

국립국어원 2023-01-51

발간등록번호

11-1371028-000979-01

2023년 한국수어 말뭉치 수집 및 분석

연구 책임자 | 홍성은

2023. 12. 8.



국립국어원

제출문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 「2023년 한국수어
말뭉치 수집 및 분석」에 관한 연구보고서를 작성하여 제출합니다.

■ 사업 기간: 2023년 5월 2일 ~ 2023년 12월 8일

2023년 12월 8일

연구 책임자 : 홍성은(강남대학교)

연구수행기관	강남대학교 산학협력단
연구 책임자	홍성은
공동 연구원	강창욱, 원성욱, 이은영, 우내리
연구 보조원	이한솔, 유재은, 손설표, 지혜원, 김효정

2023년 한국수어 말뭉치 수집 및 분석

■ 연구 기관 : 강남대학교 산학협력단

■ 연구 기간 : 2023년 5월 2일 ~ 2023년 12월 8일

■ 연구 진 : 총 10명

구분	이름	소속
책임 연구원	홍성은	강남대학교 대학원 특수교육학과 초빙교수
공동 연구원	강창욱	강남대학교 특수교육재활연구소 선임연구원
	원성욱	한경국립대학교 재활복지학부 교수
	이은영	프리랜서 농통역사
	우내리	강남대학교 대학원 박사과정
연구 보조원	이한솔	강남대학교 대학원 석사과정
	유재은	한국복지대학교 학생
	손설표	한국복지대학교 학생
	지혜원	수어민들레
	김효정	강남대학교 석사과정

<국문 요약>

2023년 한국수어 말뭉치 수집 및 분석

이 연구의 목적은 과학적이고 체계적인 한국수어 연구 기반 마련을 위해 원시 수어 말뭉치를 수집하고, 한국수어 말뭉치 공개를 위해 구축된 한국수어 말뭉치의 기본 주석 및 번역 주석을 정비하는 데 있다. 또한 한국수어 말뭉치 공개 및 홍보를 위해 수어 영상을 제작한다.

이 연구에서의 한국수어 원시 말뭉치 수집은 아직 수어 말뭉치가 수집된 적이 없는 곳을 대상으로 하여 지역의 특성이 반영된 다양한 수어 형태를 확보할 수 있게 할 것이다. 이미 구축된 한국수어 원시 말뭉치의 기본 주석과 번역 주석 정비는 그간의 축적된 한국수어 말뭉치를 공개할 수 있는 환경을 조성하여 향후 한국수어 연구에 필요한 기초 정보를 제공한 데 핵심적인 역할을 할 것이다.

‘한국수어 원시 말뭉치 수집’은 미구축 지역인 충북지역에서 수집이 이루어졌는데, 32명의 언어제공자로부터 69시간 11분 55초 분량의 영상을 촬영·수집하였다. 충북지역의 수어 말뭉치 수집을 위해 사용된 유도 과제는 2015년부터 활용한 13종과 2021년 신규 개발한 5종의 유도자료를 더하여 총 18종이었다. 수집한 영상은 유도 과제별로 편집하여 초별 가공한 이후, 이를 다시 세 차례 이상 편집 및 검증 과정을 거쳐 최종적으로 가공 시간 기준 59시간 31분 29초 분량의 원시 말뭉치를 확보하였다.

‘기 구축된 한국수어 말뭉치 공개를 위한 기본 주석 및 번역 주석 정비’는 정비 지침 및 선정 기준을 먼저 설정한 다음 실시하였으며, 63분 54초 분량의 번역 정비가 된 수어 말뭉치, 95분 6초의 주석과 번역이 모두 정비된 수어 말뭉치를 확보하였고, 공개가 가능한 원시 수

어 말뭉치 273분 30초를 합하여, 총 432분 30초 분량의 공개용 한국수어 말뭉치를 확보하였다.

‘한국수어 말뭉치 안내 및 홍보 동영상 제작’에서는 한국수어 말뭉치에 대한 접근이 용이하도록 한국수어 말뭉치를 공개·배포하는 국립국어원 ‘모두의 말뭉치’의 회원가입 및 말뭉치 신청 방법 안내 수어 동영상, ‘모두의 말뭉치’ 내 한국수어 말뭉치 소개 자료 수어 동영상, 그리고 한국수어 말뭉치가 공개되기까지의 전 과정을 보여주며 한국수어 말뭉치를 전체적으로 홍보하는 영상 세 가지를 제작하여 처음으로 공개될 한국수어 말뭉치가 대중에게 잘 알려지도록 하였다.

주요어: 한국수어, 한국수어 연구, 한국수어 말뭉치, 주석, 번역, 공개

<영문 요약>

2023 Data Collection and Analysis of the Korean Sign Language (KSL) Corpus

The purpose of this study is to collect sign language data for the Korean Sign Language (KSL) corpus and to lay the foundation for scientific and systematic Korean Sign Language research, and to verify basic annotations and translation annotations of the already established KSL corpus in order to publish it. We also produced sign language videos to publicize and promote the KSL corpus.

The data collection of the sign language data in this study will be conducted in places where no corpus data has yet been collected, and will enable the collection of various forms in sign language that reflect the characteristics of the region. The maintenance of basic annotations and translation annotations of the existing KSL corpora will play a key role in creating an environment in which the existing KSL corpora can be made public and provide basic information needed for future KSL research.

The data collection of corpus data was conducted in the Chungbuk region. 69 hours, 11 minutes, and 55 seconds of video was filmed and collected from 32 language informants. There were a total of 18 elicitation tasks used to collect sign language data in the Chungbuk region, including 13 tasks used since 2015 and 5 tasks newly developed in 2021. The collected videos were edited and rough-processed for each elicitation task, then again edited and reviewed more than three times. Finally, the raw corpus contains 59 hours, 31 minutes, and 29 seconds.

The 'Maintenance of basic annotations and translation annotations for the publication of the existing KSL corpus' was conducted by first establishing maintenance guidelines and selection criteria. Following these criteria we reviewed 273 minutes and 30 seconds of sign language video data, 63 minutes and 54 seconds of sign language data with translation, and 95 minutes and 6 seconds of annotated and translated data, which makes a total of 432 minutes and 30 seconds of KSL corpus data available for the public.

In order to promote and explain the KSL Corpus three videos were produced. The first video guide explains everything from the registration process to how to download materials of the KSL corpus. The second video contains explanations which help to understand the composition of the KSL corpus. And the last video shows the entire process of building a sign language corpus until its publication and has the aim to promote the KSL corpus.

Keywords: Korean Sign Language, Korean Sign Language research, Korean Sign Language corpus, annotations, translations, publication

목 차

I. 서론	1
1. 연구의 필요성과 목적	1
1) 연구 필요성	1
2) 연구 목적과 과업	6
2. 연구 방법	7
1) 연구 내용	7
2) 연구 방법	7
3) 연구 절차	9
4) 과업별 추진 일정	10
II. 한국수어 원시 말뭉치 수집	11
1. 언어 자료 수집 절차	12
2. 유도자료	13
1) 유도자료	14
2) 유도자료 설명 동영상	22
3. 언어제공자 모집 및 선정	24
1) 언어제공자 선정 조건	24
2) 언어제공자 모집	25
3) 언어제공자 기본 사항	25
4) 언어제공자 질문지	31
5) 촬영 동의	31
4. 사회자 및 촬영·편집자 교육	32
1) 사회자의 역할	32
2) 사회자 선정 및 교육	32

3) 유도 과제 설명 동영상 제작	33
4) 촬영 및 편집자 교육	33
5. 언어 수집 촬영	35
1) 촬영실 조건 및 촬영 시스템 지침	35
2) 촬영장 설치	36
3) 촬영 설명서	38
4) 촬영 시간	40
6. 영상 편집	42
1) 편집 지침	42
2) 편집 영상 검수	47
3) 수집 영상 분량	48
7. 후속 자료 수집 시 고려 사항	50

Ⅲ. 공개를 위한 한국수어 말뭉치 정비 53

1. 한국수어 말뭉치 공개 배포 대상 말뭉치 선정.....	53
2. 한국수어 말뭉치 공개를 위한 검토	53
1) 공개 배포를 위한 검토 기준	53
2) 검토 방법 및 결과	54
3. 한국수어 말뭉치 공개를 위한 정비	57
1) 주석 정비 절차	57
2) 번역 정비 절차	62
4. 한국수어 말뭉치 공개 자료	74
1) 공개 파일 및 시간	74
2) 지역별	75
3) 성별	75
4) 과제별	76
5) 언어제공자 연령별 분표	76
6) 공개 파일의 티어명	78

IV. 한국수어 말뭉치 공개 및 홍보를 위한 동영상	80
1. 홍보영상 제작 절차	81
2. 홍보영상 촬영 조건 및 촬영 시스템	82
3. 홍보영상 스토리보드	83
V. 요약 및 제언	91
1. 연구 결과 요약	91
1) 한국수어 원시 말뭉치 수집	91
2) 기 구축 한국수어 말뭉치 공개를 위한 정비	93
3) 한국수어 말뭉치 공개 및 홍보 영상제작	95
2. 활용 방안 및 제언	96
1) 연구의 활용 및 기대 효과	96
2) 제언	97

참고문헌	104
------------	-----

[부록 1] 국가 언어 자원(말뭉치) 구축 및 활용 저작권 등 이용 허락 계약서	106
[부록 2] 한국수어 말뭉치 사업 언어제공자 개인정보 수집·이용 및 제공 동의서	110
[부록 3] 언어제공자 질문지	112
[부록 4] 언어제공자와 사회자 관계에 관한 질문지	119
[부록 5] 원시 말뭉치 수집 점검표	120
[부록 6] 사회자 일지(서식)	124
[부록 7] 촬영 일지(서식)	126
[부록 8] 한국수어 말뭉치 주석 규약	128

표 목 차

<표 II -1> ‘H-주제 영역’ 과제	19
<표 II -2> 언어제공자 선정 조건	24
<표 II -3> 언어제공자의 인적 사항	25
<표 II -4> 언어제공자 연령	26
<표 II -5> 언어제공자 청각보조기기 착용 여부	28
<표 II -6> 아버지의 농인 여부	28
<표 II -7> 어머니의 농인 여부	29
<표 II -8> 배우자의 농인 여부	29
<표 II -9> 한국수어 습득 시기	30
<표 II -10> 한국수어 습득 장소	30
<표 II -11> 최종 학력	31
<표 II -12> 유도 과제별 권장 촬영 시간	40
<표 II -13> 충북 지역 유도 과제별 평균 촬영 시간	40
<표 II -14> 충북 지역 말뭉치 촬영 시간	40
<표 II -15> 충북 지역 언어제공자 쌍별 말뭉치 원본 영상 시간	48
<표 II -16> 충북 지역 언어제공자 쌍별 말뭉치 말뭉치 편집본 영상 시간 ..	49

그림 목 차

[그림 I -1] 연구 절차	9
[그림 II -1] 한국수어 말뭉치 언어 자료 수집 현황	11
[그림 II -2] 언어 자료 수집 절차	12
[그림 II -3] 언어제공자 모집 및 선정 절차	25
[그림 II -4] 언어제공자의 연령	27
[그림 II -5] 언어제공자의 성별	27
[그림 II -6] 언어제공자 거주 지역 분포	27
[그림 II -7] 언어제공자 촬영실 배치도	36
[그림 II -8] 언어제공자 휴식 공간	37
[그림 II -9] 연구진 모니터 공간	37

I. 서론

1. 연구의 필요성과 목적

1) 연구의 필요성

○ 한국수어의 언어적 위치 확보와 한국수어의 연구 기반 마련

2016년 제정된 한국수화언어법은 한국수어를 농인의 공용어로 공포하면서 농인이 모든 생활영역에서 한국수어를 통하여 삶을 영위하고 필요한 정보를 제공받을 권리가 있음을 천명함으로써 한국수어는 그 언어적 위치를 확보하게 되었다.

그러나 수어의 경우는 음성언어와 달리 농부모 밑에 태어난 소수의 수어 사용자를 제외하고는 모어로 수어를 습득하는 경우가 거의 없다. 이로 인해 동일 의미의 수어 어휘가 여러 형태로 사용되어 수어 연구에서 수어 어휘나 문법을 직관적으로 기술하는 데 어려움을 주는 원인이 되기도 한다. 즉, 수어는 음성언어(한국어)와는 달리 습득과 학습 환경이 매우 열악하여 많은 변이가 나타나므로 이러한 다양한 수어 변이현상을 기록하고 기술한 자료들이 축적되어야 실증적인 수어 연구가 이루어질 수 있다. 따라서 수어의 구조와 사용, 수어 습득과 발달, 수어 학습에 대해 이해하고 연구를 활성화시키기 위해서는 수어 말뭉치를 기반으로 하여 한국수어의 체계를 과학적이고 체계적으로 밝힐 필요가 있다.

○ 한국 수어와 농문화 보존을 위한 수어언어학 연구의 활성화

미국, 독일, 네덜란드, 영국, 호주 등 해외 여러 나라에서는 문자가 없이 광범위하게 사용되는 수어의 보존과 농문화의 기록, 수어 연구 기반의 제공 등을 목적으로 수어 말뭉치를 구축하고 있다. 또한 우리

보다 앞서 수어 연구를 시작한 나라들에서는 수어언어학에서 수어 말뭉치를 기반으로 하여 분류사나 일치동사 등 형태론적 주제에 대한 연구가 이미 많이 이루어졌다. 최근 ‘수어 연구의 이론적 이슈(Theoretical Issues in Sign Language Research, TISLR)’라는 수어학 전문 학술지에서는 사회언어학, 비수지 기호, 심리 언어학, 토착 수어(Indigenous Sign Language)와 이중 언어 등에 대한 연구가 발표되는 등 다양한 주제의 수어 연구가 활발하게 진행되고 있다.

이에 우리나라도 2016년 한국수화언어법 제정의 취지와 목적을 살려 우리 고유의 한국수어를 보존하기 위한 과학적인 연구가 체계적이고 지속적으로 이루어질 필요가 있다.

○ 구축 완료된 한국수어 말뭉치의 공개 및 활용을 위한 환경 구축

우리나라에서도 2015년에 한국수어 연구 활성화를 위한 과학적인 연구 기반을 제공하기 위해 ‘한국수어 말뭉치 기반 조사 및 시범 구축’ 사업을 진행하였다. 이 시범 구축 사업에서는 한국수어 원시 말뭉치 구축을 위한 지침서를 개발하고, 서울 지역의 농인 30쌍을 대상으로 약 71시간의 원시 말뭉치를 구축하였으며, 그중에서 약 2,800분 분량의 영상을 번역하고 780분 분량의 기본 주석을 실시하였다(국립국어원, 2015).

2016년에는 ‘한국수어 자료 통합 지원 시스템 구축’ 사업이 진행되어 주석 지침을 수정 보완하고, 수어 말뭉치 자료 관리를 위한 수어 자료 시스템을 마련하였다. 또한 이 사업에서는 한국수어 말뭉치를 활용한 실증적인 수어 연구 사례를 제시하기 위해 입 움직임과 관련된 주석 지침을 마련하고, 한국수어에 나타난 입 움직임(mouthing)에 대한 연구 사례를 제시하기도 하였다. 그와 더불어 해외 수어 자원과의 연계 방안을 제시하였으며, 농사회를 대상으로 한국수어 말뭉치 구축에 대한 이해를 향상시키기 위해 2016년 9월 9일과 10일 두 차례의 워크숍도 개최하였다(국립국어원, 2016).

그 후 2017년에는 기구축한 한국수어 원시 말뭉치를 대상으로 약 260분 분량의 영상에 대한 기본 주석을 실시하고, 타입 정비와 더불어 한국수어 자료 통합 지원 시스템에 연결하는 작업을 완성하였다(국립국어원, 2017).

2018년에는 한국수어 말뭉치의 확대를 위해 네 지역(전북, 전남, 강원, 경남)의 농인 총 14쌍으로부터 한국수어를 수집하여 약 35시간 분량의 원시 말뭉치를 구축하였으며, 약 80분 분량의 기본 주석을 실시하였다(국립국어원, 2018).

2019년에는 한국수어 원시 말뭉치의 확대 및 지역 균형성 확보를 위해 경남 지역 농인 총 11쌍으로부터 약 31시간 분량의 한국수어 원시 말뭉치를 구축하였으며 2015년 구축한 주석 말뭉치에 대한 정비의 일환으로 135분 분량의 의미 주석을 실시하였다(국립국어원, 2019).

2020년에는 한국수어 말뭉치의 확대 및 지역 균형성 확보를 위해 전남 지역(광주 포함) 농인 총 19쌍을 대상으로 약 51시간 분량의 한국수어 원시 말뭉치를 구축하였으며, 2019년에 구축한 원시 말뭉치 중 약 80분 분량의 영상에 대한 기본 주석과 의미 주석을 실시하였다. 또한 국내외 수어 말뭉치 구축 및 연구 현황을 분석함과 동시에 산재한 한국수어 자료 현황을 파악하여 한국수어 말뭉치 구축 사업의 방향을 탐색하고 중장기 계획(안)을 제시하였다(국립국어원, 2020).

2021년에는 2015년 개발된 13종의 유도자료를 보완할 수 있는 5개의 신규 유도 과제를 개발하였다. 지역 수어 자료 수집 양이 균등하도록 전북, 강원, 경기, 광주·전남 총 4개 지역에서 총 41쌍의 농인을 대상으로 추가로 원시 말뭉치를 수집하여 약 147시간의 원시 말뭉치를 구축하였다. 완료 시점에 신규 유도 과제 중 1종이 저작권 문제가 발생할 가능성이 있어 이 유도 과제로 수집된 자료는 폐기하였다. 이에 따라 최종적으로 전북과 강원 지역의 각 13쌍 및 경기 지역의 9쌍에게는 17개의 유도 과제로, 광주·전남 지역의 6쌍에게는 새로 개발된 4개의 유도 과제만을 활용하여 한국수어 말뭉치를 수집하였다. 이후 폐기된

1종의 유도 과제를 대체하는 유도 과제 1종을 보완하여 5종의 신규 유도 과제 개발을 완수하였다. 또한 2015년에 구축한 서울지역 원시 수어 말뭉치 중 H과제 약 123분 분량의 영상에 대해 기본 주석과 의미 주석을 실시하였으며, 2019년 구축한 경남 지역 원시 말뭉치 중 H과제 약 100분 분량의 영상을 번역하였다(국립국어원, 2021).

2022년도 한국수어 원시 말뭉치 수집은 미 구축지역인 대전·충남 지역에서 이루어졌으며, 대전 지역 5쌍과 충남지역 11쌍, 총 16쌍 32명의 언어제공자로부터 58시간 분량의 영상을 수집하였다. 또한 전남, 경남, 강원 3개 지역의 4개 파일, 총 53분 분량의 ‘H-주제 영역’ 자료에 대해 기본 주석을 하였다(국립국어원, 2022).

이처럼 8회에 걸쳐 시행된 한국수어 말뭉치 구축 사업은 불모지와 같았던 한국수어 연구 생태계에 과학적이고 체계적인 한국수어 연구 기반을 갖추 수 있게 하였고 소수의 주석 인력이나마 확보함으로써 수어 연구 및 교육을 위한 한국수어 말뭉치 자료를 제공할 수 있게 되었다.

○ 한국수어 원시 말뭉치 추가 수집을 통한 지역적 균형 유지와 한국수어의 대표성 확보

한국수어는 음성언어의 방언과 동일한 이유로 특정 지역에서 사는 사람들끼리 수어가 분화되어 지역별로 변이된 형태의 수어가 존재하고 있어 원시 수어 말뭉치 수집 지역을 확대하여 지역별 수어 변이형 수집이 균형을 이루도록 할 필요가 있다. 한국수어 원시 말뭉치는 2015년 서울 지역 농인 30쌍을 시작으로, 2018년에는 전북, 전남, 강원, 경남 지역 농인 14쌍을 대상으로 수어 말뭉치를 수집하였다. 2019년에는 2018년에 2쌍밖에 수집하지 못한 경남 지역 농인의 수어 자료를 보완하기 위해 추가로 11쌍을 대상으로 경남 지역 원시 말뭉치를 수집하였으며, 2020년에는 전남·광주 지역 농인 19쌍을 대상으로 수어 말뭉치를 추가 수집하였다. 2021년에는 수집된 자료의 양적 균형을 이루기 위해 2018년에 4쌍씩만 수집하였던 전북과 강원 지역에서 13쌍, 2015

년, 서울 지역 수집 시 2쌍밖에 수집하지 못한 경기 지역에서 9쌍, 광주 지역에서 6쌍, 모두 41쌍을 대상으로 수어 말뚱치를 추가 수집하였다. 2022년에는 대전·충남 지역 농인 16쌍을, 2023년에는 충북 지역 농인 16쌍을 대상으로 한국수어 말뚱치를 수집하였다.

이러한 지역별 균형을 고려한 원시 수어 말뚱치 수집은 한국수어의 지역별 변이를 밝힐 수 있게 할 것이다. 아울러 지역적 균형을 이루어 수집된 한국 원시 수어 말뚱치들은 우리나라 수어의 언어적 특성을 대표하는 정당성을 갖게 될 것이다. 이런 맥락에서 17개 시·도 중 아직 한국수어를 수집하지 못한 대구·경북 지역과 제주 지역의 수어도 시급히 수집함으로써 지역적 균형을 완성할 필요가 있다.

○ 기 구축된 한국수어 말뚱치 공개를 위한 정비 실시

한국수어 말뚱치가 한국수어 연구의 기초 자료로 사용되기 위해서는 공개가 이루어져 모든 사람이 한국수어 말뚱치 접근이 가능해야 한다. 이를 위해 기 구축된 한국수어 말뚱치 중에서 공개에 적합한 한국수어 말뚱치를 선정하는 기준과 지침을 마련하여야 한다. 수집된 한국수어 말뚱치에는 원시 한국수어 말뚱치 상태로 있는 것들과 주석이 되어 있는 한국수어 말뚱치, 그리고 주석과 번역이 모두 되어 있는 한국수어 말뚱치 자료들이 있어 이 자료들에 대한 공개 적합성을 판단하고, 공개가 적합하다고 판단된 한국수어 말뚱치들에 대해 기본 주석과 번역 주석 정비를 실시할 필요가 있다.

○ 말뚱치 공개 및 홍보를 위한 동영상 제작

공개되는 한국수어 말뚱치는 국립국어원의 ‘모두의 말뚱치’를 통해 공개될 예정이다. 한국수어 말뚱치에 대한 수어 언중과 연구자들의 접근이 용이하도록 할 필요가 있다. 이를 위해 한국수어 말뚱치를 공개·배포하는 국립국어원 ‘모두의 말뚱치’의 회원가입 및 말뚱치 신청

방법 안내 수어 동영상, ‘모두의 말뚱치’ 내 한국수어 말뚱치 소개 자료 수어 동영상, 한국수어 말뚱치 수집과 주석 과정 등의 전 과정이 담긴 홍보 영상을 제작함으로써 공개되는 한국수어 말뚱치가 수어 연구를 위한 기초 자료로 널리 활용될 수 있도록 해야 할 것이다.

2) 연구 목적과 과업

○ 본 연구의 목적과 과업은 다음과 같다.

첫째, 충북 지역의 농인 15쌍을 대상으로 한국수어 원시 말뚱치를 수집한다.

- 수집 과정에서 발생할 수 있는 문제에 대비하여 언어제공자 1쌍을 추가한 16쌍(32명)을 대상으로 수집
- 18종의 유도 과제를 활용하며 일부 과제는 지역 특성을 반영하여 내용 수정·보완
- 수집된 언어 자료를 총 40시간 이상의 원시 말뚱치 형태로 편집 가공

둘째, 구축된 한국수어 말뚱치 공개를 위한 정비를 실시한다.

- 정비 및 선정 기준 마련
- 공개용 한국수어 말뚱치 360분 선정
- 기본 주석 및 번역 주석 정비

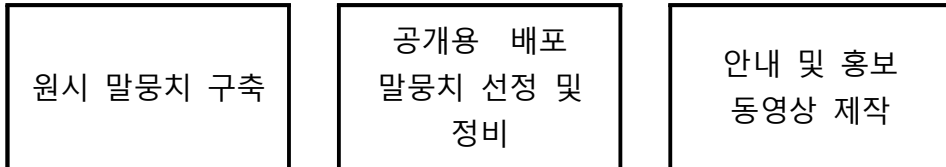
셋째, 한국수어 말뚱치 안내 및 홍보 영상을 제작한다.

- ‘모두의 말뚱치’ 회원가입 및 말뚱치 신청 방법 안내 수어 동영상
- ‘모두의 말뚱치’ 내 한국수어 말뚱치 소개 자료 수어 동영상
- 한국수어 말뚱치 수집과 분석 과정을 담은 홍보 동영상

2. 연구 방법

1) 연구 내용

- 본 연구의 내용은 크게 세 가지 영역에서 진행되었다.



2) 연구 방법

(1) 충북 지역 한국수어 원시 말뭉치 구축

- 17개 시·도 중 한국수어 원시 말뭉치 구축이 되지 않은 충북 지역을 대상으로 하였다.
- 2021년 새로 개발한 유도 과제 5종을 포함한 총 18종의 유도 과제를 활용하였으며, 지역의 특성과 시의성을 고려하여 유도자료를 수정·보완하였다.
- 자료 수집 시의 기계적 오류, 언어제공자의 당일 심리 및 신체적 상태 등의 돌발적 상황에 따른 수집 실패의 위험에 대비하여 1쌍의 예비 언어제공자를 추가하여 구축하였다.

(2) 공개를 위한 한국수어 말뭉치 정비

- 기 구축된 한국수어 말뭉치 300시간 8분 분량에 대한 공개 가능 여부 판단을 위한 검토 기준 설정
- 한국수어 말뭉치 중 기본 주석과 번역이 되어 있는 영상에 대한 공개 가능 여부를 결정

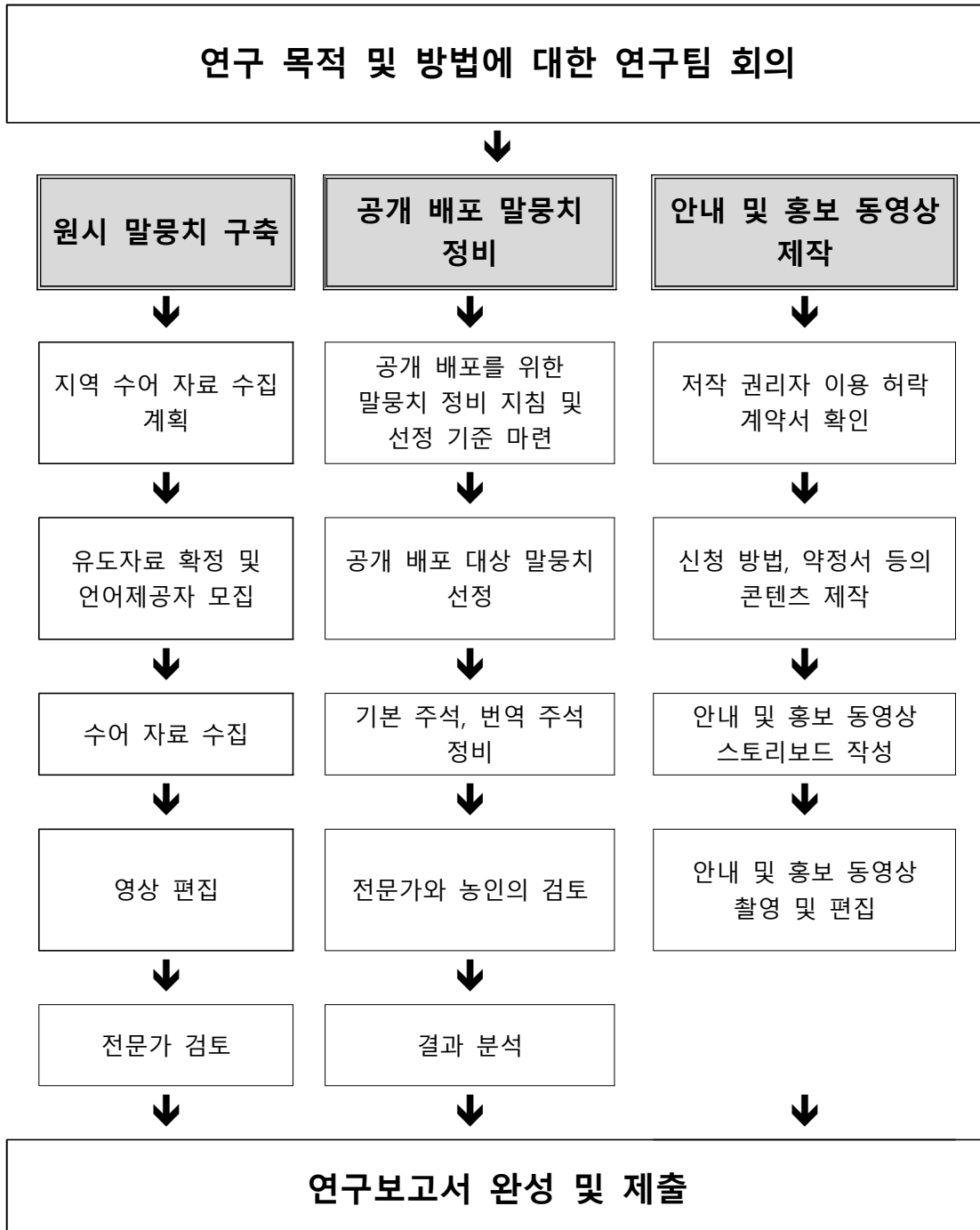
- 지역, 성별, 나이를 고려하여 공개를 위한 한국수어 말뭉치 최종 선정
- 주석의 수정 및 보완은 원래 주석이 실시된 층렬에서 하고, 수정 보완한 내용은 정비 층렬에 12가지 통제 어휘를 사용하여 기록
- 번역 정비는 1차 자문회의를 거쳐 시범 번역을 실시
- 시범 번역 결과를 바탕으로 번역 정비 지침 작성하고 2차 자문회의를 거쳐 확정
- 직역과 의역으로 나누어 번역하고, 직역은 수어의 언어적 구조 특성이 잘 드러나게 하는 것에 주안점을 둠
- 의역은 도착어인 한국어가 가능한 한 자연스럽게 정확하게 전달되도록 하는 데 주안점을 둠. 단, 2023년에는 의역 정비는 하지 않고 그 기준을 마련함

(3) 한국수어 말뭉치 공개 및 홍보를 위한 수어 영상 제작

- 한국수어와 한국수어 말뭉치에 대한 이해가 있는 ‘수어 민들레’를 제작 업체로 선정
- 1차에서 3차 회의에서 1편, 2편, 3편의 영상 콘셉트 설정
- 필요한 경우 음성을 넣되 한국수어와 자막으로 내용을 전달
- 1편, 2편, 3편에 대한 연구진과 업체 간 협업과 국어원의 피드백으로 스토리보드 작성
- 스토리보드를 기초로 1편과 3편 영상을 우선 제작
- 2편은 내용의 성격상 주석 정비와 번역 정비가 모두 마쳐진 다음 제작
- 1편과 2편은 설명을 중심으로, 3편은 다큐 형식으로 제작

3) 연구 절차

본 연구는 과업별로 다음과 같은 과정으로 수행되었다.



[그림 1-1] 연구 절차

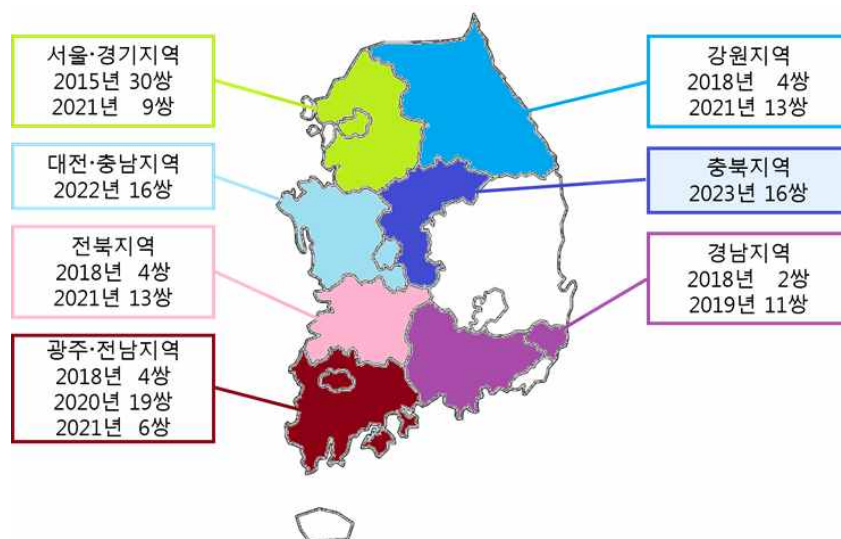
4) 과업별 추진 일정

내용	2023							
	5	6	7	8	9	10	11	12
o 연구계획 수립								
o 한국수어 원시 말뭉치 구축								
- 사전 준비(기획 및 협의)								
- 언어제공자 모집 및 사회자 교육								
- 촬영 및 편집								
- 검수 및 수정								
- 시스템 탑재(메타정보, 영상)								
o 공개 한국수어 말뭉치 정비								
- 정비 지침 및 선정 기준 마련								
- 기본 주석, 번역 정비								
- 주석, 번역 결과 분석								
o 공개 및 홍보 동영상 제작								
- 사전 준비(기획 및 협의)								
- 촬영 및 편집								
- 시스템 탑재(영상)								
o 과업별 점검(중간, 최종)								
o 연구 보고서 완성 및 제출								

II. 한국수어 원시 말뭉치 수집

한국수어 원시 말뭉치는 2015년 서울(경기)지역 농인 30쌍을 대상으로 처음 구축되었다. 그 후, 2018년 전북 4쌍, 전남(광주 포함) 4쌍, 강원 4쌍, 경남 2쌍, 총 14쌍의 농인을 대상으로 네 개 지역에서 수집하였다. 수집된 원시 말뭉치를 분석하거나 자료로 활용하기 위해서는 다양한 지역에서의 수집과 함께 수집된 자료량의 균형 또한 중요하다. 이에 2019년에는 2쌍만 수집된 경남 지역에서 11쌍을 수집하였으며, 2020년에는 전남(광주 포함) 지역에서 19쌍을 대상으로 수집하였고, 2021년에는 전북과 강원 지역에서 각각 13쌍, 경기 지역에서 9쌍, 광주·전남 지역에서 6쌍을 수집하여 총 7개 지역에서 총 115쌍(230명)의 한국수어 원시 말뭉치를 수집하였다. 2022년에는 대전·충남 지역에서 대전 지역 5쌍, 충남 지역 11쌍, 총 16쌍을 대상으로 수집하였다.

2023년 연구에서는 현재까지 수어 수집이 이루어지지 않은 지역 중 충북지역에서 16쌍을 대상으로 18종의 유도 과제를 활용하여 수집하였다. 2015년부터 2023년까지 한국수어 말뭉치 수집 현황을 제시하면 그림 II-1과 같다.



[그림 II -1] 한국수어 말뭉치 언어 자료 수집 현황

1. 언어 자료 수집 절차

충북 지역 한국수어 말뭉치 자료 수집은 [그림 II-2]와 같이 추진하였다.

~ 7월 3주	촬영 및 편집 업체(지역 농아인협회) 선정 유도자료 수정 보완 및 제작
~ 7월 4주	사회자 교육 언어제공자 계약서 동영상 촬영 조건에 맞는 언어제공자 모집 및 검토
~ 7월 5주	언어제공자 최종 확정(16쌍, 32명)
~ 8월 1주	촬영장 세팅 및 사전 점검
8월 2주 ~ 8월 3주	촬영
8월 4주 ~ 10월 4주	동영상 편집, 검수 및 수정
11월 1주 ~ 11월 3주	최종본 검토

[그림 II-2] 한국수어 말뭉치 자료 수집 절차

2. 유도자료

수어를 연구하거나 수어 말뭉치를 구축하기 위해서는 실제 사용하는 수어를 자연스러운 상황에서 다양하게 수집하여야 한다. 이를 위해 한국어를 배제한 수어 동영상, 사진이나 그림, 영상자료 등의 유도자료를 사용한다. 매년 지역을 달리하여 그 지역의 원시 수어 말뭉치를 수집해야 하는 특성상 유도자료는 2015년도에 개발된 유도자료를 바탕으로 수집 지역의 특성과 시의성 등을 고려하여 매년 수정·보완하여 사용하였다. 특히, 구축된 한국수어 말뭉치가 시기별, 지역별로 변이형을 연구할 수 있다는 점에 유념하여 각 과제의 목적에 부합하도록 수정하였다. 본 연구에서는 2021년에 새롭게 개발한 유도 과제 N - 가장 기억에 남는 날, O - 취미, P - 우리 고장 이야기, Q - 수어 어휘 찾기, R - 농예피소드 5종을 추가하여, 총 18개의 유도 과제를 활용하였으며, 그 내용은 다음과 같다.

과제 A - 수어 이름 말하기

과제 B - 준비해 온 농담

과제 C - 다시 이야기하기(애니메이션)

과제 D - 수어 보고 다시 이야기(넥타이)

과제 E - 토론

과제 F - 표지판

과제 G - 특정 사건 묘사

과제 H - 주제 영역

과제 I - 개별 수어의 다양한 표현

과제 J - 농사회 행사

과제 k - 그림 이야기 설명(홍부와 놀부)

과제 L - 옛날 수어와 현재 수어

과제 M - 농학교 이야기

과제 N - 가장 기억에 남는 날

과제 O - 취미

과제 P - 우리 고장 이야기

과제 Q - 수어 어휘 찾기

과제 R - 농에피소드

1) 유도자료

(1) 과제 번호 A: 수어 이름 말하기

이 과제는 언어제공자들이 자신의 수어 이름과 그 수어 이름이 어디서 유래되었는지에 대해 설명하는 것이다. 농문화에서 대표적인 것 중 하나가 수어 이름이라 할 수 있다. 따라서 이 과제는 언어제공자가 평소에 자주 설명하는 자신의 수어 이름을 소개함으로써 어색한 분위기를 풀고 자연스럽게 대화로 들어가도록 하는 데 유용한 과제이다.

(2) 과제 번호 B: 준비해 온 농담

이 과제는 사전에 언어제공자에게 연락하여 촬영 당일 다른 언어제공자에게 보여줄 만한 재미있는 이야기를 준비해 오도록 하여 진행하는 과제이다. 이 과제는 언어제공자들이 미리 준비해 온 이야기이므로 편안한 마음으로 자신감 있게 표현할 수 있다는 장점이 있다. 또한 농사회에서 많이 이야기되는 농담이나 유머가 나오므로 농문화의 일면을 볼 수도 있다.

(3) 과제 번호 C: 다시 이야기하기

다시 이야기하기는 한 명의 언어제공자가 주어진 영상자료를 보고 그 내용을 상대 언어제공자에게 수어로 설명해 주는 과제이다. 이 과제에서는 언어제공자가 동작만 있는 무성 만화영화 ‘Tweety & Sylvester’를 보고 그 내용을 상대방에게 이야기하도록 한다. 언어 없

이 행동으로만 된 장면이므로 분류사 수어 등 수어 본연의 특색이 드러나는 자료를 끌어낼 수 있다는 장점이 있다. 또한 대부분의 나라에서 유도자료로 사용하고 있으므로 다른 나라의 말뭉치와 비교가 가능한 자료이다.

이 과제는 총 3분 45초 분량의 영상으로 2015년 최초 제작 당시 6분 가량의 영상을 사용하였으나, 영상이 길어 언어제공자가 지루해하는 경향이 있어 2019년 한국수어 원시 말뭉치 수집부터 반복되거나 비교적 덜 중요하다고 판단되는 부분들을 편집하여 사용하였다. 처음에는 전체 영상을 보여주고, 두 번째는 3분 45초 분량의 영상을 7개의 영상으로 잘라서 제시하였다. 이는 언어제공자가 전체 영상을 한 번만 보고 상대방에게 설명하기에는 내용이 너무 많아 부담을 줄 수 있고, 전체 이야기에 대해 설명을 하다 보면 세부적인 내용이 누락되어 과제의 목표가 이루어지지 않기 때문이었다. 또한 각각의 과제 영상 사이에 검은색 슬라이드를 삽입해서 언어제공자가 설명할 때는 모니터에 검은색 슬라이드만 보이도록 하였다. 이는 수어를 할 때 언어제공자의 눈의 움직임을 촬영하기 위하여 언어제공자의 시선이 모니터에 가지 않도록 하여 상대방만을 바라보고 영상자료의 내용을 설명하도록 하기 위함이었다.

(4) 과제 번호 D: 수어 보고 다시 이야기하기

수어로 주어지는 영상자료를 보고 이야기를 하는 과제로 유도자료 중에 유일하게 수어로 제시되는 과제이다. 이 과제는 한국농아방송의 Deaf UCC ‘나라별 넥타이 고르는 법’의 영상 중에서 일부분을 추출하여 사용하였다. 영상의 길이는 총 2분 21초이다. 2015년에 제작한 유도자료의 경우, 제시된 전체 영상의 내용이 언어제공자가 영상의 내용을 다시 이야기하는 데 부담이 될 정도로 길어 2분 21초의 전체 영상을 3개의 영상으로 나누어 제시하도록 수정하였다.

(5) 과제 번호 E : 토론

과제의 목적은 수어 언어제공자가 자신이 표현하는 수어의 형식보다는 치열한 토론 내용에 집중하게 하여 자연스러운 수어 산출을 유도하는 것으로 다양한 주제를 제공하였다. 토론 주제는 ‘찬반 논쟁’이 가능한 주제로 하고, 언어제공자의 연령대나 지역 특성에 따라 선호하는 주제가 다를 수 있으므로 여러 주제 중에서 선택할 수 있도록 하였다. 또한 언어제공자들의 적극적인 참여를 유도하기 위해 다양하고 시의성이 있는 주제들로 구성하였다.

두 명의 언어제공자가 제시된 여러 주제 중 각자 하나씩 고른 2개의 주제를 갖고 차례대로 토론을 진행하였다. 언어제공자에게 주제를 제시할 때는 한국어 문장으로 하지 않고 수어 동영상으로 하였다. 그러나 수어 동영상만 제시함으로써 언어제공자가 느낄 주제 선정의 어려움을 해소하고자 언어제공자가 토론 주제 전체를 한눈에 볼 수 있도록 마지막 페이지에 전체 주제를 목록화하여 제시하였다.

‘E-토론’ 과제의 주제는 2015년에는 ① 1종 면허 농, ② 농학교 농인 교사, ③ 한국수어 통역사협회 설립, ④ 메르스 등 4가지였으며, 2018년에는 ③ 한국수어 통역사협회 설립과 ④ 메르스가 시의적절하지 않아 ③ 수어통역사 센터 분리 ④ 폭염으로 교체한 바 있다. 2019년에는 ① 1종 면허 농, ② 농학교 농인 교사, ③ 수어통역센터 분리, ④ TV 수어통역 대 문자통역, ⑤ 남북통일, ⑥ 일회용품 사용 ⑦ 수어 교원 자격증 문제로 진행하였고, 2020년에는 코로나 19와 관련하여 코로나 개학 연기를 추가하여 진행하였으며, 2021년에는 시의성을 고려하여 ‘코로나 개학 연기’ 주제를 ‘코로나 백신 접종’ 주제로 변경하여 진행하였고 2022년에는 ‘수어교원 자격증 문제’와 ‘코로나 백신 접종’ 주제를 ‘듀크족과 덩크족’, ‘장애인 감형 문제’로 변경하여 진행하였다. 이번 연구에서는 ‘수화통역센터 통합유지, 분리’, ‘수어방송, 자막방송’, ‘일회용품 사용반대’ 주제를 삭제 후 진행하였다.

(6) 과제 번호 F: 표지판

세계의 표지판을 모아서 제시하고 그 의미를 유추하게 하여 서로 이야기를 나누도록 하는 과제이다. 이 과제에서 제시하는 표지판은 다양한 유추가 가능하므로 언어제공자들의 풍부한 대화를 유도할 수 있었으며, 특히 금지어 등 부정문을 유도하는 데 유용하였다.

이번 연구에서는 기존에 사용하던 그림 자료가 추후 저작권 문제가 발생할 수 있는 점을 고려하여 전문가에 의뢰하여 그림을 제작하였다.

(7) 과제 번호 G: 특정 사건 묘사

이 과제는 제시된 사건들 중에 하나의 사건을 선택하여 설명하고, 그당시에 본인이 무엇을 하고 있었는지를 이야기하는 과제이다. 언어제공자는 다양한 연령대로 구성되어 있어 태어난 시대에 따라 특정 사건에 대한 정보를 모를 수도 있기 때문에 오래된 사건이라도 기억할 만한 역사적 사건을 제시하였다.

언어제공자들의 연령을 고려하여 1970년대부터 2010년대 사이에 일어난 역사적인 사건들을 시대별로 골고루 구성하였다. 이 과제는 사진으로 제시되는데, 사건의 대표적인 사진을 제시하여 해당 사건을 쉽게 파악할 수 있도록 하였다. 언어제공자는 사건을 묘사하며 사건이 일어난 당시 본인이 무엇을 하고 있었는지를 설명하게 했다. 이 과제를 통해 언어자료 수집 이외에도 간접적으로 농인이 어떻게 정보를 접하게 되는지 확인할 수 있었다. 두 명의 언어제공자가 제시된 사건 중 각각 하나의 사건을 선택하여 교대로 이야기를 나누었다. 언어제공자의 참여를 유도하기 위해 지역성을 고려하여 해당 지역 사건을 추가할 수 있게 했으며, 이전 지역 사건은 삭제하였다.

‘G-특정 사건 묘사’ 과제를 수행하기 위해 지난 2015년에는 ① 1974년 육영수 여사 피살 사건, ② 1995년 삼풍백화점 붕괴, ③ 2002년 월드컵 스페인전, ④ 2008년 남대문 화재 사건, ⑤ 2009년 노무현 대통령 자살 사건, ⑥ 2014년 세월호 침몰을 제시하였다. 그러나 2018년에

는 위의 사건들이 10년 이상 오래된 사건이므로 세월호 사건을 제외하고 전면 교체하여 사용한 바 있다. 2019년에는 ① 1974년 육영수 여사 피살 사건, ② 1980년 5·18 민주화 운동, ③ 1995년 삼풍백화점 붕괴, ④ 2002년 월드컵 스페인전, ⑤ 2009년 노무현 대통령 자살, ⑥ 2014년 세월호 침몰, ⑦ 2017년 행복팀 사건, ⑧ 2018년 남북 정상회담, ⑨ 2019년 안인득 사건으로 구성하였으며, 2020년에는 2019년 경남 지역에서 사용한 ‘2019년 안인득 사건’을 제외하고 수집 지역인 전남 지역의 지역적 특성을 반영하여 ‘2011년 도가니 사건’으로 대체하였으며, 2021년에는 강원 지역의 큰 행사였던 ‘2018년 평창 동계 올림픽’을 추가, 2022년에는 ‘2018년 평창 동계올림픽’을 삭제하고 충남 지역의 큰 사건이었던 ‘천안함 사건’과 ‘태안 기름 유출 사건’을 추가하여 총 11개의 사건 주제를 제시하였다. 이번 연구에서는 충북 지역의 큰 사건이었던 ‘오송 지하철도 차량 침수 사건’을 추가하여 총 12개의 사건 주제를 제시하였다.

(8) 과제 번호 H : 주제 영역

이 과제는 사전 제작에 필요한 어휘를 수집하기에 유용한 과제로 언어제공자가 제시된 주제에 대해 자유롭게 대화를 나누는 과제이다. 주제는 날씨, 경제(은행, 증권, 노후 등), 감정, 직업, 가족과 친척, 시간(일정 등), 음식, 정치, 문화생활, 행사(잔치), 의사소통(전화, 스마트폰 등), 법, 도시와 시골, 주택과 아파트, 공공기관, 휴가(여행, 자유시간), 패션, 종교, 외모, 건강, 사랑(성, 결혼), 교통, 계절 등이었다. 2021년 연구에서는 2020년까지 수집된 원시 말뭉치 자료를 확인하여 언어제공자가 가장 많이 선택했던 주제인 ‘도시와 시골’, ‘의사소통’, 신규 유도 과제로 개발된 것과 유사한 ‘운동과 취미’, ‘관광명소’ 등은 제외하고 총 16개의 주제를 선정하였다.

이번 연구에서는 선정한 16개의 주제 중 제시된 그림이 저작권에 민감할 수도 있는 점을 고려하여 ‘대중매체’를 제외하고 총 15개의 주

제를 3개의 언어제공자 모둠을 나누어 다르게 제시하였다. 언어제공자 16쌍 중 1번 모둠 6쌍에게는 날씨, 직업, 종교, 휴가, 음식 주제를, 2번 모둠 5쌍에게는 경제, 여행, 연애와 결혼, 건강, 패션 주제를, 3번 모둠 5쌍에게는 감정, 계절, 교통, 행사, 부동산 주제를 묶어서 제시하였다. 각 언어제공자 쌍은 자기가 속한 모둠에 제시된 주제 중에서 2개의 주제를 각자 골라 이야기를 나누는 방식으로 진행하였다. 언어제공자들의 대화가 매끄럽게 이어지지 않거나, 무슨 말을 해야 할지 모르는 경우, 사회자가 미리 준비해 놓은 질문을 언어제공자들에게 함으로써 과제가 원활히 진행될 수 있도록 하였다.

〈표 II -1〉 ‘H-주제 영역’ 과제

1번 그룹	날씨	직업	종교	휴가	음식
2번 그룹	경제	여행	연애와 결혼	건강	패션
3번 그룹	감정	계절	교통	행사	부동산

(9) 과제 번호 I: 개별 수어의 다양한 표현

이 과제는 지역 방언의 표현을 유도해 내기 위한 과제이다. ‘가족, 미원, 색깔, 콩나물’ 등 지역별로 다양하게 사용되는 수어를 유도하기 위해 관련 그림을 준비하여 언어제공자에게 보여주고 수어로 표현하게 하였다. 많은 종류의 유도자료 중에 대화가 아니라 단어를 유도해 내는 과제로 다른 지역에서도 이 유도자료를 통해 수어를 수집한다면 지역 방언 연구가 가능할 것이다. 언어제공자의 출신 학교나 고향에 따라 개별적으로 다르게 사용할 것으로 예측되는 ‘7가지 색깔’ (보라색, 분홍색, 빨간색, 회색, 밤색, 주황색, 회색)을 보여주었으며, ‘고모, 이모, 외삼촌’은 가계도 그림을 함께 제시하였다. 그리고 ‘콩나물’과 ‘청소’는 사진으로 제시하였으며, 이외에 지역 방언을 유도하기에 적절하다고 판단된 하위 주제를 사회자가 질문하였다.

(10) 과제 번호 J: 농 사회 행사

농 사회의 주요 행사에 대한 포스터를 제시하고, 3개의 행사 중에 자신이 참석했던 행사 한 가지를 골라서 이야기하도록 하는 과제이다. 이 과제는 ‘농학교 경험 이야기하기’ 과제와 함께 가장 농인다운 대화를 유도할 수 있는 자료로서 언어제공자들에게 편안하고 즐거운 분위기에서 이야기를 끌어낼 수 있다는 장점이 있다. 수집을 위해 제시한 농사회 관련 행사 포스터는 이번 연구의 수집 대상 지역인 충북 지역에서 개최된 것으로 교체하였다.

(11) 과제 번호 K: 그림 이야기 보고 설명하기

이 과제는 ‘흥부와 놀부’ 그림을 본 다음 한 명의 언어제공자가 상대방에게 ‘흥부와 놀부’ 이야기를 설명하는 과제이다. 한국 사람에게 ‘흥부와 놀부’는 매우 친숙한 이야기이므로 언어제공자에게 큰 부담을 주지 않는다. 총 10장의 그림으로 구성하였는데 한 슬라이드에 한 장씩 총 10장의 그림을 보여주고 난 후, 한 슬라이드에 4장의 그림을 넣어 다시 보여주었다. 두 번째로 그림을 보여줄 때는 각 슬라이드 사이에 검은색 슬라이드를 삽입하였는데, 언어제공자는 검은색 슬라이드가 있는 동안 상대방에게 자신이 본 그림 이야기를 설명하였다.

(12) 과제 번호 L: 옛날 수어와 현재 수어

이 과제는 옛날에는 사용했으나 현재는 사용되지 않는 수어나, 옛날과 형태가 달라진 수어를 생각나는 대로 이야기해 보게 하는 과제이다. 이 과제를 통해 사회언어학적 변화에 대한 자료와 메타언어 자료를 수집할 수 있었다.

(13) 과제 번호 M: 농학교 경험 이야기하기

2013년에 실시한 기초 연구에서는 ‘어떤 일(특정 사건)이 발생했을

때, 본인이 무엇을 했는지’를 이야기하는 과제였었는데, 언어제공자의 경험을 끌어내기가 어려웠다. 따라서 본 연구에서는 ‘농학교에서의 자기 경험’을 이야기해 보는 과제로 수정하여 제시하였다. 농학교는 농문화를 접했던 핵심적인 곳으로 언어제공자들이 오랜 시간의 촬영을 통해 지쳐갈 때쯤 농학교에서의 경험을 이야기하는 과제를 제시함으로써 즐겁게 자기 경험을 이야기할 수 있도록 하였다. 과제는 별도의 그림 자료 없이 수집하고자 하는 지역의 사회자가 과제 설명 동영상을 제작하여 제시하였다.

(14) 과제 번호 N: 가장 기억에 남는 날

이 과제는 2021년 새로 개발된 유도 과제로 언어제공자가 지금껏 살아오면서 겪었던 일 중 가장 기억에 남는 날에 대해 이야기하는 과제이다. 명절, 사고, 결혼, 장례식, 출산, 졸업, 밸런타인데이, 생일, 성탄절 등의 그림 자료를 제시함으로써 가장 기억에 남는 날을 떠올리도록 유도하였으며, 지역의 사회자가 과제 설명 동영상을 통해 그림 자료로 제시되지 않은 다양한 기억에 대해서도 자유롭게 이야기할 수 있도록 하였다.

(15) 과제 번호 O: 취미

취미를 주제로 언어제공자 간에 대화를 진행하는 과제이다. 취미와 관련된 총 45개의 그림 자료를 유형별로 6개 그룹으로 유목화하여 제시하였으며, 현재 가지고 있는 취미뿐 아니라 과거에 가지고 있었던 취미, 미래에 가져보고 싶은 취미 또는 경제적으로 여유가 생기면 갖고 싶은 취미 등에 대해 자유롭게 이야기할 수 있도록 하였다.

(16) 과제 번호 P: 우리 고장 이야기

수집 지역에 10년 이상 거주한 언어제공자가 자기 지역을 주제로 하여 이야기하는 과제이다. 지역의 유명한 농산물, 음식, 축제, 풍습, 인

물, 관광지 등을 이야기함으로써 해당 지역 수어 및 지역의 고유 명사 등을 수집하는 데 목적이 있다. 과제는 해당 지역의 관광 관련 홈페이지에서 제공하고 있는 지역의 지도와 관광지, 축제 등의 사진을 하나의 슬라이드에 넣어 제시하였다.

(17) 과제 번호 Q: 수어 어휘 찾기

명확한 수어 어휘를 수집하기 위한 유도 과제이다. 신체 부위, 동물, 채소, 가전 및 가구, 사무용품, 주방용품, 나라와 관련된 어휘 그림을 슬라이드에 넣어 제시하였고, 나라의 경우 해당 국가의 그림을 넣어 제시하였다. 개인이 사용하는 수어 어휘를 표현하도록 유도하였고 그 수어가 맞는지, 틀렸는지를 판단하는 것이 아니라고 설명하였다. 또한 해당 수어가 없는 경우, 억지로 수어를 만들어서 표현하지 않도록 하여 최대한 자연스러운 의사소통 상황에서 사용되는 어휘를 수집할 수 있도록 하였다.

(18) 과제 번호 R: 농에피소드

농인으로서 겪는 다양한 경험과 농문화에 대한 이야기를 유도하는 과제이다. 이 과제는 첫째, 농인으로 겪었던 억울하거나 황당한 사건, 둘째, 농인에 대한 사회의 편견 때문에 겪은 경험, 셋째, 농인이어서 좋은 점이라는 3개의 하위 주제로 구성하였다. 별도의 그림 자료 없이 사회자의 설명 동영상으로 3가지 하위 주제에 대한 질문과 사례를 중심으로 설명하였다.

2) 유도자료 설명 동영상

확정된 유도자료와 더불어 과제를 수행하기 위한 별도의 설명이 필요하다. 그러나 촬영일을 기준으로 사회자의 심리적, 신체적 상태나 다양한 환경적 요인에 따라 유도자료에 대한 설명이 달라질 수 있다.

이에 따라 유도자료에 대한 설명이 모든 언어제공자들에게 동일하게 제공될 수 있도록 사전에 유도 과제 설명 영상을 제작하였다. 또한 같은 과제를 수행하더라도 각각의 언어제공자에게 서로 다른 역할이 부여되는데, 동영상 방식의 경우 각각의 언어제공자에게 동시에 다른 설명이 가능하다는 점에서 동영상 형태의 설명 방식이 유용하다. 언어제공자들에게서 지역 수어를 효과적으로 유도해 내기 위해 사회자를 해당 지역에 거주하는 농인으로 선정하였고, 유도자료 설명 동영상도 같은 이유로 해당 지역의 사회자가 촬영하도록 하였다. 수집 대상 지역이 변경되는 경우, 이에 따라 유도자료 설명 동영상을 다시 제작하였다.

이번 연구에서는 충북 지역에서 각각의 설명 동영상을 제작하였으며, 제작한 설명 동영상은 기존 유도자료의 틀과 동일한 형식으로 편집하여 수집에 활용하였다.

3. 언어제공자 모집 및 선정

1) 언어제공자 선정 조건

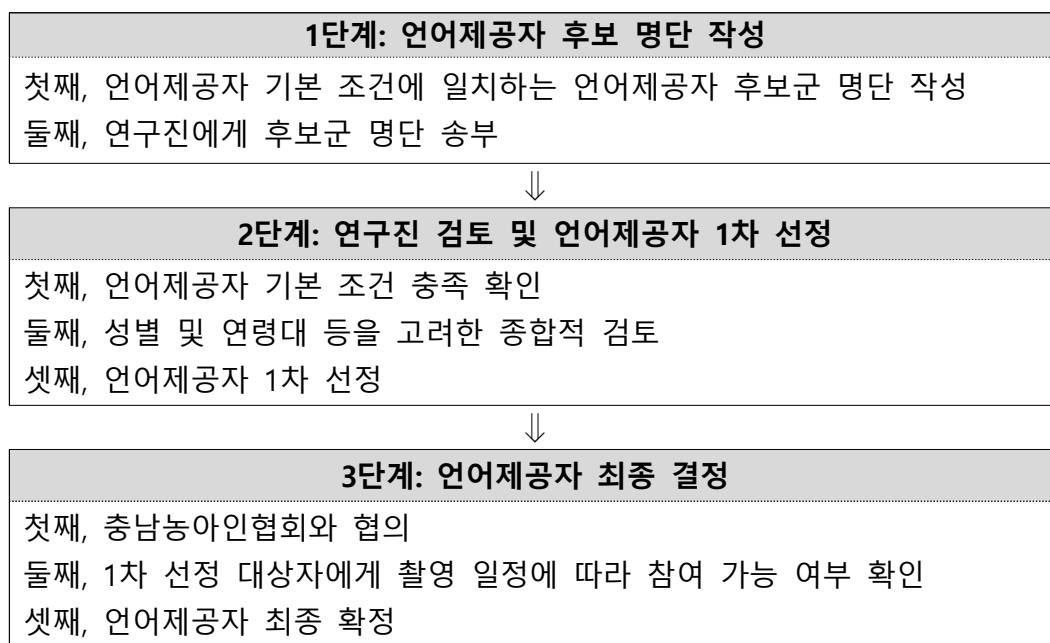
언어제공자 선정은 사전에 연구진에서 충북농아인협회에 ‘언어제공자 선정 기준’을 제공하고, 이에 적합한 언어제공 후보자를 협회가 연구진에게 추천하였다. 추천된 후보자 중에서 연구진과 협회가 협의하여 지역 내 농학교 졸업 여부, 성별, 나이대 그리고 사는 지역을 고려하여 최종 선정하였다. 언어제공자 선정 조건은 <표 II-2>과 같다. 특히, 농부모 혹은 농형제, 농배우자가 있는 농인이거나 초등학교 이전에 수어를 습득한 경우에 우선 선정하였다.

<표 II-2> 언어제공자 선정 조건

번호	조 건	가부
1	• 19세 이상 성인 농인	(예, 아니오)
2	• 한국수어를 모어로 하는 농인 - 의사소통을 위해 수어를 가장 편하고 자주 사용함	(예, 아니오)
3	• 농학교 졸업 ※ 부모가 모두 농인이면 일반학교 졸업한 농인도 가능	(예, 아니오)
4	• 현재 충북 지역에서 10년 이상 거주 혹은 직장생활을 하고 있음	(예, 아니오)
5	• 주 3회 이상 농인을 만남	(예, 아니오)

2) 언어제공자 모집

언어제공자 모집은 충북농아인협회의 협조로 진행되었으며, 충북농아인협회가 추천한 언어제공자 후보군 중 언어제공자 선정 조건과 우선 선정 조건을 고려하여 언어제공자와 예비 언어제공자를 선정하였다. 이후 촬영 가능 일정 등을 조정하여 최종 언어제공자를 결정하였다.



[그림 II -3] 언어제공자 모집 및 선정 절차

3) 언어제공자 기본 사항

<표 II -3> 언어제공자의 인적 사항

지역	성명	성별	출생년도	지역	성명	성별	출생년도
충북	최OO	남	1950	충북	이OO	여	1970
충북	정OO	남	1960	충북	오OO	여	1980
충북	이OO	남	1960	충북	김OO	여	1940
충북	정OO	남	1960	충북	서OO	여	2000

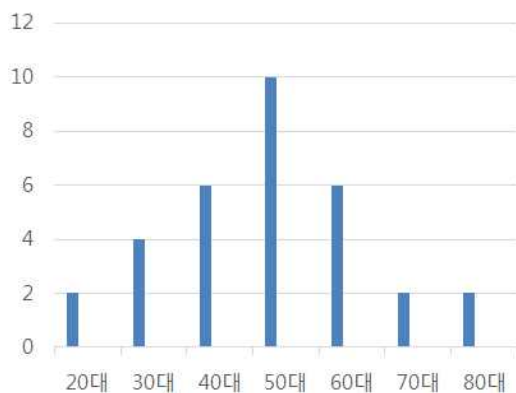
지역	성명	성별	출생년도	지역	성명	성별	출생년도
충북	장OO	여	1950	충북	송OO	남	1960
충북	김OO	여	1960	충북	이OO	남	1940
충북	이OO	남	1940	충북	김OO	여	1950
충북	한OO	여	1940	충북	임OO	여	1950
충북	이OO	남	1970	충북	신OO	남	1970
충북	김OO	남	1960	충북	박OO	여	1980
충북	송OO	남	1990	충북	송OO	남	1960
충북	양OO	여	1980	충북	홍OO	남	1960
충북	김OO	남	1960	충북	정OO	여	1960
충북	김OO	남	1960	충북	서OO	남	1990
충북	박OO	남	1980	충북	정OO	여	1980
충북	한OO	남	1980	충북	임OO	여	1970

(가) 언어제공자 인구통계학적 정보

① 연령

원시 말뭉치의 언어제공자는 50대가 10명으로 가장 많았으며, 다음으로 40대와 60대가 각각 6명으로 많았다.

지역별 언어제공자의 연령 분포는 <표 II-4>와 같다.



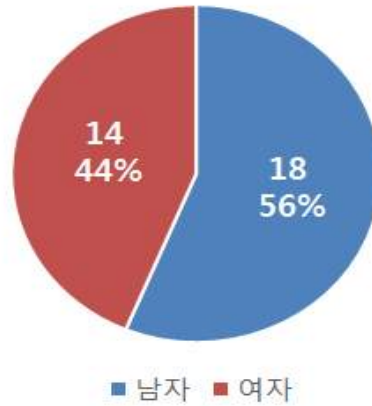
[그림 II-4] 언어제공자 연령

<표 II-4>
언어제공자 연령

연령	충북
20대	2
30대	4
40대	6
50대	10
60대	6
70대	2
80대	2
계	32명

② 성별

전체 언어제공자 중 남성이 56%(18명), 여성이 44%(14명)이었다.

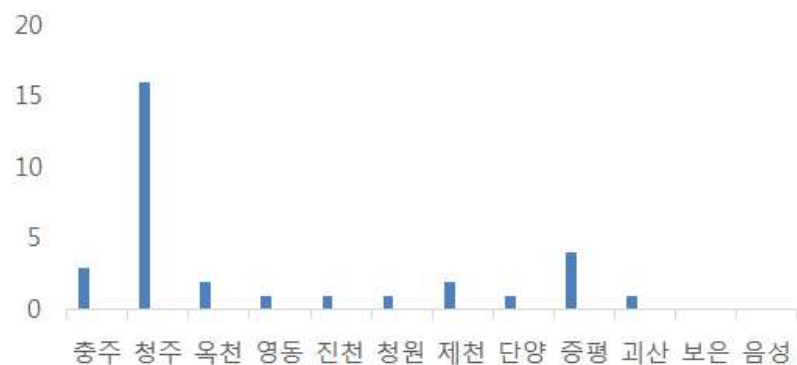


[그림 II -5] 언어제공자 성별

③ 언어제공자 거주 지역

충북 지역 언어제공자의 거주 지역 분포는 <그림 II -6>과 같다.

충북 지역 내 12개 시·군 지역 중 청주지역이 16명으로 가장 많았다.



[그림 II -6] 언어제공자 거주 지역 분포

④ 청각 보조기기 착용 여부

청각 보조기기 착용 여부를 묻는 질문에 ‘보청기 등의 보장구를 사용하지 않는다.’ 라는 답변이 69%(22명)로 가장 많았고, ‘규칙적으로 보청기를 사용한다.’ 라는 답변이 22%(7명), ‘가끔 보청기를 사용한

다.’ 라는 답변이 9%(3명) 순으로 나타났으며, 이번 언어제공자 중에는 인공와우를 착용한 사람은 없었다.

<표 II -5> 언어제공자 청각 보조기기 착용 여부

청각 보조기기 착용 여부		
인공와우	0	0%
보청기-규칙적 사용	7	22%
보청기-가끔 사용	3	9%
사용하지 않음	22	69%
계	32명	100%

⑤ 아버지의 농인 여부

언어제공자의 아버지가 청인인 경우가 81%(26명)로 압도적으로 많았고, 아버지가 농인인 경우는 19%(6명)로 나타났다.

<표 II -6> 아버지의 농인 여부

아버지 농인 여부		
농인	6	19%
난청인	0	0%
청인	26	81%
잘 모르겠다	0	0%
계	32명	100%

⑥ 어머니의 농인 여부

언어제공자의 어머니가 청인인 경우가 아버지의 경우와 마찬가지로 81%(26명)로 압도적으로 많았고, 어머니가 농인인 경우는 19%(6명)로 나타났다.

〈표 II -7〉 어머니의 농인 여부

어머니 농인 여부		
농인	6	19%
난청인	0	0%
청인	26	81%
잘 모르겠다	0	0%
계	32명	100%

⑦ 배우자의 농인 여부

언어제공자의 배우자가 농인인 경우가 75%(24명), 배우자가 없는 경우 19%(6명), 배우자가 청인인 경우 6%(2명)로 나타났다.

〈표 II -8〉 배우자의 농인 여부

배우자 농인 여부		
농인	24	75%
난청인	0	0%
청인	2	6%
CODA	0	0%
배우자 없음	6	19%
계	32명	100%

⑧ 한국수어 습득 시기

언어제공자 중 한국수어 습득 시기에 대한 답변으로 3세부터 8세가 50%(16명)로 가장 많았고, 9세에서 14세가 44%(14명), 15세 이상이 6%(2명)로 나타났다.

〈표 II -9〉 한국수어 습득 시기

한국수어 습득 시기			
3세	1	3%	16명 (50%)
4세	1	3%	
6세	2	6%	
7세	4	13%	
8세	8	25%	
9세	4	19%	14명 (44%)
10세	3	13%	
11세	1	3%	
12세	1	3%	
13세	1	3%	
14세	4	13%	
15세	1	3%	2명 (6%)
16세	1	3%	
계	32명	100%	32명(100%)

⑨ 한국수어 습득 장소

언어제공자의 한국수어 습득 장소에 대한 답변으로 농학교가 91%(29명)로 압도적으로 많았고, 가정/부모님으로부터 배웠다는 응답이 9%(3명)였다. 이전의 다른 지역과 달리 농인 교회나 농아인협회에서 배웠다는 응답은 없었다.

〈표 II -10〉 한국수어 습득 장소

한국수어 습득 장소		
가정/부모님	3	9%
농학교	29	91%
농인 교회	0	0%
농아인협회	0	0%
기타	0	0%
계	32명	100%

⑩ 최종 학력

언어제공자의 최종 학력에 대한 답변으로 고등학교 졸업 34%(11명), 전문대학 졸업 25%(8명), 초등학교 졸업 13%(4명), 기타 13%(4명), 중

학교 졸업 9%(3명), 4년제 대학교 졸업 3%(1명) 순으로 나타났다.

〈표 II -11〉 최종 학력

최종 학력		
초등학교 졸업	4	13%
중학교 졸업	3	9%
고등학교 졸업	11	34%
전공과 졸업	1	3%
전문대학 졸업	8	25%
4년제 대학교 졸업	1	3%
대학원 석사 졸업	0	0%
기타	4	13%
계	32명	100%

4) 언어제공자 질문지

언어제공자가 한글로 제공되는 질문지를 작성하는 데 어려움이 있을 수 있어 촬영을 지원하는 농인 담당자 및 연구진이 질문지의 내용을 수어로 설명하고 언어제공자가 수어로 답을 할 수 있도록 하였다.

5) 촬영 동의

촬영 전 촬영 영상에 대한 이용 동의를 구하기 위하여 사전에 제작해 둔 ‘국가 언어 자원(말뭉치) 구축 및 활용 저작권 등 이용 허락 계약서’에 서명하도록 안내하였으며, 필요한 경우 내용을 이해할 수 있도록 연구진이 수어로 추가 설명하였다. 계약서에 서명한 후 수집에 참여한 언어제공자는 촬영 초반에 언어제공자가 말뭉치 영상 촬영 및 촬영된 영상을 연구에 이용을 허락한다는 내용의 수어 동영상을 본 후, 수어로 동일 내용에 동의한다는 영상을 촬영하였다.

4. 사회자 및 촬영·편집자 교육

1) 사회자의 역할

- 언어제공자에게 파워포인트로 제작된 유도자료를 제시한다.
- 설명 동영상만으로 언어제공자가 과제를 제대로 이해하지 못했을 경우 과제에 대한 설명을 면대면으로 다시 한다.
- 과업 수행 시간을 관리한다.
- 대화가 자연스럽게 진행될 수 있는 분위기를 조성한다.
- 언어제공자의 대화를 중재하고 촉진한다.
- 사회자 일지를 기록한다.

2) 사회자 선정 및 교육

- 사회자 선정: 충북 2명(최00, 노00)을 선정하였다.
- 사회자 교육은 1회의 집합 교육을 한 후에 수시로 질문과 답을 주고받는 등의 보충 교육을 하였다.
- 교육 일정
 - 집합교육: 총 1회 실시
일시: 7/25(화) 10:00~13:00
장소: 한경국립대학교 평택캠퍼스
 - 개별교육: 수시 진행
일시: 수시
장소: 충북농아인협회 및 온라인

○ 교육 내용

(1차)

- 수어 말뭉치와 유도자료의 내용 및 목적에 대한 이해
- 언어제공자가 동영상 설명 자료를 이해하지 못할 시 필요한 보충 설명 방법
- 언어제공자가 자연스러운 분위기에서 편하게 대화할 수 있는 분위기 조성 방법
- 두 언어제공자의 수어가 균형 있게 산출될 수 있도록 조절하는 전략
 - 그 외 사회자가 지켜야 할 태도
 - 언어제공자와 사회자 관계 질문지, 사회자 일지 작성 방법
 - 유도자료 설명 동영상 촬영 시 고려 사항 및 촬영 진행

(수시)

- 언어제공자와 사회자 관계 질문지, 사회자 일지 작성 방법
- 각 과제별 적정 진행 시간 및 조절 전략

3) 유도 과제 설명 동영상 제작

- 19개(동의서 및 유도 과제 18개)의 유도 과제 설명 동영상을 촬영하였다.
- 유도자료 설명 영상 촬영 시에는 사회자 교육 담당자의 서울지역 방언의 간섭 현상이 발생하지 않도록 촬영장 접근을 금지하였다.

4) 촬영 및 편집자 교육

- 2021년까지는 촬영 및 편집 지침에 따라 연구진이 직접 교육을 진행하였으나 촬영 및 편집은 전문적인 지식과 기술이 필요하다.

○ 이에 작년도 수집에 참여한 촬영 및 편집 담당자 중 1명을 강사로 하여 교육을 진행하였다.

○ 경험을 바탕으로 실제적이고 구체적인 기술을 공유하였다.

○ 교육 일정

- 집합교육: 총 1회 실시

일시: 7/25(화) 14:00~16:00

장소: 한경국립대학교 평택캠퍼스

- 개별교육: 수시 진행

일시: 수시

장소: 충북농아인협회 및 온라인

○ 교육 내용

(1차)

- 수어 말뭉치와 유도자료의 내용 및 목적에 대한 이해
- 촬영실 세팅 시 고려 사항
- 촬영 장비(카메라 및 기타 장비) 구비 시 고려 사항
- 촬영 시 주의 사항
- 편집 시 주의 사항

(수시)

- 촬영 장비 세팅 시 주의 사항 및 문제점 해결
- 편집 시 주의 사항

5. 언어 수집 촬영

1) 촬영실 조건 및 촬영 시스템 지침

언어 수집 촬영을 위한 촬영실 조건과 촬영 시스템에 관한 지침은 다음과 같다.

- ① 촬영실의 크기는 39m² (6.3m x 6m) 이상이어야 한다.
- ② 촬영실 창문은 빛을 차단할 수 있는 커튼이나 암막을 설치한다.
- ③ 촬영 중 청인은 촬영실에 들어갈 수 없으므로 촬영 감독과 보조 인력이 촬영을 준비하고 진행 상황을 모니터할 수 있는 별도의 방이 필요하다.
- ④ 촬영 장소는 언어제공자가 오기 편하고 연구진들이 24시간 접근이 가능한 곳이어야 한다.
- ⑤ 두 언어제공자를 각각 촬영할 수 있는 카메라 2대와 두 언어제공자와 사회자를 다 촬영할 수 있는 카메라 1대 등, 총 3대 이상의 카메라가 필요하다.
- ⑥ 카메라의 메모리 용량은 128G 이상으로 하며, 해당 카메라에 지원되는 메모리 카드의 규격인지 확인한다.
- ⑦ 카메라를 전원에 연결하여 사용하며 카메라의 전원 연결이 중간에 끊어지지 않도록 유의한다.
- ⑧ 카메라는 언어제공자들이 방해받지 않고 대화를 할 수 있도록 최대한 눈에 띄이지 않게 설치한다.
- ⑨ 촬영을 위한 조명, 유도자료 제시를 위한 컴퓨터와 모니터의 설치 는 언어제공자 촬영실 <그림 II-7>과 같이 한다.

2) 촬영장 설치

위에 제시된 지침에 따라 설치한 촬영실과 촬영 시스템에 관한 사항은 다음과 같다.

(1) 언어제공자 촬영실

- 촬영실 구비 장비는 카메라 3대, 조명 5대, 노트북 3대, 모니터 5대로 구성하였다.
- 촬영실의 모습은 다음과 같다.



[그림 II -7] 언어제공자 촬영실 배치도

(2) 휴게 공간

- 테이블과 의자를 마련해 두고, 동의서와 질문지를 비치하였다.
- 계약서 및 질문지의 수어 영상 버전을 시청할 수 있도록 노트북을 설치하였다.
- 휴식 시간에 섭취할 수 있는 다과를 준비해 두었다.



[그림 II -8] 언어제공자 휴게 공간

(3) 모니터링 공간

- 3대의 카메라와 연결된 모니터를 설치하였다.
- 언어제공자 촬영 모습을 관찰할 수 있도록 하였다.



[그림 II -9] 연구진 모니터링 공간

3) 촬영 설명서

언어제공자의 수어 사용에 영향을 줄 수 있으므로 촬영 전과 촬영 중 쉬는 시간에는 가급적 언어제공자가 청인이나 다른 지역 농인(연구진 포함)과 대화를 나누지 않도록 하였다.

(1) 촬영 전

사회자	1. 언어제공자와 사회자 간의 관계에 관한 질문지 작성 2. 준비물 확인 - 사회자 일지, 타임 벨, 탁상시계, 핸드폰, 펜, 물병, 색종이
촬영자	1. 촬영실의 조명과 에어컨을 켜고 햇빛이 들어오거나 기타 문제 사항이 없는지 확인 2. 촬영 준비 1) 컴퓨터 및 모니터 정상 작동 여부 확인 2) 카메라 정상 작동 여부 확인 3) 언어제공자가 카메라 안에 다 들어오는지 확인 (언어제공자마다 의자의 높이를 조절하거나 카메라의 각도를 조절함)
협회 담당자	1. 당일 촬영할 언어제공자에게 문자나 영상으로 연락 - 촬영 시작 시간 안내 및 촬영 시작 1시간 전 도착 당부 - 촬영실 오기 전 통장, 신분증을 지참할 것을 안내 - 오시는 길 안내 2. 음료와 다과 준비 3. 언어제공자가 오면 대기실로 안내하여 관련 서류를 받음 1) 국가 언어 자원(말뭉치) 구축 및 활용 저작권 등 이용 허락 계약서 2) 개인정보 수집, 이용, 제공 및 조회 동의서 3) 자료 수집을 위한 질문지 4) 통장 사본, 신분증 사본
연구원	1. 유도자료 정상 작동 여부 확인 2. 언어제공자의 언어 사용 양상 파악 후 A(대화 나누는 것을 좋아하고 상황에 대한 묘사를 잘하는지), B(다른 사람의 이야기에 집중하는 것을 즐기는지)로 배치 3. 기타 제반 사항 확인

(2) 촬영 중

사회자	<ol style="list-style-type: none"> 1. 촬영자 준비가 완료되었는지 확인 2. 언어제공자 준비가 완료되었는지 확인 3. 촬영자에게 촬영 준비가 완료됨을 알림 4. 유도자료를 제시함 5. 사회자 일지를 작성함
촬영자	<ol style="list-style-type: none"> 1. 3개 카메라의 녹화 버튼을 눌러 촬영 시작
협회 담당자	<ol style="list-style-type: none"> 1. 촬영 중에는 청인이 촬영 공간 안에 들어갈 수 없지만, 촬영 공간에 들어가야 할 돌발 상황에 대처하기 위해 대기
연구원	<ol style="list-style-type: none"> 1. 촬영실 바깥에 있는 모니터로 녹화가 잘 되고 있는지 수시로 확인 2. 촬영일지 작성 3. 자료 수집을 위한 질문지가 잘 작성되었는지 확인하고 누락된 답변 메모 4. 사회자 진행 보완 사항 확인

(3) 촬영 후

사회자	<ol style="list-style-type: none"> 1. 촬영자에게 촬영종료를 알림 2. 작성한 사회자 일지를 연구원에게 전달
촬영자	<ol style="list-style-type: none"> 1. 3개의 카메라의 녹화 버튼을 눌러 촬영종료
협회 담당자	<ol style="list-style-type: none"> 1. 언어제공자 영수증에 서명 받음
연구원	<ol style="list-style-type: none"> 1. 자료 수집을 위한 질문지에서 누락 된 답변을 질문하여 보완 작성 2. 사회자 일지를 복사하여 촬영자에게 전달(편집 작업에 필요) 3. 사회자 진행에서의 보완 사항 공유 4. 언어제공자에게 말뭉치 수집의 목적에 관해 설명

4) 촬영 시간

- 언어제공자의 기질, 낯선 환경 등으로 인해 촬영 시간이 불충분해질 수 있는 위험성을 제거하기 위해 언어제공자 쌍별로 유도 과제별 권장 최소 촬영 시간을 설정하고 가급적 이 시간을 지키도록 안내하였다.

〈표 II -12〉 유도 과제별 권장 촬영 시간

과제명	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
최소시간 (분)	5	10	10	10	25	20	30	25	10	10	10	10	15	20	20	15	10	25

- 촬영 시간은 휴식 시간을 제외한 촬영이 지속된 시간으로 사회자 일지를 기준으로 작성하였다.
- 충북 지역의 총 말뭉치 촬영 시간은 69시간 11분이며, 언어제공자 한 쌍을 촬영하는데 평균 4시간 19분이 소요되었다.
- 유도 과제별 평균 소요 시간은 〈표 II -13〉과 같으며, 하나의 유도 과제를 촬영하는 데 평균 13분이 소요되었다.

〈표 II -13〉 충북 지역 유도 과제별 평균 촬영 시간

과제명	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
평균시간 (분)	3	9	10	11	23	16	23	25	4	7	7	11	22	26	14	13	9	13

〈표 II -14〉 충북 지역 말뭉치 촬영 시간

(단위: 분)

쌍코드 유도 과제	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
동의서	1	1	2	1	1	1	0	1	1	1	0	1	1	1	1	1

쌍둥이 유도 과제	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	1	5	2	3	5	3	3	1	3	2	14	3	4	7	6	3
B	9	11	8	1	10	8	9	12	8	7	7	7	8	6	6	7
C	12	11	12	12	12	13	10	13	12	15	9	8	14	7	10	12
D	18	7	16	6	8	7	7	7	12	9	9	11	12	10	9	10
E	33	24	23	24	37	24	27	29	28	23	15	25	25	29	11	25
F	32	19	23	19	29	23	35	20	26	36	26	22	32	21	46	26
G	42	46	39	30	54	38	57	37	35	47	41	37	40	48	23	42
H	18	36	25	25	32	19	32	29	35	27	15	20	20	25	16	28
I	5	2	3	5	3	3	2	2	3	1	4	3	1	2	2	2
J	5	14	9	6	8	5	7	9	9	12	13	10	11	9	26	10
K	15	7	6	4	5	5	4	5	8	10	5	6	5	9	3	6
L	8	10	11	13	8	10	10	16	8	7	11	14	6	19	20	12
M	11	18	11	27	16	19	17	15	17	19	18	18	12	17	9	16
N	11	20	17	14	15	16	14	11	19	16	10	20	15	19	10	13
O	10	11	15	14	14	10	16	10	27	9	16	9	10	10	11	11
P	5	7	11	7	6	8	12	10	13	9	12	16	8	13	10	10
Q	5	5	8	7	8	7	7	5	8	5	10	8	6	7	8	5
R	14	21	13	7	8	25	15	17	10	13	9	18	17	11	20	17
총 시간	4,151분(총 69시간 11분)															

6. 영상 편집

1) 편집 지침

○ 연구자와 편집자 간의 의사소통을 위하여 다음과 같이 편집 지침을 마련하였다.

A. 파일명

(a) 동의 영상 파일명 : [지역 코드]-[언어제공자 코드]-[촬영 방향 코드]

- 동의 영상이란, A 과제(이름 말하기) 시작 전 언어제공자의 촬영 동의를 재확인하는 영상을 의미함

지역 코드		언어제공자쌍 코드 (촬영순)	촬영 방향 코드 (풀샷 카메라 기준)	
지역명	코드			
충북	KSLCB	1 ~ 16	왼쪽(A)	Lt
			중앙	Md
			오른쪽(B)	Rt

예) 충북 지역 첫 번째 언어 쌍의 오른쪽(B) 언어제공자의 동의 영상 파일명

- “KSLCB1Rt”

(b) 과제별 영상 : [지역 코드]-[언어제공자 코드]-[과제 코드]-[촬영 방향 코드]

- 과제는 총 18가지이며, 과제명 코드는 A~R

지역 코드		언어제공자 쌍 코드 (촬영 순)	과제명		촬영 방향 코드 (풀샷 카메라 기준)	
지역명	코드		과제명	코드		
충북	KSLCB	1 ~ 16	수어 이름	A	왼쪽(A)	Lt
			다시이야기	D		
			표지판	F		
			흥부와놀부	K		
			가장 기억에 남는 날	N	풀샷	Md
			농학교 경험	M		
			주제영역	H		
			농담	B		

			특정사건	G	오른쪽(B)	Rt
			토론	E		
			취미	O		
			농사회 행사	J		
			고장 이야기	P		
			개별 수어	I		
			수어 이야기	C		
			어휘 찾기	Q		
			옛날 수어	L		
			농에피소드	R		

(예시 1) 충북 지역 1번 쌍 A과제 영상



B. 파일 편집 기준

(a) 영상 화질

■ 원본 영상

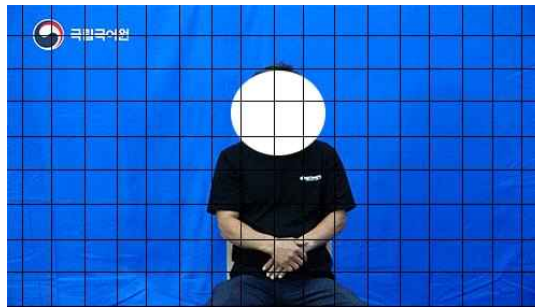
- 해상도: 1920*1080(용량 제한 없음)
- 프레임수: 30
- 화면 화질: FHD
- 확장자명: MOV, AVI, MP4

■ 편집 영상

- 해상도: 1920*1080(500MB 이하)
- 프레임 수: 30~60
- 화면 화질: FHD
- 규격: MPEG4
- 확장자명: MP4

(b) 국립국어원 로고

- 국어원 로고 지침서 참조
- 테두리가 있는 로고 사용, 글자 크기 및 이미지, 글자 간 간격 주의
- 로고 위치: 화면을 64:36 비율로 나누었을 때, 9~15:5~6에 해당하는 지점에 글자가 들어오게 함. 아래와 같이 화면을 나누었을 때, 빨간색으로 표시한 곳에 ‘국립국어원’ 글씨가 들어오게 함



: 단, 화면 비율에 따라 해당 지점에 로고를 삽입하였으나 로고의 위치가 적절하지 않다고 판단되는 경우 위치를 변경할 수 있음. 이것은 해당 연도에 원시 자료 촬영을 위한 카메라 설치에 따라 위치가 변경될 수 있음을 의미함

(c) 과제별 영상 편집 시점

- 수어를 발화하기 시작하는 지점과 발화가 종료된 지점 전후에 여유가 있어야 함.

: 영상의 시작점은 첫 수어 발화 시점 전 5초의 여유가 필요함. 단, 유도자료 설명 동영상으로 제시된 부분은 제외

: 마지막 발화가 끝난 후 5초의 여유를 두고 종료

(d) 과제별 편집 파일에는 소리 삭제

C. 영상 검수 주요 내용

(1) 누락된 영상이 있는가?

지역명	원본 영상	편집 영상	총 개수
충북	48개(16쌍)	864개	912

※ 유도 과제 설명 동영상 42개, 동의 영상 32개 별도, 원본 영상은 정면(사회자와 두 언어제공자)이 보이도록, 왼쪽과 오른쪽 언어제공자를 정면에서 촬영한 영상 총 16쌍 48개의 영상

- (2) 파일명이 제대로 되어 있는가?
 - (3) 파일명과 내용이 다른 것이 있는가?
 - (4) 한 쌍의 3개(Rt, Lt, Md) 영상 시간이 동일한가?
 - (5) 시작점과 끝점이 동일한가? (엘란 확인)
 - (6) 동영상의 시작과 끝에 여유가 있는가? (수어 발화 및 종료 전후 5초의 여유 필요. 단, 여유 시간 5초 내 불필요한 영상이 있을 시 더 짧게 단축할 수 있음)
 - (7) 편집 실수 여부 (앞뒤 영상 간에 여유가 없는 것, bit rate 변환 오류 여부)
- : bit rate의 경우, 2015년 검수 내용에 있었던 것으로, [파일-속성-자세히]에서 비트 속도 전송이 5,000으로 된 경우 엘란에서 사진으로 나타나거나 검은 창으로 뜨는 경우가 발생

(8) 검수표

코드	세부과제	언어제공자 쌍								
		1	2	3	4	5	6	7	...	16
A	수어 이름									
D	넥타이									
F	표지판									
K	흥부와 놀부									
M	농학교 경험 이야기									
H	여러 가지 주제									
	날씨, 직업, 종교, 휴가, 음식									
	경제, 여행, 연애와 결혼, 건강, 패션									

	감정, 계절, 교통, 행사, 부동산									
B	농담 이야기									
G	사건									
	1) 육영수 여사 피살									
	2) 민주화 운동									
	3) 삼풍백화점									
	4) 2002 월드컵									
	5) 노무현 대통령									
	6) 도가니 사건									
	7) 세월호 침몰									
	8) 행복팀 사건									
	9) 남북 정상회담									
	10) 천안함 사건									
	11) 태안 기름 유출 사건									
	12) 오송 지하차도 차량 침수 사건									
E	토론									
	1) 1종 면허 농									
	2) 농학교 교사									
	3) 남북통일									
	4) 듀크족, 딩크족									
	5) 장애인 감형									
J	농사회 행사									
I	개별 수어									
C	애니메이션									
L	옛날 수화와 현재 수화									
N	가장 기억에 남는 날									
O	취미									
P	우리 고장 이야기									
Q	수어 어휘 찾기									
R	농에피소드									
無										

○ 파일명은 지역 코드-언어제공자 쌍 코드-촬영 방향 코드의 조합으

로 부여하였다. 언어제공자 쌍 코드는 1번부터 16번까지 부여하였다.

2) 편집 영상 검수

수집된 원시 말뭉치 원본 영상 및 편집본 영상 검수는 위에 제시된 영상 검수 기준을 토대로 2023년 8월 21일부터 종료 시점까지 여러 차례 집합 검수 및 개별 검수로 진행하였다. 검수 대상은 원본 파일 48개, 편집 영상 파일 864, 총 912개로 나왔다. 세부적인 검수 내용은 다음과 같다.

- 1차 검수에서는 연구진에서 제공한 ‘편집 지침’을 준수하지 않은 것으로 확인되어 전체 재편집을 요청하였다. 예를 들어, 과제별 편집본 영상은 파일명이 [지역 코드]-[언어제공자 코드]-[과제명 코드]-[촬영 방향 코드]로 지정되어야 하나 기준과는 무관한 파일명으로 설정되어 연구진이 직접 방문하여 편집 재교육을 한 후, 수정 편집을 요청하였다.
- 2차 검수에서는 원본 영상 Lt, Rt, Md가 동기화 오류, 과제별 편집본 영상의 파일명 오류, 수어 발화 시작 지점과 종료 지점 전후 5초의 여유가 없음, 편집 영상일지 작성 오류 등으로 수정을 요청하였다.
- 3차 검수에서는 원본 영상의 Lt, Rt, Md 간의 동기화 오류 및 수어 발화 시작 지점과 종료 지점 전후 5초 여유 없음 등으로 연구진이 직접 방문하여 함께 편집을 진행하였다.

3) 수집 영상 분량

(1) 원본 영상

- 원본 영상은 촬영된 영상을 1차 가공한 파일로 촬영 시 휴식 시간, 메모리 교체 등의 이유로 나누어져 촬영된 여러 개의 영상을 하나의 영상 파일로 편집한 것이다.
- 원본 영상은 언어제공자 쌍별 방향에 따른 카메라 3대(Lt, Rt, MD)로 촬영된 영상을 각각 하나의 파일로 편집하여 동기화하였다.
- 충북 지역의 총 말뭉치 원본 영상 시간은 69시간 11분 55초이며, 원본 영상 시간은 다음과 같다.

<표 II -15> 충북 지역 언어제공자 쌍별 말뭉치 원본 영상 시간

1번	2번	3번	4번	5번	6번	7번	8번	9번	10번
04:23:36	04:37:29	04:11:15	04:03:19	4:42:47	04:07:01	04:50:31	04:11:44	04:12:08	04:31:52
11번	12번	13번	14번	15번	16번	총계			
04:16:10	04:16:51	04:11:33	04:33:11	03:46:57	04:15:31	69시간 11분 55초			

(2) 편집본 영상

- 편집본 영상은 1차 가공한 원본 영상에서 유도 과제 설명 시간과 과제와 상관없는 발화 등의 내용은 삭제하여 편집한 것으로, 이것을 유도 과제별로 다시 분할 편집하였다.
- 편집본 영상의 시간은 과제 내에서 발화한 시간만을 포함하기 때문에 전체 편집본 시간과 전체 원본 시간 간에 차이가 있다.
- 충북 지역에서 수집하여 편집 완료된 영상의 시간은 총 59간 31분 29초이며, 각 언어제공자 쌍의 편집본 영상 시간은 다음과 같다.

<표 II -16> 충북 지역 언어제공자 쌍별 말뭉치 편집본 영상 시간

	1번	2번	3번	4번	5번	6번	7번	8번	9번	10번
A	00:01:41	00:03:53	00:01:12	00:01:55	00:04:29	00:01:46	00:02:25	00:01:46	00:03:11	00:01:56
B	00:08:21	00:09:30	00:04:36	00:11:24	00:08:48	00:06:39	00:08:45	00:09:38	00:06:46	00:06:48
C	00:06:40	00:06:01	00:06:35	00:06:06	00:07:49	00:08:29	00:08:40	00:07:30	00:06:49	00:08:53
D	00:13:34	00:03:54	00:05:29	00:02:58	00:03:49	00:03:31	00:03:03	00:02:27	00:09:07	00:05:19
E	00:30:31	00:21:26	00:20:32	00:19:09	00:33:42	00:19:28	00:23:29	00:26:59	00:25:28	00:19:40
F	00:30:53	00:19:34	00:21:43	00:18:14	00:28:44	00:22:06	00:33:52	00:18:55	00:24:17	00:34:32
G	00:37:58	00:42:37	00:34:20	00:24:59	00:49:29	00:33:30	00:53:40	00:33:20	00:31:44	00:42:48
H	00:15:53	00:33:50	00:23:08	00:22:30	00:29:31	00:16:17	00:31:06	00:27:32	00:03:14	00:24:12
I	00:05:16	00:02:10	00:02:57	00:04:31	00:01:41	00:02:12	00:01:30	00:01:13	00:01:31	00:01:06
J	00:05:12	00:13:03	00:07:33	00:04:22	00:07:42	00:03:46	00:08:02	00:08:43	00:08:15	00:10:38
K	00:12:20	00:04:51	00:03:27	00:01:09	00:01:55	00:03:11	00:01:38	00:03:32	00:06:37	00:08:36
L	00:08:20	00:08:43	00:09:57	00:11:56	00:08:03	00:11:16	00:10:00	00:14:57	00:07:24	00:05:58
M	00:10:24	00:16:24	00:11:00	00:26:11	00:14:18	00:12:38	00:16:00	00:14:22	00:15:24	00:17:34
N	00:04:24	00:14:18	00:16:24	00:13:20	00:15:31	00:14:20	00:12:19	00:09:07	00:17:03	00:14:52
O	00:09:18	00:09:00	00:12:12	00:12:14	00:11:42	00:09:02	00:14:23	00:07:49	00:23:54	00:07:12
P	00:05:19	00:06:15	00:09:28	00:06:01	00:05:00	00:05:51	00:11:40	00:08:38	00:11:45	00:08:17
Q	00:05:06	00:03:55	00:07:41	00:06:25	00:07:30	00:05:50	00:06:36	00:05:04	00:06:50	00:04:16
R	00:13:29	00:19:38	00:11:48	00:04:50	00:07:01	00:22:48	00:13:40	00:15:13	00:08:44	00:12:03
계	03:44:39	03:59:02	03:30:02	03:18:14	04:06:44	03:22:40	04:20:48	03:36:45	03:38:03	03:54:40

	11번	12번	13번	14번	15번	16번	총계
A	00:12:32	00:02:34	00:03:10	00:07:15	00:04:42	00:02:10	59시간 31분 29초
B	00:06:17	00:05:26	00:07:10	00:05:32	00:05:01	00:05:59	
C	00:04:08	00:03:14	00:08:46	00:01:39	00:04:24	00:06:54	
D	00:04:47	00:06:48	00:08:21	00:07:20	00:05:12	00:05:56	
E	00:11:29	00:21:53	00:22:12	00:24:21	00:06:12	00:21:58	
F	00:24:51	00:21:54	00:31:37	00:20:24	00:45:20	00:24:58	
G	00:36:56	00:33:05	00:36:40	00:44:30	00:18:27	00:37:21	
H	00:13:12	00:19:28	00:18:28	00:23:56	00:16:46	00:26:43	
I	00:02:53	00:01:39	00:00:54	00:01:37	00:01:39	00:01:12	
J	00:11:35	00:08:35	00:09:59	00:07:33	00:06:02	00:10:08	
K	00:03:19	00:04:15	00:03:31	00:05:25	00:02:54	00:04:33	
L	00:09:51	00:12:50	00:06:09	00:18:14	00:09:41	00:10:52	
M	00:17:08	00:16:53	00:11:02	00:15:26	00:08:03	00:15:07	
N	00:08:54	00:19:11	00:13:33	00:19:17	00:08:02	00:12:00	
O	00:14:10	00:08:25	00:12:37	00:09:49	00:19:35	00:08:28	
P	00:10:41	00:14:18	00:07:44	00:11:03	00:08:08	00:09:33	
Q	00:08:43	00:06:38	00:04:50	00:04:42	00:06:35	00:03:58	
R	00:17:13	00:16:36	00:15:04	00:10:14	00:16:39	00:16:14	
계	03:38:39	03:43:42	03:41:48	03:58:17	03:13:22	03:44:04	

7. 후속 자료 수집 시 고려 사항

(1) 설명 동영상의 한국어 제목 제시 방법 재고

- 유도 과제 제시 시, 설명 동영상의 시작 부분에 한글로 제목이 제시되는 것이 농인 언어제공자가 유도 과제를 잘못 이해하거나 이해를 방해하는 요소가 되기도 한다.
- 예를 들어 ‘P - 우리 고장 이야기’의 경우 ‘고장’이라는 단어를 ‘고장이 나다’의 의미로 이해하는 경우가 발생하기도 하였으며, ‘B-준비해 온 농담 이야기’의 경우 ‘농담’이라는 단어를 이해하지 못하거나, ‘농담을 해 본 경험이 없다’라고 이야기하였다.
- 추후 설명 동영상 제작 시, 이해하기 쉽게 풀어서 설명하는 것이 필요하다. ‘P - 우리 고장 이야기’의 경우 ‘당신이 살고 있는 지역의 유명한 곳이나 맛집 등을 이야기해 보세요.’, ‘B-준비해 온 농담 이야기’의 경우 ‘당신이 속았던 경험이나 다른 사람을 속였던 경험이 있다면 이야기해 보세요.’라고 유도하는 것이 언어제공자가 과제를 이해하고 수행하는 데 도움이 될 것이다.
- 한글 제목은 이해하기 쉬운 한글 표현으로 대체하거나 유도자료 제시 시, 한글 제목이 크게 드러나지 않도록 화면의 하단에 연구진만이 구별할 수 있도록 표시하는 것을 고려할 필요가 있다.

(2) 언어제공자 모집 및 선정 기준 재고

- 언어제공자 적합성 여부를 판단하기 위해 ‘언어제공자 선정 조건’을 두고 있다. 이 중 ‘번호 3. 농학교 졸업’은 농학교 졸업 여부만을 한정하고 있으며, 부모가 모두 농인이면 예외를 두어 일

반 학교 졸업 농인도 포함하고 있다.

- 우리나라는 비교적 지역 간의 이동이 용이한 특성을 가지고 있어 수집 지역의 농학교를 졸업한 농인보다 타 지역의 농학교를 졸업한 언어제공자가 다수 참여하는 경우가 많다. 이에 유도 과제 중 ‘M-농학교 경험 이야기’ 나 ‘I-다양한 수어 표현’, ‘P-우리 고장 이야기’ 에서 수집 지역의 수어보다는 타 지역의 수어가 수집될 가능성이 있다.
- ‘언어제공자 선정 조건’ 에서 ‘농학교 졸업’ 을 ‘수집 해당 지역 또는 권역의 농학교 졸업’ 을 우선 선정 조건으로 하고, 다른 지역의 농학교를 졸업한 경우는 후 순위 조건으로 수정할 필요가 있다.
- 말뚝치 수집 시, 수집 지역의 농아인협회에 ‘언어제공자 선정 조건’ 에 부합하는 언어제공자를 추천받아 선정하고 있다. 말뚝치 수집은 한정된 사업 기간 내에 진행되어야 해서 한 지역의 농아인협회에 의존할 수밖에 없는 상황으로 다양한 언어제공자를 모집하는데 한계가 있다.
- 충분한 시간을 가지고 다양한 언어제공자를 발굴하는 것이 필요하며, 농아인협회 산하 지부에 홍보 및 공문을 통해 다양한 언어제공자가 참여할 수 있도록 유도하는 것이 필요하다.

(3) 사회자 및 촬영·편집자 교육방법 재고

- 2021년까지는 사회자 교육과 촬영·편집 담당자 교육이 따로 진행되었다. 각각의 교육은 해당 지침에 따라 연구진이 개별적으로 진행하여 서로 간의 역할을 충분히 이해하지 못해 어려움을 겪기도 하였다.
- 이번 사업에서는 사회자와 촬영·편집 담당자가 함께 교육에 참여하였으며, 사회자 교육은 사회자 경험이 있는 연구진이, 촬영·편집

담당자 교육은 작년도 수집에 참여한 촬영·편집 담당자가 진행하였다. 실제적 경험을 바탕으로 구체적인 노하우를 공유하여 말뚝치 수집을 준비하는 데 도움을 주었다.

- 후속 말뚝치 수집에서는 사회자와 촬영·편집자뿐 아니라 행정지원 등 말뚝치 사업에 참여하는 관계자가 모두 참여하여 말뚝치를 이해하고 사업을 진행할 수 있도록 준비하는 것이 필요할 것이다.

(4) 언어 수집을 위한 촬영 방법 재고

- 2019년부터 한국수어 원시 말뚝치는 수집 해당 지역의 농아인협회와 계약을 체결하고 진행한다. 지역별 농아인협회는 해당 지역에 거주하는 농인들에 대한 정보를 가장 많이 갖고 있어 수집에 있어 가장 중요한 언어제공자 모집 및 선정, 그리고 사회자 섭외 등에서 큰 도움을 받을 수 있다. 또한 협회의 일부 공간을 촬영 장소로 활용할 수 있으므로 언어제공자들이 익숙한 공간에서 편안하게 촬영에 참여할 수 있으며, 이동에도 큰 불편을 겪지 않을 수 있다.
- 이번 사업에서는 충북 지역의 수어를 수집함에 있어 충북농아인협회와 계약을 하고 진행하였다. 촬영장은 충북종합사회복지센터 공간에 설치하였으며, 수집에 필요한 전반적인 사항을 충북농아인협회의 도움을 받아 진행하였다.

Ⅲ. 공개를 위한 한국수어 말뭉치 정비

1. 한국수어 말뭉치 공개 대상 말뭉치 선정

수어 말뭉치 연구는 2015년부터 시작하여 해마다 원시 수어말뭉치 수집과 더불어 주석 말뭉치 구축을 진행해왔다. 2022년까지 축적된 원시 말뭉치는 약 395시간 분량이며, 수지 글로스로 주석이 된 기본 주석 말뭉치는 1,438분(23시간), 수지 주석에 대한 의미 주석까지 수행된 자료는 445분(7시간)이다. 영상의 내용을 번역한 자료는 5,782분(96시간)이고, 의미를 포함한 기본 주석과 번역 주석이 모두 진행된 자료는 157분으로 파악되었다.

본 연구에서는 기 구축된 한국수어 말뭉치 중에서 공개 가능 여부를 판단하기 위한 검토 기준을 결정하였다. 1차로 기본 주석과 번역이 있는 영상 파일을 연구진들이 공개 검토 기준에 따라 공개 여부를 결정한 후에 지역, 성별, 나이대를 고려하여 공개를 위한 한국수어 말뭉치를 최종 선정하였다. 최종 선정된 말뭉치 중에서 기본 주석과 번역이 있는 파일은 정비 작업을 실시하였다. 이번 한국수어 말뭉치 공개 영상은 273분 30초 분량의 원시 수어 말뭉치, 63분 54초 분량의 번역된 말뭉치, 그리고 95분 6초 분량의 주석과 번역이 모두 된 수어 말뭉치이다.

2. 한국수어 말뭉치 공개를 위한 검토

1) 공개 배포를 위한 검토 기준

공개할 영상의 선정을 위해 다음과 같은 기준으로 검토하였다.

1. 성(性) 관련 부정적인 내용 및 성적 지향과 관련된 민감한 주제 포함 여부

2. 얼굴이름 노출 여부 및 특정인에 대한 부정적인 내용 포함 여부
3. 개인정보 노출 (병, 금전 등) 여부
4. 농협회 관련 부정적인 내용 여부
5. 종교 및 농교회 관련 부정적인 내용 여부
6. 정치적으로 편향된 내용 여부
7. 법적 혹은 도덕적으로 문제가 되는 내용 여부

2) 검토 방법 및 결과

말뭉치 자료가 공개 자료로 적합한지 검토하기 위해서 각 엘란 파일에 공개 검토A(A언어제공자)와 공개 검토B(B언어제공자) 층렬을 만들어 공개 결과를 기록하였다. 해당 파일의 공개 적합 여부에 따라 층렬의 맨 앞부분에 “공개” 또는 “비공개”로 기록했으며 비공개인 경우, 그 이유도 함께 기록하였다.

	00.000	00:00:01.000	00:00:02.000	00:00:03.000
공개검토A [4]	비공개: 농협회관련 부정적인 내용은			
공개검토B [5]	비공개: 농협회관련 부정적인 내용은			

또한 공개 기준 검토 결과에 따라 문제가 될 만한 구체적인 내용을 엘란 파일 맨 앞에만 기록하지 않고 해당 영상 부분을 분절하여 공개 검토 층렬에 주석을 남겼다. 이는 해당 부분을 빠르게 다시 찾을 수 있도록 하기 위한 것이다. 본 연구에서는 공개를 위해 1차 선정된 엘란 파일 46개, 총 11시간이 넘는 자료를 검토하였으며 검토 결과는 아래와 같다.

파일 이름	공개 검토 결과	시간
강원지역 10번쌍 P-우리고장이야기	비공개: 불법 도박 이야기 포함	0:14:09
강원지역 10번쌍 R-농에피소드	비공개: 다수의 얼굴이름 노출 및 부정적인 내용	0:33:41
강원지역 1번쌍 H-주제	공개	0:14:22

강원지역 7번쌍 O-취미	공개	0:17:32
경남지역 11번쌍 H-주제	공개	0:09:03
경남지역 12번쌍 H-주제	공개	0:13:30
경남지역 2번쌍 J-농사회 행사	공개	0:02:41
경남지역 4번쌍 H-주제	공개	0:13:30
경남지역 7번쌍 H-주제	공개	0:22:16
경남지역 8번쌍 L-옛날 수어와 현재 수어	공개	0:04:06
경남지역 9번쌍 H-여러가지 주제	비공개: 성적 소수자에 대한 부정적인 내용	0:30:12
서울지역 10번쌍 G-특정사건	공개	0:18:15
서울지역 14번쌍 H-주제	공개	0:09:03
서울지역 14번쌍 M-농학교 공험이야기	비공개: 성희롱	0:11:13
서울지역 19번쌍 J-농사회행사	비공개: 행사에 대한 부정적인 내용	0:11:37
서울지역 1번쌍 E-토론	공개	0:17:03
서울지역 20번쌍 E-토론	비공개: 개인정보 포함	0:11:21
서울지역 20번쌍 J-농행사	공개	0:12:13
서울지역 24번쌍 J-농사회 행사	비공개: 금전적으로 부정적인 내용 포함	0:09:42
서울지역 31번쌍 J-농행사	공개	0:07:59
전남지역 10번쌍 A-수어이름	공개	0:04:47
전남지역 10번쌍 B-준비해 온 농담 이야기하기	공개	0:06:08
전남지역 10번쌍 E-토론	비공개: 부정적인 내용 포함	0:12:36
전남지역 14번쌍 M-농학교 경험 이야기	비공개: 학교폭력, 비행행동, 자살에 대한 이야기	0:46:41
전남지역 15번쌍 D-다시 이야기하기	공개	0:08:46
전남지역 29번쌍 P-우리고장이야기	공개	0:36:53
전남지역 29번쌍 R-농에피소드	비공개: 특정 개인에 대한 부정적인 내용 포함	0:19:15
전남지역 2번쌍 H-주제영역	비공개: 개인정보 포함	0:20:22
전남지역 7번쌍 M-농학교 경험 이야기 (일부)	공개 00:30:52까지만	0:41:12
전북지역 15번쌍 R-농에피소드	공개	0:20:40
전북지역 15번쌍로 C-영상을 보고	비공개: A가 수어를 할 때 마다	0:05:57

수어 이야기하기	모자가 나와서 공개하기 어렵다고 판단	
충남지역 11번쌍 G-사건	공개	0:10:20
충남지역 11번쌍 N-가장 기억에 남는 날	공개	0:24:23
충남지역 15번쌍 K-흥부와 놀부	공개	0:07:31
충남지역 1번쌍 N-가장 기억에 남는 날	비공개: 험담이야기는 아니지만 특정한 수어통역사 얼굴이름이 나와서 공개하기 애매하다고 판단	0:19:33
충남지역 1번쌍 R-농에피소드	공개	0:05:59
충남지역 8번쌍 F-표지판	비공개: 여성 신체부위에 대한 정밀한 묘사	0:25:10
충남지역 8번쌍 N-가장 기억에 남는 날	비공개: 종교에 대한 비방이야기	0:54:10
충남지역 8번쌍 P-우리 공장 이야기	비공개: 사회자가 너무 많은 대화 참여로 사회자 영상 빼고 공개 시 내용 연결이 안됨.	0:17:09
전북지역 15번쌍 Q-수어 어휘말하기	공개	0:11:42
강원지역 7번쌍 B-준비해온 농담 이야기하기	공개	0:07:47
강원지역 10번쌍 F-표지판	공개	0:37:47
충남지역 8번쌍 L-옛날 수어와 현재 수어	공개	0:18:44

파일을 검토한 결과, 특정 인물에 대한 부정적인 내용이 있고, 얼굴 이름이 노출되어서 공개하기에는 부적합 경우가 여러 번 나타났다. 성희롱, 신체 부위에 대한 정밀한 묘사와 성적 소수자에 대한 부정적인 내용은 성 관련 민감한 주제로 공개에 적합하지 않은 것으로 결정하였다. 학교폭력, 비행 행동, 자살 이야기, 도박 이야기 또는 도덕적으로 법적으로 문제 될 만한 언어제공자의 경험 또는 언어제공자 지인의 이야기는 공개 시, 언어제공자의 이미지를 나쁘게 보이게 하든가 불이익을 줄 수 있기 때문에 비공개로 결정하였다. 종교에 대한 부정적인 이야기는 한 번밖에 안 나타났지만 농사회 행사를 부정적으로 언급하는 경우는 여러 차례 있었고 그 내용이 공개되면 언어제공자한테 부정적인 영향을 미칠 수 있어서 공개하지 않기로 결정하였다. 그

외 사회자가 너무 많은 이야기를 하는 경우도 비공개로 결정하였다. 사회자를 보여주는 폴샷의 중앙 영상은 공개하지 않을 예정인데 사회자 영상 없이는 이야기 연결이 안 될 정도라 판단되어 해당 파일도 비공개로 결정하였다.

모든 공개 및 비공개 구분은 논의를 거쳐 결정하였고, 공개가 가능한 자료는 현재 432분 30초이다.

3. 한국수어 말뭉치 공개를 위한 정비

1) 주석 정비

○ 주석 정비 방법

주석에 대한 수정 또는 보완 사항이 있는 경우 원래 주석을 실시한 층렬에서 직접 수정 또는 보완하였다. 수정 보완한 내용에 대한 것은 정비 층렬에 통제 어휘를 사용하여 기록하였다. 이것은 정비 흔적을 남기고 수정된 내용을 파악하여 주석 시에 어떤 오류가 많이 나타나는지를 추후에라도 확인할 수 있게 하기 위함이었다. 정비 층렬은 정비A(언어제공자A)와 정비B(언어제공자B), 그리고 정비코멘트A와 정비코멘트B로 이루어졌다.

정비한 내용은 정비 층렬에 12가지의 통제 어휘를 사용하여 기록하였다. 통제 어휘는 다음과 같다.

- 분절수정: 잘못된 분절을 수정한다.
- 손바뀜: 한 언어제공자 내에서 우세손과 비우세손이 바뀌어 입력되어 있는 경우 수정한다.
- A/B바뀜: 언어제공자A의 수어가 언어제공자B의 수어로 입력되어 있거나 그 반대의 경우 주석을 수정한다.
- 타입명 적합성: 타입명의 대표성이 부적절할 경우 새로운

타입명을 제안한다.

- 토큰삽입: 토큰 누락 시 토큰을 삽입한다.
- 타입수정 고정수어: 고정된 수어가 잘못된 타입으로 주석된 경우 정비를 한다.
- 타입수정 생산적 수어: 생산적 수어가 잘못된 타입으로 주석된 경우 정비를 한다.
- 의미수정: 의미를 수정한다.
- 지문자 수정: 지문자를 수정한다.
- 합성어: 합성어 주석이 잘못된 경우 수정한다.
- ###: 모르는 수어를 해결한다.
- 기타: 그 외 오류를 수정한다.

정비코멘트 층렬에는 다음과 같은 경우에 기록하였다.

- 수정 사항이 2개 이상인 경우(통제 어휘가 중복 선택이 되지 않으므로)
- 우세손과 비우세손이 다른 경우 - 수정한 손에 대한 내용으로서 예를 들어, 우세손만 토큰을 삽입한 경우는 ‘우세손 토큰 삽입’으로 기록한다.
- 의미 층렬이나 코멘트 층렬에 있던 내용을 없애거나 삽입한 경우 - 정비코멘트 층렬에 ‘의미삽입 또는 의미삭제’, ‘코멘트 삽입 또는 코멘트 삭제’ 라고 기록한다.
- 통제 어휘에 토큰 삭제가 없으므로 토큰을 삭제한 경우 - 정비 층렬에서 통제어휘 기타를 선택하고 정비코멘트 층렬에 ‘토큰 삭제’로 기록한다.

○ 주석 정비 파일 선정 및 주석 정비 교육

2022년까지 기 구축된 주석 말뭉치 중 아래와 같은 파일들을 선정하여 주석 정비 작업자한테 분배하여 정비를 진행하였다.

	서울지역 10번쌍 G	서울지역 14번쌍 H	서울지역 14번쌍 J	서울지역 31번쌍 J	서울지역 1번쌍 E	강원지역 1번쌍 H	경남지역 2번쌍 J	서울지역 20번쌍 J
영상 시간	18:15	9:03	4:03	7:59	17:03	14:22	02:41	12:13
주석자	주석자 A		주석자 B			주석자 C		

본 연구에 투입된 3명의 주석 정비 작업자는 오랜 기간의 주석 경험을 지니고 있어 주석 정비를 수월하게 실시할 수 있었다. 주석 정비 교육은 2023년 6월 26일 비대면 형식으로 진행하였고, 주석 정비 회의는 2023년 8월 3일 비대면 형식으로 진행하였다. 그 외 문의는 구글 드라이브를 활용하여 실시간으로 연구진과 피드백을 주고받았다.

No.	날짜	구분	질문자	파일 이름	영상 타임코드	질문 내용	답변자	답문 내용	연구진 메모
3	7:30	타입명	임하리	서울1번E토론	16:40~16:44	[지시2]가 2개 나오는데, 2개 다 [지시2]로 보이지 않습니다. 두번씩에 나오는 [지시2]는 2개를 가르는 게 어렵다고 하는 것처럼 보이는데 이 수어는 타입명으로 등록돼 있지 않습니다.	홍상은	지시2 두개 다 구형으로 보는 것이 좋을 것 같습니다.	
4	7:30	타입명	임하리	서울1번E토론	1:16	[지시2]가 2개 다 확인 부탁드립니다.			
5	7:31	타입명	유진민	서울20번J행사	8분44	[2중]이 나오는데요, [일중]은 타입이 등록돼 있습니다. 그런데 2중은 한글자문자로 돼 있습니다. 1중이 고정수어라면 1중도 고정수어로 봐야할 것 같은데...	홍상은 >이은영	[일중]이 고정수어라면 2중도 고정된 수어로 타입을 만들어야한다는 뜻이죠? 살펴보니 2중을 표현할 때는 [일중]1수어와 다르게 보입니다. 우세선과 비우세선의 역할도 다르고 '중'을 실제로 다 산출하지 않는 것으로 보입니다. 그래서 한국자문자로 처리를 한 것 같습니다. 홍이 [약속1]어글1을 [늦었다2] 변이형으로 보기는 어렵습니다. 왜냐하면 수어2개 중 하나만 등장하니까요. 그래서 사실 늦었다3으로 등록하는 것이 제일 쉬운 방법인데 합의여인지 아닌지 확신이 쓰지 않아서 [약속1](물1)로 놔두기로 했습니다. 합성어가 아닌 2개의 수어로 정비해주세요.	
6	7:31	타입명	유진민	경남2번J행사	7초1	[한적있다1](상일예정)1(경형1) 저 두 수어를 동시에 써서 '경형'이라는 의미로도 쓰는데요, [경형1]이나 [한적있다1]의 변이형?인지 아니면 새로 등록을 해야하는지? 새로 등록하는 거라면 합성어인지? 궁금합니다.	이은영 >홍상은	[한적있다1]1(경형1)을 사용해서 '경형'한 적 있다를 표현합니다. 다른 의미로는 '제함'으로도 사용합니다. [제함]은 등록되어 있지 않은 것으로 보여서 새로 등록이 필요할 것 같습니다. [한적있다1]에서 음지를 회고하는 수형은 변이형으로 보입니다. 홍: 제함1로 정비를 하면 따로 합성어 입성이 기록은 필요없음 저는 [사백오십1](생존1)(오십1)이라고 하는데... 어렵네요. 이은영: 비우세선으로 [사백오십1](생존1)(오십1)이라고 하는데... 어렵네요. 이은영: 선생님께서 도움을 요청이 필요해보입니다.	제함1 등록 필요

○ 주석 정비 결과

주석 정비를 실시한 파일을 연구자가 검토하는 과정에서 일괄적으로 최종 수정 보완을 한 내용은 다음과 같다.

- 정비코멘트 층렬에 있는 ###(모르는 수어)를 해결하였다.

- 원 주석 층렬 중 ‘의미’ 층렬이나 ‘코멘트’ 층렬에 의미를 삽입하거나 삭제한 경우 정비코멘트 층렬에 “토큰 삭제” 또는 “토큰 삽입” 이라고 되어 있는 것을, 이는 토큰의 삭제나 삽입이 아니므로 정비 층렬에 “기타” 로 하고 정비코멘트 층렬에 ‘의미삽입’ 또는 ‘의미 수정’ 으로 기록하였다.
- 글로스 층렬에 %(수어를 하다가 중지)로 기록되어 있는 경우, 의미를 알 수 있는 경우는 의미 층렬에 의미를 기록하지 않고 = 의미코멘트 층렬에 그 의미를 기록하는 것으로 수정하였다.
- 정비와 정비코멘트의 분절이 윗 층렬(글로스 층렬, 의미 층렬 등)의 분절과 맞지 않는 경우에는 정비 층렬의 분절을 수정하였다.
- 같은 수어가 반복(예를 들어, 3번 반복)되는 경우는 각 토큰별로 분절 후 정비 층렬에 ‘토큰 삽입’ 으로 기록하고, 정비코멘트 층렬에는 반복되는 수어를 ‘분절’ 이라고 기록하였다.

연구자 검토 후 최종 주석 정비 작업자들이 실시한 정비 건을 종류별로 정리하면 아래와 같다.

	정비A	정비B	합계
A/B 바뀜	0	0	0
손바뀜	16	10	26
토큰삽입	101	99	200
타입수정 고정수어	106	116	222
타입수정 생산적 수어	15	12	27
타입명 적합성	0	1	1
의미수정	722	537	1259
지문자 수정	81	15	96
합성어	2	0	2
###	11	8	19
기타	98	107	205
			2,057

표에 제시된 바와 같이 의미 수정에 대한 정비가 제일 많이 이루어졌다. 주석자가 토큰에 대한 주석을 달고 토큰의 의미를 따로 의미 층렬에 기록해야 하는 데 이 때 의미가 누락이 되기도 하고 잘못된 의미를 기록한 경우도 있었다. 의미 층렬은 2019년에 추가된 층렬이라 주석자들이 익숙하지 않아 의미 주석 누락이 있었던 것으로 보인다.

다음으로 많은 정비가 진행된 것은 타입수정 고정수어와 토큰 삽입이다. 타입수정 고정수어는 토큰 타입 매칭이 잘못된 오류다. 언어제공자가 하는 수어를 주석자가 잘못 봤든가 타입명을 잘못 기록한 오류이다. 또한 생산적 수어를 고정된 수어로 잘 못 주석된 경우도 있었다. 누락된 토큰에 대한 토큰 삽입 정비도 적지 않게 이루어졌다. 수어의 발화 속도가 빠르기 때문에 자칫하면 수어를 놓치는 경우가 발생할 수 있다. 지문자 수정은 한국수어자료 통합지원시스템이 도입되기 전의 주석 규정과 그 후의 규정이 달라서 많은 사례가 나타난 것으로 보인다. 2015년 주석 작업을 시작할 때는 지문자를 (예:ㅎㅏㄱ고) 우세Gloss 아니면 비우세Gloss 층렬에 주석했지만 (한국수어 말뭉치 기반 조사 및 시범 구축 최종보고서, 2015) 한국수어자료 통합지원시스템이 지문자 인식이 어려워 모든 지문자를 ‘한글지문자’로 주석하고 개별 지문자는 코멘트 층렬에 기록하였다(국립국어원, 2020).

A/B 바뀔, 타입명 적합성과 합성어 정비가 필요한 사례는 거의 나타나지 않았다. 분절에 대한 정비는 계획과 달리 실시하지 못하였다. 컴퓨터 상향 따라 분절된 영상이 일관되게 재생이 되지 않았기 때문이다.

‘기타’의 경우는 ‘토큰 삭제’가 많았다. 이는 토큰이 아닌 단순한 손짓을 수어로 오인하여 주석한 것이다. 따라서 향후 정비에서는 “토큰 삭제” 항목을 추가할 필요가 있다. 이외에도 수정한 내용이 2개 이상인 경우, 통제 어휘 중복 선택이 되지 않아 정비코멘트 층렬에 기입하였는데, 향후 정비에서는 이에 대한 보완이 필요하다.

○ 주석 지침 보완

원래 주석 규약에는 외국수어를 [미국수어], [국제수어]로 주석하고 의미를 코멘트 층렬에 기록하는 것으로 되어 있지만 코멘트 층렬은 공개가 안 될 예정이라 외국수어의 의미를 코멘트 층렬에서 의미 층렬로 옮겨야 할 필요가 있다. 그러므로 주석 지침은 아래와 같이 수정을 하였다.

[수정 전]

③ 외국수어/국제수화

- 수어 화자가 미국수어를 하는 경우 [미국수화]로 입력하고, 국제수어를 할 경우 [국제수화]로 입력한다. 그리고 그 수어의 의미는 코멘트 층렬에 입력한다.

[수정 후]

③ 외국수어/국제수어

- 수어 화자가 미국수어를 하는 경우 [미국수어]로 입력하고, 국제수어를 할 경우 [국제수어]로 입력한다. 그리고 그 수어의 의미는 의미 층렬에 입력한다.

2) 번역 정비 절차

○ 번역 정비 자문회의

한국수어말뭉치 연구에서 번역을 실시한 적은 있으나 체계적인 정비 작업은 첫 시도이기에 번역에 대한 기준과 지침을 제공하는 것이 매우 중요하다.

한국어 문법과 수어 문법에 차이가 있기 때문에 한국어 사용자 기준으로 가독성 있는 번역을 제공하기 위해서는 두 언어의 차이를 이해하고 구체적인 지침이 필요하다. 예를 들어, 수어에서 지시 수어는 매우

다양한 역할을 가지고 있기 때문에 같은 방식으로 번역해서는 안 된다. 그리고 수어 문장에 대한 정의와 수어 문장의 종결은 수어언어학에서도 충분히 연구되지 않은 분야라 본 연구에서도 번역 분절에 대한 명확한 기준이 현재 없는 상태이다. 이 같은 문제를 논의하고 번역 정비에 앞서 전문가들의 자문을 구하기 위해 번역 정비 자문회의를 2회 개최하였다.

1차 자문회의

- ◆ 일시: 2023.6.8. 10:00~11:30
- ◆ 장소: 강남대학교 창조산학관 120호
- ◆ 참가자: 자문위원 3명, 국립국어원 사업 담당자, 연구원 4명
- ◆ 주요 안건: 기존 번역 지침의 수정 및 보완 방향

1차 자문회의에서는 수어 번역 시에 고려해야 할 점 및 번역 지침에 대해 논의한 결과 공개 자료를 볼 다양한 사람들에게는 의역과 직역을 모두 제공하는 것이 수어를 이해하는 데 더 많은 도움이 될 것이라는 의견이 제시되었다. 직역은 수어의 언어적 특성을 잘 살리고 의역은 도착어로 자연스러운 번역을 제공해주어 효과적인 내용 전달이 가능하다는 특성이 있다. 현재 대한민국에서 수어 번역을 직역과 의역으로 나누어 번역한 사례는 없었기에 본 연구에서 5분의 수어말뭉치 자료를 직역과 의역으로 구분하여 시범 번역하고 그 시범 번역을 기반으로 지침을 만들기로 결정하였다.

연구진은 수어말뭉치 자료 중 강남지역 12번쌍 H 파일 00:00:000 ~ 05:00:000 자료를 시범 번역 자료로 선정하고 그 자료에 대한 직역과 의역을 작성하였다. 또한 시범 번역을 참고하여 번역 정비 지침을 작성하였다. 시범 번역을 한 것에 대한 피드백은 2차 자문회의를 통해서

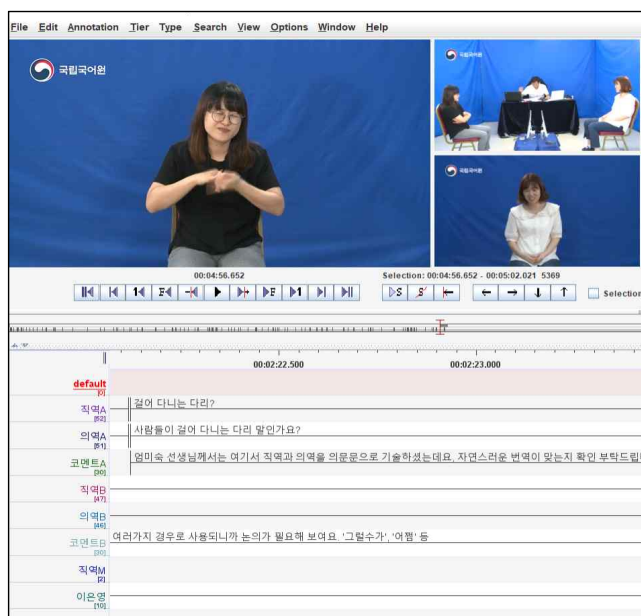
받았고 번역 정비 지침에 대한 자문은 서면으로 받았다.

2차 자문회의

- ◆ 일시: 2023.8.10. 10:00~13:00
- ◆ 장소: 가산동 알토비전 회의실
- ◆ 참가자: 자문위원 2명, 연구원 2명
- ◆ 주요 안건: 시범 번역 정비

○ 시범 번역

번역은 통역과 달리 여러 차례의 수정이 필요하다. 연구진이 작성한 시범 번역(강남지역 12번쌍 H 파일 00:00:000 ~ 05:00:000)은 6차 수정까지 이루어졌다. 구축된 초벌 번역을 외부 번역자를 통해 수정하고 여러 차례 연구진이 검토와 수정을 하였다. 하지만, 5분이라는 자료를 6차례 수정작업 후에도 완성도를 보장하기 어려웠다.



○ 번역 정비 방법

본 연구에서는 기존 번역을 정비만 하는 것이 아니라 번역문을 직역과 의역으로 구분하여 기록하도록 그 기준을 마련하였다. 기존 번역의 정비는 번역 오역을 수정하고, 적절하지 않은 어휘를 대체하고, 잘못된 분절을 새로 잡았다. 직역은 수어 하나하나의 의미를 충실하고 정확하게 번역하는 것이 핵심이고, 수어의 언어적 구조 특성이 잘 드러나게 하는 것에 주안점을 두었다. 의역은 원문의 단어나 구절에 지나치게 얽매이지 않고 맥락을 바탕으로 전체적인 의미가 한국어로 가능한 자연스럽고 정확하게 전달되도록 하는 데 주안점을 두었다.

번역 정비에 해당되는 층렬은 아래와 같다.

- ◆ 번역A/B: 직역에 해당되는 층렬이며 기존 번역을 직접 수정하는 층렬이다.
- ◆ 의역A/B: 의역을 새로 기록하는 층렬이다.
- ◆ 번역정비_코멘트_A/B: 번역 정비자가 번역문에 대한 설명 및 표시를 남겨야 할 필요가 있을 때 사용하는 층렬이다.

○ 번역 정비 지침

자문회의와 여러 차례 연구진의 협의 수정한 시범 번역을 기반으로 작성된 번역 정비 지침은 아래와 같다.

번역 정비는 「주석 규약」의 번역 지침에 따라 최종적으로는 한국어 독자에게 수어 발화의 정확한 형태 및 통사 구조를 제시하되 한국어 문장으로 어색하지 않도록 번역함을 원칙으로 한다. 「한국수어 말뭉치 주석 규약」(강창욱 외, 2020)에서는 다음과 같이 번역에 관한 지침을 설명하고 있다.

1. 번역의 종류

「주석 규약」의 번역 지침에 따르면, ‘한국수어-한국어 번역’은 향후 한국수어 말뭉치 공개 시, 수어를 모르는 사람들도 자료에 접근할 수 있도록 제공하는 것을 그 목적으로 밝히고 있다.

번역은 다음과 같이 의역(Free Translation)과 직역(Narrow/Literal Translation)으로 구분될 수 있다.

① 의역(Free Translation)

- 의역은 내용 전달성을 핵심으로 하여 내용이 잘 이해될 수 있도록 목표어(한국어)의 문법에 맞게 완결된 형식으로 제공한다.
- 의역은 내용 전달을 목적으로 하되, 출발어인 수어의 내용을 과도하게 주관적으로 해석하여 본래의 의미를 왜곡하지 않도록 한다.
- 출발어인 수어와 목표어인 한국어의 문장 길이가 일치하지 않을 수 있다.
- 수어의 지시 표현([지시])은 이해하기 쉽도록 한국어에서 가장 자연스러운 표현으로 번역한다.
- 일반적인 수어 담화의 맥락 안에서 그 내용을 이해하는 데 도움이 되는 번역을 제공한다.
- 문장 끊김, 감탄사(오!, 헉 등), 어순 뒤바뀔 등의 구어체에서 나타나는 전형적인 특징은 생략될 수 있다.

② 직역(Narrow/Literal Translation)

- 출발어인 수어를 절과 문장 수준으로 잘라 번역하므로 비교적 짧은 절 또는 문장으로 번역된다.

- 직역은 수어 담화 구조를 이해하고 기계 번역 알고리즘을 학습하고 테스트하는 데 활용할 수 있도록 수어의 언어적 구조 특성이 드러나게 함을 원칙으로 한다.
- 직역은 목표어인 한국어로서는 부자연스럽거나 비문법적 문장의 형태를 띌 수 있다.
- 수어의 지시 표현([지시])은 대명사를 사용하여 직접적으로 번역한다.

2. 구체적 번역 지침

본 연구에서는 향후 일반인을 대상으로 한국수어 말뭉치를 공개하고 주석 작업자의 지역 수어 접근성을 높여 주석 작업에 도움이 될 수 있는 번역 정비를 하고자 한다. 다년간에 걸친 지역 수어 말뭉치에 대한 주석을 진행하면서 숙련된 주석자들이 겪었던 가장 큰 난제 중 하나가 그 지역에서 나타나는 생소한 수어 형태 처리에 관한 부분이었다. 이에 의역을 하되 향후 본 번역 파일에 대한 주석을 진행할 때, 주석 작업자가 잘 모르는 수어의 의미를 유추하는 데 도움이 될 수 있도록 번역하도록 기준을 마련하였다. 이러한 번역의 지향점을 고려하여 다음과 같은 구체적인 번역 지침을 설정한다.

① 출발어(수어)의 분절

- 수어를 분절할 때, 비수지가 먼저 나타나더라도 시작점은 우세손과 비우세손을 기준으로 한다. 끝점도 비수지가 아닌 우세손과 비우세손의 움직임의 해제나 손 내림, 손 풀림, 손 유지 등의 휴지가 나타날 때를 기준으로 한다.
- 수어의 의미 전환이 명료하게 나타나는 지점에서 분절한다.

② 직역티어: 발화된 수어의 형태 및 통사 구조가 드러나도록 번역한 문장

- 수어와 일대일로 대응되는 의미에서 크게 벗어나지 않도록 하여 수어 발화 내용의 정확성을 확보한다.
- 최대한 수어의 절과 문장 수준으로 분절하여 수어의 구조적 특성이 드러나도록 한다.

③ 의역티어: 발화된 수어의 의미를 자연스러운 목표어(한국어)로 적절하게 번역한 문장

- 수어 발화 내용을 그대로 옮겨 놓은 듯한 직역 중심의 번역을 하지 않는다.
- 자연스러운 한국어 번역을 추구하되 과도한 의역으로 인해 수어 발화 내용이 왜곡되지 않도록 한다.
- 한국어로 번역된 단일 문장이 부자연스러울 정도의 여러 개의 절로 구성되지 않도록 한다.

④ 위의 직역과 의역이 구분될 수 있도록 한다.

- 직역: 수어와 일대일로 대응되는 의미
- 의역: 맥락과 뉘앙스를 고려하여 문법에 맞는 완전한 목표어(한국어)로 번역
- 단, 직역과 의역이 같은 경우 두 티어에 같은 내용을 기재한다.

예시1: 직역과 의역의 차이를 잘 드러내는 예시

직역B [48]	<유등>축제. 항상 10월, 늘 첫째 주에 해요.
의역B [47]	유등축제는 늘 매년 10월 첫째 주에 열려요.

설명: [항상] 수어는 ‘항상’ 이라는 의미를 가지고 있지만 위의 문장에서는 ‘매년’ 을 뜻하기 때문에 의역에서 ‘매년’ 이라고 기술하도록 한다. [전통] 수어는 ‘늘 변함없이, 똑같이’ 라는 의미를 가지고 있으므로 ‘늘’ 이라고 기술한다. 또한 목표어(한국어)에서 자연스러운 어순과 문법을 고려하여 의역에서는 단일 문장으로 기술한다.

예시2: 직역과 의역의 차이를 잘 드러내는 예시

직역B [48]	"야, 축제에 와! 축제에 와!"	가서 보니, 오! 제법 발전했더라고요.
의역B [47]	축제에 하도 오라고 해서 가 보았더니 예상외로 눈에 띄게 발전했더라고요.	

설명: 구성된 행위(constructed action)로 나타나는 수어문은 직역에서는 직접인용으로 번역하고 의역에서는 간접인용으로 번역함을 원칙으로 한다. 또한, ‘가서 보니’의 경우 비수지로만 표현이 되었는데, 이처럼 비수지로 의미가 표현될 때 따로 제스처로 표기하지 않고 직역과 의역에 그 의미를 반영하여 기술하도록 한다.

예시3: 직역과 의역이 같은 경우

직역A [52]	당신 차례예요.
의역A [51]	당신 차례예요.

설명: 마우스 제스처가 수어 없이 독립적으로 나타나는 경우 분절은 하되 제스처의 내용을 따로 표기하지 않으며, 이 때 직역과 의역은 같을 수 있다.

⑤ “(())” 기호 사용

- 번역 내용이 확실하지 않거나 이해되지 않는 경우 “(())” 으로 표기하고 추측되는 내용을 번역정비_코멘트_A/B 티어에 기술한다.

⑥ [지시]에 대한 구체적인 내용 동반 표기

- [지시]가 나타나는 수어문의 번역 시, 직역에서는 ‘이것’, ‘저것’, ‘여기’, ‘저기’ 등으로 번역하나, 의역에서는 한국어 지시어휘 다음에 ‘()’로 지시가 가리키는 구체적 내용을 표기한다. 이는 [지시]를 통해 이루어지는 발화의 내용에 대한 독자의 이해를 돕기 위함이다.

예시: 직역과 의역에서 지시수어 번역

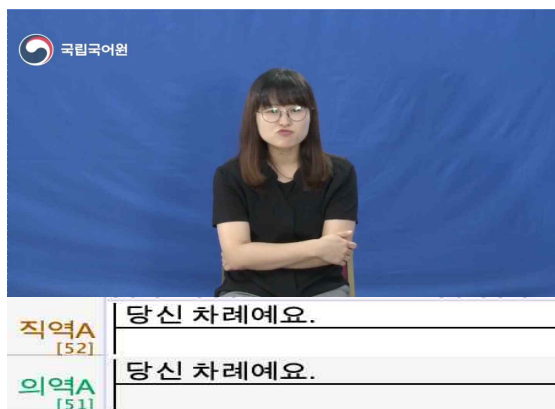
직역B [48]	여기처럼 강에 유등들이 없었어요. 예전에는 없었어요.
의역B [47]	여기(모니터)처럼 강에 떠 있는 유등들이 예전에는 없었어요.

설명: [지시]수어를 직역과 의역에서 모두 ‘여기’로 번역하였으나, 의역에서는 ‘여기’가 무엇을 뜻하는지 구체적인 내용을 괄호 안에 기술한다. 위 예시에서는 언어제공자가 [지시]수어로 촬영실에 있는 모니터를 가리키고 있으므로 의역에 ‘(모니터)’를 추가적으로 기술하였다.

⑦ 제스처가 사용된 문장의 번역 시 직역과 의역에 반영

- 언어제공자 간의 고개 끄덕임, 고개 가로저음과 같은 제스처가 빈번하게 나타나는 경우가 있는데, 이 경우 제스처를 따로 코멘트 티어에 기술하지 않으며, 그 의미를 번역하여 직역과 의역에 반영한다.
- 마우스 제스처가 수어 없이 독립적으로 나타나는 경우 분절은 하되 제스처를 따로 코멘트 티어에 기술하지 않으며, 그 의미를 번역하여 직역과 의역에 반영한다.

예시: 마우스 제스처가 독립적으로 나타나는 경우



① 기타

- 구어체에서 문장의 높임 표현은 비격식체인 ‘해요체(-어요, -해요)로 통일한다.
- 감탄사 “아”가 긴 경우 (예: 아~~~~!) 한국어 맞춤법 기준에 따라 ‘아하’로 기술한다.
- 번역 대안은 번역정비_코멘트_A/B 티어에 추가할 수 있다.
- 지문자나 수어 없이 마우스가 독립적으로 나타나는 경우

번역정비_코멘트_A/B 티어에 ‘마우딩: OOO’ 으로 기술한다.

- 언어제공자가 지문자를 사용하는 경우 <ㄱㅏㄴㅏㄴㅏ>가 아닌 <가나다>로 기술한다.

예시: 직역과 의역에서 지문자 번역

직역A [52]	아! <유등> 축제.
의역A [51]	아! 유등 축제요.

설명: 위 예시에서는 언어제공자가 ‘유등’이라는 한국어 단어를 지문자를 사용하여 발화하였다. 이를 번역할 때, ‘<ㅇㅠㄷㅡㅇ>’ 으로 기술하지 않고 ‘<유등>’ 으로 기술한다.

- ㅎㅎ: (웃음)으로 대체하여 의미정보로 추가한다.
- ‘(황당해 함)’, ‘(웃음)’, 등 맥락적 정보에 의해 나타나는 의미는 괄호로 처리한다.
- ㅎ, 헉, ^^ 와 같은 인터넷 용어 혹은 채팅어를 사용하지 않는다. 단, ‘헉’의 경우 지문자로 사용되었을 시 직역에서는 사용을 허용한다. 또한 ‘헉’과 같은 감탄사는 한국어에서 ‘어쩜’, ‘이럴 수가’, ‘어머나’, ‘세상에’ 등 여러 가지 어휘로 번역될 수 있으므로 맥락과 뉘앙스를 고려하여 가장 적절한 한국어 대응 표현으로 번역한다.

예시: 지문자 ‘헉’의 예외적인 허용 및 번역의 예

직역B [48]	내가 또? 당신은요? 내가 또 해요? <헉>.
의역B [47]	내가 또 해요? 당신이 아니라 또 내가 해요? (황당해 함)

- 설명: 언어제공자가 ‘헉’이라는 한국어 표현을 지문자로 발화하여 직역에서는 ‘<헉>’으로 기술하였다. 의역에서는 ‘헉’을 맥락적 정보로 풀이하여 괄호 안에 그 의미를 기술하였다.
- 수어이름(SN)의 경우 직역에서는 OOO(수어이름), 의역에서는 OO O으로 기재한다.

○ 번역 정비 파일 선정 및 번역 정비 교육

2022년까지 번역된 말뭉치 자료 중 아래와 같은 파일들을 선정하여 번역 정비 작업자한테 분배하여 정비를 진행하였다.

	번역/주석	영상시간	과제	A00
서울지역 10번쌍 G	번역/주석	18분 15초	특정사건	명00
서울지역 14번쌍 H	번역/주석	9분 3초	주제영역	안00
경남지역 4번쌍 H	번역/주석	13분 30초	주제영역	송00
서울지역 31번쌍 J	번역/주석	7분 59초	농사회행사	명00
서울지역 1번쌍 E	번역/주석	17분 3초	토론	손00
강원지역 1번쌍 H	번역/주석	14분 22초	주제영역	정00
경남지역 2번쌍 J	번역/주석	2분 41초	농사회행사	송00
서울지역 20번쌍 J	번역/주석	12분 13초	농사회행사	송00
경남지역 7번쌍 H	번역	22분 16초	주제영역	안00
경남지역 11번쌍 H	번역	27분 6초	주제영역	명00
경남지역 12번쌍 H	번역	14분 32초	주제영역	손00

표에 제시되어 있듯이 파일 11개 중 마지막 파일 3개는 수지 및 의미 주석이 없는 자료이다. 번역 정비 작업을 한 파일 11개는 총 159분이다. 번역 정비는 주석 정비와 달리 본 연구에서 처음으로 진행되는 작업이라 경험 있는 사람을 구할 수가 없었다.

최대한 많은 번역 정비 작업자들이 교육에 참여할 수 있도록 동일한 교육을 3번 하였다.

번역 정비 교육

- ◆ 일시: 2023.9.12. 14:00~15:30시, 2023.9.15. 10:30~12:00, 2023.9.22. 21:00~22:30
- ◆ 장소: 강남대학교 창조산학관 120호, 비대면(줌)
- ◆ 참가자: 번역 정비 작업자 총 6명, 연구원 4명
- ◆ 주요 안건: 번역 정비 교육 및 번역 정비 지침 설명, 필요 시 엘란 교육
번역 교육에서는 번역 정비 및 번역 정비 지침을 설명하고 필요 시,

엘란 교육도 실시하였다. 교육 후의 문의는 구글 드라이브를 활용하여 실시간으로 연구진과 피드백을 주고받았다.

[illegible]

○ 번역 정비 중간 검토 및 결과

정비 진행 상황을 중간에 검토한 결과, 엘란 프로그램 사용이 익숙하지 않거나 번역 분절에 대한 설명이 명확하지 않아 분절 수정이 제대로 이루어지지 않은 문제점들이 발견되었다. 또한 번역 지침에 대한 교육과 더불어 시범 번역 파일이 제공되었지만 번역자마다 지침에 대한 해석이 달라 지침을 준수하지 않은 문제들이 다수 발견되었다. 이에 연구진은 각 번역자마다 발견되는 문제들을 파악하여 개별적으로 재교육 및 수정 요청을 했다. 일관성 있는 번역 결과물을 위해 번역 분절, 직역과 의역의 차이에 대한 더욱 구체적인 설명 등 번역 지침 보완이 필요할 것으로 보인다.

4. 한국수어 말뭉치 공개 자료

최종 공개할 자료의 선정은 다음과 같은 기준을 고려하였다.

- 언어제공자의 동의서를 받은 자료
- 공개 검토에서 공개 적합으로 판단 된 자료 (2.1 참고)
- 자연스러운 한국수어 사용이 잘 드러나는 자료
- 지역 균형을 반영한 자료
- 성별 균형을 반영한 자료
- 연령 비율을 반영한 자료

1) 공개 파일 및 시간

2024년 공개할 파일 28개는 (총 432분 30초) 아래와 같다.

파일이름	시간	영상	번역	주석
서울지역 1번쌍 E-토론	0:17:03	○	○	○
서울지역 10번쌍 G-특정사건	0:18:15	○	○	○
강원지역 1번쌍 H-주제	0:14:22	○	○	○
경남지역 4번쌍 H-주제	0:13:30	○	○	○
서울지역 14번쌍 H-주제	0:09:03	○	○	○
경남지역 2번쌍 J-농행사	0:02:41	○	○	○
서울지역 20번쌍 J-농행사	0:12:13	○	○	○
서울지역 31번쌍 J-농행사	0:07:59	○	○	○
경남지역 11번쌍 H-주제	0:27:06	○	○	
경남지역 12번쌍 H-주제	0:14:3	○	○	
경남지역 7번쌍 H-주제	0:22:16	○	○	
강원지역 7번쌍 B-준비해온 농담이야기	0:07:47	○		
강원지역 10번쌍 F-표지판	0:37:47	○		
전북지역 15번쌍 Q-수어 어휘말하기	0:11:42	○		
전남지역 10번쌍 A-수어이름	0:04:47	○		
전남지역 10번쌍 B-준비해 온 농담	0:06:08	○		
전남지역 15번쌍 D-다시이야기	0:08:46	○		
충남지역 11번쌍 G-사건	0:10:20	○		

충남지역 15번쌍 K-흥부와 놀부	0:07:31	O		
경남지역 8번쌍 L-옛날 수어와 현재 수어	0:04:06	O		
충남지역 8번쌍 L-옛날 수어와 현재 수어	0:18:44	O		
전남지역 7번쌍 M-농인이야기 (일부)	0:30:52	O		
충남지역 11번쌍 N-가장 기억에 남는 날	0:24:23	O		
강원지역 7번쌍 O-취미	0:17:32	O		
전남지역 29번쌍 P-우리고장이야기	0:36:53	O		
전북지역 15번쌍 R-농에피소드	0:20:40	O		
충남지역 1번쌍 N-가장 기억에 남는 날	0:19:33	O		
충남지역 1번쌍 R-농에피소드	0:05:59	O		

2) 지역별

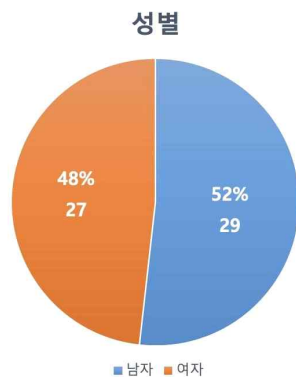
공개 파일들을 지역별로 정리하면 아래와 같다.



지역	파일 개수	시간
합계	28	432분 30초

3) 성별

공개 파일의 언어제공자 성별 비율은 아래와 같다.



성별	비율(명)
남자	52%(29)
여자	48%(27)
합계	100%(56)

4) 과제별

과제는 최대한 다양하게 선정하도록 하였다. 공개 파일들을 과제별로 정리하면 다음과 같다.

과제	파일 개수	시간
A- 수어이름	1	4분 47초
B-준비해온 농담 이야기	2	13분 55초
D-다시이야기	1	8분 46초
E-토론	1	17분 3초
F-표지판	1	37분 47초
G-특정사건	2	28분 35초
H-주제	6	100분 49초
J-농행사	3	22분 53초
K-흥부와 놀부	1	7분 31초
L-옛날 수어와 현재 수어	2	22분 50초
M-농이야기	1	30분 52초
N-가장 기억에 남는 날	2	43분 56초
O-취미	1	17분 32초
P-우리 고장 이야기	1	36분 53초
Q-수어어휘 말하기	1	11분 42초
R-농에피소드	2	26분 39초
합계	28	432분 30초

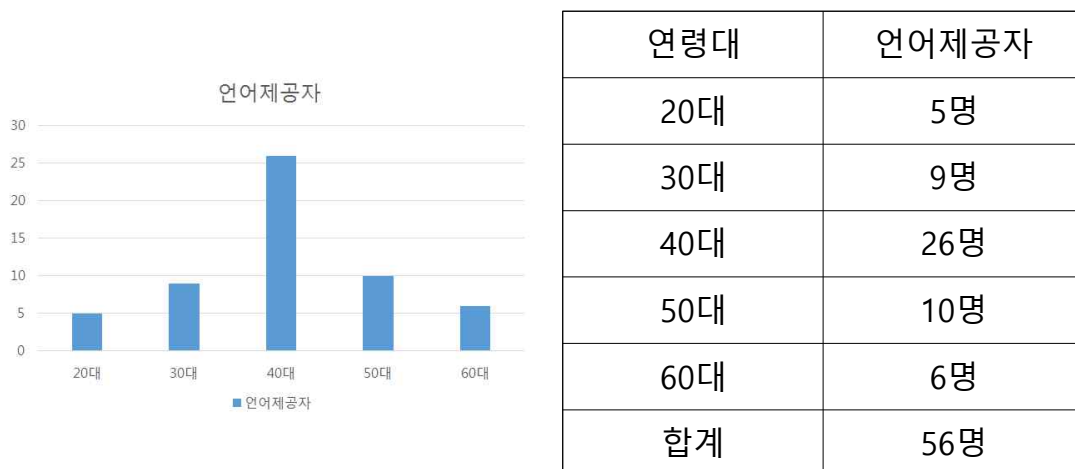
5) 언어제공자 연령별 분포

언어제공자의 연령은 촬영했던 시기에 연령으로 제시한다. 예를 들어, 1967년에 태어난 언어제공자는 현재 56살이지만 2015년 촬영 당시에는 48살이었기 때문에 40대로 분류하였다.

수집년도	생년A	수집 시 연령 A	연령대A	생년B	수집 시 연령 B	연령대B
2015	1973	42	40대	1987	28	20대
2015	1967	48	40대	1973	42	40대
2018	1978	40	40대	1963	55	50대

2019	1977	42	40대	1973	46	40대
2015	1954	61	60대	1953	62	60대
2018	1958	60	60대	1960	58	50대
2015	1963	52	50대	1955	60	60대
2015	1991	24	20대	1995	20	20대
2019	1974	45	40대	1972	47	40대
2019	1981	38	30대	1980	39	30대
2019	1979	40	40대	1973	46	40대
2021	1984	37	30대	1976	45	40대
2021	1972	49	40대	1972	49	40대
2021	1979	42	40대	1970	51	50대
2020	1973	47	40대	1981	39	30대
2020	1973	47	40대	1981	39	30대
2020	1965	55	50대	1974	46	40대
2022	1980	42	40대	1980	42	40대
2022	1972	50	50대	1981	41	40대
2019	1962	57	50대	1973	46	40대
2022	1966	56	50대	1979	43	40대
2020	2000	20	20대	1990	30	30대
2022	1980	42	40대	1980	42	40대
2021	1984	37	30대	1976	45	40대
2021	1960	61	60대	1955	66	60대
2021	1979	42	40대	1970	51	50대
2022	1991	31	30대	2001	21	50대
2022	1991	31	30대	2001	21	20대

공개되는 언어제공자들의 연령별 분포는 아래와 같다.



자료가 공개되는 언어제공자의 연령은 20대가 5명, 30대가 9명, 40대가 26명, 50대가 10명 그리고 60대가 6명이다. 40대가 제일 많은 이유는 수집 때부터 40대가 제일 많이 참여했었으며, 또한 한국수어 사용이 제일 잘 드러나는 연령대이기 때문이다.

6) 공개 파일의 티어명

최종 공개에 앞서 엘란파일 내에서 사용되는 층렬명을 일정한 규칙에 따라 일관성 있게 설정 및 수정할 예정이다. 기존 티어명에 있던 ‘_’ 는 삭제하고 영어 단어 ‘Gloss’ 를 한글로 바꾸어 기재하였으며 띄어쓰기가 없는 것을 원칙으로 하여 수정하였다. 수정된 층렬명 목록은 다음과 같다.

현 층렬명	수정 후
번역A	직역A
의역A	의역A
우세GlossA	글로스우세A
의미_우세A	의미우세A
비우세GlossA	글로스비우세A
의미_비우세A	의미비우세A

번역B	직역B
의역B	의역B
우세GlossB	글로스우세B
의미_우세B	의미우세B
비우세GlossB	글로스비우세B
의미_비우세B	의미비우세B

Ⅳ. 한국수어 말뭉치 공개 및 홍보를 위한 동영상 제작

한국수어 말뭉치는 다양한 수어 연구와 기술 개발 등에 활용될 수 있어야 한다. 2019년부터 구축된 주석 말뭉치에서는 의미 주석이 함께 수행됨으로써 한국수어 말뭉치 기반 한국수어 연구 및 인공지능 기술 개발에 대한 과학적이고 체계적인 방법론을 제공할 것이라는 기대가 있었다. 이번 한국수어 말뭉치 배포·공개는 학계, 연구기관, 산업체 등에서 한국수어 말뭉치 자료를 이용한 연구를 할 수 있는 기반을 조성하는 계기가 될 것이다. 즉, 한국수어 말뭉치의 공개는 급속히 발전되고 있는 컴퓨터 기술과 관련 소프트웨어를 활용한 인공지능과 접목되면서 효율적인 한국수어 학습 방법의 개발과 각종 사회 전반에 걸쳐 수어 의사소통 가능성을 높일 수 있는 길을 열게 될 것이며, 특히 수어 연구자들에게 생생한 수어 연구자료를 제공함으로써 한국수어 발전에 이바지하게 될 것이다.

하지만 아무리 좋은 자료라 할지라도 수어 언중에게 이것이 알려지지 않는다면 그 가치를 드러내지 못할 것이다. 이에 한국수어 말뭉치를 공개·배포하는 국립국어원 ‘모두의 말뭉치’의 회원가입 및 신청 방법 안내 수어 동영상, ‘모두의 말뭉치’ 내 한국수어 말뭉치 소개 자료 수어 영상, 그리고 한국수어 말뭉치 수집에서 주석 작업을 하고 시스템에 탑재되기까지의 전 과정을 설명하는 홍보 동영상을 제작하였다.

1. 홍보영상 제작 절차

~ 8월 4주	제작업체 계약: 수어민들레 신청 방법, 약정서 콘텐츠 제작
~ 9월 1주	가입 안내 및 홍보 동영상 스토리 보드 제작
~ 10월 2주	가입 안내 및 홍보 동영상 스토리 보드 수정
10월3주~10월 4주	1편 수어말뭉치 가입 안내 동영상 1차 시안
11월 1주 ~ 11월 3주	3편 수어말뭉치 홍보 동영상 1차 시안
11월 4주 ~ 12월 1주	제1차 1편 및 3편 동영상 수정 보완 작업
12월 2주 ~ 12월 3주	제2차 1편 및 3편 동영상 수정 보완 작업 및 제출
12월 4주 ~ 1월 1주	2편 동영상 제작 및 제출

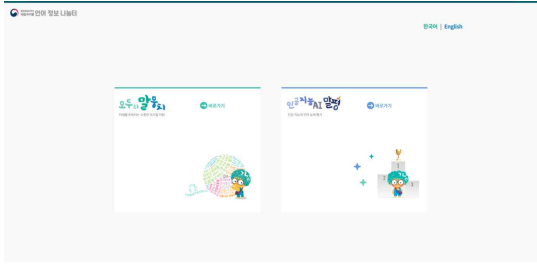
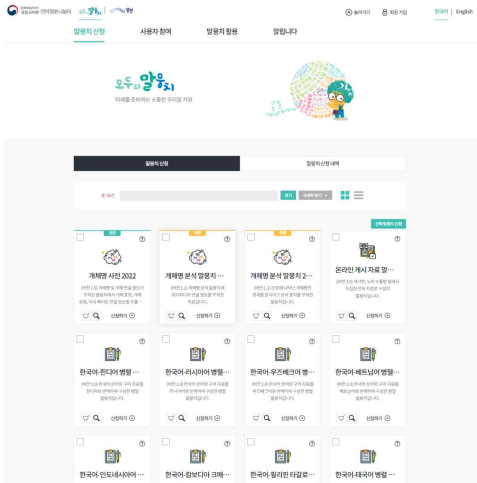
[그림 IV-1] 홍보영상 제작 절차

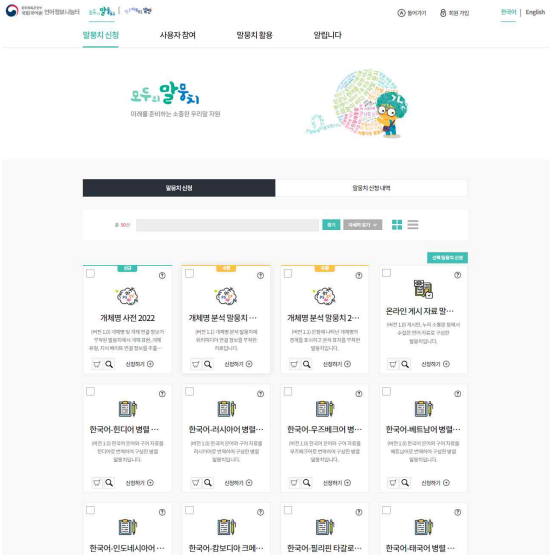
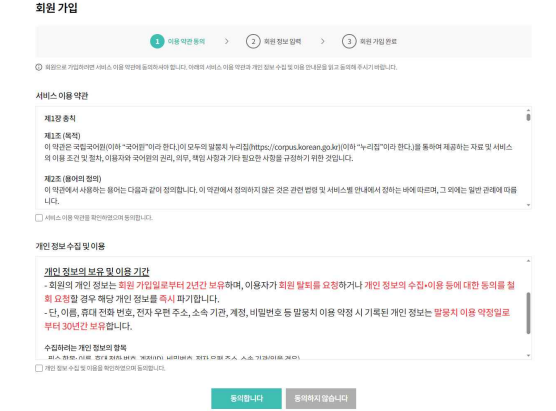
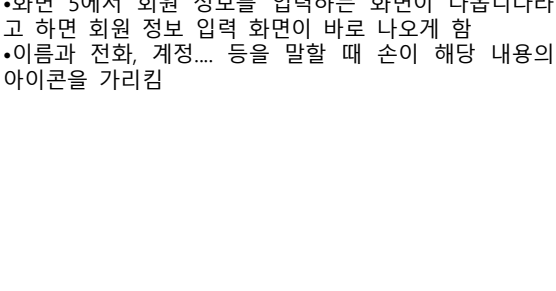
2. 홍보영상 촬영 조건 및 촬영 시스템


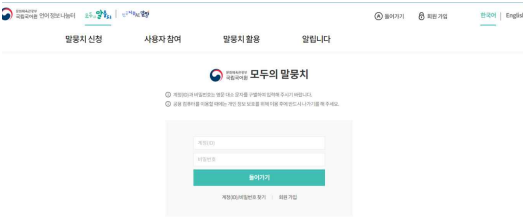
- 촬영 장소: 수어민들레 사무실, 한경국립대학교, 충청북도장애인복지관
- 촬영 환경
 - 수어민들레 사무실
 - 촬영실 크기 : 32㎡ 이상 (5.3m * 6m)
 - 촬영 공간 : 암막 커튼으로 빛 차단, 배경막 설치, 카메라 2대(사양full-HD,※ 소형 핸디캠 및 휴대폰 카메라 안됨), 조명 2대, 프롬프터, 삼각대 1대, 노트북 1대
 - 기타 준비물: 배경 천
 - 대학 연구실
 - 촬영 준비물: 카메라 2대, 조명 1대, 삼각대 1대, 노트북 1대
 - 충청북도장애인복지관
 - 촬영 준비물 : 촬영 준비물: 카메라 1대, 삼각대 1대, 노트북 1대
- 필요 인력 : 촬영 감독, 보조 인력 감수자
- 사전 촬영
 - 예비 촬영 : 촬영 시연 및 이상 유무 확인
 - 수연가, 음성통역자, 언어제공자 촬영 동의 영상
- 본 촬영
 - ‘모두의 말뭉치’ 회원가입 및 말뭉치 신청 방법 안내 수어 동영상
 - ‘모두의 말뭉치’ 내 한국수어 말뭉치 소개 수어 영상
 - 수집 및 분석과정이 드러나면서 한국수어 말뭉치의 활용 등, 전반적인 이해를 돕는 동영상
 - 주의사항: 추가 여분의 메모리카드 및 배터리 준비, 주석 과정은 화면녹화로 촬영

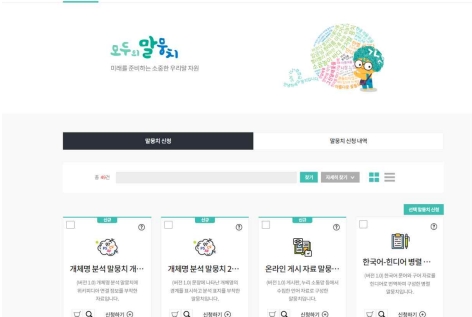












3. 홍보영상 스토리보드







1번째 홍보영상 : < ‘모두의 말뭉치’ 회원가입 및 말뭉치 신청 방법
안내 수어 동영상> : 한국수어 말뭉치 시스템을 통한 누구나 쉽게 접
근하고 이용하는 방법안내

영상 순서	비디오(Video, 화면) ※비디오(영상) 구현을 어떻게 할 건지 아이디어를 적어주세요.	내용(content)	해설 자막(Audio)
1	인트로(도입) 화면 제작을 만들레에서 해주세요	모두의 말뭉치 가 입 방법 안내	자막: 모두의 말뭉치 가입 및 한국수어 말뭉치 자료 신청 안내 해설: 안녕하세요. 모두의 말뭉치에 가입해서 한국 수어 말뭉치 자료를 신청 하는 방법에 대해 말씀드 리겠습니다.
2	<ul style="list-style-type: none"> 모두의 말뭉치 초기 화면을 제시하고, 화면 클릭하기 를 손가락 그림으로 보여주기 	[가입 과정]	<p>자막: 모두의 말뭉치 접속</p> <p>해설: 모두의 말뭉치 가입 을 위해서는 먼저 '모두의 말뭉치 사이트' https://kli.korean.go.kr/에 들어갑니다.</p> <p>모두의 말뭉치는 국립국 어원에서 우리나라의 다 양한 종류의 말뭉치를 모 아놓은 곳입니다.</p>
3	<ul style="list-style-type: none"> 국어원 모두의 말뭉치 사이트 캡처해서 보여주고, 손 가락 그림으로 한국수어 말뭉치 창 가리키면서 해설을 함 		<p>자막: 모두의 말뭉치 내에 있는 한국수어 말뭉치 창</p> <p>해설: '모두의 말뭉치' 사 이 트 (https://kli.korean.go.kr/co rpus/main/requestMain.d o?lang=ko)에는 한국수 어 말뭉치 창이 있습니다.</p>

4	<p>•회원가입 아이콘을 확대하면서 빨간색으로 동그라미 치고, 손 아이콘으로 클릭하기</p> 	회원 가입하기	<p>자막: 회원 가입하기</p> <p>해설: 한국수어 말뭉치 자료를 이용하기 위해서는 회원가입을 먼저 해야 합니다. 회원가입 버튼을 누르면 서비스 이용 약관과 개인 정보 수집 및 이용 안내문이 나옵니다.</p>
5	<p>•화면 4에서 개인정보 수집 및 이용 약관 화면...이 나온다고 하면 이용 약관과 정부 수집 및 이용 화면이 나오게 함</p> <p>서비스 이용 약관 <input type="checkbox"/>와 개인 정보 수집 및 이용 <input type="checkbox"/>에 클릭한 한 수</p> <p>동의합니다. 에 클릭하는 손 그림 넣기</p> 	이용 약관과 개인 정보 수집 및 이용 동의하기	<p>자막: 이용 약관과 개인 정보 수집 및 이용 동의</p> <p>해설: 서비스 이용 약관과 개인 정보 수집 및 이용 내용에 동의를 해야 회원가입 절차가 시작됩니다. 동의를 하시면 회원 정보를 입력하라는 화면이 나타납니다.</p>
6	<p>•화면 5에서 회원 정보를 입력하는 화면이 나옵니다라고 하면 회원 정보 입력 화면이 바로 나오게 함</p> <p>•이름과 전화, 계정.... 등을 말할 때 손이 해당 내용의 아이콘을 가리킴</p> 	회원 정보 입력	<p>자막: 회원 정보 입력</p> <p>해설: 회원 가입을 위해 본인인증을 위한 이름과 전화번호, 계정(ID), 비밀번호, 전자우편주소, 소속기관을 입력합니다. 입력을 다 한 후 다음을 누르면 입력하신 전자우편주소로 가입 승인 여부를 알려줍니다.</p>

	<div> <div>회원 가입</div> <div> <div>1</div> <div>이름·약관 동의</div> <div>2</div> <div>회원 정보 입력</div> <div>3</div> <div>회원 가입 완료</div> </div> </div> <div> <div>회원 정보 입력</div> <div> <div>이름</div> <div> <div> <div>필수 입력</div> </div> <ul style="list-style-type: none"> • 알뜰한이름(이름)을 입력해 본인 인증 후 입력 가능합니다. (로그인 전용으로 본인 인증을 해주셔야 합니다.) • 한 14세 미만(이)아이는 회원 가입이 불가능합니다. </div> </div> <div> <div>최대 전자 연배</div> <div> <div> <div>필수 입력</div> </div> <ul style="list-style-type: none"> • 영아(0~12개월), 소아(1~12개월) 중 2명 중 1명 이상 12세 이하(이)아이를 등록해야 합니다. • 글자(이)가 포함된 계정(ID)은 등록할 수 없습니다. (예: system_admin@naver.com) </div> </div> <div> <div>* 비밀번호</div> <div> <div> <div>필수 입력</div> </div> <ul style="list-style-type: none"> • 영아(0~12개월), 소아(1~12개월), 청소년(13~19세) 중 2명 중 1명 이상 12세 이하(이)아이를 등록해야 합니다. • 계정의료(과) 정보에 등록할 수 있습니다. </div> </div> <div> <div>* 비밀번호 확인</div> <div> <div> <div>필수 입력</div> </div> <ul style="list-style-type: none"> • 비밀번호(12자 이상)를 반드시 입력해 본인 인증 후 입력 가능합니다. (로그인 전용으로 본인 인증을 해주셔야 합니다.) • 글자(이)가 포함된 계정(ID)은 등록할 수 없습니다. (예: system_admin@naver.com) </div> </div> <div> <div>* 현재 주민등록 주소</div> <div> <div> <div>필수 입력</div> </div> <ul style="list-style-type: none"> • 알뜰한(이)를 반드시 입력해 본인 인증 후 입력 가능합니다. (로그인 전용으로 본인 인증을 해주셔야 합니다.) • 글자(이)가 포함된 계정(ID)은 등록할 수 없습니다. (예: system_admin@naver.com) </div> </div> <div> <div>* 소속 기관</div> <div> <div> <div>선택</div> <div>기관 정보를 입력해 현재 소속 기관(단체)을 등록할 수 있습니다.</div> </div> </div> <div> <div>이전</div> <div>다음</div> </div> </div> </div>		
7	한국수어 말뭉치 자료 화면	한국수어 말뭉치 이용 안내	<p>자막: 한국수어 말뭉치 자료 이용</p> <p>해설: 가입 승인을 받으셨으면 이제부터 자료 신청을 하실 수 있습니다.</p>
8	한국수어 말뭉치 배경에 해설자가 클로즈업 된 동영상	자료 신청 안내	<p>자막: 한국수어 말뭉치 자료 이용 방법</p> <p>해설: 한국수어 말뭉치 자료를 내려받으려면 몇 가지 단계를 거쳐야 합니다.</p>
9	<p>들어가기 아이콘에 붉은색으로 동그라미를 그리고 손 아이콘으로 클릭하기</p> 	자료 신청 안내	<p>자막: 첫째, 들어가기</p> <p>해설: 첫째, 모두의 말뭉치 전체 창의 위쪽에 있는 '들어가기'를 클릭합니다.</p>
10		자료 신청 안내	<p>자막: 둘째, 계정(ID)과 비밀번호 입력하기</p> <p>해설: 들어가기를 누르면 계정과 비밀번호를 입력하라는 창이 뜹니다. 여기에 자신의 계정(ID)과 비밀번호를 입력합니다.</p>

<p>11</p>		<p>자료 신청 안내</p>	<p>자막: 셋째, ‘한국수어 말뭉치’ 창으로 가기</p> <p>해설: 홈페이지 상단에 있는 메뉴에서 [말뭉치 신청]을 누릅니다.</p>
<p>12</p>	<p>·한국수어 말뭉치 그림 보여주면서 해설이 진행되는 것이 따라 해당 아이콘</p> <p>· ①, , , 신청하기  에 붉은 색으로 동그라미 치기</p>	<p>한국수어 말뭉치 창 소개</p>	<p>자막: ‘한국수어 말뭉치’ 창 소개</p> <p>해설: 한국수어 말뭉치 창에 있는 아이콘들의 기능을 설명드리겠습니다.</p> <p>‘한국수어 말뭉치’ 창 위쪽에는 ①이, 아래쪽에는 , , 신청하기  가 있습니다.</p>
<p>13</p>	<p>·한국수어 말뭉치 창 위쪽 ②에 붉은색으로 동그라미 그리기</p>	<p>한국수어 말뭉치 창 소개</p>	<p>자막: ‘한국수어 말뭉치’ 창 소개</p> <p>해설: ②는 한국수어 말뭉치를 간단하게 소개하는 자료를 PDF 파일로 내려받을 수 있습니다.</p>
<p>14</p>	<p>·한국수어 말뭉치 창 아래쪽  에 붉은색으로 동그라미 그리기</p>	<p>한국수어 말뭉치 창 소개</p>	<p>자막: ‘한국수어 말뭉치’ 창 소개</p> <p>해설:  는 한국수어 말뭉치를 신청해두는 장바구니 담기입니다. 여러 개의 말뭉치 자료를 장바구니에 담을 수 있습니다.</p>
<p>15</p>	<p>·한국수어 말뭉치 창 아래쪽  에 붉은색으로 동그라미 그리기</p>	<p>한국수어 말뭉치 창 소개</p>	<p>자막: ‘한국수어 말뭉치’ 창 소개</p> <p>해설:  는 한국수어 말뭉치 예시 자료를 보여줍니다.</p>
<p>16</p>	<p>·한국수어 말뭉치 창 아래쪽 신청하기  에 붉은색으로 동그라미 그리기</p>	<p>한국수어 말뭉치 창 소개</p>	<p>자막: ‘한국수어 말뭉치’ 창 소개</p> <p>해설: 신청하기  를 클릭하면 한국수어 말뭉치 자료를 신청할 수 있습니다.</p>

17	<div><div>개인정보 수집 및 이용 내역</div><div><div>국민권익위원회(이하 권익위)가 처리한 개인정보를 수집하여 이용합니다. 내역을 자세히 알으신 후 동의 여부를 결정하시어 주십시오.</div><table><tr><th>개인정보 수집·이용 내역</th><th>수집·이용 항목</th><th>수집·이용 목적</th><th>보유 기간</th></tr><tr><td>*개인정보 수집·이용 내역 이름, 생년월일, 전자 우편, 전화번호, 주소</td><td>- 맞춤형 이용목적에 따른 본인 확인 및 개인식별 - 맞춤형 이용 목적에 따른 정보 제공(응답)</td><td>30년</td></tr></table><div><small>※본 개인정보 수집·이용에 대한 동의를 신청할 순서가 없습니다. 다른 동의항목을 신청하는 경우, 등록된 이용자의 개인정보를 제공할 수 있습니다. (다. 동의 없이 개인정보를 수집 및 이용하시거나 동의하지 않으신 경우, <input type="checkbox"/> 동의 <input type="checkbox"/> 거부하</small></div><div>신청 정보</div><table><tr><td>*신청 발원처</td><td>한국어·인도네시아어·일본어 발원처 2021</td><td>*신청 사유</td><td>신청 사유를 선택해 주세요.</td></tr><tr><td>*이용 예정 기간</td><td>연월일 - <input type="text"/></td><td></td><td></td></tr><tr><td></td><td colspan="3"><input type="radio"/> 일용적 이용 기간을 일정에 주세요. 최대 이용 기간은 일용적 용량과 이용 목적 기간 이내로 제한됩니다.</td></tr><tr><td>*사업명/연구 과제명</td><td colspan="3"><input type="text"/></td></tr><tr><td>*사용자/사용기관 목적</td><td colspan="3">국적을 선택해 주세요.</td></tr><tr><td>*수령 기준</td><td colspan="3">수령 기준을 선택해 주세요.</td></tr><tr><td></td><td colspan="3">- 연구 개발 내용 요약 <input type="text"/></td></tr><tr><td></td><td colspan="3">- 연구 개발 기간 <input type="text"/></td></tr><tr><td>*구체적 신청 사유 (최대 300자)</td><td colspan="3">- 연구용 데이터 서비스 활용: <input type="text"/></td></tr><tr><td></td><td colspan="3"><input type="radio"/> 신청 사유를 자세히 입력해 주세요. 신청 사유가 불분명할 경우 반려될 수 있습니다. <input type="button" value="신청 사유 예시"/></td></tr></table><div><input type="button" value="신청"/> <input type="button" value="취소"/></div></div></div> <tr><td>18</td><td></td><td>자료 신청 안내</td><td>자막: 넷째, 자료 신청을 위한 정보 입력</td></tr> <tr><td>19</td><td><div><div>·엘란 사이트 그림의 <div>Download</div>에 붉은 색 동그라미 그리기</div><div><div>ELAN</div><div><div>IntroRelease notesDocumentationDownloadCitingSupportResources</div><div>ELAN is an annotation tool for audio and video recordings.</div><div>Screenshot 1Screenshot 2</div><div></div><div><small>A sample from the ACLEW project#.</small><small>A sample from the ACLEW project#.</small></div></div></div></div><tr><td>20</td><td>·인사 영상 + 한국수어 말뭉치 그래픽 + 감사합니다(텍스트)</td><td>마무리</td><td>국립국어원은 한국수어 말뭉치를 통해 한국수어가 더욱 발전하기를 기대합니다. 감사합니다.</td></tr></td></tr>	개인정보 수집·이용 내역	수집·이용 항목	수집·이용 목적	보유 기간	*개인정보 수집·이용 내역 이름, 생년월일, 전자 우편, 전화번호, 주소	- 맞춤형 이용목적에 따른 본인 확인 및 개인식별 - 맞춤형 이용 목적에 따른 정보 제공(응답)	30년	*신청 발원처	한국어·인도네시아어·일본어 발원처 2021	*신청 사유	신청 사유를 선택해 주세요.	*이용 예정 기간	연월일 - <input type="text"/>				<input type="radio"/> 일용적 이용 기간을 일정에 주세요. 최대 이용 기간은 일용적 용량과 이용 목적 기간 이내로 제한됩니다.			*사업명/연구 과제명	<input type="text"/>			*사용자/사용기관 목적	국적을 선택해 주세요.			*수령 기준	수령 기준을 선택해 주세요.				- 연구 개발 내용 요약 <input type="text"/>				- 연구 개발 기간 <input type="text"/>			*구체적 신청 사유 (최대 300자)	- 연구용 데이터 서비스 활용: <input type="text"/>				<input type="radio"/> 신청 사유를 자세히 입력해 주세요. 신청 사유가 불분명할 경우 반려될 수 있습니다. <input type="button" value="신청 사유 예시"/>			18		자료 신청 안내	자막: 넷째, 자료 신청을 위한 정보 입력	19	<div><div>·엘란 사이트 그림의 <div>Download</div>에 붉은 색 동그라미 그리기</div><div><div>ELAN</div><div><div>IntroRelease notesDocumentationDownloadCitingSupportResources</div><div>ELAN is an annotation tool for audio and video recordings.</div><div>Screenshot 1Screenshot 2</div><div></div><div><small>A sample from the ACLEW project#.</small><small>A sample from the ACLEW project#.</small></div></div></div></div> <tr><td>20</td><td>·인사 영상 + 한국수어 말뭉치 그래픽 + 감사합니다(텍스트)</td><td>마무리</td><td>국립국어원은 한국수어 말뭉치를 통해 한국수어가 더욱 발전하기를 기대합니다. 감사합니다.</td></tr>	20	·인사 영상 + 한국수어 말뭉치 그래픽 + 감사합니다(텍스트)	마무리	국립국어원은 한국수어 말뭉치를 통해 한국수어가 더욱 발전하기를 기대합니다. 감사합니다.
개인정보 수집·이용 내역	수집·이용 항목	수집·이용 목적	보유 기간																																																							
*개인정보 수집·이용 내역 이름, 생년월일, 전자 우편, 전화번호, 주소	- 맞춤형 이용목적에 따른 본인 확인 및 개인식별 - 맞춤형 이용 목적에 따른 정보 제공(응답)	30년																																																								
*신청 발원처	한국어·인도네시아어·일본어 발원처 2021	*신청 사유	신청 사유를 선택해 주세요.																																																							
*이용 예정 기간	연월일 - <input type="text"/>																																																									
	<input type="radio"/> 일용적 이용 기간을 일정에 주세요. 최대 이용 기간은 일용적 용량과 이용 목적 기간 이내로 제한됩니다.																																																									
*사업명/연구 과제명	<input type="text"/>																																																									
*사용자/사용기관 목적	국적을 선택해 주세요.																																																									
*수령 기준	수령 기준을 선택해 주세요.																																																									
	- 연구 개발 내용 요약 <input type="text"/>																																																									
	- 연구 개발 기간 <input type="text"/>																																																									
*구체적 신청 사유 (최대 300자)	- 연구용 데이터 서비스 활용: <input type="text"/>																																																									
	<input type="radio"/> 신청 사유를 자세히 입력해 주세요. 신청 사유가 불분명할 경우 반려될 수 있습니다. <input type="button" value="신청 사유 예시"/>																																																									
18		자료 신청 안내	자막: 넷째, 자료 신청을 위한 정보 입력																																																							
19	<div><div>·엘란 사이트 그림의 <div>Download</div>에 붉은 색 동그라미 그리기</div><div><div>ELAN</div><div><div>IntroRelease notesDocumentationDownloadCitingSupportResources</div><div>ELAN is an annotation tool for audio and video recordings.</div><div>Screenshot 1Screenshot 2</div><div></div><div><small>A sample from the ACLEW project#.</small><small>A sample from the ACLEW project#.</small></div></div></div></div> <tr><td>20</td><td>·인사 영상 + 한국수어 말뭉치 그래픽 + 감사합니다(텍스트)</td><td>마무리</td><td>국립국어원은 한국수어 말뭉치를 통해 한국수어가 더욱 발전하기를 기대합니다. 감사합니다.</td></tr>	20	·인사 영상 + 한국수어 말뭉치 그래픽 + 감사합니다(텍스트)	마무리	국립국어원은 한국수어 말뭉치를 통해 한국수어가 더욱 발전하기를 기대합니다. 감사합니다.																																																					
20	·인사 영상 + 한국수어 말뭉치 그래픽 + 감사합니다(텍스트)	마무리	국립국어원은 한국수어 말뭉치를 통해 한국수어가 더욱 발전하기를 기대합니다. 감사합니다.																																																							

2번째 홍보영상 : < ‘모두의 말뭉치’ 내 한국수어 말뭉치 소개 자료 수어 동영상> : 공개되는 한국수어 말뭉치의 형식 등에 대한 안내 보기.

비밀	장면 설명	자막
1	영상 시작	사회자: 이 자료의 이름은 국립국어원 한국수어 말뭉치입니다.
2	자료 유형: MP4, 엘란 파일, JSON 파일	사회자: 자료 유형은 MP4, 엘란 파일, JSON 파일

		로 구성되어 있습니다.
3	한국수어 말뭉치 기반 조사 및 시범 구축(2015) 한국수어 말뭉치 연구 및 구축(2017, 2018, 2019, 2020) 한국수어 말뭉치 수집 및 분석(2021, 2022, 2023)	사회자: 관련 사업은 자막의 내용과 같습니다.
4	자료에 대한 설명	사회자: 농인 사회자가 제시하는 13개의 유도 과제(2021년 이후는 추가 개발한 5개의 유도 과제 포함 총 18개)를 통해 농인 언어제공자 2명이 자연스럽게 발화한 내용 18개의 유도과제를 보고 농인 언어제공자 2명이 자연스럽게 수어로 이야기하는 것을 수집한 것입니다. 자료의 종류에는 영상만 있는 원시 한국수어 말뭉치 자료, 주석과 번역이 제공되는 자료, 그리고 번역만 제공되는 자료의 3가지가 있습니다.
5	분량 - 원시 말뭉치 영상: 273분 30초 - 번역과 주석(글로스)이 제공되는 영상: 95분 6초 - 번역만 제공되는 영상: 63분 54초	사회자: 공개되는 자료의 분량은 자막에 나와 있는 것과 같습니다.
6	영상 및 주석 파일 명명 규칙 주석 파일: [지역명]+[언어제공자 쌍번호]+[과제 코드-과제명] <예시> 강원지역+3번쌍+E-토론 인용 (국문) 국립국어원(2024). 국립국어원 한국수어 말뭉치 (버전 1.0) URL: https://kli.korean.go.kr/ (영문) https://kli.korean.go.kr/?lang=en	사회자: 영상파일과 주석파일의 이름을 붙여놓은 규칙입니다. 영상 파일: [지역 코드]-[언어제공자 코드]-[과제 코드]-[촬영 방향 코드] 영상 파일의 이름은 지역, 언어제공자, 과제, 촬영 방향의 순서로 구성되어 있습니다. 사회자 : 화면의 예를 보시기 바랍니다. KSLGW3ELt: KSL-한국수어, GW-강원지역 3-3번째 언어제공자쌍, E-토론 과제, Lt- 왼쪽 언어제공자 영상, KSL은 한국수어, GW는 강원지역, 3은 3번째 언어제공자쌍, E는 토론 과제, Lt- 왼쪽 언어제공자 영상임을 코드로 나타내었습니다.
7	주석 및 번역 파일(ELAN) 예	자막: 사회자: 화면의 그림은 엘란을 이용한 주석 및 번역 파일의 예입니다. 해설: 내려 받으신 한국수어 말뭉치 자료는 엘란(ELAN, https://archive.mpi.nl/tla/elan) 프로그램을 통해 수어학 정보를 볼 수 있습니다.
8	주석 - 주석은 특정 수어 발화에 글로스로 수어학적 정보를 부여한 것이다. - 글로스는 수어 발화에 한글 라벨을 붙인 것으로 번역과는 다르다(한국수어 말뭉치 주석 규약 참조).	사회자: 주석이란 수어에 글로스로 수어학적 정보를 입력하는 것입니다. 글로스는 번역이 아니고 한글 이름을 붙인 것입니다.
9	번역 - 지역: 수어의 구조를 중심으로 번역	사회자: 번역은 수어의 구조를 중심으로 번역한 지역을 제공합니다. 한국어로 의미가 자연스럽게 전달되도록 한 의역은 추후 공개 예정입니다.
10	총괄 설명 지역A: 왼쪽 언어제공자의 발화에 대한 한국어 지역 글로스우세A: 왼쪽 언어제공자의 우세손에서 발화되는 수어의 글로스 의미우세A: 우세GlossA에 제시되는 수어 단어의 의미 글로스비우세A: 왼쪽 언어제공자의 비우세손에서 발화되는 수어의 글로스 의미비우세A: 비우세GlossA에서 제시되는 수어 단어의 의미 지역B: 오른쪽 언어제공자의 발화에 대한 한국어 지역 의역B: 오른쪽 언어제공자의 발화에 대한 한국어 의역 글로스우세B: 오른쪽 언어제공자의 우세손에서 발화되는 수어의 글로스 의미우세B: 우세GlossB에 제시되는 수어 단어의 의미 글로스비우세B: 오른쪽 언어제공자의 비우세손에서 발화되는 수어의 글로스 의미비우세B: 비우세GlossB에 제시되는 수어 단어의 의미	사회자 :주석을 기록하는 총괄에 대한 설명은 화면의 내용과 같습니다.
	자료 내용 문의: 02-2669-9613, kwakee2022@korea.kr	사회자: 자료에 대한 문의는 02-2669-9613, kwakee2022@korea.kr로 하시기 바랍니다.

3번째 홍보영상 : <한국수어의 말뭉치 소개> 한국수어 말뭉치가 공개되기까지의 전 과정을 보여주며 한국수어 말뭉치를 전체적으로 홍보하는 영상

동영상 번호	내용(content)	해설 자막(Audio)
1	한국수어 말뭉치 소개	한국수어 말뭉치 소개
2	들어가기 농인의 언어, 한국수어 말뭉치	사회: 국립국어원은 한국수어의 발전을 위해 2015년부터 한국수어 말뭉치를 구축해오고 있습니다. 오늘은 한국수어 말뭉치가 어떻게 만들어지며, 이것이 농사회와 한국수어 발전에 어떤 영향을 미칠지에 대해 알아보겠습니다.
3	한국수어 말뭉치 정의	사회: 국립국어원은 여러 종류의 음성언어 말뭉치들과 함께 한국수어 말뭉치도 구축하고 있는데, 한국수어 말뭉치가 무엇일까요? 전문가: 한국수어 말뭉치란 농인이 실제 사용하는 수어 자료를 모아 컴퓨터로 분석하고 처리할 수 있도록 데이터베이스에 모아둔 대규모 자료를 말합니다.
4	한국수어 말뭉치 자료를 만드는 목적	사회: 국가에서 한국수어 말뭉치를 구축하는 목적은 무엇일까요? 전문가: 농인의 언어문화를 보전하고, 실제로 사용되고 있는 수어를 체계적으로 이해하려는 데 있습니다.
5	장면 전환	사회: 한국수어 말뭉치는 여러 단계의 복잡한 과정을 거쳐 구축됩니다. 함께 한국수어 말뭉치가 만들어지는 과정을 살펴보실까요?
6	말뭉치 수집을 위한 지역 결정	사회: 한국수어 말뭉치 구축은 먼저 수어를 수집할 지역을 결정합니다. 2015년 서울/경기 지역을 시작으로 2023년까지 전북, 광주/전남, 강원, 경남, 대전/충남, 충북 등 7개 지역의 수어를 수집하였습니다.
7	지역 농아인협회지부의 역할	사회: 지역의 수어 수집은 해당 지역의 농아인협회가 중심 역할을 합니다. 지역농인협회관계자: 지역의 수어를 수집하는데 어려움이 많았지만 우리의 언어인 한국수어가 발전하는 데 기여한다는 생각으로 기쁘게 일했습니다.
8	지역 언어제공자 모집	사회: 한국수어 말뭉치 수집을 위해서 어떤 조건을 갖춘 언어제공자를 모으셨나요? 전문가: 언어제공자가 되기 위한 여러 가지 기준이 있지만 무엇보다 특정 지역의 수어를 잘 보존하고 있어야 합니다. 화면에 나오는 기준에 맞는 언어제공자를 찾기 위해 지역 농아인협회와 전문가들이 많은 의논을 하였습니다. 지금까지 7개 지역의 294명의 농인이 참여하였습니다.
9	유도자료 제작	사회: 언어제공자들이 자연스럽고 풍부한 한국수어를 말하도록 유도자료를 이용한다고 합니다. 유도자료란 무엇일까요? 전문가: 유도자료는 자연스러운 상황에서 실제 사용하는 수어를 수집하기 위해 수어 동영상, 사진이나 그림 등을 제시하여 언어제공자가 수어를 하도록 유도하는 자료입니다.
10	한국수어 말뭉치 수집 촬영	사회: 한국수어 말뭉치 수집은 약 4시간에 걸쳐 힘들게 이루어 집니다.

11	촬영	촬영은 저작권 동의서 등을 먼저 작성하고 시작됩니다. 한 쌍의 촬영은 대개 4시간 정도가 걸립니다.
12	원시 말뭉치	사회: 4시간 내외의 힘든 수집 활동을 통해 처음 수집된 한국수어 말뭉치를 한국수어 원시말뭉치라고 합니다. 이제 주석과 번역을 하는 중요한 과정이 기다리고 있습니다.
13	주석	사회: 한국수어 말뭉치에 주석을 한다는 것이 무엇일까요? 전문가: 주석은 수집된 원시 한국수어 말뭉치에 들어 있는 여러 가지 수어학 정보를 분석하여 기록하는 것입니다. 이런 주석은 수어 연구자들과 수어에 관심이 있는 사람들에게 소중한 정보를 제공할 것입니다. 주석은 '엘란'이라는 언어분석 프로그램을 활용하여 분석합니다.
14	번역	사회: 한국수어를 모르는 사람도 한국수어 말뭉치를 볼 수 있나요? 전문가: 수어를 잘 알고 있는 농이이나 연구자 이외에 수어를 모르는 사람도 이용할 수 있도록 한국어로 번역한 것도 있습니다.
15	모어 화자에 의한 검증	사회: 주석과 번역 작업이 마쳐진 한국수어 말뭉치는 수어 모어 화자인 농인들과 연구진들에 의해 검증을 합니다. 이 작업이 끝나면 드디어 한국수어 말뭉치가 데이터 베이스에 말뭉치 자료로서 보관이 됩니다.
16	한국수어 말뭉치의 쓰임새	사회: 이런 한국수어 말뭉치는 어디에 유용하게 사용될까요? 전문가: 구축된 한국수어 말뭉치는 수어연구의 기초 자료를 제공할 것입니다. 또한 수어사전 제작, 수어교육을 위한 기초 자료, 수어 관련 AI 개발 등에 기여할 것입니다.
17	마무리	사회: 한국수어 말뭉치는 농인의 문화와 언어를 더욱 발전시키고 수어를 귀한 유산으로 물려줄 수 있게 할 것입니다. 한국수어 말뭉치에 더 많은 관심을 가져 주시기를 부탁드립니다.

V. 요약 및 제언

1. 연구 결과 요약

2023년에는 아직까지 한국수어 말뭉치가 구축되지 않은 충북 지역의 원시말뭉치를 수집하였으며, 이와 함께 이번 연구에서는 지금까지 수집해 오던 수어 말뭉치를 공개하기 위해서 공개 검토 기준을 마련하고 이 기준에 부합하는 원시 한국수어 말뭉치 273분 30초, 번역만 정비된 한국수어 말뭉치 63분 54초, 번역과 주석이 모두 정비된 한국수어 말뭉치 95분 5초, 총 432분 30초를 공개함으로써 앞으로 한국수어 연구와 교육의 기초자료로 의미 있게 사용될 것으로 기대된다.

1) 한국수어 원시 말뭉치 수집

본 사업에서는 충북 지역의 한국수어 원시 말뭉치를 수집하였다. 수집 과정은 ‘지역 선정 - 유도자료 제작 - 언어제공자 선정 - 사회자 선정 및 교육, 촬영·편집자 교육 - 원시 말뭉치 촬영 - 편집 - 검수의 순서로 이루어졌다. 이에 따른 결과는 다음과 같다.

첫째, 기존 수집된 지역별 자료의 양과 유사하도록 총 15쌍의 언어제공자로부터 수어 말뭉치를 수집하는 과제였다. 하지만 수어 수집의 안정성을 확보하기 위하여 1쌍의 예비 언어제공자를 추가하여 총 16쌍의 언어제공자로부터 한국수어 말뭉치를 수집하였다.

둘째, 유도자료는 2021년에 추가 개발된 5개의 신규 유도 과제를 포함하여 총 18개의 유도 과제를 사용하였으며, 시의성과 지역성을 고려하여 ‘E-토론’, ‘G-특정 사건 묘사’, ‘H-주제 영역’, ‘I-개별 수어의 다양한 표현’, ‘J-농사회 행사’의 일부 주제를 수정하여 사

용하였다. 또한 추후 한국수어 말뭉치 공개 시 발생할 수 있는 저작권 문제를 해결하고자 ‘F-표지판’, ‘P-홍부 놀부 이야기’, ‘E-토론’의 그림은 기존의 유도 과제와 유사한 그림을 자체 제작하여 사용하였다.

설명 동영상 제작 시에는 다른 지역 수어의 간섭을 배제하고 지역 수어를 구사하는 언어제공자들이 안정감을 느끼도록 수집 지역 출신 사회자가 유도자료 설명 동영상을 촬영하였다.

셋째, 언어제공자 선정은 충북농아인협회의 도움을 받아 언어제공자 우선 선정 조건 등을 고려하여 예비 언어제공자 1쌍을 포함한 최종 16쌍의 언어제공자를 확정하였다.

넷째, 사회자 선정 및 교육에서는 언어제공자들의 대화를 돕기 위한 사회자를 선정하였는데, 언어제공자의 수, 촬영 일정 등을 고려하여 사회자 2명을 선정하였다. 수어 말뭉치 수집에 있어 사회자의 역할이 매우 중요하므로 사회자를 한곳에 모아 대면으로 본 사업에 대한 전체적인 내용과 함께 사회자의 역할에 대한 집합 교육을 하였다. 이번 사업에서는 촬영·편집자도 함께 교육을 받았으며, 2021년 사업에 참여하여 촬영·편집을 담당했던 담당자를 섭외하여 교육함으로써 보다 실질적인 내용을 교육할 수 있었다.

다섯째, 총 8일간 약 69시간 11분 57초 분량을 촬영하였으며, 이를 편집하여 최종적으로 총 59시간 31분 29초 분량의 한국수어 원시 말뭉치 영상을 확보하였다. 수집된 원시 말뭉치 영상은 이전에 수집된 영상과 추후 이루어질 수집 영상과의 통일성을 고려하여 편집 설명서에 따라 편집하였으며, 영상의 오류를 최소화하기 위하여 4명의 연구원에 의해 총 9차례에 걸친 검증 과정을 거쳤다.

2) 기 구축된 한국수어 말뭉치 공개를 위한 정비

둘째, 구축된 한국수어 말뭉치 공개를 위한 정비를 실시한다.

첫째, 한국수어 말뭉치 공개를 위한 정비에 앞서 기 구축된 한국수어 말뭉치 중에서 공개가 가능한 자료를 선별하였다, 특정 인물에 대한 부정적인 내용이나, 얼굴 이름이 노출되었거나, 성과 관련된 민감한 주제, 농협하나 농사회 행사에 대한 비난성 발언 내용 등은 공개가 불가능한 것으로 판정하여 28개 파일, 최종 432분 30초 분량의 공개가 가능하다고 판단되는 한국수어 말뭉치 자료를 선정하였다.

둘째, 주석 정비는 정비코멘트 층렬에 ###(모르는 수어)로 기록되어 있는 것들을 해결하였고, 원 주석 층렬 중 ‘의미’ 층렬이나 ‘코멘트’ 층렬에 의미를 삽입하거나 삭제한 경우, 정비코멘트 층렬에 “토큰 삭제” 또는 “토큰 삽입” 이라고 기록되어 있는 것을 정비 층렬에 “기타”로 기록하고 정비코멘트 층렬에 ‘의미삽입’ 또는 ‘의미 수정’으로 기록하였다.

그리고, 글로스 층렬에 %(수어를 하다가 중지)로 기록되어 있는 경우, 의미 층렬에 의미를 기록하지 않고 의미코멘트 층렬에 그 의미를 기록하였다. 정비와 정비코멘트의 분절이 윗 층렬(글로스 층렬, 의미 층렬 등)의 분절과 맞지 않는 경우에는 정비 층렬의 분절을 수정하였다. 같은 수어가 반복(예를 들어, 3번 반복)되는 경우는 각 토큰별로 분절 후 정비 층렬에 ‘토큰 삽입’으로 기록하고, 정비코멘트 층렬에는 반복되는 수어를 ‘분절’이라고 기록하였다.

셋째, 정비한 내용은 정비 층렬에 12가지의 통제 어휘를 사용하여 기록하였다. 통제 어휘는 분절수정, 손바뀜, A/B바뀜, 타입명 적합성, 토큰삽입, 타입수정, 의미수정, 지문자 수정, 합성어, ###, 기타로 설

정하였다. 총 주석 정비가 이루어진 사례는 2,057건이었다. 이중 ‘의미수정’이 1,259건으로 반 이상으로 많았으며, 그 다음 ‘타입수정’이 고정수어와 생산적 수어를 합쳐 249건으로 나타났으며, ‘토큰삽입’이 200건으로 나타났다, 이 세가지 수정 사례를 합하면 1,699건, 82.6%를 차지하여 주석 정비가 대부분 이 세 가지에서 필요하다는 것을 보여준다.

넷째, 번역 정비는 1차 자문회의 결과와 이를 토대로 강남지역 12번 쌍 H 파일을 갖고 5분 분량의 시범 번역을 하였으며, 시범 번역에서 확인된 시사점을 반영하여 번역 지침을 만들었다. 번역은 직역과 의역으로 구분하여 번역 지침을 마련하고, 올해 번역 정비에서는 직역번역을 실시하였다. 구체적인 번역 지침에는 다음과 같은 것들이 있다.

수어를 분절할 때, 비수지가 먼저 나타나더라도 시작점은 우세손과 비우세손을 기준으로 하는 출발어(수어)의 분절, 직역티어에는 발화된 수어의 형태 및 통사 구조가 드러나도록 한 번역문을 기록, 의역티어에는 발화된 수어 의미가 자연스러운 목표어(한국어)로 드러나도록 한 번역문을 기록하도록 그 기준을 마련하였다. 번역 내용이 확실하지 않거나 이해되지 않는 경우 “(())”으로 표기하고 추측되는 내용을 번역정비_코멘트_A/B 티어에 기술하였으며, [지시]가 나타나는 수어문의 번역 시, 직역에서는 ‘이것’, ‘저것’, ‘여기’, ‘저기’ ‘등으로 번역하나, 의역에서는 한국어 지시 어휘 다음에 ‘()’로 지시가 가리키는 구체적 내용을 표기하도록 하였다. 언어제공자 간의 고개 끄덕임, 고개 가로저음과 같은 제스처가 나타나는 경우, 제스처를 따로 코멘트 티어에 기술하지 않고 그 의미를 직역과 의역 번역에 반영한다. 이외에 문장의 높임 표현은 비격식체인 해요체(-어요, -해요)로 통일하고, ㅎㅎ, ㅋㅋ, ^^ 와 같은 인터넷 용어 혹은 채팅어를 사용하지 않으며, 수어이름(SN)의 경우 직역에서는 OOO(수어이름), 의역에서는 OOO으로 기재하도록 하였다.

3) 한국수어 말뭉치 공개 및 홍보 영상 제작

첫째, 홍보 동영상은 한국수어 말뭉치에 어려움 없이 접근할 수 있도록 말뭉치를 공개하는 국립국어원의 ‘모두의 말뭉치’ 회원가입 및 신청 방법 안내 수어 동영상, ‘모두의 말뭉치’ 내 한국수어 말뭉치 소개 자료 수어 동영상, 그리고 한국수어 말뭉치 수집에서 주석 작업을 하고 시스템에 탑재되기까지의 전 과정을 설명하는 홍보 동영상의 3가지 로 제작하였다.

둘째, 홍보 영상의 촬영은 32㎡ 이상 (5.3m * 6m) 면적의 촬영 공간에서 암막 커튼으로 빛을 차단하고, 배경막을 설치하여 수어가 선명하게 보이도록 하고, full-HD 사양의 카메라 2대와 조명 기기 2대, 프롬프터, 삼각대 1대, 노트북 1대로 이루어졌다.

셋째, 1편과 2편은 한국수어 말뭉치 접근을 위한 안내 동영상은 설명 위주로 제작하였으며 ‘모두의 말뭉치’에 접속하여 한국수어 말뭉치 창을 통해 한국수어 말뭉치 자료를 내려받는 내용으로 구성하였다. 2편은 한국수어 말뭉치의 구성 설명 동영상은 자료 유형, 관련 사업, 자료 설명, 영상 및 주석 파일 명명 규칙, 주석 및 번역 파일 (ELAN) 예시, 주석과 번역, 총렬 설명으로 구성되었다. 3편 한국수어 말뭉치 홍보 동영상은 한국수어 말뭉치의 성격과 정의, 한국수어 말뭉치 수집, 한국수어 말뭉치의 주석과 번역, 가공된 한국수어 말뭉치의 시스템 탑재에 이르기까지의 전 과정을 보여주고 있다.

2. 활용 방안 및 제언

1) 연구의 활용 및 기대 효과

본 연구는 한국수어 말뭉치 확장을 위해 2021년 신규 개발한 유도 과제 5종을 추가하여 총 18개의 유도 과제를 활용하여 미구축 지역인 충북 지역의 원시 말뭉치를 수집하였다. 이와 함께 기존에 수집된 한국수어 말뭉치를 공개하기 위해서 정비 절차를 진행하였다. 일련의 이런 연구의 결과는 다음과 같이 한국수어 발전을 위해 활용될 것이다.

첫째, 본 연구에서는 미구축 지역인 충북 지역 16쌍의 말뭉치를 수집하였다. 이는 충북 지역의 농인이 참여하는 원시 말뭉치의 추가 수집 편집 시간 기준 59시간 31분 29초를 통하여 지역 수어의 대표성을 갖추는 자료를 추가 확보하게 되었다는 점에서 그 의의가 크다고 할 수 있다. 수집에는 2021년 신규 개발된 유도 과제 5종을 포함한 총 18개의 유도 과제를 사용하여 다양한 한국수어 담화 양식을 수집하였다. 이로써, 서울·경기, 전북, 강원, 경남, 전남·광주, 대전·충남, 충북 7개(10개)의 지역 수어 사용자의 다양한 수어 형태를 확인할 수 있을 것으로 기대된다. 특히, 충북 지역은 기숙제 농학교가 있는 지역으로 농학교에서 사용하는 다양한 변이 형태가 존재할 것으로 예상되는데, 이에 대한 실증적인 자료를 제공함으로써 향후 수어 연구에 기여하게 될 것이다.

둘째, 본 연구에서는 기 구축된 한국수어 말뭉치 공개를 위한 정비 작업을 실시하였다. 수어 말뭉치 연구는 2015년부터 시작하여 해마다 원시 수어 말뭉치 수집과 더불어 주석 말뭉치 구축을 진행하였다. 2022년까지 구축된 원시 말뭉치는 약 395시간이며, 수지 글로스로 주석이 된 기본 주석 말뭉치는 1,438분(23시간), 수지 주석에 대한 의미

주석까지 수행된 자료는 445분(7시간)이다. 영상의 내용을 번역한 자료는 5,782분(96시간)이고 의미를 포함한 기본 주석과 번역 주석이 동시에 진행된 자료는 현재 157분으로 파악된 상태이다. 기 구축된 한국수어 말뭉치 중에서 공개를 위한 검토를 거쳐 최종 선정 후에 정비 작업을 실시하였다. 이번 한국수어 말뭉치 정비 영상은 273분 30초 분량의 원시 수어 말뭉치, 63분 54초 분량의 번역된 말뭉치, 그리고 95분 6초 분량의 주석과 번역된 말뭉치, 총 432분 30초 분량이다. 본 연구를 통해 정비되어 공개하는 한국수어 말뭉치는 향후 한국수어를 연구하고자 하는 전문가들에게 더 체계적이고 실증적인 연구를 위한 기초 자료와 더불어 한국수어 교육이나 교재 제작을 위해 유의한 정보를 제공할 것이다. 또한 기본 주석과 번역이 정비되어서 함께 제공되는 한국수어 말뭉치는 수어에 익숙하지 않은 언어학자들에게도 수어 연구를 위한 기초 자료를 제공함으로써 수어 연구의 지평을 넓히는 데 기여할 것이다.

셋째, 본 연구에서는 한국수어 말뭉치 공개 및 홍보 영상을 제작하였다. 이 영상들은 한국수어 연구나 교육에 관심이 있는 사람들뿐 아니라 한국수어를 제1 언어로 사용하고 있는 농민들에게 한국수어 말뭉치를 소개하고, 향후 한국수어 말뭉치에 접근하고자 할 때 편의를 제공할 것이다.

2) 제언

본 연구에서는 충북 지역의 원시 말뭉치를 구축하고, 기존에 구축된 한국수어 말뭉치 공개를 위한 정비를 진행하였으며, 공개된 한국수어 말뭉치를 홍보하기 위해서 홍보 동영상 제작하였다.

시간의 흐름에 따라 기존 수어 세대가 점점 사라져 가고 있는 현실에서 가능한 이른 시간 내에 지역별 수어 말뭉치 수집을 완료하는 것은 매우

시급하다. 그리고 원시 말뭉치가 수어 연구의 기초 자료로 활용되기 위해서는 기본 주석과 번역 역시 시급하게 이루어져야 한다. 그러나 제한된 시간과 전문가의 수로 이 모든 작업을 동시에 진행하는 데에는 어려움이 있다. 따라서 짧은 사업 기간에 수어 말뭉치를 수집하고 기본 주석을 진행했던 과정에서 나타난 문제점을 중심으로 향후 한국수어 말뭉치의 효과적인 구축을 위해 다음의 것들을 제언하고자 한다.

첫째, 수집 지역이 복수인 경우, 각각의 농아인협회와 사업 계약을 체결할 필요가 있다. 2019년부터 한국수어 말뭉치 수집은 수집 해당 지역의 농아인 협회와 사업 계약을 체결하고 진행해 왔다. 이번 사업은 충북 지역을 수집하였고, 충북농아인협회와 사업 계약을 체결하고 진행하였다. 한국수어 말뭉치 수집을 보다 효율적이고 성공적으로 수행하기 위해서는 수집 지역의 농아인협회와 협업을 하는 것이 중요하다. 지역 수어 사용자의 다양한 수어 형태를 수집하기 위해서는 지역의 언어제공자를 모집하고 사회자를 섭외에 하는 것에서부터, 촬영 및 편집, 현장에서의 원활한 소통이 매우 중요하기 때문이다.

둘째, 한국수어 원시 말뭉치의 질 보장을 위한 적정 촬영 시간을 확보해야 한다. 이번 연구에서는 2021년 새로 개발한 5종의 유도 과제를 포함하여 총 18종의 유도 과제를 사용하여 원시 말뭉치를 수집하였다. 2020년까지 13종의 유도 과제를 사용한 원시 말뭉치 수집에서는 언어제공자 쌍별 촬영 시간이 약 3시간으로 충분한 휴식을 가지며 촬영을 진행할 수 있었다. 이번 연구에서는 언어제공자 쌍별 평균 약 4시간 15분 이상이 소요되었으며, 휴식 시간을 포함하면 약 5시간에서 6시간에 걸쳐 촬영이 이루어졌다. 짧은 사업 기간으로 하루에 2팀을 촬영하였으며, 촬영 준비 및 진행을 위해 하루에 12시간 이상을 소요해야 했다. 다행히 사회자는 오전과 오후로 나눠 진행하였으나 촬영 시간이 길어진 만큼 수집에 참여하는 사회자나 언어제공자의 피로도는 증가

하였다. 언어제공자의 경우 시간이 길어질수록 수집을 위한 대화의 몰입도가 떨어지는 등의 문제가 발생하기도 하였다. 따라서 효과적 원시 말뭉치 수집을 위해 유도 과제의 수를 10개 정도로 조정하는 등의 적정 촬영 시간을 위한 대안 마련이 필요하다.

셋째, 주석 전문 인력의 확보 및 전문 인력이 안정적으로 유지될 수 있는 장치가 필요하다.

이번 연구에서는 다년간 주석 경험이 있는 인력을 확보할 수 있었을 뿐만 아니라, 주석 및 타입 정비 기준을 구체화한 2019년 정비 제안 과정부터 참여하며 본 연구에 대한 중요성을 충분히 이해하는 주석자였기에 연구진과의 유기적인 협력을 통해 주석에서의 어려움을 덜 수 있었다.

그러나 향후 주석의 양이 많아지면 더 많은 주석자가 필요할 것이며, 양질의 주석을 위해 경험 많은 주석자를 확보하는 것이 중요하다. 단순히 양적 효율만 생각하며 주석 작업을 하게 되면 수어의 형태를 충실히 반영하여 주석을 하지 못하는 문제를 낳을 수 있다. 따라서 주석 작업 내용을 일관된 관점으로 진행할 수 있는 주석자 확보가 이루어진다는 것은 주석의 일관성과 신뢰성을 담보할 수 있다는 의미이며, 향후 자료가 공개되는 시점에서 별도의 재정비 과정을 최소화할 수 있을 것이다. 그러므로 주석 전문팀을 상시 운용하는 방안에 대한 고려가 필요하다. 반면, 매년 과업으로 제시되는 연구에서는 수어의 기본 형태 정보보다는 수어 연구를 위한 추가 정보(예, 문법 정보)를 기록하는 것과 함께 주석 전문 인력의 저변 확대를 위한 교육과 홍보에 진력할 필요가 있다.

넷째, 주석 작업의 효율성을 높일 수 있도록 「한국수어 자료 통합 지원 시스템」의 개선이 필요하다.

주석 작업은 「한국수어 자료 통합 지원 시스템」을 통해 이뤄지게 된

다. 수어 말뭉치 연구를 하는 동안 수어 자료 탐색, 탐색, 호출 등, 전 과정에서 동일한 문제에 대한 처리 불가 현상이 반복되거나 작업을 지연시키는 등 시스템의 불안정성으로 인해 여러 가지 문제점이 노출된 바 있다. 그러나 금년도 연구 사업에서는 이러한 문제가 비교적 해결된 상태로 진행되어 시간 지체 등이 다소 해결될 수 있었다.

다만 주석 작업을 하는 동안 실제적인 작업에 도움이 될 수 있는 기능 추가의 필요성이 제기되었다. 최우선으로 ‘토큰 관리’ 기능의 개선이었다. 주석 말뭉치 구축은 동일한 수어의 형태를 동일한 글로스로 입력하며 글로스명은 ‘타입’으로 관리되고 있다. 따라서 주석자가 기존에 등록된 ‘타입명’을 검색하거나 새로운 타입을 등록할 때, 타입명에 쉽게 접근할 수 있도록 가장 보편적인 한글 텍스트로 정하도록 하였다. 그럼에도 주석자마다 가지고 있는 개인적 어휘망이 다르므로 인해 동일한 형태가 다른 ‘타입명’으로 중복 등록되기도 하였다. 다년간 주석 경험이 있는 주석자들은 이러한 오류를 피하고자 ‘시스템’ 내 ‘토큰 관리’ 메뉴에서 해당 수어의 의미로 검색하며 동일한 수어가 있는지를 검색하였다. 이때, 해당 수어의 의미는 찾아볼 수 있으나, 주석자가 찾는 의미의 수어가 어떤 타입명을 가졌는지는 확인할 수 없기 때문에 주석자는 해당 의미로 주석된 파일을 주석 파일(엘란)을 내려받아 재검색하는 과정을 거쳐야 했다. 이러한 불편을 감소시키기 위해서는 ‘토큰 관리’ 메뉴에서 수어 의미에 해당하는 타입명이 함께 쌍으로 검색될 수 있는 기능이 필요하다.

두 번째로는, 주석 작업 파일 내에서 「시스템」에서 관리되고 있는 타입 검색 기능이 필요하다. 주석자는 ‘수어 형태별 분절-타입명 검색-타입명 기록’의 과정으로 거쳐 주석 파일에 기록한다. 이때, 타입명 기록에 앞서 시스템에 해당 수어의 글로스가 타입명으로 관리되고 있는지를 확인하는 과정은 반드시 필요하며 중요한 과정이다. 이를 위해 주석자는 주석 파일과 시스템을 동시에 확인하며 진행할 수 있어야 하는데, 주석 파일 내에서 「시스템」에서 관리되고 있는 타입을 바로 검색하

는 기능을 활용할 수 있다면 작업 효율성이 높일 수 있을 것이다.

아울러 첫 공개 시점에서는 공개 자료가 많지 않아 관리상의 문제가 크게 드러나지 않지만 앞으로 공개 파일들이 많아지면 공개 파일 관리 시스템이 필요하다. 현재 통합시스템은 공개 파일 관리를 할 수 없는 상황이다. 구체적으로 언제 어떤 파일이 공개됐으며, 어떤 주석이 있는지(영상만 있는지, 번역이 있는지, 주석이 있는지)가 확인이 가능해야 한다. 또한 향후 공개될 수어 말뭉치의 균형성을 담보하기 위해서는 언어제공자의 연령, 지역, 과제 등에 대한 통계기능이 보완되어야 할 것이다.

다섯째, 2019년도 정비 제안된 타입과 연계된 주석 파일의 정비가 필요하다.

2019년도 연구 사업에서 주석 파일 및 타입 정비를 한 바 있다. 특히, 타입 정비를 통해 타입명이 중복된 타입의 삭제(98건), 부적절한 타입명의 수정(60건), 타입으로 등록한 영상 수정(38건) 등을 진행하였으나, 합성어가 아닌 형태를 합성어 형태로 기록함으로써 재분절이 필요한 타입도 상당 부분 있음을 확인하였다. 이에 따라 다시 주석해야 할 파일 등에 대한 추가 정비가 필요한 상황이다. 향후 공개를 앞둔 상황에서 이러한 부분에 대한 정비를 통해 구축된 주석 파일의 질을 높여야 할 것이다.

여섯째, 정확하고 효율적인 번역을 위해 농인과 청인 쌍으로 된 번역진 구성이 필요하다.

이번 번역 정비를 하면서 청인이 번역한 초벌 번역에서 참나적으로 지나가는 수어를 놓치거나 수어를 오인하는 경우가 허다하게 발생할 수 있음을 확인하였다. 이에 따라 수어 문장의 오역 결과를 초래하여 번역의 신뢰성을 떨어뜨릴 가능성이 있다. 또한 의역 시, 자연스러우면서 정확한 한국어로의 번역을 위해 번역 훈련이 된 청인 번역사의 참여도 매우 중요

함을 확인할 수 있었다. 따라서 이번 번역 정비에서와 같이 경력이 오래된 매우 뛰어난 수어통역사라 할지라도 혼자서 번역하는 것에는 오류가 일어날 위험성이 있을 뿐만 아니라 청인 번역사가 한 것을 나중에 놓인 번역사가 재확인하는 것은 작업의 효율성을 떨어뜨리므로 애초부터 농인과 청인 쌍을 이루어 번역하는 것이 번역의 정확성을 담보하고 번역 작업의 효율성을 높일 수 있을 것이다.

일곱째, 한국수어 말뭉치 구축을 다년도 사업으로 전환해야 한다.

현재 매년 모집 공고를 내고 연구팀을 공모하여 본 사업을 진행해 오고 있어 사업을 수행할 기간이 짧고 전문 인력 확보 또한 어렵다. 이런 연구 환경에서는 한국수어 말뭉치 구축 연구를 중장기 계획을 갖고 수행하기 어렵다. 다년 사업으로 연구 위탁이 이루어질 때, 위탁 받은 연구진은 중장기 연구계획을 수립하고 의미 있는 연구 결과물을 도출해 낼 수 있을 것이다.

여덟째, 한국수어 말뭉치 기반의 연구 과업 수행이 필요하다.

한국수어 말뭉치는 구축만을 목적으로 하는 것이 아니며, 말뭉치 기반의 다양한 연구가 이루어짐으로써 그 가치를 높이 인정받을 수 있을 것이다. 현실적으로 수어 연구 인력의 저변 확대를 위한 시간이 필요한 만큼 원시 수어 말뭉치 구축 및 분석 연구 사업을 통해 다양한 연구 사례를 제시할 필요가 있다.

아홉째, 공개 파일을 효율적으로 관리할 수 있도록 한국수어 자료 통합 지원 시스템을 개선하거나, 새로운 관리 시스템의 도입이 필요하다. 현재 한국수어 자료 통합 지원 시스템은 공개된 파일과 아직 공개되지 않은 파일을 구분하지 않는다. 내년 이후 계속해서 공개될 파일들을 연령, 지역 과제별로 균등하게 선정하기 위해서는 공개 파일을 따로 관리할 수 있는 별도의 기능이 필요하다. 또한 파일이 공개된 시

기와 각 파일별로 포함된 주석 정보, 즉 영상만 있는 파일인지, 번역 또는 의미 주석이 있는지 등을 확인하기 위해서는 공개 파일 관리 기능 및 검색이 필요하다. 공개 말뭉치의 균형성을 보장하기 위해 언어 제공자 연령, 지역, 과제, 포함된 주석 정보 등 항목별 파일 개수와 시간을 파악할 수 있는 통계 및 검색 기능이 반드시 필요하다.

참고문헌

<국내외 연구>

- 국립국어원(2014). 한국수어 코퍼스 구축을 위한 기초 연구. 서울: 국립국어원.
- 국립국어원(2015). 한국수어 말뭉치 기반 조사 및 시범 구축. 서울: 국립국어원.
- 국립국어원(2016). 한국수어 자료 통합 지원 시스템 구축 완료 보고서. 서울: 국립국어원.
- 국립국어원(2017). 한국수어 자료 통합 지원 시스템 고도화. 서울: 국립국어원.
- 국립국어원(2017). 한국수어 말뭉치 연구 및 구축. 서울: 국립국어원.
- 국립국어원(2018). 2018 한국수어 말뭉치 연구 및 구축. 서울: 국립국어원.
- 국립국어원(2019). 2019 한국수어 말뭉치 연구 및 구축. 서울: 국립국어원.
- 국립국어원(2020). 2020 한국수어 말뭉치 연구 및 구축. 서울: 국립국어원.
- 국립국어원(2021). 2021 한국수어 말뭉치 연구 및 구축. 서울: 국립국어원.
- 홍성은(2010). 수화언어학의 유도자료와 그 사용. 제12회 한국수화학회 학술대회 자료집, 9-17.
- 홍성은, 니시오 리에(2010). 독일수어 코퍼스 프로젝트에서의 유도 방법. 제12회 한국수화학회 학술대회 자료집, 50-58.
- Crasborn, Onno & Zwitserlood, Inge (2008). The Corpus NGT: an online corpus for professionals and laymen. In Crasborn, Onno et al. (Eds.): Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation

- (LREC 2008), Marrakesch, Marokko, 1.6.2008. Paris, Frankreich: European Language Resources Association (ELRA), 44-49.
- de Vos, C., Torreira, F., and Levinson, S. C. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Front. Psychol.* 6: 268.
- Nishio, Rie / Hong, Sung-Eun / König, Susanne / Konrad, Reiner / Langer, Gabriele / Hanke, Thomas / Rathmann, Christian (2010). Elicitation methods in the DGS (German Sign Language) Corpus Project. In Dreuw, Philippe et al. (Eds.): Proceedings of the LREC 2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta, 22.-23.5.2010. Paris, Frankreich: European Language Resources Association (ELRA), 178-185.
- Schembri, Adam / Fenlon, Jordan / Rentelis, Ramas / Reynolds, Sally / Cormier, Kearsy (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation* 7, 136-154.

<온라인 자료>

한국수어사전 [https://sldict.korean.go.kr/front/main/main.do#\(2023.12.8\)](https://sldict.korean.go.kr/front/main/main.do#(2023.12.8))
 ELAN 다운로드 경로 ‘막스프랑크심리학연구소의 언어 기록 보관소’
[https://archive.mpi.nl/tla/elan/download\(2023.12.8\)](https://archive.mpi.nl/tla/elan/download(2023.12.8))

[부록 1] 국가 언어 자원(말뭉치) 구축 및 활용 저작권 등 이용 허락계약서

국가 언어 자원(말뭉치) 구축 및 활용 저작권 등 이용 허락 계약서

권리자 _____(이하 “권리자”이라 함)와 이용자 국립국어원(이하 “이용자”이라 함)은 아래 저작물에 관한 저작권과 권리자의 초상권, 기타 이에 부수한 권리 등(이하 ‘계약 대상 권리’라고만 한다) 일체의 권리에 대한 이용 허락과 관련하여 다음과 같이 계약을 체결한다.

다 음

제1조 (계약의 목적)

본 계약은 계약 대상 권리의 원만한 이용 허락과 관련하여 권리자와 이용자 사이의 권리관계를 정하는 것을 목적으로 한다.

제2조 (계약의 대상)

① 본 계약의 대상이 되는 저작권산권은 다음 각 호에서 정하는 바와 같다.

1. 대상저작물은 한국수어 말뭉치로 한다(이하 ‘대상저작물’이라고만 한다)
2. 이용 대상 권리는 다음 각 목에서 정하는 일에 필요한 일체의 저작권산권과 저작인접권 중 재산권 부분을 말한다.
 - 가. 이용자 및 이용자가 발주한 용역 사업의 수행자가 대상저작물을 일정한 형식으로 전자적 기록 매체에 담아 보존하는 일
 - 나. 이용자 및 이용자가 발주한 용역 사업의 수행자가 한국수어 연구와 언어 정보 처리 분야에 응용하기 위해 대상저작물을 복제·변형(목차·머리말·도표·그림·각주 등의 편집 및 삭제, 언어 단위별 분리, 언어적·비언어적 정보 부착 등)하는 일
 - 다. 이용자 및 이용자가 발주한 용역 사업의 수행자가 대상저작물 및 그 복제·변형물을 연구 및 기술 개발용으로 학계·연구기관·산업체 등이 이용할 수 있도록 제공·배포하는 일
 - 라. 대상저작물 및 그 복제·변형물을 제공·배포받은 학계·연구기관·산업체 등이 한국수어 연구와 언어 정보 처리 분야 응용을 위하여 대상저작물 및 그 복제·변형물을 분석 및 처리하여 사용하는 것을 허락하는 일

② 본 계약의 대상이 되는 초상권은 다음 각 호에서 정하는 바와 같다.

- 가. 초상권의 대상은 제①항 제1호에 기재된 저작물에 포함된 권리자의 초상 전체로 한다.
- 나. 이용 대상 권리는 제①항 제2호 가목과 나목에 기재된 일체의 행위에 필요한 초

상권 및 퍼블리시터권으로 한다.

- ③ 기타 본 계약의 원활한 이행을 위하여 제1항 또는 제2항에서 정한 권리의 행사를 위하여 권리자는 이용자 또는 이용자가 허락한 자에 대하여 저작권(저작권법 제17조 제1항에 의한 인격권 부분을 포함한다) 및 기타 법적 권리를 부당하게 행사하지 아니하기로 한다.

제3조 (이용 허락 기간)

- ① 계약 대상 권리에 대한 이용 허락 기간은 계약체결일부터 2052년 12월 31일까지로 한다.
- ② 권리자가 이용 허락 중지 의사를 밝히지 아니하면 이용 허락이 5년 단위로 자동 갱신된다.
- ③ 계약기간 중 권리자의 사망이 발생하면 특별한 사정이 없는 한 그의 적법한 상속인이 이 계약의 지위를 상속하는 것으로 한다.

제4조 (권리자의 의무)

- ① 권리자는 이용자에게 계약 대상 권리에 관하여 본 계약서 제2조에서 정한 계약 대상 권리를 제3조의 기간 동안 비독점적으로 허락한다.
- ② 권리자는 이용자에게 계약 체결일로부터 10일 이내에 계약 대상 권리의 이용을 위해 필요한 상당한 자료를 인도하여야 한다. 다만, 제2조 제①항 제1호의 저작물이 한국저작권위원회에 등록되어 있지 않은 경우 이용자가 요청하면 권리자는 이 저작물의 저작재산권을 등록한 후 위 의무를 이행한다.
- ③ 권리자는 계약 대상 권리에 제3자의 이용 허락권, 질권 등이 존재하는 경우, 이용자에게 그 사실을 사전에 알려야 한다.
- ④ 권리자는 계약 대상 권리의 저작재산권 전부 또는 일부를 제3자에게 양도하거나 이에 대하여 질권을 설정하고자 하는 경우, 사전에 이용자에게 이 사실을 통보하여야 한다.
- ⑤ 제4항의 경우 권리자는 계약기간 중 계약 대상 권리 중 일부 또는 전부를 제3자에게 양도할 경우, 이에 대하여 사전에 이용자에게 통지하여야 하여 제3자와 이용자가 계약을 체결할 수 있도록 적극 협조하여야 한다.

제5조 (이용자의 권리 및 의무)

- ① 이용자는 계약 대상 권리를 제3조의 이용 허락 기간 동안 제2조의 이용 허락을 받은 범위 내에서 비독점적으로 자유롭게 이용할 수 있다.
- ② 계약 대상 권리의 이용료는 무상으로 한다.
- ③ 이용자는 관례적으로 저작자 및 저작재산권자의 성명 등 표시를 허용하는 대상 저작물을 이용하는 경우, 그 저작자 및 저작재산권자의 성명 등을 표시하여야 한다.
- ④ 이용자는 대상저작물을 이용함에 있어서 저작권(저작권법 제17조 제1항에 의한 인격권)을 침해하지 아니한다. 다만, 대상저작물의 본질적인 내용을 변경하지 않는 범위 내에서 권리자에게 그 사실을 사전에 고지한 후 사소한 수정 및 편집을 할 수 있다.

제6조 (확인 및 보증)

① 권리자는 이용자에게 다음 각 호의 사항을 확인하고 보증한다.

1. 대상저작물의 저작권 이용 허락을 체결하는 데 필요한 권리 및 권한을 적법하게 보유하고 있다는 것
2. 대상저작물의 내용이 제3자의 저작권, 상표권, 인격권을 비롯한 일체의 권리를 침해하지 아니한다는 것
3. 대상저작물에 대하여 이용자에게 사전에 알린 제3자의 권리 외에는 이용자의 이용을 제한할 수 있는 부담이 더 이상 존재하지 아니한다는 것

② 이용자는 권리자에게 다음 각 호의 사항을 확인하고 보증한다.

1. 대상저작물에 적용된 이용 허락 조건에 의해서만 대상저작물의 재이용을 허락할 것
2. 대상저작물을 권리자 및 제3자의 인격적 권리를 침해하는 방식으로 이용하지 아니할 것

제7조 (계약내용의 변경)

본 계약 내용 중 일부를 변경할 필요가 있는 경우에는 권리자와 이용자의 서면합의에 의하여 변경할 수 있으며, 그 서면합의에서 달리 정함이 없는 한, 변경된 사항은 그 다음날부터 효력을 가진다.

제8조 (계약의 해지)

- ① 당사자는 천재지변 또는 기타 불가항력으로 계약을 유지할 수 없는 경우에 본 계약을 해지할 수 있다.
- ② 당사자는 상대방이 정당한 이유 없이 본 계약을 위반하는 경우에 4주 이상의 상당한 기간을 정하여 상대방에게 그 시정을 최고하고, 상대방이 그 기간이 지나도록 이행하지 아니하는 경우에는 계약을 해지할 수 있다. 다만, 상대방이 명백한 시정거부의사를 표시하였거나 위반 사항의 성격상 시정이 불가능하다는 것이 명백히 인정되는 경우에는 위와 같은 최고 없이 계약을 해지할 수 있다.
- ③ 본 계약에 대한 해지권의 행사는 상대방에 대한 손해배상청구권 행사에 영향을 미치지 아니한다.

제9조 (손해배상)

당사자가 정당한 이유 없이 본 계약을 위반하는 경우, 그로 인하여 상대방에게 발생한 모든 손해를 배상할 책임이 있다. 다만, 제8조 1항의 사유로 본 계약을 이행하지 못한 경우에는 손해배상책임을 면한다.

제10조 (비용의 부담)

계약 체결에 따른 비용은 이용자가 전부 부담한다.

제11조 (분쟁해결)

- ① 본 계약에서 발생하는 모든 분쟁은 권리자와 이용자가 상호 원만한 합의에 이르도록 노력하여야 하며, 분쟁이 원만히 해결되지 않는 경우에는 소제기에 앞서 한국

저작권위원회에 조정을 신청할 수 있다.

- ② 제1항에 따라 해결되지 아니할 때에는 서울남부지방법원을 제1심 소송의 전속관할로 정하여 소송에 의해 해결토록 한다.

제12조 (비밀유지)

양 당사자는 본 계약의 체결 및 이행과정에서 알게 된 상대방에 관한 정보, 본 계약의 내용 및 대상저작물의 내용을, 상대방의 서면에 의한 승낙 없이 제3자에게 공개하여서는 아니 된다.

제13조 (기타부속합의)

- ① 권리자와 이용자는 본 계약의 내용을 보충하거나, 이 계약에서 정하지 아니한 사항을 규정하기 위하여 부속합의서를 작성할 수 있다.
- ② 제1항에 따른 부속 합의는 본 계약의 내용과 배치되거나 위반하지 않는 범위 내에서 유효하다.

제14조 (계약의 해석 및 보완)

본 계약서에서 명시되어 있지 아니하거나 해석상 이견이 있을 경우에는 저작권법, 민법 등을 준용하고 사회 통념과 조리에 맞게 해결한다.

제15조 (계약 효력 발생일)

본 계약의 효력은 계약 체결일로부터 발생한다.

2023년 월 일

권리자

이용자

성명:

(인)

성명: 국립국어원장

(인)

생년월일:

주소: 서울특별시 강서구 금남화로 154

주소:

[부록 2] 한국수어 말뭉치 사업 언어제공자 개인정보
수집·이용 및 제공 동의서

한국수어 말동치 사업 언어제공자 개인정보 수집·이용 및 제공 동의서

국립국어원 2023 한국수어말동치 연구 및 분석팀 귀중

한국수어말동치 연구 및 분석팀은 아래와 같이 귀하의 개인정보를 수집하여 이용하고 제3자에게 제공하고자 합니다. 내용을 자세히 읽으신 후 동의 여부를 결정하여 주십시오.

□ 개인정보 수집·이용 내역 (필수)

수집·이용 항목	수집·이용 목적	보유기간
(필수항목) 성명, 생년월일, 주소, 전자우편, 전화번호, 성별, 살았던 장소	한국수어 말동치 연구 및 분석	<u>동의일로부터 30년,</u> <u>동의 철회 시 즉시 삭제</u>

※ 위의 개인정보 수집·이용에 대한 동의를 거부할 권리가 있습니다.

다만 동의하지 않으시는 경우 말동치 구축을 위한 자료 제공자에서 배제됩니다.

※ 보유기간: 본 동의서의 효력은 본 계약이 갱신 또는 변경되는 경우에도 유효하며, 위 개인 정보는 수집·이용에 관한 동의일로부터 동의 철회 시 또는 제공된 목적을 달성할 때까지 보유·이용됩니다. 동의 철회는 개인정보 보호법 제36조(개인정보의 정정·삭제)에 따릅니다.

☞ 위와 같이 개인정보를 수집·이용하는 데 동의하십니까?

동의	<input type="checkbox"/>	미동의	<input type="checkbox"/>
----	--------------------------	-----	--------------------------

년 월 일
본인 성명 (서명 또는 인)

□ 민감정보 처리 내역 (필수)

수집·이용 항목	수집·이용 목적	보유기간
(필수항목) 본인청력 가족정보 (직계 존비속 및 친척 포함) - 가족 및 친척의 농인 여부 - 가족 및 친척과의 의사소통 방법 본인의 주된 의사소통 방법 한국수어 습득 시기 한국수어 습득 장소 한글지문자 사용 유무 지역수어 사용 농(난청)인과의 정기적인 만남 여부 본인의 한국어 능력 정도 본인의 외국어 사용 여부 본인의 학교 및 직업 정보(취미 포함)	한국수어 말동치 연구 및 분석	<u>동의일로부터 30년</u> <u>동의 철회 시 즉시 삭제</u>

※ 위의 민감정보 수집·이용에 대한 동의를 거부할 권리가 있습니다.

다만 동의하지 않으시는 경우 말동치 구축을 위한 자료 제공자에서 배제됩니다.

※ 보유기간: 본 동의서의 효력은 본 계약이 갱신 또는 변경되는 경우에도 유효하며, 위 개인 정보는 수집·이용에 관한 동의일로부터 동의 철회 시 또는 제공된 목적을 달성할 때까지 보유·이용됩니다. 동의 철회는 개인정보 보호법 제36조(개인정보의 정정·삭제)에 따릅니다.

☞ 위와 같이 민감정보를 수집·이용하는 데 동의하십니까?

동의	<input type="checkbox"/>	미동의	<input type="checkbox"/>
----	--------------------------	-----	--------------------------

년 월 일
본인 성명 (서명 또는 인)

□ 개인정보 제3자 제공 내역 (필수)

제공받는 기관	제공목적	제공하는 항목	보유기간
OO연구소, OO대학교	한국수어 말뭉치 연구 및 분석	(필수항목) 성명, 생년월일, 주소, 전자우편, 전화번호, 성별, 살았던 장소 본인경력, 가족정보 (직계 존비속 및 친척 포함) - 가족 및 친척의 농인 여부 - 가족 및 친척과의 의사소통 방법 본인의 주된 의사소통 방법 한국수어 습득 시기 한국수어 습득 장소 한글지문자 사용유무 지역수어 사용 농(난청)인과의 정기적인 만남여부 본인의 한국어 능력 정도 본인의 외국어 사용 여부 본인의 학교 및 직업 정보(취미포함)	<u>동의일로부터 30년</u> 동의 철회 시 즉시 삭제 제공받은 기관 동시 삭제

※ 위의 개인정보 제3자 제공에 대한 동의를 거부할 권리가 있습니다.

다만 동의하지 않으시는 경우 말뭉치 구축을 위한 자료 제공자에서 배제됩니다.

※ 보유기간: 본 동의서의 효력은 본 계약이 갱신 또는 변경되는 경우에도 유효하며, 위 개인 정보는 수집·이용에 관한 동의일로부터 동의 철회 시 또는 제공된 목적을 달성할 때까지 보유·이용됩니다. 동의 철회는 개인정보 보호법 제36조(개인정보의 정정·삭제)에 따릅니다.

☞ 위와 같이 개인 민감정보를 제공하는 데 동의하십니까?

동의

☐

미동의

☐

년 월 일
본인 성명

(서명 또는 인)

□ 개인정보 취급업무 위탁 내역

위탁 받는 자(수탁업체)	업무내용
	한국수어 말뭉치 시스템 유지보수

[부록 3] 언어제공자 질문지

자료 수집을 위한 질문지

개인정보

1. 이름: _____
2. 생년월일: _____
3. 주소: _____
4. 성별: ☐ 여성 ☐ 남성
5. 현재 사는 곳: _____ 거주 기간(예: 08-현재): _____
6. 현재까지 살았던 장소들
_____년부터 _____년까지 거주지 _____ (지역명)
_____년부터 _____년까지 거주지 _____ (지역명)
_____년부터 _____년까지 거주지 _____ (지역명)
_____년부터 _____년까지 거주지 _____ (지역명)
7. 청력상태/청력손실 시기

☐ 농 (☐ 태어날 때부터 ☐ ____살 이후부터)
☐ 난청 (☐ 태어날 때부터 ☐ ____살 이후부터)
8. 청각보조기기 착용 여부

☐ 인공와우 시술을 받았다. (____년 이후부터 사용)
☐ 규칙적으로 보청기를 사용한다.
☐ 가끔 보청기를 사용한다.
☐ 보청기 등의 보장구를 사용하지 않는다.

가족 정보

9. 아버지의 농인 여부: ☐ 농인 ☐ 난청인 ☐ 청인 ☐ 잘 모르겠다

10. 어머니의 농인 여부: ☐ 농인 ☐ 난청인 ☐ 청인 ☐ 잘 모르겠다

11. 부모와의 동거 여부(어린시절): ☐ 예 ☐ 아니오

12. 부모 외의 양육자: [11번 질문에 '아니오'라고答한 경우]

☐ 할아버지 ☐ 할머니 ☐ 친척 ☐ 기타 _____

이들 양육자 중 남성의 농인 여부: ☐ 농인 ☐ 난청인 ☐ 청인

이들 양육자 중 여성의 농인 여부: ☐ 농인 ☐ 난청인 ☐ 청인

13. 어머니와의 주된 의사소통 방법

☐ 한국수어 ☐ 한국어대응수어+말 ☐ 제스처 ☐ 말 ☐ 필담
☐ 홈싸인 ☐ 기타 _____

14. 아버지와의 주된 의사소통 방법

☐ 한국수어 ☐ 한국어대응수어+말 ☐ 제스처 ☐ 말 ☐ 필담
☐ 홈싸인 ☐ 기타 _____

15. 형제 및 자매의 농인 여부/나이

- ① ☐ 농인 ☐ 난청인 ☐ 청인 (나보다 나이가 ____살 ☐ 많다 ☐ 적다)
② ☐ 농인 ☐ 난청인 ☐ 청인 (나보다 나이가 ____살 ☐ 많다 ☐ 적다)
③ ☐ 농인 ☐ 난청인 ☐ 청인 (나보다 나이가 ____살 ☐ 많다 ☐ 적다)
④ ☐ 농인 ☐ 난청인 ☐ 청인 (나보다 나이가 ____살 ☐ 많다 ☐ 적다)
⑤ ☐ 형제 및 자매 없음

16. 형제 및 자매들과의 주된 의사소통 방법

☐ 한국수어 ☐ 한국어대응수어+말 ☐ 제스처 ☐ 말 ☐ 필담
☐ 홈싸인 ☐ 기타 _____

17. 가족 및 친척의 농인 여부

할머니 ☐ 농인 ☐ 난청인 ☐ 청인
할아버지 ☐ 농인 ☐ 난청인 ☐ 청인
_____ ☐ 농인 ☐ 난청인 ☐ 청인
_____ ☐ 농인 ☐ 난청인 ☐ 청인

18. 배우자의 농인 여부

☐ 농인 ☐ 난청인 ☐ 청인 ☐ 코다(CODA, 농부모 청인자녀) ☐ 배우자 없음

19. 배우자와의 주된 의사소통 방법

☐ 한국수어 ☐ 한국어대응수어+말 ☐ 제스처 ☐ 말 ☐ 필담
☐ 홈싸인 ☐ 기타 _____

20. 자녀의 농인 여부

☐ 농인 (_____ 명)
☐ 난청인 (_____ 명)
☐ 청인 (_____ 명)
☐ 자녀 없음

21. 자녀와의 주된 의사소통 방법

☐ 한국수어 ☐ 한국어대응수어+말 ☐ 제스처 ☐ 말 ☐ 필담
☐ 홈싸인 ☐ 기타 _____

언어

22. 언어제공자(본인)의 주된 의사소통 방법

☐ 한국수어 ☐ 한국어대응수어+말(음성언어) ☐ 말(음성언어)

23. 한국수어 습득 시기: _____살 때부터

24. 한국수어 습득 장소

- ☐ 가정/부모님
☐ 농학교
☐ 농인 교회
☐ 농아인협회
☐ 기타 _____

25. 한글지문자(지화) 사용 유무

- ☐ 전혀 사용하지 않음 ☐ 거의 사용하지 않음 ☐ 가끔 사용
☐ 자주 사용

26. 지역수어 사용: 지역명 _____

27. 농(난청)인과의 정기적인 만남 여부

- ☐ 일주일에 1~2번 ☐ 일주일에 3~4번 ☐ 일주일에 5번 이상

28. 본인의 한국어 능력 정도

- 읽기 ☐ 전혀 못한다 ☐ 조금 한다 ☐ 보통 ☐ 잘 한다 ☐ 아주 잘
한다
쓰기 ☐ 전혀 못한다 ☐ 조금 한다 ☐ 보통 ☐ 잘 한다 ☐ 아주 잘
한다

29. 외국수어 사용 여부

- 미국수어 ☐ 전혀 못한다 ☐ 조금 한다 ☐ 보통 ☐ 잘 한다 ☐ 아주
잘 한다
일본수어 ☐ 전혀 못한다 ☐ 조금 한다 ☐ 보통 ☐ 잘 한다 ☐ 아주
잘 한다
국제수화 ☐ 전혀 못한다 ☐ 조금 한다 ☐ 보통 ☐ 잘 한다 ☐ 아주
잘 한다
_____ ☐ 전혀 못한다 ☐ 조금 한다 ☐ 보통 ☐ 잘 한다 ☐ 아주
잘 한다

학교 및 직업

30. 출신 학교

출신학교1: _____살부터 _____살까지 _____학교 _____지역

학교에서의 주된 의사소통 방법:

- ☐ 한국수어 ☐ 한국어대응수어+말 ☐ 한국수어 + 한국어 ☐ 구화
☐ 수어통역사와 함께 ☐ 기타 _____

출신학교1: _____살부터 _____살까지 _____학교 _____지역

학교에서의 주된 의사소통 방법:

- ☐ 한국수어 ☐ 한국어대응수어+말 ☐ 한국수어 + 한국어 ☐ 구화
☐ 수어통역사와 함께 ☐ 기타 _____

출신학교1: _____살부터 _____살까지 _____학교 _____지역

학교에서의 주된 의사소통 방법:

- ☐ 한국수어 ☐ 한국어대응수어+말 ☐ 한국수어 + 한국어 ☐ 구화
☐ 수어통역사와 함께 ☐ 기타 _____

출신학교1: _____살부터 _____살까지 _____학교 _____지역

학교에서의 주된 의사소통 방법:

- ☐ 한국수어 ☐ 한국어대응수어+말 ☐ 한국수어 + 한국어 ☐ 구화
☐ 수어통역사와 함께 ☐ 기타 _____

출신학교1: _____살부터 _____살까지 _____학교 _____지역

학교에서의 주된 의사소통 방법:

- ☐ 한국수어 ☐ 한국어대응수어+말 ☐ 한국수어 + 한국어 ☐ 구화
☐ 수어통역사와 함께 ☐ 기타 _____

31. 학교 기숙사 입사 여부: ☐ 예 ☐ 아니오

32. 학교 기숙사 입사에 대한 추가정보 [31번에서 '예'라고 답한 경우]

- ☐ 기숙사에서 생활하고, 주말에 집에 규칙적으로 갔다
☐ 기숙사에서 생활하고, 방학 때만 집에 갔다

33. 최종 학력

- ☐ 초등학교 졸업 ☐ 중학교 졸업
☐ 고등학교 졸업 ☐ 전공과 졸업
☐ 전문대학 졸업 ☐ 4년제 대학교 졸업
☐ 대학원 석사 졸업 ☐ 대학원 박사 졸업
☐ 기타 _____

34. 대학명/전공

대학명 _____ 전공 _____ 지역 _____

35. 직업훈련 기관명/지역

기관명 _____ 지역 _____

36. 재직 여부: ☐ 예 ☐ 아니오

37. 직장 정보 (현재 재직 중인 직장 또는 마지막 직장)

직종(직장) _____ 지역 _____
근무 기간 _____년부터 _____년까지

직장에서의 주된 의사소통 방법:

- ☐ 한국수어 ☐ 한국어대응수어+말 ☐ 한국수어 + 한국어 ☐ 구화
☐ 수어통역사와 함께 ☐ 기타 _____

38. 직장에서 작업 형태

- ☐ 팀에서 일함 ☐ 혼자 일함

39. 취미와 특별히 흥미가 있는 영역(예, 축구, 자동차, 요리, 독서, 등산 등)

농 사회에서의 활동

40. 한국수어 강사 경험

현재 한국수어를 가르치고 있다 ☐ 예 ☐ 아니오
과거에 한국수어를 가르쳤다 ☐ 예 ☐ 아니오

41. 한국수어 강사 활동 정도 [40번에서 '예'라고 답한 경우]

☐ 1년 이하 ☐ 1~2년 ☐ 3~4년 ☐ 5년 이상

42. 수어 예술분야 활동에서 수어사용 여부(예, 수어 예술분야: 연극, 시, 퍼포먼스 등)

☐ 수어를 사용한다. ☐ 수어를 사용하지 않는다.

43. 우세손

	오른손사용	왼손사용	가끔 왼손 사용
글쓰기	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
양치질	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
칼질하기	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
수어할 때	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

44. 마지막으로 하고 싶은 말

날짜: 2022년 ____월 ____일

서명: _____

[부록 4] 언어제공자와 사회자 관계에 관한 질문지

언어제공자와 사회자 관계에 관한 질문지

자료수집 날짜:

사회자:

언어제공자 A:

언어제공자 B:

언어제공자 A와 언어제공자 B

- 1) ● 우리는 서로 매우 잘 안다
 ● 우리는 서로 잘 안다
 ● 우리는 서로 알지만, 자주 만나지 않는다
 ● 우리는 과거에 만난 적이 있다
 ● 우리는 오늘 처음 만났다
- 2) 우리는 서로 친척 관계이거나 결혼을 통해 친척 관계이다 ☐ 예 ☐ 아니오
 ('예'라고 답한 경우) 어떤 친척 관계인지(예, 조카 등)

사회자와 언어제공자 A

- 1) ● 우리는 서로 매우 잘 안다
 ● 우리는 서로 잘 안다
 ● 우리는 서로 알지만, 자주 만나지 않는다
 ● 우리는 과거에 만난 적이 있다
 ● 우리는 오늘 처음 만났다
- 2) 우리는 서로 친척 관계이거나 결혼을 통해 친척 관계이다 ☐ 예 ☐ 아니오
 ('예'라고 답한 경우) 어떤 친척 관계인지(예, 조카 등)

사회자와 언어제공자 B

- 1) ● 우리는 서로 매우 잘 안다
 ● 우리는 서로 잘 안다
 ● 우리는 서로 알지만, 자주 만나지 않는다
 ● 우리는 과거에 만난 적이 있다
 ● 우리는 오늘 처음 만났다
- 2) 우리는 서로 친척 관계이거나 결혼을 통해 친척 관계이다 ☐ 예 ☐ 아니오
 ('예'라고 답한 경우) 어떤 친척 관계인지(예, 조카 등)

[부록 5] 원시 말뭉치 수집 점검표

원시 말뭉치 수집 점검표

1. 촬영 전 사전 준비

구분	세부내용	체크 (O, X)	참고 파일
지역협회	1. 사회자 모집 및 추천		
	- 언어제공자 조건 부합하는 지역의 농인 중 담화 촉진 능력이 탁월한 자 - 말뭉치 수집 인원수에 따라 1~2명		
	2. 언어제공자 모집 및 추천		
	- 언어제공자 조건에 부합하는 지역의 농인 - 촬영 인원보다 많은 명단 확보(예비인원 포함) ※ 지역별, 연령별 수집 현황을 연구진에서 제공하며 이를 참고하여 지역의 특성을 고려하여 모집함(단, 고령 언어제공자 확보 필요).		1-1. (서식)
	3. 언어제공자 확정 및 촬영 스케줄 확인		
	- 연구진에 언어제공자 목록 제출(검토 및 논의)		
	- 언어제공자 확정 및 언어제공자 매칭 • 언어제공자 매칭 시 나이대 고려(자연스러운 담화 환경 제공 목적) • 언어제공자 매칭 시 성별로 구분되지 않아도 됨		3-1. 3-2.
	- 촬영 스케줄 작성(일일 촬영 인원 조율)		1-2. (서식)
	- 연구진에 촬영 스케줄 공유		
	4. 사회자 유도자료 설명 영상 제작		
	- 유도자료 삽입용으로 지역에서 선정된 사회자가 유도자료 설명하는 영상을 제작하여 연구진에 제출 ① 공동: 동의 영상 1개 ②-1 유도 과제 18종 지역: 18종 22개 영상 촬영 및 편집		

2. 촬영 준비 및 촬영 당일

※ 수집(촬영) 시 방역 유의사항을 준수하도록 한다.

※ 농인의 코드 스위칭 방지 목적을 위해 촬영종료 시까지 촬영실 내 청인의 출입을 자제하며 청인과의 대화를 금지하도록 한다.

구분	세부내용		체크 (O, X)	참고 파일
촬영	1. 촬영장 공간 확보			
	공간규격	• 촬영실 : 32㎡(5.3m×6m) 이상		2-1. 2-2. 2-3.
		• 모니터실		
		• 휴게실		
	장비	• 카메라 3대 이상		
		• 메모리 128G 이상		
		• 노트북(컴퓨터) 3대 이상		
		• 모니터 5대 이상		
	2. 촬영실 설치(연구진 확인 여부)			
	- 조명, 에어컨, 햇빛(암막) 등 시설 완비 여부			
	- 배경: 파란색			
	- 카메라 및 메모리카드: 촬영 사양 확인			
	- 촬영 높이 및 각도 점검			
	- 컴퓨터(본체, 모니터 연결 상태)점검			
	- 유도자료 정상 작동 및 유도자료용 색지 구비 여부			
	- 사회자용 시계 차임벨, 사회자 일지 준비 여부			
	3. 모니터실 설치			
	- 촬영장 카메라와 모니터실 모니터 간의 연결상태 점검			
	- 촬영 일지 준비			
	- 돌발 상황에 대처할 인력			
	4. 휴게실 설치			
	지침	• 방문 명부 작성		7-1. (서식)
	촬영 전	• 저작권 이용 허락 계약서 수령		3-3. (서식)
		• 개인정보 수집 동의서 작성		3-4. (서식)
		• 자료 수집을 위한 질문지 작성		3-5. (서식)
		• 보관용 파일 및 필기구 등		
		※ 촬영 전, 두 언어제공자가 서로 서먹하지 않도록 약간의 담화를 나눈 후, 촬영을 시작할 수 있도록 함.		

	촬영 후	• 사회자와의 관계 질문지 작성		3-6.
		• 언어제공자 수당 지급용 영수증 작성		3-7.
		• 언어제공자 수당 지급용 서류(신분증, 통장 사본)		
		• 언어제공자 촬영 확인 명부 작성		3-9
		• 보관용 파일 및 필기구 등		
	※ 답변 미비 사항 재확인 필수			
	5. 예비 촬영			
	- 본 촬영 시 문제 발생을 최소화할 목적으로 실시 - 촬영팀에서 자체적으로 조율			
	6. 본 촬영			
	- 사회자 및 언어제공자 복장 확인 <ul style="list-style-type: none"> • 단정한 옷차림 • 어두운색 또는 손이 잘 드러날 수 있는 옷차림 • 배경색과 동일한 파란색이나 체크 무늬의 옷은 피할 것 - 사회자 일지 작성			4-1. (서식) 4-2. 4-3.

- 유도 과제 18종 지역: 최소 4시간 30분이상 촬영
☞ 최소 편집시간 3시간 30분 확보되어야 함

※ 편집 시간은 편집 지침 참고

3. 편집

구분	세부내용	체크 (O, X)	참고 파일
연구진 사전 준비	1. 편집 교육 실시		
	- 연구진 주체 촬영 첫날 지침 설명		
	2. 언어제공자 쌍별 원본 영상		
	- 카메라별(3대) 영상 1쌍 기준: 3개(Rt, Md, Lt)		6-1.
	- 편집 영상 파일명 (지역별 상이)		6-2. (서식)
	예 • 충북: KSLCB5원본Rt, KSLCB5원본Md, KSLCB5원본Lt		6-3.
	※ 파란색 영어 는 지역 코드이며, 숫자는 첫 번째 언어제공자 쌍의 번호임		6-4. (서식)
	3. 원본 영상 활용 과제별 편집 영상		6-5.
	- 카메라별(3대) 동의 영상 1쌍 기준: 1개 총 3개		6-6.
	예 • 충남: KSLCB5동의서Rt, KSLCB5동의서Md, KSLCB5동의서Lt		6-7.

- 유도 과제 18종(A~R) 지역 1쌍 기준:

	카메라별(3대) 과제 18종 총 54개 ※ 초록색 영어 는 유도 과제 번호임		
	예 • 강원: KSLGW5ARt, KSLGW5AMd, KSLGW5Alt		
	4. 편집일지 작성		
	- 편집 시 발생한 특이사항에 대해 기록		
	5. 편집 영상 및 편집 일지 1차 제출		
	- 외장하드에 담아 우편으로 교류		
	6. 연구진 검수		
	- 편집 영상 및 편집 일지 토대로 영상 검토		
	- 수정 부분 표시 후 수정 요청		
	- 검수 및 수정 과정 반복		
	7. 최종 납품		
	<ul style="list-style-type: none"> • 산출물 <ul style="list-style-type: none"> - 사회자 설명 동영상 19종(19개) ☞ 사회자 교육 실시 후 즉시 제출 <ul style="list-style-type: none"> - 언어자료 촬영(카메라 Lt, Mt, Rt) 원본 영상(48개) - 언어자료 편집 영상(864개) - 언어제공자 동의 영상(32개) • 해상도 : 1024*768, 언어제공자 촬영분 편집 영상 1개당 500 M • 규격 : 촬영 원본 - AVI 혹은 MOV 편집 영상 - MPEG4 • 제출 매체 및 수량 : 저장 매체 (외장하드 등) 2개 		

[부록 6] 사회자 일지(서식)

사회자 일지	
일 시	
장 소	
참석자	
내용 및 결과	

과제명	소요 시간	시작 시간	종료 시간	기타사항
전체	273"			
1. 촬영 동의서	3"			
2. 수어 이름 말하기(A)	5"			
3. 다시 이야기하기(벡타이(D)	10"			
4. 표지판(F)	20"			
5. 흥부와 놀부(K)	10"			
6. 가장 기억에 남는 날(N)	20"			
7. 농학교 경험 이야기하기(M)	15"			

8. 여러 가지 주제(H)	25"			A: B:
9. 준비해 온 농담 이야기하기(B)	10"			
10. 사건(G)	30"			A: B:
11. 토론(E)	25"			A: B:
12. 취미(O)	20"			
13. 농사회 행사(J)	11"			
14. 우리고장이야기(P)	15"			
15. 다양한 수어 표현(I)	10"			
16. 영상을 보고 수어로 이야기하기 [애니메이션(Q)]	10"			
17. 수어어휘찾기(Q)	10"			
18. 옛날 수화와 현재 수화(L)	10"			
19. 농에피소드(R)	25"			

[부록 7] 촬영 일지(서식)

언어자료 동영상 촬영 일지			
일시			
장소			
참석자	사회자: 언어제공자:	작성자	
내용 및 결과			

과제명(코드)	예상 시간	소요시간 (시작-종료)	내용 및 결과
전체	273"	-	
1. 촬영 동의서	3"		
2. 수어 이름 말하기(A)	5"		
3. 다시 이야기하기_넥타이(D)	10"		
4. 표지판(F)	20"		
5. 흥부와 놀부(K)	10"		
6. 가장 기억에 남는 날(N)	20"		
7. 농학교 경험 이야기하기(M)	15"		
8. 여러 가지 주제(H)	25"		A: B:
9. 준비해 온 농담 이야기하기(B)	10"		
10. 사건(G)	30"		A: B:
11. 토론(E)	25"		A: B:
12. 취미(O)	20"		
13. 농사회 행사(J)	11"		
14. 우리고장이야기(P)	15"		
15. 다양한 수어 표현(I)	10"		
16. 영상을 보고 수어로 이야기하기 _애니메이션(C)	10"		
17. 수어어휘찾기(Q)	10"		
18. 옛날 수화와 현재 수화(L)	10"		
19. 농에피소드(R)	25"		

[부록 8] 한국수어 말뭉치 주석 규약

2019. 12. 03.

2020. 11. 27.

2021. 12. 10.

2022. 12. 16.

1. 개요

이 규약은 2015년 국내 최초로 한국의 원시 수어 말뭉치를 수집하면서 생성된 동영상에 주석하면서 제작된 ‘전사 지침’을 수정·보완한 것이다. 이 규약은 해마다 추가로 수행되는 주석 작업에 따라 새로운 내용이 추가되거나 수정될 수 있다.

이러한 이유로 규약의 명칭은 몇 차례 개정의 과정을 거쳤다. 한국수어 말뭉치 주석 작업이 처음 시도될 당시에는 ‘전사’, ‘전사 지침’ 등의 용어로 사용되었다. 그러나 문자가 없는 수어의 기록은 한국어 문자를 빌려와 이름(gloss)을 붙이는 방식임에도 언어의 형태를 그대로 받아 적는다는 의미의 ‘전사(transcription)’로 사용되어 언어학자 간 용어의 혼동을 초래하였다. 따라서 본 규약은 2019년에 ‘전사’를 ‘주석(annotation)’으로 통일하여 사용하기로 하고 ‘전사 지침’에서 ‘주석 지침’으로 개칭하였다.

또한 2020년에는 주석에 대한 지침을 제공하기보다는 현재까지 이루어진 주석에 대한 주석 규칙 및 약속을 공유한다는 차원에서 ‘주석 규약’으로 개정하였다.

이에 본 주석 규약은 지금까지 수행된 번역 및 주석을 수행하기 위해 알아야 할 구체적인 규칙과 약속을 다루고 있다. 2015년부터 수행된 의미 있는 손 활동에 대한 기본 주석, 2016년에 실시한 비수지 중 입 움직임에 대한 주석, 그리고 2019년에 실시한 결합 수어(합성어), 생산적 수어의 도상성, 의미 주석에 대한 기본 규약을 기술하였다.

향후 새로운 주석 과업에 수행되면 그에 따라 주석 규약은 수정 보완

될 것이며, 기타 항목에는 앞으로 수행할 수 있는 주석 과업 내용에 대한 제목만을 제시하였다.

1.1. 주석의 기본 원칙

한국수어 말뭉치 구축을 위한 주석은 엘란(EUDICO Linguistic Annotator, ELAN) 프로그램을 사용한다. 엘란은 멀티미디어 기록물(녹음 또는 영상자료 등)에 필요한 정보를 주석하여 탐색할 수 있도록 네덜란드 막스플랑크 심리언어학 연구소에서 개발한 프로그램이다.

주석의 가장 기본은 모든 의미 있는 손 활동에 이름을 붙이는 것이다. 따라서 주석을 하고자 할 때는 가장 먼저 엘란에 수어 화자별로 양손 활동을 기록할 수 있는 두 개의 수지 층렬을 만든다.

주석을 할 때는 각 수어 단어가 시작되는 지점에서 끝나는 지점까지를 분절하여 해당 부분에 글로스를 단다. 이때, 수지 층렬을 주석 작업에서는 실제 언어자료에서 나타나는 수어 어휘(토큰)의 기본형¹⁾을 타입으로 등록하며 체계적으로 관리하게 된다. 그래서 주석 작업을 수행할 때는 토큰-타입 매칭을 실행하여야 한다.

그런데 한국수어 말뭉치가 처음 구축될 당시 한국수어사전은 표제화 원칙에 따라 체계적으로 이루어진 것이 아니어서 본 주석 작업에서 사용되는 타입 목록은 한국수어사전을 참고할 수 없었다. 이에 주석 과정에서 새로운 토큰이 나타날 때마다 타입을 등록하면 타입 목록에 누적된다. 현재 타입 목록은 「한국수어 자료 통합지원 시스템」에서 관리하고 있다.

1.2. 번역

향후 한국수어 말뭉치를 공개할 경우, 수어를 모르는 사람들도 자료에 접근할 수 있도록 한국어 번역을 제공한다. 번역 목적에 따라 의역과 직역으로 제공할 수 있다. 다음은 두 번역의 특징이다.

1) 기본형(citation form)이란 농인에게 특정 수어를 물었을 때 제시하는 수어의 형태를 말하는데, 이때의 수어는 무표적인 형태이다.

① 의역(Free Translation)

- 목표어인 한국어로서 가장 자연스러운 형태로 제공한다.
- 출발어인 수어와 목표어인 한국어의 문장 길이가 일치하지 않을 수 있다.
- 수어의 지시 표현([지시])은 이해하기 쉽도록 한국어에서 가장 자연스러운 표현으로 번역한다.
- 일반적으로 수어 담화의 내용을 이해하는 데 도움이 되는 번역을 제공한다.

② 직역(Narrow Translation)

- 출발어인 수어의 절과 문장 수준으로 잘라 번역함으로 비교적 짧은 절 또는 문장으로 번역된다.
- 목표어인 한국어로 매끄럽게 번역되지 않을 수 있다.
- 수어의 지시 표현([지시])은 대명사를 사용하여 직접적으로 번역한다.
- 수어 담화 구조를 이해하는 데 도움이 되며 기계번역 알고리즘을 학습하고 테스트하는 데 유용하다.

2. 주석 총렬

한국수어 말뭉치는 모두 두 명의 언어제공자가 서로 마주보며 대화하는 형태로 구성되어 있다. 화면에 보이는 왼쪽 수어 화자는 A, 오른쪽 수어 화자는 B로 한다. 주석을 하기 위해 목적에 따라 총렬을 추가할 수 있다. 현재까지 만들어진 총렬은 다음과 같다.

2.1. 번역 총렬

지역의 원시 말뭉치를 수집하는 단계에서 번역 총렬은 향후 주석 작업자의 지역 수어 접근성을 높이기 위해 말뭉치 수집 지역의 수어 전문가(농민, 수어통역사 등)를 활용하여 기록한다. 즉 한국수어 말뭉치 수집

팀은 촬영, 편집, 탑재 및 번역 작업을 수행하여야 한다.

번역 층렬은 언어제공자당 한 개의 층렬로 구성되며 「한국수어 자료 통합지원 시스템」에서 주석 파일과 영상을 내려받기하여 기록하고 저장된 파일은 다시 시스템에 등록하여야 한다.

2.2. 수지 글로스 층렬

수지 글로스 층렬은 우세손 수지 글로스 층렬과 비우세손 수지 글로스 층렬로 나눈다. 층렬의 이름은 <표1>과 같다.

수지 글로스 층렬에는 해당 수어의 글로스를 입력한다.

<표1> 한 개의 수어 영상자료의 수지 글로스 층렬 예시

언어제공자	구분	층렬의 이름
수어 화자 A (화면에 보이는 왼쪽)	우세손	우세GlossA
	비우세손	비우세GlossA
수어 화자 B (화면에 보이는 오른쪽)	우세손	우세GlossB
	비우세손	비우세lossB

2.3. 의미 층렬

의미 층렬은 수지 글로스 층렬 바로 밑에 두고 수지 글로스 층렬과 동일하게 분절한다. 의미 층렬의 기록은 해당 수어의 맥락 속 의미를 입력하되, 한국어 표현이 다양한 경우 여러 개를 입력할 수 있다.

<표2> 수어 화자 A의 의미 층렬 예시

언어제공자	구분		층렬의 이름
	손	층렬	
수어 화자 A (화면에 보이는 왼쪽)	우세손	수지	우세GlossA
		의미	의미_우세A
	비우세손	수지	비우세GlossA
		의미	의미_비우세A

2.4. 코멘트 층렬

코멘트 층렬은 주석 시에 발견되는 특이사항 등을 자유롭게 텍스트로 입력할 수 있다. 단, 주석 목적에 따라 특정 내용을 기록하기로 약속하고 입력할 수 있다. 예를 들어 수지 글로스 층렬을 주석할 때 ‘음운변동’, ‘비우세손의 유지 현상’ 등이 나타나는 경우 이를 코멘트 층렬에 입력하였다.

2.5. 비수지 층렬

비수지 층렬은 연구 목적에 따라 다양하게 나눌 수 있지만, 주석의 편리를 위해서 눈과 머리 방향, 눈썹, 얼굴과 입 등으로 나누는 것이 좋다. 예를 들어, 눈과 머리의 방향은 주로 함께 움직이며 문법적인 기능을 나타내므로 하나의 층렬을 할당하고, 눈썹은 의문문, 화제화 등 문장의 형태를 나타내므로 다른 층렬을 할당한다.

3. 분절

3.1. 분절 원칙

수어 단어는 다음과 같이 분절한다.

- ① 이동이 없는 수어는 수형이 보이기 시작하는 첫 번째 프레임이 수어의 시작이다.
- ② 이동이 있는 수어 단어의 경우는 손이 움직이기 시작하는 첫 번째 프레임이 수어의 시작이다.
- ③ 이동도 수형의 변화도 없이 이전 수어에서 새로운 수어가 나타나는 경우는 수향이 변화하기 시작하는 첫 번째 프레임이 수어의 시작이다.
- ④ 수어 단어의 끝은 수형이 변화하기 시작하는 첫 번째 프레임의 바로 이전 프레임이다.
- ⑤ 이동이 시작되는 첫 번째 프레임의 바로 이전 프레임이 이전 수어 단어의 끝이다.

⑥ 양손 수어에서 두 손의 시작과 끝이 다를 경우 별도로 수어 영상을 확인하여 정확히 분절한다.

예) 우세손과 비우세손의 구현된 [공부1]을 표현한 후, 비우세손이 약간 내린 상태로 유지하고, 우세손은 다른 단어 [가다1]를 할 때 그 유지된 비우세손은 잘리지 않고 완전히 내릴 때까지 기다리다가 분절한다.

⑦ 양손 수어의 경우 두 손의 분절이 일치하지 않을 수 있다. 즉, 비우세손이 이전의 수어를 유지하고, 우세손이 다음 수어로 바뀔 때는 따로 분절한 후, 비우세손 코멘트 층렬에 ‘유지’라고 입력한다.

	 QR코드
<p>제스처 이동 단계의 예 ‘한 [가족]’ ※ 출처: De Vos, C.,Torreira, F. & Levinson, S C, 2015</p>	

분절 기준에
대한 설명 영상

4. 글로스

글로스는 해당 수어에 이름을 붙이는 것으로 같은 형태의 수어는 동일한 이름을 갖게 된다. 이름은 일련번호를 붙일 수도 있으나 해당 수어의 형태를 빠르게 떠올리기 위해 수어의 핵심 의미를 바탕으로 짓는다. 이러한 글로스는 수어의 의미도 번역도 아니다. 따라서 글로스를 부여하기 위해서 규칙에 대한 명확한 이해가 필요하다.

글로스를 부여할 때는 다음의 원칙에 따른다.

첫째, 글로스는 한글로 표기한다. 일반적으로 하나의 한국어 단어는 표준국어대사전에 등재된 어휘를 기본으로 한다.

둘째, 글로스는 가능한 한 해당 수어의 형태가 떠오를 수 있도록 핵

심 의미를 기반으로 하되 익숙하고 알기 쉬운 표현으로 선택한다.
 셋째, 동일한 수어 형태는 항상 같은 글로스를 가진다. 즉 맥락 안에서 의미가 달라진다고 하더라도 글로스는 항상 같다.
 넷째, 핵심 의미가 같고 형태가 다른 수어를 구분하기 위하여 한국어 단어 뒤에 번호를 붙인 ID 글로스를 사용한다.
 다섯째, 몸과 얼굴(비수지표지)은 종종 추가적인 의미 또는 전혀 다른 의미를 표현하기도 하지만, 수지 총렬 글로스 주석에서는 손의 움직임 정보만 기록한다. 예를 들어, 수어 화자가 머리를 좌우로 가로짓는 움직임과 함께 수지 기호를 표현한다면, 수지 기호에 대한 주석만 하고 부정 표현은 기록하지 않는다. 이는 글로스가 실제 번역이 아니라 단지 특정 수어에 대한 이름일 뿐이기 때문이다.

한국수어 말뭉치 구축에서는 상기 글로스 부여 원칙에 따라 다음과 같은 방법으로 글로스를 확정하였다.

① 글로스 선택 시 하나의 수어가 여러 개의 한국어 단어로 번역될 때는 가능한 많은 사람이 가장 쉽게 해당 수어를 떠올릴 수 있는 한국어 단어를 선택한다.

예 소개, 통역, 사회 → [소개1]

꼼꼼하다, 자세하다, 구체적 → [꼼꼼1]

② 영어는 한글로 표기한다.

예 ‘엘지’ → [엘지1]

[LG1] × (잘못된 글로스 표기)

예 ‘케이티엑스’ → [케이티엑스1]

[KTX1] × (잘못된 글로스 표기)

③ 문법적 표시는 가장 중립적인 것, 즉 무표적인 것으로 한다. 동사나 형용사의 경우 기본형으로 글로스를 붙인다.

예 잡다, 잡는다, 잡았다, 잡겠다 → [잡다1]

예쁜, 예쁘고, 예뻐다 → [예쁘다1]

④ 글로스는 하나의 단어로 하는 것을 원칙으로 한다. 단, 하나의 단어로 의미가 불충분한 경우 조사와 띄어쓰기 없이 붙여 쓴다.

예



[버스가다1]

⑤ 수어의 형태가 다르고 의미는 같은 경우 한국어 단어 뒤에 숫자 1, 2, 3, 등을 붙인 글로스를 한다.

예



[없다1]



[없다2]



[없다3]

⑥ 글로스로 사용한 한국어 단어가 동음이의어인 경우는 타입의 혼동을 방지하기 위해 다음과 같이 구분하여 글로스를 부여한다.

예 [예시1]

- ‘예’의 경우 ‘윗사람의 부름에 대답하거나 묻는 말에 긍정하여 대답할 때 쓰는 말’과 혼동을 방지하기 위하여 같은 의미의 다른 어휘인 [예시1]로 글로스를 정한다.

예 [값싸다1]

- ‘싸다’의 경우 ‘물건을 싸다’, ‘값이 싸다’ 등 다른 의미가 있기 때문에 이에 혼동을 방지하기 위하여 앞에 ‘값’을 붙여 [값싸다1]로 글로스를 정한다.

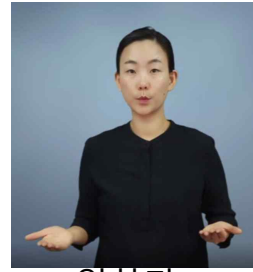
예 [일하다1]

- ‘일’의 경우 ‘자연수의 맨 처음 수’와 혼동을 방지하기 위하여 [일하다1]로 글로스를 정한다.

예



[값싸다1]



[일하다1]

⑦ 2015년 처음 주석에서는 각 토큰에 대해서 형태론적 변형과 음운 변이를 기록하여 구분하였으나, 그 후에는 상세 기록을 하지 않았다. 향후 연구 목적에 따라 형태론적인 변형과 음운 변이를 주석할 수 있다.

※ 2015년 글로스 예시

㉠ [주다1]: 글로스 기본형

㉢ [주다11]: [주다1]의 형태론적 변형([주다1]에서 주체와 대상이 바뀐 경우)

㉣ [주다11B]: [주다11]의 음운적 변이([주다11]에서 수형이 바뀐 경우)

4.1. 고정된 수어

사전에 등재되었거나 형태가 고정되어 어휘화된 것을 고정된 수어라고 한다. 고정된 수어의 글로스 예시는 다음과 같다.

① 수 또는 단위와 관련된 수어의 글로스는 다음과 같이 한다.

- 수를 나타내는 수어는 한국어 단어로 글로스를 붙인다. 즉 숫자 ‘1’과 ‘2’는 각각 [일1]과 [이1]로 글로스를 붙인다. 숫자 ‘10’의 두 형태는 각각 [십1] 또는 [십2] 등으로 글로스를 붙인다.

- ‘첫 번째’, ‘두 번째’와 같은 서수는 [첫번째1], [두번째1] 등으로



[십1]

글로스를 붙인다.

- ‘1년’, ‘2분’과 같이 숫자와 단위가 혼합된 수어의 경우는 다음과 같이 글로스를 붙인다.

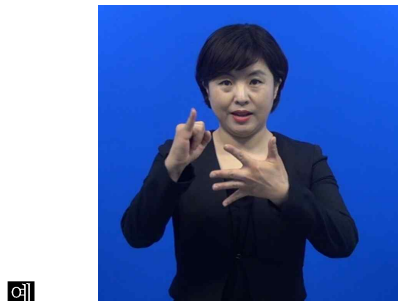
예 1년 → [일년1]

2분 → [이분1]

- 알파벳 지문자로 나타내는 단위 수어는 한국어로 글로스를 붙인다.

예 ‘mm’ → [밀리미터1]

- 한 손에 나타나는 ‘counting hand(수를 세는 손)’는 펼쳐진 손가락의 수와 상관없이 모두 [counting hand]로 글로스를 붙인다. 한 손 ‘counting hand(수를 세는 손)’와 함께 나타나는 다른 손 수어는 해당 수어의 글로스를 붙인다.



[counting hand 딸]



[counting hand 남동생]

예 우세: [남자4]

비우세: [counting hand]

예 우세: [딸1]

비우세: [counting hand]

- ② 형태가 같은 수어에서 경로 수동의 여부에 따라 다음과 같이 글로스를 한다.

예 동일한 수어 형태에서 [승용차1]이 경로 수동이 있는 경우는 [승용차가다1]로 글로스를 붙인다.

예 [승용차1]이 유지된 상태에서 경로 수동이 나타나는 경우 경

로 수동이 시작되는 지점부터 분절하고 타입명 [승용차가다1]로 글로스를 기록한다.



[승용차1]



[승용차가다1]

③ 양손을 사용하는 일치동사는 비우세손이 나타나지 않아도 되는 경우에는 다음과 같이 우세손과 비우세손 글로스를 따로 붙인다.

예 수어 의미 ‘흔내다’의 수지 글로스 예시



우세손: [야단치다1]

비우세손: [사람4]

4.2. 생산적 수어

생산적 수어는 관습적인 의미를 가지지 않고 맥락에 따라 자연스럽게 생성되는 수어로 맥락을 바탕으로 이해되는 수어이다. 그러므로 한 단어로 번역 또는 한 단어로 글로스를 정하기 어렵다.

능숙한 수어 화자들 사이에서 이해가 가능한 이 수어들은 단순히 사물을 수형이나 수동으로 나타내는 것이 아니라, 생산적 수어에서 나타나는 체계적인 규칙이 적용된다.

4.2.1. 생산적 수어의 주석 원칙

생산적 수어를 주석을 할 때는 수형을 기준으로 글로스(타입명)를 정하며 수형이 변화하는 경우 시작 수형을 기준으로 정한다. ‘생산적 수어’

의 글로스는 수형의 변별자질에 따라 결정된다.

첫째, 선택된 손가락에 따라 일형, 이형, 육형, 구형 등의 지숫자형 나타내는 수형으로 글로스를 붙인다.

둘째, 선택된 손가락의 벌림 유무에 따라 글로스는 달라진다. 지숫자형 이름 뒤에 숫자 1, 2, 3을 붙이며 타입명을 구분한다. 이때 손가락 벌림의 정도는 변이로 보아 별도로 구분하지 않는다.

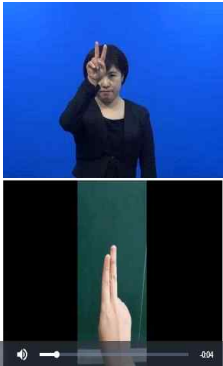






셋째, 선택된 손가락의 자세 즉, 손가락 관절의 구부림 유무에 따라 지숫자형 이름 뒤에 숫자 1, 2, 3을 붙이며 타입명을 구분한다. 이때 손가락 구부린 정도는 변이로 본다.



넷째, 선택된 손가락의 접촉 시 손의 모양이 평평하거나 원형인 경우는 다른 수형인 것으로 판단하여 타입명을 구분한다.

단, 이러한 변별자질을 고려할 때, ‘장향’, ‘손목의 구부림’까지는 고려하지 않는다.

다음은 한국수어 말뭉치 구축에서 나타난 생산적 수어의 글로스(타입)이다.

연번	타입명	그림	설명
1	일형2		일형 수형의 고리 형태 ※ [고리2]와 같지만 [고리2]는 고정된 수어이고, [일형2]는 생산적 수어라서 문제 없음.
2	일형3		일형 수형을 편 형태 양손 일형 수형으로 네모를 그리거나 한손 일형 수형으로 세모 혹은 원을 그리는 형태도 해당됨.

연번	타입명	그림	설명
3	이형1		<p>접촉이 없는 이형 수형</p> <p>접촉이 있는 이형 수형</p>
4	육형1		<p>양손 육형 수형으로 초승달과 같은 모양을 표현하는 것</p> <p>육형 수형으로 네모를 표현하는 것과 양손 육형 수형을 붙여 세모 모양으로 표현하는 것이 해당됨.</p>
5	육형2		<p>육형 수형이 벌어지고 줄어드는 것</p> <p>구부린 양손 육형 수형으로 네모를 표현하는 것 해당됨</p>
6	육형3		<p>구부린 육형 수형</p>
7	구형1		<p>구형 수형</p> <p>구형 수형으로 선이나 면으로 그리거나 양손으로 세모 그리거나 혹은 손가락을 벌린 구형으로 그리거나, 손가락을 벌린 구형이 선을 나타내는 것 해당됨</p>
8	구형2		<p>구부린 구형 수형으로 그리는 것</p> <p>'장소형' 수형의 양손으로 공, 덩어리 등을 나타내는 그 수형 포함됨</p>
9	구형3		<p>구형 수형을 평평하게 접촉하여 움직이는 형태</p>

연번	타입명	그림	설명
10	구형4		구형 수형을 원형으로 접촉하여 움직이는 형태 모아진 구형 수형으로 움직이는 형태
11	십형1		십형 수형으로 붓을 그리는 형태

4.2.2. 생산적 수어의 도상성

생산적 수어 어휘(productive signs)는 의미를 전달하기 위해 사물과 상황, 행위를 묘사한다. 수어 사용자는 이미지 형성 기법들을 사용해 기저 이미지를 환기시키며 맥락과 어떤 기법인지 확인하는 것은 생산적 수어 표현의 해석에 도움이 된다(Koenig, Konrad & Langer, 2008).

수어 사용자는 여러 형태적 특징 중에서 표현하고자 하는 지시체의 가장 핵심적이고 시각적인 특성을 선택하여 조어하는데, 이미지 형성 기법을 알면 생산적 수어 어휘를 적절하게 산출하는 데 도움이 될 뿐만 아니라, 생산적 수어 어휘의 의미를 이해하기 위해 기저 이미지를 인식하는 데에도 도움이 된다. 이미지 형성 기법들은 고정된 수어 어휘뿐만 아니라 생산적인 수어 어휘에도 적용된다. 2019년 한국수어 말뭉치 사례 연구에서 생산적 수어의 도상성을 분석하기 위해 활용된 이미지 형성 기법은 다음과 같다. 한국수어 말뭉치 연구에서는 독일 함부르크 대학교의 독일수어 연구소에서 제안한 이미지 형성 기법 중 손대체 기법, 손이용 기법, 스케치 기법, 측정 기법 총 4개의 기법만 활용되었다.

① 대체 기법(Substitutive technique)

- 손이 지시 대상 혹은 대상 물체의 주요 부분을 상징하는 기법

- 수어 공간 내의 수위, 수향, 수동은 대상 물체의 공간적 위치, 방향, 움직임에 상응함

예 [육형3]

[구형1]과 [주먹형1]



손대체 기법
(우) **육형3** (비) **육형3**
물체를 상하좌우로 옮기는
크레인 같은 기계의 고리로
물건을 집은 형태

손대체 기법
(우) **주먹형1** (비) **구형1**
컨테이너, 선적물, 화물

② 손이용 기법(Manipulative technique)^[SEP]

- 수어 사용자의 손이 만지거나, 잡거나 혹은 대상물을 사용하는 행위를 표현하는 기법

예) 손이 칼을 사용하는 모습을 보여 주거나 손으로 부채를 잡고 부채질하는 행위

- 손이 특정 위치에 가만히 있는 모습도 포함됨

예 [구형3]

[주먹형]



손이용 기법
(우) **구형3** (비) **구형3**
이불을 끌어당겨 잡은 손

손이용 기법
(두손) **주먹형1**
아이의 손을 잡은 손

③ 스케치 기법(Sketching technique)

- 손이 연필 혹은 붓 같은 스케치 도구이며, 수동을 통해서 사물의 모양을 그리는 기법
- 수동은 표현하고자 하는 사물의 움직임이 아니라 움직임이 없는 사물을 그리는 것임
- 사물의 모양을 수어 공간에 선으로 표현함

예 [구형2]

[십형1]



스케치 기법
(우) 구형2 (비) 구형2
구명보트 모양을 그리는 손



스케치 기법
(우) 십형1 (비) 십형1
백화점 원기둥이 큰크리트 기둥 또는 군데군데 세워진 모양

④ 측정 기법(Measuring technique)

- 손이나 손가락이 사물의 크기나 길이를 표현하는 기법
- 손이나 손가락이 지시 대상의 크기와 그 경계선이나 테두리를 보여줌
- 측정 기법의 수형은 대부분 무표 수형(unmarked hand shapes)으로 표현 (예: 펼친 수형, 검지 수형).
- 표현하고자 하는 사물의 모양을 보여주지 않고 그 사물의 크기나 길이를 나타냄

예 [구형1]



측정 기법

(우) 구형1 (비) 구형1

맥락정보 : [버스1] [구형1] [무겁다1] [몇1]
[구형1] 의미 : 버스(차)의 덩치가 큼

4.3. 결합 수어(합성어)

둘 이상의 수어가 순차적으로 결합된 형태의 경우, 하나의 단어인 합성어는 하나의 글로스로, 구로 나타나는 형태는 각 단어에 글로스를 붙인다. 구체적인 예시는 다음과 같다.

① 둘 이상의 수어가 결합된 형태로 명확하게 하나의 단어로 인지되는 경우에는 하나의 글로스로 주석한다. 예를 들어 ‘손’+‘돈’으로 된 수어는 두 수어를 묶어 분절한 후에 [수당1]로 글로스를 붙인다.

② 둘 이상의 수어가 결합된 형태에서 실제로 하나의 단어로 합성된 형태인지 의심스러운 경우는 두 개의 수어로 분리하여 각각 글로스를 붙인다. 예를 들어 ‘분홍’이라는 의미를 나타내는 수어 중 [빨강]+[약하다]로 된 수어는 [빨강1], [약하다1]로 글로스를 붙인다.

③ 둘 이상의 수어가 결합하여 하나의 합성어가 된 어휘는 다음과 같은 특성을 가진다.

- 두 수어의 조합이 특정한 의미를 갖는다.
- 조합된 두 수어 사이에 다른 수어나 신호를 삽입할 수 없다.
- 두 수어의 결합 시에 음운론적 특성은 다음과 같다.

■ 예 [알려주다1]

- [알다1]과 [주다1]이라는 두 수어가 결합 될 때 [알다1]의 반복 동작이 생략되고 두 수어를 연결하는 움직임(M)이 첨가되면서 수어

화자들은 하나의 단어로 인식함으로 합성어 [알려주다1]이 된다.

예 [안산1]

- [편안하다1]라는 수어와 [산1]이라는 수어가 결합되면서 [편안하다1]에서 마지막 정지(H)가 탈락하면서 [산1]이라는 수어 사이에 연결 움직임(M)이 첨가되면서 합성어 [안산1]이 된다.

예 [만약1]

- [거짓1]과 [그러면1]이라는 수어가 결합되면서 앞 수어가 뒤 수어의 수형에 동화되어 [거짓1]의 수형이 [그러면1]의 수형과 같아져 합성어 [만약1]이 된다.

4.4. 기타

① 수어 이름

• 수어 이름의 경우 대상자의 이름 앞에 [SN]을 붙인다. 수어 이름이지만 그 사람의 한국어 이름을 모르는 경우는 [SN]으로만 글로스를 붙인다.

예 ‘김영삼’ → [SN_김영삼]

예 ‘이름을 모르는 경우’ → [SN]

② 지역명

• 지역명의 경우 동일한 수어의 형태가 타입으로 존재하더라도 [지역명]으로 글로스를 붙인다.

예 ‘옥수수’와 ‘강원도’



[옥수수1]



[강원도1]

③ 외국수어/국제수화

- 수어 화자가 미국수어를 하는 경우 [미국수화]로 입력하고, 국제수어를 할 경우 [국제수화]로 입력한다. 그리고 그 수어의 의미는 코멘트 층렬에 입력한다.

④ 지시 수어

- 지시 수어는 수형에 따라 두 종류로 글로스를 정한다. ‘1지’만 편 수형은 [지시1], 손가락을 모두 편 수형은 [지시2]로 한다.

- 자세한 지시 글로스는 다음과 같다.

연번	타입명	그림	설명
1	지시1		1지만 편 수형으로 지시하는 형태
2	지시1호		1지만 편 수형으로 반원 또는 원을 그리는 형태 '반원'의 의미: 너희(의미 층렬) '원'의 의미: 우리(의미 층렬)
4	지시1_1		1지만 편 수형으로 1인칭 지시
5	지시1_1A		1지와 5지를 편 수형으로 1인칭 지시
7	지시1_1C		5지만 편 수형 1인칭 지시

번호	타입명	그림	설명
11	지시2		손가락을 모두 편 수형으로 지시하는 형태
12	지시2호		손가락을 모두 편 수형(손바닥이 위로, 아래로)으로 반원 또는 원을 그리며 지시
13	지시2_1		손가락을 모두 편 수형으로 1인칭 지시
14	지시2_1A		손가락을 모두 편 수형에서 손가락 끝으로 1인칭을 지시
15	지시2_1B		손가락을 모두 편 수형의 손가락을 모아 1인칭을 지시
	지시2그림		이형, 구형이 형태를 그리는 것 이형과 구형이 섞여 있음

⑤ 제스처

- 다음과 같은 기준에 의해 수어와 제스처를 구분한다.

㉠ 수어가 아니다.

㉡ 수어를 모르는 사람이 봐도 그 의미를 이해할 수 있다.

㉔ 수어를 모르는 청인도 사용하는 몸짓이다.

㉕ 제스처는 수지 글로스 층렬에 [제스처]로 입력하고, 코멘트 층렬에 구체적인 내용을 설명한다.

⑥ 지문자

- 한글의 자음과 모음을 나타내는 형태는 타입명 [한글지문자]로 글로스를 기록한다.

- [한글지문자]는 자음과 모음을 분리하여 코멘트 층렬에 기록한다.

예 수지 글로스 층렬: [한글지문자]

코멘트 층렬: [ㅎㅏ ㄱㅑㅓ]

⑦ 음운변동

- 수어가 앞뒤 수어의 영향을 받아 음운변동이 나타난 경우에는 수어의 원형에 해당하는 글로스를 달고 코멘트 층렬에 ‘음운변동’으로 주석을 단다.

⑧ 토큰으로 처리되지 않는 수어(Non-tokens)

- 수어 말뭉치 자료에서는 실제 사람들의 발화 형태를 반영하는 것이 아닌 형태들이 포함될 수 있다(Gabriele Langer 외. 2016). 즉, 대화 대상자의 발화를 그대로 반복하거나 따라 하는 것과 같은 것은 실제 수어의 사용 빈도에 포함시켜서는 안 된다. 따라서 이러한 수어는 토큰으로 처리하지 않으며 비토큰(non-token)이라 한다.

- 메타 언어적 언급은 토큰으로 처리하지 않는다.

- 손으로는 수어를 하면서 비수지로 부정을 표현하는 수어는 토큰으로 처리하지 않는다. 예를 들면 어떤 ‘특정 수어’를 하면서 이렇게 수어를 하지 않는다는 의미로 머리를 흔들면 이때의 ‘특정 수어’는 토큰이 아니다.

- 대화에서 이미 앞에서 발화한 수어를 그대로 모방해서 반복하거나 유도자료에서 제시된 수어를 따라 한 것은 토큰으로 처리하지 않는다. 이는 이미 앞에서 서술된 수어를 반복한 것이기 때문에 사용 빈도에 포함되어서는 안 되기 때문이다.

⑨ 그 외

- [%]: 수어가 완전히 조음되지 않은 경우 글로스 대신에 [%]로 주석을 입력하여 거짓 시작임을 나타낸다.

- [###]: 전사자가 수어로 인지했지만 무슨 수어인지 모르는 경우는 글로스 대신에 [###]로 주석을 단다. 그 후에 온라인 문의 공간에 문의를 하거나 전사 회의에 참석하여 의문을 해결하게 되면 ID 글로스로 바꿔 입력한다. 전사자들이 모두 모여 논의를 한 후에도 어떤 수어인지 명확하지 않을 경우에는 [###]를 그대로 둔다. 수어가 영상화면 바깥에서 조음이 된다고 하더라도 보이지 않는 경우에도 [###]로 입력한다.

- [@]: 수어 화자가 실수로 수어를 잘못 표현했고, 그 수어가 사용되지 않는 수어일 때 [@]로 입력한다.

5. 비수지 주석

한국수어 말뭉치 주석에는 ‘손 움직임’ 없이 나타나는 비수지 표지는 기록하지 않았다. 다만, 다음의 경우에만 코멘트 층렬에 입력을 하였다. 수어 없이 긍정의 의미로 고개를 끄덕인 경우 코멘트 층렬에 ‘끄덕임’으로, 수어 없이 부정의 의미로 고개를 흔드는 경우 ‘가로저음’으로 입력하였다.

2016년 한국수어 말뭉치를 활용한 사례연구를 위해 ‘입 움직임’과 관련된 주석을 실시하였으며 이와 관련된 규약은 아래와 같다.

5.1. 입 움직임

입의 움직임은 2개의 범주 즉, 마우딩(Mouthings)과 마우스 제스처(Mouth gestures)로 나눌 수 있다. 마우딩은 구어를 조음할 때처럼 입의 모양을 만드는 것을 말하며, 마우스 제스처는 구어의 단어를 조음하는 것 외의 모든 입의 움직임을 의미한다. 마우딩은 음성언어로부터 온 것으로 단어 전체 또는 한 부분을 조음하는 것처럼 보이는데 이 입의 움직임은 해당 수어 단어의 앞 또는 뒤까지 지속되어 나타나기도 한다. 그러나 마우스 제스처는 음성언어의 조음과는 관련이 없는 입의 움직임이다

(Crasborn, 2016).

5.1.1. 입 움직임 주석의 기본 원칙

입 움직임을 주석할 때 기본 원칙은 다음과 같다.

- ① 마우스 제스처는 타입을 추가하며, 마우딩은 별도의 타입이 없다.
- ② 마우스 제스처 관련 전사는 2개의 층렬 즉, 마우스 제스처 글로스와 마우스 제스처 분류(class) 층렬로 나눈다.
- ③ 마우딩 관련 전사는 4개의 층렬로, 마우딩을 기입하는 층렬 외에 스프레딩, 마우딩의 완성도, 마우딩의 분류 층렬로 나눈다.

5.1.2. 입 움직임 층렬

총 8개의 층렬이 2명의 언어제공자에게 각각 할당된다. 층렬의 이름은 띄어쓰기 없이 입력하며 추가된 층렬의 속성은 다음과 같다.

- ① Mouthing: 한국어 단어를 조음하는 것처럼 보이는 입의 움직임을 해당 단어의 철자법에 따라 한글로 입력하는 층렬이다.
- ② Spreading: 마우딩이 하나의 수어를 넘어서 얼마나 지속되는지를 입력하는 층렬이다.
- ③ Mouthing완성도: 마우딩으로 나타난 구어 단어의 완성도를 입력하는 층렬이다.
- ④ Mouthing분류: 마우딩을 특성에 따라 분류하고 이를 입력하는 층렬이다.
- ⑤ MouthGesture: 구어 단어를 발음하는 것 외의 모든 입의 움직임에 대한 글로스를 입력하는 층렬이다.
- ⑥ MouthGesture분류: 마우스 제스처를 특성에 따라 분류하고 이를 입력하는 층렬이다.
- ⑦ 수어분류: 마우딩 또는 마우스 제스처와 함께 나타난 수어의 품사 중에 명사와 동사를 구분하여 입력하는 층렬이다.
- ⑧ Mcoment: 입의 움직임과 관련한 주석 시에 생성되었으나 층렬에

답을 수 없는 다양한 의견 등을 자유롭게 입력하는 층렬이다.

5.1.3. 입 움직임 분절

① 수어가 시작되고 끝나는 지점과 상관없이 마우딩과 마우스 제스처가 시작하고 끝나는 지점을 분절의 기준으로 삼는다.

② 마우딩 또는 마우스 제스처의 마지막 모양이 의미 없이 오랜 시간 유지되는 경우는 분절에 포함하지 않는다.

③ 마우딩 또는 마우스 제스처 관련 속성의 층렬은 해당 마우딩 또는 마우스 제스처의 분절 프레임과 동일하게 분절한다. 즉, 스프레딩, 마우딩 완성도, 마우딩 분류 층렬은 마우딩 층렬 분절과 마우스 제스처 분류는 마우스 제스처 층렬의 분절과 동일하게 분절한다. M코멘트 층렬 역시 해당 마우딩 또는 마우스 제스처 분절과 동일하게 분절한다.

5.1.4. 마우스 제스처 주석

(1) 마우스 제스처 기록

① 마우딩 이외의 모든 입의 움직임을 특정 기호를 사용하여 입력한다.

② 입의 움직임을 나타내는 기호는 컴퓨터 자판에서 선택이 가능한 기호 중 최대한 도상성을 살린 기호로 정하였다. 가장 기본적인 기호는 다음과 같다.

③ 입의 모양은 기본 기호를 조합하여 입력한다.

- ‘입술을 내밀고 있는데 이가 보이는 경우’는 [BH]
- ‘입을 다 물고 양쪽 입꼬리가 약간 내려간 경우’는 [-^1]
- ‘이가 보이는 상태로 왼쪽 입꼬리만 올라 간 경우’는 [<H]
- ‘입이 약간 벌어지고 혀가 조금 나온 경우’는 [O1U1]
- ‘혀를 여러 번 내미는 경우’는 [U1~]

④ 연속적으로 일어나는 입의 움직임은 움직임의 순서대로 기호를 조합하여 입력한다.

- ‘입을 약간 벌리고 혀를 내밀었다가 집어넣으면서 바람을 내뿜는 경우’는 [O1UJ(=)]

- ‘입을 다물었다가 이가 보이게 양쪽으로 벌리면서 바람을 내뿜는 경우’는 [-H1(=)]

(2) 마우스 제스처 특성

① 마우스 제스처의 특성에 따른 분류를 입력하는 총렬이다.

② 마우스 제스처 분류는 다음의 특성에 따라 A, E, 4, W로 입력한다.

- A: 입의 움직임이 손으로 하는 수어의 의미를 꾸미거나 그 의미 정보에 대한 부가적인 정보를 추가해 주므로 문장 안에서 부사어(Adverbial)와 같은 역할을 하는 경우이다. 예를 들어 [바람/불다]라는 수어를 하면서 입으로 내는 바람에 의해 바람의 정도를 표현한다.

- E: 손의 움직임과 함께 조음소로 작용하는 입의 움직임으로 이 입의 움직임만으로는 추가적이거나 독립적인 의미는 없는 (Empty) 경우이다. 특히 이 입의 움직임이 없는 경우 농인들은 그 수어가 자연스럽게 있다고 느낀다. 예를 들면 [알다]라는 수어를 하면서 혀를 움직이는 동작, [있다]라는 수어를 하면서 입을 벌리는 것 등이 이 분류에 속한다. 또한 [끝]이라는 수어를 할 때 우세손 끝이 비우세 손바닥에 닿는 순간 ‘혀끝을 살짝 이로 무는’ 입의 움직임처럼 손의 움직임을 그대로 투영(Eco)시키는 경우도 이 분류에 속한다. 이 분류에 속하는 입의 움직임은 [가능하다]라는 수어에서처럼 손의 움직임은 없이 ‘입을 닫았다가 열면서 바람을 순식간에 내는(파)는 입의 움직임만으로도 의미를 생성하기도 한다.

- 4: 입이 실제 입의 역할을 하는 경우이다. 예를 들면 [먹다] [씹다] [외치다]와 같은 수어를 하면서 실제 먹거나 씹거나 외치는 입 모양을 만드는 것이다.

- W: 입의 모양이 독립적으로 나타나는 것이 아니라, 전체 얼굴 표정(Whole-face)의 한 부분으로 나타나는 경우이다. 이 분류의 입 움

직임은 주로 감정을 표현한다. 예를 들면 [징그럽다]라는 수어를 하면서 만드는 입의 움직임은 얼굴 전체의 움직임과 함께 나타나므로 이 분류에 속한다.

5.1.5. 마우딩 주석

(1) 마우딩 기록

① 구어의 단어를 조음하는 대로 입력하되, 조음점이 보이지 않는 경우는 마우딩을 하면서 함께 나타나는 수어를 참조해서 한글 맞춤법에 맞추어 입력한다. 예를 들어 ‘고기’와 ‘공기’, ‘오이’의 입 모양은 구분할 수가 없으나, 해당 분절의 관련 수어가 [고기]일 경우는 [고기]로 관련 수어가 [공기]일 경우는 [공기]로 입력한다.

② 마우딩이 단어의 일부분만 나타난 경우는 입 모양으로 표현된 부분을 입력하고 추론이 가능한 나머지 부분은 괄호를 하고 입력한다. 예를 들어 수어 [조심]을 하면서 입 모양은 ‘조’만 나타난 경우에 [조(심)]으로 주석을 입력한다.

③ 수어와 입 모양이 다른 경우에는 입 모양이 나타난 그대로 주석을 입력한다. 예) [일본]이라는 수어를 하면서 입의 움직임은 ‘일반’으로 나타난 경우, [일반]으로 입력하고 M코멘트(Mcomment) 층렬에 [일본]이라고 입력한다.

(2) 스프레딩(Spreading) 특성

① 마우딩은 지속 시간이 길어 하나 이상의 수어에서 나타나기도 한다. 이러한 현상을 스프레딩이라고 하며 마우딩을 기록할 때는 이러한 스프레딩 현상을 확인하기 위하여 별도의 층렬을 만들어서 기록하여야 한다.

② 마우딩이 지속되는 시간을 분절한 후에 다음의 특성에 따라 오른쪽, 왼쪽, 양쪽, 일치, 부분 등으로 스프레딩을 입력한다.

- 오른쪽: 마우딩의 시작점이 의미와 관련된 수어의 시작점과 같

으나, 아무 관련이 없는 다음 수어까지 지속되는 경우이다. 즉, 마우딩의 지속 시간이 해당 수어의 지속 시간보다 길어 다음 수어까지 덮는다. 예를 들어 [학교]라는 수어를 하면서 마우딩으로 ‘학교’를 하는데, 수어 [학교] 다음 수어인 [지시] 수어를 할 때까지 ‘학교’의 조음이 지속되는 경우에 [오른쪽]이라고 입력한다.

- 왼쪽: 마우딩이 의미와 관련된 해당 수어보다 먼저 나타나는 경우이다. 예를 들어, [장소]라는 수어와 함께 마우딩 ‘학교’가 시작되고 입 모양이 수어 [학교]가 나타날 때까지 지속되는 경우에 [왼쪽]이라고 입력한다.

- 양쪽: 마우딩이 의미와 연관된 해당 수어보다 먼저 나타나서 다음 수어를 할 때까지 지속되어 늦게 끝나는 경우이다. 예를 들어 [장소][학교][지시] 순으로 수어를 하는데, [장소]라는 수어를 하면서 ‘학교’라는 마우딩이 시작되어 [지시]수어를 할 때까지 지속되는 경우, [양쪽]이라고 입력한다.

- 일치: 마우딩의 시작점과 끝점이 의미와 연관된 해당 수어의 시작점과 끝점과 같은 경우에는 [일치]라고 입력한다.

- 부분: 마우딩이 의미와 연관된 해당 수어와 함께 나타나지만 지속되는 구간이 수어보다 짧은 경우이다. 예를 들어 ‘학교’ 마우딩이 [학교]수어와 함께 나타나긴 하지만 수어가 끝나기 이전에 입 모양이 먼저 끝나는 경우에는 [부분]이라고 입력한다.

③ 스프레딩 현상은 오른쪽, 왼쪽, 양쪽으로 나타날 수 있다. 이러한 스프레딩을 결정하는 시간 기준은 ‘스프레딩이 되었다고 보이는 부분이 10프레임 이상일 때’ 또는 ‘스프레딩이 앞 혹은 뒤 수어의 50% 이상을 덮을 때’이다. 즉, 오른쪽 스프레딩은 마우딩이 해당 수어와 함께 시작되어 다음 수어의 10프레임 이상 혹은 다음 수어 지속 시간의 50% 이상까지 지속될 때이다.

(3) 마우딩 완성도 특성

① 마우딩은 해당 구어 단어의 일부분만 나타나기도 한다. 이 총렬은 마우딩이 해당 단어의 모든 음절이 나타났는지 부분적으로 나타났는지에 대한 완성도를 확인하는 총렬이다.

② 마우딩 완성도는 다음의 특성에 따라 완, 앞, 뒤, 앞뒤 등으로 입력한다.

- 완: 구어 단어의 모든 음절을 조음한 경우이다. 예를 들어 ‘어린이’ ‘조심’ 등 모든 음절이 나타난 경우 [완]으로 입력한다.

- 앞: 구어 단어의 첫 음절만 조음한 경우이다. 예를 들어 ‘조심’에서 ‘조’ 음절만 나타난 경우 [앞]으로 입력한다. ̈ 뒤: 구어 단어의 마지막 음절만 조음한 경우이다. 예를 들어 ‘조심’에서 ‘심’ 음절만 나타난 경우 [뒤]로 입력한다. ̈ 앞뒤: 구어 단어의 첫 음절과 마지막 음절만 조음한 경우이다. 예를 들어 ‘자동차’에서 ‘동’ 음절은 탈락되고 ‘자차’라는 음절만 나타난 경우 [앞뒤]로 입력한다.

5.2. 기타 (시선, 몸, 어깨...)

향후 시선, 몸, 어깨 등 비수지 총렬에 대한 주석 작업이 이루어지면 추가되어갈 것이다.

연구 책임자 홍성은

공동 연구원 강창욱, 원성욱, 이은영, 우내리

연구 보조원 이한솔, 유재은, 손설표, 지혜원, 김효정

담당 연구원 곽정란(주무관)
 차예진(연구원)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9697

인쇄일: 2023년 12월 8일

발행일: 2023년 12월 8일

인 쇠: 우리동네출력소

※ 이 책은 국립국어원의 용역비로 수행한 ‘2023년 한국수어 말뭉치 수집 및 분석’ 연구 사업의 결과물을 발간한 것입니다.