Youngeun Koo

Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, Korea
✉ sarah8835@skku.edu

Research Interest

- Pragmatics, Computational Linguistics
- Dialogue summarization
- Speech act, intention recognition
- Conversational Analysis
- Dialog system, Chatbot
- Simultaneous interpretation system

Education

2018.3 – 2023.2 **Sungkyunkwan University**, Department of German Linguistics and Literature
Ph.D. of Arts (German Linguistics)
(Advisor: *Prof.* Munpyo Hong)

A Computerlinguistic Study on Automatic Summarization of German Dialog

This paper analyzed the sequential and hierarchical structure of German dialog and proposed a machine learning-based automatic summarization model. Dialog summarization is to shorten a given dialogue into short sentences containing the most important content of the dialogue. In this paper, I pointed out the problem that the existing dialogue summarization methodology relies only on the surface information of utterances, such as words. It is difficult to automatically summarize dialog using only surface information because phenomena such as omissions, corrections, and fillers often appear in dialogs. In addition, since a dialog is a special type of language use in which two or more speakers participate, it is very important to consider the interaction between speakers.
The automatic summarization model proposed in this paper was implemented using sequential and hierarchical structure analysis of dialogs. This model goes through the stages of data preprocessing, sequential structure analysis, hierarchical structure analysis, and summary generation. Analyzing the sequential structure of a dialog is to segment a dialog into the interactions to find out which interactions are made in the dialog. Analyzing the hierarchical structure of a dialog is to extract the most important and dominant utterance pairs from each interactional units, to know which are the most important utterances of each speaker in order to carry out the interaction.

2016.3 – 2018.2 **Sungkyunkwan University**, Department of German Linguistics and Literature
M.S. of Arts (German Linguistics)
(Advisor: *Prof.* Munpyo Hong)

A Study on Speech Act and Automatic Speech Act Classification of German Dialog

This paper proposes linguistic features of speech act classification, builds an automatic speech act classification model, and, based on linguistic analysis and machine learning result, figures out the principle of conveying and understanding conversation. We understand other's utterance intention with no big difficulty, though it is not explicitly mentioned in an utterance or is different from what is mentioned in utterance. This paper started from a curiosity about this. This paper proposed speech act categories by analyzing German tutorial dialogue. Then, by applying these categories, I built a speech act annotated corpus. Based on this, sentential features and contextual features are proposed as linguistic features that determine the speech act. Sentential features contain information that is based on the sentence itself and contextual features contain one that is based on the previous sentence.
To validate the proposed method, this paper conducted a machine learning experiment using WEKA, a Java-based machine learning toolkit. The 'Support vector machine(SVM)' was used as the machine learning algorithm and experiment results were obtained based on the '10-fold cross validation' method. We set unigram and bigram as a baseline of the experiment since this study aims to point out that what is explicitly mentioned in an utterance is not enough to accurately classify speech act. As a result, when all of the features proposed in this study are used, the accuracy was 75.13%, 32%p higher than the baseline.

2012.3 – 2016.2 **Sungkyunkwan University**, Department of German Linguistics and Literature
Bachelor of Arts (German Language and Literature)
Bachelor of Economics (International Trade and Policy)

2009.3 – 2012.2 **Jeonnam Foreign Language High School**, Department of Chinese

Publication

Automatic Classification of Speech Acts in German Daily Conversation using Machine Learning
Youngeun Koo
Journal of German Linguistics, Vol. 48, pp. 33-53, 2023
This paper collected and analyzed MoCoDa2 corpus, which consists of SNS messenger WhatsApp conversation data, to analyze German daily conversations. First, we compared SNS messenger conversations, which have strong characteristics of daily conversation, with tutorial dialogue, which are goal-oriented conversations, in the aspect of speech act types that appear in each conversation. Then we conducted a machine learning experiment applying the speech act classification methodology, proposed in previous research, to German daily conversations. Through the experiment, we intend to verify applicability of the methodology proposed in previous research. Furthermore, this study conducts additional experiments to classify utterances into few representative speech act types rather than many detailed speech acts, and find optimal feature combinations using dimensionality reduction techniques.

A Study on Writing Practice of Beginner German Learners using ChatGPT
Youngeun Koo and Munpyo Hong
Journal of German Language and Literature, Vol. 100, pp. 47-69, 2023
This study discusses two ways to utilize ChatGPT for practicing writing skills for German learners of beginner level. First, it is used as a proofreading tool to review the learner's written text. Second, it is used to create a learning materials of example texts that beginners can use as a reference when writing German text. Furthermore, this study analyzes the linguistic characteristics of ChatGPT's German proofreading results and written texts and then analyzes the effectiveness of ChatGPT as a learning tool. To this end, we evaluate the written texts of beginner German learners using various linguistic indicators and identify the impact of using ChatGPT on the writing practice of beginner learners.

A Study on Key-utterance Pair Extraction for Automatic Summarization of German Dialog
Youngeun Koo and Munpyo Hong
Journal of German Linguistics, Vol. 45, pp. 21-39, 2022
This paper proposes an automatic extraction method of a key-utterance from a dialog. To utilize extracted key-utterances as a summarization, this paper argues for a key-utterance pair extraction. This method bears some resemblance to extractive summarization, but it differs in various aspects. Especially, the main purpose in this work is to emphasize conversational features of an utterance and to utilize them for better dialogue understanding and summarization. Also, this work extracts utterances in-pair from a dialogue, whereas previous studies extracted key utterances individually. It is intended to consider an interaction between speakers.

Automatic Segmentation of German Dialog – Towards Automatic Dialogue Summarization
Youngeun Koo and Munpyo Hong
Journal of German Language and Linguistics, Vol. 160, pp. 77-98, 2021
This paper proposes linguistically motived features for automatic segmentation of german dialog. In particular, for a dialogue summarization, this paper emphasizes segmentationa into interactonal unit, instead of topical unit.

Automatic Speech Act Classification for Domain-independent Dialogue
Youngeun Koo and Munpyo Hong
Journal of German Linguistics, Vol. 39, pp. 25-48, 2019
This paper presents examples showing that some utterances are difficult to be understood as one certain speech act type, but can be understood as several types of speech act. To overcome this problem, this paper proposes hierarchical structure of speech act, which consists of representative and concrete speech act type, by comparing two different dialogue, German tutorial dialogue and German telephone call. Then, this paper applies it to classify speech act of an utterance. Ultimately, this paper asserts that this hierarchical structure will be effective for domain-independent speech act classification.

A Linguistic Study of Speech Act and Automatic Speech Act Classification for Korean Tutorial Dialog ***[Invited Paper]***
Youngeun Koo, Jiyoun Kim, Munpyo Hong, and Youngkil Kim
Journal of KIISE, Vol. 45(8), pp. 807-815, 2018
This paper proposes linguistically motived features for speech act classification: 9 sentential features and 4 contextual features.
(This paper is an invited paper, extended from Koo et al.(2017) which was nominated as best paper of HCLT 2017.)

Conferences (Talk/Poster)

Computational Pragmatics based Approach for Conversation Processing of Artificial Intelligence
Youngeun Koo
In Proceeding of the World Convention of the Korean Language, 2023.
This talk attempts to analyze speech acts of utterances and extract conversation summaries based on computational pragmatics, which combines pragmatic theory and computer technologies such as machine learning and deep learning, so that humans can understand the intent of utterances and the overall conversation. The mechanism for understanding key content was discussed. Furthermore, we attempted to examine the applicability of this computer-aided conversation study in conversation processing of artificial intelligence.

Towards a Linguistically Motivated Segmentation for a Simultaneous Interpretation System
Youngeun Koo
In Proceeding of The Korea Association of Foreign Languages Education, 2023.
This presentation explained the operating principles of ChatGPT and discussed ways to use ChatGPT in teaching German writing. We proposed ways for beginner learners at A1 to A2 levels to practice German writing using ChatGPT, and discussed an experiment conducted on 15 students at Sungkyunkwan University.

Linguistic Research using Deep Learning Models – focused on BERT and GPT
Youngeun Koo
In Proceeding of the Korean Psycholinguistics Association PIP Workshop, 2023.
In this presentation, major deep learning methodologies such as ANN, DNN, RNN, LSTM, Attention, and Transformer were explained to examine trends in language processing technology using deep learning. In particular, we focused on natural language processing research using Transformer-based BERT and GPT, and demonstrated a Korean chatbot model using sbert and a question-and-answer model using kogpt.

A Linguistic Analysis on Automatic Summarization of German Dialog
Youngeun Koo and Munpyo Hong
In Proceeding of the Koreanische Gesellschaft für Germanistik(KGG), 2022.
This paper discusses a difference between text summarization and dialogue summarization and emphasizes key characteristics of dialogue summarization. In particular, this paper argues for linguistic, especially conversational analysis based, approach to dialogue summarization.

Automatic Dialogue Segmentation for German Dialogue Summarization
Youngeun Koo
In Proceeding of the Fall Linguistic Society of Korea(LSK) Young Scholar Symposium, 2021.
This paper deals with a method of segmenting a dialogue into an interactional unit for dialogue summarization. As an useful features for dialogue segmentation, this paper proposes linguistically motivated features, such as prosodic, lexical, semantic, and conversational features. To validate proposed method, a machine learning based experiment was conducted. As a result, by the help of proposed features, the accuracy of dialogue segmentation experiment showed 83.56%.

Segmentation Methods for Different Speech Rate in Simultaneous Interpretation
Youngeun Koo, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, and Sung-Kwon Choi
In Proceeding of the 32nd Annual Conference on Human & Cognitive Language Technology(HCLT), Vol. 32, pp. 369-374, 2020.
This paper points out diverse circumstances in simultaneous interpretation, such as different speech rates varied by speakers and change of speech rate during a talk, and searches for a method to adapt to them. To do so, this paper proposes a dynamic segmentation method, a 'personalization method': measure a standard speech rate of a speaker through previous n-segmentation units and, based on the current speech rate compared with the standard speech rate, change segmentation method in real-time.

Towards a Linguistically Motivated Segmentation for a Simultaneous Interpretation System ***[SCOPUS]***
Youngeun Koo, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, Sung-Kwon Choi
In Proceeding of the 34th Pacific Asia Conference on Language, Information and Computation(PACLIC), 2020.
This paper deals with various segmentation methods for a simultaneous interpretation. Through experiments on Korean-to-English and English-to-Korean simultaneous interpretation, this paper investigates optimal segmentation method and analyses it based on typological aspects that each language possesses.

A Study of Segmentation Unit for the Real-time Simultaneous Interpretation System
Youngeun Koo, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, and Sung-Kwon Choi
In Proceeding of the 31th Annual Conference on Human & Cognitive Language Technology(HCLT), Vol. 31, pp. 229-235, 2019.
This paper aims to investigate translation unit for simultaneous interpretation system. To do so, this paper focuses on 'segmentation' technique and proposes various features for segmentation based on general linguistics and cognitive linguistics: prosodic, syntactic, semantic, pragmatic and cognitive feature.

Metonymy Resolution on Neural Approach
Taesun Whang, Chanhee Lee, Kisu Yang, Dongyub Lee, **Youngeun Koo**, Taehee Jeon, Heuiseok Lim
In Proceeding of the 31th Annual Conference on Human & Cognitive Language Technology(HCLT), Vol. 31, pp. 375-379, 2019.
This paper deals with metonymy, a linguistic concept for referring 'a thing' by the name of 'something' closely associated with that thing. This figurative expression is problematic in language processing because one expression can be interpreted in various meaning. This paper applies 'deep learning' approach to metonymy resolution: LSTM, BERT, XLNet and RoBERTa.

A Study for Categorizing Relations Between Headword and Aliases

Jiyoun Kim, **Youngeun Koo**, and Yongjun Zhu

In Proceeding of the 82nd Annual Meeting of The Association for Information Science and Technology(ASIS&T), 2019.

This paper focuses on terms, which have different forms, but denote the same single object, such as 'United Nation' and 'UN': headword and aliases. By using Korean Wikipedia data, this paper extracts data of headword and aliases and then categorizes relations between headword and aliases based on extracted data.

Automatic Speech Act Classification of Korean Dialogue based on the Hierarchical Structure of Speech Act Categories *[SCOPUS]*

Youngeun Koo, Jiyoun Kim, and Munpyo Hong

In Proceeding of the 33rd Pacific Asia Conference on Language, Information and Computation(PACLIC), pp. 432-441, 2019.

This paper points out that some utterances are difficult to be understood as one certain speech act type, but can be understood as several types of speech act, with some examples mentioned in paper. To overcome this problem, this paper proposes that speech act can be divided into two types: representative speech acts and concrete speech acts. By comparing two different type of dialogue, Korean tutorial dialogue and Korean telephone call, this paper structuralize hierarchical structure of speech act categories.

A Linguistic Study of Speech Act and Automatic Speech Act Classification for German Dialog

Youngeun Koo

In Proceeding of the Koreanische Gesellschaft für Germanistik(KGG), 2017.

This talk deals with speech act classification for German tutorial dialog. Especially, this talk focuses on linguistic features that decides speech act of an utterance. In this talk, sentential features, such as sentence type, subject and tense, and contextual features, such as previous speech act and adjacency pair, are discussed.

A Study of Speech Act Classification for German Dialog

Youngeun Koo and Munpyo Hong

In Proceeding of the Korean Society for Language and Information(KSLI), 2017.

This talk deals with speech act classification for German tutorial dialog. Especially, this talk focuses on linguistic features that decides speech act of an utterance. In this talk, sentential features, such as sentence type, subject and tense, and contextual features, such as previous speech act and adjacency pair, are discussed.

A Linguistic Study of Automatic Speech Act Classification for Korean Dialog *[nominated as 'Best Paper']*

Youngeun Koo, Jiyoun Kim, Munpyo Hong, and Youngkil Kim

In Proceeding of the 29th Annual Conference on Human & Cognitive Language Technology(HCLT), Vol. 29, pp. 17-22, 2017.

This paper deals with speech act classification for Korean tutorial dialog. Especially, this paper focuses on linguistic features that decides speech act of an utterance. In this paper, sentential features, such as sentence type, subject and tense, and contextual features, such as previous speech act and adjacency pair, are proposed.

SCHOLARSHIPS

- **Global Ph.D. Fellowship Program** (2018.03~2021.02)
  Funded by National Research Foundation of Korea (Ministry of Education)
  (Research scholarship (KRW 60 million) + Tuition (KRW 30 million))
- **Graduate School Excellent Student Scholarship** (2018.03~2020.02)
  Funded by Sungkyunkwan University
  (app. KRW 8 million))
- **Graduate School Excellent Student Scholarship** (2016.03~2018.02)
  Funded by Sungkyunkwan University
  (app. KRW 8 million))

AWARDS

- **Best Professor Award**
  Language Technology and Cultural Content (CNT3002-01) [Spring 2023]
  Department of Culture and Technology, Sungkyunkwan University, Seoul, Korea
- **Best Professor Award**
  Cultural contents and natural language processing (CNT3054-01) [Fall 2022, Fall 2023]
  Department of Culture and Technology, Sungkyunkwan University, Seoul, Korea

- **Best Paper Award**
  Koo et al.(2017), "A Linguistic Study of Automatic Speech Act Classification for Korean Dialog
  Oral Presentation at the 29th Annual Conference on Human & Cognitive Language Technology
  Invited Paper at Journal of KIISE(Korean Institute of Information Scientists and Engineers)

Research Experiences

2023.09 – 2024.08 **National Research Foundation of Korea(NRF)**, Ministry of Education, Seoul, Korea
"A Study on Communication Patterns in German Daily Conversation Using Deep Learning Methodology"
Project Leader

2023.06 – 2026.05 **National Research Foundation of Korea(NRF)**, Ministry of Education, Seoul, Korea
"Cognitive-semantic interpretation model of modern German noun compounds using deep learning and neuroscience methodologies"
Researcher

2021.09 – 2021.12 **National Institute of Korean Language(NIKL)**, Seoul, Korea & **Buzz Metrix**, Seoul, Korea
"Research on Construction of Sentiment Analysis Corpus and Usage Trend"
Researcher
This study examines various up-to-date corpus of english and korean sentiment analysis and its usage. Through the analysis on them, this study proposed a guideline for construction of korean sentiment analysis courpus, i.e. corpus size, domain, and usage. This study also deals with emotional analysis and proposed a general guideline for the construction of its corpus.

2020.03 – 2020.12 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea
"A Study on Segmentation Unit for Real-time Simultaneous Interpretation"
Project Manager & Researcher
This study attempts to find a segmentation point that achieves balance between translation accuracy and translation speed, based on various linguistic features. Moreover, this study proposes a method that can adapt to various circumstances in simultaneous interpretation, such as different speech rates varied by speakers and change of speech rate during a talk.

2019.05 – 2019.12 **SK Telecom Co., Ltd**, Seoul, Korea
"2019 NLP Core Test and Data Construction"
Project Manager & Researcher
This study aims to improve performance of SKT KLP engine: frequent queries of NUGU and new entity. To be specific, this study deals with sentence structure and KMA(Korean morphological analysis) results of queries and news headlines.

2019.04 – 2019.11 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea
"A Study on Cognitive-Pragmatic Model of Simultaneous Interpretation"
Project Manager & Researcher
This study applies 'turn construction unit(TCU)', a topic of 'Conversational analysis(CA)', to segmentation unit of simultaneous interpretation. For this, this study analyzes relation between them and proposes various features based on general linguistics and cognitive linguistics: prosodic, syntactic, semantic, pragmatic and cognitive feature.

2018.05 – 2018.12 **SK Telecom Co., Ltd**, Seoul, Korea
"NLP Core Test and Data Construction"
Project Manager & Researcher
This study aims to improve performance of SKT KLP engine. To be specific, this study deals with KMA(Korean morphological analysis) results of headwords in Wikipedia and POI(place of interest) in 'T-map' service.

2018.03 – 2020.12 **National Research Foundation of Korea(NRF)**, Ministry of Education, Seoul, Korea
"Automatic Speech Act Classification of German Dialog using Machine Learning"
Project Leader
This study deals with speech act recognition of an utterance in german dialog. To do so, this study analysed speech act categories in german dialog and scrutinized various issues in speech classification.

2018.03 – 2018.11 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea
"A Study on Segmentation Unit of Simultaneous Interpretation Considering Language-dependent Features"
Project Manager & Researcher

This study aims to discover segmentation unit, considering language-dependent features, for simultaneous interpretation. To find appropriate segmentation rules, this study analyzes English, Korean and Chinese. In addition, this study proposes 'rhetorical structure marker' as one of the most important feature for segmentation and collects rhetorical structure markers of English, Korean and Chinese.

2017.04 – 2017.11 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea
"A Study on Correlation between Segmentation Unit of Simultaneous Interpretation and Translation Rate"
Project Manager & Researcher
This study explains two main characteristic of simultaneous translation, translation quality and translation latency. When a translation unit is short, translation latency is short but quality is relatively low, and vice versa. This study aims to find translation unit with suitable length and quality. In this respect, this study proposes optimal segment length.

2017.03 – 2017.10 **SK Telecom Co., Ltd**, Seoul, Korea
"QA Corpus Construction"
Project Manager & Researcher
This study aims to design and construct data for QA system based on Wikipedia. To be specific, a 'natural' questions asking for a definition, property and yes/no answer were collected.

2016.06 – 2016.11 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea
"A Linguistic Study on Translation Unit for Simultaneous Interpretation System"
Project Manager & Researcher
This study analyzes translation techniques of human simultaneous translator and stresses 'segmentation' technique. In this sense, this study proposes some rules for segmenting translation unit for simultaneous interpretation system based on linguistic analysis.

TEACHING EXPERIENCES

German Writing for German Studies Ⅰ (GER2025-02)
Department of German Language and Literature, Sungkyunkwan University, Seoul, Korea
Instructor
- Spring 2023
- German Grammar, German sentence patterns

Language Technology and Cultural Content (CNT3002-01)
Department of Culture and Technology, Sungkyunkwan University, Seoul, Korea
Instructor
- Fall 2021, Spring 2022, Spring 2023
- Linguistics(Morphology, Syntax, Sematics, Pragmatics)
- Python, NLTK(Natural Language Toolkit), KoNLPy(Korean NLP in Python), Dialogflow
- Machine learning using Weka, Scikit-learn

Cultural Content and Natural Language Processing (CNT3054-01)
Department of Culture and Technology, Sungkyunkwan University, Seoul, Korea
Instructor
- Fall 2022, Fall 2023
- Python, Data crawling, Text preprocessing & representation
- POS tagging, Dependency parsing, Semantic role labeling, WSD, NER, Keyword extraction, Topic modeling
- Sentiment analysis using Machine learning

Artificial Intelligence and Social Communication (FSE5013-01) (*Graduate)
Department of Future Humanities Social Entrepreneurship, Sungkyunkwan University, Seoul, Korea
Teaching Assistant
- Fall 2022
- History, Methodology and Applications of Artificial Intelligence
- Python, Text analysis
- Conversational Analysis and Dialogflow, ChatGPT and Prompt engineering

Introduction to German Linguistics Ⅱ (GER3025-01)
Department of German Language and Literature, Sungkyunkwan University, Seoul, Korea
Teaching Assistant
- Fall 2019
- Semantics, Pragmatics(Entailment, Presupposition, Implicature, Speech Act, Politeness Theory)
- Computational Pragmatics

Introduction to Computational Linguistics (ILI2002-01)

Department of Interdisciplinary Linguistics, Sungkyunkwan University, Seoul, Korea

Teaching Assistant

- Spring 2017
- Speech Act, Automatic Speech Act Classification

OTHER EXPERIENCES

2023.09 – 2024.01 **2023-2 Co-Deep Learning Project - Advisor Professor**

Hosted by The Knowledge Center for Innovative Higher Education of Sungkyunkwan University

- Advisee : 4 students (Department of Convergence Software, Business Administration, Artificial Intelligence Convergence/History, System Management)
- Project Thema: Korean pronunciation solutions provided by AI: Customized correction service based on tongue position

2023.11.13 **Lecture at SKKU Data Analysis Club 'DScover'**

- Natural Language Processing

2023.06.10 **Lecture at Tutorial of MIRAEELL Spring Conference**, Jeju, Korea

- Linguistic studies using deep learing models (Linguistics, Literature, History, Education)
- Tutorial 1: Keyword extraction using korean BERT embedding
- Tutorial 2: Text classification(Metonomy classification) using korean BERT model

2023.01.30 **Lecture at Workshop of KGDS Winter Conference**, Tongyeong, Korea

- Computer-linguistic approach on German Linguistics
- Natural Language Processing
- Key extraction and Hate speech detection using German BERT

2022.05.31 **Lecture at Bu-Gae Girls High School**, Incheon, Korea

2021.11.02

2021.06.15

- Language and Linguistics
- Applied Linguistics(Cognitive Linguistics, Computational Linguistics)

2019.11.21 **6th SGPF Annual Conference**

Society of Global Ph.D. Fellows(SGPF)

2019.08.05~16 **European Summer School in Logic, Language and Information(ESSLLI 31)**

University of Latvia, Liga, Latvia

2019.07.16 **Talk at Ji-Pyeong-Seon Middle School**, Gimje, Korea

Hosted by National Research Foundation of Korea (Ministry of Education)

- Language and Linguistics
- Interdisciplinary works of linguistics and computer science

2018.11.10 **5th SGPF Annual Conference**

Society of Global Ph.D. Fellows(SGPF)

2014.02 – 2014.08 **Internship at Samsung C&T Deutschland GmbH**

Schwalbach am Taunus, Germany

OTELINOX Sales Team

SKILLS

- Foreign Language
  - ✓ English(TOEFL 105, TOEIC 970) – Daily & professional conversation available
  - ✓ German – Daily & professional conversation available
  - ✓ Chinese(新HSK Grade 4) – Daily conversation available
- Programming Language
  - ✓ Python(Advanced level) – Data Analysis, Machine Learning, Deep Learning