Personal Information

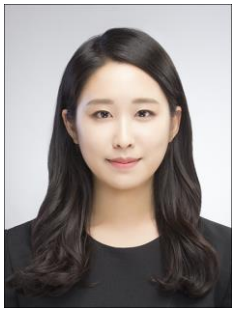
Name : Youngeun Koo

Date of Birth : 1993.12.06

Address : 25, Yangnyeongsi-ro, Dongdaemun-gu, Seoul, Korea

Tel : 010-6690-8835

E-mail : sarah8835@skku.edu



Education

2018.3 – **Sungkyunkwan University**, Department of German Linguistics and Literature

Ph.D. Student

(Advisor: Munpyo Hong)

2016.3 – 2018.2 **Sungkyunkwan University**, Department of German Linguistics and Literature

M.S. of Arts (German Linguistics)

(Advisor: Munpyo Hong)

Thesis: A Study on Speech Act and Automatic Speech Act Classification of German Dialog_based on German Tutorial 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, figure out the principle of conveying and understanding conversation. People, well and without much difficulty, understand other's utterance intention, 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.

First, this paper proposed speech act categories by analyzing German tutorial dialogue. Referring to existing tag-set, such as DAMSL, and preceding works, we suggested 25 speech act categories. Then, by applying these categories, we 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)

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

Research Interest

- Segmentation Unit for Simultaneous Interpretation System
- Pragmatics, Computational Pragmatics
- Conversational Analysis
- Dialog system, Chatbot

Publication

- **Youngeun Koo** and Munpyo Hong, "Automatic Speech Act Classification for Domain-independent Dialogue", 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.

- **Youngeun Koo**, Jiyoun Kim, Munpyo Hong, and Youngkil Kim, "A Linguistic Study of Speech Act and Automatic Speech Act Classification for Korean Tutorial Dialog", Journal of KIISE, Vol. 45(8), pp. 807-815, 2018. [Invited Paper]

  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)

- **Youngeun Koo**, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, and Sung-Kwon Choi, "Segmentation Methods for Different Speech Rate in Simultaneous Interpretation", 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.

- **Youngeun Koo**, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, Sung-Kwon Choi, "Towards a Linguistically Motivated Segmentation for a Simultaneous Interpretation System", 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.

- **Youngeun Koo**, Jiyoun Kim, Jungpyo Hong, Munpyo Hong, and Sung-Kwon Choi, "A Study of Segmentation Unit for the Real-time Simultaneous Interpretation System", 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.

- Taesun Whang, Chanhee Lee, Kisu Yang, Dongyub Lee, **Youngeun Koo**, Taehee Jeon, Heuiseok Lim, "Metonymy Resolution on Neural Approach", 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.

- Jiyoun Kim, **Youngeun Koo**, and Yongjun Zhu, "A Study for Categorizing Relations Between Headword and Aliases", 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.

- **Youngeun Koo**, Jiyoun Kim, and Munpyo Hong, "Automatic Speech Act Classification of Korean Dialogue based on the Hierarchical Structure of Speech Act Categories", 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.

- **Youngeun Koo**, "A Linguistic Study of Speech Act and Automatic Speech Act Classification for German Dialog", 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.

- **Youngeun Koo** and Munpyo Hong, "A Study of Speech Act Classification for German Dialog", 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.

- **Youngeun Koo**, Jiyoun Kim, Munpyo Hong, and Youngkil Kim, "A Linguistic Study of Automatic Speech Act Classification for Korean Dialog", In Proceeding of the 29th Annual Conference on Human & Cognitive Language Technology(HCLT), Vol. 29, pp. 17-22, 2017. [nominated as 'Best Paper']

  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 20 million) + tuition (Up to KRW 10 million))

Awards

- **Best Paper** (**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

2018.03 – **National Research Foundation of Korea(NRF)**, Ministry of Education, Seoul, Korea

"Automatic Speech Act Classification of German Dialog using Machine Learning"

Project Leader

2020.03 – 2020.12 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea

"A Study on Segmentation Unit for Real-time Simultaneous Interpretation"

Project Manager

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

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

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

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 – 2018.11 **Electronics and Telecommunications Research Institute(ETRI)**, Daejeon, Korea

"A Study on Segmentation Unit of Simultaneous Interpretation Considering Language-dependent Features"

Project Manager

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

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

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

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

2020.09 – 2020.12 **Language Technology and Cultural Content**

Department of Culture and Technology, Sungkyunkwan University, Seoul, Korea

Teaching Assistant

- Python, NLTK(Natural Language Toolkit), KoNLPy(Korean NLP in Python)
- Data Crawling, Twitter Crawling(Tweepy)
- Code Review

2019.09 – 2019.12 **Introduction to German Linguistics Ⅱ**

Department of German Language and Literature, Sungkyunkwan University, Seoul, Korea

Teaching Assistant

- Semantics, Pragmatics(Entailment, Presupposition, Implicature, Speech Act, Politeness Theory)
- Computational Pragmatics

2017.03 – 2017.06 **Introduction to Computational Linguistics**

Department of Interdisciplinary Linguistics, Sungkyunkwan University, Seoul, Korea

Teaching Assistant

- Speech Act, Automatic Speech Act Classification

Other Experiences

2021.03 – 2021.12 **Lecture at Bu-Gae Girls High School**, Incheon, Korea

- 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 app. 105, TOEIC app. 950) – Daily & professional conversation available
  - ✓ German(Level B1) – Daily conversation available
  - ✓ Chinese(新HSK Grade 4) – Daily conversation available
- Programming Language
  - ✓ Python