

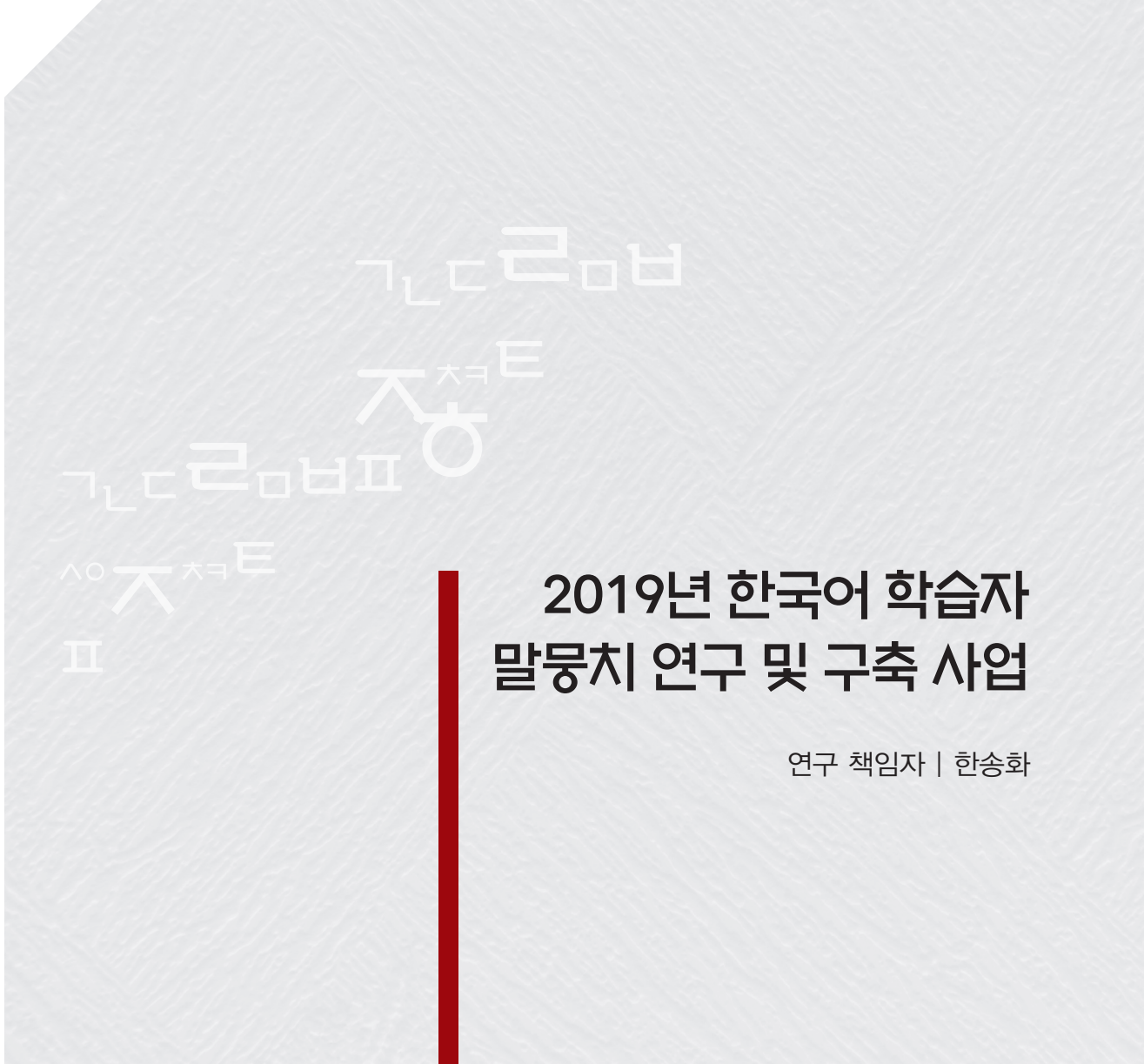
국립국어원 2019-01-38

2019년 한국어 학습자 말뭉치 연구 및 구축 사업

국립국어원



NATIONAL INSTITUTE OF KOREAN LANGUAGE



# 2019년 한국어 학습자 말뭉치 연구 및 구축 사업

연구 책임자 | 한송화

발 간 등 록 번 호
-------------

11-1371028-000746-10
----------------------

# 2019년 한국어 학습자 말뭉치 연구 및 구축 사업

연구 책임자: 한 송 화



국립국어원

# 제 출 문

국립국어원장 귀하

“2019년 한국어 학습자 말뭉치 연구 및 구축 사업”에 관하여  
귀 원과 체결한 연구용역 계약에 의하여 연구보고서를 작성하여  
제출합니다.

2019년 12월 31일

연구 책임자: 한송화(연세대학교)

연구 기관	연세대학교 산학협력단
연구 책임자	한송화(연세대)
공동 연구원	김선정(계명대), 김정숙(고려대), 김일환(성신여대), 김한샘(연세대), 안경화(서울대), 홍혜란(연세대), 김수현(국립국어원)
연구 보조원	공나형(연세대), 김동은(연세대), 김미선(연세대), 박은현(연세대), 김선영(연세대), 배미연(연세대), 송지혜(연세대), 유소영(연세대), 윤현애(연세대), 허희정(연세대)

## 2019년 한국어 학습자 말뭉치 연구 및 구축 사업

이 연구는 한국어 학습자 말뭉치 구축 중장기 사업 계획에 따른 제3단계, 5차 연도 연구이다. 5차 연도 연구는 효율적인 자료의 구축과 관리, 질적 제고를 위해 말뭉치 구축 지침을 정교화하고 그 지침에 따라 기구축된 2,602,780어절(문어 2,023,438어절, 구어 579,342어절) 규모의 말뭉치를 수정·보완함과 동시에 약 800,000어절(문어 550,000어절, 구어 250,000어절) 규모의 말뭉치를 새롭게 구축하는 데에 목적이 있다. 이에 따른 주요 과업과 연구의 성과는 다음과 같다.

**한국어 학습자 말뭉치 연구:** 한국어 학습자 말뭉치 연구는 말뭉치 구축 지침의 정교화와 말뭉치 활용을 위한 연구 체계의 수립으로 나뉜다. 5차 연도는 자료의 균형성 확보를 위한 국내 기관 자료와 이주민 자료의 계속 구축, 국외 학습자 자료의 집중 구축이 함께 이루어졌다. 이에 따라 자료의 효율적인 수집과 구축, 그리고 지침의 완결성을 갖출 수 있도록 체계성과 정합성의 측면에서 수집 지침, 자료 처리 지침, 문어 입력 지침, 구어 전사 지침, 형태 주석 지침과 오류 주석 지침을 보완하였다. 아울러 2015-2019년 구축 말뭉치를 활용한 분석 연구, 학습자 말뭉치 구축 이론 및 활용 연구의 집대성, 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구를 통해 말뭉치 활용을 위한 연구 체계 구축이 이루어졌다. 이는 구축된 말뭉치가 사용자들에게 효율적으로 사용되고 질적으로 완성도 높은 말뭉치를 구축해 나가기 위한 것이다.

**한국어 학습자 말뭉치 관련 교육 및 홍보:** 한국어 학습자 말뭉치 관련 교육은 효율적이고 체계적인 말뭉치 구축을 위하여 말뭉치 구축의 각 단계에 참여하는 인력들을 대상으로 지침 교육과 도구 사용 교육을 중심으로 이어져 오고 있다. 5차 연도는 이러한 교육 외에도 파이썬(Python)을 활용한 말뭉치 자료 처리 특강을 실시하여 말뭉치 구축 인력으로서의 전문성을 제고하였다. 한편, 3회의 학습자 말뭉치 아카데미를 통해 많은 사용자들이 학습자 말뭉치를 폭넓게 활용할 수 있도록 널리 알리고 말뭉치를 보다 손쉽게 활용할 수

있도록 하였다. 또한 학술대회 발표를 통해 학습자 말뭉치 연구와 구축에 대한 학계의 관심을 고조시켰다.

**한국어 학습자 말뭉치 수집 및 가공:** 5차 연도의 한국어 학습자 말뭉치 구축은 1-4차 연도 사업에서의 집중 구축 대상인 국내 교육 기관과 이주민 자료의 균형성을 맞추기 위한 계속 구축과 함께 국외 자료의 본격 구축이 진행되었다. 한국어 학습자 말뭉치 구축 및 가공은 1-4차 연도에 구축한 말뭉치의 수정·보완과 함께 원시 말뭉치 1,048,052어절(문어 795,925어절, 구어 252,127어절), 형태 주석 말뭉치 738,220어절(문어 527,744어절, 구어 210,476어절), 오류 주석 226,136어절(문어 113,789어절, 구어 112,347어절) 규모의 말뭉치가 새롭게 구축되었다. 그 결과 1-5차 연도(2015-2019년)에 구축한 전체 말뭉치의 규모는 원시 말뭉치 3,650,832어절(문어 2,819,363어절, 구어 831,469어절), 형태 주석 말뭉치 2,629,251어절(문어 2,037,748어절, 구어 591,503어절), 오류 주석 말뭉치 799,157어절(문어 462,322어절, 구어 336,835어절)이 되었다.

**말뭉치 구축 지원 도구 관리 및 구축 말뭉치 검증 정교화:** 한국어 학습자 말뭉치는 표본 등록에서 말뭉치 주석 가공까지 전체 작업 공정을 관리하고 수행할 수 있는 말뭉치 구축 지원 도구를 활용하여 구축되고 있다. 5차 연도에는 구축 실무 연구원들을 중심으로 성능 피드백팀을 구성하고 구축 지원 도구에 대한 정기적인 피드백을 실시하였다. 또한 격월로 국립국어원, 구축 지원 도구 개발팀, 구축팀 실무회의를 통해 구축 지원 도구의 사용과 말뭉치 구축에 관한 다양한 쟁점에 대해 논의하고 해결 방안을 모색하였다. 그럼으로써 보다 효율적으로 구축 지원 도구를 관리하고 말뭉치를 구축해 나갈 수 있었다. 그 외에도 구축된 말뭉치 자료의 질적 제고를 위해 문어 입력과 구어 전사, 형태 주석, 오류 주석의 각 팀에서 3단계의 검수, 시스템 기반의 데이터 검증과 검수 작업을 지속적으로 실시하였다.

<한국어 학습자 말뭉치>는 한국어 교육 연구, 교수, 학습에 광범위하게 활용됨으로써 한국어의 세계화 및 국제 경쟁력 강화에 이바지할 것이다.

주요어: 한국어 학습자 말뭉치, 국내 교육 기관의 학습자 자료,  
이주민 자료, 국외 교육 기관의 학습자, 문어 말뭉치, 구어 말뭉치,  
원시 말뭉치, 형태 주석 말뭉치, 오류 주석 말뭉치

# 차 례

<b>I. 연구 개요</b>	<b>1</b>
1. 연구의 목적 및 필요성	1
1.1. 연구의 목적	1
1.2. 연구의 필요성	1
2. 연구의 범위	2
3. 연구 방법	2
4. 연구 수행 기간 및 추진 경과	3
5. 연구 결과	4
<b>II. 학습자 말뭉치 연구</b>	<b>7</b>
1. 말뭉치 구축 지침의 정교화	7
1.1. 수집 및 자료 처리 지침	10
1.2. 자료 구축 지침	12
1.3. 자료 가공 지침	15
2. 말뭉치 활용 연구 체계 구축	22
2.1. 2015-2019년 구축 말뭉치를 활용한 분석 연구	22
2.2. 학습자 말뭉치 활용 연구	42
2.3. 한국어 학습자 말뭉치의 균형성 확보를 위한 구축 방법 연구	44
<b>III. 학습자 말뭉치 교육 및 홍보</b>	<b>51</b>
1. 말뭉치 구축/가공 인력 실무 교육	51
1.1. 교육 대상	51
1.2. 교육 방법	51
1.3. 교육 내용	52
1.4. 참여 인력	53
2. 한국어 학습자 말뭉치 아카데미 개최	55
3. 학술대회 발표	56

## IV. 학습자 말뭉치 구축 및 가공 ..... 57

1. 수집 .....	57
1.1. 구축 목표 및 수집 대상 .....	57
1.2. 수집 네트워크 .....	57
1.3. 자료 수집 현황 .....	60
2. 구축 및 가공 .....	64
2.1. 원시 말뭉치 .....	64
2.2. 형태 주석 말뭉치 .....	69
2.3. 오류 주석 말뭉치 .....	72
2.4. 1-4차 연도 구축 말뭉치 자료 정제 .....	75
3. 구축 지원 도구 검증 .....	77
3.1. 구축 지원 도구 성능 피드백 실시 .....	77
3.2. 기능 추가 및 개선 시 도구 사용 연계 .....	78

## VI. 구축 말뭉치 검증 정교화 ..... 79

1. 문어 입력 및 구어 전사 작업 검수 시스템 정교화 .....	79
2. 형태 주석 및 오류 주석 작업 검수 시스템 정교화 .....	80
3. 데이터베이스 검증을 통한 검수 시스템 정교화 .....	81
3.1. 주석별 불일치 항목의 검수 .....	81
3.2. 작업 중 생성된 오조작 데이터 검증 .....	81

## VII. 결론 ..... 83

1. 연구 요약 .....	83
2. 연구의 의의 및 기대 효과 .....	86
3. 보고서 활용 방안 .....	89
4. 정책 제안 .....	90

## 참고문헌

## 부록. 2019년 한국어 학습자 말뭉치 구축 지침

## 표 차례

<표 1> 연구의 범위와 세부 과업 내용 .....	2
<표 2> 연구 수행 방법 .....	3
<표 3> 연구 추진 경과 .....	3
<표 4> 2015-2019년 학습자 말뭉치의 구축 규모 .....	5
<표 5> 1-4차 연도 말뭉치 구축 지침 정교화의 주요 쟁점과 내용 .....	7
<표 6> 구어 자료의 다양성과 균형성 확보를 위한 수집 과제 개발 내용 .....	12
<표 7> 구축 지침의 정교화 내용 .....	13
<표 8> 가공 지침의 정교화 내용 .....	16
<표 9> 문어 형태 주석 결과_숙달도 단계별 .....	23
<표 10> 문어 형태 주석 결과_언어권별 .....	24
<표 11> 구어 형태 주석 결과_숙달도 단계별 .....	26
<표 12> 구어 형태 주석 결과_언어권별 .....	27
<표 13> 문어 오류 주석 결과_분석 불능 여부: 숙달도 단계별 .....	29
<표 14> 문어 오류 주석 결과_분석 불능 여부: 언어권별 .....	29
<표 15> 문어 오류 주석 결과_오류 위치: 숙달도 단계별 .....	29
<표 16> 문어 오류 주석 결과_오류 위치: 언어권별 .....	31
<표 17> 문어 오류 주석 결과_오류 층위: 숙달도 단계별 .....	32
<표 18> 문어 오류 주석 결과_오류 층위: 언어권별 .....	33
<표 19> 문어 오류 주석 결과_오류 양상: 숙달도 단계별 .....	34
<표 20> 문어 오류 주석 결과_오류 양상: 언어권별 .....	35
<표 21> 구어 오류 주석 결과_분석 불능 여부: 숙달도 단계별 .....	35
<표 22> 구어 오류 주석 결과_분석 불능 여부: 언어권별 .....	36
<표 23> 구어 오류 주석 결과_오류 위치: 숙달도 단계별 .....	36
<표 24> 구어 오류 주석 결과_오류 위치: 언어권별 .....	37
<표 25> 구어 오류 주석 결과_오류 층위: 숙달도 단계별 .....	39
<표 26> 구어 오류 주석 결과_오류 층위: 언어권별 .....	40
<표 27> 구어 오류 주석 결과_오류 양상: 숙달도 단계별 .....	41
<표 28> 구어 오류 주석 결과_오류 양상: 언어권별 .....	41
<표 29> 말뭉치 구축/가공 인력 교육 내용 .....	52
<표 30> 학습자 말뭉치 활용 아카데미 개최 .....	55
<표 31> 수집 네트워크 현황 .....	57



<표 32> 수집 대상에 따른 과제 유형 .....	59
<표 33> 국외 자료 수집 현황 .....	60
<표 34> 종적 자료 수집 현황: 주차별 누적 자료 수 .....	62
<표 35> 종적 자료 수집 현황: 학습자 유형 및 국적별 분포 .....	62
<표 36> 문어 원시 말뭉치의 숙달도별 자료 분포 .....	64
<표 37> 문어 원시 말뭉치의 국적별 자료 분포 .....	65
<표 38> 구어 원시 말뭉치의 숙달도별 자료 분포 .....	67
<표 39> 구어 원시 말뭉치의 국적별 자료 분포 .....	67
<표 40> 문어 형태 주석 말뭉치의 숙달도별 자료 분포 .....	69
<표 41> 문어 형태 주석 말뭉치의 국적별 자료 분포 .....	69
<표 42> 구어 형태 주석 말뭉치의 숙달도별 자료 분포 .....	71
<표 43> 구어 형태 주석 말뭉치의 국적별 자료 분포 .....	72
<표 44> 문어 오류 주석 말뭉치의 숙달도별 자료 분포 .....	73
<표 45> 문어 오류 주석 말뭉치의 국적별 자료 분포 .....	73
<표 46> 구어 오류 주석 말뭉치의 숙달도별 자료 분포 .....	74
<표 47> 구어 오류 주석 말뭉치의 국적별 자료 분포 .....	75
<표 48> 성능 평가 회의 개최 일시와 주요 안건 .....	77

## 그림 차례

<그림 1> 작업 공정에서의 3단계 검수 체계 .....	79
<그림 2> 가공 말뭉치 검수 체계 .....	80

# I. 연구 개요

## 1. 연구의 목적 및 필요성

### 1.1. 연구의 목적

○ 본 연구의 목적은 3단계 6개년 간의 중장기 사업 계획에 근거하여 한국어 학습자의 수준별·언어권별 문어·구어 말뭉치 자료를 구축함으로써 과학적인 한국어 교육의 기반을 조성하는 데에 있다. 즉, 한국어 학습자 말뭉치를 한국어 교육을 위한 기초 연구, 교재 개발, 교수·학습에 활용하여 더욱 체계적인 교육 환경을 제공하고자 하는 것이다. <2019년 한국어 학습자 말뭉치 연구 및 구축 사업>은 3단계 5차 연도의 사업으로 다음과 같은 세부 목표에 따라 수행되었다.

- 말뭉치 구축 및 활용 연구
- 학습자 말뭉치 교육 및 홍보
- 말뭉치 수집 및 구축 가공
- 구축 지원 도구 검증
- 구축 말뭉치 검증 정교화

### 1.2. 연구의 필요성

- 국가 주도의 한국어 학습자 균형 말뭉치 구축
- 국어 빅데이터의 한 축인 한국어 학습자 말뭉치 구축
- 한국어 학습자의 다양화에 따른 교수·학습 환경의 과학화
- 사용자 편의성을 고려한 학습자 말뭉치 구축 및 배포
- 학습자 말뭉치 활용 연구를 통한 국제 수준의 학술 교류 활성화

## 2. 연구의 범위

○ 본 연구의 범위와 세부 내용은 다음과 같다.

<표 1> 연구의 범위와 세부 과업 내용

연구의 범위	세부 내용
학습자 말뭉치 연구	<ul style="list-style-type: none"> <li>○ 한국어 학습자 말뭉치 균형성 확보를 위한 구축 방법 연구</li> <li>○ 말뭉치 구축 지침 정교화</li> <li>○ 말뭉치 활용 연구</li> </ul>
학습자 말뭉치 교육 및 홍보	<ul style="list-style-type: none"> <li>○ 말뭉치 구축/가공 인력 실무 교육</li> <li>○ 한국어 학습자 말뭉치 아카데미 개최(2회)</li> <li>○ 학회 발표(1회)</li> </ul>
말뭉치 구축 가공	<ul style="list-style-type: none"> <li>○ 학습자 자료 수집 및 말뭉치 구축</li> <li>○ 구축된 말뭉치 가공(형태분석, 오류분석)</li> <li>○ 기구축(2015~2018년) 학습자 말뭉치 수정 및 정교화</li> </ul>
구축 지원 도구 검증	<ul style="list-style-type: none"> <li>○ 구축 지원 도구 성능 피드백 실시</li> <li>○ 기능 추가 및 개선 시 도구 사용 교육 연계</li> </ul>
구축 말뭉치 검증 정교화	<ul style="list-style-type: none"> <li>○ 구어·문어 입력 검수, 형태·오류 검수 시스템 정교화</li> <li>○ 검수 결과 및 현황 2회 보고 (중간보고회 1회, 최종보고회 전 1회)</li> </ul>

## 3. 연구 방법

○ 본 연구는 말뭉치 구축을 위한 지침 정교화, 말뭉치 활용에 관한 기반 연구와 실제 말뭉치 구축 연구가 순환적으로 이루어진다는 특성이 있다. 연구 방법은 문헌 연구, 현황 조사, 전문가 의견 수렴의 세 가지 방법에 따르며, 각 연구 방법에 따른 연구 내용과 목적은 다음과 같다.

<표 2> 연구 수행 방법

	내용	목적 및 의의
문헌 연구	○한국어 학습자 말뭉치의 활용에 관한 연구 사례 조사	○말뭉치 활용 연구 체계 구축의 기초 자료
현황 조사	○수집 대상 기관의 교육과정 및 학습자 분포 등 말뭉치 구축을 위한 기초 조사	○현실적인 자료 수집 가능성 검토 및 수집 계획 수립
의견 수렴	○한국어 학습자 말뭉치 수집 인력풀을 통한 자문 및 의견 수렴 ○사업 착수 보고회의, 중간보고회의, 최종보고회의를 통한 정기 자문 및 평가	○수집 교사의 참여 유도 ○연구 방향 및 방법, 절차의 타당성 검증

## 4. 연구 수행 기간 및 추진 경과

○ 연구 수행 기간 및 추진 경과는 다음과 같다.

- 연구 수행 기간: 2019년 3월 19일-2019년 12월 31일
- 연구 추진 경과

<표 3> 연구 추진 경과

단계	작업 내용	1 개 월	2 개 월	3 개 월	4 개 월	5 개 월	6 개 월	7 개 월	8 개 월	9 개 월	10 개 월
준비	2018년 구축 자료 정리 및 연구 계획 수립	○									
연구	구축 지침의 정교화	○	○	○	○	○	○	○	○	○	○
	말뭉치 균형성 확보 및 활용 연구	○	○	○	○	○	○	○	○	○	○
수집	자료 수집 지침 교육	○	○	○	○	○	○	○	○	○	○

	자료 수집 및 처리	○	○	○	○	○	○	○	○	○	○
구축	말뭉치 자료 처리, 입력/전사 지침 및 도구 교육	○	○	○	○	○	○	○	○	○	○
	스캔 및 파일 분류, 등록		○	○	○	○	○	○	○	○	○
	입력 및 전사		○	○	○	○	○	○	○	○	○
가공	말뭉치 가공 지침 및 도구 교육	○	○	○	○	○	○	○	○	○	○
	형태 주석		○	○	○	○	○	○	○	○	○
	오류 주석		○	○	○	○	○	○	○	○	○
	기구축 말뭉치 자료 정제	○	○	○	○	○	○	○	○	○	○
홍보	한국어 학습자 말뭉치 아카데미 개최(2회)					○			○		
	학회 발표					○					
관리	말뭉치 구축 지원 도구 성능 피드백	○	○	○	○	○	○	○	○	○	○
	구축 말뭉치 검증 정교화	○	○	○	○	○	○	○	○	○	○
마무리	2019년 연구 마무리 및 2020년 연구 계획 수립										○

## 5. 연구 결과

○ 본 사업의 연구 결과는 다음과 같다.

- 학습자 말뭉치 구축 연구 체계 구축
- 한국어 학습자 말뭉치 아카데미 3회 개최 및 학회 발표
- 1-4차 연도 구축 말뭉치의 정제
- 2019년 신규 말뭉치 자료 수집 및 구축, 가공
- 기구축 말뭉치의 정합성 검증

<표 4> 2015-2019년 학습자 말뭉치의 구축 규모

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
원시 말뭉치									
2015 - 2018	문어	278,595 (4,013)	316,860 (3,165)	337,531 (2,676)	338,801 (2,429)	333,584 (2,011)	278,457 (1,553)	139,610 (136)	2,023,438 (15,983)
	구어	84,831 (250)	98,155 (236)	114,948 (254)	109,762 (245)	83,834 (152)	66,700 (99)	21,112 (15)	579,342 (1,251)
2019 신규	문어	76,368 (1,331)	147,461 (1,787)	212,429 (2,133)	167,893 (1,674)	120,022 (951)	59,902 (481)	11,850 (17)	795,925 (8,374)
	구어	47,106 (162)	68,484 (130)	48,219 (103)	56,062 (102)	15,311 (28)	16,698 (33)	247 (1)	252,127 (559)
합계	문어	354,963 (5,344)	464,321 (4,952)	549,960 (4,809)	506,694 (4,103)	453,606 (2,962)	338,359 (2,034)	151,460 (153)	2,819,363 (24,357)
	구어	131,937 (412)	166,639 (366)	163,167 (357)	165,824 (347)	99,145 (180)	83,398 (132)	21,359 (16)	831,469 (1,810)
형태 주석 말뭉치									
2015 - 2018	문어	227,290 (3,292)	252,564 (2,469)	273,140 (2,195)	267,267 (2,037)	235,434 (1,558)	215,260 (1,325)	39,049 (28)	1,510,004 (12,904)
	구어	50,880 (165)	51,279 (124)	73,160 (171)	68,490 (165)	68,525 (117)	47,581 (63)	21,112 (15)	381,027 (820)
2019 신규	문어	67,161 (966)	89,682 (927)	99,513 (828)	88,729 (695)	89,190 (573)	59,749 (327)	33,720 (46)	527,744 (4362)
	구어	31,251 (78)	44,629 (83)	34,354 (78)	39,486 (73)	27,580 (58)	33,176 (65)	-	210,476 (435)
합계	문어	294,451 (4,258)	342,246 (3,396)	372,653 (3,023)	355,996 (2,732)	324,624 (2,131)	275,009 (1,652)	72,769 (74)	2,037,748 (17,266)
	구어	82,131 (243)	95,908 (207)	107,514 (249)	107,976 (238)	96,105 (175)	80,757 (128)	21,112 (15)	591,503 (1,255)
오류 주석 말뭉치									
2015 - 2018	문어	57,852 (826)	57,251 (552)	62,856 (508)	61,169 (492)	5,9087 (408)	50,318 (307)	-	348,533 (3,093)
	구어	23,683 (85)	28,869 (78)	36,340 (107)	42,468 (119)	51,741 (83)	38,341 (45)	3,046 (3)	224,488 (520)

2019 신규	문어	17,012 (223)	17,862 (195)	19,585 (179)	16,997 (148)	17,551 (137)	24,782 (174)	-	113,789 (1,056)
	구어	21,492 (55)	25,306 (45)	25,695 (58)	19,841 (50)	8,371 (18)	9,796 (19)	1,846 (2)	112,347 (247)
합계	문어	74,864 (1,049)	75,113 (747)	82,441 (687)	78,166 (640)	76,638 (545)	75,100 (481)	-	462,322 (4,149)
	구어	45,175 (140)	54,175 (123)	62,035 (165)	62,309 (169)	60,112 (101)	48,137 (64)	4,892 (5)	336,835 (767)

- 단위: 어절 수, ( ) 안은 파일 수



## II. 학습자 말뭉치 연구

### 1. 말뭉치 구축 지침의 정교화

- 한국어 학습자 말뭉치 구축 지침은 2015년 1단계 1차 연도에 기본 지침을 수립한 이후 5차 연도 사업을 수행하고 있는 현재에 이르기까지 수집 대상(국내 기관, 이주민, 국외 학습자), 자료 유형(문어, 구어), 구축 단계(수집 및 자료 처리, 문어 입력, 구어 전사, 형태 주석, 오류 주석)의 특성을 반영하고 작업의 일관성과 체계성을 제고하는 방향으로 지속적인 수정과 보완을 해 왔다. 다음은 1-4차 연도에 지침을 정교화해 나가는 과정에서 제기되었던 주요한 쟁점과 지침을 수정·보완한 내용이다.

<표 5> 1-4차 연도 말뭉치 구축 지침 정교화의 주요 쟁점과 내용

시기	지침 정교화의 쟁점과 내용
1차 연도	○ 한국어 학습자 말뭉치 구축 지침의 기준(standard)으로서의 수집 및 자료 처리, 문어 입력, 구어 전사, 형태 주석, 오류 주석 지침 마련
2차 연도	<ul style="list-style-type: none"> <li>○ 문어 입력 지침: 1차 연도에 텍스트 에디터를 기반으로 한 입력 작업을 시스템 기반으로 전환하면서 발생하는 문제 해결               <ul style="list-style-type: none"> <li>- 입력과 식별이 어려운 문자의 표기 방법</li> <li>- 일반 문자 외의 기호 처리 방식</li> </ul> </li> <li>○ 구어 전사 지침: 학습자의 발음 오류 및 한국어의 음운 체계와 맞지 않는 발음의 표기 문제 해결               <ul style="list-style-type: none"> <li>- 음운규칙을 적용하지 못하여 발생한 오류</li> <li>- 불분명한 발음, 중간 소리의 표기</li> <li>- 음절 발음에서의 오류</li> <li>- 외래어의 규범 발음과 현실 발음이 다른 경우</li> </ul> </li> <li>○ 형태 주석 지침: 21세기 세종 계획 한국어 균형 말뭉치 현대 문어 형태 주석 말뭉치 분석 지침과의 호환성과 분석의 일관성 제고               <ul style="list-style-type: none"> <li>- 21세기 세종계획과의 호환성                   <ul style="list-style-type: none"> <li>▶ 명사 + ‘하다’ 류의 분리 문제</li> </ul> </li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>▶ ‘척하다’ 류의 분석</li> <li>▶ 지정사의 처리</li> <li>▶ 부사격조사와 접속조사와 구분</li> <li>▶ 접두사 목록의 확장</li> <li>▶ 접미사 목록의 확장</li> <li>▶ XR(어근) 표지 복원</li> <li>- 분석의 일관성 <ul style="list-style-type: none"> <li>▶ 유사 접사류 목록의 제시</li> <li>▶ 학습자 오류 어절의 처리</li> </ul> </li> </ul> <p>○ 오류 주석 지침: 오류 주석 작업의 효율성 제고, 형태 주석 작업과 연계한 오류 주석 틀과 표지의 정비, 오류 판정과 주석 지침의 상세화</p> <ul style="list-style-type: none"> <li>- 오류 주석 체계와 틀 수정 <ul style="list-style-type: none"> <li>▶ 오류 주석의 세부적인 체계 수정</li> <li>▶ 오류 주석 범주인 [오류 영역]을 [오류 위치]로 용어 변경</li> <li>▶ 오류 주석 범주인 [오류 양상]의 세부 오류 유형 변경</li> <li>▶ 오류 주석 표지 추가</li> </ul> </li> <li>- 지침의 상세화 <ul style="list-style-type: none"> <li>▶ 오류의 식별 기준 보완</li> <li>▶ 오류의 교정 기준 보완</li> <li>▶ 시제 오류 처리 기준 보완</li> <li>▶ 사동과 피동 오류 처리 기준 보완</li> <li>▶ 높임 오류 처리 기준 보완</li> <li>▶ 구어 오류 처리 기준 보완</li> </ul> </li> </ul>
3차 연도	<p>○ 수집 및 자료 처리 지침: 3차 연도 사업의 집중 수집 대상인 이주민의 특성을 반영한 수집 과제와 학습자 동의서 수정·보완</p> <ul style="list-style-type: none"> <li>- 이주민 학습자의 교육과정을 반영한 수집 과제 설계</li> <li>- 이주민 학습자의 특성을 반영한 학습자 정보 수집 항목 추가</li> <li>- 이주민 학습자의 국적 분포를 고려한 동의서 번역 언어 추가</li> </ul> <p>○ 문어 입력 지침: 최고급 학습자 자료 추가에 따른 본문 주석 추가 개발 및 관련 지침 보완, 작업자들의 이해를 돕기 위한 지침의 구체화</p> <ul style="list-style-type: none"> <li>- 최고급 자료 처리를 위한 본문 주석 체계 개발</li> <li>- 개인 정보 처리에 관한 지침의 구체화</li> </ul>

	<ul style="list-style-type: none"> <li>○ 구어 전사 지침: 1차 전사한 텍스트를 말뭉치 구축 지원 도구에 업로드하는 과정에서 발생하는 문제 해결과 형태 주석, 오류 주석과의 연계성을 고려한 지침의 정교화 <ul style="list-style-type: none"> <li>- 발화자 표시 위치 수정</li> <li>- 전사 기호의 중복 문제 해소</li> <li>- 학습자의 발음 오류 표기에 관한 지침 상세화</li> </ul> </li> <li>○ 형태 주석 지침: 비정형의 학습자 언어 처리의 효율성과 일관성 제고를 위한 분석 기준, 학습자 오류 어절 처리에 관한 기술과 예시의 상세화 <ul style="list-style-type: none"> <li>- 분석 불가능(NA) 항목 처리 지침의 구체화</li> <li>- 학습자 오류 어절 처리 지침의 구체화</li> <li>- 학습자 오류 어절의 분할 기준 보완</li> <li>- 기타 세부 지침의 보완</li> </ul> </li> <li>○ 오류 주석 지침: 오류 주석 작업자 간의 일관성과 작업의 효율성 제고, 구어 전사 지침과 연동한 사례별 오류 식별과 판정 기준에 대한 지침 상세화 <ul style="list-style-type: none"> <li>- 오류 층위에서 형태와 관련된 단어 형성[합성법], [파생법], 굴절[곡용], [활용] 및 품사 오류 지침 보완</li> <li>- 오류 층위에서 통사와 관련된 높임, 시제, 사동과 피동, 부정, 어순 오류 지침 보완</li> <li>- 오류 층위에서 담화와 관련된 지시, 접속, 담화표지, 구어/문어 오류 지침 보완</li> <li>- 음소(PP) 오류 지침 보완</li> <li>- 음절(PS) 오류 지침 보완</li> <li>- 음운규칙(PC) 오류 지침 보완</li> <li>- 원어식 발음(PN) 오류 지침 보완</li> <li>- 중간발음(변이음 포함)(PA) 오류 지침 보완</li> </ul> </li> </ul>
4차 연도	<ul style="list-style-type: none"> <li>○ 수집 및 자료 처리 지침: 국내 기관, 이주민, 국외 학습자 자료의 효율적 수집과 통합 <ul style="list-style-type: none"> <li>- 말뭉치의 균형성 확보 및 자료 수집의 효율성 제고를 위한 기획 수집 과제 설계</li> <li>- 말뭉치 구축 대상 자료 선별 지침 보완</li> <li>- 파일 분류 및 파일명 부여 체계 보완</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>○ 문어 입력 지침: 입력 과정에서 부정확하게 생성된 오형태 어절의 분리로 인한 문제, 작업자 간의 일관성 제고를 위한 지침 보완 <ul style="list-style-type: none"> <li>- 학습자의 언어 특성에 따른 오형태 문자 처리 지침 보완</li> </ul> </li> <li>○ 구어 전사 지침: 말뭉치 구축 지원 도구 내에 음성 동기화 및 전사 기능 사용과 관련한 작업 지침의 보완 <ul style="list-style-type: none"> <li>- 구축 도구 기반의 직접 주석 작업을 위한 지침 보완</li> </ul> </li> <li>○ 형태 주석 지침: 오류 형태 처리에 관한 지침 상세화 및 구어형 분석에 대한 세부 지침 보완 <ul style="list-style-type: none"> <li>- 잉여적 요소가 추가된 어절의 형태 분석 지침 수정</li> <li>- 불규칙 용언의 활용 오류의 형태 분석 지침 보완</li> <li>- 구어에서 나타나는 구어형의 형태 분석 지침 보완</li> <li>- 기타 세부 지침 보완</li> </ul> </li> <li>○ 오류 주석 지침: 주석자 간의 일관성, 작업의 효율성 제고를 위한 처리 지침 상세화 <ul style="list-style-type: none"> <li>- 문어 오형태 활용 오류 처리 기준 보완</li> <li>- 음소(PP) 오류와 오형태(MIF) 오류의 구분</li> <li>- 불완전한 발화 처리 문제</li> <li>- 음소(PP) 오류와 음운규칙(PC) 오류의 구분</li> </ul> </li> </ul>
--	---

- 2019년 5차 연도에는 각 단계별 구축 작업을 진행하는 과정에서 작업자 간의 일관성을 제고하기 위하여 지침의 세부 항목별 사례를 추가로 제시하거나 기술을 상세화하는 방향으로 지침을 보완하였다. 아울러 2015년부터 현재까지 지침을 기술해 오는 과정에서 기술 내용이 서로 충돌하거나 입력 및 전사, 형태 주석, 오류 주석의 구축 단계별로 연관성이 있는 지침 내용이 상충되는 부분을 점검하여 지침의 정교함을 높이고자 하였다.

## 1.1. 수집 및 자료 처리 지침

### (1) 지침 수정 및 보완의 쟁점

- 본 사업에서는 국외 학습자 자료를 집중적으로 구축하면서 동시에 균형성 확보를 위한 국내 교육 기관 학습자와 이주민 학습자 자료를 기획하여 구

축하는 것을 목표로 하였다. 이에 따라서 국외의 수집 환경에 맞춰 수집 및 자료 처리 지침을 상세화하였다. 아울러 국내 자료 중 구어 자료의 다양성과 균형성 확보를 위하여 기획 과제를 추가로 개발하였다.

- 국외 자료 수집 및 처리 지침의 상세화 및 분리
- 구어 자료의 다양성과 균형성 확보를 위한 기획 과제 추가 개발

## (2) 수정 및 보완 내용

### ① 국외 자료 수집 및 처리 지침의 상세화 및 분리

- 국외 자료 수집과 처리에 있어서 효율성을 제고하는 방향으로 지침을 차별화하였다. 즉, 국내와 달리 물리적인 거리가 있고 인터넷 사용 환경이 불안정하며, 수집 자료의 처리를 위한 복사기나 스캐너 등의 기기 사용이 쉽지 않은 상황을 고려하여 국외에서 자료를 수집하여 전송하기까지의 처리 과정을 차별화한 별도의 지침을 구성하였다. 또한 현재 14개 언어로 번역되어 있는 학습자 동의서에 스페인어를 추가하였다. 그 결과 학습자 동의서를 영어, 중국어, 일본어, 러시아어, 몽골어, 베트남어, 필리핀어, 인도네시아어, 몽골어, 태국어, 캄보디아어, 네팔어, 파키스탄어, 방글라데시어, 스페인어로 총 15개의 번역어로 제공할 수 있게 되었다.

### ② 구어 자료의 다양성과 균형성 확보를 위한 기획 과제 추가 개발

- 현재까지 구축된 구어 자료는 교육기관의 성취도 평가의 일환으로 시행된 말하기 자료가 주를 이룬다. 이 자료는 한국어 교육과정에서 중점적으로 다루는 주제와 과제를 기반으로 했기 때문에 구축 자료의 간결성 확보에 많은 장점이 있다. 반면에 시험이라는 제한된 상황 맥락과 환경에서 수집된 다소 인위적이고 통제된 발화 자료이기 때문에 학습자의 자연 발화에서 확인 가능한 다양한 언어적 오류를 다양하게 확보할 수 없다는 점에서 한계가 있다. 따라서 본 연구에서는 구어 자료의 다양성과 과제의 균형성을 확보하기 위하여 학습자의 자연스러운 언어 수행을 관찰할 수 있는 과제를 <표6>과 같이 개발하였고, 이를 기획 자료 수집에 활용하였다. (☞ 상세한 과제는 지침 참고).

<표 6> 구어 자료의 다양성과 균형성 확보를 위한 수집 과제 개발 내용

	초급	중급	고급
준비하기(공통)	자기소개		
과제 1. 제시 자료 보고 말하기	1. 두 장의 제시 그림을 보고 공통점과 차이점을 비교해서 말하기 2. 두 장소 중 좋아하는 곳, 그 이유 말하기	1. 한국 전래 동화 그림 보고 스토리텔링하기 2. 이어지는 내용 상상해서 말하기	1. 비디오 자료 보고 스토리텔링하기 2. 이어지는 내용 상상해서 말하기
과제 2. 대화하기	친구와 여행 계획 세우기 - 서로 가고 싶은 여행지 얘기하기 - 자신이 말한 장소가 좋은 이유를 설명하고 여행지 결정하기(설명하기, 설득하기) - 교통편, 숙박, 여행지에서 할 일, 음식 등에 대한 계획 세우기(제안하기, 의사 결정하기)	좋아하는 드라마/영화에 대해서 이야기하기 - 좋아하는 드라마/영화, 최근에 재미있게 본 드라마, ..... - 출연 배우, 줄거리, 기억에 남는 장면, ..... - 친구에게 드라마/영화 추천하기	결혼에 대한 찬반 의견 나누기

## 1.2. 자료 구축 지침

### (1) 지침 수정 및 보완의 쟁점

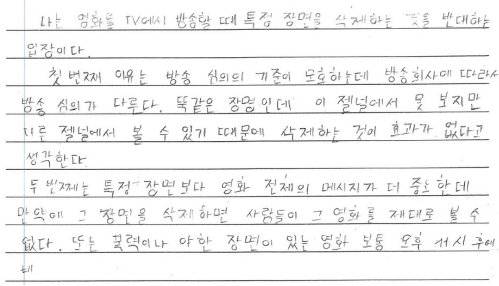
- 구축 지침은 문어 입력 지침과 구어 전사 지침으로 구분되어 있다. 두 지침은 국가 수준의 표준 말뭉치로서의 체계성과 호환성을 고려하여 <21세기 세종 한국어 균형 말뭉치>의 지침을 기본으로 하되, 비모어 화자인

한국어 학습자의 언어 특성을 고려하여 세부적인 내용을 수정·보완하는 방식으로 정교화해 왔다. 그간 문어 입력 지침의 경우 주로 입력 과정에서 발생하는 비정형의 문자 처리 지침을 해결하기 위한 세부 지침을 보완하는 데에 초점이 있었고, 구어 전사 지침의 경우는 구축 지원 도구 내에서의 주석 작업 지침을 정비하는 데에 초점을 두었다. 본 사업에서는 그간에 수정된 지침들의 정합성을 검증함과 동시에 입력과 전사 작업이 가공 작업과 긴밀한 관계가 있음을 고려하여 작업 단계별 연계성을 강화하는 데에 초점을 두어 작업 지침을 점검하고 관련 사항을 보완하였다. 아울러 작업자 간의 일관된 처리를 위해 기술 내용을 상세화하거나 구체적인 예시 자료를 추가하였다.

## (2) 수정 및 보완 내용

○ 문어 입력 지침과 구어 전사 지침의 수정·보완 내용은 다음과 같다.

<표 7> 구축 지침의 정교화 내용

단계	수정 및 보완 내용	해당 지침
문어 입력	미완결 텍스트에 대한 작업자 간의 일관된 판단을 위해 구체적인 예시 자료 추가	<p>&lt;신규 추가&gt;</p> 
	문장 부호 및 특수 기호 처리에 대한 기술의 상세화	<p>&lt;수정 전&gt;</p> <p>○ 문장부호는 학습자가 적어 넣은 대로 입력한다. 즉, 문장부호의 누락이나 생략, 중복 등을 그대로 반영한다.</p> <p>&lt;수정 후&gt;</p> <p>○ 문장부호는 학습자가 적어 넣은 대로 입력한다. 즉, 문장부호의 누락이나 생략,</p>

		중복 등을 그대로 반영한다. 단, 학습자가 부적절한 위치에 습관적으로 찍은 온점은 문장부호로 보기 어려우므로 반영하지 않는다.
	입력 지침 중 문장 및 단락 처리에 관한 사항 추가	<p>&lt;신규 추가&gt;</p> <p>○ 단락 구분은 하지 않으며 한 행에 한 문장을 입력하는 것을 원칙으로 한다.</p>
구어 전사	형태 주석, 오류 주석과의 연계를 고려한 '사람 이름 + 접사 '-이'의 처리 지침 추가	<p>&lt;신규 추가&gt;</p> <p>○ 사람 이름 뒤에 붙는 접사 '-이'는 &lt;name&gt; 태그 시, 함께 포함시킨다. 구어 전사 후에 수행되는 형태소 분석 시 발생할 수 있는 오류를 최소화시키기 위하여 구어 전사 과정에서 접사 '-이'를 포함시켜서 &lt;name&gt; 태그를 붙이기로 한다.</p> <div style="border: 1px solid black; padding: 10px; margin: 10px 0;"> <p>&lt;예&gt; 학습자의 이름이 '민영'일 때          학습자 발화: 민영이가, 민영이는, 민영이도, 민영이를          → 접사 '-이'를 포함시켜서 각각          '&lt;name&gt;가', '&lt;name&gt;는',          '&lt;name&gt;도', '&lt;name&gt;를'로 태그를 붙인다.</p> </div> <p>○ 대학교 이름은 '&lt;name&gt;대학교'로 붙여서 전사한다. 형태소 분석 시 해당 어절을 하나의 고유명사로 처리하기 때문에 팀 간의 작업상 편의를 위해 구어 전사 과정에서 '&lt;name&gt;대학교'로 붙이기로 한다. 만약 '대학교'에 발화 오류가 발생하면 ( ) 안에 &lt;name&gt;태그를 넣어서 마크업 처리를 한다.</p>



		<예> 학습자가 ‘대학교’를 [대학교]로 발음한 경우: <name1>대학교(<name1>대학 교)
--	--	---

### 1.3. 자료 가공 지침

#### (1) 지침 수정 및 보완의 쟁점

- 가공 지침은 형태 주석과 오류 주석 지침으로 나뉜다. 형태 주석 지침은 <21세기 세종 한국어 균형 말뭉치>의 현대 문어 형태 주석 말뭉치 분석 지침과의 호환성을 갖추고 분석의 일관성 제고를 목표로 일부 항목을 보완하고 추가해 왔다. 아울러 비정형의 학습자 언어를 일관성 있게 처리하기 위한 세부적인 지침을 마련하는 것에 주력해 왔다. 오류 주석 역시 오류에 대한 작업자 간의 처리 기준이 다르기 때문에 이를 일관성 있게 조정하는 것이 가장 큰 쟁점이었다. 본 사업에서는 이러한 관점을 지속적으로 반영하여 지침을 점검하였으며 구체적인 사례를 추가하거나 기술을 상세화하였다. 특히 다양하고 방대한 자료가 축적되기 시작하면서 이전에는 없었던 새로운 사례들이 출현하기도 하였는데, 이들 또한 지침에 추가되었다. 이와 함께 입력/전사-형태 주석-오류 주석으로 이어지는 작업 공정상 작업 단계별 지침 간의 연계성을 강화하면서 정합성을 제고하는 방향으로 지침을 수정·보완하였다.

#### (2) 수정 및 보완 내용

- 형태 주석 지침과 오류 주석 지침의 수정·보완 내용은 다음과 같다.

<표 8> 가공 지침의 정교화 내용

단계	수정 및 보완 내용	해당 지침
형태 분석	용언의 활용과 관련한 오류의 일관성 있는 형태 분석을 위해 분석 기준과 예시 추가	<p>&lt;신규 추가&gt;</p> <p>(3) 학습자가 활용 규칙을 몰라서 어미를 제대로 선택하지 못한 경우, 어간을 복원하지 않고 잘못된 어간 형태를 살려 분석한다.</p> <p>■■■ 공부가 <u>힘드는</u> [힘드/VA+는/ETM]</p> <p>(4) 학습자가 활용 규칙은 알지만 형태소 결합 규칙을 몰라서 잘못된 어절을 만든 경우에는 어간을 복원해서 분석한다.</p> <p>■■■ 공부가 <u>힘든지만</u> [힘들/VA+ㄴ/NA+지만/EC]</p> <p>■■■ 공부가 <u>힘든입니다</u>. [힘들/VA+ㄴ/NA+이/VCP+ㅂ니다/EF]</p> <p>(6) 학습자가 용언의 활용에서 동일한 형태(혹은 이형태)를 중복해서 사용한 경우는 같은 형태 표지로 분석한다.</p> <p>■■■ 풍선이 <u>부푼는</u> 것이 [부풀/VV+ㄴ/ETM+는/ETM]</p> <p>■■■ <u>다른는</u> 게 되게 많이 [다르/VA+ㄴ/ETM+는/ETM]</p> <p>■■■ <u>불평등한는</u> 생각도 많아요. [불/XPN+평등/NNG+하/XSA+ㄴ/ETM+는/ETM]</p> <p>■■■ 이권엔 <u>지나지지</u> 않았습니다 [지나/VV+지/EC+지/EC]</p>
	형태소 경계를 분할하기 어려운 오류 어절의 일관성 있는 형태 분석을 위해 분석 기준을 구체화하여 서술하고 관련된 예시 추가	<p>&lt;수정 전&gt;</p> <p>○ 일부 오류의 경우는 분석한 형태의 결합이 원어절의 형태가 되지 않더라도 기본적으로는 분할을 원칙으로 다음과 같이 처리한다.</p> <p>■■■ 쓰레기를 <u>버려도</u> 되면 좋겠습니다. [버리/VV+어도/EC]</p> <p>■■■ 많은 친구를 <u>사귀서</u> 재밌었어요. [사구/VV+어서/EC]</p>

단계	수정 및 보완 내용	해당 지침
		<p>■■■ 소치에 2시간 <u>걸레요</u>. [걸레/VV+어요/EF]</p> <p>■■■ 점점 <u>심해질</u> 것이다. [심해/VA+어/EC+지/VX+ㄴ/ETM]</p> <p>&lt;수정 후&gt; (3) 기본적으로는 분석한 형태의 결합이 원어절이 되도록 분할해야 하지만 그 경계 분할이 어려운 오류 어절의 경우, 어간에 학습자가 잘못 쓴 오형태를 주며 어미는 해당 오류 어절이 취해야 하는 어미의 형태를 부여한다.</p> <p>■■■ 쓰레기를 <u>버러도</u> 되면 좋겠습니다. [버러/VV+어도/EC]</p> <p>■■■ 많은 친구를 <u>사귀서</u> 재밌었어요. [사귀/VV+어서/EC]</p> <p>■■■ 소치에 2시간 <u>걸레요</u>. [걸레/VV+어요/EF]</p> <p>■■■ 점점 <u>심해질</u> 것이다. [심해/VA+어/EC+지/VX+ㄴ/ETM]</p> <p>■■■ A의 주장에 <u>반에</u> [반에/VV+어/EC]</p> <p>■■■ 제 꿈을 <u>위에</u> [위에/VV+어/EC]</p> <p>■■■ 제 꿈을 <u>위에서</u> [위에서/VV+어서/EC]</p>
	지침 기술의 오류 수정, 본 지침에서는 세종 현대 구어 말뭉치의 분석 결과를 따라 (2)와 같은 경우를 ‘이요/JX’로 분석하는 것으로 수정	<p>&lt;수정 전&gt; (1) A: 선생님이 집에 오셨었어요. B: <u>선생님이요?</u> [선생/NNG+님/XSN+이/JKS+요/JX]</p> <p>A: 커서 선생님이 되는 게 어떠니? B: <u>선생님이요?</u> [선생/NNG+님/XSN+이/JKC+요/JX]</p>

단계	수정 및 보완 내용	해당 지침
		<p>(2) A: 선생님에 대해 알고 있니?  B: <u>선생님이요?</u>  [선생/NNG+님/XSN+이/VCP+요/EF]</p> <p>&lt;수정 후&gt;  (1) A: 선생님이 집에 오셨었어.  B: <u>선생님이요?</u>  [선생/NNG+님/XSN+이/JKS+요/JX]</p> <p>A: 커서 선생님이 되는 게 어떠니?  B: <u>선생님이요?</u>  [선생/NNG+님/XSN+이/JKC+요/JX]</p> <p>(2) A: 선생님에 대해 알고 있니?  B: <u>선생님이요?</u>  [선생/NNG+님/XSN+이요/JX]</p> <p>A: 가장 좋아하는 명절이 언제예요?  B: <u>설날이요.</u> [설날/NNG+이요/JX]</p>
오류 주석	구어 발음 층위에서 음운 규칙 오류와 음소 오류 판별 기준 추가	<p>&lt;신규 추가&gt;  ○ 학습자가 음운규칙을 몰라서 발생시킨 오류 인지 음소를 변별하지 못하여 발생시킨 오류 인지를 판단하여 오류 층위에 음운규칙과 음소를 구분하여 주석한다.</p> <div style="border: 1px solid black; padding: 10px; margin-top: 10px;"> <p>&lt;예&gt;     한국어(√ 한국어)  ☞ 음운규칙 미적용 오류. 오류 위치: 명사, 층위: 음운규칙(PC)</p> <p>한국거(√ 한국어)  ☞ 오류 위치: 명사, 층위: 음운규칙(PC)</p> <p>한구껴(√ 한국어)  ☞ 음운규칙은 적용하였으나 음</p> </div>

단계	수정 및 보완 내용	해당 지침
		<div data-bbox="665 295 1173 556" style="border: 1px solid black; padding: 10px; margin-bottom: 10px;"> <p style="text-align: center;">소 변별 실패 오류 위치: 명사, 층위: 음소(PP) 한국계(√ 한국어) ☞ 오류 위치: 명사, 층위: 음운규칙(PC), 음소(PP)</p> </div> <p>○ 한 단어 내에서 음운 규칙 또는 음소와 관련된 오류가 나타났을 때에는 그 단어의 품사를 주석하면 되지만, 체언과 조사, 용언의 어간과 어미 사이에서 음운 규칙이 적용되면서 나타난 오류의 경우, 오류 위치를 무엇으로 삼을 것인가가 문제가 된다. 이때에는 오류가 나타난 발음이 어느 부분의 영향을 받았는지를 고려하여 처리한다. 예를 들어, 모음으로 시작되는 조사나 어미의 경우 본래 음가를 가지고 있지 않았으나, 선행하는 음절의 받침 발음에 영향을 받아 초성에서 그 발음이 실현된다. 따라서 조사나 어미의 초성 위치에서 오류가 나타나더라도 오류 위치를 영향을 미친 체언과 용언의 어간으로 주석한다.</p> <div data-bbox="665 1209 1173 1659" style="border: 1px solid black; padding: 10px;"> <p>&lt;예&gt; 한국계(√ 한국어) 갔어요. ☞ 오류 위치: 고유명사, 오류 층위: 음운규칙(PC) 한국계(√ 한국어) 갔어요. ☞ 오류 위치: 고유명사, 오류 층위: 음운규칙, 음소(종성이 그 대로 실현됨과 동시에 [ㄱ]와 [ㄱ] 음소를 변별하지 못한 오류로 해석한다.) 네 말이 맞쵸요(√ 맞아요). ☞ 오류 위치: 형용사, 오류 층</p> </div>

단계	수정 및 보완 내용	해당 지침
		<div> <p>위: 음운규칙(PC), 음소(PP) 음식글(√음식을) 먹어요. ☞ 오류 위치: 명사, 오류 층위: 음운규칙(PC)</p> </div> <p>○ 연음이 적용되는 단어에서 연음은 시켰지만 음소를 잘못 발음하였을 때에는 음소 오류로 처리한다.</p> <div> <p>&lt;예&gt; 네 말이 마차요(√맞아요). ☞ 오류 위치: 형용사, 오류 층위: 음소(PP) 음식글(√음식을) 먹어요. ☞ 오류 위치: 명사, 오류 층위: 음소(PP)</p> </div> <p>○ 종성이 다음 음절의 초성으로 넘어가서 발음되는 연음과 달리, 경음화나 격음화 오류는 오류의 양상에 따라, 발음된 위치를 고려하여 오류 위치를 주석한다.</p> <div> <p>&lt;예&gt; 물건을 찰코(√찾고[찰꼬]). ☞ 오류 위치: 연결어미, 오류 층위: 음소(PP) [찰꼬]로 발음되어야 하는데, 연결어미 ‘고’를 ‘코’로 발음하였다. 이것이 앞의 종성 [ㄷ]의 영향을 받은 것인지, 학습자가 초성 위치에서 [ㄱ]와 [ㅋ]를 변별하지 못한 것인지를 판별하기는 어렵다. 본래 음가를 가지고 있지 않은 어미가 연음 규칙에 따라 발음을 잘못하였을 때에는 선행하는 어간의 품사를 오류 위치로</p> </div>

단계	수정 및 보완 내용	해당 지침
		<p>주석한다. 그러나 ‘고’처럼 본래 음가를 가지고 있던 위치에서 음소를 변별하지 못하였을 때에는 해당 위치에 오류 위치를 주석한다. 따라서 이러한 경우 오류 위치: 연결어미, 오류 층위: 음소(PP)로 주석한다.</p> <p>응답뻘(√응답한) 결과        ☞ 오류 위치: 명사, 오류 층위: 음운규칙(PC)        ☞ 오류 위치: 동사파생접시마, 오류 층위: 음소(PP)        격음화가 적용되어 [응다판]으로 발음해야 하지만, [응답]을 그대로 발음하는 동시에 격음으로 발음해야 하는 [판]을 [뻘]으로 발음하였다. 따라서 ‘응답’과 동사파생접미사 ‘하’를 형태 분리하므로 ‘응답’에는 음운규칙을 주석하고, ‘하’에는 음소를 주석한다.</p>
	담화 표지 오류의 실제 사례 추가	<p>&lt;신규 추가&gt;</p> <p>&lt;예&gt; 아노(√담화 표지)... 음..        ☞ 일본인 학습자가 한국어로 발화하기 전, L1의 간투사를 그대로 사용한 경우이다. L1을 사용하는 것이 오류라고 할 수는 없지만, 학습자가 구어 발화에서 L1을 어떻게 사용하고 있는지, 발화 전략 등을 파악할 수 있다는 차원에서 이에 대해 담화 표지를 부착한다. 해당 용례의 경우 감탄사(IC)로 형태 주석하므</p>

단계	수정 및 보완 내용	해당 지침
		로 오류 위치를 감탄사로 주석하고, 오류 증위에 담화 표지를 주석한다.

## 2. 말뭉치 활용 연구 체계 구축

- 말뭉치 활용 연구 체계 구축은 1) 2015-2019년 구축 말뭉치를 활용한 분석 연구, 2) 학습자 말뭉치 활용 연구, 3) 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구로 나누어 진행하였다.

### 2.1. 2015-2019년 구축 말뭉치를 활용한 분석 연구<sup>1)</sup>

- 한국어 학습자 말뭉치는 다양한 변인에 따른 학습자의 중간언어의 특성이나 언어 발달 양상을 관찰하기 위한 연구 자료로의 활용도가 무엇보다도 가장 높다. 이에 따라 본 절에서는 2015-2019년까지 구축된 형태 주석 말뭉치와 오류 주석 말뭉치의 주석 통계를 말뭉치 유형별(구어/문어), 숙달도 단계별, 국적별로 분석하여 제시하기로 한다.

1) 본 장의 분석 결과는 2019년 11월 23일 한국어 학습자 말뭉치 구축 도구에 구축된 자료의 통계를 기준으로 한 것으로 향후 기구축 말뭉치 수정·보완 과정에서 세부적인 통계는 변경될 수 있음.



## (1) 형태 주석 결과의 분석<sup>2)</sup>

### ① 문어

#### 가. 숙달도 단계별

<표 9> 문어 형태 주석 결과\_숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	합계
표본 수	4,256	3,397	3,034	2,727	2,134	1,644	74	17,266
어절 수	294,657	342,255	374,356	356,401	323,615	273,695	72,769	2,037,748
주석 수 합계	552,662	650,873	724,314	698,389	641,079	544,199	133,488	3,945,004
일반명사	111,724	127,502	148,171	154,261	153,493	134,058	40,149	869,358
고유명사	26,489	24,710	14,132	6,586	5,771	7,116	3,942	88,746
의존명사	18,614	19,331	24,501	24,739	22,478	19,966	4,678	134,307
대명사	20,665	20,557	18,548	15,133	9,994	8,322	1,105	94,324
수사	3,446	942	1,022	915	931	712	276	8,244
긍정지정사	9,378	8,240	10,158	10,351	10,403	9,108	2,367	60,005
부정지정사	48	329	1,128	1,414	1,355	1,413	192	5,879
동사	51,613	63,846	73,741	69,997	58,624	47,641	8,927	374,389
형용사	16,422	25,821	25,359	23,074	19,869	13,801	2,130	126,476
보조용언	6,991	10,682	15,039	14,546	13,097	10,895	1,949	73,199
관형사	4,283	5,859	9,616	9,545	8,623	7,481	1,654	47,061
일반부사	22,985	32,806	33,489	29,313	22,802	17,760	3,447	162,602
접속부사	9,323	8,277	7,748	5,999	4,445	3,306	615	39,713
감탄사	49	146	76	55	38	38	13	415
주격조사	15,808	21,954	26,292	25,327	23,860	18,584	3,514	135,339
보격조사	135	584	1,145	1,794	1,900	1,935	311	7,804
관형격조사	4,376	5,268	6,363	7,582	9,075	8,908	3,994	45,566
목적격조사	24,634	23,479	26,854	27,685	24,611	22,181	5,412	154,856
부사격조사	42,389	40,711	36,476	31,295	28,565	24,914	7,577	211,927

2) 본 보고서에 제시한 형태 주석 통계는 숫자, 기호, 분석 불능 처리된 것을 제외한 것임.

	1급	2급	3급	4급	5급	6급	6급 이상	합계
호격조사	4	58	13	8	9	2	1	95
인용격조사	21	317	290	206	222	247	44	1,347
보조사	29,371	31,608	31,839	30,832	27,953	22,217	4,732	178,552
접속조사	4,669	3,791	3,651	3,543	3,722	3,076	1,359	23,811
선어말어미	16,655	20,856	21,159	13,666	9,078	10,262	2,169	93,845
종결어미	60,879	52,560	47,201	39,252	32,277	24,979	4,986	262,134
연결어미	22,056	44,077	60,284	60,918	53,887	44,286	8,168	293,676
명사형전성어미	590	3,961	4,525	5,096	5,250	3,633	823	23,878
관형형전성어미	12,677	26,487	39,203	42,849	42,607	36,168	8,223	208,214
체언접두사	21	44	137	630	685	673	258	2,448
명사파생접미사	3,747	6,207	8,240	10,935	12,108	12,176	3,302	56,715
동사파생접미사	8,505	12,792	18,346	20,665	22,482	20,122	5,264	108,176
형용사파생접미사	2,887	5,364	7,502	8,092	8,242	5,963	1,319	39,369
어근	1,208	1,707	2,066	2,086	2,623	2,256	588	12,534

## 나. 언어권별

<표 10> 문어 형태 주석 결과\_언어권별

	중국어	일본어	베트남어	영어	러시아어	기타 <sup>3)</sup>	합계
표본 수	6,175	3,058	1,670	1,500	851	4,012	17,266
어절 수	742,321	375,308	183,105	178,485	91,995	466,534	2,037,748
주석 수 합계	1,420,750	746,121	356,320	345,201	177,874	898,738	3,945,004
일반명사	326,440	156,182	78,847	74,392	38,764	194,733	869,358
고유명사	30,574	15,882	7,943	8,707	3,896	21,744	88,746
의존명사	46,507	27,956	11,562	11,869	5,904	30,509	134,307
대명사	34,661	14,284	9,243	7,703	4,741	23,692	94,324
수사	3,034	1,139	937	812	326	1,996	8,244
긍정지정	20,327	12,206	5,796	4,847	2,715	14,114	60,005
부정지정사	1,925	1,259	488	565	284	1,358	5,879
동사	129,940	74,836	32,374	33,578	17,443	86,218	374,389

	중국어	일본어	베트남어	영어	러시아어	기타 <sup>3)</sup>	합계
형용사	46,183	25,106	11,125	11,264	5,525	27,273	126,476
보조용언	24,701	16,428	6,624	6,192	3,122	16,132	73,199
관형사	16,463	8,235	3,742	4,403	2,328	11,890	47,061
일반부사	62,384	26,576	15,222	14,529	7,384	36,507	162,602
접속부사	14,774	6,426	3,852	3,335	1,682	9,644	39,713
감탄사	163	43	31	63	14	101	415
주격조사	51,090	25,369	11,508	12,074	6,144	29,154	135,339
보격조사	2,870	1,918	603	566	351	1,496	7,804
관형격조사	17,350	7,455	3,980	4,134	2,014	10,633	45,566
목적격조사	56,222	28,116	14,575	13,287	7,217	35,439	154,856
부사격조사	76,011	38,464	18,404	18,879	9,962	50,207	211,927
호격조사	41	7	14	16	3	14	95
인용격조사	370	394	70	141	65	307	1,347
보조사	63,049	38,682	15,766	14,025	7,399	39,631	178,552
접속조사	7,708	4,159	2,482	2,291	1,117	6,054	23,811
선어말어미	29,597	18,083	8,387	9,743	4,354	23,681	93,845
종결어미	99,503	42,955	24,820	22,334	12,342	60,180	262,134
연결어미	98,379	63,629	27,096	25,852	12,883	65,837	293,676
명사형전성어미	7,832	5,216	2,103	2,224	1,180	5,323	23,878
관형형전성어미	71,749	46,103	17,336	18,317	8,928	45,781	208,214
체언접두사	1,110	469	133	214	106	416	2,448
명사파생접미사	20,770	8,583	4,884	5,461	2,958	14,059	56,715
동사파생접미사	39,535	21,066	11,071	8,875	4,485	23,144	108,176
형용사파생접미사	14,835	6,792	3,934	3,398	1,721	8,689	39,369
어근	4,653	2,103	1,368	1,111	517	2,782	12,534

3) 문어 형태 주석 말뭉치는 총 82개 언어권의 자료로 이루어져 있으며 기타에는 몽골어, 타이어, 프랑스어, 광둥어, 스페인어, 인도네시아어, 카자흐어, 아랍어, 우즈베크어, 말레이어, 타갈로그어, 크메르어, 심할라어, 독일어, 스웨덴어, 이탈리아어, 키르기스어, 버마어, 벵골어, 네팔어, 한국어, 우르두어, 터키어, 포르투갈어, 네덜란드어, 투르크멘어, 페르시아어, 아제르바이잔어, 노르웨이어, 우크라이나어, 힌디어, 폴란드어, 암하라어, 르완다어, 핀란드어, 라오어, 스와힐리어, 타밀어, 루마니아어, 불가리아어, 세르비아어, 체코어, 타지크어, 테툼어, 헝가리어, 히브리어, 덴마크어, 노르웨이어 (뉘노르스크), 아르

## ② 구어

### 가. 숙달도 단계별

<표 11> 구어 형태 주식 결과\_숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	계
표본 수	244	206	251	237	174	128	15	1,255
어절 수	82,738	95,301	109,421	107,170	95,175	80,586	21,112	591,503
주식 수 합계	124,891	146,374	173,807	172,302	160,093	136,899	33,934	948,300
일반명사	21,045	23,412	29,569	30,463	31,909	25,495	6,284	168,177
고유명사	4,966	5,342	4,526	3,980	3,318	1,905	812	24,849
의존명사	3,270	4,121	4,582	5,133	5,361	4,380	1,118	27,965
대명사	3,820	3,597	4,522	4,192	3,385	3,349	1,283	24,148
수사	1,186	1,138	1,063	1,035	1,196	439	302	6,359
긍정지정	1,714	1,737	2,057	2,318	2,539	1,955	600	12,920
부정지정사	101	217	382	333	306	353	150	1,842
동사	11,507	13,257	16,475	15,373	13,234	11,298	2,901	84,045
형용사	4,184	5,985	5,898	5,651	4,418	4,512	1,037	31,685
보조용언	1,596	1,994	2,573	2,973	2,368	2,295	516	14,315
관형사	1,442	2,035	2,779	3,390	3,275	2,924	789	16,634
일반부사	6,142	8,023	9,262	9,138	7,103	7,167	2,128	48,963
접속부사	1,650	2,199	2,582	2,439	2,023	1,800	618	13,311
감탄사	19,069	21,413	22,162	19,980	15,058	12,958	2,744	113,384
주격조사	2,187	3,227	4,417	4,483	4,518	3,927	642	23,401
보격조사	20	62	142	169	224	214	52	883
관형격조사	350	377	600	813	1,271	1,221	195	4,827
목적격조사	2,197	2,222	3,267	3,017	3,910	3,311	580	18,504
부사격조사	5,653	5,917	6,867	6,157	6,201	4,723	1,249	36,767

메니아어, 쿠르드어, 마다가스카르어, 카탈루냐어, 텔루구어, 자바어, 조지아어, 이그보어, 리투아니아어, 말라얄람어, 슬로바키아어, 구자라트어, 위구르어, 룩셈부르크어, 간다어, 마오리어, 판테어(Fanti), 아프리카ンス어, 그리스어, 비술라마어, 벨라루스어, 티그리냐어, 슬로베니아어, 편자브어, 라틴어, 파슈토어, 에스토니아어, 덩카어, 마라티어가 포함된다.

	1급	2급	3급	4급	5급	6급	6급 이상	계
호격조사	3	10	2	6	4	5	9	39
인용격조사	4	18	39	38	96	37	3	235
보조사	4,323	5,352	6,472	7,161	6,732	6,364	1,485	37,889
접속조사	654	542	695	597	828	592	148	4,056
선어말어미	3,315	4,298	5,660	4,989	3,941	3,112	888	26,203
종결어미	14,049	13,458	13,145	10,970	8,903	7,094	2,215	69,834
연결어미	4,920	7,882	11,576	12,895	11,358	10,625	2,456	61,712
명사형전성어미	61	257	330	513	581	569	92	2,403
관형형전성어미	1,995	3,856	5,808	6,677	7,399	6,560	1,250	33,545
체언접두사	3	12	14	50	77	117	23	296
명사파생접미사	658	1,097	1,526	1,612	2,286	2,257	431	9,867
동사파생접미사	2,005	2,338	3,203	4,084	4,606	3,961	721	20,918
형용사파생접미사	585	716	1,204	1,200	1,312	1,026	163	6,206
어근	217	263	408	473	353	354	50	2,118

## 나. 언어권별

<표 12> 구어 형태 주식 결과\_언어권별

	중국어	일본어	베트남어	영어	러시아어	기타 <sup>4)</sup>	합계
표본 수	398	247	223	83	70	249	1,270
어절 수	191,669	112,717	113,949	41,735	26,643	104,790	591,503
주식 수 합계	305,833	191,321	175,790	65,758	42,180	175,009	955,891
일반명사	55,214	32,977	31,189	10,953	7,378	31,526	169,237
고유명사	7,712	4,715	5,473	2,010	926	4,231	25,067
의존명사	9,243	6,214	4,669	1,977	1,132	4,999	28,234
대명사	7,966	3,932	4,623	1,435	1,442	5,020	24,418
수사	1,827	1,119	1,235	477	294	1,460	6,412
긍정지정	4,107	3,177	2,190	846	516	2,180	13,016
부정지정사	709	363	235	101	89	354	1,851
동사	27,020	17,707	15,833	5,808	3,592	14,840	84,800

	중국어	일본어	베트남어	영어	러시아어	기타 <sup>4)</sup>	합계
형용사	10,259	6,401	5,881	2,241	1,547	5,719	32,048
보조용언	4,511	3,334	2,830	927	572	2,249	14,423
관형사	5,872	3,361	2,661	989	816	3,074	16,773
일반부사	16,558	8,224	10,115	3,532	2,374	8,558	49,361
접속부사	4,550	2,368	2,112	937	640	2,882	13,489
감탄사	34,811	19,696	23,310	9,354	5,124	22,131	114,426
주격조사	7,625	5,753	3,359	1,673	1,092	4,078	23,580
보격조사	336	242	100	34	31	143	886
관형격조사	1,737	900	701	298	204	1,016	4,856
목적격조사	6,106	4,598	2,030	1,431	887	3,580	18,632
부사격조사	11,526	7,813	6,321	2,722	1,705	7,013	37,100
호격조사	19	2	8	2	1	14	46
인용격조사	45	111	19	18	5	38	236
보조사	12,381	7,967	6,528	2,179	1,616	7,452	38,123
접속조사	1,239	903	813	275	152	684	4,066
선어말어미	7,804	5,481	5,134	1,866	1,128	4,942	26,355
종결어미	21,431	12,371	14,872	5,591	3,039	13,184	70,488
연결어미	20,138	14,359	11,224	3,527	2,752	10,164	62,164
명사형전성어미	806	624	315	146	90	460	2,441
관형형전성어미	11,346	8,465	4,862	1,991	1,368	5,764	33,796
체언접두사	137	65	25	11	7	51	296
명사파생접미사	3,431	1,997	1,456	692	450	1,865	9,891
동사파생접미사	6,756	4,228	4,145	1,156	884	3,867	21,036
형용사파생접미사	1,943	1,304	1,162	386	269	1,159	6,223
어근	668	550	360	173	58	312	2,121

4) 구어 형태 주석 말뭉치는 35개 언어권의 자료로 이루어져 있으며 타이어, 몽골어, 키르기스어, 타갈로그어, 인도네시아어, 스페인어, 크메르어, 프랑스어, 우즈베크어, 세부아노어, 아랍어, 광둥어, 독일어, 이탈리아어, 싱할라어, 말레이어, 스웨덴어, 카자흐어, 버마어, 네팔어, 우르두어, 아제르바이잔어, 벵골어, 불가리아어, 페르시아어, 네덜란드어, 터키어, 아르메니아어, 헝가리어, 세르비아어가 포함된다.

## (2) 오류 주식 결과의 분석

### ① 문어

가. 분석 불가능 여부

ㄱ. 숙달도 단계별

<표 13> 문어 오류 주식 결과\_분석 불가능 여부: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
표본 수	1,048	749	685	642	544	481	4,149
어절 수	74,850	75,281	82,047	78,519	76,492	75,133	462,322
분석 불가능	397	862	1,004	911	761	340	4,275

ㄴ. 언어권별

<표 14> 문어 오류 주식 결과\_분석 불가능 여부: 언어권별

	중국어	영어	베트남어	러시아어	타이어	기타 <sup>5)</sup>	합계
표본 수	902	932	452	266	193	1,404	4,149
어절 수	102,001	106,612	50,274	29,672	24,095	149,668	462,322
분석 불가능	1,218	1,100	869	393	295	400	4,275

나. 오류 위치

ㄱ. 숙달도 단계별

<표 15> 문어 오류 주식 결과\_오류 위치: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
표본 수	1,048	749	685	642	544	481	4,149

5) 문어 오류 주식 말뭉치는 51개 언어권으로 이루어져 있으며 기타에는 일본어, 우즈베크어, 우르두어, 독일어, 타갈로그어, 몽골어, 아랍어, 인도네시아어, 크메르어, 텔루구어, 스페인어, 카자흐어, 네팔어, 광둥어, 프랑스어, 벵골어, 말레이어, 네덜란드어, 스웨덴어, 폴란드어, 한국어, 이탈리아어, 버마어, 키르기스어, 터키어, 힌디어, 페르시아어, 루마니아어, 아제르바이잔어, 헝가리어, 카탈루냐어, 파슈토어, 마다가스카르어, 르완다어, 자바어, 암하라어, 핀란드어, 싱할라어, 히브리어, 라오어, 불가리아어, 우크라이나어, 노르웨이어, 리투아니아어, 세르비아어, 슬로바키아어가 포함된다.

	1급	2급	3급	4급	5급	6급	합계
어절 수	74,850	75,281	82,047	78,519	76,492	75,133	462,322
주석 수 합계	14,882	15,996	17,486	16,908	12,438	11,336	89,046
고유명사	439	223	151	78	62	62	1,015
일반명사	2,210	2,012	2,354	2,426	2,135	2,068	13,205
의존명사	460	451	550	646	396	403	2,906
대명사	247	433	449	426	205	194	1,954
수사	92	35	28	20	24	24	223
동사	968	1,278	1,478	1,585	1,143	1,067	7,519
형용사	448	404	435	379	260	227	2,153
보조용언	137	207	351	322	226	205	1,448
지정사	124	132	125	107	91	106	685
관형사	76	80	117	138	116	91	618
일반부사	604	615	613	570	356	311	3,069
접속부사	287	182	158	128	77	52	884
감탄사	2	1	1		1	1	6
체언접두사	1	2	5	10	7	10	35
명사파생접미사	47	67	87	109	140	79	529
동사파생접미사	41	67	104	127	151	160	650
형용사파생접미사	13	11	23	38	21	40	146
어근	29	35	72	63	53	58	310
주격조사	1,122	1,159	1,317	1,267	912	816	6,593
관형격조사	170	213	239	240	284	273	1,419
목적격조사	1,060	1,006	1,078	1,048	790	654	5,636
부사격조사	2,061	1,697	1,512	1,327	959	845	8,401
호격조사	1	3				1	5
인용격조사	2	16	19	8	4	11	60
보조사	975	1,053	962	993	758	651	5,392
연결어미	715	1,265	1,582	1,498	1,020	878	6,958
종결어미	635	750	713	637	323	240	3,298
선어말어미	545	921	803	624	311	341	3,545



	1급	2급	3급	4급	5급	6급	합계
명사형 전성어미	40	109	125	106	111	92	583
관형사형 전성어미	294	538	788	810	644	656	3,730
접속조사	583	264	207	136	87	60	1,337
보격조사	8	36	56	63	43	39	245
구단위표현	25	41	59	58	65	61	309
표현 문형	421	690	925	921	663	560	4,180

#### ㄴ. 언어권별

<표 16> 문어 오류 주석 결과\_오류 위치: 언어권별

	일본어	영어	중국어	베트남어	러시아어	기타 <sup>6)</sup>	합계
표본 수	1,045	932	902	452	266	552	4,149
어절 수	112,428	106,612	102,001	50,274	29,672	61,335	462,322
주석 수 합계	14,122	22,670	21,806	11,370	6,078	13,000	89,046
고유명사	145	367	183	95	77	148	1,015
일반명사	2,456	3,151	3,208	1,573	884	1,933	13,205
의존명사	506	742	698	360	231	369	2,906
대명사	383	474	389	282	127	299	1,954
수사	23	61	50	30	15	44	223
동사	1,312	1,935	1,563	1,027	565	1,117	7,519
형용사	360	593	458	239	141	362	2,153
보조용언	259	355	334	217	89	194	1,448
지정사	125	159	172	84	53	92	685
관형사	95	177	123	83	36	104	618
일반부사	528	802	699	385	216	439	3,069
접속부사	136	243	205	103	61	136	884
감탄사		1	4			1	6
체언접두사	3	7	19	3	2	1	35
명사파생접미사	79	131	136	68	35	80	529
동사파생접미사	107	154	200	80	42	67	650
형용사파생접미사	30	40	32	16	7	21	146

	일본어	영어	중국어	베트남어	러시아어	기타 <sup>6)</sup>	합계
어근	55	64	93	44	19	35	310
주격조사	629	1,804	1,927	811	508	914	6,593
관형격조사	229	387	368	140	96	199	1,419
목적격조사	495	1,537	1,625	795	358	826	5,636
부사격조사	1,433	2,155	1,952	995	567	1,299	8,401
호격조사		1	1	2		1	5
인용격조사	19	14	13	1	3	10	60
보조사	545	1,389	1,410	729	413	906	5,392
연결어미	1,290	1,683	1,630	932	466	957	6,958
종결어미	468	761	914	363	264	528	3,298
선어말어미	508	861	1,009	457	165	545	3,545
명사형 전성어미	87	158	120	95	41	82	583
관형사형 전성어미	797	891	907	502	213	420	3,730
접속조사	206	365	316	170	70	210	1,337
보격조사	39	82	39	31	16	38	245
구단위표현	40	95	62	45	18	49	309
표현 문형	735	1,031	947	613	280	574	4,180

## 다. 오류 층위

### ㄱ. 숙달도 단계별

<표 17> 문어 오류 주석 결과\_오류 층위: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
표본 수	1,048	749	685	642	544	481	4,149
어절 수	74,850	75,281	82,047	78,519	76,492	75,133	462,322
주석 수 합계	3,760	4,133	4,118	3,618	2,218	2,023	19,870

6) 기타에는 타이어, 스페인어, 몽골어, 인도네시아어, 프랑스어, 아랍어, 이탈리아어, 독일어, 스웨덴어, 카자흐어, 말레이어, 광둥어, 키르기스어, 버마어, 우즈베크어, 싱할라어, 페르시아어, 터키어, 크메르어, 네덜란드어, 헝가리어, 세르비아어, 네팔어, 아제르바이잔어, 불가리아어가 포함된다.

		1급	2급	3급	4급	5급	6급	합계
형태	단어 형성 (합성법)	24	25	22	25	18	12	126
	단어 형성 (파생법)	220	193	248	193	185	154	1,193
	굴절(곡용)	234	166	123	129	74	62	788
	굴절(활용)	654	1,027	982	893	478	353	4,387
	품사	270	266	285	197	146	135	1,299
통사	높임	247	320	367	417	159	133	1,643
	시제	393	817	844	581	336	377	3,348
	사동	3	30	53	85	123	133	427
	피동	12	72	139	160	179	188	750
	부정	42	86	100	90	49	63	430
	어순	323	252	207	169	85	93	1,129
담화	지시	45	61	53	59	45	35	298
	접속	193	109	109	84	62	47	604
	담화표지	2	1	2				5
	구어/문어 오류	1,098	708	584	536	279	238	3,443

#### ㄴ. 언어권별

<표 18> 문어 오류 주석 결과\_오류 층위: 언어권별

	중국어	영어	일본어	베트남어	러시아어	기타	합계
표본 수	902	932	1,045	452	266	552	4,149
어절 수	102,001	106,612	112,428	50,274	29,672	61,335	462,322
주석 수 합계	5,038	4,826	3,515	2,452	1,235	2,804	19,870

형태	단어 형성 (합성법)	19	37	24	25	7	14	126
	단어 형성 (파생법)	312	295	159	163	84	180	1,193
	굴절(곡용)	205	171	176	106	24	106	788
	굴절(활용)	1,249	997	769	452	308	612	4,387
	품사	303	337	174	194	97	194	1,299
통사	높임	361	357	285	255	110	275	1,643
	시제	981	800	565	403	143	456	3,348
	사동	136	77	51	69	29	65	427
	피동	173	179	155	107	59	77	750
	부정	88	101	85	68	19	69	430
	어순	230	287	156	185	72	199	1,129
담화	지시	42	81	64	49	23	39	298
	접속	127	151	106	81	47	92	604
	담화표지		2	3				5
	구어/문어 오류	812	954	743	295	213	426	3,443

## 라. 오류 양상

### ㄱ. 숙달도 단계별

<표 19> 문어 오류 주석 결과\_오류 양상: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
표본 수	1,048	749	685	642	544	481	4,149
어절 수	74,850	75,281	82,047	78,519	76,492	75,133	462,322
주석 수 합계	14,776	15,613	17,085	16,420	12,218	11,070	87,182
누락	2,576	2,811	2,792	2,611	1,874	1,533	14,197
첨가	883	1,154	1,427	1,566	1,105	1,001	7,136
대치	6,499	7,486	8,659	8,279	6,493	6,146	43,562
오 형태	4,818	4,162	4,207	3,964	2,746	2,390	22,287

## ㄴ. 언어권별

<표 20> 문어 오류 주석 결과\_오류 양상: 언어권별

	중국어	일본어	영어	베트남어	러시아어	기타 <sup>7)</sup>	합계
표본 수	932	902	1,045	452	266	552	4,149
어절 수	106,612	102,001	112,428	50,274	29,672	61,335	462,322
주석 수 합계	21,994	21,733	13,688	11,004	5,924	12,839	87,182
누락	3,824	3,938	1,531	1,796	992	2,116	14,197
첨가	1,916	1,416	1,106	1,112	478	1,108	7,136
대치	11,039	10,403	6,990	5,870	2,999	6,261	43,562
오형태	5,215	5,976	4,061	2,226	1,455	3,354	22,287

## ② 구어

### 가. 분석 불가능 여부

#### ㄱ. 숙달도 단계별

<표 21> 구어 오류 주석 결과\_분석 불능 여부: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	합계
표본 수	140	123	166	170	100	63	5	767
어절 수	45,175	54,175	62,577	63,474	59,182	47,360	4,892	336,835
분석 불가능	123	263	170	350	188	79	7	1,180

7) 구어 오류 주석 말뭉치는 30개 언어권으로 이루어져 있으며 기타에는 타이어, 스페인어, 몽골어, 인도네시아어, 프랑스어, 아랍어, 이탈리아어, 말레이어, 광둥어, 독일어, 스웨덴어, 카자흐어, 버마어, 우즈베크어, 싱할라어, 키르기스어, 헝가리어, 세르비아어, 네팔어, 아제르바이잔어, 불가리아어, 페르시아어, 터키어, 크메르어, 네덜란드어가 포함된다.

## ㄴ. 언어권별

<표 22> 구어 오류 주석 결과\_분석 불능 여부: 언어권별

	일본어	베트남어	중국어	영어	러시아어	기타	합계
표본 수	224	176	163	78	30	96	767
어절 수	97,752	75,508	74,971	36,748	12,980	38,876	336,835
분석 불가능	260	321	331	82	34	152	1,180

## 나. 오류 위치

### ㄱ. 숙달도 단계별

<표 23> 구어 오류 주석 결과\_오류 위치: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	합계
표본 수	140	123	166	170	100	63	5	767
어절 수	45,175	54,175	62,577	63,474	59,182	47,360	4,892	336,835
주석 수 합계	5,486	6,633	8,731	9,354	9,299	5,551	637	45,691
고유명사	344	321	347	227	220	104	45	1,608
일반명사	1,784	1,877	2,332	2,789	3,057	1,803	219	13,861
의존명사	131	176	208	209	266	119	16	1,125
대명사	95	107	168	124	87	80	5	666
수사	100	79	39	51	133	48	6	456
동사	360	429	681	747	606	447	34	3,304
형용사	162	269	289	323	239	140	17	1,439
보조용언	43	45	127	140	108	62	7	532
지정사	20	34	29	43	55	25	4	210
관형사	44	74	92	128	133	92	11	574
일반부사	319	371	562	543	426	281	24	2,526
접속부사	90	85	100	93	93	37	6	504
감탄사	7	12	19	11	15	6	3	73
체인접두사	1		2	11	1	10		25

	1급	2급	3급	4급	5급	6급	6급 이상	합계
명사파생접미사	64	62	64	83	99	74	6	452
동사파생접미사	11	37	39	42	52	50	5	236
형용사파생접미사	4	11	15	7	13	17	1	68
어근	27	38	56	54	40	29	5	249
주격조사	180	247	356	452	466	264	16	1,981
관형격조사	21	27	27	56	86	44	4	265
목적격조사	303	300	469	452	470	404	26	2,424
부사격조사	376	472	494	467	443	231	19	2,502
호격조사	1	4	1			1		7
인용격조사	1	1	2		3	2		9
보조사	263	307	438	478	524	236	21	2,267
연결어미	206	367	684	594	501	337	22	2,711
종결어미	152	276	153	195	152	90	74	1,092
선어말어미	138	190	271	201	199	112	8	1,119
명사형 전성어미	1	13	26	27	17	12	2	98
관형사형 전성어미	143	244	404	485	472	234	25	2,007
접속조사	43	27	48	47	53	15	4	237
보격조사	4	4	9	20	10	2		49
구 단위 표현	3	7	9	8	6	4	0	37
표현 문형	45	120	171	247	254	139	2	978

## ㄴ. 언어권별

<표 24> 구어 오류 주석 결과\_오류 위치: 언어권별

	일본어	베트남어	중국어	영어	러시아어	기타	합계
표본 수	224	176	163	78	30	96	767
어절 수	97,752	75,508	74,971	36,748	12,980	38,876	336,835
주석 수 합계	14,454	12,153	9,754	3,142	1,489	4,707	45,699
고유명사	638	374	338	91	31	136	1,608

	일본어	베트남어	중국어	영어	러시아어	기타	합계
일반명사	4,841	3,414	2,909	797	434	1,466	13,861
의존명사	400	269	199	73	45	139	1,125
대명사	181	170	138	64	37	76	666
수사	120	115	100	30	23	68	456
동사	980	921	663	248	106	386	3,304
형용사	514	360	257	114	48	146	1,439
보조용언	176	173	83	41	24	35	532
지정사	66	43	50	19	3	29	210
관형사	150	137	154	41	26	66	574
일반부사	801	734	463	188	109	231	2,526
접속부사	153	118	93	50	20	70	504
감탄사	18	30	9	8	5	3	73
체인접두사	8	5	9			3	25
명사파생접미사	109	125	131	30	14	43	452
동사파생접미사	101	38	62	13	10	12	236
형용사파생접미사	35	8	18	3	1	3	68
어근	99	42	54	20	7	27	249
주격조사	396	505	593	170	67	250	1,981
관형격조사	50	68	87	21	5	34	265
목적격조사	754	790	452	170	43	215	2,424
부사격조사	618	702	554	238	101	289	2,502
호격조사		4	2			1	7
인용격조사	2	3	1	1		2	9
보조사	668	669	493	117	59	261	2,267
연결어미	1,044	620	578	164	94	211	2,711
종결어미	375	242	271	99	35	70	1,092
선어말어미	233	354	276	130	29	97	1,119
명사형 전성어미	21	21	26	10	5	15	98
관형사형 전성어미	674	598	415	101	47	172	2,007
접속조사	51	69	66	14	10	27	237



	일본어	베트남어	중국어	영어	러시아어	기타	합계
보격조사	8	21	12	1		7	49
구단위표현	4	6	7	11	4	5	37
표현 문형	166	405	191	65	47	112	986

## 다. 오류 층위

### ㄱ. 숙달도 단계별

<표 25> 구어 오류 주석 결과\_오류 층위: 숙달도 단계별

		1급	2급	3급	4급	5급	6급	6급 이상	합계
표본 수		140	123	166	170	100	63	5	767
어절 수		45,175	54,175	62,577	63,474	59,182	47,360	4,892	336,835
주석 수 합계		4,697	5,081	6,620	6,620	6,294	3,942	473	33,727
발음	음소	3,293	3,449	4,538	4,516	4,544	3,018	377	23,735
	음절	215	291	316	211	145	52	11	1,241
	음운규칙	349	321	510	531	452	274	20	2,457
	원어식 발음	139	107	125	107	118	40	14	650
	중간 발음 (변이음 포함)	339	180	333	471	166	115	12	1,616
형태	단어 형성 (합성법)	2	5	8	5	2	1		23
	단어 형성 (파생법)	32	48	56	71	64	32		303
	굴절(곡용)	57	79	75	86	120	39	25	481
	굴절(활용)	51	147	157	155	148	97	4	759
	품사	38	58	60	81	80	26		343
통사	높임	50	146	83	97	84	37	2	499
	시제	91	166	217	156	206	110	4	950
	사동	1	3	15	14	16	10		59
	피동		5	20	46	29	37		137

	부정	11	21	20	11	20	10		93
	어순	12	25	37	28	48	11		161
담화	지시	3	14	16	13	16	18		80
	접속	9	13	24	18	30	15	4	113
	담화표지	5	3	10	1	6			25
	구어/문어 오류				2				2

## ㄴ. 언어권별

<표 26> 구어 오류 주석 결과\_오류 층위: 언어권별

		일본어	베트남어	중국어	영어	러시아어	기타	합계
표본 수		224	176	163	78	30	96	767
어절 수		97,752	75,508	74,971	36,748	12,980	38,876	336,835
주석 수 합계		12,529	8,283	6,671	2,166	960	3,118	33,727
발음	음소	8,918	5,691	4,960	1,311	678	2,177	23,735
	음절	824	146	126	45	10	90	1,241
	음운규칙	739	849	375	153	77	264	2,457
	원어식 발음	332	41	63	111	27	76	650
	중간 발음 (변이음 포함)	807	489	70	101	36	113	1,616
형태	단어 형성 (합성법)	4	6	6	1	2	4	23
	단어 형성 (파생법)	41	107	67	33	10	45	303
	굴절(곡용)	142	80	168	23	6	62	481
	굴절(활용)	228	188	200	65	28	50	759
	품사	51	102	93	35	11	51	343
통사	높임	110	151	123	72	15	28	499
	시제	162	290	248	126	30	94	950
	사동	19	16	10	4	4	6	59
	피동	38	21	35	27	1	15	137

	부정	35	19	21	7	4	7	93
	어순	19	53	44	21	7	17	161
담화	지시	16	15	29	3	7	10	80
	접속	30	18	27	23	7	8	113
	담화표지	14		6	5			25
	구어/문어 오류		1					1

## 라. 오류 양상

### ㄱ. 숙달도 단계별

<표 27> 구어 오류 주석 결과\_오류 양상: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	합계
표본 수	140	123	166	170	100	63	5	767
어절 수	45,175	54,175	62,577	63,474	59,182	47,360	4,892	336,835
주석 수 합계	1,444	2,555	3,154	3,839	4,022	2,153	106	17,273
누락	358	557	712	825	893	421	15	3,781
첨가	121	218	290	384	330	243	13	1,599
대치	685	1,375	1,644	2,040	2,271	1,173	46	9,234
오형태	280	405	508	590	528	316	32	2,659

### ㄴ. 언어권별

<표 28> 구어 오류 주석 결과\_오류 양상: 언어권별

	일본어	베트남어	중국어	영어	러시아어	기타	합계
표본 수	224	176	163	78	30	96	767
어절 수	97,752	75,508	74,971	36,748	12,980	38,876	336,835
주석 수 합계	3,528	5,130	4,347	1,487	673	2,108	17,273
누락	587	1,431	897	360	98	408	3,781
첨가	375	399	389	129	100	207	1,599

대치	1,942	2,524	2,356	838	360	1,214	9,234
오형태	624	776	705	160	115	279	2,659

## 2.2. 학습자 말뭉치 활용 연구

○ 한국어 학습자 말뭉치 구축이 마무리되는 3단계 5차 연도, 6차 연도는 말뭉치의 구축과 정제 작업의 완결 외에도 6년간에 걸친 과제 수행 결과를 기반으로 학습자 말뭉치 구축과 가공, 활용에 관한 이론과 실제에 관한 논의를 집대성해야 하는 시기라고 할 수 있다. 이는 6개년 동안 수행해 온 사업 결과물을 종합하는 것뿐만 아니라 학습자 말뭉치의 구축에서 활용까지 전반적인 내용에 관한 실제적인 모델과 지침을 제시함으로써 학문적으로도 기여한다는 점에서 의미가 있다. 이에 따라 2019년부터 2020년까지 3단계 사업에서는 다음과 같은 성과물 산출을 목표로 삼았다.

- [이론서] 한국어 학습자 말뭉치 구축과 가공(이론서)
- [연구에의 적용] 한국어 학습자의 중간언어 발달 연구
  - 영역별/품사별 언어 사용 양상(참고 자료집)
- [교육에의 적용] 한국어 학습자 오류 사전(교수 현장에의 적용)

○ 그리고 2019년 사업에서는 이 중 중국어권 한국어 학습자의 중간언어 발달 양상에 관한 참고 자료집을 기획하여 연구를 수행하였다. 주요 내용은 다음과 같다.

- 제목: 중국어권 한국어 학습자의 중간언어 발달 연구
- 구성

- 1장 들어가며
  - 1. 학습자 말뭉치
  - 2. 학습자 말뭉치 기반 중간언어 발달 연구
- 2장 중국어권 학습자 말뭉치의 구성

1. 학습자 말뭉치의 구축
  - 1.1. 학습자 말뭉치의 설계와 구축
  - 1.2. 한국어 학습자 말뭉치의 구축 현황
    - 1.2.1. 한국어 학습자 말뭉치
    - 1.2.2. 외국어 교육에서의 학습자 말뭉치
2. 연구 말뭉치의 구성
  - 2.1. 말뭉치의 규모
  - 2.2. 말뭉치 수집 과제
  - 2.3. 말뭉치의 장르
  - 2.4. 말뭉치의 주제
  
- 3장 형태소별 사용 양상
  1. 학습자 말뭉치 형태 분석의 원칙과 실제
  2. 형태소별 사용 분포
  3. 키워드 분석을 통해 살펴본 주요 형태소의 사용 양상
    - 3.1. 키워드의 개념과 분석 방법
    - 3.2. 수준에 따른 키워드의 차이
  4. 어휘 범주별 사용 양상
    - 4.1. 체언
      - 4.1.1. 일반명사
      - 4.1.2. 의존명사
      - 4.1.3. 대명사
    - 4.2. 용언
      - 4.2.1. 동사
      - 4.2.2. 형용사
      - 4.2.3. 보조 용언
    - 4.3. 수식언 및 독립언
      - 4.3.1. 관형사
      - 4.3.2. 일반부사
      - 4.3.3. 접속부사
      - 4.3.4. 감탄사
    - 4.4. 관계언
      - 4.4.1. 격조사
      - 4.4.2. 보조사
    - 4.5. 어미

- 4.5.1. 선어말어미
- 4.5.2. 연결어미
- 4.5.3. 종결어미
- 4.5.4. 관형사형 전성어미
- 4.5.5. 명사형 전성어미
- 4.6. 접사
  - 4.6.1. 체언접두사
  - 4.6.2. 명사파생접미사
  - 4.6.3. 동사파생접미사
  - 4.6.4. 형용사파생접미사
- 4.7. 표현문형

참고문헌

## 2.3. 한국어 학습자 말뭉치의 균형성 확보를 위한 구축 방법 연구

- 균형성은 대표성과 함께 말뭉치의 기본 요건으로서 학습자 말뭉치의 완성도와 활용도를 높이기 위해 설계와 구축 과정에서 세심하게 고려되어야 할 부분이다. 학습자 말뭉치는 학습자의 제1 언어, 수준, 자료 유형(문어/구어), 장르 등의 변인이 매우 다양하기 때문에 이러한 특성을 고려하여 대표성 있는 자료들이 고르게 분포되도록 해야 한다. 4차 연도까지의 연구에서 이를 위한 방안으로서 학습자의 분포를 고려한 현실적인 균형성 확보와 특수 말뭉치(sub corpus) 구축안을 제시한 바 있다. 본 연구에서는 이러한 방향성에 맞춰 실제 자료를 구축해 나아가되 국외 학습자 말뭉치 자료를 수집하면서 그 타당성을 검증하고 필요 시 이전의 계획을 조정하고자 하였다.
- 학습자 말뭉치의 균형성 확보를 위한 구축 방법 연구는 3-4차 연도의 과업을 통해 기초적인 논의가 이루어진 바 있다. 3차 연도 연구에서는 현황 조사를 통해 첫째, 국내외 한국어 학습자의 분포에서 국내 유학생이 35만 명으로 가장 많으며 둘째, 국적 분포에서 국내 거주 외국인의 경우 중국,

베트남, 대만, 일본 등 아시아권 국적 소지자의 비중이 높고(언어교육원의 경우 중국, 일본, 기타 아시아 국가의 학습자 비중이 대체적으로 높고 그 중 중국이 40% 이상임), 셋째, 언어교육원의 숙달도 분포 1급(23.2%), 2급(24.1%), 3급(19.0%), 4급(15.6%), 5급(10.7%), 6급(5.9%), 최고급 이상(1.4%)로 초급이 50%에 달하며 고급 단계로 갈수록 수가 급격하게 감소함을 살펴보았다. 그리고 그 결과를 바탕으로 한국어 학습자 말뭉치의 균형성 확보를 위한 방안으로 다음과 같은 안을 제안하였다.

- 실제적인 학습자 분포를 고려한 균형성 확보
  - 선 구축, (귀납적 수집 결과에 따른) 후 균형
  - 서브 말뭉치 구축으로 다양성 확보
  - 다양한 수집 과제를 통한 균형성과 다양성 확보
- 그리고 실제적인 구축 방법론으로서 ‘특수 말뭉치(sub corpus)’ 구축을 핵심적인 과제로 삼게 되었다. 이에 따라 4차 연도에는 2020년까지의 사업이 완료된 후 구축될 말뭉치의 최종 규모가 약 410여절이라는 사실에 주목하고 전체 구축 말뭉치의 학습자 변인, 자료 변인에 따른 분포를 참고하여 학습자 말뭉치의 특성을 고려한 현실적인 균형을 맞추기 위해 구축 비율을 조정하였다. 이는 지난 4년간 수렴한 자문위원들의 의견을 반영한 결과이기도 하다.
- 2019년에는 이러한 방향성을 그대로 수용하여 구축 작업을 수행하되 이 방향성이 타당한지를 검증하고 국외에서 학습자 말뭉치의 균형성을 확보하기 위해 어떠한 방법을 강구하였는가를 조사하기 시작하였다. 이 작업은 2020년 상반기까지 이어질 예정이다. 다음은 현재 사례 분석 대상이 되는 국외 학습자 말뭉치 목록이며 분석 결과는 2020년까지 구축 말뭉치의 분석을 완료한 후에 종합하여 제시할 예정이다.

- The Arabic Learner Corpus(ALC)
- The Pilot Arabic Learner Corpus
- The AKCES/CZESL corpus (Akvizièni korpusy èetiny - Acquisition corpora of the Czech language/Czech as a second language)
- Leerdercorpus Nederlands als Vreemde Taal
- The Advanced Learner English Corpus(ALEC)

- The ANGLISH corpus
- Asao Kojiro's Learner Corpus Data
- The Barcelona English Language Corpus (BELC)
- The BATMAT Corpus
- The Bilingual Corpus of Chinese English Learners (BICCEL)
- The Br-ICLE corpus (Brazilian component of ICLE)
- The British Academic Written English (BAWE) corpus
- The BUiD Arab Learner Corpus (BALC)
- The Cambridge Learner Corpus (CLC)
- The Corpus of Academic Learner English (CALE)
- The Corpus of English Essays Written by Asian University Students (CEEAAUS)
- The Chinese Academic Written English (CAWE) corpus
- The Chinese Learner English Corpus (CLEC)
- The City University Corpus of Academic Spoken English (CUCASE)
- The Cologne-Hanover Advanced Learner Corpus (CHALC)
- College Learners' Spoken English Corpus (COLSEC)
- The Corpus Archive of Learner English in Sabah/Sarawak (CALES)
- The Corpus of Young Learner Interlanguage (CYLIL)
- The Eastern European English learner corpus
- The EFLCorpus (ETC)
- The English of Malaysian School Students corpus (EMAS)
- The English Speech Corpus of Chinese Learners (ESCCL)
- The ETS Corpus of Non-Native Written English
- The EVA Corpus of Norwegian School English
- The Gachon Learner Corpus
- The GICLE corpus (German component of ICLE)
- The Giessen-Long Beach Chaplin Corpus (GLBCC)
- The Hong Kong University of Science & Technology (HKUST) learner corpus
- The Indianapolis Business Learner Corpus (IBLC)
- The International Corpus of Crosslinguistic Interlanguage (ICCI)
- The International Corpus Network of Asian Learners of English



(ICNALE)

- The International Corpus of Learner English (ICLE)
- The International Teaching Assistants corpus (ITAcorp)
- The ISLE speech corpus
- The Israeli Learner Corpus of Written English
- The Japanese English as a Foreign Language Learner (JEFLL) Corpus
- The Janus Pannonius University (JPU) Corpus
- Lancaster Corpus of Academic Written English (LANCAWE)
- The Lang-8 Learner Corpora
- The LeaP Corpus :Learning Prosody in a Foreign Language
- The Learner Corpus of Engineering Abstracts(LCEA)
- The Learner Corpus of English for Business Communication
- The Learner Corpus of Essays and Reports
- A Learners' Corpus of Reading Texts
- The LONGDALE project: LONGitudinal DAtabase of Learner English
- The Longman Learners' Corpus
- The Louvain International Database of Spoken English Interlanguage (LINDSEI)
- The Malaysian Corpus of Learner English (MACLE)
- The Malaysian Corpus of Students' Argumentative Writing (MCSAW)
- The Michigan Corpus of Academic Spoken English (MICASE)
- The Michigan Corpus of Upper-level Student Papers (MICUSP)
- The Montclair Electronic Language Database (MELD)
- The Multimedia Adult ESL Learner Corpus (MAELC)
- The Neungyule Interlanguage Corpus of Korean Learners of English (NICKLE)
- The NICT JLE (Japanese Learner English) Corpus
- The NON-native Spanish corpus of English (NOSE)
- The NUS Corpus of Learner English
- The PELCRA Learner English Corpus (PLEC)
- The PICLE corpus (Polish component of ICLE)
- The Qatar learner corpus
- The Québec learner corpus

- The Romanian Corpus of Learner English (RoCLE)
- The Russian Learner Translator Corpus(RusLTC)
- The Santiago University Learner of English Corpus (SULEC)
- The Scientext English Learner Corpus
- Second Language Research Tasks(SLRT)
- The Seoul National University Korean-speaking English Learner Corpus (SKELC)
- The SILS Learner Corpus of English
- The Soochow Colber Student Corpus (SCSC)
- The Spoken and Written English Corpus of Chinese Learners (SWECCCL)
- The Taiwanese Corpus of Learner English (TLCE)
- The Tawainese learner academic writing corpus (TaiwanLAWC)
- The TELEC Secondary Learner Corpus (TSLC)
- The Telecollaborative Learner Corpus of English and German Telekorp
- The Tswana Learner English Corpus (TLEC)
- The Uppsala Student English Corpus (USE)
- The UPF Learner Translation Corpus
- The UPV Learner Corpus
- The Varieties of English for Specific Purposes dAtabase (VESPA) learner corpus
- The WriCLE (Written Corpus of Learner English) corpus
- The Yonsei English Learner Corpus (YELC)
- The Young Learner Corpus of English(YOLECORE)
- The Estonian Interlanguage Corpus (EIC) of Tallinn University
- Linguistic Basis of the Common European Framework for L2 English and L2 Finnish(CEFLING)
- Paths in Second Language Acquisition(TOPLING)
- The Advanced Finnish Learner Corpus(LAS2)
- The Finnish National Foreign Language Certificate Corpus (YKI)
- The International Corpus of Learner Finnish(ICLFI)
- The Chy-FLE (Cypriot Learner Corpus of French)
- The COREIL corpus

- The "Dire Autrement" corpus
- French Interlanguage Database (FRIDA)
- French Learner Language Oral Corpora (FLLOC)
- The InterFra corpus
- The "Interphonologie du Français Contemporain" (IPFC) corpus
- The LCF corpus (Learner Corpus French)
- The Lund CEFLE Corpus (Corpus Écrit de Français Langue Étrangère)
- The UWi (University of the West Indies) learner corpus
- Comasan Labhairt ann an Gàidhlig (CLAG)
- Gaelic Adult Proficiency
- The AleSKO corpus
- Analyzing Discourse Strategies: A Computer Learner Corpus
- The Corpus of Learner German (CLEG13)
- The FALKO corpus (Fehlerannotiertes Lernerkorpus 'error annotated learner corpus')
- The KOLIPSI corpus
- The LeaP Corpus (Learning the Prosody of a foreign language)
- The LeKo (Lernerkorpus) corpus
- The LINCOS Corpus
- MERLIN(Multilingual Platform for the European Reference Levels: Exploring Interlanguage in Context)
- The Telecollaborative Learner Corpus of English and German Telekorp
- The Langman corpus
- Corpus parlato di italiano L2
- The KOLIPSI corpus
- The LIPS Corpus (Lexicon of Sopoken Italian by Foreigners)
- Varietà di Apprendimento della Lingua Italiana: Corpus Online (VALICO)
- The Korean learner corpus
- The ASK (Andrespråkskorpus = Second Language Corpus) corpus
- The PIKUST pilot learner corpus
- The Anglia Polytechnic University (APU) Learner Spanish Corpus
- Aprescillov ("Aprendera Escribiren Lovaina")

- CEDEL2 (Corpus Escrito del Español L2)
- The Corpus of Taiwanese Learners of Spanish (Corpus de Apprentices Taiwanaises de Español, CATE)
- The DIAZ corpus
- The Japanese learner corpus of Spanish
- Spanish Learner Language Oral Corpus (SPLLOC)
- Spanish Learner Oral Corpus
- The ASU corpus
- The ESF(European Science Foundation Second Language) Database
- The Foreign Language Examination Corpus (FLEC)
- The MeLLANGE Learner Translator Corpus (LTC)
- The MiLC Corpus
- The Multilingual Learner Corpus (MLC)
- The Padova Learner Corpus
- The PAROLE corpus(corpus PARallèle Oral en Langue Etrangère)
- The University of Toronto Romance Phonetics Database (RPD)

### Ⅲ. 학습자 말뭉치 교육 및 홍보

#### 1. 말뭉치 구축/가공 인력 실무 교육

- 말뭉치 구축/가공 인력 실무 교육은 실무 작업자에게 말뭉치 구축에 관한 기본 소양과 기술을 익히도록 하고 각 구축 단계별 작업자로서의 전문성을 제고하여 체계적인 말뭉치를 구축해 나가기 위한 것이다. 이 교육은 또한 자료 처리 방식의 일관성 확보를 위해서도 필요하다. 다양한 변이형을 포함한 학습자 말뭉치의 특성상 작업자의 직관에 의해 세부적인 자료 처리 방식이 달라질 우려가 있기 때문이다. 이와 같은 필요성으로 본 사업에서는 온라인/오프라인, 정기/비정기 워크숍을 통해 실무자 간에 원활한 소통이 이루어지도록 하였다.

##### 1.1. 교육 대상

- 한국어 학습자 말뭉치 수집, 구축, 가공 실무 작업자

##### 1.2. 교육 방법

- 수집, 입력, 전사, 형태 주석, 오류 주석 팀별 정기 워크숍(주 1회)  
: 작업 관련 쟁점에 대한 토론 및 지침 교육
- 온라인 카페, 채팅 프로그램을 활용한 수시 질의응답  
: 작업 과정에서 발생하는 문제점이나 궁금증을 실시간으로 해결
- 학습자 말뭉치 구축 시스템을 활용한 피드백 제공  
: 미해결 주석 항목을 검토 요청 항목으로 남겨 전체 회의를 통해 해결하고 지침에 반영함

### 1.3. 교육 내용

- 교육 내용은 크게 지침 교육과 도구 사용 교육/실습으로 나뉜다.

<표 29> 말뭉치 구축/가공 인력 교육 내용

	지침 교육	도구 교육 및 실습
수집	<ul style="list-style-type: none"> <li>○ 자료 수집 과제 유형 및 수집 방법</li> <li>○ 학습자 동의서 수집 및 처리</li> <li>○ 수집 자료의 처리와 관리</li> </ul>	<ul style="list-style-type: none"> <li>○ 온라인 구축 도구: 수집 표본 등록 및 표본 관리</li> </ul>
자료 처리 및 파일 등록	<ul style="list-style-type: none"> <li>○ 자료의 분류</li> <li>○ 스캔 및 음성 파일 변환</li> <li>○ 파일명 부여 체계</li> <li>○ 학습자 정보 및 파일 정보 등록(헤더 마크업)</li> </ul>	<ul style="list-style-type: none"> <li>○ 온라인 구축 도구: 스캔/음성 원본 파일 업로드 및 파일 등록, 파일명 생성</li> <li>○ 스캐너 사용</li> <li>○ 음성 파일 변환</li> </ul>
입력	<ul style="list-style-type: none"> <li>○ 문어 입력 및 검수 방법, 쟁점</li> </ul>	<ul style="list-style-type: none"> <li>○ 온라인 구축 도구: 파일 입력 및 마크업, 할당 받은 작업 파일 관리 및 작업</li> </ul>
전사	<ul style="list-style-type: none"> <li>○ 구어 전사 및 검수 방법, 쟁점</li> </ul>	<ul style="list-style-type: none"> <li>○ 온라인 구축 도구: 도구 내 전사 및 마크업, 할당 받은 작업 파일 관리 및 작업</li> </ul>
형태 주석	<ul style="list-style-type: none"> <li>○ 형태 분석 방법 및 절차</li> <li>○ 형태 주석 체계</li> <li>○ 형태 분석 자료 검수 및 검수 관련 쟁점</li> </ul>	<ul style="list-style-type: none"> <li>○ 온라인 구축 도구: 형태 주석, 할당 받은 작업 파일 관리 및 작업</li> </ul>
오류 주석	<ul style="list-style-type: none"> <li>○ 오류 식별, 판정 및 교정의 기준</li> <li>○ 오류 주석 체계</li> <li>○ 오류 분석 자료 검수 및 검수 관련 쟁점</li> </ul>	<ul style="list-style-type: none"> <li>○ 온라인 구축 도구: 오류 주석, 할당 받은 작업 파일 관리 및 작업</li> </ul>

## 1.4 참여 인력

### (1) 자료 스캔 및 처리

김지영(연세대학교 국어국문학과 석사과정)

### (2) 문어 입력 작업 및 검수

김동은(연세대학교 국어국문학과 박사과정/문어 작업 관리 및 검수)

박현진(연세대학교 국어국문학과 석사과정/문어 검수)

류지현(연세대학교 국어국문학과 석사과정/문어 검수)

유근선(연세대학교 국어국문학과 박사과정/문어 검수)

이주아(연세대학교 국어국문학과 석사과정/문어 검수)

기달림(연세대학교 국어국문학과 학사과정/문어 입력)

김지원(연세대학교 국어국문학과 학사과정/문어 입력)

박재익(연세대학교 국어국문학과 박사과정/문어 입력)

손영랑(연세대학교 국어국문학과 학사과정/문어 입력)

신예진(연세대학교 국어국문학과 학사과정/문어 입력)

여은정(서울대학교 언어학과 석사과정/문어 입력)

최단비(연세대학교 국어국문학과 학사과정/문어 입력)

### (3) 구어 전사 작업 및 검수

김미선(연세대학교 국어국문학과 박사과정/구어 작업 관리 및 검수)

윤현애(연세대학교 국어국문학과 박사과정, 구어 검수)

김정현(연세대학교 국어국문학과 박사과정, 구어 검수)

주향아(연세대학교 국어국문학과 박사수료, 구어 검수)

김지영(연세대학교 국어국문학과 석사과정, 구어 전사 작업)

양현지(연세대학교 국어국문학과 석사과정, 구어 전사 작업)

오세원(연세대학교 국어국문학과 석사과정, 구어 전사 작업)

이미영(연세대학교 국어국문학과 석사과정, 구어 전사 작업)

전화연(연세대학교 언어정보학협동과정 석사과정, 구어 전사 작업)

#### (4) 형태 주석 작업 및 검수

배미연(연세대학교 국어국문학과 박사과정/형태 주석 작업 관리 및 작업·검수)  
김미경(연세대학교 국어국문학과 박사과정/형태 주석 검수)  
김수경(연세대학교 국어국문학과 박사과정/형태 주석 검수)  
이보미(연세대학교 국어국문학과 박사과정/형태 주석 검수)  
임상은(연세대학교 국어국문학과 박사과정/형태 주석 검수)  
강예지(연세대학교 언어정보학협동과정 석사과정/형태 주석 작업)  
김다희(연세대학교 국어국문학과 석사과정/형태 주석 작업)  
김하늘(연세대학교 국어국문학과 석사졸업/형태 주석 작업)  
박은현(연세대학교 언어정보학협동과정 박사과정/형태 주석 작업)  
서세령(연세대학교 한국학협동과정 석사과정/형태 주석 작업)  
서지혜(연세대학교 국어국문학과 박사과정/형태 주석 작업)  
성미향(연세대학교 국어국문학과 박사과정/형태 주석 작업)

#### (5) 오류 주석 작업 및 검수

허희정(연세대학교 국어국문학과 박사과정/오류 주석 작업 관리 및 작업·검수)  
공나형(연세대학교 국어국문학과 박사과정/오류 주석 작업 및 검수)  
김선영(연세대학교 한국학협동과정 박사과정/오류 주석 작업 및 검수)  
송지혜(연세대학교 국어국문학과 박사과정/오류 주석 작업 및 검수)  
유소영(연세대학교 국어국문학과 박사과정/오류 주석 작업 및 검수)



## 2. 한국어 학습자 말뭉치 아카데미 개최

- 한국어 학습자 말뭉치 아카데미는 학습자 말뭉치의 폭넓은 활용을 유도하고 학습자 말뭉치의 손쉬운 활용 방법 교육을 목표로 개최한다. 한국어 학습자 말뭉치 아카데미는 2015년부터 2018년까지 총 9회에 걸쳐 개최했다. 개최 때마다 많은 사용자들의 관심과 호응을 얻어 온 아카데미는 2019년에도 계속되었다.
- 2019년 사업에서는 다음과 같이 총 3회의 말뭉치 아카데미를 개최하였다. 2019년 아카데미는 학습자 말뭉치 나눔터를 통해 제공되는 검색 기능을 넘어서 사용자 스스로 학습자 말뭉치 나눔터에서 내려받은 말뭉치를 활용 목적에 맞게 가공하여 분석할 수 있도록 심화된 교육 프로그램을 제공함으로써 무르익어 가는 학습자 말뭉치에 대한 관심을 더욱 고조시킬 수 있었다<sup>8)</sup>. 아울러 교육 내용의 측면에서 질적으로 한 단계 더 나아갈 수 있었다.

<표 30> 학습자 말뭉치 활용 아카데미 개최

일시 및 장소	프로그램	대상
7월 9일(화) 연세대학교	○ 한국어 학습자 말뭉치 소개 ○ 한국어 학습자 말뭉치의 정량적 분석: 빈도, N-그램, 연어, 핵심어성(keywordness)를 중심으로	한국어교육 연구자 및 대학원생
8월 21일(수) 건국대학교 (언어학회 여름학교 연계)	○ 한국어 학습자 말뭉치 소개 ○ 한국어 학습자 말뭉치의 정량적 분석: 빈도, N-그램, 연어, 핵심어성(keywordness)를 중심으로	한국어교육 연구자 및 대학원생
11월 8일(금)	○ 한국어 학습자 말뭉치 소개	한국어교육

8) 당초 계획에서는 서울 1회, 지방 1회 개최 예정이었으나 이미 앞선 말뭉치 아카데미에서 주요 거점 지역의 대학에서 아카데미를 개최한 바 있으며 파급력이나 참가자의 편의성 측면에서 서울이 더 효율적이라는 점을 고려하여 수요 기관과의 합의하에 모두 서울에서 개최하였다.

고려대학교	○ 한국어 학습자 말뭉치의 활용 - 학습자 말뭉치를 대상으로 한 공기어 분석 과 네트워크 시각화	연구자 및 대학원생
-------	---	---------------

### 3. 학술대회 발표

- 학술대회 발표는 학습자 말뭉치 구축 방법론 및 쟁점에 관한 학술적 교류와 향후 학습자 말뭉치 구축 결과물을 널리 확산시키는 것을 목표로 한다. 이론적으로는 학습자 말뭉치 구축에 관한 이론적 논의가 부족한 상황에서 구축 방법론에 대한 이론적 토대를 제시하고 학계의 공통적인 관심사로 이끌어 낼 수 있다는 점에서 의의가 있다. 또한 활용의 측면에서 말뭉치 활용의 실제적인 방법론과 모델을 보임으로써 사용자의 범위를 확대할 수 있다는 점에서 가치가 있다. 2019년에는 다음의 발표를 통해 학습자 말뭉치 구축에 관한 쟁점과 활용에 관한 논의를 불러일으켰다.
- 한송화(2019), 한국어 학습자의 한국어 부사 사용 양상에 대한 연구, 한국언어문화교육학회 학술대회.
- 유소영 · 강현화(2019), 한국어 학습자 말뭉치의 오류 주석의 쟁점과 실제 언어사실과 관점 48, 연세대학교 언어정보연구원.

## IV. 학습자 말뭉치 구축 및 가공

### 1. 수집

#### 1.1. 구축 목표 및 수집 대상

- 3단계 5차 연도와 6차 연도는 국외 자료를 집중적으로 구축하며, 전체 말뭉치의 균형성 확보를 위하여 국내 자료와 이주민 자료를 추가 구축하는 것을 목표로 한다.
- 집중 구축: 국외 학습자 자료
- 균형성 확보를 위한 추가 구축: 국내 대학 교육 기관의 정규 교육과정 수강 학습자, 학문 목적 학습자, 결혼이주민, 이주 노동자, 중도입국청소년

#### 1.2. 수집 네트워크

##### (1) 수집 네트워크 확장

- 2019년도는 국외 자료 집중 구축의 해로 효율적인 자료 수집을 위해 2018년의 시범 수집 결과를 바탕으로 구체적인 수집 계획을 수립하고 이를 수행해 나가기 위한 자료 수집 네트워크를 마련하였다. 이로써 한국어 학습자 말뭉치는 다소 미비하나마 국내와 국외의 자료를 두루 수집할 수 있는 종합적인 네트워크를 갖추게 되었다.
- 2017년까지 한국어 학습자 말뭉치 수집 네트워크는 국내 대학 부설 한국어 교육 기관과 이주민 교육 기관, 종적 자료 수집을 위한 교사 인력풀로 구성되어 있었다. 그러나 2018년에는 국내 교육 기관을 대상으로 한 대규모 수집을 종료하고 그간 구축된 자료의 균형성 확보를 위한 기획 자료 구축과 이주민 자료의 집중 수집, 국외 자료의 시험 수집을 목표로 하였다. 2019년에는 2018년까지의 수집 네트워크를 기반으로 하되, 자료 구축 목표와 그간 수집을 하면서 발생한 쟁점들을 해소하면서 국외 수집 네트워크를 확장하는 방향으로 네트워크를 개편하였다.

<표 31> 수집 네트워크 현황

구분	수집 참여 기관 및 인력풀	비고
국내 학습자	<ul style="list-style-type: none"> <li>○ 기관: 1, 2단계 사업에서 수집에 참여하였던 18개 기관에 특정 언어권 학습자 자료 수집을 위한 자율적 참여 독려</li> <li>○ 교사 인력풀: 종적 자료 및 기획 자료 수집 교사</li> </ul>	<p>기획 자료 수집 종료 후 기획 자료</p>
이주민	<ul style="list-style-type: none"> <li>○ 기관: 2단계 사업에서 수집에 참여하였던 12개 기관을 대상으로 한 수집 참여 독려</li> <li>○ 교사 인력풀: 종적 자료 및 기획 자료 수집 교사 10명</li> </ul>	<p>수집을 위한 교사 인력풀 보강</p>
국외 학습자	<ul style="list-style-type: none"> <li>○ 대학: 6개국 12개 대학 <ul style="list-style-type: none"> <li>- 대만 문화대</li> <li>- 대만 정치대</li> <li>- 싱가포르 난양공대</li> <li>- 미국 에모리대</li> <li>- 미국 클레어몬트대</li> <li>- 태국 부라파대</li> <li>- 태국 치앙마이대</li> <li>- 태국 마하사라캄대</li> <li>- 중국 남경대</li> <li>- 중국 화중사범대</li> <li>- 중국 길림외국어대</li> <li>- 인도네시아 가자마다대</li> </ul> </li> <li>○ 세종학당: 14개국 21개 세종학당 <ul style="list-style-type: none"> <li>- 중국 선양 세종학당</li> <li>- 중국 린이 세종학당</li> <li>- 중국 칭다오2 세종학당</li> <li>- 중국 윈저우 세종학당</li> <li>- 중국 시안 세종학당</li> <li>- 몽골 울란바토르 세종학당</li> <li>- 주베트남한국문화원 세종학당</li> <li>- 베트남 다낭 세종학당</li> <li>- 베트남 하노이3 세종학당</li> <li>- 스페인 라스팔마스 세종학당</li> </ul> </li> </ul>	<p>2018년 시험 수집 기관과 2019년 신규 참여 기관이 포함된 것으로 수집 참여 의사 밝힌 후 본 보고서 작성 시점까지 자료 수집이 미완료 상태인 기관이 있으며, 세종학당의 경우 세종학당재 단의</p>

	<ul style="list-style-type: none"> <li>- 포르투갈 리스본 세종학당</li> <li>- 이탈리아 세종학당</li> <li>- 바레인 마나마 세종학당</li> <li>- 태국 마하사라캄 세종학당</li> <li>- 우루과이 몬테비데오 세종학당</li> <li>- 키르기스스탄 비슈케크 세종학당</li> <li>- 주독일 한국 문화원</li> <li>- 아랍에미리트 한국문화원 세종학당</li> <li>- 아부다비 세종학당</li> <li>- 인도네시아 거점 세종학당</li> <li>- 주카자흐스탄 한국문화원 세종학당</li> </ul>	공문을 통해 수집 홍보를 하였으며 수집 교사의 자율적 참여 의사에 따라 수집 기관이 확정됨.
--	--	--

## (2) 수집 경로

○ 수집 교사에 의해 1차 수집이 완료된 자료는 수집 지역이 국내에서 국외로 확대됨에 따라 다음의 세 가지 방식을 통해 수합하였다.

- 우편 또는 택배: 국내 자료(원본 자료)
- 전자 메일: 국내/국외 자료(스캔/음성 녹음 파일)
- 웹하드: 국외 자료(스캔/음성 녹음 파일)

## (3) 수집 과제

○ 수집 과제는 수집 대상에 따라 다음과 같이 차별화하였다.<sup>9)</sup>

<표 32> 수집 대상에 따른 과제 유형

과제 유형	세부 과제 유형	수집 대상
기획 자료	문어/구어 통합 수집 과제 (2018년 기획 자료)	국내 학습자 (이주민 포함) 국외 학습자

9) 세부 수집 과제는 수집 및 자료 처리 지침 참고

	구어 자료 수집 과제 (2019년 신규 기획)	국내 학습자 (이주민 포함)
	종적 말뭉치	국내 학습자 (이주민 포함)
교육과정 자료	성취도 평가, 수업 활동, 수업 과제 자료	국외 학습자

### 1.3. 자료 수집 현황

#### (1) 국외 자료

- 국외 자료의 수집 현황은 다음과 같다. 제시된 현황 자료는 현재 구축 팀에 자료가 전송되어 1차 정리가 된 자료들이며 그 외 기관의 자료는 지속적으로 수합하고 있다.

<표 33> 국외 자료 수집 현황

수집 기관		구어	문어
세종학당	리스본 세종학당	13개	57개
	몬테비데오 세종학당	77개	77개
	선양 세종학당	58개	58개
	라스팔마스 세종학당	12개	28개
	시안 세종학당	15개	15개
	아부다비 세종학당	14개	14개
	울란바토르 세종학당	50개	50개
	이탈리아 세종학당	20개	17개
	중국 칭다오2 세종학당	34개	34개
	카자흐스탄 세종학당	52개	36개
	키르기스스탄 비슈케크 세종학당	17개	17개
	몬테비데오 세종학당	15개	15개

	인도네시아 거점 세종학당	16개	16개
대학	싱가포르 난양공대	4개	72개
	미국 예모리대	46개	46개
	태국 부라파대	76개	76개
	태국 치앙마이대	15개	205개
	중국 화중사범대	-	179개
합계		534개	1,012개

☞ 국외 자료의 경우 수집 참여 기관의 수는 많으나 IRB 문제나 교육과정 운영상의 문제, 교육 기관 내부의 사정 등으로 인해 기관 내 참여율이 높지 않은 편이며 수집 계획 수립 시 추정하는 자료 수, 실제 수집 후 수합되는 자료의 수에 차이가 많아 수집에 어려움이 따른다. 또한 수집되어 오는 자료들 중에는 발화 산출량이 자료 구축이 어려울 만큼 적거나 스캔 등의 자료 처리 과정에서 자료가 손상이 되어 구축 대상에서 제외되는 자료의 비중이 높다.

☞ 이에 따라 2020년 사업에서는 현재의 수집 네트워크를 계속 유지하되, 국가별 한국어교육자 협의회나 주요 기관과의 협약을 통해 자료 수집에 대한 협조를 이끌어 내고, 말하기 대회 또는 글쓰기 대회 등을 유치하여 기획 과제 외의 자료를 수집하는 안을 마련할 계획이다.

## (2) 국내 자료

### ① 종적 자료

○ 종적 자료는 2018년 수집이 시작되어 2019년 계속 수집 대상이 된 7명을 대상으로 수집이 진행되었다. 7명 중 3명은 60주차까지 수집 완료하였으며, 2명은 48주차, 2명은 52주차까지 수집하고 중도에 수집을 철회하였다. 이렇게 해서 2015년부터 시작된 종적 자료의 수집이 모두 완료되었으며, 2주차부터 60주차까지의 누적 자료 수가 2,358세트(작문 1편, 말하기 음성 녹음 1개/1세트) 수집되었다. 수집에 참여한 학습자의 수는 191명으로 다음은 주차별 누적 자료 수이다.<sup>10)</sup>

<표 34> 종적 자료 수집 현황: 주차별 누적 자료 수

주차	2	4	6	8	10
자료 수	178	165	157	147	144
주차	12	14	16	18	20
자료 수	131	126	121	117	114
주차	22	24	26	28	30
자료 수	103	96	92	88	89
주차	32	34	36	38	40
자료 수	73	66	61	51	53
주차	42	44	46	48	50
자료 수	25	24	24	23	21
주차	52	54	56	58	60
자료 수	14	14	13	14	15

○ 다음은 종적 자료의 학습자 유형과 국적별 분포를 나타낸 것이다.

<표 35> 종적 자료 수집 현황: 학습자 유형 및 국적별 분포

	국내 기관	결혼이주	이주노동	이주노동자	중도입국	합계
남수단	1					1
네팔			6			6
러시아		4		1	5	10
르완다	1					1
말레이시아	3					3

10) 본 사업 당초 계획에서는 종적 자료 수집 대상자를 신규로 모집하여 추가 수집을 할 계획이었으나 2015-2019년 수집 현황에서 보는 바와 같이 주차별로 학습자의 수가 급격하게 감소하며 사실상 60주차까지 수집을 완료하는 비율이 매우 낮다. 이에 따라 더 이상 수집 대상자를 추가로 모집하지 않기로 하였다. 아울러 2주차의 자료 수가 178명으로 총 참여 학습자 수보다 적은데, 이는 교육과정상의 문제로 일부 학습자의 자료가 중도에 수집이 개시되어 생긴 차이이다.



몽골	4					4
미얀마		1				1
방글라데시	1		1			2
베트남	45	11	2			58
사우디	1					1
사우디아라비아	1					1
스리랑카	3					3
스웨덴	1					1
싱가포르	1					1
영국	1					1
우즈베키스탄	1	1	1			3
우즈벡	1	5			2	8
이탈리아	1					1
인도네시아	2	1	1			4
일본	1					1
중국	38	6			2	46
카메룬	1					1
카자흐스탄	1		1		2	5
캄보디아	1	3	3	1		7
콜롬비아	1					1
키르기즈스탄	1					1
태국	2	4				6
투르크메니스탄	1					1
튀니지	1					1
프랑스	1					1
필리핀		6				6
한국					2	2
호주	1					1
홍콩	1					1
합계	118	42	15	2	13	191

② 국내 기획 과제 자료

- 국내 기획 자료는 이주민 학습자와 국내 교육 기관 학습자를 대상으로 1) 문어/구어 통합 과제와 2) 구어 자료 수집 과제를 사용하여 수집하였다.

- 문어/구어 통합 과제: 이주민 학습자 250명 자료 수집

- 구어 수집 과제: 국내 대학 교육기관 학습자 450명 자료 수집

☞ 구어 수집 과제의 경우는 1단계 국내 교육기관 집중 구축 시 수집에 참여했던 기관 중 각 지역별 거점 기관과 연계하여 수집을 확대하고 있다.

## 2. 구축 및 가공

### 2.1. 원시 말뭉치

#### (1) 문어

- 2019년 문어 원시 말뭉치는 795,925어절이 구축되었다. 그 결과 1-4차 연도에 기구축한 말뭉치와 합산하여 누적 합계 2,819,363어절 규모의 문어 원시 말뭉치가 구축되었다.

#### ① 속달도별 자료 분포

- 문어 원시 말뭉치는 1급에서 6급, 그리고 최고급인 6급 이상의 자료가 구축되었다. 다음은 각 속달도별 구축 규모를 집계한 것이다.

<표 36> 문어 원시 말뭉치의 속달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	어절 수	278,595	316,860	337,531	338,801	333,584	278,457	139,610	2,023,438
	파일 수	4,013	3,165	2,676	2,429	2,011	1,553	136	15,983
2018	어절 수	76,368	147,461	212,429	167,893	120,022	59,902	11,850	795,925
	파일 수	1,331	1,787	2,133	1,674	951	481	17	8,374

합계	어절 수	354,963	464,321	549,960	506,694	453,606	338,359	151,460	2,819,363
	파일 수	5,344	4,952	4,809	4,103	2,962	2,034	153	24,357

② 국적별 자료 분포

- 문어 원시 말뭉치는 이중 국적 학습자를 포함하여 총 142개국의 자료가 구축되었다.

<표 37> 문어 원시 말뭉치의 국적별 자료 분포

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	876,128	6,493	284,112	3,369	1,160,240	9,862
일본	277,484	2,246	114,446	919	391,930	3,165
베트남	159,506	1,452	25,830	236	185,336	1,688
홍콩	52,094	553	61,020	209	113,114	762
대만	81,651	383	26,626	707	108,277	1,090
미국	63,436	392	22,487	179	85,923	571
말레이시아	54,196	428	19,448	160	73,644	588
몽골	49,051	376	16,439	180	65,490	556
태국	45,234	532	18,311	319	63,545	851
러시아	41,449	377	15,850	150	57,299	527
인도네시아	31,535	229	11,398	50	42,933	279
카자흐스탄	26,908	71	9,052	208	35,960	279
프랑스	20,772	116	13,229	138	34,001	254
싱가포르	16,773	125	12,701	27	29,474	152
우즈베키스탄	22,079	253	5,837	109	27,916	362
스웨덴	5,328	68	18,601	57	23,929	125
캐나다	15,889	90	3,493	13	19,382	103
이탈리아	13,143	63	4,220	41	17,363	104
스리랑카	15,066	174	1,926	113	16,992	287
한국	10,599	101	6,009	23	16,608	124

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
호주	11,887	64	3,748	37	15,635	101
키르기스스탄	8,503	90	4,895	56	13,398	146
영국	10,544	67	2,272	53	12,816	120
필리핀	9,376	209	2,386	81	11,762	290
독일	7,012	51	4,458	17	11,470	68
스페인	5,626	99	4,684	37	10,310	136
터키	4,192	31	4,239	42	8,431	73
캄보디아	6,419	46	1,532	43	7,951	89
인도	3,237	48	4,158	16	7,395	64
미얀마	5,353	86	2,035	17	7,388	103
기타 <sup>11)</sup>	72968	670	75,607	801	148,575	1,471
합계	2,023,438	15,983	801,049	8,407	2,824,487	24,390

11) 기타에는 마카오, 방글라데시, 멕시코, 브라질, 이란, 사우디아라비아, 네팔, 노르웨이, 투르크메니스탄, 포르투갈, 라오스, 파키스탄, 우크라이나, 에콰도르, 콜롬비아, 아제르바이잔, 우루과이, 아랍에미리트연합, 모로코, 우간다, 벨기에, 네덜란드, 폴란드, 나이지리아, 튀니지아, 브루나이, 헝가리, 이집트, 아프가니스탄, 에티오피아, 뉴질랜드, 베네수엘라, 이라크, 케냐, 가나, 볼리비아, 가봉, 아르헨티나, 파나마, 엘살바도르, 동티모르, 르완다, 칠레, 스위스, 핀란드, 벨라루스, 페루, 아르메니아, 루마니아, 탄자니아, 불가리아, 콩고, 시리아, 세네갈, 에스토니아, 타지키스탄, 요르단, 오스트리아, 슬로바키아, 덴마크, 도미니카 공화국, 체코, 예멘, 카메룬, 니카라과, 코트디부아르, 룩셈부르크, 잠비아, 알제리, 마다가스카르, 남아프리카, 남수단, 수단, 이스라엘, 카타르, 세르비아, 앙골라, 트리니다드 토바고, 파라과이, 팔레스타인, 코스타리카, 모잠비크, 과테말라, 자메이카, 북한, 신트마르텐, 말리, 슬로베니아, 쿠웨이트, 스와질란드, 몰도바, 저지, 도미니카 연방, 쿠바, 그리스, 온두라스, 리투아니아, 바베이도스, 리비아, 알바니아, 아일랜드, 바하마, 콩고 민주 공화국, 라트비아, 아이슬란드, 조지아, 보츠와나, 바레인, 라이베리아, 소말리아, 오만, 정보 없음이 포함된다.

## (2) 구어

- 2019년 구어 원시 말뭉치는 252,127어절이 구축되었다. 그 결과 1-4차 연도에 기구축한 말뭉치와 합산하여 누적 합계 831,469어절 규모의 구어 원시 말뭉치가 구축되었다.

### ① 숙달도별 자료 분포

- 구어 원시 말뭉치는 1급에서 6급, 그리고 최고급인 6급 이상의 자료가 구축되었다. 다음은 숙달도별 구축 규모를 집계한 것이다.

<표 38> 구어 원시 말뭉치의 숙달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	어절 수	84,831	98,155	114,948	109,762	83,834	66,700	21,112	579,342
2018	파일 수	250	236	254	245	152	99	15	1,251
2019	어절 수	47,106	68,484	48,219	56,062	15,311	16,698	247	252,127
	파일 수	162	130	103	102	28	33	1	559
합계	어절 수	131,937	166,639	163,167	165,824	99,145	83,398	21,359	831,469
	파일 수	412	366	357	347	180	132	16	1,810

### ② 국적별 자료 분포

- 구어 원시 말뭉치는 이중 국적 학습자를 포함하여 총 72개국의 자료가 구축되었다.

<표 39> 구어 원시 말뭉치의 국적별 자료 분포

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	151,477	319	22,583	87	174,060	406
베트남	120,989	226	16,806	31	137,795	257
일본	54,133	101	56,442	141	110,575	242
태국	24,218	32	35,361	0	59,579	32

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
필리핀	11,291	5	38,435	25	49,726	30
우즈베키스탄	29,658	63	10,960	0	40,618	63
스리랑카	30,473	5	0	0	30,473	5
러시아	15,659	61	10,489	21	26,148	82
대만	20,763	1	0	0	20,763	1
미국	11,550	65	8,940	50	20,490	115
인도네시아	20,449	78	0	20	20,449	98
미얀마	17,924	52	1,365	0	19,289	52
키르기스스탄	13,560	4	1,122	6	14,682	10
스페인	1,777	3	10,724	3	12,501	6
말레이시아	11,462	27	0	0	11,462	27
카자흐스탄	2,992	3	7,233	0	10,225	3
우루과이	0	1	9,644	0	9,644	1
캄보디아	7,178	4	0	11	7,178	15
영국	2,951	3	2,892	0	5,843	3
몽골	5,759	27	0	0	5,759	27
기타 <sup>12)</sup>	25,079	171	19,131	164	44,210	335
합계	579,342	1,251	252,127	559	831,469	1,810

12) 기타에는 한국, 이탈리아, 포르투갈, 이집트, 아랍에미리트연합, 싱가포르, 호주, 프랑스, 독일, 네팔, 캐나다, 스위스, 스웨덴, 네덜란드, 사우디아라비아, 세르비아, 벨기에, 터키, 콜롬비아, 소말리아, 코스타리카, 알제리, 멕시코, 방글라데시, 예멘, 요르단, 벨라루스, 아르메니아, 불가리아, 오만, 파라과이, 아르헨티나, 아제르바이잔, 헝가리, 모로코, 파키스탄, 홍콩, 폴란드, 노르웨이, 마카오, 인도, 수단, 투르크메니스탄, 우크라이나, 엘살바도르, 동티모, 콩고, 오스트리아, 슬로바키아, 트리니다드 토바고, 그리스, 보츠와나가 포함된다.

## 2.2. 형태 주식 말뭉치

### (1) 문어

- 2019년 문어 형태 주식 말뭉치는 527,744어절이 구축되었다. 그 결과 1-4차 연도에 기구축한 말뭉치와 합산하여 누적 합계 2,037,748어절 규모의 형태 주식 말뭉치가 구축되었다.

#### ① 숙달도별 자료 분포

- 문어 형태 주식 말뭉치는 1급에서 6급, 그리고 최고급인 6급 이상의 자료가 구축되었다. 다음은 각 숙달도별 구축 규모를 집계한 것이다.

<표 40> 문어 형태 주식 말뭉치의 숙달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	어절 수	227,290	252,564	273,140	267,267	235,434	215,260	39,049	1,510,004
2018	파일 수	3,292	2,469	2,195	2,037	1,558	1,325	28	12,904
2019	어절 수	67,161	89,682	99,513	88,729	89,190	59,749	33,720	527,744
	파일 수	966	927	828	695	573	327	46	4,362
합계	어절 수	294,451	342,246	372,653	355,996	324,624	275,009	72,769	2,037,748
	파일 수	4,258	3,396	3,023	2,732	2,131	1,652	74	17,266

#### ② 국적별 자료 분포

- 문어 형태 주식 말뭉치는 128개국의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 41> 문어 형태 주식 말뭉치의 국적별 자료 분포

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	613,137	5,071	61,958	567	675,095	5,638
일본	257,383	2,117	108,341	853	365,724	2,970

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
베트남	129,596	1,244	53,965	429	183,561	1,673
대만	58,566	509	23,845	219	82,411	728
미국	33,205	337	31,245	188	64,450	525
태국	36,238	301	23,989	282	60,227	583
홍콩	56,817	311	0	217	56,817	528
말레이시아	33,559	443	22,100	0	55,659	443
러시아	26,500	204	20,364	123	46,864	327
우즈베키스탄	40,533	155	636	134	41,169	289
몽골	19,050	307	16,854	7	35,904	314
인도네시아	13,947	186	18,927	92	32,874	278
스리랑카	12,045	99	20,592	184	32,637	283
필리핀	17,788	121	9,991	158	27,779	279
프랑스	14,931	111	10,191	24	25,122	135
키르기스스탄	10,134	77	8,768	53	18,902	130
스페인	9,665	82	5,579	12	15,244	94
미얀마	9,110	71	5,818	44	14,928	115
한국	6,852	69	7,157	19	14,009	88
이탈리아	5,948	56	7,522	25	13,470	81
스웨덴	7,328	46	4,691	41	12,019	87
싱가포르	6,419	75	4,770	56	11,189	131
카자흐스탄	8,099	106	2,156	79	10,255	185
영국	4,510	58	2,888	12	7,398	70
캄보디아	6,233	39	956	22	7,189	61
기타 <sup>13)</sup>	72,411	709	54,441	522	126,852	1,231
합계	1,510,004	12,904	527,744	4362	2,037,748	17,266

13) 기타에는 캐나다, 호주, 독일, 우루과이, 터키, 방글라데시, 인도, 네팔, 멕시코, 포르투갈, 사우디아라비아, 마카오, 브라질, 노르웨이, 이집트, 콜롬비아, 투르크메니스탄, 라오스, 이란, 우크라이나, 아제르바이잔, 네덜란드, 벨기에, 아랍에미리트연합, 폴란드, 스위스, 튀니지, 브루나이, 아프가니스탄, 아르헨티나, 벨라루스, 에콰도르, 우간다, 에티오피아, 뉴질랜드, 베네수엘라, 헝가리, 이라크, 모로코, 케냐, 가나, 볼리비아, 나이지리아, 파키스탄, 파나마, 르완다, 아르메니아, 칠레, 요르단, 핀란드, 페루, 세르비아, 루마니아,



## (2) 구어

- 2019년 구어 형태 주식 말뭉치는 210,476어절이 구축되었다. 그 결과 1-4차 연도에 기구축한 말뭉치와 합산하여 누적 합계 591,503어절 규모의 구어 형태 주식 말뭉치가 구축되었다.

### ① 숙달도별 자료 분포

- 구어 형태 주식 말뭉치는 1급에서 6급, 그리고 최고급인 6급 이상의 자료가 구축되었다. 다음은 숙달도별 구축 규모를 집계한 것이다.

<표 42> 구어 형태 주식 말뭉치의 숙달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	어절 수	50,880	51,279	73,160	68,490	68,525	47,581	21,112	381,027
2018	파일 수	165	124	171	165	117	63	15	820
2019	어절 수	31,251	44,629	34,354	39,486	27,580	33,176	0	210,476
	파일 수	78	83	78	73	58	65		435
합계	어절 수	82,131	95,908	107,514	107,976	96,105	80,757	21,112	591,503
	파일 수	243	207	249	238	175	128	15	1,255

### ② 국적별 자료 분포

- 구어 형태 주식 말뭉치는 63개국의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

- 
- 탄자니아, 가봉, 예멘, 알제리, 콩고, 코스타리카, 시리아, 세네갈, 에스토니아, 타지키스탄, 오스트리아, 슬로바키아, 덴마크, 도미니카 공화국, 체코, 소말리아, 파라과이, 불가리아, 카메룬, 니카라과, 코트디부아르, 룩셈부르크, 잠비아, 오만, 마다가스카르, 남아프리카, 동티모르, 남수단, 엘살바도르, 수단, 이스라엘, 카타르, 앙골라, 트리니다드 토바고, 팔레스타인, 모잠비크, 과테말라, 자메이카, 북한, 신트마르텐, 말리, 슬로베니아, 쿠웨이트, 스와질란드, 몰도바, 저지, 도미니카 연방, 쿠바, 그리스, 정보 없음이 포함된다.
- 14) 기타에는 우즈베키스탄, 카자흐스탄, 인도네시아, 영국, 몽골, 한국, 캄보디아, 미얀마, 싱가포르, 호주, 이탈리아, 프랑스, 스페인, 스리랑카, 독일, 네팔, 캐나다, 스위스, 스웨덴, 네덜란드, 사우디아라비아, 세르비아, 벨기에, 콜롬비아, 코스타리카, 터키, 알제리,

<표 43> 구어 형태 주식 말뭉치의 국적별 자료 분포

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	127,082	276	40,021	76	167,103	352
베트남	68,852	140	43,935	81	112,787	221
일본	48,650	92	61,925	150	110,575	242
태국	16,530	27	7,570	0	24,100	27
대만	19,442	43	1,321	11	20,763	54
필리핀	6,277	4	13,736	1	20,013	5
미국	9,611	17	10,379	22	19,990	39
키르기스스탄	12,641	3	2,041	6	14,682	9
말레이시아	11,462	27	0	0	11,462	27
러시아	10,156	30	457	2	10,613	32
기타 <sup>14)</sup>	50,324	161	29,091	86	79,415	247
합계	381,027	820	210,476	435	591,503	1,255

## 2.3. 오류 주식 말뭉치

### (1) 문어

- 2019년 문어 오류 주식 말뭉치는 113,789어절이 구축되었다. 그 결과 1-4차 연도에 기구축한 말뭉치와 합산하여 누적 합계 462,322어절 규모의 문어 원시 말뭉치가 구축되었다.

#### ① 속달도별 자료 분포

- 문어 오류 주식 말뭉치는 1급에서 6급의 자료가 구축되었다. 다음은 각 속달도별 구축 규모를 집계한 것이다.

멕시코, 방글라데시, 이집트, 벨라루스, 아르메니아, 불가리아, 파라과이, 아르헨티나, 아제르바이잔, 헝가리, 모로코, 파키스탄, 홍콩, 투르크메니스탄, 노르웨이, 튀니지, 브루나이, 가나, 칠레, 핀란드, 세네갈, 아프가니스탄, 루마니아, 니카라과, 덴마크, 슬로바키아가 포함된다.

<표 44> 문어 오류 주식 말뭉치의 속달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	합계
2015 - 2018	어절 수	57,852	57,251	62,856	61,169	59,087	50,318	348,533
	파일 수	826	552	508	492	408	307	3,093
2019	어절 수	17,012	17,862	19,585	16,997	17,551	24,782	113,789
	파일 수	223	195	179	148	137	174	1,056
합계	어절 수	74,864	75,113	82,441	78,166	76,638	75,100	462,322
	파일 수	1,049	747	687	640	545	481	4,149

② 국적별 자료 분포

- 문어 오류 주식 말뭉치는 총 85개 국적의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 45> 문어 오류 주식 말뭉치의 국적별 자료 분포

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
일본	61,698	523	47,696	491	109,394	1,014
중국	75,789	686	518	5	76,307	691
미국	31,933	271	25,026	218	56,959	489
베트남	43,861	418	6,434	35	50,295	453
태국	16,423	180	8,206	0	24,629	180
대만	21,758	136	0	63	21,758	199
러시아	18,340	170	0	0	18,340	170
싱가포르	8,435	59	6,339	57	14,774	116
말레이시아	8,964	73	4,427	39	13,391	112
카자흐스탄	9,333	79	0	0	9,333	79
캐나다	4,863	45	3,352	32	8,215	77
영국	4,136	38	3,387	37	7,523	75
호주	3,702	31	2,868	23	6,570	54
우즈베키스탄	5,302	50	0	0	5,302	50
기타 <sup>15)</sup>	33,996	334	5,536	56	39,532	390

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
합계	348,533	3,093	113,789	1,056	462,322	4,149

## (2) 구어

- 2019년 구어 오류 주석 말뭉치는 112,347어절이 구축되었다. 그 결과 1-4차 연도에 기구축한 말뭉치와 합산하여 누적 합계 336,835어절 규모의 구어 오류 주석 말뭉치가 구축되었다.

### ① 숙달도별 자료 분포

- 구어 오류 주석 말뭉치는 1급에서 6급의 자료가 구축되었다. 다음은 숙달도별 구축 규모를 집계한 것이다.

<표 46> 구어 오류 주석 말뭉치의 숙달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015-2018	어절 수	23,683	28,869	36,340	42,468	51,741	38,341	3,046	224,488
	파일 수	85	78	107	119	83	45	3	520
2019	어절 수	21,492	25,306	25,695	19,841	8,371	9,796	1,846	112,347
	파일 수	55	45	58	50	18	19	2	247
합계	어절 수	45,175	54,175	62,035	62,309	60,112	48,137	4,892	336,835
	파일 수	140	123	165	169	101	64	5	767

- 15) 기타에는 한국, 프랑스, 필리핀, 홍콩, 캄보디아, 인도네시아, 몽골, 사우디아라비아, 독일, 키르기스스탄, 파키스탄, 스페인, 뉴질랜드, 미얀마, 스웨덴, 인도, 우크라이나, 네팔, 이탈리아, 방글라데시, 벨라루스, 헝가리, 가나, 나이지리아, 네덜란드, 터키, 아르헨티나, 모로코, 잠비아, 남수단, 마카오, 니카라과, 아제르바이잔, 포르투갈, 폴란드, 이란, 자메이카, 우간다, 슬로바키아, 케냐, 마다가스카르, 멕시코, 이집트, 볼리비아, 루마니아, 도미니카 공화국, 아프가니스탄, 가봉, 콜롬비아, 오스트리아, 과테말라, 노르웨이, 핀란드, 브루나이, 스리랑카, 아일랜드, 바하마, 세르비아, 스위스, 남아프리카, 불가리아, 르완다, 리투아니아, 페루, 라오스, 에티오피아, 세네갈, 벨기에, 코스타리카, 알제리, 파라과이가 포함된다.

## ② 국적별 자료 분포

- 구어 오류 주석 말뭉치는 총 53개 국적의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 47> 구어 오류 주석 말뭉치의 국적별 자료 분포

국적	2015-2018		2019		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
일본	43,056	90	52,247	129	95,303	219
베트남	49,362	128	26,146	48	75,508	176
중국	53,509	124	4,299	6	57,808	130
대만	17,285	22	0	3	17,285	25
미국	8,457	15	8,551	20	17,008	35
말레이시아	3,365	10	7,844	16	11,209	26
러시아	9,828	21	0	0	9,828	21
태국	8,521	29	676	0	9,197	29
영국	2,514	4	2,507	6	5,021	10
우즈베키스탄	1,156	6	1,852	2	3,008	8
기타 <sup>16)</sup>	27,435	71	8,225	17	35,660	88
합계	224,488	520	112,347	247	336,835	767

## 2.4. 1-4차 연도 구축 말뭉치 자료 정제

- 본 연구에서는 1-4차 연도에 구축한 말뭉치의 질적 제고를 위하여 2019년도 말뭉치 구축 작업 과정에서 발견되는 오류를 수정하였다. 아울러 원시 말뭉치와 형태/오류 주석 말뭉치, 형태 주석 말뭉치와 오류 주석 말뭉치 간의 작업 처리 지침의 일관성, 효율성이라는 쟁점을 해소하기 위한 방향

16) 기타에는 카자흐스탄, 몽골, 인도네시아, 싱가포르, 호주, 이탈리아, 스페인, 프랑스, 한국, 캐나다, 미얀마, 스리랑카, 독일, 스웨덴, 네덜란드, 키르기스스탄, 사우디아라비아, 세르비아, 벨기에, 콜롬비아, 코스타리카, 터키, 알제리, 멕시코, 이집트, 벨라루스, 필리핀, 불가리아, 파라과이, 아르헨티나, 아제르바이잔, 헝가리, 캄보디아, 모로코, 홍콩, 네팔, 파키스탄, 인도, 가나, 니카라과, 슬로바키아, 아프가니스탄, 핀란드가 포함된다.

으로 수정 작업을 진행하였다.

### (1) 원시 말뭉치

- 2019년 형태 주석과 오류 주석 대상 표본은 2018년까지 구축된 원시 말뭉치를 대상으로 선정하였다. 이에 따라 형태 주석과 오류 주석 작업 과정에서 원시 말뭉치에 남아 있는 오류나 지침 업데이트에 의해 수정이 필요한 항목들을 함께 수정하였다.

### (2) 형태 주석 말뭉치

- 형태 주석 말뭉치 자료의 정제는 학습자 오류를 포함한 비정형 자료를 처리하는 과정에서 발생할 수밖에 없는 작업자 간의 일관성 문제를 해소하는 데에 초점을 두어 작업을 진행하였다. 주석 유형별로 자료를 추출하여 불일치 항목을 추출한 후 수정 단계에 차등을 두어 수정을 하였다.

### (3) 오류 주석

- 오류 주석 말뭉치의 정제는 전체 자료를 대상으로 오류의 식별과 판정 단계에서 발생할 수밖에 없는 작업자 간의 일관성 문제를 해소하고, 보다 체계적인 오류의 식별과 판정을 위한 지침을 보완하는 방향으로 진행하였다. 새로운 지침을 반영하여 수정한 결과는 공동 연구원들이 최종 검토를 하였다.

### 3. 구축 지원 도구 검증

#### 3.1. 구축 지원 도구 성능 피드백 실시

- 본 연구에서는 구축 지원 도구를 검증하기 위하여 구축 지원 도구 성능 평가팀을 운영하여 정기적/비정기적으로 구축 지원 도구 성능 피드백을 실시하였으며, 구축 지원 도구의 효율적 활용을 위하여 구축 지원 도구의 기능이 추가되거나 개선될 시에 작업자 교육을 실시하였다.

##### ① 구축 지원 도구 성능 평가팀 운영

- 본 연구에서는 표본 등록, 입력, 전사, 형태 주석, 오류 주석 작업 과정에서 구축 도구를 실질적으로 사용하는 팀장급 관리자들을 중심으로 구축 지원 도구 성능 평가 팀을 구성하여 구축 지원 도구를 모니터링하고 성능을 평가하였다. 성능 평가 팀의 위원들은 직접 구축 도구의 성능을 모니터링하고 평가하는 것 외 각 팀원들의 사용 의견을 수렴하는 역할도 담당하였다.

##### ② 구축 지원 도구 성능 평가 및 피드백

- 성능 평가 팀의 모니터링 및 평가 결과는 구축 지원 도구 개발 팀에 상시 보고됨은 물론 격월 단위 구축 팀, 개발 팀, 국립국어원 사업 담당 연구사가 모여 정기 회의를 통해 보고되었다. 구축 지원 도구의 성능 평가는 안정적인 환경에서 말뭉치 구축 작업을 수행할 수 있도록 오작동과 다양한 유형의 버그를 관리하며, 자동 형태 주석 기능을 포함한 세부 기능을 고도화하는 데에 초점이 주어졌다.
- 다음은 성능 평가 회의 개최 일시와 주요 안건이다.

<표 48> 성능 평가 회의 개최 일시와 주요 안건

구분	일시 및 주요 안건
1차	1. 회의 일시: 2019년 5월 2일 14:00 2. 회의 장소: 연세대학교 위당관 301호 3. 안건

	1) 2019년 구축 지원팀과의 협업 업무 내용 점검 2) 2019년 '학습자 맞춤형 자료 나눔터' 자료 서비스 관련 논의 3) 2019년 구축 계획 관련 논의
2차	1. 회의 일시: 2019년 7월 21일 10:00 2. 회의 장소: 연세대학교 위당관 514호 3. 안건 1) 작업 플로우 관리 2) 작업 이력 관리 기능 개선 3) 자동 형태 주식 기능 개선 4) 오류 주식 기능 개선
3차	1. 회의 일시: 2019년 9월 2일 오후 3시 2. 회의 장소: 연세대학교 위당관 514호 3. 안건 1) 구축 지원 도구 내 자동 형태 주식 성능 피드백 2) 기구축 자료 수정 및 데이터 검증 관련 사항 논의
4차	1. 회의 일시: 2019년 11월 5일 오후 3시 2. 회의 장소: 연세대학교 위당관 514호 3. 안건 1) 구축 지원 도구 유지 보수 경과 및 쟁점 점검 2) 데이터 배포 형식 검토 3) 데이터 검증 방식 협의

### 3.2. 기능 추가 및 개선 시 도구 사용 연계

- 구축 지원 도구의 성능 피드백은 궁극적으로 구축 작업자들이 안정적인 환경에서 효율적으로 작업을 하도록 하는 데에 목적이 있다. 따라서 피드백 결과를 반영하여 구축 도구에 새로운 기능이 추가되거나 개선되면 작업자들에게 새로운 구축 도구를 신속하게 배포하고 관련된 기능에 대한 교육을 즉각적으로 실시하였다.

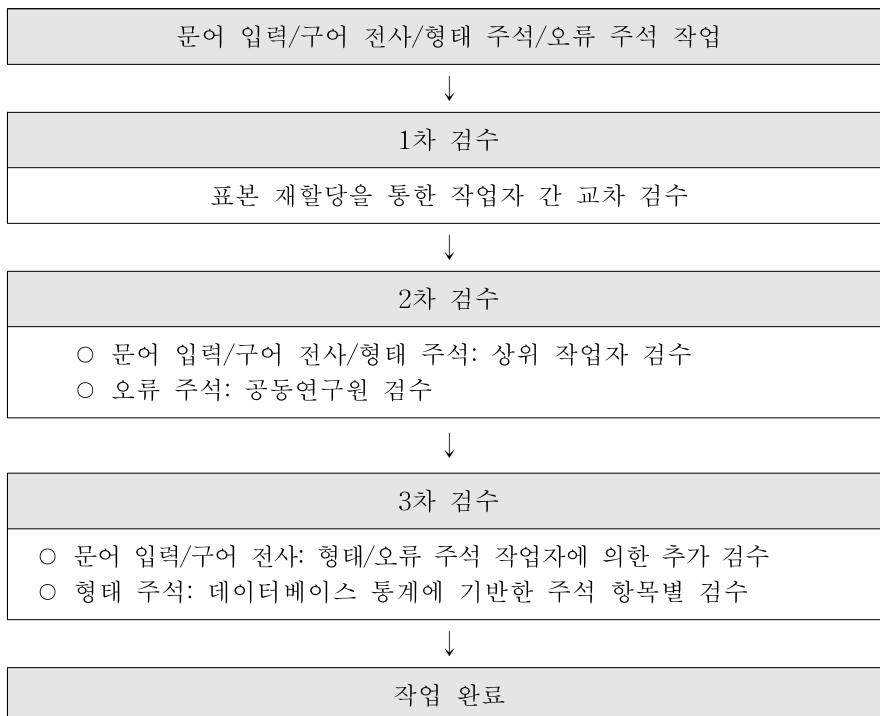


## VI. 구축 말뭉치 검증 정교화

- 구축 말뭉치 검증은 구축 과정에 있는 데이터의 무결성을 확보하기 위한 것이다. 본 연구에서는 이를 위해 각 구축 단계별로 3단계의 검수 작업 절차를 두었으며 이는 전체 데이터를 대상으로 진행되었다.

### 1. 문어 입력 및 구어 전사 작업 검수 시스템 정교화

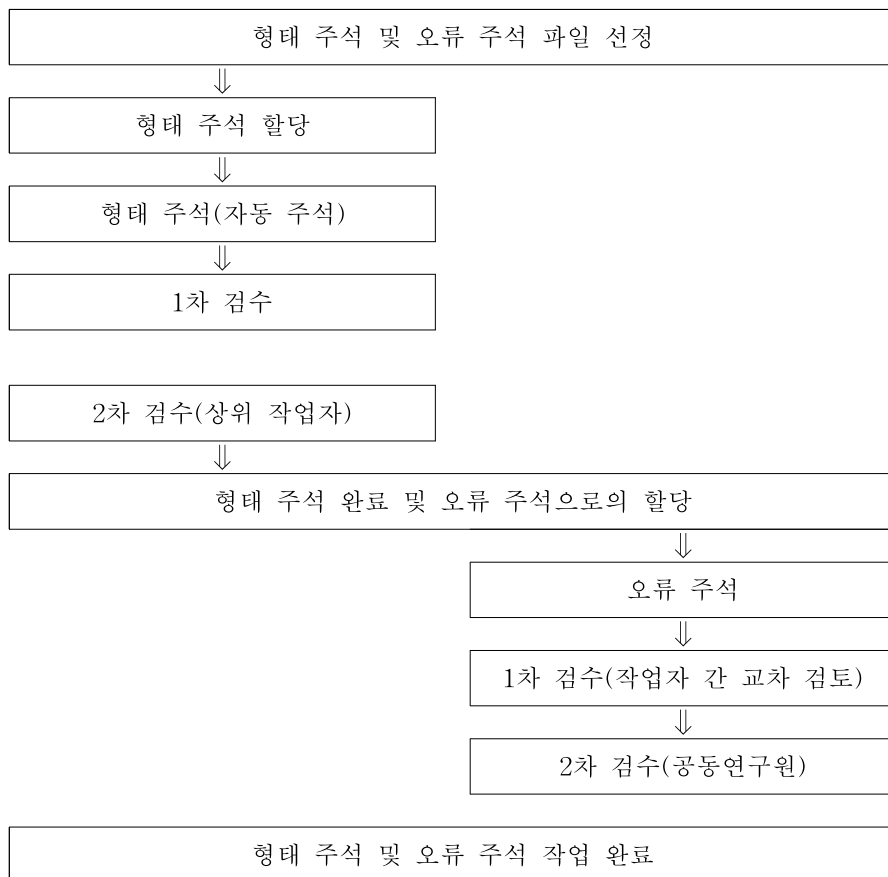
- 한국어 학습자 말뭉치는 구어 전사, 문어 입력, 형태 주석, 오류 주석의 각 단계에서 최소 두세 차례의 전수 검수 과정을 거친다. 이 중 구어 전사와 문어 입력 텍스트의 경우는 다음의 작업 단계인 형태 주석과 오류 주석 단계에서 의해 두 차례의 검수에서 미처 수정하지 못한 오류들을 다시 수정하게 되기 때문에 사실상 3차 검수 작업을 수행하게 된다.



<그림 1> 작업 공정에서의 3단계 검수 체계

## 2. 형태 주석 및 오류 주석 작업 검수 시스템 정교화

- 형태 주석 말뭉치와 오류 주석 말뭉치 구축은 서로 긴밀하게 연계된다. 형태 주석은 구축 지원 도구 내에서의 자동 주석, 자동 주석 결과에 대한 전수 검수(1차 검수), 1차 검수 결과에 대한 상위 작업자의 2차 검수로 순차적으로 이루어진다. 이후 오류 주석 단계로 표본을 할당하면 오류 주석 작업이 이루어진다. 이후 작업자 간의 교차 검토를 통한 1차 점수와 공동연구원에 의한 2차 검수가 이루어진다. 오류 주석 과정에서 형태 주석 검수 과정에서 미처 발견하지 못한 오류들을 수정함으로써 형태 주석은 세 차례의 검수 작업을 하였다



<그림 2> 가공 말뭉치 검수 체계

### 3. 데이터베이스 검증을 통한 검수 시스템 정교화

#### 3.1. 주식별 불일치 항목의 검수

- 주식별 불일치 항목의 검수는 형태 주식의 일관성을 확보하기 위한 통합 검수이다. 데이터베이스에서 실질어, 어미, 조사, 접사, 기호, NA, NONE 등으로 항목별 주식 불일치 검색하여 엑셀로 출력한 후, 불일치 항목을 유형화하여 수정 작업을 진행하였다.

#### 3.2. 작업 중 생성된 오조작 데이터 검증

- 작업 중 생성된 오조작 데이터 검증은 작업 로그나 표본 정보, 주식 정보, 구축 시기 등에 관한 통계 정보를 추출한 후 전체 데이터의 구조나 작업 공정상 논리적으로 타당하지 않는 통계 정보를 확인하여 해당 표본을 집중 검수하는 방식이다.
- **작업 진행 이상 표본 검증:** 작업 로그(log)를 검색하여 작업 진행상의 이상이 의심되는 표본을 검증하는 방식이다. 이에 따라 작업 이력이 남아 있지 않거나 작업이나 검수에 소요된 시간이 30초 미만인 표본을 추출하여 이상 여부를 검증하였다.
- **작업 결과 이상 표본 검증:** 원형태와 주식, 교정 형태와 주식의 존재 여부, 주식이 누락된 항목을 살펴 작업 결과의 이상이 의심되는 표본을 검증하는 방식이다. 원형태가 있으나 주식이 없는 경우, 원형태도 없고 주식도 없는 경우, 교정 형태는 있으나 주식이 없는 경우는 해당 표본을 확인하여 검수하였다.
- **파일의 표본 정보와 파일 정보 간 이상 표본 검증:** 파일의 표본 정보와 파일 정보 간 이상 항목이 의심되는 표본을 추출하여 검증하는 방식이다. 표본 등록 과정에서 해당 정보의 등록이나 저장이 정상적으로 이루어지지 않은 경우로 파일 정보가 있으나 파일 표본 정보가 없는 경우, 파일 정보가 없으나 파일 표본 정보가 있는 경우의 표본을 추출하여 검

수 작업을 실시하였다.

- **표본 정보 중 선행 정보 없이 후행 정보가 있는 표본 검증:** 각 단계별 작업 공정의 결과가 부분적으로 누락되었는지를 확인하는 과정이다. 표본 정보 중 표본 등록, 입력/전사와 같은 선행 정보가 없이 형태 주석이나 오류 주석 정보가 있는 표본을 대상으로 이상 유무를 확인하였다.

## Ⅶ. 결론

### 1. 연구 요약

이 연구는 한국어 학습자 말뭉치 구축 중장기 사업 계획에 따른 제3단계 2차 연도 연구이다. 본 연구는 효율적인 자료의 구축과 관리, 질적 제고를 위해 말뭉치 구축 지침을 정교화하고 그 지침에 따라 기구축된 2,602,780어절(문어 2,023,438어절, 구어 579,342어절) 규모의 말뭉치를 수정·보완함과 동시에 약 800,000어절(문어 550,000어절, 구어 250,000어절) 규모의 말뭉치를 새롭게 구축하는 데에 목적이 있다. 이에 따른 주요 과업과 연구의 성과는 다음과 같다.

#### ○ 한국어 학습자 말뭉치 연구

한국어 학습자 말뭉치 연구는 크게 말뭉치 구축 지침의 정교화, 말뭉치 활용을 위한 연구 체계 구축으로 나뉜다. 5차 연도는 말뭉치의 균형성 확보를 위해 1-4차 연도에 구축한 국내 기관 자료와 이주민 자료의 지속적인 구축과 함께 5차 연도의 집중 구축 대상인 국외 학습자 자료의 수집 및 구축·가공이 이루어졌다. 이에 따라 5차 연도에는 변인이 다양한 자료를 수집하여 이를 효율적으로 구축·가공하는 것이 가능해졌다. 또한 다양한 변인을 고려한 구축·가공 자료의 지속적인 수정·보완이 이루어졌다.

이러한 구축·가공 작업을 수행하면서 동시에 각 단계별 구축 지침이 정합성과 체계성을 갖출 수 있도록 수집 지침, 자료 처리 지침, 문어 입력 지침, 구어 전사 지침, 형태 주석 지침과 오류 주석 지침을 보완하였다. 특히, 수집에서 구축, 가공에 이르기까지 순차적이며 순환적인 작업 과정이 매우 유기적으로 이루어지는 점을 고려하여 각 단계별 작업 지침의 기술에 있어 서로 배치되는 부분이 없는지 검토하였다.

말뭉치 활용을 위한 연구 체계 구축은 2015-2019년 구축 말뭉치를 활용한 분석 연구, 학습자 말뭉치 구축 이론 및 활용 연구의 집대성, 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구로 나누어 이루어졌다. 이 중 2015-2019년 구축 말뭉치를 활용한 분석 연구는 현재까지 구축된 말뭉치를 활용한 계량적 분석 연구로 학습자 말뭉치의 주석 체계를 기반으로 하여 형태 주석 말뭉치와 오류 주석 말뭉치의 세부 주석 통계를 숙달도 단계별,

언어권별로 제시하여 사용자들이 목적에 따라 참조 자료로 활용할 수 있도록 하였다.

그다음으로 진행할 예정인 학습자 말뭉치 구축 이론 및 활용 연구의 집대성은 4차 연도까지 제시해 온 다양한 유형의 활용 모형을 넘어서 학습자 말뭉치 구축 방법론과 말뭉치를 활용한 연구 성과를 체계적으로 종합하기 위한 연구이다. 이를 위해 본 연구에서는 그 첫 번째 단계로 중국어권 한국어 학습자의 자료를 계량적으로 분석하여 비중국어권 또는 중국어권 외의 언어권 학습자의 중간언어 사용 양상을 연구하는 데에 활용할 수 있는 참조 자료집을 기획하여 집필하였다. 이는 학습자 말뭉치 사용자들에게 학습자 말뭉치 활용의 실제적 사례가 될 뿐만 아니라 한국어 학습자의 중간언어 연구를 위한 기초 자료를 제공한다는 점에서 의의가 있다.

마지막으로 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구는 4차 연도까지의 연구에서 제시한 학습자의 분포를 고려한 현실적인 균형성 확보, 특수 말뭉치(sub corpus) 구축이라는 방향성이 과연 타당한가에 대한 검증에 초점을 맞추어 진행되었다. 또 국외에서 학습자 말뭉치 수집 시 균형성을 확보하기 위해 어떠한 방법론을 적용하였는지 사례 조사를 진행하였다. 이 작업은 2020년까지 계속 진행할 예정이다.

## ○ 한국어 학습자 말뭉치 관련 교육 및 홍보

한국어 학습자 말뭉치 관련 교육은 구축 실무자와 사용자를 대상으로 하여 이루어졌다. 먼저 구축 실무자 교육은 효율적이고 체계적인 말뭉치 구축을 위한 것으로, 말뭉치 구축의 각 단계에 참여하는 인력들이 그 대상이며 지침 교육과 도구 사용 교육을 중심으로 1차 연도부터 현재까지 지속되고 있다. 그 밖에도 5차 연도에는 파이썬(Python)을 활용한 말뭉치 자료 처리 특강을 통해 말뭉치 구축 인력으로서의 전문성을 제고하였다.

한편, 사용자를 대상으로 한 교육은 학습자 말뭉치 아카데미를 통해 이루어졌다. 학습자 말뭉치 아카데미는 총 3회에 걸쳐 이루어졌으며, 그간 ‘학습자 말뭉치 나눔터’에서 제공하는 검색 기능을 활용하는 것에 중점을 두어 왔던 것에서 한 단계 더 나아가 말뭉치 분석 도구인 AntConc와 R을 활용하여 ‘학습자 말뭉치 나눔터’에서 내려받은 자료를 보다 다양하게 활용할 수 있는 방법을 제공하였다. 그 외에도 학술대회 발표를 통해 학습자 말뭉치 연구와 활용 모형을 제시하고 학습자 말뭉치를 널리 알림으로써 학계의 관심을 고조시켰다.

## ○ 한국어 학습자 말뭉치 수집 및 가공

5차 연도의 한국어 학습자 말뭉치 구축은 1-4차 연도 사업의 집중 구축 대상인 국내 교육 기관과 이주민 자료의 균형성을 맞추기 위한 구축을 계속 진행하였다. 이와 함께 국외 자료의 본격 구축이 진행되었다. 수집은 1-4차 연도에 국내 학습자와 이주민 자료 수집에 참여하였던 교사와 함께 국외 대학, 세종학당 소속 교원등이 수집에 참여하였다. 이에 따라 한국어 학습자 말뭉치 구축 및 가공은 1-4차 연도에 구축한 말뭉치의 수정·보완과 함께 원시 말뭉치 1,048,052어절(문어 795,925어절, 구어 252,127어절), 형태 주석 말뭉치 738,220어절(문어 527,744어절, 구어 210,476어절), 오류 주석 226,136어절(문어 113,789어절, 구어 112,347어절) 규모의 말뭉치가 새롭게 구축되었다. 그 결과 1-5차 연도(2015-2019년)에 구축한 전체 말뭉치의 규모는 원시 말뭉치 3,650,832어절(문어 2,819,363어절, 구어 831,469어절), 형태 주석 말뭉치 2,629,251어절(문어 2,037,748어절, 구어 591,503어절), 오류 주석 말뭉치 799,157어절(문어 462,322어절, 구어 336,835어절)이 되었다. 그 외에도 새롭게 수정·보완된 지침에 따라 1-4차 연도에 구축한 말뭉치를 수정하였고, 여기에 5차 연도 구축 말뭉치를 포함시켜 전체 말뭉치가 일관성과 체계성을 갖추도록 하였다.

## ○ 말뭉치 구축 지원 도구 관리 및 구축 말뭉치 검증 정교화

한국어 학습자 말뭉치는 표본 등록에서 말뭉치 주석 가공까지 전체 작업 공정을 관리하고 수행할 수 있는 말뭉치 구축 지원 도구를 활용하여 구축하고 있다. 5차 연도에는 구축 실무 연구원들을 중심으로 성능 피드백팀을 구성하여 구축 지원 도구의 성능 개선과 안정화를 위한 정기적인 피드백을 실시하였다. 또한 격월로 국립국어원, 구축 지원 도구 개발팀, 구축팀의 실무회의를 통해 구축 지원 도구 사용과 말뭉치 구축에 관한 다양한 쟁점에 대해 논의하고 해결 방안을 모색하였다. 그럼으로써 보다 효율적으로 구축 지원 도구를 관리하고 말뭉치를 구축해 나갈 수 있었다. 한편, 구축 말뭉치 검증은 구축된 말뭉치 자료의 질적 제고를 위한 것으로, 문어 입력과 구어 전사, 형태 주석, 오류 주석 공정에서의 3단계 검수, 시스템 기반의 데이터 검증을 통한 오조작 데이터와 이상 데이터 검수를 상호보완적으로 적용하였다. 이러한 검증 정교화를 통해 방대한 자료 구축 과정에서 필연적으로 발생할 수밖에 없는 문제들을 최소화하고자 하였다.

## 2. 연구의 의의 및 기대 효과

2015년부터 2020년까지 3단계 6개년의 중장기 계획을 통해 구축되는 <한국어 학습자 말뭉치>는 외국어 또는 제2 언어로서 한국어를 학습하는 비모어 화자의 말뭉치로서 국가 주도의 공공 말뭉치이다. 이 한국어 학습자 말뭉치 구축은 한국어교육의 선진화와 과학화의 기반 자료이자 한국어교육 분야의 숙원 사업으로 한국어교육 연구자, 교수자, 학습자에게 매우 활용도가 높고 유용한 자료라는 점에서 큰 의미가 있다. <한국어 학습자 말뭉치>의 의의 및 기대 효과를 구체적으로 설명하자면 다음과 같다.

### ○ 국가 주도의 한국어 학습자 균형 말뭉치 구축

한국어 학습자의 언어를 관찰하기 위해서는 체계적으로 설계되고 구축된 학습자 산출 자료가 적정 규모만큼 필요하다. 한국어 학습자 말뭉치가 구축되기 이전까지는 이러한 자료를 필요로 하는 개인 연구자들이 자료를 수집하고 자료 사용에 필요한 제반 처리를 해 왔다. 그러나 현실적으로 학습자 자료는 접근이 쉽지 않고 수집과 처리에 많은 시간과 노력이 소요될 뿐만 아니라 이를 위한 설계에 상당한 전문 지식과 경험이 필요하다. 그렇기 때문에 연구자 또는 교수자 개인이 자료를 구축하는 데에는 한계가 있다. 이에 따라 2002년 문화체육관광부의 주도로 구축된 50만 어절의 학습자 말뭉치를 구축한 바 있으나 저작권 문제로 실질적인 자료의 배포와 활용이 불가하였다. 이 연구에서는 이러한 문제를 해소하기 위하여 IRB 규정에 따라 자료 제공 및 사용에 관한 학습자의 동의를 얻어 자료를 수집함으로써 국가 주도의 공공 자료로서 자유로운 배포와 폭넓은 활용이 가능한 한국어 학습자 균형 말뭉치를 구축하고 있다.

### ○ 한국어교육 학계와 협업을 통한 말뭉치 구축

한국어 학습자 말뭉치 구축은 한국어교육 학계의 숙원 사업인 만큼 한국어교육 연구자와 교수자의 많은 관심과 주목 속에서 사업이 착수되어 현재까지 성공적으로 추진해 오고 있다. 1차 연도에서 5차 연도까지 구축한 약 360만 어절 규모의 말뭉치와 함께 큰 의미를 부여할 수 있는 또 하나의 성과는 국내외의 한국어학당, 대학(원), 세종학당 등의 한국어 교육 기관과 교사의 적극적인 협조와 지지를 기반으로 대규모의 말뭉치를 수집했다는 것이다. 교육과정 운영, 수집 교사 및 학습자의 참여 독려, 실질적인 수집 과정에 많은 어



려움이 있음에도 불구하고 지난 5년간 자료 수집에 국내외의 한국어 교육 기관과 1,000여 명 이상의 교사가 함께해 주었다. 또한 이 과정에서 수만 명의 학습자가 자료 제공과 사용에 대한 동의와 함께 자료를 제공해 주었다. 한국어 학습자 말뭉치는 이처럼 많은 사람들의 참여로 구축된 공동의 연구 성과물이라는 점에서 매우 값진 자료이며 큰 의미가 있다.

### ○ 한국어교육 이론의 체계화 및 교육 자료 구축의 기반 조성

한국어 학습자 말뭉치는 연구자 또는 교수자의 경험과 직관만으로 모두 알기 어려운 학습자의 중간언어 특성을 실증적으로 보여 주며, 이는 체계적인 교육 자료와 교수법, 평가 도구 등을 개발하는 데에 중요한 기초 자료가 된다. 한국어교육 학계와 교수 현장에서는 이미 오래전부터 이와 같은 학습자 말뭉치의 의의와 중요성, 필요성에 대해 상당한 공감대가 형성되어 있었다. 그렇기 때문에 이를 활용한 연구와 교수·학습에 대한 관심 또한 매우 큰 편이다. 이러한 관심은 현재까지 구축된 학습자 말뭉치는 시범 운영 중인 학습자 말뭉치 나눔터(<https://kcorpus.korean.go.kr>)의 접속자 수와 자료 배포 요청수가 지속적으로 증가하는 것을 통해서 알 수 있다. 또한 최근에는 학습자 말뭉치를 활용한 연구 성과로 학위 논문이나 학술지 논문이 출간되고 있기도 하다. 이러한 연구 성과는 한국어 학습자의 중간언어 발달과 습득의 특성을 밝히고, 교수 현장에서의 교육 자료와 교수·학습 방법을 개발하기 위한 기초 자료로 활용될 수 있다.

### ○ 한국어 학습자의 다양화에 따른 교수·학습 환경의 과학화

말뭉치를 이용한 제2 언어 또는 외국어 교수·학습은 세계적인 추세이며 영어 교육을 비롯한 외국어교육에서는 이미 오랜 경험과 노하우를 바탕으로 상당한 성과를 거두고 있다. 한국어 학습자 말뭉치는 한국어를 모국어로 하지 않는 외국인 학습자의 언어 발달이나 습득의 특성을 실증적으로 관찰할 수 있는 자료로 학습자의 제1 언어, 숙달도 단계, 수집 과제, 장르 등의 변인을 고려하여 구축되었다. 이러한 자료는 변인에 따라 매우 세분화된 특성을 가진 학습자의 언어 사용 양상을 비교·분석하는 데에 최적화된 자료로 활용할 수 있다. 그리고 그 결과는 점차 다변화·세분화되어 가는 한국어 학습자의 요구에 부응하기 위한 맞춤형 교육의 설계·개발·운영을 위한 참고 자료로 활용될 수 있다. 아울러 최근 들어 4차 산업혁명과 함께 교육 환경과 교수 매체, 교수 방법 등에서 큰 변화가 요구되고 있는데, 학습자 말뭉치는 이러한

흐름에 맞는 교수 자료로도 활용될 수 있다. 이는 학습자에게 체계적이고 과학적인 교육을 제공하기 위한 발판이 되며, 궁극적으로 한국어교육의 질적 제고에 기여할 것이다.

#### ○ 학습자 말뭉치 활용 연구를 통한 국제 수준의 학술 교류 기반 조성

최근 외국어 또는 제2 언어 교육 분야에서 학습자 말뭉치를 활용한 연구가 활발하게 이루어지고 있다. 말뭉치 기반 연구를 선도하는 영어 교육 분야의 경우 CLC(Cambridge Learner Corpus), ICLE(International Corpus of Learner English)와 같은 대규모 말뭉치를 비롯한 말뭉치가 다양하게 구축되어 있으며, 학습자 말뭉치 연합회(Learner Corpus Association)와 같은 학술 단체를 중심으로 자료를 공유하고 정기적인 학술 모임 등을 개최하는 등 학술 교류가 매우 활발하게 이루어지고 있다. 한국어교육 분야에서도 학습자 말뭉치에 대한 요구가 매우 커지고 있지만 사용 목적에 따라 폭넓게 활용 가능한 말뭉치의 부재, 말뭉치 활용 방법에 대한 부담감으로 인해 연구자들이 쉽게 접근하지 못하였던 것이 사실이다. 그러나 이제 국가 주도의 공공 자료로서 구축된 한국어 학습자 말뭉치는 학습자 말뭉치 나눔터의 검색 기능을 사용하여 누구나 쉽게 접근할 수 있다. 또한 학습자 말뭉치 아카데미를 통해 기초 과정에서 심화 과정까지 말뭉치 활용법에 대한 사용자 교육을 실시함으로써 말뭉치에 대한 활용 능력을 제고하였다. 이러한 노력은 한국어 학습자 말뭉치에 대한 사용자의 관심과 활용의 폭을 넓히고 학습자 말뭉치 기반 연구에 관한 다양한 논의거리들을 이끌어 내 한국어교육 연구자 간의 학술 교류는 물론 국제 수준의 학술 교류 기반을 조성하는 데에 이바지할 것이다.

#### ○ 한국어의 세계화 및 국제 경쟁력 강화

한국 언어·문화에 대한 세계인의 관심이 나날이 커지고 있다. 이는 한국어 학습자의 양적인 증가 외에도 점점 확대되고 있는 지역 분포, 다양해지는 학습 목적 등을 통해 어렵지 않게 확인할 수 있다. 한국어 학습자 말뭉치는 이처럼 서로 다른 환경, 다양한 목적으로 한국어를 배우고자 하는 한국어 학습자들의 요구에 맞는 체계적이고 질 높은 교육을 제공하기 위한 기초 자료로 활용할 수 있다. 그럼으로써 일차적으로는 한국 언어·문화를 세계화하고, 더 나아가 한국 언어·문화를 널리 알리고 정치·경제·사회의 각 분야에서 폭넓게 활동할 수 있는 국제 인력을 양성함으로써 해서 한국의 국제 경쟁력 강화에 기여할 수 있을 것이다.

### 3. 보고서 활용 방안

본 연구는 한국어 학습자 말뭉치 구축을 위한 기초 연구와 활용 모형 연구, 실제 말뭉치 구축을 주요한 과업으로 하였다. 본 보고서에는 그러한 방법론, 절차와 결과, 그와 관련된 쟁점들이 체계적으로 기술되어 있으며 말뭉치 구축의 각 단계별 지침이 첨부되어 있다. 이는 다양한 목적의 사용자들에게 한국어 학습자 말뭉치 구축에 관한 이론과 적용, 그리고 말뭉치의 활용을 위한 지침으로 활용될 수 있다.

#### ○ 한국어 학습자 말뭉치 구축에 관한 이론적 지침

한국어 학습자 말뭉치는 한국어를 모국어로 하지 않는 외국인 한국어 학습자의 자료를 수집하여 구축한 자료로 비정형의 발화를 포함하고 있으며 매우 다양한 변인이 전제된다는 점에서 특수 말뭉치로 분류할 수 있다. 따라서 말뭉치를 구축함에 있어 범용 말뭉치 구축에 관한 이론적 체계를 따르되 학습자 자료의 특성에 따른 다양한 쟁점에 대한 세심한 고려가 필요하다. 본 연구는 학습자 말뭉치 구축에 관한 이론적 논의를 찾아보기 힘든 상황에서 많은 시행착오를 거치며 1차 연도에서 5차 연도까지 5년간의 연구를 수행해 왔다. 본 보고서는 그 과정에서 학습자 말뭉치 구축을 위한 자료 수집과 구축, 가공에 관한 수많은 쟁점을 해결해 나가는 과정을 체계적으로 기술하였다. 그러한 점에서 본 보고서는 지금까지 체계화되지 못하였던 한국어 학습자 말뭉치 구축에 관한 실제적인 모형이자 이론적 지침으로서 활용될 수 있다.

#### ○ 한국어 학습자 말뭉치 구축의 실제를 위한 실용적 지침

본 보고서는 자료 수집 및 처리, 입력과 전사, 형태 주석, 오류 주석 지침이 포함되어 있다. 각 단계별 구축 지침에는 자료 처리에 관한 다양한 쟁점과 해결 방안이 구체적인 사례와 함께 상세하게 제시되어 있다. 이러한 지침은 자료의 호환성을 위해 <21세기 세종 한국어 균형 말뭉치>의 구축 지침을 기반으로 하되 비모어 화자 자료인 학습자 말뭉치의 특성을 반영하고 5년간의 구축 과정에서 대두되는 수많은 쟁점들을 반영하여 정교화해 온 것으로 학술적으로는 물론 향후 학습자 말뭉치를 구축하고자 하는 기관이나 연구자들에게 실용적인 지침으로 활용될 수 있다.

## 4. 정책 제안

### ○ 자료의 질적 제고를 위한 정비 사업 계획으로의 확장

3단계 6개년의 중장기 계획이 마무리되는 2020년 이후에는 본격적인 서비스에 앞서 말뭉치 전체를 대상으로 한 검증 작업이 필수적으로 이루어져야 한다. 지난 5년간의 연구에서는 매해 신규 말뭉치를 구축하는 작업과 함께 수정·보완된 지침에 따라 기구축 말뭉치를 추가적으로 검수하는 작업을 지속적으로 수행해 왔다. 그럼에도 440만 어절에 이르는 방대한 자료를 모든 공정에서 많은 작업자들이 단계별 수작업을 통해 구축한 자료이기 때문에 일정 정도의 완결성을 갖추기까지 한층 심화된 종합적인 검수 작업, 즉 정비 작업이 잘 이루어져야 한다. 이는 명실공히 국가 주도의 공공 자료로서 학습자 말뭉치 전반의 질을 높이고 신뢰성 있는 자료로 견고하게 자리매김하도록 하기 위한 것이다.

### ○ 중장기 계획에서 장기 사업 계획으로의 확대

4차 산업혁명 시대를 맞이하여 인공지능을 기반으로 한 다양한 산업이 빠르게 발전하고 있다. 이러한 시대적 흐름에 발맞추어 국립국어원은 2019년에 ‘4차 산업혁명 대비 국어 빅데이터(말뭉치)’ 구축 사업을 본격적으로 시작하였다. 한국어 학습자 말뭉치는 한국어 비모어 화자들의 자료를 기반으로 한 특수 말뭉치로 국어 빅데이터의 한 축을 이루고 있기에 2020년까지의 중장기 계획에서 한발 더 나아가야 한다.

즉, 지속적으로 구축 규모를 확대하고 다양한 자료들을 확보할 수 있는 장기적인 계획을 수립해야 할 필요가 있다. 2020년 1차 구축 완료 시점에 달성할 것으로 보이는 약 440만 어절 규모의 학습자 말뭉치는 그 규모면에서 충분하지 않다. 학습자의 매우 세분화된, 그리고 다양한 변인이 포함된 균형 말뭉치로서는 충분치 않은 규모이며, 5천만 어절에 달하는 The Cambridge Learner Corpus(CLC)와 비교해 보아도 매우 작은 규모이다. 또한 최근 산업계와 교육계에서 각광 받고 있는 인공지능 기반 언어 교육 매체를 개발하기 위한 기초 /자료로 활용하기에도 매우 부족한 규모이다. 그렇기 때문에 말뭉치 전체의 규모를 지속적으로 확대해 나아가기 위한 장기 사업 계획은 반드시 필요하며, 이 계획에는 반드시 다음과 같은 사항을 고려해야 한다.

- **구어 말뭉치 확충:** 2019년까지 5년간 구축된 말뭉치는 원시 말뭉치 3,650,832어절 중 문어 말뭉치 2,819,363어절, 구어 말뭉치 831,469어절로 구어 말뭉치가 전체 말뭉치의 약 22.8%로 그 비중이 매우 적음을 알 수 있다. 이는 첫째, 구어 말뭉치의 수집과 구축, 가공 작업 공정이 문어 말뭉치에 비해 상대적으로 복잡하고 둘째, 많은 비용과 시간, 노력이 소요되는 작업이며, 셋째, 한정된 예산이라는 한계 속에서 이루어 낸 결실이다. 그럼에도 자료의 활용도 면에서는 충분치 않은 규모이며 다양한 변인으로 이루어진 학습자 말뭉치의 특성을 고려할 때 최소한 문어에 준하는 규모의 자료가 구축되어야 한다.
- **오류 주석 말뭉치 확충:** 한국어 학습자 말뭉치는 원시 말뭉치, 형태 주석 말뭉치, 오류 주석 말뭉치로 구성된다. 이 중 오류 주석 말뭉치는 특수 말뭉치로서 한국어 학습자 말뭉치의 변별적 특성을 가장 분명하게 드러내 주는 것으로 2019년까지 구축된 말뭉치를 기준으로 799,157어절(문어 462,322어절, 구어 336,835어절)이 구축되어 있다. 이는 3,650,832어절 규모의 원시 말뭉치 중 약 21.9%에 해당하는 자료로 매우 적은 규모이다. 오류 주석 말뭉치에 부착된 주석 정보는 학습자 말뭉치 나눔터에서 사용자들이 한국어 학습자의 언어 사용에 나타나는 오류 유형을 검색하도록 하는 색인의 역할을 하며, 이러한 자료의 양이 많을수록 다양한 유형의 오류 정보를 제공할 수 있을 뿐만 아니라 그러한 결과를 객관화하고 일반화할 수 있게 된다. 그러나 현재 구축된 분량은 검색 결과의 객관성과 신뢰성을 담보하기에는 턱없이 부족한 상황이다. 객관성과 신뢰성 확보뿐만 아니라 한국어 교사나 학습자, 일반인(국외 외국인 학습자 및 국내 이주민 학습자 등) 사용자의 활용도를 높이기 위해서는 오류 주석 말뭉치의 규모를 지속적으로 확장해 나가야만 한다.<sup>17)</sup>

---

17) 대규모의 오류 주석을 효율적으로 부착할 수 있는 방안으로 오류 주석을 간소화하거나 구축된 데이터 정보를 기반으로 하여 오류 주석을 자동 부착하는 것에 대한 가능성을 모색한 바 있다. 그러나 구축 규모가 적어 테스트가 이루어지기 어려워 주석 자동 부착의 효율성을 가늠하기 어렵다. 물론, 학습자 말뭉치의 특성상 주석 부착의 자동화도 높은 성공률을 보장하기 어렵기는 하지만, 대량의 오류 주석 말뭉치 구축을 위해서는 구축 규모를 지속적으로 확장하기 위한 방안을 마련해야 한다.

## ○ 한국어 학습자 말뭉치를 활용한 교수·학습 자료 개발 사업으로의 연계

2015년부터 3단계 6개년 계획으로 진행해 온 한국어 학습자 말뭉치의 구축은 2020년에 마치게 되며, 그 성과물로 원시 말뭉치를 기준으로 약 440만 어절 규모의 말뭉치가 구축된다. 이렇게 구축된 말뭉치는 이미 시범 서비스 중인 한국어 학습자 말뭉치 나눔터를 통해 사용자들에게 배포되어 한국어교육 연구와 교수·학습에 활용되기 시작하였으며 학술지 논문과 학위 논문 등이 출간되고 있다. 이러한 성과들은 실증적 자료를 기반으로 한 한국어교육에 관한 논의로서 의의가 있으나, 실제적인 활용의 측면에서 교수·학습 현장의 교사와 학습자에게까지 파급력이 미치기까지 어느 정도 시간이 필요하다. 따라서 구축 결과를 바탕으로 하여 국가 주도로 교수·학습 현장에서 활용이 가능한 교육 자료를 개발할 필요가 있다. 이는 학습자 말뭉치 구축에 협조한 수요 기관에 구축 자료가 활용된 결과물을 제공하고, 한국어 학습자 말뭉치의 통합적인 활용 모형을 보인다는 점에서 큰 의미가 있을 것이다. 이렇게 구축 말뭉치를 활용하여 한국어 교수·학습 현장의 과학화를 선도하는 교육 자료 개발 사업을 진행하는 등의 사업 연계는 큰 의의가 있을 것이다.

## ○ 자료의 지속적인 업데이트를 위한 수집 네트워크와 수집 경로의 혁신

한국어 학습자 말뭉치 구축 연구의 성패는 말뭉치 구축의 첫 단계인 수집 단계에서 결정된다. 즉, 구축·가공 가능한 질이 확보된 학습자의 언어 자료를 일정 규모 이상으로 구축할 수 있는 효율적인 방법이 매우 절실하다. 2020년 사업까지는 한국어교육 학계를 중심으로 국내외의 한국어 교육 기관과 교수자의 도움으로 자료를 수집할 수밖에 없는 실정이다. 그러나 빅데이터 구축의 일환으로 한국어 학습자 말뭉치를 지속적으로 업데이트하고 양적으로 확충해 나가기 위해서는 다음과 같이 수집 네트워크와 수집 경로의 혁신이 필요하다.

- 정부 부처 간의 공조 체계를 통한 수집 네트워크 확대: 영어 학습자 말뭉치 중 하나인 The Cambridge Learner Corpus(CLC)의 경우 영어 능력 평가 자료를 지속적으로 수집하고 있다. 이처럼 숙달도 평가와 연계한 자료 수집의 장점은 학습자의 수준, 과제 등이 균질한 자료를 수집할 수 있으며, 지속적으로 자료를 업데이트하면서 전세계 학습자의 자료를 대규모로 수집할 수 있다는 것이다. 한국어의 경우 국립국제교육원에서 주관하고 있는 한국어능력시험(TOPIK), 법무부 사회통합프로그램의 성취도 평가, 세종학당

재단의 성취도 평가나 숙달도 평가 등을 들 수 있다. 앞으로 이와 같은 국가 공인 언어 능력 평가 시행 기관의 협조를 이끌어 낼 수 있다면 자료 수집을 효율적으로 지속해 나갈 수 있을 것이다.

- **온라인 수집 시스템 개발 및 활용:** 본 연구에서는 주로 우편, 이메일, 웹하드 등의 경로를 통해 자료를 수집하였다. 이는 주로 특정 교육 기관이나 교사 인력풀을 통해 자료를 수집하였기 때문에 가능한 일이었으며 현실적으로 선택한 최선의 효율적인 선택이었다. 그러나 이러한 방식으로는 빅데이터를 구축하기 매우 어렵다. 전세계의 학습자를 대상으로 학습자 말뭉치를 알리고 자율적인 수집 참여를 독려하며 지속적으로 자료를 수집해 나가기 위해서는 온라인 수집 시스템을 매개로 한 자료 수집이 필요하다.

#### ○ 효율적 사용을 한국어 학습자 말뭉치 나눔터의 기능 개선

한국어 학습자 말뭉치 나눔터를 통해 배포 서비스가 시작되면서 학습자 말뭉치에 대한 사용자들의 다양한 의견이 수렴되고 있다. 현재 시범 운영 중인 한국어 학습자 말뭉치 나눔터는 사용자들이 비교적 손쉽게 자료를 검색하고 내려받을 수 있는 기능이 비교적 잘 구현되어 있으며 이는 5년간 사용자는 물론 구축팀과 수요 기관의 끊임없는 소통의 결과라고 할 수 있다. 그럼에도 보다 편리함과 활용도를 제고한다는 측면에서 다음과 같은 기능이 추가될 필요가 있다. 이러한 기능이 추가된다면 사용자들의 다양한 목적과 요구에 부응한 보다 만족도 높은 서비스를 제공할 수 있을 것이다.

- **사용자 목적에 따른 검색 기능의 차별화:** 학습자 말뭉치 사용자의 범위는 연구자, 교사, 학습자, 그리고 한국어교육에 관심을 가지는 일반 국민까지 상당히 넓다고 할 수 있다. 이는 다양한 사용 목적, 사용 능력 등으로 이어진다. 예를 들어 한국어 교사나 학습자는 학습자 언어 사용에서 빈번히 나타나는 오류에 관심 가질 것이다. 이러한 사용자들은 오류를 검색할 수 있는 단순 검색 기능만으로도 원하는 정보를 손쉽게 찾아볼 수 있을 것이다. 반면, 연구자는 보다 정교화된 검색 기능과 통계 정보, 더 나아가 제공하는 자료에 대한 2차 가공까지 요구할 수 있다. 이와 같이 사용 목적에 따라 세부적인 사용 기능이 달라짐을 고려하여 사용자 집단 또는 사용 목적에 따라 접속할 수 있는 경로를 달리하여 검색 기능이나 정보를 차별화한다면 보다 효율적인 활용이 가능할 것이다.

- **말뭉치 바꾸니 기능 추가:** 현재 학습자 말뭉치 나눔터는 자료를 검색하고 검색한 자료를 내려받는 기능을 주로 제공하고 있다. 그러나 특정 사용자들, 즉 연구자들은 언어 숙달도 단계, 언어권 등과 같은 학습자 변인, 장르, 과제 유형 등과 같은 자료 변인에 따라 특정 자료를 새롭게 재구성하여 자료를 검색하거나 사용하기를 원한다. 이러한 요구를 반영하여 사용자가 사용 목적에 따라 필요로 하는 말뭉치를 재구성하여 내려받을 수 있는 말뭉치 바꾸니 기능을 추가한다면 학습자 말뭉치를 더욱 효율적으로 활용할 수 있을 것으로 기대된다.
  
- **사용자를 위한 말뭉치 구축 지원 도구 제공:** 학습자 말뭉치 나눔터에서 제공하는 말뭉치는 본 연구를 통해 구축된 자료로 구성되며 검색 기능은 연구진에 의한 주식 정보에 기반을 두고 있다. 그러나 학습자 말뭉치 나눔터를 통해 서비스되는 자료는 변인이 매우 다양한 데에 비해 전체 규모가 상대적으로 크지 않아 소수 언어권 혹은 특정 집단의 학습자 자료를 관찰하고자 할 경우 그 규모가 충분하지 못하다는 아쉬움이 있다. 이에 따라 사용자들은 개인이 수집한 자료를 추가로 구축하여 양을 확충하거나 사용 목적에 따라 추가적인 주식 가공을 하고자 하는 경우가 많다. 이는 학습자 말뭉치 구축 지원 도구에 대한 관심과 배포에 대한 요구로 이어지고 있다. 학습자 말뭉치의 폭넓은 활용을 위해서는 사용자를 위한 말뭉치 구축 지원 도구의 개발에 대한 고려가 필요해 보인다.



## 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 사업 보고서, 국립국어원.
- 강현화 외(2016), 2016년 한국어 학습자 말뭉치 연구 및 구축 사업 보고서, 국립국어원.
- 강현화 외(2017), 2017년 한국어 학습자 말뭉치 연구 및 구축 사업 보고서, 국립국어원.
- 강현화 외(2018), 2018년 한국어 학습자 말뭉치 연구 및 구축 사업 보고서, 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), 한국어문화교육학회.
- 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 교육 14(3), 국제한국어교육학회.
- 김미옥·정희정(2003), 한국어 학습자 작문에 나타난 어휘 오류 분석, 제3회 한국어

- 교육 국제 워크숍 발표 요지, 연세대 언어정보연구원 외국어로서의 한국어교육 연구센터, 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 처리 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 외(2015), 2015년 한국어 학습자 말뭉치 기초 연구 및 구축 사업, 국립국어원.
- 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, 한국어교육 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.

조철형 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 사

- 전의 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.
- 한송화(2018), 한국어 학습자의 종결어미 사용 양상과 오류 연구, 문법교육 33, 한국문법교육학회, 166-210.
- 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) Comprthensible 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. 2019년 한국어 학습자  
말뭉치 구축 지침

## 차 례

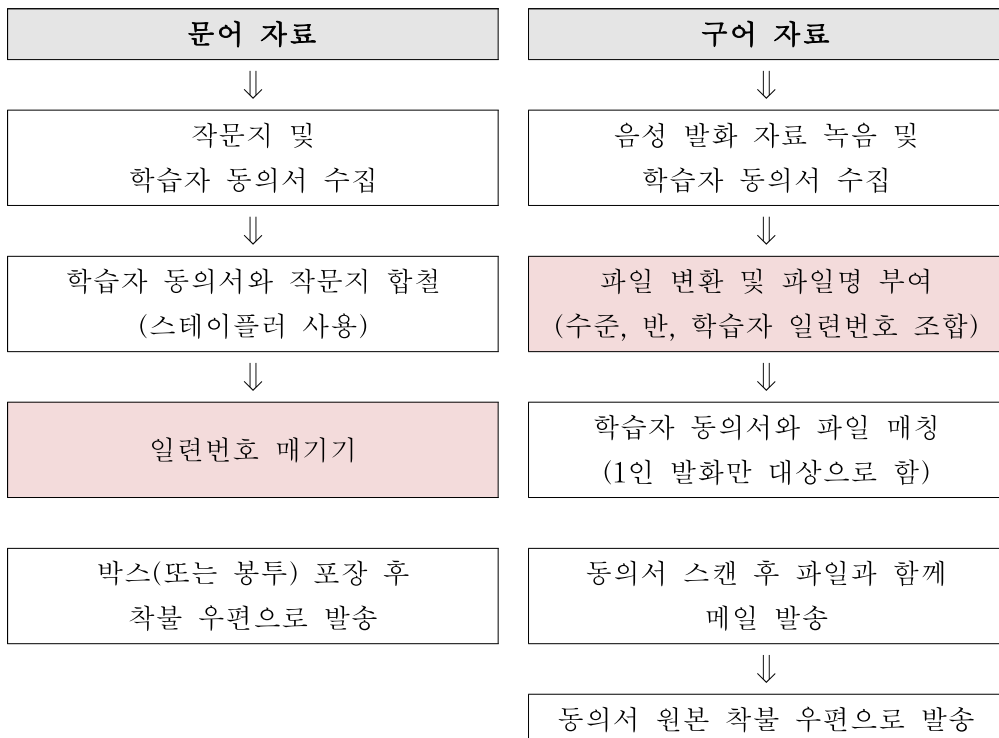
한국어 학습자 말뭉치 수집 지침 .....	1
한국어 학습자 말뭉치 자료 처리 지침 .....	43
한국어 학습자 말뭉치 문어 입력 지침 .....	48
한국어 학습자 말뭉치 구어 전사 지침 .....	55
한국어 학습자 말뭉치 형태 주석 지침 .....	77
한국어 학습자 말뭉치 오류 주석 지침 .....	143

## 한국어 학습자 말뭉치 자료 수집 지침

### 1. 자료 수집 대상 및 수집 자료

- ▶ 대상: 국내 한국어 교육 기관의 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







통지한다(착불 결제용).

120-749 서울시 서대문구 연세로 50 연세대학교 연세우유사무소  
언어정보연구원 한국어 학습자 말뭉치 연구실  
홍혜란 (전화 010-8727-9024)

## 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 받

도록 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 교육 기관의 교육과정 실러버스에 이미 포함되어 있는 작문을 활용하여 이를 수집함. 글의 종류 및 주제는 각 기관의 교육과정에 따름

###### ② 성취도 평가 수집

- 각 교육 기관의 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 주제

로 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, 게시문 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 학습

자 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 진사자를 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 졸업좌담회, 말하기 대회 등의 자료 수집

- 졸업좌담회, 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 싶습니까? 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



수집 시기	문제	수준
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"> <li>○ 가장 아끼는 물건은 무엇인가?</li> <li>○ 왜 그 물건을 아끼는가?</li> <li>○ 어떻게 그 물건을 가지게 되었는가?</li> </ul>	
26주차	<p>여러분은 취미로 무엇을 배우고 싶습니까? ‘내가 취미로 배우고 싶은 것’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 취미로 배우고 싶은 것은 무엇인가?</li> </ul> <p>(※ 한국어를 배우고 싶다는 내용은 쓰지 마십시오.)</p> <ul style="list-style-type: none"> <li>○ 왜 그것을 배우고 싶은가?</li> <li>○ 그것을 배운 후에 무엇을 하고 싶은가?</li> </ul>	
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> <ul style="list-style-type: none"> <li>○ 나의 나쁜 생활 습관</li> <li>○ 습관 때문에 생기는 불편하거나 안 좋은 점</li> <li>○ 습관을 고치기 위해 해야 할 일</li> </ul>	
30주차	<p>잊지 못할 추억’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 어떤 추억인가요?</li> <li>○ 왜 지금까지 기억에 남아 있는가?</li> <li>○ 언제 그 추억이 떠오르는가?</li> </ul>	
32주차	<p>갖고 싶은 직업’이라는 제목으로 글을 써 보십시오. 단 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 직업명, 하는 일, 그 일을 하려는 이유, 그 일에 필요한 조건</li> </ul>	
34주차	<p>나의 성격’이라는 제목으로 글을 써 보십시오. 단 아래에 제시된 내용에 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 성격의 특징, 장점과 단점, 고치고 싶은 부분과 그 이유</li> </ul>	
36주차	<p>내가 생각하는 행복’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 행복을 위해서 어떤 노력을 하는가?</li> </ul>	

수집 시기	문제	수준
	<ul style="list-style-type: none"> <li>○ 언제 행복하다고 느끼는가?</li> <li>○ 행복은 무엇이라고 생각하는가?</li> </ul>	
38주차	<p>‘내가 좋아하는 책’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 가장 좋아하는 책은 무엇인가?</li> <li>○ 그 책은 어떤 내용인가?</li> <li>○ 그 책을 좋아하는 이유는 무엇인가?</li> </ul>	
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> <ul style="list-style-type: none"> <li>○ 닮고 싶은 사람은 누구인가?</li> <li>○ 왜 그 사람처럼 되고 싶은가?</li> <li>○ 그 사람처럼 되기 위해서 어떻게 해야 하는가?</li> </ul>	
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> <ul style="list-style-type: none"> <li>○ 체질에 맞는 음식</li> <li>○ 수면의 질</li> <li>○ 적당한 운동</li> <li>○ 긍정적인 사고</li> </ul>	
44주차	<p>올바른 인터넷 사용 태도’에 대한 자신의 견해를 서술하십시오. 단, 아래 제시한 &lt;올바른 인터넷 사용 태도의 예&gt; 중에서 세 가지를 선택하여 쓰되, 각각의 태도를 지키지 않았을 경우에 나타나는 부작용의 예를 포함해야 합니다.</p> <p>&lt;올바른 인터넷 사용 태도의 예&gt;</p> <ul style="list-style-type: none"> <li>○ 상대방의 의견 존중하기</li> <li>○ 타인의 사생활 보호하기</li> <li>○ 의견 차이 인정하기</li> <li>○ 바른 언어 사용하기</li> <li>○ 정확한 정보 올리기</li> </ul>	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>&lt;신문의 기능&gt;</p>	

수집 시기	문 제	수준				
	<ul style="list-style-type: none"> <li>○ 사건 보도</li> <li>○ 여론 조성</li> <li>○ 정보 제공</li> <li>○ 소통의 분위기 조성</li> </ul>					
48주차	<p>다음 글을 읽고 '감시 카메라 설치 확대'에 대한 자신의 견해를 서술하십시오. (찬성하거나 반대하는 입장중 하나를 선택하여 서술 할 것. 단 아래 제시된 각 입장의 논거 중 두 개 이상을 제시할 것.)</p> <div style="border: 1px solid black; padding: 10px; margin: 10px 0;"> <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 사생활 침해            범죄 예방 효과 불분명            가해자의 인권 보호         </td> </tr> </table> </div>	찬성	반대	사회 안전 유지 범죄 예방 인권보다 공인이 우선	개인의 사생활 침해 범죄 예방 효과 불분명 가해자의 인권 보호	
찬성	반대					
사회 안전 유지 범죄 예방 인권보다 공인이 우선	개인의 사생활 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"> <li>○ 내가 생각하는 성공이란 무엇인가?</li> <li>○ 그것을 이루기 위해 필요한 것은 무엇인가?</li> <li>○ 그 이유는 무엇인가?</li> </ul>					
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>&lt; 선의의 거짓말이란 &gt;</p> <ul style="list-style-type: none"> <li>○ 선의의 거짓말이란 무엇인가?</li> <li>○ 선의의 거짓말은 언제 필요한가?</li> </ul>					

수집 시기	문제	수준
	○ 선의의 거짓말이 가질 수 있는 문제점은 무엇인가?	
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>&lt; 예술 교육의 필요성 &gt;</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>&lt;자연 보존과 자연 개발&gt;</p> <p>○ 자연 보존과 자연 개발 중 어느 것이 더 중요하다고 생각하는가?</p> <p>○ 그렇게 생각하는 이유는 무엇인가? (2가지 이상 쓰시오.)</p>	
58주차	<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 꾸려 나가는 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 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 속</p> </div>	

수집 시기	문제	수준
	<div> <p>에 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.</p>	
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	

수집 시기	글의 종류	주제	수준
34주차	인터뷰	만남	
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.	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	문 제	수준
	다. ○ 10년 후에 어떻게 살고 싶은가? ○ 그 이유는 무엇인가? ○ 무엇을 준비해야 하는가?	
24주차	여러분이 소중하게 생각해서 사랑하는 물건은 무엇입니까? ‘내가 가장 아끼는 물건’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다. ○ 가장 아끼는 물건은 무엇인가? ○ 왜 그 물건을 아끼는가? ○ 어떻게 그 물건을 가지게 되었는가?	
26주차	‘한국의 첫인상’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다. ○ 한국에 언제 왔는가? ○ 시내, 길거리는 어떤 모습이었는가? ○ 한국 사람들은 어땠는가? ○ 한국 음식은 어땠는가? ○ 가장 기억에 남는 것은 무엇인가?	
28주차	여러분은 고치고 싶은 생활 습관이 있습니까? ‘고치고 싶은 나의 생활 습관’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다. ○ 나의 나쁜 생활 습관 ○ 습관 때문에 생기는 불편하거나 안 좋은 점 ○ 습관을 고치기 위해 해야 할 일	
30주차	‘잊지 못할 추억’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다. ○ 어떤 추억인가요? ○ 왜 지금까지 기억에 남아 있는가? ○ 언제 그 추억이 떠오르는가?	
32주차	‘나의 한국 생활’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다. ○ 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"> <li>○ 어디에 살고 싶은가?</li> <li>○ 어떤 집에 살고 싶은가? 왜 그런가?</li> <li>○ 집은 어떻게 꾸미고 싶은가?</li> <li>○ 집에서 누구와 무엇을 하고 싶은가?</li> </ul>	
36주차	<p>고향의 음식을 소개하는 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 음식 이름</li> <li>○ 주로 언제 먹는 음식인가?</li> <li>○ 어떻게 만드는가?</li> <li>○ 한국 음식과 비슷한 음식이 있는가? 어떤 점이 비슷하고 어떤 점이 다른가?</li> </ul>	
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"> <li>○ 닮고 싶은 사람은 누구인가?</li> <li>○ 왜 그 사람처럼 되고 싶은가?</li> <li>○ 그 사람처럼 되기 위해서 어떻게 해야 하는가?</li> </ul>	
40주차	<p>취업을 하려고 합니다. 무슨 일을 하고 싶은지 생각해 보고 자기 소개서를 쓰십시오.</p> <ul style="list-style-type: none"> <li>○ 살아온 과정</li> <li>○ 성격의 장단점</li> <li>○ 지금까지의 경험 또는 경력</li> <li>○ 앞으로의 계획</li> </ul>	
42주차	<p>‘내가 생각하는 행복’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 행복을 위해서 어떤 노력을 하는가?</li> <li>○ 언제 행복하다고 느끼는가?</li> <li>○ 행복은 무엇이라고 생각하는가?</li> </ul>	고급
44주차	<p>‘절약과 저축’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 절약하기 위해 무엇을 하는가?</li> <li>○ 저축을 하고 있는가? 왜 그런가?</li> </ul>	

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 좋을까?)</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 생각하십니까? 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>&lt; 선의의 거짓말이란 &gt;</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"> <li>○ 노후 준비가 왜 필요한가?</li> <li>○ 노후를 위해 무엇을 준비해야 하는가?</li> <li>○ 여러분은 노후를 어떻게 준비하고 있는가?</li> </ul>	
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"> <li>○ 칭찬이 미치는 긍정적인 영향은 무엇입니까?</li> <li>○ 부정적인 영향은 무엇입니까?</li> <li>○ 효과적인 칭찬의 방법은 무엇입니까?</li> </ul>	
58주차	<p>‘여성의 사회적 지위와 역할’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> <li>○ 여성이 일을 해야 한다고 생각하는가? 왜 그런가?</li> <li>○ 한국에서 여성의 사회적 지위는 어떻다고 생각하는가? 고향과 비교해서 높은 편인가? 낮은 편인가?</li> <li>○ 여성이어서 좋은 점 혹은 좋지 않은 점이 있다고 생각하는가?</li> </ul>	
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 꾸려 나가는 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 안</p> </div>	

종적 자료 수집 시기	문제	수준
	<div>정감을 잃어버릴 위험이 많고, 이런 변화 속에 적응하지 못하는 이들은 사회 변화의 뒷전으로 밀려날 수밖에 없게 된다.</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,	

수집 시기	발화 유형	주제	수준
		싫어하는 음식 등)	
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 사람이예요? 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 소중해요? 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 만들어요? 한국 음식과 비슷한 음식이 있어요? 어떤 점이 비슷하고 어떤 점이 달라요? 고향 음식을 소개하는 글을 쓰세요.	
40주차	그림을 보고 이야기를 순서대로 써 보세요.	고급
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</p> </div>	

종적 자료 수집 시기	문 제	수준
	<div>있어요. 그리고 저는 눈이 작고 쌍꺼풀이 없어요. 그래서 성형 수술을 하고 싶어요.</div> <p>여러분도 자신의 외모에 대해 고민을 해 봤어요? 여러분은 자신의 아름다움을 위해 어떤 노력을 하고 있어요? 글에서 읽은 친구의 고민에 대한 여러분의 생각을 써 보세요.</p>	
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 될까요? 부정적인 영향은 될까요? ‘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	중급
22주차	인터뷰	★ 식사 관련 텍스트 읽기 ★ 나의 식사 생활	
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><sup>18)</sup>

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 핍니다. 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.

---

18) 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><sup>19)</sup>



19) 그림은 『엄마와 함께 읽어요. 지식 쑥쑥 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><sup>20)</sup>

※ 다음 그림을 보고 이야기를 만들어 보세요.



1



2

20) 그림은 '키즈짱 잼쟁동화-끓어지는 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 과제1

## 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? 국제 결혼은 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 주제로의 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.

## 한국어 학습자 말뭉치 구축을 위한 기획 자료 수집 과제2

- 수준별로 다음의 과제를 차례대로 진행하도록 합니다.

	초급	중급	고급
준비하기(공통)	자기소개		
과제 1. 제시 자료 보고 말하기	1. 두 장의 제시 그림을 보고 공통점과 차이 점을 비교해서 말하기 2. 두 장소 중 좋 아하는 곳, 그 이유 말하기	1. 한국 전래 동 화 그림 보고 스토리텔링 하 기 2. 이어지는 내용 상상해서 말하 기	1. 비디오 자료 보 고 스토리텔링하 기 2. 이어지는 내용 상상해서 말하기
과제 2. 대화하기	친구와 여행 계획 세우기 - 서로 가고 싶은 여행지 얘기하 기 - 자신이 말한 장 소가 좋은 이유 를 설명하고 여 행지 결정하기 (설명하기, 설득 하기) - 교통편, 숙박, 여행지에서 할 일, 음식 등에 대한 계획 세우 기(제안하기, 의 사 결정하기)	좋아하는 드라마/ 영화에 대해서 이 야기하기 - 좋아하는 드라 마/영화, 최근 에 재미있게 본 드라마, ..... - 출연 배우, 줄 거리, 기억에 남는 장면, ..... - 친구에게 드라 마/영화 추천 하기	결혼에 대한 찬반 의견 나누기

- 교사가 수집 준비, 수집 진행을 돕는 진행자로서 참여할 것을 권장하나 여  
건에 따라 학습자가 단독으로 과제를 수행해도 무방합니다(자료 참고).
- 학습자의 자연스러운 언어 사용을 관찰하는 것이 목표이므로 연습을 하거  
나 발화할 내용을 메모하여 보면서 말하지 않도록 합니다.

## 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): \_\_\_\_\_

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?

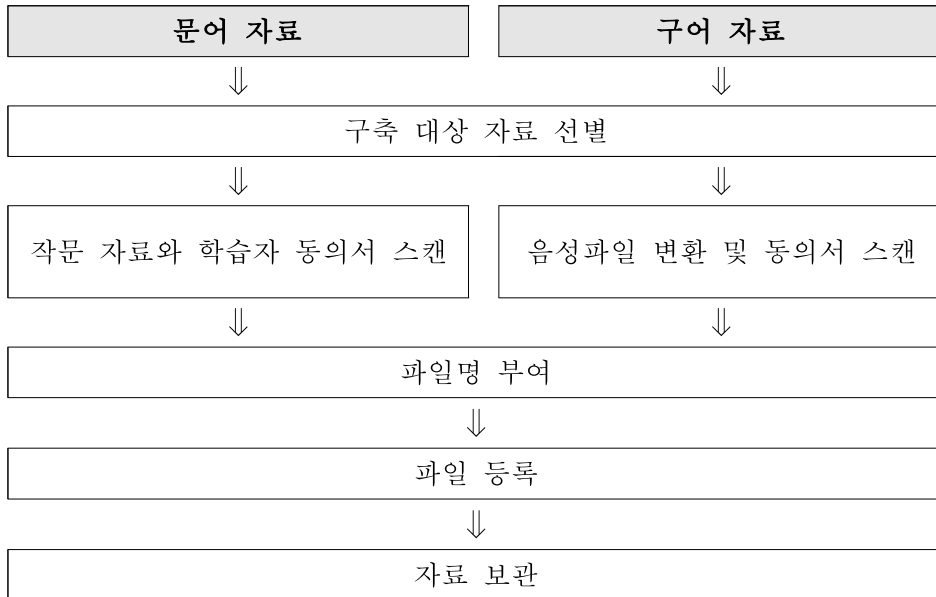
☐ 거의 없음    ☐ 하루 1시간~하루 3시간    ☐ 하루 3시간~ 5시간    ☐ 하루 5시간 이상



## 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
<ul style="list-style-type: none"> <li>○ 학습자 동의서에 서명한 자료</li> <li>○ 학습자 동의서의 개인 정보 모두 입력된 항목 선정</li> <li>○ 동일 학습자의 자료 2개 이하로 제한</li> <li>○ 영어권, 일본어권 자료/1, 5, 6급 단계의 자료 우선 선정</li> </ul>	
<ul style="list-style-type: none"> <li>○ 완결된 텍스트 작문 자료 선정</li> <li>○ 텍스트의 길이 평균 100어절 이상의 자료 선정. 단, 숙달도 단계를 고려하여 1, 2급은 50어절 내외의 자료를 포함함</li> <li>○ 복사 또는 스캔 파일의 경우 화질이 좋은 자료 선정</li> </ul>	<ul style="list-style-type: none"> <li>○ 완결된 담화 단위의 발화 자료 선정</li> <li>○ 발화 길이 2분 이상의 자료 선정</li> <li>○ 음질이 좋은 자료 선정</li> <li>○ 교사의 개입이 많지 않고 학습자의 발화가 중심인 자료를 우선 선정</li> </ul>

## 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.

접수 날짜	기관명	자료 유형	자료 내용	자료 수(수준별)						합계	자료 관련 메모
				1급	2급	3급	4급	5급	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							
				수준	자료 번호	페이지 번호							
					수집 기관: 코드화하여 비공개 처리됨								
		국적											
		자료 유형											
자료 및 학습자 유형													

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
학습자 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.

나는 영화를 TV에서 방송할 때 특정 장면을 삭제하는 것을 반대하는  
입장이다.

첫 번째 이유는 방송 심의의 기준이 모호하는데 방송 회사에 따라서  
방송 심의가 다르다. 똑같은 장면인데 이 채널에서 못 보지만  
다른 채널에서 볼 수 있기 때문에 삭제하는 것이 효과가 없다고  
생각한다.

두 번째는 특정 장면보다 영화 전체의 메시지가 더 중요한데  
만약에 그 장면을 삭제하면 사람들이 그 영화를 제대로 볼 수  
없다. 단는 폭력이나 야한 장면이 있는 영화 보통 모두 커시 후에  
네

- 필적을 알아보기 어려운 것은 일단 가장 가까운 상태로 입력한다.
- 단락을 구분하여, 문장 단위로 입력한다. 단락은 자판의 엔터키로 구분하  
고, 들여쓰기는 반영되지 않는다.
- 전체 본문 입력이 끝나면 ‘주석 자동 생성’을 클릭하여 본문 주석을 확인  
하고 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 오른쪽을 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 <sup>(BARFLY)</sup> '바프라이' 술집에 ~~가야~~ 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.

<sup>(세)</sup> 3달전에 미국에서 한국까지

- 학습자가 작문 중간에 교정 기호를 사용하거나 교정에 관한 문구를 적어 넣은 경우 이를 반영해서 수정 입력한다. 단, 학습자의 답안에 교사가 같

은 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.

- 단락 구분은 하지 않으며 한 행에 한 문장을 입력하는 것을 원칙으로 한다.

### 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. 단, 학습자가 부적절한 위치에 습관적으로 찍은 온점은 문장부호로 보기 어려우므로 반영하지 않는다.
- 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>ㄱㄱㄱㄱㄱ</EX\_Alpha>

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

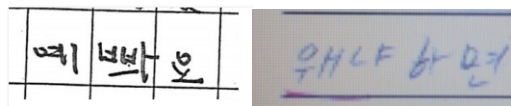
<예> <CNI>ㄱㄱ</CNI>

- ‘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 ㄱ기호를 입력하지 않도록 주의한다.)

<예> 1) <EX\_Symbol>『』「」《》</EX\_Symbol>  
: [괄호기호] ‘ㄴ’ 입력 후 ‘한자키’ 눌러서 선택  
2) <EX\_Symbol>※★</EX\_Symbol>  
: [일반기호] ‘ㄴ’ 입력 후 ‘한자키’ 눌러서 선택  
3) <EX\_Symbol>m’ kg kcal</EX\_Symbol>  
:[단위기호] ‘ㄴ’ 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>



- 예쁘다 : 시스템 ‘예ㄱ요’로 입력 후 ‘한글기호’ 마크업 처리  
- 예<NSS>ㄱ</NSS>요
- 우애나하면 : 시스템 ‘ㄱ나하면’으로 입력 후 ‘한글기호’ 마크업 처리

- <NSS>ㄱ</NSS>냐하면
- 3) 좌우대칭된 ㄱ이 포함된 ‘가’
- <NSS>ㄱ</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> 번지에 삽니다.

#### 5. 기타

- 스캔 과정에서 일부분이 잘린 경우, 잘린 부분이 한두 글자, 또는 한두 단

어 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 및 주제어	<Forabs>
	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 심양에 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 淸國產 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
억양 단위	하강	.	2:네.
	상승	?	2:어디 갈 거예요?
	약한 상승이나 하강	,	1:그래서 그랬는데 이번에,
	활기, 기운찬 어조	!	1:아!
	하나의 억양 단위가 끼어들 에 의해 끊어진 경우	-	6:기자가 와서 - 2:응. 6:- 그 사람한테 인터뷰를 시작했어.
	두 억양 단위가 휴지 없이 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 03:52.2 아 왜 거짓말을 하나요? ☞ 발화 겹침이 있음을 알 수 있음
잘 들리지 않는 부분	잘 들리지 않는 부분	<X X>	<X통X>
	전혀 들리지 않는 부분	<note>안들림</note>	1: 거기까지 <note>안들림</note> 2:<note>안들림</note> 너무한 거 같더라.
	들리지 않는 음절수만큼	X	2: 근데 그거 진짜 XX해야 되겠더라
전사자의 설명	-	<note>여기에 전사자의 설명을	1:응.<note>장소 이동으로 인해 잠시 멈춤</note>

대분류	소분류	기호	예시
		입력하세요</note>	2:우리 때는 그런 거 없었잖아.
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)	소리 나는 대로 적고, 원래의 형태 없이는 내용을 이해하기 어려울 때는 ( ) 안에 학습자의 발화를, 괄호 밖에는 규범 표기를 밝힘	1: 친구와(칭구와) 강남에(간남에) 갔습니다.(갔슨니다.)
	숫자 표기	발음에 따라 한글로 표기, 만 단위 이상도 모두 붙여 씀	7:오늘 제 동생이 이케 하나 오백 원이라고 사 가지고 왔더라구.
	외래어·외국어 표기	발음 규범에 따라 한글로 표기	2: 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.
	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,

대분류	소분류	기호	예시
	담화표지	~	1:그::~~ 음::~~
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='목청가다듬 는소리'>	-
	들이마시는 숨(쓰)	<vocal desc='들이마시는 숨'>	-
	내쉬는 숨(후우)	<vocal desc='내쉬는숨'>	-
	혀 차는 소리(쓰)	<vocal desc='혀차는소리' >	-
	헛기침(에 험)	<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.@>
	박수치면서 말하는 부분		<# #>	5:우우 <vocal desc='박수'> <#이리 와 이리 와.#>
	노래를 부르는 부분		<M M>	<M동해물과 백두산이 마르고 닳도록M>
	박수나 손가락 부딪치는 소리		<kinesics desc=' '>	<kinesics desc='박수'>
	대화 흐름에 영향을 주는 전화벨 소리라든지 기타 음성 아닌 소리		<event desc=' '>	<event desc='전화벨소리'>
2차 전사	구어적 변이형	( ) 안에는 학습자의 발 화를, ( ) 밖 에는 철자 보 충. 단, 주 등 장하거나 쉽 게 원래 형태 로 이해될 수 있는 것들은 일일이 철자 형을 붙여주 지 않음	( )	1:책상 위에 놔 뒀.  → 놓아(놔) 로 전사하지 않고 변이형 그대로 전사해도 무방함
	발음 오류	한국어 모어 화자의 발음과 음운적으로 구분이 될 정도로 발음에 오류가 있는	( )	저는 어제 밥을 먹었어요.(먹어요.)



대분류	소분류		기호	예시
		경우		
		한국어 모어 화자의 발음과 음성 혹은 변이음의 구분이 모호한 경우	(    )	가구(가구)<note>‘가’의 ㄱ에서 유성음으로 발음</note>  밥을(밥을)<note>ㄹ 권설음</note>
		음운규칙으로 인해 한국어의 철자대로 발음되지 않으나, 학습자가 이를 철자대로 발음하는 경우와 같이 철자 전사를 통해 학습자의 발음 오류를 반영하기 어려운 경우	(    )	무조건(무조건)<note>경음화 안됨</note> 같이(같이)<note>구개음화 안됨</note>
	외국어, 외래어 발음	외국어나 외래어를 원어에 가깝게 발음할 경우	(    )	인터뷰(인틸뷰)<note>원어식 발음</note>  외래어표기법을 우선으로 하여 교정어절을 제시하고, 학습자의 원어를 살려 적는 것을 원칙으로 한다.
		한국어 모어 화자와 다른 규범을 할 경우	(    )	카페: 현실 발음 [까페], 학습자 발음 [카페] 버스: 현실 발음 [빠스], 학습자 발음 [버스]

대분류	소분류		기호	예시
				→ 각각 ‘카페(카페)’, ‘버스(버스)’로 처리함  카페(카페)<note>현실발음 과 다름</note>
	방언형 표시		확실한 방언형(대응하는 표준형 형태소가 없는 것)의 경우 태그 부착	차는 <dia>여일</dia> 있어.
	긴 휴지		(1초 이상의 쉼은) 0.1초 단위까지 표시	2:{1.2}그럴까?
	짧은 휴지		한 어절 안에서의 짧은 쉼은 ‘.’로 표시	2:아::~ 그리고 어::~ 남의 의견을 잘 듣고 수용하고 대화..로 타협해야 된다고 하면서,
	인용		<Q Q>	후자들은 현대 사회에 대하여 <Q불확실성의 시대는 아니다.Q> 라고 말하죠.
	텔레비전 방송이나 강의 등 텍스트 종류 표현		<R R>	1:그다음에 <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> :	5: 그게 어찌면 <name1> 선배님이라든지 다른 선배님들 말:: 들은 걸 생각해 보면,

대분류	소분류	기호	예시
		신용카드 번호 <address> : 주소  <tel-num> : 전화 번호	
세그멘테이션	분할 단위는 어절 단위로 한다		내{1.2}/가 (X) 내{1.2}가 (O)  *빗금은 세그멘테이션 분할을 나타내는 표시이다.

## 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

#### 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 윤곽에서 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) !

### 나. 끊어진 억양단위(불임표의 사용)

- 계속해서 말을 할 의향이 있는데, 끼어들음을 당해서(혹은 적극적인 호응에 의해서) 말끝이 잘린 경우는 다음의 예와 같이 불임표(-)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 엘란에서 자동으로 처리된다.

#### 4) 잘 들리지 않는 부분

- 잘 들리지 않는 부분은 <X X>안에 전사한다.

<예> 그때도,  
 <X보통X> 그런 자만심이 있었다.

- 화자의 발화 내용이 전혀 들리지 않는 부분은 <note>안들림</note>으로 전사한다. 이때는 억양을 확인할 수 없으므로 문장부호를 넣지 않는다.

<예> 근데 그거 <note>안들림</note>

- 들리지 않는 음절은 그 음절의 수만큼 X를 붙인다.

<예> 근데 그거 진짜 XX해야 되겠더라.

#### 5) 전사 기호의 중복

- 전사 기호가 서로 중복이 될 경우에는 기본적으로 문장부호를 우선하고 (즉, 어절의 가장 가까운 곳에 붙이고) 웃음이나 박수 등 준음성의 표현을 2차로 한다.

<예> 2:<@길섭이가 그거를 딱 보더니,  
허 <Q맛이 없어Q> 그러는 거 알어?@>  
7:1<vocal desc='웃음'> 그까 -  
2:놀래 애가 놀래 가지고 <Q맛이 없어,Q>  
7:- 그게 주식이에요 언니.

#### 6) 전사자의 설명

- 전사자가 전사의 어려움에 대한 것이나 특별한 설명을 붙일 필요가 있을 때는, <note> </note> 태그를 사용하여 표현한다. 녹음 상태, 잡음, 동시 다발대화 설명, 특이한 발음 상태 등의 설명이 덧붙을 수 있다. 주석 태그는 다음과 같이 붙여야 한다.

<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 철자법대로 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: 친구와(칭구와) 강남에(간남에) 갔습니다(갔습니다)

- ☞ 한국어 학습자 발음치는 일반 발음치와 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 학습자의 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 왔더라(왔더라구)  
: 500원으로 적지 않는다.

### 3) 외래어·외국어 표기

- 외래어나 외국어는 아래의 예에서처럼 발음에 따라 한글로 적는다.

<예> 2:어떻게 이거 크립 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,

- 하나의 형태소가 완전히 반복되는 발화라 하더라도 학습자가 자가 수정을 하는 경우 ‘=’를 붙인다.

<예> 1:좋아= 좋았어요.

- 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)
- 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 된다거나, 두 음절이 한 음절 겹핥소리가 되는 것 등이다. 구어 말뭉치에서는 발음되는 음절 수와 표기상의 음절 수를 맞추는 것이 원칙이므로 축약형의 경우 모두 표기에 반영한다. 그런데 모음의 축약형의 경



우 대부분 현재 국어의 모음 체계상 표기할 글자가 존재하지만, 반홀소리 된 /ㄱ/, /ㄴ/의 표기가 문제가 된다. /ㄱ/, /ㄴ/가 반홀소리가 되어 /ㄱ/, /ㄴ/와 축약되는 현상은 구어에서 자주 나타나는데, 한글의 현재 글자 체계상 이러한 현상을 반영할 방법이 없으므로 구어 전사에서는 '(apostrophe, 영문따옴표)를 사용해서 두 음소를 연결해 준다.

<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 쉼표 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 물

결표(~, 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.
- 이때 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 상대방이 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> 친구와(칭구와) 강남에(간남에) 갔습니다(갔슨니다)

<예> 같이(같이)<note>구개음화 안됨</note> 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 붙여 주지 않는다.

<예> 책상 위에 놔 뒤.

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

<예> [민족에]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,)<note>‘의’로 발음</note>  
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 2, 3과 같이 철자 전사를 통해 이를 반영하기 어려우므로 원래의 표기를 먼저 적고, 학습자의 실제 발음을 ( )에 남겨 표기는 동일하나 해당 발화에 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(성생님)

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>경음화 안 됨</note>

같이(같이)<note>구개음화 안 됨</note>

신라(신라)<note>유음화 안 됨</note>

먹는(먹는)<note>자음동화 안 됨</note>

종고(종고)<note>유기음화 안 됨</note>

앞에(앞에)<note>연음 안 됨</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 ( ) 안에 원어 표기를 적어 밝힌다. 이때 한국어의 음운 체계로 전사가 불가능한 발음은 표기에 반영하지 않는다.

<예> 인터뷰: 영어식 발음으로 [인털류]에 가까운 소리가 남  
센터: 영어식 발음으로 [세너]에 가까운 소리가 남  
파트너: 영어식 발음으로 [팔너]에 가까운 소리가 남

→ 각각 ‘인터뷰(인털류)’, ‘센터(센너)’, ‘파트너(팔너)’로 전사한 후 <note>원어식 발음</note>을 적는다.

- 외국인 학습자 발화의 경우 외래어 또는 외국어 발화 시 원어식의 발음을 하거나 한국어 모어 화자의 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(카페)’, ‘버스(버스)’로 전사한 후 <note>현실발음과 다름</note>을 적는다.

## 2) 방언형 표시

- 확실한 방언형(대응하는 표준형 형태소가 없는 것)의 경우는 다음의 예와 같은 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 타협해야 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 종류 표현을 위한 전사 부호

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

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

6) 익명성 보장을 위한 마크업

- 대화자들의 신분 보장을 위해 이름, 주민등록번호, 카드 번호, 전화번호 등은 노출  
되지 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> : 전화번호

- 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::들  
은 걸 생각해 보면,  
<name2> 사장이 <name2> 회장이 있으니까 보안이 있어도  
눈치 보여서 얘기를 못한다,

- 사람 이름 뒤에 붙는 접사 ‘-이’는 <name> 태그 시, 함께 포함시킨다. 구  
어 전사 후에 수행되는 형태소 분석 시 발생할 수 있는 오류를 최소화시  
키기 위하여 구어 전사 과정에서 접사 ‘-이’를 포함시켜서 <name> 태그  
를 붙이기로 한다.



<예> 학습자의 이름이 ‘민영’일 때  
 학습자 발화: 민영이가, 민영이는, 민영이도, 민영이를  
 → 접사 ‘-이’를 포함시켜서 각각 ‘<name>가’, ‘<name>는’,  
 ‘<name>도’, ‘<name>를’로 태그를 붙인다.

- 대학교 이름은 ‘<name>대학교’로 붙여서 전사한다. 형태소 분석 시 해당 어절을 하나의 고유명사로 처리하기 때문에 팀 간의 작업상 편의를 위해 구어 전사 과정에서 ‘<name>대학교’로 붙이기로 한다. 만약 ‘대학교’에 발화 오류가 발생하면 ( ) 안에 <name>태그를 넣어서 마크업 처리를 한다.

<예> 학습자가 ‘대학교’를 [대학교]로 발음한 경우  
 <name1>대학교(<name1>대학교)

#### 7) 세그멘테이션 분할

- 2015년 전사 작업의 경우 시스템을 전제한 전사라기보다는 학습자 발화의 흐름에 따른 기계적인 전사였다. 따라서 세그멘테이션의 경우 구체적인 지침은 마련하지 않았다. 이는 학습자의 발화에 있어 세그멘테이션의 분할은 모어 화자의 직관에 의하여 이루어지기 때문에 구체적으로 규정을 하기가 매우 모호하기 때문이다. 하지만 시스템을 통한 작업이 이루어지다 보니, 발화 단위의 끊김이 빈번하게 일어날 경우 이를 직관적으로 발화의 의미를 판단하지 못하는 경우가 많았다. 따라서 전사 지침에 독립된 세그멘테이션을 보았을 때 맥락을 보지 않고서도 의미가 판단될 수 있는 단위로 분할한다는 내용을 기재하였다. 즉 세그멘테이션은 어절 단위로 분할한다. 1초 이상의 휴지는 {}라는 기호를 사용하여 그 길이를 기재하기 때문에, 이는 어절 단위에 따라 세그멘테이션을 나누어도 휴지를 알아보는 데에는 문제가 없다. 이는 아래와 같이 정리할 수 있다.

2016년 세그멘테이션 지침 추가 부분
○ 세그멘테이션은 학습자의 발화의 흐름을 반영하여 전사하되 분할의 단위는 어절 단위로 한다.

<예> 5:네.  
내{1.2}가 학교에,/

빗금은 세그멘테이션의 분할을 나타낸다. 아래와 같이 전사  
하지 않는다.

5:네.  
내/가 학교에,/

## 한국어 학습자 말뭉치 형태 주석 지침

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

### I. 학습자 말뭉치의 형태 분석 표지<sup>21)</sup>

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

21) **【수정】** 기존 세종 지침에 있었던 NF(명사추정범주), NV(용언추정범주)를 삭제하고 대  
부분 추정하여 해당 표지로 분석하거나 NA(분석불능범주)로 분석함

	목적격조사	JKO	JKO
	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
	명사형 전성어미	ETN	ETN
	관형형 전성어미	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
	불임표(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 이형태를 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) 표준국어대사전을 따른다.

■■ 중요하다 [중요/NNG+하/XSA+다/EF]  
- 중요 표준국어대사전: 중요/NNG ✓

※ 단, 조사나 어미의 결합형은 지침에 목록으로 제시된 것을 제외하고는 표준국어대사전에 등재되어 있어도 각각 분리해서 분석한다.

■■ 먹었으니까는 [먹/VV+었/EP+으니까/EC+는/JX]  
■■ 가고는 [가/VV+고/EC+는/JX]  
■■ 좋으니만큼 [좋/VA+으니/EC+만큼/JKB]

- 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. <b>교정 어절</b> 이 취할 표지를 준다.	
■■ 문제를 <b>쉬게</b> 풀어요.	[쉬/VA+게/EC]
■■ 너 <b>때문내</b> 죽겠어.	[때문/NNB+내/JKB]
■■ <b>여러까지</b> 문제가 생겼다.	[여러/MM Ⅱ 까지/NNB]
■■ 강에 <b>패수르</b> 버렸다.	[패수/NNG+르/JKO]
2) 형태와 뜻(어휘적 의미+기능적 의미)을 혼동한 경우 → 오류 유형에서는 대치에 해당하는 경우로 <b>보이는 대로</b> 분석한다. <b>오류 어절</b> 만 고려해서 분석한다.	
■■ 내가 <b>고기</b> 가 먹어요.	[고기/NNG+가/JKS]
■■ 입학하자마자 교과서를 <b>팔려고</b> 서점에 갔어요.	[팔/VV+려고/EC]

마. [학습자] 분석 기준의 적용

1) 학습자 말뭉치에서는 종종 문맥에서 전혀 의미를 유추할 수 없는 경우가 나타난다. 이 경우는 주석자가 교정어절을 상정하는 것이 불가능하며, 특정한 형태의 오형태로 추측하기도 어렵다. 이처럼 교정 어절을 상정하기 어려운 경우는 ‘**분석불가능(NA)**’으로 처리한다.

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

→ 표현 문형의 구성과 인접한 경우 교정어절을 상정하기 어렵다 해도 표현 문형 구성에 포함되는 형태까지는 분석을 한다.<sup>22)</sup>

2) 학습자의 오류 어절에서 문맥적 의미의 추측이 가능할 때는, 최소 교정을 원칙으로 교정 어절을 상정해 형태 분석을 한다.

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

가) 상정한 교정어절에 없는 형태가 추가된 경우의 형태 분석은 다음과 같다.

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

\* 교정어절: √ 학교를

■■ 나는 OO <u>학</u> 곡을 갔다.	[학곡/NNG+을/JKO]
■■ 나는 OO <u>학</u> 교 <u>르</u> 갔다.	[학교/NNG+른/JKO]
■■ 나는 OO <u>학</u> 교 <u>고</u> 를 갔다.	[학교/NNG+고/NA+를/JKO]

→ 마지막 예시와 같이 정확한 형태의 체언과 조사가 분리되는데 어절 내에 정체를 알기 힘든 형태가 있을 경우는 해당 형태를 분리해 분석불가능(NA)으로 처리한다.

(2) 용언과 어미 결합의 경우는 용언의 어간 혹은 어근과 어미를 먼저 확보해 분리한 후 잉여적 요소에 대해서는 분석불가능(NA)으로 처리한다.

\* 교정어절: √ 많다

■■ 교실에 학생이 <u>만</u> 다.	[만/VA+다/EF]
■■ 내일은 행사가 더 <u> 많</u> 타.	[많/VA+타/EF]
■■ 오늘은 수업이 <u> 많</u> 나 <u>다</u> .	[많/VA+나/NA+다/EF]

■■ 내 꿈을 <u>이</u> 뤄진 <u>고</u> 나 이루기 위해	[이뤄지/VV+ㄴ/NA+고나/EC]
■■ 꼭 잘해야 <u>되</u> 겠 <u>다</u> .	[되/VV+ㄴ/NA+겠/EP+다/EF]

■■ 교통이 꼭 <u>편</u> 립 <u>습</u> 니다.	[편리/NNG+ㅁ/NA+습니다/EF]
■■ 나는 선생님이 <u>되</u> 겠 <u>다</u> .	[되/VV+ㅁ/NA+겠/EP+다/EF]

■■ 공부를 <u>해</u> 고	[하/VV+아/NA+고/EC]
■■ 7시 30분까지 <u>운</u> 동 <u>했</u> 니다.	[운동/NNG+하/XSV+아/NA+ㅁ니다/EF]
■■ 꽃 가게 주인 <u>돼</u> 면	[되/VV+어/NA+면/EC]
■■ 식사가 <u>준</u> 비 <u>되</u> 고 있었다.	[준비/NNG+되/XSV+어/NA+고/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> 은 <u>다</u> .	[많/VA+은/NA+다/EF]
■■ 교실에 사람이 <u> 많</u> 는 <u>다</u> .	[많/VA+는다/EF]

나) 용언의 활용과 관련한 오류의 형태 분석은 다음과 같다.

(1) 학습자의 오류가 활용 규칙과 직접적인 관련이 없는 경우는, 기존 형태 분석 지침대로 어간을 복원해 형태를 분석한다.

- |                             |                     |
|-----------------------------|---------------------|
| ■■ 한국말이 너무 <u>아렵다</u> .     | [아렵/VA+다/EF]        |
| ■■ 저녁식사도 준비하기가 <u>번거러워서</u> | [번거/XR+럽/XSA+어서/EC] |
| ■■ 공간에 <u>해로원데</u>          | [해롭/VA+어/NA+ㄴ데/EC]  |
| ■■ 아 <u>크다름</u> 꿈이네요.       | [크다름/VA+ㄴ/ETM]      |

(2) 학습자의 오류가 활용 규칙과 직접적인 관련이 있는 경우, 기존의 형태 분석과 달리 어간을 복원하지 않고 오류 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]  |
| ■■ 내가 <u>게으르서</u>                | [게으러/VV+어서/EC] |
| ■■ 우리 나라하고 <u>다라서</u> 싫어지만       | [다라/VA+아서/EC]  |

(3) 학습자가 활용 규칙을 몰라서 어미를 제대로 선택하지 못한 경우, 어간을 복원하지 않고 잘못된 어간 형태를 살려 분석한다.

- |                   |               |
|-------------------|---------------|
| ■■ 공부가 <u>힘드는</u> | [힘드/VA+는/ETM] |
|-------------------|---------------|

(4) 학습자가 활용 규칙은 알지만 형태소 결합 규칙을 몰라서 잘못된 어절을 만든 경우에는 어간을 복원해서 분석한다.

- |                       |                           |
|-----------------------|---------------------------|
| ■■ 공부가 <u>힘든지만</u>    | [힘들/VA+ㄴ/NA+지만/EC]        |
| ■■ 공부가 <u>힘든입니다</u> . | [힘들/VA+ㄴ/NA+이/VCP+ㅂ니다/EF] |

(5) 다음과 같은 활용 오류형이 나타날 경우 어미를 누락한 오류로 보고 어간의 형태만 분석한다.

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

(6) 학습자가 용언의 활용에서 동일한 형태(혹은 이형태)를 중복해서 사용한



경우는 같은 형태 표지로 분석한다.

■■ 풍선이 <u>부풀</u> 는 것이	[부풀/VV+ㄴ/ETM+는/ETM]
■■ <u>다른</u> 는 게 되게 많이	[다르/VA+ㄴ/ETM+는/ETM]
■■ <u>불평등한</u> 는 생각도 많아요.	[불/XPN+평등/NNG+하/XSA+ㄴ/ETM+는/ETM]
■■ 이권엔 <u>지나지</u> 지 않았습니 <del>다</del>	[지나/VV+지/EC+지/EC]

다) 형태소 경계를 분할하기 어려운 오류 어절의 형태 분석은 다음과 같다.

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

■■ 내 유학생활을 <u>아프로</u> 미래에게	[아/NNG+프로/JKB]
■■ 다언에도 소개해 <u>주게서</u> 요.	[주/VX+게/EP+서요/EF]
■■ <u>어려</u> 슬 데 가을에 좋은 기억이	[어리/VA+어/EP+슬/ETM]
■■ <u>자시</u> 느이 마음대로 했다.	[자시/NNG+느이/JKG]

(2) 오류의 형태가 형태 단계에서도 표시될 수 있도록 가능한 경우 자모 단위로도 형태를 분할한다.

■■ 정말 신기하다고 <u>생각</u> 했다.	[생각/NNG+하/XSV+ㅅ/EP+다/EF]
■■ 어제 영화를 <u>봔</u> 는데	[보/VV+ㅅ/EP+는데/EC]

(3) 기본적으로는 분석한 형태의 결합이 원 어절이 되도록 분할해야 하지만 그 경계 분할이 어려운 오류 어절의 경우, 어간에 학습자가 잘못 쓴 오형태를 주며 어미는 해당 오류 어절이 취해야 하는 어미의 형태를 부여한다.

■■ 쓰레기를 <u>버려</u> 도 되면 좋겠습니다.	[버러/VV+어도/EC]
■■ 많은 친구를 <u>사귀</u> 서 재밌었어요.	[사구/VV+어서/EC]
■■ 소치에 2시간 <u>걸</u> 레요.	[걸레/VV+어요/EF]
■■ 점점 <u>심</u> 해질 것이다.	[심해/VA+어/EC+지/VX+ㄹ/ETM]
■■ A의 주장에 <u>반</u> 에	[반에/VV+어/EC]
■■ 제 꿈을 <u>위</u> 에	[위에/VV+어/EC]
■■ 제 꿈을 <u>위</u> 에서	[위에/VV+어서/EC]

라) 다음의 축약형에서 나타나는 오류형은 이후에 이뤄질 오류 주석을 고려해 형태를 분할하지 않고 오형태로만 분석한다.

■■ 내일은 네 생일이라서 소포를 받았 <del>다</del> .	[네/NP]
--------------------------------------	--------

■ ■ 재 장소 중에서 제일 좋아하는 곳은

[재/NP]

■ ■ 세 아버지는 키가 큼니다.

[세/NP]

■ ■ 또 궁금한 게 있으면

[개/NNB]

■ ■ 저는 OO어학당 6급 학생이에요.

[학생/NNG+이/VCP+예요/EF]

■ ■ 나는 학생이었다.

[학생/NNG+이/VCP+였/EP+다/EF]

### 3. 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]<sup>23)</sup>

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

■■ 그것은 “는”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.

23) [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||웅/NNB,  
케네디/NNP||씨/NNB<sup>24)</sup>, 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

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

(라) 사람 이름의 뒤에 접사 ‘-이’가 붙는 경우는 이름과 함께 하나의 단위로 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, 인창동/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/NNG, 조선조/NNG

(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/NNG||차이/NNG
- 남북/NNP||적십자회담/NNG
- 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 미분리 (NN 구성 포함)	사전 등재	전체 NNP	(책) 삼국사기/NNP, 손자병법/NNP
	사전 미등재	전체 NNP	(드라마) 전원일기/NNP, 가을동화/NNP (영화) 어벤저스/NNP, 쿵푸팬더3/NNP (TV프로그램) 런닝맨/NNP, 가족오락관/NNP
어절 분리	사전 등재	나누어 분석	(책) 안네/NNP+의/JKG    일기/NNG
	사전 미등재	나누어 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 나뉠 수 있으나, 본 분석에서는 이를 세분화하지 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 <u>걸</u> 가져왔니?	[것/NNB+ㄹ/JKO]
■■ 연습할 <u>건</u> 있니?	[것/NNB+ㄴ/JX]
■■ 먹을 <u>게</u> 모자르다	[것/NNB+이/JKS]

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

■■ 밥을 먹을 <u>건</u> 다.	[거/NNB+이/VCP+ㄴ 다/EF]
----------------------	----------------------

※ [보완] 학습자의 오류로 인해 축약된 ‘게’의 형태가 유지되지 않는 경우는, 분할하지 않고 분석한다.

■■ 또 궁금한 <u>개</u> 있으면	[개/NNB]
-----------------------	---------

## 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 <u>그</u> 걸 믿어 왔단다.	[그것/NP+ㄹ/JKO]

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

■■ 내	이제부터는 <u>내</u> 명령을 따라라.	[나/NP+의/JKG]
■■ 내게	<u>내</u> 게 전자우편으로 알려 다오.	[나/NP+에게/JKB]
■■ 네게	어제 <u>네</u> 게 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) ‘뉘’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]

※ [보완] 다음의 대명사와 조사의 축약형에서 나타나는 오류의 경우 원형을 밝혀 분석하지 않고 오형태로 분석한다.

■ ■ 내일은 네 생일이라서 소포를 받았다.	[네/NP]
■ ■ 재 장소 중에서 제일 좋아하는 곳은	[재/NP]
■ ■ 세 아버지는 키가 큼니다.	[세/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, 백한들, 두서넛, 서넛, 너댓, 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/MMⅡ번/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 **수사**로 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, 너
- 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 본용언일 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	꼬치꼬치 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.<sup>25)</sup>

■■■ <u>철수만 한</u> 인재가 없다	[철수/NNP+만/JX    하/VV+ㄴ/ETM]
■■■ 이 음식은 먹을 <u>만하다</u> .	[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 <u>버릇하다</u> .	[버릇/NNG+하/XSV+다/EF+./SF]
-------------------------	--------------------------

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

■■■ 열심히 <u>공부하도록</u> 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.

---

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

- 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]

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 OO여학당 6급 학생이예요. [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 알리오. [이내/MM]  
 ■■ 아침 들판에 이내가 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/VA + 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.

건테, 고로01 「2」, 그래서, 그러나, 그러니까, 그러면, 그러므로, 그런데, 그럼01 「1」, 그렇지마는, 그렇지만, 그리고, 그리하여, 근테01, 단06, 따라서, 연이나, 연중에, 연즉, 이리하여, 하건만, 하기는, 하기가, 하긴, 하물며, 하지만, 한테03

※ 용언의 활용형

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

※ 일반 부사

- 게다가, 곧, 다만, 또, 또는, 또한, 밋, 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 동일한 형태로 구분이 어려운 경우가 있다. 이들은 뒤에 조사가 결합하느냐 여부와, 문맥에서 후행 명사를 수식하느냐의 여부에 따라 부사와 명사로 분석될 수 있다.

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

- ② 부사적인 용법을 가졌음에도 불구하고 일반부사가 아닌 일반명사로만 표준국어대사전에 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 <u>덜달아</u> 왜 난리니?	[덜달/VV+아/EC]
■■ 우리 함께 <u>더불어</u> 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, 흉내를 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. 관계언<sup>26)</sup>

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

### 1) 격조사(JK)

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

#### 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	부대장님께오서	[부대장/NNG+님/XSN+께오서/JKS]
■■ 께옵서	황제께옵서 드나드신다.	[황제/NNG+께옵서/JKS]

---

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

■■ 닭이가 울었다.	[닭/NNG+이/JKS+가/JKS]
■■ 책상이가 있다.	[책상/NNG+이/JKS+가/JKS]

---

26) 지침에 제시된 조사 목록에서 빠진 이형태와 예시를 추가함

■■ 친구들이 6월에 일이를 찾아있었다.	[일/NNG+이/JKS+를/JKO]
■■ 좋아하는 거는 옷입니다.	[옷/NNG+이/JKS+이/VCP+ㅂ니다/EF]
■■ <u>어른이들</u> 이 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]
■■ 로써/으로써	장관으로써 책임을 다해야 한다.	[장관/NNG+으로써/JKB]
	돌로써 지붕을 만든다고?	[돌/NNG+로써/JKB]
■■ 같이	콩으로써 메주를 쏜다고 해도	[콩/NNG+으로써/JKB]
■■ 더러	바보같이 웃고 다닌다.	[바보/NNG+같이/JKB]
■■ 랑/이랑	나더러 이것도 하라고 한다.	[나/NP+더러/JKB]
	너랑 많이 닮았다.	[너/NP+랑/JKB]
	오늘 동생이랑 싸웠다.	[동생/NNG+이랑/JKB]
■■로부터/ 으로부터	TV로부터 받는 영향력이	[TV/SL+로부터/JKB]
■■ 마냥	시험으로부터 해방되다	[시험/NNG+으로부터/JKB]
■■ 마따나	기영이마냥 놀 수만은 없다.	[기영이/NNP+마냥/JKB]
■■ 만큼	네 말마따나 나도 그래야 한다.	[말/NNG+마따나/JKB]
■■ 보고	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/VC+야/EF+./SF]

## 사) 인용격조사(JKQ)

- 인용문이나 인용구를, 동사에 대한 부사적 성분으로 도입하는 조사이다.

■■ 고	그는 "이제 가도 좋다"고 말했다.	[좋/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 닫았다.	[가/VV+자/EF+ "/SS+하고/JKQ]

## <주의사항>27)

- ① 인용격조사는 연결어미와 구별하기 어려운 경우가 있으므로 주의한다. 인용 기호가 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.  
[좋/VA+다/EF+ "/SS+고/JKQ]
- 먼저 "주민등록증이 있냐?"고 묻는다.  
[있/VV+냐/EF+?/SF+ "/SS+고/JKQ]

---

27) <세종> 분석 결과를 바탕으로 다시 정리함

## (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]

### ※ [참고]

■■ 시골 아이라고 그것도 모르겠니?

[아이/NNG+라고/JX]

- 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]

## 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+이/고/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	잔치상에는 배며 대추며 여러 가지 과일이 차려 져 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]

- 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.

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

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

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

- 친구에게 줄 꽃과 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.

■ 그러/그래	좋습니다그러.	[좋/VA+습니다/EF+그러/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.	[그것/NP+이나/JX]
■ 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 건졌다	[빨리/MAG+ㄴ커녕/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. 사람이란 분수를 지킬 줄 알아야 한다.	[코알라/NNG+란/JX] [사람/NNG+이란/JX]
■■ ㄹ랑/일랑	강엘랑 가지 마라. 그 일에 대해선 걱정일랑 하지 말아라.	[강/NNG+에/JKB+ㄹ랑/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. 남이사 무슨 상관이야.	[내/NP+사/JX] [남/NNG+이사/JX]
■■ 야/이야	그야 그렇지. 그가 인간성이야 그만이지.	[그/NP+야/JX] [인간성/NNG+이야/JX]
■■ 야말로/이야말로	사과야말로 가을의 과일이다. 통일이야말로 최대의 과업이지.	[사과/NNG+야말로/JX] [통일/NNG+이야말로/JX]
■■ 요	나는 그림을요 잘 그립니다.	[그림/NNG+을/JKO+요/JX]
■■ 조차	이젠 봄조차 빼앗기는구나.	[봄/NNG+조차/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)ㄹ랑, (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.<sup>28)</sup>

■■ 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]
	코알라란 매우 귀여운 동물이다.	[코알라/NNG+란/JX]
■■ (이)나	밥이나 빵을 먹도록 해라.	[밥/NNG+이나/JC]
	그가 비록 열심히 하나 능력은 부족하다.	[하/VV+나/EC]
	어제 내가 술을 마셨나?	[마시/VV+었/EP+나/EF+?/SF]

28) [삭제] 말고

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

■■ (이)야	철수야 그 일을 할 수 있지.	[철수/NNP+야/JX]
	내가 좋아하는 것은 철수야.	[철수/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 오셨었어.

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)<sup>29)</sup> A: 선생님에 대해 알고 있니?

B: 선생님이요? [선생/NNG+님/XSN+이요/JX]

A: 가장 좋아하는 명절이 언제예요?

B: 설날이요 [설날/NNG+이요/JX]

(마) ‘말고’는 용언 ‘말다’의 활용형으로 처리한다.

■■ 돈말고 지혜가 필요하다. [돈/NNG+말/VV+고/EC]

29) 본 지침에서는 세종 현대 구어 말뭉치의 분석 결과를 따라 (2)와 같은 경우를 ‘이요/JX’로 분석한다.

## 바. 의존형태

### 1) 어미<sup>30)</sup>

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

30) 지침에 제시된 어미 목록에서 빠진 이형태와 예시를 추가함

■■ -(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 멋지구나!	[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]
■■ -구면/는구면	학교가 참 크구면.	[크/VA+구면/EF+./SF]
	공부를 잘하는구면.	[잘/MAG+하/XSV+는구면/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]
	키가 얼마나 크냐?	[크/VA+냐/EF+./SF]
■■ -냐/으냐/느냐	물이 얼마나 깊으냐?	[깊/VA+으냐/EF+?/SF]
	그것보다 이것이 낫느냐?	[낫/VA+느냐/EF+?/SF]
■■ -냐고/으냐고 /느냐고	그가 누구냐고?	[누구/NP+이/VCP+냐고/EF+?/SF]
	그렇게 싫어? 싫으냐고?	[싫/VA+으냐고/EF+?/SF]
	너 뭐 해? 뭐 하느냐고?	[하/VV+느냐고/EF+?/SF]
■■ -네	정말 큰일 났네!	[나/VV+았/EP+네/EF+!/SF]
■■ -니	그게 없니?	[없/VA+니/EF+?/SF]

■■ -다/ㄴ다/는다	그게 사실이다. 이건 말도 안 된다. 아이가 글을 잘 읽는다.	[사실/NNG+이/VCP+다/EF+./SF] [되/VV+ㄴ다/EF+./SF] [읽/VV+는다/EF+./SF]
■■ -다구/ㄴ다구 /는다구	돈이 많다구? 너도 간다구? 소설책을 읽는다구?	[많/VA+다구/EF+?/SF] [가/VV+ㄴ다구/EF+?/SF] [읽/VV+는다구/EF+?/SF]
■■ -다나/ㄴ다나 /는다나	그도 가겠다나. 나를 잘 안다나. 건강한 여자를 찾는다나.	[가/VV+겠/EP+다나/EF+./SF] [알/VV+ㄴ다나/EF+./SF] [찾/VV+는다나/EF+./SF]
■■ -다네/ㄴ다네 /는다네	일을 망쳤다네 우리네 짧은 인생도 간다네. 평소에도 한복을 잘 입는다네.	[망치/VV+었/EP+다네/EF+./SF] [가/VV+ㄴ다네/EF+./SF] [입/VV+는다네/EF+./SF]
■■ -다니까/ㄴ다니까 /는다니까	돈이 없단니까! 어머니가 오늘은 꼭 오신다니까. 내 말을 믿지를 않는다니까.	[없/VA+다니까/EF+!/SF] [오/VV+시/EP+ㄴ다니까/EF+!/SF] [않/VX+는다니까/EF+./SF]
■■ -다니/ㄴ다니 /는다니	서울이 이렇게 변화하다니. 이 긴 시를 어떻게 외운다니? 이 많은 책을 언제 읽는다니?	[변화/XR+하/XSV+다니/EF+?/SF] [외우/VV+ㄴ다니/EF+?/SF] [읽/VV+는다니/EF+?/SF]
■■ -다면서/ㄴ다면서 /는다면서	술은 싫다면서? 니가 축구를 잘한다면서? 달팽이도 먹는다면서?	[싫/VA+다면서/EF+?/SF] [잘/MAG+하/XSV+ㄴ다면서/EF+?/SF] [먹/VV+는다면서/EF+?/SF]
■■ -다오/ㄴ다오 /는다오	그가 가지고 있다오. 꽃은 이른 봄에 핀다오. 이 나무는 열매를 많이 맺는다오.	[있/VX+다오/EF+./SF] [피/VV+ㄴ다오/EF+./SF] [맺/VV+는다오/EF+./SF]
■■ -단다/ㄴ단다 /는다단다	나도 슬프단다. 선생님께서 공부를 가르쳐 주신단다. 누에는 뽕잎을 먹는단다.	[슬프/VA+단다/EF+./SF] [주/VX+시/EP+ㄴ단다/EF+./SF] [먹/VV+는다단다/EF+./SF]
■■ -도다/는다다	꽃이 아름답도다. 짐이 조서를 내리는도다.	[아름답/VA+도다/EF+./SF] [내리/VV+는다다/EF+./SF]
■■ -ㄹ걸/을걸	모른다고 할걸. 생각만큼 쉽지 않을걸.	[하/VV+ㄹ걸/EF+./SF] [않/VX+을걸/EF+./SF]
■■ -ㄹ게/을게	그렇게 할게. 남은 밥은 내가 먹을게.	[하/VV+ㄹ게/EF+./SF] [먹/VV+을게/EF+./SF]
■■ -ㄹ까/을까	이제 밥을 할까? 이 과자는 내가 먹을까?	[하/VV+ㄹ까/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.	[아니/VCN+라니까/EF+./SF]

	가만히 있으라니까.	[있/VV+으라니까/EF+./SF]
■■ -ㅁ세/음세	그날 꼭 음세.	[오/VV+ㅁ세/EF+./SF]
	곧 밥을 먹음세.	[먹/VV+음세/EF+./SF]
■■ -ㅂ니까/습니까	이제야 옵니까?	[오/VV+ㅂ니까/EF+?/SF]
	그래도 되겠습니까?	[되/VV+겠/EP+습니까/EF+?/SF]
■■ -ㅂ니다/습니다	이렇게 합니다.	[하/VV+ㅂ니다/EF+./SF]
	정말 재미있습니다.	[재미있/VA+습니다/EF+./SF]
■■ -ㅂ시다/읍시다	다시 만납시다.	[만나/VV+ㅂ시다/EF+./SF]
	여기 앉읍시다.	[앉/VV+읍시다/EF+./SF]
■■ -ㅂ시오/읍시오	서둘러 주십시오.	[주/VX+시/EP+ㅂ시오/EF+./SF]
	여기 앉읍시오.	[앉/VV+읍시오/EF+./SF]
■■ -ㅂ디까/습디까	신부가 예쁘디까?	[예쁘/VA+ㅂ디까/EF+?/SF]
	보기에 좋습디까?	[좋/VA+습디까/EF+?/SF]
■■ -ㅂ디다/습디다	참 좋은 곳입디다.	[곳/NNB+이/VCP+ㅂ디다/EF+./SF]
	덕수궁에 사람이 많습디다	[많/VA+습디다/EF+./SF]
■■ -세/으세	제대로 좀 하세.	[하/VV+세/EF+./SF]
	이 책을 우리 함께 읽으세.	[읽/VV+으세/EF+./SF]
	함께 가.	[가/VV+아/EF+./SF]
■■ -아/어/여	밥 먹어!	[먹/VV+어/EF+!/SF]
	같이 해.	[하/VV+아/EF+./SF]
■■ -야	그건 사실이 아니야.	[아니/VCN+야/EF]
■■ -아라/어라	웃기지 말아라.	[말/VX+아라/EF+./SF]
	천천히 먹어라.	[먹/VV+어라/EF+./SF]
	물이 깨끗하오.	[깨끗/XR+하/XSA+오/EF+./SF]
■■ -오/으오/소	나는 요즘 논어를 읽으오.	[읽/VV+으오/EF+./SF]
	그 곳에는 내가 가겠소.	[가/VV+겠/EP+소/EF+./SF]
■■ -자	잠이나 자자.	[자/VV+자/EF+./SF]
■■ -자꾸나	약속을 좀 미루자꾸나.	[미루/VV+자꾸나/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 <u>했는걸요</u> .	[하/VV+았/EP+는걸요/EF+./SF]
■■ 뭐 <u>먹었는데요?</u>	[먹/VV+었/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 경우는 <표준>을 따라 종결어미로 분석한다.

■■■ 아픈데 밥을 먹을까 싶다. [먹/VV+을까/EF]

■■■ 진짜 부자유친이 아닐까 생각합니다. [아니/VCN+ㄹ까/EF]

■■■ 돈을 어떻게 쓰느냐에 따라 [쓰/VV+느냐/EF+에/JKB]

■■■ 무슨 일이 있었는가 했다. [있/VV+었/EP+는가/EF]

■■■ 보통 사용할 때는 뭔가 게임을 할 때 [뭐/NP+이/VCP+ㄴ가/EF]

■■■ 언제 고향에 갈지 잘 모르겠습니다. [가/VV+ㄹ지/EF]

■■■ 왜 좋았는지 알아요. [좋/VA+았/EP+ㄴ지/EF]

■■■ 자기 적성에 맞는지 안 맞는지 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]
	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]
■■ -ㄴ 데/은데/ 는데	예쁜데 미워한다. 방이 좁은데 가구는 많다. 눈이 오는데 차를 가져가지 말까?	[예쁘/VV+ㄴ 데/EC] [좁/VV+은데/EC] [오/VV+는데/EC]
■■ -ㄴ 들/는들	간다 한들 아주 같까? 그걸 먹는들 뭐가 달라지겠나.	[하/VV+ㄴ 들/EC] [먹/VV+는들/EC]
■■ -ㄴ 즉/은즉	배가 고평즉 속이 쓰리다. 물이 맑은즉 고기가 많기는 어렵소.	[고프/VV+ㄴ 즉/EC] [맑/VV+은즉/EC]
■■ -ㄴ 지라/은지라/ 는지라	눈이 온지라 길이 미끄럽다. 기분이 좋은지라 다정하다. 선생님께서 고집을 굽히지 않으시는지라	[오/VV+ㄴ 지라/EC] [좋/VV+은지라/EC] [않/VX+으시/EP+는지라/EC]
■■ -나/으나	눈이 오나 비가 오나 밥을 먹으나 마나이다.	[오/VV+나/EC] [먹/VV+으나/EC]
■■ -나니	멀리 보이나니 넓은 들이로다.	[보이/VV+나니/EC]
■■ -나마/으나마	도와주지는 못하나마 방해를 해서는 맛은 없으나마 많이 드세요.	[못하/VX+나마/EC] [없/VV+으나마/EC]
■■ -노니	물노니, 포부가 무엇이냐?	[물/VV+노니/EC+,,/SP]
■■ -니/으니	밥을 다 먹고 보니 배가 불렀다. 이 옷은 작으니 큰 것으로 바꿔 주세요.	[보/VX+니/EC] [작/VV+으니/EC]
■■ -느니	앉아서 걱정하느니 나가서 하겠다.	[걱정/NNG+하/XSV+느니/EC]
■■ -니까/으니까	웃기니까 좋다. 약속을 했으니까 만나야 한다.	[웃기/VV+니까/EC] [하/VV+았/EP+으니까/EC]
■■ -다가	자랑하다가 망신당했다.	[자랑/NNG+하/XSV+다가/EC]
■■ -다기에/ㄴ 다기에 /는다기에	그녀가 예쁘다기에 보러 왔소. 앞으로 잘 한다기에 승낙했다. 빵을 먹는다기에 주었다.	[예쁘/VV+다기에/EC] [하/VV+ㄴ 다기에/EC] [먹/VV+는다기에/EC]
■■ -다손/ㄴ 다손 /는다손	밑다손 치더라도 구박하지 말자. 그가 제시간에 온다손 하더라도 내 앞의 음식은 다 먹는다손 치더라도	[밑/VV+다손/EC] [오/VV+ㄴ 다손/EC] [먹/VV+는다손/EC]
■■ -대도/ㄴ 대도 /는대도	시간이 있대도 만나 주질 않는다. 늦으면 큰일 난대도 서두르질 않아요. 떠들면 야단맞는대도 계속 떠들었다.	[있/VV+대도/EC] [나/VV+ㄴ 대도/EC] [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]
■■ -ㄹ뿐더러/ 을뿐더러	비가 올뿐더러 바람도 분다. 그는 재산이 많을뿐더러 재능도 많다	[오/VV+ㄹ뿐더러/EC] [많/VV+을뿐더러/EC]
■■ -ㄹ수록/ 을수록	갈수록 태산이다. 이 책은 읽을수록 감동을 준다. 비가 얼마나 올지 천둥이 다 친다.	[가/VV+ㄹ수록/EC] [읽/VV+을수록/EC] [오/VV+ㄹ지/EC]
■■ -ㄹ지/올지	내일은 얼마나 날씨가 좋을지 오늘 밤하늘에 별이 유난히 빛난다.	[좋/VV+을지/EC]
■■ -ㄹ지라도/ 을지라도	이길지라도 명예롭지는 않다.	[이기/VV+ㄹ지라도/EC]
■■ -ㄹ지언정/ 을지언정	마음에 걱정이 있을지라도 내색하지 마라. 그것은 무모한 행동일지언정 죽을지언정 그 일은 못하겠다.	[있/VV+을지라도/EC] [행동/NNG+이/VCP+ㄹ지언정/EC] [죽/VV+을지언정/EC]
■■ -라고	바보라고 생각한다.	[바보/NNG+이/VCP+라고/EC]
■■ -락	오르락 내리락	[오르/VV+락/EC]
■■ -랍시고	그는 반장이랍시고 행패만 부린다.	[반장/NNG+이/VCP+랍시고/EC]
■■ -러/으러	청소하러 가자. 점심 먹으러 집에 간다.	[청소/NNG+하/XSV+러/EC] [먹/VV+으러/EC]
■■ -려/으려	학교에 가려 한다. 웃으려 한다.	[가/VV+려/EC] [웃/VV+으려/EC]
■■ -려니와/ 으려니와	비용도 문제려니와 일꾼도 문제다. 이 마을은 경치도 좋으려니와	[문제/NNG+이/VCP+려니와/EC] [좋/VV+으려니와/EC]
■■ -련마는/ 으련마는	보면 반가우련마는 볼 수가 없네. 별써 제 잘못을 알았으련마는 노래하며 춤을 춘다.	[반갑/VV+련마는/EC] [알/VV+았/EP+으련마는/EC] [노래/NNG+하/XSV+며/EC]
■■ -며/으며	강물이 맑으며 깊다.	[맑/VV+으며/EC]
■■ -면/으면	지옥이 존재하면 만원일 것이다. 내일 날씨가 좋으면 소풍을 가겠다.	[존재/NNG+하/XSV+면/EC] [좋/VV+으면/EC]
■■ -면서/으면서	푸르면서 검은 물빛 밥을 먹으면서 신문을 본다.	[푸르/VV+면서/EC] [먹/VV+으면서/EC]
■■ -므로/으므로	비가 오므로 가지 않겠다. 강이 깊으므로 배 없이 건널 수 없다.	[오/VV+므로/EC] [깊/VV+으므로/EC]
■■ -아/어	입을 막아 버렸다. 밥을 먹어 버렸다.	[막/VV+아/EC] [먹/VV+어/EC]
■■ -아도/어도	암만 봐도 모르겠다. 나는 부자가 아니어도 행복하다.	[보/VV+아도/EC] [아니/VCN+어도/EC]
■■ -아서/어서	덧을 놓아서 꿩을 잡았다. 그는 걸어서 학교에 갔다.	[놓/VV+아서/EC] [걸/VV+어서/EC]
■■ -아야/어야	이 일은 잘해야 한다. 사람은 먹어야 산다.	[잘/MAG+하/XSV+아야/EC] [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 <u>데</u> 가 어디지?	[있/VV+는/ETMⅡ데/NNB+가/JKS]
■■ 다들 <u>만족</u> 하는지 아무런 불평이 없다.	[만족/NNG+하/XSV+는지/EC]
■■ 너를 <u>만난</u> <u>지도</u> 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/NNG+하/XSA+기/ETN]
■■ -ㅁ/-음	학생 신분임을 밝히다.	[신분/NNG+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 명사형이나 아니면 굳어진 명사이냐 하는 것은 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 분석을하기로 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, 이들을 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 연

구자에 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, 댁(宅)-청주댁, 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)-복고풍, 화(化)-도구화, 기-기름기, 께-10분께, 꿀-십 원꿀, 끼리-전우끼리, 궂-노름궂, 네-동이네, 님-선생님, 들-우리들, 들이-1ㄹ들이, 배기-열 살배기, 뽕-조카뽕, 씹-만원씹, 장이-간판장이, 쟁이-심술쟁이, 쯤 -내일쯤, 질-서방질, 짜리-백 원짜리, 째1 -이틀째, 째2-옹기째, 치레-인사치레, 투성이-먼지투성이

## <주의사항>

(가) 명사과생접미사인 ‘-들’은 그 분포가 매우 다양하여 일부에서는 이를 보조사와 접미사로 나누어 분석하기도 한다. 그러나, 본 분석에서는 이들을 모두 명사과생접미사로 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 사용됐다면 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 동사파생 접미사로 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 벉락치기 공부 하는 것을 [공부/NNG || 하/VV+는/ETM]

■■ 카페에서 벉락치기 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 <u>먹</u> 하다.	[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+ㄴ 다면/EC]

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.

■■ 따듯도 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.

■■ **듬직하다** [듬직/XR+하/XSA+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+ㄴ 다/EF]

■■ **음식 맛이 저번보다 못하다.** [못하/VA+다/EF]

■■ **밥을 먹지 못한다.** [못하/VX+ㄴ 다/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>쪼매턴게</u> 하마 자라서 빠나?	[쪼매턴게/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 등재되며, 이들은 앞뒤에 함께 쓰인 명사와 합쳐서 명사로 분석한다. 이들의 목록은 다음과 같다.

가01 「04」	(일부 명사 뒤에 붙어) ‘주변4’의 뜻을 나타내는 말.	강가 //넷가 //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 //떨감 //양념감 //안춧감 //장난감//웃음감//사형감//노벨상감//마느질감.
값 「07」	(일부 명사 뒤에 붙어) ‘가격’, ‘대금’, ‘비용’의 뜻을 나타내는 말.	기름값 //물값 //물건값 //부식값 //신문값 //우옷값 //음식값.
값 「08」	(일부 명사 뒤에 붙어) ‘수치’의 뜻을 나타내는 말.	변숫값//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//사극//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//해외 토픽난.
놀이01 「04」	(일부 명사 뒤에 붙어) ‘모방4’, ‘장난’, ‘흉내’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 일으키는 사람이나 사물을 나타내는 말).	꿀칫덩어리 //심술덩어리 //애꿎덩어리//제주덩어리.
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.	곡명//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//꿀 뱅이무침//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.

	나타내는 말.	
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) ‘안건’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.	옥제떨이 //옥매트 //옥침대.
왜03 「03」	(일부 명사 앞에 붙어) ‘일본식의’, ‘일본의’의 뜻을 나타내는 말.	왜간장 //왜모시.
은04 「02」	(일부 명사 앞에 붙어) ‘은색’, ‘은제3’의 뜻을 나타내는 말.	은갈치 //은귀고리 //은목걸이 //은찰잔.
자08 「03」	(일부 명사 앞에 붙어) 모체에 딸려 있음을 나타내는 말.	자회사.
잡이01 「04」	(일부 명사 뒤에 붙어) 민속놀이나 전통 음악에서 기술이나 재주, 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.
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
즙 「02」	(먹을 것을 나타내는 일부 명사 뒤에 붙어) ‘농축액’을 나타내는 말.	미나리즙 //석류즙 //배즙 //양파즙 //쥬즙.
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…….”
찜01 「02」	(일부 명사 뒤에 붙어) 찜 음식의 뜻을 나타내는 말.	갈비찜 //아귀찜.
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

### ※ 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 종결어미를 대부분 연결어미로 분석하는데, 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 주의해야 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>쥘</u> 줘는데	[주/VX+L/ETM    주/VX+어/EC]

## 2) 끊어진 어절

- 끊어진 어절은 어절의 일부만 발화된 경우나 불분명한 경우이다. 끊어진

채로 발화된 어절은 형태 단위가 온전히 발화되었을 때만 분석하고, 형태 단위가 온전히 발화되지 못해 형태소가 확보되지 못한 경우는 분석불가능(NA)으로 처리한다.

#### 가) 어절의 일부만 발화되었지만 분석이 가능한 경우

■■ 어 손의 관절에 통증= 통증이	[통증/NNG]
■■ 미국과 같= 같은	[같/VA]
■■ 정말 아름다우= 아름다운 곳이에요.	[아름다우/VA]
■■ 그랬= 그랬어요.	[그러/VV+었/EP]
■■ 비결인=인 것 같은데.	[이/VCP+ㄴ/ETM]

#### 나) 어절의 일부만 발화되어 분석이 불가능하거나 중의성이 발생하는 경우

■■ 음 플라스 될 수 아이다, 마이너= 아, 플라스는	[마이너/NA]
■■ 어 운동할 시= 힘도 부족해서	[시/NA]
■■ 어제 약= 약= 약국에 갔어요	[약/NA  약/NA]
■■ 한국인과 가= 같은	[가/NA]
■■ 의사 선생님도 곧 나= 나올 수 있다고	[나/NA]
■■ 경복궁에 가 봐= 봤다.	[보/VV+아/NA]
■■ 스페인에서 있어= 있어요.	[있/VV+어/NA]
■■ 슬퍼=퍼 가지구	[퍼/NA]

#### [보완] <주의사항>

- 합성어와 파생어 앞에서 이들 어휘의 일부가 끊어진 채로 발화된 경우 형태 단위가 온전해 분석이 가능한 경우는 분석하고 그렇지 못한 경우는 NA 표지를 부여한다.

■■ 많= 많이 주세요	[많/VA]
■■ 일본인 학습= 학습자들이	[학습/NNG]
■■ 일상= 일상생활에서 주로 공부를 해요	[일상/NNG]
■■ 한국 음식을 좋아= 좋아하지만	[좋/VA+아/EC]
한국 음식을 좋아하= 좋아하지만	[좋아하/VV]
■■ 한국어 공부가 힘= 힘들었지만	[힘/NNG]



### 다) 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]

### 라) 어절 중간에 간투사 따위가 들어가는 경우

- 심약 어 하다 [심약/NNG || 어/IC || 하/VV+다/EF]  
■■ 좋아 어 하다 [좋/VV+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/VV+아/EF]  
라고 대답했지. [라고/JKQ]  
■■ 주부 우울증. [우울증/NNG]  
이라고 말할 수 있겠습니까. [이/VCP+라고/EC]  
■■ 공부. [공부/NNG]  
한다고 [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. [어/NA]  
제는 별일 없었어. [제/NA+는/JX]

■■ 선두주  
자가 도착했다.

[선두/NNG || 주/NA]  
[자/NA+가/JKS]

#### 4) 불분명한 어절 (X)

- 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 쥘 봐.	[그/MM    책/NNG]
■■ 그 무엇인가를 알아내고자 했지만	[그/MM  무엇/NP+이/VCP+ㄴ가/EF+를/JKO]
■■ 그 왜 있잖아요.	[그/IC]

■■ 이도 저도 다 싫다.	[저/NP+도/JX]
■■ 저 둘 중에 하나를 선택해라.	[저/MM    둘/NR    중/NNB+에/JKB]
■■ 저, 뭐라더라..	[저/IC]
■■ 저 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 : 아니?  
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 대응으로 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/V+어요/EF+?SF]  
[그러/VV+어요/EF+?SF]

#### <주의사항>

(가) [보완] 맥락에 따라 감탄사로 쓰였는지 판단이 어려운 경우가 있다. 이때 그 형태가 선·후행 형태소와 같을 때는 선·후행 형태소를 반복한 것으로 분석하고, 그렇지 않은 경우에는 감탄사로 처리한다.

- 학교에 에 [에/JKB]  
가서 에 반 친구를 만났어요 [에/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.

# 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. 오류는 오류 위치 검색으로 찾을 수 있다.<sup>31)</sup>

	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

31) 구 단위 주석과 표현 문형 주석은 구 전체와 구 구성요소에 각각 주석함을 원칙으로 한다.

	오류 유형		주석 표지
		접두사	CXPN
		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	
	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.<sup>32)</sup>

	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

32) 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 문



장이 초급보다 높은 수준이지만 정확한 문장 생성에 실패 하였으므로 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,

한국어 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 앞부분에서 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.<sup>33)</sup>

<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. 다만, 교정할 위치가 어절이나 구 단위 차원에서 처리하여야만 용이하거나 구 단위로만 교정이 가능한 한 경우는 구 단위로 처리할 수 있다. 또한 표현 문형의 경우는 <국립국어원 2> 목록을 확인하여 해당 표현이 목록에 있을 경우는 표현 문형도 함께 주석한다. 즉, 기본적으로 형태 단위로 분석하여 주석하되, 구 단위와 표현 문형도 중복 주석할 수 있다. (☞ 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.

33) 이후 웹사이트에서는 교정어절을 중심으로 한 검색도 가능하게 하여, 미사용으로 인한 오류를 파악할 수 있게 한다.

즉, [오류 위치-표현 문형, 관형사형 전성어미], [오류 양상-대치] 오류로 주석한다.

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 좋춤하고(√/좋고? 조용하고?, IMP) 예쁘기 대문에(때문에, MIF) 결혼(√/결혼, MIF)하고 싫어요.

## 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.<sup>34)</sup>
- 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 독요(√/도쿄, CNNP, MIF)에 가고 싶어요.
	일반 명사	CNNG	일반 명사 어휘의 사용에서 나	애기와(√/아기와, CNNG,

34) 단, 누락 오류의 경우에는 원 어절이 없으므로 교정 어절에 따라 주석한다.

위치	주석 표지	포함 범위	예시
		타난 오류	MIF) 노인들한테 건강이 나빠졌다.
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

위치	주석 표지	포함 범위	예시
		오류	돈이나 개인 정보를 잃어버릴 수도 있다.
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 사용에서 나타난 오류	‘네(√네, CIC, MIF), 알겠습니다’라고 대답했다.
접두사	CXPN	접두사 어휘의 사용에서 나타난 오류	최초 임금을 실행하면 처소득층(√저소득층, CXPN, MIF) 사람의 살기가 보증할 수 있다.
명사파생 접미사	CXSN	명사 파생 접미사 어휘의 사용에서 나타난 오류	두 번째(√번제, CXSN, MIF)에 간 곳이는 경주였습니다.
동사파생 접미사	CXSV	동사 파생 접미사 어휘의 사용에서 나타난 오류	그 꿈을 위해서 매일 운동해다(√운동한다, CXSV, MIF).
형용사파생 접미사	CXSA	형용사 파생 접미사 어휘의 사용에서 나타난 오류	이러한 사회에서 자신이 하고 싶은 직업을 할 수 있으면 너무 행복안(√행복한, CXSA, MIF) 것이다.
어근	CXR	어근 어휘의 사용에서 나타난 오류	시원(√시원한, CXR, REP) 옷을 준비하세요.



위치		주석 표지	포함 범위	예시
기능 어휘	주격 조사	FNP	주격조사의 사용 에서 나타난 오 류	그리고 여행이(√을, FNP, REP) 너무 좋아합니다. 고시원에서 많이 학생(√이, FNP, OM) 살았다.
	관형격 조사	FGP	관형격조사의 사 용에서 나타난 오류	‘론딩리’의(√를, FGP, REP) 소개합니다.
	목적격 조사	FOP	목적격조사의 사 용에서 나타난 오류	그래서 저는 한국 사람하고 다른 외국 사람을(√과, FOP, MIF/REP) 교류하고 싶습니다.
	부사격 조사	FAP	부사격조사의 사 용에서 나타난 오류	차 안에(√에서, FAP, REP) 잤어요.
	접속 조사	FJC	접속조사의 사용 에서 나타난 오 류	한국어 문법와(√과, FJC, MIF) 중국어 문법이 비슷하지 않았다.
	호격 조사	FVP	호격조사의 사용 에서 나타난 오 류	친구아(√친구야, FVP MIF), 같이 가자.
	인용격 조사	FQP	인용격조사의 사 용에서 나타난 오류	내가 감사하다고 말한다는 게 ‘고맙다’다고(√라고, FQP, REP) 했다.
	보조사	FXP	보조사의 사용에 서 나타난 오류	론딩리는 맛있는 음식은(√이, FXP, REP) 많습니다.
	연결 어미	FED	연결어미의 사용 에서 나타난 오 류	그리고 친구들과 같이 노래방 가고(√가서, FED, REP) 노래를 부르고 싶습니다.
	종결	FFE	종결어미의 사용	특히 아랫목에 정말

위치		주석 표지	포함 범위	예시
	어미		에서 나타난 오류	따뜻한다(√ 따뜻하다, FFE, MIF).
	선어말 어미	FPE	선어말어미의 사용에서 나타난 오류	내일부터 수업이 시작됐어요(√ 시작돼요, FPE, REP).
	명사형 전성 어미	FNE	명사형 전성어미의 사용에서 나타난 오류	우리 계획은 저녁을 먹기(√ 먹은, FNE, REP) 후에 우리 만든 신분증을 가지고 갈 겁다.
	관형사형 전성어미	FAE	관형사형 전성어미의 사용에서 나타난 오류	중요하는(√ 중요한, FAE, MIF) 것은 사람마다 다르다는 것을 인정하는 것이다.
구 단위 표현		PHE	구 단위 표현 사용에서 나타난 오류	가족들이 관심하면(√ 관심을 가지면, PHE, REP) 생존이 행복해지겠다.
표현 문형		PE	보조 용언이나 여러 요소의 결합 구성으로 이루어진 표현 문형의 사용에서 나타난 오류(제시 목록 참고)	신세대는 기상세대와 가치관이 달라서 세대 차이가 생기기 마련이다(생기기 마련이다, PE, MIF).

### 3) 오류 양상

- 오류 양상은 표면적으로 드러난 오류의 모습으로, 누락, 첨가, 대체, 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
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 많게 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 규칙	PC	음운변동에 관한 오류로 구어에서	[구어] 한국계(✓ 한국에, CNNG,

층위		주석 표지	포함 범위	예시
			필수적 음운 규칙의 일탈 또는 음운 규칙을 적용하지 않고 절음화하여 발음한 경우 예) 연음규칙, 비음화 유음화, 구개음화, 경음화 등	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,
	중 간 발음(변이음 포함)	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	단 어 형성[합성법]	MCP	단어 합성에서 나타나는 오류	해물고기(√물고기, CNNG, MCP)가 많았어요.
	단 어 형성[파생법]	MDV	단어 파생에서 나타나는 오류	작년 방학 때는 LG 전자에서 통역사로 일한 경험이 있고 한국에 대한

층위	주석 표지	포함 범위	예시
			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
	시제	ST	시제를 나타내는 문법 형태소의 오류
			할머니께서 우유를 마시십니다(√드십니다, CVV, REP, SH).
			어제부터 항상 시제를 확인하기로 한다(√했다, FPE, REP, ST).

층위	주석 표지	포함 범위	예시
사동	SC	사동사, 사동 표현의 오류	갈릴레이는 새로 발명된 망원경을 사용하여 연구를 깊었다. (✓ 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 문법정보보다는 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 연구는 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
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) 실절어휘

##### 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 이야기를 해 뉘야 된다(√ 된다)  
      ☞ ‘뉘다’는 보조용언 ‘된다’를 잘못 사용한 것이기 때문에 [오류 위치-보조용언]으로 처리하는 동시에, ‘-어/아야 되다’가 표현 문형의 목록에 있기 때문에 표현 문형 위치도 중복 주석한다. 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 지정사 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 것다(√ 것이다).  
      ☞ 지정사 ‘이다’가 생략되었으므로 [오류 위치-지정사], [오류 양상-누락]으로 처리한다.  
          남자이예요(√ 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]

학생이어서(√ 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 학곤데(√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 향유하고 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(√불합리적인) 행동을 제약할 수 있는 효과적인 방법이다.  
산세대(√신세대) 사람들이 부모님 입장에 많이 생학하고 부모님도 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 것이 큰 원인이라고 할 수 있다.

⑯ 동사파생접미사(CXSV)

- 동사파생접미사의 형태, 의미, 사용 오류를 말한다.

<예> 제 한국어를 좋아합니다(√좋아합니다).  
이에 따라서 노동사의 인권이 보장하게(√보장되게) 되어 안정한 생활을 할 수 있게 되었다.

⑰ 형용사파생접미사(CXSA)

- 형용사파생접미사의 형태, 의미, 사용 오류를 말한다.

<예> 고독스러운(√고독한) 중학생

이러한 사회에서 자신이 하고 싶은 직업을 할 수 있으면 너무 행복안(√행복한) 것이다.

⑱ 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 오렌지가(√오렌지를) 샀어요.

- ‘나는’과 ‘내가’가 상호 교정 어절이 될 때에는 조사 오류로 한 번만 처리한다. 즉, 주격조사나 보조사의 대치로 인하여 대명사의 형태가 바뀌는 경우에는 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.  
기숙사의(√기숙사에) 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 "수험 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 있다.  
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(-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 고팠다(√고팠다).

⑬ 명사형 전성어미(FNE)

- 명사형 전성어미의 형태, 의미, 사용과 관련된 오류를 말한다.

<예> 과학자들의 의식에 의하면 아이들은 어른들보다 배우기(√배우는) 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 먹을(√먹는) 것이 중요해요.

- ☞ 관형사형 전성어미 대치 오류로 처리한다.  
한국에 온(√오는) 비행기에서 친구를 만났어요.
- ☞ 관형사형 전성어미 대치 오류로 처리하며, 시제와 관련된 어 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(√ 끝에) 남편과 5살 어린 아이를 두고 더 좋은 인생을 찾으려 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, ‘나타나다’ 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 해당 품사의 대치로 처리한다.

<예> 태도를 키워야 한다 (√ 길러야 한다)  
☞ ‘태도를 키워야 한다’를 ‘태도를 길러야 한다’로 교정할 때, 이를 연어 오류로 볼 수 있다. 그러나 연어의 정의가 판단

자마다 다를 수 있다는 문제가 있다. 따라서 연어 오류를 주석하기 위해서는 연어 목록이 선행되어야 하므로 본 연구에서는 연어 오류를 별도로 주석 하지 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.<sup>35)</sup>
- 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],

35) 표현 문형 목록 기준을 모든 표현 문형의 합집합으로 할 경우, 다양한 형태들이 표현 문형 안으로 들어가기 때문에 본 연구에서는 어느 정도 정제된 목록으로서 <한국어 문법 사전>(국립국어원, 2005)을 기준으로 정한다. 표현 문형 목록은 <부록>으로 첨부하였다.

[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 않

고, 필수적인 성분이 생략됐을 경우에만 [누락] 오류로 처리한다.

- 누락 오류는 하나의 유의미한 교정 어절에만 누락 오류로 주석하고, 뒤따라 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 기분이 왔는데요.  
 ☞ ‘있을’이 누락되었는데, 형태주석을 기본단위로 오류주석을 할 때, 동사 ‘있’과 관형사형 전성어미 ‘을’을 각각 누락 오류로 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 누락은 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.

- 동사 누락은 다음과 같은 것들이 해당된다.

<예> 한번 (√먹어) 보야 한다.  
 ☞ 본동사 ‘먹다’가 누락되었으므로 [오류 위치-동사], [오류 양상-누락]으로 주석한다. 뒤따라오는 연결어미는 교정어절만 써 주고 누락으로 주석하지 않는다.

○ 조사 누락은 다음과 같은 것들이 해당된다.

<예> 그로 인해 평소 일상생활에서 말할 수 없는 말 욱하는 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 누락은 다음과 같은 것들이 해당된다.

<예> 10년 후에 내가 가(√ 갈) 수 있다.  
 ☞ 관형사형 전성어미를 누락한 오류로, 관형사형 전성어미 ‘-(으)ㄴ’의 [누락]으로 주석한다. 경우에 따라서 받침을 제대로 쓰지 못한 오철자 오류로 볼 수도 있으나, 주석의 일관성을 위해 누락 오류로 주석한다. 단, 구어에서는 받침 발음을 실현시키지 못한 것인지, 문법 요소를 누락시킨 것인지에 대한 구분이 필요하므로 주의하여 처리한다.

○ 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 ‘ㄹ’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 경

우나 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 첨가는 다음과 같은 것들이 해당된다.

<예> 대만의 수도가(√수도) 타이베이

☞ 주격조사 ‘가’가 첨가된 오류로 주석한다.

※참고: 이때 ‘수도가’를 ‘수도인’으로 교정할 경우에는, 주격조사 ‘가’와 서술격조사 ‘이’의 대치 오류로도 볼 수 있으



나 본 연구에서는 ‘최소 수정 원칙’으로 주격조사 첨가 오류로 처리한다.

- 표현 문형 첨가는 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 보존하려고(√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, ‘어드레스’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> 누가 돈이 없다면 행복할 수 없다고 생학했는데 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 바

뛰는 어머니 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 대치는 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.

○ 명사 대치는 다음과 같은 것들이 해당된다.

<예> 보고 결정하는 이야기가(✓일이) 많이 있다.

☞ 명사 ‘이야기’를 ‘일’로 대치하므로 [오류 위치-명사], [오류 양상-대치] 오류로 주석한다. 명사가 바뀔므로 해서 뒤따라 바뀌는 조사 ‘가’와 ‘이’는 대치로 주석하지 않는다.

○ 연결어미 대치는 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(‘-니까’의 활용형도 잘못 사용)], [오류 층위-활용] 오류로 주석한다.

- 주의: 대치와 오형태 오류 판단 시, 의미적 대치인지, 오형태 오류인지 혼동되는 경우가 있다. 다시 말해서, 오류의 원인이 의미의 문제인지 형태의 문제인지를 고려해야 하는데, 이럴 경우 문맥을 고려하여 오류를 판단하며, 학습자들이 얼마나 이러한 오류를 보일 수 있는가를 고려한다. 예를 들어, 한국어교육 과정에서 다루고 있는 어휘인지, 숙달도를 고려했을 때 사용할 수 있는 어휘인지, 해당 맥락에서 출현할 수 있는 어휘인지 등을 고려하여 처리할 수 있다. 아울러 교정은 최소 수정 원칙을 기본으로 하되, 모어 화자에게 자연스러운 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 용

언 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 ‘니다’에 대한 철자 오류로, [오류 위치-종결어미], [오류 양상-오형태(MIF)] 오류로 주석한다.

- [처리 예시] 오형태 오류의 오류 위치별 일부 예시는 다음과 같다.
- 명사 오형태는 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 오형태는 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 오형태는 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 ‘쓰’를 누락시킨 경우는 선어말 어미 오형태(오철자) 오류로 처리한다. [오류 위치-선어말 어미], [오류 양상-오형태] 오류로 처리한다.

○ 연결어미 오형태는 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 오형태는 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)  
 ☞ ‘비니다/습니다/니다/입니다’는 종결어미 오형태 활용 오류로 처리한다.

#### 4) 오류 층위

##### (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 층위에 음소(PP)를 주석한다.
- [처리 기준] 단어 내에서 명확하게 다른 음소로 발음한 경우나 음소를 발음하지 못한 경우, 음소 오류로 주석한다.
- 구어에서 받침과 같은 특정 음소를 실현시키지 못하였을 때에는 오류 양상에 누락을 주석하지 않도록 주의한다. 문법적 요소를 누락시킨 경우에는 누락을 주석할 수 있지만, 학습자가 발음하지 제대로 하지 못하여 실현되지 않은 음소에 대해서는 오류 층위에서 음소만을 주석한다.

<예> 팔이(√빨리), 말을::, 즐= 아:: 잘해서::,  
 ☞ 자음 ‘ㅃ’을 ‘ㅍ’로 발음하여, 음소 오류로 주석한다. [오류 위치-일반부사], [오류 양상-없음], [오류 층위-음소]  
 예:: 코피(√커피)= 카페에서::, 예:: 공부를, 합니다::.  
 ☞ 모음 ‘아’와 ‘오’를 교체하여 발음하므로 음소 오류로 주석한다. [오류 위치-명사], [오류 양상-없음], [오류 층위-음소]

부모니(√부모님)

☞ 한 단어 안에서 받침을 발음하지 못한 경우도 마찬가지로 음소 오류로 주석한다. [오류 위치-명사], [오류 양상-없음]

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 ‘ㄱ’과 모음 ‘으’를 결합하여 3음절로 발음하고 있으므로 음절 오류로 주석한다. [오류 위치-의존명사], [오류 양상-없음], [오류 층위-음절]로 주석한다.

- 학습자가 단어를 완전히 발화하지 못하였을 때에는 음절이 줄어든 것처럼 보이나 이는 형태를 제대로 구현시키지 못한 오류로 음절이 아닌 오형태 오류로 처리한다.

<예> 길가에 아 음 쓰레기통을 쓰레(√쓰레기) 버리고

☞ ‘쓰레기’의 음절이 줄어든 것처럼 보일지라도 이는 발음 차원의 문제로 보기 힘들기 때문에 형태 오류로 처리한다.

③ 음운규칙(PC)

- [정의] 음운규칙 오류는 구어 발화에 나타난 필수적 음운 규칙의 일탈을



말한다. 유음화, 연음화하여 발음해야 하는데, 글자 그대로 절음화하여 발음한 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 [설랄]로 발음하지 않고 [설]과 [날]을 각각 끊어서 개별 음절 발음에 충실하였다면 ‘음운규칙’ 오류로 처리한다. 이밖에 ‘육학년’을 [유강년]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)).

- 학습자가 음운규칙을 몰라서 발생시킨 오류인지 음소를 변별하지 못하여 발생시킨 오류인지를 판단하여 오류 층위에 음운규칙과 음소를 구분하여 주석한다.

<예> 한국어(√ 한국어)

☞ 음운규칙 미적용 오류. 오류 위치: 명사, 층위: 음운규칙(PC)

한국거(√ 한국어)

☞ 오류 위치: 명사, 층위: 음운규칙(PC)

한구꺼(√ 한국어)

☞ 음운규칙은 적용하였으나 음소 변별 실패  
오류 위치: 명사, 층위: 음소(PP)

한국꺼(√ 한국어)

☞ 오류 위치: 명사, 층위: 음운규칙(PC), 음소(PP)

- 한 단어 내에서 음운 규칙 또는 음소와 관련한 오류가 나타났을 때에는 그 단어의 품사를 주석하면 되지만, 체언과 조사, 용언의 어간과 어미 사이에서 음운 규칙이 적용되면서 나타난 오류의 경우, 오류 위치를 무엇으로 삼을 것인가가 문제가 된다. 이때에는 오류가 나타난 발음이 어느 부분의 영향을 받았는지를 고려하여 처리한다. 예를 들어, 모음으로 시작되는 조사나 어미의 경우 본래 음가를 가지고 있지 않았으나, 선행하는 음절의 받침 발음에 영향을 받아 초성에서 그 발음이 실현된다. 따라서 조사나 어미의 초성 위치에서 오류가 나타나더라도 오류 위치를 영향을 미친 체언과 용언의 어간으로 주석한다.

<예> 한국계(√ 한국에) 갔어요.

☞ 오류 위치: 고유명사, 오류 층위: 음운규칙(PC)

한국계(√한국에) 갔어요.

☞ 오류 위치: 고유명사, 오류 층위: 음운규칙, 음소(종성이 그대로 실현됨과 동시에 [ㄱ]와 [ㄱᄃ] 음소를 변별하지 못한 오류로 해석한다.)

네 말이 맞쵸요(√맞아요).

☞ 오류 위치: 형용사, 오류 층위: 음운규칙(PC), 음소(PP)  
음식글(√음식을) 먹어요.

☞ 오류 위치: 명사, 오류 층위: 음운규칙(PC)

- 연음이 적용되는 단어에서 연음은 시켰지만 음소를 잘못 발음하였을 때에는 음소 오류로 처리한다.

<예> 네 말이 마차요(√맞아요).

☞ 오류 위치: 형용사, 오류 층위: 음소(PP)

음시글(√음식을) 먹어요.

☞ 오류 위치: 명사, 오류 층위: 음소(PP)

- 종성이 다음 음절의 초성으로 넘어가서 발음되는 연음과 달리, 경음화나 격음화 오류는 오류의 양상에 따라, 발음된 위치를 고려하여 오류 위치를 주석한다.

<예> 물건을 찰코(√찰고찰꼬).

☞ 오류 위치: 연결어미, 오류 층위: 음소(PP)

[찰꼬]로 발음되어야 하는데, 연결어미 ‘고’를 ‘코’로 발음하였다. 이것이 앞의 종성 [ㄷ]의 영향을 받은 것인지, 학습자가 초성 위치에서 [ㄱ]와 [ㄱᄃ]를 변별하지 못한 것인지를 판별하기는 어렵다. 본래 음가를 가지고 있지 않은 어미가 연음 규칙에 따라 발음을 잘못하였을 때에는 선행하는 어간의 품사를 오류 위치로 주석한다. 그러나 ‘고’처럼 본래 음가를 가지고 있던 위치에서 음소를 변별하지 못하였을 때에는 해당 위치에 오류 위치를 주석한다. 따라서 이러한 경우 오류 위치:

연결어미, 오류 층위: 음소(PP)로 주석한다.

응답뻘(√응답한) 결과

☞ 오류 위치: 명사, 오류 층위: 음운규칙(PC)

☞ 오류 위치: 동사파생접시마, 오류 층위: 음소(PP)

격음화가 적용되어 [응다판]으로 발음해야 하지만, [응답]을 그대로 발음하는 동시에 격음으로 발음해야 하는 [판]을 [뻘]으로 발음하였다. 따라서 ‘응답’과 동사파생접시마 ‘하’를 형태 분리하므로 ‘응답’에는 음운규칙을 주석하고, ‘하’에는 음소를 주석한다.

#### 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 같 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) 줌:: 없어,지고:: 있습니다 지금.

☞ 외국어 발음 오류로 외래어 표기법과 ‘다르다’는 차원에서 [오류 층위-원어식 발음] 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 구어 전사 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)

- [정의] 단어 형성[합성법] 오류는 조어 과정에서 발생하는 오류를 말한다. 즉, 학습자가 존재하지 않는 어휘를 생산해 내는 오류가 포함된다.
- [주석 방식] 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 오

류 양상을 필수적으로 주석하지 않고 오류 층위에서 파생(MDV)과 합성(MCP)만을 주석하도록 한다.

- 따라서 학습자가 생산해 낸 형태가 한국어에는 없는 합성어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 합성(MCP)으로 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)

- [정의] 단어 형성[파생법] 오류는 조어 과정에서 접사를 잘못 사용한 오류를 말한다.
- [주석 방식] 학습자가 파생접사(유사파생접사 포함)를 사용하여 생산해 낸 형태가 한국어에는 없는 파생어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 파생(MDV)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.

- [처리 기준] 동사 어간에 ‘하다’를 붙여 한국어에는 없는 동사를 만들어냈을 경우는 파생 오류로 처리한다. 이는 합성 오류로도 볼 수 있지만 형태주석에서 ‘-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 동사로 볼 수도 있고 파생접사로도 볼 수 있다. 형태주석에서는 이를 동사파생접미사로 주석하기 때문에 오류주석에서도 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 형태주석에서 접사로 따로 분리하여 처리하기 때문에 형태주석에서 분리하는 접사들은 ‘누락/첨가’로 처리하고, 피동/사동은 어휘 대치로 처리한다. 또한 어휘적 의미를 더해주는 접사의 경우, 예를 들어 ‘사과’를 ‘뽕사과’로 썼을 때에는 형태분석에서 ‘뽕’을 분리하지 않기 때문에 어휘적 의미를 더해주는 접사가 덧붙여진 단어는 대치 오류로 처리한다.
- 그밖에 형태주석에서 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)

- [정의] 굴절[곡용] 오류는 조사의 이형태를 잘못 사용한 경우를 말한다.
- [주석 방식] 굴절[곡용] 오류의 오류 양상은 기본적으로 오형태(MIF) 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 잘못된 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)을 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 ‘ㄷ’ 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 양쪽에 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.(√덥습니다.)

☞ ‘ㄷ불규칙 활용’ 형용사인 ‘덥다’를 종결어미 ‘ㄴ니다’ 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 주

석한다.

- [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 좋아한  
다.  
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 어간으  
로 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 재  
구성한 중간언어로 보고 용언의 활용 오류로 처리한다.  
이는 학습자들이 연결어미(예: 해고), 종결어미(예: 됩니  
다), 관형사형 전성어미(예: 해는) 앞에서 동일한 형태를  
생산해 내는 것으로 보아, 학습자들이 ‘해’를 하나의 단위  
로 인식하고 활용을 잘못 적용하였다고 판단했기 때문이  
다. 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 될고(√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 앞에서 ‘ㄴ, ㄹ, ㅂ’ 요소들이 첨가된 경우 오형태 활용 오류로 보아, 이 역시 오형태 활용 오류로 처리한다.

- [어미 활용 오류 1] 용언의 받침 유무에 따라 어미의 이형태 선택이 달라지는 경우는 활용 오류로 처리한다.

<예> 이렇게 살으면(√살면) 정말 행복할 수 있다.  
 ➡ 받침의 유무에 따라 연결어미의 이형태를 선택하여 사용해야 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 오류가 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 쓰지 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 오류로 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 시키야 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 잘못한 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 듣면서(√ 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)

- [정의] 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.
- [주석 방식] 품사 오류의 오류 양상은 기본적으로 대치(REP) 오류로 주석하고 층위에 품사(POS)를 주석한다.
- [처리 기준] 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.

<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 생가기는 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 방법을 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 나타냈다. 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 주석

한다.

<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 바뀔 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, -다 고, -하니까’ 등 관형사형 전성어미나 연결어미 등은 ‘누

락’ 처리하지 않는다.

### (3) 통사

#### ① 높임(SH)

- [정의] 높임 오류는 조사, 선어말어미, 종결어미 부분의 높임 관련 문법 형태소, 높임 어휘의 사용이 잘못된 경우를 말한다.
- [주석 방식] 높임 요소를 잘못 사용하였을 때(예를 들어 주격 조사 ‘가’와 ‘께서’의 교체)에는 오류 양상을 대치(REP)로 주석한다. 높임 요소를 사용해야 하는데 사용하지 않은 경우는 누락(OM)을, 불필요한 높임 요소를 사용했을 때에는 첨가(ADD)로 오류 양상을 주석한다. 총위는 모두 높임(SH)으로 주석한다.
- [처리 기준] 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 ‘게시다, 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 주석한다.
- 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.

- 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.

- 학습자가 글이나 발화에서 ‘-다/ㄴ다’와 ‘-ㅂ니다/습니다’ 등 종결어미를 혼용하여 사용한 경우에는 학습자의 종결어미 선택에 일관성이 부족한 것으로 판단하여 더 자주 쓴 종결어미로 교정하고 대치 오류로 주석한다. 그런데 두 종결어미가 대치될 때 이를 높임 오류로 볼 수 있는가에 대해서는 문어와 구어에서 다르게 접근할 수 있다. 문어의 경우 구어에 비해서 학습자가 높임 범주를 인식하지 못하였다고 일괄적으로 판단하기 어려운 측면이 있는 데 반해, 구어에서는 발화 상대를 고려하였을 때 높임에 대한 인식을 고려해야 한다고 보고 말뭉치의 유형에 따라 달리 처리한다.
- 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)

- [정의] 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) 오느날 갑자기 끝에 왔다.

- [주석 방식] 시제 오류는 오류 양상을 두 가지로 처리한다. 하나는 시제 요소를 사용하지 않고 기본형을 사용했을 경우, 시제를 사용하지 않았다고 판단하여 누락(OM) 오류로 주석한다. 다른 하나는 시제를 제대로 인식하지 못하여 과거 시제나 미래 시제 자리에 현재 시제를 사용한 경우, 그 반대의 경우 등은 대치(REP) 오류로 처리한다. 즉, 시제를 사용했으나 현재와 과거, 과거와 미래처럼 잘못 사용한 경우는 시제 간 대치로 처리한다. 양상은 달라질 수 있지만 층위에는 모두 시제(ST)를 주석한다.
- 이때, ‘이다/아니다’, 형용사, 연결어미 앞에서의 용언은 기본형이 현재를 나타내기 때문에 기본형을 사용했을 경우, 현재 시제로 인식하고 대치(REP)로 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 대치 오류로 처리한다. 그러나 시제 선어말 어미의 문법적 제약이 있는 연결어미 앞에서 ‘-었-’과 ‘-겠-’을 사용한 경우는 첨가(ADD) 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.

- ‘-겠-’의 경우, 미래보다는 추측이나 의지와 같은 양태 의미를 나타내는 경우가 많다. 본 연구에서는 ‘-겠-’이 양태 의미를 나타내는 경우에는 시제를 주석하지 않는다.

<예> 앞으로 학교 규칙을 안 어긴다(√어기겠다).

☞ 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 경우와 시제로 인한 오류일 경우를 구분하여 처리한다. 예를

들어, 시제를 나타내지 않는 경우로 ‘-(으)ㄴ 때, -(으)ㄴ 따름이다, -(으)ㄴ/는 편이다, -(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)ㄴ’)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)ㄴ지’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)ㄴ지’의 대치의 경우, 오류 층위에 시제를 주석하지 않도록 주의한다.

### ③ 사동(SC)

- [정의] 사동 오류는 사동사, 사동 표현의 사용, 사동문 생성에서 발생한 오



류를 말한다.

- [주석 방식] 사동사, 사동 표현을 사용해야 하는데 사용하지 않은 경우는 기본적으로 오류 양상을 대치(REP) 오류로 주석하고, 층위에는 사동(SC)을 주석한다.
- [처리 기준] 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 밥들 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 관계없이, 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 층위

는 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)

- [정의] 피동사, 피동 표현의 사용, 피동문 생성에서 발생한 오류를 말한다.
- [주석 방식] 피동사, 피동 표현을 사용해야 하는데 사용하지 않은 경우는 기본적으로 대치(REP) 오류로 주석하고, 층위에는 피동(SP)을 주석한다.
- [처리 기준] 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 분석에 따라 처리하므로, 동사로 분석했을 때에는 그 품사를 따르고, 연결어미, 보조 용언으로 분리하였을 때에는 분리된 형태대로 오류 위치를 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 오류로 처리한다.

##### ⑤ 부정(SN)

- [정의] 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.

- [주석 방식] 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.
- [처리 기준] ‘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)

- [정의] 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.

- [주석 방식] 어순 오류는 오류 위치와 오류 층위만 주석하고, 오류 양상은 주석하지 않는다. 또한 교정어절을 줄 필요가 없어 교정된 어순을 반영하여 앞이나 뒤에 추가하지 않는다. 아울러, 2개의 문장 성분이 상호 교체될

때에는 2개 모두 대치 어순(WO) 오류로 주석한다.

<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.

- [쟁점] 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.
- [처리 기준] 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.

<예> 8반 시(√8시 반)에 학교에 가서 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.

○ 용인 가능성을 적용하여 다음은 어순 오류로 처리하지 않는다.

<예> 저는 우즈베키스탄에서 한국에 2014년 7월에 왔어요.

☞ 시간부사와 장소부사가 함께 올 때 한국어에서는 보통 시간 부사를 먼저 쓴다. 그리하여 ‘저는 2014년 7월에 우즈베키니스탄에서 한국에 왔어요.’가 더 자연스러울 수 있지만 문법적으로는 용인 가능한 것으로 보고 어순 오류로 처리하지 않는다.

○ 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 처리함

을 원칙으로 한다. 이에 따라, 지시 표현에서 의미 간 대치(REP) 오류를 중심으로 하여 오류 층위에 지시(DR)를 주석한다. 지시 표현에서 나타난 단순 오철자 오류는 오류 양상에 오형태(MIF)만 주석하고, 오류 층위에서 지시를 주석하지 않도록 주의한다.

- [처리 기준] 지시 오류는 앞 뒤 문장과 연결, 상황 맥락을 통해서 오류 판단이 가능하기 때문에 문장 단위를 기본원칙으로 삼으나 지시 오류의 경우는 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 해당되므로 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 포함된다.
- [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 오형태(MIF)만 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 주석하고, 층위에 담화표지((DM)를 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. 구어 발화에서 학습자가 L1의 영향으로 인한 간투사 사용과 모어 화자와는 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 발음하고 있어 오형태 오류로 주석하고, 오류 층위에서 담화표지 오류로 처리한다.

- 학습자가 L1을 간투사처럼 사용한 경우 담화 표지를 주석한다. L1을 사용하는 것이 오류라고 할 수는 없지만, 학습자가 구어 발화에서 L1을 어떻게 사용하고 있는지, 발화 전략 등을 파악할 수 있다는 차원에서 이에 대해 담화 표지를 부착한다.

<예> 아노(✓담화 표지)... 음..  
 ☞ 일본인 학습자가 한국어로 발화하기 전, L1의 간투사를 그대로 사용한 경우이다. 감탄사(IC)로 형태 주석하므로 [오류

위치-감탄사], [오류 층위-담화표지]를 주석한다.

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 식별하고 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:: = 음식’과 같이 수정 전 발화에서 조사를 잘못  
 사용하였지만, 다시 수정하여 조사를 고쳐 제대로 사용한  
 경우에 앞부분 ‘음식이’는 오류로 처리하지 않는다.

- [주석 방식] 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.
- [처리 기준] 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 한다. 그러나 조사의 경우, 문법 오류를 우선 처리하도록 하고, 관형사형 전성어미를 사용해야 할 자리에 사용하지 못했을 경우, 문법 오류로 우선 처리한다. 문어와 마찬가지로 문법 오류를 우선 처리하도록 하되, 둘을 구분해야 할 때에는 학습자의 발음 양상을 살핀다. 만약, 학습자가 반복적으로 특정 음소를 다른 음소로 발음하거나 받침을 실현시키지 못하는 경우는 문법 누락 오류와 발음 오류를 구분하여 처리한다.

<예> 가(√갈) 수 있는 방법으(√을) 모라서(√몰라서)

☞ 원칙적으로 관형사형 전성어미를 사용해야 할 자리에 사용하지 못했을 경우, 문법 오류로 보고 누락 오류로 처리한다. 그러나 예시처럼 학습자가 관형사형 전성어미 ‘-르’

뿐만 아니라 목적격 조사 ‘을’과 동사 ‘몰라서’에서 반복적으로 ‘ㄹ’ 받침을 제대로 발음하지 못하는 양상을 보이면 누락이 아닌, 음소 오류로 처리한다. 이는 발음 층위의 문제이기 때문이다.

- 발음 오류는 문법뿐 아니라 어휘 오류와의 구분도 모호할 때가 있다. 구어 발음 오류와 어휘 형태 오류의 구분이 어려울 때에는 학습자가 형태를 제대로 모르는 것인지 발음의 문제인지를 판단한다. 구어에서 발음 차원이 아닌, 형태를 잘못 발화한 경우는 오행태(MIF) 오류로 처리한다.

<예> 사잉(√사건)  
보석필(√보살핌)  
그 부모들이 그 고절(√걱정)이 많이 되어서  
아무리 보다도 요즘 여성들이 사회 화동 더 많이 참가하고  
그 자아시선(√자아실현) 이런 거도 어 요구 많이 생겨서

☞ 발음의 위치를 고려해 보면 동일 조음 기관이나 유사한 부분에서 발음되었다고 보기 힘들다. 이러한 형태들은 음소의 변별보다는 단어의 형태를 잘못 알고 있거나 형태를 잘못 만들어낸 것이다. 유사 발음과도 떨어져 음소 오류로 볼 수 없고, 한국어에 없는 형태들을 발음했다고 판단하여 오행태 오류로 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](쫌), [바래요](바라요) 등 구어에서 허용되는 발음은 오류로 처리하지 않는다.

- 구어의 특성이나 표현 의도에 의한 발음 특성은 오류로 처리하지 않는다.

<예> 표현 의도에 의한 수의적 경음화: 쪼금

구어에 의한 발음 특성: ~먹었구요 / ~했어여

- 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::

[글구](√그리고)

☞ 그러나 <표준>에 등재되어 있지 않더라도, ‘그쵸’와 같이 구어에서 축약된 형태로 많이 나타나는 용례들은 오류로 처리하지 않는다.

- 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 전사한 것으로 판단이 불분명하기 때문에 분석불능으로 처리한다.

- 구어 오류는 오류 층위에서 발음 오류와 가장 밀접하다. 오류 층위 [발

음]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> [보롱](√복용),하는 뜻이느(√뜻은)::, 그 마약을 쓰는 아니면 마약을, 하,는:: 것입니다.

☞ 복용을 [보공]으로 발음해야 하나 [보롱]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 ‘뜻이느’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		
-ㄴ 데도 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, -ㄴ 대로	
-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 채하다		
-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, -ㄴ데도 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
-르 뻔하다		
-르 뿐만 아니라		
-르 수밖에 없다		
-르 줄		
-르 테고		
-르 테냐		
-르 테니		
-르 테다		
-르 테면		
-르 테야		
-르 테지만		
-르 텐데		
-르까 보다		
-르락 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	-ㄴ 데도 불구하고, -는데도 불구하고	
을 가지고	-ㄴ 가지고	
-을 것 같다	-ㄴ 것 같다	
-을 것1	-ㄴ 것	
-을 것2	-ㄴ 것	
-을 것이 아니라	-ㄴ 것이 아니라	
-을 나름이다		
-을 대로	-ㄴ 대로	
-을 듯	-ㄴ 듯	
-을 듯하다	-ㄴ 듯하다	
-을 따름이다	-ㄴ 따름이다	
-을 때	-ㄴ 때	
-을 리가 없다	-ㄴ 리가 없다	
-을 리가 있다		
을 막론하고	를 막론하고	
-을 만큼	-ㄴ 만큼	
-을 만하다	-ㄴ 만하다	

표제어	형태 정보	대표형
-을 모양이다	-리 모양이다	
-을 바에	-리 바에	
-을 법하다	-리 법하다	
을 불문하고		
-을 뻔하다	-리 뻔하다	
-을 뿐만 아니라	-리 뿐이다	
-을 뿐이다		
-을 수 없다		
-을 수 있다		
-을 수밖에 없다	-리 수밖에 없다	
을 위해(서)	를 위해(서)	
-을 줄	-리 줄	
-을 테고	-리 테고	
-을 테냐	-리 테냐	
-을 테니	-리 테니	
-을 테니까		
-을 테다	-리 테다	
-을 테면	-리 테면	
-을 테야	-리 테야	
-을 테지만	-리 테지만	
-을 텐데	-리 텐데	
-을까 보다	-리까 보다	
-을락 말락 하다	-리락 말락 하다	
-지 말다		-지 말다
-지 못하다		-지 못하다
-지 않다		-지 않다

## 2019 Project on Research and Construction of the Korean Learner Corpus

This project is a follow-up study in its third stage—the fifth year of research on the construction of a learner corpus for Korean language learners in mid- and long-term situations. Constructing effective data and improving the quality of corpus construction established in the fifth-year research, this research aims to revise and elaborate on 2,602,780 raw corpora (written language 2,023,438 items, spoken language 579,342 items) constructed during previous research, and simultaneously add approximately 800,000 new raw corpora (written language 550,000 items, spoken language 250,000 items) to the learner corpus. The major tasks and goals of the project are as follows:

**Research on Korean Language Learner Corpus:** Research on the construction of Korean language learner corpora is divided into foundation building for systematic corpus construction and model demonstration for corpus application. In the fifth year of research, the team conducted continuous construction of domestic institutional data, construction of immigrant data, and intensive collection of international Korean language learner data to ensure the balance of data. In terms of systemicity and consistency, the data were efficiently collected and constructed through supplementing guidelines regarding data collection, data processing, written language, spoken language, as well as morph tagged guidelines and error annotated guidelines. In addition, the research system for utilizing corpora was established through an analysis of collective corpus on construction theory and application research constructed from 2015 to 2019 and a methodological study for securing the balance of Korean language learner corpora.

### **Korean Language Learner Corpus Education and Promotion:**

Education about and promotion of learner corpora in the Korean language have focused on instructional education and tool use to facilitate efficient and systematic corpus construction for the worker participating in each stage of corpus construction. The fifth year research provided a special lecture on data processing of a corpus by using Python to enhance their expertise as corpus-building personnel. Many users were informed of this through the three-time Korean learner corpus academy, thus enabling them to utilize the corpus more easily and widely. In addition, the conferences raised academic interest in Korean learner corpus research and construction.

### **Korean Language Learner Corpus Collection, Revision and Construction:**

The construction of the Korean learner corpus for the fifth-year of the project proceeded in earnest with the establishment of intensive construction to balance domestic education institutional and immigrant data which was the target of intensive construction in the first to four years of the projects (1st-4th year). The construction of the Korean language learner corpus built off of the foundation established in the previous four years of research (1st-4th year), and new items were constructed in the form of 1,048,052 raw corpora (written 795,925 items, spoken 252,127 items), 738,220 morph tagged corpus (written 527,744 items, spoken 210,476 items), and 226,136 error annotated corpus (written 113,789 items, spoken 112,347 items). In total, the first to fifth year of research (1st-5th year) from 2015-2019 amount to 3,650,832 raw corpus (written 2,819,363 items, spoken 831,469 items), 2,629,251 morph tagged corpus (written 2,037,748 items, spoken 591,053 items), and 799,157 error annotated corpus (written 462,322 items, spoken 336,835 items).

**Corpus construction support tool management and construction corpus verification elaboration:** A Korean learner corpus is built using corpus-building support tools that can manage and execute the whole work process from sample registration to corpus annotation processing. In

the fifth year, a performance feedback team was organized around the research staff to allow for regular feedback on construction support tools. In addition, bimonthly discussions were made on various issues related to the use of construction support tools and the construction of the corpus with the National Institute of Korean language, the development support team of the construction team, and the construction team at the working-level. Therefore, this allowed us to manage deployment support tools and construct the corpus more efficiently. On top of this, the quality of the corpus data was continuously implemented by each team through written input, colloquial transcription, shape annotation, and error annotation. In order to improve the quality of the corpus data, three teams comprised of inspection, system-based data verification, and inspection work were created.

As the successor to the “Korean Learner Corpus”, this study may be applied broadly by education researchers, Korean language Instructors, and Korean language learners. This broad application will then pave the way for strengthening the Korean language’s systemization and its international competitiveness.

Key terms: Korean learner corpus, continuously collection of domestic institutional data, intensive of immigrant data, collection of international Korean language learner data, written language corpus, spoken language corpus, raw corpus, morph tagged corpus, error annotated corpus

연구 책임자: 한송화(연세대학교 글로벌인재대학 교수)  
 공동 연구원: 김선정(계명대학교 한국문화정보학과 교수)  
                   김일환(성신여자대학교 국어국문학과 교수)  
                   김정숙(고려대학교 국어국문학과 교수)  
                   김한샘(연세대학교 문과대학 교수)  
                   안경화(서울대학교 언어교육원 대우 부교수)  
                   홍혜란(연세대학교 국어국문학과 BK 연구교수)  
                   김수현(국립국어원 한국어진흥과 학예연구사)

연구 보조원: 공나형(연세대학교 대학원 박사과정)  
                   김동은(연세대학교 대학원 박사과정)  
                   김미선(연세대학교 대학원 박사과정)  
                   김선영(연세대학교 글로벌인재대학 강사)  
                   박은현(연세대학교 대학원 박사과정)  
                   배미연(연세대학교 대학원 박사과정)  
                   송지혜(연세대학교 대학원 박사과정)  
                   윤현애(연세대학교 대학원 박사과정)  
                   유소영(연세대학교 대학원 박사과정 수료)  
                   허희정(연세대학교 대학원 박사과정)

담당 연구원: 김수현(국립국어원 한국어진흥과 학예연구사)

## 2019년 한국어 학습자 말뭉치 연구 및 구축 사업

발 행 인	소 강 춘
발 행 처	국립국어원 서울시 강서구 금낭화길 148(방화 3동 827) 전화: 02-2669-9743~4 전송: 02-2669-9727
인 쇄 일	2019년 12월 31일
발 행 일	2019년 12월 31일
인 쇄	학위사

※ 이 보고서는 국립국어원 누리집(<http://www.korean.go.kr>)에서도 내려받을 수 있습니다.