

# 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

## Areas of Interest:

Extracting and structuring semantic information from spoken language; adapting to individual users and user groups by automatically interpreting user behavior as feedback; speech summarization; natural language processing.

<b>Education:</b>	<b>PhD in Language Technologies:</b> Language Technologies Institute, School of Computer Science Carnegie Mellon University Tentative thesis: <i>Using Implicit Human Supervision to Automatically Understand Meetings</i>	May 2010 (expected)
	<b>Master of Language Technologies:</b> Language Technologies Institute, School of Computer Science Carnegie Mellon University Result: GPA: <b>3.96/4.00</b>	Dec 2004
	<b>Master of Science, Computer Science:</b> University of Minnesota, Duluth Result: GPA: <b>3.964/4.000</b> Thesis: <i>Adapting the Lesk Algorithm for Word Sense Disambiguation to WordNet.</i>	Dec 2002
	<b>Bachelor of Engineering, Computer Engineering:</b> University of Pune, India Result: First Class with Distinction	May 2000
<b>Work Experience:</b>	<b>Graduate Research Fellowships/Assistantships:</b>	
	Working with Dr. Rudnicky on PhD thesis: Extracting implicit supervision from normal human interaction with systems through passive observation and active querying. Exploring this topic within the context of assisting meeting participants in taking notes, and learning to do so from their note-taking behavior in previous meetings.	Aug 2003 – present
	Worked with Dr. Lavie on developing METEOR, a new metric for automatic machine translation evaluation.	2004 – 2005
	Worked with Dr. Mostow and Dr. Beck on learning language models from speech recognition errors.	Jun 2002 – Aug 2003
	Worked with Dr. Ted Pedersen on word sense disambiguation. Created and released Ngram Statistics Package and Sense Tools.	Sep 2000 – May 2002
	<b>Summer Internship at Microsoft Research, Redmond:</b> Worked with the Speech Research Group on improving a phone-based dialog system.	Summer 2007
	<b>Graduate Teaching Assistantships:</b>	
	For undergraduate course on Data Structures and Algorithms in Computer	Spring 2006

Science. Held weekly recitations for 30 students, and designed and graded homework programming assignments.

For undergraduate course on Entrepreneurship in Computer Science. Fall 2006  
Mentored student projects.

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

#### **Most Cited Papers:**

Banerjee and Pedersen: **Extended Gloss Overlaps as a Measure of Semantic Relatedness**. In Proceedings of the 18<sup>th</sup> International Conference on Artificial Intelligence (IJCAI 03). Aug 9-15, 2003, Acapulco, Mexico. *[200+ citations]*.

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 43<sup>rd</sup> Annual Meeting of the Association of Computational Linguistics, Ann Arbor, MI, Jun 2005. *[200+ citations]*.

Banerjee and Pedersen: **The Design, Implementation and Use of the Ngram Statistics Package**. In Proceedings of the 4<sup>th</sup> International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 03). Feb 17-21, 2003, Mexico City, Mexico. *[100+ citations]*.

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 10<sup>th</sup> Conference on Human-Computer Interaction. Sep, 2005, Rome. *[30+ citations]*.

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 8<sup>th</sup> International Conference on Spoken Language Processing (Interspeech 2004 - ICSLP). Oct 4-8, 2004, Jeju Island, Korea. *[30+ citations]*.

#### **Other Papers, by Topic:**

##### **On Meeting Understanding:**

Banerjee and Rudnicky: **Extracting Implicit Supervision from Meeting Notes to Train a Noteworthy-Utterance Detection Model**. In submission.

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, 2009, 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.

**Publications  
Continued:**

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.

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 6<sup>th</sup> International Conference on Multimodal Interfaces, Oct 14-15, 2004, State College, PA.

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 20<sup>th</sup> National Conference on Artificial Intelligence - Demonstration track. Jul 2005, Pittsburgh, PA. Also in Proceedings of the Demonstration and Interactive Poster Session of the 43<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics. Jun 26, 2005, Ann Arbor, MI.

Patwardhan, Banerjee and Pedersen: **Using Measures of Semantic Relatedness for Word Sense Disambiguation.** In Proceedings of the 4<sup>th</sup> 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 3<sup>rd</sup> International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 02). Feb 17-21, 2003, Mexico City, Mexico.

**On Other Topics:**

Marge, Banerjee and Rudnicky: **Using the Amazon Mechanical Turk for Transcription of Spoken Language.** To appear in Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Spoken Language (ICASSP). March 14-19, 2010, Dallas, TX.

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.

**Publications Continued:** 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 program helps meeting participants record audio and take notes during meetings, and share and access them afterwards. It is the implementation of my PhD thesis, and has been used for meeting recording and research purposes for other projects at CMU. Webpage: <http://ww.cmusmartnotes.org>

**METEOR System for Automatic Evaluation of Machine Translation.**  
Co-developed the original version of this system with Dr. Alon Lavie for automatic evaluation of machine translation. This has become a popular metric in the translation community. Webpage: <http://www.cs.cmu.edu/~alavie/METEOR>

**Ngram Statistics Package.**  
Co-developed the original version of this Perl toolkit with Dr. Ted Pedersen for detecting collocations from text corpora using various statistical methods. This toolkit is used widely for both research and education. Webpage: <http://ww.d.umn.edu/~tpederse/code.html>.

**SenseTools.**  
Co-developed this Perl toolkit that perform word sense disambiguation using a supervised approach. Webpage of the software: [www.d.umn.edu/~tpederse/code.html](http://www.d.umn.edu/~tpederse/code.html).

**Computer Skills:** Languages: Proficient in Java, Perl, CGI, C. Medium proficiency in JavaScript, C#, C++.  
Toolkits: Proficient in Weka Machine Learning Toolkit, Amazon Mechanical Turk.  
IDEs: Proficient in Eclipse. Medium proficiency in Visual Studio.  
Operating systems worked on: Windows 2000, XP, Vista, 7, and UNIX (Solaris, Linux).

**Professional Activities:** **IEEE SLTC Newsletter.** 2005  
I am a Senior Staff Reporter at the [IEEE Speech and Language Technical Committee's quarterly newsletter](#). I regularly contribute articles on new developments in speech research, and mentor junior reporters. – present

**Served as Paper Reviewer.** 2004  
Reviewed papers for journals such as Transactions on Pattern Analysis and Machine Intelligence (TPAMI), and Knowledge Engineering Review (KER), and for conferences such as Multimodal Interaction (ICMI), Natural Language Processing (ICON) and Hawaii Conference on System Sciences. – present

**Organized Young Researchers' Roundtable on Spoken Dialog System.** 2005  
Co-organized the [first annual Young Researchers' Roundtable on Spoken Dialog System workshop](#). This is now an annually held research workshop.

**Conference Roommate Finding Service.** 2006  
Co-created and ran a web-based service to help conference attendees find roommates with whom to share their hotel rooms. This service was deployed at the Interspeech conferences in 2006, and 2007. – 2007

**Other  
Projects:**

**Artificial Intelligence Based Othello Player.**

1998

Designed and implemented the game engine to play the board game Othello against human opponents. This program was awarded the first and second prizes at two separate inter-college competitions in India.

**References:**

**Alexander I. Rudnicky**

Principal Systems Scientist  
Carnegie Mellon University  
Email: [air@cs.cmu.edu](mailto:air@cs.cmu.edu)  
Web: <http://www.cs.cmu.edu/~air>

**Carolyn Rosé**

Assistant Professor  
Carnegie Mellon University  
Email: [cprose@cs.cmu.edu](mailto:cprose@cs.cmu.edu)  
Web: <http://www.cs.cmu.edu/~cprose>

**Alon Lavie**

Associate Research Professor  
Carnegie Mellon University  
Email: [alavie@cs.cmu.edu](mailto:alavie@cs.cmu.edu)  
Web: <http://www.cs.cmu.edu/~alavie>

**Ted Pedersen**

Professor  
University of Minnesota Duluth  
Email: [tpederse@d.umn.edu](mailto:tpederse@d.umn.edu)  
Web: <http://www.d.umn.edu/~tpederse>