

XIAO, Tong (肖桐)
Postdoctoral researcher
Institute of Computer Software
College of Information Science & Engineering
Northeastern University
Shenyang, P.R.China.
Postal Code: 110819

Tel: (+86) 24-83672481
Fax: (+86) 24-83672481
Email: xiaotong@mail.neu.edu.cn
Web: <http://www.nlplab.com/members/~xiaotong.html>

- Education**
- Ph.D. in Computer Science 09/2008 - 07/2012
Advisor: Prof. Jingbo Zhu
External co-advisor: Dr. Keh-Yih Su
Natural Language Processing Lab
Northeastern University, Shenyang, China
 - M.S. in Computer Science 09/2005 - 03/2008
Supervisor: Dr. Li Zhang
Natural Language Processing Lab
Northeastern University, Shenyang, China
 - B.S. in Computer Science 09/2001 - 07/2005
Northeastern University, Shenyang, China

- Research Interests**
- Natural Language Processing:
 - Machine Translation (My recent work has been focused on syntax-based MT, in particular the approaches/models that make use of the syntax of both source and target languages)
 - Syntactic Alignment
 - Language Modeling

- Experience**
- **Postdoctoral Researcher** 07/2012 - Now
Natural Language Processing Lab, Northeastern University, Shenyang, China
Advisor: Prof. Jingbo Zhu
Working on "Syntax-based and semantics-augmented machine translation"
 - **Visiting Postdoctoral Researcher** 03/2013 - Now
Department of Engineering, University of Cambridge, UK
Advisor: Prof. Bill Byrne
Working on "Source-language Modeling for Chinese-English MT"
 - **Ph.D. Candidate** 09/2008 - 07/2012
Natural Language Processing Lab, Northeastern University, Shenyang, China
Working on "Training and Decoding for Tree-to-tree Translation"
 - **Research Intern** 08/2008 - 07/2009
Microsoft Research Asia, Beijing, China
Mentor: Dr. Mu Li
Working on "Better Synchronous Binarization for Machine Translation"
 - **Visiting Fellowship** 09/2006 - 09/2007
Fuji Xerox, Hadano, Japan
Mentor: Dr. Shaoming Liu
Working on "Chinese Chunking" and "Example-based Approaches to Chinese-Japanese Machine Translation"

- **Master** 09/2005 - 03/2008
Natural Language Processing Lab, Northeastern University, Shenyang, China
Working on "Comparison of Word-based and Phrase-based Approaches to Machine Translation".

Professional Activities

- **PC members** of CWMT2013, ACL2013, IJCNLP2011, SWCL2010, SEWM2009 and SWCL2008
- **Reviewer** for Journal of the Acoustical Society of America (2011-2012)
- **Reviewer** for International Journal of Pattern Recognition and Artificial Intelligence (2012-2013)
- **Reviewer** for International Journal of Computational Linguistics and Chinese Language Processing (2007-2012)
- **Reviewer** for International Journal of Computer Processing Of Languages (2010-2012)
- **Reviewer** for Chinese Journal of Electronics (2012-2013)
- **Secondary Reviewers** for EACL2012, EMNLP2012 and AAAI2011

Honors and Awards

- **Excellent Paper Award** of The 5th National Young Workshop on Computational Linguistics, 2010
- **Excellent Paper Award** of The 10th Chinese National Conference on Computational Linguistics, 2009
- **Excellent Master Thesis Award** of Northeastern University, 2008
- **Excellent Undergraduate Student Fellowship**, Northeastern University, 2001-2005

Evaluation Tasks

- [1] **WMT13 Russian-English MT track - 2nd/1st place** (case-insensitive/sensitive BLEU)
Juan Pino, Aurelien Waite, Tong Xiao, Adri de Gispert, Federico Flego and William Byrne. 2013. The University of Cambridge Russian-English System at WMT13. In *Proc. of the Eighth Workshop on Statistical Machine Translation (WMT)*, Sofia, Bulgaria, pages 200-205.
- [2] **NTCIR-9 Chinese-English Patent MT Track - 2nd place** (human evaluation)
Tong Xiao, Qiang Li, Qi Lu, Hao Zhang, Haibo Ding, Shujie Yao, Xiaoming Xu, Xiaoxu Fei, Jingbo Zhu, Feiliang Ren and Huizhen Wang. 2011. The NiuTrans Machine Translation System for NTCIR-9. In *Proc. of NTCIR-9 Workshop Meeting*, Tokyo, Japan, pages 593-599.
- [3] **CWMT2011 English-Chinese/Chinese-English News-Domain MT Tracks - 1st and 4th places** (BLEU)
Tong Xiao, Hao Zhang, Qiang Li, Qi Lu, Jingbo Zhu, Feiliang Ren and Huizhen Wang. 2011. The NiuTrans Machine Translation System for CWMT2011. In *Proc. of The 6th China workshop on Machine Translation (CWMT)*, Xiamen, China.
- [4] **CWMT2009 Chinese-English Single System Track - 2nd place** (BLEU)
Tong Xiao, Rushan Chen, Tianning Li, Muhua Zhu, Jingbo Zhu, Huizhen Wang and Feiliang Ren. 2009. NEUTrans: a Phrase-Based SMT System for CWMT2009. In *Proc. of The 5th China workshop on Machine Translation (CWMT)*, Nanjing, China.
- [5] **NTCIR-7 English Patent Mining Track - 1st place** (MAP)
Tong Xiao, Feifei Cao, Tianning Li, Guolong Song, Ke Zhou, Jingbo Zhu and Huizhen Wang. 2008. KNN and Re-ranking Models for English Patent Mining at NTCIR-7. In *Proc. of NTCIR-7 Workshop Meeting*, Tokyo, Japan, pages 333-340.

Publications (selected)

- [1] Tong Xiao and Jingbo Zhu. Unsupervised Sub-tree Alignment for Tree-to-tree Translation. To appear in *Journal of Artificial Intelligence Research*.

-
- [2] Jingbo Zhu, Qiang Li and Tong Xiao. Improving Syntactic Rule Extraction through Deleting Spurious Links with Translation Span Alignment. To appear in *Natural Language Engineering*.
- [3] Ji Ma, Jingbo Zhu, Tong Xiao and Nan Yang. 2013. Easy-First POS Tagging and Dependency Parsing with Beam Search. In *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (ACL, short paper)*, Sofia, Bulgaria, pages 110-114.
- [4] Tong Xiao, Jingbo Zhu and Tongran Liu. Bagging and Boosting Statistical Machine Translation Systems. *Artificial Intelligence (AI)*, Volume 195, February 2013, 496 - 527.
- [5] Ji Ma, Tong Xiao, Jingbo Zhu, and Feiliang Ren. 2012. Easy-First Chinese POS Tagging and Dependency Parsing. In *Proc. of the 24rd International Conference on Computational Linguistics (COLING)*, Mumbai, India.
- [6] Tong Xiao, Jingbo Zhu, Hao Zhang and Qiang Li. 2012. NiuTrans: An Open Source Toolkit for Phrase-based and Syntax-based Machine Translation. In *Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (ACL, system demonstration)*, Jeju, Korea, pages 19-24.
- [7] Jingbo Zhu, Tong Xiao and Chunliang Zhang. 2012. Learning Better Rule Extraction with Translation Span Alignment. In *Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (ACL, short paper)*, Jeju, Korea, pages 280-284.
- [8] Tong Xiao, Jingbo Zhu and Shujie Yao. 2011. Document-level Consistency Verification in Machine Translation. In *Proc. of MT summit XIII*, Xiamen, China, pages 131-138.
- [9] Jingbo Zhu and Tong Xiao. 2011. Improving Decoding Generalization for Tree-to-String Translation. In *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics (ACL, short paper)*, Portland, USA, pages 418-423.
- [10] Tong Xiao, Jingbo Zhu and Muhua Zhu. 2011. Language Modeling for Syntax-based Machine Translation Using Tree Substitution Grammars: A Case Study on Chinese-English Translation. *ACM Transactions on Asian Language Information Processing (TALIP)*, 10(4):1-29.
- [11] Muhua Zhu, Jingbo Zhu and Tong Xiao. 2011. Automatic Treebank Conversion via Informed Decoding - A Case Study on Chinese Treebanks. *ACM Transactions on Asian Language Information Processing (TALIP)*, 10(3):1-23.
- [12] Tong Xiao, Jingbo Zhu, Muhua Zhu and Huizhen Wang. 2010. Boosting-based System Combination for Machine Translation. In *Proc. of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, Uppsala, Sweden, pages 739-748.
- [13] Tong Xiao, Jingbo Zhu, Hao Zhang and Muhua Zhu. 2010. An Empirical Study of Translation Rule Extraction with Multiple Parsers. In *Proc. of the 23rd International Conference on Computational Linguistics (COLING, poster session)*, Beijing, China.
- [14] Muhua Zhu, Jingbo Zhu and Tong Xiao. 2010. Heterogeneous Parsing via Collaborative Decoding. In *Proc. of the 23rd International Conference on Computational Linguistics (COLING)*, Beijing, China.
- [15] Shujie Yao, Tong Xiao and Jingbo Zhu. 2010. Selection of SMT Training Data Based on Sentence Pair Quality and Coverage. In *Proc. of The 5th National Young Workshop on*

Computational Linguistics, China, pages 221-227 (In Chinese. **Excellent Paper Award**).

- [16] Hao Zhang, Huizhen Wang, Tong Xiao and Jingbo Zhu. 2010. The impact of parsing accuracy on syntax-based SMT. In *Proc. of IEEE NLPKE 2010*, Beijing, China.
- [17] Tong Xiao, Mu Li, Dongdong Zhang, Jingbo Zhu and Ming Zhou. 2009. Better Synchronous Binarization for Machine Translation. In *Proc. of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Singapore, pages: 362-370.
- [18] Nan Duan, Mu Li, Tong Xiao and Ming Zhou. 2009. The Feature Subspace Method for SMT System Combination. In *Proc. of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Singapore, pages 1096-1104.
- [19] Tong Xiao, Tianning Li, Rushan Chen, Jingbo Zhu and Huizhen Wang. 2009. Word Realignment for Statistical Machine Translation. In *Proc. of The 10th Chinese National Conference on Computational Linguistics, China*, pages 439-445 (In Chinese).

Papers in Submission

- [1] Tong Xiao, Jingbo Zhu, Keh-Yih Su. Beam-width limited and BLEU Oriented Training for Tree-to-tree Translation. Submitted to a major conference.

Project Experience

- **Co-principal Investigator and Core Developer** 09/2009 - Present
NiuTrans Open Source Project (<http://www.nlplab.com/NiuPlan/NiuTrans.html>)
- **Principal Investigator** 01/2010 - 12/2011
 Project: *Towards Better Use of Tree Structure for Machine Translation*
 Funded by the Ministry of Education of the People's Republic of China
- **Principal Investigator** 07/2013 - 06/2014
 Project: *Better Use of Source-Language Syntax for Chinese-English Translation*
 Funded by the Ministry of Education of the People's Republic of China
- **Principal Investigator** 01/2013 - 12/2015
 Project: *On Skeleton-based Models for Syntactic Statistical Machine Translation*
 Funded by the National Natural Science Foundation of China

Software

- NiuTrans open source training and decoding platform for statistical machine translation.
<http://www.nlplab.com/NiuPlan/NiuTrans.html>

Personal Information

- Citizen of the P. R. China.
- Born June 1982, Shenyang, China.
- Languages: Mandarin (native); English (read/write/speak); Japanese (basic conversational).
- Programming Languages: C/C++, C#, Perl

References

- available upon request