

국립국어원 국회 회의록 말뭉치 2021

(버전 1.1)

- **자료명:** 국립국어원 국회 회의록 말뭉치 2021
- **공개일**
 - (버전 1.0) 2022. 3. 31.
 - (버전 1.1) 2023. 7. 19.
 - JSON 형식 수정
- **자료 유형:** 텍스트
- **관련 사업:** 없음.
- **자료 설명**
 - **내용**
 - 2003년~2020년 국회 소위원회 회의록을 대상으로 구축한 말뭉치
 - 발화 내 한자로 기록된 용어나 이름을 한글로 변환함.
 - 행동이나 상황 묘사와 원어 삽입구를 정제함.
 - 자세한 내용은 국립국어원 2021년 회의록 요약 말뭉치 연구 분석 사업의 결과 보고서 중 원시 말뭉치 부분을 참고할 수 있음.
- **분량**
 - 5,395건(73,478,080어절)
- **파일 형식:** JSON(UTF-8 인코딩)
- **파일 수 및 크기:** 파일 5,395개, 총 307MB(ZIP 파일 기준)
- **인용:**
 - (국문) 국립국어원(2022). 국립국어원 국회 회의록 말뭉치 2021(버전 1.1). URL: <https://kli.korean.go.kr/corpus>
 - (영문) National Institute of Korean Language (2022). NIKL Korean Parliamentary Corpus 2021 (v.1.1). 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	S: 구어	B: 공적 대화 (국회 소위원회 회의)		RW: 원시 말뭉치		21: 2021년	00000001 ~ 99999999 (여덟 자리 일련번호)							

※ 예시: SBRW2100000001.json 2021년에 구축한 공적 대화(국회 소위원회 회의) 말뭉치 파일

· 예시

```
{
  "id": "SBRW2100000001",
  "metadata": {
    "title": "국립국어원 국회 회의록 원시 말뭉치 SBRW2100000001",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2021",
    "category": "회의록 > 국회소위원회 > 女性委員會會議錄",
    "annotation_level": [
      "원시"
    ],
    "sampling": "본문 전체"
  },
  "document": [
    {
      "id": "SBRW2100000001.1",
      "metadata": {
        "title": "第284回國會(定期會)女性委員會會議錄(豫算決算基金審查小委員會)第1號",
        "author": "國會事務處",
        "publisher": "國會事務處",
        "date": "20090925",
        "topic": "1. 2008회계연도 세입세출결산 > 가. 여성부 소관 / 2. 2008회계연도  
기금결산 > 가. 여성발전기금",
        "speaker": [
          {
            "id": "김상희",
            "occupation": "소위원장"
          },
          {
            "id": "김귀순",
            "occupation": "수석전문위원"
          },
          {
            "id": "金玉伊",
            "occupation": "위원"
          }
        ]
      }
    }
  ]
}
```

```

    {
      "id": "김태석",
      "occupation": "여성부기획조정실장"
    },
    {
      "id": "이정선",
      "occupation": "위원"
    },
    {
      "id": "곽정숙",
      "occupation": "위원"
    },
    {
      "id": "손숙미",
      "occupation": "위원"
    },
    {
      "id": "이복실",
      "occupation": "여성부권익증진국장"
    },
    {
      "id": "정봉협",
      "occupation": "여성부여성정책국장"
    },
    {
      "id": "김춘진",
      "occupation": "위원"
    },
    {
      "id": "이성선",
      "occupation": "여성부인력개발사업과장"
    },
    {
      "id": "김권영",
      "occupation": "여성부복지지원과장"
    },
    {
      "id": "차인순",
      "occupation": "입법심의회"
    }
  ],
  "utterance": [
    {
      "id": "SBRW210000001.1.1.1",
      "form": "회의를 시작하도록 하겠습니다.",
      "original_form": "회의를 시작하도록 하겠습니다.",
      "speaker_id": "김상희",
      "note": "(09시40분 개의)"
    }
  ],

```

```

    {
      "id": "SBRW2100000001.1.1.2",
      "form": "제284회 국회(정기회) 여성위원회 제1차 예산결산기금심사소위원회
를 개의하겠습니다.",
      "original_form": "제284회 국회(정기회) 여성위원회 제1차 예산결산기금심
사소위원회를 개의하겠습니다.",
      "speaker_id": "김상희",
      "note": ""
    },
    {
      "id": "SBRW2100000001.1.1.3",
      "form": "바쁘신 중에도 이렇게 회의에 참석해 주신 여러분들께 감사를 드
립니다.",
      "original_form": "바쁘신 중에도 이렇게 회의에 참석해 주신 여러분들께 감
사를 드립니다.",
      "speaker_id": "김상희",
      "note": ""
    },
    {
      "id": "SBRW2100000001.1.1.4",
      "form": "그러면 바로 의사일정에 들어가도록 하겠습니다.",
      "original_form": "그러면 바로 의사일정에 들어가도록 하겠습니다.",
      "speaker_id": "김상희",
      "note": ""
    },
    {
      "id": "SBRW2100000001.1.1.5",
      "form": "의사일정 제1항 2008회계연도 여성부 소관 세입세출결산, 의사일
정 제2항 2008회계연도 여성발전기금 결산, 이상 2건을 일괄하여 상정합니다.",
      "original_form": "의사일정 제1항 2008회계연도 여성부 소관 세입세출결
산, 의사일정 제2항 2008회계연도 여성발전기금 결산, 이상 2건을 일괄하여 상정합니다.",
      "speaker_id": "김상희",
      "note": ""
    },
    {
      "id": "SBRW2100000001.1.1.6",
      "form": "오늘 소위원회 회의 진행은 지난번 전체회의 대체토론 시에 위원
님들께서 지적하신 사항과 검토보고를 토대로 수석전문위원이 정리한 결산시정요구사항을 중심으
로 해서 수석전문위원으로부터 설명을 듣고 위원님들의 의견과 정부 측 답변을 들은 다음 사안별
로 정리하는 방법으로 회의를 진행하고자 합니다.",
      "original_form": "오늘 소위원회 회의 진행은 지난번 전체회의 대체토론 시
에 위원님들께서 지적하신 사항과 검토보고를 토대로 수석전문위원이 정리한 결산시정요구사항을
중심으로 해서 수석전문위원으로부터 설명을 듣고 위원님들의 의견과 정부 측 답변을 들은 다음
사안별로 정리하는 방법으로 회의를 진행하고자 합니다.",
      "speaker_id": "김상희",
      "note": ""
    },
    {
      "id": "SBRW2100000001.1.1.7",
      "form": "그러면 먼저 수석전문위원, 설명해 주시기 바랍니다.",

```

```

        "original_form": "그러면 먼저 수석전문위원, 설명해 주시기 바랍니다.",
        "speaker_id": "김상희",
        "note": ""
    },
    {
        "id": "SBRW2100000001.1.1.8",
        "form": "시간절약 차원에서 결산시정요구사항을 중심으로 논의를 전개하도
        록 하겠습니다.",
        "original_form": "시간절약 차원에서 결산시정요구사항을 중심으로 논의를
        전개하도록 하겠습니다.",
        "speaker_id": "김귀순",
        "note": ""
    },
    {
        "id": "SBRW2100000001.1.1.9",
        "form": "먼저 결산 등 심사기준은 유인물에 나와 있는 바와 같습니다. 지난
        번과 다른 점은 시정이 변상과 징계 다음으로 올라갔는데 올해부터 예결위에서 이 기준을 정비했
        다고 합니다. 그래서 위에서 밑으로 갈수록 강도가 약하다고 합니다. 참고해 주시면 감사하겠습니
        다.",
        "original_form": "먼저 결산 등 심사기준은 유인물에 나와 있는 바와 같습
        니다. 지난번과 다른 점은 시정이 변상과 징계 다음으로 올라갔는데 올해부터 예결위에서 이 기준
        을 정비했다고 합니다. 그래서 위에서 밑으로 갈수록 강도가 약하다고 합니다. 참고해 주시면 감
        사하겠습니다.",
        "speaker_id": "김귀순",
        "note": ""
    },
    {
        "id": "SBRW2100000001.1.1.10",
        "form": "그다음에 타 위원회의 시정요구 현황을 보면 보건복지가족부 15건,
        지식경제부 23건, 국토해양부 66건으로 되어 있습니다.",
        "original_form": "그다음에 타 위원회의 시정요구 현황을 보면 보건복지가
        족부 15건, 지식경제부 23건, 국토해양부 66건으로 되어 있습니다.",
        "speaker_id": "김귀순",
        "note": ""
    },
    {
        "id": "SBRW2100000001.1.1.11",
        "form": "그리고 예결위 협조요청 사항을 말씀드리겠습니다.",
        "original_form": "그리고 예결위 협조요청 사항을 말씀드리겠습니다.",
        "speaker_id": "김귀순",
        "note": ""
    },
    {
        "id": "SBRW2100000001.1.1.544",
        "form": "산회를 선포합니다.",
        "original_form": "산회를 선포합니다.",
        "speaker_id": "김상희",
        "note": "(11시44분 산회)"
    }
}
]

```

```
    }  
  ]  
}
```

- ※ "topic": 상정된 안건
- "speaker": 발화자 포함 회의 참석자
- "utterance": 발화
 - "form": 정제 후 발화(비발화 삽입구 삭제, 한자-한글 변환 등)(예. SBRW2100000201.1.1.2)
 - "original_form": 정제 전 발화(원문 그대로의 형태)
- "note": 행동 및 상황 묘사, 불특정인의 발언, 회의 진행 사항 등 기타 내용(예. SBRW2100000201.1.1.20)

· 자료 내용 문의: 02-2669-9638