

국립국어원 의미역 분석 말뭉치

(버전 1.0)

- **자료명:** 국립국어원 의미역 분석 말뭉치
- **공개일**
 - (버전 1.0) 2021. 12. 1.
- **자료 유형:** 텍스트
- **관련 사업:** 의미역 분석 말뭉치 구축(2019)
- **자료 설명**
- **내용**
 - 문장의 술어가 가지는 논항을 분석하고 의미 역할을 부착한 말뭉치임.
 - 2018년 구축 국립국어원 신문 말뭉치에서 추출한 원시 말뭉치를 분석 대상으로 함.
 - 한국전자통신연구원(ETRI)의 ‘한국어 의존 의미역 주석 가이드라인’을 기반으로 하여 분석하였음.
 - 세부 분석 방법은 2019년 의미역 분석 말뭉치 구축 사업 결과 보고서를 참고할 것.
(국립국어원 누리집 > 자료 > 연구·조사 자료 > ‘의미역 분석 말뭉치 구축’ 사업 보고서)
- **분석 표지**

의미역 표지	정의
ARG0	서술어의 동작주, 행위자(agent), 경험주(experiencer)
ARG1	서술어의 피동작주(patient), 대상(theme) 등
ARG2	시작점(starting point), 수혜자(benefactive) 등
ARG3	도착점(ending point) 등
ARGA	장형 사동 구문(-게 하다)의 사동주
ARGM-LOC	장소 (locatives)
ARGM-DIR	방향 (directional)
ARGM-CND	조건 (condition)
ARGM-MNR	방법 (manner)
ARGM-INS	도구 (instrument)
ARGM-TMP	시간 (temporal)
ARGM-CAU	이유/원인 (cause)
ARGM-EXT	범위 (extent)

ARGM-PRD	보조 서술 (secondary predication)
ARGM-PRP	목적 (purpose clauses)
ARGM-DIS	담화 연결 (discourse)
ARGM-ADV	부사적 어구 (adverbials)
ARGM-NEG	부정 (negation)
ARGM-AUX	보조 용언

· 분량: 200만 어절

· 파일 형식: JSON(UTF-8 인코딩)

· 파일 수 및 크기: 파일 1개, 총 54.9MB(ZIP 파일 기준)

· 인용:

(국문) 국립국어원(2021). 국립국어원 의미역 분석 말뭉치(버전 1.0). URL: <https://kli.korean.go.kr/corpus>

(영문) National Institute of Korean Language (2022). NIKL Semantic Role Labeling Corpus (v.1.0).

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

· 예시

```

"sentence": [
  {
    "id": "NWRW1800000021.117.2.1",
    "form": "비타민 성분이 들어 있지 않은 비타민 음료가 무더기로 적발됐다.",
    "word": [
      {
        "id": 1,
        "form": "비타민",
        "begin": 0,
        "end": 3
      },
      {
        "id": 2,
        "form": "성분이",
        "begin": 4,
        "end": 7
      },
      {
        "id": 3,
        "form": "들어",
        "begin": 8,
        "end": 10
      },
      {
        "id": 4,
        "form": "있지",
        "begin": 11,

```

```

        "end": 13
      },
      {
        "id": 5,
        "form": "않은",
        "begin": 14,
        "end": 16
      },
      {
        "id": 6,
        "form": "비타민",
        "begin": 17,
        "end": 20
      },
      {
        "id": 7,
        "form": "음료가",
        "begin": 21,
        "end": 24
      },
      {
        "id": 8,
        "form": "무더기로",
        "begin": 25,
        "end": 29
      },
      {
        "id": 9,
        "form": "적발됐다.",
        "begin": 30,
        "end": 35
      }
    ],
    "SRL": [
      {
        "predicate": {
          "form": "들어",
          "begin": 8,
          "end": 10,
          "lemma": "들다"
        },
        "argument": [
          {
            "form": "비타민 성분",
            "label": "ARG1",
            "begin": 0,
            "end": 6,
            "word_id": 2
          },
          {
            "form": "있지",
            "label": "AGRM-AUX",
            "begin": 11,
            "end": 13,
            "word_id": 4
          }
        ]
      }
    ]
  }

```

```
{
  {
    "form": "않은",
    "label": "ARGM-NEG",
    "begin": 14,
    "end": 16,
    "word_id": 5
  },
  {
    "form": "비타민 음료",
    "label": "ARG2",
    "begin": 17,
    "end": 23,
    "word_id": 7
  }
]
},
{
  "predicate": {
    "form": "적발됐다.",
    "begin": 30,
    "end": 35,
    "lemma": "적발되다"
  },
  "argument": [
    {
      "form": "비타민 음료",
      "label": "ARG1",
      "begin": 17,
      "end": 23,
      "word_id": 7
    }
  ]
}
],
}
```

· 자료 내용 문의: 02-2669-9638