

Chanjun Park

UPSTAGE · KOREA UNIVERSITY

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 & Data Manager at Upstage. Also I am a Ph.D candidate in the Department of Computer Science and Engineering at Korea University (Advisor: Heuiseok Lim). I am the founder and leader of KU-NMT Group. I have published more than 110 papers in the field of Natural Language Processing, especially Machine Translation, and applied for 13 patents.

- Data-Centric AI
- Neural Machine Translation
- Language Resource and Evaluation
- Data Management and Engineering

Education

Korea University

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

Seoul, South Korea

Sep. 2019 - Aug. 2023

- 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

Busan, South Korea

Mar. 2015 - Exp. Aug. 2018

- Advisor: Pummo Ryu
- Full Scholarship

Selected Publications

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

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

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

Priming Ancient Korean Neural Machine Translation

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

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

A Self-Supervised Automatic Post-Editing Data Generation Tool

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Honors & Awards

2021.12	Naver Ph.D. Fellowship, 2021 , Naver	South Korea
2021.10	Best Paper Award (NLP Application 2 Section), 2021 , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.10	Best Paper Award (Language Resource Section), 2021 , The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)	South Korea
2021.10	Best Paper Award (QA and Speech Section), 2021 , 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	Best Paper Award, 2020 , 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	Best Paper Award, 2019 , The 31st Annual Conference on Human & Cognitive Language Technology (HCLT2019)	South Korea
2019.10	1st Place Microsoft AI Accessibility Hackathon in Korea (Team Lead) , Microsoft	South Korea
2019.03	Graduate School Associate Scholarship , Sungkyunkwan University	South Korea
2018.10	Next Generation Information Processing NLP Competition 2018: Participation Award , Next-generation information computing technology development business	South Korea
2017.06	Bit Computer Excellence Award (President Award) , Bit Computer	South Korea
2015-2019	Full Scholarship , BUFS	South Korea
2016-2017	Scholarship for academic excellence , Sooyoungro Church	South Korea

Selected Research Experience

(Project) LGU + VOD Search POC Project	Seoul, South Korea
RESEARCHER IN UPSTAGE	June. 2022. - Present
<ul style="list-style-type: none">Gold Testset and evaluation system design for tag extraction model comparisonResponsible for designing search system evaluation. Planning and designing blind tests for comparison with commercial systems.Skills: Customer Experience (CX), Quality Assurance (QA)	
(Project) Amorepacific Search commercialization Project	Seoul, South Korea
RESEARCHER IN UPSTAGE	March. 2022. - Present
<ul style="list-style-type: none">Participated in Amorepacific's AP Mall search commercialization projectStability test planning and designResponse time testing planning and designEstablishment of systematic QA process based on bug testingSkills: Customer Experience (CX), Quality Assurance (QA)	

(Project) NAVER PAPAGO - Parallel Corpus Filtering and Mining

Seoul, South Korea

RESEARCHER IN KOREA UNIV

December. 2021. - June. 2022.

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

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

Seoul, South Korea

RESEARCHER IN KOREA UNIV

June. 2021. - December. 2021.

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

(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://nlpplab.iptime.org:32285/>

(Project) Domain Specialized Neural Machine Translation (Security)

Seoul, South Korea

RESEARCHER IN KOREA UNIV

May. 2020. - November.2020

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

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

Seoul, South Korea

RESEARCHER IN KOREA UNIV

April. 2020. - Present

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

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

Seoul, South Korea

RESEARCHER IN KOREA UNIV

September. 2019. - August. 2022

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

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

Seoul, South Korea

RESEARCHER IN SYSTRAN

June. 2018. - July.2019

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

(Project) CJ Logistics Speech Recognition Agent

Seoul, South Korea

RESEARCHER IN SYSTRAN

Aug. 2018. - July. 2019

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

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

Seoul, South Korea

RESEARCHER IN BUFS

Sep. 2016. - June. 2018

- 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 BUFS

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

Professional Experience

Upstage

Seoul, South Korea

AI RESEARCH ENGINEER, DATA MANAGER

March. 2022 - Present

- (Concurrent Position) Data & Product Search Team
- Responsible roles related to search system evaluation
- 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
- VOD based 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
- 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

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

Skills

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

Patents

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

Invited Talk & Teaching Experience

Kyungsung University

PRESENTER

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

South Korea

August. 2022

Hankuk University of Foreign Studies

PRESENTER

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

South Korea

July. 2022

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

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
- Lecture materials(Introduction of Machine Translation): 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