

AI Hub 한국어-영어 번역 말뭉치를 통한 영어 수동문의 한국어 번역 양상 분석

2021-14058 언어학과 이예은

1 연구 개요

1.1 연구 배경

영어를 한국어로 번역할 때의 어려움 중 하나는, 한국어에서는 능동문으로 쓰이는 문장들이 영어에서 빈번하게 수동문으로 쓰이는 경우일 것이다. 특히 영어 수동문에서의 'by+행위자'는 한국어로 '~에 의해'로 직역되어 번역투가 발생하기도 한다.

1.2 연구 목적

이 프로젝트에서는 AI Hub 한국어-영어 번역 말뭉치를 활용해 영어 수동문이 한국어에서 각각 피동문과 능동문으로 번역되는 양상을 비교 분석하고자 한다.

1.3 데이터 선정

이 목적을 달성하기 위해서는 실제 번역 양상을 담고 있고, 한국어와 영어 병렬 데이터가 문장 단위로 구성되어 있는 코퍼스를 사용해야 한다. 이 목적을 달성하기 위해서 AI Hub에서 공개한 한국어-영어 번역 말뭉치 중 대화체 데이터를 선택했다.

2 데이터 구축 및 활용 계획

2.1 데이터 수집

AI Hub에서 인증 절차를 거쳐 데이터를 다운로드하고, Korpora 패키지로 로딩할 수 있다.

2.2 데이터 가공

영어 수동문은 spaCy 라이브러리의 auxpass 태그를 통해 탐지하고, 한국어 피동문은 KoNLPy를 사용하여 피동 접사와 특정 수동 동사를 확인하여 감지할 수 있다.

2.3 데이터 활용

구축된 데이터프레임을 **pandas**로 분석하여 영어 수동문의 한국어 번역 양상을 비교합니다. 피동문과 능동문의 비율을 계산하고, 시각화 기법을 활용할 수 있다.

참고 문헌

김순영. (2012). <영한 번역에 나타난 번역투 문장>. 《새국어생활》 22(1), 67-79.

이승아. (2010). <영한 병렬 코퍼스에 나타난 영어 수동문의 한국어 번역>. 《영어영문학》, 56(5), 871-905.