

국립국어원 추론_확신성 분석 말뭉치 2020

(버전 1.1)

· **자료명:** 국립국어원 추론_확신성 분석 말뭉치 2020

· **공개일**

· (버전 1.0) 2021. 12. 1.

· (버전 1.1) 2022. 8. 25.

- 국립국어원 문어 말뭉치에서 추출한 담화 전체 제외

· **자료 유형:** 텍스트

· **관련 사업:** 말뭉치 합의 분석 및 연구(2020)

· **자료 설명**

※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구·조사 자료 > ‘말뭉치 합의 분석 및 연구’ 사업 결과보고서 참조(발간 예정)

· **내용**

- 국립국어원 신문, 문어, 준구어 말뭉치에서 대상 담화 추출

- 화자가 내포문에서 추출한 명제(가설)을 확인하는 정도를 설문 조사하여, 그 결과를 바탕으로 ‘합의(Entailment)’, ‘모순(Contradiction)’ 표지 부착

※ 합의(Entailment): 화자가 내포문의 명제(가설)이 참이라고 확인하는 경향

※ 모순(Contradiction): 화자가 내포문의 명제(가설)이 거짓이라고 확인하는 경향

- 모문은 선별된 (비)사실성 술어 목록과 합의 취소 운용소(부정, 의문, 조건, 양태 등)를 포함

- SuperGLUE의 과제 중 CommitmentBank를 참고

· **분량:** 총 796건 중 훈련 및 검증 말뭉치 공개

	훈련	검증	시험(비공개)
규모	243	42	511

· **파일 형식:** csv(UTF-8 인코딩)

· **파일 수 및 크기:** 파일 1개, 총 272KB

· **인용:** 국립국어원(2021). 국립국어원 추론_확신성 분석 말뭉치 2020(버전 1.0).

URL: <https://corpus.korean.go.kr>

· **데이터 구조와 예시**

	설명	예시
Genre	말뭉치 구분(문어, 신문, 구어)로 구분	문어
KCB_num	합의 분석 일련 번호	NE_003
IDX	원시 말뭉치 번호	NWRW1900000018.25972.9
AN3	거짓 확산(모순)(-3)으로 답한 응답자의 수	0
AN2	거짓 확산(모순)(-2)으로 답한 응답자의 수	0
AN1	거짓 확산(모순)(-1)으로 답한 응답자의 수	0
A0	모름 (0)으로 답한 응답자의 수	0
AP1	참 확산(합의)(+1)으로 답한 응답자의 수	1
AP2	참 확산(합의)(+2)으로 답한 응답자의 수	3
AP3	참 확산(합의)(+3)으로 답한 응답자의 수	4
cnt	총 응답자수	8
Mean	평균	2.375
SD	표준편차	0.744023809
CI_low	95% 신뢰구간의 하한 값	1.75298053
CI_high	95% 신뢰구간의 상한 값	2.99701947
*Class_Restrict	합의: 응답자의 80% 이상이 +1, +2, +3을 선택 모순: 응답자의 80% 이상이 -1, -2, -3 선택 중립: 응답자의 80% 이상이 0 선택	Entailment
Discourse	담화	성조와 발음이 아무리 중요하다 해도 '언어는 소통을 위한 도구임'을 기억하면 문제는 의외로 쉽게 풀린다.
Proposition	화자가 명시되고 원형이 복원된 내포 명제	언어는 소통을 위한 도구이다.
Factivity	술어 유형(fact, counter, none)	fact
Speaker	화자 유형	필자
Mat_Tense	모절 시제	현재
Inner_Tense	내포 명제 시제	현재
Mat_Person	모절 인칭	1
Inner_Person	내포 명제 인칭	3
Subject_Equal	모절과 내포 명제의 주어 일치 여부 (Y: 일치, N: 불일치)	N
ECO	합의 취소 운용소(부정, 조건, 의문, 양태의 조합)	조건
Modal	양태 구분	null
Verb	술어 원형	기억하다
Complement	보문소	ㅁ/음
Mat_Tense_Op	모절 시제소	null
Inner_Tense_Op	내포 명제 시제소	ㅣ
Modal_Rep	양태표지	null
Negation_Rep	부정표지	null

*de Marneff and Jiang(2019)의 기준에 따라 분류. 결과적으로 본 자료에는 ‘합의’와 ‘모순’ 표지만 부착

· **자료 내용 문의:** 02-2669-9614