

국립국어원 한국어-외국어 병렬 말뭉치 2021 (버전 1.0)

- **자료명:** 국립국어원 한국어-외국어 병렬 말뭉치 2021
- **공개일**
 - (버전 1.0) 2023. 3. 31.
- **자료 유형:** 텍스트
- **관련 사업:** 2021년 한국어-외국어 병렬 말뭉치 구축(2021)
- **자료 설명**
 - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구·조사 자료 > ‘2021년 한국어-외국어 병렬 말뭉치 구축 사업’ 보고서 참고
 - **내용**
 - 한국어 문어 데이터(신문 기사, 시나리오 지시문)와 구어 데이터(자체 구축 일상 대화, 시나리오 대화문)를 8개 외국어(베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어)로 번역하여 구성한 병렬 말뭉치
 - **분량 및 구성**
 - 한국어-외국어 병렬 말뭉치 언어별 1,060,122어절(원천 자료 기준)
 - 원천 자료 구성

구분		수량
문어체	신문 기사	390,757어절
	시나리오 지시문	33,601어절
구어체	일상 대화	620,479어절
	시나리오 대화문	15,285어절

- **파일 형식:** JSON(UTF-8 인코딩)
- **파일 수 및 크기:** 파일 36개, 총 100MB

· 파일 명명 규칙

자리	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
속성	유형	매체 및 장르	주석 단계	병렬	원자료(src)	결과 자료(trg)	구축 연도	일련번호(8자리)												
정의 값	N: 신문 말뭉치	I: 신문 기사	OR: 원문 자료	PA: 병렬	KO: 한국어	VI: 베트남어 ID: 인도네시아어 TH: 태국어 HI: 인도 힌디어 KM: 캄보디아 크메르어 TL: 필리핀 타갈로그어 RU: 러시아어 UZ: 우즈베크어	21	00000001 ~ 99999999 (8자리 일련번호)												
	W: 문어 말뭉치	Z: 시나리오 지시문																		
	S: 구어 말뭉치	E: 시나리오 대화문 D: 일상 대화																		
※ 예시: NIORPAKOVI2102205260.json 2021년에 구축한 한국어-베트남어 병렬 말뭉치 파일 (장르: 신문기사) WZORPAKOTH2102205260.json 2021년에 구축한 한국어-태국어 병렬 말뭉치 파일 (장르: 시나리오 지시문) SDORPAKOH2102205260.json 2021년에 구축한 한국어-인도 힌디어 병렬 말뭉치 파일 (장르: 일상 대화) SEORPAKOID2102205260.json 2021년에 구축한 한국어-인도네시아어 병렬 말뭉치 파일 (장르: 시나리오 대화문)																				

· 인용:

- (국문) 국립국어원(2022). 국립국어원 한국어-외국어 병렬 말뭉치 2021(버전 1.0).
URL: <https://kli.korean.go.kr/corpus>
- (영문) National Institute of Korean Language(2022), Korean-Foreign Language Parallel Corpus 2021(v.1.0).
URL: <https://kli.korean.go.kr/corpus>

· 예시

```
{
  "language_info": {
    "source_language": "ko",
    "target_language": "vi"
  },
  "parallel": [
    {
      "id": "NIOR2102303070.1.1",
      "source": "이로 인해 경유 없이 비행시간 12시간 50분이면 일본에서 이탈리아 밀라노에 도착할 수 있다.",
      "target": "Do đó, có thể bay từ Nhật Bản đến Milan, Ý, mà không cần quá cảnh, với thời gian bay là 12 tiếng 5 phút.",
      "revision": {
        "revision1": "Do đó, có thể bay từ Nhật Bản đến Milan, Ý, mà không cần quá cảnh, với thời gian bay là 12 tiếng 5 phút.",
        "revision2": "Do đó, có thể bay từ Nhật Bản đến Milan, Ý, mà không cần quá cảnh, với thời gian bay là 12 tiếng 5 phút."
      }
    },
    {
      "id": "NIOR2102303070.1.2",
      "source": "일본 학생들 사이의 인기 여행지 대부분이 유럽 지역인 것으로 조사됐다.",
      "target": "Kết quả điều tra cho thấy hầu hết điểm đến du lịch yêu thích của sinh viên Nhật Bản là các địa điểm tại châu Âu.",
      "revision": {
        "revision1": "Kết quả điều tra cho thấy hầu hết điểm đến du lịch yêu thích của sinh viên Nhật Bản là các địa điểm tại châu Âu.",
        "revision2": "Kết quả điều tra cho thấy hầu hết điểm đến du lịch yêu thích của sinh viên Nhật Bản là các địa điểm tại châu Âu."
      }
    }
  ],
}
```

· 자료 내용 문의: 02-2669-9754