

## Shu-Jian HUANG (黄书剑)

---

Address: Dept. Computer Science and Technology, Nanjing University,  
22 Hankou Road, Nanjing P.R.China, 210093.

Office: Room 409, MengMinWei Building, Main campus of Nanjing University

Cellphone: +86-138-1386-6587

Email: [huangsj@nlp.nju.edu.cn](mailto:huangsj@nlp.nju.edu.cn) or [huangshujian@gmail.com](mailto:huangshujian@gmail.com)

Homepage: <http://nlp.nju.edu.cn/huangsj>

### Personal Information:

➤ Birthday: Apr. 1<sup>st</sup> 1984                      Gender: Male                      Nationality: P. R. China

### Education:

- Ph.D. Candidate  
Sep. 2008 - Now: Department of Computer Science and Technology, Nanjing University
- M.S. Sep. 2006 – Jul. 2008: Department of Computer Science and Technology, Nanjing University
- B.S. Sep. 2002 – Jul. 2006: Department of Computer Science and Technology, Nanjing University  
GPA: 4.3/5.0    Rank: Top 10%

### Working Experiences:

- Sep. 2005 - Now              Member of **Natural Language Processing Group (NLP)**, **Nanjing University (NJU)**, supervised by Prof. Jiajun Chen.
- Oct. 2007 - Aug. 2008    Intern at **Natural Language Computing group (NLC)**, **Microsoft Research Asia (MSRA)**, working with Long Jiang and Mu Li.

### Research Experiences and Projects:

Currently, I'm focusing on the research of Statistical Machine Translation. Besides, I'm also interested at the research of Coreference Resolution and other NLP problems, especially using various machine learning techniques to solve complicated NLP problems.

- **Statistical Machine Translation:**                      **Sep. 2008 - Now @NLP.NJU**  
I'm leading a group of 5 students working on SMT. We are planning researches and experiments about various parts of the whole SMT procedure, for example word alignment, phrase extraction and scoring, phrase reordering, etc. Some of our work are presented in Publication [2][4][6]. We participated two MT evaluations in 2009, namely NIST Open MT Evaluation 2009 and the evaluation of 5th China Workshop on Machine Translation. See our two TRs for details. Currently, I'm working on supervised and semi-supervised methods on word alignment.
- **Statistical Machine Translation:**                      **Apr. 2008 - Aug. 2008 @NLC.MSRA**  
Collaborating with Mu Li, I tried to improve the performance of word alignment by combining the outputs of different systems.  
I was also responsible for the rule extraction part of MSRA's SMT system for NIST Open MT evaluation 2008. Firstly, I maintained the code for "Hiero" rule extraction and did some rule filtering experiments. Later on, I implemented some syntactic rule extraction procedures according to some papers from ISI, USC.
- **Chinese Couplet project:**                      **Oct. 2007 - Mar. 2008 @NLC.MSRA**

This project aims at building automatic second-line and banner generation systems for Chinese Couplet. I mainly worked with Long Jiang, on the automatic evaluation of second-line generation system.

➤ **Coreference Resolution:**

**Jul. 2006 – Jul. 2007 @NLP.NJU**

We firstly worked on unsupervised methods based on graph cut. Then some supervised learning techniques such as perceptron and Markov Logic Networks were employed. We also worked on automatic generation of first order logic rules and clustering techniques. Some of these work are presented in Publication [1][3][5][7].

➤ **Chinese word segmentation and POS tagging:**

**Nov. 2005 - Apr. 2006 @NLP.NJU**

I implemented a Chinese word segmenter and a POS tagger based on HMM.

## **Publications:**

2009

7. Yabing Zhang, Junsheng Zhou, Shujian Huang and Jiajun Chen. Combining ILP and MLN for Coreference Resolution. in International Conference on Asian Language Processing (IALP 2009), Singapore, Dec 7-9, 2009
6. Biping MENG, Shujian Huang, Xinyu Dai and Jiajun Chen. Segmenting Long Sentence Pairs for Statistical Machine Translation. in International Conference on Asian Language Processing (IALP 2009), Singapore, Dec 7-9, 2009
5. Liu Weipeng, Zhou Junsheng, Huang Shujian and Chen Jiajun. Global Optimization Based On Clustering for Coreference Resolution. in The 10th Chinese National Conference on Computational Linguistics (CNCCL-2009), Yantai, China, July 24-26, 2009. (In Chinese)
4. Shujian Huang, Ning Xi, Yinggong Zhao, Xinyu Dai, Jiajun Chen. An Error-Sensitive Metric for Word Alignment in Phrase-based SMT. Journal of Chinese Information Processing, 2009, vol. 23, no. 3. (Revised version of CWMT2008 paper, In Chinese)
3. Shujian Huang, Yabing Zhang, Junsheng Zhou, Jiajun Chen. Coreference Resolution using Markov Logic Networks. (**Best Poster Award (1/25)**) in The 10th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'2009), Mexico city, Mexico, 2009. in Research in Computing Science: Advances in Computational Linguistics, Alexander Gelbukh Ed., vol. 41, page 157~168, ISSN: 1870-4069)

2008

2. Shujian Huang, Ning Xi, Yinggong Zhao, Xinyu Dai, Jiajun Chen. An Error-Sensitive Metric for Word Alignment in Phrase-based SMT. presented in The 4th China Workshop on Machine Translation, CWMT'2008, Beijing, China, 2008. (In Chinese)

2007

1. Junsheng Zhou, Shujian Huang, Jiajun Chen and Weiguang Qu. A New Graph Clustering Algorithm for Chinese Noun Phrase Coreference Resolution. Journal of Chinese Information Processing, 2007, vol. 21, no. 2. (In Chinese)

### **Technical Reports:**

Shujian Huang, Yinggong Zhao, Boyuan Li, Qiufeng Wu, Xinyu Dai, Jiajun Chen. Nanjing University's System Report for NIST MT09 Workshop. Included in the materials of NIST Open MT 2009, Ottawa, ON, Canada. Aug 31-Sep 1, 2009.

Shujian Huang, Yinggong Zhao, Boyuan Li, Qiufeng Wu, Xinyu Dai, Jiajun Chen. NJU-NLP's Technique Report for the 5th China Workshop on Machine Translation. Included in the materials of CWMT 2009, Nanjing, China. Oct. 15-16, 2009. (In Chinese)

## **Honors and Awards:**

- 2006-2007: Excellent Member Award of Graduate Students Union (Nanjing University)
- 2003-2006: Renmin Scholarship (2rd Prize, Nanjing University, three times)
- 2002-2004: Excellent Student Award (Dept. of Computer Science and Technology, twice)
- 2002-2003: Renmin Scholarship (1nd Prize, Nanjing University)

## **Self Evaluation:**

- Passionate, hard-working, team work spirits with good communication skills.

## **Skills:**

- Programming languages: C#, C++, Java, Perl, Matlab.
- Fluent speaking and writing English.