

## CURRICULUM VITAE

### PERSONAL INFORMATION

---

|                                    |   |
|------------------------------------|---|
| Name                               | LEE, John Sie Yuen  |
| Office Address                     | Department of Linguistics and Translation<br>College of Liberal Arts and Social Sciences<br>City University of Hong Kong<br>Kowloon Tong<br>Hong Kong |
| Telephone                          | (852) 3442-5597   |
| Email                              | jsylee@cityu.edu.hk   |
| Research and<br>Teaching Interests | Computational linguistics<br>Computer-assisted language learning<br>Digital humanities  |

### ACADEMIC AND PROFESSIONAL QUALIFICATIONS

---

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)

### EMPLOYMENT HISTORY

---

Assistant Professor, Department of Linguistics and Translation, City University of Hong Kong (2010 – current)

Visiting Research Assistant, Perseus Digital Library, Tufts University, USA (2009)

## **RESEARCH GRANTS**

---

### **External grants**

Innovation and Technology Support Program (ITSP) Tier 3 Project, Innovation and Technology Fund (ITF), “Automatic Generation of Exercises for Learning Chinese”, 2018 – 2019, PI, HK\$1,305,020, On-going

Thomas Nelson, "The NKJV Conversation Bible", 03/2018 – 11/2018, HK\$312,000, On-going

Nanyang Angelz Limited, "Assessment Package for GENUS - English Reading Program for Underprivileged Children", 2017 – 2018, Co-PI, HK\$60,000, On-going

Teaching and Learning Projects (UGC), “Meeting the Challenge of Teaching and Learning Language in the University: Enhancing Linguistic Competence and Performance in English and Chinese”, 2017 – 2020, Co-I, HK\$3,398,646, On-going

Language Fund, Standing Committee on Language Education and Research (SCOLAR), “A Mobile, Personal Reading Tutor for Learners of Chinese”, 01/02/2016 – 31/01/2018, PI, HK\$922,680, Completed

Procore-France/HK Joint Research Scheme (UGC), “Semi-automatic Creation of a Parallel Treebank of Cantonese and Mandarin”, 2016 – 2017, PI, HK\$90,000, Completed

Innovation and Technology Support Program (ITSP) Tier 3 Project, Innovation and Technology Fund (ITF), “An Automatic Text Analysis and Simplification Tool for Learners of English”, 01/12/2015 – 31/05/2017, PI, HK\$726,800, Completed

General Research Fund / Early Career Scheme (UGC), “Semantic Networking and Imagery Configuration: A Computer-aided Study of ‘Poetic Visions’ with Special Reference to Wang Wei and Meng Haoren”, 2015 – 2017, Co-I, HK\$532,860, Completed

General Research Fund / Early Career Scheme (UGC), “A Treebank for the Chinese Buddhist Canon”, 01/11/2012 – 30/04/2015, PI, HK\$746,000, Completed

Germany/HK Joint Research Scheme (UGC), “Textual Revision in Second Language Writing”, 2012 – 2013, PI, HK\$60,000, Completed

### **Internal grants (CityU)**

Strategic Research Grant for unfunded GRF, "Towards Automatic Generation of Exercises for Learning Chinese", 01/09/2017-31/08/2019, PI, HK\$100,000, On-going

Applied Research Grant, “Mobile Learning for Students of English as a Foreign Language”, 01/06/2017 – 31/05/2019, PI, HK\$239,940, On-going

Knowledge Transfer Initiatives in Business, Law and Liberal Arts and Social Sciences Disciplines, “Towards Automatic Translation between Mandarin Chinese and Cantonese”, 01/07/2017 – 30/06/2018, PI, HK\$199,710, On-going

Knowledge Transfer Initiatives in Business, Law and Liberal Arts and Social Sciences Disciplines, “EnGenius – Designing and Implementing a Game-based Language Acquisition and Evaluation Package for Hong Kong Primary School Students”, 01/07/2018 – 30/06/2019, Co-I, HK\$200,000, On-going

Research Grant Writing Fund, “Language and the aging brain: spatial-motional coding, understanding and processing”, 30/06/2016 – 30/06/2017, Co-I, HK\$249,565, Completed

CityU Internal Grant for ITF, “Natural Language Processing for Pedagogical Purposes”, 01/12/2015 – 31/05/2021, PI, HK\$265,020, On-going

Strategic Research Grant for unfunded GRF, “A Parallel Treebank of Cantonese and Mandarin”, 01/09/2016 – 31/08/2018, PI, HK\$100,000, On-going

Applied Research Grant, “Automatic Generation of Personalized Language-Learning Exercises on Mobile Devices”, 01/05/2015 — 30/04/2017, PI, HK\$199,080, Completed

Strategic Research Grant for unfunded GRF, “Toward a Parallel Treebank of Cantonese and Mandarin”, 01/09/2015 – 31/08/2017, PI, HK\$100,000, Completed

Strategic Research Grant for unfunded GRF, “Data-driven Analysis of Syntax and Parallelism in Classical Chinese Poems”, 10/01/2014 – 30/09/2015, PI, HK\$100,000, Completed

Teaching Development Grant, “Data-driven Evaluation of Computer-assisted Translation Systems for Pedagogical Purposes”, 30/06/2014 – 29/06/2015, PI, HK\$180,000, Completed

Strategic Research Grant for unfunded GRF, “Manual and Automatic Creation of Learner Corpora”, 2012 – 2014, PI, HK\$180,000, Completed

Teaching Development Grant, “Evidence-based Learning for Students of Linguistics and Language Technology”, 2011 – 2013, PI, HK\$238,140, Completed

Start-up Grant for New Staff, “Automatic Discovery of Social Networks in Literary Texts”, 2011 – 2012, PI, HK\$100,000, Completed

Strategic Research Grant, “A Reading Environment for Classical Chinese”, 2010 – 2012, PI, HK\$179,326, Completed

Strategic Research Grant, “An Annotated Corpus of Hong Kong ESL Learners”, 2010 – 2011, PI, HK\$90,000, Completed

Interdisciplinary Professional Development Award, “Data is Beautiful: Visualization in the Humanities”, 2011 – 2013, Co-I, HK\$250,000, Completed

## **PUBLICATIONS AND RESEARCH OUTPUT**

---

### **Journal Articles**

1. Wong, T S, & Lee, J (2018). Vernacularization in Medieval Chinese: A Quantitative Study on Classifiers, Demonstratives, and Copulae in the Chinese Buddhist Canon. *Digital Scholarship in the Humanities*. doi:10.1093/llc/fqy012 [SSCI-indexed]
2. Lee, J, & Luo, M (2018). Inducing Word Clusters from Classical Chinese Poems. *International Journal of Asian Language Processing*, 28(1):13-30.
3. Hao, T, Xie, W, & Lee, J. (2018). A Semantic-Context Ranking Approach for Community-Oriented English Lexical Simplification. In X. Huang et al. (Eds.), *NLPCC 2017, Lecture Notes in Artificial Intelligence (LNAI) 10619*, p.1-13.
4. Lee, J, Kong, Y H, & Luo, M (2018). Syntactic Patterns in Classical Chinese Poems: A Quantitative Study. *Digital Scholarship in the Humanities*, 33(1):82-95 [SSCI-indexed]
5. Wong, T S, & Lee, J (2016). Corpus-based Learning of Cantonese for Mandarin Speakers. *ReCALL*, 28(2):187-206 [SSCI-indexed]
6. Lee, J, & Wong, T S (2016). Conversational Network in the Chinese Buddhist Canon. *Open Linguistics*, 2(1).
7. Lee, J, & Kong, Y H (2016). A Dependency Treebank of Chinese Buddhist Texts. *Digital Scholarship in the Humanities*, 31(1):140-151. [SSCI-indexed]
8. Lee, J, Hui, Y C, & Kong, Y H (2016). Knowledge-rich, Computer-assisted Composition of Chinese Couplets. *Digital Scholarship in the Humanities*, 31(1):152-163. [SSCI-indexed]
9. Lee, J, Yeung, C Y, Zeldes, A, Reznicek, M., Lüdeling, A, & Webster, J (2015). CityU Corpus of Essay Drafts of English Language Learners: a Corpus of Textual Revision in Second Language Writing. *Language Resources and Evaluation*, 49(3):659-683 [SCI-indexed]
10. Lee, J (2011). Verb Tense Generation. *Procedia - Social and Behavioral Sciences*, 27:122-130.
11. Webster, J, Chan, A, & Lee, J (2011). Online Language Learning for Addressing Hong Kong Tertiary Students' Needs in Academic Writing. *Asia Pacific World*, 2(2):44-65.
12. Lee, J, & Knutsson, O (2008). The Role of PP Attachment in Preposition Generation. *Lecture Notes in Computer Science*, 4919:643-654.

### **Book Chapters**

1. Lee, J (2015). Uma abordagem do vizinho mais próximo para a análise automática da morfologia do grego antigo. In Ferreira, D'O. A. (org). *Introdução aos textos clássicos na era digital do terceiro milênio*, p.111-128. Araraquara-SP: Letraria.

- Lee, J (2015). Linguistic Annotations for Literary Texts. In Dos Santos, F B, & De Oliveira, J K (eds.), *Estudos Clássicos e seus desdobramentos*, p.207-227. Araraquara, Brazil: Laboratorio Editorial.

## **Refereed conference papers**

### **Top-tier conferences in computational linguistics**

- Lee, J, & Yeung, C Y (2018). Personalizing Lexical Simplification. *Proc. 27<sup>th</sup> International Conference on Computational Linguistics (COLING)*.
- Yeung, C Y, & Lee, J (2018). A Personalized Text Retrieval System for Learners of Chinese as a Foreign Language. *Proc. 27<sup>th</sup> International Conference on Computational Linguistics (COLING)*.
- Yeung, C Y, & Lee, J (2017). Attributing Quotes to Speakers and Listeners in Literary Texts. To appear in *Proc. 8th International Joint Conference on Natural Language Processing (IJCNLP)*.
- Pereira, L, Liu, X, & Lee, J (2017). Lexical Simplification with the Deep Structure Similarity Model. To appear in *Proc. 8th International Joint Conference on Natural Language Processing (IJCNLP)*.
- Lee, J, Liu, M, Lam, C Y, Lau, T O, Li, B, & Li, K. (2017). Automatic Difficulty Assessment for Chinese Texts. *Proc. International Joint Conference on Natural Language Processing: System Demonstrations (IJCNLP)*.
- Lee, J, Sturgeon, D, & Luo, M (2016). A CALL System for Learning Preposition Usage. *Proc. 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Lee, J, & Luo, M (2016). Personalized Exercises for Preposition Learning. *Proc. Demo Session at the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Lee, J, Lam, C Y, & Jiang, S (2016). A Reading Environment for Learners of Chinese as a Foreign Language. *Proc. 26<sup>th</sup> International Conference on Computational Linguistics (COLING): System Demonstrations*.
- Lee, J, Zhao, W, & Xie, W (2016). A Customizable Editor for Text Simplification. *Proc. 26<sup>th</sup> International Conference on Computational Linguistics (COLING): System Demonstrations*.
- Yeung, C Y, & Lee, J (2015). Automatic Detection of Fragments. *Proc. 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Lee, J, & Wong, T S (2012). Glimpses of Ancient China from Classical Chinese Poems. *Proc. 24th International Conference on Computational Linguistics (COLING)*.
- Lee, J, & Webster, J (2012). A Corpus of Textual Revisions in Second Language Writing. *Proc. 50th Annual Meeting of the Association for Computational Linguistics (ACL)*.

13. Lee, J, & Kong Y H (2012). A Dependency Treebank of Classical Chinese Poems. *Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.
14. Lee, J (2011). Toward a Parallel Corpus of Spoken Cantonese and Written Chinese. *Proc. 5th International Joint Conference on Natural Language Processing (IJCNLP)*.
15. Lee, J, Naradowsky, J, & Smith, D A (2011). A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing. *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL)*.
16. Lee, J (2008). A Nearest-Neighbor Approach to the Automatic Analysis of Ancient Greek Morphology. *Proc. Conference on Computational Natural Language Learning (CoNLL)*.
17. Lee, J, & Seneff, S (2008). Correcting Misuse of Verb Forms. *Proc. 46th Annual Meeting of the Association for Computational Linguistics (ACL)*.
18. Lee, J (2007). A Computational Model of Text Reuse in Ancient Literary Texts. *Proc. 45th Annual Meeting of the Association for Computational Linguistics (ACL)*.
19. Sun, G, Liu, X, Cong, G, Zhou, M, Xiong, Z, Lee, J, & Lin, C Y (2007). Detecting Erroneous Sentences using Automatically Mined Sequential Patterns. *Proc. 45th Annual Meeting of the Association for Computational Linguistics (ACL)*.
20. Lee, J, Zhou, M, & Liu, X (2007). Detection of Non-native Sentences using Machine-translated Training Data. *Proc. Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*.
21. Wang, Y Y, Acero, A, Mahajan, M, & Lee, J (2006). Combining Statistical and Knowledge-based Spoken Language Understanding in Conditional Models. *Proc. 44th Annual Meeting of the Association for Computational Linguistics (ACL)*.

### **Other conferences**

1. Lee, J, & Yeung, C Y (2018). Automatic Prediction of Vocabulary Knowledge for Learners of Chinese as a Foreign Language. *Proc. International Conference on Natural Language and Speech Processing (ICNLSP)*.
2. Li, K, & Lee, J (2018). L1-L2 Parallel Treebank of Learner Chinese: Overused and Underused Syntactic Structures. *Proc. Conference on Language Resources and Evaluation (LREC)*.
3. Wong, T S, & Lee, J (2018). Register-sensitive Translation: A Case Study of Mandarin and Cantonese. *Proc. Association for Machine Translation in the Americas (AMTA)*.
4. Jiang, S, & Lee, J (2017). Carrier Sentence Selection for Fill-in-the-blank Items. *Proc. 4th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA)*.

5. Lee, J, Li, K, & Leung, H (2017). L1-L2 Parallel Dependency Treebank as Learner Corpus. *Proc. 15th International Conference on Parsing Technologies.*
6. Lee, J, Pathirage Don, J B K (2017). Splitting Complex English Sentences. *Proc. 15th International Conference on Parsing Technologies.*
7. Jiang, S, & Lee, J (2017). Distractor Generation for Chinese Fill-in-the-blank Items. *Proc. 12th Workshop on Innovative Use of NLP for Building Educational Applications.*
8. Wong, T S, Gerdes, K, Lee, J, & and Leung, H (2017). Quantitative Comparative Syntax on the Cantonese-Mandarin Parallel Dependency Treebank. *Proc. International Conference on Dependency Linguistics.*
9. Lee, J, Leung, H, & Li, K (2017). Towards Universal Dependencies for Learner Chinese. *Proc. Workshop on Universal Dependencies.*
10. Wong, T S, & Lee, J (2017). 國粵平行依存樹庫為本之量化研究. *Proc. 22nd International Conference on Yue Dialects.*
11. Leung, H, Poiret, R, Wong, T S, Chen, X, Gerdes, K, & Lee, J (2016). Developing Universal Dependencies for Mandarin Chinese (2016). *Proc. 12<sup>th</sup> Workshop on Asian Language Resources.*
12. Lee, J, & Luo, M (2016). Word Clustering for Parallelism in Classical Chinese Poems. *Proc. International Conference on Asian Language Processing (IALP).*
13. Lee, J, & Yeung C Y (2016). An Annotated Corpus of Direct Speech. *Proc. Conference on Language Resources and Evaluation (LREC).*
14. Wong, T S, & Lee, J (2016). A Dependency Treebank of the Chinese Buddhist Canon. *Proc. Conference on Language Resources and Evaluation (LREC).*
15. Lee, J, & Wong, T S (2016). Hierarchy of characters in the Chinese Buddhist Canon. *Proc. Florida Artificial Intelligence Research Society (FLAIRS).*
16. Lee, J (2016). Semantic Parallelism in Classical Chinese Poems. *Proc. Florida Artificial Intelligence Research Society (FLAIRS).*
17. Sturgeon, D, & Lee, J (2015). Translation Quality and Effort: Options versus Post-editing. *Proc. 12th International Conference on Natural Language Processing (ICON).*
18. Lee, J, Yeung, C Y, & Chodorow, M (2014). Automatic Detection of Comma Splices. *Proc. 28th Pacific Asia Conference on Language, Information and Computation (PACLIC).*
19. Lee, J, & Kong, Y H (2014). Imagistic and Propositional Languages in Classical Chinese Poetry. *Proc. International Conference on Asian Language Processing (IALP).*
20. Lee, J, Camargo, C, & Kong, Y H (2014). A Corpus of Commented Editions of Literary Texts. *Proc. Workshop on semantic technologies for research in the humanities and social sciences (STRiX).*

21. Sturgeon, D, & Lee, J (2014). Phrase-based alignment of classical Chinese and English. *Proc. Workshop on Greek and Latin in an age of Open Data*.
22. Lee, J, & Wong, T S (2014). Corpus-based Learning of Cantonese for Mandarin Speakers. In S. Jager, L. Bradley, E. J. Meima, & S. Thouèsny (Eds), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference*, Groningen, The Netherlands (pp. 1-6). Dublin: Research-publishing.net.
23. Lee, J, Hui, Y C, & Yeung, C Y (2013). Toward a Digital Library with Search and Visualization Tools. *Proc. 6th Language and Technology Conference*.
24. Lee, J., Hui, Y C, & Kong, Y H (2013). Treebanking for Data-driven Research in the Classroom. *Proc. 4th Workshop on Teaching Natural Language Processing*.
25. Lee, J, & Yeung, C Y (2012). Extracting Networks of People and Places from Literary Texts. *Proc. 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC)*.
26. Lee, J (2012). A Corpus-based Analysis of Mixed Code in Hong Kong Speech. *Proc. International Conference on Asian Language Processing (IALP)*.
27. Lee, J, Wong, S S M, Tang, P K, & Webster, J (2012). A Greek-Chinese Interlinear of the New Testament Gospels. *Proc. Workshop on Language Resources and Evaluation for Religious Texts*.
28. Lee, J (2012). A Classical Chinese Corpus with Nested Part-of-Speech Tags. *Proc. Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*.
29. Lee, J (2010). Dependency Parsing using Prosody Markers from a Parallel Text. *Proc. 9th International Workshop on Treebanks and Linguistic Theories (TLT)*.
30. Lee, J, & Haug, D (2010). Porting an Ancient Greek and Latin Treebank. *Proc. Conference on Language Resources and Evaluation (LREC)*.
31. Lee, J, Tetreault, J, & Chodorow, M (2009). Human Evaluation of Article and Noun Number Usage: Influences of Context and Construction Variability. *Proc. Linguistic Annotation Workshop (LAW)*.
32. Lee, J, & Seneff, S (2008). An Analysis of Grammatical Errors in Non-native Speech in English. *Proc. IEEE Workshop on Spoken Language Technology (SLT)*.
33. Lee, J, & Seneff, S (2007). Automatic Generation of Cloze Items for Prepositions. *Proc. Interspeech*.
34. Lee, J, & Seneff, S (2006). Automatic Grammar Correction for Second-Language Learners. *Proc. Interspeech*.
35. Seneff, S, Wang, C, & Lee, J (2006). Combining Linguistic and Statistical Methods for Bi-directional English-Chinese Translation in the Flight Domain. *Proc. Conference of the Association for Machine Translation in the Americas (AMTA)*.



36. Wang, Y Y, Lee, J, & Acero, A (2006). Speech Utterance Classification Model Training without Manual Transcriptions. *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
37. Lee, J, & Seneff, S (2005). Interlingua-based Translation for Language Learning Systems. *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*.
38. Lee, J, & Seneff, S (2004). Translingual Grammar Induction. *Proc. Interspeech*.
39. Lee, J (2004). Automatic Article Restoration. *Proc. Student Research Workshop at the Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

## COMMUNITY AND PROFESSIONAL SERVICES

| Community / Professional Services  | Role                        | Period                      |
|--|-----------------------------|-----------------------------|
| Botzup Limited   | Technical Adviser           | 2018 – present              |
| <b><u>Top-tier conferences in computational linguistics</u></b>  |                             |                             |
| Empirical Methods in Natural Language Processing (EMNLP) Conference  | Program Committee Member    | 2012 – 2013;<br>2016 – 2018 |
| Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) Annual Conference of the Association for Computational Linguistics (ACL) | Program Committee Member    | 2016; 2018                  |
| Conference on Computational Linguistics (COLING)   | Program Committee Member    | 2013, 2017                  |
| International Joint Conference on Natural Language Processing (IJCNLP)   | Program Committee Member    | 2010                        |
|  |                             | 2017                        |
| <b><u>Other conferences and workshops</u></b>  |                             |                             |
| Workshop on Innovative Use of NLP for Building Educational Applications  | Program Committee Member    | 2008 - 2018                 |
| Workshop on Statistical Parsing of Morphologically Rich Languages  | Program Committee Member    | 2013, 2015                  |
| Linguistic Annotation Workshop (LAW)   | Program Committee Member    | 2015; 2017-2018             |
| International Conference on Corpus Linguistics and Technological Advancement   | Scientific Committee Member | 2015                        |
| Workshop on Greek and Latin in an age of Open Data   | Steering Committee Member   | 2014                        |
| Demonstration Session, Annual Conference of the Association for Computational Linguistics (ACL)  | Program Committee Member    | 2016 - 2018                 |
| Demonstration Session, Empirical Methods in Natural Language Processing (EMNLP) Conference   | Program Committee Member    | 2018                        |

|  |                                |             |
|--|--------------------------------|-------------|
| FLAIRS Conference  | Program Committee<br>Member    | 2017        |
| International Conference on Translation Education (ICTE)   | Scientific Committee<br>Member | 2013        |
| Annual Conference of the International Association of Chinese Linguistics (IACL)                     | Reviewer                       | 2012        |
| Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2016) | Program Committee<br>Member    | 2016 - 2017 |
| Workshop on Corpus-based Research in the Humanities (CRH)  | Program Committee<br>Member    | 2015        |
| Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA)             | Program Committee<br>Member    | 2016 - 2018 |
| China National Conference on Computational Linguistics   | Program Committee<br>Member    | 2017        |