Zhiyi Song

Linguistic Data Consortium
3600 Market Street, Suite 810
TEL: 215-573-4108
EMAIL: zhiyi@ldc.upenn.edu

Education

M.A. Linguistics, Dept of Linguistics, School of Arts and Science, University of Pennsylvania (2005)

B.A. English Language & Literature, Dept of English, School of Foreign Languages, Wuhan University (1996)

Training

Graduate Program in Applied Linguistics, School of Foreign Languages, Huazhong University of Science and Technology (2000-2001)

Certificate Program of Essentials of Management
Human Resource, University of Pennsylvania (2009)

Experience

Research Project Manager (2012-present)
Linguistic Data Consortium, University of Pennsylvania

- Managing multiple LDC projects
  - BOLT collection, translation, transliteration and transcription
  - DEFT collection and multilingual ERE annotation
  - LORELEI annotation
  - Chinese named entity dictionaries
  - Big Mechanism annotation
- Diverse technical, linguistic and management responsibilities
  - Continual communication with government sponsors and partners
  - Project specification writing and maintenance
  - Statistical analysis of inter-annotator-agreement
  - Workflow design and management
  - Annotation efficiency analysis
  - Direct and indirect supervision of a staff of over 5 FT and over 35 PT
  - Direct and indirect supervision of 5 vendors
- HLT resource creation experience in diverse languages
  - English, MSA, Egyptian Arabic, Chinese, Spanish, multiple low resource languages

Project Manager (2008-2012)
Linguistic Data Consortium, University of Pennsylvania

- Manage multiple LDC projects
  - BOLT translation and MT post editing
  - GALE translation, MT post editing, distillation
  - MADCAT collection, transcription and translation
  - Simple event annotation
- HLT resource creation experience in diverse languages
  - English, MSA, Iraqi Arabic, Chinese

Research Coordinator (2006-2007)

Linguistic Data Consortium, University of Pennsylvania
- Diverse project responsibilities
  - Automatic Content Extraction (ACE)
  - Reflex Entity cross-doc cross-lingual extraction
- HLT resource creation experience in diverse languages
  - English, MSA, Chinese, Spanish

Lead Annotator (2005-2006)
Linguistic Data Consortium, University of Pennsylvania
- Diverse project responsibilities
  - Automatic Content Extraction (ACE)
  - EARS metadata extraction
  - Language, Variation and Dialect Identification (LVDID)

Research Assistant (2002-2004)
Linguistic Data Consortium, University of Pennsylvania
Project: ACE

Annotator (2003)
Computer and Information Science, University of Pennsylvania
Project: Chinese Proposition Bank

Teaching Assistant (2003)
LING-102: *Introduction to Sociolinguistics* (Prof. William Labov)
School of Arts and Sciences, University of Pennsylvania

Teaching Assistant (2003)
LING-115: *Writing System* (Prof. Gene Buckley)
School of Arts and Sciences, University of Pennsylvania

Lecturer of English Language (1996-2001)
Foreign Language Department
Huazhong University of Science and Technology

Language Proficiency

Chinese: Native
English: Near-native
German: Elementary proficiency
Arabic: Elementary proficiency

Computer Proficiency

OS: Windows, Linux
Software: Microsoft Office suite, Unix shell script
Computer languages: Basic knowledge of AWK and sed

Publications

Zhiyi Song, Ann Bies, Justin Mott, Xuansong Li, Stephanie Strassel, Christopher Caruso. Cross-Document, Cross-Laguage Event Coreference Annotation Using Event Hoppers. Proc. LREC 2018

Justin Mott, Ann Bies, **Zhiyi Song**, Stephanie Strassel. *Parallel Chinese-English Entities, Relations and Events Corpora.* Proc. LREC, 2016

Ann Bies, **Zhiyi Song**, Jeremy Getman, Joe Ellis, Justin Mott, Stephanie Strassel, Martha Palmer, Teruko Mitamura, Marjorie Freedman, Heng Ji, Tim O'Gorman. *A Comparison of Event Representations in DEFT.* Proc. NAACL HLT 2016.

Zhiyi Song, Ann Bies , Stephanie Strassel , Joe Ellis , Teruko Mitamura , Hoa Dang , Yukari Yamakawa , Sue Holm. *Event Nugget and Event Coreference Annotation*. Proc. NAACL HLT 2016.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant and Xiaoyi Ma. *From Light to Rich ERE: Annotation of Entities, Relations, and Events*. Proc. NAACL HLT 2015.

Teruko Mitamura, Yukari Yamakawa, Susan Holm, **Zhiyi Song**, Ann Bies, Seth Kulick, Stephanie Strassel. *Event Nugget Annotation: Processes and Issues*. Proc. NAACL HLT 2015.

Mariona Taulé, M Antonia Martí, Ann Bies, Aina Garí, Montserrat Nofre, **Zhiyi Song**, Stephanie Strassel and Joe Ellis. *Spanish Treebank Annotation of Informal Non-Standard Web Text*. NLPIT 2015. 1st International Workshop on Natural Language Processing for Informal Text, Rotterdam, Netherlands, June 23

Bies. A., **Song Z.,** M. Maamouri, S. Grimes, H. Lee, J. Wright, N. Habash, R. Skander, O. Rambow. *Transliteration of Arabizi into Arabic Orthography: Developing a Parallel Annotated Arabizi-Arabic Script SMS/Chat Corpus*. Proc. EMNLP, 2014.

Aguilar J, C. Beller, P. McNamee, B. Van Durme, S. Strassel, **Z. Song** and J. Ellis. *A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards*. Proc. ACL, 2014.

Song Z., S. Strassel, H. Lee, K. Walker, J. Wright, J. Garland, D. Fore, B. Gainor, P. Cabe, T. Thomas, B. Callahan, A. Sawyer. 2014. *Collecting Natural SMS and Chat Conversations in Multiple Languages: The BOLT Phase 2 Corpus*. Proc. LREC, 2014.

Song Z., S. Ismael, S. Grimes, D. Doermann, S. Strassel. 2012. *Linguistic Resources for Handwriting Recognition and Translation Evaluation*. Proc. LREC, 2012.

Song, Z., S. Strassel, G. Krug and K. Maeda. 2010. *Enhanced Infrastructure for Creation and Collection of Translation Resources*. Proc. LREC, 2010.

Strassel S., J. Kolár, **Z. Song**, L. Barclay., M. Glenn. 2005. *Structural Metadata Annotation: Moving Beyond English*. *INTERSPEECH-2005*, 1545-1548

Song, Z. 2004. *A Study of Chinese EFL learners' Response to Indirect Request*. 33rd annual New Ways of Analyzing Variation (NWAV) Meeting

Huang, S., A. Mitchell, S. Strassel, **Z. Song**. 2004. *Shared Resources for Multilingual Information Extraction and Challenges in Named Entity Annotation*. The First International Joint Conference on Natural Language Processing (IJCNL-04)

Song, Z. 2003. *A Comparative Study of Subject Pro-drop in Old Chinese and Modern Chinese*. 32nd annual New Ways of Analyzing Variation (NWAV) Meeting

Cheng, Y., Y. Hu, Q. Yu, **Z. Song**. 2003. *Xinshiji Daxue Yingyu Kuaishu Yuedu (New Century English Fast Reading)*. HUST Publish, Wuhan, China

Zhang, L., **Z. Song**. 2000. *Gunanhua Yingyu Kaoshi Chongshu – CET4 (Gu Nanhua English Test Series—CET4)*. HUST Publish, Wuhan, China

Selected Corpora Publication

BOLT English SMS/Chat
LDC2018T19. Web Download. Philadelphia: Linguistic Data Consortium, 2018

BOLT Egyptian Arabic SMS/Chat and Transliteration
LDC2017T07. Web Download. Philadelphia: Linguistic Data Consortium, 2017

BOLT Chinese Discussion Forum Parallel Training Data
LDC2017T05. Web Download. Philadelphia: Linguistic Data Consortium, 2017

MADCAT Chinese Pilot Training Set LDC2014T13. DVD.
Philadelphia: Linguistic Data Consortium, 2014.

ACE 2007 Multilingual Training Corpus LDC2014T18. Web Download. Philadelphia: Linguistic Data Consortium, 2014.

GALE Phase 2 Chinese Broadcast Conversation Parallel Text Part 1
LDC2013T11. Web Download. Philadelphia: Linguistic Data Consortium, 2013.

MADCAT Phase 1 Training Set LDC2012T15. Web Download. Philadelphia: Linguistic Data Consortium, 2012.

GALE Phase 2 Arabic Broadcast News Parallel Text LDC2012T18. Web Download. Philadelphia: Linguistic Data Consortium, 2012.

REFLEX Entity Translation Training/DevTest LDC2009T11. Web Download. Philadelphia: Linguistic Data Consortium, 2009.

GALE Phase 1 Distillation Training LDC2007T20. Web Download. Philadelphia: Linguistic Data Consortium, 2007.