

국립국어원 상호 참조 해결 말뭉치 2019 (버전 1.0)

- **자료명:** 국립국어원 상호 참조 해결 말뭉치 2019
- **공개일**
 - (버전 1.0) 2021. 3. 30.
- **자료 유형:** 텍스트
- **관련 사업:** 상호 참조 해결 말뭉치 구축(2019)
- **자료 설명**
 - **내용**
 - 한 문서에서 같은 대상을 가리키는(상호 참조 또는 공지시 관계) 것들을 서로 연결한 말뭉치
 - **분량**
 - 총 300만 어절(문어 200만 어절, 구어 100만 어절)
 - **파일 형식:** JSON(UTF-8 인코딩)
 - **파일 수 및 크기:** 파일 2개, 총 883MB
- **인용:**

(국문) 국립국어원(2021). 국립국어원 상호 참조 해결 말뭉치 2020(버전 1.0). URL: <https://kli.korean.go.kr/corpus>

(영문) National Institute of Korean Language (2021). NIKL Korean Coreference Resolution Corpus 2020 (v.1.0). 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	매체	장르	주석 단계		구축 연도		일련번호(8자리)							
정의값	N: 신문 S: 구어	X: 추출	CR: 상호 참조 해결 말뭉치		19: 2019년		00000001 ~ 99999999 (여덟 자리 일련번호)							

※ 예시: NXCR1902103160.json 신문 전국 종합지 매체의 기사 원시 말뭉치를 추출하여 2019년도에 구축한 상호 참조 해결 말뭉치 파일

· 예시

```

    {
      "id": "NWRW1800000033.77.10.3",
      "form": "캐나다 회사인 노틸러스미네랄은 2000제공마일의 파푸아뉴기니 심
해를 20년간 임차하는 계약을 맺고 내년부터 본격적인 발굴에 들어간다.",
      "word": [
        {
          "id": 1,
          "form": "캐나다",
          "begin": 0,
          "end": 3
        },
        {
          "id": 2,
          "form": "회사인",
          "begin": 4,
          "end": 7
        },
        {
          "id": 3,
          "form": "노틸러스미네랄은",
          "begin": 8,
          "end": 16
        },
        {
          "id": 4,
          "form": "2000제공마일의",
          "begin": 17,
          "end": 26
        },
        {
          "id": 5,
          "form": "파푸아뉴기니",
          "begin": 27,
          "end": 33
        },
        {
          "id": 6,
          "form": "심해를",
          "begin": 34,
          "end": 37
        }
      ]
    }
  
```

```
{
  "id": 7,
  "form": "20년간",
  "begin": 38,
  "end": 42
},
{
  "id": 8,
  "form": "임차하는",
  "begin": 43,
  "end": 47
},
{
  "id": 9,
  "form": "계약을",
  "begin": 48,
  "end": 51
},
{
  "id": 10,
  "form": "맷고",
  "begin": 52,
  "end": 54
},
{
  "id": 11,
  "form": "내년부터",
  "begin": 55,
  "end": 59
},
{
  "id": 12,
  "form": "본격적인",
  "begin": 60,
  "end": 64
},
{
  "id": 13,
  "form": "발굴에",
  "begin": 65,
  "end": 68
},
{
  "id": 14,
  "form": "들어간다.",
  "begin": 69,
  "end": 74
}
]
},
```

```

    {
      "id": "NWRW1800000033.77.10.4",
      "form": "이 회사에 따르면 추정 매장량만 금 10t과 구리 12만500t에 이른
다.",
      "word": [
        {
          "id": 1,
          "form": "이",
          "begin": 0,
          "end": 1
        },
        {
          "id": 2,
          "form": "회사에",
          "begin": 2,
          "end": 5
        },
        {
          "id": 3,
          "form": "따르면",
          "begin": 6,
          "end": 9
        },
        {
          "id": 4,
          "form": "추정",
          "begin": 10,
          "end": 12
        },
        {
          "id": 5,
          "form": "매장량만",
          "begin": 13,
          "end": 17
        },
        {
          "id": 6,
          "form": "금",
          "begin": 18,
          "end": 19
        },
        {
          "id": 7,
          "form": "10t과",
          "begin": 20,
          "end": 24
        },
        {
          "id": 8,
          "form": "구리",

```

```

        "begin": 25,
        "end": 27
      },
      {
        "id": 9,
        "form": "12만500t에",
        "begin": 28,
        "end": 36
      },
      {
        "id": 10,
        "form": "이른다.",
        "begin": 37,
        "end": 41
      }
    ]
  },
  "CR": [
    {
      "mention": [
        {
          "sentence_id": "NWRW1800000033.77.10.3",
          "form": "캐나다 회사",
          "begin": 0,
          "end": 6,
          "NE_id": -1
        },
        {
          "sentence_id": "NWRW1800000033.77.10.3",
          "form": "캐나다 회사인 노틸러스미네랄",
          "begin": 0,
          "end": 15,
          "NE_id": -1
        },
        {
          "sentence_id": "NWRW1800000033.77.10.4",
          "form": "이 회사",
          "begin": 0,
          "end": 4,
          "NE_id": -1
        }
      ]
    }
  ]
}
]

```

· 자료 내용 문의: 02-2669-9638