

한국어의 두가지 문장형 부정의 분포에 대한 코퍼스 연구

Sungwoo S. Kwon

Linguistics and Data Science 2023F

November 17, 2023

현상 설명

- 한국어 문장형 부정 *안-* 과 *-지 않-*.
 - ▶ 점심 안 먹었어.
 - ▶ 점심 먹지 않았어.
- 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'\b안 (\w+)').sum()))
```

[28]

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

+ Code + Marquage

```
from collections import Counter

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

[7]

```
Counter({'원': 818, '하': 613, '있': 456, '되': 366, '나쁜': 341, '좋은': 300, '말': 293, '싫': 264, '그럴': 237, '어렵': 201, '안': 199, '못': 198, '아니': 197, '아니고': 196, '아니야': 195, '아니예요': 194, '아니에요': 193, '아니거든': 192, '아니잖아': 191, '아니잖아': 190, '아니잖아': 189, '아니잖아': 188, '아니잖아': 187, '아니잖아': 186, '아니잖아': 185, '아니잖아': 184, '아니잖아': 183, '아니잖아': 182, '아니잖아': 181, '아니잖아': 180, '아니잖아': 179, '아니잖아': 178, '아니잖아': 177, '아니잖아': 176, '아니잖아': 175, '아니잖아': 174, '아니잖아': 173, '아니잖아': 172, '아니잖아': 171, '아니잖아': 170, '아니잖아': 169, '아니잖아': 168, '아니잖아': 167, '아니잖아': 166, '아니잖아': 165, '아니잖아': 164, '아니잖아': 163, '아니잖아': 162, '아니잖아': 161, '아니잖아': 160, '아니잖아': 159, '아니잖아': 158, '아니잖아': 157, '아니잖아': 156, '아니잖아': 155, '아니잖아': 154, '아니잖아': 153, '아니잖아': 152, '아니잖아': 151, '아니잖아': 150, '아니잖아': 149, '아니잖아': 148, '아니잖아': 147, '아니잖아': 146, '아니잖아': 145, '아니잖아': 144, '아니잖아': 143, '아니잖아': 142, '아니잖아': 141, '아니잖아': 140, '아니잖아': 139, '아니잖아': 138, '아니잖아': 137, '아니잖아': 136, '아니잖아': 135, '아니잖아': 134, '아니잖아': 133, '아니잖아': 132, '아니잖아': 131, '아니잖아': 130, '아니잖아': 129, '아니잖아': 128, '아니잖아': 127, '아니잖아': 126, '아니잖아': 125, '아니잖아': 124, '아니잖아': 123, '아니잖아': 122, '아니잖아': 121, '아니잖아': 120, '아니잖아': 119, '아니잖아': 118, '아니잖아': 117, '아니잖아': 116, '아니잖아': 115, '아니잖아': 114, '아니잖아': 113, '아니잖아': 112, '아니잖아': 111, '아니잖아': 110, '아니잖아': 109, '아니잖아': 108, '아니잖아': 107, '아니잖아': 106, '아니잖아': 105, '아니잖아': 104, '아니잖아': 103, '아니잖아': 102, '아니잖아': 101, '아니잖아': 100, '아니잖아': 99, '아니잖아': 98, '아니잖아': 97, '아니잖아': 96, '아니잖아': 95, '아니잖아': 94, '아니잖아': 93, '아니잖아': 92, '아니잖아': 91, '아니잖아': 90, '아니잖아': 89, '아니잖아': 88, '아니잖아': 87, '아니잖아': 86, '아니잖아': 85, '아니잖아': 84, '아니잖아': 83, '아니잖아': 82, '아니잖아': 81, '아니잖아': 80, '아니잖아': 79, '아니잖아': 78, '아니잖아': 77, '아니잖아': 76, '아니잖아': 75, '아니잖아': 74, '아니잖아': 73, '아니잖아': 72, '아니잖아': 71, '아니잖아': 70, '아니잖아': 69, '아니잖아': 68, '아니잖아': 67, '아니잖아': 66, '아니잖아': 65, '아니잖아': 64, '아니잖아': 63, '아니잖아': 62, '아니잖아': 61, '아니잖아': 60, '아니잖아': 59, '아니잖아': 58, '아니잖아': 57, '아니잖아': 56, '아니잖아': 55, '아니잖아': 54, '아니잖아': 53, '아니잖아': 52, '아니잖아': 51, '아니잖아': 50, '아니잖아': 49, '아니잖아': 48, '아니잖아': 47, '아니잖아': 46, '아니잖아': 45, '아니잖아': 44, '아니잖아': 43, '아니잖아': 42, '아니잖아': 41, '아니잖아': 40, '아니잖아': 39, '아니잖아': 38, '아니잖아': 37, '아니잖아': 36, '아니잖아': 35, '아니잖아': 34, '아니잖아': 33, '아니잖아': 32, '아니잖아': 31, '아니잖아': 30, '아니잖아': 29, '아니잖아': 28, '아니잖아': 27, '아니잖아': 26, '아니잖아': 25, '아니잖아': 24, '아니잖아': 23, '아니잖아': 22, '아니잖아': 21, '아니잖아': 20, '아니잖아': 19, '아니잖아': 18, '아니잖아': 17, '아니잖아': 16, '아니잖아': 15, '아니잖아': 14, '아니잖아': 13, '아니잖아': 12, '아니잖아': 11, '아니잖아': 10, '아니잖아': 9, '아니잖아': 8, '아니잖아': 7, '아니잖아': 6, '아니잖아': 5, '아니잖아': 4, '아니잖아': 3, '아니잖아': 2, '아니잖아': 1})
```