

Youngmeen Kim

---

Ph.D.  
Department of Applied Linguistics  
Georgia State University  
25 Park Place, Atlanta, GA 30303  
Email: ykim140@gsu.edu  
Phone: 1-808-798-6272

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |  |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| RESEARCH AREAS      | <ul style="list-style-type: none"><li>- Computational Linguistics and Corpus Linguistics</li><li>- Explainable Machine Learning and AI for Interdisciplinary Research</li><li>- Language and Human Behavior (cognitive, social, and psychological aspects)</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
| EDUCATION           | <p><b>Doctor of Philosophy, Applied Linguistics</b><br/><b>Georgia State University, Atlanta, GA</b> 2021 - 2026<br/><i>Dissertation:</i> Language as a Reflection of Human Minds: Machine Learning-Based Corpus Analysis of Linguistic Patterns in Online Mental Health Communities<br/><i>Committee Members:</i> Drs. Ute Römer-Barron, YouJin Kim, Daniel Dixon (Georgia State University), and Munmun De Choudhury (Georgia Institute of Technology)<br/><i>Specialization:</i> Computational Linguistics and Corpus Linguistics</p> <p><b>Master of Arts, Second Language Studies</b><br/><b>University of Hawai'i at Mānoa, Honolulu, HI</b> 2019 - 2021<br/><i>Graduation Paper:</i> Computational Analysis of Cohesion in EFL Writing: A Comparison between L1 and EFL Writers<br/><i>Specialization:</i> Computational Linguistics and Language Education</p> <p><b>Bachelor of Arts, English Education (Summa Cum Laude)</b><br/><b>Bachelor of Economics, Economics and Law (Summa Cum Laude)</b><br/><b>Hankuk University of Foreign Studies, Seoul, Korea</b> 2011 - 2018</p>                          |  |
| RESEARCH EXPERIENCE | <p><i>Research Assistant</i> January 2026 - May 2026<br/>Georgia State University, College of Law</p> <ul style="list-style-type: none"><li>• Advising eleven Juris Doctor (J.D.) candidates in designing and conducting research for the course <i>Linguistic Interpretation and Investigation of U.S. Legal Documents</i></li></ul> <p><i>Research Assistant, Software Developer</i> May 2024 - May 2025<br/>Georgia State University, Department of Applied Linguistics</p> <p>Research on the Challenges of Acquiring Language &amp; Literacy (RCALL)<br/>Seed grants, Georgia State University (\$48,500)<br/>YouJin Kim (PI) and Daniel Dixon (Co-PI)<br/><i>Generative AI for Supporting Recently Immigrated English Learners' Language Development through Alignment: The Role of Input Complexity and Learner Proficiency</i></p> <ul style="list-style-type: none"><li>• Developed Generative AI-based human-computer interaction system for language learners</li><li>• Implemented Text-to-Speech, Speech-to-Text, and Speech-to-Speech functionalities to enhance human-computer interaction</li></ul> |  |

- Designed and managed data collection and developed NLP scripts for data analysis

*Research Assistant*

January 2024 - May 2024

Georgia State University, College of Law

- Advised six J.D. candidates on the use of NLP tools and corpus-linguistic approaches for the course *Linguistic Interpretation and Investigation of U.S. Legal Documents*

*Research Assistant, Software Developer*

August 2023 - December 2023

Georgia State University, Department of Applied Linguistics

Cambridge English Funded Research Programme (FRP) (£13,000)

YouJin Kim (PI) and Daniel Dixon (Co-PI)

*Generative AI for Supporting Recently Immigrated English Learners' L2 Development through Alignment: The Role of Task Input Complexity and Interaction Modalities*

- Developed GPT-based chatbot designed to help the immigrant population
- Designed NLP scripts for data analysis including measuring lexical sophistication, syntactic complexity, and large-language-model-based semantic alignment
- Developed NLP scripts for collocation and semantic analysis

*Research Assistant*

August 2022 - May 2023

Georgia State University, Department of World Languages and Cultures

- Managed a virtual exchange program between Korean students and Georgia State University students
- Conducted data collection and analysis of virtual interactions

*Research Assistant*

May 2021 - August 2022

Georgia State University and The Learning Agency

- Analyzed demographic data linked to writing samples from U.S. K-12 students
- Performed data cleaning and preprocessing
- Conducted bias and discrimination analyses

*Research Assistant*

August 2020 - May 2021

Georgia State University, Department of Applied Linguistics

- Developed a new text readability formula using part-of-speech tags
- Analyzed traditional readability formulas and the correlation between readability scores and human-based holistic readability ratings
- Assisted in developing a machine learning-based readability formula
- Examined the validity and reliability of the new readability formula

*Research Assistant*

May 2017 - December 2018

Cha and Kim English Education Research Center, Seoul, Korea

- Investigated the effects of an extensive reading program on undergraduate students
- Designed and developed the extensive reading program
- Managed research participants and analyzed data

Research Assistant

May 2013 - December 2017

HUFS Education Center, Seoul, Korea

- Conducted a longitudinal study examining factors influencing campus life adjustment and satisfaction
- Collected data spanning four years from over 20,000 undergraduate students
- Analyzed the relationship between admission types and academic performance, school adjustment, and other needs

**PUBLICATIONS** **Kim, Y.** & Römer-Barron, U. (2025) Detection and analysis of depression-related language in an online community: Machine learning, topic modeling, and corpus-linguistic approaches. *Corpus-based Studies across Humanities*.

Crossley, S., Tian, Y., Baffour, P., Franklin, A., **Kim, Y.**, Morris, W., Benner, M., Picou, A., & Boser, U. (2024) The English Language Learner Insight, Proficiency and Skills Evaluation (ELLIPSE) Corpus. *International Journal of Learner Corpus Research*.

**Kim, Y.** (Under review) Machine Learning Insights into L2 Development: Tracing Collocation Use Across Various Proficiency Levels. *Applied Linguistics*.

**Kim, Y.** & Vogel, A. (Under review) Lexical and Syntactic Predictors of Human-Judged Readability: An Interpretable Machine Learning Analysis of Main Effects and Genre Interactions. *Applied Corpus Linguistics*.

Dixon, D., Kim, YJ., & **Kim, Y.** (Under review) The Power of the Prompt: The Linguistic Effects of Prompt Engineering on GenAI and Learners. *Research Methods in Applied Linguistics*.

**IN PREPARATION** **Kim, Y.** & Kubek, M. Understanding the language of distress: Identification of distinctive language and topics in online PTSD and anxiety disorder communities. *Computers in Human Behavior*.

**Kim, Y.** Explainable Machine Learning of L2 Development: Linking Lexico-Syntactic Features to Dynamic Systems Theory. *System*.

**INVITED TALKS** *Guest speaker*

**Kim, Y.** (November 2025) *Human versus Machine Language Learning*. Special Lecture for the course *Language and Mind*. Georgia State University, Georgia

*Keynote speaker*

**Kim, Y.** (November 2024) *Automatic Language Evaluation Using Machine Learning: Integrating LLM with Lexical and Syntactic Features*. Linguistics Symposium, University of North Georgia, Georgia

*Invited speaker*

**Kim, Y.** (July 2024) *Research Applications and Possibilities of Natural Language Processing and Machine Learning in the Fields of Linguistics and Education*. Korea Association of Multimedia-Assisted Language Learning, Seoul, Korea

**MEDIA  
SPOTLIGHT**

**NBC 11Alive** (December 2025)

*“Georgia State study uses AI to better spot depression.”*

Televised and online news coverage. <https://www.11alive.com/article/tech/georgia-state-study-uses-ai-to-better-spot-depression/85-4fa0abde-4871-4a13-956e-875b0be15089>

**The Korea Times Atlanta** (December 2025).

*“AI-based research identifies linguistic markers of depression in online communities.”*

Coverage of AI-driven mental health research and its social implications.

<https://higoodday.com/news/1010550>

**Republished in: The Korea Times Florida** (December 2025).

**Atlanta JoongAng Daily** (December 2025).

*“Georgia State study highlights AI’s role in detecting depression through language.”*

Korean-language media coverage of AI-based interdisciplinary research.

<https://www.atlantajoongang.com/124515>

**Georgia State University Newsroom** (October 2025)

*“Study: AI Can Spot Signs of Depression in Reddit Posts.”*

<https://news.gsu.edu/2025/10/21/study-ai-can-spot-signs-of-depression-in-reddit-posts>

**Science X Network (Phys.org)** (October 2025)

*“AI Can Spot Signs of Depression in Reddit Posts.”*

<https://phys.org/news/2025-10-ai-depression-reddit.html>

**PRESENTATIONS** Kim, YJ., Dixon, D., **Kim, Y.**, & Cathey, R. (May 2025) *The development of ‘ReaL2Chat’, a GenAI TBLT chatbot platform: Prompt engineering for tasks designed with communicative goals and effects on learning outcomes* Computer-Assisted Language Instruction Consortium, San Diego State University, California

Kim, YJ., Dixon, D., **Kim, Y.**, & Cathey, R. (April 2025) *Task Performance with Generative AI as an Interlocutor: Input Complexity, Task Type, and Student Perceptions.* TBLT Conference, University of Groningen, Netherlands.

**Kim, Y.** & Römer-Barron, U. (March 2025) *How different is language use across online mental health communities? Machine-learning and corpus-analytic approaches* American Association for Applied Linguistics, Denver, Colorado

Kim, YJ., Dixon, D., **Kim, Y.**, & Cathey, R. (March 2025) *Effects of a customized GenAI platform on the writing development of recently immigrated English learners* American Association for Applied Linguistics, Denver, Colorado

Jordan, J., Dixon, D., & **Kim, Y.** (March 2025) *Defining “rhetoric” in second language writing contexts: A corpus-based approach* American Association for Applied Linguistics, Denver, Colorado

**Kim, Y.** & Römer-Barron U. (September 2024) *Detection and Analysis of Depression-related Language in a Corpus of Online Texts: Machine Learning and Topic Modeling Approaches* American Association for Corpus Linguistics, Oregon

**Kim, Y.** (September 2024) *Machine learning insights into L2 development: Tracing Collocation Use Across Proficiency Levels* American Association for Corpus Linguistics, Oregon

Dixon, D., Kim, YJ., & **Kim, Y.** (May 2024) *Generative AI for Supporting Recently Immigrated English Learners' L2 Lexical Development: Linguistic Alignment, Lexical Complexity, and Learner Perceptions* Computer-Assisted Language Instruction Consortium, Carnegie Mellon, Pittsburgh, Pennsylvania

**Kim, Y.** (April 2024) *Distinguishing PTSD from Anxiety Disorders: A Machine Learning Investigation of Linguistic Patterns in Online Mental Health Communities* South NLP Symposium, Emory University, Georgia

**Kim, Y.** & Crossley S. (July 2022) *The relationship between POS tag and human judgment-based text readability score* The Society for Text and Discourse, Georgia

Crossley S., Scherber, Y., Choi, J. S., Lucka, M., & **Kim, Y.** (July 2022) *Using sentence embeddings to predict text readability: A comparison* The Society for Text and Discourse, Georgia

## SERVICE TO ACADEMIA

*Manuscript Reviewer* February 2026  
Journal of Affective Disorders

*Conference Reviewer* October 2025  
American Association for Corpus Linguistics

*Manuscript Reviewer* November 2023  
Computers in Human Behavior

*Conference Reviewer* August 2023  
American Association for Applied Linguistics

## PROGRAMMING SKILLS Python and R

### Natural Language Processing

#### **Feature Engineering Methods:**

Bidirectional Encoder Representations from Transformers (BERT), text vectorization using N-gram, Term Frequency-Inverse Document Frequency (TF-IDF), parts-of-speech (POS) tagging, and dependency parsing, psycholinguistic feature engineering, sentiment analysis using Valence Aware Dictionary and sEntiment Reasoner (VADER), Emotion Lexicon (EmoLex), data preprocessing, linguistic feature extraction, and modeling using spaCy and NLTK modules

### Machine Learning and Deep Learning

**Machine Learning Models:** Support Vector Machines (SVM), Logistic Regression (LR), Elastic Net, Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and CatBoost (CAT), and ensemble modeling techniques (e.g., stacking, blending, voting)

**Deep Learning Models:** Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), Long Short-Term Memory Networks (LSTMs), and Temporal Convolutional Networks (TCNs)

**Unsupervised Methods:** Topic modeling (LDA, BERTopic)

### Interpretable Machine Learning and Explainable AI

Feature ablation methods, permutation importance, coefficient-based analysis for interpretable machine learning (IML), SHapley Additive exPlanations (SHAP), and Local Interpretable Model-Agnostic Explanations (LIME) for explainable AI (XAI) and black-box model interpretation from linguistic perspectives

## Machine Learning Usages

Text analysis (e.g., classification, clustering, time series forecasting, trend analysis, sentiment analysis), dimensionality reduction (e.g., Principal Component Analysis, PCA), feature selection (e.g., Recursive Feature Elimination, RFE), topic modeling (e.g., Latent Dirichlet Allocation, LDA, and BERTopic), handling imbalanced datasets through oversampling (random oversampling, Synthetic Minority Oversampling Technique, SMOTE) and undersampling (random undersampling, NearMiss-1, NearMiss-2, NearMiss-3, Edited Nearest Neighbors, ENN, and Neighborhood Cleaning Rule, NCR), and ensemble methods

## Modules frequently used

TensorFlow, Scikit-learn, PyTorch, NVIDIA CUDA, RAPIDS Toolkit (including cuDF for GPU-accelerated DataFrame operations, cuML for GPU-accelerated machine learning algorithms, and cuPY for GPU-accelerated arrays), Optuna for hyperparameter tuning, and NSGA-II (Non-dominated Sorting Genetic Algorithm II) for multi-objective optimization

## Statistical Computing

**Proficient in:** Correlation test, t-test, ANOVA, MANOVA, Single/Multiple linear regression, Linear Mixed-effects models, Path analysis, Structural equation modeling (SEM), Factor analysis (CFA, EFA), Rasch model, Discriminant function analysis (DFA), Multiple Discrimination Analysis (MDA)

## Libraries and Packages

Numpy, SciPy, Pandas, Statsmodels, Scikit-learn, StatsPy, pingouin, scipy.stats for data analysis, and Matplotlib, Seaborn, YellowBrick, Plotly, Scikit-plot, and more

## TEACHING EXPERIENCE

*Applied Linguistics Instructor* January 2026 - May 2026  
Georgia State University, Department of Applied Linguistics  
*Course name:* Language in Society

*Applied Linguistics Instructor* August 2025 - Present  
Georgia State University, Department of Applied Linguistics  
*Course name:* Introduction to English Linguistics

*Applied Linguistics Instructor* January 2025 - May 2025  
Georgia State University, Department of Applied Linguistics  
*Course name:* Introduction to Linguistics

*Applied Linguistics Instructor* June 2024 - August 2024  
Georgia State University, Department of Applied Linguistics  
*Course name:* Introduction to English Linguistics

*Applied Linguistics Instructor* January 2024 - May 2024  
Georgia State University, Department of Applied Linguistics  
*Course name:* Introduction to Linguistics

*College of Law Teaching Assistant* January 2023 - May 2024  
Georgia State University, College of Law  
*Course name:* Linguistic Interpretation and Investigation of U.S. Legal Documents

*World Languages & Cultures Instructor* August 2022 - December 2023  
Georgia State University, Department of World Language & Cultures  
*Course name:* Intermediate Korean

*World Languages & Cultures Instructor* August 2021 - May 2022  
Georgia State University, Department of World Language & Cultures  
*Course name:* Elementary Korean

*East Asian Languages & Literature Tutor* August 2019 - August 2020  
University of Hawai'i at Mānoa, Department of East Asian Languages & Literature  
*Course name:* Korean Language

**LANGUAGES**

- Korean (Native)
- English (Near-native)
- French (Intermediate)