

John S. Y. Lee

EDUCATION

- PhD in Electrical Engineering and Computer Science, MIT, USA (2009)
- S.M. in Electrical Engineering and Computer Science, MIT, USA (2004)
- Bachelor of Mathematics in Computer Science, University of Waterloo, Canada (2002)

ACADEMIC POSITIONS

- *Assistant Professor*, Department of Chinese, Translation and Linguistics, City University of Hong Kong, Hong Kong (2010 – current)
- *Visiting Research Assistant*, Perseus Digital Library, Tufts University, USA (2009)

RESEARCH PROJECTS

- PI: “Data-driven Analysis of Syntax and Parallelism in Classical Chinese Poems”, *Strategic Research Grant*, City University of Hong Kong (2014 – 2015, HK\$100,000)
- PI: “Data-driven Evaluation of Computer-assisted Translation Systems for Pedagogical Purposes”, *Teaching Development Grant*, City University of Hong Kong (2014 – 2015, HK\$180,000)
- PI: “A Treebank for the Chinese Buddhist Canon”, *General Research Fund / Early Career Scheme*, University Grants Committee, Hong Kong (2012 – 2014, HK\$746,000)
- PI: “Textual Revision in Second Language Writing”, *Germany/HK Joint Research Scheme*, University Grants Committee, Hong Kong (2012 – 2013, HK\$60,000)
- PI: “Manual and Automatic Creation of Learner Corpora”, *Strategic Research Grant*, City University of Hong Kong (2012 – 2014, HK\$180,000)
- PI: “Evidence-based Learning for Students of Linguistics and Language Technology”, *Teaching Development Grant*, City University of Hong Kong (2011 – 2013, HK\$238,140)
- PI: “Automatic Discovery of Social Networks in Literary Texts”, *Start-up Grant for New Staff*, City University of Hong Kong (2011 – 2012, HK\$100,000)
- Co-I: “Data is Beautiful: Visualization in the Humanities”, *Interdisciplinary Professional Development Award*, City University of Hong Kong (2011 – 2013, HK\$250,000)
- PI: “A Reading Environment for Classical Chinese”, *Strategic Research Grant*, City University of Hong Kong (2010 – 2012, HK\$179,326)
- PI: “An Annotated Corpus of Hong Kong ESL Learners”, *Strategic Research Grant*, City University of Hong Kong (2010 – 2011, HK\$90,000)

REFEREED PUBLICATIONS

- John Lee, Ying Cheuk Hui and Yin Hei Kong. Knowledge-rich, Computer-assisted Composition of Chinese Couplets. In *Literary and Linguistic Computing* (doi: 10.1093/lc/fqu052), 2014.
- John Lee and Yin Hei Kong. A Dependency Treebank of Chinese Buddhist Texts. In *Literary and Linguistic Computing* (doi: 10.1093/lc/fqu048), 2014.
- John Lee and Yin Hei Kong. Imagistic and Propositional Languages in Classical Chinese Poetry. In *Proc. International Conference on Asian Language Processing (IALP)*, 2014.

- John Lee and Tak-sum Wong. Corpus-based learning of Cantonese for Mandarin Speakers. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands* (pp. 1-6). Dublin: Research-publishing.net, 2014.
- John Lee, Ying Cheuk Hui, and Chak Yan Yeung. Toward a Digital Library with Search and Visualization Tools. In *Proc. 6th Language and Technology Conference*, 2013.
- John Lee, Ying Cheuk Hui, and Yin Hei Kong. Treebanking for Data-driven Research in the Classroom. In *Proc. 4th Workshop on Teaching Natural Language Processing*, 2013.
- John Lee and Tak-sum Wong. Glimpses of Ancient China from Classical Chinese Poems. In *Proc. 24th International Conference on Computational Linguistics (COLING)*, 2012.
- John Lee and Chak Yan Yeung. Extracting Networks of People and Places from Literary Texts. In *Proc. 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC)*, 2012.
- John Lee and Jonathan Webster. A Corpus of Textual Revisions in Second Language Writing. In *Proc. 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2012.
- John Lee and Yin Hei Kong. A Dependency Treebank of Classical Chinese Poems. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2012.
- John Lee, Jason Naradowsky, and David A. Smith. A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing. In *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011.
- John Lee. Toward a Parallel Corpus of Spoken Cantonese and Written Chinese. In *Proc. 5th International Joint Conference on Natural Language Processing (IJCNLP)*, 2011.
- John Lee. Verb Tense Generation. In *Proc. 12th Conference of the Pacific Association for Computational Linguistics (PACLING)*, 2011.
- John Lee. Dependency Parsing using Prosody Markers from a Parallel Text. In *Proc. 9th International Workshop on Treebanks and Linguistic Theories (TLT)*, 2010.
- John Lee and Stephanie Seneff. Correcting Misuse of Verb Forms. In *Proc. 46th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2008.
- John Lee and Stephanie Seneff. An Analysis of Grammatical Errors in Non-native Speech in English. In *Proc. IEEE Workshop on Spoken Language Technology (IWSLT)*, 2008.
- John Lee. A Nearest-Neighbor Approach to the Automatic Analysis of Ancient Greek Morphology. In *Proc. Conference on Computational Natural Language Learning (CoNLL)*, 2008.
- John Lee. A Computational Model of Text Reuse in Ancient Literary Texts. In *Proc. 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2007.
- John Lee and Stephanie Seneff. Automatic Generation of Cloze Items for Prepositions. In *Proc. Interspeech*, 2007.
- John Lee, Ming Zhou, and Xiaohua Liu. Detection of Non-native Sentences using Machine-translated Training Data. In *Proc. Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2007.
- John Lee and Stephanie Seneff. Automatic Grammar Correction for Second-Language Learners. In *Proc. Interspeech*, 2006.
- Ye-Yi Wang, Alex Acero, Milind Mahajan, and John Lee. Combining Statistical and Knowledge-based Spoken Language Understanding in Conditional Models. In *Proc. 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2006.

- Ye-Yi Wang, John Lee and Alex Acero. Speech Utterance Classification Model Training without Manual Transcriptions. In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2006.
- Stephanie Seneff, Chao Wang, and John Lee. Combining Linguistic and Statistical Methods for Bi-directional English-Chinese Translation in the Flight Domain. In *Proc. Conference of the Association for Machine Translation in the Americas (AMTA)*, 2006.
- John Lee and Stephanie Seneff. Interlingua-based Translation for Language Learning Systems. In *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2005.
- John Lee and Stephanie Seneff. Translingual Grammar Induction. In *Proc. Interspeech*, 2004.
- John Lee. Automatic Article Restoration. In *Proc. Student Research Workshop at the Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, 2004.

INVITED TALKS

- “Linguistic Annotations for Literary Studies”. *Semana de Estudos Clássicos*, Sao Paulo State University, Araraquara, Brazil, May 2014.
- “Creating a Treebank for the Chinese Buddhist Canon”. *Dharma Drum Buddhist College*, Taiwan, November 2013.
- “A Corpus Study on Syntactic Parallelism in Classical Chinese Poetry”. *North American Conference on Chinese Linguistics (NACCL)*, University of Michigan, USA, June 2013.
- “Beyond the Concordance: Tools for text analysis and visualization”. *Open University of Hong Kong*, January 2013.
- “Toward a Treebank for the Chinese Buddhist Canon”. *Pacific Neighbourhood Consortium Annual Conference and Joint Meetings (PNC)*, University of California, Berkeley, December 2012.
- “Use of Corpora in Teaching and Research: Three Examples”. *NEH Institute for Advanced Technology in the Digital Humanities*, Perseus Project. Tufts University, USA, August 2012.
- Discussion Panelist, *Workshop on Syntactic Analysis of Non-Canonical Language*, at the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Montreal, Canada, June 2012.
- “A Corpus of Textual Revisions in Second Language Writing”. *Conference on Language Learning Centers and Internationalization: Supporting Global Connections*, Southwest Association for Language Learning and Technology (SWALLT). Phoenix, USA, 2012.
- “Developing Visual Analytics for Text Data”. *Taiwan e-Learning and Digital Archives Program (TELDAP) International Conference*, Taipei, Taiwan, 2011.
- “Introducing an Online Language Learning Environment and its Corpus of Tertiary Student Writing”. *Workshop on Corpora in Teaching Languages and Linguistics*, Humboldt University, Berlin, Germany, 2011.
- “Textual Analysis and Visualization”. *Workshop on Visualization Strategies for Large-scale Heterogeneous Data, 16th International Conference on Virtual Systems and Multimedia*, Seoul, Korea, 2010.
- “Text Visualization: Retrospect and Prospect”. *The Electronic Cultural Atlas Initiative Workshop*, City University of Hong Kong, 2010.

PROFESSIONAL SERVICE

- Reviewer, *IEEE Transactions on Audio, Speech and Language Processing*; *Journal of Artificial Intelligence Research*; *Literary and Linguistic Computing Journal*; *Software: Practice and Experience*
- Reviewer, *Annual Conference of the International Association of Chinese Linguistics (IACL) 2012*
- Scientific Committee Member, *International Conference on Translation Education (ICTE) 2013*
- Steering Committee Member, *Workshop on Greek and Latin in an age of Open Data 2014*
- Program Committee Member, ACL 2013; COLING 2010; EMNLP 2012—2013; *Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages* (2013); *Workshop on Innovative Use of NLP for Building Educational Applications* (2008—2014)