



## Introducing Prior Knowledge into Machine Translation 机器翻译中先验知识的使用

**Prof. Tong Xiao**

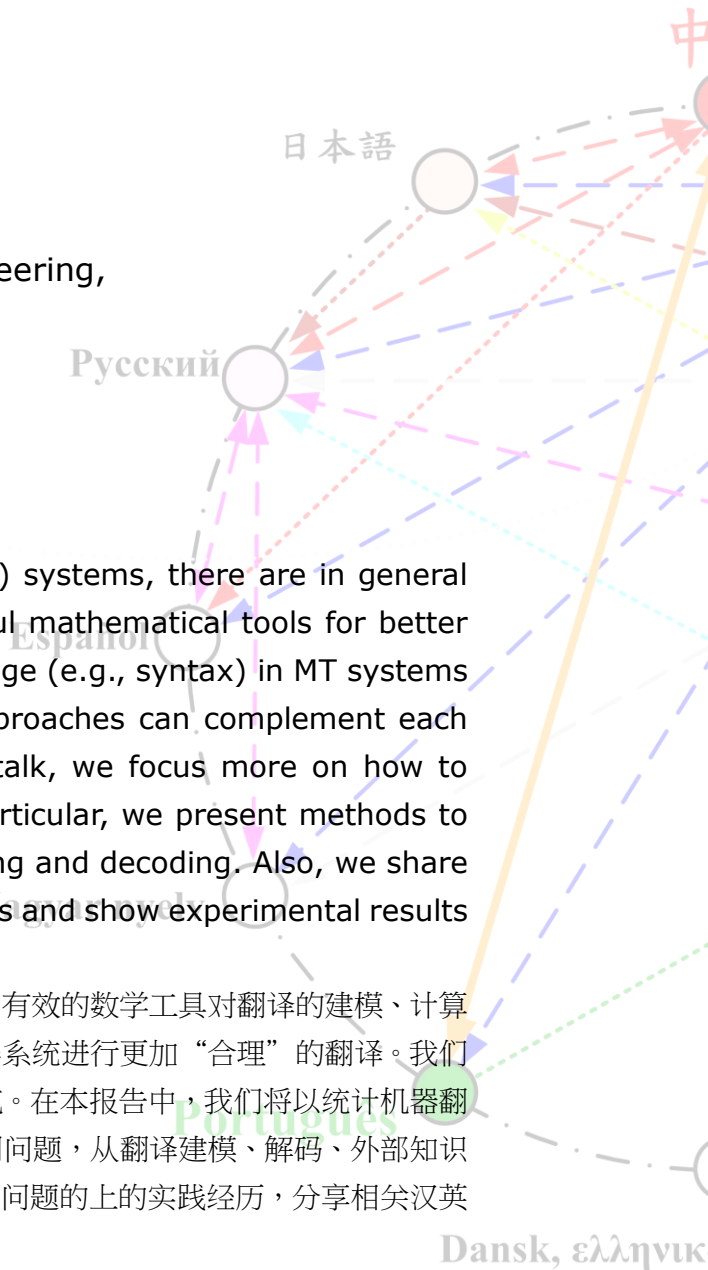
*Associate Professor*

College of Computer Science and Engineering,  
Northeastern University, China

### Abstract

To develop better machine translation (MT) systems, there are in general two paradigms: 1) exploring more powerful mathematical tools for better modeling; and 2) using more prior knowledge (e.g., syntax) in MT systems to translate more like a person. These approaches can complement each other but conflict in some cases. In this talk, we focus more on how to incorporate prior knowledge into MT. In particular, we present methods to use different types of knowledge in modeling and decoding. Also, we share our experiences in developing these systems and show experimental results on Chinese-English tasks.

机器翻译的发展有两个重要的维度，一是探索更为有效的数学工具对翻译的建模、计算等问题进行求解，二是使用丰富的先验知识来引导系统进行更加“合理”的翻译。我们发现二者在一定程度上存在矛盾，但是又相互补充。在本报告中，我们将以统计机器翻译为背景，重点分析在机器翻译中先验知识的使用问题，从翻译建模、解码、外部知识的使用多个层面展开讨论。最终，结合我们在相关问题的上的实践经历，分享相关汉英翻译实验结果及系统开发经验。



## Biography



Dr. Tong Xiao is an associate professor at the College of Computer Science and Engineering, Northeastern University (CN). He obtained his Ph.D. from Northeastern University (CN) in 2012, and was elected as a candidate of the Excellent Ph.D. Dissertation Award established by CIPS (Chinese Information Processing Society of China). He also visited several research institutes for machine translation research, including Fuji Xerox, Microsoft Research Asia, and the University of Cambridge. He was and is a co-PI of the NiuTrans open source MT project (<http://www.niutrans.com/niutrans/NiuTrans.html>). By now, Dr. Xiao has published over twenty papers on AI, JAIR, AACL, ACL, EMNLP and COLING. His current research interests include statistical machine translation and language modeling.

肖桐，博士，东北大学计算机科学与工程学院副教授，中国中文信息学会首届优秀博士学位论文提名奖获得者，NiuTrans 团队技术负责人。2012 年博士毕业于东北大学，并先后在日本富士施乐公司、微软亚洲研究院进行访问学习。2013-2014 赴英国剑桥大学开展博士后研究。作为技术负责人主持 NiuTrans 开源统计机器翻译系统的研发 (<http://www.niutrans.com/niutrans/NiuTrans.html>)，曾主持多套机器翻译评测（比赛）系统的研发，并在国内机器翻译评测 CWMT 和国际专利机器翻译评测 NTCIR PatentMT 中取得优异成绩。主要从事基于句法的统计机器翻译、语言建模方面的研究。在人工智能及自然处理语言领域重要期刊及会议 AI、JAIR、TALIP、AAAI、ACL、EMNLP、COLING 发表论文 20 余篇。社会学术兼职包括：中国中文信息学会青年工作委员会委员、中国中文信息学会信息检索与内容安全专业委员会委员等。

**Venue:** E11-1012  
**Date:** 29 August, Monday  
**Time:** 16:00~17:00  
**Language:** Mandarin

**Contacts:** Dr. Derek F. Wong / Dr. Lidia S. Chao  
**Telephone:** (853) 8822-8051  
**Website:** <http://nlp2ct.cis.umac.mo/>

**Address:**  
E11-2023, University of Macau,  
Avenida da Universidade,  
Taipa, Macau, China