

# Szu-ting Yi

## EDUCATION

**University of Pennsylvania, School of Engineering and Applied Science, Philadelphia, Pennsylvania**  
**PhD, Computational Linguistics, Department of Computer and Information Science** in August 2007

**Advisor:** Martha Stone Palmer

**Dissertation Topic: ROBUST SEMANTIC ROLE LABELING USING PARSING VARIATIONS AND SEMANTIC CLASSES**

- To correctly identify semantic entities and successfully disambiguate the relations between them and their predicates in order to serve as an important and necessary intermediate step towards successful natural language processing applications, such as text summarization, question answering, and machine translation.

**University of Pennsylvania, School of Engineering and Applied Science, Philadelphia, Pennsylvania**  
**Master of Science in Engineering Degree** in December 1999

**Major:** Computer and Information Science

**Projects:**

- Designed a Web-resume retrieval system to autonomously locate, extract, and categorize on-line resumes.  
- Evaluated word-sense disambiguation approaches. Examined the individual contributions made by syntax, word classes, and statistical context to the task of word-sense disambiguation. Tested the hypothesis that more coarse-grained, hierarchical sense inventories would simplify this task.

**National Taiwan University, College of Engineering, Taipei, Taiwan**

**Bachelor of Science in Engineering Degree** in June 1997

**Major:** Computer Science and Information Engineering

**Project:** JDBC/ODBC bridge

## RESEARCH EXPERIENCE

**University of Colorado, Boulder, Colorado** 2008  
**Research Consultant in Computational Linguistics**

**University of Pennsylvania, Philadelphia, Pennsylvania**  
**Chinese Automatic Content Extraction (CACE)** 2004  
**Chinese Automatic Content Extraction (CACE)** 2003

The objective of the ACE Program is to develop extraction technology to support automatic processing of source language data. This includes classification, filtering, and selection based on the language content of the source data, i.e., based on the meaning conveyed by the data. Thus the ACE program requires the development of technologies that automatically detect and characterize this meaning. The ACE research objectives are viewed as the detection and characterization of Entities, Relations, and Events. The CACE team designed a pipelined system of automatic taggers providing annotations for more well known and tractable problems to support the system to process Chinese text.

**Institute for Research in Cognitive Science, Philadelphia, Pennsylvania** 1999  
**Research Associate**

Applied results from cross-linguistic research to the semantic tagging of English and Chinese newswire texts. Incorporated this tagged text to use as a gold standard (a corpus of correct training examples) for machine-learning purposes.

## TEACHING EXPERIENCE

**University of Pennsylvania, Department of Computer and Information Science, Philadelphia, Pennsylvania**

**Teaching Assistant of CSE391 Artificial Intelligence** Spring 2000

**Teaching Assistant of CSE110 Introduction to Programming** Fall 2000

**Teaching Assistant of CSE391 Artificial Intelligence** Spring 2001

## WORK EXPERIENCE

**Butler Hill Group**, Seattle Washington

**Linguistic Consultant**

2010

Wrote a text normalization specification document for a Text-to-Speech (TTS) system. Collected and annotated a corpus of 10K postal addresses in Taiwanese Mandarin Chinese. Evaluated the TTS system.

**Theatre Institute On Line**, Taipei, Taiwan

**Project Leader**

1998

Consulted with clients on requirements and technologies. Designed and analyzed distance-learning project architecture. Assembled and led a four-member team that performed coding, testing, implementation, and documentation.

**3P Information Ltd.**, Taipei, Taiwan

**System Administrator, Internet Developer**

1997 - 1998

Installed and maintained UNIX platform to create operating environment for Internet services. Established several Web, database, and name servers. Created and implemented CGI applications for company Web site.

**National Taiwan University, Computer and Information Network Center**, Taipei, Taiwan

**Software Engineer**

1996 - 1997

Assisted in computerizing and incorporating university administrative systems. Designed and developed client/server applications in Windows environment.

## PUBLICATIONS/ PRESENTATIONS

Szu-ting Yi. 2010. *Robust Semantic Role Labeling Using Parsing Variation and Semantic Classes*. Published by VDM Verlag Dr. Muller.

Szu-ting Yi. 2007. *Robust Semantic Role Labeling Using Parsing Variation and Semantic Classes*. Ph.D. thesis, University of Pennsylvania.

Szu-ting Yi, Edward Loper, and Martha Palmer. 2007. Can semantic roles generalize across genres? In *Proceedings of NAACL-HLT*.

Edward Loper, Szu-ting Yi, and Martha Palmer. 2007. Empirical evidence for useful semantic role categories. In *Proceedings of the International Workshop on Computational Linguistics*.

Olga Babko-Malaya, Ann Bies, Ann Taylor, Szu-ting Yi, Martha Palmer, Mitch Marcus, Seth Kulick, and Libin Shen. 2006. Issues in synchronizing the English treebank and propbank. In *Frontiers in Linguistically Annotated Corpora, A Merged Workshop with 7th International Workshop on Linguistically Interpreted Corpora (LINC-2006) and Frontiers in Corpus Annotation III, Coling/ACL*.

Szu-ting Yi and Martha Palmer. 2005. The integration of syntactic parsing and semantic role labeling. In *Proceedings of the Conference on Natural Language Learning (CoNLL)*. Ann Arbor, MI.

Szu-ting Yi and Martha Palmer. 2004. Pushing the boundaries of semantic role labeling with svm. In *Proceedings of the International Conference on Natural Language Processing*. Hyderabad, India.

Scott Cotton and Szu-ting Yi. Relation Detection and Characterization in Chinese Newswire and Broadcast News. 2004. In *Proceedings of the International Conference on Natural Language Processing*. Hyderabad, India.

Martha Palmer and Szu-ting Yi. 2000. Computational Lexical Semantics and its Constraints. In *Proceedings of Research on Research on Computational Linguistics Conference XIII (ROCLING XIII)*, Taipei, Taiwan.

The Impact of the Internet on Contemporary Performing Arts. *Third Taiwan-Area Network Conference on Society and Network*, Taipei, Taiwan, October 1997.