

## 국립국어원 형태 분석 말뭉치

(버전 1.1)

- **자료명:** 국립국어원 형태 분석 말뭉치
- **공개일**
  - (버전 1.0) 2020. 8. 25.
  - (버전 1.1) 2023. 10. 19.
    - 형태 오분석 수정
- **자료 유형:** 텍스트
- **관련 사업:** 형태 분석 말뭉치 구축(2019)
- **자료 설명**
  - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구조사 자료 > 연구 보고서 > ‘형태 분석 말뭉치 구축’ 사업 보고서 참고
- **내용**
  - 총 300만 어절 규모(문어 200만 어절, 구어 100만 어절)의 형태 분석 말뭉치
- **분량**
  - 총 300만 어절(문어 200만 어절, 구어 100만 어절)
- **파일 형식:** JSON(UTF-8 인코딩)
- **파일 수 및 크기:** 파일 2개, 총 2.31GB
- **인용:**
  - (국문) 국립국어원(2020). 국립국어원 형태 분석 말뭉치(버전 1.1). URL: <https://kli.korean.go.kr/corpus>
  - (영문) National Institute of Korean Language (2020). NIKL Part-of-speech Tagged Corpus (v.1.1). URL: <https://kli.korean.go.kr/corpus>

· 예시

```
{
  "id": "NXMP1902004240",
  "metadata": {
  },
  "document": [
    {
      "id": "NWRW1800000022.417",
      "metadata": {
      },
      "sentence": [
        {
          "id": "NWRW1800000022.417.1.1",
          "form": "[제주·서울] \"세계환경수도 조성위해 10개년 실천계획 만들겠다\" 김태환 지
사 밝혀",
          "word": [
            {
              "id": 1,
              "form": "[제주·서울]",
              "begin": 0,
              "end": 7
            },
            {
              "id": 2,
              "form": "\"세계환경수도\"",
              "begin": 8,
              "end": 15
            },
            {
              "id": 3,
              "form": "조성위해",
              "begin": 16,
              "end": 20
            },
            {
              "id": 4,
              "form": "10개년",
              "begin": 21,
              "end": 25
            },
            {
              "id": 5,
              "form": "실천계획",
              "begin": 26,
              "end": 30
            },
            {
              "id": 6,
              "form": "만들겠다\"",
              "begin": 31,
              "end": 36
            }
          ]
        }
      ]
    }
  ]
}
```

```
    },  
    {  
      "id": 7,  
      "form": "김태환",  
      "begin": 37,  
      "end": 40  
    },  
    {  
      "id": 8,  
      "form": "지사",  
      "begin": 41,  
      "end": 43  
    },  
    {  
      "id": 9,  
      "form": "밝혀",  
      "begin": 44,  
      "end": 46  
    }  
  ],  
  "morpheme": [  
    {  
      "id": 1,  
      "form": "[",  
      "label": "SS",  
      "word_id": 1,  
      "position": 1  
    },  
    {  
      "id": 2,  
      "form": "제주",  
      "label": "NNP",  
      "word_id": 1,  
      "position": 2  
    },  
    {  
      "id": 3,  
      "form": ".",  
      "label": "SP",  
      "word_id": 1,  
      "position": 3  
    },  
    {  
      "id": 4,  
      "form": "서울",  
      "label": "NNP",  
      "word_id": 1,  
      "position": 4  
    },  
    {  
      "id": 5,  
      "form": "]",  
      "label": "SS",  
      "word_id": 1,  
      "position": 5  
    }  
  ],  
}
```

```
{
  "id": 6,
  "form": "\",
  "label": "SS",
  "word_id": 2,
  "position": 1
},
{
  "id": 7,
  "form": "세계",
  "label": "NNG",
  "word_id": 2,
  "position": 2
},
{
  "id": 8,
  "form": "환경",
  "label": "NNG",
  "word_id": 2,
  "position": 3
},
{
  "id": 9,
  "form": "수도",
  "label": "NNG",
  "word_id": 2,
  "position": 4
},
{
  "id": 10,
  "form": "조성",
  "label": "NNG",
  "word_id": 3,
  "position": 1
},
{
  "id": 11,
  "form": "위하",
  "label": "VV",
  "word_id": 3,
  "position": 2
},
{
  "id": 12,
  "form": "아",
  "label": "EC",
  "word_id": 3,
  "position": 3
},
{
  "id": 13,
  "form": "10",
  "label": "SN",
  "word_id": 4,
  "position": 1
},
}
```

```
{
  "id": 14,
  "form": "개년",
  "label": "NNB",
  "word_id": 4,
  "position": 2
},
{
  "id": 15,
  "form": "실천",
  "label": "NNG",
  "word_id": 5,
  "position": 1
},
{
  "id": 16,
  "form": "계획",
  "label": "NNG",
  "word_id": 5,
  "position": 2
},
{
  "id": 17,
  "form": "만들",
  "label": "VV",
  "word_id": 6,
  "position": 1
},
{
  "id": 18,
  "form": "짚",
  "label": "EP",
  "word_id": 6,
  "position": 2
},
{
  "id": 19,
  "form": "다",
  "label": "EF",
  "word_id": 6,
  "position": 3
},
{
  "id": 20,
  "form": "\",",
  "label": "SS",
  "word_id": 6,
  "position": 4
},
{
  "id": 21,
  "form": "김태환",
  "label": "NNP",
  "word_id": 7,
  "position": 1
},
}
```

```

    {
      "id": 22,
      "form": "지사",
      "label": "NNG",
      "word_id": 8,
      "position": 1
    },
    {
      "id": 23,
      "form": "밝히",
      "label": "VV",
      "word_id": 9,
      "position": 1
    },
    {
      "id": 24,
      "form": "어",
      "label": "EF",
      "word_id": 9,
      "position": 2
    }
  ],
  "WSD": []
},

```

· 형태 분석 태그 정보

대분류	소분류	세분류	태그
체언	명사	일반명사	NNG
		고유명사	NNP
		의존명사	NNB
	대명사	대명사	NP
	수사	수사	NR
용언	동사	동사	VV
	형용사	형용사	VA
	보조용언	보조용언	VX
		지정사	긍정지정사 부정지정사
수식언	관형사	성상 관형사	MMA
		지시 관형사	MMD
		수 관형사	MMN
	부사	일반부사	MAG
		접속부사	MAJ
독립언	감탄사	감탄사	IC
관계언	격조사	주격조사	JKS
		보격조사	JKC
		관형격조사	JKG
		목적격조사	JKO
		부사격조사	JKB
		호격조사	JKV
		인용격조사	JKQ

	보조사	보조사	JX
	접속조사	접속조사	JC
의존형태	어미	선어말어미	EP
		종결어미	EF
		연결어미	EC
		명사형전성어미	ETN
		관형형전성어미	ETM
		접두사	체인접두사
	접미사	명사파생접미사	XSN
		동사파생접미사	XSV
		형용사파생접미사	XSA
	어근	어근	XR
기호	일반기호	마침표, 물음표, 느낌표	SF
		쉼표, 가운뎃점, 콜론, 빗금	SP
		따옴표, 괄호표, 줄표	SS
		줄임표	SE
		붙임표(물결)	SO
		기타 기호	SW
	외국어	외국어	SL
	한자	한자	SH
	숫자	숫자	SN
	분석불능범주	분석불능범주	NA
		명사추정범주	NF
		용언추정범주	NV

· 자료 내용 문의: 02-2669-9638