NAJOUNG KIM

najoung@bu.edu / najoung.kim
Office 808, 665 Commonwealth Ave.
Boston, MA, USA 02215

Updated March 2024

ACADEMIC POSITIONS

Boston University 2023–current
Assistant Professor, Department of Linguistics
Affiliated Faculty, Department of Computer Science
Faculty Fellow, Center for Data Science Until 2026

New York University 2021–2022
Faculty Fellow/Assistant Professor, Center for Data Science

EDUCATION

Johns Hopkins University 2021
Ph.D. Cognitive Science
Advisors: Paul Smolensky and Kyle Rawlins

Johns Hopkins University 2017
M.A. Cognitive Science

University of Oxford 2015
M.St. General Linguistics & Comparative Philology Distinction

Seoul National University 2014
B.A. English Linguistics & Literature Summa Cum Laude
B.A. Linguistics (double major)
Minor: Computer Science & Engineering

EXPERIENCE

Visiting Faculty Researcher, Google Research 2022–current
Host: Deepak Ramachandran
Reasoning and abstraction in large language models
Evaluation of large multimodal models

Student researcher, Google AI Aug–Nov 2020
Research intern, Google AI May–Aug 2020
Hosts: Deepak Ramachandran, Ellie Pavlick
Improving Question-Answering systems with presupposition verification

Research intern, IBM Research AI May–Aug 2019
Hosts: Song Feng, Chulaka Gunasekara
Evaluating sentence representations for discourse

Graduate student researcher, Jun–Aug 2018
Jelinek Memorial Summer Workshop (JSALT)
PIs: Sam Bowman, Ellie Pavlick
General-purpose Sentence Representation Learning team

Visiting researcher, 2015–2016
Korea Advanced Institute of Science and Technology (KAIST)

PI: Jong C. Park
Computational modeling of semantic fluency

Intern developer, NAVER Corporation 2013–2014
Knowledge Extraction Team

AWARDS

CDS Faculty Fellowship 2023–2026
Center for Data Science, Boston University ($50,000)

Area Chair Award (Interpretability and Analysis Track), ACL 2023 2023
For *Entity Tracking in Language Models*

NSF Doctoral Dissertation Research Improvement Grant 2021–2022
Title: *Compositional Linguistic Generalization in Human and Machine Learning*
BCS-2041221 ($12,771) with Paul Smolensky, Geraldine Legendre & Tal Linzen.

NeurIPS 2019 Travel Award 2019

Best Paper Award, *SEM 2019 2019
For *Probing what different NLP tasks teach machines about function word comprehension*

Owen Scholarship 2016–2019
Awarded to select incoming doctoral students in natural sciences ($18,000)

Mica and Ahmet Ertegun Graduate Scholarship Programme in the Humanities 2014–2015
Full scholarship for Master's degree at University of Oxford (∼$42,000)

SNU Undergraduate Research Program 2013
Grant for select undergraduate research proposals (∼$3000)

SNU Foundation Funds Scholarship 2013
SNU Superior Academic Performance Scholarship 2010–2011

PREPRINTS

Ashwin Daswani, Rohan Sawant, and **Najoung Kim** (2024). Syn-$(QA)^2$: Evaluating False Assumptions in Long-tail Questions with Synthetic QA Datasets. In submission.

Najoung Kim and Paul Smolensky (2024). Structural Generalization of Modification in Adult Learners of Artificial Language. In submission.

Nitish Joshi, Javier Rando, Abulhair Saparov, **Najoung Kim**, and He He (2023). Personas as a Way to Model Truthfulness in Language Models. In submission.

Zhaofeng Wu, Linlu Qiu, Alexis Ross, Ekin Akyürek, Boyuan Chen, Bailin Wang, **Najoung Kim**, Jacob Andreas, and Yoon Kim (2023). Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks. In submission.

Najoung Kim, Tal Linzen, and Paul Smolensky (2022). Uncontrolled Lexical Exposure Leads to Overestimation of Compositional Generalization in Pretrained Models. Under revision.

PEER-REVIEWED ARTICLES

Bingzhi Li, Lucia Donatelli, Alexander Koller, Tal Linzen, Yuekun Yao, and **Najoung Kim** (2023). SLOG: A Structural Generalization Benchmark for Semantic Parsing. *Empirical Methods in Natural Language Processing (EMNLP)*.

Jason Wei,* **Najoung Kim**,* Yi Tay, and Quoc V. Le (2023). Inverse scaling can become U-shaped. *Empirical Methods in Natural Language Processing (EMNLP)*. (*Equal contribution)

Abulhair Saparov, Richard Yuanzhe Pang, Vishakh Padmakumar, Nitish Joshi, Seyed Mehran Kazemi, **Najoung Kim**,* and He He* (2023). Testing the General Deductive Reasoning Capacity of Large Language Models Using OOD Examples. *Conference on Neural Information Processing Systems (NeurIPS)*. (*Equal contribution)

Mehran Kazemi, Quan Yuan, Deepti Bhatia, **Najoung Kim**, Xin Xu, Vaiva Imbrasaite, and Deepak Ramachandran (2023). BoardgameQA: A Dataset for Natural Language Reasoning with Contradictory Information. *Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmark Track*.

Ian McKenzie, Alexander Lyzhov, Michael Pieler, Alicia Parrish, Ameya Prabhu, Aaron Mueller, Euan McLean, Aaron Kirtland, Alexis Ross, Alisa Liu, Andrew Gritsevskiy, Daniel Wurgaft, Derik Kauffman, Gabriel Recchia, Jiacheng Liu, Joe Cavanagh, Max Weiss, Sicong Huang, The Floating Droid, Tom Tseng, Tomasz Korbak, Xudong Shen, Yuhui Zhang, Zhengping Zhou*, **Najoung Kim**, Samuel R. Bowman, and Ethan Perez (2023). Inverse Scaling: When Bigger Isn't Better. *Transactions on Machine Learning Research (TMLR)*. (*Winning task authors) **Featured Certification**

Wentao Wang, Wai Keen Vong, **Najoung Kim**, and Brenden Lake (2023). Finding Structure in One Child's Linguistic Experience. *Cognitive Science*.

Najoung Kim* and Sebastian Schuster* (2023). Entity Tracking in Language Models. *Annual Conference of the Association for Computational Linguistics (ACL)*. (*Equal contribution) **Area Chair Award**

Najoung Kim*, Phu Mon Htut*, Samuel R. Bowman, and Jackson Petty (2023). $(QA)^2$: Question Answering with Questionable Assumptions. *Annual Conference of the Association for Computational Linguistics (ACL)*. (*Equal contribution)

Seyed Mehran Kazemi, **Najoung Kim**, Deepti Bhatia, Xin Xu, and Deepak Ramachandran (2023). LAMBADA: Backward Chaining for Automated Reasoning in Natural Language. *Annual Conference of the Association for Computational Linguistics (ACL)*.

Najoung Kim, Jatin Khilnani, Alex Warstadt, and Abed Qaddoumi (2023). Reconstruction Probing. *Findings of the Annual Conference of the Association for Computational Linguistics (Findings of ACL)*.

Najoung Kim, Ellie Pavlick, Burcu Karagol Ayan, and Deepak Ramachandran (2021). Which Linguist Invented the Lightbulb? Presupposition Verification for Question-Answering. *Annual Conference of the Association for Computational Linguistics (ACL)*.

Najoung Kim and Tal Linzen (2020). COGS: A Compositional Generalization Challenge Based on Semantic Interpretation. *Empirical Methods in Natural Language Processing (EMNLP)*. [talk]

Najoung Kim, Song Feng, Chulaka Gunasekara, and Luis A. Lastras (2020). Im-

plicit Discourse Relation Classification: We Need to Talk About Evaluation. *Annual Conference of the Association for Computational Linguistics (ACL)*.

Najoung Kim and Tal Linzen (2019). Compositionality as directional consistency in sequential neural networks. *Context and Compositionality in Biological and Artificial Neural Systems Workshop, NeurIPS*.

Najoung Kim, Roma Patel, Adam Poliak, Alex Wang, Patrick Xia, Tom McCoy, Ian Tenney, Alexis Ross, Tal Linzen, and Benjamin Van Durme, Samuel R. Bowman, and Ellie Pavlick (2019). Probing what different NLP tasks teach machines about function word comprehension. *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM)*. **Best Paper Award**

Alex Wang, Jan Hula, Patrick Xia, Raghavendra Pappagari, R. Thomas McCoy, Roma Patel, **Najoung Kim**, Ian Tenney, Yinghui Huang, Katherin Yu, Shuning Jin, Berlin Chen, Benjamin Van Durme, Edouard Grave, Ellie Pavlick, and Samuel R. Bowman (2019). How to Get Past Sesame Street: Sentence-Level Pretraining Beyond Language Modeling. *Proceedings of the Annual Conference of the Association for Computational Linguistics (ACL)*.

Najoung Kim, Jung-Ho Kim, Maria K. Wolters, Sarah E. MacPherson, and Jong C. Park (2019). Automatic Scoring of Semantic Fluency. *Frontiers in Psychology*.

Najoung Kim, Kyle Rawlins, Benjamin Van Durme, and Paul Smolensky (2019). Predicting the Argumenthood of English Prepositional Phrases. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, **Najoung Kim**, Benjamin Van Durme, Samuel R. Bowman, Dipanjan Das, and Ellie Pavlick (2019). What do you learn from context? Probing for sentence structure in contextualized word representations. *International Conference on Learning Representations (ICLR)*.

Maria K. Wolters, **Najoung Kim**, Jung-Ho Kim, Sarah E. MacPherson, and Jong C. Park (2016). Prosodic and Linguistic Analysis of Semantic Fluency Data: A Window into Speech Production and Cognition. *Interspeech*.

Jung-Ho Kim, **Najoung Kim**, Hancheol Park, and Jong C. Park (2016). Enhanced Sign Language Transcription System via Hand Tracking and Pose Estimation. *Journal of Computing Science and Engineering vol 10.3*.

Najoung Kim and Jong C. Park (2016). A Morphological Approach to the Longitudinal Detection of Dementia. *Proceedings of HCI Korea 2016, The HCI Society of Korea*.

WORK PEER-REVIEWED BY ABSTRACT

Hayley Ross, **Najoung Kim**, and Kathryn Davidson (2024). Fake reefs are sometimes reefs and sometimes not, but are always compositional. Accepted to Experiments in Linguistic Meaning (ELM) 3 (selected as talk).

Kanishka Misra and **Najoung Kim** (2023). Abstraction via exemplars? A representational case study on lexical category inference in BERT. Presented at *The 47th Boston University Conference on Language Development (BUCLD)*.

Najoung Kim and Paul Smolensky (2021). Testing for Grammatical Category Abstraction in Neural Language Models. *The Society for Computation in Linguistics (SCiL)*

(extended abstract, selected as talk).

Sadhwi Srinivas,* **Najoung Kim**,* and Kyle Rawlins (2020). Maximize Presupposition! and the Korean demonstrative *ku*. Presented at *The 94th Annual Meeting of the Linguistic Society of America (LSA)*. [poster] (*Equal contribution)

Najoung Kim, Benjamin Van Durme, Ellie Pavlick, and Paul Smolensky (2018). Linguistically informed tasks for evaluating structure encoded by sentence representations. *WeCNLP Summit*, Facebook HQ, Menlo Park, CA (selected as spotlight talk).

Najoung Kim, Kyle Rawlins, and Paul Smolensky (2018). A gradient blend analysis of English PP verbal dependents. *Conference on Interdisciplinary Approaches to Linguistic Theory (CiALT) 2*, Humboldt-Universität zu Berlin.

Najoung Kim, Kyle Rawlins, and Paul Smolensky (2018). A gradient blend analysis of English PP verbal dependents. *Acceptability judgments in current linguistic theory*, Universitat Autònoma de Barcelona.

OTHER ARTICLES

Yonatan Belinkov, Sophie Hao, Jaap Jumelet, **Najoung Kim**, Arya McCarthy, and Hosein Mohebbi (2023). Proceedings of the 6th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP. Association for Computational Linguistics.

Teven Le Scao, Angela Fan, ... **Najoung Kim** (Evaluation and Interpretability) ... (2022), BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. Preprint on arXiv.

Najoung Kim (2022). Compositional Generalization in Artificial Neural Networks. PhD Dissertation, Johns Hopkins University. Committee: Paul Smolensky, Kyle Rawlins, Benjamin Van Durme, Tal Linzen, and Bob Frank.

SELECTED TALKS / TUTORIALS

Entity Tracking in Language Models.

- Invited talk at South by Semantics Workshop at UT Austin. Feb 2024.
- Invited talk at Penn NLP CLunch Seminar. Oct 2023.

Challenges in logical reasoning with LLMs. Invited talk at Yale/Google ML Workshop: Theory and Practice of Foundation Models. Oct 2023.

Evidence for abstraction-via-exemplars from lexical category inference in neural language models. Invited talk at the Virtual Psycholinguistics Forum. Sep 2023.

Compositional Linguistic Generalization in Artificial Neural Networks: Taking Stock. Invited talk at Harvard Universals Workshop, Apr 2023.

Questions with Questionable Assumptions as a Challenge to QA Systems. Invited talk at Colgate University. Mar 2023.

Compositional Generalization in Neural Networks. Interview at NLP Highlights Podcast hosted by Allen Institute for AI, Jan 2023.

Compositional Linguistic Generalization in Artificial Neural Networks: Taking Stock.

- Invited talk at Seminars on Formal Languages and Neural Networks (FlaNN), Nov 2022.

- Invited talk at CUNY Graduate Center Computational Linguistics Talk Series, Nov 2022.

Compositional Linguistic Generalization in Artificial Neural Networks.

- Invited talk at CompLing Lab, University of Chicago, May 2022.
- 'Tech Talk' at Google, May 2022.
- Invited talk at Boston University Linguistics Colloquium Series, Mar 2022.
- Invited talk at MIT Complang. Mar 2022.
- Invited talk at Seminar on the interactions between formal and computational linguistics (ILFC). Feb 2022.
- Invited talk at Linguistics Colloquium, Seoul National University. Dec 2021.
- Invited talk at KAIST Electrical Engineering Colloquium Lecture Series. Nov 2021.
- Invited talk at NYU ConCats. Nov 2021.
- Invited talk at the University of Michigan Cognitive Science Community Colloquium. Oct 2021.
- Invited talk at the New York Philosophy of Language Workshop. Oct 2021.
- Invited talk at NYU Center for Data Science. Sep 2021.

Compositional Linguistic Generalization in Contemporary Neural Models of Language. Invited talk at Cornell Computational Psycholinguistics Group. Apr 2021.

What Aspects of Meaning are Missing from Current Natural Language Understanding Systems? Invited talk at Boston University, Mar 2021.

COGS: A Compositional Generalization Challenge Based on Semantic Interpretation.

- Invited talk at NERT lab, Georgetown University, Nov 2020.
- Invited poster presentation at the Microsoft Research AI Breakthroughs Workshop, Sep 2020.

Probing what different NLP tasks teach machines about function word comprehension and where to go next. Invited talk at JHU CLSP Seminar, Oct 2019.

The complement-adjunct distinction as gradient blends: the case of English prepositional phrases. Invited talk at the Gradient Symbolic Computation Workshop, Baltimore, Sep 2019.

Semantic Role Labeling Tutorial (with Diego Marcheggiani, Michael Roth, and Benjamin Van Durme), EMNLP, Copenhagen, Sep 2017.

Approximating the Semantic Structures behind Category Fluency Sequences (Poster). MACSIM 6 at CUNY, New York, Oct 2016.

Detection and Categorisation of Neograms in Korean Text (+a). Invited talk at Ertegun House, Oxford, 2015.

ADVISING / MENTORING

PhD Advisees:

- Aditya Yedetore (BU Linguistics), 2022–present
- Hayley Ross (Harvard Linguistics, co-advised with Kathryn Davidson), 2023–present
- Jing Liu (École Normale Supérieure, co-advised with Emmanuel Dupoux & David Harwarth), 2023–present

PhD Committees:

- Zhongping Zhang (Prospectus, BU Computer Science), 2024
- Andrea Burns (Dissertation, BU Computer Science), 2023
- Isidora Tourni (Dissertation, BU Computer Science), 2023
- Afra Feyza Akyürek (Qualifying exam, BU Computer Science), 2023

MS Committees:

- Ashwin Daswani (Thesis for MS in AI, BU Computer Science), 2023
- Rohan Sawant (Thesis for MS in AI, BU Computer Science), 2023

Other Advising & Mentoring:

- Alara Balcisoy, BU Undergraduate Research Opportunities Program (UROP), Spring 2024.
- Bingzhi Li (co-advised with Tal Linzen), Visiting PhD student from Université de Paris, Fall 2022.
- Pablo Santos (co-mentored with Phu Mon Htut), CDS-Courant Undergraduate Research Program (CURP), Spring 2022.

TEACHING

Boston University:

- Spring 2024, Computational Linguistics
- Spring 2024, Topics in Linguistics: Metrics and Evaluation in Natural Language Processing
- Fall 2023, Introduction to Programming for Computational Linguistics
- Spring 2023, Topics in Linguistics: Metrics and Evaluation in Natural Language Processing

New York University:

- Fall 2022, Capstone Project in Data Science
- Fall 2021, Capstone Project in Data Science

Johns Hopkins University:

- Fall 2019, Introduction to Computational Cognitive Science (TA/Co-Instructor)
- Spring 2019, Foundations of Neural Networks (TA/Lab Instructor)
- Spring 2018, Foundations of Cognitive Science (TA)
- Fall 2017, Semantics I (TA/Lab Instructor)
- Spring 2017, Language and Advertising (TA)

Seoul National University:

- Winter 2014, Samsung Convergence Software Course (Instructor)

PROFESSIONAL SERVICE

Publicity Chair:

- NAACL 2024

Senior Area Chair:

- ACL 2023 (Linguistic Theories, Cognitive Modeling, and Psycholinguistics)

Area Chair:

- *SEM 2022 (Psycholinguistics, cognitive linguistics and semantic processing)

Organization:

- BlackboxNLP 2024 (Co-located with EMNLP 2024)
- BlackboxNLP 2023 (Co-located with EMNLP 2023)
- Inverse Scaling Prize (2022)

Ad-hoc Journal Reviews:

- Cognitive Science (2022, 2023)
- Nature Machine Intelligence (2022, 2023)

Conference/Workshop Reviews:

- *SEM (2024, 2023*, 2022, 2019)
- Scale-LLM Workshop (2024)
- EACL Student Research Workshop (2023, 2021)
- EMNLP (2022, 2020*, 2018)
- COLING (2022)
- Bridges and Gaps between Formal and Computational Linguistics (Workshop at ESSLLI 2022)
- CogSci (2022)
- ACL Student Research Workshop, (2019–2022)
- ACL-IJCNLP (2021)
- ACL (2020)
- EMNLP-IJCNLP (2019)

*Acknowledged as outstanding reviewer

PROFESSIONAL MEMBERSHIP

Association for Computational Linguistics
Cognitive Science Society
Linguistic Society of America

COMPUTER SKILLS

Python, R, HTML (proficient)
Java, C, Scheme/Racket, OCaml (basic)
Tools: PyTorch, PCIBex, Mechanical Turk, Prolific

HUMAN LANGUAGES

Korean (native); English (near-native); Spanish (intermediate); Japanese (reading knowledge)

COMMUNITY ENGAGEMENT

Panelist, Learning to Think after ChatGPT: A Panel Discussion — 2023
Center for Data Science, Boston University

Panelist, KASELL Fall Conference — 2021

Diversity & Representation Committee — 2021
Colloquium Student Committee — 2018–2021
Department of Cognitive Science, Johns Hopkins University

Panelist, BrainIAC Professional Development Event — 9/21/2020

Mentor, Samsung Convergence Software Course — 2014
Department of Computer Science, Seoul National University

Peer tutor, English (volunteer work) — 2011
Seoul National University

English tutor (volunteer work) — 2010
Youngnak High School