

국립국어원 2017-01-53

발간등록번호
11-1371028-000690-01

2017년 한국어 학습자 말뭉치 연구 및 구축 사업

연구 책임자
강 현 화



제 출 문

국립국어원장 귀하

“2017년 한국어 학습자 말뭉치 연구 및 구축 사업”에 관하여
귀 원과 체결한 연구용역 계약에 의하여 연구보고서를 작성하여
제출합니다.

2017년 12월 15일

연구 책임자: 강현화(연세대학교)

연구 기관 연세대학교 산학협력단

연구 책임자 강현화(연세대)

공동 연구원 김선정(계명대), 김일환(고려대),
김정숙(고려대), 김한샘(연세대),
안경화(서울대), 이동은(국민대),
이정희(경희대), 한송화(연세대),
김수현(국립국어원)

실무 연구원 홍혜란(연세대)

연구 보조원 공나형(연세대), 김동은(연세대),
김선영(연세대), 김정현(연세대),
배미연(연세대), 송지혜(연세대),
유소영(연세대), 허희정(연세대)

2017년 한국어 학습자 말뭉치 연구 및 구축 사업

이 연구는 한국어 학습자 말뭉치 구축 중장기 사업 계획에 따른 2단계, 3차 연도 연구이다. 이 연구는 2차 연도의 학습자 말뭉치 구축 지침을 자료의 특성에 맞게 정교화하고 그 지침에 따라 기구축된 115만 어절(문어 100만 어절, 구어 15만 어절) 규모의 말뭉치를 수정·보완함과 동시에 약 66만 어절(문어 50만 어절, 구어 16만 어절) 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. 2단계는 이주민 자료의 구축이 시작되는 시기로 이주민 자료의 특성과 실제적인 자료 구축에 따른 주요한 쟁점들을 반영하여 2차 연도에 마련된 수집, 자료 처리, 문어 입력과 구어 전사, 형태 주석, 오류 주석의 각 작업 단계별 지침을 정교화하였다. 또한 구축된 말뭉치가 한국어 교육 연구자와 교사, 학습자에게 효율적으로 활용될 수 있도록 연구 체계를 마련하였다. 연구 체계는 1-3차 연도에 구축된 말뭉치 분석 결과에 따른 기초 통계, 학습자 말뭉치 나눔터를 기반으로 한 자료의 검색 및 활용 모형, 실제적 연구에의 적용 사례와 함께 한국어 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구로 구성된다.

한국어 학습자 말뭉치 관련 교육 및 홍보: 한국어 학습자 말뭉치 관련 교육 및 홍보에서는 체계적인 말뭉치 구축을 위하여 자료의 수집과 수집 자료의 처리, 문어 입력, 구어 전사, 형태 주석, 오류 주석의 각 단계에 참여하는 인력들을 대상으로 지침 교육과 도구 사용 교육을 수행하였다. 아울러 정기/비정기 워크숍을 통해 작업자 간의 일관성과 통일성을 높임으로써 구축 자료의 신뢰성을 높이려 하였다. 한편 한국어 학습자 말뭉치 사업을 널리 알리고 사용자들의 활용 능력 제고하기 위하여 한국어 교육 연구자와 교사를 대상으로 한 워크숍을 서울에서 2회, 지방에서 1회 개최하였다. 또한 학술대회 발표를 통해 한국어 학습자 말뭉치를 연구자들에게 널리 알리고 실제적인 활용 모형을 소개하였다.

한국어 학습자 말뭉치 수집 및 가공: 한국어 학습자 말뭉치 수집은 국내 거주 어학 연수생과 유학생, 이주민(결혼이주민, 중도입국청소년, 이주노동자)을 대상으로 하였으며, 전국의 한국어 교육 기관 14개와 다문화가족지원센터, 사회통합프로그램 운영 기관 등 20개의 이주민 교육 기관의 협조를 얻어 수행되었다. 한국어 학습자 말뭉치 가공은 1-2차 연도에 구축한 원시 말뭉치 115만 어절(문어 100만 어절, 구어 15만 어절), 형태 주석 말뭉치 77만 어절(문어 70만 어절, 구어 7만 어절), 오류 주석 말뭉치 20만 어절(문어 14만 어절, 구어 6만 어절) 규모의 말뭉치를 새로운 지침에 따라 수정·보완하는 작업이 선행되었다. 그리고 원시 말뭉치 66만 어절(문어 50만 어절, 구어 16만 어절), 형태 주석 말뭉치 50만(문어 40만, 구어 10만 어절), 오류 주석 16만 어절(문어 10만, 구어 6만) 규모의 말뭉치가 새롭게 구축되었다. 그 결과 1-3차 연도(2015-2017년)에 구축한 전체 말뭉치의 규모는 원시 말뭉치 181만 어절(문어 150만, 구어 31만 어절), 형태 주석 말뭉치 127만 어절(문어 110만 어절, 구어 17만 어절), 오류 주석 말뭉치 36만 어절(문어 24만 어절, 구어 12만 어절)이 되었다.

구축 말뭉치 검증 정교화: 한국어 학습자 말뭉치는 각 단계별로 작업자 간의 교차 검수, 상위 작업자의 검수, 작업 단계별 연계를 통한 심층 검수를 순차적으로 거치게 된다. 그 외에도 3차 연도 사업에서는 말뭉치의 검수 체계를 한 단계 더 정교화하기 위하여 개발팀과의 협업을 통해 말뭉치 구축 도구에 통계적으로 데이터의 신뢰도를 검증하는 기능을 추가하여 시험 사용하고 개선안을 개발팀에 제안하였다. 그 외에도 작업 로그와 표본 정보, 주석 정보 등을 기반으로 하여 주석 작업이나 교정 작업 중에 생성된 오조작 데이터를 추출하여 검수하였다.

<한국어 학습자 말뭉치>는 <21세기 세종 한국어 균형 말뭉치> 이후로 구축되는 국가 주도의 말뭉치로서 한국어 교육 연구, 교수, 학습에 광범위하게 활용됨으로써 한국어의 세계화 및 국제 경쟁력 강화에 이바지할 것이다.

주요어: 한국어 학습자 말뭉치, 국내 교육기관의 학습자 자료,
이주민 자료, 문어 말뭉치, 구어 말뭉치, 원시 말뭉치
형태 주석 말뭉치, 오류 주석 말뭉치

차 례

I. 연구 개요	1
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 추진 경과	3
5. 연구 결과	4
 II. 학습자 말뭉치 연구	 6
1. 말뭉치 구축 지침의 정교화	6
1.1. 수집 지침	6
1.2. 문어 입력 지침	8
1.3. 구어 전사 지침	11
1.4. 형태 주석 지침	13
1.5. 오류 주석 지침	18
2. 말뭉치 활용 연구 체계 구축	44
2.1. 2015-2017년 구축 말뭉치를 활용한 분석 연구	44
2.2. 한국어 학습자 말뭉치 활용 실제 사례 연구	63
2.3. 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구	74
 III. 학습자 말뭉치 교육 및 홍보	 81
1. 말뭉치 구축/가공 인력 실무 교육	81
1.1. 교육 대상	81
1.2. 교육 방법	81
1.3. 교육 내용	82

1.4 참여 인력	83
2. 한국어 학습자 말뭉치 아카데미 개최	84
3. 학술대회 발표	85
4. 이주민 워크숍 개최	86

IV. 학습자 말뭉치 구축 및 가공 86

1. 수집	86
1.1. 수집 대상	86
1.2. 교육 기관 참여 방안 마련: 수집 네트워크 마련	87
1.3. 자료 수집 결과	88
2. 구축 및 가공	95
2.1. 원시 말뭉치	95
2.2. 형태 주석 말뭉치	100
2.3. 오류 주석 말뭉치	104
2.4. 1-2차 연도 구축 말뭉치의 정제	107

V. 구축 말뭉치 검증 정교화 108

1. 문어 입력 및 구어 전사 검수	108
1.1. 형태소 주석, 오류 주석 작업 단계에서의 심층 검수	108
1.2. 작업 중 생성된 오조작 데이터 검증	109
2. 형태·오류 주석 검수 시스템 정교화	110
2.1. 작업 중 생성된 오조작 데이터 검증	110
2.2. 형태소별 주석의 종합 검수	113
2.3. 시스템 기반의 데이터 신뢰도 검증	114

VI. 결론	119
1. 연구 요약	119
2. 연구의 의의 및 기대 효과	121
3. 보고서 활용 방안	123
4. 정책 제언	124

참고문헌

부록. 2017년 한국어 학습자 말뭉치 구축 지침

표 차례

<표 1> 연구의 범위와 세부 과업	2
<표 2> 연구 수행 방법	2
<표 3> 연구 추진 경과	3
<표 4> 2017년 학습자 말뭉치의 구축 규모	4
<표 5> 이주민 수집 과제 설계의 방향 및 내용	7
<표 6> 학습자 동의서의 학습자 정보 수집 항목	8
<표 7> 최고급 자료 처리를 위한 본문 주석 체계	9
<표 8> 지침 보완 내용: 문어 입력 지침	10
<표 9> 지침 보완 내용: 구어 전사 지침 1	11
<표 10> 지침 보완 내용: 구어 전사 지침 2	12
<표 11> 지침 보완 내용: 구어 전사 지침 3	12
<표 12> ‘못하다’의 분석 지침의 수정	17
<표 13> 지침 보완 내용: 단어 형성[합성법]과 [파생법] 오류	18
<표 14> 지침 보완 내용: 굴절[곡용](MDC)과 [활용](MCJ) 오류	20
<표 15> 지침 보완 내용: 품사(POS) 오류	24
<표 16> 지침 보완 내용: 높임(SH) 오류	27
<표 17> 지침 보완 내용: 시제(ST) 오류	28
<표 18> 지침 보완 내용: 사동(SC)과 피동(SP) 오류	31
<표 19> 지침 보완 내용: 부정(SN) 오류	35
<표 20> 지침 보완 내용: 어순(WO) 오류	37
<표 21> 지침 보완 내용: 담화 오류	38
<표 22> 문어 형태 주석 결과_숙달도 단계별	45
<표 23> 문어 형태 주석 결과_국적별	46
<표 24> 구어 형태 주석 결과_숙달도 단계별	48
<표 25> 구어 형태 주석 결과_국적별	49
<표 26> 문어 오류 주석 결과_분석 불능 여부: 숙달도 단계별	51
<표 27> 문어 오류 주석 결과_분석 불능 여부: 국적별	51
<표 28> 문어 오류 주석 결과_오류 위치: 숙달도 단계별	51
<표 29> 문어 오류 주석 결과_오류 위치: 국적별	53
<표 30> 문어 오류 주석 결과_오류 층위: 숙달도 단계별	54
<표 31> 문어 오류 주석 결과_오류 층위: 국적별	55

<표 32> 문어 오류 주석 결과_오류 양상: 숙달도 단계별	56
<표 33> 문어 오류 주석 결과_오류 양상: 국적별	56
<표 34> 구어 오류 주석 결과_분석 불능 여부: 숙달도 단계별	56
<표 35> 구어 오류 주석 결과_분석 불능 여부: 국적별	57
<표 36> 구어 오류 주석 결과_오류 위치: 숙달도 단계별	57
<표 37> 구어 오류 주석 결과_오류 위치: 국적별	58
<표 38> 구어 오류 주석 결과_오류 층위: 숙달도 단계별	60
<표 39> 구어 오류 주석 결과_오류 층위: 국적별	61
<표 40> 문어 오류 주석 결과_오류 양상: 숙달도 단계별	62
<표 41> 문어 오류 주석 결과_오류 양상: 국적별	62
<표 42> 변인에 따른 ‘아주’와 ‘너무’의 사용 양상 비교	65
<표 43> 한국어 학습자가 될 수 있는 외국인의 분포	76
<표 44> 한국어 학습자가 될 수 있는 외국인의 국적 분포	76
<표 45> 자료 변인	79
<표 46> 말뭉치 구축/가공 인력 교육 내용	82
<표 47> 한국어 학습자 말뭉치 아카데미 개최 일시 및 장소	85
<표 48> 이주민 워크숍 프로그램	86
<표 49> 이주민 수집 네트워크	88
<표 50> 문어 자료 수집 현황: 국내 교육 기관	89
<표 51> 구어 자료 수집 현황: 국내 교육 기관	90
<표 52> 이주민 자료 시험 수집 현황: 숙달도별(자료 수, 개)	90
<표 53> 이주민 자료 시험 수집 현황: 국적별(자료 수, 개)	91
<표 54> 2017년 종적 자료 학습자 분포	94
<표 55> 문어 원시 말뭉치의 숙달도별 자료 분포	95
<표 56> 문어 원시 말뭉치의 국적별 자료 분포	96
<표 57> 구어 원시 말뭉치의 숙달도별 자료 분포	98
<표 58> 구어 원시 말뭉치의 국적별 자료 분포	99
<표 59> 형태 주석 말뭉치의 숙달도별 자료 분포	100
<표 60> 문어 형태 주석 말뭉치의 국적별 자료 분포	101
<표 61> 구어 형태 주석 말뭉치의 국적별 자료 분포	103
<표 62> 오류 주석 말뭉치의 숙달도별 자료 분포	104
<표 63> 문어 오류 주석 말뭉치의 국적별 자료 분포	105
<표 64> 구어 오류 주석 말뭉치의 국적별 자료 분포	106

<표 65> 1-2차 연도 기구축 학습자 말뭉치의 규모	107
<표 66> 표본 정보에 생성된 오조작 데이터 검증	109
<표 67> 오조작 데이터 검증: 작업 진행 이상 여부 검증	110
<표 68> 오조작 데이터 검증: 작업 결과 이상 표본 검증	111
<표 69> 오조작 데이터 검증: 파일의 표본 정보와 파일 정보 간 이상 표본 검증	112
<표 70> 오조작 데이터 검증: 표본 정보 중 선행 정보 없이 후행 정보가 있는 표본 검증	112
<표 71> 주식 유형별 검수 항목: 1-2차 연도 기구축 자료 대상	113
<표 72> 검수 작업의 단계화	113
<표 73> 주식 유형별 검수 항목: 1-3차 연도 전체 자료 대상	114

그림 차례

<그림 1> 한국어 학습자 말뭉치 검색 화면 예시1	63
<그림 2> 한국어 학습자 말뭉치 검색 화면 예시2	64
<그림 3> 한국어 학습자 말뭉치 검색 화면 예시3	64
<그림 4> 한국어 학습자 말뭉치 검색 화면 예시4	65
<그림 5> 한국어 학습자 말뭉치 검색 화면 예시5	66
<그림 6> 한국어 학습자 말뭉치 검색 화면 예시6	66
<그림 7> 한국어 학습자 말뭉치 검색 화면 예시7	67
<그림 8> 한국어 학습자 말뭉치 검색 화면 예시8	67
<그림 9> 한국어 학습자 말뭉치 검색 화면 예시9	68
<그림 10> 한국어 학습자 말뭉치 검색 화면 예시10	68
<그림 11> 한국어 학습자 말뭉치 검색 화면 예시11	69
<그림 12> 한국어 학습자 말뭉치 검색 화면 예시12	70
<그림 13> 한국어 학습자 말뭉치 검색 화면 예시13	70
<그림 14> 한국어 학습자 말뭉치 검색 화면 예시14	70
<그림 15> 한국어 학습자 말뭉치 검색 화면 예시15	71
<그림 16> 한국어 학습자 말뭉치 검색 화면 예시16	71
<그림 17> 한국어 학습자 말뭉치 검색 화면 예시17	72
<그림 18> 시스템 기반의 데이터 신뢰도 검증 절차	114
<그림 19> 형태 주식 말뭉치 검증을 위한 가상 말뭉치 설정	116
<그림 20> 형태 주식 말뭉치 품질 확인	116
<그림 21> 불일치 항목의 형태 확인	117
<그림 22> 불일치 항목의 문맥 확인	117
<그림 23> 오류 주식 말뭉치 검증을 위한 가상 말뭉치 설정	118
<그림 24> 불일치 항목의 교정 형태 확인	118

I. 연구 개요

1. 연구의 목적 및 필요성

1.1. 연구의 목적

○ 본 연구의 목적은 ‘한국어 학습자 말뭉치 구축 사업 중장기 계획’에 따라 한국어 학습자의 숙달도별·언어권별 문어·구어 말뭉치 자료를 구축하여 과학적인 한국어교육의 기반을 조성하는 데에 있다. 즉, 한국어교육 연구는 물론 한국어교육 현장에서 교재를 개발하거나 교수·학습 방법을 구안하는 데에 학습자 말뭉치를 활용함으로써 보다 체계적이고 과학적인 교육 환경을 제공하기 위한 것이다. 본 연구는 2015년-2016년에 수행된 <한국어 학습자 말뭉치 구축 사업>에 이은 제3 차 연도의 사업으로 이를 위한 세부 목표는 다음과 같다.

- 말뭉치 구축 지침의 정교화와 말뭉치 활용 체계 구축을 위한 연구
- 학습자 말뭉치 구축 관련 교육 및 홍보
- 말뭉치 수집 및 구축 가공
- 구축 말뭉치 검증 정교화

1.2. 연구의 필요성

- 국가 주도의 한국어 학습자 균형 말뭉치 구축
- 한국어 학습자의 다양화에 따른 교수·학습 환경의 과학화 필요
- 학습자 말뭉치를 활용한 연구 방법론과 도구 사용 교육에 대한 요구
- 학습자 말뭉치 활용 연구를 통한 국제 수준의 학술 교류에 대한 요구

2. 연구의 범위

○ 본 연구의 범위와 세부 내용은 다음과 같다.

<표 1> 연구의 범위와 세부 과업

연구의 범위	세부 내용
학습자 말뭉치 연구	○ 말뭉치 구축 지침 정교화 ○ 말뭉치 활용 연구 체계 구축
학습자 말뭉치 구축 관련 교육 및 홍보	○ 말뭉치 구축/가공 인력 실무 교육 ○ 한국어 학습자 말뭉치 아카데미 개최(2회) ○ 학술대회 발표
학습자 말뭉치 수집 및 구축 가공	○ 한국어 학습자 실제 말뭉치 자료 수집 및 구축 (66만 어절) ○ 1-2차 연도 말뭉치 정제 및 2017년 구축 말뭉치 가공 (형태분석, 오류분석)
구축 말뭉치 검증 정교화	○ 구어·문어 입력 검수, 형태·오류 검수 시스템 정교화 ○ 검수 결과 및 현황 보고 (중간보고회 1회, 최종보고회 전 1회)

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, 사례 조사, 현황 조사, 전문가 의견 수렴의 네 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 활용	○ 말뭉치 활용 연구 체계 구축

	용에 관한 선행 연구	의 기초 자료
사례 분석	○국내외 학습자 말뭉치 활용 사례	○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 협의체 구성을 통한 자문 및 의견수렴 ○사업 착수 보고회의, 중간보고회의, 최종보고회의를 통한 정기 자문 및 평가	○교육 기관의 참여 유도 ○연구 방향 및 방법, 절차의 타당성 검증

4. 연구 추진 경과

○ 연구 추진 경과는 다음과 같다.

<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
준비	1-2차 연도 구축 자료 정리 및 사업 계획 수립	○								
연구	구축 지침의 정교화	○	○	○	○	○	○	○	○	○
	말뭉치 활용 연구 체계 구축	○	○	○	○	○	○	○	○	○
수집	자료 수집 지침 교육	○	○	○	○	○	○	○	○	○
	자료 수집	○	○	○	○	○	○	○	○	○
구축	말뭉치 입력/전사 지침	○	○	○	○	○	○			

	및 도구 교육									
	스캔 및 파일 분류, 등록	○	○	○	○	○	○	○	○	○
	입력 및 전사	○	○	○	○	○	○			
가공	말뭉치 가공 지침 및 도구 교육	○	○	○	○	○	○	○	○	
	형태 분석	○	○	○	○	○	○	○	○	
	오류 주석	○	○	○	○	○	○	○	○	
홍보	한국어 학습자 말뭉치 아카데미 개최						○		○	○
	학술대회 발표		○					○		
마무리	2차 년도 사업 마무리 및 3차 년도 기초 계획 수립									○

5. 연구 결과

○ 본 사업의 연구 결과는 다음과 같다.

- 구축 단계별 지침 정교화
- 학습자 말뭉치 활용 연구 체계 구축
- 한국어 학습자 말뭉치 아카데미 개최 및 학술대회 발표
- 1-2차 연도 구축 말뭉치의 정제
- 2017년 신규 말뭉치 자료 수집 및 구축, 가공

<표 4> 2017년 학습자 말뭉치의 구축 규모

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
원시	문어	83,545 (1,302)	63,611 (842)	52,753 (505)	34,801 (293)	63,310 (436)	78,392 (502)	129,112 (115)	505,524 (3,995)
	구어	22,123 (70)	28,648 (57)	37,121 (82)	24,088 (50)	31,510 (52)	29,317 (34)	5,765 (4)	178,572 (349)
형태	문어	60,736	70,413	48,717	72,211	52,093	48,431	52,598	405,199

		(847)	(652)	(352)	(590)	(348)	(288)	(55)	(3,132)
주석	구어	26,827 (69)	13,632 (34)	11,823 (36)	8,700 (29)	18,838 (24)	20,453 (22)	5,765 (4)	106,038 (218)
오류 주석	문어	16,593 (260)	16,516 (154)	17,993 (141)	17,805 (132)	14,431 (88)	16,609 (102)	-	101,947 (877)
	구어	7,948 (28)	10,905 (30)	11,823 (36)	10,724 (32)	12,019 (14)	6,925 (7)	-	60,344 (147)

단위: 어절 수, () 안은 파일 수

- 구축 말뭉치 검증 체계의 정교화

Ⅱ. 학습자 말뭉치 연구

1. 말뭉치 구축 지침의 정교화

1.1. 수집 지침

(1) 지침 수정 및 보완의 쟁점

○ 수집 지침은 3차 연도 사업의 집중 수집 대상인 이주민의 특성을 반영하여 수집 과제와 학습자 동의서를 수정·보완하는 데에 주력하였다. 구체적인 사항은 다음과 같다.

- 이주민 학습자의 교육과정을 반영한 수집 과제 설계
- 이주민 학습자의 특성을 반영한 학습자 정보 수집 항목 추가
- 이주민 학습자의 국적 분포를 고려한 동의서 번역 언어 추가

(2) 수정 및 보완 내용

- ① 이주민 학습자의 교육과정을 반영한 수집 과제 설계
- 이주민 학습자는 결혼이주민, 이주노동자, 중도입국청소년의 세 집단으로 구성된다. 이들 중 성인 학습자인 결혼이주민과 이주노동자는 다문화가족지원센터나 사회통합프로그램의 교육과정에 편입되어 한국어를 학습하고 있고, 중도입국청소년의 경우 초·중·고 각 급 학교나 비정규 교육 기관에서 한국어를 학습하고 있다. 이들이 속해 있는 한국어 교육과정은 사회통합프로그램을 기준으로 할 때 약 400시간 정도로 국내 대학 교육 기관 학습자에 비해 매우 짧고, 학습 목적이나 환경, 그에 따른 교육 내용에도 큰 차이를 보인다. 이에 따라 2차 연도의 수집 과제를 기초로 하되, 아래와 같은 방법과 절차로 학습자의 특성에 맞게 수집 과제를 수정·보완하였다. 이렇게 마련된 과제는 실제 수집 과정에서 수집 교사로부터 다양한 환경 변인, 학습자 변인에 의해 조정이 필요한 부분에 대한 피드백을 받아 조정하여 자료 수집의 효율성을 높일 수 있도록 탄력적으로 적용하고 있다.

<표 5> 이주민 수집 과제 설계의 방향 및 내용

구분		설계 방향 및 내용
종적 수집 과제	성인	<ul style="list-style-type: none"> - 최소 1년(48주)에서 최대 1년 6개월(60주) 수집을 목표로 초급, 중급, 고급 단계로 구분하여 60주차 30회분의 말하기/쓰기 과제 설계 - 토픽 기출문제를 기반으로 한 국내 교육 기관 학습자의 종적 자료의 수집 과제를 참고하되, 최신 기출 문제를 추가하고 이주민 교재(<여성결혼이민자와 함께하는 한국어>, <EPS-TOPIK을 위한 표준한국어>)를 분석하여 이주민에게 적합한 주제들을 추가하여 보완함
	중도입국 청소년	<ul style="list-style-type: none"> - 성인과 동일하게 30회분의 말하기/쓰기 과제 설계 - 성인용 수집 과제 중 초급의 공통 주제를 포함하되 <초/중/고등학생을 위한 표준한국어 교재>의 과제와 활동을 분석하여 항목을 추가함
기관 수집 과제	성인	<ul style="list-style-type: none"> - 교육과정에 따른 시험 자료 수집 - 필요 시 종적 수집 과제를 기획 과제로 활용
	중도입국 청소년	<ul style="list-style-type: none"> - 각 급 학교의 교육과정 내에 언어 평가가 포함되지 않으므로 중도입국청소년 수집 과제를 별도로 설계함 - <KSL 교육과정 진단도구>의 말하기, 쓰기 문항을 활용하여 작성하였으며, 초등, 중·고등으로 나누어 과제의 유형이나 난이도에 차등을 둠

② 이주민 학습자의 특성을 반영한 학습자 정보 수집 항목 추가

- 이주민 학습자의 경우 국내 대학의 교육 기관과 교육과정이 다르고 각 기관 간, 학습자 개인 간 편차가 커 한국어 수준을 측정할 수 있는 준거가 분명하지 않다는 특성이 있다. 또한 한국어 사용 환경에 노출되어 있는 시간이 길고 일상적으로 사용할 수 있는 기회가 많아 한국어 습득 양상이 국내 대학 교육 기관의 학습자와 다른 양상을 보일 것으로 판단된다. 따라서 학습자가 수강하고 있는 교육과정의 등급 외에 학습자의 언어 발달에 영향을 미칠 수 있는 환경 변인에 관한 다음의 항목들을 추가 수집하여 자료의 활용도를 높이하고자 하였다.

<표 6> 학습자 동의서의 학습자 정보 수집 항목

학습자 정보 항목	국내 교육 기관	이주민 학습자
성별, 출생년, 현재 등급, 국적, 제1언어, 한국에서의 거주 기간, 한국어 학습 목적, 직업, 한국어 외의 사용 가능 외국어	공통	공통
한국어 학습 기간		학습 기관, 학습 교재명 추가
평상시에 가장 많이 사용하는 언어	-	신규 추가
한국어 대화 상대	-	신규 추가
한국어로 말하는 시간	-	신규 추가
한국어를 듣는 시간	-	신규 추가
한국어로 나오는 방송 매체(TV, 라디오, 인터넷 동영상)를 보는 시간	-	신규 추가

③ 이주민 학습자의 국적 분포를 고려한 동의서 번역 언어 추가

- 이주민 학습자는 베트남, 태국, 필리핀, 캄보디아, 네팔, 파키스탄 등 동남아시아 지역 학습자의 비중이 높다. 이는 국내 대학 교육 기관 학습자가 주로 중국, 일본, 대만, 베트남, 미국, 홍콩 등에서 높은 비중을 보이는 것과 다르다. 이에 따라 학습자 동의서의 번역본을 추가하여 영어, 중국어, 일본어, 러시아어, 몽골어, 베트남어, 필리핀어, 인도네시아어, 몽골어, 태국어, 캄보디아어, 네팔어, 파키스탄어, 방글라데시어로 총 14개 언어로 번역된 동의서가 수집 지침에 포함되었다. 이는 수집 교사의 편의성을 제고하고 학습자 본인 정보를 정확히 수집하기 위한 것이다.

1.2. 문어 입력 지침

(1) 지침 수정 및 보완의 쟁점

- 3차 연도 문어 입력 지침에서 가장 큰 변화는 소논문 형식의 최고급 학습

자 자료가 추가됨에 따라 이를 처리하기 위한 본문 주석을 추가 개발하고 이와 관련된 처리 지침이 추가되었다는 것이다. 그 외에는 2차 연도에 구축된 지침을 토대로, 말뭉치 구축 시스템을 사용하는 과정에서 발생하는 추가적인 문제점들을 해결하고 그에 대한 작업자들의 이해를 돕기 위해 지침을 구체화하는 데에 초점이 주어졌다.

- 최고급 자료 처리를 위한 본문 주석 체계 개발
- 개인 정보 처리에 관한 지침의 구체화

(2) 수정 및 보완 내용

- ① 최고급 자료 처리를 위한 본문 주석 체계 개발
 - 최고급 자료의 경우, 주로 소논문 형식의 텍스트로 기본적인 입력 원칙은 동일하나 텍스트의 구조를 고려하여 아래와 같은 본문 주석을 추가하고 이와 관련된 사항을 지침에도 추가하였다.

<표 7> 최고급 자료 처리를 위한 본문 주석 체계

	주석	내용	주석 표지
형식 구분	보고서 제목	전체 보고서의 제목	<head>
	본문 앞	앞부분의 부속물	<front>
	본문	여러 개의 장절 제목과 본문	<body>
		장절 제목	<title>
	본문 뒤	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	<q>

	주석	내용	주석 표시
		예시문	
기타	표, 그림, 그래프, 수식 등	문어 입력 과정에서 표, 그림, 그래프 수식 등의 생략을 나타내 주는 표시	<gap reason>

② 개인 정보 처리에 관한 지침의 구체화

- 텍스트에 포함된 이름이나 소속 기관, 전화번호 등은 학습자의 개인 정보 보호 차원에서 모두 기호화하여 말뭉치 자료에서 노출되지 않도록 처리하고 있다. 그런데 개인 정보 여부를 판단하는 데 있어 작업자 간의 이견이 있어 일관된 처리가 이루어지지 못할 우려가 있다. 이에 3차 연도 지침에서는 이를 구체적으로 명시하여 자료 안에 포함된 개인 정보들을 일관성 있게 처리하도록 하였다.

<표 8> 지침 보완 내용: 문어 입력 지침

	2차 연도 지침	3차 연도 지침	비고
개인정보 처리	-	학습자 이름, 가족 이름, 친구 이름, 거주지, 소속 기관명, 학번, 직장명, 전화번호, 외국인등록번호 또는 주민등록번호 등 글을 작성한 사람의 구체적인 정보나 특정인을 규정할 수 있는 정보는 개인정보 처리함.	추가

1.3. 구어 전사 지침

(1) 지침 수정 및 보완의 쟁점

- 구어 전사 지침의 수정 및 보완은 ELAN에서의 1차 전사 작업 결과물을 말뭉치 구축 시스템으로 옮기는 과정에서 발생하는 문제점 해결에 중점을 두고 이루어졌다. 또한 형태 주석 및 오류 주석으로 이어지는 일련의 작업 과정에서의 효율성을 고려하여 전사가 이루어질 수 있도록 기존의 지침을 정교화하였다.

- 발화자 표시 위치 수정
- 전사 기호의 중복 문제 해소
- 학습자의 발음 오류 표기에 관한 지침 상세화

(2) 수정 및 보완 내용

① 발화자 표시 위치 수정

- 2차 연도에는 발화자 정보를 파일의 첫 부분에 명시하였다. 3차 연도에는 말뭉치 구축 도구에서의 구어 자료 처리 환경을 고려하여 발화자 정보를 표시하는 위치를 변경하였다.

<표 9> 지침 보완 내용: 구어 전사 지침 1

2차 연도 지침	3차 연도 지침	비고
모든 발화자에 관한 정보는 동의서의 개인 정보에 있는 내용을 참고하여 이름, 연령대, 성별, 직업을 간단하게 표시하여 파일 맨 처음에 붙인다.	시스템 등록 시 주석 입력 창에서 발화자의 기본 정보를 선택하여 입력한다.	수정

② 전사 기호의 중복 문제 해소

- 3차 연도에는 말뭉치 구축 시스템에서의 구어 자료 처리가 본격화됨에 따라 ELAN에서의 전사 지침과 시스템에서의 자료 지침 간의 연계성과 호

환성이 중요한 쟁점으로 대두되었다. 대표적으로 ELAN에서의 전사 결과물을 시스템으로 옮겨오는 과정에서 기호가 중복되면서 시스템 내에서 충돌이 일어난 것이 그 예이다. 이에 따라서 기존의 중복된 전사 기호 처리 관련 지침을 구체화할 필요성이 제기되어, 다음과 같이 전사 기호의 체계를 유형화하고 기호가 중복되어 사용되는 경우의 기입 순서를 조정하였다.

<표 10> 지침 보완 내용: 구어 전사 지침 2

2차 연도 지침	3차 연도 지침	비고
전사 기호가 서로 중복이 될 경우에는 기본적으로 문장부호를 우선하고(즉, 어절의 가장 가까운 곳에 붙이고) 웃음이나 박수 등 준음성의 표현을 2차로 한다.	<p>전사 기호가 서로 중복될 경우에는 기호의 유형을 고려하여 우선 순위에 따라 기입한다. 기호의 유형은 다음과 같이 네 가지로 나눌 수 있다.</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>< >, (), 문장부호 (온점, 느낌표, 물음표/쉽표), 기타</p> </div> <p>1) < > 기호 안에는 하나의 문장 단위만 입력가능하다. (억양을 나타내는 문장부호와 동시에 사용 불가하므로 분리하여 입력한다.)</p> <p>2) 억양기호는 바로 앞의 음절의 억양을 표시하므로 다른 기호보다 우선하여 사용한다.</p>	보완

③ 학습자의 발음 오류 표기에 관한 지침 상세화

- 한국어 모어 화자에게서도 필수적으로 나타나는 음운 변화는 철자대로 전사하되 학습자가 일으킨 오류에 대해서는 소리 나는 대로 전사하는 것을 원칙으로 한다. 다만 원래의 표기와 오류 표기가 동일한 경우 메모를 남겨 오류의 유형을 확인할 수 있도록 한다.

<표 11> 지침 보완 내용: 구어 전사 지침 3

2차 연도 지침	3차 연도 지침	비고
모든 경우 음운적 구분이 모호하거나 특징적인 사항이 있을 때에 메모를 남기도록 한다.	모든 경우 음운적 구분이 모호하거나 특징적인 사항이 있을 때, 또한 교정어절과 원어절의 표기가 동일한 경우 오류의 유형에 대한 메모를 남기도록 한다. <note> 기호를 사용하여 그 내용을 상세히 기술한다. 예) 앞에(앞에)<note>‘앞’과 ‘에’ 사이에서 연음되지 않음</note> 먹는(먹는)<note>‘먹’과 ‘는’ 사이에서 동화되지 않음</note>	수정

1.4. 형태 주석 지침

(1) 지침 수정 및 보완의 쟁점

- 2차 연도의 형태 분석 지침 수정 및 보완 작업에서 가장 중점이 되었던 부분은 ‘21세기 세종 계획 현대 문어 형태 주석 말뭉치 분석 지침과의 호환성’을 도모하는 작업이었다. 하지만 세종 형태 분석 지침의 경우 모어 화자를 위한 형태 분석 지침이기 때문에 학습자 언어의 특성을 지침에서 충분히 반영하지 못하는 한계가 있다. 이에 3차 연도에는 오류가 포함된 학습자 언어를 형태 주석함에 있어 실제적이고 일관성 있는 분석 기준의 적용 방안을 마련하는 작업이 가장 중점이 되었다. 기본 원칙에서 분석 기준의 적용과 관련한 항목을 추가하였으며, 세부 지침에서도 학습자 오류 어절의 처리에 대한 보완 항목을 마련하고 예시를 추가했다.

- 분석 불가능(NA) 항목 처리 지침의 구체화
- 학습자 오류 어절 처리 지침의 구체화
- 학습자 오류 어절의 분할 기준 보완

(2) 수정 및 보완 내용

① 분석 불가능(NA)의 처리 지침의 구체화

- 학습자 오류 어절이 나타난 경우 최소 교정을 원칙으로 교정 어절을 상정해 형태 분석을 한다. 이때, ‘교정 어절을 상정하기 어려운 경우’는 분석 불가능(NA)로 처리한다. 즉, 분석 불가능은 문맥에서 전혀 의미를 유추할 수 없을 경우의 처리 방식이다. 분석 불가능한 학습자 오류 어절이 표현 문형의 구성과 인접한 경우에는 교정 어절을 상정하기 어렵다 해도 표현 문형 구성에 포함되는 형태까지는 분석을 한다.

- 경전철 타로 필어 50분쯤 경싯역 있습니다. [필어/NA]
 - 저는 친구와 같이 순열전 수열고 싫어요. [순열전/NA]
[순열/NA+고/EC]

② 학습자 오류 어절의 처리 지침의 구체화

- 학습자의 오류 어절에서 교정 어절의 상정이 가능한 경우 교정 어절이 취할 형태 표지를 부여한다. 우선적으로 형태 분할이 가능한 것은 최대한 분할하는 것을 원칙으로 한다.

- 휴대 전화가 꺼서 모자도 보낼 수 있고 아무도 몰라다.(√몰랐다)
[무르/VV+아/EP+다/EF]
 - 술을 많이 마시는은 공간에 해로원데.(√해로운데)
[해롭/VA+언데/EC]
 - 저는 중국 사람은데.(√사람인데) [사람/NNG+으/VCP+ㄴ데/EC]

- 다음의 경우는 지침에 따라 형태를 분할해 분석하지 않고 학습자의 오류 어절에서 나타나는 오형태로 보고 분석한다.

가. 용언 어간에 포함된 파생접미사에서 오형태가 나타날 경우 형태 분할을 하지 않고 파생 접미사 자체의 오형태로 분석한다.

- 오후 7시 30분까지 운동합니다 [운동/NNG+해/XSV+ㅂ니다/EF]
- 시간이 많이 걸려 안 도착해서서
[도착/NNG+해/XSV+시/EP+어서/EC]

나. 다음의 축약형에서 나타나는 오형태는 분할하지 않고 오류 어절 자체에서 오형태 분석을 한다.

- 보통 저하고 걔 친구들이 홍대에 자주 갑니다. [재/NP]
- 또 궁금한 걔 있으면 어제든지 물어보세요. [개/NNB]

다. 복합어의 구성 요소인 어근이나 접사가 누락된 경우 단어의 오형태로 보고 복합어 전체의 품사로 분석한다. 하지만 용언의 어간이 어미와 결합한 활용형에서 어미가 누락된 경우는 용언의 어간으로 분석한다.

- 눈이 많(√ 많이) 왔다. [많/MAG]
- 줄임말은 젊은들이(√ 젊은이들이) [젊은/NNG+들/XSN+이/JKS]
- 많(√ 많은) 음식이 있었어요. [많/VA]

③ 학습자 오류 어절의 분할 기준 보완

- 오형태 분할에서 문제가 되는 부분은 상정한 교정어절에 없는 형태가 추가된 경우이다. 작업자 간 분할의 일관성을 위해 마련한 기준은 다음과 같다.
- 용언의 경우 어간 혹은 어근을 확보한 후, ‘연결/종결어미’ 앞뒤에 잉여적 요소가 추가된 경우 분리하지 않고 어미의 오형태로 처리한다.(기본적인 원칙은 이와 같지만, 학습자의 오류 양상에 따라 연결/종결어미 앞에 나타났어도 NA로 분석되는 경우가 존재할 수 있다.)

- 교통이 꼭 편립습니다. [편리/NNG+ㅂ습니다/EF]
- 어학당 교실에 사람이 많은다. [많/VA+은다/EF]

- 한편 어간과 ‘전성/선어말어미’ 사이에 잉여적 요소가 추가된 경우 분리하고 NA 처리한다. 전성어미는 연결/종결어미와 더 비슷한 특성(어말어미)을 가진다고 할 수 있지만, 형태적인 면에서는 선어말어미와 같이 타입

(type)이 제한되어 있고 형태 자체가 더 고정적이다. 이러한 특성으로 인해 상정한 교정어절에 없는 형태소가 추가된 경우 직관적으로 ‘연결/종결 어미’에서는 어미 이형태를 잘못 쓴 오류로 판단되는 반면, ‘전성/선어말어미’에서는 어간과 어미 사이에 잉여적 요소가 첨가된 오류로 판단되는 것을 반영했다.

- 하숙집인기 때문에 사람이 많았다. [하숙집+이/VCP+ㄴ/NA+기/ETN]
- 나는 선생님이 되겠다. [되/VV+ㅂ/NA+겠/EP+다/EF]

- 교정 어절을 상정할 수 있지만 형태소 경계를 구분 짓기 어려운 경우, 교정 어절의 형태소 음절에 따라 앞에서부터 형태를 분할한 후 분석한다.

- 다 무시하고 자시느이 마음대로 했다. [자시/NNG+느이/JKG]
- 아프로 미래에게 살이면 좋을 것 같다. [아/NNG+프로/JKB]
- 다언에도 소개해 주게서요. [주/VX+게/EP+서요/EF]

④ 기타 세부 지침 보완

- 학습자 언어에서 오류가 나타나는 것을 반영해 세부 지침에서의 보완 작업이 이뤄졌다. 구체적으로 분석 표지별로 보완한 사항은 다음과 같다.

가. 명사 상당어의 분석 : 학습자 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]

나. 대명사의 이형태 분석 : 학습자가 ‘제/NP+가/JKS’에서 ‘가/JKS’를 누락해서 쓴 경우 ‘저/NP+의/JKG’로 분석하지 않도록 주의한다.

- 제 먹었습니다. [제/NP]

다. 인용격조사의 분석 : 인용격조사는 인용 기호가 있을 경우에만 인용격조사로 분석하고, 인용기호가 없는 경우 연결어미로 분석한다. 다만 학습자 말뭉치에서는 생산자가 외국인 학습자이기 때문에 한국어에서 인

용 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.

- [보완] 내 일이다라고 말했다.
[일/NNG+이/VCP+다/EF+라고/JKQ] (○)
[일/NNG+이/VCP+다/EF+이/VCP+라고/EC] (×)

라. 접속 조사의 분석 : 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]

마. 접미사의 분석 : 학습자가 잘못 접미사를 사용한 경우 교정어절을 상정했을 때 교정어절의 품사가 동사일 때는 동사파생접미사(XSV), 교정어절의 품사가 형용사일 때는 형용사파생접미사(XSA)를 부여한다

- 음식을 먹하다. [먹/VV+하/XSV+다/EF+./SF]
- 마음이 아프한 아주머니가 [아프/VA+하/XSA+ㄴ/ETM]

- ‘못하다’의 분석 : ‘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 접미사 ‘하’를 분석하지 않는다.

<표 12> ‘못하다’의 분석 지침의 수정

수정 전 처리	수정 후 처리
[못/MAG+하/XSV+다/EF]	[못하/VV+다/EF]
[못/MAG+하/XSA+다/EF]	[못하/VA+다/EF]
[못하/VX+다/EF]	[못하/VX+다/EF]

1.5. 오류 주석 지침

(1) 지침 수정 및 보완의 쟁점

- 3차 연도에는 오류 주석 작업자 간의 일관성과 작업의 효율성을 위하여 보다 체계적이고 세부적인 오류 주석 지침 보완에 중점을 두었다. 이를 위하여 1, 2차 연도에 작업한 오류 주석 용례를 다시 재점검하여 귀납적으로 오류 용례에 따른 오류 판정 여부 및 오류 판정 기준에 대한 세부 지침을 수정·보완하였다.
- 특히, 언어학적 층위에 해당하는 발음, 형태, 통사, 담화상의 오류인 ‘오류 층위’에서의 오류 판정의 기준을 구체적으로 추가 보완하였다.
 - 오류 층위에서 형태와 관련된 단어 형성[합성법], [파생법], 굴절[곡용], [활용] 및 품사 오류 지침 보완
 - 오류 층위에서 통사와 관련된 높임, 시제, 사동과 피동, 부정, 어순 오류 지침 보완
 - 오류 층위에서 담화와 관련된 지시, 접속, 담화표지, 구어/문어 오류 지침 보완
- 구어 오류 주석 지침의 정교화 작업으로 오류 층위에서 발음과 관련된 오류 판정 지침을 보완하였다.
 - 음소(PP) 오류 지침 보완
 - 음절(PS) 오류 지침 보완
 - 음운규칙(PC) 오류 지침 보완
 - 원어식 발음(PN) 오류 지침 보완
 - 중간발음(변이음 포함)(PA) 오류 지침 보완

(2) 수정 및 보완 내용

- ① 오류 층위에서의 형태 오류 처리 지침 보완
- 가. 단어 형성[합성법](MCP)과 [파생법](MDV) 오류 주석

<표 13> 지침 보완 내용: 단어 형성[합성법]과 [파생법] 오류

	2차 연도 지침	3차 연도 지침
오류 양상 수정	○ 학습자가 조어 과정에서 형태를 잘못 만들어 낸 오류이기 때문에 오류 양상은 오형태 오류로 주석한다.	○ 학습자가 조어 과정에서 형태를 잘못 만들어 낸 경우, 즉, 파생과 합성 오류는 오류 양상을 필수적으로 주석하지 않고 오류 층위에서 파생과 합성만을 주석하도록 한다. ○ 따라서 학습자가 생산해 낸 형태가 한국어에는 없는 합성어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 합성으로 주석한다. ○ 학습자가 파생접사(유사파생접사 포함)를 사용하여 생산해 낸 형태가 한국어에는 없는 파생어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 파생으로 주석한다. ○ 단, 접사와 접사의 대치의 경우나 접사의 불필요한 첨가 또는 생략은 오류 양상에 대치, 첨가, 생략으로 주석한다.
접사 ‘하다’ 처리		○ 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.
유사 파생 접사 ‘파생’ 처리		○ 형태 주석에서 접사로 분리하지는 않지만 유사파생접사로 볼 수 있는 형태들을 사용하여 어휘를 파생시킨 경우는 오류 양상은 주석하지 않고, 오류 층위에 파생으로 주석한다.

- 2016년 오류 주석에서는 학습자가 조어 과정에서 잘못 생산해 낸 형태에 대해 오류 양상을 오형태로 주석하였다. 그러나 오형태 오류가 오철자 오류와 이형태 활용 오류에 해당하는 오류 양상이기 때문에 합성과 파생 관련한 오류를 형태 오류에 포함시킬 수 있는가가 문제가 된다. 이에 본 연구에서는 조어 과정에서 나타난 형태가 오형태로 보기에 적절하지 않다고 판단하였다. 따라서 오류 양상은 수의적 주석이므로 학습자가 생산한 형태가 한국어에 없는 어휘일 경우, 오류 양상은 필수적으로 주석하지 않는 것으로 수정하여, 오류 층위에서 파생과 합성만을 주석하는 것으로 변경하였다.
- 한국어 학습자들은 명사, 어근 또는 동사 어간에 ‘하다’를 과잉 적용하여 용언을 생산하는 오류를 범하는데, 이때 ‘하다’를 동사로 보느냐 접사로 보느냐에 따라 합성과 파생으로 처리할 수 있다. 본 연구에서는 이러한 경우 형태 주석에서 ‘하다’를 접사로 분리하여 처리하고 있기 때문에 형태주석과의 연계성을 고려하여 접사로 처리하고 오류 층위에서 파생 오류로 주석한다는 내용을 추가하였다.
- 접사는 형태 주석에서 생산성이 강한 접사에 한해서 목록을 마련하여 처리하고 있다. 이 목록에 해당하지 않는 경우는 분리하여 처리하지 않는데, 오류 주석에서는 형태 주석에서 접사로 분리하지는 않지만 유사파생접사로 볼 수 있는 형태들을 사용하여 어휘를 파생시킨 경우는 오류 양상은 주석하지 않고, 오류 층위에 파생으로 주석한다는 지침을 추가하였다.
- 그밖에 수정된 사항에 대해 보다 적절한 예문으로 교체하거나 추가된 보완 사항에 해당하는 예문들을 추가하였다.

나. 굴절[곡용](MDC)과 [활용](MCJ) 오류 주석

<표 14> 지침 보완 내용: 굴절[곡용](MDC)과 [활용](MCJ) 오류

	2차 연도 지침	3차 연도 지침
굴절 활용 오류 정의	○ 굴절[활용] 오류는 용언의 활용과 어미의 활용을 잘못된 경우를 말한다. 즉, 용언의 규칙 및	○ 굴절[활용] 오류는 용언의 활용과 어미의 활용을 잘못된 경우를 말한다. 즉, 용언의 규칙 및 불규칙 활용, 어미의 이형태 오류가 포함된다. 따라서 용언의 규칙/불규칙 활용과 어미 이형태에 따라 활용 오류를 처리하고, 활용 양상

	2차 연도 지침	3차 연도 지침
	불규칙 활용, 어미의 이형태 오류가 포함된다.	에 따라 오류 위치를 판단한다. 단순 오철자 오류의 경우 오형태(MIF)만 처리하며, 오류 층위에 활용(MCJ)을 주석하지 않는다.
활용 오류 범위		<ul style="list-style-type: none"> ○ 어미 앞에서 ‘아/어’, ‘으’가 삽입된 경우 어미 활용과 관련된 것으로 보고 활용 오류로 처리한다. 그러나 ‘ㄴ, ㄹ’ 등의 기타 요소는 용언의 오철자 오류로 처리한다. ○ 어미 앞에서 ‘ㄴ, ㄹ, ㅁ’ 등의 불필요한 요소를 첨가했을 경우, 기본적으로 용언 어간의 오철자 오류로 처리한다. 단, ‘ㄴ, ㄹ’이 시제의 의미를 나타낸 것이라고 판단될 경우는 이들 요소를 분리하여 관형사형 어미(ETM) 첨가(ADD)로 처리하며, 층위에 시제(ST)를 주석한다. ○ ‘아서/어서’에서 ‘서’만 쓴 경우, ‘아도/어도’에서 ‘도’만 쓴 경우, ‘(으)니, (으)면’ 등에서 ‘으’를 쓰지 않은 경우 연결 어미 활용 오류로 처리한다. 마찬가지로 종결어미에서도 ‘아요/어요’에서 ‘요’만 쓴 경우 종결어미 활용 오류로 처리한다. ○ 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. ○ 종결어미에서 ‘ㅂ니다/습니다’ 외에 ‘니다’ 또는 ‘입니다’를 잘못 사용한 경우

	2차 연도 지침	3차 연도 지침
		도 종결어미 오형태 활용 오류로 처리한다.
지정사 관련 오류 처리		○ 지정사와 종결어미가 결합할 때, 지정사를 누락시키거나 첨가했을 경우 또는 축약을 잘못된 경우는 오철자 오류로 처리한다.
‘하다/ 되다’ 활용 관련 오류 처리		○ ‘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 처리한다.

- 오류 층위에서 곡용과 활용 오류는 오류 양상에서 오형태(MIF) 오류에 해당한다. 그러나 오형태 오류는 음소 단위 형태를 잘못 쓴 오철자 오류까지 포함하기 때문에 단순 오철자 오류와 곡용과 활용 오류를 구분하여 처리해야 한다. 따라서 용언의 규칙/불규칙 활용과 조사 및 어미 이형태에 따른 오류는 오류 층위에 곡용과 활용을 주석하고, 단순 오철자 오류의 경우 오류 층위는 주석하지 않고 오류 양상에 오형태만 주석함으로써 구분한다는 내용을 명시하였다.
- 한국어 학습자는 연결어미 또는 종결어미 앞에 ‘아/어’, ‘으’, ‘ㄴ, ㄹ, ㅂ’ 등의 기타 요소를 첨가하는 경우가 많다. 이때 이들을 활용과 관련된 활용 오류로 처리할지, 오철자 오류로 처리할지 문제가 되는데, ‘아/어’, ‘으’의 경우 어미 활용과 밀접한 오류로 판단해, 활용 오류로 처리하고, ‘ㄴ, ㄹ, ㅂ’ 등의 기타 요소는 활용과 관련된 것으로 보기 어려워 오철자 오류로 처리한다는 내용을 추가하였다.
- 아울러 ‘아/어’, ‘으’가 첨가된 것을 활용 오류로 처리할 때, 오류 위치를 용언 어간에 줄 것인지, 어미에 줄 것인지 또한 문제된다. 예를 들어, 이형태가 없는 어미 ‘고, 지만’ 앞에 ‘아/어’ 또는 ‘으’가 첨가된 경우, 어간 활용과 어미 활용의 구분이 어렵다. 학습자들이 동사 어간에 ‘아/어’를 첨가해 하나의 어간으로 재구성했다고 볼 수도 있고, ‘아고, 아지만’의 형태로 어미를 잘못 활용한 것으로 볼 수도 있다. 본 연구에서는 이와 같은 경우 이

형태 어미의 영향을 받았다고 보고, 일괄 어미의 활용 오류로 처리한다는 내용을 추가하였다.

- ‘아/어/으’ 외에 ‘ㄴ, ㄷ, ㅂ’ 등의 기타 요소가 첨가된 것은 오철자 오류로 처리하는데, 이때 위치는 기본적으로 용언 어간의 오철자 오류로 처리한다. 그리고 ‘ㄴ, ㄷ’이 시제와 관련되어 나타난 오류로 판단 될 경우는 ‘ㄴ, ㄷ’을 분리하여 관형사형 어미(ETM) 첨가(ADD)로 처리하며, 층위에 시제(ST)를 주석할 수 있다는 내용을 추가하였다.
- 용언 어간과 어미의 축약상의 오류, 즉 ‘가아서(√가서)’와 같이 축약을 시켜야 하는데 축약을 시키지 않은 경우와, ‘쉬요(√쉬어요)’처럼 과도하게 축약을 시킨 경우, 활용 오류로 처리할지 오철자 오류로 처리할지 문제가 된다. 본 연구에서는 필수적으로 탈락시켜야 하는데 시키지 않은 경우 또는 과도하게 축약을 시켜버린 경우 모두 어미의 활용을 제대로 모르는 것으로 판단하여 활용 오류에 포함한다는 내용을 추가하였다.
- 종결어미 ‘ㅂ니다/습니다’를 잘못 사용한 경우는 이형태 활용 오류이나, ‘니다’만을 쓴 경우, 이형태 활용 오류로 볼 수 있는지가 문제가 된다. 특히, ‘ㅂ’ 받침으로 끝나는 용언 어간의 경우, 학습자들이 ‘아름답니다’와 같은 형태를 만들어 내는데, 이것이 용언 어간 ‘아름다’와 종결어미 ‘ㅂ니다’를 결합시킨 것인지, ‘아름답’과 ‘니다’의 형태로 결합한 것인지 불분명하다. 본 연구에서는 용언어간 ‘아름답’과 ‘니다’로 결합한 것으로 일괄 처리하고, ‘ㅂ니다/습니다’와 ‘니다’만을 결합한 경우, 종결어미의 오형태 활용 오류로 처리한다는 내용을 추가하였다.
- 지정사와 관련된 오류에서 지정사를 누락하거나 첨가했을 때, 또는 종결어미에서 잘못 축약된 형태를 썼을 때는 발음과 철자 불일치로 판단하여 활용 오류가 아닌 오철자 오류로 처리한다는 내용을 추가하였다.
- ‘되다/하다’의 경우 어미와 결합할 때, ‘되/돼’, ‘하/해’로 형태를 바꾸기 때문에 오형태 활용 오류에 해당한다. 그러나 오류 위치를 어간 어미 중 무엇으로 처리하느냐가 문제가 된다. ‘해(√하), 돼(√되)’로 잘못 쓴 경우는 형태주석에서 ‘해’와 ‘돼’를 분리하여 처리하지 않기 때문에 용언의 활용 오류로 처리한다. 그러나 ‘하(√해), 되(√돼)’의 경우는 학습자가 용언 어간, 어미 둘 중 어느 곳을 잘못 사용하였는지 정확히 분리하기 어려워 어간과 어미 양쪽 모두 오형태 활용 오류로 주석한다는 내용을 첨가하였다.
- 그밖에 수정된 사항에 대해 보다 적절한 예문으로 교체하거나 추가된 보완 사항에 해당하는 예문들을 추가하였다.

다. 품사(POS) 오류 주석

<표 15> 지침 보완 내용: 품사(POS) 오류

	2차 연도 지침	3차 연도 지침
품사 오류 정의	○ 품사 오류는 동일 의미 단어의 품사를 잘못 사용한 경우를 말한다. 즉, 같은 의미인 단어를 동사형으로 사용해야 하는데 명사형으로 잘못 사용하거나, 명사형을 동사형으로 잘못 사용한 경우가 해당된다.	○ 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.
품사 오류 범위		○ 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, 품사가 달라진 것보다 품사를 모르고 있는 것을 우선적으로 적용하여, 품사에 대한 인식이 없어서 생겨난 오류만 품사 오류로 보았다. 따라서 문맥에 따라 교정 어절이 바뀌면서 품사가 달라진 경우(단순히 원어절과 교정어절의 품사가 상이한 경우)는 품사 오류로 처리하지 않는다. ○ 즉, 원어절과 교정 어절이 의미를 공유하면서 품사를 제대로 사용하지 못한 경우, 품사에 대한 인식이 부재한 것으로 간주하고 이를 품사 오류로 처리한다.
품사		○ 품사 오류에서 ‘파생/합성’과 관련된

	2차 연도 지침	3차 연도 지침
오류와 합성 /파생 중복주 석		<p>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.</p> <p>○ 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.</p> <p>○ 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.</p>
‘외국어 +하다’ 처리		<p>○ ‘외국어’에 파생접사 ‘하다’를 결합하여 동사/형용사를 만들 수 있는데, 외국어 명사만 사용한 경우도 동일하게 품사 오류와 파생 오류로 중복 주석한다.</p>

- 2016년 오류 주석에서 품사 오류는 동일 의미의 명사형과 동사형만으로 제한하였으나 명사-동사 대치 외에도 품사를 바꾸어 사용하는 경우 모두

를 품사 오류로 확장하는 것으로 수정하였다.

- 연결어미 앞에서 ‘명사+하다’ 동사에서 파생접사 ‘하’의 누락 또는 ‘명사+있다’ 형용사에서 ‘있’을 누락한 경우 어떤 오류로 처리할지 문제가 된다. 이 경우, 학습자의 오류 원인을 예측해봤을 때, 크게 3가지로 해석할 수 있다.¹⁾ 이 중 원어절과 교정어절상 표면적으로 드러나는 것은 동사를 써야하는데 명사를 썼기 때문에 품사 대치 오류로 우선 처리하는 것으로 결정하였다. 즉, 원어절과 교정어절에서 드러나는 차이에 주목하여 품사 오류로 처리한다. 동시에 이러한 품사 오류는 조어법과도 긴밀하게 관련된다. 파생과 합성을 하면서 품사가 바뀌게 되기 때문에 파생/합성에는 품사의 의미도 포함되어 있다고 볼 수 있는 것이다. 따라서 이러한 오류에 대해서는 품사를 바꾸는 접미사에 대한 인식이 없다고 보고 오류 층위에서 ‘품사 대치’ 오류로 주석한 후, ‘파생/합성’ 오류도 함께 주석한다.
- 위의 논의를 확장하여, ‘N+하다, 되다, 시키다, 있다, 없다, 나다’에서 어간만 사용한 경우, 오류 위치[명사 또는 어근] - 오류 양상[대치] - 오류 층위[품사, 파생/합성(파생접사가 아닌 경우)] 중복 주석 처리한다는 내용을 추가하였다.
- 지정사 ‘이다’와 파생접미사 ‘하다’가 대치될 경우, 파생을 잘못된 오류와도 관련된다. 따라서 ‘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 것도 모르는 것으로 판단하여 오류 층위에 품사와 파생을 중복 주석한다는 내용을 추가하였다.
- 그밖에 수정된 사항에 대해 보다 적절한 예문으로 교체하거나 추가된 보완 사항에 해당하는 예문들을 추가하였다.

1) 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 있음

② 오류 층위에서의 통사 오류 처리 지침 보완

가. 높임(SH) 오류 주석

<표 16> 지침 보완 내용: 높임(SH) 오류

	2차 연도 지침	3차 연도 지침
높임 오류 범위	관련 내용 없음	<p>○ 높임법은 주체 높임법과 객체 높임법, 상대 높임법이 있으며, 이는 다시 문법적 높임과 어휘적 높임으로 나뉜다. 이 중 문법적 높임 오류는 주격조사와 서술에서의 높임 호응 관계 불일치, 주체 높임을 나타내는 선어말어미 ‘-시-’의 잘못된 사용, 상대 높임을 나타내는 대명사와 종결어미의 호응 관계 불일치, 객체 높임을 나타내는 부사격조사 ‘께’의 잘못된 사용이 해당되며, 어휘적 높임 오류는 ‘계시다, 드리다, 모시다, 잠수시다, 주무시다’ 등의 특수한 높임말을 써야 할 자리에 쓰지 않은 경우 또는 그 반대의 경우가 해당된다.</p> <p>○ 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.</p> <p>○ 따라서 주된 높임 오류의 대상이 되는 품사는 대명사, 조사(주격조사 ‘께서’, 부사격 조사 ‘께’), 선어말어미, 종결어미이다.</p> <p>○ 높임의 오류에는 높임법을 써야 하는 환경에서 낮춤말을 쓴 경우와 반대로 낮춤말</p>

	2차 연도 지침	3차 연도 지침
		을 써야 하는 환경에서 높임말을 사용한 경우를 모두 포함한다.
높임 오류 세부 처리 방안		<ul style="list-style-type: none"> ○ 문장 이상의 단위인 담화 상에서 높임말과 반말을 혼용하여 사용하고 있는 경우에는 문장 단위로 오류를 판단한다는 원칙하에 한 문장 내에서 높임 표현에 문제가 없으면 오류로 주석하지 않는다. ○ 상대높임법에서 대명사와 종결어미 양쪽 모두 교정이 가능할 때에는 종결어미를 기준으로 대명사 오류로 일괄 처리한다. 즉, 종결어미에 따라 대명사 ‘나’와 ‘저’의 대치 오류로 우선 처리한다.

- 높임 오류를 무엇으로 볼 것인가는 높임 오류의 쟁점이 된다. 따라서 높임 오류의 범위를 명시하고, 문법적 높임과 어휘적 높임 중, 문법적 높임을 우선 적용한다는 내용을 추가하였다.
- 높임 오류는 적절성과 용인 가능성이 주석자마다 다르게 적용될 수 있어, 세부적 오류 처리기준을 명확히 하기 위해 동일 문장 내 높임 표현의 호응을 필수적인 높임 오류 판정 기준으로 삼았다. 그밖에 세부 높임 표현 오류 처리 기준으로 상대높임법에서 대명사와 종결어미 양쪽 모두 교정이 가능할 때에는 종결어미를 기준으로 대명사 오류로 일괄 처리한다는 내용을 추가하였다.

나. 시제(ST) 오류 주석

<표 17> 지침 보완 내용: 시제(ST) 오류

	2차 연도 지침	3차 연도 지침
시제 오류 범위	○ 시제 오류는 시제 또는 시상을 나타내기 위한 선어말어미 오류 등을 말한다.	○ 시제 오류는 시제 또는 시상을 나타내기 위한 선어말어미, 관형사형 전성어미, ‘-(으)ㄴ 것’ 등의 오류 등을 말한다.

	2차 연도 지침	3차 연도 지침
시제 오류 범위	<p>○ 시제 오류는 오류 양상을 대치 오류로 주석한다. 형태상으로 보면 누락이지만 시제를 인식하지 못한 것으로 판단하여, 현재와 과거, 현재와 미래 등으로 대치한 것으로 판단한다. 단, 동사 기본형을 썼을 경우에는 누락 오류로 주석한다. 그 외에 과거 시제나 미래 시제 자리에 현재 시제를 사용한 경우, 그 반대의 경우 등은 대치 오류로 처리한다. 형용사, 연결어미의 시제는 현재 시제로 인식하고 대치로 주석한다.</p>	<p>○ 시제 오류는 오류 양상을 두 가지로 처리한다. 하나는 시제 요소를 사용하지 않고 기본형을 사용했을 경우, 시제를 사용하지 않았다고 판단하여 누락 오류로 주석한다. 다른 하나는 시제를 제대로 인식하지 못하여 과거 시제나 미래 시제 자리에 현재 시제를 사용한 경우, 그 반대의 경우 등은 대치 오류로 처리한다. 즉, 시제를 사용했으나 현재와 과거, 과거와 미래처럼 잘못 사용한 경우는 시제 간 대치로 처리한다.</p> <p>○ 이때, ‘이다/아니다’, 형용사, 연결어미 앞에서의 용언은 기본형이 현재를 나타내기 때문에 기본형을 사용했을 경우, 현재 시제로 인식하고 대치로 주석한다.</p>
시제 오류 세부 처리 방안		<p>○ 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 오류로 처리한다.</p> <p>○ 미래시제를 나타내는 ‘-겠-’의 경우, 추측이나 의지와 같은 양태 의미와도 관련되는데, 본 연구에서는 추측을 제외한 경우 오류 층위에 시제를 주석한다.</p> <p>○ 관형사형 전성어미 ‘-(으)ㄴ, 는, (으)ㄹ’의 경우, 뒤에 오는 (의</p>

	2차 연도 지침	3차 연도 지침
		<p>존)명사에 따라 시제를 나타내는 경우가 있고 그렇지 않은 경우가 있다. 시제를 나타내지 않을 경우는 대치 또는 오형태 활용 오류로 처리하고, 시제를 나타낼 때에는 관형사형 전성어미의 시제 대치 오류로 처리한다.</p> <p>○ 시제를 나타내지 않는 경우로 ‘-(으)르 때, -(으)르 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.</p> <p>○ 관형사형 전성어미에서 시제 대치 오류 판단에 어려움이 있을 경우, 현재(‘-는’)와 미래(‘-(으)르’)가 둘 다 가능할 때에는 용인 가능한 것으로 판단하여 오류로 처리하지 않고, 명확하게 과거형을 써야 하는데 쓰지 않은 경우나 반대의 경우 시제 오류로 처리한다.</p> <p>○ 연결어미 ‘-(으)ㄴ지/는지/-(으)르지’는 시제 대치 오류로 처리하지 않도록 주의한다.</p>

- 시제 오류는 선어말 어미(‘-었-’, ‘-겠-’) 외에 관형사형 전성어미에 의해서도 나타나기 때문에 시제 오류 정의에서 관형사형 전성 어미와, ‘-(으)르 것’과 같은 표현도 시제 관련 오류로 처리한다는 내용을 추가하였다.
- 시제 오류 양상을 처리하는 데 있어, 학습자가 시제를 인식하고 있는지, 없는지에 따라 대치와 누락으로 처리한다. 그렇기 때문에 동사의 기본형을

사용했을 때는 시제를 인식하지 못한 것으로 판단하여 시제 누락으로 처리하고 나머지는 대치 오류로 주석한다. 그러나 연결어미에서 시제 선어말어미의 문법적 제약이 있는데 ‘-었-’과 ‘-겠-’을 썼을 때는 첨가 오류로 처리한다는 내용을 추가하였다.

- 관형사형 전성어미의 경우, 시제를 나타내는 경우와 그렇지 않은 경우를 구분하는 데 어려움이 있다. 뒤에 오는 (의존)명사에 따라 시제를 나타내는 경우가 있고 그렇지 않은 경우가 있기 때문이다. 따라서 ‘-(으)ㄴ/는/(으)ㄴ’이 대치될 때 오류 층위에서 시제로 판단하는 경우를 제한하였다.

다. 사동(SC)과 피동(SP) 오류 주석

<표 18> 지침 보완 내용: 사동(SC)과 피동(SP) 오류

	2차 연도 지침	3차 연도 지침
사동 오류 범위	관련 내용 없음	<ul style="list-style-type: none"> ○ 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.
피동 오류 범위	관련 내용 없음	<ul style="list-style-type: none"> ○ 피동 표현은 접미사 ‘-이/히/리/기-’에 의한 피동, ‘-아/어지다, -게 되다’에 의한 통사적 피동, ‘-되다, 받다, 당하다’ 등 어휘적 피동으로 나

	2차 연도 지침	3차 연도 지침
		<p>타낼 수 있다. 이 연구에서는 형태를 중심으로 접미사 ‘-이/히/리/기-’에 의한 피동사와 ‘-어지다’ 피동 표현에 의한 피동으로 제한한다.</p> <p>○ 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 학습기관에서 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.</p> <p>○ ‘-게 되다’의 경우도 변화의 의미를 나타내는 경우가 많으며, 대응되는 능동문이 없을 때도 존재하며, ‘-게 하다’ 사동 표현과 짝을 이루지 않는 경우도 있다는 근거로 피동에 포함시키지 않는 논의에 근거해 본 연구에서도 제외하였다.</p> <p>○ 즉, ‘형용사+아/어지다’와 ‘-게 되다’는 기본 의미를 변화로 보고, 피동으로 다루지 않으며, 맥락에 따라 다르게 판단할 수 있는 어휘적 피동도 제외한다.</p>
사동 피동 오류		<p>○ 사동사, 사동 표현에서 나타난 오류를 모두 볼 수 있도록 오류 양상에 관계없이, 철자를 잘못 사용한 오형</p>

	2차 연도 지침	3차 연도 지침
세부 처리 방안		<p>태 오류도 오류 층위에서 사동으로 주석한다.</p> <p>○ 원어절에서 사동사, 사동 표현을 사용한 경우와 교정 어절이 사동사, 사동 표현이어야 하는 경우 모두 오류 층위에서 사동으로 주석한다.</p> <p>○ 사동을 쓸 자리에 피동을 썼거나 반대의 경우는 사동과 피동으로 중복 주석한다. 오류 위치와 오류 양상은 원어절 기준으로 주석하지만 오류 층위는 원어절과 교정어절 양쪽에서 주석함에 따라 사동과 피동을 중복 주석한다.</p> <p>○ 이중피동을 사용한 경우, 첨가 오류로 처리한다.</p>
오류 위치		<p>○ 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.</p> <p>○ 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</p>

	2차 연도 지침	3차 연도 지침
		오류 위치는 형태 주석에 따라 일괄 처리하여, 동사로 분석했을 때는 그 품사를 따르고, 연결어미, 보조 용언으로 분리하였을 경우 해당 품사를 오류 위치로 주석한다.
교정어절에 따른 처리	○ 오류 영역에서 교정어절로 인해 조사나 어미가 바뀌는 경우는 오류로 처리하지 않는다. 기본적으로 맥락을 살펴 되도록 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.	○ 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 주석한다.

- 사동과 피동은 크게 문법적 사동/피동과 어휘적 사동/피동 표현으로 나뉜다. 문법적 사동/피동에 비해 의미를 고려한 어휘적 사동/피동의 경우, 그 대상을 한정 짓기 어렵다. 주석자 간의 일관성을 고려하여 문법적 사동/피동을 사동과 피동 오류의 범위로 한정하였다. 그러나 어휘적 사동에 해당하는 ‘시키다’의 경우 한국어 교육 현장에서 사동으로 교육하고 있다는 점을 고려하여 사동 오류 범위에 포함하였다. 그밖에 ‘형용사+-아/어지다’, ‘-게 되다’와 같이 변화의 의미를 나타내는 표현은 피동 오류로 처리하지 않는다는 내용을 추가하였다.
- 사동을 쓸 자리에 피동을 썼거나 반대의 경우, 오류 층위를 원어절과 교

정 어절 기준 가운데 무엇을 기준으로 삼을 것인가가 문제가 된다. 기본적으로 본 연구에서는 원어절과 교정어절 양쪽을 모두 고려하여 오류 층위를 주석하기 때문에 이러한 경우 사동과 피동을 중복 주석한다는 내용을 추가하였다.

- 오류 위치와 관련하여 사동 표현 ‘-게 하다’와 피동 표현 ‘-어지다’에서 오류가 발생했을 경우, 오류 위치를 해당 품사로 주석할지, 표현 문형으로 주석할지 문제가 된다. 이때에는 형태 주석에 따라 일관되게 처리함을 원칙으로 한다. 따라서 ‘-게 하다’의 경우, 용언의 품사(연결어미+보조용언)를 주석하는 동시에 표현 문형 목록에 해당하기 때문에 표현문형(PE)도 중복 주석한다. ‘-어지다’의 경우 <표준>을 근거로 형태 주석에서 분리하거나 하나의 동사로 처리하기 때문에 형태 주석에서 처리한대로 오류 위치를 주석한다는 내용을 추가하였다.
- 피동 오류에서 이중 피동 표현도 오류 대상으로 삼는다는 내용을 추가하였다.
- 2차 연도 지침에서는 사동과 피동 오류에서 용언보다 격조사를 대치시키는 것을 우선적으로 처리하고 오류 층위에서 사동과 피동을 주석하지 않았다. 또한 용언을 바꾸는 것으로 인해 조사가 달라지는 경우, 조사 밑에 교정어절만 써주고 오류로 처리하지 않았다. 그러나 3차 연도 지침에서는 조사를 우선 교정할 때에도 사동과 피동의 구조를 모르고 있다는 측면에서 조사 밑에도 오류 층위에 사동과 피동을 주석하는 것으로 수정하였다.

라. 부정(SN) 오류 주석

<표 19> 지침 보완 내용: 부정(SN) 오류

	2차 연도 지침	3차 연도 지침
부정 오류 범위	관련 내용 없음	<ul style="list-style-type: none"> ○ 일반적으로 부사 ‘아니(안), 못’이나 부정의 의미를 가진 용언 ‘아니다, 아니하다(않다), 못하다, 말다’를 써서 부정문을 만드는 방법에 근거하여 부정 부사를 잘못 사용하거나 해당 용언에서 오류가 났을 경우, 오류 층위에 부정(SN)을 주석한다. ○ ‘없다, 모르다’, 부정 의미의 접두사는

	2차 연도 지침	3차 연도 지침
		<p>부정 오류에 포함하지 않는다.</p> <p>○ 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.)</p> <p>○ 장형부정이 더 자연스럽지만 단형부정을 썼을 때 용인가능하기도 하다. 따라서 단형부정을 장형부정으로 반드시 바꿔야하는 경우 기준 마련 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.</p>
부정 오류 세부 처리 방안	관련 내용 없음	<p>○ ‘-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.</p>

- 부정 오류의 경우, 형태 기준으로 처리할 것인지 의미 기준으로 처리할 것인지에 따라 오류의 범위가 달라질 수 있다. 본 연구에서는 부정을 나타내는 부사 ‘아니(안), 못’이나 부정의 의미를 가진 용언 ‘아니다, 아니하다(않다), 못하다, 말다’를 써서 부정문을 만드는 부정법에 근거하여 부정 오류로 처리한다는 내용을 추가하였다. 그밖에 ‘없다, 모르다’, 부정 의미의 접두사는 부정 오류에 포함하지 않는 것으로 부정 오류의 범위를 제한하

였다.

- 부정의 경우, 장형부정이 더 자연스럽지만 단형부정을 써도 용인 가능하다고 판단할 수 있다. 따라서 단형부정을 장형부정으로 반드시 바꿔야 하는 경우, 즉 부정 오류로 처리할 때 그 기준을 명확히 할 필요가 있다. 이를 위해서 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 한다는 내용을 추가하였다.
- ‘N+하다’ 파생동사 앞에 부정부사 ‘안’을 사용한 경우는 하다 앞에 부정부사 ‘안’을 써주는 것과 장형부정 ‘-지 않다’ 두 가지로 교정이 가능하다. 이러한 경우, 최소 수정 원칙에 의해서 단형부정을 장형부정으로 바꾸는 것보다 단형부정의 위치를 잘못 사용한 것으로 보고 오류 층위에서 어순 오류와 부정 오류로 중복 주석한다는 세부 처리 기준을 추가하였다.

마. 어순(WO) 오류 주석

<표 20> 지침 보완 내용: 어순(WO) 오류

	2차 연도 지침	3차 연도 지침
어순 오류 범위	관련 내용 없음	<ul style="list-style-type: none"> ○ 한국어 어순의 특징 중 하나는 문장성분의 자리 이동이 비교적 자유롭다는 것이다. 그렇기 때문에 그만큼 용인가능성이 크다고 할 수 있다. 이에 따라 주석자 간의 일치도도 다르게 나타날 수 있어, 어순 오류의 경우 최소 수정의 기준을 마련하여 처리한다. ○ 문장 부사는 자리 이동이 자유롭지만 성분 부사의 경우는 제한되기 때 특정한 성분을 수식해야 하는 성분부사의 위치를 잘못 사용했을 때는 오류로 주석한다. ○ 관형사의 경우, ‘지시관형사-수관형사-성상관형사’ 순의 기준을 적용하여 처리한다.
어순	관련 내용 없음	<ul style="list-style-type: none"> ○ 시간을 나타내는 표현의 배열이 잘못되었

	2차 연도 지침	3차 연도 지침
오류 세부 처리 방안		<p>을 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.</p> <p>○ 2개의 문장 성분이 상호 교체될 때에는 2개 모두 대치 어순 오류로 주석한다.</p> <p>○ 조사의 경우, 어순 오류로 처리하지 않고 조사 첨가 또는 누락 오류로 처리한다.</p>

- 어순 오류에서의 쟁점은 적절성에 따라 주석자 간 용인 가능성이 달라져 오류 판단이 다를 수 있다는 점이다. 이를 위해서 어순 오류의 판단 기준을 마련하기 위하여 성분부사의 위치, 관형사의 순서 등 특정 어순의 기준을 추가하였다.
- 어순 오류의 세부 처리 방안으로 시간 표현, 조사 외 기본 어순 위배 등에 대한 주석자 간 용인 가능성 판단 조사를 통해 지침을 추가 보완하였다.

③ 오류 층위에서의 담화 오류 처리 지침 보완

<표 21> 지침 보완 내용: 담화 오류

	2차 연도 지침	3차 연도 지침
지시 오류	관련 내용 없음	○ 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.
접속 오류	관련 내용 없음	○ 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.

	2차 연도 지침	3차 연도 지침
담화 표지 오류	관련 내용 없음	○ 담화표지는 미시 담화표지와 거시 담화표지로 나눌 수 있으나 연구자마다 그 정의가 다르고, 해당 형태도 다르기 때문에 담화표지의 목록을 마련하기 쉽지 않다. 이러한 이유로 미시 담화표지에 초점을 두고 오류를 판단하도록 한다. 미시 담화표지의 경우, 학습자의 L1의 영향으로 인한 간투사 사용과 모어화자와는 다른 위치에서 담화표지를 사용한 경우를 오류로 주석한다.
구어/ 문어 오류	○ 구어체/문어체, 격식체/비격식체의 혼용에 의해 담화 맥락에서의 일관성이 떨어지는 경우를 말한다.	○ 구어체(구어성)/문어체(문어성), 격식체/비격식체의 혼용에 의해 담화 맥락에서의 일관성이 떨어지는 경우를 말한다.

- 본 연구는 문장 내에서의 오류 판단을 기본 원칙으로 하기 때문에 문장 단위를 넘어서는 담화 오류는 최소한으로 제한하여 주석함을 명시하였다. 그러나 지시(DR)와 접속(DC) 오류에 한해서 선행문과 후행문과의 의미적 연결을 고려해 오류 여부를 판단한다는 내용을 추가하였다.
- 담화표지 오류의 경우, 무엇을 담화표지로 정의하며 담화표지의 오류가 무엇인지 그 기준을 마련할 필요가 있다. 본 연구에서는 미시 담화표지에 초점을 두고, 학습자의 L1의 영향으로 인한 간투사 사용과 모어 화자와는 다른 위치에서 담화표지를 사용한 경우를 오류로 주석한다는 기준을 추가하였다.
- 구어/문어 오류는 사용역에 따라 용인 가능성이 달리 적용할 수 있기 때문에 엄격하게 그 기준을 적용하기가 어렵다. 이에 구어체(구어성)/문어체(문어성)를 판단하는 기준은 <표준국어대사전>으로 삼고, 해당하는 목록을 일부 추가하였다.

④ 구어 오류 처리 기준 보완

- 구어에서는 발음 오류와 어휘 및 문법 오류의 구분이 명확하지 않을 수 있다. 즉, 학습자가 어휘와 문법을 잘못 사용한 것인지 단순히 발음을 잘못한 것으로 인해 나타난 오류인지 판별하는 데 어려움이 있다. 구어에서는 발음의 영향과 함께 어휘 및 문법 오류를 표시해주는 차원에서 오류 층위에서 중복 주석을 하도록 한다. 그러나 조사의 경우, 문법 오류를 우선 처리하도록 하고, 관형사형 전성어미를 사용해야 할 자리에 사용하지 못했을 경우도 받침을 발음하지 못한 음소 오류보다는 누락 오류로 문법 오류를 우선 처리한다는 내용을 추가하였다.
- 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.

<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.

- 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 <표준국어대사전>에서 ‘~의 준말’로 등재되어 있는 것을 기준으로 삼았다. 그리하여 <표준>을 기준으로 ‘준말’로 등재되어 있는 형태는 오류로 처리하지 않고, 등재되지 않은 형태는 오류로 처리한다는 내용을 추가하였다.

<예> 그래서 맘 먹고 여기 왔어요.

- ☞ ‘맘’의 경우, <표준>에 ‘마음의 준말’로 등재되어 있다. 따라서 오류로 처리하지 않는다.
- ☞ 젤(제일), 담(다음), 침(처음) 등과 같이 ‘준말’로 등재되어 있는 형태들은 오류로 처리하지 않는다.

- 구어에서 발음의 차원이 아닌, 형태를 잘못 발화한 경우는 오형태(MIF) 오류로 처리한다.

- <예> 사잉(√사건)
보석필(√보살핌)
- ☞ 이와 같은 예시들은 발음을 잘못했다기보다 형태를 잘못 만들어낸 것이다. 유사 발음과도 떨어져 음소 오류로 볼 수 없고, 한국어에 없는 형태들을 발음했다고 판단하여 오형태 오류로 처리한다.

- 구어 오류는 오류 층위에서 발음 오류와 가장 밀접하다. 발음 오류에는 음소(PP), 음절(PS), 음운규칙(PC), 원어식 발음(PN), 중간 발음(변이음 포함(PA))이 있는데, 각각 해당하는 오류의 유형을 다음과 같이 구분하여 처리한다는 내용을 추가하였다.

가. 음소(PP) 오류 주석

- 명확하게 다른 음소로 발음한 경우와 단어 내에서 음소를 발음하지 못한 경우, 음소 오류로 주석한다.

- <예> 성생닌(√선생님)
☞ 받침을 모두 다른 음소로 발음한 음소 오류로 처리한다.
- 부모니(√부모님)
☞ 한 단어 안에서 받침을 발음하지 못한 경우도 마찬가지로 음소 오류로 주석한다.

나. 음절(PS) 오류 주석

- 음절 오류는 음절 단위에서 발화가 잘못 사용된 경우를 말한다. 음절 오류는 원래 음절보다 적게 또는 더 많이 발화한 경우가 해당된다.

<예> 기무치(√ 김치)
해스므니다(√ 했습니다)
☞ 받침에 ‘으/우’를 결합하여 한 음절을 더 많이 발음하고 있는 형태들은 음절 오류로 주석한다.

다. 음운규칙(PC) 오류 주석

- 음운규칙 오류는 음운규칙을 적용하지 않은 경우와 적용했으나 잘못 발음한 경우 2가지로 나눌 수 있다.
- 음운규칙 미적용 오류는 음운규칙으로 인해 철자와 다르게 발음해야 하나, 학습자가 철자대로 절음화 하여 발음한 경우다. 학습자가 철자대로 발음했기 때문에 원어절과 교정어절의 형태는 동일하다.

<예> 무조건(√ 무조건) ☞ 경음화 미적용
같이(√ 같이) ☞ 구개음화 미적용
신라(√ 신라) ☞ 유음화 미적용
앞에(√ 앞에) ☞ 연음 미적용
먹는(√ 먹는) ☞ 비음화 미적용

- 음운규칙 미적용 외에 음운규칙을 적용시켜야 하는 단어에서 잘못 적용한 경우도 음운규칙 오류로 주석한다. 그러나 이때에는 음운규칙과 음소 오류를 중복 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)).

라. 원어식 발음(PN) 오류 주석

- 원어식 발음은 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 이를 표시해 주는 차원에서 외국어와 외래어의 경우, 한국어와 다르게 발음했을 경우 ‘원어식 발음’ 오류로 주석한다.

<예> 인텔류(√인터류)
 세너(√센터)
☞ [학쿄]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)).

- 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> 팔너(√파트너)
 그대 처음에 갈 뻔남(√베트남)에서
☞ 원어식 발음의 차이로 인해 한국어의 외래어 표기와 다르게 음절이 줄어들었다. 이때는 음절 오류가 아닌 원어식 발음 오류로 주석한다.

마. 중간발음(변이음 포함)(PA) 오류 주석

- 중간발음은 한국어 모어 화자와는 다른 발음, 변이음과 관련된 오류로, 구어 전사 과정에서 유성음, 무성음으로 표기해준 경우에 근거해 변이음 오류로 처리한다. 이때는 음성과 관련된 문제로, 원어절과 교정어절의 형태는 동일하다.

<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.

- 변이음 외에 학습자가 원어절과 교정어절 사이의 발음, 즉 음소 간의 중간 형태로 발음한 경우도 중간발음 오류로 처리한다.

<예> 요자(√여자)

- ☞ ‘ㄷ’의 ‘ㄲ’의 중간 발음

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. 말뭉치 활용 연구 체계 구축

2.1. 2015-2017년 구축 말뭉치를 활용한 분석 연구²⁾

- 한국어 학습자 말뭉치는 다양한 변인에 따른 학습자의 중간언어의 특성이나 언어 발달 양상을 관찰하기 위한 연구 자료로의 활용도가 무엇보다도 가장 높다. 이에 따라 본 절에서는 2015-2017년까지 구축된 형태 주석 말

2) 본 장의 분석 결과는 2017년 12월 8일 한국어 학습자 말뭉치 구축 도구에 구축된 자료의 통계를 기준으로 한 것으로 향후 심화 검수 과정에서 변경될 수 있음

묻치와 오류 주석 말뭉치의 주석 통계를 말뭉치 유형별(구어/문어), 숙달도 단계별, 국적별로 분석하여 제시하기로 한다.

(1) 형태 주석 결과의 분석

① 문어

가. 숙달도 단계별

<표 22> 문어 형태 주석 결과_숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	합계
일반명사	65,149	69,632	73,543	86,017	78,187	75,241	26,873	474,642
고유명사	14,539	12,375	6,678	3,379	2,705	4,680	2,882	47,238
의존명사	10,159	10,482	12,161	13,606	11,545	11,543	2,892	72,388
대명사	11,976	11,321	9,304	8,366	4,712	4,355	753	50,787
수사	1,368	602	603	470	476	379	208	4,106
동사	30,007	33,656	35,718	38,476	29,534	26,291	5,818	199,500
형용사	9,789	14,675	12,942	13,475	11,036	7,814	1,396	71,127
보조용언	4,601	5,698	7,434	8,258	6,536	6,066	1,331	39,924
부정지정사	89	209	383	687	721	762	127	2,978
긍정지정사	5,534	4,135	4,736	5,411	5,213	5,009	1,523	31,561
관형사	2,329	3,353	4,850	5,234	4,495	4,249	1,098	25,608
일반부사	13,643	17,998	17,370	16,392	11,790	9,804	2,221	89,218
접속부사	5,328	4,685	4,012	3,379	2,292	1,825	348	21,869
감탄사	29	70	32	30	19	18	12	210
주격조사	9,326	12,565	13,471	14,642	12,573	10,611	2,383	75,571
접속조사	2,363	1,916	1,611	1,612	1,681	1,505	891	11,579
보격조사	127	354	476	927	949	1,019	191	4,043
관형격조사	2,733	3,013	3,043	4,273	4,383	4,796	2,887	25,128
목적격조사	14,477	12,318	12,714	15,285	12,332	12,328	3,625	83,079
부사격조사	23,906	21,570	18,527	17,628	14,776	14,102	5,021	115,530

	1급	2급	3급	4급	5급	6급	6급 이상	합계
호격조사	3	23	6	6	2	0	1	41
인용격조사	18	211	151	97	102	173	26	778
보조사	16,935	17,016	15,519	16,745	14,294	12,128	3,156	95,793
선어말어미	8,264	9,356	10,040	7,494	4,411	5,970	1,464	46,999
어말어미(연결)	14,664	23,931	29,552	34,275	27,738	25,201	5,663	161,024
어말어미(종결)	33,454	27,834	23,519	21,136	15,772	13,216	3,336	138,267
명사형정선어미	576	2,180	2,039	3,010	2,716	1,956	572	13,049
관형형전성어미	8,565	14,223	18,787	23,513	21,721	19,851	5,127	111,787
체인 접두사	35	24	57	422	367	463	211	1,579
명사파생접미사	2,226	3,221	3,876	5,877	6,043	6,543	2,183	29,969
동사파생접미사	5,248	6,941	8,914	11,360	11,013	11,251	3,553	58,280
형용사파생접미사	1,740	2,829	3,746	4,240	3,905	2,927	907	20,294
어근	732	932	941	1,158	1,232	1,116	418	6,529
분석불능	414	393	342	250	250	209	13	1,871
합계	320,346	349,741	357,097	387,130	325,521	303,401	89,110	2,132,346

나. 국적별

<표 23> 문어 형태 주석 결과_국적별

	중국	일본	대만	베트남	홍콩	기타	계
일반명사	238,432	67,369	23,398	18,767	17,122	109,554	474,642
고유명사	22,028	7,126	2,109	1,670	1,545	12,760	47,238
의존명사	34,146	12,068	3,374	2,787	2,578	17,435	72,388
대명사	25,386	5,918	2,034	2,464	1,517	13,468	50,787
수사	2,086	369	152	203	113	1183	4,106
동사	94,280	31,308	9,181	7,716	6,770	50,245	199,500
형용사	35,415	10,805	3,266	2,834	2,161	16,646	71,127
보조용언	18,488	7,209	1,811	1,631	1,281	9,504	39,924
부정지정사	1,326	477	172	115	129	759	2,978
궁정지정사	14,687	5,342	1,446	1,343	1,124	7,619	31,561

	중국	일본	대만	베트남	홍콩	기타	계
관형사	11,826	3,540	1,176	948	1,017	7,101	25,608
일반부사	46,044	11,151	4,374	3,625	3,062	20,962	89,218
접속부사	11,054	2,715	957	915	652	5,576	21,869
감탄사	122	22	6	5	5	50	210
주격조사	37,962	10,906	3,826	2,814	2,605	17,458	75,571
접속조사	5,010	1,871	572	543	463	3,120	11,579
보격조사	1,901	809	231	113	134	855	4,043
관형격조사	12,692	3,375	1,284	972	1,013	5,792	25,128
목적격조사	41,109	11,826	3,617	3,537	2,765	20,225	83,079
부사격조사	55,763	16,499	5,352	4,542	3,873	29,501	115,530
호격조사	28	2	1	4	0	6	41
인용격조사	275	164	50	19	28	242	778
보조사	45,562	16,503	4,132	3,902	3,168	22,526	95,793
선어말어미	20,838	7,014	2,207	2,039	1,630	13,271	46,999
어말어미(연결)	73,629	27,953	7,739	6,713	5,661	39,329	161,024
어말어미(종결)	71,414	16,939	5,942	5,694	4,245	34,033	138,267
명사형정선어미	5,938	2,204	591	539	423	3,354	13,049
관형형전성어미	52,350	19,953	5,569	4,267	3,966	25,682	111,787
체언접두사	892	221	113	21	54	278	1,579
명사과생접미사	14,803	3,868	1,570	993	1,098	7,637	29,969
동사과생접미사	28,716	9,088	3,017	2,670	2,102	12,687	58,280
형용사과생접미사	10,332	2,688	923	886	746	4,719	20,294
어근	3,138	926	320	331	267	1,547	6,529
분석불능	1,096	112	40	58	56	509	1,871
합계	1,038,768	318,340	100,552	85,680	73,373	515,633	2,132,346

② 구어

가. 숙달도 단계별

<표 24> 구어 형태 주석 결과_숙달도 단계별

	1급	2급	3급	4급	5급	6급	6급 이상	계
일반명사	9,472	6,235	5,754	6,546	10,836	11,573	1,832	52,248
고유명사	1,885	1,372	1,001	657	1,141	1,040	195	7,291
의존명사	1,535	1,039	951	1,087	1,852	2,256	366	9,086
대명사	1,682	939	943	911	919	1,617	426	7,437
수사	514	239	159	186	520	300	87	2,005
동사	5,142	3,506	3,173	2,962	4,274	5,429	840	25,326
형용사	1,676	1,668	1,301	1,049	1,199	2,298	236	9,427
보조용언	850	480	629	516	654	1,031	137	4,297
부정지정사	123	47	64	68	107	246	40	695
긍정지정사	726	465	427	505	792	993	157	4,065
관형사	836	529	660	756	1,151	1,628	253	5,813
일반부사	2,736	2,016	1,902	1,617	2,065	3,857	497	14,690
접속부사	793	614	512	551	606	1,030	143	4,249
감탄사	8,467	4,977	3,129	3,926	3,824	6,040	574	30,937
주격조사	1,174	996	1,033	988	1,590	1,726	143	7,650
접속조사	289	138	147	112	293	240	50	1,269
보격조사	25	17	49	42	71	110	18	332
관형격조사	216	140	184	226	493	469	44	1,772
목적격조사	1,106	811	790	875	1,396	1,226	162	6,366
부사격조사	2,522	1,752	1,418	1,272	2,302	2,195	325	11,786
호격조사	1	7	0	0	0	3	6	17
인용격조사	3	7	9	8	44	20	0	91
보조사	2,146	1,230	1,315	1,351	1,936	2,822	426	11,226
선어말어미	1,309	916	1,262	753	1,524	1,456	254	7,474
어말어미(연결)	2,835	2,012	2,777	2,590	3,437	5,333	664	19,648
어말어미(종결)	5,040	3,502	2,208	1,849	2,646	3,296	585	19,126

	1급	2급	3급	4급	5급	6급	6급 이상	계
명사형정선어미	75	95	115	126	232	245	28	916
관형형전성어미	1,367	1,127	1,418	1,608	2,431	2,929	371	11,251
체언접두사	10	4	3	11	65	111	2	206
명사파생접미사	414	268	299	421	782	1,058	75	3,317
동사파생접미사	917	577	744	988	1,499	1,613	187	6,525
형용사파생접미사	282	202	284	221	358	386	66	1,799
어근	103	67	111	77	93	166	22	639
분석불능	1,192	717	492	626	645	855	160	4,687
합계	57,463	38,711	35,263	35,481	51,777	65,597	9,371	293,663

나. 국적별

<표 25> 구어 형태 주석 결과_국적별

	중국	일본	대만	베트남	미국	기타	계
일반명사	24,474	8,068	4,416	2,448	2,035	10807	52,248
고유명사	3,075	1,007	587	567	281	1774	7,291
의존명사	4,049	1,600	768	437	394	1838	9,086
대명사	3,733	946	582	457	217	1502	7,437
수사	902	299	115	91	69	529	2,005
동사	11,631	4,171	1,849	1,428	1,010	5237	25,326
형용사	4,441	1,565	653	490	396	1882	9,427
보조용언	1,969	777	348	233	184	786	4,297
부정지정사	404	116	48	12	18	97	695
긍정지정사	1,769	834	332	171	138	821	4,065
관형사	2,677	918	583	266	225	1144	5,813
일반부사	7,302	1,999	1,186	790	704	2709	14,690
접속부사	1,912	622	407	240	129	939	4,249
감탄사	14,333	4,197	1,868	2,429	1,099	7011	30,937
주격조사	3,242	1,447	632	334	326	1669	7,650

	중국	일본	대만	베트남	미국	기타	계
접속조사	553	235	131	55	43	252	1,269
보격조사	148	80	32	6	8	58	332
관형격조사	808	293	181	53	73	364	1,772
목적격조사	2,676	1,169	485	255	320	1461	6,366
부사격조사	5,054	1,933	941	670	513	2675	11,786
호격조사	11	0	0	2	0	4	17
인용격조사	24	33	3	0	13	18	91
보조사	5,627	1,820	923	448	410	1998	11,226
선어말어미	3,344	1,218	462	427	335	1688	7,474
어말어미(연결)	9,018	3,748	1,555	818	775	3734	19,648
어말어미(종결)	8,883	2,660	1,302	1,389	751	4141	19,126
명사형정선어미	368	193	61	24	45	225	916
관형형전성어미	5,112	2,153	970	406	413	2197	11,251
체언접두사	107	33	35	0	5	26	206
명사파생접미사	1,551	531	275	59	148	753	3,317
동사파생접미사	3,028	1,096	578	312	208	1303	6,525
형용사파생접미사	884	264	151	90	66	344	1,799
어근	287	114	55	34	27	122	639
분석불능	2,464	534	325	307	103	954	4,687
합계	135,860	46,673	22,839	15,748	11,481	61,062	293,663

(2) 오류 주석 결과의 분석

① 문어

가. 분석 불가능 여부

ㄱ. 숙달도 단계별

<표 26> 문어 오류 주석 결과_분석 불가능 여부: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
분석 불가능	174	194	293	296	143	108	1,208

ㄴ. 국적별

<표 27> 문어 오류 주석 결과_분석 불가능 여부: 국적별

	중국	베트남	미국	러시아	대만	기타	합계
분석 불가능	595	148	106	61	56	242	1,208

나. 오류 위치

ㄱ. 숙달도 단계별

<표 28> 문어 오류 주석 결과_오류 위치: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
감탄사	1	1	0	0	1	0	3
일반부사	329	324	278	267	171	169	1,538
접속부사	145	78	64	49	31	23	390
관형사	33	42	65	65	50	41	296
의존명사	184	187	295	259	180	173	1,278
일반명사	1,181	984	1,203	1,269	1,019	1,167	6,823
고유명사	265	111	60	46	11	33	526
대명사	110	158	219	173	105	71	836
수사	34	15	12	10	15	15	101
형용사	235	191	205	165	108	111	1,015
지정사	68	58	59	57	40	55	337

	1급	2급	3급	4급	5급	6급	합계
동사	456	599	707	710	518	508	3,498
보조용언	55	71	121	112	76	84	519
접두사	0	1	4	6	6	9	26
어근	23	14	45	29	21	32	164
형용사파생접미사	7	10	12	18	11	14	72
명사파생접미사	32	24	40	56	73	42	267
동사파생접미사	29	38	64	83	77	91	382
관형사형 전성어미	125	239	414	352	299	298	1,727
부사격조사	1,159	844	743	684	456	427	4,313
보격조사	4	12	26	24	19	9	94
연결어미	345	632	828	685	470	429	3,389
종결어미	256	320	356	271	129	96	1,428
관형격조사	80	103	122	142	137	138	722
접속조사	282	121	106	65	51	33	658
명사형 전성어미	17	50	45	40	38	45	235
주격조사	589	596	704	684	495	455	3,523
목적격조사	568	521	576	547	405	384	3,001
선어말어미	223	472	393	272	155	173	1,688
인용격조사	1	5	11	2	3	6	28
호격조사	1	2	0	0	0	0	3
보조사	501	527	463	473	365	341	2,670
표현문형	374	629	862	758	558	489	3,670
구 단위 표현	26	24	46	49	78	56	279
합계	7,738	8,003	9,148	8,422	6,171	6,017	45,499

ㄴ. 국적별

<표 29> 문어 오류 주석 결과_오류 위치: 국적별

	중국	일본	미국	베트남	대만	기타	합계
감탄사	1	0	0	0	0	2	3
일반부사	489	227	183	135	131	373	1,538
접속부사	129	47	27	31	35	121	390
관형사	93	38	34	25	28	78	296
의존명사	385	163	126	137	127	340	1,278
일반명사	2,209	1,009	605	632	655	1,713	6,823
고유명사	130	75	95	36	43	147	526
대명사	253	94	80	114	78	217	836
수사	32	9	10	12	10	28	101
형용사	341	150	112	93	72	247	1,015
지정사	109	56	28	25	29	90	337
동사	1,041	507	367	313	322	948	3,498
보조용언	145	55	48	83	49	139	519
접두사	14	1	4	2	2	3	26
어근	57	16	17	18	20	36	164
형용사파생접미사	21	11	8	5	5	22	72
명사파생접미사	87	22	28	28	29	73	267
동사파생접미사	153	59	43	33	22	72	382
관형사형 전성어미	534	284	151	160	205	393	1,727
부사격조사	1,364	606	430	384	359	1,170	4,313
보격조사	21	11	6	11	10	35	94
연결어미	1,023	531	305	338	301	891	3,389
종결어미	592	121	124	102	134	355	1,428
관형격조사	244	90	66	68	84	170	722
접속조사	204	82	84	56	58	174	658
명사형 전성어미	79	25	24	40	20	47	235
주격조사	1,308	280	353	344	371	867	3,523

	중국	일본	미국	베트남	대만	기타	합계
목적격조사	1,107	233	301	287	292	781	3,001
선어말어미	613	184	134	145	216	396	1,688
인용격조사	9	5	1	0	4	9	28
호격조사	2	0	0	0	0	1	3
보조사	924	197	273	279	269	728	2,670
표현문형	1,206	399	346	450	304	965	3,670
구 단위 표현	77	19	33	22	26	102	279
합계	14,996	5,606	4,446	4,408	4,310	11,733	45,499

다. 오류 층위

ㄱ. 숙달도 단계별

<표 30> 문어 오류 주석 결과_오류 층위: 숙달도 단계별

		1급	2급	3급	4급	5급	6급	합계
형태	단어 형성 [합성법]	8	4	8	2	5	3	30
	단어 형성 [파생법]	90	79	104	70	68	67	478
	굴절[곡용]	137	84	66	61	49	35	432
	굴절[활용]	364	494	573	447	241	205	2,324
	품사	111	114	135	67	58	65	550
통사	높임	84	111	167	127	76	45	610
	시제	141	445	468	280	182	189	1,705
	사동	2	15	26	56	46	82	227
	피동	7	47	91	90	62	108	405
	부정	21	30	41	36	28	25	181
	어순	171	132	120	96	42	64	625
담화	지시	16	13	17	16	8	3	73
	접속	99	38	42	34	21	16	250
	담화표지	0	1	0	0	0	0	1
	구 어 / 문 어	533	352	293	206	138	113	1,635

	오류							
합계		1,784	1,959	2,151	1,588	1,024	1,020	9,526

ㄴ. 국적별

<표 31> 문어 오류 주식 결과_오류 총위: 국적별

		중국	일본	대만	베트남	미국	기타	합계
형태	단어 형성 [합성법]	7	1	4	8	4	6	30
	단어 형성 [파생법]	173	52	54	48	37	114	478
	굴절[곡용]	140	81	45	50	26	90	432
	굴절[활용]	833	307	245	177	195	567	2,324
	품사	161	56	77	74	50	132	550
통사	높임	201	62	58	75	65	149	610
	시제	604	231	216	135	144	375	1,705
	사동	65	16	37	28	17	64	227
	피동	87	77	45	39	51	106	405
	부정	44	34	16	28	16	43	181
	어순	149	82	32	99	72	191	625
담화	지시	7	9	9	9	11	28	73
	접속	69	37	24	24	19	77	250
	담화표지	0	0	0	1	0	0	1
	구 어 / 문 어 오류	523	271	147	112	173	409	1,635
합계		3,063	1,316	1,009	907	880	2,351	9,526

라. 오류 양상

ㄱ. 숙달도 단계별

<표 32> 문어 오류 주석 결과_오류 양상: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
누락	1,259	1,392	1,379	1,257	795	745	6,827
첨가	415	486	590	718	469	453	3,131
대치	3,169	3,577	4,427	3,998	3,197	3,139	21,507
오형태	2,745	2,223	2,334	2,071	1,423	1,432	12,228
합계	7,588	7,678	8,730	8,044	5,884	5,769	43,693

ㄴ. 국적별

<표 33> 문어 오류 주석 결과_오류 양상: 국적별

	중국	일본	미국	대만	베트남	기타	합계
누락	2,587	544	677	673	610	1,736	6,827
첨가	870	348	342	297	375	899	3,131
대치	6,861	2,785	2,092	2,027	2,252	5,490	21,507
오형태	4,292	1,770	1,123	1,118	874	3,051	12,228
합계	14,610	5,447	4,234	4,115	4,111	11,176	43,693

② 구어

가. 분석 불가능 여부

ㄱ. 숙달도 단계별

<표 34> 구어 오류 주석 결과_분석 불능 여부: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
분석 불가능	74	115	108	113	96	39	545

ㄴ. 국적별

<표 35> 구어 오류 주석 결과_분석 불능 여부: 국적별

	중국	일본	대만	베트남	미국	기타	합계
분석 불가능	147	134	58	39	34	133	545

나. 오류 위치

ㄱ. 숙달도 단계별

<표 36> 구어 오류 주석 결과_오류 위치: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
감탄사	4	1	1	0	6	4	16
일반부사	89	135	164	156	161	126	831
접속부사	42	25	22	19	41	24	173
관형사	18	39	29	60	60	39	245
의존명사	48	41	41	55	101	34	320
일반명사	597	674	570	771	1,191	676	4,479
고유명사	132	109	66	40	135	47	529
대명사	15	28	52	60	23	20	198
수사	42	16	4	14	76	13	165
형용사	45	94	93	74	73	43	422
지정사	10	8	10	15	26	9	78
동사	112	143	175	234	185	137	986
보조용언	7	11	26	18	33	9	104
접두사	1	1	0	3	0	10	15
어근	12	15	14	13	18	14	86
형용사파생접미사	0	3	5	2	2	1	13
명사파생접미사	30	20	21	24	40	17	152
동사파생접미사	5	9	11	14	23	10	72
관형사형 전성어미	39	75	96	122	130	63	525
부사격조사	138	151	132	137	169	69	796
보격조사	3	0	2	5	3	2	15

	1급	2급	3급	4급	5급	6급	합계
연결어미	78	117	210	181	148	124	858
종결어미	50	97	19	47	137	42	392
관형격조사	6	5	14	14	41	16	96
접속조사	17	9	12	9	23	5	75
명사형 전성어미	0	9	12	8	6	2	37
주격조사	56	95	135	162	144	101	693
목적격조사	76	131	91	108	136	110	652
선어말어미	40	50	84	37	56	26	293
인용격조사	0	1	0	0	2	1	4
호격조사	1	2	0	0	0	0	3
보조사	69	71	105	136	135	78	594
표현문형	27	66	93	68	80	69	403
구 단위 표현	4	0	9	9	7	0	29
합계	1,813	2,251	2,318	2,615	3,411	1,941	14,349

ㄴ. 국적별

<표 37> 구어 오류 주석 결과_오류 위치: 국적별

	중국	일본	대만	베트남	미국	기타	합계
감탄사	6	0	2	0	0	8	16
일반부사	247	202	77	79	35	191	831
접속부사	46	18	32	9	18	50	173
관형사	91	41	11	25	14	63	245
의존명사	82	93	32	30	8	75	320
일반명사	1,454	1,003	488	281	128	1,125	4,479
고유명사	188	129	69	51	8	84	529
대명사	66	39	16	5	9	63	198
수사	58	32	14	11	0	50	165
형용사	121	90	43	31	22	115	422
지정사	27	16	10	3	6	16	78
동사	278	186	101	67	40	314	986

	중국	일본	대만	베트남	미국	기타	합계
보조용언	43	19	11	6	6	19	104
접두사	8	4	2	0	0	1	15
어근	34	22	6	2	1	21	86
형용사파생접미사	5	4	2	1	0	1	13
명사파생접미사	56	21	26	11	2	36	152
동사파생접미사	26	14	11	6	5	10	72
관형사형 전성어미	169	96	80	70	7	103	525
부사격조사	235	120	99	81	28	233	796
보격조사	6	3	1	0	0	5	15
연결어미	282	210	89	63	35	179	858
종결어미	192	64	28	20	14	74	392
관형격조사	44	16	12	3	3	18	96
접속조사	18	7	23	1	3	23	75
명사형 전성어미	16	3	1	1	1	15	37
주격조사	260	97	90	45	24	177	693
목적격조사	210	137	69	51	17	168	652
선어말어미	131	25	30	26	18	63	293
인용격조사	1	2	0	0	1	0	4
호격조사	1	0	0	1	0	1	3
보조사	215	82	92	48	20	137	594
표현문형	152	54	61	21	22	93	403
구 단위 표현	15	0	0	0	3	11	29
합계	4,783	2,849	1,628	1,049	498	3,542	14,349

다. 오류 층위
 ㄱ. 숙달도 단계별

<표 38> 구어 오류 주석 결과_오류 층위: 숙달도 단계별

		1급	2급	3급	4급	5급	6급	합계
발음	음소	1,264	1,287	1,241	1,439	2,061	1,187	8,479
	음절	75	127	60	53	71	26	412
	음운규칙	70	123	104	101	88	64	550
	중간 발음 (변이음 포함)	5	21	3	2	17	0	48
	원어식 발음	44	29	14	40	83	7	217
형태	단어 형성 [합성법]	1	2	5	3	0	0	11
	단어 형성 [파생법]	3	15	16	22	9	9	74
	굴절[곡용]	13	25	21	28	64	19	170
	굴절[활용]	16	42	54	43	49	38	242
통사	높임	16	35	11	32	43	16	153
	품사	11	16	26	25	18	8	104
	사동	0	0	1	1	4	7	13
	부정	5	2	6	4	7	8	32
	피동	0	4	8	17	10	22	61
	시제	21	38	68	39	70	32	268
	어순	2	5	22	9	29	5	72
담화	지시	0	0	4	4	7	2	17
	접속	5	7	5	4	19	10	50
	담화표지	0	0	0	0	5	0	5
	구어/문어 오류	0	0	0	1	0	0	1
합계		1,551	1,778	1,669	1,867	2,654	1,460	10,979

ㄴ. 국적별

<표 39> 구어 오류 주석 결과_오류 층위: 국적별

		중국	일본	대만	베트남	미국	기타	합계
발음	음소	2,767	1,916	942	678	215	1,961	8,479
	음절	81	174	27	29	14	87	412
	음운규칙	199	117	43	35	24	132	550
	중간 발음 (변이음 포함)	27	0	11	0	1	9	48
	원어식 발음	33	71	19	4	38	52	217
형태	단어 형성 [합성법]	3	0	3	0	0	5	11
	단어 형성 [파생법]	20	3	12	10	3	26	74
	굴절[곡용]	50	28	55	2	6	29	170
	굴절[활용]	108	45	21	11	9	48	242
통사	높임	79	9	13	12	2	38	153
	품사	36	5	13	9	5	36	104
	사동	7	4	0	0	0	2	13
	부정	13	10	2	2	1	4	32
	피동	13	15	3	4	9	17	61
	시제	108	24	34	24	20	58	268
	어순	24	5	6	8	6	23	72
담화	지시	10	1	3	0	0	3	17
	접속	17	4	7	0	11	11	50
	담화표지	5	0	0	0	0	0	5
	구어/문어 오류	0	0	0	0	0	1	1
합계		3,600	2,431	1,214	828	364	2,542	10,979

라. 오류 양상

ㄱ. 숙달도 단계별

<표 40> 문어 오류 주석 결과_오류 양상: 숙달도 단계별

	1급	2급	3급	4급	5급	6급	합계
누락	102	168	239	215	240	133	1,097
첨가	41	52	75	108	109	71	456
대치	210	425	518	644	675	394	2,866
오형태	73	141	124	120	150	119	727
합계	426	786	956	1,087	1,174	717	5,146

ㄴ. 국적별

<표 41> 문어 오류 주석 결과_오류 양상: 국적별

	중국	일본	대만	베트남	미국	기타	합계
누락	405	120	109	86	63	314	1,097
첨가	130	89	58	28	24	127	456
대치	1,033	391	333	191	108	810	2,866
오형태	278	115	119	37	21	157	727
합계	1,846	715	619	342	216	1,408	5,146

2.2. 한국어 학습자 말뭉치 활용 실제 사례 연구

(1) 한국어 검색 서비스 활용 예시

① 한국어 학습자와 모어 화자의 언어 사용 양상 비교

- 한국어 학습자 말뭉치를 활용하여 한국어 학습자와 모어 화자의 언어 사용을 비교 분석할 수 있다. 실제 사례로서, 이미 기발표된 ‘모어 화자와 한국어 학습자 말뭉치에서의 보조사 ‘은/는’의 분포와 조사 선택 요인 분석’이 있다. 한국어 학습자와 모어 화자의 언어 사용 양상의 비교 분석을 위해, 21세기 세종 계획 문어 말뭉치에서의 조사 ‘이/가’와 ‘은/는’의 사용례를 검색한 후 이 중 무작위로 각 1,000개씩의 용례를 추출하며, 학습자의 보조사 ‘은/는’의 사용례를 확인하기 위해 다음과 같은 작업을 시행하였다.

[형태 주석 말뭉치 검색] -> [보조사, 은/는] 검색 -> [문어] 상세 검색 -> [내려받기]

<그림 1> 한국어 학습자 말뭉치 검색 화면 예시1

- 내려받은 파일에서 무작위로 2000개의 용례를 추출하여, 모어 화자와 학습자의 조사 결합구 분포와 사용 양상을 살펴보기 위해 이들 각 2,000개의 용례에 대해 결합 명사구와 후행 용언에 대한 코딩화 작업을 거쳤다. 이를 통해 모어 화자와 한국어 학습자가 ‘이/가’와 ‘은/는’ 결합구를 사용함에 있어, 이들 조사와 결합한 결합구의 문장에서의 위치, 즉 문두에 위치하는지 아닌지에 따라 조사 선택이 달리 나타나는지, 또 ‘이/가’와 ‘은/는’과 결합하는 명사의 특성에 따라 결합 양상을 달리 하는지, 그리고 ‘은/는’이 주어인 경우 ‘이/가’의 후행 용언과

그 분포가 달리 나타나는지 통계 분석을 실시하여 분포적 차이를 검증하였다. 또한 이러한 경향성이 모어 화자와 학습자 간에 차이가 나타나는지 등을 살펴 보기 위해 이항 로지스틱 회귀분석 등 다양한 통계 분석을 실시하여 선택 요인 에 차이가 있음을 검증한 바 있다.

② 학습자 변인에 따른 어휘 사용 양상 분석

○ 원시 말뭉치를 활용하여 한국어 학습자의 어휘 사용 양상을 숙달도와 학습자의 국적, 학습자의 모국어, 학습 목적, 나이, 말뭉치 유형, 국내외의 학습 환경 변인에 따라 상세히 살필 수 있다. 예를 들어, 국적이 미국이며 모어는 영어이며, 1급 학습자들의 어휘 ‘친구’의 사용 양상을 살펴보고자 하면 다음과 같은 작업을 시행하여 검색 결과를 문서로 내려받을 수 있다.

[illegible]

<그림 2> 한국어 학습자 말뭉치 검색 화면 예시2

표본 검색 결과				검색 내용 보기 ▼	20개씩 보기 ▼
번번	문역어	참조	문역		
1번	영어	1번	그동안 북 물에서 전국 정보 및 선례 .		
2번	영어	1번	수출에 전국 같이 한국 으로 이제까지 와 선례 .		
3번	영어	1번	전기 의 로동 한 시 사실 을 후대 전국 을 만나 .		
4번	영어	1번	전기 에서 보고 삼촌 친구 하고 가족 을 만날 선례 .		
5번	영어	1번	고향 에서 주말 에 전국 와 만나 선례 .		
6번	영어	1번	그다음 에 여기 시 한편 전기 하고 전국 을 만나 선례 .		
7번	영어	1번	그다음 에 전국 집 에 갈 선례 .		
8번	영어	1번	그다음 에 전국 을 만나 카피 하는 가 .		
9번	영어	1번	그다음 에 전국 하고 저녁 식사 하러 일 근로 직업 에 갈 선례 .		
10번	영어	1번	그러서 휴일 친구 하고 같이 발도 달리 에 반 분 준비 를 보 선례 .		
11번	영어	1번	그러서 친구 하고 전화 할 .		
12번	영어	1번	그런데 다들 화가 전국 을 만날 질 .		
13번	영어	1번	그런데 저는 전국 전국 을 발도 루시 에 갈 선례 .		
14번	영어	1번	그러고 부 전국 은 고향 에 돌아 갈 질 .		
15번	영어	1번	그러고 전국 을 같이 이 만 질 .		

<그림 3> 한국어 학습자 말뭉치 검색 화면 예시3

○ 이러한 원시말뭉치의 상세 검색을 활용하여, 유의어 관계에 있는 다양한 어휘의 다양한 변인에 따른 사용 양상의 차이를 살펴볼 수 있다. 예를 들어, 부사 ‘아주’와 ‘너무’의 숙달도, 연령, 문구어 변인에 따른 차이를 다음

과 같이 살펴볼 수 있다.

<표 42> 변인에 따른 ‘아주’와 ‘너무’의 사용 양상 비교

	변인	아주	너무
숙달도별	1급	970	491
	2급	656	830
연령별	20대	1,833	1,917
	30대	121	118
	40대	14	11
자료 유형	문어	2,404	2,484
	구어	75	138

③ 한국어 학습자의 언어 사용에 나타난 공기 관계 연구

- 원시말뭉치의 공기관계 검색을 활용하여 한국어 학습자들의 어휘 간 공기 양상 혹은 연어 양상의 파악이 가능하다. 예를 들어 한국어 학습자들의 많은 연어 오류 중 ‘친구를 사귀다’ 대신 ‘친구를 만들다’를 많이 사용하고 있는데, 이를 공기관계 검색을 활용하여 확인할 수 있다. 또한 상세검색을 활용하여 다양한 변인에 따른 차이도 살펴볼 수 있다. ‘친구+사귀(다)’의 공기관계 검색 실행 화면과 ‘친구+만들(다)’의 공기관계 검색 실행 화면은 다음과 같다.

<그림 4> 한국어 학습자 말뭉치 검색 화면 예시4

발문지 검색

- 통합 발문지 검색
- 원시 발문지 검색
- 형태 주석 발문지 검색
- 오타 주석 발문지 검색

원시 발문지 검색

검색 조건에 해당하는 문례를 검색합니다.

검색어

+

말씀

-

검색어 포함 조건

AND

검색 조건

검색

표본 검색 결과

검색 결과66건

번번	오목어	글수	문맥
1	중국어	6글	시간이 지나가면 동 사할의 가작권이 앞설지 안 앞는지 불명확 할 수 있는 한편 새로운 친구를 만들고 싶을지 때문에 직접 친구를 그릴게 중요하지 않는다는 것도 이해할 수 있다.
2	중국어	5글	그리고 대학교 때 환경에 대해 관심이 많은 친구를 같이 학교의 환경에 동아회를 만들었습니다.
3	베트남어	5글	친구를 만들고 말고 밖에 둘라 있다.
4	일본어	4글	그 별명은 유학에 오기 전에 친구가 나한테 만들어 준 것임니다.
5	일본어	4글	그리고 앞으로 구매행가 더 깊어지면 좋은 사람들에게도 보낼이 크니까 노익을도 친구나 이웃과 좋은 일 것있기를 믿으면서 서로 이마기하거나 응원하거나 도와주어야 한다.
6	일본어	4글	그리고 친구를 만들기 위해서는 SNS나 유튜브보다 직접 만나게 게 더 좋으려다.
7	일본어	4글	그리고 함께 친구와 서로 도와서 우정뿐만 좋은 관계를 만들고 싶습니다.

○ 이와 같은 공기관계 검색을 활용하여 학습자들의 사회 문화적 인식 등을 검토할 수 있다. 예를 들어 ‘친절하다’와 공기하는 어휘를 살펴봄으로써 학습자 주변에서 학습자들이 친절하다고 인식하는 대상을 확인할 수 있으며, 한국어와 ‘어렵다, 재미있다’와의 공기 관계를 살펴봄으로써 한국어에 대한 학습자들의 인식을 확인할 수 있다.

④ 형태소 범주별 사용 양상 연구

표본 검색 결과					
검색 결과 94222건			글수 내림차순	20개씩 보기	
연번	오류어	글수	원래 문맥	공정어	오른쪽 문맥
1	일본어	6글	그리고 주변 사람들의 따뜻한 한국	이것은	반드시 필요한 것이라고 생각한다.
2	중국어	6글	따라서	이것	교육 방식에 대한 논란이 뜨겁다.
3	중국어	6글	(와 문맥 없음)	나눔	수준별 이용수입을 운영한다면 학생...
4	중국어	6글	(와 문맥 없음)	자기의	부족한 점이 있고 더 열심히 공부할...
5	비유개	6글	이렇게 하면	자기의	경쟁 대상은 다른 사람이 아니라 자기...
6	비유개	6글	...이렇게 하면 자기의 경쟁 대상은 다른	자기이다	(무 문맥 없음)
7	비유개	6글	수준에 따라	자기	비슷한 수준의 친구를 하고 수업 들으...
8	비유개	6글	...무엇 학생들에게 이런 효과가 있지만	나보다	얼마나 잘하는지를 알 수가 없다.
9	비유개	6글	학생들이	자기	자신의 잘하고 할수 높은 과학에는 무...
10	비유개	6글	(와 문맥 없음)	이에	대학 교실 이용으로 인해 수업에 대한...

<그림 7> 한국어 학습자 말뭉치 검색 화면 예시7

- 이 또한 상세검색을 통해 숙달도, 연령, 국적, 언어권, 문/구어 등 다양한 변인에 따른 용례를 검색할 수 있다. 이 외에도 각 품사별로 학습자의 사용례를 한꺼번에 검색할 수 있어, 각 품사의 한국어 학습자 언어 사용 양상의 연구에 활용할 수 있다.

⑤ 변인에 따른 개별 형태소의 사용 양상 연구

- 형태 주석 말뭉치 검색을 활용하여 각 형태소의 한국어 학습자의 언어 사용 양상을 숙달도와 연령, 국적, 언어권 등 다양한 변인별로 확인할 수 있다. 예를 들어 1급에서의 관형사형 어미 ‘-은/는’의 학습자 언어 사용 양상을 살펴보고자 하면 형태소 표시의 [관형형전성어미]와 검색어 ‘-은’ 혹은 ‘-ㄴ’의 입력을 통해 검색할 수 있다. 검색 화면은 다음과 같다.

말뭉치 검색

- 통합 말뭉치 검색
- 일시 말뭉치 검색
- **형태 주석 말뭉치 검색**
- 오류 주석 말뭉치 검색

형태 주석 말뭉치 검색

검색 조건에 해당하는 용례를 검색합니다.

검색어

은

국적

국적을 선택하세요.

모국어

모국어를 선택하세요.

한국어 등급

☒ 1급
☐ 2급
☐ 3급
☐ 4급
☐ 5급
☐ 6급
☐ 6급 이상

나이

☐ 10대
☐ 20대
☐ 30대
☐ 40대
☐ 50대

학습 목적

☐ 진학
☐ 취업
☐ 거주
☐ 취미
☐ 결혼

말뭉치 유형

☐ 문어
☐ 구어

학습 환경

☐ 국내
☐ 국외

검색

<그림 8> 한국어 학습자 말뭉치 검색 화면 예시8

표본 검색 결과					
검색 결과 966건					
		중심어 가나다순		207쪽 보기	
연번	모국어	급수	원문 문맥	중심어	오른쪽 문맥
1	일본어	1급	(좌 문맥 없음)	같은	사람들과 같이 이야기를 해 보고 싶어...
2	중국어	1급	여름에 여행 치마와	가볍은	티셔츠를 입을 수 있습니다.
3	말레이어	1급	아 한국 실날에	가주들은	(무 문맥 없음)
4	중국어(광둥어)	1급	(좌 문맥 없음)	갈생전은	남자 가수 소년공화국의 민주 씨예요.
5	영어	1급	키고 더 크고 아 패패로	같은	물이 더 갈아요.
6	중국어(광둥어)	1급	구등학교하고 대학교도	같은	반이었었습니다.
7	스웨덴어	1급	또 지금 나	같은	언어를 공부하고 있어요.
8	일본어	1급	이제는 친구가 아니라 가족	같은	존재입니다.
9	일본어	1급	절은 연니하고 저는	같은	7반 학생입니다.
10	일본어	1급	저와	같은	나이이지만. 아들이 2명 있습니다..
11	중국어(광둥어)	1급	우리의 많이	같은	취미 있습니다.
12	일본어	1급	그때 구리로 인형과	같은	포즈에서 사진을 찍으면 좋습니다.
13	일본어	1급	나도 매주 금요일에 호텔에서	같은	방 친구와 같이 삼겹살과 소주를 먹습...

<그림 9> 한국어 학습자 말뭉치 검색 화면 예시9

- 이와 같은 형태소 분석 말뭉치 검색을 활용하여 원시 말뭉치 검색에서는 다소 어려웠던 문법 형태소인 조사나 어미 등을 검색할 수 있어, 학습자들의 문법 형태소 사용 양상의 분석에 활용할 수 있다.

⑥ 오류 유형별 용례 검색

- 오류 주석 말뭉치 검색을 활용하여, [오류 위치], [오류 양상], [오류 층위]의 오류 유형에 따른 오류 용례를 검색할 수 있다. 예를 들어 한국어 학습자의 존대법과 관련한 오류 용례만을 검색하고자 하면, [오류 층위]의 [높임]만을 검색할 수 있음. 검색 화면은 아래와 같다.

말뭉치 검색

·용인 말뭉치 검색

·원시 말뭉치 검색

·형태 주석 말뭉치 검색

·오류 주석 말뭉치 검색

오류 주석 말뭉치 검색

검색 조건에 해당하는 용례를 검색합니다.

복합 주석 검색

자소 검색

문법 검색

검색 조건

원 형태

오류 위치

오류 양상

오류 층위

원 형태

오류 위치

오류 양상

오류 층위

원 형태

오류 위치

오류 양상

오류 층위

상세 조건

검색

<그림 10> 한국어 학습자 말뭉치 검색 화면 예시10

원색 결과407건

급수 내림자순

20개씩 보기

연번	모국어	급수	원색 문맥	중심어	교정 형태	오류측 문맥	오류 주석
1	일본어	6급	...씨는 예는 프로로 출	소개하셨다	ADD/EP	(주 문맥 없음)	FPE ADD SH
2	영어	6급	(좌 문맥 없음)	저는	나/NP	TV에서 영화를 방영...	CNP REP SH
3	일본어	6급	(좌 문맥 없음)	저는	나/NP	그것에 대해 반대하...	CNP REP SH
4	비공개	6급	(좌 문맥 없음)	저는	나/NP	유학생으로 온 외국...	CNP REP SH
5	러시아어	6급	(좌 문맥 없음)	저는	나/NP	인생관에 대해서 말...	CNP REP SH
6	러시아어	6급	(좌 문맥 없음)	저는	나/NP	자신 아이들에게 귀...	CNP REP SH
7	중국어	6급	(좌 문맥 없음)	저는	나/NP	최저임금제 설정에...	CNP REP SH
8	중국어	6급	(좌 문맥 없음)	제가	내/NP	생각하기에는 아무...	CNP REP SH
9	일본어	6급	...그 중등 중등적인	같다	아요/EF	(주 문맥 없음)	FPE REP SH
10	일본어	6급	그래서	저는	나/NP	나중에 엄마에게 행...	CNP REP

<그림 11> 한국어 학습자 말뭉치 검색 화면 예시11

○ 위의 용례를 [내려받기]하여 엑셀에서 다양한 방식으로 소트하여 연구자의 필요에 따른 통계 처리가 가능하다. 이를 활용하여 오류 위치, 즉 고유명사, 일반명사, 의존명사, 대명사, 수사, 동사, 형용사, 보조용언, 지정사, 관형사, 일반부사, 접속부사, 감탄사 등 각 품사별 오류를 검색할 수 있으며, 명사과생접미사, 동사과생접미사, 형용사과생접미사, 어근 등 형태소의 오류를 별도로 검색할 수 있다. 또한 조사와 어미도 주격조사, 관형격조사, 목적격조사, 부사격조사, 호격조사, 인용격조사, 보조사, 연결어미, 종결어미, 선어말어미, 명사형 전성어미, 관형사형 전성어미, 접속조사, 보격조사에 따라서 각각 검색이 가능하여 오류 용례를 각 품사나 형태소 문법 범주별로 분리하여 검색할 수 있다. 이 외에 구 단위표현과 표현문형 등 단어 이상의 범주에 대한 검색이 가능하여 학습자의 오류 용례를 총체적으로 파악할 수 있다. 물론 이 또한 숙달도, 언어권 등 상세조건 검색을 통해 오류 용례를 별도로 검색하고 내려받을 수 있어, 학습자의 숙달도나 언어권에 따른 오류 양상을 살펴볼 수 있다.

⑦ 특정 항목의 오류 용례 분석

○ 오류 주석 말뭉치 검색을 활용하여 연구자가 연구하고자 하는 특정 언어 형태의 오류 용례를 검색할 수 있다. 예를 들어, 연결어미 ‘-어서’와 ‘-니까’를 서로 혼용하여 사용하는 학습자 언어에서의 오류 양상을 살펴보고자 할 때 [원형태]와 [교정형태]에 언어 형태를 입력함으로써 필요한 용례만을 검색할 수 있다. 검색 실행 화면은 다음과 같다.

오류 주석 말뭉치 검색 검색 조건에 해당하는 용례를 검색합니다.

말뭉치 검색

- 통합 말뭉치 검색
- 원시 말뭉치 검색
- 형태 주석 말뭉치 검색
- 오류 주석 말뭉치 검색

복합 주석 검색 자소 검색 문형 검색

검색 조건

원 형태 어서 AND 교정 형태 니까

오류 위치 오류 양상 오류 종류

오류 종류

- ☐ 발음
 - ☐ PP(종소)
 - ☐ PS(종절)
 - ☐ PC(종운규칙)
 - ☐ 종어식 발음
 - ☐ 중간 발음(변이음 포함)
- ☐ 형태
 - ☐ MCP(단어 형성(합성법))
 - ☐ MDV(단어 형성(파생법))
 - ☐ MDC(굴절(어종))
 - ☐ MC(굴절(형용))
 - ☐ POS(품사)

상세 조건

<그림 12> 한국어 학습자 말뭉치 검색 화면 예시12

표본 검색 결과

검색 결과4건 급수 내림자순 20개씩 보기

연번	모국어	급수	왼쪽 문맥	중심어	교정 형태	오른쪽 문맥	오류 주석
1	중국어	4급	...장애인이냐 비장애	멤버이어서	니까/EC	사람마다 생각이 다...	FED REP
2	중국어	4급	...장애인은 생활이 너	힘들어서	니까/EC	표값이 싸야 되다.	FED REP
3	중국어	3급	...대학교 입학한 후에	피아노여서	니까/EC	음악에 관심이 있는...	FED REP
4	중국어	3급	이사할 때 정말	힘들어서	니까/EC	이사가 끝나자마자...	FED REP

<< < 1 > >>

[내려받기](#)

<그림 13> 한국어 학습자 말뭉치 검색 화면 예시13

오류 주석 말뭉치 검색 검색 조건에 해당하는 용례를 검색합니다.

말뭉치 검색

- 통합 말뭉치 검색
- 원시 말뭉치 검색
- 형태 주석 말뭉치 검색
- 오류 주석 말뭉치 검색

복합 주석 검색 자소 검색 문형 검색

검색 조건

원 형태 니까 AND 교정 형태 어서

오류 위치 오류 양상 오류 종류

오류 위치

- ☐ FAP(부사격조사)
- ☐ PVP(목적조사)
- ☐ PQP(인용격조사)
- ☐ PVP(보조사)
- ☒ FFE(단결어미)
- ☐ FFE(종결어미)
- ☐ FFE(선어말어미)
- ☐ FNE(명사형 전성어미)
- ☐ FAE(관형사형 전성어미)
- ☐ FJC(형속조사)
- ☐ FCP(보격조사)
- ☐ PHE(구단위표현)

오류 양상

- ☐ OM(누락)
- ☐ ADD(첨가)
- ☐ REP(대치)
- ☐ MIF(오형태)

오류 종류

- ☐ 발음
 - ☐ PP(종소)
 - ☐ PS(종절)
 - ☐ PC(종운규칙)
 - ☐ 종어식 발음
 - ☐ 중간 발음(변이음 포함)
- ☐ 형태
 - ☐ MCP(단어 형성(합성법))
 - ☐ MDV(단어 형성(파생법))
 - ☐ MDC(굴절(어종))
 - ☐ MC(굴절(형용))

<그림 14> 한국어 학습자 말뭉치 검색 화면 예시14

표본 검색 결과							
검색 결과 12건				급수 내림차순	20개씩 보기		
연번	모국어	급수	원문 문맥	증설어	교정 형태	오른쪽 문맥	오류 주석
1	일본어	4급	...구분하지 않는 나라	일이니까	어서/EC	하나도 안 힘들다.	FED REP
2	아랍어	4급	...영화 처음에는 재미	보니까	어서/EC	계속 스마트폰을 보...	FED REP
3	일본어	2급	은 성격이 활발한	변이니까	어서/EC	재 친구하고 바로...	FED REP
4	베트남어	2급	바다에	근처이니까	어서/EC	정말 좋았다.	FED REP
5	중국어	2급	정말	예쁘니까	어서/EC	시간 있으면 꼭 갑...	FED REP
6	터키어	2급	전혀 재미있는	사람이니까	어서/EC	나의 마음에도 똑같은...	FED REP
7	베트남어	2급	...한국 물건값이 베트남	비싸니까	어서/EC	안 먹고 안 쓸 실용...	FED REP
8	중국어	1급	단종이	예쁘니까	어서/EC	등산하는 사람이 많...	FED REP

<그림 15> 한국어 학습자 말뭉치 검색 화면 예시15

- 위에서와 같이 ‘-니까’의 ‘-어서’ 대치 오류는 4건이 검색됨에 반해, ‘-니까’의 ‘-어서’ 대치 오류는 12건이 검색되어 학습자들이 ‘니까’를 ‘-어서’로 더 많이 대치해서 사용하고 있음을 알 수 있다. 따라서 오류 분석 말뭉치 검색을 활용하여 각 언어 형태별로 오류 양상을 다양한 변인별로 살펴볼 수 있다.

⑧ 자소 검색을 통한 철자, 발음 오류 양상 분석

- 오류 주석 말뭉치의 자소 검색을 활용하여 학습자들의 철자 혹은 발음(음운) 상의 오류 양상을 편리하게 살펴볼 수 있다. 예를 들어 중국어권 학습자들의 어말 종성 폐쇄음 받음 오류와 관련하여, 중국어권 학습자들이 어말 종성 폐쇄음인 /ㄱ, ㄷ, ㄴ/를 실제로 어떻게 철자화하고 혹은 발음하는지 편리하게 살펴볼 수 있다. 다음은 중국어권 학습자의 어말 종성 ‘ㄱ’의 오류를 살펴보기 위한 실행 화면이다.

말뭉치 검색

- 통합 말뭉치 검색
- 원시 말뭉치 검색
- 형태 주석 말뭉치 검색
- 오류 주석 말뭉치 검색

오류 주석 말뭉치 검색

검색 조건에 해당하는 결과를 검색합니다.

복합 주석 검색

자소 검색

문법 검색

검색 유형

교정 형태

어두 조성

ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅈ ㅊ ㅋ ㆁ ㅅ ㅈ ㅊ ㆁ ㅅ ㅈ ㅊ ㆁ

어말 종성

ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅈ ㅊ ㅋ ㆁ ㅅ ㅈ ㅊ ㆁ ㅅ ㅈ ㅊ ㆁ

국적

국적을 선택하세요.

모국어

중국어

중국어(광둥어)

중국어(만다린어)

중국어(민난어)

중국어(우어)

중국어(타이완어)

<그림 16> 한국어 학습자 말뭉치 검색 화면 예시16

표본 검색 결과							
검색 결과576건				급수 내림차순	20개씩 보기		
연번	모국어	급수	왼쪽 문맥	중심어	교정 형태	오른쪽 문맥	오류 주석
1	중국어	6급	2008년에	베이징어협에서	올림픽/NNG	베드민턴 단식은 2...	CNNG MF
2	중국어	6급	...2218년부터 한국 ·	자택에서	채택/NNG	2220년에 부산과...	CNNG MF
3	중국어	6급	...2218년부터 한국 ·	시작했다.	시작/NNG	{후 문맥 없음}	CNNG MF
4	중국어	6급	...TV는 공공성을 가진	판단현이	판단력/NNG	호된 아동과 청소년...	CNNG MF
5	중국어	6급	{좌 문맥 없음}	가군시	씩/XSN	그 방송에서 같은...	CXSN MF
6	중국어	6급	{좌 문맥 없음}	가정	가족/NNG	관념 그 변하기 때...	CNNG REP
7	중국어	6급	{좌 문맥 없음}	본영에	번역/NNG	대한 책을 그런 거...	CNNG PP
8	중국어	6급	{좌 문맥 없음}	생각하고	생각/NNG	아마 이런 성취력...	CNNG PP
9	중국어	6급	{좌 문맥 없음}	아직	아직/MAG	심막하지 알아지만...	CNAG PP

<그림 17> 한국어 학습자 말뭉치 검색 화면 예시17

- 위에서 보는 바와 같이 중국어권 학습자의 오류 중 어말 종성이 ‘ㄱ’인 모든 용례를 추출함으로써, 내려받은 후 엑셀에서 철자나 발음과 관련한 오류만 추출이 가능하다. 그리고 이를 활용하여 중국어권 학습자의 어말 종성 폐쇄음 발음 및 철자 오류 양상을 파악할 수 있다.

(2) 학습자 말뭉치를 활용한 연구 사례

- ① 중국인 한국어 학습자 말뭉치에 나타난 중간언어 분석 연구
 - 이 연구는 형태 주석 말뭉치 분석하여 중국인 학습자의 숙달도별 품사별 언어 사용 실태 및 오류 분석의 결과를 객관화·일반화하고자 하였다. 특히, 구어 말뭉치를 구축, 형태 주석하여 구어에서 드러나는 학습자 언어적 특성을 규명함으로써 한국어 교육에서의 학습자 언어 연구에 있어서 자료의 다각화, 사용역의 다양화를 이루었다는 점에서 학습자 말뭉치를 활용한 연구로서의 효용성을 살펴볼 수 있다.
- ② 모어 화자와 한국어 학습자 말뭉치에서의 ‘은/는’과 ‘이/가’의 분포와 조사 선택 요인 분석
 - 이 연구는 한국어 학습자의 ‘이/가’와 ‘은/는’의 분포 양상을 살펴보고 모어 화자의 사용 양상과 비교함으로써 한국어 학습자의 조사 선택 요인을 객관적으로 규명하고자 하였다. 1,000만 어절 규모의 모어 화자 말뭉치 및 77만 어절 규모의 한국어 학습자 말뭉치를 이용하여 실제 언어 사용 양상을 객관적으로 분석함으로써 학습자 말뭉치를 활용한 학습자 언어 연구(Learner Corpus Research)의 구체적인 실례로서 의의가 있다고 할 수 있다.

- ③ 학문 목적 학습자의 객관화 전략 사용 양상 연구 - 중국인 학습자의 학술 텍스트를 중심으로
- 이 연구는 객관화 전략 구현의 언어 표지로 지속 표현, 비인칭 표현, 피동 표현, 명사화 구문이 학술 텍스트에서 사용되고 있는 양상을 고찰함으로써 중국인 최고급 학습자들의 지속 표현 선택의 오류, 연구 주체의 부적절한 노출, 과도한 복수 표지 사용 등을 확인하고자 하였다. 최고급 학습자가 산출한 보고서 텍스트 구조에 대한 마크업과 상적 표현, 피동, 복수 등의 문법 범주를 주석한 결과를 활용한 연구로 학문 목적 학습자 연구에 대한 마크업 체계와 태깅 유형을 확립하는 데 기여하였다는 점에서 의의가 있다.
- ④ 중국어권 한국어 학습자의 조사 오류 및 사용 양상 연구
- 이 연구는 학습자 말뭉치를 기반으로 한 중간언어 연구로 중·고급의 중국어권 한국어 학습자의 조사 ‘이/가’와 ‘은/는’ 오류 및 사용 양상을 분석하였다. 실제 학습자 자료의 분석을 통해서 ‘이/가’와 ‘은/는’의 사용 및 미사용 오류를 살피고, 두 조사 선택에 영향을 미치는 잠재적 요인들에 대한 검증을 실시함으로써 학습자의 중간언어 규칙을 탐구하였다는 점에서 말뭉치 활용 연구로서의 의의를 찾아볼 수 있다.
- ⑤ 중국어권 학습자의 접사 사용 연구
- 이 연구는 중국어권 학습자의 접사 사용에서 나타나는 중간언어 체계를 고찰하고자 한 연구이다. 다양한 숙달도의 학습자 자료로 이루어진 학습자 말뭉치를 활용하여 한국어 숙달도에 따른 접사 사용 양상을 살피고, 사용 과정에서 발생하는 접사 오류를 분석하여 오류의 원인을 규명하고자 하였다는 점에서 말뭉치 활용 연구로서의 의의를 찾아볼 수 있다.

2.3. 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구

(1) 균형성의 개념과 쟁점

○ 신뢰도가 높은 말뭉치를 구축하기 위해 말뭉치 설계에서 가장 중요하게 고려되는 것은 말뭉치의 대표성과 균형성이다. 이 중 균형성이란 ‘일반 말뭉치의 경우 텍스트 영역들 간의 균형과 가중치의 문제’로 정의된다. 이때의 균형성이란 단순히 각 영역의 텍스트 비율을 맞추는 것을 의미하지 않으며 실제 언어 사용에서의 비중이나 그 텍스트가 사람들에게 미칠 수 있는 영향력이 고려되어야 한다(서상규·한영균, 1999: 34-35). 이와 같은 맥락에서 학습자 말뭉치의 설계와 구축 시 균형성을 확보하기 위해 고려해야 하는 것은 학습자와 자료에 관련된 다양한 변인이 된다. 일반 말뭉치와 비교할 때 학습자 말뭉치는 이러한 변인들이 매우 다양하기 때문에 이들을 어디까지 수용하느냐가 균형성 확보를 위한 우선적인 과제가 된다. 다음은 한국어 학습자 말뭉치의 균형성의 측면에서 고려해야 할 변인 정보가 될 수 있는 정보의 유형들이다.

- 수집 대상: 국내 교육 기관 학습자, 이주민(결혼이주민, 이주노동자, 중도입국청소년, 국외 교육 기관 학습자)
 - 학습자 변인: 이름, 성별, 연령(출생년), 현재 등급, 국적, 제1 언어, 한국어 학습 기간, 한국에서의 거주 기간, 한국어 학습 목적, 직업. 한국어 외의 사용 가능 언어
 - 자료 변인: 발화 유형(문어/구어), 수집 시기(횡적/종적), 수집 과제(교육과정/기획), 장르
- 이상적인 학습자 말뭉치는 위의 변인별로 일정 규모 이상의 말뭉치가 고르게 분포하여 다양한 목적으로 말뭉치를 사용하려는 많은 사람들의 요구를 만족시킬 수 있는 말뭉치일지도 모른다. 그러기 위해서는 각 변인 간의 관계가 고려된 다양한 말뭉치가 일정 규모 이상 구축되어야 하는데, 그리 크지 않은 규모의 말뭉치에서는 전체 말뭉치를 이루는 하위 말뭉치의 규모가 적어지게 되는 결과를 낳는다. 이를 보완하려면 결과적으로 대규모의

말뭉치가 구축되어야 하는데 학습자가 산출한 자료를 일일이 수집하는 것부터 시작되는 학습자 말뭉치의 수집 여건상 많은 시간과 비용, 노동력을 필요로 하기 때문에 현실적으로 쉽지 않다. 결국 다양성과 균형성의 문제가 서로 배치되는 것이다.

- 결국 모든 변인을 수용할 만큼 대규모의 말뭉치를 구축하는 것이 아니라면 학습자 말뭉치의 설계와 구축에서 다양한 변인 어떠한 변인을 우선적인 수집 대상으로 삼을지를 결정해야 하며, 이에 대한 타당성 있는 근거가 필요하다. 이로부터 학습자 말뭉치의 균형성이란 모든 변인에 따른 표본을 산술적으로 균형적으로 갖춘 말뭉치가 아니라 자료의 다양성을 훼손하지 않은 범위 내에서 일정한 근거에 따라 변인의 종류와 비율을 조정하는 것을 말한다. 본 연구에서는 한국어 학습자의 변인과 자료의 변인과 관련된 특성을 살펴봄으로써 그 근거가 무엇이 될 수 있는지 검토해 보고 그것을 바탕으로 한국어 학습자 말뭉치의 균형성 확보를 위한 방법론을 마련해 보고자 한다.

(2) 학습자 말뭉치의 균형성 확보를 위한 주요 변인 검토

① 학습 대상 및 목표

- 본 연구에서는 국내 학습자, 이주민, 국외 학습자의 자료를 수집 대상으로 한다. 국내외 학습자의 분포에 관한 통계 분석 자료에 따르면 세 집단은 국내 유학생 351,823명, 국내 거주 이주민 456,187명, 국외 학습자 237,482+a(대학)로 약 100만 명에 이르며, 그 중 국내 거주 이주민의 수가 가장 많다³⁾. 국내 거주 이주민의 분포를 다시 살펴보면 외국인 근로자의 수가 약 25만 명으로 가장 많은데, 이들의 경우 적극적으로 한국어 학습에 참여하지 않는 경우가 많다는 점을 고려할 때 한국어 학습자로서의 국내 거주 이주민의 수는 사실상 대폭 줄어들 가능성이 높다. 또한 약 8만 명에 달하는 다문화가정 자녀의 경우 한국어 학습을 필요로 하는 중도입국청소년의 비율이 한국 출생의 외국인 부모 가정의 자녀 비율에 비해 매우 낮음을 고려한다면 국내 거주 이주민의 수는 더욱 줄어들 것으로 예상된다. 이로부터 사실상 한국어 학습자로서의 외국인 분포는 국내 학습자가 가장 많다는 결론을 이끌어낼 수 있다. 이에 따라 자료의 활용

3) 2014년 세종학당 관련 주요 통계, 2014년 국내 외국인 유학생 통계, 2014 재외동포 교육기관 현황, 2013년 법무부 체류 외국인 통계 참고.

측면에서 국내 학습자의 비중을 높이고 나머지 비중에서 이주민과 국외 학습자 자료의 비율을 조정하는 것이 타당하다고 볼 수 있다. 이는 자료 수집의 용이성 측면에서도 합리적인 면이 있다.

<표 43> 한국어 학습자가 될 수 있는 외국인의 분포

구분	수집 대상	학습자 수
국내 학습자	어학 연수생	18,543명
	대학, 대학원 진학 유학생	60,639명
	기관 외 외국인	272,641명
	소계	351,823명
국내 거주 이주민	결혼 이민자	124,679명
	외국인 근로자	248,794명
	다문화가정 자녀	82,714명
	소계	456,187명
국외 학습자	세종학당	37,177명
	한국학교	12,322명
	한글학교	105,097명
	초중등학교 한국어반	82,886명
	소계	237,482+α(대학)
합계		1,045,492명

② 학습자의 국적 및 제1 언어에 따른 자료의 균형성

- 통계 분석 자료에 따르면 한국어 학습자가 될 가능성이 있는 외국인의 국적 분포를 살펴보면 국내 학습자, 국내 이주민, 국외 학습자의 세 집단 모두 중국이 가장 높은 것을 알 수 있다. 다음 표는 집단별 국적 분포를 순위별로 나타낸 것이다. 집단별로 세부적인 국적에서는 차이가 있지만 대체로 중국, 베트남, 대만, 일본 등의 아시아권 국적을 소지한 외국인의 비중이 높음을 볼 수 있다.

<표 44> 한국어 학습자가 될 수 있는 외국인의 국적 분포

순 위	국내 학습자		국내 이주민			국외 학습자	
	유학생	기관 미소속	결혼이민자	외국인 근로자	외국계 주민 자녀	재외 동포	국내 거주 동포
1	중국 (66.37)	중국 (64.75)	중국 (51.91)	중국 (46.19)	중국 (40.51)	북미 (32.05)	중국 (67.5)
2	기타 (6.61)	대만 (7.70)	베트남 (23.18)	베트남 (10.15)	베트남 (26.81)	기타 (12.09)	북미 (19.16)
3	몽골 (5.13)	베트남 (6.13)	필리핀 (6.64)	남미 (10.10)	필리핀 (9.58)	태국 (10.22)	기타 (9.32)
4	남미 (4.94)	기타 (5.73)	일본 (5.19)	아시아 기타 (6.85)	일본 (8.91)	일본 (1.10)	중앙아시아 (2.00)
5	베트남 (4.55)	남미 (4.06)	아시아 기타 (2.86)	인도네시아 (5.77)	아시아 기타 (3.64)	중앙아시아 (8.95)	러시아 (0.34)
6	일본 (2.69)	북미 (2.76)	기타 (2.02)	중앙아시아 (5.26)	기타 (1.91)	중국 (7.70)	일본 (0.34)
7	아시아 기타 (2.67)	일본 (2.72)	남미 (1.51)	필리핀 (4.53)	중앙아시아 (1.75)	남미 (4.62)	인도네시아 (0.04)
8	중앙아시아 (1.77)	필리핀 (1.58)	중앙아시아 (1.41)	태국 (4.19)	몽골 (1.45)	러시아 (4.18)	아시아 기타 (0.02)
9	북미 (1.22)	몽골 (1.22)	몽골 (1.26)	북미 (2.44)	태국 (1.36)	인도네시아 (2.23)	대만 (0.02)
10	인도네시아 (1.19)	아시아 기타 (1.09)	태국 (1.25)	기타 (2.32)	남미 (1.25)	아시아 기타 (2.08)	남미 (0.01)
11	대만 (0.90)	러시아 (0.91)	북미 (1.19)	몽골 (1.63)	대만 (0.93)	베트남 (1.95)	필리핀 (0.01)
12	러시아 (0.73)	중앙아시아 (0.90)	대만 (0.66)	일본 (0.32)	북미 (0.91)	대만 (1.64)	태국 (0.00)
13	필리핀 (0.70)	태국 (0.23)	러시아 (0.65)	러시아 (0.20)	러시아 (0.65)	필리핀 (1.35)	몽골 (0.00)
14	태국 (0.53)	인도네시아 (0.23)	인도네시아 (0.28)	대만 (0.04)	인도네시아 (0.37)	몽골 (0.85)	베트남 (0.00)

- 이러한 현상은 국내 언어교육원의 학습자 분포에서도 확인되었는데, 언어 권별로는 중국(40.5%), 일본(16.8%), 기타 아시아 국가(19.2%)로 아시아 국가의 학습자가 주를 이루며, 아메리카(8.6%), 유럽(11.8%), 오세아니아(1.6%), 중동(0.7%), 아프리카(0.8%) 국가의 학습자는 매우 적음을 알 수 있다.
- 국내에 거주하는 외국인의 국적이 최소 150개 이상임을 고려할 때 국가별 자료를 산술적으로 균등한 규모로 구축하는 것은 현실적으로 불가능하며 실제 사용의 측면에서도 그 효용성이 크지는 않을 것으로 본다. 이에 따라 본 연구에서는 한국어 학습자가 될 가능성이 있는 외국인의 국적 분포에 따라 비중이 높은 국적의 학습자 자료 구축 비중을 높이는 것이 바람직하다고 판단하였으며 현재까지의 수집 결과 또한 이를 뒷받침해 주고 있다.

③ 학습자의 언어 숙달도에 따른 자료의 균형성

- 숙달도 단계는 교육 기관마다 다르다. 외국인 유학생을 대상으로 한 국내 교육 기관에서는 보편적으로 1-6등급 체계를 선택하고 있고, 이주민 교육 기관에서는 다문화가족지원센터의 경우 1-4단계, 사회통합프로그램 운영 기관에서는 1-5단계의 체계에 따르고 있다. 또한 많지는 않지만 기관을 불문하고 초급, 중급, 고급의 3등급 체계를 채택하고 있는 기관도 있다. 한 국내 교육 기관 학습자의 숙달도 분포에 따르면 개원 이래 누적된 학생 분포가 1급(23.2%), 2급(24.1%), 3급(19.0%), 4급(15.6%), 5급(10.7%), 6급(5.9%), 최고급 이상(1.4%)으로 초급에 해당하는 학습자가 50%에 달하는 반면, 고급 학습자는 20% 이내로 매우 낮아지는 것을 볼 수 있다. 이러한 현상은 학습 대상이나 등급 체계와는 상관없이 공통적으로 나타나는 현상이며, 상대적으로 교육의 기회가 적고 교육과정 시수가 짧은 이주민과 국외 학습자의 경우 더욱 심한 경향이 있다⁴⁾. 현실적인 학습자의 분포는 이와 같지만 숙달도에 따른 중간언어 특성의 차이 비교, 언어 발달 양상 비교 등 자료 활용을 고려할 때 숙달도 단계에 따른 자료에 대한 요구가 다르지는 않을 것으로 판단된다. 이에 따라 본 연구에서는 언어 숙달도에 따른 자료의 양적 분포를 고르게 하여 균형성을 확보하고 있다.

4) 연세대학교 언어교육원 홈페이지 참고. 2014년 기준

④ 자료의 변인에 따른 균형성

- 현재까지의 한국어 학습자 말뭉치 구축 결과를 귀납적으로 살펴보면 발화 유형, 수집 기간, 수집 방식, 장르별로 다음과 같은 자료들이 수집·구축되었다. 본 연구에서는 대학 부설 언어교육원, 다문화가족지원센터, 사회통합프로그램운영기관 등의 교육 기관 말뭉치를 중점적으로 수집하여 구축하였기 때문에 교육과정에 종속된 자료가 주를 이루고 있다.

<표 45> 자료 변인

구성	범주 구성
발화 유형	문어, 구어
수집 기간	종적 자료, 횡적 자료
수집 방식	교육과정 말뭉치(시험 작문, 과제 작문) 기획 말뭉치(주제 작문, 백일장)
장르	문어: 생활문, 편지글, 논설문, 설명문, 기행문, 안내문, 광고문, 기사문, 인터뷰, 발표, 자유 대화, 감상문, 투고문, 영화평, 자기 소개서, 수필, 시, 일기, 평전, 보고서, 자전적 에세이, 건의문, 답사기 구어: 인터뷰, 발표, 내러티브

(3) 한국어 학습자 말뭉치의 균형성 확보 방안

① 실제적인 학습자 분포를 고려한 균형성 확보

- 한국어 학습자 말뭉치의 균형성은 학습자의 국적, 수준 등의 학습자 분포를 참고하여 수집 비율을 조정되는 것이 현실적이다. 따라서 귀납적인 자료 수집 결과에 따라 실제적인 비중이 높은 국적의 자료의 구축 비율을 높이고 그 외의 국적을 소지한 학습자의 자료를 고르게 구축하여 자료의 다양성 또한 확보하도록 한다. 한편, 숙달도별 구축 비율은 숙달도에 따른 중간언어 특성의 차이 비교, 언어 발달 양상 비교 등 자료 활용을 고려하여 자료의 양적 분포를 고르게 조정하도록 한다.

② 다양한 수집 과제를 활용한 균형성 확보

- 종적 자료, 주제 작문이나 백일장 자료 등의 기획 말뭉치, 내러티브와 같

은 다양한 장르의 말뭉치 비율을 높임으로써 자료의 다양성을 추구한다.

③ 선 구축, 후 균형

- 거시적인 구축 계획에 따라 자료를 수집한 후 수집 결과를 귀납적으로 정리하여 균형성 확보를 위해 추가가 필요한 말뭉치를 기획하여 보완하도록 한다. 예를 들어, 자료 수집과 구축의 용이성, 비용 등을 고려하여 문어 대비 구어의 비중을 낮추었으나 자료의 양이 적어 특정 숙달도 단계, 특정 국적의 표본 수가 매우 적어 연구 결과의 일반화가 어려운 점, 구어 자료에 대한 사용자의 요구가 높은 점 등을 고려하여 구어 자료의 비중을 높일 수 있다. 또한 자료의 다양성 측면에서 백일장 자료나 상대적으로 비율이 낮은 특정 장르의 자료를 추가할 수 있다.

④ 하위 말뭉치(sub corpus)의 구축을 통한 다양성의 확보

- 학습자 말뭉치의 설계와 구축에서 고려해야 하는 변인이 매우 다양함을 앞에서 살펴보았다. 개별 범주의 변인 수도 많지만 ‘중국인 학습자의 1급 자료’와 같이 둘 이상의 변인이 결합하면 실제적인 변인의 수는 통제하기 어려운 수준으로 증가한다. 이에 따라 대규모의 말뭉치가 아닐 경우 지나치게 다양한 말뭉치를 구축 대상에 포함시키게 되면 하위 범주별 말뭉치의 수가 줄어들어 실제적인 활용도가 매우 낮아지는 결과를 초래한다. 이러한 점을 고려하여 한국어 학습자 말뭉치는 학습자의 분포가 가장 높은 국내 기관 학습자의 자료를 국적, 숙달도 단계별로 구축하되 이주민 자료나 국외 기관의 자료는 별도의 하위 말뭉치(sub corpus)로 구축하여 자료의 균형성과 다양성을 모두 확보하는 방안을 고려해 볼 수 있다.

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 하여 전문성을 제고하여 체계적인 말뭉치 구축을 위한 것이다. 또한 다양한 변이형을 포함한 학습자 말뭉치의 특성상 작업자의 직관에 의해 달라질 수밖에 없는 세부적인 자료 처리 방식을 최대한 일관성 있게 맞추기 위한 것으로 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 46> 말뭉치 구축/가공 인력 교육 내용

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

1.4 참여 인력

(1) 자료 수집

① 국내 교육기관

김지은(서강대학교 언어교육원), 김지혜(동아대학교 언어교육원),
김태범(배재대학교 한국어교육원), 노미숙(부산외국어대학교 한국어학당),
노미연(국민대학교 국제교육원), 박진철(이화여자대학교 언어교육원),
안민지(경북대학교 한국어문화원), 유순천(선문대학교 한국어교육원),
이수미(연세대학교 언어교육원), 이지민(경희대학교 국제캠퍼스 언어교육원),
이지영(경희대학교 국제교육원), 이혜지(서울대학교 언어교육원),
최유진(고려대학교 국제어학원), 홍종호(계명대학교 한국어학당)

② 이주민 자료 수집

강진숙(성주군 다문화가족지원센터), 권소라(고려대학교 교육대학원),
권혜경(경남이주민문화센터 라함), 김성경(대구 서구 다문화가족지원센터),
김은숙(전남대 언어교육원/광주외국인력지원센터),
박명숙(충북대학교 국제교류본부), 박수연(조선대학교 언어교육원),
백선지(경남이주민문화센터 라함), 유영주(송담대학교 한국어학원),
윤미희(연세대학교 교육대학원), 이덕향(호수돈여자고등학교),
이미란(고양시 다문화가족지원센터), 이재경(구로구 다문화가족지원센터),
이현주(충남대학교 국제언어교육원/부여다문화센터),
인민지(서울대동초등학교), 임수정(구미대학교 국제교류처),
임정혜(울산 북구 다문화가족지원센터), 장준영(평화 다문화센터),
정지윤(송파 다문화가족지원센터), 정훈(김포외국인주민센터/인천다문화센터),
주월랑(동두천시 다문화가족지원센터), 지숙이(경남이주민문화센터 라함),
최현정(경기글로벌센터)

(2) 자료 스캔 및 처리

문형빈, 이지언(연세대학교 국어국문학과)

(3) 문어 입력

김동은(연세대학교 국어국문학과 박사 과정),
박현진, 채여진(이상 연세대학교 국어국문학과 석사 과정),
정빛나(연세대학교 교육대학원 석사 과정),
김태환, 여은정, 이종현, 최미희, 김나연, 노장호, 박현진(이상 연세대학교 국어국문학과)

(4) 구어 전사

김정현(연세대학교 국어국문학과 박사 과정),
김미선, 김예슬, 김윤하, 임효진, 진승현
(이상 연세대학교 국어국문학과 석사 과정)

(5) 형태 주석

김미경, 박상경, 배미연, 이창봉, 임상미, 임상은
(이상 연세대학교 국어국문학과 박사 과정),
윤이운, 이보미, 이주희(이상 연세대학교 국어국문학과 석사 과정)

(6) 형태 주석/오류 주석

공나형, 김선영, 송지혜, 유소영, 허희정(연세대학교 국어국문학과 박사 과정)

2. 한국어 학습자 말뭉치 아카데미 개최

- 한국어 학습자 말뭉치 아카데미는 2016년에 개최한 두 차례에 아카데미에 이어 올해에도 두 차례의 아카데미를 개최하게 된다. 2017년 한국어 학습자 말뭉치 아카데미는 구어 말뭉치 구축을 위한 전사 도구 사용 방법과 한국어 학습자 말뭉치 나눔터를 소개하여 사용자들이 말뭉치 구축의 방

법과 절차를 배우고, 효율적으로 활용할 수 있도록 하는 데에 목표를 두고 실습형 아카데미를 실시하였다. 세 차례 모두 많은 참가자의 관심과 호응이 있었으며 이를 통해 한국어 학습자 말뭉치에 대한 사용자의 관심과 요구를 파악할 수 있었다. 개최 일시 및 장소는 다음과 같다.

<표 47> 한국어 학습자 말뭉치 아카데미 개최 일시 및 장소

구분	일시	장소
제3회	2017년 8월 4일(금) 14:00-18:00	경희대학교 호텔관광대학 401호
제4회	2017년 11월 10일(금) 16:00-18:00	계명대학교 영암관 457호
제5회	2017년 12월 2일(토) 16:30-18:30	연세대학교 백양관 S221호

- 프로그램은 다음과 같다.
 - 한국어 학습자 말뭉치 소개
 - 한국어 학습자 말뭉치 구축을 위한 구어 전사의 실제
 - 학습자 말뭉치의 검색과 활용의 실제

3. 학술대회 발표

- 3차 연도의 학술대회 발표는 그간 구축된 말뭉치를 활용한 연구 성과들을 발표함으로써 학습자 말뭉치 활용 방법론을 공유하고 학습자 말뭉치 기반의 연구의 효용성을 알리는 데에 목적이 있다. 상반기에는 중국 온주대학교에서 열린 ‘온주대 인도국별 연구 국제학술대회’(2017년 4월 21-23일)에서의 발표를 통해 국내뿐만 아니라 국외에도 그 성과를 알리고자 하였다.
 - 중국인 한국어 학습자 말뭉치에 나타난 중간언어 분석 연구(강현화, 2017)
 - 모어 화자와 한국어 학습자 말뭉치에서의 ‘은/는’과‘이/가’의 분포와 조사 선택 요인 분석(한송화, 2017)
 - 학문 목적 학습자의 객관화 전략 사용 양상 연구 - 중국인 학습자의 학술 텍스트를 중심으로(김한샘·배미연, 2017)

4. 이주민 워크숍 개최

- 이주민 워크숍은 이주민 대상의 한국어 교육 현황을 파악하여 수집 방향을 점검하기 위한 목적으로 개최되었다. 결혼이주민, 이주노동자, 중도입국청소년의 한국어교육을 담당하고 있는 기관에서 활동하고 있는 교사, 기관 관계자 등을 패널로 하여 교육 현황을 듣고 효율적인 자료 수집을 위한 의견을 수렴할 수 있었다.

<표 48> 이주민 워크숍 프로그램

일시	주제	발표자	소속
5월 18일	결혼이주민의 한국어교육 현황과 학습자 맞춤형 구축	안 영	경남이주민문화센터 라함
		이현주	충남다문화센터
		박명숙	충북대국제교류센터
5월 25일	이주노동자의 한국어교육 현황과 학습자 맞춤형 구축	문주연	한국외국인력지원센터
		강수정	평화다문화센터
		곽순복	대구인력지원센터
6월 1일	중도입국청소년의 한국어교육 현황과 학습자 맞춤형 구축	인민지	서울대동초등학교
		김수영	서울온드림교육센터
		이영심	서울다솜학교

IV. 학습자 맞춤형 구축 및 가공

1. 수집

1.1. 수집 대상

- 국내 대학 교육 기관의 정규 교육과정 수강 학습자 및 학문 목적 학습자
- 결혼이주민, 이주 노동자, 중도입국청소년

1.2. 교육 기관 참여 방안 마련: 수집 네트워크 마련

(1) 국내 대학 교육 기관

- 3차 연도에는 수집 기관의 자율적인 의사를 반영하여 수집 참여 철회 기관을 제외하고 교육과정상의 등급 체계가 상이한 기관을 제외하여 다음의 14개 기관으로 집약하여 자료를 수집하였다.

고려대학교, 경희대학교, 국민대학교, 연세대학교, 서울대학교, 서강대학교, 이화여자대학교, 경희대학교(국제캠퍼스), 배재대학교, 선문대학교, 계명대학교, 경북대학교, 부산외국어대학교, 동아대학교

(2) 이주민 기관

- 3차 연도부터 본격 구축이 이루어지는 이주민 자료의 수집은 다음과 같이 두 가지 수집 네트워크를 통해 이루어졌다⁵⁾.

5) 이주민 수집 네트워크 마련을 위한 섭외 절차는 다음과 같으며, 수집 개시 이후에도 수집 기관의 추가 섭외를 위한 접촉이 동시에 진행되었다.

섭외 절차	일정	내용
1차 공문 발송	4월	1. 발신: 중앙건강가정진흥원, 법무부 이민통합과 2. 수신: 다문화센터, 사통프로그램 운영 기관 3. 내용: 수집 협조 요청 4. 경로: 부처 간 협의
2차 공문 발송	5월	1. 발신: 한국어학습자맞춤치구축본부 2. 수신: 다문화센터 218개, 사통프로그램 거점 기관 47개, 초·중·고 다문화 예비학교 20개교, 외국인력지원센터 8개 3. 내용: 수집 안내 및 수집 기관 현황 조사지, 수집 지침 4. 경로: 기관 이메일, 팩스 발송 전후 유선 통화
수집 가능 여부 파악 및 섭외	6월-7월	1. 2차 공문 수신 후 자율적으로 참여 여부를 회신 기관 정리 2. 지역별 분포를 고려하여 각 집단별로 10-15개 기관 섭외를 목표로 유선 접촉
수집 교육 및 수집 시작	7월-9월	8월 중순-9월 이후 개강

<표 49> 이주민 수집 네트워크

구분	수집 네트워크	수집 자료
교육 기관	<p>다문화가족지원센터, 사회통합프로그램 운영 기관, 외국인지원센터, 초·중·고 각 급 학교에 속한 12개의 기관에서 자료를 제공한다</p> <p>인천다문화센터, 부여다문화센터, 동두천다문화센터, 대구서구다문화센터, 성주군다문화센터, 평화다문화센터, 경남이주민센터라함, 경기글로벌센터, 서울대동초등학교, 충북대학교/구미대학교 사회통합프로그램센터, 김포외국인주민센터</p> <p>☞ 자료 수집 참여에 동의한 기관 수는 20개 기관이었으나 실제 자료를 수집하여 제공한 기관은 12개 기관임</p>	수업 및 과제 활동, 시험 자료
교사 인력풀	학회 공고를 통해 17명의 수집 교사를 모집함	종적 자료

1.3. 자료 수집 결과

(1) 국내 교육 기관 자료

- 국내 교육 기관 자료는 2017년 봄 학기부터 자료 수집이 시작되어 현재 문어 11,741개, 구어 2,173개 파일이 수집되었다. 문어의 경우 파일 하나당 평균 어절 수를 100어절로 추산하였을 때 약 1,174,10어절 규모의 자료이며 구어의 경우 파일 하나당 평균 어절 수를 350어절로 추산하였을 때 760,500어절에 해당된다⁶⁾.

6) 위의 수집 규모는 2017년 수집된 자료의 단순 누적량이며, 실제 구축 대상이 될 자료를 선별하는 질적 검사 과정에서 구축 불가로 판정되는 자료들이 생김으로 해서 변동이 생

① 문어

- 문어 자료는 현재까지 95개국의 자료가 수집되었으며 숙달도별, 국적별 분포는 다음과 같다.

<표 50> 문어 자료 수집 현황: 국내 교육 기관

	1급	2급	3급	4급	5급	6급	기타 ⁷⁾	합계
중국	986	1,629	1,926	970	358	90	234	6,193
베트남	585	402	430	250	80	14	171	1,932
일본	90	159	178	135	78	25	39	704
대만	81	82	110	102	48	5	4	432
홍콩	56	54	42	54	52	13		271
몽골	73	60	47	28	16	8	6	238
미국	23	65	51	43	11	7	10	210
러시아	22	34	36	32	15	5	4	148
태국	52	36	24	6	5	6	10	139
우즈베키스탄	8	35	19	11	4	3	28	108
말레이시아	20	27	20	15	4	5	4	95
프랑스	9	16	23	25	8	4		85
인도네시아	31	12	15	9	2	3	12	84
미상	5	6	12	24	8	6	18	79
카자흐스탄	8	10	10	22	8	2		60
한국	6	8	5	16	3		13	51
영국	12	11	9	10	2		2	46
노르웨이	11	1	10	7	4			33
싱가포르	7	11	6	2	4		2	32
독일		8	7	6	7	2		30
멕시코	12	10	4	2			2	30
기타	157	102	217	101	47	26	91	741
합계	2,254	2,778	3,201	1,870	764	224	650	11,741

기게 된다. 1-2차 연도 구축 경험을 기준으로 하였을 때 분량 미달, 미완성 작문 또는 발화, 복사 자료나 음질 불량 등의 이유로 구축 불가 판정되는 자료는 전체 자료의 약 30% 이상이 된다.

② 구어

- 구어 자료는 현재까지 38개국의 자료가 수집되었으며 숙달도별, 국적별 분포는 다음과 같다.

<표 51> 구어 자료 수집 현황: 국내 교육 기관

	1급	2급	3급	4급	5급	6급	기타	합계
중국	241	326	288	169	40	4	31	1,099
베트남	193	157	179	76	22	4		631
일본	6	27	30	20	6			89
대만	22	4	20	29	6			81
우즈베키스탄	4	12	4					20
러시아	3	2	11	2				18
몽골	6	6		3	3			18
인도네시아	16		1					17
태국	4	8			3			15
말레이시아	2	2	5	2				11
기타	27	14	83	34	16			174
합계	524	558	621	335	96	8	31	2,173

(2) 이주민 자료

- 이주민 자료는 하반기부터 자료 수집이 본격 시작되어 문어 1,008개 파일, 구어 259개 자료가 수집되었다.

가. 숙달도별

- 이주민 자료의 숙달도별 분포는 다음과 같다⁸⁾.

7) 기타에는 초급, 중급, 고급 등으로 3등급 체계로 구분된 자료, 6급 이상의 최고급 자료, 등급 미확인 자료 등이 포함된다.

8) 이주민 자료의 경우 자료에서 보는 바와 같이 학습자 대상별 소속 기관에 따라 등급 기준이 상이함을 볼 수 있다.

<표 52> 이주민 자료 시험 수집 현황: 숙달도별(자료 수, 개)

결혼이주민			이주노동자			중도입국청소년		
수준	문어	구어	수준	문어	구어	수준	문어	구어
1단계	121	25	1단계	33		초1	3	3
2단계	161	32	2단계	4	5	초2	8	8
3단계	150	54	3단계	17		초3	11	11
4단계	118	10	4단계	8		초4	11	11
5단계	23	7	5단계	4		초5	2	2
기초			기초	10	12	초6	5	5
초급	17	1	초급	22	14	중1	5	
중급	13	1	중급	5		중2	3	
고급	17	8	고급			중3	10	
토픽2급			토픽2급	1		고1	7	
토픽3급	14	7	토픽3급			고2	4	
토픽4급	11	10	토픽4급	2		고3	3	
토픽5급	9	7	토픽5급			1단계	2	
토픽6급	4	3	토픽6급			4단계	2	
토픽반	5	5	토픽반			등급 미상	1	
등급 미상	54	12	등급 미상	108	6	-	-	-
합계	717	182	합계	214	37	합계	77	40

나. 국적별

○ 이주민 자료의 국적별 분포는 다음과 같다.

<표 53> 이주민 자료 시험 수집 현황: 국적별(자료 수, 개)

국적	결혼이주민		이주노동자		중입국청소년	
	문어	구어	문어	구어	문어	구어
베트남	254	80	40	18	2	
중국	125	23	15		45	32
캄보디아	58	14	29	1	1	

러시아	65	1	10		12	
필리핀	44	12	3		1	
네팔	11	3	32	6		
일본	27	21	2			
우즈베키스탄	28	5	1	1	2	
파키스탄	22	10	2			
방글라데시	7	2	22	2	1	
미얀마	4	2	23			
태국	15	1	3	1		
한국	3	1			7	7
인도네시아	5		6	7		
스리랑카	3		12			
영국	8	1	3			
키르기스스탄	1	1	7			
대만	8					
몽골	7				1	
카자흐스탄	4		1		2	
페르시아	3	1	2			
미국	4	1				
아프가니스탄	3	1				
에리트리아	1	1				
아랍	1		1			
에티오피아	2					
인도	2					
리투아니아	1					
수단					1	
시리아					1	
프랑스	1					
국적 미상				1	1	1
합계	717	182	214	37	77	40

<이주민 교육 기관 자료 수집의 문제>

- 여성가족부 중앙건강가정지원센터, 법무부 이민통합국, 교육청 등과의 협의를 통해 자료 수집 협조 공문을 보내고 이메일, 유선 통화 등을 통해 수집 기관 섭외를 진행하였으나 강제 사항이 아니므로 공문의 영향력이 크게 미치지 못하였으며, 오히려 잦은 공문으로 반감을 보이는 기관도 일부 있었음
- 대규모 기관의 경우 자료 수집, 설문 요청을 받은 경험이 많고 운영의 어려움, 부가적인 업무로 인한 피로 등을 이유로 수집을 거절하는 하는 경우가 많음
- 기관장 및 관리자 - 자료 수집 총괄 교사 - 각 학급의 담임교사의 3단 체계로 기관장 및 관리자가 수집 협조에 응하였어도 자원봉사 또는 시간 강사들이 자료 수집 의지가 적고, 전반적으로 사업에 대한 이해가 부족한 편임
- 일부 기관을 제외하고 기관 규모가 그리 크지 않고 기관에 따라 등록 인원과 실제 출석 인원의 차가 커서 수집 가능 자료의 규모를 기준으로 산술적으로 계산할 때 국내 대학의 교육 기관에 비해 수집 기관의 섭외, 교육, 관리에 약 10배 이상의 노력이 필요함
- 실질적으로 학습자와 대면하여 동의서, 자료를 수집해야 하는 담임교사가 수집 총괄 교사를 통해 수집 지침을 전달받아 수집을 하고 있음. 어학당과 달리 수집 총괄 교사가 영향력을 행사할 수 있는 구조가 아니라 자료 수집의 양과 질을 담보하기가 쉽지 않음

(3) 종적 자료

- 종적 자료는 17명의 교사를 통해 총 42명의 자료를 수집하고 있다. 최초에 국내 기관 17명, 결혼이주민 22명, 이주노동자 6명, 중도입국청소년 초등 4명, 중등 2명, 고등 2명으로 총 53명을 선발하여 수집을 시작하였으나 임신, 이사, 장기 결석, 부모의 이혼 등의 사유로 인해 11명의 학습자가 수집을 철회하였다.

<표 54> 2017년 종적 자료 학습자 분포

	국내 기관	결혼 이주민	이주 노동자	중도입국청소년			합계
				초등	중등	고등	
우즈베키스탄	2	3		1			6
중국	3	2				1	6
러시아		3		2			5
베트남	3	1	1				5
필리핀		4					4
스리랑카	2						2
인도네시아	1	1	1				2
카자흐스탄	2						2
캄보디아		1	1				2
태국		2					2
한국					1	1	2
몽골			1				1
미얀마	1						1
방글라데시		1	1				2
호주	1						1
합계	14	17	5	3	1	2	42

<이주민 종적 자료 수집의 문제>

- 종적 자료는 학습자의 언어 발달 과정을 추적 관찰하기 위한 자료로 2주 간격을 두고 정기적으로 수집할 것으로 기본 지침으로 하고 있다. 그러나 이주민의 경우 육아, 가계 상황, 직장 생활, 질병, 농번기 등과 같이 피하기 어려운 개인 사유로 인해 정기적 수집이 이루어지지 못하는 경우가 많다. 이에 따라 2주 간격으로 수집하는 것을 원칙으로 하되 다소 늦어지더라도 지속적인 수집을 하도록 탄력적으로 운영하고 있다.
- 이주민 학습자가 수강 중인 교육과정은 국내 대학의 교육 기관에서 운영 중인 집중 코스와 달리 수업 시수가 총 400시간 내외로 매우 짧고 주당 수업 시간이 1회당 2시간씩 총 2회로 학습 내용이 많지 않다. 따라서 이주민 교재 분석 등을 통해 수집 과제를 마련하였지만 난이도가 높다는 피드백이 많다. 이에 따라 수집 지침의 과제를 기본으로 하되 학습

자가 최대한 발화를 산출할 수 있도록 상황에 맞추어 주제를 자유롭게 확장하여 수집을 진행하도록 권고하고 있다.

2. 구축 및 가공

2.1. 원시 말뭉치

(1) 문어

- 2017년 문어 원시 말뭉치는 505,524어절이 구축되었다. 그 결과 1-2차 연도에 기구축한 말뭉치와 합산하여 누적 합계 1,507,948어절 규모의 문어 원시 말뭉치가 구축되었다.

① 문어 말뭉치의 숙달도별 자료 분포

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

<표 55> 문어 원시 말뭉치의 숙달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	어절 수	49,358	50,210	50,705	49,974	50,194	50,556	0	300,997
	파일 수	697	451	422	385	345	316	0	2,616
2016	어절 수	113,479	125,359	122,196	141,259	111,125	88,009	0	701,427
	파일 수	1,536	1,068	876	1,058	722	508	0	5,768
2017	어절 수	83,545	63,611	52,753	34,801	63,310	78,392	129,112	505,524
	파일 수	1,302	842	505	293	436	502	115	3,995
합계	어절 수	246,382	239,180	225,654	226,034	224,629	216,957	129,112	1,507,948
	파일 수	2,838	1,910	1,381	1,351	1,158	1,010	115	12,379

② 문어 말뭉치의 국적별 자료 분포

- 문어 말뭉치는 이중 국적 학습자를 포함하여 총 118개국의 자료가 구축되었다.

<표 56> 문어 원시 말뭉치의 국적별 자료 분포

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	177,617	1,597	290,671	2,436	246,895	1,607	715,183	5,640
일본	47,154	384	111,923	834	44,425	441	203,502	1,659
베트남	6,923	61	35,699	302	41,720	455	84,342	818
대만	13,032	96	38,093	301	12,937	69	64,062	466
미국	5,351	48	28,770	236	14,771	143	48,892	427
홍콩	6,975	35	29,097	103	8,664	128	44,736	266
말레이시아	4,624	53	14,381	223	18,037	65	37,042	341
태국	4,151	34	16,552	136	12,109	118	32,812	288
러시아	1,844	17	15,656	140	8,979	116	26,479	273
몽골	1,563	14	10,725	91	10,268	87	22,556	192
카자흐스탄	4,120	35	10,302	80	7,770	71	22,192	186
이탈리아	1,915	5	2,473	18	11,909	76	16,297	99
우즈베키스탄	1,988	18	4,817	37	8,451	110	15,256	165
인도네시아	1,577	11	4,913	71	7,482	23	13,972	105
싱가포르	3,524	2	6,836	7	2,565	71	12,925	80
프랑스	1,276	23	9,001	46	1,812	15	12,089	84
한국	1,242	1	5,421	16	2,108	30	8,771	47
필리핀	483	11	1,632	41	4,985	20	7,100	72
호주	831	12	4,253	34	1,792	21	6,876	67
독일	1,345	14	4,017	25	1,499	23	6,861	62
스리랑카	122	11	2,903	42	3,766	12	6,791	65
캐나다	1,324	9	3,598	28	1,729	18	6,651	55
키르기스스탄	69	14	2,321	46	4,246	62	6,636	122
영국	681	6	3,536	32	2,257	29	6,474	67

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
캄보디아	199	13	994	50	4,214	4	5,407	67
스웨덴	1,276	10	3,426	35	330	2	5,032	47
사우디아라비아	875	1	2,963	7	277	26	4,115	34
투르크메니스탄	0	6	46	12	3,447	8	3,493	26
터키	670	3	1,670	12	1,081	16	3,421	31
미얀마	351	5	1,425	22	1,525	2	3,301	29
스페인	445	3	2,628	8	161	25	3,234	36
방글라데시	111	1	437	22	2,559	3	3,107	26
파키스탄	92	0	909	21	1,548	0	2,549	21
튀니지아	0	2	2,110	13	0	6	2,110	21
노르웨이	94	1	1,125	12	768	5	1,987	18
인도	149	5	1,244	6	517	2	1,910	13
네덜란드	276	0	1,046	1	391	20	1,713	21
아제르바이잔	652	2	596	8	428	3	1,676	13
브라질	112	1	1,274	9	108	2	1,494	12
콜롬비아	0	3	843	9	485	6	1,328	18
멕시코	283	1	679	11	351	1	1,313	13
우크라이나	167	4	823	4	274	1	1,264	9
헝가리	173	1	756	9	303	0	1,232	10
이란	236	1	855	5	121	3	1,212	9
마카오	143	2	875	6	181	1	1,199	9
이라크	0	0	162	3	960	4	1,122	7
에콰도르	162	1	960	7	0	0	1,122	8
폴란드	78	3	1,020	4	0	0	1,098	7
네팔	149	2	252	5	668	2	1,069	9
기타 ⁹⁾	4,573	34	14,719	142	3,651	43	22,943	219
합계	300,997	2,616	701,427	5,768	505,524	3,995	1,507,948	12,379

9) 기타에는 엘살바도르, 칠레, 벨기에, 뉴질랜드, 핀란드, 스위스, 모로코, 포르투갈, 아르메니아, 벨라루스, 이집트, 르완다, 페루, 나이지리아, 아프카니스탄, 이디오피아, 루마니아, 볼리비아, 불가리아, 가나, 브루나이, 룩셈부르크, 베네수엘라, 도미니카 공화국, 라오스,

(2) 구어

- 2017년 구어 원시 말뭉치는 178,572어절이 구축되었다. 그 결과 1-2차 연도에 기구축한 말뭉치와 합산하여 누적 합계 327,723어절 규모의 구어 원시 말뭉치가 구축되었다.

① 구어 말뭉치의 숙달도별 자료 분포

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

<표 57> 구어 원시 말뭉치의 숙달도별 자료 분포

구분		1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	어절 수	4,780	9,847	8,334	9,056	10,147	7,236	0	49,400
	파일 수	27	17	17	14	13	6	0	94
2016	어절 수	34,354	12,936	11,302	13,330	11,050	16,789	0	99,761
	파일 수	91	40	37	48	22	17	0	255
2017	어절 수	22,123	28,648	37,121	24,088	31,510	29,317	5,765	178,572
	파일 수	70	57	82	50	52	34	4	349
합계	어절 수	61,257	51,431	56,757	46,474	52,707	53,342	5,765	327,733
	파일 수	188	114	136	112	87	57	4	698

② 구어 말뭉치의 국적별 자료 분포

- 구어 말뭉치는 이중 국적 학습자를 포함하여 총 62개국의 자료가 구축되었다.

가봉, 체코, 세르비아, 남수단, 아르헨티나, 오스트리아, 동티모, 트리니다드 토바고, 아랍 에미리트연합, 마다가스카르, 이스라엘, 잠비아, 니카라과, 예멘, 덴마크, 탄자니아, 파라과이, 코스타리카, 모잠비크, 시리아, 우루과이, 자메이카, 슬로바키아, 쿠웨이트, 저지, 케냐, 세네갈, 파나마, 우간다, 온두라스, 리투아니아, 과테말라, 알제리, 타지키스탄, 쿠바, 콩고, 리비아, 아일랜드, 정보 없음, 콩고 민주 공화국, 슬로베니아, 카메룬, 그루지아, 남아프리카가 포함된다.

<표 58> 구어 원시 말뭉치의 국적별 자료 분포

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	5,946	15	52,333	133	66,624	123	124,903	271
베트남	1,802	5	8,376	22	33,011	54	43,189	81
일본	7,905	14	14,589	34	13,633	24	36,127	72
대만	4,055	5	6,782	15	8,458	9	19,295	29
키르기스스탄	0	0	0	3	13,000	0	13,000	3
태국	1,348	3	618	2	10,125	23	12,091	28
말레이시아	1,220	1	1,060	0	9,373	0	11,653	1
미국	3,803	6	3,171	5	1,645	4	8,619	15
우즈베키스탄	1,828	2	478	3	4,768	10	7,074	15
필리핀	0	1	331	1	5,958	2	6,289	4
러시아	2,071	3	1,167	4	1,323	5	4,561	12
몽골	763	2	383	1	3,216	20	4,362	23
캄보디아	0	3	217	0	2,886	0	3,103	3
카자흐스탄	2,492	3	504	2	0	0	2,996	5
영국	2,240	3	0	0	705	2	2,945	5
한국	729	0	232	0	1,057	25	2,018	25
호주	910	1	0	1	1,032	1	1,942	3
이탈리아	0	0	1,718	1	169	13	1,887	14
스페인	955	0	796	0	0	1	1,751	1
프랑스	1,255	2	467	1	0	0	1,722	3
스위스	1,212	0	0	1	0	0	1,212	1
미얀마	650	1	448	3	0	0	1,098	4
캐나다	161	3	492	0	417	2	1,070	5
기타 ¹⁰⁾	8,055	21	5,599	22	1,172	31	14,826	74
합계	49,400	94	99,761	255	178,572	349	327,733	698

10) 기타에는 스리랑카, 싱가포르, 독일, 스웨덴, 인도네시아, 네덜란드, 사우디아라비아, 세르비아, 벨기에, 콜롬비아, 코스타리카, 터키, 알제리, 멕시코, 방글라데시, 이집트, 벨라루스, 정보 없음, 불가리아, 파라과이, 아르헨티나, 아제르바이잔, 헝가리, 모로코, 파키스탄, 홍콩, 네팔, 투르크메니스탄, 인도, 우크라이나, 이라크, 폴란드, 엘살바도르, 뉴질랜드

2.2. 형태 주식 말뭉치

- 2016년 형태 주식 말뭉치는 문어 405,199어절, 구어 106,038어절이 구축되었다. 1-2차 연도에 기구축한 말뭉치와 합산한 형태 주식 말뭉치의 누적합계는 문어 1,107,299어절, 구어 178,256어절이다.

(1) 형태 주식 말뭉치의 숙달도별 자료 분포

- 형태 주식 말뭉치의 숙달도별 자료 분포는 다음과 같다.

<표 59> 형태 주식 말뭉치의 숙달도별 자료 분포

구분			1급	2급	3급	4급	5급	6급	6급 이상	합계
2015	문어	어절 수	31,664	30,422	35,948	37,422	34,153	31,381	0	200,990
		파일 수	456	269	311	292	235	203	0	1,766
	구어	어절 수	3,033	3,687	2,032	3,941	3,511	4,985	0	21,189
		파일 수	16	5	5	7	5	4	0	42
2016	문어	어절 수	77,173	84,225	100,489	87,887	78,007	73,329	0	501,110
		파일 수	1,044	718	724	598	506	433	0	4,023
	구어	어절 수	7,154	7,179	6,820	8,852	7,430	13,594	0	51,029
		파일 수	22	26	25	33	18	14	0	138
2017	문어	어절 수	60,736	70,413	48,717	72,211	52,093	48,431	52,598	405,199
		파일 수	847	652	352	590	348	288	55	3,132
	구어	어절 수	26,827	13,632	11,823	8,700	18,838	20,453	5,765	106,038
		파일 수	69	34	36	29	24	22	4	218
합계	문어	어절 수	169,573	185,060	185,154	197,520	164,253	153,141	52,598	1,107,299
		파일 수	2,347	1,639	1,387	1,480	1,089	924	55	8,921
	구어	어절 수	37,014	24,498	20,675	21,493	29,779	39,032	5,765	178,256
		파일 수	107	65	66	69	47	40	4	398

드, 아르메니아, 베네수엘라, 동티모, 트리니다드 토바고, 슬로바키아가 있다.

(2) 형태 주식 말뭉치의 국적별 자료 분포

① 문어

- 문어 형태 주식 말뭉치는 114개국의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 60> 문어 형태 주식 말뭉치의 국적별 자료 분포

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	122,101	1,110	193,909	1,610	231,675	1,633	547,685	4,353
일본	28,442	227	109,310	810	22,170	187	159,922	1,224
대만	7,636	59	23,323	173	21,064	171	52,023	403
베트남	5,502	50	29,173	222	9,080	104	43,755	376
홍콩	4,669	35	17,224	121	16,123	129	38,016	285
미국	3,740	31	23,792	200	6,716	54	34,248	285
말레이시아	3,024	27	5,710	63	13,184	81	21,918	171
태국	3,089	24	8,141	43	9,597	100	20,827	167
러시아	1,048	10	11,675	100	5,105	49	17,828	159
카자흐스탄	3,323	10	7,967	62	3,132	51	14,422	123
몽골	964	28	8,139	65	4,467	22	13,570	115
프랑스	751	6	6,855	52	3,724	37	11,330	95
싱가포르	1,743	9	6,406	22	2,211	59	10,360	90
인도네시아	1,034	13	2,107	40	6,477	16	9,618	69
우즈베키스탄	1,390	10	3,395	13	2,020	40	6,805	63
한국	891	13	2,750	30	3,022	12	6,663	55
독일	1,345	11	1,941	25	2,674	20	5,960	56
이탈리아	575	5	2,560	22	2,773	30	5,908	57
키르기스스탄	69	8	1,480	13	4,269	24	5,818	45
호주	218	6	3,184	27	1,844	13	5,246	46
캐나다	765	3	2,529	28	1,628	7	4,922	38
스웨덴	853	1	1,049	10	2,800	30	4,702	41

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
영국	367	7	3,090	22	760	23	4,217	52
사우디아라비아	670	3	1,525	25	1,643	10	3,838	38
투르크메니스탄	0	1	46	20	3,447	2	3,493	23
스페인	366	1	1,748	3	959	24	3,073	28
스리랑카	122	4	2,752	15	151	8	3,025	27
방글라데시	111	0	163	1	2,404	20	2,678	21
터키	670	5	789	5	881	13	2,340	23
필리핀	483	6	385	5	1,247	7	2,115	18
튀니지아	0	1	122	6	1,988	9	2,110	16
미얀마	245	0	1,449	1	82	20	1,776	21
인도	32	1	795	3	714	9	1,541	13
브라질	112	2	247	6	1,027	3	1,386	11
네덜란드	203	2	374	12	745	1	1,322	15
아제르바이잔	230	2	696	3	322	7	1,248	12
멕시코	283	2	296	4	653	6	1,232	12
노르웨이	94	2	283	4	842	3	1,219	9
캄보디아	199	1	474	4	520	7	1,193	12
에콰도르	162	1	50	8	910	1	1,122	10
폴란드	78	0	964	2	56	6	1,098	8
이란	0	1	472	3	619	3	1,091	7
마카오	0	1	638	3	380	2	1,018	6
파키스탄	0	2	871	4	130	1	1,001	7
기타 ¹¹⁾	3,391	25	10,262	113	8,964	78	22,617	216
합계	200,990	1,766	501,110	4,023	405,199	3,132	1,107,299	8,921

11) 기타에는 우크라이나, 엘살바도르, 헝가리, 콜롬비아, 뉴질랜드, 모로코, 칠레, 벨기에, 벨라루스, 핀란드, 이집트, 르완다, 나이지리아, 루마니아, 볼리비아, 불가리아, 이디오피아, 페루, 브루나이, 룩셈부르크, 베네수엘라, 가나, 도미니카 공화국, 네팔, 라오스, 가봉, 세르비아, 아프가니스탄, 남수단, 아르헨티나, 포르투갈, 오스트리아, 스위스, 동티모, 트리니다드 토바고, 마다가스카르, 잠비아, 체코, 니카라과, 탄자니아, 파라과이, 이스라엘, 모잠비크, 우루과이, 슬로바키아, 이라크, 아르메니아, 코스타리카, 케냐, 세네갈, 예멘, 우간다, 파나마, 쿠웨이트, 온두라스, 과테말라, 타지키스탄, 쿠바, 시리아, 자메이카,

② 구어

- 구어 형태 주식 말뭉치는 52개국의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 61> 구어 형태 주식 말뭉치의 국적별 자료 분포

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	3,914	10	15,600	40	63,990	124	83,504	174
일본	3,146	4	10,068	25	13,558	34	26,772	63
대만	989	1	4,967	12	7,770	11	13,726	24
베트남	520	1	5,783	15	3,875	11	10,178	27
미국	1,565	3	1,871	3	3,442	7	6,878	13
러시아	568	1	1,167	4	1,886	3	3,621	8
말레이시아	405	1	554	3	1,736	3	2,695	7
태국	356	2	1,177	2	1,139	4	2,672	8
영국	660	0	0	1	1,854	0	2,514	1
카자흐스탄	1,583	1	504	1	0	2	2,087	4
이탈리아	0	0	347	1	1,540	3	1,887	4
몽골	575	2	383	2	403	0	1,361	4
우즈베키스탄	841	2	478	0	0	1	1,319	3
미얀마	0	0	448	0	650	0	1,098	0
캐나다	0	1	492	0	578	3	1,070	4
호주	910	0	0	1	108	2	1,018	3
기타 ¹²⁾	5,157	13	7,190	28	3,509	10	15,856	51
합계	21,189	42	51,029	138	106,038	218	178,256	398

콩고, 리비아, 아일랜드, 콩고 민주 공화국, 리투아니아, 슬로베니아, 카메룬, 그루지아, 알제리가 있다.

- 12) 기타에는 스리랑카, 프랑스, 싱가포르, 독일, 스웨덴, 인도네시아, 네덜란드, 스페인, 사우디아라비아, 세르비아, 벨기에, 코스타리카, 터키, 알제리, 멕시코, 이집트, 벨라루스, 불가리아, 키르기스스탄, 파라과이, 아르헨티나, 필리핀, 아제르바이잔, 헝가리, 한국, 캄보디아, 모로코, 홍콩, 네팔, 투르크메니스탄, 브라질, 노르웨이, 이란, 마카오, 엘살바도르, 칠레, 나이지리아, 룩셈부르크, 체코, 이스라엘, 아르메니아가 있다.

2.3. 오류 주석 말뭉치

- 2017년 오류 주석 말뭉치는 문어 101,947어절, 구어 60,344어절이 구축되었다. 1-2차 연도에 기구축한 말뭉치와 합산한 누적 합계는 문어 244,831어절, 구어 124,313어절이다.

(1) 오류 주석 말뭉치의 숙달도별 자료 분포

- 오류 주석 말뭉치의 숙달도별 자료 분포는 다음과 같다.

<표 62> 오류 주석 말뭉치의 숙달도별 자료 분포

구분			1급	2급	3급	4급	5급	6급	합계
2015	문어	어절 수	6,804	7,651	6,604	7,141	7,399	7,285	42,884
		파일 수	96	64	48	51	48	45	352
	구어	어절 수	1,675	1,890	2,032	1,917	1,978	2,286	11,778
		파일 수	9	3	5	4	3	2	26
2016	문어	어절 수	14,843	16,138	17,747	16,120	18,590	16,562	100,000
		파일 수	233	152	140	133	136	96	890
	구어	어절 수	7,860	7,179	6,820	8,852	8,963	12,517	52,191
		파일 수	25	26	25	33	20	13	142
2017	문어	어절 수	18,593	16,516	17,993	17,805	14,431	16,609	101,947
		파일 수	260	154	141	132	88	102	877
	구어	어절 수	7,948	10,905	11,823	10,724	12,019	6,925	60,344
		파일 수	28	30	36	32	14	7	147
합계	문어	어절 수	40,240	40,305	42,344	41,066	40,420	40,456	244,831
		파일 수	589	370	329	316	272	243	2,119
	구어	어절 수	17,483	19,974	20,675	21,493	22,960	21,728	124,313
		파일 수	62	59	66	69	37	22	315

(2) 오류 주석 말뭉치의 국적별 자료 분포

① 문어

- 문어 오류 주석 말뭉치는 총 73개 국적의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 63> 문어 오류 주석 말뭉치의 국적별 자료 분포

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	22,591	193	32,128	288	16,734	156	71,453	637
일본	11,113	88	24,358	197	14,867	132	50,338	417
미국	1,410	5	8,786	33	12,159	131	22,355	169
대만	1,111	10	3,624	78	16,285	99	21,020	187
베트남	456	8	4,130	36	15,688	131	20,274	175
태국	556	4	1,361	12	5,907	48	7,824	64
러시아	379	3	3,085	29	3,357	31	6,821	63
말레이시아	0	0	2,117	18	4,426	34	6,543	52
카자흐스탄	2,381	18	665	6	3,341	32	6,387	56
캐나다	116	2	494	8	2,558	17	3,168	27
영국	287	1	824	6	1,799	24	2,910	31
프랑스	199	0	2,496	7	0	15	2,695	22
호주	0	1	674	18	1,834	0	2,508	19
홍콩	645	1	1,567	16	0	0	2,212	17
사우디아라비아	0	2	389	12	1,480	1	1,869	15
몽골	213	5	1,288	14	105	0	1,606	19
독일	124	0	1,173	4	0	12	1,297	16
우즈베키스탄	348	2	819	5	0	0	1,167	7
인도네시아	77	0	1,007	3	0	8	1,084	11
싱가포르	135	1	926	10	0	0	1,061	11
스페인	0	1	289	8	755	0	1,044	9
기타 ¹³⁾	743	7	7,800	82	652	6	9,195	95

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
합계	42,884	352	100,000	890	101,947	877	244,831	2,119

② 구어

- 구어 오류 주석 말뭉치는 총 52개 국적의 자료가 구축되었으며 국적별 자료 분포는 다음과 같다.

<표 64> 구어 오류 주석 말뭉치의 국적별 자료 분포

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
중국	925	5	16,129	42	17,629	49	34,683	96
일본	1,297	1	10,068	25	15,407	37	26,772	63
대만	989	2	4,967	3	5,613	8	11,569	13
베트남	520	1	5,783	12	3,875	10	10,178	23
미국	927	1	1,871	15	4,080	11	6,878	27
러시아	568	1	1,167	4	1,886	3	3,621	8
말레이시아	286	1	554	2	1,855	5	2,695	8
태국	356	1	1,177	3	1,139	3	2,672	7
영국	660	0	0	1	1,854	2	2,514	3
몽골	0	1	383	0	978	0	1,361	1
카자흐스탄	627	1	504	2	0	0	1,131	3
미얀마	0	0	448	3	650	0	1,098	3
캐나다	0	1	492	0	578	3	1,070	4
호주	648	0	262	2	108	2	1,018	4
기타 ¹⁴⁾	3,975	10	8,386	28	4,692	14	17,053	52

- 13) 기타에는 미얀마, 스웨덴, 키르기스스탄, 이탈리아, 한국, 네덜란드, 아르헨티나, 모로코, 우크라이나, 터키, 인도, 마카오, 니카라과, 네팔, 아제르바이잔, 헝가리, 포르투갈, 폴란드, 이란, 슬로바키아, 뉴질랜드, 가나, 마다가스카르, 멕시코, 이집트, 볼리비아, 루마니아, 도미니카 공화국, 가봉, 콜롬비아, 오스트리아, 과테말라, 노르웨이, 핀란드, 캄보디아, 세르비아, 스위스, 필리핀, 불가리아, 나이지리아, 파키스탄, 르완다, 페루, 라오스, 세네갈, 이디오피아, 스리랑카, 벨기에, 코스타리카, 알제리, 벨라루스, 파라과이가 있다.

국적	2015		2016		2017		합계	
	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수	어절 수	파일 수
합계	11,778	26	52,191	142	60,344	147	124,313	315

2.4. 1-2차 연도 구축 말뭉치의 정제

- 1-2차 연도 구축 말뭉치의 정제는 2017년 말뭉치 구축 작업 과정에서 발견되는 오류를 소급하여 수정하고, 지침에 새롭게 반영되는 사항들을 추가적으로 반영하는 방식으로 진행되었다. 1-2차 연도에 기구축한 정제 대상 말뭉치의 구축 규모는 다음과 같다.

<표 65> 1-2차 연도 기구축 학습자 말뭉치의 규모

구분		1급	2급	3급	4급	5급	6급	합계
원시	문어	162,837 (2,233)	175,569 (1,519)	172,901 (1,298)	191,233 (1,443)	161,319 (1,067)	138,565 (824)	1,002,424 (8,384)
	구어	39,134 (118)	22,783 (57)	19,636 (54)	22,386 (62)	21,197 (35)	24,025 (23)	149,161 (349)
형태 주석	문어	108,837 (1,500)	114,647 (987)	136,437 (1,035)	125,309 (890)	112,160 (741)	104,710 (636)	702,100 (5,789)
	구어	10,187 (38)	10,866 (31)	8,852 (30)	12,793 (40)	10,941 (23)	18,579 (18)	72,218 (180)
오류 주석	문어	21,647 (329)	23,789 (216)	24,351 (188)	23,261 (184)	25,989 (184)	23,847 (141)	142,884 (1,242)
	구어	9,535 (34)	9,069 (29)	8,852 (30)	10,769 (37)	10,941 (23)	14,803 (15)	63,969 (168)

- 14) 기타에는 스리랑카, 프랑스, 싱가포르, 독일, 스웨덴, 인도네시아, 네덜란드, 스페인, 사우디아라비아, 세르비아, 벨기에, 이탈리아, 코스타리카, 터키, 알제리, 멕시코, 이집트, 우즈베키스탄, 벨라루스, 불가리아, 키르기스스탄, 파라과이, 아르헨티나, 필리핀, 아제르바이잔, 헝가리, 한국, 캄보디아, 모로코, 홍콩, 네팔, 마카오, 니카라과, 이란, 오스트리아, 과테말라, 르완다, 라오스가 포함된다.

① 원시 말뭉치

- 2017년 형태 주석과 오류 주석은 2016년에 구축된 원시 말뭉치를 대상으로 한다. 형태 주석과 오류 주석 작업 과정에서 원시 말뭉치에 남아 있는 오류나 지침 업데이트에 의해 수정이 필요한 항목들을 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 진행되었다. 새로운 지침을 반영하여 수정한 결과는 공동 연구원들이 최종 검토를 통해 마지막 검증 단계를 거쳤다.

V. 구축 말뭉치 검증 정교화

1. 문어 입력 및 구어 전사 검수

1.1. 형태소 주석, 오류 주석 작업 단계에서의 심층 검수

- 문어 입력과 구어 전사 텍스트는 1차 검수(작업자 간의 교차 검수)와 2차 검수(상위 작업자의 검수)를 통해 형태 주석 단계로 넘어가게 된다. 이들 자료는 형태 주석 또는 그 다음 단계인 오류 주석 작업자에 의해 3차 검수의 과정을 거치게 된다. 이 단계에서 사소한 오류는 형태 주석 작업자와 오류 주석 작업자가 원문 수정 기능을 사용하여 직접 수정하고, 전반적으로 수정이 필요한 경우는 문어 입력과 구어 전사 단계로 되돌려 보내 수정 작업을 하도록 하였다.

1.2. 작업 중 생성된 오조작 데이터 검증

- 표본 정보는 자료의 관리, 구축 통계, 변인별 검색 등을 위한 기초 정보로 정확성을 요구한다. 따라서 최초의 등록과 검수 단계에서도 신중한 작업이 필요하지만 전체 구축 자료의 표본 정보에 대한 검증도 중요하다. 이에 따라 본 연구에서는 다음과 같이 표본 정보 중 각 변인별로 정상 범위의 값이 아닌 값이 입력되거나 미입력 항목을 표본을 추출하여 해당 표본 정보 검수 작업을 진행하였다.

<표 66> 표본 정보에 생성된 오조작 데이터 검증

변인	정상 등록 값	비정상 표본 수	비고
성별	남성/여성	50	미입력
연령	7-80(정상 언어 수집 범위)	208	미입력, 해당범위 밖
현재등급	1, 2, 3, 4, 5, 6, 6급 이상	47	미입력, 해당범위 밖
국적	1개 국적	14	미입력
교포 여부	교포/외국인	1,529	미입력
모국어	1개 모국어	170	미입력
학습 기간	개월 단위	268	미입력, 년/개월 등 입력단위 확인 불가
거주 기간	개월 단위	271	미입력, 년/개월 등 입력단위 확인 불가
학습 목적	다수의 목적	58	미입력
직업	다수의 직업	788	미입력
소속	1개 학교명	2,859	미입력, 학교 등 기관명이 아닌 경우
학습자 유형	일반/학습/종적/이주	48	미입력, 해당범위 밖(표본 정보 있는 표본 중)
학습 환경	국내/국외	33	미입력
자료 유형	문어/구어	13	미입력

2. 형태·오류 주식 검수 시스템 정교화

2.1. 작업 중 생성된 오조작 데이터 검증

- 작업 중 생성된 오조작 데이터 검증은 작업 로그나 표본 정보, 주식 정보, 구축 시기 정보 유형을 추출하여 추가적인 검수가 필요할 것으로 판단되는 표본에 대한 검수 작업을 진행하는 것이다.

(1) 작업 진행 이상 여부 검증

- 작업 로그(log)를 검색하여 작업 진행상의 이상이 의심되는 표본을 검증하는 것이다. 다음의 결과에서 작업자가 없이 검수자만 있는 표본의 경우 검토해 본 결과 구축 도구 사용 초기의 작업 이력이 남지 않은 경우로 실제 표본의 문제는 아닌 것으로 확인되었다. 한편, 형태 주식 또는 오류 주식 작업 연도는 없지만 오류 주식 이력이 있는 표본은 작업 시 구축 시기 정보 할당이 누락된 경우로 해당 정보를 등록하여 수정하였다. 마지막으로 검수 시간이 30초 미만인 표본이 925개가 발견되었는데, 이들은 지침 중 미확정 사항이 있어 검수 작업 후 곧바로 [완료] 버튼을 누르지 못하고 지침이 최종 확정된 후에 [완료] 버튼을 누른 항목들이므로 확인되었다.

<표 67> 오조작 데이터 검증: 작업 진행 이상 여부 검증

	작업자	검수자	결과(건)	비고
원문 전사 작업 표본 수	×	O	502	log 생성:
형태 주식 작업 표본 수	×	O	202	
오류 주식 작업 표본 수	×	O	2	
원문 전사 작업 표본 수	×	×	-	표본 정보 중 각 단계 작업이 없는데 이력이 있는 표본
형태 주식 작업 표본 수	×	×	-	
오류 주식 작업 표본 수	×	×	-	
원문 전사 작업 작업 수	×	O	-	표본 번호 참조

형태 주석 작업 작업 수	×	O	-	
오류 주석 작업 작업 수	×	O	-	
원문 전사 작업 작업 수	×	×	-	
형태 주석 작업 작업 수	×	×	-	
오류 주석 작업 작업 수	×	×	-	
원문 전사 작업 연도는 없지만 전사 이력은 있는 표본			-	
형태 주석 작업 연도는 없지만 형태 주석 이력은 있는 표본			2	
오류 주석 작업 연도는 없지만 오류 주석 이력은 있는 표본			42	
검수 시간 30초 미만 표본 수			925	검수작업 할당 후 작업완료까지의 시 간

(2) 작업 결과 이상 표본 검증

- 작업 결과 이상 여부는 다음과 같이 원형태와 주석, 교정 형태와 주석의 존재 여부, 주석이 누락된 항목을 살펴 작업 결과의 이상이 의심되는 표본을 검증하는 단계이다. 다음의 결과 중 원형태가 있으나 주석이 없는 경우, 원형태도 없고 주석도 없는 경우, 교정 형태는 있으나 주석이 없는 경우는 해당 표본을 확인하여 수정 조치하였다. 한편, 21,921건으로 많은 표본에서 발견된 ‘오류 주석이 모두 태깅되지 않은 형태’는 오류 주석 지침에서 오류 층위의 경우 필수 주석인 반면, 오류 양상이나 오류 위치의 경우 수의적 주석으로 두고 있어 하나의 형태에 세 가지 주석이 반드시 있어야 하는 것이 아니므로 오류 의심 항목이 아닌 것으로 파악되었다.

<표 68> 오조작 데이터 검증: 작업 결과 이상 표본 검증

	형태	주석	결과(건)	비고
원형태	O	×	-	
원형태	×	O	20	
원형태	×	×	12	

교정 형태	O	×	12	
교정 형태	×	O	-	
교정 형태	×	×	-	
오류 주석이 모두 태깅되지 않은 형태			21,921	오류주석이 있을 경우 최소 3개(오류 충위, 양상, 위치)가 있어야 하나 1개 또 는 2개만 주석이 된 경우

(3) 파일의 표본 정보와 파일 정보 간 이상 표본 검증

- 파일의 표본 정보와 파일 정보 간 이상 항목이 의심되는 표본을 추출한 결과는 다음과 같았다. 이들은 표본 등록 과정에서 해당 정보의 등록이나 저장이 정상적으로 이루어지지 않은 경우로 파일 정보가 있으나 파일 표본 정보가 없는 경우가 276건, 파일 정보가 없으나 파일 표본 정보가 있는 경우 26건은 해당 표본을 확인하여 수정 조치하였다.

<표 69> 오조작 데이터 검증: 파일의 표본 정보와 파일 정보 간 이상 표본 검증

	파일 표본 정보 있음	파일 표본 정보 없음
파일 정보 있음	-	276
파일 정보 없음	26	-

(4) 표본 정보 중 선행 정보 없이 후행 정보가 있는 표본 검증

- 표본 정보 중 선행 정보 없이 후행 정보가 있는 표본의 수를 추출한 결과는 다음과 같이 네 가지 유형 모두 0으로 이상이 없는 것으로 확인되었다.

<표 70> 오조작 데이터 검증: 표본 정보 중 선행 정보 없이 후행 정보가 있는 표본 검증

등록 연도	×	O	O	O
전사 연도	-	×	×	O
형태주식 작업 연도	-	O	×	×
오류주식 작업 연도	-	×	O	O
결과 표본 수	0	0	0	0

2.2 형태소별 주식의 종합 검수

(1) 1차 작업: 1-2차 연도 기구축 자료

- 상반기에는 1-2차 연도에 구축된 형태 주식 말뭉치의 DB에서 실질어, 어미, 조사, 접사, 기호, NA, NONE 등으로 항목별 주식 불일치 검색하여 엑셀로 출력한 후, 불일치 항목을 유형화하여 수정 작업을 진행하였다. 다음은 주식 유형별 검토 항목의 타입 수이다. 실질어와 어미, 조사는 그 안에서 다시 세부 주식 항목으로 구분되었다.

<표 71> 주식 유형별 검수 항목: 1-2차 연도 기구축 자료 대상

실질어	어미	조사	어근/접사	기호	NA	NONE
22,588	1,375	650	765	1,129	1,966	20

- 검수 자료는 아래와 같이 등급화하였다. 이 중 2, 3단계로 분류된 항목은 2017년 자료 구축 완료 후 맥락을 확인하면서 추가적인 검수 작업을 진행하기로 하여 수정 대상에 제외하고 1단계로 분류된 항목에 대하여 수정 작업을 진행하였다.

<표 72> 검수 작업의 단계화

구분	설명	시스템 반영 여부
1단계	맥락과 무관하게 확실히 오류로 판정된 항목으로	시스템

	[수정 전 형태소]-[수정 후 형태소]가 1:1로 대응되어 시스템 팀에서 기술적으로 수정 가능한 항목	반영 완료
2단계	오류인 것은 확실하나 [수정 전 형태소]-[수정 후 형태소]가 1: 다수, 혹은 다수:1이 되어 기술적인 수정을 하게 될 경우 전체 데이터에 손상이 가해질 우려가 있는 항목	작업자가 맥락을 확인하여 직접 반영 예정
3단계	표본 확인 후 오류 여부를 확정해야 하는 항목	

(2) 2차 작업: 1-3차 연도 전체 자료

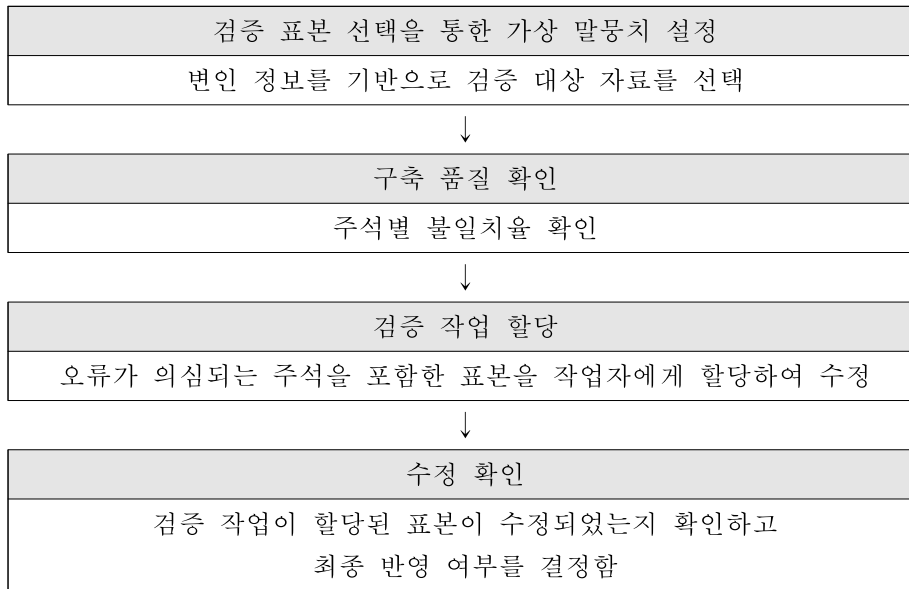
- 하반기에는 2017년 구축 말뭉치의 형태 주석 작업이 완료된 후 위와 같은 방법으로 1-3차 연도에 구축된 전체 말뭉치를 대상으로 하여 형태소별 주석의 종합 검수 작업을 진행하였다. 다음은 주석 유형별 검토 항목의 투입 수로 총 370개(토큰 1,037개)가 우선적인 수정 대상으로 확인되어 수정 조치되었다.

<표 73> 주석 유형별 검수 항목: 1-3차 연도 전체 자료 대상

실질어	어미	조사	어근/접사	기호	NA	NONE
36,577	1,812	804	1,040	3,915	2,549	167

2.3. 시스템 기반의 데이터 신뢰도 검증

- 시스템 기반의 데이터 신뢰도 검증은 학습자 말뭉치 구축 도구 내에 추가된 구축 데이터 품질 확인 기능을 통해 형태 주석 말뭉치와 오류 주석 말뭉치 표본을 추출하여 주석 간 일치율을 통해 오류 가능성이 있는 자료를 추출하여 검수 작업을 진행하는 방식이다. 다음과 같은 과정을 반복적으로 수행하여 데이터의 신뢰도를 높이는 방법으로 반복 횟수가 많아질수록 데이터의 신뢰도가 올라간다.



<그림 18> 시스템 기반의 데이터 신뢰도 검증 절차

- 구축 데이터 품질 확인 기능은 이제 막 개발이 완료된 상태로 안정적인 사용을 위해서는 여러 차례의 시험 사용과 개선이 요구되므로 본 절에서는 시험 사용 예시와 개선 방향을 제시하기로 한다.

(1) 형태 주석 말뭉치의 검증

- 2017년에 구축된 표본 중 구어 형태 주석 말뭉치를 가상 말뭉치로 설정하여 표본 수 219개, 29,091개 문장, 85,921어절, 149,636개의 형태소에 대한 검증 작업을 실시하였다.

구속 대위 품종 확인		구속 품종									
2017		구속년도		구속 품종		구속 수		구속 품종		구속 수	
총 품종 수		4,291		13,032							
합계 품종 수		25,844		188,398							
총 대위 품종 수		147,314		1,244,749							
		품종 수		대위 품종 수							
합계 품종 수		4,291		147,314							
합계 대위 품종 수		3,980		487,777							
교배 품종 수		1,022		150,086							
오류 품종 수		1,022		150,086							
오류 품종 수		1,022		150,086							
오류 품종 수		955		143,849							
		품종 수		대위 품종 수							
연월		10 대위 이하		678							
		20 대		2,837							
		30 대		567							
		40 대		128							
		50 대 이상		79							
		미분류		11							
금수		1 금		1,380							
		2 금		896							
		3 금		587							
		4 금		952							
		5 금		487							

<그림 19> 형태 주식 말뚝치 검증을 위한 가상 말뚝치 설정

○ 그 결과 가상 말뭉치 설정 과정에서 자동으로 제외 처리된 것을 제외한 20,580어절 중 19,184어절의 주석이 일치하고, 나머지 1,396어절에서 0.16%에서 최대 32.77%의 주석이 서로 일치하지 않는 것으로 확인되었다. 가령, 다음 그림의 경우 ‘살’이라는 어절에서 31.56%의 주석이 서로 일치하지 않는 것으로 나타나 추가적인 검수가 필요한 항목이 되었다.

구독 데이터 품질 확인 (월대 주석)					
월대 주석(구어)	연번	매일	총 회수	표본 수	불합치율
성성자	관리자(admin)	불	8	5	30.76 %
성성일	2017.12.10	불	6	6	30.06 %
표본 수	219 건	20.573	하	76	50 30.91 %
문장 수	29,091	20.574	바	7	6 31.24 %
매일 수	86,921	20.575	개	7	5 31.24 %
월대 수	149,636	20.576	차	23	17 31.24 %
		20.577	산	14	13 31.33 %
구속년도	2017 년	20.578	살	43	27 31.56 %
말뭉치 유형	월대 주석	20.579	해요.	121	95 31.68 %
NLI 대	관계	20.580	못했어요.	5	5 32.77 %
급수	관계				
국적	관계				
언어권	관계				
학습 환경	관계				
학습 목적	관계				
자료 유형	관계				
검증 샘플 검색	검증 작업자 선택				
4948	공니탈(interid)				
4949	락유전(silhouette)				
4952	관리자(admin)				
4953	집인리(written...				
4954	집인리(written...				
검증 작업 발발					

연번	원쪽 문맥	중심어	오른쪽 문맥	표본번호
1	그리죠 가까이 애성	살	애성할 것이지.	4397
2	아 열대	살	따부터 마	4997
3	열두	살	태 어머니가 어머니가 마	5110
4	이 그 바아오된 모함은 다섯	살	다섯 살 태	8953
5	이 그 바아오된 모함은 다섯 살 다섯	살	태	8953
6	열곱	살	태 이혼했습니다	8953
7	오아 열 열열열 열 열열열	살	마이는	8954
8	물 스루 머 열열	살	머 이후	8954
9	스름 스름아름	살	(우 문맥 없음)	8954
10	스름 스름아름	살	이후 열분하된 중점어	8954
11	마아인 것은 제가 보아은 나이가 몇	살	쳐어 없으면은 그냥 존재하는 것도 그 경태방을 존재하여 돼	13637
12	머 제가 어미 스름	살	(우 문맥 없음)	14042

<그림 20> 형태 주석 말뭉치 품질 확인

○ 해당 항목을 선택하여 주식 결과를 살펴본 결과 ‘살’은 다음과 같이 6가지 유형으로 주식이 되었으며 각각의 항목 내에서 주식 간 불일치율이 서로 다른 것으로 나타났다. 가령, 7개의 표본에서 12회 출현한 ‘살/NNB’의 경

우 불일치율이 72.09%로 나타났으며, 세부 주식 항목은 이어 제시한 그림과 같다¹⁵⁾.

빈도	원 형태 구조	불일치율	표본 수
15	사/VV+e/ETM	65.12 %	11
12	살/NNB	72.09 %	7
7	살/VV+e/ETM	83.72 %	7
5	살/VV	88.37 %	4
3	살/NA	93.02 %	3
1	사르/NNB	97.67 %	1

<그림 21> 불일치 항목의 형태 확인

연번	왼쪽 문맥	중심어	오른쪽 문맥	표본번호
1	그리고 차이의 여섯	살	여섯살 있습니다.	4957
2	아 열네	살	때부터 어	4997
3	열두	살	때 어머니가 어머니가 마	5110
4	이 그 바이오린 모형을 다섯	살	다섯 살 때	8953
5	이 그 바이오린 모형은 다섯 살 다섯	살	때	8953
6	일곱	살	때 이혼했습니다	8953
7	오미 볼 열여덟 세 열여덟	살	미미는	8954
8	음 스물 여 여덟	살	여 이후	8954
9	스 스물마흔	살	{우 문맥 없음}	8954
10	스물 스물마흔	살	이후 결혼하면 좋겠어	8954
11	마 이런 것은 제가 보면은 나이가 열	살	차이 없으면은 그냥 존경하는 것도 그 상대방을 존경해야 돼	13637
12	여 제가 여 미미 스물	살	{우 문맥 없음}	14042

<그림 22> 불일치 항목의 문맥 확인

- 형태 주식 말뭉치의 경우 학습자의 불규칙한 오류 어절을 제외한 경우 주식 값이 정해져 있기 때문에 위와 같은 검증 방식이 어느 정도 효용성이 있다고 판단된다. 다만, 현재는 중심어를 포함해 해당 주식의 문맥을 확인하는 창에서 중심어의 어절 정보만 제시되고 있어 모든 표본을 작업자에게 할당하여 일일이 주석을 확인해야 하는 불편함이 있다. 따라서 해당 창에서 주식 정보가 제시되도록 수정하면 오류 항목을 해당 창에서 직접 확인할 수 있어 보다 효율적인 검증 작업을 진행할 수 있을 것으로 보인다.

(2) 오류 주식 말뭉치의 검증

- 2017년에 구축된 표본 중 1급의 문어 오류 주식 말뭉치를 가상 말뭉치로 설정하여 표본 수 260개, 3,980개의 문장, 원어절 18,597개, 교정 어절 3,823개에 대한 검증 작업을 실시하였다. 그 결과 검증 대상 중 2,057개 유형의 어절 주석이 일치하고, 나머지 1,820건은 주식 간의 불일치율이

15) 다음 단계는 불일치율을 참고하여 주석의 수정이 필요하다고 판단되는 표본을 작업자에게 할당한 후 수정을 하도록 하고, 수정 사항을 확인하여 반영 여부를 결정하는 것이나 본 절에서는 생략하였다.

4.03%에서 최대 32.98%의 주석이 서로 일치하지 않는 것으로 확인되었다.

2017년 오류 주석-문어-금

작성자: 관리자(admin)
작성일: 2017.12.10

표본 수: 290 건
문장 수: 3,980
어절 수: 18,597
교정 어절 수: 3,823

구속년도: 2017 년
발행처 유형: 오류 주석
나이 대: 전체
급수: 1 급
국적: 전체
연령: 전체
학급: 전체
학수: 전체
자료 유형: 문어

검증 샘플 선택: 58, 60, 61, 63, 64
검증 작업자 선택: 김나연(inter06), 박용진(silhouette), 김나연(written...), 김용은(written...)

검증 작업 할당

연번	어절	총 회수	표본 수	불일치율
1.817	카자흐스탄	1	1	0.00 %
1.818	집에	1	1	0.00 %
1.819	카라테를	1	1	0.00 %
1.820	카메라	1	1	0.00 %
1.821	친구하고	95	73	4.03 %
1.822	기대고	15	5	4.22 %
1.823	우리	46	29	7.96 %
1.824	음식을	10	9	9.00 %
1.825	말이	18	14	9.91 %
1.826	머머니하고	6	5	13.89 %
1.827	한국어가	6	4	13.89 %

연번	원문	원문	오류	오류	오류	오류	표본번호
1	저 중국 친구를 한국에 줄	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
2	저 중국 친구를 한국에 줄	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
3	저 친구를 소개하고 싶으면 명동에 갈	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
4	저 친구를 소개하고 싶으면 명동에 갈	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
5	친구를 바다를 보고 싶으면 부산에 갈	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
6	친구를 바다를 보고 싶으면 부산에 갈	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
7	우리가 부산의 바다에서 수영할	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
8	우리가 부산의 바다에서 수영할	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
9	저 친구는 활동 실을 일해 한국에 줄	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325
10	저 친구는 활동 실을 일해 한국에 줄	기대	(우 문맥 없음)	기대	(우 문맥 없음)	기대	2325

<그림 23> 오류 주석 말뭉치 검증을 위한 가상 말뭉치 설정

- 다음 그림에서 26개의 표본에 33회가 나타난 ‘친구’라는 어절은 교정 형태 구조에서 32.99%의 불일치율을 보이며, 7가지의 교정 형태에서 각각 다른 불일치율을 보이고 있다. 오류 주석의 경우 각 어절에 대한 교정 형태는 문맥에 따라 달라질 수 있고 달라지는 것이 자연스러우므로 이와 같은 교정 형태 구조 자체의 불일치율은 데이터 검증에 그다지 유의미한 정보라고 보기 어렵다. 보다 효율적인 검증을 위해 오류 위치, 오류 양상, 오류 층위의 각 주석별 불일치율을 볼 수 있도록 정보가 추가되어야 할 것으로 판단된다.

연번	어절	총 회수	표본 수	불일치율
2.048	언니	4	4	31.64 %
2.049	일	4	4	31.64 %
2.050	가요.	4	4	31.64 %
2.051	마주	7	7	32.13 %
2.052	싫어요.	7	7	32.13 %
2.053	그래서	20	15	32.16 %
2.054	씨	14	8	32.19 %
2.055	제	12	12	32.36 %
2.056	있습니다.	10	9	32.66 %
2.057	친구	33	26	32.98 %

연번	원문	원문	오류	오류	오류	오류	표본번호
1	그런데 한한	친구	다 머만에 있어서 파티를 할 수 없어요.	친구/NNG	FNP	OM	2354
2	제가	친구	한국에 오면 오리 영동에 갈 거예요.	친구/NNG	FNP	OM	3823
3	제	친구	한국에 가려고 해요.	친구/NNG	FNP	OM	3993
4	지난 여름에 부산에 가서 제	친구	한국에 오면 거기에 또 가고 싶어요.	친구/NNG	FNP	OM	5175
5	매일 수업 후에	친구	와서 같이 밥을 먹습니다.	친구/NNG	FNP	OM	6020
6	저는	친구	뭘 할 겁니다.	친구/NNG	FNP	OM	6024
7	중국	친구	있습니다.	친구/NNG	FNP	OM	6143
8	빌리 씨는 여자	친구	없어요.	친구/NNG	FNP	OM	6143

<그림 24> 불일치 항목의 교정 형태 확인

VI. 결론

1. 연구 요약

이 연구는 한국어 학습자 말뭉치 구축 중장기 사업 계획에 따른 제2단계, 3차 연도 연구이다. 2017년에 새롭게 구축하기 시작한 이주민 자료의 특성에 맞게 학습자 말뭉치 구축 지침을 보완하고 그 지침에 따라 기구축된 115만 어절(문어 100만 어절, 구어 15만 어절) 규모의 말뭉치를 수정·보완함과 동시에 약 66만 어절(문어 50만 어절, 구어 16만 어절) 규모의 말뭉치를 새롭게 구축하는 데에 목적이 있다. 이에 따른 주요 과업과 연구의 성과는 다음과 같다.

○ 한국어 학습자 말뭉치 연구

한국어 학습자 말뭉치 연구는 체계적이고도 활용도 높은 말뭉치 구축을 위한 구축 지침의 정교화와 효율적인 말뭉치 활용을 위한 연구 체계의 수립, 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구로 나뉜다. 2단계는 이주민 자료의 구축이 본격화되는 시기로 이주민 자료의 특성에 따른 주요한 쟁점들을 반영하여 2차 연도에 마련된 수집, 자료 처리, 문어 입력과 구어 전사, 형태 주석, 오류 주석의 각 작업 단계별 지침을 보완하고 정교화하였다. 또한 구축된 말뭉치가 한국어 교육 연구자와 교사, 학습자에게 효율적으로 활용될 수 있도록 다양한 측면에서의 연구 체계를 마련하였다. 연구 체계는 1-3차 연도에 구축된 말뭉치 분석 결과에 따른 기초 통계, 학습자 말뭉치 나눔터를 기반으로 한 자료의 검색 및 활용 모형, 실제적 연구에의 적용 사례로 구성된다. 각 연구 체계는 한국어교육 연구자, 교사, 학습자들이 목적에 따라 참조 자료로 활용하여 다양하게 응용할 수 있을 것이다. 한국어 학습자 말뭉치의 균형성 확보를 위한 (조사) 방법론 연구는 학습자 말뭉치의 특성을 바탕으로 하여 균형성의 개념과 쟁점을 점검해 보고, 학습자 말뭉치의 구성 요소와 직결되는 학습 대상 및 목표, 학습자의 국적 및 제1 언어, 언어 숙달도, 자료의 변인을 살펴보았다. 그리고 지난 3년간의 구축 과정에서 쟁점이 된 사실들을 바탕으로 실제적인 학습자 분포를 고려한 균형성 확보, 다양한 수집 과제를 활용한 균형성 확보, 선 구축, 후 균형, 하위 말뭉치(sub corpus)의 구축을 통한 다양성의 확보라는 방안을 구체화하여 제시하였다.

○ 한국어 학습자 말뭉치 관련 교육 및 홍보

한국어 학습자 말뭉치 관련 교육은 효율적이고 체계적인 말뭉치 구축을 위하여 자료의 수집과 수집 자료의 처리, 문어 입력, 구어 전사, 형태 주석, 오류 주석의 각 단계에 참여하는 인력들을 대상으로 지침 교육과 도구 사용 교육을 중심으로 이루어졌다. 아울러 정기/비정기 워크숍을 통해 작업자 간의 일관성과 통일성을 높임으로써 구축 자료의 신뢰성을 높이하고자 하였다. 한편, 한국어 학습자 말뭉치 홍보는 한국어 학습자 말뭉치 사업을 널리 알리고 사용자들의 활용 능력 제고하는 것을 목표로 한 사용자 워크숍을 통해 이루어졌다. 워크숍은 한국어 교육 연구자와 교사를 대상으로 이루어졌으며 서울에서 2회, 지방에서 1회 개최되었다. 사용자들이 직접 학습자의 구어 자료를 전사해 보고, 학습자 말뭉치 나눔터에 접속하여 자료를 검색해 보는 실습 중심의 워크숍이 이루어졌으며 매회 많은 참가 희망자들의 뜨거운 관심과 호응을 확인할 수 있었다. 또한 워크숍에서의 질의응답 시간을 통해 학습자 말뭉치 구축에 대한 다양한 의견수렴을 할 수 있었다. 그 외에도 학술대회 발표를 통해 한국어 학습자 말뭉치를 연구자들에게 널리 알리고 실제적인 활용 모형을 소개하였다.

○ 한국어 학습자 말뭉치 수집 및 가공

3차 연도의 한국어 학습자 말뭉치 구축은 1-2차 연도 사업에서의 집중 구축 대상인 국내 교육 기관 자료의 균형성을 맞추기 위한 계속 구축과 함께 이주민 자료의 본격 구축이 동시에 진행되었다. 이에 따라 한국어 학습자 말뭉치 수집은 국내 거주 어학 연수생과 유학생, 이주민(결혼이주민, 중도입국청소년, 이주노동자)을 대상으로 하였으며, 전국의 한국어 교육 기관 14개와 다문화가족지원센터, 사회통합프로그램 운영 기관 등 20여 개의 이주민 교육기관의 협조를 얻어 수행되었다. 한국어 학습자 말뭉치 가공은 1-2차 연도에 구축한 원시 말뭉치 115만 어절(문어 100만 어절, 구어 15만 어절), 형태 주석 말뭉치 77만 어절(문어 70만 어절, 구어 7만 어절), 오류 주석 말뭉치 20만 어절(문어 14만 어절, 구어 6만 어절) 규모의 말뭉치를 새로운 지침에 따라 수정·보완하는 작업이 선행되었다. 그리고 원시 말뭉치 66만 어절(문어 50만 어절, 구어 16만 어절), 형태 주석 말뭉치 50만(문어 40만, 구어 10만 어절), 오류 주석 16만 어절(문어 10만, 구어 6만) 규모의 말뭉치가 새롭게 구축되었다. 그 결과 1-3차 연도(2015-2017년)에 구축한 전체 말뭉치의 규모는 원시 말뭉치 181만 어절(문어 150만, 구어 31만 어절), 형태 주석 말뭉치 127만

어절(문어 110만 어절, 구어 17만 어절), 오류 주석 말뭉치 36만 어절(문어 24만 어절, 구어 12만 어절)이 되었다.

○ 구축 말뭉치 검증 정교화

한국어 학습자 말뭉치는 자료의 질적 제고를 위해 각 단계별로 작업자 간의 교차 검수, 상위 작업자의 검수, 작업 단계별 연계를 통한 심층 검수를 순차적으로 거친다. 3차 연도 사업에서는 그 외에 작업 로그와 표본 정보, 주석 정보 등을 기반으로 하여 주석 작업이나 교정 작업 중에 생성된 오조작 데이터를 추출하여 검수하였다. 또한 말뭉치의 검수 체계를 한 단계 더 정교화하기 위하여 개발팀과의 협업을 통해 말뭉치 구축 도구에 통계적으로 데이터의 신뢰도를 검증하는 기능을 추가하였으며, 이를 시험 사용하고 실제적인 적용을 위한 개선안을 개발팀에 제안하였다.

2. 연구의 의의 및 기대 효과

<한국어 학습자 말뭉치>는 외국어 또는 제2 언어로서 한국어를 학습하는 비모어 화자의 말뭉치로서 모어 화자 말뭉치인 <21세기 세종 한국어 균형 말뭉치>와 함께 국가 주도의 말뭉치로서의 위상을 가진다. <한국어 학습자 말뭉치>는 다음과 같은 의의가 있다.

○ 국가 주도의 한국어 학습자 균형 말뭉치 구축

한국어 학습자 언어 연구를 위해서는 학습자가 산출한 자료가 필요하다. 그러나 2002년 문화체육관광부의 주도로 구축된 50만 어절의 말뭉치 외에는 한국어 학습자 말뭉치 구축이 시도된 바 없으며, 문화체육관광부(2002)의 경우 자료 사용에 대한 학습자의 동의 절차를 거치지 못해 배포와 활용이 불가능하여 이러한 문제들을 해결한 일정 규모 이상의 학습자 말뭉치 구축에 대한 사용자들의 요구가 매우 컸다. 이 연구에서는 이전의 말뭉치에서 해결하지 못하였던 자료의 배포와 활용을 고려하여 IRB 규정에 따라 자료 제공 및 사용에 관한 학습자의 동의를 얻어 총 6년간 약 370만 어절 규모의 말뭉치를 구축하는 것을 목표로 하고 있다. 한국어 학습자 말뭉치 구축은 한국어 교육 학계의 오랜 숙원 사업으로 2015년에 시작되어 2017년까지 한국어 교육 기관 및 교사의 적극적인 협조와 지지를 얻어 성공적으로 수행되어 왔다.

○ 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. 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 학술적인 교류가 매우 활발하게 이루어지고 있다. 한국어 학습자 말뭉치는 개인 연구자가 접근하기 어려운 자료를 국가가 주도하여 대규모로 구축하고 공유함으로써 한국어 교육 관련 연구를 활성화할 수 있다는 점에서 큰 의의가 있다. 또한 학습자 말뭉치뿐만 아니라 사용자를 대상으로 한 홍보와 교육, 연구 활용 체계를 마련하여 제공함으로써 그간 학습자 말뭉치 연구에의 요구가 큼에도 불구하고 실질적인 사용 기반이 마련되지 않아 제한적으로 활용되었던 말뭉치 기반 연구와 교육의 확산이 기대된다. 이는 한국어교육 연구자 간의 학술 교류, 더 나아가 세계 규모의 학술대회 대회 참여 등을 통한 국제 수준의 학술 교류를 활성화에도 기여할 수 있다.

○ 한국어교육 이론의 체계화 및 교육 자료 구축의 기반 조성

기존의 언어 교수법은 교수자의 경험과 직관에 상당 부분을 의존함으로써 언어 교육의 효율성 및 교수법의 적합성을 평가하기 어려웠다. 반면, 학습자 말뭉치를 활용한 연구는 외국어로서 혹은 제2 언어로서의 한국어 학습의 측면에서 한국어 사용에 대한 실질적이며 구체적인 설명과 예시를 제시할 수 있기 때문에 이러한 문제를 체계적으로 보완할 수 있게 된다. 학습자 말뭉치를 활용하여 학습자의 모국어, 국적, 숙달도 단계 등의 변인에 따른 언어 사용 양상을 분석하고 이를 기반으로 한 교수·학습 방법 및 교육 자료 개발을 통해 점점 다양해져 가는 학습자의 요구에 부응한 과학적인 교수 환경을 제공할 수 있다. 뿐만 아니라 학습자 말뭉치는 사전뿐만 아니라 학습 교재 편찬, 다양한 교수 이론을 체계화할 수 있으며 그에 필요한 내용 자료를 구축하는 데에까지 광범위하게 활용될 수 있다.

○ [학습] 한국어 학습자의 다양화에 따른 교수·학습 환경의 과학화

세계적으로 말뭉치를 이용한 언어 교육이 활발하게 이루어지고 있으며 이미 상당한 성과를 거두고 있다. 학습자 말뭉치는 학습자의 성별이나 나이, 학습

기간, 모국어, 사용 교재 등과 같은 여러 가지 변인이 언어 학습에 미치는 영향을 비교·분석하는 데에 활용할 수 있다. 그리고 그 결과는 학습자의 수준, 제1 언어, 학습 목적 등의 변인이나 상황에 맞는 맞춤형 교육을 위한 자료로 활용할 수 있다. 그 외에도 또한 각광받고 있는 학습자 중심의 컴퓨터 보조 언어 학습(CALL) 도구 개발을 위한 기초 자료로 활용할 수 있다. 그럼으로써 보다 체계적이고 과학적인 환경에서 보다 수준 높은 학습 환경을 제공할 수 있다.

○ 한국어의 세계화 및 국제 경쟁력 강화

한국어 학습자 맞춤형은 한국어 교육 연구와 교수·학습을 위한 기초 자료로 연구자, 교사, 학습자에 의해 광범위하게 활용될 수 있다. 한국어 학습자 맞춤치를 활용한 연구는 제2 언어, 외국어교육 연구의 주축으로서 국제 수준의 학술 교류 기반을 조성하는 데에 기여할 것이다. 아울러 한국어교육 이론을 체계화하고 교육 자료를 구축함으로써 다변화되어 가고 있는 한국어 교육 환경에 맞는 양질의 교육 서비스를 제공할 수 있게 된다. 그 결과 한국어의 세계화와 함께 한국의 국제 경쟁력 강화에 이바지할 것이다.

3. 보고서 활용 방안

본 연구는 한국어 학습자 맞춤형 구축을 위한 연구와 실제 구축 작업 수행을 주요한 과업으로 하였으며, 본 보고서에는 그러한 과정과 결과, 그와 관련된 쟁점들이 체계적으로 기술되어 있다. 한국어 학습자 맞춤형 구축을 위한 연구의 성과는 한국어교육 연구자를 포함한 예비 사용자들에게 한국어 학습자 맞춤형 구축의 방법과 절차, 각 단계에서의 지침 수립과 관련한 쟁점을 통해 한국어 학습자 맞춤형 구축에 관한 이론과 실제 구축을 위한 가이드라인을 제공해 준다. 또한 학습자 맞춤치의 효율적인 활용을 위한 방안들을 제시해 준다. 본 보고서는 다음과 같이 활용될 수 있다.

○ 한국어 학습자 맞춤형 구축에 관한 이론적 지침

한국어 학습자 맞춤형은 비모어 화자의 자료를 수집하여 구축한 자료로 특수 맞춤치로 분류할 수 있다. 따라서 맞춤치 구축의 일반적인 체계는 참조할 수 있지만 비모어 화자가 산출한 작문이나 발화 자료를 구축 대상으로 하는

만큼 수집, 구축, 가공의 전 단계에서 고려해야 할 사항과 쟁점이 매우 많다. 그러나 일반적인 말뭉치 구축 이론을 소개하는 논저가 비교적 많이 출간되어 있는 것과 달리, 학습자 말뭉치 구축에 관한 실제적인 내용을 다룬 논의는 그리 많지 않으며 그간 한국어 학습자 말뭉치 구축이 활발하게 이루어지지 못하였기 때문에 사실상 구축 방법에 관한 이론이 체계화되었다고 보기는 어렵다. 본 보고서에는 1차 연도부터 축적된 한국어 학습자 말뭉치 구축의 노하우와 쟁점, 그것을 풀어가는 과정이 체계적으로 기술되어 있다. 이는 지금까지 체계화되지 못하였던 한국어 학습자 말뭉치 구축에 관한 이론적 지침으로서 자리매김할 수 있다.

○ 한국어 학습자 말뭉치 구축의 실재를 위한 실용적 지침

본 보고서는 각 단계별 구축 지침을 포함하여 구축 지침에는 한국어 학습자 말뭉치의 수집, 자료 처리, 입력과 전사, 형태 주석, 오류 주석의 방법과 절차가 상세하게 기록되어 있다. 각 단계별 지침은 비모어 화자 자료인 학습자 말뭉치의 여러 가지 특성을 반영한 것으로 학술적으로는 물론 향후 학습자 말뭉치를 구축하고자 하는 기관이나 연구자들에게 실용적인 지침으로 활용될 수 있다. 한편, 한국어 학습자 말뭉치의 주요한 특징 중 하나는 학습자 말뭉치 구축을 위한 표본 등록에서 입력/전사, 형태 주석, 오류 주석의 전 과정을 통합하여 작업할 수 있는 말뭉치 구축 도구를 활용하였다는 것이다. 본 연구에서 사용한 말뭉치 구축 도구는 작업자의 편의성이나 말뭉치 구축과 가공, 자료 관리의 효율성을 고려한 것으로 국외에서도 찾아보기 어려운 첨단 기능들을 포함하고 있으며 매우 선진화되어 있다. 본 보고서에서는 한국어 학습자 말뭉치 구축에 관한 이론적인 지침 외에도 이처럼 말뭉치 구축 도구를 활용한 실제적인 구축 방법과 절차가 충실하게 기술되어 있다. 한국어 학습자 말뭉치 구축 도구는 향후 구축 결과물과 함께 애플리케이션으로 개발하여 배포하여 연구자들이 학습자 말뭉치를 폭넓게 활용할 수 있는 기반을 제공할 예정이다.

4. 정책 제언

○ 하위 말뭉치(sub corpus) 구축을 통한 균형 말뭉치 구축

말뭉치의 필요 요건 중 하나는 대표성과 균형성이다. 본 연구에서는 한국

어 학습자의 말뭉치 균형성 확보를 위하여 한국어 학습자의 변인과 관련 쟁점들을 살펴보고 몇 가지 방안을 제안하였다. 그 중 하나로 하위 말뭉치 구축을 통한 균형 말뭉치 구축 방안을 제안하였는데, 이는 약 300만 어절 규모의 학습자 말뭉치를 효율적으로 활용하기 위한 것이다. 즉, 모든 변인을 수용하여 소규모의 변인별 자료의 총합으로 300만 어절의 말뭉치를 구축하기보다는 실질적인 학습자의 분포를 고려하여 중점 구축 대상을 결정하여 자료를 구축하고, 그 밖의 자료는 말뭉치의 대표성과 다양성 확보 차원에서 일정 규모의 하위 말뭉치로 구축하는 방법이다. 이는 2015년 1차 연도 사업에 착수하면서 본 사업에 이르기까지 수차례의 자문 의견을 통해서도 타당성이 확인된 바 있다. 이에 대한 합의가 이루어진다면 2020년까지의 최종 구축 규모는 300만 어절+a로 당초 계획보다 확장될 필요가 있으며, 구체적인 구축 계획에 대한 사전 논의가 필요하다.

○ 이주민 자료 및 국외 자료 수집 네트워크 마련을 위한 공조 체계 강화

한국어 학습자 말뭉치 구축 중장기 계획에 따르면 2단계(2017-2018년)는 이주민 자료, 3단계(2019-2020년)는 국외 학습자 자료의 집중 수집과 구축이 이루어지는 시기이다. 이에 따라 2017년 3차 연도 사업은 결혼이주여성, 이주노동자, 중도입국청소년의 세 집단으로 이루어진 이주민 자료의 수집과 가공을 본격적으로 시작하였다. 수집 네트워크 마련을 위해 국립국어원과 여성가족부 건강가정지원센터(다문화가족지원센터), 법무부(사회통합 프로그램 운영기관), 교육부(각 급 학교) 등과의 업무 협의, 공문 발송 절차를 거쳐 실질적인 수집을 위한 섭외와 교육을 실시하였다. 그러나 1단계(2015-2016년) 연구에서 자료 수집을 위해 학계의 공통적인 관심과 요구에 의한 공조 체제를 이끌어낼 수 있었던 것과 달리, 이들 기관의 경우 자료 수집의 필요성에 대한 공감대를 이끌어 내기가 쉽지 않았고 자료 수집 협조를 요청할 수는 있으나 강제할 수 없어 실질적인 수집으로 이어지는 기관이 매우 적었다. 이와 같은 문제에 대한 대안으로 이주민 관련 기관에서 활동하는 교사들을 공개 모집하여 교사 인력풀을 통한 수집을 동시에 시도하기도 하였는데, 개개인으로서 이루어진 네트워크이기 때문에 관리의 어려움이 따랐다. 그러나 무엇보다도 큰 어려움은 학습자들의 신변에 변동이 잦아 지속적인 수집이 어렵다는 점이었으며, 학습자의 변인이나 교육 환경 등이 매우 다양하여 자료가 균질하지 못하다는 문제가 그러한 어려움을 가중시켰다. 이러한 문제는 향후 직접 접촉이 쉽지 않고 다양한 환경에서 교육이 이루어지는 국외 자료 수집에서는 더

육 심화될 것으로 예측된다. 이에 따라 교육 현장에 있는 교사 커뮤니티와의 연계, 공식적인 교사 연수와의 연계를 통해 학습자 말뭉치에 대한 기초 이해 교육과 함께 체계적인 수집 교육을 실시하고, 공식적이고 지속적인 네트워크를 이어갈 필요가 있다. 이를 위해 국립국어원이 주관하는 국내외 교사 연수회를 적극적으로 이용하는 등 국립국어원과 구축 사업자 간의 긴밀한 공조 체계가 필요하다.

○ 사용자의 요구를 반영한 구어 말뭉치, 오류 주석 말뭉치 비율 확대

한국어 학습자 말뭉치는 현재 원시 말뭉치 181만 어절(문어 150만 어절, 구어 31만 어절), 형태 주석 말뭉치 127만 어절(문어 110만 어절, 구어 17만 어절), 오류 주석 말뭉치 36만 어절(문어 24만 어절, 구어 12만 어절)로 원시 말뭉치의 약 70%가 형태 주석 말뭉치로 구축되었으며, 약 20%가 오류 주석 말뭉치로 구축되었다. 한편, 구어 말뭉치의 비율은 원시 말뭉치 17.5%, 형태 주석 말뭉치 13.3%, 오류 주석 말뭉치 33.3%로 문어 말뭉치에 비해 매우 적다.

그 결과 현재 시범 서비스 중인 한국어 학습자 말뭉치 나눔터의 오류 주석 말뭉치와 구어 말뭉치의 검색 시 실제로는 학습자들의 오류가 빈번한 항목인데도 검색 결과가 나오지 않는다는 문제가 발생하고 있다. 이는 사용자의 불편으로 이어지며 학습자 말뭉치에 대한 사용자들의 관심과 활용 빈도를 저하시키는 직접적인 원인이 된다. 한국어교육 연구 동향이나 학습자 말뭉치 아카데미를 통한 사용자들의 요구를 볼 때 장기적으로 오류 주석 말뭉치와 구어 말뭉치 구축 비율의 확대는 반드시 필요하다.

○ 한국어 학습자 말뭉치 나눔터의 효율적 사용을 위한 검색 속도

개선 및 검색 기능 정교화

한국어 학습자 말뭉치 아카데미를 통해 한국어 학습자 말뭉치 나눔터가 공개되면서 학습자 말뭉치에 대한 다양한 의견이 수렴되고 있다. 현재 시범 서비스 중인 한국어 학습자 말뭉치 나눔터는 지속적인 업데이트를 통해 다양한 조건에 따른 자료의 검색 정보를 제공하는 기능이 잘 구현되어 있다. 여기에 사용자의 사용 목적과 요구를 고려하여 다음과 같은 기능을 추가함으로써 활용도를 높일 수 있을 것이다.

검색 속도 개선: 보다 효율적인 사용을 위해 국가 정보화 보안 방침 및 하드웨어 구성을 고려한 검색 속도의 개선이 필요하다.

통계 정보의 정교화: 검색 자료에 대한 통계 정보의 정교화가 필요하다. 특히, 학습자의 중간언어 특성을 파악하기 위해서는 오류 주석 말뭉치에서 특정 문법 형태소를 검색하였을 때 검색 대상이 된 말뭉치 대비 오류형과 비 오류형의 비율이 검색 결과와 함께 제시되어야 한다.

○ 지속적인 자료의 업데이트 및 다양화를 위한 온라인 수집 시스템 운영

한국어 학습자 말뭉치는 그간 연구자, 교사, 학습자 등 다양한 목적의 사용자들의 요구가 컸음에도 불구하고 원천 자료에의 접근이 쉽지 않고 구축과 가공에 많은 경험과 시간, 노력이 필요함으로 해서 실제 구축으로 이어지지 못한 측면이 있다. 그런 만큼 본 연구에 대한 학계와 사용자들의 기대와 호응이 매우 크며 이는 본 연구에서 2017년에 개최한 세 차례의 워크숍을 통해서도 확인된 바 있다. 따라서 단발적인 연구 용역에 그쳐서는 안 되며 학습자 집단의 변화나 사용자의 요구에 따라 자료가 지속적으로 업데이트될 수 있는 수집 체계가 마련되어야 한다. 이에 본 연구에서는 온라인 수집 시스템의 개발과 운영을 제안하는 바이다. 온라인 수집 시스템은 2차 연도 최종 보고서를 통해 제안한 바와 같이 두 가지 장점이 있다.

첫째, 지속적인 온라인 수집 시스템을 구축함으로써 수집 자료의 다양성과 양적 규모를 추가적으로 확보할 수 있다. 한국어 교육 환경은 다양한 국적, 학습 목적 등의 학습자 변인에 따라 지속적으로 다변화되고 있다. 그리고 그러한 흐름에 따라 학습자들의 요구가 달라지며, 그러한 요구에 부응하여 교육과정, 교육 자료, 교수 방법 또한 달라져야 한다. 더불어 학습자들이 산출하는 자료의 내용이나 특성도 달라진다. 따라서 학습자 말뭉치는 이러한 변화를 반영할 수 있도록 6개년의 중장기 계획이 마무리된 후에도 지속적인 자료 수집이 이루어질 수 있는 환경을 마련할 필요가 있다.

둘째, 온라인 수집 방법은 자료 제공자인 학습자의 동의를 전제로 하므로 간단한 방법과 절차를 통해 IRB 규정에 따른 자료 수집의 제한점을 보다 쉽게 해결할 수 있는 장점이 있다. 다만, 시스템의 실제 운영을 위해서는 시스템 개발과 함께 학습자들의 자료 제공을 독려하기 위한 보상 방안의 마련이 필요하며, 이를 위해 행정적 지원과 예산 배정이 수반되어야 할 것이다.

○ 균형 말뭉치 구축을 위한 추가 예산 필요

말뭉치 구축 작업은 노동집약적인 작업이기 때문에 많은 인력과 시간, 비용과 노력이 소요된다. 학습자 말뭉치의 경우는 한국어를 비모어로 하는 화

자들의 발화 산출 자료를 기반으로 한 특수 말뭉치로 그것이 배가된다. 매해 일정 규모의 말뭉치 구축을 위해 수집 단계에서 1,000여 명, 구축과 가공 단계에서 50여 명의 인력이 투입되고 있다. 향후 구축 말뭉치의 균형성을 위해 이주민 자료와 국외 학습자 자료를 구축하고, 오류 주석 말뭉치와 구어 말뭉치의 구축 비율을 확대하게 된다. 이때에는 반드시 구축의 각 단계에서 필요 인력 구성을 위한 비용이 뒷받침되어야 할 것이다.

부록. 2017년 한국어 학습자
말뭉치 구축 지침

차 례

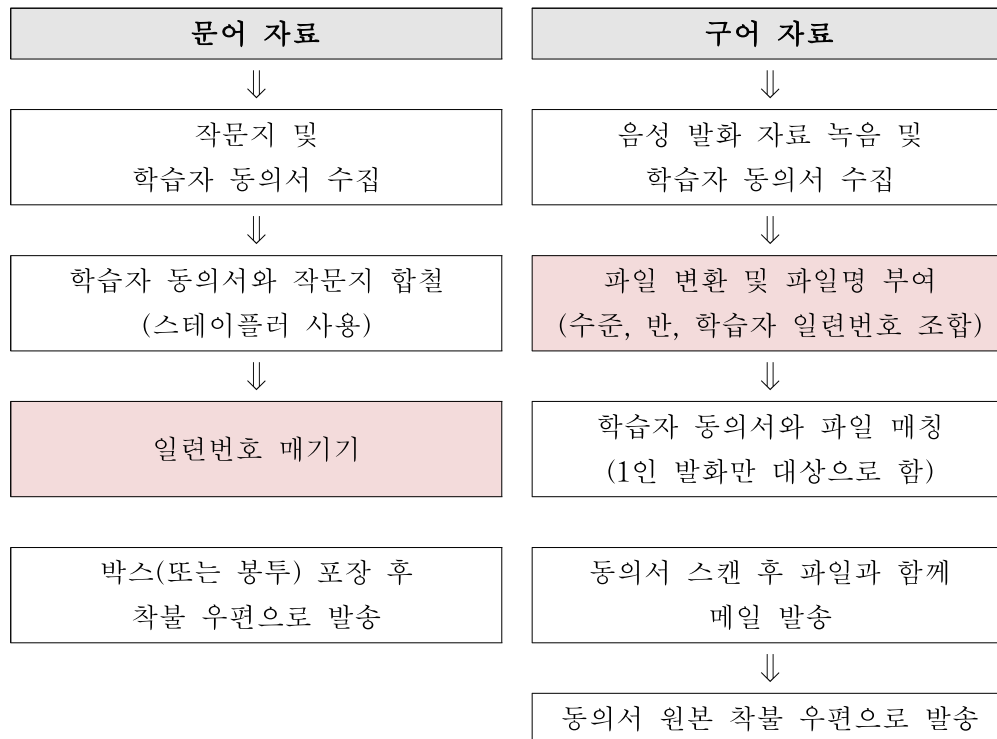
한국어 학습자 말뭉치 수집 지침	1
한국어 학습자 말뭉치 자료 처리 지침	40
한국어 학습자 말뭉치 문어 입력 지침	44
한국어 학습자 말뭉치 구어 전사 지침	51
한국어 학습자 말뭉치 형태 주석 지침	72
한국어 학습자 말뭉치 오류 주석 지침	134

한국어 학습자 말뭉치 자료 수집 지침

1. 자료 수집 대상 및 수집 자료

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

2. 자료 수집 절차



1) 문어

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

한국어교육과정연구위원회(한글)

0001

建立韓國語學習者語料庫的個人資料使用同意書

韓國國立公設局為了精進語言習得的提高，正在蒐集語學資料（語料庫）以便促進本學系的教育。【本系執行：基於大學平權合作協定】，大家期望的預知必須有科學語法教學方法的改善，和國語教材的開發。韓國語教學領域及知識進步，參加本研究的所有人員皆須知悉本人尊重危險，和其國立公立以上包括中等學校等均可任意利用同意書。此外，蒐集之個人資料中除教學事業以外無其他目的。並且，為了保護信息，所有集得之個人信息是他人不能識別的狀態下進行使用，希望！

申請校： 國立大學 戶字合作校 02-222-4159

● 請將欄位併使用以下信息至2015年度/秋季學業考試止。金銀資料。

日期 7/9
姓名 SAH JUNG
母姓 MU BAEU

(印)

下面是研究時所应用的个人资料，此處選擇保證無下列事項。

- 性別：☐男 ☒女
- 年齡： 29
- 現在居住所在地： _____
- 國籍： Japanese () 中國人 () 外國人 ()
- 母語： Chinese
- 韓國語的學習時間： 0 年 2 月
- 在韓國的居住時間： b 年 2 月
- 學習韓國語的目的：
☐ 旅行 ☒ 就學 ☐ 工作 ☐ 結婚 ☐ 其他 ()
- 職業： Student
- 新到韓國以內可以看見的第四語言(按順序排列)：Chinese English Japanese Korean

001

부속 151 부활에 무엇을 했습니까? 어디에 갔습니까? 누구를 만났습니까? 무엇을 먹었습니까? 여겼습니까?
부활이요. 아예의 요행을 모두 사해에서 하시오. (150~300대) [30점]

월	년	세	일	시	분	초	지역

주말에 한국어로 공부했습니다. 토요일에 친구를 만나서 배웠습니다. 어머니를 데리고 가서 판매했습니다. 저는 백화점에서 소스 빼먹기 불공근하고 먹었습니다. 저는 시장에서 꽃 하나나 시와 기원했습니다. 모토 / 오오 원입니다.

일요일에 저는 극장에서 영화를 봤습니다.

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

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

② 성취도 평가 수집

- 각 교육기관의 성취도 평가(중간 및 기말) 쓰기 시험에 포함된 작문을 활용하여 이를 수집함. 글의 종류 및 주제는 각 기관의 성취도 평가에 따름

③ 교육과정 외 프로젝트를 위한 기획 작문 수집(수집 가능 기관)

- 각 등급에 맞추어 수업 시간(1시간) 내에 다음과 같은 글의 종류와 주제로 작문을 하게 하여 이를 수집함. 세부 주제는 종적 말뭉치의 과제 활동

자료를 참고함

수준	추천 글의 종류	기타	주제
초급	체험적 글(생활문)	일기, 편지, 이메일 등	소개(자신, 가족 등), 취미, 한국생활, 주말, 계절, 좋아하는 음식, 학교생활, 여행, 일상사 등
중급	체험적 글(생활문) 설명적 글(설명문)	안내문, 감상문 등	소개(가족, 문화, 풍습 등), 취미, 여행, 여가생활, 한국생활, 추억, 영화, 만남, 직업, 후회, 사회문제(환경문제 등), 등
고급	설명적 글(설명문) 논리적 글(논설문)	기사문, 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 학습자 간의 2인 발화의 경우에는 각 학습자가 녹음하여 이를 교사에게 전송

하게 하여 이를 수집함

- 토론 등 다인 발화의 경우에는 발화자의 정보를 확인 가능하도록 비디오로 녹화하거나 녹음 및 전사자를 일치시킬 것을 권유함
- 이의 담화 유형과 주제, 시간은 각 기관의 교육과정에 따름

② 성취도 평가 수집

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

③ 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"> ○ 가장 아끼는 물건은 무엇인가? ○ 왜 그 물건을 아끼는가? ○ 어떻게 그 물건을 가지게 되었는가? 	
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"> ○ 취미로 배우고 싶은 것은 무엇인가? <p>(※ 한국어를 배우고 싶다는 내용은 쓰지 마십시오.)</p> <ul style="list-style-type: none"> ○ 왜 그것을 배우고 싶은가? ○ 그것을 배운 후에 무엇을 하고 싶은가? 	
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"> ○ 나의 나쁜 생활 습관 ○ 습관 때문에 생기는 불편하거나 안 좋은 점 ○ 습관을 고치기 위해 해야 할 일 	
30주차	<p>잊지 못할 추억’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 어떤 추억인가요? ○ 왜 지금까지 기억에 남아 있는가? ○ 언제 그 추억이 떠오르는가? 	
32주차	<p>갖고 싶은 직업’이라는 제목으로 글을 써 보십시오. 단 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 직업명, 하는 일, 그 일을 하려는 이유, 그 일에 필요한 조건 	
34주차	<p>나의 성격’이라는 제목으로 글을 써 보십시오. 단 아래에 제시된 내용에 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 성격의 특징, 장점과 단점, 고치고 싶은 부분과 그 이유 	
36주차	<p>내가 생각하는 행복’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 행복을 위해서 어떤 노력을 하는가? 	

수집 시기	문제	수준
	<ul style="list-style-type: none"> ○ 언제 행복하다고 느끼는가? ○ 행복은 무엇이라고 생각하는가? 	
38주차	<p>‘내가 좋아하는 책’이라는 제목으로 글을 쓰십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 가장 좋아하는 책은 무엇인가? ○ 그 책은 어떤 내용인가? ○ 그 책을 좋아하는 이유는 무엇인가? 	
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"> ○ 닮고 싶은 사람은 누구인가? ○ 왜 그 사람처럼 되고 싶은가? ○ 그 사람처럼 되기 위해서 어떻게 해야 하는가? 	
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"> ○ 체질에 맞는 음식 ○ 수면의 질 ○ 적당한 운동 ○ 긍정적인 사고 	
44주차	<p>올바른 인터넷 사용 태도’에 대한 자신의 견해를 서술하십시오. 단, 아래 제시한 <올바른 인터넷 사용 태도의 예> 중에서 세 가지를 선택하여 쓰되, 각각의 태도를 지키지 않았을 경우에 나타나는 부작용의 예를 포함해야 합니다.</p> <p><올바른 인터넷 사용 태도의 예></p> <ul style="list-style-type: none"> ○ 상대방의 의견 존중하기 ○ 타인의 사생활 보호하기 ○ 의견 차이 인정하기 ○ 바른 언어 사용하기 ○ 정확한 정보 올리기 	고급
46주차	<p>다음 글을 읽고, ‘현대 사회에서 바람직한 신문의 기능’에 대한 자신의 견해를 서술하십시오. 단 아래에 제시한 기능 중에서 두 가지 이상을 선택하여 쓰되, 그 기능이 현대 사회에 중요하다고 생각하는 이유를 포함해야 합니다.</p> <p><신문의 기능></p>	

수집 시기	문제	수준				
	<ul style="list-style-type: none"> ○ 사건 보도 ○ 여론 조성 ○ 정보 제공 ○ 소통의 분위기 조성 					
48주차	<p>다음 글을 읽고 '감시 카메라 설치 확대'에 대한 자신의 견해를 서술하십시오. (찬성하거나 반대하는 입장에서 하나를 선택하여 서술 할 것. 단 아래 제시된 각 입장의 논거 중 두 개 이상을 제시할 것.)</p> <div 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 사상·행위 침해 범죄 예방 효과 불분명 가해자의 인권 보호 </td> </tr> </table> </div>	찬성	반대	사회 안전 유지 범죄 예방 인권보다 공인이 우선	개인의 사상·행위 침해 범죄 예방 효과 불분명 가해자의 인권 보호	
찬성	반대					
사회 안전 유지 범죄 예방 인권보다 공인이 우선	개인의 사상·행위 침해 범죄 예방 효과 불분명 가해자의 인권 보호					
50주차	<p>여러분은 성공이 무엇이라고 생각하십니까? 그리고 그러한 성공을 이루기 위해 필요한 것이 무엇이라고 생각하십니까? 이와 관련된 자신의 견해를 서술하십시오. 단, 아래에 제시한 내용이 모두 포함되어야 합니다.</p> <ul style="list-style-type: none"> ○ 내가 생각하는 성공이란 무엇인가? ○ 그것을 이루기 위해 필요한 것은 무엇인가? ○ 그 이유는 무엇인가? 					
52주차	<p>여러분은 무엇이 선의의 거짓말이라고 생각하십니까? 어떤 경우에 그런 거짓말을 할 수 있다고 생각하십니까? 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p>< 선의의 거짓말이란 ></p> <ul style="list-style-type: none"> ○ 선의의 거짓말이란 무엇인가? ○ 선의의 거짓말은 언제 필요한가? 					

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>< 예술 교육의 필요성 ></p> <p>○ 예술 교육이 왜 필요한가?</p> <p>○ 예술 교육을 통해 얻을 수 있는 효과는 무엇인가?</p>	
56주차	<p>자연을 그대로 보존해야 한다는 주장과 인간을 위해 자연을 개발해야 한다는 주장이 있습니다. 이에 대한 자신의 견해를 서술하십시오. 단, 아래에 제시된 내용이 모두 포함되어야 합니다.</p> <p><자연 보존과 자연 개발></p> <p>○ 자연 보존과 자연 개발 중 어느 것이 더 중요하다고 생각하는가?</p> <p>○ 그렇게 생각하는 이유는 무엇인가? (2가지 이상 쓰시오.)</p>	
58주차	<div 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>에 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.</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> <ul style="list-style-type: none"> ○ 현대 사회에서 필요한 인재는 어떤 사람입니까? ○ 이러한 인재가 되기 위해서 어떤 노력이 필요합니까? 	

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

종적 자료 수집 시기	문제	수준
	○ 돈을 모으면 무엇을 하고 싶은가?	
46주차	<p>‘텔레비전이 우리 생활에 미치는 영향’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <p>○ 텔레비전을 자주 보는가? 왜 그런가?</p> <p>○ 무슨 프로그램을 자주 보는가? 왜 그런가?</p> <p>○ 텔레비전이 우리 생활에 미치는 긍정적인 영향은 무엇인가? 부정적인 영향은 무엇인가?</p>	
48주차	<p>‘효과적인 자녀 교육법’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <p>○ 현재 자녀가 있는가?</p> <p>○ 자녀가 말을 듣지 않을 때 어떻게 하는가? (현재 자녀가 없는 경우, 부모의 말을 듣지 않는 아이를 어떻게 하면 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>< 선의의 거짓말이란 ></p> <p>○ 선의의 거짓말이란 무엇인가?</p> <p>○ 선의의 거짓말은 언제 필요한가?</p> <p>○ 선의의 거짓말이 가질 수 있는 문제점은 무엇인가?</p>	
54주차	<p>‘노후 준비’라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p>	

종적 자료 수집 시기	문제	수준
	<ul style="list-style-type: none"> ○ 노후 준비가 왜 필요한가? ○ 노후를 위해 무엇을 준비해야 하는가? ○ 여러분은 노후를 어떻게 준비하고 있는가? 	
56주차	<p>‘칭찬은 고래도 춤추게 한다’는 말처럼 칭찬에는 강한 힘이 있습니다. 그러나 칭찬이 항상 긍정적인 영향을 주는 것은 아닙니다. 아래의 내용을 중심으로 칭찬에 대한 자신의 생각을 쓰십시오.</p> <ul style="list-style-type: none"> ○ 칭찬이 미치는 긍정적인 영향은 무엇입니까? ○ 부정적인 영향은 무엇입니까? ○ 효과적인 칭찬의 방법은 무엇입니까? 	
58주차	<p>‘여성의 사회적 지위와 역할’이라는 제목으로 글을 쓰십시오. 단, 다음의 내용이 모두 포함되어 있어야 합니다.</p> <ul style="list-style-type: none"> ○ 여성이 일을 해야 한다고 생각하는가? 왜 그런가? ○ 한국에서 여성의 사회적 지위는 어떻다고 생각하는가? 고향과 비교해서 높은 편인가? 낮은 편인가? ○ 여성이어서 좋은 점 혹은 좋지 않은 점이 있다고 생각하는가? 	
60주차	<div style="border: 1px solid black; padding: 10px;"> <p>오늘날 직업에 대한 생각은 크게 두 가지로 나뉘는 것 같다. 하나는 여러 방면으로 사회에 도움을 주거나, 공헌할 수 있는 직업을 택해 봉사하는 마음으로 일하고, 그것을 통해 얻어지는 대가로 자신과 가정을 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	문제	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 뭘까요? 부정적인 영향은 뭘까요? ‘텔레비전’이라는 제목으로 글을 쓰세요.	
48주차	여러분은 거짓말을 한 적이 있어요? 무슨 거짓말을 했어요? 거짓말을 한 후 어떤 일이 일어났어요? ‘거짓말’이라는 제목으로 글을 쓰세요.	
50주차	여러분은 언제 행복해요? 그리고 언제 슬퍼요? ‘행복한 일과 슬픈 일’이라는 제목으로 글을 쓰세요.	
52주차	학교에서 일어난 큰 실수나 사고를 생각해 보세요. 기억에 남는 일을 쓰세요.	
54주차	여러분은 무슨 놀이(게임, 수업 활동 등)를 좋아해요? 어떻게 해요? 여러분들이 좋아하는 놀이 방법을 소개하는 글을 쓰세요.	
56주차	선생님과 부모님께 칭찬을 받아 본 적이 있지요? 언제 칭찬을 받았어요? 무슨 일로 칭찬을 받았어요? 기분이 어땠어요? 칭찬 받은 일에 대해서 글을 써 보세요.	
58주차	지금까지 읽은 책 중에 가장 재미있었던 책이 뭐예요? 무슨 내용이에요? 읽고 무슨 생각을 했어요? ‘내가 가장 좋아하는 책’이라는 제목으로 글을 써 보세요.	
60주차	그림을 보고 이야기를 순서대로 써 보세요.	

2. 구어

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

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

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

<2주차 읽기 자료>¹⁶⁾

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

<12주차 읽기 자료>

한국에는 봄, 여름, 가을, 겨울 사계절이 있습니다.

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

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

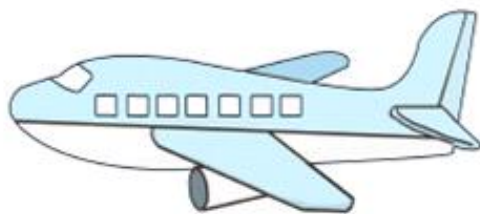
가을에는 날씨가 선선해집니다. 산에 가서 단풍 구경을 합니다. 빨간 단풍이 참 아름답습니다. 맛있는 과일도 많이 먹을 수 있습니다.

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

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

16) 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>¹⁷⁾



17) 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>¹⁸⁾

※ 다음 그림을 보고 이야기를 만들어 보세요.

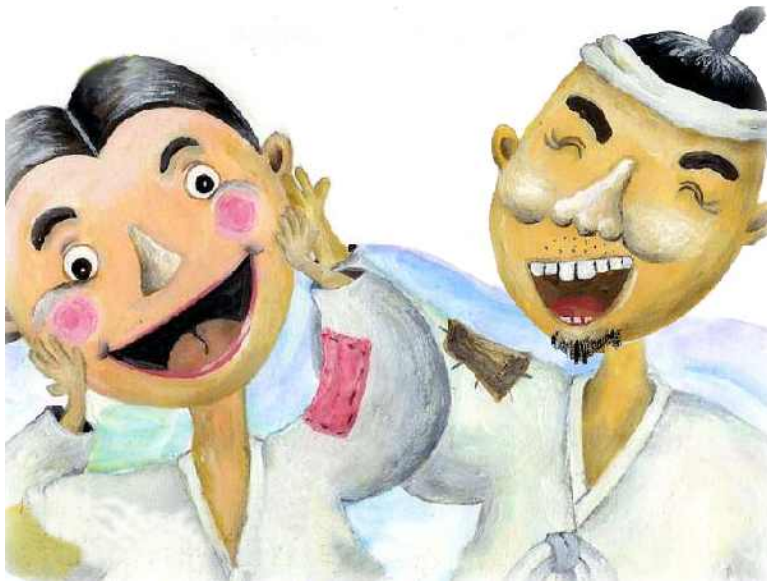


1



2

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



3



4



5



6

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, 비밀유지를 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

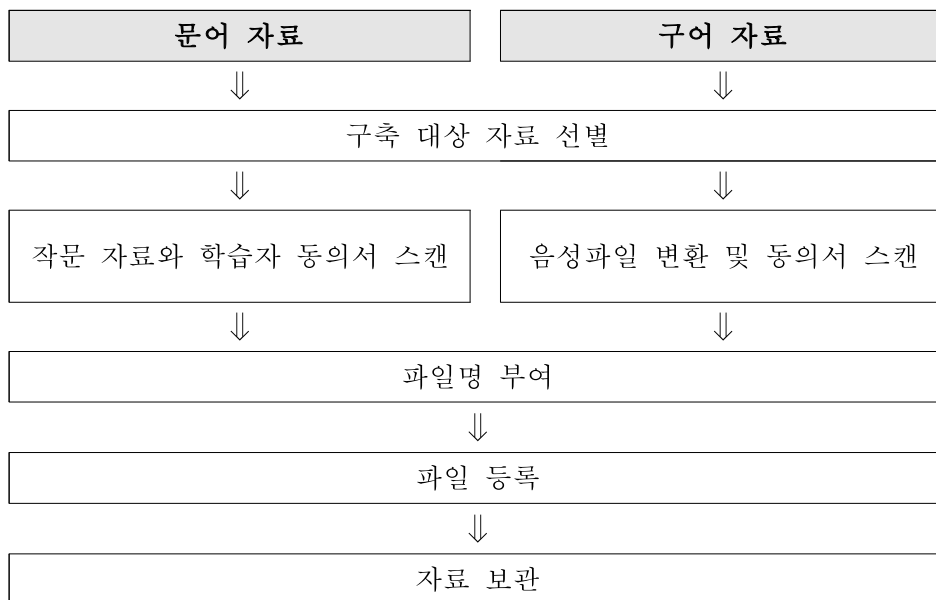
15. 한국어로 나오는 방송 매체(TV, 라디오, 인터넷 동영상)를 보는 시간은 얼마나 됩니까?

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

2) 학습자 동의서 확인 및 스캔

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

집수 날짜	기관명	자료 유형	자료 내용	자료 수(수준별)							합계	자료 관련 메모
				1급	2급	3급	4급	5급	6급	6급 이상		
2015.07.14	동국대학교(경주)	일반	여름 학기 중간고사 쓰기 자료								105	
2015.07.21	한남대학교	일반	여름 학기 중간고사 쓰기 자료	13	12	23	8				56	
2015.07.22	한양대학교	일반	여름 학기 중간고사 쓰기 자료	39	71	78	70	43	9		310	
2015.07.23	호남대학교	일반	여름 학기 기말고사 쓰기 자료								104	
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.

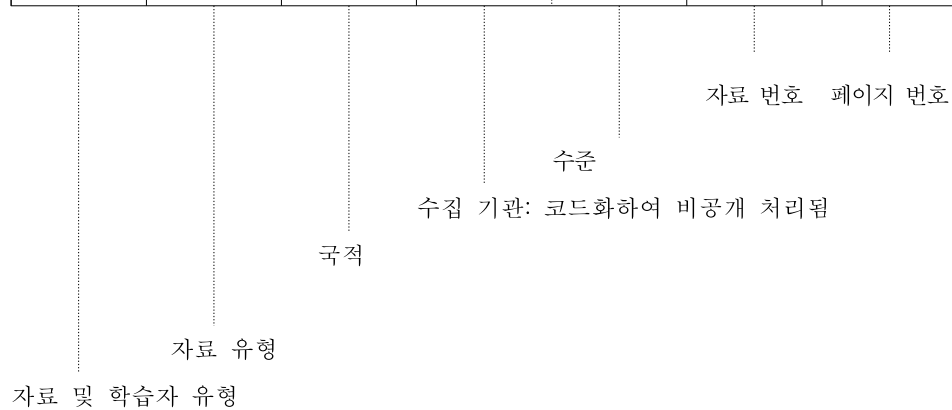
4) 파일명 부여

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

예) 종적_문어_중국_서울대_1급_0001_01.txt

자료 코드	학습자 변인 정보 코드
-------	--------------

종적	문어	중국	서울대	1급	0001	01
----	----	----	-----	----	------	----



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)	국내 일반 국내 종적 결혼이주 일반 결혼이주 종적 중도입국 일반 중도입국 종적 이주노동 일반 이주노동 종적
	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	동일한 학습자가 두 개 이상의 자료를 제공할 경우 자료를 구분하기 위한 일련번호	01 02 ...	01 02 ...

한국어 학습자 말뭉치 문어 입력 지침

1. 전체적인 형식 원칙

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

2. 입력 지침

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

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

- 원문의 영어와 한자는 모두 유지한다. 한자는 시스템 입력창에서 글자를 선택 후 마우스 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 ^(BARFLY) ‘바프라이’ 술집에 ~~가요~~ 가요.

- 숫자와 한글 표기를 함께 쓴 경우 원문에 따라 병기한다.

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

^(세) 3달전에 미국에서 한국까지

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

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

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

3. 문장 부호 및 기호류 마크업

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

<예> <EX_Alpha>¿¿¿¿¿</EX_Alpha>

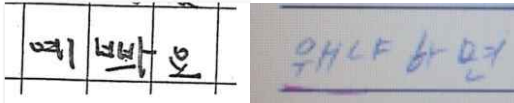
- ‘식별불가’는 원본에서 다양한 이유로 확인이 어려운 문자나 기호에 대해 ¿로 입력 후 마크업한다.

<예> <CNI>¿¿</CNI>

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

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

- ‘한글기호’는 두벌식 한글과 같이 자판에서 하나의 음절로 입력이 불가능한, 우리말에 없는 글자를 입력할 때 사용하는 마크업 기호이다.

<예> 

1) 예ㅼㅅ요 : 시스템 '예ㅼㅅ요'로 입력 후 ‘한글기호’ 마크업 처리
 - 예<NSS>ㅼㅅ</NSS>요

2) 우냐하면 : 시스템 '우냐하면'으로 입력 후 ‘한글기호’ 마크업 처리
 - <NSS>우</NSS>냐하면

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

<예> <MP> • </MP>

4. 익명성 보장을 위한 개인 정보의 처리

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

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

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

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

5. 기타

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

6. 최고급 자료 마크업

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

	주석	내용	주석 표시
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	<q>

	주석	내용	주석 표시
		제시된 인용 구절과 예시문	
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. 1786년에 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.

<p><예> <body> <title>1. 서론 </title> <sentence>본 연구의 목적은...</p>

○ 문장 중간에 나타나는 계산식이나 그림은 작업자 메모를 달아 준다.

한국어 학습자 말뭉치 구어 전사 지침

I. 구어 전사 기호 체계

대분류	소분류	기호	예시
발화자 정보	화자 1의 표시	1	<person id=1 sex=M age=20s> 1:
	화자가 불분명할 때	?	P?:
	동시 발화	모두/ 나머지/ 2,3	2,3:네.
억양 단위	하강	.	2:네.
	상승	?	2:어디 갈 거예요?
	약한 상승이나 하강	,	1:그래서 그랬는데 이번엔,
	활기, 기운찬 어조	!	1:아!
	억양 단위 경계의 처리	엘란에서 발화 단위를 분할하면 줄 바꿈 처리됨	2:어디 갈 거예요? 1:안 아직까지 그냥 계획만 잡아 났는데, 2:음.
	하나의 억양 단위가 끼어들어 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?

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

대분류	소분류	기호	예시
	축약형	'(apostrophe, 영문따옴표)를 사용해서 두 음소를 연결	사귀'어, 바뀌'어, ...
	표현적 장음	::	1:많은 경우에:: 논문, 저::~ 어::~ 연구는 네이션,
	담화표지	~	1:많은 경우에 논문, 그::~ 어::~ 연구는 네이션,
준음성	웃음	<vocal desc='웃음'>	6:어우 야 <vocal desc='웃음'>
	기침	<vocal desc='기침'>	-
	하품	<vocal desc='하품'>	-
	재채기	<vocal desc='재채기'>	-
	목청 가다듬는 소리(음, 으음)	<vocal desc='목청가다듬는소리'>	-
	들이마시는 숨(쓰)	<vocal desc='들이마시는숨'>	-
	내쉬는 숨(후우)	<vocal desc='내쉬는숨'>	-
	혀 차는 소리(쯔)	<vocal desc='혀차는소리'>	-
	헛기침(에 험)	<vocal desc='헛기침'>	-
	한숨	<vocal desc='한숨'>	-
	노래	<vocal desc='노래'>	-
	웃으면서 말하는 부분	<@ @>	2:[<vocal desc='웃음'>]

대분류	소분류	기호	예시
			너무 줌 <@오버한다.@>
	박수 치면서 말하는 부분	<# #>	5:[우우 <vocal desc='박수'>] <#이리 와 이리 와.#>
	노래를 부르는 부분	<M M>	-
	박수나 손가락 부딪치는 소리	<kinesics desc=' '>	박수 <kinesics desc='박수'>
	대화 흐름에 영향을 주는 전화벨 소리라든지 기타 음성 아닌 소리	<event desc=' '>	<event desc='전화벨소리'>
2차 전사	구어적 변이형	()	1:브릿지를 한 가닥을 넣어(너) 줬어요.
	발음 오류	()	친구(칭구)와 강남(간남)에 갔습니다(갔슨니다). 같이(가티) 가자. 과반수(화반수)
		()	가구(가구)<note>'가'의 ㄱ 을 유성음으로 발음 </note> 가구(가구)<note>'구'의 ㄱ

대분류	소분류		기호	예시
		변이음의 구분이 모호한 경우		을 무성음으로 발음 </note>
		음운규칙으로 인해 한국어의 철자대로 발음되지 않으나, 학습자가 이를 철자대로 발음하는 경와 같이 철자 전사를 통해 학습자의 발음 오류를 반영하기 어려운 경우	()	무조건(무조건)<note>경 음화되지 않음</note> 같이(가티)<note>구개음 화되지 않음</note> 신라(신라)<note>자음동 화되지 않음</note> 앞에(앞에)<note>연음되 지 않음</note> 먹는(먹는)<note>자음동 화되지 않음</note>
	외국어, 외래어 발음	외국어나 외래어를 원어에 가깝게 발음할 경우	()	인터뷰(이너뷰)
		한국어 모어 화자와 다른 규범을 할 경 우	()	카페: 현실 발음 [까페], 학습자 발음 [카페] 버스: 현실 발음 [빠스], 학습자 발음 [버스] → 각각 ‘카페(카페)’, ‘버스(버스)’로 처리함
	방언형 표시		확실한 방언형(대응하는 표준형 형태소가 없는 것)의 경우 태그 부착	차는 <dia>여일</dia> 있어.

대분류	소분류	기호	예시
	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>	혹자들은 현대 사회에 대하여 <Q불확실성의 시대는 아니다.Q> 라고 말하죠.
	텔레비전 방송이나 강의 등 텍스트 종류 표현	<R R>	1:그 다음에 <R생각과 느낌이 유기적으로 잘 짜여져 조직체를 이룰 때 좋은 글이 될 수 있다R>라고 돼 있어요.
	익명성 보장을 위한 마크업	<name> : 사람 이름, 단체 이름, 학교 이름 등 <social security number> : 주민등록번호 <card-num> : 신용카드 번호 <address> : 주소 <tel-num> : 전화 번호	5: 그게 어찌면 <name1> 선배님이라든지 다른 선배님들 말:: 들은 걸 생각해 보면,
발화 단위 분할(segmentation)	분할 단위는 어절 단위로 한다		내{1.2}/가 (X) 내{1.2}가 (O) 빗금은 분할 단위를 나타내는 단위이다.

II. 항목별 세부 설명

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

1. 전체적인 형식과 규정

1) 발화자 표시

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

<예> 1:학습자
 2:원어민

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

<예> 1:어~ 물건을 바꾸고 싶은데요,
 2:네,

- 발화자 표시에 스페이스를 넣지 않는다.

2) 억양 단위

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

가. 억양 단위의 개념

- 구어는 문어와는 달리 정보의 흐름이 통사적인 단위로 이루어지지 않는다. 즉, 문어의 기본단위인 문장은 종결어미로 마무리되고 마침표라는 문장부호로 인해 명확하게 그 단위를 설정할 수 있지만, 구어는 종결어미를 사용하여 발화를 끝내는 경우가 많지 않고, 억양이나 휴지 등의 운율적인 요소에 영향을 받기 때문에 기본단위를 운율적인 단위 곧 억양 단위로 본다.
- 억양 단위는 하나의 통일된 억양 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) !
- 하나의 억양단위 경계에 스페이스 없이 엔터(enter)를 친다.
- 전사 도구에서는 세그멘테이션을 분할하는 것으로 대체한다.

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

나. 끊어진 억양단위(붙임표의 사용)

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

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

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

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

다. 억양단위의 연속성

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

<예> 3:요거는 교수 학습의 개요지,
&요 표는,
4:아::,

3) 겹침 현상

- 겹침 발화의 표시는 겹친 부분을 따로 표시하지 않고 전사 도구를 통해 표시되는 발화 시간을 함께 제시한다¹⁹⁾.

4) 잘 들리지 않는 부분

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

<예> 그때도,
<X보통X> 그런 자만심이 있었다.

19) 21세기 세종계획 말뭉치 구어 전사 지침에서는 연속적인 겹침과 비연속적인 겹침, 동시 다발적 발화 등 다양한 겹침 상황에 따른 세부 전사 지침을 제시하고 있다. 한국어 학습자 말뭉치에서 이러한 지침을 포함하지 않은 것은 전사 도구를 통해 제시되는 발화 시간 정보가 겹침에 관한 정보들을 대체하기 때문이다.

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

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

- 들리지 않는 음절은 그 음절의 수만큼 X를 붙인다.

<예> 근데 그거 진짜 XX해야 되겠더라.

5) 전사 기호의 중복

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

< >, (), 문장부호(온점, 느낌표, 물음표/쉽표), 기타

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

6) 전사자의 설명

- 전사자가 특정 발화 구간에 대한 설명을 붙일 필요가 있을 경우 <note> </note> 태그를 사용하여 표현한다. 녹음 상태, 잡음, 동시다발적인 대화,

특이한 발음 상태 등의 설명이 덧붙을 수 있다. 주석 태그는 다음과 같이 붙여야 하고 앞뒤로 엔터를 쳐 준다.

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

7) 혼잣말

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

<예> <monologue>미치겠네.</monologue>

2. 표기 지침

1) 대원칙

- 발화 내용은 기본적으로 철자법 수준의 전사를 한다. 다만, 구어의 발음 특성, 외국인 학습자의 발음 특성이나 오류, 지역적인 특성 등에 의해 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: 친구(칭구)와 강남(간남)에 갔습니다(갔슨니다).

- ☞ 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 전

사하고 그 밖의 수의적인 발음은 위의 예와 같이 소리 나는 대로 1차 전사를 한다. 그리고 소리 나는 대로 전사한 표기 형태만으로 그 의미를 파악하기 어려운 경우대로 ()의 밖에 원래의 표기를 보충하여 넣도록 한다. (보충적 표기 관련 지침은 ‘4. 2차 전사/철자법 보충’ 참고)

2) 숫자 표기

- 숫자는 아래의 예에서처럼 발음에 따라 한글로 적는다.

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

3) 외래어·외국어 표기

- 외래어나 외국어는 아래의 예에서처럼 발음에 따라 한글로 적는다.

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

4) 끊어진 단어(불완전하게 발화된 단어)

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

<예> 4:사실 학습 자료랑 학= 형태는 떨어져도 되는데,

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

<예> 1:주부 우울증,
=이라고 말할 수 있겠습니다.

- 그러나 완전한 어절이 완전 반복되는 발화의 경우는 ‘=’ 표시를 하지 않고

전사만 한다.

- 발화가 끝나지 않았는데, 말끝을 흐릴 경우 메모를 남긴다.

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

5) 띄어쓰기

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

6) 축약형의 표기

- 구어에서는 발음의 축약 현상이 많이 나타나는데, 두 음절이 한 음절 사잇소리가 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 경우 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 음소가 연결됨을 표시하고 구축 도구 내에서 마크업 처리한다.

<예> 사귀'어, 바뀌'어, ...

7) 장음 처리

- 표기와 관련해서 문제가 되는 간투사는 감정을 나타내는 감탄사와 특별한 의미 없이 말버릇 및 머뭇거림의 표지(담화표지)로 사용되는 간투사 유형이다. 이러한 발화는 그 특성상 원 음절보다 길게 발음되는데, 이 경우에는 ‘::’를 같이 사용해서 표기한다. 참고로 쉼표 표시는 담화표지를 나타냄

으로써 후술한다.

- 마지막 음소를 길게 발음하는 경우 역시 ‘:’로 표기하여 준다.

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

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

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

8) 담화표지

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

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

3. 기타

1) 준음성과 기타 소리들

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

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

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

- 웃으면서 말하는 부분, 박수 치면서 말하는 부분 등도 표시한다.

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

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

4. 2차 전사

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

1) 철자법 보충

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

<예> 친구(칭구)와 강남(간남)에 갔습니다(갈슨니다).

<예> 같이(가티) 가자.

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

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

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

<예> 책상 위에 놔 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.)
무궁한 영광을 위하여,

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

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

선생님(성생닌)

여자(요자)

회사(회사)

과반수(화반수)

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

가구(가구)<note>‘가’의 ㄱ을 유성음으로 발음</note>

가구(가구)<note>‘구’의 ㄱ을 무성음으로 발음</note>

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

무조건(무조건)<note>경음화되지 않음</note>

같이(가티)<note>구개음화되지 않음</note>

신라(신라)<note>자음동화되지 않음</note>

앞에(앞에)<note>연음되지 않음</note>

먹는(먹는)<note>자음동화되지 않음</note>

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

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

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

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

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) 쉼

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

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

- 한 어절 안에서의 짧은 쉼은 ‘.’로 표시한다. 하나의 억양 단위 내부에서의 짧은 쉼은 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 읽기 인용

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

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

6) 익명성 보장을 위한 마크업

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

<name> : 사람 이름, 단체 이름, 학교 이름 등

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

<card-num> : 신용카드 번호

<address> : 주소

<tel-num> : 전화 번호

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

<예> 5:네.

거 어떻게 보면 가장 실망::스런 일 중 하난데요,

헤럴드 쪽에서도 그다지 뽕족한,

대안을 가지고 있는 것 같지는 않더라구요,

그게 어찌면 <name1> 선배님이라든지 다른 선배님들 말::들은 걸 생각해 보면,

<name2> 사장이 <name2> 회장이 있으니까 보안이 있어도 눈치 보여서 얘기를 못한다,

한국어 학습자 말뭉치 형태 주석 지침

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

I. 학습자 말뭉치의 형태 분석 표지²⁰⁾

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

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

	호격조사	JKV	JKV
	인용격조사	JKQ	JKQ
	보조사	JX	JX
(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
	붙임표(숨김, 빠짐)	SO	SO
	외국어	SL	SL
	한자	SH	SH
	기타 기호	SW	SW
	숫자	SN	SN
	분석불능범주	NA	NA

II. 기본 원칙

가. 분석 대상 : 형태분석은 하나의 어절을 분석 대상으로 한다.

나. 분석 원리 : 본 분석은 ‘형태소’ 차원이 아닌 ‘형태’ 차원의 분석이므로 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) <u>표준국어대사전</u> 을 따른다. ■■ <u>중요하다</u> [중요/NNG+하/XSA+다/EF] - 중요 표준국어대사전: 중요/NNG ✓ ※ 단, 조사나 어미의 결합형은 <u>지침에 목록으로 제시된 것을 제외하고는</u> 표준국어대사전에 등재되어 있어도 각각 분리해서 분석한다. ■■ <u>먹었으니까는</u> [먹/VV+었/EP+으니까/EC+는/JX] 나) 사전으로 해결이 불가능한 경우 세종 말뭉치를 확인하고 분석한다.

○ 분석 기준 :

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

마. [학습자] 분석 기준의 적용

- 학습자 언어에서 나타나는 오류의 유형에 따라 다음과 같이 분석한다.

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

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

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

→ 표현 문형의 구성과 인접한 경우 교정어절을 상정하기 어렵다 해도 표현 문형 구성에 포함되는 형태까지는 분석을 한다.²¹⁾

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

가) 형태 분할이 가능한 것은 최대한 분할하는 것을 원칙으로 한다.

■■ 아무도 <u>몰라다</u> (√몰랐다)	[무르/VV+아/EP+다/EF]
--------------------------	-------------------

21) 표현 문형의 경우 <외국인을 위한 한국어 문법2>의 표현 문형 목록을 기준으로 한다.
(부록 참고)

■■ 공간에 해로원데(√해로운데)

[해롭/VA+언데/EC]

■■ 저는 중국 사람은데(√사람인데)

[사람/NNG+으/VCP+ㄴ데/EC]

나) 다음의 경우는 형태 분할을 하지 않고 오형태로 분석한다.

(1) 용언 어간에 포함된 파생접미사에서 오형태가 나타날 경우 형태 분할을 하지 않고 파생 접미사 자체의 오형태로 분석한다.

■■ 7시 30분까지 운동합니다.

[운동/NNG+해/XSV+ㅂ니다/EF]

■■ 시간이 많이 걸려 안 도착해서서

[도착/NNG+해/XSV+시/EP+어서/EC]

※ [참고]

■■ 한국 친구 열악해 기다리게요.

[열악/NNG+하/XSV+아/EC]

■■ 지금까지 말해던 이런 태도가

[말/NNG+하/XSV+아/EP+던/ETM]

→ 위와 같이 파생접미사의 오형태가 아닌 활용으로 인한 경우도 나타나기 때문에, 맥락에 따라 파생접미사의 오형태를 판단한다.

(2) 다음의 축약형에서 나타나는 오형태는 분할하지 않고 오류 어절 자체에서 오형태 분석을 한다.

■■ 내일은 네 생일인데 어제 어머니한테서 소포를 받았다.

[네/NP]

■■ 재 장소 중에서 제일 좋아하는 장소가 홍대입니다.

[재/NP]

■■ 보통 저하고 재 친구들이 홍대에 자주 갑니다.

[재/NP]

■■ 세 아버지는 키가 큼니다.

[세/NP]

■■ 또 궁금한 개 있으면 어제든지 물어보세요.

[개/NNB]

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

■■ 눈이 많(√많이) 왔다.

[많/MAG]

■■ 줄임말은 젊은들이(√젊은이) 많이 사용하는

[젊은/NNG+들/XSN+이/JKS]

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

■■ 많(√많은) 음식이 있었어요.

[많/VA]

다) 상정한 교정어절에 없는 형태소가 추가된 경우: 용언의 경우 어간 혹은 어근을 확보한 후,

(1) 연결/종결어미 앞뒤에 잉여적 요소가 추가된 경우 분리하지 않고 어미의
오형태로 처리한다.²²⁾

■■ 내 꿈을 <u>이뤄진</u> 고나 이루기 위해	[이뤄지/VV+ㄴ 고나/EC]
■■ 교통이 꼭 <u>편립</u> 습니다.	[편리/NNG+ㅂ습니다/EF]
■■ 어학당 교실에 사람이 <u>많은</u> 다.	[많/VV+은다/EF]
■■ 어학당 교실에 사람이 <u>많은</u> 단.	[많/VV+단/EF]

(2) 어간과 전성²³⁾/선어말어미 사이에 잉여적 요소가 추가된 경우 분리하고
NA 처리한다.

■■ <u>하숙집</u> 인기 때문에 사람이 많았다.	[하숙집+이/VCP+ㄴ/NA+기/ETN]
■■ 이번 시험을 볼 때 꼭 잘해야 <u>되</u> 겠다.	[되/VV+ㄴ/NA+겠/EP+다/EF]
■■ 나는 선생님이 <u>되</u> 겠다.	[되/VV+ㅂ/NA+겠/EP+다/EF]
■■ 할 수 있는 일이 <u>많아져</u> 했는데	[많아지/VV+어/NA+시/EP+었/EP+는데/EC]

라) 교정 어절을 상정할 수 있지만 형태소 경계를 구분 짓기 어려운 경우, 교정
어절의 형태소 음절에 따라 앞에서부터 형태를 분할한 후 분석한다.

■■ <u>자시느</u> 이 마음대로 했다.	[자시/NNG+느이/JKG]
■■ 내 유학생활을 <u>아프</u> 로 미래에게	[아/NNG+프로/JKB]
■■ 다언에도 소개해 <u>주</u> 게서요.	[주/VX+게/EP+서요/EF]
■■ <u>어려</u> 슬 데 가을에 좋은 기억이	[어리/VA+어/EP+슬/ETM]
■■ 그것을 <u>마가</u> 려고 하는 것들 중에	[마/VV+가려고/EC]

마) 오류 포함 용언의 활용형을 분석할 때 형태를 경계 짓기 어렵거나 자모 단위
로 쪼개지는 경우가 있는데, 분석한 형태의 결합이 원 어절의 형태가 되지
않더라도 분할을 원칙으로 다음과 같이 분석한다.

22) 기본적인 원칙은 이와 같지만, 전체 텍스트에서 나타나는 학습자의 오류 양상에 따라
연결/종결어미 앞에 나타났어도 NA로 분석되는 경우가 존재할 수 있다.

23) 전성어미는 연결/종결어미와 더 비슷한 특성(어말어미)을 가진다고 할 수 있지만, 형태
적인 면에서는 선어말어미와 같이 타입(type)이 제한되어 있고 형태 자체가 더 고정적이
다. 이러한 특성으로 인해 상정한 교정어절에 없는 형태소가 추가된 경우 직관적으로 (1)
은 어미 이형태를 잘못 쓴 오류로 판단되는 반면, (2)는 어간과 어미 사이에 잉여적 요소
가 첨가된 오류로 판단되는 것을 반영했다.

■■ 쓰레기를 버려도 되면 좋겠습니다.	[버리/VV+어도/EC]
■■ 많은 친구를 사귀어서 재밌었어요.	[사구/VV+어서/EC]
■■ 문화 등 우리 나라하고 달라서 싫어지만	[다르/VA+아서/EC]
■■ 소치에 2시간 걸려요.	[걸레/VV+어요/EF]
■■ 정말 신기하다고 생각했다.	[생각/NNG+하/XSV+ㄴ/EP+다/EF]
■■ 어제 영화를 봤는데	[보/VV+ㄴ/EP+는데/EC]
※ 참고	
■■ 어제 영화를 봤는데	[바/VV+ㄴ/EP+는데/EC]

III. 표지별 분류 기준 및 세부 지침

가. 체언

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

1) 명사(NN)

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

가) 일반명사(NNG)

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

(1) 일반명사로 분석할 수 있는 단어

(가) 표준국어대사전의 명사 표제어

■■ 국어/NNG, 연구/NNG

(나) 1음절 한자어가 독립된 단위로 사용되는 경우

■■ 서울초등학교 졸 [졸/NNG]

※ [참고]

■■ 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]

(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]²⁴⁾

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

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

24) [수정] '21세기세종계획' 지침에는 '느냐/EC'로 되어 있지만 오류이므로 수정함.

(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.

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

나) 고유명사(NNP)

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

(1) 인명, 종족명

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

■ 남수/NNP||군/NNB, 김/NNP||씨/NNB, 최치원/NNP||옹/NNB,
케네디/NNP||씨/NNB²⁵⁾, 정/NNP||과장/NNG, 최/NNP||선생/NNG

(나) 성과 이름, 호가 함께 쓰이면 하나의 단위로 분석한다.

■ 김철수/NNP, 이태백/NNP

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

■ 김/NNP+가/XSN

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

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

(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 인정

하여 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 함께 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 나뉘 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 년도에 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 월 30만원을 받는다.

[월/NNG]

■■ 달러의 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 사람이 교실로 들어왔다.

[사람/NNG+이/JKS]

■■ 자장면 한 그릇만 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 거를 준비해 왔니?

[거/NNB+를/JKO]

■■ 공부할 걸 가져왔니?

[것/NNB+ㄹ/JKO]

■■ 연습할 건 있니?

[것/NNB+ㄴ/JX]

■■ 먹을 게 모자르다

[것/NNB+이/JKS]

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 <u>저거</u> 를 먹을래.	[저거/NP+를/JKO]
■■ 나는 여태 <u>그걸</u> 믿어 왔단다.	[그것/NP+를/JKO]

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

■■ 내	이제부터는 내 명령을 따라라.	[나/NP+의/JKG]
------	------------------	--------------

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

※ [참고]

■■ 내가	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 얘기이냐?	[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 ‘제/NP+가/JKS’에서 ‘가/JKS’를 누락해서 쓴 경우 ‘저/NP+의/JKG’로 분석하지 않도록 주의한다.

■■ 제 먹었습니다.	[제/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, 일

- 흔, 여든, 아흔, 백한들, 두서넛, 서넛, 너덧, 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 수사로 분석한다.

- 광에 가서 물건 하나만 가져오렴. [하나/NR+만/JX]
- 우리는 하나로 뭉쳤다. [하나/NR+로/JKB]

(다) [보완] 때로 수사와 수관형사의 구별이 애매한 경우가 있다. 이 분석에서는 임흥빈(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, 넉
- 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 동사로 처리한다. (세종 말뭉치 기준)

※ [보완] ‘감사하다’는 모두 동사로 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/EF+는/JX
■ ■ 계시다	손님께서 와 계십니다.	계시/VX+입니다/EF+./SF
■ ■ 나가다	추진해 나가는 과정에서 문제가 생겼다.	나가/VX+는/ETM
■ ■ 나다	아침에 깨어 나 보니 그가 없어졌다.	나/VX+아/EF
■ ■ 내다	힘들겠지만 잘 견뎌 내야 한다.	내/VX+아야/EF
■ ■ 놓다	약속을 잡아 놓고 출장을 가다니	놓/VX+고/EF
■ ■ 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/EF
■ ■ 들다	도무지 내 말은 믿으려 들지 않는다.	들/VX+지/EF
■ ■ 말다	어렵더라도 희망을 잃지 말아야 한다.	말/VX+아야/EF
■ ■ 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/체하다/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.²⁶⁾

■■ 그 일을 처리하는 데 철수만한 인재가 없다 [철수/NNP+만/JX+하/VV+ㄴ/ETM]
 ■■ 이 음식은 먹을 만하다. [만/NNB+하/XSA+다/EF+./SF]

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

■■ 자꾸 물어 버릇하다. [버릇/NNG+하/XSV+다/EF+./SF]

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

■■ 열심히 공부하도록 하자. [공부/NNG+하/XSV+도록/EC ㄴ 하/VV+자/EF]

4) 지정사(VC)

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

■■ 철수는 매우 우수한 학생이다. [학생/NNG+이/VCP+다/EF+./SF]
 ■■ 철수는 모범적인 학생이 아니다. [아니/VCN+다/EF+./SF]

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

※ [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]

※ [학습자] 학습자가 지정사 ‘이’를 몰라서 누락한 경우 (오류의 경우)

-> 복원하지 않는다.

■■ 방법은 한 가지예요 [가지/NNB+예요/EF]

■■ 이것은 책상라며 나를 가르쳤다. [책상/NNG+라며/EC]

※ ‘아니다’는 부정 지정사(VCN)으로 분석한다.

다. 수식언

1) 관형사(MM)

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

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

<주의사항>

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

① 관형사, 명사 통용

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

② 관형사, 부사 통용

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

③ 관형사, 명사, 부사 통용

■■ 이내 마음을 어찌 알리요.	[이내/MM]
■■ 아침 들판에 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 [그리/VV+어도/EC]

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

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

※ 용언의 활용형

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

※ 일반 부사

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

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

■■■ 그니까(그러니까), 글고(그리고)/MAJ

■■■ 왜냐면(왜냐하면)/MAG

나) 일반부사(MAG)

<주의사항>

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

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

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 덜달아 왜 난리니?	[덜달/VV+아/EC]
■■■ 우리 함께 더불어 살아가자.	[더불/VV+어/EC]

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

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

라. 독립언

1) 감탄사(IC)

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

■■■ 그림(요), 야호, 어머, 앓, 아, 예, 그래(요), 아니(요), 글썸, 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 뛰어들어갔다. [멍멍/MAG]

- ③ 욕이나 욕설을 나타내는 말은 전체를 감탄사로 분석한다.

■■■ 빌어먹을!

[빌어먹을/IC+!/SF]

- ④ ‘뉘’는 문맥에 따라 대명사와 감탄사의 두 가지 쓰임이 있다.

■■■ 뭔지도 모른 채

[뉘/NP+이/VCP+ㄴ지/EC+도/JX]

■■ ■■ 신문에 뭐 대단한 특종이라도 실렸습니까? [뭐/IC]

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

■■ ■■ 그러어엿/NA, 으 ~ 어 ~ 이/NA

마. 관계언²⁷⁾

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

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

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

1) 격조사(JK)

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

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]

27) 지침에 제시된 조사 목록에서 빠진 이형태와 예시를 추가함

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.	[바람/NNG ㄹ 소리/NNG+를/JKO]

라) 관형격조사(JKG)

- 선행 체언으로 하여금 관형어가 되게 하는 조사이다.

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

마) 부사격조사(JKB)

- 선행 체언으로 하여금 부사어가 되게 하는 조사이다.

■■ 로/으로	망치로 못을 박아야지.	[망치/NNG+로/JKB]
	음식으로 장난치지 마.	[음식/NNG+으로/JKB]
■■ 로서/으로서	교사로서 책임을 다해야 한다.	[교사/NNG+로서/JKB]
	장관으로서 책임을 다해야 한다.	[장관/NNG+으로서/JKB]
■■ 로써/으로써	돌로써 지붕을 만든다고?	[돌/NNG+로써/JKB]
	콩으로써 메주를 쑤다고 해도	[콩/NNG+으로써/JKB]

■■ 같이	바보같이 웃고 다닌다.	[바보/NNG+같이/JKB]
■■ 더러	나더러 이것도 하라고 한다.	[나/NP+더러/JKB]
■■ 랑/이랑	너랑 많이 닮았다. 오늘 동생이랑 싸웠다.	[너/NP+랑/JKB] [동생/NNG+이랑/JKB]
■■ 로부터/ 으로부터	TV로부터 받는 영향력이 시험으로부터 해방되다	[TV/SL+로부터/JKB] [시험/NNG+으로부터/JKB]
■■ 마냥	기영이마냥 놀 수만은 없다.	[기영이/NNP+마냥/JKB]
■■ 마따나	네 말마따나 나도 그래야 한다.	[말/NNG+마따나/JKB]
■■ 만큼	눈물만큼 콧물도 흐른다니까.	[눈물/NNG+만큼/JKB]
■■ 보고	영자보고 놀자고 좀 해라.	[영자/NNP+보고/JKB]
■■ 보다	직관보다는 논리가 동원돼야 한다.	[직관/NNG+보다/JKB+는/JX]
■■ 에	나는 너에 대해 아무것도 모른다.	[너/NP+에/JKB]
■■ 에게	너에게 말하기 싫다.	[너/NP+에게/JKB]
■■ 에게서	나는 철수에게서 그 말을 들었다.	[철수/NNP+에게서/JKB]
■■ 에서	집에서 학교까지 너무 멀다.	[집/NNG+에서/JKB]
■■ 에서부터	연구소에서부터 가게까지는	[연구소/NNG+에서부터/JKB]
■■ 와/과	경미와 함께 다닌다면, 동생과 함께 다닌다면,	[경미/NNP+와/JKB] [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]
■■ 시여/이시여	전능자시여 자비를 베풀어 주옵소서	[전능자/NNG+시여/JKV+!/SS]
	신이시여! 우리를 저버리지 마소서	[신/NNG+이시여/JKV+!/SS]

<주의사항>

- 호격조사와 어말어미는 구분해서 분석해야 한다.

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

사) 인용격조사(JKQ)

- 인용문이나 인용구를, 동사에 대한 부사적 성분으로 도입하는 조사이다.

■■ 고	그는 "이제 가도 좋다"고 말했다.	[좋/VA+다/EF+ "/SS+고/JKQ]
■■ 라고/이라고	문제가 심각하다"라고 보고했다.	[심 각 / X R + 하 / X S A + 다 /EF+ "/SS+라고/JKQ]
	팻말에는 "금지구역"이라고 쓰여 있었다.	[" / S S + 금 지 / N N G + 구 역 /NNG+ "/SS+이라고/JKQ]
■■ 하고	영수는 "이제 가자"하고 말문을 닫았다.	[가/VV+자/EF+ "/SS+하고/JKQ]

<주의사항>28)

- 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] (○)
 ["/SS+금지/NNG+구역/NNG+ "/SS+이/VCP+라고/EC] (×)
- 철수는 "다음 주에 놀러 가도 좋다"고 말하였다.
 [좋/VA+다/EF+ "/SS+고/JKQ] (○)
 [좋/VA+다/EF+ "/SS+고/EC] (×)
- 먼저 "주민등록증이 있냐?"고 묻는다.
 [있/VV+냐/EF+ ?/SF+ "/SS+고/JKQ] (○)
 [기쁘/VV+냐/EF+ "/SS+고/EC] (×)

(2) 연결어미

- 철수는 자기가 학생이라고 말했다.
 [학생/NNG+이/VCP+라고/EC] (○)
 [학생/NNG+이라고/JKQ] (×)
- 자장면을 시킨 뒤 집에 가겠다고 우기는 할머니를 달래기 시작했다.
 [가/VV+겠/EP+다고/EC] (○)

28) <세종> 분석 결과를 바탕으로 다시 정리함

[가/VV+겠/EP+다/EF+고/JKQ] (×)

■■■ 내가 안 기쁘냐고 다그쳐 물었을 때,

[기쁘/VA+냐고/EC] (○)

[기쁘/VA+냐/EF+고/JKQ] (×)

- ② 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.

■■■ [보완] 내 일이다라고 말했다.

[일/NNG+이/VCP+다/EF+라고/JKQ] (○)

[일/NNG+이/VCP+다/EF+이/VCP+라고/EC] (×)

※ [참고] 다음은 간접 인용의 경우로 보고 분석한다.

■■■ 내 일이라고 말했다.

[일/NNG+이/VCP+다/EC] (○)

[일/NNG+이/VCP+다/EF+고/JKQ] (×)

2) 접속조사(JC)

- 두 단어를 같은 자격으로 이어 주는 구실을 하는 조사를 말한다.

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

	연습이다 레슨이다 시간이 하나도 없다.	[연습/NNG+이다/JC]
■■ 랑/이랑	머루랑 다래랑 먹으며 청산에 살고 싶어라.	[머루/NNG+랑/JC]
	떡이랑 과일이랑 많이 먹었다.	[떡/NNG+이랑/JC]
■■ 며/이며	잔칫상에는 배며 대추며 여러 가지 과일이 차려져 있었다.	[배/NNG+며/JC]
	그림이며 조각이며 미술품으로 가득 찬 화실	[그림/NNG+이며/JC]
■■ 에	아버지가 책에, 연필에 이것저것 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]

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

- ③ [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]
■■ 아버지가 나에게 준 배려심이나 사람을 얼마나 많은지 어떻게 계산하는지 이제 마음속에는 다 알게 되었다.	[배려심/NNG+이나/JC]

- ④ [보완] 접속 조사 중에서 ‘고/이고’, ‘니/이니’, ‘다/이다’, ‘며/이며’, ‘에’의 경우는 주로 ‘-고 -고’, ‘-니 -니’와 같은 구성에서 쓰인다. 이들 접속 조사는 연결어미와 동일한 형태인 경우가 있으므로 주의할 필요가 있다.

■■ 슬픔이고 기쁨이고 다 느끼지 못한다.	[슬픔/NNG+이고/JC]
■■ 다른 어떤 무늬보다 개성적이고 색다른 현대 감각을 보여준다면,	[개성/NNG+적/XSN+이/VCP+고/EC]
■■ 웃이며 신이며 죄다 흩어져 있었다.	[웃/NNG+이며/JC]

■■ 내부는 어지럽게 설계된 미로 공간이며, [공간/NNG+이/VCP+며/EC]
같은 건물 안에 반드시 식당가가 있다.

3) 보조사(JX)

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

■■ 그려/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 걸겠다	[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 호수에 사는 초식동물이다.	[코알라/NNG+란/JX]
	사람이란 분수를 지킬 줄 알아야 한다.	[사람/NNG+이란/JX]
■■ ㄹ랑/일랑	강엘랑 가지 마라.	[강/NNG+에/JKB+ㄹ랑/JX]
	그 일에 대해선 걱정일랑 하지 말아라.	[걱정/NNG+일랑/JX]
■■ 마다	꽃마다 독특한 향기가 있다.	[꽃/NNG+마다/JX]
■■ 마저	장미마저 시들고 말았다.	[장미/NNG+마저/JX]
■■ 만	사람은 빵만으로 살 수 없다.	[빵/NNG+만/JX+으로/JKB]
■■ 밖에	이제는 떠날 수밖에 없다.	[수/NNB+밖에/JX]

■■ 부터	우선 노약자부터 태워야 한다.	[노약자/NNG+부터/JX]
■■ 뿐	가진 것은 집 한 채뿐이다.	[채/NNB+뿐/JX+이/VCP+다/EF]
■■ 서꺼	국물이나 동치미서꺼 아무 거나	[동치미/NNG+서꺼/JX]
■■ 사/이사	내사 그걸 이미 했지.	[내/NP+사/JX]
	남이사 무슨 상관이야.	[남/NNG+이사/JX]
■■ 야/이야	그야 그렇지.	[그/NP+야/JX]
	그가 인간성이야 그만이지.	[인간성/NNG+이야/JX]
■■ 야말로/이야말로	사과야말로 가을의 과일이다.	[사과/NNG+야말로/JX]
	통일이야말로 최대의 과업이지.	[통일/NNG+이야말로/JX]
■■ 요	나는 그림을요 잘 그립니다.	[그림/NNG+을/JKO+요/JX]
■■ 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)ㄴ랑, (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.²⁹⁾

■ ■ 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 보조사
로 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+였/EP+나/EF+?/SF]
■ ■ (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.

■ ■ 마음이 예쁜 사람은 좋아하는 사람이 많으니까요. [많/VA+으니까요/EF+./SF]

29) [삭제] 말고,

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

(다) [보완] 보조사 ‘요’의 분석

(1) A: 선생님이 집에 오셨어요.

B: 선생님이요? [선생/NNG+님/XSN+이/JKS+요/JX]

A: 커서 선생님이 되는 게 어떠니?

B: 선생님이요? [선생/NNG+님/XSN+이/JKC+요/JX]

(2) A: 선생님에 대해 알고 있니?

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

(라) ‘비중결어미+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]

(마) ‘말고’는 용언 ‘말다’의 활용형으로 처리한다.

■■ 돈말고 지혜가 필요하다. [돈/NNG||말/VV+고/EC]

바. 의존형태

1) 어미³⁰⁾

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.

30) 지침에 제시된 어미 목록에서 빠진 이형태와 예시를 추가함

■■ -겠-	그 일은 내일 처리하겠다.	[처리/NNG+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+았/EP+었/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 하셨습니다. [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/EC+하/VV+았/EP+다/EF+./SF]

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

③ 위의 선어말어미가 포함되지 않은 어미 형태는 그대로 연결어미로 분석한다.

■■ -랄까-, -대야-, -래야-

④ [보완] ‘-여’나 ‘-었-’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.³¹⁾

■■ -거든	나는 이것이 좋거든!	[좋/VA+거든/EF+!/SF]
■■ -게	그만한 돈이 있으면 좋게.	[좋/VA+게/EF+./SF]
■■ -구나/는구나	넌 정말 멋지구냐! 앞이 잘 안 보이는구나.	[멋지/VA+구나/EF+?/SF] [보이/VV+는구나/EF+./SF]
■■ -구려/는구려	당신도 가시겠구려.	[가/VV+시/EP+겠/EP+구려/EF+./SF]
	잘도 먹는구려.	[먹/VV+는구려/EF+./SF]
■■ -구먼/는구먼	학교가 참 크구먼. 공부를 잘하는구먼. 이것이 무엇인가?	[크/VA+구먼/EF+./SF] [잘/MAG+하/XSV+는구먼/EF+./SF] [무엇/NP+이/VCP+ㄴ가/EF+?/SF]
■■ -ㄴ가/은가/는가	그것이 좋은가? 그가 집에 있는가? 이제 시작인걸.	[좋/VA+은가/EF+?/SF] [있/VV+는가/EF+?/SF] [시작/NNG+이/VCP+ㄴ걸/EF+./SF]
■■ -ㄴ걸/은걸/는걸	그 책은 벌써 다 읽은걸. 그는 벌써 갔는걸.	[읽/VV+은걸/EF+./SF] [가/VV+았/EP+는걸/EF+./SF]
■■ -나	자네 그리로 가냐? 키가 얼마나 크냐?	[가/VV+나/EF+?/SF] [크/VA+나/EF+./SF]
■■ -냐/으냐/느냐	물이 얼마나 깊으냐? 그것보다 이것이 낫느냐?	[깊/VA+으냐/EF+?/SF] [낫/VA+느냐/EF+?/SF]
■■ -냐고/으냐고 /느냐고	그가 누구냐고? 그렇게 싫어? 싫으냐고? 너 뭐 해? 뭐 하느냐고?	[누구/NP+이/VCP+냐고/EF+?/SF] [싫/VA+으냐고/EF+?/SF] [하/VV+느냐고/EF+?/SF]
■■ -네	정말 큰일 났네!	[나/VV+았/EP+네/EF+!/SF]
■■ -도다/는도다	꽃이 아름답도다. 짐이 조서를 내리는도다.	[아름답/VA+도다/EF+./SF] [내리/VV+는도다/EF+./SF]
■■ -니	그게 없니? 그게 사실이다.	[없/VA+니/EF+?/SF] [사실/NNG+이/VCP+다/EF+./SF]
■■ -다/ㄴ다/는다	이건 말도 안 된다. 아이가 글을 잘 읽는다.	[되/VV+ㄴ다/EF+./SF] [읽/VV+는다/EF+./SF]
■■ -다구	돈이 많다구?	[많/VA+다구/EF+?/SF]
■■ -다나	그도 가겠다나.	[가/VV+겠/EP+다나/EF+./SF]
■■ -다네	일을 망쳤다네	[망치/VV+었/EP+다네/EF+./SF]
■■ -다니까	돈이 없다니까!	[없/VA+다니까/EF+!/SF]
■■ -다니	무엇을 읽었다니?	[읽/VV+었/EP+다니/EF+?/SF]
■■ -다면서	술은 싫다면서?	[싫/VA+다면서/EF+?/SF]
■■ -다오	그가 가지고 있다오.	[있/VX+다오/EF+./SF]
■■ -단다	애들이 다쳤단다.	[다치/VV+었/EP+단다/EF+./SF]

31) [수정] 세종 지침에서는 ‘SF’ 앞에서만 종결어미를 인정한다고 하였으나, SF 앞이 아니더라도 한 문장을 끝맺는 역할을 하면 종결 어미로 분석을 하며, 사전에는 종결 어미로 등재되어 있으나 연결 어미로 사용된 경우에는 연결 어미로 분석한다.(종결어미를 연결어미로 분석하는 경우와 관련해서는 <주의사항> (라) 참고)

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

(라) [보완] 사전에 종결어미로 제시되어 있지만 연결어미로 처리하는 경우는 용언이 연결되는 구성에서 앞에 위치한 어절의 끝에 오는 어말어미의 경우로 제한한다.

■■ 아픈데 밥을 먹을까 싶다. [먹/VV+을까/EC]

■■ 무슨 일이 있었는가 했다. [있/VV+었/EP+는가/EC]

※ [참고] 위의 경우를 제외하고는 <표준>을 따라 종결어미로 분석한다.

■■ 돈을 어떻게 쓰느냐에 따라 [쓰/VV+느냐/EF+에/JKB]

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

※ [보완] ***은 사전에 종결어미로만 등재되어 있지만 경우에 따라 연결어미로 분석한 어미이다.

■■ -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	숙제한 것도 빌려가곤 한다.	[빌리/VV+어/EC+가/VX+곤/EC]
■■ -고자	병을 낫고자 몸부림쳤다.	[낫/VV+고자/EC]
■■ -기에	실수했기에 용서해 주었다.	[실수/NNG+하/XSV+았/EP+기에/EC]
■■ -ㄴ 다손/다손	밑다손 치더라도 구박하지 말자.	[밧/VV+다손/EC]
	그가 제시간에 온다손 하더라도	[오/VV+ㄴ 다손/EC]
■■ -ㄴ 다기에/	앞으로 잘 한다기에 승낙했다.	[하/VV+ㄴ 다기에/EC]
는다기에	빵을 먹는다기에 주었다.	[먹/VV+는다기에/EC]
■■ -ㄴ 데/은데/	예쁜데 미워한다.	[예쁘/VV+ㄴ 데/EC]
는데	방이 좁은데 가구는 많다.	[좁/VV+은데/EC]
	눈이 오는데 차를 가져가지 말까?	[오/VV+는데/EC]
■■ -ㄴ 들/는들	간다 한들 아주 갈까?	[하/VV+ㄴ 들/EC]
	그걸 먹는들 뭐가 달라지겠나.	[먹/VV+는들/EC]
■■ -ㄴ 족/은족	배가 고프족 속이 쓰리다.	[고프/VV+ㄴ 족/EC]
	물이 맑은족 고기가 많기는 어렵소.	[맑/VV+은족/EC]
■■ -ㄴ 지라/은지라/	눈이 온지라 길이 미끄럽다.	[오/VV+ㄴ 지라/EC]
는지라	기분이 좋은지라 다정하다.	[좋/VV+은지라/EC]
	선생님께서 고집을 굽히지 않으시는지라	[않/VX+으시/EP+는지라/EC]
■■ -나/으나	눈이 오나 비가 오나	[오/VV+나/EC]
	밥을 먹으나 마나이다.	[먹/VV+으나/EC]
■■ -나니	멀리 보이나니 넓은 들이로다.	[보이/VV+나니/EC]
■■ -나마/으나마	도와주지는 못하나마 방해를 해서는	[못/MAG+하/XSV+나마/EC]
	맛은 없으나마 많이 드세요.	[없/VV+으나마/EC]
■■ -노니	묻노니, 포부가 무엇이냐?	[묻/VV+노니/EC+/,/SP]
■■ -니/으니	멍하니 서있다.	[멍하/VV+니/EC]
	이 옷은 작으니 큰 것으로 바꿔 주세요.	[작/VV+으니/EC]
■■ -느니	앉아서 걱정하느니 나가서 하겠다.	[걱정/NNG+하/XSV+느니/EC]
■■ -니까/으니까	웃기니까 좋다.	[웃기/VV+니까/EC]
	약속을 했으니까 만나야 한다.	[하/VV+았/EP+으니까/EC]
■■ -다 ***	공부를 잘하는 편이다 보니	[편/NNB+이/VCP+다/EC]
■■ -다가	자랑하다가 망신당했다.	[자랑/NNG+하/XSV+다가/EC]
■■ -다기에	그녀가 예쁘다기에 보러 왔소.	[예쁘/VV+다기에/EC]
■■ -대도	시간이 있대도 만나 주질 않는다.	[있/VV+대도/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?	[어떠하/VA+든지/EC]
■■ -되	싸우되 꼭 지도록 해라.	[싸우/VV+되/EC]
■■ -ㄹ까/을까	영화를 볼까 싶다.	[보/VV+ㄹ까/EC]
***	아픈데 밥을 먹을까 싶다.	[먹/VV+을까/EC]
■■ -ㄹ뿐더러/ 을뿐더러	비가 올뿐더러 바람도 분다.	[오/VV+ㄹ뿐더러/EC]
■■ -ㄹ수록/ 을수록	그는 재산이 많을수록 재능도 많다	[많/VA+을뿐더러/EC]
	갈수록 태산이다.	[가/VV+ㄹ수록/EC]
	이 책은 읽을수록 감동을 준다.	[읽/VV+을수록/EC]
■■ -ㄹ지/을지	비가 얼마나 올지 천둥이 다 친다.	[오/VV+ㄹ지/EC]
	내일은 얼마나 날씨가 좋을지	[좋/VA+을지/EC]
■■ -ㄹ지라도/ 을지라도	이길지라도 명예롭지는 않다.	[이기/VV+ㄹ지라도/EC]
■■ -ㄹ지언정/ 을지언정	마음에 걱정이 있을지라도 내색하지 마라.	[있/VA+을지라도/EC]
	그것은 무모한 행동일지언정	[행동/NNG+이/VCP+ㄹ지언정/EC]
	죽을지언정 그 일은 못하겠다.	[죽/VA+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.	[청소/NNG+하/XSV+러/EC]
	점심 먹으러 집에 간다.	[먹/VV+으러/EC]
■■ -려/으려	학교에 가려 한다.	[가/VV+려/EC]
	웃으려 한다.	[웃/VV+으려/EC]
■■ -려니와/ 으려니와	비용도 문제려니와 일꾼도 문제다.	[문제/NNG+이/VCP+려니와/EC]
■■ -련마는/ 으련마는	이 마을은 경치도 좋으려니와	[좋/VA+으려니와/EC]
	보면 반가우련마는 볼 수가 없네.	[반갑/VA+련마는/EC]
■■ -ㄴ다 ***	벌써 제 잘못을 알았으련마는	[알/VV+았/EP+으련마는/EC]
	안 그럼 때릴 거야.	[그리/VV+ㄴ다/EC]
■■ -며/으며	노래하며 춤을 춘다.	[노래/NNG+하/XSV+며/EC]
	강물이 맑으며 깊다.	[맑/VA+으며/EC]
■■ -면/으면	지옥이 존재하면 만원일 것이다.	[존재/NNG+하/XSV+면/EC]
	내일 날씨가 좋으면 소풍을 가겠다.	[좋/VA+으면/EC]
■■ -면서/으면서	푸르면서 검은 물빛	[푸르/VA+면서/EC]
	밥을 먹으면서 신문을 본다.	[먹/VV+으면서/EC]
■■ -므로/으므로	비가 오므로 가지 않겠다.	[오/VV+므로/EC]
	강이 깊으므로 배 없이 건널 수 없다.	[깊/VA+으므로/EC]
■■ -아/어	입을 막아 버렸다.	[막/VV+아/EC]
	밥을 먹어 버렸다.	[먹/VV+어/EC]
■■ -아도/어도	암만 봐도 모르겠다.	[보/VV+아도/EC]

	나는 부자가 아니어도 행복하다.	[아니/VCN+어도/EC]
■■ -아서/어서	땀을 흘려서 땀을 잡았다.	[놓/VV+아서/EC]
	그는 걸어서 학교에 갔다.	[걸/VV+어서/EC]
■■ -아야/어야	이 일은 잘해야 한다.	[잘/MAG+하/XSV+아야/EC]
	사람은 먹어야 산다.	[먹/VV+어야/EC]
■■ -자마자	오자마자 당했다.	[오/VV+자마자/EC]
■■ -지	우기지 못해 버렸다.	[우기/VV+지/EC]
■■ -지마는	비가 오지마는 가야 한다.	[오/VV+지마는/EC]

<주의사항>

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

■■ 너는 내가 왔는데 기쁘지도 않니?	[오/VV+았/EP+는데/EC]
■■ 내가 지금 있는 데가 어디지?	[있/VV+는/ETMⅡ데/NNB+가/JKS]
■■ 다들 만족하는지 아무런 불평이 없다.	[만족/NNG+하/XSV+는지/EC]
■■ 너를 만난 지도 꽤 오래구나.	[만나/VV+ㄴ/ETMⅡ지/NNB+도/JX]

(나) ‘-음직’은 “음직/EC”로 분석한다. 그러나 ‘바람직, 먹음직’ 등은 그 자체가 하나의 어근이므로 더 이상 분석할 수 없다는 것에 유의한다.

■■ 어른답고 믿음직하게 행동해라.	[믿/VV+음직/EC+하/VX+게/EC]
■■ 그것 참 먹음직스럽다.	[먹음직/XR+스럽/XSA+다/EF+./SF]
■■ 그것은 매우 바람직한 일이다.	[바람직/XR+하/XSA+ㄴ/ETM]

라) 명사형 전성 어미(ETN)

- 한 문장의 성격을 임시로 바꾸어 다른 문장 속에서 명사적인 역할을 하게 하는 어미를 말한다.

■■ -기	그 일은 정말 중요하기 때문이다.	[중요/XR+하/XSA+기/ETN]
	학생 신분임을 밝히다.	[신분/NNG+이/VCP+ㅁ/ETN]
■■ -ㅁ/-음	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	읽어버린 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) 연결 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(在)-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-풋살, 피(被)-피고소인, 한-한가운데, 헛-헛고생

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 연구자에 따라 그 목록이 다르며, 실제로도 구분이 애매한 경우가 많다. 본 분석에서는 접미사의 생산성과 접미사를 제외한 형태의 독립성을 기준으로 다음과 같이 목록을 마련하였다.

가(價)-매매가, 가(哥)-김가, 경(頃)-두 시경, 계(系)-몽고계, 계(界)-교육계, 광(狂)-메모광, 권(圈)-운동권, 권(權)-참정권, 당(當)-한 사람당, 대(臺)-억대, 댁(宅)-청주댁, 론(論)-비평론, 별(別)-가구별, 여(餘)-삼십여, 류(類)-자연류, 룰, 율(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, 껌-10분께, 꿀-십 원꿀, 끼리-전우끼리, 껌-노름꾼, 네-동이네, 님-선생님, 들-우리들, 들이-1ㄹ들이, 배기-열 살배기, 뽕-조카뽕, 씹-만원씩, 장이-간판장이, 쟁이-심술쟁이, 쯤 -내일쯤, 질-서방질, 짜리-백 원짜리, 쩌1 -이틀쨌, 쨌2-옹기쨌, 치레-인사치레, 투성이-먼지투성이

<주의사항>

(가) 명사파생접미사인 ‘-들’은 그 분포가 매우 다양하여 일부에서는 이를 보조사와 접미사로 나누어 분석하기도 한다. 그러나, 본 분석에서는 이들을 모두 명사파생접미사로 처리한다. ‘먹고들’의 ‘-들’도 선행성분이 어미이긴 하나, 일치하는 대상은 선행하는 명사로 해석할 수도 있기 때문이다.

- 사람들이 우리 집에 왔다. [사람/NNG+들/XSN]
- 그들은 밥을 먹고들 싶었다. [먹/VV+고/EC+들/XSN]

(나) ‘-님’은 다음과 같이 세 가지의 분석 중의성을 가지므로 주의해서 분석한다.

① ‘임’의 의미로 쓰인 경우: 보통명사

- 님과 이별하다. [님/NNG+과/JKB]

② 사람의 ‘이름’이나 ‘성’ 뒤에서 쓰인 경우: 의존명사

- 김철수님께서 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]

※ [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.

예) ‘공부하다’와 ‘공부 하다’의 구분

1. 공부/NNG+하/XSV

- ■ 외국에서 [공부하는] 일이 쉬운 것은 아니다. [공부/NNG+하/XSV+는/ETM]
- ■ 외국에서 [공부 하는] 일이 쉬운 것은 아니다. [공부/NNG+하/XSV+는/ETM]

→ 띄어쓰기가 잘못 되어 있는 경우 어절 경계를 조정하여 붙인다.(기능 구현 이후, 이전에는 메모 남기기)

2. 공부/NNG || 하/VV

- ■ 절에서 [선 공부] 하는 것을 본 사람이 있다. [공부/NNG || 하/VV+는/ETM]
- ■ 서울에 [학당 공부] 하러 올라갔다. [공부/NNG || 하/VV+러/EC]
- ■ 서울에 [학당 공부]하러 올라갔다. [공부/NNG | 하/VV+러/EC]

※ [보완] ‘NA+하다’의 접사 ‘-하’는 XSV로 분석한다.

※ [보완] 학습자가 잘못 접미사를 사용한 경우 교정어절을 상정했을 때 교정어절의 품사가 동사일 때는 동사파생접미사, 교정어절의 품사가 형용사일 때는 형용사파생접미사로 분석한다.

- ■ 음식을 먹하다. [먹/VV+하/XSV+다/EF+./SF]

■■ 마음이 아픈 아주머니가 집에 돌아왔다.	[아프/VA+하/XSA+ㄴ/ETM]
■■ 그렇지 않다면	[않/VX+하/XSA+ㄴ 다면/EC]
■■ 한국어 공부를 열심히 하다.	[열심/NNG+하/XSV+다/EF]

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]
■■ 롬	어려운 일일수록 슬기롭게 대처하라.	[슬기/NNG+롬/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]
■■ 이리하다	[이리/XR+하/XSA+다/EF]

<주의사항>

- 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.

■■ 듻직하다	[듻직/XR+하/XSV+다/EF]	
■■ 취하다	[취하/VV+다/EF]	‘취’는 어근 (위하다, 반하다, 강하다, 약하다, 중하다, 대하다, 의하다, 통하다 등)
■■ 밥하다	[밥/NNG+하/XSV+다/EF]	‘밥’은 명사

(절하다, 인하다, 비하다, 한하다 등)

■■ 잘되다

[잘/MAG+되/XSV+다/EF]

'잘'은 부사

※ [보완] “못하다”의 경우 ‘못하/VV, 못하/VA, 못하/VX’의 세 가지 경우가 존재하는데, 이때 ‘-하’를 분석할 경우 본용언의 분석과 보조용언의 분석이 동형이 되기 때문에 예외로 취급해서 ‘-하’를 분석하지 않는다.

■■ 못하다

[못하/VV+다/EF]

[못하/VA+다/EF]

[못하/VX+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]

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] (○)

3) 분석불능범주(NA)

※ [보완] 그 자체가 사전에 등재되어 있지도 않으면서, 축약의 정도가 심하거나 분석하기 어려운 방언형의 경우 분석불능범주로 처리한다.

■■ 담배가 쪼매턴게 하마 자라서 빼나?	[쪼매턴게/NA]
■■ 친구한테 전화를 적긴 일이었다.	[적긴/NA]
■■ "부산국제영화"제가니와	[제가니와/NA]
■■ 있잖아요	[있/VV+잖/NA+아요/EF]
■■ ㅋㅋ	[ㅋㅋ/NA]
■■ ππ	[ππ/NA]
■■ ㅇㅋㅇㅋ	[ㅇㅋㅇㅋ/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 붙어 쓰이는 명사와 붙여 쓴다. 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 //멜감 //양념감 //안춧감 //장난감//웃음감//사형감// 노벨상감//바느질감.
값 「07」	(일부 명사 뒤에 붙어) ‘가격’, ‘대금’, ‘비용’의 뜻을 나타내는 말.	기름값 //물값 //물건값 //부식값 //신문값 //우윳값 //음식값.
값 「08」	(일부 명사 뒤에 붙어) ‘수치’의 뜻을 나타내는 말.	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//사기극//실험극//특집극
금06 「04」	(일부 명사 앞에 붙어) ‘금색1’, ‘금제1’의 뜻을 나타내는 말.	금두꺼비 //금목걸이 //금수저.
급04 「05」	(직급 따위를 나타내는 일부 명사 뒤에 붙어) ‘그 직급’의 뜻을 나타내는 말.	과장급 //부장급 //간부급.
길01 「10」	(일부 명사 뒤에 붙어) ‘과정’, ‘도중’, ‘중간’의 뜻을 나타내는 말.	산책길//시장길
꽃01 「07」	(일부 명사 뒤에 붙어) ‘그 꽃’의 뜻을 나타내는 말.	도라지꽃//무궁화꽃//목련꽃//민들레꽃//사과꽃//유채꽃.
난05 「02」	(고유어와 외래어 명사 뒤에 붙어) ‘구분된 지면’의 뜻을 나타내는 말.	어린이난//가십난//컴퓨터난//해외 토픽난.
놀이01 「04」	(일부 명사 뒤에 붙어) ‘모방4’, ‘장난’, ‘흥내’의 뜻을 나타내는 말.	시장놀이//병원놀이//엄마놀이//학교놀이.
대15 「03」	(일부 명사 뒤에 붙어) 받침이 되는 시설이나 이용물의 뜻을 나타내는 말.	급수대 //조희대 //독서대.
택01 「04」	(일부 명사 뒤에 붙어) ‘택호’를 나타내는 말.	윤 판서택
덩어리 「03」	(일부 명사 뒤에 붙어) [같은 말] 덩이(3. 그러한 성질을 가지거나 그런 일을 일으키는 사람이나 사물을 나타내는 말).	꿀칫덩어리 //심술덩어리 //애꿎덩어리//채주덩어리.
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 //접포망//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//골뱅이무침//파래무침.
문06 「01」	(일부 명사 뒤에 붙어) 학술 전문의 종류를 나타내는 말.	어학문 //법학문.
문06 「02」	(일부 명사 뒤에 붙어) 씨족에 따른 집안을 나타내는 말.	강씨문(姜氏門) //이씨문(李氏門).
미14 「02」	(일부 명사 앞 또는 뒤에 붙어) ‘아름다움’의 뜻을 나타내는 말.	미소년 //승고미 //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 //잠기름병 //플라스틱병.
볶음 「02」	(일부 명사 뒤에 붙어) 볶아서 만든 음식의 뜻을 나타내는 말.	쇠고기볶음 //야채볶음.
불09 「02」	(일부 명사 뒤에 붙어) ‘부처1’의 뜻을 나타내는 말.	무량수불 //아미타불.
비05 「03」	(일부 명사 뒤에 붙어) ‘비율2’의 뜻을 나타내는 말.	농도비 //혼합비.
비19 「03」	(일부 명사 뒤에 붙어) 기념하여 세운 물건의 뜻을 나타내는 말.	문학비 //문인비.
빛 「07」	(일부 명사 뒤에 붙어) ‘빛깔’의 뜻을 나타내는 말.	능금빛 //산빛.
상04 「03」	(일부 명사 뒤에 붙어) ‘상차림’을 나타내는 말.	다과상 //생신상 //차례상.
상23 「02」	(일부 명사 뒤에 붙어) 조각이나 그림을 나타내는 말.	성당의 성모 마리아상.
상23 「03」	(일부 명사 뒤에 붙어) ‘모범2’, ‘본보기’의 뜻을 나타내는 말.	교사상 //어머니상.
상25 「02」	(일부 명사 뒤에 붙어) ‘상장10’, ‘상패4’, ‘상품4’ 따위의 뜻을 나타내는 말.	감독상 //봉사상 //선행상 //작품상 //효행상.
색03 「05」	(일부 명사 뒤에 붙어) ‘색깔’의 뜻을 나타내는 말.	딸기색 //바이올렛색.
선14 「07」	(일부 명사 뒤에 붙어) ‘광선1’의 뜻을 나타내는 말.	감마선 //엑스선.
식04 「04」	(일부 명사 뒤에 붙어) ‘수법’, ‘수식’을 나타내는 말.	곱셈식 //덧셈식 //나눗셈식 //뺄셈식.
쌍02 「03」	(일부 명사 앞에 붙어) ‘두 짝으로 이루어짐.’의 뜻을 나타내는 말.	쌍가락지 //쌍가마 //쌍권총.
씨01 「05」	(일부 식물이나 동물을 나타내는 명사 뒤에 붙어) 그 식물이나 동물의 씨를 나타내는 말.	배추씨//살구씨//굴씨//조개씨.

안04 「04」	(일부 명사 뒤에 붙어) ‘안전’의 뜻을 나타내는 말.	개정안 //채택안 //협상안.
알01 「09」	(일부 식물이나 동물을 나타내는 명사 뒤에 붙어) 그 식물이나 동물의 알을 나타내는 말.	머루알//은행알//타조알.
액03 「02」	(일부 명사 뒤에 붙어) ‘액체’의 뜻을 나타내는 말.	냉각액 //링거액 //수정액.
양20 「02」	(고유어와 외래어 명사 뒤에 붙어) 분량이나 수량을 나타내는 말.	구름양//알칼리양.
옥03 「02」	(일부 명사 앞에 붙어) ‘옥색1’, ‘옥제2’의 뜻을 나타내는 말.	옥제털이 //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.	은감치 //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) ‘저적2’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	막내티 //소녀티 //중년티 //춘티.

	뜻을 나타내는 말.	
표05 「07」	(일부 명사 뒤에 붙어) ‘그 사람이 만든 물건’의 뜻을 더하는 말.	엄마표 //아빠표 //신랑표 //주부표.
후08 「03」	(일부 명사 앞에 붙어) ‘뒤나 다음’의 뜻을 나타내는 말.	후더침 //후보름 //후서방.

아. [보완] 구어 형태 분석 말뭉치(추가)

- 구어 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 하여 하고 [하/VV+아/EC]
■■ 충고를 해 준 쥐 줄 [주/VX+ㄴ/ETM | 주/VX+어/EC | 주/VX+ㄹ/ETM]

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 NA 표지를 부여한다.

가) 어절의 일부만 발화되었지만 분석이 가능한 경우

■■ 어 손의 관절에 통증 통증이 [통증/NNG]
■■ 그러지 변화 그렇게 변화하지 않는다고 [변화/NNG]
■■ 미국과 같 같은 [같/VA]
■■ 그랬 그랬어요. [그러/VV+았/EP]
■■ 의사 선생님도 곧 나 나올 수 있다고 [나/VV]
■■ 비결인 인 것 같은데. [이/VCP+ㄴ/ETM]

나) 어절의 일부만 발화되어 분석이 불가능한 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]
■■ 경복궁에 가 봐 봤다.	[봐/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]
c.f 일본인 학스 학습자들이	[학스/NA]
■■ 일상 일상생활에서 주로 공부를 해요	[일상/NNG]
■■ 한국 음식을 좋아 좋아하지만	[좋/VA+아/EC]
한국 음식을 좋아하 좋아하지만	[좋아하/VA]
■■ 한국어 공부에 힘 힘들었지만	[힘/NNG]

다) 어절의 일부를 더듬으며 반복하는 경우 (용언의 경우)

■■ 결혼하 하기 한	[결혼/NNG+하/XSV하/XSV+기/ETN하/XSV+L/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/XSV+다/EF]
■■ 좋아 어 하다	[좋/VA+아/EC 어/IC 하/VX+다/EF]

3) 억양 단위가 바뀐 어절

가) 억양 단위가 형태소 경계로 바뀐 경우

- 발화자가 불완전하게 발화한 것은 아니지만 한 어절을 발화하는 도중에 억양 단위가 바뀌어서 조사나 어미 등 문법 형태소가 실질 형태소와 다른 억양 단위로 전사될 때, 억양 단위를 통합하지 않고 경계를 살려 형태 주석한다. 하지만 주석은 통합했을 때의 표지를 부여한다.

■■ 캔 유 두 미어 페이버?	[페이버/NNG]
가 무슨 뜻?	[가/JKS]
■■ 좋아.	[좋/VA+아/EF]
라고 대답했지.	[라고/JKQ]
■■ 주부 우울증.	[우울증/NNG]
이라고 말할 수 있겠습니까.	[이/VCP+라고/EC]
■■ 공부.	[공부/NNG]
한다고	[하/XSV+ㄴ다고/EC]

나) 억양 단위가 형태소를 가르는 경우

- 형태소 중간에 억양 단위가 바뀌어서 다른 억양 단위로 전사될 때, 각각 NA로 분석한다.

■■ 어,	[어/NA]
제는 별일 없었어.	[제/NA+는/JX]
■■ 선두주,	[선두/NNG 주/NA]
자가 도착했다.	[자/NA+가/JKS]

4) 불분명한 어절

- 잘 들리지 않아 추측하여 전사한 어절은 최대한 분석하고 분석이 불가능한 경우에는 NA 표지로 분석한다.
- 들리지 않는 음절수만큼 붙인 X에도 NA 표지를 부여한다. X가 한 어절의 일부인 경우에는 해당 단어 부분을 NA로 분석한다.

■■ 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 'NP', 조사 없이 확실하게 뒤의 명사를 수식할 때는 'MM', 위의 경우가 아니거나 확실하게 감탄사로 사용된 경우에는 'IC'로 처리한다. (※ 구분이 애매한 경우 IC로 분석한다.)

■■ 그는 참으로 좋은 사람이다.	[그/NP+는/JX]
■■ 그 책 좀 이리 줘 봐.	[그/MM Ⅱ 책/NNG]
■■ 그 무엇인가를 알아내고자 했지만	[그/MM Ⅱ 무엇/NP+이/VCP+L 가/EF+를/JKO]
■■ 그 왜 있잖아요.	[그/IC]
■■ 이도 저도 다 싫다.	[저/NP+도/JX]
■■ 저 둘 중에 하나를 선택해라.	[저/MM Ⅱ 둘/NR Ⅱ 중/NNB+에/JKB]
■■ 저, 뭐라더라..	[저/IC]
■■ 저 말씀 중에 잠시 실례하겠습니다..	[저/IC]

나) 아니

- 아니 : 대답이나 감탄일 때는 'IC', 부정이나 반대의 뜻을 나타낼 때나 명사와 명사 또는 문장과 문장 사이에서 강조할 때는 'MAG'로 처리한다.

■■ 잠 자니? / 아니, 안 자.	[아니/IC]
■■ 아니, 그럴 수가 있니?	[아니/IC]
■■ 아침까지만 해도, 아니 점심 먹을 때만 해도...	[아니/MAG]

[보완] <주의사항>

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

■■ 학교에 예	[예/JKB]
----------	---------

가서 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) 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]

자. [보완] 구어 형태 분석 말뭉치 (시스템 관련)

1) 문장 분리

구어 말뭉치에서 마침표는 하나의 문장이 끝났음을 나타내는 것이 아니라 억양 단위를 나타내는 기호이므로 임의로 추가할 수 없다. 아래와 같은 예에서, ‘왔어요’ 뒤에 마침표를 찍어 두 억양 단위로 구분하지 않도록 한다. 하지만 ‘오/VV+왔/EP+어요/EF || 음/IC’와 같이 ‘어요’를 종결어미로 분석한다.

■■ 그때 이 년 전에 친구랑 같이 여행 왔어요 음 극 때...

2) 교정 어절이 있는 경우 형태 주석

- 구어 말뭉치는 미리 ‘교정 어절’과 ‘교정 주석’이 들어간다. 이 경우 아래와 같이 처리한다.

가) 통합이나 분할이 필요 없이 수정이 가능한 경우

작업 문장 정보					
	4	15	16	17	
원어절		많이	목꼬		문화
형태주석		MAG	VV	EC	NNG
교정어절			먹	고	
교정주석			VV	EC	



작업 문장 정보					
	4	15	16	17	
원어절		많이	목	꼬	문화
형태주석		MAG	VV	EC	NNG
교정어절			먹	고	
교정주석			VV	EC	

나) 통합이든 분할이든 수정했을 때 분석이 가능한 경우

- ‘그때’는 한 단어로 NNG로 분석해야 한다. 먼저 원어절에서 어절 경계가 분할된 1번과 2번의 어절을 통합해서 1번을 ‘극때’로 수정하고 NNG로 분석을 한 후, 2번을 삭제한다. 이때 아래 교정 어절에 ‘그/MM’만 남게 되는데 ‘그때’로 수정하고 NNG로 교정 주석을 수정한다.

작업 문장 정보				
	1	2	3	
원어절	극	때	부터	한
형태주석	MM	NNG	JX	↑
교정어절	그			
교정주석	MM			



작업 문장 정보			
	1	2	
원어절	극때	부터	한
▶ 형태주석	NNG	JX	N
교정어절	그때		
교정주석	NNG		

다) 통합이든 분할이든 수정할 수 없으며, 원어절의 형태주석을 NA 처리해야 하는 경우

- 구어 팀에서 제시한 교정어절대로 형태 분석하기에 무리가 있다고 판단되는 경우는 NA로 분석한다. 아래 제시된 경우처럼 ‘스고시’는 ‘것이’의 오류로 ‘것+이’로 분리하여 분석하기에는 무리가 있으므로 NA로 처리해야 한다. 이 경우 교정어절에 ‘것’ 위의 형태주석 1번 칸의 ‘스고시’는 NA로 수정하고, 2번 칸의 형태 주석은 시스템 왼쪽 아래의 ‘주석 삭제’ 기능을 이용하여 삭제하고 빈칸으로 남겨둔다.

작업 문장 정보					
	1	2	3	4	5
▶ 원어절	스고시		별로	안	꼼꼼
형태주석	NA		MAG	MAG	MAG
교정어절	것	이			
교정주석	NNB	JKS			

- 위와 같이 구어에서 제시한 교정어절대로 형태 분석을 하기 어려운 경우는, 메모를 남겨 다음 단계 작업에서 확인할 수 있도록 한다.

<메모 형식> 구어 팀에서 제시한 교정어절대로 ‘○○’을 ‘△△’으로 교정할 가능성이 낮아 보임

한국어 학습자 말뭉치 오류 주석 지침

I. 학습자 말뭉치 오류 주석 체계 틀

1. 기본 주석

- 오류 위치는 오류가 나타난 부분의 품사를 주석한다. 오류 위치는 기본 주석으로 형태소 분석 단계에서 모든 오류에 대해 전수 주석한다. 모든 오류는 오류 위치 검색으로 찾을 수 있다.³²⁾

	오류 유형		주석 표지
분석 불가능	전체적 오류 포함		IMP
오류 위치	실질어휘	고유명사	CNNP
		일반명사	CNNG
		의존명사	CNNB
		대명사	CNP
		수사	CNR
		동사	CVV
		형용사	CVA
		보조용언	CVX
		지정사	CVC
		관형사	CMM
		일반부사	CMAG
		접속부사	CMAJ
		감탄사	CIC
		접두사	CXPN

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

	오류 유형		주석 표지
		명사파생접미사	CXSN
		동사파생접미사	CXSV
		형용사파생접미사	CXSA
		어근	CXR
	기능어휘	주격조사	FNP
		관형격조사	FGP
		목적격조사	FOP
		부사격조사	FAP
		접속조사	FJC
		보격조사	FCP
		호격조사	FVP
		인용격조사	FQP
		보조사	FXP
		연결어미	FED
		종결어미	FFE
		선어말어미	FPE
		명사형 전성어미	FNE
		관형사형 전성어미	FAE
	구 단위 표현		PHE
	표현 문형		PE

2. 확장 주석

- 확장 주석은 한국어교육의 선행 오류 연구에서 유의미한 주석에 초점을 두어 교수자의 활용에 초점을 둔 주석이다. 연구자들은 필요한 주석을 추가하여 스스로 주석할 수 있다. 교정 어절에 대한 형태 주석에 기대어 주석한다.

2.1. 오류 양상

○ 어휘나 문법의 층위에서 발생하는 오류 양상만을 주석한다.³³⁾

	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
		시제	ST

33) 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
		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.
- 오류 판단 시 어문 규범에 어긋나는 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 문장이 초급보다 높은 수준이지만 정확한 문장 생성에 실패하였으므로 오류로 보아 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 보아 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 않지

만, 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.

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.

3) 오류의 교정(교정 어절 제시 원칙)

- 오류의 교정은 오류로 식별된 부분을 올바르게 고치는 것을 말한다. 따라서 오류로 식별된 것은 교정의 대상이 된다.
- 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 수정을 한다.
- 정보가 소실되지 않는 차원에서 최소한의 교정을 원칙으로 한다. 즉, 앞부

분의 오류를 수정하는 것으로 인해 뒷부분에까지 영향을 미쳐 뒷부분까지 바뀌게 되면 안 된다. 따라서 앞부분을 수정하는 것 때문에 뒷부분까지 수정하도록 영향을 미치게 되면 수정하지 않는다.

<예> 제가(√저는) 가족에서 3남매 중 막내로 태어났다(√태어나다) 보니 성격은 좀 활발하고 어릴 때부터 오빠랑 언니 보고 좀 눈치가 있는 편입니다.

☞ 자연스러운 문장으로 교정하기 위해 ‘저는 가족 중 3남매의 막내로 태어났다’로 수정하게 되면, ‘가족에서’를 ‘가족 중’으로 수정함에 따라 뒤의 ‘3남매 중’까지 수정하게 된다. 따라서 ‘3남매 중 막내로 태어났다’는 올바른 문장이므로 교정하지 않는다. 이처럼 앞부분을 교정함으로써 뒷부분까지 교정하게 되면 앞부분은 수정하지 않는다. 따라서 이 경우 ‘저는 가족에서 3남매 중 막내로 태어났다’로 정보가 소실되지 않는 차원에서 최소로 교정한다. 그리고 ‘태어났다 보니’를 ‘오형태’와 ‘시제’ 오류로 처리한다.

- 오류 영역에서 교정 어절로 인해 조사나 어미가 바뀌는 경우는 오류로 처리하지 않는다. 기본적으로 맥락을 살펴 되도록 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.

<예> 아파트가 평형이(√평수가) 많으면(√넓으면, √크면) 친구들을 부를 수 있다.

☞ 교정형으로 인한 조사 교정 어절은 조사 오류로 처리하지 않는다. ‘평형’과 ‘많다’만 대치 오류로 처리하고, 조사 ‘이’는 오류로 처리하지 않는다.

나라가(√나라를) 발전하다(√발전시키다)

☞ 사동 ‘시키다’로 교정해야 할 경우, 조사까지 교정해야 한다. 그러나 이 경우에는 조사 ‘를’은 대치로 처리하지 않고 ‘가’ 아래에 교정어절만 써준다. 따라서 [오류 위치-접사], [오

류 양상-대치], [오류 층위-사동]오류로 처리한다.

- 한국어 모어 화자의 보편적인 언어 사용 방식에 따라 수정한다.

4) 오류 판정의 대상

- 오류 판정은 오류 주석을 위해서 식별되어 수정된 부분이 어떤 범주의 오류인지를 판정하는 것을 말한다.
- 교정 어절의 품사가 바뀌는 경우에도, 오류에 대한 주석이므로 교정 어절이 아닌 오류 어절을 기준으로 한다.

<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.

- 즉, 오류를 판정할 때에는 학습자가 산출한 언어 형태와 오류 발생 위치를 기준으로 한다. 하지만 교정어절을 중심으로 한 검색도 가능하게 하여, 미사용으로 인한 오류를 파악할 수 있게 한다.

<예> 이런 심한 문제는 사람들이 과도 개발하고 환경문제를 등한시한 결과였다(√결과였다).
☞ 이 경우 과거 시제 선어말어미의 형태 오류로 볼 수도 있으나 일관되게 지정사 오류로 처리한 후 그 안에서 선어말어미 오류를 다시 살필 수 있도록 한다.

- 오류 주석은 형태 단위를 기본으로 한다. 따라서 구 단위 이상의 어휘나 표현은 구성 요소를 형태로 나누어 분석한다. 단, 표현 문형의 경우는

<국립국어원 2> 목록을 확인하여 해당 표현이 목록에 있을 경우는 표현 문형도 함께 주석한다. 따라서 형태 단위로 분석하여 오류로 처리하는 동시에 ‘표현 문형’ 오류도 중복 주석한다. (☞ 3. 범주별 세부 오류 유형의 처리 예시-2) 오류 위치-(4) 표현 문형(PE) 참고)

<예> 내일은 비가 온 것(√올 것) 같아요.

☞ ‘(으)ㄴ 것’, ‘(으)ㄹ 것’은 표현 문형 목록에 포함되어 있기 때문에 이 경우에는 오류가 발생한 ‘온 것’을 하나의 덩어리 표현으로 처리하는 동시에 구성 요소인 관형사형 전성어미로도 분석하여 오류가 나타난 위치를 중복 주석한다. 즉, [오류 위치-표현 문형, 관형사형 전성어미], [오류 양상-대치] 오류로 주석한다.

은행에 저축한(√저축할) 겨우에는(√경우에는) 얼마정도 이익을 얻을지 미리 알아서 더 편할 것 같아요.

☞ ‘-(으)ㄹ 경우에는’를 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)

- 입력이나 전사 단계의 실수로 추정되는 오류는 원문을 찾아 확인한 후 처리한다.

<예> 주식에 대한 잘 알아야 되고 주식에 대해(✓대해)
 ☞ 학습자가 동일한 본문 내에서 ‘에 대해’를 정확하게 사용하였고 이 문장에서만 이러한 실수가 나타났다. 따라서 이는 입력 단계의 실수로 판단하고, 원문을 확인하여 학습자가 쓴 형태대로 수정하여 처리한다.

2. 오류의 범주

- 오류 위치는 모든 오류에 대해 1:1로 주석하고(분석이 불가능한 오류는 분석 불가능 표시) 이를 기본 주석이라고 한다. 오류 양상과 오류 층위는 관련 오류가 있는 경우에만 주석되며 이를 확장 주석이라 한다. 확장 주석의 경우, 한 형태에 2개 이상의 오류가 나타나면 중복 주석이 가능하다.

1) 분석 여부

- ‘분석 여부’의 판단은 오류로 식별되었을 때 이를 올바르게 교정하고 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.

영역	주석 표시	포함 범위	예시
분석 불가능	IMP	문맥 내에서 해석이 불가능한 경우	한국여자 좋좁하고(✓좋고? 조용하고?, IMP) 예쁘기 대문에(때문에, POS) 결혼(✓결혼, POS)하고 싶어요.

2) 오류 위치

- 기본 주석으로 [오류 위치]를 주석하는데 오류가 일어난 부분, 즉 오류가 발생한 위치의 품사에 대해 주석한다. 오류가 발생한 위치 표지로서 오류 어절의 품사에 초점을 둔다.³⁴⁾ 아래는 오류로 식별된 부분의 품사 위치를 표시한 주석 표지이다.
- 오류 위치는 기본적으로 형태 주석에 따라 처리한다. 형태 주석에서 <표준국어대사전>에 근거하여 형태소 분석을 하기 때문에, 오류 주석은 이에 입각하여 오류 위치를 주석한다.

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 나타난 오류 애기와(√아기와, CNNG, 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

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

위치	주석 표지	포함 범위	예시
			있다.
형용사	CVA	형용사 어휘의 사용에서 나타난 오류	불고기 먹기 때문에 기분이 기쁩니다(√ 좋습니다, CVA, REP).
보조 용언	CVX	보조용언 어휘의 사용에서 나타난 오류	다른 사람에게 아픈다운 모습을 보여 싶기(√ 주기, CVX, REP) 위하여 노력하세요.
지정사	CVC	지정사 어휘의 사용에서 나타난 오류	그러니까 저는 외모지상주의가 위험이라고(√ 위험하다고, CVC, REP)생각한다.
관형사	CMM	관형사 어휘의 사용에서 나타난 오류	그렇게 되면 어느(√ 어떤, CMM, REP) 사람은 돈이나 개인 정보를 잃어버릴 수도 있다.
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 가적을 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 저소득층(√ 저소득층, CXPN, MIF) 사람의 살기가 보증할 수 있다.
명사파생 접미사	CXSN	명사 파생 접미 사 어휘의 사용	두 번째(√ 번째, CXSN, MIF)에 간 곳이는

위치	주석 표지	포함 범위	예시
		에서 나타난 오류	경주였습니다.
	동사파생 접미사	CXSV 동사 파생 접미사 어휘의 사용에서 나타난 오류	그 꿈을 위해서 매일 운동해다(✓ 운동한다, CXSN, 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.
기능 어휘	주격 조사	FNP 주격조사의 형태, 의미, 사용 오류	그리고 여행이(✓ 을, FNP, REP) 너무 좋아합니다. 고시원에서 많이 학생(✓ 이, FNP, OM) 살았다.
	관형격 조사	FGP 관형격조사의 형태, 의미, 사용 오류	‘론딩리’의(✓ 를, FGP, REP) 소개합니다.
	목적격 조사	FOP 목적격조사의 형태, 의미, 사용 오류	그래서 저는 한국 사람하고 다른 외국 사람을(✓ 과, FOP, MIF/REP) 교류하고 싶습니다.
	부사격 조사	FAP 부사격조사의 형태, 의미, 사용 오류	차 안에(✓ 에서, FAP, REP) 잤어요.
	접속 조사	FJC 접속조사의 형태, 의미, 사용 오류	한국어 문법와(✓ 과, FJC, MIF) 중국어 문법이 비슷하지 않았다.

위치	주석 표지	포함 범위	예시
호격 조사	FVP	호격조사의 형태, 의미, 사용 오류	친구아(✓ 친구야, FVP MIF), 같이 가자.
인용격 조사	FQP	인용격조사의 형태, 의미, 사용 오류	내가 감사하다고 말한다는 게 ‘고맙다’다고(✓ 라고, FQP, REP) 했다.
보조사	FXP	보조사의 형태, 의미, 사용 오류	론딩리는 맛있는 음식은(✓ 이, FXP, REP) 많습니다.
연결 어미	FED	연결어미의 형태, 의미, 사용 오류	그리고 친구들과 같이 노래방 가고(✓ 가서, FED, REP) 노래를 부르고 싶습니다.
종결 어미	FFE	종결어미의 형태, 의미, 사용 오류	특히 아랫목에 정말 따뜻한다(✓ 따뜻하다, FFE, MIF).
선어말 어미	FPE	선어말어미의 형태, 의미, 사용 오류	내일부터 수업이 시작됐어요(✓ 시작돼요, FPE, REP).
명사형 전성 어미	FNE	명사형 전성어미의 형태, 의미, 사용 오류	우리 계획은 저녁을 먹기(✓ 먹은, FNE, REP) 후에 우리 만든 신분증을 가지고 갈 겁니다.
관형사형 전성어미	FAE	관형사형 전성어미의 형태, 의미, 사용 오류	중요하는(✓ 중요한, FAE, MIF) 것은 사람마다 다르다는 것을 인정하는 것이다.
구 단위 표현	PHE	구 단위 표현을 잘못 사용한 경우	앞으로는 얼마나 어려운 일이 생기는 때(✓ 생겨, PHE, REP)도 포기하지 않고 사소한 일에도

위치	주석 표지	포함 범위	예시
			최선을 다하겠다.
표현 문형	PE	보조 용언이나 여러 요소의 결합 구성으로 이루어진 표현 문형을 잘못 사용한 경우(제시 목록 참고)	신세대는 기상세대와 가치관이 달라서 세대 차이가 생기기 마련이다(생기기 마련이다, PE, MIF).

3) 오류 양상

- 확장 주석 영역에 [오류 양상]을 두어 표면적으로 드러난 오류의 모습을 분석한다.
- 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 발화에서 나타나야 할 형태가 빠져 있는 경우	한 나라의 중요한 것 (✓은, FXP, OM) 뒷 사람이 일을 잘한다.
첨가	ADD	완전한 발화에서 나타나지 말아야 할 형태가 쓰인 경우, 중복된 형태를 반복해서 사용한 경우	한국말은 동경에 있었을 때, 일년간 동안(✓일년 간, CNNG, ADD) 한국 YMCA에서 공부했습니다.
대치	REP	한국어에 없는 표현이나 학습자의 모국어를 사용한 경우	나는 cousin(✓사촌, CNNG, REP)한테 이야기했어요.
		어휘 의미나 용법을 잘못 이해하여 다른 어휘를 사	용서를 줄(✓할, CVV, REP) 수 있게

양상	주석 표지	포함 범위	예시
		용한 경우	
오형태	MIF	오형태 오류: 어휘나 문법의 조합 양상과 활용 형태가 잘못된 형태로 제시된 경우. 즉, 활용 또는 곡용을 잘못하여 다른 이형태를 사용한 경우와 의미적으로 전혀 관련이 없는 항목이 선택된 경우	이 시간은 별로 덤지 않고 시원하셔(√/시원해서, EC, MIF, MCJ) 숙제하기에 좋습니다.
		맞춤법 오류: 어휘나 문법을 사용함에 있어서 다른 어휘나 문법으로 대체하여 이해할 가능성이 없는 경우로 이형태 등과 관련이 없이 철자를 잘못 사용한 경우	우리는 피간했어요(√/피곤했어요, CNNG, MIF).

4) 오류 층위

- 오류 층위는 오류로 식별된 부분이 언어학적 측면에서 어느 영역의 오류 인지를 판단한 것으로 언어학적 층위에 따라 나눈 오류의 범주이다. 오류 층위는 교수자나 학습자들이 자주 활용할 일부 영역에만 주석하였다.
- 오류 층위는 원어절과 교정 어절 양쪽에서 주석한다. 즉, 학습자가 원어절에서 인식하지 못한 경우와 교정어절을 인식하지 못한 경우를 모두 고려하여 해당 층위를 주석한다.
- 발음 층위는 주로 구어에서의 발음 오류를 다루는 영역이다. 발음 층위에서는 음소, 음절, 음운 규칙에서 발음상에서 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	[구어] 생일(√/썩일, NNG, PP)

층위	주석 표지	포함 범위	예시
		예) 평음, 격음, 경음의 구분	선물로(션물로, NNG, PP) 갓(간, MAG, PP) 태어난 강아지를 한 마리 기릅니다 ³⁵⁾ .
음절	PS	음절 차원에서 발생하는 오류. 구어에서 음절의 발음을 정확하게 하지 못한 경우로 원래 음절보다 더 적게 혹은 많이 발음한 경우	[구어] 우리::= 우리나라도:: 마약(√마야크, NNG, PS),.. 어:: 판매::, 될 수 있..=있긴 한데::
음운 규칙	PC	연음규칙, 비음화 등 음운변동에 관한 오류. 구어에서 필수적 음운 규칙의 이탈 오류로 음운 규칙을 적용하지 않고 절음화하여 발음한 경우	[구어] 그래서:: 서울레(√서울에::, , NNG, PC)) 가고 싶어요.
원어식 발음	PN	원어식 발음으로 발생하는 외국어 오류	[구어] 예를 들면 자기 계발, 재미, 수업 등 그래서 아래 그래프(√그래프, NNG, PN)를 보며는,
중 간 발음(변이음 포함)	PA	변이음을 포함한 중간 발음으로 발생하는 오류.	[구어] 전공(√전공, NNG, PA) ('저'와 '조'의 중간발음)

- 형태 층위는 주로 어휘 오류를 다루는 영역이다. 형태 층위에서는 합성어, 파생어 등의 조어 과정에서 발생하는 오류와 어미의 활용, 조사의 사용 등에서 나타나는 오류를 주석하도록 한다.

층위		주석 표지	포함 범위	예시
형태	단 어 형성[합 성법]	MCP	합성어 생산 및 사용에서 오류	그리고 나는 계속 매일 지각했을 때 나는 번금도 내고 반성서(✓반성문, NNG, MIF, MDP)도 썼다.
	단 어 형성[파 생법]	MDV	접사의 과잉 적 용과 같은 파생 어 사용 오류	작년 방학 때는 LG 전자에서 통역사로 일한 경험이 있고 한국에 대한 사이트에서 번역사(✓번역가, NNG, MIF, MDV)로 일한 경험도 있기 때문에 <name>에서 일할 수 있는 자신을 가지고 있다.

- 35) 한국어 학습자의 구어 자료는 구어적 특성에 따른 변이형이 많을 뿐만 아니라 비모어 화자의 구어라는 자료의 특성에 의해 더욱 다양한 변이형이 발생한다. 이에 따라 구어 전사에서는 자동 형태 주석의 편이성을 고려하여 다음과 같이 () 안에 여러 가지 중간언어 발음을 포함한 학습자의 발화를 제시하는 방식을 취하였으며, 오류 주석 작업자들은 두 가지 표기를 모두 보면서 오류 주석을 하게 된다.

예) 음 생일(뽕일) 선물(썸물로) 갓(간) 태어난 강아지를 한 마리를 기릅니다.

한편, 구어에서의 발음을 전사자가 문자로 완전하게 표기하는 데에는 여러 가지 한계가 있으므로, 다음과 같이 한국어 모어 화자의 발음과는 다르나 한국어의 음운 체계로 표기가 불가능한 발음의 경우 전사자의 설명을 달아 오류 주석 시 참고할 수 있도록 하였다.

예) 열한(여한) 살<note>살을 [sar]로 발음</note> 나이답지 않게

또한 오류 주석 작업자들은 전사자들이 기록한 음성 정보 외에도 필요 시 실제 음성 자료를 찾아 들을 수 있도록 해당 파일을 공유하였으며, 향후 구축 도구 내에서 음성 파일이 동기화되어 해당 부분을 간편하게 들을 수 있도록 구축 도구의 기능 개선 작업을 진행 중이다.

층위	주석 표지	포함 범위	예시
굴절 [곡용]	MDC	조사의 사용에서 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, EC, MIF, MCJ) 좋겠다
품사	POS	품사 사용에서 나타나는 오류	주말에 친구하고 같이 유명의를(√유명한, NNG, REP, POS) 곳이 가고 싶습니다.

- 통사 층위는 문법 오류를 다루는 영역이다. 높임, 시제, 사동, 피동, 부정 등의 문법 범주에서 발생하는 오류가 그 대상이 된다.
- 오류 층위에서 통사 층위는 문법범주와 관련되어, 해당 문법범주를 알지 못해서 오류가 발생했을 경우 주석한다.

층위		주석 표지	포함 범위	예시
통사	높임	SH	조사, 선어말어미, 종결어미 부분의 높임 관련 문법 형태소, 높임 어휘의 사용이 잘못된 경우	할머니께서 우유를 마십니다(√드립니다 CVV, REP, SH).
	시제	ST	시제 또는 시상을 나타내기 위한 선어말어미의 사용이 잘못된 경우	어제부터 항상 시계를 확인하기로 한다(√했다, FPE, REP, ST).
	사동	SC	사동사, 사동표현의 사용, 사동문 생성에서 발생한	갈릴레이는 새로 발명된 망원경을 사용하여 연구를 깊였다. (√깊게

층위	주석 표지	포함 범위	예시
		오류	하였다, CVA, MIF/REP, SC)
피동	SP	피동사, 피동표현의 사용, 피동문 생성에서 발생한 오류	교실 문이 닫아(√ 닫혀, CVV, REP, SP) 있었습니다.
부정	SN	부정 표현의 사용, 부정문의 생성에서 발생한 오류	우리는 항상 같이 있고 사우지 않해요. (√ 싸우지 않아요, PE, MIF, SN)
어순	WO	한국어의 통사 구조에 맞지 않는 방식으로 문장 전체 또는 일부가 배열된 경우	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 지시사 나는 롯데월드

층위	주석 표지	포함 범위	예시
		의 선택으로 선행문과 후행문의 관계를 결속성 있게 나타내지 못한 경우	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(MAG, ADD, DM) 10분쯤 걸렸어요..
구어/문어	DS	구어체/문어체, 격식체/비격식체의 혼용으로 인한 오류	근데(✓그런데, MAJ, 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 생각하
고 귀가 아주 밝다(√IMP).

- 분석 여부는 기본 값이 ‘분석 가능’으로 설정되어 있으므로, 오류임에 분명
하지만 교정어절을 주기 어려워 오류의 판정이 불가능한 경우에만 주석을
한다.

2) 오류 위치

(1) 실질어휘

① 고유명사(CNNP)

- 고유명사 어휘의 형태, 의미, 사용 오류를 말한다.

<예> 중구(√중국) 요리를 맛있었습니다.
하지만 독요(√도쿄) 쇼핑은 조금 비싸요.

② 일반명사(CNNG)

- 일반명사 어휘의 형태, 의미, 사용 오류를 말한다.

<예> 애기와(√아기와) 노인들한테 건강이 나빠졌다.
현대(√현재) 세계적으로 환경문제가 대두되고 있다.

③ 의존명사(CNNB)

- 의존명사 어휘의 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)

- 대명사 어휘의 형태, 의미, 사용 오류를 말한다.

<예> 내(√우리) 아버지가 남편하고 친하게 되면 좋겠다.
저(√나)는 유학생으로 온 외국인이다.

⑤ 수사(CNR)

- 수사 어휘의 형태, 의미, 사용 오류를 말한다.

<예> 이 셋(√세) 가지 단어의 뜻에 따라 이 외모지상주의라는 말을 충분히 이해할 수 있다.
이 세(√삼) 번, 문제가 뭐예요?

⑥ 동사(CVV)

- 동사 어휘의 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)

- 형용사 어휘의 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> 그리고 만약 한국에서 살다가 보면(√살다가) 좋아하는 남자가 사귀어서 같이 여행을 가고 싶다.

단, 보조용언과 결합된 구성이 표현 문형 목록에 있을 경우에는 표현 문형 위치로도 중복 주석한다.

<예> 세대 차이를 극복하기 위해서 신세대와 가상세대는 자주 이야기를 해 줘야 된다(√ 된다)
 ☞ ‘된다’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> 그러니까 저는 외모지상주의가 위험이라고(√ 위험하다고)생각한다.
 매일 저는 한국어를 공부합니다(√ 공부합니다).
 ☞ 지정사 ‘이다’가 접사 ‘하다’의 위치에 잘못 결합되었으므로 지정사 대치 오류로 처리한다.

<예> 다른 사람들에게 도와주고 싶기 때문에 한국어를 열심히 공부할 것이다(√ 것이다).
 ☞ 지정사 ‘이다’가 생략되었으므로 지정사 오류로 처리한다.

○ 문어에서 지정사를 축약해서 사용한 경우에는 오류로 보기 어려운 측면이 있으나 모어 화자의 보편적인 언어 사용 방식에 있어서 어색한 것으로 보고 지정사 오류로 처리한다. 이때에는 오류 양상은 주석하지 않고, [오류 위치-지정사], [오류 층위-문어/구어] 오류로 처리한다.

<예> 여기는 우리 학콘데(√ 학교인데) 정말 아름답다.
 ☞ 문어(격식체)에서 축약형으로 사용하는 것은 어색하기 때문에 [문어/구어] 오류로 처리한다. 단, 이때에는 오류 양상은 주석하지 않는다.

⑩ 관형사(CMM)

- 관형사 어휘의 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)

- 부사 어휘의 형태, 의미, 사용 오류를 말한다.

<예> 내 남편은 꼭(√정말) 멋있게 생겼다.
어히려(√오히려) 남성보다 여성의 힘이 더 강하는 경우도 있는 정도다.

⑫ 접속부사(CMAJ)

- 접속부사 어휘의 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)

- 감탄사 어휘의 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 부

모님도 신세대 입장 색학하면 세대 차이를 줄일 수 있다.

⑮ 명사파생접미사(CXSN)

- 명사파생접미사의 형태, 의미, 사용 오류를 말한다.

<예> 저의 첫 번째(√번째) 고민은 어떻게 시간을 잘 지킨 좋은 습관은 될 수 있는 것이다.
여성들의 사회 진출에 따라서 이혼률(√이혼율)이 높아진 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 행복한(√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.

<예> 나는(√내가) 이렇게 된 이유는 가족 때문이다.
☞ [오류 위치-보조사], [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 안
됐다.
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 "수협 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 활용 오류로 처리한다. 그러나 이형태가 없는 어미인데도(‘-고, -지만’) 연결어미 앞에 ‘아/어’나 ‘으’와 같은 매개요소를 사용했을 경우도 연결 어미의 활용 오류로 처리한다.

<예> 해고(√하고)
☞ [오류 위치-동사], [오류 양상-오형태], [오류 층위-활용]
와고(√오고)
☞ [오류 위치-연결어미], [오류 양상-오형태], [오류 층위-활용]
예쁘니까(√예쁘니까) : 오류 위치: ‘예쁘다’
☞ [오류 위치-형용사], [오류 양상-오형태], [오류 층위-활용]
예뻐으니까(√예쁘니까)
☞ 그러나 ‘예뻐으니까’와 같이 형용사와 연결어미 양쪽 모두 잘못 쓴 경우에는 ‘예쁘다’ 형용사와 연결어미 ‘-으니까’ 두 곳 모두 오형태 활용 오류로 처리한다. 즉, [오류 위치-형용사, 연결어미], [오류 양상-오형태], [오류 층위-활용]

⑪ 종결어미(FFE)

- 종결어미의 형태, 의미, 사용과 관련된 오류를 말한다.
- 종결어미 기본형 “-니다/다”는 오형태 활용 오류로 처리한다.
- 용언의 기본형 사용은 시제와도 관련되므로 ‘니다’가 현재를 나타낼 경우에는 오류 층위에 활용(MCJ)과 시제(ST)를 중복 주석한다.
- 종결어미 이형태 활용 오류는 오형태 오류로 처리한다.

<예> 그래서 일본에서 웃어른은 노약자석에서 꼭 앉는다(√앉는다).
그때부터 시간을 지키다(√지킨다).

⑫ 선어말어미(FPE)

- 선어말어미의 형태, 의미, 사용과 관련된 오류를 말한다.

<예> 인종 차별 때문에 많은 사람들이 죽었으니까 앞으로 인종 차별이 없었으면 좋게다(√좋겠다).
그래서 너무 배가 고파다(√고팠다).

⑬ 명사형 전성어미(FNE)

- 명사형 전성어미의 형태, 의미, 사용과 관련된 오류를 말한다.

<예> 과학자들의 의식에 의하면 아이들은 어른들보다 배우기(√배우는) 능력이 6배 뛰어나다고 한다.
언어를 배울 때는 언어만 말고 그 나라의 문화도 공부하기(√공부하는 것) 중요하다.

⑭ 관형사형 전성어미(CIC)

- 관형사형 전성어미의 형태, 의미, 사용과 관련된 오류를 말한다. 관형사형 어미의 이형태 선택과 관련된 오류가 포함된다.

<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 관련되어 오류 층위에 시제(ST)도 주석한다.

(3) 구 단위 표현(PHE)

- 구 단위 표현은 어절 단위로 이루어진 표현을 잘못 사용한 경우를 말한다. 교정 어절 주석은 구 단위 표현 단위를 묶어서 처리한다.

<예> 남자의 아내가 더 이상 돈 없는 힘든 인생을 살고 싶지 않아 오랜 고민과 망설임을 한 나머지(√ 끝에) 남편과 5살 어린 아이를 두고 더 좋은 인생을 찾으러 다른 도시로 이사를 간다.

- 구 단위 표현에 연어 오류가 있을 경우, 별도 처리하지 않고 동사 대치로 처리한다(본 연구는 연어 오류를 주석하지 않는다).

<예> 태도를 키워야 한다 (√ 길러야 한다)
 ※ 논의: ‘태도를 키워야 한다’를 ‘태도를 길러야 한다’라는 연어 오류를 어떻게 할 것인가? 처리하는 데 있어서, 구 단위 또는 연어 표현에서 용인 가능성 기준이 판단자마다 다를 수 있다는 문제가 있다.
 ☞ 연어 주석은 연어 목록이 선행되어야 하고 연어에 대한 판단이 정해져야 하므로 본 연구에서는 주석하지 않는다. 따라서 연어 오류의 경우에는 동사 대치로 처리한다. 다음과 같이 ‘이사를 옮기다’의 경우, ‘짐을 옮기다’, ‘이사를 가다’로 수정할 수 있다. 교정 어절을 주고 [오류 위치-동사], [오류 양상-대치]로 처리한다.

<예> 이상한 날씨로 악화가 나타났다(√악화가 되었다).
 ☞ 구 단위 오류로 볼 것인지, ‘나타나다’ 동사의 대치 오류로

불 것인가가 문제가 되는데, 이 경우 가능한 분리하여 ‘나 타났다’를 ‘되다’로 교정하여 [오류 위치-동사], [오류 양상-대치]로 처리한다.

(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.³⁶⁾

<예> 무슨 파티하면 우리 학생들이 열심히 공부한 연세대학교 열심히 공부해서 조금 피곤한 것인데요(√ 피곤해요).

<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’이 표현 문형 목록에 있기 때문에 표현 문형 누락 오류로 처리한다.

3) 오류 양상

(1) 누락(OM)

주 누락 오류는 완전한 발화에서 나타나야 할 형태가 빠져 있는 경우를 말한다. 누락 오류는 필수적인 성분이 생략됐을 경우에만 처리한다. 따라서 누락 오류의 경우는 조사나 어미 누락 정도를 잡아주는 것으로 한다. 필수적인 성분이 아닌 수의적이거나 없는 정보를 더 넣어주어서는 안 된다.

36) 표현 문형 목록 기준을 모든 표현 문형의 합집합으로 할 경우, 다양한 것이 표현 문형 안으로 들어오기 때문에 어느 정도 정제한 것으로서 <한국어 문법 사전>(국립국어원, 2005) 목록을 기준으로 정한다. 표현 문형 목록은 <부록>으로 첨부하였다.

<예> 저는 여러 까지(√가지) 능력서(√자격증)가 취득하지만 그
중에서 영어를 능수하는(√능숙한) 정도입니다.
※ ‘그 중에서 영어를 가장 능숙한 편입니다’라고 교정하여, ‘
가장’이라는 부사를 첨가하고, ‘능숙한 편입니다’라고 교정
해도 되는가?
☞ 필수적인 성분이 아닌 것을 추가하여 [누락] 오류로 처리
해서는 않는다.
☞ 최소한의 교정 원칙에 따라 ‘능숙한 정도입니다’를 교정
어절로 삼는다. 수의적인 것은 [누락]으로 처리하지는 않
고, 필수적인 성분이 생략됐을 경우에만 [누락] 오류로 처
리한다.

- 누락은 교정 예측 단위이다. 유의미한 교정 어절 하나(어절 단위)에만 [누락]으로 한 번에 처리하도록 한다. 누락의 경우, 오류 위치는 교정 어절이
기준이 된다. 따라서 교정 어절을 입력하고, 누락된 품사를 오류 위치로
주석하여 [오류 위치-조사], [오류 양상-누락] 또는 [오류 위치-동사], [오
류 양상-누락] 등으로 처리한다.

<예> 돈이 많은 사람들(√은) 투자 할 수도 있어요
우리는 함께 (√있을) 때 좋은 기본이 왔는데요.
☞ ‘있을’이 누락되었는데, 형태주석을 기본단위로 오류주석을
할 때, 동사 ‘있’과 관형사형 전성어미 ‘을’을 각각 누락오
류로 처리할 수 있다. 그러나 동사 뒤에 오는 관형사형 어
미는 동사에 의해 따라오는 것으로 판단하여, 동사 ‘있’ 하
나만을 누락 오류로 처리한다.

- 누락된 품사(교정 어절) 위치를 오류 위치로 주석하고, 오류 양상을 누락
(OM)으로 주석한다. 일부 예시는 다음과 같다.
- [명사 누락]은 다음과 같은 것들이 해당된다.

<예> 저는 하고 싶은 일이 있습니다. 여행사, 호텔, 하우스 (√등)
많이 있습니다.

☞ 나열 맨 마지막에 ‘등’이 [누락]된 것으로 처리한다. [오류 위치-명사], [오류 양상-누락] 오류로 처리한다.

○ [조사 누락]은 다음과 같은 것들이 해당된다.

<예> 그로 인해 평소 일상생활에서 말할 수 없는 말 욕하는 말 등 비우는 말 등(√ 을) 흔히 볼 수 있다.
 ☞ ‘등’을 써줬기 때문에 목적격 조사 ‘을’을 사용하지 않아도 된다고 용인할 수도 있으나, 조사 누락은 엄격하게 적용하여 누락 오류로 처리한다. 이 경우는 ‘등’ 맨 마지막에만 [오류 위치-목적격조사], [오류 양상-누락] 오류로 처리한다.

○ [관형사형 전성어미 누락]은 다음과 같은 것들이 해당된다.

<예> 10년 후에 내가 가(√ 갈) 수 있다.
 ☞ 관형형 어미 오류의 경우 오형태 오류인지 누락 오류인지 헷갈릴 수 있으나 이때에는 관형형 어미 ‘-(으)ㄴ’의 [누락]으로 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 ‘ㅁ’이 누락되었으나, 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.

○ 시제를 나타내는 선어말어미 ‘-었-’이나 ‘-겠-’을 써야할 자리에 선어말어미가 생략된 경우 기본적으로 대치 오류로 처리하나, 동사의 기본형을 사

용한 경우에만 누락으로 처리한다. (자세한 사항은 4) 오류 층위 (3) 통사 시제 참고).

<예> 나 한국어를 1년 만에 배우다(√배웠다).

(2) 첨가(ADD)

- 첨가 오류는 완전한 발화에서 나타나지 말아야 할 형태가 쓰인 경우나 중복된 형태를 사용한 경우를 말한다.

<예> 종일에(√종일) 반 친구들과 나는 만나서 언제나 재미있는 시간을 하고 있어요.
그리고 저녁은 기숙사에서 식사를 하니까 7시까지(√까지) 공부를 끝을 내서 친구들이랑 헤어져요.
머지않은 장래에 장래(√장래) 인간 복제도 가능하게 될 것이다.

- 첨가된 부분은 해당 부분의 교정 어절 없이 첨가된 위치를 오류 위치로 주석하고, 오류 양상을 첨가(ADD)로 주석한다. 일부 예시는 다음과 같다.
- [조사 첨가]는 다음과 같은 것들이 해당된다.

<예> 대만의 수도가(√수도) 타이베이
☞ 이는 ‘수도인’으로 교정할 경우에는 대치 오류로 볼 수도 있으나 최소 수정 원칙으로 주격조사 ‘가’가 첨가된 오류로 처리한다.

- [표현 문형 첨가]는 다음과 같은 것들이 해당된다.

<예> 저는 <name> 한국어센터에서 공부하고 있는 동안 선생님한테서 도움이 많이 받아 공부하고 있는(√공부하는) 시간이 아주 즐거웠습니다.
☞ ‘-고 있’이 첨가된 오류로 [오류 위치-표현 문형], [오류 양상-첨가] 오류로 처리한다.

(3) 대치(REP)

- 대치 오류는 의미적 오류로 서로 다른 의미의 어휘를 바꾸어 쓴 경우를 말한다. 즉, 학습자가 어휘나 문법 요소의 의미나 용법을 잘못 이해하여 다른 어휘나 문법 요소를 사용한 경우이다.

<예> 그래서 한국 사람들이 토시야는 일본보다 한국이 맞춰보일(√어울릴) 거라고 합니다.
전통의 아름다움이 사람들에게 알려주는 것도 전통을 보존하려고(√보존하려면) 해야 할 일이다.
물론 계획표대로 일을 하다가 보니(√하더라도) 일을 잘 할 때도 있을 잘못 할 때도 있어요.
직접 비판을 받을 때보다 상처가 더 많은(√큰) 것이다.

- 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.

- 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 주석 처리한다.

- 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> 누구가 돈이 없다면 행복할 수 없다고 생각했는데 한편에
 일부 사람은 가족이 가장 중용 생각했는데 그 이유 점은
 돈으로 바뀔(√바꿀) 수 없다고 지적을 했다.
 ☞ ‘바꾸다’를 써야 하는데, 피동형을 사용하였기 때문에 동사
 대치 피동 오류로 처리한다.

- 대치되어야 할 품사를 오류 위치로 주석하고, 오류 양상을 대치(REP)로 주석한다. 일부 예시는 다음과 같다.
- [대명사 대치]는 다음과 같은 것들이 해당된다.

<예> 그들은 환경문제보다 자기가(√자신의) 먹고 살기가 더 중요
 하다고 생각한다.

- [조사 대치]는 다음과 같은 것들이 해당된다.

<예> 한국 영화를(√영화는) 재미있습니다.
 ☞ 보조사 ‘는’ 자리에 목적격 조사 ‘를’을 잘못 사용한 경우이
 므로 [오류 위치-보조사], [오류 양상-대치] 오류로 처리
 한다.

- [연결어미 대치]는 다음과 같은 것들이 해당된다.

<예> 저는 지난 주말에 날씨가 좋니까(√좋아서) 기숙사 친구하고
 같이 한강공원에 갔습니다.
 ☞ ‘-아서/어서’를 사용해야 할 곳에 ‘-니까’를 잘못 사용하였
 으므로 [오류 위치-연결어미], [오류 양상-대치, 오형태(‘-
 니까’의 형태도 잘못 사용)] 오류로 처리한다.

- ※ 주의: 대치와 오형태 오류를 판단할 때, ① 문맥으로 오류를 판단한다.
 ② 학습자들이 얼마나 이러한 오류를 보일 수 있는가를 생각한다. 즉, 수
 준별로 사용할 수 있는 어휘를 고려하여 오류를 처리한다. ③ 최소 수정

원칙을 기본으로 하되, 용인 가능한 교정 어절로 수정해야 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 처리한다.

(4) 오형태(MIF)

- 오형태 오류는 어휘나 문법의 조합 양상과 활용 형태가 잘못된 형태로 제시된 경우를 말한다. 즉, 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.
- 또한 오형태 오류는 음소 단위 형태를 잘못 쓴 오철자 오류와 용언 활용, 조사 이형태 곡용, 어미 활용 등 형태를 잘못 활용한 경우를 포함한다. 따라서 용언의 규칙/불규칙 활용과 조사 및 어미 이형태에 따라 활용 오류를 처리하고, 활용 양상에 따라 오류 위치를 판단한다. 단순 오철자 오류의 경우 오형태(MIF)만 처리하며, 오류 층위에 활용(MCJ)을 주석하지 않는다.

<예> 어떤 초대하고(√초대하고) 싶은 사람,
 멜버른에 어떻게(√어떻게) 가는지 알아요?
 우리나라에서 과일들 하고 야채들도 많아서 과일와(√과일
 과) 야채도 다른 나라에 팔아요.
 지금 친구 같이 등산에 가시다(√갑니다).

<예> 저녁에 커피를 마시면서 간단한(√/간단한) 책을 읽고 싶다.
 ☞ 최소 교정을 원칙으로 하여, 이 경우 ‘간단한’ 책이라고 하면 언어적으로 어색하고, 말이 안 된다고 생각할 수도 있지만 ‘가벼운’의 의미로 모어 화자들도 사용한다고 볼 수도 있다. 따라서 최소 교정 원칙을 따라 ‘간단한’의 오철자로 보고, [오형태] 오류로 처리한다.

- 오철자 및 잘못된 활용이 나타난 부분을 오류 위치로 주석하고, 오류 양상을 오형태(MIF)로 주석한다. 단순한 오철자 외에 활용 오류의 경우, 조사 이형태를 잘못 사용한 경우는 오류 층위에서 굴절(곡용)(MDC)으로 주석하고, 어미 활용을 잘못된 경우는 굴절(활용)(MCJ)으로 주석한다. 일부 예시는 다음과 같다.
- [명사 오형태]는 다음과 같은 것들이 해당된다.

<예> 우리 집에 물을(√/문을) 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 처리한다.

- [조사 오형태]는 다음과 같은 것들이 해당된다.

<예> 20, 30대 남녀는 친구을(√친구를) 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 처리한다.

- [선어말어미 오형태]는 다음과 같은 것들이 해당된다.

<예> 제주 친구하고 옥등산에서 등산을 가세요(√ 갔어요).
 ☞ 과거 시제를 나타내는 선어말 어미 ‘-았-’이 생략된 형태이다. 그러나 ‘가어요’로 쓰지 않고 ‘가세요’로 썼기 때문에 이것은 과거를 인식하고 있다고 보고 오형태 오류로 처리한다. 중국인 학습자의 경우 ‘해습니다’, ‘마셔지만’, ‘와지만’ 으로 잘못 사용하는 경우가 많다. 이러한 예들은 [오류 위치-선어말어미], [오류 양상-오형태], [오류 층위-활용] 오류로 처리한다.

- [연결어미 오형태]는 다음과 같은 것들이 해당된다.

<예> 갈 수 있는다면(√ 있다면) 언제까지도 기다린다”고 해서 희망자들이 속출하고 있다.
 ☞ ‘있다면’을 써야할 자리에 ‘있는다면’으로 연결어미를 잘못 활용하여 쓴 것이기 때문에 [오류 위치-연결어미], [오류 양상-오형태], [오류 층위-활용] 오류로 처리한다.

- [종결어미 오형태]는 다음과 같은 것들이 해당된다.

<예> 그 이유는 제가 우라 아내보다 한국에 돈을 잘 못 벌읍니다(√ 법니다).
 ☞ 동사 ‘벌다’를 활용하여 ‘법니다’로 써야하는데 ‘벌’을 그대로 사용하고 있기 때문에 [오류 위치-동사], [오류 양상-오형태], [오류 층위-활용] 오류로 처리한다.

- 지정사와 종결어미가 결합할 때, 지정사를 누락시키거나 첨가했을 경우 또는 축약을 잘못된 경우는 오철자 오류로 처리한다.

<예> 남자이에요(√ 남자예요)
 학생예요(√ 학생이에요)
 학생예요(√ 학생이에요)

학생이에요(√학생이에요)

아니예요(√아니예요)

☞ 지정사와 관련된 오류에서 지정사를 쓰고 어미와 축약하지 않거나 잘못 축약을 시킨 경우는 오철자 오류로 처리한다. 또한, 지정사를 누락하거나 첨가했을 경우도 발음의 영향으로 인한 오철자 오류로 판단하여 오철자 오류로 처리한다. 활용 오류로 처리하지 않도록 주의한다.

4) 오류 층위

(1) 발음

① 음소(PP)

○ 음소 오류는 음소 단위에서 발화가 잘못 사용된 경우를 말한다.

<예> 빨리(√팔이) 말을::, 즐= 아:: 잘해서::,

☞ 자음 ‘ㅁ’을 ‘ㅍ’로 발음하여, 음소 오류로 주석한다.

예:: 커피(√코피)= 카페에서::, 예:: 공부를, 합니다::.

☞ 모음 ‘아’와 ‘오’를 교체하여 발음하므로 음소 오류로 주석한다.

○ 명확하게 다른 음소로 발음한 경우와 단어 내에서 음소를 발음하지 못한 경우, 음소 오류로 주석한다.

<예> 선생님(√성생닌)

☞ 받침을 모두 다른 음소로 발음한 음소 오류로 처리한다.

부모님(√부모니)

☞ 한 단어 안에서 받침을 발음하지 못한 경우도 마찬가지로 음소 오류로 주석한다.

② 음절(PS)

- 음절 오류는 음절 단위에서 발화가 잘못 사용된 경우를 말한다. 음절 오류는 원래 음절보다 더 많게 혹은 더 적게 발화한 경우가 해당된다.

<예> 예:: 매일(√매이루) 노무::, 즐겁..습니다::.
제가, 한국에 와서.. 사 개월(√사 개워르) 정도:: 살았습니다::
☞ 2음절인 ‘개월’을 받침 ‘르’과 모음 ‘으’를 결합하여 3음절로 발음하고 있으므로 음절 오류로 주석한다.

③ 음운규칙(PC)

- 음운규칙 오류는 구어 발화에 나타난 필수적 음운 규칙의 일탈을 말한다. 유음화, 연음화하여 발음해야 하는데, 글자 그대로 절음화하여 발화한 경우가 해당된다.

<예> [설랄](√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 경우, ‘학교’를 [학교]로 발음하지 않고 글자 그대로 [학교]라고 발음한 경우를 음운규칙 오류로 처리한다.

- 음운규칙 오류는 음운규칙을 적용하지 않은 경우와 적용했으나 잘못 발음한 경우 2가지로 나눌 수 있다.
- 음운규칙 미적용 오류는 음운규칙으로 인해 철자와 다르게 발음해야 하나, 학습자가 철자대로 절음화 하여 발음한 경우다. 학습자가 철자대로 발음했기 때문에 원어절과 교정어절의 형태는 동일하다.

<예> 무조건(√무조건) ☞ 경음화 미적용
같이(√같이) ☞ 구개음화 미적용
신라(√신라) ☞ 유음화 미적용
앞에(√앞에) ☞ 연음 미적용

먹는(√먹는) ⇨ 비음화 미적용

- 음운규칙 미적용 외에 음운규칙을 적용시켜야 하는 단어에서 잘못 적용한 경우도 음운규칙 오류로 주석한다. 그러나 이때에는 음운규칙과 음소 오류를 중복 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. 학교(√핵교).

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

<예> 표현 의도에 의한 수의적 경음화: 조금

구어에 의한 발음 특성: ~먹었구요 / ~했어여

④ 원어식 발음(PN)

- 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.

<예> 인터뷰(√인털뷰)

센터(√세너)

⇨ 한국어 외래어 표기법과 다르게 원어식 발음으로 발화한 경우 원어식 발음 오류로 처리한다.

- 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> 파트너(√ 팔너)
베트남(√ 뵤남)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> 요하네스버그(√ 요한스버그)
이.. 기자는:: 한국에서:: 이:: 마약, 없는::, 아:: 이미지(√ 이페지) 줌:: 없어,지고:: 있습니다 지금.
☞ 외국어 발음 오류로 외래어 표기법과 ‘다르다’는 차원에서 [오류 양상-오형태] 오류로 처리한다.

⑤ 중간 발음(변이음포함) (PA)

- 중간 발음은 변이음을 포함하여 학습자의 외국어와 한국어의 중간 발음으로 인한 발화 오류를 말한다.
- 중간 발음은 한국어 모어 화자와는 다른 발음, 변이음과 관련된 오류로 구어 전사 과정에서 유성음, 무성음으로 표기해준 경우에 근거해 변이음 오류로 처리한다. 이때는 음성과 관련된 문제로, 원어절과 교정어절의 형태는 동일하다.

<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.

- 변이음 외에 학습자가 원어절과 교정어절 사이의 발음, 즉 음소 간의 중간 형태로 발음한 경우도 중간발음 오류로 처리한다.

<예> 여자(√ 요자)

☞ ‘ㄷ’의 ‘교’의 중간 발음

과반수(√ 화반수)

☞ ‘ㅎ’과 ‘ㄱ’의 중간 발음

☞ 구어 전사에서 위와 같이 ‘음소’와 ‘음소’의 중간 발음으로 기술한 메모에 근거해 중간 발음(PA) 오류로 주석한다.

- 구어전사에서 변이음(음성 대치)이 분명하게 식별될 경우에만 ‘중간 발음’으로 오류 주석한다. 변이음 식별이 분명하지 않은 경우에는 주석하지 않는다.

<예> 여자가<note>‘여’를 ‘으’와 ‘유’의 중간 발음으로 발음</note>

☞ 구어 전사 시, 중간 발음으로 들릴 때 괄호 안에 (‘X’와 ‘Y’와 중간 발음)으로 표기한다. 이를 바탕으로 하여 중간 발음으로 표기 된 경우, 오류 층위에서 중간 발음 오류로 처리한다.

(2) 형태

① 단어 형성[합성법](MCP)

- 단어 형성[합성법] 오류는 조어 과정에서 발생하는 오류를 말한다. 즉, 학습자가 존재하지 않는 어휘를 생산해 내는 오류가 포함된다.
- 학습자가 조어 과정에서 형태를 잘못 만들어 낸 경우, 오형태 오류로 볼 수도 있다. 그러나 오형태 오류가 오철자 오류와 이형태 활용 오류에 해당하는 오류 양상일 때, 합성과 파생 관련한 오류를 형태 오류에 포함시킬 수 있는가가 문제가 된다. 본 연구에서는 파생과 합성 오류의 경우 오형태 오류로 보기 어렵고, 오류 양상은 수의적 주석이므로 오류 양상을 필수적으로 주석하지 않고 오류 층위에서 파생과 합성만을 주석하도록 한다.
- 따라서 학습자가 생산해 낸 형태가 한국어에는 없는 합성어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 합성으로 주석한다.

<예> 해물 고기(√물고기)가 많 많았어요.
☞ ‘물고기’를 ‘해물’과 ‘고기’로 잘못 합성하였으므로 [오류 위치-명사], [오류 양상-없음], [오류 층위-합성법]으로 처리한다.

② 단어 형성[파생법](MDV)

- 단어 형성[파생법] 오류는 조어 과정에서 접사를 잘못 사용한 오류를 말한다. 접사의 과잉 적용과 같은 경우가 포함된다.
- 학습자가 파생접사(유사파생접사 포함)를 사용하여 생산해 낸 형태가 한국어에는 없는 파생어일 경우, 오류 양상은 주석하지 않고 오류 층위에서 파생으로 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 형태-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 오류초 처리한다. [오류 위치-접미사], [오류 양상-대치], [오류 층위-파생법]으로 주석한다.

이런 데다가 의사 선생님이 의하면 균형 깨진 영양성(✓영양) 바람에 났던 여드름이 더 날 계속한다고 걱정했는데도 그렇지 않습니다.

☞ 접미사 ‘-성’을 과잉적용한 오류로 [오류 위치-접미사], [오류 양상-첨가], [오류 층위-파생법]으로 주석한다.

- 동사 어간에 ‘하다’를 붙여 한국어에는 없는 동사를 만들어냈을 경우는 파생 오류로 처리한다. 이는 합성 오류로도 볼 수 있지만 형태 주석에서 ‘-하다’를 파생접미사로 처리하고 있어 처리의 연계성과 일관성을 고려하여 파생 오류로 처리한다.

<예> 그날 수업 후 집에 도착하자마자 어머니가 나한테 혼했다(혼냈다).

☞ 이는 ‘혼내다’를 명사 ‘혼’에 ‘하다’를 붙여 ‘혼+하다’로 한국어에는 없는 어휘를 생산한 것이다. 이 경우 ‘하다’를 동사로 볼 수도 있고 파생접사로도 볼 수 있다. 형태 주석에서는 이를 동사파생접미사로 주석하기 때문에 오류 주석에서도 파생 오류로 판단하도록 한다. [오류 위치-동사], [오류 양상-없음], [오류 층위-파생법]으로 처리한다.

- 접사는 문법범주에서 논의하는 존대, 피동/사동, 복수 표지 중 피동/사동만 대치 오류로 처리한다. 접사 중 문법적인 성격 강한 존대(님), 복수 표지(들) 등은 형태 주석에서 접사로 따로 분리하여 처리하기 때문에 형태

주석에서 분리하는 접사들은 ‘누락/첨가’로 처리하고, 피동/사동은 어휘 대치로 처리한다. 또한 어휘적 의미를 더해주는 접사의 경우, 예를 들어 ‘사과’를 ‘꽃사과’로 썼을 때에는 형태 분석에서 ‘꽃’을 분리하지 않기 때문에 어휘적 의미를 더해주는 접사가 덧붙여진 단어는 대치 오류로 처리한다.

- 그밖에 형태 주석에서 접사로 분리하지는 않지만 유사파생접사로 볼 수 있는 형태들을 사용하여 어휘를 파생시킨 경우는 오류 양상은 주석하지 않고, 오류 층위에 파생으로 주석한다.

<예> 그리고 나는 계속 매일 지각했을 때 나는 번금도 내고 반성서(√반성문)도 썼다.

☞ 반성서에서 ‘서’는 형태주석에서 분리하여 처리하지 않고, 반성서를 하나의 명사로 주석한다. 이를 오류 주석에서는 반성문을 한국어 어휘에는 없는 ‘반성+서’로 파생시킨 것으로 보고 [오류 위치-명사], [오류 형태-없음], [오류 층위-파생법]으로 처리한다.

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 파생으로 주석한다.

③ 굴절[곡용](MDC)

- 굴절[곡용] 오류는 조사의 이형태를 잘못 사용한 경우를 말한다.
- 굴절[곡용] 오류의 오류 양상은 기본적으로 오형태 오류로 주석한다.

<예> 지금 가족가(√가족이) 너무 보고 싶습니다.

☞ 받침 뒤에서 주격조사 ‘가’로 잘못 사용하였으므로 [오류 위치-주격조사], [오류 양상-오형태], [오류 층위-곡용] 오류로 처리한다.

④ 굴절[활용](MCJ)

- 굴절[활용] 오류는 용언의 활용과 어미의 활용을 잘못된 경우를 말한다. 즉, 용언의 규칙 및 불규칙 활용, 어미의 이형태 오류가 포함된다. 따라서 용언의 규칙/불규칙 활용과 어미 이형태에 따라 활용 오류를 처리하고, 활용 양상에 따라 오류 위치를 판단한다. 단순 오철자 오류의 경우 오형태(MIF)만 처리하며, 오류 층위에 활용(MCJ)을 주석하지 않는다.
- 굴절[활용] 오류의 오류 양상은 기본적으로 오형태 오류로 주석한다.

<예> 미국에 영어 제일 중요하다(√중요하다).
 ≡ 형용사 ‘중요하다’를 ‘중요한다’로 잘못 활용하였으므로 [오류 위치-종결어미], [오류 양상-오형태], [오류 층위-활용] 오류로 처리한다.
 이렇게 살으면(√살면) 정말 행복할 수 있다.
 ≡ 받침 ‘ㄴ’ 뒤에서 ‘-으면’의 형태로 잘못 활용하였으므로 오형태 활용 오류로 처리한다.

- 어미 앞에서 ‘아/어’, ‘으’가 삽입된 경우 어미 활용과 관련된 것으로 보고 활용 오류로 처리한다.

<예> 2015년에는 친구하고 같이 많이 놀아고(√놀고) 싶습니다.
 저는 좋아하는 프로그램은 많아지만(√많지만) 다른 방송에 비해서 동물에 대한 다큐멘터리는 프로그램이 제일 좋아한다.
 ≡ 이형태가 없는 어미 ‘고, 지만’ 앞에 ‘아/어’ 또는 ‘으’가 첨가된 경우, 어간 활용과 어미 활용의 구분이 어렵다. 학습자들이 동사 어간에 ‘아/어’를 첨가해 하나의 어간으로 재구성했다고 볼 수도 있고, ‘아고, 아지만’의 형태로 어미를 잘못 활용한 것으로 볼 수도 있다. 본 연구에서는 이와 같은 경우도, 이형태의 어미의 영향을 받았다고 보고, 어미의 활용 오류로 처리한다.

- 어미 앞에서 ‘ㄴ, ㄷ, ㅂ’ 등의 불필요한 요소를 첨가했을 경우, 기본적으로 용언 어간의 오철자 오류로 처리한다. 단, ‘ㄴ, ㄷ’이 시제의 의미를 나타낸

것이라고 판단될 경우는 이들 요소를 분리하여 관형사형 어미(ETM) 첨가(ADD)로 처리하며, 층위에 시제(ST)를 주석한다.

<예> 왜냐하면 어렸을 때부터 커피숍이나 호텔의 사장님 될고(√
되고) 싶어 하기 때문이다.

☞ 연결어미 앞에 ‘ㄴ’ 요소가 첨가된 경우, 동사 오철자 오류로 처리한다. 그러나 ‘ㄴ’이 미래의 의미를 나타내고자 쓴 것으로 판단될 때에는 ‘ㄴ’을 관형사형 전성어미 첨가(ADD) 오류로 처리하고, 오류 층위에 시제(ST)를 줄 수 있다.

달른(√다른)

알랴다(√알았다)

☞ ‘ㄴ’ 앞에서 ‘ㄴ’이 첨가된 경우는 [ll] 발음의 영향으로 인한 오철자 오류로 처리한다. 따라서 ‘달른’은 형용사의 오철자, ‘알랴다’는 선어말 어미의 오철자 오류로 처리한다.

- ‘아서/어서’에서 ‘서’만 쓴 경우, ‘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.
- 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)

- 종결어미에서 ‘ㅂ니다/습니다’ 외에 ‘니다’ 또는 ‘입니다’를 잘못 사용한 경우도 종결어미 오형태 활용 오류로 처리한다.

<예> 힘들습니다(√힘듭니다)

아름답니다(√아름답습니다)
 시끄럽니다(√시끄럽습니다)
 재미있입니다(√재미있습니다)
 맛있입니다(√맛있습니다)

⑤ 품사(POS)

- 품사 오류는 동일 의미 단어의 품사를 잘못 사용한 경우를 말한다. 즉, 같은 의미인 단어의 품사를 제대로 인식하지 못하여 명사를 동사로 사용하거나, 부사를 형용사로 잘못 사용한 경우를 말한다.
- 오류 층위에서 품사에 해당하는 오류는 품사에서 나타난 오류(품사가 달라진 경우)와 품사를 몰라서 생겨난 오류(품사 혼동으로 인한 오류)로 구분할 수 있다. 이때, 품사가 달라진 것보다 품사를 모르고 있는 것을 우선적으로 적용하여, 품사에 대한 인식이 없어서 생겨난 오류만 품사 오류로 보았다. 따라서 문맥에 따라 교정 어절이 바뀌면서 품사가 달라진 경우(단순히 원어절과 교정어절의 품사가 상이한 경우)는 품사 오류로 처리하지 않는다.
- 즉, 원어절과 교정 어절이 의미를 공유하면서 품사를 제대로 사용하지 못한 경우, 품사에 대한 인식이 부재한 것으로 간주하고 이를 품사 오류로 처리한다.
- 품사 오류의 오류 양상은 기본적으로 대치 오류로 주석한다.

<예> 그래서 우리는 빠른(√빨리) 우리 집에 다가왔다.
 ☞ 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.

- 내 생가기는 한국 사람들 다른 나라 사람들보다 친절(√/친절한) 것 같다.
- 그리고 한국 친구를 인사고(√/인사하고) 싶습니다.
- 한국 공부가 너무 재미(√/재미있)기는 하지만, 단어 위우가기 힘들다.

- 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).

- ‘외국어’에 파생접사 ‘하다’를 결합하여 동사/형용사를 만들 수 있는데, 외국어 명사만 사용한 경우도 동일하게 품사 오류와 파생 오류로 중복 주석한다.

<예> 그리고 장애인의 사회참여를 스마트(√/스마트한) 사람이 필요합니다.

(3) 통사

① 높임(SH)

- 높임 오류는 조사, 선어말어미, 종결어미 부분의 높임 관련 문법 형태소, 높임 어휘의 사용이 잘못된 경우를 말한다.
- 높임법은 주체 높임법과 객체 높임법, 상대 높임법이 있으며, 이는 다시 문법적 높임과 어휘적 높임으로 나뉜다. 이 중 문법적 높임 오류는 주격 조사와 서술에서의 높임 호응 관계 불일치, 주체 높임을 나타내는 선어말어미 ‘-시-’의 잘못된 사용, 상대 높임을 나타내는 대명사와 종결어미의 호응 관계 불일치, 객체 높임을 나타내는 부사격조사 ‘께’의 잘못된 사용이 해당되며, 어휘적 높임 오류는 ‘계시다, 드리다, 모시다, 잡수시다, 주무시다’ 등의 특수한 높임말을 써야할 자리에 쓰지 않은 경우 또는 그 반대의 경우가 해당된다.
- 그러나 특수 어휘에 의해서 표현되는 높임법의 경우 존대와 겸양을 나타내는 특수 어휘가 다양하다. 높여야 할 대상인물을 직접적으로 높이는 어휘, 대상과 관계있는 것을 간접적으로 높이는 어휘, 객체를 높이는 동사 어휘, 접미사나 접두사가 붙어 존대나 겸양을 나타내는 어휘 등 다양하기 때문에 높임과 관계된 모든 어휘를 오류 주석 대상으로 삼기 어렵다. 이러한 이유로 본 연구에서는 어휘적 높임 오류보다 문법적 높임 오류를 우선 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.

- 문장 이상의 단위인 담화 상에서 높임말과 반말을 혼용하여 사용하고 있는 경우에는 문장 단위로 오류를 판단한다는 원칙하에 한 문장 내에서 높임 표현에 문제가 없으면 오류로 주석하지 않는다.
- 상대높임법에서 대명사와 종결어미 양쪽 모두 교정이 가능할 때에는 종결어미를 기준으로 대명사 오류로 일괄 처리한다. 즉, 종결어미에 따라 대명사 ‘나’와 ‘저’의 대치 오류로 우선 처리한다.

<예> 저는(√나는) 10년 후에 생활이 부유하고 싶다.
 ☞ 상대높임법 체계에서 해라체를 사용했는데, 대명사는 자신을 낮추는 ‘저’를 잘못 사용하였다. 따라서 상대높임법 체계에 맞추어 ‘나’로 교정한다. 반대로, 합쇼체를 사용하고, 대명사 ‘나’를 사용한 경우도 마찬가지로 [오류 위치-대명사], [오류 양상-대치], [오류 층위-높임]으로 주석한다.

- 높임을 나타내는 접사 ‘님’의 경우, 생산성이 강한 접사로 형태 주석에서 접미사로 따로 분석 처리한다. 이와 연계하여 오류 주석에서는 높임을 나타내는 접사 ‘님’의 과잉사용 또는 미사용의 경우 ‘첨가’와 ‘누락’ 오류로 처리한다.

<예> 의사님(√의사), 책상님(√책상), 선생(√선생님)

② 시제(ST)

- 시제 오류는 시제 또는 시상을 나타내기 위한 선어말어미, 관형사형 전성어미, ‘-(으)ㄴ 것’ 등의 오류 등을 말한다.

<예> 옛날에 한국의 정통 난방법을 온돌이다(√온돌이었다).
때로는 한국말을 공부할 때 끝이 없는 것 같은데(√같았는데) 오느날 갑자기 끝에 왔다.

- 시제 오류는 오류 양상을 두 가지로 처리한다. 하나는 시제 요소를 사용하지 않고 기본형을 사용했을 경우, 시제를 사용하지 않았다고 판단하여 누락 오류로 주석한다. 다른 하나는 시제를 제대로 인식하지 못하여 과거 시제나 미래 시제 자리에 현재 시제를 사용한 경우, 그 반대의 경우 등은 대치 오류로 처리한다. 즉, 시제를 사용했으나 현재와 과거, 과거와 미래 처럼 잘못 사용한 경우는 시제 간 대치로 처리한다.
- 이때, ‘이다/아니다’, 형용사, 연결어미 앞에서의 용언은 기본형이 현재를 나타내기 때문에 기본형을 사용했을 경우, 현재 시제로 인식하고 대치로 주석한다.

<예> 먹는다(√먹었다)
예쁘다(√예뻤다)
이다(√이었다)
아니다(√아니었다)
밥을 먹지만(√먹었지만),

밥을 먹고(√/먹었고)

- 미래시제를 나타내는 ‘-겠-’의 경우, 추측이나 의지와 같은 양태 의미와도 관련되는데, 본 연구에서는 추측을 제외한 경우 오류 층위에 시제를 주석한다.

<예> 앞으로 학교 규칙을 안 어긴다(√어기겠다).

☞ 미래의 의지를 나타내야 하는데 현재형 ‘ㄴ다’로 잘못 사용한 것으로, 선어말 어미 ‘-겠-’ 대치 오류로 처리한다.

사람들이 항상 물건을 어떻게 선택할지 모르겠다(√모른다).

아마 가을에 또 가겠습니다(√갈 것입니다).

☞ ‘모르겠다’의 경우 ‘-겠-’이 시제의 의미를 나타낸다고 보기 어렵다. 따라서 이때의 ‘-겠-’은 오류 층위에 시제를 처리하지 않는다.

- 관형사형 전성어미 ‘-(으)ㄴ, 는, (으)ㄹ’의 경우, 뒤에 오는 (의존)명사에 따라 시제를 나타내는 경우가 있고 그렇지 않은 경우가 있다. 시제를 나타내지 않을 경우는 대치 또는 오형태 활용 오류로 처리하고, 시제를 나타낼 때에는 관형사형 전성어미의 시제 대치 오류로 처리한다.
- 시제를 나타내지 않을 경우로 ‘-(으)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.

- 관형사형 전성어미에서 시제 대치 오류 판단에 어려움이 있을 경우, 현재(‘-는’)와 미래(‘-(으)ㄹ’)가 둘 다 가능할 때에는 용인 가능한 것으로 판단

하여 오류로 처리하지 않고, 명확하게 과거형을 써야 하는데 쓰지 않은 경우나 반대의 경우 시제 오류로 처리한다.

- 연결어미 ‘-(으)ㄴ지/는지/-(으)르지’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)르지’의 대치의 경우, 오류 층위에 시제를 주석하지 않도록 주의한다.

③ 사동(SC)

- 사동 오류는 사동사, 사동 표현의 사용, 사동문 생성에서 발생한 오류를 말한다.
- 사동사, 사동 표현을 사용해야 하는데 사용하지 않은 경우는 기본적으로 오류 양상을 대치 오류로 주석한다.
- 사동 표현은 접미사 ‘-이/히/리/기/우/구/추-’에 의한 사동, ‘-게 하다’에 의한 통사적 사동, ‘-내다, 만들다, 시키다’ 등 어휘적 사동으로 나타낼 수 있다. 이 연구에서는 접미사 ‘-이/히/리/기/우/구/추-’에 의한 사동사와 ‘-게 하다’, ‘시키다’ 사동 표현에 의한 사동으로 제한한다. ‘내다, 만들다, 시키다’ 중 ‘시키다’는 한국어 교육에서 ‘사동’을 나타내는 표현으로 교수하고 있는 상황을 고려하여 포함하나 나머지 어휘에 의한 사동은 맥락에 따라 다르게 처리될 수 있기 때문에 주석자 간 일관성을 유지하기 위해 제외한다.
- ‘형용사 - 게 하다’의 경우, ‘사동’으로 처리하지 않도록 주의한다. 예) 방을 깨끗하게 해야 한다.

<예> 왜냐하면 중국 법률 중에서 노동법에 따라서 소득 격차 등 불공평 제도를 감소할 수 있다(√감소시킬 수 있다).

- 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 층위는 원어절과 교정어절 양쪽에서 주석함에 따라 사동과 피동을 중복 주석한다.
- 사동 표현 ‘-게 하다’와 일반 사동사가 대치된 경우, 오류 위치는 형태 주석에 따라 일관되게 처리한다. 동사 또는 ‘연결어미+보조용언’으로 분리되어 처리되었을 경우는 각각의 품사로 오류 위치를 주석하며, 사동 표현의 경우 표현 문형 목록에 해당되기 때문에 표현문형(PE)도 중복 주석한다.

<예> 더 간단하게 하려고 생각하면 에어컨의 온도를 조금만 높게 하는(√높이는) 것만 한 방법이 없다.
 ☞ ‘-게 하다’가 사동사로 대치된 경우로 오류 위치는 ‘연결어미, 보조용언, 표현 문형으로 주석하며, 오류 양상은 대치로 주석한다.

④ 피동(SP)

- 피동사, 피동 표현의 사용, 피동문 생성에서 발생한 오류를 말한다.
- 피동사, 피동 표현을 사용해야 하는데 사용하지 않은 경우는 기본적으로 대치 오류로 처리한다.
- 피동 표현은 접미사 ‘-이/히/리/기-’에 의한 피동, ‘-아/어지다, -게 되다’에 의한 통사적 피동, ‘-되다, 받다, 당하다’ 등 어휘적 피동으로 나타낼 수 있다. 이 연구에서는 형태를 중심으로 접미사 ‘-이/히/리/기-’에 의한 피동사와 ‘-어지다’ 피동 표현에 의한 피동으로 제한한다.
- 단, 통사적 피동 ‘-아/어지다’의 경우, ‘형용사+아/어지다’는 피동보다는 상태변화의 의미를 나타낸다고 보고 ‘피동’으로 주석하지 않는다. 상태변화가 일어나게 된 요인이 타의에 의해 발생하여 피동의 의미가 내포되어 있더라도 한국어 교재 및 학습 기관에서 피동과 상태변화를 분리하여 교수하고 있으며, 맥락에 따라 피동과 상태변화를 구분하여 주석할 경우, 주석자 간 일관성이 떨어질 수 있기 때문에 ‘형용사+아/어지다’는 일괄적으로 오류 층위에서 피동으로 주석하지 않는다.
- ‘-게 되다’의 경우도 변화의 의미를 나타내는 경우가 많으며, 학교문법에서 피동에 포함시키지 않는 논의에 근거해 본 연구에서도 제외하였다.

- 즉, ‘형용사+아/어지다’와 ‘-게 되다’는 기본 의미를 변화로 보고, 피동으로 다루지 않으며, 맥락에 따라 다르게 판단할 수 있는 어휘적 피동도 제외한다.

<예> 우리 집에 물을 열리면(√열면) 계단을 있다.

☞ 피동표현을 사용하지 말아야 하는데, ‘열리면’으로 피동형을 사용했기 때문에 피동 대치 오류로 처리한다.

기술이 발달해서 멋진 영화나 공연이 많아질수록 전통문화를 점점 잊어버리게 했다.(√잊어버리게 된다).

☞ 사동 표현 ‘-게 하다’와 피동 표현 ‘-게 되다’의 대치 오류로 처리한다. 오류 층위는 원어절과 교정 어절 양쪽 모두를 기준으로 하기 때문에 이때에는 오류 층위에 사동(SC)과 피동(SP)을 중복 주석한다.

- 조사와 용언 교정이 모두 가능한 경우, 격조사 오류를 우선적으로 처리하나, 문맥에 따라 양쪽을 모두 바꿔야 하는 경우는 양쪽 모두 오류 주석한다. 특히, 피동문에서 용언을 교정하여 바뀌게 되는 조사의 경우, 교정어절만 써주고 오류로 처리하지 않았으나 피동/사동 구조를 모르고 있다는 측면에서 오류 층위에서 사동/피동을 주석하는 것으로 수정하여 처리한다.

<예> 둘째, 의학 기술의 발전에 따라 수명을(√수명이) 연장하지만(√연장되지만) 노인층 증가도 할 수 있다.

☞ ‘되다’의 피동으로 용언을 교정함에 따라 조사도 바뀌게 된다. 이 경우는 양쪽 모두 오류로 처리하고 조사에도 오류 층위에 ‘피동’을 주석한다.

- 피동 오류를 처리하는 데 있어, ‘동사+아/어지다’의 경우, <표준국어대사전>에 한 단어로 등재되어 있는 동사가 있는 반면, 등재되지 않은 단어가 있다. 이 경우, 형태소 분석에서 사전에 등재되어 있는 단어는 동사로 주석하고, 그렇지 않은 경우는 ‘연결어미+보조 용언’으로 분리하여 주석한다. 따라서 오류 주석에서 오류 위치는 형태 주석에 따라 일괄 처리하여, 동사로 분석했을 때는 그 품사를 따르고, 연결어미, 보조 용언으로 분리하였을

경우 해당 품사를 오류 위치로 주석한다.

- 이중피동을 사용한 경우, 첨가 오류로 처리한다.

<예> 환경 오염이 심해지게 되고(√심해지고) 있지만 더 이상 심하기 전에 여러분의 도움이 필요하다고 생각한다.
☞ ‘심해지다’에 ‘-게 되다’까지 첨가된 이중 피동표현으로 ‘-게 되다’의 ‘연결어미+보조용언’, 표현문형(PE)의 첨가 오류로 처리한다.

⑤ 부정(SN)

- 부정 표현의 사용, 부정문의 생성에서 발생한 오류를 말한다.

<예> 한국에서 혼자서 살다가보니 외로울 때가 많이 있으니까 그냥 혼자 있지 말고(√않고) 친구들이랑 같이 공부를 해요.

- 일반적으로 부사 ‘아니(안), 못’이나 부정의 의미를 가진 용언 ‘아니다, 아니하다(않다), 못하다, 말다’를 써서 부정문을 만드는 방법에 근거하여 부정 부사를 잘못 사용하거나 해당 용언에서 오류가 났을 경우, 오류 층위에 부정(SN)을 주석한다.
- ‘없다, 모르다’, 부정 의미의 접두사는 부정 오류에 포함하지 않는다.
- 장형부정인 ‘-지 않다’, ‘-지 못하다’, ‘-지 말다’의 경우, 표현문형 목록에 해당되기 때문에 오류 위치는 보조용언과 표현문형을 중복 주석한다. (※ ‘-고(야) 말다’는 부정의 의미를 나타내는 것이 아니기 때문에 부정으로 처리하지 않도록 주의한다.)
- 장형부정이 더 자연스럽지만 단형부정을 썼을 때 용인가능하기도 하다. 따라서 단형부정을 장형부정으로 반드시 바꿔야하는 경우 기준 마련 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 체인과 ‘-하다’가 분리될 때 ‘하다’ 앞에 아니(안)를 넣어 단형부정문을 만들 수 있다. 따라서 ‘-하다’ 파생동사 앞에 부정부사를 쓴 경우는 오류로 처리하고, 이때는 어순 오류와 부정 오류로 중복 주석한다.

<예> 스페인어 안 사용해서(√ 사용 안 해서) 스페인어만 말하기 저금 어렵습니다.

☞ ‘N+하다’ 파생동사 앞에 부정부사 ‘안’을 사용한 경우로, 이때에는 ‘사용 안 해서’와 ‘사용하지 않아서’ 두 가지로 교정이 가능하다. 그러나 이 경우, 최소 수정 원칙에 의해서 단형부정을 장형부정으로 바꾸는 것보다 단형부정의 위치를 잘못 사용한 것으로 보고 오류 층위에서 어순 오류와 부정 오류로 중복 주석한다.

⑥ 어순(WO)

- 어순 오류는 한국어의 통사 구조에 맞지 않는 방식으로 문장 전체 또는 일부가 배열된 경우를 말한다.

<예> 그래서 잘 아직까지(√ 아직까지 잘) 몰라요.
저녁까지 많이 이야기도(√ 이야기도 많이) 합니다.

- 한국어 어순의 특징 중 하나는 문장성분의 자리 이동이 비교적 자유롭다는 것이다. 그렇기 때문에 그만큼 용인가능성이 크다고 할 수 있다. 이에

따라 주석자 간의 일치도도 다르게 나타날 수 있어, 어순 오류의 경우 최소 수정의 기준을 마련하여 처리한다.

- 문장 부사는 자리 이동이 자유롭지만 성분 부사의 경우는 제한되기 때 특정한 성분을 수식해야 하는 성분부사의 위치를 잘못 사용했을 때는 오류로 주석한다.
- 관형사의 경우, ‘지시관형사-수관형사-성상관형사’ 순의 기준을 적용하여 처리한다.
- 어순 오류는 오류 위치와 오류 층위만 주석하고, 오류 양상은 주석하지 않는다. 또한 교정어절을 줄 필요가 없어 교정된 어순을 반영하여 앞이나 뒤에 추가하지 않는다.

<예> 저는 많이 여행을 (√많이) 가고 싶습니다.
 저는 한국 여행에서 자주 서울만 (√자주) 갔습니다.
 ☞ 일반적으로 성분 부사는 서술어 앞에서 수식해야 하는데, 이처럼 ‘많이’, ‘자주’가 명사 앞에 온 경우, [오류 위치-일반부사], [오류 양상-없음(빈칸)], [오류 층위-어순] 오류로 처리한다.

- 시간을 나타내는 표현의 배열이 잘못되었을 경우, 어순 오류로 처리한다. 시간을 나타내는 표현은 ‘년도-월-일-오전/오후/밤/낮/아침/점심/저녁-시-분-초’의 순서로 배열되는 것이 일반적으로 이를 기준으로 시간의 배열 어순 오류를 판단하여 처리한다.
- 2개의 문장 성분이 상호 교체될 때에는 2개 모두 대치 어순 오류로 주석한다.

<예> 8반 시(√8시 반)에 학교에 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.

- 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)

- 지시 오류는 부적절한 지시사의 선택으로 선행문과 후행문의 관계를 결속성 있게 나타내지 못한 경우를 말한다.
- 지시 오류는 앞 뒤 문장과의 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.

② 접속(DC)

- 접속 오류는 선행문과 후행문의 의미 관계를 나타내는 데에 부적절한 접속사를 사용한 경우를 말한다. 접속 부사 및 접속 표지의 오류가 포함된다.
- 접속 오류는 앞 뒤 문장과의 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.

③ 담화표지(DM)

- 담화표지 오류는 담화표지와 간투사의 오류로, 부적절한 담화 표지를 선택하거나, 잘못된 형태로 이들을 사용한 경우를 말한다.
- 담화표지는 미시 담화표지와 거시 담화표지로 나눌 수 있으나 연구자마다 그 정의가 다르고, 해당 형태도 다르기 때문에 담화표지의 목록을 마련하기 쉽지 않다. 이러한 이유로 미시 담화표지에 초점을 두고 오류를 판단하도록 한다. 미시 담화표지의 경우, 학습자의 L1의 영향으로 인한 간투사 사용과 모어화자와는 다른 위치에서 담화표지를 사용한 경우를 오류로 주석한다.

<예> 아~ 그럼 제가 음~ 오늘 밤에, 데~(√에~) 잊어버리지 않으면 추대할게요.
☞ ‘에~’는 간투사로 볼 수 있는데, 이를 ‘데~’로 잘못 발음하고 있어 오형태 오류로 주석하고, 오류 층위에서 담화표지 오류로 처리한다.

④ 구어/문어 오류(DS)

- 구어체(구어성)/문어체(문어성), 격식체/비격식체의 혼용에 의해 담화 맥락에서의 일관성이 떨어지는 경우를 말한다.

<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.

- 구어/문어 오류는 상황에 따라 용인 가능성을 적용할 수 있기 때문에 엄격하게 그 기준을 적용하기가 어렵다. 이에 구어체/문어체를 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.
- ‘하고’, ‘한테’ 등을 구어적 표현으로 보고, 문어에서 사용했을 경우 오류로 처리한다. 구어/문어 오류에 해당하는 목록은 다음과 같다.

<예> 한테(√ 에게)
 하고(√ 와/과)
 거/게(√ 것/것이)
 아무거(√ 아무것)
 근데(√ 그런데)

- 구어/문어 오류의 양상은 기본적으로 대치 오류로 처리한다.

<예> 제가 이 광고를 보고 지원하게 된 동기는 외국 학생 상담을 통해 많은 것을 배울 수 있을 거라고(√ 것이라고) 생각하기 때문입니다.

☞ ‘-것’을 ‘-거’, ‘-게(√ 것이)’로 사용한 경우는 [오류 위치-의존명사], [오류 양상-대치], [오류 층위-구어/문어] 오류로 처리한다.

저는 옷하고(√ 과) 화장품 아주 좋아해요.

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.

- 말더듬은 오류로 처리하지 않는다. 다만, 전사 단계에서 특정 표시를 하므로 이를 통해 향후 검색이 가능하게 한다.
- 자기 수정 발화의 경우, 수정 전 앞부분의 발화는 오류로 주석하지 않는다. 수정 후 발화에 초점을 두고 오류 여부를 판정한다.

<예> 친= 친구가 한국 음식이:: = 음식을 좋아해서

☞ ‘친= 친구’와 같은 말더듬거림은 오류로 처리하지 않는다. ‘음식이:: = 음식을’과 같이 수정 전 발화에서 조사를 잘못 사용하였지만, 다시 수정하여 조사를 고쳐 제대로 사용한 경우에 앞부분 ‘음식이’는 오류로 처리하지 않는다.

- 구어 오류 주석과 문어 오류 주석의 기본 원칙 및 처리 방법은 동일하다.

그러나 발화 상에서 나타나는 발음 오류의 경우에는 오류 양상(대치, 누락, 첨가, 오형태)을 주석하지 않고, 오류 위치와 오류 층위 [발음]만 주석한다.

- 구어에서는 발음 오류와 어휘 및 문법 오류의 구분이 명확하지 않을 수 있다. 즉, 학습자가 어휘와 문법을 잘못 사용한 것인지 단순히 발음을 잘못된 것으로 인해 나타난 오류인지 판별하는 데 어려움이 있다. 구어에서는 발음의 영향과 함께 어휘 및 문법 오류를 표시해주는 차원에서 오류 층위에서 중복 주석을 하도록 한다. 그러나 조사의 경우, 문법 오류를 우선 처리하도록 하고, 관형사형 전성어미를 사용해야 할 자리에 사용하지 못했을 경우도 받침을 발음하지 못한 음소 오류보다는 누락 오류로 문법 오류를 우선 처리한다.
- 구어 오류 주석에서의 또 다른 쟁점은 구어의 특성으로 볼 수 있는 현실 발음과 준말을 오류로 처리해야 하는가이다. 예를 들어, ‘김밥[김밥]’으로 발음했을 때 ‘적절성’을 기준으로 하여 ‘한국어 모어 화자’와 다르게 발음한다는 차원에서 오류로 볼 수도 있을 것이다. 그러나 현실 발음은 그 기준을 확정하기가 쉽지 않다. 한국어 화자도 표준 발음으로 발음하는 경우가 있고, 현실 발음을 어느 범위까지 인정해야 하는지도 문제가 되기 때문에 본 연구에서는 현실 발음에 어긋난다고 해서 오류로 처리하지 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> 하려고(√할려고)

☞ ‘ㄱ’로 시작하는 단어 앞에 받침 ‘ㄱ’을 첨가하여 발음하는 것은 한국인 모어 화자에게서도 많이 나타나는 현상이다. 현실 발음을 고려하여 이러한 경우는 오류로 처리하지 않는다.

그리고(√그리구)::, 음::

☞ ‘그리고’를 [그리구]라고 발음하는 것은 한국어 모어 화자들에게도 많이 나타난다. 이처럼 한국어 모어 화자들의

현실 발음을 고려하여, 줌(쫌), 바라요(바래요) 등 구어에서 허용되는 발음은 오류로 처리하지 않는다.

- 구어에서는 준말이 용인가능하기 때문에 오류로 보기 어려운 측면이 있다. 그러나 모든 준말을 허용할 수 있는 것은 아니기 때문에 구어에서 준말의 오류 판단 기준이 필요하다. 이에 따라 본 연구에서는 <표준국어대사전>에서 ‘~의 준말’로 등재되어 있는 것을 기준으로 삼는다. <표준>을 기준으로 ‘준말’로 등재되어 있는 형태는 오류로 처리하지 않고, 등재되지 않은 형태는 오류로 처리한다.

<예> 그래서 맘 먹고 여기 왔어요.

☞ ‘맘’의 경우, <표준>에 ‘마음의 준말’로 등재되어 있다. 따라서 오류로 처리하지 않는다.

그렇죠(√ 그쵸)::, 음::

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

- 구어에서 발음의 차원이 아닌, 형태를 잘못 발화한 경우는 오형태(MIF) 오류로 처리한다.

<예> 사건(√ 사잉)

보살핌(√ 보석필)

☞ 이와 같은 예시들은 발음을 잘못했다기보다 형태를 잘못 만들어낸 것이다. 유사 발음과도 떨어져 음소 오류로 볼 수 없고, 한국어에 없는 형태들을 발음했다고 판단하여 오형태 오류로 처리한다.

- 구어에서 조사의 생략이나 축약된 형태의 사용은 한국어에서 일반적으로 나타나는 현상이다. 또한 구어 오류 주석을 억양 단위로 했을 경우, 조사

의 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 듣지 못한 부분을 전사한 것으로 판단이 불분명하기 때문에 분석불능으로 처리한다.

- 구어 오류는 오류 층위에서 발음 오류와 가장 밀접하다. 오류 층위 [발음]에는 음소(PP), 음절(PS), 음운규칙(PC), 원어식 발음(PN), 중간 발음(변이음 포함(PA) 총 5가지가 있는데, 그중 ‘음소, 음절, 음운규칙’ 3가지를 우선적으로 주석한다. 원어식 발음과 중간 발음은 오류의 원인에 해당하는 문제이기 때문에 외래어(및 모국어 화자의 원어식 발음)의 경우에만 ‘원어식 발음’ 오류로 주석하고, 변이음(음성대치)이 분명하게 식별될 경우에만 ‘중간 발음’ 오류로 주석한다. 변이음 식별이 분명하지 않은 경우에는 주석하지 않는다.(☞ 세부 처리 방법은 ‘3. 범주별 세부 오류 유형의 처리, 4) 오류 층위, (1) 발음’을 참고한다.)

<예> 복용(√보룡),하는 뜻은(√뜻이느)::, 그 마약을 쓰는 아니면
마약을, 하,느:: 것입니다.

☞ 복용을 [보공]으로 발음해야 하나 [보룡]으로 발음하였으
므로 음소 오류로 처리한다. 오류 층위[발음]에 해당하는
오류이기 때문에 오류 양상은 주석하지 않고, 오류 위치
와 오류 층위만 주석한다. 따라서 [오류 위치-명사], [오
류 양상-없음(빈칸)], [오류 층위-음소]로 주석한다.

또한 ‘뜻이느’은 ‘뜻은’에 조사 ‘이’를 첨가한 것이기 때문
에 발음의 오류가 아닌, 문어와 마찬가지로 조사 첨가 오
류로 처리한다. 이 경우, [오류 위치-주격조사], [오류 양
상-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 나뉘이다		###기 나뉘이다
-기 때문		###기 때문
-기 마련이다		###기 마련이다
-기 십상이다		###기 십상이다
-기 위한		###기 위한
-기 위해(서)		###기 위해(서)
-기 일쑤이다		###기 일쑤이다
-기 전에		###기 전에
-기 짝이 없다		###기 짝이 없다
-기가 무섭게		###기가 무섭게
-기가 바쁘게		###기가 바쁘게
-기가 쉽다		###기가 쉽다
-기나 하다		###기나 하다
-기로 들다		###기로 들다
-기로 하다		###기로 하다

표제어	형태 정보	대표형
-기만 하다		###기만 하다
-기에 따라		###기에 따라
-기에 앞서(서)		###기에 앞서(서)
-ㄴ 것		
-ㄴ 것 같다		
-ㄴ 결과		
-ㄴ 김에		
-ㄴ 나머지		
-ㄴ 대로1		
-ㄴ 대로2		
-ㄴ 대신에		
-ㄴ 데요		
-ㄴ 듯		
-ㄴ 듯하다		
-ㄴ 마당에		
-ㄴ 모양이다		
-ㄴ 법이다		
-ㄴ 이상		
-ㄴ 줄		
-ㄴ 지2		
-ㄴ 채로		
-ㄴ 척하다		
-ㄴ 탓		
-ㄴ 편이다		
-ㄴ 후에		
-ㄴ가 보다		
-ㄴ다는 것이		
-ㄴ데도 불구하고		
-나 보다		###나 보다
-나 싶다		###나 싶다
-는 가운데		###는 가운데

표제어	형태 정보	대표형
-는 것		###는 것
-는 것 같다		###는 것 같다
는 고사하고	은 고사하고	
-는 길에		###는 길에
-는 김에		###는 김에
-는 대로	###은 대로, ###, ㄴ대로	
-는 대신에	###은 대신에, ###ㄴ 대신에	
-는 덕분에/이다		
-는 데다가	###은 데다가, ###ㄴ 데다가	
-는 도중에		
-는 동시에		###는 동시에
-는 동안		###는 동안
-는 등 마는 등		###는 등 마는 등
-는 듯	###은 듯, ###ㄴ 듯	
-는 듯하다	###은 듯하다, ###ㄴ 듯하다	
-는 마당에	###은 마당에, ###ㄴ 마당에	
-는 만큼	###은 만큼, ###ㄴ 만큼	
는 말할 것도 없고		
-는 모양이다	###은 모양이다, ###ㄴ 모양이다	
는 물론	###은 물론	
-는 바람에		###는 바람에
-는 반면에	###은 반면에, ###ㄴ 반면에	
-는 법이다	###은 법이다, ###ㄴ 법이다	
-는 사이		###는 사이
-는 수밖에 없다		
-는 이상	###은 이상, ###ㄴ 이상	
-는 적이 있다/없다		###는 적이 있다/없다

표제어	형태 정보	대표형
-는 줄	###은 줄, ###ㄴ 줄	
-는 중이다		###는 중이다
-는 척하다	###은 척하다, ###ㄴ 척하다	
-는 체하다		
-는 탓	###은 탓, ###ㄴ 탓	
-는 통에		###는 통에
-는 편이다	###은 편이다, ###ㄴ 편이다	
-는 한		###는 한
-는 한이 있어도/있더라도		###는 한이 있어도/있더라도
-는 한편		###는 한편
-는가 보다	###은가 보다, ###ㄴ가 보다	
-는다는 것이	###ㄴ다는 것이	
-는데도 불구하고	###은데도 불구하고, ###ㄴ데도 불구하고	
-도록 하다		###도록 하다
-ㄹ 것 같다		
-ㄹ 것1		
-ㄹ 것2		
-ㄹ 것이 아니라		
-ㄹ 대로		
-ㄹ 듯		
-ㄹ 듯하다		
-ㄹ 따름이다		
-ㄹ 때		
-ㄹ 리가 없다		
-ㄹ 만큼		
-ㄹ 만하다		
-ㄹ 모양이다		
-ㄹ 바에		
-ㄹ 법하다		

표제어	형태 정보	대표형
-ㄴ 뻔하다		
-ㄴ 뿐만 아니라		
-ㄴ 수밖에는 없다		
-ㄴ 줄		
-ㄴ 테고		
-ㄴ 테냐		
-ㄴ 테니		
-ㄴ 테다		
-ㄴ 테면		
-ㄴ 테야		
-ㄴ 테지만		
-ㄴ 텐데		
-ㄴ까 보다		
-ㄴ락 말락 하다		
-려고 하다		
-려나 보다		
로 인하다	으로 인하다	
를 가지고	을 가지고	
를 막론하고	을 막론하고	
를 불문하고		
를 위해(서)	을 위해(서)	
만 같아도		만 같아도
만 아니면		만 아니면
-면 되다		
-면 몰라도		
-면 안 되다		
-면 좋겠다		
-아 가다	###어 가다, ###여 가다	
-아 가지고	###어 가지고, ###여 가지고	
-아 계시다		
-아 내다	###어 내다, ###여 내다	

표제어	형태 정보	대표형
-아 놓다	###어 놓다, ###여 놓다	
-아 대다	###어 대다, ###여 대다	
-아 두다	###어 두다, ###여 두다	
-아 드리다	###어 드리다, ###여 드리다	
-아 버리다	###어 버리다, ###여 버리다	
-아 보다	###어 보다, ###여 보다	
-아 보이다	###어 보이다, ###여 보이다	
-아 오다	###어 오다, ###여 오다	
-아 있다	###어 있다, ###여 있다	
-아 주다	###어 주다, ###여 주다	
-아 치우다	###어 치우다, ###여 치우다	
-아도 되다	###어도 되다, ###여도 되다	
-아서는 안 되다		
-아야 되다	###어야 되다, ###여야 되다	
-아야 하다	###어야 하다, ###여야 하다	
-어 가다		
-어 가지고		
-어 내다		
-어 놓다		
-어 대다		
-어 두다		
-어 드리다		
-어 버리다		
-어 보다		
-어 보이다		
-어 오다		
-어 있다		

표제어	형태 정보	대표형
-어 주다		
-어 치우다		
-어도 되다		
-어야 되다		
-어야 하다		
에 관하여		에 관하여
에 관한		에 관한
에 대하여		
에 대한		
에 따라		에 따라
에 따르면		에 따르면
에 비하여		에 비하여
에 의하면		에 의하면
에 의하여		에 의하여
에도 불구하고		에도 불구하고
-여 가다		
-여 가지고		
-여 내다		
-여 놓다		
-여 대다		
-여 두다		
-여 드리다		
-여 버리다		
-여 보다		
-여 보이다		
-여 오다		
-여 있다		
-여 주다		
-여 치우다		
-여도 되다		
-여야 되다		

표제어	형태 정보	대표형
-여야 하다		
-으려고 하다	###려고 하다	
-으려나 보다	###려나 보다	
으로 인하다	로 인하다	
-으면 되다	###면 되다	
-으면 몰라도	###면 몰라도	
-으면 안 되다	###면 안 된다	
-으면 좋겠다	###면 좋겠다	
-은 가운데		###은 가운데
-은 것	###ㄴ 것	
-은 것 같다	###ㄴ 것 같다	
-은 결과	###ㄴ 결과	
은 고사하고		
-은 김에	###ㄴ 김에	
-은 나머지	###ㄴ 나머지	
-은 다음에	###ㄴ 다음에	
-은 다음에야	###ㄴ 다음에야	
-은 대로1	###ㄴ 대로, ###는 대로	
-은 대로2	###ㄴ 대로, ###는 대로	
-은 대신에	###ㄴ 대신에, ###는 대신에	
-은 데다가1	###ㄴ 데다가	
-은 데다가2	###ㄴ 데다가, ###는 데다가	
-은 뒤에		
-은 듯	###ㄴ 듯, ###는 듯	
-은 듯하다	###ㄴ 듯하다, ###는 듯하다	
-은 마당에	###ㄴ 마당에, ###는 마당에	
-은 만큼	###ㄴ 만큼, ###는 만큼	
-은 모양이다	###ㄴ 모양이다, ###는 모양이다	
은 물론	는 물론	

표제어	형태 정보	대표형
-은 반면에	###ㄴ 반면에, ###는 반면에	
-은 법이다	###ㄴ 법이다, ###는 법이다	
-은 이상	###ㄴ 이상, ###는 이상	
-은 줄	###ㄴ 줄, ###는 줄	
-은 지2	###ㄴ 지	
-은 채로	###ㄴ 채로	
-은 척하다		###은 척하다
-은 체하다		
-은 탓	###ㄴ 탓, ###는 탓	
-은 편이다	###ㄴ 편이다, ###는 편이다	
-은 후에	###ㄴ 후에	
-은가 보다	###ㄴ가 보다, ###는가 보다	
-은데도 불구하고	###ㄴ데도 불구하고, ###는데도 불구하고	
을 가지고	###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	###르 모양이다	
-을 바에	###르 바에	
-을 법하다	###르 법하다	
을 불문하고		
-을 뻔하다	###르 뻔하다	
-을 뿐만 아니라	###르 뿐이다	
-을 뿐이다		
-을 수 없다		
-을 수 있다		
-을 수밖에 없다	###르 수밖에 없다	
을 위해(서)	###를 위해(서)	
-을 줄	###르 줄	
-을 테고	###르 테고	
-을 테냐	###르 테냐	
-을 테니	###르 테니	
-을 테니까		
-을 테다	###르 테다	
-을 테면	###르 테면	
-을 테야	###르 테야	
-을 테지만	###르 테지만	
-을 텐데	###르 텐데	
-을까 보다	###르까 보다	
-을라 말라 하다	###르라 말라 하다	
-지 말다		###지 말다
-지 못하다		###지 못하다
-지 않다		###지 않다

참고 문헌

- 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(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, 외국어로서의 한국어교육 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

- 준계학술발표논문집, 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, 우리말 글 52, 우리말글학회.
- 이승연(2006), 한국어 학습자 말뭉치 오류 표지 방안 제고, 이중언어학 31, 이중언어학회.
- 이승연(2007), 한국어 학습자 오류 판정 및 수정 기준 연구-교사, 비교사 집단간 오류 판별 비교 실험을 바탕으로, 이중언어학 33, 이중언어학회.
- 이승연(2007), 한국어교육을 위한 한국어 학습자 말뭉치의 구축과 활용 연구, 고려대 박사학위논문.
- 이유림, 김영주(2013), 교사의 피드백 방법이 한국어 학습자의 작문 내 어휘 오류 감소에 미치는 영향, 외국어로서의 한국어교육 39, 연세대학교 언어연구교육원 한국어학당.
- 이은서(2017), 중국어권 학습자의 접사 사용 연구, 연세대학교 대학원 석사학위 논문.
- 이정희(2002), 한국어 오류 판정과 분류 방법에 관한 연구, 한국어교육 13(1), 국제한국어교육학회.
- 이정희(2003), 초급 단계 한국어 학습자의 어휘 오류, 이중언어학 22, 이중언어학회.
- 이정희(2009), 중국어권 한국어 학습자의 어휘 오류 연구, 한국어 교육 19(3), 1-23쪽, 국제한국어교육학회.
- 이화진·이지연(2016), 학습자 말뭉치 구축과 음성 인식 활용, 한국언어문화교육학회 학술대회 발표 자료집, 한국언어문화교육학회.
- 이훈호(2015), 한국어 오류 분석 연구의 동향 분석 연구, 외국어교육연구 29(2), 107-135쪽, 한국외국어대학교 외국어교육연구소.
- 전영옥(2010), 여성결혼이민자의 한국어 어휘 오류 분석, 한말 연구 27, 한말연구학회.
- 조철현 외(2002), 한국어 학습자의 오류 유형 조사 연구, 문화관광부.
- 최원평, 유효려(2010), 중국 대학생 글쓰기에 나타난 어휘 오류 연구, 언어와 문화 6(3), 한국언어문화교육학회.
- 한상미(2014), 중급 한국어 학습자의 구어 담화에 나타난 조사 오류 연구, 한국어교육 25(3), 국제한국어교육학회.
- 한송화(2001), 말뭉치와 학습자 오류를 이용한, 외국인 학습자를 위한 한국어 어휘 사전의 의미 기술, 한국어정보학 4, 한국어정보학회.
- 한송화, 원미진(2017), 모어 화자와 한국어 학습자 말뭉치에서의 ‘은/는’과 ‘이/가’의 분포와 조사 선택 요인 분석, 언어사실과 관점 41, 연세대학교 언어정보연구원, 5-47.
- 한송화·강현화(2016), 학습자 말뭉치에서의 구어 전사와 오류 주석의 쟁점과 실제, 한국언어문화교육학회 학술대회 발표 자료집, 한국언어문화교육학회.
- Brock, C , Crookes, C , Day, R., and Long, M. (1986). The differential effects of corrective feedback in native speaker-non-native speaker conversation. In R. Day (Ed.), Talking to learn. Rowley, MA: Newbury House. pp. 229-236.

- Brock, C. (1986). The effects of referential questions on ESL classroom discourse. *TESOL Quarterly*, 20, pp. 47-59.
- Corder. S. P.(1981), *Error Analysis and Interlanguage*, Oxford University Press.
- Foster, P. and Skehan, P. (1996) The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition* 18. pp. 299 - 324.
- Foster, P., Tonkyn, A. and Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics* 21:3. pp. 354-375.
- Hunt, K. (1965). Grammatical structures written at three grade levels. NCTE Research report No. 3. Champaign, IL, USA: NCTE. pp. 1467-1770.
- James, C.(1998), *Errors in Language Learning and Use*. New York : Addison Welsey Longman Inc. pp. 144-154.
- Pica, T., Holliday, L., Lewis, L. and Morgenthaler, L. (1989) Comprhensible Output As An Outcome of Linguistic Demandes On the Learner, *Studies in Second Language Acquisition* 11:1. pp. 63-90.
- Young, R. (1995). Conversational Styles in Language Proficiency Interviews. *Language Learning* 45:1. pp. 3 - 42.

2017 Project on Research and Construction of the Korean Learner Corpus

This research is a follow-up study and the third year of research on the construction of learner corpus for Korean language learners in mid and long-term situations. Following the guidelines of corpus construction established in the previous round's research, this research aims to revise and elaborate on 1,150,000 raw corpus (written language 1,000,000 items, spoken language 150,000 items) constructed in previous research, and simultaneously add approximately 660,000 new raw corpus (written language 500,000 items, spoken language 160,000 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 corpus is divided into foundation building for systematic corpus construction and model demonstration for corpus application. Building off of the data and properties of corpus construction in immigrants obtained in the previous research, this year's round of research elaborated on the current issues of corpus construction, e.g., acquisition instruction, data collection instruction, written language instruction, spoken language instruction, morph tagged instruction, and error annotated instruction. Additionally, a research system allowing for the practical construction of Korean learner corpus by Korean language educators and researchers was established. This research system was constructed from analysis of the data and statistics obtained in the previous rounds of corpus research.

Korean Language Learner Corpus Education and Promotion: In the field of corpus education and promotion, Vital acquisition instruction, written language instruction, spoken language instruction, morph tagged instruction, and error annotated instruction for systematic corpus construction tool use and education were undertaken. Additionally, through periodic and non-periodic workshops, construction data reliability was assured due to higher rates of

data collector consistency. Korean learner corpus has also become widely known and a workshop (held twice in Seoul and once in the outer provinces) on user data collection targeting Korean language educators and researchers was held. Also, a presentation on Korean language learner corpus was held in an academic conference, promoting an actualized model of research to academics.

Korean Language Learner Corpus Collection, Revision and Construction: Korean language learner corpus was collected through international students undergoing language training in Korea and immigrants (marriage migrants, immigrant children, migrant laborers), and with the help of 14 Korean education institutions, multicultural family centers, and 20 various immigrant social welfare facilities. The construction of Korean language learner corpus built off of the foundation established in the previous two rounds of research, and consisted of 1,150,000 raw corpus (written 1,000,000 items, spoken 150,000 items) 770,000 morph tagged corpus (written 700,000 items, spoken 70,000 items), and 200,000 error annotated corpus (written 140,000 items, spoken 60,000 items), along with the correction and formation of new corpus. Additionally, new items in the form of 660,000 raw corpus (written 500,000 items, spoken 160,000 items), 500,000 morph tagged corpus (written 400,000 items, spoken 100,000 items), and 160,000 error annotated corpus (written 100,000 items, spoken 60,000 items) were constructed. In total, the three rounds of research from 2015–2017 amount to 1,810,000 raw corpus (written 1,500,000 items, spoken 310,000 items), 1,270,000 morph tagged corpus (written 1,100,000 items, spoken 170,000 items), and 360,000 error annotated corpus (written 240,000 items, spoken 120,000 items).

Elaboration on Inspection of Corpus Construction: Korean learner corpus undertakes the process of worker cross inspection, followed by high profile worker inspection, then deep inspection through worker association. Furthermore, in order to further elaborate on the inspection system in the third round of research, a tool examining the statistical reliability of the corpus data was added, and test usage and improvement plans were suggested to the development team. Aside from this, based off of the work logs, sample information, and annotated information, errors in the annotation or correction processes were inspected and eliminated.

As the successor to “21st century Sejong Project,” “Korean Learner

Corpus” aims for broad application by education researchers, Korean language teachers, and Korean language learners. This broad application will then allow for the strengthening of the Korean language’s systemization and its international competitiveness.

Key terms: Korean learner corpus, Korean education data, collection of immigrant test data, written language corpus, spoken language corpus, raw corpus, morph tagged corpus, error annotated corpus

연구 책임자: 강현화(연세대학교 국어국문학과 교수)
 공동 연구원: 김선정(계명대학교 한국문화정보학과 교수)
 김일환(고려대학교 민족문화연구원 HK연구교수)
 김정숙(고려대학교 국어국문학과 교수)
 김한샘(연세대학교 언어정보연구원 HK교수)
 안경화(서울대학교 언어교육원 대우 부교수)
 이동은(국민대학교 국어국문학과 부교수)
 이정희(경희대학교 국제교육원 부교수)
 한송화(연세대학교 언어정보연구원 HK교수)
 김수현(국립국어원 한국어진흥과 학예연구사)
 실무 연구원: 홍혜란(연세대학교 언어정보연구원 전문연구원)
 연구 보조원: 공나형(연세대학교 대학원 박사과정)
 김동은(연세대학교 대학원 박사과정)
 김선영(연세대학교 대학원 박사과정)
 김정현(연세대학교 대학원 박사과정)
 배미연(연세대학교 대학원 박사과정)
 송지혜(연세대학교 대학원 박사과정)
 유소영(연세대학교 대학원 박사과정)
 허희정(연세대학교 대학원 박사과정)
 담당 연구원: 김수현(국립국어원 한국어진흥과 학예연구사)

2017년 한국어 학습자 말뭉치 연구 및 구축 사업

발 행 인	송 철 의
발 행 처	국립국어원 서울시 강서구 금낭화길 148(방화 3동 827) 전화: 02-2669-9743~4 전송: 02-2669-9727
인 쇄 일	2017년 12월 15일
발 행 일	2017년 12월 15일
인 쇄	학위사

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