

# Seongmin Mun

[simon.seongmin.mun@gmail.com](mailto:simon.seongmin.mun@gmail.com), <https://seongminmun.com/>

## INTERESTS

Main area: Natural Language Processing, Machine learning, Data visualization, Data analysis, Statistics

Current Focus: Word sense disambiguation, language acquisition, Classification/visualization system development, Word segmentation

## SKILLS

Programming Languages: Python, R, Java, JavaScript, SQL, PHP, HTML/CSS Web Application: Bitnami, Flask, Docker, Node.js, Shiny, Spring Natural Language Processing: nltk, spaCy, KoNLPy, stringr, tm, KoNLP, CoreNLP, KKMA Machine Learning: TensorFlow, Transformers, Keras, PyTorch, Gensim, scikit-learn, SciPy, MASS, Imtest Data Visualization: D3.js, Google Charts, Chart.js, matplotlib, ggplot2 Statistics Software: R (RStudio), SAS, SPSS, MATLAB DBMS: MySQL, MongoDB Web Server: Apache2 Web Client: Pure JavaScript Applications AWS: EC2, RDS, Route 53, Certificate Manager

## EDUCATION

Ph.D. in Sciences du Langage, Université Paris Nanterre, France, JUL 2021

Dissertation: Polysemy resolution with word embedding models and data visualization: the case of adverbial postpositions -ey, -eyse, -(u)lo in Korean (Chairperson: [Dr. Guillaume Desagulier](#))

M.S. in Media Contents, Ajou University, South Korea, FEB 2016

Dissertation: Data analysis by Integrating statistics and visualization: Visual verification for the prediction model (Chairperson: Prof. Kyungwon Lee)

B.S. in Statistics, Pyeongtaek University, South Korea, FEB 2014

## EXPERIENCE

PostDoctoral Researcher, Chosun University, AUG 2021 - PRESENT

Designed and developed an [eye tracking-based web experimental application](#), and word segmentation model by using syllable-based transitional probability

Graduate Student Researcher, Université Paris Nanterre, MAR 2016 - JUL 2021

Designed and developed language model-based recognition systems covering various topics (e.g., [similarity-based estimator](#), [multiword parser](#)) and data visualization systems (e.g., [PostEmbedding](#), [PostBERT](#))

Research Assistant, Palacký University Olomouc, FEB 2021 - JUN 2021

Developed and designed a web-based [Korean test system](#) for L2 learner

Research Assistant, Université Paris 3, NOV 2018 - FEB 2019

Designed and developed a [web-based data archiving system](#) for efficient handling and safe management of child language development records

Research Assistant, Université Paris 8, JUL 2018 - SEP 2018

Developed and launched COHA [pre-trained models](#) by using RoBERTa, and a [data visualization system](#) to evaluate the COHA pre-trained models

Graduate Student Researcher, Ajou University, MAR 2014 - FEB 2016

Conducted statistical analysis for various data types (e.g., diagnosis records, movie reviews), and evaluated the obtained results by using data visualization systems (e.g., [ParallelSentimentVis](#), [BubbleJoseon](#))

## AWARDS

Best Paper Award, Journal of Design Research, 2018 KSDS Fall International Conference, NOV 2018

Best Researcher Award from the Minister of Education, The Education Ministry of Korea, National Research Foundation of Korea, MAR 2018

Kantar Information is Beautiful Awards, Kantar, NOV 2016

## PAPERS AND PROCEEDINGS

### Major

Mun, S. & Desagulier, G. (2022). How do Transformer-Architecture Models Address Polysemy of Korean Adverbial Postpositions? In Proceedings of Deep Learning Inside Out (DeeLIO): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, 1–11, Dublin, Ireland. Association for Computational Linguistics (ACL).

Classification model, Visualization system, GPT-2, BERT, Korean

Shin, G-H. & Mun, S. (2022). Limits on Neural Networks: Agent-First Strategy in Child Comprehension. In Proceedings of the 44th Annual Meeting of the Cognitive Science Society (CogSci 2022). 1-10, Toronto, Canada.

Child language development, LSTM, BERT, GPT-2, Korean

Shin, G-H. & Mun, S. (2022). Korean-speaking children's constructional knowledge about a transitive event: Corpus analysis and Bayesian modelling. Journal of Child Language, 1-27.

Bayesian modeling, Language development, Child language data, Korean

Ha, H., Han, H., Mun, S., Bae, S., Lee, J. & Lee, K. (2019). An improved study of multilevel semantic network visualization for analyzing sentiment word of movie review data. Applied Sciences, 9, 2419.

Sentiment analysis, Recommendation system, Network visualization, Movie reviews, Korean

Lee, K., Lee, J., Kim, D., Park, J., Mun, S., Jang, Y. & Park, J. (2017). Controversy Visualization- How Controversial Public Discourse in Wikipedia Articles Evolves over Time. Archives of Design Research, 30(4), 57-69.

Wikipedia Revision History Visualization, Public Discourse Analysis, Flow Circle, Integrated Design

## **Others**

Jung, B. K., Shin, G-H., & Mun, S. (under review). What Natural Language Processing techniques tell us about text quality of learner writing: Relationship between similarity scores and proficiency.

Shin, H-J., & Mun, S. (Upcoming). A Text Mining Analysis of Translator's Style: Comparing Translating Styles of 'as if' Constructions. T&I REVIEW.

Kim, S-H., Mun, S., & Ko, E-S. (2022). Sound symbolism according to gender: Focusing on the names of residents' characters in the Animal Crossing game. *Studies in Phonetics, Phonology and Morphology*, 28(1).

Shin, G-H., & Mun, S. (2021). Bayesian simulation of clause-level constructional knowledge in child language development: Active transitives and suffixal passives in Korean. In D. Dionne and L-A. Vidal Covas (Eds.), *Proceedings of the 45th Annual Boston University Conference on Language Development* (pp. 679–692). Boston, MA: Cascadilla Press.

Ha, H., Mun, S., Choi, G., Hong, E., Lee, S., & Lee, K. (2019). Visual analysis on the political orientation of historical characters in the Joseon Dynasty: Focusing on Seo, Geojeong. *Archives of Design Research*, 32(1), 147-161.

Mun, S., Choi, G., Desagulier, G. & Lee, K. (2018). PreechVis: Visual profiling using multiple-word combinations. *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 97-107.

Mun, S., Choi, G., Lee, S., & Lee, K. (2017). Visual analysis for voting relationships in Joseon Dynasty, Korea. *Proceedings of the 10th International Conference on ICT, Society and Human Beings*, 111-118.

Mun, S., Desagulier, G. & Lee, K. (2017). A visual approach for text analysis using multiword topics. *IEEE, EuroVis 2017, Proceedings of the 19th EG/VGTC Conference on Visualization*, 57-59.

Mun, S., & Lee, K. (2016). Data analysis by integrating statistics and visualization: visual verification for the prediction model. *Journal of Society of Design Convergence*, 15(6), 195-214.

Park, S., Jo, E., Mun, S., & Lee, K. (2016). Analysis of corporate twitter usage types and Influence visualization on key players. *Journal of Society of Design Convergence*, 15(6), 129-144.

Mun, S., Kim, G., Choi, G., & Lee, K. (2016). Visual trend analysis method for ontology-based opinion mining on movie review. *Proceedings of the 13th International Conference on Web Based Communities and Social Media 2016*, 287-290.

Ha, H., Han, H., Mun, S., Bae, S., Lee, J., & Lee, K. (2016). Visualization of movie recommendation system using the sentimental vocabulary distribution map. The Korea Society of Computer and Information, 21(5), 19-29.

Mun, S., Kim, G., Choi, G., & Lee, K. (2016). Movie Recommended system based on analysis for the user review utilizing ontology visualization. Journal of Society of Design Convergence, 15(2), 347-368.

Park, S., Jo, E., Lee, H., Mun, S., & Lee, K. (2015). Analysis of corporate twitter usage types and influence visualization on key players. Proceedings of the IADIS International Conference Web Based Communities, 122-131.

Ha, H., Mun, S., Kwon, O., & Lee, K. (2015). Proximity based circular visualization for similarity analysis of UNGA voting patterns. Proceedings of the IEEE International Symposium on Big Data Visual Analytics (BDVA 2015), 127-128.

Choi, H. Ha, H., Mun, S., & Lee, K. (2015). Proximity based circular visualization for similarity analysis of voting patterns between nations in UN General Assembly. Journal of Society of Design Convergence, 14(4), 133-150.

Mun, S., Ha, H., & Lee, K. (2015). Relationship analysis between the box office performance and sentimental words in movie review. Journal of Society of Design Convergence, 14(4), 1-16.

Mun, S., Han, H., Ha, H., & Lee, K. (2014). A visual analysis on factors affecting repurchase intention in social commerce. Journal of Society of Design Convergence, 13(6), 139-152.

## **PRESENTATIONS Major**

Mun, S. (2021, NOV). How does BERT address polysemy of Korean adverbial postpositions? Paper presented at the 2021 LSK Young Scholar Symposium, The linguistic society of Korean, Korea.

Classification model, Neural language models, BERT, KoBERT, t-SNE, Web-based visualization system

Shin, G-H., & Mun, S. (2020, JUN). Bayesian modelling of clause-level constructional knowledge for Korean-speaking preschool children. Poster presented at the IMPRS Conference 2020, Max Planck Institute for Psycholinguistics.

Bayesian modeling, Language development, Child language data, Korean

Mun, S., & Shin, G-H. (2020, JUN). Context window and polysemy interpretation: A case of Korean adverbial postposition -(u)lo. Poster presented at the IMPRS Conference 2020, Max Planck Institute for Psycholinguistics.

Distributional semantic models, Context window, Similarity-based estimate

Guo, W., Kim, S., Mun, S., & Lee, K. (2018, OCT). BubbleUp: Toward better analysis for temporal event data. Poster presented at the IEEE VIS2018.

Longitudinal data, Diagnosis records, Visualization system, K-means clustering

Mun, S., Wang, I., Desagulier, G., Choi, G., & Lee, K. (2018, JUL). Time flies like an arrow and fruit flies like a banana; parsing multiword constructions with DepVis. Paper presented at the 10th International Conference on Construction Grammar (ICCG10), Paris Sorbonne University.

Multiword expressions, N-grams, POS patterns, Dependency structures, Web application, Language recognition system, English

### **Others**

Shin, G-H., Jung, B. K., & Mun, S. (2022, AUG). Transformer-architecture-based text similarity and L2 proficiency. EuroSLA 31. University of Fribourg, Switzerland.

Mun, S., & Ko, E-S. (2022, JUL). Adaptation of maternal speech in statistical word segmentation of Korean. The 2022 Congress of the International Congress of Infant Studies (ICIS2022), Ottawa, Canada.

Shin, G-H., Jung, B. K., & Mun, S. (2022, June). To what extent neural network models reveal L2 constructs? Relationship between text similarity and learner proficiency. Paper presenting at the 9th edition of the Grammar and Corpora (CaC) conference.

Mun, S. (2022, June). Polysemy resolution by using similarity-based estimation. LSK2022 Summer Conference: Artificial Intelligence and Linguistics, Online, South Korea.

Shin, G-H., & Mun, S. (2021, OCT). Korean-speaking children's constructional knowledge about transitive events: Corpus analysis and Bayesian modelling. Paper presented at the 30th Association for Korean Studies in Europe, La Rochelle University, France.

Jung, B. K., Shin, G-H., & Mun, S. (2021, AUG). NLP-based measurement of text quality for learner writing: Relationship between text similarity and proficiency. Paper presented at the 22nd Biennial Meeting of the International Circle of Korean Linguistics.

Mun, S., & Shin, G-H. (2021, AUG). How does context window size address polysemy of adverbial postposition -(u)lo in Korean? Paper presented at the 11th International Conference on Construction Grammar (ICCG11).

Shin, G-H., & Mun, S. (2021, AUG). Bayesian simulation of Korean-speaking children's constructional knowledge about transitive events. Paper presented at the 11th International Conference on Construction Grammar (ICCG11).

Shin, G-H., & Mun, S. (2020, NOV). Bayesian simulation of clause-level constructional knowledge in child language development: Active transitives and suffixal passives in Korean. Poster presented at the 45th Annual Boston University Conference on Language Development (BUCLD).

Mun, S. (2019, NOV). A computational approach to resolve the polysemy of postpositions in Korean. Poster presented at the COLDOC 2019, Colloque de doctorants et de jeunes chercheurs en Sciences du langage.

Mun, S., Wang, I., Desagulier, G., Choi, G., & Lee, K. (2018, DEC). Fruit flies like a banana: Parsing multiword constructions with DepVis. Paper presented at the Digits, internal reflection, and the study of linguistics., Hankuk University of Foreign Studies.

Choi, G., Yoon, T., Park, C., Lee, S., Choi, H., Park, C., Kang, J., Mun, S., & Lee, K. (2018, DEC). A visual analysis of powerRelations among people on time-series data using network visualization: Focusing on 19th national assembly data in Korea. Paper presented at the 43rd Annual Meeting of the Social Science History Association.

Desagulier, G., Lacheret-Dujour, A., Isel, F., & Mun, S. (2018, MAY). Characterizing discourse genres with prosodic features in a reference treebank of spoken French. Paper presented at the Aflicojet2018, Corpora and Representativeness.

Mun, S., Desagulier, G., & Lee, K. (2017, OCT). How can we capture multiword expressions? Poster presented at the 5th International Conference on Statistical Language and Speech Processing (SLSP2017).

Mun, S., Desagulier, G., Lacheret-Dujour, A., & Lee, K. (2017, JUN). How can we capture multiword expressions? Paper presented at the day of Ph.D. students in MODYCO.

Mun, S., Han, H., Choi, G., Yoon, T., Kim, J., & Lee, K. (2016, NOV). Hierarchical network visualization approach to power mechanism in medieval Joseon Korea. Paper presented at the 41st Annual Meeting of the Social Science History Association.

Bae, S., Mun, S., Choi, G., Lim, S., Bang, S., Son, S., Hong, C., Shin, H., & Lee, K. (2016, OCT). Decision tree analysis utilizing sankey diagram. Poster presented at the Academic Association of Global Cultural Contents (AAGCC).

Choi, S., Kim, J., Kim, S., Mun, S., Kim, K., & Lee, K. (2016, AUG). Analysis of linguistic difference in writing primed with regulatory focus. Poster presented at the 6th International Conference on Language, Literature, Culture and Education (ICLLCE) 2016.

Mun, S., Choi, G., Han, H., Kim, J., & Lee, K. (2016, JUN). VoteStreamVis: Visual analysis of congressional votes in the annals of the Joseon Dynasty. Paper presented at the Interdisciplinary Research Methodology for 'Big Data' in Historical Studies; The beginning of 'Digital History'.

Mun, S., Kim, G., Karim, R. M., & Lee, K. (2015, OCT). Internet review opinion mining and data visualization. Poster presented at the Symposium on Visualization in Data Science (VDS at IEEE VIS 2015).

Lee, S., Choi, H., Mun, S., Han, H., Ha, H., & Lee, K. (2014, DEC). Visualization approach to a Korean genealogy data. Paper presented at the 1st International Symposium on Digital Humanities, Digital Humanities and Culture: Trends and Prospects.

## **TEACHING EXPERIENCE**

Research Methods for English Studies | Chosun University, MAR 2022 ~ Present

- Aim: To introduce how to analyze language data by using R & Python
- Target: Graduate
- Language(s) / package(s) used: R & Python / wordcloud, KoNLP, stringr, RCurl, rsconnect, shiny, pandas, konlpy, BeautifulSoup, selenium, etc.
- Contents: Web crawling / POS tagging / Regular expression, Data processing, morpheme analysis

R Data/Text mining & Visualization | Ajou University, JAN 2022 ~ FEB 2022

- Aim: To introduce how to analyze language data in an R environment
- Target: Undergraduate
- Language(s) / package(s) used: R / wordcloud, KoNLP, stringr, RCurl, rsconnect, shiny, etc.
- Contents: POS tagging / dependency parsing / web crawling / word cloud

Take recommendations for lunch by using Python | Chosun University, JAN 2022 ~ JAN 2022

- Aim: Take recommendations for lunch by using Python
- Target: Undergraduate
- Language(s) / package(s) used: Python / colab, pandas, konlpy, BeautifulSoup, selenium, etc.
- Contents: Web crawling / POS tagging / Regular expression, Data processing, morpheme analysis



Exploring History through Statistics | Korea University, OCT 2018 ~ NOV 2018

- Aim: To create a dataset on Korean history and analyze the data automatically
- Target: Graduate
- Language(s) / package(s) used: R / MASS, ggplot2, scatterplot3d, lme4, Rcurl, stringr, etc.
- Contents: Web crawling / POS tagging / regression analysis / cluster analysis

Recipes: Data Handling and Analysis with R | Ajou University, JUL 2018 ~ AUG 2018

- Aim: To introduce how to handle/analyze data through common data mining techniques
- Target: Undergraduate
- Language(s) / package(s) used: R / KoNLP, stringr, Rcurl, d3Network, wordcloud, etc.
- Contents: N-gram extraction / word cloud / network visualization

Recipes: Text Mining with R | Ajou University, MAY 2018 ~ JUN 2018

- Aim: To introduce how to analyze language data in an R environment
- Target: Undergraduate
- Language(s) / package(s) used: R / wordcloud, KoNLP, stringr, Rcurl, rsconnect, shiny, etc.
- Contents: POS tagging / dependency parsing / web crawling / word cloud

Text Mining for Historical Data | Ajou University, MAR 2018 ~ JUN 2018, MAR 2016 ~ JUN 2016, SEP 2015 ~ DEC 2015

- Aim: To introduce data mining techniques for historical data
- Target: Undergraduate
- Language(s) / package(s) used: R / wordcloud, KoNLP, stringr, Rcurl, rsconnect, shiny, etc.
- Contents: Word cloud / POS tagging / dependency parsing / network visualization

Data Analysis through Visualization | Lecipes, JAN 2016 ~ MAR 2016

- Aim: To train people for acquiring commonly used data visualization techniques, in combination with statistical analysis, in order to apply this knowledge to their workplace
- Target: Industry workers
- Language(s) / package(s) used: R / MASS, ggplot2, lme4, Rcurl, stringr, corrplot, hebin
- Contents
  - Visualization: Hitmap, Hebin, MDS
  - Statistics: Correlation, Regression, Cluster





Research Assistance Scholarship (EUR 2,500; app. USD 3,000) |  
Department of Linguistics, Université Sorbonne Nouvelle Paris  
3; NOV 2018 ~ FEB 2019

Research Assistance Scholarship (KRW 1,000,000; app. USD 900) |  
Ajou University; AUG 2018

Research Grant (KRW 2,000,000; app. USD 1,800) | National  
Research Foundation of Korea, Brain Korea 21; JUL 2018 ~  
AUG 2018

Research Assistance Scholarship (KRW 3,000,000; app. USD 2,700)  
| Department of History, Ajou University; JUN 2018 ~ SEP  
2018

Ph.D. Research Scholarship (KRW 7,000,000; app. USD 6,300) |  
National Research Foundation of Korea, Brain Korea 21, FEB  
2018 ~ AUG 2018

Research Assistance Scholarship (KRW 1,500,000; app. USD 1,350)  
| Department of History, Seongbuk-gu government public  
relations division; NOV 2017 ~ MAR 2018

Full Scholarship for Registration (KRW 5,000,000; app. USD 4,500) |  
Ajou University; SEP 2017

Full Scholarship for Registration (KRW 5,000,000; app. USD 4,500) |  
Ajou University; FEB 2017

Ajou Undergraduate Research Award (KRW 1,000,000; app. USD  
900) | Ajou University; DEC 2016

Full Scholarship for Registration (KRW 5,000,000; app. USD 4,500) |  
Ajou University; SEP 2016

Research Assistance Scholarship (KRW 1,800,000; app. USD 1,620)  
| Leaders in INdustry-university Cooperation (LINC); MAR  
2016 ~ AUG 2016

Ph.D. Research Scholarship (KRW 7,000,000; app. USD 6,300) |  
National Research Foundation of Korea, Brain Korea 21, FEB  
2016 ~ AUG 2016

80% Scholarship for Registration (KRW 4,000,000; app. USD 3,600) |  
National Research Foundation of Korea, Brain Korea 21; FEB  
2016

Research Assistance Scholarship (KRW 1,000,000; app. USD 900) |  
Sungshin Women's University; DEC 2015

Half Scholarship for Registration (KRW 2,500,000; app. USD 2,250) |  
National Research Foundation of Korea, Brain Korea 21; SEP  
2015

Research Assistance Scholarship (KRW 4,000,000; app. USD 3,600)  
| Korea Meteorological Administration; MAR 2015 ~ DEC 2015

Half Scholarship for Registration (KRW 2,500,000; app. USD 2,250) |  
National Research Foundation of Korea, Brain Korea 21; FEB  
2015

Full Scholarship for Registration (KRW 5,000,000; app. USD 4,500) |  
Ajou University; SEP 2014

Special Admission Scholarship (KRW 5,000,000; app. USD 4,500) |  
Ajou University; FEB 2014

Half Scholarship for Registration (KRW 2,000,000; app. USD 1,800) |  
Pyeongtaek University; SEP 2013

Full Scholarship for Registration (KRW 4,000,000; app. USD 3,600) |  
Pyeongtaek University; FEB 2013

Full Scholarship for Registration (KRW 4,000,000; app. USD 3,600) |  
Pyeongtaek University; SEP 2012

**CERTIFICATIONS** Oracle Certified JAVA Programmer, Oracle, OCT 2014

Word Processor Lv.1, Korea chamber of Commerce & Industry, OCT  
2013

SAS Certified Base Programmer, SAP Korea, AUG 2013

Survey Analyst (Junior), Human Resources Development Service of  
Korea, MAY 2013