

국립국어원 한국어-한국수어 병렬 말뭉치 2022 (버전 1.0)

- **자료명:** 국립국어원 한국어-한국수어 병렬 말뭉치 2022
- **공개일:**
 - (버전 1.0) 2024. 3. 29.
- **자료 유형:** 텍스트, 한국수어 영상
- **관련 사업:** 2022년 한국어-한국수어 병렬 말뭉치 구축
- **자료 설명**
 - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구·조사 자료 > ‘2022년 한국어-한국수어 병렬 말뭉치 구축’ 사업 보고서 참고
 - **내용**
 - 한국어 구어 데이터(생활 분야, 문화 분야)를 한국수어로 번역하여 구성한 병렬 말뭉치
 - 한국수어로 번역한 영상에 전사 정보 포함(수지 주식, 비수지 주식, 키포인트)
 - **분량 및 구성**
 - 한국어-한국수어 데이터 총 1,001,087어절(원천 자료 기준), 수어 120,295문장
 - 원천 자료 구성

구분		한국어 어절 수량	한국수어 문장 수량
구어체	생활 분야	의료	273,663
		민원 행정	303,713
	문화 분야	쇼핑	203,370
		관광	220,341
합계		1,001,087	120,295

- **파일 형식:** JSON(UTF-8 인코딩), 영상(MP4)
- **파일 수 및 크기:** 파일 120,295쌍(JSON과 영상), 총 585GB

· 파일 명명 규칙

자리	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
속성	유형	분야		장르		병렬		원자료		결과자료			구축연도		일련번호(8자리)							
정의 값	S: 구어	LI: 생활	ME: 의료		PA: 병렬	KO: 한국어	KSL: 한국수어	22	00000001 ~ 99999999 (8자리 일련번호)													
			CC: 민원 행정																			
		CU: 문화	SH: 쇼핑																			
			TO: 관광																			
※ 예시 : SLICCPAKOKSL2200019426.json 2022년에 구축한 한국어-한국수어 병렬 말뭉치 파일 (분야: 생활, 장르: 민원 행정) SCUTOPAKOKSL2200009741.json 2022년에 구축한 한국어-한국수어 병렬 말뭉치 파일 (분야: 문화, 장르: 관광)																						

· 인용:

- (국문) 국립국어원(2023). 국립국어원 한국어-한국수어 병렬 말뭉치 2022(버전 1.0). URL: <https://kli.korean.go.kr/corpus>
- (영문) National Institute of Korean Language(2023). Korean-Korean Sign Language Parallel Corpus 2022(v.1.0). URL: <https://kli.korean.go.kr/corpus>

· 예시

```
{
  "id":"SLICCPAKOKSL2200040323",
  "opertor":"cubeworker15",
  "krlgg_sntenc":{
    "koreanText":"신고서를 작성하여 접수하면 처리기관의 검토와 결재를 거치게 됩니다",
    "realm":"생활",
    "thema":"민원/행정"
  },
  "sign_lang_sntenc":"신고/서류/작성/후/접수[하다]/하면(어미)/처리/기관(기구)/검토/또(and)/결재/진행",
  "sign_lang_trnslator":"DB",
  "sign_script":{
    "sign_gestures_strong":[
      {
        "start":9.806,
        "end":10.296,
        "gloss_id":"진행",
        "express":"s",
        "direction":{
          "source":"","
          "target":""
        },
        "sentence_loc":{
          "start":"","
          "end":""
        }
      },
      {
        "start":8.947,
        "end":9.543,
        "gloss_id":"결재",
        "express":"s",
        "direction":{
          "source":"","
          "target":""
        },
        "sentence_loc":{
          "start":"","
          "end":""
        }
      },
      {
        "start":8.328,
        "end":8.53,
```

```

"gloss_id":"ㅏ(and)",
"express":"s",
"direction":{
  "source":"","
  "target":""
},
"sentence_loc":{
  "start":"","
  "end":""
}
},
{
  "start":7.68,
  "end":8.103,
  "gloss_id":"검토",
  "express":"s",
  "direction":{
    "source":"","
    "target":""
  },
  "sentence_loc":{
    "start":"","
    "end":""
  }
},
{
  "start":6.36,
  "end":7.368,
  "gloss_id":"기관(기구)",
  "express":"s",
  "direction":{
    "source":"","
    "target":""
  },
  "sentence_loc":{
    "start":"","
    "end":""
  }
},
{
  "start":5.664,
  "end":6.077,
  "gloss_id":"처리",
  "express":"s",

```

```

"direction":{
  "source": "",
  "target": ""
},
"sentence_loc":{
  "start": "",
  "end": ""
}
},
{
  "start": 5.035,
  "end": 5.401,
  "gloss_id": "하면(어미)",
  "express": "s",
  "direction":{
    "source": "",
    "target": ""
  },
  "sentence_loc":{
    "start": "",
    "end": ""
  }
},
{
  "start": 4.099,
  "end": 4.772,
  "gloss_id": "접수[하다]",
  "express": "s",
  "direction":{
    "source": "2",
    "target": "1"
  },
  "sentence_loc":{
    "start": "",
    "end": ""
  }
},
{
  "start": 3.49,
  "end": 3.884,
  "gloss_id": "후",
  "express": "s",
  "direction":{
    "source": "",

```

```
"target":""
},
"sentence_loc":{
  "start": "",
  "end": ""
}
},
{
  "start": 2.957,
  "end": 3.284,
  "gloss_id": "작성",
  "express": "s",
  "direction": {
    "source": "",
    "target": ""
  },
},
"sentence_loc": {
  "start": "",
  "end": ""
}
},
{
  "start": 2.232,
  "end": 2.684,
  "gloss_id": "서류",
  "express": "s",
  "direction": {
    "source": "",
    "target": ""
  },
},
"sentence_loc": {
  "start": "",
  "end": ""
}
},
{
  "start": 1.392,
  "end": 1.835,
  "gloss_id": "신고",
  "express": "s",
  "direction": {
    "source": "2",
    "target": "1"
  },
},
```

```
"sentence_loc":{
  "start": "",
  "end": ""
}
},
"sign_gestures_weak": [
  {
    "start": 1.392,
    "end": 1.835,
    "gloss_id": "신고",
    "express": "s",
    "direction": null,
    "sentence_loc": null
  },
  {
    "start": 2.232,
    "end": 2.684,
    "gloss_id": "서류",
    "express": "s",
    "direction": null,
    "sentence_loc": null
  },
  {
    "start": 2.957,
    "end": 3.284,
    "gloss_id": "작성",
    "express": "s",
    "direction": null,
    "sentence_loc": null
  },
  {
    "start": 3.49,
    "end": 3.884,
    "gloss_id": "후",
    "express": "s",
    "direction": null,
    "sentence_loc": null
  },
  {
    "start": 4.099,
    "end": 4.772,
    "gloss_id": "접수[하다]",
    "express": "s",
```

```
"direction":null,
"sentence_loc":null
},
{
  "start":5.035,
  "end":5.401,
  "gloss_id":"하면(어미)",
  "express":"s",
  "direction":null,
  "sentence_loc":null
},
{
  "start":5.664,
  "end":6.077,
  "gloss_id":"처리",
  "express":"s",
  "direction":null,
  "sentence_loc":null
},
{
  "start":6.36,
  "end":7.368,
  "gloss_id":"기관(기구)",
  "express":"s",
  "direction":null,
  "sentence_loc":null
},
{
  "start":8.947,
  "end":9.543,
  "gloss_id":"결재",
  "express":"s",
  "direction":null,
  "sentence_loc":null
},
{
  "start":9.806,
  "end":10.296,
  "gloss_id":"진행",
  "express":"s",
  "direction":null,
  "sentence_loc":null
}
]
```



```
},  
"nms_script":{  
  "Mmo":null,  
  "Hno":[  
    {  
      "descriptor": "",  
      "start": 3.49,  
      "end": 3.99  
    },  
    {  
      "descriptor": "",  
      "start": 9.806,  
      "end": 10.306  
    }  
  ],  
  "Mo1":null,  
  "Hs":null,  
  "EBf":[  
    {  
      "descriptor": "",  
      "start": 5.664,  
      "end": 6.164  
    },  
    {  
      "descriptor": "",  
      "start": 7.68,  
      "end": 8.18  
    }  
  ],  
  "Mctr":[  
    {  
      "descriptor": "",  
      "start": 9.806,  
      "end": 10.306  
    }  
  ],  
  "Ci":null,  
  "Tbt":[  
    {  
      "descriptor": "",  
      "start": 1.392,  
      "end": 2.717  
    },  
    {
```

```

        "descriptor": "",
        "start": 3.49,
        "end": 5.448
    },
    {
        "descriptor": "",
        "start": 8.327,
        "end": 9.499
    }
]
}.
"vido_file_nm": "SLICCPAKOKSL2200040323",
"potogrf": {
    "createdTime": "2023-01-18 00:00:00",
    "sentence_ID": "SLICCPAKOKSL2200040323",
    "photographer": "A",
    "location": "a",
    "devide": "FDR_AX700",
    "iris": 2.8,
    "Gain": 6,
    "WhiteBalance": 6500,
    "ShutterSpeed": 2000,
    "fps": 30,
    "width": 1920,
    "height": 1080,
    "fileSize": 6.75,
    "length": "00:00:11:15",
    "format": "mp4",
    "codec": "avc1",
    "sl_speaker_id": "J",
    "sl_speaker_age": 40,
    "sl_speaker_sex": "Female",
    "sl_speaker_legion": "Seoul",
    "sl_speaker_hand": "right_handed"
},
"landmarks" : [ {
    "frame" : 1,
    "predictions" : [ {
        "keypoints" : [ 962.13, 222.58, 0.87, 1000.0, 190.04, 0.87, 929.95, 190.35, 0.87, 1051.44,
218.42, 0.79, 891.93, 216.26, 0.68, 1112.08, 415.99, 0.77, 826.29, 418.2, 0.78, 1156.39, 667.78, 0.81,
772.38, 650.95, 0.81, 1158.38, 880.53, 0.82, 748.77, 865.54, 0.83, 1060.75, 899.58, 0.7, 861.32,
892.52, 0.69, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0,
0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 896.15, 193.88,
-(중략)- ],

```

```
"frame" : 345,  
"predictions" : [ {  
  "keypoints" : [ 965.07, 226.13, 0.86, 1002.7, 191.68, 0.87, 932.8, 191.28, 0.87, 1052.38, 217.2,  
0.78, 893.91, 215.37, 0.71, 1116.63, 416.95, 0.77, 826.65, 418.29, 0.78, 1161.04, 668.54, 0.81, 772.2,  
651.4, 0.81, 1160.83, 882.43, 0.8, 752.59, 866.66, 0.82, 1064.06, 916.71, 0.67, 867.55, 903.98, 0.68,  
0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17,  
0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 0.0, -36.17, 0.0, 898.18, 193.27, 0.88, -(중  
략)-],  
  "hand_pos" : [ 6, 4 ]  
} ]  
} ]  
}
```

· 자료 내용 문의: 02-2669-9613