

국립국어원 2025-01-45

발 간 등 록 번 호
11-1371028-100041-01

2025년

지시문 기반 침삭 지원 말뭉치 연구 분석

연구 책임자
곽 용 진



국립국어원

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 '2025년 지시문 기반 첨삭 지원
말뭉치 연구 분석'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2025년 5월 ~ 2025년 12월

2025년 12월 18일

연구책임자: 곽용진((주)이르테크)

연구 기관: (주)이르테크 컨소시엄

연구책임자: 곽용진

공동연구원: 김한샘, 함영균, 이지연,
임경태, 오누리, 안의정,
이진, 봉미경, 김선희,
윤영민

연구보조원: 천성호, 진예솔, 이소미,
이운복, 정해영, 홍은기,
박우정, 박정혁, 박하영,
이현숙, 김윤희, 오수연,
하승목, 이담허, 윤찬혁,
안소현, 이승빈, 김미영,
박현이, 이정은, 신현규,
정은수, 임수용, 주민재,
안혜지, 강예지, 이이슬

<국문 요약>

2025년 지시문 기반 첨삭 지원 말뭉치 연구 분석

이 연구는 인공지능을 이용해 국어생활 상담, 글쓰기 첨삭 등 한국어에 특화된 서비스를 구현하기 위한 말뭉치 구축 방법론을 연구하고 구축 지침을 마련해 국어생활 상담 질문 유형별 세부 질문-답변쌍 말뭉치와 국어 지식 기초 자료를 구축하고, 2024년에 구축된 지시문 기반 글쓰기 첨삭 지원 말뭉치와 기초 자료를 보완하는 것이 목적이다.

이번에 구축한 지시문 기반 첨삭 지원 말뭉치는 기존에 구축된 한국어 말뭉치와 여러 면에서 다른 구축 방법론을 통해 구축됐다. 국립국어원이 보유한 국어생활 상담 자료를 분석해 832종의 질문 유형을 도출함으로써 상담 사례를 모사하고 각 질문 유형별로 관련 국어 지식 및 어문 규범과 관련한 수십 건의 질문-답변을 AI를 활용해 생성한 뒤, 이를 국어 전문가가 실제 상담 답변에 근접하도록 다듬고 검수했다. 이러한 구축 방식은 기존에 정답 데이터(labeled data)를 통해 학습 효율을 높이거나 초대규모 데이터와 거대 모델로 지식을 확장하는 방식과 달리 현실 데이터를 분석해 인공지능이 모방하도록 하는 과정에서 얻어지는 현실 데이터 특성을 강화해 고품질 고효율의 말뭉치를 만들어 내는 발견적 말뭉치 구축 방법이라고 할 수 있다. 실제로 지시문 기반 첨삭 말뭉치를 구축하는 과정은 인공지능이 국어생활 상담 질문-답변을 더 잘 생성하도록 하는 과정이면서 동시에 인공지능이 생성한 답변의 부족한 부분을 찾아 보완하는 것을 반복하는 과정이었다.

구축된 말뭉치가 AI 기반 국어생활 상담·첨삭 서비스 구현에 활용될 수 있도록 4종의 공개 거대언어모델(LLM) 4종에 대해 훈련을 진행하고 국어 상담 능력이 형성되는지에 중점을 두고 검증했다. 그 결과 구축된 말뭉치로 학습한 4종의 LLM 모두 국어 지식 관련 질문-답변에 대해 학습 전과 대비해 평균 약 63.6%의 응답 성능 향상을 보였다. 또한, 표준발음법 등의 어문 규정과 사례 84,134건을 국어 지식 기초 자료로 구축하고 2024년에 구축된 지시문 기반 글쓰기 첨삭 지원 말뭉치와 기초 자료도 보완했다.

이러한 성과에도 불구하고 빠르게 발전하는 글로벌 빅테크 기업의 인공지능 관련 기술은 이제 언어의 장벽을 넘어 현실 세계, 가상 세계, 일상의 모든 영역으로 확장하며 새로운 기회와 함께 인공지능 종속의 위협을 가속하고 있다. 한국은 AI 반도체, 독자 파운데이션 모델, 디지털 인프라 기반 인공지능 혁신 등 사회 전방위에 걸친 인공지능 기술 경쟁에 돌입했다. 이번 연구가 초기 인공지능 연구부터 지속해 온 한국어 말뭉치 구축 경험과 역량을 새로운 고품질 고효율 한국어 언어자원으로 전환하는 시금석이 될 것으로 기대한다.

주요어: 한국어 어문 규범 인공지능, 국어생활 상담 자료, 국어 지식 기초 자료

<Abstract>

2025 Research on the Construction of Instruction Corpus for writing editing support(Orthography and Norms)

This study explored a corpus construction methodology for implementing Korean-specific services, such as Korean language counseling and writing editing, using AI. The methodology developed included detailed question-answer pair corpora and Korean language knowledge data for each type of Korean language counseling question.

These instruction-based editing support corpora were constructed using a methodology that differed significantly from existing Korean corpora. By analyzing Korean language counseling data held by the National Institute of the Korean Language, 832 question types were identified. These corpora were then modeled after consultation cases. For each question type, dozens of question-answer pairs related to relevant Korean language knowledge and grammar norms were generated using AI. These were then refined and reviewed by Korean language experts to approximate real-world counseling responses. Unlike existing methods that enhance learning efficiency through labeled data or expand knowledge through ultra-large data sets and large models, this heuristic corpus construction method leverages the real-world data analysis process to generate a high-quality, highly efficient corpus. In fact, the process of building a directive-based editing corpus involved repeatedly identifying and supplementing deficiencies in the AI-generated answers while simultaneously enabling AI to better generate Korean language counseling questions and answers.

To ensure that the corpus could be utilized in the implementation of AI-based Korean language counseling and editing services, four publicly available large-scale language models (LLMs) were trained and tested, with a focus on developing Korean language counseling skills. As a result, all four LLMs trained with the corpus showed an average improvement of approximately 63.6% in response performance for Korean language knowledge-related questions and answers compared to pre-training. And we've also constructed 84,134 of Korean knowledge(ex. Standard Pronunciation)'s rules and examples as a basic data

Despite these achievements, the rapidly evolving AI technologies of global big tech companies are now transcending language barriers and expanding into the real world, virtual worlds, and all areas of daily life, accelerating the threat of AI dependence while creating new opportunities. Korea has entered a race for AI technology across all aspects of society, including AI semiconductors, proprietary foundation models, and digital infrastructure-based AI innovation. This study is expected to serve as a touchstone for transforming the experience and capabilities gained from early AI research into a Korean corpus-building resource, building a new, high-quality, and efficient Korean language resource.

Keywords: Korean grammar normative AI, Korean language counseling materials

차 례

제1장 사업 개요

1.1. 사업의 목적과 범위	2
1.2. 사업 추진 방법	4
1.3. 사업 추진 일정	8

제2장 지시문 기반 첨삭 지원 말뭉치

2.1. 지시문 기반 첨삭 지원 말뭉치 설계	10
2.2. 지시문 기반 첨삭 지원 말뭉치 구축	35
2.3. 말뭉치 검증	47

제3장 국어 지식 기초 자료

3.1. 국어 지식 기초 자료 개요	61
3.2. 국어 지식 기초 자료 유형별 구축 사례	62
3.3. 국어 지식 기초 자료 데이터 구조	65

제4장 2024년 구축 결과 개선 및 보완

4.1. 글쓰기 첨삭 지원을 위한 기초 자료 개선	67
4.2. 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치 개선	71

제5장 결론

5.1. 사업 결과 요약	74
5.2. 발전 제언	75

[부록1] 지시문 기반 첨삭 지원 말뭉치 구축 지침	83
------------------------------------	----

차 례

표 차례

<표 1> 사업의 범위와 세부 과업 내용	3
<표 2> 예측 위험 요소 및 대응 방안	6
<표 3> 국어생활 상담 유형 분류 자료 예시	13
<표 4> 국어생활 상담 유형 분류 결과	14
<표 5> 상담 자료 유형(유형 분류)별 상담 사례 하위 분류 프롬프트 예시	15
<표 6> 언어모델 출력 결과를 활용한 세부 질문 유형 재분류 결과	16
<표 7> 질문 유형 범주별 세부 질문 유형과 예시 질문-답변쌍의 수	16
<표 8> 지시문 기반 탐색 지원 말뭉치의 기본 구성 요소	17
<표 9> 지시문 기반 탐색 지원 말뭉치 데이터 구조 (json)	18
<표 10> 설계 말뭉치 작성 단계별 작업 내용	22
<표 11> 생성 질의-답변 평가 지표	22
<표 12> 말뭉치 설계 자료 예시	23
<표 13> 설계 말뭉치 질문 유형별 작성 양식	24
<표 14> 이용자 질문 유형	26
<표 15> 국어 지식 작성 예시	27
<표 16> 국어 지식 작성 형식	28
<표 17> 모범 답변의 형식과 구성 요소	29
<표 18> 꼬리 질문 예시	31
<표 19> 꼬리 질문 유형	31
<표 20> 초기 답변 생성용 프롬프트	32
<표 21> 모범 답변 구성 및 작성 방식	31
<표 22> 모범 답변 표현 예시	32
<표 23> 모범 답변 작성 규칙을 적용한 답변 생성용 프롬프트	34
<표 24> 초기 질문-답변 생성 프롬프트	37
<표 25> 1차 질문-답변 증강 생성 프롬프트	38
<표 26> 1차 개선 프롬프트 모델별 생성 답변 부적합률	39
<표 27> 질문 생성용 프롬프트	40
<표 28> 답변 생성용 프롬프트	41
<표 29> 2차 개선 프롬프트 모델별 생성 답변 부적합률	42
<표 30> 작업 도구 주요 기능	44
<표 31> 말뭉치 검증을 위한 AI모델 베이스라인 테스트	48
<표 32> 평가 특화 프롬프트 전문	51

차 례

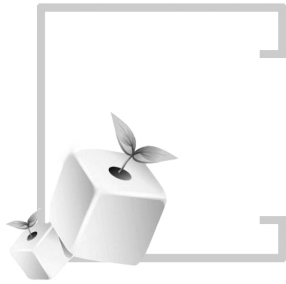
표 차례

<표 33> 지시문 기반 첨삭 지원 말뭉치 학습데이터 변환 예시	52
<표 34> AI 모델별 성능 평가 결과	56
<표 35> 모델별 학습 후 출력 사례1	56
<표 36> 모델별 학습 후 출력 사례2	56
<표 37> 국어 지식 기초 자료 데이터 구성 항목	61
<표 38> 표준 발음법 국어 지식 기초 자료 구축 사례	63
<표 39> 로마자 표기법 국어 지식 기초 자료 구축 사례	64
<표 40> 국어 지식 기초 자료 데이터 구조 (json)	65
<표 41> 기초 자료 구축 지침중 [부록]오류 유형 및 관련 규정의 세부 규정 작성 기준	68
<표 42> ‘표현-문단’의 추가 오류 유형	71
<표 43> 지시문 기반 첨삭 지원 말뭉치 구축 결과	74
<표 44> 지시문 기반 첨삭 지원 말뭉치 AI모델 적용(학습) 결과	74

차 례

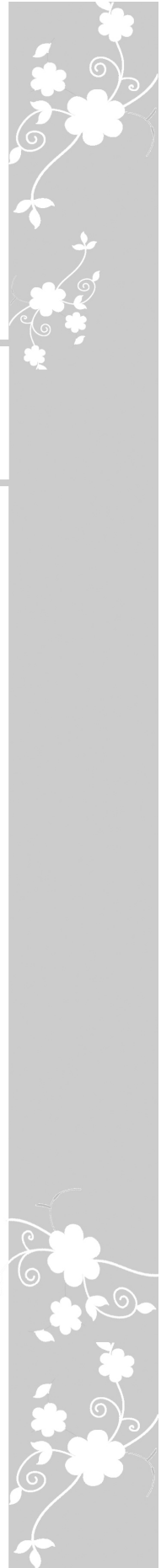
그림 차례

<그림 1> 사업의 목적과 범위	2
<그림 2> 총괄 추진 체계	4
<그림 3> 사업 수행 체계 구성	5
<그림 4> 지시문 기반 탐색 지원 말뭉치 구축 절차	7
<그림 5> 사업 추진 일정	8
<그림 6> 지시문 기반 탐색 지원 말뭉치 설계 절차	11
<그림 7> 언어모델 기반 하위 그룹화를 위한 질문과 답변 형태분석 처리	15
<그림 8> 설계 말뭉치 구축 과정	21
<그림 9> 이용자 세부 질문 유형 파일 예시	24
<그림 10> 지시문 기반 탐색 지원 말뭉치 구축 과정	35
<그림 11> 지시문 기반 탐색 지원 말뭉치 구축 도구(설계 정보 확인)	44
<그림 12> 지시문 기반 탐색 지원 말뭉치 구축 도구(질문-답변 편집)	45
<그림 13> LLM 모델 학습 성과 기반 말뭉치 평가 계획	47



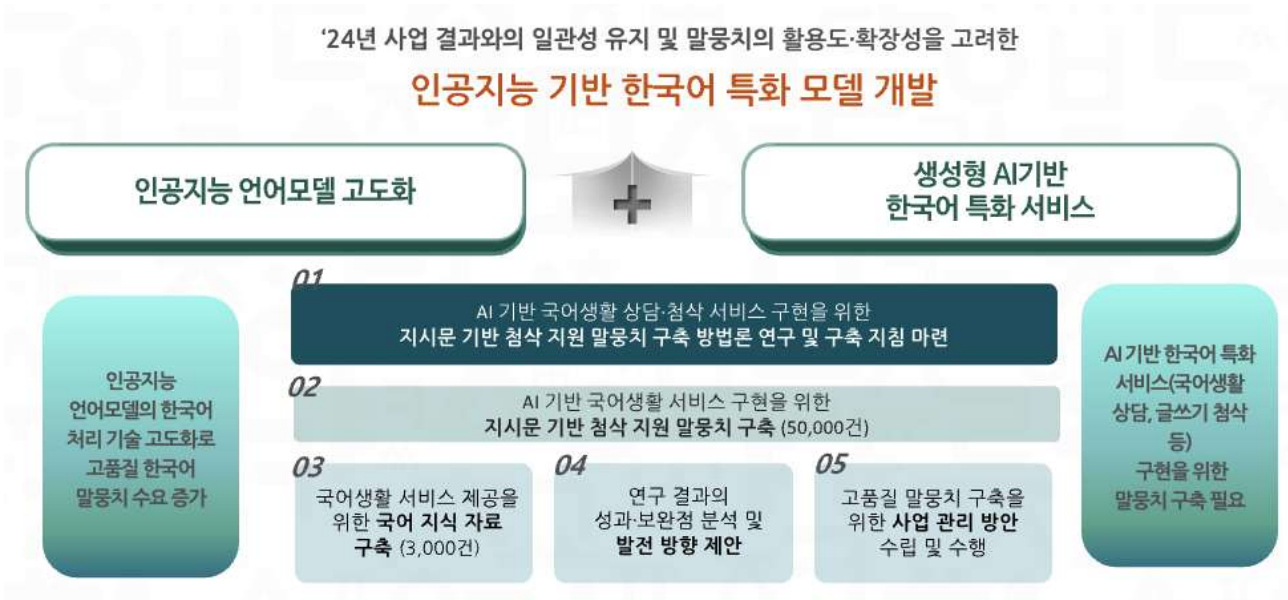
제 1 장

사업 개요



1.1. 사업의 목적과 범위

1.1.1. 사업의 목적



<그림 1> 사업의 목적과 범위

최근 인공지능 기술 분야에서는 빠르게 발전하고 있는 글로벌 빅테크의 영어 중심 모델에 대응해 자국 언어에 강점을 가진 독자 인공지능 모델 개발이 주목받고 있다. 이와 함께 고도의 한국어 처리 기술을 갖는 언어모델 개발을 지원하는 고품질의 한국어 말뭉치 수요도 증가하고 있다. 이런 수요에 대응하고자 본 사업은 인공지능을 이용해 국어생활 상담, 글쓰기 첨삭 등 한국어에 특화된 서비스를 구현하기 위한 지시문 기반 첨삭 지원 말뭉치와 국어 지식 기초 자료를 구축하고, 그 방법론을 연구하는 데 목적을 두고 있다.

1.1.2. 사업의 범위

사업의 수행 범위는 AI 기반 국어생활 상담·첨삭 서비스 구현을 위한 지시문 기반 첨삭 지원 말뭉치 구축 방법론 연구 및 구축 지침을 확립하고, 이에 따라 글쓰기 첨삭 지원 지시문 기반 생성 말뭉치 구축 방법론을 설계해서 국어생활 상담 질문 유형별 세부 질문-답변 50,000건, 국어 지식 기초 자료 3,000건을 구축하는

것이다.

지시문 기반 첨삭 지원 말뭉치는 국어원에서 제공하는 국어생활 상담 자료(가나다전화, 온라인가나다)를 활용해 실제 사례를 중심으로 이용자 질문 유형을 세분하고 각 질문 유형별로 AI와의 상세 질문과 답변을 대화형으로 구성한다. 질문 유형별로 실제 챗봇(인공지능)에 문의할 것으로 예상되는 다양한 상세 질문들을 만들고 각 질문에 대해 모범 답안이 되도록 답변의 품질 요소를 정의하고 답변의 상세 작성 지침을 마련하여 답변을 작성한다. 도출된 이용자 질문 유형별로 50건 내외의 대화 사례를 작성해 총 50,000건 이상의 질문-답변 쌍을 말뭉치로 구축한다.

구축된 말뭉치는 거대언어모델(LLM: Large Language Model)을 2종 이상을 활용해 유용성을 검증하고 개선 사항을 반영해 최종 결과물로 제출한다. 한국어에 맞춰 미세 조정(fine-tuning)된 한국어 특화 소형 언어모델을 검증에 활용하여 학습 전후 평가 결과를 비교하여 제시한다. 또한, 전년도 사업(2024년 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치 구축 연구)의 지침 개선을 수행하고, 그 결과를 반영하여 기구축 말뭉치를 수정·보완(1,500건)한다.

마지막으로 말뭉치 구축 과정 및 구축 결과를 바탕으로 사업의 성과 및 보완점을 분석하고 관련 자문 회의를 개최하여 발전 방향을 제시한다. 이상의 사업 범위와 과업 내용을 정리하면 다음과 같다.

<표 1> 사업의 범위와 세부 과업 내용

사업의 범위	세부 과업 내용
지시문 기반 첨삭 지원 말뭉치 구축	○ AI 기반 국어생활 상담 서비스 구현을 위한 지시문 기반 대화 유형 연구 설계와 지침 ○ 지시문 첨삭 지원 말뭉치 구축 방법론 설계(1,000개 질의 유형 및 샘플) ○ 질의 유형별 세부 질의-응답 말뭉치 50,000건 구축
국어 지식 기초 자료 구축	○ 국어 지식 기초 자료 3,000건 구축
말뭉치 검증	○ 초거대 언어모델을 활용한 성능 평가 및 말뭉치 품질 검증(2개 이상의 언어모델 활용)
전년도 구축 결과 보완	○ '2024년 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치 구축 연구' 사업에서 마련한 지침 분석 및 개선 ○ 전년도 구축 지침 개선 결과를 반영하여 기구축 말뭉치 수정·보완(1,500건)
연구 결과 분석 및 발전 방향 제안	○ 시범 구축 과정 및 결과를 분석하여 성과 및 보완점 분석 ○ 국어생활 서비스 제공을 위한 지시문 기반 말뭉치 관련 자문 의견을 반영하여 결과물 구축(자문회의 1회 이상 개최)

1.2. 사업 추진 방법

1.2.1. 추진 체계 및 수행 방안

효과적인 추진을 위해 말뚝치 구축 과업을 말뚝치 설계, 가공, 검증의 3개 영역으로 구분하고 각 영역을 전담팀으로 구성했다. 또한 각각의 연구 결과물을 상호 검토하면서 사업을 추진할 수 있도록 추진 체계를 조직했다.

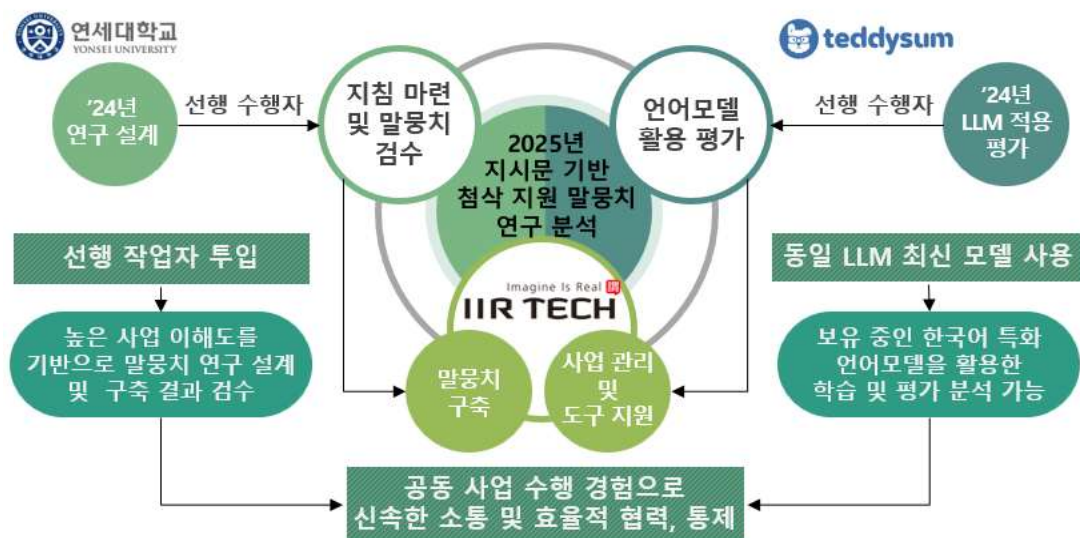


<그림 2> 총괄 추진 체계

말뭉치 설계팀은 국어원에서 제공하는 자료를 분석해 국어생활 상담 자료(가나다 전화, 온라인가나다)의 이용자 질문 유형을 도출하고 각 질문 유형별로 질문-답변의 사례를 작성해 설계 결과물로 제시하고, 말뭉치 가공팀에서 상세 질문과 답변을 추가할 수 있도록 작성 지침을 제작했다.

말뭉치 가공팀은 설계 자료를 토대로 인공지능을 이용해 유사 질문과 답변을 생성하고 생성된 질문과 답변을 말뭉치 설계팀의 지침에 맞춰 검토하면서 부적절한 질문을 제거, 추가, 수정한 후, 각 질문에 대한 답변이 모범 답변이 되도록 수정, 보완했다. 또한 5만 개 이상의 질문-답변 쌍을 효과적으로 가공하고 검증하기 위한 전용 구축 도구를 제작하고 운영관리를 지원했다.

말뭉치 검증팀은 말뭉치 품질 검증을 위한 LLM 평가 지표를 마련한 후 선정된 LLM 모델의 베이스라인(말뭉치 학습 전) 성능과 구축된 말뭉치를 학습한 후 성능 개선 정도를 비교하여 말뭉치의 활용성을 검증했다.



<그림 3> 사업 수행 체계 구성

1.2.2. 위험 요소 및 대응 방안

‘지시문 기반 첨삭 지원 말뭉치’는 1개의 객관적인 답이 있는 소위 레이블 데이터(Labeled Data)와 다른 새로운 유형의 말뭉치이다. 그러므로 초기에 다양한 위험 요소를 도출하고 대응 방안을 모색해 대규모 자료 구축 과정에서 발생하는 자료량 부족, 품질 불량, 일정 지연 등의 위험에 대비했다.

〈표 2〉 예측 위험 요소 및 대응 방안

	예상 어려움 및 쟁점 사항	대책 및 고려 사항
이용자 질문 유형 세분화	<ul style="list-style-type: none"> •1,000개 이상의 유형 도출 가능성 •질문 유형 선정 기준이 필요 	<ul style="list-style-type: none"> •질문 유형의 빈도, 범위(range) 등을 고려하여 질문의 중요도를 평가한 후 중요도가 높은 질문 유형부터 구축 대상에 포함
구축 말뭉치 검증 평가	<ul style="list-style-type: none"> •구축 말뭉치 성능 평가 방안과 언어 모델(LLM)의 능력 평가가 혼동 우려 •전문가의 평가 시 개인적 편차로 인한 평균 점수 변동 우려 •LLM 모델 간의 학습성과 편차 발생 가능성 	<ul style="list-style-type: none"> •LLM의 평가 지표가 말뭉치 구축 지침과 일관성을 갖도록 개선 •LLM 응답 결과에 대해 평가 전문가 확대 또는 자동화된 LLM 평가 적용 •검증용 LLM 모델 다양화(2종 → 4종)
인공지능 생성 응답의 품질	<ul style="list-style-type: none"> •인공지능이 생성한 질문과 답변에 오류 분포가 광범위할 경우(대부분의 질문 답변에 오류가 있는 경우) •유사 질문-답변 많은 경우 	<ul style="list-style-type: none"> •전수 검수, 수정이 가능하도록 말뭉치 가공 일정 수립 •상세 질문-답변별 유사도 측정. 기준값 이상 유사도 수정 또는 제거
작업자 간 일관성 및 정확도	<ul style="list-style-type: none"> •여러 명이 동시에 작업하므로 일관성 문제 발생 •작업 기간 동안 지침 숙지 정도를 평가하기 어려움 •작업 진행에 따라 작업자들의 정확도를 높일 수 있는 방안 필요 	<ul style="list-style-type: none"> •검수자를 중심으로 정기 교육 실시 •지침 추가·수정 시 이를 교육에 즉각적으로 반영 •검수 샘플을 교차 확인하여 일관성 확보 •검수자는 작업자별 오류를 전달하고, 다음 교육 시 해당 오류 유형 개선 여부 확인
품질 지원	<ul style="list-style-type: none"> •최종 납품 품질의 협의를 위한 품질 기준 협의 •공개 전까지 품질 개선/정제를 위한 인력 운용 방안 	<ul style="list-style-type: none"> •국립국어원과 검증 기준 및 방안 협의, 실증 •사업 종료 후 정제 및 개선 대비 핵심 인력 30% 이상 유지

1.2.3. 사업 추진 절차

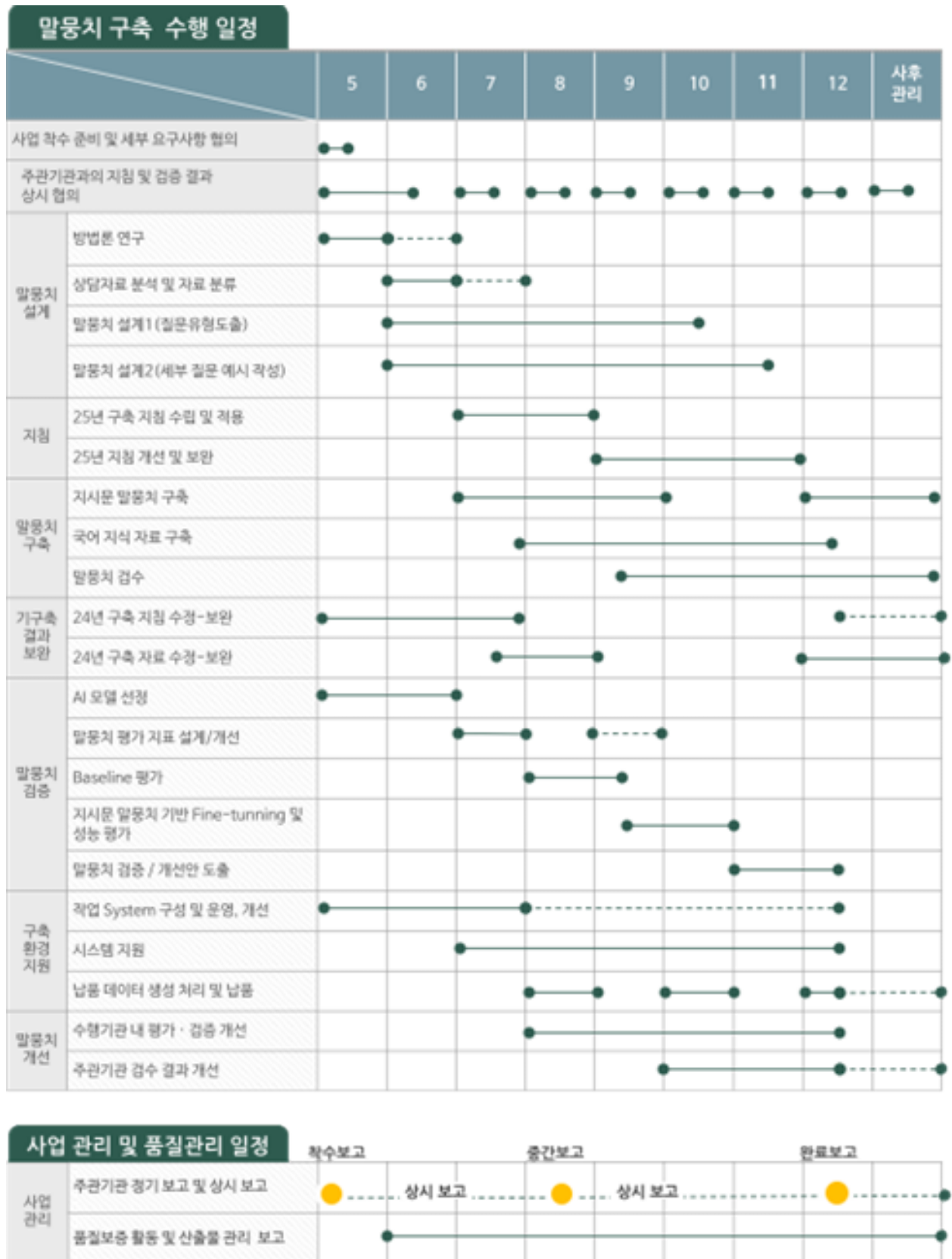
본 사업은 말뚝치 설계, 구축, 검증을 중심으로 추진하되, 구축 결과를 분석하고 자문 의견을 수렴하여 구축 과정 및 결과를 보완하고 발전 방향을 제시하는 과업을 병행 수행한다. 특히 말뚝치 설계, 구축, 검증의 각 과정이 사업 수행 기간 중 계속해서 상호 연계되기 때문에, 구축 공정상 병목현상이 발생하지 않도록 추진 절차를 설계하였다.

예를 들어, 말뚝치 설계팀이 설계 작업에 착수하였을 때 말뚝치 가공팀은 도구 개발과 구축 작업 기초 교육을 시행하였으며, 말뚝치 검증팀은 모델 선정 및 검증 체계를 정비하였다. 말뚝치 설계팀에서 초기 지침 및 상세 질문에 대한 모범 답변 및 예시를 제작하는 대로 자연 없이 말뚝치 가공팀에서 구축 작업이 이루어질 수 있도록 작업 절차를 구성하였다. 말뚝치 품질 검증을 위한 LLM 학습용 말뚝치의 형식과 내용을 말뚝치 구축팀과 말뚝치 검증팀 간 사전 협의하여 확정함으로써 업무 수행에서 발생할 수 있는 공백을 최소화하고 전처리 작업 소요를 낮추도록 했다.

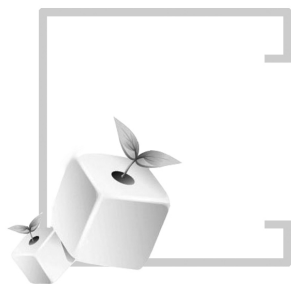


<그림 4> 지시문 기반 검색 지원 말뚝치 구축 절차

1.3. 사업 추진 일정



〈그림 5〉 사업 추진 일정



제 2 장

지시문 기반 첨삭 지원 말뭉치 구축



2.1. 지시문 기반 첨삭 지원 말뭉치 설계

2.1.1. 개요

본 사업에서 “지시문 기반 첨삭 지원 말뭉치”란 일반 국민이 국립국어원 ‘온라인가나다’나 네이버 지식인 등에 자주 문의하는 맞춤법과 문법 설명부터 글쓰기 과정에서 부딪히는 어문 규범이나 작성된 글의 교정과 같은 첨삭에 이르기까지 국어생활 상담·첨삭에 전반에 대한 지시를 정확하게 수행하는 한국어 특화 인공지능 개발용 언어자원이다. 이와 같은 언어자원을 구축하기 위하여 본 연구에서는 우선 국립국어원에서 제공한 국어생활 상담 자료(가나다전화, 온라인가나다)를 분석하여 ‘이용자 질문-국어 지식-모범 답변’을 중심으로 말뭉치를 설계했다. 이 말뭉치는 한국어 어문 규범 및 언어 사용과 관련된 다양한 질문에 대해 인공지능이 정확하고 일관된 답변을 생성할 수 있도록 구축하고 검증하는 데 중점을 두었다.

2.1.2. 설계 방법

말뭉치 설계는 구축 결과가 활용 목적에 맞는 대표성과 균형성을 갖도록 표본(sample text)을 구성하고, 필요한 주석 체계와 규모, 데이터 형식 등을 포함하는 과정이다.¹⁾ 특히 최근 10년간 인공지능에 활용되는 말뭉치는 주로 주석 말뭉치(Annotation Corpus 또는 Labeled Data)로 주석 체계, 구축 데이터의 규모, 데이터 형식(format) 등이 말뭉치 설계의 중요한 고려 사항이었다.

그런데 “지시문 기반 첨삭 지원 말뭉치”는 직접적인 가공 대상이 되는 표본 단위도 가지지 않고 명확한 답을 갖는 주석 체계도 갖지 않는 등 지금까지의 말뭉치와는 양상이 사뭇 다르다. 그럼에도 불구하고 이 말뭉치는 국어생활 상담·첨삭과 같은 질문에 대하여 한국어 어문 규범을 이해하고 정확한 설명을 제시할 수 있는 인공지능 모델 학습(훈련)에 구축 목적이 있으므로, 국어생활 상담 자료의 질문 유형을 도출하는 것이 대표성과 균형성을 갖춘 표본 구성에 해당한다고 할 수 있다. 더불어 별도의 주석 체계는 없으나 하나의 질문 유형을 구성하는 유형명, 관련 어문 규범과 국어 지식, 세부 질문과 답변에 대한 구성과 데이터 구조가 기존 설계

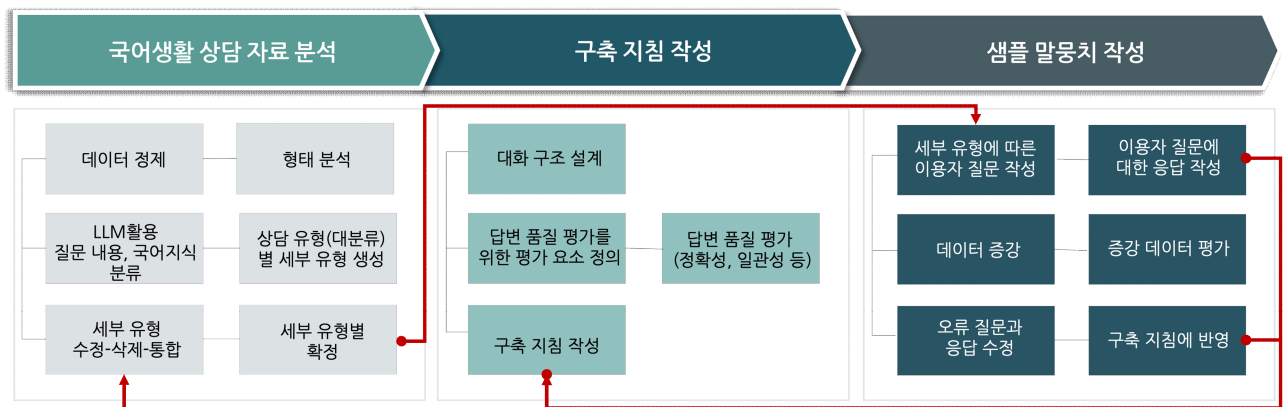
1) Sinclair, J. McH. (2004) ‘Corpus and Text: Basic Principles’, in M. Wynne (ed.) *Developing Linguistic Corpora: A guide to good practice*. Available at <https://users.ox.ac.uk/~martinw/dlc/chapter1.htm#section4>.

결과물에 대응한다.

한편, 1차 설계가 완료되었더라도 분류된 질문 유형별로 50~60개 이상의 세부 질문 유형과 각 질문에 대해 일관성 있는 답변을 생성하기 위해서는 시행착오를 거듭하면서 설계 결과물을 수정하는 것이 필요하다. 그러므로 이러한 특성을 고려해 설계 절차를 수립하고 가공과 검증 계획에도 반영하였다.

2.1.2.1. 설계 절차

설계 절차는 국어생활 상담 자료의 분석, 말뭉치 샘플(예시)의 작성 그리고 설계 문서화 결과물에 대한 말뭉치 구축 지침 작성으로 이루어진다. 진행 과정은 초기에는 자료 분석, 지침 작성, 샘플 말뭉치 작성 순으로 진행되지만, 샘플 말뭉치가 증가함에 따라 질문 유형 분류가 조정되거나 샘플 말뭉치의 질문-답변이 수정-보완됨에 따라 구축 지침도 변경되기 때문에 약 20%(질문 유형 100개, 질문-답변 쌍 500개) 규모로 분류 변경과 지침 검토를 반복해 전체 말뭉치 구축에 반영하도록 했다.



<그림 6> 지시문 기반 침삭 지원 말뭉치 설계 절차

2.1.2.2. 국어생활 상담 자료 분석 : 질문 유형 분류

국어생활 상담 자료 분석의 목표는 상담 자료의 질문과 답변을 어문 규범을 중심으로 한 국어 지식과 이용자 질문 유형으로 상세화하는 것이다. 이 과정에서 이용자가 질문들을 통해 얻고자 질문 의도와 상담 답변들이 제공하고자 한 국어 지

식과 답변 방법, 특징을 파악해 ‘이용자 질문-국어 지식-모범 답변’ 쌍의 내용 구성과 형식을 결정했다. 또한, 샘플을 작성하고 반복적으로 교차 검토를 진행하면서 질문 유형 간 질문의 실제성, 정확성, 일관성 유지에 중점을 두었다. 이를 정리해 생성 증강된 질문과 답변의 품질 기준이 되도록 구축 지침을 작성하고 구축 결과를 검토하여 지침을 반복적으로 보완했다.

분석 대상 자료는 국립국어원으로부터 2021년 11월부터 2025년 5월까지 취합된 ‘가나다전화’, ‘온라인가나다’, ‘우리말365’의 상담 자료 567,340건²⁾을 대상으로 하였다. 분석 대상 자료는 질문, 답변, 분류 유형(이하 대분류 또는 상담 자료 유형)으로 구성되어 있다.

2) 상담 자료의 건수는 더 많았으나 텍스트가 손상되거나 상담 건별로 분리가 되지 않는 자료들을 1차 정제한 결과이다.

<표 3> 국어생활 상담 유형 분류 자료 예시

구분	질문 내용	답변 내용	분류 유형
가나다 전화 (상담자 요약 정리본)	'기가 막힌다는 듯 인상을 썼다.'에서 동사 '막하다'의 의미.	'기운 따위의 흐름이 원활하지 않게 되다'를 뜻함.	의미/화용
	'기막히다는 듯 인상을 썼다.'에 쓰인 형용사 '기막하다'의 의미.	'어떠한 일이 놀랍거나 언짢아서 어이없다'를 뜻함.	의미/화용
	상사분께서 문서를 읽으셨는데 잘못 인지하신 부분이 있었다. '오해가 있으신 부분에 대해 다시 설명드리고자 메일 드렸습니다.'와 같이 썼는데 상사분께서 '오해'라는 말을 상사에게 쓰면 안 된다고 하셨다. 이와 같은 상황이라면 상사분께서 '오해'라는 표현을 개인적으로 안 좋게 느낀 것으로 볼 수 있겠는가.	그와 같이 볼 수 있을 듯함. 상사분께서 안 좋게 느끼셨다면 다른 표현으로 바꿔 쓰실 수도 있겠음.	형태/통사
온라인 가나다	지금 열추 계획했던 대로 맞추어 간다고 볼 수 있고, 차근차근 단계를 밟아가고 있습니다. 맞추어가다, 밟아가다 / 맞추어 가다, 밟아 가다 어떤 게 맞나요?	안녕하십니까? '맞추어 가다', '밟아 가다'처럼 띄어 쓰임이 원칙이되, '밟아 가다'는 붙여 쓰임을 허용합니다. 고맙습니다.	띄어쓰기
	5개년 계획에 비춰봤을 때 지금까지 한 결과가 어떻습니까? 잘되고 있습니까? 계획에 비춰봤을 때 / 비춰봤을 때 어떤 게 맞나요?	안녕하십니까? 문의하신 표현은 '비춰봤을 때'로 표기하는 것이 바릅니다.	맞춤법/부호
	히뿌엿다, 히뿌엿고 / 히뿌였다, 히뿌였고어떤 게 맞나요?	안녕하십니까? 문의하신 표현은 '히뿌엿다/히뿌엿고'로 쓰는 것이 바릅니다.	맞춤법/부호
우리말 365	저변과 지난의 차이가 궁금합니다.	'저변'은 '말하는 때 이전의 지나간 차례나 때'를 뜻하고 '지난'은 '시간이러 그 시기에서 벗어나다'를 뜻하는 사 '지나다'의 관형사형 표현입니다.	의미/화용
	[다저준 답니다] 띄어쓰기, 표기법 확인 부탁드립니다!	'다저 준답니다'로 띄어 쓰임이 원칙이되, 본용언과 보조 용언의 구성이라면 붙여 쓰임을 허용합니다.	띄어쓰기
	엄지발가락, 엄지손가락은 엄지 발가락, 엄지 손가락처럼 띄어 쓰는게 아니라, 붙여 쓰는 게 맞나요?	'엄지발가락', '엄지손가락'은 모두 붙여 씁니다.	띄어쓰기

상담 자료에 식별된 13개 유형 분류를 상세 질문 유형으로 구분하기 위해 각 유형 분류별로 자료를 분리하는 한편, 각 유형 분류별 자료 분포를 산출하였다.³⁾ 사전, 음성/음운, 국어사/문자, 로마자, 건의 등은 1만 개 미만의 상담 사례이지만, 그 외 유형 분류는 모두 1만 개 이상으로 육안으로 모든 자료를 확인할 수는 없었다. 그러므로 이들 상담 사례를 상세 질문 유형으로 구분하기 위해서는 질문 내용과 의도(질문 형태)를 기반으로 그룹화가 필요했다.

3) 1,2차 구분은 자료 수령 일자에 따른 분석 진행상의 구분이기도 하지만, 자료 차수별로 유형 분류 분포에 차이가 있는지 확인하기 위한 구분이기도 했다. 약 2년 간격으로 누적된 2개의 데이터에 대한 유형 분류 분포 비율의 차이는 10% 미만으로 순위의 변동에는 영향이 없는 수준이었다.

<표 4> 국어생활 상담 유형 분류 결과

유형 분류	1차 (2021.11~2023.10)				2차 (2023.11~2025.05)				전체	
	가나다 전화	온라인 가나다	우리말 365	총	가나다 전화	온라인 가나다	우리말 365	총	빈도(개)	비율(%)
1. 띄어쓰기	25,897	6,121	85,711	117,729	14,214	3,031	66,308	83,553	201,282	35.48
2. 형태/통사	24,493	14,687	43,215	82,395	12,959	9,936	35,442	58,337	140,732	24.81
3. 맞춤법/부호	10,570	2,505	22,433	35,508	5,501	1,804	17,962	25,267	60,775	10.71
4. 기타	26,421	2,295	2,143	30,859	10,807	980	2,860	14,647	45,506	8.02
5. 의미/화용	9,184	6,799	8,682	24,665	5,811	3,923	7,369	17,103	41,768	7.36
6. 외래어	2,846	1,429	8,421	12,696	1,202	711	7,291	9,204	21,900	3.86
7. 표준어/표준발음	4,097	2,464	5,830	12,391	2,156	1,305	5,957	9,418	21,809	3.84
8. 언어예절/순화	2,832	863	3,171	6,866	1,490	425	3,316	5,231	12,097	2.13
9. 사전	2,535	489	619	3,643	1,619	220	741	2,580	6,223	1.1
10. 음성/음운	483	606	554	1,643	200	463	493	1,156	2,799	0.49
11. 국어사/문자	406	629	309	1,344	163	371	330	864	2,208	0.39
12. 로마자	419	181	239	839	279	92	164	535	1,374	0.24
13. 건의	210	210	22	442	183	147	10	340	782	0.14
총	110,440	47,279	181,386	339,105	56,584	23,408	148,243	228,235	567,340	100

상담 자료 유형(유형 분류)별 상담 사례 자료를 하위 그룹화할 때는 주제(Topic) 모델 기반의 군집화와 거대언어모델(LLM)을 활용한 그룹화를 적용했다. 주제 모델 기반 군집화는 질문과 답변에 대해 문장 임베딩(sentence embedding) 처리 후 ‘LDA’, ‘BERTopic’, ‘KeyBert’ 3종의 알고리즘을 이용해 군집화를 진행했으나 군집화가 제대로 진행되지 않았다. 대부분의 상담 사례는 “나한테될수록 빨리 이야기해 보도록 해 📖띄어쓰기좀알려주세요.”와 같이 상담 대상(예문)과 상담 의도(적용 규범, 유형 분류)가 함께 포함되어 있는데, 상담 의도에 포함된 유형 분류가 강한 잡음으로 작용해 하위 그룹화가 잘 형성되지 않은 것으로 판단하였다.

거대언어모델(LLM, 이하 언어모델)을 활용한 그룹화는 공동 연구진이 보유한 자체 모델(llama-3-Korean-Blossom-8B)을 활용했다. 상담 자료 유형(유형 분류)별로 프롬프트(prompt)를 작성해 상담 사례의 질문과 답변에 대해 질문 내용과 세부 유형(품사 또는 세부 문법 사항)이 무엇인지 출력하도록 하고, 출력 결과를 기반으로 그룹화를 진행했다.⁴⁾

4) 보다 효과적으로 그룹화할 수 있는 출력 결과를 얻기 위해 질문과 답변에 대한 형태분석 결과를 포함시켰다.

<표 5> 상담 자료 유형(유형 분류)별 상담 사례 하위 분류 프롬프트 예시

Instruction

[Input]은 띄어쓰기에 대한 질문입니다. [Input]에서 [Example]과 같이 구체적인 띄어쓰기 항목과 세부 유형(품사 등)만 출력하세요. 설명은 하지 마세요.

Input

다이어트한 지 얼마나 됐어? 이제 오 일 차야. 위 예문에서 '5(오) 일 차'로 띄어 쓰는 게 바른가요? 그렇다면 '일차'는 어떨 때 쓰이는 건가요?

Example

[Input]

보조용언 띄어쓰기 알려 주세요. 적은 반드시 나가게 되어 있다. 적은 반드시 나가게 되어 있다. (허용인가요?) 커튼이 드리워져 있다. 커튼이 드리워져 있다. (허용인가요?)

[Output]

어 있다, 보조용언

Bllossom출력 : 오 일 차, 의존 명사

언어모델을 이용한 하위 그룹화는 상담 사례를 직접 그룹화하지 못하고 출력에 환각 현상을 포함하기도 하였으나 질문 내용에 대한 사례 부분(오일 차)과 범주(의존 명사)를 출력함으로써 하위 유형화를 위한 후보군을 형성할 수 있다는 장점이 있었다.

질문 내용	답변 내용	분류 유형	구분	질문 내용+답변 내용	질문 내용+답변 내용_형태분석	띄어쓰기 항목	띄어쓰기 항목 형태소 분석(kiwi)
'OO해 주다'는 앞의 활용형이 3음절 이상일 때 'OO해 주다' 띄어쓰기	프로그래밍의 오류임. '공감해 주다' 띄어쓰기	가나다전화_20211	'OO해 주다'는 앞의 활용형이 3음절 이상일 때 'OO해 주다' 띄어쓰기	'OO해 주다'는 앞의 활용형이 3음절 이상일 때 'OO해 주다' 띄어쓰기	공감해 주다	공감/NNG 하/XSV 어/EC 주/VX 다/EC	공감/NNG 하/XSV 어/EC 주/VX 다/EC
'내 옷에 손 댄지 마'에서 '댄지 마'의 띄어쓰기	'댄지 마'로 띄어 씀.	가나다전화_20211	'내 옷에 손 댄지 마'에서 '댄지 마'의 띄어쓰기	'내 옷에 손 댄지 마'에서 '댄지 마'의 띄어쓰기	댄지 마	댄/VV 지/EC 말/VX 어/EF	댄/VV 지/EC 말/VX 어/EF
'부하 직원'의 띄어쓰기	'부하 직원'으로 띄어 씀.	가나다전화_20211	'부하 직원'의 띄어쓰기	'부하 직원'의 띄어쓰기	부하 직원	부하/NNG 직원/NNG	부하/NNG 직원/NNG
'획득 작전'의 띄어쓰기	한 단어가 아니므로 '획득 작전' 띄어쓰기	가나다전화_20211	'획득 작전'의 띄어쓰기	'획득 작전'의 띄어쓰기	획득 작전	획득/NNG 작전/NNG	획득/NNG 작전/NNG
'샐러리맨'의 띄어쓰기	'샐러리맨'으로 붙여 씀.	가나다전화_20211	'샐러리맨'의 띄어쓰기	'샐러리맨'의 띄어쓰기	샐러리맨	샐러리맨/NNG	샐러리맨/NNG
'다리를 건너려는데 저쪽에서도 개가 나타났' '건너려는데'로 붙여 씀.	'건너려는데'로 붙여 씀.	가나다전화_20211	'다리를 건너려는데 저쪽에서도 개가 나타났' '건너려는데'로 붙여 씀.	'다리를 건너려는데 저쪽에서도 개가 나타났' '건너려는데'로 붙여 씀.	건너려는데	건너/VV 려는데/EC	건너/VV 려는데/EC
'생선 많던데'에서 '많던데'의 띄어쓰기	'많던데'로 붙여 씀.	가나다전화_20211	'생선 많던데'에서 '많던데'의 띄어쓰기	'생선 많던데'에서 '많던데'의 띄어쓰기	많던데	많/VX 던데/EF	많/VX 던데/EF
'친구가 널 못 알아보'에서 '알아보'로 띄어 씀.	'알아보'로 띄어 씀.	가나다전화_20211	'친구가 널 못 알아보'에서 '알아보'로 띄어 씀.	'친구가 널 못 알아보'에서 '알아보'로 띄어 씀.	알아보	알아보/VV 었/EP 나/EC 보/VX 어/EF	알아보/VV 었/EP 나/EC 보/VX 어/EF
'종 걸리적거리지 마'에서 '걸리적거리지 마'로 띄어 씀.	'걸리적거리지 마'로 띄어 씀.	가나다전화_20211	'종 걸리적거리지 마'에서 '걸리적거리지 마'로 띄어 씀.	'종 걸리적거리지 마'에서 '걸리적거리지 마'로 띄어 씀.	걸리적거리지 마	걸리적거리지/VV 지/EC 말/VX 어/EF	걸리적거리지/VV 지/EC 말/VX 어/EF
'근데 이게 말이 돼요? 중개가 이렇게 딱하니' '딱하니'는 '보란 듯이 의젓하게' 띄어쓰기	'딱하니'는 '보란 듯이 의젓하게' 띄어쓰기	가나다전화_20211	'근데 이게 말이 돼요? 중개가 이렇게 딱하니' '딱하니'는 '보란 듯이 의젓하게' 띄어쓰기	'근데 이게 말이 돼요? 중개가 이렇게 딱하니' '딱하니'는 '보란 듯이 의젓하게' 띄어쓰기	딱하니	딱하/VX MAG	딱하/VX MAG
'한 번만 해 봐 봐요'에서 '해 봐 봐요'의 띄어 씀	'해 봐 봐요'로 띄어 씀	가나다전화_20211	'한 번만 해 봐 봐요'에서 '해 봐 봐요'의 띄어 씀	'한 번만 해 봐 봐요'에서 '해 봐 봐요'의 띄어 씀	해 봐 봐요	해/VV 어/EC 보/VX 어요/EF	해/VV 어/EC 보/VX 어요/EF
'어떻게 이혼 서류 받으려 직접 오냐?'에서 'C'이혼 서류(들) 받으려'의 구조어 띄어쓰기	'어떻게 이혼 서류 받으려 직접 오냐?'에서 'C'이혼 서류(들) 받으려'의 구조어 띄어쓰기	가나다전화_20211	'어떻게 이혼 서류 받으려 직접 오냐?'에서 'C'이혼 서류(들) 받으려'의 구조어 띄어쓰기	'어떻게 이혼 서류 받으려 직접 오냐?'에서 'C'이혼 서류(들) 받으려'의 구조어 띄어쓰기	어떻게 이혼 서류 받으려	어떻게/NNG 이혼/NNG 서류/NNG 받으려	어떻게/NNG 이혼/NNG 서류/NNG 받으려
'내가 일을 뒷같이 해서'에서 '뒷같이'의 띄어 씀	'뒷같이'로 붙여 씀.	가나다전화_20211	'내가 일을 뒷같이 해서'에서 '뒷같이'의 띄어 씀	'내가 일을 뒷같이 해서'에서 '뒷같이'의 띄어 씀	뒷같이	뒷/NP 같이/KB	뒷/NP 같이/KB
'그 사람 와이프의 통화 기록을 뒤져 봤어요.' '뒤져 봤어요'로 띄어 씀	'뒤져 봤어요'로 띄어 씀	가나다전화_20211	'그 사람 와이프의 통화 기록을 뒤져 봤어요.' '뒤져 봤어요'로 띄어 씀	'그 사람 와이프의 통화 기록을 뒤져 봤어요.' '뒤져 봤어요'로 띄어 씀	뒤져 봤어요	뒤져/VV 어/EC 보/VX 었/EP 어요/EF	뒤져/VV 어/EC 보/VX 었/EP 어요/EF
'다섯 쌍둥이'의 띄어쓰기	'다섯 쌍둥이'와 같이 띄어 씀. -> 띄어쓰기	가나다전화_20211	'다섯 쌍둥이'의 띄어쓰기	'다섯 쌍둥이'의 띄어쓰기	다섯 쌍둥이	다섯/NR 쌍둥이/NNG	다섯/NR 쌍둥이/NNG
'여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기	'여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기	가나다전화_20211	'여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기	'여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기	여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기	여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기	여러 경우에 명사 뒤 '하다'를 붙여 쓰는가. 동작성/서술성이 있는 명사 뒤에 띄어쓰기
'명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기	'명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기	가나다전화_20211	'명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기	'명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기	명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기	명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기	명사 뒤에 '하다'가 쓰였을 때 띄어쓰기' '완성하다', '배출하다'는 한 단어 띄어쓰기
'밀어붙이다'의 띄어쓰기	'밀어붙이다'와 같이 붙여 씀.	가나다전화_20211	'밀어붙이다'의 띄어쓰기	'밀어붙이다'의 띄어쓰기	밀어붙이다	밀어붙이/VV 다/EF	밀어붙이/VV 다/EF
'너희 집 생선 가게 하는구나'에서 '하는구나' 띄어 씀.	'하는구나'로 붙여 씀.	가나다전화_20211	'너희 집 생선 가게 하는구나'에서 '하는구나' 띄어 씀.	'너희 집 생선 가게 하는구나'에서 '하는구나' 띄어 씀.	하는구나	하/VV 는구나/DAF	하/VV 는구나/DAF
'내가 예민하게 구냐?'에서 '예민하게 구냐' 띄어 씀.	'예민하게 구냐'로 띄어 씀.	가나다전화_20211	'내가 예민하게 구냐?'에서 '예민하게 구냐' 띄어 씀.	'내가 예민하게 구냐?'에서 '예민하게 구냐' 띄어 씀.	예민하게 구냐	예민/XR 하/XSA 게/EC 골/VV 나/EC	예민/XR 하/XSA 게/EC 골/VV 나/EC
'쏘아붙이다'의 바른 표기와 띄어쓰기	'쏘아붙이다'로 쓰고, 한 단어 띄어쓰기	가나다전화_20211	'쏘아붙이다'의 바른 표기와 띄어쓰기	'쏘아붙이다'의 바른 표기와 띄어쓰기	쏘아붙이다	쏘아붙이/VV 다/EF	쏘아붙이/VV 다/EF
'고를막지/심술막지' 등은 붙여 쓰는가	'고를막지/심술막지' 등은 붙여 씀.	가나다전화_20211	'고를막지/심술막지' 등은 붙여 쓰는가	'고를막지/심술막지' 등은 붙여 씀.	고를막지/심술막지	고를/NNG 막지/NNG	고를/NNG 막지/NNG
'휴대 및 공중 전화'의 띄어쓰기	'휴대 전화 및 공중전화'와 같이 띄어쓰기	가나다전화_20211	'휴대 및 공중 전화'의 띄어쓰기	'휴대 전화 및 공중전화'와 같이 띄어쓰기	휴대 전화 및 공중전화	휴대/NNG 전화/NNG 및/MAG 공중/NNG 전화/NNG	휴대/NNG 전화/NNG 및/MAG 공중/NNG 전화/NNG
'접수 및 처리 절차' 띄어쓰기	'접수 및 처리 절차'로 띄어 씀.	가나다전화_20211	'접수 및 처리 절차' 띄어쓰기	'접수 및 처리 절차'로 띄어 씀.	접수 및 처리 절차	접수/NNG 및/MAG 처리/NNG 절차/NNG	접수/NNG 및/MAG 처리/NNG 절차/NNG
'차르락차르락'의 띄어쓰기	표준어는 아니나 쓴다면 '차르락' 띄어쓰기	가나다전화_20211	'차르락차르락'의 띄어쓰기	'차르락' 띄어쓰기	차르락	차르락/NNG	차르락/NNG
'김 회장의 띄어쓰기	'김 회장'과 같이 띄어 씀.	가나다전화_20211	'김 회장의 띄어쓰기	'김 회장'과 같이 띄어 씀.	김 회장	김/NNG 회장/NNG	김/NNG 회장/NNG
'김 목사의 띄어쓰기	'김 목사'와 같이 띄어 씀.	가나다전화_20211	'김 목사의 띄어쓰기	'김 목사'와 같이 띄어 씀.	김 목사	김/NNG 목사/NNG	김/NNG 목사/NNG
'김 사장의 띄어쓰기	'김 사장'과 같이 띄어 씀.	가나다전화_20211	'김 사장의 띄어쓰기	'김 사장'과 같이 띄어 씀.	김 사장	김/NNG 사장/NNG	김/NNG 사장/NNG
'돼지고기찌개'의 띄어쓰기	'돼지고기찌개'와 같이 붙여 씀.	가나다전화_20211	'돼지고기찌개'의 띄어쓰기	'돼지고기찌개'와 같이 붙여 씀.	돼지고기찌개	돼지고기/NNG 찌개/NNG	돼지고기/NNG 찌개/NNG

<그림 7> 언어모델 기반 하위 그룹화를 위한 질문과 답변 형태분석 처리

그러나 언어모델의 출력 결과가 어느 정도 하위 유형화 그룹을 형성하더라도 여전히 충분한 그룹을 형성하지 못하거나 형성된 그룹에 오류나 예외적인 사례들이 포함되어 있었다. 그러므로 하위 유형화를 형성한 상담 자료들을 기반으로 설계 말뭉치 작성자들이 세부 질문 유형을 목록화하고 샘플 말뭉치를 작성하면서 세부 질문 유형을 재분류하는 과정을 반복했다.

〈표 6〉 언어모델 출력 결과를 활용한 세부 질문 유형 재분류 결과

질문 유형 분류 범주	언어모델 출력 결과		세부 질문 유형
띄어쓰기	용언	보조용언	보조 동사 '마지않다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 동사 '말다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 동사 '못하다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 형용사 '싫다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 형용사 '만하다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 형용사 '직하다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 형용사 '듯하다' 띄어쓰기
띄어쓰기	용언	보조용언	본용언이 합성어나 파생어이고 그 활용형이 3음절 이상인 경우에 보조용언 띄어쓰기
띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어오다', '아/어가다' 띄어쓰기
띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어지다' 띄어쓰기
띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어하다' 띄어쓰기
띄어쓰기	용언	보조용언	보조 용언 띄어쓰기 관련 어문 규정 설명 요청
띄어쓰기	용언	보조용언, 동사	사전에 등재된 한 단어 합성 동사의 띄어쓰기

세부 질문 유형의 분류는 상위 분류인 질문 유형 분류 1개(띄어쓰기 등)에 대해 모든 참조 국어 지식과 5개의 질문-답변 예시가 모두 작성될 때까지 반복했다. 또한 13개의 질문 유형 분류 중 답변이 참조하는 어문 규범이나 국어 지식이 현저히 다른 표준어와 표준발음, 언어 예절과 다듬은 말(순화어), 국어사와 문자, 맞춤법과 부호는 유형 분류를 별도로 구분했다. 최종 선별된 세부 질문 유형은 총 872개이다.⁵⁾

〈표 7〉 질문 유형 범주별 세부 질문 유형과 예시 질문-답변쌍의 수

질문 유형 분류 범주	세부 질문 유형 수(개)	예시 질문-답변쌍 수(개)	비율(%)
의미/화용	120	600	13.44%

5) 질문 유형 목록 전체는 [부록 1]에 제시하였다.

외래어	118	590	13.21%
표준발음	103	515	11.53%
맞춤법	89	445	9.97%
띄어쓰기	82	410	9.18%
로마자	73	365	8.17%
기호	61	260	5.82%
음성/음운	52	260	5.82%
사전	13	215	4.82%
형태/통사	37	185	4.14%
표준어	35	175	3.92%
언어예절	33	165	3.70%
순화어	30	150	3.36%
건의	13	65	1.46%
국어사	8	40	0.90%
문자	5	25	0.56%
합계	872	4,465	100%

2.1.2.3. 말뭉치 구성 요소 및 데이터 형식

도출된 세부 질문 유형은 ‘이용자 질문’과 ‘모범 답변’(이하 대화 쌍)을 핵심 데이터로 한다. 각 대화 쌍은 이용자가 실제로 챗봇을 이용하는 상황을 가정하여 실제 국어상담 자료를 분석하여 ‘이용자 질문’을 작성하고 이용자 질문에 답변하기 위한 ‘국어 지식’을 구축한 후 이를 바탕으로 ‘모범 답변’을 작성하는 방식으로 구축되었다. 지시문 기반 탐색 지원 말뭉치의 기본 구성 요소는 아래와 같다.

<표 8> 지시문 기반 탐색 지원 말뭉치의 기본 구성 요소

기본 구성 요소	내용
유형 번호	상세 질문 유형별 고유 번호
예시 번호	유형별 전문가 작성 이용자 질문, 모범 답변의 쌍 번호(1~5)
이용자 질문 유형	이용자 질문의 유형
국어 지식	답변 작성 시 근거로 삼은 규범·사전·문법 정보
이용자 질문	이용자 예상 질문
모범 답변	이용자 질문에 대응하는 모범 답변
꼬리 질문 여부	앞 번호 질문·답변과 연관되는 질문인지 여부

“지시문 기반 첨삭 지원 말뭉치”의 데이터 형식은 “2024년 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치 구축 연구”의 JSON 규격을 준용하되 질문-답변 쌍을 활용하기 용이하도록 수정했다. 지시문 특성은 유지하되 질문-답변 정보가 효과적으로 활용될 수 있도록 AI 모델이 생성한 출력(output) 데이터 구조는 생성된 질문, 답변, 작업자 수정 답변으로 간소화했다.

1수준은 국립국어원 말뭉치의 데이터 규격을 승계하여 ‘id, metadata, document’로 구성된다. ‘metadata’는 말뭉치로써의 데이터 설명값을 기록하고, ‘document’는 표본으로서의 데이터값을 기록한다. 표본 구성 단위는 국어 생활 상담 질문 유형으로 총 872개의 유형이다.

2수준은 ‘document’의 값만 변화하며 ‘id, category, contents’로 구성된다. ‘document.id’는 표본의 구분자이므로 분류된 상담 유형을 나타내고 유효값 범위는 ‘GWIG2502511300.1~GWIG2502511300.872’까지이다. ‘document.category’는 질문 유형의 명칭값이다. ‘document.contents’는 질문 유형별 말뭉치 구성 데이터를 표현한다.

‘document.contents’는 ‘id, model, instruction’으로 구성되며, ‘id’는 현재 질문 유형에 대한 질문-답변 쌍의 일련번호이고, ‘model’은 질문과 답변 생성에 사용된 AI 언어모델 API의 서비스명이다. ‘instruction’은 국어 생활 상담 질문-답변 쌍을 구성하는 정보들로 ‘knowledge(국어 지식 또는 어문 규범의 내용), knowledge_info(국어 지식 또는 어문 규범의 출처), 질문-답변 쌍(question, answer), linked-question(꼬리질문인 경우 원질문 번호: metadata.id의 마지막 숫자값), is_generated(생성된 질문-답변 쌍 여부)⁶⁾, revised_comment(수정된 답변)’으로 구성해 현재 표본의 질문 유형에 대해 질문-답변 쌍의 추가, 확장이 용이하고, AI에 의한 증강이나 사람에 의한 수정을 통한 활용에도 용이하다.

<표 9> 지시문 기반 첨삭 지원 말뭉치 데이터 구조 (json)

1수준	2 수준	3 수준	4 수준	타입	예시	비고
id				str	CXIG2502511300	원문: 상담자료-무작위(CX) 분석 층위: 지시문-

6) 생성된 질문 여부는 질문 생성 시 AI모델로부터 응답을 받는 시점에 생성된다. ‘metadata.model’이 ‘null’이 아닌 경우 ‘is_generated’는 반드시 ‘true’여야 한다.

						생성(IG) 사업 연도: 25 구축 유형: 0 구축 날짜: 251130 수정 횟수: 0
metadata				obj		
	title			str	국립국어원 2025년 지시문 기반 첨삭 지원 말뭉치 CXIG2502511300	
	creator			str	국립국어원	
	distributor			str	국립국어원	
	year			str	2025	
	category			str	비출판물	
	annotation_level			str	지시문 기반 생성	
	sampling			str	부분 추출-임의 추출	
document				arr(obj)		
	id			str	CXIG2502511300.1	(1수준 id).number
	category			str	형용사파생접미사 '-스럽다' 띄어쓰기	질문 유형
	contents			arr(obj)		
		id		str	GWIG2502511300.29.6	상세 질문 유형 id
		model		str	gpt-4o	생성AI모델(cf. 설계 말뭉치의 경우 전문 가에 의해 작성되므 로 null로 표시)
		instruction		obj		
			knowledge	str	-스럽다'는 일부 어근의 뒤에 붙어 '그러한 성질이 있음'의 뜻을 더하고 형용사를 만드는 접미사이다. '-스럽다'가 결합한 파생어의 예로는 '복스럽다, 걱정스럽다, 자랑스럽다, 거북스럽다, 조잡스럽다' 등이 있다.	관련 국어 지식 / 어문 규범
			knowledge_info	str	표준국어대사전	관련 국어 지식 / 어문 규범의 출처
			question	str	그렇다면, '복스럽다'와 '복 스럽다' 중 어떤 표현이 맞나요?	is_generated : false(설계 말뭉치 질문 또는 작업자 수정) true(AI모델이 생성한 질문)
			answer	str	안녕하세요? '복스럽다'와 '복 스럽다' 중 어떤 표현이 맞는지에 관한 질문에 답변드리겠습니다. '-스럽다'는 형용사를 만드는 접미사로서, 앞에 오는 어근과 붙여서 하나의 단어로 써야 합니다. 따라서 '복스럽다'는 '복'이라는 어근에 '-스럽다'가 붙어 '복이 많은 듯한' 뜻을 나타내는 하나의 형용사입니다. 반면에 '복 스럽다'처럼 띄어 쓰는 것은 어문	is_generated : false(설계 말뭉치 질문 또는 작업자 수정) true(AI모델이 생성한 질문)

					<p>규범상 맞지 않습니다.</p> <p>[한국어 어문 규범]에 따르면 ‘-스럽다’는 일부 어근 뒤에 붙어 ‘그러한 성질이 있음’을 뜻하는 형용사 접미사이며, ‘복스럽다’는 그 예시 중 하나입니다. 따라서 ‘복스럽다’로 붙여 쓰는 것이 바릅니다.</p> <p>예를 들어, ‘복스럽게 밥을 먹다’, ‘복스러운 얼굴’과 같이 쓸 수 있습니다. 띄어 쓰면 의미가 분리되어 어색하게 표현됩니다.</p> <p>다만, 접미사가 아닌 다른 경우로 ‘복’과 ‘스럽다’가 문장 성분상 독립적으로 쓰이는 특별한 경우는 거의 없으므로 이를 띄어 쓰는 일은 드뭅니다.</p> <p>요약하자면, ‘복스럽다’는 ‘-스럽다’ 접미사가 붙은 형용사이므로 붙여 쓰는 것이 옳고, ‘복 스럽다’로 띄어 쓰는 것은 잘못된 표기입니다.</p> <p>참고하시기 바랍니다. 감사합니다.</p>	
			linked_question	str	GWIG2502511300.29.5	꼬리 질문 여부를 나타내며 연관된 원 질문 번호를 부여(꼬리 질문이 아닌 경우 : 0)
			is_generated		true	AI모델에 의해 생성된 질문-답변 여부 (설계말뭉치에 해당하는 id: 1~5는 false, 질문-답변에 대해 증강생성을 시도한 경우는 true)
			revised_comment	str	<p>안녕하세요? '-스럽다'의 띄어쓰기에 관한 질문에 답변드리겠습니다. '-스럽다'는 '그러한 성질이 있음'의 뜻을 더하는 접미사로, 앞말과 반드시 붙여 써야 합니다. 따라서 '복스럽다'가 올바른 표현이며, '걱정스럽다', '자랑스럽다' 등도 모두 붙여 씁니다. 감사합니다.</p>	수정 답변

2.1.3. 설계 말뭉치 구축

“지시문 기반 첨삭 지원 말뭉치”의 설계 결과물은 여타 말뭉치와 달리 최종 말뭉치 표본의 일부를 구축하는 것을 포함한다. 단순히 말뭉치 구축 사양(규격)으로써의 설계 결과물이 아니라 AI가 설계 정보를 기반으로 질문을 이해하고 답변을 생성할 수 있도록 해야 하기 때문이다. 그러므로 앞서 도출한 세부 질문 유형과 이에 대한 예시 질문-답변, 그리고 관련 국어 지식을 함께 작성해 AI가 유사 질문과 그에 대한 정확한 답변을 생성할 수 있도록 하였다.

2.1.3.1. 구축 방법

설계 말뭉치는 말뭉치 구축 공정상 AI가 유사 질문과 답변을 생성하는 데 사용되므로 세부 질문 유형별로 예시 질문과 모범 답변을 정교하게 구축해야 한다. 또한, 각 질문-답변 쌍은 챗봇과 같은 대화형 AI의 응답을 고려해 하나의 완결된 대화 쌍으로 작성하는 등 다양한 고려 사항을 포함하고 있다. 그러므로 설계 말뭉치 구축 과정은 아래 그림과 같이 반복적인 단계를 거쳤다. 특히 이용자 질문과 답변의 실제성, 정확성, 일관성을 유지하기 위해 구축 지침을 작성하고 질문-답변 쌍을 검수하기 위한 점검 항목도 수립해 제시했다.



<그림 8> 설계 말뭉치 구축 과정

설계 말뭉치 구축은 1개의 세부 질문 유형별로 5개의 예시 질문-답변 쌍을 작성하는 과정을 통해 세부 유형 분류가 적절한지, 적용된 국어 지식이 적합한지를 확인하는 과정이기도 하다. 그러므로 5개의 질문-답변을 작성하며 세부 질문 유형, 국어 지식, 질문, 답변이 일관되고 정확한지 점검표를 통해 반복적으로 확인해 하나의 질문 유형으로 확정하였다.

<표 10> 설계 말뭉치 작성 단계별 작업 내용

	단계	작업 내용 요약
①	이용자 질문 유형 확인	이용자 질문 유형이 적합한지, 이용자 질문 유형 목록의 명칭과 일치하는지 확인.
②	국어생활 상담 자료 확인	국어생활 상담 자료에서 해당 오류 유형을 검색하여 실제로 나타난 질문 유형과 답변을 살핌.
③	이용자 질문 작성	국어생활 상담 자료에 나타난 질문 유형을 참고하여 실제로 이용자가 할 질문을 작성함.
④	출처 및 국어 지식 작성	참고한 어문 규정이나 자료를 제시함. 참고 규정이나 자료를 최대한 제시하되, 참고 규정이나 자료를 찾을 수 없다면 해당 셀에는 'x(소문자)'로 제시함.
⑤	모범 답변 작성	국어생활 상담 자료의 대답과 어문 규정, 표준국어대사전 등을 참고하여 답변을 작성함.
⑥	점검표 항목 점검	첨부된 점검표의 점검 항목을 점검하여 작성한 내용을 점검하고 오류가 있으면 수정함.

설계 말뭉치 특성상 반복적인 수정이 이루어지기 때문에 평가 지표를 정하고, 부적합한 부분에 대하여 수정한 후 검수 과정을 한 번 더 거쳤다. 평가 지표와 기준은 아래 표와 같다.

<표 11> 생성 질의-답변 평가 지표

번호	평가 지표	정의 및 설명
1	표현 적합성	질문 및 답변의 문장이 자연스럽고 의도가 명확하게 전달되는가?
		→ 어색한 문장 구조, 의미가 불분명한 표현, 화용적 오류 등 여부를 판단함
2	범위 적합성	질문과 답변 전체가 평가 대상이 되는 지정 주제나 오류 유형 범위 안에 포함되는가?
		→ 주제와 무관한 내용, 다른 오류 유형이나 범위를 벗어난 질문/답변이 포함되어 있는지 여부를 평가함
3	답변의 적합성	질문에 대해 제공된 국어 지식을 활용하여 정확하게 답했는가?
		→ 설명의 논거가 제공된 국어 지식에 근거해야 하며, 단정적 오류, 엉뚱한 지식 사용 여부 등을 포함
4	답변의 수준 적합성	답변이 이해하기 쉽고 친절하게 작성되었는가?
		→ 전문 용어 사용 시 설명 포함 여부, 학술적 또는 일반 독자가 이해하기에 적절한 난이도인지 평가함
5	맞춤법 및 표기법 적합성	질문과 답변에 문법, 맞춤법, 띄어쓰기 오류가 없는가?
		→ 질문이나 대답 자체에 문법 오류를 포함하여 혼란을 주지 않는지 확인
6	질문-답변의	질문과 답변이 내용적으로 잘 맞물려 있는가?
		→ 질문의 핵심에 정확히 대응하는 답변이며, 논리적 연결이 부자연스럽지

	상호 보완성	않은지 판단
7	질문-답변의 중복성	질문과 답변 내에서 같은 내용이 불필요하게 반복되고 있지 않은가?
		→ 답변이 단순히 질문 내용을 반복하거나, 새로운 정보나 설명을 제공하지 않는지 판단

완성된 설계 말뭉치의 세부 질문 유형(이용자 질문 유형) 1개의 예시는 아래와 같다. 하나의 ‘이용자 질문 유형’에는 5개의 ‘이용자 질문-모범 답변’ 대화 쌍을 포함하였다. 3개 대화 쌍은 단일 질문으로, 2개 대화 쌍은 꼬리질문 유형으로 구성하였다.

<표 12> 말뭉치 설계 자료 예시

구분		내용
이용자 질문 유형		동사 '못쓰다' 띄어쓰기
출처		한글 맞춤법 제5장 제2절 제46항, 표준국어대사전
국어 지식		단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다. '못쓰다'는 '((‘못쓰게’ 꼴로 쓰여)) 얼굴이나 몸이 축나다', '((주로 ‘-으면’, ‘-어서’와 함께 쓰여)) 옳지 않다. 또는 바람직한 상태가 아니다.'의 의미로 쓰인다.
예시	이용자 질문(1)	‘못 쓰다’는 언제 붙여 쓰나요?
	모범 답변(1)	안녕하세요? ‘못쓰다’의 띄어쓰기에 대해 답변드립니다. ‘못쓰다’는 ‘얼굴이 상하다’, ‘행동이 바람직하지 않다’, ‘옳지 않다’와 같은 의미로 쓰일 때 ‘못쓰다’와 같이 붙여 씁니다. 감사합니다.
	이용자 질문(2)	그럼, '이 망치는 더 못 쓴다.'는 '더못 쓴다'로 붙여 쓰나요?
	모범 답변(2)	앞선 질문에 이어 답변드립니다. 한글 맞춤법 제46항에서는 단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다는 규정이 있으나 표현이 의미적으로 자연스럽게 하나로 이어질 때에만 붙여 쓸 수 있습니다. 그러므로 '다 못 쓰다'로 띄어 쓰는 것이 바릅니다. 감사합니다.
	이용자 질문(3)	‘먹어 버릇하면 버릇이 돼서 못 쓴다.’에서는 ‘못 쓴다’를 붙여 써야 하나요?
	모범 답변(3)	안녕하세요? ‘못쓰다’의 띄어쓰기에 대해 답변드립니다. ‘옳지 않다.’ 또는 ‘바람직한 상태가 아니다.’를 뜻한다면 ‘못쓴다’로 붙여 씁니다. 그러므로 ‘먹어 버릇하면 버릇이 돼서 못쓴다.’로 붙여 쓰는 것이 바릅니다. 감사합니다.
	(…)	

2.1.3.2. 이용자 세부 질문 유형 선별

설계 말뭉치는 하나의 이용자 질문 세부 유형(이하 질문 유형)별로 1개의 결과

대분류	중분류	소분류	질문 유형	국어학당 상담 자료 예시	비고
띄어쓰기	용언	보조용언	보조 동사 '주다' 띄어쓰기	이용자 질문 유형	주요??? 문헌(70쪽, 84쪽)182, 문법(91쪽), 언어지식, 내역(54쪽) ●●●●●, 표준국어대사전, 네이버(2008), 달인(2008), 국어(1982), 등
띄어쓰기	용언	보조용언	보조 동사 '보다' 띄어쓰기		쓰기가 어떻게 하는 표현인가? 문헌(70쪽, 84쪽)182, 문법(91쪽), 언어지식, 내역(54쪽) ●●●●●, 표준국어대사전, 네이버(2008), 달인(2008), 국어(1982), 등
띄어쓰기	용언	보조용언	보조 동사 '버리다' 띄어쓰기		의 띄어쓰기. 문헌(70쪽, 84쪽)182, 문법(91쪽), 언어지식, 내역(54쪽) ●●●●●, 표준국어대사전, 네이버(2008), 달인(2008), 국어(1982), 등
띄어쓰기	용언	보조용언	보조 동사 '있다' 띄어쓰기		보조 용언 '있다'의 띄어쓰기가 궁금하다. '열려 있다'는 붙여 쓰는가. 문헌(70쪽, 84쪽)182, 문법(91쪽), 언어지식, 내역(54쪽) ●●●●●, 표준국어대사전, 네이버(2008), 달인(2008), 국어(1982), 등
띄어쓰기	용언	보조용언	보조 동사 '내다' 띄어쓰기		'내다': '~나다' 단어를 쓸 경우 ~ 내다. ~ 나므로 띄어 쓴 건가요? 내다를 붙여 써서 내었다 문헌(70쪽, 84쪽)182, 문법(91쪽), 언어지식, 내역(54쪽) ●●●●●, 표준국어대사전, 네이버(2008), 달인(2008), 국어(1982), 등
띄어쓰기	용언	보조용언	보조 동사 '없다' 띄어쓰기		'밀어'의심치않다'의 띄어쓰기는 어떻게 하나요? 말다(64), 않다(44), 못하다(36)
띄어쓰기	기	용언	보조용언	보조 동사 '마지않다' 띄어쓰기	바라마지않다'로 붙여 쓸 수 있나? 마지않다(12)
띄어쓰기	기	용언	보조용언	보조 동사 '말다' 띄어쓰기	차이 마요의 띄어쓰기./ '하지 마'를 띄어 쓰는 이유가 무엇인가? 말다(64), 않다(44), 못하다(36)
띄어쓰기	기	용언	보조용언	보조 동사 '못하다' 띄어쓰기	히조차지 못한다'에서 '못한다'의 띄어쓰기. 말다(64), 않다(44), 못하다(36)
띄어쓰기	용언	보조용언	보조 형용사 '싫다' 띄어쓰기	싫어 하다'를 띄우고 했는데~하더라'는 앞말에 붙이는 것이 아닌가? 싫다(172), 적하다(28), 만나다(28), 못하다(20),	
띄어쓰기	용언	보조용언	보조 형용사 '만하다' 띄어쓰기	흔들흔들 흔들흔들 흔들어떨던 것은 흔들 흔들어떨기인가? 같은 문장 가운데 앞에 주어진 일자를줄였다? 싫다(172), 적하다(28), 만나다(28), 못하다(20),	
띄어쓰기	용언	보조용언	보조 형용사 '적하다' 띄어쓰기	미웠을 적하다는 띄어 쓰는 것이 원칙이지 붙여 씀도 허용되는가? 싫다(172), 적하다(28), 만나다(28), 못하다(20),	
띄어쓰기	용언	보조용언	보조 형용사 '듯하다' 띄어쓰기	문헌(70쪽)에 실려있으나 파생어이고 그 활용형이 3종 이상일 경우에 보조용언 띄어쓰기 싫다(172), 적하다(28), 만나다(28), 못하다(20),	
띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어도', '-아/어가다' 띄어쓰기	'설명해 주다'도 띄어 쓸만이 바른가? '역사가 흘러 오다'에서 '흘러 오다'의 띄어쓰기.	
띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어지다' 띄어쓰기	<신동철>이라는 붙여 쓰나요? 붙여 쓴다면 문법 근거가 무엇이지요?	
띄어쓰기	용언	보조용언	보조 용언 띄어쓰기 관련 여러 규정 설명 요청	문헌(70쪽)에 실려있으나 파생어이고 그 활용형이 3종 이상일 경우에 보조용언 띄어쓰기 '역사가 흘러 오다'에서 '흘러 오다'의 띄어쓰기.	
띄어쓰기	용언	보조용언, 동사	사전에 등재된 한 단어 합성 동사의 띄어쓰기	가장 쉬운은 건강 방법'에서 '순쉬운'의 띄어쓰기.	
띄어쓰기	용언	보조용언, 동사	사전에 등재된 한 단어 합성 동사의 띄어쓰기	가장 쉬운은 건강 방법'에서 '순쉬운'의 띄어쓰기.	
띄어쓰기	용언	보조용언, 형용사	사전에 등재된 한 단어 합성 동사의 띄어쓰기	가장 쉬운은 건강 방법'에서 '순쉬운'의 띄어쓰기.	
띄어쓰기	의존태	동사파생접미사	동사파생접미사 '-하다' 띄어쓰기	'남방하다'가 한 단어가 아니므로 '남방 하다'로 띄어 쓰는가? 하다(156), 당하다(60), 시키다(60), 되다(24), 받다(16)	
띄어쓰기	의존태	동사파생접미사	동사파생접미사 '-당하다' 띄어쓰기	괴롭힘당하다 를 형용사로 쓸 때도 괴롭힘당하는 이렇게 붙여 쓰나요? 하다(156), 당하다(60), 시키다(60), 되다(24), 받다(16)	
띄어쓰기	의존태	동사파생접미사	동사파생접미사 '-시키다' 띄어쓰기	괴롭힘당하다 를 형용사로 쓸 때도 괴롭힘당하는 이렇게 붙여 쓰나요? 하다(156), 당하다(60), 시키다(60), 되다(24), 받다(16)	

작성된 설계 말뭉치는 구축 도구 내에서 AI로 추가 질문과 답변을 생성하면서 확장해야 하기 때문에 공통 양식을 활용했다.⁷⁾ 작성 양식 파일의 파일명 뒤에 유형 index를 추가하여(지시문 기반 말뭉치_작성 양식_유형 index.xlsx) 저장하였다.

유형 번호	예시 번호	이용자 질문 유형	출처	국어 지식	이용자 질문	모범 답변	표리 질문 여부
1	1	보조 동사 '주다' 띄어쓰기	한글 맞춤법 제5장 제3절 제47항	보조 용언은 띄어 씀을 원칙으로 하되, 경우에 따라 붙여 씀도 허용한다. '보용언+아/어+보조 용언' 구성과 관형사형+보조 용언(이존 명사+하다/싫다) 구성의 경우는 붙여 쓰는 것이 허용된다. 앞말에 조사가 붙거나 앞 단어가 합성어이거나 파생어인 경우 보조 용언을 앞말에 붙여 쓰지 않는다. 본용언이 합성어나 파생어라도 그 활용형이 2음절인 경우에는 붙여 쓴 말이 너무 긴 것을 이유로 본용언과 보조 용언을 붙여 쓸 수 있다.	설명해 주다'는 왜 띄어 써야 하나요?	안녕하세요? '설명해 주다'의 띄어쓰기에 대한 질문에 답변하겠습니다. '설명해 주다'는 본용언 '설명하다'와 보조 용언 '주다'로 구성된 보조 용언 구성입니다. 본용언과 보조 용언은 띄어 쓰는 것이 원칙이나, '보용언+아/어+보조 용언' 구성의 경우에는 붙여 쓰는 것이 허용됩니다. 그러나 본용언이 파생어인 경우는 본용언에 붙여 쓰면 전체가 너무 길어지므로 이때는 보조 용언을 앞말에 붙여 쓰지 않습니다. '설명하다'는 명사 '설명'에 동사 파생 접미사 '-하다'가 붙은 파생어입니다. 따라서 '설명해 주다'는 띄어 써야 합니다. 고맙습니다.	X
1	2	보조 동사 '주다' 띄어쓰기	한글 맞춤법 제5장 제3절 제47항	보조 용언은 띄어 씀을 원칙으로 하되, 경우에 따라 붙여 씀도 허용한다. '보용언+아/어+보조 용언' 구성과 관형사형+보조 용언(이존 명사+하다/싫다) 구성의 경우는 붙여 쓰는 것이 허용된다. 앞말에 조사가 붙거나 앞 단어가 합성어이거나 파생어인 경우 보조 용언을 앞말에 붙여 쓰지 않는다. 본용언이 합성어나 파생어라도 그 활용형이 2음절인 경우에는 붙여 쓴 말이 너무 긴 것을 이유로 본용언과 보조 용언을 붙여 쓸 수 있다.	보여 주다/보여 주다' 중 어느 쪽 띄어쓰기가 맞나요?	안녕하세요? '보여 주다'의 띄어쓰기에 대한 질문에 답변하겠습니다. '보여 주다'는 본용언 '보이다'와 보조 용언 '주다'로 구성된 보조 용언 구성입니다. 본용언과 보조 용언은 띄어 쓰는 것이 원칙이나, '보용언+아/어+보조 용언' 구성의 경우에는 붙여 쓰는 것이 허용됩니다. 따라서 '보여 주다'로 쓰는 것이	X

- 24 -

				언어 합성어나 파생어라도 그 활용형이 2음절인 경우에는 붙여 쓴 말이 너무 긴 것은 아니므로 본용언과 보조 용언을 붙여 쓸 수 있다.		원칙이나 '보여준다'로 쓰는 것도 가능합니다. 고맙습니다.	
1	3	보조 동사 '준다' 띄어쓰기	한글 맞춤법 제5장 제3절 제47항	보조 용언은 띄어 쓰음을 원칙으로 하되, 경우에 따라 붙여 씀도 허용한다. '본용언+아/어+보조 용언' 구성과 관형사형+보조 용언(의존 명사+하다/시다) 구성의 경우는 붙여 쓰는 것이 허용된다. 앞말에 조사가 붙거나 앞 단어가 합성어이거나 파생어인 경우는 보조 용언을 앞말에 붙여 쓰지 않는다. 본용언이 합성어나 파생어라도 그 활용형이 2음절인 경우에는 붙여 쓴 말이 너무 긴 것은 아니므로 본용언과 보조 용언을 붙여 쓸 수 있다.	보여는준다'의 띄어쓰기는 어떻게 하나요?	안녕하세요? '보여는준다'의 띄어쓰기에 대한 질문에 답변하겠습니다. '보여 준다'는 본용언 '보이다'와 보조 용언 '준다'로 구성된 보조 용언 구성입니다. 본용언과 보조 용언은 띄어 쓰는 것이 원칙이나, '본용언+아/어+보조 용언' 구성의 경우에는 붙여 쓰는 것이 허용됩니다. 그러나 앞말에 조사가 붙어 있을 때는 보조 용언을 앞말에 붙여 쓰지 않습니다. 따라서 '보여는 준다'는 띄어 써야 합니다. 고맙습니다.	X
1	4	보조 동사 '준다' 띄어쓰기	한글 맞춤법 제5장 제3절 제47항	보조 용언은 띄어 쓰음을 원칙으로 하되, 경우에 따라 붙여 씀도 허용한다. '본용언+아/어+보조 용언' 구성과 관형사형+보조 용언(의존 명사+하다/시다) 구성의 경우는 붙여 쓰는 것이 허용된다. 앞말에 조사가 붙거나 앞 단어가 합성어이거나 파생어인 경우는 보조 용언을 앞말에 붙여 쓰지 않는다. 본용언이 합성어나 파생어라도 그 활용형이 2음절인 경우에는 붙여 쓴 말이 너무 긴 것은 아니므로 본용언과 보조 용언을 붙여 쓸 수 있다.	보조 용언 '준다'의 띄어쓰기가 궁금해요. '확인해 준다'는 붙여 써도 되나요?	안녕하세요? '확인해 준다'의 띄어쓰기에 대한 질문에 답변하겠습니다. '확인해 준다'는 본용언 '확인하다'와 보조 용언 '준다'로 구성된 보조 용언 구성입니다. 본용언과 보조 용언은 띄어 쓰는 것이 원칙이나, '본용언+아/어+보조 용언' 구성의 경우에는 붙여 쓰는 것이 허용됩니다. 그러나 본용언이 파생어인 경우는 본용언에 붙여 쓰면 전체가 너무 길어지므로 이때는 보조 용언을 앞말에 붙여 쓰지 않습니다. '확인하다'는 명사 '확인'에 동사 파생 접미사 '-하다'가 붙은 파생어입니다. 따라서 '확인해 준다'는 띄어 써야 합니다. 고맙습니다.	X
1	5	보조 동사 '준다' 띄어쓰기	한글 맞춤법 제5장 제3절 제47항	보조 용언은 띄어 쓰음을 원칙으로 하되, 경우에 따라 붙여 씀도 허용한다. '본용언+아/어+보조 용언' 구성과 관형사형+보조 용언(의존 명사+하다/시다) 구성의 경우는 붙여 쓰는 것이 허용된다. 앞말에 조사가 붙거나 앞 단어가 합성어이거나 파생어인 경우는 보조 용언을 앞말에 붙여 쓰지 않는다. 본용언이 합성어나 파생어라도 그 활용형이 2음절인 경우에는 붙여 쓴 말이 너무 긴 것은 아니므로 본용언과 보조 용언을 붙여 쓸 수 있다.	그럼 '구해 준다'도 반드시 띄어 써야 하나요?	앞선 질문에 이어서 답변드립니다. '구해 준다'는 본용언 '구하다'와 보조 용언 '준다'로 구성된 보조 용언 구성입니다. 본용언+아/어+보조 용언 구성에서 본용언이 파생어인 경우는 본용언에 붙여 쓰면 전체가 너무 길어지므로 보조 용언을 앞말에 붙여 쓰지 않습니다. 그러나 본용언의 활용형이 2음절인 경우는 붙여 쓴 말이 긴 것이 아니므로 붙여 쓸 수 있습니다. '구해 준다'의 본용언 '구하다'는 어근 '구'에 동사 파생 접미사 '-하다'가 결합된 파생어이지만 활용형 '구해'가 2음절이라 보조 용언을 붙여 쓸 수 있습니다. 따라서 '구해 준다'로 쓰는 것이 원칙이나 '구해준다'로 쓰는 것도 허용됩니다. 고맙습니다.	O

- 1) 유형 index : '이용자 질문 유형.xlsx'의 '유형 index' 열에서 확인함.
- 2) 예시 index : 유형별로 5개의 샘플을 작성하여 1에서 5까지 index를 붙임.
대화 쌍이 하나인 샘플을 3개(index 3~4), 대화 쌍이 둘인 샘플을 1개(index 1~2) 작성함. 꼬리 질문은 반드시 원 질문의 바로 아래 행에 작성함. 꼬리 질문은 하나만 작성함.
- 3) 이용자 질문 유형: '이용자 질문 유형.xlsx'의 '질문 유형' 열에서 확인함.
- 4) 출처 : 국어 지식 작성을 위한 참고 규정이나 자료를 제시함. 참고 규정을 제시할 때는 규정명과 구체적인 장, 절, 항의 번호를 함께 제시함. 예) '한글 맞춤법 제4장 제2절 제44항'. 참고 규정이나 자료를 최대한 제시하되, 참고 규정이나 자료를 찾을 수 없다면 해당 셀에는 'x(소문자)'로 제시함.
- 5) 국어 지식: 이용자 질문을 해결하기 위한 국어 지식을 작성함.
- 6) 이용자 질문: 국어생활 상담 자료의 실제 이용자 질문을 참고하여 다양한 유형으로 작성함.
- 7) 모범 답변: 이용자 질문에 대한 답변을 작성함. 질문 유형별로 제시된 작성 형식에 맞춰 일관성 있게 작성함.
- 8) 꼬리 질문 여부: 꼬리 질문이 있는 경우 꼬리 질문 행에만 'o(소문자)'로 표시함. 나머지는 모두 'x(소문자)'로 표시함.
- 9) 메모: 질문 유형과 지침에 반영이 필요한 내용이 있다면 메모에 기록함.

2.1.3.3. 이용자 질문 작성

이용자의 질문은 한 질문 유형에 대해서 5가지 형태를 하나씩 작성했다. 이용자 질문 유형은 국어생활 상담 자료의 질문 자료로부터 도출했다.⁸⁾ 실제 이용자의 질문은 어문 규정에 맞지 않을 수 있으나 본 말뭉치에서 이용자 질문은 어문 규정에 맞게 쓰며 질문 그 자체로 질문의 요지를 이해할 수 있도록 작성하는 것을 원칙으로 하였다.

<표 14> 이용자 질문 유형

	이용자 질문 유형	이용자 질문 유형 예시
①	어문 규범 설명형	<ul style="list-style-type: none"> 어문 규범에 대한 설명을 요청하는 질문 예) ‘협조해 주다.’는 왜 띄어 쓰는 건가요???
②	선택형	<ul style="list-style-type: none"> 두 개 이상의 예시 중 바른 표현을 묻는 질문 예) 안녕하세요. 나무 심어봐요./나무 심어 봐요. 띄어쓰기 어떤 게 맞는 표현인가요?
③	교정형	<ul style="list-style-type: none"> 제시한 표현 교정을 요청하는 질문 예) ‘믿어의심치않다’의 띄어쓰기는 어떻게 하나요?
④	확인형	<ul style="list-style-type: none"> 제시한 표현이 맞는지 확인을 요청하는 질문 예) 보조 용언 ‘있다’의 띄어쓰기가 궁금해요. ‘열려 있다’는 붙여 쓰나요?
⑤	꼬리 질문형	<ul style="list-style-type: none"> 앞선 답변에 대한 추가 설명을 요구하는 질문 예) 근데, ‘넓어졌어요’는 왜 붙여 써요?

2.1.3.4. 출처 작성

국어 지식의 출처를 작성하기 위해 사용할 수 있는 규정과 사전, 자료는 아래와 같다. 국어 지식의 출처를 작성할 때는 표기 변동이 없도록 아래 목록을 복사해 활용하도록 했다.

- 1) 한글 맞춤법
- 2) 표준어 규정 제1부 표준어 사정
- 3) 표준어 규정 제2부 표준 발음법
- 4) 문장 부호 해설
- 5) 국어의 로마자 표기법

8) 제공된 국어생활 상담 자료 중 상담 질문이 포함된 ‘온라인가나다’와 ‘우리말365’를 주로 참조했다.

- 6) 외래어 표기법
- 7) 국립국어원 다듬은 말: 순화어는 ‘국립국어원 다듬은 말’만 활용함.
- 8) 표준국어대사전
- 9) 우리말샘
- 10) 표준 국어 문법, 한국어 표준 문법

2.1.3.5. 국어 지식 작성

이용자 질문에 답변을 작성하기 위해 사용되는 모든 국어 지식을 작성하였다. 국어 지식은 아래와 같은 형식으로 작성하였다. 한 질문에 여러 개의 국어 지식이 필요한 경우, 여러 개를 작성할 수 있도록 하였다.

<표 15> 국어 지식 작성 예시

구분	내용	설명
이용자 질문 유형	동사 '못쓰다' 띄어쓰기	
출처	한글 맞춤법 제5장 제2절 제46항, 표준국어대사전	두 개 이상의 국어 지식을 사용할 경우 출처는 쉼표로 구분하며 규정을 앞에, 사전을 뒤에 씀.
국어 지식	단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다. '못쓰다'는 '(((‘못쓰게’ 꼴로 쓰여))) 얼굴이나 몸이 축나다', '(((주로 ‘-으면’, ‘-어서’와 함께 쓰여))) 옳지 않다. 또는 바람직한 상태가 아니다.'의 의미로 쓰인다.	두 개 이상의 국어 지식을 활용할 경우 앞선 국어 지식 뒤에 공백을 하나 넣고 붙여서 씀.
이용자 질문	'이 망치는 더 못 쓴다.'는 '더못 쓴다'로 붙여 쓰나요?	
모범 답변	안녕하세요. ‘더 못 쓴다’의 띄어쓰기에 대해 답변드립니다. ‘못쓴다’는 ‘얼굴이 상하다’, ‘행동이 바람직하지 않다’, ‘옳지 않다’와 같은 의미로 쓰일 때만 붙여 씁니다. 또한, 한글 맞춤법 제46항에서는 단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다는 규정이 있으나 표현이 의미적으로 자연스럽게 하나로 이어질 때에만 붙여 쓸 수 있습니다. 그러므로 '다 못 쓰다'로 띄어 쓰는 것이 바릅니다. 고맙습니다.	

국어 지식을 제공할 수 없는 경우는 크게 두 가지로 나눌 수 있는데, 1) 어원이나 의미 차이, 순화어가 없는 단어 등과 같이 국어 지식으로 삼고 있는 자료에 정확히 규정되어 있지 않은 경우, 2) 건의 사항과 같은 이용자의 의견과 관련된 경

우이다. 이러한 경우에는 별도로 모아 정리했다.

<표 16> 국어 지식 작성 형식

국어 지식 종류	작성 형식 예시
사전	<ul style="list-style-type: none"> • 여러 의미가 있는 경우 쉽표로 구분하여 제시한다. • 다 제시하기에 의미가 많은 경우, 이용자 질문에 해당하는 의미만 제시한다. • 뜻풀이 외에 이용자 질문에 필요한 정보가 있다면 더 제시할 수 있다. <p><형식 예시></p> <ul style="list-style-type: none"> • ‘[단어]’은/는 ‘[뜻풀이1]’, ‘[뜻풀이2]’(이)라는 의미로 쓰인다. • ‘[단어]’은/는 ‘[뜻풀이]’(이)라는 뜻이다. • ‘[단어]’은/는 ‘[뜻풀이]’(이)라는 의미이다. <p>예) ‘메우다’는 ‘뚫려 있거나 비어 있는 곳을 막거나 채우다.’, ‘어떤 장소를 가득 채우다.’, ‘부족하거나 모자라는 것을 채우다.’, ‘시간을 적당히 또는 그럭저럭 보내다.’라는 의미로 쓰인다.</p> <p>예) ‘드러나다’는 ‘알려지지 않은 사실이 널리 밝혀지다.’라는 의미이다. ‘드러나다’는 ‘들다’와 ‘나다’가 결합한 말이지만 ‘들어나다’로 적지 않고 ‘드러나다’로 적는다. 이는 두 개의 용언이 어울려 한 개의 용언이 될 적에, 앞말의 본뜻이 유지되고 있는 것은 그 원형을 밝혀 적고, 그 본뜻에서 멀어진 것은 밝혀 적지 않는다는 규정에 따른 것이다.</p>
한글 맞춤법 외래어 표기법 국어의 로마자 표기법 문장 부호 해설	<ul style="list-style-type: none"> • 이용자 질문에 답변을 작성하기 위한 모든 규정을 그대로 적는다. <p>예) 단위를 나타내는 의존 명사는 띄어 쓴다. 다만, 수 관형사 뒤에 단위 명사가 붙어서 차례를 나타내는 경우에는 앞말과 붙여 쓸 수 있다.</p> <p>예) 받침에는 ‘ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ’만을 쓰는 것이 원칙이다. 짧은 모음 다음의 어말 무성 파열음 [p]은 받침 ‘ㅂ’으로 적는다.</p> <p>예) ‘ㄱ, ㄷ, ㅂ’은 모음 앞에서는 ‘g, d, b’로, 자음 앞이나 어말에서는 ‘k, t, p’로 적는다.</p> <p>예) 줄임표는 앞말에 붙여 쓰는 것이 원칙이며 문장의 앞에 붙여 쓰지 않는다. 문장이나 글의 일부를 생략할 때는 앞뒤 쉽표나 마침표 없이 띄어 쓰며 문장이 끝나면 마침표나 느낌표를 쓰는 것이 원칙이다. 가운데 여섯 점을 찍는 것이 원칙이나 아래에 여섯 점을 찍거나 세 점으로 찍을 수 있다.</p>
국립국어원 다듬은 말	<ul style="list-style-type: none"> • 순화어 정보는 ‘국립국어원 다듬은 말’의 내용만 활용한다. <p><형식 예시></p> <ul style="list-style-type: none"> • ‘[단어]’의 원어를 ‘[원어]’(으)로, 다듬은 말을 ‘[다듬은 말]’(으)로 제시하고 있다.

	<p>예) 'SNS'의 원어를 'social network service(영)'로, 다듬은 말을 '누리소통망, 누리 소통망 서비스, 사회 관계망, 사회 관계망 서비스'로 제시하고 있다.</p>
국어 지식을 제공할 수 없는 경우	<p>1) 국어 지식 자료에 정확히 규정되어 있지 않은 경우</p> <p>- 정확히 규정되어 있지 않음을 언급하고 참고 정보가 있다면 '다만'을 쓴 후에 제시함. 참고할 정보가 없는 경우에는 끝인사로 끝냄.</p> <p>예) 안녕하세요. 특정 어휘의 어원에 관한 질문에 답변하겠습니다. 우리말샘의 '역사 정보'란에 '생때같다'의 역사 및 어원 정보가 실려 있지 않습니다. 다만, 사전의 형태 분석을 통해 '생때'와 '같다'가 결합된 말임을 짐작할 수 있습니다. 감사합니다.</p> <p>예) 안녕하세요? 알고리즘 순화어에 관해 말씀드리겠습니다. 현재 알고리즘의 순화어는 따로 정해져 있지 않습니다. 다만, 생명과학 분야에서 사용되는 '지네틱알고리즘'을 '유전연산방식'으로 다듬은 바가 있어 이를 참고하실 수 있습니다. 국립국어원의 온용어, 우리말샘을 참고하시길 바랍니다. 감사합니다."</p> <p>2) 이용자의 의견과 관련된 경우</p> <p>- 이용자의 의견을 언급하고 관련 부서에 전달하겠다고 답변함.</p> <p>예) 안녕하세요. 사전의 뜻풀이에 관한 건의 사항에 답변하겠습니다. [이용자 의견 언급] 제안해 주신 사항을 관련 부서에 전달해 두겠습니다. 사전 뜻풀이에 깊은 관심을 가지고 건의해 주셔서 감사합니다.</p>

2.1.3.6. 모범 답변 작성

모범 답변은 아래 제시된 형식을 참고하여 일관된 형식으로 작성하였다. 모범 답변은 어문 규범 전문가가 아닌 일반 이용자를 고려하여 한 단락 길이에서 완결된 의미를 전달하도록 하였다. 일관된 답변을 제시하도록 국어 생활 상담 자료를 기반으로 답변 형식과 구성 요소를 정의해 활용하도록 했다.

<표 17> 모범 답변의 형식과 구성 요소

구성 요소	역할	예시
인사	대화의 시작	안녕하세요? 안녕하십니까?
질문 요약	어떤 주제의 질문인지 요약	[질문 내용] 질문에 답변하겠습니다.

	하여 질문의 주제를 명확히 함.	[질문 내용]에 대해서 답변드리겠습니다. [질문 내용]에 대한 질문에 답변하겠습니다.
규범 설명	해당 질문의 핵심 규범이나 원칙을 설명함.	~은/는 ~하는 것이 바랍니다./적절합니다. ~(으)로 쓰는 것을 권장합니다. 예) '1만 원대'라고 띄어 쓰는 것이 바랍니다.
문법 설명	판단의 근거를 '국어 지식' (문법, 맞춤법, 어휘 등)에 근거하여 설명함.	~는 ~의미를 가진 ~입니다. 예) '원'은 화폐 단위를 뜻하는 의존 명사입니다.
예시 제시	실제 예를 통해 일반화된 규범을 구체적으로 제시하여 독자의 이해 보완	예를 들어 ~와 같이 씁니다. 아래와 같이 씁니다. 예) [예시]
예외 안내	특수한 상황이나 예외적 사용을 안내하여 규범의 유연성 제시	다만, ~의 경우에는 ~할 수 있습니다. 예) 계약서 등 특수 상황에서는 붙여 쓰기도 합니다.
맺음말	답변의 마무리	참고하시기 바랍니다. 고맙습니다.

꼬리 질문을 작성할 때는 이어지는 질문이므로 '안녕하세요.'와 같은 인사말과 질문 요약을 생략하고 '앞선 질문에서 말씀드린 바와 같이', '앞선 질문에 이어서 답변드립니다.' 등과 같이 앞선 질문과 이어진다는 것을 나타내는 표현을 사용하도록 했다.

<표 18> 꼬리 질문 예시

질문 유형	질문	대답
질문	'한번'은 언제 붙여 쓰고 언제 띄어 써야 하나요?	<ul style="list-style-type: none"> • [인사]. [질문 내용]에 대한 질문에 답변하겠습니다. [질문에 대한 대답]. 감사합니다. <p>예) 안녕하세요. '한번'의 띄어쓰기에 관한 질문에 답변하겠습니다. '한 번', '한번'은 맥락에 따라 띄어쓰기가 달라집니다. 1회를 말하는 것이라면 '한 번'으로 띄어 쓰고, '지난 어느 때나 기회', '어떤 일을 시험 삼아 시도함을 나타내는 말', '기회 있는 어떤 때에', '어떤 행동이나 상태를 강조하는 뜻을 나타내는 말', '일단 한 차례'라는 뜻으로 쓸 때는 '한번'으로 붙여 씁니다. 이에 따라, '한번 해 보다'의 경우 붙여 쓰고 '떡볶이를 딱 한 번 먹어 보았다'의 경우 띄어 씁니다. 감사합니다.</p>
꼬리 질문	그럼, "한번만 용서해 주세요"는 붙여 써야 하나요?	<ul style="list-style-type: none"> • 앞선 질문에서 말씀드린 바와 같이 • 앞선 질문에 이어 답변드립니다. • 이어서 답변드립니다. • 추가 질문에 이어서 답변드립니다. • 앞서 말씀드린 것처럼 <p>예) 앞선 질문에서 말씀드린 바와 같이 '한번만 용서해 주세요'의 경우 횟수의 의미가 드러나므로, '한 번만 용서해 주세요'로 띄어 쓰는 것이 적절합니다. 감사합니다.</p>

꼬리 질문은 아래와 같이 4개의 유형으로 분류했다. 그 외에 질문 유형에 따라 다른 유형이 있을 수 있으므로 국어생활 상담 자료를 확인하여 새로운 유형을 추가하여 작성할 수 있도록 했다.

<표 19> 꼬리 질문 유형

꼬리 질문 유형	설명	질문 예시
분할 질문형	본래 하나의 답변으로 설명되어 있는데, 답변에 따라 추가 질문이 가능한 경우	<p>Q1: '한 개'처럼 수량을 나타내는 말은 왜 띄어 쓰는 건가요? Q2: 그럼 '두 마리', '세 송이' 같은 수량 명사도 띄어 쓰는 게 맞나요?</p>
추가 질문형 (예시 변경)	질문 의도는 동일하나 다른 예시를 이용해 묻는 경우	<p>Q1: 한글 맞춤법에 따르면 '카페'가 표준어이고 '까페'는 틀린 표기라고 하던데, 왜 그런 건가요? Q2: 그렇다면 '카페'처럼 'ㄱ, ㅌ, ㅍ' 대신 'ㅋ, ㅌ, ㅍ'을 쓰는 외래어 표기법에는 어떤 원칙들이 더 있나요? 혹시 예외적인 경우는 없나요?</p>
질문 보충형	"고유명사는 붙여쓰나요? 띄어쓰나요?"와 같이 막연한 질문에 대하여 답변을 얻은 후, 구체적인 예시를 요청하는 추가 질문	그러면, "한번만 용서해 주세요"는 붙여 써야 하나요?
답변 보충형	답변에 부족한 부분이 있다고 판단되거나, 추가 답변을 유도할 수 있는 질문	<p>Q1: 한 단어일 경우 띄어 쓰지 않나요? Q2: 그러니까 '한번'처럼 한 단어일 경우 띄어 쓰지 않는 게 원칙이라는 말씀이시죠?</p>

설계 말뭉치는 AI를 활용해 추가 질문과 답변을 생성하는 데 사용된다. 그런데, 초기에는 AI가 생성한 답변에 일관성이 떨어지는 경우가 자주 발견됐다. 설명이 지나치게 긴 경우, 예시가 있는 답변과 없는 답변, 국어 지식 제시 여부 등이 차이가 났고, 질문이 반말인 경우 답변도 반말로 생성하기도 했다. 이러한 문제는 AI의 기술적 특성인 면도 있었지만, 연구진이 상담 답변에 익숙해지면서 형성된 암묵지가 실제 프롬프트에는 반영되지 않았던 면이 있었다. 초기 답변 생성 프롬프트에서는 설계 말뭉치에 작성된 예시 모범 답변 5개를 직접 참고하도록 하고 쉬운 용어의 사용, 명확한 답변 등 작업자용 지침을 그대로 사용했다.

<표 20> 초기 답변 생성용 프롬프트

<p>[example]</p> <p>Question: “깨끗하다”와 “깨끗 하다”는 의미 차이가 있나요? 둘 다 맞는 말인가요?</p> <p>Answer: 안녕하세요? 형용사를 만드는 접미사 '-하다'의 띄어쓰기에 관한 질문에 답변하겠습니다. '깨끗하다'는 어근 '깨끗'에 접미사 '-하다'가 결합하여 만들어진 파생어입니다. 접미사는 앞말에 붙여 쓰는 것이 원칙이므로, '깨끗하다'로 적는 것이 적절한 표현입니다. '똑똑하다, 강렬하다, 조용하다'의 경우에도 이와 동일하게 붙여 씁니다. 감사합니다.</p> <p>###Task3###</p> <p>[Question]에 대해 [Knowledge]를 이용해서 생성하세요.</p> <p>이용자가 잘 이해할 수 있도록 어려운 문법 용어 사용은 피하고 명확하게 답변을 제공하세요.</p> <p>답변을 생성할 때는 [example]의 양식을 따르세요.</p>

답변 작성 지침은 연구진이나 작업자에게는 충분히 형식화됐다고 여겨졌으나 짧은 글로 어문 규범이나 국어 지식을 명확히 전달하기에는 경험적인 부분에 의존하는 부분이 많았다. 이런 점에 착안해 먼저 모범 답변을 구성하는 요소들을 도출하고 순서와 길이, 문체까지 예시를 포함해 지침으로 구성하고 이에 맞춰 설계 말뭉치의 답변을 작성하면서 보완했다.

<표 21> 모범 답변 구성 및 작성 방식

항목	작성 방식
문장 구성 구조	<ul style="list-style-type: none"> 인사 → 질문 요약 → 규범 제시 → 설명 → 예시 → 예외 및 보충 → 맺음말

답변 길이	<ul style="list-style-type: none"> 한 단락 수준
문체	<ul style="list-style-type: none"> 존댓말, 평이한 문장 사용 (지나친 전문용어 지양, 예시로 보완) 용어나 설명의 수준은 학교 문법 수준으로 작성 인용할 때는 작은따옴표를 씀.
내용의 근거	<ul style="list-style-type: none"> 표준국어대사전, 한글 맞춤법, 표준 발음법 등의 ‘국어 지식’ 명시
예시 활용	<ul style="list-style-type: none"> 원칙 설명 후 실제 사용 예시 포함(잘못된 예보다는 바른 예 위주로)
자주 쓰는 표현 유형화	<ul style="list-style-type: none"> ‘~는 ~입니다.’ ‘~는 ~의 뜻을 더하는 접미사입니다.’ ‘따라서 ~라고 써야 바랍니다.’
답변이 어려운 경우	<ul style="list-style-type: none"> 이용자의 의견이나 감정적 질문에는 “답변드리기 어렵습니다.”로 응대

모범 답변 구성과 작성 방식을 수립한 후에는 위 지침을 답변 작성 규칙으로 대입해 답변 생성 프롬프트에 적용하였으며, 그 결과 답변 일관성이 현저히 높아졌다. 더불어 상담 답변 시 권고되는 표현을 답변 구성 부분별로 제시하여 설계 말뭉치의 모범 답변에 더 근접한 답변을 생성할 수 있게 했다.

<표 22> 모범 답변 표현 예시

상황	문장 예시
인사	<ul style="list-style-type: none"> 안녕하세요? 안녕하세요? [질문 주제]에 관한 질문에 답변드리겠습니다.
규범 제시	<ul style="list-style-type: none"> ~는 ~하는 것이 바랍니다. ~로 표기하는 것이 적절합니다. ~는 [이유]로 인해 ~하는 것이 바랍니다.
근거 설명	<ul style="list-style-type: none"> ~는 [의존 명사/접미사/조사]이므로 ~하게 처리합니다. ‘~’는 [문법적 기능/역할]을 가지므로 ~하게 표기합니다.
예시 제시	<ul style="list-style-type: none"> 예) ~ 예를 들어 ~처럼 쓸 수 있습니다. 예를 들어, ‘~’과 같이 씁니다.
예외 언급	<ul style="list-style-type: none"> 다만, ~하는 경우에는 ~할 수도 있습니다. 다만, ~한 특수 상황에서는 ~하게 쓰는 것도 허용됩니다.
맺음말	<ul style="list-style-type: none"> 고맙습니다. 참고하시기 바랍니다.

모범 답변 작성 지침을 구조화하고 답변 프롬프트에 적용한 것은 AI 답변 생성에서도 좋은 결과를 얻었지만, 향후 모범 답변이 변경되거나 개선되더라도 AI가 반영할 수 있도록 데이터적인 연관성을 확립했다는 점에서 더 효용 가치가 있다.

<표 23> 모범 답변 작성 규칙을 적용한 답변 생성용 프롬프트

당신은 “한국어 교육 및 한국어 어문 규범 전문가”입니다.

[question]에 대해서 [한국어 어문 규범]를 근거로 하여 답변을 작성하세요.

[한국어 어문 규범]이 없으면, 알고 있는 "한국어 어문 규범"을 활용해 답하고 그 근거 조항을 답변에 포함하세요.

답변은 [답변 작성 규칙]을 준수해야 합니다.

[답변 작성 규칙]

<규칙1>문장 구성 구조 : [인사] → [질문 요약] → [규범 제시] → [설명] → [예시] → [예외 및 보충] → [맺음말].

<규칙2>답변 길이: 650자 내외.

<규칙3>문체: 존댓말 사용. 쉬운 문장, 예시 사용.

<규칙4>문법 지식 수준: 학교 문법 수준, [한국어 어문 규범]를 근거로 사용.

<규칙5>인용: 일관된 작은따옴표를 사용할 것.

<규칙6>예시 활용: 원칙을 설명한 후 올바른 사용 예시를 제시할 것.

<규칙7>자주 쓰는 표현 유형 :

[인사] : ‘안녕하세요? [질문 주제]에 관한 질문에 답변드리겠습니다.’

[규범 제시] : ‘~는 ~하는 것이 바릅니다.’, ‘~로 표기하는 것이 적절합니다.’, ‘~는 [이유]로 인해 ~하는 것이 바릅니다.’

[예시 제시] : ‘예: ~’, ‘예를 들어 처럼 쓸 수 있습니다.’, ‘예를 들어, “과 같이 씁니다.’

[예외 언급] : ‘다만, ~하는 경우에는 ~할 수도 있습니다.’, ‘다만, ~한 특수 상황에서는 ~하게 쓰는 것도 허용됩니다.’

[맺음말] : ‘고맙습니다.’, ‘참고하시기 바랍니다.’

<규칙8>한국어 어문 규범 및 한국어 사용과 무관한 질문:

한국어 어문 규범 및 한국어 사용과 무관한 질문은 답변하지 않고 아래 문장으로 대신합니다.

‘죄송하지만, 이 질문에는 답변할 수 없습니다. 한국어 어문 규범과 한국어 사용에 대해서 질문해 주세요.’

[question]

‘맏고’와 ‘맏게’의 받침 발음이 같은가요?

2.2. 지시문 기반 첨삭 지원 말뭉치 구축

지시문 기반 첨삭 지원 말뭉치의 구축은 설계 말뭉치의 질문 유형, 국어 지식, 예시 질문과 답변 쌍에 부합하는 다양한 질문과 그에 부합하는 모범 답변 쌍을 약 10배로 확대하는 과정이다. 이러한 질문-답변 쌍의 확대 구축을 위해 설계 말뭉치를 입력으로 받아 AI 모델이 생성하도록 하는 데이터 증강 방식을 적용했다. 그리고 증강 생성된 질문-답변 쌍은 국어학, 언어학, 글쓰기 전공 작업자가 중복성, 질문-답변 오류의 수정, 추가, 삭제하는 가공과 검수를 거쳐 완성하였다. 구축 전 과정은 지시문 기반 첨삭 지원 말뭉치 특성을 고려한 전용 구축 도구를 개발해 진행했다.



<그림 10> 지시문 기반 첨삭 지원 말뭉치 구축 과정

2.2.1. 설계 말뭉치 전처리 및 질문-답변 증강

설계 말뭉치 전처리란 AI가 설계 말뭉치에 부합하는 질문-답변 쌍을 생성해 반환하도록 설계 말뭉치 데이터를 프롬프트로 변환하는 일련의 데이터 처리 과정이다. 인공지능으로 질문-답변 쌍을 증강하는 과정은 일견 단순해 보이지만 설계 자료가 정확해야 하고, 설계 자료에 포함된 국어 지식 내에서 일관된 질문과 답변이 생성되어야 한다. 또한 기본적으로 생성된 각각의 질문에 대해 올바른 답변이 쌍을 이루도록 해야 한다.

그런데, 시범 구축 단계에서 각 질문 유형별로 45~70개 사이의 질문-답변을 생성해 보니 질문 유형과 이에 연계된 국어 지식과 맞지 않는 질문이 생성되거나 질문-답변 쌍이 어긋나거나(질문에 대한 답이 아닌 경우) 제시된 국어 지식을 준용하지 않는 경우가 많았다. 따라서 전처리 과정에서는 인공지능이 설계 자료를 바탕으로 지침에 충실하게 질문과 답변을 생성하도록 여러 차례 프롬프트를 조정해 초기 추가 질문-답변 쌍을 생성하게 하고 이를 확인하는 데 집중했다.

질문-답변의 생성 결과는 인공지능 언어모델의 성능과 특징에 따라 크게 달라질 수 있다. 질문-답변 생성에 사용할 인공지능 언어모델을 선정하기 위하여 Google Gemini 2.5 Pro(2025년 6월 출시), OpenAI GPT-4o(4.1)⁹⁾를 중심으로 비교 테스트를 진행하였고, GPT-4o API로 결정했다. 비교 테스트는 설계 말뭉치에 대해 동일한 질문-답변 생성 프롬프트를 통해 출력된 질문-답변 쌍들을 <표 11>의 생성된 질문-답변 평가 지표에 따라 검수하고 그 결과를 계량해 비교했다.

- 질문-답변 생성 프롬프트(설계 말뭉치 검증용)

설계 단계에서는 완성된 설계 말뭉치를 대상으로 질문 유형과 국어 지식에 관련된 질문과 모범 답변이 제대로 생성되는지 검증을 진행했다. 설계 말뭉치 검증용 프롬프트는 “2024년 글쓰기 첨삭 지원을 위한 지시문 기반 생성(인스트럭션) 말뭉치 구축 연구” 결과의 프롬프트를 참고해 작성했다.

9) GPT-4o는 공식적으로 2024년 10월까지의 데이터를 마지막 사전학습 데이터로 사용했다고 밝혀져 있으며, GPT-4.1은 2025년 4월 15일 출시로 더 개선된 성능으로 소개되었으나 GPT-4.1의 개선사항이 GPT-4o로 통합되었다고 소개되었다. Note that GPT-4.1 will only be available via the API. In ChatGPT, many of the improvements in instruction following, coding, and intelligence have been gradually incorporated into the latest version [GPT-4o](https://openai.com/index/gpt-4-1/) (opens in a new window), and we will continue to incorporate more with future releases. (<https://openai.com/index/gpt-4-1/>)

<표 24> 초기 질문-답변 생성 프롬프트

###Instruction###

당신은 한국어 어문 규범 전문가입니다.

당신의 임무는 한국어 어문 규범 및 한국어 사용과 관련된 이용자의 다양한 질문에 답변하는 것입니다.

아래 Task1, Task2, Task3를 순차적으로 수행하세요.

###Task1###

아래 [Question]에서 이용자의 질문을 확인하고 한국어 어문 규범과 한국어 사용에 대한 질문인지 확인하세요.

한국어 어문 규범이나 한국 사용과 관련된 질문이 아닌 경우에는 답변을 제공하지 말고 아래와 같이 답변하세요.

죄송하지만, 이 질문에는 답변할 수 없습니다. 한국어 어문 규범과 한국어 사용에 대해서 질문해 주세요.

[Question]

'로맨틱하다' 띄어써야 하나요?

###Task2###

아래 [Question]이 한국어 어문 규범과 한국어 사용에 대한 질문인 경우 아래 [Knowledge]를 이용하여 질문에 답변하세요.

이용자가 잘 이해할 수 있도록 어려운 문법 용어 사용은 피하고 명확하게 답변을 제공하세요. 답변을 생성할 때는 [example]의 양식을 따르세요.

[Knowledge]

'-하다'는 일부 명사나 의성 의태어, 어근, 의존 명사 뒤에 붙어 동사나 형용사를 만드는 접미사이다. 형용사 뒤에 붙어 앞말의 상태를 강조하는 보조형용사 '하다'와는 구별된다.

[example]

Question: “깨끗하다”와 “깨끗 하다”는 의미 차이가 있나요? 둘 다 맞는 말인가요?

Answer: 안녕하세요? 형용사를 만드는 접미사 '-하다'의 띄어쓰기에 관한 질문에 답변하겠습니다. '깨끗하다'는 어근 '깨끗'에 접미사 '-하다'가 결합하여 만들어진 파생어입니다. 접미사는 앞말에 붙여 쓰는 것이 원칙이므로, '깨끗하다'로 적는 것이 적절한 표현입니다. '똑똑하다, 강렬하다, 조용하다'의 경우에도 이와 동일하게 붙여 씁니다. 감사합니다.

###Task3###

위의 [Question]과 비슷한 유형의 이용자 질문을 50개 생성하세요.

생성한 [Question]의 답변도 [Knowledge]를 이용해서 생성하세요.

이용자가 잘 이해할 수 있도록 어려운 문법 용어 사용은 피하고 명확하게 답변을 제공하세요. 답변을 생성할 때는 [example]의 양식을 따르세요.

- 1차 증강 프롬프트 개선

<표 25> 1차 질문-답변 증강 생성 프롬프트

<p>당신은 “한국어 교육 및 한국어 사용에 대한 질문 생성 전문가”입니다.</p> <p>한국어 문법 및 어문 규범 오류를 중심으로, 실제 사람이 질문하는 것처럼 자연스럽게 다양한 질문을 생성하는 것이 당신의 역할입니다.</p> <p>Task 1: [오류 유형] 오류 유형 쓰기</p> <p>위의 [오류 유형]에 대해, 아래 [Question types] 중 하나를 적용하여 질문을 생성하세요.</p> <p>총 60개의 질문을 생성하며, 질문 생성 시 제시된 [Rules]을 지키세요.</p> <p>[Question types]¹⁰⁾</p> <ol style="list-style-type: none"> 1) 어문 규범에 대한 설명을 요청하는 질문 예) ‘협조해 주다.’는 왜 띄어 쓰는 건가요??? 2) 두 개 이상의 예시 중 바른 표현을 묻는 질문 예) 안녕하세요. 나무 심어봐요./나무 심어 봐요. 띄어쓰기 어떻게 맞는 표현인가요? 3) 제시한 표현 교정을 요청하는 질문 예) ‘믿어의심치않다’의 띄어쓰기는 어떻게 하나요? 4) 제시한 표현이 맞는지 확인을 요청하는 질문 예) 보조 용언 ‘있다’의 띄어쓰기가 궁금해요. ‘열려 있다’는 붙여 쓰나요? <p>[Rules]</p> <p>모든 질문은 존댓말을 사용해야 합니다.</p> <p>질문자는 나이, 직업, 말투, 상황 등이 다양¹¹⁾해 보이도록 작성하되, 직접적으로 노출하지 마세요.</p> <p>[example]의 Question 표현은 참고만 하고 그대로 내용이나 단어를 복제하지 마세요.</p> <p>Task 2: Task 1에서 생성한 질문을 검토하고, 한국어 어문 규범과 관련된 질문인 경우 [Knowledge]를 근거로 하여 답변을 작성한 뒤, 질문과 1:1로 매칭¹²⁾하여 최종 Q&A 세트를 출력하세요.</p> <ul style="list-style-type: none"> - 답변 작성 규칙:¹³⁾ <p>문장 구성 구조: 인사 → 질문 요약 → 규범 제시 → 설명 → 예시 → 예외 및 보충 → 맺음말</p> <ul style="list-style-type: none"> - 답변 길이: 한 단락 수준 - 문체: 존댓말, 평이한 문장, 지나친 전문용어 지양, 예시로 보완 - 설명 수준: 학교 문법 수준 - 인용: 작은따옴표(' ') 사용 - 예시 활용: 원칙 설명 후 올바른 사용 예시 포함 (바른 예 위주) - 자주 쓰는 표현 유형: <p>인사 : ‘안녕하세요? [질문 주제]에 관한 질문에 답변드리겠습니다.’</p> <p>규범 제시 : ‘~는 ~하는 것이 바릅니다.’, ‘~로 표기하는 것이 적절합니다.’, ‘~는 [이유]로 인해 ~하는 것이 바릅니다.’</p> <p>예시 제시 : ‘예: ~’, ‘예를 들어 처럼 쓸 수 있습니다.’, ‘예를 들어, “과 같이 씁니다.”</p> <p>예외 언급 : ‘다만 ~하는 경우에는 ~할 수도 있습니다.’, ‘다만 ~한 특수 상황에서는 ~하게 쓰는 것도 허용됩니다.’</p> <p>맺음말 : ‘고맙습니다.’, ‘참고하시기 바랍니다.’</p>

비규범 질문 처리: 한국어 어문 규범 및 한국어 사용과 무관한 질문은 답변하지 않고 아래 문장으로 대신합니다.
죄송하지만, 이 질문에는 답변할 수 없습니다. 한국어 어문 규범과 한국어 사용에 대해서 질문해 주세요.

초기 질문-답변 프롬프트는 10개 이내의 질문-답변 쌍을 생성하는 경우에는 비교적 정확하고 유의미한 질문과 답변을 생성했으나, 20개 이상의 질문-답변을 생성할 때에는 몇 가지 문제가 발생했다. 첫째, 질문에 대한 반복성이 높아졌다. 질문 유형이 품사나 문법 현상과 같은 범주성을 갖는 경우에는 어느 정도 다양한 질문을 생성했지만, 구체적인 예시가 포함된 경우에는 동일한 질문을 반복하는 경향이 많았다. 둘째, 질문 자체가 설명형, 요구형인 경우가 종종 발생했다. 즉, 답변을 위해 질문을 구체화하는 경우가 포함되었다. 셋째, 생성한 질문과 답변이 하나의 쌍을 이루지 못하는 경우가 상당수 포함되었다. 또한, 질문과 답변의 어투가 제각각인 경우도 빈번하게 발생했다. 이러한 문제를 보완하기 위해 다음과 같은 사항들을 반영해 프롬프트를 변경했다.

- 1) 상세 질문 유형과 예시를 함께 제공 : 질문 내용과 유형에 의한 다양성을 유도
- 2) 질문자 페르소나 추가 : 질문 문체를 다양하게 생성하도록 유도
- 3) 질문과 답변 매칭 요건 추가 : 질문-답변 불일치 해소를 위한 제약사항 명시
- 4) 답변 작성 규칙 추가

<표 26> 1차 개선 프롬프트 모델별 생성 답변 부적합률

질문 유형 번호	평가자	GPT-4o			Gemini 2.5 pro		
		적합 질문-답변쌍	부적합 질문-답변쌍	부적합률	적합 질문-답변쌍	부적합 질문-답변쌍	부적합률
103	설계1	24	21	46.67%	19	26	57.78%
	설계2	30	15	33.33%	19	26	57.78%
	구축1	35	10	22.22%	30	15	33.33%
114	설계1	31	14	31.11%	24	21	46.67%
	설계2	34	11	24.44%	25	20	44.44%

- 10) 상세 질문과 예시를 함께 제공함.
- 11) 질문자 페르소나 추가함.
- 12) 질문-답변 매칭 요건 추가함.
- 13) 답변 규칙 추가 : 모범 답변 구조를 구체화하여 구성 요소별 작성 순서, 내용, 표현을 명시함.

	구축1	29	16	35.56%	31	14	31.11%
126	설계1	27	18	40.00%	30	15	33.33%
	설계2	30	15	33.33%	33	12	26.67%
	구축1	27	18	40.00%	17	28	62.22%
138	설계1	27	18	40.00%	22	23	51.11%
	설계2	22	23	51.11%	24	21	46.67%
	구축1	31	14	31.11%	17	28	62.22%
162	설계1	33	12	26.67%	35	10	22.22%
	설계2	35	10	22.22%	28	17	37.78%
	구축1	35	10	22.22%	32	13	28.89%
합계(평균)		450	225	33.33%	386	289	42.81%

- 2차 증강 프롬프트 개선

1차 프롬프트 개선으로 질문-응답 쌍의 일관성이 다소 개선되었으나 질문 유형의 반복, 질문-답변의 불일치 현상은 반복되었다. 또한, 응답이 지연되는 경우 이전 응답과 현저히 다른 질문-답변을 출력하거나 거의 동일한 결과를 출력하는 등에 대량의 질문-답변 쌍을 생성하기에 불안정한 경우도 새롭게 발견되었다. 이 문제를 해소하기 위해 질문 생성과 답변 생성을 각각의 프롬프트로 나누어 처리하도록 2차 프롬프트 개선을 진행했다. 우선, 프롬프트를 분리하여 질문용 프롬프트의 역할(페르소나)과 답변용 프롬프트의 역할을 다르게 부여했다. 질문 생성용 프롬프트는 ‘질문하는 학생’ 역할을 부여하고 세부 질문 유형을 질문 주제로 제공해 실제 질문 상황의 문맥 정보(context)를 제시했다. 또한 질문 주제와 생성할 질문 유형과 예시와의 관계도 제시해 문맥 정보를 보다 풍부하게 했다.

<표 27> 질문 생성용 프롬프트

당신은 “한국어 교육 및 한국어 사용에 대한 관심이 많은 학생”입니다. 한국어 문법 및 어문 규범 오류를 중심으로, 실제 사람이 질문하는 것처럼 자연스럽게 다양한 질문을 생성하는 것이 당신의 역할입니다.
Task 1: [질문 주제] 한글 자모음 수

위의 [질문 주제]에 대해, 아래 [Question types] 중 하나를 적용하여 질문을 생성하세요.
총 60개의 질문을 생성하며, 질문 생성 시 제시된 [Rules]을 지키세요.

[Question types]

1) 어문 규범에 대한 설명을 요청하는 질문

예) ‘협조해 주다.’는 왜 띄어 쓰는 건가요???

2) 두 개 이상의 예시 중 바른 표현을 묻는 질문

예) 안녕하세요. 나무 심어봐요. / 나무 심어 봐요. 띄어쓰기 어떻게 맞는 표현인가요?

3) 제시한 표현 교정을 요청하는 질문

예) ‘믿어의심치않다’의 띄어쓰기는 어떻게 하나요?

4) 제시한 표현이 맞는지 확인을 요청하는 질문

예) 보조 용언 ‘있다’의 띄어쓰기가 궁금해요. ‘열려 있다’는 붙여 쓰나요?

[Rules]

모든 질문은 존댓말을 사용해야 합니다.

질문자는 나이, 직업, 말투, 상황 등이 다양해 보이도록 작성하되, 직접적으로 노출하지 마세요.

[example]의 Question 표현을 참고하세요. 그러나, 동일한 내용이나 단어를 사용하지 마세요.

[example]

1) ‘맑다’는 어떻게 발음하나요?

2) 그럼, ‘맑고’는 [말꼬]로 읽나요, [막꼬]로 읽나요?

3) ‘맑게’의 발음이 [말께]인가요?

4) ‘맑은 하늘’에서 ‘맑’은 어떻게 발음하나요?

5) ‘맑고’와 ‘맑게’의 받침 발음이 같은가요?

답변 생성용 프롬프트는 한국어 어문 규범 전문가로서 역할을 부여하고 생성될 모범 답변의 구조와 구성 요소에 규칙 번호를 붙여 더 엄격한 답변 생성을 유도하는 한편, 국어 지식을 답변의 근거로 제시하는 요건도 강화했다. 프롬프트를 질문 생성용과 답변 생성용으로 분리함에 따라 각각의 결과를 생성하기 위한 문맥 정보와 제약 조건 구성이 용이해졌을 뿐만 아니라, 답변 생성의 경우 질문이 모두 생성된 후에 각 질문별로 답변 생성 프롬프트를 적용해서 질문과 답변이 불일치하는 문제가 해소되었다. 질문 생성 결과에서도 평균 유사도 76.84%에서 71.53%로 낮아지며 반복적인 질문이 체감할 수 있을 정도로 낮아졌다.

<표 28> 답변 생성용 프롬프트

당신은 “한국어 교육 및 한국어 어문 규범 전문가”입니다.

<p>[question]에 대해서 [한국어 어문 규범]을 근거로 하여 답변을 작성하세요.</p> <p>[한국어 어문 규범]이 없으면, 알고 있는 "한국어 어문 규범"을 활용해 답하고 그 근거 조항을 답변에 포함하세요.</p> <p>답변은 [답변 작성 규칙]을 준수해야 합니다.</p> <p>[답변 작성 규칙]</p> <p><규칙1>문장 구성 구조 : [인사] → [질문 요약] → [규범 제시] → [설명] → [예시] → [예외 및 보충] → [맺음말].</p> <p><규칙2>답변 길이: 650자 내외.</p> <p><규칙3>문체: 존댓말 사용. 쉬운 문장. 예시 사용.</p> <p><규칙4>문법지식 수준: 학교 문법 수준. [어문 규범]을 근거로 사용.</p> <p><규칙5>인용: 일관된 작은따옴표를 사용할 것.</p> <p><규칙6>예시 활용: 원칙을 설명한 후 올바른 사용 예시를 제시할 것.</p> <p><규칙7>자주 쓰는 표현 유형 :</p> <p>[인사] : ‘안녕하세요? [질문 주제]에 관한 질문에 답변드리겠습니다.’</p> <p>[규범 제시] : ‘~는 ~하는 것이 바릅니다.’, ‘~로 표기하는 것이 적절합니다.’, ‘~는 [이유]로 인해 ~하는 것이 바릅니다.’</p> <p>[예시 제시] : ‘예. ~’, ‘예를 들어 처럼 쓸 수 있습니다.’, ‘예를 들어, “과 같이 씁니다.’</p> <p>[예외 언급] : ‘다만, ~하는 경우에는 ~할 수도 있습니다.’, ‘다만, ~한 특수 상황에서는 ~하게 쓰는 것도 허용됩니다.’</p> <p>[맺음말] : ‘고맙습니다.’, ‘참고하시기 바랍니다.’</p> <p><규칙8>한국어 어문 규범 및 한국어 사용과 무관한 질문:</p> <p>한국어 어문 규범 및 한국어 사용과 무관한 질문은 답변하지 않고 아래 문장으로 대신합니다.</p> <p>‘죄송하지만, 이 질문에는 답변할 수 없습니다. 한국어 어문 규범과 한국어 사용에 대해서 질문에 주세요.’</p> <p>[question]</p> <p>‘맑고’와 ‘맑게’의 받침 발음이 같은가요?</p>

개선된 2차 프롬프트를 적용한 결과 유사 질문 생성, 질문과 답변의 불일치를 약 30%가량 개선했고 생성된 답변의 일관성도 향상되었다. 다만, 단일 프롬프트로 질문-답변 쌍이 한 번에 생성됐던 1차 개선 결과와 달리, 질문 프롬프트로 생성된 질문 수만큼 답변 프롬프트를 실행해야 하기 때문에 인공지능에 전달하는 토큰량이 늘고, 응답 소요 시간이 생성할 답변 수만큼 증가했다. 특히 설계 자료가 변경되거나 질문 생성에 오류가 있어서 질문-답변 쌍을 다시 생성하는 설계 말뭉치가 50여 개 이상인 경우에는 1~2일 이상 소요되기도 했다.

<표 29> 2차 개선 프롬프트 모델별 생성 답변 부적합률

질문 유형 번호	평가자	GPT-4o			Gemini 2.5 pro		
		적합	부적합	부적합률	적합	부적합	부적합률

		질문-답변쌍	질문-답변쌍		질문-답변쌍	질문-답변쌍	
103	설계1	40	5	11.11%	31	14	45.16%
	설계2	30	15	33.33%	35	10	25.87%
	구축1	36	9	20.00%	29	16	55.17%
114	설계1	39	6	13.33%	37	8	21.62%
	설계2	34	11	24.44%	29	16	55.17%
	구축1	31	14	31.11%	34	11	24.44%
126	설계1	38	7	15.56%	36	9	20.00%
	설계2	31	14	31.11%	32	13	28.89%
	구축1	41	4	8.89%	38	7	18.42%
138	설계1	30	15	33.33%	31	14	31.11%
	설계2	32	13	28.89%	37	8	21.62%
	구축1	33	12	26.67%	34	11	32.35%
162	설계1	39	6	15.38%	30	15	30.00%
	설계2	40	5	11.11%	35	10	25.87%
	구축1	41	4	9.76%	36	9	25.00%
합계(평균)		535	140	20.74%	504	171	25.33%

2.2.2. 말뭉치 가공

말뭉치 가공의 핵심 업무는 인공지능이 수십 개의 질문-답변 쌍을 제대로 생성했는지 확인하고 이를 보완하는 것이다. 증강된 질문-답변이 설계 말뭉치의 질문 유형, 국어 지식, 예시 질문-답변들과 일관성을 갖는지 <표 11> 생성 질문-답변 평가 지표에 따라 평가¹⁴⁾하고 잘못 생성된 질문-답변 쌍을 수정, 삭제, 추가하는 것에 중점을 두고 진행했다. 특히, 동일한 질문이 반복되는 것을 최소화하고 가능한 다양한 질문이 제공되도록 보정하는 데 집중했다. 이를 위해 각 세부 질문 간의 유사도를 측정하고 일정 수준(질문: 80% 이상, 답변: 90%)인 경우 우선적으로 수정, 삭제하도록 했다. 또한, 질문과 답변 내용을 검토해 추가 질문(꼬리 질문)이 가능한 경우에는 질문을 추가하고 선행 질문과 연결되도록 가공했다. 이를 위해 이 사업에서 마련한 전용 작업 도구의 주요 기능은 <표 29>와 같다.

14) 이런 면에서 볼 때, 지시문 탐색 지원 말뭉치는 가공 단계 역시 검수의 일환이라고 할 수 있다. 인공지능이 생성한 결과물을 사람의 관점으로 검수한다는 것 외에는 절차와 방법 면에서 동일하다.

<표 30> 작업 도구 주요 기능

주요 기능	설명
설계 데이터 출력	질문 유형, 국어 지식, 질문-답변 예시 5개, 꼬리 질문 여부를 출력하여 작업자가 현재 질문 유형에 대한 설계 관점을 참조할 수 있게 한다.
유사도 측정	각 질문과 답변에 대해 등록된 모든 질문과 대조하여 유사도를 구하고, 최소/최대 유사도를 제시한다. 최대 유사도를 가진 항목을 우선적으로 점검한다.
수정 편집 및 이력 관리	증강된 질문과 답변이 수정되면 이전 값과 현재 값을 기록해 로그로 보관한다. 현재 작업 중인 파일(질문 유형)에서 변경 전후 값을 비교해 다음 단계 검수자가 확인할 수 있다.
JSON 변환	작업 완료된 파일은 엑셀과 JSON으로 변환해 격리 보관한다. (납품용 데이터는 격리 보관된 파일을 기반으로 별도 생성한다.)
기타	파일 등록, 작업자 로그인, 작업자/관리자/검수자 권한 제어, 단계별 작업 현황 및 통계 등

작업자는 부여받은 계정과 비밀번호로 로그인하면 관리자가 분배한 설계 말뭉치 파일 목록을 확인하고 불러올 수 있다(작업 할당 영역). 작업 할당된 설계 말뭉치는 관리자에 의해 질문-답변이 증강 생성되어 있고 작업자는 필요에 따라 질문-답변 쌍 단위로 재생성할 수 있다. 작업자는 작업 중인 말뭉치의 설계 정보(질문 유형, 국어 지식, 예시 질문-답변 쌍)를 필요에 따라 확인할 수 있다.

1. 로그인 영역

사용자 인증

사이드바를 숨기려면 사용자 인증 제목 위에 있는 < 버튼을 클릭하세요.

환영합니다, 작업자02님!

로그아웃

파일 관리

1. assign 출력의 파일 목록

- user002_file516.xlsx
- ✓ user002_file517.xlsx
- user002_file518.xlsx
- user002_file519.xlsx
- user002_file94.xlsx
- user002_file95.xlsx

선택된 파일: user002_file517.xlsx

불러오기

3. 작업영역 : 설계 정보 출력

질문유형/국어지식 표시

질문유형	서비스 개선 관련 건의
국어지식	x

예시 표시

Index	질문	답변	꼬리 질문 여부	질문 변경	답변 변경
예시1	온라인 가나다에 원래 질문자가 아니고 다른 사람도 재질문을 올릴 수 있습니다. 원래 질문자가 설정한 비밀번호를 입력했을 때에만 재질문을 올릴 수 있도록 수정해 주시기 바랍니다.	안녕하세요? 서비스 개선 관련 건의 사항에 대해서 답변드립니다. 온라인 가나다의 재질문 기능 개선에 대한 의견을 주셔서 감사합니다. 관련 부서에 전달하였으며, 비밀번호 확인 후 재질문이 가능하도록 조치 중임을 안내드립니다. 고맙습니다.	nan	nan	nan
예시2	온라인 가나다의 공지 사항에 답변 수정 내용을 달아 주시면 좋겠습니다.	안녕하세요? 서비스 개선 관련 건의 사항에 대해서 답변드립니다. 주신 건의 사항은 관련 부서에 건의 사항으로 전달하여 검토하겠습니다. 의견 주셔서 고맙습니다.	nan	nan	nan
예시3	국립국어원의 '다들은 말' 페이지 안내 문구가 헷갈립니다. '우리말 다들기 누리집'이 따로 있는 줄 알았습니다.	안녕하세요? 서비스 개선 관련 건의 사항에 대해서 답변드립니다. 주신 건의 사항은 관련 부서에 전달하겠습니다. 의견 주셔서 고맙습니다.	nan	nan	nan
예시4	<표준국어대사전>의 뜻풀이에 대한 의견을 제시하고 싶습니다. 그런 의견을 제시할 수 있는 게시판만 만들어 주시면 좋겠습니다.	안녕하십니까? 서비스 개선 관련 건의 사항에 대해서 답변드립니다. <표준국어대사전>의 뜻풀이에 대해서 건의하실 내용이 있으시면 <우리말샘>의 '의견 제시'를 이용해 주시기 바랍니다. 고맙습니다.	nan	nan	nan
예시5	'가나다 전화'의 번호가 '02'로 시작하는 일반 전화번호이면 좋겠습니다.	안녕하세요? 서비스 개선 관련 건의 사항에 대해서 답변드립니다. 주신 건의 사항은 관련 부서에 전달하겠습니다. 의견 주셔서 고맙습니다.	nan	nan	nan

<그림 11> 지시문 기반 탐색 지원 말뭉치 구축 도구(설계 정보 확인)

작업자는 하나의 질문 유형마다 40~50개의 질문-답변 쌍을 반복해서 확인하므로 질문과 답변 내용을 혼동하거나 작업할 질문-답변 쌍을 무심코 건너뛸 수 있다. 이를 방지하기 위해 구축 도구는 우측에 작업 현황을 OMR 답안지처럼 제공해 어떤 질문-답변의 값이 변경되었는지 보여준다. 그리고 질문이나 답변의 수정은 [변경] 버튼을 클릭해야 수정값이 반영되고, 반영 결과가 다음 변경 때까지 지속되어 오조작으로 인한 변경을 최소화하도록 했다.

4. 작업영역 : 질문-답변별 편집 영역

질의응답 수정/편집

질문 및 답변 변경은 비어 있지 않은 값 (nan 이 아닌 값)에 이어서 입력해 주세요

질문/답변 관리

질문/답변 1 (8)

출력: 0

꼬리질문여부

값 입력 (예시 번호 기준)

0

적용

삭제

'달'이라는 글자를 읽을 때 그냥 '달'이라고 읽는 게 맞나요, 아니면 리를 발음을 살짝 섞어야 하나요?

이처럼 단어 끝에서는 '달', 모음이 이어질 때는 '달글'과 같이 '르' 발음이 날 수 있다는 점을 함께 기억하시면 좋겠습니다. 정리하면, '달' 단독 발음은 '달'으로 읽는 것이 좋습니다. 고맙습니다.

답변 1 값이 안녕하세요? '달' 발음에 관한 질문에 답변드리겠습니다. 국립국어원의 '한글 맞춤법'과 '표준 발음법'에 따르면, 겹받침 'ㄹ'은 자음 앞이나 어말(단어 끝)에서는 보통 'ㄱ'으로 발음하는 것이 바릅니다. 다만, 겹받침 'ㄹ'이 뒤에 모음으로 시작하는 조사나 어미가 올 때 '르'로 발음되는 경우도 있습니다.

예:

- '달을': '달글[달글]
- '말은': '말르[말르]

이처럼 단어 끝에서는 '달', 모음이 이어질 때는 '달글'과 같이 '르' 발음이 날 수 있다는 점을 함께 기억하시면 좋겠습니다. 정리하면, '달' 단독 발음은 '달'으로 읽는 것이 좋습니다. 고맙습니다. 으로 변경되었습니다.

질문 변경

답변 변경

5. 작업현황: 변경 질문-답변 목록

	질문	답변	꼬리질문여부
1		true	0
2		true	0
3			0
4			0
5			0
6			0
7			0
8			0
9			0
10			0
11			0
12			0
13			0
14			0
15			0
16			0
17			0
18			0
19			0
20			0
21			0
22			0
23			0
24			0

<그림 12> 지시문 기반 침삭 지원 말뭉치 구축 도구(질문-답변 편집)

작업자에 의해 수정, 삭제, 추가된 질문-답변 쌍은 [임시저장]부터 JSON으로 변환되어 보관되고 [저장]으로 작업 완료를 확정하면 작업관리 기능으로 검수자에게 할당하였다.

2.2.3. 말뭉치 검수

지시문 기반 첨삭 지원 말뭉치 검수의 절차와 방법은 말뭉치 가공과 완전히 동일하다. 검수 작업도 <표 11> 생성 질문-답변 평가 지표에 따라 질문-답변 쌍들이 설계 말뭉치에 부합하도록 작성되었는지를 확인하고 보완하였다. 검수 작업과 가공 작업의 차이는 작업자와 검수자의 국어 능력과 모범 답변 작성 역량에 있다. 데이터 관점에서 검수자의 검수 대상은 설계 정보를 포함한 모든 데이터이므로 당연히 가공 단계에서 작업자가 수정한 질문-답변 쌍도 검수 대상이 된다. 검수자는 이전 작업자, 검수자가 수정한 이력을 통해 현재 질문-답변 쌍의 작성 주체를 확인할 수 있다.

검수 작업은 구축 도구의 작업 관리 기능으로 자동으로 처리하였다. 검수 작업은 여러 차례 반복될 수 있고 특정 파일이 더 많이 검수되어야 할 수 있으므로 작업 완료 시 다음 검수 대기 상태로 자동 전환되도록 했다.

말뭉치의 설계 자료(질문 유형, 국어 지식, 예시 질문-답변)는 설계팀만 수정할 수 있도록 했다.¹⁵⁾ 설계 자료가 질문-답변 전체를 포함하지 못하거나, 적용된 국어 지식, 답변에 사용된 지식에 이견이 존재하는 경우 설계 자료를 포함한 최종 검수를 설계 연구팀에 요청해 지침, 설계, 구축 결과에 일관성을 유지하도록 했다.

15) 즉, 설계 정보는 구축 기간 동안 구축 도구에서 수정할 수 없도록 했다. 설계 말뭉치의 변경은 질문-답변의 재생성 가능성이 높으므로 사실상 수정이라기보다 삭제 후 추가에 해당한다. 권한 위반에 의한 데이터 오염 방지를 위해 구축 도구는 설계 후 구축에만 활용했다.

2.3. 말뭉치 검증

지시문 기반 첨삭 말뭉치는 한국어 특화 인공지능 서비스(국어생활 상담, 글쓰기 첨삭) 구현에 기여하는 데이터를 만드는 데 구축 목적이 있다. 그러므로 말뭉치의 품질도 구축된 말뭉치가 AI의 한국어 특화 학습에 어느 정도 효용성이 있는지를 확인해야 효과적으로 검증될 수 있다. 이를 위해 연구진은 2개 이상의 공개 언어모델(LLM)을 대상으로 학습 전후의 성과 차이를 비교해 구축된 말뭉치를 검증하였다.



<그림 13> LLM 모델 학습 성과 기반 말뭉치 평가 계획

2.3.1. 말뭉치 검증 개요

AI 모델의 학습 성과를 기반으로 말뭉치를 검증하기 위해 말뭉치로 학습시킬 LLM을 선정했다. 대상 모델은 공개 활용이 가능한 라이선스 정책이 있어야 하며 모델 규모가 작으면서도 말뭉치 구축 규모를 감안할 때 학습 결과가 충분히 수렴될 수 있는 규모여야 했다. 또한 한국어 특화 말뭉치임을 감안해 국내 모델과 해외 모델을 모두 포함하기로 했다. 이러한 조건들을 고려해 한국어 학습 능력이 우수하다고 평가된 국내, 해외 1개 모델씩을 선택해 베이스라인(Based-line) 테스트를 진행했다.

베이스라인 테스트란 향후 평가 결과를 비교하기 위한 기준 성능을 측정하는 과정으로 이번 검증에서는 지시문 기반 첨삭 지원 말뭉치로 학습하기 전의 LLM 성능을 평가했다. 테스트에 사용된 데이터는 지시문 기반 첨삭 지원 말뭉치 구축 목표량의 약 10%로 국어생활 관련 질문 유형 114종, 질문-답변 쌍 5,670개를 사용

했다. 평가 방식은 LLM을 활용한 자동 평가 방식인 LLM-as-a-Judge, 대상 AI 모델(LLM)은 HyperCLOVA X(네이버: 국내), Qwen(알리바바: 해외)을 사용했다.

<표 31> 말뭉치 검증을 위한 AI 모델 베이스라인 테스트

	HyperCLOVA X	Qwen
평가 데이터	국어생활 상담 킷어쓰기, 외래어, 로마자 표기 등 114종 세부 질문별 질문-답변 5,670개	
평가 방식	자동 평가(LLM-as-a-Judge)	
모델 크기	3B	7B
파 인 튜 닝	없음	없음
평가 결과	2.22	2.18
오류 사례	<p>(질문)</p> <p>한국 사람들 이름 중에 '정은'이라는 이름이 많은데, 로마자로 표기할 때 'Jeong-eun'으로 쓰는 게 맞나요, 'Jengeun'으로 쓰는 게 맞나요?</p> <p>(모델 답변)</p> <p>'정은'이라는 이름을 로마자로 표기할 때는 'Jeong-eun'으로 쓰는 것이 맞습니다. 한국어 이름의 로마자 표기법에서는 성모(성)를 먼저 쓰고, 그 다음에 모음, 그리고 받침을 순서대로 씁니다. 따라서 '정은'은 'Jeong-eun'으로 표기하는 것이 올바른 표기입니다. 'Jengeun'은 잘못된 표기입니다.</p>	<p>(질문)</p> <p>잘못'과 '잘'의 킷어쓰기가 헛갈려요. '잘못' 이 항상 붙여 쓰는 건가요?</p> <p>(모델 답변)</p> <p>'잘'과 '못'은 종종 혼동되곤 합니다. 하지만 일반 적으로 다음과 같이 사용됩니다: 1. '잘': 잘이라는 의미로 사용될 때는 킷어 씁니다. - 예: 그녀는 잘 먹었습니다. 2. '잘못': 오류나 실수를 의미할 때는 붙여 씁니다. - 예: 그는 잘못 이해했습니다. 하지 만 '잘'이 다른 단어와 결합하여 새로운 의미를 가 지는 경우도 있습니다. 예를 들어: - 잘난: 잘 나 來 看中文 - 잘생긴: 아름다운 외모를 가진 이러 한 경우는 킷어 씁니다. 결론적으로, '잘'과 '못'을 분리해서 쓰는 경우가 더 많습니다. '잘못'은 주로 오류나 실수를 의미하는 경우에 붙여 씁니다.</p>

베이스라인 테스트 결과 HyperCLOVA X가 모델 크기가 더 작음에도 불구하고 미세하지만 더 높은 평가 점수를 받았다. 그러나, 평가 데이터 세트가 요구하는 국어 지식 관련 질문에 대한 답변의 품질은 둘 다 낮은 편이었다. HyperCLOVA X의 경우 사람 이름에 대한 로마자 표기법을 답변으로 생성하기는 했지만, 답변 내용은 잘못된 지식으로 출력되었다. Qwen의 경우 한국어 형태소를 다루는 질문

과 답변에서 잘못된 지식뿐만 아니라 중국어가 뒤섞여 출력되기도 했다. 이에 말뭉치 검증 테스트에서는 AI 모델 2종을 추가하고 모델 크기는 비슷하도록 조정해 총 4개의 모델로 말뭉치 검증을 시행하기로 했다.

2.3.2. 말뭉치 검증 방법 : 평가 과업

본 검증 과업에서는 검증 대상 말뭉치로 학습한 AI 모델이 단순한 텍스트 생성을 넘어, 국어 지식과 관련한 다양한 질의에 대해 규범적 근거를 바탕으로 한 올바른 답변을 도출할 수 있는지 그 성능을 심층적으로 점검했다.

본 연구를 통해 구축된 지시문 기반 첨삭 지원 말뭉치는 실제 언어생활에서 빈번하게 발생하는 오류 사례와 혼동하기 쉬운 표현부터 일반인이 알기 어려운 국어 지식까지 다양하게 구성되었으며, 모델이 지시문에 명시된 복합적인 요구사항을 논리적으로 수행하는 과정을 체계적으로 평가했다. 검증의 핵심 절차는 모델의 판단 능력과 생성 능력을 동시에 측정할 수 있도록 설계된 두 가지 단계의 과업(Task)으로 수행되었다.

우선 첫 번째 단계인 ‘질문 적합성 판별(Task 1)’ 과정에서는 입력된 이용자의 질문이 국어생활과 관련된 상담 질문인지를 실시간으로 분류하게 했다. 이는 인공지능이 무분별하게 답변을 생성하는 것을 방지하고, 어문 규범 기반 답변 제공이라는 본연의 목적에 집중하게 함으로써 시스템의 전문성을 확보하기 위한 절차이다. 만약 질문이 어문 규범과 무관한 주제일 경우, 모델은 학습된 가이드라인에 따라 정해진 거절 문구를 출력하도록 하였다. 이러한 과정을 통해 모델이 질문의 의도를 파악하고, 답변 가능 여부를 스스로 판단할 수 있는 역량을 검증하였다.

두 번째 단계인 ‘규범 기반 답변 생성(Task 2)’ 과정은 적절한 질문으로 판정된 사례에 대해 구체적인 교정과 설명을 제공하는 단계이다. 이때 모델은 단순히 기계적인 정답을 제시하는 것에 그치지 않고, 이용자가 어떠한 부분에서 오류를 범했는지 쉽고 명확하게 설명해야 한다. 특히 어문 규범을 준수하되 일반인이 이해하기 어려운 전문적인 문법 용어의 나열은 지양하며, 친절하고 정중한 어조를 유지하는 것이 중요하다. 모델은 제시된 지식과 예시를 참고하여 답변을 생성하지만, 예시의 문장을 그대로 복제하지 않고 질문의 맥락에 맞추어 어휘와 문장 구조를 새롭게 재구성하는 창의적 응용력을 보여주어야 한다. 예를 들어, 사물에 대한

잘못된 높임 표현인 ‘있는 거세요’와 같은 사례를 검증할 때, 주체가 ‘자료’임을 명시하고 왜 ‘계시다’나 ‘있으시다’가 부적절한지를 논리적으로 설명하도록 했다.

2.3.3. 말뭉치 검증 방법 : 정량적 평가

본 연구에서는 구축된 말뭉치의 객관적인 품질을 측정하고, 모델별 성능 차이를 정밀하게 비교 분석하기 위해 지표 기반의 정량적 평가를 수행하였다. 평가의 공정성과 신뢰성을 확보하기 위해 전체 구축 데이터 중 10%를 무작위로 추출하여 별도의 평가용 데이터 세트를 구성하였으며, 이때 평가용 데이터가 학습 과정에 포함되어 결과가 왜곡되는 ‘데이터 오염’을 방지하기 위해 학습 데이터 세트와 평가 데이터 세트가 엄격히 분리되도록 관리하였다. 이러한 과정을 통해 모델이 단순히 학습 데이터를 암기하여 답변하는 것이 아니라, 미학습된 새로운 질문에 대해서도 어문 규범에 근거한 올바른 추론을 수행할 수 있는지 그 일반화 성능을 검증하고자 하였다.

모델 성능 평가를 위해 국내외에서 성능이 입증된 4종의 언어 모델(QWEN3-4B16), GEMMA-3-4B, LLAMA-3.2-3B, HyperclovaX-3B)을 선정하여 비교 실험을 진행하였다. 각 모델은 동일한 지시문 체계하에서 학습되었으며, 한국어 문법과 어문 규범을 얼마나 정확하게 기술하는지를 중점적으로 비교하였다. 특히 각 모델의 규모와 구조적 특성에 따른 답변의 정확도 및 유창성 차이를 분석해 구축된 지시문 기반 첨삭 지원 말뭉치를 통해 국어 지식과 관련한 학습 성과가 있었는지를 확인하는 데 주목하였다.

평가 방식에 있어서는 주관적 판단의 개입을 최소화하고 일관된 기준을 적용하기 위해 ‘평가 특화 프롬프트’를 활용한 LLM 평가자(LLM Judge) 방식을 채택하였다. 평가자 모델(Judge Model)은 사전에 정의된 5가지 핵심 지표인 사실적 정확성, 근거 및 충실도, 명확성 및 완전성, 유창성 및 구조, 설명의 효율성을 기준으로 5점 척도 평가를 수행한다. 각 지표는 단순한 점수 부여에 그치지 않고, 해당 점수를 산출하게 된 객관적인 근거를 반드시 기술하도록 설계하여 평가 결과의 투명성을 높였다. 아래의 <표 32>는 평가 특화 프롬프트의 전체 내용이다.

16) 베이스라인 테스트에서는 파인튜닝을 진행하지 않기 때문에 해외 오픈 소스 모델의 경우 별도의 7~10B 모델을 선정했으나, 본 시험에서는 구축된 말뭉치를 활용해 추가 학습이 진행되므로 모델 규모로 인한 성능 해석 이슈가 없도록 3~4B의 유사 규모 모델들을 적용했다.

<표 32> 평가 특화 프롬프트 전문

[지시사항 및 역할 정의]

당신은 대한민국 국어, 로마자 표기법, 외래어 표기법 등 공식 **언어 규정에 대한 지식을 갖춘 전문 평가자(Expert Judge)**입니다. 당신의 역할은 특정 질문에 대해 AI 모델이 생성한 답변의 품질을 아래의 5가지 핵심 지표에 따라 매우 객관적이고 공정하게 평가하는 것입니다.

[입력 데이터]

1. 질문 (Question):

{Question_Placeholder}

2. AI 모델 답변 (Answer):

{Answer_Placeholder}

[평가 기준 및 5점 척도]

아래 5가지 지표에 대해 각각 1점(매우 나쁨)부터 5점(매우 좋음)까지 점수를 부여하고, 해당 점수를 부여한 **객관적인 근거(rationale)**를 반드시 제시해야 합니다.

| 점수 | 5점 (최고) | 3점 (보통) | 1점 (최저) |

| :--- | :--- | :--- | :--- |

| **1. 사실적 정확성** | 모든 내용이 관련 표준 및 사실에 완벽히 부합하며 오류 전무. | 사소한 사실 오류 1~2개 발견되나, 핵심 의미는 유지됨. | 핵심 내용이 완전히 잘못되었거나 심각한 오류 다수.

| **2. 근거 및 충실도** | 제시된 모든 근거가 정확하고 질문 해결에 직접적임. | 근거가 대체로 정확하나, 일부는 질문과 무관함. | 제시된 근거가 완전히 틀렸거나, 답변과 모순됨.

| **3. 명확성 및 완전성** | 질문자의 궁금증을 완전히 해소하고 내용이 매우 명료함. | 핵심은 해결했으나, 부연 설명 부족으로 일부 모호함. | 질문의 의도와 완전히 동떨어져 해결이 전혀 안 됨.

| **4. 유창성 및 구조** | 문법/맞춤법 오류 전무, 흐름 및 전문적 어조가 완벽함. | 사소한 문법/맞춤법 오류 1~2개 있으나 이해에 지장 없음. | 문장 구조가 붕괴되어 내용을 이해하기 어려움.

| **5. 설명의 효율성** | 핵심 정보만 간결하게 전달, 불필요한 내용 전무. | 핵심 정보는 있으나, 불필요한 내용이 다소 포함됨. | 핵심 내용을 찾기 어려울 정도로 불필요한 내용이 압도적임.

또한 외부 클라우드 기반의 API를 사용하지 않고, 모든 평가 과정을 내부 구축형인 ‘온프레미스(On-Premise) LLM’ 환경에서 진행하였다. 이는 검증 대상 데이

터가 외부 서비스(GPT 등)로 유출되는 보안 리스크를 원천적으로 차단하고, 민감한 언어 데이터에 대한 통제권을 확보한 상태에서 안정적인 검증 업무를 수행하기 위함이다.

2.3.4. 데이터 변환 및 학습 데이터 형식

<표 33> 지시문 기반 검색 지원 말뭉치 학습 데이터 변환 예시

```
{
  "doc_name": "file477_final",
  "sample_id": "예시49",
  "instruction": "당신은 한국어 어문 규범 전문가입니다.\n당신의 임무는 한국어 어문 규\n범 및 한국어 사용과 관련된 이용자의 다양한 질문에 답변하는 것입니다.\n아래 Task1,\nTask2를 순차적으로 수행하세요.\n\n###Task1###\n아래 [Question]에서 이용자의 질\n문을 확인하고 한국어 어문 규범과 한국어 사용에 대한 질문인지 확인하세요.\n한국어 어\n문 규범이나 한국 사용과 관련된 질문이 아닌 경우에는 답변을 제공하지 말고 아래와 같\n이 답변하세요.\n#답변: 죄송하지만, 이 질문에는 답변할 수 없습니다. 한국어 어문 규범\n과          한국어          사용에          대해서          질문해          주세\n요.\n\n[Question]\n{question}\n\n###Task2###\n아래 [Question]이 한국어 어문 규\n범과 한국어 사용에 대한 질문인 경우 아래 [Knowledge]를 이용하여 질문에 답변하세\n요.\n이용자가 잘 이해할 수 있도록 어려운 문법 용어 사용은 피하고 명확하게 답변을 제\n공하세요.\n답변을 생성할 때는 [Example]의 양식을 따르세요.\n질문 예시는 다음과 같습\n니다:\n\n[Knowledge]\n{knowledge}\n\n[Example]\n#질문: {question_first}\n#답변:\n{answer_first}\n\n[Output]\n[질문]과 [답변]을 모두 [Example] 형태에 맞게 생성해 주\n세요. 생성된 [답변]은 [Example]의 답변과 완전히 일치하지 않도록 표현과 예시를 바꾸어\n주세요. \"질문\"과 \"답변\"은 반드시 들어가야 합니다. \n#질문: {question}\n#답변:",
  "question": "\"선생님, 이 자료는 이미 있는 거세요.\"라는 표현이 올바른가요?",
  "answer": "'계시다'의 사용 확인 요청 질문에 답변드리겠습니다. '있는 거세요'는 올바\n르지 않은 표현입니다. 주체는 '자료'라는 사물이므로 높일 필요가 없으며, '있다'의 높임\n말인 '계시다' 또는 '있으시다'를 쓰는 것이 부적절합니다. 따라서 '선생님, 이 자료는 이미\n있는 것입니다.' 또는 '선생님, 이 자료는 이미 있습니다.'라고 말하는 것이 바랍니다. 고맙\n습니다.",
  "knowledge": "높임은 높여야 할 인물을 직접 높이는 것이 원칙이나 경우에 따라서는\n그의 신체 일부분이나 높여야 할 인물과 긴밀한 관련을 가진 사람 혹은 사물을 높임으로\n써 간접적으로 높이는 방법도 있는데 이를 간접 높임이라고 한다.",
  "input_text": "당신은 한국어 어문 규범 전문가입니다.\n당신의 임무는 한국어 어문 규\n범 및 한국어 사용과 관련된 이용자의 다양한 질문에 답변하는 것입니다.\n아래 Task1,\nTask2를 순차적으로 수행하세요.\n\n###Task1###\n아래 [Question]에서 이용자의 질
```

문을 확인하고 한국어 어문 규범과 한국어 사용에 대한 질문인지 확인하세요.\n한국어 어문 규범이나 한국 사용과 관련된 질문이 아닌 경우에는 답변을 제공하지 말고 아래와 같이 답변하세요.\n#답변: 죄송하지만, 이 질문에는 답변할 수 없습니다. 한국어 어문 규범과 한국어 사용에 대해서 질문해 주세요.\n\n[Question]\n\n"선생님, 이 자료는 이미 있는 거세요.\n"라는 표현이 올바른가요? \n\n###Task2###\n아래 [Question]이 한국어 어문 규범과 한국어 사용에 대한 질문인 경우 아래 [Knowledge]를 이용하여 질문에 답변하세요.\n이용자가 잘 이해할 수 있도록 어려운 문법 용어 사용은 피하고 명확하게 답변을 제공하세요.\n답변을 생성할 때는 [Example]의 양식을 따르세요.\n질문 예시는 다음과 같습니다:\n\n[Knowledge]\nx\n\n[Example]\n#질문: 안녕하세요. '스님께서 요즘 변비가 계신 것 같다.'라는 문장은 맞는 표현인가요?\n#답변: '계시다'의 사용에 관한 질문에 답변하겠습니다. '계시다'는 동사 '있다'의 높임 표현입니다. 이 경우는 스님을 직접 높이는 것이 아니므로 간접 높임을 써야 합니다. 간접 높임은 높여야 할 주체의 신체의 일부, 소유물, 생활의 필수적 조건이나 밀접한 관련이 되는 사물을 높임으로써 주체를 간접적으로 높이는 방법을 말합니다. 따라서 '있으시다'로 표현하는 것이 바릅니다. 고맙습니다.\n\n[Output]\n\n[질문]과 [답변]을 모두 [Example] 형태에 맞게 생성해 주세요. 생성된 [답변]은 [Example]의 답변과 완전히 일치하지 않도록 표현과 예시를 바꾸어 주세요. \"질문\"과 \"답변\"은 반드시 들어가야 합니다. \n#질문: \"선생님, 이 자료는 이미 있는 거세요.\n\"라는 표현이 올바른가요? \n#답변: \"label\": \"#질문: \"선생님, 이 자료는 이미 있는 거세요.\n\"라는 표현이 올바른가요? \n#답변: '계시다'의 사용 확인 요청 질문에 답변드리겠습니다. '있는 거세요'는 올바르지 않은 표현입니다. 주체는 '자료'라는 사물이므로 높임 필요가 없으며, '있다'의 높임말인 '계시다' 또는 '있으시다'를 쓰는 것이 부적절합니다. 따라서 '선생님, 이 자료는 이미 있는 것입니다.' 또는 '선생님, 이 자료는 이미 있습니다.'라고 말하는 것이 바릅니다. 고맙습니다.\"}

2.3.5. 모델 학습 방법

말뭉치 검증에서 활용된 데이터 세트는 모델이 복합적인 지시 사항을 논리적으로 수행할 수 있도록 구조화된 JSON 형식을 취하고 있다. 각 데이터는 단순한 질의응답 쌍을 넘어, 모델의 역할을 정의하는 페르소나 설정부터 단계별 수행 절차, 참조 지식, 그리고 최종 정답 라벨로 설계된 필드들로 구성되어 있다. 이는 모델이 한국어 어문 규범 전문가로서 사고하고 답변하는 과정을 데이터의 구조적 층위에서부터 지원하기 위함이다.

데이터의 핵심 필드인 ‘instruction’은 모델에게 부여된 전문적 임무를 명시한다. 여기에는 한국어 어문 규범 전문가라는 역할 정의와 함께, 질문의 적합성을 먼저 판단하는 Task 1과 구체적인 규범 답변을 생성하는 Task 2의 순차적 수행

로직이 담겨 있다. 특히 부적절한 질문에 대한 예외 처리 문구와 답변 시 지켜야 할 가이드라인(쉬운 용어 사용, 특정 양식 준수 등)을 포함하여 모델의 수행 경로를 엄격하게 제어하였다.

개별 콘텐츠를 구성하는 항목을 살펴보면, ‘question’은 사용자의 질문이고, ‘answer’는 해당 질문에 대한 전문가의 모범 답안이다. ‘knowledge’ 필드는 답변의 근거가 되는 원천 정보이다. ‘label’ 필드는 모델이 최종적으로 도출해야 하는 출력값의 표준(Gold Standard)으로서 기능한다. 특히 ‘label’은 ‘#질문’과 ‘#답변’이라는 특정 태그 구조를 포함하고 있어, 모델이 내용적 정확성뿐만 아니라 형식적 일관성까지 유지하고 있는지 검증하는 척도가 된다.

특히 본 데이터 구조에서는 ‘input_text’ 필드 내에 구체적인 수행 예시(Example)를 직접 포함하고 있다. 이는 인공지능의 ‘퓨샷 프롬프팅(Few-shot Prompting)’ 기법을 데이터 구조에 내재화한 것으로, 모델이 지시문을 이론적으로 이해하는 데 그치지 않고 실제 답변의 적정 수준과 형식을 직관적으로 습득하도록 설계되었다. 예시 데이터에는 ‘스님’과 관련된 높임법 오용 사례와 같이 한국어 규범의 핵심을 관통하는 문답을 배치하여, 모델이 Task 1의 분류 과정과 Task 2의 답변 생성 과정을 어떠한 논리적 흐름으로 연결해야 하는지 명확한 가이드라인을 제공한다. 예를 들어, 자칫 모델이 딱딱한 문법 용어만을 나열하거나 반대로 너무 일상적인 말투를 사용하지 않도록, 예시를 통해 국어 전문가다운 정중하고 신뢰감 있는 어조를 학습시켰다. 또한, ‘#질문’과 ‘#답변’이라는 특수 태그를 활용한 출력 형식을 예시로 미리 보여줌으로써, 모델이 복잡한 지시 사항 속에서도 규격화된 답변 양식을 엄격히 준수하도록 유도하였다.

해당 데이터 샘플에서는 아래의 예시가 삽입되었다.

- #질문: 안녕하세요. ‘스님께서 요즘 변비가 계신 것 같다.’라는 문장은 맞는 표현인가요?
- #답변: ‘계시다’의 사용에 관한 질문에 답변하겠습니다. ‘계시다’는 동사 ‘있다’의 높임 표현입니다. 이 경우는 스님을 직접 높이는 것이 아니므로 간접 높임을 써야 합니다. 간접 높임은 높여야 할 주체의 신체의 일부, 소유물, 생활의 필수적 조건이나 밀접한 관련이 되는 사물을 높임으로써 주체를 간접적으로 높이는 방법을 말합니다. 따라서 ‘있으시다’로 표현하는 것이 바릅니다. 고맙습니다.

2.3.6. 모델 학습 설정

본 연구에서는 구축된 말뭉치를 활용하여 인공지능 모델이 한국어 어문 규범 전문가로서의 역량을 습득할 수 있도록 미세 조정(Fine-tuning) 학습을 수행하였다. 학습 효율성을 극대화하기 위해 허깅 페이스(Hugging Face)의 TRL(Transformer Reinforcement Learning) 라이브러리에서 제공하는 SFTTrainer(Supervised Fine-Tuning Trainer)를 활용하였다.

대규모 언어 모델의 대용량 데이터를 신속하게 처리하기 위해 다중 GPU 환경에서 DDP(Distributed Data Parallel) 전략을 채택하였으며, accelerate launch 명령어를 통해 병렬 연산 학습을 구현함으로써 전체적인 학습 시간을 단축하였다. 구체적인 학습 하이퍼파라미터로는 안정적인 학습을 위한 per_device_train_batch_size=4, 문맥 유지 능력을 위한 max_length=2048, 정밀한 연산을 지원하는 bf16=True 및 adamw_8bit 최적화 알고리즘을 적용하였다. 또한 모델의 과적합을 방지하고 일반화 성능을 높이기 위해 3회 반복 학습(num_train_epochs=3)을 수행하였으며, 에포크 단위로 모델 상태를 저장하여 최적의 성능을 발휘하는 시점의 가중치를 확보하였다. 학습에 사용된 모델은 다음과 같다.

- meta-llama/Llama-3.2-3B-Instruct
- Qwen/Qwen3-4B-Instruct-2507
- naver-hyperclova/HyperCLOVAX-SEED-Vision-Instruct-3B
- google/gemma-3-4b-it

각 모델은 동일한 환경 내에서 학습되었다.

2.3.7. 모델 평가 결과

<표 34> AI 모델별 성능 평가 결과

모델	Base-line	평가 유형	통합 평균	사실적 정확성	근거 및 충실도	명확성 및 완전성	유창성 및 구조	설명의 효율성
Qwen_Qwen3-4B-Instruct-2507	2.18 ¹⁷⁾	Indist	3.7905	3.4128	3.1306	3.5990	4.7210	4.0890
		Outdist	3.7017	3.2638	3.0678	3.4742	4.6956	4.0069
google_gemma-3-4b-it	-	Indist	3.6975	3.2680	3.0070	3.4898	4.7096	4.0129
		Outdist	3.6046	3.1415	2.9253	3.3568	4.6805	3.9189
meta-llama_Llama-3.2-3B-Instruct	-	Indist	3.6067	3.1666	2.9254	3.3813	4.6549	3.9054
		Outdist	3.4543	2.9631	2.7545	3.1901	4.5967	3.7670
naver-hyperclovax_HyperC LOVA X-SEED-Vision-Instruct-3B	2.22	Indist	3.5237	3.0706	2.8071	3.2705	4.6223	3.8481
		Outdist	3.3962	2.9314	2.6803	3.1101	4.5426	3.7167

<표 35> 모델별 학습후 출력 사례1 : [질문] 선생님, 이 자료는 이미 있는 거세요."라는 표현이 올바른가요?

모델	생성답변	통합 평균	사실적 정확성	근거 및 충실도	명확성 및 완전성	유창성 및 구조	설명의 효율성
Qwen3	'게시다'의 사용 확인 요청 질문에 답변드리겠습니다. '이미 있는 거세요.'는 '있다'의 간접 높임말인 '있으시다'가 사용된 것으로 보입니다. 주체(자료)를 높이는 것이 아니라 소유를 나타내는 말에 높임을 쓴 것이므로, 간접 높임을 쓴 바른 표현입니다. 고맙습니다.	1.6	1	1	1	3	2
gemma-3	'게시다'의 사용 확인 요청 질문에 답변드리겠습니다. 이 문장에서 주체는 '자료'이고, '있는'은 '있다'의 활용형입니다. '자료'는 살아 있는 생물이 아니므로 '게시다'를 쓸 수 없습니다. 따라서 '이 자료는 이미 있다.' 또는 '이 자료는 이미 있으시다.'(주체 높임)로 표현하는 것이 바릅니다. 고맙습니다.	1.4	1	1	1	2	2
Llama-3.2	'게시다'의 사용 확인 요청 질문에 답변드리겠습니다. 네, '자료는 이미 있는 거세요.'는 올바른 표현입니다. '있으시다'는 '있다'의 높임말로, 자료나 문서와 같은 소유물을 간접적으로 높이는 데 사용됩니다. 따라서 선생님의 자료를 높여서 나타내는 데 적절합니다. 고맙습니다.	1.6	1	1	1	3	2
naver-hyperclovax	'거세요'의 사용 확인 요청 질문에 답변드리겠습니다. 네, '거세요'는 '있다'의 높임말인 '게시다'의 명령형 활용형입니다. 이 문장은 선생님께 자료가 이미 존재함을 알리는 간접적인 높임 표현이므로, '있으세요'의 명령형인 '거세요'를 사용하여 '이 자료는 이미 있는 거세요.'로 표현하는 것이 바릅니다. 고맙습니다.	1.2	1	1	1	2	1

17) 베이스라인 테스트에서는 추가 학습이 없는 점을 고려해 Qwen2.5 7b 모델로 평가했다.

<표 36> 모델별 학습후 출력 사례2 : [질문] 주변에서 다들 [인산말]이라고 하니까 저 혼자 [인사말]이라고 하면
오히려 유난 떠는 것처럼 보일까 봐 눈치 보여요. 그냥 대세에 따르는 게 나을까요?'

모델	생성답변	통합 평균	사실적 정확성	근거 및 충실도	명확성 및 완전성	유창성 및 구조	설명 의 효율성
Qwen3	‘인사말’의 올바른 발음과 표기에 관한 질문에 답변드리겠습니다.\n주변에서 많이 쓰는 [인산말]이라는 발음 때문에 혼자 [인사말]을 사용할까 봐 고민이 되시는군요.\n결론부터 말씀드리면, ‘인사말’은 표기 그대로 [인사말]로 발음하는 것이 바릅니다. 따라서 [인산말]로 발음하는 것은 잘못된 표현입니다.\n‘표준 발음법’ 제29항은 합성어에서 ‘ㄴ’ 소리가 첨가되는 현상을 다루고 있습니다. 하지만 ‘인사말’은 ‘인사’와 ‘말’이 결합할 때 ‘ㄴ’ 소리가 덧나지 않는 단어로 규정되어 있습니다. 즉, 사이시옷 현상이 적용되지 않으므로 ‘ㄴ’을 첨가하지 않고 본래의 소리대로 발음해야 합니다. 많은 사람이 [인산말]로 발음하는 것은 ‘혼잣말[혼잔말]’이나 ‘노랫말[노랜말]’처럼 비슷한 형태의 단어에 익숙해져서 잘못 유추하여 적용하기 때문입니다. 하지만 이는 ‘표준국어대사전’에 명시된 표준 발음이 아니므로 주의해야 합니다.\n예를 들어, ‘졸업식 인사말’을 말할 때는 [졸업식 인산말]이 아니라 [졸업식 인사말]과 같이 씁니다. 마찬가지로 ‘머리말’ 역시 [머린말]이 아닌 [머리말]로 발음하는 것이 올바른 용법입니다.\n다만, 언어 현실에서 [인산말]이라는 발음이 널리 퍼져 있어 혼동을 겪는 경우가 많습니다. 하지만 규범에 맞는 발음은 [인사말]이므로, 자신 있게 올바른 표준 발음을 사용하셔도 좋습니다.\n고맙습니다.	4.6	5	5	5	4	4
gemma-3	‘인사말’의 올바른 발음과 표기에 관한 질문에 답변드리겠습니다.\n주변에서 ‘인산말’로 발음하는 경우가 많아 혼란스러웠군요. 결론부터 말씀드리면, ‘인사말’은 [인사말]로 발음하고 표기하는 것이 바릅니다.\n‘표준 발음법’ 제30항은 합성어에서 ‘ㄴ’ 소리가 첨가되는 현상, 즉 ‘사이시옷 현상’을 다루고 있습니다. 많은 분이 ‘인사말’을 [인산말]로 발음하지만, 이는 ‘존댓말[존덴말]’이나 ‘혼잣말[혼잔말]’처럼 ‘ㄴ’ 소리가 덧나는 구조와 혼동하여 생긴 오류입니다. ‘인사말’은 ‘인사’와 ‘말’이 결합한 단어로, ‘ㄴ’ 소리가 덧나지 않는 것이 원칙입니다. 따라서 표기 역시 소리 나는 대로 ‘인산말’이 아닌 ‘인사말’로 적어야 합니다.\n예를 들어, ‘졸업식 인사말을 준비하다’라고 할 때는 [졸업식 인산마를]이 아니라 [졸업식 인사마를]과 같이 발음합니다. 이와 비슷한 예로 ‘머리말’도 [머린말]이 아니라 [머리말]로 발음하고 표기하는 것이 적절합니다.\n다만, 언어 현실에서 많은 사람이 [인산말]로 발음하더라도 그것이 곧바로 표준 발음이 되지는 않습니다. ‘표준 발음법’은 ‘국어의 전통성과 합리성을 고려하여 정함’을 원칙으로 하므로, 널리 쓰이더라도 규범에 맞지 않는 발음은 지향해야 합니다.\n고맙습니다.	4.6	5	5	5	4	4
Llama-3.2	‘인사말’의 올바른 발음에 관한 질문에 답변드리겠습니다.\n주변에서 ‘인사말’을 [인산말]로 발음하는 경우가 많아 혼동을 느끼셨군요. 결론부터 말씀드리면, ‘인사말’은 [인사말]로 발음하는 것이 바릅니다.\n‘표준 발음법’ 제29항은 합성어에서 ‘ㄴ’ 소리가 첨가되는 현상을 다루고 있습니다. 많은 분이 ‘인사말’을 [인산말]로 발음하는 이유는 ‘존댓말[존덴말]’이나 ‘혼잣말[혼잔말]’과 같은 단어의 발음 형태와 유사하게 ‘ㄴ’이 덧나는 것으로 착각하기 때문입니다. 하지만 국립국어원에서는 ‘인사말’을 ‘ㄴ’ 첨가가 없는 단어로 규정하고 있습니다. 따라서 표기된 글자 그대로 [인사말]이라고 발음해야 합니다.\n예를 들어, ‘졸업식 인사말’은 [졸업식 인산말]이 아니라 [졸업식 인사말]과 같이 씁니다. 이와 비슷한 예로 ‘머리말’ 역시 [머린말]이 아닌 [머리말]로 발음해야 합니다.\n다만, ‘노랫말[노랜말]’이나 ‘혼잣말[혼잔말]’처럼 사이시옷이 표기되어 있고 ‘ㄴ’ 소리가 덧나는 단어들은 실제 발음에서 ‘ㄴ’이 첨가되는 현상이 나타납니다. 이 때문에 혼동하기 쉽지만, ‘인사말’은 사이시옷이 없으므로 [인사말]로 발음하는 것이 정확한 표현입니다. 따라서 걱정하지 않으셔도 됩니다.\n고맙습니다.	4.4	5	4	4	5	4
naver-hyperclovax	‘인사말’의 올바른 발음과 사용에 관한 질문에 답변드리겠습니다.\n주변에서 흔히 듣는 발음인 [인산말] 대신, 저는 [인사말]이라고 발음하는 것이 더 적절합니다. 이는 ‘표준 발음법’ 제29항 ‘ㄴ’ 소리 첨가 규정과 관련이 있습니다.\n‘인사말’은 ‘인사’와 ‘말’이 결합한 합성어입니다. 우리말에는 ‘표준 발음법’에 따라 ‘ㄴ’ 소리가 덧나는 현상이 자주 나타나는데, 이를 ‘ㄴ’ 소리 첨가라고 합니다. ‘표준 발음법’ 제29항은 ‘합성어 및 파생어에서, 앞 단어나 접두사의 끝이 자음이고 뒤 단어나 접미사의 첫음절이 ‘이, 야, 여, 요, 유’인 경우에는, ‘ㄴ’ 소리를 첨가하여 [니, 나, 네, 뇨, 뉴]로 발음한다’라고 규정하고 있습니다. 하지만 ‘인사말’은 이러한 ‘ㄴ’ 첨가 규칙의 적용 대상이 아닙니다. 따라서 표기된 대로 [인사말]로 발음하는 것이 바릅니다.\n예를 들어, ‘머리말’은 [머린말]로, ‘노랫말’은 [노랜말]로 발음하지만, ‘인사말’은 [인사말]로 발음합니다. 이와 같은 원리로 ‘머리’와 ‘말’이 합쳐진 ‘머리말’도 [머린말]이 아닌 [머리말]로 발음하는 것이 원칙입니다.\n다만, 일상 대화에서 많은 사람이 [인산말]로 발음하는 경향이 있어 혼동하기 쉽습니다. 하지만 정확한 한국어 사용을 위해서는 표	4.2	4	3	5	5	4

	준 발음인 [인사말]을 사용하는 것이 바람직합니다. 표준 발음은 국어 생활의 통일성을 위해 매우 중요하므로, 주변 사람들의 발음과 다르더라도 자신 있게 [인사말]이라고 말하셔도 전혀 유난이 아닙니다.\n\n고맙습니다.							
--	---------------------------------------------------------------------------------------------------------------------------	--	--	--	--	--	--	--

2.3.7.1. 지표 간 상관관계 및 모델별 성능 안정성

본 검증 결과에서 주목할 점은 평가 지표 간의 일관성이다. 성능이 우수한 모델은 특정 지표에 치우치지 않고 사실적 정확성부터 설명의 효율성까지 모든 영역에서 고르게 높은 점수를 기록하였다. 특히 최상위 성능을 보인 모델의 경우, 학습 데이터와 유사한 분포를 가진 Indist 데이터뿐만 아니라, 모델이 학습 과정에서 접하지 못한 새로운 유형의 질문인 Outdist 데이터에서도 성능 저하 폭이 매우 적은 것으로 나타났다. 이는 고도화된 모델일수록 본 사업을 통해 구축된 어문 규범 데이터를 단순 암기하는 것이 아니라, 규범의 논리 구조를 내재화하여 미학습 데이터에 대해서도 유연한 추론 능력을 발휘하고 있음을 시사한다.

2.3.7.2. 모델의 유창성 및 형식 준수 능력

문장 생성의 자연스러움과 구조적 완결성을 평가하는 유창성 지표에서는 비교 대상인 4종의 모델 모두 4.5점 이상의 최상위권 점수를 기록하였다. 이는 최신 언어모델들이 지시문에 명시된 특정 출력 형식(#질문, #답변 태그 사용 등)을 엄격히 준수하며, 전문가다운 어조로 답변을 구성하는 기본 역량은 이미 충분히 갖추었음을 보여준다. 따라서 향후 모델 고도화의 초점은 형식적인 유창성보다는 답변의 실질적인 정확도와 근거의 충실도를 높이는 데 집중될 필요가 있다.

본 말뭉치 검증 과업을 통해 도출된 핵심 내용은 다음과 같다.

1) 데이터 구축의 실질적 효과 입증

구축 초기, 학습을 수행하지 않은 실험에서는 하이퍼클로바 3B 모델의 경우 2.2점, QWEN 7B의 경우 2.18점의 낮은 점수를 보인 바 있었다. 이와 비교할 때 본 사업을 통해 구축된 데이터를 학습한 후 모든 모델이 3점대 중후반에서 4점대에 이르는 괄목할 만한 성능 향상을 보였다. 이는 본사업

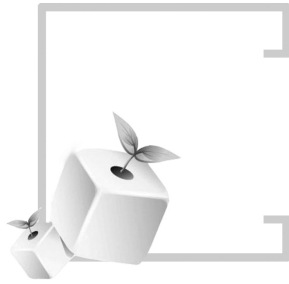
에서 구축한 한국어 어문 규범 데이터가 모델의 전문성을 강화하는 데 결정적인 역할을 수행하고 있음을 정량적으로 입증하는 결과로 볼 수 있다.

2) 기초 모델 선정의 중요성

검증 결과, 자체적인 추론 능력이 뛰어난 범용 LLM을 기반으로 학습을 진행할 때 어문 규범 안내 성능 또한 비례하여 상승하는 경향이 뚜렷하게 관찰되었다. 따라서 향후 고품질의 어문 규범 상담 서비스를 구축하기 위해서는 파라미터 최적화와 더불어 높은 기초 성능을 보유한 모델을 선별하여 활용하는 전략이 필수적이다.

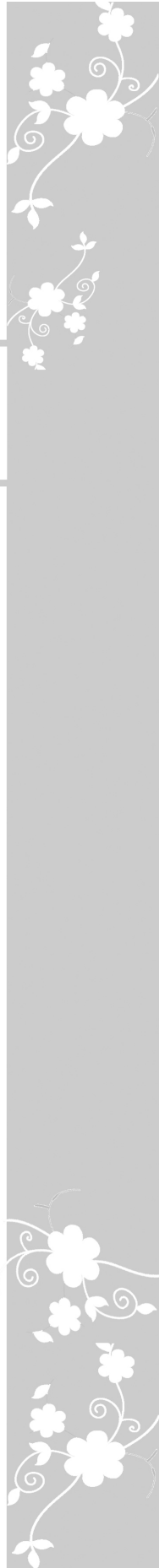
3) 한국어 특화 모델의 현시점 한계와 과제

본 검증에서 한국어 특화 모델로 기대를 모았던 HyperCLOVA X 모델이 외산 베이스 모델들과 유사한(또는 상대적으로 낮은) 성능을 기록한 점은 시사하는 바가 크다. 이는 어문 규범과 같은 고도의 논리적 추론이 필요한 영역에서 한국어 특화 모델의 미세 조정 전략이나 학습 데이터 반영 방식에 대한 재검토가 필요함을 의미한다. 향후 한국어에 최적화된 독자적인 LLM 도입 및 고도화 연구를 위한 지속적인 투자가 뒷받침되어야 할 것으로 판단된다.



제 3 장

국어 지식 기초 자료



3.1. 국어 지식 기초 자료 개요

본 연구에서 국어 지식 기초 자료는 질문-답변에 활용될 수 있는 어문 규범 및 국어 지식 기초 데이터를 형성하는 데 목표를 두고 구축되었다. 앞서 구축한 질문-답변 쌍 중심의 말뭉치는 단순한 답변 제시가 아니라, 질문에 포함된 질문 의도와 관련된 국어 지식이 결합된 모범 답변을 생성하기 위한 문맥 정보를 포함하고 있다. 그런데, 국어상담 자료 중에는 단순한 정보 제공을 요청하는 질문이 상당수 포함되어 있다. 예를 들어, 외래어, 로마자 표기 등에 많이 나타나는 ‘Pâtissier’는 파티쉐가 맞나요? 파티시에가 맞나요?’와 같은 질문처럼 정답이나 규정이 어느 것 인지를 묻는 경우가 많다. 문법적인 원리나 규칙보다 권장하는 결괏값이 무엇인지 확인하고자 하는 경우이다. 이런 질문은 입력 ‘Pâtissier’에 대한 출력 ‘파티시에’와 관련 규범을 하나의 데이터로 묶어 효과적으로 구성할 수 있다. 비슷한 사례로 규범 표기, 표준어 등도 등재된 사전, 문헌 정보를 이용해 구축할 수 있다. 따라서 국어 지식 기초 자료는 외래어, 로마자, 규범 표기, 발음 표기법 규정에 등장하는 예시를 중심으로 구축하였다. 데이터 구성은 2024년 구축된 기초 자료와 호환성을 갖도록 동일한 구조를 유지했다.

<표 37> 국어 지식 기초 자료 데이터 구성 항목

구성 항목	예시	설명
유형 (input_type, out_type)	외래어표기법(규범)	• 규범/용례로 구분
입력 (input)	C(규범), 야곱(오류), Jacob(용례)	• 사용자 질의 값(오류) • 규범 제시 값
출력 (output)	크(규범), 야코브(용례)	• 규범 또는 정답이 되는 값
분류 (category_large, medium, small)	외래어 표기법	• 규범명
	표기 세칙	• 규범 제목
	노르웨이어	• 규범 분야
비고 (category_detail)	외래어 표기법 제2장 표12 노르웨이어 자모와 한글 대조표 : 자음 앞·어말	

한글 맞춤법을 비롯한 어문 규범들은 문법적 원리와 원칙을 설명하는 한편, 예시를 통해 규범/규칙, 오류, 예외/허용 등을 제시하고 있다.

3.2. 국어 지식 기초 자료 유형별 구축 사례

기초 자료의 구축 대상인 로마자, 외래어 표기 등의 자료 특성을 입출력 값의 유형과 규범의 유무 관점에서 구분하여 적용하였다. 입출력 값의 유형은 입력과 출력이 모두 규범인 형태와 입력은 오류 또는 비규범이고 출력은 규범 또는 권장 표기인 두 가지 유형으로 구성되었다. 대부분의 규범에 대한 예시들은 설명이거나 오류에 대한 정정이기 때문이다. 예를 들어, 로마자/외래어 표기법은 한글/외국어(입력)에 대한 외국어/한글(출력)로 대부분의 예시들이 기술되어 있다. 이 경우 입력과 출력은 모두 규범(또는 정상)이다. 규범의 유무는 맞춤법 규정, 로마자/외래어 표기법 등의 장, 절, 항과 상세 항에 속하는 예시와 명시적으로 종속되는 규범 없이 그에 준하는 권장 표기로 제시된 예시 목록이다. 국립국어원에서는 여러 차례의 전문가 논의를 거쳐 권장하는 표기의 목록들을 엑셀, 오픈에이피아이, 검색 결과로 제공하고 있다.¹⁸⁾ 다만, 인공지능의 학습과 같은 대량의 데이터를 일괄적으로 처리하는 데 활용하기에는 규범 유형별로 데이터의 형식, 의미, 구조에 약간씩 차이가 있어 불편함이 있다. 본 연구에서는 이들을 일관된 형식으로 제공하여 대규모 국어 지식의 다양한 활용이 용이하도록 하였다.

3.2.1. 표준발음법

표준발음법은 표준어 규정의 제2부로 표준어 표기에 대한 발음을 제시하고 있다. 표준발음법의 7개 장, 30항에 걸쳐 발음 규칙과 원칙, 예외를 설명하고 있지만, 예시 자료 형식은 “표준어 표기[발음 표기]”로 일관되게 유지하고 있다. 그러므로 표준어 표기를 입력값, 발음 표기를 출력값으로 기초 자료를 구축하고 각각의 자료가 속한 표준발음법 목차와 설명을 한 묶음의 데이터로 표현해 활용할 수 있다. 단, 발음 표기를 나타내는 대괄호 []에 2개의 발음이 제시된 경우와 각 항에 부가된 설명에 따라 값의 양상이 달라지기도 하므로 유의해야 한다. 예시의 표현은 크게 3가지 형태이다.

18) 이번에 구축된 자료의 로마자 표기법, 외래어 표기법의 예시 자료는 '국립국어원-용례찾기'(<https://korean.go.kr/kornorms/example/exampleList.do>)에 게재된 검색결과와 엑셀자료(2025.10.29.일자)를 기준으로 구축되었다.

1) 표기[발음] : 디근이[디그시] (표준 발음법 제4장 16항)

유일한 표준 발음을 갖는 경우. 다만, 설명에 따라 예외나 허용인 경우가 있다. 국어 지식 기초 자료에서는 별도의 예외, 허용을 명확히 설명하지 않는 경우에는 입출력 값 유형을 모두 정상으로 표시한다. 설명을 통해 식별할 수 있다.

2) 표기[발음1/발음2] : 메별[메별/메별] (표준 발음법 제2장 제5항)

두 개의 발음이 인정되는 경우. 다만, 설명에 따라 둘 중 하나는 예외나 허용인 경우가 있다. 두 개의 발음이 제시되므로 2개의 기초 자료로 등록한다.¹⁹⁾

3) 표기[발음1→발음2] : 막론[막논→망논] (표준 발음법 제5장 제19항 [붙임])

[발음1]은 오류, [발음2]만 표준 발음인 경우. 두 가지를 모두 기초 등록함으로써 오류 입력에 대한 대응에 활용될 수 있다.

<표 38> 표준발음법 국어 지식 자료 구축 사례

Ind ex	input	input_type	output	output_type	category_large, medium, small	category_detail
1	가저	정상	[가저]	정상	제2장 제5항	다만 1. 용언의 활용형에 나타나는 ‘저, 쯔, 처’는 [저, 쯔, 처로 발음한다.
2	메별	정상	[메별]	정상	제2장 제5항 ²⁰⁾	다만 2. ‘예 레’ 이외의 ‘ㄹ’은 [ㄹ로 발음한다.
	메별	정상	[메별]	허용		
3	반신반의	정상	[반신 바니]	정상	제2장 제6항	다만, 합성어의 경우에는 둘째 음절 이하에서도 분명한 긴소리를 인정한다.
	반신반의	정상	[반신 바니]	정상		
4	막론	정상	[막논]	오류	제5장 제19항 붙임	받침 ‘ㄱ, ㅇ’ 뒤에 연결되는 ‘ㄹ’도 [니]로 발음한다.
	막론	정상	[망논]	정상		

19) 일반적으로 1개의 입력에 대해 서로 다른 두 값은 중의성, 모호성을 갖는 것으로 여겨졌으나, LLM의 등장으로 새로운 해석의 여지가 생겼다. 1개의 입력에 대한 두 출력값을 어떻게 해석해야 하는지가 문맥값으로 제공되고 해석될 수 있으므로 두 출력값은 중의성이나 모호성이 아니라 필연적인 값이 되었다. ‘메별’은 [메별]이거나 [메별]이 아니라 두 발음이 모두 가능해서 다른 표기와 차별성을 갖는 자료로 해석될 수 있다. 더불어, 국어 지식 기초 자료 구축 시점(2025.12.10.)까지 메별, 반신반의 표준 발음을 제시할 수 있는 인공지능 모델은 구글 제미나이3 뿐이었다. 두 단어 모두 검색을 통해 쉽게 얻을 수 있다는 점, 제미나이와 GPT가 모델 규모나 학습 데이터 규모에서 현격한 차이를 갖지 않는다는 점 등을 고려하면 이 응답 결과의 차이는 AI에 데이터를 어떻게 활용하느냐에 따라 차별화가 가능하다는 점을 시사할 수 있다.

20) 실제 구축 자료에서는 모든 입출력값 쌍이 근거(출처) 정보를 가지지만 여기서는 편의한 하나의 셀에 기록했다.

3.2.2. 로마자 표기법

현행 로마자 표기법은 2000년에 개정해 국어의 표준 발음법의 적용을 받는다. 로마자 표기법의 경우 기본적인 입력값은 한글 표기, 출력값은 로마자 표기가 되지만 데이터의 구조는 대체로 표준 발음법과 유사하다. 제2장 표기 일람에 한글 자음 모음에 대한 대응표가 있다는 점, 국어 발음 표기가 제시되는 경우가 있다는 점, “퇴계로 3가, Toegyero 3(sam)-ga”와 같이 부분 이중 표기가 있다는 점 등을 제외하면 표준 발음법의 자료와 동일하게 기초 자료로 통합할 수 있다. 다만, 로마자 표기법 전체에 제시된 예시는 표기 일람의 자모를 포함해 150여 개 수준으로 로마자 표기법이 사용되는 사례에 비해 매우 적다. 이에 대해 국립국어원에서 로마자 표기 용례 1만 개 이상을 자료로 제공하고 있어 로마자 표기법의 예시와 동등하게 기초 자료에 포함할 수 있다.

<표 39> 로마자 표기법 국어 지식 자료 구축 사례

Ind ex	input	input_type	output	output_type	category_large, medium, small	category_detail
1	ㅏ	규범	eo	규범	제2장 제1항	1. 단모음
2	ㄱ	규범	g,k	규범	제2장 제2항	2. 파열음
3	월곶 ²¹⁾	규범	Wolgot	규범	제2장 제2항	[붙임 1] 'ㄱ, ㄷ, ㅂ'은 모음 앞에서는 'g, d, b'로 자음 앞이나 어말에서는 'k, t, p'로 적는다
4	민용하	정상	Min Yongha	정상	제2장 제6항	인명은 성과 이름의 순서로 띄어 쓴다. 이름은 붙여 쓰는 것을 원칙으로 하되 음절 사이에 붙임표(-)를 쓰는 것을 허용한다
	민용하	정상	Min Yong-ha	허용	제3장 제4항	
5	종로2가	정상	Jongno 2(i)-ga	정상	제3장 제5항	제5항도 시·군·구 읍·면·리 동의 행정 구역 단위와 '가는' 각각 'do, si, gun, gu, eup, myeon, ri, dong ga'로 적고 그 앞에는 붙임표(-)를 넣는다. (이하同様)
	종로2가	정상	Jongno i-ga	정상		
6	가곡동	정상	Gagok-dong	정상	제3장 제5항	로마자 표기법 용례번호: 12848

로마자 표기법 기초 자료화의 특징은 입력 ‘ㄱ’에 대한 2개의 출력값을 구분자 쉼표를 이용해 하나의 출력으로 표현한 자료가 있는 점이다. 2개의 입력값이 하나의 출력이 되거나 1개의 입력값이 2개의 출력값이더라도 하나의 유형(특히 규범의 경우)을 갖는 경우에는 허용된다. 이와 달리 이름이나 ‘종로2가’와 같은 경우는 출

21) 로마자 표기법 문서에는 ‘월곶[월곶]’으로 발음표기가 함께 적혀 있으나 기초 자료에서는 발음기호를 제거한 철자 표기만 기록한다. 기초 자료에서는 입출력값의 범주가 정보의 다양성보다 우선되기 때문이다.

력값 유형이 다르거나 규범이 아니므로 반드시 2개의 값을 갖는다는 필연성이 없다. 입력 ‘ㄱ’은 로마자 표기법에 있어서 반드시 2개의 값으로 사용된다는 점이 데이터의 변별성을 나타낸다.²²⁾

3.3. 국어 지식 기초 자료 데이터 구조

앞서 살펴본 외래어 표기법, 표준 발음법, 로마자 표기법의 데이터를 통해 어문 규범 정보와 예시 자료를 간단한 데이터 구조로 통합해 활용할 수 있음을 확인했다. 이번 연구에서 제안하는 국어 지식 기초 자료 데이터 구조는 <표 40>과 같다. 국어 지식 기초 자료의 데이터 구조는 단순 검색으로도 활용이 용이하도록 간단한 구조로 구성하기로 했다. 모든 데이터 요소는 기본적으로 단일값을 가지지만 입출력 값의 유형(type)과 규범에 따라 여러 값을 가질 수 있다.

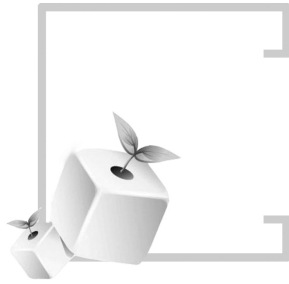
<표 40> 국어 지식 기초 자료 데이터 구조 (json)

1수준	2 수준	타입	예시	비고
index		num	71903	단순 일련번호
kind		str	로마자표기(용례)	자료 출처명(유형 : 용례, 규범)
input		obj		
	input_value	arr(str)	광화문	입력(검색)값
	input_type	str	정상	입력값 유형 (규범, 예외, 허용, 정상, 오류 ²³⁾)
output		obj		
	output_value	arr(str)	Gwanghwanmun	출력값
	output_type	str	규범	출력값 유형 (규범, 예외, 허용, 정상)
evidence		obj		자료 출처상 위치 분류
	category_large	str	로마자 표기법	대분류
	category_medium	str	null	중분류
	category_small	str	명소	소분류
	category_detail	str	로마자 표기법 용례번호 : 11809	자료(규범) 설명 또는 url

향후 본 연구에서 구축한 어문 규범들 외에 규범 표기, 순화어, 표준화 용어 등 다양한 사례 중심 규범 자료들을 통합해 활용할 수 있을 것으로 기대된다. 다만 한글 맞춤법과의 연계를 위해서는 추가적인 연구가 필요하다.

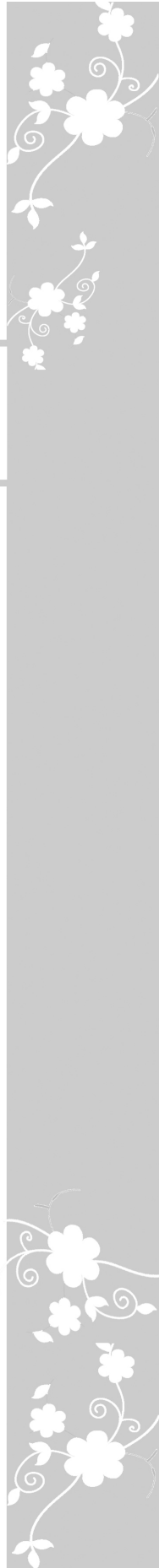
22) 다만, 자료를 활용할 때 사용하는 소프트웨어에 따라 본 자료에서 사용한 침표 등이 예약어로 사용되는 경우가 있으므로 구분자를 입출력값으로 사용할 필요가 있으면 전처리를 통해 분리하는 것을 권장한다.

23) 규범은 규범자료 상의 예시, 예외는 규범에서 예외로 규정한 것, 허용은 규범에서 허용한다고 표현한 것, 오류는 빈번하게 발생하는 오류 형태이고, 정상은 국어원 사전, 용례 자료 등에서 권고한 값 등을 의미한다.



제 4 장

2024년 구축 결과 개선 및 보완



4.1. 글쓰기 첨삭 지원을 위한 기초 자료 개선

4.1.1. 내용 보완

1) 오류 유형 목록 수정/추가

기초 자료의 데이터 설명력이 개선되도록 유형을 총 89개에서 146개 유형으로 확대하였다. 표준어-비표준어 관계(구렛나루-구레나룻)나, 혼동되는 말이지만 한글 맞춤법 57항(혼동되는 말의 구별)에 제시되지 않은 사례 등 어문 규범으로 제시되어 있지 않지만, 표준국어대사전에서 식별할 수 있는 오류는 오류 유형 ‘표준어 사정 원칙(표준국어대사전)’으로 기술하였다.

<추가된 오류 유형 예시>

· 한글 맞춤법 제56항

‘-더라, -던’과 ‘-든지’는 다음과 같이 적는다.

1. 지난 일을 나타내는 어미는 ‘-더라, -던’으로 적는다.(ㄱ을 취하고, ㄴ을 버림.)

- ㄱ: 지난겨울은 몹시 춥더라., 깊던 물이 얕아졌다., 그렇게 좋던가?, 그 사람 말 잘하던데!, 얼마나 놀랐던지 몰라.

- ㄴ: 지난겨울은 몹시 춥드라., 깊은 물이 얕아졌다., 그렇게 좋든가?, 그 사람 말 잘하든데!, 얼마나 놀랐든지 몰라.

2. 물건이나 일의 내용을 가리지 아니하는 뜻을 나타내는 조사와 어미는 ‘(-)든지’로 적는다.(ㄱ을 취하고, ㄴ을 버림.)

- ㄱ: 배든지 사과든지 마음대로 먹어라., 가든지 오든지 마음대로 해라.

- ㄴ: 배던지 사과던지 마음대로 먹어라., 가던지 오던지 마음대로 해라.

· 표준어 사정 원칙 제25항

의미가 똑같은 형태가 몇 가지 있을 경우, 그중 어느 하나가 압도적으로 널리 쓰이면, 그 단어만을 표준어로 삼는다.(ㄱ을 표준어로 삼고, ㄴ을 버림.)

- ㄱ: -게끔, 겸사-겸사, 고구마, 고치다, 골목-쟁이, 광주리, 괴통, 국-물, 군-표, 길-잡이, 까치-발, 꼬창-모, 나룻-배, 납-도리, 농-지거리, 다사-스럽다, 다오, 담배-꽁초, ...

- ㄴ: -게시리, 겸지-겸지/겸두-겸두, 참-감자, 낫우다, 골목-자기, 광우리, 호구, 멀-국/말-국, 군용-어음, 길-앞잡이, 까치-다리, 말뚝-모, 나루, 민-도리, 기롱-지거리, 다사-하다, 다구, 담배-꼬투리/담배-꽁치/담배-꽁추, ...

2) 지침에 실제 사례를 제시하여 작업자의 이해 증진

- 실제 샘플에 있는 형식 및 메타언어를 그대로 제시

예) <패턴1>

Input: “[{오류가 있는 표현/교정한 표현}]”가운데 올바른 것은?

Output: “[교정한 문장]”

<예시1>

Input: “그는 {휴게소/휴게소}에서 김밥을 사 먹었다.” 가운데 올바른 것은?

Output: “그는 휴게소에서 김밥을 사 먹었다.”

<패턴2>

Input: 다음 문장에서 어문 규범에 부합하지 않는 부분을 찾아 고치시오.

“[오류가 있는 문장]”

Output: “[교정한 문장]”

<예시2>

Input: 다음 문장에서 어문 규범에 부합하지 않는 부분을 찾아 고치시오.

“이번 과제는 지난 과제를 짜집기해 제출하고 말았다.”

Output: “이번 과제는 지난 과제를 짜집기해 제출하고 말았다.”

그 외에 원문을 각색한 경우 이력이 남도록 각색 여부와 출처 기록하고 전체적인 정서법과 형식에 일관성이 부족한 부분을 보완했다.

4.1.2. 형식 일관성 보완

2024년 기초 자료 지침의 ‘[부록]오류 유형 및 관련 규정’에서 어문 규범에 제시된 형식 기준과 가독성을 고려해 다음 표를 기준으로 형식을 통일하였다.

<표 41> 기초 자료 구축 지침중 [부록]오류 유형 및 관련 규정의 세부 규정 작성 기준

번호	유형	내용	예시
1	내용	어문 규범 내용을 수정하지 않고 그대로 적는다.	줄어질 경우 (X) // 줄어질 적 (O)
2	예시의 열거	(한글 맞춤법, 표준어 규정, 외래어 표기법) 예시를 열거할 때는 쉼표를 사용한다.	- 신여성(新女性), 공염불(空念佛), 남존여비(男尊女卑)
		(문장 부호) 예시를 열거할 때는 /를 사용	(1) 감탄문이나 감탄사의 끝에 쓴다.

		한다.	- 이거 정말 큰일이 났구나! / 어머!
3	조항 내 소괄호	마침표 뒤에 소괄호가 나올 경우, 띄어쓰기나 줄 바꿈 없이 소괄호를 바로 붙여서 적는다.	다음 단어는 모음이 단순화한 형태를 표준어로 삼는다.(ㄱ을 표준어로 삼고, ㄴ을 버림.)
4	붙임	표기는 ‘[붙임]’으로 통일한다.	한자음 ‘녀, 뇨, 뉴, 니’가 단어 첫머리에
5		[붙임]은 위 문장과 한 줄 띄우고 적는다.	올 적에는, 두음 법칙에 따라 ‘여, 요, 유, 이’로 적는다.(ㄱ을 취하고, ㄴ을 버림.)
6		[붙임2], [붙임3]…이 이어질 때는 [붙임1] 아래 한 줄 띄우지 않고 줄 바꿈만 하여 적는다.	- ㄱ: 여자(女子), 연세(年歲), 요소(尿素), 유대(紐帶), 이토(泥土), 익명(匿名) - ㄴ: 녀자, 년세, 뇨소, 뉴대, 니토, 닉명 [붙임 1] 단어의 첫머리 이외의 경우에는 본음대로 적는다. - 남녀(男女), 당뇨(糖尿), 결뉴(結紐), 은닉(隱匿) [붙임 2] 접두사처럼 쓰이는 한자가 붙어서 된 말이나 합성어에서, 뒷말의 첫소리가 ‘ㄴ’ 소리로 나더라도 두음 법칙에 따라 적는다. - 신여성(新女性), 공염불(空念佛), 남존여비(男尊女卑)
7		단, 내용이 1, 2, 3…으로 구분되어 이미 줄 바꿈 되어 있는데 해당 번호에만 해당하는 붙임 사항일 경우 한 줄 띄우지 않고 줄 바꿈만 하여 적는다.	(2) 짝을 이루는 어구들 사이에 쓴다. - 한(韓)·이(伊) 양국 간의 무역량이 늘고 있다. (3) 공통 성분을 줄여서 하나의 어구로 묶을 때 쓴다. - 상·중·하위권, 금·은·동메달, 통권 제 54·55·56호 [붙임] 이때는 가운뎃점 대신 쉼표를 쓸 수 있다. - 상, 중, 하위권, 금, 은, 동메달, 통권 제 54, 55, 56호
8	다만,	위 문장과 한 줄 띄우지 않고 줄 바꿈만 하여 적는다.	한자음 ‘녀, 뇨, 뉴, 니’가 단어 첫머리에 올 적에는, 두음 법칙에 따라 ‘여, 요, 유, 이’로 적는다.(ㄱ을 취하고, ㄴ을 버림.) - ㄱ: 여자(女子), 연세(年歲), 요소(尿素), 유대(紐帶), 이토(泥土), 익명(匿名)

			<p>- ㄴ: 녀자, 년세, 뇨소, 뉴대, 니토, 닉명 다만, 다음과 같은 의존 명사에서는 ‘냐, 녀’ 음을 인정한다.</p> <p>- 냥(兩), 냥쫙(兩-), 년(年)(몇 년)</p>
9		단, 내용이 1, 2, 3...으로 구분되어 이미 줄바꿈 되어 있는데 해당 번호에만 해당하는 사항이 아닌, 전체에 해당하는 사항일 경우에 한해 한 줄 띄우고 적는다.	어간에 ‘-이’나 ‘-음/-ㅁ’이 붙어서 명사로 된 것과 ‘-이’나 ‘-히’가 붙어서 부사로 된 것은 그 어간의 원형을 밝히어 적는다.
10	숫자 1. 2....	숫자 1. 2. ...로 항목이 구분될 때는 위 문장과 한 줄 띄우고 적는다.	1. ‘-이’가 붙어서 명사로 된 것 - 길이, 깊이, 높이, 다듬이, 땀받이, 달맞이, 먹이, 미달이, 벌이, 벼훅이, 살림살이, 쇠붙이
11	붙임표 (-)	들여쓰기 없이 제일 앞에 바로 쓰고, 내용은 붙임표 뒤에 한 칸 띄우고 적는다.	<p>2. ‘-음/-ㅁ’이 붙어서 명사로 된 것 - 걸음, 묶음, 믿음, 얼음, 엮음, 울음, 웃음, 졸음, 죽음, 앓</p> <p>다만, 어간에 ‘-이’나 ‘-음’이 붙어서 명사로 바뀐 것이라도 그 어간의 뜻과 떨어진 것은 원형을 밝히어 적지 아니한다.</p> <p>- 굽도리, 다리[髭], 목거리(목병), 무녀리, 코끼리, 거름(비료), 고름[膿], 노름(도박)</p>
12	숫자 (1) (2) ...	위 문장과 한 줄 띄우지 않고 줄 바꿈만 하여 적는다.	<p>사이시옷은 다음과 같은 경우에 받치어 적는다.</p> <p>1. 순우리말로 된 합성어로서 앞말이 모음으로 끝난 경우 (1) 뒷말의 첫소리가 된소리로 나는 것 - 고랫재, 꺾밥, 나룻배, 나뭇가지, 냇가, 댕가지, 뒷갈망, 땃돌, 머릿기름, 모깃불, 못자리, 바닷가, 뱃길, 뱃가리, 부싯돌, 선짓국, 쇠조각, 아랫집, 우렁잇속, 잇자국, 잣더미, 조갯살, 찻집, 찻바퀴, 킷값, 핏대, 햇별, 햇바늘</p> <p>(2) 뒷말의 첫소리 ‘ㄴ, ㅁ’ 앞에서 ‘ㄴ’ 소리가 덧나는 것 - 땃나물, 아랫니, 텃마당, 아랫마을, 땃머리, 잇몸, 깻묵, 냇물, 빗물</p>

13		단, 한글 맞춤법 부록(문장 부호)의 경우, 조항이 1. 2.가 아닌 (1), (2)로 구분되어 있으므로 기호를 그대로 쓰되 위 문장과 한 줄 띄우고 쓴다.	<p>(1) 열거할 어구들을 일정한 기준으로 묶어서 나타낼 때 쓴다.</p> <p>- 민수·영희, 선미·준호가 서로 짝이 되어 윗놀이를 하였다.</p> <p>(2) 짝을 이루는 어구들 사이에 쓴다.</p> <p>- 한(韓)·이(伊) 양국 간의 무역량이 늘고 있다.</p>
14	부록(문장 부호)- 띄어쓰기	어문 규범에 띄어쓰기 관련 내용이 제시되어 있는 경우, 해당 내용을 그대로 적는다.	<p>■ 가운뎃점의 띄어쓰기: 가운뎃점은 앞말과 뒷말에 붙여 쓴다.</p>

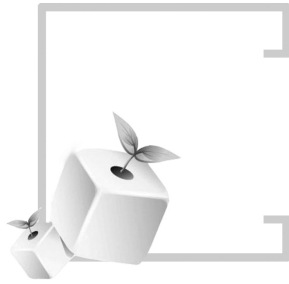
4.2. 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치 개선

“2024년 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치”의 개선은 오류 유형 목록 가운데 ‘표기’ 오류 유형을 기초 자료 지침과 통일성이 있도록 형식과 규범 정보를 수정하고 말뭉치에 포함된 작문 자료(input)에서 나타나는 개인정보처리 지침을 추가해 보완했다. 구축 데이터는 지침 준수 여부와 일관성에 중점을 두고 개선을 진행하였다.

〈표 42〉 ‘표현-문단’의 추가 오류 유형

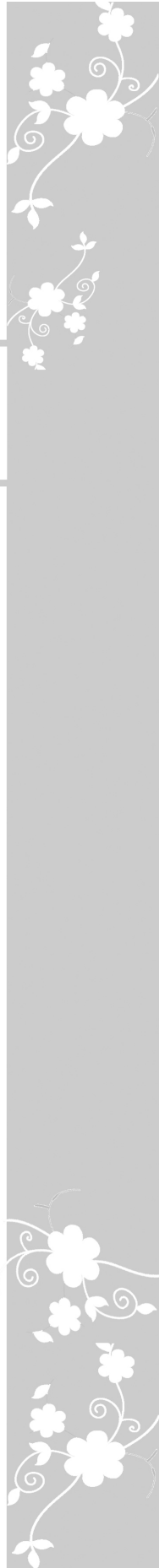
대분류	중분류	소분류	규정 내용
표현-문단	부적절한 높임	상대 높임	<p>글은 장르와 독자 등에 따라 종결어미의 상대 높임 등급이 결정된다. 이렇게 결정된 상대 높임 등급은 하나의 문서 내에서 동일하게 나타나는 것이 일반적이다.(ㄱ은 적절하지 않은 문장, ㄴ은 적절한 문장)</p> <p>ㄱ. 개나리가 곱게 피었습니다. 봄바람이 따뜻하다.</p> <p>ㄴ. 개나리가 곱게 피었습니다. 봄바람이 따뜻합니다.</p>
표현-문단	부적절한 조사	보조사	<p>보조사는 문장에서 고유한 의미를 더해 주므로 의미와 기능에 맞게 사용해야 한다.(ㄱ은 적절하지 않은 문장, ㄴ은 적절한 문장)</p> <p>ㄱ. 거란은 대군을 이끌고 고려를 침입해 왔다. 그러나 거란은 서희의 외교 활동으로 스스로 물러났다.</p> <p>ㄴ. 거란이 대군을 이끌고 고려를 침입해 왔다. 그러나 거란은 서희의 외교 활동으로 스스로 물러났다.</p> <p>새로운 정보를 나타낼 때는 ‘이/가’가 이전의 정보를 나타낼 때는</p>

			‘은/는’이 쓰이므로 ‘거란은’은 ‘거란이’로 수정해야 한다.
표현-문단	부적절한 호응 및 시제	주어와 서술어의 호응	<p>주어가 생략되어 서술어와 호응이 되지 않거나 주어와 서술어의 호응이 맞지 않는 경우 부적절한 문장이 된다. 주어와 서술어는 각각 문장의 앞과 뒤에서 실현되는 경우가 많기 때문에 중간에 들어간 다른 성분들에 이끌려 서로 호응이 잘 이루어지지 않는 경우가 많다. 이때, 문장의 주어와 서술어만을 따로 연결해 봄으로써 의미 관계가 제대로 이루어지는지 확인할 수 있다. 주어가 생략된 경우에도 마찬가지이다.(ㄱ은 적절하지 않은 문장, ㄴ은 적절한 문장)</p> <p>ㄱ. 백일홍의 특징은 줄기가 곧게 자라는 것이다. 그리고 높이가 1미터쯤 된다.</p> <p>ㄴ. 백일홍의 특징은 줄기가 곧게 자라는 것이다. 그리고 높이가 1미터쯤 되는 것이다.</p>
표현-문단	부적절한 호응 및 시제	부적절한 시제	<p>하나의 문서 내에서, 내용상 시제 이동이 없는 경우에는 문장마다 동일한 시제를 표현하는 것이 일반적이다.(ㄱ은 적절하지 않은 문장, ㄴ은 적절한 문장)</p> <p>ㄱ. 나는 사람 간의 소통이 중요하다고 생각한다. 소통을 통해 다른 사람을 이해할 수 있다고 생각했다.</p> <p>ㄴ. 나는 사람 간의 소통이 중요하다고 생각한다. 소통을 통해 다른 사람을 이해할 수 있다고 생각한다.</p>
표현-문단	부적절한 호응 및 시제	대용 표현	<p>대용 표현은 문장에서 적절하게 제시해야 한다.(ㄱ은 적절하지 않은 문장, ㄴ은 적절한 문장)</p> <p>ㄱ. 연합군은 적군과 여러 차례 전투를 벌였다. 그리하여 이를 크게 격파하였다.</p> <p>ㄴ. 연합군은 적군과 여러 차례 전투를 벌였다. 그리하여 적군을 크게 격파하였다.</p> <p>‘적군’을 ‘이를’로 대용하는 것은 자연스럽지 않으므로 ‘적군을’로 수정해야 한다.</p>



제 5 장

결론



5.1. 사업 결과 요약

5.1.1. 지시문 기반 첨삭 지원 말뭉치 구축 결과

지시문 기반 첨삭 지원 말뭉치는 국어 상담 자료 분석을 통해 872개의 질문 유형을 도출하고 각 질문 유형별로 50~65개의 질문-답변쌍을 작성해 총 50,472개의 질문-답변쌍을 구축했다.

국어 지식 기초 자료는 3,000건의 구축을 목표로 정했으나, 자료의 실질적 활용도를 고려하여 공개된 각종 어문 규범의 예시 자료를 모아 일관된 형식으로 국어 지식을 정리하는 것으로 자료의 성격이 정의되면서 표준 발음법, 로마자 표기법, 외래어 표기법과 관련 용례 자료 84,134건을 구축해 당초 목표를 크게 초과 달성하였다.

<표 43> 지시문 기반 첨삭 지원 말뭉치 구축 결과

구축 자료	계 획	구축량	
지시문 첨삭 지원 설계 말뭉치	1,000개 유형	872개 유형	4,360개 질문-답변쌍
지시문 기반 첨삭 지원 말뭉치	50,000개 질문-답변쌍	872개 유형	50,472개 질문-답변쌍
국어 지식 기초 자료	3,000건	84,134	표준발음법(428) +로마자 표기법 및 용례(12,906) +외래어 표기법 및 용례(70,800)

5.1.2. 말뭉치 검증 결과

<표 44> 지시문 기반 첨삭 지원 말뭉치 AI모델 적용(학습) 결과

모 델	Base-line	평가 유형	통합 평균	사실적 정확성	근거 및 충실도	명확성 및 완전성	유창성 및 구조	설명의 효율성
Qwen_Qwen3-4B-Instruct-2507	2.18	Indist	3.7905	3.4128	3.1306	3.5990	4.7210	4.0890
		Outdist	3.7017	3.2638	3.0678	3.4742	4.6956	4.0069
google_gemma-3-4b-it	-	Indist	3.6975	3.2680	3.0070	3.4898	4.7096	4.0129
		Outdist	3.6046	3.1415	2.9253	3.3568	4.6805	3.9189
meta-llama_Llama-3.2-3B-Instruct	-	Indist	3.6067	3.1666	2.9254	3.3813	4.6549	3.9054
		Outdist	3.4543	2.9631	2.7545	3.1901	4.5967	3.7670
naver-hyperclova_HyperCLOVA X-SEED-Vision-Instruct-3B	2.22	Indist	3.5237	3.0706	2.8071	3.2705	4.6223	3.8481
		Outdist	3.3962	2.9314	2.6803	3.1101	4.5426	3.7167

5.2. 발전 제언

구축된 지시문 기반 첨삭 지원 말뭉치는 국어학 및 인공지능 전문가(국어학 5인, 인공지능 2인)를 통해 구축 목표, 설계, 지침, 말뭉치 샘플, 활용성 등에 대해 자문을 받았다. 자문 의견과 구축 간 연구진 논의사항을 정리해 발전 제언으로 정리하였다.

5.2.1. 데이터 고도화

5.2.1.1. 서비스 고도화를 위한 데이터 다양성 및 규모 확장

지시문 기반 첨삭 지원 말뭉치는 질문과 답변을 중심으로 구축되었다. 그러나 종합적인 국어생활 상담을 위한 인공지능 개발을 위해서는 보다 다양한 유형의 국어지식 데이터로의 확장을 고려해야 한다. 특히, 2024년에 수행한 글쓰기 지시문 말뭉치와 같이 일반인이 글을 써 나가면서 발생하는 여러 상황(교정, 문체 변환, 요약, 재진술 등)에 대한 원문과 수정문이 쌍을 이루는 대규모 데이터 구축이 필요하다.

더불어 일상에 더 가까운 국어생활 지원을 위해 트렌드성 용어 및 신조어 반영도 중요하다. '우리말샘' 등에 등재되는 신조어나 사회적 이슈가 되는 언어 현상(예: 밤양갱 등)을 주기적으로 추적하여 질문 유형에 추가되는 체계를 마련해야 한다.

또한, 대화 중심의 인공지능 서비스의 고도화를 위해서는 상담 자료에 나타난 대화 유형을 고도화하여 멀티턴(Multi-turn) 대화를 강화하기 위한 연구도 필요하다. 1차 질문과 2차 질문을 결합한 형태의 데이터를 추가 생성하거나, 답변 후 이용자에게 추가 정보를 제공할지 묻는 등 인공지능과의 대화 맥락을 고려해 이용자의 참여를 강화하는 질문-답변 데이터 설계 보완도 고도화 방안이 될 수 있다.

5.2.1.2. 모호한 규정에 대한 대응 체계 및 신뢰성 강화

본 연구에서는 신뢰할 수 있는 답변을 제시하기 위해 최대한 명확한 어문 규정을 적시하도록 했다. 그러나, 명확한 어문 규정이 없는 경우 '답변 불가'로 처리하

기보다, 유연하고 신뢰성 있는 정보를 제공해야 한다. 규정이 모호한 사안에 대해 전문가(국어학자)의 유추 해석이나 견해를 받아 데이터베이스화하고, 이를 '참고 정보' 형태로 인공지능이 활용하도록 구축함으로써 활용성을 높일 수 있다.

다른 한편으로는 빅데이터 기반 근거 제시 시스템을 마련해 규범이 없는 경우에 대응하도록 할 수 있다. 대규모 말뭉치(뉴스, 공문서 등)에서의 사용 빈도나 통용되는 관례를 근거로 제시하여 이용자의 질문에 대한 답변을 생성하는 데이터 기반 시스템도 활용해야 한다. 정제 전 데이터를 전문가 검토용으로 제공한다면 국어 지식 데이터의 고도화를 위한 선순환적 체계를 마련할 수 있다. 더불어 구축된 말뭉치에 활용된 국어 지식에 대한 공신력 있는 학술 자료와 교육 자료의 출처를 추가해 기반 자료를 공고히 하는 것이 향후 인공지능 시대에 지속성 있는 데이터 공급 체계를 확립하는 데 중요하다.

5.2.1.3. 인공지능 모델의 성능 최적화 및 활용 분야 다각화

최근 인공지능 모델은 보편 지식으로의 거대화와 함께 개별 분야로의 전문성을 확보하는 두 갈래로 발전하고 있다. 따라서 인공지능이 특정 유형의 질문이나 특정 사용자에게 편중되지 않도록 모범 답변의 구성을 다양화해야 한다. 예를 들어 외국인 학습자의 질문에 대한 답변과 공공기관의 문서 업무 중에 발생하는 질문에 대한 답변은 같은 내용이라고 하더라도 서로 다른 문체와 서술이 필요하다. 이러한 부분은 해당 서비스를 구현할 때 프롬프트의 가공, 변경이나 에이전트화를 통해서도 해결할 수 있으나 데이터의 다각화를 통해 보다 효과적으로 구현될 수 있다.

5.2.2. 데이터 규모 확대

최근 인공지능 모델의 고도화, 학습 알고리즘의 성능 향상으로 데이터의 품질보다 대규모의 데이터를 선호하는 경향이 있다. 이는 고품질 데이터 구축이 비용적, 현실적 한계로 제한되는 탓이지 고품질의 데이터가 필요 없다는 것을 의미하는 것은 아니다. 고품질 데이터의 문제는 규모의 한계로 인해 모든 충분한 경우의 수를 포함하지 못하는 다른 측면의 품질 문제를 가지기 때문이다. 그러므로 말뭉치 설

계 단계에서 고품질의 중소 규모 데이터와 상대적 저품질의 대규모 데이터를 함께 구축해 활용성을 높이는 전략을 고려해야 한다.

글로벌 IT 리서치 기관 가트너(Gartner)는 2027년까지 인공지능 학습에 사용되는 데이터의 60% 이상이 합성데이터(Synthetic Data)가 될 것으로 예측하였다. 한편, 최신 보고서 “AI Services Disruptor: Hyper-Synthetic Data”(2025년 6월 발간)에서는 Hyper-Synthetic 데이터를 실제 데이터의 단순한 복제를 넘어, 정교한 시뮬레이션과 복잡한 사회적 행동을 시뮬레이션하는 보다 안전한 데이터로 소개하며 2028년까지 AI에 활용되는 데이터의 80%가 합성데이터가 될 것으로 비중을 확대했다. 그러므로 앞으로는 고품질 말뭉치 구축에만 집중할 것이 아니라 여러 품질 수준의 데이터를 염두에 두고 새로운 관점의 합성 데이터 생성, 활용 방안을 모색해야 한다.

5.2.3. 데이터 검증과 활용

말뭉치 품질 검증 결과는 본 사업에서 구축된 데이터가 인공지능 모델이 한국어 어문 규범을 파악하고 올바른 답변을 생성하는 데 유효한 역할을 수행함을 보여준다. 학습 전 2점대 수준이었던 모델들의 성능이 본 데이터를 학습한 후 3점대 후반에서 4점대까지 향상된 점은 데이터의 학습 효과를 객관적으로 뒷받침한다고 볼 수 있다. 특히 모델들이 규정된 답변 형식을 정확히 준수하며 유창하게 답변을 구성하는 역량을 갖추었음을 확인하였다.

이러한 검증 결과를 바탕으로 향후 서비스 고도화를 위한 제언은 다음과 같다.

첫째, 모델의 기초 역량에 따른 성능 차이를 고려한 모델 선정이 필요하다. 검증 결과에 따르면, 모델 자체의 추론 능력이 뛰어날수록 어문 규범 안내의 정확도 또한 비례하여 상승하는 경향을 보였다. 따라서 향후 실제 서비스 구현 시에는 한국어 규범 자료의 학습 효율이 높은 고성능 베이스 모델을 선별하여 활용하는 것이 효과적이다.

둘째, 학습데이터의 범위 밖(Out-of-distribution) 대응 능력 강화가 요구된다. 상위 모델일수록 학습하지 않은 질문에 대해서도 안정적인 성능을 유지하였으나, 여전히 전체 모델에서 학습데이터 내 질문에 비해 성능이 다소 낮아지는 현상이 관찰되었다. 이를 보완하기 위해 더욱 다양하고 폭넓은 유형의 어문 규범 사례를

지속적으로 확충하여 모델의 일반화 능력을 높여야 한다. 혹은 국립국어원의 표준 국어대사전이나 어문 규범 해설서 등 공신력 있는 외부 자료를 실시간으로 찾아 답변에 활용하는 RAG 시스템을 구축함으로써, 데이터 일반화 문제를 보완하고 답변의 사실적 정확성을 높이는 방안도 모색해야 한다.

셋째, 한국어 특화 모델의 정밀도 향상을 위한 연구가 지속되어야 한다. 이번 실험에서 한국어 특화 모델이 외산 모델 대비 다소 낮은 성능을 보인 것은 복잡한 문법 이치를 논리적으로 풀이하는 미세 조정 과정에서 보완할 점이 있음을 시사한다. 국어 전문 지식을 더욱 정교하게 학습시킨 한국어 특화 모델의 개발과 고도화 작업이 병행되어야 할 것으로 판단된다.

결론적으로 본 과업을 통해 구축된 말뭉치는 인공지능 기반의 상담 서비스 구축을 위한 기반 자료로서 충분한 가치를 지니고 있다고 판단되며, 향후 제시된 개선 방향을 바탕으로 모델의 논리적 추론 능력을 강화한다면 더욱 신뢰도 높은 국어 정보 서비스를 제공할 수 있을 것으로 보인다.

참고 문헌

- Sinclair, J. McH. (2004) ‘Corpus and Text: Basic Principles’, in M. Wynne (ed.) *Developing Linguistic Corpora: A guide to good practice*. <https://users.ox.ac.uk/~martinw/dlc/chapter1.htm#section4>.
- Ben, Fieselmann. (2025), *AI Services Disruptor : Hyper-Synthetic Data*, Gartner Research.
- Lim, Y., Chu, H., Kim, S., Jang, J. Y., Jung, M., & Shin, S. (2023). Korean Instruction Tuning Dataset. In *Annual Conference on Human and Language Technology* (pp. 591-595). Human and Language Technology.
- Zhou, J., Lu, T., Mishra, S., Brahma, S., Basu, S., Luan, Y., Zhou, D., Hou, L. (2023). Instruction-following evaluation for large language models. *arXiv preprint arXiv:2311.07911*.
- Longpre, S., Hou, L., Vu, T., Webson, A., Chung, H. W., Tay, Y., Zhou, D., Quoc V. Le, Zoph, B., Wei, J., Roberts, A. (2023, July). The flan collection: Designing data and methods for effective instruction tuning. In *International Conference on Machine Learning* (pp. 22631-22648). PMLR.
- Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., Efrat, A., Yu, P., Yu, L., Zhang, S., Ghosh, G., Lewis, M., Zettlemoyer, L., Levy, O. (2024). Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36.
- Chakrabarty, T., Padmakumar, V., & He, H. (2022). Help me write a
- Raheja, V., Kumar, D., Koo, R., & Kang, D. (2023). CoEditIT: Text editing by task-specific instruction tuning. *arXiv preprint arXiv:2305.09857*.
- Raheja, V., Alikaniotis, D., Kulkarni, V., Alhafni, B., & Kumar, D. (2024). medit: Multilingual text editing via instruction tuning. *arXiv preprint arXiv:2402.16472*.

지시문 기반 첨삭 지원 말뭉치 구축 지침

1. 구축 목표

본 구축 지침은 국립국어원 <2025년 지시문 기반 첨삭 지원 말뭉치 연구 분석>에서 인공지능(AI) 기반 국어생활 상담·첨삭 서비스 구현을 위한 지시문 기반 말뭉치를 구축하기 위해 작성되었다. 이 말뭉치는 한국어 어문 규범 및 언어 사용과 관련된 질문에 대해 인공지능이 정확하고 일관된 답변을 생성할 수 있도록 하는 것을 목표로 한다. 이를 위해 이용자가 실제로 할 수 있는 한국어 어문 규범 및 언어 사용과 관련된 질문과 그 질문에 대한 모범 답변을 ‘이용자 질문-국어 지식-모범 답변’ 형태로 구축한다.

2. 구축 형식과 규모

2.1. 구축 형식

이 말뭉치는 ‘이용자 질문’과 ‘모범 답변’으로 구성된 대화 형식으로 구축한다. 각 대화 쌍은 이용자가 실제로 챗봇을 이용하는 상황을 가정하여 실제 국어상담 자료를 분석하여 ‘이용자 질문’과 작성하고 이용자 질문에 답변하기 위한 ‘국어 지식’을 구축한 후 이를 바탕으로 ‘모범 답변’을 작성한다. 지시문 기반 말뭉치의 기본 구축 요소는 아래 <표 1>과 같다. 각 기본 구성 요소는 3장에서 자세한 구축 방법과 함께 제시한다.

<표 1> 지시문 기반 말뭉치의 기본 구성 요소

기본 구성 요소	내용
유형 Index	질문 유형별 고유 번호
예시 Index	유형별 샘플 번호(1~5)
이용자 질문 유형	이용자 질문의 유형
출처	참조 규정명 또는 자료명
국어 지식	답변 작성 시 근거로 삼은 규범·사전·문법 정보
이용자 질문	이용자 예상 질문
모범 답변	이용자 질문에 대응하는 모범 답변
꼬리 질문 여부	앞선 질문에 연결되는 꼬리 질문 여부

2.2. 구축 규모

지시문 기반 말뭉치는 국립국어원에서 제공한 국어생활 상담 자료를 분석하여 총 872개의 질문 유형을

도출하고 각 유형별로 5개 쌍의 샘플 대화를 작성한다. 그중 3개 대화 쌍은 단일 질문으로, 2개 대화 쌍은 꼬리질문 유형으로 구성한다. 꼬리질문 유형은 앞선 질문에 대해 이용자가 추가 질문을 할 경우를 상정하여 2개의 대화 쌍이 이어지도록 구성하는 질문 유형이다. 한 질문 유형에 대해 총 50개 대화 쌍을 구축하는데 그중 5개는 전문가가 작성하고 나머지 45개는 전문가가 구축한 대화 쌍을 기반으로 인공지능이 생성한 결과물을 수정하여 구축한다. 구축 형식과 규모는 아래 <표 2>와 같다.

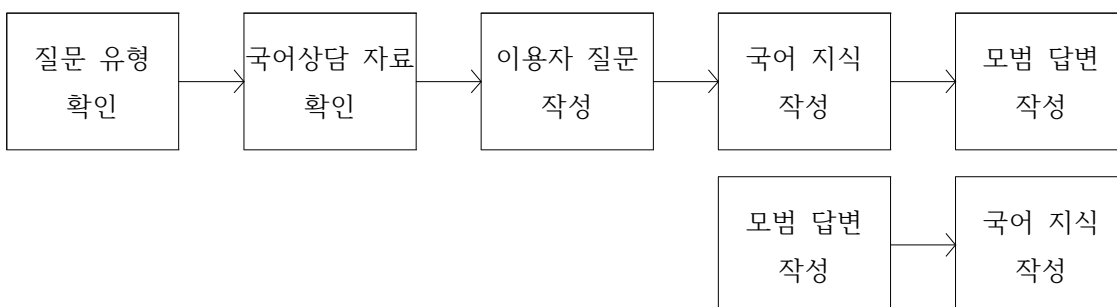
<표 2> 구축 형식에 따른 구축 규모

번호	구축 형식		구축 규모	작성 방법
1	사용자 질문 유형		872개	국어원 제공 국어생활 상담 자료 분석
2	지시문		1개	LLM을 활용하여 작성 후 성능 평가
3	국어 지식		872개	한국어 어문 규정, 표준국어대사전 등을 활용하여 이용자 질문에 답변하는 국어 지식 작성
4	샘플1	이용자 질문	[대화 쌍 1개] 유형 1,000개 × 30개 = 30,000개	총 50,000개 작성(샘플 5,000개 포함) 이용자 질문과 챗봇 답변을 작성
		모범 답안		
	샘플2	이용자 질문		
		모범 답안		
	샘플3	이용자 질문		
		모범 답안		
	샘플4	이용자 질문1	[대화 쌍 2개] 유형 1,000개 × 20개 = 20,000개	
		모범 답안1		
이용자 질문2				
모범 답안2				

3. 구축 방법

본 장에서는 지시문 기반 말뭉치의 구체적인 구축 절차와 구축 방법을 제시한다. 말뭉치 구축은 국립국어원에서 제공한 국어생활 상담 자료를 기반으로 하여 이용자의 실제 질문을 반영한 ‘이용자 질문-모범 답변’ 대화 쌍을 작성하는 과정이다. 말뭉치 구축에 필요한 기반 자료와 작성 양식에 대해 제시하고 각 말뭉치 구성 요소별 자세한 작성 방법을 설명한다. 대화 쌍 구축 과정은 아래와 같은 단계를 거친다. 이용자 질문과 답변의 실제성, 정확성, 일관성을 유지하기 위해 구축 지침을 작성하였다.

<표 3> 대화 쌍 구축 절차



먼저, 개별 작업자에게 할당된 질문 유형을 확인한다. 국어상담 자료에서 질문 유형에 해당하는 실제 이용자 질문과 답변을 확인한다. 실제 이용자 질문을 바탕으로 이용자 예상 질문을 작성한다. 이용자 예상 질문에 답변하기 위한 국어 지식을 어문 규범과 사전 등에서 선별하여 작성한다. 마지막으로 국어 지식을 바탕으로 모범 답변을 작성한다. 경우에 따라서는 모범 답변을 먼저 작성하고 모범 답안에 활용된 국어 지식을 작성할 수 있다.

<표 4> 작성 단계별 작업 내용

	작성 단계	작업 내용 요약
①	이용자 질문 유형 확인	이용자 질문 유형 확인
②	국어생활 상담 자료 확인	국어생활 상담 자료에서 해당 오류 유형을 검색하여 실제로 나타난 질문 유형과 답변을 살핌.
③	이용자 질문 작성	국어생활 상담 자료에 나타난 질문 유형을 참고하여 실제로 이용자가 할 질문을 작성함.
④	출처 및 국어 지식 작성	참고한 어문 규정이나 자료를 제시함. 참고 규정이나 자료를 최대한 제시하되, 참고 규정이나 자료를 찾을 수 없다면 해당 셀에는 'x(소문자)'로 제시함.
⑤	모범 답변 작성	국어생활 상담 자료의 대답과 어문 규정, 표준국어대사전 등을 참고하여 답변을 작성함.
⑥	점검표의 항목 점검	첨부된 점검표의 점검 항목을 점검하여 작성한 내용을 점검하고 오류가 있으면 수정함.

답변을 작성한 후에는 아래 점검표의 점검 항목을 점검해야 한다.

<표 5> 답변 작성 후 점검 항목

순번	점검 항목
1	답변의 시작을 인사말로 시작하는가? 예) 안녕하세요?
2	질문 내용에 대해서 정확하게 언급하는가? 예) 안녕하세요? [질문 주제]에 관한 질문에 답변드리겠습니다.
3	(꼬리질문이 있을 경우) 앞선 질문과 연결됨을 표현하였는가? 예) 앞선 질문에서 말씀드린 바와 같이
4	핵심 규범이나 원칙이 명확히 서술되었는가?
5	문법/어휘적 근거가 설명되었는가?
6	내용이 일반 사용자에게 이해하기 쉬운가? 지나친 전문 용어 없이 학교 문법 수준으로 설명되었는가?
7	예시가 필요한 경우 적절한 예시가 포함되었는가?
8	예외 상황이나 참고 사항이 필요한 경우 적절하게 안내되었는가?
9	문장이 자연스럽고 맞춤법에 맞는가?
10	답변이 지나치게 길지 않게 요점을 전달하는가?
11	‘감사합니다.’, ‘고맙습니다.’, ‘참고하시기 바랍니다.’로 끝맺었는가?

3.1. 말뭉치 구축에 활용되는 기반 자료

지시문 기반 대화 자료 말뭉치 구축에 활용되는 기반 자료는 총 4개로 다음과 같다.

<표 6> 자료 구축에 제공되는 파일 목록

구분	파일명	파일 내용
이용자 질문 유형	이용자 질문 유형_0월 0째 주.xlsx	<ul style="list-style-type: none"> 작업자별 담당 이용자 질문 유형 제시 질문 유형 index, 이용자 질문 유형, 국어 상담 자료 예시 제시
자료 구축 양식	지시문 기반 말뭉치_작성 양식.xlsx	<ul style="list-style-type: none"> 대화 자료 구축 양식 해당 양식을 이용하여 대화 자료 구축
국어생활 상담 자료	국어생활 상담 자료_전체.xlsx	<ul style="list-style-type: none"> 국립국어원에서 제공한 국어생활 상담 자료 전체 (2021년 11월~2025년 5월)
	온라인가나다_상담사례모음.xlsx	<ul style="list-style-type: none"> 온라인가나다의 상담 사례 모음으로 자주 하는 질문 2,452개
구축 지침	구축 지침.hwp	<ul style="list-style-type: none"> 구축 지침

먼저, 이용자 질문 유형 파일(이용자 질문 유형_0월 0째 주.xlsx)에서 각 작업자에게 할당된 질문 유형을 확인하고 국어생활 상담 자료(국어생활 상담 자료_전체.xlsx, 온라인가나다_상담사례모음.xlsx)에서 질문 유형과 관련된 실제 이용자 질문과 답변을 확인한다. 국어생활 상담 자료를 바탕으로 이용자 예상 질문과 모범 답변을 작성한다. 자료 구축 양식(지시문 기반 말뭉치_작성 양식.xlsx)에 예상 질문과 모범 답변을 구축 지침에 따라 작성한다.

3.2. 대화 자료 형식 및 작성 양식

3.2.1. 대화 자료 형식

대화 자료 구축 형식의 예시를 제시하면 아래 <표 7>과 같다. ‘이용자 질문 유형’, ‘국어 지식’, ‘이용자 질문’, ‘모범 답변’을 작성한다. 하나의 ‘이용자 질문 유형’에는 5개의 ‘이용자 질문-모범 답변’ 대화 쌍을 구축한다. 3개 대화 쌍은 단일 질문으로, 2개 대화 쌍은 꼬리질문 유형으로 구성한다.

<표 7> 인공지능 기반 국어생활 상담·탐색 챗봇 구현을 위한 대화 자료 구축 예시

구분	내용
이용자 질문 유형	동사 '못쓰다' 띄어쓰기
출처	한글 맞춤법 제5장 제2절 제46항, 표준국어대사전
국어 지식	단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다.
	'못쓰다'는 '(((‘못쓰게’ 꼴로 쓰여)) 얼굴이나 몸이 축나다', '(((주로 ‘-으면’, ‘-어서’와 함께 쓰여)) 옳지 않다. 또는 바람직한 상태가 아니다.'의 의미로 쓰인

예 시		다.
	이용자 질문(1)	‘못 쓰다’는 언제 붙여 쓰나요?
	모범 답변(1)	안녕하세요? ‘못쓰다’의 띄어쓰기에 대해 답변드립니다. ‘못쓰다’는 ‘얼굴이 상하다’, ‘행동이 바람직하지 않다’, ‘옳지 않다’와 같은 의미로 쓰일 때 ‘못쓰다’와 같이 붙여 씁니다. 감사합니다.
	이용자 질문(2)	그럼, '이 망치는 더 못 쓴다.'는 '더못 쓴다'로 붙여 쓰나요?
	모범 답변(2)	앞선 질문에 이어 답변드립니다. 한글 맞춤법 제46항에서는 단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다는 규정이 있으나 표현이 의미적으로 자연스럽게 하나로 이어질 때에만 붙여 쓸 수 있습니다. 그러므로 '다 못 쓰다'로 띄어 쓰는 것이 바릅니다. 감사합니다.
	이용자 질문(3)	‘먹어 버릇하면 버릇이 돼서 못 쓴다.’에서는 ‘못 쓴다’를 붙여 써야 하나요?
	모범 답변(3)	안녕하세요? ‘못쓰다’의 띄어쓰기에 대해 답변드립니다. ‘옳지 않다’. 또는 ‘바람직한 상태가 아니다.’를 뜻한다면 ‘못쓴다’로 붙여 씁니다. 그러므로 ‘먹어 버릇하면 버릇이 돼서 못쓴다.’로 붙여 쓰는 것이 바릅니다. 감사합니다.
(…)		

현재 샘플을 기본으로 국어 지식을 사용해 평균 45개의 대화 쌍을 증강한 후 수정을 진행한다. 수정 작업은 7가지 평가 지표를 기준으로 이루어지며, 부적합한 부분에 대하여 수정한 후 검수 과정을 한 번 더 거친다. 평가 지표는 아래 표와 같다.

<표 8> 대화 쌍 증강 생성 후 검수를 위한 평가 지표

번호	평가 지표	정의 및 설명
1	표현 적합성	질문 및 답변의 문장이 자연스럽게 의도가 명확하게 전달되는가? → 어색한 문장 구조, 의미가 불분명한 표현, 화용적 오류 등 여부를 판단함
		질문과 답변 전체가 평가 대상이 되는 지정 주제나 오류 유형 범위 안에 포함되는가? → 주제와 무관한 내용, 다른 오류 유형이나 범위를 벗어난 질문/답변이 포함되어 있는지 여부를 평가함
3	답변의 적합성	질문에 대해 [Knowledge]를 활용하여 정확하게 답했는가? → 설명의 논거가 [Knowledge]에 근거해야 하며, 단정적 오류, 엉뚱한 지식 사용 여부 등을 포함
		답변이 이해하기 쉽고 친절하게 작성되었는가?

	준 적합성	→ 전문 용어 사용 시 설명 포함 여부, 학술적 또는 일반 독자가 이해하기에 적절한 난이도인지 평가함
5	맞춤법 및 표기법 적합성	질문과 답변에 문법, 맞춤법, 띄어쓰기 오류가 없는가? → 질문이나 대답 자체에 문법 오류를 포함하여 혼란을 주지 않는지 확인
6	질문-답변의 상호보완성	질문과 답변이 내용적으로 잘 맞물려 있는가? → 질문의 핵심에 정확히 대응하는 답변이며, 논리적 연결이 부자연스럽지 않은지 판단
7	질문-답변의 중복성	질문과 답변 내에서 같은 내용이 불필요하게 반복되고 있지 않은가? → 답변이 단순히 질문 내용을 반복하거나, 새로운 정보나 설명을 제공하지 않는지 판단

3.2.2. 유형별 모범 사례 작성 방법

대화 자료 구축 예시를 참고하여 대화 자료의 구성 요소를 엑셀 파일에 작성한다. 먼저, 엑셀 파일 ‘이용자 질문 유형.xlsx’의 ‘담당’ 열에서 작성할 질문 유형을 확인한다.

담당	유형 index	대분류	중분류	소분류	질문 유형	국어형 상담 자료 예시	비고
응답	1	띄어쓰기	응답	보조용어	보조 동사 '주다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	2	띄어쓰기	응답	보조용어	보조 동사 '보다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	3	띄어쓰기	응답	보조용어	보조 동사 '버리다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	4	띄어쓰기	응답	보조용어	보조 동사 '있다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	5	띄어쓰기	응답	보조용어	보조 동사 '내다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	6	띄어쓰기	응답	보조용어	보조 동사 '알다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	7	띄어쓰기	응답	보조용어	보조 동사 '마지않다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	8	띄어쓰기	응답	보조용어	보조 동사 '말다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	9	띄어쓰기	응답	보조용어	보조 동사 '못하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	10	띄어쓰기	응답	보조용어	보조 형용사 '실다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	11	띄어쓰기	응답	보조용어	보조 형용사 '안하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	12	띄어쓰기	응답	보조용어	보조 형용사 '직하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	13	띄어쓰기	응답	보조용어	보조 형용사 '듯하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	14	띄어쓰기	응답	보조용어	보조 형용사 '맞다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	15	띄어쓰기	응답	보조용어	보조 형용사 '틀리다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	16	띄어쓰기	응답	보조용어	보조 형용사 '옳다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	17	띄어쓰기	응답	보조용어	보조 형용사 '바르다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	18	띄어쓰기	응답	보조용어	보조 형용사 '잘하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	19	띄어쓰기	응답	보조용어	보조 형용사 '못하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	20	띄어쓰기	응답	보조용어	보조 형용사 '안하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	21	띄어쓰기	응답	보조용어	보조 형용사 '직하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	22	띄어쓰기	응답	보조용어	보조 형용사 '듯하다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	23	띄어쓰기	응답	보조용어	보조 형용사 '맞다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	24	띄어쓰기	응답	보조용어	보조 형용사 '틀리다' 띄어쓰기	이요??? [요???]	이요??? [요???]
응답	25	띄어쓰기	응답	보조용어	보조 형용사 '옳다' 띄어쓰기	이요??? [요???]	이요??? [요???]

[그림 1] 이용자 질문 유형 파일 예시

작성할 질문 유형을 확인한 후 엑셀 파일 ‘지시문 기반 말뭉치_작성 양식.xlsx’의 9개의 열에 각 구성 요소를 아래와 같이 작성한다. 양식은 질문 유형별로 저장하며 작성 양식 파일의 파일명 뒤에 유형 index를 추가하여(지시문 기반 말뭉치_작성 양식_유형 index.xlsx) 저장한다.

출판 index	제시 index	이슈가 질문 유형	출처	국적/직업	이슈가 질문	보통 답변	교과관련 여부	비고
37	2	한번 / 한 번 씌어쓰기 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	○	한글맞춤법 제4장 제2절 제43항
37	3	한번 / 한 번 씌어쓰기 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	○	한글맞춤법 제4장 제2절 제43항
37	4	한번 / 한 번 씌어쓰기 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	○	한글맞춤법 제4장 제2절 제43항
37	5	한번 / 한 번 씌어쓰기 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	2학년 한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	한글맞춤법 제4장 제2절 제43항	○	한글맞춤법 제4장 제2절 제43항

- 1) 유형 index : ‘이용자 질문 유형.xlsx’의 ‘유형 index’ 열에서 확인함.
- 2) 예시 index : 유형별로 5개의 샘플을 작성하여 1에서 5까지 index를 붙임. 대화 쌍이 하나인 샘플을 3개(index 3~4), 대화 쌍이 둘인 샘플을 1개(index 1~2) 작성함. 꼬리 질문은 반드시 원 질문의 바로 아래 행에 작성함. 꼬리 질문은 하나만 작성함.
- 3) 이용자 질문 유형: ‘이용자 질문 유형.xlsx’의 ‘질문 유형’ 열에서 확인함.
- 4) 출처 : 국어 지식 작성을 위한 참고 규정이나 자료를 제시함. 참고 규정을 제시할 때는 규정명과 구체적인 장, 절, 항의 번호를 함께 제시함. 예) ‘한글 맞춤법 제4장 제2절 제44항’. 참고 규정이나 자료를 최대한 제시하되, 참고 규정이나 자료를 찾을 수 없다면 해당 셀에는 ‘x(소문자)’로 제시함.
- 5) 국어 지식: 이용자 질문을 해결하기 위한 국어 지식을 작성함.
- 6) 이용자 질문: 국어생활 상담 자료의 실제 이용자 질문을 참고하여 다양한 유형으로 작성함.
- 7) 모범 답변: 이용자 질문에 대한 답변을 작성함. 질문 유형별로 제시된 작성 형식에 맞춰 일관성 있게 작성함.
- 8) 꼬리 질문 여부: 꼬리 질문이 있는 경우 꼬리 질문 행에만 ‘o(소문자)’로 표시함. 나머지는 모두 ‘x(소문자)’로 표시함.
- 9) 메모: 질문 유형과 지침에 반영이 필요한 내용이 있다면 메모에 기록함.

본 절에서는 말뭉치 구성 요소별로 작성 방법을 설명한다. 말뭉치 구성 요소인 ‘질문 유형’, ‘출처’, ‘국어 지식’, ‘이용자 질문’, ‘모범 답변’ 순으로 자세한 작성 방법을 제시한다.

질문 유형은 국어생활 상담 자료의 범주별로 질문 유형을 선정하였다. 2021년 11월 ~ 2025년 5월까지의 국어생활 상담 자료를 바탕으로 거대언어모델을 활용하여 세부 유형 분류를 시도하였다. 국어생활 상담 자료에서 자주 나타나는 질문 유형을 선별하여 총 872개의 질문 유형을 선별하였다. 질문 유형 범주별 질

문 유형 수와 샘플 수는 아래 <표 9>과 같다. 질문 유형 목록 전체는 [부록 1]에 제시하였다.

<표 9> 질문 유형 범주별 질문 유형 수와 샘플 수

질문 유형 범주	질문 유형 수(개)	샘플 수(개)
의미/화용	120	600
외래어	118	590
표준 발음	103	515
맞춤법	89	445
띄어쓰기	82	410
로마자	73	365
기호	62	310
음성/음운	52	260
사전	13	215
형태/통사	37	185
표준어	35	175
언어예절	33	165
순화어	30	150
건의	13	65
국어사	8	40
문자	5	25
합계	873	4,515

3.3.2. 이용자 질문 작성

이용자 질문은 국어생활 상담 자료를 참고하여 실제 이용자가 챗봇을 이용하는 상황을 가정하여 다양하게 작성한다. 이용자의 질문은 한 질문 유형에 대해서 5가지 형태를 모두 작성한다. 기본적으로 아래와 같이 5개의 형식을 하나씩 작성하며, 국어생활 상담 자료에서 5가지 형식 중 나타나지 않는 형식이 있을 경우 제외하고 작성하는 것이 가능하며, 5가지 형식 외에 새로운 질문 형식이 있을 경우 추가도 가능하다. 실제 이용자의 질문은 어문 규정에 맞지 않을 수 있으나 작성하는 이용자 질문은 어문 규정에 맞게 쓰며 질문 그 자체로 질문의 요지를 이해할 수 있어야 한다.

<표 10> 이용자 질문 유형

	이용자 질문 유형	이용자 질문 유형 예시
①	어문 규범 설명형	<ul style="list-style-type: none"> 어문 규범에 대한 설명을 요청하는 질문 예) ‘협조해 주다.’는 왜 띄어 쓰는 건가요???
②	선택형	<ul style="list-style-type: none"> 두 개 이상의 예시 중 바른 표현을 묻는 질문 예) 안녕하세요. 나무 심어봐요./나무 심어 봐요. 띄어쓰기 어떻게 맞는 표현인가요?
③	교정형	<ul style="list-style-type: none"> 제시한 표현 교정을 요청하는 질문 예) ‘믿어의심치않다’의 띄어쓰기는 어떻게 하나요?

④	확인형	<ul style="list-style-type: none"> 제시한 표현이 맞는지 확인을 요청하는 질문 예) 보조 용언 ‘있다’의 띄어쓰기가 궁금해요. ‘열려 있다’는 붙여 쓰나요?
⑤	꼬리 질문형	<ul style="list-style-type: none"> 앞선 답변에 대한 추가 설명을 요구하는 질문 예) 근데, ‘넓어졌어요’는 왜 붙여 써요?

3.3.3. 출처 작성

국어 지식의 출처를 작성하기 위해 사용할 수 있는 규정과 사전, 자료는 아래와 같다. 국어 지식의 출처를 작성할 때는 아래의 명칭을 그대로(띄어쓰기까지) 쓴다.

- 1) 한글 맞춤법
- 2) 표준어 규정 제1부 표준어 사정
- 3) 표준어 규정 제2부 표준 발음법
- 4) 문장 부호 해설
- 5) 국어의 로마자 표기법
- 6) 외래어 표기법
- 7) 국립국어원 다듬은 말: 순화어는 ‘국립국어원 다듬은 말’만 활용함.
- 8) 표준국어대사전
- 9) 우리말샘
- 10) 표준 국어 문법, 한국어 표준 문법

- 규정은 장/절/항까지 씀. 예) 제5장 제1절 제41항
- 여러 규정이 쓰인 경우 쉽표로 구분함. 예) 한글 맞춤법 제1장 제2항, 제5장 제1절 제41항
- 규정과 사전을 모두 제시하는 경우 규정, 사전 순으로 제시함. 예) 한글 맞춤법 제1장 제2항, 표준국어대사전
- 한 규정에서 두 개 이상의 조항을 참고한 경우 규정명은 한 번만 쓰고 장, 절, 항을 쉽표로 구분하여 제시 예) 한글 맞춤법 제1장 제2항, 제5장 제1절 제41항
- 국어 문법과 관련된 질문의 경우에는 <표준 국어 문법>과 <한국어 표준 문법>을 참고
- 출처가 없는 경우에는 ‘x(소문자)’로 표시

3.3.4. 국어 지식 작성

이용자 질문에 답변을 작성하기 위해 사용되는 모든 국어 지식을 작성한다. 국어 지식은 아래와 같은 형식으로 작성한다. 한 질문에 여러 개의 국어 지식이 필요한 경우, 여러 개를 작성할 수 있다(아래 작성 예시 참고).

<표 11> 국어 지식 작성 예시

구분	내용	설명
이용자 질문 유형	동사 '못쓰다' 띄어쓰기	
출처	한글 맞춤법 제5장 제2절 제46항, 표준국어대사전	두 개 이상의 국어 지식을 사용할 경우 출처는 쉽표로 구분 하며 규정을 앞에, 사전을 뒤에 씀.
국어 지식	단음절로 된 단어가 연이어 나타날 때는 붙여 쓸 수 있다. '못쓰다'는 '((-못쓰게' 꼴로 쓰여)) 얼굴이나 몸이 축나다', '((주로 '-으면', '-어서'와 함께 쓰여)) 옳지 않다. 또는 바 람직한 상태가 아니다.'의 의미로 쓰인다.	두 개 이상의 국어 지식을 활용할 경우 앞선 국어 지식 뒤에 공백을 하나 넣고 붙 여서 씀.
이용자 질문	'이 망치는 더 못 쓴다.'는 '더못 쓴다'로 붙여 쓰나요?	
모범 답변	안녕하세요. '더 못 쓴다'의 띄어쓰기에 대해 답변드립니다. 다. '못쓴다'는 '얼굴이 상하다', '행동이 바람직하지 않 다', '옳지 않다'와 같은 의미로 쓰일 때만 붙여 씁니다. 또한, 한글 맞춤법 제46항에서는 단음절로 된 단어가 연 이어 나타날 때는 붙여 쓸 수 있다는 규정이 있으나 표현 이 의미적으로 자연스럽게 하나로 이어질 때에만 붙여 쓸 수 있습니다. 그러므로 '다 못 쓰다'로 띄어 쓰는 것이 바 릅니다. 고맙습니다.	

국어 지식을 제공할 수 없는 경우는 크게 두 가지로 나눌 수 있는데 어원이나 의미 차이, 순화어가 없
는 단어 등과 같이 국어 지식으로 삼고 있는 자료에 정확히 규정되어 있지 않은 경우와 건의 사항과 같은
이용자의 의견과 관련된 경우이다. 이러한 경우에는 아래 <표 12>의 마지막 열을 참고하여 작성한다.

<표 12> 국어 지식 작성 형식

국어 지식 종류	작성 형식 예시
사전	<ul style="list-style-type: none"> • 여러 의미가 있는 경우 쉽표로 구분하여 제시한다. • 다 제시하기에 의미가 많은 경우, 이용자 질문에 해당하는 의 미만 제시한다. • 뜻풀이 외에 이용자 질문에 필요한 정보가 있다면 더 제시할 수 있다. <p><형식 예시></p> <ul style="list-style-type: none"> • '[단어]'은/는 '[뜻풀이1]', '[뜻풀이2]'(이)라는 의미로 쓰인다. • '[단어]'은/는 '[뜻풀이]'(이)라는 뜻이다.

	<ul style="list-style-type: none"> • ‘[단어]’은/는 ‘[뜻풀이]’(이)라는 의미이다. <p>예) ‘메우다’는 ‘뚫려 있거나 비어 있는 곳을 막거나 채우다.’, ‘어떤 장소를 가득 채우다.’, ‘부족하거나 모자라는 것을 채우다.’, ‘시간을 적당히 또는 그럭저럭 보내다.’라는 의미로 쓰인다.</p> <p>예) ‘드러나다’는 ‘알려지지 않은 사실이 널리 밝혀지다.’라는 의미이다. ‘드러나다’는 ‘들다’와 ‘나다’가 결합한 말이지만 ‘들어나다’로 적지 않고 ‘드러나다’로 적는다. 이는 두 개의 용언이 어울려 한 개의 용언이 될 적에, 앞말의 본뜻이 유지되고 있는 것은 그 원형을 밝혀 적고, 그 본뜻에서 멀어진 것은 밝혀 적지 않는다는 규정에 따른 것이다.</p>
<p>한글 맞춤법 외래어 표기법 국어의 로마자 표기법 문장 부호 해설</p>	<ul style="list-style-type: none"> • 이용자 질문에 답변을 작성하기 위한 모든 규정을 그대로 적는다. <p>예) 단위를 나타내는 의존 명사는 띄어 쓴다. 다만, 수 관형사 뒤에 단위 명사가 붙어서 차례를 나타내는 경우에는 앞말과 붙여 쓸 수 있다.</p> <p>예) 받침에는 ‘ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ’만을 쓰는 것이 원칙이다. 짧은 모음 다음의 어말 무성 파열음 [p]은 받침 ‘ㅂ’으로 적는다.</p> <p>예) ‘ㄱ, ㄷ, ㅂ’은 모음 앞에서는 ‘g, d, b’로, 자음 앞이나 어말에서는 ‘k, t, p’로 적는다.</p> <p>예) 줄임표는 앞말에 붙여 쓰는 것이 원칙이며 문장의 앞에 붙여 쓰지 않는다. 문장이나 글의 일부를 생략할 때는 앞뒤 쉼표나 마침표 없이 띄어 쓰며 문장이 끝나면 마침표나 느낌표를 쓰는 것이 원칙이다. 가운데에 여섯 점을 찍는 것이 원칙이나 아래에 여섯 점을 찍거나 세 점으로 찍을 수 있다.</p>
국립국어원 다듬은 말	<ul style="list-style-type: none"> • 순화어 정보는 ‘국립국어원 다듬은 말’의 내용만 활용한다. <p><형식 예시></p> <ul style="list-style-type: none"> • ‘[단어]’의 원어를 ‘[원어]’(으)로, 다듬은 말을 ‘[다듬은 말]’(으)로 제시하고 있다. <p>예) ‘SNS’의 원어를 ‘social network service(영)’로, 다듬은 말을 ‘누리 소통망, 누리 소통망 서비스, 사회 관계망, 사회 관계망 서비스’로 제시하고 있다.</p>
국어 지식을 제공할 수 없는 경우	<p>1) 국어 지식 자료에 정확히 규정되어 있지 않은 경우</p> <p>- 정확히 규정되어 있지 않음을 언급하고 참고 정보가 있다면</p>

	<p>‘다만’을 쓴 후에 제시함. 참고할 정보가 없는 경우에는 끝인사로 끝냄.</p> <p>예) 안녕하세요. 특정 어휘의 어원에 관한 질문에 답변하겠습니다. 우리말샘의 ‘역사 정보’란에 ‘생때같다’의 역사 및 어원 정보가 실려 있지 않습니다. 다만, 사전의 형태 분석을 통해 ‘생때’와 ‘같다’가 결합된 말임을 짐작할 수 있습니다. 감사합니다.</p> <p>예) 안녕하세요? 알고리즘 순화어에 관해 말씀드리겠습니다. 현재 알고리즘의 순화어는 따로 정해져 있지 않습니다. 다만, 생명과학 분야에서 사용되는 ‘지네틱알고리즘’을 ‘유전연산방식’으로 다듬은 바가 있어 이를 참고하실 수 있습니다. 국립국어원의 용어, 우리말샘을 참고하시길 바랍니다. 감사합니다."</p> <p>2) 이용자의 의견과 관련된 경우</p> <p>- 이용자의 의견을 언급하고 관련 부서에 전달하겠다고 답변함.</p> <p>예) 안녕하세요. 사전의 뜻풀이에 관한 건의 사항에 답변하겠습니다. [이용자 의견 언급] 제안해 주신 사항을 관련 부서에 전달해 두겠습니다. 사전 뜻풀이에 깊은 관심을 가지고 건의해 주셔서 고맙습니다.</p>
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

3.3.5. 모범 답변 작성

모범 답변은 아래 제시된 형식을 참고하여 일관된 형식으로 작성한다. 이용자의 질문은 5가지 형식으로 작성한다. 모범 답변은 어문 규범 전문가가 아닌 일반 이용자를 고려하여 한 단락 길이에서 완결된 의미를 전달하도록 하며, 아래 7개의 내용이 포함되도록 작성한다. 모범 답변에 포함되어야 하는 요소는 아래와 같다.

<표 13> 모범 답변의 형식과 구성 요소

구성 요소	역할	예시
인사	대화의 시작	안녕하세요? 안녕하십니까?
질문 요약	어떤 주제의 질문인지 요약하여 질문의 주제를 명확히 함.	[질문 내용] 질문에 답변하겠습니다. [질문 내용]에 대해서 답변드리겠습니다. [질문 내용]에 대한 질문에 답변하겠습니다.
규범 설명	해당 질문의 핵심 규범이나 원칙을 설명함.	~은/는 ~하는 것이 바릅니다./적절합니다. ~(으)로 쓰는 것을 권장합니다. 예) '1만 원대'라고 띄어 쓰는 것이 바릅니다.
문법 설명	판단의 근거를 '국어 지식'(문법, 맞춤법, 어휘 등)에 근거하여 설명함.	~는 ~의미를 가진 ~입니다. 예) '원'은 화폐 단위를 뜻하는 의존 명사입니다.
예시 제시	실제 예를 통해 일반화된 규범을 구체적으로 제시하여 독자의 이해 보완	예를 들어 ~와 같이 씁니다. 아래와 같이 씁니다. 예) [예시]
예외 안내	특수한 상황이나 예외적 사용을 안내하여 규범의 유연성 제시	다만, ~의 경우에는 ~할 수 있습니다. 예) 계약서 등 특수 상황에서는 붙여 쓰기도 합니다.
맺음말	답변의 마무리	참고하시기 바랍니다. 고맙습니다.

꼬리 질문을 작성할 때는 이어지는 질문이므로 '안녕하세요.'와 같은 인사말과 질문 요약을 생략하고 '앞선 질문에서 말씀드린 바와 같이', '앞선 질문에 이어서 답변드립니다.' 등과 같이 앞선 질문과 이어진다는 것을 나타내는 표현을 사용한다.

<표 14> 꼬리 질문 예시

질문 유형	질문	대답
질문	'한번'은 언제 붙여 쓰고 언제 띄어 써야 하나요?	<ul style="list-style-type: none"> • [인사]. [질문 내용]에 대한 질문에 답변하겠습니다. [질문에 대한 대답]. 감사합니다. <p>예) 안녕하세요, '한번'의 띄어쓰기에 관한 질문에 답변하겠습니다. '한 번', '한번'은 맥락에 따라 띄어쓰기가 달라 집니다. 1회를 말하는 것이라면 '한 번'으로 띄어 쓰고, '지난 어느 때나 기회', '어떤 일을 시험 삼아 시도함을 나타내는 말', '기회 있는 어떤 때에', '어떤 행동이나 상태를 강조하는 뜻을 나타내는 말', '일단 한 차례'라는 뜻으로 쓸 때는 '한번'으로 붙여 씁니다. 이에 따라, '한 번 해 보다'의 경우 붙여 쓰고 '떡볶이를 딱 한 번 먹어 보았다'의 경우 띄어 씁니다. 감사합니다.</p>
꼬리 질문	그럼, "한번만 용서해 주세요"는 붙여 써야 하나요?	<ul style="list-style-type: none"> • 앞선 질문에서 말씀드린 바와 같이 • 앞선 질문에 이어 답변드립니다. • 이어서 답변드립니다. • 추가 질문에 이어서 답변드립니다. • 앞서 말씀드린 것처럼 <p>예) 앞선 질문에서 말씀드린 바와 같이 '한번만 용서해 주세요'의 경우 횟수의 의미가 드러나므로, '한 번만 용서해 주세요'로 띄어 쓰는 것이 적절합니다. 감사합니다.</p>

꼬리 질문은 아래와 같이 4개의 유형으로 분류할 수 있다. 그 외에 질문 유형에 따라 다른 유형이 있을 수 있으므로 국어생활 상담 자료를 확인하여 새로운 유형을 추가하여 작성할 수 있다.

<표 15> 꼬리 질문 유형

꼬리 질문 유형	설명
분할질문형	본래 하나의 답변으로 설명되는 있는데, 답변에 따라 추가 질문이 가능한 경우
추가질문형(예시 변경)	동일한 질문 형태이나 다른 예시를 이용해 묻는 경우
질문보충형	"고유명사는 붙여쓰나요? 띄어쓰나요?"와 같이 막연한 질문인 경우, 구체적인 예시를 요청하는 추가 질문
답변보충형	답변에 부족한 부분이 있다고 판단되거나, 추가 답변을 유도할 수 있는 질문

앞선 구축 방법을 요약 정리하면 아래 표와 같다.

항목	작성 방식
문장 구성 구조	• 인사 → 질문 요약 → 규범 제시 → 설명 → 예시 → 예외 및 보충 → 맺음말
답변 길이	• 한 단락 수준
문체	• 존댓말, 평이한 문장 사용 (지나친 전문용어 지양, 예시로 보완) • 용어나 설명의 수준은 학교 문법 수준으로 작성 • 인용할 때는 작은따옴표를 씀.
내용의 근거	• 표준국어대사전, 한글 맞춤법, 표준 발음법 등의 ‘국어 지식’ 명시
예시 활용	• 원칙 설명 후 실제 사용 예시 포함(잘못된 예보다는 바른 예 위주로)
자주 쓰는 표현 유형화	• ‘~는 ~입니다.’ • ‘~는 ~의 뜻을 더하는 접미사입니다.’ • ‘따라서 ~라고 써야 바랍니다.’
답변이 어려운 경우	• 이용자의 의견이나 감정적 질문에는 “답변드리기 어렵습니다.”로 응대

이용자 질문과 모범 답변의 예시는 다음과 같다.

예시 1	높임법의 '-시-' 사용 여부
이용자 질문	선생님한테 ‘말씀하세요’라고 해야 하나요, ‘말해요’라고 해도 되나요?
모범 답변	안녕하세요? 높임 표현 사용에 관한 질문에 답변하겠습니다. 상대방이 높여야 할 대상일 경우 ‘말씀하세요’와 같이 높임 표현을 사용하는 것이 적절합니다. ‘말씀하다’는 ‘말하다’의 높임말이며, ‘-세요’는 종결어미 ‘-시-’의 형태입니다. 반면, ‘말해요’는 일반적 평어체 표현으로 높임의 의미는 없습니다. 따라서 ‘선생님’처럼 존대해야 할 대상에게는 ‘말씀하세요’가 바른 표현입니다. 감사합니다

예시 2	‘에서’와 ‘부터’의 쓰임
이용자 질문	서울에서 출발하다와 ‘서울부터 출발하다’는 어떤 차이가 있나요?
모범 답변	안녕하세요? 조사의 쓰임에 관한 질문에 답변하겠습니다. ‘서울에서 출발하다’는 출발 장소를 나타내는 표현이며, ‘에서’는 동작이 이루어지는 장소를 나타내는 부사격 조사입니다. 반면 ‘서울부터 출발하다’는 순서상의 시작 지점을 강조하는 표현으로 ‘부터’는 범위의 시작을 나타내는 부사격 조사입니다. 따라서 문맥에 따라 두 표현은 의미상의 차이가 있습니다. 감사합니다.

자주 사용되는 문장 유형 정리하면 아래와 같다.

상황	문장 예시
인사	<ul style="list-style-type: none"> 안녕하세요? 안녕하세요? [질문 주제]에 관한 질문에 답변드리겠습니다.
규범 제시	<ul style="list-style-type: none"> ~는 ~하는 것이 바릅니다. ~로 표기하는 것이 적절합니다. ~는 [이유]로 인해 ~하는 것이 바릅니다.
근거 설명	<ul style="list-style-type: none"> ~는 [의존 명사/접미사/조사]이므로 ~하게 처리합니다. '~'는 [문법적 기능/역할]을 가지므로 ~하게 표기합니다.
예시 제시	<ul style="list-style-type: none"> 예) ~ 예를 들어 ~처럼 쓸 수 있습니다. 예를 들어, '~'과 같이 씁니다.
예외 언급	<ul style="list-style-type: none"> 다만, ~하는 경우에는 ~할 수도 있습니다. 다만, ~한 특수 상황에서는 ~하게 쓰는 것도 허용됩니다.
맺음말	<ul style="list-style-type: none"> 고맙습니다. 참고하시기 바랍니다.

■ 질문 유형 목록

유형 index	대분류	중분류	소분류	질문 유형
1	띄어쓰기	용언	보조용언	보조 동사 '주다' 띄어쓰기
2	띄어쓰기	용언	보조용언	보조 동사 '보다' 띄어쓰기
3	띄어쓰기	용언	보조용언	보조 동사 '버리다' 띄어쓰기
4	띄어쓰기	용언	보조용언	보조 동사 '있다' 띄어쓰기
5	띄어쓰기	용언	보조용언	보조 동사 '내다' 띄어쓰기
6	띄어쓰기	용언	보조용언	보조 동사 '없다' 띄어쓰기
7	띄어쓰기	용언	보조용언	보조 동사 '마지않다' 띄어쓰기
8	띄어쓰기	용언	보조용언	보조 동사 '말다' 띄어쓰기
9	띄어쓰기	용언	보조용언	보조 동사 '못하다' 띄어쓰기
10	띄어쓰기	용언	보조용언	보조 형용사 '싫다' 띄어쓰기
11	띄어쓰기	용언	보조용언	보조 형용사 '만하다' 띄어쓰기
12	띄어쓰기	용언	보조용언	보조 형용사 '직하다' 띄어쓰기
13	띄어쓰기	용언	보조용언	보조 형용사 '듯하다' 띄어쓰기
14	띄어쓰기	용언	보조용언	본용언이 합성어나 파생어이고 그 활용형이 3음절 이상인 경우에 보조용언 띄어쓰기
15	띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어오다', '-아/어가다' 띄어쓰기
16	띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어지다' 띄어쓰기
17	띄어쓰기	용언	보조용언	예외적 보조 용언 구성 '-아/어하다' 띄어쓰기
18	띄어쓰기	용언	보조용언	보조 용언 띄어쓰기 관련 어문 규정 설명 요청
19	띄어쓰기	용언	보조용언, 동사	사전에 등재된 한 단어 합성 동사의 띄어쓰기
20	띄어쓰기	용언	보조용언, 동사	사전에 등재된 한 단어이지만, 문맥의 의미를 고려하여 '본용언+보조 동사'로 볼 수 있는 합성 동사의 띄어쓰기
21	띄어쓰기	용언	보조용언, 형용사	사전에 등재된 한 단어 합성 형용사의 띄어쓰기
22	띄어쓰기	용언	보조용언, 형용사	사전에 등재된 한 단어이지만, 문맥의 의미를 고려하여 '본용언+보조 형용사'로 볼 수 있는 합성 형용사의 띄어쓰기
23	띄어쓰기	의존형태	동사파생접미사	동사파생접미사 '-하다' 띄어쓰기
24	띄어쓰기	의존형태	동사파생접미사	동사파생접미사 '-당하다' 띄어쓰기
25	띄어쓰기	의존형태	동사파생접미사	동사파생접미사 '-시키다' 띄어쓰기
26	띄어쓰기	의존형태	동사파생접미사	동사파생접미사 '-되다' 띄어쓰기
27	띄어쓰기	의존형태	동사파생접미사	동사파생접미사 '-받다' 띄어쓰기
28	띄어쓰기	의존형태	형용사파생접미사	형용사파생접미사 '-하다' 띄어쓰기
29	띄어쓰기	의존형태	형용사파생접미사	형용사파생접미사 '-스럽다' 띄어쓰기
30	띄어쓰기	의존형태	형용사파생접미사	형용사파생접미사 '-답다' 띄어쓰기
31	띄어쓰기	수식언	부사	'못+하다' 띄어쓰기
32	띄어쓰기	용언	동사, 형용사	못하다
33	띄어쓰기	수식언	부사	'못+동사/형용사'의 띄어쓰기
34	띄어쓰기	수식언	부사	부사 '못다' 띄어쓰기
35	띄어쓰기	수식언	부사	'잘못' 띄어쓰기
36	띄어쓰기	수식언	부사	'잘' 띄어쓰기
37	띄어쓰기	수식언	부사, 명사	한번/한 번 띄어쓰기
38	띄어쓰기	용언	동사	'잘하다' 띄어쓰기
39	띄어쓰기	용언	동사	동사 '못쓰다' 띄어쓰기
40	띄어쓰기	용언	동사	동사 '못살다' 띄어쓰기
41	띄어쓰기	용언	형용사	형용사 '못마땅하다' 띄어쓰기
42	띄어쓰기	용언	형용사	동사 '못생기다' 띄어쓰기

43	띄어쓰기	체언	명사	전문용어 띄어쓰기
44	띄어쓰기	체언	명사	지명 띄어쓰기
45	띄어쓰기	체언	명사	주소 띄어쓰기
46	띄어쓰기	체언	명사	한국인 성과 이름 띄어쓰기
47	띄어쓰기	체언	명사	외국 인명 띄어쓰기
48	띄어쓰기	체언	명사	성+씨 띄어쓰기
49	띄어쓰기	체언	명사	이름+씨 띄어쓰기
50	띄어쓰기	체언	명사	성+이름+직위/신분 띄어쓰기
51	띄어쓰기	체언	명사	직위/신분 + 님 띄어쓰기
52	띄어쓰기	체언	명사	숫자+의존명사(단위)
53	띄어쓰기	체언	명사	수사+의존명사(단위)
54	띄어쓰기	체언	명사	숫자+시/분 띄어쓰기
55	띄어쓰기	체언	명사	수사+시/분 띄어쓰기
56	띄어쓰기	체언	명사	금액 띄어쓰기
57	띄어쓰기	체언	명사	금액 붙여쓰기
58	띄어쓰기	관계언	조사	조사 '-같이' 띄어쓰기
59	띄어쓰기	수식언	부사	부사 '-같이' 띄어쓰기
60	띄어쓰기	관계언	조사	조사 '-만큼' 띄어쓰기
61	띄어쓰기	체언	명사	의존명사 '만큼' 띄어쓰기
62	띄어쓰기	관계언	조사	조사 '-뿐' 띄어쓰기
63	띄어쓰기	의존형태	의존명사	의존명사 '뿐' 띄어쓰기
64	띄어쓰기	관계언	조사	조사 '-처럼' 띄어쓰기
65	띄어쓰기	관계언	조사	조사 '-하고' 띄어쓰기
66	띄어쓰기	관계언	조사	조사 '-대로' 띄어쓰기
67	띄어쓰기	체언	명사	의존명사 '대로' 띄어쓰기
68	띄어쓰기	관계언	조사	조사 '-하고는' 띄어쓰기
69	띄어쓰기	관계언	조사	조사 '-은카녕' 띄어쓰기
70	띄어쓰기	용언	지정사	-이다 띄어쓰기
71	띄어쓰기	체언	명사	의존명사 '데' 띄어쓰기
72	띄어쓰기	의존형태	어미	어미 'ㄴ/은/는데' 띄어쓰기
73	띄어쓰기	의존형태	어미	어미 '(으)르수록' 띄어쓰기
74	띄어쓰기	기타		첨어의 띄어쓰기
75	띄어쓰기	기타		단음절어 연속되는 경우의 띄어쓰기
76	띄어쓰기	기호		말줄임표…… 앞뒤 띄어쓰기
77	띄어쓰기	기호		쌍반점(:) 앞뒤 띄어쓰기
78	띄어쓰기	기호		가운뎃점(·) 앞뒤 띄어쓰기
79	띄어쓰기	기호		쌍점(:) 앞뒤 띄어쓰기
80	띄어쓰기	기호		앰프 기호(&) 앞뒤 띄어쓰기
81	띄어쓰기	기호		물결표(~) 앞뒤 띄어쓰기
82	띄어쓰기	기호		퍼센트(%) + 조사 띄어쓰기
83	외래어			'capture'의 외래어 표기
84	외래어			'culture'의 외래어 표기
85	외래어			'ketchup'의 외래어 표기
86	외래어			'Avengers'의 외래어 표기
87	외래어			'signature'의 외래어 표기
88	외래어			'flash'의 외래어 표기
89	외래어			'brush'의 외래어 표기

90	외래어			'fish'의 외래어 표기
91	외래어			'setting'의 외래어 표기
92	외래어			'bridge'의 외래어 표기
93	외래어			'marketing'의 외래어 표기
94	외래어			'cutting'의 외래어 표기
95	외래어			'targeting'의 외래어 표기
96	외래어			'message'의 외래어 표기
97	외래어			'sausage'의 외래어 표기
98	외래어			'application'의 외래어 표기
99	외래어			'app'의 외래어 표기
100	외래어			'ambassador'의 외래어 표기
101	외래어			'license'의 외래어 표기
102	외래어			'allergy'의 외래어 표기
103	외래어			'ceremony'의 외래어 표기
104	외래어			'recipe'의 외래어 표기
105	외래어			'running'의 외래어 표기
106	외래어			'lemonade'의 외래어 표기
107	외래어			'sharp'의 외래어 표기
108	외래어			'shop'의 외래어 표기
109	외래어			'locker'의 외래어 표기
110	외래어			'rocker'의 외래어 표기
111	외래어			'stop'의 외래어 표기
112	외래어			'launching'의 외래어 표기
113	외래어			'solution'의 외래어 표기
114	외래어			'cheers'의 외래어 표기
115	외래어			'melon'의 외래어 표기
116	외래어			'placard'의 외래어 표기
117	외래어			'palette'의 외래어 표기
118	외래어			'pallet'의 외래어 표기
119	외래어			'chassis'의 외래어 표기
120	외래어			'The Beatles'의 외래어 표기
121	외래어			'sofa'의 외래어 표기
122	외래어			'pâtissier'의 외래어 표기
123	외래어			'déjà vu'의 외래어 표기
124	외래어			'chef'의 외래어 표기
125	외래어			'towel'의 외래어 표기
126	외래어			'kickboard'의 외래어 표기
127	외래어			'cake'의 외래어 표기
128	외래어			'squat'의 외래어 표기
129	외래어			'lobster'의 외래어 표기
130	외래어			'outlet'의 외래어 표기
131	외래어			'jamón'의 외래어 표기
132	외래어			'latte'의 외래어 표기
133	외래어			'gelato'의 외래어 표기
134	외래어			'tonkasu'의 외래어 표기
135	외래어			'café'의 외래어 표기
136	외래어			'carbonara'의 외래어 표기

137	외래어			'cardigan'의 외래어 표기
138	외래어			'massage'의 외래어 표기
139	외래어			'thumbnail'의 외래어 표기
140	외래어			'site'의 외래어 표기
141	외래어			'sign'의 외래어 표기
142	외래어			'bag '의 외래어 표기
143	외래어			'badge'의 외래어 표기
144	외래어			'workshop'의 외래어 표기
145	외래어			'leadership'의 외래어 표기
146	외래어			'manual'의 외래어 표기
147	외래어			'damage'의 외래어 표기
148	외래어			'trend'의 외래어 표기
149	외래어			'tracking '의 외래어 표기
150	외래어			'platform'의 외래어 표기
151	외래어			'navigation'의 외래어 표기
152	외래어			'penalty'의 외래어 표기
153	외래어			'pandemic'의 외래어 표기
154	외래어			'mental'의 외래어 표기
155	외래어			'juice'의 외래어 표기
156	외래어			'super '의 외래어 표기
157	외래어			'deck'의 외래어 표기
158	외래어			'catalog '의 외래어 표기
159	외래어			'good'의 외래어 표기
160	외래어			'respect'의 외래어 표기
161	외래어			'summer'의 외래어 표기
162	외래어			'catholic'의 외래어 표기
163	외래어			'chocolate'의 외래어 표기
164	외래어			'benefit'의 외래어 표기
165	외래어			'leaflet'의 외래어 표기
166	외래어			'jacket'의 외래어 표기
167	외래어			'pamphlet'의 외래어 표기
168	외래어			'shorts'의 외래어 표기
169	외래어			'remocon(remote control)'의 외래어 표기
170	외래어			'aircon(air conditioner)'의 외래어 표기
171	외래어			'mania '의 외래어 표기
172	외래어			'valentine'의 외래어 표기
173	외래어			'narration'의 외래어 표기
174	외래어			'takoyaki'의 외래어 표기
175	외래어			'barbecue'의 외래어 표기
176	외래어			'accessory'의 외래어 표기
177	외래어			'calligraphy'의 외래어 표기
178	외래어			'halloween'의 외래어 표기
179	외래어			'shrimp'의 외래어 표기
180	외래어			'Türkiye '의 외래어 표기
181	외래어			'Kyiv'의 외래어 표기
182	외래어			'propose '의 외래어 표기
183	외래어			'agenda'의 외래어 표기

184	외래어			'concept'의 외래어 표기
185	외래어			'contact'의 외래어 표기
186	외래어			'conference'의 외래어 표기
187	외래어			'collaboration'의 외래어 표기
188	외래어			'fighting'의 외래어 표기
189	외래어			'fence'의 외래어 표기
190	외래어			'business'의 외래어 표기
191	외래어			'thank you'의 외래어 표기
192	외래어			'soup'의 외래어 표기
193	외래어			enquête'의 외래어 표기
194	외래어			'and'의 외래어 표기
195	외래어			'YouTube'의 외래어 표기
196	외래어			'jet'의 외래어 표기
197	외래어			'zip line'의 외래어 표기
198	외래어			'class'의 외래어 표기
199	외래어			'portal'의 외래어 표기
200	외래어			'mug shot'의 외래어 표기
201	로마자			단모음 'ㅏ(a)'의 로마자 표기
202	로마자			단모음 'ㅓ(eo)'의 로마자 표기
203	로마자			단모음 'ㅗ(o)'의 로마자 표기
204	로마자			단모음 'ㅜ(u)'의 로마자 표기
205	로마자			단모음 'ㅡ(eu)'의 로마자 표기
206	로마자			단모음 'ㅣ(i)'의 로마자 표기
207	로마자			단모음 'ㅐ(ae)'의 로마자 표기
208	로마자			단모음 'ㅒ(e)'의 로마자 표기
209	로마자			단모음 'ㅓ(oe)'의 로마자 표기
210	로마자			단모음 'ㅗ(wi)'의 로마자 표기
211	로마자			이중모음 'ㅑ(ya)'의 로마자 표기
212	로마자			이중모음 'ㅓ(yeo)'의 로마자 표기
213	로마자			이중모음 'ㅖ(yo)'의 로마자 표기
214	로마자			이중모음 'ㅠ(yu)'의 로마자 표기
215	로마자			이중모음 'ㅒ(yae)'의 로마자 표기
216	로마자			이중모음 'ㅕ(ye)'의 로마자 표기
217	로마자			이중모음 'ㅗ(wa)'의 로마자 표기
218	로마자			이중모음 'ㅓ(wae)'의 로마자 표기
219	로마자			이중모음 'ㅗ(wo)'의 로마자 표기
220	로마자			이중모음 'ㅕ(we)'의 로마자 표기
221	로마자			이중모음 'ㅓ(ui)'의 로마자 표기
222	로마자			모음 앞에서 파열음 'ㄱ(g)'의 로마자 표기
223	로마자			자음 앞에서 파열음 'ㄱ(k)'의 로마자 표기
224	로마자			어말에서 파열음 'ㄱ(k)'의 로마자 표기
225	로마자			파열음 'ㄱ(kk)'의 로마자 표기
226	로마자			파열음 'ㅋ(k)'의 로마자 표기
227	로마자			모음 앞에서 파열음 'ㄷ(d)'의 로마자 표기
228	로마자			자음 앞에서 파열음 'ㄷ(t)'의 로마자 표기
229	로마자			어말에서 파열음 'ㄷ(t)'의 로마자 표기
230	로마자			파열음 'ㅌ(tt)'의 로마자 표기

231	로마자		파열음 'ㅌ(t)'의 로마자 표기
232	로마자		모음 앞에서 파열음 'ㅂ(b)'의 로마자 표기
233	로마자		자음 앞에서 파열음 'ㅃ(p)'의 로마자 표기
234	로마자		어말에서 파열음 'ㅃ(p)'의 로마자 표기
235	로마자		파열음 'ㅍ(pp)'의 로마자 표기
236	로마자		파열음 'ㅍ(p)'의 로마자 표기
237	로마자		파찰음 'ㅈ(j)'의 로마자 표기
238	로마자		파찰음 'ㅉ(jj)'의 로마자 표기
239	로마자		파찰음 'ㅊ(ch)'의 로마자 표기
240	로마자		마찰음 'ㅅ(s)'의 로마자 표기
241	로마자		마찰음 'ㅆ(ss)'의 로마자 표기
242	로마자		마찰음 'ㅎ(h)'의 로마자 표기
243	로마자		비음 'ㄴ(n)'의 로마자 표기
244	로마자		비음 'ㄹ(m)'의 로마자 표기
245	로마자		비음 'ㅇ(ng)'의 로마자 표기
246	로마자		모음 앞에서 유음 'ㄹ(r)'의 로마자 표기
247	로마자		어말에서 유음 'ㄹ()'의 로마자 표기
248	로마자		어말에서 유음 'ㄹㄹ()'의 로마자 표기
249	로마자		자음 사이에서 동화 작용이 일어나는 경우의 로마자 표기
250	로마자		'ㄴ, ㄹ'이 닳는 경우의 로마자 표기
251	로마자		구개음화가 되는 경우의 로마자 표기
252	로마자		끝소리 규칙이 되는 경우의 로마자 표기
253	로마자		연음이 되는 경우의 로마자 표기
254	로마자		표준 발음 외에 허용 발음의 로마자 표기 허용 여부
255	로마자		체언 외에 'ㄱ, ㄷ, ㅂ, ㅈ'이 'ㅎ'과 합하여 거센소리로 소리나는 경우의 로마자 표기
256	로마자		체언에서 'ㄱ, ㄷ, ㅂ' 뒤에 'ㅎ'이 따를 경우의 로마자 표기
257	로마자		된소리되기의 경우 로마자 표기
258	로마자		로마자 표기에서 붙임표(-) 사용
259	로마자		고유 명사의 첫 글자 대문자 표기
260	로마자		인명의 성과 이름의 로마자 표기 띄어쓰기 음절 사이에 붙임표(-) 사용
261	로마자		인명의 성과 이름의 음절 사이 붙임표(-) 사용
262	로마자		인명의 이름에서 일어나는 음운 변화의 로마자 표기
263	로마자		인명의 성과 이름 순서
264	로마자		인명의 성과 이름 띄어쓰기
265	로마자		인명의 성의 표기
266	로마자		'도, 시, 군, 구, 읍, 면, 리, 동'의 행정 구역의 로마자 표기
267	로마자		행정 구역 단위 중 '가(예: 퇴계로 3가)'의 로마자 표기
268	로마자		행정 구역 단위와 '가(예: 퇴계로 3가)'의 로마자 표기의 붙임표(-) 사용
269	로마자		'시, 군, 읍'의 행정 구역 단위의 로마자 표기에서의 생략
270	로마자		자연 지물명, 문화재명, 인공 구조물명의 로마자 표기의 붙임표(-) 사용
271	로마자		인명, 회사명, 단체명 등은 관행적 표기
272	로마자		전문 분야에서 한글 복원을 전제로 표기할 경우
273	로마자		발음상 혼동의 우려가 있을 경우의 로마자 표기
274	기호	마침표	날짜 표기에서 마침표 사용
275	기호	마침표	괄호 사용시 마침표 위치

276	기호	마침표		간접 인용 시 마침표 사용
277	기호	마침표		간접 인용 시 마침표 사용
278	기호	마침표		가운뎃점 대신 마침표 사용
279	기호	마침표		용언의 명사형이나 명사로 끝나는 문장의 마침표 사용
280	기호	물음표		물음표와 다른 기호(마침표, 느낌표 등)와 함께 사용
281	기호	물음표		제목이나 표어의 물음표 사용
282	기호	물음표		선택적인 물음이 이어질 때 물음표 사용
283	기호	물음표		물음의 정도가 약한 경우 물음표 사용
284	기호	물음표		의례적 인사(안녕하세요)에서 물음표 사용
285	기호	느낌표		문장 중간에 느낌표 사용
286	기호	느낌표		감탄사 뒤에 느낌표 사용(쉽표 사용과 비교)
287	기호	느낌표		제목의 느낌표 사용
288	기호	느낌표		의례적 인사(안녕하세요)에서 느낌표 사용
289	기호	쉽표		쉽표의 띄어쓰기
290	기호	쉽표		도치문에서 쉽표 사용
291	기호	쉽표		연결어미 뒤의 쉽표 사용
292	기호	가운뎃점		열거할 경우 가운뎃점 사용(쉽표 사용과 비교)
293	기호	가운뎃점		열거할 경우 가운뎃점 사용(빗금 사용과 비교)
294	기호	가운뎃점		특정한 의미가 있는 날(6.25, 4.19)을 표시할 경우 가운뎃점 사용
295	기호	가운뎃점		가운뎃점 띄어쓰기
296	기호	쌍점		쌍점, 반쌍점, 쉽표, 가운뎃점 쓰임의 차이
297	기호	쌍점		쌍점 띄어쓰기
298	기호	쌍점		시간 표기의 쌍점 사용
299	기호	빗금		빗금의 쓰임
300	기호	빗금		빗금의 띄어쓰기
301	기호	큰따옴표		직접 인용할 경우 큰따옴표 사용
302	기호	큰따옴표		간접 인용할 경우 큰따옴표 사용
303	기호	큰따옴표		큰따옴표 띄어쓰기
304	기호	큰따옴표		책 제목의 큰따옴표 사용
305	기호	작은따옴표		작은따옴표와 큰따옴표의 쓰임 차이
306	기호	작은따옴표		강조할 경우의 작은따옴표 사용(큰따옴표의 비교)
307	기호	작은따옴표		작은따옴표, 드래넵표, 밑줄의 쓰임 차이
308	기호	작은따옴표		예술 작품명의 작은따옴표 사용
309	기호	작은따옴표		인용한 말 안에 인용한 말을 쓸 경우 작은따옴표 사용
310	기호	소괄호		소괄호의 띄어쓰기
311	기호	중괄호		중괄호의 쓰임
312	기호	대괄호		괄호 안에 괄호를 쓸 경우 대괄호 사용
313	기호	겹낫표		겹낫표, 겹화살괄호, 큰따옴표 쓰임의 차이
314	기호	겹화살괄호		겹화살괄호의 쓰임
315	기호	활낫표		예술 작품의 제목, 상호, 법률, 규정의 활낫표 사용
316	기호	활화살괄호		예술 작품의 제목, 상호, 법률, 규정의 활낫표 사용
317	기호	줄표(-)		줄표의 띄어쓰기
318	기호	줄표(-)		줄표와 붙임표의 쓰임 차이
319	기호	붙임표(-)		붙임표의 띄어쓰기
320	기호	물결표(~)		물결표와 붙임표의 쓰임 차이
321	기호	물결표(~)		물결표의 띄어쓰기

322	기호	드래냄표(`)		드래냄표의 쓰임
323	기호	밑줄		밑줄의 쓰임
324	기호	숨김표(○, ×)		숨김표의 쓰임
325	기호	빠짐표(□)		빠짐표의 쓰임
326	기호	줄임표(……)		줄임표의 쓰임
327	기호	줄임표(……)		줄임표의 점의 위치(가운데, 아래쪽)
328	기호	줄임표(……)		줄임표의 점의 수
329	기호	줄임표(……)		줄임표의 띄어쓰기
330	기호	&		규정에 없는 기타 기호인 &의 쓰임
332	기호	별표(*)		규정에 없는 기타 기호인 별표의 쓰임
333	기호	아포스트로피		규정에 없는 기타 기호인 아포스트로피의 쓰임
334	기호	아포스트로피		아포스트로피와 작은따옴표의 쓰임 차이
335	기호	각주 번호		각주 번호 위치(마침표 앞뒤)
336	맞춤법			‘-예요/-이에요’의 맞춤법
337	맞춤법			‘-(으)르게요/-(으)르게요’의 맞춤법
338	맞춤법			‘-(으)르 거예요/-(으)르 거예요/-(으)르 꺼예요/-(으)르 꺼예요’의 맞춤법
339	맞춤법			‘-(으)로써/-(으)로서’의 맞춤법
340	맞춤법			‘-든지/-던지’의 맞춤법
341	맞춤법			‘아니오/아니요’의 맞춤법
342	맞춤법			‘아니예요/아니예요’의 맞춤법
343	맞춤법			‘어떻다’의 맞춤법 예) 어떻게요/어떻해요
344	맞춤법			‘말다’의 맞춤법
345	맞춤법			‘되다’의 맞춤법
346	맞춤법			‘뵈다’의 맞춤법
347	맞춤법			‘담그다’의 맞춤법
348	맞춤법			‘설레이다/설레다’의 맞춤법
349	맞춤법			‘파다/퍼다’의 맞춤법
350	맞춤법			‘띄다/뜨다’의 맞춤법
351	맞춤법			‘바라다’의 맞춤법
352	맞춤법			‘년/연’의 맞춤법 예) 년도/연도
353	맞춤법			‘-량/-양’의 맞춤법 예) 잔여량/잔여양
354	맞춤법			‘을/를’의 맞춤법 예) 실업률/실업률
355	맞춤법			‘횟수/회수’의 맞춤법
356	맞춤법			‘며칠/몇 일’의 맞춤법
357	맞춤법			‘웬/웬’의 맞춤법 예) 웬지/웬지, 웬일/웬일
358	맞춤법			‘데/대’의 맞춤법 예) 여행에서 간 데
359	맞춤법			‘설거지/설겅이’의 맞춤법
360	맞춤법			‘깨끗이/깨끗히’의 맞춤법
361	맞춤법			‘조용히/조용이’의 맞춤법
362	맞춤법			‘꼼꼼히/꼼꼼이’의 맞춤법
363	맞춤법			‘틈틈이/틈틈히’의 맞춤법
364	맞춤법			‘반반이/반반히’의 맞춤법
365	사전	표준국어대사전에 쓰인 기호 관련		표준국어대사전에 쓰인 기호 ‘-’의 기능
366	사전	표준국어대사전에 쓰인 기호 관련		표준국어대사전에 쓰인 기호 ‘^’의 기능
367	사전	표준국어대사전에 쓰인 기호 관련		표준국어대사전에 쓰인 기호 ‘≡’의 기능
368	사전	표준국어대사전에 쓰인 기호 관련		표준국어대사전에 쓰인 기호 ‘=’의 기능

369	사전	표준국어대사전에 쓰인 기호 관련		표준국어대사전에 쓰인 기호 '→'의 기능
370	사전	우리말샘 관련		우리말샘 로그인 및 접속 방법
371	사전	우리말샘 관련		우리말샘 단어 등재 상태(전문가 감수, 참여자 제안)
372	사전	우리말샘 관련		우리말샘 검색 방법
373	사전	우리말샘 관련		우리말샘 참여 방법
374	사전	우리말샘 관련		우리말샘 다운로드 방법
375	사전	표준국어대사전 관련		표준국어대사전과 우리말샘 차이
376	사전	표준국어대사전 관련		표준국어대사전 검색 방법
377	사전	표준국어대사전 관련		표준국어대사전 내려받기 방법
378	음성/음운	음운 현상 관련		음절의 끝소리 규칙
379	음성/음운	음운 현상 관련		연음
380	음성/음운	음운 현상 관련		된소리되기
381	음성/음운	음운 현상 관련		거센소리되기
382	음성/음운	음운 현상 관련		비음화
383	음성/음운	음운 현상 관련		유음화
384	음성/음운	음운 현상 관련		구개음화
385	음성/음운	음운 현상 관련		모음 조화
386	음성/음운	음운 현상 관련		모음 역행 동화
387	음성/음운	음운 현상 관련		자음 축약
388	음성/음운	음운 현상 관련		모음 축약
389	음성/음운	음운 현상 관련		ㅎ탈락
390	음성/음운	음운 현상 관련		ㄹ탈락
391	음성/음운	음운 현상 관련		모음 탈락
392	음성/음운	음운 현상 관련		반모음 첨가
393	음성/음운	음운 현상 관련		ㄴ첨가
394	음성/음운	개별 음운 관련		음운 개수
395	음성/음운	모음 관련		단모음
396	음성/음운	모음 관련		이중모음
397	음성/음운	모음 관련		반모음
398	음성/음운	모음 관련		전설 모음
399	음성/음운	모음 관련		후설 모음
400	음성/음운	모음 관련		고모음
401	음성/음운	모음 관련		중모음
402	음성/음운	모음 관련		저모음
403	음성/음운	모음 관련		평순 모음
404	음성/음운	모음 관련		원순 모음
405	음성/음운	자음 관련		입술소리
406	음성/음운	자음 관련		잇몸소리
407	음성/음운	자음 관련		센입천장소리
408	음성/음운	자음 관련		여린입천장소리
409	음성/음운	자음 관련		목청소리/후음
410	음성/음운	자음 관련		파열음
411	음성/음운	자음 관련		파찰음
412	음성/음운	자음 관련		마찰음
413	음성/음운	자음 관련		비음
414	음성/음운	자음 관련		유음
415	음성/음운	자음 관련		예사소리

416	음성/음운	자음 관련		된소리
417	음성/음운	자음 관련		가센소리
418	음성/음운	자음 관련		겹받침
419	음성/음운	불규칙 관련		ㄹ 불규칙
420	음성/음운	불규칙 관련		ㅎ 불규칙
421	음성/음운	불규칙 관련		ㄴ 불규칙
422	음성/음운	불규칙 관련		ㅅ 불규칙
423	음성/음운	불규칙 관련		ㄷ 불규칙
424	음성/음운	불규칙 관련		여 불규칙
425	음성/음운	불규칙 관련		우 불규칙
426	음성/음운	불규칙 관련		러 불규칙
427	음성/음운	불규칙 관련		가라 불규칙
428	음성/음운			가센소리
429	음성/음운			겹받침
430	순화어			'아재'의 순화어
431	순화어			'닭도리탕'의 순화어
432	순화어			'온라인'의 순화어
433	순화어			'유모차'의 순화어
434	순화어			'키오스크'의 순화어
435	순화어			'덕후'의 순화어
436	순화어			'오마카세'의 순화어
437	순화어			차별적 표현('눈먼 돈', '병어리 민 가슴', '절름발이')에 대한 순화어
438	순화어			'쪼리'의 순화어
439	순화어			'휴대폰'의 순화어
440	순화어			'SNS'의 순화어
441	순화어			'계란프라이'의 순화어
442	순화어			'고아원'의 순화어
443	순화어			'골든타임'의 순화어
444	순화어			'기라성'의 순화어
445	순화어			'나베'의 순화어
446	순화어			'다대기'의 순화어
447	순화어			'단도리'의 순화어
448	순화어			'메뉴'의 순화어
449	순화어			'벤치마킹'의 순화어
450	순화어			'사부사부'의 순화어
451	순화어			'스지'의 순화어
452	순화어			'쓰기다시'의 순화어
453	순화어			'알고리즘'의 순화어
454	순화어			'애매하다'의 순화어
455	순화어			'오뎅'의 순화어
456	순화어			'잉꼬부부'의 순화어
457	순화어			'클랙슨'의 순화어
458	순화어			'키워드'의 순화어
459	순화어			'탑'의 순화어
460	언어예절	호칭어	직장 관련 호칭	직장 상사 호칭
461	언어예절	호칭어	직장 관련 호칭	직장 부하 호칭

462	언어예절	호칭어	직장 관련 호칭	직장 상사의 배우자 호칭
463	언어예절	호칭어	가족 관련 호칭	작계 가족 호칭
464	언어예절	호칭어	가족 관련 호칭	배우자 가족 호칭
465	언어예절	호칭어	가족 관련 호칭	형제/자매 및 그 배우자에 대한 호칭
466	언어예절	호칭어	가족 관련 호칭	사돈 관계 호칭
467	언어예절	호칭어	가족 관련 호칭	촌수 기반 호칭
468	언어예절	호칭어	기타	직원과 손님 호칭
469	언어예절	호칭어	기타	씨의 존칭 여부
470	언어예절	호칭어	기타	친구의 아내/남편 호칭
471	언어예절	호칭어	기타	선생님의 아내/남편 호칭
472	언어예절	호칭어	기타	공식/비공식 상황에서 '님', '군', '양'의 적절성
473	언어예절	인사		고생하셨습니다', '수고하세요' 사용의 적절성
474	언어예절	인사		고맙습니다'와 '감사합니다' 사용의 적절성
475	언어예절	경어		압존법
476	언어예절	경어	주체 높임	선어말어미 '-사-'
477	언어예절	경어	주체 높임	'계시다'의 사용
478	언어예절	경어	주체 높임	'말씀하다'와 '말씀드리다'에서 '말씀'의 차이
479	언어예절	경어	객체 높임	특수 어휘 사용
480	언어예절	경어	상대 높임	존댓말(아주 높임)
481	언어예절	경어	상대 높임	존댓말(예사 높임)
482	언어예절	경어	상대 높임	존댓말(두루 높임)
483	언어예절	경어	상대 낮춤	반말(예사 낮춤)
484	언어예절	경어	상대 낮춤	반말(아주 낮춤)
485	언어예절	경어	상대 낮춤	반말(두루 낮춤)
486	언어예절	경어	겸양 표현	겸양 표현
487	언어예절	경어		과도한 높임
488	언어예절	경어		사물 존칭
489	언어예절	경어		간접 높임
490	언어예절	경어		이중 높임
491	언어예절	경어		당신의 쓰임
492	언어예절	경어		자기의 쓰임
493	문자			훈민정음 창제
494	문자			훈민정음 이체자
495	문자			훈민정음 소실 문자
496	문자			한글 자모음 수
497	문자			한글 기획 순서
498	국어사	어원		특정 어휘가 일본어, 중국어, 서구어 등에서 차용된 단어인지 여부
499	국어사	어원		특정 어휘의 어원(어원을 알 수 있는 경우)
500	국어사	어원		특정 어휘의 어원(출현 시기는 알 수 있으나 어원을 알 수 없는 경우)
501	국어사	어원		특정 어휘의 어원(출현 시기와 어원을 알 수 없는 경우)
502	국어사	어원		특정 어휘의 출현 시기
503	국어사	어원		특정 어휘의 옛말
504	국어사	어원		특정 어휘의 형태 변화
505	국어사	어원		특정 어휘의 의미 변화
506	건의			사전 등재 관련 건의
507	건의			사전 뜻풀이 수정 관련 건의

508	건의			사전 용법 관련 건의
509	건의			한자어 관련 건의
510	건의			어원 관련 건의
511	건의			띄어쓰기 관련 건의
512	건의			맞춤법 관련 건의
513	건의			표준 발음 관련 건의
514	건의			외래어 표기 관련 건의
515	건의			로마자 표기 관련 건의
516	건의			자료 제공 관련 건의
517	건의			서비스 개선 관련 건의
518	건의			시스템 개선 관련 건의
519	의미/화용	의미 차이		개발/계발
520	의미/화용	의미 차이		결제/결재
521	의미/화용	의미 차이		분배/배분
522	의미/화용	의미 차이		고맙습니다/감사합니다
523	의미/화용	의미 차이		난도/난이도
524	의미/화용	의미 차이		목제/목제
525	의미/화용	의미 차이		부딪히다/부딪치다
526	의미/화용	의미 차이		사용/이용
527	의미/화용	의미 차이		운영/운용
528	의미/화용	의미 차이		장애물/장애물
529	의미/화용	의미 차이		주관사/주관사
530	의미/화용	의미 차이		참고/참조
531	의미/화용	의미 차이		파난/파란
532	의미/화용	의미 차이		햇빛/햇볕
533	의미/화용	의미 차이		보전/보존
534	의미/화용	의미 차이		하지만/그러나
535	의미/화용	의미 차이		제한/금지
536	의미/화용	의미 차이		쫓다/쫓다
537	의미/화용	의미 차이		깨우치다/깨치다
538	의미/화용	의미 차이		사투리/방언
539	의미/화용	의미 차이		이송/운송
540	의미/화용	의미 차이		시연/시현
541	의미/화용	의미 차이		오해/착각
542	의미/화용	의미 차이		분실/망실
543	의미/화용	의미 차이		유래/유래
544	의미/화용	의미 차이		변하다/변화하다
545	의미/화용	의미 차이		성함/존함
546	의미/화용	의미 차이		지향하다/지향하다
547	의미/화용	의미 차이		배다/배다
548	의미/화용	의미 차이		끼달/이유
549	의미/화용	의미 차이		반증/생증
550	의미/화용	의미 차이		첫 번째/첫째
551	의미/화용	의미 차이		추수/수확
552	의미/화용	의미 차이		납부/납입
553	의미/화용	의미 차이		맞추다/맞히다
554	의미/화용	의미 차이		중지/정지

555	의미/화용	의미 차이		행동/행위
556	의미/화용	의미 차이		발간/발행
557	의미/화용	의미 차이		외래어/외국어
558	의미/화용	의미 차이		마법/마술/요술
559	의미/화용	의미 차이		동문/동창
560	의미/화용	의미 차이		중요하다/중시하다
561	의미/화용	의미 차이		열사/의사
562	의미/화용	의미 차이		신림/삼림
563	의미/화용	의미 차이		개연성/가능성
564	의미/화용	의미 차이		치사율/치명율
565	의미/화용	의미 차이		대비/대조
566	의미/화용	의미 차이		이제/인제
567	의미/화용	의미 차이		역행/후퇴
568	의미/화용	의미 차이		경축/축하
569	의미/화용	의미 차이		보유/소유
570	의미/화용	의미 차이		성과급/성과금
571	의미/화용	의미 차이		수요/소요
572	의미/화용	의미 차이		시간/시각
573	의미/화용	의미 차이		수익/수입
574	의미/화용	의미 차이		비혼모/미혼모
575	의미/화용	의미 차이		저감/절감
576	의미/화용	의미 차이		보강/방호/보안
577	의미/화용	의미 차이		구입/구매
578	의미/화용	의미 차이		~에 대한/~에 관한
579	의미/화용	의미 차이		절벽/낭떠러지
580	의미/화용	의미 차이		바지런하다/부지런하다
581	의미/화용	의미 차이		가지/개
582	의미/화용	의미 차이		다르다/틀리다
583	의미/화용	의미 차이		평안/편안
584	의미/화용	의미 차이		부문/부분
585	의미/화용	의미 차이		이왕/기왕
586	의미/화용	의미 차이		의미/뜻
587	의미/화용	의미 차이		괴멸/궤멸
588	의미/화용	의미 차이		원인/이유
589	의미/화용	의미 차이		일어나다/발생하다
590	의미/화용	의미 차이		분리수거/분리배출
591	의미/화용	의미 차이		지그시/지긋이
592	의미/화용	의미 차이		깍질/깍대기
593	의미/화용	의미 차이		누락
594	의미/화용	의미 차이		감지하다/탐지하다
595	의미/화용	의미 차이		실력/능력
596	의미/화용	의미 차이		곤욕/곤혹
597	의미/화용	의미 차이		연임/재선임
598	의미/화용	의미 차이		자택 근무/재택근무
599	의미/화용	의미 차이		도피/회피
600	의미/화용	의미 차이		퇴직/퇴임
601	의미/화용	의미 차이		배송/배달

602	의미/화용	의미 차이		안알하다/안이하다
603	의미/화용	의미 차이		따듯하다/따뜻하다
604	의미/화용	의미 차이		시험/실험
605	의미/화용	의미 차이		취임/부임
606	의미/화용	의미 차이		정확하다/적확하다
607	의미/화용	의미 차이		외국인/아방인
608	의미/화용	의미 차이		체제/체계
609	표준어/표준 발음	표준 발음		'의' 발음
610	표준어/표준 발음	표준 발음		많다, 많고, 많게
611	표준어/표준 발음	표준 발음		얕다, 얇고, 얇기, 얇는, 얇습니다
612	표준어/표준 발음	표준 발음		닭이, 닭을
613	표준어/표준 발음	표준 발음		흙을, 흙이, 흙하고, 흙은, 흙으로
614	표준어/표준 발음	표준 발음		밝다, 밝기
615	표준어/표준 발음	표준 발음		글개
616	표준어/표준 발음	표준 발음		뵈다, 뵈고
617	표준어/표준 발음	표준 발음		짧다, 짧네
618	표준어/표준 발음	표준 발음		넓다
619	표준어/표준 발음	표준 발음		여덟
620	표준어/표준 발음	표준 발음		넙죽하다
621	표준어/표준 발음	표준 발음		많다, 많은
622	표준어/표준 발음	표준 발음		짬닭
623	표준어/표준 발음	표준 발음		품삿을
624	표준어/표준 발음	표준 발음		값이
625	표준어/표준 발음	표준 발음		무릎이, 무릎을
626	표준어/표준 발음	표준 발음		끝을, 끝이
627	표준어/표준 발음	표준 발음		꽃아, 꽃이, 꽃 위
628	표준어/표준 발음	표준 발음		맛있다, 맛있는
629	표준어/표준 발음	표준 발음		맛없다
630	표준어/표준 발음	표준 발음		밭, 밭을, 밭아[랴], 밭이, 밭이
631	표준어/표준 발음	표준 발음		부엌, 부엌에
632	표준어/표준 발음	표준 발음		팔이
633	표준어/표준 발음	표준 발음		신고
634	표준어/표준 발음	표준 발음		날다
635	표준어/표준 발음	표준 발음		안다
636	표준어/표준 발음	표준 발음		히형
637	표준어/표준 발음	표준 발음		다금이
638	표준어/표준 발음	표준 발음		김밥
639	표준어/표준 발음	표준 발음		담임
640	표준어/표준 발음	표준 발음		효과
641	표준어/표준 발음	표준 발음		온라인
642	표준어/표준 발음	표준 발음		신리면
643	표준어/표준 발음	표준 발음		불법
644	표준어/표준 발음	표준 발음		원룸
645	표준어/표준 발음	표준 발음		공권력
646	표준어/표준 발음	표준 발음		학여울
647	표준어/표준 발음	표준 발음		줄넘기
648	표준어/표준 발음	표준 발음		밤양갱

649	표준어/표준 발음	표준 발음		띄어쓰기
650	표준어/표준 발음	표준 발음		참외
651	표준어/표준 발음	표준 발음		당소
652	표준어/표준 발음	표준 발음		교과서
653	표준어/표준 발음	표준 발음		선릉
654	표준어/표준 발음	표준 발음		버스
655	표준어/표준 발음	표준 발음		환난
656	표준어/표준 발음	표준 발음		물고기
657	표준어/표준 발음	표준 발음		깻잎
658	표준어/표준 발음	표준 발음		인간힘
659	표준어/표준 발음	표준 발음		상견례
660	표준어/표준 발음	표준 발음		깨끗이
661	표준어/표준 발음	표준 발음		일요일 월요일
662	표준어/표준 발음	표준 발음		청산리
663	표준어/표준 발음	표준 발음		신라
664	표준어/표준 발음	표준 발음		못 잊어
665	표준어/표준 발음	표준 발음		관세
666	표준어/표준 발음	표준 발음		값어치
667	표준어/표준 발음	표준 발음		해돋이
668	표준어/표준 발음	표준 발음		환율
669	표준어/표준 발음	표준 발음		쭈꾸미
670	표준어/표준 발음	표준 발음		의약품
671	표준어/표준 발음	표준 발음		솜이불
672	표준어/표준 발음	표준 발음		복유럽
673	표준어/표준 발음	표준 발음		볶음밥
674	표준어/표준 발음	표준 발음		못 읽어요
675	표준어/표준 발음	표준 발음		로서
676	표준어/표준 발음	표준 발음		들기름
677	표준어/표준 발음	표준 발음		네/니
678	표준어/표준 발음	표준 발음		끝인사
679	표준어/표준 발음	표준 발음		개수
680	표준어/표준 발음	표준 발음		환란
681	표준어/표준 발음	표준 발음		햇볕을
682	표준어/표준 발음	표준 발음		해님
683	표준어/표준 발음	표준 발음		태권도
684	표준어/표준 발음	표준 발음		창고
685	표준어/표준 발음	표준 발음		종소
686	표준어/표준 발음	표준 발음		잔불
687	표준어/표준 발음	표준 발음		있습니다
688	표준어/표준 발음	표준 발음		작열
689	표준어/표준 발음	표준 발음		인사말
690	표준어/표준 발음	표준 발음		외국
691	표준어/표준 발음	표준 발음		식용유
692	표준어/표준 발음	표준 발음		순가락
693	표준어/표준 발음	표준 발음		색연필
694	표준어/표준 발음	표준 발음		사건
695	표준어/표준 발음	표준 발음		산길

696	표준어/표준 발음	표준 발음		대구법
697	표준어/표준 발음	표준 발음		전화번호 전화
698	표준어/표준 발음	표준 발음		문화
699	표준어/표준 발음	표준 발음		의견란
700	표준어/표준 발음	표준 발음		외
701	표준어/표준 발음	표준 발음		옷 입다
702	표준어/표준 발음	표준 발음		웃어른의
703	표준어/표준 발음	표준 발음		어릴 적
704	표준어/표준 발음	표준 발음		안팎의
705	표준어/표준 발음	표준 발음		면류
706	표준어/표준 발음	표준 발음		떡볶이
707	표준어/표준 발음	표준 발음		대가
708	표준어/표준 발음	표준 발음		눈
709	표준어/표준 발음	표준 발음		놓치다
710	표준어/표준 발음	표준 발음		그렇군요
711	표준어/표준 발음	표준 발음		갈등
712	표준어/표준 발음	표준어		대거리
713	표준어/표준 발음	표준어		거꾸로/꺼꾸로
714	표준어/표준 발음	표준어		야채
715	표준어/표준 발음	표준어		깐부
716	표준어/표준 발음	표준어		닭도리탕
717	표준어/표준 발음	표준어		핸드폰/휴대폰
718	표준어/표준 발음	표준어		최애
719	표준어/표준 발음	표준어		으시대다
720	표준어/표준 발음	표준어		강낭콩
721	표준어/표준 발음	표준어		집콕
722	표준어/표준 발음	표준어		알터리
723	표준어/표준 발음	표준어		니환자
724	표준어/표준 발음	표준어		외톨이
725	표준어/표준 발음	표준어		자장면
726	표준어/표준 발음	표준어		짬짜미
727	표준어/표준 발음	표준어		덜등덜등하다
728	표준어/표준 발음	표준어		오뎅
729	표준어/표준 발음	표준어		암튼
730	표준어/표준 발음	표준어		따뜻하다/따듯하다
731	표준어/표준 발음	표준어		차년
732	표준어/표준 발음	표준어		보말
733	표준어/표준 발음	표준어		금번
734	표준어/표준 발음	표준어		쫄다
735	표준어/표준 발음	표준어		꼼짝달싹
736	표준어/표준 발음	표준어		진통
737	표준어/표준 발음	표준어		꿀값
738	표준어/표준 발음	표준어		꼬옥
739	표준어/표준 발음	표준어		날짜/날자
740	표준어/표준 발음	표준어		야구르트
741	표준어/표준 발음	표준어		날뻘치
742	표준어/표준 발음	표준어		갓다

743	표준어/표준 발음	표준어	놀래키다
744	표준어/표준 발음	표준어	개구장이/개구쟁이
745	표준어/표준 발음	표준어	희멀건
746	표준어/표준 발음	표준어	시보리
747	의미/화용	의미 차이 및 용법	겻 (으)르 거예요
748	의미/화용	의미 차이 및 용법	아/아서, 아/어 가지고
749	의미/화용	의미 차이 및 용법	고 (으)며
750	의미/화용	의미 차이 및 용법	(으)리지, (으)리 지
751	의미/화용	의미 차이 및 용법	아/어도 되다, 아/어야 하다
752	의미/화용	의미 차이 및 용법	(으)르게요, (으)르래요
753	의미/화용	의미 차이 및 용법	아/어지다, 게 되다
754	의미/화용	의미 차이 및 용법	거든요, 잡아요
755	의미/화용	의미 차이 및 용법	네요, 군요
756	의미/화용	의미 차이 및 용법	(으)르 것 같다, (으)르 모양이다
757	의미/화용	의미 차이 및 용법	(으)리/는/(으)르 줄 몰랐다
758	의미/화용	의미 차이 및 용법	(으)르결 후회, 추측
759	의미/화용	의미 차이 및 용법	느라고
760	의미/화용	의미 차이 및 용법	는 바람에
761	의미/화용	의미 차이 및 용법	(으)리/는 탓에, (으)리/는 덕분에
762	의미/화용	의미 차이 및 용법	던, 았/었던
763	의미/화용	의미 차이 및 용법	다가, 았/었다가
764	의미/화용	의미 차이 및 용법	(으)르 정도, (으)르 만큼
765	의미/화용	의미 차이 및 용법	더니, 았/었던니
766	의미/화용	의미 차이 및 용법	아/어 놀다, 아/어 두다
767	의미/화용	의미 차이 및 용법	스럽다/답다/롭다
768	의미/화용	의미 차이 및 용법	게/도록
769	의미/화용	의미 차이 및 용법	아/어 버리다, 고 말다
770	의미/화용	의미 차이 및 용법	(으)면, 다면, 더라면
771	의미/화용	의미 차이 및 용법	(으)니까, 아/아서
772	의미/화용	의미 차이 및 용법	(으)려고, (으)러
773	의미/화용	의미 차이 및 용법	고 있다, 아/어 있다
774	의미/화용	의미 차이 및 용법	아/어 오다, 아/어 가다
775	의미/화용	의미 차이 및 용법	아/어 보이다, 게 보이다
776	의미/화용	의미 차이 및 용법	자마자, 자
777	형태/통사		합성법
778	형태/통사		파생법
779	형태/통사		곡용
780	형태/통사		활용
781	형태/통사		품사(체언)
782	형태/통사		품사(용언)
783	형태/통사		품사(관형사)
784	형태/통사		품사(부사)
785	형태/통사		품사(독립언)
786	형태/통사		품사(조사)
787	형태/통사		품사(어미)
788	형태/통사		품사(접사)
789	형태/통사		품사(어근)

790	형태/통사			품사통용
791	형태/통사			높임
792	형태/통사			사제/상
793	형태/통사			문장 종결법(평서문)
794	형태/통사			문장 종결법(의문문)
795	형태/통사			문장 종결법(청유문)
796	형태/통사			문장 종결법(명령문)
797	형태/통사			문장 종결법(감탄문)
798	형태/통사			사동
799	형태/통사			파동
800	형태/통사			부정
801	형태/통사			어순
802	형태/통사			지시
803	형태/통사			접속문
804	형태/통사			내포문
805	형태/통사			담화표지
806	형태/통사			구어/문어
807	형태/통사			의성어, 의태어
808	형태/통사			문장 성분(주성분)
809	형태/통사			문장 성분(부속성분)
810	형태/통사			문장 성분(독립성분)
811	형태/통사			형태소 분석
812	형태/통사			고유어
813	형태/통사			한자어
814	맞춤법			명시형 전성어미 '-(으)ㄴ' 관련 맞춤법
815	맞춤법			네/예
816	맞춤법			니/내
817	맞춤법			남녀/남여
818	맞춤법			안/않
819	맞춤법			찌개/찌게
820	맞춤법			다행히/다행이
821	맞춤법			간간히/간간이
822	맞춤법			붙이다
823	맞춤법			베개/베게
824	맞춤법			붓다
825	맞춤법			바뀌었다/바꼈다
826	맞춤법			유월/육월
827	맞춤법			오랜만에/오랜만에
828	맞춤법			내요/내요
829	맞춤법			(의)십시오/(의)십시요
830	맞춤법			애들아/애
831	맞춤법			얼마만큼/얼마큼
832	맞춤법			헛갈리다/헛갈린다
833	맞춤법			만드려면/만들려면
834	맞춤법			아기/애기
835	맞춤법			맞추다/맞히다
836	맞춤법			토록

837	맞춤법			일부러/일부로
838	맞춤법			오뚜기/오투이
839	맞춤법			부딪히다/부딪치다
840	맞춤법			만만치 않다/만만하지 않다
841	맞춤법			매다/매다
842	맞춤법			개/재/애
843	맞춤법			결제/결재
844	맞춤법			한잔/한 잔
845	맞춤법			한번/한 번
846	맞춤법			대로/데로
847	맞춤법			니들/니들
848	맞춤법			르결/르결
849	맞춤법			덮히다/덮이다
850	맞춤법			더욱이/더우이
851	맞춤법			누이다/눠다
852	맞춤법			나누다/나눠다
853	맞춤법			치 않다/지 않다
854	맞춤법			이따가/이따
855	맞춤법			메우다/메꾸다
856	맞춤법			맞히다/맞추다
857	맞춤법			예기/얘기
858	맞춤법			알맞은/알맞은
859	맞춤법			곱빼기/곱빼기
860	맞춤법			불다/불다
861	맞춤법			들어나다/드려나다
862	맞춤법			흐트리다/흘뜨리다
863	맞춤법			되려/되레
864	맞춤법			깨닫다
865	맞춤법			애달다/애달쁘다
866	맞춤법			치루다/치르다
867	맞춤법			무엇/뭇
868	맞춤법			금세/금새
869	맞춤법			연월일 표기
870	맞춤법			대요/테요
871	맞춤법			냐고/으냐고
872	맞춤법			재요/재요
873	맞춤법			래요/래요

<기획·연구>

국립국어원 이현주 언어정보과장

국립국어원 김소희 학예연구사

국립국어원 김명주 연구원

연구 책임자 곽용진(㈜이르테크)

공동 연구원 김한샘(연세대), 함영균(㈜테디썸),
이지연(㈜이르테크), 오누리(㈜이르테크), 이진(세종대),
안익정(연세대), 임경태(연세대), 봉미경(연세대),
윤영민(연세대), 김선희(연세대)

연구 보조원 홍은기(㈜이르테크), 정해영(㈜이르테크), 박정혁(㈜이르테크),
김윤희(㈜이르테크), 박우정(㈜이르테크), 박하영(㈜이르테크),
최용기(㈜이르테크), 최주연(㈜이르테크), 오수연(㈜이르테크),
임수용(㈜이르테크), 하승목(㈜이르테크), 정은수(㈜이르테크),
박현이(㈜이르테크), 이현숙(㈜이르테크), 신현규(㈜이르테크),
이정은(㈜이르테크), 안소현(㈜이르테크), 김미영(㈜이르테크),
이승빈(㈜이르테크), 이윤경(㈜이르테크), 주민재(연세대),
안혜지(연세대), 천성호(연세대), 진예솔(연세대),
이소미(연세대), 강채안(연세대), 이윤복(연세대),
이이슬(㈜테디썸), 김민선(㈜테디썸), 정용빈(㈜테디썸),
김민선(㈜테디썸), 강예지(㈜테디썸)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2025년 12월 18일

발행일: 2025년 12월 18일

인 쇄: 세종기획

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘2025년 지시문 기반 첨삭
지원 말뭉치 연구 분석’ 사업의 결과물을 발간한 것입니다.

