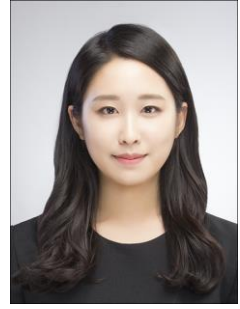


인적 사항

이름 : 구영은
생년월일 : 1993.12.06
주소 : 서울특별시 종로구 성균관로 25-2
이메일 : sarah8835@skku.edu



학력

2018.3 – 성균관대학교, 독어독문학과
박사과정(수료)
(지도교수: 홍문표)

2016.3 – 2018.2 성균관대학교, 독어독문학과
문학석사 (독어학)
(지도교수: 홍문표)

학위논문: 독일어 대화체 문장의 화행 분석과 화행 자동분류에 대한 연구_독일어 수업대화를 중심으로

본 논문은 화행 분류를 위한 언어학적 자질을 제안하고, 이를 기계학습에 적용하여 화행 자동분류 모델을 만들며, 언어학적 분석 및 기계학습 결과를 바탕으로 인간의 대화 수행 및 이해 원리를 파악하고자 하는 연구이다. 인간은 발화문 상에 드러나지 않더라도 혹은 발화문 상에 드러난 것과 다르더라도 상대방의 발화 의도를 어렵지 않게 잘 파악한다. 어떻게 그것이 가능한지에 대한 궁금증에서 본 논문은 시작되었다.

이를 연구하기 위해 먼저 본 논문에서는 독일어 수업대화에 나타나는 화행 유형을 분석하여 화행 유형 체계를 스스로 정립하였다. DAMSL 같은 기존에 제안된 발화 의도 태그셋과 기존연구에서 제안한 화행 유형을 참고하여, 총 25개의 화행 유형을 직접 수립하였다. 그 이후, 독일어 수업대화에 위의 화행 유형 체계를 적용하여 화행이 부착된 코퍼스를 만들었다. 이를 바탕으로 발화의 화행을 결정짓는 언어학적 자질을 문장자질과 문맥자질로 구분하여 제안하였다. 문장자질은 해당 문장 자체로부터 얻을 수 있는 정보이며, 문맥자질은 해당 문장이 아닌 그 이전 문장으로부터 얻을 수 있는 정보이다.

본 논문에서는 제안하는 방법론의 타당성을 검증하기 위해 기계학습 실험을 진행하였다. 실험은 자바 기반의 기계학습 툴킷을 이용하였고, 분류 알고리즘으로는 '서포트 벡터 머신(Support Vector Machine; SVM)'을, 평가 방식으로는 '10-fold 교차 검증(cross validation)'을 사용하였다. 문장에 드러난 정보만으로는 정확한 화행 분석이 어려운 점을 확인하기 위해, 유니그램(unigram)과 바이그램(bigram)을 자질로 사용한 경우를 베이스라인으로 잡았다. 그 결과 본 논문에서 제안한 다양한 화행 분류 자질들을 사용하였을 때 정확률 75.13%를 보이며, 베이스라인 대비 32%p 상승하였음을 확인할 수 있었다.

2012.3 – 2016.2 성균관대학교, 독어독문학과
문학학사 (독어독문학)
경제학학사 (국제통상학)

2009.3 – 2012.2 전남외국어고등학교, 중국어과

관심 분야

- 동시통역 시스템 분절단위
- 화용론, 전산 화용론
- 대화 분석
- 대화 시스템, 챗봇

학술지 논문

- **구영은, 홍문표, "범도메인 대화의 화행 자동분류", 한국독어학회, 독어학, Vol. 39, pp. 25-48, 2019.**

본 논문은 하나의 화행이 아닌 여러 화행으로 분석될 수 있는 예시들을 보여준다. 이 문제를 해결하기 위해 본 논문은 두개의 서로 다른 대화인 독일어 수업 대화와 독일어 전화 대화를 비교 분석하여, '대표 화행'과 '세부 화행'으로 구성된 화행 위계구조를 제안한다. 그 다음 이를 화행 분류에 적용한다. 결과적으로 본 논문은 이런 위계구조의 화행 유형 체계가 범도메인 대화의 화행 분류에 효과적일 것이라고 주장한다.

- **구영은, 김지연, 홍문표, 김영길, "한국어 수업 대화의 화행 분석과 화행 자동분류를 위한 언어학적 기반연구", 한국정보과학회, 정보과학회논문지, Vol. 45(8), pp. 807-815, 2018. [초청논문]**

본 논문은 한국어 수업 대화에 나타나는 화행을 분석하고, 화행 분류를 위한 언어학적 자질들로 9개의 문장 자질과 4개의 문맥 자질을 제안한다.

(본 논문은 우수논문으로 채택된 제 29회 한글 및 한국어 정보처리 학술대회의 '한국어 대화문 화행 자동분류를 위한 언어학적 기반연구'에서 확장된 초청논문임)

학술대회 발표(구두/포스터)

- **구영은, 김지연, 홍정표, 홍문표, 최승권, "발화자별 발화 속도를 고려한 실시간 동시통역 분절 방법론", 제 32회 한글 및 한국어 정보처리 학술대회(HCLT), Vol. 32, pp. 369-374, 2020.**

본 논문은 실시간으로 변화하는 동시통역의 상황에 적응가능한 동시통역 분절 방법론을 모색하는 데에 목표를 두고 있다. 이를 위해 이전 N개의 분절단위를 기반으로 발화자의 발화 속도를 측정하고, 발화의 빠르기에 따라 분절 방법론을 실시간으로 다르게 적용하는 개인화 기법을 제안한다.

- **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.**

본 논문은 동시통역 단위를 분절하는 다양한 분절 방법론을 다룬다. 영한, 한영 동시통역 분절 실험을 통해 언어쌍에 따른 최적의 분절 방법론을 찾고, 이를 원천 언어의 유형론적 특징에 의거하여 분석한다.

- **구영은, 김지연, 홍정표, 홍문표, 최승권, "실시간 동시통역 시스템 개발을 위한 통역 분절 단위 연구", 제 31회 한글 및 한국어 정보처리 학술대회(HCLT), Vol. 31, pp. 229-235, 2019.**

본 논문은 동시통역 시스템을 위한 번역 단위를 찾는 데에 목표를 두고 있다. 이를 위해 본 논문은 '분절' 기술에 초점을 맞추고, 일반 언어학과 인지 언어학을 기반으로 하는 다양한 분절 자질들을 제안한다. 여기에는 운율, 통사, 의미, 화용, 인지 자질이 해당한다.

- **황태선, 이찬희, 양기수, 이동엽, 구영은, 전태희, 임희석, "딥러닝 방식을 이용한 환유 해소", 제 31회 한글 및 한국어 정보처리 학술대회(HCLT), Vol. 31, pp. 375-379, 2019.**

본 논문은 어떤 것을 그것과 관련 있는 다른 것의 이름으로 지칭하는 방식을 가리키는 언어학적 개념인 환유를 다룬다. 이 수사학적 표현은 언어 처리에서 난해한 경우가 많다. 왜냐하면 한 표현이 여러 의미로 해석될 수 있기 때문이다. 본 논문에서는 LSTM, BERT, XLNet, RoBERTa 와 같은 딥러닝 방식을 환유 해소에 적용한다.

- **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.

본 논문은 하나 대상을 가리키지만 여러 형태를 갖는 용어들을 살펴본다. 예를 들어 'United Nation' 과 'UN' 이 그러하며, 이를 표제어와 알리아스라고 한다. 본 논문은 한국어 위키피디아 데이터를 활용하여 표제어와 알리아스 데이터를 추출하고 그 다음 추출된 데이터를 바탕으로 표제어와 알리아스 간의 관계를 유형화한다.

- **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.

본 논문은 예시를 통해 하나의 화행이 아닌 여러 화행으로 분석되는 발화들이 있음을 지적한다. 이런 문제를 해결하기 위해 본 논문은 화행을 두가지 유형으로 나눈다. 대표 화행과 세부 화행이다. 두개의 서로 다른 대화인 한국어 수업 대화와 한국어 전화 대화를 비교 분석하여, 본 논문은 화행 유형을 위계구조로 구조화한다.

- **구영은**, "독일어 대화체 문장의 화행 분석과 화행 자동분류를 위한 언어학적 기반연구", 한국독어독문학회(KGG), 2017년 겨울 학술대회, 2017.

본 발표는 독일어 수업 대화의 화행 분류를 다루었다. 특히 본 발표에서는 발화의 화행을 결정짓는 언어학적 자질에는 어떤 것이 있는지 논의하였다. 구체적으로는 문장 자질(문장 유형, 주어, 시제 등)과 문맥 자질(이전 발화의 화행, 인접쌍 등)에 대해서 설명하였다.

- **구영은, 홍문표**, "독일어 대화체 문장의 화행 분석과 화행 자동분류에 대한 연구", 한국언어정보학회(KSLI), 2017년 가을 정기 학술대회, 2017.

본 발표는 독일어 수업 대화의 화행 분류를 다루었다. 특히 본 발표에서는 발화의 화행을 결정짓는 언어학적 자질에는 어떤 것이 있는지 논의하였다. 구체적으로는 문장 자질(문장 유형, 주어, 시제 등)과 문맥 자질(이전 발화의 화행, 인접쌍 등)에 대해서 설명하였다.

- **구영은, 김지연, 홍문표, 김영길**, "한국어 대화문 화행 자동분류를 위한 언어학적 기반연구", 제 29회 한글 및 한국어 정보처리 학술대회(HCLT), Vol. 29, pp. 17-22, 2017. [우수논문]

본 논문은 한국어 수업 대화의 화행 분류를 다루고 있다. 특히 본 논문은 발화의 화행을 결정짓는 언어학적 자질을 찾는 것에 초점을 맞추고 있다. 본 논문에서는 문장 자질(문장 유형, 주어, 시제 등)과 문맥 자질(이전 발화의 화행, 인접쌍 등)이 제안되었다.

장학금

- **글로벌 박사 펠로우십** (2018.03~2021.02)

한국연구재단(교육부) 지원

연 3천만원 (연구 장학금 (2천만원) + 등록금 (최대 1천만원))

수상 경력

- **우수 논문** (구영은 외(2017), "한국어 대화문 화행 자동분류를 위한 언어학적 기반연구")

2017년도 제29회 한글 및 한국어 정보처리 학술대회에서 발표

『정보과학회논문지 - 소프트웨어 및 응용』에 초청논문 투고

연구 과제

2018.03 – **한국연구재단(NRF), 교육부**

"기계학습을 이용한 독일어 대화문의 화행 자동분류"

Project Leader

2020.03 – 2020.12 **한국전자통신연구원(ETRI)**

“실시간성을 고려한 동시통역의 분절단위 연구”

Project Manager

본 연구는 번역 정확도와 번역 속도를 모두 고려한 동시통역 단위의 분절 자질을 찾는 것을 목표로 한다. 또한 동시통역의 다양하고, 실시간으로 변화하는 상황에 적용할 수 있는 분절 방법론을 제안한다.

2019.05 – 2019.12 **SK Telecom Co., Ltd**

“19년 NLP Core 테스트 및 데이터 구축”

Project Manager

본 연구는 SKT KLP 엔진의 NUGU 빈출 쿼리와 신규 개체명 처리 성능 향상을 목표로 한다. 구체적으로는 쿼리와 뉴스기사 제목의 문장구조 및 형태소 분석 결과를 다룬다.

2019.04 – 2019.11 **한국전자통신연구원(ETRI)**

“동시통역 과정에 대한 인지-화용론적 모델 연구”

Project Manager

본 연구는 ‘대화 분석’의 주제인 ‘말차례 구성 단위’ 개념을 동시통역의 분절 단위에 적용한다. 이를 위해 본 연구는 말차례 구성 단위와 분절 단위의 관련성을 분석하고 일반언어학 및 인지언어학을 기반으로 하는 다양한 자질을 제안한다. 여기에는 운율, 통사, 의미, 화용, 인지 자질이 해당한다.

2018.05 – 2018.12 **SK Telecom Co., Ltd**

“NLP Core 테스트 및 데이터 구축”

Project Manager

본 연구는 SKT KLP 엔진의 성능 향상을 목표로 한다. 구체적으로는 위키피디아의 표제어와 티맵 POI의 형태소 분석 결과를 다룬다.

2018.03 – 2018.11 **한국전자통신연구원(ETRI)**

“언어별 특성을 반영한 동시통역 분절단위 연구”

Project Manager

본 연구는 언어별 특성을 반영한 동시통역 분절단위를 상정하고자 한다. 적절한 분절 규칙을 찾기 위해, 본 연구는 영어, 한국어, 중국어를 분석한다. 또한 본 연구는 ‘담화구조 표지’를 분절의 중요한 자질 중 하나로 제안하고 영어, 한국어, 중국어 담화구조 표지를 수집한다.

2017.04 – 2017.11 **한국전자통신연구원(ETRI)**

“동시통역 분절단위와 번역률간의 상관관계에 대한 연구”

Project Manager

본 연구는 동시통역의 두가지 주요 특징인 번역 정확도와 번역 지연을 설명한다. 번역 단위가 짧을 때 번역 지연은 짧지만 번역 정확도는 비교적 낮다. 번역 단위가 길 때는 그 반대이다. 본 연구는 적절한 번역 단위의 길이와 정확도를 만족하는 번역 단위를 찾고자 한다. 이러한 관점에서 본 연구는 최적의 분절 길이를 제안한다.

2017.03 – 2017.10 **SK Telecom Co., Ltd**

“QA 코퍼스 구축 연구”

Project Manager

본 연구는 위키피디아 기반의 QA 시스템을 위한 데이터 설계 및 구축을 목표로 한다. 구체적으로는 정의형, 속성형, Yes/no 질문 등의 ‘자연스러운’ 질의 데이터를 구축한다.

2016.06 – 2016.11 **한국전자통신연구원(ETRI)**

“동시통역 시스템 개발을 위한 통역단위의 언어학적 연구”

Project Manager

본 연구는 인간 동시통역사의 번역 기술을 분석하고 '분절' 기술의 중요성을 강조한다. 이러한 점에서 본 연구는 언어학적 분석을 바탕으로 동시통역 시스템의 번역 단위를 분절하는 규칙들을 제안한다.

강의 경력

2020.09 – 2020.12 언어공학과 문화콘텐츠

컬처앤티크놀로지융합전공, 성균관대학교

실습 조교

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

2019.09 – 2019.12 독어학개론 II

독어독문학과, 성균관대학교

실습 조교

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

2017.03 – 2017.06 컴퓨터언어학입문

융합언어학연계전공, 성균관대학교

실습 조교

- Speech Act, Automatic Speech Act Classification

외부 활동

2021.03 – 2021.12 인천 부개여자고등학교 강의

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

2019.11.21 6th SGPF Annual Conference

글로벌 박사 펠로우즈(Society of Global Ph.D. Fellows; SGPF)

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

University of Latvia, Liga, Latvia

2019.07.16 김제 지평선중학교 강연 및 멘토링

주최: 한국연구재단(교육부)

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

2018.11.10 5th SGPF Annual Conference

글로벌 박사 펠로우즈(Society of Global Ph.D. Fellows; SGPF)

2014.02 – 2014.08 삼성물산 독일법인 인턴쉽

Schwalbach am Taunus, Germany

오텔리눅스 판매팀(OTELINOX Sales)

개인 능력

- 외국어
 - ✓ 영어(TOEFL 105점, TOEIC 950점) – 일상 및 전문 대화 가능
 - ✓ 독일어(B1) – 일상 대화 가능
 - ✓ 중국어(新HSK 4급) – 일상 대화 가능
- 프로그래밍 언어
 - ✓ Python