

한국어의 두가지 문장형 부정의 분포에 대한 코퍼스 연구

Stanislas Kwon

Linguistics and Data Science 2023F

November 22, 2023

현상 설명

- 한국어 문장형 부정 *안-* 과 *-지 않-*.
 - ▶ 점심 안 먹었어. (sfn)
 - ▶ 점심 먹지 않았어. (lfn)
- 그렇다면 두 가지를 어떻게 구분하는가? 함께 나타나는 술어에 제약이 있는듯 함.
 - ① 음절 수에 관한 제약. ?*안 불성실하다* (Song 1988)
 - ② 음운론적 제약. 술어가 '아'로 시작하면 *-지 않-* 과 함께. ?*안 아름답다* (Suh 1996)
 - ③ 화용론적 제약. *안-* 이 표면 구조상 덜 복잡함 (morpheme vs phrase). (McClanahan 1991)
- 모두의 말뭉치 코퍼스 분석을 통해 이 가설들이 맞는지 확인할 계획.
 - ▶ 빈도 분석.
 - ▶ 술어 호환성 분석.

진행 상황

```

sfn_counts = utterances['form'].str.count(r'^안\s|\s안\s').sum()
lfn_counts = utterances['form'].str.count(r'지\s않').sum()

print(sfn_counts)
print(lfn_counts)

```

[30]

... 20373
14273

```
from collections import Counter

print(Counter(forms.str.findall(r'\b\w+')).sum())
```

[28]

```
... Counter({'하고': 231, '하는': 124, '먹고': 116, '되는': 111, '해': 111, '쫄음': 109, '먹어': 107, '돼': 101, '할': 94, '해서':
```

+ Code

+ Marquage

```
from collections import Counter

print(Counter(forms.str.findall(r'(\w+)지\s\w+').sum()))
```

[7]

```
... Counter({'쉽': 818, '하': 613, '있': 456, '되': 366, '나쁜': 341, '줄': 300, '많': 293, '싫': 264, '그럴': 237, '어렵': 201,
```