

# Chanjun Park

UPSTAGE

338, Gwanggyojungang-ro, Suji-gu, Yongin-si, Gyeonggi-do, Republic of Korea

☎ (+82) 10-2580-9066 | ✉ [chanjun.park@upstage.ai](mailto:chanjun.park@upstage.ai) / [bcj1210@naver.com](mailto:bcj1210@naver.com) | 📱 [parkchanjun](#) | 📺 [bcj1210](#)

## Research Interests

Chanjun Park is a researcher in the field of Natural Language Processing (NLP), with a focus on Data-Centric AI and Machine Translation. He is currently working as an AI Research Engineer at Upstage since March 2022. In 2023, he received Ph.D. from Korea University under the supervision of Professor Heuseok Lim for the work on “Data-Centric Neural Machine Translation”. From 2018 to 2019, he worked at SYSTRAN as a Research Engineer. Chanjun is the founder and leader of the KU-NMT Group, and has received the Naver Ph.D. Fellowship in 2021. He served as the Virtual Social Chair at COLING 2022, and is currently serving as the Program Chair for the WinLP Workshop. He has published more than 140 papers in the field of NLP. See CV for more information.

- Data-centric AI
- Machine Translation
- Large Language Model (LLM)
- Real-World-centric AI

## Education

### Korea University

PH.D IN COMPUTER SCIENCE AND ENGINEERING, GPA: 4.44 / 4.5

Seoul, South Korea

Sep. 2019 - Aug. 2023

- Advisor: Heuseok Lim
- Natural Language Processing & Artificial Intelligence Lab
- Founder and Leader of KU-NMT Group.

### Busan University of Foreign Studies (BUFS)

B.S. IN NATURAL LANGUAGE PROCESSING & CREATIVE CONVERGENCE, GPA: 4.2 / 4.5

Busan, South Korea

Mar. 2015 - Aug. 2018

- Advisor: Pummo Ryu
- Full Scholarship

## Academic Services

### Widening NLP (WinLP) Workshop

PROGRAM CHAIR

Anywhere on Earth

March.2023-Present

- WinLP: <https://www.winlp.org/>

### COLING 2022

VIRTUAL SOCIAL CHAIR

South Korea

Sep.2021-Present

- COLING 2022: <http://coling2022.org/>
- Virtual Social Chair

### Program committe

PROGRAM COMMITTEE

South Korea

January.2021-Present

- ICML 2023-ICML 2023-Data-centric Machine Learning Research (DMLR), ACL 2023-Industry Track, ACL 2023-BEA, EACL 2023, EACL 2023-Insights, ACL 2022-Insights, NAACL 2022-Industry Track, Reviewer of ARR (2022 Current)

### Reviewer

REVIEWER

South Korea

January.2021-Present

- IEEE Access, International Journal of Human-Computer Interaction, Cybernetics and Systems, Applied Sciences, Electronics, Entropy, Connection Science

## Selected Publications

### Mimicking Infants' Bilingual Language Acquisition for Domain Specialized Neural Machine Translation

Chanjun Park, Woo-Young Go, Sugyeong Eo, Hyeonseok Moon, Seolhwa Lee, Heuseok Lim

IEEE Access 10 (2022) pp. 38684–38693. IEEE, 2022

### FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue

Chanjun Park, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuseok Lim

LREC 2022 (Proceedings of the Language Resources and Evaluation Conference), 2022, Marseille, France

## **PicTalky: Augmentative and Alternative Communication for Language Developmental Disabilities**

**Chanjun Park**, Yoonna Jang, Seolhwa Lee, Jaehyung Seo, Kisu Yang, Heuseok Lim

*AACL-IJCNLP 2022: System Demonstrations*, 2022, Taipei, Taiwan

## **Priming Ancient Korean Neural Machine Translation**

**Chanjun Park**, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Heuseok Lim

*IREC 2022 (Proceedings of the Language Resources and Evaluation Conference)*, 2022, Marseille, France

## **The ASR Post-Processor Performance Challenges of BackTranScripton (BTS): Data-Centric and Model-Centric Approaches**

**Chanjun Park**, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Heuseok Lim

*Mathematics* 10.19 (2022). 2022

## **Empirical Analysis of Parallel Corpora and In-Depth Analysis Using LIWC**

**Chanjun Park**, Midan Shim, Sugyeong Eo, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Heuseok Lim

*Applied Sciences* 12.11 (2022) p. 5545. Multidisciplinary Digital Publishing Institute, 2022

## **QUAK: A Synthetic Quality Estimation Dataset for Korean-English Neural Machine Translation**

Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Gyeongmin Kim, Jungseob Lee, Heui-Seok Lim

*COLING 2022 - Proceedings of the 29th International Conference on Computational Linguistics*, 2022

## **KU X Upstage's Submission for the WMT22 Quality Estimation: Critical Error Detection Shared Task**

Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuseok Lim

*WMT 2022 - EMNLP 2022 - Proceedings of the Seventh Conference on Machine Translation*, 2022, Abu Dhabi

## **Focus on FoCUS: Is FoCUS focused on Context, Knowledge and Persona?**

SeungYoon Lee, Jungseob Lee, **Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Jaehyung Seo, Jeongbae Park, Heui-Seok Lim

*COLING 2022 - Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge*, 2022

## **A Self-Supervised Automatic Post-Editing Data Generation Tool**

Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, SeungJun Lee, Heuseok Lim

*ICML 2022 - DataPerf workshop*, 2022 (2022). 2022

## **Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing**

Hyeonseok Moon, **Chanjun Park**, Seolhwa Lee, Jaehyung Seo, Jungseob Lee, Sugyeong Eo, Heuseok Lim

*IREC 2022 (Proceedings of the Language Resources and Evaluation Conference)*, 2022, Marseille, France

## **A Dog Is Passing Over The Jet? A Text-Generation Dataset for Korean Commonsense Reasoning and Evaluation**

Jaehyung Seo, Seounghoon Lee, **Chanjun Park**, Yoonna Jang, Hyeonseok Moon, Sugyeong Eo, Seonmin Koo, Heuseok Lim

*Findings of the Association for Computational Linguistics: NAACL 2022*, 2022, Seattle, United States

## **PU-GEN: Enhancing generative commonsense reasoning for language models with human-centered knowledge**

Jaehyung Seo, Dongsuk Oh, Sugyeong Eo, **Chanjun Park**, Kisu Yang, Hyeonseok Moon, Kinam Park, Heuseok Lim

*Knowledge-Based Systems* (2022) p. 109861. Elsevier, 2022

## **Should we find another model?: Improving Neural Machine Translation Performance with ONE-Piece Tokenization Method without Model Modification**

**Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim

*NAACL 2021 - Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*, 2021

## **FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue**

**Chanjun Park**, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuseok Lim

*AAAI 2022 - Artificial Intelligence for Education(AI4EDU)* (2021). 2021

## **How should human translation coexist with NMT? Efficient tool for building high quality parallel corpus**

**Chanjun Park**, Seolhwa Lee, Hyeonseok Moon, Sugyeong Eo, Jaehyung Seo, Heuseok Lim

*NeurIPS 2021 - Data-centric AI (DCAI) workshop* (2021). 2021

## **Two Heads are Better than One? Verification of Ensemble Effect in Neural Machine Translation**

**Chanjun Park**, Sungjin Park, Seolhwa Lee, Taesun Whang, Heuseok Lim

*EMNLP 2021 - Proceedings of the Second Workshop on Insights from Negative Results in NLP*, 2021, Online and Punta Cana, Dominican Republic

## **BTS: Back TranScripton for Speech-to-Text Post-Processor using Text-to-Speech-to-Text**

**Chanjun Park**, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo, Heuseok Lim

*ACL 2021 - Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, 2021, Online

## **Dealing with the Paradox of Quality Estimation**

Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuseok Lim, **(Equal Contribution (First Co-Author))**

*MT Summit - Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021)*, 2021, Virtual

## **A New Tool for Efficiently Generating Quality Estimation Datasets**

Sugyeong Eo, Chanjun Park, Jaehyung Seo, Hyeonseok Moon, Heuseok Lim

*NeurIPS 2021 - Data-centric AI (DCAI) workshop*, 2021

## **Syntax-enhanced Dialogue Summarization using Syntax-aware information**

Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heui-Seok Lim

*NeurIPS 2021 16th Women in Machine Learning Workshop*, 2021

## **Towards Syntax-Aware Dialogue Summarization using Multi-task Learning**

Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heui-Seok Lim

*EMNLP 2021 - Widening NLP (WinLP2021) workshop*, 2021

## **Who speaks like a style of Vitamin: Towards Syntax-Aware Dialogue Summarization using Multi-task Learning**

Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heuseok Lim

#### Automatic Knowledge Augmentation for Generative Commonsense Reasoning

Jaehyung Seo, **Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim

NeurIPS 2021 - Data-centric AI (DCAI) workshop, 2021

#### Neural Spelling Correction: Translating Incorrect sentences to Correct sentences for Multimedia

**Chanjun Park**, Kuekyeng Kim, YeongWook Yang, Minho Kang, HeuiSeok Lim

Multimedia Tools and Applications (2020) pp. 1–18. Springer, 2020

#### Ancient Korean Neural Machine Translation

**Chanjun Park**, Chanhee Lee, YeongWook Yang, Heuiseok Lim

IEEE Access 8 (2020) pp. 116617–116625. 2020

#### Comparison of the evaluation metrics for Neural Grammatical Error Correction with Overcorrection

**Chanjun Park**, YeongWook Yang, Chanhee Lee, Heuiseok Lim

IEEE Access 8 (2020) pp. 106264–106272. 2020

#### Decoding Strategies for Improving Low-Resource Machine Translation

**Chanjun Park**, Yeongwook Yang, Kinam Park, Heuiseok Lim

Electronics 9.10 (2020) p. 1562. Multidisciplinary Digital Publishing Institute, 2020

## Honors & Awards

2023.02	<b>Best Paper Award</b> , Korea University	South Korea
2023.02	<b>Research Encouragement Scholarship</b> , Korea University	South Korea
2022.12	<b>1st place in WMT 2022 QE Task 3, 2022</b> , WMT 2022 - EMNLP 2022	Abu Dhabi
2022.10	<b>Best Paper Award, 2022</b> , The 34th Annual Conference on Human & Cognitive Language Technology (HCLT2022)	South Korea
2021.12	<b>Naver Ph.D. Fellowship, 2021</b> , Naver	South Korea
2021.10	<b>Best Paper Award (NLP Application 2 Section), 2021</b> , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.10	<b>Best Paper Award (Language Resource Section), 2021</b> , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.10	<b>Best Paper Award (QA and Speech Section), 2021</b> , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.07	<b>Ranked 4th on the CommonGen 1.1 Leaderboard (Nov. 2022 Ranked 7th, CommonGen 1.1)</b> , CommonGen 1.1	South Korea
2020.11	<b>1st Place in Flitto Hackathon (Team Lead), 2020</b> , Flitto	South Korea
2020.10	<b>Best Paper Award, 2020</b> , The 32nd Annual Conference on Human & Cognitive Language Technology (HCLT2020)	South Korea
2020.05	<b>Best practices for using NIA AI training data(Korean-English Neural Machine Translation model)</b> , AI Hub	South Korea
2019.10	<b>Best Paper Award, 2019</b> , The 31st Annual Conference on Human & Cognitive Language Technology (HCLT2019)	South Korea
2019.10	<b>1st Place Microsoft AI Accessibility Hackathon in Korea (Team Lead)</b> , Microsoft	South Korea
2019.03	<b>Graduate School Associate Scholarship</b> , Sungkyunkwan University	South Korea
2018.10	<b>Next Generation Information Processing NLP Competition 2018: Participation Award</b> , Next-generation information computing technology development business	South Korea
2017.06	<b>Bit Computer Excellence Award (President Award)</b> , Bit Computer	South Korea
2015-2019	<b>Full Scholarship</b> , BUFS	South Korea
2016-2017	<b>Scholarship for academic excellence</b> , Sooyoungro Church	South Korea

## Selected Research Experience

### Large Language Model (LLM)

RESEARCHER IN UPSTAGE

- Research for Large Language Model (LLM)
- Skills: Data-centric AI, LLM, NLP

Seoul, South Korea

June. 2023. - Present

### DataOps and Data Management

RESEARCHER IN UPSTAGE

- DataOps research related to OCR and Product Search
- Building training and evaluation data for OCR and Product Search
- Data Management for OCR and Product Search.
- Skills: Data Management, DataOps, OCR, Product Search

Seoul, South Korea

October.2022. - June.2023

### **(Project) LGU + VOD Search POC Project**

Seoul, South Korea

RESEARCHER IN UPSTAGE

June. 2022. - October.2022

- Gold Testset and evaluation system design for tag extraction model comparison
- Responsible for designing search system evaluation. Planning and designing blind tests for comparison with commercial systems.
- Skills: Customer Experience (CX), Quality Assurance (QA)

### **(Project) Amorepacific Search commercialization Project**

Seoul, South Korea

RESEARCHER IN UPSTAGE

March. 2022. - June.2022

- Participated in Amorepacific's AP Mall search commercialization project
- Stability test planning and design
- Response time testing planning and design
- Establishment of systematic QA process based on bug testing
- Skills: Customer Experience (CX), Quality Assurance (QA)

### **Product Search Pack**

Seoul, South Korea

RESEARCHER IN UPSTAGE

March. 2022. - October. 2022

- Develop Product Search Pack
- Establishment of evaluation system to develop Product Search Pack and responsible for data, planning, and evaluation
- Quality Assurance for Search System, Gold Testset Planning
- Skills: Data Manager, User Experience (UX), Customer Experience (CX), Quality Assurance (QA)

### **(Project) Polyglot: Large Language Models of Well-balanced Competence in Multi-languages**

Seoul, South Korea

RESEARCHER IN ELEUTHER AI

September. 2022. - May.2023

- Various multilingual models such as mBERT, BLOOM, and XGLM have been released. Therefore, someone might ask, "why do we need to make multilingual models again?" Before answering the question, we would like to ask, "Why do people around the world make monolingual models in their language even though there are already many multilingual models?" We would like to point out there is a dissatisfaction with the non-English language performance of the current multilingual models as one of the most significant reason. So we want to make multilingual models with higher non-English language performance. This is the reason we need to make multilingual models again and why we name them 'Polyglot'.
- Skills: Language Model, Data management

### **(Project) Korean-Prompt-based Engineering for Few-shot Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

March. 2022. - July. 2022.

- Plain Template Insertion: Korean-Prompt-based Engineering for Few-shot Learners
- Supported by Hyundai Motor
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9913979>
- Skills: Prompt learning, Few-shot learning, Pytorch, Huggingface

### **(Project) Return on Advertising Spend (ROAS) Prediction Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - March. 2022.

- Return on Advertising Spend (ROAS) refers to the ratio of revenue generated by advertising projects to its expense.
- We present a method for predicting ROAS that does not require controlled experiments in data acquisition and validates its effectiveness through comparative experiments. Specifically, we propose a task decomposition method that divides the end-to-end prediction task into the two-stage process; occurrence prediction and occurred ROAS regression.
- Supported by Data voucher
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/10/10/1637>
- Skills: ROAS, Regression, Pytorch

### **(Project) Hybrid Retriever System for Industrial Frequently Asked Questions Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - March. 2022.

- We propose a method of returning responses to customer inquiries using only data that can be easily obtained from companies based on the deep learning information retrieval module. In this work, we hybridize dense embedding and sparse embedding to make it robust to professional terms and suggest new functions to scale the results returned by the two modules and adjust the weight ratio.
- Supported by Data voucher
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/10/8/1335>
- Skills: Information Retrieval, Pytorch

### **Fake News Detection Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2022. - Present

- Research for Utilization Strategy of User Engagements in Korean Fake News Detection
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9841541>
- Skills: Graph Neural Network, Pytorch, Fairseq

### **Simultaneous Speech Translation Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - Present

- Research for Simultaneous Speech Translation for MCN Market
- Skills: Neural Machine Translation, Pytorch, Fairseq

## **(Project) NAVER PAPAGO - Parallel Corpus Filtering and Mining**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - June. 2022.

- Research for Static and Dynamic Parallel Corpus Filtering and Mining
- Human-inspired curriculum learning based NMT
- Supported by NAVER PAPAGO
- Skills: Neural Machine Translation, Parallel corpus filtering, Pytorch, Fairseq

## **Data-centric Neural Machine Translation Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

July. 2021. - December. 2021.

- Research for Data-centric Neural Machine Translation, Automatic Post Editing, Quality Estimation, Generative Commonsense Reasoning
- Develop Efficient tool for building high quality parallel corpus and A New Tool for Efficiently Generating Quality Estimation Datasets and A Self-Supervised Automatic Post-Editing Data Generation Tool
- Publication of 3 research paper in NeurIPS 2021 - Data-centric AI (DCAI) workshop: <https://eval.how/dcai2021/index.html>
- Publication of research paper in ICML 2022 - DataPerf workshop: <https://sites.google.com/view/dataperf2022>
- Skills: Neural Machine Translation, Web Application, Fairseq
- Improving Formality-Sensitive Machine Translation using Data-Centric Approaches and Prompt Engineering
- Publication of research paper in IWSLT 2023 : <https://sites.google.com/view/dataperf2022>
- A Survey on Evaluation Metrics for Machine Translation
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/11/4/1006>
- Skills: Neural Machine Translation, Web Application, Fairseq

## **Korean Commonsense Generation and Reasoning Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - Present

- KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation
- Automatic Knowledge Augmentation for Generative Commonsense Reasoning
- Publication of research paper in NAACL 2022: <https://aclanthology.org/2022.findings-naacl.172/>
- Publication of research paper in Knowledge-Based Systems: <https://www.sciencedirect.com/science/article/abs/pii/S0950705122009546>
- Skills: Commonsense Reasoning, Pytorch, Huggingface

## **Image Retrieval Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - September. 2021.

- Grounded Vocabulary for Image Retrieval Using a Modified Multi-Generator Generative Adversarial Network
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9585128>
- Skills: Image Retrieval, Pytorch, Huggingface

## **(Project) Development of patent document invention sentence extraction and key phrase extraction module**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - December. 2021.

- AI for Patent research
- T5 based Key sentence extraction and key phrase generation
- Supported by "LG Innotek"
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9779775>
- Skills: Automatic Summarization, Sentence classification, Information Extraction Pytorch, Huggingface

## **Dialogue Summarization Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - Present

- Syntax-Aware Dialogue Summarization using Multi-task Learning
- Skills: Dialogue system, Automatic Summarization, Pytorch, Huggingface
- Publication of research paper in EMNLP 2021 -Widening NLP (WiNLP2021) workshop
- Publication of research paper in NeurIPS 2021 - Women in Machine Learning (WiML 2021) workshop
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/abstract/document/9664379/>

## **Dialogue System Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2021. - December. 2021

- Robust Reward Estimator for Multidomain Task-Oriented Dialogue
- Publication of research paper in Applied Science (SCIE): <https://www.mdpi.com/2076-3417/11/14/6624>
- BlenderBot 2.0 Errors Analysis in terms of Model, Data and User-Centric Approach
- Publication of research paper in Arxiv: <https://arxiv.org/abs/2201.03239>
- Focus on FoCuS: Is FoCuS focused on Context, Knowledge and Persona?
- Publication of research paper in COLING 2022 Workshop: <https://aclanthology.org/2022.ccgpk-1.1.pdf>
- Skills: Dialogue system, Pytorch, Huggingface, ParlAI

## **(Project) Subtitle production and editing technology development for colloquial broadcasting contents**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2021. - December, 2021.

- Development of Colloquial Aware Neural Machine Translation (CANMT) model that is robust to colloquial expression based on broadcast content
- Consortium: Korea Univ, Hanyang Univ, ON THE IT, SBS Contents Hub, CBS
- Supported by IITP
- Korean Colloquial-Sentence-Splitter(Ko-CSS) for simultaneous speech translation
- Neural Architecture Search(NAS) based NMT Research
- Speech recognition post-processor research
- Cross lingual adaptation research
- Skills: Neural Machine Translation, Automatic Speech Recognition, simultaneous speech translation, Pytorch, Huggingface, Fairseq
- Ko-CSS Demo: <http://nlplab.iptime.org:32285/>

## **BTS: Speech recognition post-processor Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- Proposed BackTranScript(BTS). Back Transcription (BTS) has seen limited success without the role of phonetic transcription for building and creating vast datasets automatically.
- Publication of research paper in WAT 2021 (ACL Workshop): <https://aclanthology.org/2021.wat-1.10/>
- Publication of research paper in Mathematics: <https://www.mdpi.com/2227-7390/10/19/3618>
- Skills: Neural Machine Translation, Pytorch, Huggingface, Fairseq

## **Cross Lingual Pretrained Language Model Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- Research for Exploring the Data Efficiency of Cross-Lingual Post-Training in Pretrained Language Models
- Research for combining BART and Parallel Corpus Filtering
- Publication of research paper in Applied Science (SCIE): <https://www.mdpi.com/2076-3417/11/5/1974>
- Skills: Neural Machine Translation, Pytorch, Huggingface, Fairseq

## **Automatic Post Editing Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- APE research based on Cross Lingual Pretrained Language Model
- Adapter based APE research
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9528385/>
- Research for Efficient and Simple Data Generation Method - Low Resource APE and Pseudo APE data generation
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9714400>
- Research for Synthetic Data Generation Using Noising Strategies for Automatic Post-editing
- Publication of research paper in LREC 2022: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.93.pdf>
- Skills: Neural Machine Translation, Pytorch, Automatic Post Editing, Huggingface, Fairseq

## **Quality Estimation Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - Present

- Large Scale PLM based Quality Estimation development and research
- Low Resource QE and Pseudo QE data generation
- Publication of research paper in Applied Science (SCIE): <https://www.mdpi.com/2076-3417/11/14/6584>
- Publication of research paper in MT Summit 2021 - LoResMT Workshop: <https://aclanthology.org/2021.mtsummit-loresmt.1/>
- Publication of research paper in IEEE Access: <https://ieeexplore.ieee.org/document/9761258>
- Publication of research paper in COLING 2022: <https://aclanthology.org/2022.coling-1.460.pdf>
- 1st place in WMT 2022 QE Task 3: KU X Upstage's submission for the WMT22 Quality Estimation: Critical Error Detection Shared Task
- Skills: Neural Machine Translation, Pytorch, Quality Estimation, OpenKiwi, Huggingface, Fairseq

## **COVID-19 Machine Translation Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

May. 2020. - November.2020

- Neural Machine translation specialized for COVID-19 domain
- Skills: Neural Machine Translation, Pytorch, Python
- Publication of research paper in Journal of the Korea Convergence Society(KCI): <https://bit.ly/2QgQKLK>

## **(Project) Domain Specialized Neural Machine Translation (Security)**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

May. 2020. - November.2020

- Research on Neural Machine Translation specialized for security domain
- Supported by "National Security Research Institute"
- Project Manager, Tech Lead
- Development of the latest translation model based on Transformer/BERT, AI-based Korean-English translation model suitable for the security field, and research trend of the latest AI-based translation technology research
- We propose Cross Communication Method (CCM), a new DS-NMT training methodology, inspired by cognitive scientific interpretation of cross language speech perception. This is the first study to reflect cognitive science theory in DS-NMT training methodology.
- Publication of research paper in IEEE Access: <https://ieeexplore.ieee.org/document/9751075>
- Skills: Neural Machine Translation, Pytorch, Python



## (Project) A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2020. - Present

- Neural-Symbolic Model Research and Development
- Skills: Neural Symbolic Computing, Knowledge base, Pytorch, Python
- Supported by IITP
- Demo: <http://nlpplab.iptime.org:32245/>

## Ancient Korean Neural Machine Translation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

March. 2020. - Present

- Ancient Korean Neural Machine Translation Research
- Bilingual Word Embedding for Ancient Korean Neural Machine Translation
- Priming Ancient Korean Neural Machine Translation Research
- Publication of research paper in IEEE Access(SCIE): <https://ieeexplore.ieee.org/document/9125904/>
- Publication of research paper in LREC 2022: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.3.pdf>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

## Deep Learning based Grammar Error Correction Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - June.2020

- Deep learning based English grammar corrector development
- Publication of research paper in IEEE Access(SCIE): <https://ieeexplore.ieee.org/document/9102992>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

## Deep Learning based Korean Spelling Correction Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - Present

- Development of Korean spelling corrector based on deep learning
- We propose a gold test set, K-NCT (Korean Neural grammar Correction Test set), for Korean grammar error correction and a new error type classification guideline.
- Publication of research paper in multimedia tools and applications (SCIE): <https://link.springer.com/article/10.1007/s11042-020-09148-2>
- Publication of research paper in IEEE Access (SCIE): <https://ieeexplore.ieee.org/document/9938990>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

## Neural Machine Translation Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - Present

- Deep learning based Korean-English and English-Korean, Korean-Japanese machine translator development
- Research for Parallel Corpus Filtering, Data Augmentation, Various decoding strategies
- Best practices for using NIA AI training data: <http://www.aihub.or.kr/node/4525>
- Publication of research paper in ELECTRONICS (SCIE): <https://www.mdpi.com/2079-9292/9/10/1562#cite>
- Our paper is accepted in NAACL 2021 Industrial Track: "Should we find another model?: Improving Neural Machine Translation Performance with ONE-Piece Tokenization Method without Model Modification"
- How to improve the performance of NMT without changing the model through data pre- and post-processing
- Publication of research paper in NAACL 2021: <https://www.aclweb.org/anthology/2021.naacl-industry.13/>
- Verification of Ensemble Effect in Neural Machine Translation
- Publication of research paper in EMNLP 2021 - Workshop on Insights from Negative Results in NLP: <https://insights-workshop.github.io/2021/cfp/>
- Publication of research paper in Applied Sciences (SCIE): <https://www.mdpi.com/2076-3417/12/11/5545>
- Skills: Pytorch, Python, REST API,Flask,Neural Machine Translation

## AI for Education Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2019. - Present

- FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue
- Publication of research paper in AAAI 2022 - Artificial Intelligence for Education (AI4EDU) workshop: <https://arxiv.org/abs/2112.04126>
- Publication of research paper in LREC 2022: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.132.pdf>
- Automatic Essay Scoring Research
- PEEP-Talk: Deep Learning-based English Education Platform for Personalized Foreign Language Learning
- Publication of research paper in ACL 2023 - Demo Track: <https://www.mdpi.com/2227-7390/10/9/1486>
- AI student: A Machine Reading Comprehension System for the Korean College Scholastic Ability Test
- Publication of research paper in Mathematics (SCIE): <https://www.mdpi.com/2227-7390/10/9/1486>
- BERTOEIC: Solving TOEIC Problems Using Simple and Efficient Data Augmentation Techniques with Pretrained Transformer Encoders
- Publication of research paper in Applied Sciences (SCIE) : <https://www.mdpi.com/2076-3417/12/13/6686>
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

## AI for Accessibility Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2019. - Present

- PicTalky: Design of augmentative alternative communication (AAC) software for children with developmental disabilities
- 1st Place Microsoft AI Accessibility Hackathon in Korea (Team Lead)
- Award: <https://bit.ly/2vPaZ91>
- PickTalky Demo: <https://pictalky-311.de.r.appspot.com/>
- Publication of research paper in ACL 2022 Demo Track: <https://preview.aclanthology.org/acl-22-ingestion/2022.aac1-demo.3/>
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

## (Project) Research and development of human-inspired multiple intelligence

Seoul, South Korea

RESEARCHER IN KOREA UNIV

September. 2019. - August. 2022

- Human-inspired language understanding model
- Funded by the Korea Ministry of Science and ICT (MSIT)
- Site: <http://hiai.co.kr/>

## (Project) Next-generation information computing technology development business

Seoul, South Korea

RESEARCHER IN KOREA UNIV

September. 2019. - January. 2021.

- Development of Core Application Techniques for Korean Intelligent Systems
- Research and development of Korean natural language processing core technology

## (Project) Core technology development of the real-time simultaneous speech translation based on knowledge enhancement

Seoul, South Korea

RESEARCHER IN SYSTRAN

June. 2018. - July. 2019

- Collaboration with the Korea Electronics and Telecommunications Research Institute (ETRI) Language Intelligence Research Group
- Director: Dr. Young-Gil Kim (ETRI)
- 3rd year: Real-time lecture interpretation solution development
- 4th year: Busan University of Foreign Studies Field Group Test (FGT) progress and software maintenance
- Skills: JAVA, JAVA FX, Kaldi

## (Project) Neural Machine Translation service for SK Telecom's in-house education program

Seoul, South Korea

RESEARCHER IN SYSTRAN

April. 2019. - April. 2019

- Development of automation software to provide translation service to SK Telecom's in-house education program
- Skills: Python

## LCloud-API Platform Service Development

Seoul, South Korea

RESEARCHER IN SYSTRAN

January. 2019. - June. 2019

- Planning and development of natural language processing API Platform Service
- Gravitee IO-based platform design, planning, marketing materials production, strategy establishment
- REST API management using Swagger
- Transcription API, Translation API, Address Correction API, Speaker Verification API
- Learned Microservices architecture(MSA), IT Infrastructure
- Skills: Gravitee IO, Microservices architecture(MSA), IT Infrastructure, Mockup, Marketing, Strategic Planning, Technology management, Amazon AWS

## Speech recognition transcription tool development

Seoul, South Korea

RESEARCHER IN SYSTRAN

January. 2019. - July. 2019

- Development and maintenance of speech recognition transcription tool, modification and addition of various functions
- Skills: Front End, Spring Boot, Flask, AJAX, MySQL, Kaldi,

## (Project) CJ Logistics Speech Recognition Agent

Seoul, South Korea

RESEARCHER IN SYSTRAN

Aug. 2018. - July. 2019

- CJ Logistics artificial intelligence (AI) virtual assistant project
- Skills: Quality Assurance, Quality control, Mockup, Strategic Planning, Kaldi

## Automatic subtitle generation platform service development and planning

Seoul, South Korea

RESEARCHER IN SYSTRAN

June. 2018. - June. 2019

- Development and maintenance of automatic subtitle generation platform
- Establish platform service strategies such as B2B and B2C
- Various company contacts and SWOT analysis
- Skills: JAVA, JAVA FX, Kaldi, Mockup, Strategic Planning, Technology management, Marketing

## Planning of English speaking evaluation platform service

Seoul, South Korea

RESEARCHER IN SYSTRAN

June. 2018. - December. 2018

- Establishing a service strategy using English speaking evaluation technology
- English speaking evaluation technology Q.A & Q.C
- Skills: Quality Assurance, Quality control, Mockup, Strategic Planning, Technology management, Marketing



## Speech Recognition Research

RESEARCHER IN SYSTRAN

- Research for Automatic Speech Recognition Technology based on Kaldi
- Acoustic Model, Language Model
- Speaker Verification
- English speaking evaluation
- Real-time speech recognition system
- Skills: Kaldi

Seoul, South Korea

Aug. 2018. - July. 2019

## Neural Machine Translation Research

RESEARCHER IN SYSTRAN

- Research for Neural Machine Translation
- Skills: OpenNMT Tensorflow, OpenNMT Pytorch

Seoul, South Korea

Aug. 2018. - July. 2019

## (Project) Development of multilingual word translation service based on Korean semantics analysis for multicultural family

RESEARCHER IN BUFS

- Han-Tong-E: Multilingual Word Translation Service based on Word Semantic Analysis
- Collaboration with Professor Ok Cheol-young's laboratory at Ulsan University
- Project homepage development
- In-depth understanding of Utagger and seminar on WSD
- Build project-related data and conduct surveys
- Project Page: <https://bufsconv.blogspot.kr/>
- Funded by National Research Foundation of Korea(NRF)

Seoul, South Korea

Sep. 2016. - June. 2018

## (Project) Establishment of Gold Standard language resource bank for Korean machine learning and evaluation (Next-generation information computing technology development business)

RESEARCHER IN BUFS

- For the purpose of developing Korean Natural Language Toolkit
- Development of automatic Korean word spacing and Korean compound noun decomposer
- Skills: Python

Seoul, South Korea

Sep. 2017. - June. 2018

## (Project) Local industry development project in connection with public institutions

RESEARCHER IN BUFS

- ICT VAEMS: Vertical Aquaculture Energy Management
- Funded by the Korea Institute for Advancement of Technology(KIAT)

Seoul, South Korea

Sep. 2016. - Aug. 2017

## (Project) Basic Research on Semantic Frame Generation for Short Questions (ExoBrain)

RESEARCHER IN BUFS

- Exobrain Project

Seoul, South Korea

Oct. 2015. - Feb. 2016

# Professional Experience

## Upstage

AI RESEARCH ENGINEER

- About Upstage: <https://www.upstage.ai/>
- Product Search Team -> Data Team -> Personalization AI Team & LLM TF
- Responsible roles related to search system evaluation
- Establishment of evaluation system to develop Product Search Pack and responsible for data, planning, and evaluation
- Semantic search project with AMOREPACIFIC
- Semantic search project with LG U+
- Semantic Search-related DataOps
- OCR-related DataOps
- Research for Data-centric AI
- Research for Large Language Model
- Responsible roles related to LLM evaluation

Seoul, South Korea

March. 2022 - Present

## SYSTRAN

RESEARCH ENGINEER

Seoul, South Korea

Sep. 2018 - July. 2019

- About SYSTRAN: <https://www.systransoft.com/>
- Automatic Speech Recognition(ASR) Team
- Core technology development of the real-time simultaneous speech translation based on knowledge enhancement (ETRI)
- Participated in NTT Docomo speech Recognition Solution Project
- Participated in CJ Logistics speech recognition agent development project
- Participated in Samsung SDS speech recognition project
- Speech recognition service planning and development (Real-time lecture interpretation, Automatic subtitle generation service, English speaking evaluation, Customer-specific recognizer service, Web RTC-based Real-time video interpretation)
- Deep learning based speech recognition research (Acoustic Model, Language Model Training)
- Machine Translation research
- IaaS and PaaS service planning, API Platform service operation planning, B2B B2C planning and development
- Micro Service Architecture Design and Successful API Platform Service Launching Experience (LCloud)
- Governmental project proposal experience
- Reason for leaving: Graduate school

## SYSTRAN

RESEARCH INTERN

Seoul, South Korea

June. 2018 - August. 2018

- Automatic Speech Recognition(ASR) Team
- Participated a project "Core technology development of the real-time simultaneous speech translation based on knowledge enhancement (ETRI)"
- Developed Simultaneous Speech Translation (implemented using JAVA, Python and Kaldi)

## Skills

<b>Programming</b>	Python, JAVA, C / C++
<b>Machine Learning &amp; Deep Learning</b>	Pytorch, Pytorch lightning, Keras, Tensorflow
<b>Natural Language Processing</b>	Machine Translation, Speech Recognition, Information Retrieval, Huggingface, OpenNMT, Kaldi, FAISS
<b>Full Stack</b>	Html, CSS, Javascript, Java FX, Oracle, MySQL, Spring Boot, VertX, AWS
<b>Product &amp; Project &amp; Data Manager</b>	MS Office, Notion, Slack, Linear, Whimsical, Recruitee, Redmine
<b>Strategy &amp; Planning</b>	API Platform Service, Platform Service Planning, Gravitee IO, PaaS, IaaS, Q.A, Q.C, Mockup, Swagger
<b>Etc</b>	Linux, Docker, Web RTC, Latex
<b>Language Skills</b>	Korean, English, Japanese, Thai, Myanmar

## Patents

2019.12	<b>Device and method for correcting Korean spelling</b> , Granted Patent (10-2430918)	South Korea
2020.09	<b>APPARATUS AND METHOD FOR OUTPUTTING IMAGE CORRESPONDING TO LANGUAGE</b> , Granted Patent (10-2476497)	South Korea
2020.09	<b>METHOD OF TRANSLATING ANCIENT KOREAN USING MACHINE TRANSLATION</b> , Granted Patent (10-2425922)	South Korea
2022.04	<b>Correction performance evaluation metrics of neural network machine translation and method of constructing the same</b> , Granted Patent (10-2390154)	South Korea
2020.09	<b>METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION</b> , Granted Patent (10-2409667)	South Korea
2021.03	<b>Improving speech recognition performance using TTS in domain-specific environment</b> , Apply for a patent	South Korea
2021.05	<b>Method For Generating Training Data And Method For Post-Processing Of Speech Recognition Using The Same</b> , Apply for a patent	South Korea
2021.09	<b>DEVICE AND METHOD FOR GENERATING OPTIMAL TRANSLATION SUBTITLE USING QUALITY ESTIMATION</b> , Apply for a patent	South Korea
2021.09	<b>DEVICE AND METHOD FOR GENERATING TRAINING DATA FOR AUTOMATIC POST EDITING</b> , Apply for a patent	South Korea
2021.09	<b>METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION</b> , PCT International Application	South Korea
2021.09	<b>APPARATUS FOR CORPUS PROCESSING, APPARATUS AND METHOD AND MATHINE TRANSLATION</b> , Apply for a patent	South Korea
2021.11	<b>DEVICE AND METHOD FOR GENERATING OF TRAINING DATA FOR QUALITY ESTIMATION IN MACHINE TRANSLATION</b> , Apply for a patent	South Korea

## Demo

2019.12	<b>Deep Learning based English Grammar Error Correction</b> , <a href="http://nlplab.iptime.org:32292/">http://nlplab.iptime.org:32292/</a>	South Korea
2019.12	<b>Deep Learning based Korean Spelling Correction</b> , <a href="http://nlplab.iptime.org:32288/">http://nlplab.iptime.org:32288/</a>	South Korea
2020.01	<b>Korea University Neural Machine Translation Platform</b> , <a href="http://nlplab.iptime.org:32296/">http://nlplab.iptime.org:32296/</a>	South Korea
2020.02	<b>Neural Symbolic Inference</b> , <a href="http://nlplab.iptime.org:32245/">http://nlplab.iptime.org:32245/</a>	South Korea
2020.02	<b>English Free Talking and Grammar Error Correction Feed Back System using NAO Robot</b> , <a href="https://www.youtube.com/watch?v=Rl6dG_4cnFk">https://www.youtube.com/watch?v=Rl6dG_4cnFk</a>	South Korea
2020.03	<b>Korean Ancient Neural Machine Translation Platform</b> , <a href="http://nlplab.iptime.org:32242/">http://nlplab.iptime.org:32242/</a>	South Korea
2020.03	<b>PicTalky: Text to Pictogram Demo for Children with Developmental Disabilities</b> , <a href="https://pictalky-311.de.r.appspot.com/">https://pictalky-311.de.r.appspot.com/</a>	South Korea
2020.05	<b>COVID-19 Machine Translation Platform</b> , <a href="http://nlplab.iptime.org:32250/">http://nlplab.iptime.org:32250/</a>	South Korea
2020.07	<b>Ancient Korean Translation &amp; Machine Reading Comprehension with NAO</b> , <a href="https://www.youtube.com/watch?v=nXBZe-t9a-4&amp;t=8s">https://www.youtube.com/watch?v=nXBZe-t9a-4&amp;t=8s</a>	South Korea
2020.11	<b>Bilingual Education Using NAO</b> , <a href="https://www.youtube.com/watch?v=4kCMMa62LDY&amp;t=9s">https://www.youtube.com/watch?v=4kCMMa62LDY&amp;t=9s</a>	South Korea
2021.04	<b>Ko-CSS (Korean Colloquial-Sentence-Splitter) for simultaneous speech translation</b> , <a href="http://nlplab.iptime.org:32285/">http://nlplab.iptime.org:32285/</a>	South Korea
2021.06	<b>BTS (BackTranScription) for STT Post-Processor</b> , <a href="https://backtrans-317413.de.r.appspot.com/">https://backtrans-317413.de.r.appspot.com/</a>	South Korea
2021.09	<b>Translation platform with AIHUB dataset</b> , <a href="http://nlplab.iptime.org:32247/">http://nlplab.iptime.org:32247/</a>	South Korea
2021.09	<b>Con-Talky</b> , <a href="http://nlplab.iptime.org:32245/">http://nlplab.iptime.org:32245/</a>	South Korea
2021.09	<b>Korea University English Grammar Corrector KakaoTalk Channel</b> , <a href="http://pf.kakao.com/_cqxkps">http://pf.kakao.com/_cqxkps</a>	South Korea
2021.09	<b>PEEP-Talk</b> , <a href="http://pf.kakao.com/_FlDxgs">http://pf.kakao.com/_FlDxgs</a>	South Korea
2021.10	<b>A Tool for generating High-Quality Parallel Corpus considering Human Translation Market</b> , <a href="http://nlplab.iptime.org:9090/">http://nlplab.iptime.org:9090/</a>	South Korea
2021.10	<b>A New Tool for Efficiently Generating Quality Estimation Datasets</b> , <a href="http://nlplab.iptime.org:9091/">http://nlplab.iptime.org:9091/</a>	South Korea
2021.11	<b>A Self-Supervised Automatic Post-Editing Data Generation Tool</b> , <a href="http://nlplab.iptime.org:9092/">http://nlplab.iptime.org:9092/</a>	South Korea

## Invited Talk & Teaching Experience

### AI Education Alliance Policy lab (AIEDAP)

PRESENTER

- Deep Learning Understanding and Practice
- Invited by AIEDAP

South Korea

June. 2023

### KIDA (Korea Institute for Defense Analysis)

PRESENTER

- Data-Centric AI in the Large Language Model Era
- Invited by KIDA (Korea Institute for Defense Analysis)

South Korea

May. 2023

### Upstage

PRESENTER

- NLP based Large Language Model for All

South Korea

April. 2023

### BoostCamp

MASTER (INSTRUCTOR)

- Course Name: Data-centric NLP
- HomePage: [https://boostcamp.connect.or.kr/program\\_ai.html](https://boostcamp.connect.or.kr/program_ai.html)

South Korea

February. 2023-Present

### AI · DATA SUMMIT 2023

PRESENTER

- Real-World Centric AI
- Invited by AI · DATA SUMMIT 2023

South Korea

February. 2023

### Sunmoon University

PRESENTER

- Real-World Centric AI
- Invited by Sunmoon University

South Korea

December. 2022

### Kyungsung University

PRESENTER

- Language and Information Studies and the Future of Artificial Intelligence
- Invited by Kyungsung University

South Korea

August. 2022

## Hankuk University of Foreign Studies

### PRESENTER

- Basic practice of natural language processing for everyone
- Invited by Hankuk University of Foreign Studies, Interpreting and Translation Research Institute (ITRI)

## Dongguk University

### PRESENTER

- Artificial intelligence and Machine Translation lectures (3 times)
- Invited by Dongguk University, Department of English Language and Literature

## Korea University

### TEACHING ASSISTANT

- (BDC101) Introduction to Natural Language Processing in Big Data

## Busan Social Welfare Development Institute

### INVITED ADVISOR

- Attending advisory meetings and Focus Group Interview
- ICT-Based Ecosystem Construction Plan to Secure Opportunities to Enjoy Together - A Study on Developmental Disabilities

## Korea University

### HEAD TEACHING ASSISTANT

- (BDC101) Introduction to Natural Language Processing in Big Data
- (DFE610) Natural Language Processing for Digital Finance Engineering

## Korea University

### TEACHING ASSISTANT

- (COSE461) Natural Language Processing
- (DFC615) Artificial Intelligence and Natural Language Processing
- Lecture materials: <https://github.com/Parkchanjun/KU-NLP-2020-1>

## LLsoLLu

### PRESENTER

- Latest natural language processing trends
- Language Representation, Neural Symbolic Computing
- Invited by LLsoLLu Global Strategy and Planning Team

## NC SOFT

### PRESENTER

- Technology Transfer Seminar
- Korean Spelling Corrector Using Automatic Noise Generation Methodology
- Invited by NC SOFT Language AI Team

## Dongguk University

### PRESENTER

- A.I - NLP - MT for Liberal Arts
- Introduction to Natural Language Processing, RBMT, SMT, NMT, OpenNMT
- Invited by Dongguk University, Department of English Language and Literature professor. Myung-Kwan Park
- Lecture materials: [https://github.com/Parkchanjun/Dongguk\\_AI\\_NLP\\_MachineTranslation](https://github.com/Parkchanjun/Dongguk_AI_NLP_MachineTranslation)

## SKC

### INSTRUCTOR

- Deep Learning and Natural Language Processing Course
- Introduction of Text Preprocessing and NLP for Linguistics
- Introduction of Word Embedding
- Introduction of Machine Translation
- Lecture materials(Text Preprocessing): [https://github.com/Parkchanjun/SKC\\_Text\\_Preprocessing](https://github.com/Parkchanjun/SKC_Text_Preprocessing)
- Lecture materials(Introduction of Machine Translation): [https://github.com/Parkchanjun/SKC\\_MachineTranslation](https://github.com/Parkchanjun/SKC_MachineTranslation)

## Naver Tech Talk

### PRESENTER

- Machine Translation for Everyone
- Invited by Sung Kim who are the Head of Naver Clova A.I
- Lecture videos: <https://tv.naver.com/v/9906991>

## SK T Academy

### INSTRUCTOR

- Machine Translation for Everyone
- Invited by SK Planet
- Introduction to Natural Language Processing, RBMT, SMT, NMT, OpenNMT
- Lecture videos: <https://tacademy.skplanet.com/live/player/onlineLectureDetail.action?seq=165>

South Korea

July. 2022

South Korea

January. 2022

South Korea

Autumn 2021

South Korea

July. 2021

South Korea

Autumn 2020

South Korea

Spring 2020

South Korea

March. 2020

South Korea

February. 2020

South Korea

January. 2020

South Korea

Oct. 2019 - Nov.2019

South Korea

Aug. 2019

South Korea

Aug. 2019

## Sooyoungro Church

INSTRUCTOR

- Sooyoungro Church Dream Bowl School Instructor (Free Service)
- Basic English Grammar and English Conversation Education

South Korea

Aug. 2017 - Dec.2017

## Korea Press Foundation

INSTRUCTOR

- Reading mentoring

South Korea

Aug. 2015 - Dec.2015

# All Publications

---

## PREPRINTS

### Self-Improving-Leaderboard (SIL): A Call for Real-World Centric Natural Language Processing Leaderboards

Chanjun Park, Hyeonseok Moon, Seolhwa Lee, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim

arXiv preprint arXiv:2303.10888 (2023). 2023

### DMPs: Data Management Operation and Recipes

Eujeong Choi, Chanjun Park, (Corresponding Author)

arXiv preprint arXiv:2301.01228 (2023). 2023

### Empirical study on BlenderBot 2.0 Errors Analysis in terms of Model, Data and User-Centric Approach

Jungseob Lee, Midan Shim, Suhhye Son, Yujin Kim, Chanjun Park, Heuiseok Lim, (Equal Contribution (First Co-Author))

arXiv preprint arXiv:2201.03239 (2022). 2022

### Language Chameleon: Transformation analysis between languages using Cross-lingual Post-training based on Pre-trained language models

Suhhye Son, Chanjun Park, Jungseob Lee, Midan Shim, Chanhee Lee, Yoonna Jang, Jaehyung Seo, Heuiseok Lim, (Equal Contribution (First Co-Author))

arXiv preprint arXiv:2209.06422 (2022). 2022

## TOP CONFERENCE

### FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue

Chanjun Park, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuiseok Lim

LREC 2022 (Proceedings of the Language Resources and Evaluation Conference), 2022, Marseille, France

### Priming Ancient Korean Neural Machine Translation

Chanjun Park, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim

LREC 2022 (Proceedings of the Language Resources and Evaluation Conference), 2022, Marseille, France

### QUAK: A Synthetic Quality Estimation Dataset for Korean-English Neural Machine Translation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Gyeongmin Kim, Jungseob Lee, Heui-Seok Lim

COLING 2022 - Proceedings of the 29th International Conference on Computational Linguistics, 2022

### KU X Upstage's Submission for the WMT22 Quality Estimation: Critical Error Detection Shared Task

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim

WMT 2022 - EMNLP 2022 - Proceedings of the Seventh Conference on Machine Translation, 2022, Abu Dhabi

### Focus on FoCus: Is FoCus focused on Context, Knowledge and Persona?

SeungYoon Lee, Jungseob Lee, Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Jaehyung Seo, Jeongbae Park, Heui-Seok Lim

COLING 2022 - Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge, 2022

### A Self-Supervised Automatic Post-Editing Data Generation Tool

Hyeonseok Moon, Chanjun Park, Sugyeong Eo, Jaehyung Seo, SeungJun Lee, Heuiseok Lim

ICML 2022 - DataPerf workshop, 2022 (2022). 2022

### Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing

Hyeonseok Moon, Chanjun Park, Seolhwa Lee, Jaehyung Seo, Jungseob Lee, Sugyeong Eo, Heuiseok Lim

LREC 2022 (Proceedings of the Language Resources and Evaluation Conference), 2022, Marseille, France

### A Dog Is Passing Over The Jet? A Text-Generation Dataset for Korean Commonsense Reasoning and Evaluation

Jaehyung Seo, Seounghoon Lee, Chanjun Park, Yoonna Jang, Hyeonseok Moon, Sugyeong Eo, Seonmin Koo, Heuiseok Lim

Findings of the Association for Computational Linguistics: NAACL 2022, 2022, Seattle, United States

### Should we find another model?: Improving Neural Machine Translation Performance with ONE-Piece Tokenization Method without Model Modification

Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim

NAACL 2021 - Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers, 2021

### FreeTalky: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue

Chanjun Park, Yoonna Jang, Seolhwa Lee, Sungjin Park, Heuiseok Lim

AAAI 2022 - Artificial Intelligence for Education(AI4EDU) (2021). 2021

### How should human translation coexist with NMT? Efficient tool for building high quality parallel corpus

Chanjun Park, Seolhwa Lee, Hyeonseok Moon, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim

NeurIPS 2021 - Data-centric AI (DCAI) workshop (2021). 2021

## Two Heads are Better than One? Verification of Ensemble Effect in Neural Machine Translation

Chanjun Park, Sungjin Park, Seolhwa Lee, Taesun Whang, Heuseok Lim

EMNLP 2021 - Proceedings of the Second Workshop on Insights from Negative Results in NLP, 2021, Online and Punta Cana, Dominican Republic

## BTS: Back TranScripton for Speech-to-Text Post-Processor using Text-to-Speech-to-Text

Chanjun Park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo, Heuseok Lim

ACL 2021 - Proceedings of the 8th Workshop on Asian Translation (WAT2021), 2021, Online

## Dealing with the Paradox of Quality Estimation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Heuseok Lim, (Equal Contribution (First Co-Author))

MT Summit - Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021), 2021, Virtual

## A New Tool for Efficiently Generating Quality Estimation Datasets

Sugyeong Eo, Chanjun Park, Jaehyung Seo, Hyeonseok Moon, Heuseok Lim

NeurIPS 2021 - Data-centric AI (DCAI) workshop, 2021

## Syntax-enhanced Dialogue Summarization using Syntax-aware information

Seolhwa Lee, Kisu Yang, Chanjun Park, João Sedoc, Heui-Seok Lim

NerIPS 2021 16th Women in Machine Learning Workshop, 2021

## Towards Syntax-Aware Dialogue Summarization using Multi-task Learning

Seolhwa Lee, Kisu Yang, Chanjun Park, João Sedoc, Heui-Seok Lim

EMNLP 2021 - Widening NLP (WiNLP2021) workshop, 2021

## Automatic Knowledge Augmentation for Generative Commonsense Reasoning

Jaehyung Seo, Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Heuseok Lim

NeurIPS 2021 - Data-centric AI (DCAI) workshop, 2021

## INTERNATIONAL JOURNAL

### A Survey on Evaluation Metrics for Machine Translation

Seungjun Lee, Jungseob Lee, Hyeonseok Moon, Chanjun Park, Jaehyung Seo, Sugyeong Eo, Seonmin Koo, Heuseok Lim

Mathematics 11.4 (2023) p. 1006. Multidisciplinary Digital Publishing Institute, 2023

### Mimicking Infants' Bilingual Language Acquisition for Domain Specialized Neural Machine Translation

Chanjun Park, Woo-Young Go, Sugyeong Eo, Hyeonseok Moon, Seolhwa Lee, Heuseok Lim

IEEE Access 10 (2022) pp. 38684–38693. IEEE, 2022

### PicTalky: Augmentative and Alternative Communication for Language Developmental Disabilities

Chanjun Park, Yoonna Jang, Seolhwa Lee, Jaehyung Seo, Kisu Yang, Heuseok Lim

AACL-IJCNLP 2022: System Demonstrations, 2022, Taipei, Taiwan

### The ASR Post-Processor Performance Challenges of BackTranScripton (BTS): Data-Centric and Model-Centric Approaches

Chanjun Park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Heuseok Lim

Mathematics 10.19 (2022). 2022

### Empirical Analysis of Parallel Corpora and In-Depth Analysis Using LIWC

Chanjun Park, Midan Shim, Sugyeong Eo, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Heuseok Lim

Applied Sciences 12.11 (2022) p. 5545. Multidisciplinary Digital Publishing Institute, 2022

### Word-level Quality Estimation for Korean-English Neural Machine Translation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Heuseok Lim, (Equal Contribution (First Co-Author))

IEEE Access 10 (2022) pp. 44964–44973. IEEE, 2022

### Research on Utilization Strategy of User Engagements in Korean Fake News Detection

Myunghoon Kang, Jaehyung Seo, Chanjun Park, Heuseok Lim

IEEE Access (2022) pp. 1–1. 2022

### AI Student: A Machine Reading Comprehension System for the Korean College Scholastic Ability Test

Gyeongmin Kim, Soomin Lee, Chanjun Park, Jaechoon Jo

Mathematics 10.9 (2022) p. 1486. Multidisciplinary Digital Publishing Institute, 2022

### K-NCT: Korean Neural Grammatical Error Correction Gold-Standard Test Set Using Novel Error Type Classification Criteria

Seonmin Koo, Chanjun Park, Jaehyung Seo, Seungjun Lee, Hyeonseok Moon, Jungseob Lee, Heuseok Lim, (Equal Contribution (First Co-Author))

IEEE Access (2022). IEEE, 2022

### BERTOEIC: Solving TOEIC Problems Using Simple and Efficient Data Augmentation Techniques with Pretrained Transformer Encoders

Jeongwoo Lee, Hyeonseok Moon, Chanjun Park, Jaehyung Seo, Sugyeong Eo, Heuseok Lim

Applied Sciences 12.13 (2022) p. 6686. Multidisciplinary Digital Publishing Institute, 2022

### An Automatic Post Editing with Efficient and Simple Data Generation Method

Hyeonseok Moon, Chanjun Park, Jaehyung Seo, Sugyeong Eo, Heuseok Lim, (Equal Contribution (First Co-Author))

IEEE Access (2022). IEEE, 2022

### Return on Advertising Spend Prediction with Task Decomposition-Based LSTM Model

Hyeonseok Moon, Taemin Lee, Jaehyung Seo, Chanjun Park, Sugyeong Eo, Imatitkua D Aiyanyo, Jeongbae Park, Aram So, Kyoungwha Ok, Kinam Park

Mathematics 10.10 (2022) p. 1637. MDPI, 2022

### Dense-to-Question and Sparse-to-Answer: Hybrid Retriever System for Industrial Frequently Asked Questions

Jaehyung Seo, Taemin Lee, Hyeonseok Moon, Chanjun Park, Sugyeong Eo, Imatitkua D Aiyanyo, Kinam Park, Aram So, Sungmin Ahn, Jeongbae Park



**Plain Template Insertion: Korean-Prompt-Based Engineering for Few-Shot Learners**

Jaehyung Seo, Hyeonseok Moon, Chanhee Lee, Sugyeong Eo, **Chanjun Park**, Jihoon Kim, Changwoo Chun, Heuiseok Lim  
*IEEE Access* 10 (2022) pp. 107587–107597. 2022

**PU-GEN: Enhancing generative commonsense reasoning for language models with human-centered knowledge**

Jaehyung Seo, Dongsuk Oh, Sugyeong Eo, **Chanjun Park**, Kisu Yang, Hyeonseok Moon, Kinam Park, Heuiseok Lim  
*Knowledge-Based Systems* (2022) p. 109861. Elsevier, 2022

**AI for Patents: A Novel yet Effective and Efficient Framework for Patent Analysis**

Junyoung Son, Hyeonseok Moon, Jeongwoo Lee, Seolhwa Lee, **Chanjun Park**, Wonkyung Jung, Heuiseok Lim  
*IEEE Access* 10 (2022) pp. 59205–59218. IEEE, 2022

**Comparative Analysis of Current Approaches to Quality Estimation for Neural Machine Translation**

Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim, **(Equal Contribution (First Co-Author))**  
*Applied Sciences* 11.14 (2021) p. 6584. Multidisciplinary Digital Publishing Institute, 2021

**Grounded Vocabulary for Image Retrieval Using a Modified Multi-Generator Generative Adversarial Network**

Kuekyeng Kim, **Chanjun Park**, Jaehyung Seo, Heuiseok Lim  
*IEEE Access* (2021). IEEE, 2021

**Exploring the Data Efficiency of Cross-Lingual Post-Training in Pretrained Language Models**

Chanhee Lee, Kisu Yang, Taesun Whang, **Chanjun Park**, Andrew Matteson, Heuiseok Lim  
*Applied Sciences* 11.5 (2021) p. 1974. Multidisciplinary Digital Publishing Institute, 2021

**Who speaks like a style of Vitamin: Towards Syntax-Aware Dialogue Summarization using Multi-task Learning**

Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heuiseok Lim  
*IEEE Access* (2021). IEEE, 2021

**An Empirical Study on Automatic Post Editing for Neural Machine Translation**

Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim, **(Equal Contribution (First Co-Author))**  
*IEEE Access* 9 (2021) pp. 123754–123763. IEEE, 2021

**Variational Reward Estimator Bottleneck: Towards Robust Reward Estimator for Multidomain Task-Oriented Dialogue**

Jeiyeon Park, Chanhee Lee, **Chanjun Park**, Kuekyeng Kim, Heuiseok Lim  
*Applied Sciences* 11.14 (2021) p. 6624. Multidisciplinary Digital Publishing Institute, 2021

**Neural Spelling Correction: Translating Incorrect sentences to Correct sentences for Multimedia**

**Chanjun Park**, Kuekyeng Kim, YeongWook Yang, Minh Kang, Heuiseok Lim  
*Multimedia Tools and Applications* (2020) pp. 1–18. Springer, 2020

**Ancient Korean Neural Machine Translation**

**Chanjun Park**, Chanhee Lee, YeongWook Yang, Heuiseok Lim  
*IEEE Access* 8 (2020) pp. 116617–116625. 2020

**Comparison of the evaluation metrics for Neural Grammatical Error Correction with Overcorrection**

**Chanjun Park**, YeongWook Yang, Chanhee Lee, Heuiseok Lim  
*IEEE Access* 8 (2020) pp. 106264–106272. 2020

**Decoding Strategies for Improving Low-Resource Machine Translation**

**Chanjun Park**, Yeongwook Yang, Kinam Park, Heuiseok Lim  
*Electronics* 9.10 (2020) p. 1562. Multidisciplinary Digital Publishing Institute, 2020

**INTERNATIONAL CONFERENCE**

**Prompt-based Learning for English-German Critical Translation Error Detection**

Sugyeong Eo, **Chanjun Park**, Heuiseok Lim  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022)*, 2022

**A Dataset for Korean Graph-to-Text Generation**

Dahyun Jung, Lee, Seungyoon, SeungJun Lee, Sugyeong Eo, **Chanjun Park**, Danielle Aiyanyo Imatitikua  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022)*, 2022

**Masked language modeling-based Korean Data Augmentation Techniques Using Label Correction**

Myunghoon Kang, Jungseob Lee, SeungJun Lee, Hyeonseok Moon, **Chanjun Park**, Aram So  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022)*, 2022

**Automatic Generation of Learning Data for Korean Speech Recognition Post-Processor**

Seomin Koo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022)*, 2022

**Fortune Telling via Deep Learning based Generation Model**

Hyeonseok Moon, Jungseob Lee, Jaehyung Seo, Sugyeong Eo, **Chanjun Park**, Jeongbae Park  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022)*, 2022

**Efficient Way for Constructing Hate Speech-Counter Narrative Dataset**

Lee, Seungyoon, SuhYune Son, Dahyun Jung, **Chanjun Park**, Yuna Hur  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2022)*, 2022

**A Study on Jeju Dialect ASR processor using BackTranScriptio**

**Chanjun Park**, Jaehyung Seo, Seolhwa Lee, Heonseok Moon, Sugyeong Eo, Yoonna Jang, Heuiseok Lim  
*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021)*, 2021

**Cross-lingual Zero-shot Transfer for Quality Estimation**

Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Heonseok Moon, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Proposal of Automatic Essay Scoring Method Using 4-level Natural Language Analysis**

Yujin Kim, **Chanjun Park**, Seolhwa Lee, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Error Type Categorization Criteria for Korean Spelling Correction**

Seonmin Koo, **Chanjun Park**, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**A High-level Machine Reading Comprehension Model**

Jeongwoo Lee, Hyeonseok Moon, **Chanjun Park**, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Comparative study of Korean speech recognition based on SpecAugment and Kaldi**

Seounghoon Lee, **Chanjun Park**, Jaehyung Seo, Gyeongmin Kim, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Context Detector: Persona Based Natural Conversation using Conversational Agent**

Seungjun Lee, Yoonna Jang, **Chanjun Park**, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Domain Specialized Automatic Post Editing System**

Heonseok Moon, **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Korean Generative Commonsense Reasoning**

Jaehyung Seo, **Chanjun Park**, Heonseok Moon, Sugyeong Eo, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**AI for Construction: Development of communication platform for construction field using Natural Language Processing**

Midan Shim, **Chanjun Park**, Yuna Hur, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**Categorization of Common Error Types in the Off-the-shelf ASR systems for Korean**

Junyoung Son, **Chanjun Park**, Jaehyung Seo, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**A Comparative Study on Korean Relation Extraction with entity position information**

Suhylene Son, Yuna Hur, Jungwoo Lim, Midan Shim, **Chanjun Park**, HeuiSeok Lim

*International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE 2021), 2021*

**PicTalky: Deep learning-based speech therapy platform for developmental disabilities**

**Chanjun Park**, Yanghee Kim, Yoonna Jang, Umadevi G.R, HeuiSeok Lim

*The International Conference on Big data, IoT, and Cloud Computing, 2020*

**Domain-specialize Neural Machine Translation Methodology**

**Chanjun Park**, HeuiSeok Lim

*International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2020*

**Toward High Quality Parallel Corpus Using Monolingual Corpus**

**Chanjun Park**, Youngdae Oh, Jongkeun Choi, Dongpil Kim, HeuiSeok Lim

*The 10th International Conference on Convergence Technology, 2020*

**Data Augmentation Method for Korean Neural Machine Translation**

**Chanjun Park**, Kinam Park, HeuiSeok Lim

*INTERNATIONAL CONFERENCE ON INFORMATION, SYSTEM AND CONVERGENCE APPLICATIONS, 2020*

**Self- Supervised Korean Spelling Correction via Denoising Transformer**

**Chanjun Park**, Sungjin Park, HeuiSeok Lim

*INTERNATIONAL CONFERENCE ON INFORMATION, SYSTEM AND CONVERGENCE APPLICATIONS, 2020*

**Can artificial intelligence translate the Annals of the Joseon Dynasty ?**

**Chanjun Park**, Sungjin Park, Kinam Park, Jaechoon Jo, HeuiSeok Lim

*The 10th International Conference on Convergence Technology, 2020*

**Design Neural Machine Translation Model Combining External Symbolic Knowledge in translation task**

Sugyeong Eo, **Chanjun Park**, HeuiSeok Lim

*International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2020*

**Free Talking NAO Robot for Children**

**Chanjun Park**, Jaekue Choi, Sungjin Park, Yanghee Kim, HeuiSeok Lim

*International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019*

**Improved Machine Translation Performance Using Two Stage Subword Tokenization**

**Chanjun Park**, Sora Choi, HeuiSeok Lim

*International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019*

**Quality before Quantity: Improved Machine Translation performance using parallel corpus filtering**

**Chanjun Park**, Seunggyu Han, HeuiSeok Lim

*International conference on Interdisciplinary research on Computer Science, Psychology, and Education, 2019*

**Improved Machine Translation Performance Using Relative Ratio of Original and Synthetic Corpus**

## DOMESTIC JOURNAL

### **Policy-based performance comparison study of Real-time Simultaneous Translation**

Jungseob Lee, Hyeonseok Moon, Jaehyung **Chanjun Park** Seo, Sugyeong Eo, Seungjun Lee, Seonmin Koo, Heuiseok Lim

*Journal of the Korea Convergence Society* 13.3 (2022) pp. 43–54. Korea Convergence Society, 2022

### **Error Analysis of Recent Conversational Agent-based Commercialization Education Platform**

Seungjun Lee, **Chanjun Park**, Jaehyung Seo, Heuiseok Lim

*Journal of the Korea Convergence Society* 13.3 (2022) pp. 11–22. Korea Convergence Society, 2022

### **Deep Learning based Korean Dialect Machine Translation Research Considering Linguistics Features and Service**

Sangbeom Lim, **Chanjun Park**, Yeongwook Yang

*Journal of the Korea Convergence Society* 13.2 (2022) pp. 21–29. Korea Convergence Society, 2022

### **Korean and Multilingual Language Models Study for Cross-Lingual Post-Training (XPT)**

SuhYune Son, **Chanjun Park**, Jungseob Lee, Midan Shim, Chanhee Lee, Heuiseok Lim

*Journal of the Korea Convergence Society* 13.3 (2022) pp. 77–89. Korea Convergence Society, 2022

### **A study on performance improvement considering the balance between corpus in Neural Machine Translation**

**Chanjun Park**, Kinam Park, Sugyeong Eo, Hyeonseok Moon, HeuiSeok Lim

*Journal of the Korea Convergence Society* 12.5 (2021) pp. 23–29. Korea Convergence Society, 2021

### **A Study on Verification of Back TranScript(BTS)-based Data Construction**

**Chanjun Park**, Jaehyung Seo, Seolhwa Lee, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim

*Journal of the Korea Convergence Society* 12.11 (2021) pp. 109–117. Korea Convergence Society, 2021

### **Research on Subword Tokenization of Korean Neural Machine Translation and Proposal for Tokenization Method to Separate Jong-sung from Syllables**

Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, HeuiSeok Lim

*Journal of the Korea Convergence Society* 12.3 (2021) pp. 1–7. Korea Convergence Society, 2021

### **Research on Recent Quality Estimation**

Sugyeong Eo, **Chanjun Park**, Hyeonseok Moon, Jaehyung Seo, HeuiSeok Lim

*Journal of the Korea Convergence Society* 12.7 (2021) pp. 37–44. Korea Convergence Society, 2021

### **Study on Zero-shot based Quality Estimation**

Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Hyeonseok Moon, Heuiseok Lim

*Journal of the Korea Convergence Society* 12.11 (2021) pp. 35–43. Korea Convergence Society, 2021

### **Classification and analysis of error types for deep learning-based Korean spelling correction**

Seonmin Koo, **Chanjun Park**, Aram So, Heuiseok Lim

*Journal of the Korea Convergence Society* 12.12 (2021) pp. 65–74. Korea Convergence Society, 2021

### **Empirical study on BlenderBot 2.0's errors analysis in terms of model, data and dialogue**

Jungseob Lee, SuhYune Son, Midan Shim, Yujin Kim, **Chanjun Park**, Aram So, Jeongbae Park, Heuiseok Lim

*Journal of the Korea Convergence Society* 12.12 (2021) pp. 93–106. Korea Convergence Society, 2021

### **Filter-mBART Based Neural Machine Translation Using Parallel Corpus Filtering**

Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, JeongBae Park, HeuiSeok Lim

*Journal of the Korea Convergence Society* 12.5 (2021) pp. 1–7. Korea Convergence Society, 2021

### **Recent Automatic Post Editing Research**

Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, HeuiSeok Lim

*Journal of Digital Convergence* 19.7 (2021) pp. 199–208. Journal of Digital Convergence, 2021

### **The Verification of the Transfer Learning-based Automatic Post Editing Model**

Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim

*Journal of the Korea Convergence Society* 12.10 (2021) pp. 27–35. Korea Convergence Society, 2021

### **Study on Decoding Strategies in Neural Machine Translation**

Jaehyung Seo, **Chanjun Park**, Sugyeong Eo, Hyeonseok Moon, Heuiseok Lim

*Journal of the Korea Convergence Society* 12.11 (2021) pp. 69–80. Korea Convergence Society, 2021

### **Cross-Validated Ensemble Methods in Natural Language Inference**

Kisu Yang, Taesun Whang, Dongsuk Oh, **Chanjun Park**, HeuiSeok Lim

*Journal of KIISE* 48.2 (2021) pp. 154–159. Korean Institute of Information Scientists and Engineers (KIISE), 2021

### **Neural Machine translation specialized for Coronavirus Disease-19 (COVID-19)**

**Chanjun Park**, Kyeong-Hee Kim, Ki-Nam Park, Heui-Seok Lim

*Journal of the Korea Convergence Society* 11.9 (2020) pp. 7–13. Korea Convergence Society, 2020

### **An AI Service to support communication and language learning for people with developmental disability**

**Chanjun Park**, Yanghee Kim, Yoonna Jang, Umadevi G.R, HeuiSeok Lim

*Journal of the Korea Convergence Society* 11.6 (2020) pp. 51–57. 2020

### **Automatic Post Editing Research**

**Chanjun Park**, HeuiSeok Lim

*Journal of the Korea Convergence Society* 11.5 (2020) pp. 1–8. 2020

### **A Study on the Performance Improvement of Machine Translation Using Public Korean-English Parallel Corpus**

Chanjun Park, HeuiSeok Lim

## DOMESTIC CONFERENCE

### **Are Vector Size and Response Time Proportional?: Empirical Study of Response Time according to Index Type in Semantic Search**

**Chanjun Park**, Eujeong Choi, Dayone Suh Suh, Minjeong Kim, Suwon Shin, Juhwan Yoon, Seungwon Jeong, Jaekyoung Bae

*Korea Software Congress 2022 (2022) pp. 350–352. 2022*

### **Is Modeling Essential for Model Improvement? Encoder performance improvement through a Data-Centric Approach**

Seungwon Jeong, Suwan Kim, **Chanjun Park**

*Korea Software Congress 2022 (2022) pp. 377–379. 2022*

### **BackTranScription (BTS)-based Jeju Automatic Speech Recognition Post-processor Research**

**Chanjun Park**, Jaehyung Seo, Seolhwa Lee, Heonseok Moon, Sugyeong Eo, Yoonna Jang, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Word-level Korean-English Quality Estimation**

Sugyeong Eo, **Chanjun Park**, Jaehyung Seo, Heonseok Moon, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Persona-based Korean Conversational Model**

Yoonna Jang, Jungwoo Lim, Yuna Hur, Kisu Yang, **Chanjun Park**, Jaehyung Seo, Seungjun Lee, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Proposal of Automated Essay Scoring Method based on Deep-Learning**

Yujin Kim, **Chanjun Park**, Seolhwa Lee, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Error Type Categorization for deep learning-based Korean Spelling Correction**

Seonmin Koo, **Chanjun Park**, HeuiSeok Lim

*Korea Software Congress 2018 (2021) pp. 1279–1281. 2021*

### **Model Training and Data Augmentation Schemes For the High-level Machine Reading Comprehension**

Jeongwoo Lee, Hyeonseok Moon, **Chanjun Park**, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Comparative study of Korean speech recognition based on SpecAugment and Kaldi**

Seounghoon Lee, **Chanjun Park**, Jaehyung Seo, Gyeongmin Kim, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **PEEP-Talk: Deep Learning-based English Education Platform for Personalized Foreign Language Learning (Best Paper Award)**

Seungjun Lee, Yoonna Jang, **Chanjun Park**, Minwoo Kim, Bernardo N Yahya, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Deep Learning based Korean Dialect Machine Translation Research**

Sangbeom Lim, **Chanjun Park**, Jaechoon Jo, Yeongwook Yang

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Verification of the Domain Specialized Automatic Post Editing Model**

Heonseok Moon, **Chanjun Park**, Jaehyung Seo, Sugyeong Eo, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation (Best Paper Award)**

Jaehyung Seo, **Chanjun Park**, Heonseok Moon, Sugyeong Eo, Myunghoon Kang, Seounghoon Lee, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Con-Talky: Information Extraction and Visualization Platform for Communication of Construction Industry**

Midan Shim, **Chanjun Park**, Yuna Hur, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Categorization and Analysis of Error Types in the Korean Speech Recognition System (Best Paper Award)**

Junyoung Son, **Chanjun Park**, Jaehyung Seo, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **A Comparative Study on Korean Relation Extraction with entity position information**

Suhyune Son, Yuna Hur, Jungwoo Lim, Midan Shim, **Chanjun Park**, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Attention Patterns and Semantics of Korean Language Models**

Kisu Yang, Yoonna Jang, Jingwoo Lim, **Chanjun Park**, Hwanseok Jang, HeuiSeok Lim

*The 33rd Annual Conference on Human Cognitive Language Technology, 2021*

### **Quality, not Quantity? : Effect of parallel corpus quantity and quality on Neural Machine Translation**

**Chanjun Park**, Yeonsu Lee, Chanhee Lee, HeuiSeok Lim

*The 32st Annual Conference on Human Cognitive Language Technology, 2020*

### **Alleviation of Overcorrection Problem in Neural Korean Spelling Correction**

**Chanjun Park**, Yeonsu Lee, Kisu Yang, HeuiSeok Lim

*The 32st Annual Conference on Human Cognitive Language Technology, 2020*

### **Design Neural Machine Translation Model Combining External Symbolic Knowledge**

Sugyeong Eo, **Chanjun Park**, HeuiSeok Lim

*The 32st Annual Conference on Human Cognitive Language Technology, 2020*

### **Cross-Lingual Transfer of Pretrained Transformers to Resource-Scarce Languages (Best Paper Award)**

Chanhee Lee, **Chanjun Park**, Gyeongmin Kim, Dongsuk Oh, HeuiSeok Lim

*The 32st Annual Conference on Human Cognitive Language Technology, 2020*

### **MMA: Multi-modal Message Aggregation for Korean VQA**

Sungjin Park, **Chanjun Park**, Jaehyung Seo, HeuiSeok Lim

*The 32st Annual Conference on Human Cognitive Language Technology, 2020*

### **Korean Spell Correction based on Denoising Transformer**

**Chanjun Park**, Sol Jeong, Kisu Yang, Sumi Lee, Jaechoon Joe, HeuiSeok Lim

*The 31st Annual Conference on Human Cognitive Language Technology, 2019*

### **Parallel Corpus Filtering and Korean-Optimized Subword Tokenization for Machine Translation**

**Chanjun Park**, Gyeongmin Kim, HeuiSeok Lim

*The 31st Annual Conference on Human Cognitive Language Technology, 2019*

### **Optimization of Data Augmentation Techniques in Neural Machine Translation**

**Chanjun Park**, Kuekyeng Kim, HeuiSeok Lim

*The 31st Annual Conference on Human Cognitive Language Technology, 2019*

### **Comparative Study of Sentiment Analysis Model based on Korean Linguistic Characteristics**

Gyeongmin Kim, **Chanjun Park**, HeuiSeok Lim

*The 31st Annual Conference on Human Cognitive Language Technology, 2019*

### **Cross-Validated Ensemble Methods in Natural Language Inference (Best Paper Award)**

Kisu Yang, Taesun Whang, Dongsuk Oh, **Chanjun Park**, HeuiSeok Lim

*The 31st Annual Conference on Human Cognitive Language Technology, 2019*

### **Korean Spacing Module using Eojeol dictionary and Morpheme dictionary**

**Chanjun Park**, Jin-hyuk Choi, Pummo Ryu

*Korea Software Congress 2018, 2018*

### **Knowledge-enhancing Real-time Lecture interpreter service Development**

**Chanjun Park**, Minhoo Kang, Byunggun Park, Suntae Kim, Myeongjin Hwang, Dongpil Kim, Young-kil Kim

*Korea Software Congress 2018, 2018*

### **Voice recognition performance improvement using TTS system in address recognition system**

**Chanjun Park**, Suntae Kim, Minhoo Kang, Myeongjin Hwang, Byunggun Park, Yeonsu Kim

*Korea Software Congress 2018, 2018*

### **Improved speech recognition performance in business meeting environments using domain-specific method**

**Chanjun Park**, Suntae Kim, Minhoo Kang, Byunggun Park, Myeongjin Hwang, Yeonsu Kim

*Korea Software Congress 2018, 2018*

### **Spam/Ham Classification using Ensemble Technique of Machine Learning Algorithms**

**Chanjun Park**, Pummo Ryu

*Korea Computer Congress 2018, 2018*

### **Two-Stage Korea Compound Noun Decomposer**

**Chanjun Park**, Pummo Ryu

*The 30st Annual Conference on Human Cognitive Language Technology, 2018*

### **Dictionary based Compound Noun Decomposer using Resource extension**

**Chanjun Park**, Hyejin Won, Pummo Ryu

*Korea Computer Congress 2018, 2018*

## **BOOK CHAPTERS**

### **Data-Centric Neural Machine Translation - A Real-World Approaches**

**Chanjun Park**

*Ph.D. Thesis (2023). 2023*

### **Natural Language Processing Bible**

Heuiseok Lim, **Korea University NLP AI Lab**

Human Science, <https://www.aladin.co.kr/shop/wproduct.asp?partner=rss&ISBN=K412637214>, 2020

## Operate Naver's natural language processing blog

BLOGGER

- URL: <https://blog.naver.com/bcj1210>
- Linguistics for natural language processing
- Natural Language Processing (NLP)
- Machine Translation for Everyone
- Business, strategy, and planning for everyone
- Natural language processing with NLTK
- Software engineering
- Today's CCM

South Korea

2016.Dec-Present

## OpenNMT Forum Members

MEMBERS

- Active as OpenNMT Forum Members
- Develop and share OpenNMT Pytorch, Tensorflow, Library Tutorial

South Korea

2019.July-2019.July

## ETRI's machine translation quality evaluator

ESTIMATOR

- Quality evaluation of machine translation hosted by ETRI

South Korea

2019.October-2019.December

## Completion of basic curriculum for natural language processing and language intelligence

STUDENT

- Korea University Human-Inspired AI & Computing Research Center

South Korea

2019.July-2019.July

## Completed the natural language processing tutorial (2016-2019)

STUDENT

- Special Interest Group of Human and Cognitive Language Technology

South Korea

2016 - 2019

## Reporter, Korea Electronics and Telecommunications Research Institute (ETRI)

REPORTER

- Served as 3rd and 4th reporters at the Korea Electronics and Telecommunications Research Institute

South Korea

2016.Feb-2018.Feb

## Flitto Supporters

SUPPORTERS

- The first plito supporters

South Korea

2017.Dec-2017.Dec

## Journalist of the Ministry of Science and ICT

REPORTER

- 8 articles published
- Visit Exobrain Quiz Show
- URL: <https://blog.naver.com/bcj1210/221147196551>

South Korea

2016.Feb-2017.Feb

## Planning a video lecture for Natural Language Processing Basic at BUFS

PLANNER

- URL: <https://bit.ly/2NmFJV1>

South Korea

2016.July-2016.July

## BIT Computer Advanced Course Completion

STUDENT

- Bet Advanced Course Completion and Bet President Award
- Bit project level 2 license
- Development of intelligent English learning APP

South Korea

2016.January-2016.August

## BIT Computer Short Course Completion

STUDENT

- Short-term course completion (C, JAVA, Data Structure)

South Korea

2015.June-2015.August