

# Chanjun Park

NATURAL LANGUAGE PROCESSING & ARTIFICIAL INTELLIGENCE LAB · KOREA UNIVERSITY

311 Aegineung Student Center, College of Informatics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea

☎ (+82) 10-2580-9066 | ✉ bcj1210@naver.com | 📷 parkchanjun | 📺 bcj1210

## Research Interests

I am a MS/Ph.D student in the Department of Computer Science and Engineering at Korea University. My advisor is prof. Heuiseok Lim, and I am part of the Natural Language Processing & Artificial Intelligence Lab. I am the founder and leader of KU-NMT Group.

- Data-Centric AI
- Neural Machine Translation
- Domain Specialized Neural Machine Translation
- Ancient Korean Neural Machine Translation
- Simultaneous Speech Translation
- Parallel Corpus Filtering and Mining
- Automatic Post Editing
- Quality Estimation
- Korean Spelling Correction
- Grammar Error Correction
- Speech Recognition Post-Processor
- Dialogue Summarization
- AI for Accessibility
- AI for Education
- Language Resource and Evaluation

## Education

### Korea University

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

- Advisor: Heuiseok Lim
- Natural Language Processing & Artificial Intelligence Lab

Seoul, South Korea

Sep. 2019 - Current

### Busan University of Foreign Studies(BUFS)

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

- Advisor: Pummo Ryu
- Full Scholarship

Busan, South Korea

Mar. 2015 - Exp. Aug. 2018

## Academic Services

### COLING 2022

VIRTUAL SOCIAL CHAIR

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

South Korea

Sep.2021-Present

### Reviewer

REVIEWER

- ACL 2022
- NAACL 2022
- MDPI (Applied Sciences, Electronics, Entropy)

South Korea

January.2021-Present

## Selected Publications

---

### **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

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

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

*Applied Sciences* 11.14 (2021) p. 6584. Multidisciplinary Digital Publishing Institute, 2021

### **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

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

Kuekyeng Kim, **Chanjun Park**, Jaehyung Seo, Heuseok 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, Heuseok Lim

*Applied Sciences* 11.5 (2021) p. 1974. Multidisciplinary Digital Publishing Institute, 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

### **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

*IEEE Access* (2021). IEEE, 2021

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

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

*Applied Sciences* 11.14 (2021) p. 6624. Multidisciplinary Digital Publishing Institute, 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

### **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, Heuseok 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, Heuseok Lim

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

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

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

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

## Honors & Awards

|           |   |             |
|-----------|---|-------------|
| 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 |
| 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

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

RESEARCHER IN KOREA UNIV

Seoul, South Korea

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

RESEARCHER IN KOREA UNIV

Seoul, South Korea

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>
- Skills: Neural Machine Translation, Web Application, Fairseq

### Korean Commonsense Reasoning Research

RESEARCHER IN KOREA UNIV

Seoul, South Korea

June. 2021. - Present

- KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation
- Automatic Knowledge Augmentation for Generative Commonsense Reasoning
- Skills: Commonsense Reasoning, Pytorch, Huggingface

### Image Retrieval Research

RESEARCHER IN KOREA UNIV

Seoul, South Korea

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

RESEARCHER IN KOREA UNIV

Seoul, South Korea

June. 2021. - December. 2021.

- AI for Patent research
- T5 based Key sentence extraction and key phrase generation
- Supported by "LG Innotek"
- 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>
- 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 BackTranScripton(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/>
- Skills: Neural Machine Translation, Pytorch, Huggingface, Fairseq

## Cross Lingual Pretrained Language Model Research

Seoul, South Korea

RESEARCHER IN KOREA UNIV

January. 2021. - March. 2021.

- 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/>
- Low Resource APE and Pseudo APE data generation
- 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/>
- 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.
- 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/>
- 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>
- 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/>
- 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
- Automatic Essay Scoring Research
- PEEP-Talk: Deep Learning-based English Education Platform for Personalized Foreign Language Learning
- Publication of research paper in AAAI 2022 - Artificial Intelligence for Education (AI4EDU) workshop: <https://arxiv.org/abs/2112.04126>
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

## AI for Accessibility Research

RESEARCHER IN KOREA UNIV

- 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/>
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

Seoul, South Korea

June. 2019. - Present

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

RESEARCHER IN KOREA UNIV

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

Seoul, South Korea

September. 2019. - August. 2022

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

RESEARCHER IN KOREA UNIV

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

Seoul, South Korea

September. 2019. - January. 2021.

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

RESEARCHER IN SYSTRAN

- 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

Seoul, South Korea

June. 2018. - July. 2019

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

RESEARCHER IN SYSTRAN

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

Seoul, South Korea

April. 2019. - April. 2019

## LCloud-API Platform Service Development

RESEARCHER IN SYSTRAN

- Planning and development of natural language processing API Platform Service
- Gravitree 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: Gravitree IO, Microservices architecture(MSA), IT Infrastructure, Mockup, Marketing, Strategic Planning, Technology management, Amazon AWS

Seoul, South Korea

January. 2019. - June. 2019

## Speech recognition transcription tool development

RESEARCHER IN SYSTRAN

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

Seoul, South Korea

January. 2019. - July. 2019

## (Project) CJ Logistics Speech Recognition Agent

RESEARCHER IN SYSTRAN

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

Seoul, South Korea

Aug. 2018. - July. 2019

## Automatic subtitle generation platform service development and planning

RESEARCHER IN SYSTRAN

- 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

Seoul, South Korea

June. 2018. - June. 2019

## Planning of English speaking evaluation platform service

RESEARCHER IN SYSTRAN

- 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

Seoul, South Korea

June. 2018. - December. 2018

## 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

### SYSTRAN, Automatic Speech Recognition(ASR) Team

Seoul, South Korea

RESEARCH ENGINEER

Sep. 2018 - July. 2019

- 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, Automatic Speech Recognition(ASR) Team

Seoul, South Korea

RESEARCH INTERN

June. 2018 - August. 2018

- 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>                          | JAVA, Python, C / C++   |
| <b>Machine Learning &amp; Deep Learning</b> | Pytorch, Keras, Tensorflow, Huggingface, OpenNMT, Kaldi   |
| <b>Natural Language Processing</b>          | Machine Translation, Speech Recognition, Neuro-Symbolic Computing                                   |
| <b>Full Stack</b>                           | Html, CSS, Javascript, Java FX, Oracle, MySQL, Spring Boot, VertX, AWS                              |
| <b>Strategy &amp; Planning</b>              | API Platform Service, Platform Service Planning, Gravitee IO, PaaS, IaaS, Q.A, Q.C, Mockup, Swagger |
| <b>Etc</b>                                  | MS Office, Redmine, 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> , Apply for a patent  | South Korea |
| 2020.09 | <b>APPARATUS AND METHOD FOR OUTPUTTING IMAGE CORRESPONDING TO LANGUAGE</b> , Apply for a patent   | South Korea |
| 2020.09 | <b>METHOD OF TRANSLATING ANCIENT KOREAN USING MACHINE TRANSLATION</b> , Apply for a patent  | South Korea |
| 2020.09 | <b>Correction performance evaluation metrics of neural network machine translation and method of constructing the same</b> , Apply for a patent | South Korea |
| 2020.09 | <b>METHOD OF BUILDING TRAINING DATA OF MACHINE TRANSLATION</b> , Apply for a patent   | 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 MACHINE 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> ,<br><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> ,<br><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> ,<br><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> ,<br><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 platfrom 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/_cqkxps">http://pf.kakao.com/_cqkxps</a>   | South Korea |
| 2021.09 | <b>PEEP-Talk</b> , <a href="http://pf.kakao.com/_F1Dxgs">http://pf.kakao.com/_F1Dxgs</a>  | South Korea |
| 2021.10 | <b>A Tool for generating High-Quality Parallel Corpus considering Human Translation Market</b> ,<br><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

### Dongguk University

PRESENTER

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

South Korea

January. 2022

### Korea University

TEACHING ASSISTANT

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

South Korea

Autumn 2021

### 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

South Korea

July. 2021

### Korea University

HEAD TEACHING ASSISTANT

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

South Korea

Autumn 2020

### 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>

South Korea

Spring 2020

### LLsoLLu

PRESENTER

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

South Korea

March. 2020

## NC SOFT

### PRESENTER

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

South Korea

February. 2020

## 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)

South Korea

January. 2020

## 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)

South Korea

Oct. 2019 - Nov. 2019

## 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>

South Korea

Aug. 2019

## 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

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

## PREPRINTS

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

Jungseob Lee, Midan Shim, Suhyune Son, Yujin Kim, **Chanjun Park**, Heuseok Lim, **(Equal Contribution(First Co-Author))**  
*arXiv preprint arXiv:2201.03239 (2022). 2022*

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

**Chanjun Park**, Yoonna Jang, Seolhwa Lee, Jaehyung Seo, Kisu Yang, Heuseok Lim  
*arXiv preprint arXiv:2109.12941 (2021). 2021*

### Empirical Analysis of Korean Public AI Hub Parallel Corpora and in-depth Analysis using LIWC

**Chanjun Park**, Midan Shim, Sugyeong Eo, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Heuseok Lim  
*arXiv preprint arXiv:2110.15023 (2021). 2021*

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

Hyeonseok Moon, **Chanjun Park**, Sugyeong Eo, Jaehyung Seo, SeungJun Lee, Heuseok Lim  
*arXiv preprint arXiv:2111.12284 (2021). 2021*

## TOP CONFERENCE

### 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

*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

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

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

## **INTERNATIONAL CONFERENCE**

### **A Study on Jeju Dialect ASR processor using BackTranScripttion**

**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

**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**

**Chanjun Park**, Jang Mook Kang

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

**DOMESTIC JOURNAL**

**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 TranScription(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, Suh-yune 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

*Journal of Digital Convergence* 18.6 (2020) pp. 271–277. 2020

## DOMESTIC CONFERENCE

### BackTranScripton (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

### 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

Suhylene 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

**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**, Minho 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, Minho 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, Minho 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**

**Natural Language Processing Bible**

Heuiseok Lim, **Korea University NLPAL Lab**

Human Science, <https://www.aladin.co.kr/shop/wproduct.aspx?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