



NATIONAL INSTITUTE OF KOREAN LANGUAGE



2025년 한국어 학습자 말뭉치 구축

연구 책임자 | 한송화

국립국어원 2025-01-42

발 간 등 록 번 호

11-1371028-100039-01

한국어 학습자 말뭉치 구축

연구 책임자: 한 승 화



국립국어원

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구 용역 계약에 따라 '2025년 한국어 학습자 말뭉치 구축'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2025년 6월 25일~2025년 12월 15일

2025년 12월 15일

연구 책임자: 한송화(연세대학교)

연구 기관 : 연세대학교 산학협력단, (주)이르테크

공동 연구원: 곽용진(주)이르테크), 강현화(연세대),

김선정(계명대), 김한샘(연세대), 홍혜란(배재대)

실무 연구원: 김미선(연세대), 손연정(홍익대), 서지혜(연세대)

연구 보조원: 최용기, 오수연, 이순미, 오누리, 최주연, 박정혁,
윤하정, 조현수(이상 (주)이르테크), 강수진, 구교빈,
권민수, 김가람, 김가영, 김가은, 김나영, 김연수,
김예슬, 김재희, 김진경, 김현지, 김호영, 남소현,
박유경, 박호은, 배미연, 송지혜, 유민지, 윤나영,
이레, 이소정, 이슬, 임미령, 이수빈, 이영현,
이은정, 임상은, 임은정, 전은혜, 조승희, 최예슬,
하태현(이상 연세대)

<국문 요약>

2025년 한국어 학습자 말뭉치 구축

한국어 학습자 말뭉치 구축 사업은 2010년 기초 연구를 수행한 이후, 2015-2020년 1차 중장기 계획, 2021-2025년 2차 중장기 계획을 통해 실행되어 왔다. 본 사업은 2차 중장기 계획과 함께, 지난 10여 년의 사업을 마무리하고 최종 성과를 산출하는 것을 목적으로 수행되었다. 이에 따른 사업의 성과는 다음과 같다.

○ 한국어 학습자 말뭉치 구축

본 사업에서는 25개 언어권의 문어 표본 2,838개, 355,020어절 규모의 형태주석 말뭉치와 23개 언어권의 문어 표본 5,147개, 608,668어절 규모의 오류주석 말뭉치를 구축하였다. 자료의 품질 제고를 위하여 한국어 학습자 말뭉치 구축 시스템(LCMS)을 기반으로 작업을 배분하고, 작업 진행, 작업 결과를 체계적으로 관리하였으며, 모든 작업 표본에 대하여 2단계 검수 작업을 진행하였다. 2025년 형태 주석 및 오류 주석 말뭉치를 구축한 결과, 2015-2025년 구축 말뭉치의 누적 규모는 원시 말뭉치 10,655,477어절(문어 8,808,556어절, 구어 1,846,921어절), 형태 주석 말뭉치 4,368,151어절(문어 3,115,001어절, 구어 1,253,150어절), 오류 주석 말뭉치 1,957,108어절(문어 1,283,420어절, 구어 673,688어절)이 되었다.

○ 한국어 학습자 쓰기 평가 자료 구축

한국어 학습자 쓰기 평가 자료는 쓰기 자동 채점을 비롯한 기술 개발에 활용 가능한 인공지능 학습용 데이터 구축을 위한 기반 자료이다. 본 사업에서는 주제 1(10년 후 나의 계획) 1,804건, 주제 2(내가 가 본 여행지) 1,694건, 주제 3(내가 생각하는 행복과 성공) 1,502건으로, 총 5,000건의 학습자 작문을 평가하였다. 등급 체계 및 평가 구인, 평가 기준은 자료의 호환성과 일관성을 위하여 <2023-2024년 한국어 학습자 말뭉치 구축>에서 마련한 지침을 준용하였다. 한 편의 작문에 대하여 2명의 평가자가 교차 채점을 하였으며, 등급별 점수(100점 만점), 1-6급의 누적 점수(600점 만점), 최종 점수(100점 만점)와 숙달도 등급을 부여하고 한국어교육 전문가의 총평을 제시하였다. 주제별 평가 결과에 대한 신뢰도 검정 결과, 모두 높은 신뢰도를 보이는 것으로 확인되었다.

○ 한국어 학습자 말뭉치 파일 변환

한국어 학습자 말뭉치는 매해 구축이 완료된 후 한국어 학습자 말뭉치 나눔터를 통해 원시 말뭉치, 형태 주석 말뭉치, 오류 주석 말뭉치별로 텍스트(원시), 엑셀(형태), XML(오류)을 생성해 배포해 오고 있다. 그러나 최근에는 말뭉치 파일을 JSON 포맷으로 배포하는 것이 보편화되고 있으며, 한국어 학습자 말뭉치도 JSON 포맷으로 변환하여 다른 자료와의 호환성과 활용성을 제고할 필요가 있다. 이에 따라 본 연구에서는 국립국어원의 ‘모두의 말뭉치’ 배포용 데이터 규격에 학습자 말뭉치의 메타데이터를 추가하고 오류 주석의 주석 정보 특수성을 감안해 JSON 스키마를 수립하였다. 이를 토대로 2015-2025년 구축된 전체 자료를 변환하였다.

○ 한국어 학습자 말뭉치 교육 및 홍보

한국어 학습자 말뭉치 교육 및 홍보는 한국어 학습자 말뭉치의 원활한 구축과 작업자의 전문성 제고를 통한 결과물의 품질 향상, 성과의 확산과 이용자의 활용 확대를 목적으로 한다. 본 사업에서는 학습자 말뭉치 구축 실무자를 대상으로 형태 주석 말뭉치와 오류 주석 말뭉치 구축을 위한 교육을 실시하였다. 또한 한국어 학습자 말뭉치 이용자를 대상으로 총 4회의 ‘한국어 학습자 말뭉치 교실’을 개최하였다. ‘한국어 학습자 말뭉치 교실’은 2015-2024년 개최 성과에 대한 분석을 토대로 교육 대상, 수준에 따라 차등화된 다양한 프로그램으로 구성되었으며, 참가자들의 호응과 관심 속에 성공적으로 개최되었다. 매 회의 학습자 말뭉치 교실에서는 <학습자 말뭉치 나눔터>의 통계 자료를 현행화하여 참가자들에게 홍보하고 ‘학습자 말뭉치 활용 매뉴얼’을 배포함으로써 학습자 말뭉치 이용을 활성화하였다. 그 외에도 본 사업에서는 학습자 말뭉치 활용 교육을 위한 3종의 상시 동영상 교육 자료를 개발하였다. 상시 동영상 교육 자료는 텍스트 에디터, 엑셀을 이용한 학습자 말뭉치 활용법, 한국어 학습자 말뭉치를 활용한 교수 자료 개발을 주제로 하였다.

주요어 : 한국어 학습자 말뭉치, 형태 주석 말뭉치, 오류 주석 말뭉치, 학습자 쓰기 평가, JSON 파일 생성, 한국어 학습자 말뭉치 교육

<Abstract>

2025 Project on the Construction of the Korean Language Learner Corpus

The Korean Language Learner Corpus construction project has been implemented through the basic research conducted in 2010, the first mid- to long-term plan (2015-2020), and the second mid- to long-term plan (2021-2025). As the final year of the second mid- to long-term plan, this project was carried out to conclude the past decade of work and to produce the final outcomes. The major achievements of the project are as follows.

○ Construction of the Korean Language Learner Corpus

In this project, a morph-tagged corpus consisting of 2,838 written samples from 25 language backgrounds, totaling 355,020 words, and an error-annotated corpus consisting of 5,147 written samples from 23 language backgrounds, totaling 608,668 words, were constructed. To enhance data quality, tasks were assigned through the Korean Language Learner Corpus Construction System (LCMS), and the progress and results of the work were systematically managed. A two-step inspection process was conducted for all samples. As a result of constructing the morph-tagged corpus and error-annotated corpus in 2025, the cumulative size of the corpus constructed from 2015 to 2025 reached 10,655,477 words in the raw corpus (written: 8,808,556; spoken: 1,846,921), 4,368,151 words in the morph-tagged corpus (written: 3,115,001; spoken: 1,253,150), and 1,957,108 words in the error-annotated corpus (written: 1,283,420; spoken: 673,688).

○ Construction of Korean Language Learner Writing Assessment Data

The Korean language learner writing assessment data serve as

foundational data for building artificial intelligence training datasets applicable to technology development, including automated writing scoring. In this project, a total of 5,000 learner essays were evaluated: 1,804 responses for Topic 1 (“My Plans in 10 Years”), 1,694 responses for Topic 2 (“A Place I Have Traveled”), and 1,502 responses for Topic 3 (“Happiness and Success as I See Them”). The rating scale, evaluation constructs, and scoring criteria followed the guidelines established in the <2023-2024 Korean Language Learner Corpus Construction> to ensure compatibility and consistency of the data. Each essay was evaluated by two raters using a cross-scoring method. After scoring, scores by level (out of 100), cumulative scores for Levels 1-6 (out of 600), final scores (out of 100), and proficiency levels were assigned, along with overall comments by a Korean language education expert. Reliability tests of the evaluation results by topic confirmed high levels of consistency in all cases.

○ **File Conversion of the Korean Language Learner Corpus**

The Korean Language Learner Corpus has been distributed annually upon completion through the Korean Language Learner Corpus Search Engine. The raw corpus, morph-tagged corpus, and error-annotated corpus have been provided in text (raw), Excel (morph-tagged), and XML (error-annotated) formats, respectively. However, in recent years, distributing corpus files in JSON format has become common practice, and there is a need to convert the Korean Language Learner Corpus into JSON format in order to enhance its compatibility and usability with other resources. Accordingly, in this study, the metadata of the learner corpus were added to the National Institute of Korean Language’s “Modu Corpus” distribution data specifications, and a JSON schema was established in consideration of the specific characteristics of annotation information in the error annotations. Based on this schema, all data constructed from 2015 to 2025 were converted.

○ Learner Corpus Education and Promotion

Education and promotion of the Korean Language Learner Corpus aimed to facilitate smooth corpus construction, enhance the expertise of practitioners to improve the quality of outcomes, and expand dissemination and user utilization. In this project, training was provided to construction practitioners for building the morph-tagged corpus and error-annotated corpus. In addition, four sessions of the “Korean Language Learner Corpus Class” were conducted for users. Based on an analysis of the outcomes of the 2015-2024 sessions, the program was organized into differentiated modules according to target audience and proficiency level and was successfully implemented with strong participation and interest. At each session, updated statistical data from the “Korean Language Learner Corpus Search Engine” were introduced to participants, and the “Learner Corpus Utilization Manual” was distributed to promote corpus use. Furthermore, three on-demand video training materials were developed to support learner corpus utilization education. These materials covered the use of text editors and Excel for learner corpus analysis and the development of instructional materials using the Korean Language Learner Corpus.

Keywords: Korean language learner corpus, morph-tagged corpus, error-annotated corpus, learner writing assessment, JSON file generation, Korean language learner corpus education

차 례

I. 사업 개요	1
1. 사업의 목적 및 필요성	2
1.1. 사업의 목적	2
1.2. 사업의 필요성	3
2. 사업의 범위	4
3. 사업 수행 체계	5
4. 사업 추진 일정	6
5. 사업 추진 결과	8
 II. 사업 추진 내용	 9
1. 한국어 학습자 말뭉치 구축	10
1.1. 구축 목표 설정	10
1.2. 형태 주석 및 오류 주석 말뭉치 실제 구축·가공	11
1.3. 개인정보 표지 부착 등 자료 정제	17
1.4. 자료 검수 및 품질 관리	19
1.5. 2015-2025년 한국어 학습자 말뭉치 구축 통계	22
2. 한국어 학습자 쓰기 평가 자료 구축	34
2.1. 평가 체계	34
2.2. 채점자 훈련	41
2.3. 학습자 쓰기 평가 자료 구축	43

차 례

3. 한국어 학습자 말뭉치 파일 변환	50
4. 한국어 학습자 말뭉치 교육 및 홍보	53
4.1. 말뭉치 구축 실무자 교육	53
4.2. 말뭉치 이용자를 위한 연수회 개최	54
4.3. 상시 동영상 교육 자료	90
4.4. 말뭉치 소개 · 활용 자료집 현행화 및 배포	91

Ⅲ. 결론 93

1. 사업 결과 요약	94
2. 사업의 의의 및 기대 효과	96
3. 보고서 활용 방안	97
4. 정책 제언	98

참고 자료 99

부록. 한국어 학습자 말뭉치 구축 지침

표 차례

<표 1> 사업의 범위와 세부 과업 내용	4
<표 2> 사업 추진 일정	6
<표 3> 2025년 한국어 학습자 말뭉치 구축 목표	10
<표 4> 2025년 형태 주석 문어 말뭉치 구축 결과	11
<표 5> 2025년 오류 주석 문어 말뭉치 구축 결과	14
<표 6> 개인정보 표지 부착 등 자료 정제 결과	18
<표 7> 2015-2025년 한국어 학습자 말뭉치 구축 결과	22
<표 8> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 원시 문어 말뭉치	23
<표 9> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 원시 구어 말뭉치	25
<표 10> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 형태 주석 문어 말뭉치	27
<표 11> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 형태 주석 구어 말뭉치	29
<표 12> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 오류 주석 문어 말뭉치	30
<표 13> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 오류 주석 구어 말뭉치	32
<표 14> ‘국제 통용 한국어교육 표준 모형’의 쓰기 총괄 목표 (국립국어원, 2017)	34
<표 15> ‘국제 통용 한국어교육 표준 모형’의 쓰기 등급별 목표 및 내용(국립국어원, 2017)	35
<표 16> 쓰기 평가 구인	37
<표 17> 쓰기 평가 기준	37
<표 18> 척도별 점수 구간	39

표 차례

<표 19> 점수 환산표	40
<표 20> 300개 표본의 채점자 간 신뢰도 검정 결과	42
<표 21> 학습자 쓰기 평가 자료의 국적 분포	44
<표 22> 학습자 쓰기 평가 결과	47
<표 23> 주제 1의 평가 결과에 대한 채점자 간 신뢰도 검정 결과	48
<표 24> 주제 2의 평가 결과에 대한 채점자 간 신뢰도 검정 결과	48
<표 25> 주제 3의 평가 결과에 대한 채점자 간 신뢰도 검정 결과	49
<표 26> 한국어 학습자 말뭉치 데이터 구조(JSON)	51
<표 27> 말뭉치 구축 실무자 교육 내용	53
<표 28> 한국어 학습자 말뭉치 교실 개최 프로그램	54
<표 29> 한국어 교육 경력 빈도	56
<표 30> 학습자 말뭉치의 인지 여부 빈도	57
<표 31> 원시 말뭉치의 인지 여부 빈도	58
<표 32> 형태 주석 말뭉치의 인지 여부 빈도	59
<표 33> 오류 주석 말뭉치의 인지 여부 빈도	60
<표 34> 말뭉치 수집 및 구축 절차 인지 여부 빈도	61
<표 35> 학습자 말뭉치의 활용처 인지 여부 빈도	62
<표 36> 학습자 말뭉치의 활용 방법 인지 여부 빈도	63
<표 37> 활용 도구 사용 방법 인지 여부 빈도	64
<표 38> 한국어 학습자 말뭉치 활용 경험 빈도	65
<표 39> 한국어 학습자 말뭉치 활용 목적 빈도	66
<표 40> 한국어 학습자 말뭉치 향후 활용 목적 빈도	68
<표 41> 한국어 학습자 말뭉치 교실 참가 경험 빈도	69
<표 42> 한국어 학습자 말뭉치 교실 참가 프로그램 빈도	70
<표 43> 교원을 위한 교육 프로그램 참여 의향 빈도	71

표 차례

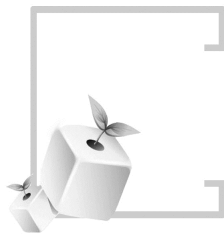
<표 44> 한국어 교수 현장에서의 활용을 위한 기술 빈도	72
<표 45> 한국어 교수 현장에서의 활용을 위한 필요 도구 빈도	74
<표 46> 교육 프로그램 수준 빈도	75
<표 47> 학습자 말뭉치 교실 교육 방식 빈도	76
<표 48> 학습자 말뭉치 교실에서 배우고 싶은 프로그램에 대한 의견 ...	76
<표 49> 학습자 말뭉치 교실 및 구축에 대한 자유 의견	77
<표 50> 회차별 응답 참여 인원 빈도	78
<표 51> 프로그램 운영 방식에 대한 만족도 빈도	79
<표 52> 강의 방식에 대한 만족도 빈도	80
<표 53> 강의 시간에 대한 만족도 빈도	81
<표 54> 강의 내용 이해도에 대한 만족도 빈도	82
<표 55> 실습 수행에 대한 만족도 빈도	83
<표 56> 한국어 학습자 말뭉치에 대한 이해도 빈도	84
<표 57> 말뭉치 활용 능력에 대한 만족도 빈도	85
<표 58> 한국어 학습자 말뭉치 교실에서 듣고 싶은 강의에 대한 의견 ·	86
<표 59> 학습자 말뭉치 교실 및 구축에 대한 자유 의견	88
<표 60> 학습자 말뭉치 상시 동영상 교육 자료 개발 개요	91

그림 차례

<그림 1> 한국어 학습자 말뭉치 구축 사업의 흐름	2
<그림 2> 사업 수행 체계: 과업 수행 체계	5
<그림 3> 2025년 형태 주석 말뭉치 수준별 분포	13
<그림 4> 2025년 형태 주석 말뭉치 언어권별 분포	14
<그림 5> 2025년 오류 주석 말뭉치 수준별 분포	16
<그림 6> 2025년 오류 주석 말뭉치 언어권별 분포	17
<그림 7> 한국어 학습자 말뭉치 작업 관리 체계	20
<그림 8> 한국어 학습자 말뭉치 2단계 검수 체제	21
<그림 9> 평가 점수 및 등급 부여 체계	40
<그림 10> 채점자 교육 절차	41
<그림 11> 채점 신뢰도 확보를 위한 쓰기 평가 절차	42
<그림 12> 학습자 말뭉치 데이터 JSON 생성 절차	50
<그림 13> 한국어 교육 경력 분포	56
<그림 14> 학습자 말뭉치의 인지 여부 분포	57
<그림 15> 원시 말뭉치의 인지 여부 분포	58
<그림 16> 형태 주석 말뭉치의 인지 여부 분포	59
<그림 17> 오류 주석 말뭉치의 인지 여부 분포	60
<그림 18> 말뭉치 수집 및 구축 절차 인지 여부 분포	61
<그림 19> 학습자 말뭉치의 활용처 인지 여부 분포	62
<그림 20> 학습자 말뭉치의 활용 방법 인지 여부 분포	63
<그림 21> 활용 도구 사용 방법 인지 여부 분포	64
<그림 22> 한국어 학습자 말뭉치 활용 경험 분포	65
<그림 23> 한국어 학습자 말뭉치 활용 목적 분포	66
<그림 24> 한국어 학습자 말뭉치 향후 활용 목적 분포	67
<그림 25> 한국어 학습자 말뭉치 교실 참가 경험 분포	69

그림 차례

<그림 26> 한국어 학습자 말뭉치 교실 참가 프로그램 분포	70
<그림 27> 교원을 위한 교육 프로그램 참여 의향 분포	71
<그림 28> 한국어 교수 현장에서의 활용을 위한 기술 분포	72
<그림 29> 한국어 교수 현장에서의 활용을 위한 필요 도구 분포	73
<그림 30> 교육 프로그램 수준 분포	75
<그림 31> 학습자 말뭉치 교실 교육 방식 분포	76
<그림 32> 회차별 응답 참여 인원 분포	78
<그림 33> 프로그램 운영 방식에 대한 만족도 분포	79
<그림 34> 강의 방식에 대한 만족도 분포	80
<그림 35> 강의 시간에 대한 만족도 분포	81
<그림 36> 강의 내용 이해도에 대한 만족도 분포	82
<그림 37> 실습 수행에 대한 만족도 분포	83
<그림 38> 한국어 학습자 말뭉치에 대한 이해도 분포	84
<그림 39> 말뭉치 활용 능력에 대한 만족도 분포	85



제 I 장

사업 개요



1. 사업의 목적 및 필요성

1.1. 사업의 목적



<그림 1> 한국어 학습자 말뭉치 구축 사업의 흐름

본 사업은 2010년 기초 연구를 시작으로 2차에 걸쳐 진행되어 온 한국어 학습자 말뭉치 중장기 계획(2015~2025년)의 최종 단계로, 지난 10여 년의 사업을 마무리하고 최종 성과를 산출하는 것을 목적으로 한다. 한국어 학습자 말뭉치 2차 중장기 계획은 인공지능 기술 개발과 함께 한국어교육 분야에서의 활용을 위한 언어 자원에 대한 수요 증가로, 국가 언어 자원으로써 대상, 수준, 언어권 등의 변인에 따라 자료를 다변화하고 자료의 규모를 1,000만 어절로 확대하는 것을 목표로 하였다. 이에 따라 2024년 1,000만 어절의 원시 말뭉치 구축을 달성하게 되었다. 본 사업에서는 한국어교육 연구자 및 교육자, 민간 기업의 개발자 등 다양한 목적의 이용자들이 한국어교육 학술 연구 및 교육, 기술 개발을 위해 이를 효율적으로 활용할 수 있는 기반을 마련하는 것을 목표로 다음의 과업을 수행하였다.

- 한국어 학습자 말뭉치(형태 주석 및 오류 주석 말뭉치) 구축
- 한국어 학습자 쓰기 평가 자료 구축
- 한국어 학습자 말뭉치 파일 변환
- 한국어 학습자 말뭉치 교육 및 홍보

1.2. 사업의 필요성

○ 공공 언어 자원으로 활용도를 갖춘 한국어 학습자 말뭉치 구축

한국어 학습자 말뭉치는 2015년 시작된 1차 중장기 계획에 따라 약 520만 어절 규모의 균형 말뭉치를 구축한 후, 2021년 수립된 2차 중장기 계획에 따라 1,000만 어절로 말뭉치의 규모가 확대되었다. 2차 중장기 계획은 인공지능 기술 개발과 함께 한국어 학습자 말뭉치의 활용 범위 확대, 이용자의 다변화에 따른 것으로, 이를 뒷받침하기 위해서는 원시 말뭉치뿐만 아니라 다양한 정보를 포함한 주석 말뭉치 구축을 확대할 필요가 있다. 본 연구의 주요 과업인 형태 주석 및 오류 주석 말뭉치 구축, 학습자 쓰기 평가 자료 구축은 그와 관련한 핵심 과제라는 점에서 의의가 있다.

○ 국가 언어 자원으로서의 품질과 신뢰성 제고

학습자 말뭉치는 비모어 화자가 산출한 비정형의 언어 자료이므로 입력과 전사, 형태 주석과 오류 주석의 전 과정에서 다양한 쟁점과 이견이 발생하며, 이로 인해 모어 화자 말뭉치보다 상세한 지침과 작업 공정, 지속적인 검증과 보완이 요구된다. 한국어 학습자 말뭉치는 지난 10여 년간 체계적인 작업 공정과 전문 인력을 통해 구축이 되었음에도, 다수의 인력이 전체 표본을 어절 단위로 직접 입력, 전사하고 주석하는 과정에서 오류가 필연적으로 남을 수밖에 없다. 본 사업에서 형태 주석 및 오류 주석 말뭉치 구축, 쓰기 평가 자료 구축과 함께 수행한 자료 검수 및 정제 작업은 국가 언어 자원으로써 한국어 학습자 말뭉치의 품질과 신뢰성 제고에 기여할 수 있다.

○ 학습자 말뭉치 기반 연구 방법론과 사용 교육에 대한 요구

2015년부터 지속되어 온 한국어 학습자 말뭉치 중장기 계획의 주요 성과 중 하나는, 한국어 학습자 말뭉치 홍보와 교육을 통해 한국어 학습자 말뭉치 이용자의 저변을 확대하고 학계에서 다수의 연구 성과를 산출한 것이라고 할 수 있다. 본 사업에서 다변화된 이용자의 요구를 수용하여 개최한 대상별, 수준별 한국어 학습자 말뭉치 교실, 상시 교육을 위한 기초 과정 동영상 자료 개발은 학습자 말뭉치의 활용도를 제고할 수 있다.

2. 사업의 범위

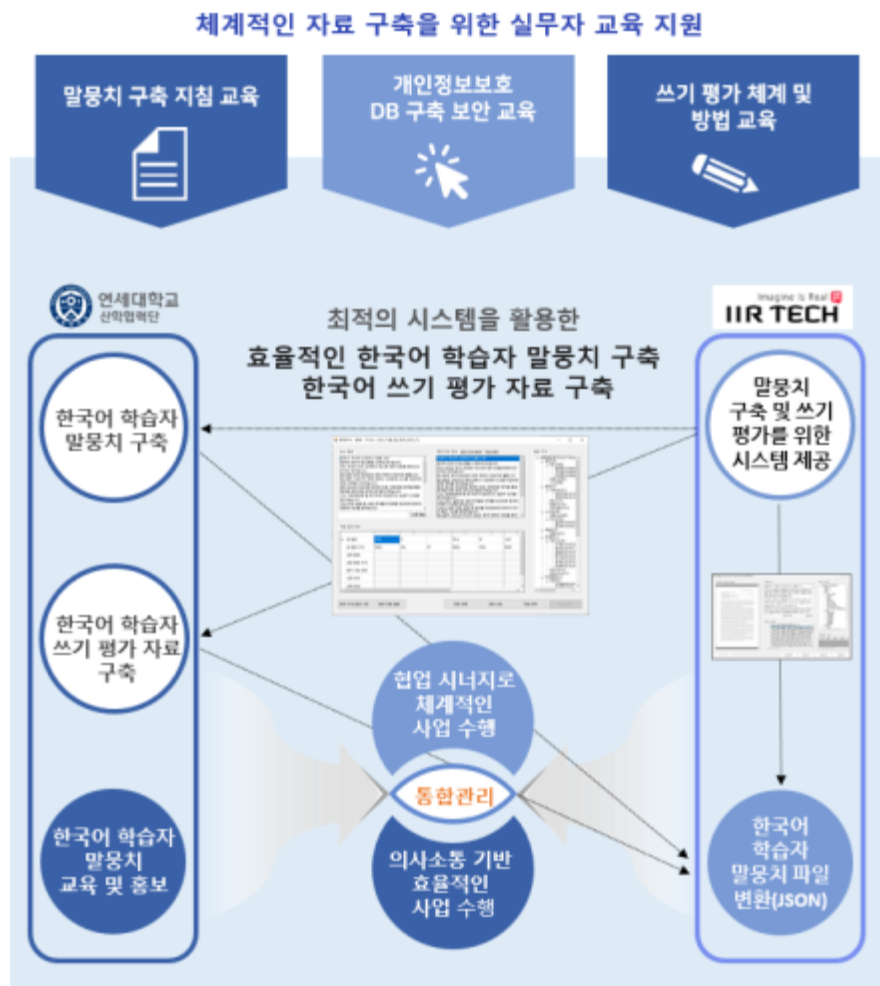
본 사업의 범위와 세부 내용은 다음과 같다.

<표 1> 사업의 범위와 세부 과업 내용

사업의 범위	세부 과업 내용
한국어 학습자 말뭉치 구축	<ul style="list-style-type: none"> ○ 형태 주석 말뭉치 350,000어절 구축 ○ 오류 주석 말뭉치 500,000어절 구축 ○ 개인정보 표지 부착 등 자료 정제 ○ 자료 검수 및 품질 관리 등
한국어 학습자 쓰기 평가 자료 구축	<ul style="list-style-type: none"> ○ 한국어교육 전문가가 평가한 쓰기 평가 자료 5,000건 이상 구축함
한국어 학습자 말뭉치 파일 변환	<ul style="list-style-type: none"> ○ 2025년까지 구축된 모든 학습자 말뭉치 파일을 JSON 파일 변환 및 생성
한국어 학습자 말뭉치 교육 및 홍보	<ul style="list-style-type: none"> ○ 한국어 학습자 말뭉치 이용자를 위한 연수회 4회 개최, 상시 동영상 3종 개발 ○ 한국어 학습자 말뭉치 구축 및 검수, 정비 인력 실무 교육 ○ 한국어 학습자 말뭉치 소개·활용 자료집 현행화 및 배포

3. 사업 수행 체계

본 사업은 연세대학교 산학협력단과 (주)이르테크에 의해 분담 이행 방식으로 수행되었다. 연세대학교 산학협력단에서는 한국어 학습자 말뭉치 구축과 한국어 학습자 쓰기 평가 자료 구축, 한국어 학습자 교육 및 홍보를 담당하였으며, (주)이르테크는 말뭉치 구축 지원 도구 제공 및 관리, JSON 형식에서의 한국어 학습자 말뭉치 자료 변환 및 생성을 담당하였다.



<그림 2> 사업 수행 체계: 과업 수행 체계

4. 사업 추진 일정

사업 추진 일정은 다음과 같다.

<표 2> 사업 추진 일정

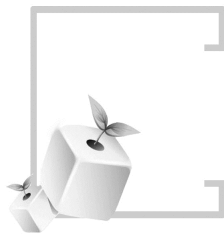
세부 과업 내용		M-2W	M	M+1	M+2	M+3	M+4	M+5
사업 계획 수립 및 착수		○						
한국어 학습자 말뭉치 구축	기구축 말뭉치 분포 분석		○					
	말뭉치 구축 설계		○					
	형태 주석 및 오류 주석 말뭉치 구축		○	○	○	○	○	
	말뭉치 구축 자료 품질 검증		○	○	○	○	○	
한국어 학습자 쓰기 평가 자료 구축	기구축 말뭉치 분포 분석	○						
	평가 표본 선정 및 작업 계획 수립		○					
	평가 작업 수행		○	○	○	○	○	○
	평가자료 품질 검증			○				○
한국어 학습자 말뭉치 파일 변환		○	○	○	○	○	○	○
한국어 학습자 말뭉치 교육 및 홍보	교육 및 홍보 계획 수립	○						
	말뭉치 구축 실무자 교육(1차)		○					

세부 과업 내용		M-2W	M	M+1	M+2	M+3	M+4	M+5
	말뭉치 구축 실무자 교육(2차)			○				
	말뭉치 구축 실무자 교육(3차)					○		
	한국어 학습자 말뭉치 교실(1차)			○				
	한국어 학습자 말뭉치 교실(2차)				○			
	한국어 학습자 말뭉치 교실(3차)					○		
	한국어 학습자 말뭉치 교실(4차)						○	
	상시 동영상 자료 기획			○				
	상시 동영상 자료 제작				○	○	○	
	한국어 학습자 말뭉치 소개·활용 자료집 현행화 및 배포		○	○	○	○	○	○
최종보고회의 및 사업 마무리								○

5. 사업 추진 결과

사업 추진 결과는 다음과 같다.

- 한국어 학습자 말뭉치 구축
 - 한국어 학습자 형태 주석 말뭉치 355,020어절(2,838개 표본) 구축
 - 한국어 학습자 오류 주석 608,668어절(5,147개 표본) 구축
- 한국어 학습자 쓰기 평가 자료 5,000건 구축
 - 주제 1(10년 후 나의 계획) 1,804건
 - 주제 2(내가 가 본 여행지) 1,694건
 - 주제 3(내가 생각하는 행복과 성공) 1,502건
- 한국어 학습자 말뭉치 JSON 파일 변환 및 생성
- 한국어 학습자 말뭉치 교실 4회 개최
- 상시 동영상 교육 자료 3종 개발
 - 한국어 학습자 말뭉치 활용을 위한 텍스트 에디터 활용 지침
 - 한국어 학습자 말뭉치 활용을 위한 엑셀 활용 지침
 - 한국어 학습자 말뭉치를 교육 자료에 활용하기
- 한국어 학습자 말뭉치 소개·활용 자료집 현행화 및 배포



제 II 장

사업 추진 내용



1. 한국어 학습자 말뭉치 구축

1.1. 구축 목표 설정

한국어 학습자 말뭉치는 2015년부터 구축을 시작하여 2024년까지 원시 말뭉치 약 1,050만 어절, 형태 주석 말뭉치 약 400만 어절, 오류 주석 말뭉치 약 135만 어절을 구축하였다. 한국어 학습자 말뭉치는 한국어교육 연구를 위한 자료로 유용하게 활용되고 있으나 원시 말뭉치 대비 형태 주석 말뭉치와 오류 주석 말뭉치의 비중이 현저하게 적은 편이다. 따라서 말뭉치의 균형성과 교육, 산업 분야로의 활용 범위를 확대하기 위하여 형태 주석 말뭉치와 오류 주석 말뭉치의 구축 규모를 확대할 필요가 있다. 본 사업에서는 기구축 말뭉치의 말뭉치 유형별, 대상별, 수준별, 언어권별(국적별) 분포 현황을 토대로, 다음과 같이 형태 주석 35만 어절, 오류 주석 말뭉치 50만 어절을 구축 목표로 설정하였다. 2025년 형태 주석과 오류 주석 대상 표본은 무엇보다도 자료의 활용도 측면에서 다양성 확보가 중요하다는 판단하에 구축 비중이 높은 언어권을 배제하고 소수 언어권의 자료를 우선 선정하였다. 구축 비중이 높은 언어권의 자료도 수준별 균형을 맞추어 필요가 있는 경우는 추가 구축 대상으로 선정하였다.

<표 3> 2025년 한국어 학습자 말뭉치 구축 목표

구분	문어			구어			합계		
	기구축	2025년	누적 합계	기구축	2025년	누적 합계	기구축	2025년	누적 합계
원시	870만	-	870만	180만	-	180만	1,050만	-	1,050만
형태 주석	275만	35만	310만	125만	-	125만	400만	35만	435만
오류 주석	67만	50만	117만	67만	-	67만	134만	50만	184만

1.2. 형태 주석 및 오류 주석 말뭉치 실제 구축·가공

1.2.1. 형태 주석 말뭉치

2025년 형태 주석 말뭉치는 25개 언어권의 문어 표본 2,838개, 355,020어절을 구축 대상으로 선정하여 작업을 진행하였다.

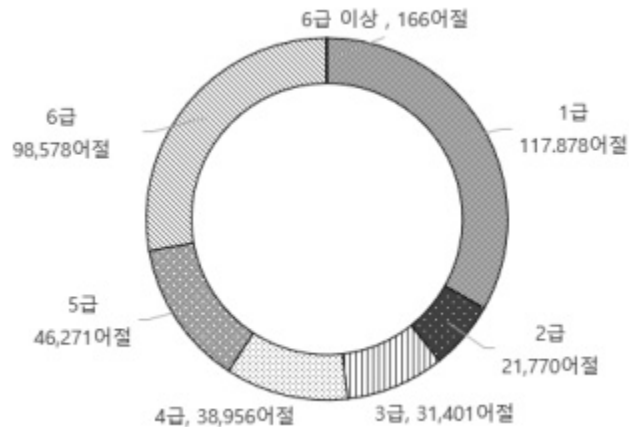
<표 4> 2025년 형태 주석 문어 말뭉치 구축 결과

모국어	구분	1급	2급	3급	4급	5급	6급	정보 없음	합계
일본어	어절 수	24,314					28,271		52,585
	표본 수	225					185		410
우즈베크 어	어절 수	23,459		104	2,729	18,023	2,022		46,337
	표본 수	193		1	22	131	13		360
베트남어	어절 수		122	676	218		38,632		39,648
	표본 수		1	6	2		267		276
몽골어	어절 수	10,944		266		7,073	14,489		32,772
	표본 수	118		2		58	125		303
인도네시아어	어절 수	11,652	7,059	3,168	5,923	778	1,562		30,142
	표본 수	83	61	26	41	6	8		225
러시아어	어절 수	13,525			6,756	5,543	1,770		27,594
	표본 수	111			49	44	11		215
광둥어	어절 수	408	1,165	4,246	7,376	3,451	2,669		19,315
	표본 수	5	11	34	54	26	16		146
프랑스어	어절 수	2,588	4,761	3,916	3,995	495	891		16,646
	표본 수	29	42	36	31	4	4		146
키르기스 어	어절 수	1,560	2,483	2,379	1,574	4,142	1,474		13,612
	표본 수	16	23	22	12	31	8		112
버마어	어절 수	668		3,196	3,882	1,151	3,367		12,264
	표본 수	7		26	29	9	21		92
아랍어	어절 수	3,722	1,432	5,338	1,050	251		166	11,959

모국어	구분	1급	2급	3급	4급	5급	6급	정보 없음	합계
	표본 수	40	14	46	7	2		1	110
스페인어	어절 수	2,178	1,479	4,080	618	1,305	849		10,509
	표본 수	23	14	34	5	10	5		91
타이어	어절 수	8,004				825	1,667		10,496
	표본 수	67				6	12		85
영어	어절 수	4,388				2,843	605		7,836
	표본 수	41				21	4		66
카자흐어	어절 수	1,657	2,298	1,747	1,811				7,513
	표본 수	13	17	15	14				59
라오어	어절 수	3,754		223	386	124			4,487
	표본 수	30		2	3	1			36
말레이어	어절 수	911	468	2,062	605				4,046
	표본 수	9	5	16	5				35
이탈리아 어	어절 수	2,148							2,148
	표본 수	23							23
싱할라어	어절 수	82			1,785	267			2,134
	표본 수	1			13	2			16
우르두어	어절 수	470	225		248		110		1,053
	표본 수	5	2		2		1		10
독일어	어절 수	545							545
	표본 수	5							5
중국어	어절 수	385	143						528
	표본 수	8	2						10
타갈로그 어	어절 수	516							516
	표본 수	5							5
벵골어	어절 수						200		200
	표본 수						1		1
포르투갈 어	어절 수		135						135
	표본 수		1						1

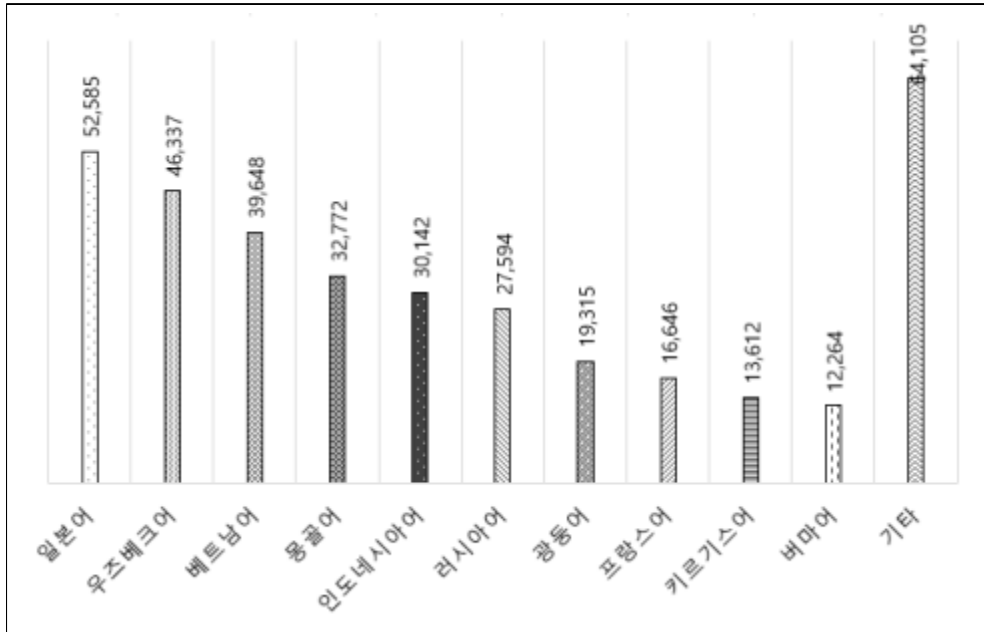
모국어	구분	1급	2급	3급	4급	5급	6급	정보 없음	합계
합계	어절 수	117,878	21,770	31,401	38,956	46,271	98,578	166	355,020
	표본 수	1,057	193	266	289	351	681	1	2,838

수준별로는 1급에서 6급 자료가 중점적으로 구축되었으며, 그중 1급이 117,878어절(33.2%)로 가장 비중이 높았으며, 6급이 98,578어절(27.8%)로 그 다음으로 높은 비중을 차지하였다. 그 외에 5급 46,271어절(13.0%), 4급 38,956어절(11.0%), 3급이 31,401어절(8.8%), 2급이 21,770어절(6.0%)를 차지하였다. 그리고 6급 이상이 166어절로 0.05%를 차지하였다.



<그림 3> 2025년 형태 주석 말뭉치 수준별 분포

언어권별로는 일본어권이 52,585어절(14.8%)로 가장 높은 비중을 차지하였으며, 이어서 우즈베크어 46,337어절(13.1%), 베트남어 39,648어절(11.2%), 몽골어 32,772어절(9.2%), 인도네시아어 30,142어절(8.5%), 러시아어 27,594어절(7.8%)로 상대적으로 높은 비중을 차지하였다. 그 외에 광둥어 19,315어절(5.4%), 프랑스어 16,646어절(4.7%), 키르기스어 13,612어절(3.8%), 버마어 12,264어절(3.5%)를 차지하였으며, 기타 언어권의 자료가 64,105어절(18.1%)를 차지하였다.



<그림 4> 2025년 형태 주석 말뭉치 언어권별 분포

1.2.2. 오류 주석 말뭉치 구축

2025년 오류 주석 말뭉치는 23개 언어권의 문어 표본 5,147개, 608,668어절을 구축 대상으로 선정하여 작업을 진행하였다.

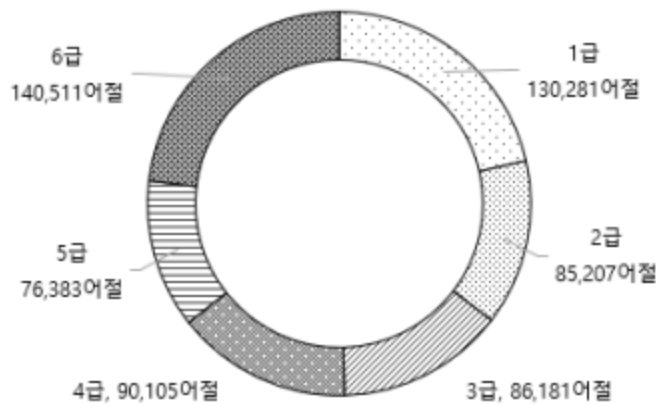
<표 5> 2025년 오류 주석 문어 말뭉치 구축 결과

모국어	구분	1급	2급	3급	4급	5급	6급	합계
러시아어	어절 수	13,427	14,349	15,324	20,168	14,339	9,843	87,450
	표본 수	112	174	147	168	119	63	783
몽골어	어절 수	13,237	11,213	10,904	10,537	11,301	20,008	77,200
	표본 수	143	122	101	82	95	164	707
베트남어	어절 수	4,507	6,540	4,714	7,072		38,668	61,501
	표본 수	47	55	37	54		267	460
일본어	어절 수	23,898					28,271	52,169
	표본 수	222					185	407

모국어	구분	1급	2급	3급	4급	5급	6급	합계
우즈베크어	어절 수	23,459		104	2,729	17,615	2,585	46,492
	표본 수	193		1	22	128	17	361
스페인어	어절 수	2,223	11,942	12,575	7,312	5,506	2,684	42,242
	표본 수	24	125	114	61	44	17	385
타이어	어절 수	8,004	9,357	5,566	5,903	3,485	6,319	38,634
	표본 수	67	105	52	53	28	40	345
인도네시아어	어절 수	9,654	7,059	4,946	5,923	2,591	6,609	36,782
	표본 수	70	61	43	41	21	41	277
프랑스어	어절 수	2,588	11,342	9,372	7,114	2,577	3,707	36,700
	표본 수	29	108	84	57	20	24	322
광둥어	어절 수	408	1,165	4,246	7,376	3,540	11,896	28,631
	표본 수	5	11	34	54	27	89	220
아랍어	어절 수	6,098	3,632	6,621	3,147	1,064	677	21,239
	표본 수	71	39	58	25	9	6	208
카자흐어	어절 수	1,657	5,299	4,042	4,445	2,346	1,855	19,644
	표본 수	13	46	36	36	16	9	156
키르기스어	어절 수	1,560	2,357	2,379	1,574	6,179	1,975	16,024
	표본 수	16	22	22	12	46	12	130
버마어	어절 수	668		3,196	3,493	2,873	4,046	14,276
	표본 수	7		26	26	21	25	105
영어	어절 수	8,211	106			2,843	605	11,765
	표본 수	90	1			21	4	116
말레이어	어절 수	911	468	1,893	605		529	4,406
	표본 수	9	5	15	5		4	38
라오어	어절 수	3,389		223	386	124		4,122
	표본 수	27		2	3	1		33
중국어	어절 수	2,621	153	76	288		124	3,262
	표본 수	31	2	1	2		1	37
이탈리아어	어절 수	2,148						2,148

모국어	구분	1급	2급	3급	4급	5급	6급	합계
상할라어	표본 수	23						23
	어절 수	82			1,785			1,867
	표본 수	1			13			14
우르두어	어절 수	470	225		248		110	1,053
	표본 수	5	2		2		1	10
독일어	어절 수	545						545
	표본 수	5						5
타갈로그어	어절 수	516						516
	표본 수	5						5
합계	어절 수	130,281	85,207	86,181	90,105	76,383	140,511	608,668
	표본 수	1,215	878	773	716	596	969	5,147

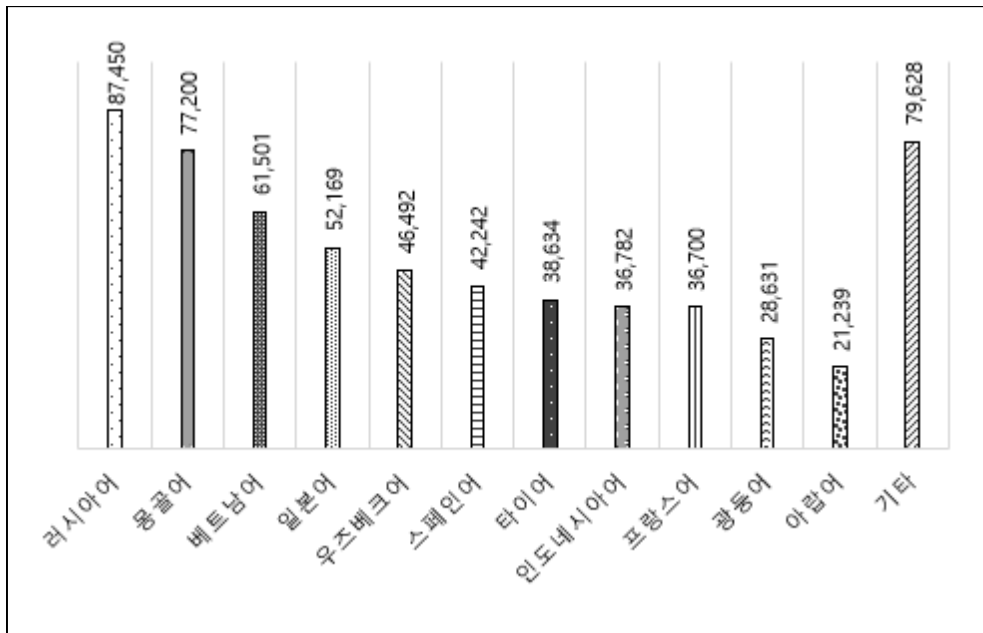
수준별로는 6급이 140,511어절(23.1%)로 가장 비중이 높았으며, 근소한 차이이지만 1급이 130,281어절(21.4%)로 그다음으로 높은 비중을 차지하였다. 이어서 4급 90,105어절(14.8%), 3급 86,181어절(14.2%), 2급이 85,207어절(14.0%), 5급이 76,383어절(12.5%)로 그 뒤를 이었다.



<그림 5> 2025년 오류 주석 말뭉치 수준별 분포

언어권별로는 러시아어권이 87,450어절(14.4%)로 가장 높은 비중을 차지하였으며, 이어서 몽골어 77,200어절(12.7%), 베트남어 61,501어절(10.1%), 일본

어 52,169어절(8.6%), 우즈베크어 46,492어절(7.6%), 스페인어 42,242어절(6.9%), 타이어가 38,634어절(6.3%)로 상대적으로 높은 비중을 차지하였다. 그 외에 인도네시아어 36,782어절(6.0%), 프랑스어가 36,700어절(6.0%)을 차지하였으며, 기타 언어권의 자료가 79,628어절(13.1%)를 차지하였다.



<그림 6> 2025년 오류 주석 말뭉치 언어권별 분포

1.3. 개인정보 표지 부착 등 자료 정제

본 사업에서는 2015-2024년의 자료 처리 및 입력 지침에 따라 구축된 원시 말뭉치를 기반으로 형태 주석 및 오류 주석 작업을 수행하였다. 원시 말뭉치에는 학습자의 국적, 모국어, 연령 등을 포함한 변인 정보와 자료 수집을 위한 과제 유형, 주제 등의 메타 정보들이 포함되어 있다. 형태 주석 및 오류 주석 정보에도 이들 정보들이 포함되도록 하였으며, 학습자의 이름, 외국인 등록번호, 카드 번호, 전화 번호 등의 개인정보는 지침에 따라 비식별화 처리되도록 하였다.

아울러 2025년 형태 주석 및 오류 주석 표본, 학습자 쓰기 평가 자료를 대상

으로 메타 정보를 중심으로 한 자료의 무결성에 대한 검수 작업을 함께 진행하여 수정 작업을 병행하였다. 검수 과정에서 수정 처리된 표본 수는 250개였으며, 등급 정보 오류, 주제 정보 오류, 유사 내용으로 중복 표본 의심, 학습자 이름 노출, 원문 입력 오류, 줄바꿈 오류의 문제로 문제 표본은 (주)이르테크와 협업을 통해 수정되었다.

<표 6> 개인정보 표지 부착 등 자료 정제 결과

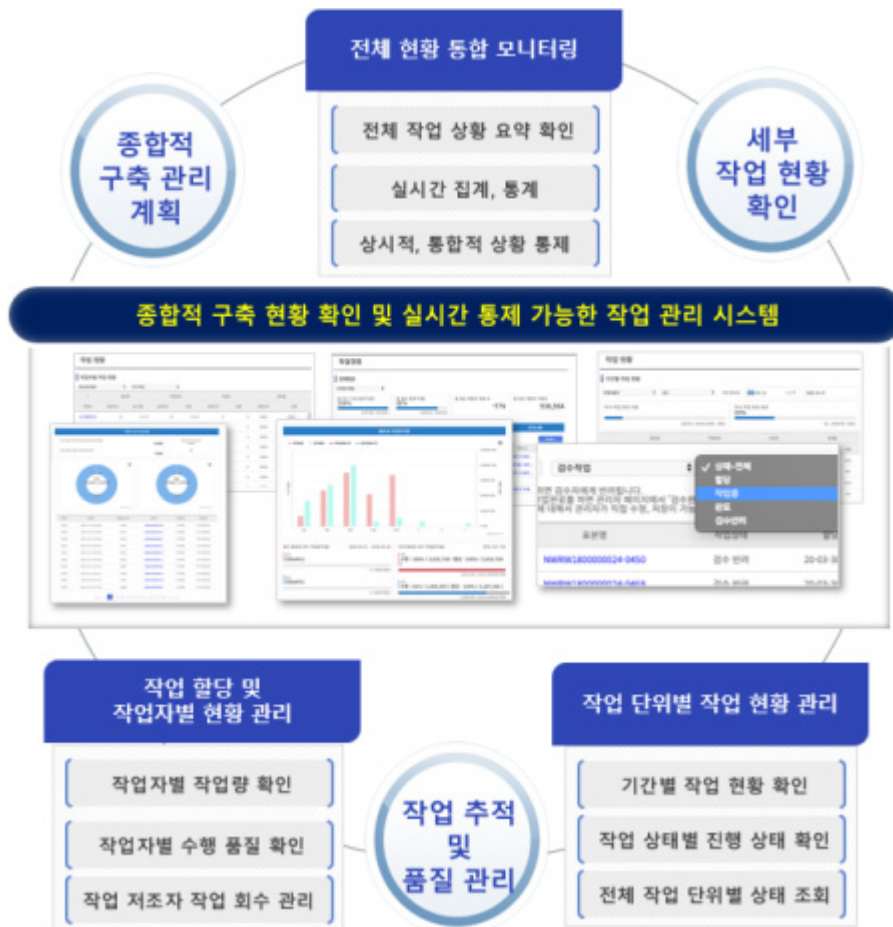
처리 유형	표본 번호	표본 수
등급 정보 수정	7273, 47711, 47735, 47841, 48394, 49120	6
주제 정보 수정	49079, 49118, 56633, 56651, 56662, 56677, 56695, 56712, 56798, 56808, 56816, 56843, 56853, 56862, 56870, 56879, 56888, 56910, 56919, 56929, 56937, 56943, 56971, 56972, 56978, 56979, 56980, 56986, 56988, 56995, 56997, 57004, 57005, 57009, 57010, 57014, 57015, 57020, 57025, 57029, 57030, 62979, 64717, 68093, 70224, 70859, 73022, 73026, 73031, 73035, 73039, 73043, 73047, 73051, 73058, 73062, 73066, 73076, 73078, 73082, 73086, 73091, 73094, 73096, 73098, 73100, 73102, 73106, 73110, 73114, 73118, 73122, 73126, 73130, 73134, 73138, 73142, 73146, 73150, 73154, 73157, 73160, 73163, 73166, 73168, 73170, 73172, 73174, 73178, 73182, 73186, 73190, 73194, 73198, 73202, 73206, 73210, 73214, 73218, 73224, 73227, 73230, 73233, 73236, 73238, 73240, 73242, 73244, 73246, 73250, 73252, 73255, 73257, 73261, 74892, 77460, 78749, 79195, 79196	119
유사 내용으로 중복 표본 여부 확인	19443, 19444, 43459, 43732, 53369, 53400, 53437, 53589, 54359, 54369, 54394, 62336, 62346, 62419, 62670, 62693, 62724, 62744, 62764, 62834, 62849, 62854, 62879, 62894, 62939, 62949, 62995, 63033, 63043, 63053, 63069, 63254, 63290, 63341, 63377, 63819, 64566, 64587, 64671, 64672, 65162, 65415, 65420, 68910, 68953, 68968, 69049, 69059, 69262, 69493	50

학습자 이름 비식별화 처리	10626, 14375, 43877, 43931, 43935, 44564, 47975, 47985, 48079, 48083, 48105, 48130, 48255, 48690, 48724, 50052, 50077, 50322, 50373, 50402, 50408, 50466, 52051, 52053, 55622, 57549, 57559, 57563, 58199, 58200, 58394, 60178, 64498, 64503, 65121, 65327, 66607, 66609, 66617, 67593, 67628, 67629, 67688, 67940, 67942, 73029, 73249, 73251, 74959, 74963, 74980, 75198, 78635, 79437, 81480	55
원문 입력 오류 수정	45197, 47072, 47188, 49368, 58768, 59064, 59125, 60101, 68174, 69521, 75125, 76077, 79056, 79061	14
줄바꿈 오류 수정	52885, 52902, 57488, 57506, 66620, 79111	6

1.4. 자료 검수 및 품질 관리

○ 한국어 학습자 말뭉치 구축 시스템(LCMS)를 기반으로 한 종합 구축 관리

본 사업에서는 자료의 품질 제고를 위하여 한국어 학습자 말뭉치 구축 시스템(LCMS)를 기반으로 종합적인 구축 관리 계획을 수립하고 작업자들에게 적정 분량의 작업 표본을 주기적으로 배분하고 관리하였다. 특히, 작업 진행 현황 및 작업 통계, 작업 이력 정보 등을 기반으로 작업 현황을 철저히 관리하여 일정 기간 내에 일정 분량의 작업을 순차적으로 진행하도록 하였으며, 상위 작업자의 작업 품질 검사를 통해 구축 자료의 질을 높이고자 하였다.



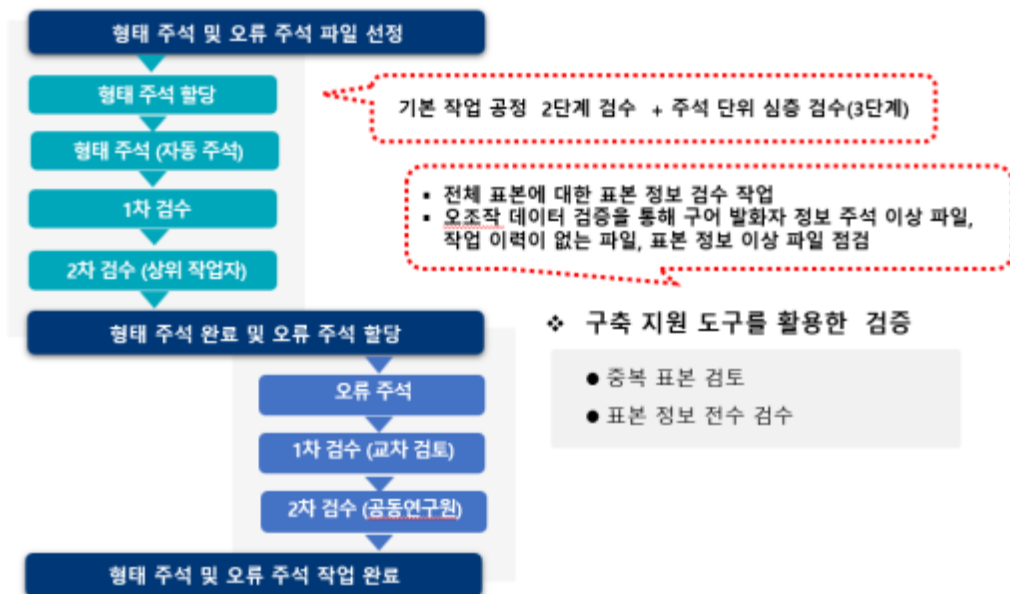
<그림 7> 한국어 학습자 말뭉치 작업 관리 체계

○ 2단계 전수 검수 체제를 통한 학습자 말뭉치의 질적 제고

본 사업에서는 형태 주석 및 오류 주석 대상이 된 전체 표본을 대상으로 2단계 검수 작업을 진행하여 작업의 질을 높이고자 하였다.

형태 주석 1차 검수는 학습자 말뭉치 구축 시스템(LCMS)에 탑재된 주석 도구를 기반으로 자동 주석한 자료를 연구 인력이 전수 검토하는 방식으로 진행하였으며, 형태 주석 2차 검수는 1차 검수가 완료된 후 자료의 무결성을 위해 상위 작업자에 의해 진행. 2차 검수 또한 전수 검토 방식으로 진행하였다. 오류 주석은 형태 주석과 달리 작업자에 의해 수동 주석을 실시한 후 2단계의 검수가 진행되었다. 1차 검수는 작업자 간의 교차 검토 형식을 통해 주석의 일관성을 제고하고 오류를 수정하는 데에 초점을 맞추었으며, 2차 검수는 1차 검수

가 완료된 후 상위 작업자가 검수를 진행하여 자료의 무결성을 확보하는 데 초점을 두었다. 2단계 검수가 완료된 후에는 최상위 검토자(팀별 총괄 공동연구원)가 주석 단위로 자료를 임의 추출하여 심층 검수를 실시하였다.



<그림 8> 한국어 학습자 말뭉치 2단계 검수 체제

1.4.2. 2025년 이전 기구축 표본

본 사업에서는 2025년 구축 대상 표본 외에도 2015-2024년 기구축 한국어 학습자 말뭉치의 표본 중 1,500건의 표본을 샘플로 추출하여 학습자가 자료를 산출하는 과정에서 번역기나 인공지능(AI) 사용이 의심되는 표본들을 확인하기 위해 검토 작업을 진행하였다. 이는 자료의 품질 검증 작업의 일환으로, 작업을 통해 자료의 품질 향상, 향후의 수집 지침 정교화를 위한 기초 자료로 활용될 수 있다.

1.5. 2015-2025년 한국어 학습자 말뭉치 구축 통계

본 사업에서는 25개 언어권의 문어 표본 2,838개, 355,020어절을 대상으로 형태 주석 작업을 진행하였고, 23개 언어권의 문어 표본 5,147개, 608,668어절을 대상으로 오류 주석 작업을 진행하였다. 그 결과 2015-2025년 구축 말뭉치의 누적 규모는 원시 말뭉치 10,655,477어절(문어 8,808,556어절, 구어 1,846,921어절), 형태 주석 말뭉치 4,368,151어절(문어 3,115,001어절, 구어 1,253,150어절), 오류 주석 말뭉치 1,957,108어절(문어 1,283,420어절, 구어 673,688어절)이 되었으며, 수준별 분포는 다음과 같다.

<표 7> 2015-2025년 한국어 학습자 말뭉치 구축 결과

말뭉치 유형	자료 유형	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
원시	문어	어절 수	1,368,376	2,372,841	1,757,938	1,358,933	1,129,167	640,015	135,555	45,731	8,808,556
		표본 수	15,287	21,061	13,833	9,924	7,432	3,918	119	258	71,832
	구어	어절 수	287,561	330,971	450,558	322,811	210,923	132,886	32,859	78,352	1,846,921
		표본 수	888	726	768	663	351	198	23	133	3,750
	합계	어절 수	1,655,937	2,703,812	2,208,496	1,681,744	1,340,090	772,901	168,414	124,083	10,655,477
		표본 수	16,175	21,787	14,601	10,587	7,783	4,116	142	391	75,582
형태 주석	문어	어절 수	498,443	502,317	531,041	501,053	503,449	506,696	63,120	8,882	3,115,001
		표본 수	6,621	5,083	4,452	3,898	3,400	3,270	43	45	26,812
	구어	어절 수	233,994	258,567	260,772	231,661	134,329	103,746	21,112	8,969	1,253,150
		표본 수	750	553	567	487	237	160	15	19	2,788
	합계	어절 수	732,437	760,884	791,813	732,714	637,778	610,442	84,232	17,851	4,368,151
		표본 수	7,371	5,636	5,019	4,385	3,637	3,430	58	64	29,600
오류 주석	문어	어절 수	229,078	198,609	201,707	197,655	196,182	256,416	0	3,773	1,283,420
		표본 수	2,598	2,065	1,821	1,643	1,472	1,771	0	21	11,391
	구어	어절 수	136,815	122,948	122,835	128,930	88,908	66,184	4,892	2,176	673,688
		표본 수	443	248	300	271	150	89	5	7	1,513
	합계	어절 수	365,893	321,557	324,542	326,585	285,090	322,600	4,892	5,949	1,957,108

		표본 수	3,041	2,313	2,121	1,914	1,622	1,860	5	28	12,904
--	--	------	-------	-------	-------	-------	-------	-------	---	----	--------

1.5.1. 원시 말뭉치

1) 문어

원시 문어 말뭉치는 106개 언어권의 71,832개 표본, 8,808,556어절이 구축되었다. 수준별로는 1급 1,368,376어절, 2급 2,372,841어절, 3급 1,757,938어절, 4급 1,358,933어절, 5급 1,129,167어절, 6급 640,015어절로 2급이 가장 많았으며, 6급이 가장 적었다. 언어권별로는 중국어가 2,227,321어절로 가장 많았으며, 이어서 베트남어 1,906,316어절, 일본어 978,255어절, 벵골어 815,847어절로 높은 비중을 차지하였다.

<표 8> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 원시 문어 말뭉치

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
중국어	어절 수	303,233	321,180	390,259	374,444	440,653	279,301	102,924	15,327	2,227,321
	표본 수	4,298	3,205	3,179	2,797	2,819	1,700	77	95	18,170
베트남어	어절 수	428,982	439,643	420,679	354,717	188,938	62,637	6,107	4,613	1,906,316
	표본 수	4,400	3,909	3,307	2,600	1,333	403	4	35	15,991
일본어	어절 수	76,172	223,290	214,319	199,068	156,506	105,568	1,188	2,144	978,255
	표본 수	939	2,103	1,766	1,476	981	668	7	16	7,956
벵골어	어절 수	2,656	809,041	2,148	1,636	166	200			815,847
	표본 수	39	6,899	22	15	2	1			6,978
타이어	어절 수	113,974	109,974	111,862	90,427	13,314	9,756	212	794	450,313
	표본 수	777	653	638	511	87	59	1	7	2,733
몽골어	어절 수	131,188	29,768	121,304	32,214	59,179	22,474	267	1,072	397,466
	표본 수	1,235	289	1,041	244	436	176	2	6	3,429
영어	어절 수	45,042	67,373	65,677	56,124	41,450	33,390	2,856	857	312,769
	표본 수	641	684	565	443	273	181	4	6	2,797
인도네시아어	어절 수	18,664	64,574	114,262	33,794	14,452	13,190	1,097	678	260,711

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
	표본 수	186	484	757	218	83	72	1	5	1,806
광둥어	어절 수	17,362	40,293	54,235	55,422	47,750	40,600			255,662
	표본 수	262	390	434	408	305	250			2,049
러시아어	어절 수	30,180	44,827	44,458	37,643	48,025	13,565	4,539	2,161	225,398
	표본 수	372	458	391	295	325	82	6	16	1,945
우즈베크어	어절 수	67,554	47,582	30,354	13,238	25,516	3,350		1,937	189,531
	표본 수	542	376	232	95	176	19		14	1,454
스페인어	어절 수	15,540	21,699	25,422	12,554	9,871	3,001		434	88,521
	표본 수	203	217	199	89	68	19		2	797
프랑스어	어절 수	13,948	14,405	14,287	9,485	7,703	4,842	933	149	65,752
	표본 수	182	134	117	73	50	30	1	1	588
우르두어	어절 수	32,053	21,861	6,395	483	428	110			61,330
	표본 수	267	178	56	4	2	1			508
라오어	어절 수	5,873	23,725	9,509	452	810	348	2,158		42,875
	표본 수	48	186	75	4	6	2	2		323
기타 ¹⁾	어절 수	65,955	93,606	132,768	87,232	74,406	47,683	13,274	15,565	530,489
	표본 수	896	896	1,054	652	486	255	14	55	4,308
합계	어절 수	1,368,376	2,372,841	1,757,938	1,358,933	1,129,167	640,015	135,555	45,731	8,808,556
	표본 수	15,287	21,061	13,833	9,924	7,432	3,918	119	258	71,832

1) 기타 언어권에는 버마어, 아랍어, 말레이어, 이탈리아어, 카자흐어, 한국어, 스웨덴어, 키르기스어, 독일어, 싱할라어, 타갈로그어, 포르투갈어, 자바어, 터키어, 크메르어, 힌디어, 폴란드어, 슬로베니아어, 페르시아어, 네팔어, 슬로바키아어, 노르웨이어, 투르크멘어, 헝가리어, 네덜란드어, 우크라이나어, 아제르바이잔어, 타지크어, 타밀어, 루마니아어, 체코어, 스와힐리어, 불가리아어, 핀란드어, 암하라어, 위구르어, 중국어(만다린어), 카탈루냐어, 르완다어, 아르메니아어, 테툼어, 세르비아어, 덴마크어, 쿠르드어, 크로아티아어, 간다어, 편자브어, 이그보어, 마라티어, 히브리어, 소말리어, 순다어, 조지아어, 트위어, 룩셈부르크어, 벨라루스어, 말라얄람어, 칸나다어, 크리올어, 에스토니아어, 알바니아어, 노르웨이어 (뉘노르스크), 니안콜어, 마다가스카르어, 보스니아어, 리투아니아어, 츠와나어, 요루바어, 그리스어, 웨일스어, 쇼나어, 하우스어, 티베트어, 벤다어, 텔루구어, 덩카어, 티그리냐어, 몰도바어, 바슈키르어, 마오리어, 아삼어, 구자라트어, 피지어, 월로프어, 룬디어, 라트비아어, 세부아노어, 파슈토어, 아이슬란드어, 판테어(Fanti), 아프리카נס어가 있다.

2) 구어

원시 구어 말뭉치는 51개 언어권의 3,750개 표본, 1,846,921어절이 구축되었다. 수준별로는 1급 287,561어절, 2급 330,971어절, 3급 450,558어절, 4급 322,811어절, 5급 210,923어절, 6급 132,886어절로 3급이 가장 많았으며, 6급이 가장 적었다. 언어권별로는 중국어가 390,306어절로 가장 많았으며, 이어서 베트남어 344,055어절, 타이어 296,958어절, 일본어 218,793어절로 높은 비중을 차지하였다.

<표 9> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 원시 구어 말뭉치

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
중국어	어절 수	51,834	68,929	69,867	66,950	43,765	54,345	21,112	13,504	390,306
	표본 수	245	207	159	153	76	71	15	33	959
베트남어	어절 수	49,949	64,064	72,859	76,341	66,309	10,502	777	3,254	344,055
	표본 수	119	133	135	135	108	17	2	7	656
타이어	어절 수	64,535	38,930	150,393	13,789	5,750	4,376		19,185	296,958
	표본 수	155	60	158	25	8	8		33	447
일본어	어절 수	14,844	30,422	34,578	52,863	42,303	37,927	1,701	4,155	218,793
	표본 수	60	57	77	118	68	53	2	6	441
스페인어	어절 수	29,334	23,538	23,346	18,790	4,225	1,988	594	375	102,190
	표본 수	64	43	38	39	8	3	1	1	197
러시아어	어절 수	12,560	17,779	21,620	18,426	5,330	4,419		3,973	84,107
	표본 수	42	49	55	44	14	12		7	223
영어	어절 수	10,644	18,766	15,084	7,002	10,860	2,282		864	65,502
	표본 수	32	28	34	17	13	3		1	128
인도네시아어	어절 수	11,302	13,237	15,122	7,881	8,307	4,365		851	61,065
	표본 수	38	26	29	17	18	11		1	140
타갈로그어	어절 수	11,948	11,615	11,635	6,632	1,682	1,711			45,223
	표본 수	31	29	13	12	2	3			90
한국어	어절 수				534			6,799	23,912	31,245

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
	표본 수				1			1	30	32
싱할라어	어절 수	6,204	5,306	6,539	6,133	3,726	2,990			30,898
	표본 수	10	11	10	11	6	5			53
몽골어	어절 수	2,046	2,102	4,370	7,229	4,062	2,343		1,966	24,118
	표본 수	10	9	12	16	8	2		3	60
버마어	어절 수	2,633	6,469	6,653	6,297	751				22,803
	표본 수	5	7	6	8	1				27
키르기스어	어절 수	1,469	4,525	4,334	4,616	3,327			2,084	20,355
	표본 수	5	9	6	9	6			4	39
우즈베크어	어절 수	1,946	7,706	2,637	4,075	666				17,030
	표본 수	6	14	5	8	1				34
기타 ²⁾	어절 수	16,313	17,583	11,521	25,253	9,860	5,638	1,876	4,229	92,273
	표본 수	66	44	31	50	14	10	2	7	224
합계	어절 수	287,561	330,971	450,558	322,811	210,923	132,886	32,859	78,352	1,846,921
	표본 수	888	726	768	663	351	198	23	133	3,750

1.5.2. 형태 주석 말뭉치

1) 문어

형태 주석 문어 말뭉치는 88개 언어권의 26,812개 표본, 3,115,001어절이 구축되었다. 수준별로는 1급 498,443어절, 2급 502,317어절, 3급 531,041어절, 4급 501,053어절, 5급 503,449어절, 6급 506,696어절로 약간의 차이는 있으나 대체로 균형적으로 구축되었다. 언어권별로는 중국어가 960,476어절로 가장 많았으며, 이어서 일본어 567,665어절, 베트남어 292,424어절, 영어 247,153어절

2) 기타 언어권에는 프랑스어, 아랍어, 독일어, 카자흐어, 크메르어, 포르투갈어, 네팔어, 이탈리아어, 광둥어, 스웨덴어, 세부아노어, 타지크어, 터키어, 말레이어, 네덜란드어, 암하라어, 폴란드어, 타타르어, 스와힐리어, 세르비아어, 루마니아어, 벵골어, 딩카어, 우르두어, 아르메니아어, 불가리아어, 에스토니아어, 자바어, 페르시아어, 르완다어, 아제르바이잔어, 테툼어, 폴라어, 헝가리어, 우크라이나어, 라틴어가 있다.

로 높은 비중을 차지하였다.

<표 10> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 형태 주석 문어 말뭉치

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
중국어	어절 수	142,139	141,121	141,382	139,203	161,979	194,280	37,662	2,710	960,476
	표본 수	2,007	1,355	1,129	1,110	1,132	1,298	23	13	8,067
일본어	어절 수	64,319	86,454	110,482	108,407	100,522	97,401		80	567,665
	표본 수	784	903	930	835	637	626		1	4,716
베트남어	어절 수	40,383	49,857	53,588	45,564	40,785	56,251	5,996		292,424
	표본 수	625	576	490	355	259	373	3		2,681
영어	어절 수	36,071	47,656	48,520	42,098	38,045	33,002	1,103	658	247,153
	표본 수	526	495	426	338	254	179	1	5	2,224
러시아어	어절 수	26,532	21,306	27,947	31,873	31,523	12,503	3,762	840	156,286
	표본 수	323	264	258	261	221	77	4	7	1,415
타이어	어절 수	23,733	28,975	22,239	12,586	10,434	9,516	212	109	107,804
	표본 수	255	218	146	90	74	58	1	1	843
광둥어	어절 수	13,656	7,105	10,342	21,615	23,047	28,847			104,612
	표본 수	214	60	83	168	156	177			858
몽골어	어절 수	21,830	11,984	15,082	12,118	18,019	20,539		511	100,083
	표본 수	279	132	128	92	134	168		3	936
인도네시아어	어절 수	18,664	14,872	10,139	15,098	7,433	6,791	1,097		74,094
	표본 수	186	141	78	102	49	42	1		599
스페인어	어절 수	11,310	16,493	19,317	9,719	7,421	3,001		293	67,554
	표본 수	155	166	161	73	52	19		1	627
우즈베크어	어절 수	27,049	1,644	2,955	5,104	20,137	3,350		333	60,572
	표본 수	245	20	26	41	141	19		2	494
프랑스어	어절 수	12,932	12,834	13,576	8,262	4,713	4,424			56,741
	표본 수	170	118	111	62	30	28			519
카자흐어	어절 수	3,641	6,139	6,990	6,330	3,281	2,396			28,777
	표본 수	45	52	54	50	22	12			235

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
말레이어	어절 수	2,550	4,777	9,834	6,420	1,907	880	2,288		28,656
	표본 수	38	49	67	36	6	6	2		204
아랍어	어절 수	7,169	5,416	7,777	3,730	1,965	1,882		684	28,623
	표본 수	97	58	69	30	14	15		4	287
기타 ³⁾	어절 수	46,465	45,684	30,871	32,926	32,238	31,633	11,000	2,664	233,481
	표본 수	672	476	296	255	219	173	8	8	2,107
합계	어절 수	498,443	502,317	531,041	501,053	503,449	506,696	63,120	8,882	3,115,001
	표본 수	6,621	5,083	4,452	3,898	3,400	3,270	43	45	26,812

2) 구어

형태 주석 구어 말뭉치는 37개 언어권의 2,788개 표본, 1,253,150어절이 구축되었다. 수준별로는 1급 233,994어절, 2급 258,567어절, 3급 260,772어절, 4급 231,661어절, 5급 134,329어절, 6급 103,746어절로 1-4급은 약간의 차이는 있으나 대체로 균형적으로 구축되었고, 5-6급은 상대적으로 규모가 적었다. 언어권별로는 중국어가 318,053어절로 가장 많았으며, 이어서 베트남어 223,688어절, 일본어 137,534어절, 타이어 118,023어절로 높은 비중을 차지하였다.

3) 기타 언어권에는 키르기스어, 이탈리아어, 버마어, 싱할라어, 스웨덴어, 독일어, 한국어, 타갈로그어, 포르투갈어, 터키어, 크메르어, 라오어, 벵골어, 네팔어, 페르시아어, 노르웨이어, 네덜란드어, 투르크멘어, 우르두어, 아제르바이잔어, 힌디어, 우크라이나어, 폴란드어, 타밀어, 헝가리어, 루마니아어, 스와힐리어, 쿠르드어, 덴마크어, 암하라어, 불가리아어, 핀란드어, 타지크어, 르완다어, 아르메니아어, 카탈루냐어, 테툼어, 이그보어, 간다어, 룩셈부르크어, 세르비아어, 체코어, 슬로바키아어, 자바어, 슬로베니아어, 위구르어, 노르웨이어 (뉘노르스크), 히브리어, 마다가스카르어, 칸나다어, 편자브어, 조지아어, 중국어 (만다린어), 마라티어, 티베트어, 말라얄람어, 에스토니아어, 그리스어, 텔루구어, 벨라루스어, 덩카어, 티그리냐어, 리투아니아어, 마오리어, 알바니아어, 구자라트어, 월로프어, 세부아노어, 파슈토어, 아이슬란드어, 츠와나어, 판테어(Fanti), 아프리카ンス어가 있다.

<표 11> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 형태 주석 구어 말뭉치

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
중국어	어절 수	47,097	55,653	54,161	49,628	37,899	51,761	21,112	742	318,053
	표본 수	211	143	133	122	68	68	15	3	763
베트남어	어절 수	49,882	47,178	45,789	49,858	21,855	8,455		671	223,688
	표본 수	118	92	90	91	41	14		2	448
일본어	어절 수	11,242	27,394	24,217	26,787	27,462	20,432			137,534
	표본 수	43	46	58	61	44	29			281
타이어	어절 수	39,603	20,683	33,473	13,259	4,672	4,376		1,957	118,023
	표본 수	128	45	79	24	7	8		5	296
스페인어	어절 수	23,820	19,148	14,730	18,790	4,225	511			81,224
	표본 수	54	37	29	39	8	1			168
러시아어	어절 수	11,510	17,477	21,348	16,483	4,610	3,682		441	75,551
	표본 수	40	48	54	42	12	11		1	208
인도네시아어	어절 수	10,801	11,186	15,122	7,613	8,307	3,732		851	57,612
	표본 수	37	23	29	16	18	10		1	134
영어	어절 수	9,204	17,635	12,798	6,458	7,088	2,282			55,465
	표본 수	26	26	28	16	9	3			108
타갈로그어	어절 수	11,948	10,987	11,635	6,632	1,682	1,711			44,595
	표본 수	31	28	13	12	2	3			89
싱할라어	어절 수	6,204	5,306	6,539	5,740	3,726	2,990			30,505
	표본 수	10	11	10	10	6	5			52
버마어	어절 수	2,633	4,889	6,653	5,224					19,399
	표본 수	5	5	6	6					22
키르기스어	어절 수	1,469	3,648	4,334	4,616	3,327			978	18,372
	표본 수	5	8	6	9	6			2	36
우즈베크어	어절 수	1,470	7,345	1,434	3,195	666				14,110
	표본 수	4	13	2	6	1				26
아랍어	어절 수		2,855	560	3,854	512	829			8,610
	표본 수		6	1	7	1	2			17

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
몽골어	어절 수	388	1,196	968	3,095	2,280			676	8,603
	표본 수	5	6	6	9	5			1	32
기타 ⁴⁾	어절 수	6,723	5,987	7,011	10,429	6,018	2,985		2,653	41,806
	표본 수	33	16	23	17	9	6	0	4	108
합계	어절 수	233,994	258,567	260,772	231,661	134,329	103,746	21,112	8,969	1,253,150
	표본 수	750	553	567	487	237	160	15	19	2,788

1.5.3. 오류 주석 말뭉치

1) 문어

오류 주석 문어 말뭉치는 54개 언어권의 11,391개 표본, 1,283,420어절이 구축되었다. 수준별로는 1급 229,078어절, 2급 198,609어절, 3급 201,707어절, 4급 197,655어절, 5급 196,182어절, 6급 256,416어절로 6급과 1급이 약간 많으나 대체로 균형적으로 구축되었다. 언어권별로는 일본어가 226,183어절로 가장 많았으며, 이어서 영어 173,790어절, 중국어 161,133어절, 베트남어 127,486어절, 러시아어가 118,612어절로 높은 비중을 차지하였다.

<표 12> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 오류 주석 문어 말뭉치

모국어	구분	1급	2급	3급	4급	5급	6급	정보 없음	합계
일본어	어절 수	46,787	22,654	23,634	25,219	39,619	68,190	80	226,183
	표본 수	493	244	225	225	288	473	1	1,949
영어	어절 수	33,400	31,922	30,834	26,512	28,494	22,628		173,790
	표본 수	455	327	265	221	192	149		1,609
중국어	어절 수	25,881	26,087	25,972	25,538	26,658	28,682	2,315	161,133

4) 기타 언어권에는 카자흐어, 크메르어, 포르투갈어, 세부아노어, 이탈리아어, 프랑스어, 독일어, 광둥어, 네팔어, 스웨덴어, 네덜란드어, 세르비아어, 터키어, 벵골어, 아르메니아어, 불가리아어, 아제르바이잔어, 헝가리어, 우르두어, 말레이어, 폴란드어, 라틴어가 있다.

모국어	구분	1급	2급	3급	4급	5급	6급	정보 없음	합계
	표본 수	365	253	220	224	218	217	12	1,509
베트남어	어절 수	12,957	13,921	18,467	18,432	11,587	52,122		127,486
	표본 수	182	146	168	157	81	350		1,084
러시아어	어절 수	17,291	19,948	21,690	25,976	20,592	12,503	612	118,612
	표본 수	171	247	201	218	165	77	4	1,083
몽골어	어절 수	13,496	11,213	11,599	11,137	11,674	20,257		79,376
	표본 수	147	122	107	88	98	166		728
타이어	어절 수	12,412	15,859	8,834	9,581	7,329	9,353		63,368
	표본 수	125	157	75	78	54	57		546
스페인어	어절 수	4,571	14,455	14,111	8,559	5,658	3,001		50,355
	표본 수	57	156	129	69	45	19		475
우즈베크 어	어절 수	24,073	408	266	3,622	17,785	2,585		48,739
	표본 수	199	6	4	28	129	17		383
프랑스어	어절 수	3,420	12,057	9,779	7,230	3,421	4,424		40,331
	표본 수	42	114	88	58	24	28		354
인도네시아어	어절 수	10,006	7,088	5,070	6,664	3,028	6,791		38,647
	표본 수	77	62	44	47	25	42		297
광둥어	어절 수	664	1,991	4,438	7,673	4,462	12,815		32,043
	표본 수	10	18	36	56	33	93		246
아랍어	어절 수	6,243	5,068	7,318	3,730	1,166	1,882		25,407
	표본 수	74	50	63	30	10	15		242
카자흐어	어절 수	1,904	5,735	5,562	5,300	2,898	2,396		23,795
	표본 수	17	50	49	43	20	12		191
키르기스 어	어절 수	1,714	3,127	2,379	1,670	6,179	2,071	389	17,529
	표본 수	19	28	22	13	46	13	3	144
기타 ⁵⁾	어절 수	14,259	7,076	11,754	10,812	5,632	6,716	377	56,626
	표본 수	165	85	125	88	44	43	1	551
합계	어절 수	229,078	198,609	201,707	197,655	196,182	256,416	3,773	1,283,420
	표본 수	2,598	2,065	1,821	1,643	1,472	1,771	21	11,391

2) 구어

오류 주석 구어 말뭉치는 35개 언어권의 1,513개 표본, 673,688 어절이 구축되었다. 수준별로는 1급 136,815어절, 2급 122,948어절, 3급 122,835어절, 4급 128,930 어절, 5급 88,908어절, 6급 66,184어절로 1-4급은 약간의 차이는 있으나 대체로 균형적으로 구축되었으며, 5-6급은 상대적으로 규모가 적었다. 언어권별로는 중국어가 155,045어절로 가장 많았으며, 이어서 일본어 129,412어절, 베트남어가 117,177어절로 높은 비중을 차지하였다.

<표 13> 2015-2025년 한국어 학습자 말뭉치 구축 결과: 오류 주석 구어 말뭉치

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
중국어	어절 수	24,033	22,422	13,901	25,437	26,701	37,440	4,892	219	155,045
	표본 수	104	57	50	56	41	43	5	2	358
일본어	어절 수	11,157	27,394	22,251	24,192	23,986	20,432			129,412
	표본 수	42	46	55	58	41	29			271
베트남어	어절 수	30,435	21,528	17,007	28,639	14,068	5,500			117,177
	표본 수	68	43	38	54	27	10			240
타이어	어절 수	30,071	3,529	22,893	4,696	850	628		1,957	64,624
	표본 수	105	13	60	12	2	2		5	199
영어	어절 수	9,204	15,925	12,728	6,458	6,136	560			51,011
	표본 수	26	24	27	16	8	1			102
스페인어	어절 수	13,631	11,082	6,773	14,435	4,225	511			50,657
	표본 수	36	20	11	28	8	1			104
인도네시아어	어절 수	8,096	5,343	9,423	4,044	4,948	284			32,138
	표본 수	25	10	18	7	9	1			70
러시아어	어절 수	2,246	2,146	7,972	4,192	1,271				17,827

5) 기타 언어권에는 버마어, 말레이어, 라오어, 타갈로그어, 크메르어, 네팔어, 이탈리아어, 포르투갈어, 한국어, 우르두어, 독일어, 싱할라어, 벵골어, 스웨덴어, 네덜란드어, 터키어, 힌디어, 우크라이나어, 루마니아어, 아제르바이잔어, 헝가리어, 폴란드어, 페르시아어, 슬로바키아어, 자바어, 세르비아어, 마다가스카르어, 히브리어, 노르웨이어, 핀란드어, 카탈루냐어, 이그보어, 텔루구어, 불가리아어, 르완다어, 세부아노어, 파슈토어, 리투아니아어, 암하라어가 있다.

모국어	구분	1급	2급	3급	4급	5급	6급	6급 이상	정보 없음	합계
	표본 수	6	6	16	8	2				38
아랍어	어절 수		2,855	560	3,854	512	829			8,610
	표본 수		6	1	7	1	2			17
타갈로그 어	어절 수	3,996	1,650	767	428					6,841
	표본 수	6	3	2	2					13
카자흐어	어절 수	758	959		3,326					5,043
	표본 수	2	1		5					8
키르기스 어	어절 수		1,830	217	2,254	549				4,850
	표본 수		4	1	5	1				11
포르투갈 어	어절 수		1,627	788	1,174					3,589
	표본 수		5	3	2					10
이탈리아 어	어절 수	1,277	372	347		1,168				3,164
	표본 수	14	1	1		2				18
몽골어	어절 수		564	188	788	1,416				2,956
	표본 수		2	1	2	4				9
기타 ⁶⁾	어절 수	1,911	3,722	7,020	5,013	3,078				20,744
	표본 수	9	7	16	9	4				45
합계	어절 수	136,815	122,948	122,835	128,930	88,908	66,184	4,892	2,176	673,688
	표본 수	443	248	300	271	150	89	5	7	1,513

6) 기타 언어권에는 프랑스어, 우즈베크어, 버마어, 광둥어, 싱할라어, 네팔어, 크메르어, 독일어, 스웨덴어, 네덜란드어, 세르비아어, 터키어, 벵골어, 불가리아어, 아제르바이잔어, 헝가리어, 우르두어, 말레이어, 폴란드어, 라틴어가 있다.

2. 한국어 학습자 쓰기 평가 자료 구축

2.1. 평가 체계⁷⁾

2.1.1. 평가 등급

한국어 학습자 쓰기 평가 자료 구축은 최근 많은 연구가 이루어지고 있는 쓰기 자동 채점을 비롯한 기술 개발에 활용 가능한 인공지능 학습용 데이터 구축을 위한 기반 자료를 마련하기 위한 것이다. 이에 따라 본 사업에서는 2023-2024년 기초 연구를 통해 마련한 평가 체계에 근거하여 ‘국제 통용 한국어교육 표준 모형’의 6등급 체계에 따라 평가 작업을 수행하였다. 다음은 본 연구의 토대가 된 ‘국제 통용 한국어교육 표준 모형’의 총괄 목표이다.

<표 14> ‘국제 통용 한국어교육 표준 모형’의 쓰기 총괄 목표(국립국어원, 2017)

등급	세부 내용
1급	정형화된 표현을 이용해 일상생활에서 매우 간단한 의사소통(자기소개, 인사, 물건 사기 등)을 할 수 있다. 기초적 어휘와 간단한 문장을 이해하고 사용할 수 있다. 가장 기본적인 한국의 일상생활 문화를 이해하고 자국의 문화와 비교할 수 있다
2급	기초 어휘와 단순한 문장을 이용해 일상생활에서 자주 마주치는 간단한 문제를 해결할 수 있다. 일상생활에서 자주 다루는 개인적·구체적 주제에 대해 간단하게 의사소통할 수 있다. 기본적인 한국의 일상생활 문화를 이해하고 자국의 문화와 비교할 수 있다.
3급	일상생활에서 자주 마주치는 문제를 대부분 해결할 수 있으며, 친숙한 사회적 맥락에서 요구되는 과제를 어느 정도 해결할 수 있다. 친숙한 사회적·추상적 주제와 자신의 관심 분야에 대해 간단하게 의사소통할 수 있다. 문어와 구어를 어느 정도 구분해 사용할 수 있다. 대부분의 한국의 일상생활 문화와 대표적인 행동 문화, 성취 문화를 이해하고 자국의 문화와 비교할 수 있다.
4급	친숙한 사회적 맥락에서 요구되는 과제를 대부분 해결할 수 있으며, 자신의

7) 평가 체계는 2023년 기초 연구를 통해 마련되었으며, 2023-2024년 쓰기 평가 자료 구축에 적용된 바 있다. 본 연구에서는 자료의 호환성과 일관성을 고려하여 동일한 체계에 따라 평가 작업을 수행하였다. 이에 따라 본 보고서의 평가 체계는 2023-2024년 수록 내용을 옮겨 제시하였다.

	직업과 관련된 기본적인 업무를 처리할 수 있다. 친숙한 사회적·추상적 주제와 자신의 관심 분야에 대해 비교적 유창하게 의사소통할 수 있다. 문어와 구어를 적절히 구분해 사용할 수 있으며, 대상과 상황에 따라 격식과 비격식을 구분하여 사용할 수 있다. 한국의 대표적인 행동 문화, 성취 문화를 이해하고 자국의 문화와 비교할 수 있다.
5급	덜 친숙한 사회적 맥락에서 요구되는 과제를 어느 정도 해결할 수 있으며, 자신의 업무나 학업과 관련된 기본적 의사소통 기능을 수행할 수 있다. 친숙하지 않은 사회적·추상적 주제 및 자신의 직업이나 학문 영역에 대해 간단하게 의사소통할 수 있다. 공식적인 맥락에서 격식을 갖추어 의사소통할 수 있다. 한국의 다양한 행동 문화, 성취 문화 및 대표적인 관념 문화를 이해하며 자국의 문화와 비교하여 문화의 다양성과 특수성을 이해할 수 있다.
6급	덜 친숙한 사회적 맥락에서 요구되는 과제를 적절히 해결할 수 있으며, 자신의 업무나 학업과 관련된 의사소통 기능을 어느 정도 수행할 수 있다. 친숙하지 않은 사회적·추상적 주제 및 자신의 직업이나 학문 영역에 대해 비교적 유창하게 다룰 수 있다. 한국인이 즐겨 사용하는 담화·텍스트 구조를 적절히 이용할 수 있다. 한국의 다양한 행동 문화, 성취 문화, 관념 문화를 이해하며 자국의 문화와 비교하여 문화의 다양성과 특수성을 이해할 수 있다.

다음은 총괄 목표에 따른 등급별 목표 및 내용 기술이다.

<표 15> '국제 통용 한국어교육 표준 모형'의 쓰기 등급별 목표 및 내용(국립국어원, 2017)

등급	목표/ 내용	세부 내용
1급	목표	기본적인 맞춤법에 맞게 글자를 정확하게 쓸 수 있으며 일상생활에 관한 간단한 글을 쓸 수 있다.
	내용	<ul style="list-style-type: none"> • 한글 자음과 모음을 결합해 글자를 쓴다. • 맞춤법에 맞게 짧은 문장을 바르게 쓴다. • 간단한 메모를 한다. • 일상생활에 관한 짧은 글을 간단한 구조로 쓴다.
2급	목표	일상생활에서 경험한 일이나 친숙한 인물에 관한 글을 쓸 수 있다.
	내용	<ul style="list-style-type: none"> • 자신의 일과를 비교적 정확하게 쓴다. • 경험한 일이나 앞으로의 계획에 관해 문장과 문장이 자연스럽게 연결되도록 쓴다. • 친숙한 인물, 사물, 장소 등을 간단하게 소개하는 글을 장르적 특성에 맞게 쓴다.

등급	목표/ 내용	세부 내용
3급	목표	친숙한 사회적·추상적 주제로 된 글을 간단한 구조로 쓸 수 있다.
	내용	<ul style="list-style-type: none"> • 자신과 관련된 생활문을 비교적 정확하게 쓴다. • 친숙한 사회적·추상적 주제(직업, 사랑, 교육 등)에 관한 글을 간단한 구조로 쓴다. • 실용문(안내문, 전자 우편(이메일) 등)을 단락과 단락이 자연스럽게 연결되도록 쓴다. • 간단한 구조의 설명문에 핵심 내용이 잘 드러나도록 쓴다.
4급	목표	친숙한 사회적·추상적 주제로 된 글을 정확하게 쓸 수 있으며 설명문, 논설문, 쉽고 짧은 감상문을 쓸 수 있다.
	내용	<ul style="list-style-type: none"> • 친숙한 사회적·추상적 주제(직업, 사랑, 교육 등)에 관하여 정확하게 설명하거나 의견을 들어 주장하는 글을 쓴다. • 친숙한 소재를 다루는 논설문의 구조에 맞게 자신의 주장과 뒷받침 내용을 쓴다. • 짧고 간단한 구조의 수필을 일관된 내용으로 쓴다. • 예시, 비교/대조 등을 활용하여 글을 쓴다.
5급	목표	친숙하지 않은 사회적·추상적 주제나 자신의 전문 분야에 관한 글을 구조에 맞게 쓸 수 있다
	내용	<ul style="list-style-type: none"> • 친숙하지 않은 사회적·추상적 주제(정치, 경제, 과학 등)에 관해 논리적 구조를 반영한 글을 쓴다. • 자신의 전문 분야에 관하여 핵심 내용이 드러나도록 글을 쓴다. • 다양한 소재의 글을 요약하고 자신의 의견을 반영한 요약문을 쓴다. • 정의, 인용 등을 활용하여 글을 쓴다.
6급	목표	친숙하지 않은 사회적·추상적 주제나 자신의 전문 분야에 관한 글을 논리적인 구조로 쓸 수 있으며 다양한 장르의 특성을 고려한 글을 쓸 수 있다.
	내용	<ul style="list-style-type: none"> • 친숙하지 않은 사회적·추상적 주제(정치, 경제, 과학 등)에 관해 논리적이고 정확하게 의견을 전개하는 글을 쓴다. • 자신의 전문 분야에 관하여 핵심 내용과 세부 내용이 연결되도록 글을 쓴다. • 평론, 학술 보고서, 학술 논문 등의 전문적인 글의 특성을 이해하고 간단하지만 일관된 내용 구조를 가진 글을 쓴다. • 비유, 분류, 분석 등을 활용하여 글을 쓴다.

2.1.2. 평가 구인

평가 구인은 연계성을 고려하여 2023-2024년 연구에서 마련한 구인 체계를 동일하게 적용하였다. 구인은 내용 조직, 글의 구조, 언어 사용의 세 가지 범주로 구성되며, 각 구인별 세부 평가 내용은 다음과 같다.

<표 16> 쓰기 평가 구인

평가 범주	평가 내용
내용 조직	<ul style="list-style-type: none"> • 글의 목적과 기능에 맞는 적절한 형식으로 글을 썼는가 • 주제에 관련된 내용으로 구성하였는가 • 주어진 내용을 풍부하게 표현하였는가
글의 구조	<ul style="list-style-type: none"> • 글의 구성이 명확하고 논리적인가 • 글의 내용에 따라 단락 구성이 잘 이루어졌는가 • 문장 간의 연결이 자연스럽고 짜임새 있는가 • 글의 내용이 긴밀하게 구성되었는가
언어 사용	<ul style="list-style-type: none"> • 문법과 어휘를 정확하게 사용하였는가 • 문법과 어휘를 다양하게 사용하였는가 • 문법과 어휘를 적절하게 사용하였는가

2.1.3. 평가 방식과 기준

본 사업에서는 채점의 신뢰도를 제고하기 위하여 종합적 채점과 분석적 채점 방식을 함께 적용하였다. 평가자가 작문과 함께 제공된 메타 정보, 경험적 직관을 토대로 잠정 등급을 부여한 후 평가 기준에 따라 분석적 채점을 하도록 하였다. 분석적 채점을 할 때에는 앞서 제시한 구인을 토대로 하여 다음과 같이 5간 척도로 루브릭을 제공하여 정교한 채점을 할 수 있도록 하였다.

<표 17> 쓰기 평가 기준

구인 척도	내용 조직	글의 구조	언어 사용
5	<ul style="list-style-type: none"> • 주제와 관련된 내용을 풍부하게 구성함 • 글의 내용이 장르의 특성에 맞게 잘 	<ul style="list-style-type: none"> • 글의 시작과 마무리, 단락이 짜임새 있게 구성되어 글의 완성도가 높음 	<ul style="list-style-type: none"> • 글의 주제와 관련된 어휘와 문법을 풍부하고 다양하게 사용함

구인 척도	내용 조직	글의 구조	언어 사용
	구성됨	<ul style="list-style-type: none"> 글의 내용이 담화적으로 긴밀하게 구성됨 	<ul style="list-style-type: none"> 어휘와 문법을 글의 특성에 맞게 적절하게 사용함 해당 수준의 어휘와 문법 사용에서 오류가 거의 없음
4	<ul style="list-style-type: none"> 주제와 관련된 내용을 비교적 풍부하게 구성함 글의 내용이 장르의 특성에 맞게 비교적 잘 구성됨 	<ul style="list-style-type: none"> 글의 시작과 마무리, 단락이 어느 정도 짜임새 있게 구성되어 글의 완성도가 비교적 높음 글의 내용이 담화적으로 비교적 긴밀하게 구성되어 글의 흐름을 방해하지 않음 	<ul style="list-style-type: none"> 글의 주제와 관련된 어휘와 문법을 비교적 풍부하고 다양하게 사용함 글의 특성을 고려하여 어휘와 문법을 비교적 적절하게 사용함 어휘와 문법 사용에서 오류가 약간 있으나 글의 내용을 이해하는 데에 문제가 없음
3	<ul style="list-style-type: none"> 주제와 관련된 내용으로 글을 구성하였으나 내용이 다소 빈약함 간혹 부적절한 부분이 있으나 글의 내용이 장르의 특성에 어느 정도 맞게 구성됨 	<ul style="list-style-type: none"> 글의 시작과 마무리, 단락이 짜임새 있게 구성되지 못하여 글의 완성도가 부족함 글의 내용이 담화적으로 어느 정도 긴밀하게 구성되었으나 간혹 글의 흐름을 방해하는 부분이 있음 	<ul style="list-style-type: none"> 제한적이지만 글의 주제와 관련된 어휘와 문법을 어느 정도 풍부하고 다양하게 사용함 간혹 부적절한 부분이 있으나 글의 특성을 고려하여 어휘와 문법을 어느 정도 적절하게 사용함 어휘와 문법 사용에서 오류가 많아 글의 내용을 이해하는 데에 다소 문제가 있음
2	<ul style="list-style-type: none"> 간혹 주제와 관련이 없는 내용이 글에 포함되거나 내용이 빈약함 장르의 특성에 맞게 글의 내용을 구성하지 못함 	<ul style="list-style-type: none"> 글의 시작과 마무리, 단락 구성에서 짜임새가 없고 각각의 내용을 나열함 글의 내용이 담화적으로 긴밀하지 못하여 글의 흐름이 	<ul style="list-style-type: none"> 글의 주제와 관련된 어휘를 다양하게 사용하지 못하고 쉽고 단순한 어휘와 문법을 반복적으로 사용함 글의 특성을 고려하여 어휘와 문법을

구인 척도	내용 조직	글의 구조	언어 사용
		매끄럽지 못함	적절하게 사용하지 못함 • 어휘와 문법 사용에서 오류가 너무 많아 글의 내용을 이해하기가 어려움
1	• 주제와 관련된 내용으로 글을 구성하지 못함	• 완결된 구조의 담화를 구성하지 못함	• 어휘와 문법 사용에 오류가 많아 글의 내용을 이해할 수 없음

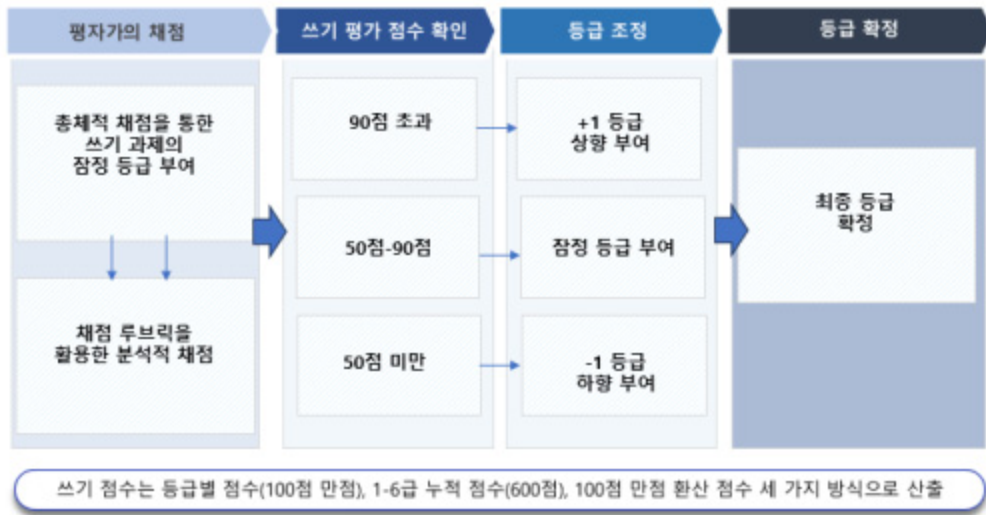
2.1.4. 평가 점수 및 등급 부여 체계

평가 점수는 각 구인별로 내용 조직 30점, 글의 구조 30점, 언어 사용 40점으로 총점을 100점으로 설정하였다. 그리고 다음과 같이 척도별 점수 구간을 설정하여 채점자가 구인별로 적정 점수를 입력하면 그 점수를 토대로 척도 값이 자동 반환되도록 하였다.

<표 18> 척도별 점수 구간

구인 척도	내용 조직(30점)	글의 구조(30점)	언어 사용(40점)
5	28점 이상	28점 이상	37점 이상
4	24-27점	24-27점	32-36점
3	19-23점	19-23점	26-31점
2	15-18점	15-18점	21-25점
1	14점 이하	14점 이하	20점이하

등급은 평가자가 분석적인 평가에 앞서 작문과 함께 제공된 메타 정보, 경험적 직관을 토대로 부여한 잠정 등급을 평가 점수에 따라 조정하는 방식으로 최종 확정되었다. 즉, 평가 기준에 따라 분석적 채점을 한 후 산출된 점수가 50-90점이면 잠정적으로 부여한 등급을 유지하고, 90점을 초과하면 등급을 +1 등급 상향 조정하고, 50점 미만이면 -1 등급 하향 조정하였다.



<그림 9> 평가 점수 및 등급 부여 체계

쓰기 점수는 각 등급별로 부여한 100점 만점의 점수, 1-6급까지 각 등급별로 100점씩 누적하여 산출한 600점 만점의 점수, 이를 다시 100점 만점으로 환산한 점수 세 가지로 제시하였다. 다음은 세 가지 유형의 점수 환산표 일부를 보인 것이다.

<표 19> 점수 환산표

등급	등급별 점수	1-6급 누적 점수 (600점 만점)	환산 점수 (100점 만점)
1급	50	50	8
1급	75	75	13
1급	100	100	17
2급	50	150	25
2급	75	175	29
2급	100	200	33
3급	50	250	42
3급	75	275	46
3급	100	300	50

4급	50	350	58
4급	75	375	63
4급	100	400	67
5급	50	450	75
5급	75	475	79
5급	100	500	83
6급	50	550	92
6급	75	575	96
6급	100	600	100

2.2. 채점자 훈련

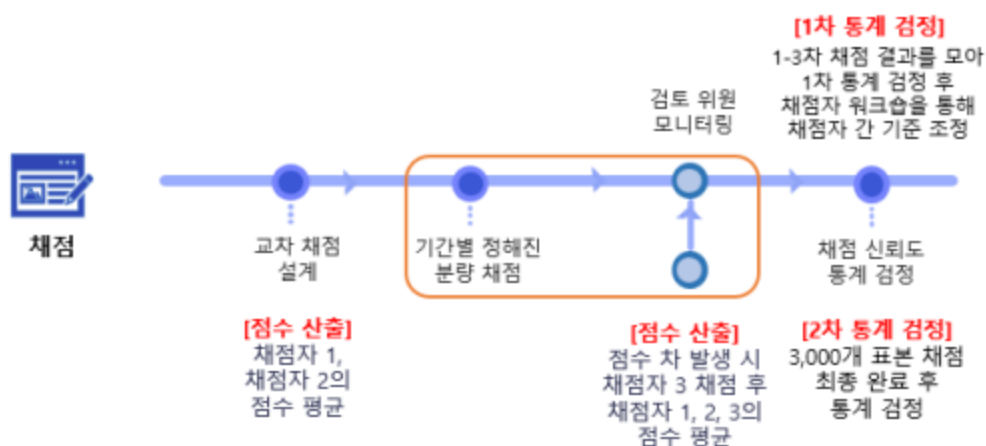
본 사업에서는 채점자 간 신뢰도와 채점자 내의 신뢰도를 확보하기 위하여 다음과 같은 절차에 따라 채점자 훈련을 실시하였다.



<그림 10> 채점자 교육 절차

본 사업을 수행하는 동안 총 3회의 채점자 워크숍을 실시하였다. 채점자 워크숍은 평가 작업의 목적과 평가 체제에 대해서 설명하고 연습 채점을 한 후 채점자 간의 차이를 좁히기 위해 채점 기준을 조정하는 방식으로 진행되었다. 그 외에도 격주로 정기 실무회의를 진행하면서 회차별 채점 결과를 검토하였으며, 평가 작업을 진행하면서 쟁점이 발생할 때마다 온라인 매체를 통해 실시간으로 소통을 하거나 수시 회의를 소집하여 문제를 해결하고 채점자들의 전문성을 제고하였다.

아울러 채점 신뢰도를 확보하기 위하여 하나의 표본에 대하여 두 명의 채점자가 교차 채점을 하도록 하고 평균 점수를 산출하여 점수를 부여하였다. 이때 채점자가 부여한 등급에 차이가 있는 경우 상위 채점자가 3차 채점을 실시하였으며, 세 명의 평균 점수를 산출하여 점수를 부여하였다. 또한 채점자 교육의 일환으로 초기 300개 표본의 채점 결과에 대한 통계 검정을 실시하였다. 그리고 그 결과를 채점자들에게 피드백함으로써 채점자 간, 그리고 채점자 내적 신뢰도를 확보하고자 하였다.



<그림 11> 채점 신뢰도 확보를 위한 쓰기 평가 절차

다음은 초기 300개 표본의 채점 결과에 대한 통계 검정 결과로 피어슨 상관계수와 급내 상관계수(intra-class correlation coefficient: ICC), 가중 카파(Weighted Kappa)를 분석하였다. 분석 결과 세 가지 지표 모두 총점에서 높은 신뢰도를 보였다.

<표 20> 300개 표본의 채점자 간 신뢰도 검정 결과

구인	ICC	상관계수	cohen의 가중 카파
내용 조직	.874	.779	.574
글의 구조	.781	.641	.563
언어 사용	.908	.839	.717
총점	.941	.894	.905

- **급내 상관계수** : 급내 상관계수는 평가의 반복성과 재현성, 평가자 간 신뢰도를 평가하기 위해 사용되는 또 다른 신뢰도 계수이다. 0부터 1 사이의 값을 가지며 0.8 이상이면 신뢰도가 높다고 할 수 있다. 분석 결과에 따르면 내용 조직 0.874, 글의 구조 0.781, 언어 사용은 0.908로 언어 점수가 가장 높은 신뢰도를 보였고 구조 점수가 상대적으로 낮은 것으로 확인되었다.
- **피어슨 상관계수** : 두 채점자가 동일한 집단의 피험자에게 얼마나 유사하게 점수를 부여하고 있는가를 분석하는 방법이다. 상관계수가 0.6 이상이면 신뢰성이 있다고 보며 점수가 높을수록 좋다. 분석 결과에 따르면 내용 조직 0.779, 글의 구조 0.641, 언어 사용은 0.839로 기준점 이상의 점수를 보여 어느 정도 신뢰성이 있는 확인되었다.
- **가중 카파(Weighted Kappa)** : 순서형 척도로 등급 간의 불일치를 반영할 수 있는 점수이다. Kappa값이 0.41-0.60 구간이면 보통, 0.60 이상이면 신뢰성이 높다고 보며 점수가 높을수록 좋다. 분석 결과에 따르면 내용 조직과 글의 조직은 각각 0.574, 0.563으로 일정 수준 이상의 신뢰도를 보였으며, 언어 사용은 0.717로 신뢰도가 높은 것으로 확인되었다.

본 사업에서는 채점자 간 신뢰도를 높이기 위하여 이러한 결과와 함께 평가자 개인의 채점 경향에 대하여 피드백을 하여 평가자들이 자신의 채점 기준을 재검검하고 조정하도록 하였다.

2.3. 학습자 쓰기 평가 자료 구축

2.3.1. 학습자 쓰기 평가 자료 분포

본 사업에서는 주제 1(10년 후 나의 계획, 생활문) 1,804건, 주제 2(내가 가 본 여행지, 생활문) 1,694건, 주제 3(내가 생각하는 행복과 성공, 논설문) 1,502건으로 총 5,000건의 학습자 작문을 평가하였다.⁸⁾ 평가 대상이 된 작문은 총 88개국 학습자가 산출한 자료로, 국적별 자료의 분포는 다음과 같다.

8) 본 사업의 과업에서는 세 가지 주제에 대하여 각 1,700건 이상의 표본을 평가 대상으로 쓰기 평가를 수행하도록 되어 있으나, 2025년 사업은 신규 수집 없이 기구축 말뭉치를 평가 대상으로 함에 따라 말뭉치에서 주제별 자료를 추출한 결과에 따라 최소 1,502건에서 최대 1,804건을 평가 대상으로 하게 되었다.

<표 21> 학습자 쓰기 평가 자료의 국적 분포

국적	10년 후 나의 계획	내가 가 본 여행지	내가 생각하는 행복과 성공	합계
가나	1	1	1	3
가봉	3			3
과테말라	1			1
나이지리아		1		1
네덜란드		1	1	2
네팔		2		2
노르웨이	1	1	1	3
뉴질랜드		1		1
대만	23	17	19	59
독일	2	3	1	6
라오스	5	17		22
러시아	17	13	31	61
르완다	1			1
마카오	3	1		4
말레이시아	15	12	12	39
멕시코	2	2	2	6
모로코			2	2
몰도바		2	1	3
몽골	111	117	176	404
미국	5	10	7	22
미얀마	35	8	20	63
바레인	1			1
방글라데시	49	102	2	153
베네수엘라		1	2	3
베트남	389	541	357	1,287
벨라루스			1	1
불가리아			1	1
브라질		5	3	8

국적	10년 후 나의 계획	내가 가 본 여행지	내가 생각하는 행복과 성공	합계
브루나이		1		1
사우디아라비아	2		1	3
세르비아		1		1
소말리아		1		1
스리랑카	10		4	14
스웨덴	1	6	1	8
스페인	1	3	2	6
슬로바키아		1		1
슬로베니아	7	1	7	15
싱가포르	1	4	4	9
싱가폴	1	2		3
아랍에미리트			2	2
아르메니아	1			1
아일랜드		4		4
아제르바이잔	1	1	1	3
알제리			3	3
에콰도르			1	1
에티오피아	1		1	2
엘살바도르		1		1
영국	5	6	1	12
요르단		1		1
우간다	1	2		3
우루과이	1			1
우즈베키스탄	191	140	75	406
우크라이나	1			1
이란	2	1	1	4
이스라엘		1		1
이집트	21			21
이탈리아	6	8	3	17

국적	10년 후 나의 계획	내가 가 본 여행지	내가 생각하는 행복과 성공	합계
인도	5	1	8	14
인도네시아	182	77	79	338
일본	223	182	144	549
일본어	5		2	7
중국	248	184	245	677
카자흐스탄	24	4	14	42
캄보디아	3	3	5	11
캐나다	1	2	1	4
케냐	2	1	1	4
코트디부아르			1	1
콜롬비아	1	2		3
콩고	1	1		2
크로아티아	1			1
키르기스스탄	4	11	15	30
타지키스탄		3	2	5
태국	136	107	179	422
터키		1	2	3
투르크메니스탄	4	2	2	8
튀니지	1		1	2
파나마		1		1
파키스탄	2	4	1	7
페루		1		1
포르투갈	1	5		6
폴란드	1	1	11	13
프랑스	10	9	3	22
핀란드		1		1
필리핀	1	5	1	7
한국	2	1	9	12
헝가리	1	2	1	4

국적	10년 후 나의 계획	내가 가 본 여행지	내가 생각하는 행복과 성공	합계
호주		1	1	2
홍콩	27	39	27	93
합계	1,804	1,694	1,502	5,000

2.3.2. 학습자 쓰기 평가 결과

5,000건의 표본을 평가한 결과, 주제별 표본은 다음과 같이 최종 등급이 부여되었다. 각 표본에 대해서는 등급 외에 등급별 점수(100점 만점), 1-6급 누적 점수(600점 만점), 최종 환산 점수(100점 만점)와 함께 총평을 제시하였다.

<표 22> 학습자 쓰기 평가 결과

주제	1급	2급	3급	4급	5급	6급	합계
10년 후 나의 계획	206	493	628	402	58	17	1,804
내가 가 본 여행지	641	837	147	55	11	3	1,694
내가 생각하는 행복과 성공	18	61	489	423	317	194	1,502
합계	865	1,391	1,264	880	386	214	5,000

쓰기 평가 결과에 대한 신뢰도 검정 결과는 다음과 같다.

○ 주제 1 - 10년 후 나의 계획

피어슨 상관계수 총점 .914, 급내 상관계수 총점 .849, 가중 카파 총점 .981로 세 가지 지표에서 모두 신뢰도가 높은 것으로 확인되었다.

<표 23> 주제 1의 평가 결과에 대한 채점자 간 신뢰도 검정 결과

구인	ICC	상관계수	cohen의 가중 카파
내용 조직	.849	.751	.556
글의 구조	.843	.733	.526
언어 사용	.888	.802	.615
총점	.914	.849	.981

- 피어슨 상관계수 : 내용 점수 0.849, 글의 구조 0.843, 언어 사용 0.888로, 세 가지 구인 모두 채점 신뢰도가 높은 것으로 확인되었다.
- 급내 상관계수(intra-class correlation coefficient: ICC) : 내용 조직이 0.751, 글의 구조 0.733, 언어 사용 0.802로 모두 신뢰도가 높은 것으로 확인되었으며, 특히 언어 사용에서 높은 신뢰도를 보이는 것으로 확인되었다.
- 가중 카파(Weighted Kappa) : 내용 조직 .556, 글의 구조 0.526, 언어 사용 0.615로 일정 수준 이상의 신뢰도가 있는 것으로 확인되었다.

○ 주제 2 - 내가 가 본 여행지

피어슨 상관계수 총점 .917, 급내 상관계수 총점 .846, 가중 카파 총점 .992로 세 가지 지표에서 모두 신뢰도가 높은 것으로 확인되었다.

<표 24> 주제 2의 평가 결과에 대한 채점자 간 신뢰도 검정 결과

구인	ICC	상관계수	cohen의 가중 카파
내용 조직	.872	.774	.655
글의 구조	.864	.764	.512
언어 사용	.896	.811	.574
총점	.917	.846	.992

- 피어슨 상관계수 : 내용 점수 0.872, 글의 구조 0.864, 언어 사용 0.896으로, 주제1에 비해 신뢰도 수치가 약간 더 높았으며 세 가지 구인 모두 채점 신뢰도가 높은 것으로 확인되었다.
- 급내 상관계수(intra-class correlation coefficient: ICC) : 내용 조직이

0.774, 글의 구조 0.764, 언어 사용 0.811로 모두 신뢰도가 높은 것으로 확인되었다.

- 가중 카파(Weighted Kappa) : 내용 조직 .655, 글의 구조 0.512, 언어 사용 0.574로 일정 수준 이상의 신뢰도가 있는 것으로 확인되었다.

○ 주제 3 - 내가 생각하는 행복과 성공

피어슨 상관계수 총점 .935, 급내 상관계수 총점 .882, 가중 카파 총점 .985로 세 가지 지표에서 모두 신뢰도가 높은 것으로 확인되었다. 또한 주제1, 주제2에 비해 채점 점수의 신뢰도와 척도 점수의 일치율이 매우 높았다.

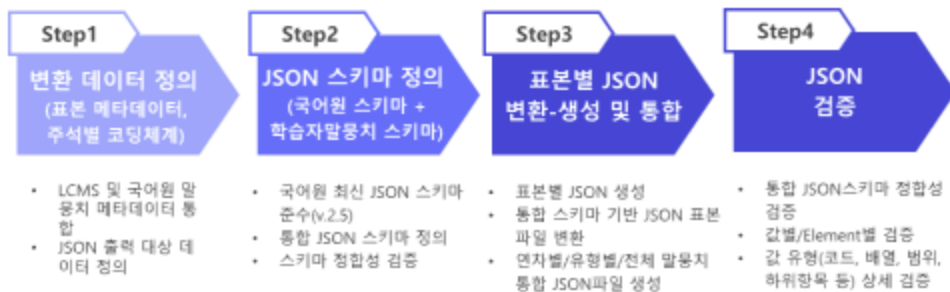
<표 25> 주제 3의 평가 결과에 대한 채점자 간 신뢰도 검정 결과

구인	ICC	상관계수	cohen의 가중 카파
내용 조직	.898	.816	.875
글의 구조	.886	.799	.884
언어 사용	.919	.852	.868
총점	.935	.882	.985

- 피어슨 상관계수 : 내용 점수 0.898, 글의 구조 0.886, 언어 사용 0.919로, 세 가지 구인 모두 채점 신뢰도가 높은 것으로 확인되었다.
- 급내 상관계수(intra-class correlation coefficient: ICC) : 내용 조직 0.816, 글의 구조 0.799, 언어 사용 0.852로 모두 신뢰도가 높은 것으로 확인되었다.
- 가중 카파(Weighted Kappa) : 내용 조직 .875, 글의 구조 0.884, 언어 사용 0.868로 다른 두 주제에 비해 신뢰도가 매우 높은 것으로 확인되었다.

3. 한국어 학습자 말뭉치 파일 변환

학습자 말뭉치는 학습자 말뭉치 구축 도구(LCMS : Learner's Corpus Management System)을 통해 구축-관리된다. 구축 도구에서는 DBMS에 최적화된 코드값 및 수치 정보로 보관되고 구축이 완료된 후에는 말뭉치 나눔터 웹 검색용 데이터와 말뭉치 주석 상태에 따라 엑셀, 텍스트, XML 포맷으로 다운로드 받을 수 있다. 매년 구축된 전체 말뭉치를 원시 말뭉치, 형태 주석 말뭉치, 오류 주석 말뭉치별로 텍스트(원시), 엑셀(형태), XML(오류)을 생성해 학습자 말뭉치 나눔터에서 배포한다. 최근에는 JSON 포맷이 보편적으로 사용되기 때문에 추가 변환기로 했다. JSON 포맷으로 학습자 말뭉치를 변환하기 위해서는 먼저 다음 절차에 따라 변환 대상 데이터와 JSON 스키마를 결정해야 한다.



<그림 12> 학습자 말뭉치 데이터 JSON 생성 절차

JSON 스키마는 국립국어원의 ‘모두의 말뭉치’ 배포용 데이터 규격에 학습자 말뭉치의 메타 데이터를 추가하고 오류 주석의 주석 정보 특수성을 감안해 수립하였다. 학습자 말뭉치는 문어와 구어 2종의 원시 말뭉치가 있고, 형태 주석과 오류 주석 2종의 주석 말뭉치, 1종의 쓰기 평가 말뭉치(문어)가 있다. 각각의 배타적인 하위 구조를 갖지만, 텍스트, 단락, 문장(발화), 구문, 어절 수준에서 공통 오프셋 정보를 가지므로 스키마를 통합하거나 분리할 수 있다. 단, 구어와 문어는 메타데이터와 발화-문장 간에 동일 의미의 서로 다른 엘리먼트를 사용하기 때문에 통합하는 데 제약이 있다.

학습자 말뭉치 데이터를 JSON으로 생성하는 방법은 크게 2가지가 있다. 하나는 구축 도구나 DBMS를 통해 직접 JSON을 생성하는 방법이고, 다른 하나는 배포된 파일로부터 변환 생성하는 방식이다. 전자는 구축 도구에 포함된 모든 데이터를 다양한 JSON 스키마로 대응시킬 수 있지만, DBMS로부터 매번

필요한 데이터를 조회해 와야 하므로 자료량에 따라 소요 시간이 크게 증가하고 개인정보 처리, 주석 정보의 매핑, 코드 데이터의 문자열 변환 등 다양한 전후 처리를 함께 수행해야 한다. 배포된 파일은 이러한 전후 처리를 거쳤기 때문에 추가적인 변환이 필요 없고, 로컬 저장소 파일을 직접 읽어 처리하기 때문에 변환 시간도 매우 작다. 그러나, 배포 과정에서 변환 처리된 원본값을 복원하거나 제외된 메타 데이터를 추가할 수는 없다.⁹⁾ 이번 JSON 데이터 생성은 기존 배포 표본 정보들과 일관성을 유지하므로 배포 파일로부터 변환한다. 최종 확정된 JSON 규격은 다음과 같다.

<표 26> 한국어 학습자 말뭉치 데이터 구조(JSON)

Lv1	Lv2	Lv3	Lv4	Lv5	타입	예시
id					str	* 말뭉치 파일 ID * 고유 ID로 중복이 없어야 함.예) 국립국어원 학습자 원시말뭉치 WLRW2200044354
metadata					obj	* 파일의 메타 정보
	title				str	* 원시 말뭉치: 국립국어원 [말뭉치 유형 구분] [파일 ID] 예) 국립국어원 신문 말뭉치 NWRW1900000001
	creator				str	* 생성자: 국립국어원
	distributor				str	* 배포자: 국립국어원
	year				str	* 말뭉치 구축 연도 예) 2019
	category				str	* 분류 예) 신문>전국 종합지
	annotation_level				arr (str)	* 분석 층위 예) 원시
	sampling				str	* 샘플링 방식 (본문 전체 / 부분 추출 - 임의 추출 / 부분 추출 - 특정 부분 추출 등)
learners_info					obj	* 학습자 정보
	learner_no				str	56_00044
	learner_type				str	일반
	learner_type_detail				str	일반
	longitudinal_learner				str	18
	longitudinal_learner_id				str	12
	learning_week				str	16
	age				str	37
	data_grade				str	2급
	nationality				str	홍콩
	overseas				str	외국인
	mother_tongue				str	광둥어
	learning_period				str	6개월
	residence_period				str	6개월
	learning_goal				str	취미
	job				str	무직
	major				str	한국어교육
	other_lang_1				str	중국어

9) 메타데이터 추가 등은 표본 고유번호를 이용해 DBMS로부터 정보를 조회해 추가할 수는 있지만 앞서 소개한 방식과 동일한 방법이다.

	other_lang_2			str	영어
	other_lang_3			str	-
	other_lang_4			str	-
	topik_grade			str	3
	gender			str	여성
document				arr (obj)	* 문서 정보
	id			str	* 문서 ID * '원시 말뭉치 파일 ID.파일 내 문서 번호'로 구성
	metadata			obj	* 문서의 메타 정보
		file_name		str	원본파일명 : 2022 기획_문어_광동어_건국대_2급_00044.jpg,
		agreement_file_name		str	동의서파일명 : AG_2022 기획_문어_광동어_건국대_2급_00044.jpg,
		learn_env		str	학습자 환경 : 국내,
		source_type		str	문어/구어
		collect_time		str	수집년도 : 2022,
		collect_agency		str	수집기관 : 건국대,
		collect_agency_detail		str	수집기관 상세
		assignment_type		str	과제 유형 : 기획 작문,
		assignment_genre		str	과제 장르 : 설명문,
		assignment_theme		str	과제 주제 : 내가 가장 좋아하는 것과 싫어하는 것.
		full_text_open		str	전문공개 : 1/0
		sentence_count		str	문장수 : 20,
		word_count		str	어절수 : 126,
		register_year		str	LCMS등록년도 : 2023,
		input_year		arr	작업년도 : 2023
	paragraph			arr (obj)	* 문단
		id		str	* 문단 ID. * '문서 ID.문서 내 문단 순서'로 구성. 문서 내 문단 순서는 1부터 시작
		form		str	* 정제된 형태
		original_form		str	* 원문 표기된 그대로의 형태
		sentence		arr (obj)	
			id	str	* 문장 id
			form	str	* 문장 문자열
			morphe me	arr (obj)	* 형태소 분석 정보
			id	num	* 형태소 ID 문장 내 1부터 시작
			form	str	* 형태소
			label	str	* 형태소 품사 태그
			word_id	num	* 형태소가 포함된 어절 ID
			position	num	* 어절 내 형태소 위치. 예) 1=>첫번째 형태소, 2=>두번째 형태소...
			error_of _learner	arr (obj)	* 학습자 오류 분석 정보
			id	num	* 오류 형태소 ID 문장 내 1부터 시작
			form	str	* 형태소
			label	str	* 오류 주석 태그
			word_id	arr (str)	* 오류가 포함된 어절 ID들 (**오류주석은 1개 어절 이 상에서 발생할 수 있음)
			position	num	* 어절 내 형태소 위치. 예) 1=>첫번째 형태소, 2=>두번째 형태소...

4. 한국어 학습자 말뭉치 교육 및 홍보

4.1. 말뭉치 구축 실무자 교육

말뭉치 구축 실무자 교육은 형태 주석 및 오류 주석, 쓰기 평가 작업자에게 구축 지침, 학습자 말뭉치 구축 지원 도구 사용 방법 등을 교육하여 자료 구축을 원활하게 하는 것을 목표로 한다. 본 사업에서는 2015-2024년까지의 학습자 말뭉치 작업 지침을 토대로 교육 실행 계획을 수립하여 사업 참여자 전원을 대상으로 말뭉치 구축 실무 교육을 실시하였다. 참여 인력 교육 지원 및 기술 지원을 위한 사업 수행 개요는 다음과 같다.

<표 27> 말뭉치 구축 실무자 교육 내용

구분	지침 교육	구축 지원 도구 교육 및 실습
기본 교육(공통)	<ul style="list-style-type: none"> 한국어 학습자 말뭉치의 이해 한국어 학습자 말뭉치 구축 개인정보보호, DB구축 보안 교육 	<ul style="list-style-type: none"> 구축 지원 도구의 체계 및 사용 방법 전반 작업 할당 및 접수 등의 절차 이력 관리
형태 주석 말뭉치 구축자	<ul style="list-style-type: none"> 형태 주석 지침 및 쟁점 형태 주석 검수 지침 및 쟁점 	<ul style="list-style-type: none"> 형태 주석 방법 및 절차 형태 주석 검수 방법 및 절차
오류 주석 말뭉치 구축자	<ul style="list-style-type: none"> 오류 주석 지침 및 쟁점 오류 주석 검수 지침 및 쟁점 	<ul style="list-style-type: none"> 오류 주석 방법 및 절차 오류 주석 검수 방법 및 절차
쓰기 평가 작업자	<ul style="list-style-type: none"> 작문 평가 기준 및 지침 작문 평가 검수 방법 및 쟁점 	<ul style="list-style-type: none"> 작문 평가 도구 사용법 작문 평가 도구 제약 사항

교육은 정기 워크숍과 세미나 형식으로 진행되었으며, 격주 간격으로 진행되는 실무 회의를 통해 작업에서 해결하지 못하는 다양한 쟁점에 대한 논의들이 이루어졌다. 또한 원활한 작업 진행을 위해 온라인 매체를 통해 작업자들이 실시간으로 질문을 할 수 있도록 하고 즉각적으로 피드백을 제공하였다.

- 정기 워크숍: 작업 단계별 지침 및 도구 사용 교육, 실습

- 정기 세미나: 작업 단계별 쟁점 발표와 토론
- 수시 피드백: 메일, 커뮤니티, 구축 지원 도구의 피드백 기능 등을 통해 업데이트되는 지침 즉시 교육
- 실무 회의: 작업 관련 쟁점 논의

4.2. 말뭉치 이용자를 위한 연수회 개최

4.2.1. 개최 내용

본 사업에서는 한국어 학습자 말뭉치 이용자를 대상으로 총 4회의 ‘한국어 학습자 말뭉치 교실’을 개최하였다. 이는 한국어 학습자 말뭉치의 확산과 이용자의 활용 능력을 제고하기 위한 것으로, 교육 프로그램은 한국어교육 학계 및 교육 분야 외에도 민간 분야에서도 폭넓게 활용할 수 있도록 다채롭게 구성하였다. 회차별 세부 프로그램은 다음과 같다.

<표 28> 한국어 학습자 말뭉치 교실 개최 프로그램

회차	일시 및 장소	프로그램	과정	주요 대상	참가 인원
1차	8월 25일(월) 오전 10-12시 서울 연세대	학습자 말뭉치를 활용한 한국어 교수·학습 자료 개발의 실제(장채린, 명지대) - 한국어 교수 자료 개발을 위한<한국어 학습자 말뭉치 나눔터> 자료 검색 - 한국어 학습자 오류 말뭉치를 활용한 교수 자료 개발	기초 과정	한국어 교원, 학습자	<ul style="list-style-type: none"> • 신청: 대면 60명/비대면 258명 • 참가: 대면 73명/비대면 203명
2차	9월 26일(금) 오후 2-4시 대전 배재대	한국어 학습자 말뭉치를 활용한 교육 콘텐츠 개발 현황과 전망 - 한국어 학습자 말뭉치 활용한 교육 콘텐츠 및 기술 개발 현황(곽용진, ㈜이르테크) - 인공지능과 한국어 학습자	심화 과정	산업계 및 학계의 연구자, 한국어 교원	<ul style="list-style-type: none"> • 신청: 대면 47명/비대면 306명 • 참가: 대면 40명/비대면 110명

회차	일시 및 장소	프로그램	과정	주요 대상	참가 인원
		말뭉치의 연계 가능성과 전망(황금하, ETRI)			
3차	10월 25일(토) 오전 10-12시 부산 부산외대	말뭉치 분석 애플리케이션을 활용한 한국어 학습자 말뭉치 자료 전 처리와 분석의 기초(최정도, 계명대) - 에디터(Editor)를 활용한 전처리	기초 과정	연구자	<ul style="list-style-type: none"> • 신청: 대면 22명/비대면 298명 • 참가: 대면 20명/비대면 133명
4차	11월 28일(금) 오후 3-5시 광주 호남대	한국어 학습자 말뭉치를 활용한 연구 주제 탐색과 적용(김한샘, 연세대) - 앤트콘크(AntConc)의 기초 - 앤트콘크(AntConc)를 활용한 자료 분석	기초 과정	연구자	<ul style="list-style-type: none"> • 신청: 대면 20명/비대면 240명 • 참가: 대면 15명/비대면 151명

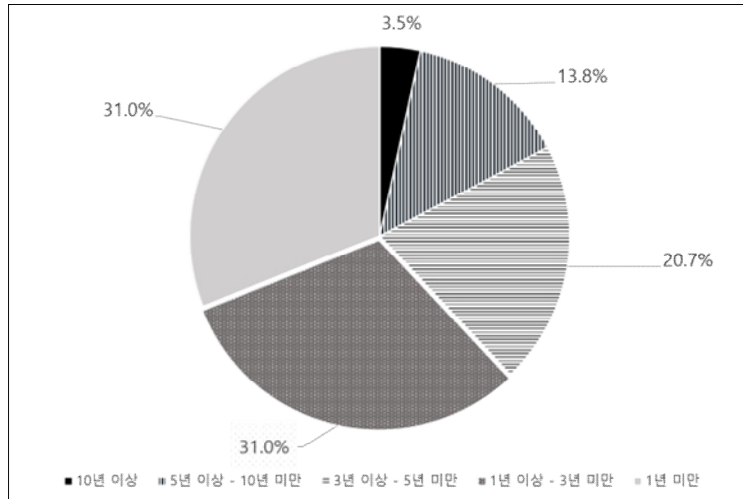
이 중 3차, 4차 학습자 말뭉치 교실에서는 한국어 교수 현장에서의 학습자 말뭉치 활용도를 제고하기 위하여 공동 주최 기관인 부산외국어대학교, 호남대학교의 한국어 교육 기관 소속 교원을 대상으로 사전 설문조사를 실시한 후 간담회를 통해 한국어 학습자 말뭉치에 대한 의견을 수렴하였다. 또한 한국어 학습자 말뭉치 교실이 끝난 후에는 참가자를 대상으로 하여 만족도 조사를 실시하였다.

4.2.2. 한국어 교육 소속 기관 교원 대상 의견수렴 결과

I. 기본 정보

① 한국어 교육 경력

- 만족도 조사에 참여한 전체 응답자의 수는 총 29명이다. 그중 한국어 교육 경력이 10년 이상이라고 밝힌 참여자는 1명, 5년 이상~10년 미만이라고 밝힌 참여자는 4명, 3년 이상~5년 미만이라고 밝힌 참여자는 6명, 1년 이상~3년 미만이라고 밝힌 참여자는 9명, 그리고 1년 미만이라고 밝힌 참여자도 9명이었다.



<그림 13> 한국어 교육 경력 분포

<표 29> 한국어 교육 경력 빈도

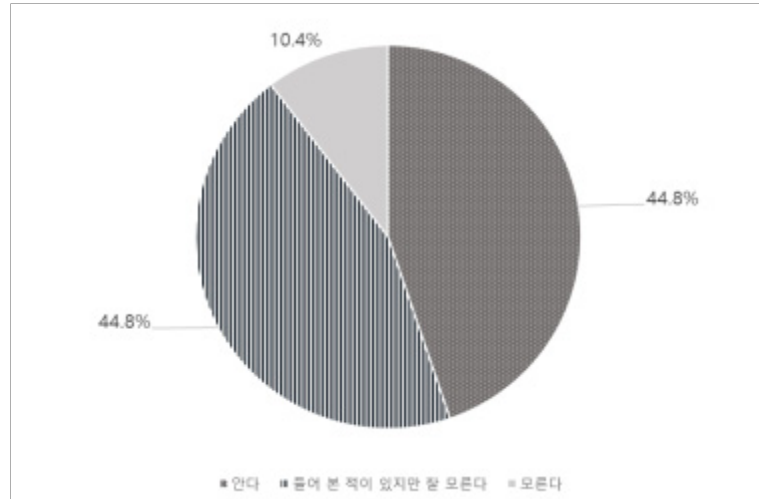
한국어 교육 경력	빈도
10년 이상	1 (3.5%)
5년 이상 - 10년 미만	4 (13.8%)
3년 이상 - 5년 미만	6 (20.7%)
1년 이상 - 3년 미만	9 (31.0%)
1년 미만	9 (31.0%)
합계	29 (100.0%)

Ⅱ. 학습자 말뭉치에 대한 이해도

① 학습자 말뭉치가 무엇인지 알고 있다.

○ 학습자 말뭉치가 무엇인지 알고 있는가에 대한 질문에 ‘안다’라고 응답한 참가자와 ‘들어 본 적이 있지만 잘 모른다’라고 대답한 응답이 각각 13명으로 동일하게 나타났으며, 두 응답을 합한 비율이 89.6%에 달한다. 이를 통해 설문에 참여한 대다수의 교원들이 학습자 말뭉치의 존재에 대해 인지하고 있다는 점을 확인할 수 있다. 그다음으로 ‘모른다’라고 답변한 응답은 3명으

로 10.4%의 비율을 보였다.



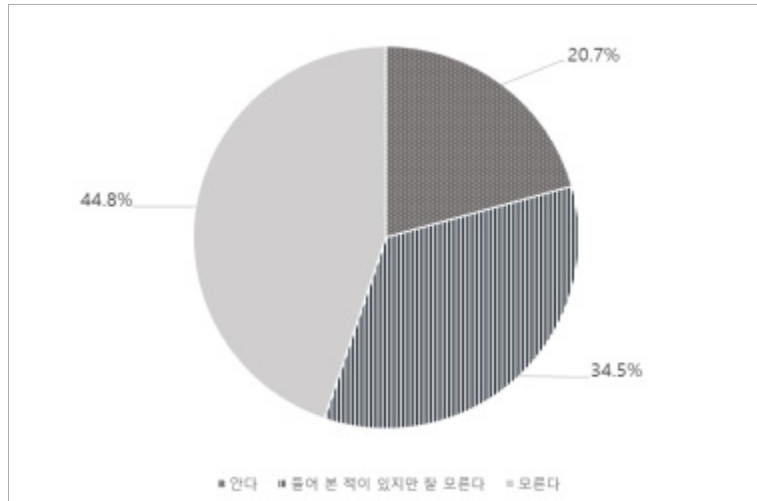
<그림 14> 학습자 말뭉치의 인지 여부 분포

<표 30> 학습자 말뭉치의 인지 여부 빈도

응답	빈도
안다	13 (44.8%)
들어 본 적이 있지만 잘 모른다	13 (44.8%)
모른다	3 (10.4%)
합계	29 (100.0%)

㉔ 원시 말뭉치가 무엇인지 알고 있다.

○ 원시 말뭉치가 무엇인지 알고 있는가에 대한 질문에 ‘안다’라고 응답한 참가자는 총 6명으로 20.7%를 차지한다. 그다음으로 ‘들어 본 적이 있지만 잘 모른다’라고 답변한 응답은 10명으로 34.5%, ‘모른다’라고 답변한 응답은 13명으로 44.8%의 가장 높은 비율을 보였다. 이에 따라 원시 말뭉치에 대하여 대체로 알지 못하며, 절반은 알고 있거나 알지만 잘 모르는 경우로 나뉜다는 것을 확인할 수 있다.



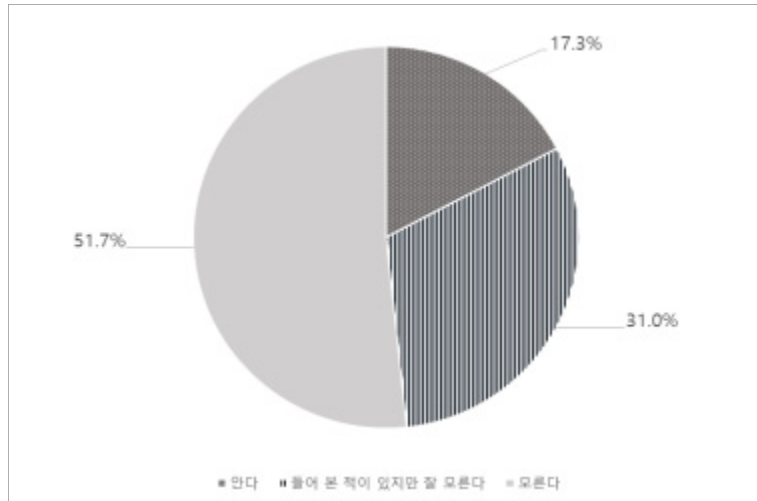
<그림 15> 원시 말뚝치의 인지 여부 분포

<표 31> 원시 말뚝치의 인지 여부 빈도

응답	빈도
안다	6 (20.7%)
들어 본 적이 있지만 잘 모른다	10 (34.5%)
모른다	13 (44.8%)
합계	29 (100.0%)

③ 형태 주석 말뚝치가 무엇인지 알고 있다.

- 형태 주석 말뚝치가 무엇인지 알고 있는가에 대한 질문에 ‘안다’라고 응답한 참가자는 총 5명으로 17.3%, 그다음으로 ‘들어 본 적이 있지만 잘 모른다’라고 답변한 응답은 9명으로 31.0%를 차지한다. ‘모른다’라고 답변한 응답은 15명으로 51.7%의 절반 이상의 비율을 보였다. 이를 통해 절반 이상의 응답자가 형태 주석 말뚝치의 존재를 인지하지 못하고 있음을 확인할 수 있다.



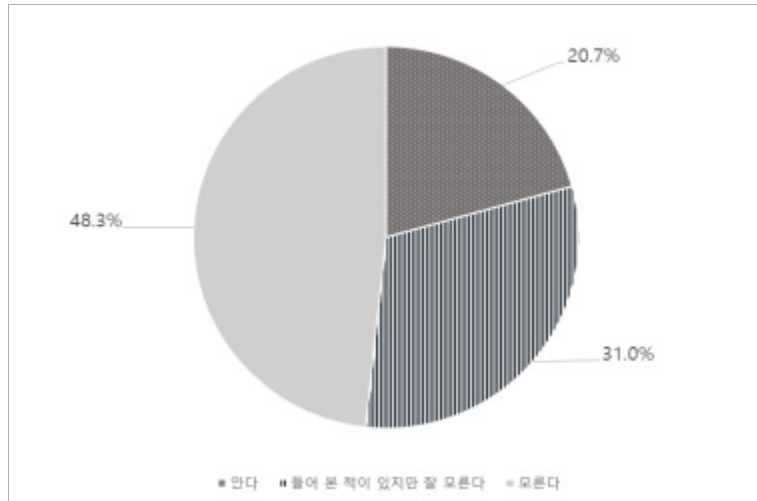
<그림 16> 형태 주석 말뭉치의 인지 여부 분포

<표 32> 형태 주석 말뭉치의 인지 여부 빈도

응답	빈도
안다	5 (17.3%)
들어 본 적이 있지만 잘 모른다	9 (31.0%)
모른다	15 (51.7%)
합계	29 (100.0%)

④ 오류 주석 말뭉치가 무엇인지 알고 있다.

- 오류 주석 말뭉치가 무엇인지 알고 있는가에 대한 질문에 ‘안다’라고 응답한 참가자는 총 6명으로 20.7%를 차지한다. 그다음으로 ‘들어 본 적이 있지만 잘 모른다’라고 답변한 응답은 9명으로 31.0%, ‘모른다’라고 답변한 응답은 14명으로 48.3%의 비율을 보였다.



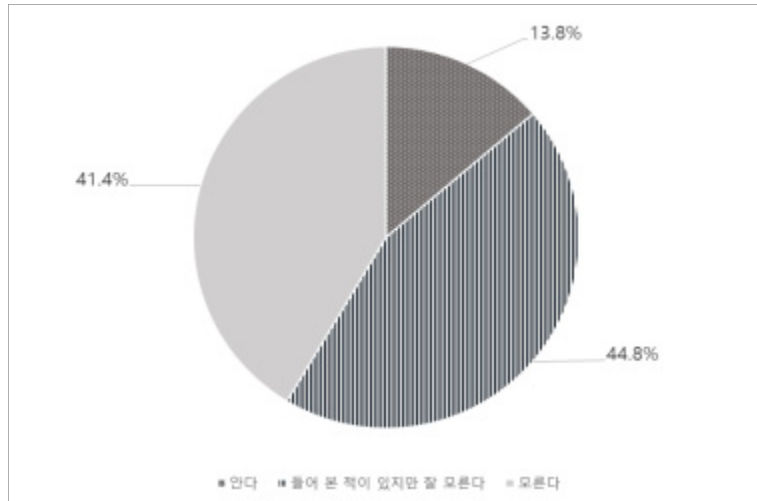
<그림 17> 오류 주석 말뭉치의 인지 여부 분포

<표 33> 오류 주석 말뭉치의 인지 여부 빈도

응답	빈도
안다	6 (20.7%)
들어 본 적이 있지만 잘 모른다	9 (31.0%)
모른다	14 (48.3%)
합계	29 (100.0%)

㉔ 학습자 말뭉치 수집 및 구축 절차를 알고 있다.

- 학습자 말뭉치 수집 및 구축 절차를 알고 있는가에 대한 질문에 ‘안다’라고 응답한 비율은 총 4명으로 13.8%를 차지한다. ‘들어 본 적이 있지만 잘 모른다’와 ‘모른다’라고 응답한 비율은 각각 44.8%와 41.4%로 비슷하게 나타났다. 이를 통해 학습자 말뭉치 수집 및 구축 절차에 대하여 들어 본 적은 있지만 대체로 잘 알지 못하는 경우가 가장 많다는 사실을 확인할 수 있다.



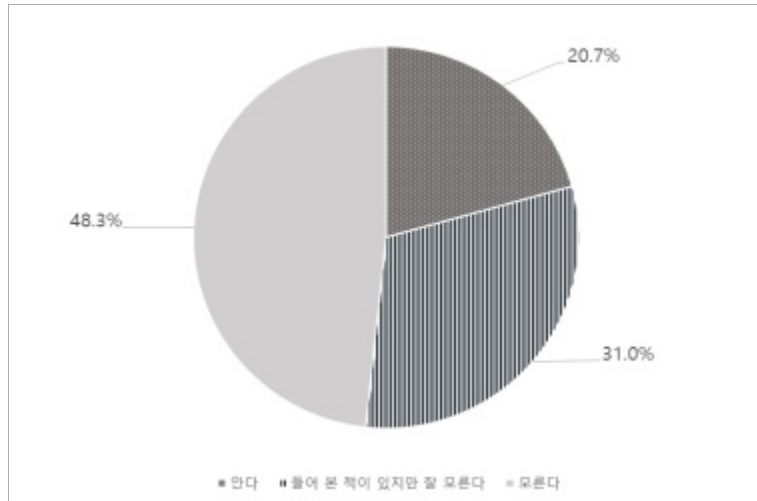
<그림 18> 말뭉치 수집 및 구축 절차 인지 여부 분포

<표 34> 말뭉치 수집 및 구축 절차 인지 여부 빈도

응답	빈도
안다	4 (13.8%)
들어 본 적이 있지만 잘 모른다	13 (44.8%)
모른다	12 (41.4%)
합계	29 (100.0%)

㉔ 학습자 말뭉치의 활용처를 알고 있다.

- 학습자 말뭉치의 활용처를 알고 있는가에 대한 질문에 ‘안다’라고 응답한 참가자는 총 6명으로 20.7%를 차지한다. 그다음으로 ‘들어 본 적이 있지만 잘 모른다’라고 답변한 응답은 9명으로 31.0%, ‘모른다’라고 답변한 응답은 14명으로 48.3%의 비율을 보였다.



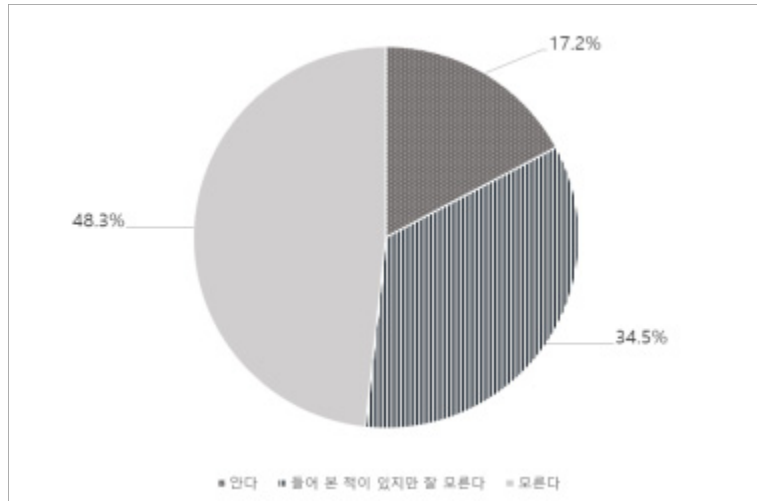
<그림 19> 학습자 말뭉치의 활용처 인지 여부 분포

<표 35> 학습자 말뭉치의 활용처 인지 여부 빈도

응답	빈도
안다	6 (20.7%)
들어 본 적이 있지만 잘 모른다	9 (31.0%)
모른다	14 (48.3%)
합계	29 (100.0%)

㉓ 학습자 말뭉치의 활용 방법을 알고 있다.

- 학습자 말뭉치의 활용 방법을 알고 있는가에 대한 질문에 ‘안다’라고 응답한 참가자는 총 5명으로 17.2%의 낮은 비율을 차지한다. 그다음으로 ‘들어 본 적이 있지만 잘 모른다’라고 답변한 응답은 10명으로 34.5%, ‘모른다’라고 답변한 응답은 14명으로 48.3%의 가장 높은 비율을 보였다.



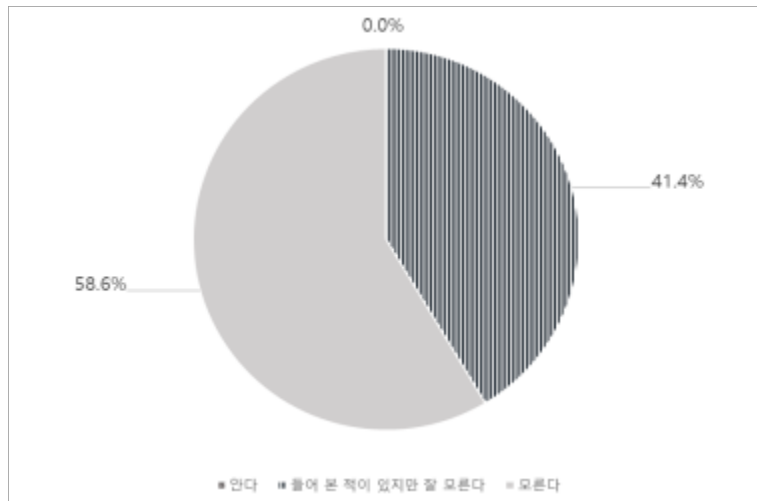
<그림 20> 학습자 말뚝치의 활용 방법 인지 여부 분포

<표 36> 학습자 말뚝치의 활용 방법 인지 여부 빈도

응답	빈도
안다	5 (17.2%)
들어 본 적이 있지만 잘 모른다	10 (34.5%)
모른다	14 (48.3%)
합계	29 (100.0%)

⑧ 학습자 말뚝치의 활용 도구 사용 방법을 알고 있다.

- 학습자 말뚝치의 활용 도구 사용 방법을 알고 있는가에 대한 질문에 ‘안다’라고 답변한 응답은 없었다. 반면 ‘모른다’라고 응답한 비율은 총 17명의 58.6%로 가장 높게 나타났으며, ‘들어 본 적이 있지만 잘 모른다’로 응답한 비율은 12명의 41.4%로 나타났다. 이러한 결과를 통해 학습자 말뚝치의 활용 도구 사용 방법에 대한 인지가 다른 항목에 비해 가장 낮은 것을 확인할 수 있다.



<그림 21> 활용 도구 사용 방법 인지 여부 분포

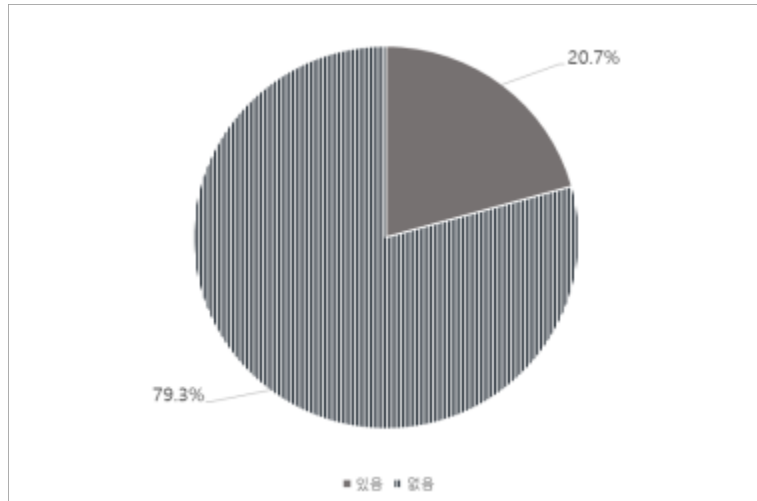
<표 37> 활용 도구 사용 방법 인지 여부 빈도

응답	빈도
안다	0 (0.0%)
들어 본 적이 있지만 잘 모른다	12 (41.4%)
모른다	17 (58.6%)
합계	29 (100.0%)

Ⅲ. 학습자 말뭉치에 대한 이해도

① 한국어 학습자 말뭉치 활용 경험

- 한국어 학습자 말뭉치의 활용 경험에 관한 질문에 대해 활용 경험이 ‘없다’고 대답한 응답이 총 23명으로 79.3%를 차지한다. 반면 활용 경험이 ‘있다’고 대답한 응답은 6명의 20.7%로 나타났다.



<그림 22> 한국어 학습자 말뭉치 활용 경험 분포

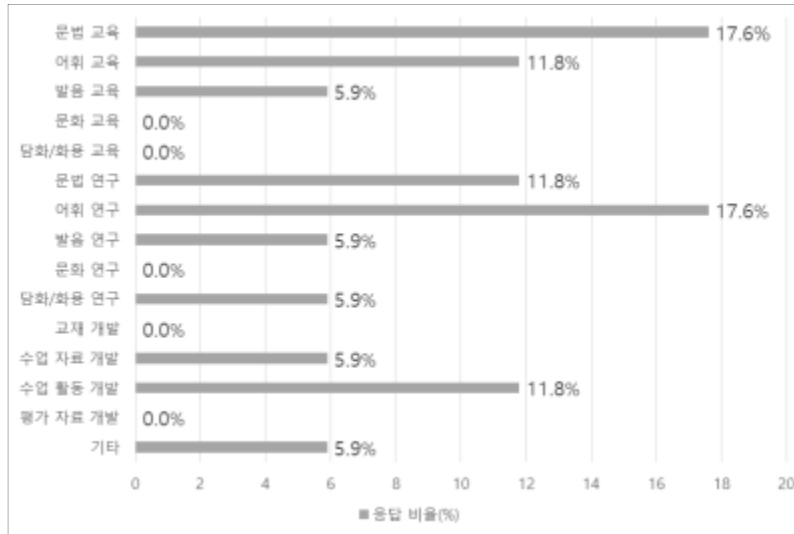
<표 38> 한국어 학습자 말뭉치 활용 경험 빈도

응답	빈도
있음	6 (20.7%)
없음	23 (79.3%)
합계	29 (100.0%)

② 한국어 학습자 말뭉치 활용 목적

- 한국어 학습자 말뭉치의 활용 경험이 있다고 답변한 응답자들을 대상으로 어떤 목적으로 활용하였는지에 대해 질문하였다.¹⁰⁾ 답항의 구성은 크게 교육 분야(문법, 어휘, 발음, 문화, 담화/화용), 연구 분야(문법, 어휘, 발음, 문화, 담화/화용), 개발 분야(교재, 수업 자료, 수업 활동, 평가 자료)로 나뉘었으며, 문법 교육과 어휘 연구가 각각 17.6%로 가장 많이 활용되는 것으로 나타났다. 다음으로 어휘 교육, 문법 연구, 수업 활동 개발이 각각 11.8%로 뒤를 이었으며, 발음 교육, 발음 연구, 담화/화용 연구, 수업 자료 개발도 각각 5.9%로 다양한 분야에서 활용이 되고 있음을 알 수 있었다.

10) 해당 문항은 중복 응답이 가능한 문항으로 응답 수의 합이 29와 일치하지 않는다.



<그림 23> 한국어 학습자 말뭉치 활용 목적 분포

<표 39> 한국어 학습자 말뭉치 활용 목적 빈도

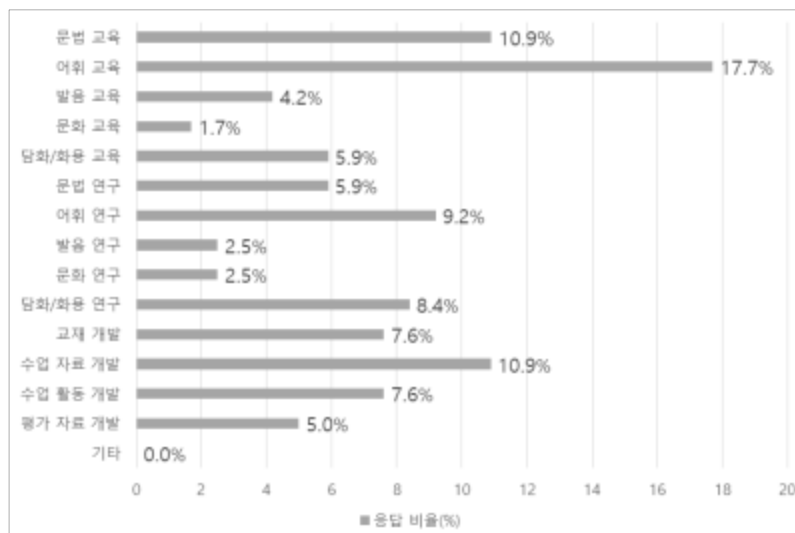
응답	빈도
문법 교육	3 (17.6%)
어휘 교육	2 (11.8%)
발음 교육	1 (5.9%)
문화 교육	0 (0.0%)
담화/화용 교육	0 (0.0%)
문법 연구	2 (11.8%)
어휘 연구	3 (17.6%)
발음 연구	1 (5.9%)
문화 연구	0 (0.0%)
담화/화용 연구	1 (5.9%)
교재 개발	0 (0.0%)
수업 자료 개발	1 (5.9%)
수업 활동 개발	2 (11.8%)
평가 자료 개발	0 (0.0%)

기타	1 (5.9%)
합계	17 (100.0%)

- 한국어 학습자 말뭉치의 활용 목적으로 가장 높은 비율을 차지한 응답은 17.6%를 차지한 문법 교육과 어휘 연구이다. 그 뒤를 이어 어휘 교육, 문법 연구, 수업 활동이 각 11.8%로 비교적 높은 비율을 차지했다. 그다음으로 발음 교육, 발음 연구, 담화/화용 연구, 수업 자료 개발이 각 5.9%로 나타났다.¹¹⁾

③ 한국어 학습자 말뭉치 향후 활용 목적

- 앞으로 한국어 학습자 말뭉치를 활용한다면 어떤 목적으로 활용하고 싶은지에 대해 질문하였다.¹²⁾ 해당 문항의 구성은 앞선 질문의 항목과 동일하며, 크게 교육 분야(문법, 어휘, 발음, 문화, 담화/화용), 연구 분야(문법, 어휘, 발음, 문화, 담화/화용), 개발 분야(교재, 수업 자료, 수업 활동, 평가 자료)로 나뉜다. 응답 결과에 따르면 어휘 교육이 17.7%, 문법 교육과 수업 자료 개발이 10.9%, 교재 개발과 수업 활동 개발이 각각 7.6%로 교육 분야에서 활용을 희망하는 이용자가 많은 것으로 나타났다. 연구의 경우는 9.2%로 어휘 연구에 활용하고자 하는 이용자의 비중이 높은 것으로 나타났다.



<그림 24> 한국어 학습자 말뭉치 향후 활용 목적 분포

11) 주관식 응답인 '기타'로 선택된 1개 응답은 공란으로 제출되었다.

12) 해당 문항은 중복 응답이 가능한 문항으로 응답 수의 합이 29와 일치하지 않는다.

<표 40> 한국어 학습자 말뭉치 향후 활용 목적 빈도

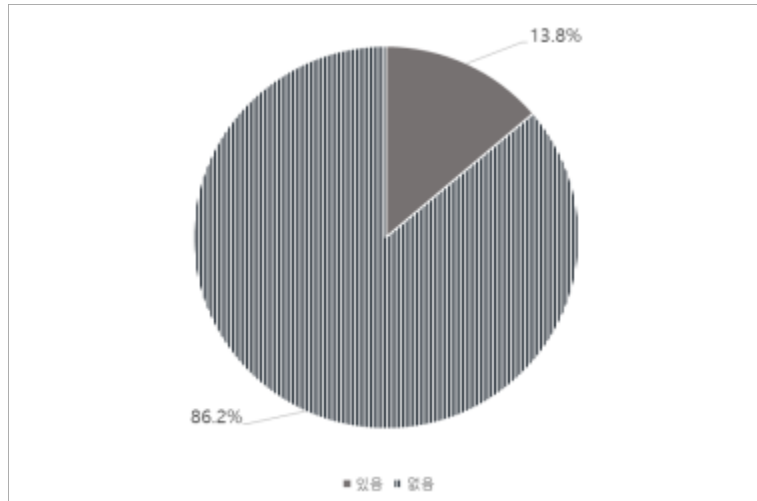
응답	빈도
문법 교육	13 (10.9%)
어휘 교육	21 (17.7%)
발음 교육	5 (4.2%)
문화 교육	2 (1.7%)
담화/화용 교육	7 (5.9%)
문법 연구	7 (5.9%)
어휘 연구	11 (9.2%)
발음 연구	3 (2.5%)
문화 연구	3 (2.5%)
담화/화용 연구	10 (8.4%)
교재 개발	9 (7.6%)
수업 자료 개발	13 (10.9%)
수업 활동 개발	9 (7.6%)
평가 자료 개발	6 (5.0%)
기타	0 (0.0%)
합계	119 (100.0%)

- 한국어 학습자 말뭉치의 향후 활용 목적으로 가장 높은 비율을 차지한 응답은 17.7%의 어휘 교육이다. 그 뒤를 이어 문법 교육과 수업 자료 개발이 각 10.9%, 어휘 연구가 9.2%, 담화/화용 연구가 8.4%, 그리고 교재 개발과 수업 활동 개발이 각 7.6%를 차지했다. 이를 통해 어느 한 분야에 치우치기보다 교육, 연구, 개발의 다양한 분야에서의 활용 기대가 있다는 점을 확인할 수 있다. 그중에서도 어휘 교육과 어휘 연구 분야에서의 학습자 말뭉치 활용에 대한 기대가 높은 것을 파악할 수 있다. 반면, 문화 교육이나 문화 연구에 대한 분야에서의 학습자 말뭉치 활용의 기대가 상대적으로 낮다고 볼 수 있다.

IV. 한국어 학습자 말뭉치 교실 참가 경험

① 한국어 학습자 말뭉치 교실 참가 경험

- 한국어 학습자 말뭉치 교실 참가 경험에 관한 질문에 대해 참가 경험이 ‘없다’고 대답한 응답이 총 25명으로 86.2%를 차지한다. 반면 참가 경험이 ‘있다’고 대답한 응답은 4명의 13.8%로 나타났다.



<그림 25> 한국어 학습자 말뭉치 교실 참가 경험 분포

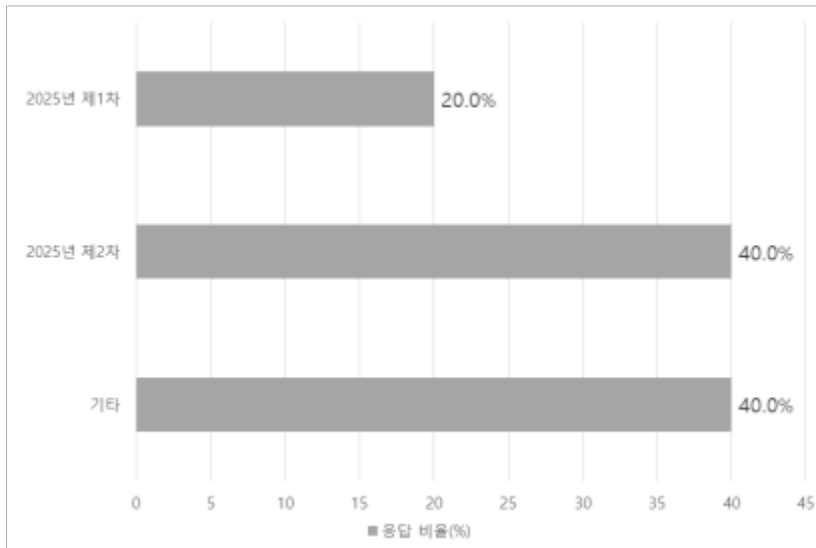
<표 41> 한국어 학습자 말뭉치 교실 참가 경험 빈도

응답	빈도
있음	4 (13.8%)
없음	25 (86.2%)
합계	29

② 한국어 학습자 말뭉치 참가 프로그램

- 한국어 학습자 말뭉치 교실의 참가 경험이 있다고 답변한 응답자들을 대상으로 어떤 프로그램에 참가하였는지에 대해 주관식으로 질문하였다.¹³⁾ 이에 대한 응답은 다음의 <그림 26>과 <표 42>에서 확인할 수 있다.¹⁴⁾

13) 이 문항은 주관식으로 자유 답변이 가능하여 응답 수의 합이 29와 일치하지 않는다.



<그림 26> 한국어 학습자 말뭉치 교실 참가 프로그램 분포

<표 42> 한국어 학습자 말뭉치 교실 참가 프로그램 빈도

응답	빈도
2025 제1차 한국어 학습자 말뭉치 교실	1 (20.0%)
2025 제2차 한국어 학습자 말뭉치 교실	2 (40.0%)
기타	2 (40.0%)
합계	5 (100.0%)

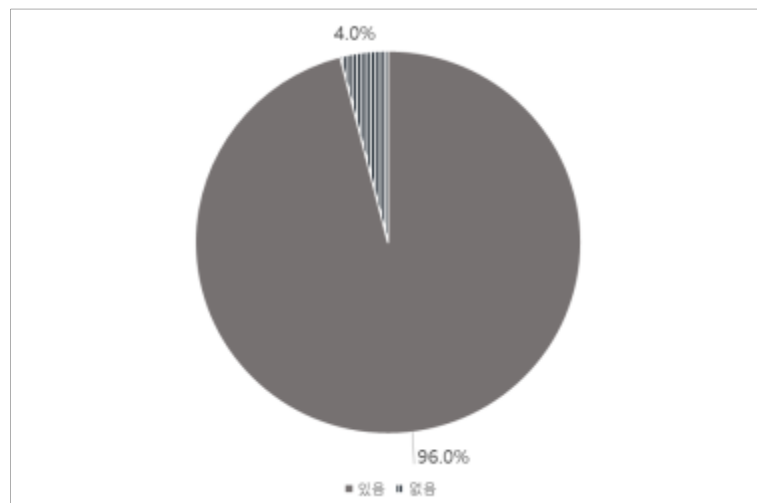
- 한국어 학습자 말뭉치 교실 프로그램 중 2025년 제1차 한국어 학습자 말뭉치 교실에 참여한 1명의 응답이 20%를 차지하였고, 제2차 한국어 학습자 말뭉치 교실에 참여한 2명의 응답이 40%를 차지하였다.¹⁵⁾ 기타 응답으로는 ‘기억이 나지 않는다’라고 응답한 답변이 2명 존재한다.

14) <그림 14>, <표 14>에서는 주관식 답변의 유형을 분류하여 제시하였다.

15) 참가 경험이 없다고 답변한 응답자는 <표 13>에서 확인할 수 있듯이 총 4명이다. <표 14>에서 합계가 5명으로 나온 것은 행사에 중복 참여한 응답자가 있기 때문이다. 제2차 행사를 참여하였다고 답변한 응답자 중 1명은 제1차 행사에 참여한 응답자로 총 2번 참여한 것으로 확인된다.

③ 한국어 교원을 위한 교육 프로그램 참가 의향 여부

- 한국어 학습자 말뭉치 교실의 참가 경험이 없다고 답변한 응답자들을 대상으로 한국어 교원을 위한 교육 프로그램이 있다면 참가할 의향이 있는지에 대해 질문하였다. 이에 의향이 '있다'고 대답한 응답이 총 24명으로 96.0%를 차지한다. 반면 **의향이** '없다'고 대답한 응답은 1명의 4.0%라는 아주 낮은 비율을 보였다. 이에 따라 한국어 교원을 위한 교육 프로그램의 수요가 매우 높다는 사실을 확인할 수 있다.



<그림 27> 교원을 위한 교육 프로그램 참여 의향 분포

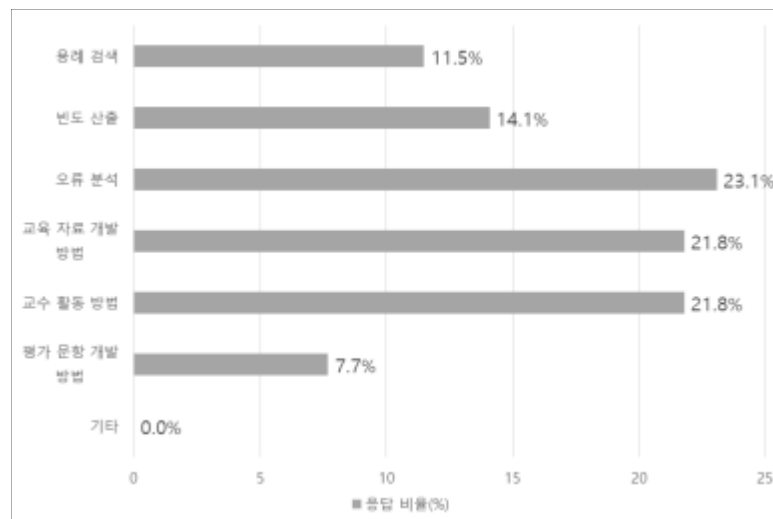
<표 43> 교원을 위한 교육 프로그램 참여 의향 빈도

응답	빈도
있음	24 (96.0%)
없음	1 (4.0%)
합계	25 (100.0%)

V. 학습자 말뭉치 교실에 대한 요구

① 한국어 교수 현장에서의 활용을 위해 필요한 기술

- 한국어 교수 현장에서의 활용을 위해 꼭 필요한 기술이 무엇이라고 생각하는지에 대해 질문하였다.¹⁶⁾ 이에 대하여 응답자의 23.1%가 오류 분석이라고 답하였다. 교육 자료 개발 방법과 교수 활동 방법이 각각 21.8%, 빈도 산출이 14.1%, 용례 검색이 11.5%, 평가 문항 개발 방법이 7.7%로, 연구 또는 교육을 위한 다양한 기술에 대해 관심이 있음을 확인할 수 있었다. 이에 대한 응답은 다음의 <그림 28>과 <표 44>에서 확인할 수 있다.



<그림 28> 한국어 교수 현장에서의 활용을 위한 기술 분포

<표 44> 한국어 교수 현장에서의 활용을 위한 기술 빈도

응답	빈도
용례 검색	9 (11.5%)
빈도 산출	11 (14.1%)
오류 분석	18 (23.1%)
말뭉치를 활용한 교육 자료 개발 방법	17 (21.8%)

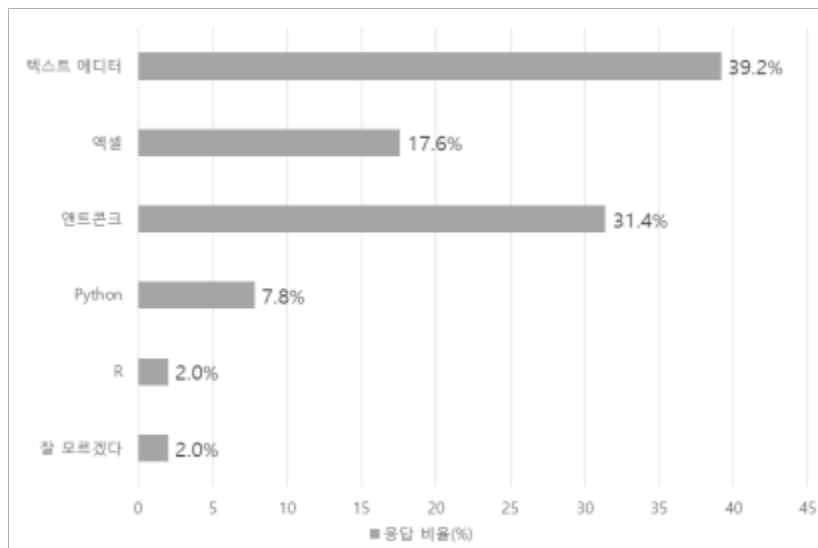
16) 해당 문항은 중복 응답이 가능한 문항으로 응답 수의 합이 29와 일치하지 않는다.

말뭉치를 활용한 교수 활동 방법	17 (21.8%)
말뭉치를 활용한 평가 문항 개발 방법	6 (7.7%)
기타	0 (0.0%)
합계	78 (100.0%)

- 한국어 교수 현장에서의 활용을 위해 꼭 필요한 기술로 가장 높은 비율을 차지한 응답은 23.1%의 오류 분석이다. 그 뒤를 이어 말뭉치를 활용한 교육 자료 개발 방법과 교수 활동 방법이 각 21.8%를 차지했다. 그다음 빈도 산출이 14.1%, 용례 검색이 11.5%, 말뭉치를 활용한 평가 문항 개발 방법이 7.7%를 차지했다.

② 한국어 교수 현장에서의 활용을 위해 필요한 도구

- 한국어 교수 현장에서의 활용을 위해 꼭 필요한 도구는 무엇이라고 생각하는지를 묻는 문항에 대해서는¹⁷⁾ 텍스트 에디터가 39.2%로 한국어 학습자 말뭉치를 활용하는 데 가장 중요한 도구라고 인식하고 있음을 알 수 있었다. 다음으로 엔트콘크가 31.4%, 엑셀이 17.6%로 높은 비중을 차지하였으며, Python과 R은 각각 7.8%, 2.0%로 덜 중요한 도구로 인식하고 있었다.



<그림 29> 한국어 교수 현장에서의 활용을 위한 필요 도구 분포

17) 해당 문항은 중복 응답이 가능한 문항으로 응답 수의 합이 29와 일치하지 않는다.

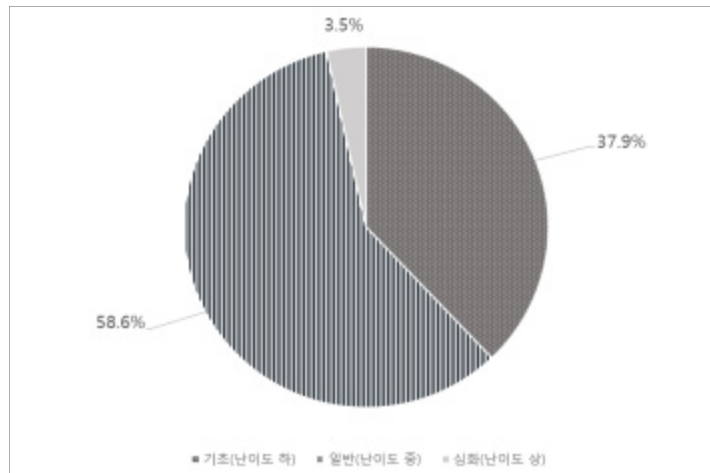
<표 45> 한국어 교수 현장에서의 활용을 위한 필요 도구 빈도

응답	빈도
텍스트 에디터	20 (39.2%)
엑셀	9 (17.6%)
애플콘크	16 (31.4%)
Python	4 (7.8%)
R	1 (2.0%)
잘 모르겠다	1 (2.0%)
합계	51 (100.0%)

- 한국어 교수 현장에서의 활용을 위해 꼭 필요한 도구로 가장 높은 비율을 차지한 응답은 39.2%의 텍스트 에디터다. 그 뒤를 이어 31.4%의 애플콘크가 상대적으로 높은 비율을 차지했다. 그다음으로 엑셀은 17.6%, Python은 7.8%, R은 2.0%로 나타났다. 그리고 ‘잘 모르겠다’고 답한 응답자의 비율도 2%로 나타났다.

③ 교육 프로그램의 수준

- 적절한 교육 프로그램의 수준에 대해 문항을 난이도를 기초(난이도 하), 일반(난이도 중), 심화(난이도 상)로 나누어 질문하였다. 그 결과 일반(난이도 중)이라고 답변한 응답이 총 17명으로 58.6%라는 가장 높은 비율을 보였다. 그다음으로 기초(난이도 하)가 총 11명의 37.9%, 심화(난이도 상)가 총 1명의 3.5%로 나타났다.



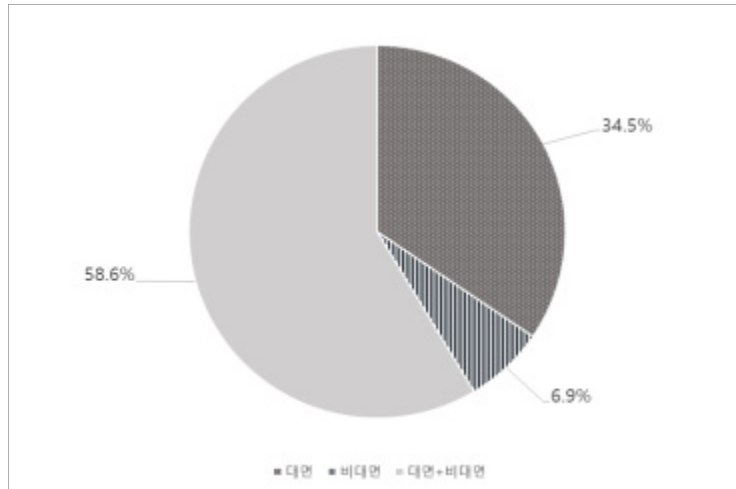
<그림 30> 교육 프로그램 수준 분포

<표 46> 교육 프로그램 수준 빈도

응답	빈도
기초(난이도 하)	11 (37.9%)
일반(난이도 중)	17 (58.6%)
심화(난이도 상)	1 (3.5%)
합계	29 (100.0%)

④ 학습자 맞춤형 교실의 교육 방식

- 학습자 맞춤형 교실은 주로 강의와 실습을 겸하는 프로그램을 운영하고 있다. 이에 학습자 맞춤형 교실에 적합한 교육 방식에 대해 대면, 비대면, 대면+비대면으로 나누어 질문하였다. 그 결과 대면+비대면 방식이 총 17명의 58.6%라는 가장 높은 비율을 보였다. 그다음으로 대면 방식이 총 10명의 34.5%, 비대면 방식이 총 2명의 6.9%로 나타났다.



<그림 31> 학습자 말뚝치 교실 교육 방식 분포

<표 47> 학습자 말뚝치 교실 교육 방식 빈도

응답	빈도
대면	10 (34.5%)
비대면	2 (6.9%)
대면+비대면	17 (58.6%)
합계	29 (100.0%)

5) 학습자 말뚝치 교실에서 배우고 싶은 프로그램 제안

- 한국어 학습자 말뚝치 교실에서 배우고 싶은 프로그램에 대한 주관식 답변으로는 다음과 같은 것들이 있었다.¹⁸⁾

<표 48> 학습자 말뚝치 교실에서 배우고 싶은 프로그램에 대한 의견

의견 유형	내용
말뚝치 활용 도구 및 방법 소개 (6명)	<ul style="list-style-type: none"> 말뚝치를 활용하는 방법 소개 (5명) - 말뚝치 자료 소개, 자료 검색 방법, 활용 방법 탐색 등 말뚝치를 구축하는 방법 소개

18) 해당 문항은 필수 응답 문항이 아닌 선택 문항으로 응답 수의 합이 29와 일치하지 않는다.

말뭉치 활용 사례 소개 (5명)	<ul style="list-style-type: none"> 말뭉치를 연구에 활용한 실제 사례 소개 (4명) 말뭉치를 수업에 활용한 실제 사례 소개
도구를 활용한 말뭉치 활용 실습 (3명)	<ul style="list-style-type: none"> 말뭉치를 활용할 때 사용하는 도구에 대한 교육 (2명) 예) Python, AntConc, Excel, Kiwi, Text editor 등 도구를 활용한 말뭉치 통계에 대한 교육
강의 주제에 대한 의견 (2명)	<ul style="list-style-type: none"> 말뭉치를 이용한 어휘 교육에 관련한 내용 많은 내용을 다루기보다 확실하고 구체적인 활용 방법에 대한 교육 필요

⑥ 학습자 말뭉치 교실 및 학습자 말뭉치 구축에 대한 자유 의견

- 앞에서 언급된 항목 외에 한국어 학습자 말뭉치 교실 및 구축에 대한 자유 의견으로는 다음과 같은 것들이 있었다.¹⁹⁾

<표 49> 학습자 말뭉치 교실 및 구축에 대한 자유 의견

의견 유형	내용
학습자 말뭉치 교실 관련 (6명)	<ul style="list-style-type: none"> 한국어 학습자 말뭉치 교실에 대한 홍보 확대 필요 (2명) 한국어 학습자 말뭉치 교실 운영 횟수 확대 필요 (2명) 한국어 학습자 말뭉치 교실의 언어권별 운영에 대한 고려 필요 학습자의 실제 언어 사용을 이해하고 교육에 반영하는 방향성에 대한 고려 필요 - 단순히 말뭉치 분석 기술만을 배우는 장으로 학습자 말뭉치 교실을 구성하는 것은 한계가 있음
학습자 말뭉치 구축 관련 (3명)	<ul style="list-style-type: none"> 한국어 학습자 말뭉치를 교육적으로 활용할 수 있는 방법에 대한 안내 및 홍보 필요 (2명) AI 시대에 맞는 말뭉치 구축 방향 설정 필요 - AI 시대에서 실효성 있는 학습 자료로 학습자 말뭉치를 사용하기 위함

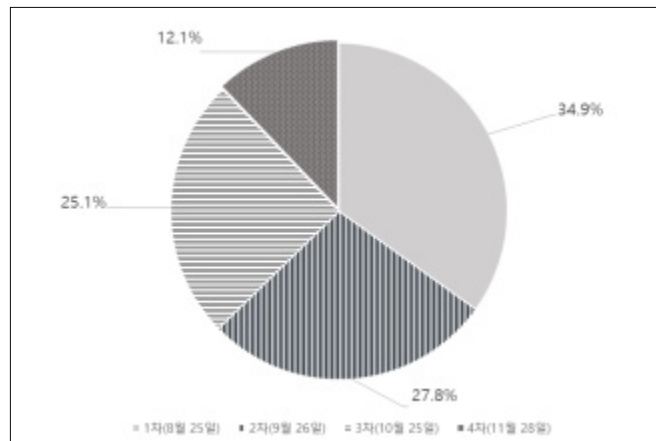
19) 해당 문항은 필수 응답 문항이 아닌 선택 문항으로 응답 수의 합이 29와 일치하지 않는다.

4.2.3. 한국어 학습자 말뭉치 교실 참가자 만족도 조사 결과

I. 기본 정보

① 참석한 한국어 학습자 말뭉치 교실 회차

- 만족도 조사에 참여한 전체 응답자의 수는 총 259명이다. 중복 참여한 인원을 고려하여 한국어 학습자 말뭉치 교실 회차별 응답 참여 인원을 조사한 결과 총 338명으로 확인되었다.²⁰⁾ 회차별 응답자의 수는 118명으로 1차가 가장 많았으며, 그다음으로 2차가 94명, 3차가 85명, 4차가 41명이었다.



<그림 32> 회차별 응답 참여 인원 분포

<표 50> 회차별 응답 참여 인원 빈도

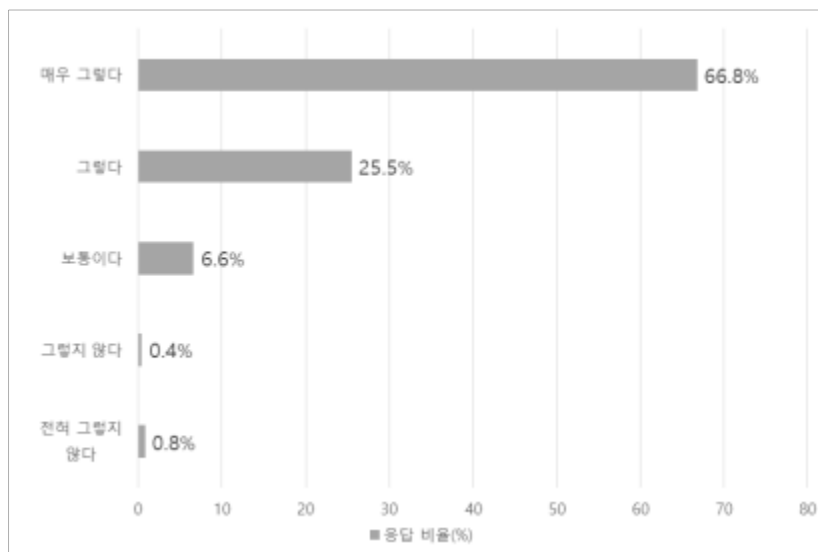
참여 회차	빈도
1차 (8월 25일)	118 (34.9%)
2차 (9월 26일)	94 (27.8%)
3차 (10월 25일)	85 (25.1%)
4차 (11월 28일)	41 (12.1.0%)
합계	338 (100.0%)

20) 1인이 다회차에 참여한 경우가 있어, 만족도 조사에 참여한 인원의 총합인 259명과 아카데미 회차별 응답 참여 인원인 338명이 서로 일치하지 않는다.

Ⅱ. 프로그램에 대한 만족도 조사

① 프로그램 운영 방식(대면+비대면)이 효율적이다

- 2025년 한국어 학습자 말뭉치 교실의 대면과 비대면 혼용 운영 방식이 효율적인지에 대한 질문에 ‘매우 그렇다’라고 답한 응답자는 총 173명(66.8%)으로 가장 많은 응답 수를 보였다. 구체적인 응답 결과는 <그림 33>과 <표 50>에서 확인할 수 있다.



<그림 33> 프로그램 운영 방식에 대한 만족도 분포

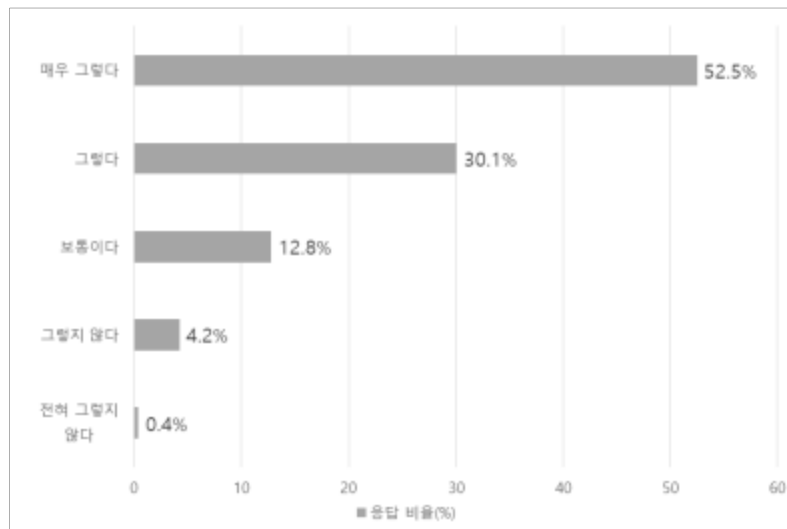
<표 51> 프로그램 운영 방식에 대한 만족도 빈도

만족도	빈도
매우 그렇다	173 (66.8%)
그렇다	66 (25.5%)
보통이다	17 (6.6%)
그렇지 않다	1 (0.4%)
매우 그렇지 않다	2 (0.8%)
합계	259 (100.0%)

- 2025년 한국어 학습자 말뭉치 교실의 대면과 비대면 혼용 운영 방식이 효율적인지에 대한 응답을 보면 ‘매우 그렇다’라는 응답이 66.8%의 가장 높은 비율로 나타났다. 그리고 ‘그렇다’라는 응답의 비율이 25.5%로 그 뒤를 이었다. 이를 통해 90% 이상의 응답자가 운영 방식에 대해 만족하고 있으며, 절반 이상이 ‘매우 그렇다’라고 긍정적으로 답변한 결과를 확인할 수 있다.

② 강의 방식(강의+실습)이 적절하다

- 강의와 실습을 결합한 강의 방식이 적절한지에 대한 질문에 ‘매우 그렇다’라고 답한 응답자는 총 136명으로 가장 많은 응답 수를 보였다. 구체적인 응답 결과는 다음의 <그림 34>와 <표 51>에서 확인할 수 있다.



<그림 34> 강의 방식에 대한 만족도 분포

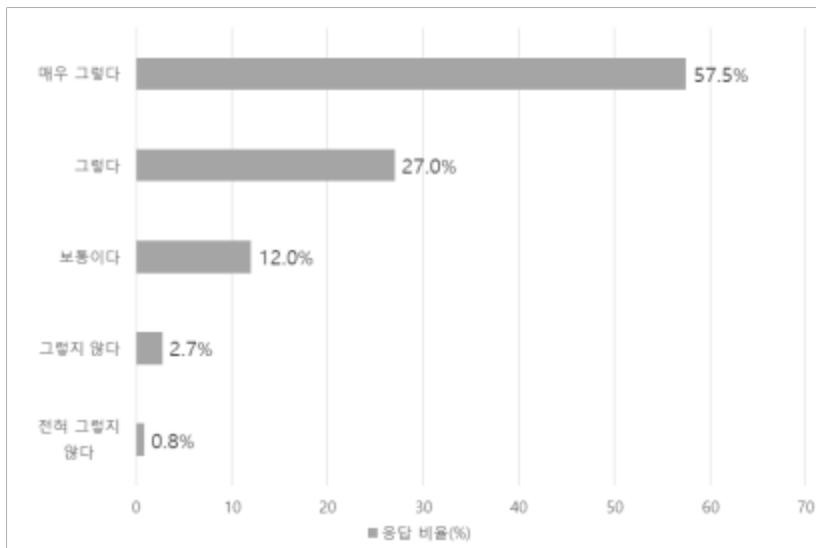
<표 52> 강의 방식에 대한 만족도 빈도

만족도	빈도
매우 그렇다	136 (52.5%)
그렇다	78 (30.1%)
보통이다	33 (12.8%)
그렇지 않다	11 (4.2%)
매우 그렇지 않다	1 (0.4%)
합계	259 (100.0%)

- 2025년 한국어 학습자 말뭉치 교실의 강의 방식은 강의와 실습을 결합한 형태로 이루어졌다. 해당 강의 방식의 적절성에 대한 질문에 ‘매우 그렇다’라는 응답이 52.5%의 가장 높은 비율로 나타났다. 그다음으로 ‘그렇다’라는 답변이 30.1%, ‘보통이다’라는 답변이 12.8%로 확인되었다.

③ 강의 시간이 적절하다

- 강의 시간이 적절한지에 대한 응답에 ‘매우 그렇다’라고 답한 응답자는 총 149명으로 가장 많은 응답 수를 보였다. 구체적인 응답의 결과는 다음의 <그림 35>와 <표 52>에서 확인할 수 있다.



<그림 35> 강의 시간에 대한 만족도 분포

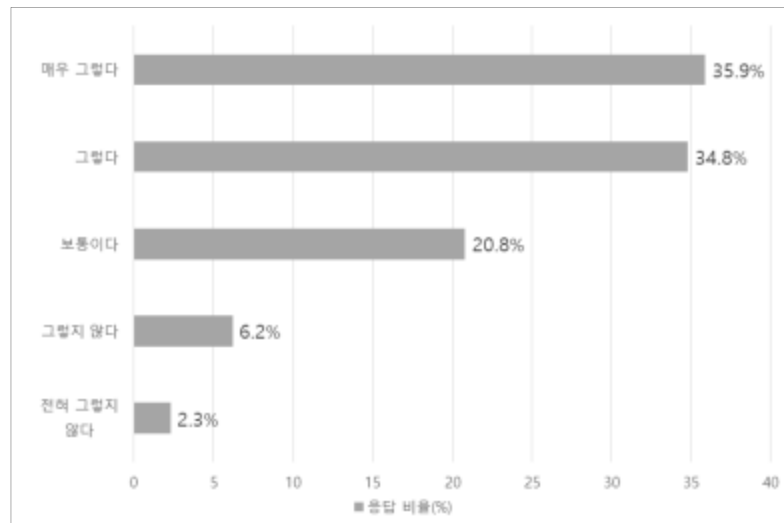
<표 53> 강의 시간에 대한 만족도 빈도

만족도	빈도
매우 그렇다	149 (57.5%)
그렇다	70 (27.0%)
보통이다	31 (12.0%)
그렇지 않다	7 (2.7%)
매우 그렇지 않다	2 (0.8%)
합계	259 (100.0%)

- 매 회차 2시간으로 이루어졌던 강의 시간의 적절성에 대한 질문에 ‘매우 그렇다’라는 응답이 57.5%의 가장 높은 비율로 나타났다. ‘매우 그렇다’ 다음으로 높은 비율을 차지한 항목은 27.0%의 ‘그렇다’로 나타났으며, ‘보통이다’라는 응답은 12.0%로 나타났다.

④ 강의 내용이 이해하기 쉬웠다.

- 강의 내용이 이해하기 쉬웠는지에 대한 질문에는 ‘매우 그렇다’라고 답변한 응답자의 비율이 35.9%로 가장 높았으며, ‘그렇다’가 34.8%, ‘보통이다’가 20.8%로 그 뒤를 이었다. 이를 통해 강의의 내용이 어렵지 않게 구성되고 참여자가 이해하기 쉽게 강의가 진행되었음을 알 수 있었다.



<그림 36> 강의 내용 이해도에 대한 만족도 분포

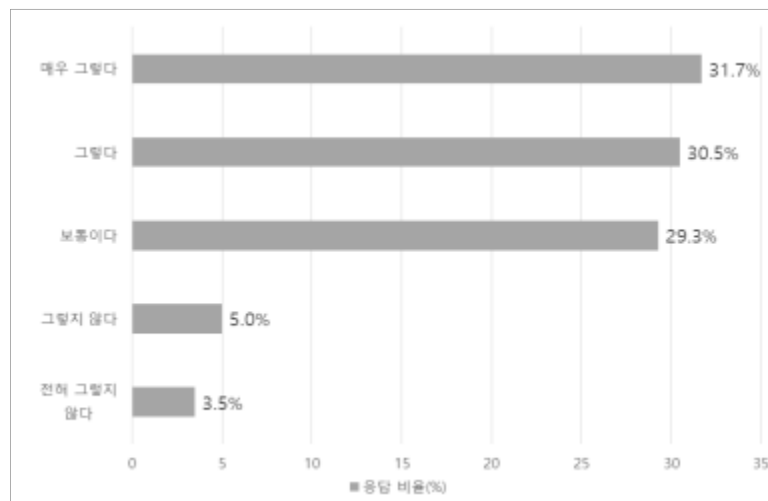
<표 54> 강의 내용 이해도에 대한 만족도 빈도

만족도	빈도
매우 그렇다	93 (35.9%)
그렇다	90 (34.8%)
보통이다	54 (20.8%)
그렇지 않다	16 (6.2%)
매우 그렇지 않다	6 (2.3%)
합계	259 (100.0%)

- 강의 내용의 이해도에 대한 응답을 살펴보면 ‘매우 그렇다’와 ‘그렇다’의 응답이 매우 비슷하게 나타났으며 두 응답의 비율을 합치면 70.7%로 대부분이 강의 내용을 잘 이해함을 알 수 있다. 이어서 ‘보통이다’가 20.8%, ‘그렇지 않다’가 6.2%의 비율을 보였다.

㉔ 실습을 잘 따라할 수 있었다.

- 실습을 잘 따라할 수 있었는지에 대한 질문에는 ‘매우 그렇다’라고 답한 응답자의 비율이 31.7%로 가장 많았고, ‘그렇다’가 30.5%, ‘보통이다’가 29.3%로, 대체로 실습에 참여하는 데에 큰 어려움이 없을 정도의 난이도로 강의가 진행되었음을 확인할 수 있었다.



<그림 37> 실습 수행에 대한 만족도 분포

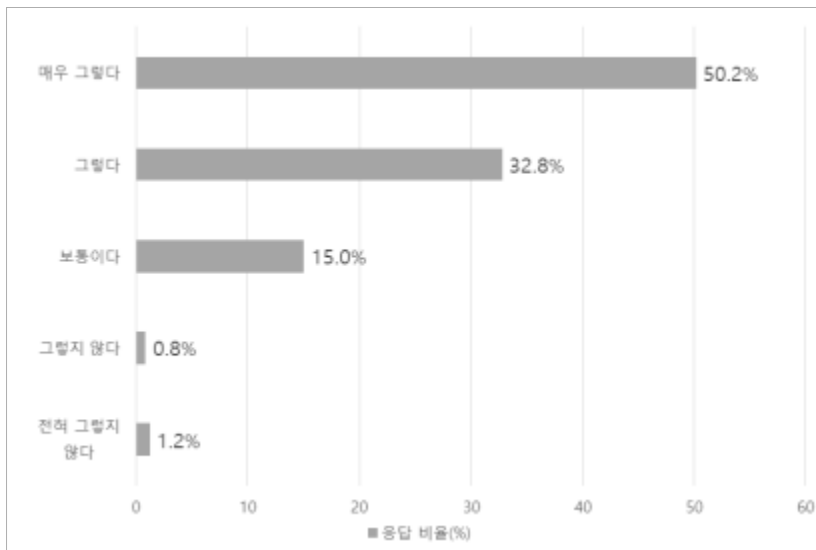
<표 55> 실습 수행에 대한 만족도 빈도

만족도	빈도
매우 그렇다	82 (31.7%)
그렇다	79 (30.5%)
보통이다	76 (29.3%)
그렇지 않다	13 (5.0%)
매우 그렇지 않다	9 (3.5%)
합계	259 (100.0%)

- 실습 수행에 관한 만족도 질문에 대해 ‘매우 그렇다’라고 답변한 응답은 31.7%로 가장 높게 나타났다. 그다음으로 ‘그렇다’가 30.5%, ‘보통이다’가 29.3%로 유사한 비율로 연이어 나타났다. 다른 만족도 조사 항목에 비하여 ‘실습 수행에 대한 만족도’는 상대적으로 낮게 나타난 결과를 보인다. 참고로 2025년 한국어 학습자 말뭉치 교실에서는 3차와 4차 행사에서 강의와 실습이 병행되었다.

⑥ 참가 후 한국어 학습자 말뭉치에 대해 더 잘 알게 되었다.

- 학습자 말뭉치 교실 참가 후, 한국어 학습자 말뭉치에 대해 더 잘 알게 되었는가라는 질문에 대해서는 응답자의 50.2%가 ‘매우 그렇다’로 가장 높은 비율을 차지하였다. 그다음으로 ‘그렇다’가 32.8%, ‘보통이다’가 15.8%로 나타났다. 반면, ‘그렇지 않다’와 ‘매우 그렇지 않다’는 각 0.8%와 1.2%이라는 낮은 비율로 나타났다.



<그림 38> 한국어 학습자 말뭉치에 대한 이해도 분포

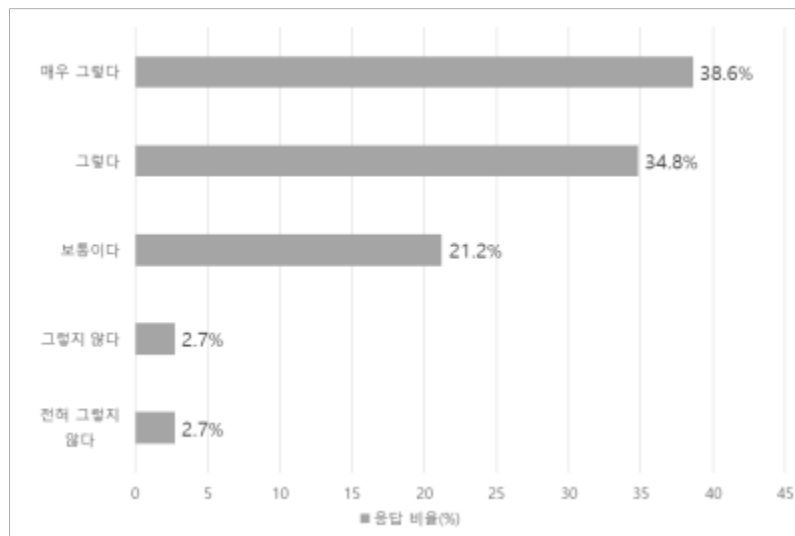
<표 56> 한국어 학습자 말뭉치에 대한 이해도 빈도

만족도	빈도
매우 그렇다	130 (50.2%)
그렇다	85 (32.8%)

보통이다	39 (15.0%)
그렇지 않다	2 (0.8%)
매우 그렇지 않다	3 (1.2%)
합계	259 (100.0%)

㉓ 참가 후 말뭉치 활용 능력이 향상되었다.

- 학습자 말뭉치 교실 참가 후, 한국어 학습자 말뭉치 활용 능력이 향상되었는가에 대한 질문에 대해서는 ‘매우 그렇다’라고 답변한 응답자가 가장 많았다. 이어서 ‘매우 그렇다’라고 답한 응답자의 비율이 38.6%, ‘그렇다’가 34.8%, ‘보통이다’가 21.2%로 학습자 말뭉치가 참가자의 말뭉치 활용 능력 향상에 긍정적인 영향을 미쳤음을 알 수 있었다.



<그림 39> 말뭉치 활용 능력에 대한 만족도 분포

<표 57> 말뭉치 활용 능력에 대한 만족도 빈도

만족도	빈도
매우 그렇다	100 (38.6%)
그렇다	90 (34.8%)
보통이다	55 (21.2%)
그렇지 않다	7 (2.7%)

매우 그렇지 않다	7 (2.7%)
합계	259 (100.0%)

- 행사 참가 후 말뭉치 활용 능력이 향상되었는가에 대한 응답 결과는 ‘매우 그렇다’라고 응답한 비율이 38.6%로 가장 높게 나타났다. 그다음으로 ‘그렇다’가 34.8%, ‘보통이다’가 21.2%로 나타났다. ‘그렇지 않다’와 ‘매우 그렇지 않다’라고 응답한 비율은 각 2.7%로 동일하게 나타났다.

Ⅲ. 추가 의견 사항

- ① 향후 학습자 말뭉치 교실에서 듣고 싶은 강의
- 한국어 학습자 말뭉치 교실에서 듣고 싶은 강의에 대한 주관식 답변으로는 다음과 같은 것들이 있었다.

<표 58> 한국어 학습자 말뭉치 교실에서 듣고 싶은 강의에 대한 의견

의견 유형	내용
말뭉치 이해 관련 (29명)	<ul style="list-style-type: none"> ○ 말뭉치의 유형별 소개 (11명) <ul style="list-style-type: none"> - 원시 말뭉치, 문어/구어 말뭉치, 오류 말뭉치, 언어권별 말뭉치, 속달도별 말뭉치, 모어 화자 말뭉치 등 ○ 말뭉치의 구축 과정에 대한 교육 및 실습 (9명) <ul style="list-style-type: none"> - 새로운 말뭉치를 직접 구축하는 방법 ○ 국립국어원 한국어 학습자 말뭉치의 구조 및 특징 소개 (4명) <ul style="list-style-type: none"> - 말뭉치의 통계 결과 포함 ○ 말뭉치 자료 검색 및 분석 방법 (3명) <ul style="list-style-type: none"> - 말뭉치 자료 검색에 대한 매뉴얼 배포 필요 ○ 말뭉치 활용 시 고려해야 하는 저작권 관련 교육 ○ 타 국가별(언어권별) 말뭉치 구축 및 구현 사례 소개
말뭉치 활용 관련 (84명)	<ul style="list-style-type: none"> ○ 말뭉치를 활용할 때 사용하는 도구에 대한 교육 및 실습 (38명) <ul style="list-style-type: none"> - 설치 방법에 대한 안내 및 도구의 사용 목적 소개 포함 예) NetMiner, Python, AntConc, UTagger, Kiwi, Lancbox, json, Excel 파일 활용 등

의견 유형	내용
	<ul style="list-style-type: none"> ○ 말뭉치의 다양한 활용 방법 소개 (32명) - 말뭉치 유형별(원시, 오류 등) 활용 방법 - 분야별(교수/학습/연구 등) 활용 방법 - 국립국어원의 자료를 활용하는 방법 - 학습자 입장에서 말뭉치를 활용할 수 있는 방법 - 학습자 유형에 따른 말뭉치의 활용 방법 예) 학습자 숙달도에 따른 말뭉치 활용, 중도입국 청소년 또는 재외동포 학습자 대상에 맞춘 말뭉치 활용 등 ○ AI 프로그램을 활용하여 말뭉치를 분석하는 방법 소개 (8명) - LLM 활용 관련 포함 ○ 말뭉치 결과를 통계적으로 분석하는 방법 소개 (3명) ○ 특정 주제를 잡고 실제로 말뭉치를 응용하는 실습 (3명)
연구 및 교수 현장에서의 말뭉치 적용 관련 (79명)	<ul style="list-style-type: none"> ○ 말뭉치를 활용한 교수 및 학습 사례 소개와 실습 (43명) - 말뭉치를 적용할 수 있는 수업 및 교수 방법에 대한 사례 소개 요구(수업 시연 포함) - 말뭉치를 활용하여 현장에서 사용할 교수 자료 개발 및 제작에 대한 사례 소개 요구 예) 숙달도별 수업 적용, 문법 강의에의 적용, 언어 기능(말하기, 듣기, 쓰기, 읽기) 수업을 위한 자료 제작, 문법 및 어휘 학습을 위한 자료 제작, 학습자 유형(어학당, 중도입국, 초등학생 등)에 맞춘 수업 적용 사례, 연습지 제작 등 ○ 말뭉치를 활용한 연구의 절차 및 결과물 소개 (25명) ○ 학습자 말뭉치를 활용한 연구 동향 소개 (6명) - 타 분야와의 결합이 이뤄진 연구 동향 포함 - 동향 분석을 통한 말뭉치 연구 방법론 및 도구의 장단점 비교 ○ 오류 말뭉치를 활용한 한국 언어 및 문화 오류의 예시 소개 (3개) ○ 오류 말뭉치를 활용한 맞춤형 피드백 전략 소개 ○ 다국어 말뭉치와 학습자 말뭉치의 비교

② 학습자 말뭉치 교실 및 학습자 말뭉치 구축에 대한 자유 의견

- 앞에서 언급된 항목 외에 한국어 학습자 말뭉치 교실 및 구축에 대한 자유 의견으로는 다음과 같은 것들이 있었다.

<표 59> 학습자 말뭉치 교실 및 구축에 대한 자유 의견

의견 유형	내용
행사 운영 관련 (72명)	<ul style="list-style-type: none"> ○ 강의 회차의 확대 필요 (10명) ○ 반복 수강을 위한 강의 내용의 영상 제공 (9명) - 실습 내용이나 지나간 회차의 내용을 다시 듣고 싶다는 다수 의견 존재 ○ 실습 위주의 학습자 말뭉치 교실 운영의 확대 (8명) - 이론 강의보다 실습을 선호하는 경향 존재 ○ 문서화된 강의 자료의 배포 필요 (7명) - 강의 자료를 미리 받아 수강할 때 참고하고자 하는 목적 ○ 저녁 및 주말 시간대의 강의 개설 필요 (6명) - 업무 시간을 고려한 시간대 설정 필요 - 다른 지역에서 대면 참가를 원하는 경우를 대비한 시간대 설정 필요 ○ 강의 및 실습 시간의 연장 필요 (6명) - 실습 기간의 연장 혹은 선택적인 실습 추가 시간 제공 ○ 정기적이고 지속적인 학습자 말뭉치 교실 개설 필요 (5명) - 정기적으로 일정을 계획하여 달별로 혹은 분기별로 진행되도록 강의를 개설하는 것에 대한 고려 필요 ○ 학습자 말뭉치 교실 및 구축에 대한 홍보 확대 필요 (5명) - 다음 회차의 학습자 말뭉치 교실에 대한 내용과 일정을 행사 마지막에 미리 홍보 ○ ‘입문-초급-중급-고급’으로 나뉘는 단계별 교육 필요 (4명) - 초보자를 위한 말뭉치 활용 강의 요구가 다수 존재 ○ 동일 주제의 강의 내용의 반복 학습 필요 (3명) - 특히 기초적인 내용에 대한 반복 학습 요구가 다수 존재 ○ 연속성이 있는 강의 주제 필요 ○ 대면 및 비대면 하이브리드 방식의 운영 유지 필요 (2명) ○ 대면 방식의 운영 확대 필요 (3명) ○ 하이브리드 방식이 아닌 비대면 별도 강의 필요 - 대면 강의를 송출하는 비대면 방식의 강의로는 현장에서 강의자가 무엇을 지칭하는지 확실하게 알기 어려운 문제 존재

의견 유형	내용
	<ul style="list-style-type: none"> ○ 외국인 대상 한국어 학습자 말뭉치 교실 운영 고려 필요 ○ 어학당에서 말뭉치를 활용하고 있는 강의자의 강의 필요
실습 관련 (29명)	<ul style="list-style-type: none"> ○ 더 쉬운 설명이 필요하다는 의견 (7명) - 현장에서 일하는 교사들의 경우 학습자 말뭉치를 다루는 것이 낯설 수밖에 없음 ○ 실습 자료 및 프로그램에 대한 사전 다운로드와 설치 안내 필요 (5명) ○ 실습 자료의 사전 공유 필요 (4명) - 실습에 참고할 수 있는 용어에 대한 사전 지도 필요 - 실습 전 개별 학습이 필요한 부분에 대한 구분 및 지도 필요 ○ 동일 내용의 반복적인 실습 제공 필요 (4명) ○ 천천히 진행되는 실습 속도에 대한 요구 (3명) ○ 다양하고 구체적인 실습 예시의 제공 필요 (2명) ○ 한 단계에서 다른 단계로 넘어가기 전 시간 간격 필요 - 참가자들이 따라 할 수 있는 시간 확보를 위함 ○ 국립국어원 한국어 학습자 말뭉치의 자료를 활용한 실습 고려 ○ Mac 사용자를 위한 실습 프로그램 제공 고려 ○ 보다 체계적이고 실질적인 구성의 실습 진행 필요
강의 진행 관련 (24명)	<ul style="list-style-type: none"> ○ [공통] 강의자의 철저한 준비 필요 (4명) - 강의 시작 전 현장에서의 도구(프로그램 등) 준비 및 숙지 필요 ○ [공통] 조별 모임의 활용 필요 ○ [공통] 질의응답 시간 필요 ○ [공통] 실습 중 지속적인 지도 및 안내 필요 ○ [공통] 어학당에서 근무하는 강의자의 강의 필요 ○ [대면] 원활한 인터넷 사용이 가능한 장소 대관 필요 (2명) ○ [비대면] 참가자 통제 필요 (3명) - 마이크를 끈 상태로 참여하도록 통제 필요 - 하울링 방지에 대한 관리 필요 ○ [비대면] 강의 중 실습에 관한 질의응답 필요 (3명) - 채팅창을 통한 질의응답 방식의 유지 선호 ○ [비대면] 조금 더 매끄러운 강의 진행 필요 (3명) - 비대면 참가자를 고려한 강의자의 강의 진행을 바람 ○ [비대면] 화면에 발표자가 비춰질 수 있도록 개선 필요 (3명) - 현장 상황을 볼 수 있는 화면과 실습 화면의 동시 송출 고려 필요

의견 유형	내용
	<ul style="list-style-type: none"> - 현재의 강의는 비대면 참여자에게는 다소 일방적인 강의 진행처럼 느껴짐 ○ [비대면] 비대면 강의를 위한 음질 개선 필요 ○ [비대면] 현장의 인터넷 연결에 대한 개선 필요
4) 학습자 말뭉치 구축 관련 (24명)	<ul style="list-style-type: none"> ○ 특수 분야의 한국어 (학습자) 말뭉치 구축 필요 (6명) 예) 중도입국/재외동포/아동/이주민/유학생 학습자의 말뭉치, 외국인 인이 주로 취업하는 분야별 말뭉치, 학습자의 모어 구어 말뭉치, 지역어 말뭉치 등 ○ 시각적으로 단순화되어 직관적으로 쉽게 사용할 수 있는 말뭉치 및 플랫폼의 UI의 개선 필요 (4명) - 현재의 말뭉치는 다소 복잡하게 느껴짐 - 직관적으로 검색 결과가 도출되어 바로 수업의 예문으로 활용할 수 있도록 개선 필요 ○ 한국어 학습자 말뭉치 사용 중 문의 및 수정 사항에 대한 소통 창구 필요 (3명) ○ 많은 양의 말뭉치 데이터 구축 필요 (2명) - 학습자들의 실제 쓰기 데이터를 활용하는 방안 고려 ○ 말뭉치 수집 대상 학습자의 국적 확대 필요 (2명) ○ 구어 말뭉치의 확대 필요 ○ 영상을 활용한 학습자 말뭉치 구축 고려 - 음성 데이터와 동시에 비언어적 요소를 확인하기 위함 ○ AI 기술을 활용한 말뭉치 기능 고려 ○ 말뭉치 검색 과정의 단순화 필요 ○ 국적과 모어에 대한 검색 편의성 보완 필요 ○ 검색 어절에 대한 말뭉치의 전체 맥락을 파악할 수 있는 정보의 제공 필요 ○ 학습자 말뭉치 구축에 대한 지속적인 정보 공유 필요

4.3. 상시 동영상 교육 자료

본 사업에서는 4회의 학습자 말뭉치 교실 외에 3종의 상시 동영상 교육 자료를 개발하였다. 상시 동영상 교육 자료에 수록된 텍스트 에디터, 엑셀은 학습자 말뭉치를 활용하기 위한 기초적인 자료 처리 프로그램으로 그간 학습자 말

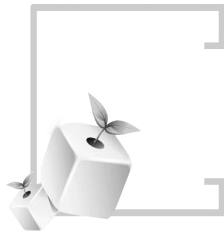
뭉치 교실을 통해 교육을 해 왔으며, 많은 수요가 있어 왔다. 교수 자료 개발을 위한 학습자 말뭉치 활용의 기초를 다룬 또 다른 프로그램은 교수 현장의 교원들을 위한 교육 자료로 이 또한 학습자 말뭉치 교실에서 이용자들의 참여가 많았던 프로그램이다. 상시 동영상 교육 자료의 개발 개요는 다음과 같다.

<표 60> 학습자 말뭉치 상시 동영상 교육 자료 개발 개요

구분	프로그램	자료 유형	러닝 타임	강의자
1	한국어 학습자 말뭉치 활용을 위한 텍스트 에디터 활용 지침 -텍스트 에디터 활용 기초-정규 표현식을 활용한 한국어 학습자 자료 전처리	동영상	약 42분*2편	최정도 (계명대)
2	한국어 학습자 말뭉치 활용을 위한 엑셀 활용 지침 - 엑셀 활용 기초 - 피벗테이블, 함수를 활용한 자료 처리 및 빈도 분석	동영상	약 30분*1편 약 45분*1편	남신혜 (경희대)
3	학습자 말뭉치를 교육 자료에 활용하기	동영상	약 30분*1편	한송화 (연세대)

4.4. 말뭉치 소개 · 활용 자료집 현행화 및 배포

본 사업에서는 <학습자 말뭉치 나눔터>의 통계 자료를 현행화하여 학습자 말뭉치 교실 참가자에게 홍보하고 학습자 말뭉치 활용 매뉴얼을 배포하여 이용을 활성화하였다. 학습자 말뭉치 활용 매뉴얼은 ‘한국어 학습자 말뭉치 나눔터’에서의 자료 검색 방법과 통계 정보를 소개한 자료로, ‘한국어 학습자 말뭉치 나눔터’를 개발한 (주) 이르테크에서 2017년에 ‘한국어 학습자 말뭉치 나눔터 검색, 통계 활용 길잡이’라는 제목으로 제작되었다. 이 자료에는 말뭉치 통합 검색, 원시 말뭉치 검색, 형태 주석 말뭉치 검색, 오류 주석 말뭉치 검색 방법이 상세히 설명되어 있으며, 말뭉치 유형별 통계 정보가 제시되어 있다. 본 사업에서는 회차별 말뭉치 교실에서 말뭉치 구축 통계를 현행화하여 소개하고, 교육 자료와 함께 배포하였다.



제 III 장

결론



1. 사업 결과 요약

본 사업은 2010년 기초 연구를 시작으로 2차에 걸쳐 진행되어 온 한국어 학습자 말뭉치 중장기 계획(2015~2025년)의 최종 단계로, 지난 10여 년의 사업을 마무리하고 최종 성과를 산출하는 것을 목적으로 수행되었다. 이에 따른 사업의 성과는 다음과 같다.

○ 한국어 학습자 말뭉치 구축

본 사업에서는 기구축 말뭉치의 말뭉치 유형별, 대상별, 수준별, 언어권별(국적별) 분포 현황을 토대로, 35만 어절의 형태 주석 말뭉치와 50만 어절의 오류 주석 말뭉치를 구축하기 위한 설계를 하고 구축 작업을 진행하였다. 그 결과 형태 주석 말뭉치는 25개 언어권의 문어 표본 2,838개, 355,020어절이 구축되었고, 오류 주석 말뭉치는 23개 언어권의 문어 표본 5,147개, 608,668어절이 구축되었다. 말뭉치를 구축하는 과정에서 학습자와 자료에 관한 메타 정보가 포함되도록 하였으며, 학습자의 이름, 외국인 등록번호, 카드 번호, 전화 번호 등의 개인정보는 지침에 따라 비식별화 처리되도록 하였다. 아울러 메타 정보를 중심으로 자료의 무결성에 대한 검수 작업을 함께 진행하여 수정 작업을 병행하였다. 그 외에도 한국어 학습자 말뭉치 구축 시스템(LCMS)를 기반으로 작업 배분, 작업 진행, 작업 결과를 체계적으로 관리하고, 2단계 검수 작업을 진행하여 작업의 질을 높이하고자 하였다. 2025년 형태 주석 및 오류 주석 결과를 포함한 2015-2025년 구축 말뭉치의 누적 규모는 원시 말뭉치 10,655,477어절(문어 8,808,556어절, 구어 1,846,921어절), 형태 주석 말뭉치 4,368,151어절(문어 3,115,001어절, 구어 1,253,150어절), 오류 주석 말뭉치 1,957,108어절(문어 1,283,420어절, 구어 673,688어절)이 되었다.

○ 한국어 학습자 쓰기 평가 자료 구축

한국어 학습자 쓰기 평가 자료 구축은 최근 많은 연구가 이루어지고 있는 쓰기 자동 채점을 비롯한 기술 개발에 활용 가능한 인공지능 학습용 데이터 구축을 위한 기반 자료를 마련하기 위한 것이다. 본 사업에서는 주제 1(10년 후 나의 계획) 1,804건, 주제 2(내가 가 본 여행지) 1,694건, 주제 3(내가 생각하는 행복과 성공) 1,502건으로 총 5,000건의 학습자 작문을 평가하였다. 등급 체계 및 평가 구인, 평가 기준은 자료의 호환성과 일관성을 위하여 <2023-2024년

한국어 학습자 말뭉치 구축>에서 마련한 지침을 준용하였다. 이에 따라 한 편의 작문에 대하여 2명의 평가자가 교차 채점을 진행한 후, 등급별 점수(100점 만점), 1-6급의 누적 점수(600점 만점), 최종 점수(100점 만점)로 구성된 3종의 점수와 함께, ‘국제 통용 한국어 교육 표준 모형’을 기준으로 한 등급을 부여하고 한국어교육 전문가의 총평을 제시하였다. 평가 대상이 된 작문은 총 88개국 학습자가 산출한 자료로, 주제별 평가 결과에 대해 신뢰도 검정을 실시한 결과는 모두 높은 신뢰도를 보이는 것으로 확인되었다.

○ 한국어 학습자 말뭉치 파일 변환

한국어 학습자 말뭉치는 매해 구축이 완료된 후 한국어 학습자 말뭉치 나눔터를 통해 원시 말뭉치, 형태 주석 말뭉치, 오류 주석 말뭉치별로 텍스트(원시), 엑셀(형태), XML(오류)을 생성해 배포해 오고 있다. 최근 JSON 포맷이 보편적으로 사용되기 때문에 추가 변환하는 것으로 결정되었다. 이에 따라 본 연구에서는 국립국어원의 말뭉치 배포용 데이터 규격에 학습자 말뭉치의 메타데이터를 추가하고 오류 주석의 주석 정보 특수성을 감안해 JSON 스키마를 수립하였다. 이를 토대로 2015-2025년 구축된 전체 자료를 변환하였다.

○ 한국어 학습자 말뭉치 교육 및 홍보

한국어 학습자 말뭉치 교육 및 홍보는 한국어 학습자 말뭉치의 원활한 구축과 작업자의 전문성 제고를 통한 결과물의 품질 향상, 성과의 확산과 이용자의 활용 확대를 목적으로 한다. 본 사업에서는 학습자 말뭉치 구축 실무자를 대상으로 형태 주석 말뭉치와 오류 주석 말뭉치 구축을 위한 교육을 실시하고, 한국어 학습자 말뭉치 이용자를 대상으로 총 4회의 ‘한국어 학습자 말뭉치 교실’을 개최하였다. ‘한국어 학습자 말뭉치 교실’은 2015-2024년 개최 성과에 대한 분석을 토대로 교육 대상, 수준에 따라 차등화된 다양한 프로그램으로 구성되었으며, 1차(서울 연세대) 276명, 2차(대전 배재대) 150명, 3차(부산 부산외대) 153명, 4차(광주 호남대) 166명의 참가자들이 참석하여 성황리에 개최되었다. 아울러 본 사업에서는 학습자 말뭉치 활용 교육에 대한 이용자의 요구에 부응하기 위하여 3종의 상시 동영상 교육 자료를 개발하였다. 상시 동영상 교육 자료는 텍스트 에디터, 엑셀을 이용한 학습자 말뭉치 활용법, 한국어 학습자 말뭉치를 활용한 교수 자료 개발을 주제로 하였다. 그 외에도 <학습자 말뭉치 나눔터>의 통계 자료를 현행화하여 학습자 말뭉치 교실 참가자에게 홍보하고 ‘학습자 말뭉치 활용 매뉴얼’을 배포하여 이용을 활성화하였다.

2. 사업의 의의 및 기대 효과

한국어 학습자 말뭉치는 한국어 학습자가 산출한 언어 자료를 체계적으로 수집·정제하여 구축한 국가 차원의 공공 언어 자원이다. 이는 한국어 교육 및 연구를 위한 기초 인프라이자, 인공지능 기반 언어 기술 개발에 활용 가능한 핵심 원천 자료라는 점에서 높은 가치를 지닌다. 한국어 학습자 말뭉치 구축을 주목적으로 하는 본 사업의 의의와 기대 효과는 다음과 같다.

○ 국가 차원의 대규모 학습자 언어 자원 구축

한국어 학습자 말뭉치 구축 사업은 2010년 기초 연구를 수행한 이후, 2015-2020년 1차 중장기 계획, 2021-2025년 2차 중장기 계획을 통해 실행되어 왔다. 본 사업은 2차 중장기 계획을 마무리하는 해로, 누적 규모 약 1,000만 어절 규모의 원시 말뭉치, 435만 어절 규모의 형태 주석 말뭉치, 195만 어절 규모의 오류 주석 말뭉치를 최종 성과로 산출하였다.

특히 2025년에는 약 35만 어절의 형태 주석 말뭉치와 50만 어절의 오류 주석 말뭉치 구축을 목표로 하여 주석 가공 말뭉치의 규모를 확대하여 자료의 활용도를 제고하였다. 아울러 소수 언어권의 자료 비중을 확대하고 수준별 균형성을 제고함으로써 자료의 대표성과 균형성을 확대하고자 하였다. 이는 한국어 교육 및 응용 연구에서 신뢰도 높은 실증 자료로 활용될 수 있는 기반을 강화한다는 점에서 중요한 의미를 지닌다.

○ 한국어교육 연구 및 교육 자료 혁신의 토대 마련

한국어 학습자 말뭉치는 학습자의 중간언어와 언어 습득의 양상을 실증적으로 확인할 수 있는 자료로, 학습자의 언어 발달 특성을 객관적으로 연구하여 한국어 교수·학습 이론을 정립하는 데에 이바지할 수 있다. 또한 교재 개발, 교수법 설계, 평가 도구 개발 등 교수 현장에서도 그 활용도가 매우 높다. 뿐만 아니라 인공지능 기술을 기반으로 한 교육 콘텐츠 개발을 주도하는 민간 분야에서도 유용한 기초 자료가 된다.

이러한 점에서 공공 언어 자료로서 한국어 학습자 말뭉치 구축과 배포, 이용자 교육을 위한 연수회 운영은 한국어교육 연구 및 교육 자료 개발의 저변을 확대하고 선진화하였다는 점에서 높이 평가할 수 있다.

○ 한국 언어·문화의 국제적 확산과 국가 경쟁력 제고

한류와 함께 시작된 한국어와 한국 문화에 대한 세계적 관심은 현재까지도 지속적으로 확대되고 있으며, 이는 한국어 학습자의 증가로 이어지고 있다. 한국어 학습자 말뭉치는 다양한 배경과 목적을 지닌 학습자들의 세분화된 요구를 반영한 교육과정 개발 및 교수요목 설계, 교육 자료, 교수법 개발의 기초 자료가 된다. 이는 보다 체계적이고 수준 높은 교육 환경 제공의 발판이 되며, 궁극적으로 한국 언어·문화의 세계적 확산을 촉진하는 데에 기여할 수 있다. 더 나아가 다양한 분야에서 한국과 전 세계 여러 나라 간의 교류를 촉진할 수 있는 국제 인재를 양성함으로써 국가 경쟁력을 강화할 수 있다.

3. 보고서 활용 방안

○ 한국어 학습자 말뭉치 구축·활용에 대한 이론적 준거 마련

본 보고서에는 2025년 사업에서 수행한 과업의 절차와 방법, 결과가 체계적으로 기술되어 있으며, 2015-2025년까지의 사업 수행 과정에서 수정·보완된 학습자 말뭉치 최종 지침에 부록으로 수록되어 있다. 한국어 학습자 말뭉치는 비모어 화자가 산출한 비정형의 자료를 구축 대상으로 한 특수 말뭉치로, 구축 설계와 구축에서 다양한 쟁점이 대두되는데, 이러한 내용을 체계화한 이론적 논의가 많지 않다. 본 보고서는 한국어 학습자 말뭉치 구축에 관한 실천적 보고서로, 학습자 말뭉치 구축, 활용에 관한 이론적 기초로 활용될 수 있다.

○ 한국어 학습자 말뭉치 구축을 위한 실천적 가이드라인 제공

보고서 부록에 제시된 지침은 자료 수집, 입력 및 전사, 형태 주석, 오류 주석 처리에 관한 기본 지침은 물론 실제 작업 과정에서 발생하는 구체적인 문제와 해결 방안이 상세하게 담겨 있다. 본 보고서의 지침은 한국어 학습자 말뭉치와 유사한 말뭉치를 구축하고자 할 때, 실제적인 작업 지침이 될 수 있다. 또한 직접 구축을 하지 않더라도 이용자가 한국어 학습자 말뭉치의 세부적인 특성을 파악함으로써 활용 범위를 무한 확장할 수 있다.

4. 정책 제안

○ 지속적인 자료 수집과 구축을 통한 학습자 말뭉치의 규모 확대

한국어 학습자 말뭉치는 2015-2025년 10여 년간의 사업을 통해 원시 말뭉치 기준 1,000만 어절 규모의 말뭉치를 구축하였다. 비모어 화자의 산출 자료라는 특성상 한 편당 평균 100어절 내외의 자료를 표본 단위로 수집하여 구축하기 때문에 모어 화자 말뭉치보다 자료의 수집 및 구축에 많은 시간과 비용, 노력이 소요된다. 지금까지 학습자 말뭉치는 국내외 한국어 교육 기관 및 학계의 수집 네트워크를 기반으로 수집되었는데, 이를 유지하되 한국어능력시험(TOPIK)이나 세종한국어평가(SAK)와 같은 공인 평가와의 연계를 통해 균질한 자료를 지속적으로 수집할 수 있도록 수집 네트워크의 변경이 필요하다. 이를 통해 새로운 자료를 계속 수집하고, 이를 통해 인공 지능 시대에 요구되는 대규모 말뭉치로의 확대하여야 할 것이다.

○ 활용도 제고를 위한 주석 가공 말뭉치의 확대

한국어 학습자 말뭉치는 원시 말뭉치에 비해 형태 주석 말뭉치와 오류 주석 말뭉치의 비중이 매우 부족하다. 비모어 화자가 산출한 비정형의 자료라는 특성으로 인해 자료의 주석 가공을 완전하게 자동화하지 못하고 국어학과 한국어 교육 지식을 겸비한 전문 인력들이 일일이 주석하고 검수하는 과정을 거치기 때문이다. 그럼에도 한국어교육 연구 및 교육, 산업계에서의 활용을 위해서는 주석 가공 말뭉치의 확대가 필수적이다.

○ 한국어 교원 및 학습자를 위한 학습자 말뭉치 기반 교수·학습 자료 개발

한국어 학습자 말뭉치는 ‘한국어 학습자 말뭉치 나눔터’를 통해 배포되고 있으며, 한국어 교수 현장의 교원이나 학습자들은 한국어 학습자 말뭉치 나눔터의 검색 기능을 활용해 학습자의 오류 자료를 검색하여 교수·학습에 활용하고 있다. 그러나 이들 자료는 검색 조건에 의해 추출되는 전체 자료로, 교수 현장에서의 활용을 위해서는 학습자의 수준과 언어권에 따른 전형적인 오류 특성을 포함한 자료로의 정제 과정이 필요하다. 이에 따라 자료 처리 능력이 부족하거나 시간적인 여유가 없는 현장의 교원들은 한국어 학습자 말뭉치 활용에 대한 요구는 있으나 실질적으로 사용하지 못하는 경우가 많다. 따라서 한국어 학습자 말뭉치 이용자의 저변 확대, 활용 범위 확대의 차원에서 학습자 말뭉치를 교수·학습 자료로 2차 가공하여 배포하는 연계 사업을 고려해 볼 수 있다.

참고 자료

- 강현화 외(2010), 한국어 학습자 말뭉치 구축 설계, 국립국어원.
- 강현화(2010) 한국어 학습자 사전 표제어 선정을 위한 자료 구축 및 선정 방법에 관한 연구, 한국사전학 16, 한국사전학회.
- 강현화 외(2015), 2015년 한국어 학습자 말뭉치 기초 연구 및 구축, 국립국어원.
- 강현화 외(2016), 2016년 한국어 학습자 말뭉치 연구 및 구축, 국립국어원.
- 강현화 외(2017), 2017년 한국어 학습자 말뭉치 연구 및 구축, 국립국어원.
- 한송화 외(2018), 2018년 한국어 학습자 말뭉치 연구 및 구축, 국립국어원.
- 한송화 외(2019), 2019년 한국어 학습자 말뭉치 연구 및 구축, 국립국어원.
- 한송화 외(2020), 2019-2020년 한국어 학습자 말뭉치 연구 및, 국립국어원.
- 한송화 외(2021), 2021년 한국어 학습자 말뭉치 연구 및 구축, 국립국어원.
- 한송화 외(2022), 2022년 한국어 학습자 말뭉치 연구 및 구축, 국립국어원.
- 곽용진 외(2023), 2023년 한국어 학습자 말뭉치 구축, 국립국어원.
- 곽용진 외(2024), 2023년 한국어 학습자 말뭉치 구축, 국립국어원.
- 강현화(2017), 학습자 말뭉치의 구축과 활용, 소통.
- 강현화(2011) 한국어 학습자 말뭉치의 자료 구축 방안 대한 기초 연구, 한국사전학 17, 한국사전학회.
- 강현화(2017), 중국인 한국어 학습자 말뭉치에 나타난 중간언어 분석 연구, 언어사실과 관점 41, 연세대학교 언어정보연구원, 5-47.
- 강현화.조민정(2003), 스페인어권 한국어 학습자의 어미, 조사 및 시상, 사동 범주의 오류 분석, 한국어교육 14(2), 국제한국어교육학회.
- 고석주(2002), 학습자 말뭉치에서 조사 오류의 특징, 외국어로서의 한국어교육 27(1), 연세대학교 한국어학당.
- 고석주(2004), 오류 유형 주석을 위한 기초 연구, 한국 문화사.
- 고승연(2013), 아랍어권 한국어 학습자의 발음 오류 분석, 한국어문화교육 7(1), 한국어 문화교육학회.
- 권기양(2006), KFL 학습자의 오류에 대하여: 중국인 학습자 중심으로, 언어과학 13(3), 한국언어과학회.
- 김경화(2013), 고급단계 한국어학습자의 오류연구, 중국조선어문 188, 길림성민족사무위원회.
- 김미경·강현화(2017), 중·고급 중국어권 한국어 학습자의 조사 '가'와 '는' 선택 요인 연구, 외국어로서의한국어교육 47, 25-52.
- 김미옥(2002), 학습 단계에 따른 한국어 학습자 오류의 통계적 분석, 외국어로서의 한국어 교육 27(1), 연세대학교 한국어학당.
- 김미옥(2003), 한국어 학습자의 단계별 언어권별 어휘 오류의 통계적 분석, 한국어 교

- 육 14(3), 국제한국어교육학회.
- 김미옥·정희정(2003), 한국어 학습자 작문에 나타난 어휘 오류 분석, 제3회 한국어 교육 국제 워크숍 발표 요지, 연세대 언어정보연구원 외국어로서의 한국어교육연구센터, 102-135쪽.
- 김아름(2014), 한국어 학습자의 문법 및 화용오류에 대한 인식, 새국어교육 100, 한국국어교육학회.
- 김유미(2002), 학습자 말뭉치를 이용한 한국어 학습자 오류 분석 연구, 외국어로서의 한국어교육 27, 연세대학교 한국어학당.
- 김유정(2005), 한국어 학습자 말뭉치 오류 분석의 기준, 한국어 교육 16(1), 국제한국어교육학회.
- 김일환(2016), 한국어 학습자 말뭉치의 주석 과정과 활용 방법, 국제한국어교육학회 춘계학술발표논문집, 국제한국어교육학회.
- 김정숙(2002), 영어권 한국어 학습자의 조사 사용 오류 분석과 교육 방법, 한국어교육 13(1), 국제한국어교육학회.
- 김정숙(2002), 한국어 학습자 말뭉치 구축을 위한 기초 연구 -개인 정보 표지 체계와 오류 정보 표지 체계를 중심으로-, 이중언어학회.
- 김정숙, 김유정(2002), 한국어 학습자 말뭉치 구축을 위한 기초 연구 -개인정보 표지 체계와 오류 정보 표지 체계를 중심으로-, 이중언어학 21, 이중언어학회.
- 김정은(2003), 한국어교육에서의 중간언어와 오류 분석, 한국어 교육 14(1), 국제한국어교육학회.
- 김지민, 신승용(2010), 어휘오류 분석의 문제점과 어휘오류 처치 방안 연구, 언어와 문화 6(2), 한국언어문화교육학회.
- 김지영(2014), 중국인 유학생의 한국어 사용 오류 분석, 시학과 언어학 26, 시학과언어학회.
- 김한샘, 배미연(2017), 학문 목적 학습자의 객관화 전략 사용 양상 연구 - 중국인 학습자의 학술 텍스트를 중심으로, 언어사실과 관점 41, 연세대학교 언어정보연구원, 5-47.
- 김한샘·곽용진(2016), 차세대 학습자 말뭉치 통합 관리 시스템 개발, 한국언어문화교육학회 학술대회 발표 자료집, 한국언어문화교육학회.
- 남길임(2007), 학습자 오류 말뭉치를 활용한 한국어 용법 사전의 편찬, 한말연구회.
- 남윤주 외(2014), L2로서의 한국어 자연발화 코퍼스의 구축과 활용, 통일인문학논총.
- 노미연(2012), 한국어 학습자의 구어 오류와 후속 상호작용 분석 연구, 동국대학교 박사학위논문.
- 민영란(2008), 부정적 전이로 인한 중국어권 학습자의 오류 분석, 한국어 교육 19(1), 국제한국어교육학회.
- 박수연(2007), 한국어 학습자 오류 말뭉치 구축과 그 문제점에 관한 연구, 언어 사실과 관점 17, 연세대학교 언어정보연구원.

서상규, 유현경, 남윤진(2002), 한국어 학습자 말뭉치와 한국어교육, 한국어교육 13(1), 국제한국어교육학회.

신성철(2002), 호주 한국어 학습자의 어휘 오류 분석 연구, 한국어 교육 13(1), 국제한국어교육학회.

신성철(2007), 영어권 한국어 학습자의 철자 오류 유형과 패턴, 한국어 교육 18(3), 국제한국어교육학회.

유석훈(2001), 외국어로서의 한국어 학습자 말뭉치 구축의 필요성과 자료 분석, 한국어 교육 12(1), 국제한국어교육학회.

이동은(2007), 한국어 학습자의 철자 오류와 개선 방안 -북미지역 청소년 교포 학습자를 대상으로-, 한국어학 35, 한국어학회.

이병운(2011), 베트남인 학습자의 작문 오류 경향 분석: 조사·어미를 중심으로, 우리말글 52, 우리말글학회.

이승연(2006), 한국어 학습자 말뭉치 오류 표지 방안 재고, 이중언어학 31, 이중언어학회.

이승연(2007), 한국어 학습자 오류 판정 및 수정 기준 연구-교사, 비교사 집단간 오류 판별 비교 실험을 바탕으로, 이중언어학 33, 이중언어학회.

이승연(2007), 한국어교육을 위한 한국어 학습자 말뭉치의 구축과 활용 연구, 고려대 박사학위논문.

이유림, 김영주(2013), 교사의 피드백 방법이 한국어 학습자의 작문 내 어휘 오류 감소에 미치는 영향, 외국어로서의 한국어교육 39, 연세대학교 언어연구교육원 한국어학당.

이은서(2017), 중국어권 학습자의 접사 사용 연구, 연세대학교 대학원 석사학위 논문.

이정희(2002), 한국어 오류 판정과 분류 방법에 관한 연구, 한국어교육 13(1), 국제한국어교육학회.

이정희(2003), 초급 단계 한국어 학습자의 어휘 오류, 이중언어학 22, 이중언어학회.

이정희(2009), 중국어권 한국어 학습자의 어휘 오류 연구, 한국어 교육 19(3), 1-23쪽, 국제한국어교육학회.

이화진·이지연(2016), 학습자 말뭉치 구축과 음성 인식 활용, 한국언어문화교육학회 학술대회 발표 자료집, 한국언어문화교육학회.

이훈호(2015), 한국어 오류 분석 연구의 동향 분석 연구, 외국어교육연구 29(2), 107-135쪽, 한국외국어대학교 외국어교육연구소.

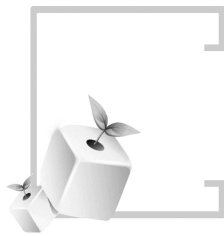
전영옥(2010), 여성결혼이민자의 한국어 어휘 오류 분석, 한말 연구 27, 한말연구학회.

조철현 외(2002), 한국어 학습자의 오류 유형 조사 연구, 문화관광부.

최원평, 유효려(2010), 중국 대학생 글쓰기에 나타난 어휘 오류 연구, 언어와 문화 6(3), 한국언어문화교육학회.

한상미(2014), 중급 한국어 학습자의 구어 담화에 나타난 조사 오류 연구, 한국어교육 25(3), 국제한국어교육학회.

- 한송화(2001), 말뭉치와 학습자 오류를 이용한, 외국인 학습자를 위한 한국어 어휘 사전의 의미 기술, 한국어정보학 4, 한국어정보학회.
- 한송화, 원미진(2017), 모어 화자와 한국어 학습자 말뭉치에서의 ‘은/는’과 ‘이/가’의 분포와 조사 선택 요인 분석, 언어사실과 관점 41, 연세대학교 언어정보연구원, 5-47.
- 한송화·강현화(2016), 학습자 말뭉치에서의 구어 전사와 오류 주석의 쟁점과 실제, 한국언어문화교육학회 학술대회 발표 자료집, 한국언어문화교육학회.
- Brock, C , Crookes, C , Day, R., and Long, M. (1986). The differential effects of corrective feedback in native speaker–non-native speaker conversation. In R. Day (Ed.), Talking to learn. Rowley, MA: Newbury House. pp. 229–236.
- Brock, C. (1986). The effects of referential questions on ESL classroom discourse. TESOL Quarterly, 20, pp. 47-59.
- Corder. S. P.(1981), Error Analysis and Interlanguage, Oxford University Press.
- Foster, P. and Skehan, P. (1996) The influence of planning on performance in task-based learning. Studies in Second Language Acquisition 18. pp. 299-324.
- Foster, P., Tonkyn, A. and Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. Applied Linguistics 21:3. pp. 354-375.
- Hunt, K. (1965). Grammatical structures written at three grade levels. NCTE Research report No. 3. Champaign, IL, USA: NCTE. pp. 1467-1770.
- James, C.(1998), Errors in Language Learning and Use. New York : Addison Welsey Longman Inc. pp. 144-154.
- Pica, T., Holliday, L., Lewis, L. and Morgenthaler, L. (1989) Comprhensible Output As An Outcome of Linguistic Demandes On the Learner, Studies in Second Language Acquisition 11:1. pp. 63-90.
- Young, R. (1995). Conversational Styles in Language Proficiency Interviews. Language Learning 45:1. pp. 3-42.



부록

한국어 학습자
말뭉치 구축 지침



차 례

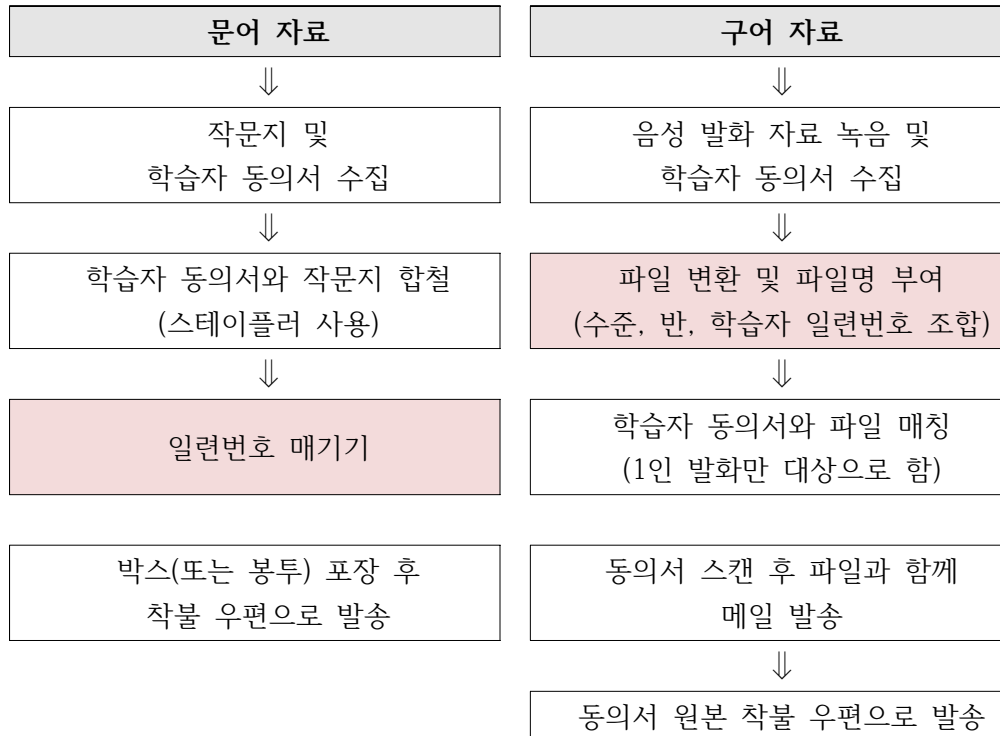
1. 2015-2022년 한국어 학습자 말뭉치 자료 수집 지침 1
2. 2023-2024년 한국어 학습자 말뭉치 자료 수집 지침 ... 42
3. 한국어 학습자 말뭉치 자료 처리 지침 52
4. 한국어 학습자 말뭉치 문어 입력 지침 57
5. 한국어 학습자 말뭉치 구어 전사 지침 64
6. 한국어 학습자 말뭉치 형태 주석 지침 86
7. 한국어 학습자 말뭉치 오류 주석 지침 152
8. 2023-2024년 학습자 동의서 양식 241

2015-2022년 한국어 학습자 말뭉치 자료 수집 지침

1. 자료 수집 대상 및 수집 자료

- ▶ 대상: 한국어 교육기관의 학습자
- ▶ 자료: 학습자가 산출한 작문과 말하기 자료
- ▶ 수집 시기: 여름 학기와 가을 학기의 각 중간, 기말의 2회(총 4회)를 원칙으로 한다. (추가 가능)

2. 자료 수집 절차



1) 문어

- ① 학습자가 손으로 쓴 작문지나 시험 답안지의 원본(사본도 가능)을 수집한다.
- ② 사본의 경우 복사가 흐릿하여 텍스트를 알아보기 힘든 경우 자료의 활용이 불가능하므로 유의한다.
- ③ 해당 자료의 출처를 파악할 수 있도록 수집된 자료는 반드시 동의서와 함께 수합하여 합철을 한다.
- ④ 합철이 된 파일에 아래와 같이 급별로 네 자리의 일련번호(0001, 0002, ...)를 붙인다.

0001

建立韓國語學習者語料庫的個人資料使用同意書

韓國語立國語院為了韓國語教學的推廣，正在搜集學習者語料(語料庫)以促進未來學者的研究。【重要事項】：蒐集學習者語料，為未來研究的資料應用於韓國語教學方法的改善、韓國語教材的開發、韓國語教育領域發展等領域。參加本研究的各位學習者沒有經濟損失和人身危險，如果提交以上表格或參與時可以隨時停止參與。另外，搜集的個人資料不會用於除本計畫以外其他目的。方法：為了保護隱私，所收集的個人資料會在他人不能識別的狀態下進行使用。謝謝！

寄件處： 國立大學 產學合作組 02-2123-4199

☒ 同意提供使用以下信息至2015年夏/秋季學期作文、會議資料。

日期： 7/9
姓名： 오빠두 (88)

下面是研究所使用的個人資料，此信息將保密不對外透露。

1. 性別： ☒ 男 ☐ 女
2. 年齡： 29
3. 現在韓國就學等級： 1
4. 國籍： Taiwan (☒ 韓國 ☐ 外國人)
5. 母語： Chinese
6. 韓國語的學習時間： 0 年 2 月
7. 資料提供最佳時間： 0 年 2 月
8. 學習韓國語的動機：
☐ 科學 ☐ 觀光 ☒ 結婚 ☐ 興趣 ☐ 其他 ()
9. 職業： Student
10. 語料提供以外可以應用的外國語(依該國語的順序來填寫)：
Chinese English Japanese Korean

001

※ (1) 주위에 무엇을 했습니까? 어디에 갔습니까? 누구를 만났습니까? 무엇을 먹었습니까? 여행했습니까?
(일시: 아래에 표백을 모두 사용해서 됩니다. (100-300자) [30점])

姓 名 年 代 性 別 國 籍 母 語

주	위	에	한	국	어	공	분	했	습	니	요	돈	요	본	에
최	구	로	역	화	는	배	움	었	습	니	요	만	여	니	로
배	구	에	서	만	났	습	니	다	저	는	백	화	집	에	서
소	스	배	백	밥	을	공	공	그	하	고	먹	었	습	니	요
저	는	시	장	에	서	꽃	빠	나	나	시	과	기	회		
추	려	요	모	도	/	0	0	0	0	원	했	니	다		
일	요	일	에	저	는	국	중	에	서	영	향	을	위	쳤	았
주	요														

- ④ 급별로 번호를 붙이며 각각 0001로 시작한다.
- ④ 한 명의 학습자 자료가 두 개 이상일 경우 0001-01, 0001-02, ... 와 같이 앞자리 수는 동의서와 동일하게 맞추고 뒤에 -01, -02,...를 붙인다.
- ⑤ 착불 우편을 이용하여 아래의 주소로 발송한다. 이때 인터넷 우체국 택배를 이용하여 신청한 후, 이메일(2016klcorpus@gmail.com)로 등기번호를 통지한다(착불 결제용).

2) 구어

- ① 학습자가 산출한 대화, 발표, 토론 등의 원음을 수집한다.
- ② 녹음을 할 때에는 양질의 음성 자료 확보를 위하여 가능하면 보이스 레코더와 같은 녹음기기를 사용한다.
- ③ 하나의 파일에 한 명의 학습자 자료가 녹음되도록 한다. 하나의 파일에 여러 명의 파일을 연이어 녹음한 경우는 학습자별로 파일을 분리한다. 만약, 파일을 분리하지 못할 경우 발화자를 알 수 있도록 녹음된 순서에 맞춰 학습자의 정보를 정리한 후 동의서와 합철한다.
- ④ 파일명은 다음과 같이 국적, 기관명, 수준, 파일 구분을 위한 번호(0001, 0002,...)를 조합하여 부여하고 언더바(_)를 사용하여 순서대로 이어 붙인다.
예) 대만_한국대_1급_0001.wav, 대만_한국대_1급_0002.wav,...
- ⑤ 발표와 같은 1인 발화에 한해서 파일명을 학습자 동의서에 적어 학습자 정보를 파악할 수 있도록 한다.
- ⑥ 파일을 이메일(2016klcorpus@gmail.com)로 발송한 후 동의서 원본은 착불 우편으로 발송한다.

3. 학습자 동의서 수집

- ▶ 모든 자료는 자료 제공과 사용에 관한 학습자의 동의서를 받은 후 수집한다.
- ▶ 동의서는 같은 학기 중의 동일한 학습자라도 자료 수집 시마다 매번 받는 것을 원칙으로 한다. 예를 들어 한 학습자가 여름 학기 과제 작문 한 편, 기말 쓰기 시험의 작문 한 편을 제공할 경우에도 2번의 동의서를 각각 받도록 한다. 다만, 자료 수집의 효율성이나 기관 내 사정 등으로 인해 매번 받는 것이 어려울 경우 처음 수집할 때 받은 동의서와 짝을 맞출 수 있도록 학습자의 이름, 수준, 학급(반) 정보를 시험지에 적는다. 구어 자료는 학습자

정보와 파일명을 함께 기록한다.

- ▶ 동의서를 수합한 후 누락된 항목이 있는지 확인한다. 국적 정보와 같이 수집 교사가 확인 가능한 항목이 누락된 경우 적어 넣는다.

- [주의]
1. 동의서는 학습자의 모국어 또는 학습자가 가장 이해하기 쉬운 언어로 번역된 것을 배부하여 자료 수집 목적과 개인 정보 제공 등에 관한 사항을 충분히 이해할 수 있도록 한다. 그 밖의 학습자가 추가적으로 궁금해 하는 사항이 있을 경우에는 설명해 준다.
 2. 학습자가 수기로 적고 사인하도록 할 수 있도록 출력하여 배포한다. 동의서와 개인 정보는 학습자의 개인 정보 보호를 위하여 자료 분류 후 절취하여 따로 보관하게 된다.
 3. 구어 자료 수집 시 2인 이상의 대화 자료를 녹음할 경우 참여 학습자 각각에게 동의서를 받는다.

[참고] 한국어 학습자 말뭉치 자료의 유형 및 수집 방법

1. 횡적 말뭉치(국내 대학 및 이주민 교육 기관)

1) 문어

(1) 수집 원칙

- 수업 활동 또는 수업 과제, 시험에서 작성한 쓰기 자료를 수집한다.
- 하나의 완결된 글이 되도록 한다.
- 모어화자(가족, 교사 포함) 혹은 동료의 피드백이 이루어지지 않은 글이어야 한다.
- 사전 사용이 배제된 작문을 원칙으로 한다.
- 보기 글을 그대로 베껴 쓰거나 주어진 다량의 어휘를 기반으로 한 작문은 되도록 배제한다.
- 구축 본부에서 제시한 기획 과제를 활용할 경우 학습자의 수준에 맞추어 제시된 글의 종류와 주제로 작문을 하게 하여 이를 수집한다(☞수집 과제는 요청 시에 별도 제공).

(2) 수집 방법

- ① 교육과정 내 과제 작문 수집
 - 각 교육기관의 교육과정 실러버스에 이미 포함되어 있는 작문을 활용하여 이를 수집함. 글의 종류 및 주제는 각 기관의 교육과정에 따름
- ② 성취도 평가 수집
 - 각 교육기관의 성취도 평가(중간 및 기말) 쓰기 시험에 포함된 작문을 활용하여 이를 수집함. 글의 종류 및 주제는 각 기관의 성취도 평가에 따름
- ③ 교육과정 외 프로젝트를 위한 기획 작문 수집(수집 가능 기관)
 - 각 등급에 맞추어 수업 시간(1시간) 내에 다음과 같은 글의 종류와 주제로 작문을 하게 하여 이를 수집함. 세부 주제는 종적 말뭉치의 과제 활동 자료를 참고함

수준	추천 글의 종류	기타	주제
초급	체험적 글(생활문)	일기, 편지, 이메일 등	소개(자신, 가족 등), 취미, 한국생활, 주말, 계절, 좋아하는 음식, 학교생활, 여행, 일상사 등
중급	체험적 글(생활문) 설명적 글(설명문)	안내문, 감상문 등	소개(가족, 문화, 풍습 등), 취미, 여행, 여가생활, 한국생활, 추억, 영화, 만남, 직업, 후회, 사회문제(환경문제 등), 등
고급	설명적 글(설명문) 논리적 글(논설문)	기사문, 계시문 등	사회문제, 경제문제, 문화, 예술, 봉사, 갈등 등

2) 구어

(1) 수집 원칙

- 발화를 유도하기 위해 유인물 등을 기반으로 할 수는 있으나 그대로 읽는 것은 배제하며 읽은 후 이야기를 할 때에는 되도록 보지 않고 발화하도록 한다.
- 해당 등급의 중반 혹은 그 이후에 발화된 것을 녹음하는 것을 원칙으로 한다.
- 교사는 되도록 자신의 발화를 통제하고, 학생이 자신의 발화를 유지할 수 있도록 안내자 정도의 역할을 하도록 한다.
- 학습자가 단어나 구를 활용한 단답형의 대답만 하지 않도록 하며 과제의 주제 또는 교사의 질문과 관련된 내용을 충분히 발화할 수 있도록 기다려 준다(☞수집 과제는 요청 시에 별도 제공).

(2) 수집 방법

① 교육과정 내 담화 수집

- 학습자와 학습자 간의 역할극 혹은 간단한 토론 등과 같이 학습자와 학습자 간의 2인 발화의 경우에는 각 학습자가 녹음하여 이를 교사에게 전송하게 하여 이를 수집함
- 토론 등 다인 발화의 경우에는 발화자의 정보를 확인 가능하도록 비디오로 녹화하거나 녹음 및 전사자를 일치시킬 것을 권유함
- 이의 담화 유형과 주제, 시간은 각 기관의 교육과정에 따름

② 성취도 평가 수집

- 교사와 학생, 혹은 학생과 학생 간에 이루어지는 성취도 평가의 담화를 수집함
- 이의 담화 유형과 주제, 시간은 각 기관의 성취도 평가에 따름

③ 졸업작담회, 말하기 대회 등의 자료 수집

- 졸업작담회, 말하기 대회 등 공식적인 구어 담화를 수집함. 비디오 녹화를 권유함

④ 교육과정 외 담화 자료로 본 프로젝트를 위한 기획 발화 수집(수집 가능 기관)

- 각 등급에 맞추어 수업 시간 내 또는 수업 시간 외에 다음과 같은 주제로 발표 또는 인터뷰 활동을 통해 자료를 수집함. 발화 시간은 5-10분 이내로 함. 세부 주제는 종적 말뭉치의 과제 활동 자료를 참고함

수준	담화 유형	주제
초급	발표, 인터뷰	소개(자신, 가족 등), 취미, 한국생활, 주말, 계절, 좋아하는 음식, 학교생활, 여행, 일상사 등
중급	발표, 인터뷰	소개(가족, 문화, 풍습 등), 취미, 여행, 여가생활, 한국생활, 추억, 영화, 만남, 직업, 후회, 사회문제 (환경문제 등), 등
고급	발표, 인터뷰	사회문제, 경제문제, 문화, 예술, 봉사, 갈등 등

2. 종적 말뭉치 (해당 기관)

1) 문어

(1) 수집 원칙

- 학습자들이 작문을 시작하기 전에 주제와 글의 장르를 충분히 이해한 후 글을 쓸 수 있도록 설명하며, 필요한 경우 쓰기 전 활동처럼 관련 질문들을 하시면서 잠시 이야기를 나눌 수 있음
- 초급 단계의 경우 10문장 이상 쓰도록 지도함(중급 15-20문장, 고급 20문장 이상)
- 완성되지 않은 작문 자료의 경우 말뭉치로 구축하기가 어려우므로 주제에 관해 완결된 글을 쓰도록 함
- 작문은 사전이나 교재 등의 자료를 참고하지 않고 쓸 수 있도록 하며, 가능하다면 숙제로 주지 않고 함께 모여서 쓸 수 있도록 함

(2) 수집 방법

- 각 등급에 맞추어 수업 시간(1시간) 외에 다음과 같은 글의 종류와 주제로 작문을 하게 하여 이를 수집함

2) 구어

(1) 수집 원칙

- 자료 수집을 시작하기 전에 발화를 유도하기 위한 도입 질문 등을 통해 학습자가 발화 주제에 대해 충분히 생각한 후 이야기할 수 있도록 유도함
- 학습자가 발화를 충분히 할 수 있도록 시간적 여유를 줌
- 학습자가 발화를 이어가지 못할 경우 간단한 유도 발화를 해서 발화를 이어갈 수 있도록 도움을 줄 수 있음
- 모든 발화에 대하여 과도하게 맞장구를 치거나 학습자가 말하는 도중에 끼어들지 않도록 함

- 학습자가 오류를 범하더라도 일일이 교정해 주지 않음
- 학습자가 발화를 이어가기 위해 특정 어휘나 표현을 생각하느라고 머뭇거리거나 다소 긴 휴지가 지속될 경우 교사가 먼저 말해 주지 않고 학습자가 스스로 발화를 이어가도록 기다려 줌

(2) 수집 방법

- 교육과정 외 담화(본 프로젝트를 위한 기획 발화) 수집
 - 각 등급에 맞추어 수업 시간 내에 다음과 같은 주제로 발표를 하게 하여 이를 수집함. 발표는 5-10분 이내로 함
 - 각 등급에 맞추어 다음과 같은 주제로 교사가 인터뷰를 하여 이를 수집함. 인터뷰는 5-10분 이내로 함. 교사는 되도록 자신의 발화를 통제하고, 학생이 자신의 발화를 유지할 수 있도록 안내자 정도의 역할을 함

학습자 말뭉치 종적 자료 수집 과제(일반)

1. 문어 수집

수집 시기	문제	수준
02주차	자기소개를 해보십시오. 이름이 무엇입니까? 어느 나라 사람입니까? 무엇을 합니까? 무엇을 좋아합니까?	초급
04주차	여러분의 가족에 대해 쓰십시오. 누가 있습니까? 무슨 일을 합니까? 무엇을 좋아합니까?	
06주차	여러분은 토요일이나 일요일에 무엇을 합니까? 어디에 갑니까? 누구를 만납니까? 여러분의 주말 이야기를 쓰십시오.	
08주차	여러분은 어떤 선물을 받고 싶습니다? 왜 그 선물을 받고 싶습니까? 선물에 대한 글을 쓰십시오.	
10주차	어느 계절을 좋아합니까? 왜 그 계절을 좋아합니까? 그 계절에 특별히 무엇을 합니까? 좋아하는 계절에 대해서 글을 쓰십시오.	
12주차	여러분은 뭐 하는 것을 좋아합니까? 왜 그것을 좋아합니까? 그것을 얼마나 자주 합니까? 여러분의 취미에 대해서 쓰십시오.	
14주차	여러분이 가장 좋아하는 친구는 누구입니까? 그 친구는 무엇을 합니까? 왜 그 친구를 좋아합니까? 여러분이 가장 좋아하는 친구를 소개해 보십시오.	
16주차	여러분은 어디에 자주 갑니까? 왜 그곳에 자주 갑니까? 거기에서 무엇을 합니까? 여러분이 자주 가는 장소에 대해서 쓰십시오.	
18주차	여러분은 올해 무엇을 하고 싶습니까? 왜 그것을 하고 싶습니까? 2011년에 하고 싶은 것에 대해 쓰십시오.	
20주차	여러분은 어디에 여행을 가 봤습니까? 그것에서 무엇을 했습니까? 어땠습니까? 여러분의 여행 경험에 대해 쓰십시오.	
22주차	여러분은 10년 후에 어떻게 살고 싶습니까? 그 이유는 무엇입니까? '10년 후의 나의 계획'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다. ○ 10년 후에 어떻게 살고 싶은가? ○ 그 이유는 무엇인가? ○ 무엇을 준비해야 하는가?	중급
24주차	여러분이 소중하게 생각해서 사랑하는 물건은 무엇입니까? '내가	

수집 시기	문제	수준
	<p>가장 아끼는 물건'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 가장 아끼는 물건은 무엇인가?</p> <p>○ 왜 그 물건을 아끼는가?</p> <p>○ 어떻게 그 물건을 가지게 되었는가?</p>	
26주차	<p>여러분은 취미로 무엇을 배우고 싶습니까? '내가 취미로 배우고 싶은 것'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 취미로 배우고 싶은 것은 무엇인가?</p> <p>(※ 한국어를 배우고 싶다는 내용은 쓰지 마십시오.)</p> <p>○ 왜 그것을 배우고 싶은가?</p> <p>○ 그것을 배운 후에 무엇을 하고 싶은가?</p>	
28주차	<p>여러분은 늦잠을 자거나 누워서 책을 보는 것과 같은 고치고 싶은 생활 습관이 있습니까? '고치고 싶은 나의 생활 습관'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 나의 나쁜 생활 습관</p> <p>○ 습관 때문에 생기는 불편하거나 안 좋은 점</p> <p>○ 습관을 고치기 위해 해야 할 일</p>	
30주차	<p>잊지 못할 추억'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 어떤 추억인가요?</p> <p>○ 왜 지금까지 기억에 남아 있는가?</p> <p>○ 언제 그 추억이 떠오르는가?</p>	
32주차	<p>갖고 싶은 직업'이라는 제목으로 글을 써 보십시오. 단 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 직업명, 하는 일, 그 일을 하려는 이유, 그 일에 필요한 조건</p>	
34주차	<p>나의 성격'이라는 제목으로 글을 써 보십시오. 단 아래에 제시된 내용에 모두 포함되어야 합니다.</p> <p>○ 성격의 특징, 장점과 단점, 고치고 싶은 부분과 그 이유</p>	
36주차	<p>내가 생각하는 행복'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 행복을 위해서 어떤 노력을 하는가?</p> <p>○ 언제 행복하다고 느끼는가?</p>	

수집 시기	문제	수준
	○ 행복은 무엇이라고 생각하는가?	
38주차	<p>‘내가 좋아하는 책’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 가장 좋아하는 책은 무엇인가? ○ 그 책은 어떤 내용인가? ○ 그 책을 좋아하는 이유는 무엇인가?</p>	
40주차	<p>여러분은 어떤 사람처럼 되고 싶습니까? 왜 그 사람처럼 되고 싶습니까? ‘내가 닮고 싶은 사람’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>○ 닮고 싶은 사람은 누구인가? ○ 왜 그 사람처럼 되고 싶은가? ○ 그 사람처럼 되기 위해서 어떻게 해야 하는가?</p>	
42주차	<p>1)~4)의 내용은 ‘피로를 예방하려면 네 가지를 실천하라’는 글의 소재입니다. 이 소재를 이용하여 글을 쓰십시오.</p> <p>‘피로를 예방하려면 네 가지를 실천하라’</p> <p>○ 체질에 맞는 음식 ○ 수면의 질 ○ 적당한 운동 ○ 긍정적인 사고</p>	
44주차	<p>올바른 인터넷 사용 태도’에 대한 자신의 견해를 서술하십시오. 단, 아래 제시한 <올바른 인터넷 사용 태도의 예> 중에서 세 가지를 선택하여 쓰되, 각각의 태도를 지키지 않았을 경우에 나타나는 부작용의 예를 포함해야 합니다.</p> <p><올바른 인터넷 사용 태도의 예></p> <p>○ 상대방의 의견 존중하기 ○ 타인의 사생활 보호하기 ○ 의견 차이 인정하기 ○ 바른 언어 사용하기 ○ 정확한 정보 올리기</p>	고급
46주차	<p>다음 글을 읽고, ‘현대 사회에서 바람직한 신문의 기능’에 대한 자신의 견해를 서술하십시오. 단 아래에 제시한 기능 중에서 두 가지 이상을 선택하여 쓰되, 그 기능이 현대 사회에 중요하다고 생각하는 이유를 포함해야 합니다.</p> <p><신문의 기능></p> <p>○ 사건 보도</p>	

수집 시기	문제	수준				
	<ul style="list-style-type: none"> ○ 여론 조성 ○ 정보 제공 ○ 소통의 분위기 조성 					
48주차	<p>다음 글을 읽고 '감시 카메라 설치 확대'에 대한 자신의 견해를 서술하십시오. (찬성하거나 반대하는 입장에서 하나를 선택하여 서술 할 것. 단 아래 제시된 각 입장의 논거 중 두 개 이상을 제시할 것.)</p> <div> <p>최근 들어 각종 범죄가 급증하면서 감시 카메라 설치가 사회적 문제로 대두되고 있다. 지금까지 감시 카메라는 은행이나 지하 주차장 등에 주로 설치되어 있었으나 이제는 설치 장소를 대폭 확대하자는 것이다. 이러한 감시 카메라 설치 확대에 어떻게 생각하는가?</p> <table> <tr> <th>찬성</th> <th>반대</th> </tr> <tr> <td> 사회 안전 유지 범죄 예방 인권보다 공인이 우선 </td> <td> 개인의 사상활동 침해 범죄 예방 효과 불분명 가해자의 인권 보호 </td> </tr> </table> </div>	찬성	반대	사회 안전 유지 범죄 예방 인권보다 공인이 우선	개인의 사상활동 침해 범죄 예방 효과 불분명 가해자의 인권 보호	
찬성	반대					
사회 안전 유지 범죄 예방 인권보다 공인이 우선	개인의 사상활동 침해 범죄 예방 효과 불분명 가해자의 인권 보호					
50주차	<p>여러분은 성공이 무엇이라고 생각하십니까? 그리고 그러한 성공을 이루기 위해 필요한 것이 무엇이라고 생각하십니까? 이와 관련된 자신의 견해를 서술하십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 내가 생각하는 성공이란 무엇인가? ○ 그것을 이루기 위해 필요한 것은 무엇인가? ○ 그 이유는 무엇인가? 					
52주차	<p>여러분은 무엇이 선의의 거짓말이라고 생각하십니까? 어떤 경우에 그런 거짓말을 할 수 있다고 생각하십니까? 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>< 선의의 거짓말이란 ></p> <ul style="list-style-type: none"> ○ 선의의 거짓말이란 무엇인가? ○ 선의의 거짓말은 언제 필요한가? ○ 선의의 거짓말이 가질 수 있는 문제점은 무엇인가? 					

수집 시기	문제	수준
54주차	<p>학교에서는 음악이나 미술과 같은 예술 교육이 이루어지고 있습니다. 이러한 예술 교육이 왜 필요하다고 생각합니까? 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다.</p> <p>< 예술 교육의 필요성 ></p> <p>○ 예술 교육이 왜 필요한가?</p> <p>○ 예술 교육을 통해 얻을 수 있는 효과는 무엇인가?</p>	
56주차	<p>자연을 그대로 보존해야 한다는 주장과 인간을 위해 자연을 개발해야 한다는 주장이 있습니다. 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p><자연 보존과 자연 개발></p> <p>○ 자연 보존과 자연 개발 중 어느 것이 더 중요하다고 생각하는가?</p> <p>○ 그렇게 생각하는 이유는 무엇인가? (2가지 이상 쓰시오.)</p>	
58주차	<div> <p>오늘날 직업에 대한 생각은 크게 두 가지로 나뉘는 것 같다. 하나는 여러 방면으로 사회에 도움을 주거나, 공헌할 수 있는 직업을 택해 봉사하는 마음으로 일하고, 그것을 통해 얻어지는 대가로 자신과 가정을 꾸려 나가는 것이다. 다른 하나는 사회에 대한 봉사나 공헌보다는 일에 대한 자기 만족과, 욕구 충족, 충분한 대가에 더 큰 비중을 두는 경우이다.</p> <p>전자의 경우, 일이 힘들거나 보수가 적다 하더라도 일에 대한 보람과 긍지 때문에 쉽게 그 일을 그만두거나 직업을 바꾸려 생각은 하지 않는다. 하지만 후자의 경우는 일에 대한 즐거움이나 자기 만족, 충분한 보상이 뒤따르지 않는다고 판단될 때는 언제라도 직장을 옮길 마음의 준비가 되어 있다. 전자의 경우에는 사회를 안정시키는 데에 기여를 하지만 보수적 경향으로 사회적 분위기를 다소 침체시킬 수도 있다. 후자의 경우에는 생동감은 있으나 급격한 변화로 안정감을 잃어버릴 위험이 많고, 이런 변화 속에 적응하지 못하는 이들은 사회 변화의 뒷전으로 밀려날 수밖에 없게 된다.</p> </div>	

수집 시기	문제	수준
	위의 글에 나타난 두 가지 유형의 직업관 중 자신의 생각은 어느 쪽인지 말하고, 그 이유를 설득력 있게 글로 나타내시오.	
60주차	<p>현대 사회는 빠르게 세계화·전문화되고 있습니다. 이러한 현대 사회의 특성을 참고하여, ‘현대 사회에서 필요한 인재’에 대해 아래의 내용을 중심으로 자신의 생각을 쓰십시오.</p> <p>○ 현대 사회에서 필요한 인재는 어떤 사람입니까?</p> <p>○ 이러한 인재가 되기 위해서 어떤 노력이 필요합니까?</p>	

2. 구어 수집

수집 시기	글의 종류	주제	수준
02주차	인터뷰	소개(자신, 가족 등)	초급
04주차	발표	취미	
06주차	인터뷰	주말	
08주차	발표	한국생활	
10주차	인터뷰	계절	
12주차	발표	좋아하는 음식	
14주차	인터뷰	학교생활	
16주차	발표	여행	
18주차	인터뷰	일상(사)	
20주차	발표	선물	
22주차	인터뷰	소개(고향, 문화, 풍습 등)	중급
24주차	발표	스트레스	
26주차	인터뷰	여가생활	
28주차	발표	추억	
30주차	인터뷰	명절	
32주차	발표	영화	
34주차	인터뷰	만남	

수집 시기	글의 종류	주제	수준
36주차	발표	진로와 직업	
38주차	인터뷰	후회	
40주차	발표	환경 문제	
42주차	인터뷰	성공적인 삶	고급
44주차	발표	경제문제	
46주차	인터뷰	문화	
48주차	발표	갈등	
50주차	인터뷰	예술	
52주차	발표	학교 교육	
54주차	인터뷰	봉사	
56주차	발표	현대인의 생활	
58주차	인터뷰	결혼	
60주차	발표	남성과 여성	

학습자 말뭉치 이주민 자료 수집 과제 (결혼이주민, 이주노동자)

1. 문어

종적 자료 수집 시기	문제	수준
02주차	자기소개를 해보십시오. 이름이 무엇입니까? 어느 나라 사람입니까? 무엇을 합니까? 무엇을 좋아합니까?	초급
04주차	여러분의 가족에 대해 쓰십시오. 누가 있습니까? 무슨 일을 합니까? 무엇을 좋아합니까?	
06주차	여러분은 토요일이나 일요일에 무엇을 합니까? 어디에 갑니까? 누구를 만납니까? 여러분의 주말 이야기를 쓰십시오.	
08주차	여러분은 어떤 선물을 받고 싶습니까? 왜 그 선물을 받고 싶습니까? 쓰십시오.	
10주차	어느 계절을 좋아합니까? 왜 그 계절을 좋아합니까? 그 계절에 특별히 무엇을 합니까?	
12주차	여러분은 뭐 하는 것을 좋아합니까? 왜 그것을 좋아합니까? 그것을 얼마나 자주 합니까? 여러분의 취미에 대해서 쓰십시오.	
14주차	여러분이 가장 좋아하는 친구는 누구입니까? 그 친구는 무엇을 합니까? 왜 그 친구를 좋아합니까? 여러분이 가장 좋아하는 친구를 소개해 보십시오.	
16주차	여러분은 어디에 자주 갑니까? 왜 그곳에 자주 갑니까? 거기에서 무엇을 합니까? 여러분이 자주 가는 장소에 대해서 쓰십시오.	
18주차	여러분은 올해 무엇을 하고 싶습니까? 왜 그것을 하고 싶습니까?	
20주차	여러분의 고향은 어디입니까? 고향을 소개하는 글을 쓰십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다.	중급
22주차	여러분은 10년 후에 어떻게 살고 싶습니까? 그 이유는 무엇입니까? '10년 후의 나의 계획'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다.	

종적 자료 수집 시기	문제	수준
	<ul style="list-style-type: none"> ○ 10년 후에 어떻게 살고 싶은가? ○ 그 이유는 무엇인가? ○ 무엇을 준비해야 하는가? 	
24주차	<p>여러분이 소중하게 생각해서 사랑하는 물건은 무엇입니까? ‘내가 가장 아끼는 물건’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 가장 아끼는 물건은 무엇인가? ○ 왜 그 물건을 아끼는가? ○ 어떻게 그 물건을 가지게 되었는가? 	
26주차	<p>‘한국의 첫인상’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 한국에 언제 왔는가? ○ 시내, 길거리는 어떤 모습이었는가? ○ 한국 사람들은 어땠는가? ○ 한국 음식은 어땠는가? ○ 가장 기억에 남는 것은 무엇인가? 	
28주차	<p>여러분은 고치고 싶은 생활 습관이 있습니까? '고치고 싶은 나의 생활 습관'이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 나의 나쁜 생활 습관 ○ 습관 때문에 생기는 불편하거나 안 좋은 점 ○ 습관을 고치기 위해 해야 할 일 	
30주차	<p>‘잊지 못할 추억’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 어떤 추억인가요? ○ 왜 지금까지 기억에 남아 있는가? ○ 언제 그 추억이 떠오르는가? 	
32주차	<p>‘나의 한국 생활’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 한국에 온 지 얼마나 되었는가? ○ 왜 한국에 오게 되었는가? ○ 한국에서 가장 재미있었던 일이 무엇인가? ○ 한국에서 가장 힘들었던 일이 무엇인가? 	

종적 자료 수집 시기	문제	수준
34주차	<p>‘살고 싶은 집’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 어디에 살고 싶은가? ○ 어떤 집에 살고 싶은가? 왜 그런가? ○ 집은 어떻게 꾸미고 싶은가? ○ 집에서 누구와 무엇을 하고 싶은가? 	
36주차	<p>고향의 음식을 소개하는 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> ○ 음식 이름 ○ 주로 언제 먹는 음식인가? ○ 어떻게 만드는가? ○ 한국 음식과 비슷한 음식이 있는가? 어떤 점이 비슷하고 어떤 점이 다른가? 	
38주차	<p>여러분은 어떤 사람처럼 되고 싶습니까? 왜 그 사람처럼 되고 싶습니까? ‘내가 닮고 싶은 사람’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 닮고 싶은 사람은 누구인가? ○ 왜 그 사람처럼 되고 싶은가? ○ 그 사람처럼 되기 위해서 어떻게 해야 하는가? 	
40주차	<p>취업을 하려고 합니다. 무슨 일을 하고 싶은지 생각해 보고 자기 소개서를 쓰십시오.</p> <ul style="list-style-type: none"> ○ 살아온 과정 ○ 성격의 장단점 ○ 지금까지의 경험 또는 경력 ○ 앞으로의 계획 	
42주차	<p>‘내가 생각하는 행복’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 행복을 위해서 어떤 노력을 하는가? ○ 언제 행복하다고 느끼는가? ○ 행복은 무엇이라고 생각하는가? 	고급
44주차	<p>‘절약과 저축’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> ○ 절약하기 위해 무엇을 하는가? ○ 저축을 하고 있는가? 왜 그런가? 	

종적 자료 수집 시기	문제	수준
	○ 돈을 모으면 무엇을 하고 싶은가?	
46주차	<p>‘텔레비전이 우리 생활에 미치는 영향’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <p>○ 텔레비전을 자주 보는가? 왜 그런가?</p> <p>○ 무슨 프로그램을 자주 보는가? 왜 그런가?</p> <p>○ 텔레비전이 우리 생활에 미치는 긍정적인 영향은 무엇인가? 부정적인 영향은 무엇인가?</p>	
48주차	<p>‘효과적인 자녀 교육법’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <p>○ 현재 자녀가 있는가?</p> <p>○ 자녀가 말을 듣지 않을 때 어떻게 하는가? (현재 자녀가 없는 경우, 부모의 말을 듣지 않는 아이를 어떻게 하면 좋을 까?)</p> <p>○ 자녀와 대화를 잘하려면 어떻게 해야 하는가?</p> <p>○ 어떤 부모가 되고 싶은가?</p> <p>○ 자녀 교육을 어떻게 하고 싶은가?</p>	
50주차	<p>여러분은 성공이 무엇이라고 생각하십니까? 그리고 그러한 성공을 이루기 위해 필요한 것이 무엇이라고 생각하십니까? 이와 관련된 자신의 견해를 서술하십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다.</p> <p>○ 내가 생각하는 성공이란 무엇인가?</p> <p>○ 그것을 이루기 위해 필요한 것은 무엇인가?</p> <p>○ 그 이유는 무엇인가?</p>	
52주차	<p>여러분은 무엇이 선의의 거짓말(좋은 거짓말)이라고 생각하십니까? 어떤 경우에 그런 거짓말을 할 수 있다고 생각하니까? 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>< 선의의 거짓말이란 ></p> <p>○ 선의의 거짓말이란 무엇인가?</p> <p>○ 선의의 거짓말은 언제 필요한가?</p> <p>○ 선의의 거짓말이 가질 수 있는 문제점은 무엇인가?</p>	
54주차	<p>‘노후 준비’라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p>	

종적 자료 수집 시기	문제	수준
	<ul style="list-style-type: none"> ○ 노후 준비가 왜 필요한가? ○ 노후를 위해 무엇을 준비해야 하는가? ○ 여러분은 노후를 어떻게 준비하고 있는가? 	
56주차	<p>‘칭찬은 고래도 춤추게 한다’는 말처럼 칭찬에는 강한 힘이 있습니다. 그러나 칭찬이 항상 긍정적인 영향을 주는 것은 아닙니다. 아래의 내용을 중심으로 칭찬에 대한 자신의 생각을 쓰십시오.</p> <ul style="list-style-type: none"> ○ 칭찬이 미치는 긍정적인 영향은 무엇입니까? ○ 부정적인 영향은 무엇입니까? ○ 효과적인 칭찬의 방법은 무엇입니까? 	
58주차	<p>‘여성의 사회적 지위와 역할’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> ○ 여성이 일을 해야 한다고 생각하는가? 왜 그런가? ○ 한국에서 여성의 사회적 지위는 어떻다고 생각하는가? 고향과 비교해서 높은 편인가? 낮은 편인가? ○ 여성이어서 좋은 점 혹은 좋지 않은 점이 있다고 생각하는가? 	
60주차	<div style="border: 1px solid black; padding: 10px;"> <p>오늘날 직업에 대한 생각은 크게 두 가지로 나뉘는 것 같다. 하나는 여러 방면으로 사회에 도움을 주거나, 공헌할 수 있는 직업을 택해 봉사하는 마음으로 일하고, 그것을 통해 얻어지는 대가로 자신과 가정을 꾸려나가는 것이다. 다른 하나는 사회에 대한 봉사나 공헌보다는 일에 대한 자기만족과, 욕구 충족, 충분한 대가에 더 큰 비중을 두는 경우이다.</p> <p>전자의 경우, 일이 힘들거나 보수가 적다고 하더라도 일에 대한 보람과 긍지 때문에 쉽게 그 일을 그만두거나 직업을 바꾸려 생각은 하지 않는다. 하지만 후자의 경우는 일에 대한 즐거움이나 자기 만족, 충분한 보상이 뒤따르지 않는다고 판단될 때는 언제라도 직장을 옮길 마음의 준비가 되어 있다. 전자의 경우에는 사회를 안정시키는 데에 기여를 하지만 보수적 경향으로 사회적 분위기를 다소 침체시킬 수도 있다. 후자의 경우에는 생동감은 있으나 급격한 변화로 안정감을 잃어</p> </div>	

종적 자료 수집 시기	문제	수준
	<div> <p>버릴 위험이 많고, 이런 변화 속에 적응하지 못하는 이들은 사회 변화의 뒷전으로 밀려날 수밖에 없게 된다.</p> </div> <p>위의 글에 나타난 두 가지 유형의 직업관 중 자신의 생각은 어느 쪽인지 말하고, 그 이유를 설득력 있게 글로 나타내시오. (200자 내외)</p>	

2. 구어

수집 시기	발화 유형	주제	수준
02주차	인터뷰	자기소개	초급
04주차	발표	가족	
06주차	인터뷰	주말	
08주차	발표	선물(받은 선물, 준 선물, 받고 싶은 선물, 주고 싶은 선물 등)	
10주차	인터뷰	계절	
12주차	발표	취미	
14주차	인터뷰	친구	
16주차	발표	자주 가는 장소	
18주차	인터뷰	올해 계획	
20주차	발표	고향	
22주차	인터뷰	나의 꿈과 미래 계획	중급
24주차	발표	소중한 것들	
26주차	인터뷰	한국(첫인상, 한국에 대한 여러 가지 생각 등)	
28주차	발표	습관	
30주차	인터뷰	추억(어린 시절, 학창 시절 등)	
32주차	발표	나의 한국 생활	
34주차	인터뷰	살고 싶은 집	
36주차	발표	음식(고향 음식, 한국 음식, 좋아하는 음식, 싫어하는 음식 등)	

수집 시기	발화 유형	주제	수준
38주차	인터뷰	존경하는 인물	
40주차	발표	나의 삶 (성격, 경험 및 경력, 앞으로의 계획)	
42주차	인터뷰	내가 생각하는 행복	고급
44주차	발표	경제문제	
46주차	인터뷰	텔레비전	
48주차	발표	자녀 교육	
50주차	인터뷰	성공적인 삶	
52주차	발표	거짓말	
54주차	인터뷰	노후	
56주차	발표	칭찬	
58주차	인터뷰	남성과 여성	
60주차	발표	직업	

학습자 말뭉치 이주민 자료 수집 과제(중도입국청소년)

1. 문어

- 20주, 40주, 60주차에는 제시된 그림을 보면서 이야기를 만들어 쓰도록 한다. 학생들의 수준에 따라 이야기를 만들어 쓴 후 이야기 내용과 관련된 학생들의 생각이나 경험담을 함께 쓰도록 할 수 있다. 말하기에서도 동일한 자료를 활용하므로 학습자의 수준을 고려하여 말하기 또는 쓰기를 먼저 하고 관련 내용을 확장함으로써 작문과 발화를 최대한 많이 할 수 있도록 유도한다.

종적 자료 수집 시기	문제	수준
02주차	이름이 뭐예요? 어느 나라 사람이에요? 몇 학년이에요? 무엇을 좋아해요? 자기소개를 해 보세요.	초급
04주차	누가 있어요? 무슨 일을 해요? 무엇을 좋아해요? 여러분의 가족에 대해 쓰세요.	
06주차	여러분의 하루 일과에 대해서 쓰세요. 아침에 몇 시에 일어나요? 그리고 무엇을 해요?	
08주차	여러분은 토요일이나 일요일에 무엇을 해요? 어디에 가요? 누구를 만나요? 여러분의 주말 이야기를 쓰세요.	
10주차	여러분은 뭐 하는 것을 좋아해요? 왜 그것을 좋아해요? 그것을 얼마나 자주 해요? 여러분의 취미에 대해서 쓰세요.	
12주차	어느 계절을 좋아해요? 왜 그 계절을 좋아해요? 그 계절에 특별히 무엇을 해요? 좋아하는 계절에 대해서 글을 쓰세요.	
14주차	여러분이 가장 좋아하는 친구는 누구예요? 여러분이 가장 좋아하는 친구에게 편지를 쓰세요.	
16주차	여러분은 어떤 선물을 받고 싶어요? 왜 그 선물을 받고 싶어요? 지금까지 받은 선물 중에 가장 좋은 선물이 뭐예요? 선물에 대해서 글을 쓰세요.	
18주차	여러분의 고향은 어디예요? 고향에서 무엇이 유명해요? 고향을 소개하는 글을 쓰세요.	
20주차	그림을 보고 이야기를 순서대로 써 보세요.	

종적 자료 수집 시기	문제	수준
22주차	여러분은 무슨 음식을 좋아해요? 무슨 음식을 좋아하지 않아요? '나의 식생활'이라는 제목으로 글을 쓰세요.	중급
24주차	여러분이 소중하게 생각하는 물건은 뭐예요? 왜 그 물건이 소중한해요? 그 물건을 어떻게 가지게 되었어요? '내가 가장 아끼는 물건'이라는 제목으로 글을 쓰세요.	
26주차	무슨 과목을 좋아해요? 왜 그래요? 여러분이 알고 있는 좋은 공부 방법이 있어요? '나의 공부 방법'이라는 제목으로 글을 쓰세요.	
28주차	여러분은 고치고 싶은 생활 습관이 있어요? 습관 때문에 생기는 불편한 점이 있어요? '고치고 싶은 나의 생활 습관'이라는 제목으로 글을 쓰세요.	
30주차	20년 후에 여러분은 어디에 있을까요? 무엇을 하고 있을까요? 20년 후 자신의 모습을 상상해 보고 '자신의 미래 모습'이라는 제목으로 글을 쓰세요.	
32주차	여러분은 어디에 여행을 가 봤어요? 누구하고 갔어요? 거기에서 무엇을 했어요? 어땠어요? 여러분의 여행 경험에 대해 쓰세요. (가족 여행, 수학여행, 체험 학습 등)	
34주차	한국에 언제 왔어요? 한국에서 가장 재미있는 일은 뭐예요? 한국에서 가장 힘든 일은 뭐예요? '나의 한국 생활'이라는 제목으로 글을 쓰세요.	
36주차	여러분은 어떤 사람처럼 되고 싶어요? 왜 그 사람처럼 되고 싶어요? '내가 닮고 싶은 사람'이라는 제목으로 글을 쓰세요.	
38주차	음식 이름이 뭐예요? 주로 언제 먹는 음식이에요? 어떻게 만 들어요? 한국 음식과 비슷한 음식이 있어요? 어떤 점이 비슷하고 어떤 점이 달라요? 고향 음식을 소개하는 글을 쓰세요.	
40주차	그림을 보고 이야기를 순서대로 써 보세요.	
42주차	<div> <p>저는 심각한 고민이 하나 있어요. 저는 3학년인데 키가 140cm이고 몸무게는 455kg예요. 저는 키도 작은 것 같고 뚱뚱한 것 같아요. 저도 가수나 탤런트처럼 더 날씬하고 키도 크고 싶어요. 그래서 요즘 다이어트를 하고 있어요. 그리고 저는 눈이 작고 쌍꺼풀이 없어요. 그래서 성형 수술을 하고 싶어요.</p> </div>	고급

종적 자료 수집 시기	문제	수준
	여러분도 자신의 외모에 대해 고민을 해 봤어요? 여러분은 자신의 아름다움을 위해 어떤 노력을 하고 있어요? 글에서 읽은 친구의 고민에 대한 여러분의 생각을 써 보세요.	
44주차	여러분은 스트레스를 잘 받는 편이에요? 여러분이 지금 받고 있거나 예전에 받았던 스트레스는 뭐예요? 스트레스를 받았을 때 어떻게 해결했어요? ‘스트레스’라는 제목으로 글을 써 보세요.	
46주차	여러분은 텔레비전을 자주 봐요? 무슨 프로그램을 자주 봐요? 텔레비전이 우리 생활에 미치는 긍정적인 영향은 뭘까요? 부정적인 영향은 뭘까요? ‘텔레비전’이라는 제목으로 글을 쓰세요.	
48주차	여러분은 거짓말을 한 적이 있어요? 무슨 거짓말을 했어요? 거짓말을 한 후 어떤 일이 일어났어요? ‘거짓말’이라는 제목으로 글을 쓰세요.	
50주차	여러분은 언제 행복해요? 그리고 언제 슬퍼요? ‘행복한 일과 슬픈 일’이라는 제목으로 글을 쓰세요.	
52주차	학교에서 일어난 큰 실수나 사고를 생각해 보세요. 기억에 남는 일을 쓰세요.	
54주차	여러분은 무슨 놀이(게임, 수업 활동 등)를 좋아해요? 어떻게 해요? 여러분들이 좋아하는 놀이 방법을 소개하는 글을 쓰세요.	
56주차	선생님과 부모님께 칭찬을 받아 본 적이 있지요? 언제 칭찬을 받았어요? 무슨 일로 칭찬을 받았어요? 기분이 어땠어요? 칭찬 받은 일에 대해서 글을 써 보세요.	
58주차	지금까지 읽은 책 중에 가장 재미있었던 책이 뭐예요? 무슨 내용이에요? 읽고 무슨 생각을 했어요? ‘내가 가장 좋아하는 책’이라는 제목으로 글을 써 보세요.	
60주차	그림을 보고 이야기를 순서대로 써 보세요.	

2. 구어

- 2주차, 12주차, 22주차, 32주차, 42주차, 52주차에는 제시한 읽기 텍스트를 큰소리로 낭독하도록 한 후 텍스트와 관련된 질문을 하는 방법으로 학생들과 주제에 관한 대화를 간단히 나눈 후에 학생들에 관한 이야기로 대화를 확장해 나간다.
- 20주, 40주, 60주차에는 제시된 그림을 보면서 이야기를 만들어 보도록 한다. 이야기를 다 만들고 난 후에는 교사가 이야기와 관련된 질문을 하여 이야기에 관한 학생의 의견이나 경험 등을 자유롭게 말하도록 한다.

수집 시기	발화 유형	주제	수준
02주차	인터뷰	자기소개	초급
04주차	발표	가족	
06주차	인터뷰	하루 일과	
08주차	발표	주말	
10주차	인터뷰	취미	
12주차	발표	★ 계절 관련 텍스트 읽기 ★ 계절	
14주차	인터뷰	친구	
16주차	발표	선물(받은 선물, 준 선물, 받고 싶은 선물, 주고 싶은 선물 등)	
18주차	인터뷰	고향	
20주차	발표	★ 그림 보고 이야기 만들기 ★ 이야기와 관련한 자유 대화	
22주차	인터뷰	★ 식사 관련 텍스트 읽기 ★ 나의 식사 생활	중급
24주차	발표	소중한 것들	
26주차	인터뷰	공부와 시험	
28주차	발표	생활 습관	
30주차	인터뷰	미래	
32주차	발표	★ 여행 관련 텍스트 읽기 ★ 여행	
34주차	인터뷰	한국	
36주차	발표	존경하는 인물	
38주차	인터뷰	한국 음식과 고향 음식	

수집 시기	발화 유형	주제	수준
40주차	발표	★ 그림 보고 이야기 만들기 ★ 이야기와 관련한 자유 대화	고급
42주차	인터뷰	★ 중등 외모 관련 텍스트 읽기 ★ 성격과 외모	
44주차	발표	스트레스	
46주차	인터뷰	텔레비전	
48주차	발표	거짓말	
50주차	인터뷰	감정	
52주차	발표	★ 실수 관련 텍스트 읽기 ★ 실수	
54주차	인터뷰	놀이	
56주차	발표	칭찬	
58주차	인터뷰	독서	
60주차	발표	★ 그림 보고 이야기 만들기 ★ 이야기와 관련한 자유 대화	

<2주차 읽기 자료>²¹⁾

안녕하세요? 저는 송안나입니다. 대한초등학교 학생입니다. 저는 1학년 5반입니다. 우즈베키스탄에서 왔어요. 지금은 대림동에 살아요. 우리 집은 학교 근처에 있어요. 저는 컴퓨터 게임을 좋아해요. 만나서 반갑습니다.

<12주차 읽기 자료>

한국에는 봄, 여름, 가을, 겨울 사계절이 있습니다.

봄에는 날씨가 따뜻합니다. 산과 들에 예쁜 꽃이 많이 피니다. 사람들은 꽃놀이를 갑니다.

여름에는 날씨가 더워집니다. 비도 많이 옵니다. 사람들은 넓은 바다로 여행을 갑니다. 우리는 수영을 하고 물총 싸움도 합니다. 시원한 팔빙수도 먹습니다.

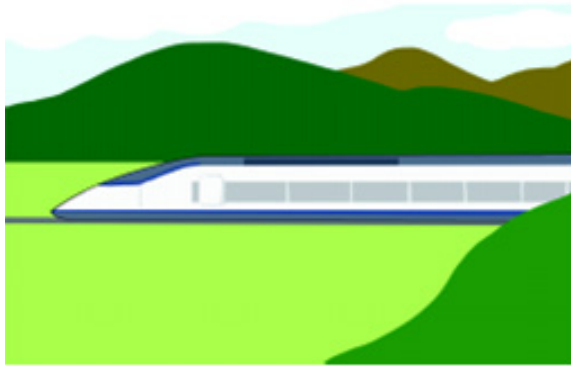
가을에는 날씨가 선선해집니다. 산에 가서 단풍 구경을 합니다. 빨간 단풍이 참 아름답습니다. 맛있는 과일도 많이 먹을 수 있습니다.

겨울에는 날씨가 추워집니다. 바람도 많이 불고 눈도 옵니다. 겨울에는 따뜻한 옷을 입고 장갑도 껍니다.

저는 사계절 중에서 추운 겨울을 제일 좋아합니다. 눈사람도 만들도 눈싸움도 할 수 있습니다.

21) 중도입국청소년 자료 수집을 위한 과제는 『초/중/고등학생을 위한 표준한국어 교재』(국립국어원), 『KSL 교육과정 진단도구』(국가평생교육원)의 자료를 발췌하거나 개작하였다. 따라서 과제와 함께 제시되는 텍스트와 그림 자료도 두 자료에서 발췌한 것이다.

<20주차 그림 자료>



<22주차 읽기 자료>

우리 엄마는 항상 ‘무엇을 요리할까?’ 하고 고민하십니다. 왜냐하면 나는 햄이나 고기반찬을 좋아해서 김치나 채소 반찬을 잘 안 먹기 때문입니다.

나는 매일 아침 바빠서 아침을 안 먹고 학교에 갑니다. 점심시간에는 제가 좋아하는 반찬이 없으면 점심을 안 먹고 빵이나 과자를 사 먹으러 매점에 갑니다. 그리고 저녁에는 배가 고파서 한꺼번에 많이 먹습니다.

또 나는 밥보다 햄버거나 치킨, 빵, 과자를 좋아하고 물보다 음료수를 더 좋아합니다. 매일 이렇게 내가 좋아하는 음식을 먹고 싶습니다. 그런데 엄마는 “그런 음식만 먹으면 건강에 안 좋아! 아침을 꼭 먹고 반찬을 골고루 먹어 봐!”라고 말씀하십니다.

나는 왜 내가 좋아하는 음식만 먹으면 안 될까요?

<32주차 읽기 자료>

지난주 토요일에 공주에 갔다 왔다. 오전에 도착해서 먼저 간 곳은 공산성이었다. 공산성은 옛날에 전쟁을 할 때 지은 성이다. 이곳은 경치가 매우 좋고 공주 시내도 잘 보였다.

공산성에서 내려오니 12시였다. 배가 너무 고파서 내려오자마자 공원에서 도시락을 먹었다. 그리고 시청에서 무료로 빌려주는 자전거를 타고 송산리 고분군으로 갔다. 그곳에는 벽화들이 많았다. 옛날 사람들은 무덤에 벽화도 그려 놓고 여러 가지 물건도 넣었다. 신기했다.

2시에 무령왕릉도 갔다. 무령왕릉은 생각보다 정말 컸다. 안에 들어갈 수 없어서 아쉬웠다. 주변에서 사진도 찍고 놀다 보니 오후 3시였다. 정문 옆에는 제기차기와 윷놀이를 할 수 있는 곳도 있었다. 거기서 친구들과 제기차기를 하면서 놀았다. 오늘은 사회 시간에 배웠던 곳에 가서 백제 시대 역사 공부를 할 수 있어서 정말 좋았다.

<40주차 그림 자료>²²⁾



22) 그림은 『엄마와 함께 읽어요. 지식 쏙쏙 만화』(한국간행물윤리위원회, 2010)에서 발췌함.

<42주차 읽기 자료>

저와 제 동생 마리는 쌍둥이 자매입니다. 우리는 머리가 금색이고 피부가 아주 하얗습니다. 키는 다른 친구들에 비해서 아주 큰 편입니다.

우리는 얼굴은 똑같이 생겼지만 성격은 아주 다릅니다. 저는 조용한 성격이라서 나가서 노는 것보다 집에서 책을 읽거나 엄마 일을 돕는 것을 좋아합니다. 그리고 성격이 좀 느린 편이라서 어떤 일을 할 때 천천히 꼼꼼하게 합니다.

그런데 동생은 활발해서 친구들과 같이 운동장에서 노는 것을 좋아합니다. 또한 성격이 급한 편이라서 무슨 일이든지 빨리 하기 때문에 실수를 자주 합니다. 호기심도 많아서 궁금한 것이 있으면 꼭 물어봅니다.

우리는 이렇게 같은 점도 있고 다른 점도 있지만 기쁠 때나 슬플 때나 늘 함께하는 사이좋은 자매입니다.

<52주차 읽기 자료>

오늘 좀 속상했다. 가장 친한 친구 라몬과 싸웠기 때문이다. 오늘 낮에 라몬과 농구를 하다가 내 실수로 라몬이 넘어졌다. 나는 일부러 한 게 아니라서 미안하다는 말을 안 했는데 그것 때문에 화가 많이 났나 보다. 라몬은 농구공을 던져 버리고, 나한테 소리를 질렀다. 그래서 나도 너무 화가 나서 소리를 질렀다. 그리고 나는 라몬과 크게 싸우게 될까 봐 혼자 집으로 와 버렸다. 그런데 오면서 생각해 보니 내가 일부러 그런 것은 아니지만 나도 넘어지면 기분이 나쁠 것 같다. 어떻게 할까 고민하고 있는데 저녁에 라몬에게서 전화가 왔다. 그리고 나에게 먼저 사과를 했다. 그때 나는 라몬에게 너무 미안했다. 내가 먼저 사과할걸 그랬다. 내일 라몬을 만나면 라몬이 좋아하는 과자를 주면서 다시 한 번 사과를 해야겠다.

<60주차 그림 자료>²³⁾

※ 다음 그림을 보고 이야기를 만들어 보세요.



1



2

23) 그림은 '키즈짱 잼쟁동화-접어지는 샘물' 동영상의 주요 장면을 캡처하여 편집한 것임
(<https://www.youtube.com/watch?v=-rHeP6eJKSM>)



3



4



5



6

한국어 학습자 말뭉치 구축을 위한 기획 자료 수집 과제(1차)

1. 문어

- 각 수준별로 다음의 주제와 장르에 따라 글을 쓰도록 하여 수집합니다.
- 다음의 주제를 제시하되 풍부한 글쓰기를 위해 관련 내용을 자유롭게 확장할 수 있도록 합니다.
- 완결된 글이 되도록 하되 글의 길이가 너무 짧지 않은지 확인합니다.
(현재까지의 수집 결과에 따르면 1급 최소 50어절[7~10문장] 이상, 2-3급 100어절 이상[15~20문장], 4급 이상 150어절 이상이 평균 길이임)
- 가족이나 친구, 동료, 한국어 모어 화자의 피드백이 이루어지지 않은 글이 되도록 합니다.

수준	주제	장르
1급	가족, 친구, 취미, 성격, 좋아하는 것과 싫어하는 것, 꿈	생활문
2급	나의 가족, 나의 친구, 나의 이웃	생활문
3급	기억에 남는 여행, 여행 경험	기행문
4급	추천하고 싶은 여행지	설명문
5급	내가 생각하는 성공적인 삶(성공적인 삶이란 무엇인가?)	논설문
6급	결혼에 대한 나의 생각(결혼을 해야 할까? 아니면 혼자 사는 게 좋을까? 한다면 어떤 사람하고 해야 할까? 국제결혼은 어떨까?)	논설문

2. 구어

- 말하기는 특정 주제에 한정하지 않고 다음과 같은 흐름으로 발화를 진행하도록 합니다. 이때 전개 부분에서는 [나의 현재-나의 과거-나의 미래]의 순으로 이야기를 진행할 수 있도록 적절한 때에 교사가 관련 질문을 해 준다.

단계	발화 내용	비고
도입	■ 자기소개(이름, 국적 등) - 교사와의 일상적 대화	초급 학습자의 경우 최대한 가능한 주제까지 대화를 이어가도록 함
전개	■ 나의 현재: 취미, 성격, 좋아하는 것, 싫어하는 것, 가족, 친구 ■ 나의 과거: 어린 시절, 학창 시절 중 기억에 남는 일 고향 소개, 추천하고 싶은 장소 ■ 나의 미래: 꿈, 내가 생각하는 성공, 결혼 계획, 결혼에 대한 생각, 죽기 전에 꼭 하고 싶은 일	
마무리	■ 학습자 격려를 위한 피드백 및 감사 인사	

- 발화 시간은 10분 내외로 합니다. 다만, 1급의 경우 어휘량이 상대적으로 부족한 시기이므로 5분 이상 10분 내외로 시간을 조정할 수 있습니다.
- 초반에는 이름, 국적 등에 대한 질문, 간단한 일상 대화를 통해 학습자의 긴장을 풀어 주면서 발화를 시작하도록 합니다.
- 발화가 시작된 후 교사는 주로 청취자로서 간단한 맞장구를 하며 학습자가 스스로 발화를 이어갈 수 있도록 합니다. 다만, 학습자가 발화를 이어가지 못하거나 도중에 끊길 경우 자연스럽게 발화를 이어갈 수 있도록 관련 주제로 의 확장을 위한 질문을 해 줍니다.
- 이때 교사의 질문이 지나치게 상세하고 길어지면 학습자가 단답형의 대답을 하게 되는 경우가 많으므로 학습자가 발화를 이어가는 데에 필요한 단서를 제공하는 정도의 질문을 합니다.

2015-2022년 한국어 학습자 말뭉치 구축 사업을 위한
학습자 자료 이용 동의서(일반)

국립국어원에서 한국어교육의 질적 향상을 위해 학습자들의 언어 자료(말뭉치)를 수집하여 활용하는 사업(사업 수행: 연세대학교 산학협력단)을 추진하고 있습니다. 여러분이 제공한 자료는 한국어 교수 방법 개선, 한국어 교재 개발, 한국어 교육 분야 및 인접 학문 분야의 연구에 사용됩니다. 이 연구에 참여하는 분들은 경제적인 손해나 신체적 위험이 없습니다. 만약 참여를 원하지 않을 때에는 참여 의사를 철회할 수 있습니다. 또한 수집하는 개인 정보는 본 사업의 목적 외로는 사용되지 않으며, 비밀 유지를 위하여 식별할 수 없는 형태로 사용될 것입니다. 감사합니다.

문의처: 연세대학교 산학협력단 02-2123-4199

☐ 저는 위의 내용을 충분히 이해하였으며 다음의 정보와 말하기/쓰기 자료를 제공하고, 쓰기 원문/말하기 음성 녹음 자료 전체의 공개와 연구 목적의 사용을 허락합니다.

날짜 _____

이름 _____ (서명)

✂-----

다음은 연구를 위한 자료로 활용될 정보입니다. 개인 신상 정보는 비밀이 보장되며 외부로 유출되지 않습니다. (가능하면 한국어로 응답해 주세요. 필요하면 영어를 사용해도 좋습니다.)

1. 성별: ☐ F ☐ M
2. 나이: _____
3. 현재 등급: _____
4. 국적: _____ (※ 교포 여부 ☐ 교포 ☐ 외국인)
5. 제1 언어: _____
6. 한국어 학습 기간(한국어를 얼마 동안 공부했습니까?): _____ 년 _____ 개월
(예. 1년 3개월)
7. 한국에서의 거주 기간(한국에서 얼마 동안 살았습니까?): _____ 년 _____ 개월
(예. 1년 3개월)
8. 한국어 학습 목적
☐ 진학 ☐ 취업 ☐ 거주 ☐ 취미 ☐ 결혼 ☐ 기타 ()
9. 직업: _____
10. 한국어 외의 사용 가능 외국어(잘하는 언어 순서대로 쓰시오): _____

2015-2022년 한국어 학습자 말뭉치 구축 사업을 위한 학습자 자료 이용 동의서
(이주민 자료/종적 자료)

국립국어원에서 한국어교육의 질적 향상을 위해 학습자들의 언어 자료(말뭉치)를 수집하여 활용하는 사업(사업 수행: 연세대학교 산학협력단)을 추진하고 있습니다. 여러분이 제공한 자료는 한국어 교수 방법 개선, 한국어 교재 개발, 한국어 교육 분야 및 인접 학문 분야의 연구에 사용됩니다. 이 연구에 참여하는 분들은 경제적인 손해나 신체적 위험이 없습니다. 만약 참여를 원하지 않을 때에는 참여 의사를 철회할 수 있습니다. 또한 수집하는 개인 정보는 본 사업의 목적 외로는 사용되지 않으며, 비밀 유지를 위하여 식별할 수 없는 형태로 사용될 것입니다. 감사합니다.

문의처: 연세대학교 산학협력단 02-2123-4199

☐ 저는 위의 내용을 충분히 이해하였으며 다음의 정보와 말하기/쓰기 자료를 제공하고, 쓰기 원문/말하기 음성 녹음 자료 전체의 공개와 연구 목적의 사용을 허락합니다.

날짜 _____
이름 _____ (서명)
(학습자와의 관계 _____)

다음은 연구를 위한 자료로 활용될 정보입니다. 개인 신상 정보는 비밀이 보장되며 외부로 유출되지 않습니다. (가능하면 한국어로 응답해 주세요. 필요하면 영어를 사용해도 좋습니다.)

학교명: _____ 학교 _____ 학년
입학/편입 학년: _____ 학년
(☞ 해당 사항이 없는 경우 쓰지 않아도 됩니다.)

1. 성별: ☐ F ☐ M
2. 출생년: _____ 년(예. 1989년)
3. 현재 등급: _____ (TOPIK: _____)
4. 국적: _____ (※ 교포 여부 ☐ 교포 ☐ 외국인)
5. 제1 언어: _____
6. 한국어 학습 기간(한국어를 얼마 동안 공부했습니까?): _____ 년 _____ 개월
(예. 1년 3개월)
 - 6-1. 학습 기관명: _____
 - 6-2. 사용 교재명: _____

7. 7-1. 입국년월: _____년_____월(예. 2015년 2월)
 7-2. 한국에서 얼마 동안 살았습니까?: _____년_____월(예. 1년 3개월)
8. 한국어 학습 목적
☐ 진학 ☐ 취업 ☐ 거주 ☐ 취미 ☐ 결혼 ☐ 기타 ()
9. 직업: _____
10. 한국어 외의 사용 가능 외국어(잘하는 언어 순서대로 쓰시오):
11. 평상시에 가장 많이 사용하는 언어는 무엇입니까? _____
12. 한국어로 대화하는 상대는 누구입니까?
☐ 부모님 ☐ 시부모님 ☐ 남편 ☐ 친척
☐ 이웃 ☐ 친구 ☐ 선생님 ☐ 직장 동료 ☐ 기타 ()
13. 한국어로 말하는 시간은 얼마나 됩니까?
☐ 거의 없음 ☐ 하루 1시간~하루 3시간 ☐ 하루 3시간~ 5시간 ☐ 하루 5시간 이상
14. 한국어로 듣는 시간은 얼마나 됩니까?
☐ 거의 없음 ☐ 하루 1시간~하루 3시간 ☐ 하루 3시간~ 5시간 ☐ 하루 5시간 이상
15. 한국어로 나오는 방송 매체(TV, 라디오, 인터넷 동영상)를 보는 시간은 얼마나 됩니까?
☐ 거의 없음 ☐ 하루 1시간~하루 3시간 ☐ 하루 3시간~ 5시간 ☐ 하루 5시간 이상

2023-2024년 한국어 학습자 말뭉치 자료 수집 지침

※ 국립국어원에서 주관하는 ‘2024년 한국어 학습자 말뭉치 구축’ 사업 수행을 위해 학습자의 쓰기 자료 수집을 요청드립니다. 협조에 응해 주셔서 진심으로 감사드립니다.

1. 자료 수집 대상 및 수집 자료

- ▶ 대상: 한국어 교육 기관의 학습자
- ▶ 자료: 학습자가 산출한 작문 자료 (문어)
 - 학습자 1인당 최대 5편 제출 가능
 - 각 작문 자료는 100어절 이상 작성
- ▶ 수집 유형: 특정 주제에 대한 기획 자료 수집
- ▶ 수집 시기: 2024년 6월 - 2024년 12월

2. 작문 자료 주제

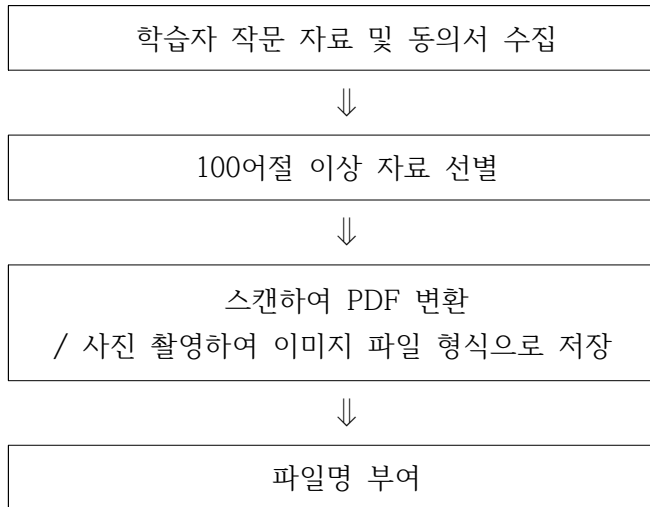
- 각 수준별로 다음의 주제에 따라 글을 쓰도록 하여 수집합니다.
- 학습자 1인당 최대 5편의 주제에 대해 글을 쓸 수 있습니다.
- 다음의 주제를 제시하되 풍부한 글쓰기를 위해 관련 내용을 자유롭게 확장할 수 있도록 합니다. (세부 질문을 넣은 예시 과제를 별첨해 드립니다. 필요 시 사용해 주십시오.)
- 하나의 완결된 글로, 글의 길이가 100어절이 넘어야 합니다.
- 가족이나 친구, 동료, 한국어 모어 화자의 피드백이 이루어지지 않은 글이 되도록 합니다.

※ 1번과 2번은 모든 수준의 공통 주제입니다.(★표) 많은 수집을 부탁드립니다.

<수준별 수집 주제>

수준	주제 번호	주제
초급	1	★내가 가장 좋아하는 것과 싫어하는 것 (일, 행동, 사람, 물건, ...)
	2	★과학 기술의 발전이 인간의 생활에 미치는 영향
	3	10년 후의 나의 계획
	4	내가 가 본 여행지 (위치, 교통편, 날씨, 음식, 간 곳, ...)
	5	올해의 계획 (하고 싶은 일, 이유)
중급	1	★내가 가장 좋아하는 것과 싫어하는 것 (일, 행동, 사람, 물건, ...)
	2	★과학 기술의 발전이 인간의 생활에 미치는 영향
	3	내가 생각하는 행복과 성공
	4	기후 위기와 해결 방안
	5	효율적인 시간 관리 방법
고급	1	★내가 가장 좋아하는 것과 싫어하는 것 (일, 행동, 사람, 물건, ...)
	2	★과학 기술의 발전이 인간의 생활에 미치는 영향
	3	기후 위기와 해결 방안
	4	인구 문제
	5	역사적 사건이나 인물 소개

3. 자료 수집 절차



- ① 학습자가 손으로 쓴 작문지와 동의서를 수집합니다.
- ② 수집된 자료가 100어절 이상인지 확인합니다.
- ③ 자료를 스캔하거나 사진 촬영하여 PDF 파일로 변환합니다. (PDF 변환이 어려울 경우, 이미지 파일 형식(jpg, png, bmp)으로 변환) 이때 화면이 흐릿하여 글자가 잘 보이지 않는 경우가 있으니 유의합니다.
- ④ 해당 자료의 출처를 파악할 수 있도록 동의서와 작문 자료에 다음과 같은 파일명을 부여합니다. (주제 번호는 2p 참조)

자료	파일명	예시
동의서	Ag_급수_국적_이름	Ag_3급_대한민국_김서준
작문 자료	급수_국적_이름_T주제번호	3급_대한민국_김서준_T1

※ 이름은 한글/영문 무관함

4. 학습자 동의서 수집

- ▶ 모든 자료는 자료 제공과 사용에 관한 학습자의 동의서를 받은 후 수집합니다.
- ▶ 동의서를 수합한 후 누락된 항목이 있는지 확인합니다. 국적 정보와 같이 수집 교사가 확인 가능한 항목이 누락된 경우 적어 넣습니다.

- [주의] 1. 동의서는 학습자의 모국어 또는 학습자가 가장 이해하기 쉬운 언어로 번역된 것을 배부하여 자료 수집 목적과 개인 정보 제공 등에 관한 사항을 충분히 이해할 수 있도록 합니다. 그 밖의 학습자가 추가적으로 궁금해 하는 사항이 있을 경우에는 설명해 줍니다.
2. 학습자가 수기로 적고 사인하도록 할 수 있도록 출력하여 배포합니다.
3. 학습자 한 명당 동의서는 한 번만 받습니다. 처음 수집할 때 받은 동의서와 짝을 맞출 수 있도록 쓰기 자료의 파일명에 들어가는 학습자 이름과 동의서 파일명에 들어가는 학습자 이름을 반드시 동일하게 적어 주십시오.
4. 사업 종료 시까지 동의서 원본을 보관하여 주시기 바랍니다.

(주)이르테크

주소: (34188) 대전 유성구 계룡로 114, 907호(봉명동, 대전유성 BYC빌딩)

전화번호: 042-824-1108

담당자: 정해영 선임연구원

5. 학습자에게 전달할 자료

① 학습자 개인 정보 수집·이용 및 저작권 이용 동의서

- 자료 수집 동의서입니다. 총 9개 언어로 번역되어 나갑니다.

(영어, 중국어, 일본어, 베트남어, 태국어, 러시아어, 스페인어, 인도네시아어, 몽골어)

② 학습자 매뉴얼 (한국어 학습자 말뭉치 수집 참여 가이드)

- 쓰기 참여 가이드입니다. 주제 및 쓰기 방법 등이 제시되어 있습니다.
- 직접 제출, 온라인 제출 매뉴얼이 각각 제시되어 있습니다.
- 초급, 중급, 고급 각 급수별로 다르게 제작됩니다.
- 초급에 한해 위의 9개 언어로 번역되어 나갑니다.

⇒ 학습자 매뉴얼은 직접 전달하셔도 되고, 설명으로 대체하셔도 됩니다.

아래의 내용을 별첨해 드립니다.

수집에 참고 부탁드립니다.

별첨1. 수준별 쓰기 과제 예시 (필요하신 경우에 사용하십시오.)

별첨2. 100어절 쓰기 예시

별첨3. 100어절 쓰기 양식 (필요하신 경우에 사용하십시오.)

별첨1. 수준별 쓰기 과제 예시

수준	번호	주제
초급	1	★내가 가장 좋아하는 것과 싫어하는 것 (일, 행동, 사람, 물건, ...)
	2	★과학 기술의 발전이 인간의 생활에 미치는 영향
	3	10년 후의 나의 계획
	4	내가 가 본 여행지 (위치, 교통편, 날씨, 음식, 간 곳, ...)
	5	올해의 계획 (하고 싶은 일, 이유)
중급	1	★내가 가장 좋아하는 것과 싫어하는 것 (일, 행동, 사람, 물건, ...) ○ 여러분이 가장 좋아하는 것은 무엇입니까? ○ 여러분은 어떤 경우를 가장 싫어합니까?
	2	★과학 기술의 발전이 인간의 생활에 미치는 영향 (스마트폰, 인터넷, 로봇, 인공지능, ...) ○ 과학 기술의 발전이 미치는 긍정적인 영향은 무엇입니까? ○ 과학 기술의 발전이 미치는 부정적인 영향은 무엇입니까? ○ 과학 기술의 발전에 따른 문제 해결 방안은 무엇입니까?
	3	내가 생각하는 행복과 성공 ○ 여러분이 생각하는 행복은 무엇입니까? ○ 여러분은 성공이 무엇이라고 생각합니까? ○ 행복과 성공은 서로 어떤 관계가 있습니까? ○ 그 이유는 무엇입니까?
	4	10년 후의 나의 계획 ○ 여러분은 10년 후에 어떻게 살고 싶습니까? ○ 그 이유는 무엇입니까? ○ 무엇을 준비해야 합니까?
	5	효율적인 시간 관리 방법 ○ 언제 시간 관리가 필요합니까? ○ 효율적으로 시간을 쓸 수 있는 방법은 무엇입니까?
고급	1	★내가 가장 좋아하는 것과 싫어하는 것 (일, 행동, 사람, 물건, ...) ○ 여러분이 가장 좋아하는 것은 무엇입니까? ○ 여러분은 어떤 경우를 가장 싫어합니까?
	2	★과학 기술의 발전이 인간의 생활에 미치는 영향 (스마트폰, 인터넷, 로봇, 인공지능, ...) ○ 과학 기술의 발전이 미치는 긍정적인 영향은 무엇입니까? ○ 과학 기술의 발전이 미치는 부정적인 영향은 무엇입니까? ○ 과학 기술의 발전에 따른 문제 해결 방안은 무엇입니까?
	3	기후 위기와 해결 방안 (폭염, 폭우, 지구 온난화, 해수면 상승, ...) ○ 기후 위기가 우리 생활에 미치는 영향은 무엇입니까? ○ 기후 위기의 원인은 무엇입니까?

		○ 기후 위기를 해결할 수 있는 방안은 무엇입니까?
	4	인구 문제 ○ 현재 무엇이 가장 심각한 인구 문제라고 생각합니까?(저출산, 고령화, 1인 가구, ...) ○ 그 문제의 원인은 무엇입니까? ○ 그 문제의 해결 방법은 무엇입니까?
	5	역사적 사건이나 인물 소개 ○ 여러분 나라 또는 다른 나라의 역사적인 사건이나 인물을 소개해 주십시오. ○ 그 사건이 갖는 역사적 의미는 무엇입니까? / 그 인물을 통해 얻을 수 있는 교훈이 있습니까?

별첨2. 100어절 쓰기 예시

<원고지> 105어절

[illegible]

<A4 용지> 111어절

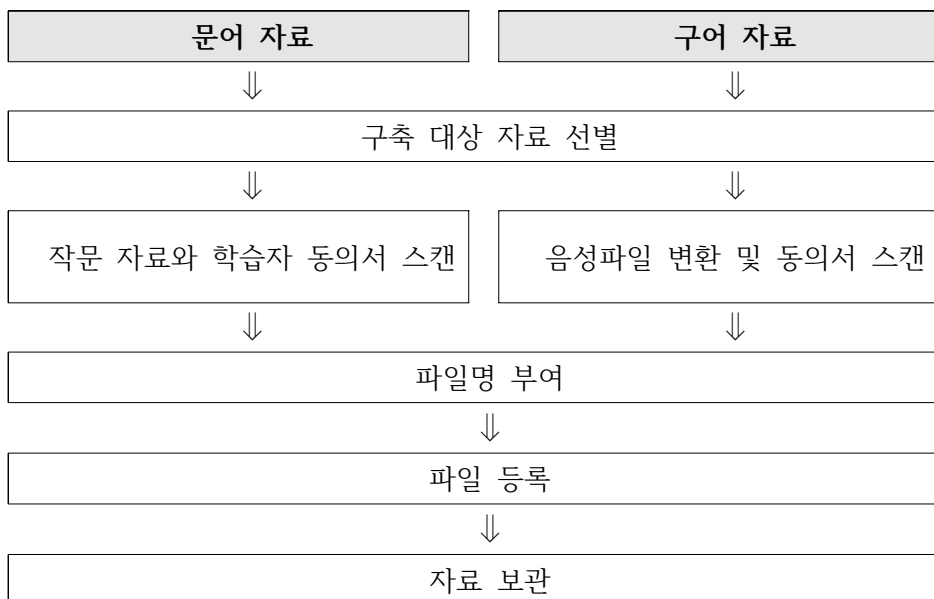
부모가 속아서를 읽는 이유는 크게 두 가지가 있습니다. 첫째는 책을 통해 위안을 얻고자 함이고, 둘째는 책을 통해 깨달음을 얻고자 함입니다. 위안은 곧, '내가 그래도 아이를 잘 키우고 있구나', '우리 아이만 이런 것은 아니구나'하는 것입니다. 이런 내용은 매우 일반적인 이야기, 가령, '무정어를 사용하지 말라', '결과만 보고 아이를 혼내지 말라'하는 것 등의 책의 내용과 관련됩니다. 깨달음은 이와는 반대로 내가 알고 싶었던 이야기, 내가 알고 있던 것과는 다른 이야기와 관련됩니다. '아이가 자만해도 내버려 두라', '아이가 적극적으로 말아도 걱정하지 말라' 등의 내용입니다. 본 책은 이 두 가지 내용 중 어느 한 가지에 저중 하지 않고 적절의 내용을 잘 조합하고 있어 흥미롭게 읽을 수 있었습니다.

결론적으로 이 책은 아이를 키우는 부모에게 도움이 될 수 있는 책이라고 생각합니다.

한국어 학습자 말뭉치 자료 처리 지침

1. 자료 처리 절차

- 자료 처리는 파일을 전산화하여 말뭉치 자료로서 본격적인 구축과 가공 작업을 하기 위한 전처리 단계로 다음과 같은 절차에 따라 처리한다.



2. 단계별 자료 처리 지침

1) 말뭉치 구축 대상 자료 선별

- 말뭉치 구축을 위해서는 IRB 규정에 따라 학습자의 서명이 완료되고 자료의 활용을 위해 필요한 개인 정보가 빠짐없이 입력이 되어야 한다. 그 외에도 다음과 같은 기준으로 우선적으로 구축할 자료를 선정하도록 한다.

문어	구어
○ 학습자 동의서에 서명한 자료	
○ 학습자 동의서의 개인 정보 모두 입력된 항목 선정	
○ 동일 학습자의 자료 2개 이하로 제한	
○ 영어권, 일본어권 자료/1, 5, 6급 단계의 자료 우선 선정	
○ 완결된 텍스트 작문 자료 선정	○ 완결된 담화 단위의 발화 자료 선정
○ 텍스트의 길이 평균 100어절 이상의 자료 선정. 단, 숙달도 단계를 고려하여 1, 2급은 50어절 내외의 자료를 포함함	○ 발화 길이 2분 이상의 자료 선정
○ 복사 또는 스캔 파일의 경우 화질이 좋은 자료 선정	○ 음질이 좋은 자료 선정
	○ 교사의 개입이 많지 않고 학습자의 발화가 중심인 자료를 우선 선정

2) 학습자 동의서 확인 및 스캔

- 학습자 동의서와 작문 자료가 제대로 짝을 이루고 있는지 확인한다. 학습자 동의서나 작문 자료 어느 한쪽이라도 누락된 자료는 구축 불가 자료로 분리하여 따로 모은다.
- 자료를 확인하는 과정에서 수집 기관에 문의가 필요한 사항이나 수집 시 주의해야 할 사항이 있을 경우 학습자 동의서 및 자료 관련 특이 사항에 메모를 남긴다.

집수 날짜	기관명	자료 유형	자료 내용	자료 수(수준별)							합계	자료 관련 메모
				1급	2급	3급	4급	5급	6급	6급 이상		
2015.07.14	동국대학교(경주)	일반	여름 학기 중간고사 쓰기 자료								105	
2015.07.21	한남대학교	일반	여름 학기 중간고사 쓰기 자료	13	12	23	8				56	
2015.07.22	한양대학교	일반	여름 학기 중간고사 쓰기 자료	39	71	78	70	43	9		310	
2015.07.23	호남대학교	일반	여름 학기 기말고사 쓰기 자료								124	
2015.07.24	충남대학교	일반	여름 학기 중간고사 쓰기 자료								466	

동의서 급수와 시험 자료의 급수가 맞지 않아 수집 기관에 확인이 필요함

동의서 급수와 시험 자료의 급수가 맞지 않아 수집 기관에 확인이 필요함

3) 일련번호 부여

- 학습자 동의서와 작문 자료에 일련번호를 부여한 후 스캔한다. 동일한 학습자가 작성한 학습자 동의서와 작문 자료는 같은 일련번호를 부여한다. 이때 학습자 동의서가 두 장으로 분리된 경우는 각각을 ‘0001-앞, 0001-뒤’로 처리하고, 작문 자료가 두 장 이상일 경우는 ‘0001-01, 0001-02, 0001-03……’으로 처리한다.

4) 파일명 부여

- 자료의 효율적인 관리를 위하여 자료의 유형과 국적, 수집 기관, 수준 등의 정보가 포함된 파일명을 부여한다. 파일 분류 및 파일명 부여 체계는 다음과 같다.

예) 국내일반종적_문어_중국_○○대_1급_0001_01.txt

자료 코드		학습자 변인 정보 코드				
국내 일반 종적	문어	중국	○○ 대	1급	0001	01
자료 및 학습자 유형		국적	수집 기관: 코드화하여 비공개 처리됨	수준	자료 번호	페이지 번호

구분	범주	설명	항목	코드
자료 코드	자료 및 학습자 유형	학습자의 특성에 따른 분류	일반 종적 이주 학문 목적 (2017년 현재 학문목적은 일반 최고급으로 분류)	국내일반횡적 국내일반종적 국내일반기획 결혼이주횡적 결혼이주종적 결혼이주기획 이주노동횡적 이주노동종적 이주노동기획 중도입국횡적 중도입국종적 중도입국기획 국외횡적 국외종적 국외기획
	자료 유형	자료의 유형을 구분하는 코드 부여	문어(Written) 구어(Spoken)	문어 구어
학습 자 정보 코드	언어권	학습자의 제1 언어를 구분하는 코드 부여	중국어 일본어 베트남어 영어 ...	중국 일본 베트남 영어 ...
	자료 수집 기관	자료 수집 기관명	서울대 경희대 ...	서울대 경희대 ...
	수준	학습자의 수준을 구분하는 코드 부여	1급 2급 3급 4급 5급 6급 최고급	1 2 3 4 5 6 7
	학습자 구분 번호	기관의 학습자 구분을 위한 일련번호	0001 0002 ...	0001 0002 ...

자료 번호	자료 번호	동일한 학습자가 두 개 이상의 자료를 제공할 경우 자료를 구분하기 위한 일련번호	01 02 ...	01 02 ...
----------	----------	---	-----------------	-----------------

한국어 학습자 말뭉치 문어 입력 지침

1. 전체적인 형식 원칙

- 기본적으로 온라인 입력/전사 시스템의 입력 창에서 입력한다.
- 자료를 입력하기 전 표본 정보와 학습자의 개인 정보를 입력한다.
(☞ ‘수집 정보 등록/검증’ 메뉴)
- 학습자가 글 하나를 스스로 완성하였을 경우에만 입력하는 것을 원칙으로 한다. 중간에 채 완성하지 못한 문장은 입력하지 않는다.
- 필적을 알아보기 어려운 것은 일단 가장 가까운 상태로 입력한다.
- 단락을 구분하여, 문장 단위로 입력한다. 단락은 자판의 엔터키로 구분하고, 들여쓰기는 반영되지 않는다.
- 전체 본문 입력이 끝나면 ‘주석 자동 생성’을 클릭하여 본문 주석을 확인하고 이후 개별 마크업을 진행한다.

2. 입력 지침

- 원본의 텍스트를 그대로 입력하는 것을 원칙으로 한다. 철자 오류가 있더라도 원본 그대로 입력한다.

<예> 특히 말할 때 춘대말을 한다는 것이 자주 반말을 말한다.
→ 수정 안 함.

- 원문의 영어와 한자는 모두 유지한다. 한자는 시스템 입력창에서 글자를 선택 후 마우스 오른쪽쪽을 클릭하여 입력한다.
- 띄어쓰기는 어문 규범과 <표준국어대사전>의 표제어에 맞춰 수정하여 입력한다. 원활한 형태소 분석 작업을 위해 띄어쓰기를 정확히 적용한다.
- 분수 표시는 다음과 같이 입력한다.

<예> 1/2, 3/4

- 영문자, 한글 자모, 괄호문자 등은 자판을 사용하여 입력한다.

<예> ㄱ ㄴ ㄷ ㄹ, (1) (2) (3)

- 외국어를 함께 쓴 경우 다음과 같이 원문에 따라 병기한다. 단, 입력과 해석의 용이성을 고려하여 영어와 한자에 한정한다.

<예> 아래의 경우 '바프라이(BARFLY)'로 입력한다.

우리는 술을 마시고 싶으면 '바프라이' 술집에 ~~가~~ ^(BARFLY) ~~가~~ ^가요.

- 숫자와 한글 표기를 함께 쓴 경우 원문에 따라 병기한다.

<예> 아래의 경우 '3(세) 달 전'으로 입력한다. 이때 '3달(세 달)'과 같이 동일한 표기가 두 번 이상 입력되도록 하지 않는다.

^(세) 3달 전에 미국에서 한국까지

- 학습자가 작문 중간에 교정 기호를 사용하거나 교정에 관한 문구를 적어 넣은 경우 이를 반영해서 수정 입력한다. 단, 학습자의 답안에 교사가 같은 색으로 수정 또는 채점을 한 경우, 학습자가 작성하면서 스스로 수정한 것인지 교사가 수정한 것인지 선별해야 한다.

<예> 반 친구도 노래를 잘 볼 수 있어요.
그래서 노래방도 자주 가요.
우리는 함께 때 좋은 기본이 왔는데요.
어떻게 가는지 알아요? 서울까지 피행기를 타야 해요.

우리는 함께 때 ~~좋은~~ ^(반 친구) 좋은 기본이 ~~왔~~ ^왔는데요.
반 친구도 노래를 잘 볼 수 있어요. 그래서 노래방도 자주 ~~가~~ ^가요.
어떻게 가는지 알아요? ^{서울까지} 피행기를 타야 해요.

3. 문장 부호 및 기호류 마크업

- 문장 부호는 원본 그대로 입력하는 것을 원칙으로 한다.
- 문장 부호 및 기호류는 기본적으로 자판 문자(기호)를 입력하며, 한글 워드 프로그램 등에서 사용하는 전각 기호나 반각 기호를 사용하여 입력하지 않도록 한다.
- 문장부호는 학습자가 적어 넣은 대로 입력한다. 즉, 문장부호의 누락이나 생략, 중복 등을 그대로 반영한다.
- 입력이 어려운 문자는 거꾸로 된 물음표(?) 기호를 사용하여 입력한다. 거꾸로 된 물음표(?) 기호는 키보드에 없는 문자, 식별되지 않는 문자 등 기본 자판에서 입력 불가능한 모든 문자와 기호 형태를 의미한다.
 - ‘외국문자’는 영어와 한자 이외의 외국어를 입력할 때 ? 기호 입력 후 마크업할 때 사용한다.

<예> <EX_Alpha>?/?/?/?/?</EX_Alpha>

- ‘식별불가’는 원본에서 다양한 이유로 확인이 어려운 문자나 기호에 대해 ?로 입력 후 마크업한다.

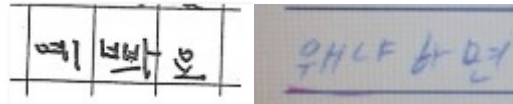
<예> <CNI>?/?</CNI>

- ‘기타기호’는 문장 앞에 붙인 블릿 기호나 다른 특수 기호들을 원본 그대로 입력 후 마크업할 때 사용한다. 키보드에서 한글 자음을 입력 후 ‘한자’ 키를 눌러서 선택하여 입력한다. (‘기타기호’는 원본 그대로 입력하므로 ? 기호를 입력하지 않도록 주의한다.)

<예> 1) <EX_Symbol>『』《》</EX_Symbol>
: [괄호기호] ‘ㄴ’ 입력 후 ‘한자키’ 눌러서 선택
2) <EX_Symbol>※★</EX_Symbol>
: [일반기호] ‘ㅁ’ 입력 후 ‘한자키’ 눌러서 선택
3) <EX_Symbol>m² kg kcal</EX_Symbol>
:[단위기호] ‘ㄹ’ 입력 후 ‘한자키’ 눌러서 선택

- 두벌식 한글과 같이 자판에서 하나의 음절이나 글자로 입력이 불가능한, 우리말에 없는 글자를 입력할 때에는 해당 글자의 위치에 거꾸로 된 물음표(?)를 입력한 후, 구축 도구 내의 ‘한글기호’ 주석을 사용하여 마크업한다.

<예>



1) 예ㅼㅅ 요 : 시스템 '예ㅅ요'로 입력 후 ‘한글기호’ 마크업 처리

- 예<NSS>ㅅ</NSS>요

2) 우ㅼ냐하면 : 시스템 'ㅅ냐하면'으로 입력 후 ‘한글기호’ 마크업 처리

- <NSS>ㅅ</NSS>냐하면

3) 좌우대칭된 ㅅ이 포함된 ‘가’

- <NSS>ㅅ</NSS>

- 기호류 중 자주 사용되는 ‘가운뎃점’은 별도의 마크업 없이 입력/전사 창 아래에서 바로 클릭하여 입력한다.

<예>

<MP> • </MP>

4. 익명성 보장을 위한 개인 정보의 처리

- 학습자들의 이름, 외국인 등록번호, 카드 번호, 전화 번호 등은 신분 보장을 위해 실제 입력 정보에 ‘개인 정보’ 태그로 마크업한다. 이렇게 마크업이 된 정보들은 기호로 자동 처리되어 공개되지 않는다.
- 다음은 마크업 과정에서 각각의 정보를 대신하는 태그들이다.
 - 이름 : 사람 이름, 단체 이름, 학교 이름 등 ⓘ <Privacy_Name> 태그

<예> 저는 태국에서 온 <Privacy_Name>사일롬</Privacy_Name>입니다.

- 전화번호 : 학습자의 휴대폰 번호 등 ⓘ <Privacy_PhoneNum> 태그

- 카드번호 : 학습자의 개인 신용카드 번호 등 ☞ <Privacy_CardNum> 태그
- 기타 : 개인식별 번호(주민등록번호, 외국인등록번호, 학번 등), 주소 등 ☞ <Privacy_Etc> 태그

<예> 저는 서대문구 신촌동 <Privacy_Etc>135</Privacy_Etc> 번
지에 삽니다.

5. 기타

- 스캔 과정에서 일부분이 잘린 경우, 잘린 부분이 한두 글자, 또는 한두 단어 이내로 누가 봐도 추정 가능한 내용일 경우에는 해당 내용을 적어 입력한다. 그 외에는 입력 대상에서 제외한다.

6. 최고급 자료 마크업

- 최고급 자료의 입력은 기존 지침을 동일하게 적용한다.
- 기존의 마크업과 더불어 형식과 내용을 구분하기 위해 아래의 마크업을 사용한다.

	주석	내용	주석 표지
형식 구분	보고서 제목	전체 보고서의 제목	<head>
	본문앞	앞부분의 부속물	<front>
	본문	여러 개의 장절 제목과 본문	<body>
		장절 제목	<title> (기존 주석)
	본문뒤	뒷부분의 부속물	<back>
내용 구분	국문 초록	한글로 된 초록 및 주제어	<Korads>
	외국어 초록	외국어로 된 초록	<Forabs>

	주석	내용	주석 표시
		및 주제어	
	각주 미주	주석 내용	<ft>
	예문 인용	단락이 구분되어 제시된 인용 구절과 예시문	<q>
기타	그림 그래프 도표 설명	문어 입력 과정에서 표, 그림, 그래프 수식 등의 생략을 나타내 주는 표지	<gap reason>

- 각주 미주: 본문과 각주 내용에 각각 각주표시 1),2)를 남기고 해당 각주는 본문뒤, 참고문헌 앞으로 이동 후 <ft> 태그
- 예문 인용: 본문 내에서 문단 구분되어 하나의 단락으로 삽입된 부분을 <q> 태그

<예>

중국인 학자인 劉爲는 당시 조선국내와 대청무역에서 유통하는 銀에 대해서 아래와 같이 설명한다.

“조선은 일본 白銀을 萊銀이라고 불렀는데 그 은의 순도가 80% 이상이다. 그 외에 조선에서 유통하는 백은은 또한 순도가 90% 이상의 청나라산 天銀이 있고 순도가 70%-80% 정도의 丁銀이 있다.”

그러나 위의 인용문에는 틀린 부분이 있다. 첫 번째는 天銀이란 것은 淸國產 은이 아닌 朝鮮產의 순도가 높은 은인 것이다. 1766년에 북경에 다녀온 홍대용과 1783년에 실양에 다녀온

중국인 학자인 劉爲는 당시 조선국내와 대청무역에서 유통하는 銀에 대해서 아래와 같이 설명한다.

<q>

<sentence>“조선은 일본 白銀을 萊銀이라고 불렀는데 그 은의 순도가 80% 이상이다.</sentence>

<sentence>그 외에 조선에서 유통하는 백은은 또한 순도가 90% 이상의 청나라산 天銀이 있고 순도가 70%-80% 정도의 丁銀이 있다.”</sentence>

</q>

그러나 위의 인용문에는 틀린 부분이 있다.

첫 번째는 天銀이란 것은 淸國産 은이 아닌

- 보고서의 장절 제목은 <title>처리한다.

<예>

<body>

<title>1. 서론 </title>

<sentence>본 연구의 목적은...

- 문장 중간에 나타나는 계산식이나 그림은 작업자 메모를 달아 준다.

한국어 학습자 말뭉치 구어 전사 지침

I. 구어 전사 기호 체계

대분류	소분류	기호	예시
발화자 정보	화자 1의 표시	1	<person id=1 sex=M age=20s> 1:
	화자가 불분명할 때	?	P?:
	동시 발화	모두/ 나머지/ 2,3	2,3:네.
억양 단위	하강	.	2:네.
	상승	?	2:어디 갈 거예요?
	약한 상승이나 하강	,	1:그래서 그랬는데 이번에,
	활기, 기운찬 어조	!	1:아!
	억양 단위 경계의 처리	엘란에서 발화 단위를 분할하면 줄 바꿈 처리됨	2:어디 갈 거예요? 1:안 아직까지 그냥 계획만 잡아 났는데, 2:음.
	하나의 억양 단위가 끼 어둠에 의해 끊어진 경우	-	6:기자가 와서 - 2:응. 6:- 그 사람한테 인터뷰를 시작했어.
	두 억양 단위가 휴지 없 이 이어질 경우	&	3:요거는 교수 학습의 개요지, &요 표는, 4:아::,
겹침 현상	겹침 현상	엘란에서 자동 표시	1 03:49.2 03:51.3 네. 다 거짓말이기 때문에. 2 03:50.8

대분류	소분류	기호	예시
			03:52.2 아 왜 거짓말을 하나요? ☞ 발화 겹침이 있음을 알 수 있음
잘 들리지 않는 부분	잘 들리지 않는 부분	<X X>	<X보통X>
	전혀 들리지 않는 부분	<note>안들림</note>	1: 거기까지 <note>안들림</note> 2:<note>안들림</note> > 너무나 거 같더라.
	들리지 않는 음절수만큼	X	2: 근데 그거 진짜 XX해야 되겠더라
전사자의 설명	-	<note>연음되지 않음</note>	1:응. 2:
혼잣말	-	<monologue></monologue>	<monologue>미치겠네.</monologue>
표기 지침	구어의 발음 특성, 개인의 발음 특성, 지역적인 특성 등에 의해 철자법대로 소리 나지 않는 발음(표준 발음이 아닌 경우)	소리나는 대로 적고, 원래의 형태가 없이 내용을 이해하기 어려울 때에는 () 안에 학습자의 발화를, 괄호 밖에는 규범 표기를 밝힘	1: 친구(칭구)와 강남(간남)에 갔습니다(갔슨니다)
	숫자 표기	발음에 따라 한글로 표기	7:오늘 제 동생이 이케 하나 오백 원이라고 사가지고 왔더라구.
	외래어.외국어 표기	발음에 따라 한글로 표기	2:어떻게 이거 크림 장난 아니야. 1:이거도 오리지날 제주도 감귤이 아니야.
	끊어진 단어(불완전하게 발화된 단어)	=	4:사실 학습 자료랑 학= 형태는 떨어져도 되는데.
	한 어절 발화 도중 다른 억양 단위로 전사될 때 조사나 어미에	=	1:주부 우울증, =이라고 말할 수 있겠습니다.

대분류	소분류	기호	예시
	띄어쓰기	맞춤법에 따름	
	축약형	'(apostrophe, 영문따옴표)를 사용해서 두 음소를 연결	사귀'어, 바뀌'어, ...
	표현적 장음	::	1:많은 경우에:: 논문, 저::~ 어::~ 연구는 네이션,
	담화표지	~	1:많은 경우에 논문, 그::~ 어::~ 연구는 네이션,
준음성	웃음	<vocal desc='웃음'>	6:어우 야 <vocal desc='웃음'>
	기침	<vocal desc='기침'>	-
	하품	<vocal desc='하품'>	-
	재채기	<vocal desc='재채기'>	-
	목청 가다듬는 소리(음, 으음)	<vocal desc='목청가다듬는 소리'>	-
	들이마시는 숨(쓰)	<vocal desc='들이마시는숨'>	-
	내쉬는 숨(후우)	<vocal desc='내쉬는숨'>	-
	혀 차는 소리(쯔)	<vocal desc='혀차는소리'>	-
	헛기침(에헴)	<vocal desc='헛기침'>	-
	한숨	<vocal	-

대분류	소분류		기호	예시
			desc='한숨'>	
	노래		<vocal desc='노래'>	-
	웃으면서 말하는 부분		<@ @>	2:[<vocal desc='웃음'>] 너무 좀 <@오버한다.@>
	박수 치면서 말하는 부분		<# #>	5:[우우 <vocal desc='박수'>] <#이리 와 이리 와.#>
	노래를 부르는 부분		<M M>	-
	박수나 손가락 부딪치는 소리		<kinesics desc=' '>	박수 <kinesics desc='박수'>
	대화 흐름에 영향을 주는 전화벨 소리라든지 기타 음성 아닌 소리		<event desc=' '>	<event desc='전화벨소리'>
2차 전사	구어적 변이형	() 안에는 학습자의 발화 를, () 밖에는 철자 보충. 단, 주 등장하 거나 쉽게 원 래 형태로 이 해될 수 있는 것들은 일일이 철자형을 붙여 주지 않음	()	1:브릿지를 한 가닥을 넣어(너) 줬어요.
	발음 오류	한국어 모어 화자의 발음과 음운적으로 구분이 될 정도로 발음에 오류가 있는 경우	()	친구(칭구)와 강남(간남)에 갔습니다(갔슨니다). 같이(가티) 가자. 과반수(화반수)

대분류	소분류		기호	예시
		한국어 모어 화자의 발음과 음성 혹은 변이음의 구분이 모호한 경우	()	가구(가구)<note>‘가’의 ㄱ을 유성음으로 발음 </note> 가구(가구)<note>‘구’의 ㄱ을 무성음으로 발음 </note>
		음운규칙으 로 인해 한국어의 철자대로 발음되지 않으나, 학습자가 이를 철자대로 발음하는 경와 같이 철자 전사를 통해 학습자의 발음 오류를 반영하기 어려운 경우	()	무조건(무조건)<note>경 음화되지 않음</note> 같이(가티)<note>구개음 화되지 않음</note> 신라(신라)<note>자음동 화되지 않음</note> 앞에(앞에)<note>연음되 지 않음</note> 먹는(먹는)<note>자음동 화되지 않음</note>
	외국어, 외래어 발음	외국어나 외래어를 원어에 가깝게 발음할 경우	()	인터뷰(이너뷰)
		한국어 모 어화자와 다른 규범을 할 경 우	()	카페: 현실 발음 [까페], 학습자 발음 [카페] 버스: 현실 발음 [빠스], 학습자 발음 [버스] → 각각 ‘카페(카페)’, ‘버스(버스)’로 처리함

대분류	소분류	기호	예시
	방언형 표시	확실한 방언형(대응하는 표준형 형태소가 없는 것)의 경우 태그 부착	차는 <dia>여일</dia> 있여.
	긴 휴지	(1초 이상의 침은) 0.1초 단위까지 표시	2:{1.2} 그럴까?
	짧은 휴지	한 어절 안에서의 짧은 침은 ‘.’로 표시	2:아::~~ 그리고 어::~~ 남의 의견을 잘 듣고 수용하고 대화..로 타협해야 된다고 하면서,
	인용	<Q Q>	흑자들은 현대 사회에 대하여 <Q불확실성의 시대는 아니다.Q> 라고 말하죠.
	텔레비전 방송이나 강의 등 텍스트 종류 표현	<R R>	1:그 다음에 <R생각과 느낌이 유기적으로 잘 짜여져 조직체를 이룰 때 좋은 글이 될 수 있다R>라고 돼 있어요.
	익명성 보장을 위한 마크업	<name> : 사람 이름, 단체 이름, 학교 이름 등 <social security number> : 주민등록번호 <card-num> : 신용카드 번호 <address> : 주소 <tel-num> : 전화 번호	5: 그게 어찌면 <name1> 선배님이라든지 다른 선배님들 말:: 들은 걸 생각해 보면,
발화 단위 분할(segmen- tation)	분할 단위는 어절 단위 로 한다		내{1.2}/가 (X) 내{1.2}가 (O)

대분류	소분류	기호	예시
entation)			빗금은 분할 단위를 나타내는 단위이다.

II. 항목별 세부 설명

- 한국어 학습자 말뭉치의 구어 전사 지침은 <21세기 세종 한국어 균형 말뭉치>의 전사 지침을 기초로 하되, 비모어 화자로서 한국어 학습자의 구어 자료에서 나타날 수 있는 여러 가지 발음 표기에 대한 지침 등을 보강하여 수정·보완한 것이다.

1. 전체적인 형식과 규정

1) 발화자 표시

- 모든 발화자에 관한 정보는 시스템 등록 시 주석 입력 창에서 발화자의 기본 정보를 선택하여 입력한다.

<예>	1:학습자 2:원어민
-----	----------------

- 본문 전사에서 발화자 정보와 발화자 표시는 반드시 일치해야 하고 발화자가 분명하지 않을 경우에는 필요에 따라서는 ‘모두’나 ‘나머지’ 등의 지칭을 사용할 수 있고, 화자 2와 화자 3이 동시에 말하는 경우는 ‘2,3’으로 표시하기도 한다.

<예>	1:어~ 물건을 바꾸고 싶은데요, 2:네,
-----	----------------------------

- 발화자 표시에 스페이스를 넣지 않는다.

2) 억양 단위

- 구어 자료는 억양 단위 전사를 한다. 다만, 학습에 의한 발화로 모어 화자와 달리 문장 단위 발화가 많고, 발화 길이가 길지 않다. 이러한 특성을 반영하여 통사 구조에 따른 절 단위 혹은 문장 단위의 전사와 억양 단위 전사를 절충하도록 한다.

가. 억양 단위의 개념

- 구어는 문어와는 달리 정보의 흐름이 통사적인 단위로 이루어지지 않는다. 즉, 문어의 기본단위인 문장은 종결어미로 마무리되고 마침표라는 문장부호로 인해 명확하게 그 단위를 설정할 수 있지만, 구어는 종결어미를 사용하여 발화를 끝내는 경우가 많지 않고, 억양이나 휴지 등의 운율적인 요소에 영향을 받기 때문에 기본단위를 운율적인 단위 곧 억양 단위로 본다.
- 억양 단위는 하나의 통일된 억양 운곽에서 나타난 발화의 연속 단위이다. 단위의 시작에서 기본적인 높이(pitch)로 시작하고, 쉼이 나타나며, 빠른 음절의 연쇄가 나타나는 특징이 있고, 단위의 끝에서는 음절이 길어진다.
- 억양 단위의 구분은 다음과 같이 문장부호를 사용한다.

하강 억양 .
 상승 억양 ?
 약한 상승이나 하강 억양 ,
 활기에 넘치는 기운찬 어조(감탄의 끝) !

- 하나의 역량단위 경계에 스페이스 없이 엔터(enter)를 친다.
- 전자 도구에서는 세그멘테이션을 분할하는 것으로 대체한다.

<예> 2: 어디 갈 거예요?
1: 안 아직까지 계획이 없는데,
2: 음.

나. 끌어진 억양단위(붙임표의 사용)

- 계속해서 말을 할 의향이 있는데, 끼어듦을 당해서(혹은 적극적인 호응에 의해서) 말끝이 잘린 경우는 다음의 예와 같이 마침표를 쓰지 않고 붙임표(-)를 사용한다(단위의 끝에서는 앞쪽에만, 단위의 시작에서는 뒤쪽에만 스페이스 있음). 시간적 순서에 의해 표현된 발화를 억양단위로 묶을 수 있게 된다.

<예> 6:그래서 세계의 매스컴에 다 집중이 되면서 기자가 와서 -
2:응.
6:- 그 사람한테 인터뷰를 시작을 했어.

- 한 명이 말을 하는 도중에 말을 끊은 것이 아니라 지속적으로 반응을 하는 경우 그 수가 많더라도 모두 반영한다.

<예> 1:어제 인사동에 갔는데 -
 2:네.
 1:- 길에서 공연을 하고 있어서 -
 2:네.
 1:- 보다가 -
 2:네.
 1:- 배고파서 호떡을 사 먹었어요.

다. 억양단위의 연속성

- 두 억양단위가 휴지 없이 빨리 이어지는 경우 뒤의 발화 앞에 띄어쓰기 없이 & 기호를 붙인다.

<예> 3:요거는 교수 학습의 개요지,
 &요 표는,
 4:아::,

3) 겹침 현상

- 겹침 발화의 표시는 겹친 부분을 따로 표시하지 않고 전사 도구를 통해 표시되는 발화 시간을 함께 제시한다²⁴⁾.

4) 잘 들리지 않는 부분

- 잘 들리지 않는 부분은 <X X>안에 전사한다. 문장부호 다음에 붙인다.

24) 21세기 세종계획 말뭉치 구어 전사 지침에서는 연속적인 겹침과 비연속적인 겹침, 동시다발적 발화 등 다양한 겹침 상황에 따른 세부 전사 지침을 제시하고 있다. 한국어 학습자 말뭉치에서 이러한 지침을 포함하지 않은 것은 전사 도구를 통해 제시되는 발화 시간 정보가 겹침에 관한 정보들을 대체하기 때문이다.

<예> 그때도,
 <X보통X> 그런 자만심이 있었다.

- 화자의 발화 내용이 전혀 들리지 않는 부분은 <안들림>으로 전사한다. 이 때는 억양을 확인할 수 없으므로 문장부호를 넣지 않는다. 억양단위의 끝 부분에 말줄임표가 올 때는 앞에만 스페이스를 두고, 억양단위의 시작 부분에 올 말줄임표가 올 때는 뒤에만 스페이스를 두며, 중간에 올 때는 양쪽에 스페이스를 둔다.

<예> 4:<vocal desc='웃음'>
 1:거기까지 <note>안들림</note>
 2:근데 그거 진짜 신고해야 되겠더라.
 6:해야 돼.
 2:<note>안들림</note> 너무한 거 같더라.

- 들리지 않는 음절은 그 음절의 수만큼 X를 붙인다.

<예> 근데 그거 진짜 XX해야 되겠더라.

5) 전사 기호의 중복

- 전사 기호가 서로 중복될 경우에는 기호의 유형을 고려하여 우선순위에 따라 기입한다. 기호의 유형은 다음과 같이 네 가지로 나눌 수 있다.

< >, (), 문장부호(온점, 느낌표, 물음표/침표), 기타

- 전사 도구를 통한 전사 시 기호가 중복되어 사용되는 경우 시스템상의 오류가 발생할 수 있으므로 < > 기호 안에는 하나의 문장 단위만 입력한다. (억양을 나타내는 문장부호와 동시에 사용 불가하므로 분리하여 입력한다.)
- 억양 기호는 바로 앞의 음절의 억양을 표시하므로 다른 기호보다 우선하여 사용한다. 웃음이나 박수 등의 준음성 표현이 있을 경우는 그 다음에 붙인다.

6) 전사자의 설명

- 전사자가 특정 발화 구간에 대한 설명을 붙일 필요가 있을 경우 <note> </note> 태그를 사용하여 표현한다. 녹음 상태, 잡음, 동시다발적인 대화, 특이한 발음 상태 등의 설명이 덧붙을 수 있다. 주석 태그는 다음과 같이 붙여야 하고 앞뒤로 엔터를 쳐 준다.

<예> 1:응.
 <note>장소 이동으로 인해 잠시 멈춤</note>
 2:우리 때는 그런 거 없었잖아,

7) 혼잣말

- 혼잣말은 반영하여 전사하되 혼잣말임을 구분하기 위해 <monologue>, </monologue> 태그를 사용하여 표현한다.

<예> <monologue>미치겠네.</monologue>

2. 표기 지침

1) 대원칙

- 발화 내용은 기본적으로 철자법 수준의 전사를 한다. 다만, 구어의 발음 특성, 외국인 학습자의 발음 특성이나 오류, 지역적인 특성 등에 의해 철자법 대로 소리 나지 않는 발음(표준 발음이 아닌 경우)에 대해서는 발음 나는 대로 적는다. 이 경우 학습자의 발화는 괄호 안에 밝혀 주되 규범 표기는 괄호 밖에 전사하여 준다.

<예> 1: 친구(칭구)와 강남(간남)에 갔습니다(갔슨니다).

- ☞ 한국어 학습자 말뭉치는 일반 말뭉치와 달리 여러 가지 유형의 발음을 포함하기 때문에 철자법 전사의 수용 범위에 대한 고려가 필요하다. 가령, 위의 문장의 경우 한국어 모어 화자의 음성 자료를 들으면서 전사를 한다면 ‘친구’로 전사하겠지만, 외국인 학습자의 발음이므로 ‘친구’와 적을 것인지 ‘칭구’로 적을 것인지 결정해야 하는 문제가 발생한다. 이 경우 외국인 학습자

의 발음에 익숙한 한국어 교사라면 발음 오류를 비교적 쉽게 판단하여 표기에 반영할 수 있지만 일반인의 경우는 어려움이 따른다. 따라서 한국어 모어 화자에게서도 필수적으로 나타나는 음운 변화는 철자대로 전사하고 그 밖의 수의적인 발음은 위의 예와 같이 소리 나는 대로 1차 전사를 한다. 그리고 소리 나는 대로 전사한 표기 형태만으로 그 의미를 파악하기 어려운 경우대로 ()의 밖에 원래의 표기를 보충하여 넣도록 한다. (보충적 표기 관련 지침은 ‘4. 2차 전사/철자법 보충’ 참고)

2) 숫자 표기

- 숫자는 아래의 예에서처럼 발음에 따라 한글로 적는다.

<예> 오늘 제 동생이 이렇게(이케) 하나 오백 원이라고 사 가지고 왔더라고(왔더라구).
 : 500원으로 적지 않는다.

3) 외래어.외국어 표기

- 외래어나 외국어는 아래의 예에서처럼 발음에 따라 한글로 적는다.

<예> 2:어떻게 이거 크림 장난 아니야.
 1:이거도 오리지널 제주도 감귤이 아니야.

4) 끊어진 단어(불완전하게 발화된 단어)

- 발화된 대로 그대로 전사하고, ‘=’를 붙여 정상적인 단어와 구별할 수 있게 한다. 발화의 수정 등으로 인하여 한 어절이 완전하게 발화되지 못하고 불완전하게 발화된 경우 불완전하게 발화된 어절에 ‘=’를 붙인다.

<예> 4:사실 학습 자료랑 학= 형태는 떨어져도 되는데,

- 발화자가 불완전하게 발화한 것은 아니지만 한 어절을 발화하는 도중에 억양 단위가 바뀌어서 조사나 어미 등 문법 형태소가 실질 형태소와 다른 억양 단위로 전사될 때, 조사나 어미에 ‘=’를 붙인다.

<예> 1:주부 우울증,
=이라고 말할 수 있겠습니다.

- 그러나 완전한 어절이 완전 반복되는 발화의 경우는 ‘=’ 표시를 하지 않고 전사만 한다.
- 발화가 끝나지 않았는데, 말끝을 흐릴 경우 메모를 남긴다.

<예> 1:제주도에 가고 싶지만 돈이.<note>말끝흐림</note>

5) 띄어쓰기

- 띄어쓰기의 경우 맞춤법에 맞게 한다.
- 의존명사는 띄어 쓴다. 단, 특정 시점이나 순서를 나타내는 수사와 함께 사용될 경우는 띄어쓰기를 하지 않는다.(예, 일학년, 일층 등)
- 수를 적을 때는 만 단위로 띄어 쓴다.(예, 십이억 삼천백만 팔백구 불 등)
- 판단하기 어려운 경우에는 회의를 거쳐 최종 판단하고 향후 유사 사례를 일관되게 처리한다. (예, 오십대, 일 대 이, 등)
- 본용언과 보조 용언도 띄어 쓴다.

6) 축약형의 표기

- 구어에서는 발음의 축약 현상이 많이 나타나는데, 두 음절이 한 음절 사잇소리가 된다면, 두 음절이 한 음절 겹소리되는 것 등이다. 구어 말뭉치에서는 발음되는 음절 수와 표기상의 음절 수를 맞추는 것이 원칙이므로 축약형의 경우 모두 표기에 반영한다. 그런데 모음의 축약형의 경우 대부분 현재 국어의 모음 체계상 표기할 글자가 존재하지만, 반홀소리된 /ㄱ/, /ㄴ/의 표기가 문제가 된다. /ㄱ/, /ㄴ/가 반홀소리가 되어 /ㅊ/, /ㅌ/와 축약되는 현상은 구어에서 자주 나타나는데, 한글의 현재 글자 체계상 이러한 현상을 반영할 방법이 없으므로 구어 전사에서는 ‘(apostrophe, 영문따옴표)’를 사용해서 두 음소가 연결됨을 표시하고 구축 도구 내에서 마크업 처리한다.

<예> 사귀'어, 바뀌'어, ...

7) 장음 처리

- 표기와 관련해서 문제가 되는 간투사는 감정을 나타내는 감탄사와 특별한 의미 없이 말버릇 및 머뭇거림의 표지(담화표지)로 사용되는 간투사 유형이다. 이러한 발화는 그 특성상 원 음절보다 길게 발음되는데, 이 경우에는 ‘::’를 같이 사용해서 표기한다. 참고로 침표 표시는 담화표지를 나타냄으로써 후술한다.
- 마지막 음소를 길게 발음하는 경우 역시 ‘::’로 표기하여 준다.

<예> 1:많은 경우에 논문,
저::~ 어::~ 연구는 네이션,
국가라는 거하고(거하구) 직결되는:: 과정이죠.

- 발화자의 여러 가지 감정을 나타내는 소리들은 실제 구어 전사에서 다양한 형태로 나타난다. ‘오, 허, 응, 어, 어우, 와, 예, 이, 어휴’ 등의 형태를 기본으로 억양이나 길이 등이 달라지면서 놀람, 기쁨, 유감의 감정을 나타내게 된다. 이는 사전에 없는 유형들일 경우가 많은데, 가능한 한 실제 발화에 가깝게 전사하는 것을 원칙으로 한다.

<예> 오, 허, 응, 어, 어우, 와, 와우, 예, 앵, 이, 어휴, 아이, 치, 씨, 헤, 에이...

8) 담화표지

- "이, 그, 저, 아, 어" 등 동일한 형태로 기존 품사의 의미 및 기능을 가지지 않으며 시간을 끌기 위한 주변적인 말일 때 이를 담화표지로 보고 물결표(~, 숫자1 key 옆에 있음)를 이용하여 표시한다(주로 머뭇거림의 표지로 사용되는 이::~, 그::~, 저::~, 어::~, 아::~ 등이 해당됨).
- 억양과 운율에 의해서만 구분이 가능할 경우는 반드시 전사 단계에서 표시해 준다.
- 이 때 담화표지는 대부분 장음을 동반하는 경우가 대부분인데, 반드시 장음 표시와 함께 기재하여 줄을 원칙으로 한다.

<예> 1:많은 경우에 논문,
 그::~ 어::~ 연구는 네이션,
 국가라는 거하고(거하구) 직결되는 과정이죠.

3. 기타

1) 준음성과 기타 소리들

- 음소가 아닌 요소 즉, 웃음, 기침, 하품, 재채기, 박수와 같은 언어 외적 소리, 전화벨 소리와 같이 사람의 음성 아닌 소리는 대화 흐름에 영향을 주는 경우에만 표기한다. 가령, 발화자가 말하는 도중에 전화벨 소리가 울려서 발화를 멈추고 전화를 받거나, 발화 도중에 웃음소리가 끼어들 경우 발화는 자연스럽게 끊어진다. 이런 경우 전화벨 소리, 웃음을 표기한다. 반면, 발화를 하면서 책장을 넘기거나 볼펜소리를 낼 경우는 말소리와 동시에 소리가 나게 되는데, 이 경우 대화 상대자가 그 소리에 대해 언급하는 등 대화의 내용이나 흐름에 영향을 미치지 않는다면 표기하지 않는다.

- 웃음 <vocal desc='웃음'>
- 기침 <vocal desc='기침'>
- 하품 <vocal desc='하품'>
- 재채기 <vocal desc='재채기'>
- 목청 가다듬는 소리(음, 으음) <vocal desc='목청가다듬는소리'>
- 들이마시는 숨(쓰) <vocal desc='들이마시는숨'>
- 내쉬는 숨(후우) <vocal desc='내쉬는숨'>
- 혀 차는 소리(쯔) <vocal desc='혀차는소리'>
- 헛기침(에헴) <vocal desc='헛기침'>
- 한숨 <vocal desc='한숨'>
- 노래 <vocal desc='노래'>

- 학습자 개인의 발화 특성으로 습관적으로 반복해서 들이마시는 숨소리나 혀 차는 소리, 헛기침 등은 반영하지 않는다.

- 웃으면서 말하는 부분, 박수 치면서 말하는 부분 등도 표시한다.

<예> 5:우우 <vocal desc='박수'>
 <#이리 와 이리 와.#>
 위는 박수만을 치는 경우이고 아래의 경우는 박수를 치며
 발화를 하는 경우를 표현한다.

- 웃으면서 말하는 부분	<@	@>
- 박수치면서 말하는 부분	<#	#>
- 노래를 부르는 부분	<M	M>

4. 2차 전사

- 2차 전사의 경우 1차 전사 지침을 참고하여 전사된 자료를 검토하고, 아래의 항목에 대해 추가로 작업한다.

1) 철자법 보충

- 1차 전사 작업에서 발음대로 적은 것 가운데, 구어의 발음 특성, 외국인 학습자의 발음 특성 등에 의해 철자대로 소리 나지 않는 발음(표준 발음이 아닌 경우), 음운 규칙이나 정확한 음절 발음을 몰라 일으킨 발음 오류는 () 안에 표기하고 () 밖에는 본래의 표기를 함께 적어 준다. 철자법에 맞는 것을 함께 적어주지 않으면 내용 이해에 어려움이 있을 수 있기 때문이다. 이는 작업자에 따라 1차 전사 과정에서 할 수도 있다.

<예> 친구(칭구)와 강남(간남)에 갔습니다(같슨니다).

<예> 같이(가티) 가자.

- 억양단위 맨 끝에 억양기호와 함께 나타나는 경우에는 기호도 함께 붙여준다.

<예> 2: 몇 살인데,
그 광은.
광희초등학교 간 사람은?
1: 서른 몇 살이나 될 거야.
2: 음 젊네 다?(다이?)

- 그러나 구어적 변이형이라 할지라도 자주 등장하거나 쉽게 원래 형태로 이해될 수 있는 것들은 일일이 철자형을 붙여주지 않는다.

<예> 책상 위에 놔 뒀.

- 소유격 조사 '의'의 경우 한글 맞춤법 통일안에서는 [의]와 [예]를 모두 표준 발음으로 인정하고 있다. 즉 <표준 발음법>제 5 항에서는 단어의 첫 음절 이외의 '의'는 [이]로, 조사 '의'는 [예]로 발음함도 허용하고 있다. 그러나 실제 발화에서 소유격 조사 '의'를 [의]라고 발화하는 모어 화자는 매우 드물기 때문에 '의'를 [예]로 발음한 경우는 '의' 그대로 전사하고, '의'를 [의]로 발화한 경우에는 '의(의)'로 전사한다.

<예> [민족예]로 발화하였을 경우

9: 나는,
자랑스런 태극기 앞에,
조국과 민족의,
무궁한 영광을 위하여,

<예> [민족의]로 발화하였을 경우

9: 나는,
자랑스런 태극기 앞에,
조국과 민족의,(민족의.)
무궁한 영광을 위하여,

- 외국인 학습자의 발화는 한국어의 음운 체계에 없는 음운의 발음이나 표기가 어려운 중간 발음, 외국인 학습자에게만 나타나는 독특한 발음이 자주 등장한다. 이 경우 () 밖에 규범 표기를 넣어 철자법을 보충하는 것을 기본으로 하나, <예>의 유형 3과 같이 철자 전사를 통해 이를 반영하기 어려우나 원래의 표기를 먼저 적고, 학습자의 실제 발음을 ()에 남겨 표기는 동일하나 오류가 있음을 알 수 있도록 한다. 모든 경우 음운적 구분이 모호하거나 특징적인 사항이 있을 때에 메모를 남기도록 한다.

<예> 유형 1. 한국어 모어 화자의 발음과 음운적으로 구분이 될 정도로 발음에 오류가 있는 경우

선생님(성생닌)
여자(요자)
회사(휘사)
과반수(화반수)

유형 2. 한국어 모어 화자의 발음과 음성 혹은 변이음의 구분이 모호한 경우

가구(가구)<note>‘가’의 ㄱ을 유성음으로 발음</note>

가구(가구)<note>‘구’의 ㄱ을 무성음으로 발음</note>

유형 3. 단, 음운규칙으로 인해 한국어의 철자대로 발음되지 않으나, 학습자가 이를 철자대로 발음하는 경우

무조건(무조건)<note>경음화되지 않음</note>

같이(가티)<note>구개음화되지 않음</note>
 신라(신라)<note>자음동화되지 않음</note>
 앞에(앞에)<note>연음되지 않음</note>
 먹는(먹는)<note>자음동화되지 않음</note>

- 외국인 학습자의 발화에서 외국어나 외래어 발음 시 원어에 가까운 소리로 발음을 하는 경우가 자주 발생한다. 이는 한국어 모어 화자에게서도 일어나는 현상이기는 하나 외국인 학습자에게서 그 빈도가 더 잦고, 발음 또한 모어 화자의 그것과 많이 다르다. 따라서 이 경우에도 표기 원칙에 맞춰 한글로 적되, 원래의 형태를 파악하기 어렵다고 판단되는 경우에는 학습자의 원어 발음을 최대한 원어에 가깝게 () 안에 적어 밝힌다. 이때 한국어의 음운 체계로 전사가 불가능한 발음은 표기에 반영하지 않는다.

<예> 인터뷰: 영어식 발음으로 [인털뷰]에 가까운 소리가 남
 센터: 영어식 발음으로 [세너]에 가까운 소리가 남
 파트너: 영어식 발음으로 [팔너]에 가까운 소리가 남
 → 각각 ‘인터뷰(인털뷰)’, ‘센터(세너)’, ‘파트너(팔너)’로 전사함

- 외국인 학습자 발화의 경우 외래어 또는 외국어 발화 시 원어식의 발음을 하거나 한국어 모어화자의 현실 발음이 아닌 규범 발음을 하여 어색하게 들리는 경우가 있다. 이 경우 철자 전사를 통해 반영하기 어려우나 원래의 표기를 먼저 적고, 학습자의 실제 발음을 ()에 남겨 발음에 차이가 있음을 알 수 있도록 한다.

<예> 카페: 현실 발음 [까페], 학습자 발음 [카페]
 버스: 현실 발음 [빠스], 학습자 발음 [버스]
 → 각각 ‘카페(까페)’, ‘버스(버스)’로 전사함.

2) 방언형 표시

- 확실한 방언형(대응하는 표준형 형태소가 없는 것)의 경우는 다음의 예와 같

은 태그를 붙인다.

<예> 2:저기여.
선거 저기 성화 차가 오는 게,
오 분마다 있어.
차는 <dia>여일</dia> 있어.

3) 씬

- (1초 이상의 씬은) 0.1초 단위까지 표시한다(전사 도구의 시간 정보를 이용한다). 씬은 발화와 발화 사이의 씬이기 때문에 다음 발화의 시작 전에 표시한다. 만약 씬이 누구의 것인지 불분명할 때는 한 줄에 표시한다.

<예> 1:이거 올라가면서 먹을까?
2:{1.2} 그럴까?
{4.3}
... 하게 먹는다.

- 한 어절 안에서의 짧은 씬은 ‘..’로 표시한다. 하나의 억양 단위 내부에서의 짧은 씬은 따로 표시하지 않는다.

<예> 2:아:: 그리고 어:: 남의 의견을 잘 듣고 수용하고 대화..로
타 협해야 된다고 하면서,

4) 인용

- 인용된 부분은 <Q Q>를 사용하여 표시한다. 여러 억양단위에 걸쳐 인용된 경우는 처음과 끝에만 표시를 한다.

<예> 1:근데 요즘 사회학자들은 또는 철학자들은 그렇게 얘기하지 않아요.
<Q현대사회는 다양성의 시대다.Q>
라고 말하죠.

5) 텍스트 읽기 인용

- 책이나 자료 등을 보고 읽은 경우는 <R R>를 사용하여 표시한다.

<예> 1:그 답에 <R생각과 느낌이 유기적으로 잘 짜여져 조직체를 이룰 때 좋은 글이 될 수 있다R>라고 돼 있어요.

6) 익명성 보장을 위한 마크업

- 대화자들의 신분 보장을 위해 이름, 주민등록번호, 카드 번호, 전화 번호 등은 노출되지 않도록 태그로 대신한다. 다음은 마크업 과정에서 각각의 정보를 대신하는 태그들이다.

<name> : 사람 이름, 단체 이름, 학교 이름 등

<id-num> : 주민등록번호, 외국인등록번호, 학번 등 개인 식별 번호

<card-num> : 신용카드 번호

<address> : 주소

<tel-num> : 전화 번호

- 여러 사람의 이름이 나올 때는 <name1>, <name2> 등으로 일련번호를 붙여 준다.

<예> 5:네.
거 어떻게 보면 가장 실망::스런 일 중 하난데요,
헤럴드 쪽에서도 그다지 뽕족한,
대안을 가지고 있는 것 같지는 않더라구요,
그게 어찌면 <name1> 선배님이라든지 다른 선배님들 말::
들은 걸 생각해 보면,
<name2> 사장이 <name2> 회장이 있으니까 보안이 있어도
눈치 보여서 얘기를 못한다,

한국어 학습자 말뭉치 형태 주석 지침

※ 본 지침은 21세기 세종 계획 현대문어 형태분석 말뭉치 구축 지침을 기본으로 한다.

I. 학습자 말뭉치의 형태 분석 표지²⁵⁾

대분류	형태 주석 내용	기호	세종 표지
(1) 체언	일반명사	NNG	NNG
	고유명사	NNP	NNP
	의존명사	NNB	NNB
	대명사	NP	NP
	수사	NR	NR
(2) 용언	동사	VV	VV
	형용사	VA	VA
	보조용언	VX	VX
	지정사	VCP/VCN	VCP/VCN
(3) 수식언	관형사	MM	MM
	일반부사	MAG	MAG
	접속부사	MAJ	MAJ
(4) 독립언	감탄사	IC	IC
(5) 관계언	주격조사	JKS	JKS
	보격조사	JKC	JKC
	관형격조사	JKG	JKG
	목적격조사	JKO	JKO

25) [수정] 기존 세종 지침에 있었던 NF(명사추정범주), NV(용언추정범주)를 삭제하고 대부분 추정하여 해당 표지로 분석하거나 NA(분석불능범주)로 분석함

	부사격조사	JKB	JKB
	호격조사	JKV	JKV
	인용격조사	JKQ	JKQ
	보조사	JX	JX
	접속조사	JC	JC
(6) 의존형태	선어말어미	EP	EP
	어말어미(연결)	EC	EC
	어말어미(종결)	EF	EF
	명사형전성어미	ETN	ETN
	관형형전성어미	ETM	ETM
	체언접두사	XPN	XPN
	명사파생접미사	XSN	XSN
	동사파생접미사	XSV	XSV
	형용사파생접미사	XSA	XSA
	어근	XR	XR
(7) 기호	마침표, 물음표, 느낌표	SF	SF
	쉼표, 가운뎃점, 콜론, 빗금, 줄표, 물결	SP	SP
	따옴표, 괄호표	SS	SS
	줄임표	SE	SE
	붙임표(숨김, 빠짐)	SO	SO
	외국어	SL	SL
	한자	SH	SH
	기타 기호	SW	SW
	숫자	SN	SN
	분석불능범주	NA	NA

II. 기본 원칙

가. 분석 대상 : 형태분석은 하나의 어절을 분석 대상으로 한다.

나. 분석 원리 : 본 분석은 ‘형태소’ 차원이 아닌 ‘형태’ 차원의 분석이므로 이 형태를 최대한 반영한다.

다. 분석 원칙

- 형태 분석은 분석 대상인 원시 말뭉치를 가급적 훼손하지 않는다.
- [보완] 띄어쓰기는 어문 규범의 한글맞춤법을 기본으로 하며 ‘허용’ 규정도 인정한다.

라. [학습자] 분석 기준

- 대원칙 : “21세기 세종 계획 형태 분석 말뭉치 구축 지침”

1) “지침”을 보고 분석이 가능하면 지침으로 해결한다.	
2) “지침”을 통해 해결이 불가능한 어휘에 대한 분석은, 가) <u>표준국어대사전</u> 을 따른다.	
■■ <u>중요하다</u>	[중요/NNG+하/XSA+다/EF]
- 중요	표준국어대사전: 중요/NNG ✓
※ 단, <u>조사나 어미의 결합형은 지침에 목록으로 제시된 것을 제외하고는</u> 표준국어대사전에 등재되어 있어도 각각 분리해서 분석한다.	
■■ <u>먹었으니까는</u>	[먹/VV+었/EP+으니까/EC+는/JX]
■■ <u>가고는</u>	[가/VV+고/EC+는/JX]
■■ <u>좋으니만큼</u>	[좋/VA+으니/EC+만큼/JX]
나) 사전으로 해결이 불가능한 경우 세종 말뭉치를 확인하고 분석한다.	

○ 분석 기준 :

1) 뜻(어휘적 의미+기능적 의미)은 알지만 정확한 형태는 모르는 경우 → 원래 품사로 분석한다. 교정 어절 이 취할 표지를 준다.	
■■ 문제를 <u>쉬게</u> 풀어요.	[쉬/VA+게/EC]
■■ 너 <u>때문내</u> 죽겠어.	[때문/NNB+내/JKB]
■■ <u>여러까지</u> 문제가 생겼다.	[여러/MM Ⅱ 까지/NNB]
■■ 강에 <u>패수른</u> 버렸다.	[패수/NNG+른/JKO]
2) 형태와 뜻(어휘적 의미+기능적 의미)을 혼동한 경우 → 보이는 대로 분석한다. 오류 어절 만 고려해서 분석한다.	
■■ 내가 <u>고기가</u> 먹어요.	[고기/NNG+가/JKS]
■■ 입학하자마자 교과서를 <u>팔려고</u> 서점에 갔어요.	[팔/VV+려고/EC]

마. [학습자] 분석 기준의 적용

- 학습자 언어에서 나타나는 오류의 유형에 따라 다음과 같이 분석한다.

- 1) 오형태가 나타난 경우 최소 교정을 원칙으로 교정 어절을 상정해 형태 분석을 한다. 교정 어절을 상정하기 어려운 경우는 **분석불가능(NA)**으로 처리한다.

■■ 그러므로 <u>부스르른</u> 광고는 물가를 인상한다.	[부스르른/NA]
■■ 그리고 경전철 타로 <u>필어</u> 50분쯤 경사턱역 있습니다.	[필어/NA]
■■ 이번 방학에 저는 친구와 같이 <u>순열전 수열고</u> 싶어요.	[순열전/NA] [순열/NA+고/EC]
■■ 그 꿈을 <u>아구할</u> 수 없을 것 같다.	[아구하/NA+ㄹ/ETM]

→ 분석 불가능은 문맥에서 전혀 의미를 유추할 수 없어 오형태로도 보기 어려운 경우이다.

→ 표현 문형의 구성과 인접한 경우 교정어절을 상정하기 어렵다 해도 표현 문형 구성에 포함되는 형태까지는 분석을 한다.²⁶⁾

- 2) 교정 어절의 상정이 가능한 경우 교정 어절이 취할 형태 표지를 부여한다.

가) 형태 분할이 가능한 것은 교정 어절을 기준으로 최대한 분할하는 것을 원칙으로 한다.

26) 표현 문형의 경우 <국립국어원2>의 표현 문형 목록을 기준으로 한다.(부록 참고)

■■ 아무도 <u>몰랐다</u> .(√몰랐다)	[무르/VV+아/EP+다/EF]
■■ 저는 중국 <u>사람인데</u> .(√사람인데)	[사람/NNG+으/VCP+ㄴ데/EC]

나) 상정한 교정어절에 없는 형태가 추가된 경우의 분석은 다음과 같다.

(1) 체언과 조사 곡용의 경우는 체언 또는 조사의 오형태로 분석한다.

■■ <u>학곡</u> 을 갔다.	[학곡/NNG+을/JKO]
■■ <u>학곤</u> 을 갔다.	[학곤/NNG+을/JKO]
■■ <u>학교으</u> 를 갔다.	[학교/NNG+으를/JKO]

(2) 용언과 어미 활용의 경우는 용언의 어간 혹은 어근과 어미를 확보해 분리한 후 잉여적 요소에 대해서는 NA로 처리한다.

■■ 내 꿈을 <u>이뤄진</u> 고나 이루기 위해	[이뤄지/VV+ㄴ/NA+고나/EC]
■■ <u>하숙집인</u> 기 때문에 사람이 많았다.	[하숙집+이/VCP+ㄴ/NA+기/ETN]
■■ 꼭 잘해야 <u>되</u> 겠다.	[되/VV+ㄴ/NA+겠/EP+다/EF]
■■ 힘이 <u>강한</u> 하	[강하/VA+ㄴ/NA+하/XSA+지/EC]
■■ 교실에 사람이 <u>많</u> 다.	[많/VA+다/EF+ㄴ/NA]
■■ 교통이 꼭 <u>편리</u> 합니다.	[편리/NNG+ㅁ/NA+습니다/EF]
■■ 나는 선생님이 <u>되</u> 겠다.	[되/VV+ㅁ/NA+겠/EP+다/EF]
■■ 공부를 <u>해</u> 고	[하/VV+아/NA+고/EC]
■■ 7시 30분까지 <u>운동</u> 합니다.	[운동/NNG+하/XSV+아/NA+ㅁ니다/EF]
■■ 교실에 사람이 <u>많</u> 아다.	[많/VA+아/NA+다/EF]
■■ 꽃 가게 주인 <u>돼</u> 면	[되/VV+어/NA+면/EC]
■■ 식사가 <u>준비</u> 돼어 있었다.	[준비/NNG+되/XSV+어/NA+어/EC]
■■ 할 수 있는 일이 <u>많</u> 아져셨는데	[많아지/VV+어/NA+시/EP+였/EP+는데/EC]

→ 이때, 교정 어절을 기준으로 했을 때 잉여적인 요소가 추가된 오류와 기존의 다른 문법 요소와 혼동한 오류의 경우를 구분해서 주석해야 한다.

■■ 교실에 사람이 <u>많</u> 은다.	[많/VA+은/NA+다/EF]
■■ 교실에 사람이 <u>많</u> 는다.	[많/VA+는다/EF]

다) 불규칙 용언의 활용과 관련한 오류의 형태 분석은 다음과 같다.

(1) 불규칙 용언의 형태 분석에서, 학습자가 불규칙 활용의 오류로 어간을 잘못 쓴 경우에는 어간을 복원하지 않고 오류가 발생한 어간의 형태를 그대로 살려 분석한다.

■■ 한국말이 너무 <u>어려우다</u> .	[어려우/VA+다/EF]
■■ 친구들과 같이 <u>즐거우게</u> 칠 수 있으면	[즐거우/VA+게/EC]
■■ 다른 사람이 다시 저에게 <u>도울</u> 수 있다.	[도오/VV+ㄹ/ETM]
■■ 정말 <u>추운</u> 경험이었다.	[추으/VA+ㄴ/ETM]

→ 하지만 활용 오류가 용언의 불규칙 활용과 직접적인 관련이 없거나 어미의 오류로 보이는 경우는 기존의 형태 분석대로 어간을 살려 형태를 부여한다.

■■ 한국말이 너무 <u>아렵다</u> .	[아렵/VA+다/EF]
■■ 저녁식사도 준비하기가 <u>번거러워서</u>	[번거/XR+럽/XSA+어서/EC]
■■ 공부가 <u>힘드는</u>	[힘들/VA+는/ETM]
■■ 풍선이 <u>부푼</u>	[부푼/VV+ㄴ/NA+는/ETM]
■■ 공부가 <u>힘든지만</u>	[힘들/VA+ㄴ/NA+지만/EC]
■■ 공부가 <u>힘든입니다</u> .	[힘들/VA+ㄴ/NA+이/VCP+ㅂ니다/EF]

(2) 또한 다음과 같이 어미가 누락한 것으로 보이는 경우는 오류가 발생한 어간 형태를 그대로 살려 분석한다.

■■ 친구들과 <u>가벼운</u> 장난을 하는 것은	[가벼우/VA]
■■ 경제력이 부족하거나 <u>힘들</u> 생활을 겪고	[힘들/VA]

라) 형태소 경계를 분할하기 어려운 오류의 형태 분석은 다음과 같다.

(1) 상정한 교정 어절의 형태소 음절에 따라 앞에서부터 형태를 분할한 후 형태 표지를 부여한다.

■■ <u>자시느이</u> 마음대로 했다.	[자시/NNG+느이/JKG]
■■ 내 유학생활을 <u>아프로</u> 미래에게	[아/NNG+프로/JKB]
■■ 다언에도 소개해 <u>주게서요</u> .	[주/VX+게/EP+서요/EF]
■■ <u>어려슬</u> 데 가을에 좋은 기억이	[어리/VA+어/EP+슬/ETM]
■■ 그것을 <u>마가려고</u> 하는 것들 중에	[마/VV+가려고/EC]

(2) 오류의 형태가 형태 단계에서도 표시될 수 있도록 가능한 경우 자모 단위로

도 형태를 분할한다.

- 정말 신기하다고 생각했다. [생각/NNG+하/XSV+ㄴ/EP+다/EF]
- 어제 영화를 봤는데 [보/VV+ㄴ/EP+는데/EC]

→ 일부 오류의 경우는 분석한 형태의 결합이 원 어절의 형태가 되지 않더라도 기본적으로는 분할을 원칙으로 다음과 같이 처리한다.

- 쓰레기를 버려도 되면 좋겠습니다. [버려/VV+어도/EC]
- 많은 친구를 사귀서 재밌었어요. [사귀/VV+어서/EC]
- 소치에 2시간 걸려요. [걸려/VV+어요/EF]
- 점점 심해질 것이다. [심해/VA+어/EC+지/VX+ㄹ/ETM]

마) 다음의 경우는 형태 분할을 하지 않고 오형태로 분석한다.

(1) 다음의 축약형에서 나타나는 오류 유형은 분할하지 않고 오형태로 분석해 형태 표지를 할당한다.

- 내일은 네 생일이라서 소포를 받았다. [네/NP]
- 재 장소 중에서 제일 좋아하는 곳은 [재/NP]
- 세 아버지는 키가 큼니다. [세/NP]
- 또 궁금한 개 있으면 [개/NNB]
- 저는 OO어학당 6급 학생이에요. [학생/NNG+이/VCP+예요/EF]
- 나는 학생이었다. [학생/NNG+이/VCP+였/EP+다/EF]

(2) 복합어의 구성 요소인 어근이나 접사가 누락된 경우 단어의 오형태로 보고 복합어 전체의 품사로 분석한다.

- 눈이 많(√많이) 왔다. [많/MAG]
- 줄임말은 젊은들이(√젊은이) 많이 사용하는 [젊은/NNG+들/XSN+이/JKS]

→ 하지만 용언의 어간이 어미와 결합한 활용형에서 어미가 누락된 경우는 용언의 어간으로 분석한다.

- 많(√많은) 음식이 있었어요. [많/VA]

Ⅲ. 표지별 분류 기준 및 세부 지침

가. 체언

- 체언은 명사, 대명사, 수사를 포괄하는 대범주로서, 조사와 결합하거나 그 자체로 다른 체언이나 용언과 어울려 하나의 문장성분이 될 수 있다.

1) 명사(NN)

- 명사는 사물의 이름을 나타내는 품사이다. 본 표지에서는 명사를 일반명사, 고유명사, 의존명사로 세분한다.

가) 일반명사(NNG)

- 사물의 이름을 나타내는 단어로서 표준국어대사전에 명사로 등재된 표제어(고유명사와 의존명사를 제외한 모든 명사)와 독립된 음절(한자어), 약어, 고사성어 등 사전 표제어는 아니나 다른 품사로 분석될 수 없는 단위들을 포함한다.

(1) 일반명사로 분석할 수 있는 단어

(가) 표준국어대사전의 명사 표제어

■■ 국어/NNG, 연구/NNG

(나) 1음절 한자어가 독립된 단위로 사용되는 경우

■■ 서울초등학교 졸 [졸/NNG]

※ [보완]

■■ 나는 환경에 '환'자도 모르는 [‘/SS+환/NA+’/SS+자/NNG+도/JX]

(다) 한자성어

■■ 백척간두(百尺竿頭) [백척간두/NNG+(/SS+百尺竿頭/SH+)/SS]

(라) 외국어를 음차한 경우

■■ 아이 러브 유(I love you) [아이/NNG]

(마) [보완] ‘명사 + (분석 목록에 없는) 접사’는 전체를 통합하여 명사로 분석한다.

■■ 2년간 [2/SN+년간/NNB]

■■ 4호선 [4/SN+호선/NNB]

■■ 상상력 [상상력/NNG]

■■ 중국식 [중국식/NNG]

(2) 명사 상당어의 분석

(가) 동사의 활용형이 따옴표 없이 문장 속에서 명사처럼 기능하는 경우는 원래 품사대로 분석한다.

■■ 어디 가느냐가 그의 물음이었다. [가/VV+느냐/EF+가/JKS]²⁷⁾

(나) 따옴표를 가진 성분이나 요소도 명사처럼 기능할 수 있으나, 원래 품사대로 분석한다.

■■ 그것은 “느”이 아니라 “를”이다. ["/SS+는/JX+"/SS+이/JKC]

(다) 부사 뒤에 격조사가 쓰이는 것도 의미론적인 따옴의 효과에 의하여 부사가 명사적인 용법을 가지는 것이므로 분석은 ‘부사’로 한다.

■■ 가족을 멀리에 보냈다. [멀리/MAG+에/JKB]

(라) [보완] 학습자의 특성상 접사를 명사적 기능으로 사용한 경우 분석하는 접사 목록에 없더라도 원래 품사대로 접사로 분석한다.

■■ 제주도에는 한국의 여명이 도예요. [도/XSN+이/VCP+예요/EF+./SF]

(3) [보완] 학생, 학교

- 대학교, 고등학교, 중학교, 대학생, 고등학생, 중학생은 모두 일반명사로 분석한다.

27) [수정] ‘21세기세종계획’ 지침에는 ‘느냐/EC’로 되어 있지만 오류이므로 수정함.

■■ 대학교	[대학교/NNG]
■■ 고등학교	[고등학교/NNG]
■■ 중학교	[중학교/NNG]
■■ 대학생	[대학생/NNG]
■■ 고등학생	[고등학생/NNG]
■■ 중학생	[중학생/NNG]

나) 고유명사(NNP)

- 고유 명사는 특정한 사물에 붙여진 이름으로, 기본적으로 최하의어에 속하는 대상을 서로 변별하기 위하여 붙인 이름이며, 원칙적으로 지시 대상만 가질 뿐 의미 내용은 가지지 않는다. 고유명사의 분석 기준은 매우 다양하므로, 본 지침에서는 다음에 제시하는 것만을 고유명사로 인정한다. 또한, 본 지침은 띄어쓰기 단위의 분석을 원칙으로 하고 있으므로, 한 단어 이상으로 구성된 고유명사(‘바람과 함께 사라지다’)와 같은 경우의 분석을 위해 전체를 아우르는 단위를 설정하지는 않는다.

(1) 인명, 종족명

- (가) ‘씨(氏), 공(公), 군(君), 양(孃), 웅(翁)’ 등 성 또는 이름 뒤에 같이 쓰이는 호칭어나 직책명은 분리해서 분석한다.

■■ 남수/NNP||군/NNB, 김/NNP||씨/NNB, 최치원/NNP||옹/NNB,
케네디/NNP||씨/NNB²⁸⁾, 정/NNP||과장/NNG, 최/NNP||선생/NNG

- (나) 성과 이름, 호가 함께 쓰이면 하나의 단위로 분석한다.

■■ 김철수/NNP, 이태백/NNP

- (다) ‘씨, 군’ 등과 달리 ‘가(哥)’는 접미사이므로, ‘김가(金哥), 이가(李哥)’는 파생어이다.

■■ 김/NNP+가/XSN

- (라) 사람 이름의 뒤에 접사 ‘-이’가 붙는 경우는 이름과 함께 하나의 단위로

28) [수정] 지침 전체적으로 띄어 써야 할 부분이 +기호로 연결되어 있어 ||기호로 수정함.

분석한다.

■ ■ 진현이/NNP + 가/JKS

(마) 특정한 종족의 이름은 고유명사가 된다.

■ ■ 알타이족/NNP, 피그미족/NNP, 돌궐족/NNP, 한족/NNP

(2) 지명

(가) 내륙, 바다, 강, 산, 산맥, 호수, 섬, 만, 계곡, 늪, 주 등의 이름

■ ■ 카스피해/NNP, 템즈강/NNP, 태백산맥/NNP, 미시시피호/NNP, 네바다주/NNP

■ ■ 한강/NNP, 한라산/NNP, 남이섬/NNP, 남극/NNP, 북극/NNP

(나) 주소를 나타내는 도(道), 시(市), 읍(邑), 면(面), 리(里), 군(郡), 구(區), 동(洞), 골, 촌, 로 등의 이름은 그 구역의 종류를 나타내는 말과 함께 전체가 고유명사가 된다.

■ ■ 서울특별시/NNP, 성북구/NNP, 강진군/NNP, 인천동/NNP, 빨래골/NNP, 해방촌/NNP

■ ■ 연세로/NNP, 세검정로/NNP, 상동로/NNP, 테헤란로/NNP

■ ■ 신촌/NNP, 여의도/NNP, 광화문/NNP, 명동/NNP

(3) 국가명 또는 왕조명

(가) 국가의 명칭, 또는 왕조의 명칭은 고유명사로 분석한다.

■ ■ 대한민국/NNP, 조선/NNP

(나) 다른 형태가 붙어 국가나 왕조의 존립 기간을 나타내는 경우 일반명사로 분석한다.

■ ■ 대한제국기/NGG, 조선조/NGG

(다) '남, 북, 남북'은 방향을 가리키는 일반명사와 '남한'과 '북한'을 의미하는 고유명사를 구별한다. 남한을 뜻하는 '남'과 북한을 뜻하는 '북'을 고유명사로 분석한다.

■ ■ 남/NNP+과/JC||북/NNP+의/JKG||의견/NGG||차이/NGG

■ ■ 남북/NNP||적십자회담/NGG

■ ■ 북/NNP+미/NNP|| 회담/NNG

(라) 어떤 국가의 국민을 나타내는 ‘국가+인’은 통합하여 일반명사로 분석한다.

■ ■ 이집트인/NNG, 아제르바이잔인/NNG, 이스라엘인/NNG, 조선인/NNG

(마) 어떤 국가의 군대를 나타내는 ‘국가+군’은 통합하여 일반명사로 분석한다.

■ ■ 미군/NNG, 북한군/NNG, 영국군/NNG, 일본군/NNG

(바) 국가명의 약어는 고유명사로 분석한다.

■ ■ 한/NNP+중/NNP+일/NNP

(4) 건축물이나 시설물 혹은 구조물의 이름

(가) [보완] 도로, 항만, 철도, 전철, 지하철 및 그 명칭과 함께 쓰이는 부대시설은 그 종류를 나타내는 말과 함께 전체가 고유명사가 된다.

■ ■ 부산항/NNP, 대전역/NNP, 서울지하철/NNP, 인천공항/NNP

■ ■ 홍대입구역/NNP, 홍대입구/NNP(준말)

(나) 빌딩, 박물관, 극장 등 건물명은 그 종류를 나타내는 말과 함께 전체가 고유명사가 된다.

■ ■ 서울역사/NNP, 세종문화회관/NNP, 개나리유치원/NNP, 연세대학교/NNP

■ ■ 국립중앙박물관/NNP, 국립민속박물관/NNP, 루브르박물관/NNP

■ ■ 신라호텔/NNP, 현대백화점/NNP, 동궁예식장/NNP, 명보극장/NNP, 세브란스병원/NNP

(다) 알파벳이나 숫자, 기호를 포함한 경우 전체가 고유명사가 된다.

■ ■ N서울타워/NNP, N-서울타워/NNP, 63빌딩/NNP

※ [보완]

■ ■ 남대문/NNP|| 시장/NNG, 한강/NNP|| 공원/NNG

(5) 회사, 학교, 정당, 기관이나 단체의 이름

(가) 특정 회사나 학교, 정당 등의 이름은 고유명사로 분석한다. 단, 특정 회사의 상품명은 고유명사가 아닌 일반명사로 취급한다.

- ■ 삼성/NNP, 연세대학교/NNP, 새누리당/NNP, 자유민주주의연합/NNP
- ■ 초코하임/NNG, 한메타자교실/NNG

(나) 정부기관의 명칭은 모두 일반명사로 처리한다. 그러나 거기에 인명, 지명 등의 고유명사가 포함된 경우 그 통합형을 고유명사로 처리한다.

- ■ 헌법/NNG|| 재판소/NNG, 대/XPN+법원/NNG, 고등/NNG|| 법원/NNG, 재정/NNG|| 경제원/NNG
- ■ 서울고등법원/NNP, 서울시경찰서/NNP, 서대문구치소/NNP

(다) 특정 기관이나 단체, 연구소 등의 경우에는 분석하는 것을 원칙으로 한다. 그러나 거기에 인명, 지명 등의 고유명이나 ‘전국’, ‘국제’, ‘세계’ 등이 포함되면 그 통합형을 고유명사로 처리한다.

- ■ 대한축구협회/NNP, 전국은행협회/NNP, 한국전자통신연구소/NNP
- ■ 생활/NNG|| 체육/NNG|| 연구소/NNG, 입주자/NNG|| 대표자/NNG|| 협의회/NNG

(라) 약어나 준말의 처리

- 고유명사가 축약된 형태(준말)로 쓰일 경우 본디말과 함께 준말도 인정하여 축약된 형태 그대로를 고유명사로 분석한다. 그리고 일반명사로 분석하는 기관명의 약자는 일반명사로 분석한다.

- ■ 육사/NNP, 연대/NNP, 자민련/NNP, 서울고법/NNP
- ■ 정보통신위/NNG (정보/NNG|| 통신/NNG|| 위원회/NNG)

(6) [보완] 아이돌 등의 그룹명은 (6) 창작물의 제목과 같게 처리한다.

- ■ 소녀시대/NNP, 걸스데이/NNP, 방탄소년단/NNP
- ■ 제국/NNG+의/JKG|| 아이/NNG+들/XSN, 서태지/NNP+와/JC|| 아이/NNG+들/XSN

※ [보완]

- ■ EXID/SL, YG/SL+Family/SL
- ■ B1A4/NNP, 2NE1/NNP

(7) [보완] 책, 연극, 영화, 드라마, TV 프로그램 등의 창작물의 제목

■■■ 삼국사기/NNP, 손자병법/NNP, 고래사냥/NNP

■■■ 슈키라(슈퍼주니어의 키스 더 라디오) 슈키라/NNG(준말)

어절 미분 리 (NN 구성 포함)	사전 등재	전 체 NNP	(책) 삼국사기/NNP, 손자병법/NNP
	사전 미등재	전 체 NNP	(드라마) 전원일기/NNP, 가을동화/NNP (영화) 어벤저스/NNP, 쿵푸팬더3/NNP (TV프로그램) 런닝맨/NNP, 가족오락관/NNP
어절 분리	사전 등재	나 누 어 분석	(책) 안네/NNP+의/JKG 일기/NNG
	사전 미등재	나 누 어 분석	(드라마) 서울/NNP+의/JKG 달/NNG (영화) 비밀/NNG+은/JX 없/VA+다/EF (TV프로그램) 남자/NNG+의/JKG 자격/NNG

(8) 언어명

- 언어명의 경우 ‘-어’의 형태만을 통합하여 고유명사로 인정한다.

■■■ 한국어/NNP, 일본어/NNP, 영어/NNP, 알타이어/NNP, 네덜란드어/NNP

■■■ 한국말/NNG, 러시아/NNP||말/NNG, 일본/NNP||말/NNG

■■■ 한글/NNG, 알파벳/NNG, 한자/NNG

(9) 웹사이트, SNS, APP

- 웹사이트, SNS, APP의 이름은 모두 고유명사로 처리한다.

■■■ 네이버/NNP, 다음/NNP, 구글/NNP

■■■ 인스타그램/NNP, 페이스북/NNP, 카카오톡/NNP, 트위터/NNP

■■■ 직방/NNP, 카카오퍼스/NNP

(10) [보완] 캐릭터의 이름

■■■ 미키마우스/NNP, 호돌이/NNP, 알라딘/NNP, 키티/NNP, 라이언/NNP

다) 의존명사(NNB)

- 의존명사는 자립해서 쓰일 수 없는 명사로, 수식 성분을 반드시 동반해야

한다. 의존명사는 비단위성 의존명사와 단위성 의존명사로 나눌 수 있으나, 본 분석에서는 이를 세분해하지 않는다. 또한 의존명사가 일반명사와 같이 독립적으로 쓰일 때는 일반명사로 분석한다. 의존명사와 일반명사의 구분은 표준국어대사전의 등재 여부에 따른다.

(1) 의존명사이지만, 일반명사처럼 쓰이는 경우

(가) “연대, 연도, 연차”는 “년대, 년도, 년차”와 달리 모두 일반명사로 분석한다.

- | | |
|------------------------|----------|
| ■■ 연도별로 정리된 자료 | [연도/NNG] |
| ■■ 몇 <u>년도</u> 에 일어난 일 | [년도/NNB] |

(나) “월, 연, 일, 주, 달러, 원” 등은 본래 의존명사이지만, 독립되어 쓰일 경우 모두 일반명사의 자격을 가지므로 일반명사로 분석해야 한다.

- | | |
|---------------------------|----------|
| ■■ 나는 <u>월</u> 30만원을 받는다. | [월/NNG] |
| ■■ <u>달러</u> 의 가치는 | [달러/NNG] |

(2) 단위를 나타내는 표현

(가) 길이, 무게, 수효, 시간 따위의 수량을 수치로 나타내는 단위들 중 “미터, 그램, 리터” 등은 의존명사(NNB)로, 외국어로 된 “m, g, l” 등은 기호(SW)로 분석한다.

(나) 일반명사가 단위적인 용법으로 쓰인 경우에는 의존명사가 아니므로 주의한다.

- | | |
|----------------------------|----------------|
| ■■ 사람, 시간, 그릇, ... | |
| ■■ 한 <u>사람</u> 이 교실로 들어왔다. | [사람/NNG+이/JKS] |
| ■■ 자장면 한 <u>그릇</u> 만 주세요. | [그릇/NNG+만/JX] |

(3) ‘것’과 구어형 ‘거’의 분석

- ‘거’의 형태를 그대로 인정하여 분석한다. 그러나 다른 형태와의 결합에서 ‘거’의 형태가 유지되지 않는다면 그 때에는 ‘것’으로 복원하여 분석한다.

- | | |
|---------------------------|---------------|
| ■■ 공부할 <u>거</u> 를 준비해 왔니? | [거/NNB+를/JKO] |
|---------------------------|---------------|

■■ 공부할 걸 가져왔니?	[것/NNB+ㄹ/JKO]
■■ 연습할 건 있니?	[것/NNB+ㄴ/JX]
■■ 먹을 게 모자르다	[것/NNB+이/JKS]

※ [보완] 학습자의 오류로 인해 ‘거’의 형태가 유지되지 않는 경우는, ‘것’으로 복원하지 않는다.

■■ 밥을 먹을 건다.	[거/NNB+이/VCP+ㄴ 다/EF]
--------------	----------------------

2) 대명사(NP)

- 대명사는 그 자체로는 자신의 본유적 지시물을 가지지 않은 채, 다만 사람이나 사물 등 어떤 대상을 간접적으로 지시하는 품사이다. 단, 동일한 대명사가 방언이나 고어의 이형태를 가진 경우에는 이들도 대명사로 같이 분석한다.

(1) 1인칭 대명사

(가) 1인칭 대명사

■■ 나, 내, 우리, 저, 제, 저희

(나) 2인칭 대명사

■■ 너, 네, 그대, 당신, 댁, 어르신

(다) 기타 대명사

■■ 이이, 이분, 그이, 그분, 저이, 저분, 아무, 아무개, 누구, 무엇, 뭐, 어디, 언제, 자기, 개, 재, 애, 이것, 저것, 그것, 이거, 저거, 그거, 여기, 저기, 거기, 이곳, 그곳, 저곳, 어디, 모(某), 모모(某某)

※ [보완] ‘자기’는 대명사로 분석한다.

※ [보완] ‘자신’, ‘아무것’은 일반명사로 분석한다.

※ [보완] ‘우리나라’는 한국인이 사용하는 경우 ‘우리 한민족이 세운 나라를

스스로 이르는 말.’의 뜻의 일반명사로 분석하지만, 외국 학생들이 사용하는 경우 ‘우리/NP ∥ 나라/NNG’로 분석해야 한다. 학습자 말뭉치의 경우 외국 학생들의 작문이나 구어 전사 텍스트이므로 ‘우리나라(우리 나라)’가 등장하는 경우 모두 ‘우리/NP ∥ 나라/NNG’으로 분석한다.

(2) 대명사와 관형사의 두 가지 분석이 가능한 단어

(가) ‘모(某)’는 관형사와 대명사로 분석될 수 있으므로 주의를 요한다.

■■ 모 기업체	[모/MM]
■■ 김 모 씨	[모/NP ∥ 씨/NNB]

(나) ‘모모(某某)’도 위와 같이 분석될 수 있다.

■■ 모모가 말했다	[모모/NP+가/JKS]
■■ 모모 기관의 조사를 마쳤다	[모모/MM]

(3) 대명사의 이형태 분석

(가) ‘이것, 그것, 저것; 이거, 그거, 저거’는 분석하지 않고 대명사로 인정한다. ‘~거’의 경우, 다른 형태와의 결합에서 ‘~거’의 형태가 유지되지 않는다면 그 때에도 ‘~것’으로 복원한다.

■■ 난 저거를 먹을래.	[저거/NP+를/JKO]
■■ 나는 여태 그걸 믿어 왔단다.	[그것/NP+ㄹ/JKO]

(나) 다음과 같이 원형을 밝힐 수 있는 대명사는 원형대로 분석한다.

■■ 내	이제부터는 내 명령을 따라라.	[나/NP+의/JKG]
■■ 내게	내게 전자우편으로 알려 다오.	[나/NP+에게/JKB]
■■ 네게	어제 네게 보낸 선물이 잘못되었다.	[너/NP+에게/JKB]
■■ 제게	문제가 있다면 제게 말씀해 주세요.	[저/NP+에게/JKB]
■■ 누가	누가 전화를 하는 지 보고해라.	[누구/NP+가/JKS]
■■ 뉘	뉘 집 애기가 울고 있는 거야?	[누구/NP+의/JKG]
■■ 뭐가	도대체 뭐가 문제라는 거야?	[뭐/NP+가/JKS]

※ [참고] ‘내가’는 모두 ‘내/NP+가/JKS’로 분석한다.

■■ 내가	내가 살던 집	[내/NP+가/JKS]
-------	---------	--------------

(다) ‘뭘’는 ‘무엇’과 대등할 정도로 자주 사용되므로 그 형태 자체를 인정해 준다. 다만, 다음과 같이 조사와 축약되었을 경우에만 원형으로 복원해 준다.

■■ 앞으로 우리가 뭘 하자는 얘기냐? [무엇/NP+ㄹ/JKO]

(라) '제'의 경우, '제/NP+가/JKS'를 제외하고는 모두 '저/NP+의/JKG'로 분석한다.

■■ 제가 갈 것입니다. [제/NP+가/JKS]

■■ 철수는 제 잘못을 안다. [저/NP+의/JKG]

■■ 제 무게를 못 견디다. [저/NP+의/JKG]

※ [보완] 학습자가 대명사 뒤에서 조사를 누락해서 쓴 경우와 형태적 유사함이 있기 때문에 분석에 주의해야 한다. 이 경우는 ‘저/NP+의/JKG’ 또는 ‘나/NP+의/JKG’로 분석하지 않는다

■■ 제 먹었습니다. [제/NP]

■■ 내 활짝 웃었다. [내/NP]

3) 수사(NR)

- 수사는 사물의 수량이나 차례를 나타내는 품사를 말한다.

(1) 수사의 종류

(가) 양수사

■■ 하나, 둘, 셋, 넷, 다섯, 여섯, 일곱, 여덟, 아홉, 열, 스물, 서른, 마흔, 쉰, 예순, 일흔, 여든, 아흔, 백한둘, 두서넛, 서넛, 너넛, 네다섯, 네댓, 대여섯, 예닐곱, 일여덟, 일고여덟, 열두서넛, 열대여섯, 열일고여덟, 스물두서넛

■■ 일, 이, 삼, 사, 오, 육, 칠, 팔, 구, 십, 백, 천, 만, 억, 조

■■ 기십, 기백, 기천,

■■ 수십, 수백, 수천, 수만, 수억, 수십만, 수백만, 수천만

(나) 서수사

■■ 첫째, 둘째, 셋째, 넷째, ..., 열째, 열한째,..., 스물한째,...,

■ ■ 아흔아홉째, 백째, 백한째, ...

※ [보완] '째'는 분석하는 접미사에 해당하지만 서수사에서 쓰인 경우 분석하지 않는다.

■ ■ 첫째

[첫째/NR]

■ ■ 첫 번째

[첫/MMⅡ 번/NNB+째/XSN]

<주의사항>

(가) 복수의 수사가 한 어절 내에 나타날 때에는 전체를 통합해서 분석한다.

■ ■ 백만오천삼십사

[백만오천삼십사/NR]

(나) ‘하나’는 표준국어대사전에 그 품사가 명사와 수사로 되어 있지만 본 지침에서는 수사로 분석한다.

■ ■ 광에 가서 물건 하나만 가져오렴.

[하나/NR+만/JX]

■ ■ 우리는 하나로 뭉쳤다.

[하나/NR+로/JKB]

(다) [보완] 때로 수사와 수관형사의 구별이 애매한 경우가 있다. 이 분석에서는 임홍빈(1998)의 견해에 따라, 다음과 같은 특이한 형식을 가진 예만을 수관형사로 취급하고, 그 밖의 것들은 모두 수사로 분석한다.

■ ■ 한, 한두, 한두어, 두, 두어, 두세, 두서너, 세, 석, 서, 서너, 네, 너, 녀

■ ■ 열한, 스물두, 서른세 등

→ 수관형사로 취급하는 특이한 형식으로 끝나는 경우는 모두 수관형사로 취급한다.

(라) ‘제일, 제이’ 등은 접두사 ‘제-’와 수사의 결합으로 분석한다.

■ ■ 제일, 제이, 제삼, 제사, 제오, ..., 제구십구, 제백, ... [제/XPN+일/NR],

[제/XPN+이/NR], ...

나. 용언

- 용언은 동사, 형용사, 지정사를 가리킨다. 용언 범주에서는 분석 대상이 본 용언일 경우에만 동사와 형용사로 구분하여 표시하고, 보조용언의 경우에는 보조동사와 보조형용사를 구분하지 않고 'VX'라는 하나의 표지만을 준다. 또한 학교 문법에서 서술격조사로 다루는 '이다'는 조사의 범주에 넣지 않고 '지정사'라는 용언의 하위범주에 넣기로 한다. 지정사는 다시 긍정 지정사(VCP)와 부정 지정사(VCN)로 세분된다.

1) 동사(VV)

- 동사는 사물의 움직임이나 작용을 나타내는 용언을 말한다. 동사는 일반적으로 목적어의 필요성 여부에 따라 자동사, 타동사로 나누기도 하지만, 본 분석에서는 그것을 위한 별도의 표지를 세분하지 않고 모두 'VV'로 표시한다.

※ [보완] '있다'는 모두 **동사**로 처리한다. (세종 말뭉치 기준)

※ [보완] '감사하다'는 모두 **동사**로 보고 '-하-'는 모두 동사파생접미사로 처리한다. (세종 말뭉치 기준)

※ [보완] '명사/어근/부사 + (분석 목록에 없는) 동사파생접미사'는 전체를 통합하여 동사로 분석한다.

■■ 말씀드리다

[말씀드리/VV+다/EF]

■■ 반짝거리다

[반짝거리/VV+다/EF]

2) 형용사(VA)

- 형용사는 사물의 성질이나 상태를 나타내는 용언을 가리킨다.

※ [보완] '명사/어근/부사 + (분석 목록에 없는) 형용사파생접미사'는 전체를 통합하여 형용사로 분석한다.

■■ 나다

[별나/VA+다/EF]

■■ 맞다

[능글맞/VA+다/EF]

3) 보조용언(VX)

1. 사전 등재

예) 가늘어지다 가늘어지/VV+다/EF

 좋아하다 좋아하/VV+다/EF

2. 사전 미등재

예) 심해지다 심하/VA+아/EC+지/VX+다/EF

 초조해하다 초조/NNG+하/XSA+아/EC+하/VX+다/EF

→ 이 분석에서는 보조용언을 보조동사와 보조형용사로 하위 구분하지 않는다.

(1) 보조용언 분석 원칙

(가) 보조용언의 후보는 표준국어대사전에 그 쓰임이 제시되어 있어야 한다.

(나) 보조용언 앞에는 반드시 다른 용언이 위치해 있어야 한다.

(다) 보조용언이 동시에 두 개 이상이 연결되어 나타날 수도 있다.

(2) 보조용언의 예와 주의사항

- 보조용언의 목록은 다음과 같다. 이 목록은 표준국어대사전을 참고한 것이다.

■■ 가다	세월이 흘러 가는 대로 떠도는 나그네	가/VX+는/ETM
■■ 가지다	그렇게 해 가지고는 기일을 맞출 수 없다.	가지/VX+고/EC+는/JX
■■ 계시다	손님께서 와 계십니다.	계시/VX+ㅂ니다/EF+./SF
■■ 나가다	추진해 나가는 과정에서 문제가 생겼다.	나가/VX+는/ETM
■■ 나다	아침에 깨어 나 보니 그가 없어졌다.	나/VX+아/EC
■■ 내다	힘들겠지만 잘 견뎌 내야 한다.	내/VX+아야/EC

■ ■ 놓다	약속을 잡아 놓고 출장을 가다니	놓/VX+고/EC
■ ■ 달다	이번 시험 문제의 정답을 알려 다오.	달/VX+오/EF+./SF
■ ■ 대다	자꾸 졸라 대는 통에 허락해 주고 말았다.	대/VX+는/ETM
■ ■ 두다	남겨 둔 쌀도 이제 바닥이 났다.	두/VX+ㄴ/ETM
■ ■ 드리다	염려를 끼쳐 드려 송구하옵니다.	드리/VX+어/EC
■ ■ 들다	도무지 내 말은 믿으려 들지 않는다.	들/VX+지/EC
■ ■ 말다	어렵더라도 희망을 잃지 말아야 한다.	말/VX+아야/EC
■ ■ 먹다	나는 오늘도 수업을 빼 먹었다.	먹/VX+었/EP+다/EF+./SF
■ ■ 못하다	그 참상을 차마 보지는 못할 것이다.	못하/VX+ㄹ/ETM
■ ■ 버리다	음식이 다 타 버렸다.	버리/VX+었/EP+다/EF+./SF
■ ■ 보다	이제는 새벽이 오는가 보다.	보/VX+다/EF+./SF
■ ■ 빠지다	썩어 빠진 생선을 사오다니	빠지/VX+ㄴ/ETM
■ ■ 싶다	너를 보고 싶다.	싶/VX+다/EF+./SF
■ ■ 쌓다	고치고치 물어 쌓는 통에 정신이 없었다.	쌓/VX+는/ETM
■ ■ 아니하다	일이 순리대로 풀리지 아니했다.	아니하/VX+았/EP+다/EF+./SF
■ ■ 앓다	시간이 지나도 기차는 오지 않았다.	앓/VX+았/EP+다/EF+./SF
■ ■ 오다	고향을 떠나 온 지 10년이 지났다.	오/VX+ㄴ/ETM
■ ■ 있다	그녀는 검정 옷을 입고 있었다.	있/VX+었/EP+다/EF+./SF
■ ■ 주다	아버지는 아기에게 동화책을 읽어 주었다.	주/VX+었/EP+다/EF+./SF
■ ■ 지다	한 번 넘어 진 아이는 일어나는 법을 안다.	지/VX+ㄴ/ETM
■ ■ 치우다	다섯 명이 10인분의 식사를 먹어 치웠다.	치우/VX+었/EP+다/EF+./SF
■ ■ 터지다	끓인 지 오래 되어서 라면이 불어 터졌다.	터지/VX+었/EP+다/EF+./SF
■ ■ 하다	나귀를 쉬게 하는 것이 좋겠다.	하/VX+는/ETM

① 다음과 같은 어절은 보조용언으로 취급되기도 하나, 여기서는 ‘의존명사+접사’로 분석한다. 이들 앞에는 항상 관형어가 온다는 분포적인 특성을 중시한 것이다.

■ ■ 양하다/채하다/척하다/듯하다/법하다/뻔하다 [양/NNB+하/XSA+다/EF]

■■■ 듯싶다

[듯싶/VX+다/EF]

※ 표준국어대사전에 따라, 기존에 접미사로 분석하던 ‘만하’의 지침을 변경하여, ‘만’을 보조사로, ‘하’를 동사로 분석한다. ‘만하’는 ‘만/NNB+하/XSA’로 분석되는 경우도 있으므로 주의해야 한다.²⁹⁾

■■■ 철수만 한 인재가 없다

[철수/NNP+만/JX || 하/VV+L/ETM]

■■■ 이 음식은 먹을 만하다.

[만/NNB+하/XSA+다/EF+./SF]

② ‘버릇하다’의 경우에는 선행 성분으로 관형형이 오는 것은 아니지만, 일반명사 ‘버릇’과 크게 구별되지 않으므로 ‘버릇’은 명사로 분석한다.

■■■ 자꾸 울어 버릇하다.

[버릇/NNG+하/XSV+다/EF+./SF]

※ [보완] ‘-도록 하다’는 형용사 일부 어간에만 사용되는 등 ‘-게 하다’와 분포가 다르므로 이때의 ‘하다’는 본용언으로 분석한다.

■■■ 열심히 공부하도록 하자.

[공부/NNG+하/XSV+도록/EC || 하/VV+자/EF]

4) 지정사(VC)

- 지정사는 학교 문법의 서술격 조사에 대응되는 것인데, 용언과 같이 활용한다는 특성을 중시한 술어이다. 여기서는 학교 문법의 ‘이다’를 긍정 지정사로, ‘아니다’를 부정 지정사로 하위 구분한다. 일반적으로 ‘아니다’는 형용사로 다루어지기도 하나, 여기서는 ‘아니다’가 ‘이다’의 부정형이라는 점을 중시하여 ‘부정지정사’로 다룬다.

■■■ 철수는 매우 우수한 학생이다.

[학생/NNG+이/VCP+다/EF+./SF]

■■■ 철수는 모범적인 학생이 아니다.

[아니/VCN+다/EF+./SF]

※ [참고] 지정사 ‘이/VCP’를 복원해야 하는 경우

① 체언에 어미가 직접 연결된 경우

■■■ 철수는 훌륭한 교사다.

[교사/NNG+이/VCP+다/EF+./SF]

29) ‘바. 3) 다) 형용사파생접미사’의 주의사항의 내용 이동함.

② 조사에 어미가 직접 연결된 경우

■ ■ 우리가 그를 본 것은 서울에서다. [서울/NNP+에서/JKB+이/VCP+다/EF+./SF]

③ ‘~였다’

■ ■ 그 당시 나는 아이였다. [아이/NNG+이/VCP+였/EP+다/EF+./SF]

④ 어미 '-라고, -라는, -라도, -라며, -라면서, -라서

■ ■ 나는 그에게 절교라고 말했다. [절교/NNG+이/VCP+라고/EC]

■ ■ 나는 친구라는 말이 좋다. [친구/NNG+이/VCP+라는/ETM]

■ ■ 거지라도 존중해 주어야 한다. [거지/NNG+이/VCP+라도/EC]

■ ■ 그는 최고라며 나를 추켜 주었다. [최고/NNG+이/VCP+라며/EC]

■ ■ 그는 실수라면서 얼버무렸다. [실수/NNG+이/VCP+라면서/EC]

■ ■ 너는 부자라서 우릴 이해하지 못할 것이다. [부자/NNG+이/VCP+라서/EC]

⑤ 인용문 뒤에 오는 “~며”는 지정사를 복원하지 않는다.

■ ■ 얼마나 친절하냐?”며 [친절/NNG + 하/XSA + 나/EF + ?/SF + "/SS + 며/EC]

⑥ [보완] ‘아서/어서’에 종결어미가 결합된 경우 (세종 말뭉치)

■ ■ 없어진 것을 확인하기 위해서다. [위하/VV+아서/EC+이/VCP+다/EF+./SF]

■ ■ 그때 그 시절의 사람들이 생각나서다. [생각나/VV+아서/EC+이/VCP+다/EF+./SF]

■ ■ 내가 개를 좋아하는 건 개가 착해서야. [착하/VA+아서/EC+이/VCP+야/EF+./SF]

<주의사항>

(가) [보완] 학습자가 지정사 ‘이’를 몰라서 누락한 경우는 ‘이/VCP’를 복원하지 않는다.

■ ■ 방법은 한 가지예요. [가지/NNB+예요/EF]

■ ■ 이것은 책상라며 나를 가르쳤다. [책상/NNG+라며/EC]

(나) [보완] 학습자가 ‘예요’를 써야하는 부분에서 ‘예요’로 쓴 경우는 ‘이/VCP+예요/EF’로 분석하지 않고 종결어미의 오형태로 분석한다.

■■ 저는 OO어학당 6급 학생이에요.

[학생/NNG+이/VCP+예요/EF]

※ ‘아니다’는 부정 지정사(VCN)으로 분석한다.

다. 수식언

1) 관형사(MM)

- 관형사는 체언 앞에서 그것을 꾸미는 품사를 말한다. 관형사는 지시관형사, 수관형사, 성상관형사로 세분될 수 있는데, 본 분석에서는 이를 세분하여 분석하지 않는다.

■■ 각(各)	각 가정	[각/MM]
■■ 그까짓	그까짓 일	[그까짓/MM]
■■ 전(全)	전 국민	[전/MM]
■■ 현(現)	현 정권	[현/MM]

<주의사항>

(가) 관형사는 때로 문맥에 따라 다른 품사로 분석될 가능성이 있으니 문맥을 잘 살펴서 분석해야 한다.

① 관형사, 명사 통용

■■ 올 예산이 다 바닥이 났다.	[올/MM]
■■ 올 들어 물가가 많이 올랐다.	[올/NNG]

② 관형사, 부사 통용

■■ 단 세 명에서 그 일을 꾸몄다.	[단/MM]
■■ 단, 그 일은 해서는 안 된다.	[단/MAJ]

③ 관형사, 명사, 부사 통용

■■ 이내 마음을 어찌 알리요.	[이내/MM]
■■ 아침 들판에 <u>이내</u> 가 끼었다.	[이내/NNG]

■ ■ 그는 이내 떠나갔다.

[이내/MAG]

(나) 수사가 명사를 단독으로 수식하는 경우 그것을 관형사로 분석하기 쉬우나, '수'를 나타내는 말 가운데서 앞서 언급한 수관형사를 제외하고는 수사는 오로지 수사로만 분석한다. 즉, 수사의 관형사적 쓰임을 인정하지 않는 것이다. 따라서 다음과 같이 '다섯'은 모든 환경에서 중의성 없이 '수사'로만 분석된다. (1.3 수사 [2]주의사항 참고)

■ ■ 다섯이 먹기에 충분하다.

[다섯/NR+이/JKS]

■ ■ 다섯 명이 앉아 있었다.

[다섯/NR]

(다) 접미사 '-적(的)'이 붙는 경우는 조사와의 결합여부와 관계없이 모두 명사로 분석한다.

■ ■ 명사의 부사적인 용법

[부사/NNG+적/XSN+이/VCP+L/ETM]

■ ■ 명사의 부사적 용법

[부사/NNG+적/XSN]

2) 부사(MA)

- 부사는 주로 용언을 꾸며서 그 뜻을 더 세밀하고 분명하게 해 주는 품사를 말한다. 여기서는 부사를 세분하지 않고, 접속부사와 일반부사로만 나누기로 한다.

가) 접속부사(MAJ)

<주의사항>

① 접속부사는 종종 용언의 활용형으로도 쓰일 수 있으므로 주의한다.

■ ■ 그래서 마지막에는 조심하라고 했지?

[그래서/MAJ]

■ ■ 영희가 그래서 결석을 했구나.

[그렇/VX + 어서/EC]

② '그리고나서', '그래도'의 분석

■ ■ 그리고 나서

[그리/MAG+하/XSV+고/EC || 나/VX+서/EC]

■ ■ 그래도

[그러/VV+어도/EC]

- ※ [보완] 접속부사는 《표준국어대사전》에 접속부사로 뜻풀이된 것만 인정한다.
아래는 《표준국어대사전》의 접속부사 목록이다.

건데, 고로01 「2」, 그래서, 그러나, 그러니까, 그러면, 그러므로, 그런데, 그럼01 「1」, 그렇지만, 그렇지만, 그리고, 그리하여, 근데01, 단06, 따라서, 연이나, 연중에, 연즉, 이리하여, 하건만, 하기는, 하기에, 하긴, 하물며, 하지만, 한데03

※ 용언의 활용형

■■ 그래, 그래도, 그래야, 그러니, 그러다가, 그러매, 그러면서, 그러자, 그렇다면, 그렇잖아도, 그리한 즉

※ 일반 부사

■■ 게다가, 곧, 다만, 또, 또는, 또한, 및, 예컨대, 요컨대, 왜냐하면, 이를테면, 한편, 혹시, 혹은

- ※ [보완] 사전에 등재되지 않은 부사의 약어는 본딴말과 같은 표지로 분석한다.

■■ 그니까(그러니까), 글고(그리고)/MAJ

■■ 왜냐면(왜냐하면)/MAG

나) 일반부사(MAG)

<주의사항>

- ① 일반부사는 종종 일반명사와 동일형태를 띠고 있어 구분이 어려운 경우가 있다. 이들은 뒤에 조사가 결합하느냐의 여부와, 문맥에서 후행 명사를 수식하느냐의 여부에 따라 부사와 명사로 분석될 수 있다.

■■ 너의 진짜 속셈이 무엇인지 말해 봐라.	[진짜/NNG]
■■ 그 수학 문제는 진짜 어려웠다.	[진짜/MAG]
■■ 지금이 공부하기 딱 좋은 때이다.	[지금/NNG+이/JKS]
■■ 나는 지금 막 집에 도착했다.	[지금/MAG]

■■■ 오늘은 내 생일이 아니다.

[오늘/NNG+은/JX]

■■■ 오늘은 그가 왔다.

[오늘/MAG+은/JX]

- ② 부사적인 용법을 가졌음에도 불구하고 일반부사가 아닌 일반명사로만 표준 국어대사전에 등재되어 있는 단어는 오로지 일반명사로만 분석한다.

■■■ 구석구석, 무작정, 여기저기, 오랫동안, 이곳저곳, 정작, 좌우간, 처음, 최근, 한때

- ③ 일반부사로 분석하기 쉬운 활용상의 불완전동사인 ‘덜달아, 더불어’는 모두 동사로 옳게 분석해야 함에 주의한다.

■■■ 너는 덜달아 왜 난리니?

[덜달/VV+아/EC]

■■■ 우리 함께 더불어 살아가자.

[더불/VV+어/EC]

- ④ ‘명사+없이’는 원칙적으로 ‘일반명사+없이/MAG’로 분석하지만, 아래와 같이 하나의 단어로 굳어져 사전에 등재된 경우는 ‘없이’ 통합형 자체를 하나의 일반부사로 분석한다.

■■■ 관계없이, 그지없이, 꾸밈없이, 끊임없이, 난데없이, 남김없이 등

라. 독립언

1) 감탄사(IC)

- 감탄사는 화자의 부름이나 느낌, 놀람이나 대답을 직접적으로 나타내는 품사를 말한다.

■■■ 그럼(요), 야호, 어머, 앓, 아, 예, 그래(요), 아니(요), 글썄, 참, 아이구, 와아, 오호, 세상에

<주의사항>

- ① 사람이 입으로 직접 내는 소리를 대상으로 하되, 흥내를 내는 의도가 없는 것과 본능적인 놀람이나 느낌을 나타내는 것을 대상으로 한다. 또한 감탄사와 혼동되는 부사로서 음성상징어류의 부사어가 있는데, 이는 감탄사가 아닌 일반부사로 분석한다.

■■ 야호! 드디어 정상이다.

[야호/IC+!/SF]

■■ 쿨럭쿨럭 기침을 했다.

[쿨럭쿨럭/MAG]

② 동물의 울음소리 등은 감탄사가 아니라 일반부사로 분석한다.

■■ 검둥이는 멍멍 짖으며 수풀 속으로 뛰어 갔다. [멍멍/MAG]

③ 욕이나 욕설을 나타내는 말은 전체를 감탄사로 분석한다.

■■ 빌어먹을! [빌어먹을/IC+!/SF]

④ ‘뭐’는 문맥에 따라 대명사와 감탄사의 두 가지 쓰임이 있다.

■■ 원지도 모른 채 [뭐/NP+이/VCP+ㄴ지/EF+도/JX]

■■ 신문에 뭐 대단한 특종이라도 실렸습니까? [뭐/IC]

⑤ 한 어절이 비정상적으로 늘어나거나 다른 기호가 개입되었을 경우 분석불능 범주(NA)로 분석한다.

■■ 그러어엄/NA, 으~어~이/NA

마. 관계언³⁰⁾

- 조사는 주로 체언과 결합하여 다른 말과의 문법적 관계를 나타내거나, 특별한 뜻을 더해 주는 품사를 말한다. 조사는 크게 격조사, 보조사, 접속조사로 나눈다. 한국어는 조사가 중첩하는 경우가 많은데, 이러한 경우 조사의 결합형은 분리해서 분석함을 원칙으로 한다.

■■ 부산에서도 대형 사고가 있었다. [부산/NNP+에서/JKB+도/JX]

■■ 그녀와의 약속이 갑자기 잡혔다. [그녀/NP+와/JKB+의/JKG]

1) 격조사(JK)

- 이는 체언과 다른 성분 간의 일정한 문법 관계를 나타내는 조사이다.

30) 지침에 제시된 조사 목록에서 빠진 이형태와 예시를 추가함

가) 주격조사(JKS)

- 선행 체언으로 하여금 주어가 되게 하는 조사이다.

■■ 이/가	책이 보인다.	[책/NNG+이/JKS]
	나무가 보인다.	[나무/NNG+가/JKS]
■■ 께서	선생님께서 오신다.	[선생/NNG+님/XSN+께서/JKS]
■■ 서/이서	둘이서 그 일을 꾸몄다고?	[둘/NR+이서/JKS]
	혼자서 그 일을 꾸몄다고?	[혼자/NNG+서/JKS]
■■ 께오서	부대장님께서 오서	[부대장/NNG+님/XSN+께오서/JKS]
■■ 께옵서	황제께옵서 드나드신다.	[황제/NNG+께옵서/JKS]

※ [보완] 다음과 같이 체언 뒤에서 ‘이’가 첨가되어 나타나는 오류의 경우, 이 때 ‘이’는 모두 주격 조사로 분석한다.

■■ 닭이가 울었다.	[닭/NNG+이/JKS+가/JKS]
■■ 책상이가 있다.	[책상/NNG+이/JKS+가/JKS]
■■ 친구들이 6월에 일이를 찾아있었다.	[일/NNG+이/JKS+를/JKO]
■■ 좋아하는 거는 옷이입니다.	[옷/NNG+이/JKS+이/VCP+ㅂ니다/EF]
■■ 어른이들이 수많은 노력을	[어른/NNG+이/JKS+들/XSN+이/JKS]

나) 보격조사(JKC)

- 선행 체언으로 하여금 서술어 ‘되다, 아니다’의 보어가 되게 하는 조사이다. ‘되다, 아니다’ 앞의 조사 ‘이, 가’는 모두 보격조사로 분석한다.

■■ 이/가	얼음이 물이 되었다.	[물/NNG+이/JKC]
	씨앗이 열매가 되었다.	[열매/NNG+가/JKC]
	철수는 범인이 아니다.	[범인/NNG+이/JKC]
	범인은 남자가 아니다.	[남자/NNG+가/JKC]

다) 목적격조사(JKO)

- 선행 체언으로 하여금 목적어가 되게 하는 조사이다.

■■ ㄹ/을/를	수지가 널 좋아해.	[너/NP+ ㄹ/JKO]
	민수는 음식을 많이 먹는다.	[음식/NNG+을/JKO]
	너는 바람 소리를 들었다.	[소리/NNG+를/JKO]

라) 관형격조사(JKG)

- 선행 체언으로 하여금 관형어가 되게 하는 조사이다.

■■ 의	나의 친구는 너 하나뿐이다.	[나/NP+의/JKG]
------	-----------------	--------------

마) 부사격조사(JKB)

선행 체언으로 하여금 부사어가 되게 하는 조사이다.

■■ 로/으로	망치로 못을 박아야지.	[망치/NNG+로/JKB]
	음식으로 장난치지 마.	[음식/NNG+으로/JKB]
■■ 로서/으로서	교사로서 책임을 다해야 한다.	[교사/NNG+로서/JKB]
	장관으로서 책임을 다해야 한다.	[장관/NNG+으로서/JKB]
■■ 로써/으로써	돌로써 지붕을 만든다고?	[돌/NNG+로써/JKB]
	콩으로써 메주를 쏜다고 해도	[콩/NNG+으로써/JKB]
■■ 같이	바보같이 웃고 다닌다.	[바보/NNG+같이/JKB]
■■ 더러	나더러 이것도 하라고 한다.	[나/NP+더러/JKB]
■■ 랑/이랑	너랑 많이 닮았다.	[너/NP+랑/JKB]
	오늘 동생이랑 싸웠다.	[동생/NNG+이랑/JKB]
■■ 로부터/	TV로부터 받는 영향력이	[TV/SL+로부터/JKB]
으로부터	시험으로부터 해방되다	[시험/NNG+으로부터/JKB]
■■ 마냥	기영이마냥 놀 수만은 없다.	[기영이/NNP+마냥/JKB]
■■ 마따나	네 말마따나 나도 그래야 한다.	[말/NNG+마따나/JKB]
■■ 만큼	눈물만큼 콧물도 흐른다니까.	[눈물/NNG+만큼/JKB]
■■ 보고	영자보고 놀자고 좀 해라.	[영자/NNP+보고/JKB]
■■ 보다	직관보다는 논리가 동원돼야 한다.	[직관/NNG+보다/JKB+는/JX]

■■ 에	나는 너에 대해 아무것도 모른다.	[너/NP+에/JKB]
■■ 에게	너에게 말하기 싫다.	[너/NP+에게/JKB]
■■ 에게서	나는 철수에게서 그 말을 들었다.	[철수/NNP+에게서/JKB]
■■ 에서	집에서 학교까지 너무 멀다.	[집/NNG+에서/JKB]
■■ 에서부터	연구소에서부터 가게까지는	[연구소/NNG+에서부터/JKB]
■■ 와/과	경미와 함께 다닌다면,	[경미/NNP+와/JKB]
	동생과 함께 다닌다면,	[동생/NNG+과/JKB]
■■ 처럼	사람처럼 행동하는 동물이 있다.	[사람/NNG+처럼/JKB]
■■ 하고	그 일하고 관련된 사람은	[일/NNG+하고/JKB]
■■ 한테	그 일은 경비한테 부탁해라	[경비/NNG+한테/JKB]

바) 호격조사(JKV)

- 주로 사람을 가리키는 체언 뒤에 연결되어 그것으로 하여금 부름의 대상이 되게 하는 조사이다.

■■ 아/야	호동아! 이제 그만 일어나거라	[호동/NNP+아/JKV+!/SF]
	철수야! 밥 먹어라	[철수/NNP+야/JKV+!/SF]
■■ 여/이여	주여, 우리에게 힘을 주소서	[주/NNG+여/JKV]
	슬픔이여, 안녕	[슬픔/NNG+이여/JKV]
■■ 시여/이시여	전능자시여 자비를 베풀어 주옵소서	[전능자/NNG+시여/JKV+!/SS]
	신이시여! 우리를 저버리지 마소서	[신/NNG+이시여/JKV+!/SS]

<주의사항>

- 호격조사와 어말어미는 구분해서 분석해야 한다.

■■ 저기 오는 것이 철수야. [철수/NNP+이/VCP+야/EF+./SF]

사) 인용격조사(JKQ)

- 인용문이나 인용구를, 동사에 대한 부사적 성분으로 도입하는 조사이다.

■■ 고	그는 "이제 가도 좋다"고 말했다.	[종/VA+다/EF+"/SS+고/JKQ]
■■ 라고/이라고	문제가 심각하다"라고 보고했다. 팻말에는 "금지구역"이라고 쓰여 있었다.	[심각/XR+하/XSA+다/EF+"/SS+라고/JKQ] ["/SS+금지/NNG+구역/NNG+"/SS+이라고/JKQ]
■■ 하고	영수는 "이제 가자"하고 말문을 달았다.	[가/VV+자/EF+"/SS+하고/JKQ]

<주의사항>31)

- ① 인용격조사는 연결어미와 구별하기 어려운 경우가 있으므로 주의한다. 인용기호가 있을 경우에만 인용격조사로 분석하고, 인용기호가 없는 경우 연결어미로 분석한다.

(1) 인용격조사

- 팻말에는 "금지구역"이라고 쓰여 있었다.
["/SS+금지/NNG+구역/NNG+"/SS+이라고/JKQ]
- 철수는 "다음 주에 놀러 가도 좋다"고 말하였다.
[종/VA+다/EF+"/SS+고/JKQ]
- 먼저 "주민등록증이 있냐?"고 묻는다.
[있/VV+냐/EF+?/SF+"/SS+고/JKQ]

(2) 연결어미

- 철수는 자기가 학생이라고 말했다.
[학생/NNG+이/VCP+라고/EC]
- 자장면을 시킨 뒤 집에 가겠다고 우기는 할머니를 달래기 시작했다.
[가/VV+겠/EP+다고/EC]
- 내가 안 기쁘냐고 다그쳐 물었을 때,
[기쁘/VA+냐고/EC]

31) <세종> 분석 결과를 바탕으로 다시 정리함

- ② [보완] 학습자 말뭉치에서는 생산자가 외국인 학습자이기 때문에 한국어에서 인용 기호로 구현되는 직접 인용, 간접 인용에 대한 지식이 없어 따옴표를 적지 못한 경우가 ‘문어’에서도 많이 발생한다. 이러한 경우는 <세종> 구어에서의 처리와 마찬가지로 인용 기호가 없더라도 직접 인용인 경우 인용격 조사로 분석한다.

■■ 내 일이다라고 말했다. [일/NNG+이/VCP+다/EF+라고/JKQ]

※ [참고] 다음은 간접 인용의 경우로 보고 분석한다.

■■ 내 일이라고 말했다. [일/NNG+이/VCP+다고/EC]

- ③ 인용 기호 중 하나인 <“ ”>은 맥락에 따라 인용이 아닌 강조를 위해 사용되기도 한다. 이때는 인용격 조사로 분석하지 않도록 주의한다.

■■ “사랑”이라는 건 뭘까? ["/SS+사랑/NNG+"/SS+이/VCP+라는/ETM]

■■ 철수는 자기가 “학생”이라고 말했다. ["/SS+학생/NNG+"/SS+이/VCP+라고/EC]

※ [참고]

■■ 시골 아이라고 그것도 모르겠니? [아이/NNG+라고/JX]

2) 접속조사(JC)

- 두 단어를 같은 자격으로 이어 주는 구실을 하는 조사를 말한다.

■■ 고/이고	그 사람은 염치고 체면이고가 없어.	[염치/NNG+고/JC]
	책이고 책상이고 다 타 버렸다.	[책/NNG+이고/JC]
■■ 와/과	그 아주머니는 딸기와 사과를 샀다.	[딸기/NNG+와/JC]
	그 기계는 사람과 컴퓨터를 구별하지 못한다.	[사람/NNG+과/JC]
■■ 나/이나	사과나 배는 모두 몸에 좋은 과일이다.	[사과/NNG+나/JC]
	바자회 물품으로 책이나 옷을 받고 있다.	[책/NNG+이나/JC]
■■ 니/이니	시장에는 사과니 배니 과일이 잔뜩 있다.	[사과/NNG+니/JC]
	떡이니 과일이니 잔뜩 먹었다.	[떡/NNG+이니/JC]
■■ 다/이다	그는 농구다 축구다 못하는 운동이 없다.	[농구/NNG+다/JC]
	연습이다 레슨이다 시간이 하나도 없다.	[연습/NNG+이다/JC]
■■ 랑/이랑	머루랑 다래랑 먹으며 청산에 살고 싶어라.	[머루/NNG+랑/JC]

	떡이랑 과일이랑 많이 먹었다.	[떡/NNG+이랑/JC]
■■ 며/이며	잔칫상에는 배며 대추며 여러 가지 과일이 차려져 있었다.	[배/NNG+며/JC]
	그림이며 조각이며 미술품으로 가득 찬 화실	[그림/NNG+이며/JC]
■■ 에	아버지가 책에, 연필에 많이 사 주셨다.	[책/NNG+에/JC]
■■ 하고	이번 준비물로 칼하고 연필을 샀다.	[칼/NNG+하고/JC]

<주의사항>

- ① ‘함께 함’의 뜻을 나타내는 접속조사는 부사격조사와 형태상 동일하므로 주의할 필요가 있다.

■■ 철수와 영희가 왔다.	[철수/NNP+와/JC]
■■ 철수와 같이 놀았다.	[철수/NNP+와/JKB]
■■ 철수랑 영희랑 왔다.	[철수/NNP+랑/JC 영희/NNP+랑/JC]
■■ 나는 철수랑 영희랑 같이 놀았다.	[철수/NNP+랑/JC 영희/NNP+랑/JC]

- ② 표준국어대사전에 조사로 등재(주로 구어체의 경우)된 ‘하며’는 조사로 인정하지 않고 ‘하/VV+며/EC’로 분석한다.

- ③ [보완] 접속 조사 중에서 ‘고/이고’, ‘니/이니’, ‘다/이다’, ‘며/이며’, ‘에’의 경우는 주로 ‘-고 -고’, ‘-니 -니’와 같은 구성에서 쓰인다. 이들 접속 조사는 연결어미와 동일한 형태인 경우가 있으므로 주의할 필요가 있다.

■■ 슬픔이고 기쁨이고 느끼지 못한다.	[슬픔/NNG+이고/JC]
■■ 그 옷은 <u>개성적이고</u> 색다른 현대 감각을 보여준다면,	[개성/NNG+적/XSN+이/VCP+고/EC]

■■ <u>옷이며</u> 신이며 흠어져 있었다.	[옷/NNG+이며/JC]
■■ 내부는 어지러운 공간이며, 같은 건물 안에 반드시 식당가가 있다.	[공간/NNG+이/VCP+며/EC]

- ④ [보완] 학습자의 오류로 인해 두 단어를 이어주는 병렬 구조가 제시되지 않더라도 의미상 접속 조사로 쓰인 경우에는 접속 조사로 분석한다.

■■ 친구에게 줄 <u>꽃과</u> 샀어요.	[꽃/NNG+과/JC]
--------------------------	--------------

■■■ 나에게 준 배려심이나 사람을 얼마나 많은지
어떻게 계산하는지 이제 마음속에는 다 알게 되었다. [배려심/NNG+이나/JC]

3) 보조사(JX)

- 체언이나 부사 또는 용언의 연결 어미나 종결 어미의 뒤에 쓰여 특별한 뜻을 더해 주는 조사를 말한다.

■■■ 그려/그래	종습니다그려.	[종/VA+습니다/EF+그려/JX+./SF]
■■■ 까지(꺼정/까장)	걸어서 하늘까지	[하늘/NNG+까지/JX]
■■■ 깨나	힘깨나 쓰게 생겼다.	[힘/NNG+깨나/JX]
■■■ 나/이나	너나 가라!	[너/NP+나/JX]
	그것이나 가져라.	[그 것 / N P + 이 나 / J X]
■■■ 나마/이나마	네 덕에 늦게나마 일을 마쳤다.	[늦/VA+게/EC+나마/JX]
	빵이나마 먹어라.	[빵/NNG+이나마/JX]
■■■ ㄴ/은/는	난 학생이다.	[나/NP+ㄴ/JX]
	오늘은 금요일이다.	[오늘/NNG+은/JX]
	이 종이는 어제 사 온 것이다.	[종이/NNG+는/JX]
■■■ ㄴ커녕/은커녕/는커녕	빨린커녕 천천히도 못 걸겠다	[빨리/MAG+ㄴ 커녕/JX]
	돈은커녕 먹을 쌀도 없다.	[돈/NNG+은커녕/JX]
	돕기는커녕 방해할 생각만 했다.	[돕/VV+기/ETN+는커녕/JX]
■■■ 다	물건을 거기다 놓아라.	[거기/NP+다/JX]
	그 물건을 거기에다 놓아라.	[거기/NP+에/JKB+다/JX]
■■■ 다가	책상을 어디다가 둘까요?	[어디/NP+다가/JX]
	집에다가 놓아 두어라.	[집/NNG+에/JKB+다가/JX]
■■■ 대로	철수는 철수대로 고민이 있다.	[철수/NNP+대로/JX]
■■■ 따라	오늘따라 버스도 안 온다.	[오늘/NNG+따라/JX]
■■■ 도/두	강아지도 주인은 알아본다.	[강아지/NNG+도/JX]
■■■ 란/이란	코알라란 호주에 사는 초식동물이다.	[코알라/NNG+란/JX]
	사람이란 분수를 지킬 줄 알아야 한다.	[사람/NNG+이란/JX]
■■■ ㄹ랑/일랑	강엘랑 가지 마라.	[강/NNG+에/JKB+ㄹ 랑]

	그 일에 대해선 걱정일랑 하지 말아라.	/JX] [걱정/NNG+일랑/JX]
■■■ 마다	꽃마다 독특한 향기가 있다.	[꽃/NNG+마다/JX]
■■■ 마저	장미마저 시들고 말았다.	[장미/NNG+마저/JX]
■■■ 만	사람은 뽕만으로 살 수 없다.	[뽕/NNG+만/JX+으로/JKB]
■■■ 밖에	이제는 떠날 수밖에 없다.	[수/NNB+밖에/JX]
■■■ 부터	우선 노약자부터 태워야 한다.	[노약자/NNG+부터/JX]
■■■ 뿐	가진 것은 집 한 채뿐이다.	[채/NNB+뿐/JX+이/VCP+다/EF]
■■■ 서꺼	국물이나 동치미서꺼 아무 거나	[동치미/NNG+서꺼/JX]
■■■ 사/이사	내사 그걸 이미 했지.	[내/NP+사/JX]
	남이사 무슨 상관이야.	[남/NNG+이사/JX]
■■■ 야/이야	그야 그렇지.	[그/NP+야/JX]
	그가 인간성이야 그만이지.	[인간성/NNG+이야/JX]
■■■ 야말로/이야말로	사과야말로 가을의 과일이다.	[사과/NNG+야말로/JX]
	통일이야말로 최대의 과업이지.	[통일/NNG+이야말로/JX]
■■■ 요	나는 그림을요 잘 그립니다.	[그림/NNG+을/JKO+요/JX]
■■■ 초차	이젠 봄초차 빼앗기는구나.	[봄/NNG+초차/JX]
■■■ 치고	값싼 물건치고 쓸 만한 게 없지.	[물건/NNG+치고/JX]

(1) 보조사 분석 기준

- 보조사는 ‘이다’의 활용어미와 구분하기 어려운 경우가 있다. 흔히 보조사로 간주되던 몇몇 형태들은 연결어미와 의미상의 차이가 없으며, 분포상으로도 구별되지 않기 때문에 이런 대상들은 보조사로 분석하지 않는다.

[기준 1] 대상 형태가 용언의 어미로 사용되는가.

[기준 2] 대상 형태가 체언에 후행할 때 서술어의 자격을 가지고 사용되는가.

(가) [기준 1, 2]에 부합하는 다음의 형태들은 모두 ‘연결어미’로 분석한다.

■■■ (이)ㄴ들, (이)ㄴ즉, (이)든, (이)든지, (이)라도, (이)라서, (이)라야

(나) [기준 1, 2]에 부합하지 않는 다음의 형태들은 ‘보조사’가 된다.

■ (이)나마, (이)야, (이)랑, (이)야말로, (이)란

(다) [기준 1]에 부합하지 않으나, [기준 2]에는 부합하는 형태는 ‘중의성’을 가진다.

■ (이)나, (이)요

(라) 다음의 형태는 서술격조사 ‘이다’의 활용형과는 관계가 없으므로 모두 보조사가 된다.³²⁾

■ 까지, 깨나, 는(은/ㄴ), 대로, 도, 따라, 마다, 마저, 만, 밖에, 부터, 뿐, 조차, 치고, ㄴ 커녕

※ [참고] ‘만’, ‘뿐’은 의존 명사로도 분석될 수 있음.

(마) [보완] 종결어미 뒤에 나타나는 ‘든지, 든가, 거나’ 등의 경우는 보조사로 분석한다.

■ 공부를 잘한다든지 운동을 잘한다든지 [잘/MAG+하/XSV+ㄴ 다/EF+든지/JX]

■ 시기라든가 질투라든가 하는 데에까지 [시기/NNG+이/VCP+라/EF+든가/JX]

■ 그녀는 예쁘다거나 귀엽다거나 하는 [예쁘/VA+다/EF+거나/JX]

<주의사항>

(가) 다음의 형태들은 분석 결과에 중의성이 생기므로, 이들을 분석할 때는 특히 주의해야 한다.

■ (이)란	코알라란 동물은 호주에 주로 서식한다.	[코알라/NNG+이/VCP+란/ETM]
■ (이)나	코알라란 매우 귀여운 동물이다. 밥이나 빵을 먹도록 해라. 그가 비록 열심히 하나 능력은 부족하다.	[코알라/NNG+란/JX] [밥/NNG+이나/JC] [하/VV+나/EC]
■ (이)야	어제 내가 술을 마셨나? 철수야 그 일을 할 수 있지.	[마시/VV+었/EP+나/EF+?/SF] [철수/NNP+야/JX]

32) [삭제] 말고,

→ ‘말고’는 ‘표준국어대사전’에 보조사로 등재되어 있지 않으며, 세종 말뭉치에서도 보조사로 분석하지 않았으므로 목록에서 삭제함.

	내가 좋아하는 것은 철수야.	[철 수 / NNP + 이 / VCP + 야 /EF+./SF]
	철수야! 부르는 소리	[철수/NNP+야/JKV]
■■ (이)요	밥을 먹다가요	[먹/VV+다가/EC+요/JX]
	밥이요, 빵이요.	[밥/NNG+이/VCP+요/EC]

(나) ‘종결어미+요(보조사)’는 종결어미로 통합하여 분석한다.

■■ 말씀대로 했는걸요. [하/VV+았/EP+는걸요/EF+./SF]

(다) ‘비종결어미+요(보조사)’는 통합하지 않고 각각 분석해 준다.

■■ 제가 몸이 좀 아파서요 지각을 했어요. [아프/VA+아서/EC+요/JX]

■■ 내가요, 왜요? [내/NP+가/JKS+요/JX]

[왜/MAG+요/JX+?/SF]

(라) [보완] 보조사 ‘요’의 분석

(1) A: 선생님이 집에 오셨어요.

B: 선생님이요? [선생/NNG+님/XSN+이/JKS+요/JX]

A: 커서 선생님이 되는 게 어떠니?

B: 선생님이요? [선생/NNG+님/XSN+이/JKC+요/JX]

(2) A: 선생님에 대해 알고 있니?

B: 선생님이요? (‘오’의 오류) [선생/NNG+님/XSN+이/VCP+요/EF]

(마) ‘말고’는 용언 ‘말다’의 활용형으로 처리한다.

■■ 돈말고 지혜가 필요하다. [돈/NNG+말/VV+고/EC]

바. 의존형태

1) 어미³³⁾

33) 지침에 제시된 어미 목록에서 빠진 이형태와 예시를 추가함

가) 선어말어미(EP)

- 용언이 활용할 때, 어간과 어말 어미 사이에 나타나는 것으로 높임법이나 시제, 양태를 나타내는 문법적인 요소이다. 선어말어미의 목록은 연구자에 따라 다를 수 있으나 이 분석에서는 아래의 것만을 선어말어미로 인정한다.

■■ -겠-	그 일은 내일 처리하겠다.	[처리/NNG+하/XSV+겠/EP+다/EF]
■■ -(으)시-	선생님께서 손수 만드신 삼촌은 형님이 있으시다.	[만들/VV+시/EP+ㄴ/ETM] [있/VV+으시/EP+다/EF]
■■ -오/으오/ 옵/으옵-	어머님께 선물을 받치오니 책을 읽으오니 어머님께 선물을 받치옵고 책을 읽으옵고	[받치/VV+오/EP+니/EC] [읽/VV+으오/EP+니/EC] [받치/VV+옵/EP+고/EF] [읽/VV+으옵/EP+고/EC]
■■ -았/었-	그는 집에 갔다. 우리가 먹었던 음식이 잘못됐다.	[가/VV+았/EP+다/EF+./SF] [먹/VV+었/EP+던/ETM]
■■ -았었/었었-	거기는 전에 갔었던 곳이다. 우리가 먹었던 음식에 문제가 있 다.	[가/VV+았었/EP+던/ETM] [먹/VV+었었/EP+던/ETM]

<주의사항>

- ① 선어말어미가 한 음절로 통합된 경우에는 각각 분리해서 분석한다.

■■ -셨- 그 일은 어머니께서 하셨다. [하/VV+시/EP+었/EP+다/EF+./SF]

- ② 다음의 선어말어미는 그 어간이 생략되었을 경우에 어간을 복원해 준다.

■■ -겠- 이것은 그대로 두어야겠다. [두/VV+어야/EC+하/VX+겠/EP+다/EF+./SF]

■■ -았/었- 철수가 그것을 가져오래다. [가져오/VV+라/EF+하/VV+았/EP+다/EF+./SF]

■■ -(으)시- 선생님께서 가자시오. [가/VV+자/EF+하/VV+시/EP+오/EF+./SF]

- ③ 위의 선어말어미가 포함되지 않은 어미 형태는 그대로 연결어미로 분석한다.

■■ -랄까-, -대야-, -래야-

④ [보완] ‘-여’나 ‘-였-’은 ‘-아’나 ‘-았-’으로 수정한 후 분석한다.

■■ 공부를 하였다. [하/VV+았/EP+다/EF]

■■ 공부를 열심히 하여 시험을 잘 보았다. [하/VV+아/EC]

나) 종결 어미(EF)

- 용언의 어간이나 선어말 어미 뒤에 연결되어 용언의 형식을 완성시키는 어미로서 한 문장을 끝맺는 역할을 하는 어미이다.

■■ -거든	나는 이것이 좋거든!	[좋/VA+거든/EF+!/SF]
■■ -게	그만한 돈이 있으면 좋게.	[좋/VA+게/EF+./SF]
■■ -구나/는구나	넌 정말 멋지구냐!	[멋지/VA+구나/EF+?/SF]
	앞이 잘 안 보이는구나.	[보이/VV+는구나/EF+./SF]
■■ -구려/는구려	당신도 가시겠구려.	[가/VV+시/EP+겠/EP+구려/EF+./SF]
	잘도 먹는구려.	[먹/VV+는구려/EF+./SF]
■■ -구먼/는구먼	학교가 참 크구먼.	[크/VA+구먼/EF+./SF]
	공부를 잘하는구먼.	[잘/MAG+하/XSV+는구먼/EF+./SF]
	이것이 무엇인가?	[무엇/NP+이/VCP+ㄴ가/EF+?/SF]
■■ -ㄴ가/은가/는가	그것이 좋은가?	[좋/VA+은가/EF+?/SF]
	그가 집에 있는가?	[있/VV+는가/EF+?/SF]
	이제 시작인걸.	[시작/NNG+이/VCP+ㄴ걸/EF+./SF]
■■ -ㄴ걸/은걸/는걸	그 책은 벌써 다 읽은걸.	[읽/VV+은걸/EF+./SF]
	그는 벌써 갔는걸.	[가/VV+았/EP+는걸/EF+./SF]
■■ -나	자네 그리로 가나?	[가/VV+나/EF+?/SF]
■■ -냐/으냐/느냐	키가 얼마나 크냐?	[크/VA+냐/EF+./SF]
	물이 얼마나 깊으냐?	[깊/VA+으냐/EF+?/SF]
■■ -냐고/으냐고/느냐고	그것보다 이것이 낫느냐?	[낫/VA+느냐/EF+?/SF]
	그가 누구냐고?	[누구/NP+이/VCP+냐고/EF+?/SF]
	그렇게 싫어? 싫으냐고?	[싫/VA+으냐고/EF+?/SF]
■■ -네	너 뭐 해? 뭐 하느냐고?	[하/VV+느냐고/EF+?/SF]
	정말 큰일 났네!	[나/VV+았/EP+네/EF+!/SF]
■■ -니	그게 없니?	[없/VA+니/EF+?/SF]
■■ -다/ㄴ다/는다	그게 사실이다.	[사실/NNG+이/VCP+다/EF+./SF]
	이건 말도 안 된다.	[되/VV+ㄴ다/EF+./SF]
	아이가 글을 잘 읽는다.	[읽/VV+는다/EF+./SF]
■■ -다구/ㄴ다구/는다구	돈이 많다구?	[많/VA+다구/EF+?/SF]
	너도 간다구?	[가/VV+ㄴ다구/EF+?/SF]

■ ■ -다나/ㄴ 다나 /는다나	소설책을 읽는다구? 그도 가겠다나. 나를 잘 안다나. 건강한 여자를 찾는다나.	[읽/VV+는다구/EF+?/SF] [가/VV+겠/EP+다나/EF+./SF] [알/VV+ㄴ 다나/EF+./SF] [찾/VV+는다나/EF+./SF]
■ ■ -다네/ㄴ 다네 /는다네	일을 망쳤다네 우리네 짧은 인생도 간단네. 평소에도 한복을 잘 입는다네.	[망치/VV+었/EP+다네/EF+./SF] [가/VV+ㄴ 다네/EF+./SF] [입/VV+는다네/EF+./SF]
■ ■ -다니까/ㄴ 다니까 /는다니까	돈이 없다니까! 어머니가 오늘은 꼭 오신다니까. 내 말을 믿지를 앓는다니까.	[없/VA+다니까/EF+!/SF] [오/VV+시/EP+ㄴ 다니까/EF+!/SF] [앓/VX+는다니까/EF+./SF]
■ ■ -다니/ㄴ 다니 /는다니	서울이 이렇게 변화하다니. 이 긴 시를 어떻게 외운다니? 이 많은 책을 언제 읽는다니?	[변화/XR+하/XSV+다니/EF+?/SF] [외우/VV+ㄴ 다니/EF+?/SF] [읽/VV+는다니/EF+?/SF]
■ ■ -다면서/ㄴ 다면서 /는다면서	술은 싫다면서? 니가 축구를 잘한다면서? 달팽이도 먹는다면서?	[싫/VA+다면서/EF+?/SF] [잘/MAG+하/XSV+ㄴ 다면서/EF+?/SF] [먹/VV+는다면서/EF+?/SF]
■ ■ -다오/ㄴ 다오 /는다오	그가 가지고 있다오. 꽃은 이른 봄에 핀다오. 이 나무는 열매를 많이 맺는다오.	[있/VX+다오/EF+./SF] [피/VV+ㄴ 다오/EF+./SF] [맺/VV+는다오/EF+./SF]
■ ■ -단다/ㄴ 단다 /는단다	나도 슬프단다. 선생님께서 공부를 가르쳐 주신단다. 누에는 뽕잎을 먹는다단다.	[슬프/VA+단다/EF+./SF] [주/VX+시/EP+ㄴ 단다/EF+./SF] [먹/VV+는다단다/EF+./SF]
■ ■ -도다/는도다	꽃이 아름답도다. 짐이 조서를 내리는도다.	[아름답/VA+도다/EF+./SF] [내리/VV+는도다/EF+./SF]
■ ■ -ㄹ걸/을걸	모른다고 할걸. 생각만큼 쉽지 않을걸.	[하/VV+ㄹ 걸/EF+./SF] [앓/VX+을걸/EF+./SF]
■ ■ -ㄹ게/을게	그렇게 할게. 남은 밥은 내가 먹을게.	[하/VV+ㄹ 게/EF+./SF] [먹/VV+을게/EF+./SF]
■ ■ -ㄹ까/을까	이제 밥을 할까? 이 과자는 내가 먹을까?	[하/VV+ㄹ 까/EF+?/SF] [먹/VV+을까/EF+?/SF]
■ ■ -렴/으렴	맘대로 해 보렴. 이것 좀 먹으렴.	[보/VX+렴/EF+./SF] [먹/VV+으렴/EF+./SF]
■ ■ -려무나/으려무나	더 놀다 가려무나. 책이나 읽으려무나.	[가/VV+려무나/EF+./SF] [읽/VV+으려무나/EF+./SF]
■ ■ -라니까/으라니까	그 사람이 아니라니까. 가만히 있으라니까.	[아니/VCN+라니까/EF+./SF] [있/VV+으라니까/EF+./SF]
■ ■ -ㄹ세/음세	그날 꼭 음세. 곧 밥을 먹음세.	[오/VV+ㄹ 세/EF+./SF] [먹/VV+음세/EF+./SF]
■ ■ -ㄹ니까/습니까	이제야 옵니까? 그래도 되겠습니까?	[오/VV+ㄹ 니까/EF+?/SF] [되/VV+겠/EP+습니까/EF+?/SF]

■■ -ㅂ니다/습니다	이렇게 합니다. 정말 재미있습니다.	[하/VV+ㅂ니다/EF+./SF] [재미있/VA+습니다/EF+./SF]
■■ -ㅂ시다/읍시다	다시 만납시다. 여기 앉읍시다.	[만나/VV+ㅂ시다/EF+./SF] [앉/VV+읍시다/EF+./SF]
■■ -ㅂ시오/읍시오	서둘러 주십시오. 여기 앉읍시오.	[주/VX+시/EP+ㅂ시오/EF+./SF] [앉/VV+읍시오/EF+./SF]
■■ -ㅂ디까/습디까	신부가 예뻐디까? 보기에 좋습디까?	[예쁘/VA+ㅂ디까/EF+?/SF] [좋/VA+습디까/EF+?/SF]
■■ -ㅂ디다/습디다	참 좋은 곳입디다. 덕수궁에 사람이 많습디다	[곳/NNB+이/VCP+ㅂ디다/EF+./SF] [많/VA+습디다/EF+./SF]
■■ -세/으세	제대로 좀 하세. 이 책을 우리 함께 읽으세.	[하/VV+세/EF+./SF] [읽/VV+으세/EF+./SF]
■■ -아/어/여	함께 가. 밥 먹어! 같이 해.	[가/VV+아/EF+./SF] [먹/VV+어/EF+!/SF] [하/VV+아/EF+./SF]
■■ -야	그건 사실이 아니야.	[아니/VCN+야/EF]
■■ -아라/어라	웃기지 말아라. 천천히 먹어라.	[말/VX+아라/EF+./SF] [먹/VV+어라/EF+./SF]
■■ -오/으오/소	물이 깨끗하오. 나는 요즘 논어를 읽으오.	[깨끗/XR+하/XSA+오/EF+./SF] [읽/VV+으오/EF+./SF]
■■ -자	그 곳에는 내가 가겠소.	[가/VV+겠/EP+소/EF+./SF]
■■ -자구나	잠이나 자자.	[자/VV+자/EF+./SF]
■■ -자니까	약속을 좀 미루자구나.	[미루/VV+자구나/EF+./SF]
■■ -자니까	그만 따지자니까.	[따지/VV+자니까/EF+./SF]
■■ -지	그가 언제 오지?	[오/VV+지/EF+?/SF]

<주의사항>

(가) ‘종결어미+요’는 통합해서 종결어미로 분석한다.

- 말씀대로 했는걸요. [하/VV+았/EP+는걸요/EF+./SF]
 ■■ 뭐 먹었는데요? [먹/VV+았/EP+는데요/EF+?/SF]

※ [참고] ‘비종결어미+요’는 통합해서 분석하지 않는다.

- 그 애는 노래는 잘 부르는데요. [부르/VV+는데/EC+요/JX+./SF]
 춤은 잘 못 춰요.
 ■■ 어제 비가 많이 와서요. 지각을 했어요. [오/VV+아서/EC+요/JX+./SF]

(나) ‘-세요’는 다음과 같이 선어말어미까지 분석한다.

■■ 어서 출근하세요. [출근/NNG+하/XSV+시/EP+어요/EF+./SF]

(다) ‘-죠’는 축약형을 그대로 분석한다.

■■ 어서 출근하죠. [출근/NNG+하/XSV+죠/EF+./SF]

※ [참고] 다음의 경우도 <표준>을 따라 종결어미로 분석한다.

■■ 아픈데 밥을 <u>먹을까</u> 싶다.	[먹/VV+을까/EF]
■■ 진짜 부자유친이 <u>아닐까</u> 생각합니다.	[아니/VCN+ㄹ까/EF]
■■ 돈을 어떻게 <u>쓰느냐</u> 에 따라	[쓰/VV+느냐/EF+에/JKB]
■■ 무슨 일이 <u>있었는가</u> 했다.	[있/VV+었/EP+는가/EF]
■■ 보통 사용할 때는 <u>뭔가</u> 게임을 할 때	[뭐/NP+이/VCP+ㄴ가/EF]
■■ 언제 고향에 <u>갈지</u> 잘 모르겠습니다.	[가/VV+ㄹ지/EF]
■■ 왜 <u>좋았는지</u> 알아요.	[좋/VA+았/EP+ㄴ지/EF]
■■ 자기 적성에 <u>맞는지</u> 안 <u>맞는지</u> 고려하지 않습니다.	[맞/VV+는지/EF 안/MAG 맞/VV+는지/EF]

다) 연결 어미(EC)

- 용언의 어간이나 선어말 어미 뒤에 연결되어 용언의 형식을 완성시키는 어미로서 문장을 종결시키지 못하고 뒤에 오는 절을 연결시켜 주는 어미를 말한다.

■■ -거나	누가 오거나 알은 체 할 것 없다.	[오/VV+거나/EC]
■■ -거니	비가 오겠거니 생각했다.	[오/VV+겠/EP+거니/EC]
■■ -거늘	이미 늦었거늘 어찌 빨리 가는가?	[늦/VV+었/EP+거늘/EC]
■■ -거든	가거든 말해라.	[가/VV+거든/EC]
■■ -건대	내가 보건대, 네 말이 옳다.	[보/VV+건대/EC]
■■ -건마는	말렸건마는 아직도 축축하다.	[말리/VV+었/EP+건마는/EC]
■■ -게	개를 굶게 하지 마라.	[굶/VV+게/EC]
■■ -고	일단 먹고 보자.	[먹/VV+고/EC]
■■ -고	일을 하고 밥을 먹자.	[하/VV+고/EC]
■■ -곤	종종 지각하곤 했다.	[지각/NNG+하/XSV+곤/EC]
■■ -고자	병을 낫고자 몸부림쳤다.	[낫/VV+고자/EC]
■■ -기에	실수했기에 용서해 주었다.	[실수/NNG+하/XSV+았/EP+기에/EC]
■■ -ㄴ데/은데/	예쁜데 미워한다.	[예쁘/VA+ㄴ데/EC]

	는데	방이 좁은데 가구는 많다.	[좁/VA+은데/EC]
		눈이 오는데 차를 가져가지 말까?	[오/VV+는데/EC]
■■	-ㄴ 들/는들	간다 한들 아주 같까?	[하/VV+ㄴ 들/EC]
		그걸 먹는들 뭐가 달라지겠나.	[먹/VV+는들/EC]
■■	-ㄴ 족/은족	배가 고프족 속이 쓰리다.	[고프/VA+ㄴ 족/EC]
		물이 맑은족 고기가 많기는 어렵소.	[맑/VA+은족/EC]
■■	-ㄴ 지라/은지라/ 는지라	눈이 온지라 길이 미끄럽다.	[오/VV+ㄴ 지라/EC]
		기분이 좋은지라 다정하다.	[좋/VA+은지라/EC]
		선생님께서 고집을 굽히지 않으시는지라	[않/VX+으시/EP+는지라/EC]
■■	-나/으나	눈이 오나 비가 오나	[오/VV+나/EC]
		밥을 먹으나 마나이다.	[먹/VV+으나/EC]
■■	-나니	멀리 보이나니 넓은 들이로다.	[보이/VV+나니/EC]
■■	-나마/으나마	도와주지는 못하나마 방해를 해서는	[못하/VX+나마/EC]
		맞은 없으나마 많이 드세요.	[없/VA+으나마/EC]
■■	-노니	문노니, 포부가 무엇이냐?	[문/VV+노니/EC+/,SP]
■■	-니/으니	밥을 다 먹고 보니 배가 불렀다.	[보/VX+니/EC]
		이 옷은 작으니 큰 것으로 바꿔 주세요.	[작/VA+으니/EC]
■■	-느니	앉아서 걱정하느니 나가서 하겠다.	[걱정/NNG+하/XSV+느니/EC]
■■	-니까/으니까	웃기니까 좋다.	[웃기/VV+니까/EC]
		약속을 했으니까 만나야 한다.	[하/VV+았/EP+으니까/EC]
■■	-다가	자랑하다가 망신당했다.	[자랑/NNG+하/XSV+다가/EC]
■■	-다기에/ㄴ 다기에 /는다기에	그녀가 예쁘다기에 보러 왔소.	[예쁘/VA+다기에/EC]
		앞으로 잘 한다기에 승낙했다.	[하/VV+ㄴ 다기에/EC]
		빵을 먹는다기에 주었다.	[먹/VV+는다기에/EC]
■■	-다손/ㄴ 다손 /는다손	밑다손 치더라도 구박하지 말자.	[밑/VA+다손/EC]
		그가 제시간에 온다손 하더라도	[오/VV+ㄴ 다손/EC]
		내 앞의 음식은 다 먹는다손 치더라도	[먹/VV+는다손/EC]
■■	-대도/ㄴ 대도 /는대도	시간이 있대도 만나 주질 않는다.	[있/VV+대도/EC]
		늦으면 큰일 난대도 서두르질 않아요.	[나/VV+ㄴ 대도/EC]
		떠들면 야단맞는대도 계속 떠들었다.	[야단맞/VV+는대도/EC]
■■	-더라도	가더라도 꼭 돌아와라.	[가/VV+더라도/EC]
■■	-던들	진작 알았던들 방법을 취했지.	[알/VV+았/EP+던들/EC]
■■	-도록	미치도록 일했다.	[미치/VV+도록/EC]
■■	-든지	외모가 어떠하든지 무슨 상관인가?	[어떠/XR+하/XSA+든지/EC]
■■	-되	싸우되 꼭 지도록 해라.	[싸우/VV+되/EC]
■■	-ㄹ 뿐더러/ 을뿐더러	비가 올뿐더러 바람도 분다.	[오/VV+ㄹ 뿐더러/EC]
		그는 재산이 많을뿐더러 재능도 많다	[많/VA+을뿐더러/EC]
■■	-ㄹ 수록/ 갈수록	태산이다.	[가/VV+ㄹ 수록/EC]

을수록	이 책은 읽을수록 감동을 준다.	[읽/VV+을수록/EC]
■ ■ -ㄹ지/을지	비가 얼마나 올지 천둥이 다 친다.	[오/VV+ㄹ지/EC]
■ ■ -ㄹ지라도/을지라도	내일은 얼마나 날씨가 좋을지 오늘 밤하늘에 별이 유난히 빛난다.	[좋/VA+을지/EC]
■ ■ -ㄹ지언정/을지언정	이길지라도 명예롭지는 않다.	[이기/VV+ㄹ지라도/EC]
■ ■ -라고	마음에 걱정이 있을지라도 내색하지 마라.	[있/VV+을지라도/EC]
■ ■ -라	그것은 무모한 행동일지언정	[행동/NNG+이/VCP+ㄹ지언정/EC]
■ ■ -라	죽을지언정 그 일은 못하겠다.	[죽/VV+을지언정/EC]
■ ■ -라고	바보라고 생각한다.	[바보/NNG+이/VCP+라고/EC]
■ ■ -라	오르락 내리락	[오르/VV+라/EC]
■ ■ -라	그는 반장이랍시고 행패만 부린다.	[반장/NNG+이/VCP+라/EC]
■ ■ -라/으러	청소하러 가자.	[청소/NNG+하/XSV+러/EC]
■ ■ -러/으러	점심 먹으러 집에 간다.	[먹/VV+으러/EC]
■ ■ -러/으러	학교에 가러 한다.	[가/VV+러/EC]
■ ■ -러니와/으러니와	웃으러 한다.	[웃/VV+으러/EC]
■ ■ -려니와/으려니와	비용도 문제려니와 일꾼도 문제다.	[문제/NNG+이/VCP+려니와/EC]
■ ■ -려니와/으려니와	이 마을은 경치도 좋으려니와	[좋/VA+으려니와/EC]
■ ■ -려니와/으려니와	보면 반가우려니와 볼 수가 없네.	[반갑/VA+려니와/EC]
■ ■ -려니와/으려니와	벌써 제 잘못을 알았으려니와	[알/VV+았/EP+으려니와/EC]
■ ■ -며/으며	노래하며 춤을 춘다.	[노래/NNG+하/XSV+며/EC]
■ ■ -며/으며	강물이 맑으며 깊다.	[맑/VA+으며/EC]
■ ■ -면/으면	지옥이 존재하면 만원일 것이다.	[존재/NNG+하/XSV+면/EC]
■ ■ -면/으면	내일 날씨가 좋으면 소풍을 가겠다.	[좋/VA+으면/EC]
■ ■ -면서/으면서	푸르면서 검은 물빛	[푸르/VA+면서/EC]
■ ■ -면서/으면서	밥을 먹으면서 신문을 본다.	[먹/VV+으면서/EC]
■ ■ -므로/으므로	비가 오므로 가지 않겠다.	[오/VV+므로/EC]
■ ■ -므로/으므로	강이 깊으므로 배 없이 건널 수 없다.	[깊/VA+으므로/EC]
■ ■ -아/어	입을 막아 버렸다.	[막/VV+아/EC]
■ ■ -아/어	밥을 먹어 버렸다.	[먹/VV+어/EC]
■ ■ -아도/어도	암만 봐도 모르겠다.	[보/VV+아도/EC]
■ ■ -아도/어도	나는 부자가 아니어도 행복하다.	[아니/VCN+어도/EC]
■ ■ -아서/어서	덧을 놓아서 꿩을 잡았다.	[놓/VV+아서/EC]
■ ■ -아서/어서	그는 걸어서 학교에 갔다.	[걸/VV+어서/EC]
■ ■ -아야/어야	이 일은 잘해야 한다.	[잘/MAG+하/XSV+아야/EC]
■ ■ -아야/어야	사람은 먹어야 산다.	[먹/VV+어야/EC]
■ ■ -자마자	오자마자 당했다.	[오/VV+자마자/EC]
■ ■ -지	우기지 못해 버렸다.	[우기/VV+지/EC]
■ ■ -지마는	비가 오지마는 가야 한다.	[오/VV+지마는/EC]

<주의사항>

(가) 어미에 따라서는 분석의 중의성이 생길 수 있으므로 문맥 확인을 통해 형태분석을 결정한다.

■■ 너는 내가 왔는데 기쁘지도 않니?	[오/VV+았/EP+는데/EC]
■■ 내가 지금 있는 데가 어디지?	[있/VV+는/ETMⅡ데/NNB+가/JKS]
■■ 다들 만족하는지 아무런 불평이 없다.	[만족/NGG+하/XSV+는지/EC]
■■ 너를 만난 지도 꽤 오래구나.	[만나/VV+ㄴ/ETMⅡ지/NNB+도/JX]

(나) '-음직'은 “음직/EC”로 분석한다. 그러나 ‘바람직, 먹음직’ 등은 그 자체가 하나의 어근이므로 더 이상 분석할 수 없다는 것에 유의한다.

■■ 어른답고 믿음직하게 행동해라.	[믿/VV+음직/EC+하/VX+게/EC]
■■ 그것 참 먹음직스럽다.	[먹음직/XR+스럽/XSA+다/EF+./SF]
■■ 그것은 매우 바람직한 일이다.	[바람직/XR+하/XSA+ㄴ/ETM]

라) 명사형 전성 어미(ETN)

- 한 문장의 성격을 임시로 바꾸어 다른 문장 속에서 명사적인 역할을 하게 하는 어미를 말한다.

■■ -기	그 일은 정말 중요하기 때문이다.	[중요/NGG+하/XSA+기/ETN]
■■ -ㅁ/-음	학생 신분임을 밝히다.	[신분/NGG+이/VCP+ㅁ/ETN]
	장사는 신용을 얻음이 제일이다.	[얻/VV+음/ETN+이/JKS]

<주의사항>

(가) 불규칙 용언 어간에 명사형 전성 어미가 붙어 있을 경우 ‘-음’이 아닌 ‘-ㅁ’으로 분석한다.

■■ 김철수 지음	[짓/VV+ㅁ/ETN]
-----------	--------------

(나) “음, 기”가 붙은 말이 단순히 명사형이냐 아니면 굳어진 명사이냐 하는 것

은 물론 문맥에 따라 결정되어야 하지만 먼저 그것이 “사전”에 등재되어 있느냐의 여부를 살펴보아야 한다.

- ■ 책을 읽기가 어렵다. [읽/VV+기/ETN+가/JKS]
- ■ 읽기 교육이 문제가 된다. [읽기/NNG]

마) 관형사형 전성 어미(ETM)

- 용언의 성격을 임시로 바꾸어 다른 문장 속에서 관형사적인 역할을 하게 하는 어미이다.

■ ■ -ㄴ/은	어제 떠난 사람	[떠나/VV+ㄴ/ETM]
	어제 먹은 빵에 이상이 있었다.	[먹/VV+은/ETM]
■ ■ -는	잃어버린 물건을 찾는 일은 어렵다.	[찾/VV+는/ETM]
■ ■ -던	이제까지 미루던 일을 오늘 해치웠다.	[미루/VV+던/ETM]
■ ■ -ㄹ/을	나에게는 아직 처리할 일이 있다.	[처리/NNG+하/XSV+ㄹ/ETM]
	물이 깊을 것이다.	[깊/VA+을/ETM]
■ ■ -런	어제런 듯하다.	[어제/NNG+이/VCP+런/ETM]

<주의사항>

- (가) 불규칙 용언 어간에 관형사형 전성 어미가 있을 경우 ‘-은, -을’이 아닌 ‘-ㄴ, -ㄹ’로 분석한다.

- ■ 그녀의 고운 얼굴 [곱/VA+ㄴ/ETM]
- ■ 그녀는 매우 아름다울 것이다. [아름답/VA+ㄹ/ETM]

- (나) 종결 어미에 이어서 전성 어미가 올 경우 통합해서 전성어미로 처리한다.

- ■ 어느 쪽에 더 비중을 두느냐는 것이 [두/VV+느냐는/ETM]

2) 접두사(XP)

- 접두사는 명사와 수사에 결합하는 접사류를 묶어서 체언접두사만을 설정하기로 한다.

가) 체언 접두사(XPN)

- 명사 접두사에는 한자어계 접두사와 고유어계 접두사가 있는데, 그 목록의 풍부함에 비해 대개가 생산성이 그리 높지 않다. 일단 여기서는 비교적 생산성이 높다고 인정되는 접두사와, 접두사를 분리했을 경우 단일한 표제어로 등재될 수 있는 경우에 한해서 접두사 분석을 하기로 한다.

가(假)-가건물, 고(高)-고물가, 과(過)-과보호, 구(舊)-구소련, 날-날음식, 노(老)-노부부, 대(大)-대선배, 만-만아들, 맨-맨몸, 무(無)-무의식, 미(未)-미완성, 반(反)-반독재, 범(汎)-범세계, 부(不)-부도덕, 불(不)-불합리, 비(非)-비논리, 생(生)-생김치, 소(小)-소강당, 신(新)-신정당, 왕(王)-왕족발, 재(再)-재충전, 저(低)-저임금, 제(第)-제13차, 준(準)-준전시, 초(超)-초만원, 최(最)-최고급, 친(親)-친러시아, 탈(脫)-탈냉전시대, 폐(廢)-폐광산, 풋-풋살구, 피(被)-피고소인, 한-한가운데, 헛-헛고생

- ※ [보완] 단, 예외적으로 ‘대부분, 대다수, 무조건’의 경우는 체언 접두사를 분리하지 않는다.

3) 접미사(XS)

- 파생 접미사에는 어기의 품사를 바꾸는 것과 그렇지 않은 것이 있는데, 이 둘을 별도로 구별하여 표지를 부여하지는 않는다.

가) 명사파생접미사(XSN)

- 명사파생접미사는 명사나 다른 어근에 후행하여 그것이 명사의 기능을 수행할 수 있도록 만들어 주는 의존 형태이다. 그러나 명사파생접미사는 연구자에 따라 그 목록이 다르며, 실제로도 구분이 애매한 경우가 많다. 본 분석에서는 접미사의 생산성과 접미사를 제외한 형태의 독립성을 기준으로 다음과 같이 목록을 마련하였다.

가(價)-매매가, 가(哥)-김가, 경(頃)-두 시경, 계(系)-몽고계, 계(界)-교육계, 광(狂)-메모광, 권(圈)-운동권, 권(權)-참정권, 당(當)-한 사람당, 대(臺)-억대, 택(宅)-청주택, 론(論)-비평론, 별(別)-가구별, 여(餘)-삼십여, 류(類)-자연류, 률(率)-경쟁률, 리(裡)-비밀리, 분(分) 분량-일인분, 분(分)-3분의, 산(産)-중국산, 상(上)-역사상, 생1(生)갑자생, 생2(生)견습생, 성(性)-인간성, 시(視)-영웅시, 용(用)-전쟁용, 적(的)-사상적, 형(型)-기본형, 형(形)-도시형, 제(制)-봉건제, 층(層)-선수층, 치(值)-보름치, 풍(風)-북고풍, 화(化)-도구화, 기-기름기, 껌-10분껌, 꿀-십 원꿀, 끼리-전우끼리, 쿤-노름쿤, 네-동이네, 님-선생님, 들-우리들, 들이-1ㄹ들이, 배기-열 살배기, 뺄-조카뺄, 씩-만원씩, 장이-간판장이, 쟁이-심술쟁이, 쯤 -내일쯤, 질-서방질, 짜리-백 원짜리, 째1 -이틀째, 째2-옹기째, 치레-인사치레, 투성이-먼지투성이

<주의사항>

(가) 명사파생접미사인 ‘-들’은 그 분포가 매우 다양하여 일부에서는 이를 보조사와 접미사로 나누어 분석하기도 한다. 그러나, 본 분석에서는 이들을 모두 명사파생접미사로 처리한다. ‘먹고들’의 ‘-들’도 선행성분이 어미이긴 하나, 일치하는 대상은 선행하는 명사로 해석할 수도 있기 때문이다.

- ■ 사람들이 우리 집에 왔다. [사람/NNG+들/XSN]
 ■ ■ 그들은 밥을 먹고들 싶었다. [먹/VV+고/EC+들/XSN]

(나) ‘-님’은 다음과 같이 세 가지의 분석 중의성을 가지므로 주의해서 분석한다.

- ① ‘임’의 의미로 쓰인 경우: 보통명사
 ■ ■ 님과 이별하다. [님/NNG+과/JKB]
 ② 사람의 ‘이름’이나 ‘성’ 뒤에서 쓰인 경우: 의존명사
 ■ ■ 김철수 님께서 오셨습니다. [김철수/NNP || 님/NNB+께서/JKS]
 ③ 그 밖의 경우: 명사파생접미사
 ■ ■ 과장님이 부르십니다. [과장/NNG+님/XSN+이/JKS]

(다) 목록에 있는 접미사라도 사전에 등재되지 않은 명사나 어근과 함께 사용됐

다면 전체를 명사로 분석한다.

■■ 획기적

[획기적/NNG]

나) 동사파생접미사(XSV) → ‘명사/부사/어근+동사파생접미사’로 분석한다.

- 동사파생접미사는 어기 또는 어근에 붙어서 그것을 동사로 만들어 주는 기능을 갖는 접미사이다.

※ [보완] 여기서는 그러한 접미사 중 생산성이 높은 아래의 넷만 동사파생접미사로 인정하여 분석한다.

■■ 당하	아군이 공격당하는 데에는 이유가 있다.	[공격/NNG+당하/XSV+는/ETM]
■■ 되	아침식사가 이미 준비되어 있었다.	[준비/NNG+되/XSV+어/EC]
■■ 시키	강아지를 운동시키려고 공원에 나갔다.	[운동/NNG+시키/XSV+려고/EC]
■■ 하	외국에서 공부하는 일이 쉬운 것은 아니다.	[공부/NNG+하/XSV+는/ETM]

<주의사항>

(가) [보완] ‘-하’ 접사는 생산성이 높기 때문에 모든 ‘N하다’가 표제어로 등재되어 있지 않다. ‘N 하다’와 같이 구로 보는 것은 의미적으로 명사를 수식하는 요소가 선행하는 것이 명확한 경우로만 한정하고 그 이외의 경우는 구로 보지 않고 ‘-하’를 접사로 처리한다.

■■ 외국에서 공부하는 것은 힘들다. [공부/NNG+하/XSV+는/ETM]
■■ 외국에서 공부 하는 것은 힘들다. [공부/NNG+하/XSV+는/ETM]

■■ 카페에서 시험 공부 하는 것을 [공부/NNG || 하/VV+는/ETM]
■■ 카페에서 시험 공부하는 것을 [공부/NNG || 하/VV+는/ETM]

(나) [보완] 학습자가 잘못 접미사를 사용한 경우 교정어절을 상정했을 때 교정어절의 품사가 동사일 때는 동사파생접미사, 교정어절의 품사가 형용사일 때는 형용사파생접미사로 분석한다.

■■ 음식을 먹한다. [먹/VV+하/XSV+다/EF+./SF]

■■■ 마음이 <u>아프</u> 한 아주머니가 집에 돌아왔다.	[아프/VA+하/XSA+ㄴ/ETM]
■■■ 그렇지 <u>않</u> 한다면	[않/VX+하/XSA+ㄴ다면/EC]
■■■ 한국어 공부를 <u>열심</u> 하다.	[열심/NNG+하/XSV+다/EF]

(다) [보완] ‘NA+하다’의 접사 ‘-하’는 XSV로 분석한다.

다) 형용사파생접미사(XSA) → ‘명사/부사/어근+형용사파생접미사’로 분석한다.

- 형용사파생접미사는 어기나 어근에 붙어서 그것을 형용사로 파생시키는 접미사이다.

※ [보완] 여기서는 그러한 접미사 중 생산성이 높은 아래의 다섯만 형용사파생접미사로 인정하여 분석한다.

■■■ 답	사람이 사람답게 행동해야 사람이지	[사람/NNG+답/XSA+게/EC]
■■■ 되	자식된 도리로 어떻게 그런 짓을..	[자식/NNG+되/XSA+ㄴ/ETM]
■■■ 롭	어려운 일일수록 슬기롭게 대처하라.	[슬기/NNG+롭/XSA+게/EC]
■■■ 스럽	그녀의 사랑스러운 표정을 보거라.	[사랑/NNG+스럽/XSA+ㄴ/ETM]
■■■ 하	멍청한 표정을 짓지 말아라.	[멍청/XR+하/XSA+ㄴ/ETM]

4) 어근(XR)

※ [보완] 표준국어대사전에 등재된 2음절 이상의 어근만 어근으로 인정하여 분석한다.

■■■ 따뜻도 하다	[따뜻/XR+도/JX] [하/VV+다/EF]
■■■ 이러하다	[이러/XR+하/XSA+다/EF]

<주의사항>(예를 보완함.)

- 어근의 분석 대상은 표준국어대사전의 표제어 중 2음절 이상의 어휘이다.
‘하다’가 결합한 어휘 중 ‘하다’에 선행하는 음절이 1음절일 경우에는 어근

분리 현상이 매우 제한적이므로 이 경우에는 통합형으로 분석한다.

■■■ 등직하다 [등직/XR+하/XSV+다/EF]

■■■ 취하다 [취하/VV+다/EF] '취'는 어근
(위하다, 반하다, 강하다, 약하다, 중하다, 대하다, 의하다, 통하다 등)

■■■ 밥하다 [밥/NNG+하/XSV+다/EF] '밥'은 명사
(절하다, 인하다, 비하다, 한하다 등)

■■■ 잘되다 [잘/MAG+되/XSV+다/EF] '잘'은 부사

※ [보완] “못하다”의 경우 ‘못하/VV, 못하/VA, 못하/VX’의 세 가지 경우가 존재하는데, 이때 ‘-하’를 분석할 경우 본용언의 분석과 보조용언의 분석이 동형이 되기 때문에 예외로 취급해서 ‘-하’를 분석하지 않는다.

■■■ 노래를 못한다. [못하/VV+ㄴ다/EF]

■■■ 음식 맛이 저번보다 못하다. [못하/VA+다/EF]

■■■ 밥을 먹지 못한다. [못하/VX+ㄴ다/EF]

※ 참고

■■■ 숙제를 못 했다. [못/MAG || 하/VV+았/EP+다/EF]

사. 기타

1) 기호

- 영문이나 한자, 기호 등이 어절 중간에 개입하여 올바른 분석이 불가능한 경우에는 각각의 요소를 분리하여 분석한다. 이 경우 표지를 줄 수 없는 불완전한 형태가 생길 수 있다.

■■■ 마이크로소프트(microsoft)사 [마이크로소프트/NNP+(/SS+microsoft/SL+)/SS+사/NNG]

■■■ 농 · 수산물 [농/NNG+ · /SP+수산물/NNG]

■■■ 초 · 중 · 고 [초/NNG+ · /SP+중/NNG+ · /SP+고/NNG]

■■■ 위, 아래 집 [위/NNG+/,SP+아랫집/NNG]

cf.

■■■ 대~박 [대/NA+~/SS+박/NA]

2) 준말

- 준말은, 그것이 본딤말과 대등하게 사용되고 분석결과가 동일한 어절 단위를 형성할 경우에 한해서만 복원한다. 그러나 다음에서처럼, 본딤말로 복원할 경우 어절 수에 변화가 생길 뿐 아니라 본딤말로 복원하는 정도가 일관성을 띠지 않게 되는 경우는 굳이 복원하지 않는다. 그러나, 이러한 원칙이 모든 경우에 일관적으로 적용될 수 있는 것은 아니다. 결국 준말의 처리는 해당 어절에 따라 임의적일 수 있다.

■■ 라는	[라는/ETM]	(○)
	[라고/JKQ+하/VV+는/ETM]	(×)
■■ 려는	[려는/ETM]	(○)
	[려고/EC+하/VX+는/ETM]	(×)

3) 분석불능범주(NA)

- ※ [보완] 그 자체가 사전에 등재되어 있지도 않으면서, 축약의 정도가 심하거나 분석하기 어려운 방언형의 경우 분석불능범주로 처리한다.

■■ 담배가 <u>쪼매튼게</u> 하마 자라서 빠나?	[쪼매튼게/NA]
■■ 친구한테 전화를 <u>적긴</u> 일이었다.	[적긴/NA]
■■ “부산국제영화” <u>제가니와</u>	[제가니와/NA]
■■ <u>있잖아</u> 요	[있/VV+잖/NA+아요/EF]
■■ ㅋㅋ	[ㅋㅋ/NA]
■■ ππ	[ππ/NA]
■■ ○ㅋ○ㅋ	[○ㅋ○ㅋ/NA]
■■ ^^	[^^/NA]

4) 합성어

- 합성어는 표준국어대사전에 등재되어 있는 것만을 인정한다.

		띄어쓰기 상태(학습자)	분석 방법
N+N 구성	사전 등재	1. 국어사전(‘-’로 등재)	국어사전/NNG
		2. 국어 사전(‘-’로 등재)	국어사전/NNG
		3. 국어 교육(‘^’로 등재)	국어/NNG 교육/NNG
		4. 국어교육(‘^’로 등재)	국어/NNG + 교육/NNG
	사전 미등재	1. 국어연구	국어/NNG 연구/NNG
		2. 국어 연구	국어/NNG 연구/NNG
본 용언 + 보조 용언 구성	사전 등재	1. 좋아하다	좋아하/VV+다/EF
		2. 좋아 하다	좋아하/VV+다/EF
	사전 미등재	1. 가보다	가/VV+아/EC+보/VX+다/EF
		2. 가 보다	가/VV+아/EC 보/VX+다/EF

[보완] <주의사항>

(가) 표제어가 사전의 표제어로 등록되어 있는 경우는 그대로 분석한다.

■■ 정치권력 (사전: 정치-권력) [정치권력/NNG]

(나) 합성어로 등재되어 있되 띄어쓰기를 허용한 합성어는 세분하여 분석하는 것을 원칙으로 한다.

■■ 학생운동 (사전표기: 학생^운동) [학생/NNG+운동/NNG]
[학생/NNG || 운동/NNG]

(다) 합성어로 등록되어 있지 않은 표제어는 분리해서 분석하되, 사전 표제어로 등록되어 있는 최대한 많은 음절수의 단어를 생성하도록 나눈다.

(라) 3음절 어휘와 같이 어느 쪽으로 나뉘어도 음절수가 같고, 양쪽 분석이 모두 사전 표제어라면 뒤쪽을 먼저 분석한다.

■■ 차창밖 [차/NNG+창밖/NNG]

■■ 이등품 [이/NR+등품/NNG]

5) [보완] 접사처럼 쓰이는 ‘명사’의 처리

- 일부 명사는 사전에 ‘(일부 명사 뒤/앞에 붙어)~의 뜻을 나타내는 말.’로 등재

되며, 이들은 앞뒤에 함께 쓰인 명사와 합쳐서 명사로 분석한다. 이들의 목록은 다음과 같다.

가01 「04」	(일부 명사 뒤에 붙어) ‘주변4’의 뜻을 나타내는 말.	강가 // 냇가 // 우물가.
감03 「02」	(옷을 뜻하는 명사 뒤에 붙어) ‘옷을 만드는 재료’의 뜻을 나타내는 말.	한복감//양복감.
감03 「04」	(일부 명사 뒤에 붙어) ‘자격을 갖춘 사람’의 뜻을 나타내는 말.	신랑감//머느릿감//사윗감//장군감.
감03 「05」	(일부 명사 뒤에 붙어) 대상이 되는 도구, 사물, 사람, 재료의 뜻을 나타내는 말.	구경감 // 놀림감 // 땀감 //양념감 // 안춧감 //장난감//웃음감//사형감 //노벨상감//바느질감.
값 「07」	(일부 명사 뒤에 붙어) ‘가격’, ‘대금’, ‘비용’의 뜻을 나타내는 말.	기름값 // 물값 // 물건값 //부식값 // 신문값 //우윳값 // 음식값.
값 「08」	(일부 명사 뒤에 붙어) ‘수치’의 뜻을 나타내는 말.	변수값//분석값//위상값//저항값
과04 「02」	(일부 명사 뒤에 붙어) 학과나 전문 분야를 나타내는 말.	국어과 // 마취과 //물리학과.
구15 「03」	(일부 명사 뒤에 붙어) ‘법령 집행을 위하여 정한 구획’의 뜻을 나타내는 말.	선거구 // 투표구.
구이01 「02」	(일부 명사 뒤에 붙어) 구운 음식의 뜻을 나타내는 말.	갈비구이//생선구이//참새구이.
군03 「02」	(일부 명사 뒤에 붙어) 왕자군을 뜻하는 말.	경녕군 // 복성군.
군05 「02」	(일부 명사 뒤에 붙어) ‘군대3’의 뜻을 나타내는 말.	시민군//예비군//유엔군//진압군.
극04 「02」	(일부 명사 뒤에 붙어) ‘연극’, ‘드라마’ 따위의 뜻을 나타내는 말.	고발극//사기극//실험극//특집극
금06 「04」	(일부 명사 앞에 붙어) ‘금색1’, ‘금제1’의 뜻을 나타내는 말.	금두꺼비 // 금목걸이 //금수저.
급04 「05」	(직급 따위를 나타내는 일부 명사 뒤에 붙어) ‘그 직급’의 뜻을 나타내는 말.	과장급 // 부장급 //간부급.
길01 「10」	(일부 명사 뒤에 붙어) ‘과정’, ‘도중’, ‘중간’의 뜻을 나타내는 말.	산책길//시장길
꽃01 「07」	(일부 명사 뒤에 붙어) ‘그 꽃’의 뜻을 나타내는 말.	도라지꽃//무궁화꽃//목련꽃//민들레꽃//사과꽃//유채꽃.
난05 「02」	(고유어와 외래어 명사 뒤에 붙어) ‘구분된 지면’의 뜻을 나타내는 말.	어린이난//가십난//컴퓨터난//해의 토픽난.
놀이01 「04」	(일부 명사 뒤에 붙어) ‘모방4’, ‘장난’, ‘흥내’의 뜻을 나타내는 말.	시장놀이//병원놀이//엄마놀이//학교놀이.
대15 「03」	(일부 명사 뒤에 붙어) 받침이 되는 시설이나 이용물의 뜻을 나타내는 말.	급수대 // 조희대 //독서대.
택01 「04」	(일부 명사 뒤에 붙어) ‘택호’를 나타내는 말.	윤 판서택
덩어리 「03」	(일부 명사 뒤에 붙어) [같은 말] 덩이(3. 그러한 성질을 가지거나 그런 일을	꿀칫덩어리 // 심술덩어리 //애꿎덩어리//제주덩어리.

	일으키는 사람이나 사물을 나타내는 말).	
덩이 「03」	(일부 명사 뒤에 붙어) 그러한 성질을 가지거나 그런 일을 일으키는 사람이나 사물을 나타내는 말. [비슷한 말] 덩어리.	꿀칫덩이 //심술덩이.
란01	(한자어 명사 뒤에 붙어) ‘알’의 뜻을 나타내는 말.	수정란//무정란.
란02	(한자어 명사 뒤에 붙어) ‘구분된 지면’의 뜻을 나타내는 말. ‘칸01’으로 순화.	광고란//독자란//투고란.
란03 「01」	(한자어 뒤에 붙어) ‘난초’의 뜻을 나타내는 말.	금자란//문주란//은란.
량05	(한자어 명사 뒤에 붙어) 분량이나 수량의 뜻을 나타내는 말.	가사량//노동량//작업량.
례01 「01」	(일부 명사 뒤에 붙어) ‘본보기’의 뜻을 나타내는 말.	인용례//판결례
마님 「02」	(일부 명사 뒤에 붙어) 상전(上典)을 높여 이르는 말.	대감마님 //영감마님.
마마 「04」	(임금 및 그의 가족과 관련된 명사 뒤에 붙어) ‘존대’의 뜻을 나타내는 말.	대비마마//대왕마마.
망09 「02」	(일부 명사 뒤에 붙어) 그물처럼 얽혀 있는 조직이나 짜임새의 뜻을 나타내는 말.	교통망 //연락망 //점포망//유통망//판매망.
명02 「02」	(일부 명사 뒤에 붙어) ‘이름’의 뜻을 나타내는 말.	꼭명//작품명//저자명
모12 「03」	(일부 명사 앞에 붙어) 어떠한 것에서 갈려 나오거나 생겨난 것의 근본이 됨의 뜻을 나타내는 말.	모기업 //모은행.
무침 「02」	(일부 명사 뒤에 붙어) ‘양념을 해서 무친 반찬’의 뜻을 나타내는 말.	시금치무침//복어무침// 골뱅이무침//파래무침.
문06 「01」	(일부 명사 뒤에 붙어) 학술 전문의 종류를 나타내는 말.	어학문 //법학문.
문06 「02」	(일부 명사 뒤에 붙어) 씨족에 따른 집안을 나타내는 말.	강씨문(姜氏門) //이씨문(李氏門).
미14 「02」	(일부 명사 앞 또는 뒤에 붙어) ‘아름다움’의 뜻을 나타내는 말.	미소년 //송고미 //우아미//각선미//교양미 //백치미//미남자.
반10 「03」	(일부 명사 뒤에 붙어) ‘작은 집단’의 뜻을 나타내는 말.	단속반//작업반
밭01 「05」	(일부 명사 뒤에 붙어) 그 식물이나 자연물, 수산물 따위가 많이 나는 곳.	고추밭 //대나무밭 //흙밭 //파래밭.
병03 「02」	(일부 명사 뒤에 붙어) ‘병사2’의 뜻을 나타내는 말.	운전병//탈영병.
병04 「02」	(일부 명사 뒤에 붙어) ‘질병2’의 뜻을 나타내는 말.	간질병 //심장병.
병05 「03」	(일부 명사 뒤에 붙어) ‘용기’를 나타내는 말.	농약병 //링거병 //요구르트병 //참기름병 //플라스틱병.
볶음 「02」	(일부 명사 뒤에 붙어) 볶아서 만든 음식의 뜻을 나타내는 말.	쇠고기볶음 //야채볶음.
불09 「02」	(일부 명사 뒤에 붙어) ‘부처1’의 뜻을 나타내는 말.	무량수불 //아미타불.

비05 「03」	(일부 명사 뒤에 붙어) ‘비율2’의 뜻을 나타내는 말.	농도비 //혼합비.
비19 「03」	(일부 명사 뒤에 붙어) 기념하여 세운 물건의 뜻을 나타내는 말.	문학비 //문인비.
빛 「07」	(일부 명사 뒤에 붙어) ‘빛깔’의 뜻을 나타내는 말.	능금빛 //산빛.
상04 「03」	(일부 명사 뒤에 붙어) ‘상차림’을 나타내는 말.	다과상 //생신상 //차례상.
상23 「02」	(일부 명사 뒤에 붙어) 조각이나 그림을 나타내는 말.	성당의 성모 마리아상.
상23 「03」	(일부 명사 뒤에 붙어) ‘모범2’, ‘본보기’의 뜻을 나타내는 말.	교사상 //어머니상.
상25 「02」	(일부 명사 뒤에 붙어) ‘상장10’, ‘상패4’, ‘상품4’ 따위의 뜻을 나타내는 말.	감독상 //봉사상 //선행상 //작품상 //효행상.
색03 「05」	(일부 명사 뒤에 붙어) ‘색깔’의 뜻을 나타내는 말.	딸기색 //바이올렛색.
선14 「07」	(일부 명사 뒤에 붙어) ‘광선1’의 뜻을 나타내는 말.	감마선 //엑스선.
식04 「04」	(일부 명사 뒤에 붙어) ‘수법’, ‘수식’을 나타내는 말.	곱셈식 //덧셈식 //나눗셈식 //뺄셈식.
쌍02 「03」	(일부 명사 앞에 붙어) ‘두 짝으로 이루어짐.’의 뜻을 나타내는 말.	쌍가락지 //쌍가마 //쌍권총.
씨01 「05」	(일부 식물이나 동물을 나타내는 명사 뒤에 붙어) 그 식물이나 동물의 씨를 나타내는 말.	배추씨//살구씨//굴씨//조개씨.
안04 「04」	(일부 명사 뒤에 붙어) ‘안전’의 뜻을 나타내는 말.	개정안 //채택안 //협상안.
알01 「09」	(일부 식물이나 동물을 나타내는 명사 뒤에 붙어) 그 식물이나 동물의 알을 나타내는 말.	머루알//은행알//타조알.
액03 「02」	(일부 명사 뒤에 붙어) ‘액체’의 뜻을 나타내는 말.	냉각액 //링거액 //수정액.
양20 「02」	(고유어와 외래어 명사 뒤에 붙어) 분량이나 수량을 나타내는 말.	구름양//알칼리양.
옥03 「02」	(일부 명사 앞에 붙어) ‘옥색1’, ‘옥제2’의 뜻을 나타내는 말.	옥채필이 //옥매트 //옥침대.
왜03 「03」	(일부 명사 앞에 붙어) ‘일본식의’, ‘일본의’의 뜻을 나타내는 말.	왜간장 //왜도시.
은04 「02」	(일부 명사 앞에 붙어) ‘은색’, ‘은제3’의 뜻을 나타내는 말.	은갈치 //은귀고리 //은목걸이 //은촛잔.
자08 「03」	(일부 명사 앞에 붙어) 모체에 달려 있음을 나타내는 말.	자회사.
잡이01 「04」	(일부 명사 뒤에 붙어) 민속놀이나 전통 음악에서 기술이나 채주, 장단 따위를 이르는 말.	
재비01	(일부 명사 뒤에 붙어) 국악에서, 악기를 연주하거나 노래를 부르거나 춤을 추는 기능자를 이르는 말.	가야금재비 //춤재비 //노래재비.
조15 「03」	(일부 명사 뒤에 붙어) 특정한 임무나 역할을 맡아 수행하기 위하여 조직하는 작은 집단을 나타내는 말.	작업조 //폭파조.
조림01 「02	(일부 명사 뒤에 붙어) 조리 음식의 뜻을 나타내는 말.	고등어조림 //연근조림.

」	나타내는 말.	
주24 「05」	(일부 명사 뒤에 붙어) ‘주식’의 뜻을 나타내는 말.	우량주//전환주.
주머니 「03」	(일부 명사 뒤에 붙어) 무엇이 유난히 많은 사람을 비유적으로 이르는 말.	고생주머니 //병주머니 //피주머니 //근심주머니
줍 「02」	(먹을 것을 나타내는 일부 명사 뒤에 붙어) ‘농축액’을 나타내는 말.	미나리즙 //석류즙 //배즙 //양파즙 //귤즙.
직06 「04」	(일부 명사 뒤에 붙어) ‘직무’, ‘직분’, ‘직업’, ‘직위’의 뜻을 나타내는 말.	사제직//사도직.
집01 「09」	(일부 명사 뒤에 붙어) 물건을 팔거나 영업을 하는 가게를 나타내는 말.	갈빗집 //고깃집 //꽃집 //피자집.
집01 「10」	(일부 명사 뒤에 붙어) ‘택호’를 나타내는 말.	“그럼, 이 집 택호는 영월집이라고 합시다. 알기 쉽게…….”
점01 「02」	(일부 명사 뒤에 붙어) 찢 음식의 뜻을 나타내는 말.	갈비점 //아귀점.
책01 「04」	(일부 명사 뒤에 붙어) ‘서적’임을 나타내는 말.	국어책//소설책//요리책.
터01 「04」	(일부 명사 뒤에 붙어) ‘자리1’나 ‘장소5’의 뜻을 나타내는 말.	낚시터 //놀이터 //일터 //휴터.
튀김01 「02」	(일부 명사 뒤에 붙어) 튀긴 음식의 뜻을 나타내는 말.	새우튀김 //오징어튀김
티02 「02」	(일부 명사 뒤에 붙어) ‘어떤 태도나 기색’의 뜻을 나타내는 말.	막내티 //소녀티 //중년티 //춘티.
표05 「07」	(일부 명사 뒤에 붙어) ‘그 사람이 만든 물건’의 뜻을 더하는 말.	엄마표 //아빠표 //신랑표 //주부표.
후08 「03」	(일부 명사 앞에 붙어) ‘뒤나 다음’의 뜻을 나타내는 말.	후더침 //후보름 //후서방.

아. [보완] 구어 형태 분석 말뭉치 (추가)

※ 구어 전사 말뭉치의 특성

- 구어 말뭉치에서 마침표는 하나의 문장이 끝났음을 나타내는 것이 아니라 억양 단위를 나타내는 기호이므로 주석할 때 주의해야 한다.

■■ 친구랑 같이 여행 왔어요 음

[오/VV+았/EP+어요/EF]

→ 구어 말뭉치에서 문장 기호는 억양 단위를 의미하기 때문에 위와 같은 예시에서 ‘왔어요’ 뒤에 임의로 마침표를 추가하지 않도록 한다.

→ 또한, 문장 기호가 없어서 자동 주석에서는 ‘어요’를 대부분 연결어미로 분석하는

네, 이 경우 종결어미로 분석해야 한다.

■■■ 그냥 매일매일 <u>쉬고</u> .	[쉬/VV+고/EC]
북경하고 고향에 갔다 왔다 갔다 <u>왔다</u> .	[오/VV+았/EP+다/EC]
<u>했어요</u> .	[하/VV+았/EP+어요/EF]
■■■ 어~ 그렇게 어 가고 싶지 <u>않았어요</u>	[않/VX+았/EP+어요/EF]
돈이 <u>없어서</u> .	[없/VA+어서/EC]
■■■ 아까 드렸던 종이를 한번 <u>살펴보고요</u> ,	[살펴보/VV+고/EC+요/JX]
다음으로 넘어갈게요.	

→ 구어 전사는 문장 단위가 아니라 억양 단위로 전사가 되기 때문에 하나의 발화가 여러 개의 억양 단위로 나뉘어져 제시될 수 있다. 따라서 연결어미도 종결부에 위치할 수 있어 분석을 할 때에 주의하여야 한다.

→ 억양 단위를 나타내는 문장 기호 역시 위의 예시처럼 연결어미 뒤에서도 나타날 수 있기 때문에, 맥락에 따라 어미를 구분해 분석해야 한다.

- 구어 형태 분석은 문어 형태 분석 지침을 따르지만 불완전하게 발화되거나 자기 수정을 하는 등의 끊어진 발화나 억양 단위 발화와 같이 구어 말뭉치의 특성을 드러내는 경우 아래와 같이 분석한다.

1) 완전한 어절

- 기본적으로 발화가 완전히 이뤄진 어절은 문어 형태 분석 지침을 따라 분석한다.

■■■ 저희가 <u>하여</u> 하고	[하/VV+아/EC]
■■■ 충고를 해 <u>준</u> <u>줘</u> <u>줄</u>	[주/VX+ㄴ/ETM 주/VX+어/EC 주/VX+ㄹ/ETM]

2) 끊어진 어절

- 끊어진 어절은 어절의 일부만 발화된 경우나 불분명한 경우인데, 앞/뒤에 완성된 발화가 나타나는 경우 끊어진 어절은 분석불가능(NA)로 처리한다.³⁴⁾

가) 어절의 일부만 발화되어 분석하지 않는 경우

34) 세종 구어에서는 끊어진 어절과 불분명한 어절을 UNT, UNC 등으로 세분하였지만, 학습자 자료의 특성과 작업자들의 혼동을 고려하여 통합하여 NA로 처리한다.

■■ 플라스 될 수 아이다, <u>마이너</u> = 아, 플라스는	[마이너/NA]
■■ 어 운동할 <u>시</u> = 힘도 부족해서	[시/NA]
■■ 어제 <u>약</u> = <u>약</u> = 약국에 갔어요	[약/NA 약/NA]
■■ 한국어 공부가 <u>힘</u> = 힘들었지만	[힘/NA]
■■ 의사 선생님도 곧 <u>나</u> = 나올 수 있다고	[나/NA]
■■ 경복궁에 가 <u>봐</u> = 봤다.	[봐/NA]
■■ <u>그랬</u> = 그랬어요.	[그랬/NA]
■■ 스페인에서 <u>있</u> = <u>있어</u> = 있어요.	[있/NA 있어/NA]
■■ 의사가= <u>가</u>	[가/NA]
■■ 슬퍼= <u>퍼</u> 가지구	[퍼/NA]

나) 어절의 일부를 더듬으며 반복하는 경우 (용언의 경우)

- 결혼하= 하기= 한 [결혼/NNG+하/XSV || 하/XSV+기/ETN || 하/XSV+ㄴ/ETM]
- 그런 게 제일 그 비결이라= 라고 하면 [비결/NNG+이/VCP+라/EF || 라고/EC]

다) 어절 중간에 간투사 따위가 들어가는 경우

- 심약 어 하다 [심약/NNG || 어/IC || 하/XSV+다/EF]
- 좋아 어 하다 [좋/VA+아/EC || 어/IC || 하/VX+다/EF]

3) 억양 단위가 바뀐 어절

가) 억양 단위가 형태소 경계로 바뀐 경우

- 발화자가 불완전하게 발화한 것은 아니지만 한 어절을 발화하는 도중에 억양 단위가 바뀌어서 조사나 어미 등 문법 형태소가 실질 형태소와 다른 억양 단위로 전사될 때, 억양 단위를 통합하지 않고 경계를 살려 형태 주석한다. 하지만 주석은 통합했을 때의 표지를 부여한다.

- 캔 유 두 미어 페이버? [페이버/NNG]
가 무슨 뜻? [가/JKS]
- 좋아. [좋/VA+아/EF]
라고 대답했지. [라고/JKQ]
- 주부 우울증. [우울증/NNG]
이라고 말할 수 있겠습니까. [이/VCP+라고/EC]

■ ■ 공부
한다고

[공부/NNG]
[하/XSV+ㄴ 다고/EC]

나) 억양 단위가 형태소를 가르는 경우

- 형태소 중간에 억양 단위가 바뀌어서 다른 억양 단위로 전사될 때, 각각 분석불가능(NA)으로 처리한다.

■ ■ 어,
제는 별일 없었어.

[어/NA]
[제/NA+는/JX]

■ ■ 선두주,
자가 도착했다.

[선두/NNG ㄹ 주/NA]
[자/NA+가/JKS]

4) 불분명한 어절

- 잘 들리지 않아 추측하여 전사한 어절은 최대한 분석하고 분석이 불가능한 경우에는 분석불가능(NA)으로 처리한다.

■ ■ 소리 중에 XXX 이게

[XXX/NA]

■ ■ 교육 개방 XX안이

[XX안/NA+이/JKS]

■ ■ XX에 제출돼

[XX/NA+에/JKB]

■ ■ XX스의 이론을

[XX스/NA+의/JKG]

■ ■ 신발을 X다

[X다/NA]

5) 간투사의 처리

가) 그, 저

- 그, 저 : 조사가 붙어 있다면 ‘대명사’, 조사 없이 확실하게 뒤의 명사를 수식할 때는 ‘관형사’, 위의 경우가 아니거나 확실하게 감탄사로 사용된 경우에는 ‘감탄사’로 처리한다. (※ 구분이 애매한 경우 감탄사로 분석한다.)

■ ■ 그는 참으로 좋은 사람이다.

[그/NP+는/JX]

■ ■ 그 책 좀 이리 줘 봐.

[그/MM ㄹ 책/NNG]

■ ■ 그 무엇인가를 알아내고자 했지만

[그/MM ㄹ 무엇/NP+이/VCP+ㄴ 가/EF+를/JKO]

- ■ 그 왜 있잖아요. [그/IC]
- ■ 이도 전도 다 싫다. [저/NP+도/JX]
- ■ 전 둘 중에 하나를 선택해라. [저/MM || 둘/NR || 중/NNB+에/JKB]
- ■ 전, 뭐라더라.. [저/IC]
- ■ 전 말씀 중에 잠시 실례하겠습니다.. [저/IC]

나) 아니

- 아니 : 대답이나 감탄일 때는 ‘감탄사’, 부정이나 반대의 뜻을 나타낼 때나 명사와 명사 또는 문장과 문장 사이에서 강조할 때는 ‘부사’로 처리한다.
- ■ A : 자니?
B : 아니, 안 자. [아니/IC]
B' : 아니요, 안 자요. [아니요/IC]
- ■ 아니, 그럴 수가 있니? [아니/IC]
- ■ 아침까지만 해도, 아니 점심 먹을 때만 해도... [아니/MAG]

다) 그래

- 그래 : 대답이나 감탄, 놀라움, 담화 표지로 쓰였을 경우는 ‘감탄사’, 서술어의 대용으로 쓰였을 경우에는 용언의 활용형으로 분석한다.
- ■ A : 점심에 같이 밥 먹을까?
B : 그래, 알겠어. [그래/IC]
- ■ A : 점심에 같이 밥 먹을까요?
B : 그래요, 뭐 먹을까요? [그래/IC+요/JX+?/SF]
- ■ 왜 그래요?
[그렇/VA+어요/EF+?/SF]
[그리/VV+어요/EF+?/SF]

<주의사항>

- (가) [보완] 맥락에 따라 감탄사로 쓰였는지 판단이 어려운 경우가 있다. 이 때 그 형태가 선·후행 형태소와 같을 때는 선·후행 형태소를 반복한 것으로 분석하고, 그렇지 않은 경우에는 감탄사로 처리한다.

- ■ 학교에 예 [예/JKB]
가서 예 반 친구를 만났어요 [예/IC]
- ■ 그 그 사람은 제 친구예요. [그/NP|| 그/NP|| 사람/NNG+은/JX]
어학당에서 그 처음 만났어요. [그/IC]

(나) [보완] 감탄사가 반복되는 경우에는 구어 전사에서 구분한 어절 경계에 따라 형태 주석한다.

■■ 네네 맞아요.

[네네/IC]

■■ 네 네 네 그래서

[네/IC || 네/IC || 네/IC]

6) 구어형의 분석

- 세종 문어 형태 분석 지침에는 구어형 분석에 대한 기술이 자세하지 않다. 따라서 기본적으로는 문어 형태 분석 지침을 중심으로 여기서도 분석을 하지만, 일부 해결할 수 없는 경우에 한해서는 세종 구어의 형태 분석을 따른다.

■■ 뭘로

[무엇/NP+으로/JKB]

■■ 걸로

[것/NNB+으로/JKB]

→ 문어 지침에서 대명사 ‘뭐’와 의존 명사 ‘거’가 그 형태가 유지되지 않고 조사와 축약되어 나타나는 경우에는 각각 ‘무엇’과 ‘것’으로 복원하고 있다. 구어에서도 위와 같은 문어 지침을 따라 원형을 복원해준다.

■■ 그쵸, 그쵸

[그쵸/IC], [그쵸/IC]

■■ 이케 하면 되나요?

[이케/MAG]

■■ 여따 집어넣어

[여따/MAG]

→ 음운적 축약이 일어나 형태 분석이 불가능한 경우는 해당 축약형 전체가 가지는 기능을 고려해 형태 표지를 할당한다.

자. [보완] 구어 형태 분석 말뭉치 (시스템 관련)

1) 교정 어절이 있는 경우 형태 주석

- 구어 말뭉치의 경우 구어 팀에서 학습자 오류 어절에 대해 일차적인 ‘교정 어절’을 주고 있다. 따라서 교정 어절이 없는 문어와 달리 구어에서는 형태

주석 단계에서 아래와 같이 ‘교정 어절’과 ‘교정 주석’이 함께 나타나는데, 교정 어절에 대해서도 형태 주석을 수정해 준다.

가) 통합이나 분할이 필요 없이 수정이 가능한 경우

작업 문장 정보					
	4	15	16	17	
원어절		많이	꼭	고	문화
형태주석		MAG	VV	EC	NNG
교정어절			꼭	고	
교정주석			VV	EC	



작업 문장 정보					
	4	15	16	17	
원어절		많이	꼭	고	문화
형태주석		MAG	VV	EC	NNG
교정어절			꼭	고	
교정주석			VV	EC	

나) 통합이든 분할이든 수정했을 때 분석이 가능한 경우

- ‘그때’는 한 단어로 NNG로 분석해야 한다. 먼저 원어절에서 어절 경계가 분할된 1번과 2번의 어절을 통합해서 1번을 ‘극때’로 수정하고 NNG로 분석을 한 후, 2번을 삭제한다. 이때 아래 교정 어절에 ‘그/MM’만 남게 되는데 ‘그때’로 수정하고 NNG로 교정 주석을 수정한다.

작업 문장 정보				
	1	2	3	
원어절	극	때	부터	한
형태주석	MM	NNG	JX	N
▶ 교정어절	그			
교정주석	MM			



작업 문장 정보			
	1	2	
원어절	극때	부터	한
▶ 형태주석	NNG	JX	N
교정어절	그때		
교정주석	NNG		

다) 통합이든 분할이든 수정할 수 없으며, 원어절의 형태주석을 NA 처리해야 하는 경우

- 구어 팀에서 제시한 교정어절대로 형태 분석하기에 무리가 있다고 판단되는 경우는 NA로 분석한다. 아래 제시된 경우처럼 ‘스고시’는 ‘것이’의 오류로 ‘것+이’로 분리하여 분석하기에는 무리가 있으므로 NA로 처리해야 한다. 이 경우 교정어절에 ‘것’ 위의 형태주석 1번 칸의 ‘스고시’는 NA로 수정하고, 2번 칸의 형태 주석은 시스템 왼쪽 아래의 ‘주석 삭제’ 기능을 이용하여 삭제하고 빈칸으로 남겨둔다.

작업 문장 정보					
	1	2	3	4	5
▶ 원어절	스고시		별로	안	꼼꼼
형태주석	NA		MAG	MAG	MAG
교정어절	것	이			
교정주석	NNB	JKS			

- 위와 같이 구어에서 제시한 교정어절대로 형태 분석을 하기 어려운 경우는, 메모를 남겨 다음 단계 작업에서 확인할 수 있도록 한다.

<메모 형식> 구어 팀에서 제시한 교정어절대로 ‘○○’을 ‘△△’으로 교정할 가능성이 낮아 보임

한국어 학습자 말뭉치 오류 주석 지침

I. 학습자 말뭉치 오류 주석 체계 틀

1. 기본 주석

- 오류 위치는 오류가 나타난 부분의 품사를 주석한다. 오류 위치는 기본 주석으로 형태소 분석에 기대어 모든 오류에 대해 오류가 발생한 품사에 전수 주석한다. 모든 오류는 오류 위치 검색으로 찾을 수 있다.³⁵⁾

	오류 유형		주석 표지
분석 불가능	전체적 오류 포함		IMP
오류 위치	실질어휘	고유명사	CNNP
		일반명사	CNNG
		의존명사	CNNB
		대명사	CNP
		수사	CNR
		동사	CVV
		형용사	CVA
		보조용언	CVX
		지정사	CVC
		관형사	CMM
		일반부사	CMAG
		접속부사	CMAJ
		감탄사	CIC
		접두사	CXPN

35) 구 단위 주석과 표현 문형 주석은 구 전체와 구 구성요소에 각각 주석함을 원칙으로 한다.

	오류 유형		주석 표지
		명사파생접미사	CXSN
		동사파생접미사	CXSV
		형용사파생접미사	CXSA
		어근	CXR
	기능어휘	주격조사	FNP
		관형격조사	FGP
		목적격조사	FOP
		부사격조사	FAP
		접속조사	FJC
		보격조사	FCP
		호격조사	FVP
		인용격조사	FQP
		보조사	FXP
		연결어미	FED
		종결어미	FFE
		선어말어미	FPE
		명사형 전성어미	FNE
		관형사형 전성어미	FAE
	구 단위 표현		PHE
	표현 문형		PE

2. 확장 주석

- 확장 주석은 한국어교육의 선행 오류 연구에서 유의미한 주석에 초점을 두어 교수자의 활용에 초점을 둔 주석이다. 연구자들은 필요한 주석을 추가하여 스스로 주석할 수 있다. 교정 어절에 대한 형태 주석에 기대어 주석한다.

2.1. 오류 양상

○ 어휘나 문법의 층위에서 발생하는 오류 양상만을 주석한다.³⁶⁾

	오류 유형	주석 표지
오류 양상	누락	OM
	첨가	ADD
	대치	REP
	오형태	MIF

2.2 오류 층위

○ 교정 어절에 대한 형태 주석에 기대어 주석한다. ‘발음’은 구어 자료에 한하여 주석한다.

	오류 유형		주석 표지
오류 층위	발음	음소	PP
		음절	PS
		음운규칙	PC
		원어식 발음	PN(임시 기호)
		중간 발음(변이음포함)	PA(임시 기호)
	형태	단어 형성[합성법]	MCP
		단어 형성[파생법]	MDV
		굴절[곡용]	MDC
		굴절[활용]	MCJ
		품사	POS
	통사	높임	SH

36) 오류의 양상은 이론적으로는 누락, 첨가, 대치 중 하나이나, 단순 철자 오류나 활용 오류 같은 것들은 이 기준으로 분류하는 것이 무의미하므로, 오형태로 별도 처리하였다.


	오류 유형		주석 표지
		시제	ST
		사동	SC
		피동	SP
		부정	SN
		어순	WO
	담화	지시	DR
		접속	DC
		담화표지	DM
		구어/문어 오류	DS

II. 오류 판정 및 수정 지침

1. 기본 원칙

1) 오류의 식별

- 오류의 식별은 오류 여부를 식별하는 것으로부터 시작된다. 교정 어절을 만들거나 교정 어절(때로는 어절을 넘는 단위)을 만들 수 있는 가능성이 있는 경우를 오류로 본다.
- 오류의 판단은 문법성을 기준으로 삼는다. 문법성이란 의미적으로나 형태적으로 완성된 형식을 갖추지 못하고 한국어의 문법 체계에 맞지 않는 비문법적 문장을 생성하는 경우를 말한다. 즉, 문법성을 기준으로 어문 규범에 어긋나며, 용인하기 어려운 일탈은 모두 오류 판정과 주석의 대상으로 삼는다.

<예> 우리는 술을 마시고 싶으면 ‘바프라이’(BARFLY)(√ ‘바이프라이’라고 하는) 술집에 가요.
 초급 학습자가 생성한 문장으로 ‘라고 하는’을 포함한 문장

이 초급보다 높은 수준이지만 정확한 문장 생성에 실패하였으므로 오류로 주석한다.

- 외국어로 표기된 것은 오류로 본다.

<예> 그리고 제 new(√새로운) 친구들은 많이 만나고 싶습니다.
 📖 'new'라고 영어를 그대로 표기한 것은 한국어와 외국어의 대치 오류로 주석한다.

- 오류의 판단에는 용인 가능성도 고려될 수 있는데, 이는 오류 주석자에 따라 달라질 수 있으므로 복수의 주석자가 지침을 통해 합의하여 판정한다. 용인 가능성이라는 기준은 '엄격하게' 적용하여 '일관되게' 처리하도록 한다.
- 어휘 혹은 문법 오류로 동시에 판정할 수 있는 경우, 기능어 중심으로 문법 오류를 우선시하여 처리한다.
- 오류의 판단은 문장 단위에서 이루어진다. 오류 판정 시 문제가 될 때에는 앞뒤 문장까지는 살펴볼 수 있지만, 주석의 일관성을 위해 담화 단위로 보지 않고 기본적으로 문장 단위에서만 처리하도록 한다. 단, 오류 층위에서 담화 오류에 해당하는 지시(DR), 접속(DC)의 경우, 선행문과 후행문과의 의미적 연결을 고려해야 오류 판단이 가능하기 때문에 앞, 뒤 문장을 고려하여 오류를 판단한다.
- 구어 자료의 경우, 문장으로 파악하지 않고 억양 단위로 끊어서 각 단위를 기준으로 오류를 식별하고 판정한다.

<예> 무슨 파티하면
 우리 학생들이.
 열심히 공부한=
 연세대학교 열심히 공부해서
 조금 피곤한,
 =것이에요.
 📖 이 경우 억양 단위로 끊어서 보면 크게 문제가 되지 않지만, 문장 단위로 보면 여러 가지 층위에서 오류 처리가 가

능하며 일관된 기준에 의한 처리가 어렵다. 구어 자료는 문장 단위가 아닌 억양 단위를 기준으로 하여 오류를 식별하고 판정한다.

2) 오류와 실수의 구분

- 오류와 실수는 구분하지 않는다. 즉, 실수인지 오류인지의 여부와 상관없이 규범상의 일탈은 모두 오류로 간주한다. 이는 연구자의 판단 영역으로 자료만으로 학습자의 의도를 파악할 수 없기 때문에 주석 작업자의 자의적인 해석을 막기 위한 것이다.
- 구어 자료에서 발화 중에 학습자의 자기 수정이 일어난 경우, 수정하기 이전의 일탈은 오류로 간주하지 않는다. 학습자 스스로 오류임을 인지하고 수정을 하였으므로 수정 후 발화에 초점을 두고 오류 여부를 판정한다. 수정 후 발화에서도 오류가 나타난 경우는 이전의 일탈도 모두 오류로 주석하고, 수정 후 발화가 제대로 되었을 경우에는 이전의 일탈은 오류로 주석하지는 않으나 교정어절은 써주도록 한다.

3) 오류의 교정(교정 어절 원칙)

- 오류의 교정은 오류로 식별된 부분을 올바르게 고치는 것을 말한다. 따라서 오류로 식별된 것은 교정의 대상이 된다.
- 오류의 교정은 학습자의 표현 의도를 고려하여 최소한의 교정을 원칙으로 한다. 즉, 학습자의 표현 의도나 의미를 자의적으로 유추하여 교정 어절을 생성하지 않으며, 학습자가 산출한 형태를 가능한 한 훼손하지 않고 최대한 원문을 유지할 수 있는 형태로 수정한다.

<예> 현대(√ 현재) 세계적으로 환경 문제가 대두되고 있다.
 ☞ ‘현대’를 ‘현재’로 수정

- 오류로 판정된 문장을 교정할 때 그것을 문법적으로 완전한 문장으로 바꿀 것인지 용인가능한 수준의 문장으로 바꿀 것인지와 관련하여서는 학습자의 표현 의도를 반영하여 용인 가능한 수준으로 최소한의 교정을 하며, 한국

어 모어 화자의 보편적인 언어 사용 방식에 따라 교정한다.

- 또한 오류의 교정은 정보가 소실되지 않는 차원에서 최소한의 교정을 원칙으로 한다. 즉, 앞부분의 오류를 수정하는 것으로 인해 뒷부분에까지 영향을 미쳐 뒷부분까지 교정이 필요한 경우, 학습자의 의도에서 벗어날 수 있으며 주석자의 자의적인 해석이 지나치게 반영될 수 있기 때문에 앞 부분에서 최소한의 교정만 하며, 전면 교정이 필요할 때에는 분석불가능(IMP)으로 주석한다.
- 오류 영역에서 교정 어절로 인해 조사나 어미가 바뀌는 경우, 교정 어절의 영향을 받아 바뀐 조사와 어미는 오류로 처리하지 않는다.
- 기본적으로 맥락을 살펴 되도록 내용어보다 기능어를 우선 교정하는 것을 원칙으로 하므로, 뒤의 용언을 바꾸지 않는 방향에서 조사 오류로 처리하는 원칙이 우선이지만 용언을 반드시 교정해야 할 경우, 용언이 대치되면서 용언 때문에 조사가 바뀌는 경우에는 용언 대치 오류로만 처리하고 조사 오류로는 처리하지 않는다. 단, 사동과 피동 오류에 한하여 사동사와 피동사로 바뀌면서 조사가 바뀔 때에는 사동/피동과 관련한 오류라는 것을 표시해주는 차원에서 조사도 오류로 주석하며, 오류 층위에 사동과 피동을 주석한다.

<예> 아파트가 평형이(√평수가) 많으면(√넓으면, √크면) 친구들을 부를 수 있다.

☞ ‘평형’을 ‘평수’로 교정함에 따라 조사 ‘이’가 ‘가’로 바뀌게 되었으므로 조사는 오류로 주석하지 않는다. 이때에는 ‘평형’과 ‘많다’만 대치 오류로 처리하고, 조사 ‘이’는 오류로 주석하지 않는다.

나라가(√나라를) 발전하다(√발전시키다)

☞ 사동 ‘시키다’로 교정해야 할 경우, 조사까지 교정해야 한다. 이 경우에는 조사 ‘가’와 ‘를’도 대치로 처리한다. 따라서 [오류 위치-주격조사], [오류 양상-대치], [오류 층위-사동]과 [오류 위치-동사파생접미사], [오류 양상-대치], [오류 층위-사동]오류로 주석한다.

4) 오류 판정의 대상

- 오류 판정은 오류로 식별되어 교정된 부분이 어떤 범주의 오류인지를 판정하는 것을 말한다.
- 오류 판정은 오류에 대한 주석이므로 교정 어절이 아닌 오류 어절(원어절)을 기준으로 한다. 즉, 학습자가 산출한 언어 형태와 오류 발생 위치를 기준으로 오류를 판정한다.³⁷⁾

<예> 가끔 술을 마시지 않아서(√않을 때는) 영화를 보러 영화극장에 갈 거예요.

☞ 이 경우 오류 어절인 ‘않아서’의 ‘아서’를 기준으로 하여 [어미 오류]로 판정한다.

호주는 어디든지(√어디인지? 어디에 있는지?) 알아요?

☞ 오류 어절을 기준으로 하면 ‘든지’의 오류로 보아 [조사 오류]로 주석한다.

- 오류 주석은 형태소 단위를 기본으로 한다. 따라서 구 단위 이상의 어휘나 표현은 구성 요소를 형태로 나누어 분석한다. 단, 표현 문형의 경우는 <국립국어원 2> 목록을 확인하여 해당 표현이 목록에 있을 경우는 표현 문형도 함께 주석한다. 따라서 형태 단위로 분석하여 오류로 처리하는 동시에 ‘표현 문형’ 오류도 중복 주석한다. (☞ 3. 범주별 세부 오류 유형의 처리 예시-2) 오류 위치-(4) 표현 문형(PE) 참고)

<예> 내일은 비가 온 것(√올 것) 같아요.

☞ ‘(으)ㄴ 것’, ‘(으)ㄹ 것’은 표현 문형 목록에 포함되어 있기 때문에 이 경우에는 오류가 발생한 ‘온 것’을 하나의 덩어리 표현으로 처리하는 동시에 구성 요소인 관형사형 전성어미로도 분석하여 오류가 나타난 위치를 중복 주석한다. 즉, [오류 위치-표현 문형, 관형사형 전성어미], [오류 양상-대치] 오류로 주석한다.

37) 이후 웹사이트에서는 교정어절을 중심으로 한 검색도 가능하게 하여, 미사용으로 인한 오류를 파악할 수 있게 한다.

은행에 저축한(√저축할) 겨우에는(√경우에는) 얼마정도 이익을 얻을지 미리 알아서 더 편할 것 같아요.

☞ ‘-(으)ㄴ 경우에는’을 한 덩어리로 처리할 수도 있으나, 본 연구에서 참고 목록으로 삼고 있는 <국립국어원 2> 목록에 표현 문형으로 제시되어 있지 않기 때문에 [오류 위치-관형사형 전성어미], [오류 양상-대치]와 [오류 위치-명사], [오류 양상-오형태] 오류로 각각 처리한다.

5) 기타

- 문장부호 사용에 관한 오류는 주석 대상에서 제외한다. 즉, 학습자가 생략 또는 누락한 문장부호가 있다고 하더라도 오류로 판정하지 않는다.

<예> 예) 광고가 주는 정보가 모두 진실이 아니라는 예방적인 생각도 필요해요(√ 온점 누락)

2. 오류의 범주

- 본 연구에서는 오류의 범주를 오류 위치와 오류 양상, 오류 층위 세 가지로 설정한다. 그리고 이를 다시 기본 주석과 확장 주석으로 이원화 하여 오류를 주석한다.
- 기본 주석은 오류 위치가 해당되며, 모든 오류에 대해 1:1로 주석하는 필수 주석이다(분석이 불가능한 오류는 분석 불가능[IMP] 표지로 주석하며, 표현 문형의 경우는 각 형태소와 표현 문형[PE] 표지가 중복 주석된다). 확장 주석은 오류 양상과 오류 층위가 해당되며, 이는 관련 오류가 있는 경우에만 주석하는 수의적 주석이다. 확장 주석의 경우, 한 형태에 2개 이상의 오류가 나타나면 중복 주석이 가능하다.

1) 분석 여부

- ‘분석 여부’의 판단은 오류로 식별된 형태에 대해 교정이 가능한지 여부와 특정 범주의 오류로 판정 가능한지를 파악하는 것을 말한다. 따라서 부적절한 표현이 연속되거나 문장 구조의 이상으로 학습자의 표현 의도를 파악하기 어려운 경우 ‘분석 불가능[IMP]’으로 판정할 수 있다.

영역	주석 표지	포함 범위	예시
분석 불가능	IMP	문맥 내에서 해석이 불가능한 경우	한국여자 종종하고(√좋고? 조용하고?, IMP) 예쁘기 때문에(때문에, MIF) 결혼(√결혼, MIF)하고 싶어요.

2) 오류 위치

- 오류가 발생한 위치 표지로서 [오류 위치]를 주석한다. 오류 위치는 오류가 일어난 부분, 즉 오류가 발생한 위치의 품사(형태소)에 대해 주석한다.³⁸⁾
- 오류 위치는 기본적으로 형태 주석에 따라 처리한다. 형태 주석에서 <표준국어대사전>에 근거하여 형태소 분석을 하기 때문에, 오류 주석은 이에 입각하여 오류 위치를 주석한다. 다음은 오류로 식별된 부분의 품사 위치를 표시한 주석 표지이다.

위치		주석 표지	포함 범위	예시
실질 어휘	고유 명사	CNNP	고유 명사 어휘의 사용에서 나타난 오류	그리고 저는 독묘(√도묘, CNNP, MIF)에 가고 싶어요.
	일반 명사	CNNG	일반 명사 어휘의 사용에서 나타난 오류	애기와(√아기와, CNNG, MIF) 노인들한테 건강이 나빠졌다.

38) 단, 누락 오류의 경우에는 원 어절이 없으므로 교정 어절에 따라 주석한다.

위치	주석 표지	포함 범위	예시
의존 명사	CNNB	의존 명사 어휘의 사용에서 나타난 오류	그럼데 아쉬운 건(√것, CNNB, MIF)도 있다.
대명사	CNP	대명사 어휘의 사용에서 나타난 오류	내(√우리, CNP, REP) 아버지가 남편하고 친하게 되면 좋겠다.
수사	CNR	수사 어휘의 사용에서 나타난 오류	이 셋(√세, CNR, MIF) 가지 단어의 뜻에 따라 이 외모지상주의라는 말을 충분히 이해할 수 있다.
동사	CVV	동사 어휘의 사용에서 나타난 오류	그로 인해 평소 일상생활에서 말할 수 없는 말, 욕하는 말, 비우는(√비웃는, CVV, MIF) 말 등 흔히 볼 수 있다.
형용사	CVA	형용사 어휘의 사용에서 나타난 오류	불고기 먹기 때문에 기분이 기쁩니다(√좋습니다, CVA, REP).
보조 용언	CVX	보조용언 어휘의 사용에서 나타난 오류	다른 사람에게 아픈다운 모습을 보여 싶기(√주기, CVX, REP) 위하여 노력하세요.
지정사	CVC	지정사 어휘의 사용에서 나타난 오류	그러니까 저는 외모지상주의가 위험이라고(√위험하다고, CVC, REP)생각한다.
관형사	CMM	관형사 어휘의 사용에서 나타난 오류	그렇게 되면 어느(√어떤, CMM, REP) 사람은 돈이나 개인 정보를 잃어버릴 수도 있다.

	위치	주석 표지	포함 범위	예시
	일반 부사	CMAG	일반부사 어휘의 사용에서 나타난 오류	내 남편은 꼭(✓정말, CMAG, REP) 멋있게 생겼다.
	접속 부사	CMAJ	접속부사 어휘의 사용에서 나타난 오류	그런데(✓그런데, CMAJ, MIF) 제 가격을 정말 보고 싶습니다.
	감탄사	CIC	감탄사 어휘의 사 용에서 나타난 오 류	‘네(✓네, CIC, MIF), 알겠습니다’라고 대답했다.
	접두사	CXPN	접두사 어휘의 사 용에서 나타난 오 류	최소 임금을 실행하면 저소득층(✓저소득층, CXPN, MIF) 사람의 살기가 보증할 수 있다.
	명사파생 접미사	CXSN	명사 파생 접미사 어휘의 사용에서 나타난 오류	두 번째(✓번째, CXSN, MIF)에 간 곳이는 경주였습니다.
	동사파생 접미사	CXSV	동사 파생 접미사 어휘의 사용에서 나타난 오류	그 꿈을 위해서 매일 운동 해다(✓운동한다, CXSV, MIF).
	형용사파 생접미사	CXSA	형용사 파생 접미 사 어휘의 사용에 서 나타난 오류	이러한 사회에서 자신이 하고 싶은 직업을 할 수 있으면 너무 행복안(✓행복한, CXSA, MIF) 것이다.
	어근	CXR	어근 어휘의 사용 에서 나타난 오류	시원(✓시원한, CXR, REP) 웃을 준비하세요.
기능 어휘	주격 조사	FNP	주격조사의 사용 에서 나타난 오류	그리고 여행이(✓을, FNP, REP) 너무 좋아합니다. 고시원에서 많이 학생(✓이,

위치	주석 표지	포함 범위	예시
			FNP, OM) 살았다.
관형격 조사	FGP	관형격조사의 사용에서 나타난 오류	‘론딩리’의(✓를, FGP, REP) 소개합니다.
목적격 조사	FOP	목적격조사의 사용에서 나타난 오류	그래서 저는 한국 사람하고 다른 외국 시람을(✓과, FOP, MIF/REP) 교류하고 싶습니다.
부사격 조사	FAP	부사격조사의 사용에서 나타난 오류	차 안에(✓에서, FAP, REP) 잤어요.
접속 조사	FJC	접속조사의 사용에서 나타난 오류	한국어 문법와(✓과, FJC, MIF) 중국어 문법이 비슷하지 않았다.
호격 조사	FVP	호격조사의 사용에서 나타난 오류	친구아(✓친구야, FVP MIF), 같이 가자.
인용격 조사	FQP	인용격조사의 사용에서 나타난 오류	내가 감사하다고 말한다는 게 ‘고맙다’다고(✓라고, FQP, REP) 했다.
보조사	FXP	보조사의 사용에서 나타난 오류	론딩리는 맛있는 음식은(✓이, FXP, REP) 많습니다.
연결 어미	FED	연결어미의 사용에서 나타난 오류	그리고 친구들과 같이 노래방 가고(✓가서, FED, REP) 노래를 부르고 싶습니다.
종결 어미	FFE	종결어미의 사용에서 나타난 오류	특히 아랫목에 정말 따뜻한다(✓따뜻하다, FFE, MIF).
선어말 어미	FPE	선어말어미의 사용에서 나타난 오류	내일부터 수업이 시작됐어요(✓시작돼요,

위치	주석 표지	포함 범위	예시
		류	FPE, REP).
명사형 전성 어미	FNE	명사형 전성어미 의 사용에서 나타 난 오류	우리 계획은 저녁을 먹기(√먹은, FNE, REP) 후에 우리 만든 신분증을 가지고 갈 겁다.
관형사형 전성어미	FAE	관형사형 전성어 미의 사용에서 나 타난 오류	중요하는(√중요한, FAE, MIF) 것은 사람마다 다르다는 것을 인정하는 것이다.
구 단위 표현	PHE	구 단위 표현 사 용에서 나타난 오 류	앞으로는 얼마나 어려운 일이 생기는 때(√생겨, PHE, REP)도 포기하지 않고 사소한 일에도 최선을 다하겠다.
표현 문형	PE	보조 용언이나 여 러 요소의 결합 구성으로 이루어 진 표현 문형의 사용에서 나타난 오류(제시 목록 참고)	신세대는 기성세대와 가치관이 달라서 세대 차이가 생기기 만련이다(생기기 마련이다, PE, MIF).

3) 오류 양상

- 오류 양상은 표면적으로 드러난 오류의 모습으로, 누락, 첨가, 대치, 오형태 4가지로 설정한다.
- 오류 양상은 확장 주석으로 수의적 주석에 해당한다. 따라서 누락, 첨가, 대치, 오형태 오류로 보기 어려운 오류 양상은 주석하지 않는다.

양상	주석 표지	포함 범위	예시
----	-------	-------	----

양상	주석 표지	포함 범위	예시
누락	OM	완전한 문장/발화에서 나타나야 할 형태가 빠져 있는 경우	저는 여덟 시부터 여덟 시 삼십분까지 저녁(✓을, FOP, OM) 먹어요.
첨가	ADD	완전한 문장/발화에서 나타나지 말아야 할 형태가 쓰인 경우나 중복된 형태를 반복해서 사용한 경우	한국말은 동경에 있었을 때, 일년간 동안(✓일년 간, CNNG, ADD) 한국 YMCA에서 공부했습니다.
대치	REP	다른 의미의 어휘를 사용하거나 적절한 품사를 사용하지 못한 경우	용서를 줄(✓할, CVV, REP) 수 있게
		한국어에 없는 표현이나 한국어가 아닌 다른 언어를 사용한 경우	나는 cousin(✓사촌, CNNG, REP)한테 이야기했어요.
오형태	MIF	오형태 오류: 한국어에 존재하지 않는 어휘를 만들어 내거나 조사와 어미의 활용 형태가 잘못된 경우 즉, 활용 또는 곡용을 잘못하여 다른 이형태를 사용한 경우	이 시간은 별로 덥지 않고 시원해서(✓시원해서, FED, MIF, MCJ) 숙제하기에 좋습니다.
		맞춤법 오류: 철자를 잘못 사용한 경우	우리는 피간했어요(✓피곤했어요, CNNG, MIF).

4) 오류 층위

- 오류 층위는 오류로 식별된 부분을 언어학적 층위에 따라 나눈 오류의 범주이다. 즉, 언어학적 측면에서 어느 영역의 오류인지를 판단하는 것으로, 본 연구에서는 오류 층위를 교수자나 학습자들이 자주 활용할 일부 영역(발음, 형태, 통사, 담화)에 한정하여 주석하였다.
- 오류 층위는 오류 어절(원어절)과 교정 어절 모두를 고려하여 해당 층위에 맞게 주석한다.
- 발음 층위는 구어에서의 발음 오류를 다루는 영역이다. 발음 층위에서는 음소, 음절, 음운 규칙에서 발생하는 오류와 학습자의 원어식 발음, 변이음을 포함한 중간 발음에서 나타나는 오류를 주석한다.

층위		주석 표지	포함 범위	예시
발음	음소	PP	음소 차원에서 발생하는 오류 예) 평음, 격음, 경음의 구분	[구어] 케이키도(√케이크도, CNNG, PP) 있고 생일파티 주인공이도 있어요.
	음절	PS	음절 차원에서 발생하는 오류. 음절의 발음을 정확하게 하지 못한 경우로 원래 음절보다 더 적게 혹은 많이 발음한 경우와 축약해야 하는데 축약하지 않고 발음한 경우 또는 그 반대의 경우	[구어] 우리::= 우리나라도:: 마야크(√마약, CNNG, PS)... 어:: 판매::, 될 수 있..=있긴 한데::
	음운	PC	음운변동에 관한	[구어]

층위		주석 표지	포함 범위	예시
형태	규칙		오류로 구어에서 필수적 음운 규칙의 이탈 또는 음운 규칙을 적용하지 않고 절음화하여 발음한 경우 예) 연음규칙, 비음화 유음화, 구개음화, 경음화 등	한국계(√한국에, CNNG, PC) 가요.
	원어식 발음	PN	원어식 발음으로 발생하는 외국어 오류	[구어] 예를 들면 자기 계발, 재미, 수업 등 그래서 아래 그래프(√그래프, CNNG, PN)를 보며는,
	중간 발음(변이음포함)	PA	변이음을 포함한 중간 발음으로 발생하는 오류.	[구어] 전공(√전공, CNNG, PA) (‘저’와 ‘조’의 중간발음)

- 형태 층위는 어휘 오류를 다루는 영역이다. 형태 층위에서는 합성어, 파생어 등의 조어 과정에서 발생하는 오류와 어미의 활용, 조사의 사용 등에서 나타나는 오류를 주석한다.

층위		주석 표지	포함 범위	예시
형태	단어 형성[합성법]	MCP	단어 합성에서 나타나는 오류	해물고기(√물고기, CNNG, MCP)가 많았어요.
	단어 형성[파생]	MDV	단어 파생에서 나타나는 오류	작년 방학 때는 LG 전자에서 통역사로 일한

층위		주석 표지	포함 범위	예시
통사	법]			경험이 있고 한국에 대한 사이트에서 번역사(√번역가, CNNG, POS/MDV)로 일한 경험도 있기 때문에 <name>에서 일할 수 있는 자신을 가지고 있다.
	굴절 [곡용]	MDC	조사 이형태 선택에서 나타나는 오류	론딩리는 지하철와(√과, FAP, MIF, MDC) 가까워서 편리합니다.
	굴절 [활용]	MCJ	용언과 어미의 활용에서 나타나는 오류	내가 10년 후에 좋하고 행복 살으면(√살면, FED, MIF, MCJ) 좋겠다
	품사	POS	동일 의미의 품사 선택에서 나타나는 오류	주말에 친구하고 같이 유명의(√유명한, CNNG, REP, POS/MDV) 곳이 가고 싶습니다.

- 통사 층위는 문법 오류를 다루는 영역이다. 통사 층위는 높임, 시제, 사동, 피동, 부정 등의 문법 범주와 관련되어 해당 문법범주를 제대로 사용하지 못했을 경우 주석한다.

층위		주석 표지	포함 범위	예시
통사	높임	SH	조사, 선어말어미, 종결어미 등 높임 관련 문법 형태소와 높임 어휘의 오류	할머니께서 우유를 마시십니다(√드십니다, CVV, REP, SH).
	시제	ST	시제를 나타내는 문법 형태소의 오류	어제부터 항상 시계를 확인하기로 한다(√했다, FPE, REP, ST).

층위	주석 표지	포함 범위	예시
사동	SC	사동사, 사동 표현의 오류	갈릴레이는 새로 발명된 망원경을 사용하여 연구를 깊였다. (√깊게 하였다, CVA, MIF/REP, SC)
피동	SP	피동사, 피동 표현의 오류	교실 문이 닫아(√닫혀, CVV, REP, SP) 있었습니다.
부정	SN	부정 표현의 오류	한국에 온 후에 한 문장도 못(√Ø, CMAG, ADD, SN) 들을 수 없었다.
어순	WO	한국어의 통사 구조에 맞지 않는 방식으로 문장이 배열된 오류	사람이 상태에 위독한(√위독한 상태에, CVA, WO) 빠집니다.

- 담화 층위는 문장 단위를 넘어서 발생하는 오류를 다루는 영역이다. 담화 연구의 경우 그 범위가 넓고 어휘와 문법, 발음 영역에서 다양한 현상과 표지를 통해 나타나므로 체계화가 쉽지는 않다. 또한 구어 담화의 경우 문법 정보다는 발화 맥락 안에서 적절하고 효과적인 의미 전달에 초점이 주어지기 때문에 오류 판정 기준을 정하기도 쉽지 않다. 이러한 이유로 본 연구에서는 담화표지, 지시, 접속으로 비교적 표지가 분명하고 판정 기준이 명확한 항목만을 주석의 대상으로 포함시켰다.
- 본 사업은 문장 내에서의 오류 판단이 기본 원칙이기 때문에 문장 단위를 넘어서는 담화 오류는 최소한으로 제한하여 주석한다. 지시(DR)와 접속(DC)에 한해서 선행문과 후행문과의 의미적 연결을 고려해 오류 여부를 판단한다.

층위	주석 표지	포함 범위	예시
----	-------	-------	----

층위		주석 표지	포함 범위	예시
담화	지시	DR	부적절한 지시사의 선택으로 선행문과 후행문의 관계를 결속성 있게 나타내지 못한 경우	나는 롯데월드 아이스링크에 자주 가요. 여기(✓거기, CNP, REP, DR)에 가면 스트레스가 풀려요.
	접속	DC	선행문과 후행문의 의미 관계를 나타내는 데에 부적절한 접속 부사 및 접속 표지를 사용한 경우	나는 이런 남자를 만나면 경혼하고 싶습니다. 그래서(✓그러면, CMAJ, REP, DC) 기분이 좋을 거예요.
	담화 표지	DM	부적절한 담화 표지를 선택하거나 누락한 경우	우리 하숙집에서 현대백화점까지 그냥(CMAG, ADD, DM) 10분쯤 걸렸어요..
	구어/문어	DS	구어체/문어체, 격식체/비격식체의 혼용으로 인한 오류	근데(✓그런데, CMAJ, REP, DS) 어떤 사람을 평가할 때 외모만 보면 그거도 안 된다..

3. 범주별 세부 오류 유형의 처리 예시

1) 분석 여부

- ‘분석 여부’는 오류를 특정 범주의 오류로 판정 가능한지를 파악하는 것을 말한다. 부적절한 표현이 연속되거나 문장 구조의 이상으로 학습자의 표현 의도를 파악하기 어려운 경우 ‘분석 불가능(IMP)’로 판정할 수 있다.

<예> 나도 한번도 많고, 기 사람하고, 밥그릇, 노래했어요.(√/IMP)
 난 졸업만 뜬다면 드디오 내가 기다리는 시간이라고 생각하고 귀가 아주 밝다(√/IMP).

- 분석 여부는 기본 값이 ‘분석 가능’으로 설정되어 있으므로, 오류임에 분명하지만 교정어절을 주기 어려워 오류의 판정이 불가능한 경우에만 주석을 한다.

2) 오류 위치

(1) 실질어휘

① 고유명사(CNNP)

- 고유명사의 형태, 의미, 사용 오류를 말한다.

<예> 중구(√/중국) 요리를 맛있었습니다.
 하지만 독요(√/도쿄) 쇼핑은 조금 비싸요.

② 일반명사(CNNG)

- 일반명사의 형태, 의미, 사용 오류를 말한다.

<예> 애기와(✓아기와) 노인들한테 건강이 나빠졌다.
현대(✓현재) 세계적으로 환경문제가 대두되고 있다.

③ 의존명사(CNNB)

- 의존명사의 형태, 의미, 사용 오류를 말한다.

<예> 희망 10명(✓년) 후에 자기 가 수 있다.
내가 한국에 온 지 7개월(✓개월)이 되었다.

④ 대명사(CNP)

- 대명사의 형태, 의미, 사용 오류를 말한다.

<예> 내(✓우리) 아버지가 남편하고 친하게 되면 좋겠다.
저(✓나)는 유학생으로 온 외국인이다.

⑤ 수사(CNR)

- 수사의 형태, 의미, 사용 오류를 말한다.

<예> 오후 일곱(✓일곱) 시에 홍콩 친구하고 저녁 식사를 했어요.
요리를 배우기가 열(✓십) 년 전이 시작했습니다.

⑥ 동사(CVV)

- 동사의 형태, 의미, 사용 오류를 말한다.

<예> 그런데, 기숙사에서 술을 미실(✓마시는) 것 안 된다.
그것을 어쩔 수 없는 것이고 누군가가 그 변화를 세우는(✓
멈추는) 것이 못한다.

⑦ 형용사(CVA)

- 형용사의 형태, 의미, 사용 오류를 말한다.

<예> 불고기 먹기 때문에 기분이 기쁩니다(√좋습니다).
인심 약박한(√야박한) 시대 속에서 법이 사람의 부합리적인
행동을 제약할 수 있는 효과적인 방법이다.

⑧ 보조용언(CVX)

- 보조용언의 형태, 의미, 사용 오류를 말한다.

<예> 태하교에 가고 싶습니다(√싶습니다).
인터넷 쇼핑을 하고 싶으면 조심한다(√조심해야 한다).

- 오류 주석 시, 보조용언과 결합된 구성이 표현 문형 목록에 있을 경우에는
표현 문형(PE)위치로도 중복 주석한다.

<예> 세대 차이를 극복하기 위해서 신세대와 가상세대는 자주 이
야기를 해 뉘야 뉘다(√된다)
☞ ‘뉘다’는 보조용언 ‘된다’를 잘못 사용한 것이기 때문에 [오
류 위치-보조용언]으로 처리하는 동시에, ‘-어/아야 되다’가
표현 문형의 목록에 있기 때문에 표현 문형 위치도 중복 주
석한다. 즉, [오류 위치-표현 문형, 보조용언], [오류 양상-
오형태] 오류로 처리한다.

⑨ 지정사(CVC)

- 지정사의 형태, 의미, 사용 오류를 말한다.

<예> 학생예요(√학생이예요)
☞ 지정사 ‘이예요’와 ‘예요’를 잘못 사용한 것이기 때문에 지
정사 오류로 처리한다.

- 지정사와 연결어미/종결어미가 결합할 때, 지정사를 누락시키거나 첨가했을 경우는 지정사 누락, 첨가 오류로 처리한다. 단, 축약을 잘못된 경우는 오철자 오류로 처리한다.

<예> 다른 사람들에게 도와주고 싶기 때문에 한국어를 열심히 공부할 것이다(√것이다).

☞ 지정사 ‘이다’가 생략되었으므로 [오류 위치-지정사], [오류 양상-누락]으로 처리한다.

남자이에요(√남자예요)

☞ 받침이 없는 명사 뒤에서 ‘이에요’를 ‘예요’로 줄여서 쓰는 것이 일반적이거나, 필수적인 표준 규범은 아니므로 오류로 처리하지 않는다.

학생예요(√학생이에요)

☞ 받침 유무에 따라 ‘이에요’와 ‘예요’를 선택하여 사용하나, 이 경우 학습자가 ‘이에요’와 ‘예요’를 잘못 선택해서 사용한 것인지, 지정사 ‘이’를 누락한 채 종결어미를 잘못 쓴 것인지 판단이 어렵다. 본 연구에서는 지정사의 경우, 써야할 자리에 쓰지 않은 경우는 누락으로 보고, ‘예요’와 ‘예요’는 종결어미의 오철자 오류로 처리한다. [오류 위치-지정사], [오류 양상-누락], [오류 위치-종결어미], [오류 양상-오형태] 오류로 주석한다.

학생예요(√학생이에요)

☞ [오류 위치-지정사], [오류 양상-누락]

학생이에요(√학생이에요)

☞ [오류 위치-종결어미], [오류 양상-오형태]

아니예요(√아니예요)

☞ [오류 위치-종결어미], [오류 양상-오형태]

학생이여서(√학생이어서)

☞ [오류 위치-연결어미], [오류 양상-오형태]

학생이었어요(√학생이었어요)

☞ [오류 위치-선어말어미], [오류 양상-오형태]

공부를 할 거예요(✓거예요)

☞ [오류 위치-지정사], [오류 양상-누락]

☞ 지정사와 관련된 오류에서 지정사를 쓰고 어미와 축약하지 않거나 잘못 축약을 시킨 경우는 오철자 오류로 처리한다.
오류 위치는 오류가 발생한 위치에 따라 주석한다.

- 문어에서 지정사를 축약해서 사용한 경우에는 오류로 보기 어려운 측면이 있으나 모어 화자의 보편적인 언어 사용 방식에 있어서 어색한 것으로 보고 지정사 오류로 처리한다. 이때에는 오류 양상은 주석하지 않고, [오류 위치-지정사], [오류 층위-문어/구어] 오류로 처리한다.

<예> 여기는 우리 학콘데(✓학교인데) 정말 아름답다.

☞ 문어(격식체)에서 축약형으로 사용하는 것은 어색하기 때문에 [문어/구어] 오류로 처리한다. 단, 이때에는 오류 양상은 주석하지 않는다.

⑩ 관형사(CMM)

- 관형사의 형태, 의미, 사용 오류를 말한다.

<예> 그렇게 되면 어느(✓어떤) 사람은 돈이나 개인 정보를 잃어버릴 수도 있다.

두(✓이) 년 한국에 있을 겁니다.

⑪ 일반부사(CMAG)

- 부사의 형태, 의미, 사용 오류를 말한다.

<예> 내 남편은 꼭(√정말) 멋있게 생겼다.
어히려(√오히려) 남정보다 여성의 힘이 더 강하는 경우도
있는 정도다.

⑫ 접속부사(CMAJ)

- 접속부사의 형태, 의미, 사용 오류를 말한다.

<예> 그래서(√그래서) 피자하고 맥주를 먹고 많이 얘기했다.
그러니까(√그래서) 안목이 높아지거니와 다양한 문화의 향
유하고 새로운 것들을 깨닫기도 한다.

⑬ 감탄사(CIC)

- 감탄사의 형태, 의미, 사용 오류를 말한다.

<예> 아침 6시에 일어나서 하는 출근 준비, 이제 안녕(√안녕)~
응(√네). 선생님.

⑭ 접두사(CXPN)

- 접두사의 형태, 의미, 사용 오류를 말한다.

<예> 인심 약박한 시대 속에서 법이 사람의 부합리적인(√불합리
적인) 행동을 제약할 수 있는 효과적인 방법이다.
산세대(√신세대) 사람들이 부모님 입장에 많이 생학하고 부
모님도 신세대 입장 색학하면 세대 차이를 줄일 수 있다.

⑮ 명사파생접미사(CXSN)

- 명사파생접미사의 형태, 의미, 사용 오류를 말한다.

<예> 저의 첫 번째(√번째) 고민은 어떻게 시간을 잘 지킨 좋은
습관은 될 수 있는 것이다.
여성들의 사회 진출에 따라서 이혼률(√이혼율)이 높아진 것
이 큰 원인이라고 할 수 있다.

⑩ 동사파생접미사(CXSV)

- 동사파생접미사의 형태, 의미, 사용 오류를 말한다.

<예> 제 한국어를 좋아합니다(✓좋아합니다).
이에 따라서 노동사의 인권이 보장하게(✓보장되게) 되어 안
정한 생활을 할 수 있게 되었다.

⑪ 형용사파생접미사(CXSA)

- 형용사파생접미사의 형태, 의미, 사용 오류를 말한다.

<예> 고독스러운(✓고독한) 중학생
이러한 사회에서 자신이 하고 싶은 직업을 할 수 있으면 너
무 행복안(✓행복한) 것이다.

⑫ 어근(CXR)

- 어근의 형태, 의미, 사용 오류를 말한다.

<예> 주말에 날씨가 따뽕(✓따뜻)하니까 산을 갔어요.
밤에도 시원(✓시원한) 옷을 준비하세요.

(2) 기능어휘

① 주격조사(FNP)

- 주격조사의 형태, 의미, 사용 오류를 말한다.

<예> 책가(✓책이) 재미있어요.
그리고 우리 친구와 만나고 같이 시장에 고기와 사과와 오
렌지가(✓오렌지를) 샀어요.

- ‘나는’과 ‘내가’가 상호 교정 어절이 될 때에는 조사 오류로 한 번만 처리한
다. 즉, 주격조사나 보조사의 대치로 인하여 대명사의 형태가 바뀌는 경우에
는 오류로 처리하지 않고 교정 어절만 써준다.

<예> 그래서 제가(√나는) 지금 열심히 공부하고 있다.
 ☞ [오류 위치-주격조사], [오류 양상-대치] 오류로 처리한다.

저는(√제가) 공부할 때도 일할 때도 늘 새로운 아이디어를 가지고 있는 것을 보면 친구와 동료는 저를 창의적이라고 많이 하였습니다.
 ☞ [오류 위치-보조사], [오류 양상-대치] 오류로 처리한다.

② 관형격조사(FGP)

- 관형격조사의 형태, 의미, 사용 오류를 말한다.

<예> 10년 후의(√에) 아버지 같은 성공한 사람이 되고 싶다.
 기속사의(√기속사에) 규칙을 있다.

③ 목적격조사(FOP)

- 목적격조사의 형태, 의미, 사용 오류를 말한다.

<예> 집에서 포도를(√포도를) 먹었습니다.
 미래에 나를 사랑하는 남편하고 귀여운 아기를(√아기가) 있으면 좋겠다.

④ 부사격조사(FAP)

- 부사격조사의 형태, 의미, 사용 오류를 말한다.

<예> 8급에(√의) 학생들은 쉬는 시간에 학교 근처 area에 가지 안됐다.
 미국에(√에서) 영어 제일 중요하다.

⑤ 접속조사(FJC)

- 접속조사의 형태, 의미, 사용 오류를 말한다.

<예> 미국고(√과) 일본이 두 나라에 가고 싶다.

외모과(√와) 노력이 다 중요하다.

⑥ 호격조사(FVP)

- 호격조사의 형태, 의미, 사용 오류를 말한다.

<예> 친구아(√친구야), 어 내가 영화를 보고 싶은데.
철수아(√철수야), 어디 가니.

⑦ 인용격조사(FQP)

- 인용격조사의 형태, 의미, 사용 오류를 말한다.

<예> 선생님이 내가 읽은 책이 봐서 나한테 "수업 후에 사무실에
와요"이라고(√라고) 말했다.
나는 원래 "생선을 먹었어요"(√라고) 말했어야 했는데 김장
해서 "선생을 먹었어요"라고 말했다.

⑧ 보격조사(FCP)

- 보격조사의 형태, 의미, 사용 오류를 말한다.

<예> 10년 후에 30살(√이) 될 것이다.
이것은 제일 큰 문제이(√문제가) 되는 이유가 무엇일까?

⑨ 보조사(FXP)

- 보조사의 형태, 의미, 사용 오류를 말한다.

<예> 네 번째 부모님은 내가 좋은 미래는(√미래를) 기대하고 있
다.
나는(√내가) 고등학교 때 우리 엄마가 고등학교 교장이었다.

⑩ 연결어미(FED)

- 연결어미의 형태, 의미, 사용과 관련된 오류를 말한다.

<예> 그리고 우리 친구와 만나고(√만나서) 같이 시장에 고기와 사과와 오렌지가 샀어요.
나는 시계를 보면(√보고) 잠깐 놀랐다.

- 용언의 받침 유무에 따라 어미의 이형태 선택이 달라지는 경우는 연결어미의 활용 오류로 처리한다.

<예> 이렇게 살으면(√살면) 정말 행복할 수 있다.
그리고 학국 음식을 먹려고(√먹으려고) 해요.

⑪ 종결어미(FFE)

- 종결어미의 형태, 의미, 사용과 관련된 오류를 말한다.
- 종결어미를 활용하지 않고 기본형을 사용한 경우(-ㄴ다/다)는 [오류 양상-오형태], [오류 층위-활용] 오류로 주석한다.
- 종결어미 이형태 활용 오류는 [오류 양상-오형태], [오류 층위-활용] 오류로 주석한다.

<예> 그때부터 시간을 지키다(√지킨다).
그래서 일본에서 웃어른은 노약자석에서 꼭 앉는다(√앉는다).

⑫ 선어말어미(FPE)

- 선어말어미의 형태, 의미, 사용과 관련된 오류를 말한다.

<예> 인종 차별 때문에 많은 사람들이 죽었으니까 앞으로 인종 차별이 없었으면 좋게다(√좋겠다).
그래서 너무 배가 고파다(√고팠다).

⑬ 명사형 전성어미(FNE)

- 명사형 전성어미의 형태, 의미, 사용과 관련된 오류를 말한다.

<예> 과학자들의 의식에 의하면 아이들은 어른들보다 배우기(√배

우는) 능력이 6배 뛰어나다고 한다.

언어를 배울 때는 언어만 말고 그 나라의 문화도 공부하기
(√공부하는 것) 중요하다.

⑭ 관형사형 전성어미(FAE)

- 관형사형 전성어미의 형태, 의미, 사용과 관련된 오류를 말한다.
- 관형사형 전성어미는 시제와도 관련되므로 시제와 관련한 오류의 경우에는 오류 층위에서 ‘시제(ST)’ 오류도 함께 주석한다.

<예> 과식이나 하지 말고 여러 가지 음식을 골고루 먹으(√먹는)
것이 중요해요.

☞ 관형사형 전성어미 대치 오류로 처리한다.

한국에 온(√오는) 비행기에서 친구를 만났어요.

☞ 관형사형 전성어미 대치 오류로 처리하며, 시제와 관련되어
오류 층위에 시제(ST)도 주석한다.

(3) 구 단위 표현(PHE)

- 구 단위 표현은 어절 단위로 이루어진 표현을 잘못 사용한 경우를 말한다.
교정 어절 주석은 구 단위 표현 단위를 묶어서 처리한다.

<예> 남자의 아내가 더 이상 돈 없는 힘든 인생을 살고 싶지 않아
오랜 고민과 망설임을 한 나머지(√ 끝예) 남편과 5살 어린 아
이를 두고 더 좋은 인생을 찾으러 다른 도시로 이사를 간다.

- 구 단위 표현은 오류의 교정이 어절을 넘어서는 구 단위로만 처리해야 할
때 주석한다. 따라서 형태소 차원에서 교정이 가능한 경우는 형태소 단위를
오류 위치로 주석한다..

<예> 이상한 날씨로 악화가 나타났다(√악화가 되었다).

☞ ‘악화가 나타났다’라는 구 단위 오류로 처리할 것인지, ‘나타

나다’ 동사의 대치 오류로만 처리할 것인지 문제가 될 수 있다. 이 경우 가능한 분리하여 ‘나타났다’를 ‘되다’로 교정하여 [오류 위치-동사], [오류 양상-대치]로 처리한다.

- 본 연구에서는 ‘연어 오류’를 별도로 설정하지 않았기 때문에 구 단위 표현에 연어 오류가 있을 경우, 별도 처리하지 않고 동사 대치로 처리한다(본 사업은 연어 오류를 주석하지 않는다).

<예> 태도를 키워야 한다 (✓ 길러야 한다)

☞ ‘태도를 키워야 한다’를 ‘태도를 길러야 한다’로 교정할 때, 이를 연어 오류로 볼 수 있다. 그러나 구 단위 또는 연어 표현에서 용인 가능성의 기준이 판단자마다 다를 수 있다는 문제가 있다. 따라서 연어 오류를 주석하기 위해서는 연어 목록이 선행되어야 하고, 연어에 대한 판단이 정해져야 하므로 본 연구에서는 연어 오류를 별도로 주석하지 않고 동사 대치 오류로 처리한다.

이사를 옮기다(✓ 이사를 가다)

☞ ‘이사를 옮기다’의 경우, ‘짐을 옮기다’, ‘이사를 가다’ 2가지로 교정이 가능하다. 본 연구에서는 동사 교정을 우선으로 하여 [오류 위치-동사], [오류 양상-대치]로 주석한다.

(4) 표현 문형(PE)

- 보조 용언 구성, 여러 요소의 결합 구성으로 이루어진 표현 문형을 잘못 사용한 경우를 말한다.
- 표현 문형의 목록은 관점에 따라 상이할 수 있으므로, 외국인을 위한 <한국어 문법 사전>(국립국어원, 2005)에 표현으로 제시된 항목 중 두 개 이상의 요소로 이루어진 결합 구성에 한정하여 처리한다.³⁹⁾

39) 표현 문형 목록 기준을 모든 표현 문형의 합집합으로 할 경우, 다양한 형태들이 표현 문형 안으로 들어오기 때문에 본 연구에서는 어느 정도 정제된 목록으로서 <한국어 문법 사전>(국립국어원, 2005)을 기준으로 정한다. 표현 문형 목록은 <부록>으로 첨부하였다.

- 오류 주석은 형태 단위를 기본으로 하기 때문에 표현 문형 오류의 경우, 형태 단위로도 분석하여 오류 위치를 주석하는 동시에 표현 문형 오류로도 중복 주석한다.

<예> 왜냐하면 환경오염이 심해지면 건강이 나빠지기 십상이다(√ 십상이기 때문이다).

☞ 앞에서 ‘왜냐하면’을 사용했기 때문에 서술어에서 ‘-기 때문이다’를 호응해서 사용해야 한다. 이러한 경우, ‘-기 때문’에 해당하는 각각의 형태소인 명사형 전성어미와 의존명사의 누락 오류로 처리하는 동시에, ‘-기 때문’이 표현 문형 목록에 있기 때문에 표현 문형도 오류 위치에 중복 주석한다. 따라서 [오류 위치-명사형 전성어미, 의존명사/표현 문형], [오류 양상-누락]으로 주석한다.

3) 오류 양상

(1) 누락(OM)

- [정의] 누락 오류는 완전한 문장 또는 발화에서 나타나야 할 형태가 빠져 있는 경우를 말한다.
- [주석 방식] 누락 오류의 경우, 오류 위치는 교정 어절이 기준이 된다. 따라서 교정 어절을 입력하고, 누락된 품사(교정 어절)를 오류 위치로 주석한 후, 오류 양상을 누락(OM)으로 주석한다.
- [처리 기준] 누락 오류는 조사나 어미 누락을 우선적으로 주석한다.

<예> 돈이 많은 사람들(√은) 투자 할 수도 있어요.

☞ 문어에서 조사의 생략은 모어 화자의 언어생활에서도 일반적이지 않다. 따라서 조사나 어미 누락을 중심으로 누락 오류를 판단하며, 이때에는 누락된 보조사 ‘은’을 오류 위치로 주석한다. [오류 위치-보조사], [오류 양상-누락(OM)]으로 주석한다.

- 누락 오류는 필수적인 성분이 생략됐을 경우에만 처리한다. 따라서 필수적인 성분이 아닌 수의적이거나 없는 정보를 더 추가해주지 않도록 한다..

<예> 저는 여러 까지 능력서가 취득하지만 그 중에서 영어를 (√ 가장, OM?) 능숙하는 정도입니다.
 ※ ‘그 중에서 영어를 가장 능숙한 편입니다’라고 교정하여, ‘가장’이라는 부사를 첨가하고, ‘능숙한 편입니다’라고 교정할 수 있는가?
 ☞ 필수적인 성분이 아닌 것을 추가하여 [누락] 오류로 처리해서는 안 된다.
 ☞ 최소한의 교정 원칙에 따라 ‘능숙한 정도입니다’를 교정 어절로 삼는다. 즉, 수의적인 것은 [누락]으로 처리하지 않고, 필수적인 성분이 생략됐을 경우에만 [누락] 오류로 처리한다.

- 누락 오류는 하나의 유의미한 교정 어절에만 누락 오류로 주석하고, 뒤 따라 오는 요소들은 누락 오류로 처리하지 않는다. 즉, 용언이나 체언(실질어휘)의 누락으로 인해서 뒤따라오는 어미나 조사(기능어휘)는 교정어절만 써주고 누락 오류로 주석하지 않도록 주의한다.

<예> 우리는 함께 (√있을) 때 좋은 기본이 왔는데요.
 ☞ ‘있을’이 누락되었는데, 형태주석을 기본단위로 오류주석을 할 때, 동사 ‘있’과 관형사형 전성어미 ‘을’을 각각 누락오류로 처리할 수 있다. 그러나 동사 뒤에 오는 관형사형 어미는 동사에 의해 따라오는 것으로 판단하여, 동사 ‘있’ 하나만을 누락 오류로 주석한다. 즉, [오류 위치-동사], [오류 양상-누락]으로 주석한다.

- [주석 예시] 누락 오류의 오류 위치별 일부 예시는 다음과 같다.
 ○ [명사 누락]은 다음과 같은 것들이 해당된다.

<예> (√환경을) 개발한 탓에 산과 나무 점점 없졌다.

☞ 문장의 필수 성분인 목적어가 누락된 명사 누락 오류로 처리한다. 이때 ‘환경’으로 인해 따라오는 ‘을’은 누락 오류로 주석하지 않고 교정어절만 써주도록 한다. [오류 위치-명사], [오류 양상-누락] 오류로 주석한다.

○ [조사 누락]은 다음과 같은 것들이 해당된다.

<예> 그로 인해 평소 일상생활에서 말할 수 없는 말 욕하는 말 등 비우는 말 등(✓을) 흔히 볼 수 있다.
 ☞ ‘등’을 써줬기 때문에 목적격 조사 ‘을’을 사용하지 않아도 된다고 용인할 수도 있으나, 조사 누락은 엄격하게 적용하여 누락 오류로 처리한다. 이 경우는 ‘등’ 맨 마지막에만 [오류 위치-목적격조사], [오류 양상-누락] 오류로 처리한다.

○ [관형사형 전성어미 누락]은 다음과 같은 것들이 해당된다.

<예> 10년 후에 내가 가(✓갈) 수 있다.
 ☞ 관형사형 전성어미를 누락한 오류로, 관형사형 전성어미 ‘-(으)ㄹ’의 [누락]으로 주석한다. 경우에 따라서 받침을 제대로 쓰지 못한 오철자 오류로 볼 수도 있으나, 주석의 일관성을 위해 누락 오류로 주석한다.

※ 주의: 한 단어에서 단순 철자(음소)가 누락된 경우는 오철자 오류로 ‘누락’이 아닌 ‘오형태’로 처리한다.

<예> 사라(✓사람)마다 님비 현상이 다 있을 것이다.
 ☞ 사람의 종성 ‘ㅁ’이 누락되었으나, 문어에서 단어 내에서의 음소 생략은 오형태 오류로 처리한다.
 아침에 친구를 만나서(✓만나서) 혼자 청소합니다.

☞ 동사 ‘만나서’에서 받침 ‘ㄴ’이 생략된 형태는 누락이 아닌 철자의 오류로 처리하여 오형태 오류로 처리한다.

(2) 첨가(ADD)

- [정의] 첨가 오류는 완전한 발화에서 나타나지 말아야 할 형태가 쓰인 경우나 중복된 형태를 사용한 경우를 말한다.
- [주석 방식] 첨가된 부분은 해당 부분의 교정 어절 없이 첨가된 위치를 오류 위치로 주석하고, 오류 양상을 첨가(ADD)로 주석한다.

<예> 종일에(√종일) 반 친구와 나는 만나서 언제나 재미있는 시간을 하고 있어요.
 ☞ 부사 ‘종일’에 불필요하게 부사격 조사 ‘에’를 첨가한 오류로, [오류 위치-부사격 조사], [오류 양상-첨가(ADD)]로 주석한다.

- [처리 기준] 첨가 오류는 하나의 유의미한 교정 어절에만 주석하고, 첨가된 요소로 인해 뒤 따라 오는 요소들은 첨가 오류로 처리하지 않는다. 즉, 용언이나 체언(실질어휘)의 첨가로 인해서 뒤따라오는 어미나 조사(기능어휘)는 첨가 오류로 주석하지 않도록 주의한다. 이때 시스템상에서의 처리 방식은 조사나 어미로 분석된 형태소를 체언 또는 용언에 결합시켜 하나의 첨가 오류로 주석한다.

<예> 머지않은 장래에 장래에(√첨가) 인간 복제도 가능하게 될 것이다.
 ☞ 문어에서 ‘장래에’를 두 번 중복하여 썼기 때문에 두 번째 ‘장래에’를 첨가 오류로 주석한다. 이때 부사격 조사 ‘에’는 명사 ‘장래’로 인해 따라온 요소로서 ‘장래’와 ‘에’ 각각을 첨가 오류로 주석하지 않고, 명사 첨가 오류로만 주석한다. 따라서 [오류 위치-명사], [오류 양상-첨가]로 주석한다.

- [처리 예시] 첨가 오류의 오류 위치별 일부 예시는 다음과 같다.
- [조사 첨가]는 다음과 같은 것들이 해당된다.

<예> 대만의 수도가(√수도) 타이베이는
 주격조사 ‘가’가 첨가된 오류로 주석한다.
 ※참고: 이때 ‘수도가’를 ‘수도인’으로 교정할 경우에는, 주격조사 ‘가’와 서술격조사 ‘이’의 대치 오류로도 볼 수 있으나 본 연구에서는 ‘최소 수정 원칙’으로 주격조사 첨가 오류로 처리한다.

○ [표현 문형 첨가]는 다음과 같은 것들이 해당된다.

<예> 저는 <name> 한국어센터에서 공부하고 있는 동안 선생님한테서 도움이 많이 받아 공부하고 있는(√공부하는) 시간이 아주 즐거웠습니다.
 ‘-고 있’이 첨가된 오류로 연결어미와 보조용언 각각을 오류 위치로 주석하며, ‘-고 있다’는 표현 문형 목록에도 있기 때문에 표현 문형도 중복 주석한다. 따라서 [오류 위치-연결어미, 보조용언/표현 문형], [오류 양상-첨가] 오류로 주석한다.

(3) 대치(REP)

- [정의] 대치 오류는 의미적 오류로 서로 다른 의미의 어휘를 바꾸어 쓴 경우를 말한다. 즉, 학습자가 어휘의 의미나 용법을 잘못 이해하여 다른 어휘를 사용한 경우이다.
- [주석 방식] 대치되어야 할 형태소(품사)를 오류 위치로 주석하고, 오류 양상을 대치(REP)로 주석한다.

<예> 직접 비판을 받을 때보다 상처가 더 많은(√큰) 것이다.
 맥락상 ‘상처가 많다’보다 ‘상처가 크다’가 더 적절한 표현으로, 형용사 ‘많다’와 ‘크다’의 대치 오류로 처리한다. 따라서 [오류 위치-형용사], [오류 양상-대치(REP)]로 주석한다.
 전통의 아름다움이 사람들에게 알려주는 것도 전통을 보존하

려고(√보존하려면) 해야 할 일이다.

☞ 연결어미 ‘려고’와 ‘려면’의 대치 오류로, [오류 위치-연결어미], [오류 양상-대치(REP)]로 주석한다.

- [처리 기준] 대치 오류는 한국어에 없는 표현이나 학습자의 모국어를 사용한 경우도 포함한다.

<예> 그런데 요즘 부모님들이 자식이 2살부터 play group(√유치원)에 보내는데 놀면서 유치원에 입학하기 위해 준비한다.

☞ 영어 단어를 그대로 사용한 경우, 오류 위치를 해당 품사로, 오류 양상을 대치 오류로 주석한다. 즉, [오류 위치-명사], [오류 양상-대치]로 주석한다.

- 일상적인 언어생활에서 보편적으로 쓰지 않는 것으로 판단되는 외국어를 사용한 경우도 대치 오류에 포함된다. 외래어인지 외국어인지 판단하기 어려운 경우는 <표준국어대사전> 등재 여부를 참고하여 판단한다.

<예> 이메일 에드레스(√주소) 다 있어요.

☞ ‘이메일’은 <표준>에도 등재되어 있고 일상적으로도 많이 쓰이는 어휘이므로 오류로 처리하지 않으나, ‘에드레스’는 <표준>에 등재되어 있지만 전산 분야와 같은 특수 분야에서 한정적으로 사용되는 의미로 등재되어 있어 ‘외국어’ 사용으로 간주하여 대치 오류로 주석한다.

- 피동과 사동은 한 단위로 보고 대치 오류로 처리한다. 즉, ‘-어지다’, ‘-이/히/리/기/우/구/추(접사)’, 사동 표현 ‘-게 하다’, 피동 표현 ‘-게 되다’ 등은 대치 오류로 처리한다.

<예> 누군가 돈이 없다면 행복할 수 없다고 생각했는데 한편에 일부 사람은 가족이 가장 중용 생각했는데 그 이유 점은 돈으로 바뀔(√바꿀) 수 없다고 지적을 했다.

☞ ‘바꾸다’를 써야 하는 자리에 피동사 ‘바뀌다’를 사용하였기 때문에 [오류 위치-동사], [오류 양상-대치], [오류 층위-피동] 오류로 주석한다.

- 대치 오류의 경우 대치된 요소로 인해 뒤 따라 바뀌는 요소들은 대치 오류로 처리하지 않는다. 즉, 용언이나 체언(실질어휘)의 대치로 인해서 바뀌는 어미나 조사(기능어휘)는 교정어절만 써주고 대치 오류로 주석하지 않도록 주의한다.

<예> 그들은 환경문제보다 자기가(√자신의) 먹고 살기가 더 중요하다고 생각한다.

☞ 대명사 ‘자기’보다 ‘자신’이 더 자연스럽다. 따라서 대명사의 [대치] 오류로 처리한다. 단, 주격조사 ‘가’도 관형격 조사 ‘의’로 대치되지만, 이는 앞의 대명사 대치로 인한 것이기 때문에 뒤의 조사의 경우는 대치 오류로 주석하지 않고 교정어절만 써주도록 한다.

- [처리 예시] 대치 오류의 오류 위치별 일부 예시는 다음과 같다.
- [조사 대치]는 다음과 같은 것들이 해당된다.

<예> 한국 영화를(√영화는) 재미있습니다.

☞ 보조사 ‘는’ 자리에 목적격 조사 ‘를’을 잘못 사용한 경우이므로 [오류 위치-목적격 조사], [오류 양상-대치] 오류로 주석한다.

- [연결어미 대치]는 다음과 같은 것들이 해당된다.

<예> 저는 지난 주말에 날씨가 좋니까(√좋아서) 기숙사 친구하고 같이 한강공원에 갔습니다.

☞ ‘-아서/어서’를 사용해야 할 곳에 ‘-니까’를 잘못 사용하였으므로 [오류 위치-연결어미], [오류 양상-대치, 오형태(‘-니

까'의 활용형도 잘못 사용)], [오류 층위-활용] 오류로 주석한다.

※ 주의: 대치와 오형태 오류 판단 시, 의미적 대치인지, 오형태 오류인지 혼동되는 경우가 있다. 이때에는 문맥을 고려하여 오류를 판단하며, 학습자들이 얼마나 이러한 오류를 보일 수 있는가를 고려한다. 다시 말해서 수준별로 사용할 수 있는 어휘를 고려하여 오류를 처리한다. 아울러 최소 수정 원칙을 기본으로 하되, 용인 가능한 교정 어절로 수정하도록 한다.

○ 문맥에 따른 유추를 통한 대치 오류 판단의 예는 다음과 같다.

<예> 긴정한(√진정한) 아름다움이란 착한 말씀(√마음씨)이다.

☞ 문맥을 통해 '말씀'은 '마음씨'라고 유추해볼 수 있다. 이처럼 전체 맥락을 통해 '마음씨'라는 교정 어절을 추정을 해볼 수 있다면 [대치] 오류로 주석한다.

<예> 저녁에 커피를 마시면서 간간한(√간단한) 책을 읽고 싶다.

☞ '간간하다'라는 어휘가 존재하나, 학습자의 수준 및 의도를 고려했을 때, '간간하다'를 사용했다고 보기 어렵다. 따라서 '간간하다'와 '간단하다'의 어휘 대치 오류로 판단하지 않고, '간단하다'의 오형태(오철자) 오류로 주석한다. 아울러 '간단한' 책이라고 하면 엄밀한 의미에서 정확한 표현이 아니라고 판단될 수도 있으나, 본 연구에서는 최소 교정을 원칙으로 하며, '가벼운 또는 단순한'의 의미로 모어 화자들도 사용할 수 있는 표현으로 보고 이와 같은 경우 '간단한'의 오철자 오류, 즉[오형태] 오류로 주석한다.

(4) 오형태(MIF)

○ [정의] 오형태 오류는 어휘나 문법의 조합 양상과 활용 형태가 잘못된 형태로 제시된 경우를 말한다. 즉, 단어 내 도치나 이형태 사용 등을 사용한 경우와 의미적으로 전혀 관련이 없는 항목이 선택된 경우, 어휘나 문법을 사용함에 있어서 다른 어휘나 문법으로 대체하여 이해할 가능성이 없는 경우로 형태가 잘못 사용된 경우를 말한다.

- [주석 방식] 오형태 오류는 음소 단위 형태를 잘못 쓴 오철자 오류와 용언 활용, 조사 이형태 곡용, 어미 활용 등 형태를 잘못 활용한 경우를 포함한다. 따라서 오철자 및 잘못된 활용이 나타난 부분을 오류 위치로 주석하고, 오류 양상을 오형태(MIF)로 주석한다. 단, 오철자 오류는 오류 층위에 활용(MCJ)을 주석하지 않도록 주의하고, 조사 이형태를 잘못 사용한 경우는 오류 층위에서 굴절(곡용)(MDC)으로 주석하고, 용언의 규칙/불규칙 활용과 어미 활용을 잘못된 경우는 굴절(활용)(MCJ)으로 주석한다.

<예> 우리나라에서 과일들하고 야채들도 많아서 과일와(√과) 야채도 다른 나라에 팔아요.
 ☞ 접속조사의 이형태를 잘못 사용한 경우로, [오류 위치-접속조사], [오류 양상-오형태(MIF)], [오류 층위-굴절(곡용)(MDC)]로 주석한다.
 지금 친구 같이 등산에 가시다(√갑니다).
 ☞ 종결어미 ‘ㅂ니다’에 대한 철자 오류로, [오류 위치-종결어미], [오류 양상-오형태(MIF)] 오류로 주석한다.

- [처리 예시] 오형태 오류의 오류 위치별 일부 예시는 다음과 같다.
 ○ [명사 오형태]는 다음과 같은 것들이 해당된다.

<예> 우리 집에 물(√문)을 열리면 계단을 있다.
 ☞ 맥락상 ‘문’을 써야하는데 ‘물’을 쓴 경우, ‘물’이라는 단어가 존재하기 때문에 의미적인 단어와 단어 간의 대치로 볼 수 있으나, ‘문’과 유사한 형태를 잘못 쓴 오철자 오류로 판단한다. 따라서 [오류 위치-명사], [오류 양상-오형태] 오류로 주석한다.

- [조사 오형태]는 다음과 같은 것들이 해당된다.

<예> 20, 30대 남녀는 친구을(√를) 중요하게 생각하는 사람들이 많았다.
 ☞ 받침으로 끝날 때 목적격 조사 ‘를’을 써야하는데, ‘을’을

썼기 때문에 조사를 잘못 활용하여 쓴 것으로 [오류 위치-목적격조사], [오류 양상-오형태], [오류 층위-곡용] 오류로 주석한다.

○ [선어말어미 오형태]는 다음과 같은 것들이 해당된다.

<예> 제주 친구하고 옥등산에서 등산을 가세요(√갔어요).
 ☞ 과거 시제를 나타내는 선어말 어미 ‘-았-’이 생략된 형태이다. 그러나 ‘가어요’로 쓰지 않고 ‘가세요’로 썼기 때문에 이것은 과거를 인식하고 있다고 보고 오형태 오류로 처리한다. 즉, ‘해습니다’, ‘마셔지만’, ‘와지만’처럼 과거 시제 선어말 어미 ‘었’에서 ‘ㅅ’을 누락시킨 경우는 선어말 어미 오형태(오철자) 오류로 처리한다. [오류 위치-선어말어미], [오류 양상-오형태] 오류로 처리한다.

○ [연결어미 오형태]는 다음과 같은 것들이 해당된다.

<예> 갈 수 있는다면(√있다면) 언제까지도 기다린다"고 해서 희망자들이 속출하고 있다.
 ☞ ‘있다면’을 써야할 자리에 ‘있는다면’으로 연결어미를 잘못 활용하여 쓴 것이기 때문에 [오류 위치-연결어미], [오류 양상-오형태], [오류 층위-활용] 오류로 주석한다.

○ [종결어미 오형태]는 다음과 같은 것들이 해당된다.

<예> 그 이유는 제가 우라 아내보다 한국에 돈을 잘 못 벌읍니다(√법니다).
 ☞ 동사 ‘벌다’를 활용하여 ‘법니다’로 써야하는데 ‘벌’을 그대로 사용하고 있기 때문에 [오류 위치-동사, 종결어미], [오류 양상-오형태], [오류 층위-활용] 오류로 처리한다.
 아름답니다(√아름답습니다), 시끄럽니다(√시끄럽습니다)

재미있입니다(√재미있습니다), 맛있입니다(√맛있습니다)
 ☞ ‘ㅂ니다/습니다/니다/입니다’는 종결어미 오형태 활용 오류
 로 처리한다.

4) 오류 총위

(1) 발음

① 음소(PP)

- [정의] 음소 오류는 음소 단위에서 발화가 잘못 사용된 경우를 말한다.
- [주석 방식] 잘못 발음된 품사에 오류 위치를 주석하며, 오류 양상은 주석하지 않고, 오류 총위에 음소(PP)를 주석한다.
- [처리 기준] 단어 내에서 명확하게 다른 음소로 발음한 경우나 음소를 발음하지 못한 경우, 음소 오류로 주석한다.

<예> 팔이(√빨리), 말을::, 줄= 아:: 잘해서::,
 ☞ 자음 ‘ㅃ’을 ‘ㅍ’로 발음하여, 음소 오류로 주석한다. [오류
 위치-일반부사], [오류 양상-없음], [오류 총위-음소]
 예:: 코피(√커피)= 카페에서::, 예:: 공부를, 합니다::.
 ☞ 모음 ‘아’와 ‘오’를 교체하여 발음하므로 음소 오류로 주석
 한다. [오류 위치-명사], [오류 양상-없음], [오류 총위-음
 소]

부모니(√부모님)
 ☞ 한 단어 안에서 받침을 발음하지 못한 경우도 마찬가지로
 음소 오류로 주석한다. [오류 위치-명사], [오류 양상-없
 음], [오류 총위-음소]

② 음절(PS)

- [정의] 음절 오류는 음절 단위에서 발화가 잘못 사용된 경우를 말한다. 음절 오류는 원래 음절보다 적게 또는 더 많이 발화한 경우가 해당된다.

- [주석 방식] 음절 오류가 발생한 품사에 오류 위치를 주석하며, 오류 양상은 주석하지 않고, 오류 층위에 음절(PS)을 주석한다.
- [처리 기준] 원래의 음절수보다 늘어 발음하거나 축약이 불가능한 단어를 축약하여 발음한 경우 음절 오류로 주석한다.

<예> 예:: 매이루(√매일) 노무::, 즐겁..습니다::.

☞ 2음절인 ‘매일’에 모음을 삽입하여 3음절로 발음하고 있으므로 음절 오류로 주석한다. [오류 위치-일반부사], [오류 양상-없음], [오류 층위-음절]로 주석한다.

 제가, 한국에 와서.. 사 개워르(개월) 정도:: 살았습니다::

☞ 2음절인 ‘개월’을 받침 ‘ㄹ’과 모음 ‘으’를 결합하여 3음절로 발음하고 있으므로 음절 오류로 주석한다. [오류 위치-의존명사], [오류 양상-없음], [오류 층위-음절]로 주석한다.

③ 음운규칙(PC)

- [정의] 음운규칙 오류는 구어 발화에 나타난 필수적 음운 규칙의 일탈을 말한다. 유음화, 연음화하여 발음해야 하는데, 글자 그대로 절음화하여 발화한 경우가 해당된다.
- [주석 방식] 음운규칙을 적용하지 못한 품사를 오류 위치로 주석하며, 오류 양상은 주석하지 않고, 오류 층위에 음운규칙(PC)을 주석한다.
- [처리 기준] 음운규칙 오류는 음운규칙을 적용하지 않은 경우와 적용했으나 잘못 발음한 경우 두 가지로 나눌 수 있다.
- 첫째, 음운규칙 미적용 오류는 음운규칙으로 인해 철자와 다르게 발음해야 하나, 학습자가 철자대로 절음화 하여 발음한 경우다. 학습자가 철자대로 발음했기 때문에 원어절과 교정어절의 형태는 동일하다.

<예> 설날(√설날)

☞ 학습자가 발화 시 유음화 규칙에 따라 [설칼]로 발음하지 않고 [설]과 [날]을 각각 끊어서 개별 음절 발음에 충실하였다면 ‘음운규칙’ 오류로 처리한다. 이밖에 ‘육학년’을 [유칸년]으로 발음하지 않고 글자 그대로 [육학년]으로 발음한

경우나 ‘학교’를 [학교]로 발음하지 않고 글자 그대로 [학교]라고 발음한 경우를 음운규칙 오류로 처리한다.

무조건(√무조건) ➡ 경음화 미적용

같이(√같이) ➡ 구개음화 미적용

신라(√신라) ➡ 유음화 미적용

앞에(√앞에) ➡ 연음 미적용

먹는(√먹는) ➡ 비음화 미적용

➡ 위의 예들은 음운규칙을 적용하지 않고, 철자 그대로 발음한 경우다. 따라서 음운규칙이 적용되어야 하는 위치에 오류 위치를 주석하고, 오류 층위에는 음운규칙을 주석한다.

- 둘째, 음운규칙 미적용 외에 음운규칙을 적용시켜야 하는 단어에서 잘못 적용한 경우도 음운규칙 오류로 주석한다. 그러나 이때에는 음운규칙과 음소 오류를 중복 주석한다. 그러나 음운규칙을 적용시켜야 하는 단어이나 음운규칙과 상관없는 위치에서 다른 음소로 발음한 경우는 ‘음소’ 오류로만 주석한다.

<예> 학교(√학교)

➡ [학교]로 발음해야 하는데 [학교]로 발음했을 경우, 음운규칙과 음소 오류를 중복 주석한다. 그러나 음운규칙이 적용되지 않는 위치에서 다른 음소로 발음한 경우는 음소(PP) 오류로만 처리한다(예 핵교(√학교)).

- 구어의 특성이나 표현 의도에 의한 발음 특성은 오류로 처리하지 않는다.

<예> 표현 의도에 의한 수의적 경음화: 조금

구어에 의한 발음 특성: ~먹었구요 / ~했어여

④ 원어식 발음(PN)

- [정의] 원어식 발음은 학습자의 외국어 발음으로 인한 발화 오류를 말한다.

즉, 외국어나 외래어 발음 시, 원어에 가까운 소리로 발음하는 경우다. 이는 한국어 모어 화자에게서도 일어나는 현상이기는 하나 외국인 학습자에게서 그 빈도가 더 잦고, 발음 또한 모어 화자의 그것과 많이 다르다. 따라서 원어식 발음은 외래어 표기법과 불일치하므로, 이를 표시해 주는 차원에서 외국어와 외래어의 경우, 한국어와 다르게 발음했을 때 ‘원어식 발음’ 오류로 주석한다.

- [주석 방식] 원어식 발음으로 발음한 품사에 오류 위치를 주석하며, 오류 양상은 주석하지 않고, 오류 층위에 원어식 발음(PN)을 주석한다.
- [처리 기준] 외국어나 외래어에서 한국어의 표준 발음과 다르게 발음한 경우 원어식 발음 오류로 주석한다. 원어식 발음(PN)의 경우, 발음의 차이로 인해 한국어의 외래어 표기와 다르게 음절이 줄거나 늘어날 수 있다. 이때는 음절 오류가 아닌 원어식 발음 오류로 주석한다.

<예> 인텔뷰(√인터뷰)

세너(√센터)

☞ 한국어 외래어 표기법과 다르게 원어식 발음으로 발화한 경우 원어식 발음 오류로 처리한다.

팔너(√파트너)

그대 처음에 갈 뻔남(√베트남)에서

☞ 원어식 발음의 차이로 인해 한국어의 외래어 표기와 다르게 음절이 줄어들었다. 이때는 음절 오류가 아닌 원어식 발음 오류로 주석한다.

- 학습자 모국어의 외래어 발음도 포함한다.

<예> [요한스버그](√ 요하네스버그)

이.. 기자는:: 한국에서:: 이:: 마약, 없는::, 아:: [이미지])(√ 이미지) 좀:: 없어,지고:: 있습니다 지금.

☞ 외국어 발음 오류로 외래어 표기법과 ‘다르다’는 차원에서 [오류 양상-오형태] 오류로 처리한다.

⑤ 중간 발음(변이음포함) (PA)

- [정의] 중간 발음은 변이음을 포함하여 학습자의 외국어와 한국어의 중간 발음으로 인한 발화 오류를 말한다.
- [주석 방식] 변이음으로 발음한 품사에 오류 위치를 주석하며, 오류 양상은 주석하지 않고, 오류 층위에 중간 발음(PA)를 주석한다.
- [처리 기준] 중간발음은 한국어 모어 화자와는 다른 발음, 즉 변이음과 관련된 오류와 음소와 음소 간의 중간 발음 두 가지가 포함된다.
- 첫째, 음성과 관련된 변이음은 구어 전사 과정에서 유성음, 무성음으로 표기해준 경우에 근거해 변이음 오류로 처리한다. 이때는 음성과 관련된 문제로, 원어절과 교정어절의 형태는 동일하다. 다만, 변이음은 구어 전사에서 전사자 특성에 따라 다르게 식별될 수 있는 문제가 있다. 따라서 구어전사 과정에서 변이음이 분명하게 식별된 메모에 근거하여 주석하도록 한다.

<예> 가구(√가구)

- ☞ ‘가’의 ㄱ을 유성음으로 발음
- ‘구’의 ㄱ을 무성음으로 발음
- ‘구’에서 ㄱ과 ㄱ의 중간 발음

파란색(√파란색)

- ☞ ‘파’에서 f로 발음됨
- ☞ 구어 전사에서 위와 같이 기술한 메모에 근거해 중간 발음 (PA) 오류로 주석한다.

- 둘째, 학습자가 원어절과 교정어절 사이의 발음, 즉 음소 간의 중간 형태로 발음한 경우도 중간발음 오류로 처리한다.

<예> 여자가<note>‘여’를 ‘으’와 ‘유’의 중간 발음으로 발음</note>

- ☞ 구어 전사 시, 음소와 음소 간의 중간 발음으로 들릴 때 괄호 안에 (‘X’와 ‘Y’와 중간 발음)으로 표기한다. 이를 바탕으로 하여 중간 발음으로 표기된 경우, [오류 위치-명사], [오류 양상-없음], [오류 층위-중간 발음]으로 주석한다.

화반수(√과반수)

☞ ‘ㅎ’과 ‘ㄱ’의 중간 발음

☞ 구어 전사에서 위와 같이 ‘음소’와 ‘음소’의 중간 발음으로 기술한 메모에 근거해 중간 발음(PA) 오류로 주석한다.

(2) 형태

① 단어 형성[합성법](MCP)

- 단어 형성[합성법] 오류는 조어 과정에서 발생하는 오류를 말한다. 즉, 학습자가 존재하지 않는 어휘를 생산해 내는 오류가 포함된다.
- 학습자가 조어 과정에서 형태를 잘못 만들어 낸 경우, 오형태 오류로 볼 수도 있다. 그러나 오형태 오류가 오철자 오류와 이형태 활용 오류에 해당하는 오류 양상이라고 할 때, 합성과 파생 관련한 오류를 형태 오류에 포함시킬 수 있는가가 문제가 된다. 본 연구에서는 파생과 합성 오류의 경우 오형태 오류로 보기 어렵고, 오류 양상은 수의적 주석이므로 오류 양상을 필수적으로 주석하지 않고 오류 층위에서 파생과 합성만을 주석하도록 한다.
- 따라서 학습자가 생산해 낸 형태가 한국어에는 없는 합성어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 합성으로 주석한다.

<예> 해물 고기(√물고기)가 많 많았어요.

☞ ‘물고기’를 ‘해물’과 ‘고기’로 잘못 합성하였으므로 [오류 위치-명사], [오류 양상-없음], [오류 층위-합성법]으로 처리한다.

② 단어 형성[파생법](MDV)

- 단어 형성[파생법] 오류는 조어 과정에서 접사를 잘못 사용한 오류를 말한다.
- 학습자가 파생접사(유사파생접사 포함)를 사용하여 생산해 낸 형태가 한국어에는 없는 파생어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 파생

으로 주석한다.

- 단, 접사와 접사의 대치의 경우나 접사의 불필요한 첨가 또는 생략은 오류 양상에 대치, 첨가, 생략으로 주석한다.

<예> 친구와 그 사람을 사귀하면(√사귀면) 제일 좋은 일 그 사람이 멋있습니다.

☞ 동사 ‘사귀다’에 다시 동사파생접미사 ‘-하다’를 붙여 한국어에는 없는 형태를 만들어 낸 것으로 [오류 위치-동사], [오류 양상-없음], [오류 층위-파생법] 오류로 주석한다.

그 다음에 여름에는 수영을 하다든가 성풍기를 사용하다든가 해서 건강적인(√건강한) 감온 방법이 선택하면 좋다.

☞ 접사 ‘하다’ 대신 ‘적’을 사용해 형용사를 파생시킨 경우로, 이때에는 접사 ‘적’과 ‘하다’ 대치 오류로 주석한다. [오류 위치-접미사], [오류 양상-대치], [오류 층위-파생법]으로 주석한다.

이런 데다가 의사 선생님께 의하면 균형 깨진 영양성(√영양) 바람에 났던 여드름이 더 날 계속한다고 걱정했는데도 그렇지 않습니다.

☞ 접미사 ‘-성’을 과잉적용한 오류로 [오류 위치-접미사], [오류 양상-첨가], [오류 층위-파생법]으로 주석한다.

- 동사 어간에 ‘하다’를 붙여 한국어에는 없는 동사를 만들어냈을 경우는 파생 오류로 처리한다. 이는 합성 오류로도 볼 수 있지만 형태 주석에서 ‘-하다’를 파생접미사로 처리하고 있어 처리의 연계성과 일관성을 고려하여 파생 오류로 처리한다.

<예> 그날 수업 후 집에 도착하자마자 어머니가 나한테 혼했다(혼냈다).

☞ 이는 ‘혼내다’를 명사 ‘혼’에 ‘하다’를 붙여 ‘혼+하다’로 한국어에는 없는 어휘를 생산한 것이다. 이 경우 ‘하다’를 동

사로 볼 수도 있고 파생접사로도 볼 수 있다. 형태 주석에서는 이를 동사파생접미사로 주석하기 때문에 오류 주석에서도 파생 오류로 판단하도록 한다. [오류 위치-동사], [오류 양상-없음], [오류 층위-파생법]으로 처리한다.

- 접사는 문법범주에서 논의하는 존대, 피동/사동, 복수 표지 중 피동/사동만 대치 오류로 처리한다. 접사 중 문법적인 성격 강한 존대(님), 복수 표지(들) 등은 형태 주석에서 접사로 따로 분리하여 처리하기 때문에 형태 주석에서 분리하는 접사들은 ‘누락/첨가’로 처리하고, 피동/사동은 어휘 대치로 처리한다. 또한 어휘적 의미를 더해주는 접사의 경우, 예를 들어 ‘사과’를 ‘פות사과’로 썼을 때에는 형태 분석에서 ‘פות’을 분리하지 않기 때문에 어휘적 의미를 더해주는 접사가 덧붙여진 단어는 대치 오류로 처리한다.
- 그밖에 형태 주석에서 접사로 분리하지는 않지만 유사파생접사로 볼 수 있는 형태들을 사용하여 어휘를 파생시킨 경우는 오류 양상은 주석하지 않고, 오류 층위에 파생으로 주석한다.

<예> 그리고 나는 계속 매일 지각했을 때 나는 번금도 내고 반성서(√반성문)도 썼다.

☞ 반성서에서 ‘서’는 형태주석에서 분리하여 처리하지 않고, 반성서를 하나의 명사로 주석한다. 이를 오류 주석에서는 반성문을 한국어 어휘에는 없는 ‘반성+서’로 파생시킨 것으로 보고 [오류 위치-명사], [오류 양상-없음], [오류 층위-파생법]으로 주석한다.

행운하면 다음 학기는 <name>대학교 어학관(√어학원)에서 4급 공부할 거야.

☞ ‘관’과 ‘원’ 모두 형태 주석에서 접사로 따로 분리하여 처리하지 않는다. 그러나 이는 유사파생접사로 볼 수 있기 때문에 ‘어학원’ 명사를 잘못 파생시킨 오류로 보고 오류 층위에 파생법으로 주석한다.

③ 굴절[곡용](MDC)

- 굴절[곡용] 오류는 조사의 이형태를 잘못 사용한 경우를 말한다.
- 굴절[곡용] 오류의 오류 양상은 기본적으로 오형태 오류로 주석한다.

<예> 지금 가족가(√가족이) 너무 보고 싶습니다.
 ☞ 받침 뒤에서 주격조사 ‘가’로 잘못 사용하였으므로 [오류 위치-주격조사], [오류 양상-오형태], [오류 층위-곡용] 오류로 처리한다.

- 굴절[곡용] 오류 주석 시, 체언의 오류로 인한 조사 오류는 오류로 주석하지 않도록 주의한다. .

<예> 종일(√종이)을(√를) 낭비할 게 아니라 절약할 것이다.
 ☞ ‘종일을’은 ‘종이를’로 교정되나, 이때는 체언을 잘못 사용한 것으로 인해 목적격 조사 ‘을’을 썼다고 보고, ‘을’은 곡용 오류로 주석하지 않는다. 따라서 이는 명사 오형태 오류로만 주석하고, 목적격 조사 ‘을’에는 교정어절 ‘를’만 써준다.

④ 굴절[활용](MCJ)

- [정의] 굴절[활용] 오류는 용언의 활용과 어미의 활용을 잘못한 경우를 말한다. 즉, 용언의 규칙 및 불규칙 활용, 어미의 이형태 오류가 포함된다.
- [주석 방식] 활용 양상에 따라 오류 위치를 판단하여 주석하며, 오류 양상은 오형태(MIF)로 주석하고, 오류 층위에 굴절[활용](MCJ)를 주석한다.
- [처리 기준] 용언의 규칙/불규칙 활용과 어미 이형태에 따라 활용 오류를 처리한다. 단순 오철자 오류의 경우 오형태(MIF)만 처리하며, 오류 층위에 활용(MCJ)을 주석하지 않는다.

<예> 미국에 영어 제일 중요하다(√중요하다).
 ☞ ‘중요하다’를 ‘중요한다’로 종결 어미를 잘못 활용하였으므로 [오류 위치-종결어미], [오류 양상-오형태], [오류 층위-활용] 오류로 주석한다.
 이렇게 살으면(√살면) 정말 행복할 수 있다.

☞ 받침 ‘ㄹ’ 뒤에서 ‘-으면’의 형태로 잘못 활용하였으므로
오형태 활용 오류로 주석한다.

- [쟁점] 활용 오류는 한국어 학습자가 생산하는 형태에 따라 오류 위치를 명확하기 판단하기 어려운 경우들이 있다. 따라서 활용의 양상에 따라 오류 위치를 용언의 어간에만 주거나 어미에만 줄 수도 있으며, 어간과 어미 양 쪽에 줄 수도 있다. 각각의 예시는 다음과 같다.
- [용언의 어간 활용 오류 1] 용언의 불규칙 활용을 시키지 않거나 과잉 적용시킨 경우 용언의 어간 활용 오류로 처리한다.

<예> 한국의 여름 날씨는 더웁니다.(√더웁니다.)
☞ ‘ㅂ불규칙 활용’ 형용사인 ‘덥다’를 종결어미 ‘ㅂ니다’ 앞에서 ‘더우’의 형태로 과잉 적용한 경우다. 이처럼 불규칙 활용을 잘못 적용시킨 경우 [오류 위치-형용사/동사]로 주석하고 [오류 양상-오형태], [오류 층위-활용]으로 주석한다.

- [용언의 어간 활용 오류 2] 한국어 학습자들은 이형태가 없는 어미 앞에서 ‘아/어’나 ‘으’와 같은 매개 요소를 사용하는 경우가 많은데, 본고에서는 이를 학습자가 하나의 어간으로 재구성하고 있는 중간언어로 보고 용언 어간의 활용 오류로 처리한다.

<예> 2015년에는 친구하고 같이 많이 놀아고(√놀고) 싶습니다.
저는 좋아하는 프로그램은 많아지만(√많지만) 다른 방송에 비해서 동물에 대한 다큐멘터리는 프로그램이 제일 좋아한다.
우리 미래 길에 꼭 잘 조심해고(√조심하고),
☞ 이형태가 없는 어미 ‘고, 지만’ 앞에 ‘아/어’ 또는 ‘으’가 첨가된 경우, 어간 활용과 어미 활용의 구분이 어렵다. 학습자들이 동사 어간에 ‘아/어’를 첨가해 하나의 어간으로 재구성했다고 볼 수도 있고, ‘아고, 아지만’의 형태로 어미를 잘못 활용한 것으로도 볼 수 있다. 본 연구에서는 이와 같은 경우 ‘놀아’, ‘많아’, ‘해’를 하나의 어간으로 재구성

한 중간언어로 보고 용언의 활용 오류로 처리한다. 이는 학습자들이 연결어미(예: 하고), 종결어미(예: 합니다), 관형사형 전성어미(예: 해는) 앞에서 동일한 형태를 생산해내는 것으로 보아, 학습자들이 ‘해’를 하나의 단위로 인식하고 활용을 잘못 적용하였다고 판단했기 때문이다. 따라서 이와 같은 형태들은 해당 품사(용언 어간)를 오류 위치로 주석하고, [오류 양상-오형태], [오류 층위-활용] 오류로 주석한다.

- ‘하다/되다’ 활용 오류에서 ‘해(√하), 돼(√되)’로 잘못 쓴 경우는 용언의 활용 오류로 처리하는 반면에 ‘하(√해), 되(√돼)’의 경우는 용언의 어간과 어미 활용 오류로 처리함에 주의한다.

<예> 우리 미래 길에 꼭 잘 조심해다(√조심하다),
특정 장면은 항상 중요 메시지가 있어서 삭제되면(√삭제되면)
☞ ‘하다, 되다’에서, ‘해, 돼’로 잘못 쓴 경우는 어간의 활용 오류로 처리한다. 위의 경우 형태소 분석에 따라, ‘해’와 ‘돼’는 동사파생 접미사의 활용 오류로 주석한다.

그리고 한국어 말하야 해요(√말해야 해요)

아무지 친하도(√친해도) 존댓말로 써야 한다.

그래서 시청자 왕따 당할까봐 걱정이 되서(√돼서) 그 물건을 사게 된다.

☞ ‘하다/되다’ 용언 어간+‘어/아’ 계열 어미‘에서 ‘어/아’를 누락시킨 경우는 어간과 어미에 모두 오형태 활용 오류로 처리한다. ‘하다/되다’의 경우 어미와 결합할 때, ‘하/해’, ‘되/돼’로 형태를 바꾸기 때문에 오형태 활용 오류로 볼 수 있다. 그러나 오류 위치를 어간 어미 중 무엇으로 처리하느냐가 문제가 된다. 이 경우, 학습자가 용언 어간, 어미 둘 중 어느 곳을 잘못 사용하였는지 정확히 분리하기 어려워 어간과 어미 양쪽에 오형태 활용 오류를 주석한다.

- [용언의 어간 활용 오류 3] 이밖에 어미 앞에서 ‘ㄴ, ㄹ, ㅂ’ 등의 불필요한 요소를 첨가했을 경우도 용언 어간의 활용 오류로 처리한다.

<예> 왜냐하면 어렸을 때부터 커피숍이나 호텔의 사장님 되고(✓
되고) 싶어 하기 때문이다.

☞ 연결어미 앞에 ‘ㄹ’ 요소가 첨가된 경우, [오류 위치-동사],
[오류 양상-오형태], [오류 층위-활용] 오류로 주석한다.

달른(✓다른)

알랐다(✓알았다)

☞ ‘ㄹ’ 앞에서 ‘ㄹ’이 첨가된 경우는 [III] 발음의 영향으로 인
한 오철자 오류로 볼 수도 있다. 그러나 본 연구에서는 앞
에서 ‘ㄴ, ㄹ, ㅂ’ 요소들이 첨가된 경우 오형태 활용 오류
로 보아, 이 역시 오형태 활용 오류로 처리한다.

- [어미 활용 오류 1] 용언의 받침 유무에 따라 어미의 이형태 선택이 달라지
는 경우는 활용 오류로 처리한다.

<예> 이렇게 살으면(✓살면) 정말 행복할 수 있다.

☞ 받침의 유무에 따라 연결어미의 이형태를 선택하여 사용하
야 하는데, 잘못 사용한 경우로 [오류 위치-연결어], [오류
양상-오형태], [오류 층위-활용 오류]로 주석한다.

- [어미 활용 오류 2] 이형태의 선택 뿐 아니라 이형태가 있는 어미에서 오류
가 나타난 경우 오형태 활용 오류로 처리한다. 따라서 ‘아서/어서’에서 ‘서’
만 쓰거나 ‘아도/어도’에서 ‘도’만 쓴 경우, ‘(으)니, (으)면’ 등에서 ‘으’를 쓰
지 않은 경우 연결어미 활용 오류로 처리한다. 마찬가지로 종결어미에서도
‘아요/어요’에서 ‘요’만 쓴 경우 종결어미 활용 오류로 처리한다.
- [어미 활용 오류 3] 또한, 용언 어간과 어미의 축약상의 오류는 활용 오류
로 처리한다. 어미 이형태 활용 오류 외에 필수적으로 탈락시켜야 하는데
시키지 않은 경우 또는 과도하게 축약을 시켜버린 경우 모두 어미의 활용을
제대로 모르는 것으로 판단하여 활용 오류에 포함한다.

<예> 가아서(√가서), 가아도(√가도), 가아요(√가요)

한 시간 쉬요(√쉬어요)

☞ ‘가아서’처럼 축약을 시켜야 하는데 축약을 하지 않은 경우와 ‘쉬요’처럼 ‘쉬어요’를 과도하게 축약시킨 경우는 [오류 위치-연결어미/종결어미], [오류 양상-오형태], [오류 층위-활용 오류]로 주석한다.

- [어미 활용 오류 4] 종결어미에서 ‘ㅂ니다/습니다’ 외에 ‘니다’ 또는 ‘입니다’를 잘못 사용한 경우도 종결어미 오형태 활용 오류로 처리한다.

<예> 힘들습니다(√힘듭니다)

☞ 형용사 ‘힘들’과 종결어미 ‘습니다’ 양쪽 모두 활용을 잘못된 것으로, [오류 위치-형용사, 종결어미], [오류 양상-오형태], [오류 층위-활용]으로 주석한다.

아름답니다(√아름답습니다)

시끄럽니다(√시끄럽습니다)

☞ ‘ㅂ’ 받침으로 끝나는 용언 어간의 경우, 용언 어간 ‘아름다’와 종결어미 ‘ㅂ니다’로 결합시킨 것인지, ‘아름답’과 ‘니다’의 형태로 결합한 것인지 불분명하다. 이때, 용언어간 ‘아름답’과 ‘니다’로 결합한 것으로 일괄 처리하고, ‘ㅂ니다/습니다’와 같이 [오류 위치-종결어미], [오류 양상-오형태], [오류 층위- 활용] 오류로 주석한다.

재미있입니다(√재미있습니다)

맛있입니다(√맛있습니다)

☞ 형용사에 ‘입니다’를 결합시킨 경우도 종결어미의 활용 오류로 주석한다.

- [용언의 어간 + 어미 활용 오류] 어간 활용과 어미 활용 모두 실패한 경우에는 오류 위치에 해당 용언의 품사와 어미를 모두 주석한다.

<예> 그 이유는 제가 우라 아내보다 한국에 돈을 잘 못 벌읍니다 (√법니다).

☞ 동사 ‘벌다’를 활용하여 ‘법니다’로 써야하는데 ‘벌’을 그대로 사용하고 있기 때문에 [오류 위치-동사, 종결어미], [오류 양상-오형태], [오류 층위-활용] 오류로 처리한다.

한국 CD를 들면서(√들으면서) 한국어를 말하다.

☞ ‘ㄷ 불규칙’을 적용시키지 못하였으며, 연결어미 ‘으면서/면서’의 이형태 또한 제대로 사용하지 못한 오류로 동사의 어간과 연결어미 양쪽 모두의 활용 오류로 처리한다. [오류 위치-동사, 연결어미], [오류 양상-오형태], [오류 층위-활용]으로 주석한다.

⑤ 품사(POS)

- 품사 오류는 동일 의미 단어의 품사를 잘못 사용한 경우를 말한다. 즉, 같은 의미인 단어의 품사를 제대로 인식하지 못하여 명사를 동사로 사용하거나, 부사를 형용사로 잘못 사용한 경우를 말한다.
- 오류 층위에서 품사에 해당하는 오류는 품사에서 나타난 오류(품사가 달라진 경우)와 품사를 몰라서 생겨난 오류(품사 혼동으로 인한 오류)로 구분할 수 있다. 이때, 본고에서는 품사가 달라진 것보다 품사를 모르고 있는 것을 우선적으로 적용하여, 품사에 대한 인식이 없어서 생겨난 오류만 품사 오류로 주석한다. 따라서 문맥에 따라 교정 어절이 바뀌면서 품사가 달라진 경우(단순히 원어절과 교정어절의 품사가 상이한 경우, 다시 말해서 의미가 다른 품사)는 품사 오류로 처리하지 않는다.
- 즉, 원어절과 교정 어절이 의미를 공유하면서 품사를 제대로 사용하지 못한 경우, 품사에 대한 인식이 부재한 것으로 간주하고 이를 품사 오류로 처리한다.
- 품사 오류의 오류 양상은 기본적으로 대치 오류로 주석한다.

<예> 그래서 우리는 빠른(√빨리) 우리 집에 다가했다.

☞ 부사 ‘빨리’를 쓸 자리에 형용사 ‘빠르다’를 사용하였다. 이는 학습자가 동일한 의미의 단어에서 부사 품사를 모르기

때문에 형용사를 관형형으로 사용한 것으로 보고 품사 대치 오류로 처리한다.

- 품사 오류에서 ‘파생/합성’과 관련된 오류는 오류 층위에서 품사(POS)와 단어 형성[합성법](MCP) 또는 단어 형성[파생법](MDV)을 중복 주석한다. 이는 표면상 원어절과 교정어절에서 드러나는 차이에 주목하여 품사 오류로 처리하는 동시에 파생과 합성을 하면서 품사가 바뀌는 경우인데, 품사를 바꾸는 접사에 대한 인식이 없다고 판단한 것이다.
- 따라서 ‘N+하다, 되다, 시키다, 있다, 없다, 나다(화나다, 겁나다, 불나다, 열나다 등)’에서 어간만 사용한 경우는 오류 위치[명사 또는 어근] - 오류 양상[대치] - 오류 층위[품사, 파생/합성(파생접사가 아닌 경우)] 중복 주석 처리한다.

<예> 이 문제들을 방지하기 위해 인터넷 사용(√사용하는) 사람들이 몇 가지 태도를 갖춰야 한다는 것을 청구한다.

☞ 동사 ‘사용하다’를 쓸 자리에 명사 ‘사용’만 사용하였다. 이는 명사와 동사 품사를 모르고 있다고 보고 품사 오류로 처리한다. 동시에 파생접미사 ‘-하다’를 붙여서 동사를 파생시키지 못한 오류로 보고 파생 오류도 중복 처리한다.

- 내 생각기는 한국 사람들 다른 나라 사람들보다 친절(√친절한) 것 같다.
- 그리고 한국 친구를 인사고(√인사하고) 싶습니다.

☞ 위의 예시들은 크게 3가지로 해석해 볼 수 있다.

- 1) 동사와 명사를 구분하지 못하여 동사를 써야할 자리에 명사를 쓴 경우로 학습자들이 품사를 제대로 인식하지 못해서 발생한 오류. 즉, (특히 중국인 학습자의 경우) 명사가 동사적 기능을 한다고 인식하여 명사를 쓴 경우
- 2) ‘명사’에 ‘하다’ 접사를 붙여서 동사를 파생시키는 것을 모르기 때문에 ‘하’를 누락시킨 것으로 파생어를 만드는 방법

을 모르는 경우

- 3) ‘N+하다’ 동사는 알고 있지만 뒤의 연결어미와 결합시키면서 기본형 ‘다’외에 ‘하다’를 같이 생략하여 활용했다고 볼 수도 있다.

내용적으로 학습자의 오류 원인을 예측해봤을 때, 크게 위의 3가지로 해석할 수 있다. 그러나 우선 원어절과 교정어절상 표면적으로 드러나는 것은 동사를 써야하는데 명사를 썼기 때문에 품사 대치 오류로 우선 처리한다. 즉, 원어절과 교정어절에서 드러나는 차이에 주목하여 품사 오류로 처리한다. 그러나 한편으로 이러한 품사 오류는 조어법과도 긴밀하게 관련된다. 파생과 합성을 하면서 품사가 바뀌게 되기 때문에 파생/합성에는 품사의 의미도 포함되어 있다고 볼 수 있다. 따라서 이러한 오류에 대해서는 품사를 바꾸는 접사에 대한 인식이 없다고 보고 ‘품사 대치’ 오류로 우선 주석한 후, ‘파생/합성’ 오류도 함께 주석한다.

오류 양상[대치] - 오류 층위[품사, 파생/합성] 중복 주석 처리한다.

- 한국 공부가 너무 재미(✓재미있)기는 하지만, 단어 위우가기 힘들다.

📖 ‘재미있다’에서 ‘재미’만을 사용하여 활용했다면, 형용사와 동사 품사 대치 오류인 동시에, ‘재미’와 ‘있’을 합성시키지 못한 것으로 보고 오류 층위에서 품사와 합성법을 중복 주석한다. 오류 위치[명사] - 오류 양상[대치] - 오류 층위[품사, 합성법]

- 마찬가지로 지정사 ‘이다’와 파생접미사 ‘하다’가 대치된 경우, 즉 ‘명사(어근)+하다’ 동사를 쓸 자리에 ‘명사+이다’를 쓴 경우, ‘동사/형용사’와 명사 품사 혼동으로 보고 품사 대치 오류로 처리한다. 동시에 접사 ‘하다’를 사용하여 동사를 파생시키는 것을 모른다고 판단하여 오류 층위에 품사와 파생을 중복 주석한다.

<예> 이번 축제에 제수님 추구하고 부활을 위미입니다(√ 의미합니다).

☞ 동사를 쓸 자리에 ‘체언+이다’의 꼴로 나타났다. 이러한 경우 품사 오류로 처리하는 동시에 ‘하다’를 사용하여 동사를 파생시키는 것을 모르는 경우라고 해석하여 파생오류도 중복 주석한다.

- ‘외국어’에 파생접사 ‘하다’를 결합하여 동사/형용사를 만들 수 있는데, 외국어 명사만 사용한 경우도 동일하게 품사 오류와 파생 오류로 중복 주석한다.

<예> 그리고 장애인의 사회참여를 스마트(√스마트한) 사람이 필요합니다.

☞ 외국어의 명사형만 사용한 경우도 오류 양상[대치] - 오류 층위[품사, 파생] 중복 주석 처리한다.

- 품사 대치 오류로 처리할 경우, 원래 뒤에서 연결되어 있던 요소들을 첨가 처리하거나 품사 대치로 인해 따라오는 요소들을 누락으로 처리하지 않는다. 이는 품사를 제대로 인식하지 못한 것이므로, 품사 오류가 발생한 위치와 함께 뒤의 요소는 한 덩어리로 묶어서 처리하고 첨가 또는 누락으로 주석하지 않도록 한다.

<예> 직장을 선택의(√선택하는) 조건들이 많이 있고 사람마다 다른 생각하고 의미도 있는 것 같다.

☞ 동사 ‘선택하다’를 명사 ‘선택’만을 썼을 때, 뒤의 관형격 조사 ‘의’를 쓴 것은 문법적으로 틀리지 않다. 따라서 관형격 조사 ‘의’를 추가로 첨가처리 하지 않는다.

한편에 일부 사람은 가족이 가장 중용(√중요하다고) 생각했는데 그 이유 점은 돈으로 바꿀 수 없다고 지적을 했다.

☞ 동사 ‘중요하다’를 써야할 자리에 명사 ‘중요’의 오형태 ‘중용’만을 썼기 때문에 명사의 품사 대치 오류로 처리한다. 아울러 ‘중요’도 ‘중용’이라고 썼기 때문에 오형태도 동시 주석한다. [오류 위치-명사], [오류 양상-대치, 오형태], [오류 층위-품사] 오류로 처리한다. 이때, 품사 대치의 경우 뒤에 따라오는 요소는 누락으로 처리하지 않는다. 즉, ‘중요하다’ 뒤에 오는 연결어미 ‘-고’는 누락 처리하지 않는다. 이는 품사를 제대로 사용하지 못한 것으로 인한 것이기 때문에 뒤에 따라오는 활용 형태들, ‘-한, -하게, -다고, -하니까’ 등 관형사형 전성어미나 연결어미 등은 ‘누락’ 처리하지 않는다.

(3) 통사

① 높임(SH)

- 높임 오류는 조사, 선어말어미, 종결어미 부분의 높임 관련 문법 형태소, 높임 어휘의 사용이 잘못된 경우를 말한다.
- 높임법은 주체 높임법과 객체 높임법, 상대 높임법이 있으며, 이는 다시 문법적 높임과 어휘적 높임으로 나뉜다. 이 중 문법적 높임 오류는 주격조사와 서술에서의 높임 호응 관계 불일치, 주체 높임을 나타내는 선어말어미 ‘-시-’의 잘못된 사용, 상대 높임을 나타내는 대명사와 종결어미의 호응 관계 불일치, 객체 높임을 나타내는 부사격조사 ‘께’의 잘못된 사용이 해당되며, 어휘적 높임 오류는 ‘계시다, 드리다, 모시다, 잡수시다, 주무시다’ 등의 특수한 높임말을 써야할 자리에 쓰지 않은 경우 또는 그 반대의 경우가 해당된다.
- 그러나 특수 어휘에 의해서 표현되는 높임법의 경우 존대와 겸양을 나타내는 특수 어휘가 다양하다. 높여야 할 대상인물을 직접적으로 높이는 어휘, 대상과 관계있는 것을 간접적으로 높이는 어휘, 객체를 높이는 동사 어휘, 접미사나 접두사가 붙어 존대나 겸양을 나타내는 어휘 등 다양하기 때문에 높임과 관계된 모든 어휘를 오류 주석 대상으로 삼기 어렵다. 이러한 이유로 본 연구에서는 어휘적 높임 오류보다 문법적 높임 오류를 우선 주석한

다.

- 따라서 주된 높임 오류의 대상이 되는 품사는 대명사, 조사(주격조사 ‘께서’, 부사격 조사 ‘께’), 선어말어미, 종결어미이다.
- 높임의 오류에는 높임법을 써야 하는 환경에서 낮춤말을 쓴 경우와 반대로 낮춤말을 써야 하는 환경에서 높임말을 사용한 경우를 모두 포함한다.

<예> 할아버지께서 사 주었어요(√주셨어요).
근데 부모님께서 나에게 유학하라고 해서(√하셔서) 부모님의 말(√말씀)대로 한국에 와서 유학했다.
☞ 주체 높임을 실현시키기 위해 ‘께서’와 ‘-시-’를 실현시켜야 하는데 높임 선어말어미는 실현시키지 않았으므로 높임 오류로 판정한다. 명사 ‘말’의 경우 ‘말씀’의 어휘 높임 대치 오류로 주석한다.

- 한국어의 높임법은 주체 높임과 상대 높임, 특수어휘에 의한 객체 높임 모두 담화층위에서 화청자 관계에 따라 실현되는 것으로 높임 표현 오류의 적절성과 용인 가능성이 다르게 적용될 수 있다. 한국어 모어 화자들도 일상적인 언어생활에서 높임 표현을 잘 지키지 않는 경우가 많으며, 그 사용이 일관적이지 않다. 즉, 높임 표현 오류는 용인 가능성으로 인하여 문법성이나 적합성 기준을 일관적으로 적용하기에 어려움이 따른다. 이러한 점을 고려해 일관된 높임 오류 주석을 위하여 동일 문장 내 높임 표현의 호응을 필수적인 높임 표현의 오류 판정 기준으로 삼았다. 예를 들어, 주격조사 ‘께서’를 사용하였지만 서술어에서는 높임 선어말어미 ‘-시-’를 사용하지 않은 경우와 같이, 문장 내 한 부분이라도 높임 요소를 실현시킨 경우에는 실현시키지 않은 부분을 높임 오류로 판정한다.

<예> 근데 부모님께서 나에게 유학하라고 해서(√하셔서) 부모님의 말(√말씀)대로 한국에 와서 유학했다.
☞ 주체 높임을 실현시키기 위해 ‘께서’와 ‘-시-’를 실현시켜야 하는데 높임 선어말어미는 실현시키지 않았으므로 높임 오류로 판정한다. 명사 ‘말’의 경우 ‘말씀’의 어휘 높임 대치 오류로 주석한다.

- 문장 이상의 단위인 담화 상에서 높임말과 반말을 혼용하여 사용하고 있는 경우에는 문장 단위로 오류를 판단한다는 원칙하에 한 문장 내에서 높임 표현에 문제가 없으면 오류로 주석하지 않는다.
- 상대높임법에서 대명사와 종결어미 양쪽 모두 교정이 가능할 때에는 종결어미를 기준으로 대명사 오류로 일괄 처리한다. 즉, 종결어미에 따라 대명사 ‘나’와 ‘저’의 대치 오류로 우선 처리한다.

<예> 저는(√나는) 10년 후에 생활이 부유하고 싶다.

☞ 상대높임법 체계에서 해라체를 사용했는데, 대명사는 자신을 낮추는 ‘저’를 잘못 사용하였다. 따라서 상대높임법 체계에 맞추어 ‘나’로 교정한다. 반대로, 합쇼체를 사용하고, 대명사 ‘나’를 사용한 경우도 마찬가지로 [오류 위치-대명사], [오류 양상-대치], [오류 층위-높임]으로 주석한다.

- 높임을 나타내는 접사 ‘님’의 경우, 생산성이 강한 접사로 형태 주석에서 접미사로 따로 분석 처리한다. 이와 연계하여 오류 주석에서는 높임을 나타내는 접사 ‘님’의 과잉사용 또는 미사용의 경우 ‘첨가’와 ‘누락’ 오류로 처리한다.

<예> 의사님(√의사), 책상님(√책상), 선생(√선생님)

☞ ‘선생님’을 ‘선생’으로 쓴 경우, 어휘 대치로 볼 수도 있으나, 접사 ‘님’ 누락으로 처리한다. 왜냐하면 ‘의사님’, ‘책상님’ 등으로 쓴 경우는 존대를 과잉생산한 것으로서 ‘첨가’로 처리해야 해야 하는데, 동일 요소에 대하여 다르게 처리하게 되는 문제가 있다. 따라서 ‘님’과 같이 존대를 나타내는 접사는 일괄 ‘누락/첨가’로 처리한다. [오류 위치-접사], [오류 양상-누락/첨가], [오류 층위-높임]으로 주석한다.

② 시제(ST)

- 시제 오류는 시제 또는 시상을 나타내기 위한 선어말어미, 관형사형 전성어미, ‘-(으)ㄴ 것’ 등의 오류를 말한다.

<예> 옛날에 한국의 정통 난방법을 온돌이다(√온돌이었다).
 때로는 한국말을 공부할 때 끝이 없는 것 같은데(√같았는데)
 데) 오느날 갑자기 끝에 왔다.

- 시제 오류는 오류 양상을 두 가지로 처리한다. 하나는 시제 요소를 사용하지 않고 기본형을 사용했을 경우, 시제를 사용하지 않았다고 판단하여 누락 오류로 주석한다. 다른 하나는 시제를 제대로 인식하지 못하여 과거 시제나 미래 시제 자리에 현재 시제를 사용한 경우, 그 반대의 경우 등은 대치 오류로 처리한다. 즉, 시제를 사용했으나 현재와 과거, 과거와 미래처럼 잘못 사용한 경우는 시제 간 대치로 처리한다.
- 이때, ‘이다/아니다’, 형용사, 연결어미 앞에서의 용언은 기본형이 현재를 나타내기 때문에 기본형을 사용했을 경우, 현재 시제로 인식하고 대치로 주석한다.

<예> 먹다(√먹었다) ㉠ 누락
 먹다(√먹는다) ㉠ 종결어미 오형태 활용
 먹는다(√먹었다) ㉠ 대치
 습니다(√있습시다) ㉠ 대치
 있다(√있었다) ㉠ 대치
 예쁘다(√예뻤다) ㉠ 대치
 이다(√이었다) ㉠ 대치
 아니다(√아니었다) ㉠ 대치
 ㉠ 형용사와 ‘이다/아니다’는 기본형과 현재가 같으므로 기본형을 현재형으로 간주하고 대치로 처리한다. 단, 예쁘다(√예뻤다)와 같은 경우는 오형태 활용 오류에 해당한다.
 밥을 먹지만(√먹었지만),
 밥을 먹고(√먹었고)
 ㉠ 연결어미에서도 현재형으로 보고, 앞에 선어말어미 ‘-았-’이 와야 하는데 사용하지 않은 경우 [오류 위치-선어말어미], [오류 양상-대치], [오류 층위-시제] 오류로 처리한다.

- 시제 선어말 어미 ‘-었-’과 ‘-겠-’이 생략된 경우, 기본형을 제외하고는 대

치 오류로 처리한다. 그러나 시제 선어말 어미의 문법적 제약이 있는 연결 어미 앞에서 ‘-었-’과 ‘-겠-’을 사용한 경우는 첨가 오류로 처리한다.

<예> 아침에 밥을 먹을 때 해물을 먹다(✓먹었다).

☞ 기본형 ‘먹다’를 사용해서 선어말 어미 ‘-었-’ 누락 오류로 처리한다.

고향에서 향주까지 4시간 걸렸서(✓걸려서) 좀 피곤했다.

몇 년 전에 영국에서 임신분에게 동물 실험을 했던 약을 주었다 보니(✓주다 보니)

☞ 연결어미 ‘어서’ 또는 ‘-다 보니’ 앞에 과거시제 선어말 어미 ‘-었-’이 올 수 없으나 ‘-었-’을 사용했다. 이처럼 ‘-었-’과 ‘-겠-’을 사용할 수 없는 문법적 제약이 있는 자리에 사용한 경우는 선어말 어미 첨가(ADD) 처리한다.

- 미래시제를 나타내는 ‘-겠-’의 경우, 추측이나 의지와 같은 양태 의미와도 관련되는데, 본 연구에서는 추측을 제외한 경우 오류 층위에 시제를 주석한다.

<예> 앞으로 학교 규칙을 안 어긴다(✓어기겠다).

☞ 미래의 의지를 나타내야 하는데 현재형 ‘ㄴ다’로 잘못 사용한 것으로, 선어말 어미 ‘-겠-’ 대치 오류로 처리한다.

사람들이 항상 물건을 어떻게 선택할지 모르겠다(✓모른다).

☞ ‘모르겠다’의 경우 ‘-겠-’이 시제의 의미를 나타낸다고 보기 어렵다. 따라서 이때의 ‘-겠-’은 오류 층위에 시제를 처리하지 않는다.

- 관형사형 전성어미 ‘-(으)ㄴ, 는, (으)ㄹ’의 경우, 뒤에 오는 (의존)명사에 따라 시제를 나타내는 경우가 있고 그렇지 않은 경우가 있다. 시제를 나타내지 않을 경우는 대치 또는 오형태 활용 오류로 처리하고, 시제를 나타낼 때에는 관형사형 전성어미의 시제 대치 오류로 처리한다.
- 시제를 나타내지 않을 경우로 ‘-(으)ㄹ 때, -(으)ㄹ 따름이다, -(으)ㄴ/는 편

이다, -(으)ㄴ 후'의 구성 등이 있다. 이때의 관형사형 전성어미는 특정한 시제의 의미가 없기 때문에 시제 오류로 처리하지 않도록 주의한다.

<예> 가(√가는/√갈) 사람

☞ '가는 사람' 또는 '갈 사람'을 써야하는데, '가 사람'으로 쓴 경우는 관형사형 전성어미 [누락]으로 처리한다. 그러나 '가는 사람'을 '갈 사람'으로 썼을 경우는 [오류 위치-관형사형 전성어미], [오류 양상-대치], [오류 층위-시제] 오류로 처리한다.

과식이나 하지 말고 여러까지 음식을 골고루 먹을(√먹는) 것이 중요해요.

☞ 이때의 '-(으)ㄴ 것'에서 관형사형 전성어미는 시제를 나타낸다고 보기 어렵다. 따라서 관형사형 전성어미 대치 오류로 처리하고 오류 층위에 시제는 주석하지 않는다.

- 관형사형 전성어미에서 시제 대치 오류 판단에 어려움이 있을 경우, 현재('는')와 미래('-(으)ㄹ')가 둘 다 가능할 때에는 용인 가능한 것으로 판단하여 오류로 처리하지 않고, 명확하게 과거형을 써야 하는데 쓰지 않은 경우나 반대의 경우 시제 오류로 처리한다.
- 연결어미 '-(으)ㄴ지/는지/-(으)ㄹ지'는 시제 대치 오류로 처리하지 않도록 주의한다.

<예> 10년 후에 어느 나라에 살고 있는지(√있을지) 잘 모르는데 그때는 좋은 일이 있었으면 좋겠다.

☞ 연결어미 '-(으)ㄴ지/는지/-(으)ㄹ지'의 대치의 경우, 오류 층위에 시제를 주석하지 않도록 주의한다.

③ 사동(SC)

- 사동 오류는 사동사, 사동 표현의 사용, 사동문 생성에서 발생한 오류를 말한다.
- 사동사, 사동 표현을 사용해야 하는데 사용하지 않은 경우는 기본적으로 오

류 양상을 대치 오류로 주석한다.

- 사동 표현은 접미사 ‘-이/히/리/기/우/구/추-’에 의한 사동, ‘-게 하다’에 의한 통사적 사동, ‘-내다, 만들다, 시키다’ 등 어휘적 사동으로 나타낼 수 있다. 이 연구에서는 접미사 ‘-이/히/리/기/우/구/추-’에 의한 사동사와 ‘-게 하다’, ‘시키다’ 사동 표현에 의한 사동으로 제한한다. ‘내다, 만들다, 시키다’ 중 ‘시키다’는 한국어 교육에서 ‘사동’을 나타내는 표현으로 교수하고 있는 상황을 고려하여 포함하나 나머지 어휘에 의한 사동은 맥락에 따라 다르게 처리될 수 있기 때문에 주석자 간 일관성을 유지하기 위해 제외한다.
- ‘형용사 -게 하다’의 경우, ‘사동’으로 처리하지 않도록 주의한다. 예) 방을 깨끗하게 해야 한다.

<예> 왜냐하면 중국 밥물 중에서 노동밥에 따라서 소득 격차 등 불공평 제도를 감소할 수 있다(√감소시킬 수 있다).

☞ ‘감소하다’와 ‘감소시키다’의 사동 대치로 처리한다. 또한 ‘-게 하다’의 사동 표현도 대치 오류로 처리한다.

전통의 아름다움이 사람들에게 알려주는(√알려주는) 것도 전통을 보존하려고 해야 할 일이다.

☞ ‘알려주는’의 경우, 사동접미사 ‘리’를 인식하였으나 형태를 잘못 사용한 것으로 보고, 이러한 경우에는 오형태 오류로 처리한다.

- 사동사, 사동 표현에서 나타난 오류를 모두 볼 수 있도록 오류 양상에 관계 없이, 철자를 잘못 사용한 오형태 오류도 오류 층위에서 사동으로 주석한다.
- 원어절에서 사동사, 사동 표현을 사용한 경우와 교정 어절이 사동사, 사동 표현이어야 하는 경우 모두 오류 층위에서 사동으로 주석한다.
- 사동을 쓸 자리에 피동을 썼거나 반대의 경우는 사동과 피동으로 중복 주석한다. 오류 위치와 오류 양상은 원어절 기준으로 주석하지만 오류 층위는 원어절과 교정어절 양쪽에서 주석함에 따라 사동과 피동을 중복 주석한다.
- 사동 표현 ‘-게 하다’와 일반 사동사가 대치된 경우, 오류 위치는 형태 주석에 따라 일관되게 처리한다. 동사 또는 ‘연결어미+보조용언’으로 분리되어 처리되었을 경우는 각각의 품사로 오류 위치를 주석하며, 사동 표현의 경우

표현 문형 목록에 해당되기 때문에 표현문형(PE)도 중복 주석한다.

<예> 더 간단하게 하려고 생각하면 에어컨의 온도를 조금만 높게 하는(√높이는) 것만 한 방법이 없다.
 ☞ ‘-게 하다’가 사동사로 대치된 경우로 오류 위치는 ‘연결어미, 보조용언, 표현 문형으로 주석하며, 오류 양상은 대치로 주석한다.

④ 피동(SP)

- 피동사, 피동 표현의 사용, 피동문 생성에서 발생한 오류를 말한다.
- 피동사, 피동 표현을 사용해야 하는데 사용하지 않은 경우는 기본적으로 대치 오류로 처리한다.
- 피동 표현은 접미사 ‘-이/히/리/기-’에 의한 피동, ‘-아/어지다, -게 되다’에 의한 통사적 피동, ‘-되다, 받다, 당하다’ 등 어휘적 피동으로 나타낼 수 있다. 이 연구에서는 형태를 중심으로 접미사 ‘-이/히/리/기-’에 의한 피동사와 ‘-어지다’ 피동 표현에 의한 피동으로 제한한다.
- 단, 통사적 피동 ‘-아/어지다’의 경우, ‘형용사+아/어지다’는 피동보다는 상태변화의 의미를 나타낸다고 보고 ‘피동’으로 주석하지 않는다. 상태변화가 일어나게 된 요인이 타의에 의해 발생하여 피동의 의미가 내포되어 있더라도 한국어 교재 및 학습 기관에서 피동과 상태변화를 분리하여 교수하고 있으며, 맥락에 따라 피동과 상태변화를 구분하여 주석할 경우, 주석자 간 일관성이 떨어질 수 있기 때문에 ‘형용사+아/어지다’는 일괄적으로 오류 층위에서 피동으로 주석하지 않는다.
- ‘-게 되다’의 경우도 변화의 의미를 나타내는 경우가 많으며, 학교문법에서 피동에 포함시키지 않는 논의에 근거해 본 연구에서도 제외하였다.
- 즉, ‘형용사+아/어지다’와 ‘-게 되다’는 기본 의미를 변화로 보고, 피동으로 다루지 않으며, 맥락에 따라 다르게 판단할 수 있는 어휘적 피동도 제외한다.

<예> 우리 집에 물을 열리면(√열면) 계단을 있다.
 ☞ 피동표현을 사용하지 말아야 하는데, ‘열리면’으로 피동형을 사용했기 때문에 피동 대치 오류로 처리한다.

기술이 발달해서 멋진 영화나 공연이 많아질수록 전통문화를 점점 잊어버리게 했다.(✓잊어버리게 된다).

☞ 사동 표현 ‘-게 하다’와 피동 표현 ‘-게 되다’의 대치 오류로 처리한다. 오류 층위는 원어절과 교정 어절 양쪽 모두를 기준으로 하기 때문에 이때에는 오류 층위에 사동(SC)과 피동(SP)을 중복 주석한다.

- 조사와 용언 교정이 모두 가능한 경우, 격조사 오류를 우선적으로 처리하나, 문맥에 따라 양쪽을 모두 바꿔야 하는 경우는 양쪽 모두 오류 주석한다. 특히, 피동문에서 용언을 교정하여 바뀌게 되는 조사의 경우, 교정어절만 써주고 오류로 처리하지 않았으나 피동/사동 구조를 모르고 있다는 측면에서 오류 층위에서 사동/피동을 주석하는 것으로 수정하여 처리한다.

<예> 둘째, 의학 기술의 발전에 따라 수명을(✓수명이) 연장하지만 (✓연장되지만) 노인층 증가도 할 수 있다.

☞ ‘되다’의 피동으로 용언을 교정함에 따라 조사도 바뀌게 된다. 이 경우는 양쪽 모두 오류로 처리하고 조사에도 오류 층위에 ‘피동’을 주석한다.

- 피동 오류를 처리하는 데 있어, ‘동사+아/어지다’의 경우, <표준국어대사전>에 한 단어로 등재되어 있는 동사가 있는 반면, 등재되지 않은 단어가 있다. 이 경우, 형태소 분석에서 사전에 등재되어 있는 단어는 동사로 주석하고, 그렇지 않은 경우는 ‘연결어미+보조 용언’으로 분리하여 주석한다. 따라서 오류 주석에서 오류 위치는 형태 주석에 따라 일괄 처리하여, 동사로 분석했을 때는 그 품사를 따르고, 연결어미, 보조 용언으로 분리하였을 경우 해당 품사를 오류 위치로 주석한다.
- 이중피동을 사용한 경우, 첨가 오류로 처리한다.

<예> 환경 오염이 심해지게 되고(✓심해지고) 있지만 더 이상 심하기 전에 여러분의 도움이 필요하다고 생각한다.

☞ ‘심해지다’에 ‘-게 되다’까지 첨가된 이중 피동표현으로 ‘-게 되다’의 ‘연결어미+보조용언’, 표현문형(PE)의 첨가 오류

로 처리한다.

⑤ 부정(SN)

- 부정 표현의 사용, 부정문의 생성에서 발생한 오류를 말한다.

<예> 한국에서 혼자서 살다가보니 외로울 때가 많이 있으니까 그냥 혼자 있지 말고(√않고) 친구들이랑 같이 공부를 해요.

- 일반적으로 부사 ‘아니(안), 못’이나 부정의 의미를 가진 용언 ‘아니다, 아니하다(않다), 못하다, 말다’를 써서 부정문을 만드는 방법에 근거하여 부정 부사를 잘못 사용하거나 해당 용언에서 오류가 났을 경우, 오류 층위에 부정(SN)을 주석한다.
- ‘없다, 모르다’, 부정 의미의 접두사는 부정 오류에 포함하지 않는다.
- 장형부정인 ‘-지 않다’, ‘-지 못하다’, ‘-지 말다’의 경우, 표현문형 목록에 해당되기 때문에 오류 위치는 보조용언과 표현문형을 중복 주석한다. (※ ‘-고(야) 말다’는 부정의 의미를 나타내는 것이 아니기 때문에 부정으로 처리하지 않도록 주의한다.)
- 장형부정이 더 자연스럽지만 단형부정을 썼을 때 용인가능하기도 하다. 따라서 단형부정을 장형부정으로 반드시 바꿔야하는 경우 기준 마련이 필요한데, 합성어나 파생어의 경우 단형부정문을 만들지 않으며, 용언의 음절이 긴 경우에도 단형부정을 허용하지 않기 때문에 이에 해당하는 용언의 경우는 장형부정으로 교정하고, 나머지의 경우 단형부정의 용인가능성을 인정하도록 한다. 단, 단형부정의 용인가능성의 경계가 분명하지 않으므로 주석자의 판단에 따라 적절하지 않다고 판단하여 장형부정으로 교정했을 경우는 적절성의 오류도 포함한 것으로 한다.

<예> 하지만 한국어는 안(√못) 잘합니다(√합니다). 그래서 한국 친구가 아직 없습니다.

☞ 능력을 부정하는 경우, 부정부사 ‘못’을 사용해야 하는데, ‘안’을 썼으므로 ‘안’을 ‘못’으로 교정하고 대치 오류로 처리한다. [오류 위치-일반부사], [오류 양상-대치], [오류 층위-

부정]. 또한 ‘안’과 ‘못’ 부정의 경우, 서술어 ‘잘합니다’도 ‘합니다’로 교정하고 대치 오류로 처리한다.

- ‘-하다’ 파생동사들의 경우는 체언과 ‘-하다’가 분리될 때 ‘하다’ 앞에 아니(안)를 넣어 단형부정문을 만들 수 있다. 따라서 ‘-하다’ 파생동사 앞에 부정부사를 쓴 경우는 오류로 처리하고, 이때는 어순 오류와 부정 오류로 중복 주석한다.

<예> 스페인어 안 사용해서(√사용 안 해서) 스페인어만 말하기 저금 어렵습니다.

☞ ‘N+하다’ 파생동사 앞에 부정부사 ‘안’을 사용한 경우로, 이때에는 ‘사용 안 해서’와 ‘사용하지 않아서’ 두 가지로 교정이 가능하다. 그러나 이 경우, 최소 수정 원칙에 의해서 단형부정을 장형부정으로 바꾸는 것보다 단형부정의 위치를 잘못 사용한 것으로 보고 오류 층위에서 어순 오류와 부정 오류로 중복 주석한다.

⑥ 어순(WO)

- 어순 오류는 한국어의 통사 구조에 맞지 않는 방식으로 문장 전체 또는 일부가 배열된 경우를 말한다.

<예> 그래서 잘 아직까지(√아직까지 잘) 몰라요.
저녁까지 많이 이야기도(√이야기도 많이) 합니다.

- 한국어 어순의 특징 중 하나는 문장성분의 자리 이동이 비교적 자유롭다는 것이다. 그렇기 때문에 그만큼 용인가능성이 크다고 할 수 있다. 이에 따라 주석자 간의 일치도도 다르게 나타날 수 있어, 어순 오류의 경우 최소 수정의 기준을 마련하여 처리한다.
- 문장 부사는 자리 이동이 자유롭지만 성분 부사의 경우는 제한되기 때 특정한 성분을 수식해야 하는 성분부사의 위치를 잘못 사용했을 때는 오류로 주석한다.

- 관형사의 경우, ‘지시관형사-수관형사-성상관형사’ 순의 기준을 적용하여 처리한다.
- 어순 오류는 오류 위치와 오류 층위만 주석하고, 오류 양상은 주석하지 않는다. 또한 교정어절을 줄 필요가 없어 교정된 어순을 반영하여 앞이나 뒤에 추가하지 않는다.

<예> 저는 많이 여행을 (√많이) 가고 싶습니다.
 저는 한국 여행에서 자주 서울만 (√자주) 갔습니다.
 ☞ 일반적으로 성분 부사는 서술어 앞에서 수식해야 하는데, 이처럼 ‘많이’, ‘자주’가 명사 앞에 온 경우, [오류 위치-일반부사], [오류 양상-없음(빈칸)], [오류 층위-어순] 오류로 처리한다.

- 시간을 나타내는 표현의 배열이 잘못되었을 경우, 어순 오류로 처리한다. 시간을 나타내는 표현은 ‘년도-월-일-오전/오후/밤/낮/아침/점심/저녁-시-분-초’의 순서로 배열되는 것이 일반적으로 이를 기준으로 시간의 배열 어순 오류를 판단하여 처리한다.
- 2개의 문장 성분이 상호 교체될 때에는 2개 모두 대치 어순 오류로 주석한다.

<예> 8반 시(√8시 반)에 학교에 거러서 가요
 ☞ 시간표현에서 시보다 분을 먼저 배열했기 때문에, ‘반’과 ‘시’의 어순 대치 오류로 주석한다. 이 때, 어순이 상호교체되는 것으로 명사 ‘반’과 의존명사 ‘시’ 모두를 대치 어순 오류로 주석한다.

- 조사의 경우, 어순 오류로 처리하지 않고 조사 첨가 또는 누락 오류로 처리한다.

<예> 그래서 한국말을(√한국말) 공부(√공부를) 참 좋아했습니다.
 ☞ ‘한국말을’에서의 목적격 조사 ‘을’을 ‘공부’ 뒤로 보내는 어순 조정으로도 교정이 가능하다. 그러나 이때는 조사의

배열 문제라기보다는 조사를 잘못 사용한 것으로 판단하여
어순 오류로 처리하지 않고 앞의 ‘을’ 첨가, 뒤의 ‘을’ 누락
오류로 주석한다.

(4) 담화

① 지시(DR)

- [정의] 지시 오류는 부적절한 지시사의 선택으로 선행문과 후행문의 관계를
결속성 있게 나타내지 못한 경우를 말한다.
- [주석 방식] 담화 층위에서의 오류는 의미 대치 오류를 중심으로 처리함을
원칙으로 한다. 이에 따라, 지시 표현에서 의미 간 대치 오류를 중심으로
하여 오류 층위에 지시(DR)를 주석한다. 지시 표현에서 나타난 단순 오철자
오류는 오류 양상에 오형태만 주석하고, 오류 층위에서 지시를 주석하지 않
도록 주의한다.
- [처리 기준] 지시 오류는 앞 뒤 문장과 연결, 상황 맥락을 통해서 오류 판
단이 가능하기 때문에 문장 단위를 기본원칙으로 삼으나 지시 오류의 경우
는 문장 이상의 단위를 고려해 오류를 판단한다.

<예> 저기에(√거기에) 가면 좋을 것 같아요.
☞ 맥락상 ‘저기’보다는 ‘거기’가 더 적절한 표현으로, 대명사
대치 오류로 주석한다. 아울러 이는 지시 표현에 해당되므
로 오류 층위에서 지시(DR)도 함께 주석하도록 한다. [오류
위치-대명사], [오류 양상-대치], [오류 층위-지시(DR)]로
주석한다.

② 접속(DC)

- [정의] 접속 오류는 선행문과 후행문의 의미 관계를 나타내는 데에 부적절한
접속사를 사용한 경우를 말한다. 접속 부사 및 접속 표지의 오류가 포함된
다.

- [주석 방식] 담화 층위에서의 오류는 의미 대치 오류를 중심으로 처리함을 원칙으로 한다. 이에 따라, 접속 표현에서 의미 간 대치 오류를 중심으로 하여, 오류 위치에 접속 부사(CMAJ), 오류 양상에 대치(REP), 오류 층위에 접속(DC)을 주석한다. 접속 표현에서 나타난 단순 오철자 오류는 오류 양상에 오형태만 주석하고, 오류 층위에서 접속을 주석하지 않도록 주의한다.
- [처리 기준] 접속 오류는 앞 뒤 문장과 의미적 연결을 통해서 오류 판단이 가능하기 때문에 문장 단위를 기본원칙으로 삼으나 접속 오류의 경우는 문장 이상의 단위를 고려해 오류를 판단한다.

<예> 그래서(√그러면) 어떻게 해야 전통을 보존할 수 있을까요?
 그래서(√그러니까) 전통을 보존하기 위해 더 많이 노력을 해서 전통은 없어지지 않도록 하세요.

☞ 접속부사 ‘그래서’를 과잉 사용하고 있는 양상으로, 앞뒤 문장을 고려했을 때, 각각 ‘그러면’과 ‘그러니까’가 더 적절하다. 따라서 접속부사의 대치 오류로 주석하여 [오류 위치-접속부사], [오류 양상-대치], [오류 층위-접속(DC)]로 주석한다.

③ 담화표지(DM)

- [정의] 담화표지 오류는 담화표지와 간투사의 오류로, 부적절한 담화 표지를 선택하거나, 잘못된 형태로 이들을 사용한 경우를 말한다.
- [주석 방식] 담화표지에 해당하는 품사의 위치를 오류 위치로 주석한다.
- [처리 기준] 담화표지는 미시 담화표지와 거시 담화표지로 나눌 수 있으나 연구자마다 그 정의가 다르고, 해당 형태도 다르기 때문에 담화표지의 목록을 마련하기 쉽지 않다. 이러한 이유로 미시 담화표지에 초점을 두고 오류를 판단하도록 한다. 미시 담화표지의 경우, 학습자의 L1의 영향으로 인한 간투사 사용과 모어화자와는 다른 위치에서 담화표지를 사용한 경우를 오류로 주석한다.

<예> 아~ 그럼 제가 음~ 오늘 밤에, 데~(√에~) 잊어버리지 않으면 추대할게요.

☞ ‘에~’는 간투사로 볼 수 있는데, 이를 ‘데~’로 잘못 발음하

고 있어 오형태 오류로 주석하고, 오류 층위에서 담화표지 오류로 처리한다.

④ 구어/문어 오류(DS)

- [정의] 구어체(구어성)/문어체(문어성), 격식체/비격식체의 혼용에 의해 담화 맥락에서의 일관성이 떨어지는 경우를 말한다.
- [주석 방식] 문어에서 구어성이 강한 어휘나 구어에서 문어성이 강한 어휘를 사용한 경우 해당하는 품사의 위치를 오류 위치로 주석하고, 오류 양상은 대치(REP), 오류 층위에 구어/문어 오류(DS)를 주석한다.
- [처리 기준] 구어/문어 오류는 상황에 따라 용인 가능성을 적용할 수 있기 때문에 엄격하게 그 기준을 적용하기가 어렵다. 이에 구어체(구어성)/문어체(문어성)를 판단하는 기준은 <표준국어대사전>으로 삼는다. <표준국어대사전>에서 ‘문어적 표현’이라고 기술되어 있을 경우 ‘문어체’로 보고, ‘구어적 표현’이라고 기술되어 있을 경우 ‘구어체’로 판단한다. 따라서 문어에서 ‘구어체’를 사용한 경우, 구어에서 ‘문어체’를 사용한 경우에는 담화층위에서 구어/문어 오류로 주석한다.
- 단, <표준국어대사전>에는 기술되지 않았지만 문어에서 구어성이 강한 표현이거나 구어에서 문어성이 강한 표현일 경우에는 ‘용인가능성’ 기준을 적용하여 주석자간 논의 후 처리하고, 처리한 것을 검토하여 다시 목록화하는 방향으로 오류를 주석한다.

<예> 근데(√그런데) 특별한 명절이 있다.

☞ ‘근데’는 일반적으로 구어에서 자주 사용하는 접속부사로, 문어에서는 ‘그런데’를 사용하는 것이 더 자연스럽다. ‘근데’는 구어체라고 보고 문어에서 사용했을 경우, 구어/문어 오류로 처리한다.

이거(√이것은) 내 꿈이다.

☞ 해라체를 사용한 문어 텍스트에서 조사를 동반하지 않은 구어형 ‘이거’가 사용되었으므로 담화 층위에서 다른 문장과 어울리지 않으므로 구어/문어 오류로 처리한다.

- ‘하고’, ‘한테’ 등을 구어적 표현으로 보고, 문어에서 사용했을 경우 오류로 주석한다. 구어/문어 오류에 해당하는 목록은 다음과 같다.

<예> 한테(√에게)
 하고(√와/과)
 거/게(√것/것이)
 아무거(√아무것)
 근데(√그런데)
 ☞ 위의 예시들을 문어에서 사용한 경우, 오류 위치에 해당 품사를 주석하고, [오류 양상-대치], [오류 층위-구어/문어 오류]로 처리한다.

5. 구어 오류 주석

1) 구어 오류 주석 기본 원칙

- 구어 자료의 경우, 문장으로 파악하지 않고 억양 단위로 끊어서 각 단위를 기준으로 오류를 식별하고 판정한다.

<예> 무슨 파티하면
 우리 학생들이.
 열심히 공부한=
 연세대학교 열심히 공부해서
 조금 피곤한,
 =것이에요.
 ☞ 이 경우 억양 단위로 끊어서 보면 크게 문제가 되지 않지만, 문장 단위로 보면 여러 가지 층위에서 오류 처리가 가능하며 일관된 기준에 의한 처리가 어렵다. 구어 자료는 문장 단위가 아닌 억양 단위를 기준으로 하여 오류를 식별하

고 판정한다.

- 말더듬거림은 오류로 처리하지 않는다. 다만, 전사 단계에서 특정 표시를 하므로 이를 통해 향후 검색이 가능하게 한다.
- 자기 수정 발화의 경우, 수정 전 앞부분의 발화는 오류로 주석하지 않는다. 수정 후 발화에 초점을 두고 오류 여부를 판정한다.

<예> 친= 친구가 한국 음식이:: = 음식을 좋아해서

☞ ‘친= 친구’와 같은 말더듬거림은 오류로 처리하지 않는다.
‘음식이:: = 음식을’과 같이 수정 전 발화에서 조사를 잘못 사용하였지만, 다시 수정하여 조사를 고쳐 제대로 사용한 경우에 앞부분 ‘음식이’는 오류로 처리하지 않는다.

- 구어 오류 주석과 문어 오류 주석의 기본 원칙 및 처리 방법은 동일하다. 그러나 발화 상에서 나타나는 발음 오류의 경우에는 오류 양상(대치, 누락, 첨가, 오형태)을 주석하지 않고, 오류 위치와 오류 층위 [발음]만 주석한다.
- 구어에서는 발음 오류와 어휘 및 문법 오류의 구분이 명확하지 않을 수 있다. 즉, 학습자가 어휘와 문법을 잘못 사용한 것인지 단순히 발음을 잘못된 것으로 인해 나타난 오류인지 판별하는 데 어려움이 있다. 구어에서는 발음의 영향과 함께 어휘 및 문법 오류를 표시해주는 차원에서 오류 층위에서 중복 주석을 하도록 한다. 그러나 조사의 경우, 문법 오류를 우선 처리하도록 하고, 관형사형 전성어미를 사용해야 할 자리에 사용하지 못했을 경우도 받침을 발음하지 못한 음소 오류보다는 누락 오류로 문법 오류를 우선 처리한다.
- 구어 오류 주석에서의 또 다른 쟁점은 구어의 특성으로 볼 수 있는 현실 발음과 준말을 오류로 처리해야 하는가이다. 예를 들어, ‘김밥[김밥]’으로 발음했을 때 ‘적절성’을 기준으로 하여 ‘한국어 모어 화자’와 다르게 발음한다는 차원에서 오류로 볼 수도 있을 것이다. 그러나 현실 발음은 그 기준을 확정하기가 쉽지 않다. 한국어 화자도 표준 발음으로 발음하는 경우가 있고, 현실 발음을 어느 범위까지 인정해야 하는지도 문제가 되기 때문에 본 연구에서는 현실 발음에 어긋난다고 해서 오류로 처리하지는 않는다. 반대로 현실

발음을 인정해 일반적으로 한국어 모어 화자에서도 많이 나타나는 발음일 경우에 오류로 처리하지 않는다.

- 즉, 구어에서는 한국어 모어 화자들의 현실 발음을 고려하여 일반적으로 많이 사용되며, 구어에서 허용되는 형태는 오류로 처리하지 않는다.

<예> 할려고(√하려고)

☞ ‘ㄹ’로 시작하는 단어 앞에 받침 ‘ㄹ’을 첨가하여 발음하는 것은 한국인 모어 화자에게서도 많이 나타나는 현상이다. 현실 발음을 고려하여 이러한 경우는 오류로 처리하지 않는다.

[그리구](√그리고)::, 음::

☞ ‘그리고’를 [그리구]라고 발음하는 것은 한국어 모어 화자들에게도 많이 나타난다. 이처럼 한국어 모어 화자들의 현실 발음을 고려하여, 쫌, [바래요](바라요) 등 구어에서 허용되는 발음은 오류로 처리하지 않는다.

- 구어에서는 준말이 용인가능하기 때문에 오류로 보기 어려운 측면이 있다. 그러나 모든 준말을 허용할 수 있는 것은 아니기 때문에 구어에서 준말의 오류 판단 기준이 필요하다. 이에 따라 본 연구에서는 <표준국어대사전>에서 ‘~의 준말’로 등재되어 있는 것을 기준으로 삼는다. <표준>을 기준으로 ‘준말’로 등재되어 있는 형태는 오류로 처리하지 않고, 등재되지 않은 형태는 오류로 처리한다.

<예> 그래서 맘 먹고 여기 왔어요.

☞ ‘맘’의 경우, <표준>에 ‘마음의 준말’로 등재되어 있다. 따라서 오류로 처리하지 않는다.

[그쵸](√그렇죠)::, 음::

☞ 그러나 <표준>에 등재되어 있지 않더라도, ‘그쵸’와 같이 구어에서 축약된 형태로 많이 나타나는 용례들은 오류로 처리하지 않는다.

- 구어에서 발음의 차원이 아닌, 형태를 잘못 발화한 경우는 오형태(MIF) 오류로 처리한다.

<예> 사잉(√사건)
보석필(√보살핌)
☞ 이와 같은 예시들은 발음을 잘못했다기보다 형태를 잘못 만들어낸 것이다. 유사 발음과도 떨어져 음소 오류로 볼 수 없고, 한국어에 없는 형태들을 발음했다고 판단하여 오형태 오류로 처리한다.

- 구어에서 조사의 생략이나 축약된 형태의 사용은 한국어에서 일반적으로 나타나는 현상이다. 또한 구어 오류 주석을 억양 단위로 했을 경우, 조사의 생략은 자연스럽고, 용인 가능성이 문어에 비해 높아지기 때문에 구어에서는 이러한 형태를 용인 가능한 것으로 보고, 엄격하게 처리하지 않도록 한다.

<예> 제 한국 생활(√생활은)
아주 재미있고
한국도 좋아요
☞ 구어에서는 조사 생략이 자연스러운 경우가 있기 때문에 엄격하게 잡지 않도록 한다. 조사 생략을 용인 가능한 것으로 보고 오류로 처리하지 않는다.

- 구어 전사 시, 분명하게 들리지 않아서 <X X>로 처리한 부분은 분석불능(IMP)으로 주석한다.

<예> 그:: 마약::, <X청국::죄::X>라는 의미는, 한국에:: 마약, 없는, 뜻입니다. 한국에서:: 그:: 마약은 불법이라서,
그리고::, 어:: 그 뒤,에는 그 <X흔들이(흔들)X>라는 단어도..
있어서, 그::
☞ <X X> 부분은 전사가 정확하게 들지 못한 부분을 전사

한 것으로 판단이 불분명하기 때문에 분석불능으로 처리한다.

- 구어 오류는 오류 층위에서 발음 오류와 가장 밀접하다. 오류 층위 [발음]에는 음소(PP), 음절(PS), 음운규칙(PC), 원어식 발음(PN), 중간 발음(변이음 포함(PA)) 총 5가지가 있는데, 그중 ‘음소, 음절, 음운규칙’ 3가지를 우선적으로 주석한다. 원어식 발음과 중간 발음은 오류의 원인에 해당하는 문제이기 때문에 외래어(및 모국어 화자의 원어식 발음)의 경우에만 ‘원어식 발음’ 오류로 주석하고, 변이음(음성대치)이 분명하게 식별될 경우에만 ‘중간 발음’ 오류로 주석한다. 변이음 식별이 분명하지 않은 경우에는 주석하지 않는다.(세부 처리 방법은 ‘3. 범주별 세부 오류 유형의 처리, 4) 오류 층위, (1) 발음’을 참고한다.)

<예> [보콩](√복용),하는 뜻이는(√뜻은)::, 그 마약을 쓰는 아니면 마약을, 하,는:: 것입니다.

☞ 복용을 [보콩]으로 발음해야 하나 [보콩]으로 발음하였으므로 음소 오류로 처리한다. 오류 층위[발음]에 해당하는 오류이기 때문에 오류 양상은 주석하지 않고, 오류 위치와 오류 층위만 주석한다. 따라서 [오류 위치-명사], [오류 양상-없음(빈칸)], [오류 층위-음소]로 주석한다.

또한 ‘뜻이는’은 ‘뜻은’에 조사 ‘이’를 첨가한 것이기 때문에 발음의 오류가 아닌, 문어와 마찬가지로 조사 첨가 오류로 처리한다. 이 경우, [오류 위치-주격조사], [오류 양상-첨가]로 주석한다.

<부록> 표현 문형 목록

표제어	형태 정보	대표형
-게 되다		###게 되다
-게 마련이다		###게 마련이다
-게 만들다		###게 만들다
-게 생겼다		###게 생겼다
-게 하다		###게 하다
-고 나다		###고 나다
-고 들다		###고 들다
-고 말다		###고 말다
-고 보다		###고 보다
-고 싶다		###고 싶다
-고 싶어 하다		###고 싶어 하다
-고 있다		###고 있다
-고 해서		###고 해서
-고는 하다		###고는 하다
-곤 하다		###곤 하다
-기 나름이다		###기 나름이다
-기 때문		###기 때문
-기 마련이다		###기 마련이다
-기 십상이다		###기 십상이다
-기 위한		###기 위한
-기 위해(서)		###기 위해(서)
-기 일쑤이다		###기 일쑤이다
-기 전에		###기 전에
-기 짝이 없다		###기 짝이 없다
-기가 무섭게		###기가 무섭게
-기가 바쁘게		###기가 바쁘게
-기가 쉽다		###기가 쉽다
-기나 하다		###기나 하다
-기로 들다		###기로 들다
-기로 하다		###기로 하다

표제어	형태 정보	대표형
-기만 하다		###기만 하다
-기에 따라		###기에 따라
-기에 앞서(서)		###기에 앞서(서)
-ㄴ 것		
-ㄴ 것 같다		
-ㄴ 결과		
-ㄴ 김에		
-ㄴ 나머지		
-ㄴ 대로1		
-ㄴ 대로2		
-ㄴ 대신에		
-ㄴ 데요		
-ㄴ 듯		
-ㄴ 듯하다		
-ㄴ 마당에		
-ㄴ 모양이다		
-ㄴ 법이다		
-ㄴ 이상		
-ㄴ 줄		
-ㄴ 지2		
-ㄴ 채로		
-ㄴ 척하다		
-ㄴ 탓		
-ㄴ 편이다		
-ㄴ 후에		
-ㄴ가 보다		
-ㄴ다는 것이		
-ㄴ데도 불구하고		
-나 보다		###나 보다
-나 싶다		###나 싶다
-는 가운데		###는 가운데

표제어	형태 정보	대표형
-는 것		###는 것
-는 것 같다		###는 것 같다
는 고사하고	은 고사하고	
-는 길에		###는 길에
-는 김에		###는 김에
-는 대로	###은 대로, ###, ㄴ대로	
-는 대신에	###은 대신에, ###ㄴ 대신에	
-는 덕분에/이다		
-는 데다가	###은 데다가, ###ㄴ 데다가	
-는 도중에		
-는 동시에		###는 동시에
-는 동안		###는 동안
-는 등 마는 등		###는 등 마는 등
-는 듯	###은 듯, ###ㄴ 듯	
-는 듯하다	###은 듯하다, ###ㄴ 듯하다	
-는 마당에	###은 마당에, ###ㄴ 마당에	
-는 만큼	###은 만큼, ###ㄴ 만큼	
는 말할 것도 없고		
-는 모양이다	###은 모양이다, ###ㄴ 모양이다	
는 물론	###은 물론	
-는 바람에		###는 바람에
-는 반면에	###은 반면에, ###ㄴ 반면에	
-는 법이다	###은 법이다, ###ㄴ 법이다	
-는 사이		###는 사이
-는 수밖에 없다		
-는 이상	###은 이상, ###ㄴ 이상	
-는 적이 있다/없다		###는 적이 있다/없다

표제어	형태 정보	대표형
-는 줄	###은 줄, ###ㄴ 줄	
-는 중이다		###는 중이다
-는 척하다	###은 척하다, ###ㄴ 척하다	
-는 체하다		
-는 탓	###은 탓, ###ㄴ 탓	
-는 통에		###는 통에
-는 편이다	###은 편이다, ###ㄴ 편이다	
-는 한		###는 한
-는 한이 있어도/있더라도		###는 한이 있어도/있더라도
-는 한편		###는 한편
-는가 보다	###은가 보다, ###ㄴ가 보다	
-는다는 것이	###ㄴ다는 것이	
-는데도 불구하고	###은데도 불구하고, ###ㄴ 데도 불구하고	
-도록 하다		###도록 하다
-ㄴ 것 같다		
-ㄴ 것1		
-ㄴ 것2		
-ㄴ 것이 아니라		
-ㄴ 대로		
-ㄴ 듯		
-ㄴ 듯하다		
-ㄴ 따름이다		
-ㄴ 때		
-ㄴ 리가 없다		
-ㄴ 만큼		
-ㄴ 만하다		
-ㄴ 모양이다		
-ㄴ 바에		
-ㄴ 법하다		

표제어	형태 정보	대표형
-ㄴ 뻔하다		
-ㄴ 뿐만 아니라		
-ㄴ 수밖에 없다		
-ㄴ 줄		
-ㄴ 테고		
-ㄴ 테나		
-ㄴ 테니		
-ㄴ 테다		
-ㄴ 테면		
-ㄴ 테야		
-ㄴ 테지만		
-ㄴ 텐데		
-ㄴ까 보다		
-ㄴ락 말락 하다		
-려고 하다		
-려나 보다		
로 인하다	으로 인하다	
를 가지고	을 가지고	
를 막론하고	을 막론하고	
를 불문하고		
를 위해(서)	을 위해(서)	
만 같아도		만 같아도
만 아니면		만 아니면
-면 되다		
-면 몰라도		
-면 안 되다		
-면 좋겠다		
-아 가다	###어 가다, ###여 가다	
-아 가지고	###어 가지고, ###여 가지고	
-아 계시다		
-아 내다	###어 내다, ###여 내다	

표제어	형태 정보	대표형
-아 놓다	###어 놓다, ###여 놓다	
-아 대다	###어 대다, ###여 대다	
-아 두다	###어 두다, ###여 두다	
-아 드리다	###어 드리다, ###여 드리다	
-아 버리다	###어 버리다, ###여 버리다	
-아 보다	###어 보다, ###여 보다	
-아 보이다	###어 보이다, ###여 보이다	
-아 오다	###어 오다, ###여 오다	
-아 있다	###어 있다, ###여 있다	
-아 주다	###어 주다, ###여 주다	
-아 치우다	###어 치우다, ###여 치우다	
-아도 되다	###어도 되다, ###여도 되다	
-아서는 안 되다		
-아야 되다	###어야 되다, ###여야 되다	
-아야 하다	###어야 하다, ###여야 하다	
-어 가다		
-어 가지고		
-어 내다		
-어 놓다		
-어 대다		
-어 두다		
-어 드리다		
-어 버리다		
-어 보다		
-어 보이다		
-어 오다		
-어 있다		

표제어	형태 정보	대표형
-어 주다		
-어 치우다		
-어도 되다		
-어야 되다		
-어야 하다		
에 관하여		에 관하여
에 관한		에 관한
에 대하여		
에 대한		
에 따라		에 따라
에 따르면		에 따르면
에 비하여		에 비하여
에 의하면		에 의하면
에 의하여		에 의하여
에도 불구하고		에도 불구하고
-여 가다		
-여 가지고		
-여 내다		
-여 놓다		
-여 대다		
-여 두다		
-여 드리다		
-여 버리다		
-여 보다		
-여 보이다		
-여 오다		
-여 있다		
-여 주다		
-여 치우다		
-여도 되다		
-여야 되다		

표제어	형태 정보	대표형
-여야 하다		
-으려고 하다	###려고 하다	
-으려나 보다	###려나 보다	
으로 인하다	로 인하다	
-으면 되다	###면 되다	
-으면 몰라도	###면 몰라도	
-으면 안 되다	###면 안 된다	
-으면 좋겠다	###면 좋겠다	
-은 가운데		###은 가운데
-은 것	###ㄴ 것	
-은 것 같다	###ㄴ 것 같다	
-은 결과	###ㄴ 결과	
은 고사하고		
-은 김에	###ㄴ 김에	
-은 나머지	###ㄴ 나머지	
-은 다음에	###ㄴ 다음에	
-은 다음에야	###ㄴ 다음에야	
-은 대로1	###ㄴ 대로, ###는 대로	
-은 대로2	###ㄴ 대로, ###는 대로	
-은 대신에	###ㄴ 대신에, ###는 대신에	
-은 데다가1	###ㄴ 데다가	
-은 데다가2	###ㄴ 데다가, ###는 데다가	
-은 뒤에		
-은 듯	###ㄴ 듯, ###는 듯	
-은 듯하다	###ㄴ 듯하다, ###는 듯하다	
-은 마당에	###ㄴ 마당에, ###는 마당에	
-은 만큼	###ㄴ 만큼, ###는 만큼	
-은 모양이다	###ㄴ 모양이다, ###는 모양이다	
은 물론	는 물론	

표제어	형태 정보	대표형
-은 반면에	###ㄴ 반면에, ###는 반면에	
-은 법이다	###ㄴ 법이다, ###는 법이다	
-은 이상	###ㄴ 이상, ###는 이상	
-은 줄	###ㄴ 줄, ###는 줄	
-은 지2	###ㄴ 지	
-은 채로	###ㄴ 채로	
-은 척하다		###은 척하다
-은 체하다		
-은 탓	###ㄴ 탓, ###는 탓	
-은 편이다	###ㄴ 편이다, ###는 편이다	
-은 후에	###ㄴ 후에	
-은가 보다	###ㄴ가 보다, ###는가 보다	
-은데도 불구하고	###ㄴ데도 불구하고, ###는데도 불구하고	
을 가지고	###르 가지고	
-을 것 같다	###르 것 같다	
-을 것1	###르 것	
-을 것2	###르 것	
-을 것이 아니라	###르 것이 아니라	
-을 나름이다		
-을 대로	###르 대로	
-을 듯	###르 듯	
-을 듯하다	###르 듯하다	
-을 따름이다	###르 따름이다	
-을 때	###르 때	
-을 리가 없다	###르 리가 없다	
-을 리가 있다		
을 막론하고	를 막론하고	
-을 만큼	###르 만큼	
-을 만하다	###르 만하다	

표제어	형태 정보	대표형
-을 모양이다	###르 모양이다	
-을 바에	###르 바에	
-을 법하다	###르 법하다	
을 불문하고		
-을 뻔하다	###르 뻔하다	
-을 뿐만 아니라	###르 뿐이다	
-을 뿐이다		
-을 수 없다		
-을 수 있다		
-을 수밖에 없다	###르 수밖에 없다	
을 위해(서)	###를 위해(서)	
-을 줄	###르 줄	
-을 데고	###르 데고	
-을 데냐	###르 데냐	
-을 데니	###르 데니	
-을 데니까		
-을 데다	###르 데다	
-을 데면	###르 데면	
-을 데야	###르 데야	
-을 데지만	###르 데지만	
-을 텐데	###르 텐데	
-을까 보다	###르까 보다	
-을락 말락 하다	###르락 말락 하다	
-지 말다		###지 말다
-지 못하다		###지 못하다
-지 않다		###지 않다

2023-2024년 학습자 동의서 양식

1. 성별: ☐ F ☐ M
2. 출생년: _____ 년 (예. 1989년)
3. 현재 등급: _____ (TOPIK: _____)
4. 국적: _____ (※ 교포 여부 ☐ 교포 ☐ 외국인)
5. 제1 언어: _____
6. 한국어 학습 기간(한국어를 얼마 동안 공부했습니까?): _____ 년 _____ 개월
(예. 1년 3개월)
7. 한국에서의 거주 기간(한국에서 얼마 동안 살았습니까?): _____ 년 _____ 개월
(예. 1년 3개월)
8. 한국어 학습 목적
☐ 진학 ☐ 취업 ☐ 거주 ☐ 취미 ☐ 결혼 ☐ 기타 (_____)
9. 직업: _____
10. 한국어 외의 사용 가능 외국어(잘하는 언어 순서대로 쓰시오):

(한국어)

**Letter of permission to use learners' language materials for
constructing and utilizing Korean language learner corpus /
Consent to collection and use of personal information and provision to third parties**

The National Institute of Korean Language promotes a project that collects language data (corpus) from Korean learners and uses this set to improve the quality of Korean education. The materials one provides are used for research and development in various fields such as improving Korean language teaching methods, developing Korean language textbooks, Korean language education, and AI model development. Participants of this project experience no economic or physical harm. One may choose to withdraw their participation if required. The personal information of participants will not be used for any other purpose, and any identifiable information will remain confidential. Please read the contents carefully and decide whether you agree.

■ **Permission to use learner's materials**

Writing materials are provided to build and utilize the Korean language learner corpus, and the disclosure of all original writing materials and their use for research and development purposes is permitted.

☞ **Do you agree to the permission to use learners' language materials as above?**

☐ Agree ☐ Disagree

■ **Details of collection and use of personal information**

Items	Purpose of collection and use	Retention/use period
Gender, Year of Birth, Current Level of Korean, Nationality, Native Language, Time of Studying Korean Language, Time of Residence in Korea, Purpose of Studying Korean Language, Occupation, Any Other Foreign Language That You Can Speak	National Institute of Korean Language Korean learner corpus construction and utilization	<u>30 years from the date of consent to the collection and use of personal information</u> * However, if personal information is requested to be disposed of, it will be deleted immediately.

※ You have the right to refuse consent to the collection and use of personal information above. However, if you refuse to provide consent, you may be restricted from participating in the Korean learner corpus construction/utilization project.

☞ **Do you agree to the collection and use of personal information as above?**

☐ Agree ☐ Disagree

■ **Details of provision of personal information to third parties**

Recipient	Purpose of provision	Provided items	Retention/use period
<u>Contractor for use the Korean language learner corpus of National Institute of Korean Language</u>	<u>Korean language information processing research and development, Korean language research and development</u>	Gender, Year of Birth, Current Level of Korean, Nationality, Native Language, Time of Studying Korean Language, Time of Residence in Korea, Purpose of Studying Korean Language, Occupation, Any Other Foreign Language That You Can Speak	<u>30 years from the date of consent to the collection and use of personal information</u> * However, if personal information is requested to be disposed of, it will be deleted immediately.

※ You have the right to refuse consent to the collection and use of personal information above. However, if you refuse to provide consent, you may be restricted from participating in the Korean learner corpus construction/utilization project.

☞ **Do you agree to provide personal information to third parties as above?**

☐ Agree ☐ Disagree

■ **Details of consignment of personal information handling (Notice)**

Consignee (consignment company)	Business details
IIR TECH Inc. / Yonsei University-Industry Foundation	Construction of the Korean learner corpus in 2024

Date: _____

Name: _____ (Signature)

To the National Institute of Korean Language

The questionnaire includes your personal information that would help us conduct the research analysis. Strict confidentiality will be maintained at all times. (Please fill out the form in Korean if you can. This is also recommended in English.)

1. Gender: ☐ F ☐ M

2. Year of Birth: _____ (eg. 1989)

3. Current Level of Korean: _____ (TOPIK: _____)

4. Nationality : _____

(※ I am a/an Overseas Korean ☐ Foreigner ☐)

5. Native Language : _____

6. Time of Studying Korean Language (How long you've been studying?) :

_____ years _____ months (eg. 1year 8months)

7. Time of Residence in Korea How long you've been staying in Korea?) :

_____ years _____ months (eg. 1year 10months)

8. Purpose of Studying Korean Language :

☐ for admission to the universities in Korea ☐ for a job

☐ for residence in Korea ☐ for a hobby ☐ for marriage

☐ others (_____)

9. Occupation : _____

10. Any Other Foreign Languages That You Can Speak

(Note: If there are multiple languages, please list them in the order of fluency) :

(영어)

为构建和利用韩国语学习者语料库的学习者语言资料使用许可书,
个人信息收集、使用和第三方提供同意书

为了提高韩国语教育的质量,国立国语院正在推进收集和利用韩国语学习者语言资料(语料库)的项目。您提供的资料将用于改善韩国语教学方法、开发韩国语教材、韩国语教育以及人工智能模型等多个领域的研究和开发。参与此研究的人员不会遭受经济损失或人身伤害。如果您不愿意参与,可以随时撤回参与意愿。此外,收集的个人信息将不会用于本项目目的以外的用途,并将以无法识别的形式使用以便保密。请仔细阅读内容后决定是否同意。

■ 学习者资料的使用许可

为了韩国语学习者语料库的建构和利用,提供写作资料,并允许公开完整的写作原文资料并以研究开发为目的加以利用。

☞ 您是否同意如上所述使用学习者语言资料的使用许可? ☐ 同意 ☐ 不同意

■ 个人信息收集及使用详情

项目	收集和使用的目的	保留/使用期限
性别、出生年份、现TOPIK等级、国籍、第一语言、韩国语学习时间、韩国居住时间、韩国语学习目的、职业、韩国语以外其他可用外语	韩国国立国语院韩国语学习者语料库的构建与利用	自同意收集和使用个人信息之日起 30 年 *但要求销毁个人信息时,将立即删除

※ 您有权拒绝上述个人信息收集和使用的同意。但如果拒绝同意,您可能会被限制参与韩国语学习者语料库的构建与利用项目。

☞ 您是否同意如上所述个人信息的收集和使用的同意? ☐ 同意 ☐ 不同意

■ 向第三方提供个人信息的详情

接受者	提供目的	提供项目	保留/使用期限
韩国国立国语院韩国语学习者语料库使用承包商	韩国语信息处理研究开发 , 韩国语研究开发	性别、出生年份、现TOPIK等级、国籍、第一语言、韩国语学习时间、韩国居住时间、韩国语学习目的、职业、韩国语以外其他可用外语	自同意收集和使用个人信息之日起 30 年 *但要求销毁个人信息时,将立即删除

※ 您有权拒绝上述个人信息收集和使用的同意。但如果拒绝同意,您可能会被限制参与韩国语学习者语料库的构建与利用项目。

☞ 您是否同意如上所述向第三方提供个人信息? ☐ 同意 ☐ 不同意

■ 个人信息处理业务委托详情(通知事项)

受托方(受托企业)	业务内容
IIR TECH Inc. / 延世大学产学合作团	2024年韩国语学习者语料库构建

年 ____ 月 ____ 日 ____
姓名 _____(签名或印)

致韩国国立国语院

以下信息将作为研究资料使用,个人信息将被保密,不会外泄。(请尽可能用韩国语回答,必要时可以使用英语。)

1. 性别: ☐ 女 ☐ 男
2. 出生年份: _____ 年(例如: 1989年)
3. 现韩国语等级: _____ (TOPIK: _____)
4. 国籍: _____ (※ 侨胞与否 ☐ 侨胞 ☐ 外国人)
5. 第一语言: _____
6. 韩国语学习时间 (学习韩国语多久了?): _____ 年 _____ 个月 (例如: 一年零三个月)
7. 在韩国的居住时间 (在韩国居住多久了?): _____ 年 _____ 个月 (例如: 一年零三个月)
8. 学习韩国语的目的:
☐ 升学 ☐ 就业 ☐ 居住 ☐ 兴趣 ☐ 结婚 ☐ 其它 (_____)
9. 职业: _____
10. 韩国语以外其他可用外语 (按照熟练程度从高到低罗列):

(중국어)

韓国語学習者言語資料構築・活用のための学習者言語資料利用承諾書
個人情報収集・利用及び第三者提供同意書

国立国語院では韓国語教育の質的向上のため学習者様の言語資料（コーパス）を収集・活用する事業を推進しております。みなさまよりご提供いただいた資料は韓国語教授方法の改善、韓国語教材開発、韓国語教育及び人工知能モデルの開発等、さまざまな分野の研究・開発に使用されます。この研究に参加することによる経済的な損害や身体的危険はありません。ご参加をご希望されない場合は撤回することも可能です。又、収集された個人情報は本事業の目的のみに使用し保護のため識別できない形態で使用されます。以下の内容をご確認のうえご署名ください。

■ 学習者資料の利用許可

韓国語学習者言語資料構築・活用のために作文資料を提供し、作文原文資料全体の公開と研究・開発目的の利用を承諾します。

☞ 上記の学習者の言語資料の利用許可に同意しますか？ ☐ 同意 ☐ 同意しない

■ 個人情報収集・利用内訳

項目	収集・利用目的	保有・利用期間
性別、出生年、級、国籍、母国語、韓国語学習期間、韓国内居住期間、韓国語の学習目的、職業、使用可能な外国語	国立国語院 韓国語学習者の言葉の構築・活用	個人情報収集・利用同意日から30年 *ただし、個人情報の廃棄を要請した場合は直ちに破棄

※ 上記の個人情報の収集・利用に同意しない権利があります。 但し、同意しない場合、韓国語学習者言語資料構築・活用事業への参加が制限されることがあります。

☞ 上記の個人情報を収集・利用することに同意しますか？ ☐ 同意 ☐ 同意しない

■ 個人情報第三者提供内訳

提供される者	提供目的	提供項目	保有・利用期間
国立国語院韓国語学習者の言葉利用規約者	韓国語情報処理研究・開発 、 国語研究・開発	性別、出生年、級、国籍、母国語、韓国語学習期間、韓国内居住期間、韓国語の学習目的、職業、使用可能な外国語	個人情報収集・利用同意日から30年 *ただし、個人情報の廃棄を要請した場合は直ちに破棄

※ 上記の個人情報の収集・利用に同意しない権利があります。 但し、同意しない場合、韓国語学習者言語資料構築・活用事業への参加が制限されることがあります。

☞ 上記の個人情報を第三者に提供することに同意しますか？ ☐ 同意 ☐ 同意しない

■ 個人情報取扱業務委託内訳（告知事項）

委託される者（受託業者）	業務内容
IIR TECH Inc.・延世大学産学協力団	2024年の韓国語学習者言語資料の構築

日付 _____
氏名 _____ (署名)

国立国語院御中

以下は研究のための資料として活用される情報です。個人情報、秘密が保証され、外部に流出されることはありません。(出来れば韓国語でお答えください。英語でも構いません。)

1. 性別: ☐ 女 ☐ 男
2. 出生年: _____ 年(例. 1989年)
3. 級: _____(TOPIK: _____)
4. 国籍: _____(* 在日韓国人の場合は、☒をご記入してください。☐)
5. 母国語: _____
6. 韓国語学習期間(どれくらい勉強しましたか?) _____ 年 _____ ヲ月 (例. 1年 3ヵ月)
7. 韓国内居住期間(どれくらい住んでいましたか?) _____ 年 _____ ヲ月 (例. 1年 3ヵ月)
8. 韓国語の学習の目的:
☐ 進学 ☐ 就職 ☐ 居住 ☐ 趣味 ☐ 結婚 ☐ その他 (_____)
9. 職業: _____
10. 使用可能な外国語(使える言葉順にお書きください):

(일본어)

Giấy cho phép sử dụng tài liệu tiếng Hàn của người học để xây dựng và sử dụng kho ngữ liệu tiếng Hàn /

Đồng ý thu thập, sử dụng thông tin cá nhân và cung cấp cho bên thứ ba

Nhằm nâng cao chất lượng giáo dục tiếng Hàn, Viện quốc ngữ quốc gia Hàn Quốc đang tiến hành dự án thu thập và sử dụng tài liệu ngôn ngữ của học viên. Tài liệu học viên cung cấp được sử dụng để nghiên cứu và phát triển trong nhiều lĩnh vực khác nhau, chẳng hạn như cải tiến phương pháp giảng dạy tiếng Hàn, phát triển sách giáo khoa tiếng Hàn, giáo dục tiếng Hàn và phát triển mô hình trí tuệ nhân tạo. Những học viên tham gia vào việc nghiên cứu này sẽ hoàn toàn không có gì nguy hại đến tính mạng, hay tổn thất về mặt kinh tế. Tuy nhiên trong trường hợp học viên không muốn tham gia thì có thể huỷ việc tham gia. Ngoài ra, những thông tin cá nhân mà học viên cung cấp, chúng tôi cam kết không dùng vào bất kỳ mục đích nào khác ngoài dự án này, và để không làm tiết lộ thông tin cá nhân của học viên tài liệu được thu thập dưới dạng không thể xác định được. Vui lòng đọc kỹ nội dung và quyết định có đồng ý hay không.

■ **Đồng ý sử dụng tài liệu của người học**

Tôi đồng ý cung cấp tài liệu viết để xây dựng và sử dụng kho tài liệu học tiếng Hàn, đồng ý việc công khai toàn bộ tài liệu viết gốc và việc sử dụng với mục đích nghiên cứu và phát triển.

☞ **Bạn có đồng ý cho phép sử dụng tài liệu ngôn ngữ của người học như trên không?**

☐ Đồng ý ☐ Không đồng ý

■ **Nội dung chi tiết thu thập và sử dụng thông tin cá nhân**

Hạng mục	Mục đích thu thập và sử dụng	Thời gian lưu giữ/sử dụng
Giới tính, ngày tháng năm sinh, cấp Topik hiện tại, Quốc tịch, Tiếng mẹ đẻ, thời gian học tiếng Hàn, thời gian cư trú tại Hàn, mục đích của việc học tiếng Hàn, nghề nghiệp, có thể sử dụng ngôn ngữ khác ngoài tiếng Hàn.	Xây dựng và sử dụng kho ngữ liệu của người học tiếng Hàn tại Viện Ngôn ngữ Quốc gia	<u>30 năm kể từ ngày đồng ý thu thập và sử dụng thông tin cá nhân</u> <small>*Tuy nhiên, nếu thông tin cá nhân được yêu cầu xóa, nó sẽ bị xóa ngay lập tức.</small>

※ Học viên có quyền từ chối sự đồng ý cho việc thu thập và sử dụng thông tin cá nhân ở trên. Tuy nhiên, nếu từ chối đồng ý, học viên có thể bị hạn chế tham gia vào dự án xây dựng/sử dụng kho ngữ liệu dành cho người học tiếng Hàn.

☞ **Bạn có đồng ý việc thu thập và sử dụng thông tin cá nhân như trên không?**

☐ Đồng ý ☐ Không đồng ý

■ **Nội dung chi tiết cung cấp thông tin cá nhân cho bên thứ ba**

Người tiếp nhận	Mục đích cung cấp	Hạng mục cung cấp	Thời gian lưu giữ/sử dụng
<u>Người giao dịch sử dụng kho dữ liệu tiếng Hàn của người học của Viện quốc ngữ quốc gia Hàn Quốc</u>	<u>Nghiên cứu và phát triển xử lý thông tin tiếng Hàn,</u> <u>Nghiên cứu và phát triển ngôn ngữ Hàn Quốc</u>	Giới tính, ngày tháng năm sinh, cấp Topik hiện tại, Quốc tịch, Tiếng mẹ đẻ, thời gian học tiếng Hàn, thời gian cư trú tại Hàn, mục đích của việc học tiếng Hàn, nghề nghiệp, có thể sử dụng ngôn ngữ khác ngoài tiếng Hàn.	<u>30 năm kể từ ngày đồng ý thu thập và sử dụng thông tin cá nhân</u> <small>* Tuy nhiên, nếu thông tin cá nhân được yêu cầu xóa, nó sẽ bị xóa ngay lập tức.</small>

※ Học viên có quyền từ chối sự đồng ý cho việc thu thập và sử dụng thông tin cá nhân ở trên. Tuy nhiên, nếu từ chối đồng ý, học viên có thể bị hạn chế tham gia vào dự án xây dựng/sử dụng kho ngữ liệu dành cho người học tiếng Hàn.

☞ **Bạn có đồng ý cung cấp thông tin cá nhân cho bên thứ ba như trên không?**

☐ Đồng ý ☐ Không đồng ý

■ **Nội dung chi tiết ủy thác công việc xử lý thông tin cá nhân (Thông báo)**

Người được ủy thác (công ty ủy thác)	Nội dung công việc
Quỹ Hợp tác giáo dục và doanh nghiệp (주)Irtech và trường Đại học Yonsei	Xây dựng kho ngữ liệu của người học tiếng Hàn năm 2024

Ngày tháng năm
Họ và tên (ký tên)

Kính gửi Viện Ngôn ngữ Quốc gia Hàn Quốc

Sau đây là những thông tin được sử dụng làm tài liệu nghiên cứu. Thông tin cá nhân sẽ được đảm bảo bí mật và không bị tiết lộ ra bên ngoài. (Nếu có thể, vui lòng trả lời bằng tiếng Hàn hoặc khi cần có thể trả lời bằng tiếng Anh)

1. Giới tính : ☐ F ☐ M

2. Năm sinh : _____ năm (Ví dụ: năm 1989)

3. Cấp học hiện tại : _____ (TOPIK: _____)

4. Quốc tịch : _____

(※ Có phải là Hàn kiều hay không ☐Hàn kiều ☐ Người nước ngoài)

5. Tiếng mẹ đẻ (ví dụ: tiếng Việt, tiếng Trung) : _____

6. Thời gian học tiếng Hàn (Bạn đã học tiếng Hàn được bao lâu rồi?): _____ năm _____ tháng
(Ví dụ: 1 năm 3 tháng)

7. Thời gian cư trú tại Hàn Quốc (Bạn đã sống ở Hàn Quốc được bao lâu rồi?) : _____ năm
tháng

(Ví dụ: 1 năm 3 tháng)

8. Mục đích của việc học tiếng Hàn :

☐ Học lên cao ☐ Đi làm ☐ Sống ở Hàn ☐ Giải trí

☐ Kết hôn với người Hàn ☐ Nội dung khác _____

9. Nghề nghiệp : _____

10. Có thể sử dụng ngôn ngữ khác ngoài tiếng Hàn

(Vui lòng ghi theo trình tự từ ngoại ngữ giỏi đến kém hơn) :

(베트남어)

หนังสืออนุญาตให้ใช้ข้อมูลทางภาษาของผู้เรียน
เพื่อการจัดการและการทำงานคลังข้อมูลผู้เรียนภาษาเกาหลี
โดยแสดงความยินยอมให้รวบรวม ใช้ข้อมูลส่วนบุคคล
และมอบข้อมูลให้กับบุคคลที่ 3

สถาบันภาษาเกาหลีแห่งประเทศไทยส่งเสริมโครงการรวบรวมข้อมูลทางภาษาของผู้เรียนภาษาเกาหลี (คลังข้อมูล) เพื่อการยก ระดับคุณภาพของการศึกษาภาษาเกาหลี โดยข้อมูลที่ได้รับจากทุกท่านจะถูกนำไปใช้ในการวิจัยและพัฒนาในสาขาต่าง ๆ เช่น การพัฒนาวิธีการสอนภาษาเกาหลี การพัฒนาหนังสือเรียนภาษาเกาหลี การพัฒนารูปแบบปัญญาประดิษฐ์ หรือการเรียนการสอน ภาษาเกาหลี เป็นต้น ซึ่งผู้เข้าร่วมโครงการวิจัยนี้ ไม่ได้รับความเสียหายทางเศรษฐกิจหรืออันตรายทางด้านกายภาพใด ๆ อันใด และหากท่านไม่ต้องการเข้าร่วมโครงการวิจัยนี้ ท่านสามารถขอยกเลิกการเข้าร่วมได้ สำหรับข้อมูลส่วนบุคคลที่รวบรวมจะไม่ถูกนำไปใช้เพื่อวัตถุประสงค์อื่นใดนอกเหนือจากวัตถุประสงค์ของโครงการวิจัยนี้ และข้อมูลส่วนบุคคลจะถูกใช้ในรูปแบบที่ไม่สามารถระบุตัวตนได้จะเป็นการรักษาความลับ ทั้งนี้ กรุณาอ่านเข้าใจละเอียดและแล้วท่านด้วยหรือไม่

■ การอนุญาตให้ใช้ข้อมูลทางภาษาของผู้เรียน

ข้าพเจ้าเป็นต้นสนให้ข้อมูลงานเขียนภาษาเกาหลีเพื่อการจัดทำและการใช้งานคลังข้อมูลผู้เรียนภาษาเกาหลี อนุญาตให้เปิดเผยข้อมูลของต้นฉบับงานเขียนภาษาเกาหลีทั้งหมด และอนุญาตให้ใช้ข้อมูลฯ ในวัตถุประสงค์เพื่อการวิจัยหรือพัฒนา

👉 ท่านอนุญาตและยินยอมให้ใช้ข้อมูลทางภาษาของนักเรียนดังรายละเอียดข้างต้นหรือไม่

☐ ยืนยันยอม

☐ ไม่ยินยอม

■ รายละเอียดการรวบรวมและการนำข้อมูลส่วนบุคคลไปใช้

รายการข้อมูล	วัตถุประสงค์การรวบรวม และการนำไปใช้	ระยะเวลาการเก็บข้อมูล และการนำไปใช้
เพศ, ปีเกิด, ระดับ TOPIK ณ ปัจจุบัน, สัญชาติ, ภาษาที่ 1, ระยะเวลาเรียนภาษาเกาหลี , ระยะเวลาที่อยู่ในประเทศเกาหลี, วัตถุประสงค์ของการเรียนภาษาเกาหลี, อาชีพ, ภาษาต่างประเทศอื่น ๆ ที่ท่านสามารถใช้ได้ นอกเหนือจากภาษาเกาหลี	เพื่อการจัดทำ และการใช้งานคลังข้อมูล ผู้เรียนภาษาเกาหลี ในสถาบันภาษาเกาหลี แห่งชาติ	30 ปี นับจากวันที่ <u>ได้รับการยินยอมให้รวบรวมและใช้ข้อมูลส่วนบุคคล</u> *กรณี ข้อมูลส่วนบุคคลจะถูกลบใน ๒ ปีถ้ามีการแจ้งขอให้ทำลายข้อมูลดังกล่าว

※ ทำหนังสือในการปฏิเสธให้หน่วยงานยินยอมเกี่ยวกับการรวบรวมและการใช้ข้อมูลส่วนบุคคลด้านบน แต่กรณีที่ไม่ปฏิเสธการยินยอม ดังกล่าว หน่วยงานได้รับการจำกัดการเข้าร่วมโครงการจัดทำและการใช้งานคลังข้อมูลฝั่เรียนภาษาเกาหลี

👉 ท่านอนุญาตและยินยอมให้รวบรวมและใช้ข้อมูลส่วนบุคคลดังรายละเอียดข้างต้นหรือไม่

☐ ยินยอม

☐ ไม่ยินยอม

■ รายละเอียดการมอบข้อมูลส่วนบุคคลให้กับบุคคลที่ 3

ผู้รับข้อมูล	วัตถุประสงค์ ในการมอบข้อมูล	รายการข้อมูลที่ได้รับมอบ	ระยะเวลาการเก็บข้อมูล และการนำไปใช้
ผู้ทำสัญญา การใช้คลังข้อมูล ผู้เรียนภาษาเกาหลี โดย สถาบันภาษาเกาหลี แห่งชาติ	เพื่อ การวิจัยและพัฒนา การจัดการข้อมูลทาง ด้านภาษาเกาหลี, การวิจัยและพัฒนา ภาษาเกาหลี	เพศ, ปีเกิด, ระดับ TOPIK ณ ปัจจุบัน, สัญชาติ, ภาษาที่ 1, ระยะเวลาเรียนภาษาเกาหลี , ระยะเวลาอยู่ในประเทศเกาหลี , วัตถุประสงค์ของการเรียนภาษาเกาหลี, อาชีพ, ภาษาต่างประเทศอื่น ๆ ที่ทํานานสามารถใช้ได้ นอกเหนือจากภาษาเกาหลี	<u>30 ปี</u> <u>นับจากวันที่ ได้รับการ</u> <u>ยินยอมให้รวบรวม</u> <u>และใช้ข้อมูลส่วนบุคคล</u> <u>*กรณีที่</u> ข้อมูลส่วนบุคคลจะถูกใช้ในกรณีที่ มีการแจ้งขอให้ทำลาย ข้อมูลดังกล่าว

※ ทำหนังสือในการปฏิเสธให้หน่วยงานยินยอมเกี่ยวกับการรวบรวมและการใช้ข้อมูลส่วนบุคคลด้านบน แต่กรณีปฏิเสธการยินยอม ดังกล่าว ท

👉 ท่านอนุญาตและยินยอมให้มอบข้อมูลส่วนบุคคลให้แก่บุคคลที่ 3 ดังรายละเอียดข้างต้นหรือไม่

ជំនួយ

☐ ໓. ໃບຢັ້ງຢືນ

■ รายละเอียดการมอบหมายให้จัดการข้อมูลส่วนบุคคล(ประกา

ผู้รับมอบ(บริษัท) รับมอบการจัดการข้อมูล	รายละเอียดหน้าที่ ที่ รับมอบ
IIR TECH Inc. และ University Industry Foundation, Yonsei University	จัดทำคลังข้อมูลผู้เรียนภาษาเกาหลี ปี 2024

วันที่ เดือน ปี
 ชื่อ-นามสกุล (ลายเซ็น)
 ขอมอบเอกสารฉบับนี้ให้กับสถาบันภาษาเกาหลีแห่งชาติ

เนื้อหาด้านล่างนี้เป็นข้อมูลที่จะถูกนำไปใช้ในการทำวิจัย โดยข้อมูลส่วนบุคคลจะถูกรักษาเป็นความลับและไม่ถูกนำไปเผยแพร่สู่ภายนอก (หากเป็นไปได้กรุณาตอบด้วยภาษาเกาหลี หรือสามารถใช้ภาษาอังกฤษในการตอบได้เมื่อจำเป็น)

1. เพศ : ☐ F ☐ M
2. ปีเกิด : ปี ค.ศ. _____ (ตัวอย่าง ปี 1989)
3. ระดับภาษาเกาหลี ณ ปัจจุบัน : _____ (TOPIK: _____)
4. สัญชาติ : _____
(* เป็นชาวเกาหลีโพ้นทะเลหรือไม่ ☐ ชาวเกาหลีโพ้นทะเล ☐ ชาวต่างชาติ)
5. ภาษาที่ 1 : _____
6. ระยะเวลาที่เรียนภาษาเกาหลี(ท่านเรียนภาษาเกาหลีเป็นระยะเวลานานเท่าใด) : _____ ปี _____ เดือน
(ตัวอย่าง 1 ปี 3 เดือน)
7. ระยะเวลาที่อยู่ในประเทศเกาหลี(ท่านพำนักอาศัยอยู่ในประเทศเกาหลีเป็นระยะเวลานานเท่าใด)
: _____ ปี _____ เดือน
(ตัวอย่าง 1 ปี 3 เดือน)
8. วัตถุประสงค์ของการเรียนภาษาเกาหลี
☐ ศึกษาต่อ ☐ ประกอบอาชีพ ☐ พำนักราชย์ ☐ งานอดิเรก ☐ แต่งงาน ☐ อื่น ๆ (_____)
9. อาชีพ : _____
10. ภาษาต่างประเทศอื่น ๆ ที่ท่านสามารถใช้ได้ นอกเหนือจากภาษาเกาหลี
(กรุณาเรียงลำดับตามภาษาที่ใช้ได้ดี) : _____

(태국어)

**Carta de permiso para utilizar materiales lingüísticos de los estudiantes para
construir y utilizar corpus de estudiantes del idioma coreano /
Consentimiento para la recopilación y el uso de información personal y su suministro a terceros**

El Instituto Nacional de la Lengua Coreana promueve el proyecto, cual recoge datos lingüísticos (Corpus) de los estudiantes de coreano y utiliza el conjunto con el fin de mejorar la calidad de la educación del idioma coreano. Los materiales que tu proporcionas se utilizarán para la investigación y el desarrollo en diversos campos, como la mejora de los métodos de enseñanza del idioma coreano, el desarrollo de libros de texto en idioma coreano, la educación del idioma coreano y el desarrollo de modelos de inteligencia artificial. Los que participan en este proyecto no tendrá ningún daño económico ni daño físico. Cuando no desees participar más, puedes optar por retirar tu participación. Tu información personal no será utilizada para ningún otro propósito y cualquier información que pueda identificarle permanecerá confidencial. Lee atentamente el contenido y decida si estás de acuerdo.

■ **Permiso para utilizar los materiales del alumno**

Se proporcionan materiales de escritura para crear y utilizar corpus de estudiantes del idioma coreano, y se permite la divulgación de todos los materiales de escritura originales y su uso con fines de investigación y desarrollo.

☞ ¿Estás de acuerdo con el permiso para utilizar los materiales lingüísticos de los estudiantes como se indica arriba?

☐ De acuerdo ☐ En desacuerdo

■ **Detalles de recopilación y uso de información personal**

Elementos	Propósito de recopilación y uso	Período de conservación/uso
Género, Año de Nacimiento, Nivel Actual de Coreano, Nacionalidad, Lengua Materna, Tiempo Estudiando el Idioma Coreano, Tiempo de Residencia en Corea, Propósito de Estudiar el Idioma Coreano, Ocupación, Cualquier Otro Idioma Extranjero Que Puedes Hablar	Construcción y utilización de corpus de estudiantes del coreanos por el Instituto Nacional del Idioma Coreano	<u>30 años desde la fecha del consentimiento para la recopilación y uso de información personal</u> * Sin embargo, si se solicita la eliminación de información personal, se eliminará inmediatamente.

※ Tiene derecho a rechazar el consentimiento para la recopilación y el uso de la información personal anterior. Sin embargo, si se niega a dar tu consentimiento, es posible que se le restrinja la participación en el proyecto de construcción/utilización del corpus de estudiantes coreanos.

☞ ¿Estás de acuerdo con la recopilación y el uso de información personal como se indica arriba?

☐ De acuerdo ☐ En desacuerdo

■ **Detalles del suministro de información personal a terceros**

Destinatario	Propósito de la provisión	Artículos proporcionados	Período de conservación/uso
<u>Contratista para utilizar el corpus de estudiantes de idioma coreano del Instituto Nacional del Idioma Coreano</u>	<u>Investigación y desarrollo del procesamiento de información en idioma coreano, Investigación y desarrollo del idioma coreano</u>	Género, Año de Nacimiento, Nivel Actual de Coreano, Nacionalidad, Lengua Materna, Tiempo Estudiando el Idioma Coreano, Tiempo de Residencia en Corea, Propósito de Estudiar el Idioma Coreano, Ocupación, Cualquier Otro Idioma Extranjero Que Puedes Hablar	<u>30 años desde la fecha del consentimiento para la recopilación y uso de información personal</u> * Sin embargo, si se solicita la eliminación de información personal, se eliminará inmediatamente.

※ Tiene derecho a rechazar el consentimiento para la recopilación y el uso de la información personal anterior. Sin embargo, si se niega a dar tu consentimiento, es posible que se le restrinja la participación en el proyecto de construcción/utilización del corpus de estudiantes coreanos.

☞ ¿Aceptas proporcionar información personal a terceros como se indica arriba?

☐ De acuerdo ☐ En desacuerdo

■ **Información personal Manejo de detalles del envío (Aviso)**

Destinatario (empresa consignadora)	Detalles del negocio
IIR TECH Inc. / Fundación Universidad-Industria de Yonsei	Construcción del corpus de estudiantes de coreano en 2024

Fecha: _____
Nombre: _____ (Firma)
Al Instituto Nacional del Idioma Coreano

El siguiente cuestionario incluye tu información personal que nos ayudará a hacer análisis de investigación científica. Estricta confidencialidad se mantendrá en todo momento. (Si es posible, llena el formulario en coreano o inglés.)

1. Género: ☐ M ☐ H

2. Año de Nacimiento: _____(p.ej. 1989)

3. Nivel actual de coreano: _____(TOPIK: _____)

4. Nacionalidad: _____

(* Yo soy un coreano que ha vivido en el extranjero ☐ Extranjero ☐)

5. Lengua Materna: _____

6. Tiempo Estudiando el Idioma Coreano(¿Durante cuánto tiempo has estado aprendiendo?) :

_____años _____meses (p.ej. 1 año 8 meses)

7. Tiempo de Residencia en Corea(¿Por cuánto tiempo has estado en Corea?) :

_____años _____meses (p.ej. 1 año 10 meses)

8. Propósito de estudiar el idioma coreano:

☐ para ingresar a una universidad en Corea ☐ para trabajar en Corea

☐ para vivir en Corea ☐ para un pasatiempo

☐ para el matrimonio

☐ otro (_____)

9. Ocupación: _____

10. Cualquier otro idioma extranjero que puedes hablar

(Nota: Si hay varios idiomas, por favor escríbelos en orden de fluidez) :

Письмо-разрешение на использование языковых материалов учащихся для создания и использования лингвистического корпуса изучающих корейский язык
Согласие на сбор и использование персональной информации и её предоставление третьим лицам

Государственный институт корейского языка в целях улучшения преподавания корейского языка продвигает проект, в рамках которого осуществляется сбор языковых данных (лингвистический корпус) изучающих корейский язык. Предоставляемые вами материалы используются для проведения исследований и разработок в различных областях, таких как совершенствование методов преподавания корейского языка, разработка учебников корейского языка, обучение корейскому языку и создание моделей искусственного интеллекта. Данный проект не несёт никакого финансового ущерба и физического вреда для принимающих в нём участие. Вы можете отказаться от участия в проекте при отсутствии соответствующего желания. Ваша персональная информация не будет использоваться для каких-либо других целей, и любая информация, которая может идентифицировать вас, останется конфиденциальной. Пожалуйста, внимательно прочтите указанное ниже содержание и примите решение об участии.

■ **Разрешение на использование материалов учащегося**

Предоставляю письменные материалы для создания и использования корпуса изучающих корейский язык. Даю разрешение на раскрытие всех оригинальных письменных материалов и их использование в целях исследований и разработок.

☞ **Вы согласны на использование языковых материалов учащихся, как указано выше?**

☐ Да ☐ Нет

■ **Описание собираемой и используемой персональной информации**

Вид	Цель сбора и использования	Период хранения / использования
Пол, дата рождения, текущий уровень TOPIK, гражданство, первый (родной) язык, период изучения корейского языка, период проживания в Республике Корея, цель изучения корейского языка, профессия, владение другими иностранными языками кроме корейского.	Создание и использование корпуса учащихся корейского языка в Государственном институте корейского языка	<u>30 лет со дня согласия на сбор и использование персональной информации</u> * Личная информация может быть немедленно удалена по запросу.

※ Вы имеете право отказаться в согласии на сбор и использование персональной информации, указанной выше. В этом случае вы можете быть ограничены в участии в проекте по созданию/использованию корейского корпуса учащихся.

☞ **Вы согласны на сбор и использование персональной информации, как указано выше?**

☐ Да ☐ Нет

■ **Описание личной информации, предоставляемой третьим лицам**

Получатель	Цель предоставления	Предоставляемая информация	Срок хранения и использования
<u>Сторона, предоставившая согласие на использование корпуса учащихся корейского языка Государственного института корейского языка</u>	<u>Исследования и разработка информации на корейском языке, исследования и разработка корейского языка</u>	Пол, дата рождения, текущий уровень TOPIK, гражданство, первый (родной) язык, период изучения корейского языка, период проживания в Республике Корея, цель изучения корейского языка, профессия, владение другими иностранными языками кроме корейского.	<u>30 лет со дня согласия на сбор и использование персональной информации</u> * Личная информация может быть немедленно удалена по запросу.

※ Вы имеете право отказаться в согласии на сбор и использование персональной информации, указанной выше. В этом случае вы можете быть ограничены в участии в проекте по созданию/использованию корейского корпуса учащихся.

☞ **Вы согласны на предоставление персональной информации третьим лицам, как указано выше?**

☐ Да ☐ Нет

■ **Подробности передачи обязанностей по обработке персональной информации (уведомление)**

Принимающая субподряд сторона (субподрядчик)	Описание задач
ИР ТЕCH Inc., Промышленный фонд Университета Ёнсе	Создание корпуса учащихся корейского языка в 2024 году

Дата _____

ФИО _____ (подпись)

Для предоставления в Государственный институт корейского языка

Ниже требуется заполнить личные данные, необходимые для исследования. Их сохранность и защита от огласки гарантируется (Желательно заполнять на корейском языке. При необходимости можно заполнить на английском).

1. Пол: ☐ Ж ☐ М

2. Год рождения: _____ г. (напр. 1989 г.)

3. Текущий уровень: _____ (ТОPIK: _____)

4. Гражданство: _____ (※ Зарубежный соотечественник: ☐ Да ☐ Нет)

5. Первый (родной) язык: _____

6. Срок изучения корейского языка: _____ г. _____ мес. (Напр.: 1 г. 3 мес.)

7. Срок проживания в Республике Корея: _____ г. _____ мес. (Напр.: 1 г. 3 мес.)

8. Цель изучения корейского языка:

- ☐ Поступление в ВУЗ ☐ Трудоустройство ☐ Проживание в Республике Корея
☐ Увлечение ☐ Брак ☐ Иное (_____)

9. Профессия: _____

10. Владение другими иностранными языками кроме корейского

(перечислить в порядке убывания уровня мастерства владения):

(러시아어)

**Солонгос хэл суралцагчийн корпус санг бүтээх, ашиглахад солонгос хэл суралцагчдын материалыг ашиглахыг зөвшөөрөх хуудас /
Хувийн мэдээлэл цуглуулах, ашиглах, гуравдагч этгээдэд дамжуулахыг зөвшөөрөх хуудас**

Үндэсний хэл шинжлэлийн төвөөс солонгос хэлний бодит сургалтыг сайжруулахын тулд солонгос хэл суралцагчдын хэлний материалыг цуглуулан ашиглах төсөл хэрэгжүүлж байна. Таны бидэнд хураалгасан материалыг солонгос хэл заах арга зүйг сайжруулах, солонгос хэлний сурах бичиг боловсруулах, солонгос хэлний боловсрол, хиймэл оюун ухааны загвар боловсруулах гэх мэт төрөл бүрийн чиглэлээр судалгаа, боловсруулалт хийхэд ашиглана. Тус судалгаанд оролцон материалаа хураалгаж буй танд санхүүгийн болон бие эрхтний хохирол ямар нэгэн байдлаар учрахгүй. Хэрэв уг төсөлд хамрагдахыг хүсээгүй тохиолдолд оролцохгүй байх бүрэн боломжтой. Мөн таны хураалгасан материалыг дээр дурдсанаас өөр зорилгоор ашиглагдахгүй, судалгаанд ашиглах үед хувь хүний мэдээллийг нууцалж ямар нэгэн байдлаар танигдуулахгүйгээр ашиглана. Агуулгыг анхааралтай уншиж, судалгаанд оролцохыг зөвшөөрөх эсэхээ шийднэ үү.

■ Суралцагчийн материалыг ашиглах зөвшөөрөл

Бичих материалыг солонгос хэл суралцагчийн корпус сан Үүсгэх, ашиглахад зориулалтаар хураалгасан бөгөөд бүх эх бичвэрийг ил болгох, судалгааны болон хөгжүүлэх зорилгоор ашиглахыг зөвшөөрч байна.

☞ **Та суралцагчдын материалыг дараах байдлаар ашиглахыг зөвшөөрч байна уу?**

☐ Зөвшөөрч байна ☐ Зөвшөөрөхгүй

■ Хувийн мэдээллийг цуглуулах, ашиглахтай холбоотой мэдээллийн дэлгэрэнгүй агуулга

Хамаарах мэдээлэл	Цуглуулах, ашиглах зорилго	Хадгалах/ашиглах хугацаа
Хүйс, төрсөн он, солонгос хэлний түвшин одоогийн байдлаар, иргэншил, Эх хэл, Солонгос хэл суралцсан хугацааг жил сараар бичнэ үү, Солонгост нийт амьдарсан хугацаа, Солонгос хэл сурч буй зорилго, Мэргэжил, Солонгос хэлнээс гадна эзэмшсэн хэлээ дэс дарааллын дагуу бичнэ үү.	Солонгос хэлний Үндэсний хүрээлэн Солонгос хэл суралцагчийн корпус сан Үүсгэх, ашиглах	Хувийн мэдээлэл цуглуулах, ашиглахыг зөвшөөрсөн өдрөөс хойш 30 жил * Хэрэв хувийн мэдээллийг устгах хүсэлт гаргавал тэр даруй устана.

※ Танд дээрх хувийн мэдээллийг цуглуулах, ашиглахыг зөвшөөрөхөөс татгалзах эрх бий. Хэрэв та мэдээллээ ашиглуулахаас татгалзвал та солонгос хэл суралцагчдын корпус санг Үүсгэх, ашиглах эрхээ хязгаарлуулж болзошгүйг анхаарна уу.

☞ **Та дээрх хувийн мэдээллийг цуглуулах, ашиглахыг зөвшөөрч байна уу?**

☐ Зөвшөөрч байна ☐ Зөвшөөрөхгүй

■ Гуравдагч этгээдэд дамжуулах хувийн дэлгэрэнгүй мэдээллийн дэлгэрэнгүй агуулга

Хүлээн авагч	Хангамжийн зорилго	Хамаарах мэдээлэл	Хадгалах/ашиглах хугацаа
Солонгос хэлний Үндэсний хүрээлэнгийн солонгос хэл суралцагчийн корпусыг ашиглах гэрээт гүйцэтгэгч	Солонгос хэлний мэдээлэл боловсруулах судалгаа, хөгжүүлэлт, Солонгос хэлний судалгаа, хөгжүүлэлт	Хүйс, төрсөн он, солонгос хэлний түвшин одоогийн байдлаар, иргэншил, Эх хэл, Солонгос хэл суралцсан хугацааг жил сараар бичнэ үү, Солонгост нийт амьдарсан хугацаа, Солонгос хэл сурч буй зорилго, Мэргэжил, Солонгос хэлнээс гадна эзэмшсэн хэлээ дэс дарааллын дагуу бичнэ үү.	Хувийн мэдээллийг цуглуулах, ашиглахыг зөвшөөрсөн өдрөөс хойш 30 жил * Хэрэв хувийн мэдээллийг устгах хүсэлт гаргавал тэр даруй устана.

※ Танд дээрх хувийн мэдээллийг цуглуулах, ашиглахыг зөвшөөрөхөөс татгалзах эрх бий. Хэрэв та мэдээллээ ашиглуулахаас татгалзвал таны солонгос хэл суралцагчдын корпус санг Үүсгэх, ашиглах эрхийг хязгаарлаж болзошгүйг анхаарна уу.

☞ **Та дээр дурдсаны дагуу гуравдагч этгээдэд хувийн мэдээллээ өгөхийг зөвшөөрөх үү?**

☐ Зөвшөөрч байна ☐ Зөвшөөрөхгүй

■ Хувийн мэдээллийг хариуцах ажлын дэлгэрэнгүй мэдээлэл (Мэдэгдэл)

Гүйцэтгэгч (гүйцэтгэгч компани)	Бизнесийн дэлгэрэнгүй мэдээлэл
IIR TECH Inc. / Йонсей их сургууль-аж Үйлдвэр академийн хамтын ажиллагааны зөвлөл	2024 он солонгос хэлний сургалтын корпус сан Үүсгэх

Он сар өдөр _____
Нэр _____/гарын Үсэг/
Солонгос хэлний Үндэсний хүрээлэнд

Дараах нь судалгаатай холбоотой мэдээллийн чанартай асуултууд юм. Таны хувийн мэдээлэл нууцлаг дах бөгөөд гадагш ямар нэгэн байдлаар ашиглагдахгүй. (Аль болох солонгосоор бөглөнө ҮҮ. Солонгосоор бөглөх боломжгүй бол англиар бөглөж болно.)

1. Хүйс : _____

2. Төрсөн он : _____

3. Таны солонгос хэлний түвшин одоогийн байдлаар : _____(TOPIK: _____)

4. Иргэншил : _____

5. Эх хэл : _____

6. Солонгос хэл суралцсан хугацааг _____ жил, сараар нь бичнэ ҮҮ :

_____ жил _____ сар (Жишээ нь: 1жил 3сар)

7. Солонгост нийт амьдарсан хугацаа :

_____ жил _____ сар (Жишээ нь: 1жил 3сар)

8. Солонгос хэл сурч буй зорилго :

☐ Их дээд сургуульд элсэн сурах

☐ Ажилд орох

☐ Оршин суух

☐ Сонирхол

☐ Гэрлэх

☐ Бусад

9. Мэргэжил : _____

10. Солонгос хэлнээс гадна эзэмшсэн хэлээ (сайн эзэмшсэн дарааллын дагуу) бичнэ ҮҮ :

(몽골어)

**Surat Pernyataan Persetujuan Penggunaan Materi Bahasa Pembelajar untuk Pembuatan dan Pemanfaatan Korpus Pembelajaran Bahasa Korea/
Persetujuan untuk Pengumpulan dan Penggunaan Informasi Pribadi dan Penyediaan kepada Pihak Ketiga.**

Untuk meningkatkan kualitas pendidikan bahasa Korea, Institut Nasional Bahasa Korea mengadakan proyek untuk mengumpulkan dan memanfaatkan materi bahasa (korpus) pembelajar bahasa Korea. Materi yang Anda berikan akan digunakan untuk penelitian dan pengembangan di berbagai bidang seperti penyempurnaan metode pengajaran bahasa Korea, pengembangan buku teks bahasa Korea, pendidikan bahasa Korea, dan pengembangan model kecerdasan buatan (Artificial Intelligence). Tidak ada kerugian finansial dan bahaya fisik yang berisiko terjadi pada partisipan dalam penelitian ini. Jika Anda tidak ingin berpartisipasi, Anda dapat mengundurkan diri dari penelitian ini. Selain itu, informasi pribadi yang dikumpulkan tidak akan digunakan untuk tujuan apa pun selain untuk proyek ini dan akan digunakan dalam bentuk yang tidak dapat diidentifikasi untuk menjaga kerahasiaan. Silakan membaca isi dari surat pernyataan ini dengan teliti dan memutuskan apakah Anda setuju atau tidak setuju dengan pernyataan di bawah ini.

■ **Persetujuan Penggunaan Materi Pembelajar**

Menyediakan materi tulis untuk pembuatan dan pemanfaatan korpus pembelajar bahasa Korea, dan setuju bahwa seluruh materi tulisan asli akan ditampilkan dan digunakan untuk tujuan penelitian dan pengembangan.

☞ **Apakah Anda setuju dengan penggunaan materi bahasa pembelajar sesuai pernyataan di atas?**

☐Setuju ☐Tidak Setuju

■ **Perihal pengumpulan dan penggunaan informasi pribadi**

Jenis informasi pribadi yang dikumpulkan dan digunakan	Tujuan pengumpulan dan penggunaan	Durasi penyimpanan/penggunaan
Jenis kelamin, tahun kelahiran, level TOPIK yang masih berlaku, kewarganegaraan, bahasa ibu, lama belajar bahasa Korea, lama tinggal di Korea, tujuan belajar bahasa Korea, pekerjaan, bahasa asing lain yang dikuasai selain bahasa Korea.	Pembuatan dan Pemanfaatan Korpus Pembelajar Bahasa Korea oleh Institut Nasional Bahasa Korea.	30 tahun sejak tanggal persetujuan pengumpulan dan penggunaan informasi pribadi * Informasi pribadi akan segera dihapus jika ada permintaan penghapusan.

※ Anda berhak untuk menolak memberikan persetujuan untuk pengumpulan dan penggunaan informasi pribadi di atas. Jika Anda menolak untuk memberikan persetujuan, Anda tidak diperbolehkan berpartisipasi dalam proyek pembuatan/pemanfaatan korpus pelajar bahasa Korea.

☞ **Apakah Anda setuju dengan pengumpulan dan penggunaan informasi pribadi sesuai pernyataan di atas?**

☐Setuju ☐Tidak Setuju

■ **Perihal penyediaan informasi pribadi kepada pihak ketiga**

Penerima	Tujuan penyediaan	Jenis informasi pribadi yang disediakan	Durasi penyimpanan/penggunaan
Kontraktor untuk penggunaan korpus pembelajar bahasa Korea dari Institut Nasional Bahasa Korea	Penelitian dan Pengembangan Pengolahan Informasi Korea, Penelitian dan Pengembangan Bahasa Korea	Jenis kelamin, tahun kelahiran, level TOPIK yang masih berlaku, kewarganegaraan, bahasa ibu, lama belajar bahasa Korea, lama tinggal di Korea, tujuan belajar bahasa Korea, pekerjaan, bahasa asing lain yang dikuasai selain bahasa Korea.	30 tahun sejak tanggal persetujuan pengumpulan dan penggunaan informasi pribadi * Informasi pribadi akan segera dihapus jika ada permintaan penghapusan.

※ Anda berhak menolak memberikan persetujuan untuk pengumpulan dan penggunaan informasi pribadi di atas. Namun, jika Anda menolak untuk memberikan persetujuan, Anda tidak diperbolehkan untuk berpartisipasi dalam proyek pembuatan/pemanfaatan korpus pelajar bahasa Korea.

☞ **Apakah Anda setuju untuk memberikan informasi pribadi kepada pihak ketiga sesuai pernyataan di atas?**

☐Setuju ☐Tidak Setuju

■ **Perihal pendelegasian penanganan informasi pribadi (pemberitahuan)**

Penerima delegasi (Perusahaan konsinyasi)	Detail
IIR TECH Inc. / Yayasan Industri-Universitas Yonsei	Pembuatan korpus pembelajar bahasa Korea tahun 2024

Tanggal:

Nama: (Tanda tangan)

kepada Institut Nasional Bahasa Korea

Informasi pribadi sebagaimana tercantum di bawah akan digunakan sebagai data untuk penelitian. Informasi pribadi yang Anda sediakan bersifat rahasia dan akan dijaga agar tidak tersebar ke luar. (Jika memungkinkan, harap menjawab dengan bahasa Korea atau bila perlu, Anda juga bisa menggunakan bahasa Inggris)

1. Jenis kelamin : ☐ Perempuan ☐ Laki-laki
2. Tahun kelahiran : Tahun _____ (Contoh: Tahun 1989)
3. Level sekarang : _____(TOPIK: _____)
4. Kewarganegaraan : _____
(※ Perihal negara asal ☐ orang Korea yang tinggal di luar negeri ☐ orang asing)
5. Bahasa ibu : _____
6. Lama belajar bahasa Korea (Berapa lama belajar bahasa Korea?): selama ____ tahun ____ bulan
(contoh: 1 tahun 3 bulan)
7. Lama tinggal di Korea (Berapa lama tinggal di Korea?): selama ____ tahun ____ bulan
(contoh: 1 tahun 3 bulan)
8. Tujuan belajar bahasa Korea:
☐ pendidikan (masuk perguruan tinggi) ☐ pekerjaan ☐ menetap di Korea
☐ hobi ☐ menikah ☐ lain (_____)
9. Pekerjaan: _____
10. Bahasa asing lain yang dikuasai, selain bahasa Korea. (Urutkan bahasa sesuai tingkat kelancaran):

(인도네시아어)

<기획·연구>

국립국어원 이현주 학예연구사

국립국어원 전경선 학예연구사

국립국어원 신종민 연구원

<연구 참여자>

연구 책임자 한송화(연세대학교)

공동 연구원 곽용진((주)이르테크)

강현화(연세대학교)

김선정(계명대학교)

김한샘(연세대학교)

홍혜란(배재대학교)

실무 연구원 김미선(연세대학교)

서지혜(연세대학교)

손연정(홍익대학교)

홍혜란(배재대학교)

연구 보조원 강수진(연세대학교) 외 32명

최용기 외((주)이르테크) 7명

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: 학위사

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘2025년 한국어 학습자 말뭉치 구축’ 사업의 결과물을 발간한 것입니다.