UPSTAGE · KOREA UNIVERSITY · NATURAL LANGUAGE PROCESSING & ARTIFICIAL INTELLIGENCE LAB

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

□ (+82) 10-2580-9066 | **□** chanjun.park@upstage.ai / bcj1210@naver.com | **□** parkchanjun | **□** bcj1210

## Research Interests

I am currently working as an AI Research Engineer at Upstage. Also 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 Al
- · 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
- Al for Accessibility
- Al for Education
- Language Resource and Evaluation
- Semantic Search and Information Retrieval
- CX (Customer Experience), UX (User Experience)
- · Quality Assurance

## **Education**

Korea University Seoul, South Korea

M.S/Ph.D in Computer Science and Engineering, GPA: 4.46 / 4.5

· Advisor: Heuiseok 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

Advisor: Pummo Ryu

Full Scholarship

Busan, South Korea

Sep. 2019 - Current

Mar. 2015 - Exp. Aug. 2018

## Academic Services

COLING 2022 South Korea

VIRTUAL SOCIAL CHAIR

Sep.2021-Present

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

### Program committe

PROGRAM COMMITTE

January.2021-Present

ACL 2022, ACL 2022-Insights, NAACL 2022, NAACL 2022-Industry Track

CHANJUN. PARK · CURRICULUM VITAE

South Korea

## **Selected Publications**

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

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

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

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

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

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

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

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

IEEE Access (2022). IEEE, 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, Imatitikua D Aiyanyo, Kinam Park, Aram So, Sungmin Ahn, Jeongbae Park

Mathematics 10.8 (2022) p. 1335. MDPI, 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, 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, Heuiseok Lim

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

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

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

#### 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, 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

Jeiyoon Park, Chanhee Lee, Chanjun Park, Kuekyeng Kim, Heuiseok 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, Heuiseok Lim *NeurIPS 2021 - Data-centric Al (DCAI) workshop*, 2021

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

Chanjun Park, Kuekyeng Kim, Yeong Wook Yang, Minho Kang, Heui Seok 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

2021.12	Naver Ph.D. Fellowship, 2021, 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	1st Place in Flitto Hackathon (Team Lead), 2020, 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	Best practices for using NIA AI training data(Korean-English Neural Machine Translation model), AI Hub	South Korea
2019.10	<b>Best Paper Award, 2019,</b> The 31st Annual Conference on Human & Cognitive Language Technology (HCLT2019)	South Korea
2010 10		
2019.10	1st Place Microsoft AI Accessibility Hackathon in Korea (Team Lead) , Microsoft	South Korea
	1st Place Microsoft Al Accessibility Hackathon in Korea (Team Lead), Microsoft  Graduate School Associate Scholarship, Sungkyunkwan University	South Korea South Korea
2019.03 2018.10	Graduate School Associate Scholarship, Sungkyunkwan University  Next Generation Information Processing NLP Competition 2018: Participation Award, Next-generation	South Korea
2019.03 2018.10 2017.06	Graduate School Associate Scholarship, Sungkyunkwan University  Next Generation Information Processing NLP Competition 2018: Participation Award, Next-generation information computing technology development business	South Korea
2019.03 2018.10 2017.06 2015-2019	Graduate School Associate Scholarship, Sungkyunkwan University  Next Generation Information Processing NLP Competition 2018: Participation Award, Next-generation information computing technology development business  Bit Computer Excellence Award (President Award), Bit Computer	South Korea South Korea South Korea

## **Selected Research Experience**

Product Search Pack
Seoul, South Korea

RESEARCHER IN UPSTAGE

- 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) Return on Advertising Spend (ROAS) Prediction Research

Seoul, South Korea

March. 2022. - Present

RESEARCHER IN KOREA UNIV

• 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 deposition 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)
- Skills: ROAS, Regression, Pytorch

CHANJUN. PARK · CURRICULUM VITAE

December. 2021. - March. 2022.

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

Seoul. South Korea

December. 2021. - March. 2022. RESEARCHER IN KORFA UNIV

· 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

#### **Simultaneous Speech Translation Research**

Seoul, South Korea December. 2021. - Present

RESEARCHER IN KOREA UNIV

- · 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

- · 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 KORFA UNIV

July. 2021. - December. 2021.

December 2021 - June 2022

- · 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 Generation and Reasoning Research**

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - Present

- KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation
- Publication of research paper in NAACL 2022
- 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

Seoul, South Korea

June. 2021. - December. 2021.

RESEARCHER IN KOREA UNIV

• Al 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 June. 2021. - Present

RESEARCHER IN KOREA UNIV

- 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**

RESEARCHER IN KOREA UNIV

Seoul South Korea

• 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

April. 2021. - December. 2021

CHANJUN, PARK · CURRICULUM VITAE

## (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 con-
- 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 January. 2021. - Present

RESEARCHER IN KOREA UNIV

· Proposed BackTranScription(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. - 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 January. 2021. - Present

RESEARCHER IN KOREA UNIV

- APE research based on Cross Lingual Pretrained Language Model
- Adpater 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
- 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
- · Skills: Neural Machine Translation, Pytorch, Quality Estimation, OpenKiwi, Huggingface, Fairseq

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

Seoul, South Korea

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

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, Al-based Korean-English translation model suitable for the security field, and research trend of the latest Al-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 April 2020 - Present

RESEARCHER IN KOREA UNIV

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

#### **Ancient Korean Neural Machine Translation Research**

Seoul, South Korea March. 2020. - Present

RESEARCHER IN KORFA UNIV

- 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
- Skills: Pytorch, Python, REST API, Flask, Neural Machine Translation

#### **Deep Learning based Grammar Error Correction Research**

Seoul, South Korea

June. 2019. - June. 2020

RESEARCHER IN KOREA UNIV

- 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
- $\bullet \ \ \text{Publication of research paper in ELECTRONICS(SCIE): } \\ \textbf{https://www.mdpi.com/2079-9292/9/10/1562\#cite} \\ \textbf{or } \\ \textbf{o$
- · 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
- 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
- · Automatic Essay Scoring Research
- · PEEP-Talk: Deep Learning-based English Education Platform for Personalized Foreign Language Learning
- · Al student: A Machine Reading Comprehension System for the Korean College Scholastic Ability Test
- · AI for TOEIC: Solving TOEIC Problems Using Simple and Efficient-based Data Augmentation Techniques
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

#### AI for Accessibility Research

RESEARCHER IN KOREA UNIV

Seoul, South Korea 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/
- Skills: Pytorch, Python, Google Cloud Platform, REST API, Nao Toolkit

CHANJUN, PARK · CURRICULUM VITAE

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

Seoul, South Korea

RESEARCHER IN KORFA 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

September. 2019. - January. 2021.

RESEARCHER IN KOREA UNIV

- 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 Groud 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

#### 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 Aug. 2018. - July. 2019

RESEARCHER IN SYSTRAN

• 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 June. 2018. - June.2019

RESEARCHER IN SYSTRAN

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

#### Planning of English speaking evaluation platform service

Seoul, South Korea

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**

Seoul South Korea Aug. 2018. - July. 2019

- Research for Automatic Speech Recognition Technology based on Kaldi
- Acoustic Model, Language Model
- Speaker Verification

RESEARCHER IN SYSTRAN

- English speaking evaluation
- · Real-time speech recognition system
- Skills: Kaldi

CHANJUN, PARK · CURRICULUM VITAE

#### **Neural Machine Translation Research**

Seoul. South Korea

RESEARCHER IN SYSTRAN Aug. 2018. - July. 2019

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

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

Seoul, South Korea

Sep. 2016. - June. 2018

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)

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

Seoul, South Korea

RESEARCHER IN BUES

Sep. 2017. - June. 2018

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

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

Seoul, South Korea Sep. 2016. - Aug. 2017

RESEARCHER IN BUFS

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

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

Seoul, South Korea

RESEARCHER IN BUFS

· Exobrain Project

Oct. 2015. - Feb. 2016

## Professional Experience \_\_\_\_\_

**Upstage** Seoul, South Korea

Al Research Engineer

March 2022 - Present

- Product Search Team
- Establishment of evaluation system to develop Product Search Pack and responsible for data, planning, and evaluation
- Proceeding with a sematic search-based commercialization project with AMOREPACIFIC
- · Semantic Search project with LG U+

**SYSTRAN** Seoul, South Korea

RESEARCH ENGINEER

Sep. 2018 - July. 2019

- 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
- · laas 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 Seoul. South Korea

RESEARCH INTERN

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
- Developed Simultaneous Speech Translation (implemented using JAVA, Python and Kaldi)



**Programming** JAVA, Python, C / C++

**Machine Learning & Deep Learning** Pytorch, Keras, Tensorflow, Huggingface, OpenNMT, Kaldi

Natural Language Processing Machine Translation, Speech Recognition, Neuro-Symbolic Computing

Full Stack Html, CSS, Javascipt, Java FX, Oracle, MySQL, Spring Boot, VertX, AWS

**Strategy & Planning** API Platform Service, Platform Service Planning, Gravitee IO, Paas, Iaas, Q.A, Q.C, Mockup, Swagger

**Etc** MS Office, Redmine, Linux, Docker, Web RTC, Latex

**Language Skills** Korean, English, Japanese, Thai, Myanmar

## Patents \_\_\_\_\_

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

## Demo\_\_\_\_\_

2019.12	Deep Learning based English Grammar Error Correction, http://nlplab.iptime.org:32292/	South Korea
2019.12	Deep Learning based Korean Spelling Correction, http://nlplab.iptime.org:32288/	South Korea
2020.01	Korea University Neural Machine Translation Platform, http://nlplab.iptime.org:32296/	South Korea
2020.02	Neural Symbolic Inference, http://nlplab.iptime.org:32245/	South Korea
2020.02	English Free Talking and Grammar Error Correction Feed Back System using NAO Robot,	South Korea
	https://www.youtube.com/watch?v=R16dG_4cnFk	South Noted
2020.03	Korean Ancient Neural Machine Translation Platform, http://nlplab.iptime.org:32242/	South Korea
2022 02	PicTalky: Text to Pictogram Demo for Children with Developmental Disabilities,	Courtle I/ conserv
2020.03	https://pictalky-311.de.r.appspot.com/	South Korea
2020.05	COVID-19 Machine Translation Platform, http://nlplab.iptime.org:32250/	South Korea
2020.07	Ancient Korean Translation & Machine Reading Comprehension with NAO,	South Korea
2020.07	https://www.youtube.com/watch?v=nXBZe-t9a-4&t=8s	
2020.11	Bilingual Education Using NAO, https://www.youtube.com/watch?v=4kCMMa62LDY&t=9s	South Korea
2021.04	Ko-CSS (Korean Colloquial-Sentence-Splitter) for simultaneous speech translation,	South Korea
2021.04	http://nlplab.iptime.org:32285/	
2021.06	BTS (BackTranScription) for STT Post-Processor, https://backtrans-317413.de.r.appspot.com/	South Korea
2021.09	Translation platfrom with AIHUB dataset, http://nlplab.iptime.org:32247/	South Korea
2021.09	Con-Talky, http://nlplab.iptime.org:32245/	South Korea
2021.09	Korea University English Grammar Corrector KakaoTalk Channel, http://pf.kakao.com/_cqxkps	South Korea
2021.09	PEEP-Talk, http://pf.kakao.com/_FlDxgs	South Korea
2021.10	A Tool for generating High-Quality Parallel Corpus considering Human Translation Market,	Courtle I/ conserv
	http://nlplab.iptime.org:9090/	South Korea
2021.10	A New Tool for Efficiently Generating Quality Estimation Datasets, http://nlplab.iptime.org:9091/	South Korea
2021.11	A Self-Supervised Automatic Post-Editing Data Generation Tool, http://nlplab.iptime.org:9092/	South Korea

**Invited Talk & Teaching Experience** 

Dongguk University South Korea

Artificial intelligence and Machine Translation lectures (3 times)

Invited by Dongguk University, Department of English Language and Literature

Korea University South Korea

TEACHING ASSISTANT

PRESENTER

Autumn 2021

January, 2022

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

**Busan Social Welfare Development Institute** 

South Korea July. 2021

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 South Korea

**HEAD TEACHING ASSISTANT** 

Autumn 2020

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

(DFE610) Natural Language Processing for Digital Finance Engineering

Korea University South Korea

TEACHING ASSISTANT Spring 2020

• (COSE461)Natural Language Processing

• (DFC615)Artificial Intelligence and Natural Language Processing

• Lecture materials: https://github.com/Parkchanjun/KU-NLP-2020-1

**LLsoLLu** South Korea

PRESENTER March. 2020

Latest natural language processing trends

· Language Representation, Neural Symbolic Computing

Invited by LLsoLLu Global Strategy and Planning Team

NC SOFT South Korea

Presenter February. 2020

• Technology Transfer Seminar

• Korean Spelling Corrector Using Automatic Noise Generation Methodology

• Invited by NC SOFT Language Al Team

Dongguk University South Korea

PRESENTER January. 2020

• 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

SKC South Korea

INSTRUCTOR Oct. 2019 - Nov.2019

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
- Lecture materials(Introduction of Machine Translation): https://github.com/Parkchanjun/SKC\_MachineTranslation

Naver Tech Talk South Korea

PRESENTER Aug. 2019

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 South Korea

SK I Academy

- Machine Translation for Everyone
- Invited by SK Planet

INSTRUCTOR

- Introduction to Natural Language Processing, RBMT, SMT, NMT, OpenNMT
- Lecture videos: https://tacademy.skplanet.com/live/player/onlineLectureDetail.action?seq=165

Aug. 2019

Sooyoungro Church South Korea

INSTRUCTOR Aug. 2017 - Dec. 2017

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

## Korea Press Foundation South Korea

INSTRUCTOR

· Reading mentoring

Aug. 2015 - Dec.2015

## **All Publications**

#### **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, Heuiseok 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, Heuiseok 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, Heuiseok 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, Heuiseok 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, 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, Heuiseok Lim

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

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

Chanjun Park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo, Heuiseok 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, Heuiseok 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, Heuiseok Lim NeurlPS 2021 - Data-centric Al (DCAI) workshop, 2021

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

Seolhwa Lee, Kisu Yang, **Chanjun Park**, João Sedoc, Heui-Seok Lim NerlPS 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, Heuiseok Lim NeurlPS 2021 - Data-centric Al (DCAI) workshop, 2021

## INTERNATIONAL JOURNAL

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

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

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

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

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim, (Equal Contribution (First Co-Author)) *IEEE Access* 10 (2022) pp. 44964–44973. IEEE, 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

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

Hyeonseok Moon, Chanjun Park, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim, (Equal Contribution (First Co-Author)) *IEEE Access* (2022). IEEE, 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, Imatitikua D Aiyanyo, Kinam Park, Aram So, Sungmin Ahn, Jeongbae Park

Mathematics 10.8 (2022) p. 1335. MDPI, 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

Jeiyoon 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 BackTranScription

**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

# Al 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

Suhyune 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 Chanjun Park, Jang Mook Kang

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

#### 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 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 Jongsung 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

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

## DOMESTIC CONFERENCE

#### 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 Chracteristics

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

#### 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 Langugae Processing Bible**

Heuiseok Lim, Korea University NLPAI Lab

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

## **Extracurricular Activity**

#### Operate Naver's natural language processing blog

South Korea 2016.Dec-Present

**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

MEMBERS

STUDENT

STUDENT

REPORTER

**OpenNMT Forum Members** 

South Korea

2019.July-2019.July

• Active as OpenNMT Forum Members

• Develop and share OpenNMT Pytorch, Tensorflow, Library Tutorial

South Korea

ETRI's machine translation quality evaluator

• Quality evaluation of machine translation hosted by ETRI

2019.October-2019.December

Completion of basic curriculum for natural language processing and language intelligence

2019.July-2019.July

South Korea

Korea University Human-Inspired Al & Computing Research Center

Completed the natural language processing tutorial (2016-2019)

South Korea

2016 - 2019

Special Interest Group of Human and Cognitive Language Technology

Reporter, Korea Electronics and Telecommunications Research Institute (ETRI)

South Korea

2016.Feb-2018.Feb

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

Flitto Supporters South Korea

• The first plito supporters

2017.Dec-2017.Dec

Journalist of the Ministry of Science and ICT

South Korea 2016.Feb-2017.Feb

REPORTER

PLANNER

STUDENT

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

Planning a video lecture for Natural Language Processing Basic at BUFS

South Korea

• URL: https://bit.ly/2NmFJV1

2016.July-2016.July

**BIT Computer Advanced Course Completion** 

South Korea 2016. January - 2016. August

• Bet Advanced Course Completion and Bet President Award

- Bit project level 2 license
- · Development of intelligent English learning APP

South Korea

BIT Computer Short Course Completion
Student

2015.June-2015.August

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