

# FEI HUANG

Language Technologies Institute  
School of Computer Sciences  
Carnegie Mellon University  
5000 Forbes Avenue  
Pittsburgh, PA 15213-3891

Phone: (412) 268-5478  
Fax: (412) 268-6298  
Email: [fhuang@cs.cmu.edu](mailto:fhuang@cs.cmu.edu)  
<http://www.cs.cmu.edu/~fhuang/>

## RESEARCH INTERESTS

---

Natural Language Processing, in particular Machine Translation and Information Extraction.  
Special focus on named entity extraction and translation from multimedia documents.  
Text Mining, Spoken Language Processing, Statistical Machine Learning.

## EDUCATION

---

|   |                |
|---|----------------|
| <b>CARNEGIE MELLON UNIVERSITY</b><br>LANGUAGE TECHNOLOGY INSTITUTE, SCHOOL OF COMPUTER SCIENCE<br><i>Ph.D. in Language and Information Technologies</i> (expected fall 2005)<br>Thesis: <i>Multilingual Named Entity Extraction and Translation from Text and Speech.</i><br>Advisor: Alex Waibel | 2001 - Present |
| <b>CARNEGIE MELLON UNIVERSITY</b><br>LANGUAGE TECHNOLOGY INSTITUTE, SCHOOL OF COMPUTER SCIENCE<br><i>M.S. in Language Technologies</i>  | 1999 - 2001    |
| <b>CHINESE ACADEMY OF SCIENCES</b><br><i>M.E. in Pattern Recognition and Intelligent System</i><br>Thesis: <i>Language Model Adaptation for Chinese Speech Recognition.</i>   | 1996 -1999     |
| <b>TIANJIN UNIVERSITY</b><br><i>B.E. (major) in Electrical Engineering and Automation</i> with High Honors<br><i>B.A. (minor) in English for Science and Technology</i><br>GPA: 90/100  | 1992-1996      |

## SELECTED HONORS AND AWARDS

---

- |   |           |
|---|-----------|
| • Graduate Student Conference Scholarship, Carnegie Mellon University | 2002      |
| • Graduate Student Research Fellowship, Carnegie Mellon University    | 1999-2005 |
| • Honored Graduate, Tianjin University                                | 1996      |
| • Liu Jianru Scholarship, Tianjin University                          | 1995      |
| • Distinguished Student of Tianjin City (top 1.5%)                    | 1993      |
| • People's Fellowship (first prize, top 3%), Tianjin University       | 1992-1996 |

## RESEARCH EXPERIENCE

---

|  |             |
|--|-------------|
| <b>CARNEGIE MELLON UNIVERSITY</b><br><b>Graduate Research Assistant in the Interactive Systems Labs</b><br>Major research member of several projects on text and speech translation and named entity (NE) extraction and translation, funded by the Defense Advanced Research Projects Agency (DARPA), the | 1999 – 2005 |
|--|-------------|

Department of the Interior, National Business Center (NBC) and the National Science Foundation (NSF).

**STEEM (Summarization, Translation and Entity Extraction of Multimedia Documents)**

Key technology contributor responsible for the entity extraction and translation component. Developed a cluster-specific name transliteration model, which reduced the transliteration character error rate from 50.29 to 12.89.

**STR DUST (Domain Unlimited Speech Translation)**

Developed core technologies to extract and translate NEs from Chinese broadcast news speech into English, and improved the NE translation accuracy by 14.8%.

**TIDES RADD-MT (Rapidly Adaptable Data Driven – Machine Translation)**

Developed effective language-independent algorithms for automatic NE extraction, alignment and translation from bilingual and monolingual corpora. The NE translation module has been contributed to the CMU statistical machine translation system, the DARPA Surprise Language Exercise and the AMTEXT project (Machine Translation based on Information Extraction).

**CPOF (Command Post Of the Future)**

Proposed a dialogue manager that uses Bayesian network for incremental fusion of multiple information cues to minimize user effort.

**CHINESE ACADEMY OF SCIENCES**

1997 – 1999

**Research Assistant in the National Laboratory of Pattern Recognition**

Conducted research on statistical language modeling for Chinese large vocabulary continuous speech recognition, including Chinese word segmentation, data smoothing for language modeling and lexicon and language model adaptation.

**TEACHING EXPERIENCES**

---

*Teaching Assistant* for Speech Recognition and Understanding Fall 2002  
Graduate level course on theoretical backgrounds and experimental practices of state-of-the-art speech recognition and understanding. Responsibilities include designing homework and course projects, holding weekly office hours and grading homework and exams.

**SELECTED PUBLICATIONS**

---

**Fei Huang**, Cluster-specific Name Transliteration, To appear in the Proceedings of the HLT-EMNLP 2005, Vancouver, BC, Canada, October 2005.

**Fei Huang**, Ying Zhang and Stephan Vogel, Mining Key Phrase Translations from Web Corpora, To appear in the Proceedings of the HLT-EMNLP 2005, Vancouver, BC, Canada, October 2005.

**Fei Huang**, Stephan Vogel and Alex Waibel, Clustering and Classifying Names by Origins, in the Proceedings of the 25<sup>th</sup> Annual Meeting of the American Association of Artificial Intelligence (AAAI-05), Pittsburgh, PA, July 2005.

Ying Zhang, **Fei Huang** and Stephan Vogel, Mining Translations of OOV Terms from the Web through Cross-lingual Query Expansion, in the Proceedings of the 28<sup>th</sup> Annual International ACM SIGIR, Salvador, Brazil, August 2005.

**Fei Huang**, Stephan Vogel and Alex Waibel, Towards Named Entity Extraction and Translation in Spoken Language Translation, in the Proceedings of the International Workshop on Spoken Language Translation, Kyoto, Japan, October 2004.

**Fei Huang**, Stephan Vogel and Alex Waibel, Improving Named Entity Translation Combining Phonetic and Semantic Similarities, in the Proceedings of the Human Language Technologies Conference (HLT/NAACL2004), Boston, USA, May 2004.

**Fei Huang**, S. Vogel and A. Waibel, Extracting NE Translingual Equivalence with Limited Resources, ACM Transactions on Asian Language Information Processing, 2(2), 124-129. 2003

Stephan Vogel, Ying Zhang, **Fei Huang**, Alicia Tribble, Ashish Venogupal, Bing Zhao, Alex Waibel, The CMU Statistical Translation System, in the Proceedings of MT Summit IX, New Orleans, LA, USA, September 2003.

**Fei Huang**, Stephan Vogel and Alex Waibel, Automatic Extraction of Named Entity Translingual Equivalence Based on Multi-feature Cost Minimization, in the Proceedings of the 41st Annual Conference of the Association for Computational Linguistics (ACL'03), Workshop on Multilingual and Mixed-language Named Entity Recognition, Sapporo, Japan, July 2003.

**Fei Huang** and Stephan Vogel, Improved Named Entity Translation and Bilingual Named Entity Extraction, in the Proceedings of the 2002 International Conference on Multimodal Interfaces (ICMI '02), Pittsburgh, USA, October 2002.

**Fei Huang** and Alex Waibel, An Adaptive Approach to Named Entity Extraction for Meeting Applications, in the Proceedings of the 2002 Human Language Technologies Conference (HLT'02), San Diego, USA, April 2002.

**Fei Huang**, Jie Yang and Alex Waibel, Dialogue Management for Multimodal User Registration, in the Proceedings of the 2000 International Conference on Spoken Language Processing (ICSLP'00), Beijing, China, October 2000.

**Fei Huang**, Language Model Adaptation for Chinese Speech Recognition, Master Thesis, Chinese Academy of Sciences, 1999.

## INVITED PRESENTATIONS

---

- University of Karlsruhe, Germany 7. 2005
- Institute of Automation, Chinese Academy of Sciences, Beijing, China 1.2005
- International Workshop on Speech Summarization for Information Extraction and Machine Translation, Painesville, OH, USA 9.2004
- DARPA TIDES Machine Translation Workshop, Gaithersburg, MD, USA 7.2003

## PROFESSIONAL AND COMMUNITY SERVICES

---

- Reviewer for ACM Transactions on Information System, ACM Transactions on Asian Language Information Processing, Machine Learning Journal, Special Issue on Learning in Speech and Language Technologies, InterSpeech 2005, ACL2005,
- Vice president of the Student Union, Department of Electrical Engineering and Automation, Tianjin University, 1993-1995
- Invited Member on the Judge Panel, LTI Student Research Symposium, Carnegie Mellon University, 2003

## PROFESSIONAL MEMBERSHIP

---

- Member of AAAI, ACL and IEEE