracts were peer–reviewed, but acceptance rates were relatively high (greater than 75%), except when otherwise noted.

[**10**] Kulkarni, A. & Pedersen, T. (2006) How many different "John Smiths", and who are they? In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pp. 1885–1886. Boston, MA. AAAI Press / MIT Press.

[**9**] Joshi, M. & Pedersen, T. & Maclin, R. & Pakhomov, S. (2006) Kernel Methods for Word Sense Disambiguation and Acronym Expansion. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pp. 1879–1880. Boston, MA. AAAI Press / MIT Press.

[**8**] Kulkarni, A. (2005) Unsupervised Discrimination and Labeling of Ambiguous Names. In *Proceedings of the Student Research Workshop of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 145–150. Ann Arbor, MI. [acceptance rate 28%]

[**7**] Thomson–McInnes, B. & Pakhomov, S. & Pedersen, T. & Chute, C. (2004) Incorporating Bigram Statistics to Spelling Correction Tools. In *MedInfo 2004: Proceedings of the 11th World Congress on Medical Informatics*. p. 1882. San Francisco, CA. IOS Press.

[**6**] Purandare, A. & Pedersen, T. (2004) Discriminating Among Word Meanings by Identifying Similar Contexts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence*, pp. 964–965. San Jose, CA. AAAI Press / MIT Press.

[**5**] Purandare, A. (2003) Discriminating Among Word Senses Using McQuitty's Similarity Analysis. In *Proceedings of the HLT-NAACL 2003 Student Research Workshop*. pp. 19–24. Edmonton, Canada.

[**4**] Pedersen, T. (1998) Raw Corpus Word Sense Disambiguation. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*. p. 1198. Madison, WI. AAAI Press / MIT Press.

[**3**] Pedersen, T. (1998) Dependent Bigram Identification. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*. p. 1197. Madison, WI. AAAI Press / MIT Press.

[**2**] Pedersen, T. (1997) Naive Mixes for Word Sense Disambiguation. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*. p. 841. Providence, RI. AAAI Press / MIT Press.

[**1**] Pedersen, T. (1997) Knowledge Lean Word Sense Disambiguation. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*. p. 814. Providence, RI. AAAI Press / MIT Press. [Doctoral Consortium]

## Book Reviews

[**4**] Pedersen, T. (2004) Polysemy: Theoretical and Computational Approaches, by Yael Ravin and Claudia Leacock. *Minds and Machines*, 14(3), pp. 419–423.

[**3**] Pedersen, T. (2002) Empirical Methods for Exploiting Parallel Texts, by I. Dan Melamed. *Computational Linguistics*, 28(2), pp. 235–237.

[**2**] Pedersen, T. (2000) One Jump Ahead: Challenging Human Supremacy in Checkers, by Jonathan Schaeffer. *Intelligence*, 11(1), pp. 56–57.

[**1**] Pedersen, T. (1999) The Balancing Act: Combining Symbolic and Statistical Approaches to Language, edited by Judith L. Klavans and Philip Resnik. *Intelligence*, 10(1), pp. 41–43.

## Technical Reports

[**7**] Burstein, J. & Shore, J. & Sabatini, J. & Moulder, B. & Holtzman, S. & Pedersen T. (2012) The Language Muse System : Linguistically Focused Instructional Authoring. Educational Testing Service Research Report ETS RR-12-21, October.

[**6**] Pedersen, T. & Burstein, J. (2010) Towards Improving Synonym Options in a Text Modification Application. University of Minnesota Supercomputing Institute Research Report UMSI 2010/165, November.

**[5]** Pedersen, T. (2010) Computational Approaches to Measuring the Similarity of Short Contexts: A Review of Applications and Methods. University of Minnesota Supercomputing Institute Research Report UMSI 2010/118, October.

**[4]** Kulkarni, A. & Pedersen, T. (2006) Unsupervised Context Discrimination and Automatic Cluster Stopping. University of Minnesota Supercomputing Institute Research Report UMSI 2006/90, August.

**[3]** Savova, G. & Pedersen, T. & Purandare, A. & Kulkarni, A. (2005) Resolving Ambiguities in Biomedical Text with Unsupervised Clustering Approaches. University of Minnesota Supercomputing Institute Research Report UMSI 2005/80 and CB Number 2005/21, May.

**[2]** Pedersen, T. & Pakhomov, S. & Patwardhan, S. (2005) Measures of Semantic Similarity and Relatedness in the Medical Domain. University of Minnesota Digital Technology Center Research Report DTC 2005/12, May.

**[1]** Pedersen, T. & Banerjee, S. & Patwardhan, S. (2005) Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. University of Minnesota Supercomputing Institute Research Report UMSI 2005/25, March.

## Editor of Proceedings

**[5]** Jurgens, D. & Kolhatkar, V. & Li, L. & Mieskes, M. & Pedersen, T., Editors (2021). Proceedings of the Fifth Workshop on Teaching NLP. Mexico City (virtual).

**[4]** Solorio, T. & Pedersen, T., Editors (2010) Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas, Los Angeles, CA.

**[3]** Nagata, M. & Pedersen, T., Editors (2005). Proceedings of the ACL Interactive Poster and Demonstration Sessions. Ann Arbor, MI.

**[2]** Koehn, P. & Martin, J. & Mihalcea, R. & Monz, C. & Pedersen, T., Editors (2005), Proceedings of the ACL Workshop on Building and Using Parallel Texts, Ann Arbor, MI.

**[1]** Mihalcea, R. & Pedersen, T., Editors (2003) Proceedings of the HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond, Edmonton, Canada.

## Participation in Shared Tasks and Comparative Evaluations

Shared tasks bring together researchers to evaluate systems on a common set of data over a fixed period of time under very controlled conditions.

**[34]** SemEval-2026 – Participant (with A. Clark and M. Bevers) in Narrative Story Similarity and Narrative Representation Learning.

**[33]** SemEval-2026 – Participant (with S. Rafai Syed) in in Unmasking Political Question Evasions.

**[32]** SemEval-2025 – Participant (with S. Rafai Syed) in Multilingual and Crosslingual Fact-Checked Claim Retrieval.

**[31]** SemEval-2023 – Participant (with S. Akrah) in AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages using Twitter Dataset.

**[30]** SemEval-2022 – Participant (with S. Akrah) in Identifying Plausible Clarifications of Implicit and Underspecified Phrases in Instructional Texts.

**[29]** SemEval-2021 – Participant (with A. Martin) in NLP Contribution Graph : Structuring Scholarly NLP Contributions in the Open Research Knowledge Graph.

**[28]** SemEval-2020 – Participant in OffensEval 2: Multilingual Offensive Language Identification in Social Media.

**[27]** SemEval-2020 – Participant (with S. Jin, Y. Yin, and X. Tang) in the Assessing Humor in Edited News Headlines task.

**[26]** SemEval-2019 – Participant in OffensEval: Identifying and Categorizing Offensive Language in Social Media.

**[25]** SemEval-2019 – Participant (with S. Sengupta) in the Hyperpartisan News Detection task.

[24] SemEval-2018 – Participant (with A. Hassan and M. Vallabhajosyula) in the Hypernym Discovery task.

[23] SemEval-2018 – Participant (with S. Jin and Z. Wang) in the Multilingual Emoji Prediction task.

[22] SemEval-2017 – Participant in the Computational Detection and Interpretation of English Puns task.

[21] SemEval-2017 – Participant (with X. Yan) in the #HashtagWars: Learning a Sense of Humor task.

[20] SemEval-2016 – Participant (with J. Rusert) in the Semantic Taxonomy Enrichment task.

[19] SemEval-2016 – Participant in the Semantic Taxonomy Enrichment task.

[18] CLPsych-2016 – Participant (with N. Rey-Villamizar et al.) in the Computational Linguistics and Clinical Psychology Shared Task.

[17] CLPsych-2015 – Participant in the Computational Linguistics and Clinical Psychology Shared Task.

[16] SemEval-2015 – Participant in the Corpus Pattern Analysis task.

[15] SemEval-2014 – Participant in the Cross Level Semantic Similarity task.

[14] SemEval-2013 – Participant in the Evaluating Word Sense Induction task.

[13] SemEval-2013 – Participant in the Disambiguation within an End-User Application task.

[12] SemEval-2012 – Participant in the Measuring Degrees of Relational Similarity task.

[11] Medical NLP Challenge (2011) – Participant in the Sentiment Classification task. Organized by the Computational Medical Center.

[10] DiSCo (2011) – Participant in the Distributional Semantics to Measure Compositionality task.

[9] SemEval-2 (2010) – Participant in the Word Sense Induction task.

[8] Second Shared Task for Challenges in Natural Language Processing of Clinical Data (2008) – Participant in the i2b2 Obesity Challenge.

[7] SensEval-4 / SemEval-1 (2007) – Participant (with S. Patwardhan & S. Banerjee) in the English Lexical Sample Task

[6] SensEval-4 / SemEval-1 (2007) – Participant in the Sense Induction Task.

[5] Medical NLP Challenge (2007) – Participant (with B. McInnes) in the Classifying Clinical Free Text Using Natural Language Processing task.

[4] First Shared Task for Challenges in Natural Language Processing of Clinical Data (2006) : Participant in the i2b2 Smoker Identification Challenge.

[3] Senseval-3 (2004) – Participant (with S. Mohammad) in the English lexical sample task.

[2] Senseval-3 (2004) – Participant in the English, Romanian, Catalan, Basque, Spanish, and Multi-Lingual lexical sample tasks.

[1] Senseval-2 (2001) – Participant in the English lexical sample task.

## Invited Keynote Addresses

[8] Automatically Identifying Islamophobia in Social Media, Muslims in Machine Learning Workshop at NeurlPS, Vancouver (virtual), December 2020.

[7] Puns Upon a Midnight Dreary, Lexical Semantics for the Weak and Weary, Minnesota Undergraduate Linguistics Symposium, University of Minnesota, Duluth, April 2017.

[6] The Horizon isn't Found in a Dictionary : Identifying Emerging Word Senses and Identities in Online Text, VII Seminario de Ingenieria Linguistica, National University of Mexico (UNAM), Mexico City, September 2016.

[5] Improving Relatedness Measurements of Biomedical Concepts by Embedding Second-Order Vectors with Similarity Measurements, Seminario de Linguistica Forense, National University of Mexico (UNAM), Mexico City, September 2016.

[4] The Horizon isn't Found in a Dictionary : Identifying Emerging Word Senses and Identities in Online Text, Crazy Futures III Workshop, Danube Delta, Romania, August 2015.

[3] The Road from Good Software Engineering to Good Science ... is a 2 way street. Software Engineering, Testing, and Quality Assurance for Natural Language Processing, A Workshop of NAACL-HLT 2009. Boulder, Colorado, June 2009.

[2] The Semantic Quilt: Contexts, Co-occurrences, Kernels, and Ontologies. Fifth International Conference on Natural Language Processing (ICON). Hyderabad, India, January 2007.

[1] Word Sense Disambiguation with Semantic Relations and Corpus–Based Evidence. Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), Mexico City, February 2003.

## Invited Talks

[25] Automatically Identifying Islamophobia in Social Media, University of Massachusetts, Lowell (virtual), Computer Science Department Colloquium, November 2020.

[24] Algorithmic Bias: What is it? Why should we care about it? What can we do about it? York University (Canada), Schulich School of Business, Department of Marketing, February 2019.

[23] Semantic Similarity and Relatedness as Scaffolding for Natural Language Processing, Thomson Reuters, Eagan, MN, June 2018.

[22] Semantic Similarity and Relatedness as Scaffolding for Natural Language Processing, University of North Carolina at Charlotte, Department of Computer Science, April 2018.

[21] Semantic Similarity and Relatedness as Scaffolding for Natural Language Processing, University of Colorado, Department of Information Science, February 2018.

[20] Improving Relatedness Measurements of Biomedical Concepts by Embedding Second-Order Vectors with Similarity Measurements (or: eating your own tail is good for you), University of Toronto, Department of Computer Science, September 2016.

[19] Measuring Semantic Similarity and Relatedness in the Biomedical Domain : Applications and Methods. Department of Computer and Information Science, University of Alabama at Birmingham, April 2013.

[18] Measuring Semantic Similarity and Relatedness in the Medical Domain : Applications and Methods. Mayo Clinic Biomedical Informatics Webinar. February 2012.

[17] The Effect of Different Context Representations on Word Sense Discrimination in Biomedical Texts. Forum for Artificial Intelligence, University of Texas at Austin. October 2010.

[16] The Semantic Quilt: Contexts, Co-occurrences, Kernels, and Ontologies. Department of Computer Science, University of Toronto, Canada. June 2008.

[15] Measuring Similarity of Meaning. Research & Development, Thomson Legal & Regulatory, Thomson – West Publishing. Eagan, MN, June 2005.

[14] Measuring Similarity of Meaning. Search Marketing, Yahoo! Los Angeles, CA, May 2005.

[13] Measures of Semantic Similarity and Relatedness in the Medical Domain. The Association for Intelligent Data Analysis, University of Minnesota, Duluth, MN, April 2005.

[12] Finding Word Meanings by Clustering Similar Contexts. Research and Development Division, Educational Testing Service. Princeton, NJ, January 2005.

[11] Unsupervised Word Sense Discrimination by Clustering Similar Contexts. Department of Computer Science, Northern Illinois University. DeKalb, IL, October, 2004.

[10] Word Sense Disambiguation Using Measures of Semantic Relatedness. Department of Computer Science, University of Toronto, Canada, October 2003.

[9] Word Sense Disambiguation Using Measures of Semantic Relatedness. Department of Computer Science, University of North Texas, Denton, TX, April 2003.

[8] Using Measures of Semantic Relatedness for Word Sense Disambiguation. Medical Informatics Group, The Mayo Clinic, Rochester, MN, December 2002.

[7] Word Sense Disambiguation: A Kitchen-Sink Problem. Graduate Colloquium, Department of Mathematics and Statistics, University of Minnesota, Duluth, MN, November 2002.

[6] Automatic Resolution of Semantic Ambiguity in Natural Language Processing. Department of Computer and Information Science, The Ohio State University, Columbus, OH, April 2002.

[5] Automatic Resolution of Semantic Ambiguity in Natural Language Processing. Department of Computer Science and Engineering, University of Minnesota–Twin Cities, Minneapolis, MN, April 2002.

[4] A Gentle Introduction to the EM Algorithm. Panel Discussion on The Efficacy of the Expectation Maximization Algorithm, 2001 Conference on Empirical Methods in Natural Language Processing, Carnegie Mellon University, Pittsburgh, PA, June 2001.

[3] Word Sense Disambiguation with Bigram Based Decision Trees. Spring 2001 Brown Bag Seminar on Pattern Recognition and Machine Learning for Natural Language Tasks, Center for Cognitive Science, The Ohio State University, Columbus, OH, April 2001.

[2] Knowledge Lean Word Sense Disambiguation. Seminar on Computational Learning and Adaptation, Center for the Study of Language and Information, Stanford University, Palo Alto, CA, February 1998.

[1] Learning Probabilistic Models of Word Sense Disambiguation. 1998 Spring Colloquium Series, Department of Mathematics and Computer Science, Gettysburg College, Gettysburg, PA, February 1998.

## Other Presentations

[17] Let's Talk About Being White! Interactive Workshop presented at the 2025 UMD Summit on Equity, Race, & Ethnicity, University of Minnesota, Duluth, March 26, 2025.

[16] ChatGPT : Implications for the Classroom (with Jill Jenson and Emily Wooster), Workshop hosted by the Academic Writing Center, Duluth, January 26, 2023.

[15] ChatGPT is ...?, A YouTube video presentation exploring use cases of ChatGPT that could be of interest to faculty, staff, and students (71 minutes), January 8, 2023.

[14] Would you trust a self-driving car?, Discussion with high–school and middle–school students during UMD STEM Discovery Days, Duluth (virtual), March 28, 2022.

[13] Annotating Natural Language Processing Shared Task Descriptions, Poster Presentation (with Anna Martin and Jennifer D'Souza) at SciNLP 2021: 2nd Workshop on Natural Language Processing for Scientific Text, Irvine, CA (virtual), October 8, 2021.

[12] Automatic Hate Speech Detection : The Case of Ilhan Omar and Islamophobia, Poster Presentation at The Histories of Artificial Intelligence Summer School, Cambridge, England (virtual), July 11, 2021.

[11] Panelist at the Mexican NLP Summer School, Mexico City (virtual), June 4, 2021.

[10] Ethics in Computer and Technology, Discussion with high–school and middle–school students during UMD STEM Discovery Days, Duluth (virtual), March 30, 2021.

[9] Detecting Islamophobia with Natural Language Processing, Talk at the UMN Twin Cities Text as Data Practice Group, Minneapolis (virtual), May 13, 2020.

[8] What Makes Hate Speech? Interactive Workshop presented at the 2020 Summit on Equity, Race, & Ethnicity: Justice and Equity for All?, University of Minnesota, Duluth, March 4, 2020.

[7] Undergraduate Research Opportunities in Hate Speech Detection, University of Minnesota, Duluth, Computer Science Research Workshop, September, 13, 2019.

[6] Algorithmic Bias: What is it? Why should we care about it? What can we do about it? Interactive Workshop presented at the 2019 Summit on Equity, Race, & Ethnicity: Justice and Equity for All?, University of Minnesota, Duluth, February 28, 2019.

[5] Who's to say what's funny? A computer using Language Models and Deep Learning, That's Who! (with Xinru Yan). Poster presentation at WiNLP 2017, The Workshop on Women and Underrepresented Minorities in Natural Language Processing, Vancouver, BC, July 30, 2017.

[4] #HashtagWars: Learning a Sense of Humor (with Xinru Yan). Poster presentation at MinneWIC 2017, The ACM-W Celebration of Women in Computing in the Upper MidWest, University of Minnesota–Twin Cities, February 17-18, 2017.

[3] Random Walks in WordNet to Measure Lexical Semantic Relatedness, (with Yanbo Xu & Kang James). Poster presentation at MinneWIC 2010, The First Regional Celebration of Women in Computing in the Upper Mid-West, University of Minnesota–Twin Cities, February 12-13, 2010.

[2] A Measure of Semantic Relatedness Using the Concept of Allowable Paths in the Unified Medical Language System (UMLS), (with Mugdha Choudhary). Poster presentation at MinneWIC 2010, The First Regional Celebration of Women in Computing in the Upper MidWest, University of Minnesota–Twin Cities, February 12-13, 2010.

[1] Developing Measures of Semantic Relatedness for the Biomedical Domain (with Serguei Pakhomov). Poster presentation at the Digital Technology Initiatives Forum, Digital Technology Center, University of Minnesota–Twin Cities, February 28, 2005.

## Free Software Distributions

All software developed in our research is made freely available in source code form. The following are selected packages that have been particularly successful in attracting users and in supporting other research that has resulted in scholarly publications.

[14] Free 24/7 Web service on http://maraca.d.umn.edu for multiple packages maintained from 2011 until 2020. Over 1 million requests serviced per year.

[13] WordNet::Extend, version 1.00, August 2017. Automatically insert new word senses into WordNet as synonyms or hyponyms. Source code available from CPAN since September 2016.

[12] UMLS::Interface, version 1.47, January 2016. A Perl API to the Unified Medical Language System (UMLS) of the National Library of Medicine. Source code available from SourceForge and CPAN since January 2009.

[11] Text-Similarity, version 0.13, October 2015. Measure the similarity of strings using a variety of overlapping (edit distance) measures. Source code available from SourceForge and CPAN since 2004.

[10] WordNet::Similarity: version 2.07, October 2015. Measure the relatedness and similarity of concepts based on information from the lexical database WordNet. Source code available from SourceForge and CPAN since April 2003. Free 24/7 web service available 2005 – 2020.

[9] Ngram Statistics Package: version 1.31, October 2015. Identify statistically significant multi-word expressions in large corpora. Source code available from SourceForge and CPAN since October 2003.

[8] SenseClusters: version 1.05, October 2015. Cluster together similar contexts in order to discover word meanings or organize short text snippets by semantic similarity. Source code available from SourceForge since January 2004. Free 24/7 web service available 2005 – 2020.

[7] UMLS::Similarity, version 1.45, June 2015. Measure the relatedness and similarity of medical concepts based on information from the Unified Medical Language System (UMLS). Source code available from SourceForge and CPAN since January 2009. Free 24/7 web service available 2011 – 2020.

**[6]** UMLS::SenseRelate, version 0.29, July 2013. Assign a meaning to every content word in a running text by finding a set of meanings that is most related according to the Unified Medical Language System (UMLS). Source code available from CPAN since January 2011.

**[5]** WebService::UMLSKS::Similarity, version 0.23, December 2011. Measure semantic relatedness of concepts based on information taken from the Unified Medical Language System (UMLS) web services. Source code available on CPAN since December 2010.

**[4]** CuiTools: version 0.29, January 2011. Perform supervised, unsupervised, and semi–supervised disambiguation of ambiguous words in biomedical text using information from the Unified Medical Language System (UMLS). Source code available from SourceForge since May 2007.

**[3]** WordNet::SenseRelate::AllWords: version 0.19, May 2009. Assign a meaning to every content word in a running text by finding a set of meanings that is most related according to WordNet::Similarity. Source code available from SourceForge and CPAN since April 2005. Free 24/7 web service available since 2008.

**[2]** WordNet::SenseRelate::WordToSet: version 0.04, April 2008. Assign a word the meaning that is most related to a given set of words according to WordNet::Similarity. Source code available from SourceForge and CPAN since May 2005.

**[1]** WordNet::SenseRelate::TargetWord: version 0.09, December 2006. Assign the meaning to a word that is most related to its neighbors in a short context according to WordNet::Similarity. Source code available from SourceForge and CPAN since June 2005.

## Free Data Distributions

We have created a number of novel sources of experimental data that we have used in our own publications and made freely available to the research community.

**[5]** WordNet::Similarity pairwise measurements - 21 billion pairs of WordNet concepts measured for similarity using the shortest path, Wu & Palmer, and Resnik measures, released August 2011, March 2012 and April 2012.

**[4]** Kulkarni Name Corpus - 1,375 manually disambiguated occurrences of five ambiguous names on the Web, released July 2006.

**[3]** Enron Email Corpus - 3,000 email messages from the Enron Corp. manually categorized by topic, released March 2006.

**[2]** Senseval-3 English-Hindi Multilingual Data - 13,000 occurrences of selected target words manually tagged with Hindi translations by volunteers solicited from the Web, released in Summer 2003.

**[1]** Newly cleaned and reformatted versions of the existing Senseval-1, Senseval-2, *line*, *hard*, *serve*, and *interest* corpora. This consists of 50,000 occurrences of 112 different target words. Released in Spring 2003.

## Tutorial Presentations

I have prepared and presented a number of specialized short courses at international conferences and summer schools. These tutorials were accepted at these events based on peer–review of a submitted proposal except where indicated.

**[9]** Measuring the Similarity and Relatedness of Concepts. At the 12th Mexican International Conference on Artificial Intelligence. (4 hours, 30 attendees). Mexico City, Mexico. November 2013.

**[8]** Measuring the Similarity and Relatedness of Concepts in the Medical Domain (with Serguei Pakhomov, Bridget McInnes, and Ying Liu). At the 2nd ACM SIGHIT International Health Informatics Symposium. (2 hours, 35 attendees). Miami, FL. January 2012.

**[7]** Language Independent Methods of Clustering Similar Contexts. At the International Joint Conference on Artificial Intelligence. (4 hours, 40 attendees). Hyderabad, India. January, 2007.

**[6]** Language Independent Methods of Clustering Similar Contexts. At the National Conference on Artificial Intelligence. (4 hours, 40 attendees). Boston, MA. July 2006.

**[5]** Language Independent Methods of Clustering Similar Contexts. At the European Conference of the Association for Computational Linguistics. (4 hours, 30 attendees). Trento, Italy. April 2006.

**[4]** Language Independent Methods of Clustering Similar Contexts. At the EuroLAN Summer School - The Multi-lingual Web. (8 hours, 60 attendees). Cluj-Napoca, Romania. July 2005. [Invited Tutorial]

**[3]** Advances in Word Sense Disambiguation (with Rada Mihalcea). At the Twentieth National Conference on Artificial Intelligence. (4 hours, 50 attendees). Pittsburgh, PA. July 2005.

**[2]** Advances in Word Sense Disambiguation (with Rada Mihalcea). At the 43rd Annual Meeting of the Association for Computational Linguistics. (4 hours, 70 attendees) Ann Arbor, MI. June 2005.

**[1]** Advances in Word Sense Disambiguation (with Rada Mihalcea). At the 9th Ibero-American Conference on Artificial Intelligence. (4 hours, 20 attendees) Puebla, Mexico. November 2004.

## Internal Funding to Support Student Research

I have secured a total of $73,100 to provide support for graduate and undergraduate research.

**[7]** Visualizing Relations Between Concepts in WordNet (with Saiyam Kohli). Funded by a VDIL Summer Research Grant. $2,000 during Summer 2005.

**[6]** Advanced Search Tools for Online Resources (with Justin Chase) Funded by the University of Minnesota Undergraduate Research Opportunities Program (UROP) $1,700 during Spring 2005.

**[5]** Developing Measures of Semantic Relatedness for the Biomedical Domain (with Jason Michelizzi and co-PI Serguei Pakhomov, Mayo Clinic). Funded by the Digital Technology Initiative of the Digital Technology Center of the University of Minnesota. $23,000 from September 2004 through May 2005.

**[4]** Biomedical Text Processing (with Jason Michelizzi). Funded by the University of Minnesota Grant–In–Aid of Research, Artistry, and Scholarship. $23,000 from September 2003 through May 2004.

**[3]** Building Resources for Languages with Scarce Resources (with Brian Rassier) Funded by the University of Minnesota Undergraduate Research Opportunities Program (UROP) $1,700 for Spring 2003.

**[2]** Tools and Techniques for Automatic Bilingual Lexicon Construction (with Brian Rassier) Funded by the University of Minnesota Undergraduate Research Opportunities Program (UROP) $1,700 during Summer 2002.

**[1]** Developing Conversational User Interfaces to Embedded and Handheld Computers (with Satanjeev Banerjee). Funded by the University of Minnesota Grant–In–Aid of Research, Artistry, and Scholarship. $20,000 from September 2000 through May 2001.

## Other Internal Funding

I have secured a total of $18,606 in various small grants to support research and teaching.

**[15]** Detecting Islamophobia with Natural Language Processing, Funded by a UMD Chancellor's Small Grant. $4056 during Spring 2021.

**[14]** Travel to the ACL-2016 conference in Berlin. Partially funded by an Office of International Programs Travel Grant ($1500) and a UMD EVCAA Research and Scholarship Grant ($1500). Summer 2016.

**[13]** Support for Apple Laptop to be used for Research Software Testing. Funded by a UMD Chancellor's Small Grant. $1000 during Spring 2012.

**[12]** Support for a Multimedia Laptop to be used for Classroom Instruction. Funded by a UMD Chancellor's Small Grant and Faculty Development Funds. $1150 during Fall 2010.

**[11]** Support for Professional Development at the Minnesota Supercomputing Institute. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2009.

**[10]** Support for Collaborative Relationships with the University of Minnesota, Twin Cities. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2008.

**[9]** Creating High Quality CDs for Distribution of Research Software. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2006.

**[8]** Expanding the Memory of a Research Server. Funded by a UMD Chancellor's Small Grant. $750 during Fall 2005.

**[7]** Creating High Quality CDs for Distribution of Research Software. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2005.

**[6]** Lexical Resources for Undergraduate Natural Language Processing. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2004.

**[5]** A Compute Server for Undergraduate Education (with Carolyn Crouch, Rich Maclin, and Tim Colburn) Funded by a UMD Chancellor's Small Grant. $2250 during Spring 2003.

**[4]** Campus Visit by Graeme Hirst, University of Toronto. Partially funded by a UMD Chancellor's Small Grant. $750 during Spring 2002.

**[3]** A Platform for Experimental Operating Systems Projects. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2001.

**[2]** Travel to the CICLing-2000 conference in Mexico City. Partially funded by an Office of International Programs Travel Grant. $400 during Spring 2000.

**[1]** Broadening the Experience of Computer Science Majors via Handheld Computing. Funded by a UMD Chancellor's Small Grant. $750 during Spring 2000.

## Resource Grants

**[2]** Core Computing Resources at the Minnesota Supercomputing Institute. Provided 70,000 service hours per year for the project *Automatically Discriminating among Word Senses in High Dimensional Spaces*. January 2016 – December 2023.

**[1]** Core Computing Resources at the Minnesota Supercomputing Institute. Provided undergraduate and graduate students approximately 100,000 hours (per semester) of computing time on distributed supercomputers at MSI for use in courses. Fall 2005, Spring 2006, Fall 2007, Spring 2008, Fall 2008, Spring 2009, Fall 2009, Spring 2010, Spring 2011, Spring 2012, Spring 2013, Spring 2014, Spring 2015, Spring 2021.

## Journal Article Reviewing

I have reviewed 32 full-length journal articles and served on the Editorial Board of *Computational Linguistics*.

**[11]** Computational Linguistics (1999–2001, 2009, 2018, 5 articles)

**[10]** Language Resources and Evaluation (2008–2011, 2018, 5 articles)

**[9]** Journal of Intelligent Information Systems (2018, 2 articles)

**[8]** Journal of Natural Language Engineering, Special Issue on Distributional Lexical Semantics (2010, 1 article)

**[7]** Literary and Linguistic Computing (2008, 1 article)

**[6]** Journal of Information Systems (2008, 1 article)

**[5]** Editorial Board, Computational Linguistics (2002–2005, 12 articles)

**[4]** TAL, Special Issue on Scaling of Natural Language Processing: Complexity, Algorithms, and Architectures (2005, 1 article)

**[3]** Journal of Natural Language Engineering, Special Issue on Parallel Text (2005, 1 article)

**[2]** Journal of Artificial Intelligence Research (2003 and 2010, 2 articles)

**[1]** Machine Learning (1999, 1 article)

## Conference and Workshop Paper Reviewing

I have reviewed for 194 conferences and workshops. Each event typically asks that 3-5 papers be reviewed, which leads to an estimate of somewhere between 582 to 970 conference and workshop papers reviewed.

**[2025] (6)** International Joint Conference on Natural Language Processing and Asia-Pacific Chapter of the Association for Computational Linguistics, 2025 (The Third Workshop for Artificial intelligence for Scientific Publications; Second Conference on Language Modeling (Ethics Review); The 63rd Annual Meeting of the Association for Computational Linguistics (AfricaNLP 2025 Workshop : Multilingual and Multicultural-aware LLMs); The 63rd Annual Meeting of the Association for Computational Linguistics (SemEval-2025 : The 19th International Workshop on Semantic Evaluation); The 31st International Conference on Computational Linguistics (The 18th Workshop on Building and Using Comparable Corpora); The 31st International Conference on Computational Linguistics

**[2024] (6)** The 62nd Annual Meeting of the Association for Computational Linguistics (Sixth Workshop on Teaching NLP); LatinX in Natural Language Processing Workshop at NAACL 2024; LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (Joint Workshop on Multiword Expressions and Universal Dependencies); LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (17th Workshop on Building and Using Comparable Corpora); The Twelfth International Conference on Learning Representations (AfricaNLP 2024 Workshop : Adaptation of Generative AI models for African Languages) The 18th Conference of the European Chapter of the Association for Computational Linguistics (Student Research Workshop)

**[2023] (5)** The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (The 2nd Workshop on Information Extraction from Scientific Publications); The 61st Annual Meeting of the Association for Computational Linguistics (SemEval-2023 : The 17th International Workshop on Semantic Evaluation); The 17th Conference of the European Chapter of the Association for Computational Linguistics (19th Workshop on Multiword Expressions); ACM FAccT Conference; International Conference on Learning Representations (AfricaNLP 2023 Workshop : African NLP in the Era of Large Language Models)

**[2022] (5)** 13th Conference on Language Resources and Evaluation (15th Workshop on Building and Using Comparable Corpora); 13th Conference on Language Resources and Evaluation (18th Workshop on Multiword Expressions); LatinX in Natural Language Processing Workshop at NAACL 2022; The Tenth International Conference on Learning Representations (Third Workshop on African Natural Language Processing : Putting Africa on the NLP Map); 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (SemEval-2022 : The 16th International Workshop on Semantic Evaluation)

**[2021] (5)** The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (17th Workshop on MultiWord Expressions); The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Theme : NLP for Social Good, Ethics Review Committee); The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (SemEval-2021 : The 15th International Workshop on Semantic Evaluation); 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (The Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access)

**[2020] (4)** The 28th International Conference on Computational Linguistics (SemEval-2020 : International Workshop on Semantic Evaluation); The 2020 Conference on Empirical Methods in Natural Language Processing (NLP-CSS) : The 4th Workshop on Natural Language Processing and Computational Social Science); 12th Edition of the Language Resources and Evaluation Conference (The 13th Workshop on Building and Using Comparable Corpora); The 58th Annual Meeting of the Association for Computational Linguistics

**[2019] (1)** The 2019 Conference of the North American Chapter of the Association for Computational Linguistics (SemEval-2019 : International Workshop on Semantic Evaluation)

**[2018] (8)** 16th Ibero-American Conference on Artificial Intelligence; The 27th International Conference on Computational Linguistics; The 2018 Conference of the North American Chapter of the Association for Computational

Linguistics (The 13th Workshop on Innovative Use of NLP for Building Educational Applications); The 2018 Conference of the North American Chapter of the Association for Computational Linguistics (SemEval-2018 : International Workshop on Semantic Evaluation); 11th Edition of the Language Resources and Evaluation Conference (The 11th Workshop on Building and Using Comparable Corpora); 19th International Conference on Intelligent Text Processing and Computational Linguistics; The 32nd AAAI Conference on Artificial Intelligence; 9th Global WordNet Conference 2018

[**2017**] **(10)** The 2017 Conference on Empirical Methods in Natural Language Processing (The 12th Workshop on Innovative Use of NLP for Building Educational Applications); The AMIA Annual Symposium; The 55th Annual Meeting of the Association for Computational Linguistics (The Fourth Computational Linguistics and Clinical Psychology Workshop); The 55th Annual Meeting of the Association for Computational Linguistics (The 10th Workshop on Building and Using Comparable Corpora); The 55th Annual Meeting of the Association for Computational Linguistics (SemEval-2017 : International Workshop on Semantic Evaluation); The 55th Annual Meeting of the Association for Computational Linguistics; Consumer Culture Theory Conference 2017; 18th International Conference on Intelligent Text Processing and Computational Linguistics; The Fifteenth Annual Meeting of the European Chapter for Computational Linguistics - Student Research Workshop; The Thirty-First National Conference on Artificial Intelligence

[**2016**] **(12)** Clinical Natural Language Processing Workshop at COLING 2016; 15th Mexican International Conference on Artificial Intelligence; Fifth Joint Conference on Lexical and Computational Semantics; The International Conference on Computational Processing of Portuguese; The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (Student Research Workshop); The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (The Third Computational Linguistics and Clinical Psychology Workshop); The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (The 11th Workshop on Innovative Use of NLP for Building Educational Applications); The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (SemEval-2016 : International Workshop on Semantic Evaluation); The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (Workshop on Discontinuous Structures in Natural Language Processing); 9th Workshop on Building and Using Comparable Corpora; 17th International Conference on Intelligent Text Processing and Computational Linguistics; Global WordNet Conference 2016

[**2015**] **(12)** 4th Mexican International Conference on Artificial Intelligence; The 10th Brazilian Symposium in Information and Human Language Technology; International Workshop on Biomedical Data Mining, Modeling, and Semantic Integration: A Promising Approach to Solving Unmet Medical Needs; 8th Workshop on Building and Using Comparable Corpora; Workshop on Adaptive NLP at IJCAI 2015; The AMIA Annual Symposium; Consumer Culture Theory Conference 2015; The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (Student Research Workshop); The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (The 10th Workshop on Innovative Use of NLP for Building Educational Applications); SemEval-2015 : International Workshop on Semantic Evaluation; The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies; 16th International Conference on Intelligent Text Processing and Computational Linguistics

[**2014**] **(10)** 13th Mexican International Conference on Artificial Intelligence; 14th Ibero-American Conference on Artificial Intelligence; SemEval-2014 : International Workshop on Semantic Evaluation; The AMIA Annual Symposium; The International Conference on Computational Processing of Portuguese; The 52nd Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (The 9th Workshop on Innovative Use of NLP for Building Educational Applications); 7th Workshop on Building and Using Comparable Corpora - Building Resources for Machine Translation Research; The 52st Annual Meeting of the Association for Computational Linguistics; 15th International Conference on Intelligent Text Processing and Computational Linguistics; Global WordNet Conference 2014

[**2013**] **(10)** 12th Mexican International Conference on Artificial Intelligence; The AMIA Annual Symposium; The 6th International Joint Conference on Natural Language Processing; The 9th Brazilian Symposium in Information and Human Language Technology; SemEval-2013 : International Workshop on Semantic Evaluation; IEEE International Conference on Healthcare Informatics 2013; The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (The 8th Workshop

on Innovative Use of NLP for Building Educational Applications); The 51st Annual Meeting of the Association for Computational Linguistics; The 2013 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies; 14th International Conference on Intelligent Text Processing and Computational Linguistics

[**2012**] **(7)** 11th Mexican International Conference on Artificial Intelligence; SemEval-2012 : International Workshop on Semantic Evaluation; The AMIA Annual Symposium; First Joint Conference on Lexical and Computational Semantics; 13th International Conference on Intelligent Text Processing and Computational Linguistics; Global WordNet Conference 2012; 2nd ACM International Conference on Health Informatics

[**2011**] **(12)** The 5th International Joint Conference on Natural Language Processing; 10th Mexican International Conference on Artificial Intelligence; I2B2 Suicide Notes Challenge; The 8th Brazilian Symposium in Information and Human Language Technology; The AMIA Annual Symposium; EMNLP 2011 Workshop on Unsupervised Learning in NLP; EMNLP 2011 Workshop on Geometrical Models of Natural Language Semantics; Conference on Empirical Methods in Natural Language Processing; ACL-HLT 2011 Workshop on Syntax, Semantics and Structure in Statistical Translation; ACL-HLT 2011 Workshop on Distributional Semantics and Compositionality; ACL-HLT 2011 Workshop on Innovative Use of NLP for Building Educational Applications; The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; 12th International Conference on Intelligent Text Processing and Computational Linguistics

[**2010**] **(12)** 1st ACM International Conference on Health Informatics; 9th Mexican International Conference on Artificial Intelligence; ACL-2010 Workshop on Geometrical Models of Natural Language Semantics; SemEval-2010 Fourth International Workshop on Semantic Evaluation; The Annual Symposium of the American Medical Informatics Association; NAACL-2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas; NAACL-2010 Workshop on Computational Approaches to Linguistic Creativity; NAACL-2010 Student Research Workshop; NAACL-2010 Demonstration Session; NAACL-2010 Long and Short Papers; 11th International Conference on Computational Linguistics and Intelligent Text Processing; The Fifth International Conference of the Global WordNet Association

[**2009**] **(12)** 8th Mexican International Conference on Artificial Intelligence; ACL 2009 Demonstration Program 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 of the Asian Federation of Natural Language Processing; Twenty-first International Joint Conference on Artificial Intelligence; North American Chapter of the Association for Computational Linguistics - Human Language Technologies 2009 Conference; NAACL-2009 Workshop on Semantic Evaluations: Recent Achievements and Future Directions; NAACL-2009 Workshop on Innovative Use of NLP for Building Educational Applications; NAACL-2009 Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics; NAACL-2009 Student Research Workshop at the North American Chapter of the Association for Computational Linguistics - Human Language Technologies 2009 Conference; The Sixth Midwest Computational Linguistics Colloquium 2009; Second Web People Search Evaluation Workshop; 10th International Conference on Computational Linguistics and Intelligent Text Processing

[**2008**] **(7)** 11th Ibero-American Conference on Artificial Intelligence; AAAI-2007 workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy; The 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; The Fifth Midwest Computational Linguistics Colloquium; 9th International Conference on Computational Linguistics and Intelligent Text Processing; IJCNLP 2008 Workshop on Named Entity Recognition for South and South East Asian Languages; The Fourth Global WordNet Conference

[**2007**] **(5)** Doctoral Consortium at EUROLAN 2007 ; SemEval-2007 4th International Workshop on Semantic Evaluations ; Eighth International Conference on Intelligent Text Processing and Computational Linguistics ; 20th International Joint Conference on Artificial Intelligence ; 5th International Conference on Natural Language Processing

[**2006**] **(8)** ACL-COLING 2006 Workshop on Linguistic Distances ; The Twenty-First National Conference on Artificial Intelligence ; Joint Human Language Technology Conference / Annual Meeting of the North American Chapter of the Association for Computational Linguistics ; 11th Conference of the European Chapter of the Association for Computational Linguistics ; EACL 2006 Workshop on Cross-Language Knowledge Induction ; Seventh International Conference on Intelligent Text Processing and Computational Linguistics; Third International WordNet Conference

**[2005] (6)** EuroLan 2005 Workshop on Cross-Language Knowledge Induction; ACL 2005 Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond; 19th International Joint Conference on Artificial Intelligence; 43rd Annual Meeting of the Association for Computational Linguistics; AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors; Sixth Annual Conference on Intelligent Text Processing and Computational Linguistics

**[2004] (8)** The 20th International Conference on Computational Linguistics; 2004 Conference on Empirical Methods in Natural Language Processing; Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text; 42nd Annual Meeting of the Association for Computational Linguistics ; American Association for Artificial Intelligence Doctoral Consortium ; Conference on Computational Natural Language Learning ; The First International Joint Conference on Natural Language Processing ; Fifth Annual Conference on Intelligent Text Processing and Computational Linguistics

**[2003] (5)** MT Summit Workshop: Towards Systematizing MT Evaluation; Annual Meeting of the Association for Computational Linguistics; HLT/NAACL Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond; Human Language Technology Conference; International Conference on Intelligence Text Processing and Computational Linguistics

**[2002] (8)** International Conference on Computational Linguistics; Annual Meeting of the Association for Computational Linguistics; ACL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions; Conference on Empirical Methods in Natural Language Processing; Annual Meeting of the American Association for Artificial Intelligence (student session); International Workshop on Computational Approaches to Collocations; Human Language Technology Conference; International Conference on Intelligence Text Processing and Computational Linguistics

**[2001] (4)** Annual Meeting of the Association for Computational Linguistics; Conference on Empirical Methods in Natural Language Processing; Conference on Human Factors and Computing Systems; International Conference on Intelligent Text Processing and Computational Linguistics

**[2000] (3)** Annual Meeting of the Association for Computational Linguistics; International Conference on Machine Learning; Conference on Human Factors and Computing

**[1999] (2)** Conference on Empirical Methods in Natural Language Processing and Very Large Corpora; Association of Computing Machinery Symposium on Applied Computing

**[1997] (1)** Annual Meeting of the Association for Computational Linguistics, student session

## Other Reviewing

**[6]** Mitacs Accelerate (Canada), Grant Proposal Review (2025)

**[5]** Vrije Universiteit Amsterdam, Dissertation Evaluation (2019)

**[4]** Royal Netherlands Academy of Arts and Sciences External Referee (2017)

**[3]** Royal Netherlands Academy of Arts and Sciences External Referee (2017)

**[2]** National Science Foundation Grant Review Panelist (2002, 2004, 2005, 2010)

**[1]** Emmanuel College Research Fellowship Competition, Cambridge, England (2003)

## Graduate Student Research Supervision

I have supervised one Ph.D. dissertation, 19 M.S. theses and eight M.S. projects. Of my former M.S. students, twelve have gone on to pursue Ph.D. degrees in Natural Language Processing: A. Martin (Minnesota, in progress) S. Sengupta (Penn State, in progress), S. Vallabhajosyula (Ohio State, in progress), Y. Xu (Georgia Tech, in progress), J. Rusert (Iowa, awarded 2022) V. Kolhatkar (Toronto, Ph.D. awarded 2015), A. Kulkarni (Carnegie–Mellon, Ph.D. awarded 2013), M. Joshi (Carnegie–Mellon, awarded 2013), B. McInnes (Minnesota, Ph.D. awarded 2009), S. Mohammad (Toronto, Ph.D. awarded 2008), S. Patwardhan (Utah, Ph.D. awarded 2010), and S. Banerjee (Carnegie–Mellon, Ph.D. awarded 2010).

[28] Shuja Rafai Syed. *Automatic Fact Checking and Evasion Detection in Text*, M.S. Project, Department of Computer Science, expected May 2026.

[27] Jack Quigley. *Is Neural Machine Translation viable for Low Resource Languages? An experimental study of the Irish Language*, M.S. Thesis, Department of Computer Science, July 2025.

[26] Samuel Akrah. *The Impact of the Pre-Training Objective on the Performance of Large Language Models on Downstream Tasks*, M.S. Project, Department of Computer Science, October 2023.

[25] Anna Martin. *Annotating and Automatically Extracting Task Descriptions from Shared Task Overview Papers in Natural Language Processing Domains*, M.S. Thesis, Department of Computer Science, May 2022.

[24] Zakaria Daud. *Implementing an LSTM seq2seq Chatbot from the Cornell Movie Database*, M.S. Project, Department of Computer Science, May 2021.

[23] Saptarshi Sengupta. *Hyperpartisan News Detection using Logistic Regression and Convolutional Neural Networks*. M.S. Project, Department of Computer Science, August 2020.

[22] Madiha Mirza. *Language Models for Interpretation of Puns*. M.S. Project, Department of Computer Science, January 2020.

[21] Xinru Yan. *Language Models for Humor Detection*. M.S. Project, Department of Computer Science, April 2019.

[20] Swathi Vallabhajosyula. *Hypernym Discovery over WordNet and English Corpora using Hearst Patterns and Word Embeddings*, M.S. Thesis, Department of Computer Science, July 2018.

[19] Zhenduo Wang. *Text Embedding Methods on Different Levels for Tweet Authorship Attribution*, M.S. Project, Department of Mathematics and Statistics, May 2018. [co-advised with Y. Li]

[18] Jon Rusert. *Language Evolves, So Should WordNet*. M.S. Thesis, Department of Computer Science, May 2017.

[17] Mugdha Choudhary. *Developing a Path Based Measure of Semantic Relatedness for the Unified Medical Language System*, M.S. Thesis, Department of Computer Science, August 2012.

[16] Yanbo Xu. *Using Random Walks to Measure Semantic Relatedness*, M.S. Project, Department of Mathematics and Statistics, July 2011. [co-advised with Kang James]

[15] Bridget McInnes. *Accurate and Scalable Word Sense Disambiguation in the Biomedical Domain*, Ph.D. Dissertation, University of Minnesota – Twin Cities, September, 2009. [co-advised with John Carlis]

[14] Varada Kolhatkar. *An Extended Analysis of a Method of All Words Sense Disambiguation*, M.S. Thesis, August 2009.

[13] Mahesh Joshi. *Kernel Methods for Word Sense Disambiguation and Abbreviation Expansion in the Medical Domain*, M.S. Thesis, August 2006. [co-advised with Rich Maclin]

[12] Apurva Padhye. *Comparing Supervised and Unsupervised Classification of Messages in the Enron Email Corpus*, M.S. Thesis, August 2006.

[11] Anagha Kulkarni. *Unsupervised Context Discrimination and Automatic Cluster Stopping*, M.S. Thesis, July 2006.

[10] Saiyam Kohli. *Introducing an Object Oriented Design to the Ngram Statistics Package*, M.S. Project, July 2006.

[9] Jason Michelizzi. *Measures of Semantic Similarity and Relatedness and their Applications*, M.S. Thesis, July 2005.

[8] Pratheepan Raveendranathan. *Identifying Sets of Related Words in the World Wide Web*, M.S. Thesis, July 2005.

[7] Bridget McInnes. *Extending the Log-Likelihood Ratio to Identify Collocations in Large Corpora* M.S. Thesis, December 2004.

[6] Amruta Purandare. *Word Sense Discrimination by Clustering Similar Contexts*, M.S. Thesis, August 2004.

**[5]** Saif Mohammad. *Combining Lexical and Syntactic Features for Supervised Word Sense Disambiguation*, M.S. Thesis, August 2003.

**[4]** Siddharth Patwardhan. *Incorporating Dictionary and Corpus Information in a Measure of Semantic Relatedness*, M.S. Thesis, August 2003.

**[3]** Satanjeev Banerjee. *Adapting the Lesk Algorithm for Word Sense Disambiguation to WordNet*, M.S. Thesis, December 2002.

**[2]** Nitin Varma. *Measures of Collocation for Finding Word Alignment in Parallel Corpora*, M.S. Thesis, December 2002.

**[1]** Aditi Paluskar. *User Level Control of Scheduling in a Micro Kernel Operating System*, M.S. Project, November 2001.

## External Examiner of Ph.D. Dissertations

Performed detailed review of written dissertation and conducted in–person examination during defenses of V. Tsang, N. Thabet, D. Inkpen, and T. Wang.

**[9]** Zhenduo Wang. *Modeling Mix-initiative Conversational Search.* University of Utah. January 2024. Advisors : Vivek Srikumar and Qingyao Ai

**[8]** Milton King. *That Ain't How I Speak : Personalizing Natural Language Processing.* University of New Brunswick, Canada. October 2021. Advisor : Paul Cook

**[7]** Tong Wang. *Exploiting Linguistic Knowledge in Linguistics and Semantic Compositional Models.* University of Toronto, Canada. November 2016. Advisor : Graeme Hirst

**[6]** Chee Wee (Ben) Leong. *Modeling Synergistic Relationships Between Words and Images.* University of North Texas. October 2012. Advisor : Rada Mihalcea

**[5]** Oana Magdalena Frunza. *Personalized Medicine through Automatic Extraction of Information from Medical Texts.* University of Ottawa, Canada. March 2012. Advisor : Diana Inkpen

**[4]** Vivian Tsang. *A Graph Approach to Measuring Text Distance.* University of Toronto, Canada. June 2008. Advisor : Suzanne Stevenson

**[3]** Naglaa Thabet. *Classifying the Suras by their Lexical Semantics : an Exploratory Multivariate Analysis Approach to Interpreting the Qur'an.* University of Newcastle, England. August 2006. Advisor : Hermann Moisl

**[2]** Gerard Escudero Bakx. *Machine Learning Techniques for Word Sense Disambiguation.* Universitat Politècnica de Catalunya, Barcelona, Spain. May 2006. Advisors : Lluís Màrquez Villodre & German Rigau Claramunt

**[1]** Diana Zaiu Inkpen. *Building a Lexical Knowledge-Base of Near-Synonym Differences.* University of Toronto, Canada. October 2003. Advisor : Graeme Hirst

## Graduate Student Thesis and Dissertation Committees

Students are listed by the year of graduation.

**[2024]** Lovis Armah (M.S.)

**[2023]** Joseph Hnatek (M.S.)

**[2021]** Shawn Savela (M.S.); Kimberly Mandery (M.S)

**[2016]** Mounika Chillamcherla (M.S.); Yan Bai (M.S.)

**[2010]** Shana Watters (University of Minnesota – Twin Cities, Ph.D.)

**[2006]** Kai Xu (M.S.)

**[2005]** Paul Gordon (M.S.); Sampanna Salunke (M.S.)

**[2003]** Meijia Guo (University of Minnesota – Twin Cities, M.S.); Alexander Kosolapov (M.S.); Sweta Sinha (M.S.)

**[2002]** Devdatta Kulkarni (M.S.); Kristy Vanhornweder (M.S.)

**[2001]** Hariprasad Bommaganti (M.S.); Gan Chen (M.S.)

**[1999]** Wilbur Hsu (Cal Poly, M.S.)

## Undergraduate Student Research Supervision

**[10]** Aidan Clark. Participation inSemeval-2026 Task 4 : Narrative Similarity, Independent Study, University of Minnesota, Duluth, Fall 2025 - Spring 2026.

**[9]** Maxwell Bevers. Participation in Semeval-2026 Task 4 : Narrative Similarity, Independent Study, University of Minnesota, Duluth, Fall 2025 - Spring 2026.

**[8]** Joshua McKelvey. *Using Natural Language Processing (NLP) to Extract Biomarker Data using Medical Labels from FDA Approved Drugs*, Undergraduate Research Assistantship (UROP), University of Minnesota, Duluth, Spring 2019.

**[7]** Shuning Jin. *Sentiment Analysis in Twitter using Deep Learning*, Undergraduate Research Assistantship (UROP + Independent Study), University of Minnesota, Duluth, January 2018 – May 2019.

**[6]** Justin Chase. *Advanced Search Tools for Online Resources*, Undergraduate Research Assistantship (UROP), University of Minnesota, Duluth, Spring 2005.

**[5]** Brian Rassier. *Building Resources for Languages with Scarce Resources*, Undergraduate Research Assistantship (Independent Study), University of Minnesota, Duluth, Spring 2003.

**[4]** Brian Rassier. *A Tool for Manual Alignment of Bilingual Text*, Undergraduate Research Assistantship (Independent Study), University of Minnesota, Duluth, Spring, Fall 2002.

**[3]** Brian Rassier. *Tools and Techniques for Automatic Bilingual Lexicon Construction*, Undergraduate Research Assistantship (UROP), Summer 2002.

**[2]** Ryan Horton. *Service Center Tracking Tool*, Senior Project, Cal Poly, 1998–1999.

**[1]** Richard Chau. *Mandarin Speech Recognition*. Independent Study Project, Cal Poly, Fall 1998.

## Graduate Student Internship Coordination and Supervision

I have assisted in the selection and supervision of graduate student interns at the Mayo Clinic, Division of Biomedical Informatics, Rochester, Minnesota. The objective of these internships was to extend thesis research done under my direction to the medical domain.

**[5]** Anagha Kulkarni. *Exploration of Three Cluster Stopping Rules for the task of Word Sense Discrimination.* Summer 2005, Mayo supervisor: Guergana Savova.

**[4]** Mahesh Joshi. *Supervised Methods for Automatic Acronym Expansion in Medical Text.* Summer 2005, Mayo Clinic supervisor: Serguei Pakhomov.

**[3]** Bridget McInnes. *Determining the Syntactic Structure of Medical Terms in Clinical Notes.* Summer 2004, Mayo Clinic supervisor: Serguei Pakhomov.

**[2]** Siddharth Patwardhan. *Measuring Semantic Relatedness Using a Medical Taxonomy.* Summer 2003, Mayo Clinic supervisor: Serguei Pakhomov.

**[1]** Bridget McInnes. *Incorporating N-gram Statistics in the Normalization of Clinical Notes.* Summer 2003, Mayo Clinic supervisor: Serguei Pakhomov.

## Tutorials, Training, and Classes Attended

**[2025]** Futures Lab Reading Group, April – December 2025, Twin Cities Innovation Alliance, 7 online sessions.

**[2025]** Civics of Technology Reading Group, February – December 2025, Civics of Technology, 8 online sessions.

**[2024]** Civics of Technology Reading Group, April – December 2024, Civics of Technology, 5 online sessions.

**[2024]** Counting Feminicide Reading Group, June – August 2024, Catherine D'Ignazio (author), 8 online sessions.

**[2024]** Racial Covenants: From Awareness to Action, February 2024, Mapping Prejudice, Rebecca Gillette, 1 online session.

**[2023]** Introduction to Somali Language and Culture, July 2023, Cloud Unite, Farhiya Iman, 3 online sessions.

**[2022]** A Cultural History of Robots, June – July 2022, Morbid Anatomy, Jason Lahman, 7 online sessions.

**[2020 – 2021]** Histories of Artificial Intelligence : A Genealogy of Power, May 2020 – July 2021, University of Cambridge (UK), Virtual Reading Group, 3 meetings per month.

**[2020]** Integrating Ethics into the Natural Language Processing Curriculum, Tutorial by Emily Bender, Dirk Hovy, and Alexandra Schofield, July 5, 2020, Virtual Event.

**[2020]** Data Feminism Reading Group, April – June 2020, Virtual Reading Group, 9 weekly meetings led by authors Catherine D'Ignazio and Lauren F. Klein.

**[2020]** Data Statements for Natural Language Processing: Towards Best Practices, May 11–13, 2020, Virtual Event.

**[2016 – 2020]** Conversational Mandarin, Duluth Public Schools Community Education class (weekly)

**[2013]** Spark Fest – The Twin Cities Digital Humanities Symposium, May 14–15, 2013, Minneapolis, MN.

**[2010]** NLM Terminology Resources in Practice, Tutorial by Olivier Bodenreider, James Case, Kin Wah Fung, John Kilbourne, Suresh Srinivasan, and Jan Willis, November 13, 2010, Washington, DC.

**[2009]** Practical Modeling Issues Representing Coded and Structured Patient Data in EHR Systems, Tutorial by Stanley Huff, November 14, 2009, San Francisco, CA.

**[2008]** Knowledge-Based Decision-Support Systems for Implementing Clinical Practice Guidelines, Tutorial by Samson Tu, Mary Goldstein, Mor Peleg, and Susana Martins, November 8, 2008, Washington, DC.

**[2007]** Evaluating Health IT Projects: A Practical Approach, Tutorial by Caitlin M. Cusack and Dan Gaylin, November 10, Chicago, IL; Design and Conduct of Evaluation Studies in Biomedical Informatics, Tutorial by Charles P. Friedman, November 11, Chicago, IL.

**[2006]** Clinical Classifications and Biomedical Ontologies: Terminology Evolution, Principles, and Practicalities, Tutorial by Christopher Chute, James Cimino, and Suzanne Bakken. November 11, Washington, DC; What's in a Name: Current Methods, Applications, and Evaluation in Multilingual Name Search and Matching, Tutorial by Sherri Condon and Keith J. Miller. June 4, New York City

**[2005]** Ontologies in Biomedicine, Tutorial by Mark A. Musen, October 23, Washington, DC ; Unified Medical Language System (UMLS) Overview, Tutorial by Jan Willis. October 22, Washington, DC ; EuroLAN 2005, Summer School on The Multilingual Web : Resources, Technologies, and Prospects. July 25–August 6, Cluj–Napoca, Romania; Empirical Methods in Artificial Intelligence, Tutorial by Paul Cohen, July 10, Pittsburgh, PA; Max-Margin Methods for NLP: Estimation, Structure, and Applications. Tutorial by Dan Klein and Ben Taskar. June 25, Ann Arbor, MI.

**[2004]** Semantic Inference for Question Answering, Tutorial by Sanda Harabagiu and Srini Narayanan. May 2, Boston, MA; Statistical Language Models and Information Retrieval, Tutorial by ChengXiang Zhai. May 2, Boston, MA.

**[2003]** AI Techniques for Personalized Recommendation. Tutorial by Joseph Konstan, John Riedl and Anthony Jameson. August 10, Acapulco, Mexico. A Workshop on Parallel Programming with MPI. Class by Minnesota Supercomputing Institute. July 22, Minneapolis, MN; Optimization, Maxent Models, and Conditional Estimation without Magic. Tutorial by Chris Manning and Dan Klein. May 27, Edmonton, Canada;

**[2001]** Empirical Methods in Natural Language Processing : What's Happened Since the First SIGDAT Meeting? Tutorial by Kenneth Ward Church. June 2, Pittsburgh, PA.

**[2000]** Intermediate VHDL. Class by Aldec Corp. September 26–27, Bloomington, MN; Machine Translation. Tutorial by Kevin Knight. April 29, Seattle, WA.

**[1999]** Getting Started as a Successful Grant Writer and Academician. Workshop by Stephen Russell. October 28–29, Minneapolis, MN.

**[1998]** Robot Building Laboratory. Class by the KISS Institute for Practical Robotics. July 26–27, Madison, WI.

**[1996]** Default Reasoning: Between Logic and Probabilities: Concepts, Models and Algorithms. Tutorial by Hector Geffner. August 5, Portland, OR; Partially Observable Markov Decision Processes. Tutorial by Thomas Dean and Leslie Pack Kaelbling. August 5, Portland, OR; Ontologies Principles, Applications and Opportunities. Tutorial by Michael Gruninger and Mike Uschold. August 4, Portland, OR.

## Conference, Workshop, and Shared Task Organization

**[10]** Co-Organizer (with Jennifer D'Souza and Sören Auer) of SemEval-2021 Task 11 : NLP Contribution Graph, August 2021, Bangkok (virtual).

**[9]** Co-Organizer (with David Jurgens, Varada Kolhatkar, Lucy Li, and Margot Mieskes) of the 5th Workshop on Teaching Natural Language Processing, June 2021, Mexico City (virtual).

**[8]** Publicity Co-Chair (with Anna Rogers) of the 28th International Conference on Computational Linguistics (COLING), December 2020, Barcelona (virtual).

**[7]** Program Chair (with Christy Doran and Thamar Solorio) of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), June 2019, Minneapolis, MN.

**[6]** Co-Organizer (with Thamar Solorio) of the NAACL 2010 Young Investigators in the Americas Workshop, June 2010, Los Angeles, CA.

**[4]** Co-Organizer (with Joel Martin & Rada Mihalcea) of the ACL 2005 Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond, June 2005, Ann Arbor, MI.

**[5]** Task Coordinator (with Joel Martin & Rada Mihalcea) of the Word Alignment shared task in conjunction with the ACL 2005 Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond, June 2005, Ann Arbor, MI.

**[4]** Co-Chair (with Masaaki Nagata) of the Demonstration and Interactive Poster Session at the 43rd Annual Meeting of the Association for Computational Linguistics, June 2005, Ann Arbor, MI.

**[3]** Task Coordinator (with Tim Chklovski, Rada Mihalcea, & Amruta Purandare) of the MultiLingual Lexical Sample task in conjunction with the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Spring 2004.

**[2]** Task Coordinator (with Rada Mihalcea) of the Word Alignment shared task in conjunction with the HLT/NAACL Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond; Human Language Technology Conference, May 2003, Edmonton, Canada.

**[1]** Co-Organizer (with Rada Mihalcea) of HLT/NAACL Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond; Human Language Technology Conference, May 2003, Edmonton, Canada.

## Other Professional Service

- Organizer, Birds of a Feather session on NLP Applications at ACL 2021

- Mentor, ACL 2021 Graduate Student outreach

- Mentor, NAACL 2021 New Faculty and Graduate Student outreach

- Ethics Review Committee, ACL-IJCNLP 2021 conference

- Organizer, Teaching NLP Birds of a Feather Session at COLING 2020

- Ethics Advisory Group, COLING 2020 conference

- Co-organizer, Ethics Birds of a Feather Session at ACL 2020 (with Emily Bender)

- Mentor, ACL 2020 Graduate Student outreach

- Mentor, NAACL 2019 New Faculty and Graduate Student outreach

- Mentor, ACL 2017 Women in NLP workshop

- Mentor, EACL 2017 Student Research workshop

- Mentor, NAACL 2015 Student Research workshop

- Member, Nominating Committee, North American Chapter of the Association for Computational Linguistics (2010–2013)

- Member, Advisory Board, ACM SIGHIT, Special Interest Group on Health Informatics, (2011 – 2012)

- Mentor, ACM IHI Doctoral Consortium, January 2012

- Elected Representative, Executive Board of the North American Chapter of the Association for Computational Linguistics (2009–2010)

- Liaison, Latin American Outreach, North American Chapter of the Association for Computational Linguistics (2001–2008)

- Developer and Maintainer, Registry of Latin American Researchers in Natural Language Processing and Computational Linguistics (2003–2013)

- Secretary (elected), ACL Special Interest Group in Lexical Semantics (SIGLEX) (2004–2006)

- Invited Participant, Semantic Annotation Planning Meeting (Sponsored by DARPA, NSA, NSF, CIA and DOD), College Park, Maryland, April 14-15 2005,

- Mentor, Ninth SIGART/AAAI Doctoral Consortium, 2004

- Member, Senseval Organizing Committee (2001–2004)

## Professional Society Memberships

- Association for Computational Linguistics (since 1994)

- Association for Computer Machinery (since 2020)

- Society for the History of Science and Technology (since 2020)

## University Teaching Experience

While at the University of Minnesota, Duluth I have taught ten different semester length classes a total of 92 times (Fall 1999 – Spring 2026 inclusive). I developed three new undergraduate classes (Algorithms, Race, and Computing, Operating Systems Practicum, and Natural Language Processing), and a new graduate level class (Advanced Natural Language Processing). I significantly revised two undergraduate classes numerous times (Computer Architecture and Computer Ethics) and one graduate level class (Advanced Computer Architecture).

- Computer Architecture (Junior/Senior, 4 semester hours, 21 times)

- Natural Language Processing (Junior/Senior, 4 semester hours, 17 times)

- Operating Systems Practicum (Junior/Senior, 4 semester hours, 16 times)

- Advanced Natural Language Processing (Graduate, 4 semester hours, 15 times)

- Computer Ethics (Junior/Senior, 4 semester hours, 9 times)

- Algorithms, Race, and Computing (First Year, 4 semester hours, 6 times)

- Advanced Computer Architecture (Graduate, 4 semester hours, 4 times)

- Computer Organization (Sophomore, 4 semester hours, 3 times)

- Operating Systems (Junior/Senior, 4 semester hours, 1 time)

- (Cal Poly) Fundamentals of Computer Science II (Freshman, 4 quarter hours, 2 times)

- (Cal Poly) Programming Languages (Junior/Senior, 4 quarter hours, 2 times)

- (Cal Poly) Artificial Intelligence (Graduate, 4 quarter hours, 1 time)

## System-wide Committee Memberships

- UMN Strategic Plan Artificial Intelligence Action Team (Research), November 2025 - December 2025

- UMN Presidential Artificial Intelligence Work Group (Research), January 2025 - October 2025

## University Committee Memberships

- Curriculum Subcommittee of the Faculty Senate, September 2025 - present

- Liberal Education Subcommittee of the Faculty Senate, September 2024 - present

- AI Tools Working Group, January 2024 - May 2024

- Educational Policy Committee, August 2012 - May 2013

## Department and College Committee Memberships

- University Education Association (UEA) representative, Department of Computer Science, September 2017 – present

- Faculty Award Committee, Swenson College of Science and Engineering, Fall 2025

- Executive Committee, Swenson College of Science and Engineering, September 2009 – May 2015, September 2021 – May 2024

- Curriculum Committee, Department of Computer Science, February 2022 – May 2023

- Scholarship Committee, Department of Computer Science, February 2022 – May 2023

- Space Committee, Swenson College of Science and Engineering, September 2015 – May 2016

- Space and Shared Resources Working Group, Swenson College of Science and Engineering, November 2014 – May 2015

- Ad-hoc Committee on BA degree, Department of Computer Science, September 2011 – May 2013

- Committee for Review of Faculty Development Fund Proposals (chair), Swenson College of Science and Engineering, September 2010 – May 2011

- Ad-hoc Committee on Accreditation Standards, Department of Computer Science, Spring 2006

- Single Semester Leave Committee, College of Science and Engineering, Fall 2005

- Visiting Faculty Search Committee, Department of Computer Science, Spring 2004

- Faculty Search Committee, Department of Computer Science, 2000–2001

- Faculty Recruiting Committee, Department of Computer Science, Cal Poly, 1998–1999

- Computer Architecture Working Group, Department of Computer Science, Cal Poly, 1998–1999

## Other University Service

- Study Abroad Advisor, Department of Computer Science, December 2020 – present

- QA Moderator at the Sieur du Lhut Creativity Conference, October 2024.

- Mentor, Tenure–Track Faculty Member in Department of Computer Science, Fall 2018 – May 2020

- Panelist, Celebration of Women in Computing Conference, Department of Computer Science, UMD (virtual), March 2020

- Reviewer, Undergraduate Research Opportunities Proposals, Swenson College of Science and Engineering, Fall 2009 – Spring 2011, Fall 2015 – Spring 2020

- Reviewer, EVCAA Research and Scholarship proposals, Swenson College of Science and Engineering, Spring 2017

- Mentor, Tenure–Track Faculty Member in Department of Computer Science, Fall 2016 – Spring 2017

- Panelist, Talk Back session after UMD Theater production of *The (Curious Case of) Watson Intelligence*, December 2015.

- Twitter coordinator, Department of Computer Science, July 2011 – May 2020

- Reviewer, Undergraduate Research Opportunities Proposals, Swenson College of Science and Engineering, Fall 2009 – Spring 2011

- Faculty Advisor, Student Chapter of the Association for Computing Machinery, Spring 2004 – Spring 2005

- Mentor, Visiting Faculty Member in Department of Computer Science, Fall 2004 – Spring 2005

- Coordinator, Department of Computer Science Colloquia Series (23 invited speakers), Fall 2001 – Spring 2005

- Mentor, Tenure–Track Faculty Member in Department of Computer Science, Fall 2003 – Spring 2004

- Director, Undergraduate Studies in Information Systems and Technology, Department of Computer Science, Fall 1999 – Fall 2000

## Media Appearances and Interviews

- Podcast Guest, The Role of Large Language Models in Amplifying Societal Issues, AI Purity Podcast, April 8, 2024.

- Radio Interview, Islamophobia, Voice of Islam (UK), June 7, 2021.

- Television Interview, Artificial Intelligence used to track COVID-19, FOX-21 Duluth, June 26, 2020.

- Radio Interview, Islamophobia in the United States, Voice of Islam (UK), July 23, 2019.

- Television Interview, Detecting Medication Reactions in Clinical Notes, Health Corner, Fox-21 Duluth, February 5, 2009.

- Newspaper Story, UMD Professor Wins Grant to Study Medication Reactions, Duluth News–Tribune, January 29, 2009.