

Yanjun MA

PERSONAL DATA

PLACE OF BIRTH: Shandong, China
DATE OF BIRTH: 11 November, 1981
GENDER: Male
ADDRESS: VB206B Postgraduate Apartments
Campus Residences, DCU
Dublin 9, Ireland
PHONE: +353-87-2431316
EMAIL: yma@computing.dcu.ie

WORK EXPERIENCE

<i>Current</i> Nov 2009	Postdoctoral Researcher Dublin City University Carrying out research in the Centre for Next Generation Localisation (i) Continued research on word alignment, especially on open-sourcing the tools developed during PhD research; (ii) Research on integrating translation memory and SMT output, and SMT output confidence estimation; (iii) Linguistically sophisticated models for machine translation, such as tree-to-tree translation
JUL-NOV 2009	Research Assistant Dublin City University Carrying out research in the Centre for Next Generation Localisation mainly focusing on IWSLT 2009 evaluation campaign.
JAN 2009-FEB 2009	Visiting Student Cambridge University Engineering Department Extending the MTTK word alignment toolkit to handle syntactic dependencies.
DEC 2004-MAR 2005	Visiting Student Microsoft Research Asia Working on an automatic Couplet generation system, which is now successfully commercialised.

EDUCATION

- NOV 2009 PhD in COMPUTER SCIENCE, **Dublin City University**, Ireland
Major: Computer Applications
Thesis: “Constrained Word Alignment Models for Statistical Machine Translation” — Advisor: Prof. Andy WAY
- JUN 2006 Master of Arts, **Tsinghua University**, China
Major: Computational Linguistics
Thesis: “Automatic Identification and Alignment of Chinese-English Phrases” — Advisor: Prof. Ying LIU
GPA: Overall 3.7/4.0 (89/100) Major 3.7/4.0 (89/100)
- JUN 2003 Bachelor of Arts, **Shandong University**, China
Major: Linguistics
Thesis: “An Investigation on the Phonetic System of Kenli Dialect”
Advisor: Prof. Lijing YUE, Prof. Zengyi QIAN
GPA: Overall 3.7/4.0 (87/100) Major 4.0/4.0 (90/100)
RANK: 3/85

HONORS AND AWARDS

- MAY 2010 EAMT Best PhD Thesis Award
JUN 2007 Don and Betty Walker Award
JUN 2006 Science Foundation Ireland Fellowship
SEP 2003 National Scholarship for Graduate Study
JUN 2006 Undergraduate Excellence Award (Jointly by Shandong University and Shandong Province)
DEC 2000 First Place in Spoken English Contest of Shandong University
SEP 2000, 2001 First-class Scholarship for Excellence
OCT 1999 Coca-Cola Scholarship
JUN 1998 Excellent Senior High School Students of Shandong Province
SEP 1995 First Place in Olypic English Games

PUBLICATIONS

Journal Articles

- Yanjun Ma** and Andy Way. Bilingually Motivated Word Segmentation for Statistical Machine Translation. In *ACM Transactions on Asian Language Information Processing, Special Issue on Machine Translation of Asian Languages*, pp.1–24.
- Yanjun Ma**. Review of Wong, Li, Xu and Zhang’s Introduction to Chinese Natural Language Processing. In *Journal of Machine Translation*. Forthcoming.

Books

- Vera Demberg, **Yanjun Ma** and Nils Reiter. Proceedings of the Student Research Workshop at EACL 2009, Athens, Greece.

Conference Papers

- Yifan He, **Yanjun Ma**, Josef van Genabith and Andy Way. TM-SMT Integration as Ranking. In *Proceedings of the COLING 2010*, Beijing, China. Forthcoming.
- Yifan He, **Yanjun Ma**, Josef van Genabith and Andy Way. Bridging SMT and TM with Translation Recommendation. In *Proceedings of the ACL 2010*, Uppsala, Sweden. Forthcoming.
- Yanjun Ma**, Tsuyoshi Okita, Özlem Çetinoğlu, Jinhua Du and Andy Way. Low-resource Machine Translation using MaTrEx: the DCU Machine Translation System for IWSLT 2009. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT 2009)*, pp.29–36. Tokyo, Japan.
- Jinhua Du, **Yanjun Ma** and Andy Way. Source-Side Context-Informed Hypothesis Alignment for Combining Outputs from Machine Translation Systems. In *Proceedings of Machine Translation Summit XII*, pp.230–237. Ottawa, ON, Canada.
- Patrik Lambert, **Yanjun Ma**, Sylwia Ozdowska and Andy Way. Tracking Relevant Alignment Characteristics for Machine Translation. In *Proceedings of Machine Translation Summit XII*, pp.268–275. Ottawa, ON, Canada.
- Yanjun Ma**, Patrik Lambert and Andy Way. Tuning Syntactically Enhanced Word Alignment for Statistical Machine Translation. In *Proceedings of the EAMT 2009, the 13th Annual Meeting of the European Association for Machine Translation*, pp.250–257. Barcelona, Spain.
- Rejwanul Haque, Sudip Naskar, **Yanjun Ma** and Andy Way. 2009. Using Supertags as Source Language Context in SMT. In *Proceedings of EAMT 2009, the 13th Annual Meeting of the European Association for Machine Translation*, pp.234–241. Barcelona, Spain.
- Yanjun Ma** and Andy Way. Bilingually Motivated Domain-Adapted Word Segmentation for Statistical Machine Translation. In *Proceedings of the EACL 2009, the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp.549–557. Athens, Greece.
- Yanjun Ma**, John Tinsley, Hany Hassan, Jinhua Du and Andy Way. 2008. Exploiting Alignment Techniques in MaTrEx: the DCU Machine Translation System for IWSLT 2008. In *Proceedings of the IWSLT 2008 Workshop (IWSLT 2008)*, pp.26–33. Waikiki, HI.
- Yanjun Ma**, Sylwia Ozdowska, Yanli Sun and Andy Way. 2008. Improving Word Alignment Using Syntactic Dependencies. In *Proceedings of ACL08 workshop on Syntax and Structure in Statistical Translation (ACL08-SSST)*, pp.69-77. Columbus, OH.
- John Tinsley, **Yanjun Ma**, Sylwia Ozdowska and Andy Way. 2008. MaTrEx: the DCU MT System for WMT 2008. In *Proceedings of ACL08 workshop on Statistical Machine Translation (ACL08-SMT)*, pp.171–174. Columbus, OH.

Hany Hassan, **Yanjun Ma**, and Andy Way. 2007. MaTrEx: the DCU Machine Translation System for IWSLT 2007. In *Proceedings of the IWSLT 2007 Workshop (IWSLT 2007)*, pp.69–75. Trento, Italy.

Yanjun Ma, Nicolas Stroppa, and Andy Way. 2007. Alignment-Guided Chunking . In *Proceedings of The 11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)* , pp.114–121. Skövde, Sweden.

Ying Zhang, Gareth J. F. Jones, Sen Zhang, Bin Wang, Yuqing Guo and **Yanjun Ma**. 2007. An Investigation of Question Translation for English-Chinese Cross-Language Question Answering. In *Proceedings of 2007 China-Ireland International Conference on Information and Communications Technologies (CIICT 2007)*, Dublin, Ireland.

Yanjun Ma, Nicolas Stroppa, and Andy Way. 2007. Bootstrapping Word Alignment via Word Packing. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, pp.304–311. Prague, Czech Republic.

Invited Talks

Presentation for EAMT 2009 best thesis award. May 28, 2010.

Kevin Knight’s Bayesian Inference with Tears. *NCLT Seminar Series*. March 31st, 2010.

Bilingually Motivated Domain-Adapted Word Segmentation for Statistical Machine Translation. *EACL 2009*. April 2nd, 2009.

Out of Giza: Efficient Models for Word Alignment. *NCLT Seminar Series*. March 3rd, 2009.

Improving Word Alignment for Statistical Machine Translation. *NCLT-CNGL workshop*. July 23-24, 2008.

Improving Word Alignment Using Syntactic Dependencies. *ACL08 workshop on Syntax and Structure in Statistical Translation*, June 20, 2008.

Generative Word Alignment Models. *NCLT seminar series*, February 13, 2008.

New Word and Phrasal Alignment Methods for Machine Translation. *Seminar for Visit of Prof. Hermann Ney*. June 15, 2007.

Toward Syntax-rich Statistical Machine Translation. *NCLT seminar series*, January 10, 2007.

Extracting Equivalent Chunks from Chinese-English Bilingual Corpus. *NCLT seminar series*, August 23, 2006.

An Introduction to Statistical Machine Translation. *Natural Language Computing Group, Microsoft Research Asia*.

PROFESSIONAL ACTIVITIES

CONFERENCE PRESENTATIONS ACL 2007, TMI 2007, ACL 2008 SMT workshop
ACL 2008 SSST workshop, EACL 2009, MT Summit 2009
EAMT 2010

CHAIRS Co-chair of EACL 2009 Student Research Workshop

REVIEWERS ACL (2009), LREC (2010), COLING (2010),
EMNLP (2009, 2010), IWSLT (2009), NAACL (2010),
Journal of Computer Science and Technology (2010),
Journal of Knowledge-Based Systems (2010)

OTHER SERVICES webmaster for TMI 2007, student volunteer at ACL 2007

MEMBERSHIP Student Board member of EACL, member of ACL and EAMT

SUPERVISION

2010- Hanna Bechara, PhD candidate, working on statistical post-editing
Jointly with Prof. Josef van Genabith, Dublin City University

TEACHING EXPERIENCE

2010 Exam Board member of module CA313
Dublin City University

2008-2009 Lecturer of module CA313: *Algorithms and Complexity*
Dublin City University

2007-2008 Tutor of module CA215: *Language and Computability*
Dublin City University

2006-2007 Tutor of module CA247: *Advanced Programming*
Dublin City University

2006-2007 Tutor of module CA218: *Introduction to Databases*
Dublin City University

FEB 2006-JUN 2006 Part-time Lecturer of *English*, Beijing City University

SEP 2005-DEC 2005 Part-time lecturer of *English*, Beijing Geely University

SEP 2005-DEC 2005 Part-time lecturer of *Modern Mandarin* to international
students, Tsinghua University