

# TAEUK KIM (김태욱)

Assistant Professor at Hanyang University

Room 719, IT/BT Bldg., 222, Wangsimni-ro, Seongdong-gu, Seoul, Korea

🏠 <https://nlp.hanyang.ac.kr> ◊ ✉ [kimtaeuk@hanyang.ac.kr](mailto:kimtaeuk@hanyang.ac.kr)

## RESEARCH INTERESTS

---

Natural Language Processing based on Machine Learning & Deep Learning. Especially interested in

- Lexical/sentence-level semantics (e.g., representation learning, language modeling)
- Utilization and analysis of large language models (e.g., in-context learning, parameter-efficient tuning)
- Associating linguistic knowledge with neural networks (e.g., unsupervised parsing, tree-based models)
- Interpretation and analysis of NLP models (e.g., probing)
- Multilingual/Korean NLP (e.g., cross-lingual transfer, tokenization)

## AFFILIATION

---

**Hanyang University (Seoul, Korea)**

*Sep. 2021 -*

*Assistant Professor*

- Main department: Department of Computer Science
- Joint appointment
  - Department of Artificial Intelligence
  - Department of AI Application (collaboration with KT)
  - Department of AI Semiconductor
  - Department of Data Science
  - Graduate School of Applied Artificial Intelligence

## EDUCATION

---

**Seoul National University (Seoul, Korea)**

*Mar. 2016 - Aug. 2021*

*M.S. & Ph.D. (integrated) in Computer Science and Engineering*

- **Advisor:** Sang-goo Lee
- **Thesis:** Learning and Analysis of Neural Sentence Representations Using Syntax (**Committee:** Hyoung-Joo Kim, Sang-goo Lee, Seung-won Hwang, Gunhee Kim (SNU), and Minjoon Seo (KAIST))

**Seoul National University (Seoul, Korea)**

*Mar. 2009 - Feb. 2016,*

*B.S. in Computer Science and Engineering*

## PUBLICATIONS

---

†: Corresponding author

\*: Equal contribution

### Preprint (arXiv)

1. Hyuhng Joon Kim, Youna Kim, Cheonbok Park, Junyeob Kim, Choonghyun Park, Kang Min Yoo, Sang-goo Lee, and **Taeuk Kim**<sup>†</sup>, Aligning Language Models to Explicitly Handle Ambiguity, *arXiv:2404.11972*

### International Conferences

1. Young Hyun Yoo, Jii Cha, Changhyeon Kim, and **Taeuk Kim**<sup>†</sup>, Hyper-CL: Conditioning Sentence Representations with Hypernetworks, *The 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*.
2. Seong Hoon Lim, Taejun Yun, Jinhyeon Kim, Jihun Choi, and **Taeuk Kim**<sup>†</sup>, Analysis of Multi-Source Language Training in Cross-Lingual Transfer, *The 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*.
3. Yejin Yoon, Jungyeon Lee, Kangsan Kim, Chanhee Park, and **Taeuk Kim**<sup>†</sup>, BlendX: Complex Multi-Intent Detection with Blended Patterns, *The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*.
4. Taejun Yoon, Jinhyeon Kim, Deokyeong Kang, Seong Hoon Lim, Jihoon Kim, and **Taeuk Kim**<sup>†</sup>, X-SNS: Cross-Lingual Transfer Prediction through Sub-Network Similarity, *Findings of the Association for Computational Linguistics: EMNLP 2023 (Findings of EMNLP 2023)*.
5. Hyuhng Joon Kim, Hyunsoo Cho, Sang-Woo Lee, Junyeob Kim, Choonghyun Park, Sang-goo Lee, Kang Min Yoo<sup>†</sup>, and **Taeuk Kim**<sup>†</sup>, Universal Domain Adaptation for Robust Handling of Distributional Shifts in NLP, *Findings of the Association for Computational Linguistics: EMNLP 2023 (Findings of EMNLP 2023)*.
6. Hyunsoo Cho, Hyuhng Joon Kim, Junyeob Kim, Sang-Woo Lee, Sang-goo Lee, Kang Min Yoo<sup>†</sup>, and **Taeuk Kim**<sup>†</sup>, Prompt-Augmented Linear Probing: Scaling Beyond The Limit of Few-shot In-Context Learners, *The 37th AAAI Conference on Artificial Intelligence (AAAI 2023)*.
7. Kang Min Yoo<sup>†</sup>, Junyeob Kim, Hyuhng Joon Kim, Hyunsoo Cho, Hwiyeol Jo, Sang-Woo Lee, Sang-goo Lee, and **Taeuk Kim**<sup>†</sup>, Ground-Truth Labels Matter: A Deeper Look into Input-Label Demonstrations, *The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*.
8. Hyunsoo Cho, Choonghyun Park, Jaewook Kang, Kang Min Yoo, **Taeuk Kim**<sup>†</sup>, and Sang-goo Lee, Enhancing Out-of-Distribution Detection in Natural Language Understanding via Implicit Layer Ensemble, *Findings of the Association for Computational Linguistics: EMNLP 2022 (Findings of EMNLP 2022)*.
9. **Taeuk Kim**<sup>†</sup>, Revisiting the Practical Effectiveness of Constituency Parse Extraction from Pre-trained Language Models, *The 29th International Conference on Computational Linguistics (COLING 2022)*.
10. **Taeuk Kim**, Bowen Li, and Sang-goo Lee, Multilingual Chart-based Constituency Parse Extraction from Pre-trained Language Models, *Findings of the Association for Computational Linguistics: EMNLP 2021 (Findings of EMNLP 2021)*.
11. **Taeuk Kim**, Kang Min Yoo, and Sang-goo Lee, Self-Guided Contrastive Learning for BERT Sentence Representations, *The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)*.
12. Bowen Li, **Taeuk Kim**, Reinald Kim Amplayo and Frank Keller, Heads-up! Unsupervised Constituency Parsing via Self-Attention Heads, *The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (AACL-IJCNLP 2020)*.
13. **Taeuk Kim**, Jihun Choi, Daniel Edmiston, and Sang-goo Lee, Are Pre-trained Language Models Aware of Phrases? Simple but Strong Baselines for Grammar Induction, *International Conference on Learning Representations 2020 (ICLR 2020)*.

14. Jihun Choi, **Taeuk Kim**, and Sang-goo Lee, Cell-aware Stacked LSTMs for Modeling Sentences, *The 11th Asian Conference on Machine Learning (ACML 2019)*.
15. **Taeuk Kim\***, Kang Min Yoo\*, and Sang-goo Lee, Don't Just Scratch the Surface: Enhancing Word Representations for Korean with Hanja, *Conference on Empirical Methods in Natural Language Processing 2019 and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)*.
16. Daniel Edmiston and **Taeuk Kim**, Intrinsic Evaluation of Grammatical Information within Word Embeddings, *The 33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33)*.
17. Jihun Choi, **Taeuk Kim**, and Sang-goo Lee, A Cross-Sentence Latent Variable Model for Semi-Supervised Text Sequence Matching, *The 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*.
18. **Taeuk Kim**, Jihun Choi, Daniel Edmiston, SangHwan Bae, and Sang-goo Lee, Dynamic Compositionality in Recursive Neural Networks with Structure-aware Tag Representations, *Thirty-Third AAAI Conference on Artificial Intelligence (AAAI 2019)*.
19. Jihun Choi, **Taeuk Kim**, and Sang-goo Lee, Element-wise Bilinear Interaction for Sentence Matching, *The Seventh Joint Conference on Lexical and Computational Semantics (\*SEM 2018) at NAACL HLT 2018*.

### International Workshops

1. Hyuhng Joon Kim, Hyunsoo Cho, Junyeob Kim, **Taeuk Kim**, Kang Min Yoo, and Sang-goo Lee, Self-Generated In-Context Learning: Leveraging Auto-regressive Language Models as a Demonstration Generator, *Workshop on Large-scale Pre-trained Language Models (LPLM 2022) at NAACL 2022*.
2. Youngju Joung and **Taeuk Kim**<sup>†</sup>, HYU at SemEval-2022 Task 2: Effective Idiomaticity Detection with Consideration at Different Levels of Contextualization, *International Workshop on Semantic Evaluation (SemEval 2022) at NAACL 2022*.
3. Jaeyoul Shin, **Taeuk Kim**, and Sang-goo Lee, IDS at SemEval-2020 Task 10: Does Pre-trained Language Model Know What to Emphasize?, *International Workshop on Semantic Evaluation (SemEval 2020) at COLING 2020*.
4. Sanghwan Bae, **Taeuk Kim**, Jihoon Kim, and Sang-goo Lee, Summary Level Training of Sentence Rewriting for Abstractive Summarization, *The Second Workshop on New Frontiers in Summarization (NewSum 2019) at EMNLP-IJCNLP 2019*.
5. **Taeuk Kim**, Jihun Choi, and Sang-goo Lee, SNU\_IDS at SemEval-2018 Task 12: Sentence Encoder with Contextualized Vectors for Argument Reasoning Comprehension, *International Workshop on Semantic Evaluation (SemEval 2018) at NAACL HLT 2018*.
6. Sanghyuk Choi, **Taeuk Kim**, Jinseok Seol, and Sang-goo Lee, A Syllable-based Technique for Word Embeddings of Korean Words, *The 1st Workshop on Subword and Character level models in NLP (SCLeM 2017) at EMNLP 2017*.

### Domestic Conferences

1. **[Best Paper Award]** Yejin Yoon\*, Jisu Kim\*, Jungmin Im, Jungyeon Lee, and **Taeuk Kim**<sup>†</sup>, Multi-Intent Detection for Korean Spoken Language, *Korea Computer Congress 2024 (KCC 2024)*.
2. Jungyeon Lee, Youngwoo Shin, Yejin Yoon, and **Taeuk Kim**<sup>†</sup>, Enhanced Prompting for Multi-Label Classification, *Korea Computer Congress 2024 (KCC 2024)*.

3. Sungmin So\*, Jiwoo Min\*, Yejin Yoon, Jungyeon Lee, and **Taeuk Kim**<sup>†</sup>, A Multi-Intent Dataset Enhanced with Implicit Concatenation, *Korea Computer Congress 2024 (KCC 2024)*.
4. Kang Min Lee, Dong Geon Seo, Eunrang Kwon, Junmo Song, Jeonghan Kang, and **Taeuk Kim**<sup>†</sup>, Korean Semantic Role Labeling with Machine Reading Comprehension, *Korea Computer Congress 2024 (KCC 2024)*.
5. Jii Cha and **Taeuk Kim**<sup>†</sup>, Enhancing Sentence Representations with Common Error Analysis and Ensemble Techniques in Contrastive Learning, *Korea Software Congress 2023 (KSC 2023)*.
6. Taejoon Yoon and **Taeuk Kim**<sup>†</sup>, Research on Features for Effective Cross-Lingual Transfer in Korean, *The 34th Annual Conference on Human and Cognitive Language Technology (HCLT 2023)*.
7. Jinhyeon Kim and **Taeuk Kim**<sup>†</sup>, MAdapter: A Refinement of Adapters by Augmenting Efficient Middle Layers, *The 34th Annual Conference on Human and Cognitive Language Technology (HCLT 2023)*.
8. Seong Hoon Lim and **Taeuk Kim**<sup>†</sup>, Enhanced Zero-Shot Cross-Lingual Transfer with the Diversification of Source Language, *Korea Computer Congress 2023 (KCC 2023)*.
9. Young Hyun Yoo, Kyumin Lee, Minjin Jeon, Jii Cha, Kangsan Kim, and **Taeuk Kim**<sup>†</sup>, Comparison and Analysis of Unsupervised Contrastive Learning Approaches for Korean Sentence Representations, *The 34th Annual Conference on Human and Cognitive Language Technology (HCLT 2022)*.

## RESEARCH PROJECTS (AS PRINCIPAL INVESTIGATOR)

---

<b>Research on Algorithms for Multi-Intent Detection</b>	<i>2023-2024</i>
<i>Funded by Hyundai Motors</i>	
<b>Advancing Vehicle-Embedded NLP Engines via Better Utilization of Pre-trained Language Models</b>	<i>2022-2023</i>
<i>Funded by Hyundai Motors</i>	
<b>Analysis and Development of Efficient Methods for Learning High-Quality Sentence Representations Based on Large-Scale Language Models</b>	<i>2021-2023</i>
<i>Funded by National Research Foundation of Korea</i>	

## AWARDS AND HONORS

---

<b>Outstanding Dissertation Award</b>	<i>2021</i>
<i>Department of Computer Science and Engineering, Seoul National University</i>	
<b>NAVER Ph.D. Fellowship Award</b>	<i>2020</i>
<i>NAVER Corp.</i>	
<b>International Conference on Learning Representations 2020 Travel Award</b>	<i>2020</i>
<i>International Conference on Learning Representations</i>	
<b>National Science &amp; Technology Scholarship</b>	<i>2015</i>
<i>Korea Student Aid Foundation (KOSAF)</i>	

## ACADEMIC SERVICES

---

### Area Chair & Action Editor

- [Action Editor] ACL Rolling Review (**ARR**): 2021-2024
  - [Area Chair]: **ACL** (2022, 2024), **EACL** (2024), **EMNLP** (2024), **NAACL** (2024)
- [Area Chair] Empirical Methods in Natural Language Processing (**EMNLP**): 2022
- [Session Chair]
  - International Conference on Computational Linguistics (**COLING**): 2022

- Korea Computer Congress (**KCC**): 2024

## Reviewer

- **Top NLP & ML Conferences**

- AAAI Conference on Artificial Intelligence (**AAAI**): 2020-2024
- Annual Meeting of the Association for Computational Linguistics (**ACL**): 2021, 2023
- International Conference on Computational Linguistics (**COLING**): 2020, 2022, 2024, 2025
- Conference on Language Modeling (**COLM**): 2024
- Empirical Methods in Natural Language Processing (**EMNLP**): 2021, 2023
- International Conference on Representation Learning (**ICLR**): 2021-2024
- International Conference on Machine Learning (**ICML**): 2022-2024
- Conference on Neural Information Processing Systems (**NeurIPS**): 2021-2024

- **Journal & Extra Conferences**

- Bioinformatics: 2022
- Journal of Cognitive Science: 2022
- Annual Conference on Human & Cognitive Language Technology (**HCLT**): 2022-2023
- Korea Software Congress (**KSC**): 2023
- Korea Computer Congress (**KCC**): 2024

## EXTRA INFORMATION

---

- **Personal Page:** <http://galsang.github.io>
- **Lab. Page:** <http://nlp.hanyang.ac.kr>
- **Github:** <https://github.com/galsang>
- **Google Scholar:** <https://scholar.google.com/citations?user=eH5uq7wAAAAJ>
- **Semantic Scholar:** <https://www.semanticscholar.org/author/Taeuk-Kim/5041757>

(Last update: 15/June/2024)