

# Thamar Solorio

---

## Personal Information

Current Job Title: Postdoctoral Research Associate  
Affiliation: University of Texas at Dallas  
Work Address: Department of Computer Science  
Richardson, TX 75083-0688  
Telephone: 469.237.4732  
Homepage: <http://www.hlt.utdallas.edu/~tsolorio>  
E-mail Address: [tsolorio@hlt.utdallas.edu](mailto:tsolorio@hlt.utdallas.edu)  
Immigration Status: Permanent U.S. Resident

## Education

**Ph.D. in Computer Science** September 2005

*Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla, México*

Dissertation title: “*Improvement of Named Entity Tagging by Machine Learning*”

Advisor: Prof. Aurelio López López

**M.S. in Computer Science** August 2002

*Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla, México*

Thesis title: “*Using Unlabeled Data to Improve Classifier Accuracy*”

Advisor: Prof. Olac Fuentes

**B.S. in Computer Systems Engineering** July 2000

*Facultad de Ingeniería, Universidad Autónoma de Chihuahua, Chihuahua, México*

## Professional Experience

September 2007- : Postdoctoral Research Associate, Department of Computer Science, University of Texas at Dallas

September 2005-August 2007: Lecturer, Department of Computer Science, University of Texas at El Paso

January 2006-January 2007: Part-time Research Associate, University of Texas at El Paso

## Research Interests

I am interested in developing statistical methods that take advantage of deep linguistic knowledge for the automated processing of interlanguages. I am also interested in applications of Natural Language Processing and Machine Learning techniques to bioinformatics.

## Fellowships and Awards

2008: NSF Travel Grant, CRA Academic Careers Workshop

2007: NSF Travel Grant, CRA-W Career Mentoring Workshop for Women in Computer Science and FCRC

2005: NSF Travel Grant, Association of Computational Linguistics (ACL) 2005 Student Research Workshop

2002-2005: Fellowship for PhD studies from the National Council for Science and Technology (CONACYT)

2000-2002: Fellowship for MS studies from the National Council for Science and Technology (CONACYT)

## Program Committees

CICLing-2008, Ninth International Conference on Intelligent Text Processing and Computational Linguistics.

NAACL-HLT 2007, The Annual Conference of the North American Chapter of the Association for Computational Linguistics.

CICLing-2007, Eighth International Conference on Intelligent Text Processing and Computational Linguistics.

CICLing-2006, Seventh International Conference on Intelligent Text Processing and Computational Linguistics.

ACL Demos and Interactive Posters, 43rd Annual Meeting of the Association for Computational Linguistics, 2005.

## Other Research Related Activities

While at UTEP I co-directed the Vision, Learning, and Language Lab ( $VL^3$ ). I supervised several undergraduates in the group working on research projects in bioinformatics and NLP-related problems.

## Publications

### Book Chapters

- [1] Thamar Solorio. *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: Documentação e actas do encontro*, chapter 15 MALINCHE: A NER system for Portuguese that reuses knowledge from Spanish. Number ISBN:978-989-20-0731-1. 2007.

### Refereed Journal Articles

- [1] Thamar Solorio, Yang Liu, and Brenda Medina. Part-of-speech tagging English-Spanish code-switched text. *Submitted to Natural Language Engineering*, 2008.
- [2] M. Taufer, M-Y. Leung, A. Licon, D. Mireles, T. Solorio, D. Gomez-Leon, R. Araiza, and K.K. Johnson. RNAVLab: A unified environment for computational RNA structure analysis based on grid computing technology. *Submitted to the Journal of Parallel Computing*, 2007.
- [3] Nigel G. Ward, Rafael Escalante, Yaffa Al Bayyari, and Thamar Solorio. Learning to show you're listening. *Computer Assisted Language Learning*, (20):385–407, 2007.
- [4] Thamar Solorio, Olac Fuentes, Roberto Terlevich, and Elena Terlevich. An active instance-based machine learning method for stellar population studies. *Monthly Notices of the Royal Astronomical Society*, 363(2), October 2005.

### Refereed Conference and Workshop Papers

- [1] Thamar Solorio and Yang Liu. Using language models to identify language impairment in Spanish-English bilingual children. In *Proc. of BioNLP-2008, ACL-HLT 2008*, Columbus, OH, 2008 (to appear). Poster presentation.
- [2] Michela Taufer, Thamar Solorio, Abel Licon, David Mireles, and Ming-Ying Leung. On the effectiveness of rebuilding RNA secondary structures from sequence chunks. In *The Seventh IEEE International Workshop on High-Performance Computational Biology, HiCOMB 2008*, accepted for publication.
- [3] Olac Fuentes, David Vera, and Thamar Solorio. A filter-based approach to detect end-of-utterances from prosody in dialog systems. In *Proceedings of the The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007)*, Rochester, NY, April 2007.
- [4] Juan C. Franco and Thamar Solorio. Baby steps towards a language model for Spanglish. In *The Eighth International Conference on Intelligent Text Processing and Computational Linguistics CICLing-2007*, Lecture Notes in Computer Science, 2007.
- [5] Thamar Solorio, Olac Fuentes, Nigel Ward, and Yaffa Al Bayyari. Prosodic feature generation for back-channel prediction. In *The Ninth International Conference on Spoken Language Processing, INTERSPEECH 2006*, Pittsburgh, Pennsylvania, September 2006.
- [6] Ted Pedersen, Anagha Kulkarni, Roxana Angheluta, Zornista Kozareva, and Thamar Solorio. Improving name discrimination : A language salad approach. In *Proceedings of the EACL 2006 Workshop on Cross-Language Knowledge Induction*, April 2006.
- [7] Ted Pedersen, Anagha Kulkarni, Roxana Angheluta, Zornista Kozareva, and Thamar Solorio. An unsupervised language independent method of name discrimination using second order co-occurrence features. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Seventh International Conference, CICLing 2006*, Lecture Notes in Computer Science, February 2006.

- [8] Thamar Solorio. Exploiting named entity taggers in a second language. In Chris Callison-Burch and Stephen Wan, editors, *Student Research Workshop at the 43<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics, ACL-2005*, pages 25–30, Ann Arbor, Michigan, June 2005.
- [9] Thamar Solorio and Aurelio López López. Learning named entity recognition in Portuguese from Spanish. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Sixth International Conference, CICLing 2005*, volume 3406 of *Lecture Notes in Computer Science*, February 2005.
- [10] Thamar Solorio, Manuel Pérez Coutiño, Manuel Montes y Gómez, Luis Villaseñor Pineda, and Aurelio López López. Question classification in Spanish and Portuguese. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Sixth International Conference, CICLing 2005*, volume 3406 of *Lecture Notes in Computer Science*, February 2005.
- [11] Olac Fuentes, Thamar Solorio, Roberto Terlevich, and Elena Terlevich. Analysis of galactic spectra using active instance-based learning and domain knowledge. In Christian Lemaître, Carlos Reyes, and Jesús A. González, editors, *Advances in Artificial Intelligence – IBERAMIA 2004: The IX Ibero-American Conference on Artificial Intelligence*, Lecture Notes in Artificial Intelligence 3315, pages 215–224, Puebla, Mexico, November 2004.
- [12] Manuel Pérez Coutiño, Thamar Solorio, Manuel Montes y Gómez, Aurelio López López, and Luis Villaseñor Pineda. Question answering for Spanish based on lexical and context annotation. In Christian Lemaître, Carlos Reyes, and Jesús A. González, editors, *Advances in Artificial Intelligence – IBERAMIA 2004*, Lecture Notes in Artificial Intelligence 3315, pages 325–333, Puebla, Mexico, November 2004.
- [13] Manuel Pérez-Coutiño, Thamar Solorio, Manuel Montes y Gómez, Aurelio López López, and Luis Villaseñor Pineda. The use of lexical context in question answering for Spanish. In Carol Peters and Francesca Borri, editors, *Working Notes for the Cross Language Evaluation Forum Workshop, (CLEF-2004)*, Bath, England, September 2004.
- [14] Thamar Solorio, Manuel Pérez Coutiño, Manuel Montes y Gómez, Luis Villaseñor Pineda, and Aurelio López López. A language independent method for question classification. In *The 20th International Conference on Computational Linguistics, COLING-04*, volume II, pages 1374–1380, Geneva, Switzerland, August 2004.
- [15] Manuel Pérez Coutiño, Thamar Solorio, Manuel Montes y Gómez, Aurelio López López, and Luis Villaseñor Pineda. Toward a document model for question answering systems. In Jesus Favela, Ernestina Menasalvas, and Edgar Chávez, editors, *Advances in Web Intelligence: Second International Atlantic Web Intelligent Conference AWIC 2004*, Lecture Notes in Artificial Intelligence 3034, pages 145–154, Cancun, Mexico, May 2004.
- [16] Olac Fuentes and Thamar Solorio. An optimization algorithm based on active and instance-based learning. In R. Monroy, G. Arroyo-Figueroa, L. E. Sucar, and H. Sossa, editors, *MICAI 2004: Advances in Artificial Intelligence, Third Mexican International Conference on Artificial Intelligence*, Lecture Notes in Artificial Intelligence 2972, pages 242–251, Mexico City, Mexico, April 2004.
- [17] Thamar Solorio and Aurelio López López. Learning named entity classifiers using support vector machines. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Fifth International Conference, CICLing 2004*, volume 2945/2004 of *Lecture Notes in Computer Science 2945*, pages 158–167, February 2004.
- [18] Thamar Solorio and Olac Fuentes. Improving classification accuracy of large test sets using the ordered classification algorithm. In F.J. Garijo, J.C. Riquelme, and M. Toro (Eds.), editors, *Advances in Artificial Intelligence - IBERAMIA 2002: 8th Ibero-American Conference on Artificial Intelligence*, Lecture Notes in Computer Science 2527, pages 70–79, Seville, Spain, November 2002.

## Invited Talks

Processing Code-Switched Text. Department of Computer Sciences, The University of Texas at Austin. May 2008.

Processing Code-Switched Text. Department of of Computer and Information Sciences, University of Delaware. February 2008.

## Teaching Experience

Elementary Data Structures and Algorithms. UTEP Fall 2005, Spring 2006, Fall 2006, Spring 2007.

Computer Programming for Science & Engineering. UTEP Fall 2006, Spring 2007.

Artificial Intelligence. UTEP Fall 2006.

Software Engineering I. UTEP Fall 2005, Spring 2006.

Computer Architecture I. UTEP Fall 2005.

Selected Topics in Artificial Intelligence. Instituto Tecnológico de Apizaco, graduate level, 2003.

Natural Language Processing. Instituto Tecnológico de Apizaco, graduate level, 2003.

Object Oriented Programming in Java. Universidad Popular Autónoma del Estado de Puebla 2002.

Teaching Assistant, Automata Theory and Formal Languages. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Aurelio López, 2004.

Teaching Assistant, Machine Learning. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Olac Fuentes, 2004.

Teaching Assistant, Research Seminar II. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Olac Fuentes, 2003.

Teaching Assistant, Automata Theory and Formal Languages. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Ariel Carrasco, 2001.

### **Professional Affiliations**

- Association for Computing Machinery (ACM)
- Association for Computational Linguistics (ACL)

### **Student Advising**

#### **M.S. Students**

1. María Petra Paredes, Sumarización Automática de Documentos, M.S in Computer Science, Instituto Tecnológico de Apizaco, 2004.

#### **Undergraduate Students**

1. Juan Carlos Franco
2. Brenda Medina (CRA-W DMP Student Awardee 2008)
3. David Mireles (Microsoft Scholar, Google Scholar, DSSI-2007 Fellow)
4. David Vera, co-supervised with Olac Fuentes.
5. Nicolle Whitman (LSMAP Scholar)