

국립국어원 2025-01-30

발간등록번호
11-1371028-100027-01

# 2025년 한국어-외국어 병렬 말뭉치 구축

연구 책임자  
이 정 희



## 제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 '2025년 한국어-외국어 병렬 말뭉치 구축'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2025년 3월 ~ 2025년 12월

2025년 12월 15일

연구 책임자: 이정희(경희대학교)

연구 기관: 경희대학교 산학협력단, (주)플리토

연구 책임자: 이정희

공동 연구원: 이정수, 김일환, 김종민, 박진욱, 이수미, 이영준,  
임채훈, 강동한, 김진구, 유성곤, 김재훈, 김세화,  
김연희, 김영근, 문진숙, 서유리, 윤세윤, 전지연,  
정성호, 지화숙, 한재민, 이지희, 문다예, 김유연,  
이지수, 송문식, 문지연

보조 연구원: 국혜민, 김인후, 민유미, 송주현, 심현보, 이두일,  
이태현, 이혜민, 최예린, 함인숙



<연구 수행자> 2025년 한국어-외국어 병렬 말뭉치 구축 사업단

연구 책임자	이정희(경희대학교)	
연구 참여자	김일환(성신여자대학교)	김종민(경희대학교)
	박진옥(대구가톨릭대학교)	이수미(성균관대학교)
	이영준(한국학중앙연구원)	임채훈(송실대학교)
	김세화(경희대학교)	김연희(경희대학교)
	김영근(경희대학교)	문진숙(경희대학교)
	서유리(서경대학교)	윤세운(경희대학교)
	전지연(강원대학교)	정성호(고려대학교)
	지화숙(경희사이버대학교)	한재민(고려대학교)
	국혜민(송실대학교)	김인후(서강대학교)
	민유미(경희대학교)	송주현(경희대학교)
	심현보(서경대학교)	이두일(경희대학교)
	이태현(경희대학교)	이혜민(서울시립대학교)
	최예린(서정대학교)	함인숙(건국대학교)
	이정수(㈜플리토)	강동한(㈜플리토)
	김진구(㈜플리토)	유성곤(㈜플리토)
	김재훈(㈜플리토)	이지희(㈜플리토)
	문다예(㈜플리토)	김유연(㈜플리토)
	이지수(㈜플리토)	송문식(㈜플리토)
	문지연(㈜플리토)	



## <국문 초록>

# 2025년 한국어-외국어 병렬 말뭉치 구축

이 사업은 ‘2021년 한국어-외국어 병렬 말뭉치 구축’과 ‘2022년 한국어-외국어 병렬 말뭉치 구축’, ‘2023년 한국어-외국어 병렬 말뭉치 구축’, ‘2024년 한국어-외국어 병렬 말뭉치 구축’ 사업에 이어 산업계에서 자동 번역 등 한국어 기반 인공지능 통번역 프로그램의 개발 수요가 높은 외국어를 대상으로 한국어-외국어 병렬 말뭉치를 구축하고, 2026년 이후 사업 추진의 기반이 될 제2차 중장기계획을 수립하는 것이 목표이다. 병렬 말뭉치 구축 대상은 베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어이다.

병렬 말뭉치 구축은 크게 ‘수집’, ‘구축’, ‘산출’의 세 단계로 진행되었다. ‘수집’ 단계에서는 한국어 문어·구어 원시 데이터를 수집하고, 수집한 데이터를 검수·정제하였다. 검수·정제가 완료된 데이터를 대상으로 ‘구축’ 단계에서 번역이 이루어졌으며, 번역문은 교정 작업자의 교정과 (주)플리토 내부 언어 전문가의 1차 검수(7% 표본 검수)를 거쳤다. 이후 번역 검수원이 번역문을 2차 전수 검수하였으며, 2차 검수가 완료된 문장은 검수팀장이 3차 검수(20% 표본 검수)를, 감수자가 감수(10% 표본 검수)를 진행하였다. 이상의 단계를 거쳐 품질이 검증된 번역문은 ‘산출’ 단계에서 메타 정보와 함께 최종 데이터로 산출되었다.

최종적으로 한국어-외국어 병렬 말뭉치 총 8,000,000어절(한국어 원문 기준)과 추가 제안 사항이었던 한국어-영어 병렬 말뭉치 1,000,000어절(한국어 원문 기준)을 구축하였다.

다음으로 한국어-외국어 말뭉치 구축 제2차 중장기계획을 수립하였다. 이를 위해 제1차 중장기계획의 성과 및 제반 환경, 산업계·학계의 요구 사항을 분석하였다. 그리고 분석 결과를 바탕으로 비전 및 전략 체계를 설정하고 전략 과제별로 구체적인 실행 계획을 수립하였다.

또한, ‘모두의 말뭉치’의 활용성을 알리는 방안으로 ‘모두의 말뭉치 전문가 지식 강연회’를 개최하고 번역의 전문성 및 일관성을 제고하기 위하여 전문용어 번역집을 제작하였다. 그리고 국외 출장(인도네시아, 사우디아라비아)을 통해 국외 말뭉치 구축 현황을 파악하고 기관 간 협력 네트워크의 기틀을 다졌다.

이 사업의 기대 효과는 다음과 같다.

첫째, 인공지능의 핵심 자원인 데이터를 확보함으로써 언어 데이터 산업의 생태계 조성에 기여할 수 있을 것이다.

둘째, 다국어 언어 처리 및 인공지능 기반 통번역 모델의 품질을 지속적으로 향상시킬 수 있을 것이다.

셋째, 국가 정책 및 관련 업계의 수요를 충족하고 전략적 국제 협력 관계의 기반을 공고히 할 수 있을 것이다.

이 사업을 통해 구축된 한국어-외국어 병렬 말뭉치는 그 자체로서 가치가 있을 뿐만 아니라 다양한 분야에 활용되는 기초 자료로 그 역할을 할 수 있을 것으로 기대된다.

**주요어:** 한국어-외국어 병렬 말뭉치, 제2차 중장기계획, 전문가 지식 강연회, 전문용어 번역집, 베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어

<Abstract>

## 2025 Korean-Foreign Language Parallel Corpus Construction Project

This project follows the “2021 Korean-Foreign Language Parallel Corpus Construction Project,” the “2022 Korean-Foreign Language Parallel Corpus Construction Project,” the “2023 Korean-Foreign Language Parallel Corpus Construction Project,” and the “2024 Korean-Foreign Language Parallel Corpus Construction Project,” with the primary objective of constructing Korean-foreign language parallel corpora for languages with high industrial demand, particularly to support the development of Korean-based AI translation and interpretation systems, and to establish the Second Mid- to Long-Term Plan, which will serve as the foundation for project implementation from 2026 onward. The languages selected for this corpus construction include Vietnamese, Indonesian, Thai, Indian Hindi, Cambodian Khmer, Filipino Tagalog, Russian, and Uzbek.

The construction of the parallel corpus was carried out in three major stages: ‘collection,’ ‘construction,’ and ‘production.’ In the ‘collection’ stage, raw data of Korean colloquial and literary texts were gathered, followed by proofreading and refinement of the collected data. In the ‘construction’ stage, translations of the proofread and refined data were completed, followed by editorial revisions and a first round review by Flitto Inc. linguists (7% sample review). The translations were then fully reviewed a second time by a translation reviewer, followed by a third round of proofreading by the proofreading leader (20% sample review) and a final review by the supervisor (10% sample review). Once the translations had passed these quality checks, the data were then finalized in the ‘production’ stage, along with their associated meta-information.

As a result, a total of 8,000,000 words (based on the original Korean) of the Korean-foreign language parallel corpus were constructed, alongside

an additional 1,000,000 words (based on the original Korean) for the requested Korean-English parallel corpus as a supplementary output.

Subsequently, the Second Mid- to Long-Term Plan for Korean-Foreign Language Parallel Corpus Construction was established. To this end, the achievements of the First Mid- to Long-Term Plan, relevant environmental factors, and the needs of industry and academia were analyzed. Based on this analysis, the project set forth a vision and strategic framework and developed detailed implementation plans for each strategic task.

Additionally, to promote the utility of the "Modu Corpus," the project held the "Modu Corpus Expert Knowledge Lecture Series" and produced a Specialized Terminology Translation Guidebook to strengthen translation expertise and consistency. Through overseas business trips (to Indonesia and Saudi Arabia), the current state of overseas corpus construction was assessed, and the foundation for an inter-agency cooperation network was established.

The expected effects of this project are as follows:

First, by securing high-quality Korean-foreign language data, the core resource of artificial intelligence, the ecosystem of the language data industry can be fostered.

Second, the quality of multilingual language processing and artificial intelligence-based translation/interpretation models can be continuously improved.

Third, the demands of national policies and related industries can be met, and the foundation for strategic international cooperation can be solidified.

The Korean-foreign language parallel corpus constructed through this

project is expected to not only have intrinsic value but also serve as foundational data that can be used across various fields.

**Keywords:** Korean-foreign language parallel corpus, Second Mid- to Long-Term Plan, Expert Knowledge Lecture Series, Terminology Translation Guidebook, Vietnamese, Indonesian, Thai, Indian Hindi, Cambodian Khmer, Filipino Tagalog, Russian, Uzbek.



# 차 례

## 제1장 사업 개요

1. 사업의 목표 .....	1
2. 사업의 범위 .....	1
3. 사업 수행 절차 및 체계 .....	2
3.1. 사업 수행 절차 .....	2
3.2. 사업 수행 체계 .....	5
4. 사업 수행 내용 .....	9

## 제2장 사업 수행 내용

1. 원문 수집 및 정제 .....	15
1.1. 원문 수집 .....	15
1.2. 원문 정제 .....	17
1.3. 개인 정보 비식별화 .....	22
1.4. 저작권 확보 .....	23
2. 번역 지침 수정·보완 .....	25
2.1. 기존 지침 검토 .....	26
2.2. 지침 수정 및 보완 .....	26
2.3. 최종본 완성 .....	27
2.4. 적용 .....	28
3. 번역 .....	28
3.1. 번역 작업자 선정 .....	28
3.2. 번역 인력 교육 .....	29
3.3. 번역 절차 .....	30
3.4. 번역 데이터 구축 환경 .....	30
3.5. 번역 품질 향상 활동 .....	32
4. 검수 및 감수 .....	34
4.1. 번역 검수팀 .....	34
4.2. 감수자 .....	35
4.3. 번역 검수팀 및 감수자 교육 .....	36
4.4. 검수 품질 향상 활동 .....	42
5. 전문용어 번역집 .....	49
5.1. 제작 목적과 필요성 .....	49
5.2. 제작 과정 .....	49

# 차 례

<b>6. 한국어-외국어 말뭉치 구축 제2차 중장기계획 수립</b> .....	<b>53</b>
6.1. 제2차 중장기계획 수립 개요 .....	53
6.2. 제1차 중장기계획 성과 분석 .....	54
6.3. 환경 및 요구 사항 분석 .....	54
6.4. 비전 및 전략 체계 수립 .....	55
6.5. 실행 계획 .....	55
<b>7. 병렬 말뭉치 구축 및 메타 정보</b> .....	<b>56</b>
7.1. JSON 포맷 .....	56
7.2. 최종 데이터 구조 .....	57
7.3. JSON 예시 .....	58
<b>8. 보안</b> .....	<b>62</b>
8.1. 보안 관리 대상 및 담당자 .....	62
8.2. 보안 관리 방법 .....	63

## 제3장 사업 수행 결과

<b>1. 말뭉치 데이터 구축 결과</b> .....	<b>73</b>
1.1. 최종 구축 데이터 .....	73
1.2. 데이터 품질 감리 결과 .....	74
1.3. 번역 품질 비교 .....	74
<b>2. 대외 활동</b> .....	<b>112</b>
2.1. 국외 출장 .....	112
2.2. 모두의 말뭉치 전문가 지식 강연회 .....	118
<b>3. 기대 효과 및 제언</b> .....	<b>128</b>
3.1. 병렬 말뭉치의 활용 방안 .....	128
3.2. 사업의 기대 효과 .....	129
3.3. 제언 .....	130

# 표 차례

<표 1> 사업 범위 .....	1
<표 2> 사업 수행 절차별 세부 내용 .....	3
<표 3> 직책별 업무 분장 .....	6
<표 4> 월별 사업 추진 경과 .....	9
<표 5> 원문 수집 목표 및 실적 .....	15
<표 6> 원문 정보 .....	16
<표 7> 원문 예시 .....	16
<표 8> 기계적 정제 기준 .....	17
<표 9> 항목별 원문 정제 예시 .....	20
<표 10> 띄어쓰기 정제 예시 .....	21
<표 11> 인용문 정제 예시 .....	21
<표 12> 비윤리 원문 정제 예시 .....	22
<표 13> 번역 지침 제작 과정 .....	25
<표 14> 번역 및 교정 절차 .....	30
<표 15> 번역 및 교정 작업자 주요 업무 .....	30
<표 16> 1차 검수 작업자 주요 업무 .....	31
<표 17> 번역 품질 향상을 위한 기계적 정제 내용 .....	33
<표 18> 번역 품질 향상을 위한 구조적 검증 기준 .....	33
<표 19> 영어 데이터 품질 향상 활동 내용 .....	33
<표 20> 언어별 번역 검수원 인원 현황 .....	34
<표 21> 언어별 감수자 .....	36
<표 22> 검수·감수 인력 보안 교육 내용 .....	36
<표 23> 지침 공통 교육 내용 .....	37
<표 24> 언어별 지침 내용 .....	38
<표 25> 검수 유형별 검수 방법 .....	39
<표 26> 오류 유형 기준 .....	40
<표 27> 번역문 오류 유형 .....	45
<표 28> 언어별 번역 오류 양상 및 원인 .....	46
<표 29> 메타데이터 JSON 형식 예시(문어) .....	58
<표 30> 메타데이터 JSON 형식 예시(구어) .....	59
<표 31> 병렬 말뭉치 JSON 형식 예시(문어) .....	60
<표 32> 병렬 말뭉치 JSON 형식 예시(구어) .....	61
<표 33> 보안 관리 사항 .....	64

## 표 차례

<표 34> 보안 점검 항목 및 분류 .....	65
<표 35> 최종 산출물 저장 공간 .....	65
<표 36> 시스템 접근 통제 관련 법률 및 규정 .....	67
<표 37> 사업 참여자 보안 교육 .....	68
<표 38> 월별 보안 점검 실시 개요 .....	69
<표 39> 최종 구축 데이터 수량(어절) .....	73
<표 40> 최종 구축 데이터 수량(문장) .....	73
<표 41> 데이터 품질 감리 결과 .....	74
<표 42> 기계 번역과 병렬 말뭉치의 번역 비교(베트남어) .....	75
<표 43> 기계 번역과 병렬 말뭉치의 번역 비교(인도네시아어) .....	79
<표 44> 기계 번역과 병렬 말뭉치의 번역 비교(태국어) .....	85
<표 45> 기계 번역과 병렬 말뭉치의 번역 비교(인도 힌디어) .....	91
<표 46> 기계 번역과 병렬 말뭉치의 번역 비교(캄보디아 크메르어) .....	94
<표 47> 기계 번역과 병렬 말뭉치의 번역 비교(필리핀 타갈로그어) .....	100
<표 48> 기계 번역과 병렬 말뭉치의 번역 비교(러시아어) .....	104
<표 49> 기계 번역과 병렬 말뭉치의 번역 비교(우즈베크어) .....	107
<표 50> 케이팝포스트(KPopPost) 보도 기사 내용 요약 .....	114
<표 51> 2025년 모두의 말뭉치 전문가 지식 강연회 일정표 .....	119
<표 52> 2025년 모두의 말뭉치 전문가 지식 강연회 1부 주제 강연 내용 ..	119
<표 53> 2025년 모두의 말뭉치 전문가 지식 강연회 2부 포럼 발표 내용 ..	120
<표 54> 2025년 모두의 말뭉치 전문가 지식 강연회 2부 포럼 토론 내용 ..	121
<표 55> 병렬 말뭉치의 연구 및 교육 분야 활용 방안 .....	129

# 그림 차례

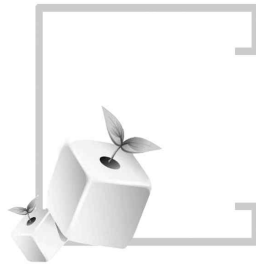
[그림 1] 한국어-외국어 병렬 말뭉치 구축 절차 요약도	3
[그림 2] 사업 수행 체계 및 인력 구성	6
[그림 3] 검수 플랫폼의 원문 신고 버튼	19
[그림 4] 1차 원문 정제 교육	20
[그림 5] 2차 원문 정제 교육	20
[그림 6] 전산 프로그램을 이용한 기계적 개인 정보 처리 과정 예시	23
[그림 7] (주)플리토 플랫폼 이용 약관 및 개인 정보 처리 방침	24
[그림 8] 양도 계약서 및 저작권 이용 허락 계약서	25
[그림 9] 번역 지침 수정 사항 예시(필리핀 타갈로그어)	27
[그림 10] 번역 지침 수정 사항 예시(러시아어)	27
[그림 11] 번역 지침 최종본 목록	28
[그림 12] 번역 작업자 선정 테스트 예시(베트남어)	29
[그림 13] 번역 및 작업 지침 교육 내용 예시	29
[그림 14] 번역(좌) 및 교정(우) 환경 예시	31
[그림 15] (주)플리토 플랫폼의 표본 검수 환경	31
[그림 16] 검수·감수 교육 절차	36
[그림 17] 검수·감수 보안 교육 예시	37
[그림 18] 지침 교육 예시(캄보디아 크메르어)	38
[그림 19] 검수 플랫폼 작업 환경 예시(인도네시아어)	39
[그림 20] 검수 플랫폼 사용 교육 예시(인도네시아어)	40
[그림 21] 감수 작업 환경 예시(필리핀 타갈로그어)	41
[그림 22] 감수 플랫폼 사용 교육 예시(인도네시아어)	41
[그림 23] 교육 이수증 예시	42
[그림 24] 3차 검수 예시 자료(필리핀 타갈로그어)	43
[그림 25] 3차 검수 결과에 대한 의견 교환 예시(러시아어)	43
[그림 26] 번역 검수원 대상 검수 재교육 예시(필리핀 타갈로그어)	44
[그림 27] 원문 이해 지원 예시	44
[그림 28] 번역 품질 개선을 위한 의견 작성 및 전달 예시(인도 힌디어)	45
[그림 29] 온용어 분야별 용어 현황	50
[그림 30] 온용어 용어집별 용어 현황	50
[그림 31] 온용어 '사회과학'의 중분류	50
[그림 32] 법제처 국가법령정보센터 법령 용어 수록 여부 확인	51
[그림 33] 전문용어 번역집 최종본 예시	53

## 그림 차례

[그림 34] metadata JSON 형식 .....	56
[그림 35] document_paragraph JSON 형식 .....	56
[그림 36] 병렬 말뭉치 JSON 형식 .....	57
[그림 37] 구축 데이터 샘플 Excel 예시 .....	57
[그림 38] 보안 담당자 역할 체계 .....	63
[그림 39] 사무실 출입문 전경(좌: 경희대학교 산학협력단, 우: ㈜플리토)	64
[그림 40] 보안 관리 대장 예시 .....	65
[그림 41] 저장 관리 예시(Cloud) .....	65
[그림 42] 저장 관리 예시(NAS) .....	66
[그림 43] 보안 교육 예시 .....	70
[그림 44] 케이팝포스트(KPopPost) 보도 기사 일부 발췌 .....	113
[그림 45] 인도네시아 출장 일정별 사진 .....	115
[그림 46] 사우디아라비아 출장 일정별 사진 .....	118
[그림 47] 2025년 모두의 말뭉치 전문가 지식 강연회 행사 사진 .....	125
[그림 48] 2025년 모두의 말뭉치 전문가 지식 강연회 홍보 예시(뉴스핌) ·	127







# 제 1 장

## 사업 개요





## 1. 사업의 목표

본 사업의 목표는 한국어-영어에 치중된 병렬 언어 자료의 불균형을 해소하고 대한민국의 교류 협력 관계로 주목받고 있는 국가 언어들을 대상으로 한국어-외국어 병렬 말뭉치를 구축하며, 2026년 이후 사업 추진의 기반이 될 제2차 중장기계획을 수립하는 것이다. 병렬 말뭉치 구축의 대상이 되는 8개 언어는 베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어이다.

본 사업에서는 고품질의 한국어-외국어 병렬 말뭉치를 구축함으로써 인공지능의 핵심 자원인 데이터를 확보하고, 다국어 언어 처리 및 인공지능 기반 통번역 기술의 성능을 향상시키며, 국가 정책과 산업계의 요구를 충족시켜 국제 협력 기반을 강화하였다. 또한 구축된 말뭉치를 다양한 분야에서 활용 가능한 기초 자료로 제공함으로써 4차 산업 혁명 시대 언어문화 산업의 성장 동력을 마련하고자 하였다.

## 2. 사업의 범위

본 사업의 범위는 먼저 2024년 한국어-외국어 병렬 말뭉치 구축 지침을 보완 및 적용하여 언어별로 100만 어절, 총 800만 어절을 구축하는 것이다. 8개 언어에는 베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어가 해당한다. 다음으로 제2차 중장기계획 수립을 위한 산업계 및 학계 전문가 자문단을 구성하여 2026년부터 2030년까지의 한국어-외국어 말뭉치 구축 제2차 중장기계획을 수립한다. 이 외에 ‘모두의 말뭉치’의 활용성을 알리는 방안으로 ‘모두의 말뭉치 전문가 지식 강연회’를 개최하고 번역의 전문성 및 일관성을 제고하기 위하여 전문용어 번역집을 제작한다.

<표 1> 사업 범위

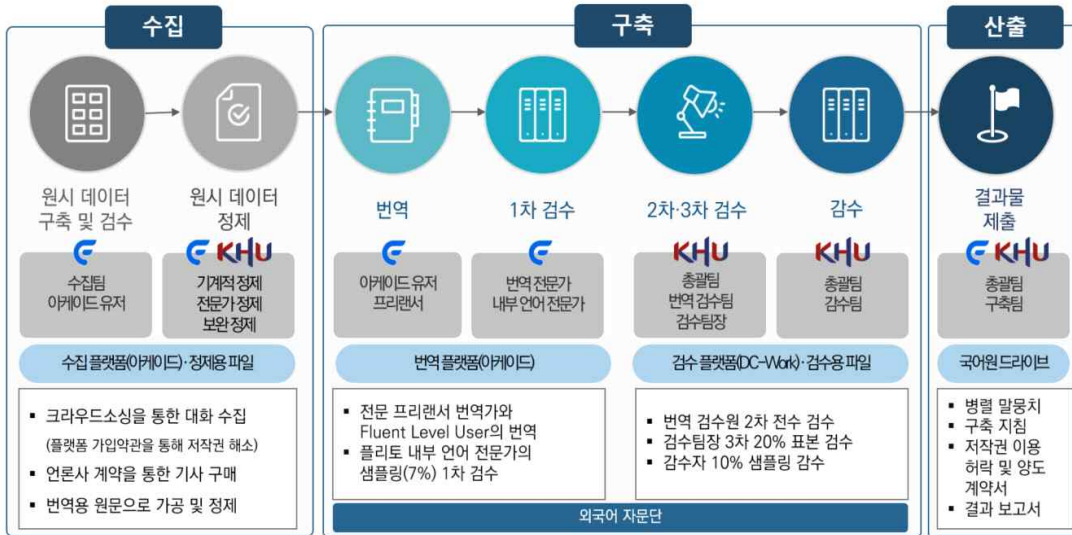
- 한국어-외국어 병렬 말뭉치 총 800만 어절 구축
  - 베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어 각 100만 어절
  - 전문가 감수·검수 체계 운영을 통한 병렬 말뭉치 품질 관리
- 한국어-외국어 말뭉치 구축 제2차 중장기(2026-2030) 계획 수립
- 모두의 말뭉치 전문가 지식 강연회 개최
- 전문용어 번역집 제작

### 3. 사업 수행 절차 및 체계

#### 3.1. 사업 수행 절차

본 사업은 ‘2021년 한국어-외국어 병렬 말뭉치 구축(이하 2021년 사업)’과 ‘2022년 한국어-외국어 병렬 말뭉치 구축(이하 2022년 사업)’, ‘2023년 한국어-외국어 병렬 말뭉치 구축(이하 2023년 사업)’, ‘2024년 한국어-외국어 병렬 말뭉치 구축(이하 2024년 사업)’의 연속 사업으로서 지난 네 차례 사업의 수행 절차를 기본 바탕으로 하되, 일부 단계들을 보완하여 사업 수행의 효율성을 제고하였다. 사업 과업은 크게 한국어-외국어 병렬 말뭉치 구축과 제2차 중장기계획 수립, 그리고 ‘모두의 말뭉치’ 활용성을 제고하는 모두의 말뭉치 전문가 지식 강연회 개최 및 전문용어 번역집 제작으로 나뉜다.

먼저, 한국어-외국어 병렬 말뭉치 구축은 데이터 수집과 정제, 번역, 검수, 감수, 결과물 제출로 이루어진다. 구체적으로 살펴보면, 원시 데이터 수집 단계에서는 ‘아주경제’와의 직접 계약을 통해 저작권이 해소된 문어 원문(인터넷 뉴스 기사)을 구매하였으며, 구어 원문은 국립국어원으로부터 ‘일상 대화 말뭉치 2024’를 제공받아 사용하였다. 원시 데이터를 수집한 후에는 기계적 정제와 전문가 정제, 보완 정제로 이루어진 총 3단계의 정제 절차를 수행하여 높은 수준의 원문 정제를 실시하였다. 정제된 원문은 번역 과정을 거친 후 (주)플리토의 내부 언어 전문가가 1차 검수 작업(7%)을 수행하였다. 1차 검수 단계를 통과한 데이터는 도착어 기준 원어민 검수원의 2차 검수(전수)와 검수팀장의 3차 검수(20%), 감수자의 감수(10%) 작업을 진행하였다. 이러한 과정을 통해서 최종적으로 총 800만 어절의 병렬 말뭉치를 산출하여 국립국어원에 제출하였다. 병렬 말뭉치 구축의 절차 요약도와 절차별 세부 내용은 다음과 같다.



[그림 1] 한국어-외국어 병렬 말뭉치 구축 절차 요약도

<표 2> 사업 수행 절차별 세부 내용

사업 수행 절차	업무 주체	세부 내용
수집	원시 데이터 수집 및 검수 (주)플리토	<ul style="list-style-type: none"> <li>다양한 분야 및 양질의 원시 데이터 수집 <ul style="list-style-type: none"> <li>원시 데이터 활용: 국립국어원 일상 대화 말뭉치 2024</li> <li>원시 데이터 구매: 경제 신문사 ‘아주경제’ 뉴스 원문</li> </ul> </li> <li>원시 데이터 저작권 확보 <ul style="list-style-type: none"> <li>원시 데이터 수집 시 해당 언론사와의 직접 계약을 통해 저작권이 해소된 문어 원문 구매</li> </ul> </li> </ul>
	원문 정제 경희대학교 및 (주)플리토	<ul style="list-style-type: none"> <li>원시 데이터의 3단계 정제 <ul style="list-style-type: none"> <li>1차 기계적 정제: 기준에 따른 데이터 기계 처리</li> <li>2차 전문가 정제: 국어학·한국어학 전공의 전문가로 구성된 원문 정제팀 운영으로 원문에 대한 집중적인 정제 실시</li> <li>3차 보완 정제: 번역·검수 과정에서 발견되는 오류에 대해 보완 정제 실시</li> </ul> </li> <li>데이터 관리 번호 및 분류·주제 등의 메타 정보 및 헤더 부착</li> </ul>

구축	번역	(주)플리토	· 번역 플랫폼 가입 절차 활용 및 별도의 저작권 이용 허락 계약 체결을 통한 번역 관련 저작권 확보
		경희대학교	· 2021~2024년 사업의 번역 지침 보완 및 제공
		(주)플리토	· (주)플리토 플랫폼 환경에서 프리랜서 전문 번역사를 활용하여 작업 수행 · 고역량 작업자가 교정을 실시
	1차 검수	(주)플리토	· 1차 표본 검수 - (주)플리토 내부 언어 전문가가 플랫폼 환경에서 번역문의 7%를 무작위로 표본 검수하여 1차 검수 수행 - 1차 검수에서 품질 기준 미달 문장은 재번역 실시
	2차 검수	경희대학교	· 2차 전수 검수 - 언어별 원어민 통번역 전문 인력을 활용하여 전체 번역문에 대한 2차 검수 수행 - 번역 품질 향상을 위한 검수 애로 사항 및 품질 관련 회의 실시
	3차 검수		· 3차 표본 검수 - 언어별 검수팀장이 검수문의 20%를 표본 검수 수행 - 검수 시 자주 발생하는 오류 정리 - 오류율이 높은 인력 대상으로 재교육 실시
	감수		· 표본 감수 - 한국어 관련 전공 교수의 10% 표본 감수 수행 - 검수의 품질 향상을 위하여 수시로 감수 의견 전달
	산출	결과물 제출	경희대학교 및 (주)플리토

다음으로, 문화체육관광부가 2021년 수행한 ‘한국 언어 문화 산업 빅데이터 구축 중장기 계획 수립’에 따라 진행된 한국어-외국어 병렬 말뭉치 구축 사업(2021~2025년)의 성과와 한계를 토대로 변화된 정책 환경과 기술 동향을 반영한 향후 5년간의 제2차 중장기계획(2026~2030년)을 수립하였다. 이를 위해 공

공 데이터 개방 전략, 디지털 플랫폼 정부 비전, 국가 인공지능 전략 등 상위 정책과의 정합성을 살피고 국내외 병렬 말뭉치 구축 사례와 인공지능 번역 기술 발전 추세를 분석하였다. 이후 중장기계획 수립을 위한 전문가 자문위원회를 3회 운영하였으며, 자문 위원단의 의견을 수렴하여 사업 추진 배경과 필요성, 목표, 범위, 방법을 구체화하였고, 타당성과 현장성을 제고하기 위해 사용자 및 산학 전문가를 대상으로 요구 사항을 분석하였다. 그리고 분석 결과를 바탕으로 비전 및 전략 체계를 설정하고 전략 과제별로 구체적인 실행 계획을 수립하였다.

또한 국립국어원 ‘모두의 말뭉치’ 활용성 제고를 위해 전문가 지식 강연회를 개최하였다. 지식 강연회는 운영 계획 수립, 장소·강연자 섭외, 홍보 및 참가 접수 등의 절차로 준비하였으며 사후 설문 조사를 통해 참여자의 만족도와 개선 의견을 파악하였다.

마지막으로 번역어의 일관성과 전문성을 확보하고 말뭉치 구축 과정에서의 용어 혼란과 번역 오류를 최소화하기 위해 국립국어원 언어정보나눔터 ‘온용어’에 수록된 ‘법률’ 분야 용어 가운데 전문용어를 선별하여 전문용어 번역집을 제작하였다. 전문용어 번역집은 주제 분야(도메인) 선정, 표제어 선별, 용어 번역, 감수 등의 절차로 진행되었다.

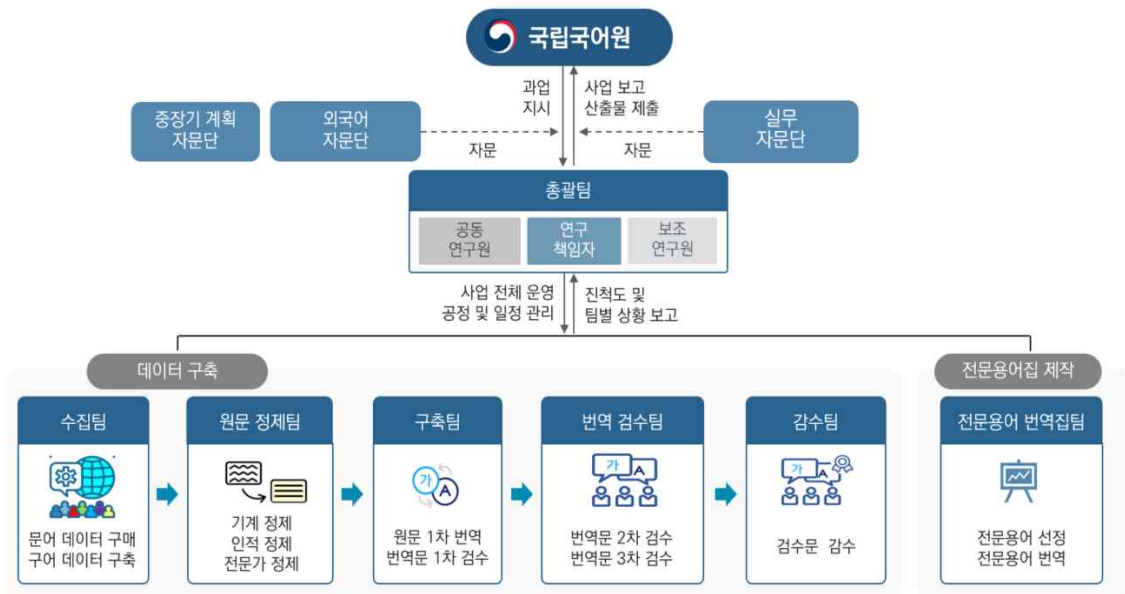
### 3.2. 사업 수행 체계

본 사업을 수행하는 조직 체계는 크게 총괄팀, 자문 위원단, 수집팀, 원문 정제팀, 구축팀, 번역 검수팀, 감수팀, 전문용어 번역집팀으로 나눌 수 있다.

총괄팀에서는 사업 계획 및 전반적 관리, 모두의 말뭉치 지식 강연회 개최, 제2차 중장기계획 수립을 수행하였다. 자문 위원단은 다시 세 분야로 나누어 사업 전반 자문을 위한 실무 자문단, 8개 언어와 관련한 외국어 자문단, 제2차 중장기계획 수립을 위한 전문가 자문단을 정기적으로 운영하여 사업 수행에 대한 자문을 구하였다.

다음으로 수집팀과 원문 정제팀, 구축팀, 번역 검수팀, 감수팀은 병렬 말뭉치 데이터에 직접적으로 관여했다. 수집팀과 원문 정제팀, 구축팀은 원시 데이터를 수집·정제·번역하였다. 번역 검수팀과 감수팀은 번역문을 검수·감수함으로써 최종 산출물 데이터의 품질을 제고하는 작업을 수행하였다.

본 사업에서는 말뭉치 구축 외에 전문용어 번역집을 제작하기 위해 전문용어 번역집팀을 조직하였다. 사업의 수행 체계와 직책별 업무 분장은 다음과 같다.



[그림 2] 사업 수행 체계 및 인력 구성

<표 3> 직책별 업무 분장

팀명	직책	업무 내용
총괄팀	연구 책임자	<ul style="list-style-type: none"> <li>· 사업 계획 전반의 수립</li> <li>· 세부 사업 과제의 조정 및 총괄</li> <li>· 사업 인력 배치 및 관리·감독</li> <li>· 사업비의 집행 전반에 대한 관리·감독</li> <li>· 사업 중간 및 최종 결과 보고</li> <li>· 국립국어원 및 컨소시엄사 상시 소통</li> </ul>
	공동 연구원	<ul style="list-style-type: none"> <li>· 사업 세부 계획의 수립 및 진행 상황 점검·조율</li> <li>· 언어별 검수 및 감수 인력 관리</li> <li>· 산출물 품질 및 사업 위험 관리</li> <li>· 각종 지침 및 교육 자료 보완</li> <li>· 사업 중간 및 최종 결과 보고 지원</li> <li>· 제2차 중장기계획 수립 및 관리</li> <li>· 모두의 말뭉치 전문가 지식 강연회 계획 수립 및 운영</li> <li>· 기타 사업 수행 지원</li> </ul>
	보조 연구원	<ul style="list-style-type: none"> <li>· 사업 관련 행정 사무 및 관리</li> <li>· 사업비 집행 및 정산, 증빙 자료 관리</li> <li>· 기타 사업 운영 지원</li> </ul>

	컨소시엄사 총괄 책임자	· 컨소시엄 사업 계획 전반 수립 및 총괄
	컨소시엄사 실무 책임자	· 컨소시엄 세부 사업 과제의 조정 · 컨소시엄 사업 인력 배치 및 관리·감독 · 컨소시엄 사업비 집행 전반에 대한 관리·감독 · 데이터 보안 관리·감독
수집팀	컨소시엄사 사업 운영 담당자	· 원시 데이터 수집 및 관리 · 데이터 보안 관리 · 컨소시엄 사업 관련 행정 사무 및 관리
원문 정제팀	공동 연구원	· 한국어 원문 정제 지침 보완 · 한국어 원문 정제 피드백 및 진척 관리
	보조 연구원	· 한국어 원문 정제 실시 및 피드백 반영
구축팀	컨소시엄사 사업 운영 담당자	· 번역 인력 및 진척 관리 · 내부 언어 전문가 관리 · 번역·검수·감수 데이터 관리 · 최종 데이터 산출 및 관리 · 데이터 보안 관리 · 컨소시엄 사업 관련 행정 사무 및 관리 · 병렬 말뭉치 구축 플랫폼 보완 설계 · 병렬 말뭉치 하자 보수
	번역사	· 번역 및 교정 실시
	내부 언어 전문가	· 1차 검수(7%) 실시 · 검수 시 논의 필요 사항 보고 · 언어별 번역 지침에 대한 피드백 제공
번역 검수팀	검수팀장	· 검수 진행 일정 점검 및 관리 · 언어별 번역 지침 보완 · 검수 전반의 오류 파악 및 분석 · 번역 검수원 대상 재교육 실시 · 2차 검수(전수) 및 3차 검수(20%) 실시
	번역 검수원	· 2차 검수(전수) 실시 · 검수 시 논의 필요 사항 보고 · 언어별 번역 지침에 대한 피드백
	보조 연구원	· 팀 내 행정 업무 보조 · 팀 회의 운영 지원 및 회의록 작성 · 한국어 원문 이해 지원 · 한국어 원문 정제 실시

감수팀	감수자	<ul style="list-style-type: none"> <li>· 감수(10%) 실시</li> <li>· 언어별 번역 지침 보완</li> <li>· 감수 상황 및 논의 필요 사항 보고</li> </ul>
전문용어 번역집팀	공동 연구원	<ul style="list-style-type: none"> <li>· 전문용어 번역집 제작 일정 점검 및 관리</li> <li>· 전문용어 번역 제작 및 보고</li> </ul>
자문 위원단	실무 자문 위원	<ul style="list-style-type: none"> <li>· 사업 추진 방향 자문</li> <li>· 세부 과업 내용에 대한 자문</li> <li>· 병렬 말뭉치에 대한 국어학적 자문</li> <li>· 제2차 중장기계획에 대한 자문</li> </ul>
	외국어 자문 위원	<ul style="list-style-type: none"> <li>· 언어별 번역 지침에 대한 자문</li> <li>· 언어별 번역 관련 문제에 대한 자문</li> <li>· 전문용어 번역집 제작에 대한 자문</li> <li>· 제2차 중장기계획에 대한 자문</li> </ul>
	중장기계획 자문 위원	<ul style="list-style-type: none"> <li>· 제2차 중장기계획에 대한 자문</li> <li>- 사업 기본 방향 및 기술·데이터 구조에 대한 자문</li> <li>- 병렬 말뭉치 구축 모델과 확장 전략에 대한 자문</li> <li>- 제2차 중장기계획 보고서 구성 및 내용 점검 자문</li> </ul>

#### 4. 사업 수행 내용

<표 4> 월별 사업 추진 경과

월 추진 과업	2025년											
	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월		
사업 계획 수립												
국립국어원 정기/비정기 보고	착수						중간					완료
각종 지침 보완												
참여 인력 교육												
원문 수집 및 정제												
산출물 중간 납품		샘플	번역문 10%					번역문 50%		검수 감수문 70%		
번역 및 1차 검수(7%)												
2차 검수(전수)												
3차 검수(20%)												
감수(10%)												
최종 점검 및 산출물 납품												
제2차 중장기계획 수립												
전문용어 번역집 제작												
컨소시엄 정례 회의												
중장기계획 수립 자문 회의												
실무 자문 회의												
외국어 자문 회의												
감리					설계						종료	

### 1) 사업 계획 수립 및 착수 보고

- 사업 계획 수립: 2025년 3월 27일 완료
- 착수 보고회: 2025년 3월 18일 실시

### 2) 원문 구축 및 정제

- 원시 데이터 수집
  - 한국어 원시 데이터: 2025년 4월 25일 완료
  - 저작권 확보: 2025년 4월 25일 완료
- 원시 데이터 검수 및 정제: 2025년 8월 4일 완료

### 3) 각종 지침 보완 및 참여 인력 교육

- 지침 보완: 2025년 4월 15일 완료
- 참여 인력 교육: 2025년 4월 29일 완료
  - 이후 인력 총원에 따라 수시로 교육 실시

### 4) 번역 및 1차 검수

- 원문 데이터 번역: 2025년 9월 10일 완료
- 번역 데이터 1차 검수(7%): 2025년 9월 17일 완료

### 5) 2·3차 검수 및 감수

- 번역 데이터 2차 검수(전수): 2025년 10월 23일 완료
- 번역 데이터 3차 검수(20%): 2025년 11월 13일 완료
- 번역 데이터 감수(10%): 2025년 11월 13일 완료

### 6) 제2차 중장기계획 수립

- 중장기계획 자문 회의 실시(3회)
  - 1차: 2025년 4월 30일
  - 2차: 2025년 6월 13일
  - 3차: 2025년 12월 3일
- 중장기계획 수립: 2025년 12월 15일 완료

### 7) 전문용어 번역집 제작

- 전문용어 번역집 제작: 2025년 12월 15일 완료

## 8) 의사소통 및 자문 의견 수렴

### ○ 단계별 보고

- 착수 보고회: 2025년 3월 18일 실시
- 중간 보고회: 2025년 8월 26일 실시
- 최종 보고회: 2025년 12월 8일 실시

### ○ 국립국어원 보고 및 컨소시엄사 회의

- 국립국어원 주간 및 월간 보고 실시
- 컨소시엄사 격주 회의 진행

### ○ 자문 의견 수렴

- 실무 자문 회의 실시(3회)
  - 1차: 2025년 4월 25일
  - 2차: 2025년 7월 19일
  - 3차: 2025년 12월 5일
- 외국어 자문 회의 실시(3회)
  - 1차: 2025년 5월 19일
  - 2차: 2025년 6월 25일
  - 3차: 2025년 11월 19일

## 8) 사업 효과 확산

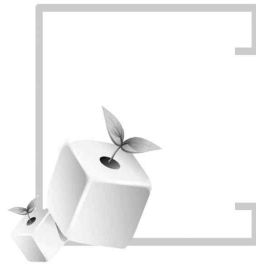
### ○ 모두의 말뭉치 전문가 지식 강연회 개최

- 일자: 2025년 10월 1일
- 주제: 인공지능 시대, 우리의 언어 자원과 미래
- 장소: 롯데월드타워 SKY31 컨벤션 오디토리움

### ○ 국외 출장

- 인도네시아: 2025년 11월 9~13일 실시
- 사우디아라비아: 2025년 11월 23~27일 실시





## 제 2 장

### 사업 수행 내용





## 1. 원문 수집 및 정제

### 1.1. 원문 수집

#### 1) 수집 목표 및 수집 결과

본 사업에서는 한국어와 외국어의 병렬 말뭉치를 구축하기 위해 먼저 한국어 원문을 수집하였다. 고품질의 원천 데이터 확보를 위해 검수·교정 과정을 거쳤으며 이 과정에서는 사용 불가 문장, 저품질 문장 등이 제거·교정되었다. 이 과정에서 수집 수량과 구축 목표 수량 간에 차이가 생길 것을 고려해 구축 목표 수량보다 10~20%가량 상향하여 수집하는 것을 목표로 하였다.

<표 5> 원문 수집 목표 및 실적

구분	구축 목표	수집 목표	수집 실적	획득 방식	세부 내용
구어	600,000어절	720,000어절	711,965어절	국립국어원 제공	국립국어원 일상 대화 말뭉치 2024
문어	400,000어절	480,000어절	492,482어절	아주경제 구매	2024년 아주경제 인터넷 뉴스 기사

#### 2) 원문 선정 과정

말뭉치는 언종의 사고방식과 의식 구조를 반영하는 자료이므로 특정 영역에 편중되지 않고 다양한 주제와 정보가 고르게 포함된 원문을 활용하는 것이 중요하다. 이에 본 사업에서는 고품질 데이터 확보를 위해 2024년의 아주경제 인터넷 뉴스 기사와 국립국어원의 일상 대화 말뭉치 2024를 각각 문어와 구어 데이터 원문으로 활용하였다.

아울러 말뭉치가 다양한 활용 목적에 따라 효과적으로 사용될 수 있도록 원문 선정 단계에서 주제, 성별, 연령대 등 다양한 메타 정보를 균형 있게 반영하여 자료의 편향성을 사전에 완화하고자 하였다. 특히 신문 말뭉치는 국내 뉴스 기사를 기반으로 정치·사회·문화 등 시의성 있는 다양한 이슈를 폭넓게 다루고 있으며, 육하원칙에 따라 작성되어 정보 전달과 의사 표현이 명확하다는 점에서 여러 산업 분야에 적용 가능한 적절한 자료로 판단하였다.

### 3) 원문 수집 절차

<표 6> 원문 정보

유형	원문	출처	세부 분야
구어	일상 대화 말뭉치 2024 (2인 일상 대화)	국립국어원	반려동물
			대인 관계
			건강과 미용
			식음료
			예술
			쇼핑
			일과 직업
			여행
			주거와 환경
			전문 분야
			여가와 오락
			사회
문어	아주경제 인터넷 뉴스 기사 2024	아주경제	IT/과학
			정치
			산업
			사회
			문화·연예/스포츠
			글로벌
			경제

구어 데이터 원문은 국립국어원으로부터 일상 대화 말뭉치 2024를 제공받아 사용하였으며, 문어 데이터 원문은 아주경제의 인터넷 뉴스 기사를 구매하여 신문 말뭉치에 포함된 문화·연예, 사회, 정치 등 다양한 분야의 신문 기사를 활용하였다. 이상의 절차로 수집된 원문의 예시는 다음과 같다.

<표 7> 원문 예시

구분	출처	수집 원문
문어	아주경제 인터넷 뉴스 기사 2024	지난주 신규 실업 수당 청구 건수가 1년 만에 최대 규모로 감소한 것으로 나타나면서 최근 경기 침체 공포를 누그러뜨렸다.
		음주 운전 뺑소니 혐의로 구속된 가수 name1의 술자리

		동석 연예인으로 name2와 name3이 지목됐다.
		또 회사는 분산된 DB들이 실시간 상호 데이터 동기화를 통해 원활한 데이터 통합 처리를 지원해, 대량 데이터 처리 속도 개선으로 업무 생산성도 향상시킬 것으로 기대하고 있다.
		이번 회의에서는 우크라이나 상황, 신흥 안보 위협 등에 대해 논의할 예정이다.
		챗GPT나 클로드 등 거대 언어 모델(LLM)을 구축하고 훈련하는 데는 막대한 비용이 든다.
구어	국립국어원 일상 대화 말뭉치 2024	요즘 ENFP 친구들을 만나면서 느끼는데 어떤 상황이든 그렇게 긍정적인 게 좋은 거 같아.
		음, 요즘에는 이제, 헬스를 name3이란 친구와 같이 다녀.
		어, 음악 중에서도 그, 기악, 성악 그런 것이 있는데 내가 악기를 다루는 것도 없고 그런데 성악은 못해도 어쨌든 본인이 목소리를 내면 되는 거잖아.
		너는 그러면 나중에 직장 생활을 하게 된다면 재택근무를 하고 싶어?
		학생들이 공부하는 이유는 시험이라고 생각을 하는데 혹시 시험을 잘 친 편이야?

## 1.2. 원문 정제

이상의 과정을 통해 수집한 원문은 번역에 용이하도록 정제 과정을 거쳤다. 본 사업에서는 3단계에 걸쳐 원문 정제를 진행하였으며 이를 통해 원문 품질을 최대한 높일 수 있도록 하였다. 정제 단계별 세부 내용은 다음과 같다.

### 1) 1차 정제(기계적 정제)

수집한 한국어 원문은 가장 먼저 기계적인 정제 과정을 통해 어절 수, 특수 문자(이모티콘), 개인 정보, 유사·일치 문장의 삭제 등 문장을 1차적으로 다듬는 작업을 진행하였다.

<표 8> 기계적 정제 기준

정제 기준	내용
어절 수	구어: 5~25어절, 문어: 7~30어절

부적합 내용	부적절하거나 자연스럽지 않은 내용, 불필요한 발화, 오류가 있는 문장(오·탈자 포함), 이모티콘 등 삭제
개인 정보	전화번호, 이메일 등의 개인 정보 비식별화 및 삭제
중복 내용	완전 일치 문장 및 유사 일치 문장 삭제
문장 형태 정제	형태소, 발화자 정보를 분석해 비구조화된 문장 정제

## 2) 2차 전문가 정제

본 사업을 통해 산출되는 자료는 국립국어원이 구축하는 공공 데이터로서 차후 8개 언어 병렬 말뭉치의 기준 틀로 활용될 수 있는 기반 자원이다. 이에 따라 자료는 규범성과 표준성을 갖추어야 하며, 표현상의 다양성을 인정하더라도 원칙과 허용 범위가 정해진 항목에 대해서는 원칙을 우선 적용하여 전체의 일관성을 유지하는 것이 필요하다.


또한 공공 데이터 특성에 부합하도록 공공성과 윤리적 기준을 충분히 확보하는 동시에 실제 언어 사용의 특성 역시 반영되어야 한다. 이러한 작업은 일반인 수준의 작업자가 수행하기에는 난도가 높으며, 언어학·한국어학 전공자들이 전문적 지식을 바탕으로 신중하게 판단해야 할 수준의 분석을 요구한다.

더불어 한국어는 문장 성분의 생략과 어순 변동이 비교적 자유롭고 고맥락적 특성이 강해 문장을 단독으로 분리할 경우 의미 파악이 어려운 경우가 적지 않다. 이러한 언어적 특징으로 인해 특정 언어와의 번역 과정에서 어려움이 발생할 수 있으므로, 이를 감안한 정제 작업이 필수적이다.

이에 본 사업에서는 공공 데이터로서 병렬 말뭉치의 규범성과 적정성을 강화하기 위해 2021년부터 2024년까지 마련된 원문 정제 지침을 재검토하고 전문성을 갖춘 원문 정제 전담팀을 별도로 구성하여 운영하였다. 정제팀은 대학교수급 공동연구원 5명과 한국어학·언어학 분야의 학사 및 석사 이상 학위를 보유한 정제원 14명으로 조직하였다.

## 3) 3차 정제(보완 정제)

전문가에 의해 정제된 원문은 번역 후 검수 과정을 거쳤다. 검수 단계에서 번역 검수원들은 원문 자체에 오류가 발견될 경우 ‘원문 신고’ 기능을 이용해 해당 문제를 보고하도록 하였다.

OriginText	Translation
<p>준비가 충분히 될 때 하고 싶다며,  </p>	<p>Anh nói là muốn làm lúc đã  sẵn sàng đầy đủ cơ mà.</p>

[그림 3] 검수 플랫폼의 원문 신고 버튼

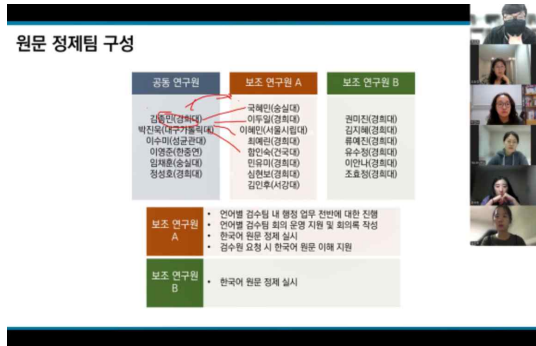
이 과정에서 번역 검수원들은 즉시 원문을 신고하는 방식이 아니라, 먼저 해당 언어를 담당하는 공동·보조 연구원과 협의한 뒤 원문 오류가 명확할 때에만 신고하도록 하였다. 신고가 접수된 원문은 공동 연구원들이 다시 검토하여 의미 변동이 없는 범위에서 적절한 문장으로 수정하거나, 수정이 불가능한 경우에는 삭제하였다. 이후 수정된 원문을 기반으로 재번역과 재검수 절차를 진행하였다.

아울러 번역 검수와 감수 작업이 이루어지는 동안 원문 정제팀에서도 원문 전체를 다시 한번 점검하였다. 이처럼 원문은 최소 세 차례 이상의 검토 절차를 거침으로써 품질을 제고하였다.

#### 4) 원문 정제 지침 교육

본 사업의 특성과 국어학적 배경지식을 갖춘 정제자라 하더라도 여러 인원이 동시에 작업에 참여하는 만큼 정제의 통일성을 확보하는 것이 중요하다. 이에 모든 정제자를 대상으로 정제 지침에 관한 교육을 실시하였다.

1차 교육은 2024년 사업 참여 경험이 있는 정제원을 대상으로 진행하였으며 정제 지침의 주요 변경 사항과 2025년 원문의 특성을 중심으로 안내하였다. 이어서 2차 교육은 2024년 사업에 참여하지 않은 정제원을 대상으로 오프라인으로 진행하였으며, 사업 개요 설명, 세부 지침 해설, 정제 실습 등을 포함하였다.



[그림 4] 1차 원문 정제 교육

[그림 5] 2차 원문 정제 교육

## 5) 원문 정제 예시

항목별 원문 정제의 예시는 다음과 같다.

<표 9> 항목별 원문 정제 예시

구분	원문 및 정제 내용
의미 명료화	<ul style="list-style-type: none"> <li>다만 아마추어 신분이라 우승 상금(2400만 엔)은 ~</li> <li>→ 다만 <b>이효송</b>은 아마추어 신분이라 우승 상금(2400만 엔)은 ~</li> </ul>
비표준어	<ul style="list-style-type: none"> <li>~ 오랜 기간 공을 들여 생성한 것마냥 너무 생생했기 ~</li> <li>→ ~ 오랜 기간 공을 들여 생성한 것처럼 너무 생생했기 ~</li> </ul>
오타	<ul style="list-style-type: none"> <li>~ 고객사의 협조를 받아 폐포장재를 회수 후 ~</li> <li>→ ~ 고객사의 협조를 받아 폐포장재를 회수 후 ~</li> </ul>
표기법	<ul style="list-style-type: none"> <li>생산성 향상을 위해선 자동화 솔루션을 도입해 ~</li> <li>→ 생산성 향상을 위해선 자동화 <b>설</b>루션을 도입해 ~</li> </ul>
두 문장 이상	<ul style="list-style-type: none"> <li>나도 수영은 이제 해 보고 싶어서 사실 나는 맥주병이라 수영을 잘 못해서.</li> <li>→ 사실 나는 맥주병이라 수영을 잘 못해.</li> </ul>

띄어쓰기의 경우 원칙과 허용 기준이 존재하지만 2022~2024년 사업에서 국립국어원과 협의하여 일관성 확보를 위해 원칙을 적용하기로 한 바 있으며, 본 사업 역시 동일한 기준을 유지하였다. 전문용어와 고유명사의 띄어쓰기는 단어 단위로 구분해 쓰는 것을 기본 방침으로 하되, 구어 표기에서는 실제 언어 사용을 반영하기 위해 붙여 쓰는 방식을 허용하였다. 다만 이러한 예외를 적용할 때에도 하나의 텍스트(대화문, 기사문) 내에서는 띄어쓰기 방식이 일관되게 유지되도록 하였다.

<표 10> 띄어쓰기 정제 예시

구분	원문 및 정제 내용
합성어	• 5월 4일부터 6일까지 이어지는 어린이날 <b>황금 연휴</b> (→ <b>황금연휴</b> )가 시작됐다.
의존 명사	• ~ 일반적인 훈련이 진행되는 가운데 눈길을 <b>끈건</b> (→ <b>끈 건</b> ) ~
접사	• 이는 1주 전보다 <b>8% 가량</b> (→ <b>8%가량</b> ) 하락한 것이다.
보조 용언	• ~ 첫 번째 시험에서 그 목표를 <b>당당히 이뤄냈다</b> (→ <b>이뤄 냈다</b> ).
전문용어	• 먼저, <b>국가 무형 유산</b> (→ <b>무형유산</b> ) 종묘 제례악 '일무' 분야의 ~
고유명사	• ~ 지난 9일 부산 해운대구 <b>카페051</b> (→ <b>카페 051</b> ) 본사에서 ~

본 사업에서 활용한 문어 원문은 2024년 사업과 동일하게 신문 기사이다. 기사 문 특성상 직접 인용과 간접 인용이 혼재하는 경우가 많다. 이에 한국어 어문 규범에 부합하도록 인용 부호와 관련 기호들을 재정비하였다. 간접 인용문에 작은따옴표를 반드시 사용해야 한다는 규정은 없지만, 한글 맞춤법(부록 9항)에서 제시하듯 ‘문장 내용 중 주의를 기울여야 하거나 특별히 부각할 부분을 드러낼 때 작은따옴표 사용이 가능하다’는 점을 고려하여 번역 과정에서 인용이 필요한 부분을 명확히 할 필요가 있는 경우에 한해 작은따옴표 사용을 허용하였다.

구어 원문은 직·간접 인용이 자주 나타나지만 대화 전사를 기반으로 하기 때문에 원시 데이터에는 이러한 구분이 명확히 표시되어 있지 않았다. 이에 정제 과정에서 한국어 어문 규범에 따라 적절한 인용 조사와 기호를 부착하여 정비하였다.

<표 11> 인용문 정제 예시

구분	원문 및 정제 내용
직접 → 간접	• ~ 최고 위원도 "이제 막 수술이 끝났으니, 의식을 회복하고 있다"며 민주당 의원들이 ~ → ~ 최고 위원도 '이제 막 수술이 끝났으니, 의식을 회복하고 있다'며 민주당 의원들이 ~
인용 표시 누락	• 대화는 현지인급으로 하는데 물 주세요라는 말은 못 해. → 대화는 현지인급으로 하는데 "물 주세요."라는 말은 못 해.

이 외에도 공공 데이터로서의 성격에 맞지 않는 비윤리적 내용이나 표현이 포함된 원문은 논의 후 삭제 처리하였다.

<표 12> 비윤리 원문 정제 예시

구분	원문 및 정제 내용
비하 (집단·계층)	<ul style="list-style-type: none"> <li>그러니까 요즘 애들 생각이 이상한 거야.</li> </ul>
비하 (국가·인종)	<ul style="list-style-type: none"> <li>근데 이제 결국 못사는 나라들 같은 경우에는 사람들이 학문이나 예술보다 일단 먹고사는 데 급급한 거지.</li> </ul>
비하 (지역)	<ul style="list-style-type: none"> <li>일반적으로 그 인천 친구들은 보면은 약간 좀 약간 거칠다.</li> </ul>
차별 표현	<ul style="list-style-type: none"> <li>아 좀 여자다 보니까 부정적인 생각이 많이 드는데.</li> </ul>
비속어	<ul style="list-style-type: none"> <li>이 티라는 성격을 가진 사람 왜 티발놈이라고 하잖아요 저희.</li> </ul>

### 1.3. 개인 정보 비식별화

본 사업에서 구축되는 병렬 말뭉치 데이터는 공개 데이터로 활용되는 만큼, 개인 정보가 외부로 노출되지 않도록 정제 단계에서 비식별화 절차를 거쳤다.

기계 정제 단계에서는 전산 프로그램의 정규식을 활용하여 수집된 원문에 개인 정보가 포함되어 있는지를 자동으로 점검하고 해당 정보가 확인될 경우 이를 제거하도록 조치하였다.

이후 전문가 정제 단계에서도 기업명이나 상표명이 상업적 홍보의 성격을 띠는 경우, 또는 개인·단체의 이름이 비윤리적이거나 민감한 내용과 연결되는 경우에는 해당 표현을 삭제하거나 비식별화 처리하였다.

```

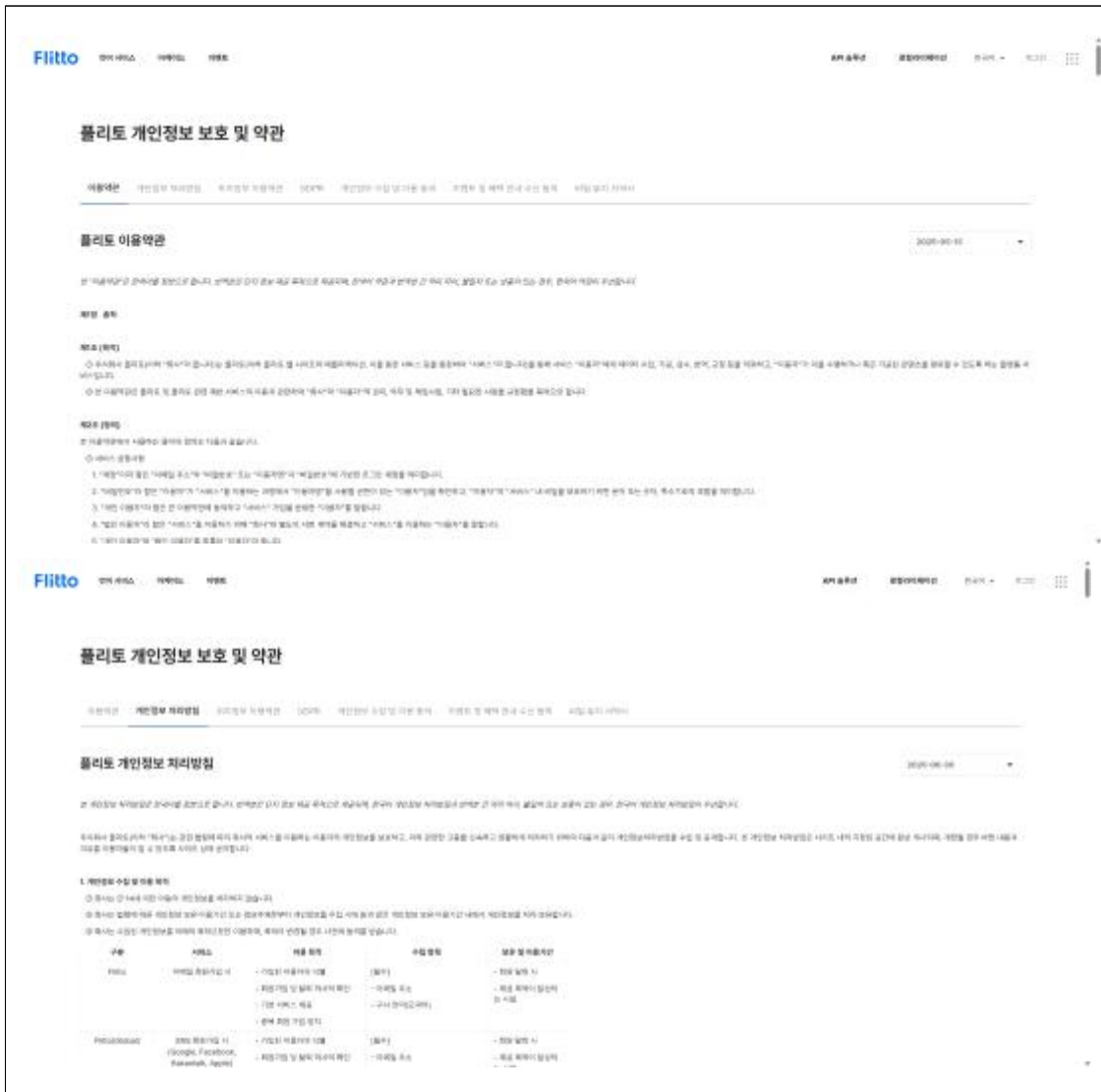
def make_personal_info_col(df, text_col, result_col):
    df = df.with_columns(
        pl.when(
            pl.col(text_col).apply(
                lambda x: re.findall(r"[0-9]{2,4}-[0-9]{3,4}-[0-9]{3,4}", x) != []
            )
        )
        .then("Y")
        .when(
            pl.col(text_col).apply(
                lambda x: re.findall(
                    r"[a-zA-Z0-9\.\@\-\_\&]{2,}\.[a-zA-Z0-9\.\@\-\_\&]{2,}", x
                )
                != []
            )
        )
        .then("Y")
        .when(pl.col(text_col).apply(lambda x: x.startswith("기획 ")))
        .then("Y")
        .otherwise("")
        .alias(result_col)
    )
    return df

```

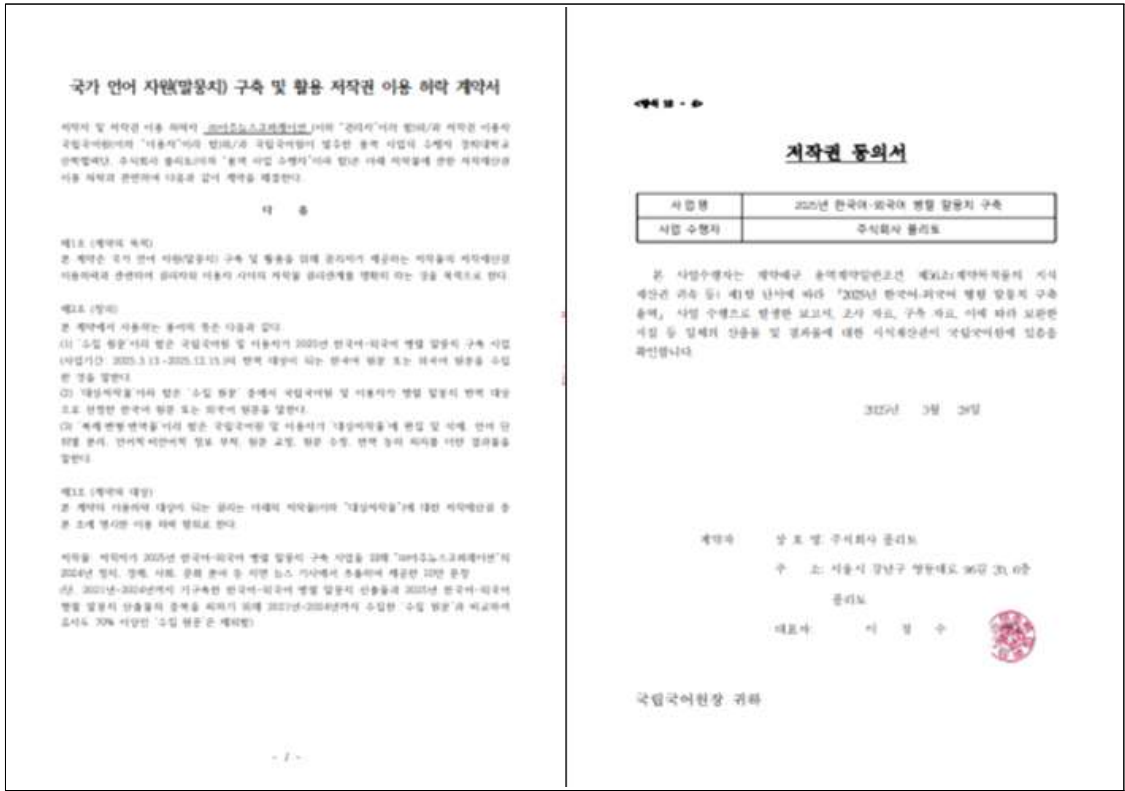
[그림 6] 전산 프로그램을 이용한 기계적 개인 정보 처리 과정 예시

#### 1.4. 저작권 확보

본 사업에서는 수집된 자료가 저작권 측면에서 안전하게 활용될 수 있도록 원문에 대한 권리를 확보하였으며, 번역물의 자유로운 배포가 가능하도록 양도 계약 및 저작권 이용 허락 계약을 체결하여 관련 문제를 정비하였다.



[그림 7] (주)플리토 플랫폼 이용 약관 및 개인 정보 처리 방침



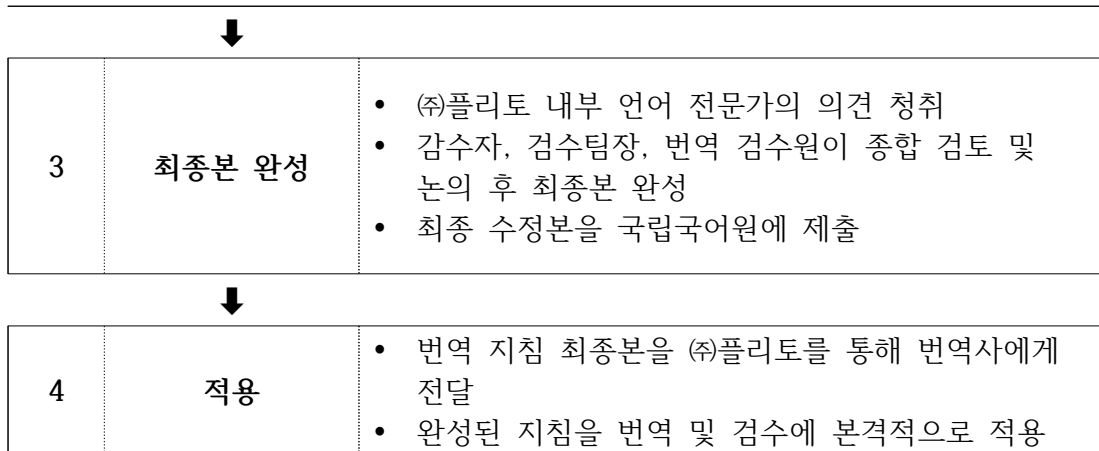
[그림 8] 양도 계약서 및 저작권 이용 허락 계약서

## 2. 번역 지침 수정·보완

2025년 한국어-외국어 병렬 말뭉치 구축을 위해 한국어를 8개 언어로 번역, 검수, 감수할 때 일관된 기준을 적용하고자 지난 사업의 8개 언어별 세부 지침을 수정 및 보완하였다. 번역 지침의 제작 과정을 구체적으로 밝히면 다음과 같다.

<표 13> 번역 지침 제작 과정

순서	작업	주요 내용
1	기존 지침 검토	<ul style="list-style-type: none"> <li>2024년 사업 지침에서의 외국어 번역(8개 언어) 지침 검토(지침에 치명적 오류나 수정 및 보완 사항이 없는지 논의)</li> <li>번역 용례 등 관련 자료 참고</li> </ul>
↓		
2	지침 수정 및 보완	<ul style="list-style-type: none"> <li>수정 필수 사항을 정리하고, 더 나은 번역과 검수, 감수를 위해 기존 지침을 상세화하거나 예시 추가</li> <li>1차로 수정된 지침을 (주)플리토에 전달</li> </ul>



## 2.1. 기존 지침 검토

2025년 사업의 번역 지침 수립을 위해, 먼저 (『한국어-외국어 병렬 말뭉치 구축을 위한 번역 지침(2023.3.)』)과 2024년 사업에서 적용된 8개 언어별 세부 지침을 종합적으로 검토하였다.

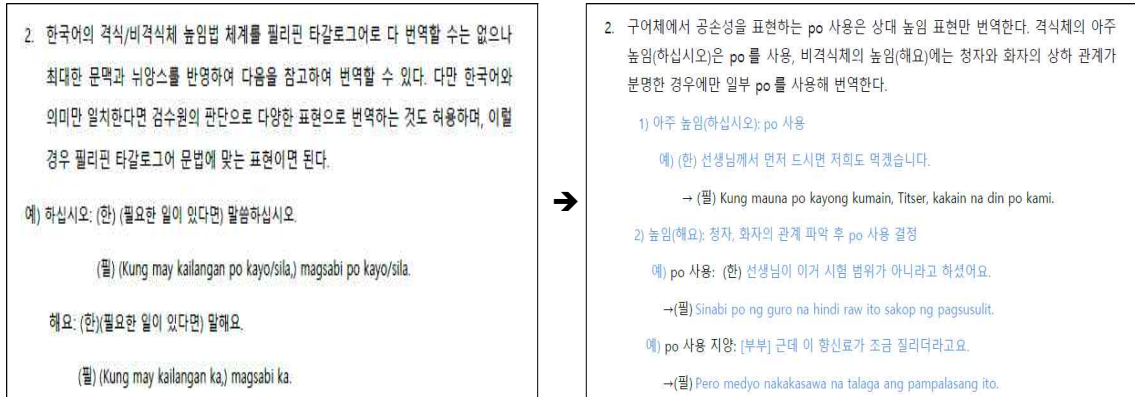
해당 지침은 1차 사업부터 단계적으로 축적된 경험과 피드백을 바탕으로 발전해 왔다. 2021년에는 문화체육관광부 훈령 「공공 용어의 외국어 번역 및 표기 지침」과 (주)플리토의 번역 가이드라인을 참고하여 기본 틀이 마련되었으며, 언어별 특징, 사용 지역의 관련 규정, 번역 관례, 실제 번역 데이터의 검수 결과 등이 반영되었다. 이후 2022년에는 실무자 및 자문위원의 검토를 통해 지침을 보완하였고, 이러한 내용을 기반으로 2023년에는 번역 지침서를 출판하여 활용도를 높였다. 2024년에는 출판된 지침을 바탕으로 실제 번역 및 검수 과정에서 확인된 사항들을 반영해 언어별 세부 지침을 추가적으로 정비하였다.

2025년 사업에서는 이러한 연속적인 개선 과정을 바탕으로, 기존 지침의 원칙과 구조를 유지하면서도, 최신 번역 사례에 맞게 실용성과 일관성을 강화하였다.

## 2.2. 지침 수정 및 보완

지난 사업에 사용된 기존 지침과 관련 자료를 검토한 후 검수팀장과 번역 검수원이 논의하여 수정·보완 사항을 도출하고 구체적인 지침 항목이나 용례를 개선하였다. 지난 사업에서 이미 여러 차례 검토된 지침이었기 때문에 대체로 더 나은 번역과 검수를 위해 기존 지침을 상세화하거나 예시를 추가하였다. 공동 연구원들은 지침 내 한국어 설명 부분이나 지침의 체계 등을 검토하고 검수팀장과 함께 논의하여 필요에 따라 수정·보완하였으며 이후 검수팀장, 번역 검수원이 함께 논의하

여 필수 수정 사항을 정리하였다. 이상의 과정을 통해 수정·보완된 지침은 (주)플리토를 통해 번역사들에게 공유하였다.



[그림 9] 번역 지침 수정 사항 예시(필리핀 타갈로그어)

3. 책, 그림, 노래, 영화와 같은 예술 작품의 경우 러시아어로 번역된 공식 제목을 사용한다. 이러한 공식 표기가 없다면 러시아어로 음역 표기하되, 제목이 간단한 단어나 짧은 문장으로 이루어진 경우에는 의미역도 허용한다.



3. 책, 그림, 노래, 영화와 같은 예술 작품의 경우 러시아어로 번역된 공식 제목을 사용한다. 이러한 공식 표기가 없다면 러시아어로 음역 표기하되, 제목이 간단한 단어나 짧은 문장으로 이루어진 경우에는 의미역도 허용한다.

구분	한국어	러시아어	번역 방식
소설	채식주의자	Вегетарианка	공식 제목 사용
영화	기생충	Паразиты	
노래	봄날	Весенний день	의미역

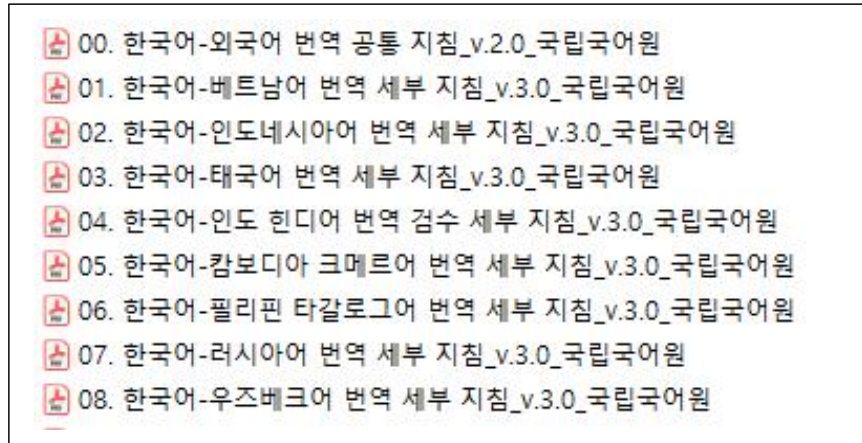
[그림 10] 번역 지침 수정 사항 예시(러시아어)

### 2.3. 최종본 완성

1차로 수정·보완된 지침에 대한 (주)플리토 번역사들의 의견을 청취한 후 감사자, 검수팀장, 번역 검수원이 다시 한번 종합적으로 검토 및 논의하여 최종본을 완성하였다. 이러한 과정에서 번역의 품질은 향상시키면서도 일관성과 통일성을 유지하기 위한 사항들을 면밀히 검토하였으며, 2024년 사업에서 사용된 지침의 원리와 규정을 대부분 유지하여 사업 간 연계성과 일관성을 최대한 높이고자 하였다. 이상의 과정을 통해 수정·보완된 최종 지침은 국어원에 전달하였다.

## 2.4. 적용

번역 지침 최종본은 (주)플리토와 공유한 후 이를 본격적으로 번역, 검수, 감수에 적용하였다. 이를 통해 번역 작업의 효율성과 일관성을 유지하면서 품질을 향상시키고자 하였다.



[그림 11] 번역 지침 최종본 목록

## 3. 번역

### 3.1. 번역 작업자 선정

번역 데이터의 품질을 제고하기 위해 본 사업에서는 사업 전반에 대한 이해도가 높고 2021~2024년 사업에 참여한 이력이 있는 프리랜서 전문 번역사(이하 ‘작업자’) 가운데 기존 결과물의 완성도가 우수한 인원을 중심으로 작업자를 선발하였다. 또한 모국어 구사 능력을 충분히 활용할 수 있는 신규 작업자도 함께 발굴하였다.

고품질 번역 수행 역량을 검증하기 위해 테스트를 실시하였으며, 번역 지침 준수 여부와 번역 완성도를 종합적으로 평가하고 그 결과를 바탕으로 신규 작업자를 최종 선정하였다.

O	R	S	T	U	V
12. 둘 다 결혼 생각이 없다고 그렇게 한 건데 요즘은 30대 초반까지만 하더라도 다들 결혼 생각이 없더라고요.	잠수	피드백	13. 세우짓 안 짜게 알라고 막 거기애다가 양파 썰어 놓고 마늘 썰어 놓고 고춧가루 섞고 막 고추 다져서 섞고 그러잖아.	잠수	피드백
<p>4.5</p> <p>의미 부정화 - nhưng mọi người cũng(반 되었는데) --&gt; là mọi người đã(반 하더라요) - 없다고 그렇게 한 건데에서 '결혼은 한 건데로 잘 해석은 하였음'</p> <p>Cả hai vốn đều không có ý định kết hôn rồi mới lấy nhau mà dạo này dù mới đến đầu độ tuổi 30 thôi nhưng mọi người cũng chẳng có ý định kết hôn đâu.</p>	4.5	의미 부정화 - nhưng mọi người cũng(반 되었는데) --> là mọi người đã(반 하더라요) - 없다고 그렇게 한 건데에서 '결혼은 한 건데로 잘 해석은 하였음'	Em định làm mắm tép không mặn nên mới thái hành tây bỏ vào đó, rồi thái tỏi bỏ vào, trộn với ớt bột, băm ớt rồi trộn vào mà	3	의미 부정화 - Em định(제가 -알라고) 첨가 힘 - rồi thái tỏi(그리고 넣고) --> và thái tỏi(넣고) 가독성이 다소 좋지 못하나 구어체는 반영 됨
<p>4</p> <p>누락 - 요즘은(hi)' 부사 누락 '그렇게 한 건데' 지역을 하여 형식을 잘 맞출</p> <p>Cả hai vốn đều không có ý định kết hôn rồi mới lấy nhau thế, dạo gần đây người ta chỉ mới bước vào độ tuổi là đã không có ý định kết hôn rồi</p>	4	누락 - '그렇게 한 건데' 누락 문장 구조 변형 - '모든 30대 초반이 결혼 생각을 가지고 있는 것은 아니다.' - 30대 초반 사람들 --> 모두가 30대 초반까지는	Chẳng phải câu củ liên tục thái hành tây với tỏi bỏ vào, cho thêm bột ớt, thái củ quả ớt cho vào là để cho mắm tép không mặn còn gì	3	의미 변형 - 주제 오류: Chẳng phải câu(나는 -그러잖아) --> củ liên tục(시속적으로) --> thường bỏ vào(되는데로) 구어체 반영 하였으며 자연스럽게 문장 구성하려고 함
<p>3</p> <p>누락 - '그렇게 한 건데' 누락 문장 구조 변형 - '모든 30대 초반이 결혼 생각을 가지고 있는 것은 아니다.' - 30대 초반 사람들 --&gt; 모두가 30대 초반까지는</p> <p>Cả hai đều không nghĩ đến việc kết hôn, mà dạo gần đây tôi thấy mọi người ở độ tuổi đầu 30 đều có suy nghĩ kết hôn</p>	3 <p>누락 - '그렇게 한 건데' 누락 문장 구조 변형 - '모든 30대 초반이 결혼 생각을 가지고 있는 것은 아니다.' - 30대 초반 사람들 --&gt; 모두가 30대 초반까지는</p>	Để mắm tép không bị mất thì mọi người thường thái hành tây bỏ vào, rồi thêm tỏi thái mỏng, bột ớt và ớt băm trộn lẫn.	3.5 <p>문장 구조 변형 - 'bột ớt và ớt băm trộn lẫn(새로운 섞는다) --&gt; rồi trộn lẫn bột ớt và ớt băm (-섞고 재료들 섞고) - 구어체 미반영: 문미 mà(chứ gì)mởa 누락 자연스럽게 문장 반영 함</p>	3.5 <p>문장 구조 변형 - 'bột ớt và ớt băm trộn lẫn(새로운 섞는다) --&gt; rồi trộn lẫn bột ớt và ớt băm (-섞고 재료들 섞고) - 구어체 미반영: 문미 mà(chứ gì)mởa 누락 자연스럽게 문장 반영 함</p>	
<p>3</p> <p>의미 변형 - 부정문 어미 '그렇게 한 건데' 누락 - 둘 다 결혼 생각이 없는데 다들 역시 결혼 생각이 없다는 것으로 해석하였음</p> <p>Cả hai đều không nghĩ đến việc kết hôn, mà dạo gần đây tôi thấy mọi người dù chỉ mới bước vào độ tuổi 30 cũng đều không có ý định kết hôn.</p>	3 <p>의미 변형 - 부정문 어미 '그렇게 한 건데' 누락 - 둘 다 결혼 생각이 없는데 다들 역시 결혼 생각이 없다는 것으로 해석하였음</p>	Để mắm tép không mất, người ta thái hành tây và tỏi bỏ vào đó, rồi trộn với bột ớt, và băm ớt tươi trộn cùng	4 <p>구어체 미반영 - người ta --&gt; người ta thường(lhay -chứ gì) 문장의 가독성이 좋음</p>	4 <p>구어체 미반영 - người ta --&gt; người ta thường(lhay -chứ gì) 문장의 가독성이 좋음</p>	

[그림 12] 번역 작업자 선정 테스트 예시(베트남어)

### 3.2. 번역 인력 교육

선정된 작업자를 대상으로 번역 지침과 작업 수행과 관련된 교육을 실시하여 본 사업에서 요구하는 번역 품질 기준을 충실히 준수할 수 있도록 하였다. 번역 인력에 대한 교육은 경희대학교 산학협력단에서 마련한 한국어-외국어 번역 공통 지침과 언어별 세부 지침을 기반으로 진행하였다.

The image displays educational content for translators, divided into three main sections:

- Left Panel:** A document titled 'CONFIDENTIAL' containing Korean text and a table with columns for '한국어' (Korean), '영어' (English), and '번역' (Translation). The text discusses translation guidelines and quality control.
- Middle Panel:** A section titled '2. Human-Translation' showing a screenshot of a translation tool interface. It includes instructions like 'Please translate the sentence below' and 'Check the Source text and write the Translation text'. Red arrows point to the 'Source' and 'Translation' input fields.
- Right Panel:** A section titled '3. Proofreading' showing another screenshot of the translation tool interface. It includes instructions like 'Compare the Source and Translation text and evaluate the Translation text' and 'Select either Bad, Good or Perfect based on the Translation quality'. Red arrows point to the evaluation options.

[그림 13] 번역 및 작업 지침 교육 내용 예시

### 3.3. 번역 절차

번역과 교정은 다음과 같은 내용으로 진행되었다.

<표 14> 번역 및 교정 절차

구분	번역	교정
주체	현지 프리랜서 및 전문 번역사	현지 프리랜서 및 전문 번역사
방식	제공된 번역 지침에 따라 원문-번역문 문장 쌍 생성	제공된 번역 지침에 따라 원문-번역문 문장 검토 및 교정
내용	표기, 어휘 및 문법 등을 종합 고려하여 번역	어색한 지역, 오류문을 수정하여 자연스러운 표현으로 교정
납품	(주)플리토 플랫폼 실시간 제출	(주)플리토 플랫폼 실시간 제출
수정·보완 의무	(주)플리토 작업 관리자의 수정·보완 지시에 따라 의무 수행	

### 3.4. 번역 데이터 구축 환경

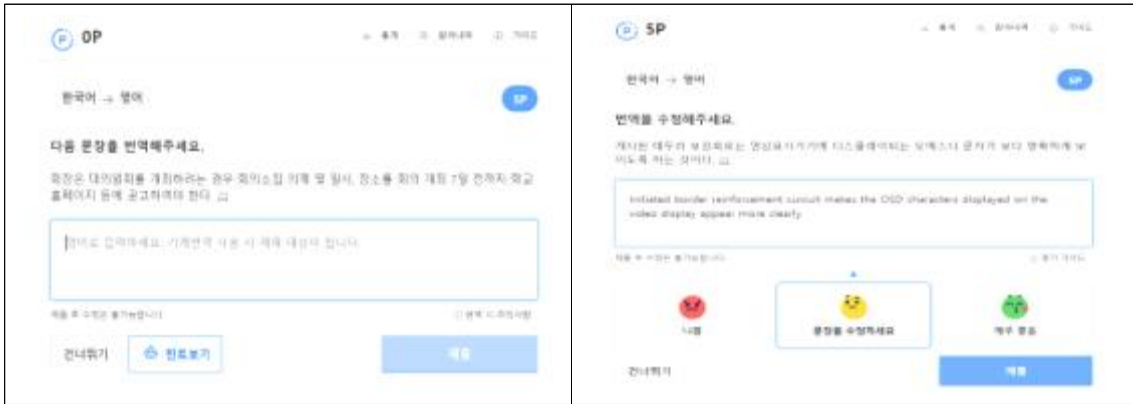
#### 1) 번역·교정

본 사업에서는 번역 품질을 제고하기 위해 전문 번역사들이 (주)플리토 번역 플랫폼을 활용하여 번역 및 교정 작업을 진행하였다.

교정에 참여한 작업자들은 해당 언어권 국가의 학위 취득자이거나 현지 출생·거주 경험을 보유하고 관련 통번역 실무 경험이 풍부한 경력자로 구성하였으며, 번역 작업자보다 높은 수준의 언어 능력을 필수 요건으로 하여 업무를 수행하도록 하였다. 이들 작업자의 주요 업무는 다음과 같다.

<표 15> 번역 및 교정 작업자 주요 업무

- |   |
|---|
| <ul style="list-style-type: none"> <li>• 해당 언어별 번역 또는 교정(작업자별 개별 계정 부여)</li> <li>• (주)플리토 플랫폼을 통해 문장 단위의 데이터를 실시간 제출</li> <li>• 완료 산출물에 대한 보완 지시 사항 대응</li> </ul> |
|---|



[그림 14] 번역(좌) 및 교정(우) 환경 예시

## 2) 1차 검수(7% 표본 검수)

전문 번역사들의 번역·교정 결과물에 대한 품질을 안정적으로 관리하기 위해 도착어(외국어)를 모국어로 하는 화자와 도착어 구사가 가능한 한국인 번역사를 작업 관리자로 채용하였다. 작업 관리자는 (주)플리토 플랫폼(또는 엑셀 파일 형식)을 활용하여 실시간으로 생성되는 산출물을 1차적으로 검수하고 작업자별 작업량과 정확도 등을 종합적으로 관리하였다. 작업 관리자의 주요 역할은 다음과 같다.

<표 16> 1차 검수 작업자 주요 업무

- 일 단위 번역·검수 완료 산출물에 대한 1차 검수(7% 표본 검수)
- (주)플리토 플랫폼(혹은 엑셀 파일)을 통해 작업자별 작업 달성도 추적·관리
- 완료 산출물에 대한 개선 필요 사항 전달 및 보완 지시

No.	User	Event Description	Create Date	Status	Update Date	Event Assess Ratio	QC Count	Assessment Count	Assessment Rate	Assessment Pass Rate
1	V01EX_PF_0001	417[Internal][Exptranso][ko-kr]NRL.LI	2022-09-07 19:02:18	Deactivate	2022-10-13 13:07:28	10	831	64	7.7%	96.9%

SID	Origin	QC Status	TR User Result	QC User Result	Admin User Result	Create Date	Update Date	Comment	Assess	Update	Assessment Date	Assessment User	Done
10365103	(Q) 기상청은 14일 제주도 인근 해역에서 발생한 규모 4.9 지진의 여진이 앞으로 1년 정도 이어질 수 있다며 각별한 주의를 당부했다.	🟢	Tổng cục Khí tượng Hàn Quốc ngày 14 đã phát đi cảnh báo đặc biệt rằng dư chấn của trận động đất mạnh 4.9 độ richter tại vùng biển gần đảo Jeju có thể kéo dài trong khoảng 1 năm.	(Last QC) Tổng cục Khí tượng Hàn Quốc ngày 14 đã phát đi cảnh báo đặc biệt rằng dư chấn của trận động đất mạnh 4.9 độ richter tại vùng biển gần đảo Jeju có thể kéo dài trong khoảng 1 năm.	Tổng cục Khí tượng Hàn Quốc ngày 14 đã phát đi cảnh báo đặc biệt rằng dư chấn của trận động đất mạnh 4.9 độ richter tại vùng biển gần đảo Jeju có thể kéo dài trong khoảng 1 năm.	2022-10-13 00:42:02	2022-10-18 15:54:46		P	Cancel	2022-10-18 15:54:46	fliito_L05	✅

[그림 15] (주)플리토 플랫폼의 표본 검수 환경

### 3.5. 번역 품질 향상 활동

본 사업에서는 번역 품질의 향상을 위해 다음과 같은 활동을 실시하였다.

#### 1) 번역 전

우선 번역 및 교정 작업자들에게 공통 번역 지침과 언어별 번역 지침의 핵심 사항 및 유의점을 사전에 안내하였다. 작업자들은 이를 숙지한 후 번역을 시작하였으며, 작업 과정에서 도착어별 특수 사항이 발견되거나 추가 주의가 필요하다고 판단되는 경우 지침을 수시로 보완·갱신하였다.

아울러 번역 작업을 시작하기 전 테스트와 간단한 교육을 실시하였고, 테스트 결과가 내부 품질 기준에 미달한 작업자는 번역 작업에서 배제하여 품질 저하를 사전에 예방하였다.

#### 2) 번역 후

번역이 완료된 이후에는 1차 번역을 수행한 작업자보다 높은 숙련도를 갖춘 작업자를 선별하여 번역문에 대한 교정을 진행하였다. 이 과정에서 단어 사용상의 오류와 어색한 직역 표현 등을 중심으로 수정하여 전반적인 번역 품질을 향상시켰다.

또한 전문 번역사들의 번역·검수 결과물에 대한 품질을 안정적으로 확보하기 위해 언어별 모국어 화자와 해당 언어 구사가 가능한 한국인 번역사를 작업 관리자로 채용하여 운영하였다. 한국인 번역사는 모국어 화자와 협업하며 한국어 원문과 한국 문화 전반에 대한 이해도를 바탕으로 보다 정확한 번역이 이루어질 수 있도록 지원하였다. 또한 작업 관리자는 병렬 말뭉치 구축 과정 전반에서 작업자별 작업량과 정확도를 관리하는 한편, 번역 진행을 위한 교육과 훈련도 함께 수행하였다.

아울러 번역 지침이나 언어 특성에 대한 이해가 미흡한 작업자에 대해서는 수시로 수정·보완을 요구하였으며, 세 차례 이상의 보완 이후에도 품질 개선이 이루어지지 않는 경우에는 번역 및 교정 작업을 중단하였다.

또한 번역문에 포함된 경미한 오류는 기계적 정제 과정을 통해 기호 유형을 일괄적으로 통일하고 불필요한 문자 등을 제거하여 정비하였다. 기계적 정제의 세부 내용은 다음과 같다.

<표 17> 번역 품질 향상을 위한 기계적 정제 내용



데이터의 형식적 일관성과 구조적 안정성을 확보하기 위해 (주)플리토 플랫폼과 엑셀 파일로 처리된 번역 파일에 대해 구조적 적합성 검증을 실시하였으며, 해당 검증에서 기준을 충족하지 못한 문장은 작업자가 다시 번역 및 교정하도록 조치하였다. 구조적 검증의 세부 기준과 불합격에 해당하는 사례는 다음과 같다.

<표 18> 번역 품질 향상을 위한 구조적 검증 기준

검증 기준	불합격 기준
중복	원문과 번역문, 서로 다른 번역문이 중복되는 경우
텍스트	원문 혹은 번역문이 누락된 경우, 번역문이 숫자 및 특수 문자로만 구성된 경우
비율	원문 길이와 번역문 길이의 비율이 극히 차이 나는 경우
언어	도착어 외의 언어가 들어간 경우
기계 번역 유사도	번역문과 기계 번역기의 번역문의 유사도가 90% 이상인 경우 ※ 단문의 경우, 전문 검수자(교정 작업자)가 재검증

### 3) (추가 제안) 영어 데이터 품질 향상 활동

추가로 제안한 한국어-영어 병렬 말뭉치는 (주)플리토에서 자체적으로 품질 향상 활동을 진행하였으며, 그 내용은 다음과 같다.

<표 19> 영어 데이터 품질 향상 활동 내용

단계	수행 내용	세부 사항
1	번역문에 대한 전수 교정(proofreading)	- 번역문 전체 대상 문법 오류, 오역 등을 교정 - 부적합 시 작업자에게 재수정 요청
2	기계 번역 유사도 및 기계적 검증	- 번역문과 번역기의 유사도 확인 - 부적합 내용, 중복 내용 등 기계적 처리 가능 부분 처리 - 부적합 시 작업자에게 재수정 요청

3	내부 언어 전문가 전수 검수	<ul style="list-style-type: none"> <li>- 전수 검수와 기계적 검증을 통과한 데이터를 도착어 기준 모국어 화자인 (주)플리토 내부 언어 전문가가 전수 검수 진행</li> <li>- 어색한 표현, 오타자, 문장 구조 등을 교정</li> <li>- 국내에서 가장 많이 사용되는 미국식 영어 표현을 위주로 수정·보완하여 번역문에 통일성을 부여</li> <li>- 부적합 시 작업자에게 재수정 요청</li> </ul>
---	--------------------	--

## 4. 검수 및 감수

### 4.1. 번역 검수팀

#### 1) 구성

본 사업의 번역 검수 인력은 모국어와 한국어에 모두 능숙한 이중 언어 화자로서 한국어를 정확히 이해하고 모국어로 번역할 충분한 능력을 갖추었다. 따라서 외국어 번역 과정에서 모국어 화자 수준의 숙달도가 아니면 판단하기 어려운 미세한 오류들도 효과적으로 방지할 수 있었다.

번역 검수팀은 팀의 총괄 관리와 검수원 교육을 담당하는 검수팀장, 전수 검수를 수행하는 번역 검수원, 한국어 원문 이해를 지원하는 보조 연구원으로 구성되었으며, 2021년부터 2024년까지 사업 경험이 있는 검수원을 우선 선발하고 필요시 번역 및 검수 테스트를 통해 추가 인력을 확보하였다<sup>1)</sup>.

팀 구성원은 국내 대학의 석박사 과정 대학원생, 해외 대학의 한국어 관련 전공 석박사 과정 대학원생 수준의 전문 인력으로 구성되었으며, 문법적 정확성뿐 아니라 맥락적 자연스러움까지 검수할 수 있는 전문성을 갖추었다.

<표 20> 언어별 번역 검수원 인원 현황

언어	베트남어	인도네시아어	태국어	인도 힌디어	캄보디아 크메르어	필리핀 타갈로그어	러시아어	우즈베크어	계
착수 인원	10	10	10	15	17	15	15	15	107
추가 인원	-	2	-	1	-	3	-	-	6
총참여 인원	10	12	10	16	17	18	15	15	113

1) 일부 언어권의 경우 검수원의 개인 사정으로 인한 인력 변동이 발생하여 검수 진척률 관리가 필요해짐에 따라 번역 및 검수 테스트를 통해 추가 검수 인력을 확보하였다.

## 2) 직책별 역할

검수팀장은 번역 검수원의 전수 검수 결과 중 20%를 표본으로 3차 검수하여 검수원별 오류 유형과 오류율을 분석하고 보고하였다. 이러한 표본 검수 결과는 즉시 검수원과 공유되었으며, 오류율이 높은 검수원에 대해서는 재교육을 실시하고 개선되지 않을 경우 교체하였다. 또한 검수 과정에서 검출된 오류를 분석하고 유형화하여 향후 재발 방지 교육 및 지침 보완의 근거로 활용하였다.

번역 검수원은 문법, 어휘, 표기법 등의 번역 오류를 검수 지침에 따라 수정하였으며, 한국어 원문의 맥락과 상황을 면밀히 고려하여 정확하고 자연스럽게 번역되도록 다듬는 역할을 수행하였다. 원문 이해를 바탕으로 번역문이 지침에 부합하는지 점검하고, 전문용어, 신조어 등 논의가 필요한 항목은 팀 내 협의를 거쳐 최종 검수문에 반영하였다. 또한 검수팀장과의 원활한 소통을 통해 전체 검수 품질 향상을 위해 노력하였다.

번역 검수원의 전수 검수와 검수팀장의 3차 검수를 병행함으로써 초기 데이터의 품질을 안정적으로 확보할 수 있었고, 지침 숙지가 미흡한 신규 검수원의 검수 정확도 또한 향상되었다. 이러한 병행 구조는 검수 과정의 효율성을 높이는 데에 기여하였다.

## 4.2. 감수자

경희대학교 산학협력단은 2021년부터 2024년까지의 네 차례 사업을 수행하며 쌓은 인적 네트워크를 통해 8개 국가의 언어 전문가 풀을 구축하고 긴밀한 협력 관계를 유지하고 있다. 이를 활용하여 도착어를 모국어로 사용하고 한국어 실력이 뛰어난 해외 대학교수 수준의 언어 전문가로 감수팀을 구성하였다. 감수자는 전체 구축 수량의 10%를 무작위로 추출하여 감수를 진행하였다.

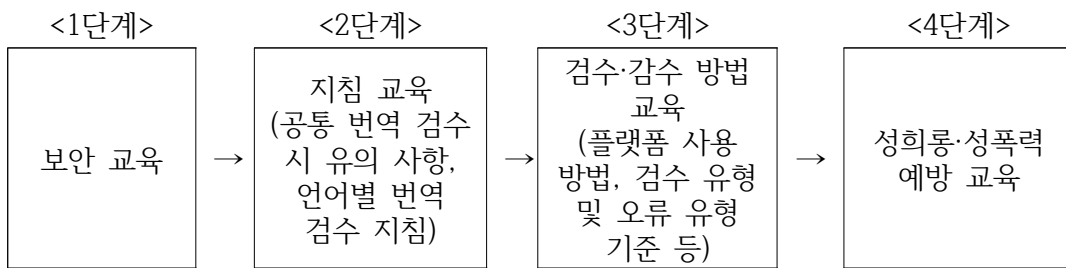
감수자로는 베트남의 호찌민시국립대학교 인문사회과학대학교, 베트남외교아카데미, 인도네시아의 인도네시아국립대학교, 가자마다대학교, 태국의 부라파대학교, 송클라대학교, 인도의 자와할랄네루대학교, 캄보디아의 왕립프놈펜대학교, 필리핀의 필리핀국립대학교, 러시아의 모스크바국립외국어대학교, 우크라이나의 우신스키국립사범대학교, 우즈베키스탄의 타슈켄트국립동방대학교 등의 아시아 및 러시아 주요 대학에 재직 중인 교수가 참여하였다.

<표 21> 언어별 감수자

언어	베트남어	인도네시아어	태국어	인도 힌디어	캄보디아 크메르어	필리핀 타갈로그어	러시아어	우즈베크어	계 <sup>2)</sup>
총참여 인원	2	3	2	2	1	1	2	2	15

### 4.3. 번역 검수팀 및 감수자 교육

사업의 보안을 유지하고 번역문의 품질과 통일성을 제고하기 위해 번역 검수원, 검수팀장, 감수자를 대상으로 다음과 같은 4단계 교육을 실시하였다.



[그림 16] 검수·감수 교육 절차

#### 1) 보안 교육

먼저 검수 및 감수 인력 전원을 대상으로 보안 교육을 실시함으로써 데이터 보안에 각별한 주의를 기울이도록 하여 보안 사고를 미연에 방지하였다. 보안 교육 내용은 교육부 산하 정보보호교육센터에서 제공하는 동영상 교육 자료로 온라인 플랫폼에서 진행하였으며 내용과 예시는 다음과 같다.

<표 22> 검수·감수 인력 보안 교육 내용

1) 1차시: 최근 정보 유출 이슈 및 동향 2) 2차시: 내부정보 보안의 중요성 및 유출 개념 3) 3차시: 내부정보 유출 경로 및 주요 사례 4) 4차시: 내부정보 유출 예방 및 유출 시 대응 방법
---

2) 감수 인력은 사업 기간 동안 추가 투입 없이 초기 구성 인원이 그대로 유지되었다.



[그림 17] 검수·감수 보안 교육 예시

## 2) 지침 교육

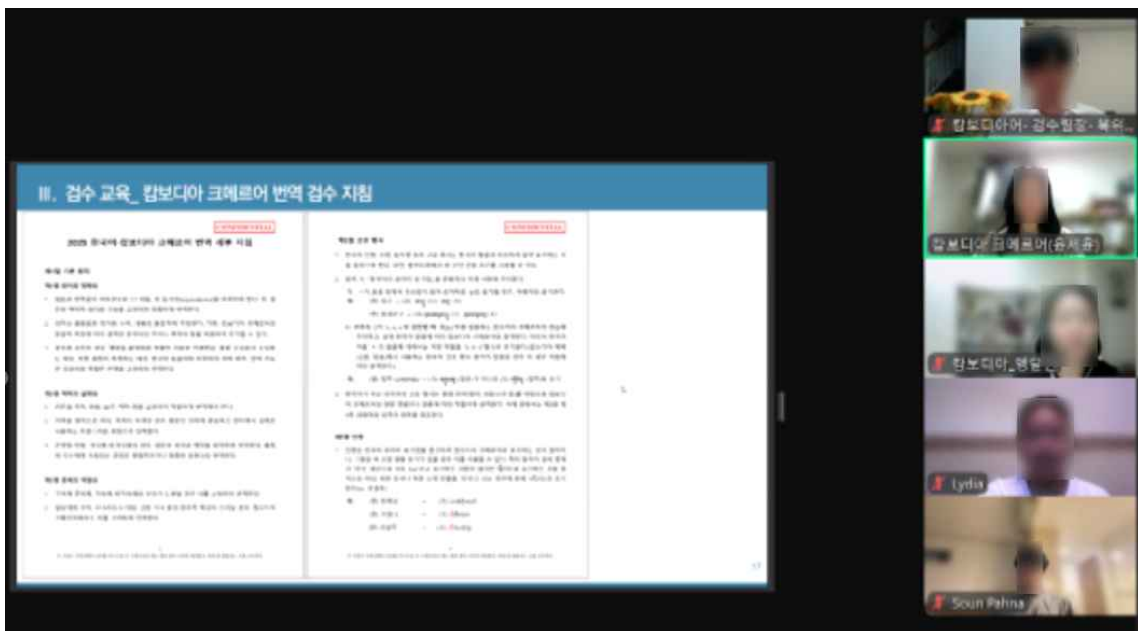
보안 교육 후 지침 교육을 진행하여 사업 수행 인력의 전문성을 함양하였다. 지침 교육은 공통 교육 내용과 언어별 지침 내용으로 나뉘어서 실시하였다. 공통 교육 내용으로는 사업의 전체적 개요와 한국어 원문의 특징 등을 교육하였고, 언어별 지침 내용으로는 번역 검수 시 따라야 할 지침 및 언어별로 수정된 지침을 교육하였다. 지침에 대한 교육을 통하여 효율적이고 체계적인 절차를 준수함으로써 일관성과 신뢰성이 높은 말뭉치를 구축할 수 있도록 하였다.

<표 23> 지침 공통 교육 내용

항목	교육 내용
사업 소개	<ul style="list-style-type: none"> <li>• 사업 개요</li> <li>• 데이터 구축 절차</li> </ul>
한국어 원문의 특징 및 번역 검수 시 유의 사항	<ul style="list-style-type: none"> <li>• 한국어 원문의 특징</li> <li>• 문어 번역 검수 시 유의 사항</li> <li>• 구어 번역 검수 시 유의 사항</li> </ul>

<표 24> 언어별 지침 내용

항목	지침 내용
내용 검수	<ul style="list-style-type: none"> <li>• 내용 일치 여부 검수</li> <li>• 오역 및 누락 내용 검수</li> <li>• 문장 구조 및 문법·관용구·어휘 일치 검수</li> <li>• 언어별 번역 특이 사항</li> </ul>
표기 검수	<ul style="list-style-type: none"> <li>• 문장 부호 및 특수 문자 일치 검수</li> <li>• 오타자 및 띄어쓰기 오류 검수</li> <li>• 언어별 번역 특이 사항</li> </ul>



[그림 18] 지침 교육 예시(캄보디아 크메르어)

### 3) 검수·감수 방법 교육

#### (1) 검수 방법 교육

검수 인력을 대상으로 검수 플랫폼에 대한 교육을 진행하였다.



[그림 19] 검수 플랫폼 작업 환경 예시(인도네시아어)

검수 플랫폼은 문장 번호(SID), 원문(Origin Text), 번역문(Translation), 검수(QC), 오류 유형(QC Reason), 검수문(QC Translation), 제출(Submit), 보류(Pending)로 구성되어 있다. 번역 검수원들은 플랫폼에서 원문(Origin Text)과 번역문(Translation)을 보고 검수(QC) 유형을 판단한 후에 그에 따라 검수 작업을 수행하였다. 검수(QC) 유형은 좋음(Perfect), 보통(So so), 나쁨(Bad)으로 구성되며, 별도의 수정이 필요 없는 경우 좋음(Perfect), 사소한 오류로 수정이 필요한 경우 보통(So so), 오류가 많아 재번역이 필요한 경우 나쁨(Bad)을 선택하도록 하였다. 또한 보류(Pending) 기능은 원문이나 검수에 대한 질의가 있을 시에, 잠시 보류할 수 있는 기능으로 원문에 대한 이해와 검수의 편의성을 높였다.

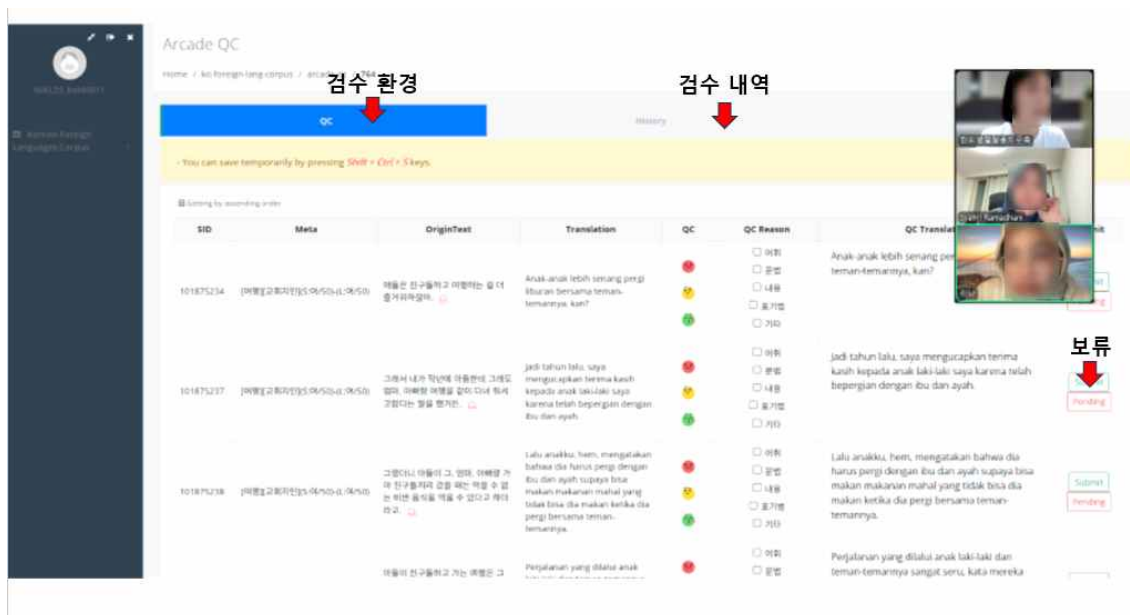
<표 25> 검수 유형별 검수 방법

검수 유형		내용
좋음 (Perfect)		· 번역문의 수정이 필요 없을 시 선택함. · 별도 수정 없이 검수문을 제출함.
보통 (So so)		· 사소한 오류가 있을 시 선택함. · 알맞은 오류 유형(QC Reason)을 선택한 후 검수문(QC Translation)을 작성하고 제출함.
나쁨 (Bad)		· 오류가 많거나 심각한 오류가 있을 시 선택함. · 오류 유형(QC Reason)만 클릭하고 검수문(QC Translation)은 작성하지 않고 제출함. · 나쁨(Bad)으로 선택한 문장은 재번역함.

번역문에 오류가 있을 시(보통, 나쁨) 선택하는 오류 유형(QC Reason)의 기준은 어휘, 문법, 내용, 표기법, 기타로 구분하였다.

<표 26> 오류 유형 기준

오류 유형	내용
어휘	· 어휘 단위에서 오류가 있을 경우
문법	· 문법(어순, 시제, 상, 수동태 등)과 관련된 오류가 있을 경우
내용	· 원문의 내용을 잘못 번역하였거나 불필요한 내용이 첨가된 경우, 즉 오역이 된 경우
표기법	· 오타자 및 문장 부호 등 표기와 관련된 오류가 있을 경우
기타	· 문법적으로 오류가 없으나 지나친 직역으로 인해 어색한 표현인 경우



[그림 20] 검수 플랫폼 사용 교육 예시(인도네시아어)

## (2) 감수 방법 교육

모든 감수 인력을 대상으로 개인 계정을 지급하여 본 사업에 최적화된 플랫폼에서 감수를 실시하였다. 원문(Origin)과 검수문(QC user result)을 보고 감수 결과(Assess)를 결정하였는데, 검수문에 오류가 없으면 P, 오류가 있으면 F를 선택하며 감수 결과가 F일 경우 검수문을 감수자가 직접 수정하고 의견(comment)란에 수정 이유를 작성하였다.

Event: 776. [Private][ko-kr] 2025 NIKL 문어제 | Status: ALL

[View Statistics](#) [Download Corpus](#)

6 results found

No.	User	Event Description	Create Date	Status	Update Date	Event Assess Ratio	QC Count	Assessment Count	Assessment Rate	Assessment Pass Rate
1	NIK125_kot0002	776. [ko-kr] 2025 NIKL 문어제	2025-04-16 17:16:30	Active	2025-10-14 19:43:15	10	507	55	10.8%	100.0%
2	NIK125_kot0004	776. [ko-kr] 2025 NIKL 문어제	2025-04-16 17:16:30	Active	2025-10-12 12:20:04	10	818	90	11.0%	100.0%
3	NIK125_kot0005	776. [ko-kr] 2025 NIKL 문어제	2025-04-16 17:16:30	Active	2025-10-12 23:32:18	10	160	16	10.0%	100.0%

Origin	QC -> Admin User Result	Comment	Assess	Update	Details
[문항]이날 오후 3시까지 금감원에 이와 관련해 접수된 민원본 총 109건이다.	Unabot sa kabuang 109 na reklamo ang nahain sa Financial Supervisory Service hanggang alas tres ng hapon ng araw na ito.		P	Cancel	SID: 2028120010 (Assmt By / Date) NIK125_kot_0002 / 2025-11-12 10:29
[부동산]계약금 입금 후 전자 계약 링크가 이메일로 발송되며 계약 완료 여부도 확인 가능하다.	Matapos ang pagbayad ng downpayment, isang link para sa elektronikong kontrata ang ipapadala sa panamagitan ng email at maari ring kumpirmahin kung netapos na ang kontrata.		P	Cancel	SID: 2027010014 (Assmt By / Date) NIK125_kot_0002 / 2025-11-12 10:29
[사회]올해는 취약 계층 교육과 '고령 업무 경감' 등 2개 분야에서 각각 2~3개 프로젝트를 공모로 선정해 2년간 최대 3억 6000만 원을 지원할 계획이다.	Sa taong ito, pinaplano nitong magbigay ng 360 milyong won sa loob ng dalawang taon sa panamagitan ng paggiling ng dalawa hanggang ratlong pangkat ng proyekto sa dalawang larangan ng "Edukasyon para sa mga Mahihirap"		P	Cancel	SID: 2009970015 (Assmt By / Date) NIK125_kot_0002 / 2025-11-12 10:29

[그림 21] 감수 작업 환경 예시(필리핀 타갈로그어)

No.	User	Event Description	Create Date	Status	Update Date	Event Assess Ratio	QC Count	Assessment Count	Assessment Rate	Assessment Pass Rate
1	NIK125_kot0001	776. [ko-kr] 2025 NIKL 문어제	2025-04-16 17:16:30	Active	2025-05-09 10:36:41	10	75	6	8.0%	0.0%

Origin	QC -> Admin User Result	Comment	Assess	Update	Details
[문항] 해당 항목을 확인하여(다) 1,000건 이상만 일한 채가 채워지실 수 있습니다. (참고)	Akan tetapi, hanya dalam waktu sebentar dia berhasil melihat, terganggu. //		P	Update	SID: 100547019
[문항] 해당 항목을 확인하여(다) 1,000건 이상만 일한 채가 채워지실 수 있습니다. (참고)	Dari antara orang yang saya temui belakangan ini, ada seorang wanita paruh baya berusia 40-an tahun dan dia ingin melakukan platex untuk hair. //		P	Update	SID: 100391124
[문항] 해당 항목을 확인하여(다) 1,000건 이상만 일한 채가 채워지실 수 있습니다. (참고)	Hein, mungkin sama, saat membuat undangan pernikahan itu yang ditanyakan siapa di atasnya, siapa yang beres-meresnya, bagian-bagian pernikahan sudah selesai atau belum, ya? //		P	Update	SID: 100767119
[문항] 해당 항목을 확인하여(다) 1,000건 이상만 일한 채가 채워지실 수 있습니다. (참고)	Akan tetapi, setelah melakukan pemeriksaan yang lebih detail, saya sadar bahwa saya pun seperti ada-ada masalah paruh baya lalu 30-an di Korea. //		P	Update	SID: 100230712
[문항] 해당 항목을 확인하여(다) 1,000건 이상만 일한 채가 채워지실 수 있습니다. (참고)	Sepertinya aku cukup sering menonton drama bertajuk seperti.		P	Update	SID: 100771126
[문항] 해당 항목을 확인하여(다) 1,000건 이상만 일한 채가 채워지실 수 있습니다. (참고)	Ya, setelah saya pikir-pikir cukup saya sama sekali tidak melihat bisa-bisa bahasa itu.		P	Update	SID: 100747045

[그림 22] 감수 플랫폼 사용 교육 예시(인도네시아어)



SID	Word Count	Origin	Translation	QC Translation	제검수 수정 문장	피드백	P/F	
2013560002	16	28일 오전 2시 20분 서울시 버스 노선조합은 서울시 서울시 버스 노선사업 조합과의 협상 결실을 선언했다.	Nagdeklara ng pagtapos ng negosasyon ang unyon ng manggagawa ng mga bus sa Seoul noong 2:20 ng umaga sa ika-28, matapos hindi magkasundo sa panig ng Seoul Bus Transportation Business Association.	Sa 2:20 am noong ika-28, inihayag ng panig ng Seoul Association of Chartered Bus Transportation ang hindi pagkakasundo sa mga negosasyon sa namamahaling panig na Seoul Automobile & Transport Workers' Federation.	Sa 2:20 ng umaga noong ika-28, inihayag ng panig ng Seoul Bus Union ang hindi pagkakasundo sa mga negosasyon sa namamahaling panig na Seoul Bus Transportation Business Association.	<a href="http://www.busunion.or.kr/">http://www.busunion.or.kr/</a> <a href="https://en.namu.wiki/w/%EC%A0%84%EA%85%AD%E8%B2%84%EC%8A%A4%EC%9A%B4%EC%88%A1%EC%82%AC%EC%97%85%EC%A1%80%ED%95%A9%EC%97%80%ED%95%A9%ED%9A%8C">https://en.namu.wiki/w/%EC%A0%84%EA%85%AD%E8%B2%84%EC%8A%A4%EC%9A%B4%EC%88%A1%EC%82%AC%EC%97%85%EC%A1%80%ED%95%A9%EC%97%80%ED%95%A9%ED%9A%8C</a>	P	
2014320017	23	[캘러리 웨]는 실크 스크린을 실 목판 적용으로 주목받은 (고스)와 (하말릭)의 작품, (캘러리 BHAK)은 지문순공순돈의 수필을 모티브로 삼아 정체성을 시각화하는 (순재)와 (민 캅)의 작품을 선보였다.	ipinakita ng Gallery Wee ang mga gawa nina Gose at Heo Pil-seok na nakatanggap ng atensyon para sa kanilang mga gawang silk screen-based, habang ipinakita ng Gallery BHAK ang mga gawa nina Sunjae at Min Kim, na nagpaapakita ng pagkakakilanlan gamit ang mga fingerprint, palm print at wrinkles bilang motif sa likod.	ipinakita ng Gallery Wee ang mga gawa nina Gosee at Heo Pil-seok na nakatanggap ng atensyon para sa kanilang mga gawang base sa iskrin na seda, habang ipinakita ng GALERIE BHAK ang mga gawa nina Sunjae at Min Kim, na nagpaapakita ng pagkakakilanlan gamit ang mga tatak-daliri, imprenta ng palad at kulubot bilang tema Ito ang pangatlong desisyon na tanggihan ang isang kahilingan para sa pagsuspensiyon ng pagpapatupad, kasunod ng pagtanggap sa kahilingan ng Korea Association of Medical College Professors (KAMCP), mga residente, mga medikal na estudyante, at mga kumukuha ng pagsusulit.	ipinakita ng Gallery Wee ang mga gawa nina Gosee at Heo Pil-seok na nakatanggap ng atensyon para sa kanilang mga gawang base sa iskrin na seda, habang ipinakita ng GALERIE BHAK ang mga gawa nina Sunjae at Min Kim, na nagpaapakita ng pagkakakilanlan gamit ang mga fingerprint, imprenta ng palad at kulubot bilang tema Ito ang pangatlong desisyon na tanggihan ang isang kahilingan para sa pagsuspensiyon ng pagpapatupad, kasunod ng pagtanggap sa kahilingan ng Korea Association of Medical College Professors (KAMCP), mga residente, mga medikal na estudyante, at kumukuha ng pagsusulit.	ipinakita ng Gallery Wee ang mga gawa nina Gosee at Heo Pil-seok na nakatanggap ng atensyon para sa iskrin na seda, habang ipinakita ng GALERIE BHAK ang mga gawa nina Sunjae at Min Kim, na nagpaapakita ng pagkakakilanlan gamit ang mga fingerprint, imprenta ng palad at kulubot bilang tema Ito ang pangatlong desisyon na tanggihan ang isang kahilingan para sa pagsuspensiyon ng pagpapatupad, kasunod ng pagtanggap sa kahilingan ng Korea Association of Medical College Professors (KAMCP), mga residente, mga medikal na estudyante, at kumukuha ng pagsusulit para sa suspensiyon ng pagpapatupad.	kwf 사진에도 나왔음	P
2014440021	17	방중 2일 차인 5일 (월)날 장군은 경계 전보게통을 만나 중국 경제의 어려움과 기회를 논의할 것으로 전해졌다.	Sa ikalawang araw ng kanyang pagbisita sa Tsina, sa ika-5, inaasahang makikipagpulong si ang kalihim na si Yellen sa mga eksperto sa ekonomiya upang talakayin ang mga hamon at oportunidad na kinakaharap ng ekonomiya ng Tsina.	Sa ikalawang araw ng kanyang pagbisita sa Tsina, sa ika-5, inaasahang makikipagpulong si ang kalihim na si Yellen sa mga eksperto sa ekonomiya upang talakayin ang mga hamon at pagkakataong kinakaharap ng ekonomiya ng Tsina.	Sa ikalawang araw ng kanyang pagbisita sa Tsina, sa ika-5, naiparing na haharapin ng kalihim na si Yellen ang mga eksperto sa ekonomiya upang talakayin ang mga hamon at pagkakataon ng ekonomiya ng Tsina.	숫자조차도 두려워요	P	
2014470005	16	미국 인플레이션 감축법(IRA)을 단 제조 생산 지역 공제 제도에 따른 세계 최대인 1889억 달러의 공 금력을 제외하면 LG 에너지 솔루션은 316억 달러의 영업 적자를 기록했다.	Umatob sa 1889 bilyong won ang mga benepiyo sa buwis sa ilalim ng U.S. Inflation Reduction Act (IRA) Advanced Manufacturing Production Tax Credit System, at kung ibabawas ang halagang Ito, nagtala ang LG Energy Solution ng 31.6 bilyong won na lugi sa operasyon.	Kung hindi isasama ang mga benepiyo sa buwis sa ilalim ng Advanced Manufacturing Production Tax Credit ng Inflation Reduction Act (IRA) ng Amerika na umatob sa 1889 bilyong won, nagtala ang LG Energy Solution ng pagkalugi sa operasyon na 31.6 bilyong won.	Sa mga benepiyo sa buwis sa ilalim ng sistema ng pagbawas sa buwis para sa produksiyon ng mga makabagong produkto sa pagmamanupaktura ng Inflation Reduction Act (IRA) ng Amerika, kung hindi isinama ang 1889 na bilyong won, nagtala ang LG Energy Solution ng pagkalugi sa operasyon na 31.6 na bilyong won.	의미역이 아닌 1:1 번역 주의하세요	P	
2014540004	25					내용 변경 주의	P	

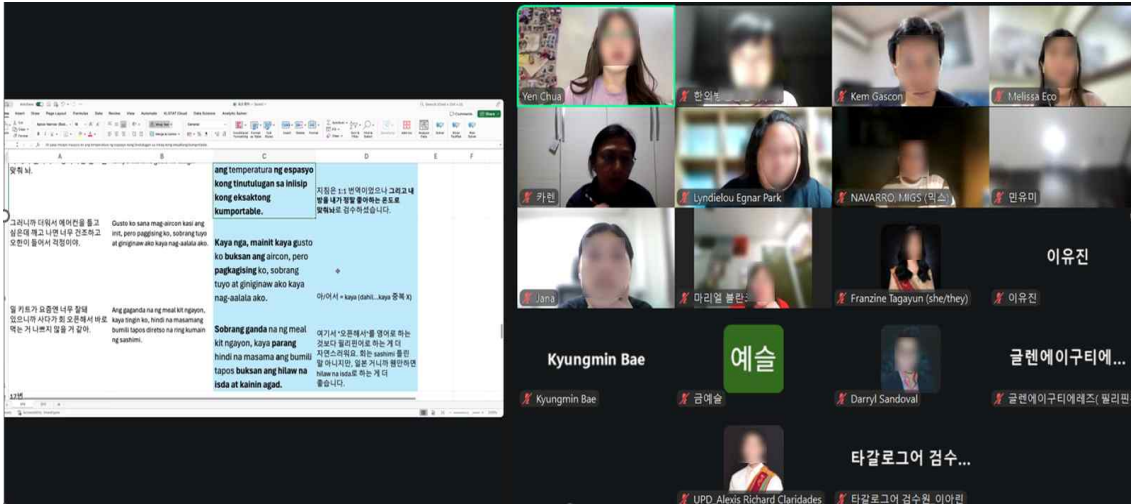
[그림 24] 3차 검수 예시 자료(필리핀 타갈로그어)

또한 검수팀장과 번역 검수원이 3차 검수 결과에 대한 의견을 자유롭게 공유함으로써 검수 능력의 상호 증진을 도모하였다.

SID	Word Count	Meta	Origin	Translation	QC Translation	수정문	제검수 결과에 대한 다른 의견	검수팀장님 최종 의견
100970133	6	[대인관계][연인](S: 6/30)-(L:0/30)	변하고 있는 모습이나, 이런 것에 대해?	О том, как это меняется или обо всем этом?	О том, как это меняется или обо всем этом?	О том, как это меняется, или обо всем этом?	ИЛИ" соединит однородные члены предложения, думаю, что запятая не нужна	Здесь запятая нужна, поскольку она отделяет придаточную часть. Однородные члены здесь "о том" и "об этом", а "как это меняется" — обособленное определение к "том".
100970161	12	[대인관계][연인](S: 0/30)-(L:남/30)	근대 내 생각에는, 그런 사람들은 온 그런 열으로 만나서 거가 조금 거부감이 들거든.	Но, по моему мнению, такие люди немного настроены относительно к знакомству через приложения.	Но, по-моему, у таких людей вызывает некоторое отторжение идея знакомства через приложения.	А тут говорящий не про себя говорит? Только начал про одно, а закончил про другое. Просто такие выражения, как "거부감이 많다, обычно в третьем лице не употребляются. Надо уточнить, кто предлагает вариант от первого лица (возможно, неправильный):	Вроде бы это обсуждение способов знакомств разных людей. Я поискала, 거부감이 많다 может и в третьем лице использоваться, к тому же это разговорная речь.	Тогда по контексту речь не о самом говорящем? В таком случае ваш вариант логичнее.
2002010007	17	[AI]	개막식은 이창윤 (과기정통부) 제1차관과 이승렬 산업부 산업정책 실장, 최수진 (국민의 힘) 의원 등이 참석한 가운데 열렸다.	Церемония открытия состоялась при участии Ли Чханюн, первого заместителя министра науки и ИКТ, Ли Синнэля, директора департамента промышленной политики Министерства промышленности, а также Чхве Суджина, депутата партии Сила народа.	Церемония открытия состоялась при участии Ли Чханюн, первого заместителя министра науки и ИКТ, Ли Синнэля, директора департамента промышленной политики Министерства промышленности, а также Чхве Суджина, депутата партии «Сила народа».	Церемония открытия состоялась при участии Ли Чханюн, первого заместителя министра науки и ИКТ, Ли Синнэля, директора департамента промышленной политики Министерства промышленности, а также Чхве Суджина, депутата партии «Сила народа».	Обычно названия партий пишутся без квычкэн, в том году тоже в роде без них писали, в этот раз с квычкэнками?	Без квычкэн пишется согласованные названия, т.е. когда родовое слово "партия" входит в состав названия и согласуется с остальными словами. Когда синтаксического согласования со словом "партия" нет, название пишется в квычкэн (Партия зеленых членовцев, но партия «Зеленый великан»)
2002110017	23	[정치]	이어 이듬해 12월 이 사건에 관련한 혐의로 name1 회장의 아들(차남)이자 company-name2 사내 이사인 name2(35) 씨와 가상 화폐 운용사 company-name3 대표 name3(48) 씨를 구속했다.	Затем в декабре следующего года были арестованы по подозрению в причастности к этому делу младший сын председателя name1, name2 (35 лет), являющийся внутренним директором компании company-name2, а также генеральный директор криптовалютной инвестиционной фирмы company-name3 — name3 (48 лет).	Затем в декабре следующего года по подозрению в причастности к этому делу были арестованы младший сын председателя name1 — name2 (35 лет), являющийся внутренним директором компании company-name2, а также генеральный директор криптовалютной инвестиционной фирмы company-name3 name3 (48 лет).	Затем в декабре следующего года по подозрению в причастности к этому делу были арестованы младший сын председателя name1 name2 (35 лет), являющийся внутренним директором компании company-name2, а также генеральный директор криптовалютной инвестиционной фирмы company-name3 name3 (48 лет).	Затем в декабре следующего года по подозрению в причастности к этому делу были арестованы младший сын председателя правления name1 и внутренний директор company-name2 name2 (35 лет), а также генеральный директор криптовалютной инвестиционной фирмы company-name3 name3 (48 лет).	동의

[그림 25] 3차 검수 결과에 대한 의견 교환 예시(러시아어)

3차 검수 결과를 바탕으로 오류율이 높은 번역 검수원은 재교육을 실시하였으며, 번역 검수원의 오류 원인에 따라 번역 검수 지침 재교육, 검수문 표본 평가 및 피드백 등을 실시하였다.



[그림 26] 번역 검수원 대상 검수 재교육 예시(필리핀 타갈로그어)

## 2) 한국어 원문 이해 지원

복잡한 상황 맥락이나 전문용어, 신조어 등으로 인해 번역 검수원이 원문을 이해하는 데에 어려움이 있어 원문에 대해 질문할 경우 공동 연구원 및 보조 연구원이 원문의 상황 맥락과 의미를 파악하여 설명하였다.

질문	질문	답변
나는 워너원에서 박우진을 제일 좋아했는데, 분방이 별로 없었는데도 데뷔를 할 정도로 열심히 하는 <b>뭔가</b> 있는 <b>무슨</b> 이 나에게 크게 다가왔던 거 같아.	'분방, 특히, 크게 다가왔던'의 의미 설명 요청	출연자의 노력이나 애력이 방송에서 얼마나 비추졌는지를 나타내는 표현, 여기서 '뭔가'는 부정적인 표현은 아니며, 목표를 이루기 위한 강한 의지와 적극적인 태도를 의미, 태도가 감정적으로 깊은 인상을 남기고, 마음에 와닿았다는 의미
거기서 많이 틀렸던 게, 인간이 이젠 개 하나라 내가 어긴 것이라는 말이 있어 => '말, 말의 차이 설명 요청과 적절한 번역어 문의	'말, 말의 차이 설명 요청과 적절한 번역어 문의	'말'은 인터넷 유행의 일부이고 '말'은 한국의 문학 책에서 만들어진 표현이며, 읽은 그대로 memo으로, 붙은 한국어만 있는 개념은 아니어서 음역은 어렵고 larawan mula sa internet 정도로 번역
그리고 음, INFP 본들이 어떻게 서로 타키타카하면서 대화하는 게 좀 편만하다고 내 스스로가 느꼈던 거 같아.	'타키타카'의 적절한 번역어 논의	음역보다는 유사한 의미의 후조책어로 번역
그, <b>현장</b> 쪽 <b>일도</b> <b>관려</b> 하시던 분이라고 하세요.	이 맥락에서 '현장'이 어떤 의미인지 질문(보통 통역하다)	여기서 '현장'은 사무직의 반대 의미로 사용되고 있습니다. 공장이나 기술직을 모두 포괄하는 의미로, 직접 일을 사용하여 하는 일을 가리키는 것으로 보입니다.
가스를 갈 줄 알아서 그랬나 봐.	'가스를 갈다'의 의미 문의	알려 맥락으로 추측이 어느 정도 가능하나 문학적 맥락에서 정확히 '갈다'가 어떤 의미인지 이해하기 어려웠다고 함, 유대용 가스버너의 가스를 교체하는 거라고 설명
근대 출발지에서 목적지까지 그 걸로써 가느냐 못 가느냐 중요한데 사람이 없으면 그냥 단순하고 어려워지잖아 <b>해</b> .	내용 의미와 문장 연결 문의	보드게임 내용이었는데, 생각이 너무 많고 고백력, 고백잡인 데다가 '단순하고 어려워진다'가 '고'로 연결되는 것도 이해가 안 가서 어려웠다고 함
3인조 밴드니까 거기서 <b>밴드</b> 얘기하는 거지만, <b>밴드는 팔이 있어야</b> 한다고 하는데, 가장 잘 구현한 게 거기, <b>팔</b> 하는 장점이 있어.	'팔이 있어야 된다'는 표현의 의미 문의	이 대화 맥락에서 '밴드는 팔이 있어야 된다'는 '밴드 음악에는 감정과 에너지가 터지는 강렬한 순간이 필요하다'는 뜻으로 해석
아, 그게 이미 들어와 있었는데 그 40시간을 지키기 위해서 일부러 자기 일한 시간으로 안 찍고 막, <b>출근할 걸로 안 치고</b> 그랬던 걸로 알아.	'자기 일한 시간으로 안 찍고' 의미 문의	회사에서 '실제로 근무한 시간을 근무시간으로 공식 기록하지 않고 누락하거나 다른 방식으로 처리하는 것'을 의미함, 이것은 '노동 시간 단축 제도의 도입'과 관련, 실제 근무 시간을 회사에 공식적으로 기록하지 않는 행위
요새 <b>통자</b> <b>틀리기</b> 라고 해서 <b>통자</b> 들어 봤어?	'통자 틀리기'가 하나의 표현으로 경제 용어인 거 같은데 어떻게 적절한 번역해야 할지 모르겠습니다.	통자 틀리기는 매달 새로운 예금이나 적금 상품에 가입하여 목돈을 모으는 채테크 방법을 의미하는데, '매달 100만원'으로 번역하는 게 적절할 것 같음.
근대 간헐적 단식 이런 것도 괜찮은 거 같기도 한데, 나는 아, <b>간헐적 단식</b> 이 나는 좀 <b>절도</b> 라고 생각할 거 같은.	이 문장에서 '절도'가 어떤 의미로 쓰였는지 문의, 이 정도까지는 못인지 아니면 좋은 방법이라는 뜻인지 궁금함	대화의 맥락을 확인해 보니, 좋은 방법, 바른 규칙이 맞는 것 같음.
그리고 <b>진짜</b> <b>주인</b> 을 잘 못 알아본지 하더라고요.	이 문장에서 '진짜 주인'인지 아니면 '진짜로 주인을 못 알아본다'인지 조금 헷갈립니다	맥락을 확인해 보니 가족이 진짜로 주인을 못 알아본다는 의미
내가 결혼 초에 돈 모았던 방법이 <b>목적</b> <b>통장</b> 이었잖아.	'목적 통장' 뜻 문의	특정한 목적을 가지고 돈을 따로 저축하는 통장으로 예를 들어, 여행 자금, 결혼 자금, 전세 자금, 자녀 교육비 등을 모으기 위해 개설하는 전용 저축 통장

[그림 27] 원문 이해 지원 예시

### 3) 번역 품질 개선을 위한 의견 취합 및 전달

번역문에서 자주 발견되는 오류, 원문 의미 오역 등의 예시를 정리하여 번역을 담당하는 구축팀에 의견을 전달하였다. 이를 향후 번역 작업에 반영되도록 하여 번역의 정확도와 검수의 효율성을 높이고자 하였다.

SID	원문	번역문	오류 내용
100767127	그런 거 보면은 '야, 나도 그 산속에 들어가서 한번 살아 봤으면 좋겠다.'라는 생각도 하거든요.	जब मैं ऐसी चीज देखता हूँ, तो सोचता हूँ, 'काश, मैं किसी दिन उन पहाड़ों में जा सकता और वहाँ रह सकता।'	이 문장에서는 스피커와 리스너가 여성이지만 번역은 남성으로 됨. यदि ऐसी चीजों को देखती हूँ तो सोचती हूँ कि 'काश, मैं भी वैसी किसी पहाड़ियों में जाकर एक बार रहती तो अच्छा होता।'
100878126	(S:여(20)/L:여(20)) 그래서 내가 케치를 느낌으로 잡고 던지는 건 해 줄 수 있어.	तो मैं इसे पकड़ने और फेंकने जैसा कुछ कर सकता हूँ।	원문에 표기된 성이 여성이지만 힂디어에는 남성로 잘못 번역됨.
100819092	질 치는 사람이 너무 많잖아.	ऐसे बहुत से लोग हैं जो खेलने में अच्छे हैं।	원문에서는 '치는 것' 뜻이 खेलना 아니고 बजाना 뜻이다. 여기 원문 내용이 "기타 치는 것"이다.
100819126	초등학교 1학년 때부터 치기 시작해서 중학교 때까지도 계속 치고 그랬지.	मैंने प्राथमिक विद्यालय की पहली कक्षा से खेलना शुरू किया और भिड़िल रकूल तक खेलती रही।	이 원문에서도 '치는 것' 뜻이 खेलना 아니고 बजाना 뜻이다. 이 원문 내용도 "기타 치는 것"이다.
100878129	언니는 가족들이랑 같이 운동하는 것도 있어요?	क्या तुम्हारी बहन अपने परिवार के साथ व्यायाम करती है?	언니 라는 주어의 번역이 잘못됨.
100878149	어, 그리고 태권도도 기회가 되면 같이해 보고 싶은데 한번 도장 알아볼까요?	हाँ, अगर मुझे मौका मिले तो मैं ताइकांडो भी आजमाना चाहूंगी और क्या मुझे जिम जाना चाहिए?	잘자 오류도 있고 원문에도 도장이라는 단어의 번역이 잘못됨.
100747051	큰딸은 뭐든지 잘 먹어도 비만이 없어.	मेरी सबसे बड़ी बेटी चाहे कुछ भी खा ले, उसे मोटापा नहीं होता।	여기는 맞춤법 오류이고, 빨간색으로 표시되어 있다. मेरी बड़ी बेटी सब कुछ अच्छी तरह से खाती है फिर भी उसे मोटापा नहीं होता।
100747072	저는 오로지 밥, 그것도 한식, 된장국, 김치 이런 거만 좋더라고요.	मुझे केवल चावल, खासकर कोरियाई भोजन, जैसे कि धनबांगमूक और किमची पसंद है।	मुझे केवल चावल, खासकर कोरियाई भोजन, जैसे कि धनबांगमूक और किमची पसंद है।
100747147	인바디 검사도 잘 다녀서 한 달 전보다 지금 어떻게 변했는지 관심을 많이 가지게 됐어요.	इनबॉडी टेस्ट अच्छे से करवाया और अब मैं एक महीने पहले की तुलना में तुमने बदलावों के बारे में बहुत उत्सुक हूँ।	문장의 의미가 잘못됨. 의미 없는 단어 추가.

[그림 28] 번역 품질 개선을 위한 의견 작성 및 전달 예시(인도 힂디어)

### 4) 번역 오류 유형 정리 및 공유

검수 과정에서 나타난 오류 유형을 범주별로 정리하고, 원인을 분석하여 번역 품질을 높이고자 하였다. 검수 과정에서 파악한 번역문의 오류 유형은 다음과 같다<sup>3)</sup>.

<표 27> 번역문 오류 유형

언어	어휘	문법	내용	표기법	기타
베트남어	46.8%	13.2%	7.7%	7.2%	25.1%
인도네시아어	41.6%	11.3%	6.9%	20.7%	19.5%
태국어	46.3%	4.0%	2.5%	29.8%	17.5%
인도 힂디어	36.1%	30.4%	19.0%	11.5%	3.0%
캄보디아 크메르어	36.5%	11.5%	15.0%	20.7%	16.3%
필리핀 타갈로그어	39.8%	20.6%	23.0%	14.8%	1.8%

3) 검수 오류 유형 정리 및 공유는 사업 기간에 수시로 진행되었으며, 여기에 제시된 표는 전수 검수 종료 후 모든 오류 유형을 정리한 것이다.

러시아어	45.7%	12.5%	5.2%	8.4%	28.1%
우즈베크어	44.8%	13.6%	13.5%	26.2%	1.9%
평균	42.2%	14.6%	11.6%	17.4%	14.2%

전반적으로 어휘 오류의 비율이 가장 높고 다음으로 표기법 오류, 문법 오류 및 내용 오류, 기타 순으로 나타났다. 언어마다 오류 유형의 비율이 다양한 양상을 보였지만 어휘 오류의 비율이 가장 높은 것이 대부분 언어의 공통적인 특징이었다. 어휘 오류는 인명, 지명, 기관명 등의 고유명사나 전문용어, 구어 담화 표지의 사용, 특정 어휘의 사용역이 부적합한 경우 등이 있었다. 문법 오류의 경우, 어순과 시제, 능·피동, 성·수 일치, 개별 문법 항목(조사, 접속사 등)에서 오류가 많이 발생하였다. 내용 오류는 문장 구조가 복잡하거나 원문을 이해하는 데에 한국의 경제, 사회, 문화적 지식을 많이 요하는 경우에 원문을 잘못 번역하거나 일부 내용을 첨가·누락한 경우가 많았다. 표기법 오류는 주로 고유명사의 음역 표기나 첩표 및 따옴표 등의 문장 부호에 오류가 나타났으며, 철자 오류와 대소문자 및 숫자 표기 오류도 나타났다. 기타 오류는 지나친 직역으로 인해 자연스럽지 않은 번역이 된 경우가 많았다.

언어별 번역 오류 양상과 주요 원인은 다음과 같다.

<표 28> 언어별 번역 오류 양상 및 원인

언어	오류 양상 및 주요 원인
베트남어	<ul style="list-style-type: none"> <li>어휘 오류의 비율이 가장 높았는데 의미는 비슷하지만 구어나 문어 맥락에 적합하지 않은 어휘를 사용한 경우가 많았으며, 오역·누락과 불필요한 어휘 추가도 확인되었다.</li> <li>그다음으로 비율이 높은 기타 오류는 의미는 통하지만 문장 구조가 잘못된 경우가 많았다.</li> <li>문법 오류는 어순이 맞지 않는 경우가 많았다.</li> <li>내용 오류는 원문 의미나 문법 구조를 잘못 해석해 다른 내용으로 번역한 경우가 많았다.</li> <li>표기법 오류는 대소문자의 구별, 성조 표기, 띄어쓰기 오류가 많았고, 번역 지침 미준수로 인한 오류도 있었다.</li> </ul>
인도네시아어	<ul style="list-style-type: none"> <li>어휘 오류가 전체 오류 중 가장 높은 비중을 차지하였다. 특히 고유명사, 직책명 등 특정 명칭을 영어식 표현으로 그대로 사용하는 경우가 많았다. 또한 문어에서 주로 쓰이는 어휘를 구어 맥락에 사용하거나, 반대로 구어적 표현을 문어 문장에서 사용하는 등 담화 양식에 부적합한 어휘 선택도 주요 오류 유형으로 확인되었다.</li> </ul>

	<ul style="list-style-type: none"> <li>• 기타를 제외하고 비율이 높은 표기법 오류의 양상은 침표 사용의 오류, 인용문에서의 따옴표 및 마침표 사용의 오류가 많았다. 이는 한국어 및 영어와의 문장 부호 사용에 차이가 있어 오류가 많이 발생한 것으로 보인다. 문장 부호 수정 시에는 인도네시아어 대사전(Kamus Besar Bahasa Indonesia, KBBI)과 EYD V의 기준을 따랐다.</li> <li>• 문법 오류는 시제 및 피·사동 표현의 부적절한 사용이 두드러졌으며, 서로 결합될 수 없는 문법 형식을 함께 사용하는 결합 제약 위반도 주로 나타났다. 내용 오류는 한국어 원문의 일부가 누락되거나 특정 단어를 잘못 번역한 사례가 주로 확인되었다. 특히 한국어와 인도네시아어의 인용문 구성 방식 차이로 인해, 인용문이 포함된 문장에서 오류가 집중되는 경향이 있었다.</li> <li>• 이 외에 기타 오류는 지나친 직역으로 인해 내용 이해는 가능하나 인도네시아어 문장이 어색한 경우가 다수를 차지하였다. 특히 인도네시아어의 특성을 고려하지 않고, 원문과 번역문의 일대일 대응에 초점을 두어 번역하여 문장이 어색해진 경우가 많았다.</li> </ul>
태국어	<ul style="list-style-type: none"> <li>• 어휘 오류가 전체 오류 중 가장 높은 비중을 차지하였다. 지칭어 번역 오류, 상위어와 하위어의 구분을 잘못하여 번역한 경우, 담화 표지의 어색한 번역, 사회·문화적 배경을 충분히 반영하지 못한 번역 등이 있었다.</li> <li>• 다음으로 비율이 높은 오류는 표기법 오류로, 태국에서 관용적으로 사용하는 표기와 다른 방식으로 번역한 경우, 태국에서 한국어 발음을 사용하지 않음에도 한국어 발음을 그대로 태국어로 옮긴 사례(예: 태국어에서 중국어 발음으로 표기해야 하는 고유명사를 한국어 발음으로 표기한 경우), 영어 표기와 태국어 표기의 혼동, '-'을 사용하지 않으면 음절 구분이 어려움에도 불구하고 생략한 사례 등이 있었다.</li> <li>• 문법 오류는 공손 표현 번역이 적절하지 않은 경우, 종결 표현 대응의 오류(적절한 태국어 종결 표현으로 옮기지 못해 화자의 태도가 충분히 드러나지 않은 경우) 등이 있었다.</li> <li>• 이 외에 번역 누락, 오역 등이 일부 있었다.</li> </ul>
인도 힌디어	<ul style="list-style-type: none"> <li>• 어휘 오류의 비율이 높았는데 고유어나 축약어 번역에서 오류가 많았다.</li> <li>• 문법 오류는 성의 불일치, 문장 어순 오류가 가장 많았다.</li> <li>• 내용 오류는 문어의 경우 일부 긴 문장에서 뒷부분은 번역하지 않은 오류, 구어에서는 원문의 의미를 이해하지 못하거나, 이해는 했으나 부자연스럽게 번역된 경우가 많았다.</li> <li>• 표기법 오류는 번역 지침을 따르지 않아서 나타난 오류로 모음 표기 오류가 많았다.</li> </ul>
캄보디아	<ul style="list-style-type: none"> <li>• 오류 유형 가운데 어휘 오류가 가장 많이 나타났는데, 주로 전</li> </ul>

<p>크메르어</p>	<p>문용어의 의미 차이나 맥락적 대응의 어려움에서 비롯된 경우가 많았다. 특히 법, 정치, 경제, 기술 등 전문 분야의 용어를 목표어로 옮기는 과정에서 의미의 정확성과 일관성을 확보하지 못해 수정이 필요가 경우가 많았다.</p> <ul style="list-style-type: none"> <li>• 다음으로 비율이 높은 표기법 오류는 단순 철자나 음역 표기의 문제에서 많이 나타났다. 인명과 지명 표기, 문장 부호 누락, 띄어쓰기 등에서 형식적 정확성이 확보되지 않아 교정이 이루어진 경우가 많았다.</li> <li>• 내용 오류는 원문 맥락에 대한 이해가 충분하지 않아 발생한 경우가 많았다. 구어에서는 고맥락적 표현이나 도치 구조로 인해 의미 파악이 어려운 경우가 있었고, 문어에서는 의미 불일치, 내용 누락 또는 불필요한 첨가 등이 문제로 나타났다.</li> <li>• 기타 오류에는 지나치게 직역되거나 의역된 문장을 수정한 경우가 많았다.</li> <li>• 문법 오류는 복잡한 문장에서의 어순 혼란, 접속 표현의 부적절한 선택 등 구조적 완결성 문제가 있었다.</li> </ul>
<p>필리핀 타갈로그어</p>	<ul style="list-style-type: none"> <li>• 어휘 오류로는 번역 과정에서 맥락과 문화적 특성을 고려하지 않은 어휘 선택이 다수 나타났다. 구어 말뭉치에서는 실제 회화에서 잘 쓰이지 않는 접속사를 사용하거나, 필수 어휘 생략으로 의미 전달이 미흡한 경우도 있었다.</li> <li>• 문법적 오류는 주로 한국어 문장을 타갈로그어 구조로 재조직하는 과정에서 발생했다. 긴 관형절을 적절히 분석하지 못해 행위의 주체가 바뀌거나, 시제·상 표현 부정확성으로 시간 개념이 모호해지거나, 품사 선택의 부적절성, 주어 생략 및 담화표지 누락으로 화자가 불명확한 문장도 있었다.</li> <li>• 내용 오류는 의미 변형과 누락이 많았다. 관용 표현을 직역해 의미가 왜곡되거나, 부정·강조 표현을 잘못 해석해 원 의미와 다르게 전달되는 경우가 있었다.</li> <li>• 표기법 오류로 고유명사와 전문용어의 철자 불일치 및 비일관적 사용, 번역 지침을 지키지 않은 표기, 대·소문자 사용 오류, 문장 부호의 누락 또는 변경으로 인해 문장의 경계와 의미 기능이 달라지는 사례, 오타자도 다수 확인되었다.</li> </ul>
<p>러시아어</p>	<ul style="list-style-type: none"> <li>• 어휘 오류의 비율이 가장 높았는데 의미의 정확성이나 자연스러움을 높이거나 번역의 일관성을 유지하기 위해 수정한 경우가 많았다.</li> <li>• 그다음으로 비율이 높은 기타 오류는 지나치게 직역되거나 의역된 문장을 수정한 경우가 많았다.</li> <li>• 문법 오류는 대명사 오류, 문장 성분의 성·수 불일치, 격 어미 및 시제, 어순의 오류 등이 많았다.</li> <li>• 표기법 오류는 음역 표기, 첨표 및 겹화살표 등의 문장 부호, 숫자 단위 표기, 철자법 등에서 오류가 많이 나타났다.</li> </ul>

	<ul style="list-style-type: none"> <li>• 내용 오류로는 한국어 원문의 의미를 잘못 이해하여 오번역한 경우가 많았다.</li> </ul>
우즈베크어	<ul style="list-style-type: none"> <li>• 어휘 오류의 비율이 가장 높게 나타났는데 고유명사(인명, 지역명, 기관명, 작품명, 프로그램명 등)의 통일을 위해 수정하는 경우가 많았고 신조어나 특정 명사의 의미를 잘못 번역한 경우가 있었다.</li> <li>• 그다음으로 높게 나타난 오류는 표기법 오류로 어휘와 마찬가지로 다양한 표기법을 하나로 통일하는 과정에서 표기법을 수정하는 경우가 많았으며 숫자를 표기하는 방식에 관한 오류, 쉼표나 작은따옴표 등 기호 사용에 관한 오류가 있었다.</li> <li>• 내용 오류는 원문의 맥락을 고려하지 않아서 발생하는 의미 불일치 오류, 내용 누락 오류, 내용 첨가 오류가 나타났다.</li> <li>• 문법 오류는 시제 및 어순의 오류, 조사 오류 등이 있었다.</li> </ul>

## 5. 전문용어 번역집

### 5.1. 제작 목적과 필요성

대규모 다국어 말뭉치를 구축하는 과정에서 전문 분야의 개념을 정확히 정립하고 그에 대응하는 용어를 표준화하는 일은 데이터 품질을 결정하는 핵심 요소이다. 특히 사회과학 분야의 하위 범주인 법률 영역은 국가별로 제도 차이가 크기 때문에 말뭉치의 번역 신뢰도와 활용도에 직접적인 영향을 미친다. 이에 본 연구에서는 국립국어원 언어정보나눔터 ‘온용어’에 수록된 ‘법률’ 분야 용어에서 전문용어를 선별하고, 표제어의 의미 구조를 명확히 분석하여 8개 언어로 번역하였다. 이를 통해 번역어 일관성과 전문성을 확보하여 향후 말뭉치 구축 과정에서 용어 혼란과 번역 오류를 최소화하고자 하였다.

### 5.2. 제작 과정

#### 5.2.1. 표제어 선별 및 확정

‘온용어’는 국립국어원이 제작한 전문 분야 용어 지식 플랫폼으로, ‘온용어’에서는 기관별·분야별로 흩어져 있던 용어 자료를 한데 모아 ‘국가과학기술표준분류체계(2023)’에 따라 대분류 22개 항목, 중분류 277개 항목으로 분류하여 통일된 분야 분류 정보를 제공한다.

전문용어 번역집은 ‘온용어’의 수록 용어와 분류 체계 내에서 제작하고자 하였





[그림 32] 법제처 국가법령정보센터 법령 용어 수록 여부 확인

### 5.2.2. 제작 시 고려 사항에 대한 전문가 자문

전문용어 번역집 제작에 앞서 용어집 구성과 표제어를 정할 때 고려해야 할 사항에 대해 외국어 자문위원의 의견을 청취하였다. 주요 자문 내용은 다음과 같다.

- ① 법체계의 차이
  - : 한국과 8개 언어의 법률 개념, 체계, 용어 구조가 다르기 때문에 일대일 대응이 어려움. 예) ‘가족’의 개념, ‘주소’의 정의 등.
- ② 한자의 존재
  - : 한국 법률에서는 개념 정확성을 위해 한자를 병기하는 경우가 많음. 그러나 번역 시 해당 한자의 의미를 정확히 해석해야 의미 왜곡을 피할 수 있음.
- ③ 동음이의어(同音異義語) 및 중의성
  - : 하나의 한국어 법률 용어가 여러 의미로 해석될 수 있음. 예) ‘결정’은 문맥에 따라 ‘판결’, ‘행정 처분’ 등 다양한 의미를 가짐.
- ④ 기계 번역의 한계
  - : 기계 번역은 법적 맥락이나 문화적 함의를 파악하기 어렵고, 전문성 부족으로 인한 오역 가능성 높음.
- ⑤ 번역자의 비전문성
  - : 번역자가 법학적 배경이 없을 경우 법률 의미 해석에 오류 발생 가능.

법률 개념과 제도의 차이로 일대일 번역이 어려운 경우에는 복수의 번역어를 제시하여 문맥에 따라 적절한 번역어를 선택할 수 있도록 하였다. 한자어의 의미를

정확히 전달하기 위해 원어와 의미 설명을 제시하고 해당 용어가 포함된 기구축 문장을 함께 제공하였다. 중의적인 표현은 문맥에 부합하는 번역어를 선정하였으며 전체 용어는 언어별 팀장이 직접 번역한 뒤 필요시 법률 전문가의 자문을 받아 법적 정확성을 확보하였다. 최종본은 외국어 자문 위원들의 전수 검토와 감수를 거쳐 완성하였다.

### 5.2.3. 각 용어의 번역어 및 번역 유형

표제어로 선정한 337개의 용어를 본 사업의 대상 언어인 베트남어, 인도네시아어, 태국어, 인도 힌디어, 캄보디아 크메르어, 필리핀 타갈로그어, 러시아어, 우즈베크어로 번역하였다. 또한 각 용어의 번역 유형을 음역, 의미역, 음역과 의미역의 혼합 형태로 나누어 정리하였다.

### 5.2.4. 전문용어 번역집 형식 및 구성

전문용어 번역집의 형식은 표제어, 원어, 정의문, 8개 언어의 번역어, 번역 유형으로 구성하였다. 정의문은 선정한 용어의 의미를 ‘우리말샘’을 바탕으로 정리하였다. 또한 수록 용어를 ‘우리말샘’ 또는 ‘온용어’에서 손쉽게 찾을 수 있도록 해당 용어의 인터넷 주소(URL)를 함께 기재하였다.

### 5.2.5. 번역 및 감수

8개 언어의 검수팀장이 전문용어 번역집에 수록된 표제어를 해당 언어에 맞게 번역하였다. 이어서 8개 언어의 외국어 자문 위원이 담당 언어의 번역 내용을 전수 감수하였으며 감수 내용은 추가 논의를 거쳐 전문용어 번역집 제작에 반영하였다.

이상의 절차를 거쳐 전문용어 번역집을 다음과 같이 구성하였다.

순번	표제어	원어(은/영어)	정의문(은/영어)	베트남어			인도네시아어			태국어			인도 힌디어			캄보디아 크메르어			필리핀 타갈로그어			
				번역(필장 최종 확정)	번역 유형	음역	의미역	혼합	번역(필장 최종 확정)	번역 유형	음역	의미역	혼합	번역(필장 최종 확정)	번역 유형	음역	의미역	혼합	번역(필장 최종 확정)	번역 유형	음역	의미역
1	가계약	<한자>假契約	정식 계약을 맺기 전에 임시로 맺는 계약.	Hợp đồng tạm thời	0		kontrak sementara	0		สัญญาชั่วคราว	0		अस्थायी संधि/करार	0		កិច្ចសន្យាសម្រេចជាមុន	0		pauunang kontrata, probisyonal na kontrata	0		
2	가압류	<한자>假押留	강제 집행을 하기 전에 재산을 숨기거나 팔아 버릴 우려가 있을 경우에 한다.	Tịch biên tạm thời	0		penyitaan sementara	0		การอายัดทรัพย์	0		अंतरिम ज़रती / अस्थायी ज़रती	0		ការបិទសម្របសម្រួល	0		probisyonal na pagkumpiska	0		
3	감경	<한자>減輕	부과 정하여진 형벌보다 가벼운 형벌에 처함 또는 그런 형벌에 처하여짐.	Giảm nhẹ	0		keringanan hukuman, peringanan hukuman	0		การลดโทษ / การขอโทษ	0		सजा कम होना	0		ការកាត់បន្ថយទោសកម្រិត	0		pagpapagaan ng parusa, pagbabawas ng parusa	0		
4	감독	<한자>監督	어떤 사람이나 기관이 다른 사람이나 기관의 행위가 잘못되지 않도록 감시하고, 필요한 경우에는 명령이나 재재를 가하는 일.	Giám sát	0		pengawasan, supervisi	0		การควบคุม / การกำกับ	0		पर्यवेक्षक	0		ការត្រួតពិនិត្យ	0		tagapangasiwa	0		
5	감염병 전문 병원	<한자>感染病専門病院	병상을 갖추어야 하며, 국가에서 설립하거나 지정하여 운영한다.	chuyên khoa về các bệnh truyền nhiễm	0		sakit khusus penyakit menular	0		โรงพยาบาลเฉพาะทางสำหรับโรคติดต่อ	0		रोग विशेषज्ञ अस्पताल, महामारी विरोधक अस्पताल	0		មន្ទីរពេទ្យជំងឺឆ្លង	0		ospital na dalubhasa sa mga nakakahawang sakit	0		
6	감형	<한자>減刑	줄여 주는 일. 대동형이 행하는데, 일반 감형과 특별 감형이 있다.	Giảm án	0		hukuman, peringanan hukuman	0		การลดโทษ / การขอโทษ	0		दंड कम करना	0		ការកាត់បន្ថយទោសកម្រិត	0		pagbabaling hatol	0		

[그림 33] 전문용어 번역집 최종본 예시

## 6. 한국어-외국어 말뭉치 구축 제2차 중장기계획 수립4)

### 6.1. 제2차 중장기계획 수립 개요

‘한국어-외국어 병렬 말뭉치 구축 사업’은 2021년부터 2025년까지 ‘한국 언어 문화 산업 빅데이터 구축 중장기 전략 계획’(문화체육관광부, 2021)에 따라 5년간 진행되어 고품질 병렬 말뭉치 데이터 제공을 위한 기반 체계를 마련하였으나, 생성형 인공지능의 급속한 확산, 다국어 데이터 수요 증가, 국가 AI 전략의 본격화 등 환경 변화에 따라 말뭉치의 범위, 구조, 활용 방식 전반에 대한 재정비 및 고도화가 필요한 상황이다.

특히 맥락이 반영된 문서 단위 데이터와 저자원 언어 지원에 대한 요구가 증대되고 있으며, 공공 데이터 개방·활용 정책 강화 등 사업 외부 환경이 빠르게 변화함에 따라 기존 성과를 계승하면서도 향후 5년간 체계적이고 지속 가능한 데이터 구축 전략을 마련할 필요가 제기되었다.

이에 본 사업에서는 변화된 기술·정책·산업 환경을 반영하여 한국어 외국어 말뭉치 구축 사업의 비전과 목표를 재정립하고, 구축 대상 언어, 데이터 구조, 데이터 활용 체계 등 핵심 영역의 중장기적인 방향성을 제시함으로써 한국어 기반 다국어

4) 본 보고서에는 제2차 중장기계획 수립의 내용을 요약하여 수록하며 세부 내용은 별책으로 제시한다.

데이터 생태계의 경쟁력 강화를 위한 ‘한국어-외국어 말뭉치 구축 제2차 중장기계획’을 수립하였다.

제2차 중장기계획에는 제1차 중장기계획 기간의 성과, 국내외 환경 변화, 산업·기술 수요 등을 종합적으로 반영하고자 하였다. 이를 위해 기존 한국어-외국어 말뭉치 구축 현황과 활용 실적을 분석하고 국내외 병렬 말뭉치 구축 사례, 관련 법·제도 및 정책, 기술 동향을 체계적으로 검토하였다. 또한 산학연 전문가 의견을 통해 목표, 전략, 과제의 타당성과 실현 가능성을 확보하였다. 아울러 연차별 구축 계획, 예산 산정 등 실행 계획의 기본 틀을 마련함으로써 향후 5개년 사업이 일관된 전략적 틀 아래 단계적으로 추진될 수 있는 기획 체계를 확립하였다.

## 6.2. 제1차 중장기계획 성과 분석

2021년~2025년 사업은 8개 언어에 대한 대규모 병렬 말뭉치 구축을 계획대로 추진하여 정량적 목표를 충족하였으며, 다단계 검수·품질 관리 체계 운영을 통해 높은 품질 수준을 유지함으로써 데이터의 활용성을 크게 제고하였다.

구축된 말뭉치는 학계·산업계·공공 기관에서 연구, 인공지능 학습, 번역 품질 개선 등 다양한 목적에 활용되어 사업의 파급 효과를 확인하였으며, 활용 현황 분석을 통해 언어별·주제별 수요 특성과 향후 확대 방향을 도출하였다.

이와 같은 성과 분석을 바탕으로 향후 중장기계획에서 반영해야 할 개선점과 전략적 보완 요소를 도출하였다.

## 6.3. 환경 및 요구 사항 분석

제1차 중장기계획의 성과 분석 이후, 제2차 중장기계획의 방향을 고도화하기 위해 국내외 환경과 이해관계자 요구를 종합적으로 조사·분석하였다.

정책적으로는 국가 AI 전략, 공공 데이터 개방 정책, 저작권·개인 정보 보호 제도 등 관련 법·제도를 검토하여 사업 추진 과정에서 고려해야 할 제약과 가능성을 확인하였다. 기술적으로는 대규모 언어모델과 멀티모달 인공지능 기술이 급속히 발전함에 따라 맥락 기반의 문서 단위 데이터, 음성·이미지 기반 자원 등 새로운 형태의 학습 데이터에 대한 수요가 증가하고 있음을 확인하였다. 산업계와 학계, 공공 기관을 대상으로 한 수요 조사와 전문가 자문에서는 언어 확장, 주제 다양화, 데이터 유형 고도화, 플랫폼 활용성 개선 등이 주요 요구로 도출되었다.

#### 6.4. 비전 및 전략 체계 수립

제2차 중장기계획에서는 사업의 비전과 목표를 재정립하고, 변화한 기술·정책 환경에 적합한 중장기적 추진 방향을 수립하기 위해 전체 사업 구조를 체계적으로 정비하였다.

이를 위해 기존 사업 성과와 외부 환경 분석 결과를 토대로 핵심 추진 영역을 재구성하고 각 영역에서 달성해야 할 전략적 방향과 중점 과제를 정리하였다. 또한 데이터 구축, 구조 고도화, 활용·확산 등 사업 전반이 유기적으로 연계되도록 전략 체계를 설계하여 중장기적으로 안정적이고 일관된 추진이 가능하도록 하였다.

아울러 전략 체계와 목표 달성 수준을 점검할 수 있도록 성과 지표를 정비하여 향후 5개년 동안의 사업 성과를 체계적으로 관리·환류할 수 있는 기반을 마련하였다.

#### 6.5. 실행 계획

비전과 전략을 구체화하기 위해 제2차 중장기계획의 목표에 부합하는 실행 계획을 마련하였다.

이를 위해 단계별 추진 구조를 정립하고 연차별 구축 규모와 주요 추진 방향을 설정하여 사업이 안정적으로 이어질 수 있는 기반을 구축하였다. 또한 데이터 구축과 구조 고도화, 활용 기반 강화 등 사업 전반의 흐름이 유기적으로 연계되도록 주요 실행 요소를 검토하고 정비하였다. 사업 성과와 관련해서는 플랫폼 개선, 협력 체계 강화, 성과 관리 등 추진 과정에서 필요한 지원 기반을 종합적으로 점검하여 향후 사업 운영의 실효성을 높일 수 있도록 하였다. 마지막으로 사업 추진에 필요한 자원을 종합적으로 산정하여 계획의 실현 가능성을 확보하였다.

## 7. 병렬 말뭉치 구축 및 메타 정보

### 7.1. JSON 포맷

국립국어원 말뭉치 metadata JSON 형식						
1수준	2수준	3수준	4수준	5수준	타입	항목 설명
id					string	말뭉치 파일 ID
metadata					object	파일의 메타 정보
	title				string	원시 말뭉치 파일 제목
	creator				string	생성자
	distributor				string	배포자
	year				string	말뭉치 구축 연도
	category				string	분류
	annotation_level				string	분석 층위: 예) 원시
	sampling				string	샘플링 방식
document					array	문서 정보
	id				string	문서 ID
	metadata				array	문서의 메타 정보
		title			string	문서 제목
		author			string	저자
		publisher			string	발간자
		date			string	생성일
		topic			string	국어원 설정 소주제
		original_topic			string	기관 설정 대주제
		speaker			array	
			id		string	발화자 ID
			age		string	발화자 나이
			sex		string	발화자 성별

[그림 34] metadata JSON 형식

국립국어원 document_paragraph JSON 형식						
1수준	2수준	3수준	4수준	5수준	타입	항목 설명
document_paragraph					array	문장 정보
	id				string	문서의 ID
	paragraph				array	문서의 메타 정보
		id			string	문장 고유의 ID
		original_form			string	정제 이전의 문장
		form			string	정제된 문장

[그림 35] document\_paragraph JSON 형식

국립국어원 말뭉치 병렬 말뭉치 JSON 형식						
1수준	2수준	3수준	4수준	5수준	타입	항목 설명
language_info					object	
	source_language				string	출발어
	target_language				string	도착어
parallel					array	
	id				string	문장 고유어 ID
	source				string	정체된 문장
	target				string	최종 번역문
	revision				array	
		revision1			string	1차 검수된 문장
		revision2			string	2차 검수된 문장

[그림 36] 병렬 말뭉치 JSON 형식

## 7.2. 최종 데이터 구조

SID	topic	original_topik	Word Count	form	revision1	revision2	target
2000010002	IT_과제+교과필	IT/과제	21	16	16	16	16
2000010004	IT_과제+교과필	IT/과제	13	13	13	13	13
2000010005	IT_과제+교과필	IT/과제	19	19	19	19	19
2000010006	IT_과제+교과필	IT/과제	15	15	15	15	15
2000010007	IT_과제+교과필	IT/과제	19	19	19	19	19
2000010008	IT_과제+교과필	IT/과제	10	10	10	10	10
2000010009	IT_과제+교과필	IT/과제	14	14	14	14	14
2000010010	IT_과제+교과필	IT/과제	24	24	24	24	24
2000010011	IT_과제+교과필	IT/과제	15	15	15	15	15
2000010012	IT_과제+교과필	IT/과제	22	22	22	22	22
2000010013	IT_과제+교과필	IT/과제	14	14	14	14	14
2000010014	IT_과제+교과필	IT/과제	10	10	10	10	10
2000010015	IT_과제+교과필	IT/과제	9	9	9	9	9
2000010016	IT_과제+교과필	IT/과제	9	9	9	9	9
2000010017	IT_과제+교과필	IT/과제	9	9	9	9	9
2000010018	IT_과제+교과필	IT/과제	9	9	9	9	9
2000010019	IT_과제+교과필	IT/과제	26	26	26	26	26
2000010020	IT_과제+교과필	IT/과제	8	8	8	8	8
2000010021	IT_과제+교과필	IT/과제	17	17	17	17	17
2000010023	경제-반도체	경/IT/과제	12	12	12	12	12
2000010026	경제-반도체	경/IT/과제	18	18	18	18	18
2000010027	경제-반도체	경/IT/과제	27	27	27	27	27
2000010028	경제-반도체	경/IT/과제	16	16	16	16	16
2000010029	경제-반도체	경/IT/과제	16	16	16	16	16
2000010030	경제-반도체	경/IT/과제	17	17	17	17	17
2000010031	경제-반도체	경/IT/과제	20	20	20	20	20
2000010032	경제-반도체	경/IT/과제	11	11	11	11	11
2000010033	경제-반도체	경/IT/과제	16	16	16	16	16
2000010034	경제-반도체	경/IT/과제	8	8	8	8	8
2000010035	경제-반도체	경/IT/과제	17	17	17	17	17
2000010036	경제-반도체	경/IT/과제	20	20	20	20	20
2000010037	경제-반도체	경/IT/과제	24	24	24	24	24
2000010038	경제-반도체	경/IT/과제	20	20	20	20	20
2000010039	경제-반도체	경/IT/과제	20	20	20	20	20
2000010040	경제-반도체	경/IT/과제	12	12	12	12	12
2000010041	경제-반도체	경/IT/과제	21	21	21	21	21

[그림 37] 구축 데이터 샘플 Excel 예시

### 7.3. JSON 예시

<표 29> 메타데이터 JSON 형식 예시(문어)

```
{
  "id": "NIPA252512150"
  "metadata": {
    "title": "국립국어원 병렬말뭉치 원천 자료 문어 NIPA252512150",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2025",
    "category": "인터넷 기반 신문",
    "annotation_level": [
      "원시"
    ],
    "sampling": ""
  },
  "document": [
    {
      "id": "NIPA252512150.1",
      "metadata": {
        "title": "",
        "author": "",
        "publisher": "아주경제",
        "date": "",
        "topic": "IT/과학>AI/IT기기",
        "original_topic": "IT/과학",
        "speaker": []
      }
    },
    {
      "id": "NIPA252512150.2",
      "metadata": {
        "title": "",
        "author": "",
        "publisher": "아주경제",
        "date": "",
        "topic": "정치>정치/정치",
        "original_topic": "정치",
        "speaker": []
      }
    }
  ]
}
```

<표 30> 메타데이터 JSON 형식 예시(구어)

```
{
  "id": "SDPA252512150",
  "metadata": {
    "title": "국립국어원 병렬말뭉치 원천 자료 구어 SDPA252512150",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2025",
    "category": "사적 대화",
    "annotation_level": [
      "원시"
    ],
    "sampling": ""
  },
  "document": [
    {
      "id": "SDPA252512150.1",
      "metadata": {
        "title": "",
        "author": "",
        "publisher": "국립국어원",
        "date": "",
        "topic": "건강과미용",
        "original_topic": "2인 일상 대화",
        "speaker": [
          {
            "id": "SD2503279",
            "age": "60이상",
            "sex": "여"
          },
          {
            "id": "SD2597197",
            "age": "60이상",
            "sex": "여"
          }
        ]
      }
    }
  ],
  {
    "id": "SDPA252512150.2",
    "metadata": {
      "title": "",
      "author": "",
      "publisher": "국립국어원",
      "date": "",
      "topic": "사회",
    }
  }
}
```

```

    "original_topic": "2인 일상 대화",
    "speaker": [
      {
        "id": "SD2532099",
        "age": "60이상",
        "sex": "여"
      },
      {
        "id": "SD2529257",
        "age": "60이상",
        "sex": "여"
      }
    ]
  }
}

```

<표 31> 병렬 말뭉치 JSON 형식 예시(문어)

```

{
  "language_info": {
    "source_language": "ko",
    "target_language": "hi"
  },
  "parallel": [
    {
      "id": "NIPA252512150.1.1",
      "source": "외신들은 가전의 꽃인 TV와 미래 모빌리티를 우선 주목하는 가운데 '온디바이스 인공 지능(AI)'이 적용된 노트북에도 많은 관심을 내비쳤다.",
      "target": "वदिशी मीडिया सबसे प्रमुख घरेलू उपकरण TV और फ्यूचर मोबिलिटी पर पहले ध्यान दे रहे हैं, ऐसे में उन्होंने 'ऑन-डवाइस आर्टफिशियल इंटेलिजेंस (AI)' से लैस लैपटॉप में भी काफी रुचि दिखाई है।",
      "revision": {
        "revision1": "वदिशी मीडिया ने 'ऑन-डवाइस आर्टफिशियल इंटेलिजेंस (AI)' वाले लैपटॉप में भी काफी रुचि दिखाई जबकि TV और भविष्य की गतिशीलता घरेलू उपकरणों के फूल पर भी ध्यान दिया।",
        "revision2": "वदिशी मीडिया सबसे प्रमुख घरेलू उपकरण TV और फ्यूचर मोबिलिटी पर पहले ध्यान दे रहे हैं, ऐसे में उन्होंने 'ऑन-डवाइस आर्टफिशियल इंटेलिजेंस (AI)' से लैस लैपटॉप में भी काफी रुचि दिखाई है।"
      }
    },
    {
      "id": "NIPA252512150.1.2",
      "source": "7일 미국 IT·가전 매체 더 버지는 'CES 2024에서 기대할 수 있는 것'이라는 기사를 통해 CES 2024의 7대 트렌드를 예측했다.",
      "target": "7 तारीख को अमेरिकी IT · होम अप्लायंस मीडिया द वर्ज ने 'CES 2024 से

```

क्या उम्मीद की जा सकती है' शीर्षक वाले लेख में CES 2024 के सात प्रमुख रुझानों की भवषिय वाणी की।",

```
"revision": {
```

```
  "revision1": "7 तारीख को अमेरिकी IT · होम अप्लायंस द वर्ज ने 'CES 2024 से क्या उम्मीद की गई है' शीर्षक वाले लेख में CES 2024 के 7 प्रमुख रुझानों की भवषियवाणी की।",
```

```
  "revision2": "7 तारीख को अमेरिकी IT · होम अप्लायंस मीडिया द वर्ज ने 'CES 2024 से क्या उम्मीद की जा सकती है' शीर्षक वाले लेख में CES 2024 के सात प्रमुख रुझानों की भवषियवाणी की।"
```

```
}
```

```
}
```

```
}
```

### <표 32> 병렬 말뭉치 JSON 형식 예시(구어)

```
{  
  "language_info": {  
    "source_language": "ko",  
    "target_language": "hi"  
  },  
  "parallel": [  
    {  
      "id": "SDPA252512150.1.1",  
      "source": "이제, 내 몸이 아프다 보니까 살기 위해서 스트레칭을 시작을 했어요.",  
      "target": "अब, चूँकि मेरे शरीर में दर्द हो रहा है मैंने ज़िदा रहने के लिए स्ट्रेचिंग करना शुरू कर दिया।",  
      "revision": {  
        "revision1": "अब, क्योंकि मेरे शरीर में दर्द हो रहा है मैंने ज़िदा रहने के लिए स्ट्रेचिंग करना शुरू कर दी है।",  
        "revision2": "अब, चूँकि मेरे शरीर में दर्द हो रहा है मैंने ज़िदा रहने के लिए स्ट्रेचिंग करना शुरू कर दिया।"  
      }  
    },  
    {  
      "id": "SDPA252512150.1.2",  
      "source": "텔레비전에서 이 자세가 좋다 그러면 스트레칭을 더 끼워 넣고 또 끼워 넣어요.",  
      "target": "अगर टीवी पर कोई कहता है कियह पोजिशन अच्छा है, तो स्ट्रेचिंग को और भी ज्यादा जोड़ देती हूँ।",  
      "revision": {  
        "revision1": "यदि तुम्हें टीवी पर यह आसन पसंद आया तो इसे और अधिक स्ट्रेच करो।",  
        "revision2": "अगर टीवी पर कोई कहता है कियह पोजिशन अच्छा है, तो स्ट्रेचिंग को
```

```
        और भी ज्यादा जोड़ देती हूँ।"
    }
}
}
```

## 8. 보안

본 사업에서는 사업 수행 과정에 참여하는 인력, 장비, 자료 등에 대한 물리적·기술적 관리를 통해 안전하고 체계적인 운영이 이루어질 수 있도록 보안 관리 대상과 보안 관리 책임자를 지정하였다. 이에 따라 관련 보안 기준을 준수하며 사업 전반을 운영하였다.

### 8.1. 보안 관리 대상 및 담당자

#### 1) 보안 관리 대상

본 사업에서는 사업장과 과업 수행과 관련된 시설 보안, 정보 보안, PC 보안, 저장 매체 보안, 산출물 보안 등 물리적·기술적 요소와 함께 사업 참여 인력을 보안 관리 대상으로 설정하였다. 또한 사업의 성격상 추가적인 관리가 필요하다고 판단 되는 대상도 보안 관리 범위에 포함하여 규정하였다.

#### 2) 담당자 지정 체계

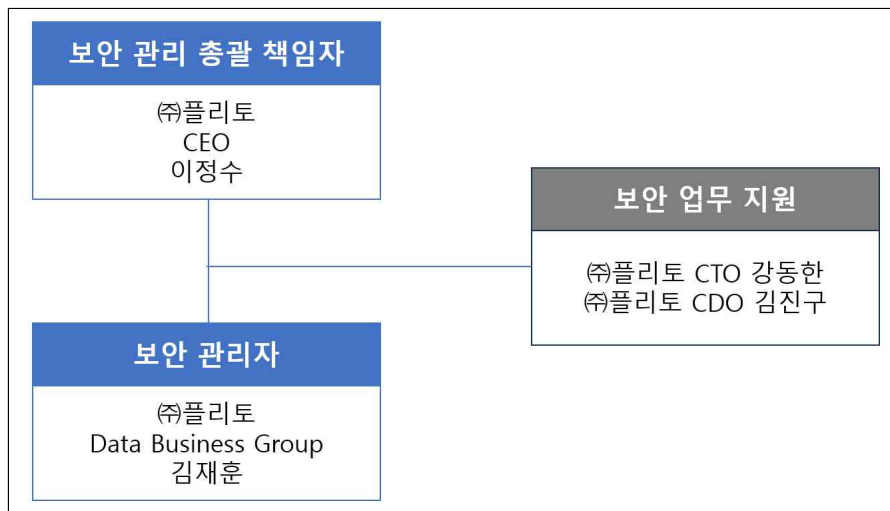
##### (1) 보안 관리 총괄 책임자

본 사업의 보안 관리에 대한 총괄 책임은 최종 데이터 산출을 담당하는 (주)플리토의 총괄 책임자에게 부여되었으며, 이에 따라 총괄 책임자는 사업 전반의 보안을 관리하기 위해 보안 관리자를 지정하여 운영하였다. 총괄 책임자는 본 사업과 관련된 인력, 장비, 자료 전반에 대한 보안 관리 업무를 총괄하였고, 보안 관리자에게 역할을 일부 위임하여 해당 업무를 수행하도록 하고 이를 감독하였다.

## (2) 보안 관리자

보안 관리 총괄 책임자는 보안 관리 활동의 원활한 수행과 통제를 위해 사업 관련 보안 관리 및 업무 수행과 연관된 인원을 보안 관리자로 지정하였다.

이에 따라 보안 관리 업무는 보안 관리자가 담당하였으며, 보안 관리자는 본 사업과 관련된 기술적·물리적 보안 조치를 책임지고 사업 수행 전반에 걸쳐 보안 활동을 실행, 점검 및 통제하는 역할을 수행하였다.



[그림 38] 보안 담당자 역할 체계

## 8.2. 보안 관리 방법

### 1) 물리적 보안

#### (1) 사무실·주요 장비 설치 장소에 대한 출입 보안

사무 공간은 CCTV와 잠금장치 등을 통해 상시 출입이 통제·관리되도록 운영하였으며 인가된 인원만 출입할 수 있도록 제한하였다. 공사나 시설 이전 등으로 불가피하게 비인가자의 출입이 필요한 경우에는 반드시 보안 관리 총괄 책임자의 사전 승인을 받은 후에 진행하도록 하였다. 또한 외부인이 방문하는 경우에는 경희대학교 산학협력단 또는 (주)플리토의 담당자가 외부인을 응대하도록 하였으며, 지정된 장소 또는 구역 내에서만 회의 및 협의를 진행하도록 관리하였다.

한편 데이터 가공·검수 및 저장이 이루어지는 핵심 공간인 서버실의 경우에도 출입이 필요한 자는 보안 관리 총괄 책임자의 사전 승인을 받은 경우에만 출입할 수

있도록 하였으며, 승인 후 출입한 인원과 출입 일시 등은 ‘서버실 출입 관리 대장’을 통해 별도로 기록·관리하였다.



[그림 39] 사무실 출입문 전경(좌: 경희대학교 산학협력단, 우: ㈜플리토)

### (2) PC 및 보조 기억 장치 반출·반입 통제

데이터 가공, 검수, 추출 등에 사용되는 모든 전산 장비(PC, 노트북 등)는 전산 장비 대장에 등록하여 관리하였으며, 인가되지 않은 장비의 사용은 원칙적으로 금지하였다. 또한 무선 랜, 외장형 HDD, USB 등 비인가 전산·통신 장비의 반입 및 사용을 금지하고, 관리 대상 장비의 반출입이 필요할 경우에는 자료의 무단 유출 여부를 점검한 뒤 관련 내용을 대장에 기록하도록 하였다.

PC와 노트북 등 주요 전산 장비는 정기적으로 보안 관리 항목을 점검하고 전체 점검 결과는 보안 관리자가 별도로 관리하도록 하였다. 아울러 전산 장비 대장을 마련하여 참여 인원의 장비 보유 현황을 체계적으로 관리하였으며, 보조 기억 장치는 보조 기억 매체 관리 대장을 통해 별도로 관리하였다.

<표 33> 보안 관리 사항

- |  |
|--|
| <ul style="list-style-type: none"> <li>• 자리 비움 시 화면 보호 기능 사용</li> <li>• 백신 프로그램 및 보안 업데이트 상시 수행</li> </ul> |
|--|

### (3) 보안 관리 대장

보안 관리는 매월 다음과 같은 5개 항목, 7개 분류에 대해 점검을 진행하였으며, 2025년 3월 2주 차부터 사업 종료 시까지 보안 점검을 실시하였다.

<표 34> 보안 점검 항목 및 분류

- 서버 보안: (주)플리토 개인 정보 암호화
- 네트워크 보안: 정보 처리 시스템 접근 통제, 시스템 접근 권한 관리
- 사업 수행 인력 관리: 전산 장비 보안
- 근무 환경 보안: 소프트웨어 패치 관리, 서버실 출입 관리
- 산출물 보안 관리

보안 관리 대장							
프로젝트 명: 2025 한·외 병렬 말뭉치 구축 사업				보안관리 책임자(정): 이정수 (플리토 CEO)			
문서번호: KFPC-007		V2.0		보안관리 책임자(부): 김재훈 (플리토 Data Business Group)			
※ 보안 관리 대장은 KFPC-008(2025년 한국어·외국어 병렬 말뭉치 구축사업 보안관리계획서) 문서를 바탕으로 함 ※ 보안 관리 대장은 월 1회 업데이트 됨(주관기관에서 요청한 보안관리 기간으로 설정함)							
NO	보안 항목 대분류	보안 항목 세부분류	참고	2025년 3월	2025년 4월	2025년 5월	2025년 6월
1	서버 보안	플리토 개인정보 암호화	KFPC-008 2.2. 나 개인정보 암호화	확인	확인	확인	확인
2	네트워크 보안	정보처리시스템 접근통제	KFPC-008 2.1 시스템관리 보안	확인	확인	확인	확인
3	네트워크 보안	시스템 접근권한 관리	KFPC-008 붙임3 계정 관리 대장	확인	확인	확인	확인
4	사업 수행 인력 관리	전산 장비 보안	보안 프로그램 설치 유무 확인 및 점검	확인	확인	확인	확인
5	근무 환경 보안	소프트웨어 패치 관리	KFPC-008 2.2. 라 소프트웨어 패치	확인	확인	확인	확인
6	근무 환경 보안	서버실 출입 관리	KFPC-008 붙임1. 서버실 출입 관리대장	확인	확인	확인	확인
7	산출물 보안 관리	산출물 보안 관리	데이터 파일 암호화	확인	확인	확인	확인
NO	보안 항목 대분류	보안 항목 세부분류	참고	2025년 7월	2025년 8월	2025년 9월	2025년 10월
1	서버 보안	플리토 개인정보 암호화	KFPC-008 2.2. 나 개인정보 암호화	확인	확인	확인	확인
2	네트워크 보안	정보처리시스템 접근통제	KFPC-008 2.1 시스템관리 보안	확인	확인	확인	확인
3	네트워크 보안	시스템 접근권한 관리	KFPC-008 붙임3 계정 관리 대장	확인	확인	확인	확인
4	사업 수행 인력 관리	전산 장비 보안	보안 프로그램 설치 유무 확인 및 점검	확인	확인	확인	확인
5	근무 환경 보안	소프트웨어 패치 관리	KFPC-008 2.2. 라 소프트웨어 패치	확인	확인	확인	확인
6	근무 환경 보안	서버실 출입 관리	KFPC-008 붙임1. 서버실 출입 관리대장	확인	확인	확인	확인
7	산출물 보안 관리	산출물 보안 관리	데이터 파일 암호화	확인	확인	확인	확인

[그림 40] 보안 관리 대장 예시

(4) 사업 수행 산출물 관리

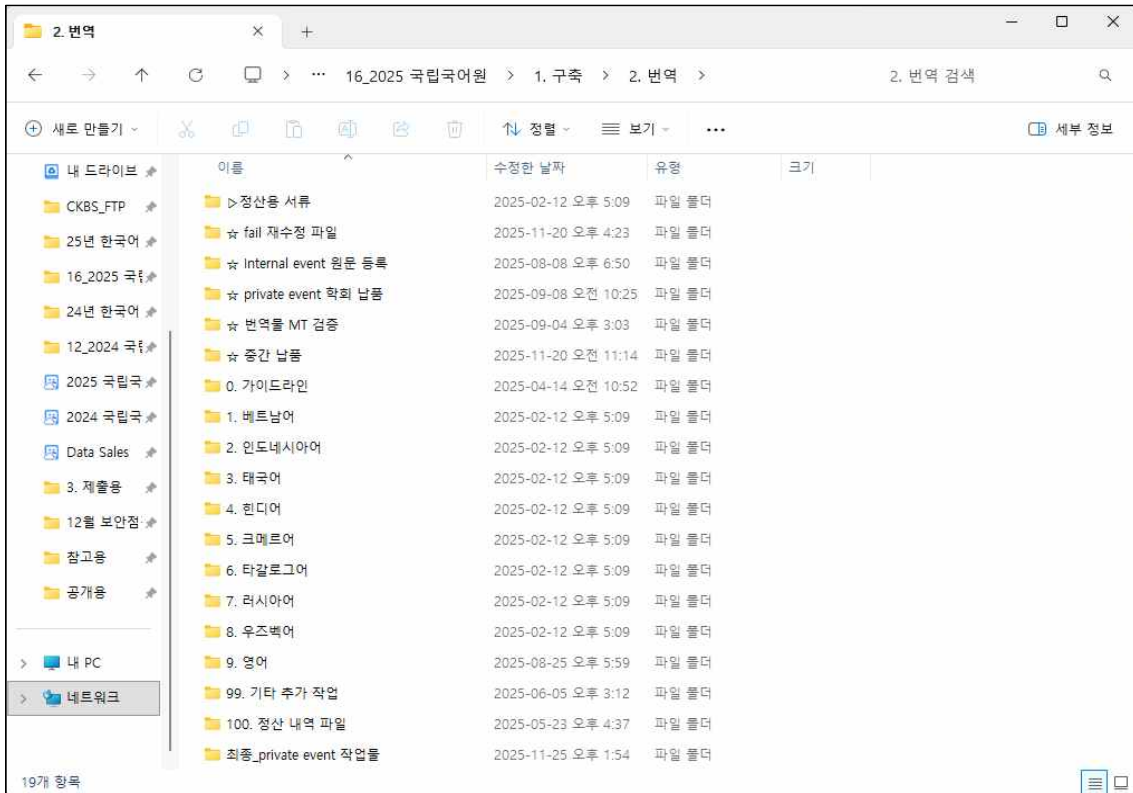
최종 산출물 도출 과정에서 생성되는 중간 결과물은 각 담당자의 PC 또는 서버에 분산 저장하였다. 최종 산출물의 경우에는 사업 수행 계획에 따라 총 3개의 보관 공간에 동일하게 이중·삼중 보관하였으며, 내·외부의 자료 요청이 있을 경우에는 반드시 보안 관리자를 통해서만 제공하도록 하였다.

<표 35> 최종 산출물 저장 공간

1차 저장 공간	데이터 세트 담당자(보안 관리자) PC
2차 저장 공간	NAS(Network Attached Storage)
3차 저장 공간	Cloud(Google Drive)



[그림 41] 저장 관리 예시(Cloud)



[그림 42] 저장 관리 예시(NAS)

국립국어원으로 자료를 송부해야 하는 경우에는 이메일을 통해 송·수신을 진행하였으며 전달되는 모든 자료는 암호화하여 안전성을 확보하였다.

사업 종료 후에는 보안 절차에 따라 최종 산출물을 제외한 사업 관련 산출물(중간 산출물 포함)을 모두 삭제하였고, 참여 인원 전원으로부터 자료 미보유 확인서를 받아 국립국어원에 제출하였다.

## 2) 기술적 보안

### (1) 시스템 관리 보안

서버를 포함한 정보 시스템에 대한 접근은 사전에 인가된 사용자만 가능하도록 제한하였으며, 정보 시스템 접근 권한은 보안 관리자의 대면 승인을 받은 후 할당된 단말기를 통해서만 부여하였다.

관리자 권한으로 접속할 수 있는 단말기는 별도로 운영하지 않았으며, 지정된 IP 기반으로 접근을 관리하고 휴대용 무선 모뎀이나 스마트폰 테더링 등 무선 인터넷 사용은 전반적으로 통제하였다.

또한 비인가자의 시스템 접속 시에는 관련 법령과 회사 내부 규정에 따라 엄격히 관리하였다.

<표 36> 시스템 접근 통제 관련 법률 및 규정

<b>관련 법률</b>	소프트웨어진흥법 (시행일 2025. 4. 23.) 통신비밀보호법 (시행일 2025. 8. 1.) 전기통신기본법 (시행일 2019. 6. 25.) 정보통신산업 진흥법 (시행일 2025. 10. 1.) 지능정보화 기본법 (시행일 2025. 3. 27.)
<b>회사 관련 규정</b>	(주)플리토 전산 관리 규정 (제정 2019.10.24.)

서버에 접속할 수 있는 모든 사용자는 계정 단위로 관리하고, 사용 목적에 따라 계정 유형을 구분하였으며, 서버 접근이 가능한 계정은 계정 관리 대장을 통해 체계적으로 관리하였다. (주)플리토 플랫폼의 경우에도 사용자별로 개별 ID를 부여하여 접속 경로를 제한적으로 운영함으로써 정보 유출 가능성을 최소화하였다. 각 계정의 비밀번호는 최대 60일 이내로 변경하도록 주기를 설정하였으며, 비밀번호 공유는 금지하고 철저히 비밀이 유지되도록 관리하였다.

아울러 보안 관리 책임자는 외부 업체 등에 기술 지원을 의뢰하는 경우에도 온라인을 통한 원격 기술 지원은 원칙적으로 허용하지 않도록 하였으며, 불가피하게 원격 지원이 필요한 경우에는 필요한 보안 대책을 충분히 마련한 후에만 제한적으로 허용하고, 해당 내역을 반드시 기록으로 남기도록 하였다.

## (2) 시스템 구축 보안

시스템 구축과 관련한 보안 정책은 (주)플리토의 기존 개발·운영 환경에 적용되고 있는 보안 정책을 준용하였다.

클라이언트와 웹 서버 간 데이터 전송 시에는 SSL 방식의 암호화를 적용하였으며, 개인 정보를 처리·관리하는 개인 정보 처리 시스템의 경우 DB에 저장된 개인 정보를 암호화하여 보관함으로써 정보의 유출, 위·변조 및 훼손을 방지하도록 하였다. DB 접근 권한은 (주)플리토의 내부 운영 정책에 따라 관리하되, 개발자별로 접속 권한을 구분하여 통제하였다.

또한 보안 취약점 개선을 위해 운영 체제와 응용 프로그램에 대한 보안 패치를 지속적으로 반영하였으며, 보안 패치는 사업 참여 인력의 전산 장비를 대상으로 관리하고 신규 패치나 업그레이드가 발표될 경우 관련 파일을 즉시 적용하여 시스

템을 최신 상태로 유지하였다.

### 3) 참여 인력 보안

#### (1) 사업 수행 인력 관리

본 사업 추진 과정에서 연구 책임자는 사업 시작 시 대표 명의의 보안 서약서와 비밀 유지 계약서를 제출하였으며, 연구 참여자 전원 역시 보안 서약서를 제출하였다. 인력 변동이 발생한 경우에도 신규 참여자를 대상으로 동일한 절차를 준수하였다. 또한 사업 참여자를 대상으로 전산 장비 반입·반출 시 악성 코드 감염 여부 및 중요 자료의 무단 반출 가능성을 확인한 후 장비 반출입 대장을 작성하여 체계적으로 관리하였다. 사업 종료 시에는 연구 책임자가 대표 명의 보안 확인서를 제출하였으며, 연구 참여자 전원은 자료 미보유 확인서와 보안 점검 결과서를 각각 제출하여 보안 관리 절차를 최종적으로 이행하였다.

#### (2) 보안 교육 실시

사업 참여자 전원을 대상으로 아래와 같이 보안 교육을 실시하였다.

<표 37> 사업 참여자 보안 교육

구분	내용
교육 방법	<ul style="list-style-type: none"> <li>• 보안 교육은 사업 기간 내 정기 교육으로 1회 이상 실시함.</li> <li>• 교육 방식은 보안 교육 자료를 활용하여 세미나 형식으로 진행함.</li> </ul>
교육 대상	<ul style="list-style-type: none"> <li>• 본 사업 참여자 전체</li> <li>• 정기 교육 미이수자는 교육일로부터 90일 이내에 보안 관리자에 의해 전달 교육함.</li> </ul>
교육 내용	<ul style="list-style-type: none"> <li>• 사업 수행에 대한 보안 및 개인 정보 보호 등 전반적인 사항</li> <li>• 법적 또는 발주 기관 규정에 따른 비밀 유지 의무 준수, 위반 시 처벌 내용, 누출 금지 대상 정보 및 정보 누출 시 부정 당업자 제재 조치 등 본 사업에 필요하다고 요구되는 내용</li> </ul>
교육 기록	<ul style="list-style-type: none"> <li>• 교육 이후 보안 교육 참석자 명단으로 관리</li> </ul>

#### 4) 월별 보안 점검

(※)플리토에서 시행하는 보안 관리와 별도로 경희대학교 산학협력단에서는 매일 관리적·물리적 보안 점검을 실시하고 각 PC의 보안 설정을 점검하였다.

점검 항목에는 최신 보안 패치 적용, 기본 보안 설정 준수, 인터넷 및 외부 접속 통제, 비인가 무선 AP 금지 등 핵심 보안 조치의 이행 여부가 포함되었다. 업무 자료는 지정된 PC에 저장하도록 관리하였으며, 인터넷 접속 전산 장비에는 백신, 보안 평가 프로그램, 네트워크 제어 등 필수 보안 정책을 적용하였다. PC 보안은 ‘AhnLab EPP Security Assessment’ 프로그램을 활용하여 정기 자가 점검을 수행하고 미흡 사항은 즉시 조치하였다.

월별 보안 점검 결과, 전반적인 보안 관리 체계가 안정적으로 운영되고 있으며, 모든 핵심 보안 조치가 지속적으로 준수되고 있음을 확인하였다.

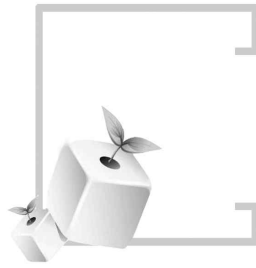
<표 38> 월별 보안 점검 실시 개요

날짜	보안 활동 내용	결과
2025.3.28.	보안 점검(종합, 정기, PC)	전원 양호
2025.4.7.	정보 보안 교육	직원 33명 완료
2025.4.25.	보안 현장 점검	전원 양호
2025.5.9.	정보 보안 교육	신규 직원 1명 완료
2025.5.30.	보안 점검(종합, 정기, PC)	전원 양호
2025.6.30.	보안 점검(종합, 정기, PC)	전원 양호
2025.7.31.	보안 점검(종합, 정기, PC)	전원 양호
2025.8.29.	보안 점검(종합, 정기, PC)	전원 양호
2025.9.30.	보안 점검(종합, 정기, PC)	전원 양호
2025.10.31.	보안 점검(종합, 정기, PC)	전원 양호
2025.11.28.	보안 점검(종합, 정기, PC)	전원 양호
2025.12.15.	정기 보안 점검 및 PC 점검, 보안 점검(종합)	전원 양호

보안 교육은 본 사업의 특성에 부합하도록 공공 기관용·공개용 보안 교육 영상 시청과 보안 담당자 강의를 병행하는 방식으로 실시하였다. 교육 내용에는 내부 정보 보안의 중요성, 내부 정보 유출 예방 방안, 개인 정보 보호 및 데이터 보안에 관한 사항뿐 아니라, 관련 법령 및 발주 기관 규정에 따른 비밀 유지 의무, 위반 시의 처벌 규정, 누출 금지 대상 정보, 정보 누출 발생 시 적용되는 부정 당업자 제재 조치 등이 포함되었다.



[그림 43] 보안 교육 예시



## 제 3 장

### 사업 수행 결과





## 1. 말뭉치 데이터 구축 결과

### 1.1. 최종 구축 데이터

본 사업에서는 언어별 100만 어절, 총 800만 어절의 한국어-외국어 병렬 말뭉치 구축을 목표로 하였다. 단계별로 살펴보면, 번역은 총 8,015,968어절, 검수는 총 8,000,000어절, 감수는 804,593어절로 목표량의 100% 이상을 달성하였다. 추가 제안 사항이었던 한국어-영어 병렬 말뭉치도 총 1,001,996어절로 100.20% 구축을 완료하였다. 최종 구축 데이터의 상세한 수량은 다음과 같다.

<표 39> 최종 구축 데이터 수량(어절)

구분		번역		검수		감수	
언어	목표	어절	비율	어절	비율	어절	비율
베트남어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	100,304	100.30%
인도네시아어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	101,251	101.25%
태국어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	101,594	101.59%
인도 힌디어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	100,764	100.76%
캄보디아 크메르어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	100,143	100.14%
필리핀 타갈로그어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	100,476	100.47%
러시아어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	100,001	100.00%
우즈베크어	1,000,000	1,001,996	100.20%	1,000,000	100.00%	100,060	100.06%
<b>합계</b>	<b>8,000,000</b>	<b>8,015,968</b>	<b>100.20%</b>	<b>8,000,000</b>	<b>100.00%</b>	<b>804,593</b>	<b>100.57%</b>
영어 (추가 제안)	1,000,000	1,001,996	100.20%				

<표 40> 최종 구축 데이터 수량(문장)

구분	문어	구어	합계
베트남어	24,105	51,036	75,141
인도네시아어	24,105	51,036	75,141
태국어	24,105	51,036	75,141
인도 힌디어	24,105	51,036	75,141
캄보디아 크메르어	24,105	51,036	75,141
필리핀 타갈로그어	24,105	51,036	75,141
러시아어	24,105	51,036	75,141
우즈베크어	24,105	51,036	75,141
<b>합계</b>	<b>192,840</b>	<b>408,288</b>	<b>601,128</b>
영어 (추가 제안)	24,105	51,036	-

## 1.2. 데이터 품질 감리 결과

사업 수행 과정에서 감리 업체가 한국어-외국어 병렬 말뭉치의 품질을 점검하였다. 품질 점검 감리는 감수까지 진행된 중간 납품 데이터 언어별 15,000어절(총 135,000어절)을 대상으로 시행되었다. 감리 결과가 언어별 전체 데이터의 적합률이 99.9~100%인 것으로 나타나 ‘고품질 언어 데이터 구축’이라는 사업 목표를 달성한 것을 확인하였다. 구체적으로는 문어는 54,000어절 중 53,994어절(99.99%), 구어는 81,000어절 중 80,989어절(99.99%)이 적합 판정을 받았으며, 이를 합산하면 135,000어절 중 134,983어절이 적합 판정을 받아 전체적으로 99.99%의 적합률을 보였다. 데이터 품질 감리 결과의 상세한 내용은 다음과 같다.

<표 41> 데이터 품질 감리 결과

구분		문어		구어		전체	
언어	어절	적합	적합률	적합	적합률	적합	적합률
베트남어	15,000	6,000	100.00%	9,000	100.00%	15,000	100.00%
인도네시아어	15,000	6,000	100.00%	9,000	100.00%	15,000	100.00%
태국어	15,000	6,000	100.00%	8,999	99.99%	14,999	99.99%
인도 힌디어	15,000	6,000	100.00%	8,999	99.99%	14,999	99.99%
캄보디아 크메르어	15,000	5,999	99.98%	9,000	100.00%	14,999	99.99%
필리핀 타갈로그어	15,000	5,999	99.98%	8,998	99.98%	14,997	99.98%
러시아어	15,000	5,999	99.98%	8,995	99.94%	14,994	99.96%
우즈베크어	15,000	5,998	99.97%	8,998	99.98%	14,996	99.97%
영어	15,000	5,999	99.98%	9,000	100.00%	14,999	99.99%
합계	135,000	53,994	99.99%	80,989	99.99%	134,983	99.99%

## 1.3. 번역 품질 비교

최근 인공지능 기술의 발전과 함께 기계 번역의 전반적인 성능도 지속적으로 향상되고 있다. 그러나 관용 표현, 동음이의어·다의어, 문장 성분의 생략과 같이 문맥 의존도가 높은 요소들은 여전히 정확한 해석이 어려워 오역이 자주 발생하는 실정이다. 특히 대규모 고품질 데이터가 확보될수록 맥락 추론의 정교함이 번역 성능에 큰 영향을 미치게 되는데, 이 때문에 저자원 언어의 경우 상대적으로 번역

정확도가 낮게 나타날 수밖에 없다.

본 사업에서는 고품질의 병렬 말뭉치 구축에 목표를 두고 여러 단계의 검증 절차를 통해 번역 품질을 체계적으로 확보하였다. 현재 대중적으로 널리 사용되는 두 개의 기계 번역 플랫폼(구글 번역, 네이버 파파고)과 본 사업의 번역 결과를 비교하기 위해, 언어별로 문어 3개, 구어 3개의 사례를 선정하여 이러한 품질 관리 활동의 효과를 확인하였다.

기계 번역에서는 문맥을 충분히 해석하지 못해 동음이의어나 다의어를 잘못 처리하거나, 관용·비유 표현·신조어를 축자적 의미로 번역해 왜곡되는 경우가 있었다. 원문의 ‘물이 차다(물이 불어 차오르다)’를 ‘물이 차갑다’로 해석하거나, ‘오이를 담다’를 ‘오이를 따다’로 번역하는 등 기본적인 문맥 해석 오류가 반복적으로 나타났다. 또한 교육 체계에 대한 이해 부족으로 ‘여상(여자상업고)’을 ‘여자 대학’으로, ‘인문계(일반계 고교)’를 ‘인문학 전공’으로 번역하는 등 한국의 제도적·문화적 맥락을 고려하지 못한 오역도 다수 발견되었다.

또한, 사회문화적 이해가 수반되어야 하는 줄임말이나 신조어의 경우에도 오역이 발생하였다. ‘야자(야간 자율 학습)’는 학교에서 이루어지는 저녁 자율 학습을 의미하지만, 기계 번역에서는 이를 ‘야자수’와 관련된 표현으로 잘못 인식하거나 의미 없는 음차 형태로 처리해 원문의 교육적 맥락을 전혀 전달하지 못하였다. 또한 부동산 신조어인 ‘영끌(가용 자금을 모두 끌어모아 투자한다는 뜻)’ 역시 기계 번역에서는 ‘취업하다’와 같이 전혀 관련 없는 의미로 번역되거나 단순히 발음을 옮긴 표기로 제시되어 원래의 사회·경제적 함의를 크게 왜곡하였다.

반면에 본 사업에서 구축한 병렬 말뭉치에서는 앞서 제시한 어휘나 표현들이 한국어 원문의 맥락을 적절히 전달할 수 있도록 번역되어 있다. 언어별로 기계 번역과 병렬 말뭉치의 번역을 비교한 내용을 상세히 제시하면 다음과 같다.

<표 42> 기계 번역과 병렬 말뭉치의 번역 비교(베트남어)

원문	문어1	한 총리는 '전공의 집단행동으로 국민들이 진료를 보는 데 불편을 겪고 있는 상황에서 적십자 병원의 비상 진료 체계 운영에 감사드립니다'고 전했다.
번역	기계 번역①	Thủ tướng Hàn Quốc phát biểu: "Tôi muốn bày tỏ lòng biết ơn tới Bệnh viện Chữ thập đỏ vì hệ thống cấp cứu y tế của bệnh viện đang hoạt động trong tình huống người dân gặp bất tiện trong việc tiếp nhận điều trị y tế do hành động tập thể của các bác sĩ nội trú." (The Korean Prime Minister stated: 'I want to express gratitude to the Red Cross Hospital in a situation where citizens face inconvenience in receiving medical treatment due to the collective action of resident doctors.)

	<b>기계 번역②</b>	<p>Thủ tướng Han nói: "Tôi cảm ơn việc vận hành hệ thống điều trị khẩn cấp của bệnh viện Chữ thập đỏ trong bối cảnh người dân đang gặp bất tiện trong việc điều trị do hành động tập thể của chuyên ngành."</p> <p>(Prime Minister Han said: 'I thank the operation of the emergency treatment system in the context that citizens are facing inconvenience in treatment due to the collective action of the specialty.)</p>
	<b>병렬 말뭉치</b>	<p>Thủ tướng Han đã nói "Cám ơn Bệnh viện Hội Chữ thập đỏ vì đã vận hành hệ thống y tế khẩn cấp trong tình hình người dân gặp bất tiện trong điều trị y tế do hành động tập thể của các bác sĩ chuyên khoa".</p> <p>(Prime Minister Han said: 'Thank you to the Red Cross Hospital for operating the emergency medical system in a situation where citizens face inconvenience in medical treatment due to the collective action of specialist doctors.)</p>
	<b>비교</b>	<p>기계 번역①에서는 '한 총리'를 성씨가 아닌 '한국의 총리'의 약어로 오역하였으며, '감사드린다'는 서술어 또한 불필요하게 장황하게 번역하여 문체의 간결성을 해쳤음.</p> <p>기계 번역②에서는 인명은 올바르게 처리했으나, '전공의(사람)'를 일반 명사 '전공(학문 분야)'의로 잘못 식별하여 문맥이 성립하지 않는 비문이 되었음.</p> <p>병렬 말뭉치에서는 고유명사(한 총리)를 정확히 식별하였고, '전공의'를 '전문의'로 번역한 미세한 차이는 있으나 행위의 주체가 의사임을 명확히 하여 문맥을 자연스럽게 살렸음.</p>
<b>원문</b>	<b>문어2</b>	<p>지난 7일 KBS가 '여자 친구 살해 20대, 의대 재학생으로 확인'이라는 제목으로 20대 남성 A 씨가 대학 수학 능력 시험에서 만점을 받았다고 보도했다.</p>
<b>번역</b>	<b>기계 번역①</b>	<p>Vào ngày 7, KBS đưa tin với tiêu đề "Người đàn ông ngoài 20 tuổi giết bạn gái được xác nhận là sinh viên trường y", rằng một người đàn ông ngoài 20 tuổi, được xác định là A, đã đạt điểm tuyệt đối trong bài kiểm tra năng lực học thuật của trường đại học.</p> <p>(On the 7th, KBS reported with the title 'Man over 20 who killed girlfriend confirmed to be med student', that a man over 20, identified as A, achieved a perfect score on the university academic ability test.)</p>
	<b>기계 번역②</b>	<p>Ngày 7/5, KBS đã đưa tin nam sinh độ tuổi 20 A đã đạt được điểm tuyệt đối trong kỳ thi năng lực toán học đại học với tiêu đề "Giết bạn gái ở độ tuổi 20 xác nhận là sinh viên y khoa".</p> <p>(On May 7th, KBS reported that male student A in his 20s achieved a perfect score in the university math capacity exam with the title 'Killing girlfriend at age 20 confirmed to be medical student'.)</p>
	<b>병렬</b>	<p>Vào ngày 7 vừa qua, KBS đưa tin rằng một nam thanh niên</p>

	<b>말뭉치</b>	<p>trong độ tuổi 20 đã đạt điểm tuyệt đối trong kỳ thi tuyển sinh đại học với tiêu đề "Nam thanh niên độ tuổi 20 giết bạn gái, được xác nhận là sinh viên đang theo học đại học y".</p> <p>(On the past 7th, KBS reported about a young man in his 20s, who achieved a perfect score in the university entrance exam, with the title 'Male youth in his 20s who killed girlfriend, confirmed to be a student studying at medical university'.)</p>
	<b>비교</b>	<p>기계 번역①에서는 '20대 남성'을 대학생이라는 맥락에 적합한 '청년'이 아닌 '성인 남자/아저씨'로 번역하여 부자연스러우며, 원문의 문장 구조를 그대로 직역하여 긴 기사 제목이 서술어 앞에 위치함으로써 가독성 또한 떨어짐.</p> <p>기계 번역②에서는 '수학 능력 시험'을 '수학(數學) 능력'으로 잘못 인식하였으며 원문에 존재하지 않는 '5월'이라는 날짜 정보를 임의로 첨가하는 오류를 범했음.</p> <p>병렬 말뭉치에서는 '수학 능력 시험'을 베트남의 교육 환경에 맞춰 이해하기 쉬운 '대학 입학 시험'으로 적절히 의역하였으며, 문장 구조를 자연스럽게 재배치하여 정보 전달력을 높였음.</p>
<b>원문</b>	<b>문어3</b>	<p>물 산업 매출액은 2022년 49조 6902억 원으로 2021년(47조 4220억 원)보다 4.8% 증가했고, GDP 2150조 6000억 원 대비 2.3% 수준이다.</p>
<b>번역</b>	<b>기계 번역①</b>	<p>Doanh số bán hàng của ngành nước năm 2022 đạt 49,6902 nghìn tỷ KRW, tăng 4,8% so với mức 47,422 nghìn tỷ KRW năm 2021 và chiếm 2,3% GDP (2.150,6 nghìn tỷ KRW).</p> <p>(Sales of the water sector in 2022 reached 49.6902 thousand billion KRW, an increase of 4.8% compared to the level of 47.422 thousand billion KRW in 2021 and accounting for 2.3% of the GDP (2,150.6 thousand billion KRW).)</p>
	<b>기계 번역②</b>	<p>Doanh thu của ngành công nghiệp hàng hóa là 49,69 nghìn tỷ won vào năm 2022, tăng 4,8% so với năm 2021 (47,42 nghìn tỷ won) và 2,3% so với GDP 2150,6 nghìn tỷ won.</p> <p>(Revenue of the goods industry was 49.69 thousand billion won in 2022, an increase of 4.8% compared to 2021 (47.42 thousand billion won) and 2.3% compared to the GDP of 2150.6 thousand billion won.)</p>
	<b>병렬 말뭉치</b>	<p>Doanh thu của ngành công nghiệp nước năm 2022 đạt 49.690,2 tỷ KRW, tăng 4,8% so với năm 2021 (47.422 tỷ KRW) và tương đương 2,3% so với GDP là 2.150.600 tỷ won.</p> <p>(Revenue of the water industry in 2022 reached 49,690.2 billion KRW, an increase of 4.8% compared to 2021 (47,422 billion KRW) and equivalent to 2.3% compared to a GDP of 2,150,600 billion won.)</p>
	<b>비교</b>	<p>기계 번역①에서는 '물 산업'을 포괄적인 '물 분야'로 번역하여 문어 말뭉치 맥락에 맞지 않게 해석되었으며, 숫자 표기에 있어 '49,6902 천억'이라는 비표준적인 방식을 사용하여 수치의 정확성과 가독성이 떨어짐.</p>

		<p>기계 번역②에서는 '물'을 '물품'으로 오역하였으며 원문의 '6902억'을 가독성을 위해 '6900억'으로 절삭하여 원문의 구체적인 수치 정보를 누락하였음.</p> <p>병렬 말뭉치에서는 '물 산업'을 전문용어로 정확히 번역하였고, 한국식 숫자 단위를 베트남어 표기법에 맞춰 정확한 수치를 전달하였음.</p>
원문	구어1	아파트 수리를 많이 해야 되니까 굳이 수리를 해 가면서 거기 들어가서 살 필요가 있나 싶어.
번역	기계 번역①	<p>Căn hộ cần sửa chữa nhiều nên tôi tự hỏi liệu có thực sự cần thiết phải sống ở đó trong khi sửa chữa không.</p> <p>(The apartment needs a lot of repairs, so I wonder if it is truly necessary to live there while repairing it.)</p>
	기계 번역②	<p>Vì phải sửa chữa nhiều chung cư nên tôi nghĩ có nhất thiết phải vừa sửa vừa vào đó sống không.</p> <p>(Because the apartment needs many repairs, I think 'is it necessary to repair and live there at the same time?')</p>
	병렬 말뭉치	<p>Vì phải sửa chữa nhiều cho căn chung cư nên chị tự hỏi có nhất thiết cần phải sửa chữa để vào đó sống không.</p> <p>(Because the apartment unit needs a lot of repairs, I wonder if it is strictly necessary to repair it in order to move in and live there.)</p>
	비교	<p>기계 번역①에서는 '수리를 해 가면서'라는 구문을 '수리하는 기간 동안(while)'이라는 물리적 시간의 의미로 번역하여 오역이 되었음.</p> <p>기계 번역②에서는 이를 '수리와 거주를 동시에 병행하다(at the same time)'는 병렬 동작의 의미로 어색하게 번역되었으며, 원문의 '굳이'가 내포한 '번거로움을 감수하다'라는 뉘앙스를 살리지 못함.</p> <p>병렬 말뭉치에서는 문맥상 '거주를 목적으로 수고를 들이다'라는 인과 관계의 의미로 잘 살려서 번역하였음.</p>
원문	구어2	언니 학교는 야자 활동을 무조건적으로 안 해도 됐었어?
번역	기계 번역①	<p>Chị ơi, chị có nhất thiết phải cấm các hoạt động liên quan đến cây cọ ở trường không?</p> <p>(Did you necessarily have to ban activities related to palm trees at school?)</p>
	기계 번역②	<p>Trường của chị là Yaja không nhất thiết phải hoạt động à?</p> <p>(Is it not necessary to operate Yaja at your school?)</p>
	병렬 말뭉치	<p>Trường chị, không bắt buộc tự học buổi tối cũng được á hả?</p> <p>(At your school, it was okay for evening self-study not to be mandatory?)</p>
	비교	<p>기계 번역①에서는 '야자(야간 자율 학습)'라는 한국어 줄임말을 '야자수'로 잘못 인식하여 학교에서 야자수 관련 활동을 금지하냐는 엉뚱한 의미로 오역이 되었음.</p> <p>기계 번역②에서는 '야자'의 의미를 파악하지 못해 발음 그대로 표기하거나 고유명사로 처리하였으며, 문장의 서술어 구조 또한 파악하지 못해 비문이 되었음.</p>

		병렬 말뭉치에서는 '야자'를 문맥에 맞게 '저녁 자을 학습'으로 정확히 풀어서 해석하였으며, '안 해도 됐다'는 허용의 뉘앙스를 '강제성이 없다'는 표현으로 자연스럽게 잘 살려서 번역하였음.
원문	구어3	근데 너 게임을 하면서 이제 그, 게임에 현질을 좀 한 적 있니?
번역	기계 번역①	Nhưng bạn đã bao giờ nghiện trò chơi điện tử khi chơi chúng chưa? (But have you ever been addicted to video games while playing them?)
	기계 번역②	Nhưng mà cậu có bao giờ vừa chơi game vừa nhận tiền không? (But have you ever received money while playing games?)
	병렬 말뭉치	Nhưng mà khi cậu chơi game thì ừm, cậu từng nạp tiền vào game bao giờ chưa? (But when you play games, um, have you ever topped up money into the game?)
비교		기계 번역①에서는 '현질'이라는 신조어를 문맥과 전혀 관계없는 '중독'이라는 의미로 추리하여 오역이 되었음. 기계 번역②에서는 돈을 쓰는 것이 아닌 반대로 '돈을 받다'라는 의미로 잘못 번역하였으며, 이는 원문의 의도와 정반대의 해석이 되었음. 병렬 말뭉치에서는 게임 용어인 '현질'을 베트남어의 상응하는 표현인 '돈을 충전하다'로 정확히 번역하였음.

<표 43> 기계 번역과 병렬 말뭉치의 번역 비교(인도네시아어)

원문	문어1	김정훈 줄라 코리아 대표는 전 국가대표 탁구 선수 출신으로, 현역 시절에 바른 세상 병원에서 꾸준히 치료를 받았던 인연이 해당 대회까지 이어진 것으로 알려졌다.
번역	기계 번역①	Kim Jeong-hoon, perwakilan Jula Korea, adalah mantan pemain tenis meja nasional, dan diketahui bahwa hubungannya dengan Barun World Hospital selama tahun-tahun aktifnya berlanjut hingga kompetisi. (Kim Jeong-hoon, the representative of Jula Korea, is a former national table tennis player, and it is known that his relationship with Barun World Hospital during his active years continued until the competition.)
	기계	Perwakilan Korea Kim Jong-hun Julia berasal dari pemain

	<b>번역②</b>	ping pong yang pernah menjadi perwakilan nasional, karena hubungan yang terus-menerus dirawat di rumah sakit dunia yang tepat pada saat saat saat ini, dikatakan telah menyebabkan kompetisi tersebut. (Korean representative Kim Jong-hun Julia comes from a ping pong player who was once a national representative, because a relationship that was constantly treated in the right world hospital at the present time, is said to have caused the competition.)
	<b>병렬 말뭉치</b>	CEO JOOLA KOREA, Kim Jung Hoon, merupakan mantan atlet nasional tenis meja yang selama masa aktifnya secara rutin menjalani perawatan di Rumah Sakit Bareun Sesang dan hubungan tersebut diketahui berlanjut hingga ajang turnamen ini. (CEO of JOOLA KOREA, Kim Jung Hoon, is a former national table tennis athlete who during his active period routinely underwent treatment at Bareun Sesang Hospital and this relationship is known to continue until this tournament event.)
	<b>비교</b>	기계 번역①의 'perwakilan Jula Korea'는 '줄라 코리아 대표'의 의미를 전달하지만, 병렬 말뭉치의 'CEO JOOLA KOREA'가 직책을 더 정확히 표현하며 공식 문서에 적합하며, 기계 번역②는 주어와 서술어의 연결이 매우 어색하고('Perwakilan Korea Kim Jong-hun Julia berasal dari pemain ping pong...'), 시제와 어순이 엉켜 있어 원문의 의미를 파악하기 어려움. 또한, 기계 번역①의 'Barun World Hospital'은 '바른 세상 병원'을 영어로 직역한 것으로, 고유명사 표기로는 부적절하며 이에 반해 병렬 말뭉치는 고유명사를 그대로 음차하여 적절하게 처리함.
<b>원문</b>	<b>문어2</b>	왕이 외교부장은 '한반도 문제 처방전은 이미 준비돼 있다'며 '바로 중국이 제시한 쌍궤병진(雙軌並進·비핵화와 북미 평화 협정 동시 추진)과 단계적·동시적 원칙'이라고 기존의 원론적 입장을 재차 내놓았다.
<b>번역</b>	<b>기계 번역①</b>	Menteri Luar Negeri Wang Yi menyatakan, "Resep untuk masalah Semenanjung Korea telah disiapkan," dan menegaskan kembali posisi berprinsip yang ada, dengan mengatakan, "Ini adalah kemajuan paralel dua jalur (upaya

		<p>simultan untuk denuklirisasi dan perjanjian damai antara AS dan Korea Utara) dan prinsip bertahap dan simultan yang diusulkan oleh Tiongkok."</p> <p>(Foreign Minister Wang Yi stated, "The recipe for the Korean Peninsula issue has been prepared," and reiterated the existing principled position, saying, "This is the parallel progress of the two tracks (simultaneous efforts for denuclearization and a peace agreement between the US and North Korea) and the principle of gradual and simultaneous proposed by China.")</p>
	<p>기계 번역②</p>	<p>Menteri Luar Negeri Wang Yi menyatakan bahwa "resep untuk masalah Semenanjung Korea sudah disiapkan," dan kembali mengulangi posisi prinsip yang ada, yaitu "prinsip paralel ganda yang diusulkan oleh Tiongkok (mendorong denuklirisasi dan perjanjian damai Korea Utara-Amerika Serikat secara bersamaan) dan prinsip bertahap dan simultan."</p> <p>(Foreign Minister Wang Yi stated that "the recipe for the Korean Peninsula issue is ready," and reiterated the existing principled positions, namely "the dual parallel principle proposed by China (promoting denuclearization and a North Korea-United States peace agreement simultaneously) and the gradual and simultaneous principle.")</p>
	<p>병렬 말뭉치</p>	<p>Menteri Luar Negeri, Wang Yi, menyatakan, "Resep untuk masalah Semenanjung Korea sudah disiapkan," dan menambahkan, "Pendekatan jalur ganda yang diajukan Tiongkok (denuklirisasi dan perjanjian damai antara Amerika Serikat dan Korea Utara yang dilakukan secara bersamaan) serta prinsip bertahap dan simultan," sambil kembali mengemukakan posisi prinsip yang telah diambil sebelumnya.</p> <p>(Foreign Minister Wang Yi stated, "The recipe for the Korean Peninsula issue is ready," adding, "China's dual-track approach (simultaneous denuclearization and a peace agreement between the United States and North Korea) and the principle of gradualism and simultaneity," while reiterating previously adopted principled positions.)</p>

<b>비교</b>		<p>기계 번역①, ②는 핵심 용어인 '쌍궤병진(雙軌竝進)'을 번역하는 데 기술적 문제와 용어 선택의 부정확함을 보임. 기계 번역①의 'kemajuan paralel dua jalur'은 외교 정책 용어로서의 적합성이 낮고 기계 번역②의 'prinsip paralel ganda'는 문장 구조가 다소 어색함.</p> <p>병렬 말뭉치는 '쌍궤병진'을 'Pendekatan jalur ganda'로 번역하여 정치·외교 용어로서의 의미를 가장 정확히 전달했으며, 문장 연결이 자연스럽고 공식적인 논조를 가장 잘 살림.</p>
<b>원문</b>	<b>문어3</b>	<p>시행령에서 이러한 부당 이득의 산정 방식을 실현 이익·미실현 이익과 회피 손실액으로 구분해 불공정 거래 행위 유형별(미공개 중요 정보 이용·시세 조종·부정 거래)로 구체적으로 규정했다.</p>
<b>번역</b>	<b>기계 번역①</b>	<p>Keputusan penegakan tersebut secara khusus menetapkan metode penghitungan keuntungan tidak adil tersebut dengan membaginya menjadi keuntungan yang direalisasikan, keuntungan yang belum direalisasikan, dan kerugian yang dihindari, dan berdasarkan jenis praktik perdagangan tidak adil (penggunaan informasi penting yang tidak diungkapkan, manipulasi harga, dan perdagangan tidak adil).</p> <p>(The enforcement decision specifically sets out the method of calculating such unfair advantage by dividing it into realized gains, unrealized gains, and avoided losses, and based on the type of unfair trade practice (use of material undisclosed information, price manipulation, and unfair trading).)</p>
	<b>기계 번역②</b>	<p>Peraturan pelaksanaan secara spesifik mengatur metode perhitungan keuntungan tidak wajar ini dengan membaginya menjadi keuntungan terealisasi, keuntungan belum terealisasi, dan jumlah kerugian yang dihindari, berdasarkan jenis perilaku transaksi tidak adil (pemanfaatan informasi penting yang tidak diungkapkan, manipulasi harga pasar, transaksi tidak sah).</p> <p>(The implementing regulations specifically regulate the method of calculating these unfair profits by dividing them into realized profits, unrealized profits, and the amount of avoided losses, based on the type of unfair transaction behavior (exploitation of material undisclosed information,</p>

		market price manipulation, unauthorized transactions).)
	<b>병렬 말뭉치</b>	Dekret pelaksanaan secara rinci menetapkan metode perhitungan laba tidak adil seperti itu dengan mengelompokkannya ke dalam laba yang direalisasikan, laba yang belum direalisasikan, serta kerugian yang berhasil dihindari sesuai dengan jenis praktik perdagangan yang tidak adil (penggunaan informasi penting yang dirahasiakan, manipulasi harga, dan perdagangan tidak adil lainnya). ("The implementing decree specifies in detail the method for calculating such unfair profits by categorizing them into realized profits, unrealized profits, and avoided losses according to the type of unfair trade practice (use of confidential information, price manipulation, and other unfair trade practices).")
	<b>비교</b>	기계 번역①의 'Keputusan penegakan (집행 결정)'이나 기계 번역②의 'Peraturan pelaksanaan (시행 규정)'보다 병렬 말뭉치의 'Dekret pelaksanaan (시행령/시행 포고령)'이 공식적인 법규 명칭에 더 적합함. 기계 번역①의 'perdagangan tidak adil'이나 기계 번역②의 'transaksi tidak sah'는 원문의 '부정 거래'가 갖는 포괄적인 의미를 충분히 담지 못하지만, 병렬 말뭉치는 'perdagangan tidak adil lainnya (기타 불공정 거래)'로 처리하여 원문이 포괄하는 범위를 가장 정확하게 표현함.
<b>원문</b>	<b>구어1</b>	택시 타고 집에 가는데 택시 기사님이 여기 완전 순 그거, 돌팔이들만 있다고 하더라고요.
<b>번역</b>	<b>기계 번역①</b>	Saya naik taksi pulang dan sopir taksi mengatakan bahwa tempat ini penuh dengan orang idiot dan penipu. (I took a taxi home and the taxi driver said this place was full of idiots and scammers.)
	<b>기계 번역②</b>	Saya naik taksi pulang ke rumah, dan sopir taksi bilang di sini semuanya adalah dokter palsu, hanya ada dokter abal-abal. (I took a taxi home, and the taxi driver said that all the doctors here are fake, there are only fake doctors.)
	<b>병렬 말뭉치</b>	Aku naik taksi untuk pulang ke rumah dan sopir taksinya bilang di sini itu benar-benar, hem, hanya ada para

		gadungan. (I took a cab home and the cab driver said here it was really, um, just a bunch of fakes.)
<b>비교</b>		기계 번역은 핵심 표현인 '돌팔이'의 의미를 정확히 포착하지 못하고 '바보/사기꾼'이나 '가짜 의사'로 오역하여 의미를 한정시킴. 반면, 병렬 말뭉치는 '돌팔이'를 가장 포괄적이고 자연스러운 구어인 'para gadungan'으로 정확히 번역했고, '완전 순 그거' 등의 구어 표현을 'benar-benar, hem, hanya ada...'와 같이 살려 택시 기사 말을 전달함.
<b>원문</b>	<b>구어2</b>	언니, 혹시 고등학교 때 보충 수업이나 {야자} 수업 같은 거 했었어?
<b>번역</b>	<b>기계 번역①</b>	Kakak, waktu SMA dulu pernah ikut les tambahan atau les palm nggak? (Sis, when you were in high school, did you ever take extra lessons or palm lessons?)
	<b>기계 번역②</b>	Adik, apakah kamu pernah melakukan hal seperti kelas tambahan atau kelas ayaza di SMA? (Little brother, have you ever done anything like extra classes or ayaza classes in high school?)
	<b>병렬 말뭉치</b>	Kakak, apakah saat SMA pernah mengikuti kelas tambahan atau sesi belajar malam mandiri seperti itu? (Sis, did you ever take extra classes or independent evening study sessions like that in high school?)
<b>비교</b>		기계 번역①은 '야자'를 'les palm (야자수 수업)'으로, 기계 번역②는 'kelas ayaza (ayaza 수업)'로 번역하여 야간 자율 학습의 의미를 전혀 전달하지 못함. 기계 번역②는 청자를 'Adik (여동생/남동생)'으로 잘못 지칭하여 원문의 '언니 (Kakak)'라는 호칭과 일치하지 않음. 병렬 말뭉치는 야자를 'sesi belajar malam mandiri (독립적인 저녁 자율 학습 시간)'으로 번역하여 정확한 뜻을 살렸음.
<b>원문</b>	<b>구어3</b>	너도 알겠지만 나는 항공 점퍼 참 자주 입는데, 이게 시대를 안 타서 제일 좋은 옷인 거 같아.
<b>번역</b>	<b>기계 번역①</b>	Seperti yang Anda ketahui, saya cukup sering mengenakan seragam maskapai penerbangan, dan menurut saya seragam itu yang terbaik karena tidak pernah ketinggalan zaman. (As you know, I wear airline uniforms quite often, and I think they are the best because they never go out of style.)

	기계 번역②	Kau juga tahu, tapi aku sering memakainya di toko penerbangan, ini adalah pakaian terbaik yang tidak pernah berlangsung. (You know too, but I wear it a lot in the flight shop, it's the best outfit that never lasts.)
	병렬 말뭉치	Kamu juga pasti mengetahuinya, tetapi aku sering memakai jaket penerbangan dan sepertinya ini memang pakaian terbaik karena tidak ketinggalan zaman. (You must know it too, but I often wear flight jackets and it seems like this is indeed the best clothing because it doesn't go out of style.)
비교		기계 번역①은 '항공 점퍼'를 'seragam maskapai penerbangan (항공사 제복/유니폼)'으로 오역하여 옷의 종류를 완전히 잘못 전달했고 기계 번역②의 'toko penerbangan (항공 가게)' 착용 표현은 문맥상 의미가 통하지 않음. 기계 번역②의 문장 끝 'pakaian terbaik yang tidak pernah berlangsung (결코 지속되지 않는 최고의 옷)'은 문법적으로나 의미적으로 원문('시대를 안 타서 제일 좋은 옷')과 완전히 반대되는 심각한 오류를 포함함. 병렬 말뭉치에서는 '항공 점퍼'를 'jaket penerbangan (비행/항공 재킷)'으로 정확히 번역하여 옷의 종류를 제대로 전달하고 있음.

<표 44> 기계 번역과 병렬 말뭉치의 번역 비교(태국어)

원문	문어1	왕 주임은 '대만 내에서 '대만 독립'을 시도하면 역사의 심판을 받을 것이며, 국제적으로 대만 독립을 지지하는 것을 용인한다면 반드시 화를 자초하고 쓴맛을 볼 것'이라고도 경고했다.
번역	기계 번역①	ผู้อำนวยการหวังยังเตือนด้วยว่า "ความพยายามใดๆ ในการ 'แยกตัวเป็นเอกราชของไต้หวัน' ภายในไต้หวันนั้น จะถูกตัดสินด้วยประวัติศาสตร์ และการสนับสนุนใดๆ ในระดับนานาชาติต่อการแยกตัวเป็นเอกราชของไต้หวันนั้น จะนำมาซึ่งปัญหาและความขมขื่นอย่างหลีกเลี่ยงไม่ได้" (Director Wang also warned that 'any attempt at 'Taiwan independence' within Taiwan will be judged by history, and any international support for Taiwan independence will inevitably bring trouble and bitterness.)
	기계 번역②	วังจும்เตือนว่า 'การพยายาม 'เอกราชของไต้หวัน' ในไต้หวันจะถูกตัดสินจากประวัติศาสตร์ และหากเขายอมรับการสนับสนุนเอกราชของไต้หวันในระดับนานาชาติ เขาจะ

		<p>ต้องได้รับความเดือดร้อนและขมขื่น'</p> <p>(Wang warned that 'the attempt at 'Taiwan independence' in Taiwan will be judged by history, and if he accepts international support for Taiwan independence, he will suffer and taste bitterness.')</p>
	병렬 말뭉치	<p>นายหวัง รัฐมนตรี ยังเตือนอีกว่า "หากไต้หวันเดินหน้าผลักดัน 'การแยกตัวเป็นเอกราช' ก็จะต้องถูกประวัติศาสตร์ตัดสิน และหากประเทศอื่น ๆ ยอมรับหรือสนับสนุนการแยกตัวของไต้หวัน ก็จะต้องนำภัยมาสู่ตนเองและลิ้มรสความขมขื่นอย่างแน่นอน"</p> <p>(Minister Wang also warned that 'if Taiwan proceeds to push for 'independence,' it will surely be judged by history, and if other countries accept or support Taiwan's separation, they will definitely bring calamity upon themselves and taste bitterness.)</p>
	비교	<p>기계 번역①은 '역사의 심판을 받을 것'이라는 비유적 정치 표현을 문자 그대로 번역하면서 태국어 문법상 부자연스러운 구조가 되었고, 표현 역시 매우 딱딱해졌음. 이는 기계 번역이 비유적·수사적 표현을 해석하지 못하고 직역 수준에서 처리함으로써 원문이 지닌 경고의 뉘앙스를 충분히 반영하지 못했다는 점을 보여 줌.</p> <p>기계 번역②는 '왕 주임'을 음역하여 마치 인명처럼 보이도록 번역한 점이 가장 큰 문제로 꼽힘. 이로 인해 독자가 직책을 사람 이름으로 오해할 가능성이 있으며, 정치 기사 문맥에서는 특히 부적절한 처리 방식임. 또한 '대만 독립을 시도하면'이라는 구절을 직역하면서 태국어 문법에서 어색하고 부자연스러운 문장이 된 점, 그리고 일부 문장이 누락되거나 의미가 축소된 점도 문제임. 특히 국제적 지지에 대한 경고 대상을 잘못 설정함으로써 원문의 핵심 메시지가 왜곡되는 심각한 오역도 발생했음.</p> <p>병렬 말뭉치는 '주임'을 실제 인물의 직책과 맞게 자연스럽게 정확한 것으로 처리했으며, '역사의 심판'이나 '화를 자초하고 쓴맛을 본다'와 같은 비유적 표현도 정치적 문맥에 맞게 자연스러운 태국어 표현으로 변환했음. 문장 구조 또한 태국어 문법과 언론 보도 관행을 충실히 반영하고 있어, 원문의 의미와 의도가 가장 정확하게 표현된 번역으로 평가됨.</p>
원문	문어2	<p>그러면서 대전협은 '행정부 최고 수장을 만나 전공의 의견을 직접 전달한다는 것에 의미를 두는 만남'이라고 설명했다.</p>
번역	기계 번역①	<p>สมาคมการแพทย์เกาหลีอธิบายว่าการประชุมครั้งนี้มีความสำคัญเพราะ "เป็นการเปิดโอกาสให้ประชาชนได้พบปะกับเจ้าหน้าที่รัฐระดับสูงสุดและแสดงความคิดเห็นโดยตรง"</p>

		(The Korean Medical Association (KMA) explained that the meeting was significant because 'it provided an opportunity for the people/public to meet with the highest-ranking government officials and express their opinions directly.')
	기계 번역②	<p>ในขณะเดียวกัน แดจอนฮยอบ อธิบายว่า 'การประชุมที่มีความหมายในการพบกับหัวหน้าฝ่ายบริหารระดับสูงและถ่ายทอดความคิดเห็นของผู้เชี่ยวชาญโดยตรง'</p> <p>(Meanwhile, Daejeonhyeop (or Daejeon Medical Association) explained that the meeting was 'meaningful in that it allowed them to meet with the high-level head of the administration and directly convey the opinions of experts.')</p>
	병렬 말뭉치	<p>พร้อมกันนั้น สมาคมแพทย์ประจำบ้านแห่งเกาหลีได้อธิบายว่า "การพบปะมีนัยสำคัญตรงที่การถ่ายทอดความคิดเห็นของแพทย์ประจำบ้านโดยตรงเมื่อได้พบกับผู้นำสูงสุดของฝ่ายบริหาร"</p> <p>(At the same time, the Korean Intern and Resident Association (KIRA) explained that 'the meeting was significant in that it allowed for the direct relaying of residents' opinions upon meeting with the supreme leader of the administration.')</p>
	비교	<p>기계 번역①은 '대전협(전공의협의회)'을 '한국 의사협회'로 잘못 번역해 주체가 완전히 달라졌음. 또한 인용 부호 처리도 부정확하여 원문의 구조가 달라졌음. 그 결과, '한국 의사협회에서 이번 만남의 의의가 있는 것은 전공의 의견을 전달할 수 있기 때문이라고 설명했다'라는 의미로 잘못 번역되었음.</p> <p>기계 번역 ②는 '대전협'을 음역하여 단체명이 아니라 사람 이름처럼 보이게 번역했음. 또한 문장이 불완전하고 태국어 문법상 부자연스러워 의미 전달이 명확하지 않음. 특히 '전공의 의견'을 '전문가 의견'으로 바꾸어 원문이 지칭하는 집단의 정체성을 정확히 반영하지 못했음.</p> <p>병렬 말뭉치는 '대전협'을 정확한 공식 명칭으로 번역해 주체를 명확히 했음. 또한 '의의를 두는 만남'이라는 원문의 표현을 자연스럽게 정확한 태국어 문체로 전달했음. 전반적으로 문장 구조와 의미 전달이 가장 자연스럽게 원문 의도에 충실한 번역임.</p>
원문	문어3	뿌리 기업이란 주조, 금형, 소성 가공, 용접, 표면 처리, 열처리 등 공정 기술을 활용한 뿌리 산업을 영위하는 기업이다.
번역	기계 번역①	บริษัทรากเป็นบริษัทที่ดำเนินกิจการอุตสาหกรรมรากโดยใช้เทคโนโลยีกระบวนการต่าง ๆ เช่น การหล่อ การขึ้นรูป การแปรรูปพลาสติก การเชื่อม การเคลือบผิว และการอบ

		<p>ชุบด้วยความร้อน</p> <p>(A Root Company is a company that operates the root industry by using various process technologies, such as casting, molding, plastic processing, welding, surface coating, and heat treatment.)</p>
	기계 번역②	<p>บริษัทราก คือบริษัทที่ดำเนินการอุตสาหกรรมรากโดยใช้เทคโนโลยีกระบวนการ เช่น การหล่อแม่พิมพ์ การแปรรูปพลาสติก การเชื่อม การบำบัดผิว และการบำบัดความร้อน</p> <p>(A Root Company is a company that conducts the root industry by using process technologies, such as die casting, plastic processing, welding, surface treatment, and heat treatment.)</p>
	병렬 말뭉치	<p>บริษัทฐานรากหมายถึงบริษัทที่ดำเนินธุรกิจใอุตสาหกรรมรากฐาน โดยใช้เทคโนโลยีการผลิต เช่น การหล่อ การทำแม่พิมพ์ การขึ้นรูป การเชื่อม การเคลือบผิว และการอบชุบ เป็นต้น</p> <p>(A Foundational Company refers to a company that operates in the foundational industry, using manufacturing technologies such as casting, die making, molding, welding, surface coating, and heat treatment, etc.)</p>
	비교	<p>기계 번역①은 '뿌리 산업'을 그대로 'อุตสาหกรรมราก(뿌리 산업)'으로 직역해 용어의 산업적 의미가 충분히 전달되지 않았고 태국 산업 용어 체계와도 맞지 않음. 또한 '소성 가공'을 'การแปรรูปพลาสติก(플라스틱 가공)'으로 번역해 의미가 과도하게 축소되었음. 전반적으로 각 공정 용어가 정확한 기술적 의미를 반영하지 못한 번역문임.</p> <p>기계 번역②는 기계 번역①과 같이 '뿌리 산업'을 'อุตสาหกรรมราก'으로 직역해 용어의 전문적 의미가 부족함. 또한 '금형(금속 주형)'을 'การหล่อแม่พิมพ์(주형을 주조하다)'으로 잘못 옮기는 등 공정명 번역의 정확성이 떨어짐. 전반적으로 용어 선택이 원문 기술적 의미와 일치하지 않아 산업 분야 번역으로서 부정확함.</p> <p>병렬 말뭉치는 '뿌리 산업'을 태국에서 통용되는 'อุตสาหกรรมรากฐาน(기반 산업)'으로 자연스럽게 의미에 맞게 번역했음. 또한 각 공정 기술도 정확한 기술 용어(หล่อ, ทำแม่พิมพ์, ขึ้นรูป, เคลือบผิว, อบชุบ 등)를 사용해 원문의 산업적 의미를 충실히 반영했음. 전체적으로 가장 자연스럽게 전문용어 사용이 정확한 번역임.</p>
원문	구어1	거기는 {전 사랑}이라고 해 가지고 닭을 이렇게 포를 떠요.

번역	기계 번역①	ตรงนั้นเขาบอกว่า ฉันรักคุณ แล้วก็ทอดไก่แบบนี้ (Over there, he said I love you, and then fried the chicken like this.)
	기계 번역②	ที่นั่น (อดีตความรัก)ทำแบบนี้ ไก่ก็เลยลอยแบบนี้ครับ (There (past love/ex-lover) did it like this, so the chicken floated like this.)
	병렬 말뭉치	ร้านนั้นที่ชื่อว่าช็อนซารัง เขาจะใช้วิธีการแยกเนื้อไก่ออกมาแบบนี้ (That restaurant named 'Cheonsarang' uses this method to separate the chicken meat.)
비교	<p>기계 번역①은 '전 사랑'을 '나는 당신을 사랑해요'로 번역하여, 실제로는 가게 이름을 전혀 인식하지 못했음. '닭을 포를 뜨다'라는 요리 동작을 'ทอดไก่(닭을 튀긴다)'로 번역해 조리 방식이 완전히 잘못되었음. 전체적으로 맥락 이해 부족으로 원문 의미가 크게 왜곡되었음.</p> <p>기계 번역②는 '전 사랑'을 'อดีตความรัก(지난 사랑/전 연애)'으로 번역해 의미가 더욱 심각하게 잘못 전달되었음. 또한 문장 구조가 부자연스럽고 'ไก่ก็เลยลอยแบบนี้(닭이 이렇게 떠 있다)'처럼 원문과 무관한 해석이 추가되어 문맥이 크게 흐트러짐. 원문의 의미, 상황, 조리 방식 모두 정확히 반영되지 못한 번역문임.</p> <p>병렬 말뭉치는 '전 사랑'을 음역을 해서 '실제 가게 이름'으로 정확하게 이해하여 자연스럽게 번역했음. 또한 '닭을 포를 뜨다'를 'แยกเนื้อไก่ออกมาแบบนี้(이렇게 살을 발라낸다)'로 상황과 조리 동작을 정확히 반영했음. 전체적으로 문맥, 의미, 자연스러운 태국어 표현까지 모두 올바른 번역문임.</p>	
원문	구어2	언니 학교는 {야자} 활동을 무조건적으로 안 해도 됐었어?
번역	기계 번역①	โรงเรียนของพี่สาว... มักจะไม่มีเงื่อนไขก็ได้ใช่ไหม? (Older sister's school... can usually be without conditions, right?)
	기계 번역②	โรงเรียนของพี่สาว... ไม่ต้องทำกิจกรรมอย่างไม่มีเงื่อนไขได้หรือ? (Older sister's school... doesn't need to do the activity unconditionally, right?)
	병렬 말뭉치	ที่โรงเรียนพี่ กิจกรรมการทบทวนบทเรียนตอนกลางคืนไม่ได้เป็นกิจกรรมบังคับเหรอ? (At older sister's school, isn't the nighttime study review activity not compulsory?)

<b>비교</b>		<p>기계 번역①은 '야자(야간자율학습)'이라는 단어를 전혀 인식하지 못하고 '점'으로만 나와 있음. 또한, 'ไม่มีเงื่อนไขก็ได้(조건 없이 괜찮다)'라는 일반적인 표현 방식으로 번역해 의미가 크게 왜곡되었음. 그리고 문장 구조가 불완전하여 질문의 초점인 '야자 활동의 의무 여부'를 전혀 드러내지 못함. 결과적으로 원문의 핵심 의미를 거의 전달하지 못한 번역문임.</p> <p>기계 번역②는 기계 번역①과 같이 '야자'라는 단어를 전혀 인식하지 못하고 있어 '점'으로만 나와 있음. 'ทำกิจกรรมอย่างไร้เงื่อนไข(조건 없이 활동을 해야 한다/안 해도 된다)'와 같이 모호하고 어색한 표현을 사용했음. 질문의 '의무성 여부'가 태국어 문장에서 자연스럽게 드러나지 않으며, 여전히 '야자'의 실제 의미가 번역되지 않았음. 전반적으로 기계 번역①보다 나아 보이지만 의미 전달은 여전히 부정확함. 병렬 말뭉치는 '야자'를 정확히 '야간 자율 학습'으로 번역해 한국 교육 문화의 맥락을 올바르게 전달했음. 또한 '무조건적으로'라는 부분을 자연스럽게 표현하여 원문의 질문 의도를 완전히 반영했음. 전체적으로 정확성, 자연스러움, 맥락 이해 모두 가장 뛰어난 번역문임.</p>
<b>원문</b>	구어3	지금 좀 좋은 데서 나오면 하나 살까 하고 저도 우물쭈물 기다리고 있어요.
<b>번역</b>	<b>기계 번역①</b>	ฉันเองก็ลังเลอยู่เหมือนกัน และกำลังรอที่จะซื้อถ้ามันออกมาในราคาที่ดี (I'm also hesitating, and I'm waiting to buy one if it comes out at a good price.)
	<b>기계 번역②</b>	ตอนนี้ถ้าออกมาจากที่ดีๆ ก็คิดว่าอาจจะซื้อสักอันดีมัย ผมเองก็กำลังรออยู่ครับ (If a good one comes out, I'm thinking I might buy one. I'm waiting as well.)
	<b>병렬 말뭉치</b>	ตอนนี้ถ้าออกมาจากที่ดี ๆ หน่อยก็คิดว่าอาจจะซื้อสักเครื่อง ดีฉันเองก็กำลังสองจิตสองใจและรออยู่ค่ะ (If it comes out from a good place, I'm thinking of getting one. I'm still in two minds about it, so I'm waiting for now.)
<b>비교</b>		<p>기계 번역①은 '우물쭈물'이라는 표현을 단순히 'ลังเล(주저하다)'로 처리해 원문의 망설임과 기대감이 공존하는 상태라는 정서적 분위기를 충분히 살리지 못했음. 또한 '좋은 데서 나온다'를 '좋은 가격'으로 잘못 해석해 제품 출시 맥락이 왜곡되었음. 이로 인해 화자가 '망설이면서도 기대하는 상태'가 제대로 전달되지 않는 번역문이 됨.</p> <p>기계 번역②는 마이야목(๗: 강조, 복수형, 정도 완화의 기능을 갖는</p>

	<p>부호로 앞뒤를 띄어 씀)의 띄어쓰기에 오류가 있음. 또한, '우물쭈물'이라는 단어가 누락되어 결국 '기다리고 있다'라는 내용만 일부 번역되면서 원문이 전달하는 미묘한 감정과 망설임의 뉘앙스가 반영되지 못했음.</p> <p>병렬 말뭉치는 '우물쭈물'의 의미를 '사고 싶지만 확신이 없어서 계속 망설이며 기다리는 상태'라는 원문의 감정을 정확하게 담아 자연스럽게 번역함.</p>
--	--

<표 45> 기계 번역과 병렬 말뭉치의 번역 비교(인도 힌디어)

원문	문어1	이러한 강구조는 강도가 크고 내구성과 내진성이 뛰어나며 공사 기간을 단축시킬 수 있는 강점을 가지고 있다.
번역	기계 번역①	इन इस्पात संरचनाओं में उच्च शक्ति, स्थायत्व, भूकंप प्रतिरोध और निर्माण अवधि को कम करने की क्षमता होती है। (These steel structures have high strength, durability, earthquake resistance and the ability to reduce construction time.)
	기계 번역②	इस इस्पात संरचना में मजबूत, टिकाऊ और भूकंप प्रतिरोधी होने तथा निर्माण अवधि को छोटा करने की खूबियाँ हैं। (This steel structure has the advantages of being strong, durable and earthquake resistant and shortening the construction period.)
	병렬 말뭉치	इस तरह की स्टील संरचनाएँ मजबूत, टिकाऊ और उत्कृष्ट भूकंप-प्रतिरोधक क्षमता वाली होती हैं, और यह निर्माण अवधि को कम करने की क्षमता रखती हैं। (Such steel structures are strong, durable and have excellent earthquake resistance, and have the potential to reduce construction time.)
비교		기계 번역에서는 원문의 '뛰어나다'에 해당하는 단어가 힌디어 번역에서 누락되어 '강도·내구성·내진성'이 단순한 성질로만 표현됨. 병렬 말뭉치는 원문의 '뛰어나다'에 해당하는 'उत्कृष्ट(excellent)'을 'भूकंप-प्रतिरोधक क्षमता(earthquake resistance)' 앞에 사용해 내구성과 내진성이 우수하다는 원문의 평가적 의미를 명확하게 반영함.
원문	문어2	그러나 연체율 20%가 방증하듯 취약 계층 지원 실효성은 떨어진다 는 지적이 있다.
번역	기계 번역①	हालाँकि, जैसा कि 20% अपराध दर से स्पष्ट है, इस बात की आलोचना भी है कि कम जोर समूहों के लिए सहायता की प्रभावशीलता कम है। (However, as evidenced by the 20% crime rate, there is also



		<p>यापक बनाया जाना चाहिए। (Secondary financial institutions are reducing lending to middle- and low-credit-scoring people due to the financial stability crisis, so there has been criticism that the financial supply mechanism of the general public should be strengthened and expanded to help weaker sections.)</p>
비교		<p>기계 번역에서는 원문의 ‘중저신용자’를 ‘नमिन और मध्यम ऋण उधारकर्ताओं(low and medium loan borrowers)’로 번역해 신용 등급을 의미하는 금융 용어가 단순 ‘대출 규모가 낮은 사람들’로 왜곡되거나, ‘मध्यम ऋणदाताओं(middle-tier lenders)’로 번역되어 ‘대출을 받는 사람’이 아니라 ‘대출을 해 주는 사람’으로 의미가 완전히 바뀌는 오역이 발생함.</p> <p>병렬 말뭉치는 ‘중저신용자’를 ‘मध्यम और नमिन क्रेडिट स्कोर वाले लोग (people with medium and low credit scores)’로 정확히 옮겨 신용 등급이 낮아 대출 접근성이 떨어지는 계층을 의미한다는 원문의 금융적 맥락을 명확히 반영함.</p>
원문	구어1	<p>그래서 {영끌}을 해서 그 집에 들어가야 되나 하는 생각을 하게 되잖아.</p>
번역	기계 번역①	<p>तो, आप सोचने लगते हैं किक्या आपको नौकरी करनी चाहिए और उस घर में जाना चाहिए। (So, you start wondering if you should take the job and move into that house.)</p>
	기계 번역②	<p>तो, मुझे आश्चर्य है किक्या मुझे येऑंगकेउल और उस घर में प्रवेश करना चाहिए। (So, I wonder if I should go to Yeongkeul and enter that house.)</p>
	병렬 말뭉치	<p>इसलिए हम ऐसा सोचते हैं न कहियें अपनी सारी जमा-पूँजी लगाकर उस घर में चले जाना चाहिए। (That is why we think that we should invest all our savings and move into that house.)</p>
비교		<p>기계 번역에서는 원문의 ‘영끌(영혼까지 끌어모아 대출·자금 마련을 한다는 뜻)’을 ‘नौकरी करनी चाहिए(should take the job)’로 번역해 전혀 다른 의미로 바뀌는 심각한 오역이 발생하거나 고유명사처럼 ‘ये ऑंगकेउल(Yeongkeul)’로 음역하여 원문의 금융적 의미가 완전히 전달되지 않음.</p> <p>병렬 말뭉치는 ‘영끌’을 ‘अपनी सारी जमा-पूँजी लगाकर(investing all our savings)’으로 번역해 가용 자금을 모두 끌어모아 집을 사려 한다는</p>

		원문의 핵심 의미를 정확하게 반영함.
원문	구어2	언니, 혹시 고등학교 때 보충 수업이나 {야자} 수업 같은 거 했었어?
번역	기계 번역①	बहन, क्या आपने हाई स्कूल में कभी अतिरिक्त कक्षाएं या ट्यूशन कक्षाएं लीं? (Sister, did you ever take extra classes or tutoring classes in high school?)
	기계 번역②	बहन, क्या आपने हाई स्कूल में कभी कोई अतिरिक्त कक्षाएं या हस्तरेखा कक्षाएं लीं? (Sister, did you ever take any extra classes or palmistry classes in high school?)
	병렬 말뭉치	दीदी, क्या तुमने हाई स्कूल में कोई अतिरिक्त कक्षाएँ या रात की पढ़ाई की थी? (Didi, did you take any extra classes or night studies in high school?)
비교		기계 번역은 ‘야자’를 ‘ <b>ट्यूशन कक्षाएं(tutoring classes)</b> ’로 번역해 방과 후 자율 학습의 의미를 사교육(튜터링)으로 변경하거나 ‘ <b>हस्तरेखा कक्षाएं(palmistry classes)</b> ’로 오역해 원문 의미에서 완전히 벗어남. 병렬 말뭉치는 ‘야자’를 ‘ <b>रात की पढ़ाई(night study)</b> ’로 번역하여 한국 학교 맥락에서의 ‘야간 자율 학습’이라는 의미를 정확히 반영함.
원문	구어3	근데 너 게임을 하면서 이제 그, 게임에 현질을 좀 한 적 있니?
번역	기계 번역①	लेकिन क्या आप कभी गेम खेलते समय उसके आदी हो गए हैं? (But have you ever gotten addicted to a game while playing it?)
	기계 번역②	लेकिन गेम खेलते समय क्या आपने कभी गेम खेला है? (But have you ever played a game while playing a game?)
	병렬 말뭉치	क्या तुमने कभी गेम खेलकर नकद भुगतान किया है? (Have you ever gotten paid cash playing games?)
비교		기계 번역은 ‘현질’을 ‘ <b>उसके आदी हो गए हैं(game addiction)</b> ’으로 해석해 ‘돈을 쓰는 행위’가 아닌 ‘중독’의 의미로 왜곡하거나 ‘게임을 하면서 게임을 했다’는 식의 의미 중복 문장을 생성함. 병렬 말뭉치는 ‘현질’을 ‘ <b>नकद भुगतान(cash payment)</b> ’으로 번역해 게임 내에서 실제 돈을 쓰는 행위를 정확하게 반영함.

<표 46> 기계 번역과 병렬 말뭉치의 번역 비교(캄보디아 크메르어)

원문	문어1	{권병훈} {HD 한국 조선 해양 전동화 센터}장은 '무탄소 대형 선박과 차세대 전기 추진 함정에는 {MVDC} 기술이 선택이 아닌 필수'라며 '국제 표준 정립을 통해 대형선 전기 추진 시장을 선도할 것'이라고 말했다.
----	-----	--

번역	기계 번역①	<p>លោក Kwon Byung-hoon ប្រធាន HD Hyundai Heavy Industries Electrification Center បានអះអាងថា "បច្ចេកវិទ្យា MVDC គឺជាការជ្រើសរើសមិនមែនជាជម្រើសទេ ប៉ុន្តែជាការចាំបាច់សម្រាប់សំពៅធំដែលមិនបញ្ចេញកាបូន និងសំពៅដែលប្រើប្រាស់កម្លាំងអគ្គិសនីជំនាន់ក្រោយ" ហើយបានបន្ថែមថា "យើងនឹងនាំទីផ្សារសំពៅដែលប្រើប្រាស់កម្លាំងអគ្គិសនីធំៗតាមរយៈការបង្កើតស្តង់ដារអន្តរជាតិ"។</p> <p>(Mr. Kwon Byung-hoon, head of the HD Hyundai Heavy Industries Electrification Center, stated that "MVDC technology is not an option but a necessity for large carbon-free ships and next-generation electric-powered vessels," adding, "We will lead the market for large electric-powered ships by establishing international standards.")</p>
	기계 번역②	<p>លោក Kwon Byung-hoon ប្រធាន HD Korea Shipbuilding &amp; Offshore Electrification Center បាននិយាយថា "បច្ចេកវិទ្យា MVDC មិនមែនជាជម្រើសទេ ប៉ុន្តែជាការចាំបាច់សម្រាប់កំប៉ាល់ធំៗដែលគ្មានកាបូន និងកំប៉ាល់ជំរុញថាមពលអគ្គិសនីជំនាន់ក្រោយ" ហើយ "យើងនឹងនាំទីផ្សារជំរុញថាមពលកំប៉ាល់ធំៗដោយបង្កើតស្តង់ដារអន្តរជាតិ"។</p> <p>(Kwon Byung-hoon, President of HD Korea Shipbuilding &amp; Offshore Electrification Center, said, "MVDC technology is not an option but a necessity for large carbon-free balls and next-generation electric propulsion balls," and "we will lead the large balls propulsion market by setting international standards.")</p>
	병렬 말뭉치	<p>ក្នុងប្រព័ន្ធប្រធានមជ្ឈមណ្ឌលថាមពលកំប៉ាល់ និងអគ្គិសនីនៅល្វែងសមុទ្រក្នុង HD និយាយថា បច្ចេកវិទ្យា 'MVDC មិនមែនជាជម្រើសទេ តែជាការចាំបាច់សម្រាប់កំប៉ាល់ធំៗដែលគ្មានកាបូន និងកំប៉ាល់ថាមពលអគ្គិសនីជំនាន់ក្រោយ' ហើយថា 'យើងនឹងនាំទីផ្សារជំរុញថាមពលកំប៉ាល់ធំៗ ដោយបង្កើតស្តង់ដារអន្តរជាតិ'។</p> <p>(Kwon Byung-hoon, Head of the Korea Offshore &amp; Shipbuilding Power System Center of HD, said that 'MVDC technology is not an option, but a necessity for large zero-carbon ships and next-generation electric power ships' and that 'we will lead the large ship power propulsion market by establishing international standards'.)</p>
비교	<p>기계 번역①은 회사명(HD 한국 조선 해양) 표기에 오류가 있고, 기계 번역②에는 오타·오역이 있어 의미가 왜곡됨.</p> <p>병렬 말뭉치는 회사명을 음역하여 크메르어 표기로 전환했고, '무탄소 대형 선박', '전기 추진 함정' 등 전문용어를 가장 자연스럽게 적절한 크메르어 용어로 사용하여 원문의 기술적 내용을 명확히 전달하고 있음.</p>	

원문	문어2	<p>2022년 4월 주민 신청으로 자문 방식의 신속 통합 기획에 착수해 2023년 7월 기획을 완료하고, 주민 공람, 구의회 의견 청취 등 절차를 거쳐 서울특별시 도시 계획 위원회에 정비 계획이 상정됐다.</p>
번역	<p>기계 번역①</p>	<p>នៅខែមេសា ឆ្នាំ 2022 ដំណើរការរៀបចំផែនការរួមបញ្ចូលគ្នាយ៉ាងរហ័សត្រូវបានផ្តើមតាមរយៈដំណើរការពិគ្រោះយោបល់ដោយផ្អែកលើពាក្យសុំស្នាក់នៅ ហើយផែនការនេះត្រូវបានបញ្ចប់នៅខែកក្កដា ឆ្នាំ 2023។ បន្ទាប់ពីឆ្លងកាត់នីតិវិធីដូចជាការជូនដំណឹងជាសាធារណៈ និងការស្តាប់យោបល់ពីក្រុមប្រឹក្សាស្រុក ផែនការអភិវឌ្ឍន៍ឡើងវិញត្រូវបានដាក់ជូនគណៈកម្មាធិការរៀបចំផែនការទីក្រុងសេដ្ឋីល។</p> <p>(In April 2022, a rapid integrated planning process was initiated through a consultation process based on residency applications, and this plan was completed in July 2023. After undergoing procedures such as public notification and hearings from the district council, the redevelopment plan was submitted to the Seoul Metropolitan Government's Planning Committee.)</p>
	<p>기계 번역②</p>	<p>២០២២ ឆ្នាំខែមេសា តាមរយៈការដាក់ពាក្យរបស់ប្រជាពលរដ្ឋ បានចាប់ផ្តើមធ្វើផែនការរួមបញ្ចូលគ្នាយ៉ាងរហ័សតាមរបៀបប្រឹក្សាពិគ្រោះ ហើយបានបញ្ចប់ផែនការនៅខែកក្កដា ២០២៣។ បន្ទាប់ពីឆ្លងកាត់ដំណើរការដូចជាការបង្ហាញជាសាធារណៈដល់ប្រជាពលរដ្ឋ ការស្តាប់មតិពីសភាស្រុក ផែនការកែលម្អត្រូវបានដាក់ជូនគណៈកម្មការគម្រោងទីក្រុងរបស់ទីក្រុងសេដ្ឋីលជាពិសេស។</p> <p>(In April 2022, through citizen applications, a rapid integrated planning process began in a consultative manner and was completed in July 2023. After undergoing procedures such as public disclosure to citizens and listening to opinions from the district council, the revised plan was submitted to the Seoul Metropolitan Government's special urban project committee.)</p>
	<p>병렬 말뭉치</p>	<p>តាមរយៈការស្នើសុំពីពលរដ្ឋ ក្នុងខែ៤ ឆ្នាំ២០២២ គម្រោងបានចាប់ផ្តើមអនុវត្តផែនការរួមបញ្ចូលគ្នាយ៉ាងរហ័សក្នុងទម្រង់ប្រឹក្សាយោបល់ ហើយបានបញ្ចប់ការរៀបចំផែនការក្នុងខែ៧ ឆ្នាំ២០២៣ ហើយបន្ទាប់ពីឆ្លងកាត់នីតិវិធីដូចជាការផ្សព្វផ្សាយជាសាធារណៈដល់ពលរដ្ឋ និងការស្តាប់យោបល់របស់ក្រុមប្រឹក្សាសង្កាត់ជាដើម ផែនការកែទម្រង់ គឺត្រូវបានដាក់ជូនគណៈកម្មាធិការផែនការទីក្រុងនៃទីក្រុងសេដ្ឋីល។</p> <p>(Following requests from citizens in April 2022, the project quickly began implementing an integrated plan in a consultative format and completed the planning in July 2023. After going through procedures such as public dissemination to citizens and listening to the opinions of the Sangkat councils, the reform plan was submitted to the Seoul</p>

		Metropolitan Government's Urban Planning Committee.)
<b>비교</b>		원문의 핵심 용어인 '신속 통합 기획'에 대해 기계 번역①과 기계 번역②는 정확하지 않게 번역함. 기계 번역②는 크메르어의 자연스러운 어순을 따르지 않고 한국어식 어순을 유지하여 부자연스러움. 병렬 말뭉치는 '신속 통합 기획'을 정확한 어휘로 번역하여 '자문 방식'의 의미를 명확히 포함하였고 낱자 표기 역시 크메르어의 자연스러운 어순에 맞춰 매끄럽게 처리했음.
<b>원문</b>	<b>문어3</b>	김정훈 줄라 코리아 대표는 전 국가대표 탁구 선수 출신으로, 현역 시절에 바른 세상 병원에서 꾸준히 치료를 받았던 인연이 해당 대회까지 이어진 것으로 알려졌다.
<b>번역</b>	<b>기계 번역①</b>	វាត្រូវបានគេរាយការណ៍ថា CEO Kim Jeong-hoon ដែលជាអតីតកីឡាករជម្រើសជាតិកំពុងទទួលការព្យាបាលជាប់លាប់នៅមន្ទីរពេទ្យ Barun Barun។ (It is reported that CEO Kim Jeong-hoon, a former national team player, is receiving ongoing treatment at Barun Barun Hospital.)
	<b>기계 번역②</b>	យោងតាមផ្លូវព័ត៌មានរបស់គេហទំព័រ លោក Kim Jung-hoon ប្រធាន Zula Korea ដែលធ្លាប់ជាអតីតអត្តពលិកតិននីស្តរូបជាតិ ក្នុងអំឡុងពេលដែលលោកនៅក្នុងវិស័យកីឡាមានទំនាក់ទំនងជាប់ពាក់ព័ន្ធជាមួយមន្ទីរពេទ្យ Barun Sesang ដោយបានទទួលការព្យាបាលជាប្រចាំ ហើយទំនាក់ទំនងនេះបានបន្តរហូតដល់ការប្រកួតប្រជែងនេះ។ (According to news from the website, Mr. Kim Jung-hoon, president of Zula Korea, who was a former national tennis athlete, was associated with Barun Sesang Hospital during his time in the sports field, receiving regular treatment, and this relationship continued until this competition.)
	<b>병렬 말뭉치</b>	នាយកប្រតិបត្តិរបស់ ជុលកូរ៉េ ឌូឡា គីមចុងហ្គ័ន ដែលជាអតីតកីឡាករវាយតុល្លើលើតុជម្រើសជាតិ ត្រូវបានគេដឹងថាទំនាក់ទំនងដែលគាត់ធ្លាប់ទទួលបានការព្យាបាលជាប់លាប់នៅមន្ទីរពេទ្យ ជារីន សេសាង ក្នុងអំឡុងពេលជាតីឡាករអាជីព បានបន្តរហូតដល់ការប្រកួតនេះ។ (Choo La Korea CEO Kim Jong-hoon, a former national table tennis player, is known to have maintained the relationship he had with Pharun Sesang Hospital for continuous treatment during his professional career until this competition.)
<b>비교</b>		기계 번역①은 '줄라 코리아 대표', '탁구 선수 출신', '해당 대회까지 이어진 인연' 등 핵심 정보를 대부분 누락했으며, '바른 세상 병원'을 중복해서 표기했음. 기계 번역②는 '탁구 선수'를 국가대표 테니스 선수로 오역함. 병렬 말뭉치의 경우 '탁구 선수'를 정확히 명시했고 '줄라 코리아'와

		'바른 세상 병원' 등 고유명사를 크메르어 음차 표기로 정확히 처리했음.
원문	구어1	어차피 나는 대학이나 그런 쪽은 진학을 아예 생각을 안 했으니깐 군대를 카투사로 갈 것도 아니고 굳이 필요하단 생각이 안 들어서 아직까지는 구체적인 계획은 없어.
번역	기계 번역①	<p>បោះយ៉ាងណាក៏ដោយ ខ្ញុំមិនទាន់បានគិតចង់ទៅរៀននៅមហាវិទ្យាល័យ ឬអ្វីទាំងអស់នោះទេ ដូច្នេះហើយខ្ញុំនឹងមិនចូលបម្រើយោធាក្នុងនាមជា KATUSA នោះទេ ហើយខ្ញុំគិតថាវាមិនចាំបាច់ទេ ដូច្នេះហើយខ្ញុំមិនទាន់មានផែនការច្បាស់លាស់នៅឡើយទេ។</p> <p>(However, I haven't thought about going to college or anything, so I won't be joining the military as a KATUSA, and I don't think it's necessary, so I don't have a definite plan yet.)</p>
	기계 번역②	<p>បោះជាយ៉ាងណាក៏ដោយ ខ្ញុំមិនគិតថា សាកលវិទ្យាល័យ និងផ្នែកទៀតនឹងទៅសិក្សាទេ ដូច្នេះខ្ញុំមិនគិតថា កងទ័ពនឹងទៅ Katusa ទេ ហើយខ្ញុំមិនគិត</p> <p>(However, I don't think universities and other departments will go to study, so I don't think the army will go to Katusa, and I don't think)</p>
	병렬 말뭉치	<p>គង់តែខ្ញុំមិនដែលគិតថា ចង់ទៅសាកលវិទ្យាល័យ ឬបន្តឡើងទៅក្នុងស្ថាប័ន ដូច្នេះខ្ញុំមិនទាន់មានគម្រោងជាក់លាក់ណាមួយនៅឡើយទេ ព្រោះខ្ញុំនឹងមិនចូលបម្រើយោធាក្នុងនាមជា កាតូសា នោះឡើយ។</p> <p>(I never thought about going to university or continuing my studies, so I don't have any specific plans yet, because I will not be enlisting in the military as a KATUSA.)</p>
비교		<p>기계 번역①은 원문의 '어차피'를 'ទោះយ៉ាងណាក៏ដោយ' (However, 그럼에도 불구하고)라는 부적절한 연결어로 번역하였음.</p> <p>기계 번역②는 문장이 미완성된 채로 끝나버리는 심각한 문법적 오류를 포함하고 있음.</p> <p>병렬 말뭉치는 '어차피'를 'គង់តែ' 로 정확히 처리하여 원문의 뉘앙스를 살렸음.</p>
원문	구어2	좀 특이하고 소쿠리에다가 하나를 전을 부쳐 가지고 내놨더라고요.
번역	기계 번역①	<p>វាខុសធម្មតាបន្តិចហើយគាត់ដាក់មួយក្នុង colander ហើយបម្រើវា។ (It was a little unusual and he put one in a colander and served it.)</p>
	기계 번역②	<p>វា ជា រឿង មិន ធម្មតា បន្តិច ហើយ ពួក គេ បាន ធ្វើ ផ្កា កុលាប ហើយ យក វា ចេញ។ (It was a little unusual, and they made a rose and took it out.)</p>
	병렬 말뭉치	<p>វាពិសេសខុសប្លែកគេបន្តិច ហើយបានចៀសចំនុះមួយ រួចដាក់នៅក្នុងកញ្ជ្រែង។</p>

		(It is a little bit special and unique, and has been stir-fried once, then put in a basket.)
<b>비교</b>		기계 번역은 핵심 단어인 '전(煎)'을 이해하지 못하고 'colander(채반)에 넣었다'거나 '장미'로 오역하여 의미 전달에 실패함. 병렬 말뭉치는 '전을 부쳐'를 정확히 번역하고 '소쿠리에 담아 내놓았다'는 상황을 자연스럽게 연결하여 원문의 상황을 가장 정확하게 묘사함.
<b>원문</b>	<b>구어3</b>	바닷물 색깔이 그, 옥색인데 거기서 자고 일어나니까 '와, 내가 이런 데 와서 자다니. 대한민국에서도 못 자 봤는데 일본에 와서 이런 데서 자 보는구나.'라는 생각이 들었어.
<b>번역</b>	<b>기계 번역①</b>	ទឹកសមុទ្រមានពណ៌ត្រងទទឹម ហើយពេលខ្ញុំភ្ញាក់ពីដំណេកនៅទីនោះ ខ្ញុំគិតថា 'Wow ខ្ញុំមិនជឿថាខ្ញុំកំពុងដេកនៅកន្លែងបែបនេះទេ។ ខ្ញុំមិនធ្លាប់គេងនៅក្នុងប្រទេសកូរ៉េ ប៉ុន្តែខ្ញុំដេកនៅកន្លែងបែបនេះក្នុងប្រទេសជប៉ុន។' (The sea water was the color of ruby, and when I woke up there, I thought, 'Wow, I can't believe I'm sleeping in a place like this. I've never slept in Korea before, but I've slept in a place like this in Japan.') ពណ៌ ទឹក សមុទ្រ គឺ ពណ៌ ត្រង ពេល ហើយ ពេល ខ្ញុំ ភ្ញាក់ ពី ដំណេក នៅ ទី នោះ ខ្ញុំ គិត ថា (BulsenWow ខ្ញុំ មិន ជឿ ថា ខ្ញុំ មក ដេក នៅ កន្លែង បែប នេះ ទេ។ ខ្ញុំ មិន អាច ដេក នៅ ក្នុង ប្រទេស កូរ៉េ ប៉ុន្តែ ខ្ញុំ មក ប្រទេស ជប៉ុន ហើយ ឃើញ ខ្លួន ឯង ដេក នៅ កន្លែង បែប នេះ។ ខ្ញុំ បាន គិត។
	<b>기계 번역②</b>	(The color of the sea is diamond, and when I woke up from the bed there I thought (BulsenWow, I can't believe I came to sleep in a place like this. I can't sleep in Korea, but I came to Japan and found myself sleeping in a place like this.I thought.)
	<b>병렬 말뭉치</b>	ពណ៌ទឹកសមុទ្រ គឺពណ៌បៃតងខ្ចី ដោយសារខ្ញុំគេង ហើយភ្ញាក់ពីដំណេកនៅទីនោះ ខ្ញុំបានគិតថា 'វ៉ូ ខ្ញុំបានមកដេកនៅកន្លែងបែបនេះ។ ខ្ញុំមិនធ្លាប់គេងបែបនេះនៅប្រទេសកូរ៉េពីមុនមកទេ ប៉ុន្តែខ្ញុំមកជប៉ុន ហើយខ្ញុំបានដេកនៅកន្លែងបែបនេះ។' (The color of the ocean is a shade of green because I was sleeping and when I woke up there, I thought, 'Wow, I slept in a place like this. I have never slept like this before in Korea, but I came to Japan and I slept in a place like this'.)
<b>비교</b>		기계 번역은 핵심 표현인 '옥색'을 '루비'나 '다이아몬드' 색으로 오역하여 바닷물 색깔의 정확성을 잃음. 병렬 말뭉치는 '옥색'을 '연한 초록색'으로 정확히 번역하여 원문의 시각적 정보를 가장 정확하게 전달하고, 문장 전체의 감탄 표현도 크메르어에 맞게 자연스럽게 처리함.

<표 47> 기계 번역과 병렬 말뭉치의 번역 비교(필리핀 타갈로그어)

원문	문어1	손 회장은 이날 {경총}을 방문한 안 장관을 만나 기업 경쟁력 제고를 위한 지원을 요청했다.
번역	기계 번역①	Nakipagpulong si Chairman Son kay Minister Ahn, na bumibisita sa Federation of Korean Industries sa araw na ito, at humiling ng suporta upang mapahusay ang pagiging mapagkumpitensya ng korporasyon. (Chairman Son met with Minister Ahn, who is visiting the Federation of Korean Industries today, and requested support to enhance corporate competitiveness.)
	기계 번역②	Nakipagpulong si Son kay Minister Ahn, na bumisita sa Korea Federation of Korean Industries noong araw na iyon, at humiling ng suporta upang mapabuti ang pagiging mapagkumpitensya ng korporasyon. (Son met with Minister Ahn, who was visiting the Korea Federation of Korean Industries that day, and requested support to improve corporate competitiveness.)
	병렬 말뭉치	Nakipagpulong si Chairman Son kay Minister Ahn, na bumibisita sa Korea Employers Federation sa araw na ito, at humiling ng suporta para sa pagpapabuti ng pagiging mapagkumpitensya ng korporasyon. (Chairman Son met with Minister Ahn, who is visiting the Korea Employers Federation today, and requested support for improving corporate competitiveness.)
비교		기계 번역에서는 원문의 ‘경총’을 각각 ‘Federation of Korean Industries’ 또는 ‘Korea Federation of Korean Industries’로 번역해 실제와 다른 기관명으로 잘못 옮겨졌으며, 특히 기계 번역②는 기관 명칭을 중복 구성해 존재하지 않는 이름으로 번역되는 오류가 발생함. 병렬 말뭉치는 ‘경총’을 ‘Korea Employers Federation’으로 정확히 번역해 고유명사에 해당하는 기관명을 원문 의미에 맞게 올바르게 반영함.
	원문	문어2
번역	기계 번역①	Nagtala ang Our Card ng average na rate ng interes na 14.87%p noong Pebrero, noong nakaraang buwan, ngunit

		tumalon ito ng 0.33%p sa isang buwan. (Our Card recorded an average interest rate of 14.87%p in February, last month, but it jumped 0.33%p in a month.)
	<b>기계 번역②</b>	Ang aming Card ay nagtala ng average na rate ng interes na 14.87%p noong Pebrero, noong nakaraang buwan, ngunit tumaas ng 0.33%p sa isang buwan. (Our Card recorded an average interest rate of 14.87%p in February, last month, but increased by 0.33%p in a month.)
	<b>병렬 말뭉치</b>	Ang Woori Card ay nagtala ng karaniwang antas ng interes na 14.87%p noong Pebrero na nakaraang buwan, pero tumaas ito ng 0.33%p sa isang buwan. (Woori Card recorded an average interest rate of 14.87%p in February last month, but it increased by 0.33%p in a month.)
	<b>비교</b>	기계 번역에서는 고유명사인 ‘우리 카드’를 각각 ‘Our Card’, ‘Ang aming Card’로 번역해 한국의 실제 카드사 명칭이 일반 명사로 바뀌는 오류가 발생함. 반면 병렬 말뭉치는 이를 ‘Woori Card’로 옮겨 회사명 고유성을 정확히 반영함.
<b>원문</b>	<b>문어3</b>	이 방문이 소원했던 한·중 관계의 분위기가 달라지는 계기가 될지 주목된다.
<b>번역</b>	<b>기계 번역①</b>	Nakatuon ang atensyon sa kung ang pagbisitang ito ay magiging dahilan ng pagbabago sa kapaligiran ng relasyon ng Korea-China. (Attention is focused on whether this visit will cause a change in the atmosphere of Korea-China relations.)
	<b>기계 번역②</b>	Binibigyan ng pansin kung ang pagbisitang ito ay magiging isang pagkakataon upang baguhin ang kapaligiran ng hiwalay na relasyon ng Korea-China. (Attention is being paid to whether this visit will be an opportunity to change the atmosphere of the isolated Korea-China relationship.)
	<b>병렬 말뭉치</b>	Nakatuon ang atensyon sa kung magiging dahilan ng pagbabago sa kapaligiran ng matagal nang malamig na relasyon ng Korea at Tsina ang pagbibisitang ito. (Attention is focused on whether this visit will cause a change in the atmosphere of the long-cold relations between

		Korea and China.)
<b>비교</b>		<p>기계 번역①에서는 원문의 ‘소원했던 한·중 관계’를 단순히 ‘relasyon ng Korea-China’로만 표현해 관계가 오랜 기간 소원했다는 뉘앙스가 누락됨.</p> <p>기계 번역②에서는 이를 ‘hiwalay na relasyon(이격된·분리된 관계)’로 번역해 ‘소원하다’의 외교적 냉각 상태를 ‘분리됨’으로 잘못 해석하는 의미 왜곡이 발생함.</p> <p>병렬 말뭉치는 원문의 ‘소원했던’을 ‘matagal nang malamig na relasyon(long-cold relations)’으로 번역해 두 나라 관계가 장기간 냉각되어 있었다는 의미를 가장 정확하게 반영함.</p>
<b>원문</b>	<b>구어1</b>	나하고 같이 다락방에서 공부하던 애는 여상 가기 싫다고 인문계를 가야 되겠다면서 인문계를 갔는데 떨어졌어.
<b>번역</b>	<b>기계 번역①</b>	<p>Ang bata na nag-aral sa akin sa attic ay hindi nais na pumunta sa isang kolehiyo ng kababaihan at pumunta sa departamento ng humanities, ngunit siya ay nabigo.</p> <p>(The boy who studied with me in the attic didn't want to go to a women's college and went to the humanities department, but he was disappointed.)</p>
	<b>기계 번역②</b>	<p>Ang batang nag-aaral sa attic kasama ko ay nagsabing ayaw niyang pumunta at kailangang pumunta sa larangan ng humanities, ngunit nabigo siya.</p> <p>(The boy studying in the attic with me said he didn't want to go and had to go into the humanities field, but he failed.)</p>
	<b>병렬 말뭉치</b>	<p>Yung kasama kong nag-aaral sa attic, ayaw daw niyang pumasok dun sa girls' commercial high school at kailangan daw niyang pumasok sa general academic high school, kaya doon siya nag-apply pero hindi siya nakapasa.</p> <p>(My roommate who was studying in the attic said she didn't want to go to the girls' commercial high school and needed to go to the general academic high school, so she applied there but didn't pass.)</p>
<b>비교</b>		<p>기계 번역에서는 원문의 ‘여상(여자상업고등학교)’이 모두 ‘women's college’ 또는 단순 ‘pumunta’와 같은 표현으로 번역되어 ‘상업계 여자 고등학교’라는 교육 체계 내 고유 의미가 누락됨. 또한 원문의 ‘인문계(일반계 고등학교)’가 ‘humanities department/field’로 번역되어 고등학교 계열을 의미하는 맥락이 대학 전공처럼 왜곡됨.</p>

		병렬 말뭉치는 ‘여상’을 ‘girls’ commercial high school’로, ‘인문계’를 ‘general academic high school’로 옮겨 학교 계열 구분을 정확하게 반영함.
원문	구어2	음, 너무 뜬구름 잡는 소리인 것 같네.
번역	기계 번역①	Well, parang medyo may kahabaan. (Well, that seems a bit of a stretch.)
	기계 번역②	Hmm, sa tingin ko iyon ay masyadong malabo. (Hmm, I think that's too vague.)
	병렬 말뭉치	Hmm, parang nakalutang talaga sa hangin ang sinasabi mo. (Hmm, what you're saying really seems to be floating in the air.)
비교		기계 번역에서는 원문의 ‘뜬구름 잡는 소리’가 ‘may kahabaan(a bit of a stretch)’으로 번역되어 근거 없고 허황된 말이라는 의미가 충분히 드러나지 않거나, ‘masyadong malabo(too vague)’로 단순히 모호함만 강조되어 원문의 비유적 의미가 누락됨. 병렬 말뭉치는 이를 ‘nakalutang sa hangin(floating in the air)’으로 옮겨 근거 없이 실체가 없다는 원문의 비유적 표현을 정확하게 반영함.
원문	구어3	비가 와서 물이 차기 시작하면 엄청 차요.
번역	기계 번역①	Kapag umuulan at nagsimulang tumaas ang tubig, talagang lumalamig. (When it rains and the water starts to rise, it gets really cold.)
	기계 번역②	Kapag umuulan at nagsimulang mapuno ang tubig, lumalamig talaga. (When it rains and the water starts to fill up, it gets really cold.)
	병렬 말뭉치	Kapag umulan at nagsimulang tumaas ang tubig, talagang punong-puno ito. (When it rains and the water starts to rise, it really fills up.)
비교		기계 번역에서는 원문의 ‘엄청 차요(차오르다)’를 ‘lumalamig(gets cold)’으로 번역해 ‘차다’의 의미를 온도 변화로 오해하는 오역이 발생함. 병렬 말뭉치는 이를 ‘차다(차오르다)’로 반영해 ‘talagang punong-puno ito(really fills up)’로 옮겨 물이 가득 찬다는 원문의 의미를 정확하게 반영함.

<표 48> 기계 번역과 병렬 말뭉치의 번역 비교(러시아어)

원문	문어1	검찰은 {이} 대표가 변호사법에서 정한 변호사로서의 품위 유지 의무를 위반했다는 사유를 제시했다.
번역	기계 번역①	Обвинение представило причину, по которой представитель Ли нарушил обязанность сохранять достоинство адвоката, как это предусмотрено в Законе о адвокатуре. (The prosecution presented the reason why Representative Lee violated his duty to maintain the dignity of a lawyer, as stipulated in the Attorney Act.)
	기계 번역②	Обвинение Представитель привел основания для того, что он нарушил обязанности адвоката по сохранению достоинства. (Charge The representative cited grounds for violating the attorney's duty to preserve dignity.)
	병렬 말뭉치	Прокуратура указала, что председатель Ли нарушил обязанность по сохранению профессионального достоинства, установленную Законом об адвокатуре. (The prosecutor's office stated that Chairman Lee had violated his duty to maintain professional dignity, as established by the Law on the Bar.)
비교		기계 번역은 '대표'를 단순한 대리인(представитель)으로 직역함. 병렬 말뭉치는 문맥을 살려 당 대표(председатель)라는 직함으로 정확히 번역했음.
	원문	문어2
번역	기계 번역①	Сообщается, что компания Apple намеревалась создать «неравные условия игры» в интересах своей платформы по токовой передачи музыки Apple Music. (It is reported that Apple intended to create an “unlevel playing field” in favor of its Apple Music streaming platform.)
	기계 번역②	Apple намеревалась создать 'светлую спортивную площадку', чтобы принести пользу своей платформе потокового музыкального потока Apple Music.

		(Apple intended to create a ‘bright sports arena’ to benefit its Apple Music streaming platform.)
	<b>병렬 말뭉치</b>	Предполагается, что Apple сознательно создала «неравные условия», чтобы извлечь выгоду для собственной музыкальной стриминговой платформы Apple Music. (It is alleged that Apple deliberately created “unequal conditions” in order to benefit its own music streaming platform, Apple Music.)
	<b>비교</b>	기계 번역②는 관용구인 기울어진 운동장을 물리적인 스포츠 경기장(светлую спортивную площадку)으로 엉뚱하게 오역함. 병렬 말뭉치는 이를 불공정한 경쟁 환경을 뜻하는 불평등한 조건(неравные условия)으로 정확히 의역했음.
<b>원문</b>	<b>문어3</b>	'명품 끼 패키지' {김소연}이 또다시 시청자들을 매료시켰다.
<b>번역</b>	<b>기계 번역①</b>	«Роскошный пакет» Ким Со Ён снова очаровала зрителей. (Kim So-yeon's “luxury package” charmed viewers once again.)
	<b>기계 번역②</b>	«Пакет подач предметов роскоши» Ким Со Ён Это снова привлекло зрителей. (“Luxury Item Delivery Package” Kim So-yeon This once again attracted viewers.)
	<b>병렬 말뭉치</b>	Обладательница «полного набора талантов» Ким Соён снова очаровала зрителей. (Kim So-yeon, who possesses a “full range of talents,” has once again charmed viewers.)
	<b>비교</b>	기계 번역은 명품 끼 패키지를 물리적인 ‘호화 소포(Роскошный пакет)’ 등 엉뚱한 표현으로 오역함. 병렬 말뭉치는 이를 다재다능한 매력을 뜻하는 ‘재능의 완전체(полный набор талантов)’로 정확히 의역했음.
<b>원문</b>	<b>구어1</b>	너는 {수시}에 대한 아쉬움은 없어?
<b>번역</b>	<b>기계 번역①</b>	Есть ли у вас сожаления по поводу процесса поступления? (Do you have any regrets about the admission process?)
	<b>기계 번역②</b>	У тебя нет сожалений о часах? (Do you have any regrets about the watches?)
	<b>병렬 말뭉치</b>	У тебя нет сожалений насчет досрочного поступления? (Do you have any regrets about enrolling early?)

<b>비교</b>		<p>기계 번역은 ‘수시’라는 한국의 입시 전문용어를 문맥에 맞게 식별하지 못해 일반적인 입학 과정이나 엉뚱한 단어(시간/시계)로 오역했음.</p> <p>병렬 말뭉치는 이를 러시아인이 이해하기 쉬운 ‘조기 입학(досрочное поступление)’이라는 정확한 대응어로 옮겨 원문의 의도를 완벽하게 살렸음.</p>
<b>원문</b>	<b>구어2</b>	<p>지금 학생들이 어떤지는 제가 잘 모르겠는데 야자를 이렇게 의무적으로 한다는 그런 거는 제가 모르겠습니다.</p>
<b>번역</b>	<b>기계 번역①</b>	<p>Я не знаю, как обстоят дела у нынешних студентов, но не знаю, считают ли они обязательным рисовать такие пальмы. (I don't know how things are with current students, but I don't know if they consider it mandatory to draw such palm trees.)</p>
	<b>기계 번역②</b>	<p>Я не знаю, что сейчас делают студенты, но я не знаю, что такое обязательное занятие пальмами. (I don't know what students are doing now, but I don't know what a mandatory palm tree class is.)</p>
	<b>병렬 말뭉치</b>	<p>Я не знаю, как обстоят дела у сегодняшних школьников, и не уверен насчет обязательной вечерней самоподготовки. (I don't know how things are for today's schoolchildren, and I'm not sure about compulsory evening self-study.)</p>
<b>비교</b>		<p>기계 번역은 야자(야간 자율 학습)를 식물인 야자수(пальмы)로 엉뚱하게 오역함.</p> <p>병렬 말뭉치는 이를 학교 문맥에 맞는 저녁 자습(вечерняя самоподготовка)으로 정확히 번역했음.</p>
<b>원문</b>	<b>구어3</b>	<p>그래서 {영끌}을 해서 그 집에 들어가야 되나 하는 생각을 하게 되잖아.</p>
<b>번역</b>	<b>기계 번역①</b>	<p>Итак, вы начинаете думать о том, стоит ли вам устроиться на работу и пойти в этот дом. (So, you start thinking about whether you should get a job and go to that house.)</p>
	<b>기계 번역②</b>	<p>Так что ты думаешь, мне нужно ли войти в тот дом, дела Яёнкыль. (So what do you think, should I go into that house doing</p>

		Yongkil?)
	<b>병렬 말뭉치</b>	И поэтому ведь начинаешь думать: может, надо вложить все до последней копейки, чтобы купить ту квартиру? (And so you start thinking: maybe you should invest every last penny to buy that apartment?)
<b>비교</b>		기계 번역은 영혼까지 끌어모아 투자한다는 신조어 영끝을 엉뚱하게 취업(устроиться на работу)으로 번역하거나 발음 그대로 음역함. 병렬 말뭉치는 이를 ‘마지막 한 푼까지 털어 넣다(вложить все до последней копейки)’라는 표현으로 의미를 정확히 살려 번역했음.

<표 49> 기계 번역과 병렬 말뭉치의 번역 비교(우즈베크어)

<b>원문</b>	<b>문어1</b>	캐나다 가압 중수로형 모델인 루마니아 {체르나보다 원전 1호기}는 오는 2026년 운전 허가일이 만료돼 30년 추가 운전을 위한 설비 개선 사업이 진행될 예정이다.
<b>번역</b>	<b>기계 번역①</b>	Ruminiyaning Sernavoda 1-atom elektr stantsiyasi, Kanada bosimli og'ir suv reaktori modeli, 2026 yilda foydalanish litsenziyasi tugashi sababli, qo'shimcha 30 yillik foydalanish imkonini beruvchi ob'ektni takomillashtirish loyihasidan o'tadi. (Romania's Sernavoda 1 nuclear power plant, a Canadian pressurized heavy water reactor model, will undergo a facility improvement project that will allow for an additional 30 years of operation, as its operating license expires in 2026.)
	<b>기계 번역②</b>	Ruminiyaning Cernaboda atom elektr stantsiyasining 1, Kanada bosimli og'ir suv reaktorining modeli, 2026 yilda foydalanishga ruxsat berish muddati tugashi rejalashtirilgan va ob'ektni yaxshilash loyihasi qo'shimcha 30 yillik faoliyat uchun amalga oshiriladi. (Romania's Cernaboda Nuclear Power Plant 1, a model of the Canadian pressurized heavy water reactor, is scheduled to expire in 2026, and a facility improvement project will be implemented for an additional 30 years of operation.)
	<b>병렬 말뭉치</b>	Kanadaning bosimli og'ir suv reaktor modeliga asoslangan

		<p>Rumaniyadagi Cernavoda AES 1-blok atom elektr stansiyasi kelayotgan 2026-yilda amaldagi foydalanish ruxsatnomasining muddati tugaganda yana 30 yilgacha ishlash imkonini yaratish maqsadida uskuna va tizimlarni yangilash loyihasi amalga oshirilishi rejalashtirilmoqda.</p> <p>(The Cernavoda NPP Unit 1 nuclear power plant in Romania, which is based on the Canadian pressurized heavy water reactor model, is planning to implement a project to upgrade equipment and systems in order to enable operation for up to another 30 years when the current operating license expires in 2026.)</p>
	<b>비교</b>	<p>기계 번역은 주어 처리 및 문장 연결이 부자연스럽고 어색함. 병렬 말뭉치는 문장 도입부('...ga asoslangan')를 우즈베크어 문법에 맞게 자연스럽게 처리했으며, 운전 연장 목적을 명확히 설명('...imkonini yaratish maqsadida...')하여 원문의 계획과 의도를 가장 정확하게 전달함.</p>
<b>원문</b>	<b>문어2</b>	{파리 바게뜨} 제빵 기사 노조 탈퇴 강요 의혹으로 name1 {SPC} 대표 이사가 4일 구속됐다.
<b>번역</b>	<b>기계 번역①</b>	<p>Paris Baguette bosh direktori name1 SPC 4-da novvoylarni kasaba uyushmasini tark etishga majburlaganlikda gumonlanib hibsga olingan.</p> <p>(Paris Baguette CEO name1 was arrested on suspicion of forcing bakers to leave the union in SPC 4.)</p>
	<b>기계 번역②</b>	<p>Parij Baget name1 SPC bosh direktori pishirish ritsarlarini ittifoqdan chiqishga majbur qilganlikda gumon qilinib, to'rt kun davomida hibsga olingan.</p> <p>(The CEO of Paris Baguette name1 SPC was arrested for four days on suspicion of forcing the Knights of Baking to leave the union.)</p>
	<b>병렬 말뭉치</b>	<p>Paris Baguette novvoylarni kasaba uyushmasidan chiqishga majburlashda gumonlangan SPC bosh direktori name1 4-sanada hibsga olindi.</p> <p>(SPC CEO name1, suspected of forcing Paris Baguette employees to withdraw from the labor union, was arrested on the 4th.)</p>
<b>비교</b>		기계 번역①의 'name1 SPC 4-da'와 기계 번역②의 'name1 SPC

		<p>bosh direktori... to'rt kun davomida hibsga olingan (4일 동안 구속되었다)'는 원문의 '4일 구속'이라는 시점을 잘못 전달하거나(기계 번역②), 주어와 술어의 배치가 어색함.</p> <p>기계 번역②의 '제빵 기사'에 대한 번역 'pishirish ritsarlarini (제빵 기사들)'은 직역투가 강하며 부자연스러움.</p> <p>병렬 말뭉치는 자연스러운 어순을 따르고 '제빵 기사'를 'novvoylarni (제빵사들)'로, '노조 탈퇴 강요'를 'kasaba uyushmasidan chiqishga majburlash (노조 탈퇴 강요)'로 정확히 표현함.</p>
<b>원문</b>	<b>문어3</b>	{우리 카드}의 '트래블 월렛 우리 카드'는 환전 가능 통화가 45종에 달하며 달러, 엔, 유로 3종에 한해 100% 환율 우대를 받을 수 있다.
<b>번역</b>	<b>기계 번역①</b>	<p>Bizning kartamizning "Sayohat hamyonimiz" kartasi 45 xil valyutani qo'llab-quvvatlaydi va faqat uchta valyuta uchun 100% imtiyozli ayirboshlash kursini taklif qiladi: AQSh dollari, Yaponiya iyenasi va evro.</p> <p>(Our card's "Travel Wallet" card supports 45 different currencies and offers a 100% preferential exchange rate for only three currencies: US dollars, Japanese yen, and euros.)</p>
	<b>기계 번역②</b>	<p>Our Card ning "Travel Wallet Woori Card" 45 ta valyutaga ega bo'lib, ular almashtirilishi mumkin va faqat 3 turdagi dollar, iyena va evroga 100% valyuta kursi imtiyozli rejimga ega.</p> <p>(Our Card's "Travel Wallet Woori Card" has 45 currencies that can be exchanged, and only 3 types have a 100% exchange rate preferential treatment for dollars, yen, and euros.)</p>
	<b>병렬 말뭉치</b>	<p>Woori Card kompaniyasining "Travel Wallet Woori Card" kartasi 45 turdagi valyutani ayirboshlash imkoniyatini beraib, AQSh dollari, Yaponiya yeni hamda yevro bo'yicha 100% kurs imtiyozidan foydalanish mumkin.</p> <p>(Woori Card's "Travel Wallet Woori Card" allows you to exchange 45 currencies, with a 100% exchange rate advantage for US dollars, Japanese yen, and euros.)</p>
<b>비교</b>		기계 번역①은 '우리 카드'와 상품명('트래블 월렛 우리 카드')을 단순 직역하여 'Bizning kartamizning "Sayohat hamyonimiz" kartasi (우리 카드의 "여행 지갑" 카드)'라고 번역해 고유 상품명을

		<p>정확히 전달하지 못함.</p> <p>기계 번역② 역시 '우리 카드'를 'Our Card'라고 영어식으로 직역함.</p> <p>병렬 말뭉치에서는 '우리 카드'와 상품명을 'Woori Card kompaniyasining "Travel Wallet Woori Card" kartasi'와 같이 음차 표기를 활용하여 정확하게 명시하였음.</p>
원문	구어1	너는 수학여행의 경험에 대해 어떤 기억이 있어?
번역	기계 번역①	Matematik sayohatingiz haqida qanday xotiralaringiz bor? (What memories do you have of your mathematical journey?)
	기계 번역②	Sen ekskursiya tajribang haqida qanday xotiralarga egasan? (What memories do you have of your excursion experience?)
	병렬 말뭉치	Senda maktab ekskursiyasi haqida qanday xotiralar qolgan? (What memories do you have of the school field trip?)
비교		<p>기계 번역①은 '수학여행'을 'Matematik sayohatingiz (수학 여행)'이라고 단순 직역하였고, 기계 번역②는 'ekskursiya (견학/소풍)'로 번역했지만, '수학여행'이 가진 '학교 차원의 단체 여행'이라는 의미를 담지 못함.</p> <p>기계 번역①의 'qanday xotiralaringiz bor?'는 직역투가 강하며, 기계 번역②의 'qanday xotiralarga egasan?' 역시 부자연스러움.</p> <p>병렬 말뭉치에서는 '수학여행'을 'maktab ekskursiyasi (학교 견학/소풍)'로 번역하여 학교 행사라는 문맥을 명확히 함.</p>
원문	구어2	그래서 저도 지나가다 인제, 오이지를 담았어요.
번역	기계 번역①	Xullas, men ham tasodifan o'tib qoldim va bodring terib oldim. (So, I happened to pass by and picked some cucumbers.)
	기계 번역②	Shuning uchun o'tib ketayotganimda Inje va Oijini ham qo'shdim. (So I added Inje and Oijin as I went along.)
	병렬 말뭉치	Shuning uchun men ham o'tib, bodringni tuzlab qo'ydim. (So I went ahead and pickled the cucumbers.)
비교		<p>기계 번역은 '오이지를 담다'라는 핵심 행위를 '오이를 따다' 또는 '인제와 오이지를 추가했다'고 완전히 오역하여 의미 전달에 실패함.</p> <p>병렬 말뭉치는 'bodringni tuzlab qo'ydim (오이지를 담갔다)'로 원문의 행위를 가장 정확하게 전달했으며, 구어 말뭉치 흐름도 자연스</p>

		럽게 살렸음.
<b>원문</b>	<b>구어3</b>	언니, 혹시 고등학교 때 보충 수업이나 {야자} 수업 같은 거 했었어?
<b>번역</b>	<b>기계 번역①</b>	Opa, siz o'rta maktabda qo'shimcha darslar yoki xurmo daraxti darslarida qatnashganmisiz? (Sister, did you take any extra classes or palm tree classes in high school?)
	<b>기계 번역②</b>	Opa, maktabda grim darslari yoki kaft darslari bormi? (Sister, are there makeup classes or palmistry classes at school?)
	<b>병렬 말뭉치</b>	Opa, siz maktabda o'qib yurganingizda qo'shimcha darslar yoki "kechki o'qish" kabilarni qilganmisiz? (Sister, did you take any extra classes or "after-school" classes while you were in school?)
<b>비교</b>	<p>기계 번역①은 '야자'를 'xurmo daraxti darslarida (야자수 수업)'로, 기계 번역②는 'kaft darslari (손금/야자수 수업)', 'grim darslari (화장/메이크업 수업)'으로 오역함.</p> <p>병렬 말뭉치는 '야자'를 'kechki o'qish (저녁 공부)'라는 일반적인 우즈베크어 표현으로 번역한 후, 따옴표를 사용하여 한국의 '야간 자율 학습'을 가장 근접하게 설명함.</p>	

## 2. 대외 활동

### 2.1. 국외 출장

한국어-외국어 병렬 말뭉치 구축 사업 홍보 및 국가 간 협력 도모를 위해 인도네시아와 사우디아라비아의 유관 기관을 공식 방문하였다.

#### 2.1.1. 인도네시아 출장

인도네시아의 말뭉치 구축 사업 관련 유관 기관을 방문하여 사업 홍보 및 상호 협력 방안을 마련하고자 2025년 11월 9일부터 11월 13일까지 3박 5일간의 일정으로 자카르타 및 반둥에서 일정을 수행하였다.

먼저 자카르타에서는 인도네시아국립대학교 한국학과 및 고영향 혁신 및 연구국(Directorate of High Impact Innovation&Research), 인도네시아 언어개발육성원(Badan Pengembangan dan Pembinaan Bahasa)을 방문하였다. 인도네시아 국립대학교 한국학과에서는 교수진과 면담을 진행하고 사업의 목적과 말뭉치 구축 현황, 번역 및 검수 체계와 지침서 내용을 설명하였다. 이어 한국학과 학생들을 대상으로 말뭉치 활용 가능성과 사업 취지를 안내한 후 한국학과와의 협력 방향을 논의하였다. 또한 인도네시아국립대학교 내의 고영향 혁신 및 연구국(Directorate of High-Impact Innovation&Research)을 방문하여 말뭉치 데이터의 연구 및 교육 분야의 활용 가능성을 논의하고 데이터 교류 및 학술 협력 등 구체적인 협력 방안을 모색하였다.

다음으로 인도네시아 언어개발육성원(Badan Pengembangan dan Pembinaan Bahasa)을 방문하여 국립국어원의 역할과 한국어-외국어 병렬 말뭉치 구축 사업의 추진 배경, 구축 절차, 활용 가능성을 설명하였다. 방문 기관은 인도네시아의 국가 말뭉치인 인도네시아어 말뭉치(Korpus Indonesia, KOIN) 구축 현황과 지역어 말뭉치, 사전 편찬 등 자국어 인프라 강화 사업을 공유하며 한국과의 언어 데이터 분야 협력에 높은 관심을 보였다. 특히 향후 자료 제공 등 실질적 협력 가능성과 공식 업무 협약(MOU) 체결 의사를 제안하였으며, 이에 따라 국립국어원과 기관 간 협력 방안을 함께 모색하기로 하였다.

반둥공과대학교 인공지능센터에서는 국립국어원과 한국어-인도네시아어 병렬 말뭉치 구축 사업의 추진 배경, 구축 절차 및 활용 가능성을 설명하였다. 해당 센터는 인도네시아어 자연어 이해 평가 데이터셋(IndoNLU), 인도네시아어 BERT 모델(IndoBERT) 등 인도네시아어 기반 자연어 처리 자원 구축과 인공지능 모델 개발

을 수행하고 있으며, 회의록 자동 생성, 아동 보호용 인공지능 기술, 실시간 음성-텍스트 변환(Speech-to-Text, STT) 프로그램 등 다양한 연구를 진행하고 있음을 공유하였다. 또한 한국어-인도네시아어 병렬 말뭉치를 활용한 한국어 어휘 학습 애플리케이션 개발 의사를 밝히며 협력 가능성을 논의하였다.

또한 방문 일정 중 케이팝포스트(KPopPost) 기자와의 인터뷰를 통해 국립국어원의 '모두의 말뭉치'와 한국어-외국어 병렬 말뭉치 구축 사업의 추진 배경, 주요 성과, 인공지능 기술 발전에의 기여 등을 알리며 사업을 대외적으로 홍보하였다.

이와 함께 인도네시아 현지에서 감수를 하고 있는 감수자와의 면담 자리를 별도로 마련하여 최근 검수 진행 상황과 품질 관리 결과를 공유하였다. 그리고 감수 과정에서 주로 발견되는 오류 유형과 반복적 문제를 점검하고 향후 발전 방향에 대해 논의하였다.

## Corpus Language for All: The Hidden AI Project Powering the Future of K-Content

by Jessica H — November 21, 2025 in Interview

0



Prof. Lee Jung Hee visits Indonesia

On a warm, gentle morning in Bandung, sunlight filtered through the glass windows of a small café inside **Institut Teknologi Bandung (ITB)**. There, in that cozy cafeteria, I sat across from **Prof. Lee Jung Hee** and researcher **Kim Yeon Hee**. They are the key figures behind a groundbreaking **Korean-Foreign Language Parallel Corpus Project** that could transform the way millions around the world experience **K-dramas, K-pop, and Korean culture**.

[그림 44] 케이팝포스트(KPopPost) 보도 기사 일부 발췌

<표 50> 케이팝포스트(KPopPost) 보도 기사 내용 요약

기사 제목	모두를 위한 말뭉치: K-콘텐츠의 미래를 이끄는 숨겨진 AI 프로젝트
주요 내용	<p>한국어-외국어 병렬 말뭉치 구축 사업과 이를 기반으로 한 인공지능 번역 기술의 발전 가능성을 소개하였다.</p> <p>해당 사업에서는 한국어와 인도네시아어, 베트남어, 태국어 등 여러 언어 간 병렬 말뭉치를 구축하고 있으며, 이러한 언어 데이터가 AI 번역 성능 향상에 기여할 수 있음을 설명하였다. 특히 일상 대화와 구어 표현을 포함한 말뭉치 구축이 K-드라마와 K-팝 등 K-콘텐츠 자막 번역의 품질과 자연스러움을 높이는 데 기여할 수 있음을 강조하였다. 또한 인도네시아 관련 기관 및 대학과의 협력을 통해 다국어 언어 데이터 구축과 AI 연구 협력이 확대될 가능성도 언급하였다.</p>



인도네시아국립대학교 한국학과 학과장 면담



인도네시아국립대학교 한국학과 학생 간담회



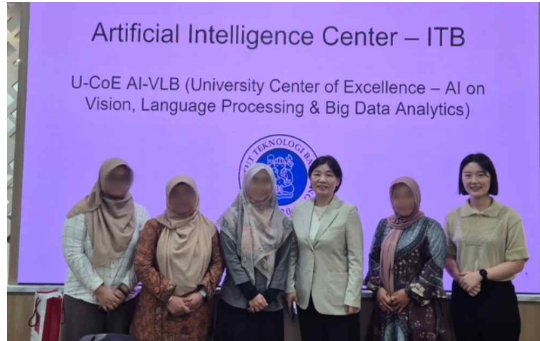
인도네시아국립대학교 고영향 혁신 및 연구국(Directorate of High-Impact Innovation&Research) 방문



인도네시아국립대학교 한국학과 교수 간담회



인도네시아 언어개발육성원(Badan Pengembangan dan Pembinaan Bahasa) 방문



반등공과대학교 인공지능센터(Artificial Intelligence Center) 방문



케이팝포스트(KPopPost) 인터뷰



감수자 면담

[그림 45] 인도네시아 출장 일정별 사진

### 2.1.2. 사우디아라비아 출장

사우디아라비아의 독자 인공지능과 대규모 언어 모형의 구축 현황을 조사하고 협력 방안을 마련하여 제2차 중장기계획 시행 과정에서 한국어-아랍어 말뭉치를 효과적으로 구축하기 위한 방안을 모색하고자 2025년 11월 23일부터 11월 27일까지 3박 5일의 일정으로 사우디아라비아 리야드에 출장을 다녀왔다.

먼저 사우디아라비아 리야드 및 중부 지역 한인회장과 면담하였다. 사업단은 한국어-외국어 말뭉치 사업의 목적과 취지를 소개하고 제2차 중장기계획 시행 과정에서 아랍어 구축의 필요성이 제기되어 데이터 구축의 중요성과 가능성 등을 타진하고자 왔음을 언급하였다. 한인회장은 사우디 및 중동의 문화적 특징을 설명하는 한편 아랍어 말뭉치 구축 시 유의해야 할 사항을 전해 주었다.

다음으로 사우디아라비아 데이터인공지능청(Saudi Data & AI Authority, SDAIA)를 방문하여 데이터 담당자와 논의를 진행했다. 한국 정부의 국가 인공지능 구축 프로젝트의 일환으로 한국어-외국어 말뭉치 구축 사업을 진행하고 있음을 설명하고 향후 한국어-아랍어 말뭉치 구축 가능성을 모색하기 위해 사우디 내 협력

기관을 탐색하고 있음을 밝혔다. SDAIA 측은 데이터 수집과 교환 범위 등에 관한 기술팀 논의를 거쳐 고도화된 협력 모델을 만들기를 원하며 이를 위해서는 고위급의 논의가 필요하다고 언급했다.

이어 리야드 글로벌 비즈니스 센터(Global Business Center, GBC)를 방문하여 소장과 면담을 했다. 소장은 중동 지역에서 국가 단위의 사업을 진행할 때 사우디아라비아가 특히 중요하다고 주장하였다. 또한 번역 검수 인력 섭외 시 현지의 비싼 인건비를 고려할 필요가 있다고 조언했으며 감수 교수 섭외 인력으로 킹 사우드 대학교(King Saud University)에 오래 재직한 한국인 교수를 추천했다.

그리고 주사우디아라비아 대한민국 대사관에서 사업단을 초청하여 만찬을 했다. 대사 대리는 아랍어는 데이터 표준화가 매우 어렵고, 특히 구어는 국가·지역마다 차이가 커서 큰 예산을 쓰더라도 효율이 매우 낮으니 표준 아랍어 혹은 문어 중심으로 데이터를 구축할 것을 제안했다. 또한 문화와 네트워크 이해가 중동 사업 성공의 핵심 요소임을 강조했다.

둘째 날에는 사우디아라비아 왕세자가 주도하여 세운 인공지능 기업인 휴메인(HUMAIN)을 방문하여 부사장을 비롯한 데이터 담당자들과 논의를 했다. 사업단은 사우디 현지 기관과 협력해 아랍어 말뭉치 및 발음 데이터 구축이 가능한지를 탐색하는 것이 이번 회의의 핵심 목적임을 언급하였고, 부사장은 기존 아랍어 데이터 구축에서 합성 데이터를 활용하였으나 인간의 검증으로도 합성 데이터 사용에 한계가 있음을 언급하며 고품질 데이터 구축의 필요성과 중요성에 공감하였다.

다음으로 빈시혼 타이어 컴퍼니(BIN-SHIHON TIRES COMPANY)를 방문하여 최고 경영자(CEO)와 면담하였다. 한국 기업과 사업을 진행하는 과정에서 언어로 인해 발생하는 문제는 없는지 청취하고 현지 기업이 한국 기업과 교류하는 과정에서 필요한 말뭉치 주제가 무엇인지 조사하였다.

마지막 날에는 프린스 술탄 대학교(Prince Sultan University)에 있는 리야드 세종학당을 방문했다. 사업단은 번역 검수 가능 인력의 현황을 확인하기 위해 학당을 방문했으며, 운영 요원은 한국어-아랍어 말뭉치 구축에 참여할 수 있는 인력을 추천해 주었으나 한국어와 영어 모두 능통한 학생들이 다수 존재함에도 별도의 직업을 갖고 있는 경우가 많고 현지 고급 인력의 급여가 높아 고용에 난항이 있을 수 있음을 언급했다. 또한 대한민국과 사우디아라비아 간 공식 행사에서 사우디 국적 전문 통역 인력이 거의 없기 때문에 전문 인력 양성이 필요함을 강조했다.

세종학당 학생 및 수료생과의 간담회를 통해서도 자연스러운 한국어 구사가 가능한 사우디아라비아 학습자들을 만날 수 있었다. 간담회 참석자들은 독학 기간을 포함하여 8~17년간 한국어를 학습하였으며 미디어의 영향으로 어려서부터 한국어와 한국 문화를 접할 수 있었다고 말했다. 사업단은 이들을 대상으로 사업의 취지

와 번역 검수 방식을 설명한 후 사업 참여 의사를 타진하였다. 참가자 다수는 긍정적 의사를 전달하였으나 본업으로 인해 사업 참여가 어렵다고 언급한 참가자도 있었다. 참가자들은 이 외에도 아랍어 구어 구축 시 표준 아랍어로 번역할 때와 사우디 방언으로 구축할 때의 장단점을 언급해 주었다.

이후 진행된 세종학당 교원과의 간담회에서는 현지의 한국어 및 한국 문화 교육 상황과 함께 문화 차이로 인해 현지에서 수업을 진행하면서 겪은 어려움에 대해 들을 수 있었으며 마지막 일정인 리야드 세종학당장과의 면담에서는 향후 한국어-아랍어 말뭉치 구축 시 협조를 부탁하였다.



리야드 및 사우디 중부지역 한인회장 면담



사우디아라비아 데이터인공지능청(Saudi Data & AI Authority, SDAIA) 방문



리야드 글로벌 비즈니스 센터(Global Business Center, GBC) 방문



주사우디아라비아 대한민국 대사관 초청 만찬



휴메인(HUMAIN) 방문



빈시혼 타이어 컴퍼니(BIN-SHIHON TIRES COMPANY) 방문



리야드 세종학당 운영 요원 면담



리야드 세종학당 학생 및 수료생 간담회



리야드 세종학당 교원 간담회



리야드 세종학당장 면담

[그림 46] 사우디아라비아 출장 일정별 사진

## 2.2. 모두의 말뭉치 전문가 지식 강연회

2025년 모두의 말뭉치 전문가 지식 강연회는 2025년 10월 1일(수) 롯데월드타워 SKY31 컨벤션 오디토리움에서 국립국어원 주최, 경희대학교와 (주)플리토 주관으로 개최되었다. 강연회는 ‘인공지능 시대, 우리의 언어 자원과 미래’라는 주제로 인공지능 기술의 발전과 언어 자원의 융합이 가속화되는 시점에서 언어 자원의 공공적

가치와 활용 방향을 국민과 함께 모색하고 학계와 산업계 전문가의 전문 지식을 대중과 공유하기 위한 소통의 장으로 마련되었다. 모두의 말뭉치 전문가 지식 강연회는 1부 주제 강연과 2부 포럼으로 진행되었으며, 현장 참여가 어려운 국민도 실시간으로 참여할 수 있도록 유튜브 생중계를 병행 운영하였다.

<표 51> 2025년 모두의 말뭉치 전문가 지식 강연회 일정표

세션	시 간	내 용	발표자
1부 주제 강연 사회: 이재성 아나운서(KBS)	14:00-14:55	말뭉치, 유창성을 넘어 AI 3대 강국의 지름길	조용민 대표 (주)언바운드랩데브
	14:55-15:10	휴식 시간	
2부 포럼 사회: 이재성 아나운서(KBS)	15:10-15:30	[발표] 한국형 인공지능 현황	이정수 대표 (주)플리토
	15:30-16:00	[토론] 한국형 인공지능의 기술, 데이터 그리고 언어문화적 요인	이정수 대표 (주)플리토), 조용탁 수석 (한국지능정보사회진흥원), 장정훈 CTO (주)와이즈넷
	16:00-16:15	청중 질의응답	
	16:15-16:30	만족도 조사 및 기념품 배부	

### 1) 1부(주제 강연)

1부는 ‘언어 자원과 인공지능 기술 개발의 미래’를 주제로, 조용민 대표가 ‘말뭉치, 유창성을 넘어 AI 3대 강국의 지름길’이라는 제목으로 주제 강연을 진행하였다.

<표 52> 2025년 모두의 말뭉치 전문가 지식 강연회 1부 주제 강연 내용

강연자	제목	내용
조용민 [주]언바운드랩데 브]	말뭉치, 유창성을 넘어 AI 3대 강국의 지름길	한국이 AI 3대 강국으로 도약하기 위해서는 기술 경쟁만이 아니라 한국어의 콘텍스트(맥락) 를 기반으로 한 전략이 필요하다. 데이터는 양 보다 말뭉치의 품질과 맥락 해석이 중요하며,

		<p>한국어는 단어 다의성과 문화적 맥락이 풍부하므로 이를 정확히 반영하는 학습이 핵심이다. 이를 위해 비가독 문서를 머신 리더블 데이터(기계가 이해할 수 있는 형식의 데이터)로 변환하는 기술 기반이 마련되어야 하며, 출판·언론·플랫폼 등 민간의 참여를 유도할 데이터 공유·보상 체계도 필요하다. 또한 플루언시(자연스러움)와 데이터 주권을 구분해 접근하고, 금융·상거래·음성·이미지 등 글로벌 수요가 높은 분야부터 한국어 특화 애플리케이션을 개발해 실사용 환경에서 성능을 검증해야 한다. 국립국어원은 말뭉치의 도메인 다양화와 투명한 이용 인프라를 통해 ‘말뭉치-모델-응용’의 선순환 구조를 구축함으로써, 한국어의 정밀한 맥락 해석 능력을 세계적 경쟁력으로 전환해야 한다.</p>
--	--	---

## 2) 2부(포럼)

2부는 ‘한국형 인공지능의 중요성과 과제’를 주제로 한 포럼이 진행되었다. 먼저 이정수 대표가 ‘한국형 인공지능 현황’을 발표하였으며, 이어 ‘한국형 인공지능의 기술, 데이터 그리고 언어문화적 요인’을 주제로 이정수 대표, 조용탁 수석, 장정훈 CTO가 참여한 토론이 이어졌다. 토론의 의제는 ‘독자 AI 경쟁력 확보’, ‘한국형 AI 발전을 위한 데이터 구축’, ‘최적의 민관 협력 모델’이었다. 토론 참여자들은 한국형 인공지능이 글로벌 경쟁 속에서 성장하기 위해서는 기술과 제도뿐 아니라 사회적 신뢰와 활용 가능한 데이터 생태계, 그리고 공공·민간·학계의 역할 분담이 필요하다는 점을 논의했다.

<표 53> 2025년 모두의 말뭉치 전문가 지식 강연회 2부 포럼 발표 내용

발표자	제목	내용
이정수 [㈜플리토]	한국형 인공지능 현황	<p>인공지능 발전의 핵심은 알고리즘이나 하드웨어보다 데이터의 질과 다양성에 있으며, 특히 한국의 경쟁력은 언어와 문화적 맥락을 정밀하게 반영한 고품질 데이터 구축에 달려 있다. 미국과 중국이 자본과 인재, 방대한 데이터를 기반으로 AI를 주도하는 가운데, 한국은 언어 자원의 정제와 데이터 관리 역량 강화를 통해 차별화된 경쟁력을 확보할 필요가 있다. 현재 해외 LLM과 인공지능은 각국의 법·윤리·문화 규</p>

		<p>범과 데이터로 학습되어 있기 때문에, 이를 그대로 사용할 경우 한국 사회의 행정·의료·문화 영역에서 법적·윤리적 충돌이나 정보 왜곡이 발생할 가능성이 있다. 따라서 한국의 언어·문화·법 체계를 반영한 독자적인 AI 생태계 구축이 필요하다. 이를 위해 국립국어원이 오랜 기간 축적해 온 방대한 말뭉치를 세분화하고 고도화하여 산업적 활용성을 높여야 하며 나아가 국내의 외국인 노동자와 관광객 등 다양한 언어 사용자를 포용할 수 있는 다언어 데이터 확장이 병행되어야 한다. 또한 한국은 파운데이션 모델 자체의 경쟁보다 한국어 기반의 실용적 솔루션 개발과 응용 확장을 통해 AI 생태계의 자립성과 글로벌 확장성을 확보해야 한다.</p>
--	--	--

<표 54> 2025년 모두의 말뭉치 전문가 지식 강연회 2부 포럼 토론 내용

토론자	제목	내용
<p>이정수 [㈜플리토]</p> <p>조용탁 [한국지능정보사 회진흥원]</p> <p>장정훈 [㈜와이즈넷]</p>	<p>의제 1 ‘독자 AI’의 경쟁력 확보</p>	<p>‘독자 AI 전략’이 글로벌 파운데이션 모델 중심의 경쟁 구도 속에서 실질적인 경쟁력을 확보하기 위해서는 기술과 제도뿐 아니라 사회적 공감대와 문화적 기반이 함께 마련되어야 한다. 정부는 세부 기술을 통제하기보다 산업이 자율적으로 성장할 수 있는 생태계를 설계해야 하며, 영화 산업의 스크린 쿼터처럼 성장의 방향과 여건을 조성하는 역할에 집중해야 한다. 독자 AI의 기반이 되는 컴퓨팅 자원과 데이터센터, 학습 데이터는 각각 땅과 씨앗, 재배 기술에 비유될 만큼 필수적인 요소로, 이를 체계적으로 구축하고 확장해야 한다. 동시에 인공지능과 데이터 학습 구조에 대한 이해를 높이는 반복 학습 중심의 인적 역량 강화가 요구된다. 무엇보다 한국의 데이터와 기술로 세계 시장에서 통할 수 있는 성공적인 사례를 만들어 내는 것이 중요하며, 이러한 실질적 성과가 국민적 신뢰와 산업의 자신감을 형성하는 출발점이 된다.</p>
	<p>의제 2 한국형 인공지능 기술 발전을 위한 데이터 구축</p>	<p>한국형 인공지능의 경쟁력을 높이기 위해서는 방대한 양의 데이터 확보보다 목적에 맞고 활용 가능한 데이터 구축 전략이 우선되어야 한다. 현재 한국어 데이터는 전반적으로 부족하며, 전문가가 생산하는 지식 기반 데이터와 일상 언어</p>

		<p>데이터 모두 급격히 줄고 있어 기본 데이터의 기반을 먼저 마련하는 것이 중요하다. 이후 과학·기술 등 전문 분야로 확장하여 단계적으로 고도화된 학습 데이터를 축적해야 한다. 데이터 구축은 부처별로 분산된 방식이 아니라, 각 기관이 자신의 전문 영역에서 고품질 데이터를 생산하고 이를 상호 연계해 학습 가능한 형태로 통합하는 방향으로 이루어져야 한다. 동시에 데이터의 양적 확대보다 명확한 목적과 사용처에 따른 구축 우선순위 설정이 필요하며, 사회와 산업에서 실제로 활용될 수 있는 소비 중심의 데이터 생태계로 전환해야 한다. 또한 외국인 노동자나 다양한 언어권 사용자를 포용할 수 있는 병렬 말뭉치와 다문화 언어 데이터의 확충은 한국형 인공지능의 확장성과 포용성을 높이는 핵심 과제로 제시되었다. 결국 데이터 전략의 초점은 ‘얼마나 많이 모으는가’가 아니라 ‘무엇을, 왜, 어떻게 쓸 것인가’에 있으며, 이러한 목표 중심의 데이터 구축 체계가 한국형 인공지능 발전의 실질적 경쟁력을 결정하게 될 것이다.</p>
	<p>의제 3 최적의 민관 협력 모델</p>	<p>민관 협력 모델이 효과적으로 작동하기 위해서는 공공과 민간, 그리고 학계가 각자의 강점을 살려 역할과 책임을 명확히 분담하는 구조가 필요하다. 공공은 누구나 활용할 수 있는 공통 기반 데이터를 구축하고 안정적인 제도적 환경을 조성해야 하며, 민간은 이를 바탕으로 창의적 아이디어와 산업적 경험을 결합한 혁신적 서비스를 개발해야 한다. 이러한 협력 구조는 공공이 만든 기반 위에서 민간이 부가가치를 창출하고, 그 성과가 다시 공공 영역으로 환류되어 재투자되는 순환형 발전 구조로 이어져야 지속 가능성을 확보할 수 있다. 또한 인공지능 기술의 고도화를 위해 대학이 협력의 중심축으로 참여해야 하며, GPU 등 고성능 연구 인프라를 확충해 학계와 산업이 상호 보완적으로 성장할 수 있는 생태계를 마련해야 한다. 정부는 인공지능 개발 과정에서 발생하는 저작권, 공정 이용, TDM 등과 관련된 규제의 장벽을 완화하고, 데이터 활용에 대한 명확한 법적 기준과 책임 체계를 정비해 민간이 자유롭게 개발에 집중할 수 있는 환경을 만들어야 한다. 결국 공공은 기반</p>

	<p>을 설계하고, 민간은 산업화를 주도하며, 학계는 인재와 연구를 뒷받침하는 삼자 협력 구조가 한국형 인공지능의 지속적 성장을 이끌 현실적 모델이 될 것이다.</p>
--	---

### 3) 행사 사진



식전 환담회



식전 기념사진



식전 라운지 전경



모두의 말뭉치 검색대



사회 - 이재성 아나운서



1부 주제 강연 - 조용민 대표



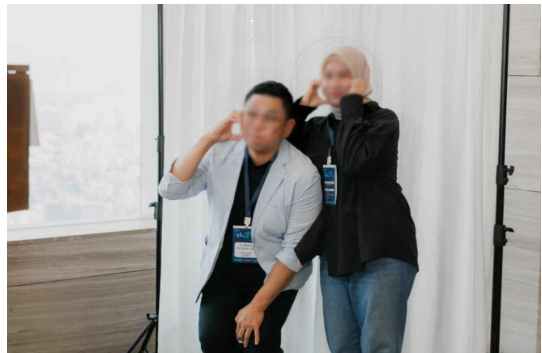
1부 주제 강연 전경



1부 주제 강연 - 청중 질문



1부 기념사진



휴식 시간 - 말뚝치 인생 네 컷 사진



2부 발표 - 이정수 대표



2부 - 포럼 토론



2부 포럼 토론 - 조용탁 수석



2부 포럼 토론 - 장정훈 CTO



2부 포럼 - 청중 질문



유튜브 생중계 화면

[그림 47] 2025년 모두의 말뭉치 전문가 지식 강연회 행사 사진

#### 4) 성과 및 기대 효과

경희대학교 산학협력단은 국립국어원의 주최 및 총괄 아래 ‘2025년 모두의 말뭉치 전문가 지식 강연회’를 기획·운영하였다. 국립국어원은 행사 전반의 기획 방향 설정, 프로그램 구성 검토, 연사 섭외 자문, 홍보 등 핵심 과정에 적극 참여하였으며, 이를 바탕으로 경희대학교 산학협력단과 (주)플리토는 실무 운영을 담당하여 행사를 안정적으로 추진하였다. 또한 사전 접수를 위한 홍보 메일 발송과 온라인 포스터 게시 등 홍보 활동을 공동으로 추진하여 ‘모두의 말뭉치’에 대한 대국민 인지도 제고를 도모하였다.

본 강연회에는 대면 124명, 비대면 110명 등 총 234명이 참여하였으며, 국립국어원의 기획 의도에 따라 유튜브 생중계를 병행 운영하여 지역적·시간적 제약으로 현장 참석이 어려운 국민도 실시간으로 강연에 참여할 수 있도록 하였다. 이를 통해 전문가와 일반 국민이 동시에 참여하는 열린 공공 강연의 형식을 구현하고, 공공 행사의 접근성과 참여 기회를 확대하였다.

본 강연회는 학계·산업계·공공 기관 전문가와 일반 국민이 함께 참여하는 개방형 공론의 장으로 운영되었으며, 인공지능 시대 언어 자원의 전략적 가치와 한국형 독자 AI 구현을 위한 데이터 기반의 중요성에 대해 전문가 중심의 논의와 함께 일반 국민의 이해 증진이 병행되었다. 이를 통해 공공 언어 자원의 사회적 역할과 국가 인공지능 경쟁력과 연계성에 대한 국민 인식 확산이 이루어졌다.

특히 ‘모두의 말뭉치’의 구축 성과, 실제 활용 사례, 향후 고도화 방향을 중심으로 발표 및 토론이 이루어져, 전문가에게는 연구 및 기술 활용과 관련한 정보 제공이 이루어졌으며, 일반 국민에게는 언어 자원의 활용 가능성과 사회적 가치에 대한 이해를 제고하는 효과가 나타났다. 이는 향후 학계의 연구 활용, 산업계의 기술 실증 및 서비스 개발, 국민의 공공 데이터 인식 확산으로 연계될 수 있는 정책

적·산업적·사회적 기반을 마련한 성과로 평가된다.

종합적으로 본 강연회는 전문가와 일반 국민이 함께 참여하는 구조 속에서 공공 언어 자원의 인지도 및 활용 수요를 동시에 제고하고, 산·학·연·관 협력 네트워크를 강화함으로써 한국형 독자 AI 생태계의 데이터 경쟁력 제고에 기여하였다. 또한 향후 국가 언어 자원 정책에 대한 국민 수용성 확대와 공공 데이터 기반 민간 활용 활성화에 기여할 것으로 기대된다.

## 5) 대외 홍보

2025년 모두의 말뭉치 전문가 지식 강연회는 행사 개최에 앞서 사전 보도를 통해 본 강연회를 대외적으로 홍보하였다. 언론 홍보는 인공지능 시대에 언어 자원의 중요성을 조명하고 국립국어원의 언어 정보화 사업을 대국민 차원에서 알리는 것을 목적으로 추진되었다. 보도 자료에는 ‘인공지능 시대, 우리의 언어 자원과 미래’라는 강연회 주제와 취지를 중심으로 모두의 말뭉치 사업의 비전, 주요 강연자 및 프로그램 구성, 온라인 생중계 참여 안내 등의 내용이 포함되었다. 이를 통해 본 강연회가 학계·산업계 전문가뿐만 아니라 일반 국민에게도 개방된 공공 지식 행사임을 효과적으로 부각하였다. 해당 보도 자료는 뉴시스, 뉴스핌 등 주요 언론 매체를 통해 게재되었으며, 각 매체의 기사에서는 국립국어원의 인공지능 언어 자원 구축 사업의 추진 방향과 이번 강연회의 사회적 의의가 중점적으로 다루어졌다. 이를 통해 강연회 및 ‘모두의 말뭉치’에 대한 대국민 인지도 제고와 관심 확산의 성과를 거두었다.

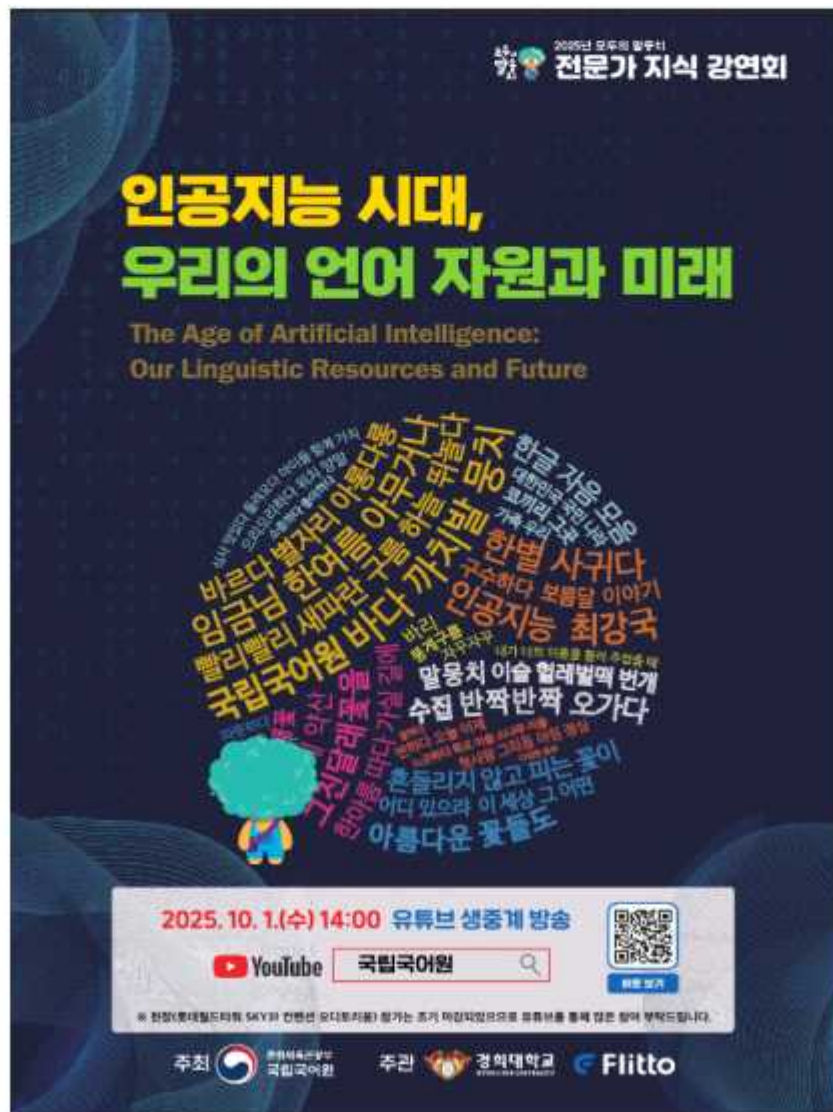
# 국립국어원, '모두의 말뭉치 전문가 지식 강연회' 개최

기사일련번호 : 2025년09월30일 08:22 | 최종수정 : 2025년09월30일 08:22



[서울=뉴스핌] 최문선 기자 = 국립국어원은 1일, 잠실 롯데월드타워 스카이 31 컨벤션에서 2025년 '모두의 말뭉치 전문가 지식 강연회'를 개최한다고 밝혔다.

이번 강연회는 인공지능 시대를 맞아 한국어 말뭉치와 한국어언어문화 자원의 가치를 새롭게 조명하고, 그 미래를 함께 논의하기 위해 마련된 자리로, 참가 신청을 한 학계와 산업계의 신진 연구자 및 학생 등 150여명이 현장에 참석한다.



[서울=뉴스핌] 최문선 기자 = 국립국어원, '2025년 모두의 말뭉치 전문가 지식 강연회' 포스터. [사진 =국립국어원] 2025.09.30 moonddo00@newspim.com

[그림 48] 2025년 모두의 말뭉치 전문가 지식 강연회 홍보 예시(뉴스핌)

### 3. 기대 효과 및 제언

#### 3.1. 병렬 말뭉치의 활용 방안

인공지능 기술은 자연어 처리 기술의 발전과 밀접하게 연결되어 있으며 한국어를 기반으로 하는 병렬 말뭉치 구축의 중요성이 날로 커지고 있다. 일정 규모의 병렬 말뭉치가 구축되었을 때 산업계 및 학계, 교육 분야에 활용할 수 있는 방안은 다음과 같다.

##### 1) 거대 언어 모델 및 기계 번역 시스템의 고도화

신경망 기계 번역 시스템을 고도화하는 데 병렬 말뭉치는 필수적인 학습 자원이다. 고품질의 대규모 병렬 말뭉치를 학습에 활용함으로써 한국어와 외국어 간의 문맥을 더 정확히 파악하고 자연스러운 번역 결과를 생성할 수 있다. 이러한 고성능의 기계 번역 시스템은 다국어로 진행되는 화상 회의 및 국제 행사, 외국인 관광 서비스나 이주민을 위한 실생활 서비스 등 다양한 분야에서 활용할 수 있다. 아울러 법률, 의료, 금융 등 특정 분야에 특화된 병렬 말뭉치를 활용한다면 해당 분야에 최적화된 전문 번역 모델을 개발할 수 있다.

##### 2) 기업 실무 및 비즈니스 운영에 적용

기업에서는 병렬 말뭉치를 업무 효율화와 국제 경쟁력 강화를 위한 자원으로 활용할 수 있다. 예를 들어 다국어로 작성된 계약서나 보고서와 같은 문서에서 핵심 정보를 추출하거나 다국어 문서를 자동으로 요약하고 분류하는 모델을 개발함으로써 업무 자동화를 구현할 수 있다. 이러한 시스템 도입을 통해 문서 관리에 소요되는 비용을 절감하고 업무 처리 속도를 크게 높일 수 있다. 이뿐만 아니라 기업이 새로운 해외 시장에 진출하기 위해 해당 국가와 지역의 언어로 된 병렬 말뭉치를 활용하여 고도화된 감성 분석 모델을 개발할 수 있다. 이를 통해 현지 고객의 브랜드 인식과 소비자의 반응을 심층적으로 파악할 수 있으며, 분석 결과를 바탕으로 제품 및 서비스 전략을 현지화하는 등 객관적인 데이터에 기반한 의사 결정이 가능해진다.

### 3) 언어 연구 및 교육 기반 구축

병렬 말뭉치는 언어 및 외국어 번역 연구와 교육을 위한 기초 자료나 도구를 개발하는 데 활용할 수 있다. 상세한 내용은 다음과 같다.

<표 55> 병렬 말뭉치의 연구 및 교육 분야 활용 방안

언어 연구 및 교육	<ul style="list-style-type: none"> <li>- 한국어 학습용 챗봇, 모바일 학습 응용 프로그램, 용례 검색기</li> <li>- 한국어 교재 개발(특히 현지 맞춤 교재)을 위한 언어 자료</li> <li>- 해외 한국어 수업·자가 학습용 보조 교재 제작을 위한 자료</li> <li>- 한국어와 외국어의 대조 연구를 위한 자료</li> <li>- 한국어 학습자의 중간 언어 및 오류 분석을 위한 분석 자료</li> <li>- 이중 언어 사전 편찬에서 용례, 언어, 대역어 선정을 위한 자료</li> <li>- 이주민 및 근로자 등을 위한 사회 통합 교육의 기초 자료</li> </ul>
외국어 번역 연구 및 교육	<ul style="list-style-type: none"> <li>- 번역학에서 번역 텍스트의 특성 분석을 위한 자료</li> <li>- 통번역 전문가 양성을 위한 실무 교육 자료</li> <li>- 다양한 실무 현장에서 정확한 번역을 위한 참고 자료</li> <li>- 자동 번역 등 기계 번역 분야 연구 자료</li> </ul>

### 3.2. 사업의 기대 효과

#### 1) 핵심 인공지능 자원의 확보 및 언어 데이터 산업의 생태계 조성

국가 기관 주도로 구축되는 대규모·고품질의 병렬 말뭉치는 4차 산업 혁명 시대의 핵심 인공지능 자원을 선제적으로 확보하고, 언어 데이터 산업의 건전한 생태계를 조성한다는 점에서 중요한 의미를 지닌다. 병렬 말뭉치는 국가 차원의 인공지능 경쟁력을 강화하는 데 기여할 수 있으며, 정부 기관과 산업계·학계에서 다양한 혁신의 기반으로 활용될 수 있다. 특히 병렬 말뭉치는 관광이나 교육 등 여러 산업 분야에 접목되어 새로운 비즈니스 모델을 창출하는 마중물의 역할을 할 수 있다. 예를 들어, 공공 데이터베이스를 활용해 지역 특화 관광 서비스를 개발하거나 전 세계적으로 증가하는 한국어 학습 수요에 발맞춰 고도화된 교육 서비스 및 학습 자료를 생성하는 데 활용될 수 있다.

#### 2) 다국어 언어 처리 및 인공지능 기반 통·번역 모델의 품질 향상

인공지능 기술의 발전으로 번역 자동화 비율이 높아지면 학습용 데이터를 효과적으로 재생산할 수 있게 되며, 결과적으로 번역 자동화 비용을 절감하는 선순환

적 구조를 형성할 수 있다. 그리고 번역 품질을 자동으로 예측하는 기술을 접목한다면 인공지능이 초별 번역을 진행하고 전문 번역가가 검수하는 협업 모델을 구축함으로써 번역 업무의 효율성을 크게 제고할 수 있다. 이러한 자동 번역 기술의 고도화는 언어 장벽을 낮춰 신한류 콘텐츠의 글로벌 확산에도 크게 기여할 수 있을 것이다.

### 3) 국가 간 미래 지향적 경제 협력 기반의 구축

국립국어원이 주도하는 한국어-외국어 병렬 말뭉치 구축 사업은 그동안 영어, 중국어, 일본어 등에 비해 데이터 확보가 미진했던 언어 자원을 확보하였다는 점에서 큰 의의를 지닌다. 이러한 데이터 구축은 국가 정책 및 산업계·학계의 구체적인 수요에 효과적으로 대응하게 함으로써 해당 국가를 대상으로 하는 외교 및 통상 정책 수립에 실질적인 기여를 할 수 있다. 또한, 국가 간 교류를 활성화하고 상호 이해를 증진시킴으로써 미래 지향적인 경제 협력의 기반을 공고히 할 수 있을 것이다.

### 3.3. 제언

#### 1) 한국어-외국어 병렬 말뭉치의 지속 구축

전 세계적인 인공지능 기술 경쟁 속에서 자국의 인공지능 주권을 확보하기 위해서는 고품질 언어 데이터가 필수적이다. 국립국어원에서는 2021년부터 8개 언어를 대상으로 대규모 병렬 말뭉치 구축을 추진해 왔다. 기계 번역 모델이 실질적인 성능을 갖추기 위해서는 한 언어당 최소 1,000만 어절의 번역 쌍이 필요하지만 아직 이 기준에는 미치지 못한 실정이다. 앞으로 인공지능 기술 경쟁력을 더욱 강화하기 위해 후속 사업에서 병렬 말뭉치를 지속적으로 구축해야 한다.

#### 2) 병렬 말뭉치 구축 전략의 다변화

인공지능의 성능을 획기적으로 개선하기 위해서는 단순히 데이터의 수량만 늘리는 것보다 구축 전략이 다변화되는 것이 중요하다. 우선, 구축 타당성 및 난이도, 활용도 등을 바탕으로 신규 구축 언어를 선정함으로써 병렬 말뭉치의 언어 다양성을 확대하는 것이 필요하다. 또한, 구축 단위를 문장에서 문서 단위로 확장하거나 텍스트와 음성 데이터 등이 결합한 멀티모달 데이터를 구축하여 인공지능 학습 환

경의 변화에 효과적으로 대응해야 한다. 아울러 의료 및 법률, 금융 등 산업계·학계에서 수요가 높은 분야에 특화된 데이터를 집중적으로 구축한다면 전문 분야의 번역 품질을 높일 수 있는 자원으로 활용할 수 있다.

### 3) 말뭉치 사용 신청 절차의 개선

제2차 중장기계획 수립을 위한 요구 조사에서 현행 말뭉치 배포 과정이 복잡하고 번거롭다는 지적들이 있었다. 아무리 고품질의 데이터라 하더라도 접근성이 떨어지면 활용도가 떨어질 수밖에 없다. 따라서 현재보다 사용자가 배포 플랫폼에서 데이터에 쉽게 접근할 수 있도록 배포 절차를 간소화하고 사용자의 편의를 높이는 것이 필요하다. 이를 통해 말뭉치 이용률을 높이고 산업계·학계의 즉각적인 참여를 이끌어 내는 선순환을 기대할 수 있다.

### 4) 국외 기관과의 협력 관계 구축

2022년부터 2025년 사업까지 태국, 인도, 필리핀, 우즈베키스탄, 인도네시아, 사우디아라비아를 방문하여 해당 국가의 언어 자원 관련 기관, 대학, 민간 업체 등을 방문하여 한국어-외국어 병렬 말뭉치를 홍보하고 협력 가능성을 확인하였다. 각 기관과 MOU 체결, 공동 프로젝트 기획, 데이터 공유 플랫폼 연계 등 다각적이고 실질적인 협력 활동을 추진한다면 상호 호혜적인 상승 효과를 창출할 수 있다. 이를 바탕으로 말뭉치 지식 교류를 촉진하고 한국어-외국어 병렬 말뭉치를 글로벌 표준에 부합하도록 발전시키는 토대를 마련할 수 있을 것이다.

<기획·연구>

국립국어원 이현주 언어정보과장

국립국어원 박미영 학예연구관

국립국어원 강정미 연구원

<사업 참여자>

연구 기관   경희대학교 산학협력단, (주)플리토

연구 책임자   이정희

공동 연구원   이정수, 김일환, 김종민, 박진욱, 이수미, 이영준,  
                  임채훈, 강동한, 김진구, 유성곤, 김재훈, 김세화,  
                  김연희, 김영근, 문진숙, 서유리, 윤세윤, 전지연,  
                  정성호, 지화숙, 한재민, 이지희, 문다예, 김유연,  
                  이지수, 송문식, 문지연

보조 연구원   국혜민, 김인후, 민유미, 송주현, 심현보, 이두일,  
                  이태현, 이혜민, 최예린, 함인숙

---

---

발행인: 국립국어원장

발행처: 국립국어원

          서울시 강서구 금남화로 154

          전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2025년 12월 15일

발행일: 2025년 12월 15일

인    쇄: 해림복사

---

---

※ 이 보고서는 국립국어원의 용역비로 수행한 '2025년 한국어-외국어 병렬  
말뭉치 구축' 사업의 결과물을 발간한 것입니다.