

# Yejin Cho

<http://yejincho.info>  
ycho@utexas.edu

## INTERESTS

### Computational Semantics

- Representation learning, language modeling, natural language understanding
- Machine learning and neural approaches for modeling semantics

## EDUCATION

### University of Texas at Austin

Sep. 2018 – Present  
Austin, TX

*Ph.D. student*

- Ph.D. in Computational Linguistics
- Advisors: *Katrin Erk* and *Jessy Li*
- Total GPA of 3.93 / 4.0

### Korea University

Sep. 2015 – Feb. 2018  
Seoul, Korea

*Masters student*

- M.A. in English Language and Literature
- Specialization in Phonetics (*Advisor: Hosung Nam*)
- Total GPA of 4.5 / 4.5

### Yonsei University

Mar. 2011 – Aug. 2015  
Seoul, Korea

*Undergraduate student*

- B.A. in Korean Language and Literature
- B.A. in English Language and Literature (Double Major)
- Total GPA of 3.87 / 4.3

### University of California, Los Angeles

Sep. 2013 – June. 2014  
Los Angeles, CA

*Undergraduate Exchange student*

- Major in Linguistics
- Total GPA of 3.97 / 4.0 (*Dean's Honors list for full three quarters*)

## RESEARCH EXPERIENCE

### Hypernym Prediction using WordNet, University of Texas, Austin

Spring, 2019 – Fall, 2019  
Austin, TX

*Research Assistant (Advisor: Katrin Erk)*

*Collaborated with Juan Diego Rodriguez and Yifan Gao*

- Read relevant papers in the literature and participated in weekly meetings.
- Replicated four state-of-the-art systems in relation prediction task on WN18RR dataset by adapting the source codes by the original authors to our dataset and task setup:
  - Translating embeddings for modeling multi-relational data (Bordes et al., 2013 NeurIPS)
  - Predicting semantic relations using global graph properties (Pinter & Eisenstein, 2018 EMNLP)
  - Poincaré embeddings for learning hierarchical representations (Nickel & Kiela, 2018 NeurIPS)
  - CRIM at SemEval-2018 Task 9: A hybrid approach to hypernym discovery (Bernier-Colborne and Barriere, 2018)
- Ran experiments using **seq2seq** model with pre-trained fastText embeddings that predicts a sense-sensitive hypernym (e.g., *wildflower*) from its hyponym (e.g., *woolly daisy*) using full taxonomy paths in WordNet.
- Achieved the **new state-of-the-art performance** in hypernym prediction task on WN18RR dataset.
- Submitted a long paper to ACL 2020 (under review).

**EMCS Laboratory**, Korea University  
(*Education, Mathematics, Computer science and Speech Laboratory*)  
Graduate Researcher (Advisor: Hosung Nam)

Aug. 2015 – Jan. 2018  
Seoul, Korea

- Built Korean Large Vocabulary Continuous Speech Recognition (LVCSR) system (800k vocabulary) from raw text and audio corpora with transcription using Kaldi speech recognition toolkit
- **Subword** (Pseudo-morpheme) **language modeling** for building Korean LVCSR system
- Language modeling experiments using SRILM and RNNLM Toolkit
- Designed and developed Korean text normalization and language preparation package for LM in Kaldi-based ASR system (KoLM) [[code](#)]
- Designed and developed rule-based Korean Grapheme-to-Phone conversion system (KoG2P) [[code](#)]

PUBLICATIONS  
\*: Equal contribution

Yejin Cho\*, Juan Diego Rodriguez\*, Yifan Gao, and Katrin Erk. (under review). **Leveraging WordNet Paths for Neural Hypernym Prediction**. *Submitted to ACL 2020 (long paper)*.

Yejin Cho. 2017. **Functional Awareness of RNNLM-based Word Embeddings**. *Master's thesis*. [[pdf](#)]

Heejo You, Hyungwon Yang, Jaekoo Kang, Youngsun Cho, Sunghah Hwang, Yeonjung Hong, Yejin Cho, Seohyun Kim, and Hosung Nam. 2016. **Development of Articulatory Estimation Model Using Deep Neural Network**. *Phonetics and Speech Sciences*. 8:31-38. [[pdf](#)]

HONORS AND  
AWARDS

*As Graduate Student:*

- Graduate School Fellowship, 2018
- **National Humanities Scholarship**: Graduate Research Scholarship for Humanities and Social Sciences, Korea Student Aid Foundation (KOSAF), 2016-2017
- Honors Scholarships, Korea University, 2016
- Teaching Assistant Scholarships, Korea University, 2015-2017

*As Undergraduate Student:*

- **National Humanities Scholarship**: The Next Century Humanities Scholarship, Korea Student Aid Foundation (KOSAF), 2013-2014
- Semester High Honors, Yonsei University, 2014 Fall, 2012 Spring
- Dean's Honors List, University of California, Los Angeles, 2013 Fall, 2013 Winter, 2014 Spring
- Best Honors Scholarships, Yonsei University, 2012 Fall

TEACHING  
EXPERIENCE

Undergraduate Courses in Computational Linguistics, *UT Austin*

- LIN313 Language and Computers Fall, 2018, Fall, 2019
- LIN350 Computational Semantics Spring, 2020

Undergraduate Courses in English Linguistics, *Korea University*

- ENGL238 English Phonetics Fall, 2015, Fall, 2016
- ENGL399 Studies on English Linguistics Spring, 2016

MATLAB Programming for Humanities majors, *Korea University*

- Covered basics of MATLAB programming for undergraduate students of humanities majors without previous experience in computer programming.
- Ran hands-on classes for over 40 students, answered questions in person and online, scored and commented on student assignments for 4 weeks.
- Led team project LyricsAnalyzer on web crawling and text mining of Korean lyrics. [[slide](#)]

TECHNICAL SKILLS	<b>Advanced</b>	Python, MATLAB, Kaldi Speech Recognition Toolkit, SRILM, HTK
	<b>Moderate</b>	PyTorch, TensorFlow, UNIX shell scripting, R
	<b>Other Toolkits</b>	NLTK, Stanford CoreNLP, Scikit-learn, DyNet
SELECTED COURSES	<i>At Korea University:</i>	
	BRI515 Introduction to Applied Mathematics (audited)	Fall, 2015
	BRI516 Introduction to Neural Networks	Spring, 2016
	AAA625 Numerical Linear Algebra	Fall, 2016
	BRI629 Applied Mathematics for Brain and Cognitive Engineering	Spring, 2017
	AAA606 Natural Language Processing Applications	Spring, 2017
	<i>At the University of Texas at Austin:</i>	
	LIN393S Computational Discourse	Fall, 2018
	CS388 Natural Language Processing (Graduate level)	Fall, 2019
	LANGUAGE PROFICIENCY	Native in Korean
Fluent in English (IBT TOEFL 116; Reading: 30, Listening: 29, Speaking: 29, Writing: 28)		
Intermediate in French (DELF B1)		
Basic conversation and readability in Japanese and Modern Standard Arabic (MSA)		