

## CURRICULUM VITAE

### PERSONAL INFORMATION

---

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

---

Associate Professor, Department of Linguistics and Translation, City University of Hong Kong (2019 – current)

Assistant Professor, Department of Linguistics and Translation, City University of Hong Kong (2010 – 2019)

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

### AWARDS AND ACHIEVEMENTS

---

Award	Awarding Body	Year of Award
Certificate of Merit, <i>Excellence in Knowledge Transfer Award</i>	College of Liberal Arts and Social Sciences	2019
Certificate of Merit, <i>Excellence in Knowledge Transfer Award</i>	College of Liberal Arts and Social Sciences	2016

## RESEARCH GRANTS

---

### External grants

Language Fund, Standing Committee on Language Education and Research (SCOLAR), “A Text Difficulty Analysis Tool for Developing Extra-Curricular Reading Materials”, 2021 – 2023, PI, HK\$1,215,840, On-going

General Research Fund (UGC), “Semantic Modeling for Sentence-level Readability Assessment”, 2021 – 2022, PI, HK\$401,565, On-going

Health and Medical Research Fund (HMRF), “CityPal: a chatbot for enhancing young adults' psychological well-being”, 2020 – 2021, PI, HK\$981,000, On-going

Thomas Nelson, "The NKJV Conversation Bible", 03/2018 – 07/2019, HK\$312,000, Completed; Research Matching Grant Scheme (UGC), 2020 – 2021, PI, HK\$133,527, On-going

Innovation and Technology Support Program (ITSP) Tier 3 Project, Innovation and Technology Fund (ITF), “Automatic Generation of Exercises for Learning Chinese”, 2018 – 2020, PI, HK\$1,305,020, Completed

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, Completed

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

Language Fund, Standing Committee on Language Education and Research (SCOLAR), “A Mobile, Personal Reading Tutor for Learners of Chinese”, 2016 – 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”, 2015 – 2017, PI, HK\$726,800, Completed

General Research Fund (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

Early Career Scheme (UGC), “A Treebank for the Chinese Buddhist Canon”, 2012 – 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)*

Applied Research Grant, “Magic Vocabulary Puzzle”: An Innovative and Pedagogical Chinese Characters e-learning Package”, 01/05/2019-30/04/2021, PI, HK\$249,980, On-going

Hong Kong Institute for Data Science, “Towards Linguistically-motivated Text Readability Assessment for Chinese Learning in Hong Kong”, 01/09/2019-31/08/2021, Co-PI, HK\$1,859,420, 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, "Towards Automatic Generation of Exercises for Learning Chinese", 01/09/2017-31/08/2019, PI, HK\$100,000, Completed

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

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, Completed

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, Completed

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

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

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. Lee, J (2021). An Editable Learner Model for Text Recommendation for Language Learning. *ReCALL* [SSCI Q2].
2. Lee, J, & Yeung, C Y (2021). Assisted Discovery-based Learning for Literature Studies. *Innovations in Education and Teaching International* [SSCI Q4].
3. Lee, J, Cheung, L M E, Saberi, D, & Webster, J (2019). Expanding Students’ Registerial Repertoire with a Writing Assistance Tool. *Journal of English for Academic Purposes*, 42. [SSCI Q1: Linguistics, 32/184]
4. Yeung, C Y, & Lee, J (2019). Dialogue Analysis: A Case Study on the New Testament. *Language Resources and Evaluation*. [SCI Q4: Computer Science Interdisciplinary Applications, 98/105]
5. Lee, J, Wong, S S M, Tang, P K, & Webster, J (2018). A Greek-Chinese Interlinear of the New Testament Gospels and its Applications. *Linguistics and the Human Sciences*, 12(1):67-87 [ERIH Plus; ESCI]
6. 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*, 34(1):64–81 [SSCI Q3: Linguistics, 113/181]
7. Lee, J, & Luo, M (2018). Inducing Word Clusters from Classical Chinese Poems. *International Journal of Asian Language Processing*, 28(1):13-30.
8. 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 Q3: Linguistics, 113/181]
9. Wong, T S, & Lee, J (2016). Corpus-based Learning of Cantonese for Mandarin Speakers. *ReCALL*, 28(2):187-206 [SSCI Q1: Linguistics, 13/181; Education & Education Research, 46/238]
10. Lee, J, & Wong, T S (2016). Conversational Network in the Chinese Buddhist Canon. *Open Linguistics*, 2(1). [ERIH Plus; ESCI]

11. Lee, J, & Kong, Y H (2016). A Dependency Treebank of Chinese Buddhist Texts. *Digital Scholarship in the Humanities*, 31(1):140-151. [SSCI Q3: Linguistics, 113/181]
12. 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 Q3: Linguistics, 113/181]
13. 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 Q4: Computer Science Interdisciplinary Applications, 98/105]
14. Lee, J (2011). Verb Tense Generation. *Procedia - Social and Behavioral Sciences*, 27:122-130.
15. 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.
16. 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, & Yeung C Y (2020). Computer-Assisted Learning for Chinese Based on Character Families. In: Lossio-Ventura J., Condori-Fernandez N., Valverde-Rebaza J. (eds) Information Management and Big Data. SIMBig 2019. *Communications in Computer and Information Science*, vol 1070. Springer.
2. Lee, J, Li, K, & Leung, H (2018). 汉语中介语依存关系的标注 [*Dependency Annotation of Learner Chinese*]. In: Kit, C, & Liu, M (eds.) 实证和语料库语言学前沿 [*Frontiers of Empirical and Corpus Linguistics*]. Beijing, China: China Social Sciences Press.
3. Hao, T, Xie, W, & Lee, J (2018). A Semantic-Context Ranking Approach for Community-Oriented English Lexical Simplification. In: Huang, X, Jiang, J, Zhao, D, Feng, Y, Hong, Y (eds.) Natural Language Processing and Chinese Computing. NLPCC 2017. *Lecture Notes in Computer Science*, 10619:784-796.
4. 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.
5. 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

1. Lee, J, Lim, H H, & Webster, C (2021). Paraphrasing Compound Nominalizations. *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
2. Lee, J, Tsou, B, & Cai, T (2020). Using Bilingual Patents for Translation Training. *Proc. 28th International Conference on Computational Linguistics (COLING)*.
3. Saberi, D, Lee, J, & Webster, J (2020). Automatic Assistance for Academic Word Usage. *Proc. 28th International Conference on Computational Linguistics (COLING)*.
4. Lee, J, & Yeung, C Y (2018). Personalizing Lexical Simplification. *Proc. 27th International Conference on Computational Linguistics (COLING)*.
5. Yeung, C Y, & Lee, J (2018). A Personalized Text Retrieval System for Learners of Chinese as a Foreign Language. *Proc. 27th International Conference on Computational Linguistics (COLING)*.
6. Yeung, C Y, & Lee, J (2017). Attributing Quotes to Speakers and Listeners in Literary Texts. *Proc. 8th International Joint Conference on Natural Language Processing (IJCNLP)*.
7. Pereira, L, Liu, X, & Lee, J (2017). Lexical Simplification with the Deep Structure Similarity Model. *Proc. 8th International Joint Conference on Natural Language Processing (IJCNLP)*.
8. 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)*.
9. 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)*.
10. 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)*.
11. 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*.
12. 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*.
13. Yeung, C Y, & Lee, J (2015). Automatic Detection of Fragments. *Proc. 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*.
14. Lee, J, & Wong, T S (2012). Glimpses of Ancient China from Classical Chinese Poems. *Proc. 24th International Conference on Computational Linguistics (COLING)*.

15. 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)*.
16. 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)*.
17. Lee, J (2011). Toward a Parallel Corpus of Spoken Cantonese and Written Chinese. *Proc. 5th International Joint Conference on Natural Language Processing (IJCNLP)*.
18. 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)*.
19. Lee, J (2008). A Nearest-Neighbor Approach to the Automatic Analysis of Ancient Greek Morphology. *Proc. Conference on Computational Natural Language Learning (CoNLL)*.
20. Lee, J, & Seneff, S (2008). Correcting Misuse of Verb Forms. *Proc. 46th Annual Meeting of the Association for Computational Linguistics (ACL)*.
21. Lee, J (2007). A Computational Model of Text Reuse in Ancient Literary Texts. *Proc. 45th Annual Meeting of the Association for Computational Linguistics (ACL)*.
22. 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)*.
23. 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)*.
24. 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. Xie, W, Lee, J, Zhan, F, Han, X, & Chow, C-Y (2021). Unsupervised Adverbial Identification in Modern Chinese Literature. To appear in *Proc. 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL)*.
2. Lee, J, & Yeung, C Y (2021). Text Retrieval for Language Learners: Graded Vocabulary vs. Open Learner Model. *Proc. Recent Advances in Natural Language Processing (RANLP)*.
3. Lee, J, Liang, B, & Fong, H H M (2021). Restatement and Question Generation for Counsellor Chatbot. *Proc. 1st Workshop on NLP for Positive Impact*.

4. Lee, J, & Liang, B (2021). Response Selection for a Virtual Counsellor. *Proc. 9th International Workshop on Natural Language Processing for Social Media*.
5. Yeung, C Y, & Lee, J (2021). Character Set Construction for Chinese Language Learning. *Proc. 16th Workshop on Innovative Use of NLP for Building Educational Applications*.
6. Pilán, I, Lee, J, Yeung, C Y, Webster, J (2020). A Dataset for Investigating the Impact of Feedback on Student Revision Outcome. *Proc. Conference on Language Resources and Evaluation (LREC)*.
7. Yimam, S M, Venkatesh, G, Lee, J, & Biemann, C (2020). Automatic Compilation of Resources for Academic Writing and Evaluating with Informal Word Identification and Paraphrasing System. *Proc. Conference on Language Resources and Evaluation (LREC)*.
8. Lee, J, Liu, M, & Cai, T (2020). Using Verb Frames for Text Difficulty Assessment. *Proc. International FrameNet Workshop*.
9. Lee, J, Cai, T, Xie, W, & Xing, L (2020). A Counselling Corpus in Cantonese. *Proc. Joint Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL) Workshop*.
10. Tsou, B K, Ji, Y, Yip, K-F, Chow, K P, & Lee, J (2020). Bilingual Multi-word Expressions, Multiple-correspondence, and their cultivation from parallel patents: The Chinese-English case. *Proc. PACLIC Workshop on Multiword Expression in Asian Languages*.
11. Yeung, C Y, Lee, J, & Tsou, B (2019). Difficulty-aware Distractor Generation for Gap-Fill Items. *Proc. 17th Annual Workshop of the Australasian Language Technology Association (ALTA)*.
12. Wong, T S, & Lee, J (2019). Character Profiling in Low-Resource Language Documents. *Proc. 24th Australasian Document Computing Symposium (ADCS)*.
13. Lee, J, & Yeung, C Y (2019). Personalized Substitution Ranking for Lexical Simplification. *Proc. 12th International Conference on Natural Language Generation (INLG)*.
14. Saberi, D, & Lee, J (2019). Noun Generation for Nominalization in Academic Writing. *Proc. Workshop on Computational Creativity in Natural Language Generation (CCNLG)*.
15. Yeung, C Y, Lee, J, & Tsou, B (2019). Carrier Sentence Selection with Word and Context Embeddings. *Proc. International Conference on Asian Language Processing (IALP)*.
16. Lee, J (2019). Example Sentence Selection for Feedback on Preposition Usage. *Proc. 13th International Conference on Advances in Semantic Processing (SEMAPRO)*.
17. Wong, T S, & Lee, J (2019). 香港粵語中 '係' 之用法——以依存樹庫為本之量化研究 [The use of haih in Hong Kong Cantonese: A quantitative study based on dependency treebank]. *Proc. 19th Workshop on Cantonese (WOC-19)*.
18. Lee, J (2019). An Emotion Detection System for Cantonese. *Proc. Florida Artificial Intelligence Research Society (FLAIRS)*.

19. Lee, J, Saberi, D, Lam, M, & Webster, J (2018). Assisted Nominalization for Academic English Writing. *Proc. Workshop on Intelligent Interactive Systems and Language Generation*.
20. 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)*.
21. 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)*.
22. 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)*.
23. 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)*.
24. Lee, J, Li, K, & Leung, H (2017). L1-L2 Parallel Dependency Treebank as Learner Corpus. *Proc. 15th International Conference on Parsing Technologies*.
25. Lee, J, Pathirage Don, J B K (2017). Splitting Complex English Sentences. *Proc. 15th International Conference on Parsing Technologies*.
26. 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*.
27. 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*.
28. Lee, J, Leung, H, & Li, K (2017). Towards Universal Dependencies for Learner Chinese. *Proc. Workshop on Universal Dependencies*.
29. 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*.
30. Leung, H, Poiret, R, Wong, T S, Chen, X, Gerdes, K, & Lee, J (2016). Developing Universal Dependencies for Mandarin Chinese. *Proc. 12<sup>th</sup> Workshop on Asian Language Resources*.
31. Lee, J, & Luo, M (2016). Word Clustering for Parallelism in Classical Chinese Poems. *Proc. International Conference on Asian Language Processing (IALP)*.
32. Lee, J, & Yeung C Y (2016). An Annotated Corpus of Direct Speech. *Proc. Conference on Language Resources and Evaluation (LREC)*.
33. Wong, T S, & Lee, J (2016). A Dependency Treebank of the Chinese Buddhist Canon. *Proc. Conference on Language Resources and Evaluation (LREC)*.
34. Lee, J, & Wong, T S (2016). Hierarchy of characters in the Chinese Buddhist Canon. *Proc. Florida Artificial Intelligence Research Society (FLAIRS)*.

35. Lee, J (2016). Semantic Parallelism in Classical Chinese Poems. *Proc. Florida Artificial Intelligence Research Society (FLAIRS)*.
36. Sturgeon, D, & Lee, J (2015). Translation Quality and Effort: Options versus Post-editing. *Proc. 12th International Conference on Natural Language Processing (ICON)*.
37. Lee, J, Yeung, C Y, & Chodorow, M (2014). Automatic Detection of Comma Splices. *Proc. 28th Pacific Asia Conference on Language, Information and Computation (PACLIC)*.
38. Lee, J, & Kong, Y H (2014). Imagistic and Propositional Languages in Classical Chinese Poetry. *Proc. International Conference on Asian Language Processing (IALP)*.
39. 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)*.
40. 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*.
41. Lee, J, & Wong, T S (2014). 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.
42. Lee, J, Hui, Y C, & Yeung, C Y (2013). Toward a Digital Library with Search and Visualization Tools. *Proc. 6th Language and Technology Conference*.
43. 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*.
44. 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)*.
45. Lee, J (2012). A Corpus-based Analysis of Mixed Code in Hong Kong Speech. *Proc. International Conference on Asian Language Processing (IALP)*.
46. 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*.
47. 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)*.
48. Lee, J (2010). Dependency Parsing using Prosody Markers from a Parallel Text. *Proc. 9th International Workshop on Treebanks and Linguistic Theories (TLT)*.
49. Lee, J, & Haug, D (2010). Porting an Ancient Greek and Latin Treebank. *Proc. Conference on Language Resources and Evaluation (LREC)*.

50. 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)*.
51. Lee, J, & Seneff, S (2008). An Analysis of Grammatical Errors in Non-native Speech in English. *Proc. IEEE Workshop on Spoken Language Technology (SLT)*.
52. Lee, J, & Seneff, S (2007). Automatic Generation of Cloze Items for Prepositions. *Proc. Interspeech*.
53. Lee, J, & Seneff, S (2006). Automatic Grammar Correction for Second-Language Learners. *Proc. Interspeech*.
54. 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)*.
55. 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)*.
56. Wang, Y, Lee, J, Mahajan, M, & Acero, A (2005). Statistical Spoken Language Understanding: from Generative Model to Conditional Model. *Proc. NIPS Workshop: Advances in Structured Learning for Text and Speech Processing*.
57. Lee, J, & Seneff, S (2005). Interlingua-based Translation for Language Learning Systems. *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*.
58. Lee, J, & Seneff, S (2004). Translingual Grammar Induction. *Proc. Interspeech*.
59. 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*.

## **Presentations**

1. Lee, J, & Wong, T S (2021). A Treebank of Buddhist Chinese Texts and its Applications. *Fred Jelinek Seminar Series*, Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic.
2. Lee, J, & Liu, M (2020). Automatic Readability Assessment for Chinese Text. Invited speaker at *Data Science Day*, 7 August 2020, City University of Hong Kong.
3. Lee, J, & Chin, A (2019). Online Resources for Cantonese Studies. Invited speaker at *School of Cantonese Studies*, 31 May 2019, The Education University of Hong Kong.
4. Lee, J (2019). Dialogue Analysis: A Case Study on the New Testament. Invited speaker at *Department Seminar Series*, January 2019, Department of English Language and Literature, Hong Kong Baptist University.

5. Lee, J (2018). A Corpus of Textual Revision in Second Language Writing. July 2018, Sun Yat-sen University, Guangzhou, China.
6. Lee, J, Leung, H, & Li, K (2017). Towards a Dependency Treebank of Learner Chinese. Presentation at *Workshop on Learning and Teaching Chinese as a Second Language*, April 2017, City University of Hong Kong, Hong Kong.
7. Lee, J (2017). Syntactic Patterns in Classical Chinese Poems: A Quantitative Study. Presentation at *Linguistic Annotation and Philology Workshop*, July 2017, University of Leipzig, Germany.
8. Lee, J (2016). Toward a Parallel Treebank of Cantonese and Mandarin. Presentation at *2016/17 Linguistics Seminar Series*, November 2016, Education University of Hong Kong, Hong Kong.
9. Lee, J & Wong, T S (2016). “Conversational network: A preliminary study on the Chinese Buddhist Canon”. Presentation at the *8th International Conference on Corpus Linguistics*, March 2016, Malaga, Spain.
10. Lee, J (2015). “Automatic Generation of Language Learning Exercises”. Presentation at *Guangdong University of Foreign Studies*, December 2015, Guangzhou, China.
11. Lee, J (2015). “A Dependency Treebank of the Chinese Buddhist Canon”. Presentation at *Workshop on Treebanking Ancient Languages: Current and Prospective Research*, University of Leipzig, December 2015, Leipzig, Germany.
12. Lee, J, & Wong, T S (2015). “A Dependency Treebank for the Chinese Buddhist Canon”. Invited Speaker, Workshop on Digital Humanities in Chinese at the *International Conference on Corpus Linguistics and Technological Advancement*, December 2015, Hong Kong Institute of Education.
13. Lee, J (2015). “Conversational Network in the Chinese Buddhist Canon”. Presentation at the *ECAI Workshop at the Pacific Neighbourhood Consortium Annual Conference and Joint Meetings (PNC)*, September 2015, University of Macao, Macao.
14. Lee, J, & Wong, T S (2014). “A Treebank for the Chinese Buddhist Canon”. Presentation at the *Electronic Cultural Atlas Initiative (ECAI) Workshop* at the 20th International Conference on Visual Systems and Multimedia (VSMM), December 2014, City University of Hong Kong.
15. Lee, J (2014). “Linguistic Annotations for Literary Studies”. Presentation at *Semana de Estudos Clássicos*, Sao Paulo State University, Araraquara, Brazil, May 2014.
16. Lee, J (2013). “Creating a Treebank for the Chinese Buddhist Canon”. Presentation at *Dharma Drum Buddhist College*, November 2013, Taiwan.
17. Lee, J (2013). “A Corpus Study on Syntactic Parallelism in Classical Chinese Poetry”. Presentation at *North American Conference on Chinese Linguistics (NACCL)*, June 2013, University of Michigan, USA.
18. Lee, J (2013). “Beyond the Concordance: Tools for text analysis and visualization”. Presentation at *Open University of Hong Kong*, January 2013.

19. Lee, J (2012). “Toward a Treebank for the Chinese Buddhist Canon”. Presentation at *Pacific Neighbourhood Consortium Annual Conference and Joint Meetings (PNC)*, December 2012, University of California, Berkeley, USA.
20. Lee, J (2012). “Use of Corpora in Teaching and Research: Three Examples”. Presentation at *NEH Institute for Advanced Technology in the Digital Humanities*, August 2012, Tufts University, USA.
21. Lee, J (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), June 2012, Montreal, Canada,.
22. Lee, J, Webster, J, & Chan, A (2012) “A Corpus of Textual Revisions in Second Language Writing”. Presentation at the *Conference on Language Learning Centers and Internationalization: Supporting Global Connections*, Southwest Association for Language Learning and Technology (SWALLT), April 2012, Arizona State University, Phoenix, USA.
23. Lee, J & Webster, J (2011). “Developing Visual Analytics for Text Data”. Presentation at the *Taiwan e-Learning and Digital Archives Program (TELDAP) International Conference*.
24. Lee, J, Webster, J, & Chan, A (2011). “Introducing an Online Language Learning Environment and its Corpus of Tertiary Student Writing”. Presentation at the *Workshop on Corpora in Teaching Languages and Linguistics*, Humboldt University, January 2011, Berlin, Germany.
25. Lee, J (2010). “Textual Analysis and Visualization”. Presentation at the *Workshop on Visualization Strategies for Large-scale Heterogeneous Data, 16th International Conference on Virtual Systems and Multimedia*, 2010, Seoul, Korea.
26. Lee, J (2010). “Text Visualization: Retrospect and Prospect”. Presentation at *The Electronic Cultural Atlas Initiative Workshop*, 2010, City University of Hong Kong, Hong Kong.

## Patents

1. Lee, J (2020). “Method of Facilitating Natural Language Interactions, a Method of Simplifying an Expression and a System thereof”. U.S. Patent #10,635,862 B2.
2. Wang, Y Y, Lee, J, & Acero, A (2006). “Training system for a speech recognition application”. U.S. Patent #7813926 B2.

## COMMUNITY AND PROFESSIONAL SERVICES

Community / Professional Services	Role	Period
Hong Kong Council for Accreditation of Academic and Vocational Qualifications	Panel Member, Accreditation Exercise	February 2020
Botzup Limited	Technical Adviser	2018 – 2021
<b><u>Conference organization</u></b>		
Workshop on Automatic Readability Assessment	Co-organizer	2021
21 <sup>st</sup> Chinese Lexical Semantics Workshop (CLSW)	Organizing Committee Member	2020
Special Track “Natural Language Processing of Ancient and other Low-resource Languages”, 30th International FLAIRS Conference	Co-organizer	2017
Corpus and Empirical Linguistics Workshop	Co-organizer	2016
<b><u>Top-tier conferences in computational linguistics</u></b>		
Empirical Methods in Natural Language Processing (EMNLP) Conference	Program Committee Member	2012-2013; 2016-2018; 2020
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; 2020
Conference on Computational Linguistics (COLING)	Program Committee Member	2013, 2017, 2019-2021
International Joint Conference on Natural Language Processing (IJCNLP)	Program Committee Member	2010, 2020
Demonstration Session, Annual Conference of the Association for Computational Linguistics (ACL)	Program Committee Member	2017; 2020
Demonstration Session, Empirical Methods in Natural Language Processing (EMNLP) Conference	Program Committee Member	2016-2019
	Program Committee Member	2018-2020
<b><u>Other conferences and workshops</u></b>		
China National Conference on Computational Linguistics, “Linguistics and Cognitive Science” area	Area Chair	2019
Language Resources and Evaluation Conference (LREC)	Program Committee Member	2020
Workshop on Innovative Use of NLP for Building Educational Applications	Program Committee Member	2008 - 2020
Workshop on Statistical Parsing of Morphologically Rich Languages	Program Committee Member	2013, 2015
Linguistic Annotation Workshop (LAW)	Program Committee Member	2015; 2017-2020
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

FLAIRS Conference	Program Committee Member	2017, 2019
International Conference on Translation Education (ICTE)	Scientific Committee 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)	Program Committee Member	2016-2017, 2019-2020
Workshop on Corpus-based Research in the Humanities (CRH)	Program Committee Member	2015
Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA)	Program Committee Member	2016-2018
NLP4CALL Workshop	Program Committee	2018-2019
SyntaxFest	Program Committee	2019
China National Conference on Computational Linguistics	Program Committee Member	2017

***Journal and grant proposal reviewing***

Faculty Development Scheme (FDS), University Grants Committee (UGC), Hong Kong	Reviewer
--	----------

Journals: ACM Transactions on Intelligent Systems and Technology; Applied Linguistics; British Journal of Educational Technology; Computational Intelligence; IEEE Transactions on Audio, Speech and Language Processing; Journal of Information Science; Journal of Artificial Intelligence Research; Literary and Linguistic Computing; Sage Open; Software: Practice and Experience; Studi e Saggi Linguistici; Transactions on Asian and Low-Resource Language Information Processing; WORD Journal

<b>Campus Services at City University of Hong Kong</b>	<b>Role</b>	<b>Period</b>
Research Impact Committee (formerly Knowledge Transfer Committee), College of Liberal Arts and Social Sciences (CLASS)	Member	2016 - 2022
College Board	Member	2021 - 2022
“Smart City” Research Cluster	Member	2020 - current
School of Data Science Undergraduate Committee for the Bachelor of Science in Data Science (BScDS)	Member	2020 - 2022
Departmental Executive Committee	Member	2021 - 2022
Research Degrees Programme Committee (Department of Linguistics and Translation)	Member	2021 - current
Research Assessment Exercise Committee (Department of Linguistics and Translation)	Member	2021 - current
E-learning	Departmental Coordinator	2013 - current

Knowledge Transfer	Departmental Coordinator	2016 - current
Information Day	Departmental Coordinator	2015 - 2016
Gateway Education	Departmental Coordinator	2014 - 2015
Minor in Linguistics and Language Applications (formerly Minor in Language Technology)	Departmental Coordinator	2013 - current
Major in Linguistics and Language Applications	Deputy Admissions Tutor	2014 - current

---