

한국 문화 유산 설명에 사용된 단어들과 수화 단어 간의 비교 연구

서종우¹, 박현정¹, 김지용¹, 김도현¹, 김태욱¹, 전문구^{1, 4*}
한국문화기술연구소¹, 광주과학기술원^{2*}

sjw007s@korea.ac.kr, jjeong6230@gmail.com, kako000@naver.com, thdfks1128@gmail.com,
goor4@nate.com, mgjeon@gist.ac.kr

A study for comparison between words used to explain Korean cultural heritage and sign language words

Jongwoo Seo¹, Hyunjeong Pak¹, Jiyong Kim¹, Dohyeon Kim¹, Taewook Kim¹, Moongu Jeon^{1, 4*}

Korea Culture Technology Institute¹, Gwangju Institute of Science and Technology^{2*}

요 약

약 2 만 5 천개에 달하는 수화 단어를 이용해서 훨씬 많은 단어를 가진 한국어로 쓰여진 문화 유산 설명들을 전달하기 위해서는 문화해설문에 사용된 한국어 단어와 수화의 동작으로 존재하는 단어를 비교하여 그대로 번역할 수 있는지 확인할 필요성이 있다. 본 논문은 문화 유산 설명에서 사용되는 단어와 수화 단어들을 비교하기 위해 형태소 분석을 진행하였다. 실험을 통해서 Komoran 형태소 분석기가 가장 성능이 좋았음을 알 수 있었다. 또한 문화 유산 해설에 사용된 단어들 중 거의 절반에 해당하는 단어들이 수화 단어로 존재하지 않았음을 확인하였다.

I. 서론

소리를 듣지 못하는 청각장애인들에게 수화로 제공되는 문화 유산에 대한 정보가 매우 적기 때문에 기존 정보를 수화로 번역하여 전달하는 기술이 필요하다. 그러나 약 2 만 5 천개에 달하는 수화 단어를 이용해서 훨씬 많은 단어를 가진 박물관에 있는 한국 문화 유산 설명들을 전달하기 위해서는 문화 유산 설명들에 사용된 한국어 단어와 수화의 동작으로 존재하는 단어를 비교하여 그대로 번역할 수 있는지 확인할 필요성이 있다. 본 논문은 문화해설문에서 사용되는 단어와 수화 단어들을 비교하기 위해 형태소 분석을 진행하였다. 실험을 위해 국립국어원에 존재하는 약 2 만 5 천개의 수화 단어들을 수집하였고, 수화 사전에 있는 문화정보 수어 동영상 1600 개에 있는 문화 유산 설명들을 수집하였다. 실험을 통해서 Hannanum[1], Kkma[1], Twitter[1], Komoran[1] 중에서 Komoran 형태소 분석기가 가장 좋았음을 확인하였다.

II. 관련 연구

수화와 관련되어 많은 연구가 진행된 것 중에 하나로 수화 동작 인식이 있다. [2-8]들을 포함하여 매우 많은 논문들이 수화 동작 인식에 관한 방법을 제안하였다. 수화 동작 인식은 손 동작에 관련한 데이터를 통해 손 동작이 어떤 단어를 의미하는지 인식하는 분야이다. 이 분야에서는 어떠한 뜻을 가진 단어에 해당하는 수화 동작이 있다고 전제한다. 그래서 한국어 단어들 중에 수화 동작이 없는 단어들에 대해 어떻게 처리해야 할지에 대한 고려가 없다. 그 외에 다른 수화 관련

기술은 거의 사례를 찾아보기 힘들다. 그러므로 한국어로 구성된 문장을 수화 동작으로 번역할 때 일반적으로 얼마나 많은 단어들이 수화 단어로 있는지 연구해볼 필요성이 있다.

한국어 자연어 처리는 한국어를 컴퓨터가 자동적으로 처리하는 기술을 연구하는 분야이다. 영어는 가장 자연어처리가 발달한 언어이다. 이 언어를 처리하기 위해 여러 기술들이 해외에서 연구되었는데, 이러한 기술들을 한국어에 사용하기 위해서는 [9]와 같이 형태소 처리를 일반적으로 한다. 한국어는 뜻의 최소 단위인 형태소가 매우 중요한 역할을 하기 때문이다. 예를 들어 ‘먹’ 형태소는 ‘먹다’, ‘먹었다’, ‘먹을 것이다’ 등으로 활용되므로 형태소 단위로 분리하여 한국어 자연어 처리를 하는 것이 효과적이다. 그러므로 본 연구에서도 단어들을 형태소 단위로 분리할 필요성이 있다.

III. 실험

실험을 위해 데이터 수집을 다음과 같이 진행하였다. 본 연구에서는 문화 유산 설명들에 대한 단어들과 수화 단어들의 비교를 하기 위해 한국 수화 사전에 있는 문화정보 수어¹ 1625 개 동영상에 있는 문장들을 파이썬²으로 수집하였다. 또한 한국 수화 사전³에 있는

¹ http://sldict.korean.go.kr/front/museum/museumList.do?to_p_category=MUE

² <https://www.python.org/>

³ <http://sldict.korean.go.kr/front/main/main.do>

일상생활 수어와 전문용어 수어 약 2 만 5 천개 단어들을 수집하였다.

형태소 분석기로 파이썬에서 동작하는 Hannanum, Kkma, Twitter, Komoran 를 문화정보 수어 1625 개 동영상에 있는 문장들의 단어들에 적용시키고, 약 2 만 5 천개의 단어들에 적용시켰다. 그리고 수화 사전에 없는 형태소들의 비율을 퍼센트로 표 1 에 나타내었다.

표 1. 수화 단어 사전에 없는 단어 비율

형태소 분석기	Hannanum	Kkma	Twitter	Komoran
수화 사전에 없는 형태소 비율	69.08	56.27	69.34	54.8

표 1 에서 나타나듯이, 동일한 데이터셋에 대해서 형태소 분석기마다 다른 성능을 보였다. 없는 형태소 비율이 가장 적은 Komoran 형태소 분석기가 수화 단어 사전의 데이터들과 관련하여 가장 좋은 성능을 보여줌을 표 1 을 통해 알 수 있다.

구체적으로 어떤 형태소들이 없는지 나타내기 위해 표 2 를 만들었다.

표 2. 수화 사전에 없는 형태소 예시

수화 사전에 없는 형태소			
춘향전	국채보상운동	담양군	무등산
춘향	조선총독부	1684	플라스크
이몽룡의	태종실록	월출산	프로젝터
명종	태조	해남군	갈라파고스
대한매일신보	이성계	완도군	다윈
7 월	11 일	인조	km
베델	코리아나	은행나무	서울대

표 2 에서 주로 볼 수 있는 것들은 고유명사들이다. ‘춘향’, ‘베델’과 같은 이름들과 ‘월출산’, ‘완도군’과 같은 지명들은 수화 사전에 수화 동작으로 저장되어 있지 않다는 것이다. 실제로 문화 유산 설명에서는 역사적 인물들과 지명이 많이 나오기 때문에 이런 형태소들을 포함하는 문장을 수화로 표현하는 것이 다른 일반적인 문장보다 어려움을 알 수 있다. 그리고 형태소 분석기 자체의 오류로 인해 ‘이몽룡의’, ‘7 월’ 등과 같은 형태소들도 얻게 되었다.

IV. 결론

청각장애인들에게 수화 동작들로 이루어진 문화 유산 설명들은 적다. 그러므로 기존의 문화 유산 설명들을 전달하기 위한 기술 기초 연구가 필요하다. 본 논문에서는 기초 연구를 위해 먼저 문화 유산 설명에 사용된 단어들과 수화 단어 사전에 있는 단어들이 얼마나 일치되는지 형태소 분석기를 통해 조사하였다. 약 절반에 해당하는 형태소들이 수화 동작으로 정의되어 있지 않았고, 대부분은 역사적 인물을 뜻하는 형태소들이었다. 향후 연구로 없는 단어들을 어떤 수화 단어들로 대체할 수 있는지를 포함한 없는 단어 처리에 대한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2020 년도 문화기술연구개발 지원사업(R2020060002)으로 수행되었음.

참 고 문 헌

- [1] 박은정, 조성준, “KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지”, 제 26 회 한글 및 한국어 정보처리 학술대회 논문집, 2014.
- [2] 양희덕 and 이.J.정.소.및. 응용, 연속된 수화 인식을 위한 자동화된 Coarticulation 검출. 2009. 36(1): p. 82-91.
- [3] 양희덕 and 이.J.정.소.및. 응용, 연속된 수화 인식을 위한 자동화된 수화 및 지화 적출. 2011. 38(2): p. 102-107.
- [4] 정택위 and 김.J.한. 논문지, 딥러닝 방식의 웨어러블 센서를 사용한 미국식 수화 인식 시스템. 2020. 15: p. 291-297.
- [5] 전자공학회논문지-SC, 박.J., 모바일 수화 인식 시스템의 개선에 관한 연구. 2006. 43(4): p. 1-10.
- [6] 이재홍, 정은석, and 김.J.한.학. 논문집, 수화 인식에 대한 센서 시스템. 2011: p. 4-5.
- [7] 김상우, 김인숙, and 정.J.대. 학술대회, CNN 을 활용한 수화 번역 시스템. 2019: p. 754-755.
- [8] 논문지, 김.J.한., 수화 패턴 인식을 위한 2 단계 신경망 모델. 2012. 22(3): p. 319-327.
- [9] Joo, W.-T., Y.-S. Jeong, and K. Oh. Political orientation detection on Korean newspapers via sentence embedding and deep learning. in 2016 International Conference on Big Data and Smart Computing (BigComp). 2016. IEEE.