

## 국립국어원 문법성 판단 말뭉치

(버전 1.1)

- 자료명: 국립국어원 문법성 판단 말뭉치
- 공개일
  - (버전 1.0) 2020. 8. 25.
  - (버전 1.1) 2021. 12. 8.
    - 도메인 내, 도메인 외 평가용 문장 및 답안 공개
- 자료 유형: 텍스트
- 관련 사업: 문장 문법성 판단을 위한 기초 자료 구축(2019)
- 자료 설명
  - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구·조사 자료 > 연구 보고서 > ‘문장 문법성 판단을 위한 기초 자료 구축’ 사업 보고서 참고
  - 내용
    - 학술 간행물(학술지 논문, 박사학위 논문, 학술 저서 등)에서 추출한 19,940개 문장(간행물 내 문법성, 비문법성 판단 예문)의 문법성 판단을 연령별(20~50대)로 조사하여 그 결과를 제시.
    - 평가용 말뭉치로 활용될 수 있도록 학술 간행물의 대표 저자를 기준으로 도메인 내 자료(623명), 도메인 외 자료(73명)로 분리함. 도메인 내 자료는 학습용 자료, 개발용 자료, 평가용 자료로, 도메인 외 자료는 개발용 자료, 평가용 자료로 분리함.
  - 분량
    - 총 19,940개 문장(문법적 문장 9,970개 문장, 비문법적 문장 9,970개 문장)
    - 도메인 내 학습용 15,876개 문장
    - 도메인 내 개발용 1,060개 문장
    - 도메인 내 평가용 1,060개 문장(비공개)
    - 도메인 외 개발용 972개 문장
    - 도메인 외 평가용 972개 문장(비공개)
- 파일 형식: TSV(UTF-8 인코딩)
- 파일 수 및 크기: 파일 4개, 총 3.19MB
- 인용:
  - (국문) 국립국어원(2020). 국립국어원 문법성 판단 말뭉치(버전 1.1). URL: <https://kli.korean.go.kr/corpus>
  - (영문) National Institute of Korean Language (2020). NIKL Grammaticality Judgment Corpus (v.1.1).

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

· 예시

source	acceptability _label	source _annotation	sentence	mean (total)	sd (total)	mean (20s)	sd (20s)	mean (30s)	sd (30s)	mean (40s)	sd (40s)	mean (50s)	sd (50s)
T00001	1		높은 달이 떴다.	4.981	2.273	5.371	2.059	5.273	2.198	4.095	2.508	4.667	2.41
T00001	0	*	달이 뜰이 높았다.	2.223	1.794	2.5	1.905	2.065	1.459	1.85	1.785	2.32	2.036
T00002	1		실없는 사람이 까불까불한다.	5.884	1.493	6.306	1.238	5.71	1.575	5.95	1.276	5.44	1.781
T00003	1		나는 철수에게 공을 던졌다.	6.784	0.836	6.971	0.169	6.923	0.27	6.55	1.395	6.353	1.367
T00004	1		내가 순이와 돌이서 다툰다.	4.286	2.187	4.735	2.079	4.152	2.181	3.583	1.976	4.714	2.644
T00004	0	*	내가 순이와 우리가 다툰다.	1.621	1.112	1.595	1.257	1.688	1.148	1.476	0.68	1.769	1.235
T00005	1		나는 부지런히 뛰었다.	6.68	0.942	6.892	0.315	6.781	0.608	6.429	1.326	6.231	1.691
T00005	0	?	나는 부지런히 뛰어졌다.	2.351	1.795	2.775	2.118	2.314	1.694	1.833	1.543	2.048	1.359
T00006	1		사랑이 죄는 아니다.	6.605	1.086	6.85	0.427	6.2	1.712	6.556	0.922	6.857	0.359
T00006	0	*	죄는 사랑이 아니다.	4.339	2.433	4.576	2.513	4.706	2.316	4.393	2.283	3.059	2.536
T00007	1		철수는 영미를 가면적으로 미워하였다.	3.554	2.151	3.909	2.31	3.059	1.969	3.464	2.027	4	2.345
T00007	0	*	철수에게는 영미가 가면적으로 미웠다.	3.02	2.044	2.813	1.975	3	2.118	3.5	1.96	2.867	2.232
T00008	1		그가 그녀에게 감기를 옮겼다.	6.465	1.162	6.563	1.19	6.206	1.321	6.85	0.489	6.333	1.291
T00008	0	*	그녀가 그에 의해서 감기를 옮겨졌다.	2.382	1.752	2.472	1.89	2.394	1.657	2.3	1.78	2.231	1.739

· 자료 내용 문의: 02-2669-9638