

국립국어원 2025-01-32

발간등록번호

11-1371028-100029-01

국립국어원
정호
국립국어원
호정
호

2025년 한국언어문화 표·사진 설명 말뭉치 구축 연구

연구 책임자 | 차정원

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 '2025년 한국언어문화 표·사진 설명
말뭉치 구축 연구'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2025년 06월 12일 ~ 2025년 12월 19일

2025년 12월 19일

연구책임자: 차정원(데이터리)

연구 기관: 데이터리, 한림대학교 산학협력단

연구책임자: 차정원

공동연구원: 신서인

보조연구원: 오진영, 최윤진, 이민희, 서민지

최수찬, 조지원, 박송원, 박수연

유승빈, 전광훈, 김민혜, 정상미

이한범, 이수현, 박진현, 전상호, 김나현

이준희, 김푸른솔, 정민교, 강주은

<국문 요약>

2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구

이 사업은 ‘한국어언어문화’와 관련된 표·사진 설명 문장 기반의 대규모 한국어 말뭉치를 구축하여, 생성형 인공지능의 한국문화 이해도를 고도화할 수 있는 고품질 언어 자원을 생산하는 것을 목표로 한다. 이를 위해 한국어언어문화 분류 체계 연구, 표·사진 자료 수집, 설명 문장 작성, 지침 수립 및 보완, 말뭉치 품질 검증 등 전 과정을 체계적으로 수행했다. 본 사업의 주요 추진 내용은 아래와 같다.

첫째, 한국어언어문화 주제 체계 확립을 위한 기초 연구를 수행했다. 표·사진 설명 문장 말뭉치 구축에 앞서, 한국어언어문화 관련 자료를 수록하고 있는 수집처들의 분류 체계를 전수 조사하여 분류 체계 구성, 적용 가능성 등을 검토했다. 특히 ‘한국민족문화대백과사전’의 분류 체계를 기준으로 삼아 이를 체계적 지표로 활용했으며, 수집된 실제 자료의 특성과 분포를 기반으로 실제 적용 가능한 수정 분류 체계를 도출했다. 더불어, 구축된 데이터가 향후 확대될 경우를 고려하여 장기적으로 확장 가능한 범주 구조와 적용 방향을 함께 제시함으로써, 한국어언어문화 기반 멀티모달 말뭉치 구축의 기반 체계를 마련했다.

둘째, 한국어언어문화 관련 표·사진 자료를 수집하고 작업 대상을 선별했다. 자료 수집 단계에서는 저작권 조건이 공공누리 제1유형 또는 CC BY에 해당하여 상업적 활용과 2차적 저작물 작성이 가능한 안전한 자료만을 선별했다. 표 자료는 한국어언어문화와 관련된 전 범위의 통계 기반 자료를 폭넓게 수집하되, 수립한 일정 기준에 부합하며 작업이 가능한 총 2,000건 이상의 표 자료를 확보했다. 사진 자료는 한국어언어문화 분류 체계의 구축 우선순위(식생활-의생활-주생활-건축-공예)에 따라 단계적으로 수집했으며, 객체가 뚜렷하게 드러나는 유형 사진을 총 13,000건 이상 확보했다. 수집된 표·사진 자료는 말뭉치 작성에 활용 가능한지 여부를 기준으로 선별 및 정제 과정을 거쳤으며, 이를 통해 고품질의 구축 대상 원천 데이터를 확보했다.

셋째, 수집된 자료를 바탕으로 설명 문장을 작성했다. 표 자료에 대해서는 표 전체, 행, 열, 불연속 영역의 4종의 설명 영역을 지정한 뒤, 각 영역의 핵심 내용을 1문장으로 서술한 설명 문장 4개를 작성하여, 표 2,037건에 대해 총 8,148개의 설명 문장을 구축했다. 사진 자료에 대해서는 사진 속 대상을 명확히 제시하는 대상 식별 문장 1개와 분류

별 특성을 반영한 속성 기술 문장 4개를 작성하여, 사진 13,005건을 대상으로 총 65,025개의 설명 문장을 구축했다. 속성 기술 문장은 가시적 정보(형태, 색채 등)와 비가시적 정보(용도, 역사 등)로 구분하여 구성했으며, 가시적 정보를 우선적으로 서술하는 방식을 적용하여 기술의 일관성과 설명 구조의 명료성을 확보했다. 모든 설명 문장은 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>를 활용하여 구축했으며, 이후 분석자와 검토자의 교차 검토 과정을 통해 품질을 강화했다.

넷째, 표·사진 설명 문장 작성 지침을 수립 및 보완했다. 전년도 ‘2024년 표 설명 문장 작성 말뭉치 연구 분석’ 사업에서 마련된 ‘표 설명 문장 작성 지침’을 재검토하여, 설명 영역 지정 기준과 문장 작성 원칙을 중심으로, 보다 명료하고 일관된 체계로 보완했다. 사진 설명 문장 작성에 활용할 지침은 본 사업에서 새롭게 수립했다. 대상 식별 문장 작성 방법, 분류별 가시적·비가시적 속성의 구분, 속성별 기술 구성 원칙 등을 체계적으로 정리하여 작업자가 혼란 없이 일관된 방식으로 문장을 작성할 수 있도록 하는 ‘사진 설명 문장 작성 지침’을 마련했다.

다섯째, 납품 자료의 품질 검증 및 보완 체계를 수립했다. 구축된 설명 문장 말뭉치는 2차 이상의 교차 검수 절차를 거쳐 오류를 정정하고 문장 품질을 지속적으로 향상했다. 최종 구축 말뭉치에 대해서는 말뭉치의 성능을 구체적인 수치로 검증하기 위해, 전문가가 구축·검수한 정답 말뭉치 10%와의 비교를 통한 유사도 평가, 유의어 확장 기반의 확장 유사도 평가, 유창성 평가를 실시했다. 또한, 표 설명 문장 말뭉치의 경우 표의 실제 내용과 설명 문장 간의 일치도를 확인하기 위해 정합성 및 관계 추론 검증을 수행했다. 이 과정에서 검출된 오류 작업물은 재수정 절차를 통해 보완함으로써 말뭉치의 정확성을 확보했다.

본 사업팀은 국립국어원과의 긴밀한 협력 아래 사업의 전 과정을 체계적으로 수행하여, 한국어언어문화 기반 연구와 산업 전반에서 활용 가능한 고품질 대규모 한국어 말뭉치를 구축했다. 이를 통해 ‘2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구’ 사업을 성공적으로 완료했으며, 향후 한국문화 기반 데이터의 확장과 멀티모달 언어 기술 발전에 기여할 수 있는 실질적 기반을 마련했다.

주요어: 한국어언어문화, 표, 사진, 설명 문장, 말뭉치, 분류 체계, 수집, 한국어 데이터

<Abstract>

2025 Research and Analysis on the Korean Language and Culture Table·Image Description Corpus

This project aimed to build a large-scale Korean-language corpus based on table and image description sentences related to Korean Language and Culture, with the goal of producing high-quality language resources that can enhance generative AI's understanding of Korean culture. To this end, the project systematically carried out the entire process, including research on the Korean Language and Culture classification system, collection of table and image materials, description sentence writing, guideline establishment and refinement, and corpus quality verification. The main components of the project are as follows.

First, a foundational study was conducted to establish a thematic classification system for Korean Language and Culture. Prior to corpus construction, existing classification systems used by major data sources containing Korean Language and Culture materials were comprehensively reviewed to examine their structure and applicability. In particular, the classification system of the Encyclopedia of Korean Culture was adopted as a primary reference and utilized as a systematic framework. Based on the characteristics and distribution of the collected data, a revised classification system suitable for practical application was derived. In addition, a scalable category structure and long-term application strategy were proposed to support future corpus expansion, thereby establishing a foundational framework for building a Korean Language and Culture-based multimodal corpus.

Second, table and image materials related to Korean Language and Culture were collected and screened as construction targets. During the data collection stage, only materials with licenses permitting commercial use and derivative works—specifically those under Korea Open Government License Type 1 or CC BY—were selected. For table data, a broad range of statistics-based materials related to Korean Language and Culture were collected, and more than 2,000 tables were secured after evaluating their compliance with predefined criteria and construction feasibility. Image data were collected in stages according to the priority order of the classification system (food culture, clothing culture, housing culture, architecture, and crafts), resulting in over 13,000 images that clearly depict identifiable objects. All collected materials underwent selection and refinement processes based on their suitability for corpus construction, ensuring high-quality source data.

Third, description sentences were written based on the collected materials. For table data, four description areas—entire table, rows, columns, and non-contiguous regions—were defined, and one sentence was written for each area, resulting in 2,037 table description entries and 8,148 sentences. For image data, one target identification sentence clearly specifying the subject of the image and four attribute description sentences reflecting classification-specific characteristics were written, yielding 13,005 image description entries and 65,025 sentences. Attribute description sentences were structured by distinguishing between visible information (e.g., shape, color) and non-visible information (e.g., usage, history), with visible information described first to ensure consistency and clarity in description structure. All description sentences were created using the National Institute of Korean Language Multimodal Corpus Construction Support Tool, and their quality was further enhanced through cross-review by analysts and reviewers.

Fourth, guidelines for writing table and image description sentences were established and refined. The existing Table Description Sentence Writing Guidelines developed in the previous year's project, 2024 Table Description Sentence Corpus Research and Analysis, were reviewed and improved to provide clearer and more consistent standards, particularly regarding description area designation and sentence writing principles. In addition, new guidelines for image description sentence writing were developed in this project. These guidelines systematically define methods for writing target identification sentences, distinguishing visible and non-visible attributes by classification, and organizing attribute descriptions, enabling annotators to write sentences consistently and without ambiguity.

Fifth, a quality verification and refinement framework was established for the delivered corpus. The constructed description sentence corpus underwent at least two rounds of cross-review to correct errors and continuously improve sentence quality. For the final corpus, quantitative performance verification was conducted through similarity evaluation against a gold-standard corpus comprising 10% of expert-constructed and reviewed data, synonym-based extended similarity evaluation, and fluency evaluation. In addition, for the table description sentence corpus, consistency and relational inference verification was performed to assess the alignment between the actual table content and the corresponding description sentences. Detected errors were corrected through a revision process, thereby ensuring the accuracy of the final corpus.

Through close collaboration with the National Institute of Korean Language, the project team systematically carried out all stages of the project and successfully constructed a high-quality, large-scale Korean corpus applicable to both research and industry in the field of Korean Language and Culture. As a result, the 2025 Korean Language and Culture Table and Image Description Corpus Construction and Analysis Project was successfully completed,

providing a practical foundation for the future expansion of Korean culture-based data and the advancement of multimodal language technologies.

Key-words: Korean Language and Culture, table, image, description sentence, corpus, classification system, data collection, Korean language data

차 례

제1장 사업 개요

1.1. 사업 목적	3
1.2. 사업 범위	3
1.3. 사업 수행 기관	5

제2장 한국어문화 표·사진 설명 말뭉치 구축

2.1. 한국어문화 분류 체계 범주 연구	9
2.1.1. 한국어문화 분류 체계 선행 연구	9
2.1.2. ‘한국민족문화대백과사전’ 기반 분류 체계 연구	11
2.1.3. 인공지능 학습용 한국어문화 데이터 구축 분류 체계 연구	15
2.1.4. 범주 우선순위 선정	17
2.2. 표·사진 자료 수집 및 가공	19
2.2.1. 자료 수집 및 검수	19
2.2.2. 자료 가공	22
2.2.3. 자료 수집 내역	24
2.3. 표·사진 설명 말뭉치 구축	25
2.3.1. 구축 방법	25
2.3.2. 구축 도구	31
2.3.3. 구축 말뭉치 가공	37
2.3.4. 구축 내역	49
2.4. 표·사진 설명 말뭉치 평가	55
2.4.1. 평가 내용	55
2.4.2. 평가 결과	57

차례

제3장 표·사진 설명 문장 작성 지침

3.1. 지침 수정·보완 및 수립 방안	61
3.1.1. 표 설명 문장 작성 지침 수정·보완	61
3.1.2. 사진 설명 문장 작성 지침 수립	62
3.2. 지침 수정·보완 및 수립 내용	66
3.2.1. 표 설명 문장 작성 지침 수정·보완	66
3.2.2. 사진 설명 문장 작성 지침 수립	69
3.3. 지침 수정·보완 및 수립 결과	82

제4장 결론

4.1. 사업 요약	89
4.2. 의의 및 기대 효과	90
4.3. 향후 연구	92

참고 문헌

부록

- [붙임 1] 표 설명 문장 작성 지침
- [붙임 2] 사진 설명 문장 작성 지침

표 차례

<표 1> 유네스코 「무형문화유산 보호 협약」(2003)과 한국의 「문화재보호법」(2015) 비교 분석	10
<표 2> ‘한국민족문화대백과사전’ 분류 체계	11
<표 3> 분류 체계 개선안	13
<표 4> 수집처 6개 분류 체계	17
<표 5> 표·사진 자료 구축 전체 과정	19
<표 6> 표 수집 json 구조	22
<표 7> 사진 수집 json 구조	23
<표 8> 표 설명 문장 구축 교육 내용	26
<표 9> 사진 설명 문장 구축 교육 내용	27
<표 10> 사진 설명 문장 작성 시 자주 발생하는 오류	27
<표 11> 표 설명 문장 작성 및 교차 검수 작업 내역	49
<표 12> 사진 설명 문장 작성 및 교차 검수 작업 내역	50
<표 13> 표 설명 말뭉치 구축 데이터 전문가 자문 내용	52
<표 14> 사진 설명 말뭉치 구축 데이터 전문가 자문 내용	54
<표 15> 사진 설명 문장 작성 지침 수립 및 수정·보완 회의 내역	64
<표 16> 설명 문장 작성 지침 전문가 자문 내용	64
<표 17> <2025년 한국어문화 표 설명 문장 작성 지침> 표 설명 문장 작성 예시 ...	67
<표 18> <2025년 한국어문화 표 설명 문장 작성 지침> 주요 수정 내용 ...	69
<표 19> <2025년 한국어문화 사진 설명 문장 작성 지침> 주요 수정 내용 ...	81

그림 차례

[그림 1] 표 설명 문장 작성 및 검수 절차	28
[그림 2] 사진 설명 문장 작성 및 검수 절차	29
[그림 3] 표 작업 프로젝트 생성 화면	31
[그림 4] 표 설명 문장 구축 도구 전체 화면	31
[그림 5] 표 작업 목록 추가 방법	32
[그림 6] 표 설명 문장 작성 방법	32
[그림 7] 표 설명 문장 검수 도구 전체 화면	33
[그림 8] 표 설명 문장 검수 및 수정 방법	33
[그림 9] 사진 작업 프로젝트 생성 화면	34
[그림 10] 사진 설명 문장 구축 도구 전체 화면	34
[그림 11] 사진 카테고리 지정 방법	35
[그림 12] 사진 설명 문장 작성 방법	35
[그림 13] 사진 설명 문장 검수 도구 전체 화면	36
[그림 14] 사진 설명 문장 검수 및 수정 방법	36
[그림 15] 국가통계포털(KOSIS) 표 자료 제공 형식	61
[그림 16] 한국어언어문화 분류별 설명 속성 구축을 위한 사전 조사	63
[그림 17] 사진 설명 문장 지침 수립을 위한 샘플 작업 예시	63
[그림 18] 사진 설명 문장 작성 예시	70



제 1 장

사업 개요



1.1. 사업 목적

‘2025년 한국언어문화 표·사진 설명 말뭉치 구축 연구’ 사업은 대규모 표·사진 자료를 한국언어문화 분류 체계에 따라 수집하고, 이를 기반으로 설명 문장을 작성하여 생성형 인공지능의 한국문화 이해도를 고도화할 수 있는 고품질 학습 자료를 구축하는 것을 목표로 한다. 본 사업은 기존 언어 자원에서 상대적으로 부족했던 한국언어문화 영역의 정보를 체계적으로 보완함으로써 한국문화 기반 데이터의 의미적 다양성과 심층성을 확대하는데 의의를 둔다. 또한 설명 문장 작성에 필요한 지침을 마련하고 표준화하여 유사 사업의 수행 기반을 확보하고, 국어 자원의 신뢰성과 확장성을 강화하고자 한다. 구축된 표·사진 설명 문장 말뭉치는 텍스트와 시각 정보의 정합성을 기반으로 멀티모달 언어 처리 기술의 발전을 지원하고, 생성형 인공지능의 환각 현상을 줄이는 데 기여하는 핵심 자원으로 활용될 수 있다. 아울러 본 사업을 통해 산출된 말뭉치, 지침, 평가 결과 등은 민간에도 개방하여 데이터 산업과 언어 인공지능 생태계의 품질 제고와 기술 고도화에 기여하는 것을 목표로 한다.

1.2. 사업 범위

본 사업의 수행 범위는 다음과 같다.

- 한국언어문화 분류 체계 연구 및 범주 우선 순위 고려
 - 한국언어문화 관련 자료를 제공하는 수집처의 분류 체계를 조사하여 분류 체계의 구성 방식, 적용 가능성, 체계 간 구조적 관계 등을 분석한다.
 - ‘한국민족문화대백과사전’의 분류 체계를 기준으로 삼아 실제 수집 데이터의 분포를 검토하고 본 사업의 목적에 적합한 한국언어문화 분류 체계를 도출한다.
 - 구축 범위의 우선순위를 설정하여 본 사업을 통해 우선적으로 구축할 범주를 단계별로 확정한다.

- 표·사진 자료 수집 및 선별
 - 공공누리 제1유형 또는 CC BY 저작권을 가진 안전한 자료만 수집하여 상업적 활용

과 2차 저작물 작성이 가능한 원천 데이터를 확보한다.

- 표 자료는 한국어문화와 관련된 통계 기반 자료를 중심으로 2,000건 이상 수집하며 작성 가능 여부를 기준으로 작업 대상을 선정한다.
- 사진 자료는 설정된 분류 체계의 우선순위에 따라 객체 식별이 가능한 이미지를 13,000건 이상 단계적으로 수집한다.
- 수집된 자료는 형식, 내용, 품질 기준을 검토하여 말뭉치 구축에 활용할 수 있는 작업 대상만을 선별한다.

□ 표·사진 자료 정제

- 수집된 표 자료를 멀티모달 도구에 입력하기 위해 메타데이터를 JSON 형태로 변환하고 정비한다.
- 사진 및 표 자료에 대해 파일명 부여, 중복 제거, 주제 분류 등 후처리 작업을 수행하여 설명 문장 작성이 가능한 형태로 정제한다.
- 정제 과정에서 오류 자료 또는 기준에 부합하지 않는 자료는 제외한다.

□ 표·사진 설명 문장 작성 및 교차 검수

- 표 자료에 대해 표 전체, 행, 열, 불연속 영역의 4종 설명 영역을 지정한 후, 각 영역의 핵심 내용을 반영한 설명 문장 1문장씩 총 4문장을 작성한다.
- 사진 자료에 대해 사진 속 대상을 명확히 제시하는 대상 식별 문장 1문장과, 가시적·비가시적 속성에 기반한 속성 기술 문장 4문장을 작성한다.
- 모든 설명 문장은 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>를 활용하여 작성한다.
- 분석자와 검토자가 상이한 2차 이상의 교차 검수 절차를 반드시 거쳐 설명 문장의 품질을 보완한다.
- 1차 설명 문장 작성 이후에는 반드시 2차 교차 검수가 수행되어야 하며, 이 절차까지 완료되어야 하나의 프로젝트 단위가 최종적으로 마무리된다.

□ 표 설명 문장 작성 지침 보완

- ‘2024년 표 설명 문장 작성 말뭉치 연구 분석’ 사업에서 수립된 지침을 검토하여 설명 영역 지정 기준, 문장 작성 방식 등을 보다 명료하고 일관되게 보완한다.

□ 사진 설명 문장 작성 방법론 및 지침 수립

- 사진 속 대상을 식별하는 문장 작성 방식, 가시적·비가시적 속성 구분 체계, 속성 기술 구성 원칙 등 사진 설명 문장 작성에 필요한 요소를 정의한다.
- 다양한 유형의 한국어언어문화 사진 자료에 적용 가능한 사진 설명 문장 지침을 신규로 수립하여 작업자의 혼란을 방지하고 문장 작성의 일관성을 확보한다.

□ 납품 자료의 품질 보증 및 보완 체계 수립

- 고품질 말뭉치 구축을 위해 작업자 모집 기준, 교육 계획, 역할 분담 등 인력 운영 체계를 설계·수행한다.
- 사업 공정과 산출물의 체계적 관리를 위한 사업 관리 방안을 마련한다.
- 구축 완료 후 전문가 구축 정답 말뭉치 10%와의 비교를 기반으로 유사도 평가, 유의어 확장 평가, 유창성 평가 등을 실시하여 말뭉치의 품질을 정량적으로 검증한다.
- 표 설명 문장은 정합성 및 관계 추론 검증을 추가로 수행하여 표 내용과 설명 문장 간의 일치도를 확인한다.
- 오류 자료가 발견될 경우 하자 보수를 위한 재수정 절차를 운영하여 최종 납품 자료의 신뢰성과 정확성을 보장한다.

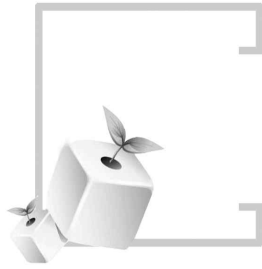
1.3. 사업 수행 기관

본 사업은 데이터리(Data.ly)와 한림대학교가 공동으로 참여하여 수행됐으며, 두 기관은 각자의 전문 역량에 따라 역할을 분담했다. 데이터리는 전체 사업 관리와 도구 테스트, 말뭉치의 수집·구축·관리 등 기술적 업무 전반을 담당했으며, 한림대학교는 국어학적 전문성을 기반으로 지침 수립 및 정답 말뭉치 구축 업무를 전담했다.

데이터리는 주로 기술적 기반을 담당하며, 한국어언어문화 분류 체계에 따라 수집된 표·사진 자료를 정제하고 구축 가능한 형태로 변환하는 작업을 수행했다. 변환된 자료는 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>에 업로드됐으며, 작업자들을 관리·감독하여 주 단위 구축 및 검수 작업을 총괄했다. 또한 작업자 모집과 배치, 일정 관리, 공정 모니터링 등 운영 전반을 관리하며 작업 흐름이 안정성 있게 유지될 수 있도록 조정했다.

더불어, <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구> 테스트 과정에서 발견된 오류와 개선 사항을 국립국어원과 공유하며 도구의 품질 개선에 적극 기여했다. 구축이 완료된 후에는 결과물을 납품 형식에 맞게 정리 및 변환하고, 정량적 평가 체계를 설계하여 말뭉치 품질을 검증했다. 이 과정에서 데이터리는 사업 진행 상황을 국립국어원에 주기적으로 보고하고 기관 간 소통과 협의를 조율하는 역할도 수행했다.

한림대학교는 언어학적 전문성을 바탕으로 표·사진 설명 문장의 언어적 품질을 책임지는 역할을 맡았다. 한림대학교는 표 설명 문장 작성 지침을 보완하고, 사진 설명 문장 작성을 위한 신규 지침을 수립하여 구축 과정 전반에서 활용 가능한 언어적 기준을 정립했다. 특히 대상 식별 방식, 가시적·비가시적 속성 구분, 속성 기술 방식 등 설명 문장 작성 과정에서 발생하는 다양한 질의를 검토하고, 작업 과정에서 드러난 모호한 표현이나 오류 가능성을 분석하여 지침을 더욱 명료하게 구체화했다. 또한 정답 말뭉치 구축을 통해 실제 설명 문장을 직접 작성하고 검토하며, 지침과 구축 결과물의 정합성과 일관성을 확보했다.



제 2 장

한국언어문화 표·사진 설명 말뭉치 구축



2.1. 한국어언어문화 분류 체계 범주 연구

본 사업의 진행 과정과 향후 확장 가능성을 고려하여 한국어언어문화 분류 체계에 대한 기초 연구를 수행했다. 본 사업은 첫 연차로서 기존에 참고할 만한 한국어언어문화 분류 체계 연구가 부재했기 때문에 사업을 안정적으로 추진하기 위한 기준 체계가 필요했다. 이에 따라 초기 단계에서는 한국문화 관련 범주 구성의 체계성과 완성도가 가장 높게 평가되는 ‘한국민족문화대백과사전’의 분류 체계를 본 사업의 기준 분류 체계로 채택했다.

이후 보다 적합한 분류 체계를 마련하기 위해 주요 표·사진 수집처에서 사용하고 있는 분류 체계를 비교하고 분석하여 자료 수집 과정에서 확인한 실제 데이터의 분포를 토대로 ‘한국민족문화대백과사전’의 분류 체계를 본 사업 목적에 맞게 조정한 수정 분류 체계를 제안했다.

아울러, 한국어언어문화 관련 멀티미디어 데이터를 효과적으로 수집·관리·활용하기 위한 매트릭스형 하이브리드 분류 체계를 별도로 설계하여 인공지능 학습용 한국어언어문화 멀티미디어 데이터 구축에 활용할 수 있는 또 하나의 분류 체계로 제시했다.

2.1.1. 한국어언어문화 분류 체계 선행 연구

한국어언어문화 및 문화재를 대상으로 한 분류 체계 연구는 주로 문화재의 활용 목적, 활용 방식, 정책적 관리를 중심으로 이루어져 왔다. 「문화재 활용을 위한 정책기반 조성 연구」(문화재청, 2006)와 「문화재 활용 가이드북」(문화재청, 2007)은 문화재 활용을 관광화, 교육화, 산업화, 상품화 등 활용 목적 중심으로 분류하였으며, 유형문화재에 국한된 분류를 시도했다. 또한 「문화재 유형별 활용 길라잡이」(문화재청, 2010)는 문화재의 종류 및 성격, 향유 효과, 향유 방식이라는 세 가지 기준에 따라 분류 체계를 제시하였다. 「문화재 활용 기본계획 연구」(류호철 외, 2012)는 문화재 활용 유형을 현재화, 대중화, 산업화, 세계화 등 활용 방법과 목적 중심으로 구분하였다.

한편, 「무형문화유산과 무형문화재 유형별 분류체계에 관한 연구」(최종호, 2016)는 유네스코 「무형문화유산 보호 협약」과 국내 「문화재보호법」을 비교 분석하여 국내 무형문화재 분류 체계가 국제 기준과 구조적으로 불일치함을 지적하였다. 해당 연구는 구전 전통, 문화공간 등 유네스코가 핵심 범주로 포함하는 요소들이 국내 체계에서는 유형문화재나

민속자료로 분산되어 관리되고 있음을 문제로 제시하였다.

개정전 문화재보호법	개정후 문화재보호법	유네스코 보호협약	중국	일본	대만
	구전 전통 및 표현	무형문화유산의 전달체로서의 언어를 포함한 구전 전통 및 표현	전통구전문학 및 그 전달체로서의 언어	민간전승	
음악 무용	전통적 공연.예술	공연 예술	전통기예, 음악, 무용, 희극, 곡예와 잡기	향토예능, 민속예능	표현예술
음식제조	의식주 등 전통적 생활관습	사회관습, 의례, 축제 행사			
의식	민간신앙 등 사회적 의식		전통예절과 의식, 명절 등 민속	풍속관습	풍속, 신앙, 절경, 상관문물, 기타
놀이	전통적 놀이.축제 및 기예.무예				
무예					
	한의약, 농경.어로 등에 관한 전통지식	자연과 우주에 관한 사회적 관습	의약, 역법		
공예기술	공예, 미술 등에 관한 전통기술	전통 기술	전통미술, 서예	공예기술, 민속기술	전통공예미술

<표 1> 유네스코 「무형문화유산 보호 협약」(2003)과 한국의 「문화재보호법」(2015) 비교 분석

이러한 선행 연구들은 한국의 문화재와 무형문화유산을 체계적으로 정리하고 제도화하는 데 중요한 기초를 제공했다는 점에서 의미가 있다. 특히 문화재의 활용 목적, 향유 방식, 정책적 관리 기준을 중심으로 분류 체계를 정립함으로써 문화재 행정과 활용 정책 수립에 실질적인 기준 틀을 마련했으며, 유네스코 기준과의 비교를 통해 국내 분류 체계의 구조적 특성과 한계를 진단하고 국제 기준과의 정합성 문제를 공론화했다는 점에서도 학술적·제도적 기여가 있다.

그러나 이러한 선행 연구들은 문화재의 보호·활용·정책 관리를 목적으로 한 분류 체계 연구라는 공통점을 가지며, 실제 데이터 구축이나 인공지능 학습을 위한 표·사진 기반 언어 데이터 분류 체계 연구는 거의 이루어지지 않았다는 한계를 지닌다. 특히 하나의 자료에 복수의 문화 요소가 혼합된 표 자료나, 시각적 해석이 필요한 사진 자료를 대상으로

한 인공지능 학습 데이터 구축 목적의 분류 체계 검토는 선행 사례가 부족한 상황이다.

이에 본 사업은 ‘한국민족문화대백과사전’의 분류 체계를 기준으로 삼되, 실제 수집된 표·사진 자료의 특성과 설명 문장 작성 가능성을 중심으로 분류 체계를 재검토하고 수정·보완하여, 실제 데이터 구축과 멀티모달 학습에 적합한 방향으로 분류 체계를 조정하였다. 이는 기존의 제도·정책 중심 분류와 달리, 언어 데이터 구축과 인공지능 학습 활용을 직접적으로 고려한 실증적 분류 체계 연구라는 점에서 차별성을 가진다.

2.1.2. ‘한국민족문화대백과사전’ 기반 분류 체계 연구

본 사업에서 기준으로 삼은 ‘한국민족문화대백과사전’의 분류 체계는 아래의 <표 2>와 같다. 수집된 표 자료를 분석한 결과, ‘예술·체육’ 분야에서는 하나의 표에 ‘무용’, ‘대중음악’, ‘국악’ 등 여러 중분류 요소가 혼합된 사례가 다수 확인됐다. 이에 기존 중분류 체계로는 명확한 분류가 어려운 표를 포괄하기 위해 ‘예술·체육/기타’ 항목을 추가했으며, 이를 제외한 나머지 분류 체계는 기존 기준을 유지하여 적용했다.

대분류	중분류
경제·산업	경제, 교통, 산업, 통신
과학	과학기술, 동물, 식물, 의약학
교육	교육
문학	고전산문, 고전시가, 구비문학, 한문학, 현대문학
예술·체육	건축, 공예, 국악, 대중음악, 무용, 서예, 연극, 영화, 조각, 회화, 체육, 기타
정치·법제	국방, 법제·행정, 외교, 정치
생활	민속·인류, 식생활, 의생활, 주생활
언론·출판	언론·방송, 출판
언어	언어·문자
역사	선사문화, 고대사, 고려시대사, 조선시대사, 근대사, 현대사
종교·철학	개신교, 대종교, 도교, 민간신앙, 불교, 신종교, 원불교, 유교, 천도교, 천주교
지리	인문지리, 자연지리
사회	가족, 사회구조, 촌락

<표 2> ‘한국민족문화대백과사전’ 분류 체계

본 연구에서는 위 분류 체계를 중심으로 한국언어문화 데이터의 수집 현황과 수집처 분류 체계를 조사하여 한국언어문화 관련 멀티모달 자료의 분포와 현행 분류 체계의 구조적 특징을 분석했다. 이를 바탕으로 다음의 사항을 중점적으로 검토·제안했다.

- 현행 분류 체계의 적절성 평가
 - 범주 간 누락·중복 여부, 용어의 모호성, 분류 기준의 일관성 등을 분석하여 개선 필요 영역을 검토했다.
- 수집된 실제 자료의 특성에 기반한 하위 분류 보완 및 세분화 방안
 - 기존 분류 체계만으로는 포착되지 않는 객체류, 사진 기반 속성, 통계표 기반 내용 등을 반영하기 위한 조정 방안을 마련했다.
- 멀티모달 말뭉치 구축에 적합한 분류 규칙 및 적용 지침
 - 텍스트·이미지 간 정합성을 고려한 분류 기준, 멀티모달 데이터의 속성(가시적·비가시적 정보)을 반영한 적용 원칙 등을 제안했다.

□ 한국어언어문화 수집 자료 분석

한국 공공기관을 중심으로 약 30여 개 수집처의 ‘한국언어문화’ 관련 데이터를 조사한 결과, 다음과 같은 특징이 확인됐다.

첫째, 자료는 유형·무형으로 구분된다. 사진 데이터는 사물 중심의 유형 자료와 행위·현상 중심의 무형 자료로 구성되어 있었다. 본 사업은 대상의 실체가 명확한 유형 자료를 중심으로 수집했으나, 향후 무형 자료까지 포함할 수 있는 가능성을 확인했다.

둘째, 수집처 간 공통 분류와 독자적 분류가 혼재한다. ‘식생활’, ‘의생활’, ‘주생활’은 대부분의 수집처에서 공통적으로 나타났으며, ‘예술’, ‘문화’, ‘사회’, ‘종교’ 등도 반복적으로 등장했다. 반면, 일부 기관은 자료 특성에 따라 고유한 분류 체계를 운영하고 있었다. 예를 들어, ‘국가유산 디지털 서비스’는 지정 종목별로 세분화된 분류를 운영하고 있었고, ‘지역어 종합 정보’는 독자적인 주제 분류를 적용했다. ‘한국민속대백과사전’은 표제어 기반의 분류를, ‘공유마당’은 자료의 장르나 생산 주체에 따른 분류 체계를 사용하고 있었다.

셋째, 표·사진 데이터는 하나의 항목에 명확히 귀속되기 어려운 중첩 분류 사례가 빈번하다. 표 데이터의 경우 ‘공연예술 장르별 공연 건수 및 횟수’, ‘17개 시도별 문화예술 활동 내역’ 등은 대분류상 ‘예술·체육’으로 분류할 수 있으나, 세부 내용에는 ‘무용’, ‘대중음악’, ‘국악’ 등 여러 중분류 요소가 혼합되어 있어 단일 항목으로 구분하기 어려웠다. 사진 데이터에서도 분야가 중첩되는 예가 있었다. 예컨대 ‘장롱’은 의복과 관련된 도구로 보아

‘의생활’에 분류할 수도 있지만, 가정의 가구로 보아 ‘주생활’에 포함할 수도 있다. 본 사업에서는 이러한 혼선을 줄이기 위해 실제 용도와 속성을 기준으로 한 분류 원칙을 마련하여 적용했다. 다만 일부 자료는 여전히 복합적인 속성을 지니고 있어, 중첩 데이터 처리에 대한 보다 정교한 방안 마련의 필요성도 확인했다.

□ 분류 체계 개선 방향 제안

객체 분류	대분류	중분류	예시	기존 분류	비고
유형	생활문화	음식	비빔밥, 김치, 송편, 탕, 전 등	생활>식생활	조리의 결과물 또는 섭취의 대상
		의복·장신구	한복, 두루마기, 비녀, 노리개, 귀걸이	생활>의생활 예술·체육>공예	인체 착용/장식용 도구
		식기·조리도구	분청사기 그릇, 놋그릇, 젓가락, 맷돌	생활>식생활 예술·체육>공예	음식의 담음/조리용 실사용 도구
		가구·생활용품	장롱, 상, 바느질상자, 촛대	생활>주생활 생활>의생활 예술·체육>공예	주거 내 생활용 도구
	건축·주거문화	전통가옥	한옥, 사랑채, 안채, 초가집	생활>주생활	거주용 전통 건축·실내/외 공간
		종교건축물	사찰, 법당, 탑(불탑), 성당, 교회	예술·체육>건축	종교 의례용 공간·구조물(기능 기준)
		궁궐·관아·누정	창덕궁, 정자, 누정, 행각	예술·체육>건축	공공·의례·휴식 기능의 전통 건축
		공공·교육건축	향교, 서원, 서당, 학당, 사당	예술·체육>건축	전통 건축물에 한함
		현대건축	남산타워, 청와대, 국회의사당, 시청, 국립박물관	예술·체육>건축	근·현대 재료/양식의 공공·상징 건축, 전망/통신 구조 포함
	예술문화	공예품	도자기, 나전칠기, 금속·목공예	생활>식생활 생활>주생활 예술·체육>공예	작품성·유물성 중심의 조형물
		악기	가야금, 대금, 장구, 북	예술·체육>공예	연주용 실물 악기
		회화·서예 작품	병풍(회화), 서예작품, 민화	예술·체육>공예	평면 예술 중심
		조각·기념조형물	인물 동상, 충혼탑·독립기념탑	예술·체육>건축 예술·체육>공예	비종교적 기념·예술 조형물
	종교·의례문화	의례용품	향로, 제기, 제단	예술·체육>공예	종교 행위의 매개 사물(도구)

		성상·유물	경전함, 불상·성모상	예술·체육>건축 예술·체육>공예	종교 상징물(설치물)· 전시 유물
무형	풍속문화	전통놀이	옷놀이, 널뛰기, 제기차기	-	행위 중심(사물보다 활동이 피사체)
		세시풍속	김장, 모내기, 명절 제사	-	특정 계절·시기마다 행해지는 절차
		노동 행위	다듬이질		노동 관련 행위
		의례	세배, 제사, 석전대 제, 혼례 절차	-	의례·제의 절차 중심

<표 3> 분류 체계 개선안

대량의 한국언어문화 데이터를 수집하여 실제 등재된 멀티모달 자료를 조사·분석한 결과를 바탕으로 제시한 수정 방안을 반영하여 새롭게 구성한 분류 체계는 <표 3>과 같다. 수정된 분류 체계의 주요 특징은 다음과 같다.

첫째, 사업의 목적과 데이터 특성을 반영했다. 기존 분류 체계는 유형·무형 자료가 혼재되어 있어 데이터의 성격을 명확히 구분하기 어려웠다. 이에 따라 수정 체계에서는 사진 데이터의 피사체를 1차 기준으로 삼아 유형과 무형을 이원화하여 구조적 명확성을 확보했다. 본 사업은 유형 자료를 중심으로 하지만, 향후 멀티모달 사업 확장 시 무형 자료 포함 가능성을 고려하여 전통놀이·세시풍속 등 무형 문화의 대표 사례를 예시로 제시했다. 또한 이미지·영상·음성·텍스트 등 다양한 매체 유형에 동일한 분류 원리가 적용될 수 있도록 확장성을 고려한 설계를 반영했다.

둘째, 중첩 문제를 해소하기 위한 구조 개선을 진행했다. 기존 분류 체계에서는 일부 항목이 너무 광범위하거나, 또는 중복되어 하나의 데이터가 둘 이상의 분류에 중첩되는 사례가 빈번했다. 예를 들어 ‘의생활-공예’, ‘주생활-건축’은 실제 사진 데이터 해석 과정에서 구분이 모호했다. 이에 따라 수정된 체계는 대분류 단계에서 유사 영역을 통합하고, 중분류 단계에서는 기능·형태 기반으로 세분화하여 분류 기준을 명확화했다. 이로써 데이터가 일관된 대분류 아래에서 관리되면서도 중분류 수준에서는 용도·형태 차이를 구체적으로 반영할 수 있게 됐으며 중복 분류로 인한 혼선을 크게 줄일 수 있었다.

셋째, 본 사업에서 기존에 활용해 온 분류 체계와의 호환성을 유지했다. 수정된 분류 체계는 본 사업에서 기준으로 삼은 ‘한국민족문화대백과사전’ 분류 체계의 구조와 연결될 수 있도록 설계했다. 수정 분류 체계 표의 ‘기존 분류’ 열에는 본 사업에서 활용된 항목을 병기하여 양 체계 간 매핑(mapping)이 가능하도록 했다. 이를 통해 후속 사업에서 새로운 분류 체계를 적용하더라도 기존 데이터와의 연속성을 유지하고, 데이터베이스 전환 과정에서 발생할 수 있는 호환성 문제를 최소화할 수 있다.

2.1.3. 인공지능 학습용 한국어문화 데이터 구축 분류 체계 연구

본 연구에서는 인공지능 학습용 한국어문화 데이터 구축을 위해 기존 국내외 분류 체계를 조사·분석하고, 멀티미디어 데이터의 구조적 특성과 적용 가능성을 검토하여 새로운 ‘매트릭스형 하이브리드 분류 체계’의 방향을 도출했다. 특히 유형(physical form)과 주제적 의미(content theme)를 결합하는 다차원 분류 방식의 필요성을 검증하고, 이를 기반으로 실효성 있는 분류 원칙과 적용 구조를 제안했다. 주요 제안 내용은 다음과 같다.

- 국내외 기존 분류 체계의 구조 및 특성 분석
 - 유네스코 국제 표준, 국가유산청 행정 분류, 한국민족문화대백과사전 등 국내외 문화유산 분류 체계를 비교·분석하여, 멀티미디어 데이터 구축 관점에서의 장단점과 시사점을 도출했다.
- 하이브리드 분류 체계 설계를 위한 기준 정립
 - 기존 분류 방식이 갖는 단일 축(수직적) 구조의 한계를 보완하기 위해, 데이터의 물리적 형태를 구분하는 ‘유형 축’과 의미적 범주를 설명하는 ‘주제 축’을 결합한 다차원 분류 모델의 개념을 정립했다.
 - 텍스트·이미지·영상·음성 등 다양한 매체 형식에 공통적으로 적용 가능한 범용적 분류 기준을 마련했다.
- 제안 분류 체계의 적용 가능성 검증 및 활용 방안 도출
 - 설계된 분류 체계의 실효성을 확인하기 위해 실제 문화유산 데이터에 적용한 사례를 검토했다.
 - 인공지능 학습 데이터 구축, 교육·연구 콘텐츠 개발, 디지털 아카이브 구축 등 다양한 분야에서의 활용 모델과 기대 효과를 제시했다.

□ 분류 체계 제안

제안하는 분류 체계는 멀티모달 자료의 특성과 인공지능 학습 데이터의 활용성을 고려하여, 유형 축, 주제 축, 데이터 메타데이터 축이라는 세 축으로 구성된다. 각 축은 독립적으로 작동하면서도 상호 보완적으로 연결되며, 데이터의 실제·의미·기술적 속성을 동

시에 반영하는 구조를 갖는다.

- 제1축: 유형 축(Type Axis)

데이터가 다루는 대상의 객관적 실체가 무엇인지를 규정하는 계층형 구조이다. 유형 축은 멀티모달 데이터의 기본 정체성을 정의하며, 다음의 네 가지 대분류로 구성된다.

- 인공물(Artifacts): 인간이 제작한 공간·건조물, 예술·공예품, 생활 자료
 - 행위 및 재현(Practices & Performances): 공연예술, 사회적 관습, 기예·기술 등
 - 기록물 및 저작물(Documents & Works): 문헌, 시청각 기록, 언어·문학 자료
 - 자연물 및 경관(Natural Objects & Landscapes): 동식물, 지질, 자연 경관 등
- 유형 축은 실제 자료의 형태·기능·용도를 반영해 데이터 분류의 기본 틀을 제공한다.

- 제2축: 주제 축(Thematic Axis)

데이터가 무엇에 관한 것인지를 나타내는 비계층형 태그 구조이다. 한 데이터에 다수의 주제를 부여할 수 있어 분석·검색의 확장성이 높다. 주요 주제 영역은 다음과 같다.

- 역사 (History): 시대·사건·전쟁 등 (#조선, #일제강점기 등)
 - 인물 (Figures): 관련 인물·집단 (#세종, #이순신 등)
 - 사상 및 종교 (Philosophy & Religion): 가치관·종교·철학 (#불교, #유교 등)
 - 사회 및 제도 (Society & Systems): 제도·생활상·관습 (#과거제도, #신분제 등)
 - 과학 및 기술 (Science & Technology): 기술·지식·원리 (#건축기술, #인쇄술 등)
- 주제 축은 데이터의 의미적 맥락을 보완하는 역할을 수행한다.

- 제3축: 데이터 메타데이터 축(Data Metadata Axis)

데이터 파일 자체가 어떤 형태인지를 기술하는 구조로, 검색·품질관리·활용에 필수적이다. 주요 메타데이터 항목은 다음과 같다.

- 데이터 유형 (Data Type): 텍스트, 이미지, 음성, 영상, 3D 모델 등
 - 파일 형식 (File Format): JPG, PNG, MP4, PDF 등
 - 품질 정보 (Quality Information): 해상도·음질 등 (예: 4K, 48kHz 등)
 - 생성 정보 (Creation Information): 생성일·촬영일·제작자
 - 저작권 정보 (Copyright Information): 공공누리 1유형, CC-BY 등
 - 언어 (Language): 한국어(현대어/고어), 영어 등
- 이 축은 데이터베이스 구축 및 AI 학습 환경과의 호환성을 높인다.

- (예시) 안동 하회탈춤(중요무형문화재 제69호) 공연을 촬영한 4K 영상 데이터

- 유형 축: 행위 및 재현 > 공연 예술 > 연희 (하회별신굿탈놀이)
- 주제 축: #고려, #민속, #무속신앙, #공동체문화
- 메타데이터 축:
 - 데이터 유형: 영상
 - 파일 형식: MP4
 - 품질 정보: 4K(3840x2160)
 - 저작권 정보: 공공누리 1유형
 - 언어: 한국어-현대어(ko-KR)

2.1.4. 범주 우선순위 선정

□ 수집처 분류 체계 비교 및 분석

한국언어문화 관련 자료를 전반적으로 수록하며 분류 체계가 비교적 명확하게 정립된 주요 수집처 6개의 분류 체계를 나열하여 비교·분석했다.

수집처	분류 체계			
공공데이터포털	1. 공공행정 2. 과학기술 3. 교육 4. 교통물류	5. 국토관리 6. 농축수산 7. 문화관광 8. 법률	9. 보건의료 10. 사회복지 11. 산업고용 12. 식품건강	13. 재난안전 14. 재정금융 15. 통일외교안보 16. 환경기상
국립민속박물관 민속아카이브	1. 문화예술 2. 사회생활	3. 산업/생업 4. 의	5. 식 6. 주	7. 종교신앙 8. 기타자료
e뮤지엄	1. 의생활 2. 식생활 3. 주생활 4. 산업/생업	5. 교통/통신 6. 전통과학 7. 사회생활 8. 종교신앙	9. 문화예술 10. 군사 11. 보건의료 12. 과학기술	13. 미디어 14. 기타자료
국립고궁박물관	1. 회화 2. 서예 3. 악기 4. 어보 5. 어책 6. 현판 7. 제구 8. 금속공예	9. 의상 10. 장신구 11. 고문서 12. 전적 13. 조각 14. 가구 15. 도자기 16. 무구	17. 노부 18. 과학기기 19. 치초직 20. 기탁 21. 구입 22. 의궤 23. 궁궐 24. 과학	25. 공예 26. 복식 27. 의장 28. 근대매체 29. 기타 30. 왕실복식 31. 궁중악무 32. 국가제례

	33. 의장	34. 무구	35. 피지초직	36. 실록
세종대학교박물관	1. 장신구 2. 도자기 3. 고고유물	4. 목가구 5. 서화 6. 민속품		
국립경주박물관	1. 의생활 2. 식생활 3. 주생활 4. 산업/생업	5. 교통/통신 6. 전통과학 7. 사회생활 8. 종교신앙	9. 문화예술 10. 군사 11. 보건의료 12. 과학기술	13. 미디어 14. 기타자료

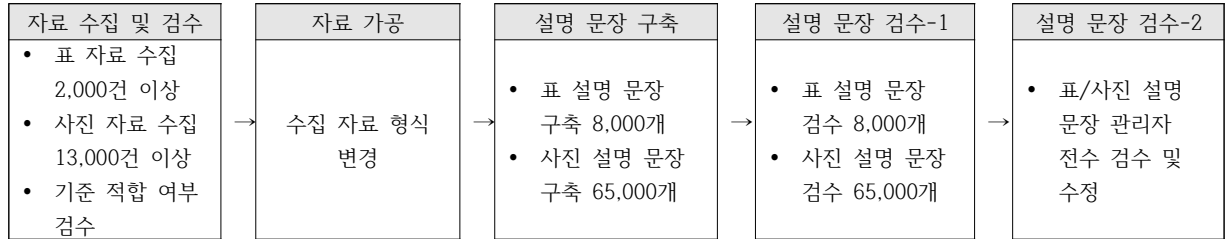
<표 4> 수집처 6개 분류 체계

<표 4>를 살펴보면, ‘식생활’, ‘의생활’, ‘주생활’과 같은 생활 관련 범주가 여러 수집처에서 공통적으로 나타나며, ‘종교신앙’과 ‘사회생활’ 역시 수집처 간 공통적으로 확인된다. 공예품 관련 분류 또한 구체화 수준의 차이는 있으나 전반적으로 유사한 범주 구조를 형성하고 있음을 알 수 있다. 다만 이는 대분류 수준에서의 공통점이며, 중분류와 소분류 단계로 내려가면 기관별 목적과 자료 특성에 따라 세부 범주가 더욱 다양하게 세분화된다.

□ 범주 우선순위 선정

범주 우선순위는 주요 수집처의 분류 체계를 비교·분석한 결과와 자료 수집의 용이성, 설명 문장 작성의 적합성 등을 종합적으로 고려하여 설정했다. ‘식생활’, ‘의생활’, ‘주생활’은 모든 수집처에서 공통적으로 확인되는 핵심 생활문화 범주로, 한국문화 내 중요도와 충분한 데이터 보유량을 고려해 우선 구축 범주로 선정했다. ‘건축’은 대분류·중분류·소분류 전반에서 건물, 건축물, 건축양식, 건축유산 등의 형태로 반복적으로 등장하여 자료 분포가 안정적이고 사진 기반 설명 문장 작성에도 적합한 범주로 판단했다. 또한 ‘공예’는 다수의 수집처에서 공통적으로 확인되며, 한국언어문화의 시각적·물질적 특성을 잘 반영하는 대표적인 예술문화 범주로서 수집 가능성과 설명 가능성이 높아 초기 구축 범주에 포함했다. 이러한 분석을 바탕으로 본 사업에서는 식생활-의생활-주생활-건축-공예 순으로 구축 우선순위를 최종 확정했다.

2.2. 표·사진 자료 수집 및 가공




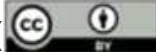
<표 5> 표·사진 자료 구축 전체 과정

본 사업의 말뭉치 구축은 <표 5>와 같은 절차에 따라 단계적으로 수행됐다. 첫 단계는 자료 수집으로, 설명 문장 작성 과정에서 일부 표·사진 자료가 제외될 가능성을 고려하여 최종 구축 목표보다 충분한 여유분을 확보하는 방식으로 수집을 진행했다. 이후 수집된 자료가 작업 기준에 부합하는지 확인하기 위해 관리자 검수 과정을 거쳤으며, 자료 수집 및 검수는 다음과 같은 절차로 이루어졌다.

2.2.1. 자료 수집 및 검수

표 자료 수집 기준 및 수집 방법은 다음과 같다.

□ 표 자료 수집 기본 조건



- 상업적 활용과 변형 등 2차적 저작물 작성이 가능한 것
- 공공누리 제1유형(, CC BY() 부착 문서
- 한국문화와 관련된 통계 수치가 포함된 표
- '24년 발행된 최신 표 자료를 우선으로 연도별 역순 수집
- 행, 열이 2개 이하인 표 제외
- 표 크기 1쪽 이상 제외
- 제목 행 또는 제목 열이 없는 표 제외
- 주제별 균등 분포

□ 표 자료 수집 방안

- 수집처 : 문화센터(<https://stat.mcst.go.kr>), KOSIS국가통계포털(<https://kosis.kr>)
- 수집 방법 :
 - 수집자는 각 수집처에서 원문 표를 엑셀 파일 형태로 다운로드한 뒤, 표의 구조와 내용을 검토한다.
 - 다운로드 과정에서 발생하는 '.0'과 같은 불필요한 수치 표기, 표의 핵심 주제와 관련 없는 전체 통계 행·열 등 잉여 정보는 제거하여 표 내용을 정제한다.
 - 정제된 표를 기반으로 메타데이터를 목록형으로 작성하여 후속 작업이 용이하도록 한다.
- 메타데이터 : 파일 번호, 표 제목, 연도, 출처, URL, 주제, 저작권 표시

사진 자료 수집 기준 및 수집 방법은 다음과 같다.

□ 사진 자료 수집 기본 조건

- 상업적 활용과 변형 등 2차적 저작물 작성이 가능한 것
- 공공누리 제1유형() , CC BY() 부착 문서
- 텍스트가 없고, 객체만 있는 사진 우선 수집
- 한국어언어문화 중 '식생활, 의생활, 주생활, 건축, 공예' 관련 사진
- 권장 해상도 : 150~300 DPI
- 최소 이미지 크기 : 650x650 이상
- 적정 용량 : 10MB 이내
- 주제별 균등 분포

□ 사진 자료 수집 방안

- 수집처
 - 한국민족문화대백과사전(<https://encykorea.aks.ac.kr>)
 - 한국관광공사 포토코리아(<https://phoko.visitkorea.or.kr>)
 - 국립국어원 지역어 종합 정보(<https://dialect.korean.go.kr>)
 - 국립민속박물관 파주 민속아카이브(<https://www.nfm.go.kr/paju/main>)
 - 국립민속박물관 한국민속대백과사전(<https://folkency.nfm.go.kr>)
 - 국립중앙박물관 e뮤지엄(<https://www.emuseum.go.kr>)

- 국가유산청 국가유산디지털서비스(<https://digital.khs.go.kr>)
- 국립국악원 국악아카이브(<https://archive.gugak.go.kr>)
- 한식진흥원(<https://www.hansik.or.kr>)
- 공유마당(<https://gongu.copyright.or.kr/gongu/main/main.do>)
- 데이터리 자체 수집
- 수집 방법
 - 수집자는 각 수집처에서 한국어언어문화와 관련된 다섯 개 범주(식생활·의생활·주생활·건축·공예)에 해당하는 분류를 확인한 뒤, 해당 자료를 대상으로 다운로드 신청 또는 크롤링을 진행한다.
 - 수집된 전체 이미지는 관리자가 전수 검토하며, 저작권 충족 여부, 한국어언어문화 해당 여부, 메타데이터의 명확성, 객체의 식별 가능성, 설명 문장 작성 가능성 등을 기준으로 최종 작업 대상 적합성을 판단한다.
 - 최종적으로 선별된 이미지를 기반으로 메타데이터를 목록형으로 정리하여 이후 설명 문장 작성 및 도구 업로드 과정에서 활용할 수 있도록 한다.
- 메타데이터는 ‘한국민족문화대백과사전’의 형식을 참고하여 다음과 같이 수집한다.
 - 파일 번호, 대상 용어 번호, 대상 용어, 자체 분야 정보, 원자료 번호, 멀티미디어 ID, 저작권, URL, 이미지 제목, 이미지 설명, 저작권자

□ 사진 자료 작업 제외 대상

사진 자료의 경우, 설명 문장 작성에 부적합한 사례들이 다수 확인됐다. 이러한 자료들은 수집 직후 관리자 전수 검토 단계와 구축 과정에서 모두 제외했다.

- 행위 중심 사진 : 인물의 동작이나 활동이 중심이 되어 객체 자체가 주제가 아닌 사진
- 관련 문서·도서 사진 : 관련 문헌, 도서, 간행물 등을 촬영한 사진
- 원재료 사진 : 완성된 문화 객체가 아닌 원재료 또는 재료의 일부만 촬영된 사진
- 중복 사진 : 동일 수집처 내 중복 이미지, 또는 수집처 간 중복 수록된 사진 (예: 한국민족문화대백과사전에 민족아카이브 자료가 재등재된 경우 등)
- 설명 불가 사진 : 객체의 정체성이 불명확하거나 설명 문장 작성이 불가능한 사진
- 관련 없는 사진 : 분류 체계와 관련성이 없거나 의미적으로 부적합한 사진
- 객체의 세부 사진 : 객체의 일부만 촬영되어 전체 형태를 파악할 수 없는 사진

- 각도만 다른 반복 사진 : 동일 객체를 극히 미세한 각도만 달리하여 반복 촬영한 사진
- 흑백 사진 : 색채 정보가 없어 가시적 속성(색채, 재질 등) 기술이 불가능한 사진
- 다른 나라 문화 사진 : 한국 문화와 무관하거나 외국 문화 객체가 촬영된 사진

2.2.2. 자료 가공

수집된 자료는 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>에 업로드 할 수 있는 형태로 가공했다.

표 자료의 경우 원문 표는 개별 xlsx 파일로, 메타데이터는 전체 표를 하나의 목록으로 정리한 xlsx 파일로 수집됐다. 도구 업로드를 위해서는 표 원문과 해당 메타데이터가 하나의 구조로 통합되어야 했기 때문에, 표 1건당 원문(xlsx)과 메타데이터를 결합한 단일 JSON 파일로 변환했다. 실제 변환 구조는 아래 <표 6>과 동일하며, 최종 구축 데이터와 형식도 동일하다. 다만 표 수집 단계에서는 EX 영역 값만 비어 있는 점이 차이점이다. 각 스키마 상세는 ‘2.3.3. 구축 말뭉치 가공’에서 설명했다.

1수준	2수준	3수준	4수준	타입	설명 및 예시
id				str	국립국어원 표 설명 문장 말뭉치 id
metadata				obj	
	title			str	국립국어원 표 설명 문장 말뭉치
	creator			str	국립국어원
	distributor			str	국립국어원
	year			str	2025
	category			str	멀티모달 > 표
	annotation_level			str	설명 문장 작성
	sampling			str	본문 일부
document				obj	
	id			str	TC00001-001/TK00001-001 - T : 표(Table) 데이터 - C/K : 수집처 • C = 문화센터 • K = 국가통계포털 - 00001 : 작업자 번호(1자리, 0으로 시작) + 문서 번호 (4자리, 0001로 시작) - 001 : 문서 내 표 번호(3자리, 001부터 시작)
	metadata			obj	
		title		str	표 제목

		date		str	연도
		publisher		str	저작권
		url		str	표가 포함되어 있는 url 주소
		topic		str	한국언어문화 주제 분류
		unit		str	단위
		copyright		str	출처 표기
	EX			obj	
		reference		obj	
			reference_type	str	table_ref, row_ref, col_ref, cell_ref
			highlighted_cells	arr	설명 영역 셀 위치 값 ex) [0,0], [0,1]....
		exp_sentence		str	작성된 설명 문장
	table			obj	
		value		str	셀 값
		is_header		str	제목 행/열 여부
		row		int	행 위치
		row_span		int	병합 정보
		column		int	열 위치
		column_span		int	병합 정보

<표 6> 표 수집 json 구조

사진 자료 역시 원문 이미지는 개별 JPG 파일로, 메타데이터는 수집처별로 목록화된 xlsx 파일로 수집됐다. 업로드를 위해서는 사진의 메타데이터를 담은 JSON 구조가 필요했기 때문에, 사진 1장(JPG)과 해당 메타데이터 1건을 매칭하여 <표 7>과 같이 사진별 JSON 파일을 생성했다. 사진 단계에서 생성된 JSON 역시 최종 구축 데이터와 동일한 형식을 갖고 있으며, 스키마 구성은 ‘2.3.3. 구축 말뭉치 가공’에서 상세히 기술했다.

1수준	2수준	타입	설명 및 예시
id		str	사진파일명/id (P100001) - P: 사진(Picture) 데이터 - 수집처 번호(2자리, 1~11) + 사진 번호(4자리, 0001로 시작)
metadata		obj	
	term_id	str	대상 용어 번호
	term	str	대상 용어
	domain	str	자체 분야 정보
	source_id	str	원자료번호
	media_id	str	멀티미디어 id
	copyright	str	저작권
	url	str	제공 url
	title	str	캡션(이미지제목)/원시파일명
	note	str	이미지 설명
	publisher	str	이미지 저작권자
	image	str	분석 대상 사진 파일명 (id)

<표 7> 사진 수집 json 구조

2.2.3. 자료 수집 내역

위에서 설명한 방식으로 표·사진 자료를 수집한 뒤, 수집된 원자료를 다시 한번 점검해 실제 구축에 활용할 수 있는 데이터만을 남겼다. 초기에는 충분한 양을 확보하기 위해 폭넓게 수집했으나, 검토 과정에서 중복 자료, 작업 기준에 맞지 않는 자료, 메타데이터가 불완전한 자료 등을 다시 한번 더 선별해 작업 대상에서 제외했다.

이러한 절차를 거쳐 최종 구축에 활용 가능한 표·사진 데이터를 확정했으며, 그 수집 내역은 아래와 같다.

□ 표 수집 내역

• 수집처별

수집처	수량
국가통계포털	1,873
문화센터	359
총합	2,232

• 주제별

주제	수량	주제	수량
경제·산업 > 산업	666	예술·체육 > 국악	15
생활 > 민속·인류	386	정치·법제 > 법제·행정	14
생활 > 식생활	256	언론·출판 > 출판	13
공연·예술 > 기타	223	예술·체육 > 기타	8
사회 > 사회구조	188	교육 > 교육	7
생활 > 주생활	114	예술·체육 > 대중음악	6
생활 > 의생활	100	예술·체육 > 무용	5
예술·체육 > 기타	97	정치·법제 > 정치	4
정치·법제 > 외교	44	예술·체육 > 연극	1
예술·체육 > 건축	40	지리 > 자연지리	1
언어 > 언어·문자	24	과학 > 과학기술	1
언론·출판 > 언론·방송	19	총합	2,232

□ 사진 수집 내역

• 수집처별

수집처	수량	수집처	수량
한국민족문화대백과사전	3,132	국가유산디지털서비스	1,462
포토코리아	4,365	국악아카이브	126
지역어종합정보	310	한식진흥원	693
민속아카이브	1,930	공유마당	197
한국민속대백과사전	105	데이터리	70
e뮤지엄	2,204	총합	14,594

- 주제별

주제	수량
식생활	3,640
의생활	2,997
주생활	3,608
건축	3,247
공예	1,102
총합	14,594

2.3. 표·사진 설명 말뭉치 구축

2.3.1. 구축 방법

2.3.1.1. 표 설명 문장 작업자 선정 및 교육

□ 작업자 선정

표 설명 문장 구축 및 검수 작업자는 다음 기준에 따라 선발했다. 먼저, 작업의 품질을 확보하기 위해 국어국문학 및 언어학 관련 전공(재학 이상 또는 학사 이상)을 보유한 인력을 중심으로 구성했다. 여기에 2024년 표 설명 문장 구축 사업 참여 경험자, 국립국어원 관련 사업 참여 작업자, 데이터리 클라우드소싱 프로젝트 참여 경험자 등을 우선적으로 선발하여 실제 작업에 적합한 인력을 선발했다. 또한, 사전 테스트를 통과한 검증된 작업자만을 포함하여, <표 설명 문장 작성 지침>에 대한 숙지 여부와 실제 작업 역량을 확인했다.

이와 같은 기준을 바탕으로 최종 선발된 작업자는 표 설명 문장 구축 10명, 검수 8명으로 구성됐다.

□ 작업자 교육

본격적인 구축 작업에 앞서 작업자들이 표 설명 영역 선정 방법과 작성 원칙을 정확히 이해할 수 있도록, 작업 시작 1주 전에 사전 지침 교육을 실시했다.

차수	교육 일자	교육 내용
1	2025.07.07	- '표 설명 문장 작성 지침' 내용 교육 - 구글 폼을 이용한 실습 - 멀티 모달 도구 사용 방법 안내 - 기타 제출 서류 (보안각서 포함) 제출 안내

<표 8> 표 설명 문장 구축 교육 내용

교육을 통해 지침 전반에 대한 설명과 함께 구글 폼을 활용한 실습을 통해 작업자의 이해도를 확인했으며, 이후 멀티모달 도구 사용 방법을 상세히 안내했다. 교육은 지침을 수립한 한림대학교 전문가가 진행하여, 작업자들이 지침의 내용을 자세히 익힐 수 있도록 했다. 멀티모달 도구 사용 방법, 작업 일정, 제출 방식 등 운영 절차는 데이터리 관리자가 별도로 안내했다.

2.3.1.2. 사진 설명 문장 작업자 선정 및 교육

□ 작업자 선정

사진 설명 문장 구축 및 검수 작업자는 다음 기준에 따라 선발했다. 먼저, 작업의 품질을 확보하기 위해 데이터리 클라우드소싱 프로젝트 참여 경험자와 사전 테스트 합격자를 중심으로 인력을 구성했다. 또한 사진 설명 문장 작성 지침을 충분히 숙지했는지 여부를 필수 기준으로 적용하여 실제 구축 작업에 적합한 인력을 선발했다.

이와 같은 기준을 바탕으로 선발된 인력은 사진 설명 문장 구축 작업자 25명, 검수 작업자 15명으로 최종 구성됐다.

□ 작업자 교육

사진 설명 문장 구축 작업은 초기 사업으로, 이전 경험이 없는 작업자가 다수였기 때문에 보다 구체적이고 체계적인 교육이 필요했다. 이에 따라 작업자들이 사진 설명 영역 선정 기준과 작성 원칙을 정확히 이해할 수 있도록 작업 시작 1주 전 1차 지침 교육을 실시했으며, 이후 범주 확장(건축·공예)과 작업 품질 피드백을 중심으로 한 2차 교육을 추가로 진행했다. 이와 별도로 작업 기간 동안 수시로 주의사항 및 공지 사항을 공유하여 작업 방향성을 지속적으로 안내했다.

차수	교육 일자	교육 내용
1	2025.07.14	- '사진 설명 문장 작성 지침' 내용 교육 - 구글 폼을 이용한 실습 - 멀티 모달 도구 사용 방법 안내 - 기타 제출 서류 (보안각서 포함) 제출 안내
2	2025.09.11	- 범주 추가 (건축, 공예) 내용 교육 - 구글 폼을 이용한 실습 - 구축·검수 시 주의사항 안내 - 오류 사례 중심의 피드백 제공

<표 9> 사진 설명 문장 구축 교육 내용

1차 교육은 '사진 설명 문장 작성 지침'을 수립한 한림대학교 전문가가 직접 진행하여, 작업자들이 지침의 구조와 작성 원칙을 충분히 이해할 수 있도록 내용을 상세히 설명했다. 멀티모달 도구 사용 방법, 작업 일정, 제출 방식 등은 데이터리 관리자가 별도로 안내했다. 또한 구글 폼 실습을 통해 작업자의 지침 이해도를 확인했으며, 멀티모달 도구 시연과 제출 과정 안내를 통해 실제 구축 과정에서 필요한 작업 과정을 상세히 안내했다.

2차 교육에서는 새롭게 포함된 건축·공예 범주에 대한 설명과 함께 작업물의 오류 사례를 기반으로 한 보완 교육을 진행하여 작업 품질을 한층 강화했다. 작업물의 대표적인 오류 사례는 다음과 같다.

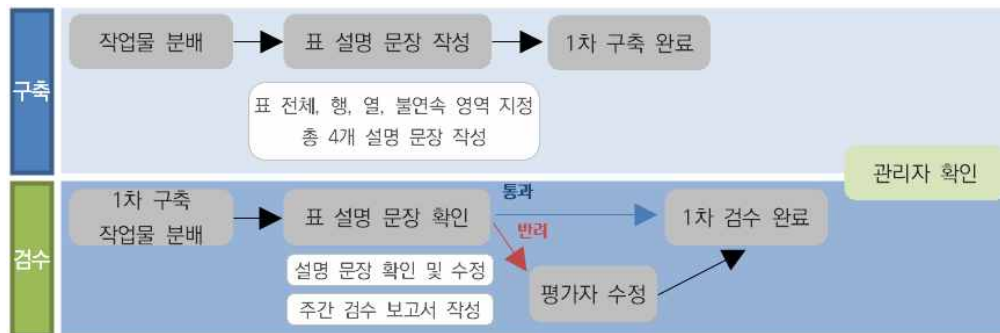
구분	오류 유형	예제
1	띄어쓰기 오류 및 오타자	- (오류) 이 어묵의 색채는 전체적으로 노릇한 갈색 빛을 띠며, 기름에 튀겨진 표면이 반들반들 하게 빛난다. - (수정) 이 어묵의 색채는 전체적으로 노릇한 갈색 빛을 띠며, 기름에 튀겨진 표면이 반들반들하게 빛난다.
2	비문 및 문장 내 주술 호응 오류	- (오류) 이 애호박볶음은 반달 모양의 애호박과 당근과 양파는 길게 채 썬 모양이다. - (수정) 이 애호박볶음의 애호박은 반달 모양이며, 당근과 양파는 길게 채 썬 모양이다. - (오류) 이 대바구니의 소재는 가늘게 쪼갠 대나무를 얹은 것이다. - (수정) 이 대바구니의 소재는 가늘게 쪼갠 대나무이다.

3	조사 오류	<ul style="list-style-type: none"> - (오류) 이 초가집은 벽은 흙으로, 기둥과 서까래, 문은 나무, 지붕은 짚으로 지어졌고, 집은 에워싸고 있는 담은 돌을 쌓아 만들었다. - (수정) 이 초가집의 벽은 흙으로, 기둥과 서까래, 문은 나무, 지붕은 짚으로 지어졌고, 집을 에워싸고 있는 담은 돌을 쌓아 만들었다.
4	속성 선정 오류	<ul style="list-style-type: none"> - (오류) [취식 방법] 송이밥은 송이버섯을 먹기 좋게 썰어 볶은 후, 지어진 밥과 함께 섞어 만든다. - (수정) [조리법] 송이밥은 송이버섯을 먹기 좋게 썰어 볶은 후, 지어진 밥과 함께 섞어 만든다.
5	가시적/비가시적 문장 구분	<ul style="list-style-type: none"> - (오류) 추어탕의 재료는 미꾸라지, 시래기, 팽이버섯, 고춧가루 등이다. -> 가시적 속성 중 재료에 대한 설명으로 지시 대명사 필요 - (수정) 이 추어탕의 재료는 미꾸라지, 시래기, 팽이버섯, 고춧가루 등이다.
6	반복적으로 동일한 구조의 문장을 사용한 경우	<ul style="list-style-type: none"> - (오류) A 나막신 : 이 나막신은 나무를 파내어 만든 직사각형 모양으로, 신발 코가 위로 솟아있는 형태이다. B 나막신 : 이 나막신은 나무를 파내어 만든 직사각형 모양으로, 신발 코가 위로 솟아있는 형태이다. - (수정) A 나막신 : 이 나막신은 나무를 파내어 만든 직사각형 모양으로, 신발 코가 위로 솟아있는 형태이다. B 나막신 : 이 나막신은 나무를 깎아낸 직사각형 모양이며, 신발 앞부분이 위로 올라간 모습이다.

<표 10> 사진 설명 문장 작성 시 자주 발생하는 오류

2.3.1.3. 표 설명 문장 구축 방법

표 설명 문장 말뭉치 구축 작업은 표 설명 문장 작성과 이에 대한 교차 검수 및 수정 작업이 순서대로 진행됐다. 전체적인 작업 과정은 아래 [그림 1]과 같다.



[그림 1] 표 설명 문장 작성 및 검수 절차

구축 작업자들은 매주 20건의 표를 배분받아 원문 내용을 충분히 검토한 뒤, ‘표 전체, 행, 열, 불연속 영역’ 네 가지 설명 영역을 각각 모두 선정하고, 각 영역별로 1개의 설명 문장을 작성했다. 작업자들은 표의 구조와 의미를 분석하며 중요 셀을 파악한 후, 지침에서 제시한 방식을 활용하여 설명 문장을 생성했다. 설명 영역 선정 및 문장 작성 방식에 관한 세부 절차는 3장에서 다시 기술하며, 부록의 ‘표 설명 문장 작성 지침’에서도 상세히 확인할 수 있다.

작업 과정에서 발생한 질의는 관리자에게 1:1로 문의하거나 구글 시트를 이용해 질의를 남긴 뒤, 지침 작성자로부터 직접 답변을 받아 해결했으며, 작업자는 구축 대상이 아닌 표를 선별하여 제외함으로써 작업물 검수 또한 함께 진행했다.

1차 구축이 완료된 후, 다음 주부터 검수 작업을 진행했다. 검수 작업자는 매주 25건씩 검수물을 배분받아, 작성 지침 준수 여부, 오타자 및 띄어쓰기 오류, 문장 구조 적절성, 수치 해석의 정확성 등을 전수 확인한 뒤 검수물을 승인하거나, 오류가 있을 경우 직접 수정하여 제출했다. 수정된 작업물에는 반드시 수정 사유를 기재하도록 했다.

검수 과정에서 검수자들의 성실도와 검토 품질을 확인하기 위해 검수 보고서를 별도로 제출받았으며, 초기에는 전체적 평가 중심으로 작성했으나 이후에는 작업물 개별 평가 방식으로 강화하여 말뭉치 품질 관리를 강화했다.

1차 검수까지 완료된 작업물은 관리자가 2차 검수를 통해 재확인했으며, 주간 수합 이후 국립국어원의 검토 의견을 반영하여 전체 작업물에 대해 전수 검토를 실시함으로써 최종 말뭉치의 품질을 체계적으로 관리했다.

2.3.1.4. 사진 설명 문장 구축 방법

사진 설명 문장 말뭉치 구축 작업 또한 사진 설명 문장 작성과 이에 대한 교차 검수 및 수정 작업이 순서대로 진행됐다. 전체적인 작업 과정은 아래 [그림 2]와 같다.



[그림 2] 사진 설명 문장 작성 및 검수 절차

사진 설명 문장 구축 작업은 매주 30건 이상의 사진을 배분했다. 작업자들은 우선 각 사진이 작업 대상 기준에 적합한지, 선택된 카테고리 및 메타데이터가 정확한지를 확인했다. 사진 수집 과정의 특성상 작업 대상이 아닌 이미지가 다수 포함되어 있어 작업자들이 선별하여 제외했으며, 관리자가 제외된 사진을 전수로 재확인하여 재작업이 가능한 경우에 한해 다시 배포했다.

작업자들은 사진의 주요 객체를 규정하는 대상 식별 문장 1개를 작성한 뒤, 객체의 가시적 정보를 중심으로 분류별 속성 설명 문장 4개를 작성했다. 속성 설명은 형태·색채·재질 등 눈으로 확인 가능한 정보를 우선하며, 필요할 경우 최대 2개의 비가시적 정보를 추가하여 객체의 특성을 보완하도록 했다. 작성 방식과 세부 절차는 3장에서 상세히 다루며, 부록의 ‘사진 설명 문장 작성 지침’에서 구체적으로 확인할 수 있다.

작업 과정에서 발생한 질의는 관리자에게 1:1 문의하거나 구글 시트를 통해 공유했고, 지침 작성자가 직접 답변을 제공했다. 이번 사업은 초기 구축 사업으로 참고 가능한 선행 사례가 부족했기 때문에, 특정 객체의 작업 대상 여부와 같은 기본적 판단에 대한 질의가 비교적 많이 발생했다.

1차 구축이 완료된 이후 차주부터 검수 작업을 진행했다. 검수 작업자들은 매주 55건 이상의 검수 대상을 배분받아 카테고리 지정의 적절성, 지침 준수 여부, 오탈자 및 띄어쓰기 오류, 문장 구조의 적절성, 설명 문장의 완성도, 내용 보완 필요 여부 등을 전수 확인했다. 오류가 발견될 경우 검수자가 직접 수정하여 제출했으며, 모든 수정 건에는 반드시 수정 사유를 기재하도록 하여 검수 과정의 신뢰성을 확보했다.

검수 품질 관리를 위해 검수자에게는 별도의 검수 보고서를 제출하도록 했다. 초기에는 작업물에 대한 전체적 평가 중심으로 보고서를 작성했으나, 이후에는 작업물 개별 평가 방식으로 전환하여 말뭉치 품질을 더욱 정밀하게 관리했다.

1차 검수까지 완료된 작업물은 관리자가 2차 검수를 통해 다시 한번 확인했으며, 주간 수합 후에는 국립국어원의 검토 의견을 반영해 전체 작업물을 전수 재검토함으로써 사진 설명 말뭉치의 전반적인 품질을 체계적으로 관리했다.

2.3.2. 구축 도구

2.3.2.1. 표 설명 문장 구축 도구

본 사업의 표 설명 문장 작성 작업에는 국립국어원이 보유한 웹 기반 도구인 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>를 사용했다. 해당 도구는 표 데이터를 시각적으로 확인하면서 설명 영역을 지정하고, 각 영역에 대응하는 설명 문장을 작성·검수할 수 있도록 구성되어 있다.



[그림 3] 표 작업 프로젝트 생성 화면

작업자는 도구 접속 후 자신에게 할당된 프로젝트를 먼저 확인한다. 관리자가 주차별로 배포한 작업만 수행해야 하며, 자신의 과업에 맞는 (구축 또는 검수) 프로젝트를 할당 받았는지 확인한다.



[그림 4] 표 설명 문장 구축 도구 전체 화면

구축 작업 화면은 [그림 4]와 같으며, 화면은 크게 좌측 영역 지정 메뉴, 중앙 표 화면, 우측 설명 문장 입력창으로 구성되어 있다. 좌측 아이콘 메뉴는 설명 영역 지정을 위한 기능으로 위에서부터 불연속 영역 지정, 행 영역 지정, 열 영역 지정, 표 전체 지정, 지정 영역 전체 삭제, 표 화면 이동·확대·축소 기능 순서로 배치되어 있다. 중앙에는 실제 표 데이터가 표시되며 하단에는 표 제목, 출처, 단위 등 주요 메타데이터가 함께 제공되어 설명 문장 작성 시 참고할 수 있다. 우측 입력창은 총 4개의 작업 목록 생성칸과 설명 문장 입력칸으로 구성되어 있다.



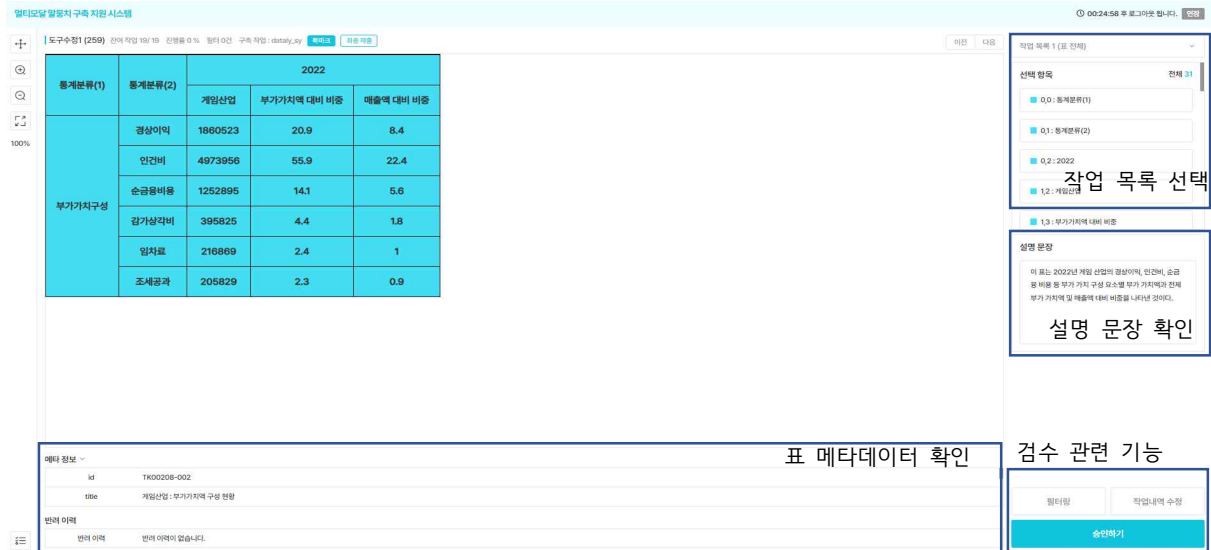
[그림 5] 표 작업 목록 추가 방법

작업자는 우측 상단의 ‘작업 목록 추가’ 기능을 이용해 총 4개의 작업 목록을 생성한 뒤, 각 목록에 해당하는 설명 영역을 순서대로 지정한다. [그림 5]의 2번처럼 표에서 특정 영역을 선택하면 해당 부분이 파란색으로 표시되며, 3번과 같이 선택된 영역 정보가 자동으로 작업 목록과 연동된다. 이후 4번과 같이 네 가지 설명 영역(표 전체·행·열·불연속 영역) 모두가 작업 목록에 정상적으로 추가·지정됐는지 확인한 뒤 설명 문장 작성 단계로 넘어간다.



[그림 6] 표 설명 문장 작성 방법

작업 목록을 모두 생성한 뒤에는 표의 내용과 하단에 제공되는 표 정보를 참고하여 우측 입력란에 설명 문장을 작성한다. 설명 문장 작성을 완료하면 ‘제출하기’를 눌러 작업을 마무리하며, 표의 구조가 불량하거나 값이 지나치게 부적절해 설명 문장을 작성하기 어렵다고 판단되는 경우에는 ‘필터링’ 기능을 사용해 해당 작업을 제외할 수 있다.



[그림 7] 표 설명 문장 검수 도구 전체 화면

검수 과정 역시 구축 화면과 유사한 구성으로 이루어졌으며, 작성된 설명 문장을 확인하고 수정·승인하는 기능이 제공된다. 검수 작업자는 우측 상단에서 작업 목록을 선택해 각 문장을 확인하고 지침 준수 여부, 설명의 타당성, 문장 구조, 오타자 등의 항목을 검토했다.



[그림 8] 표 설명 문장 검수 및 수정 방법

문제가 있는 경우 ‘작업내역 수정’ 기능을 사용해 직접 문장을 수정하고 제출했으며,

이때 반드시 수정 사유를 함께 기록하도록 했다. 수정 사유는 ‘오타자 오류, 지침과 다르게 작성됨, 설명 영역 지정 오류’ 등과 같이 간단하게 기술한다.

표 자체에 문제가 있는 경우에는 ‘필터링’을 통해 작업을 제외할 수 있었으며 문제가 없는 작업물은 ‘승인하기’를 통해 검수 절차를 마무리했다. 이후 관리자는 2차 검수를 통해 전체 작업물을 다시 한번 확인했으며, 주간 수합 후에는 국립국어원의 검토 의견을 반영해 전체 작업물을 전수 재검토함으로써 구축된 말뭉치의 품질을 체계적으로 관리했다.

2.3.2.2. 사진 설명 문장 구축 도구

사진 설명 문장 작성 구축에는 표 작업과 동일하게 국립국어원이 보유한 웹 기반 도구인 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>를 사용했다.



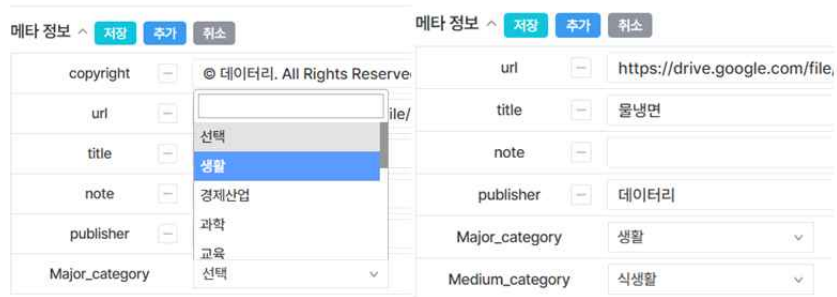
[그림 9] 사진 작업 프로젝트 생성 화면

작업자는 도구 접속 후 자신에게 할당된 프로젝트를 먼저 확인하며, 관리자가 공지한 주차의 작업만을 진행해야 한다. 또한 구축 작업자는 구축 화면에서, 검수 작업자는 검수 화면에서 작업이 이루어져야 하므로, 배정된 작업 유형을 반드시 확인해야 한다.



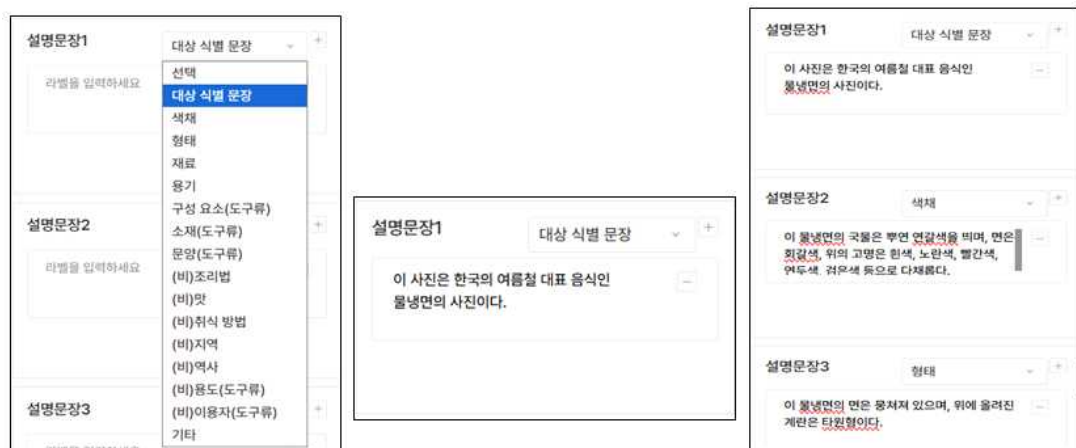
[그림 10] 사진 설명 문장 구축 도구 전체 화면

구축 작업 화면은 [그림 10]과 같다. 사진 데이터의 경우 하나의 이미지 안에 객체가 여러 개 등장하는 사례가 있기 때문에, 설명의 대상이 되는 객체를 지정하기 위한 바운딩 기능이 필요하다. 이를 위해 좌측에는 바운딩 박스를 생성·삭제하는 기능과 사진 이동, 확대, 축소 기능이 제공된다. 화면 중앙에는 원본 사진이 출력되며, 하단에는 사진의 제목·출처·분류 등 기본 정보가 표시된다. 우측 영역에는 속성을 선택하는 메뉴와 설명 문장을 작성하는 입력란이 배치되어 있다.



[그림 11] 사진 카테고리 지정 방법

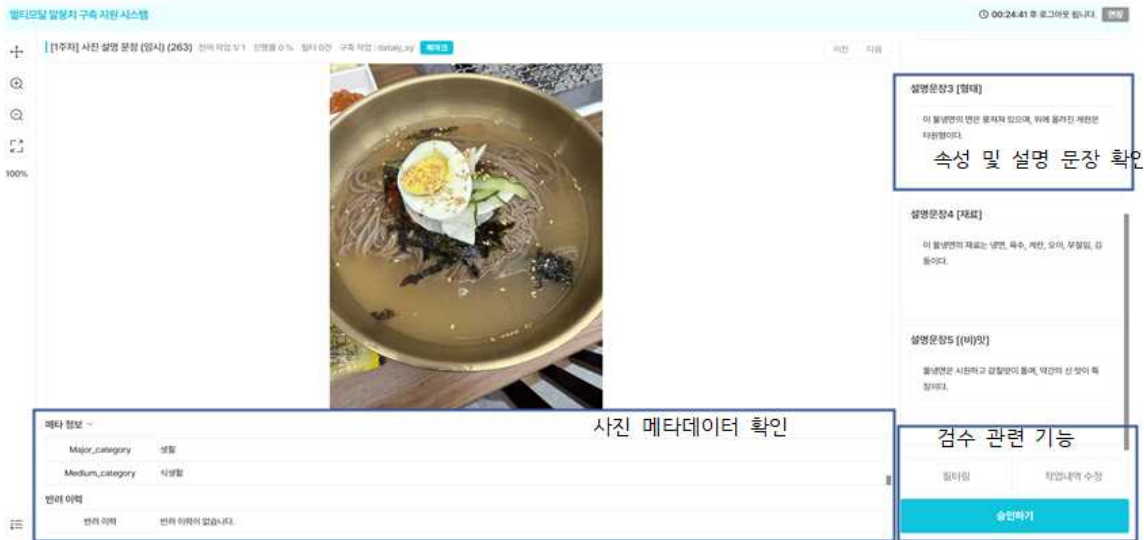
작업자는 우선 사진의 메타데이터가 정확한지 확인한 뒤, 사진에 적용할 대분류 (Major_category)와 중분류(Medium_category)를 지정한다. 분류 선택은 드롭다운 형태로 제공되며, 적절한 항목을 선택한 후 수정 정보를 저장한다.



[그림 12] 사진 설명 문장 작성 방법

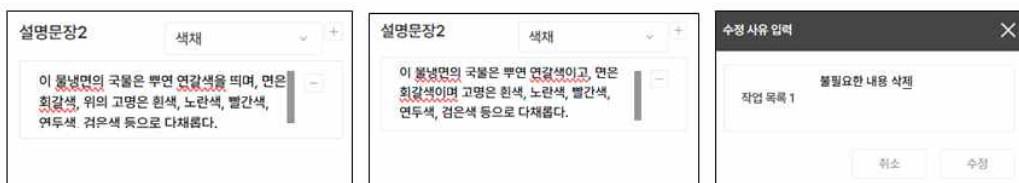
이후 화면 우측에서 설명 문장을 작성할 속성을 선택하고, 해당 속성에 따라 문장을 기

술한다. 작업자는 사진의 객체에 대해 정의하는 대상 식별 문장 1개와 속성에 기반한 설명 문장 4개를 작성해야 하며, 비가시적 속성은 최대 2개까지만 작성할 수 있다. 모든 문장 작성이 완료되면 ‘제출하기’를 눌러 작업을 마무리하며, 사진이 작업 기준에 부합하지 않는 경우에는 ‘필터링’ 기능을 사용하여 해당 작업을 제외할 수 있다.



[그림 13] 사진 설명 문장 검수 도구 전체 화면

검수 작업 화면은 [그림 13]과 같다. 검수자는 화면 중앙에서 사진과 메타데이터를 확인하고 대분류·중분류가 적절하게 지정됐는지 먼저 확인한다. 화면 우측에서 작성된 모든 사진 설명 문장을 검토하여 오타자 및 띄어쓰기 오류, 문장 구조 오류, 반복적 문장 구조 사용 여부, 비가시적 속성이 2개를 초과하여 작성된 경우, 사진 내용과 설명 문장 사이의 불일치 여부 등을 전반적으로 확인한다. 오류가 확인되면 ‘작업내역 수정’을 통해 직접 문장을 수정하여 제출하며, 설명이 적절하게 작성된 경우에는 ‘승인하기’를 눌러 검수를 완료한다. 구축된 사진의 경우에도 작업 대상 기준에서 벗어나는 사례가 발견될 수 있으므로, ‘필터링’을 이용해 제외 처리한다.



[그림 14] 사진 설명 문장 검수 및 수정 방법

2.3.3. 구축 말뭉치 가공

<국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>에서 작업한 결과물은 기본적으로 JSON 형태로 내려받을 수 있다. 그러나 도구에서 제공하는 기본 JSON 구조는 국립국어원 최종 납품 규격과 일부 차이가 있었기 때문에, 데이터리에서 제출 규격에 맞도록 JSON을 재구성하는 후처리 작업을 별도로 수행했다.

표 구축 자료의 경우, HTML 형태의 파일 제출이 필수였기 때문에, 구축된 표 데이터를 HTML로 변환하는 추가 작업을 진행했다. 이를 통해 표 원문과 설명 문장이 모두 포함된 형태의 최종 결과물을 생산했다. 최종 구축물의 제출 형식은 다음과 같다.

□ 표 구축 말뭉치 json 구조

```
{
  "id": "GTEX2502512110",
  "metadata": {
    "title": "국립국어원 표 설명 문장 말뭉치",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2025",
    "category": "멀티모달 > 표",
    "anntation_level": "설명 문장 작성",
    "sampling": "본문 일부"
  },
  "document": [
    {
      "id": "TC00003-002",
      "metadata": {
        "title": "국제문화교류 사업(활동)을 하는 주된 목적-종사자 규모별",
        "date": "2022",
        "publisher": "문화센터",
        "url": "https://stat.mcst.go.kr/portal/subject/subject01/STBL-1029902/stat/STBL-1030076",
        "topic": "경제·산업 > 산업",
        "unit": "개, %"
      },
      "copyright": "「국제문화교류실태조사」, 문화센터"
    },
    {
      "EX": [
        {
          "reference": {
            "reference_type": "table_ref",
            "highlighted_cells": [
              [0.0],[0.1],[0.2],
              [1.2],[1.3],[1.4],[1.5],[1.6],[1.7],[1.8],[1.9],[1.10],
              [2.0],[2.1],[2.2],[2.3],[2.4],[2.5],[2.6],[2.7],[2.8],[2.9],[2.10],
              [3.0],[3.1],[3.2],[3.3],[3.4],[3.5],[3.6],[3.7],[3.8],[3.9],[3.10],
              [4.1],[4.2],[4.3],[4.4],[4.5],[4.6],[4.7],[4.8],[4.9],[4.10],
              [5.1],[5.2],[5.3],[5.4],[5.5],[5.6],[5.7],[5.8],[5.9],[5.10],
              [6.1],[6.2],[6.3],[6.4],[6.5],[6.6],[6.7],[6.8],[6.9],[6.10],
              [7.1],[7.2],[7.3],[7.4],[7.5],[7.6],[7.7],[7.8],[7.9],[7.10]
            ]
          },
          "exp_sentence": "이 표는 2022년 국제 문화 교류 사업(활동)을 하는 주된 목적을 사업체의 종사자 규모별로 나타낸 표이다."
        },
        {
          "reference": {
            "reference_type": "row_ref",
            "highlighted_cells": [
              [3.0],[3.1],[3.2],[3.3],[3.4],[3.5],[3.6],[3.7],[3.8],[3.9],[3.10],
            ]
          }
        }
      ]
    }
  ]
}
```

```

    ],
    {
      "exp_sentence": "2022년 국제 문화 교류 사업(활동)의 주된 목적에 대한 응답 결과를 보면 종사자 규모 1~9명 사업체는 '정보·지식 공유 혹은 전문 역량 강화'가 28.1%로 가장 높았고 '한국·지역에 대한 친밀감 제고 및 이미지 개선'이 15.5%, '한국 문화를 발전시키고 국제 사회에 널리 알리는 것'이 15.2%로 그 뒤를 이었다."
    },
    {
      "reference": {
        "reference_type": "col_ref",
        "highlighted_cells": [
          [0,2],[1,4],[2,4],[3,4],[4,4],[5,4],[6,4],[7,4]
        ]
      },
      "exp_sentence": "2022년 국제 문화 교류 사업(활동)의 주된 목적이 '정보·지식 공유 혹은 전문 역량 강화'라고 응답한 비율을 종사자 규모별로 살펴보면 100~299명 사업체가 35.4%로 가장 높았고 10~49명 사업체가 34.3%, 50~99명 사업체가 29.7%로 그 뒤를 이었다."
    },
    {
      "reference": {
        "reference_type": "cell_ref",
        "highlighted_cells": [
          [6,5],[7,5],[4,6],[7,6]
        ]
      },
      "exp_sentence": "2022년 국제 문화 교류 사업(활동)의 주된 목적 중 '문화교류를 통한 경제 활동 활성화'와 '한국 문화를 발전시키고 국제사회에 널리 알리는 것'을 종사자 규모별로 비교해보면 가장 높은 응답은 각각 100~299명이 18.4%, 10~49명이 25.3%로 후자가 6.9%p 더 높았으며, 가장 낮은 응답은 두 항목 모두 300명 이상에서 1.4%로 동일했다."
    }
  ],
  "table": [
    [
      {
        "value": "구분",
        "is_header": "true",
        "row": 1,
        "row_span": 2,
        "column": 1,
        "column_span": 1
      },
      {
        "value": "COL2",
        "is_header": "true",
        "row": 1,
        "row_span": 2,
        "column": 2,
        "column_span": 1
      },
      {
        "value": "2022",
        "is_header": "true",
        "row": 1,
        "row_span": 1,
        "column": 3,
        "column_span": 9
      }
    ]
  ],
  [
    [
      {
        "value": "사례수",
        "is_header": "true",
        "row": 2,
        "row_span": 1,
        "column": 3,
        "column_span": 1
      },
      {
        "value": "한국/지역에 대한 친밀감 제고·이미지 개선",
        "is_header": "true",
        "row": 2,
        "row_span": 1,
        "column": 4,
        "column_span": 1
      }
    ]
  ]
]

```

```

{
  "value": "정보/지식 공유 혹은 전문역량 강화",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 5,
  "column_span": 1
},
{
  "value": "문화교류를 통한 경제 활성화(이익 증대)",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 6,
  "column_span": 1
},
{
  "value": "문화예술인 상호교류를 통한 창작활동 활성화",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 7,
  "column_span": 1
},
{
  "value": "한국 문화를 발전시키고 국제사회에 널리 알리는 것",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 8,
  "column_span": 1
},
{
  "value": "문화를 통한 상호이해 증진",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 9,
  "column_span": 1
},
{
  "value": "해당사항없음",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 10,
  "column_span": 1
},
{
  "value": "기타",
  "is_header": "true",
  "row": 2,
  "row_span": 1,
  "column": 11,
  "column_span": 1
}
],
[
{
  "value": "전체",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 1,
  "column_span": 1
},
{
  "value": "-",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 2,
  "column_span": 1
}

```

```

},
{
  "value": "5932",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 3,
  "column_span": 1
},
{
  "value": "15.5",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 4,
  "column_span": 1
},
{
  "value": "28.7",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 5,
  "column_span": 1
},
{
  "value": "12",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 6,
  "column_span": 1
},
{
  "value": "15.5",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 7,
  "column_span": 1
},
{
  "value": "15",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 8,
  "column_span": 1
},
{
  "value": "10.2",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 9,
  "column_span": 1
},
{
  "value": "1.5",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 10,
  "column_span": 1
},
{
  "value": "1.7",
  "is_header": "false",
  "row": 3,
  "row_span": 1,
  "column": 11,
  "column_span": 1
}
}

```

```

    ],
    [
      {
        "value": "중사업자 규모",
        "is_header": "false",
        "row": 4,
        "row_span": 5,
        "column": 1,
        "column_span": 1
      },
      {
        "value": "1~9명",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 2,
        "column_span": 1
      },
      {
        "value": "5113",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 3,
        "column_span": 1
      },
      {
        "value": "15.5",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 4,
        "column_span": 1
      },
      {
        "value": "28.1",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 5,
        "column_span": 1
      },
      {
        "value": "12.9",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 6,
        "column_span": 1
      },
      {
        "value": "14.9",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 7,
        "column_span": 1
      },
      {
        "value": "15.2",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 8,
        "column_span": 1
      },
      {
        "value": "10.9",
        "is_header": "false",
        "row": 4,
        "row_span": 1,
        "column": 9,
        "column_span": 1
      }
    ]
  ]

```

```

    },
    {
      "value": "1.4",
      "is_header": "false",
      "row": 4,
      "row_span": 1,
      "column": 10,
      "column_span": 1
    },
    {
      "value": "1",
      "is_header": "false",
      "row": 4,
      "row_span": 1,
      "column": 11,
      "column_span": 1
    }
  ],
  [
    {
      "value": "10~49명",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 2,
      "column_span": 1
    },
    {
      "value": "547",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 3,
      "column_span": 1
    },
    {
      "value": "12.5",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 4,
      "column_span": 1
    },
    {
      "value": "34.3",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 5,
      "column_span": 1
    },
    {
      "value": "6.3",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 6,
      "column_span": 1
    },
    {
      "value": "25.3",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 7,
      "column_span": 1
    },
    {
      "value": "10.9",
      "is_header": "false",
      "row": 5,
      "row_span": 1,
      "column": 8,

```

```

"column_span": 1
},
{
"value": "4.1",
"is_header": "false",
"row": 5,
"row_span": 1,
"column": 9,
"column_span": 1
},
{
"value": "0.9",
"is_header": "false",
"row": 5,
"row_span": 1,
"column": 10,
"column_span": 1
},
{
"value": "5.7",
"is_header": "false",
"row": 5,
"row_span": 1,
"column": 11,
"column_span": 1
}
],
[
{
"value": "50~99명",
"is_header": "false",
"row": 6,
"row_span": 1,
"column": 2,
"column_span": 1
},
{
"value": "85",
"is_header": "false",
"row": 6,
"row_span": 1,
"column": 3,
"column_span": 1
},
{
"value": "14.3",
"is_header": "false",
"row": 6,
"row_span": 1,
"column": 4,
"column_span": 1
},
{
"value": "29.7",
"is_header": "false",
"row": 6,
"row_span": 1,
"column": 5,
"column_span": 1
},
{
"value": "2.2",
"is_header": "false",
"row": 6,
"row_span": 1,
"column": 6,
"column_span": 1
},
{
"value": "15.2",
"is_header": "false",
"row": 6,
"row_span": 1,

```

```

        "column": 7,
        "column_span": 1
    },
    {
        "value": "23.8",
        "is_header": "false",
        "row": 6,
        "row_span": 1,
        "column": 8,
        "column_span": 1
    },
    {
        "value": "5.4",
        "is_header": "false",
        "row": 6,
        "row_span": 1,
        "column": 9,
        "column_span": 1
    },
    {
        "value": "3.1",
        "is_header": "false",
        "row": 6,
        "row_span": 1,
        "column": 10,
        "column_span": 1
    },
    ... 중략 ...
    {
        "value": "106",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 3,
        "column_span": 1
    },
    {
        "value": "28.9",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 4,
        "column_span": 1
    },
    {
        "value": "23.4",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 5,
        "column_span": 1
    },
    {
        "value": "1.4",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 6,
        "column_span": 1
    },
    {
        "value": "1.4",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 7,
        "column_span": 1
    },
    {
        "value": "17.5",
        "is_header": "false",
        "row": 8,
        "row_span": 1,

```

```

        "column": 8,
        "column_span": 1
    },
    {
        "value": "13.8",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 9,
        "column_span": 1
    },
    {
        "value": "4.6",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 10,
        "column_span": 1
    },
    {
        "value": "9.2",
        "is_header": "false",
        "row": 8,
        "row_span": 1,
        "column": 11,
        "column_span": 1
    }
    ]
}

```

맨 상단의 'id'는 국립국어원 표 설명 문장 말뭉치 전체의 기본 정보를 담고 있다.

'document'의 'id'는 각 표의 고유 식별자이며, 표 고유 id는 TC00001-001 / TK00001-001과 같은 형식으로 구성됐다. id의 첫 글자인 T는 해당 데이터가 표(Table) 자료임을 의미하며, 뒤에 오는 C 또는 K는 수집처를 구분하는 값으로, C는 문화센터, K는 국가통계포털을 나타낸다. 이후 이어지는 다섯 자리 숫자 00001은 앞의 한 자리가 작업자 번호(0으로 시작)를, 뒤의 네 자리가 문서 번호(0001부터 시작)를 의미한다. 마지막 세 자리 001은 해당 문서 안에서의 표 번호를 나타내며 001부터 순차적으로 부여됐다.

'metadata'에는 표 자체에 대한 상세 정보가 포함됐다. 'title'은 표의 제목이고, 'date'는 표 발행 연도, 'publisher'는 발행 기관을 나타내며, 'url'에는 원문 문서를 확인할 수 있는 링크 정보가 포함됐다. 'topic'은 표의 주제 분류를, 'unit'은 표에서 사용되는 단위 정보를 나타내며, 'copyright'는 출처를 나타낸다.

'EX'에는 설명 문장 관련 정보가 정리됐다. 여기서 'reference_type'은 설명 문장의 유형을 나타내며, 'table_ref'는 표 전체, 'row_ref'는 행, 'col_ref'는 열, 'cell_ref'는 불연속 영역을 의미한다. 'highlighted_cells'은 설명 영역으로 지정된 셀의 좌표 정보를 담고 있으며, 이는 멀티모달 구축 도구의 규칙에 따라 0부터 시작하는 좌표값(0-base)으로 기록됐다. 'exp_sentence'에는 해당 영역을 기반으로 작성한 설명 문장이 포함됐다.

'table'에는 표의 구조와 셀 값이 순차적으로 기록됐다. 표의 셀 정보는 수집 단계에서

확보한 원자료를 기준으로 하며, 이때 셀의 위치는 1부터 시작하는 좌표값(1-base)으로 기록됐다. 수집된 표의 base(1-base)와 설명 영역 지정 값의 base(0-base)가 서로 다르지만 이는 멀티모달 구축 도구의 입력 구조와 일치하기 때문에 문제없이 활용할 수 있다. 'value'는 셀의 실제 값을 의미하고, 'is_header'는 해당 셀이 표의 헤더인지 여부를 나타낸다. 'row'는 행 번호, 'row_span'은 행 확장 범위, 'column'은 열 번호, 'column_span'은 열 확장 범위를 각각 의미한다.

위 표 구축 json을 html로 변환하면 다음과 같은 형식이다.

□ 표 구축 말뭉치 html 구조

id		TC00003-002									
metadata	title	국제문화교류 사업(한동)을 하는 주된 목적 중사자 규모별									
	date	2022									
	publisher	문화재단									
	url	https://stat.mcst.go.kr/portal/subject/subject01/STBL-1029602/star/STBL-1030076									
	topic	경제-산업 > 산업									
	unit	개, %									
	copyright	「국제문화교류실태조사」, 문화재단									
EX	reference_type	table ref									
	highlighted_cells										
	exp_sentence	이 표는 2022년 국제 문화 교류 사업(한동)을 하는 주된 목적을 사업체의 중사자 규모별로 나타낸 표이다.									
	reference_type	row ref									
	highlighted_cells	highlighted_cells (3, 0)									
		highlighted_cells (3, 1)									
		highlighted_cells (3, 2)									
		highlighted_cells (3, 3)									
		highlighted_cells (3, 4)									
		highlighted_cells (3, 5)									
	highlighted_cells (3, 6)										
	highlighted_cells (3, 7)										
	highlighted_cells (3, 8)										
	highlighted_cells (3, 9)										
	highlighted_cells (3, 10)										
	exp_sentence	2022년 국제 문화 교류 사업(한동)의 주된 목적에 대한 응답 결과별 보면 중사자 규모 1~9명 사업체는 '정보·지식 공유 혹은 전문 역량 강화'가 28.1%로 가장 높았고 '본국·지역에 대한 선진화 제고 및 이미지 개선'이 15.5%, '본국 문화에 발전시키고 국제 사회에 널리 알리는 것'이 15.2%로 그 뒤를 이었다.									
	reference_type	col ref									
	highlighted_cells	highlighted_cells (0, 2)									
		highlighted_cells (1, 4)									
		highlighted_cells (2, 4)									
		highlighted_cells (3, 4)									
		highlighted_cells (4, 4)									
		highlighted_cells (5, 4)									
		highlighted_cells (6, 4)									
		highlighted_cells (7, 4)									
	exp_sentence	2022년 국제 문화 교류 사업(한동)의 주된 목적에 '정보·지식 공유 혹은 전문 역량 강화'라고 응답한 비율을 중사자 규모별로 살펴보면 100~299명 사업체가 35.4%로 가장 높았고 10~49명 사업체가 34.3%, 50~99명 사업체가 29.7%로 그 뒤를 이었다.									
	reference_type	cell ref									
	highlighted_cells	highlighted_cells (6, 5)									
		highlighted_cells (7, 5)									
		highlighted_cells (4, 6)									
		highlighted_cells (7, 6)									
	exp_sentence	2022년 국제 문화 교류 사업(한동)의 주된 목적 중 '문화교류'를 통한 경제 활동 활성화 및 '본국 문화에 발전시키고 국제사회에 널리 알리는 것'을 중사자 규모별로 비교해보면 가장 높은 응답은 각각 100~299명이 18.4%, 10~49명이 25.3%로 우세가 6.9%p 더 높았으며, 가장 낮은 응답은 두 항목 모두 300명 이상에서 1.4%로 동일했다.									
Table	구분	COL2	2022								
		사태수	한국/지역에 대한 선진화 제고-이미지 개선	정보/지식 공유 혹은 전문 역량 강화	문화교류를 통한 경제 활성화(이익 증대)	문화예술인 상호교류를 통한 창작활동 활성화	본국 문화에 발전시키고 국제사회에 널리 알리는 것	문화생활 향유를 위한 상호이해 증진	매당사항없음	기타	
	연령	전체	5832	15.5	28.7	12	15.5	15	10.2	1.5	1.7
		10~49명	413	28.1	28.1	6.3	19.9	10.9	0.9	0.9	5.7
		50~99명	547	12.5	34.3	6.3	23.1	10.9	4.1	0.9	5.7
		100~299명	82	15.3	29.7	22	15.2	23.8	5.4	2.1	6.3
		300명 이상	106	28.9	35.4	14.1	7.6	18.6	2.1	2.6	0
		106	28.9	23.4	14.1	18	17.5	13.8	4.6	9.2	

□ 사진 구축 말뭉치 json 구조

```

    "id": "P100193",
    "metadata": {
      "term_id": "223703951",
      "term": "고깔",
      "domain": "생활/의생활",
      "source_id": "E0003362",
      "media_id": "80d45fa0-9659-49db-943a-cfdf92188d1a",
      "copyright": "KOGL1",
      "url": "https://devin.aks.ac.kr/image/80d45fa0-9659-49db-943a-cfdf92188d1a?preset=orig",
      "title": "승무 / 이매방,승무 - 이매방",
      "note": "승복을 입고 추는 춤으로 우리나라의 대표적인 민속춤 가운데 하나이다. 중요무형문화재 제27호. 1987년
중요무형문화재로 지정된 이매방.",
      "publisher": "한국학중앙연구원",
      "image": "P100193",
      "Major_category": "생활",
      "Medium_category": "의생활"
    },
    "EX": [
      {
        "reference": {
          "select_all": "N",
          "references": [
            {
              "reference_words": "선택",
              "bbox": {
                "x": 398,
                "px": 480,
                "y": 240,
                "py": 346
              },
              "reference_type": "rect"
            }
          ]
        },
        "exp_sentence": {
          "설명 문장1": {
            "feature": "[대상 식별 문장]",
            "sent": "이 사진은 한국의 전통 모자 중 하나인 고깔의 사진이다."
          },
          "설명 문장2": {
            "feature": "[크기]",
            "sent": "이 고깔은 얼굴 측면 대부분을 가릴 만큼 크기가 크다."
          },
          "설명 문장3": {
            "feature": "[형태]",
            "sent": "이 고깔은 윗부분이 뾰족한 삼각 모양이다."
          },
          "설명 문장4": {
            "feature": "[색채]",
            "sent": "이 고깔은 백색이다."
          },
          "설명 문장5": {
            "feature": "[(비)착용자]",
            "sent": "고깔은 과거 승려나 사헌부 나장 등이 착용하였고, 현대에는 주로 농악대나 무당이 착용한다."
          }
        }
      }
    ]
  }
}

```

맨 상단의 'id'는 국립국어원 사진 설명 문장 말뭉치 전체의 기본 정보를 담고 있다.

'document'의 'id'는 각 사진의 고유 식별자이며, 사진 ID는 P000000과 같은 형식의

로 구성됐다. ID의 첫 글자인 P는 사진(Picture) 데이터임을 의미하며, 뒤이어 나오는 여섯 자리 숫자 중 앞의 두 자리는 수집처 번호(1부터 시작)를, 뒤의 네 자리는 사진 번호(0001부터 시작)를 나타낸다. 수집처 번호는 다음과 같이 정의했다.

- 한국민족문화대백과사전 (1)
- 포토코리아 (2)
- 지역어종합정보 (3)
- 민속아카이브 (4)
- 한국민속대백과사전 (5)
- e뮤지엄 (6)
- 국가유산디지털서비스 (7)
- 국악아카이브 (8)
- 한식진흥원 (9)
- 공유마당 (10)
- 데이터리 (11)¹⁾

‘metadata’에는 사진 자체에 관한 상세 정보가 포함됐다. ‘term_id’는 대상 용어 번호, ‘term’은 대상 용어, ‘domain’은 수집 과정에서 지정한 자체 분야 정보이다. ‘source_id’는 원자료 번호를, ‘media_id’는 수집처에서 부여한 멀티미디어 ID를 의미한다. ‘copyright’에는 저작권 정보가 기입되었고, ‘url’에는 원문 자료를 확인할 수 있는 링크가 포함됐다. ‘title’은 사진의 제목, ‘note’는 사진에 대한 설명 또는 부가 정보, ‘publisher’는 이미지 제공 기관 또는 저작권자를 나타내며, ‘image’에는 도구에 업로드 되는 JPG 파일명이 기록된다. 또한 ‘major_category’와 ‘medium_category’는 작업자가 도구 내에서 직접 지정한 사진의 대분류·중분류 정보이다. 사진 ‘metadata’의 경우는 ‘한국민족문화대백과사전’의 구조를 기준으로 형식이 잡혀 있어, 다른 수집처에는 존재하지 않거나 성격이 다른 항목들이 있다. 예를 들어 ‘term_id’와 ‘term’은 한국민족문화대백과사전에만 존재하는 특수 항목이며, 다른 수집처 자료에는 해당 값이 없다. ‘source_id’와 ‘media_id’ 또한 수집처별로 성격이 달라, ‘source_id’는 원 파일명을, ‘media_id’는 수집처 내부 고유 번호를 기록했으며, 없는 경우는 공란으로 두었다. ‘note’ 역시 수집처 특

1) 본 자료는 주관기관(데이터리)에서 직접 촬영 및 메타데이터를 구축한 성과물임.

성에 따라 기재 가능한 경우에 한해 작성했고, 제공되지 않은 경우는 비워두었다.

‘EX’에는 바운딩 정보와 설명 문장 관련 정보가 함께 포함된다. 먼저 ‘reference’ 항목은 사진 설명 문장을 작성할 때 지정한 바운딩 영역 정보를 의미한다. ‘select_all’은 전체 이미지를 선택했는지 여부를 나타내며, 값이 ‘N’이면 사진의 특정 부분만 선택되었음을 뜻한다. 금년도 사업에서는 전체 영역에 대한 바운딩 작업은 진행하지 않았으므로 모든 결과물의 값은 ‘N’으로 입력되어 있다. ‘references’에는 하나 이상의 바운딩 영역 정보가 배열 형태로 기록되며, 각 항목의 ‘reference_words’는 해당 영역에 부여된 라벨, ‘reference_type’은 지정된 영역의 형태를 나타낸다. ‘bbox’는 바운딩 박스를 구성하는 좌표값으로, ‘x’, ‘y’는 시작 지점, ‘px’, ‘py’는 끝 지점을 의미한다.

설명 문장 부분은 총 1~5개의 항목으로 구성되며, 각 항목의 ‘feature’에는 선택된 속성 정보가, ‘sent’에는 해당 속성을 기반으로 작성된 설명 문장이 기록된다.

2.3.4. 구축 내역

□ 표 설명 문장 작성 및 교차 검수 작업 내역

차수	배포량	구축량	검수량
1	200	198	194
2	202	197	197
3	200	186	193
4	219	215	209
5	204	203	203
6	201	201	201
7	200	200	198
8	200	182	184
9	218	208	211
10	210	210	209
추가	-	40	38
합계	2,054	2,040	2,037

<표 11> 표 설명 문장 작성 및 교차 검수 작업 내역

표 설명 문장 구축 및 검수 작업은 총 10주차 동안 진행됐다. 표 자료는 작업에서 제외되는 필터링 수량이 적어 주차별 배포량과 실제 구축 및 검수량의 차이가 거의 없었다. 전체 수합 이후 관리자가 필터링된 표와 구축은 됐으나 도구 상 필터링으로 처리된 표를 전수 확인하여 추가 수량을 확보했고, 그 결과 총 2,037건의 구축 작업물을 산출했다. 해당 2,037건에 대한 통계는 다음과 같다.

□ 표 설명 문장 구축 내역

• 수집처별

수집처	수량
국가통계포털	1680
문화센터	357
총합	2,037

• 주제별

주제	수량	주제	수량
생활 > 민속·인류	713	언론·출판 > 언론·방송	17
경제·산업 > 산업	325	정치·법제 > 법제·행정	8
생활 > 식생활	245	예술·체육 > 국악	7
공연·예술 > 기타	205	예술·체육 > 대중음악	6
사회 > 사회구조	187	예술·체육 > 무용	5
생활 > 주생활	111	교육 > 교육	4
생활 > 의생활	100	정치·법제 > 정치	4
정치·법제 > 외교	39	언론·출판 > 출판	2
예술·체육 > 기타	33	예술·체육 > 연극	1
언어 > 언어·문자	24	과학 > 과학기술	1
총합		총합	2,037

□ 사진 설명 문장 작성 및 교차 검수 작업 내역

차수	배포량	구축량	검수량
1	861	692	700
2	957	875	881
3	913	819	829
4	914	772	774
5	967	714	679
6	591	490	486
7	794	736	703
8	986	907	916
9	561	530	485
10	792	716	727
11	868	803	814
12	852	818	819
13	826	769	769
14	856	792	804
15	856	839	836
16	612	599	605
17	887	879	884
18	193	185	185
추가	150	150	109
합계	14,436	13,085	13,005

<표 12> 사진 설명 문장 작성 및 교차 검수 작업 내역

사진 설명 문장 구축 및 검수 작업은 총 17주차 동안 진행됐다. 사진 자료는 특성상 작업 대상에서 제외되는 이미지가 많이 발생하여 주차별 배포량과 실제 구축 및 검수량 간 편차가 컸다. 필터링된 수량은 차주 배포 계획에 반영하여 전체 목표 수량을 맞출 수 있도록 조정했다. 정규 작업 종료 후 추가 구축이 필요한 분량은 우수 작업자 5인에게 18주차 작업으로 배정했다. 전체 수합 이후 관리자가 필터링된 사진을 전수 검토하여 재작업 가능 여부를 판단한 뒤, 구축 가능하다고 확인된 자료는 추가 구축과 검수를 통해 최종 결과물에 포함했다. 또한 구축이 완료됐으나 도구에서 필터링으로 처리된 작업물도 별도로 확인하여 수합했다. 이러한 작업을 거쳐 최종적으로 13,005건의 사진 설명 문장을 구축했으며, 통계는 다음과 같다.

□ 사진 설명 문장 구축 내역

• 수집처별

수집처	수량	수집처	수량
한국민족문화대백과사전	2,831	국가유산디지털서비스	1,405
포토코리아	3,432	국악아카이브	112
지역어종합정보	297	한식진흥원	676
민속아카이브	1,781	공유마당	197
한국민속대백과사전	96	데이터리	70
e뮤지엄	2,108	총합	13,005

• 주제별

주제	수량
식생활	3,310
의생활	2,728
주생활	2,503
건축	3,370
공예	1,094
총합	13,005

사진의 경우, 사업 초기에는 각 분류가 전체 구축량의 최대 20%(약 2,600건)를 넘지 않도록 설정했다. 그러나 실제 수집 과정에서 사용 가능한 자료의 수가 예상보다 적어 제외되는 사례가 다수 발생했고, 특히 공예 범주는 공개된 자료 자체가 매우 적다는 문제가 확인됐다. 이에 국어원과 회의를 통해 이러한 수집 한계와 문제 상황을 공유한 뒤, 대안을 마련하여 각 분류별 구축 가능 비율을 조정하기로 했으며, 최종 구축량을 위와 같이 확정했다.

이와 같이 구축된 데이터의 신뢰성과 객관성을 점검하기 위해 사업 수행 과정에서 전

문가 자문을 실시했다. 해당 자문은 말뭉치 데이터의 구축 방향, 평가 방법의 적절성, 향후 활용 가능성 등에 대한 검토를 목적으로 진행되었으며, 본 사업에서 구축된 말뭉치의 전반적인 구성과 활용 적합성을 확인하는 절차로 활용됐다. 자문 과정에서 도출된 주요 의견은 향후 말뭉치의 확장 및 고도화를 위한 참고 자료로 활용할 수 있도록 정리하였다.

□ 표 설명 문장 말뭉치 자문 내용

날짜	2025년 11월 21일(금)
자문 위원	경기대학교 최성필 교수
자문 내용	<p>1. 표 설명 문장 말뭉치에서 ‘표 전체·행·열·불연속 영역’을 구분하여 각 영역별 설명 문장을 구축하는 방식이 실제 언어 연구나 인공지능 학습에서 유의미한 차별성을 가질 수 있을지 궁금합니다. 만약 그렇지 않다면, 구축 방향을 어떻게 조정하는 것이 좋을지 의견을 부탁드립니다.</p> <ul style="list-style-type: none"> - 현재 지침은 표를 단순한 텍스트가 아닌 2차원 구조 데이터로 인식하게 만드는 데 초점이 맞춰져 있음. 이는 표의 구조적 좌표 학습 (Coordinate Understanding), 다각도 정보 추출 (Multi-view Extraction), 복합 추론의 기초 (Chain-of-Thought) 와 같은 초기 단계의 멀티모달 LLM 학습에 이점을 가짐. - 그러나 실제 사용자 경험(UX)이나 고성능 LLM의 활용 측면에서는 행/열 구분의 모호성과 중복성, 단순 낭독(Read-out) 수준의 한계, 자연스러움 결여와 같은 한계가 예상되므로, 구축되는 데이터의 가치를 높이기 위해, 물리적 구조(행/열) 중심에서 ‘의미적 기능(Semantic Function)’ 중심으로 라벨링을 변경하거나 확장하는 것을 고려할 필요가 있음. <p>2. 현재는 1개의 표당 4개의 설명 문장을 작성하는 방식을 적용하고 있습니다. 이때 작성된 4개의 설명 문장을 기반으로 유사 문장(Paraphrasing) 데이터를 추가 구축하는 것이 인공지능 학습에 도움이 될지, 도움이 된다면 1세트당 어느 정도의 유사 문장 세트를 확보하는 것이 적정할지 자문을 부탁드립니다.</p> <ul style="list-style-type: none"> - 유사 문장(Paraphrasing) 데이터 구축은 인공지능 학습 효율을 높이는 데 매우 긍정적이며 필수적인 요소이며, 특히 현재의 지침처럼 특정 포맷(Template)에 맞춰 정형화된 설명 문장만 학습할 경우, AI는 데이터의 의미를 이해하기보다 문장 패턴을 암기(Overfitting)하는 경향을 보일 수 있음. 이를 방지하고 일반화(Generalization) 성능을 높이기 위해 유사 문장 구축은 필수적임. <p>3. 표 설명 말뭉치를 향후 ‘데이터 설명 자동 생성’이나 ‘표 기반 질의응답(QA)’ 모델 학습에 활용하기 위해 추가로 필요한 메타데이터 요소가 있을지 궁금합니다. 또한 이러한 활용 외에 본 말뭉치가 적용될 수 있는 다른 연구나 산업적 활용 방안이 있다면</p>

함께 의견을 부탁드립니다.

- 현재 메타데이터는 검색 중심 정보에 치우쳐 있어, 표 해석과 추론 학습을 위해서는 구조적·논리적 메타데이터 보강이 필요함. 표 유형(시계열, 구성비, 비교 등), 헤더 계층 구조, 설명 문장 생성 시 사용된 연산 유형을 추가할 경우 AI 학습 효과를 높일 수 있음
- 또한 본 말뭉치는 웹 접근성용 표 대체 텍스트 생성, 한국어 교육 평가, 자동 팩트 체크, 검색 증강 생성(RAG) 등 다양한 연구 및 산업 분야로 확장 활용이 가능함.

4. 현재 구축 중인 표 설명 문장 말뭉치를 향후 그래프·차트 자동 설명 생성, 시각자료 자막 자동 생성, 표-문장 매칭 데이터 등으로 확장하려 할 때, 기술적 관점에서 어떤 추가 정보(예: 수치 좌표, 증감 방향, 비교 관계 등)를 함께 구축해야 실질적인 인공지능 학습 데이터로 활용될 수 있을지 조언을 부탁드립니다.

- 설명 문장의 특정 구절이 표의 어느 셀에서 비롯되었는지를 나타내는 근거 셀 좌표 정보 구축이 필요함. 또한 최댓값, 차이, 추세 등 설명 문장에 사용된 논리 연산 유형과 변수 유형, 적합한 시각화 방식 정보를 함께 구축할 경우 차트 및 시각자료 설명 생성으로의 확장이 가능할 것으로 판단됨.

5. 구축된 표 설명 말뭉치의 성능을 평가하기 위해 다음과 같은 방식을 검토 중입니다. 이러한 평가 방식이 타당한지, 혹은 더 적절한 검증 지표나 평가 절차가 있을지 의견을 부탁드립니다.

1) 유사도 평가: 정답 말뭉치(약 200건)와 구축 말뭉치(약 2,000건) 간의 BERTScore 계산

- 모집단 불일치 문제: 정답 말뭉치(200건)와 구축 말뭉치(2,000건)를 단순 비교 시, 서로 다른 표에 대한 문장이 섞여 평가 신뢰도 하락함.
- 1:1 매칭 필수: 유사도 평가는 반드시 '동일한 표'를 보고 작성된 '정답 문장'과 '생성 문장' 간의 비교여야 유의미함.
- 개선 방향: 전체 2,000건 비교가 아닌, 정답 데이터가 존재하는 200건(또는 별도 검증셋)에 한정하여 점수 산출 필요.

2) 유의어 확장 유사도 평가: 동일 데이터 간 METEOR 점수 계산

- 의견 없음.

3) 유창성 검사: 구축 말뭉치의 Perplexity 역수를 활용한 언어 유창성 평가

- 유창성의 한계: 한국어 문법과 자연스러움 측정에는 적합하나, '거짓 정보(Hallucination)' 탐지 불가함.

- 데이터 특성 고려: 표 설명 데이터는 문장의 매끄러움보다 '수치와 팩트'가 핵심이나, PPL은 내용이 틀려도 문장이 자연스러우면 높은 점수가 나오는 맹점 존재.

4) 추가 검증 지표 및 절차 제안

- 사실 부합성(Fidelity) 검증 추가: '수치 일치율(Number Exact Match)' 도입 권장.
- 문장 내 포함된 숫자(연도, 값, 비율)가 원본 표 데이터에 실제 존재하는지 기계적으로 대조.

	<ul style="list-style-type: none"> - 표 설명 특화 지표 도입: 여력 허용 시, 정답 유사도와 원본 표 정보 함유량을 동시 평가하는 PARENT 지표 활용 추천. - 투트랙 평가 절차 수립: <ul style="list-style-type: none"> ○ (전수 평가) 전체 데이터 대상 ‘수치 정확도’ 및 ‘유창성(PPL)’ 검사로 기본 품질 확보. ○ (샘플 평가) 정답셋이 확보된 데이터에 한해 ‘유사도(BERTScore/METEOR)’ 측정하여 성능 입증.
--	---

<표 13> 표 설명 말뭉치 구축 데이터 전문가 자문 내용

□ 사진 설명 문장 말뭉치 자문 내용

날짜	2025년 11월 19일(수)
자문 위원	ETRI 손정우 책임연구원
자문 내용	<ol style="list-style-type: none"> 1. 본 사업에서는 [사진 설명 문장 작업 대상]에 부합하는 유형 객체 사진을 대상으로 설명 문장을 작성하고 있습니다. ‘식생활, 의생활, 주생활, 건축, 공예’의 5개 범주로 분류하여 데이터를 구축하고 있는데, 이 구성이 한국어문화 학습 데이터로 충분하다고 보시는지, 혹은 추가로 포함되면 좋을 범주가 있을지 의견을 부탁드립니다. <ul style="list-style-type: none"> - 현재 제시된 5개 범주는 매우 다양한 대상을 포함하고 있어 학습 데이터로서 충분하다고 판단됨. 다만, 향후 확장이 가능하다면 봉산탈춤, 처용무, 북청사자놀음 등 무형유산 카테고리를 추가하여 한국의 지역적 특색을 보다 강하게 드러낼 수 있도록 하는 방안을 검토해볼 수 있겠음. 2. 본 사업에서는 가시적 속성(형태·색채·문양·재료 등)을 우선으로 구축하고, 비가시적 정보(역사·기법·용도 등)는 5문장 중 최대 2문장으로 제한하고 있습니다. 이러한 원칙이 인공지능 학습용 멀티모달 데이터로 충분히 기능할지, 혹은 비가시적 정보의 비율을 달리해야 할지 의견을 부탁드립니다. <ul style="list-style-type: none"> - 가시적 속성과 비가시적 속성을 함께 태깅하여 제공하는 방식은 매우 적절한 방향으로 보임. 가시적 속성에 비해 비가시적 속성은 태깅의 자유도를 다소 낮출 필요가 있어 보임. 또한, 가시적 속성을 태깅할 때 대상 객체가 여러 요소로 구성되어 있는 경우 (예: 비빔밥의 계란지단, 호박, 밥 등) 각 속성 문장에 포함할 수 있는 요소를 최대한 충실히 기술할 것을 권함. 이러한 방식은 작업자의 임의 선택을 최소화하고, 데이터 일관성을 높이는 데 도움이 될 것으로 판단됨. 3. 본 사업에서 구축한 사진 설명 문장은 ‘대상 식별 문장 1개 + 속성 문장 4개’로 구성되어 있습니다. 이러한 구성만으로 GPT 등 대형 언어모델의 한국문화 이해 능력 향상을 위한 학습 데이터로 충분하다고 보시는지, 만약 불충분하다면 어떤 추가 메타데이터나 보강 정보가 포함되어야 할지 자문을 부탁드립니다. <ul style="list-style-type: none"> - 현재 구축된 구성(대상 식별 문장 1개 + 속성 문장 4개)은 대형 언어모델이 한국

문화나 양식에 대한 지식을 추출하는 데 충분히 활용 가능한 것으로 판단됨. 다만, 필요하다면 동일 객체에 대한 복수의 이미지를 함께 제공하는 방향도 검토할 수 있을 것으로 보임. 이는 모델이 대상 객체의 변이·맥락·다양한 형태를 학습하는 데 도움이 될 것이라 판단됨.

4. 본 사업에서는 공공기관 등재 데이터(CC-BY, 공공누리 1유형)를 중심으로 사진을 수집했습니다. 향후 유사 사업 추진 시, 추천할 만한 신규 데이터 수집처가 있을지, 그리고 필요 시 직접 촬영을 통한 데이터 확보에 대해 어떻게 생각하시는지 의견을 부탁드립니다.

- 최근 다양한 데이터셋과 수집처에서 학습 데이터 활용 시 라이선스 제약을 강화하는 추세여서, 이미지를 확보하는 것이 이전보다 더 어려워질 것으로 판단됨. 특히 한국적 특색이 강한 이미지의 경우 수량 자체가 한정적이라는 점도 고려해야 함. 따라서 향후 유사 사업에서는, 수량이 줄더라도 직접 촬영을 통한 데이터 구축을 적극적으로 추천함. 이 방식은 활용성 측면에서도 유리하며, 학계나 연구 분야에서도 파급력이 더 클 것으로 판단됨.

5. 한국어언어문화 사진 설명 문장은 향후 어떤 한국어 생성형 모델 학습(예: 문화유산 자동 기술, 이미지 캡셔닝, 질의응답 등)에 활용될 수 있을지, 현재의 구축 정보로 충분한지, 혹은 추가로 구축이 필요한 언어 데이터(예: 텍스트 요약, 유사 문장 생성 등)가 있을지 의견을 부탁드립니다.

- 현재 구축된 사진 설명 문장은 이미지 캡셔닝, 질의응답 등 기존 태스크에 바로 활용 가능한 수준으로 판단됨. 또한 한국어·한국문화 지식맵 구축, 이미지 생성 모델의 조건부 정보 제공 등 다양한 활용 가능성도 있음.

추가 구축이 가능하다면, 조리법·취식방법 등 절차적 정보를 다수의 단문 형식으로 기술하는 방식을 고려할 수 있음. 이러한 형태의 데이터는 모델 활용 편의성을 더욱 높일 수 있을 것으로 보임.

<표 14> 사진 설명 말뭉치 구축 데이터 전문가 자문 내용

2.4. 표·사진 설명 말뭉치 평가

2.4.1. 평가 내용

본 사업에서는 설명 말뭉치의 품질을 확인하기 위하여 다양한 평가를 실시했다. 평가의 객관성을 높이기 위해 전문가가 구축한 정답 말뭉치를 활용했으며, 이는 전체의 10%인 표 설명 200건, 사진 설명 1,300건으로 구성됐다. 해당 정답 말뭉치는 BLEU,

BERTScore, METEOR 값을 산출하는 대조군으로 사용됐다. 이 밖에도 말뭉치 자체의 유창성 검사와 정합성 및 관계 추론 검증을 병행하여 다각적인 품질 평가를 완료했다.

2.4.1.1. 유사도 평가 (BLEU)

구축된 설명 문장과 지침 작성 전문가가 구축·검수한 정답 말뭉치 간의 BLEU 값을 계산하여, 설명 문장이 의미를 얼마나 정확히 재현했는지를 평가했다. BLEU는 문장 간 n-gram 일치도를 기반으로 한 지표로, 0과 1 사이의 값을 가지며 1에 가까울수록 의미적 재현성이 높음을 의미한다. 다만 표현 다양성이나 동의어 사용에 따라 의미가 동일하더라도 점수가 낮게 산출될 수 있다는 한계가 있다. 이러한 특성을 고려해, 선행 연구 및 자연어 생성 평가 사례에서 의미 전달이 안정적으로 이루어진 수준으로 해석되는 BLEU 0.4~0.6 범위를 참고하여, 본 사업에서는 BLEU 0.5 이상을 최종 목표치로 설정했다.

2.4.1.2. 유사도 평가 (BERTScore)

구축된 설명 문장과 정답 말뭉치 간의 의미적 유사성을 평가하기 위해 BERTScore 검사를 수행했다. BERTScore는 사전 학습된 BERT 계열 언어 모델을 활용하여 문장 간 의미적 유사도를 계산하는 지표로, n-gram 일치에 기반한 BLEU와 달리 문맥과 의미 수준의 일치 여부를 평가할 수 있다. 다만 BERTScore 역시 완전한 의미 동일성을 절대적으로 보장하는 지표는 아니며, 문장 길이, 표현 밀도, 모델 특성 등에 따라 점수 편차가 발생할 수 있다. 이러한 특성을 고려해, 선행 연구 및 자연어 생성 평가 사례에서 의미 전달이 안정적으로 이루어진 수준으로 해석되는 기준인 BERTScore 0.85 이상을 본 사업의 최종 목표치로 설정했다.

2.4.1.3. 유의어 확장 유사도 평가 (METEOR)

구축된 설명 문장과 지침 작성 전문가가 구축·검수한 정답 말뭉치 간 METEOR 값을 계산하여, 단순 표면 구조 일치 여부를 넘어 의미 기반 유사도가 충분히 확보됐는지 평가했다. METEOR는 BLEU의 한계를 보완하기 위해 개발된 지표로, 다음과 같은 언어학적 요소를 반영하여 보다 정밀한 평가가 가능하다.

- 어간 추출(stemming): 형태는 달라도 공통 어간을 가진 단어를 동일하게 처리
- 동의어 매칭(synonym matching): WordNet 등 사전 기반 의미 유사 단어 연결
- 패러프레이징(paraphrasing): 문장 구조가 다르더라도 의미가 같으면 유사하게 평가
- 단어 순서 고려: 문장 배열의 자연스러움까지 반영
- 다중 정답 참조: 하나의 정답에만 고정되지 않고 안정적인 평가 가능

METEOR는 0~1 사이의 값을 산출하며, 1에 가까울수록 높은 품질의 문장 생성을 의미한다. 본 사업의 최종 목표치는 METEOR 0.7 이상으로 설정했다.

2.4.1.4. 유창성 검사

설명 문장의 유창성을 평가하기 위해 언어 모델이 산출한 혼란 점수(Perplexity score)의 역수를 활용했다. Perplexity는 문장을 구성하는 각 토큰이 이전 토큰을 조건으로 생성될 확률을 기반으로 산출되는 지표로, 값이 낮을수록 문장이 자연스럽게 유창함을 의미한다. 본 사업에서는 해석의 직관성을 높이기 위해 Perplexity의 역수를 유창성 점수로 사용하여, 값이 높을수록 유창성이 높은 구조로 평가했다. 유창성 점수 0.5 미만을 수정 기준으로 설정했으며, 평가 결과 전 문장이 기준을 충족했다.

2.4.1.5. 정합성 및 관계 추론 검증

표 설명 문장 말뭉치의 경우 표의 실제 내용과 설명 문장 간의 일치도를 확인하기 위해 정합성 및 관계 추론 검증을 2차에 나누어 수행했다. 검증은 행, 열, 불연속 영역 설명 문장을 대상으로 동일한 LLM을 활용해 단일 검증 과정에서 함께 수행했다. 표에 제시된 수치와 설명 문장에 기술된 수치를 자동으로 대조해 수치 기입 오류 여부를 검출함과 동시에 최댓값·최솟값 지정 오류 및 비교 관계 오류 등 설명 문장의 논리적 타당성을 점검했다. 1차 검증 후 도출된 오류를 수정 보완하고, 2차 추가 검증을 통해 최종 말뭉치의 품질을 고도화하였다.

2.4.2. 평가 결과

위에서 설명한 평가 과정을 거친 후 산출된 결과는 다음과 같다.

유사도 평가 (BLEU)

표	
BLEU	
목표치	평가 결과
0.5 이상	0.554

사진	
BLEU	
목표치	평가 결과
0.5 이상	0.556

유사도 평가 (BERTScore)

표	
BERTScore	
목표치	평가 결과
0.85 이상	0.943

사진	
BERTScore	
목표치	평가 결과
0.85 이상	0.939

유의어 확장 유사도 (METEOR)

표	
METEOR	
목표치	평가 결과
0.7 이상	0.836

사진	
METEOR	
목표치	평가 결과
0.7 이상	0.803

유창성 검사

표
fluency
0.734

사진
fluency
0.797

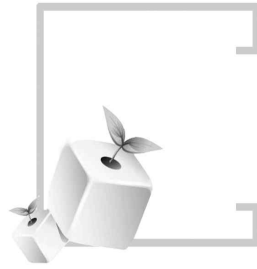
정합성 및 관계 추론 검증 (1차)

표		
전체 문장 수	오류 문장 수	오류율
6,111	667	10.92%

정합성 및 관계 추론 검증 (2차)

표		
전체 문장 수	오류 문장 수	오류율
6,111	197	3.22%

‘정합성 및 관계 추론 검증’ 결과, 확인이 필요한 행, 열, 불연속 영역 설명 문장은 총 6,111개로, 그중 1차 검증 시 발견된 오류 667문장과 2차 검증 시 발견된 오류 197문장은 관리자가 확인하여 전부 수정했다.



제 3 장

표·사진 설명 문장 작성 지침



3.1. 지침 수정·보완 및 수립 방안

3.1.1. 표 설명 문장 작성 지침 수정·보완

<2025년 한국어문화 표 설명 문장 작성 지침>은 한국어문화와 관련된 표에 대한 설명 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 한다. 지침에서는 표의 핵심 내용이 드러날 수 있는 설명 영역을 선정하고 표 전체, 행, 열, 칸(불연속 영역)에 대한 서술문을 작성하는 방안을 다양한 예시와 함께 제시하였다. 또한 ‘2025년 한국어문화 표·사진 설명 말뭉치 구축 연구’ 제안 요청서의 내용에 따라 ‘2024년 표 설명 문장 작성 말뭉치 연구 분석’에서 수립한 <2024년 표 설명 문장 작성 지침>을 수정·보완하기 위해 지침을 분석하고 전년도 사업에서 말뭉치를 구축하면서 나타난 문제점을 해소하고자 하였다.

먼저 2024년도 사업과 2025년도 사업의 차이점을 살펴보면, ‘2024년 표 설명 문장 작성 말뭉치 연구 분석’에서 설명 대상으로 삼은 표 자료의 주제는 ‘공공 행정, 교육, 사회 복지, 과학 기술, 보건 의료, 문화 관광’ 등 17개 주제였으나, ‘2025년 한국어문화 표·사진 설명 말뭉치 구축 연구’에서 설명 대상으로 삼은 표 자료의 주제는 ‘한국어문화’로 한정되었다.

다음으로 전년도 사업에서는 보도 자료나 연구 보고서 등에 삽입된 표를 수집하였으나, 2025년도 사업에서는 통계 포털에서 한국어문화 관련 표를 수집하였다.

지역별(1)	2022		구분1		구분2		기초		군		구				
	전체	소계	광역	소계	광역	특별·광역시	광역시	광역시	시	시	시	구			
전체	10,846		66,496		6,697		85,649		49,470		13,148		2,912		3,996
서울	14,544		223,521		6,185		223,521		-		-		-		6,185
부산	6,530		91,537		1,217		91,537		-		-		494		1,265
대구	12,874		81,423		4,305		81,423		-		-		1,740		4,672
인천	13,494		95,183		5,325		95,183		-		-		350		6,569
광주	14,164		78,698		1,258		78,698		-		-		-		1,258
대전	11,348		62,006		1,217		62,006		-		-		-		1,217
울산	7,435		34,311		2,060		34,311		-		-		5,798		1,125
세종	18,517		18,517		-		18,517		-		-		-		-
경기	18,513		140,193		14,588		-		140,193		15,912		2,228		-
강원	8,440		37,399		6,831		-		37,399		11,147		4,085		-
충북	8,544		31,270		6,479		-		31,270		16,663		2,659		-
충남	8,184		31,337		6,640		-		31,337		9,207		3,706		-
전북	10,807		65,034		6,934		-		65,034		13,457		2,042		-
전남	5,895		31,598		4,727		-		31,598		9,635		3,283		-
경북	7,164		45,439		5,500		-		45,439		9,913		2,105		-
경남	9,723		35,373		8,298		-		35,373		14,709		3,169		-
제주	15,675		27,587		9,718		-		27,587		9,718		-		-

[그림 15] 국가통계포털(KOSIS) 표 자료 제공 형식

이에 따라 아래와 같은 주요한 차이가 발생하였다.

- 보도 자료나 연구 보고서 등에는 ‘기사 제목, 보도 일자, 표 제목, 저작권’ 등의 정보가 제공되어 있으나, 통계 포털은 표 제목을 제외한 정보가 없다.
- 보도 자료나 연구 보고서 등에는 설명 문장 작성 시 참고할 수 있는 수집 문장이 있으나, 통계 포털에는 수집 문장이 없다.
- 보도 자료나 연구 보고서 등은 동일한 제목의 표를 연도, 내용 등에 따라 별도의 표로 제공하나, 통계 포털은 주로 하나의 표로 제공한다.

표 수집처가 달라짐에 따라 위와 같은 차이가 발생하여, 이를 지침에 반영하여야 했다. 이에 따라 ‘기사(보고서) 제목’과 ‘수집 문장’을 활용하여 설명 문장을 작성한다는 기존 지침의 내용을 삭제하고, 표의 주제 및 내용을 표현하는 가장 기본적인 정보인 표 제목을 우선적으로 활용하여 설명 문장을 작성할 수 있도록 지침을 수정·보완하였다.

표의 유형은 전년도 사업에서 ‘제목 행과 제목 열에 따른 분류, 셀 값의 데이터 형식에 따른 분류, 합계 제시 방식에 따른 분류’ 등으로 체계적으로 유형화한 바가 있어 2025년도 사업에서도 이 체계를 그대로 따랐으며, 한국어언어문화 관련 표로 각 유형의 예시를 교체하였다.

3.1.2. 사진 설명 문장 작성 지침 수립

<2025년 한국어언어문화 사진 설명 문장 작성 지침>은 한국어언어문화와 관련된 사진에 대한 설명 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 한다. 지침에서는 사진 자료에 대한 설명 문장을 작성하는 방안을 제시하였다. 사진 속 대상이 무엇인지 식별하는 설명 문장인 ‘대상 식별 문장’과 사진 속 대상의 특정한 속성을 기술하는 설명 문장인 ‘속성 기술 문장’을 작성하는 구체적인 방안을 다양한 예시와 함께 제시하였다.

또한 ‘2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구’ 제안 요청서의 내용에 따라 한국어언어문화 분류별 설명 속성을 마련하기 위해 <한국민족문화대백과사전>과 <위키피디아>에 등재된 한국어언어문화 요소의 뜻풀이를 분석하였다. 한국어언어문화의 분류 중 식생활, 의생활, 주생활, 건축, 공예 분야의 등재 요소를 중심으로 두 사전의 공통 기술 속성을 도출하였으며, 사례 검토 결과를 바탕으로 분야별 특화 속성을 보완하였다. 특히 건축과 공예의 특수성을 고려하여 건축의 비가시적 속성인 ‘건축가’와 ‘건축 양식’, 공예

의 ‘기법’ 속성을 추가하였다.

[그림 16] 한국어언어문화 분류별 설명 속성 구축을 위한 사전 조사

사진 설명 문장 작성 지침 수립을 위해 유사 사업 및 선행 연구를 검토하였으나, ‘한국언어문화’ 관련 사진에 대한 설명 문장 작성 방법을 연구한 사례는 없었다. 따라서 이 사업에서는 사진을 내려받을 수 있고 출처를 확인할 수 있는 <한국민족문화대백과사전>에서 사진을 수집하여 사진 설명 문장 작성 샘플 작업을 수행하였다. 식생활, 의생활, 주생활, 건축, 공예에 속하는 한국어언어문화 요소들에 대한 설명 문장을 작성하며 한국어언어문화 분류별 설명 속성을 마련하고, 지침의 초안을 수립하였다.

분류	하위 분류	표제어	사진	속성	설명문장	비고	질문/제안 사항
주	건물	귀틀집		정의	이 사진은 한국 전통 가옥인 귀틀집 사진이다.		
				형태	이 귀틀집의 벽면은 사각형 형태이며 지붕은 곡선 형태이다.		
				구조	이 귀틀집은 정(井)자 구조로 되어 있다.		사진으로는 식별할 수 없음
				자재	이 귀틀집은 통나무, 짚, 흙 등으로 지어졌다.		
				분포(주)	귀틀집은 산림이 많은 중부 이북의 산간 지역에 많이 분포한다.	속성 추가	
주	건물	ㄱ자집		정의	이 사진은 한국 전통 가옥인 ㄱ자집 사진이다.		
				형태	이 ㄱ자집은 용마루가 ㄱ자 형태로 엮여 있다.		
				구성(주)	이 ㄱ자집은 방과 대청, 분합문 등으로 이루어져 있다.	속성 추가	
				자재	이 ㄱ자집은 돌, 나무, 흙벽 등으로 만들어졌다.		
				분포(주)	ㄱ자집은 중부 지방에서 가장 널리 이용된다.	속성 추가	
주	건물	안채		정의	이 사진은 한국 전통 가옥인 안채를 이루는 여러 건물 중 하나인 안채의 사진이다.		
				형태	이 안채의 기둥은 원통 형태이며, 벽면은 직사각형 형태이다.		
				구조	이 안채는 두개의 기둥이 보를 받치고 있는 구조이다.		
				자재	이 안채는 통나무, 돌, 흙벽 등으로 만들어졌다.		
				용도(주)	안채는 여성들이 일상생활, 가사일 등을 하는 공간이다.		

[그림 17] 사진 설명 문장 지침 수립을 위한 샘플 작업 예시

지침 초안 작성 이후에는 정기 회의를 진행하며 설명 문장 작성 시에 발생한 문제와 말뭉치 구축 작업자들의 질의 내용, 국립국어원의 지침 검토 결과, 전문가의 자문 등을 수용하여 지속적으로 지침을 보완하였다. 사진 설명 문장 작성 지침 수립 및 수정·보완을 위한 회의는 아래와 같이 총 19회 진행되었다.

구분	일자	내용	지침
초안 작성	2025.06.24. 2025.07.01. 2025.07.08. 2025.07.15.	<ul style="list-style-type: none"> • 사진 설명 문장 작성 지침 수립 방안 논의 • 1차 샘플 작업: 54 개의 사진에 대한 설명 문장 작성 • 의, 식, 주, 주제별 설명 속성 및 우선순위 확정 • 사진 설명 문장 작성 방법 보완, 추가 사항 논의 	지침_1.0(20250715)
1차 수정	2025.07.22.	<ul style="list-style-type: none"> • 설명 문장 작성 문제 사례 검토 및 보완 방안 논의 • 국립국어원 지침 검토 결과 분석 및 지침 반영 	지침_1.1(20250723)
2차 수정	2025.07.28. 2025.08.06.	<ul style="list-style-type: none"> • 식생활 관련 도구 설명 문장 작성을 위한 설명 속성 마련 및 보완 방안 논의 	지침_1.2(20250811)
3차 수정	2025.08.12. 2025.08.19. 2025.08.26. 2025.09.03. 2025.09.10.	<ul style="list-style-type: none"> • 설명 문장 작성 문제 사례 검토 및 보완 방안 논의 • 사진 설명 문장 추가 범주(건축, 공예)에 대한 설명 속성 마련 및 지침 작성 방향 논의 • 사진 설명 문장 작성 지침 1.3 버전 수정 사항 검토 및 보완 방향 논의 	지침_1.3(20250916)
최종 수정	2025.09.17. 2025.09.24. 2025.10.01. 2025.10.15. 2025.10.22. 2025.11.05. 2025.11.19.	<ul style="list-style-type: none"> • 사진 설명 문장 작성 지침 1.3 버전 보완 방향 논의 • 국립국어원 지침 검토 결과 분석 및 지침 반영 	지침_1.3_저용량(20251027)

<표 15> 사진 설명 문장 작성 지침 수립 및 수정·보완 회의 내역

전문가 자문은 1, 2차로 나누어 진행하였다. 1차 자문 회의에서는 한국어언어문화 표·사진 설명 말뭉치 구축을 위한 분류 체계를 다루었고, 2차 자문 회의에서는 표·사진 설명 문장 작성 지침 및 말뭉치의 활용을 다루었다. 여기에서는 지침에 관한 자문인 2차 자문 회의의 내용을 제시한다.

날짜	2025년 11월 26일(수)
자문 위원	마인즈 솔루션 이숙의 대표
자문 내용	1. 「2025년 한국어언어문화 사진 설명 문장 작성 지침」에서는 생활>식생활/의생활/주생활, 예술>건축/공예 각각의 분류에 대해 눈으로 식별되는 속성인 '가시적 속성'과 눈으로 식별되지 않는 백과사전적 정보인 '비가시적 속성'으로 나누어 설명 속성을 마련하였습니다. 이렇게 설명 속성을 정하고 이에 따라 설명 문장을 작성하는 것이 말뭉치로서 효용성이 있을지 견해를 듣고 싶습니다. 또한 추가적으로 필요한 설명 속성이 있다면 어떤 것이 있을지도 궁금합니다.

- 가시적 속성과 비가시적 속성을 나누는 형식적 효용성은 있어 보입니다. 가시적 속성은 사물의 명칭, 지시 기능을 담당하고 비가시적 속성은 개념을 담당하고 있습니다. 백과사전적 지식을 담는다는 점에서 비가시적 속성을 설명부로 활용하는 것은 사전 내용을 풍부히 한다는 점에서 유용하게 보입니다. 비가시적 속성의 추가에 대해서는 아래 3번 내용과 유사하므로 참조 가능합니다.

2. 이 사업에서는 한국어언어문화 관련 사진 자료에 대해 인간 작업자들이 설명 문장을 잘 작성할 수 있도록 「2025년 한국어언어문화 사진 설명 문장 작성 지침」을 마련하였습니다. 사진 설명 문장 작성 지침에서 수정해야 하거나 보완해야 할 점을 말씀해 주시면 지침 고도화에 큰 도움이 될 것입니다.

- 역사 부분이 매우 범위가 넓고, 사용자, 소유자 등 문맥의 주체로 등장할 만한 주어를 의미를 기준으로 나누었는데 창조자 등이 없는 이유는 무엇인가요. 또 사용자, 소유자 등도 일부분 역사로 편입 가능하므로 평가, 사건 등으로 역사 분야의 세분화를 고려해 볼 만합니다.

3. 「2025년 한국어언어문화 사진 설명 문장 작성 지침」에서 제시한 설명 속성은 ‘한국언어문화’를 한정하여 마련한 것으로, 범용으로 활용하기에는 부족한 측면이 있습니다. 만일 범용의 사진 설명 말뭉치를 구축한다면 어떠한 속성들이 필요할지 견해를 여쭙고 싶습니다.

- 조건, 변화, 출처, 관계, 평가, 재료, 도구, 방법, 시간 등의 속성 등 세분화를 고려할 수 있습니다.

- 사진 키워드는 들어가지 않는지 궁금합니다.

4. <2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구>에서는 생성형 인공지능의 한국 문화에 대한 이해를 고도화할 수 있는 신뢰도 높은 학습 자료를 구축하기 위해 한국어언어문화 관련 사진을 설명하는 말뭉치를 구축하고 있습니다. 즉, 한국어언어문화 관련 사진에 대해 1종의 대상 식별 문장과 4종의 속성 기술 문장을 작성하고 있습니다. 기존에 이와 유사한 사진 설명 말뭉치가 있는지 궁금합니다. 다른 사진 설명 말뭉치에 비추어 보았을 때, 이 설명 말뭉치의 특징은 무엇인지, 이 설명 말뭉치에 보완할 점은 무엇인지 말씀해 주시면 감사하겠습니다.

- 이미지 투 텍스트 멀티모달은 기본적으로 이미지의 의미 전달이 주요 목적이므로 별다른 문제가 없어 보입니다. 이 크로스모달 데이터 구축의 목적에 맞게 작성하면 될 듯합니다.

- 이미지가 한국문화대백과 사전에 기술된 내용을 반영하여 작성되고 있는지 궁금합니다.

5. <2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구>는 향후 유사 사업에 대한 수행 기반 마련, 국어 자료의 신뢰성과 확장성 제고, 데이터 기반 기술 생태계의 고도화 등의 효과를 기대하고 있습니다. 이러한 연구 결과를 실제로 어떠한 분야에서 어떻게

활용할 수 있을지 구체적인 활용 방안이 궁금합니다.

- 이러한 데이터가 동일 분야, 동일 개체에 관해 축적된다면 문화 정보 축적 및 전달 뿐 아니라, 시계열적 측면에서 문화 주기 데이터 등으로 활용도 가능할 것으로 추측됩니다. 문화재 정규 타입과 변이 타입의 범주화 정보, 메타데이터의 시간 정보를 통해 웹손 데이터 등으로 쓰일 수 있을 듯합니다.

6. 앞으로 이와 유사한 사업을 진행한다면 대상 자료를 어떤 방향으로 확장하면 좋을지 의견을 주시면 감사하겠습니다.

- 이미지 데이터 속성 항목을 통해 데이터 간 관계를 연결하고 이미지 베이스 지식 그래프 축적도 가능할 것입니다.

- 동영상 데이터로의 확장을 통해 모달리티의 다양성 확보도 고려해 볼 만합니다. 물품, 인물, 작품, 제도, 지명, 단체, 문헌 등 고정형 외에 놀이, 사건, 생물, 음식·약, 의례·행사 등 시계열 정보를 담고 있는 데이터로 확장할 수 있을 듯합니다.

- 개념 유형보다 회화, 음식, 민속신앙, 의생활 등에서 특히 활용성이 높을 듯합니다.

<표 16> 설명 문장 작성 지침 전문가 자문 내용

3.2. 지침 수정·보완 및 수립 내용

3.2.1. 표 설명 문장 작성 지침 수정·보완

<2025년 한국어언어문화 표 설명 문장 작성 지침>은 ‘2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구’의 제안 요청에 따라 ‘2024년 표 설명 문장 작성 말뭉치 연구 분석’에서 수립한 <2024년 표 설명 문장 작성 지침>을 수정·보완하였다.

첫째, 지침의 목적과 관련 사업을 밝히고 주요 용어의 개념을 정리하여 제시하였다. <2024년 표 설명 문장 작성 지침>에서 제시한 용어 중 ‘문서 제목’과 ‘수집 문장’을 삭제하고 아래와 같은 용어 목록을 제시하였다.

- 표: 행과 열로 배열된 정보의 집합
- 행: 표에서 가로로 배열된 원소들의 집합
- 열: 표에서 세로로 배열된 원소들의 집합

- 칸: 표에서 행과 열이 만나 이루어지는 단위로서 정보를 담는 공간(=셀)
- 제목 행: 각 열의 제목을 표시한 행
- 제목 열: 각 행의 제목을 표시한 열
- 불연속 영역: 표 안의 정보를 조합하여 기술하기 위한 둘 이상의 칸
- 표 설명 문장(서술문): 표의 핵심 내용이 드러나도록 표의 정보들을 구체적으로 설명한 문장. ‘행, 열, 불연속적인 둘 이상의 칸, 표 전체’를 기술한 문장
- 설명 영역: 표의 중요 내용을 담은 영역으로서 문장으로 표현하는 대상이 되는 영역

둘째, ‘2025년 한국언어문화 표·사진 설명 말뭉치 구축 연구’에서는 설명 대상 표 자료가 한국언어문화 관련 표로 한정됨에 따라 2024년도 지침의 예시를 ‘한국언어문화’ 관련 표로 교체하였다. 대표적인 예시는 아래와 같다.

(단위: %, 명) 동반 인원수(본인 제외)_방한 횟수별

통계 분류(1)	통계 분류(2)	2023				
		1명	2명	3명 이상	평균	
방한 횟수	1회	50.6	16.5	32.9	4.3	행 설명 영역
	2회	51.7	17.5	30.8	3.5	
	3회	54.9	19.7	25.5	2.6	
	4회 이상	61.4	18	20.6	2.2	

열 설명 영역 불연속 영역

[표 전체] 설명 문장	이 표는 2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 표이다.
[행] 설명 문장	2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 결과, 방한 횟수 1회의 동반 인원수는 1명이 50.6%로 가장 많았고 3명 이상이 32.9%, 2명이 16.5%로 그 뒤를 이었다.
[열] 설명 문장	2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 결과, 동반 인원이 1명이었던 방한 외래 관광객의 방한 횟수는 4회 이상이 61.4%로 가장 많았고 3회가 54.9%, 2회가 51.7%, 1회가 50.6%로 그 뒤를 이었다.
[불연속 영역] 설명 문장	2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 결과, 동반 인원이 3명 이상이었던 방한 외래 관광객의 방한 횟수는 1회가 32.9%로 가장 많았고 동반 인원이 2명이었던 방한 외래 관광객의 방한 횟수는 3회가 19.7%로 가장 많았다.

<표 17> <2025년 한국언어문화 표 설명 문장 작성 지침> 표 설명 문장 작성 예시

셋째, ‘2024년 표 설명 문장 작성 말뭉치 연구 분석’에서 표 설명 문장 말뭉치를 구축

하며 가장 문제가 되었던 ‘불연속 영역’에 대한 지침을 수정·보완하였다. <2024년 표 설명 문장 작성 지침>에서는 ‘불연속 영역 선정’에 대해 아래와 같이 설명하였다.

□ 불연속 영역 선정

- 불연속 영역은 떨어져 있는 두 칸 이상을 선정해야 한다.
- 이미 행 영역이나 열 영역으로 선정한 영역 내에서 불연속 영역을 선정하지 않는다.
- 새롭게 지정한 칸과 이미 지정한 행 영역 혹은 열 영역 안의 칸을 비교하는 것은 가능하다.
- 비교나 대조하여 기술할 만한 내용이 있는 두 칸 이상을 선정한다.
- 수집 문장을 참고하여 불연속 영역을 선정한다.

실제 표 설명 문장 말뭉치를 구축하는 과정에서 ‘이미 행 영역이나 열 영역으로 선정한 영역에 속하지 않으면서도, 연속되지 않는 둘 이상의 칸’을 선정하는 것이 쉽지 않았을 뿐더러, 이러한 기준으로 불연속 영역을 선정하다 보니 비교나 대조를 통해 표의 핵심적인 내용을 드러낼 수 있는 설명 문장을 작성하지 못하고, 둘 이상의 칸의 내용을 단순하게 설명하는 문장을 작성하게 되는 문제가 발생하였다. 이러한 문제를 해소하기 위해 <2025년 한국어언어문화 표 설명 문장 작성 지침>에서는 ‘불연속 영역 선정’에 관한 내용을 아래와 같이 수정·보완하였다.

□ 불연속 영역 선정

- 불연속 영역은 떨어져 있는 두 칸 이상을 선정해야 한다.
- 이미 행 영역이나 열 영역으로 선정한 영역 내에서 불연속 영역을 선정하지 않는다.
- 새롭게 지정한 칸과 이미 지정한 행 영역 혹은 열 영역 안의 칸을 비교하는 것은 가능하다.
- 행, 열의 수가 적어 이미 선정한 행 영역, 열 영역과 중복되지 않으면서 떨어져 있는 둘 이상의 칸을 선정할 수 없는 경우, 둘 이상의 연속적인 칸을 선정할 수 있다.
- 비교나 대조하여 기술할 만한 내용이 있는 두 칸 이상을 선정한다.

<2025년 한국어언어문화 표 설명 문장 작성 지침>은 <2024년 표 설명 문장 작성 지침>에서 두 차례 수정·보완의 과정을 거쳤으며, 최종적으로 <2025년 한국어언어문화 표 설명 문장 작성 지침_ver.1.1>을 완성하였다. 주요 변경 내용은 아래와 같다.

구분	일자	지침	주요 수정 내용
초안	-	2024년 표 설명 문장 작성 지침 ver.1.4	-
1차 수정	2025.07.04	2025년 표 설명 문장 작성 지침 ver.1.0	한국언어문화 표 설명 문장 작성 지침 수립 ✓ 2024년 지침에서 ‘문서 제목’, ‘수집 문장’ 관련 내용 삭제 ✓ 불연속 영역을 ‘둘 이상의 연속적인 칸’도 선정할 수 있게 확대
최종 수정	2025.12.03	2025년 표 설명 문장 작성 지침 ver.1.1	한국언어문화 관련 표로 지침의 예시 교체

<표 18> <2025년 한국언어문화 표 설명 문장 작성 지침> 주요 수정 내용

3.2.2. 사진 설명 문장 작성 지침 수립

<2025년 한국언어문화 사진 설명 문장 작성 지침>은 한국언어문화 관련 사진 설명 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 하여, 한국언어문화 중 생활 분류인 식생활, 의생활, 주생활과 예술 분류인 건축, 공예 관련 문화 요소의 사진에 대한 설명 문장을 작성하는 방안을 기술하였다.

첫째, 지침의 목적과 관련 사업을 밝히고 주요 용어의 개념을 정리하여 제시하였다. 사진 설명 문장 작성에 사용되는 주요 용어는 아래와 같이 정의하였다.

- 한국언어문화: 한국의 대표적인 유무형 문화, 한국 전통문화 및 한류를 이루는 문화
- 설명 대상: 사진 속의 사물 중 설명의 대상으로 삼은 사물
- 설명 속성: 설명 문장을 작성하기 위해 구분해 놓은 대상의 특성
- 설명 문장: 사진 속 대상에 대해 설명한 문장
- 대상 식별 문장: 사진 속 대상이 무엇인지 식별하는 설명 문장
- 속성 기술 문장: 사진 속 대상의 특정한 속성에 대해 기술하는 설명 문장


둘째, 사진 설명 문장 작성의 대표적인 예를 제시하였다. 이번 사업에서는 ‘2025년 한국언어문화 표·사진 설명 말뭉치 구축 연구’의 제안 요청서에 따라 1종의 대상 식별 문장과 4종의 속성 기술 문장을 작성하도록 하였다. 한국언어문화 사진 설명 문장의 작성

예시는 아래와 같다.

생활 > 식생활

↓

분류 태그



전주비빔밥

사진 제목(캡션)

대상 식별 문장	이 사진은 한국 전통 음식인 전주비빔밥의 사진이다.
속성 기술 문장	색채 이 전주비빔밥 재료의 색채는 노란색, 녹색, 하얀색, 붉은색 등으로 다채롭다.
	형태 이 전주비빔밥에 들어 있는 달걀지단과 오이는 긴 직사각형 모양이다.
	재료 이 전주비빔밥에는 콩나물, 시금치, 오이, 달걀지단, 깨 등이 들어 있다.
	취식 방법 전주비빔밥은 밥, 소고기, 콩나물, 시금치, 고사리 등의 모든 재료를 고추장 양념과 함께 비벼서 먹는다.

↓

설명 속성

[그림 18] 사진 설명 문장 작성 예시

셋째, 사진 설명 문장의 유형을 사진의 객체를 설명 대상으로 식별하는 문장인 ‘대상 식별 문장’과 사진 속 대상에서 관찰되는 한국언어문화 요소로서의 속성을 설명한 ‘속성 기술 문장’으로 구분하고, 각 유형의 설명 문장을 어떻게 작성해야 하는지 제시하였다. ‘대상 식별 문장’은 사진 속에 사물이 한 개인 경우와 여러 개인 경우로 나누어 대상을 식별하고 그 식별 문장을 작성하는 방안을 기술하였고, ‘속성 기술 문장’은 설명 속성의 특성에 따라 크게 ‘가시적 속성 기술 문장’과 ‘비가시적 속성 기술 문장’으로 나누어 작성하는 방안을 기술하였다.

□ 대상 식별 문장 작성 방법

- 대상 식별 문장은 한국언어문화 요소로서의 ‘유형 및 명칭’을 기술한다.

- 사진 속 한국어언어문화 요소의 명칭이 대상 식별 문장에 포함되도록 작성한다.
- 가능하면 유개념과 종 차의 관계가 드러나도록 작성한다.

□ 속성 기술 문장 작성 방법

○ 가시적 속성 기술 문장 작성 요령

- 사진 속 대상에서 관찰되는 한국어언어문화 요소로서의 특징을 설명한다. 즉, 사진에서 가시적으로 드러나는 대상의 특징을 기술하는 문장으로 작성한다.
- 사진 속 한국어언어문화 요소에서 두드러지게 드러나는 특성을 묘사한다.
- 한국어언어문화 요소의 유형에 따라 그 특성을 잘 드러낼 수 있는 속성을 우선적으로 기술한다.

○ 비가시적 속성 기술 문장 작성 요령

- 사진에서는 가시적으로 드러나지 않지만, 사진 속 한국어언어문화 요소를 설명할 때 긴요한 백과사전적 정보를 기술한다.
 - ✓ 단, 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 선택하여 설명 문장을 작성할 수 있다. 이때에는 설명 문장에서 그 속성이 해당 한국어언어문화 요소의 일반적인 속성임을 드러내야 한다.

넷째, 사진 설명 문장을 작성하기 위한 설명 속성을 생활 > 식생활 / 의생활 / 주생활, 예술 > 건축, 공예 등의 한국어언어문화 분류별로 마련하고, 각각에 대한 설명 문장을 작성하는 방안을 제시하였다.

□ 식생활 설명 속성

가시적 속성	비가시적 속성
색채 형태 재료(주재료, 부재료, 양념, 고명) * 용기(액체류, 장류/소스류 한정) * 구성 요소(도구류 한정) * 소재(素材)(도구류 한정) * 문양(도구류 한정)	조리법 맛 취식 방법 지역 역사 * 용도(도구류 한정) * 이용자(도구류 한정)

○ 가시적 속성

- 색채: 사진 속 음식에서 두드러지게 드러나는 빛깔을 기술한다. 음식의 전체적인 색채를 설명할 수도 있고 음식 재료의 색채를 설명할 수도 있다.
- 형태: 사진 속 음식의 모양을 기술한다. 음식의 전체적인 모양을 설명할 수도 있고 음식 재료의 모양이나 손질된 형태를 설명할 수도 있다.
- 재료: 사진 속 음식에서 두드러지게 드러나는 재료를 기술한다. 하위 설명 속성으로 주재료, 부재료, 양념, 고명 등을 별도의 설명 문장으로 작성할 수 있다.
- 용기: 사진 속 설명 대상을 담고 있는 그릇에 대해 기술한다. 가시적 속성 기술 문장으로 '색채'만 작성할 수 있는 '액체류'와 '장류' 및 '소스류'에 한해 예외적으로 그릇에 대한 설명 문장을 작성할 수 있다.
- 구성 요소: 사진 속 식생활 관련 도구를 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 소재(素材): 사진 속 식생활 관련 도구를 만드는 데에 바탕이 되는 재료를 기술하되, 사진에서 도구의 소재가 식별되는 경우에 한해 설명 문장을 작성한다. 도구의 전체적인 소재를 설명할 수도 있고 구성 요소의 소재를 설명할 수도 있다.
- 문양: 사진 속 식생활 관련 도구 표현의 무늬를 기술한다. 도구에서 전체적으로 나타나는 문양 및 무늬를 설명할 수도 있고 구성 요소에서 나타나는 문양을 설명할 수도 있다.

○ 비가시적 속성

- 조리법: 사진 속 음식을 만드는 방법을 기술하되, 핵심적인 조리법 위주로 간단하게 작성한다.
- 맛: 사진 속 음식에서 두드러지게 느껴지는 맛을 기술한다. 해당 음식의 일반적인 맛을 기술할 수 있도록 한다.
- 취식 방법: 사진 속 음식을 먹는 방법 및 순서를 기술한다. 구절판, 열구자탕 등과 같이 특수한 방법으로 먹는 음식의 경우 취식 방법을 설명할 수 있다.
- 지역: 사진 속 음식이 유래한 지역이나 발달한 지역을 기술한다. 닭갈비, 갓김치 등 지역색이 강한 음식의 경우 지역을 설명할 수 있다.
- 역사: 사진 속 음식과 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와

관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.

- 용도: 사진 속 도구의 쓰임새를 기술한다. 해당 도구의 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
- 이용자: 사진 속 도구를 주로 이용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 이용자가 한정적인 경우 이용자를 설명 문장으로 작성할 수 있다.

□ 의생활 설명 속성

가시적 속성	비가시적 속성
구성 요소 크기 길이 소재(素材) 형태 문양 색채	용도 착용 방법 착용자 역사

○ 가시적 속성

- 구성 요소: 사진 속 의류 및 잡화를 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 크기: 사진 속 의류 및 잡화의 넓이, 부피 등이 크고 작은 정도를 기술한다. 구성 요소들의 크기를 비교, 대조하여 설명하거나 착용 시 품을 고려하여 설명 문장을 작성할 수 있다.
- 길이: 사진 속 의류 및 잡화의 한끝에서 다른 한끝까지의 거리를 기술한다. 구성 요소들의 길이를 비교, 대조하여 설명하거나 착용 시 기장을 활용하여 설명 문장을 작성할 수 있다.
- 소재(素材): 사진 속 의류 및 잡화를 만드는 데에 바탕이 되는 재료를 기술하되, 사진에서 의류 및 잡화의 소재가 식별되는 경우에 설명 문장을 작성한다. 소재의 얇고 두꺼운 정도가 식별이 되는 경우, 소재와 함께 소재의 두께도 함께 기술할 수 있다. 의류 및 잡화의 전체적인 소재를 설명할 수도 있고 구성 요소의 소재를 설명할 수도 있다.

- 형태: 사진 속 의류 및 잡화에서 두드러지게 드러나는 모양을 기술한다. 의류 및 잡화의 전체적인 모양을 설명할 수도 있고 구성 요소의 모양을 설명할 수도 있다.
- 문양: 사진 속 의류 및 잡화 표면의 무늬를 기술한다. 의류 및 잡화에서 전체적으로 나타나는 문양을 설명할 수도 있고 구성 요소에서 나타나는 문양을 설명할 수도 있다.
- 색채: 사진 속 의류 및 잡화에서 두드러지게 드러나는 빛깔을 기술한다. 의류 및 잡화의 전체적인 색채를 설명할 수도 있고 구성 요소의 색채를 설명할 수도 있다.

○ 비가시적 속성

- 용도: 사진 속 의류 및 잡화의 쓰임새를 기술한다. 해당 의류 및 잡화의 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
- 착용 방법: 사진 속 의복 및 장신구를 입거나 차는 방법 및 순서를 기술한다. 특수한 방식으로 착용하는 의복 및 장신구의 경우 착용 방법을 설명할 수 있다.
- 착용자: 사진 속 의복 및 장신구를 주로 착용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 착용자가 한정적인 경우 착용자를 설명 문장으로 작성할 수 있다.
- 역사: 사진 속 의류 및 잡화와 관련된 역사적 사실을 기술한다. 해당 한국어언어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.

□ 주생활 설명 속성

가시적 속성	비가시적 속성
형태 구성 요소 구조(외부/내부) 자재(건축물)/소재(素材)(주생활 관련 도구) 색채 문양 배치(건축물의 집합 건축물에서의 배치) 위치(건축물 일부의 건축물 안에서의 위치)	용도 소재지(所在地) 분포 이용자 소유자 역사

○ 가시적 속성

- 형태: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구의 모양을 기술한다. 건축물 및 건축물 일부, 주생활 관련 도구의 전체적인 모양을 설명할 수도 있고 구성 요소의 모양을 설명할 수도 있다.
- 구성 요소: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 구조(외부/내부): 사진 속 건축물 및 건축물 일부의 구성 요소들을 엮은 체계를 기술한다. 외부 구조를 설명할 수도 있고 내부 구조를 설명할 수도 있다.
- 자재/소재(素材): 사진 속 건축물 및 건축물 일부에 쓰인 재료, 주생활 관련 도구의 재료를 기술하되, 사진에서 자재 혹은 소재가 식별되는 경우에 설명 문장을 작성한다. 건축물 및 건축물 일부, 주생활 관련 도구의 주자재를 설명할 수도 있고 부자재를 설명할 수도 있다.
- 색채: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구에서 두드러지게 드러나는 빛깔을 기술한다. 건축물 및 건축물 일부, 주생활 관련 도구의 전체적인 색채를 설명할 수도 있고 구성 요소의 색채를 설명할 수도 있다.
- 문양: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 장식하기 위해 사용된 표면의 문양 혹은 무늬를 기술한다. 사진 속 대상에서 전체적으로 나타나는 문양 및 무늬를 설명할 수도 있고 구성 요소에서 나타나는 문양 및 무늬를 설명할 수도 있다.
- 배치(건축물의 집합 건축물에서의 배치): 사진 속 건축물이 집합 건축물 내에서 벌여 놓여 있는 형태를 기술하되, 사진에서 건축물이 어떻게 배치되어 있는지 식별되는 경우에 설명 문장을 작성한다.
 - ✓ 단, 해당 건축물 일반의 배치 형태가 한국어언어문화 요소로서 유의미한 경우(안채-집의 가장 안쪽, 사랑채-대문과 가까운 곳) 사진에서 건축물의 배치가 식별되지 않아도 속성 기술 문장으로 배치를 작성할 수 있다.
- 위치(건축물 일부의 건축물 안에서의 위치): 사진 속 건축물 일부의 건축물 안에서의 위치를 기술한다. 지붕 아래, 처마 밑 등과 같이 설명할 수 있다.

○ 비가시적 속성

- 용도: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구의 쓰임새를 기술한다. 사진 속 대상이 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
- 소재지(所在地): 사진 속 건축물 및 건축물 일부가 자리 잡고 있는 곳을 기술한다. 갈암 종택, 만성 고택 등과 같은 고유 건축물의 경우 건축물의 소재지를 설명할 수 있다.
- 분포: 사진 속 건축물이 주로 나타나는 지역을 기술한다. 특정 지역에서 주로 관찰되는 건축물의 경우 분포를 설명할 수 있다.
- 이용자: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 주로 이용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 이용자가 한정적인 경우 이용자를 설명 문장으로 작성할 수 있다.
- 소유자: 사진 속 건축물을 소유했던, 소유하고 있는 사람을 기술한다. 갈암 종택, 추산 고택 등 특정인이 거주했던 건축물의 경우 소유자를 기술할 수 있다.
- 역사: 사진 속 건축물 및 건축물 일부와 관련된 역사적 사실을 기술한다. 해당 한국언어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.

□ 건축 설명 속성

가시적 속성	비가시적 속성
형태	용도
구성 요소	소재지(所在地)
구조(외부/내부)	이용자
자재/소재(素材)	소유자
색채	건축가
문양	건축 양식
배치(집합 건축물을 이루는 단일 건축물들의 배치)	역사

○ 가시적 속성

- 형태: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 모양을 기술한다. 전체적인 모양을 설명할 수도 있고 구성 요소의 모양을 설명할 수도 있다.

- 구성 요소: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 구조(외부/내부): 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 구성 요소들을 엮은 체계를 기술한다. 외부 구조를 설명할 수도 있고 내부 구조를 설명할 수도 있다.
 - ✓ 건축 관련 전문 용어를 활용하여 설명 문장을 작성할 경우, 잘못된 정보가 전달되지 않도록 해당 용어의 개념을 명확하게 파악하고 설명 문장을 작성하도록 한다.
- 자재/소재(素材): 사진 속 집합 건축물, 단일 건축물 및 건축물 일부에 쓰인 재료를 기술하되, 사진에서 자재가 식별되는 경우에 한해 설명 문장을 작성한다. 건축물 및 건축물 일부의 주자재를 설명할 수도 있고 부자재를 설명할 수도 있다.
- 색채: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부에서 두드러지게 드러나는 빛깔을 기술한다. 집합 건축물, 단일 건축물 및 건축물 일부의 전체적인 색채를 설명할 수도 있고 구성 요소의 색채를 설명할 수도 있다.
- 문양: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 장식하기 위해 사용된 표면의 문양 혹은 무늬를 기술한다. 사진 속 대상에서 전체적으로 나타나는 문양 및 무늬를 설명할 수도 있고 구성 요소에서 나타나는 문양 및 무늬를 설명할 수도 있다.
- 배치(집합 건축물을 이루는 단일 건축물들의 배치): 사진 속 집합 건축물을 이루는 단일 건축물들이 벌여 놓여 있는 모양을 기술하되, 사진에서 건축물이 어떻게 배치되어 있는지 식별되는 경우에 설명 문장을 작성한다.

○ 비가시적 속성

- 용도: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 쓰임새를 기술한다. 사진 속 대상이 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
- 소재지(所在地): 사진 속 집합 건축물, 단일 건축물 및 건축물 일부가 자리 잡고 있는 곳을 기술한다. ‘충청북도 보은군 속리산면’ 등과 같이 소재가 명확한 경우에

소재지를 설명 문장으로 작성할 수 있다.

- 이용자: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 주로 이용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 이용자가 한정적인 경우 이용자를 설명 문장으로 작성할 수 있다.
- 소유자: 사진 속 건축물을 소유했던, 소유하고 있는 사람을 기술한다. 석파정, 팔락정 등 특정인이 거주하거나 이용했던 건축물의 경우 소유자 설명 문장을 작성할 수 있다.
- 건축가: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 설계하거나 지은 사람을 기술한다. 나상기, 박길룡 등 건축가의 구체적인 명칭이 확인되는 경우 건축가 설명 문장을 작성할 수 있다.
- 건축 양식: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부가 지어진 시대나 종류에 따라 독특하게 지니는 건축의 형식을 기술한다. 사진 속 건축물이 특정 양식으로 지어진 경우 건축 양식 설명 문장을 작성할 수 있다.
- 역사: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부와 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사 설명 문장을 작성할 수 있다.

□ 공예 설명 속성

가시적 속성	비가시적 속성
구성 요소 크기 길이 소재(素材) 형태 문양 색채	용도 기법 소재지(所在地) 역사

○ 가시적 속성

- 구성 요소: 사진 속 공예품을 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.

- 크기: 사진 속 공예품의 넓이, 부피 등이 크고 작은 정도를 기술한다. 해당 공예품의 구성 요소들의 크기를 비교, 대조하는 방식으로 크기 설명 문장을 작성할 수 있다.
- 길이: 사진 속 공예품의 한끝에서 다른 한끝까지의 거리를 기술한다. 해당 공예품 부류의 일반적인 길이와 비교하거나 구성 요소들의 길이를 비교, 대조하는 방식으로 길이 설명 문장을 작성할 수 있다.
- 소재(素材): 사진 속 공예품을 만드는 데에 바탕이 되는 재료를 기술하되, 사진에서 해당 공예품의 소재가 식별되는 경우에 설명 문장을 작성한다. 소재의 질감이 식별이 되는 경우, 소재와 함께 소재의 질감도 함께 기술할 수 있다. 공예품의 전체적인 소재를 설명할 수도 있고 구성 요소의 소재를 설명할 수도 있다.
- 형태: 사진 속 공예품에서 두드러지게 드러나는 모양을 기술한다. 해당 공예품의 전체적인 모양을 설명할 수도 있고 구성 요소의 모양을 설명할 수도 있다.
- 문양: 사진 속 공예품 표면의 무늬 및 문양을 기술한다. 해당 공예품에서 전체적으로 나타나는 무늬 및 문양을 설명할 수도 있고 구성 요소에서 나타나는 무늬 및 문양을 설명할 수도 있다.
- 색채: 사진 속 공예품에서 두드러지게 드러나는 빛깔을 기술한다. 해당 공예품의 전체적인 색채를 설명할 수도 있고 구성 요소의 색채를 설명할 수도 있다.

○ 비가시적 속성

- 용도: 사진 속 공예품의 쓰임새를 기술한다. 해당 공예품의 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
 - ✓ 그릇, 연적 등과 같이 실용적인 목적에서 만들어진 공예품이나 외형만으로는 용도를 알 수 없는 공예품의 경우 되도록 용도 설명 문장을 작성하도록 한다.
- 기법: 사진 속 공예품을 만들거나 공예품에 문양을 넣는 데에 활용한 기교와 방법을 기술한다. 해당 공예품이 특정 기법을 사용해 만들어진 경우 기법을 설명 문장으로 작성할 수 있다.
 - ✓ 기법은 ‘상감’, ‘은입사’ 등과 같은 특정 기법을 설명하는 것으로, 해당 공예품을 만드는 일반적인 방법을 작성하지 않도록 주의한다.

- 소재지(所在地): 사진 속 공예품이 위치하고 있는 장소, 혹은 해당 공예품을 소장하고 있는 기관을 기술한다. 동종, 범종 등의 경우에는 해당 공예품이 위치하고 있는 장소, 즉 소재지를 기술하고, 백자, 사리함, 연적 등의 경우에는 해당 공예품을 소장하고 있는 기관을 기술한다.
- 역사: 사진 속 공예품과 관련된 역사적 사실을 기술한다. 해당 한국언어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.

다섯째, 사진 설명 문장 말뭉치 구축자가 참고할 수 있도록 사진 설명 문장 작성 시 주의해야 할 사항을 첨부하였다.

□ 설명 문장 관련 주의 사항

- 백과사전적 정보는 이미 많이 축적되어 있고, 국립국어원 온용어 등의 사업에서 구축하고 있으므로 되도록 백과사전적 정보는 배제하고 최대한 사진의 대상을 묘사하는 가시적 속성 기술 문장을 작성한다.
- 사진 제목에 ‘정면’, ‘정측면’, ‘내부’, ‘전경’ 등과 같은 정보가 포함되어 있는 경우, 이를 제외하고 한국언어문화 요소인 대상의 명칭만을 사용하여 설명 문장을 작성한다.
- 사진 제목이 한국에서 두루 사용되는 일반적인 명칭이 아닌 경우 일반적인 명칭으로 바꾸어 설명 문장을 작성한다.
- 사진 제목이 특정 대상의 고유한 명칭이어서 한국언어문화 요소의 특성을 기술하는 것의 효용이 떨어지는 경우, 일반적인 명칭으로 설명 문장을 작성한다.

□ 맞춤법 관련 주의 사항

- 설명 문장은 어문 규범을 준수하여 작성한다.
- 전문 용어 및 고유명은 경우에 따라 붙여쓰기와 띄어쓰기를 모두 허용하는 경우가 있으나, 원칙과 허용 중 원칙에 따라 띄어쓰기한다.

<2025년 한국언어문화 사진 설명 문장 작성 지침>은 초안 작성 이후 네 차례에 걸쳐 수정하여 최종적으로 <2025년 한국언어문화 사진 설명 문장 작성 지침_ver.1.3>을 완성하였다. 버전별 주요 수정 내용은 아래와 같다.

구분	일자	지침	주요 수정 내용
초안 작성	2025.07.15	사진 설명 문장 작성 지침 ver.1.0	<ul style="list-style-type: none"> 1) 한국어언어문화 대분류 생활>중분류 식생활, 의생활, 주생활에 대한 지침 수립 ✓ 대상 식별 문장, 가시적 속성 기술 문장, 비가시적 속성 기술 문장 등의 개념 정리 ✓ 식생활, 의생활, 주생활에 대한 설명 속성 마련 ✓ 대상 식별 문장, 가시적 속성 기술 문장, 비가시적 속성 기술 문장 작성 방법 마련
1차 수정	2025.07.23	사진 설명 문장 작성 지침 ver.1.1	<ul style="list-style-type: none"> 1) 외부적 요인: 국어원의 검토 의견 반영 내부적 요인: 한림대 언어팀 자체 회의 2) 주요 변경 사항 <ul style="list-style-type: none"> ✓ 액체류,장/소스류 설명 문장 작성 문제의 해소를 위해 식생활 설명 속성에 '용기'추가 ✓ 소재지 정보가 포함된 사진의 설명 문장 작성 시 주의 사항 추가 ✓ 사진 제목이 특정 대상의 고유 명칭인 경우의 예외 조항 추가
2차 수정	2025.08.11	사진 설명 문장 작성 지침 ver.1.2	<ul style="list-style-type: none"> 1) 외부적 요인: 국어원의 검토 의견 반영 내부적 요인: 한림대 언어팀 자체 회의 2) 주요 변경 사항 <ul style="list-style-type: none"> ✓ 식생활 관련 도구 설명 문장 작성을 위한 속성 추가(구성 요소, 소재, 문양/용도, 이용자) ✓ 식, 의, 주 각 설명 문장 작성 내용에 '도구' 추가 ✓ 가시적 속성을 비가시적 속성 설명 문장을 작성할 때 활용할 수 있다는 내용 추가
3차 수정	2025.09.16	사진 설명 문장 작성 지침 ver.1.3	<ul style="list-style-type: none"> 1) 외부적 요인: 국어원의 검토 의견 반영 내부적 요인: 한림대 언어팀 자체 회의 2) 주요 변경 사항 <ul style="list-style-type: none"> ✓ 한국어언어문화 대분류 예술 >중분류 건축, 공예에 대한 지침 수립 ✓ 건축, 공예에 대한 설명 속성 마련 ✓ 대상 식별 문장, 가시적 속성 기술 문장, 비가시적 속성 기술 문장 작성 방법 마련 ✓ 국어원의 검토 의견을 반영하여 '공예'의 '특징' 설명 속성 삭제
4차 수정	2025.10.27	사진 설명 문장 작성 지침 ver.1.3(저용량)	<ul style="list-style-type: none"> 1) 외부적 요인: 국어원의 검토 의견 반영 2) 주요 변경 사항 <ul style="list-style-type: none"> ✓ 사진 용량 축소 ✓ 오탈자 등 최종 교정

<표 19> <2025년 한국어언어문화 사진 설명 문장 작성 지침> 주요 수정 내용
<2025년 한국어언어문화 사진 설명 문장 작성 지침_ver.1.0>에서는 한국어언어문화 분류

중 ‘생활’의 ‘식생활, 의생활, 주생활’에 대한 사진 설명 문장 작성 지침을 우선적으로 수립하였다. ‘대상 식별 문장’, ‘가시적 속성 기술 문장’, ‘비가시적 속성 기술 문장’ 등의 개념을 정리하고, 식생활, 의생활, 주생활 각각에 대한 설명 속성 및 설명 문장 작성 방안을 마련하였다.

<2025년 한국언어문화 사진 설명 문장 작성 지침_ver.1.1>은 지침 1.0에 대한 국립국어원의 검토 의견을 반영하고 구축 작업 시 발생한 문제점을 해소하는 방향으로 지침을 수정·보완한 결과물이다. 가시적 속성 기술 문장을 작성하기 어려운 ‘액체류, 장/소스류’를 위해 식생활 설명 속성에 ‘용기’를 추가하였다. 또한 주생활에서 소재지 정보가 포함된 사진 설명 문장 작성 시 주의 사항을 지침에 추가하였다.

<2025년 한국언어문화 사진 설명 문장 작성 지침_ver.1.2>는 말뭉치 구축 과정에서 식생활, 의생활, 주생활 관련 도구들이 구축 대상에 추가됨에 따라 이를 보완하기 위해 기존 지침을 수정한 버전이다. 식생활 관련 도구에 대한 설명 문장을 작성하기 위해 ‘구성 요소, 소재, 문양, 용도, 이용자’의 설명 속성을 추가하였고, 설명 문장 작성 내용에 ‘도구’를 추가하였다.

<2025년 한국언어문화 사진 설명 문장 작성 지침_ver.1.3>에서는 한국언어문화 분류 중 ‘예술’의 ‘건축, 공예’에 대한 사진 설명 문장 작성 지침 수립이 주로 이루어졌다. 건축과 공예에 대한 설명 속성을 마련하였고, 각 분류별로 ‘대상 식별 문장’, ‘가시적 속성 기술 문장’, ‘비가시적 속성 기술 문장’을 작성하는 구체적인 방안을 제시하였다. 마지막으로 사진 용량을 줄이고 오탈자 등을 최종적으로 교정하여 <2025년 한국언어문화 사진 설명 문장 작성 지침_ver.1.3(저용량)>을 완성하였다.

3.3. 지침 수정·보완 및 수립 결과

<2025년 한국언어문화 표 설명 문장 작성 지침>과 <2025년 한국언어문화 사진 설명 문장 작성 지침>은 다음과 같은 목차로 이루어져 있으며 지침의 전문은 부록에서 확인할 수 있다.

□ <2025년 한국언어문화 표 설명 문장 작성 지침> 목차

<목차>

1. 개요
 - 1.1. 지침의 목적
 - 1.2. 관련 사업
 - 1.3. 주요 용어
 - 1.4. 표의 다양한 유형
 - 1.5. 표 설명 문장 작성 예시
2. 표 설명 문장 작성 방법
 - 2.1. 표 전체에 대한 서술문
 - 2.2. 행에 대한 서술문
 - 2.2.1. 행 영역 선정
 - 2.2.2. 행 설명 문장 작성
 - 2.3. 열에 대한 서술문
 - 2.3.1. 열 영역 선정
 - 2.3.2. 열 설명 문장 작성
 - 2.4. 불연속 영역에 대한 서술문
 - 2.4.1. 불연속 영역 선정
 - 2.4.2. 불연속 영역 설명 문장 작성
 - 2.5. 주의 사항
 - 2.5.1. 설명 문장 관련 주의 사항
 - 2.5.2. 기호 관련 주의 사항

□ <2025년 한국언어문화 사진 설명 문장 작성 지침> 목차

<목차>

1. 개요
 - 1.1. 지침의 목적
 - 1.2. 관련 사업
 - 1.3. 주요 용어
 - 1.4. 사진 설명 문장 예시
2. 사진 설명 문장의 유형
 - 2.1. 대상 식별 문장
 - 2.2. 속성 기술 문장
3. 대상 식별 문장 작성의 실제
 - 3.1. 대상 식별 방법
 - 3.1.1. 사물이 한 개인 경우
 - 3.1.2. 사물이 여러 개인 경우
 - 3.2. 대상 식별 문장 작성 방법
4. 속성 기술 문장 작성의 실제
 - 4.1. 속성 기술 문장 작성의 일반 원리
 - 4.1.1. 가시적 속성 기술 문장 작성
 - 4.1.2. 비가시적 속성 기술 문장 작성
 - 4.2. 식생활 설명 문장

4.2.1. 식생활 설명 속성

4.2.2. 식생활 속성 기술 문장 작성 방법

4.2.2.1. 가시적 속성 기술 문장 작성 방법

4.2.2.1.1. 색채

4.2.2.1.2. 형태

4.2.2.1.3. 재료

4.2.2.1.4. 용기

4.2.2.1.5. 구성 요소

4.2.2.1.6. 소재(素材)

4.2.2.1.7. 문양

4.2.2.2. 비가시적 속성 기술 문장 작성 방법

4.2.2.2.1. 조리법

4.2.2.2.2. 맛

4.2.2.2.3. 취식 방법

4.2.2.2.4. 지역

4.2.2.2.5. 역사

4.2.2.2.6. 용도

4.2.2.2.7. 이용자

4.3. 의생활 설명 문장

4.3.1. 의생활 설명 속성

4.3.2. 의생활 속성 기술 문장 작성 방법

4.3.2.1. 가시적 속성 기술 문장 작성 방법

4.3.2.1.1. 구성 요소

4.3.2.1.2. 크기

4.3.2.1.3. 길이

4.3.2.1.4. 소재(素材)

4.3.2.1.5. 형태

4.3.2.1.6. 문양

4.3.2.1.7. 색채

4.3.2.2. 비가시적 속성 기술 문장 작성 방법

4.3.2.2.1. 용도

4.3.2.2.2. 착용 방법

4.3.2.2.3. 착용자

4.3.2.2.4. 역사

4.4. 주생활 설명 문장

4.4.1. 주생활 설명 속성

4.4.2. 주생활 속성 기술 문장 작성 방법

4.4.2.1. 가시적 속성 기술 문장 작성 방법

4.4.2.1.1. 형태

- 4.4.2.1.2. 구성 요소
- 4.4.2.1.3. 구조
- 4.4.2.1.4. 자재/소재(素材)
- 4.4.2.1.5. 색채
- 4.4.2.1.6. 문양
- 4.4.2.1.7. 배치(건축물의 집합 건축물에서의 배치)
- 4.4.2.1.8. 위치(건축물 일부의 건축물 안에서의 위치)
- 4.4.2.2. 비가시적 속성 기술 문장 작성 방법
 - 4.4.2.2.1. 용도
 - 4.4.2.2.2. 소재지(所在地)
 - 4.4.2.2.3. 분포
 - 4.4.2.2.4. 이용자
 - 4.4.2.2.5. 소유자
 - 4.4.2.2.6. 역사
- 4.5. 건축 설명 문장
 - 4.5.1. 건축 설명 속성
 - 4.5.2. 건축 속성 기술 문장 작성 방법
 - 4.5.2.1. 가시적 속성 기술 문장 작성 방법
 - 4.5.2.1.1. 형태
 - 4.5.2.1.2. 구성 요소
 - 4.5.2.1.3. 구조
 - 4.5.2.1.4. 자재/소재(素材)
 - 4.5.2.1.5. 색채
 - 4.5.2.1.6. 문양
 - 4.5.2.1.7. 배치(집합 건축물을 이루는 단일 건축물들의 배치)
 - 4.5.2.2. 비가시적 속성 기술 문장 작성 방법
 - 4.5.2.2.1. 용도
 - 4.5.2.2.2. 소재지(所在地)
 - 4.5.2.2.3. 이용자
 - 4.5.2.2.4. 소유자
 - 4.5.2.2.5. 건축가
 - 4.5.2.2.6. 건축 양식
 - 4.5.2.2.7. 역사
- 4.6. 공예 설명 문장
 - 4.6.1. 공예 설명 속성
 - 4.6.2. 공예 속성 기술 문장 작성 방법
 - 4.6.2.1. 가시적 속성 기술 문장 작성 방법
 - 4.6.2.1.1. 구성 요소
 - 4.6.2.1.2. 크기

4.6.2.1.3. 길이

4.6.2.1.4. 소재(素材)

4.6.2.1.5. 형태

4.6.2.1.6. 문양

4.6.2.1.7. 색채

4.6.2.2. 비가시적 속성 기술 문장 작성 방법

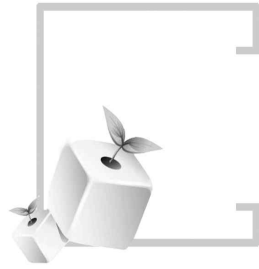
4.6.2.2.1. 용도

4.6.2.2.2. 기법

4.6.2.2.3. 소재지(所在地)

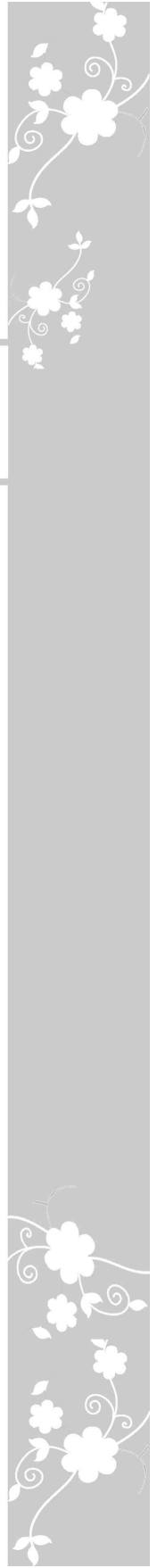
4.6.2.2.4. 역사

5. 주의 사항



제 4 장

결론



4.1. 사업 요약

본 사업은 국립국어원이 주관하고 데이터리와 한림대학교가 공동 참여하여 약 6개월 동안 추진되었으며, 한국어언어문화 기반의 표·사진 설명 문장 말뭉치 구축을 위한 체계적 과정을 수립했다. 전체 과정은 멀티모달 말뭉치 구축을 위한 기반 체계를 마련하는 데 중점을 두었고, 다음과 같은 성과를 달성했다.

첫째, 한국어언어문화 관련 표·사진 자료 대규모 수집과 정제를 진행했다. 사업 초기부터 다양한 분야의 자료를 폭넓게 수집한 뒤, 중복·부적합·메타데이터 불완전 자료를 전수 확인하여 실제 구축에 활용 가능한 데이터만을 선별했다. 표 자료는 원문(xlsx)과 메타데이터를 결합한 단일 JSON 구조로 재가공했고, 사진 자료는 JPG 원문과 메타데이터를 매칭해 사진 단위 JSON으로 정리하여 멀티모달 도구에서 활용 가능한 형태로 일원화했다.

둘째, 수집된 자료를 활용하여 표·사진 각각의 특성을 반영한 설명 문장 말뭉치를 구축했다. 표의 경우 ‘표 전체, 행, 열, 불연속 영역’ 4개 영역에 대한 설명 문장을 모두 작성하도록 구성하여 표 해석에 필요한 정보를 다각도로 제공하는 말뭉치를 구축했고, 사진은 대상 식별 문장과 가시·비가시 속성을 아우르는 속성 설명 문장을 작성해 멀티모달 학습에 적합한 구조로 정제된 데이터를 마련했다. 모든 작업은 <국립국어원 한국어 멀티모달 말뭉치 구축 지원 도구>에서 수행되었으며, 작업-검수-재검수-전수 검토의 단계적 절차를 통해 품질을 체계적으로 관리했다. 이를 통해 표 2,037건, 사진 13,005건의 최종 구축 데이터를 확보했다.

셋째, 표·사진 설명 문장 작성을 위한 지침을 개선하고 확장했다. 표 설명 문장 지침은 기존 2024년 작성 지침을 기반으로 설명 영역 선정 방식과 서술 규칙을 보완하여 실제 작업 가능성을 높였다. 사진 설명 지침은 이번 사업에서 새롭게 정립되었으며, 속성 체계 정의, 가시·비가시적 설명 문장 작성 규칙 등 사진 데이터 특성을 반영한 실용적인 지침이 완성됐다. 두 지침 모두 실제 구축 과정에서 수차례 검토·보완되며 현장에서 활용 가능한 형태로 발전했다.

넷째, 구축된 말뭉치의 품질을 검증하기 위해 BLEU 기반 유사도 평가, BERTScore 기반 유사도 평가, METEOR 기반 의미 확장 유사도 평가, Perplexity 기반 유창성 평가를 적용했다. 모든 평가지표는 사전에 설정한 목표치를 충족했으며, 이를 통해 본 사업에서 구축한 설명 문장이 전문가 기준과 높은 일치도를 보이고 의미적 정확성과 문장 자연성 측면에서도 충분한 품질을 확보했음을 확인했다. 또한 표 설명 문장의 경우 정합성

및 관계 추론 검증을 추가로 수행하여 수치 기입 오류와 최댓값·최솟값 지정 오류 등을 검출하고 수정함으로써 말뭉치의 정확성을 높였다.

본 사업은 한국어 멀티모달 말뭉치 구축을 위한 핵심 인프라를 마련하고, 인공지능이 한국어언어문화를 보다 정교하게 이해하고 학습할 수 있는 기반을 조성했다는 점에서 중요한 의의를 가진다. 나아가 향후 AI 학습, 연구, 서비스 개발 등 다양한 분야에서 활용될 수 있는 토대를 마련했으며, 국립국어원-데이터리-한림대학교의 협력을 통해 구축된 이번 말뭉치는 앞으로 한국어 기반 멀티모달 연구의 표준 자료로서 중요한 역할을 수행할 것으로 기대된다.

4.2. 의의 및 기대 효과

□ 의의

‘2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구’는 기존의 일반 분야를 대상으로 한 표·사진 설명 말뭉치에서 한 단계 더 나아가 ‘한국어언어문화’라는 특정 주제에 특화된 설명 말뭉치를 구축하고, 이를 위한 지침을 새롭게 수립했다는 점에서 의의가 있다.

<2025년 한국어언어문화 표 설명 문장 작성 지침>은 국가통계포털(KOSIS) 등의 통계 포털의 표를 설명 대상으로 삼아 표 제목 이외에는 별도의 설명이나 부가 정보가 거의 없는 환경을 지침에 반영했다. 이는 실제 공공 데이터 환경을 충실히 반영한 것으로, 향후 통계 포털 기반의 다른 말뭉치 구축 사업에도 적용할 수 있는 지침을 제시했다는 점에서 의미가 있다. 또한 2024년도 사업에서 가장 크게 부각된 문제점인 ‘불연속 영역 선정 기준’을 재정비하여 말뭉치 구축 작업자가 표의 핵심 내용을 비교·대조하는 설명 문장을 보다 유연하게 작성할 수 있도록 개선했다.

<2025년 한국어언어문화 사진 설명 문장 작성 지침>은 식생활, 의생활, 주생활, 건축, 공예 다섯 개의 한국어언어문화 분류에 따른 설명 속성과 설명 문장 작성 방안을 마련하고, 한국어언어문화 요소에 대한 멀티모달 말뭉치를 구축하는 데 기여했다는 점에서 의미가 있다.

특히 설명 속성을 사진 속에서 눈으로 확인할 수 있는 속성인 ‘가시적 속성’과 사진에 드러나지 않지만 백과사전적 정보로서 중요한 ‘비가시적 속성’으로 구분하고, 식생활, 의생활, 주생활, 건축, 공예 각각에 대해 ‘색채, 형태, 구성 요소, 소재, 문양’ 등의 시각적

묘사 중심의 속성과 ‘용도, 이용자, 기법, 소재지, 역사’ 등의 배경지식 중심의 속성을 체계적으로 정리한 점이 주목할 만하다. 이는 단순한 ‘사진 캡션’의 수준을 넘어, 한국어언어 문화 요소의 시각적 특징과 문화적 맥락을 함께 포착하여 정보를 제공하는 다층적 설명 구조를 마련했다는 점에서 의미가 크다.

‘2025년 한국어언어문화 표·사진 설명 말뭉치 구축 연구’는 한국어언어문화 관련 표와 사진에 대한 설명 문장을 체계적으로 설계된 지침에 따라 구축함으로써, 향후 생성형 인공지능이 한국어언어문화를 보다 잘 이해할 수 있고 한국어언어문화에 대한 문장을 생성해 낼 수 있게 하기 위한 신뢰도 높은 학습 자원을 마련했다는 점에서 의의가 있다. 이는 향후에 유사 사업 및 데이터 기반 기술의 고도화에 유의미하게 활용될 수 있을 것이다.

□ 한계

<2025년 한국어언어문화 표 설명 문장 작성 지침>은 한국어언어문화와 관련된 표에 대한 설명 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 했다. 이 지침은 별도의 설명이나 부가 정보가 거의 없는 표를 설명 대상으로 삼아 설명 문장을 작성할 수 있도록 고안되었으므로 범용적 성격을 띠고 있다. 그러나 주제를 한국어언어문화로 한정하여 표를 수집했을 때 표의 내용이 숫자 값으로 되어 있는 경우가 많았는데, 이러한 표가 과연 인공지능이 한국어언어문화에 대한 이해를 높이는 데 기여할 수 있을지에 대해서는 의문이 든다. 표 설명 문장 작성 지침은 ‘표’라고 하는 특정한 형식으로 된 자료 구조를 자연어로 된 설명 문장과 매핑시키는 것이 관건이므로 주제를 한정하기보다는 다양한 표를 수집하는 것이 표 설명 문장 생성에 도움이 될 것이다.

<2025년 한국어언어문화 사진 설명 문장 작성 지침>은 설명 대상과 속성 체계가 ‘한국어언어문화’ 중에서도 생활(식생활, 의생활, 주생활)과 예술(건축, 공예)에 한정되어 있다는 점에서 한계가 있다. 사진에서 대상이 명확하게 표현되어야 하기에 ‘놀이, 의례, 행사, 공연, 음악, 대중문화’ 등의 한국어언어문화 영역은 이번 사업에서 다루지 못했다. 또한 설명 속성이 ‘한국어언어문화’라는 특정 주제를 대상으로 마련됐기 때문에 범용적인 설명 말뭉치에 이를 활용하기에는 적용 범위가 제한적이라는 점도 한계로 지적할 수 있다.

이밖에도 ‘가시적 속성’과 ‘비가시적 속성’의 구분이 이론적으로는 분명하지만, 실제 말뭉치 구축 과정에서는 구분하기 모호한 경우가 있다. 사진에서 가시적으로 드러나지 않더라도 특정 대상의 일반적인 특성을 어느 범위까지 비가시적 속성으로 기술할 것인지, 여러 분류에 공통적으로 나타나는 속성인 ‘역사, 이용자, 소유자’와 같은 비가시적 속

성을 어떻게 일관되게 처리할 것인지 등에 대한 세부 기준이 보완될 필요가 있다. 한국언어문화 분류별 설명 속성을 추가적으로 도입할 가능성을 열어 두고, 향후에 단계적으로 속성 체계를 재구조화할 필요가 있다.

4.3. 향후 연구

본 사업은 한국언어문화의 표·사진 기반 멀티모달 말뭉치를 구축하는 데 중점을 두었기 때문에, ‘놀이, 의례, 행사, 공연’과 같이 시간성과 행위성이 강한 자료는 다루지 못했다. 따라서 향후에는 이러한 동적 문화 요소까지 포함한 자료로 구축 대상을 확장하여, 한국언어문화를 보다 입체적으로 반영할 수 있는 멀티모달 데이터가 마련될 필요가 있다.

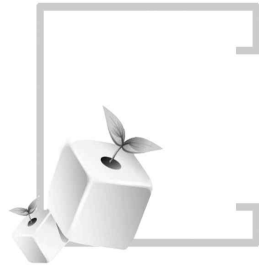
또한 자료가 부족한 희소 분야에 대해서는 직접 자료 수집을 진행하거나, 유관 기관과의 협의를 통해 안정적인 자료 확보 체계를 마련하는 후속 연구가 요구된다.

아울러, 본 사업에서는 설명 문장 작성에 초점을 두었으나, 향후에는 구축된 말뭉치를 기반으로 표·사진 유사 문장 생성 모델을 개발하는 연구가 필요하다. 이는 단순한 기술 확장을 넘어, 한국언어문화에 특화된 멀티모달 생성 능력을 고도화하고 실제 서비스·교육·연구 등 다양한 분야에서 활용성을 크게 높이는 중요한 단계가 될 것이다.

이와 같은 확장 연구를 통해 한국어 멀티모달 말뭉치는 더욱 풍부한 문화적 맥락을 담아낼 수 있으며, 인공지능이 한국언어문화를 정교하게 이해하고 재현하는 데 기여하는 핵심 기반으로 자리매김할 것으로 기대된다.

참고 문헌

- 국립국어원(2021), “2021년 유사 문장 생성 말뭉치 연구 분석 사업”, 국립국어원 연구용역 결과보고서(연구책임: 김일환).
- 국립국어원(2024), “2024년 표 설명 문장 작성 말뭉치 연구 분석”, 국립국어원 연구용역 결과보고서(연구책임: 차정원).
- 대한민국(2015), 「문화재보호법」.
- 문화재청(2006), 「문화재 활용을 위한 정책기반 조성 연구」.
- 문화재청(2007), 「문화재 활용 가이드북」.
- 문화재청(2010), 「문화재 유형별 활용 길라잡이」.
- 설현우, 찬리스, 권장우(2019), “한국어 데이터 셋을 이용한 이미지 캡션에 대한 연구”, 재활복지공학회논문지 제13권 제4호, 한국재활복지공학회.
- 유네스코(2003), 「무형문화유산 보호 협약」.
- 이현주, 위진, 강미영 외 1명(2024), “용어지식의 통합 및 확장 방안-국립국어원 온용어 플랫폼을 중심으로-”, 우리말글 제103집, 우리말글학회.
- 최종호(2016), 「무형문화유산과 무형문화재 유형별 분류체계에 관한 연구」.
- 류호철 외(2012), 「문화재 활용 기본계획 연구」.
- McHugh, Mary L. (2012), “Interrater reliability: The kappa statistic”, Biochemia Medica.
- Parikh et al.(2020), “ToTTo: A Controlled Table-To-Text Generation Dataset”, EMNLP, pp.1173-1186.
- Tianyi Zhang et al.(2020), “BERTSCORE: EVALUATING TEXT GENERATION WITH BERT”, ICLR 2020.
- Wang et al.(2022), “Robust (Controlled) Table-to-Text Generation with Structure-Aware Equivariance Learning”, NAACL, pp.5037-5048.
- Yao et al.(2018), “Exploring Visual Relationship for Image Captioning”, ECCV, pp.684-699.



부록

표 설명 문장
작성 지침



표 설명 문장 작성 지침

<목 차>

1. 개요	2
1.1. 지침의 목적	2
1.2. 관련 사업	2
1.3. 주요 용어	2
1.4. 표의 다양한 유형	3
1.5. 표 설명 문장 작성 예시	8
2. 표 설명 문장 작성 방법	9
2.1. 표 전체에 대한 서술문	12
2.2. 행에 대한 서술문	16
2.2.1. 행 영역 선정	16
2.2.2. 행 설명 문장 작성	18
2.3. 열에 대한 서술문	23
2.3.1. 열 영역 선정	23
2.3.2. 열 설명 문장 작성	25
2.4. 불연속 영역에 대한 서술문	29
2.4.1. 불연속 영역 선정	29
2.4.2. 불연속 영역 설명 문장 작성	32
2.5. 주의 사항	34
2.5.1. 설명 문장 관련 주의 사항	34
2.5.2. 기호 관련 주의 사항	37

1. 개요

1.1. 지침의 목적

이 지침은 표 설명 문장 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 한다. 이 지침에서는 표 자료에 대한 설명 문장을 작성하는 방안을 기술한다. 표 설명 문장을 작성할 때에는 표의 내용이 드러나도록 행과 열, 칸(불연속 영역), 표 전체에 대한 서술문을 각 1개의 문장으로 기술하되, 표의 핵심 내용이 드러나는 설명 영역을 선정하고 이를 담은 문장을 작성한다.

1.2. 관련 사업

이 지침은 국립국어원의 '2025년 한국어문화 표·사진 설명 말뭉치 구축 연구' 사업의 일환으로서, 표 설명 문장 말뭉치를 구축하기 위한 지침이다.

1.3. 주요 용어

표 설명 문장 작성에 사용되는 용어는 아래와 같다.

- 표: 행과 열로 배열된 정보의 집합
- 행: 표에서 가로로 배열된 원소들의 집합
- 열: 표에서 세로로 배열된 원소들의 집합
- 칸: 표에서 행과 열이 만나 이루어지는 단위로서 정보를 담는 공간(=셀)
- 제목 행: 각 열의 제목을 표시한 행
- 제목 열: 각 행의 제목을 표시한 열
- 불연속 영역: 표 안의 정보를 조합하여 기술하기 위한 둘 이상의 칸
- 표 설명 문장(서술문): 표의 핵심 내용이 드러나도록 표의 정보들을 구체적으로 설명한 문장. '행, 열, 불연속적인 둘 이상의 칸, 표 전체'를 기술한 문장
- 설명 영역: 표의 중요 내용을 담은 영역으로서 문장으로 표현하는 대상이 되는 영역

표제목 무용 세부 장르별 공연 건수 및 횟수_재공연_2022년

세부 장르별(1)	2022		
	재공연		
	공연 건수 (건)	공연 횟수 (회)	1건당 공연 횟수 (회)
계	2,118	3,847	1.8
현대 무용	401	704	1.8
한국 무용	1,084	1,634	1.5
발레	301	561	1.9
복합	161	239	1.5
기타	171	709	4.1

이 표는 2022년 기준 무용 세부 장르별 재공연 건수 및 횟수를 나타낸 표이다.

1.4. 표의 다양한 유형

- 제목 행과 제목 열에 따른 표 유형 분류
 - 제목 행과 제목 열 모두 연속적인 숫자 값인 경우

구분	2% 미만	2-3%	3-5%	5-7%	7-10%	10-13%	13%-	총 합계
300개 미만	0	1	5	8	1	0	0	15
300-599개	0	11	55	9	5	0	1	81
600-899개	0	3	40	20	2	1	3	69
900-1199개	0	5	33	12	2	1	0	53
1200-1499개	1	0	6	1	0	0	0	8
1500개 이상	0	2	1	0	0	0	0	3
총 합계	1	22	140	50	10	2	4	229

- 제목 행은 개별 항목, 제목 열은 연속적인 숫자 값인 경우

구분	치킨	한식	커피	제과제빵	피자	주점	외식업 전체
2021년	279	277	209	453	255	179	279
2022년	297	325	225	490	284	298	314
증감률	6.5%	17.2%	7.4%	8.1%	11.6%	66.2%	12.7%

- 제목 열은 개별 항목, 제목 행은 연속적인 숫자 값인 경우

구분	2019년	2020년	2021년	2022년	2023년	합계	비율
미끄러짐·넘어짐	98	315	100	226	398	1,137	92.1
부딪힘	7	11	4	5	27	54	4.4
예리함·마감처리 불량	1	4	1	4	3	13	1.1
추락	-	1	1	-	-	2	0.2
파열·파손·끼여짐	1	3	2	1	-	7	0.5
기타 위험 및 위해내용	1	1	3	1	15	21	1.7
합계	108	335	111	237	443	1,234	100.0

- 제목 행과 제목 열 모두 개별 항목인 경우

구분	복숭아	자두	매실
조사 과수원 수	20	36	8
발생 과수원 수	9	25	8
발생 과수원 비율(%)	45	69.4	100
조사 나무 수	600	1,080	240
발생 나무 수	100	239	120
발생 나무 비율(%)	16.7	22.1	50

● 셀 값의 데이터 형식에 따른 표 유형 분류

◦ 값이 숫자인 경우

구분	2020년	2021년	2022년	2023년(1~9월)	계
온라인판매	104 (63.4)	100 (66.6)	216 (76.3)	144 (65.4)	564 (69.0)
일반판매	38 (23.2)	25 (16.7)	37 (13.1)	38 (17.3)	138 (16.9)
방문판매	22 (13.4)	25 (16.7)	30 (10.6)	38 (17.3)	115 (14.1)
계	164 (100.0)	150 (100.0)	283 (100.0)	220 (100.0)	817 (100.0)

◦ 값이 문자(구)인 경우

구분	내용
대학 입시설명회	- 2025학년도 대입 지원전략 및 전형 안내 - 지역별 소재 2·4년제 대학 입시전형 및 학과 소개 - 진로진학박람회 개최 등
1:1 맞춤형 상담	- 대학별 진학 상담 및 입시 상담 - 진로·진학 박람회 공간 운영 - 자기소개서 및 모의면접 특강 등

◦ 값이 문자(문장)인 경우

질문	답변
Q. 중간 물떼기란?	A. 무효분얼기(햇새끼치기) 중 부리를 건전하게 하기 위해 이삭패기 전 40일부터 30일 사이에 논바닥이 갈라질 정도로 물떼기 하는 것
Q. 탄소중립 프로그램 시범사업중간 물떼기 인정기준은?	A. 활동기간 중 2주 이상 용수 공급을 중단하고, 배수물꼬를 개방하여 논을 마른 상태로 유지
Q. 탄소중립 프로그램 시범사업중간 물떼기 증빙방법은?	A. 중간 물떼기 시작 종료시기에 필지 내 배수물꼬 개방 또는 논바닥이 말라 있는 것이 확인 가능한 사진을 각 1회씩 촬영하여 이행점검 웹에 제출 * 활동기간 중 배수물꼬를 개방했다면 강우일도 활동 인정
Q. 논물 얇게 걸러대기란?	A. 중간 물떼기 이후(여름기) 논물을 얇게(2~5cm) 관개 후 자연적으로 말리고, 다시 얇게 대기를 이삭이 익을 때까지 반복하는 것
Q. 탄소중립 프로그램 시범사업논물 얇게 걸러대기 인정 기준은?	A. 중간물떼기 종료 후 ①2~5cm 깊이로 용수공급 후 ②자연소모로 논물을 말리는 [①→②]의 과정을 3~5일 간격으로 4회 이상 반복
Q. 탄소중립 프로그램 시범사업논물 얇게 걸러대기 증빙 방법은?	A. 배수물꼬 부근 2~5cm 이하의 높이의 톱을 조성하거나 배수물꼬를 조절하고, 논물이 자연건조되어 논바닥이 가장 말라 있는 시기의 사진을 4회 이상촬영(3~5일 간격)하여 이행점검 웹에 제출
Q. 논물관리의 효과는?	A. ① 토양에 산소공급을 원활하게 해줌으로써 물을 가뭄될 때 생기는 메탄가스 발생 저감 ② 토양 중 각종 유기산의 배제, 벼 뿌리의 활력 증진에 의한 양·수분 흡수 증대 ③ 뿌리의 심층 분포 비율 증대 및 지상부 조직을 강건히 하여 쓰러짐에 대한 저항성 증대 등

- 값에 문자(단어, 구)와 숫자가 혼합된 경우

브랜드	제품명	판매업체	10mL 가격(원)
과일나라	바세린 모이스처 이엑스 핸드 앤 네일 에센스 크림	(주)엘코스	1,000
꽃을든남자	우유 핸드크림	(주)코스모코스	375
네이처리퍼블릭	핸드 앤 네이치 시어버터 핸드크림	(주)네이처리퍼블릭	2,200
뉴트로지나	인텐스 리페어 핸드크림	한국존슨앤드존슨판매(유)	1,411
니베아	모이스춰 케어 올리브 핸드 크림	바이어스도르프코리아(유)	760
룩시땅	2022 시어 버터 드라이스킨 핸드 크림	룩시땅코리아(유)	3,733
바세린	인텐티브 케어 헬씨 핸드 앤 스트롱거 네일 로션	유니레버코리아(주)	533
이니스프리	올리브 리얼 모이스처 핸드크림	(주)이니스프리	1,200
이솝	레저렉션 아로마틱 핸드 밤	이솝코리아(유)	4,133
카밀	핸드 & 네일 크림 클래식	(주)지엘커뮤니케이션즈	720

- 합계 제시 방식에 따른 표 유형 분류

- 마지막 행이나 열에 합계가 있는 경우

불만·피해유형	응답자 수	비중
부당가입 유도	96	37.5
가입 시 주요 내용 설명 및 고지 미흡	90	35.2
약정 해지 변경에 따른 위약금 부과	80	31.3
서비스 품질 미흡	70	27.3
계약 시 안내받은 정보와 실제 서비스 상이	52	20.3
계약 후 약정 해지·변경 불가	37	14.5
계약불이행(불완전이행)	36	14.1
미신청 부가서비스 요금 청구	33	12.9
개인정보 유출	31	12.1
가입 지연 및 누락	13	5.1
기타	12	4.7
합계 (※ 중복응답 포함)	256(550)	100.0(214.8)

- 마지막 행이나 열에 합계가 없는 경우

조사대상	조사 결과			유럽표준 준용	
	등반벽 높이	매트 최소 폭*	시설 수		
25개소	3.0m 초과	전면	2.5m 미만	22개소(88.0%)	미흡
			2.5m 이상	3개소(12.0%)	충족
		측면	측면 매트리스 없음	5개소(20.0%)	미흡
			1.5m 미만	19개소(76.0%)	미흡
			1.5m 이상	1개소(4.0%)	충족

- 소계도 있고 총계도 있는 경우

구분	2022년		2023년				
	건수	비율	건수	비율	발달단계별 연앙인구	인구 천 명당 접수 건수	
취학전	영아기(0세)	2,389	(11.0)	2,916	(13.0)	235,093.0	12.4
	걸음마기(1~3세)	8,235	(38.1)	8,240	(36.8)	824,659.5	10.0
	유아기(4~6세)	4,822	(22.3)	4,661	(20.9)	1,058,566.0	4.4
	소계	15,446	71.4	15,817	70.7	2,118,318.5	7.5
학령기(7~14세)	6,196	(28.6)	6,554	(29.3)	3,654,240.5	1.8	
총계	21,642	100	22,371	100	5,772,559	3.9	

1.5. 표 설명 문장 작성 예시

(단위: %, 명)

동반 인원수(본인 제외)_방한 횟수별

통계 분류(1)	통계 분류(2)	2023				
		1명	2명	3명 이상	평균	
방한 횟수	1회	50.6	16.5	32.9	4.3	행 설명 영역
	2회	51.7	17.5	30.8	3.5	
	3회	54.9	19.7	25.5	2.6	
	4회 이상	61.4	18	20.6	2.2	
		열 설명 영역	불연속 영역			

[표 전체] 설명 문장	이 표는 2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 표이다.
[행] 설명 문장	2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 결과, 방한 횟수 1회의 동반 인원수는 1명이 50.6%로 가장 많았고 3명 이상이 32.9%, 2명이 16.5%로 그 뒤를 이었다.
[열] 설명 문장	2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 결과, 동반 인원이 1명이었던 방한 외래 관광객의 방한 횟수는 4회 이상이 61.4%로 가장 많았고 3회가 54.9%, 2회가 51.7%, 1회가 50.6%로 그 뒤를 이었다.
[불연속 영역] 설명 문장	2023년 기준 방한 외래 관광객의 방한 횟수별 동반 인원수를 조사한 결과, 동반 인원이 3명 이상이었던 방한 외래 관광객의 방한 횟수는 1회가 32.9%로 가장 많았고 동반 인원이 2명이었던 방한 외래 관광객의 방한 횟수는 3회가 19.7%로 가장 많았다.

2. 표 설명 문장 작성 방법

- 셀 값이 숫자일 경우 주어진 단위를 확인하여 설명 문장 작성에 활용한다.

교류 방법_주요 사업_업종별						
(단위: 개, %)						
구분	COL2	2022				
		사례 수	일방향 (발신형)	일방향 (수신형)	쌍방향	다자간
	전체	5,932	37.8	9.6	46.6	6.0
업종	공공 행정, 국방 및 사회보장 행정	296	20.8	10.6	61.0	7.6
업종	예술, 스포츠 및 여가 관 련 서비스업	1,604	41.2	13.5	37.6	7.7
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	37.7	7.9	49.1	5.2

[행 영역] 설명 문장	2022년 국제 문화 교류 사례 수는 총 5,932개로, 이 가운데 쌍방향 교류의 비중이 46.6%로 절반 가까이 되는 비중을 차지했고, 일방향(발신형) 교류의 비중이 37.8%, 일방향(수신형) 교류는 9.6%, 다자간 교류는 6%를 차지했다.
-----------------	---

- 셀 값이 숫자일 경우 주어진 단위를 확인하고 만 단위로 띄어 쓴다. 이때 ‘만, 억, 조’ 등은 한글로 표현한다.

공연 시설: 항목별 평균 수입 내역_운영주체2별

(단위: 개, %, 천 원)

통계 분류 (1)	통계 분류 (2)	2022						
		시설 수	수입 총액	자체 수입	정부 예산 및 보조금	기타 수입	기부/ 후원금	전년도 이월금
		소계	소계	평균	평균	평균	평균	평균
		소계	소계	소계	소계	소계	소계	소계
전체	소계	992	1,481,950.6	258,494.7	1,068,232.9	43,067	37,893.9	74,262
운영 주체2	공공	461	2,734,961.3	304,466	2,193,017.2	48,016.9	32,233	157,228.1
	민간	531	394,120.2	218,583.6	91,725.3	38,769.6	42,808.5	2,233.1

[불연속 영역] 설명 문장	2022년 공연 시설 운영 주체의 주 수입 내역을 살펴보면, 공공 기관은 정부 예산 및 보조금 명목 수입이 21억 9,301만 7,200원 으로 가장 많았고, 민간 기관은 기부 및 후원금 명목 수입이 4,280만 8,500원 으로 가장 많았다.
-------------------	--

- 다른 문장을 참조하지 않아도 해당 설명 문장을 통해 표에 대한 이해가 가능하도록 작성한다.

관광객이용시설업 한옥체험업 객실 판매 현황_존속기간별

(단위: 개, 만 원, 일)

통계 분류(1)	통계 분류(2)	2022		
		판매 객실 수	객실당 평균 가격	이용객 평균 체류 일수
전체	소계	344,063	19.1	1.8
존속기간별	5년 미만	96,814	23.2	1.6
	5년~10년 미만	147,513	13.9	1.8
	10년 이상	99,736	20.4	2.1

<p>[행] 설명 문장</p>	<p>2022년 기준 존속 기간별 관광객 이용 시설업의 한옥 체험업 객실 판매 현황 조사 결과, 5년 미만 시설의 판매 객실 수는 9만 6,814개로 객실당 평균 가격은 23만 2,000원, 이용객 평균 체류일 수는 1.6일로 나타났다.</p>	<p>(O)</p>
<p>[행] 설명 문장</p>	<p>5년 미만 시설의 판매 객실 수는 9만 6,814개로 객실당 평균 가격은 23만 2,000원, 이용객 평균 체류일 수는 1.6일로 나타났다.</p>	<p>(X)</p>

- 표 전체, 행, 열, 불연속 영역의 설명 문장이 각각 독립된 문장으로 이해될 수 있게 작성한다. 설명 문장 작성 시 필요한 문장 성분을 생략하거나 대응어를 사용하지 않도록 한다.

<p>[표 전체] 설명 문장</p>	<p>이 표는 2022년부터 2024년까지의 국가유산진흥원 공연 사업 운영 현황을 나타낸 표이다.</p>	
<p>[행] 설명 문장</p>	<p>2022년부터 2024년까지의 국가유산진흥원 공연 사업 운영 현황 조사 결과를 살펴보면, 전수교육관의 2022년 공연 횟수는 160회, 관람 인원은 1만 1,204명이었고 2023년 공연 횟수는 187회에 관람 인원 1만 3,358명, 2024년 공연 횟수는 165회에 관람 인원은 1만 2,068명이었다.</p>	<p>(O)</p>
<p>[행] 설명 문장</p>	<p>최근 2년간 국가유산진흥원 공연 사업 운영 현황 조사 결과를 살펴보면, 전수교육관의 2022년 공연 횟수는 160회, 관람 인원은 1만 1,204명이었고 2023년 공연 횟수는 187회에 관람 인원 1만 3,358명, 2024년 공연 횟수는 165회에 관람 인원은 1만 2,068명이었다.</p>	<p>(X)</p>

2.1. 표 전체에 대한 서술문

- 표 제목(캡션)을 참고하여 작성한다.

무형유산 전수교육관 현황_행정구역별				
(단위: 개)				
행정구역별(1)	2024			
	전수교육관 수	입주 종목수		
	소계	소계	국가 지정	시도 지정
합계	168	356	112	244
서울	3	38	20	18
부산	5	22	5	17
대구	1	19	2	17
인천	2	28	5	23
광주	3	14	1	13
대전	4	21	1	20
울산	-	-	-	-
세종	1	4	1	3
경기	20	25	5	20
강원	16	16	3	13
충북	8	8	3	5
충남	18	22	5	17
전북	14	19	7	12
전남	25	37	19	18
경북	18	25	10	15
경남	25	44	18	26
제주	5	14	7	7

[표 전체] 설명 문장	이 표는 2024년 무형유산 전수 교육관 현황을 서울, 부산, 대구 등 행정 구역별로 나타낸 표이다.	(0)
--------------	--	-----

- 행과 열의 관계, 즉 행에 따른 열, 열에 따른 행이 무엇을 나타내는지 기술한다.

(1) 행 제목이나 열 제목이 연속적인 경우

- 시작 값과 끝 값을 명시하고 연속된 값 사이의 간격을 명시한다.

항목별 만족도_식도락_방한 횟수별

(단위: %, 점)

통 계분류(1)	통계 분류(2)	통계 분류(1)	2023
방한 횟수	1회	① 매우 불만족	0.1
		② 불만족	0.4
		③ 보통	4.4
		④ 만족	30.8
		⑤ 매우 만족	64.2
		5점 평균	4.6
		100점 평균	89.7
		2회	① 매우 불만족
	② 불만족		0.3
	③ 보통		4
	④ 만족		33.9
	⑤ 매우 만족		61.6
	5점 평균		4.6
	100점 평균		89.2
	3회		① 매우 불만족
		② 불만족	0
		③ 보통	3.2
		④ 만족	33.4
		⑤ 매우 만족	63.3
		5점 평균	4.6
		100점 평균	90
		4회 이상	① 매우 불만족
	② 불만족		0.1
	③ 보통		3.2
	④ 만족		30.2
	⑤ 매우 만족		66.3
	5점 평균		4.6
	100점 평균		90.6

[표 전체]
설명 문장

이 표는 2023년 방한 외국인의 식도락 항목 만족도를 **1회부터 4회 이상까지 각 방한 횟수별로** 조사한 표이다.

(2) 행 제목이나 열 제목이 연속적이지 않은 항목인 경우

- 제목을 참고하여 행이나 열이 나타내는 바를 기술한다. 항목의 개수가 3개인 경우 모두 나열하며, 항목의 개수가 많은 경우 3개까지만 나열한다.

공공지원 : 지방자치단체 : 공연 예술 예산 규모별 분포 현황

(단위: 개)

예산 규모별(1)	2022							
	지자체 수	구분1		구분2				
	소계	광역	기초	광역		기초		
	소계	소계	소계	특별-광역시	광역시도	시	군	구
전체	245	17	228	8	9	77	82	69
10억 원 미만	31	-	31	-	-	-	12	19
10억 원 이상 ~ 50억 원 미만	100	-	100	-	-	17	54	29
50억 원 이상~ 100억 원 미만	46	-	46	-	-	18	16	12
100억 원 이상~ 200억 원 미만	39	1	38	1	-	29	-	9
200억 원 이상	29	16	13	7	9	13	-	-

[표 전체]
설명 문장

이 표는 2022년 광역 및 기초 지방 자치 단체의 공연 예술 예산 분포 현황을 10억 원 미만, 10억 원 이상~50억 원 미만, 50억 원 이상~100억 원 미만 등의 예산 규모별로 조사한 표이다.

- 숫자 데이터와 문자 데이터에 따라 다른 전략을 구사한다.
 - 행/열 제목의 값이 숫자인 경우: 시작 값과 끝 값을 명시하여 기술한다.
 - 행/열 제목의 값이 문자인 경우: 행/열 제목의 값이 많은 경우 3개까지 나열하거나, 표 내용을 참고하여 제목 행/열을 아우를 수 있도록 기술한다.

영업 현황: 존속 기간별 외국인 유치 인원[인바운드]_3분기_소재지별

(단위: 명)

구분(1)	구분(2)	2024.3/4			
		전체	5년 미만	5년~10년 미만	10년 이상
전체	소계	502,946	31,796	293,422	177,728
모바일 콘텐츠 보유 여부	보유	254,881	2,888	195,661	56,332
소재지별	서울	484,408	31,315	278,881	174,212
	부산	5,087	0	5,087	0
	대구	0	0	0	0
	인천	8,438	37	8,401	0
	광주	37	0	0	37
	대전	0	0	0	0
	울산	0	0	0	0
	세종	0	0	0	0
	경기	1,621	444	942	235
	강원	0	0	0	0
	충북	0	0	0	0
	충남	0	0	0	0
	전북	0	0	0	0
	전남	0	0	0	0
	경북	0	0	0	0
	경남	0	0	0	0
제주	3,355	0	111	3,244	

[표 전체]
설명 문장

이 표는 2024년 3분기 기준, 5년 미만부터 10년 이상까지 존속한 국내 여행사의 외국인 유치 인원을 서울, 부산, 대구 등 국내 체류 소재지별로 나누어 정리한 표이다.

2.2. 행에 대한 서술문

2.2.1. 행 영역 선정

- 다음과 같은 기준으로 설명의 대상으로 삼을 행 영역을 선정한다.
 - 각 행의 합계를 비교하여 가장 높은 값을 가지는 행을 선정할 수 있다.

추진 방식_주요 사업 1_업종별

(단위: 개, %)

구분	COL2	2022			
		사례수	현지 초청	한국 주관	상호 주관
전체		5,932	24.3	47.2	28.5
업종	공공 행정, 국방 및 사회보장 행정	296	17.4	43.6	39.0
업종	예술, 스포츠 및 여가관련서비스업	1,604	47.6	28.6	23.8
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	15.6	54.9	29.5

- 첫 행이나 마지막 행에 각 행의 합계가 제시되어 있으면 그 행을 설명 영역으로 선정할 수 있다.

추진 방식_주요 사업 1_업종별

(단위: 개, %)

구분	COL2	2022			
		사례수	현지 초청	한국 주관	상호 주관
전체		5,932	24.3	47.2	28.5
업종	공공 행정, 국방 및 사회보장 행정	296	17.4	43.6	39.0
업종	예술, 스포츠 및 여가관련서비스업	1,604	47.6	28.6	23.8
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	15.6	54.9	29.5

- 해당 행의 제목이 여러 층위로 구성된 경우 상위 층위까지 선정한다.

[지역어 경험] 주로 사용하는 지역어(사투리)_거주지역별

(단위: %)

표측 구분	표측 상세	2024						
		강원도	경상도	전라도	제주도	충청도	서울/경기	이북(북한)
전체		5.1	43.0	24.3	2.2	15.6	18.1	0.3
거주 지역	수도권	5.7	22.5	25.7	0.8	15.6	41.7	0.5
	충청권	1.7	14.8	9.9	0.7	75.0	12.1	0.3
	호남권	0.3	3.2	94.5	0.5	1.8	1.8	0.0
	영남권	2.0	94.2	2.7	0.5	2.5	3.2	0.1
	강원/제주	48.3	8.7	2.8	35.3	6.1	9.0	0.3

2.2.2. 행 설명 문장 작성

- 각 행의 값이 숫자이면 행 설명 문장은 다음을 참고하여 작성한다.
 - 해당 행의 행의 합계가 다른 행의 합계보다 높은 값을 가지면 해당 행의 제목과 합계를 중심으로 기술한다.

전시 서비스 사업자의 고용 현황_고용 형태별

(단위: 개, 명, %)

특성별 (1)	특성별 (2)	2023								
		사례 수 (개)	총종사자 수 (명)				전시산업 종사자 수 (명)			총종사 자 수 대비 전 시산업 종사자 수 (%)
			소계	소계	남자	여자	소계	남자	여자	
전체	소계	1,324	60,625	39,544	21,081	4,942	3,516	1,426	8.2	
고용 형태별	정규직	1,324	49,841	31,392	18,449	3,914	2,648	1,266	7.9	
	비정규직	1,324	10,364	7,815	2,549	812	685	127	7.8	
	자영업자	1,324	420	336	83	216	183	33	51.4	

[행] 설명 문장	2023년 전시 서비스 사업자의 정규직 고용 현황을 살펴보면, 전시 서비스 사업 정규직 종사자는 남자 31,392명, 여자 18,449명, 총 49,841명이었는데, 이 중 전시 산업 정규직 종사자 수는 남자 2,648명, 여자 1,266명, 총 3,914명으로 총종사자 수 대비 전시 산업 종사자 수는 7.9%로 나타났다.
--------------	---

- 해당 행의 값들이 증가세 혹은 감소세와 같은 일정한 추이를 보이면 그 내용을 기술한다.

국가유산진흥원 주요사업 추진 현황_한국의집 운영별

(단위: 명, 건)

사업별(1)	사업별(2)	2022	2023	2024
한국의집 운영	전통음식 보존 보급(명)	69,118	47,194	39,401
	전통예술 보존 보급(명)	1,148	720	-
	전통 혼례 보급(건)	268	171	88
	전통문화 체험(명)	60	-	-
국가 무형유산 전수교육관 운영(명)	전통예술 공연	11,204	13,358	12,068
	전통공예 및 전통미술 전시	12,816	20,430	13,805
한국 문화의 집 운영(명)	전통예술 공연	12,736	12,389	15,254
	교육프로그램 운영	6,624	11,462	6,143
문화 유산 관광 자원화 (명)	소계	3,191,411	6,442,272	6,868,357
인천공항 전통문화체험관 운영(명)	소계	211,795	1,079,308	1,372,400

[행] 설명 문장	국가 유산 진흥원의 한국의 집 운영 사업 가운데 전통 음식 보존 보급 사업은 2022년에 69,118명, 2023년에 47,194명, 2024년에 39,401명으로, 해마다 그 규모가 줄어들고 있다.
--------------	---

- 해당 행에서 높은 값을 가지는 칸을 중심으로 기술한다.

국제회의 근로자의 존속 기간별 현황

(단위: 명, %)

존속 기간별(1)	2023							
	국제회의 관련 총근로자 (명)	구성비 (%)	국제회의 전담 근로자		국제회의 주 근로자		국제회의 보조 근로자	
			소계	구성비 (%)	근로자수 (명)	구성비 (%)	근로자수 (명)	구성비 (%)
전체	756	100.0	750	100.0	4	100.0	2	100.0
5년 미만	90	11.9	90	12.0	-	-	-	-
5-10년 미만	147	19.4	147	19.6	-	-	-	-
10년 이상	520	68.8	514	68.5	4	100.0	2	100.0

[행] 설명 문장	2023년 국제회의의 관련 총근로자는 756명이었는데, 그중 국제회의의 전담 근로자는 750명으로 대부분을 차지했고, 국제회의의 주 근로자는 4명, 국제회의의 보조 근로자는 2명 순으로 나타났다.
--------------	---

- 해당 행에서 낮은 값을 가지는 칸을 중심으로 기술한다.

공연 단체: 분장실 보유 및 운영 현황_주요 활동 장르별

(단위: 개, %)

통계 분류 (1)	통계 분류 (2)	2022								
		단체수	소유 또는 임대		소유		임대		해당 사항 없음	
			소계	사례수	비중	사례수	비중	사례수	비중	사례수
전체	소계	4,294	571	13.3	171	4	422	9.8	3,723	86.7
주요 활동 장르	연극	1,076	157	14.6	56	5.2	115	10.7	919	85.4
	뮤지컬	275	56	20.2	10	3.7	45	16.5	219	79.8
	서양 음악 (클래식)	1,463	187	12.8	48	3.3	140	9.6	1,276	87.2
	한국 음악 (국악)	756	109	14.4	41	5.4	73	9.6	647	85.6
	무용	475	35	7.3	11	2.3	24	5	440	92.7
	복합	249	28	11.1	5	2.1	25	10.1	221	88.9

[행] 설명 문장	공연 단체의 분장실 보유 현황 및 운영 현황을 주요 활동 장르별로 조사한 결과, 2022년 기준 연극 공연 단체는 1,076개로 분장실을 소유하고 있는 공연 단체는 5.2%인 56개에 불과했으며, 분장실을 임대하고 있는 연극 공연 단체는 10.7%인 115개로 조사되었다.
--------------	--

- 제목 행의 값이 문자 값이면 해당 내용을 나열하여 기술한다.

추진 방식_주요 사업 1_업종별

(단위: 개, %)

구분	COL2	2022			
		사례수	현지 초청	한국 주관	상호 주관
전체		5,932	24.3	47.2	28.5
업종	공공 행정, 국방 및 사회보장 행정	296	17.4	43.6	39.0
업종	예술, 스포츠 및 여가 관련 서비스업	1,604	47.6	28.6	23.8
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	15.6	54.9	29.5

[행] 설명 문장	2022년 국제 문화 교류 사업체의 국제 문화 교류 사업 추진 방식을 조사한 결과, 총 5,932개의 사례 중 한국 주관이 47.2%로 가장 높았고, 상호 주관 28.5%, 현지 초청 24.3% 등이 그 뒤를 이었다.
--------------	---

- 각 행의 값이 문자이면 행 설명 문장은 다음을 참고하여 작성한다.
 - 해당 행에서 둘 이상의 값을 선정하여 비교 혹은 대조하는 문장을 작성한다.

[지역어 경험] 주로 사용하는 지역어(사투리)_거주 지역별

(단위: %)

표측 구분	표측 상세	2024						
		강원도	경상도	전라도	제주도	충청도	서울/경기	이북(북한)
전체		5.1	43.0	24.3	2.2	15.6	18.1	0.3
거주 지역	수도권	5.7	22.5	25.7	0.8	15.6	41.7	0.5
	충청권	1.7	14.8	9.9	0.7	75.0	12.1	0.3
	호남권	0.3	3.2	94.5	0.5	1.8	1.8	0.0
	영남권	2.0	94.2	2.7	0.5	2.5	3.2	0.1
	강원/제주	48.3	8.7	2.8	35.3	6.1	9.0	0.3

[행] 설명 문장	2022년 주로 사용하는 지역어를 거주 지역별로 조사한 결과, 강원/제주 지역은 강원도 지역어의 사용 비중이 48.3%로 가장 높았던 반면, 북한 지역어의 사용 비중이 0.3%로 가장 낮았다.
--------------	---

- 각 행의 값이 문자와 숫자이면 다음을 참고하여 작성한다.
 - 숫자가 높은 값을 가지는 행을 선정하여 내용을 기술한다.

국가별 수출입

(단위: %)

구분	수출			수입		
	1위	2위	3위	1위	2위	3위
하이브리드	미국 (50.6)	프랑스 (7.6)	영국 (6.6)	독일 (35.1)	미국 (32.3)	일본 (12.1)
플러그인 하이브리드	미국 (67.6)	캐나다 (9.3)	영국 (8.6)	미국 (33.0)	스웨덴 (23.4)	일본 (21.9)
전기차	미국 (45.0)	캐나다 (9.5)	영국 (8.8)	중국 (58.9)	독일 (32.9)	미국 (4.2)

[행]
설명 문장

플러그인 하이브리드 차량의 주요 수출국은 미국이 67.6%로 가장 높게 나타났으며 이어서 캐나다가 9.3%, 영국이 8.6%로 나타났고, 플러그인 하이브리드 차량의 주요 수입국은 미국이 33.0%로 가장 높은 것으로 나타났고 스웨덴이 23.4%, 일본이 21.9%를 차지한 것으로 나타났다.

2.3. 열에 대한 서술문

2.3.1. 열 영역 선정

- 다음과 같은 기준으로 설명의 대상으로 삼을 열 영역을 선정한다.
 - 각 열의 합계를 비교하여 가장 높은 값을 가지는 열을 선정할 수 있다.

공연 시설: 항목별 평균 수입 내역_운영주체2별

(단위: 개, %, 천 원)

통계 분류 (1)	통계 분류 (2)	2022						
		시설 수	수입 총액	자체 수입	정부 예산 및 보조금	기타 수입	기부/ 후원금	전년도 이월금
		소계	소계	평균	평균	평균	평균	평균
		소계	소계	소계	소계	소계	소계	소계
전체	소계	992	1,481,950.6	258,494.7	1,068,232.9	43,067	37,893.9	74,262
운영 주체2	공공	461	2,734,961.3	304,466	2,193,017.2	48,016.9	32,233	157,228.1
	민간	531	394,120.2	218,583.6	91,725.3	38,769.6	42,808.5	2,233.1

- 첫 열이나 마지막 열에 각 열의 합계가 제시되어 있으면 그 열을 설명 영역으로 선정할 수 있다.

공연 단체: 분장실 보유 및 운영 현황_주요 활동 장르별

(단위: 개, %)

존속 기간별(1)	2023							
	국제회의 관련 총근로자 (명)	구성비 (%)	국제회의 전담 근로자		국제회의 주 근로자		국제회의 보조 근로자	
			근로자수 (명)	구성비 (%)	근로자수 (명)	구성비 (%)	근로자수 (명)	구성비 (%)
전체	756	100.0	750	100.0	4	100.0	2	100.0
5년 미만	90	11.9	90	12.0	-	-	-	-
5-10년 미만	147	19.4	147	19.6	-	-	-	-
10년 이상	520	68.8	514	68.5	4	100.0	2	100.0

- 해당 열의 제목이 여러 층위로 구성된 경우 상위 층위까지 선정한다.

전시 서비스 사업자의 고용 현황_고용 형태별

(단위: 개, 명, %)

특성별 (1)	특성별 (2)	2023								
		사례 수 (개)	총종사자 수 (명)				전시산업 종사자 수 (명)			총종사 자 수 대비 전 시산업 종사자 수 (%)
			소계	소계	남자	여자	소계	남자	여자	
전체	소계	1,324	60,625	39,544	21,081	4,942	3,516	1,426	8.2	
고용 형태별	정규직	1,324	49,841	31,392	18,449	3,914	2,648	1,266	7.9	
	비정규직	1,324	10,364	7,815	2,549	812	685	127	7.8	
	자영업자	1,324	420	336	83	216	183	33	51.4	

2.3.2. 열 설명 문장 작성

- 각 열의 값이 숫자이면 열 설명 문장은 다음을 참고하여 작성한다.
 - 해당 열의 합계가 다른 열의 합계보다 높은 값을 가지면 해당 열의 제목과 합계를 중심으로 기술한다.

공연 시설: 항목별 평균 수입 내역_운영주체2별

(단위: 개, %, 천 원)

통계 분류 (1)	통계 분류 (2)	2022						
		시설 수	수입 총액	자체 수입	정부 예산 및 보조금	기타 수입	기부/ 후원금	전년도 이월금
		소계	소계	평균	평균	평균	평균	평균
		소계	소계	소계	소계	소계	소계	소계
전체	소계	992	1,481,950.6	258,494.7	1,068,232.9	43,067	37,893.9	74,262
운영 주체2	공공	461	2,734,961.3	304,466	2,193,017.2	48,016.9	32,233	157,228.1
	민간	531	394,120.2	218,583.6	91,725.3	38,769.6	42,808.5	2,233.1

[열]
설명 문장

2022년 공연 시설 운영 주체의 수입 총액은 공공 기관이 27억 3,496만 1,300원, 민간 기관이 3억 9,412만 200원으로 나타났다.

- 해당 열의 값들이 증가세 혹은 감소세와 같은 일정한 추이를 보이면 그 내용을 기술한다.

국가유산진흥원 공연 사업 운영 현황

(단위: 명, 회)

시점	항목	합계	전수교육관	한국문화의 집	한국의 집	굿보러가자	기타
		소계	소계	소계	소계	소계	소계
2022	공연 횟수 (회)	327	160	136	22	6	3
	관람 인원 (명)	31,657	11,204	12,736	1,148	3,160	3,409
2023	공연 횟수 (회)	346	187	143	7	6	3
	관람 인원 (명)	37,346	13,358	12,389	720	4,510	6,369
2024	공연 횟수 (회)	331	165	157	-	6	3
	관람 인원 (명)	36,067	12,068	15,254	-	4,738	4,007

[열]
설명 문장

2022년부터 2024년까지 국가유산진흥원 공연 사업 운영 현황을 살펴본 결과, '굿보러가자'의 관람 인원은 2022년에 3,160명이었는데 2023년에 4,510명, 2024년에 4,738명으로 점차 증가하였다.

- 해당 열에서 높은 값을 가지는 칸을 중심으로 기술한다.

국제회의 근로자의 존속 기간별 현황

(단위: 명, %)

존속 기간별(1)	2023							
	국제회의 관련 총근로자 (명)	구성비 (%)	국제회의 전담 근로자		국제회의 주 근로자		국제회의 보조 근로자	
		소계	소계	근로자수 (명)	구성비 (%)	근로자수 (명)	구성비 (%)	근로자수 (명)
전체	756	100.0	750	100.0	4	100.0	2	100.0
5년 미만	90	11.9	90	12.0	-	-	-	-
5-10년 미만	147	19.4	147	19.6	-	-	-	-
10년 이상	520	68.8	514	68.5	4	100.0	2	100.0

[열]
설명 문장

2023년 국제 회의 관련 총 근로자는 756명으로, 이 가운데 존속 기간이 10년 이상인 근로자가 520명으로 대부분을 차지했고, 존속 기간이 5~10년 미만인 근로자가 147명, 5년 미만인 근로자가 90명의 순으로 그 뒤를 이었다.

- 해당 열에서 낮은 값을 가지는 칸을 중심으로 기술한다.

추진 방식_주요 사업 1_업종별

(단위: 개, %)

구분	COL2	2022			
		사례수	현지 초청	한국 주관	상호 주관
전체		5,932	24.3	47.2	28.5
업종	공공 행정, 국방 및 사회보장 행정	296	17.4	43.6	39.0
업종	예술, 스포츠 및 여가 관련 서비스업	1,604	47.6	28.6	23.8
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	15.6	54.9	29.5

[열] 설명 문장	2022년 국제 문화 교류 사업체의 국제 문화 교류 사업 추진 방식을 업종별로 살펴본 결과, 한국 주관으로 국제 문화 교류 사업을 하는 업종은 예술, 스포츠 및 여가 관련 서비스업이 28.6%로 가장 낮았으며, 협회 및 단체, 수리 및 기타 개인 서비스업이 54.9%로 가장 높았다.
--------------	--

- 제목 열의 값이 문자 값이면 해당 내용을 나열하여 기술한다.

[지역어 경험] 주로 사용하는 지역어(사투리)_거주지역별

(단위: %)

표측 구분	표측 상세	2024						
		강원도	경상도	전라도	제주도	충청도	서울/경기	이북(북한)
전체		5.1	43.0	24.3	2.2	15.6	18.1	0.3
거주 지역	수도권	5.7	22.5	25.7	0.8	15.6	41.7	0.5
	충청권	1.7	14.8	9.9	0.7	75.0	12.1	0.3
	호남권	0.3	3.2	94.5	0.5	1.8	1.8	0.0
	영남권	2.0	94.2	2.7	0.5	2.5	3.2	0.1
	강원/제주	48.3	8.7	2.8	35.3	6.1	9.0	0.3

[열] 설명 문장	2024년 주로 사용하는 지역어 중에 이북어를 사용하는 사람들을 거주 지역별로 살펴본 결과, 수도권이 0.5%로 가장 높았으며, 충청권과 강원 및 제주가 0.3%, 영남권이 0.1% 순이었다.
--------------	---

- 각 열의 값이 문자이면 열 설명 문장은 다음을 참고하여 작성한다.
 - 해당 열에서 둘 이상의 값을 선정하여 비교 혹은 대조하는 문장을 작성한다.

추진 방식_주요 사업 1_업종별

(단위: 개, %)

구분	COL2	2022			
		사례수	현지 초청	한국 주관	상호 주관
전체		5,932	24.3	47.2	28.5
업종	공공 행정, 국방 및 사회보장 행정	296	17.4	43.6	39.0
업종	예술, 스포츠 및 여가 관련 서비스업	1,604	47.6	28.6	23.8
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	15.6	54.9	29.5

[열] 설명 문장	2022년 국제 문화 교류 사업체의 국제 문화 교류 사업 추진 방식을 업종별로 살펴본 결과, 한국 주관으로 국제 문화 교류 사업을 하는 업종은 예술, 스포츠 및 여가 관련 서비스업이 28.6%로 가장 낮았으며, 협회 및 단체, 수리 및 기타 개인 서비스업이 54.9%로 가장 높았다.
-----------	--

- 각 열의 값이 문자와 숫자이면 다음을 참고하여 작성한다.
 - 숫자가 높은 값을 가지는 열을 선정하여 내용을 기술한다.

국가별 수출입

(단위: %)

구분	수출			수입		
	1위	2위	3위	1위	2위	3위
하이브리드	미국 (50.6)	프랑스 (7.6)	영국 (6.6)	독일 (35.1)	미국 (32.3)	일본 (12.1)
플러그인 하이브리드	미국 (67.6)	캐나다 (9.3)	영국 (8.6)	미국 (33.0)	스웨덴 (23.4)	일본 (21.9)
전기차	미국 (45.0)	캐나다 (9.5)	영국 (8.8)	중국 (58.9)	독일 (32.9)	미국 (4.2)

[열] 설명 문장	하이브리드 및 전기차의 국가별 수입 현황을 살펴보면, 전기차 수입은 중국이 58.9%로 비중이 가장 컸고, 독일과 미국은 각각 35.1%와 33%로 유사했다.
-----------	--

2.4. 불연속 영역에 대한 서술문

2.4.1. 불연속 영역 선정

- 불연속 영역은 떨어져 있는 두 칸 이상을 선정해야 한다.

교류 방법_주요 사업1_업종별

(단위: 개, %)

구분	COL2	2022				
		사례 수	일방향 (발신형)	일방향 (수신형)	쌍방향	다자간
	전체	5,932	37.8	9.6	46.6	6.0
업종	공공 행정, 국방 및 사회보장 행정	296	20.8	10.6	61.0	7.6
업종	예술, 스포츠 및 여가 관 련 서비스업	1,604	41.2	13.5	37.6	7.7
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	37.7	7.9	49.1	5.2

- 이미 행 영역이나 열 영역으로 선정한 영역 내에서 불연속 영역을 선정하지 않는다.

국제회의 근로자의 존속 기간별 현황

(단위: 명, %)

존속 기간별(1)	2023							
	국제회의 관련 총근로자 (명)	구성비 (%)	국제회의 전담 근로자		국제회의 주 근로자		국제회의 보조 근로자	
			근로자수 (명)	구성비 (%)	근로자수 (명)	구성비 (%)	근로자수 (명)	구성비 (%)
전체	756	100.0	750	100.0	4	100.0	2	100.0
5년 미만	90	11.9	90	12.0	-	-	-	-
5-10년 미만	147	19.4	147	19.6	-	-	-	-
10년 이상	520	68.8	514	68.5	4	100.0	2	100.0

- 새롭게 지정한 칸과 이미 지정한 행 영역 혹은 열 영역 안의 칸을 비교하는 것은 가능하다.

전시 서비스 사업자의 고용 현황_고용 형태별

(단위: 개, 명, %)

특성별 (1)	특성별 (2)	2023								
		사례 수 (개)	총종사자 수 (명)				전시산업 종사자 수 (명)			총종사 자 수 대비 전 시산업 종사자 수 (%)
			소계	소계	남자	여자	소계	남자	여자	
전체	소계	1,324	60,625	39,544	21,081	4,942	3,516	1,426	8.2	
고용 형태별	정규직	1,324	49,841	31,392	18,449	3,914	2,648	1,266	7.9	
	비정규직	1,324	10,364	7,815	2,549	812	685	127	7.8	
	자영업자	1,324	420	336	83	216	183	33	51.4	

- 행, 열의 수가 적어 이미 선정한 행 영역, 열 영역과 중복되지 않으면서 떨어져 있는 둘 이상의 칸을 선정할 수 없는 경우, 둘 이상의 연속적인 칸을 선정할 수 있다.

공연 시설: 항목별 평균 수입 내역_운영주체2별

(단위: 개, %, 천 원)

통계 분류 (1)	통계 분류 (2)	2022						
		시설 수	수입 총액	자체 수입	정부 예산 및 보조금	기타 수입	기부/ 후원금	전년도 이월금
		소계	소계	평균	평균	평균	평균	평균
		소계	소계	소계	소계	소계	소계	소계
전체	소계	992	1,481,950.6	258,494.7	1,068,232.9	43,067	37,893.9	74,262
운영 주체2	공공	461	2,734,961.3	304,466	2,193,017.2	48,016.9	32,233	157,228.1
	민간	531	394,120.2	218,583.6	91,725.3	38,769.6	42,808.5	2,233.1

- 비교나 대조하여 기술할 만한 내용이 있는 두 칸 이상을 선정한다.

[지역어 경험] 주로 사용하는 지역어(사투리)_거주지역별

(단위: %)

표측 구분	표측 상세	2024						
		강원도	경상도	전라도	제주도	충청도	서울/경기	이북(북한)
전체		5.1	43.0	24.3	2.2	15.6	18.1	0.3
거주 지역	수도권	5.7	22.5	25.7	0.8	15.6	41.7	0.5
	충청권	1.7	14.8	9.9	0.7	75.0	12.1	0.3
	호남권	0.3	3.2	94.5	0.5	1.8	1.8	0.0
	영남권	2.0	94.2	2.7	0.5	2.5	3.2	0.1
	강원/제주	48.3	8.7	2.8	35.3	6.1	9.0	0.3

2.4.2. 불연속 영역 설명 문장 작성

- 불연속 영역으로 지정한 칸의 값을 비교 혹은 대조하여 작성한다.

교류 방법_주요 사업_업종별						
(단위: 개, %)						
구분	COL2	2022				
		사례 수	일방향 (발신형)	일방향 (수신형)	쌍방향	다자간
전체		5,932	37.8	9.6	46.6	6.0
업종	공공 행정, 국방 및 사회보장 행정	296	20.8	10.6	61.0	7.6
업종	예술, 스포츠 및 여가 관 련 서비스업	1,604	41.2	13.5	37.6	7.7
업종	협회 및 단체, 수리 및 기타 개인 서비스업	4,032	37.7	7.9	49.1	5.2
[불연속 영역] 설명 문장	2022년 국제 문화 교류 사례를 업종별로 보면 공공 행정, 국방 및 사회 보장 행정 사업, 협회 및 단체, 수리 및 기타 개인 서비스업은 쌍방향 교류의 비중이 각각 61%, 49.1%를 기록해 가장 컸고, 예술, 스포츠 및 여가 관련 서비스업은 일방향 (발신형) 교류의 비중이 41.2%로 가장 컸다.					

- 불연속 영역 설명 문장 작성 시, 두 칸 이상의 복합적인 정보를 활용하여 표의 내용을 작성한다.

공연 시설: 항목별 평균 수입 내역_운영주체2별

(단위: 개, %, 천 원)

통계 분류 (1)	통계 분류 (2)	2022						
		시설 수	수입 총액	자체 수입	정부 예산 및 보조금	기타 수입	기부/ 후원금	전년도 이월금
		소계	소계	평균	평균	평균	평균	평균
		소계	소계	소계	소계	소계	소계	소계
전체	소계	992	1,481,950.6	258,494.7	1,068,232.9	43,067	37,893.9	74,262
운영 주체2	공공	461	2,734,961.3	304,466	2,193,017.2	48,016.9	32,233	157,228.1
	민간	531	394,120.2	218,583.6	91,725.3	38,769.6	42,808.5	2,233.1

[불연속 영역] 설명 문장	2022년 공연 시설 운영 주체의 주 수입 내역을 살펴보면, 공공 기관은 정부 예산 및 보조금 명목 수입이 21억 9,301만 7,200원으로 가장 많았고, 민간 기관은 기부 및 후원금 명목 수입이 4,280만 8,500원으로 가장 많았다.
-------------------	--

- 총합, 비율, 평균, 차이 등 기본 통계를 활용하여 불연속 영역 설명 문장을 작성할 수 있다.

"아침 식사 결식률 추이_성별 전체_소득 수준별

(단위: 명, %)

성별(1)	특성별(1)	특성별(2)	2023		
			응답자수 (명)	비율 (%)	표준오차
전체	소득수준별	하	1,341	33.3	1.9
		중하	1,365	31.3	1.6
		중	1,358	30.2	1.6
		중상	1,357	28.9	1.6
		상	1,344	25.4	1.5

[불연속 영역] 설명 문장	2023년 아침 식사 결식률을 소득 수준별로 조사한 결과, 총응답자 1,341명 중 33.3%의 비율을 보인 소득 분위 '하'가 아침 식사 결식률이 가장 높았고, 총응답자 1,344명 중 25.4%의 비율을 보인 소득 분위 '상'이 아침 식사 결식률이 가장 낮았다.
-------------------	--

2.5. 주의 사항

2.5.1. 설명 문장 관련 주의 사항

- 전문 용어는 경우에 따라 붙여쓰기와 띄어쓰기를 모두 허용하는 경우가 있으나, 맞춤법의 원칙과 허용 중 원칙에 따라 기술한다.

원칙	도시 개발 사업	(O)
허용	도시개발사업	(X)
원칙	손해 배상 청구	(O)
허용	손해배상청구	(X)

- 단, 고유 명사는 어문 규범에 어긋나더라도 하나의 단어로 보고 붙여 쓴다.

한국대학교 사범대학	(O)
국립국어원 기획연수부 기획운영과	(O)

- 어문 규범을 준수하여 작성한다.

<p>모 연령과 부 연령이 30-34세일 때 출생아 수가 58677명으로 가장 많고, 두 번째로 출생아 수가 많은 부모 연령대는 모 연령이 30-34세, 부 연령이 35-39세일 때이다.</p> <p style="text-align: center;">▼</p> <p>모 연령과 부 연령이 30-34세일 때 출생아 수가 58,677명으로 가장 많고, 두 번째로 출생아 수가 많은 부모 연령대는 모 연령이 30-34세, 부 연령이 35-39세일 때이다.</p>

- 문법적인 문장으로 작성한다.

<p>최근 1주의 캠퍼스 기준 시도별 학생 오미크론 발생 현황을 살펴보면 4월 12일에 1,534명, 4월 13일에 1,081명, 4월 14일에 823명으로 감소하는 추세를 보이다가 4월 17일에 226명이었으나 4월 18일에 432명으로 다시 증가하였다.</p> <p style="text-align: center;">▼</p> <p>최근 1주의 캠퍼스 기준 시도별 학생 오미크론 발생 현황을 살펴보면 4월 12일에 1,534명, 4월 13일에 1,081명, 4월 14일에 823명, 4월 17일에 226명으로 감소하는 추세를 보이다가 4월 18일에 432명으로 다시 증가하였다.</p>

- 표 설명 문장을 작성할 때에는 시제를 특정할 수 있는 경우를 제외하고 과거 시제로 표현한다.

4월 12일부터 4월 18일까지 시도별 발생 현황을 살펴보면 경기도가 2만 1,424명으로 가장 많은 학생이 **확진되고** 서울은 1만 2,202명, 경남은 6,298명의 학생이 **확진이다**.

▼

4월 12일부터 4월 18일까지 시도별 발생 현황을 살펴보면 경기도가 2만 1,424명으로 가장 많은 학생이 **확진되었으며**, 이어 서울은 1만 2,202명, 경남은 6,298명의 학생이 **확진되었다**.

- 표 제목을 활용하여 설명 문장을 작성할 때 문법, 표현 오류를 그대로 옮기지 않는다.

국가유산진흥원 주요사업 추진 현황_한국의집 운영별

(단위: 명, 건)

사업별(1)	사업별(2)	2022	2023	2024
한국의집 운영	전통음식 보존 보급(명)	69,118	47,194	39,401
	전통예술 보존 보급(명)	1,148	720	-
	전통 혼례 보급(건)	268	171	88
	전통문화 체험(명)	60	-	-
국가 무형유산 전수교육관 운영(명)	전통예술 공연	11,204	13,358	12,068
	전통공예 및 전통미술 전시	12,816	20,430	13,805
한국 문화의 집 운영(명)	전통예술 공연	12,736	12,389	15,254
	교육프로그램 운영	6,624	11,462	6,143
문화 유산 관광 자원화 (명)	소계	3,191,411	6,442,272	6,868,357
인천공항 전통문화체험관 운영(명)	소계	211,795	1,079,308	1,372,400

[행] 설명 문장	국가 유산 진흥원의 한국의 집 운영 사업 가운데 전통 음식 보존 보급 사업은 2022년에 69,118명, 2023년에 47,194명, 2024년에 39,401명으로, 해마다 그 규모가 줄어들고 있다.
---------------------	--

- 표의 내용과 관련 없는 내용이 서술문에 포함되지 않도록 주의한다.

"아침 식사 결식률 추이_성별 전체_소득수준별

(단위: 명, %)

성별(1)	특성별(1)	특성별(2)	2023		
			응답자수 (명)	비율 (%)	표준오차
전체	소득수준별	하	1,341	33.3	1.9
		중하	1,365	31.3	1.6
		중	1,358	30.2	1.6
		중상	1,357	28.9	1.6
		상	1,344	25.4	1.5

[불연속 영역]
설명 문장

2023년 아침 식사 결식률을 **온라인 설문 조사를 통해** 소득수준별로 조사한 결과, 총응답자 1,341명 중 33.3%의 비율을 보인 소득 분위 '하'가 아침 식사 결식률이 가장 높았고, 총응답자 1,344명 중 25.4%의 비율을 보인 소득 분위 '상'이 아침 식사 결식률이 가장 낮았다.

2.5.2. 기호 관련 주의 사항

- 표에서 사용된 증감 표시 기호에 유의하여 설명 문장을 작성한다. △, △은 음수(-)를 나타낸다.

23.9월 말 금융권 부동산 PF 대출 현황

(단위: 천 통행)

구분		'20말	'21말	'22말	'23.6말 (a)	'23.9말 (b)	변동 (a-b) 전년도 이월금
금융권 합계	대출 잔액	92.5	112.9	130.3	133.1	134.3	+1.2
	연체율	0.55	0.37	1.19	2.17	2.42	+0.24
은행	대출 잔액	26.1	32.5	39.4	43.1	44.2	+1.1
	연체율	0.29	0.02	0.01	0.23	-	△0.23
증권	대출 잔액	5.2	4.6	4.5	5.5	6.3	+0.8
	연체율	3.37	3.71	10.38	17.28	13.85	△3.43
보험	대출 잔액	36.4	42.0	44.3	43.7	43.3	△0.3
	연체율	0.11	0.07	0.60	0.73	1.11	+0.38
저축 은행	대출 잔액	6.9	9.5	10.5	10.0	9.8	△0.2
	연체율	2.43	1.22	2.05	4.61	5.56	+0.95
여신 전문	대출 잔액	13.8	19.5	26.8	26.0	26.0	△0.0
	연체율	0.28	0.47	2.20	3.89	4.44	+0.55
상호 금융	대출 잔액	4.1	4.9	4.8	4.8	4.7	△0.2
	연체율	0.30	0.09	0.09	1.12	4.18	+3.05

[불연속 영역]
설명 문장

2023년 9월 말 기준 증권의 연체율은 13.5%로 2023년 6월 말에 대비하여 3.43% 감소하였으나 상호 금융의 연체율은 2023년 9월 말 4.18%로 6월 말에 대비하여 3.05% 증가하였다.

- 단위에 통화 기호로 제시되어 있더라도 화폐 단위를 한글로 적는다.

교육단계별 학생 1인당 공교육비 지출 금액을 살펴보면 중등 교육이 \$17,038로 가장 많았고, 이어 초등 교육과 고등 교육이 각각 \$13,278, \$12,225로 조사되었다.

▼

교육 단계별 학생 1인당 공교육비 지출 금액을 살펴보면 중등 교육이 1만 7,038달러로 가장 많았고, 이어 초등 교육과 고등 교육이 각각 1만 3,278달러, 1만 2,225달러로 조사되었다.

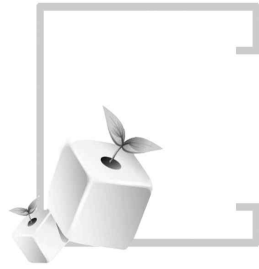
- 표에서 사용된 단위에 유의하여 설명 문장을 작성한다. ‘%’는 백분율을 나타내고 ‘%p’는 두 퍼센트 값 사이의 차이를 나타내는 단위로서 기존 퍼센트 대비 증감률을 나타낸다. 따라서 ‘%p’를 ‘%’로 대치하여 쓰지 않도록 주의해야 한다.

주요 세부 업종별 개·폐점률

(단위: %, %p)

구분		치킨	한식	커피	제과제빵	피자	주점	외식업 전체
개점률	2021년	17.3	35.1	25.1	13.4	20.3	21.9	26.2
	2022년	14.4	29.9	22.7	12.0	13.5	21.7	22.4
	증감(%p)	△2.9	△5.2	△2.4	△1.4	△6.8	△0.2	△3.8
폐점률	2021년	13.7	14.5	7.8	8.9	8.5	19.1	12.6
	2022년	14.2	18.2	9.2	9.8	12.3	13.1	14.5
	증감(%p)	0.5	3.7	1.4	0.9	3.8	△6.0	1.8

[불연속 영역] 설명 문장	2022년 한식 사업의 개점률은 29.9%로 2021년 35.1%에 비해 5.2%p 감소한 반면 2022년의 폐점률은 18.2%로 2021년 14.5%에 대비하여 3.7%p 증가하였다.
-------------------	---



부록

사진 설명 문장 작성 지침



사진 설명 문장 작성 지침

<목차>

1. 개요	1
1.1. 지침의 목적	1
1.2. 관련 사업	1
1.3. 주요 용어	1
1.4. 사진 설명 문장 예시	2
2. 사진 설명 문장의 유형	3
2.1. 대상 식별 문장	3
2.2. 속성 기술 문장	4
3. 대상 식별 문장 작성의 실제	5
3.1. 대상 식별 방법	5
3.1.1. 사물이 한 개인 경우	5
3.1.2. 사물이 여러 개인 경우	6
3.2. 대상 식별 문장 작성 방법	7
4. 속성 기술 문장 작성의 실제	11
4.1. 속성 기술 문장 작성의 일반 원리	11
4.1.1. 가시적 속성 기술 문장 작성	11
4.1.2. 비가시적 속성 기술 문장 작성	12

4.2. 식생활 설명 문장	13
4.2.1. 식생활 설명 속성	13
4.2.2. 식생활 속성 기술 문장 작성 방법	14
4.2.2.1. 가시적 속성 기술 문장 작성 방법	14
4.2.2.1.1. 색채	14
4.2.2.1.2. 형태	15
4.2.2.1.3. 재료	17
4.2.2.1.4. 용기	19
4.2.2.1.5. 구성 요소	20
4.2.2.1.6. 소재(素材)	21
4.2.2.1.7. 문양	23
4.2.2.2. 비가시적 속성 기술 문장 작성 방법	25
4.2.2.2.1. 조리법	25
4.2.2.2.2. 맛	25
4.2.2.2.3. 취식 방법	26
4.2.2.2.4. 지역	27
4.2.2.2.5. 역사	27
4.2.2.2.6. 용도	28
4.2.2.2.7. 이용자	29
4.3. 의생활 설명 문장	29
4.3.1. 의생활 설명 속성	29
4.3.2. 의생활 속성 기술 문장 작성 방법	30
4.3.2.1. 가시적 속성 기술 문장 작성 방법	30
4.3.2.1.1. 구성 요소	30
4.3.2.1.2. 크기	32
4.3.2.1.3. 길이	33
4.3.2.1.4. 소재(素材)	35
4.3.2.1.5. 형태	37

4.3.2.1.6. 문양	39
4.3.2.1.7. 색채	40
4.3.2.2. 비가시적 속성 기술 문장 작성 방법	42
4.3.2.2.1. 용도	42
4.3.2.2.2. 착용 방법	42
4.3.2.2.3. 착용자	43
4.3.2.2.4. 역사	44
4.4. 주생활 설명 문장	44
4.4.1. 주생활 설명 속성	44
4.4.2. 주생활 속성 기술 문장 작성 방법	45
4.4.2.1. 가시적 속성 기술 문장 작성 방법	45
4.4.2.1.1. 형태	45
4.4.2.1.2. 구성 요소	47
4.4.2.1.3. 구조	48
4.4.2.1.4. 자재/소재(素材)	50
4.4.2.1.5. 색채	51
4.4.2.1.6. 문양	53
4.4.2.1.7. 배치(건축물의 집합 건축물에서의 배치)	54
4.4.2.1.8. 위치(건축물 일부의 건축물 안에서의 위치)	56
4.4.2.2. 비가시적 속성 기술 문장 작성 방법	57
4.4.2.2.1. 용도	57
4.4.2.2.2. 소재지(所在地)	58
4.4.2.2.3. 분포	59
4.4.2.2.4. 이용자	59
4.4.2.2.5. 소유자	60
4.4.2.2.6. 역사	61
4.5. 건축 설명 문장	61
4.5.1. 건축 설명 속성	61

4.5.2. 건축 속성 기술 문장 작성 방법	62
4.5.2.1. 가시적 속성 기술 문장 작성 방법	62
4.5.2.1.1. 형태	62
4.5.2.1.2. 구성 요소	64
4.5.2.1.3. 구조	65
4.5.2.1.4. 자재/소재(素材)	67
4.5.2.1.5. 색채	68
4.5.2.1.6. 문양	70
4.5.2.1.7. 배치(집합 건축물을 이루는 단일 건축물들의 배치)	71
4.5.2.2. 비가시적 속성 기술 문장 작성 방법	73
4.5.2.2.1. 용도	73
4.5.2.2.2. 소재지(所在地)	74
4.5.2.2.3. 이용자	75
4.5.2.2.4. 소유자	75
4.5.2.2.5. 건축가	76
4.5.2.2.6. 건축 양식	76
4.5.2.2.7. 역사	77
4.6. 공예 설명 문장	78
4.6.1. 공예 설명 속성	78
4.6.2. 공예 속성 기술 문장 작성 방법	79
4.6.2.1. 가시적 속성 기술 문장 작성 방법	79
4.6.2.1.1. 구성 요소	79
4.6.2.1.2. 크기	80
4.6.2.1.3. 길이	82
4.6.2.1.4. 소재(素材)	83
4.6.2.1.5. 형태	85
4.6.2.1.6. 문양	86
4.6.2.1.7. 색채	88

4.6.2.2. 비가시적 속성 기술 문장 작성 방법	89
4.6.2.2.1. 용도	89
4.6.2.2.2. 기법	90
4.6.2.2.3. 소재지(所在地)	91
4.6.2.2.4. 역사	91
5. 주의 사항	93

1. 개요

1.1. 지침의 목적

이 지침은 한국어문화 관련 사진 설명 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 한다. 이 지침에서는 한국어문화 중 생활 분야인 식생활, 의생활, 주생활과 예술 분야인 건축, 공예 관련 문화 요소의 사진¹⁾에 대한 설명 문장을 작성하는 방안을 기술한다. 한국어문화를 담은 사진에 대한 설명 말뭉치를 구축함으로써 생성형 인공지능의 한국어문화에 대한 이해 능력 및 한국어 문장 생성 능력을 고도화할 수 있다.

1.2. 관련 사업

이 지침은 국립국어원의 '2025년 한국어문화 표·사진 설명 말뭉치 구축 연구' 사업의 일환으로서, 사진 설명 문장 말뭉치를 구축하기 위한 지침이다.

1.3. 주요 용어

사진 설명 문장 작성에 사용되는 용어는 아래와 같다.

- 한국어문화: 한국의 대표적인 유무형 문화, 한국 전통문화 및 한류를 이루는 문화
- 설명 대상: 사진 속의 사물 중 설명의 대상으로 삼은 사물
- 설명 속성: 설명 문장을 작성하기 위해 구분해 놓은 대상의 특성
- 설명 문장: 사진 속 대상에 대해 설명한 문장
 - 대상 식별 문장: 사진 속 대상이 무엇인지 식별하는 설명 문장
 - 속성 기술 문장: 사진 속 대상의 특정한 속성에 대해 기술하는 설명 문장


1) 이 지침에 사용된 모든 사진은 <한국민족문화대백과사전>의 공공 저작물이다.

1.4. 사진 설명 문장 예시

생활 > 식생활

↓

분류 태그



전주비빔밥

사진 제목(캡션)

대상 식별 문장	이 사진은 한국 전통 음식인 전주비빔밥의 사진이다.	
속성 기술 문장	색채	이 전주비빔밥 재료의 색채는 노란색, 녹색, 하얀색, 붉은색 등으로 다채롭다.
	형태	이 전주비빔밥에 들어 있는 달걀지단과 오이는 긴 직사각형 모양이다.
	재료	이 전주비빔밥에는 콩나물, 시금치, 오이, 달걀지단, 깨 등이 들어 있다.
	취식 방법	전주비빔밥은 밥, 소고기, 콩나물, 시금치, 고사리 등의 모든 재료를 고추장 양념과 함께 비벼서 먹는다.

↓

설명 속성

설명 문장

2. 사진 설명 문장의 유형

사진 설명 문장은 사진의 객체를 설명 대상으로 식별하는 문장인 '대상 식별 문장'과 사진 속 대상에서 관찰되는 한국언어문화 요소로서의 속성을 설명한 '속성 기술 문장'으로 구분한다. 사진 설명 문장을 작성할 때에는 대상 식별 문장 한 문장과 속성 기술 문장 네 문장을 작성한다.

속성 기술 문장은 사진 속 대상의 눈에 보이는 특성을 기술한 '가시적 속성 기술 문장'과 백과사전적 정보를 기술한 '비가시적 속성 기술 문장'으로 구분한다. 속성 기술 문장을 작성할 때에는 가능한 한 가시적 속성을 중심으로 기술하되 꼭 필요한 경우 비가시적 속성을 선택하여 기술한다.

2.1. 대상 식별 문장

대상 식별 문장은 사진 속 대상이 무엇인지, 사진 속에 있는 여러 대상 중 설명 대상이 무엇인지 식별하기 위해 작성하는 문장이다. 대상 식별 문장은 사진 속 한국언어문화 요소의 여러 속성들을 설명하기 전에 설명 대상을 식별하고 지정하는 역할을 한다. 사진 설명 문장 다섯 문장 중에서 대상 식별 문장은 한 문장을 작성한다.

생활 > 식생활



전주비빔밥

대상 식별 문장

이 사진은 한국 전통 음식인 전주비빔밥의 사진이다.

2.2. 속성 기술 문장

속성 기술 문장은 설명 대상으로 삼은 사물의 특정한 속성을 기술하는 문장이다. 한국어언어문화 요소는 주제 및 유형에 따라 공통적인 속성이 있는데, 이 지침에서 제시하고 있는 공통 설명 속성 위주로 설명 대상의 특성이 잘 드러나도록 설명 문장을 작성한다. 속성 기술 문장은 사진에서 눈으로 확인할 수 있는지 여부에 따라 가시적 속성 기술 문장과 비가시적 속성 기술 문장으로 구분한다.

생활 > 식생활



전주비빔밥

속성 기술 문장	색채	이 전주비빔밥 재료의 색채는 노란색, 녹색, 하얀색, 붉은색 등으로 다채롭다.
	형태	이 전주비빔밥에 들어 있는 달걀지단과 오이는 긴 직사각형 모양이다.
	재료	이 전주비빔밥에는 콩나물, 시금치, 오이, 달걀지단, 깨 등이 들어 있다.
	취식 방법	전주비빔밥은 밥, 소고기, 콩나물, 시금치, 고사리 등의 모든 재료를 고추장 양념과 함께 비벼서 먹는다.

3. 대상 식별 문장 작성의 실제

3.1. 대상 식별 방법

이 절에서는 사진 유형별 설명 문장 작성 방법을 기술한다. 사진에는 사물이 하나인 경우도 있고, 사물이 다수인 경우도 있다. 사진에 사물이 하나인 경우 그 사물에 대한 설명 문장을 작성하고, 사물이 다수인 경우 설명 대상을 특정한 후 설명 문장을 작성한다.

3.1.1. 사물이 한 개인 경우

- 사진 속 사물이 한 개인 경우, 해당 대상에 대한 대상 식별 문장과 속성 기술 문장을 작성한다.
- 사진 속에 있는 단일한 사물 자체에 대한 설명 문장을 작성한다. 즉, 배경이나 환경 등에 대한 설명 문장은 작성하지 않는다.



라면

대상 식별 문장		이 사진은 한국의 대표적 즉석식품인 라면의 사진이다.
속성 기술 문장	색채	이 라면의 면은 흐린 노란색이며 국물은 주황색이다.
	형태	이 라면의 면발은 구불구불한 모양이다.
	재료	이 라면에는 면, 달걀, 미역이 들어 있다.
	조리법	라면은 적정량의 물을 끓이다가 스프와 면을 넣고 면이 익을 때까지 끓여서 만든다.

3.1.2. 사물이 여러 개인 경우

○ 사진 속 사물이 여러 개인데 종류가 동일한 경우

- 사진 속 사물이 여러 개인데 종류가 동일한 경우, 해당 사물들의 특성을 종합하여 설명 문장을 작성할 수도 있고, 해당 사물들 중 하나를 설명 대상으로 삼아 설명 문장을 작성할 수도 있다.



옛강정

대상 식별 문장		이 사진은 한국의 전통 과자인 옛강정의 사진이다.
속성 기술 문장	색채	이 옛강정의 색채는 노란색, 진한 갈색, 초록색, 검은색 등이다.
	형태	이 옛강정은 한 입 크기의 원 모양과 사각 모양이다.
	재료	이 옛강정은 잣, 깨 등을 주재료로 사용했다.
	조리법	옛강정은 땅콩, 잣, 참깨, 들깨, 콩 등을 볶고 조청에 버무려 굳힌 다음 알맞은 두께로 밀어 칼로 적당하게 잘라 만든다.

○ 사진 속 사물이 여러 개인데 종류가 다른 경우

- 사진 속 사물이 여러 개인데 종류가 다른 경우, 해당 사진의 제목(캡션)을 참고하여 사진 제목이 가리키는 사물을 설명 대상으로 삼아 설명 문장을 작성한다.



따로국밥

대상 식별 문장		이 사진은 한국의 향토 음식 중 하나인 따로국밥의 사진이다.
속성 기술 문장	색채	이 따로국밥 재료의 색채는 녹색, 붉은색 등이다.
	재료	이 따로국밥에는 무, 파 등이 들어 있다.
	맛	따로국밥은 파, 마늘, 부추 등이 들어가 매운맛이 난다.
	조리법	따로국밥은 소뼈를 장시간 고아 낸 국물에 고춧가루, 파, 부추, 마늘 등을 통해 만든 양념을 섞어 다시 끓여 내는 방식으로 만든다.

3.2. 대상 식별 문장 작성 방법

○ 대상 식별 문장 작성 요령

- 대상 식별 문장은 한국언어문화 요소로서의 '유형 및 명칭'을 기술한다.



석화로

대상 식별 문장	이 사진은 한국 전통 난방 도구인 석화로의 사진이다.
----------	-------------------------------

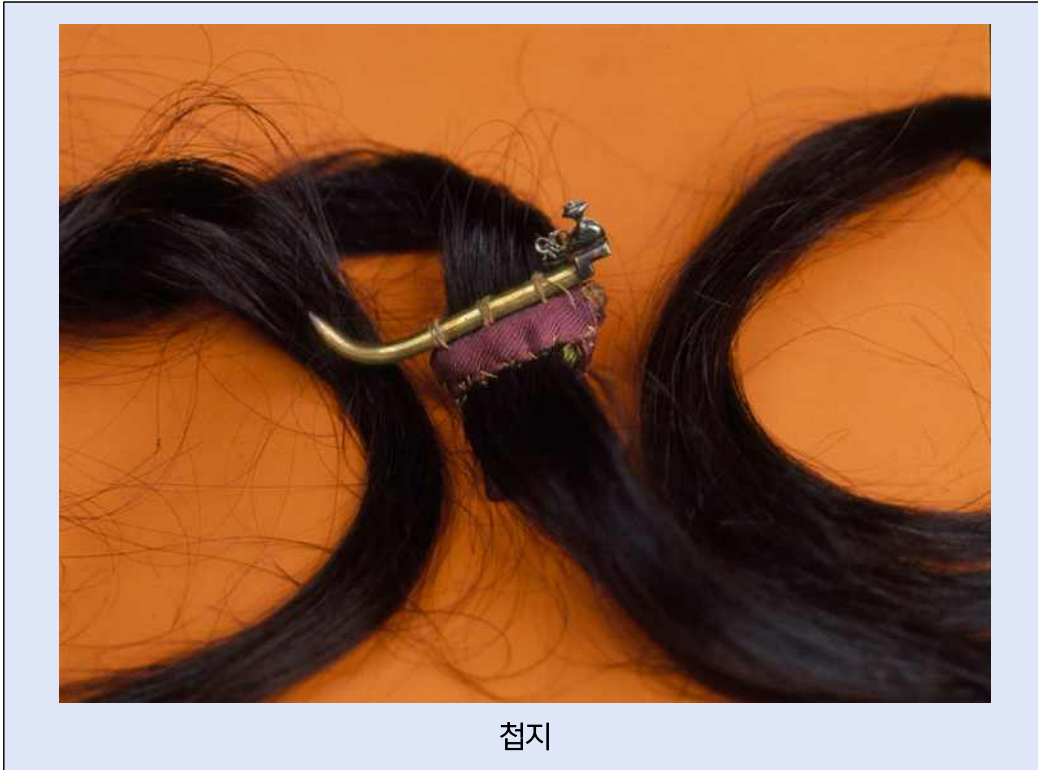
- 사진 속 한국언어문화 요소의 명칭이 대상 식별 문장에 포함되도록 작성한다.



평상

대상 식별 문장	이 사진은 한국 전통 침상인 평상의 사진이다.
----------	---------------------------

- 가능하면 유개념과 종 차의 관계가 드러나도록 작성한다.



첩지

대상 식별 문장	이 사진은 머리에 착용하는 장신구인 첩지의 사진이다.
----------	-------------------------------

○ 대상 식별 문장의 형식

- 대상 식별 문장의 주제부는 '이 사진'으로 한다.



마고자

대상 식별 문장	이 사진은 한국 전통 의복인 마고자의 사진이다.
----------	----------------------------

- 대상 식별 문장의 서술부에는 '설명 대상'이 포함되도록 한다.



백목화

대상 식별 문장

이 사진은 한국 전통 신발 중 하나인 백목화의 사진이다.

4. 속성 기술 문장 작성의 실제

이 절에서는 생활-식생활·의생활·주생활, 예술-건축·공예 등의 한국어문화 분야에 따른 속성별 설명 문장 작성 방법을 기술한다. 이 절에서 제시하는 설명 문장 틀은 설명 문장 작성자가 설명 문장을 보다 쉽게 작성할 수 있게 하기 위한 예시로서 제공하는 것이므로, 반드시 이 설명 문장 틀에 따라 문장을 작성해야 하는 것은 아니다. 설명 문장 작성자는 이 절에서 제시하는 설명 문장 틀을 적절하게 활용하여 사진 설명 문장을 작성할 수 있다.

4.1. 속성 기술 문장 작성의 일반 원리

4.1.1. 가시적 속성 기술 문장 작성

○ 가시적 속성 기술 문장 작성 요령

- 설명 문장은 사진 속 대상에서 관찰되는 한국어문화 요소로서의 특징을 설명한다. 즉 사진에서 가시적으로 드러나는 대상의 특징을 기술하는 문장으로 작성한다.
- 사진 속 한국어문화 요소에서 두드러지게 드러나는 특성을 묘사한다.
- 한국어문화 요소의 유형에 따라 그 특성을 잘 드러낼 수 있는 속성을 우선적으로 기술한다.

○ 가시적 속성 기술 문장의 형식

- 가시적 속성 기술 문장의 주제부는 '지시 대명사+대상'으로 한다.

	
마고자	
색채	이 마고자 는 연한 분홍색이다.

4.1.2. 비가시적 속성 기술 문장 작성

○ 비가시적 속성 기술 문장 작성 요령

- 사진에서는 가시적으로 드러나지 않지만, 사진 속 한국어문화 요소를 설명할 때 긴요한 백과사전적 정보를 기술한다.
 - ✓ 단, 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 선택하여 설명 문장을 작성할 수 있다. 이때에는 설명 문장에서 그 속성이 해당 한국어문화 요소의 일반적인 속성임을 드러내야 한다.(OO은 주로~, 일반적으로~ 등)

	
한천	
(비)재료	한천은 주로 우뚝가사리 과의 해초로 만들어진다.

○ 비가시적 속성 기술 문장의 형식

- 비가시적 속성 기술 문장의 주제부는 '대상'으로 한다.

	
안채	
용도	안채는 여성들이 일상생활, 가사일 등을 하는 공간이다.

4.2. 식생활 설명 문장

4.2.1. 식생활 설명 속성

가시적 속성	비가시적 속성
색채 형태 재료(주재료, 부재료, 양념, 고명) * 용기(액체류, 장류/소스류 한정) * 구성 요소(도구류 한정) * 소재(素材)(도구류 한정) * 문양(도구류 한정)	조리법 맛 취식 방법 지역 역사 * 용도(도구류 한정) * 이용자(도구류 한정)

- 음식 및 식생활 관련 도구의 종류에 따라 위 표의 설명 속성 중 적절한 속성을 선택하여 기술한다.
- 색채, 형태, 재료 등의 가시적 속성은 사진 속 대상의 전체와 부분을 설명할 때 모두 활용할 수 있다.
- 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 작성할 수 있다.
- 가시적 속성
 - 색채: 사진 속 음식의 빛깔
 - 형태: 사진 속 음식의 모양
 - 재료: 사진 속 음식에 쓰인 재료(주재료, 부재료, 양념, 고명)
 - 용기: 사진 속 음식이 담겨 있는 그릇
 - 구성 요소: 사진 속 도구를 이루는 데 필요한 요소
 - 소재(素材): 사진 속 도구를 만드는 데에 바탕이 되는 재료
 - 문양: 사진 속 도구 표면의 무늬
- 비가시적 속성
 - 조리법: 사진 속 음식을 만드는 방법
 - 맛: 사진 속 음식의 일반적인 맛
 - 취식 방법: 사진 속 음식을 먹는 방법
 - 지역: 사진 속 음식이 발달한 지역 혹은 유래된 지역
 - 역사: 사진 속 음식과 관련된 역사적 사실
 - 용도: 사진 속 도구의 쓰임새
 - 이용자: 사진 속 도구를 주로 이용하는 사람(부류)

4.2.2. 식생활 속성 기술 문장 작성 방법

4.2.2.1. 가시적 속성 기술 문장 작성 방법

4.2.2.1.1. 색채

- 사진 속 음식에서 두드러지게 드러나는 빛깔을 기술한다. 음식의 전체적인 색채를 설명할 수도 있고 음식 재료의 색채를 설명할 수도 있다.
- 색채 설명 문장의 주제부는 '지시 대명사+대상' 혹은 '지시 대명사+대상(재료)의 색채' 등으로 한다.



전주비빔밥

색채	이 전주비빔밥 재료의 색채는 노란색, 녹색, 하얀색, 붉은색 등으로 다채롭다.
----	---

- 색채 설명 문장의 서술부는 색채 명사 또는 색채 형용사를 이용하여 기술한다.



라면

색채	이 라면의 면은 흐린 노란색 이며 국물은 주황색 이다.
----	--

- 색채 설명 문장 틀

색채	이 [대상]은 [색채 명사+이다/색채 형용사]
	이 [대상]의 색채는 [색채 명사+이다/색채 형용사]
	이 [대상] [재료]의 색채는 [색채 명사+이다/색채 형용사]

4.2.2.1.2. 형태

- 사진 속 음식의 모양을 기술한다. 음식의 전체적인 모양을 설명할 수도 있고 음식 재료의 모양이나 손질된 모양을 설명할 수도 있다.
- 형태 설명 문장의 주제부는 '지시 대명사+대상' 혹은 '지시 대명사+대상의 재료' 등으로 한다.



메밀묵

<p>형태</p>	<p>이 메밀묵은 직사각형 모양이다.</p>
-----------	--------------------------



호박선

<p>형태</p>	<p>이 호박선의 호박은 단면이 깔끔하게 잘린 원통 모양이다.</p>
-----------	--



호박선

재료	이 호박선에는 호박, 달걀지단, 고기, 김, 실고추 등이 들어 있다.
	이 호박선의 재료는 호박, 달걀지단, 고기, 김, 실고추 등이다.

- 재료 설명 문장의 서술부는 '~이/가 들어 있다'; '~(등)이다' 등으로 작성한다.



전주비빔밥

재료	이 전주비빔밥에는 콩나물, 시금치, 오이, 고기, 계란 등이 들어 있다.
	이 전주비빔밥의 재료는 콩나물, 시금치, 오이, 고기, 계란 등이다.



담복장

용기	이 담복장은 검은색의 항아리에 담겨 있다.
	이 담복장은 검은색의 항아리에 들어 있다.

- 용기 설명 문장 틀

용기	이 [대상]은 []에 담겨 있다/들어 있다
----	----------------------------

4.2.2.1.5. 구성 요소

- 사진 속 식생활 관련 도구를 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 구성 요소 설명 문장의 주제부는 '지시 대명사+대상'으로 한다.



맷돌

구성 요소	이 맷돌은 작은 구멍이 뚫린 돌과 손잡이로 구성되어 있다.
-------	----------------------------------



놋그릇

소재(素材)

이 놋그릇은 놋쇠로 만들어졌다.



맷돌

소재(素材)

이 맷돌의 손잡이는 나무로 만들어졌다.



떡살

문양	이 떡살에는 꽃, 빗살 등의 문양이 새겨져 있다.
----	-----------------------------

- 문양 설명 문장의 서술부는 '~문양이/무늬가 있다', '~이/가 그려져 있다', '~이/가 새겨져 있다' 등으로 작성한다.



독배기

문양	이 독배기의 윗부분에는 요철 형태가 연결된 기하학적 문양이 있다.
----	--------------------------------------



따로국밥

맛	따로국밥은 고춧가루, 마늘 등이 들어가 매운맛이 난다.
---	--------------------------------

4.2.2.2.3. 취식 방법

- 사진 속 음식을 먹는 방법 및 순서를 기술한다. 구절판, 열구자탕 등과 같이 특수한 방법으로 먹는 음식의 경우 취식 방법을 설명할 수 있다.
- 취식 방법 설명 문장의 주제부는 ‘대상’으로 한다.



구절판

취식 방법	구절판은 구절판잔합에 담긴 여덟 가지 음식을 밀전병에 조금씩 싸서 먹는다.
-------	---

4.2.2.2.4. 지역

- 사진 속 음식이 유래한 지역이나 발달한 지역을 기술한다. 닭갈비, 갓김치 등 지역색이 강한 음식의 경우 지역을 설명할 수 있다.
- 지역 설명 문장의 주제부는 '대상'으로 한다.



갓쌈김치(갓김치)

지역	갓김치는 갓의 주요 재배지인 여수에서 발달했다.
----	----------------------------

4.2.2.2.5. 역사

- 사진 속 음식과 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.
- 역사 설명 문장의 주제부는 '대상'으로 한다.



감자밥

역사	감자밥은 한때 한국 식생활에서 주식이었다.
----	-------------------------

4.2.2.2.6. 용도

- 사진 속 식생활 관련 도구의 쓰임새를 기술한다. 해당 도구의 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
- 용도 설명 문장의 주제부는 '대상'으로 한다.



맷돌

용도	맷돌은 쌀, 밀, 옥수수 등의 곡식을 가는 데에 사용하는 도구이다.
----	---------------------------------------

4.2.2.2.7. 이용자

- 사진 속 식생활 관련 도구를 주로 이용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 이용자가 한정적인 경우 이용자를 설명 문장으로 작성할 수 있다.
- 이용자 설명 문장의 주제부는 '대상'으로 한다.

	
놋그릇	
이용자	놋그릇은 조선시대에 주로 중상류층에서 식기로 사용하였다.

4.3. 의생활 설명 문장

4.3.1. 의생활 설명 속성

가시적 속성	비가시적 속성
구성 요소 크기 길이 소재(素材) 형태 문양 색채	용도 착용 방법 착용자 역사

- 의류, 장신구류, 잡화류 및 의생활 관련 도구의 종류에 따라 위 표의 설명 속성 중 적절한 속성을 선택하여 기술한다.
- 크기, 길이, 소재, 형태, 문양, 색채 등의 사진 대상 설명 속성은 사진 대상의 전체와 부분을 설명할 때 모두 활용할 수 있다.
- 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 작성할 수 있다.
- 가시적 속성
 - 구성 요소: 사진 속 의류 및 잡화를 이루는 데 필요한 요소
 - 크기: 사진 속 의류 및 잡화의 넓이, 부피 등이 크고 작은 정도
 - 길이: 사진 속 의류 및 잡화의 한끝에서 다른 한끝까지의 거리
 - 소재(素材): 사진 속 의류 및 잡화를 만드는 데에 바탕이 되는 재료
 - 형태: 사진 속 의류 및 잡화의 모양
 - 문양: 사진 속 의류 및 잡화 표면의 무늬
 - 색채: 사진 속 의류 및 잡화의 빛깔
- 비가시적 속성
 - 용도: 사진 속 의류 및 잡화의 쓰임새
 - 착용 방법: 사진 속 의류 및 잡화를 입거나 차는 방법
 - 착용자: 사진 속 의류 및 잡화를 착용하는 사람(부류)
 - 역사: 사진 속 의류 및 잡화와 관련된 역사적 사실

4.3.2. 의생활 속성 기술 문장 작성 방법

4.3.2.1. 가시적 속성 기술 문장 작성 방법

4.3.2.1.1. 구성 요소

- 사진 속 의류 및 잡화를 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 구성 요소 설명 문장의 주제부는 '지시 대명사+대상'으로 한다.



두루마기

구성 요소	이 <u>두루마기</u> 는 <u>솜, 무, 깃, 옷고름</u> 등으로 이루어져 있다.
-------	--

- 구성 요소 설명 문장의 서술부는 '~(으)로 이루어져 있다', '~(으)로 구성되어 있다' 등으로 작성한다.



삼작노리개

구성 요소	이 <u>노리개</u> 는 <u>매듭, 술, 은장식</u> 등으로 이루어져 있다.
-------	---



치마

크기	이 치마는 폭이 넓다 .
----	----------------------

- 크기 설명 문장 틀

크기	이 [대상]은/는 []
	이 [대상]의 [구성 요소]은/는 []
	이 [대상] [구성 요소]의 크기/넓이/폭은 []

4.3.2.1.3. 길이

- 사진 속 의류 및 잡화의 한끝에서 다른 한끝까지의 거리를 기술한다. 구성 요소들의 길이를 비교, 대조하여 설명하거나 착용 시 기장을 활용하여 설명 문장을 작성할 수 있다.
- 길이 설명 문장의 주제부는 '지시 대명사+대상(대상의 구성 요소)' 혹은 '지시 대명사+대상(대상의 구성 요소)의 길이' 등으로 한다.

4.3.2.1.4. 소재(素材)

- 사진 속 의류 및 잡화를 만드는 데에 바탕이 되는 재료를 기술하되, 사진에서 의류 및 잡화의 소재가 식별되는 경우에 설명 문장을 작성한다. 소재의 얇고 두꺼운 정도가 식별이 되는 경우, 소재와 함께 소재의 두께도 함께 기술할 수 있다. 의류 및 잡화의 전체적인 소재를 설명할 수도 있고 구성 요소의 소재를 설명할 수도 있다.
- 소재 설명 문장의 주제부는 '지시 대명사+대상(대상의 구성 요소)' 혹은 '지시 대명사+대상(대상의 구성 요소)의 소재' 등으로 한다.



족두리

소재(素材)	이 <u>족두리의 장식</u> 은 옥, 진주 등으로 만들어졌다.
	이 <u>족두리 장식의 소재</u> 는 옥, 진주 등이다.



휘항

소재(素材)	이 휘항은 두꺼운 모피로 만들어졌다.
--------	----------------------

- 소재 설명 문장의 서술부는 ‘~(으)로 만들어졌다’, ‘~이/가 쓰였다’; ‘~등이다’ 등으로 작성한다.



두루마기

소재(素材)	이 두루마기는 흰색과 검은색의 비단으로 만들어졌다.
--------	------------------------------



타래버선

형태	이 타래버선의 앞코는 뽀족하게 위로 솟아 있고 뒷버선목은 곡선 형태로 오목하게 들어가 있다.
----	---

- 형태 설명 문장의 서술부는 '~모양이다', '~형태이다' 등으로 작성한다.



오각대

형태	이 오각대는 직사각형으로 각진 모양이다.
----	------------------------



전모

문양	이 전모의 윗부분에는 태극 문양이 그려져 있다.
----	----------------------------

- 문양 설명 문장 틀

문양	이 [대상]에는 []이/가 있다
	이 [대상]의 [문양의 위치]에는 [문양/무늬]이/가 [] 아/어 있다

4.3.2.1.7. 색채

- 사진 속 의류 및 집회에서 두드러지게 드러나는 빛깔을 기술한다. 의류 및 집회의 전체적인 색채를 설명할 수도 있고 구성 요소의 색채를 설명할 수도 있다.
- 색채 설명 문장의 주제부는 '지시 대명사+대상' 혹은 '지시 대명사+대상(대상의 구성 요소)의 색채' 등으로 한다.



구군복

색채	이 구군복의 소매는 붉은색과 노란색이고, 앞판은 검은색이다.
----	-----------------------------------

- 색채 설명 문장의 서술부에는 다양한 색채 명사 및 색채 형용사가 포함되도록 한다.



족두리

색채	이 족두리는 검은색으로 되어 있고, 족두리에 달린 장식의 색채는 금색, 주황색, 초록색 등으로 다채롭다.
----	--

- 색채 설명 문장 틀

색채	이 [대상]은/는 [색채 명사+이다/색채 형용사]
	이 [대상]의 색채는 [색채 명사+이다/색채 형용사]
	이 [대상] [구성 요소]의 색채는 [색채 명사+이다/색채 형용사]

4.3.2.2. 비가시적 속성 기술 문장 작성 방법

4.3.2.2.1. 용도

- 사진 속 의류 및 잡화의 쓰임새를 기술한다. 해당 의류 및 잡화의 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
- 용도 설명 문장의 주제부는 ‘대상’으로 한다.
 - ✓ 사진 제목이 특정 대상의 고유한 명칭이어서 한국어문화 요소의 특성을 기술하는 것의 효용이 떨어지는 경우, 일반적인 명칭으로 바꾸어 설명 문장을 작성한다.

	
영락봉잠	
용도	비녀 는 과거 여성들이 쪽을 진 머리를 풀어지지 않게 하기 위하여 꽂던 장신구이다.

4.3.2.2.2. 착용 방법

- 사진 속 의복 및 장신구를 입거나 차는 방법 및 순서를 기술한다. 특수한 방식으로 착용하는 의복 및 장신구의 경우 착용 방법을 설명할 수 있다.
- 착용 방법 설명 문장의 주제부는 ‘대상’으로 한다.



오각대

<p>착용 방법</p>	<p>오각대는 관복의 겨드랑 밑에 있는 고리에 끼워 착용한다.</p>
--------------	--

4.3.2.2.3. 착용자

- 사진 속 의복 및 장신구를 주로 착용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 착용자가 한정적인 경우 착용자를 설명 문장으로 작성할 수 있다.
- 착용자 설명 문장의 주제부는 '대상'으로 한다.



구군복

<p>착용자</p>	<p>구군복은 조선시대 무관들이 입었다.</p>
------------	----------------------------

4.3.2.2.4. 역사

- 사진 속 의류 및 잡화와 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.
- 역사 설명 문장의 주제부는 ‘대상’으로 한다.

	
<p>고무신</p>	
<p>역사</p>	<p>고무신은 20세기 초부터 생산되었다.</p>

4.4. 주생활 설명 문장

4.4.1. 주생활 설명 속성

가시적 속성	비가시적 속성
형태 구성 요소 구조(외부/내부) 자재(건축물)/소재(素材)(주생활 관련 도구) 색채 문양 배치(건축물의 집합 건축물에서의 배치) 위치(건축물 일부의 건축물 안에서의 위치)	용도 소재지(所在地) 분포 이용자 소유자 역사

- 집합 건축물, 단일 건축물, 건축물 일부(부분) 등 건축물의 유형, 주생활 관련 도구의 유형에 따라 위 표의 설명 속성 중 적절한 속성을 선택하여 기술한다.
- 형태, 구성 요소, 구조, 자재, 색채, 문양 등의 사진 대상 설명 속성은 사진 대상의 전체와 부분을 설명할 때 모두 활용할 수 있다.
- 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 작성할 수 있다.
- 가시적 속성
 - 형태: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구의 모양
 - 구성 요소: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 이루는 데 필요한 요소
 - 구조(외부/내부): 사진 속 건축물 및 건축물 일부의 구성 요소들을 엮은 체계
 - 자재/소재(素材): 사진 속 건축물 및 건축물 일부, 주생활 관련 도구에 쓰인 재료
 - 색채: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구의 빛깔
 - 문양: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 장식하기 위한 여러 가지 모양
 - 배치(건축물의 집합 건축물에서의 배치): 사진 속 건축물이 집합 건축물 내에서 벌여 놓여 있는 모양
 - 위치(건축물 일부의 건축물 안에서의 위치): 사진 속 건축물 일부(부분)의 건축물 안에서의 위치
- 비가시적 속성
 - 용도: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구의 쓰임새
 - 소재지(所在地): 사진 속 건축물 및 건축물 일부가 자리 잡고 있는 곳
 - 분포: 사진 속 건축물이 주로 나타나는 지역
 - 이용자: 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 주로 이용하는 사람(부류)
 - 소유자: 사진 속 건축물을 소유한 사람
 - 역사: 사진 속 건축물 및 건축물 일부와 관련된 역사적 사실

4.4.2. 주생활 속성 기술 문장 작성 방법

4.4.2.1. 가시적 속성 기술 문장 작성 방법

4.4.2.1.1. 형태

- 사진 속 건축물 및 건축물 일부, 주생활 관련 도구의 모양을 기술한다. 건축물 및 건축물 일부, 주생활 관련 도구의 전체적인 모양을 설명할 수도 있고 구성 요소의 모양을 설명할 수도 있다.
- 형태 설명 문장의 주제부는 '지시 대명사+대상' 혹은 '지시 대명사+대상의 구성 요소' 등으로 한다.



무주 귀틀집

형태	이 귀틀집의 지붕은 곡선 형태이다.
----	---------------------

- 형태 설명 문장의 서술부는 '~모양이다', '~형태이다' 등으로 작성한다.



김치광

형태	이 김치광은 안쪽이 비어 있는 세모뿔 형태이며, 김치광의 입구는 세로로 긴 병 모양이다.
----	---



사랑채

구조	이 사랑채는 다섯 개의 기둥이 보를 받치고 있는 구조이다.
----	----------------------------------

- 구조 설명 문장의 서술부는 '~구조이다', '~(으)로 되어 있다', '~게 되어 있다' 혹은 구조를 나타내는 다양한 서술어를 이용하여 작성한다.



굴피집

구조	이 굴피집의 지붕은 양쪽으로 경사가 지도록 덮은 맞배지붕 구조이다.
----	---------------------------------------



무주 귀틀집

색채	이 귀틀집은 전체적으로 황갈색을 띤다.
----	-----------------------

- 색채 설명 문장의 서술부에는 다양한 색채 명사 및 색채 형용사가 포함되도록 작성한다.



대청


색채	이 대청은 전체적으로 짙은 고동색과 연한 고동색을 띠며, 기둥은 적갈색이다.
----	--

- 색채 설명 문장 틀

색채	이 [대상]은/는 [색채 명사+이다/색채 형용사]
	이 [대상]의 색채는 [색채 명사+이다/색채 형용사]
	이 [대상] [구성 요소]의 색채는 [색채 명사+이다/색채 형용사]

4.4.2.1.6. 문양

- 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 장식하기 위해 사용된 표면의 문양 혹은 무늬를 기술한다. 사진 속 대상에서 전체적으로 나타나는 문양 및 무늬를 설명할 수도 있고 구성 요소에서 나타나는 문양 및 무늬를 설명할 수도 있다.
- 문양 설명 문장의 주제부는 '지시 대명사+대상' 혹은 '지시 대명사+대상의 문양의 위치' 등으로 한다.

	
소로	
문양	이 소로 에는 연꽃과 구름 문양 등의 단청이 그려져 있다.

- 문양 설명 문장의 서술부는 '~문양이/무늬가 있다', '~이/가 그려져 있다', '~이/가 새겨져 있다' 등으로 작성한다.



거창 윤경남 생가 안채

배치	이 안채는 부엌 옆에 배치되어 있다.
----	----------------------

- 배치 설명 문장의 서술부는 '~에 배치되어 있다', '~게 배치되어 있다', '~(으)로 배치되어 있다' 등으로 작성한다.



연립주택

배치	이 연립주택은 다른 건물들과 나란하게 배치되어 있다.
----	-------------------------------



안채

용도	<u>안채</u> 는 여성들이 일상생활, 가사일 등을 하는 공간이다.
----	--

4.4.2.2. 소재지(所在地)

- 사진 속 건축물 및 건축물 일부가 자리 잡고 있는 곳을 기술한다. 갈암 종택, 만성 고택 등과 같은 고유 건축물의 경우 건축물의 소재지를 설명할 수 있다.
- 소재지 설명 문장의 주제부는 '대상'으로 한다.



단계 박씨 고가

소재지(所在地)	<u>단계 박씨 고가</u> 는 경상남도 산청군 신등면에 있다.
----------	-------------------------------------

4.4.2.2.3. 분포

- 사진 속 건축물이 주로 나타나는 지역을 기술한다. 특정 지역에서 주로 관찰되는 건축물의 경우 분포를 설명할 수 있다.
- 분포 설명 문장의 주제부는 '대상'으로 한다.
 - ✓ 단, 대상에 특정 지역의 이름이 포함되어 있는 경우, 이를 제외하고 분포 설명 문장을 작성한다.

	
<p>무주 귀틀집</p>	
<p>분포</p>	<p>귀틀집은 산림이 많은 중부 이북의 산간 지역에 많이 분포한다.</p>

4.4.2.2.4. 이용자

- 사진 속 건축물 및 건축물 일부, 주생활 관련 도구를 주로 이용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 이용자가 한정적인 경우 이용자를 설명 문장으로 작성할 수 있다.
- 이용자 설명 문장의 주제부는 '대상'으로 한다.



안채

이용자	안채는 주로 여성들이 사용하는 중심 건물이다.
-----	---------------------------

4.4.2.2.5. 소유자

- 사진 속 건축물을 소유했던, 소유하고 있는 사람을 기술한다. 갈암 종택, 추산 고택 등 특정인이 거주했던 건축물의 경우 소유자를 기술할 수 있다.²⁾



영덕 갈암종택 전경

소유자	영덕 갈암종택은 갈암 이현일 종가가 대대로 거주했던 집이다.
-----	-----------------------------------

2) 지역 이름이 포함되어 있는 사진의 설명 문장 작성 시 주의 사항은 지침 96쪽을 참조하기 바람.

4.4.2.2.6. 역사

- 사진 속 건축물 및 건축물 일부와 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.
- 역사 설명 문장의 주제부는 ‘대상’으로 한다.

	
<p>달성 삼가헌 고택</p>	
<p>역사</p>	<p>달성 삼가헌 고택은 조선 중기에 건축된 지방 양반가의 특징을 간직한 주택으로, 1757년(영조 23년)에 건축되었다.</p>

4.5. 건축 설명 문장

4.5.1. 건축 설명 속성

가시적 속성	비가시적 속성
형태 구성 요소 구조(외부/내부) 자재/소재(素材) 색채 문양 배치(집합 건축물을 이루는 단일 건축물들의 배치)	용도 소재지(所在地) 이용자 소유자 건축가 건축 양식 역사

- 예술적, 문화적, 역사적 가치가 있는 비거주 집합 건축물(경복궁, 덕수궁, 불국사 등), 단일 건축물(경암루, 부여 정림사지 오층석탑 등) 및 건축물 일부(경산 환성사 대웅전 수미단, 경주 서악동 귀부 등) 등의 유형에 따라 위 표의 설명 속성 중 적절한 속성을 선택하여 기술한다.

- 형태, 구성 요소, 구조, 자재, 색채, 문양 등의 사진 대상 설명 속성은 사진 대상의 전체와 부분을 설명할 때 모두 활용할 수 있다.
- 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 작성할 수 있다.
- 가시적 속성
 - 형태: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 모양
 - 구성 요소: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 이루는 데 필요한 요소
 - 구조(외부/내부): 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 구성 요소들을 엮은 체계
 - 자재/소재(素材): 사진 속 집합 건축물, 단일 건축물 및 건축물 일부에 쓰인 재료
 - 색채: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부에서 두드러지게 나타나는 빛깔
 - 문양: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 장식하기 위한 여러 가지 모양
 - 배치(집합 건축물을 이루는 단일 건축물들의 배치): 사진 속 집합 건축물을 이루는 단일 건축물들이 벌여 놓여 있는 모양
- 비가시적 속성
 - 용도: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 쓰임새
 - 소재지(所在地): 사진 속 집합 건축물, 단일 건축물 및 건축물 일부가 자리 잡고 있는 곳
 - 이용자: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 주로 이용하는 사람(부류)
 - 소유자: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 소유한 사람
 - 건축가: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 설계하거나 지은 사람
 - 건축 양식: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부가 지닌 독특한 형식
 - 역사: 사진 속 집합 건축물, 단일 건축물 및 건축물 일부와 관련된 역사적 사실

4.5.2. 건축 속성 기술 문장 작성 방법

4.5.2.1. 가시적 속성 기술 문장 작성 방법³⁾

4.5.2.1.1. 형태

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부의 모양을 기술한다. 전체적인 모양을 설명할 수도 있고 구성 요소의 모양을 설명할 수도 있다.

3) '건축' 범주에 속하는 한국어문화 요소는 '경복궁, 불국사, 석파정' 등 고유한 이름이 붙은 건축물들이므로, 가시적 속성 기술 문장을 작성할 때 지시 대명사를 사용하지 않고, 대상의 명칭만으로 주제부를 작성한다. 또한 '건축' 범주에 속하는 '사찰, 탑, 석등' 등의 건축물들은 '갈계리 삼층 석탑', '방내리 삼층 석탑' 등과 같이 '지명'으로 대상을 구별 지어 부르므로, 명칭에 '갈계리, 방내리' 등의 지명이 포함되어 있는 경우에 지명을 포함하여 주제부를 작성한다.

단, 일반 명사의 경우 기존대로 지시 대명사 '이'를 포함하여 주제부를 작성한다.



구례 화엄사 각황전 앞 석등

구조	구례 화엄사 각황전 앞 석등은 장구 모양으로 된 기단부의 간석이 등불을 밝히는 부분인 화사석을 받치고 있는 구조이다.
----	---

- 구조 설명 문장의 서술부는 ‘~구조이다’, ‘~(으)로 되어 있다’, ‘~게 되어 있다’ 혹은 구조를 나타내는 다양한 서술어를 이용하여 작성한다.



구리 동구릉 송릉 정자각

구조	구리 동구릉 송릉 정자각은 기둥의 중심선에서 한 자리 나았아서 도리를 받친 일출목 구조이다.
----	---



서울 경희궁 흥화문

<p>색채</p>	<p>서울 경희궁 흥화문의 기둥은 적갈색이고, 지붕은 기둥의 색깔과 대비를 이루는 녹색이다.</p>
-----------	---

- 색채 설명 문장의 서술부에는 다양한 색채 명사 및 색채 형용사가 포함되도록 작성한다.



경포대

<p>색채</p>	<p>경포대의 벽면은 녹색과 노란색이고 기둥은 적갈색이며, 지붕의 단청은 녹색, 붉은색, 파란색, 하얀색 등으로 화려하다.</p>
-----------	--

- 색채 설명 문장 틀

색채	[대상]은/는 [색채 명사+이다/색채 형용사]
	[대상]의 색채는 [색채 명사+이다/색채 형용사]
	[대상] [구성 요소]의 색채는 [색채 명사+이다/색채 형용사]

4.5.2.1.6. 문양

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 장식하기 위해 사용된 표면의 문양 혹은 무늬를 기술한다. 사진 속 대상에서 전체적으로 나타나는 문양 및 무늬를 설명할 수도 있고 구성 요소에서 나타나는 문양 및 무늬를 설명할 수도 있다.
- 문양 설명 문장의 주제부는 ‘대상’ 혹은 ‘대상의 문양의 위치’ 등으로 한다.

	
여주 고달사지 승탑	
문양	여주 고달사지 승탑의 중대석에는 거북 문양이 새겨져 있는데 거북의 양옆에는 용과 구름이 새겨져 있다.

- 문양 설명 문장의 서술부는 ‘~문양이/무늬가 있다’, ‘~이/가 그려져 있다’, ‘~이/가 새겨져 있다’ 등으로 작성한다.



사의당

배치	사의당의 건물들은 ㄷ자로 배치되어 있다.
----	------------------------

✓ 단, 집합 건축물을 이루는 단일 건축물의 명칭이 별도로 있는 경우에는 이를 활용하여 설명 문장을 작성할 수 있다.



덕수궁

배치	덕수궁의 중화문과 중화전은 마당을 사이에 두고 마주 보도록 배치되어 있다.
----	---

- 배치 설명 문장의 서술부는 '~게/도록 배치되어 있다', '~(으)로 배치되어 있다' 등으로 작성한다.



경주 석빙고

<p>용도</p>	<p>경주 석빙고는 얼음이 녹지 않도록 보관하기 위해 만든 얼음 창고이다.</p>
-----------	---

4.5.2.2.2. 소재지(所在地)

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부가 자리 잡고 있는 곳을 기술한다. ‘충청북도 보은군 속리산면’ 등과 같이 소재가 명확한 경우에 소재지를 설명 문장으로 작성할 수 있다.
- 소재지 설명 문장의 주제부는 ‘대상’으로 한다.



종원 미륵리 삼층석탑

<p>소재지(所在地)</p>	<p>종원 미륵리 삼층석탑은 충청북도 충주시 상모면에 있다.</p>
-----------------	---------------------------------------

4.5.2.2.3. 이용자

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 주로 이용하는 사람(부류)을 기술한다. 남성, 여성, 아이 등 이용자가 한정적인 경우 이용자를 설명 문장으로 작성할 수 있다.
- 이용자 설명 문장의 주제부는 '대상'으로 한다.



강릉 대도호부 관아

<p>이용자</p>	<p>강릉 대도호부 관아는 고려 시대 중앙의 관리들이 강릉에서 머무를 때 주로 이용했던 객사이다.</p>
------------	--

4.5.2.2.4. 소유자

- 사진 속 건축물을 소유했던, 소유하고 있는 사람을 기술한다. 석파정, 팔락정 등 특정인이 거주하거나 이용했던 건축물의 경우 소유자 설명 문장을 작성할 수 있다.
- 소유자 설명 문장의 주제부는 '대상'으로 한다.



서울 석파정 별당

<p>소유자</p>	<p>서울 석파정 별당은 흥선대원군 이하응이 소유했던 별장의 사랑채이다.</p>
------------	--

4.5.2.2.5. 건축가

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부를 설계하거나 지은 사람을 기술한다. 나상기, 박길룡 등 건축가의 구체적인 명칭이 확인되는 경우 건축가 설명 문장을 작성할 수 있다.
- 건축가 설명 문장의 주제부는 '대상'으로 한다.

	
<p>서울 경교장</p>	
<p>건축가</p>	<p>서울 경교장은 지하 1층, 지상 2층의 건물로, 근대 건축가인 김세연이 설계하였다.</p>

4.5.2.2.6. 건축 양식

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부가 지어진 시대나 종류에 따라 독특하게 지니는 건축의 형식을 기술한다. 사진 속 건축물이 특정 양식으로 지어진 경우 건축 양식 설명 문장을 작성할 수 있다.
- 건축 양식 설명 문장의 주제부는 '대상'으로 한다.



서울 고려대학교 중앙도서관

건축 양식	서울 고려대학교 중앙도서관은 화강암을 주자재로 하여 서양 중세풍의 고딕양식으로 지어졌다.
-------	---

4.5.2.2.7. 역사

- 사진 속 집합 건축물, 단일 건축물 및 건축물 일부와 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사 설명 문장을 작성할 수 있다.
- 역사 설명 문장의 주제부는 '대상'으로 한다.



수원 화성 서북공심돈

역사	수원 화성 서북공심돈은 성역의 안팎에서 일어나는 일을 감시하던 망루로, 1796년(정조 20년) 3월 10일에 완공되었다.
----	--

4.6. 공예 설명 문장

4.6.1. 공예 설명 속성

가시적 속성	비가시적 속성
구성 요소 크기 길이 소재(素材) 형태 문양 색채	용도 기법 소재지(所在地) 역사


- 도자기, 금속, 나무, 종이, 유리, 칠기 등의 재료를 사용해 예술적 가치가 있게 만든 공예품의 종류에 따라 위 표의 설명 속성 중 적절한 속성을 선택하여 기술한다.
- 크기, 길이, 소재, 형태, 문양, 색채 등의 설명 속성은 사진 속 대상의 전체와 부분을 설명할 때 모두 활용할 수 있다.
- 가시적 속성으로 분류된 속성이 사진에서 식별되지 않는 경우, 해당 속성을 비가시적 속성으로 작성할 수 있다.
- 가시적 속성
 - 구성 요소: 사진 속 공예품을 이루는 데 필요한 요소
 - 크기: 사진 속 공예품의 넓이, 부피 등이 크고 작은 정도
 - 길이: 사진 속 공예품의 한끝에서 다른 한끝까지의 거리
 - 소재(素材): 사진 속 공예품을 만드는 데에 바탕이 되는 재료
 - 형태: 사진 속 공예품의 모양
 - 문양: 사진 속 공예품 표면의 무늬
 - 색채: 사진 속 공예품의 빛깔
- 비가시적 속성
 - 용도: 사진 속 공예품의 쓰임새
 - 기법: 사진 속 공예품을 만들거나 공예품의 문양을 넣는 데에 활용한 기교와 방법
 - 소재지(所在地): 사진 속 공예품이 위치하고 있는 곳, 혹은 해당 공예품을 소장하고 있는 기관
 - 역사: 사진 속 공예품과 관련된 역사적 사실

4.6.2. 공예 속성 기술 문장 작성 방법

4.6.2.1. 가시적 속성 기술 문장 작성 방법⁴⁾

4.6.2.1.1. 구성 요소

- 사진 속 공예품을 이루는 데 필요한 요소들을 기술하되, 사진에서 가시적으로 드러나는 요소 위주로 작성한다.
- 구성 요소 설명 문장의 주제부는 ‘대상’으로 한다.

	
<p>청자 상감 앵무문 표주박 모양 주전자</p>	
<p>구성 요소</p>	<p>청자 상감 앵무문 표주박 모양 주전자는 표주박 모양의 몸통과 마개, 주구, 손잡이 등으로 이루어져 있다.</p>

- 구성 요소 설명 문장의 서술부는 ‘~(으)로 이루어져 있다’, ‘~(으)로 구성되어 있다’ 등으로 작성한다.

4) ‘공예’ 범주에 속하는 한국언어문화 요소 역시 ‘건축’ 범주와 마찬가지로 고유한 이름이 붙은 경우가 많으므로, 가시적 속성 기술 문장을 작성할 때 지시 대명사를 사용하지 않고, 대상의 명칭만으로 주제부를 작성한다. 또한 공예품의 명칭에 ‘밀양, 양산’ 등의 지명이 포함되어 있는 경우, 지명을 포함하여 주제부를 작성한다.

단, 일반 명사의 경우 기존대로 지시 대명사 ‘이’를 포함하여 주제부를 작성한다.



밀양 표충사 청동 은입사 향완

크기	밀양 표충사 청동 은입사 향완은 받침의 넓이가 몸체의 넓이보다 넓다.
----	--

- 크기 설명 문장의 서술부는 '크다', '작다', '넓다', '좁다', '깊다', '얕다' 등의 다양한 형용사를 이용하여 작성한다.



백자 청화 모란문 호

크기	백자 청화 모란문 호는 입구가 바닥 부분보다 좁으며, 하부에 비해 상부가 상대적으로 폭이 넓다.
----	---

- 크기 설명 문장 틀

크기	[대상]은/는 []
	[대상]의 [구성 요소]은/는 []
	[대상] [구성 요소]의 크기/넓이/폭은 []

4.6.2.1.3. 길이

- 사진 속 공예품의 한끝에서 다른 한끝까지의 거리를 기술한다. 해당 공예품 부류의 일반적인 길이와 비교하거나 구성 요소들의 길이를 비교, 대조하는 방식으로 길이 설명 문장을 작성할 수 있다.
- 길이 설명 문장의 주제부는 ‘대상(대상의 구성 요소)’ 혹은 ‘대상(대상의 구성 요소)의 길이’ 등으로 한다.

	
임경업 추련도	
길이	임경업 추련도는 칼날 길이에 비해 손잡이 길이가 상대적으로 매우 짧게 만들어졌다.

- 길이 설명 문장의 서술부에는 ‘길다’, ‘짧다’ 등의 다양한 형용사가 포함되도록 작성한다.



금동향로

길이	금동향로의 다리는 일반적인 향로의 다리보다 길이가 긴 편이다 .
----	--

- 길이 설명 문장 틀

길이	[대상]은/는 []
	[대상]의 [구성 요소]은/는 []
	[대상] [구성 요소]의 길이는 []

4.6.2.1.4. 소재(素材)

- 사진 속 공예품을 만드는 데에 바탕이 되는 재료를 기술하되, 사진에서 해당 공예품의 소재가 식별되는 경우에 설명 문장을 작성한다. 소재의 질감이 식별이 되는 경우, 소재와 함께 소재의 질감도 함께 기술할 수 있다. 공예품의 전체적인 소재를 설명할 수도 있고 구성 요소의 소재를 설명할 수도 있다.
- 소재 설명 문장의 주제부는 ‘대상(대상의 구성 요소)’ 혹은 ‘대상(대상의 구성 요소)의 소재’ 등으로 한다.



목조 문각 과등

소재(素材)	목조 문각 과등은 주재료로 나무와 유리가 쓰였다.
--------	-----------------------------

- 소재 설명 문장의 서술부는 ‘~(으)로 만들어졌다’, ‘~이/가 쓰였다’; ‘~등이다’ 등으로 작성한다.



자정국사 사리함

소재(素材)	자정국사 사리함은 매끄러운 질감의 도자기로 만들어졌다.
--------	--------------------------------



화성 용주사 청동 향로

문양	화성 용주사 청동 향로에는 산촌의 풍경이 양각으로 정교하게 새겨져 있다.
----	--

- 문양 설명 문장의 서술부는 '~문양이/무늬가 있다', '~이/가 그려져 있다', '~이/가 새겨져 있다' 등으로 작성한다.



청자철회절지문조충문병

문양	청자철회절지문조충문병에는 날고 있는 새와 덩굴 식물이 그려져 있다.
----	---------------------------------------



- 색채 설명 문장 틀

색채	[대상]은/는 [색채 명사+이다/색채 형용사]
	[대상]의 색채는 [색채 명사+이다/색채 형용사]
	[대상] [구성 요소]의 색채는 [색채 명사+이다/색채 형용사]

4.6.2.2. 비가시적 속성 기술 문장 작성 방법

4.6.2.2.1. 용도

- 사진 속 공예품의 쓰임새를 기술한다. 해당 공예품의 특징적인 쓰임새가 있는 경우 용도를 설명 문장으로 작성할 수 있다.
 - ✓ 그릇, 연적 등과 같이 실용적인 목적에서 만들어진 공예품이나 외형만으로는 용도를 알 수 없는 공예품의 경우 되도록 용도 설명 문장을 작성하도록 한다.
- 용도 설명 문장의 주제부는 '대상'으로 한다.



청동 금강저

용도	청동 금강저는 번뇌를 깨뜨리는 보리심을 상징하는 법구로, 승려가 불도를 닦을 때 사용했다.
----	--

4.6.2.2.2. 기법

- 사진 속 공예품을 만들거나 공예품에 문양을 넣는 데에 활용한 기교와 방법을 기술한다. 해당 공예품이 특정 기법을 사용해 만들어진 경우 기법을 설명 문장으로 작성할 수 있다.
 - ✓ 기법은 '상감', '은입사' 등과 같은 특정 기법을 설명하는 것으로, 해당 공예품을 만드는 일반적인 방법을 작성하지 않도록 주의한다.
- 기법 설명 문장의 주제부는 '대상'으로 한다.



청자철회절지문조충문병

기법	청자철회절지문조충문병은 일반 청자에서 주로 사용되는 환원번조법이 아닌 산화번조법을 사용해 문양을 표현했다.
----	---

4.6.2.2.3. 소재지(所在地)

- 사진 속 공예품이 위치하고 있는 장소, 혹은 해당 공예품을 소장하고 있는 기관을 기술한다. 동종, 범종 등의 경우에는 해당 공예품이 위치하고 있는 장소, 즉 소재지를 기술하고, 백자, 사리함, 연적 등의 경우에는 해당 공예품을 소장하고 있는 기관을 기술한다.
- 소재지 설명 문장의 주제부는 '대상'으로 한다.



청자 모자 원숭이 모양 연적

소재지(所在地)	청자 모자 원숭이 모양 연적은 서울특별시 성북구 간송미술관에서 소장하고 있다.
----------	---

4.6.2.2.4. 역사

- 사진 속 공예품과 관련된 역사적 사실을 기술한다. 해당 한국어문화 요소와 관련한 특징적인 역사적 사실이 있는 경우 역사를 설명할 수 있다.
- 역사 설명 문장의 주제부는 '대상'으로 한다.



양산 내원사 금고

역사	양산 내원사 금고는 11세기에 동량승인 정묘가 금인사에 시납한 것으로 추정된다.
----	--

5. 주의 사항

- 백과사전적 정보는 이미 많이 축적되어 있고, 국립국어원 온용어 등의 사업에서 구축하고 있으므로 되도록 백과사전적 정보는 배제하고 최대한 사진의 대상을 묘사하는 가시적 속성 기술 문장을 작성한다.
- ‘담양 고재선 가옥’, ‘화순 양참사택’ 등과 같은 고유 건축물의 경우, 다음과 같이 설명 문장을 작성한다.
 - 사진 제목에 집합 건축물에서의 특정 건축물(사랑채, 안채 등)을 지시하는 단어가 포함되어 있지 않은 경우에는 ‘담양 고재선 가옥’, ‘화순 양참사택’ 등의 명칭 그대로 설명 문장을 작성한다.



합천 팔심리 윤씨 고가

역사	<u>합천 팔심리 윤씨 고가</u> 는 1995년에 민속문화유산으로 지정되었다.
----	--

- 사진 제목에 건축물 일부(사랑채, 안채 등)를 지칭하는 단어가 포함되어 있고, 사진 또한 건축물의 일부만이 (단일 건축물에서의 일부, 집합 건축물에서의 단일 건축물) 드러나는 경우에는 고유명을 제외하고 ‘안채’, ‘사랑채’ 등의 명칭을 사용하여 설명 문장을 작성한다.



화순 양동호 가옥 안채

형태	이 안채의 지붕은 ㅉ자 형태이다.
----	--------------------

- 사진 제목에 '정면', '정측면', '내부', '전경' 등과 같은 정보가 포함되어 있는 경우, 이를 제외하고 한국어문화 요소인 대상의 명칭만을 사용하여 설명 문장을 작성한다.
- 대상의 명칭 앞에 '용인 한국민속촌 중 한옥의 부엌 뒤편에 소재한' 등과 같은 정보가 포함되어 있는 경우에도 이를 제외하고 대상의 명칭만을 사용하여 설명 문장을 작성한다.



강릉 만성 고택 내부 전경

대상 식별 문장	이 사진은 한국 전통 가옥인 강릉 만성 고택의 사진이다. (O)
	이 사진은 한국 전통 가옥인 강릉 만성 고택 내부 전경의 사진이다. (X)

○ 사진 제목에 '강릉', '서울' 등의 소재지 정보가 포함되어 있는 경우, 다음과 같이 작성한다.

- 특정 개체를 지시하기 위해 사용된 경우에는 '강릉', '서울' 등의 소재지를 포함하여 설명 문장을 작성한다.



영덕 갈암 종택 전경

소유자	영덕 갈암 종택은 갈암 이현일 종가가 대대로 사용했던 집이다.
-----	------------------------------------

- ✓ 단, 한국어문화 요소의 명칭이 지나치게 길어서 설명 문장의 명료성을 해치는 경우, 대상 식별 문장에만 '강릉', '서울' 등의 소재지를 포함하여 작성하고 속성 기술 문장에는 이를 제외하고 설명 문장을 작성할 수 있다.



합천 팔심리 윤씨 고가

대상 식별 문장	이 사진은 한국 전통 가옥인 <u>합천 팔심리 윤씨 고가</u> 의 사진이다.	
속성	색채	<u>이 윤씨 고가</u> 의 지붕은 남색이고 벽은 황토색이다.

- 일반적인 문화 요소의 소재지를 표기하기 위해 사용된 경우에는 '강릉', '서울' 등의 소재지를 제외하고 한국

언어문화 요소의 명칭만을 사용하여 설명 문장을 작성한다.



무주 귀틀집

색채	이 귀틀집은 전체적으로 황갈색을 띤다.
----	-----------------------

○ 사진 제목이 한국에서 두루 사용되는 일반적인 명칭이 아닌 경우 일반적인 명칭으로 바꾸어 설명 문장을 작성한다.



갯쌈김치(갯김치)

지역	갯김치는 갯의 주요 재배지인 여수에서 발달했다.
----	----------------------------

○ 사진 제목이 특정 대상의 고유한 명칭이어서 한국어문화 요소의 특성을 기술하는 것의 효용이 떨어지는 경우, 일반적인 명칭으로 설명 문장을 작성한다.



영락봉잠

<p>용도</p>	<p>비녀는 과거 여성들이 쪽을 짚은 머리를 풀어지지 않게 하기 위하여 꽂던 장신구이다.</p>
-----------	---



거창 윤경남 생가 안채

<p>배치</p>	<p>이 안채는 지면보다 높게 배치되어 있다.</p>
-----------	-------------------------------

- ✓ 단, 문화유산이나 보물로 지정된 한국어문화 요소의 경우, 고유명을 그대로 살려 설명 문장을 작성한다.



봉화 만회고택

역사	봉화 만회고택 은 1984년에 국가민속문화유산으로 지정되었다.
----	---

- 설명 문장은 어문 규범을 준수하여 작성한다.
- 전문 용어 및 고유명은 경우에 따라 붙여쓰기와 띄어쓰기를 모두 허용하는 경우가 있으나, 원칙과 허용 중 원칙에 따라 띄어쓰기한다.

<기획·연구>

국립국어원 이현주 언어정보과장

국립국어원 박미영 학예연구관

국립국어원 주향아 학예연구사

국립국어원 김연수 연구원

국립국어원 장연지 연구원

<연구 참여자>

연구 책임자 차정원(데이터리)

공동 연구원 신서인(한림대학교 산학협력단)

연구 보조원 오진영, 최윤진, 이민희, 서민지, 최수찬, 조지원,

박송원, 박수연, 유승빈, 전광훈, 김민혜, 정상미 (데이터리)

이한범, 이수현, 박진현, 전상호, 김나현,

이준희, 김푸른솔, 정민교, 강주은 (한림대학교 산학협력단)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 25년 12월 19일

발행일: 25년 12월 19일

인 쇄: 서경기획

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘2025년 한국언어문화 표·사진 설명 말뭉치 구축 연구’ 사업의 결과물을 발간한 것입니다.