

JACKSON L. LEE

jacksonlunlee@gmail.com · jacksonlee.com · github.com/jacksonlee
Pennsylvania, USA

Last updated: June 5, 2026

Research interests

Computational linguistics, with an empiricist approach toward discovery procedures of linguistic structure. My research is operationalized by asking how children and algorithms learn the structure of language from the data they encounter. Three connected threads:

- **Morphological paradigms.** Paradigm cell alignment; inflection classes; unsupervised paradigm induction; learning from acquisition data.
- **Segmentation.** Stem extraction in paradigms; truncation; morpheme and word segmentation.
- **Morphology–phonology interfaces.** Grapheme-to-phoneme correspondence; tone; reduplication.

Open-source software and data infrastructure — the means by which discovery procedures become reproducible — are treated as first-class research outputs.

Education

- 2022 Ph.D. in Linguistics, University of Chicago.
Dissertation: *Morphological Paradigms: Computational Structure and Unsupervised Learning*.
Adviser: John Goldsmith.
- 2010 M.A. in Linguistics, University of Manchester.
Thesis: *Morphological Doubling and Lexical Phonology: The Case of Cantonese*.
Adviser: Yuni Kim.
- 2008 B.A. in Linguistics and French, University of Hong Kong.

Non-degree programs:

- LSA Summer Institutes (University of California at Berkeley, 2009; University of Colorado at Boulder, 2011; University of Chicago, 2015)
- Undergraduate Exchange Program (Université Paris Dauphine, France, Fall 2007)

Positions

- 2016–present Civis Analytics
 Engineering Manager (2024–present)
 Tech Lead (2022–2024)
 Staff Data Engineer (2021–2022)
 Senior Data Engineer (2020–2021)
 Senior Data Scientist (2019–2020)
 Data Scientist (2016–2019)

Publications

REFEREED ARTICLES

- 2026 Lam, Charles, Chaak-ming Lau, and Jackson L. Lee. To appear. Chinese language corpora. In *International Encyclopedia of Language and Linguistics*, 3rd edition. Edited by Hilary Nesi and Petar Milin. Elsevier.
- 2024 Lam, Charles, Chaak-ming Lau, and Jackson L. Lee. Multi-tiered Cantonese word segmentation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 11993–12002.
- 2023 McCarthy, Arya D., Jackson L. Lee, Alexandra DeLucia, Travis Bartley, Milind Agarwal, Lucas F. E. Ashby, Luca Del Signore, Cameron Gibson, Reuben Raff, and Winston Wu. The SIGMORPHON 2022 shared task on cross-lingual and low-resource grapheme-to-phoneme conversion. In *Proceedings of the 20th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pp. 230–238.
- 2022 Jackson L. Lee, Litong Chen, Charles Lam, Chaak Ming Lau, and Tsz-Him Tsui. PyCantonese: Cantonese linguistics and NLP in Python. In *Proceedings of the 13th Language Resources and Evaluation Conference 2022*, pp. 6607–6611.
- 2020 Jackson L. Lee, Lucas F. E. Ashby, M. Elizabeth Garza, Yeonju Lee-Sikka, Sean Miller, Alan Wong, Arya D. McCarthy, and Kyle Gorman. Massively multilingual pronunciation modeling with WikiPron. In *Proceedings of the 12th Language Resources and Evaluation Conference 2020*, pp. 4223–4228.
- 2018 Pham, Mike, and Jackson L. Lee. Mincing words: Balancing recovery and deletion in word truncation. *Glossa: a journal of general linguistics* 3(1), 36.
- 2017 Goldsmith, John A., Jackson L. Lee, and Aris Xanthos. Computational learning of morphology. *Annual Review of Linguistics* 3, 85–106.
- 2016 Jackson L. Lee and John A. Goldsmith. Linguistica 5: unsupervised learning of linguistic structure. In *Proceedings of NAACL HLT 2016*, pp. 22–26.
- 2015 Beltrama, Andrea, and Jackson L. Lee. Great pizzas, ghost negations: the emergence and persistence of mixed expressives. In *Proceedings of Sinn und Bedeutung 19*, pp. 143–160.
- 2015 Jackson L. Lee and Stephen Matthews. When French becomes tonal: prosodic transfer from L1 Cantonese and L2 English. In *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*, pp. 63–72.
- 2015 Jackson L. Lee. Morphological paradigms: computational structure and unsupervised learning. In *Proceedings of NAACL HLT 2015 Student Research Workshop*, pp. 161–167.
- 2014 Yu, Alan C. L., Hyunjung Lee, and Jackson L. Lee. Variability in perceived duration: pitch dynamics and vowel quality. In *Proceedings of the 4th International Symposium on Tonal Aspects of Languages*, pp. 41–44.
- 2014 Jackson L. Lee. The representation of contour tones in Cantonese. In *Proceedings of the 38th Annual Meeting of the Berkeley Linguistics Society*, pp. 272–287.
- 2012 Jackson L. Lee. Fixed-tone reduplication in Cantonese. In *McGill Working Papers in Linguistics* 22(1). (Proceedings from the Montreal–Ottawa–Toronto Phonology Workshop 2011.)

EDITED VOLUMES

- 2018 Brentari, Diane, and Jackson L. Lee, eds. *Shaping Phonology*. Chicago: University of Chicago Press. (Volume in honor of John Goldsmith.)
- 2014 Beltrama, Andrea, Tasos Chatzikonstantinou, Jackson L. Lee, Mike Pham, and Diane Rak, eds. *Proceedings of the Forty-eighth Annual Meeting of the Chicago Linguistic Society*.

TECHNICAL REPORTS

- 2016 Jackson L. Lee, Ross Burkholder, Gallagher B. Flinn, and Emily R. Coppess. Working with CHAT transcripts in Python. Technical report TR-2016-02, Department of Computer Science, University of Chicago, January 2016.
- 2014 Pham, Mike, and Jackson L. Lee. Combining successor and predecessor frequencies to model truncation in Brazilian Portuguese. Technical report TR-2014-15, Department of Computer Science, University of Chicago, October 2014.
- 2014 Jackson L. Lee. Automatic morphological alignment and clustering. Technical report TR-2014-07, Department of Computer Science, University of Chicago, May 2014.

Open-source research infrastructure

- | | |
|---------------|--|
| WikiPron | A large-scale grapheme-to-phoneme dataset and toolkit covering over 300 languages; basis for the SIGMORPHON 2020–2022 G2P shared tasks.
https://github.com/CUNY-CL/wikipron |
| Linguistica 5 | An unsupervised learner of morphological paradigms.
https://github.com/linguistica-uchicago/lxa5 |
| PyLangAcq | Tooling for language acquisition data from TalkBank and CHILDES.
https://github.com/jacksonlee/pylangacq |
| Rustling | A high-performance library for computational linguistics.
https://github.com/jacksonlee/rustling |
| PyCantonese | A Python library for Cantonese linguistics and NLP.
https://github.com/jacksonlee/pycantonese |

Presentations

INVITED TALKS

- 2015 “PyCantonese: Cantonese linguistic research in the age of big data”. Childhood Bilingualism Research Center, Chinese University of Hong Kong. September 15, 2015.
- 2013 “Automatic morphological alignment and clustering”. Workshop on Language, Cognition, and Computation. University of Chicago. February 22, 2013.
- 2008 “Tone assignment as a strategy in learning French by Cantonese speakers”. Linguistics Department Colloquium. University of Hong Kong. November 21, 2008.

CONFERENCE PRESENTATIONS

- 2024 Lam, Charles, Chaak-ming Lau, and Jackson L. Lee. “Multi-tiered Cantonese word segmentation”. LREC-COLING 2024, Turin, Italy.
- 2022 Jackson L. Lee, Litong Chen, Charles Lam, Chaak Ming Lau, and Tsz-Him Tsui. “PyCantonese: Cantonese linguistics and NLP in Python”. LREC 2022, Marseille, France.
- 2022 Jackson L. Lee. “Working with Cantonese CHILDES data”. The 6th Workshop on Innovations in Cantonese Linguistics (WICL-6).
- 2020 Jackson L. Lee, Lucas F. E. Ashby, M. Elizabeth Garza, Yeonju Lee-Sikka, Sean Miller, Alan Wong, Arya D. McCarthy, and Kyle Gorman. “Massively multilingual pronunciation modeling with WikiPron”. LREC 2020.
- 2016 Jackson L. Lee and John A. Goldsmith. “Linguistica 5: Unsupervised Learning of Linguistic Structure”. NAACL-HLT 2016. San Diego. June 12–17, 2016.
- 2016 Jackson L. Lee, Litong Chen, and Tsz-Him Tsui. “PyCantonese: Developing computational tools for Cantonese linguistics”. The 3rd Workshop on Innovations in Cantonese Linguistics. The Ohio State University. March 12–13, 2016.
- 2015 Jackson L. Lee, John Goldsmith, and Simon Jacobs. “Learning linguistic structure: Integrating big data analysis and visualization”. Mind Bytes: Research Computing Expo and Symposium. The Research Computing Center, University of Chicago. October 27, 2015.
- 2015 Pham, Mike, and Jackson L. Lee. “Making short cuts at the intersection of recoverability and deletability to truncate words”. American International Morphology Meeting 3, University of Massachusetts Amherst. October 2–4, 2015.
- 2015 Jackson L. Lee. “Morphological paradigms: Computational structure and unsupervised learning”. NAACL 2015 Student Research Workshop, Denver, Colorado. May 31–June 5, 2015.
- 2015 Jackson L. Lee and John Goldsmith. “Solving the paradigmatic alignment problem in unsupervised morphological paradigm induction”. Midwest Speech and Language Days 2015, Toyota Technological Institute at Chicago. May 7–8, 2015.
- 2015 Matthews, Stephen, and Jackson L. Lee. “Cantonese tones are tone clusters: evidence from pop music”. HKU Tone Conference, University of Hong Kong. January 17, 2015.
- 2015 Pham, Mike, and Jackson L. Lee. “Combining successor and predecessor frequencies to model truncation in Brazilian Portuguese”. The 89th Annual Meeting of the Linguistic Society of America, Portland, Oregon. January 8–11, 2015.
- 2015 Matthews, Stephen, and Jackson L. Lee. “The representation of Cantonese tone: evidence from music”. The 89th Annual Meeting of the Linguistic Society of America, Portland, Oregon. January 8–11, 2015.
- 2014 Jackson L. Lee and John Goldsmith. “Towards a fully unsupervised approach to learning morphological paradigms from a raw text”. The 11th High Desert Linguistics Society, University of New Mexico. November 13–15, 2014.
- 2014 Jackson L. Lee, John Goldsmith, and Simon Jacobs. “Learning natural language morphology from a raw text”. Mind Bytes: Research Computing Expo and Symposium. The Research Computing Center, University of Chicago. October 28, 2014.
- 2014 Beltrama, Andrea, and Jackson L. Lee. “Great pizzas, ghost negations: The emergence and persistence of mixed expressives”. The 19th Annual Meeting of Sinn und Bedeutung, Georg August University at Göttingen. September 15–17, 2014.

- 2014 Jackson L. Lee and Stephen Matthews. “When French becomes tonal: prosodic transfer of L1 Cantonese speakers”. The 6th Annual Conference of the Pronunciation in Second Language Learning and Teaching, University of California, Santa Barbara. September 5–6, 2014.
- 2014 Jackson L. Lee and John Goldsmith. “Complexity across morphological paradigms: A minimum description length approach to identifying inflectional stems”. MorphologyFest: Symposium on Morphological Complexity, Indiana University, Bloomington. June 16–20, 2014.
- 2014 Yu, Alan C. L., Hyunjung Lee, and Jackson L. Lee. “Variability in perceived duration: pitch dynamics and vowel quality”. The 4th International Symposium on Tonal Aspects of Languages, Nijmegen, the Netherlands. May 13–16, 2014.
- 2014 Pham, Mike, Jackson L. Lee, and Bruna da Costa Moreira. “Modeling Brazilian Portuguese truncation by successor and predecessor frequencies”. Midwest Speech and Language Days 2014, University of Illinois at Urbana-Champaign. May 2–3, 2014.
- 2014 Jackson L. Lee and John Goldsmith. “Algorithmic approaches to identifying inflectional stems: implications from Spanish”. Linguistic Symposium on Romance Languages, Western University, London, Ontario. May 2–4, 2014.
- 2014 Jackson L. Lee and John Goldsmith. “Algorithmic approaches to identifying inflectional stems”. The 6th annual Illinois Language and Linguistics Society conference, University of Illinois at Urbana-Champaign. April 4–6, 2014.
- 2014 Beltrama, Andrea, and Jackson L. Lee. “Big pizzas, ghost negations: The emergence and persistence of mixed expressives”. The 38th Penn Linguistics Conference, University of Pennsylvania. March 28–30, 2014.
- 2013 Jackson L. Lee and John Goldsmith. “Automatic morphological alignment and clustering”. American International Morphology Meeting 2, University of California, San Diego. November 8–10, 2013.
- 2013 Jackson L. Lee. “NP ellipsis may not be licensed non-locally”. The 87th Annual Meeting of the Linguistic Society of America, Boston, Massachusetts. January 4, 2013.
- 2012 Jackson L. Lee. “NP ellipsis may not be licensed by Agree”. ELLIPSIS2012: Crosslinguistic, formal, semantic, discursive and processing perspectives, Universidade de Vigo, Spain. November 9–10, 2012.
- 2012 Jackson L. Lee. “The representation of contour tones in Cantonese”. The 36th Penn Linguistics Colloquium, University of Pennsylvania. March 23–25, 2012.
- 2012 Jackson L. Lee. “The representation of contour tones in Cantonese”. Workshop on Innovations in Cantonese Linguistics, Ohio State University. March 16–17, 2012.
- 2012 Jackson L. Lee. “The representation of contour tones in Cantonese”. The 38th Annual Meeting of the Berkeley Linguistics Society. February 11–12, 2012.
- 2012 Sonderegger, Morgan, Andrea Beltrama, Tasos Chatzikonstantinou, Erin Franklin, Brett Kirken, Jackson L. Lee, Maria Nelson, Krista Nicoletto, Talia Penslar, Hannah Provenza, Natalie Rothfels, Maximilian Bane, Peter Graff, and Jason Riggle. “Coronal stop deletion on reality TV”. The Annual Meeting of the Linguistic Society of America, Portland, Oregon. January 6, 2012.
- 2011 Jackson L. Lee. “Fixed-tone reduplication in Cantonese”. The Annual Research Forum of the Linguistic Society of Hong Kong. The City University of Hong Kong. December 3, 2011.

- 2011 Sonderegger, Morgan, Andrea Beltrama, Tasos Chatzikonstantinou, Erin Franklin, Brett Kirken, Jackson L. Lee, Maria Nelson, Krista Nicoletto, Talia Penslar, Hannah Provenza, Natalie Rothfels, Maximilian Bane, Peter Graff, and Jason Riggle. "Coronal stop deletion on reality TV". The 40th New Ways of Analyzing Variation (NWAV 40). Georgetown University, Washington DC. October 30, 2011.
- 2011 Jackson L. Lee. "The (non-)blocking of non-TETU tonal overwriting in Cantonese attenuative reduplication". Phonology in the 21st Century: In Honour of Glyne Piggott. McGill University, Montréal, Canada. May 8, 2011.

Awards and fellowships

- 2015 The Mellon Foundation–University of Chicago Dissertation Completion Fellowship.
- 2014 The Research Computing Center Mind Bytes Award for Novel Use of Research Resources, University of Chicago.
- 2014 Top abstract, the 6th Annual Meeting of the Illinois Language and Linguistics Society, University of Illinois at Urbana-Champaign.
- 2010–2015 University of Chicago Fellowship.
- 2011 Linguistic Society of America Summer Institute Fellowship, University of Colorado at Boulder.
- 2006–2008 Dean's Honors Lists, University of Hong Kong.
- 2007–2008 HSBC Hong Kong Scholarships, University of Hong Kong.
- 2008 Sir Man Kam Lo / Jardine Scholarship, University of Hong Kong.
- 2007 The University of Hong Kong Foundation Scholarship; CV Starr Scholarship; Overseas Language Course Study Grant; Dragages Prize in French.

Grants

- 2014 Rella Cohn research funds, Department of Linguistics, University of Chicago.
- 2011, 2014 Graduate Student Travel Grant, Center for East Asian Studies, University of Chicago.
- 2012 Graduate Student Travel Grant, Confucius Institute, University of Chicago.

Teaching

INSTRUCTORSHIPS

University of Chicago:

- Spring 2015 LING 21000 Morphology.

TEACHING ASSISTANTSHIPS

University of Chicago:

- Summer 2015 Computational Linguistics (Sharon Goldwater) at the LSA Summer Institute.
- Spring 2014 LING 20001/30001 Introduction to Linguistics (Tim Grinsell).
- Winter 2014 LING 20101/30010 Introduction to Phonetics and Phonology (Jason Riggle).
- Winter 2013 LING 21000/31000 Morphology (John Goldsmith).

University of Hong Kong:

Winter 2009 LING 2009 Languages of the World; LING 2012 Experimental Phonetics.
 Fall 2008 LING 2004 General Phonetics.

Research assistantships

2011–2014 PI: Alan Yu (University of Chicago).
 “Understanding perceptual compensation in sound change” (NSF #0949754).
 Designed and constructed Mechanical Turk experiments; Praat and Python scripting for audio and data processing.

Summer 2012 PI: John Goldsmith (University of Chicago).
Linguistica and computational morphology.
 Computational modeling of structure across morphological paradigms.

2008–2009 PIs: Stephen Matthews (University of Hong Kong), Virginia Yip (Chinese University of Hong Kong).
 “Rethinking Cantonese grammar: typology, processing, and acquisition” (HKU 748207H).
 Assisted with the revision of *Cantonese: A Comprehensive Grammar* (Routledge, 2011).

Service

SHARED TASK AND WORKSHOP ORGANIZATION

2022 SIGMORPHON 2022 Shared Task on Cross-lingual and Low-Resource Grapheme-to-Phoneme Conversion, co-located with NAACL 2022.

2022 Analyzing Cantonese Corpus Data, at the 6th Workshop on Innovations in Cantonese Linguistics.

2015 Workshop on Computational Phonology and Morphology, LSA Summer Institute, University of Chicago.

2013–2016 Workshop on Language, Cognition, and Computation, University of Chicago.

2012 The 48th Annual Meeting of the Chicago Linguistic Society.

REVIEWING FOR JOURNALS AND CONFERENCES

- *Computational Linguistics* (2016)
- *Phonology* (2016)
- *Language and Cognition* (2026)
- *Language Resources and Evaluation* (2022, 2024)
- *ACM Transactions on Asian and Low-Resource Language Information Processing* (2014, 2015)
- Association for Computational Linguistics (2023)
- Workshop on Dataset Creation for Low-Resource Languages at LREC 2022
- Chicago Linguistic Society (2011–2016)

Professional affiliations

2013–present Association for Computational Linguistics.
2010–present Linguistic Society of America.