

# 한국어 문장형 부정의 분포에 관해: 코퍼스 활용 연구

권성우 (KWON SUNGWOO) 2023-82030

November 6, 2023

## 1 연구개요

### 1.1 연구배경

한국어의 두가지 문장형 부정(안+동사, 동사+지 않)은 그 통사적/의미-화용론적 특성때문에 언어학내에서 많은 연구의 주제가 되어왔다. 하지만 이러한 연구들의 대부분은 이 두 부정형의 추상적 특성을 연구자의 직감이나 몇몇 한정적인 자료에 근거하여 이루어졌다(Kim 2009, Cho 1975). 이 연구에서는 실제 모국어 화자들의 발화로 구성된 코퍼스를 활용해 두 부정형태의 코퍼스내 분포를 기준으로 두 가지가 어떻게 다른지 연구해 볼 것이다.

### 1.2 연구목적

이 프로젝트를 통하여 살펴 보려고 하는 주제는 두 가지이다. 첫번째로는 전체적인 코퍼스내에서의 빈도분석을 통해 두 가지 부정형의 전체분포를 관찰/비교 해 보는 것이고, 두번째로는 특정 유형의 술어들(ex. 순수 한국어 술어 vs 한자 기원 술어)과의 관계 포착을 통해 특정 부정형과 서술어 유형간의 호환성을 이야기 해 볼 수 있을것이다.

### 1.3 데이터 선정

이 목적을 달성하기 위해서는 현실의 발화로 구성된, 자연스러운 대화맥락의 코퍼스를 사용해야한다. 이를 위해 국립국어원에서 공개하는 '모두의 말뭉치' 중 '구어 말뭉치'와 '일상 대화 말뭉치 2021'를 사용할 것이다.

## 2 데이터 구축 및 활용 계획

### 2.1 데이터 수집

'모두의 말뭉치'는 공식 사이트(<https://corpus.korean.go.kr>)에서 신청한 후 승인을 받아 다운로드할 수 있다.

## 2.2 데이터 가공

데이터 전처리와 추출에는 언어 데이터과학 시간에 배운 기술(정규표현식, Pandas library, 파이썬함수)들을 활용할 계획이다.

## 2.3 데이터 활용

필요한 데이터가 모두 수집되면 파이썬과 통계적 방법들을 사용해 빈도 분석과 호환성 분석을 진행할 것이다. 예를들어, 호환성 분석의 경우, 두개의 부정형을 각각 가장 많이 사용되는 술어들과 같이 단어풍선을 만들어 볼 수 있겠다.

## 3 참고문헌

- 국립국어원 (2022). 국립국어원 일상 대화 말뭉치 2021(버전 1.0).  
URL: <https://corpus.korean.go.kr>
- 국립국어원 (2022). 국립국어원 구어 말뭉치 (버전 1.2).  
URL: <https://corpus.korean.go.kr>
- Kim, T. 2009. Negation in Korean: its Grammaticalization and Lexicalization Process. *The Journal of Studies in Language* 24.4, 685-705
- Cho, C-H. 1975. The Scope of Negation in Korean. In H-M. Sohn (ed). *The Korean Language: Its Structure and Social Projection*, 63-80. Center for Korean Studies: University of Hawaii.