

SATANJEEV (Bano) BANERJEE

School email: banerjee@cs.cmu.edu

Other email: satanjeev@gmail.com

Website: <http://www.cs.cmu.edu/~banerjee>

School address
5000 Forbes Avenue, 6223 GHC
Pittsburgh, PA-15213

Home address
5562 Hobart Street, #519
Pittsburgh, PA-15217

Objective:

Seeking a research position in the areas of natural language processing, speech and/or dialog systems, starting spring 2010.

Areas of Interest:

Speech understanding, natural language processing and adaptive systems.

Education:	PhD in Language Technologies: Language Technologies Institute, School of Computer Science Carnegie Mellon University Tentative thesis: <i>Using Implicit Human Supervision to Automatically Understand Meetings</i>	Dec 2009 (expected)
	Master of Language Technologies: Language Technologies Institute, School of Computer Science Carnegie Mellon University Result: GPA: 3.96/4.00	Dec 2004
	Master of Science, Computer Science: University of Minnesota, Duluth Result: GPA: 3.964/4.000 Thesis: <i>Adapting the Lesk Algorithm for Word Sense Disambiguation to WordNet.</i>	Dec 2002
	Bachelor of Engineering, Computer Engineering: University of Pune, India Result: First Class with Distinction	May 2000
Computer Skills:	Languages: Proficient in Java, Perl, C, C++, Visual C++. Basic knowledge of Visual Basic. Operating systems: UNIX (Solaris, Linux), Windows 95, 98, 2000, XP.	
Work Experience:	Graduate Research Fellow Working with Dr. Rudnicky on implicit human supervision applied to the task of automatic meeting understanding. Created and deployed SmartNotes.	Aug 2003 – present
	Summer Internship at Microsoft Research, Redmond Did a summer research internship with the Speech Research Group at Microsoft Research, Redmond.	Summer 2007
	Graduate Teaching Assistant Teaching assistant for Data Structures and Algorithms in Computer Science. Held weekly recitations for 30 students, designed homework assignments, graded homework assignments and exams.	Spring 2006
	Graduate Teaching Assistant Teaching assistant for Entrepreneurship in Computer Science. Mentored students and graded homework assignments.	Fall 2006

Graduate Research Fellow Jun 2002
Worked with Dr. Mostow and Dr. Beck on machine learning approaches to – Aug 2003
creating language models from speech recognition errors.

Summer Internship at UMD School of Business Summer
Software Programmer in the School of Business and Economics, University 2001
of Minnesota, Duluth. Created software Neuro-Red in Visual Basic that uses
neural network concepts to assist auditors in detecting fraudulent audits.

Graduate Research Assistant Sep 2000
Worked with Dr. Ted Pedersen on supervised and unsupervised approaches - May 2002
to Word Sense Disambiguation that led to a Masters Thesis. Created and
released Ngram Statistics Package and Sense Tools.

**Selected
Publications:**

All papers are available from my webpage: <http://www.cs.cmu.edu/~banerjee>

On Meeting Understanding:

Banerjee and Rudnicky: **Detecting the Noteworthiness of Utterances in Human Meetings.** In Proceedings of the 2009 Conference of the Special Interest Group on Discourse and Dialogue (SIGDial). Sep 11-12, 2008, London, UK.

Banerjee and Rudnicky: **An Extractive-Summarization Baseline for the Automatic Detection of Noteworthy Utterances in Multi-Party Human-Human Dialog.** In Proceedings of the 2008 IEEE Workshop on Spoken Language Technology. Dec 15-18, 2008, Goa, India.

Banerjee and Rudnicky: **Segmenting Meetings into Agenda Items by Extracting Implicit Supervision from Human Note-Taking.** In Proceedings of the 2007 Conference on Intelligent User Interfaces. Jan 28-31, 2007, Hawaii.

Banerjee and Rudnicky: **TextTiling Based Approach to Topic Boundary Detection in Meetings.** In Proceedings of the 2006 Interspeech (ICSLP) Conference. Sep 17-21, 2006, Pittsburgh, PA.

Banerjee and Rudnicky: **SmartNotes: Implicit Labeling of Meeting Data through User Note-Taking and Browsing.** In Proceedings of the Conference of the North American Association of Computational Linguistics - Human Languages Technology - Demonstration Track. Jun 5-7, 2006, New York, NY.

Banerjee and Rudnicky: **You Are What You Say: Using Meeting Participants' Speech to Detect their Roles and Expertise.** In the NAACL-HLT Workshop on Analyzing Conversations in Text and Speech, Jun 8, 2006, New York, NY.

Banerjee, Rose, and Rudnicky: **The Necessity of a Meeting Recording and Playback System, and the Benefit of Topic-Level Annotations to Meeting Browsing.** In Proceedings of the 10th Conference on Human-Computer Interaction. Sep, 2005, Rome.

Rudnicky, et al: **Intelligently Integrating Information from Speech and Vision Processing to Perform Light-Weight Meeting Understanding.** In the Workshop on Multimodal Multiparty Meeting Processing, Oct 7, 2005, Trento, Italy.

Banerjee and Rudnicky: **Aspects of the Virtuality Continuum and Multi-Participant Interaction Modeling in the Artificial Agent-Assisted Meeting Scenario.** In the workshop on The Virtuality Continuum Revisited, Apr 2-7, 2005, Portland, OR.

Rybski, et al: **Segmentation and Classification of Meetings using Multiple Information Streams.** In Proceedings of the 6th International Conference on Multimodal Interfaces, Oct 14-15, 2004, State College, PA.

Publications Continued...

Banerjee and Rudnicky: **Using Simple Speech-Based Features to Detect the State of a Meeting and the Roles of the Meeting Participants**. In Proceedings of the 8th International Conference on Spoken Language Processing (Interspeech 2004 - ICSLP), Oct 4-8, 2004, Jeju Island, Korea.

Banerjee, et al: **Creating Multi-Modal, User-Centric Records of Meetings with the Carnegie Mellon Meeting Recorder Architecture**. In the ICASSP 2004 Workshop on Meeting Recognition, May 17, 2004, Montreal, Canada.

On Word Sense Disambiguation:

Patwardhan, Banerjee and Pedersen: **SenseRelate::TargetWord - A Generalized Framework for Word Sense Disambiguation**. In the 20th National Conference on Artificial Intelligence - Demonstration track. Jul 2005, Pittsburgh, PA. Also in Proceedings of the Demonstration and Interactive Poster Session of the 43rd Annual Meeting of the Association for Computational Linguistics. Jun 26, 2005, Ann Arbor, MI.

Banerjee and Pedersen: **Extended Gloss Overlaps as a Measure of Semantic Relatedness**. In Proceedings of the 18th International Conference on Artificial Intelligence (IJCAI 03). Aug 9-15, 2003, Acapulco, Mexico.

Patwardhan, Banerjee and Pedersen: **Using Measures of Semantic Relatedness for Word Sense Disambiguation**. In Proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 03). Feb 17-21, 2003, Mexico City, Mexico.

Banerjee and Pedersen: **The Design, Implementation and Use of the Ngram Statistics Package**. In Proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 03). Feb 17-21, 2003, Mexico City, Mexico.

Banerjee and Pedersen: **An Adapted Lesk Approach to Word Sense Disambiguation using WordNet**. In Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 02). Feb 17-21, 2003, Mexico City, Mexico.

On Other Topics:

Banerjee and Lavie: **METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments**. In the Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics, Ann Arbor, MI, Jun 2005.

Harris, Banerjee and Rudnicky: **Heterogeneous Multi-Robot Dialogues for Search Tasks**. In the AAAI Spring Symposium: Dialogical Robots: Verbal Interaction with Embodied Agents and Situated Devices, March 21-23, 2005, Stanford, CA.

Harris, Banerjee, et al: **A Research Platform for Multi-Agent Dialogue Dynamics**. In the 13th International Workshop on Robot and Human Interactive Communication (RO-MAN), September 20-22, 2004, Kurashiki, Japan.

Banerjee, et al: **Improving Language Models by Learning from Speech Recognition Errors in a Reading Tutor that Listens**. In Proceedings of the Second International Conference on Applied Artificial Intelligence, December 15-16, 2003, Kolhapur, India.

Banerjee, Beck and Mostow: **Evaluating the Effect of Predicting Oral Reading Miscues**. In Proceedings of the Eighth European Conference on Speech Communication and Technology (Eurospeech-03), September 1-4, 2003, Geneva, Switzerland.

Tam, et al: **Training a Confidence Measure for a Reading Tutor that Listens**. In Proceedings of the Eighth European Conference on Speech Communication and Technology (Eurospeech-03), September 1-4, 2003, Geneva, Switzerland.

Released Software:

The SmartNotes system. This is a system to help meeting participants take multimodal notes during meetings. As the human takes notes, the system automatically acquires supervision to perform topic segmentation of meetings. Read more in the 2007 IUI paper titled "Segmenting Meetings into Agenda Items...". Webpage of the software: www.cmusmartnotes.org.

Ngram Statistics Package (NSP). I was one of the original authors of this Perl toolkit that is aimed at detecting collocations from a corpus of text using various statistical methods. NSP, released under the GPL, has been incorporated as an educational tool into the graduate course curriculum at the University of California, Berkeley and the University of Uppsala, Sweden. Webpage of the software: <http://www.d.umn.edu/~tpederse/code.html>.

SenseTools. I was one of the original authors of this Perl toolkit that is aimed at taking a supervised approach to Word Sense Disambiguation using the NSP tool above. Webpage of the software: www.d.umn.edu/~tpederse/code.html.

Other Projects:

Conference Roommate Finding Service. Aug 2006
Created a web-based service to help conference attendees find roommates to share their hotel rooms with. Was used at the Interspeech conferences in 2006, and 2007. – Aug 2007

Automatic Evaluation of Machine Translation Systems. Jan 2004
Implemented the METEOR automatic machine translation evaluation system under the guidance of Dr. Alon Lavie. – Jun 2005

Dialog Management System for Robots. Jan – May 2003
Used the Ravenclaw dialog management architecture to create a dialog system that enables spoken natural language based communication with robots to perform human guided robot navigation, and natural language based reporting of sensor data by robots.

Artificial Intelligence Based Othello Player. Jan – Feb 1998
Designed and implemented the game engine to play Othello against a human player. Used an adaptation of the mini-max algorithm with alpha-beta pruning. This program was awarded the first and second prizes at two separate inter-college competitions.