

국립국어원 한국어-외국어 병렬 말뭉치 2022 (버전 1.0)

- **자료명:** 국립국어원 한국어-외국어 병렬 말뭉치 2022
- **공개일**
 - (버전 1.0) 2023. 12. 29.
- **자료 유형:** 텍스트
- **관련 사업:** 2022년 한국어-외국어 병렬 말뭉치 구축(2022)
- **자료 설명**
 - ※ 자세한 내용은 국립국어원 누리집 > 자료 > 연구·조사 자료 > ‘2022년 한국어-외국어 병렬 말뭉치 구축 사업’ 보고서 참고
 - **내용**
 - 한국어 문어 데이터(신문 기사)와 구어 데이터(자체 구축 일상 대화, 유튜브 대본)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 1,301,074어절(원천 자료 기준)
 - 원천 자료 구성

구분		수량	
문어체	신문 기사	520,873어절	
구어체	일상 대화	자체 구축	540,033어절
		유튜브 대본	240,168어절

- **파일 형식:** JSON(UTF-8 인코딩)
- **파일 수 및 크기:** 파일 18개, 총 825MB

· 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	22	00000001 ~ 99999999 (8자리 일련번호)												
	S: 구어 말뭉치	D: 일상 대화																		
※ 예시: NIORPAKOVI2202312290.json 2022년에 구축한 한국어-베트남어 병렬 말뭉치 파일 (장르: 신문기사) SDORPAKOH12202312290.json 2022년에 구축한 한국어-인도 힌디어 병렬 말뭉치 파일 (장르: 일상 대화)																				

· 인용:

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

· 예시

```
{
  "language_info": {
    "source_language": "ko",
    "target_language": "vi"
  },
  "parallel": [
    {
      "id": "SDOR2202312290.1.1",
      "source": "밥, 빵, 면 중에 뭐 먹을지부터 고르자.",
      "target": "Chọn sẽ ăn gì trước giữa cơm, bánh mì và mì nhé.",
      "revision": {
        "revision1": "Cậu chọn giữa cơm, bánh mì và mì xem ăn cái gì trước đi.",
        "revision2": "Chọn sẽ ăn gì trước giữa cơm, bánh mì và mì nhé."
      }
    },
    {
      "id": "SDOR2202312290.1.5",
      "source": "아, 찌개 얘기 하니까 김치찌개 먹고 싶은데?",
      "target": "À, vì nhắc đến Jjigae nên em muốn ăn canh kimchi đấy ạ?",
      "revision": {
        "revision1": "Ôi, tự đừng nhắc đến canh làm em thèm ăn canh kimchi thế nhỉ?",
        "revision2": "À, vì nhắc đến Jjigae nên em muốn ăn canh kimchi đấy ạ?"
      }
    }
  ]
}
```

· 자료 내용 문의: 02-2669-9754