

국립국어원 어휘 의미 분석 말뭉치 2020 (버전 2.0)

- **자료명:** 국립국어원 어휘 의미 분석 말뭉치 2020
- **공개일**
 - (버전 1.0) 2021. 12. 1.
 - (버전 2.0) 2022. 4. 1.
 - 메신저 대화 말뭉치 분석 추가
- **자료 유형:** 텍스트
- **관련 사업:** 어휘 의미 분석 말뭉치 구축(2019),
2020년 어휘 의미 말뭉치 연구 분석 사업(2020)
- **자료 설명**
 - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구조사 자료 > 연구 보고서 > ‘어휘의미 분석 말뭉치 구축’, ‘2020년 어휘의미 말뭉치 연구 분석’ 사업 보고서 참고
- **내용**
 - 2019년도에 구축된 300만 어절 규모(문어 200만, 구어 100만 어절)와 메신저 대화 말뭉치 (92만 어절)를 대상으로 형태 분석과 어휘 의미(체언류와 용언류)를 분석한 말뭉치
- **분량**
 - 총 392만 어절(문어 200만 어절, 구어 100만 어절, 메신저 92만 어절)
- **파일 형식:** JSON(UTF-8 인코딩)
- **파일 수 및 크기:** 파일 3개, 총 3.43GB
- **인용:**
 - (국문) 국립국어원(2022). 국립국어원 어휘 의미 분석 말뭉치 2020(버전 2.0). URL: <https://kli.korean.go.kr/corpus>
 - (영문) National Institute of Korean Language (2022). NIKL Word Sense Tagged Corpus 2020 (v.2.0). URL: <https://kli.korean.go.kr/corpus>

· 예시

```
{
  "id": "NXLS2002104060",
  "metadata": {
    "title": "국립국어원 신문 말뭉치 추출 NXLS2002104060",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2020",
    "category": "신문 > 전국종합지",
    "annotation_level": [
      "어휘의미 분석"
    ],
    "sampling": "본문 전체"
  },
  "document": [
    {
      "id": "NWRW1800000022.417",
      "metadata": {
        "title": "조선일보 2009년 기사",
        "author": "최경운",
        "publisher": "조선일보사",
        "date": "20091130",
        "topic": "NA"
      },
      "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": "졌",
        "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": [
        {
            "word": "제주",
            "sense_id": 8,
            "pos": "NNP",
            "begin": 1,
            "end": 3,
            "word_id": 1
        },
        {
            "word": "서울",
            "sense_id": 2,
            "pos": "NNP",
            "begin": 4,
            "end": 6,
            "word_id": 1
        },
        {
            "word": "세계",
            "sense_id": 2,
            "pos": "NNG",
            "begin": 9,
            "end": 11,

```

```
    "word_id": 2
  },
  {
    "word": "환경",
    "sense_id": 1,
    "pos": "NNG",
    "begin": 11,
    "end": 13,
    "word_id": 2
  },
  {
    "word": "수도",
    "sense_id": 12,
    "pos": "NNG",
    "begin": 13,
    "end": 15,
    "word_id": 2
  },
  {
    "word": "조성",
    "sense_id": 5,
    "pos": "NNG",
    "begin": 16,
    "end": 18,
    "word_id": 3
  },
  {
    "word": "위하",
    "sense_id": 3,
    "pos": "VV",
    "begin": 18,
    "end": 20,
    "word_id": 3
  },
  {
    "word": "개년",
    "sense_id": 3,
    "pos": "NNB",
    "begin": 23,
    "end": 25,
    "word_id": 4
  },
  {
    "word": "실천",
    "sense_id": 1,
    "pos": "NNG",
    "begin": 26,
    "end": 28,
    "word_id": 5
  },
  {
    "word": "계획",
    "sense_id": 1,
    "pos": "NNG",
    "begin": 28,
    "end": 30,
```

```

        "word_id": 5
      },
      {
        "word": "만들",
        "sense_id": 1,
        "pos": "VV",
        "begin": 31,
        "end": 33,
        "word_id": 6
      },
      {
      {
        "word": "김태환",
        "sense_id": 777,
        "pos": "NNP",
        "begin": 37,
        "end": 40,
        "word_id": 7
      },
      {
      {
        "word": "지사",
        "sense_id": 10,
        "pos": "NNG",
        "begin": 41,
        "end": 43,
        "word_id": 8
      },
      {
      {
        "word": "밝히",
        "sense_id": 7,
        "pos": "VV",
        "begin": 44,
        "end": 46,
        "word_id": 9
      }
    }
  ]
},
{
  "id": "SXLS2002104060",
  "metadata": {
    "title": "국립국어원 구어 말뭉치 추출 SXLS2002104060",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2020",
    "category": "공적대화, 공적독백, 사적대화",
    "annotation_level": [
      "어휘의미 분석"
    ],
    "sampling": "본문 전체"
  },
  "document": [
    {
      "id": "SARW1800000001.1",
      "metadata": {
        "title": "EBS 정오뉴스 2018년 1월",

```



```

    "author": "박민영 외",
    "publisher": "EBS",
    "date": "20180000",
    "topic": "커피 위의 예술, 크리마트에 대한 기사"
  },
  "sentence": [
    {
      "id": "SARW1800000001.1.1.1",
      "form": "요즘처럼 추운 날씨에는",
      "word": [
        {
          "id": 1,
          "form": "요즘처럼",
          "begin": 0,
          "end": 4
        },
        {
          "id": 2,
          "form": "추운",
          "begin": 5,
          "end": 7
        },
        {
          "id": 3,
          "form": "날씨에는",
          "begin": 8,
          "end": 12
        }
      ]
    },
    {
      "morpheme": [
        {
          "id": 1,
          "form": "요즘",
          "label": "NNG",
          "word_id": 1,
          "position": 1
        },
        {
          "id": 2,
          "form": "처럼",
          "label": "JKB",
          "word_id": 1,
          "position": 2
        },
        {
          "id": 3,
          "form": "춡",
          "label": "VA",
          "word_id": 2,
          "position": 1
        },
        {
          "id": 4,
          "form": "ㄴ",
          "label": "ETM",
          "word_id": 2,

```

```

        "position": 2
      },
      {
        "id": 5,
        "form": "날씨",
        "label": "NNG",
        "word_id": 3,
        "position": 1
      },
      {
        "id": 6,
        "form": "에",
        "label": "JKB",
        "word_id": 3,
        "position": 2
      },
      {
        "id": 7,
        "form": "는",
        "label": "JX",
        "word_id": 3,
        "position": 3
      }
    ]
  },
  "WSD": [
    {
      "word": "요즘",
      "sense_id": 1,
      "pos": "NNG",
      "begin": 0,
      "end": 2,
      "word_id": 1
    },
    {
      "word": "춡",
      "sense_id": 1,
      "pos": "VA",
      "begin": 5,
      "end": 7,
      "word_id": 2
    },
    {
      "word": "날씨",
      "sense_id": 1,
      "pos": "NNG",
      "begin": 8,
      "end": 10,
      "word_id": 3
    }
  ]
},
{
  "id": "MXLS2002107210",
  "metadata": {

```

```

"title": "국립국어원 메신저 어휘 의미 분석 말뭉치 MXLS2002107210",
"creator": "국립국어원",
"distributor": "국립국어원",
"year": "2020",
"category": "메신저",
"annotation_level": [
    "형태분석",
    "어휘의미"
]
},
"document": [
    {
        "id": "MDRW1900000068.1",
        "metadata": {
            "title": "메신저 대화",
            "author": "개인 대화 참여자",
            "publisher": "카카오톡",
            "date": "20191219",
            "topic": "주거와 생활 (집안일, 육아, 부동산, 경제 활동, 생활 정보)",
            "speaker": [
                {
                    "id": "1",
                    "age": "30대",
                    "occupation": "사무 종사자",
                    "sex": "여성",
                    "birthplace": "전남",
                    "pricipal_residence": "전남",
                    "current_residence": "전남",
                    "device": "스마트폰",
                    "keyboard": "천지인"
                },
                {
                    "id": "2",
                    "age": "30대",
                    "occupation": "가정 주부",
                    "sex": "여성",
                    "birthplace": "경기",
                    "pricipal_residence": "경기",
                    "current_residence": "경북",
                    "device": "스마트폰",
                    "keyboard": "천지인"
                }
            ],
            "setting": {
                "relation": "직장 : 선후배/상사-부하",
                "intimacy": 5,
                "contact_frequency": "주 3회 이상"
            }
        },
        "sentence": [
            {
                "id": "MDRW1900000068.1.1.1",
                "form": "name2야~~",
                "word": [
                    {
                        "id": 1,

```

```

        "form": "name2야~~",
        "begin": 0,
        "end": 8
      }
    ],
    "morpheme": [
      {
        "id": 1,
        "form": "name2",
        "label": "NNP",
        "word_id": 1,
        "position": 1
      },
      {
        "id": 2,
        "form": "야",
        "label": "JKV",
        "word_id": 1,
        "position": 2
      },
      {
        "id": 3,
        "form": "~",
        "label": "SO",
        "word_id": 1,
        "position": 3
      },
      {
        "id": 4,
        "form": "~",
        "label": "SO",
        "word_id": 1,
        "position": 4
      }
    ],
    "WSD": [
      {
        "word": "name2",
        "sense_id": 777,
        "pos": "NNP",
        "begin": 0,
        "end": 5,
        "word_id": 1
      }
    ]
  },
  {
    "id": "MDRW1900000068.1.2.1",
    "form": "그저께부터 name3이 중기이유식 들어가서 하루에
두끼하는데",
    "word": [
      {
        "id": 1,
        "form": "그저께부터",
        "begin": 0,
        "end": 5
      }
    ]
  }

```

```

    },
    {
      "id": 2,
      "form": "name3이",
      "begin": 6,
      "end": 12
    },
    {
      "id": 3,
      "form": "중기이유식",
      "begin": 13,
      "end": 18
    },
    {
      "id": 4,
      "form": "들어가서",
      "begin": 19,
      "end": 23
    },
    {
      "id": 5,
      "form": "하루에",
      "begin": 24,
      "end": 27
    },
    {
      "id": 6,
      "form": "두끼하는데",
      "begin": 28,
      "end": 33
    }
  ],
  "morpheme": [
    {
      "id": 1,
      "form": "그저께",
      "label": "NNG",
      "word_id": 1,
      "position": 1
    },
    {
      "id": 2,
      "form": "부터",
      "label": "JX",
      "word_id": 1,
      "position": 2
    },
    {
      "id": 3,
      "form": "name3",
      "label": "NNP",
      "word_id": 2,
      "position": 1
    },
    {
      "id": 4,

```

	"form": "이", "label": "JKS", "word_id": 2, "position": 2
},	
{	"id": 5, "form": "중기", "label": "NNG", "word_id": 3, "position": 1
},	
{	"id": 6, "form": "이유식", "label": "NNG", "word_id": 3, "position": 2
},	
{	"id": 7, "form": "들어가", "label": "VV", "word_id": 4, "position": 1
},	
{	"id": 8, "form": "아셔", "label": "EC", "word_id": 4, "position": 2
},	
{	"id": 9, "form": "하루", "label": "NNG", "word_id": 5, "position": 1
},	
{	"id": 10, "form": "에", "label": "JKB", "word_id": 5, "position": 2
},	
{	"id": 11, "form": "두", "label": "MMN", "word_id": 6, "position": 1
},	
{	"id": 12,

```

        "form": "끼",
        "label": "NNG",
        "word_id": 6,
        "position": 2
    },
    {
        "id": 13,
        "form": "하",
        "label": "VV",
        "word_id": 6,
        "position": 3
    },
    {
        "id": 14,
        "form": "는데",
        "label": "EC",
        "word_id": 6,
        "position": 4
    }
],
"WSD": [
    {
        "word": "그저께",
        "sense_id": 1,
        "pos": "NNG",
        "begin": 0,
        "end": 3,
        "word_id": 1
    },
    {
        "word": "name3",
        "sense_id": 777,
        "pos": "NNP",
        "begin": 6,
        "end": 11,
        "word_id": 2
    },
    {
        "word": "중기",
        "sense_id": 7,
        "pos": "NNG",
        "begin": 13,
        "end": 15,
        "word_id": 3
    },
    {
        "word": "이유식",
        "sense_id": 1,
        "pos": "NNG",
        "begin": 15,
        "end": 18,
        "word_id": 3
    },
    {
        "word": "들어가",
        "sense_id": 3,

```

```

        "pos": "VV",
        "begin": 19,
        "end": 22,
        "word_id": 4
    },
    {
        "word": "하루",
        "sense_id": 1,
        "pos": "NNG",
        "begin": 24,
        "end": 26,
        "word_id": 5
    },
    {
        "word": "끼",
        "sense_id": 2,
        "pos": "NNG",
        "begin": 29,
        "end": 30,
        "word_id": 6
    },
    {
        "word": "하",
        "sense_id": 4,
        "pos": "VV",
        "begin": 30,
        "end": 31,
        "word_id": 6
    }
}
    ],
},

```

· 자료 내용 문의: 02-2669-9638