

## 국립국어원 구문 분석 말뭉치 (버전 2.0)

- **자료명:** 국립국어원 구문 분석 말뭉치
- **공개일**
  - (버전 1.0) 2020. 8. 25.
  - (버전 2.0) 2021. 3. 30.
    - 문어 구문 분석 말뭉치(버전 1.0)에서 CMP, DP 등에 대한 분석 오류 수정
    - 구어 구문 분석 말뭉치 추가
- **자료 유형:** 텍스트
- **관련 사업:** 구문 분석 말뭉치 구축(2019), 구문 및 무형 대용어 복원 말뭉치 연구 분석(2020)
- **자료 설명**
  - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구·조사 자료 > 2019년 ‘구문 분석 말뭉치 구축’과 2020년 ‘구문 및 무형 대용어 복원 말뭉치 연구 분석’ 사업 보고서 참고
  - **내용**
    - 문장의 구조를 파악하여 단어 간 관계를 분석한 말뭉치임.
    - 2018년 구축 국립국어원 신문 말뭉치와 구어 말뭉치에서 추출한 원시 말뭉치를 분석 대상으로 함.
    - 한국정보통신기술협회 표준 <의존 구문분석 말뭉치 구축을 위한 의존 관계 태그 세트 및 의존 관계 설정 방법> 지침에 따라 분석하였음.
  - **분석 표지**
    - 구문 표지: NP(체언), VP(용언), AP(부사구), VNP(긍정 지정사구), DP(관형사구), IP(감탄사구), X(의사구), L(부호-왼쪽 괄호 및 따옴표), R(부호-오른쪽 괄호 및 따옴표)
    - 기능 표지: SBJ(주어), OBJ(목적어), MOD(관형어), AJT(부사어), CMP(보어), CNJ(접속어)
  - **분량:** 총 300만 어절
    - 문어 200만 어절
    - 구어 200만 어절
  - **파일 형식:** JSON(UTF-8 인코딩)
  - **파일 수 및 크기:** 파일 2개

- 문어 1.07GB
- 구어 583MB

· 인용:

(국문) 국립국어원(2021). 국립국어원 구문 분석 말뭉치(버전 2.0). URL: <https://kli.korean.go.kr/corpus>

(영문) National Institute of Korean Language (2021). NIKL Dependency-Parsed Corpus (v.2.0). URL: <https://kli.korean.go.kr/corpus>

· 파일 명명 규칙

자리	1	2	3	4	5	6	7	8	9	10	11	12	13	14
속성	유형	구성 방법	주석 단계	구축 연도	일련번호(8자리)									
정의 값	N: 신문 S: 구어	X: 추출	DP: 구문 분석	19: 2019년 20: 2020년	00000001 ~ 99999999 (여덟 자리 일련번호)									
※ 예시: NXDP1902103231.json 2019년에 구축한 구어 자료 가공 구문 분석 파일 NXDP2002103180.json 2020년에 구축한 신문 자료 가공 구문 분석 파일														

· 예시

```

{
  "id": "NWRW1800000021.1.2.1",
  "form": "경기 성남시 판교신도시에서 이달 분양하는 중대형 아파트의 3.3m²당 분양가가 2006년보다 200만 원 정도 싼 1500만 원 후반대로 결정될 것으로 보인다.",
  "word": [
    {
      "id": 1,
      "form": "경기",
      "begin": 0,
      "end": 2
    },
    {
      "id": 2,
      "form": "성남시",
      "begin": 3,
      "end": 6
    },
    {
      "id": 3,
      "form": "판교신도시에서",
      "begin": 7,
      "end": 14
    },
    {
      "id": 4,
      "form": "이달",
      "begin": 15,

```

```
"end": 17
},
{
  "id": 5,
  "form": "분양하는",
  "begin": 18,
  "end": 22
},
{
  "id": 6,
  "form": "중대형",
  "begin": 23,
  "end": 26
},
{
  "id": 7,
  "form": "아파트의",
  "begin": 27,
  "end": 31
},
{
  "id": 8,
  "form": "3.3m2당",
  "begin": 32,
  "end": 38
},
{
  "id": 9,
  "form": "분양가가",
  "begin": 39,
  "end": 43
},
{
  "id": 10,
  "form": "2006년보다",
  "begin": 44,
  "end": 51
},
{
  "id": 11,
  "form": "200만",
  "begin": 52,
  "end": 56
},
{
  "id": 12,
  "form": "월",
  "begin": 57,
  "end": 58
},
{
  "id": 13,
  "form": "정도",
  "begin": 59,
  "end": 61
},
}
```

```

    {
      "id": 14,
      "form": "싼",
      "begin": 62,
      "end": 63
    },
    {
      "id": 15,
      "form": "1500만",
      "begin": 64,
      "end": 69
    },
    {
      "id": 16,
      "form": "원",
      "begin": 70,
      "end": 71
    },
    {
      "id": 17,
      "form": "후반대로",
      "begin": 72,
      "end": 76
    },
    {
      "id": 18,
      "form": "결정될",
      "begin": 77,
      "end": 80
    },
    {
      "id": 19,
      "form": "것으로",
      "begin": 81,
      "end": 84
    },
    {
      "id": 20,
      "form": "보인다.",
      "begin": 85,
      "end": 89
    }
  ],
  "DP": [
    {
      "word_id": 1,
      "word_form": "경기",
      "head": 2,
      "label": "NP",
      "dependent": []
    },
    {
      "word_id": 2,
      "word_form": "성남시",
      "head": 3,
      "label": "NP",

```

```

    "dependent": [
      1
    ]
  },
  {
    "word_id": 3,
    "word_form": "판교신도시에서",
    "head": 5,
    "label": "NP_AJT",
    "dependent": [
      2
    ]
  },
  {
    "word_id": 4,
    "word_form": "이달",
    "head": 5,
    "label": "NP_AJT",
    "dependent": []
  },
  {
    "word_id": 5,
    "word_form": "분양하는",
    "head": 7,
    "label": "VP_MOD",
    "dependent": [
      3,
      4
    ]
  },
  {
    "word_id": 6,
    "word_form": "중대형",
    "head": 7,
    "label": "NP",
    "dependent": []
  },
  {
    "word_id": 7,
    "word_form": "아파트의",
    "head": 9,
    "label": "NP_MOD",
    "dependent": [
      5,
      6
    ]
  },
  {
    "word_id": 8,
    "word_form": "3.3m²당",
    "head": 9,
    "label": "NP",
    "dependent": []
  },
  {
    "word_id": 9,

```

```

        "word_form": "분양가가",
        "head": 18,
        "label": "NP_SBJ",
        "dependent": [
            7,
            8
        ]
    },
    {
        "word_id": 10,
        "word_form": "2006년보다",
        "head": 14,
        "label": "NP_AJT",
        "dependent": []
    },
    {
        "word_id": 11,
        "word_form": "200만",
        "head": 12,
        "label": "NP",
        "dependent": []
    },
    {
        "word_id": 12,
        "word_form": "원",
        "head": 13,
        "label": "NP",
        "dependent": [
            11
        ]
    },
    {
        "word_id": 13,
        "word_form": "정도",
        "head": 14,
        "label": "NP_AJT",
        "dependent": [
            12
        ]
    },
    {
        "word_id": 14,
        "word_form": "싼",
        "head": 16,
        "label": "VP_MOD",
        "dependent": [
            10,
            13
        ]
    },
    {
        "word_id": 15,
        "word_form": "1500만",
        "head": 16,
        "label": "NP",
        "dependent": []
    }

```

```

    },
    {
      "word_id": 16,
      "word_form": "원",
      "head": 17,
      "label": "NP",
      "dependent": [
        14,
        15
      ]
    },
    {
      "word_id": 17,
      "word_form": "후반대로",
      "head": 18,
      "label": "NP_AJT",
      "dependent": [
        16
      ]
    },
    {
      "word_id": 18,
      "word_form": "결정될",
      "head": 19,
      "label": "VP_MOD",
      "dependent": [
        9,
        17
      ]
    },
    {
      "word_id": 19,
      "word_form": "것으로",
      "head": 20,
      "label": "NP_AJT",
      "dependent": [
        18
      ]
    },
    {
      "word_id": 20,
      "word_form": "보인다.",
      "head": -1,
      "label": "VP",
      "dependent": [
        19
      ]
    }
  ]
}

```

· 자료 내용 문의: 02-2669-9638