



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities
BLDG 14 RM 520,
Seoul National
University, 1
Gwanak-ro,
Gwanak-gu
Seoul
South Korea

Last Revised

2026-06-06

Sumin Park

Computational Linguist

About Me I am a visiting assistant professor at the College of Humanities at Seoul National University.

Education

2015–2021, Seoul National University, Seoul, South Korea

PhD in Linguistics

- **Dissertation Title:** *Fusion models for news quality prediction: Combining textual features with sentence embeddings*
– <https://s-space.snu.ac.kr/handle/10371/177766>
- **Supervisor:** Prof. Hyopil Shin
- **Committee:** Professors Seungho Nam (Chair), Hyopil Shin, Seungshik Kang, and Sujin Choi, and Dr. Munhyong Kim

2012–2014, Seoul National University, Seoul, South Korea

MA in Linguistics

- **Dissertation Title:** *Retrieval of Twitter messages without an explicit query term by means of serialization and discourse segmentation*
– <https://s-space.snu.ac.kr/handle/10371/131950>
- **Supervisor:** Prof. Hyopil Shin
- **Committee:** Professors Seungho Nam (Chair), Hyopil Shin, and Minhwa Chung

2006–2012, Seoul National University, Seoul, South Korea

BA in Linguistics and BSc in Mathematics

Work History

2024–present, Visiting Assistant Professor, Seoul National University

Data Science for the Humanities Minor, College of Humanities

2017–2024, (Part-time) Lecturer, Seoul National University

Department of Linguistics

2024, Senior Researcher, Seoul National University Hospital

Biomedical Research Institute

2022–2024, Senior Researcher, Seoul National University

Medical Research Center



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities

BLDG 14 RM 520,

Seoul National

University, 1

Gwanak-ro,

Gwanak-gu

Seoul

South Korea

Last Revised

2026-06-06

Publications & Presentations

Peer-reviewed Journal Articles

Seo, J., **Park, S.**, Byun, S., Choi, J., Choi, J., & Shin, H. (2025). Advancing Korean Medical Large Language Models: Automated Pipeline for Korean Medical Preference Dataset Construction. *Healthcare informatics research*, 31(2), 166–174.

– <https://doi.org/10.4258/hir.2025.31.2.166>

Kim, K., Park, S., Min, J., **Park, S.**, Kim, J.Y., Eun, J., Jung, K., Park, Y. E., Kim, E., Lee, E. Y., Lee, J., & Choi, J. (2024). Multifaceted Natural Language Processing Task-Based Evaluation of Bidirectional Encoder Representations From Transformers Models for Bilingual (Korean and English) Clinical Notes: Algorithm Development and Validation, *JMIR Medical Informatics* 2024, 12:e52897.

– <https://doi.org/10.2196/52897>

박수민. (2023). <조선 후기 한글 음식조리서에 나타난 미각형용사 ‘달다’의 분포 변화>. 《어문론총》 98, 37–62.

– <http://dx.doi.org/10.22784/eomun.2023..98.37>

김민석, **박수민**, 신호필. (2023). <한국어 사전학습모델의 토큰화가 문장 임베딩에 끼치는 영향 분석>. 《언어》 48(4), 917–947.

– <https://doi.org/10.18855/lisoko.2023.48.4.007>

이상아, 장한솔, 백연미, **박수지**, 신호필. (2020). <소규모 데이터 기반 한국어 버트 모델>. 《정보과학회논문지》 47(7), 682–692.

– <https://doi.org/10.5626/JOK.2020.47.7.682>

Park, S., & Shin, H. (2019). Leveraging More Fine-grained Representation to Reduce Instability within Word Embeddings. 《언어와 정보》 23(3), 1–18.

– <https://doi.org/10.29403/LI.23.3.1>

신호필, 김문형, **박수지**. (2016). <한국어 감정분석 코퍼스를 활용한 양상 정보 기반의 감정분석 연구>. 《언어학》 74, 93–114.

– <https://doi.org/10.17290/jlsk.2016..74.93>



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities

BLDG 14 RM 520,

Seoul National

University, 1

Gwanak-ro,

Gwanak-gu

Seoul

South Korea

Last Revised

2026-06-06

Publications & Presentations (continued)

Peer-reviewed Conference Papers

Byun, S., Hong, J., **Park, S.**, Jang, D., Seo, J., Kim, M., Oh, C., & Shin, H. (2024). Korean Bio-Medical Corpus (KBMC) for Medical Named Entity Recognition. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*.

– <https://aclanthology.org/2024.lrec-main.868/>

Park, S., & Shin, H. (2018, May). Grapheme-level Awareness in Word Embeddings for Morphologically Rich Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

– <https://aclanthology.org/L18-1471/>

Park, S., & Shin, H. (2014, August). Identification of Implicit Topics in Twitter Data Not Containing Explicit Search Queries. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*.

– <https://aclanthology.org/C14-1007/>

Kim, Y., Kim, M., Cattle, A., Otmakhova, J., **Park, S.**, & Shin, H. (2013, October). Applying Graph-based Keyword Extraction to Document Retrieval. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP 2013)*.

– <https://aclanthology.org/I13-1108/>

Conference Presentations

박수지. (2021). <뉴스 기사 제목에 나타난 ‘○○女’와 ‘○○男’의 사용 양상 비교: 1990–2021년 사회면 기사를 중심으로>. 한국사회언어학회 가을학술대회.

박수지. (2017). <한글 자모를 반영한 한국어 단어 의미의 벡터 표상>. 한국언어학회 가을학술대회.

김미경, 박수지, 이상아. (2016). <형태소 깎는 노인: 국어사 자료를 위한 형태분석 보조기>. 한글 및 한국어 정보처리 학술대회.



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities

BLDG 14 RM 520,

Seoul National

University, 1

Gwanak-ro,

Gwanak-gu

Seoul

South Korea

Last Revised

2026-06-06

Teaching Experience

2024–present, *Visiting Assistant Professor*, College of Humanities, Seoul National University, Seoul, South Korea

Courses

- **Introduction to Data Science for the humanities.** Spring 2026, 2025.
- **Quantitative Analysis in the Humanities.** Fall 2025, 2024.
- **Seminar in Data Science for the Humanities.** Spring 2026, Fall 2025, Spring 2025, Fall 2024.



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities
BLDG 14 RM 520,
Seoul National
University, 1
Gwanak-ro,
Gwanak-gu
Seoul
South Korea

Last Revised

2026-06-06

Teaching Experience (continued)

2017–2024, (*Part-time*) *Lecturer*, Department of Linguistics, Seoul National University, Seoul, South Korea

Courses

- **Digital Computer Concept and Practice.** Spring 2024, 2023, 2018.
 - Undergraduate General Education in Linguistics (035.001)
 - Designed for first-year students in the humanities
 - * Fundamentals of computer science
 - * How digital computers represent and process various types of data, such as numbers, text, images, and sound
 - * Python programming language, from the basics to the use of classes
- **Linguistics and Data Science.** Fall 2024, 2023, 2022.
 - <https://github.com/suparklingmin/LingDataSci2023>
 - Undergraduate Major Elective in Linguistics (M1246.001400)
 - Teaches data analysis techniques applicable to various subfields of linguistic research
 - * Fundamental concepts of data science
 - * Methods for managing and utilizing linguistic data
 - * Python libraries such as NumPy, pandas, seaborn, and scikit-learn
- **Language and Computer.** Fall 2021, 2020, 2018, 2017.
 - <https://github.com/suparklingmin/LangComp2021>
 - Undergraduate Major Exploration in Linguistics (100.130)
 - An introductory course in Natural Language Processing (NLP) using Python
 - * Regular expressions
 - * N-gram language models
 - * Naive Bayes classifiers
 - * Logistic regression
 - * Word embeddings
- **Computational Linguistics.** Spring 2022, 2021, 2019.
 - <https://github.com/suparklingmin/CompLing2022>
 - Undergraduate Major Requirement in Linguistics (108.413A)
 - Explores deep learning techniques in NLP with practical sessions using PyTorch
 - * Perceptron & neural network language models
 - * Neural network training techniques
 - * Architectures of CNNs, RNNs, and LSTMs
 - * Contextualized embeddings such as BERT



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities

BLDG 14 RM 520,

Seoul National

University, 1

Gwanak-ro,

Gwanak-gu

Seoul

South Korea

Last Revised

2026-06-06

Research Experience

May 2024 – August 2024, *Senior Researcher*, Biomedical Research Institute at Seoul National University Hospital, South Korea

- **Project Investigator:** Prof. Jinwook Choi
- **Project Title:** 생성형 언어모델과 CDW를 활용한 자연어 검색 시스템 및 인공지능 Doctor Assistant 개발

Aug 2022 – May 2024, *Senior Researcher*, Medical Research Center at Seoul National University, South Korea

- **Project Investigator:** Prof. Jinwook Choi
- **Project Title:** Pointer-generator network과 로봇저널리즘 기술을 이용한 의료문서 자동 요약 시스템 개발

Nov 2019 – Nov 2020, *Research Assistant*, Korea Technology & Information Promotion Agency for SMEs, South Korea

- **Project Investigator:** Prof. Hyopil Shin
- **Project Title:** AI를 활용한 웹 기반의 Legal RPA 솔루션 개발

Jul 2017 – Jun 2018, *Research Assistant*, National Research Foundation of Korea, South Korea

- **Project Investigator:** Prof. Hyopil Shin
- **Project Title:** 한국어 처리를 위한 효율적인 단어 임베딩 모델 구축 및 활용에 관한 연구

Apr 2016 – Nov 2016, *Research Assistant*, National Institute of Korean Language, South Korea

- **Project Investigator:** Prof. Bonkwan Koo
- **Project Title:** 2016년 중앙행정기관 공공언어 진단사업 용역

Mar 2015 – Oct 2015, *Research Assistant*, Samsung Electronics

- **Project Investigator:** Prof. Hyopil Shin
- **Project Title:** 한영, 한중 기계번역 전처리 (Re-ordering, 토큰변형) 기법 연구

May 2014 – Apr 2015, *Research Assistant*, National Research Foundation of Korea, South Korea

- **Project Investigator:** Prof. Hyopil Shin
- **Project Title:** 담화관계를 기반으로 한 온라인 토론에서의 입장 분석 연구

Apr 2014 – Nov 2014, *Research Assistant*, Samsung Electronics

- **Project Investigator:** Prof. Hyopil Shin
- **Project Title:** 상이어순 단어쌍 번역요소기술 선행연구



Sumin PARK

박수민

朴修敏

Formerly Suzi PARK

개명 전 성명: 박수지

改名之前的姓名: 朴修智

Email

mam3b@snu.ac.kr

Website

suminpark.kr

Github

suparklingmin

Address

College of Humanities

BLDG 14 RM 520,

Seoul National

University, 1

Gwanak-ro,

Gwanak-gu

Seoul

South Korea

Last Revised

2026-06-06

Research Experience (continued)

Sep 2012 – Dec 2012, *Research Assistant*, Electronics and Telecommunications Research Institute, South Korea

- **Project Investigator:** Prof. Hyopil Shin
- **Project Title:** SNS와 웹 텍스트 데이터에서의 콘텐츠 객체 수집 및 시험 검증 모듈 개발

Academic Service

2024, *Reviewer*, EMNLP 2024

2024, *Reviewer*, ACL 2024

Language Skills

Natural Languages

- Korean (Native)
- English (Proficient)
- Classical Chinese (Reading)
- Chinese (Reading)
- Japanese (Reading)
- French (Reading)
- Portuguese (Reading)
- Spanish (Reading)

Programming Languages

- Python
- R
- $\text{\LaTeX}$