

# CHEONKAM JEONG

 philhelenina.github.io |  cheonkaj@hs.uci.edu |  Google Scholar |  LinkedIn

## EDUCATION

<b>The University of Arizona</b> <i>Ph.D. in Linguistics</i>	Tucson, AZ 2018–2024
• Dissertation: The Effects of Lexical Competition on Hyperarticulation of Voice Onset Time and Fundamental Frequency in Korean Stops: Experimental and Big Data Approaches (Advisor: Dr. Andrew Wedel)	
• Research areas: Phonology, Computational Linguistics	
<b>The University of Arizona</b> <i>M.S. in Human Language Technology</i>	Tucson, AZ 2020–2021
• Internship Report: A semantic analysis of FindingFive support tickets using machine learning-based classifiers (Advisor: Dr. Andrew Wedel)	
• Research areas: Natural Language Processing, Machine Learning	
<b>Hankuk (Korea) University of Foreign Studies (HUFS)</b> <i>M.A. in English Linguistics</i>	Seoul, South Korea 2015–2017
• Thesis: Noisy Harmonic Grammar Modeling of Flapping in American English Based on a Statistical Analysis (Advisor: Dr. Sung-Hoon Hong)	
• Research areas: Phonology, Corpus Linguistics	
<b>Hankuk (Korea) University of Foreign Studies (HUFS)</b> <i>B.A. in English Linguistics</i>	Seoul, South Korea 2010–2015
• Research areas: English Linguistics, Philosophy	
<b>University of Leicester</b> <i>Visiting Student Program</i>	Leicestershire, United Kingdom Fall 2013
• Department of English Literature (13th representative scholar sponsored by MIRAE ASSET, a South Korean financial services company)	

## PROFESSIONAL EXPERIENCE

<b>Postdoctoral Scholar - AI/ML Lead</b> <i>Irvine, United States</i>	Sue & Bill Gross School of Nursing, University of California, Irvine 2024–Present
• Research areas: Healthcare AI/ML, NLP, Reinforcement Learning	
<b>Visiting Scholar</b> <i>West Lafayette, United States</i>	Department of Computer Science, Purdue University 2024–Present
• Research areas: NLP, Reinforcement Learning	

## PUBLICATIONS

- Pyarelal, A., Culnan, J. M., Qamar, A., Krishnaswamy, M., Wang, Y., Jeong, C., Chen, C., Miah, M. M., Hormozi, S., Tong, J., & Huang, R. (2025). MultiCAT: Multimodal communication annotations for teams. In *Proceedings of the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. Association for Computational Linguistics (ACL). Proceedings available here.
- Nyamathi, A., Dutt, N., Lee, J.-A., Rahmani, A. M., Rasouli, M., Krogh, D., Krogh, E., Sultzer, D., Rashid, H., Liqat, H., Jawad, R., Azhar, F., Ahmad, A., Qamar, B., Bhatti, T. Y., Khay, C., Ludlow, J., Gibbs, L., Rousseau, J., Abbasian, M., Song, Y., Jeong, C., & Brunswicker, S. (2024). Establishing the foundations of emotional intelligence in care companion robots to mitigate agitation among high-risk patients with dementia: Protocol for an empathetic patient-robot interaction study. *JMIR Research Protocols*. Paper available here.
- Dumitru, R., Alexeeva, M., Alcock, K., Ludgate, N., Jeong, C., Abdurahaman, Z. F., Puri, P., Krichhoff, B., Sadhu, S., & Surdeanu, M. (2024). Retrieval augmented generation of subjective explanations for

socioeconomic scenarios. In *Proceedings of the Sixth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, pages 68–85, Mexico City, Mexico. Association for Computational Linguistics (ACL). Proceedings available here.

4. Jeong, C., & Wedel, A. (2024). Applying big data and automation techniques in phonetics: A case study on hyperarticulation in Korean word-initial stops. In *Proceedings of the 19th Conference on Laboratory Phonology (LabPhon 19)*. Hanyang University, Seoul, South Korea. Proceedings available here.
5. Kwak, A., Jeong, C., Vincent Forte, G., Bambauer, D., Morrison, C. T., & Surdeanu, M. (2023). Information extraction from legal wills: How well does GPT-4 do? In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore. Association for Computational Linguistics (ACL). Proceedings available here.
6. Jeong, C., Schmitz, D., Kakolu Ramarao, A., Stein, A., & Tang, K. (2023). Linear discriminative learning: A competitive non-neural baseline for morphological inflection. In *Proceedings of the 20th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 138-150, Toronto, Canada. Association for Computational Linguistics (ACL). Proceedings available here.
7. Jeong, C., & Wedel, A. (2023). The effect of cue-specific lexical competitors on hyperarticulation of VOT and F0 contrasts in Korean stops. In *Proceedings of the 2022 Annual Meeting on Phonology (AMP)*, Los Angeles, California. Proceedings available here.
8. Jeong, C., & Wedel, A. (2023). The effect of cue-specific lexical competitors on hyperarticulation of VOT and F0 contrasts in Korean stops. Poster presented at *Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2023 (HISPhonCog 2023)*. Hanyang University, Seoul, South Korea.
9. Data Study Group Team. (2023). Data Study Group final report: Global Witness - Identifying unregulated mining sites using historical satellite data (Version 1). *Zenodo*. Paper available here.
10. Jeong, C., & Park, S. (2018). A corpus-based study on the prosodic features of *com* in Korean. Poster presented at *92nd Annual Meeting of Linguistic Society of America (LSA)*, Salt Lake City, Utah. Proceedings available here.
11. Jeong, C., & Park, S. (2017). The role of prosody in discourse: A case study of Korean *com*. In *Proceedings of 2017 Seoul International Conference on Speech Sciences (SICSS)*, pages 109-110, Seoul, South Korea. Proceedings available here.
12. Jeong, C., & Hong, S. (2017). Corpus-based noisy harmonic grammar modeling of word-medial /t/-flapping in American English. *Studies in Phonetics, Phonology and Morphology*, 23(1), 117-143. Paper available here.

## MANUSCRIPTS

1. Jeong, C., Kim, S., & Park, J. (Under review). Beyond BPM Automation: Verifiable Neuro-Symbolic Agents with Montague Semantics. *The 41st ACM/SIGAPP Symposium On Applied Computing (SAC 2026)*.
2. Nyamathi, A., Rasouli, E., Jeong, C., Brunswicker, S., Agapie, E., Ludlow, D., Jawad, A., Khay, N., Lu, L., Ding, C., Liao, F., Farahani, F., Lee, J., Dutt, A., Rahmani, A., & Krogh, P. (Under review). Evaluation of personas and daily activity maps for improved clinical care and training models for effective human-robot interaction for persons with dementia: A qualitative study. *Journal of Medical Internet Research (JMIR)*. Preprint available here.
3. Jeong, C., & Wedel, A. (Under second-round review). The effect of contrast-specific minimal pair competitor in hyperarticulation of VOT and F0 phonetic cues in Korean initial stops in tonogenetic sound change. *Journal of Phonetics*. PHONETICS-D-24-00131.
4. (Preprint) Jeong, C., Kim, S., & Park, J. (2025). The algebra of meaning: Why machines need Montague more than Moore's Law. *arXiv preprint*, Paper available here
5. (Preprint) Kwak, A., Jeong, C., Lim, J., & Min, B. (2024). A Korean legal judgment prediction dataset for insurance disputes. *arXiv preprint*, Paper available here.
6. (Preprint) Yoo, Y., Jeong, C., Gim, S., Lee, J., Schimeke, Z., & Seo, D. (2023). A novel patent similarity measurement methodology: Semantic distance and technological distance. *arXiv preprint*, Paper available here.

## INVITED TALKS

**Jeong, C.** (2025, May 20). Introduction to NLP and its Applications in Healthcare. *Sue & Bill Gross School of Nursing, University of California, Irvine*. Irvine, California, United States. [Slides](#)

**Jeong, C.**, Darling, M., & Dely, A. (2023). Enterprise AI: How are you addressing the new wave of computing? *Tucson's IT Networking Group LAMBDA*. Tucson, Arizona, United States.

**Jeong, C.**, & Romero, D. (2022). Student internships in industry: Experiences and resources. *The Arizona Research Bazaar 2022*. Tucson, Arizona, United States. [Slides](#)

**Jeong, C.** (2022). 2020 NLP internship at FindingFive. *Master of Science in Human Language Technology*, The University of Arizona. Tucson, Arizona, United States.

## PROJECTS

### Emotional Intelligence for Empathetic Patient–Robot Interaction (Dementia Care)

*UCI Noyce Initiative*

June 2024–Present

- Developing real-time models to forecast agitation/gait and infer emotional state from multimodal streams (audio, video, wearable signals).
- Designing an empathy-focused conversational agent that conditions on personal history and prosodic cues for de-escalation.
- Collecting and curating synchronized visual, audio, and physiological data using Personicle; implementing reproducible analysis pipelines.

### Global Witness: Identifying Unregulated Mining Sites from Historical Satellite Data

*The Alan Turing Institute, United Kingdom*

March 2023

- Performed satellite image augmentation and quality-controlled to improve data diversity and signal consistency.
- Trained and evaluated a Random Forest baseline over temporal composites to flag likely unregulated sites.

### Korean Education Dataset for Learners of Korean as a Foreign Language

*Ministry of Science and ICT & National Information Society Agency (Korea)*

Oct

2022–May 2023

- Audited and corrected prosodic/phonological tags; established reviewer checklists for consistency.
- Evaluated learner productions and documented recurrent error patterns to inform rubric refinements.

### Subtlex-KR: A Korean Subtitle Corpus

*Collaboration with Dr. Kevin Tang, Heinrich Heine University Düsseldorf (Dept. of English Language & Linguistics)*

May 2019–Present

- Build a large lemmatized subtitle corpus; generate morpheme-analyzed files with KoNLPy.
- Implement alignment tools (tgt for TextGrid, Parselmouth for Praat) and corpus cleaning scripts.
- Ongoing: analyze probabilistic reduction with informativity-based predictors.

## RESEARCH EXPERIENCE

- NLP Engineer | FindingFive, New Jersey** June–August 2020
- [Python · Zoho Desk API] Built an automated pipeline to collect/clean and categorize support tickets; established basic analytics reporting.
  - [Classical ML + Neural] Prototyped priority/triage classifiers (handoff/reassignment signals) and set baseline benchmarks.
  - Closed the loop with operations via lightweight dashboards and feedback reports.
- Research Associate (NLP Engineer / Developer) | School of Information & Dept. of Computer Science, The University of Arizona** 2021–2024
- [Legal IE] Defined annotation guidelines (entities, relations,  $n$ -ary events) and quality-control procedures for a wills IE dataset (*Findings of EMNLP 2023*).
  - [Docker Compose · PostgreSQL · Prodigy] Deployed a reproducible annotation infrastructure (containers, DB schema, deployment scripts).
  - [Multimodal: Text+Audio] Assembled LSTM/CNN/DistilBERT multitask experiments; collaborated on unsupervised vocal entrainment detection (*Findings of NAACL 2025*).
  - Reviewed LLM-generated scenarios and labeled causal-question datasets; supported data-engineering pipelines (*NLP+CSS 2024*).

- Research Associate (Phonetician) | Douglass Phonetics Laboratory, The University of Arizona** Spring 2021
- [FindingFive · MTurk] Designed and launched an online speech experiment with attention/quality checks; performed data cleaning and documentation.
  - Delivered reproducible analysis scripts and procedure notes for follow-up studies.

## TEACHING EXPERIENCE

- Instructor | Department of Linguistics** 2019–2021
- LING 341: Language Development (Psycholinguistics) Winter 2021
  - LING 314: Phonetics Spring 2020, Fall 2019, Fall 2020
- Teaching Assistant/Associate | Department of Linguistics & College of Social and Behavioral Sciences** 2018–2022
- SBS 200: Introduction to Statistics for the Social Sciences Spring 2022, Fall 2021
  - LING 315: Introduction to Phonology Spring 2019
  - LING 341: Language Development Spring 2019
  - LING 150: Language in the World Fall 2018

## GRANTS, AWARDS, AND CERTIFICATES

- Saturn Cloud – NVIDIA Academic Grants**  
*Purdue University* April–December 2025
- Computational Social Sciences Certificate**  
*College of Social & Behavioral Sciences, The University of Arizona* Spring 2022
- Certificate in Essay Writing**  
*Dept. of Philosophy, HUFS* 2015
- MIRAE ASSET Student Exchange Scholarship**  
*MIRAE ASSET, South Korea* 2013
- Second Distinction (Dean's recognition)**  
*Dept. of English Linguistics, HUFS* 2011

## SKILLS

**Languages:** English, Korean, Japanese, French.  
**Programming:** Python, R, Praat/Parselmouth, SQL, Prolog, Perl; Docker, PostgreSQL, KoNLPy

ACADEMIC  
SERVICES AND  
VOLUNTEER-  
ING

Paper Reviewer for:

- IEEE Access: Specializing in NLP Papers July 2023–Present
- Arizona Linguistics Circle 16: Phonetics & Statistical Modeling Proceedings Reviewer 2022
- Journal of Phonetics: Co-reviewer of PHONETICS-D-21-00128 with Dr. Andrew Wedel 2022
- Arizona Linguistics Circle 14, 15: Phonology & Phonetics Coyote Paper Reviewer 2021–2022
- Arizona Linguistics Circle 15: Phonetics & Corpus Linguistics Proceedings Reviewer 2021
- Arizona Linguistics Circle 14: Phonology & Phonetics Proceedings Reviewer 2020
- Arizona Linguistics Circle 13: Phonology Proceedings Reviewer 2019

Leadership and Communication:

- NAVER Boost Course: Coaching Study AI Basics 2023 - Team Leader January 2023
- FindingFive: Outreach and Marketing Coordinator March–August 2022
- Arizona CatCloud Service: Helped design the service with fellow student leaders.
- Arizona Linguistics Circle 15: Communication Coordinator 2021
- Arizona Linguistics Circle 13: Phonetics Section Chair 2019
- Education Supporter at Seoul Urban Science Technical High School in Seoul, South Korea 2013

REFERENCES

Research Advisors and Collaborators

- Dr. Andrew Wedel — Professor, Linguistics, The University of Arizona  
*wedel@arizona.edu*
- Dr. Adey Nyamathi — Distinguished Professor and Founding Dean, School of Nursing, University of California, Irvine  
*anyamath@uci.edu*
- Dr. Michael Hammond — Professor, Linguistics & HLT, The University of Arizona  
*hammond@arizona.edu*
- Dr. Natasha Warner — Professor & Dept. Head; Director, Douglass Phonetics Lab, The University of Arizona  
*nwarner@arizona.edu*
- Dr. Miquel Simonet — Professor, Spanish & Portuguese; Cognitive Science, The University of Arizona  
*simonet@arizona.edu*
- Dr. Sandiway Fong — Associate Professor, Linguistics & Computer Science, The University of Arizona  
*sandiway@arizona.edu*
- Dr. Gus Hahn-Powell — Assistant Professor, Linguistics; Director, HLT Online MS & NLP Certificate, The University of Arizona  
*habnpowell@arizona.edu*
- Dr. Mihai Surdeanu — Associate Professor, Computer Science, The University of Arizona  
*msurdeanu@arizona.edu*
- Dr. Clayton Morrison — Associate Professor, School of Information, The University of Arizona  
*claytonm@arizona.edu*
- Dr. Adarsh Pyarelal — Assistant Professor, School of Information, The University of Arizona  
*adarsh@arizona.edu*

- Dr. Kevin Tang — University Professor, English Linguistics, Heinrich Heine University Düsseldorf (Germany)  
*kevin.tang@hhu.de*

## Industry Collaborators

- Dr. Ting Qian — President; VP of Engineering, FindingFive  
*ting.qian@findingfive.com*

## Teaching References

- Dr. Cecile McKee — Professor, Linguistics, The University of Arizona  
*mckee@arizona.edu*
- Dr. Diane Ohala — Associate Professor of Practice, Linguistics, The University of Arizona  
*ohalad@arizona.edu*
- Dr. Amy Fountain — Associate Professor of Practice, Linguistics, The University of Arizona  
*avf@arizona.edu*
- Dr. Suzanne Delaney — Associate Professor of Practice, Social & Behavioral Sciences, The University of Arizona  
*delaney@arizona.edu*