

발간등록번호
11-1371028-100049-01

# 2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구

연구 책임자 | 신 서 인

국립국어원 2025-01-59

발간등록번호
11-1371028-100049-01

# 2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구

연구책임자  
신 서 인

## 제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 '2025년 한국어 구조 분석 및 무형  
대용어 복원 말뭉치 구축 연구'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2025년 04월 15일 ~ 2025년 12월 16일

2025년 12월 16일

연구책임자: 신서인(한림대학교 산학협력단)

연구 기관: 한림대학교 산학협력단

연구책임자: 신서인

공동연구원: 차정원

보조연구원: 이한범, 이수현, 박진현, 전상호

김나현, 이준희, 정민교, 최윤진

이민희, 서민지, 박수연, 유승빈

전광훈

## <국문 요약>

### 2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구

이 사업은 한국어 구조 분석 말뭉치(형태 분석 말뭉치, 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치)와 무형 대용어 복원 말뭉치를 구축하고 각각의 분석 지침을 재정비하는 것을 목적으로 한다.

이를 위해 국립국어원에서 배포한 ‘국립국어원 신문 말뭉치 2024’에서 ‘사회, 경제생활, 정치, IT/과학, 미용, 건강, 스포츠, 문화, 연예’ 등 다양한 주제 분야를 고르게 반영하여 20만 어절을, ‘국립국어원 일상 대화 말뭉치 2024’에서 2인 대화 위주로 10만 어절을 선별하여 총 30만 어절을 분석 대상으로 삼았다. 형태 분석 말뭉치와 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치, 무형 대용어 복원 말뭉치를 통합한 다층위 구조 분석 말뭉치 구축을 위하여 비식별화 기호가 포함되지 않은 문장을 위주로 분석 대상을 선별하였다. 또한 문어 자료의 경우 한 문서 내 문장의 길이 분포를 고려하여 10 어절 미만 문장이 70% 이상이거나 60 어절을 초과하는 문장이 30% 이상인 문서는 분석 대상에서 제외하고 주제별로 약 25,000 어절씩을 추출하여 분석을 진행하였다. 구어 자료의 경우에는 구조 분석이 어려운 4 어절 미만의 짧은 발화가 한 문서 내에 70% 이상인 경우 혹은 30 어절을 초과하는 문장이 문서의 30% 이상인 문서를 제외하고 주제별로 약 6,700 어절씩을 추출하여 분석을 진행하였다.

선별한 문장들은 오타자, 띄어쓰기 오류, 비표준 표기나 구어 자료에 나타나는 불필요한 중복 발화 등을 중심으로 정제 작업을 진행하였다. 또한 고유명과 <우리말샘>의 복합 명사나 구로 이루어진 등재어를 대상으로 띄어쓰기 방식을 변경하였다.

형태 분석 말뭉치 구축 시에는 <우리말샘>의 복합 명사나 구로 이루어진 등재어와 고유명을 하나의 어절로 분석하고 ‘-화, -적, -성’이나 ‘-하다/-되다/-받다/-당하다/-시키다’ 등의 접사를 분리하지 않고 어근에 통합된 형태로 분석하였다.

어휘의미 분석 말뭉치 구축 시에는 형태 분석 말뭉치를 바탕으로 어휘의미 분석 대상을 선정하고 한국어를 표기하기 위해 사용된 한자나 로마자, 숫자 표기에 대응되는 <우리말샘> 등재어에 대해서도 어휘의미 번호를 부여하였다.

구문 분석 말뭉치 구축 시에는 정제된 원문을 사용함에 따라 기존 지침에서 많은 부분을 차지하는 오류 처리 방안을 삭제하여 지침을 간소화하였으며 인용절을 부사절에 포함시켜

분석하였다.

의미역 분석 말뭉치 구축 시에는 번호형 의미역 분석 표지를 표지형으로 표시하고 필수 의미역과 부가 의미역 목록을 정비하여 논항과 부가어에 동일한 의미역 표지가 공유될 수 있도록 처리하였다.

무형 대용어 복원 말뭉치 구축 시에는 기존 지침에서 보이던 용어 사용의 혼란상을 정리하여 체계를 확립하고 무형 대용어를 지시 대명사와 조사를 결합한 형태로 복원하여 복원 성분의 가독성을 높였다.

다층위 구조 분석 말뭉치를 구축을 위해 개발된 워크벤치는 작업 중 오류를 발견하면 언제든지 이전 단계로 돌아가 작업 내용을 수정할 수 있는 작업 환경을 구축하였으며 교육과 실습으로 이루어진 작업자 교육과 작업 결과 정확도를 바탕으로 한 작업자 개인별 맞춤 피드백을 통해 말뭉치 정확도를 향상시킬 수 있는 작업 환경을 조성하였다.

**주요어:** 형태 분석, 어휘의미 분석, 구문 분석, 의미역 분석, 무형 대용어 복원, 다층위 구조 분석 말뭉치

<Abstract>

## A Research on the Construction of a Korean Structural Analysis and Zero Anaphora Resolution Corpus in 2025

This research aims to construct Korean structural analysis corpora—including morphological analysis, lexical semantic analysis, syntactic analysis, semantic role labeling—and an anaphora resolution corpus for zero pronouns, and to refine the respective annotation guidelines.

For this purpose, a total of 300,000 eojeols were selected as the analysis target: 200,000 eojeols from the National Institute of Korean Language (NIKL) Newspaper Corpus 2024, evenly representing topics such as society, economy, lifestyle, politics, IT/science, beauty, health, sports, culture, and entertainment, and 100,000 eojeols from two-person dialogues in the NIKL Daily Conversation Corpus 2024. To build a Korean Multi-layered Structural Analysis Corpus combining morphological, lexical semantic, syntactic, and semantic role annotations, sentences containing de-identification symbols were excluded. Furthermore, for written texts, documents in which sentences under 10 eojeols accounted for 70% or more or sentences exceeding 60 eojeols accounted for 30% or more were excluded. Approximately 25,000 eojeols per topic were then sampled for analysis. For spoken texts, documents were excluded if short utterances under 4 eojeols accounted for 70% or more or if sentences exceeding 30 eojeols were 30% or more; approximately 6,700 eojeols per topic were sampled.

All selected sentences underwent refinement focused on correcting typos, spacing errors, non-standard spellings, and unnecessary repeated utterances in spoken data. The spacing of proper nouns and multiword lexical entries listed in the *Urimalsaem* dictionary was also modified.

In constructing the Part-of-speech Tagged Corpus, proper nouns and multiword lexical entries in *Urimalsaem* were treated as single eojeols, and

affixes such as *-hwa*, *-jeok*, *-seong* and *-hada/doeda/batda/danghada/sikida* were not segmented but analyzed as integrated components of the stem.

In constructing the Word Sense Tagged Corpus, target lexemes were selected based on the morphologically annotated corpus, and lexical semantic sense numbers were assigned to *Urimalsaem* entries corresponding to Sino-Korean, Romanized, and numeric expressions.

In constructing the Dependency-Parsed Corpus, the syntactic annotation guideline was streamlined by removing error-handling sections made unnecessary due to the use of refined raw texts, and quotation clauses were incorporated into adverbial clauses to ensure consistency with the principle that subordinate clauses are analyzed as adverbial clauses.

In constructing the Semantic Role Labeling Corpus, the semantic role labeling guideline converted numeric semantic role markers into symbolic ones and reorganized the list of core and adjunct semantic roles so that both arguments and adjuncts may share the same role marker.

In constructing the Zero Anaphora Resolution Corpus, the guideline established a unified terminology system by eliminating inconsistencies found in previous versions, and represented zero anaphora as a combination of demonstratives and case particles to enhance the readability of the reconstructed elements.

A dedicated workbench developed for the construction of the Korean Multi-layered Structural Analysis Corpus provided an environment in which annotators could revisit previous stages to correct errors at any time. Training sessions combining instruction and hands-on practice, coupled with personalized feedback based on annotation accuracy, contributed to improving the overall quality and reliability of the corpus.

Keywords: Part-of-speech Tagged Corpus, Word Sense Tagged Corpus, Dependency-Parsed Corpus, Semantic Role Labeling Corpus, Zero Anaphora Resolution Corpus, Korean Multi-layered Structural Analysis Corpus

# 차 례

제 1 장 서론 .....	1
1. 사업의 목적 .....	2
2. 사업의 범위 .....	2
3. 사업 추진 체계 .....	4
4. 사업 추진 일정 .....	7
제 2 장 원문 선별 및 정제 .....	9
1. 원문 선별 .....	10
1.1. 대상 말뭉치 및 규모 .....	10
1.2. 원문 선별 기준 .....	10
2. 원문 정제 .....	12
2.1. 원문 정제 내용 .....	12
2.2. 작업 단위 변환 .....	16
2.3. 매핑 테이블 .....	18
제 3 장 형태 분석 말뭉치 구축 .....	20
1. 형태 분석 말뭉치 구축 지침 개선 .....	21
1.1. 기존 지침 보완 사항 .....	21
1.2. 형태 분석 말뭉치 구축 지침 .....	24
2. 형태 분석 말뭉치 구축 절차 .....	27
2.1. 형태 분석 초별 작업 .....	27



# 차 례

2.1.1. 형태소 분석 모델 및 입력 방식 .....	27
2.1.2. 말뭉치 초벌 결과 .....	27
2.2. 형태 분석 작업 중 확인된 문제점 및 개선 방향 .....	29
 제 4 장 어휘의미 분석 말뭉치 구축 .....	31
1. 어휘의미 분석 말뭉치 구축 지침 개선 .....	32
1.1. 기존 지침 보완 사항 .....	32
1.2. 어휘의미 분석 말뭉치 구축 지침 .....	34
2. 어휘의미 분석 말뭉치 구축 절차 .....	36
2.1. 어휘의미 분석 초벌 작업 .....	36
2.2. 어휘의미 분석 작업 중 확인된 문제점 및 개선 방향 .....	38
 제 5 장 구문 분석 말뭉치 구축 .....	40
1. 구문 분석 말뭉치 구축 지침 개선 .....	41
1.1. 기존 지침 보완 사항 .....	41
1.2. 구문 분석 말뭉치 구축 지침 .....	43
2. 구문 분석 말뭉치 구축 절차 .....	46
2.1. 구문 분석 초벌 작업 .....	46
2.1.1. 지배소(head) 후보 탐색 및 통사 기능 판별 .....	46
2.1.2. 의존 관계 설정 및 DP 구조 생성 .....	47
2.1.3. 구문 분석 말뭉치 초벌 결과 .....	48

# 차 례

2.2. 구문 분석 작업 중 확인된 문제점 및 개선 방향 .....	48
제 6 장 의미역 분석 말뭉치 구축 .....	50
1. 의미역 분석 말뭉치 구축 지침 개선 .....	51
1.1. 기존 지침 보완 사항 .....	51
1.2. 의미역 분석 말뭉치 구축 지침 .....	55
2. 의미역 분석 말뭉치 구축 절차 .....	58
2.1. 의미역 분석 초별 작업 .....	58
2.1.1. 구문 분석(DP) 기반 술어 후보 대상 탐지 .....	58
2.1.2. 의미역 정보 부착 .....	59
2.2. 의미역 분석 작업 중 확인된 문제점 및 개선 방향 .....	60
제 7 장 무형 대용어 복원 말뭉치 구축 .....	62
1. 무형 대용어 복원 말뭉치 구축 지침 개선 .....	63
1.1. 기존 지침 보완 사항 .....	63
1.2. 무형 대용어 복원 말뭉치 구축 지침 .....	71
2. 무형 대용어 복원 말뭉치 구축 절차 .....	72
2.1. 무형 대용어 복원 초별 작업 .....	72
2.1.1. 문장별 무형 대용어 복원 후보 생성 .....	73
2.1.2. LLM 기반 무형 대용어 복원 예측 .....	73
2.1.3. 복원 결과 정규화 및 ZA 구조 부착 .....	74
2.2. 무형 대용어 복원 작업 중 확인된 문제점 및 개선 방향 .....	78

# 차 례

제 8 장 다층위 구조 분석 말뭉치 구축 .....	80
1. 다층위 구조 분석 말뭉치 구축 지침 작성 .....	81
2. 다층위 구조 분석 말뭉치 구축 지침 .....	81
3. 다층위 구조 분석 말뭉치 구축의 의의와 확인된 문제점 .....	83
제 9 장 납품 자료의 품질 관리 .....	85
1. 시스템 개발 .....	86
1.1. 워크벤치 .....	86
1.2. 도구 개선 방향 .....	91
1.3. 기타 도구 개발 내역 .....	95
1.4. 기타 정보 추출 문서 목록 .....	96
2. 말뭉치 성능 평가 .....	98
2.1. 평가 방안 - 참조 기반 평가 .....	98
2.2. 평가 방안 - 다층 분석 말뭉치 간 일관성 검증 방안 .....	98
2.3. 품질 관리 및 검증 체계 .....	100
2.4. 기타 작업 형식 오류 목록 .....	100
3. 작업자 관리 및 의사소통 .....	102
3.1. 작업자 관리 .....	102
3.2. 구글 드라이브 운영 .....	107
3.2.1. 작업 지침 및 교육 자료 저장 .....	107
3.2.2. 작업 관련 이슈 관리 .....	107
3.3. 검수 관리 .....	108

# 차 례

3.4. 파일 관리 .....	110
3.4.1. 주 차별 구축 경과 .....	110
3.3.2. 중간 납품 경과 .....	111
3.4.3. 말뭉치 구축 결과 .....	112
 제 10 장 결론 .....	115
1. 요약 .....	116
2. 말뭉치 구축의 의의 .....	117
3. 제언 .....	118
 부록 .....	119
[붙임1] 형태 분석 말뭉치 구축 지침 .....	119
[붙임2] 어휘의미 분석 말뭉치 구축 지침 .....	178
[붙임3] 구문 분석 말뭉치 구축 지침 .....	195
[붙임4] 의미역 분석 말뭉치 구축 지침 .....	242
[붙임5] 무형 대용어 복원 말뭉치 구축 지침 .....	297
[붙임6] 다층위 구조 분석 말뭉치 구축 지침 .....	315
[붙임7] JSON 형식 .....	333
[붙임8] 대조 목록(원문 수정) .....	342
 참고문헌 .....	348

## 표 차례

<표 1> 문어 자료 주제별 문서 수와 어절 수 .....	10
<표 2> 구어 자료 주제별 문서 수와 어절 수 .....	11
<표 3> 원문 정제 내역 .....	13
<표 4> 문장 당 평균 어절 수 감소량 통계 .....	15
<표 5> 어근-접사 결합 예시 .....	15
<표 6> 단어절 구성의 형태 분석 .....	22
<표 7> 숫자, 로마자, 한자의 형태 분석 .....	22
<표 8> 고유명의 형태 분석 .....	23
<표 9> 형태소 분석 작업 중 발견된 초벌 오류 .....	29
<표 10> 단어절 구성의 어휘의미 분석 .....	33
<표 11> 로마자의 어휘의미 분석 .....	34
<표 12> 피인용문의 구문 분석 .....	42
<표 13> 주제어의 구문 분석 .....	43
<표 14> 장형 사동 구문의 구문 분석 .....	43
<표 15> 지배 의존 관계의 수 감소량 내역 예시 .....	48
<표 16> 지배 의존 관계의 수 감소량 통계 .....	48
<표 17> 격표지 실현에 따른 의미역 분석 .....	51
<표 18> 논항과 부가어의 의미역 분석 .....	52
<표 19> 표지형 의미역 분석 .....	52
<표 20> 논항과 부가어의 범위(스팬) .....	53

## 표 차례

<표 21> 이중주어문의 의미역 분석 .....	54
<표 22> 사동문과 피동문의 의미역 분석 .....	54
<표 23> ‘받다’ 구문의 의미역 분석 .....	55
<표 24> 연결어미의 의미에 따른 의미역 분석 .....	55
<표 25> 무형 대용어 복원의 선행어 탐색 범위 .....	63
<표 26> 무형 대용어 복원 관련 용어 정비 .....	64
<표 27> 무형 대용어 복원 말뭉치 JSON 구조 .....	65
<표 28> 서술성 명사 나열 구성과 서술어 분리 구성의 처리 .....	66
<표 29> 복원 대상 범위 .....	67
<표 30> 비지시적 대명사를 이용한 무형 대용어 복원 .....	67
<표 31> 복원 성분의 문장성분과 실현 형태 .....	68
<표 32> 복원 성분의 격표지 표시 .....	68
<표 33> 이중주어문의 무형 대용어 복원 .....	69
<표 34> 수량 표현의 무형 대용어 복원 .....	70
<표 35> ‘화자’ 및 ‘청자’를 이용한 무형 대용어 복원 .....	70
<표 36> 의존명사 구성의 무형 대용어 복원 .....	70
<표 37> 문어/구어 무형 대용어 복원 서술어 수량 .....	75
<표 38> 선행어 및 비지시 대명사 복원 비율 .....	75
<표 39> 선행어 유형별 복원 비율 .....	76
<표 40> 비지시 대명사 유형별 복원 비율 .....	76

## 표 차례

<표 41> 비지시 보어, 필수적 부사어 대명사 목록(문어) .....	76
<표 42> 비지시 보어, 필수적 부사어 대명사 목록(구어) .....	76
<표 43> 선행어와 서술어 간 문장 거리 비율 .....	78
<표 44> 한국어 통합 분석 도구 개선 내역 .....	92
<표 45> 말뭉치 관리용 관리자 도구(admin) 기능 .....	95
<표 46> 기타 정보 추출 문서 목록 .....	96
<표 47> 문어 말뭉치 성능 평가(F1) 결과 .....	98
<표 48> 구어 말뭉치 성능 평가(F1) 결과 .....	98
<표 49> 일관성 검사 목록 및 내용 .....	99
<표 50> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 형식 오류 .....	100
<표 51> 한국어 구조 분석 말뭉치 작업자 선정 기준 .....	102
<표 52> 한국어 구조 분석 말뭉치 작업자 명단 .....	102
<표 53> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 교육 일정 ·	104
<표 54> 의미역 분석 및 무형 대용어 추가 교육 일정 .....	106
<표 55> 주요 회의 일자 .....	109
<표 56> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 내역 .....	110
<표 57> 중간 납품 일정 .....	112
<표 58> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 결과 .....	113

## 그림 차례

[그림 1] 추진 체계 .....	4
[그림 2] 구축 프로세스 .....	5
[그림 3] 전체 사업 일정 .....	8
[그림 4] 우리말샘 샷갓표 포함 어휘 .....	14
[그림 5] 우리말샘 띄어쓰기 포함 어휘 .....	14
[그림 6] 말뭉치 정제 결과 어절 수 변경 내역 .....	14
[그림 7] 국립국어원 신문 말뭉치 2024 json 구조 .....	16
[그림 8] 문어 작업 말뭉치 json 구조 .....	17
[그림 9] 국립국어원 일상 대화 말뭉치 2024 json 구조 .....	17
[그림 10] 구어 작업 말뭉치 json 구조 .....	18
[그림 11] 매핑 테이블 SUB_Info_sheet .....	19
[그림 12] 매핑 테이블 GEN_Info_sheet .....	19
[그림 13] 매핑 테이블 REF_Info_sheet .....	19
[그림 14] 형태 분석 말뭉치 초벌 출력값 예시 .....	27
[그림 15] 어휘의미 분석 모델 구조 .....	36
[그림 16] 어휘의미 분석 말뭉치 초벌 말뭉치 출력값 예시 .....	37
[그림 17] 무형 대용어 복원 말뭉치 구축 모델 구조 .....	72
[그림 18] 무형 대용어 말뭉치 초벌 출력값 예시 .....	74
[그림 19] 선행어와 서술어 간 문장 거리 분포 .....	78
[그림 20] 통합 분석 도구 전체 화면 .....	86



## 그림 차례

[그림 21] 형태소 분석 행 추가/삭제 기능 .....	88
[그림 22] 형태소 분석 토큰 수 자동 검수 기능 .....	88
[그림 23] 어휘의미 분석 가이드라인 <우리말샘> .....	88
[그림 24] 구문 분석 작업 결과 화면 .....	89
[그림 25] 의미역 분석 작업 방법 .....	90
[그림 26] 무형 대용어 복원 작업 방법 .....	91
[그림 27] 말뭉치 관리용 관리자 도구(admin) 화면 .....	96
[그림 28] 관리자용 웹앱 화면 .....	96
[그림 29] 작업자에게 전달된 오류 리포트 화면 .....	105
[그림 30] 정확률 대조 확인 화면 .....	105
[그림 31] 의미역 분석의 주차별 정확률 비교 결과 .....	106
[그림 32] 구글 드라이브 화면 .....	107
[그림 33] 질의응답 시트 관리 .....	108
[그림 34] 질의응답 시트 내용 .....	108
[그림 35] 어휘의미 분석 검수 화면 .....	114
[그림 36] 의미역 분석 통합 검수 진행 화면 .....	114



# 제 1 장

## 서론



# 1. 사업의 목적

본 사업은 한국어 구조 분석 말뭉치(한국어 형태 분석 말뭉치, 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치)와 무형 대용어 복원 말뭉치를 구축하고 분석 지침을 재정비하는 것을 목적으로 한다. 최근 인공지능 언어 모델의 한국어 처리 기술이 고도화되고 한국어 특화 모델을 개발하는 데에 고품질 한국어 말뭉치의 수요가 증가하게 되었다. 또한, 국내 표준화 및 참조 기반 자료가 될 수 있는 정밀한 언어 정보 부가 말뭉치를 분석하고 배포하는 일이 필요하게 되었다. 이 사업에서 구축하는 한국어 구조 분석 말뭉치와 무형 대용어 복원 말뭉치는 한국어에 특화된 인공지능 언어 모델 개발에 기여할 수 있을 것이며, 분석 지침은 국내 표준화 및 참조 기반 자료로 활용될 수 있을 것이다.

# 2. 사업의 범위

이 사업에서는 (1) 한국어 구조 분석 말뭉치를 구축하고, (2) 무형 대용어 복원 말뭉치를 구축하였으며, (3) 분석 지침을 재정비하고, (4) 납품 자료의 품질 보증 및 보완 체계를 수립하였다.

첫째, 국립국어원에서 제공한 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’를 대상으로 문어 20만 어절 및 구어 10만 어절로 구성된 총 30만 어절 이상 규모의 원문을 선정하여 한국어 구조 분석을 진행하였다.

‘국립국어원 신문 말뭉치 2024’에서 ‘사회, 경제생활, 정치, IT/과학, 미용, 건강, 스포츠, 문화, 연예’ 등 9개 주제에서 분석 대상 자료를 25,000 어절씩 주제별로 균등하게 선별하여 문어 말뭉치 원문 20만 어절을 구성하고, ‘국립국어원 일상 대화 말뭉치 2024’에서 화자와 청자의 복원 등을 위하여 2인 이하의 대화만을 대상으로 ‘영화/드라마/전시회/공연, 먹거리/맛집/요리법, 스포츠/레저/취미/게임/만화, 건강/다이어트/질병’ 등 15개 주제에서 분석 대상 자료를 6,700 어절씩 주제별로 균등하게 선별하여 구어 말뭉치 원문 10만 어절을 구성하였다.

선별한 원문 자료는 <우리말샘>의 등재어를 기준으로 오류 및 띄어쓰기를 자동으로 수정하는 절차를 거치고 다시 전문 작업자들의 검토와 오류 수정 과정을 통해 양질의 한국어 구조 분석 말뭉치 구축이 가능한 원문으로 정제하였다.

한국어 구조 분석 말뭉치는 한국어 구조 분석 각 단계의 기반이 되는 형태 분석 말뭉

치와 체언류 및 용언류 대상 다의어 수준의 어휘의미 분석 말뭉치, 문장 내 각 어절의 지배소와 의존 관계 정보를 부착하는 구문 분석 말뭉치, 문장 내 필수 의미역과 부가 의미역 등 논항 정보를 부착하는 의미역 분석 말뭉치 등 총 4종의 말뭉치를 구축하였다.

둘째, 한국어 구조 분석 말뭉치 결과물을 대상으로 서술어의 필수 논항인 주어(SBJ), 목적어(OBJ), 보어(CMP), 필수적 부사어(AJT)의 생략 여부를 판단하고 생략어에 대응하는 선행어를 복원하여 무형 대용어 복원 말뭉치를 구축하였다. 문어와 구어 모두 분석 대상 서술어가 포함된 앞 15 문장과 뒤 2 문장 이내에서 생략어에 대응하는 선행어를 복원하였으며, 이후 무형 대용어 복원 말뭉치와 층위별 한국어 구조 분석 말뭉치 사이의 형식을 검증하고 분석의 일관성을 검증하였다.

또한 층위별로 형태 분석 정보와 어휘의미 분석 정보, 구문 분석 정보, 의미역 분석 정보, 무형 대용어 복원 정보가 하나로 통합된 다층위 통합 분석 말뭉치를 구축하고 각 층위별 분석 정보들이 충돌하는지 여부를 점검하여 문제점을 수정하는 데 힘썼다.

셋째, 국립국어원에서 진행한 기존 연구 사업의 성과를 검토하고 각 연구 사업의 결과로 공개된 말뭉치 구축 지침을 분석하였으며 층위별 한국어 구조 분석 말뭉치 구축을 위한 통합 지침을 수립하였다.

형태 분석 및 어휘의미 분석에는 국립국어원의 ‘모두의 말뭉치’ 누리집에 공개된 ‘2020년 어휘의미 말뭉치 연구 분석 사업(2020)’을 검토하여 그 결과를 말뭉치 구축에 반영하였으며, 의미역 분석에는 ‘의미역 분석 말뭉치 구축(2019)’ 사업 결과를, 구문 분석과 무형 대용어 복원에는 ‘2024년 구문 분석 및 무형 대용어 복원 말뭉치 연구’ 사업의 결과를 반영하여 통합 지침의 질적 향상을 도모하였다.

지난 말뭉치 구축 사업과 달리 본 사업에서는 원시 말뭉치의 오류 수정이 가능해짐에 따라 기존 지침에서 오류 포함 어절을 처리하기 위해 마련되었던 상당 부분을 배제하고 층위별 분석의 본령에 집중할 수 있는 환경이 마련되었다. 이에 이번 사업에서는 언어학적 정합성과 분석의 타당성을 확보하기 위해 지침의 질적 향상을 도모하였다. 개선된 지침은 층위별 한국어 구조 분석 말뭉치 통합을 고려하여 분석 지침 간 일관성을 유지하도록 하였으며 그에 더하여 각 분석 작업의 효율성과 정확성을 제고하기 위한 보완책을 마련하고 국립국어원과의 협의를 통하여 통합 지침을 확정하였다.

넷째, 완성된 산출물 납품 후 품질 보증과 사후 보완 체계를 수립하였다. 한국어 구조 분석 말뭉치 5종(형태 분석, 어휘의미 분석, 구문 분석, 의미역 분석, 무형 대용어 복원)을 구축해 본 경험이 있는 국어학 전공 석박사 과정 이상 학력의 전문 작업자를 선정하여

기계 분석 결과를 바탕으로 형태 분석 정보와 어휘의미 분석 정보, 구문 분석 정보, 의미역 분석 표지를 부착하였고 무형 대용어를 복원한 후, 일관성 검사를 진행하고 그 결과를 바탕으로 오류를 수정하여 납품 자료 전체에 대한 일관성 및 정확성을 확보하였다.

반복적인 오류를 자동적으로 수정할 수 있는 관리 체제를 설계하였고, 구축 데이터의 전체 일관성 및 정확성을 검증할 수 있는 방안을 마련하였다. 납품 자료에 문제가 발생할 시 즉각적인 대처를 통해 합리적인 대안을 마련하고 최종 분석 자료 납품 이후에도 보완 작업을 통해 납품 자료의 질을 보장할 수 있는 여건을 조성하였다.

### 3. 사업 추진 체계

본 사업은 국립국어원이 주관하고 한림대학교와 데이터리의 두 기관이 함께 진행하였다.

국어학 전문가 인력을 보유한 한림대학교는 한국어 구조 분석 방법론을 수립하여 말뭉치 구축 지침을 작성하고 작업자를 모집하여 교육, 관리하였으며, 데이터리는 말뭉치 구축 대상의 원문 선별과 정제, 말뭉치 구축을 위한 도구 개발과 자동화된 말뭉치 품질 관리 체계를 구축하였다.

그리고 사업을 주관하는 국립국어원은 사업의 방향을 설정하고 논의 사항을 최종 결정하는 것으로 최종 결과물의 수준을 판단하였다.



[그림 1] 추진 체계

사업 수행은 크게 ‘원문 선정 단계’, ‘분석 말뭉치 구축 단계’, ‘품질 검증 단계’로 나누

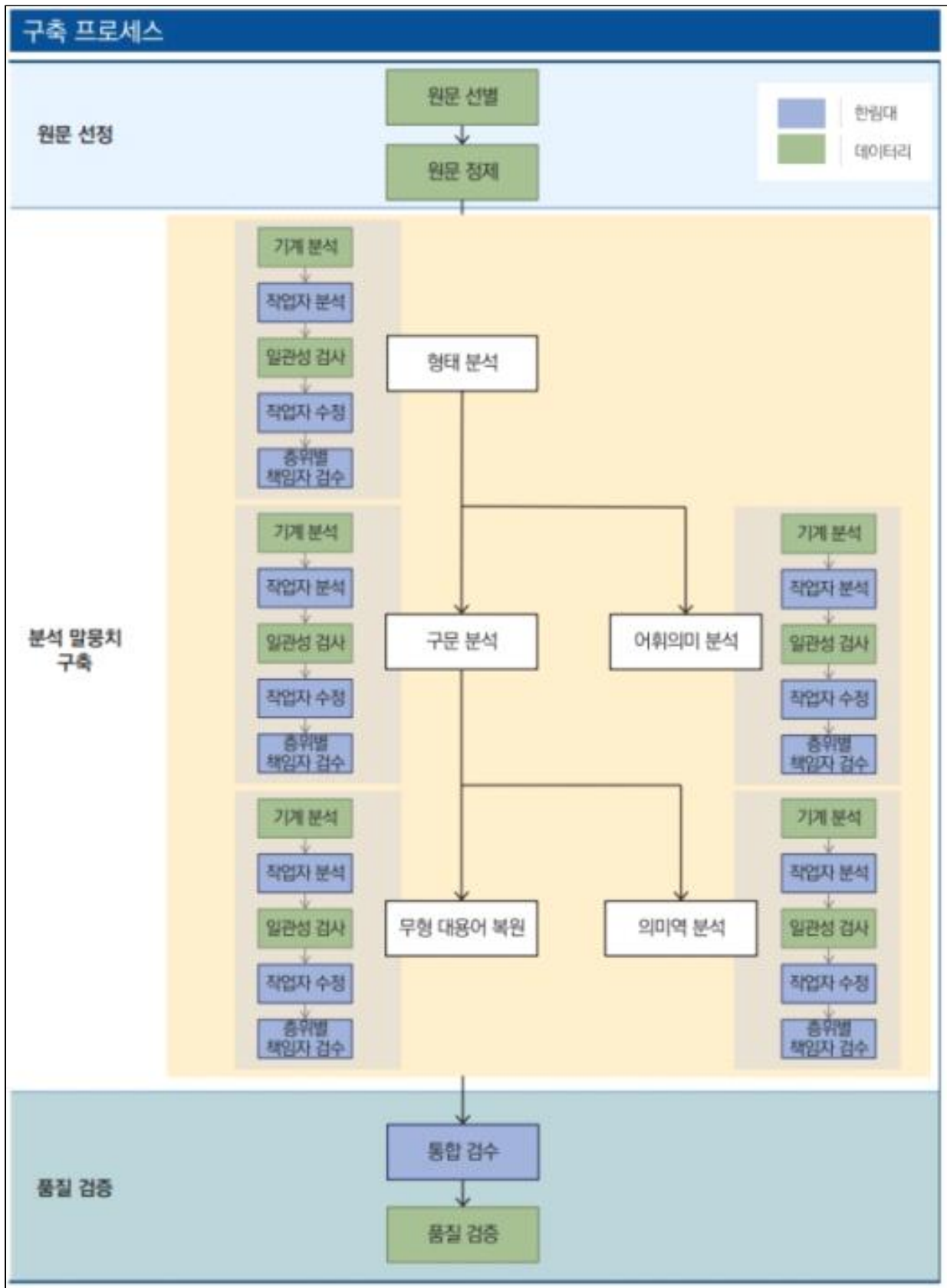
어 진행하였다.

원문 선정 단계는 ‘원문 선별’과 ‘원문 정제’의 과정을 통해 이루어졌다. 데이터리가 대상 말뭉치에서 분석 가능한 원문을 자동으로 선별하고 1차 정제하면 한림대는 정제된 원문을 다시 검토하고 수정이 필요한 부분이 있는지 확인하여 분석이 가능한 형태의 말뭉치를 확보하였다.

분석 말뭉치 구축 단계는 ‘분석’과 ‘검사 및 수정’, ‘검수’로 이루어졌다. 데이터리에서 보유하고 있는 층위별 한국어 분석 기술을 통해 기계 분석이 완료되면 국어학 분야의 전문 지식을 가진 작업자들에 의해 기계 분석 결과 검토 및 작업자에 의한 분석이 진행되었다. 분석된 결과는 기계 처리를 통한 일관성 검사를 거쳐 수정 요소를 탐지하고 작업자 검토 및 수정 과정을 통해 분석의 정확성을 높였다. 다음으로 작업자에 의해 수정이 완료된 작업 결과는 층위별 검수 책임자에 의한 검수를 통해 완성도 높은 분석 말뭉치를 구축할 수 있었다.

이와 같은 분석 말뭉치 구축 단계는 형태 분석이 완료된 시점부터 ‘구문 분석’과 ‘어휘 의미 분석’, ‘무형 대용어 복원’과 ‘의미역 분석’ 과정으로 이원화되어 진행되었다. 각 분석 과정은 층위별 전담 작업자들이 관리하게 하여 작업 집중도를 높이고 작업 결과물의 완성도를 제고하였다.

품질 검증 단계는 ‘품질 검수’와 ‘품질 검증’ 과정으로 진행하였다. 전문가에 의한 품질 검수를 거친 작업 결과는 전문 구축 인력에 의해 구축된 정답 말뭉치와의 비교를 통해 성능 검증 과정을 거쳐 F1 0.95~0.99 이상의 평균 성능 점수를 나타내는 최종 구축 품질을 확보하였다.



[그림 2] 구축 프로세스

## 4. 사업 추진 일정

사업 전체 일정은 아래와 같다.

2025년 4월 22일 착수보고회를 시작으로 2주간의 원문 정제와 도구 개발 기간을 거치며 작업자 관리 지침을 마련하였다. 이후 5월 한 달 동안 총 30만 어절의 형태 분석 말뭉치 구축을 완료하였다.

형태 분석 말뭉치 구축 종료 후 6월 첫째 주에 어휘의미 분석과 구문 분석 작업자 교육을 실시하였고 이후 8주에 걸쳐 어휘의미 분석 말뭉치와 구문 분석 말뭉치 구축 및 검수 작업을 진행하였다. 어휘의미 분석과 구문 분석 말뭉치 검수 종료 후 8월 4일부터 8월 18일까지는 구축된 어휘의미 분석과 구문 분석 말뭉치의 정비를 진행하였는데 정비 기간에는 원문 오류, 형태 분석 오류 등을 확인하고 수정하였다.

사업 시작 4개월이 되는 9월 5일에는 형태 분석, 어휘의미 분석, 구문 분석 말뭉치 100% 납품을 완료하였다.

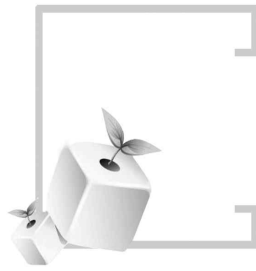
8월 18일과 19일에 의미역 분석 및 무형 대용어 복원 말뭉치 작업자 교육을 진행하고 8월 26일부터 11월 11일까지 의미역 분석 및 무형 대용어 복원 말뭉치 구축을 진행하였으며 사업 6개월 차인 10월 13일에는 형태, 어휘의미, 구문 분석 말뭉치 100%와 의미역 분석, 무형 대용어 복원 말뭉치 50%를 완료하였다. 의미역 분석은 총 9주간 진행하였고, 무형 대용어 복원 말뭉치 구축은 의미역 분석 말뭉치 검수가 끝난 작업물을 바탕으로 11주 동안 진행하였다. 의미역 분석 및 무형 대용어 복원 말뭉치 구축 작업 모두 각 주 차별로 검수를 병행하였으며, 11월 18일 무형 대용어 복원 말뭉치 검수 작업을 끝으로 총 위별 말뭉치 5종(형태 분석 말뭉치, 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치, 무형 대용어 복원 말뭉치)의 구축 및 검수 작업을 마무리하였다.

무형 대용어 검수 완료 이후인 11월 18일부터 12월 3일까지는 최종 납품을 위한 통합 검수 작업을 진행하였다.



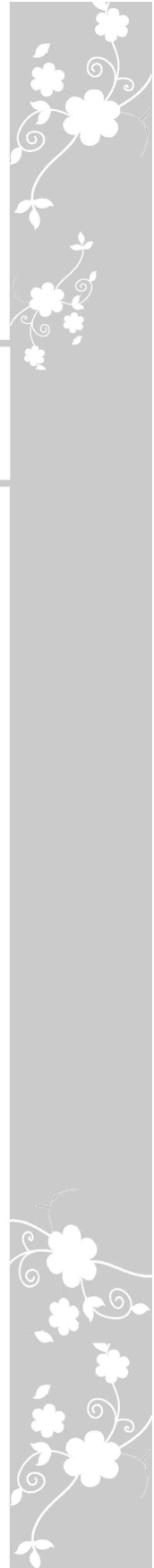
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

[그림 3] 전체 사업 일정



## 제 2 장

### 원문 선별 및 정제



# 1. 원문 선별

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업의 원문 선정 단계는 ‘원문 선별’과 ‘원문 정제’의 과정을 통해 이루어졌다. 데이터리에서는 대상 말뭉치에서 분석 가능한 원문을 자동으로 선별하고 1차 정제하였으며 한림대에서는 정제된 원문을 다시 검토하고 수정이 필요한 부분이 있는지 확인하는 과정을 통해 분석이 가능한 형태의 말뭉치를 확보하였다.

## 1.1. 대상 말뭉치 및 규모

본 사업은 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’를 대상으로 한다. 문어 자료는 ‘국립국어원 신문 말뭉치 2024’에서 20만 어절을, 구어 자료는 ‘국립국어원 일상 대화 말뭉치 2024’에서 10만 어절을 선별하여 총 30만 어절을 분석 대상으로 삼았으며, 정밀한 형태분석 말뭉치, 다의어 수준의 어휘의미 분석 말뭉치를 구축해야 하므로 최대한 비식별화 기호가 포함되지 않은 문장을 선별하였다.

## 1.2. 원문 선별 기준

문어 자료의 경우 ‘사회, 경제생활, 정치, IT/과학, 미용, 건강, 스포츠, 문화, 연예’ 등 다양한 주제 분야를 고르게 반영하여 실제 사용 빈도가 높은 일반 언어 사용 양상을 확보하고자 하였다. 정치·사회적으로 민감한 사안이나 편향 가능성이 큰 자료는 국립국어원과 협의를 통해 제외하였으며, 한 문서 내 문장 길이 분포를 고려하여 10 어절 미만 문장이 70% 이상이거나 60 어절을 초과하는 장문이 한 문서 내 30% 이상인 문서는 분석 대상에서 제외하였다. 이후 주제별 약 25,000 어절씩 균등하게 선별한 문서 수와 상세 어절 수는 아래 <표 1>과 같다.

주제	문서 수 (단위: 건)	어절 수 (단위: 어절)
생활	119	25,263
경제	106	25,090
사회	110	25,244
정치	108	25,164
문화	120	25,013
IT/과학	107	25,099
스포츠	113	25,175

미용/건강	119	25,234
연예	125	25,024
총합	1,027	226,306

<표 1> 문어 자료 주제별 문서 수와 어절 수

구어 자료는 2인 대화 위주로 선별하여 화자·청자 복원과 무형 대용어 복원에 적합한 맥락 정보를 확보하였으며, 4 어절 미만의 극단적으로 짧은 발화가 과도하게 집중된 대화는 해당 발화가 한 문서 내에 70% 이상인 경우로 기준을 두어 제외하였다. 또한, 하나의 문장 길이가 30 어절을 초과하는 경우가 한 문서의 30% 이상인 문서도 제외하였다. 선별 기준에 맞는 문서들을 대상으로 주제별 최소 6,700 어절씩 추출하였으며 선별된 문서 수와 상세 어절 수는 아래 <표 2>와 같다.

주제	문서 수 (단위: 건)	어절 수 (단위: 어절)
방송_연예_예능_아이돌_한류	5	8,377
취직_아르바이트	5	8,261
반려동물_반려용품	6	8,030
여행_휴가_휴일_자연휴양지	7	7,971
우정_성격_MBTI	5	7,961
먹거리_맛집_요리법	5	7,785
건강_다이어트_질병	5	7,702
스포츠_레저_취미_게임_만화_책_독서	5	7,571
생활_주거환경	5	7,493
영화_드라마_전시회_공연	5	7,380
스마트기기_인공지능_메타버스등_IT	4	7,350
쇼핑_선물	4	6,943
경제_재테크_부동산_금융	4	6,806
회사_학교_학창시절	5	6,741
연애_결혼_가족_관혼상제	5	6,734
총합	75	113,105

<표 2> 구어 자료 주제별 문서 수와 어절 수

## 2. 원문 정제

### 2.1. 원문 정제 내용

원문 선별 이후에는 분석의 일관성과 정확도에 영향을 미치는 오타자, 띄어쓰기 오류, 비표준 표기 등을 중심으로 정제 작업을 수행하였다. 구어 말뭉치의 경우 수집된 데이터 중 전사자의 오류, 불필요한 중복 발화 등도 포함되어 있어 원문 수정을 진행하였다. 문어 말뭉치는 원문을 그대로 유지하되 우리말샘 등재 단어와 비교하여 고유명사 및 복합명사 표기를 변경하였다.

#### (1) 말뭉치 오류(오타자·비표준 발화)

정제 과정에서 가장 많이 발견된 오류는 ‘오타자 및 띄어쓰기 오류, 비표준 표기’ 등이었다. 전체 정제 내역 중 해당 오류가 가장 많았으며, 대부분 아래와 같은 형태의 오류 수정에 해당한다.

- 전사자의 오류(오타자)
  - 오타자 및 조사 사용 오류
  - 예: 또 대사 활동도 원활해지기 때문에 → 또 대사 활동도 원활해지기 때문에
- 구어 특성으로 인한 비정형 표기 수정
  - 발음 기반 표기 수정
  - 예: 그런 게 쪼끔 고민인데 → 그런 게 조금 고민인데
- 규칙적 띄어쓰기 오류 정상화
  - 예: “그래서그냥” → “그래서 그냥”

#### (2) 불필요한 발화

정제 과정에서 총 2,799 문장(전체의 17.0%)이 불필요한 반복 발화로 확인되어 삭제·병합 처리되었다. 이는 구어 말뭉치 수집 시 흔히 발생하는 동일 문장의 반복 발화, 자연스러운 대화 중 중복 확인 응답, 수집 장치의 중복 기록 등으로 확인되었다.

- 동일 화자의 반복 어절/문장  
예: “네 네 네…”, “아 예 예 그거요 그거요”
- 발화 중단·머뭇거림 제거  
예: “그… 그제…” → “그제…”
- 비언어적 표기 삭제  
예: “그런 방 방편으로다가 그 흡연하는 것이 xxx” → “그런 방편으로다가 그 흡연하는 것이”
- 구어적 반복어미·중첩어 제거 또는 일관화  
예: “했었었어요” → “했어요”

### (3) 우리말샘 등재 단어 비교

위 과정으로 정제된 말뭉치 대상, 복합 명사 및 고유명사 형식을 수정하였다. 복합 명사는 <우리말샘>에 ‘^’로 연결되어 등재된 단어를 기준으로 하였으며, 고유명사는 띄어 쓰기로 등재된 단어 목록을 추출하여 비교하는 과정을 거쳤다.

<우리말샘>의 등재어 중 ‘^’로 형태가 나뉜 경우는 공백 없이 붙여쓰기로 처리하여 불필요한 분석 소요를 줄이고, 구나 절로 이루어진 <우리말샘> 등재어 및 고유명사나 로마자에 대해서는 언더바 기호 ‘\_’로 띄어쓰기 공백을 대체하여, 띄어쓰기를 유지하면서도 분석 대상을 한 어절로 취급하여 분석의 용이성을 확보하고 분석 단위의 일관된 처리를 가능하게 하였다.

정제 후	<우리말샘> 표제어	비고
갈색_머리_001/NNP	갈색 머리_001	2003년에 발표된 윤건의 앨범 수록곡
갈색가전_001/NNG	갈색^가전_001	텔레비전, 오디오 따위 주로 거실과 안방에서 사용하는 가전제품군.
indigo_blue_001/NNG	인디고-블루(indigo blue)_001	어두운 파랑.
brownbagmeeting_001/NNG	브라운^백^미팅(brown bag meeting)+001	점심 식사를 곁들이면서 편하고 부담 없이 하는 토론.
I_still_believe_001/NNP	아이 스틸 빌리브(I still believe)	2022년에 개봉한 미국 영화

<표 3> 원문 정제 내역

<우리말샘>에서 추출한 복합 명사, 고유명사 목록과 매치된 말뭉치 단어 목록은 아래 [그림 4], [그림 5]와 같다. 비교 내역은 한림대와 데이터리가 함께 확인한 후, 대상 말뭉치에 일괄 반영하였다.

우리말샘 어휘	매치된 단어
공개^매수	공개매수
법적^조치	법적조치
유상^증자	유상증자
지분^가치	지분가치
신주^발행	신주발행
청관^변경	청관변경
신주^발행	신주발행
기관^투자	기관투자
기관^투자자	기관투자자
기업^설명회	기업설명회
투자^금용	투자금용
국부^펀드	국부펀드
싱가포르^투자청	싱가포르투자청
기술^도입	기술도입
서빙^로봇	서빙로봇
인공^지능	인공지능
자율^주행	자율주행
자율^주행^로봇	자율주행로봇

우리말샘 어휘	매치된 단어
진품달 싸움	진품달싸움
중 이	중이
다음 달	다음달
현지 시각	현지시각
국풍 요리	국풍 요리
고층 건물	고층 건물
선두 주자	선두주자
면접 심사	면접심사
심사 위원	심사위원
주거 공간	주거 공간
하락 폭	하락폭
대단지 아파트	대단지 아파트
대단지 아파트	대단지 아파트
아파트 단지	아파트 단지
경제 단체	경제단체
선물 세트	선물세트
팬 사인회	팬사인회
사진 촬영	사진 촬영
현장 체험	현장 체험
친필 사인	친필 사인
사인 볼	사인볼

[그림 4] 우리말샘 붙여쓰기 허용 어휘      [그림 5] 우리말샘 띄어쓰기 포함 어휘

위 정제를 거친 말뭉치를 대상으로 고유명 명사구를 하나의 어절로 통합하면서 줄어든 어절 수 내역 및 통계는 아래 [그림 6]과 같다.

doc_id	sent_id	original_form	form	original_phase	modified_phase	original_wc	form_wc	diff
NPRW2400000	NPRW2400000	심상정 정의당 의원이 10월 지난 21대 총선에서	심상정 정의당 의원이 10월 지난 21대 총선에서					
001.24418	001.24418.2.1	나왔던 '위성정당의 출현을 막기 위한 이른바 '위	나왔던 '위성정당의 출현을 막기 위한 이른바 '위					
		성정당 방지법을 발의했다고 밝혔다.	성정당방지법을 발의했다고 밝혔다.	'위성정당 방지법을	'위성정당방지법을	17	16	1
NPRW2400000	NPRW2400000	정당 투표용지에는 비례대표 국회의원 선거에 후보	정당투표용지에는 비례대표 국회의원 선거에 후보					
001.24418	001.24418.4.2	보자를 추천하지 않은 정당을 포함한 모든 정당의	자를 추천하지 않은 정당을 포함한 모든 정당의					
		기호와 정당명이 표기돼야 한다.	호와 정당명이 표기돼야 한다.	정당 투표용지에는	정당투표용지에는	16	15	1
NZRW2400000	NZRW2400000	1956년 한 해 동안 군에서 2986명이 사망했다는	1956년 한 해 동안 군에서 2986명이 사망했다는					
002.35953	002.35953.2.1	다소 이해하기 어려운 기록에 대한 진상 규명이 이	다소 이해하기 어려운 기록에 대한 진상규명이 이	진상 규명이	진상규명이	15	14	1

[그림 6] 말뭉치 정제 결과 어절 수 변경 내역

어절 수 변경 내역의 original\_wc는 원시 말뭉치의 어절 수를 의미하며, form\_wc는 정제 말뭉치 어절 수, diff는 두 어절 수의 차이이다. 전체 말뭉치 바탕으로 문장 당 평균 어절 수 감소량 통계는 아래 <표 4>와 같다.

구분	전체 문장 수	문장 단위 어절 변경 비율	어절 수 감소량
문어	13,907	0.56	0.26
구어	16,439	0.19	0.17

<표 4> 문장 당 평균 어절 수 감소량 통계

또한 이어지는 어휘의미 정보와 문형 정보 파악이 용이하도록 이후 작업과의 연계성을 높이기 위하여 새로운 단어 형성에 참여하는 접두사와 접미사는 어근에 결합한 형태로 분석을 진행하였는데, 예를 들어 ‘-하다/-되다’ 등 접사가 붙은 동사나 형용사는 ‘자랑하다’, ‘과다하다’ 등 결합된 형태를 하나의 단위로 삼아 분석을 진행하였다.

‘-화, -적, -성’ 등의 접미사와 ‘-하다/-되다/-받다/-당하다/-시키다’ 등의 파생접미사는 어근에 결합하여 새로운 단어를 형성하고 그렇게 만들어진 단어는 접두사와 접미사가 결합된 형태로 <우리말샘> 사전에 등재된다. 접사를 분리하지 않고 어근에 통합된 형태로 분석하면 어휘의미 분석을 위한 기초 자료의 성격을 갖출 수 있다는 장점이 있으며 이어지는 구문 분석과 무형 대용어 복원 과정에서는 서술어와 그 논항의 파악이 무엇보다 중요하므로 접사가 통합된 하나의 어휘로 어절을 분리하는 것이 분석 대상 파악과 분석에 매우 유리하다고 할 수 있다.

그러나 <우리말샘>에 등재되어 있지 않은 경우나 조사의 삽입으로 어근과 접사가 분리된 경우는 <우리말샘>의 등재 단위를 기준으로 하여 어근과 접사를 분리하여 분석하였다.

변경 전	변경 후	<우리말샘> 표제어
자랑_001/NNG+하_044/XSV	자랑하_001/VV	자랑하다_001
과다_001/NNG+하_054/XSA	과다하_001/VA	과다하다_001
문화_001/NNG+적_031/XSN	문화적_001/NNG	문화적_001
반_023/XPN+국가_001/NNG	반국가_001/NNG	반-국가_001
가끔/MAG + 씩/XSN	가끔/MAG + 씩/XSN	가끔_001, -씩_003
토/XR+를/JKO	토/XR+를/JKO	토하다_001
하/VV+았/EP+어/EF+./SF	하/VV+았/EP+어/EF+./SF	

<표 5> 어근-접사 결합 예시



## 2.2. 작업 단위 변환

문서 단위 자료를 문장 단위 작업 포맷으로 변환하여 후속 형태소·어휘의미·구문·의미역 분석 및 무형 대용어 복원 작업에 활용하였다. 이 과정에서 문어 말뭉치의 경우 paragraph를 sentence 단위로 분리 후 각 문장에는 고유한 문장 ID를 부여하였으며 original\_form에는 분리된 형태 그대로 입력하고, form에는 표기 변경 결과를 기입하여 변경 내역을 확인할 수 있도록 하였다. 이를 통해 형태소 분석·어휘의미 분석·구문 분석·의미역 분석·무형 대용어 복원 층위를 공통으로 참조할 수 있도록 통합 ID 체계를 구성하였다.

```
"paragraph": [
  {
    "id": "NWRW2400000001.47174.1",
    "form": "원도심과 흥망성쇠 함께한 34년 된 '동인천역 민자역사' 헐린다"
  },
  (중략)
  {
    "id": "NWRW2400000001.47174.5",
    "form": "지하 2층, 지상 6층 규모인 동인천역 민자역사는 2009년 쇼핑몰 영업을 중단된 뒤 4·5층에 있는 화상경륜장이 영업 중이다. 민자역사는 현재도 임차인 보증금과 건설사들 유치권 대금 등이 정리되지 않아 각종 소송 등이 진행되고 있다."
  },

```

[그림 7] 국립국어원 신문 말뭉치 2024 json 구조

```

"sentence": [
  {
    "id": "NWRW2400000001.47174.5.1",
    "form": "지하 2층, 지상 6층 규모인 동인천역 민자역사는 2009년 쇼핑물 영업이 중단된 뒤 4·5층에 있는 화상 경륜장이 영업 중이다.",
    "original_form": "지하 2층, 지상 6층 규모인 동인천역 민자역사는 2009년 쇼핑물 영업이 중단된 뒤 4·5층에 있는 화상경륜장이 영업 중이다.",
    "word": [
    ],
  },
  {
    "id": "NWRW2400000001.47174.5.2",
    "form": "민자역사는 현재도 임차인 보증금과 건설사들 유치권 대금 등이 정리되지 않아 각종 소송 등이 진행되고 있다.",
    "original_form": "민자역사는 현재도 임차인 보증금과 건설사들 유치권 대금 등이 정리되지 않아 각종 소송 등이 진행되고 있다.",
    "word": [
    ]
  }
]

```

[그림 8] 문어 작업 말뭉치 json 구조

구어 말뭉치는 발화 단위로 분리된 원시 말뭉치와의 대응 가능성을 유지하기 위해 원시 말뭉치의 utterance > form을 구축 말뭉치의 sentence > original\_form으로 두고, 정제된 말뭉치는 form으로 기재하였으며 원문 수정 내역은 별도의 오류 리포트로 관리하였다.

```

"utterance": [
  {
    "id": "SCRW2400000011.1.1.1",
    "form": "어렸을 적부터 아빠가 강조해서 저와 오빠에게 어 했던 말씀 중의 하나가",
    "original_form": "어렸을 적부터 아빠가 강조해서 저와 오빠에게 어~ 했던 말씀 중에 하나가",
    "speaker_id": "SD2401220",
    "start": 0.66,
    "end": 9.13142,
    "note": ""
  }
]

```

[그림 9] 국립국어원 일상 대화 말뭉치 2024 json 구조

```

"sentence": [
  {
    "id": "SCRW2400000011.1.1.1",
    "form": "어렸을 적부터 아빠가 강조해서 저와 오빠에게 했던 말씀 중의 하나가 ",
    "original_form": "어렸을 적부터 아빠가 강조해서 저와 오빠에게 어 했던 말씀 중의 하나가",
    "speaker_id": "SD2401220",
    "word": [

```

[그림 10] 구어 작업 말뭉치 json 구조

위 순서로 정제한 말뭉치는 데이터리에서 개발한 한국어 통합 분석 도구에 업로드되어 형태소, 어휘의미, 구문, 의미역 분석 총 4개의 한국어 구조 말뭉치 및 무형 대용어 복원 말뭉치 구축에 사용되었다. 이후 구축 과정에서 원문 정제의 추가 필요성이 반복적으로 제기됨에 따라, 통합 분석 도구에는 원문 수정 기능을 새롭게 추가하였다. 이를 통해 작업자는 형태소 분석 작업을 진행하며 테이블 형식으로 된 분석값 입력 도구에서 행 추가/삭제 기능을 활용하여 원문을 추가로 정제하였다. 도구를 통해 정제된 말뭉치는 이후 검수자가 2차로 확인하였고, 최종 산출물은 원시 말뭉치 및 정제본과의 비교 검증을 거쳐 수정 내용의 적절성을 판단하였다. 이러한 절차를 통해 원문-정제본-최종본 간의 체계적인 정합성이 확보되었으며, 분석 말뭉치 구축의 신뢰성과 정확성을 높일 수 있었다.

## 2.3. 매핑 테이블

원시 말뭉치와 작업 말뭉치의 참조 관계를 확인할 수 있는 매핑 테이블을 생성하였다. 매핑 테이블의 구조는 작업 대상 문장이 포함되어 있는 파일의 id와 작업 대상 문장의 id를 매핑한 GEN\_Info\_sheet, 원시 말뭉치 문장의 전체 혹은 부분이 사용된 경우를 나타내는 REF\_Info\_sheet, 문장의 일부분이 작업 말뭉치의 문장 id를 부여받은 경우를 보여주는 'SUB\_Info\_sheet'로 구성된다.

문어 말뭉치의 경우 하나의 paragraph 내에 포함된 문장들을 sentence 단위로 분리하였으므로, GEN/REF/SUB\_Info\_sheet가 모두 생성되었으며, 구어 말뭉치의 경우 원시 말뭉치의 sentence 형식 그대로 사용되었으므로 REF\_Info\_sheet까지 생성되었다. 매핑 테이블 형식 예시는 아래 [그림 11]과 같다.

index	subindex	form_id	origincorpus	referencedSourceId	begin	end
1	1	NIRW2400000001.1165.2.1	신문 발원지 2024	NIRW2400000001.1165.2	0	59
1	1	NIRW2400000001.1165.2.2	신문 발원지 2024	NIRW2400000001.1165.2	60	111
1	1	NIRW2400000001.1165.3.1	신문 발원지 2024	NIRW2400000001.1165.3	0	40
1	1	NIRW2400000001.1165.3.2	신문 발원지 2024	NIRW2400000001.1165.3	41	77
1	1	NIRW2400000001.1165.3.3	신문 발원지 2024	NIRW2400000001.1165.3	78	109
1	1	NIRW2400000001.1165.4.1	신문 발원지 2024	NIRW2400000001.1165.4	0	75
1	1	NIRW2400000001.1165.4.2	신문 발원지 2024	NIRW2400000001.1165.4	76	112
1	1	NIRW2400000001.1165.5.1	신문 발원지 2024	NIRW2400000001.1165.5	0	51
1	1	NIRW2400000001.1165.5.2	신문 발원지 2024	NIRW2400000001.1165.5	52	122
1	1	NIRW2400000001.1165.6.1	신문 발원지 2024	NIRW2400000001.1165.6	0	57
1	1	NIRW2400000001.1165.6.2	신문 발원지 2024	NIRW2400000001.1165.6	58	121
1	1	NIRW2400000001.1165.7.1	신문 발원지 2024	NIRW2400000001.1165.7	0	47
1	1	NIRW2400000001.1165.7.2	신문 발원지 2024	NIRW2400000001.1165.7	48	131
1	1	NIRW2400000001.1165.9.1	신문 발원지 2024	NIRW2400000001.1165.9	0	61
1	1	NIRW2400000001.1165.9.2	신문 발원지 2024	NIRW2400000001.1165.9	62	123
1	1	NIRW2400000001.1165.11.1	신문 발원지 2024	NIRW2400000001.1165.11	0	34
1	1	NIRW2400000001.1165.11.2	신문 발원지 2024	NIRW2400000001.1165.11	35	127
1	1	NIRW2400000001.1165.12.1	신문 발원지 2024	NIRW2400000001.1165.12	0	44
1	1	NIRW2400000001.1165.12.2	신문 발원지 2024	NIRW2400000001.1165.12	45	102
1	1	NIRW2400000001.1165.12.3	신문 발원지 2024	NIRW2400000001.1165.12	103	131
1	1	NIRW2400000001.1165.13.1	신문 발원지 2024	NIRW2400000001.1165.13	0	19
1	1	NIRW2400000001.1165.13.2	신문 발원지 2024	NIRW2400000001.1165.13	20	87
1	1	NIRW2400000001.1165.13.3	신문 발원지 2024	NIRW2400000001.1165.13	88	116
1	1	NIRW2400000001.1165.14.1	신문 발원지 2024	NIRW2400000001.1165.14	0	32

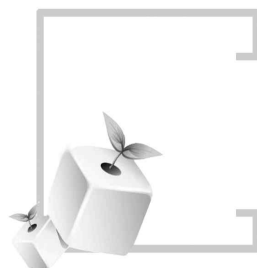
[그림 11] 매핑테이블 SUB\_Info\_sheet

index	file_id	form_id	NumOfRefer
1	NIRW2400000001	NIRW2400000001.1165.7.1	1
1	NIRW2400000001	NIRW2400000001.11075.6.1	1
1	NIRW2400000001	NIRW2400000001.11075.5.1	1
1	NIRW2400000001	NIRW2400000001.11075.4.1	1
1	NIRW2400000001	NIRW2400000001.11075.3.1	1
1	NIRW2400000001	NIRW2400000001.11075.2.1	1
1	NIRW2400000001	NIRW2400000001.11075.1.1	1
1	NIRW2400000001	NIRW2400000001.1165.13.1	1
1	NIRW2400000001	NIRW2400000001.1165.13.2	1
1	NIRW2400000001	NIRW2400000001.1165.13.3	1
1	NIRW2400000001	NIRW2400000001.14315.9.1	1
1	NIRW2400000001	NIRW2400000001.14315.8.1	1
1	NIRW2400000001	NIRW2400000001.14315.7.1	1
1	NIRW2400000001	NIRW2400000001.14315.6.1	1
1	NIRW2400000001	NIRW2400000001.11075.6.2	1
1	NIRW2400000001	NIRW2400000001.1165.14.1	1
1	NIRW2400000001	NIRW2400000001.1165.14.2	1
1	NIRW2400000001	NIRW2400000001.1165.15.1	1
1	NIRW2400000001	NIRW2400000001.1165.15.2	1
1	NIRW2400000001	NIRW2400000001.1165.16.1	1
1	NIRW2400000001	NIRW2400000001.1165.16.2	1
1	NIRW2400000001	NIRW2400000001.1165.17.1	1
1	NIRW2400000001	NIRW2400000001.10555.1.1	1
1	NIRW2400000001	NIRW2400000001.10555.2.1	1

index	subindex	form_id	origincorpus	referencedSourceId	fullOrPartial
1	1	NIRW2400000001.1165.3.1	신문 발원지 2024	NIRW2400000001.1165.1	F
1	1	NIRW2400000001.1165.2.1	신문 발원지 2024	NIRW2400000001.1165.2	P
1	1	NIRW2400000001.1165.2.2	신문 발원지 2024	NIRW2400000001.1165.2	P
1	1	NIRW2400000001.1165.3.1	신문 발원지 2024	NIRW2400000001.1165.3	P
1	1	NIRW2400000001.1165.3.2	신문 발원지 2024	NIRW2400000001.1165.3	P
1	1	NIRW2400000001.1165.3.3	신문 발원지 2024	NIRW2400000001.1165.3	P
1	1	NIRW2400000001.1165.4.1	신문 발원지 2024	NIRW2400000001.1165.4	P
1	1	NIRW2400000001.1165.4.2	신문 발원지 2024	NIRW2400000001.1165.4	P
1	1	NIRW2400000001.1165.5.1	신문 발원지 2024	NIRW2400000001.1165.5	P
1	1	NIRW2400000001.1165.5.2	신문 발원지 2024	NIRW2400000001.1165.5	P
1	1	NIRW2400000001.1165.6.1	신문 발원지 2024	NIRW2400000001.1165.6	P
1	1	NIRW2400000001.1165.6.2	신문 발원지 2024	NIRW2400000001.1165.6	P
1	1	NIRW2400000001.1165.7.1	신문 발원지 2024	NIRW2400000001.1165.7	P
1	1	NIRW2400000001.1165.7.2	신문 발원지 2024	NIRW2400000001.1165.7	P
1	1	NIRW2400000001.1165.8.1	신문 발원지 2024	NIRW2400000001.1165.8	F
1	1	NIRW2400000001.1165.8.2	신문 발원지 2024	NIRW2400000001.1165.8	P
1	1	NIRW2400000001.1165.9.1	신문 발원지 2024	NIRW2400000001.1165.9	P
1	1	NIRW2400000001.1165.10.1	신문 발원지 2024	NIRW2400000001.1165.10	F
1	1	NIRW2400000001.1165.11.1	신문 발원지 2024	NIRW2400000001.1165.11	P
1	1	NIRW2400000001.1165.11.2	신문 발원지 2024	NIRW2400000001.1165.11	P
1	1	NIRW2400000001.1165.12.1	신문 발원지 2024	NIRW2400000001.1165.12	P
1	1	NIRW2400000001.1165.12.2	신문 발원지 2024	NIRW2400000001.1165.12	P
1	1	NIRW2400000001.1165.12.3	신문 발원지 2024	NIRW2400000001.1165.12	P
1	1	NIRW2400000001.1165.13.1	신문 발원지 2024	NIRW2400000001.1165.13	P

[그림 12] 매핑테이블  
GEN\_Info\_sheet

[그림 13] 매핑테이블  
REF\_Info\_sheet



## 제 3 장

# 형태 분석 말뭉치 구축



# 1. 형태 분석 말뭉치 구축 지침 개선

## 1.1. 기존 지침 보완 사항

기존 형태 분석 말뭉치 구축 지침은 ‘2019년 형태 분석 말뭉치 구축’ 사업의 결과 보고서를 통해 마련되었다. 이 지침은 이후 ‘2020년 어휘의미 말뭉치 연구 분석 사업’을 통해 메신저 대화 부분이 추가되는 방식으로 일부 보완이 이루어졌다.

본 사업에서는 위의 두 지침을 검토하며 본 사업의 목적인 다층위 구조 분석 말뭉치 구축에 필요한 방식으로의 수정을 진행하였다. 또한 말뭉치 정제가 가능해짐에 따라 각종 오류 처리 부분을 지침에서 제외할 수 있었으며 본 사업이 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’를 대상으로 진행되는 만큼 메신저 대화 관련 각종 처리 지침 등을 삭제하여 보다 간명하고 정교한 말뭉치 구축이 이루어질 수 있는 기반을 마련하였다.

우선 말뭉치 정제 단계를 통해 불필요해진 ‘분석 불능 범주’를 대폭 축소할 수 있었다. 기존 지침에서는 분석 불능 범주를 ‘분석 불능 범주(NA)’와 ‘명사 추정 범주(NF)’, ‘용언 추정 범주(NV)’로 세분하는 방식을 통해 분석 불능 범주의 분석 가능성을 판단하기 위한 노력을 다하였는데, 본 사업에서는 형태 분석 이전에 말뭉치 정제 단계를 거침으로써 분석 불능 범주를 최소화할 수 있었으며 그 결과 세부 분류 없이 ‘분석 불능 범주(NA)’만으로도 말뭉치 분석 과정에서 발생하는 오류들을 처리할 수 있었다.

또한 메신저 대화에 포함된 개인정보를 처리하기 위한 ‘개인 정보 처리 필요 요소(NAP)’는 원시 말뭉치에서 삭제되지 않고 남아 있는 일부 개인정보의 처리를 위한 표지로, 대상 말뭉치를 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’로 한정하고 원문 정제까지 이루어진 상황에서는 불필요한 형태 분석 표지로 판단하여 삭제하였다.

지침의 보완도 동시에 진행되었다. 형태 분석과 구문 분석 정교화를 위한 1차 자문회의(채숙희 인천대 교수 자문)에서 제안된 인용조사의 처리에 대한 내용을 국립국어원과 지침 관련 논의를 통해 지침에 반영하였는데, 인용격조사 목록에 ‘(이)라’를 추가함으로써 인용조사가 통합된 형태의 연결어미 목록을 상당수 줄여 지침을 간소화하였으며 형태 분석 과정에서 이루어졌던 긍정지정사(VCP) ‘이’의 복원 역시 상당 부분 감소시켜 보다 정교한 형태 분석이 이루어질 수 있었다.

형태 분석 말뭉치 구축 지침은 ‘2019년 형태 분석 말뭉치 구축’ 사업의 결과 보고서를 바탕으로 수정이 필요한 사항을 파악하여 지침 개선 작업을 진행하였다.

<우리말샘>의 등재어는 단어부터 문장까지 다양한데, <우리말샘> 등재어의 내부 구조를 분석하게 되면 <우리말샘>을 기반으로 한 어휘의미 분석이 올바르게 이루어지지 않게 되며 불필요한 의존 관계 분석 과정이 이루어지게 된다. 고유명을 하나의 단위로 통합하여 분석하는 것은 향후 개체명 분석 및 개체 연결 말뭉치 구축 시에도 개체명 전체를 하나의 단위로 처리하게 되어 유리하게 작용할 수 있다. 다만 구나 절, 문장으로 이루어진 등재어나 로마자 등재어의 경우에는 등재어 내부 띄어쓰기 공백이 사라지게 되면 가시성이 떨어지고 어문 규정에 맞지 않는 구조 분석이 이루어지게 된다. 따라서 이와 같은 경우에는 원문 정제 단계에서 언더바 기호 ‘\_’로 띄어쓰기 공백을 대체하여 분석 대상 어절 내부 구조를 투명하게 하고 어문 규정과 일치하는 분석 단위를 유지할 수 있게 하였다.

변경 전	변경 후	<우리말샘> 표제어
템플/NNG, 스테이/NNG	템플스테이/NNG	템플 ^ 스테이 ( ▼ temple stay)_001
블루/NNP, 이즈/NNP, 더/NNP, 워미스트/NNP, 컬러/NNP,	블루_이즈_더_워미스트_컬러 /NNP	-
바람/NNP+과/JKB, 함께/MAG, 사라지/VV+다/EF	바람과_함께_사라지다/NNP	바람과 함께 사라지다_001
IMF/SL, reserve/SL, position/SL,	IMFreserveposition_NNP	아이엠에프^리저브^포지션(IMF reserve position)_001
로버트다우니주니어/NNP 로버트/NNP, 다우니/NNP, 주니어/NNP	로버트_다우니_주니어/NNP	-

<표 6> 단어절 구성의 형태 분석

한글 표기뿐만 아니라 숫자, 로마자, 한자도 한국어를 표기하기 위한 일부로 쓰였거나 <우리말샘>에 한글 형태로 등재된 경우 형태 분석 및 어휘의미 분석의 대상으로 삼아 현실에서 나타나는 표기의 다양성을 포착하였다. 이들을 SN(숫자)이나 SL(외국 문자), SH(한자)로 처리해서는 한국어의 다양한 표기 양상을 담아낼 수 없으며 이어지는 어휘의미 분석과의 연관성도 확보하기 어렵기 때문이다.

변경 전	변경 후	<우리말샘> 표제어
3.1절/NNG	3.1절/NNG	삼일-절(三一絶)
6/SN+./SP+25/SN	6.25/NNG	육-이오(六二五)
1/SN+월/NNB	1월/NNG	일월(一月)
LG전자/NNP, LG/SL+에서/JKB	LG전자/NNP, LG/NNP+에서/JKB	-
F/SL+./SP+I/SL+./SP+F/SL+./SP+A/SL	F.I.F.A/NNP	피파(FIFA) 에프아이에프에이(FIFA)

<표 7> 숫자, 로마자, 한자의 형태 분석

하나의 어절 안에 기호가 삽입되어 있는 경우에는 <우리말샘> 등재 여부를 고려하여 분석을 달리하였다.

먼저 ‘마이크로소프트(microsoft)’와 같이 한글 고유명에 이어 괄호 안에 원어가 표기된 경우에는 한글 번역된 이름과 원어 이름이 모두 특정 회사를 나타내는 고유명이므로 어절 내 삽입된 괄호를 분리하여 기호(SS)로 분석하고 한글명과 원어명을 각각 고유명사(NNP)로 분석하였다.

그러나 ‘대중(對中) 정책’, ‘대중(對中) 무역’ 등에 쓰이는 <우리말샘> 등재어 ‘대중(對中)\_006’이 ‘대(對)중’과 같이 쓰인 경우 삽입된 괄호를 분리하여 기호로 분석하면 접두사 ‘대(對)’를 접두사와 한자로 분석하게 되어 어휘의미 분석과 구문 분석 결과의 연속성을 파악하기 어려워진다. 따라서 이러한 경우는 삽입된 괄호를 따로 분리하여 분석하지 않고 기호가 포함된 전체 어절을 하나의 어절로 분석하였다.

변경 전	변경 후	<우리말샘> 표제어
마이크로소프트/NNP+ (/SS+microsoft/SL+)/SS	마이크로소프트/NNP+ (/SS+microsoft/NNP+)_777/SS	-
대 / N A + ( / S S + 對 / S H + ) / S S + 중 _014/NNP]	대(對)중_006/NNP	대중(對中)_006

<표 8> 고유명의 형태 분석



## 1.2. 형태 분석 말뭉치 구축 지침

형태 분석 말뭉치 구축 지침의 목차는 아래와 같다. 지침 세부 내용은 [붙임1] 참조.

### 형태 분석 말뭉치 구축 지침

#### 목차

1. 총론
  - 1.1. 지침의 목적
  - 1.2. 용어 정의
  - 1.3. 기본 원칙
2. 형태 분석의 원리
  - 2.1. 형태 분리의 원칙
  - 2.2. 형태 분석 표지 부착
3. 형태 분석의 실제
  - 3.1. 체언
    - 3.1.1. 명사
      - 3.1.1.1. 일반명사(NNG)
      - 3.1.1.2. 고유명사(NNP)
      - 3.1.1.3. 의존명사(NNB)
    - 3.1.2. 대명사(NP)
    - 3.1.3. 수사(NR)
  - 3.2. 용언
    - 3.2.1. 동사(VV)
    - 3.2.2. 형용사(VA)
    - 3.2.3. 지정사
      - 3.2.3.1. 긍정지정사(VCP)
      - 3.2.3.2. 부정지정사(VCN)

3.2.4. 보조용언(VX)

3.3. 수식언

3.3.1. 관형사

3.3.1.1. 성상 관형사(MMA)

3.3.1.2. 지시 관형사(MMD)

3.3.1.3. 수 관형사(MMN)

3.3.2. 부사

3.3.2.1. 일반부사(MAG)

3.3.2.2. 접속부사(MAJ)

3.4. 독립언

3.4.1. 감탄사(IC)

3.5. 관계언

3.5.1. 격조사(JK)

3.5.1.1. 주격조사(JKS)

3.5.1.2. 보격조사(JKC)

3.5.1.3. 목적격조사(JKO)

3.5.1.4. 관형격조사(JKG)

3.5.1.5. 부사격조사(JKB)

3.5.1.6. 호격조사(JKV)

3.5.1.7. 인용격조사(JKQ)

3.5.2. 보조사(JX)

3.5.3. 접속조사(JC)

3.6. 의존형태

3.6.1. 어미

3.6.1.1. 선어말어미(EP)

3.6.1.2. 종결어미(EF)

3.6.1.3. 연결어미(EC)

3.6.1.4. 명사형전성어미(ETN)

3.6.1.5. 관형사형전성어미(ETM)

3.6.2. 접사

3.6.2.1. 체언접두사(XPN)

3.6.2.2. 명사파생접미사(XSN)

3.6.2.3. 동사파생접미사(XSV)

3.6.2.4. 형용사파생접미사(XSA)

3.6.3. 어근(XR)

3.7. 기호

3.7.1. 일반 기호

3.7.1.1. 마침표, 물음표, 느낌표(SF)

3.7.1.2. 쉼표, 가운뎃점, 콜론, 빗금(SP)

3.7.1.3. 따옴표, 괄호, 줄표(SS)

3.7.1.4. 줄임표(SE)

3.7.1.5. 붙임표(물결)(SO)

3.7.1.6. 기타 기호(SW)

3.7.2. 외국 문자(SL)

3.7.3. 한자(SH)

3.7.4. 숫자(SN)

3.8. 기타

3.8.1. 분석 불능 범주(NA)

4. 구어 분석에 따른 주의 사항

4.1. 구어에서 나타나는 준말과 형태 변이 현상의 처리

4.2. 구어 전사 시 이용된 마크업과 표지의 처리

4.3. 전사 오류 및 해석 불능 어절의 처리

## 2. 형태 분석 말뭉치 구축 절차

### 2.1. 형태 분석 초벌 작업

본 사업의 말뭉치 구축은 자체 개발한 한국어 구조 분석 도구를 활용하여 기계 분석을 선행하고, 이후 인간 전문가의 검토 및 수정 과정을 거치는 2단계 방식으로 수행하였다. 형태소 분석 말뭉치는 데이터리에서 보유 중인 딥러닝 기반 통합 분석 도구, espresso5를 활용하여 초벌 작업을 진행하였다.

#### 2.1.1. 형태소 분석 모델 및 입력 방식

형태소 분석 초벌 말뭉치는 자체 개발 형태소 분석기를 활용하여 일괄적으로 생성하였다. 입력 문장은 원문 정제 단계에서 부여된 문장 ID와 어절 단위를 그대로 유지한 상태에서 분석기에 전달되며, 분석기는 어절을 형태소 열로 분해하고 품사 태그 세트에 따라 표지화된 결과를 반환하도록 설계되었다. 초벌 분석 코드는 문단 단위가 아니라 말뭉치 파일 단위로 처리할 수 있도록 배치(batch) 형태로 구현하였으며, 분석 결과는 최종 납품 형식인 JSON v1.6 스키마의 morph 필드를 자동으로 채우는 형태로 저장하였다. 이후 작업자는 구축 도구 상에서 형태소 분석 결과를 확인하고, 오류가 있는 셀만을 수정함으로써 전체 작업 시간을 단축하고 분석 품질을 높였다.

#### 2.1.2. 말뭉치 초벌 결과

위 모델을 활용한 초벌 말뭉치의 출력값 예시는 아래와 같다.

```
>>> from nltkor.tagimport EspressoTagger
>>> tagger = EspressoTagger(task='pos')
Downloading Espresso5 model...
>>> sent="3월 출시를 앞둔 넥슨의 던전애파이터모바일이 새로운 신화에 도전한다."
>>> print(tagger.tag(sent))
('3', 'SN'), ('월', 'NN'), (' ', 'SP'), ('출시', 'NN'), ('를', 'JJ'), (' ', 'SP'), ('앞두/VB', 'VB'), ('ㄴ', 'EE'), (' ', 'SP'), ('넥슨', 'NN'), ('의', 'JJ'), (' ', 'SP'), ('던전애파이터', 'NN'), (' ', 'SP'), ('모바일', 'NN'), ('"', 'SY'), ('이', 'JJ'), (' ', 'SP'), ('새롭', 'VB'), ('ㄴ', 'EE'), (' ', 'SP'), ('신화', 'NN'), ('에', 'JJ'), (' ', 'SP'), ('도전', 'NN'), ('하/XV', 'XV'), ('ㄴ다', 'EE'), ('.', 'SY')
```

[그림 14] 형태 분석 말뭉치 초벌 출력값 예시

Espresso5의 기본 태깅 체계 및 출력 형태가 2025년 최신 형태분석 지침(버전 1.2)의

표기 규칙과 일부 상이하여, 지침에 부합하는 최종 형태소 초별 말뭉치를 구축하기 위해 추가적인 정제·보정 절차가 필요하였다. 특히 Espresso5가 내부 처리 과정에서 사용하는 특수 표지(BOS, EOS 등)와 비표준 태그가 분석 결과에 직접 포함되는 문제가 확인되어, 도구 수정 및 후처리 과정을 통해 이를 제거하고 지침에 맞춘 형태로 결과를 일원화하였다. 전체 절차는 다음과 같이 요약할 수 있다.

#### (1) 표지(tagset) 체계의 지침 기반 재정렬

Espresso5는 기본적으로 세종계획의 품사를 축약한 형태로 입력 방식을 지정해 두었기에 금년도에 보완된 형태분석 지침의 표지로의 변경을 진행하였으며 변경된 내역은 아래와 같다.

- Espresso5 출력 품사를 지침 표지로 매핑하는 변환 사전 구축  
예: NN → NNG/NNP/NNB 구분, VB → VV·VA·VCP 등 재분류
- 조사(JKS/JKO/JKB 등), 어미(EC/EF/ETM/ETN), 접사 계열(XS\*, X\*) 등은 지침 규정에 맞게 세분화
- 비표준 표지 또는 복합 표지가 포함된 경우 지침 단위로 분해하여 단일 형태소 구조로 재구성

#### (2) BOS, EOS 등 불필요 특수 표지 제거 및 오류 결과 정리

Espresso5는 음절 기반 태깅 과정에서 문장 시작과 끝을 표시하기 위해 내부적으로 BOS(beginning of sentence), EOS(end of sentence) 표지를 생성한다. 기본 설정에서는 이 표지가 최종 출력에도 포함되는 문제가 확인되었다. 이에 이미 생성된 초별 형태소 분석 결과에서 BOS/EOS가 포함된 라인 자동 탐지하고, 문장 내부에 잘못 삽입된 경우 후처리 스크립트로 표지를 정정하였다.

또한, 간헐적으로 발견된 다음과 같은 오류 유형도 함께 정리하였다. 음절 단위의 잘못된 분절로 인한 외래어·숫자·기호 인식 오류와 말뭉치 정제 과정에서 임의 수정한 복합 명사 분해 오류(예: ‘불법촬영’ → ‘불법\_촬영’ 형태로 정제 후 적용) 및 인용부호, 괄호 등 특수문자와 결합된 형태소 분석 오류 등의 경우 목록화하여 실제 작업자에게 공유하고 작업자들은 이를 참고하여 초별 말뭉치 수정, 추가 부착 등의 작업을 진행하였다.

작업 중 발견된 유사 오류들은 아래와 같이 정리하여 형태소 작업 중간 기계적인 수정을

진행했으며 작업 말뭉치 동기화를 위해 작업 완료된 주 차별 작업물 수정과 형태소 분석 모델 내부 코드를 동시에 변경하였다. 이를 통해 구축 말뭉치와 형태소 분석 모델 성능을 모두 향상시켰다.

오류 예시	수정 방향
NNB:만 + XSA:하	‘-듯, -만, -법’ 등 의존명사 뒤에 ‘하-’가 결합한 것은 지침에 따라 분리
서술성 명사가 아닌 지시적 명사 뒤에 ‘하-’가 결합	-접속조사 ‘하고’로 변경
XSN:화 + XSV:되	‘-적, -화, -성, -시’ 등의 명사파생접미사, 명사의 일부로 분석 →명사파생접미사를 먼저 병합한 이후 동사파생접미사나 형용사파생접미사를 병합
XSN:들 + XSV:하	-접미사 ‘들’은 병합 대상이 아니며 이 역시 접속조사 ‘하고’로 변경
SS:’ + XSA:답	-기호 뒤 동사 파생 접미사나 형용사 파생 접미사는 미병합
원문 단어의 변형 및 분리 예)NNG:전략핵잠수 + XSV:하 NNG:심은 + XSV:하 미세먼, 건강검 등 -> ‘지, 진 누락	-오분석, 고유명 분리, 일부 음절 누락 모두 형태소 분석 모델 수정 후 재검토

<표 9> 형태소 분석 작업 중 발견된 초벌 오류

## 2.2. 형태 분석 작업 중 확인된 문제점 및 개선 방향

본 사업에서는 형태 분석 단계에서 고유명과 구로 이루어진 <우리말샘> 등재어를 통합된 한 어절로 처리하여 불필요한 분석 과정의 에너지 소모를 줄이고자 하였다. 이는 본 사업의 목적이 각 단계의 분석 말뭉치 구축과 동시에 각 분석 말뭉치들을 하나의 파일로 통합하는 다층위 구조 분석 말뭉치 구축에 있음을 고려한 것이었다.

그러나 이 과정에서 <우리말샘>의 등재어와 미등재어의 구분, 각종 대회 및 회의의 이름과 각종 제목과 주제, 각종 법률명과 개정안에 대해서 어느 정도까지 한 단위로 처리할 수 있는가 하는 문제에 직면하게 되었다.

국립국어원에서 구축하고 있는 <우리말샘>은 사용자들의 제안과 전문가 감수를 통해 단어의 등재가 이루어지는 동적 사전이므로 특정 시점을 기준으로 <우리말샘> 등재 여부에 따라 형태 분리를 진행하면 일관된 기준 적용이 어려워지는 것은 당연하다. 예를 들어 사업 진행 시점인 2025년 상반기 기준 <우리말샘>에는 ‘서울올림픽’과 ‘서울올림픽대회’는 등재되어 있으나 ‘베이징올림픽’은 2008년에 개최된 ‘베이징올림픽’만이 등재어로 올라 있고 ‘베이징올림픽대회’는 등재어로 올라 있지 않으며, ‘파리올림픽’은 1900년과 1924년

에 개최된 대회만이 등재어의 의미로 올라 있어 2024년에 개최된 ‘파리올림픽’은 등재어가 아니며 2026년으로 예정된 ‘2026 FIFA 북중미 월드컵’에 대해서는 <우리말샘>에서 확인할 수 없다. 따라서 올림픽과 같이 연속되어 나타나는 ‘대회’를 올림픽과 하나의 어절로 다룰 수 있는지, 개최 예정인 대회에 대해서도 하나의 어절로 취급할 수 있는지, 올림픽이나 월드컵보다 작은 단위의 대회명이나 회의명 역시 하나의 고유명으로 처리할 수 있는지 판단이 불명확해지는 문제가 발생하였다.

이러한 문제는 국가에서 정하고 있는 법률명에서도 확인할 수 있는데 속칭 김영란법으로 불리는 ‘부정청탁및금품등수수금지법’은 <우리말샘>에 등재되어 있지만 ‘중대재해처벌등에관한법률’은 등재되어 있지 않다.

법률명에는 띄어쓰기의 문제도 포함되어 있는데 현재 기준으로 법률명은 띄어쓰기를 하지 않고 하나의 어절로 붙여 쓰는 것을 원칙으로 삼고 있지만 가독성 관련 불편함이 꾸준히 제기되고 있어 언론 매체에서도 최근 들어 언론사에 따라 법률명을 띄어 쓰려는 시도들도 확인되고 있다.

본 사업에서는 다층위 구조 분석 말뭉치 구축이라는 목적을 위해 단어의 내부 구조 분석을 간소화하여 각종 대회 및 회의의 이름과 제목, 주제, 각종 법률명과 개정안에 대해서는 <우리말샘> 등재 여부를 어휘의미로 반영하되 각각의 이름을 하나의 어절로 분석하였으나 차후에 이루어지는 말뭉치 구축 과정에서는 이러한 동적 사전의 활용 방안과 지향에 대한 고민이 반드시 이루어져야 할 것이라 판단된다.



## 제 4 장

# 어휘의미 분석 말뭉치 구축





# 1. 어휘의미 분석 말뭉치 구축 지침 개선

## 1.1. 기존 지침 보완 사항

기존 어휘의미 분석 말뭉치 구축 지침은 ‘2020년 어휘의미 말뭉치 연구 분석’ 사업의 결과 보고서를 통해 마련되었다. 이 지침은 ‘2019년 어휘의미 분석 말뭉치 구축’ 사업을 통해 마련된 체언류를 대상으로 한 ‘어휘의미 분석 말뭉치 구축 지침’에 용언류의 어휘의미 분석 지침과 메신저 대화 관련 분석 지침을 보완하여 작성되었다.

본 사업에서는 위의 두 지침을 검토하며 본 사업의 목적인 다층위 구조 분석 말뭉치 구축에 필요한 방식으로의 수정을 진행하였는데, 이전 단계인 형태 분석 구축 시에 어근과 접사가 통합된 형태로의 형태 분석을 지향한 결과 어휘의미 분석 대상에서 어근(XR)의 비중이 대폭 줄어들게 되었다.

<우리말샘>에 분석 대상 의미가 없거나 분석 대상 형태가 등재되어 있지 않은 경우 어휘의미 번호를 각각 ‘888’과 ‘777’로 분석하는 방식은 그대로 유지되었으나 본 사업에서는 원문의 사전 정제 단계를 거친 후 형태 분석을 진행한 관계로 기존 지침에서 ‘말뭉치 원어절의 오타나 탈자 등 오류로 인해 의미 분석이 불가능할 경우에 부여하던 어휘의미 ‘999’의 사용은 크게 제한되었다.

또한 대상 말뭉치가 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’로 한정됨에 따라 ‘메신저 대화’를 위한 분석 세부 지침과 음절을 첨가하여 장음을 표현한 경우와 같이 비표준어를 위한 분석 세부 지침 등을 삭제할 수 있어 지침을 간소화하여 작업자들의 부담을 줄이고 지침 이해도를 높여 분석 말뭉치의 정확성 향상에 기여할 수 있었다.

어휘의미 분석 말뭉치는 ‘2020년 어휘의미 말뭉치 연구 분석’ 사업의 결과 보고서를 참고하여 수정이 필요한 사항을 파악한 후 지침 개선 작업을 진행하였다.

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업에서는 원문의 수정이 가능해짐에 따라 <우리말샘>의 등재 여부를 최우선으로 고려하여 <우리말샘>에 고유명으로 등재되어 있는 대상에 대해서는 고유명으로 분석하여 어휘의미를 부여하였다.

<우리말샘>이 등재 어휘를 의미(sense) 단위로 구분하여 정보를 제공하고 있는 점은 동음이의어와 다의어의 구분을 어렵게 하는 측면이 있다. 따라서 어휘의미 분석을 정교하게 하기 위해서는 <우리말샘>에 기술되어 있는 사전적 정의 및 예문과 어휘의미 분석 말뭉치를 이용하여 기계학습을 진행하고 연어 관계 확인을 통한 인간 작업자의 어휘의미

분석이 이루어져야 할 필요가 있다.

원문 정제 및 형태 분석 말뭉치 구축 방식 변경에 따라 한자, 로마자, 숫자 표기 형태에 대해서도 대응되는 <우리말샘> 등재어의 어휘의미 번호를 부여하는 것이 가능해졌으며 형태 분석 표기를 통해 해당하는 어휘의미를 대응시키는 것이 가능해졌다. 이러한 작업을 통해 ‘아샷추’, ‘리셀’, ‘웨이트하다’, ‘마이너하다’ 등 어휘 미등재(777)와 ‘키트’(밀키트), ‘클라이밍’(벽을 오르는 운동) 등 의미 미등재(888) 사례를 다수 확보할 수 있었는데 이는 차후 <우리말샘> 사전의 등재어 추가와 의미 보완에 활용할 수 있다.

반면 아래 예와 같이 기존 지침에서 ‘템플\_\_888/NNG, 스테이\_\_888/NNG’와 같이 <우리말샘> 등재어의 내부 구조를 분석하여 777과 888이 과도하게 발생되던 문제는 <우리말샘> 등재어의 형태와 동일하게 한 단위로 다루어 어휘의미를 001로 부여하는 방식을 통해 일부 해소할 수 있었다.

또한 어휘의미 분석의 기반이 되는 형태 분석 말뭉치 구축 단계에서부터 어휘의미 분석을 고려하여 <우리말샘> 등재어를 하나의 어절로 분석하여 분석 대상 어절의 어휘의미를 정확하게 부여할 수 있었다.

이러한 지침 수정에 따른 어휘의미 처리의 전 과정은 지침 관련 협의 과정을 통한 국립국어원의 최종 승인하에 이루어졌으며 원문의 어휘 정보와 수정된 내용의 대조 정보는 말뭉치와 함께 제출하였다.

변경 전	변경 후	<우리말샘> 표제어
템플__888/NNG, 스테이__888/NNG	템플스테이__001/NNG	템플^스테이(template stay)_001
블루__888/NNP, 이즈__888/NNP, 더__888/NNP, 워미스트__777/NNP, 컬러__001/NNP	블루_이즈_더_워미스트_컬러__7 77/NNP	-
바람__001/NNG+과/JKB, 함께/MAG, 사라지__001/VV+다/EF	바람과_함께_사라지다__001/NN P	바람과 함께 사라지다_001

<표 10> 다어절 구성의 어휘의미 분석

이러한 원문 정제 및 어휘의미 분석 지침 수정 결과에 따라 띄어쓰기, 오타자 등 표기 오류에 사용된 ‘999’ 표지는 그 쓰임이 제한되어 삭제가 가능하게 되었다.

로마자 약자로 이루어진 단어 중에는 <우리말샘>에 여러 형태로 등재되어 있는 단어들이 존재하여 어휘의미 연결에 일대일로 대응되지 않는 문제가 확인되었다. 어휘의미 분석 말뭉치 구축 시, ‘FIFA’, ‘UNESCO’ 등의 기관명에 대해 ‘피파’와 ‘에프아이에프에이’, ‘유네스코’와 ‘유엔이에스시오’처럼 <우리말샘>의 등재어가 여러 개인 경우, ‘에프아이에프에이’, ‘아이엠에프’ 등 문자의 이름으로 이루어진 등재어의 어휘의미 번호를 우선적으로 부여하고, 그러한 형태가 등재되어 있지 않은 경우에는 ‘피파’, ‘유네스코’와 같이 단어로 읽는 등재어의 어휘의미 번호를 부여하였다.

분석 대상 어절	<우리말샘> 표제어
FIFA/NNP	피파__003
	에프아이에프에이__001
	국제축구연맹__001
IMF/NNP	국제통화기금__001
	아이엠에프__001

<표 11> 로마자의 어휘의미 분석

## 1.2. 어휘의미 분석 말뭉치 구축 지침

어휘의미 분석 말뭉치 구축 지침의 목차는 아래와 같다. 지침 세부 내용은 [붙임 2] 참조.

어휘의미 분석 말뭉치 구축 지침	
목차	
1.	총론
1.1.	지침의 목적
1.2.	용어 정의
1.3.	기본 원칙
2.	어휘의미 분석의 원리
2.1.	어휘의미 분석 대상
2.2.	어휘의미 분석 표지 부착

2.2.1. 등재어: 등재 번호

2.2.2. 어휘 미등재: 777

2.2.3. 의미 미등재: 888

2.2.4. 분석 불가: 999

### 3. 어휘의미 분석의 실제

#### 3.1. 체언

##### 3.1.1. 명사(NN)

3.1.1.1. 일반명사(NNG)

3.1.1.2. 고유명사(NNP)

3.1.1.3. 의존명사(NNB)

##### 3.1.2. 대명사(NP)

##### 3.1.3. 수사(NR)

#### 3.2. 용언

##### 3.2.1. 동사(VV)

##### 3.2.2. 형용사(VA)

##### 3.2.3. 지정사(VC)

3.2.3.1. 긍정지정사(VCP)

3.2.3.2. 부정지정사(VCN)

##### 3.2.4. 보조용언(VX)

### 4. 주의 사항

4.1. 북한어 및 방언인 경우

4.2. 외래어인 경우

4.3. 숫자나 알파벳, 한자가 포함된 경우

4.4. 기호가 포함된 경우

4.5. 문자 모양의 유사성에 기반한 경우

4.6. 구 등재어인 경우

4.7. <우리말샘> 등재어가 여러 개인 경우

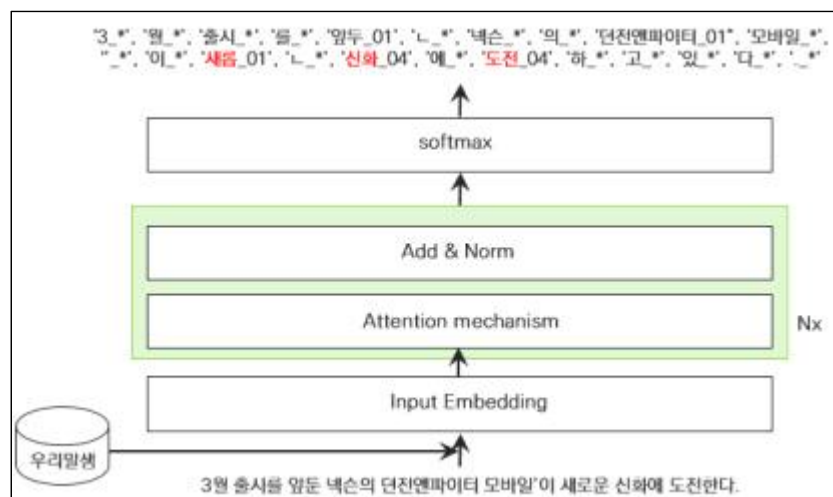
4.8. 참여자 제안 정보인 경우

4.9. 맥락이 불완전한 경우

## 2. 어휘의미 분석 말뭉치 구축 절차

### 2.1. 어휘의미 분석 초벌 작업

어휘의미 분석 초벌 말뭉치는 형태소 분석 결과를 기반으로 단어별 의미번호 (sense\_id)를 자동 부착하는 방식의 어휘의미 태깅 모듈을 통해 생성하였다. 해당 모듈은 <우리말샘>의 의미 구분 체계를 기반으로 사전 정의·예문·언어 정보를 학습 데이터로 활용하였으며, 다의어·동음이의어 구분이 필요한 단어에 대해 문맥 정보를 활용하는 통계·신경망 기반 모델을 적용하였다. 먼저 형태소 분석에서 제공된 기본형(form)과 품사 정보를 입력으로 받아 다의어·동음이의어 후보를 탐색하고, 사전 기반 의미 후보와 문맥 정보를 결합하여 각 어절에 가장 적합한 의미번호를 부여하였다. 초벌 태깅 코드 결과는 JSON 스키마의 어휘의미 층위에 기입되도록 구현하였다. 어휘의미 말뭉치 초벌에 사용된 모델의 구조는 아래 [그림 15]와 같다.



[그림 15] 어휘의미 분석 모델 구조

위 모델을 활용한 초벌 말뭉치의 출력값 예시는 아래와 같다.

```

>>> from nltkcor.tagimport EspressoTagger
>>> tagger = EspressoTagger(task='wsd')
Downloading Espresso5 model...
>>> sent="3월 출시를 앞둔 넥슨의 던전애파이터모바일이 새로운 신화에 도전한다."
>>> print(tagger.tag(sent))
'3_*', '월_*', '출시_*', '를_*', '앞두_01', 'ㄴ_*', '넥슨_*', '의_*', '"_*', '던전애파이터_01', '모바일_*', '"_*', '이_*', '새롭_01', 'ㄴ_*', '신화_04', '에_*', '도전_04', '하_*', '고_*', '있_*', '다_*', '._*'

```

[그림 16] 어휘의미 분석 말뭉치 초벌 말뭉치 출력값 예시

어휘의미 부착 과정에서는 (1) 형태소 분석 기반의 기본 의미 후보 추출, (2) 사전 매칭 및 문맥 계산을 통한 최종 의미 결정, (3) word\_id 표시 방식이 다양한 원시 결과(예: '3', '03', '3-1')를 정규화하여 동일한 문장 내 어절과 1대1로 대응시키는 절차를 수행하였다. 또한, 의미번호는 향후 의미역 분석 단계에서 술어 의미 보정에 활용되므로, 선행 0을 유지한 문자열 형태('003', '014' 등)로 저장하여 정보 손실을 방지하였다.

이와 같은 절차를 통해 구축된 어휘의미 초벌 말뭉치는 형태소-구문-의미역-무형 대용어 복원 계층 간의 연계성을 유지하는 기반 역할을 하였다. 해당 말뭉치로 작업자는 구축 도구의 <우리말샘> 검색 기능을 활용해 자동 부착 결과를 검토·수정하였으며, 빈도가 높은 오류 유형은 추가 사전 보완 및 모델 재학습을 통해 순환적으로 개선하였다. 최종적으로 워크벤치의 검수 과정을 통해 빈도가 높게 관찰되는 오류 유형을 중심으로 수정을 진행하였다. 주요 오류 중 하나는 고유명사 오분석으로, 우리말샘 사전에 띄어쓰기 형태로 등재된 단어에 대해 각 구성 어절에 개별 의미번호가 잘못 부착된 사례이다. '피부 관리'가 피부\_002 + 관리\_007로 분석되거나, '여자 친구'가 여자\_002 + 친구\_002로 분해되어 의미가 부착된 경우가 있었으나, 이는 각각 피부\_관리\_001, 여자\_친구\_001과 같이 단일 고유명사 또는 고정 결합 표현으로 처리하는 것이 적절하다고 판단하여 수정하였다. 이후 동일 유형의 오류가 반복되지 않도록 일관성 검사를 수행하여 전반적인 오류를 정비하였다. 또한 의존명사 및 보조용언 관련 오류로는 의존명사와 조사의 기능 구분이 명확하지 않아 부적절한 의미번호가 부착된 사례와, '놓다', '보다', '버리다' 등 보조용언의 의미를 본용언 의미와 혼동하여 의미번호를 잘못 부여한 사례가 확인되었다. 이러한 오류는 형태소 분석 결과와 구문 정보를 함께 참조하여 용언의 기능적 역할을 재판단하는 방식으로 보정하였다. 마지막으로 sense\_id 777, 888 부착 오류이다. 해당 의미번호는 <우리말샘>에 어휘 또는 의미가 등재되지 않은 경우에 한해 사용하도록 정의되어 있으나, '귀찮아지

다’, ‘많아지다’, ‘심해지다’와 같이 사전에 등재된 어휘에도 777 또는 888이 부착된 사례가 발견되었다. 이에 해당 어휘들을 재검토하여 적절한 의미번호로 교체하고, 반복적인 일관성 검사와 검수를 통해 동일 유형의 오류가 남아 있지 않도록 정비하였다. 이외 오류들에 대해서도 워크벤치 기반의 검수 과정을 통해 체계적인 검토와 수정이 이루어졌으며, 그 결과 최종적으로 정확성과 일관성을 갖춘 어휘의미 말뭉치를 구축하였다.

## 2.2. 어휘의미 분석 작업 중 확인된 문제점 및 개선 방향

어휘의미 분석은 형태 분석 결과를 바탕으로 이루어지며 어휘의미 분석 결과는 이어지는 의미역 분석과 무형 대용어 분석의 기초 자료로 활용된다. 어휘의미 분석은 실제 말뭉치에 나타난 형태가 <우리말샘>에 의미(sense) 단위로 구분되어 등재된 어휘 중 어느 것인지를 확인하는 작업으로, 이를 통해 확인한 격틀 정보로 필수 논항과 부가어를 구별하여 각각에 맞는 의미역을 부여하고 실현되지 않은 논항은 복원하게 된다. 따라서 각각의 의미와 격틀 정보가 담겨 있는 사전의 역할이 무엇보다 중요하다고 할 수 있다.

본 사업에서는 형태 분석 단계에서 <우리말샘> 등재어와 구로 등재된 형태, 고유명은 내부 분석을 진행하지 않고 하나의 어절로 다루어 이어지는 구문 분석에서 불필요한 의존 관계 분석이 이루어지지 않도록 하였다. 그러나 이 과정에서 어휘의미 777과 888을 부여할 수 있는 대상의 목록이 기존 연구와 상당 부분 달라지게 되었다.

‘템플스테이’처럼 통합된 형태만 <우리말샘>에 등재된 경우는 ‘템플\_\_888’, ‘스테이\_\_888’이 ‘템플스테이\_\_001’로 처리되어 어휘의미 777이나 888이 줄어들게 되었고 ‘파리올림픽’과 같이 그 세부 의미가 등재되지 않은 경우에는 ‘파리\_\_012’와 ‘올림픽\_\_001’이 ‘파리올림픽\_\_888’으로 처리되어 어휘의미 777이나 888을 부여받는 대상이 증가하게 되었다.

그러나 똑같이 어휘의미 777이나 888을 부여받고 있지만 이들은 구분되어야 하는 대상으로 판단된다. 사전의 등재어 정보가 늘어난다면 그 방향은 ‘템플\_\_888’과 ‘스테이\_\_888’를 보충하는 것이 아니라 ‘파리올림픽\_\_888’의 의미를 보충하는 방식이 될 것이기 때문이다. ‘템플’과 ‘스테이’는 한국어를 표현하기 위해 ‘템플스테이’로 사용되었을 경우에만 사전의 등재어로서 그 의의를 찾을 수 있다. 따라서 본 사업의 어휘의미 분석 결과로 발생한 어휘의미 777이나 888을 부여받는 대상들을 검토하고 그 의미를 <우리말샘>에 반영하는 과정을 통해 <우리말샘>은 한층 더 현대 한국어의 여러 모습을 포착할 수 있도록 진화할 수 있을 것으로 판단된다.

또한 이어지는 의미역 분석 단계에서는 <우리말샘>의 격틀 정보와 다른 쓰임을 보이는 어휘들이 상당수 확인되기도 하였는데 그 예로 ‘기대하다’는 <우리말샘>에 [-가 -을 기대하다] 혹은 [-가 -기를 기대하다]로 격틀 정보가 제시되어 있으나 실제 말뭉치의 사례를 확인한 결과 ‘재건축 단지의 고민이 해소될 것으로 기대했다’와 같이 [-가 -으로 기대하다]의 쓰임이 다수 확인되었다. 따라서 이때의 [-으로 기대하다]는 <우리말샘>의 격틀 정보를 지니는 ‘기대하다’로 볼 수 없으므로 ‘기대하다\_\_888’을 부여하거나 기존 격틀을 지닌 ‘기대하다\_\_001’로 처리한 후 선행하는 ‘-으로’를 부가역으로 처리하게 되는 문제가 드러났다.

따라서 말뭉치 구축 작업과 함께 말뭉치 구축 과정에서 확인되는 새로운 어휘들과 그 의미, 격틀 정보를 <우리말샘>에 반영시키는 과정이 함께 이루어진다면 실생활에서 사용되는 한국어의 모습을 보다 정교하게 담을 수 있는 말뭉치 자료와 사전을 구축할 수 있을 것이라 기대한다.





## 제 5 장

# 구문 분석 말뭉치 구축



# 1. 구문 분석 말뭉치 구축 지침 개선

## 1.1. 기존 지침 보완 사항

구문 분석 말뭉치는 ‘2024년 구문 분석 및 무형 대용어 복원 말뭉치 연구’ 사업의 결과 보고서를 바탕으로 수정이 필요한 사항을 파악하여 지침 개선 작업을 진행하였다.

본 사업은 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상대화 말뭉치 2024’를 대상으로 원문 정제가 이루어짐에 따라 원문의 맞춤법과 띄어쓰기 오류는 어문 규정에 맞게 정제 단계에서 수정하여 지침에 의한 일관된 분석을 진행할 수 있었다.

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업에서는 21세기 세종계획의 ‘현대 문어 구문 분석 말뭉치 구축 지침’, TTA의 ‘의존 구문 분석 말뭉치 구축을 위한 의존 관계 태그 세트 및 의존 관계 설정 방법’, 2024년 구문 분석 및 무형 대용어 복원 말뭉치 연구 사업의 ‘구문 분석 말뭉치 구축 지침’에서 변하지 않고 계속 유지되어 온 ‘접속과 내포를 구분하지 않고 접속절은 모두 부사절로 분석한다’는 기본 원칙에 따라 서술어의 논항으로 실현된 필수적 부사어 외 수의적 부사어와 절 접속에도 기능표지 AJT(부사어)를 부여하였다. 기존 2024년 지침에서는 내포문을 명사절, 관형사절, 부사절, 인용절의 네 가지 유형으로 나누었는데 이는 TTA에서 명사절, 관형절, 부사절, 서술절, 직접 인용절, 간접 인용절의 여섯 가지 유형으로 나누었던 것에서 서술절을 제외하고 직접 인용절과 간접 인용절을 하나로 묶은 것이었다. 그러나 이는 ‘접속절을 부사절로 분석한다’는 기본 원칙에 위배되는 분석이 이루어지는 결과를 초래한다. 이러한 지침 내 모순을 해결하기 위하여 본 사업의 지침에서는 복문을 명사절, 관형사절, 부사절의 세 가지 유형으로 나누고 인용절은 부사절의 하위 유형으로 다루어 지침 내 분석의 일관성을 유지하였다.

직접 인용절과 간접 인용절은 TTA지침에서 모두 기능표지 CMP를 부여하였는데 2024년 지침에서는 이에 더해 <표준국어대사전>의 문형 정보 조건이 추가되어 보어의 범위가 분명히 제시되지 않고 형태 분석 결과와의 연계도 이루어지기가 어려운 문제가 확인되었다. 이에 본 사업에서는 기능표지 CMP(보어)는 ‘되다’와 ‘아니다’에 선행하여 형태 분석 표지 JKC(보격조사)가 붙은 어절로 한정하고, 2024년 지침에서 사전의 격들에 따라 CMP와 AJT로 다르게 처리되던 인용절은 기능표지 AJT(부사어)를 부여하여 부사절과 보어의 분석 일관성을 확보하였다.

기능표지 CMP(보어) 부여 범위는 ‘보어’를 어떻게 정의할 것인지와 관련되는 문제로,

21세기 세종계획의 ‘현대 문어 구문 분석 말뭉치 구축 지침’에서는 보격조사 JKC가 결합하거나 인용격 조사 JKQ가 결합된 모든 구, 보문동사 및 보문 동사류 표현의 보문 용언 구와 보문절의 분석에 기능표지 CMP를 부여한 바가 있다.

변경 전				변경 후			
그녀가 그 일을 했다고 스스로 말했다.							
<표준국어대사전 격틀 정보에 ‘-고’가 있는 경우>							
1	그녀가	NP_SBJ	4	1	그녀가	NP_SBJ	4
2	그	DP	3	2	그	DP	3
3	일을	NP_OBJ	4	3	일을	NP_OBJ	4
4	했다고	VP_CMP	6	4	했다고	VP_AJT	6
5	스스로	AP	6	5	스스로	AP	6
6	말했다.	VP	0	6	말했다.	VP	0
형사가 용의자에게 사건 시각에 어디에 있었느냐고 신문하고 있었다.							
<표준국어대사전 격틀 정보에 ‘-고’가 없는 경우>							
1	형사가	NP_SBJ	7	1	형사가	NP_SBJ	7
2	용의자에게	NP_AJT	7	2	용의자에게	NP_AJT	7
3	사건	NP	4	3	사건	NP	4
4	시각에	NP_AJT	6	4	시각에	NP_AJT	6
5	어디에	NP_AJT	6	5	어디에	NP_AJT	6
6	있었느냐고	VP_AJT	7	6	있었느냐고	VP_AJT	7
7	신문하고	VP	8	7	신문하고	VP	8
8	있었다.	VP	0	8	있었다.	VP	0

<표 12> 피인용문의 구문 분석

또한 본 사업에서는 주제부각형 언어인 한국어의 특성을 반영함과 동시에 이어지는 무형 대용어 복원 말뭉치 구축과의 연계성을 높이려는 의도에 따라, ‘보조사 ‘은/는’이 결합된 어절에 대응되는 문장 성분(주어, 목적어, 부사어 등)이 있으면 해당 기능 태그를 부여하고, 문장 안의 서술어를 기준으로 대응되는 문장 성분이 없는 경우에는 일괄적으로 SBJ로 분석한다.’는 기존 지침의 설명은 유지하면서, 주제어로 실현된 요소는 최상위 서술어인 모절 서술어에 의존하는 것으로 분석하였다.

변경 전				변경 후			
위원장은 “아직 준비가 미흡하다”며 “빠른 시일 안에 준비를 마치겠다”고 강조했다.							
1	위원장은	NP_SBJ	4	1	위원장은	NP_SBJ	10
2	“아직	AP	4	2	“아직	AP	4
3	준비가	NP_SBJ	4	3	준비가	NP_SBJ	4
4	미흡하다”며	VP	10	4	미흡하다”며	VP	10
5	“빠른	VP_MOD	6	5	“빠른	VP_MOD	6
6	시일	NP	7	6	시일	NP	7
7	안에	NP_AJT	9	7	안에	NP_AJT	9
8	준비를	NP_OBJ	9	8	준비를	NP_OBJ	9
9	마치겠다”고	VP_CMP	10	9	마치겠다”고	VP_CMP	10
10	강조했다.	VP	0	10	강조했다.	VP	0

<표 13> 주제어의 구문 분석

또한 2025년 지침에서는 장형 사동 구문의 피사동주가 대격으로 실현된 경우 최상위 서술어인 모절 서술어에 의존하도록 분석하여 각각의 문형 차이가 드러나도록 하였다. 2024년 지침에서는 장형 사동 구문의 피사동주가 주격, 대격, 여격으로 달리 실현되어도 기능표지만 바꿀 뿐 지배 의존 관계는 동일하게 분석하였다.

변경 전				변경 후			
그는 철수를 집에 가게 하였다.							
1	그는	NP_SBJ	5	1	그는	NP_SBJ	5
2	철수를	NP_OBJ	4	2	철수를	NP_OBJ	5
3	집에	NP_AJT	4	3	집에	NP_AJT	4
4	가게	VP	5	4	가게	VP	5
5	하였다.	VP	0	5	하였다.	VP	0

<표 14> 장형 사동 구문의 구문 분석

## 1.2. 구문 분석 말뭉치 구축 지침

구문 분석 말뭉치 구축 지침의 목차는 아래와 같다. 지침 세부 내용은 [붙임 3] 참조.

구문 분석 말뭉치 구축 지침	
목차	
1. 총론	
1.1. 지침의 목적	

1.2. 용어 정의

1.3. 기본 원칙

## 2. 구문 분석 표지 부착

### 2.1. 구문 분석 표지 집합

2.1.1. 구문 표지 집합

2.1.2. 기능 표지 집합

### 2.2. 구문 분석 표지의 부착

2.2.1. 구문 표지의 부착

2.2.2. 기능 표지의 부착

## 3. 지배-의존 관계 분석의 원리

### 3.1. 논항-서술어 관계 분석

3.1.1. 서술어의 성립

3.1.2. 필수 논항의 성립

3.1.2.1. 주어의 성립

3.1.2.2. 목적어의 성립

3.1.2.3. 보어의 성립

3.1.2.4. 필수적 부사어의 성립

### 3.2. 수식어-피수식어 관계 분석

3.2.1. 관형어의 성립

3.2.2. 수의적 부사어의 성립

### 3.3. 접속 관계 분석

3.3.1. 명사구 접속

3.3.2. 절 접속

## 4. 지배-의존 관계 분석의 실제

### 4.1. 문장 유형별 의존 관계 분석

4.1.1. 단문

4.1.2. 접속문

- 4.1.2.1. 일반적인 접속문
- 4.1.2.2. 선행절에 서술어가 실현되지 않은 경우
- 4.1.3. 포유문
  - 4.1.2.1. 명사절 포유문
  - 4.1.2.2. 관형사절 포유문
  - 4.1.2.3. 부사절 포유문
- 4.2. 구 구성별 의존 관계 분석
  - 4.2.1. 명사구
    - 4.2.1.1. 관형어+명사+명사
    - 4.2.1.2. 수사 및 수 관형사를 포함한 명사구
    - 4.2.1.3. 명사구 접속 구성
  - 4.2.2. 용언구
    - 4.2.2.1. 본용언 + 본용언 구성
    - 4.2.2.2. 본용언 + 보조 용언 구성
    - 4.2.2.3. 의존명사 구성
    - 4.2.2.4. 'NP 중이다' 구성
  - 4.2.3. 부호 포함 어절
  - 4.2.4. 외국 문자/외국어 포함 어절
- 4.3. 세부 유형별 분석 방법
  - 4.3.1. 보조사적 용법의 '이/가', '을/를'
  - 4.3.2. 부사성 명사구
  - 4.3.3. 격 조사가 결합한 수량 표현
  - 4.3.4. 품사와 문장 성분이 일치하지 않는 경우
  - 4.3.5. 서술어적 용법의 '~으로' 명사구
  - 4.3.6. 부사 '없이', '같이', '달리', '함께'
  - 4.3.7. 연결된 부사구 '~부터 ~까지', '~에서 ~으로'
  - 4.3.8. 장형 사동 구문
  - 4.3.9. 명사-부사 통용어
  - 4.3.10. 서술어 생략 문장
- 4.4. 구어 구문 분석을 위한 추가 지침

#### 4.4.1. 단어, 구 단위

##### 4.4.1.1. 감탄사구

###### 4.4.1.1.1. 부사, 대명사와 감탄사의 구별

###### 4.4.1.1.2. 감탄사의 주석

##### 4.4.1.2. 호칭어

##### 4.4.1.2. 주제어

#### 4.4.2. 절, 문장 단위

##### 4.4.2.1. 후보층 구문

##### 4.4.2.2. 구조적 중의성이 있는 명사구

##### 4.4.2.3. 주격/목적격 조사 결합형이 절 표지로 기능하는 경우

##### 4.4.2.4. 문장 종결 부호가 포함된 경우

##### 4.4.2.5. 서술어가 나타나지 않은 경우

## 2. 구문 분석 말뭉치 구축 절차

### 2.1. 구문 분석 초벌 작업

구문 분석 초벌 말뭉치는 형태소 단위 분석 결과를 입력으로 하여, 한국어 문장의 지배-의존 관계(head-dependent)를 자동으로 산출하는 방식으로 구축되었다. Espresso 기반 의존 구문 분석기를 활용하되, 말뭉치 정제·형태소 분석·어휘의미 분석 결과와의 정합성을 고려하여, 문장 단위로 일관된 지배 의존 관계 구조를 생성하도록 설계하였다. 구문 분석 초벌 말뭉치는 이후 의미역 분석(SRL), 무형 대용어 복원(ZA) 등 상위 계층 분석의 기반이 되므로, 어절 경계·형태소 정보와의 연계가 가장 핵심적인 절차로 설정하였다.

#### 2.1.1. 지배소(head) 후보 탐색 및 통사 기능 판별

먼저 각 문장의 형태소 분석 결과에서 VV·VA·VX·VCP·VCN 등 동사·형용사 계열의 품사 태그를 포함하는 어절을 중심으로 지배소(head) 후보를 선별하였다. 한국어의 통사 구조 특성상 용언 또는 용언성 명사(VNP 계열)가 문장 중심(head) 역할을 수행하는 경우가 많기 때문에, 이들 형태소를 포함한 어절을 우선적으로 head 후보로 설정하였다. 이후 다음과 같은 절차를 통해 통사 기능을 판단하였다.

(1) 형태소 분석에서 조사가 부착된 어절(JKS, JKO, JKB 등)을 확인하여 해당 어절이 주어, 목적어, 부사어 등 어떤 통사 기능을 수행하는지를 판별한다. 조사의 유형은 의존 관계를 결정하는 주요 단서이므로, 각각의 조사 범주를 통해 기능 후보를 선별한다.

(2) 관형절·부사절·접속절을 포함하는 경우, EC·ETM 등 절 표지 정보를 활용해 해당 절이 수식 기능을 수행하는지 또는 종속 관계를 형성하는지를 결정한다.

(3) 이렇게 추출된 정보는 문장 내 후보 지배소를 기준으로 정렬되며, 이후 의존 관계를 산출하기 위한 구조적 기반으로 활용된다.

이러한 1단계 처리를 통해 형태소 분석에서 추출된 문법 단서들을 통합한 지배소 후보 집합을 구성할 수 있으며, 이는 이후 의존 관계 생성의 기준으로 사용되었다.

### 2.1.2. 의존 관계 설정 및 DP 구조 생성

두 번째 단계에서는 1단계에서 도출된 지배소 후보와 형태소·조사 정보를 기반으로 실제 의존 관계를 확정하고 구문 분석 구조를 생성한다. 이 단계는 문장 내 모든 어절의 통사적 위치를 지배-의존 관계 형태로 표현하며, 아래 절차에 따라 이루어진다.

(1) 조사의 유형을 기준으로 주어(SBJ), 목적어(OBJ), 보어(CMP) 등 필수 논항을 지배소에 연결한다. 예를 들어 JKS가 부착된 어절은 SBJ로, JKO가 부착된 어절은 OBJ로 연결되며, 보격 표지를 가진 어절은 CMP로 설정된다.

(2) 부사격 조사(JKB), 수식절(ETM/EC 기반), NP\_AJT 등 부사어·수식어 역할을 수행하는 어절을 지배소에 연결하여 MOD 또는 AJT 계열의 dependent로 설정한다.

(3) 접속 조사(JC) 및 접속 어미(EC) 정보를 활용하여 병렬 구조 또는 종속 접속 구조를 판별하고, 병렬 명사구의 경우 동일 head에 병렬 dependent로 연결한다.

이와 같이 산출된 지배-의존 관계는 word\_id를 기준으로 정렬·정규화되며, 각 어절은 {word\_id, head, label} 구조의 DP 항목으로 기록된다. 이렇게 생성된 의존 구조는 JSON v1.7에서 요구하는 DP 필드에 반영되며, 이후 의미역 분석과 무형 대용어 복원 단계에서 통사적 기반으로 활용된다.

이 2단계 과정을 거쳐 형태소 분석-구문 분석-상위 계층 분석 간의 통사적 연계성을 확보하였으며, 문장 단위로 일관된 DP 기반 통사 구조를 갖는 구문 분석 초벌 말뭉치를 완성하였다.



### 2.1.3. 구문 분석 말뭉치 초벌 결과

말뭉치 정제 단계를 거친 결과, 복합 명사와 고유명사의 분석 단위 변경으로 원시 말뭉치의 지배 의존 관계의 수 대비 구축 말뭉치의 지배 의존 관계 수 감소량 내역 및 통계를 확인하였다. 기준은 하나의 문장으로 두었으며 예시의 original\_form은 원시 말뭉치의 form이고 form은 정제 말뭉치의 형태로, 감소량은 diff로 추출하였다. 내역을 바탕으로 말뭉치 전체를 기준으로 감소량 통계를 추출하였다.

문서 id	문장 id	original_form	form	지배 의존 관계 수 (원시)	지배 의존 관계 수	diff
SDRW24000 01336.1	SDRW24000013 36.1.1.38	뭐 펫숍에서 구매하게 다 되면은	뭐 펫_숍에서 구매하게 되면은	4	3	1
SDRW24000 01336.1	SDRW24000013 36.1.1.120	난 그게 너무 싫고 강아지 개 털 날릴 까 봐	난 그게 너무 싫고 강아지 개 털 날릴까 봐	9	8	1

<표 15> 지배 의존 관계의 수 감소량 내역 예시

구분	전체 문장 수	지배 의존 관계 감소 문장 수	문장 감소 비율	지배 의존 관계 감소 비율
문어	13,907	3,636	0.26	0.1
구어	16,439	2,826	0.17	0.15

<표 16> 지배 의존 관계의 수 감소량 통계

## 2.2. 구문 분석 작업 중 확인된 문제점 및 개선 방향

본 사업에서는 기존 구문 분석 지침에서 계속 언급되어 왔지만 실제 적용에는 혼선을 보여 왔던 ‘원칙적으로 접속과 내포를 구별하지 않으며, 접속절은 모두 부사절로 분석한다.’는 구문 분석의 기본 원칙을 엄격히 적용하였으며 기능표지 CMP(보어)를 ‘되다’와 ‘아니다’에 선행하는 형태 분석 표지 JKC(보격조사)가 붙은 어절로만 한정하여 기존 지침에서 사전의 격들에 따라 CMP와 AJT로 다르게 처리되던 인용절에 기능표지 AJT(부사어)를 부여하였다. 그 결과 기능표지 AJT(부사어)를 부여받는 어절의 수가 상당수 증가하는 결과가 나타났다.

그러나 문제는 기능표지 AJT(부사어)를 부여받는 어절들 중에는 서술어의 필수 논항인

필수 부사어와 필수적으로 실현될 필요 없는 수의적 부사어가 혼재해 있다는 사실이다. 물론 이러한 문제는 구문 분석 과정에서는 전혀 문제되는 것이 없으며 그 결과물의 정확성을 떨어뜨리지도 않는다. 다만 본 사업의 목적이 각각의 분석 단계를 통합한 다층위 구조 분석 말뭉치를 구축하는 것이므로 다음 단계인 의미역 분석과 무형 대용어 복원 과정을 고려한다면 기능표지 AJT(부사어)를 부여받는 어절들을 좀 더 세분할 필요성이 제기되는 것이다.

기능표지 AJT(부사어)를 부여받는 어절들이 필수 부사어와 수의적 부사어로 구분된다면 의미역 분석 과정에서는 추가적인 검토 없이도 필수 부사어만을 확인하여 논항의 필수 의미역을 부여하고 수의적 부사어에 대해서는 부가어의 의미역을 부여할 수 있게 된다. 또한 무형 대용어 복원 과정에서는 수의적 부사어는 생략한 채 필수 부사어만을 검토하여 그 생략 여부를 확인하고 복원 과정을 진행할 수 있어 무형 대용어 복원의 작업 난이도를 줄일 수 있게 된다.

따라서 이후에 이루어지는 구문 분석 과정에서는 부사어를 세분하여 한국어 연구에 보탬이 됨과 동시에 다른 말뭉치 구축의 기초 자료로서의 가치를 높이는 작업이 이루어질 필요가 있을 것으로 판단된다.



## 제 6 장

# 의미역 분석 말뭉치 구축



# 1. 의미역 분석 말뭉치 구축 지침 개선

## 1.1. 기존 지침 보완 사항

의미역 분석 말뭉치는 ‘2019년 의미역 분석 말뭉치 구축’ 사업과 ‘2020년 의미역 기술 모형 연구 분석’ 사업의 결과 보고서를 바탕으로 수정이 필요한 사항을 파악하여 국립국어원과의 논의 및 승인 과정을 통해 지침 개선 작업을 진행하였다.

‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상대화 말뭉치 2024’를 대상으로 원문을 정제한 후 형태 분석과 어휘의미 분석, 구문 분석을 진행한 말뭉치를 대상으로 서술어의 논항과 부가어에 대한 의미역 분석을 진행하였다.

본 사업팀은 문형이 달라지면 그에 따라 의미역도 다르게 해석되고 분석되어야 한다는 전제하에 의미역 분석을 진행하였다. 한국어에서는 동일한 하나의 서술어에 대해 서로 다른 두 문형이 존재하는 경우가 자주 확인되는데 이때 다르게 실현된 두 문형에 대해 일관되게 동일한 의미역을 부여할 수도 있고 달라진 격틀에 따라 각각의 문형에 해당하는 서로 다른 의미역을 부여할 수도 있다. 문형의 변화가 화자의 발화 의도에 따라 그에 맞게 선택된 것임을 고려한다면 문형 실현 양상에 따라 의미역 또한 다르게 분석되는 것이 마땅할 것이다. 이러한 의미역 분석 지향을 통해 현실에서 한국어가 사용되는 양상을 보다 정밀하게 관찰할 수 있었으며 일관된 분석 또한 가능하였다.

	변경 전	변경 후
지침	ARG2, GOL	GOL
예문	철수가 학교에(GOL) 갔다.	철수가 학교에(GOL) 갔다.
지침	ARG1, GOL	THM
예문	철수가 학교를(GOL) 갔다.	철수가 학교를(THM) 갔다.
지침	ARG2, GOL	DIR
예문	철수가 학교로(GOL) 갔다.	철수가 학교로(DIR) 갔다.

<표 17> 격표지 실현에 따른 의미역 분석

2019년 의미역 분석 말뭉치 구축 지침과 2020년 의미역 기술 모형 지침은 한국전자통신연구원(ETRI)의 의미역 표지 목록을 그대로 받아들여 사용하였다. 한국전자통신연구원(ETRI)의 의미역 표지 목록은 K-propbank의 의미역 표지 체계를 그대로 따른 것으로, 서술어의 논항은 ARG0, ARG1, ARG2, ARG3처럼 숫자 번호로 표시하고 부가어는

ARGM-LOC, ARGM-DIR, ARGM-INS처럼 표지로 의미역을 표시한다. 그러나 이러한 번호형 의미역 분석 표지 체계를 따르게 되면 논항과 부가어의 의미역 표시 방식이 달라 혼란스럽다는 문제뿐만 아니라 장소(LOC), 도구(INS), 방식(MNR) 등의 논항과 부가어 의미역이 동일한 표지를 공유하고 있다는 사실도 포착하기 어렵게 한다.

	변경 전	변경 후
지침	ARG2, ARG3 ARGM-LOC, ...	LOC, ... M-LOC, ...
예문	철수는 방에(ARG2) 있다. 봄이 되자 공원에(ARGM-LOC) 꽃이 만발했다.	철수는 방에(LOC) 있다. 봄이 되자 공원에(M-LOC) 꽃이 만발했다.

<표 18> 논항과 부가어의 의미역 분석

또한 번호형 분석 표지로 의미역을 표시할 경우 ARG0, ARG1은 대체로 행위주나 대상을 표시하게 되지만 ARG2나 ARG3은 서술어에 따라 수혜주 또는 장소, 혹은 시작점이나 도착점에도 대응될 수 있어 의미역 분석 결과가 불분명해진다는 문제가 있다.

따라서 논항과 부가어의 의미역을 동일한 표지형 분석 표지로 표시하되 부가어에 대해서만 ‘M-’을 결합시키는 것으로 논항과 부가어를 구분하여, 의미역 분석의 정확성을 높이고 논항과 부가어에 동일한 의미역 표지가 공유될 수 있도록 처리하였다.

논항의 의미역 표지 목록과 부가어의 의미역 표지 목록은 [붙임 4]로 제시하였다.

	변경 전	변경 후
지침	ARG0, ARG1, ...	AGT, THM, ...
예문	내가(ARG0) 시장에(ARG2) 갈 테니 내가(ARG1) 집에(ARG2) 있어라. 그는(ARG0) 세입자를(ARG1) 집에서(ARG2) 내보냈다.	내가(AGT) 시장에(GOL) 갈 테니 내가(THM) 집에(LOC) 있어라. 그는(AGT) 세입자를(THM) 집에서(SRC) 내보냈다.

<표 19> 표지형 의미역 분석

논항과 부가어의 의미역을 분석할 때 고려해야 할 문제는 표지에만 국한되지 않는다. 여러 어절로 이루어진 논항과 부가어의 범위는 어떻게 표시되어야 하며 표지는 어디에 부착되어야 하는가 하는 문제 또한 의미역 분석 작업 이전에 고려해야 할 문제 중의 하나이다.

‘2019년 의미역 분석 말뭉치 구축 지침’에서는 여러 어절로 이루어진 구성에 하나의 의미역을 부여해야 하는 경우 논항이 시작되는 어절에는 해당 의미역의 표지를 부착하고

논항이 끝나는 어절에 ‘>>>’ 기호를 부착하는 방식으로 나타내는 방식을 취하였다. 그러나 의미역의 대상 범위는 구문 분석 결과에 따라 자동적으로 도출되는 것이므로 JSON 결과물에는 논항과 부가어의 스패를 표시하되, 실제 작업에는 따로 스패를 지정할 필요 없이 마지막 어절에 의미역 표지를 부착하는 방식을 취하면 해당 지침과 그에 따른 작업 과정을 간소화할 수 있다는 장점이 있다.

또한 의미역 분석 도구에서 논항과 부가어의 스패를 자동으로 표시해 주는 기능을 추가하여 작업 간에 발생할 수 있는 입력 오류를 줄여 완성도 높은 말뭉치 구축 환경을 조성할 수 있었다.

	변경 전	변경 후
지침	첫 어절에 의미역, 마지막 어절에 ‘>>>’ 표시	마지막 어절에 의미역 표시
예문	나는 국민과(ARG2) 언제나 <u>함께하겠다고</u> (>>>) 연설했다.	나는 국민과 언제나 <u>함께하겠다고</u> (CNT) 연설했다.

<표 20> 논항과 부가어의 범위(스패)

의미역 분석 작업을 간소화하려는 노력은 의미역 표지의 간소화 단계로 이어졌다. ‘2019년 의미역 기술 모형 지침’에서는 ETRI의 의미역 체계를 유지하면서도 부가 의미역에 ‘ARGM-PSR(소유자)’ 표지를 새로 설정하여 이중주어문에서 대주어가 소주어의 소유자로 해석되는 경우 그 구문의 대주어를 ‘ARGM-PSR(소유자)’로 분석하였는데 이러한 분석은 이중주어문의 첫 번째 논항을 논항이 아닌 부가어로 분석하는 것이다.

이러한 처리는 의미역 분석 말뭉치 구축 과정에서는 문제될 것이 없으나 다층위 구조 분석 말뭉치 구축을 목표로 삼고 있는 상황에서는 한국어 구조 분석의 이전 단계인 구문 분석 과정에서 해당 어절에 주어(SBJ)의 기능표지를 부여했던 부분과 충돌하는 결과를 초래하게 된다.

따라서 ‘ARGM-PSR(소유자)’를 부가어 의미역 목록에서 제외하고 필수 의미역 목록의 PRT(부분), FTR(속성) 표지로 이를 대체하는 과정을 통해 각 단계의 한국어 구조 분석 말뭉치 간 충돌을 방지하면서도 불필요한 의미역 분석 표지 설정을 제한할 수 있었다. 이에 따라 이중주어문의 첫 번째 논항은 THM(대상)으로, 두 번째 논항은 PRT(부분), 또는 FTR(속성)로 분석할 수 있었다.

	변경 전	변경 후
지침	ARGM-PSR-THM	THM-PRT
예문	코끼리는(ARGM-PSR) 코가(THM) 길다.	코끼리는(THM) 코가(PRT) 길다.
지침	ARGM-PSR-THM	THM-FTR
예문	이 제품은(ARGM-PSR) 가격이(THM) 저렴하다.	이 제품은(TMh) 가격이(FTR) 저렴하다.

<표 21> 이중주어문의 의미역 분석

타동사문에서 도출된 사동문의 주어를 CSR(사동주)로, THM이 이미 부여된 피동문의 주어를 PAT(피동주)로 분석하던 이전 사업 지침 또한 불필요한 의미역 분석 표지로 판단하여 대체 가능한 의미역 표지로의 분석 방식으로 수정하였다.

이전 사업에서 CSR(사동주)과 PAT(피동주)를 설정한 이유는 CSR(사동주)과 PAT(피동주)의 의미역 분석 필요성에 따른 것이 아니라, AGT(행동주), THM(대상)이 이미 부여된 상황에서 사동주와 피동주에 부여할 수 있는 대체 의미역이 필요했기 때문이었다.

따라서 AGT(행동주), THM(대상)를 부여했던 피사동주와 타동문의 목적어를 GOL(도착점)과 PRT(부분)로 분석하고 사동문의 주어를 AGT(행동주), 피동문의 주어를 THM(대상)으로 분석하는 방안을 통해 CSR(사동주)과 PAT(피동주)의 필요성이 해소된다면 불필요한 의미역 표지를 삭제하여 지침을 간소화할 수 있을 뿐만 아니라 각 의미역이 구문에 따라 달리 실현되는 양상을 포착할 수 있는 환경을 조성할 수 있다.

	변경 전	변경 후
지침	CSR-AGT	AGT-GOL
예문	상급자가(CSR) 담당자에게(AGT) 모든 부품을(THM) 검수시켰다.	상급자가(AGT) 담당자에게(GOL) 모든 부품을(THM) 검수시켰다.
지침	PAT-THM	THM-PRT
예문	철수가(PAT) 영희에게(AGT) 손을(THM) 잡혔다.	철수가(THM) 영희에게(GOL) 손을(PTR) 잡혔다.

<표 22> 사동문과 피동문의 의미역 분석

의미역 분석 관련 자문회의(오민석 상명대 교수 자문)를 통해서도 지침의 개선 방향을 확인할 수 있었다. 먼저 서술어 ‘받다’의 논항은 SRC(출발점)와 GOL(도착점)로 분석하였는데 이는 서술어 ‘받다’가 의도성을 가지고 있다고 보기 어렵다는 자문 의견을 반영한 것이다. 명정희·이정훈(2022)을 비롯한 몇몇 연구에서는 ‘선물을 받다’의 주어 처리를 위한

‘BNF(수혜자)’ 의미역이나, ‘가르침/귀염/별/지배를 받다’의 주어 처리를 위한 ‘PAT(피동주)’ 의미역을 제안하기도 하지만 이러한 의미역의 설정은 의미역 목록의 증가를 야기하며 본 사업단의 의미역 목록으로는 GOL(도착점)로 처리하는 방안을 고려할 수 있었다.

	변경 전	변경 후
지침	ARG0, ARG2	GOL, SRC
예문	회장이(ARG0) 동창들한테(ARG2) 협찬금(ARG1) 받아낼 걸 그랬다.	회장이(GOL) 동창들한테(SRC) 협찬금(THM) 받아낼 걸 그랬다.

<표 23> ‘받다’ 구문의 의미역 분석

또한 동일한 연결어미가 사용된 경우에 대해서도 각각의 연결된 절의 의미를 살피 M-CAU(원인), M-PRE(선행) 등으로 그 의미역을 다르게 부여하였다. 의미 발달 과정이 계기에서 이유로 다시 배경으로 발달한 것을 고려하면 M-BCK(배경)를 먼저 분석하고 이어서 M-CAU(원인) 그리고 M-PRE(선행)로의 분석을 고려하는 방식으로 의미역 분석을 적용할 수 있다는 자문 의견을 반영하고 ‘-고서’ 등 비슷한 의미를 가지는 어미들과의 교체 가능성 등을 검토함으로써 보다 정교한 의미역 분석을 적용할 수 있었다.

	변경 전	변경 후
지침	ArgM-CAU	M-CAU
예문	내가 일을 해서(M-CAU) 가족들이 살아간다.	내가 일을 해서(M-CAU) 가족들이 살아간다.
지침	ArgM-CAU	M-PRE
예문	나는 창호지에 구멍을 뚫어서(M-PRE) 방 안의 사람들을 엿보았다.	나는 창호지에 구멍을 뚫어서(M-PRE) 방 안의 사람들을 엿보았다.

<표 24> 연결어미의 의미에 따른 의미역 분석

## 1.2. 의미역 분석 말뭉치 구축 지침

의미역 분석 말뭉치 구축 지침의 목차는 아래와 같다. 지침 세부 내용은 [붙임 4] 참조.

의미역 분석 말뭉치 구축 지침



## 목차

1. 총론
  - 1.1. 지침의 목적
  - 1.2. 용어 정의
  - 1.3. 기본 원칙
2. 의미역 분석 대상 및 범위
  - 2.1. 서술어의 성립
    - 2.1.1. 대상 서술어
    - 2.1.2. 제외 서술어
  - 2.2. 논항의 성립
  - 2.3. 부가어의 성립
  - 2.4. 의미역 분석 대상 범위 설정
3. 의미역 표지 집합
  - 3.1. 필수 의미역
  - 3.2. 부가 의미역
4. 필수 의미역 분석의 실제
  - 4.1. AGT (Agent, 행동주)
  - 4.2. EXP (Experiencer, 경험주)
  - 4.3. MAG (Mental agent, 심리행위주)
  - 4.4. EFF (Effector, 영향주)
  - 4.5. THM (Theme, 대상)
  - 4.6. COM (Companion, 동반주)
  - 4.7. GOL (Goal, 도착점)
  - 4.8. LOC (Location, 장소)
  - 4.9. SRC (Source, 출발점)
  - 4.10. CNT(Content, 내용)
  - 4.11. INS(Instrument, 도구)
  - 4.12. CRT(Criterion, 기준)

4.13. FNS(Final state, 결과상태)

4.14. DIR(Direction, 방향)

4.15. QLF(Qualification, 자격)

4.16. MNR(Manner, 방식)

4.17. PRT(Part, 부분)

4.18. FTR(Feature, 속성)

5. 부가 의미역 분석의 실제

5.1. M-TMP (Temporal, 시간)

5.2. M-LOC (Location, 장소)

5.3. M-DIR (Direction, 방향)

5.4. M-ROU (Route, 경로)

5.5. M-INS (Instrument, 도구)

5.6. M-MNR (Manner, 방식)

5.7. M-QNT (Quantity, 수량)

5.8. M-DGR (Degree, 정도)

5.9. M-EXT (Extent, 범위)

5.10. M-COM (Comitative, 공동)

5.11. M-CRT (Criterion, 기준)

5.12. M-QLF (Qualification, 자격)

5.13. M-CAU (Cause, 원인)

5.14. M-PRP (Purpose, 목적)

5.15. M-CND (Condition, 조건)

5.16. M-CNC (Concession, 양보)

5.17. M-PRE (Preceding, 선행)

5.18. M-BCK (Background, 배경)

5.19. M-LST (List, 나열)

5.20. M-CTR (Contrast, 대조)

5.21. M-SEL (Selective, 선택)

5.22. M-NEG(Negation, 부정)

5.23. M-MOD (Modal, 양태)

#### 5.24 M-CNJ(Conjunction, 접속)

#### 6. 주의 사항

- 6.1. 의미역 분석 표지 부착 시 주의 사항
- 6.2. 구문 분석 결과 검토 시 주의 사항
- 6.3. 논항 여부 판별 및 격틀 구조 분석 시 주의 사항
- 6.4. 부사절 의미역 부착 시 주의 사항

## 2. 의미역 분석 말뭉치 구축 절차

### 2.1. 의미역 분석 초별 작업

의미역 분석 초별 말뭉치는 형태소 분석, 어휘의미 분석, 구문 분석 결과를 통합하여 술어-논항 구조를 재구성하고 의미역 정보를 부착하는 2단계 방식으로 구축하였다. 의미역(SRL)은 문장 내 사건 구조를 체계적으로 표현하고, 구문·의미 계층을 통합하는 핵심 역할을 수행하므로, 술어 후보 탐색-논항 범위 결정-의미역 표지 부착-술어 의미 보정의 절차를 정교하게 설계하였다. 이를 통해 앞선 한국어 구조 분석 층위와 일관된 구조를 가질 수 있도록 하였다.

#### 2.1.1. 구문 분석(DP) 기반 술어 후보 대상 탐지

먼저 각 문장의 구문 분석 결과(DP)에서 VP 계열 노드를 술어 후보로 선별한다. 이때 의존소가 전혀 없는 VP는 실질적인 술어 역할을 하지 않는 경우가 많으므로 제외하고, 하나 이상의 dependent를 가진 VP 노드만을 술어로 채택한다. 선택된 각 술어에 대해서는 다음과 같은 절차로 논항을 구성한다.

(1) 술어의 word\_id를 기준으로 해당 단어의 표층형(form)을 가져오고, 형태소 분석 결과(morph)에서 VV·VA·VX·VCP·VCN 등 동사·형용사 계열 태그가 부착된 형태소를 우선적으로 찾아 사전형(‘어간 + -다’)을 복원함으로써 lemma를 추정한다. 만약 형태소 분석에서 적절한 후보가 없으면 동일 word\_id의 WSD 결과에서 form을 가져와 보충한다.

(2) 술어 노드의 각 직접 의존소(dependent)를 시작점으로 삼아, 구문 트리에서 하위 노드를 모두 수집하여 하나의 논항 span으로 묶는다. 이때 수집된 word\_id 집합을 오름

차순으로 정렬하여 argument에 속하는 토큰들을 순서대로 병합하고, 토큰의 word\_form 혹은 원 단어의 form을 결합하여 argument 표층 표현을 구성한다.

(3) 이렇게 얻어진 술어 정보와 논항 목록을 이용해 문장 내 SRL 구조를 새로 생성하고, 생성 시점에는 모든 술어의 sense\_id를 일단 정숫값 '1'로 고정하여 부여한다. 이 단계에서는 의미역(label) 자체는 비워 두거나, 필요시 별도의 규칙에 따라 부여할 수 있도록 설계하였다.

이와 같은 1단계 처리를 통해, 기존 SRL 결과와 무관하게 구문 분석 구조를 기준으로 일관된 술어-논항 구조를 재구성하였다.

### 2.1.2. 의미역 정보 부착

의미역 분석의 두 번째 단계는 1단계에서 탐지한 술어-논항 후보를 기반으로 실제 의미역 정보를 부착하는 과정으로, 논항의 범위를 체계적으로 설정하고, 각 논항의 의미역 표지를 정의하며, 술어의 원형 정보와 의미번호를 확정하는 절차로 구성된다.

#### (1) 논항 및 부가어 범위 설정(SRL Span 산출)

의미역 부착의 첫 단계는 구문 분석(DP)에서 파악된 술어의 의존 구조를 바탕으로 한 논항의 범위(span) 결정이다. 먼저 각 술어 노드가 가진 직접 의존소(dependent)를 시작점으로 삼아, 구문 트리 하위의 모든 노드를 수집함으로써 논항의 구성 요소를 결정하였다. 수집된 word\_id 목록은 오름차순으로 정렬되어 구성 요소 간의 선형 순서를 보존하며, 이에 대응하는 표층형(form)을 연결하여 논항 표층 표현을 생성하였다. 논항의 범위(span)는 시작 위치값부터 종료 위치값까지 모두 작성되도록 구조화하였으며, 작업자의 작업 중 값 누락 및 기타 발생 오류는 최종 검수 과정에서 수정할 수 있도록 하였다.

#### (2) 의미역 표지(SRL Label) 부착

논항 범위가 확정되면, 다음 단계에서는 각 논항이 술어와 어떠한 의미적 관계를 가지는지 판단하여 의미역 표지(SRL Label)를 부착한다. 초벌 단계에서는 구조적 일관성 확보를 우선시하기 위해 의미역 자체를 자동 분류하지 않고, 의미역 표지는 일단 공란 또는 기본 값으로 부여하고, 술어-논항 구조만을 완성된 형태로 생성하였다. 실제 의미역 표지 판단 및 부착은 후속 작업자 구축 및 전문가 검수 단계에서 정밀하게 수행되도록 설계하였다.

이로써 초별 단계에서는 논항의 후보군과 구조적 틀을 정확하게 확보하는 것을 목표로 하고, 의미역 분류로 인한 초기 오분석을 최소화하였다.

### (3) 술어 원형(SRL Predicate Lemma) 및 의미번호

의미역 구성의 마지막 단계는 술어 정보의 정규화이다. 술어는 논항 구조의 중심이므로, 술어의 표면형과 원형(lemma), 의미번호(sense\_id)를 일관되게 확정하는 과정이 필요하다. 이는 앞서 구축된 어휘의미 분석값을 활용하였으며, 술어 원형의 경우 해당 술어의 word\_id에 연결된 형태소 분석 결과에서 용언 계열 태그가 부착된 형태소를 우선적으로 탐색하고, 해당 형태소의 어간에 ‘-다’를 결합하여 사전형(lemma)을 복원하였다. 또한, 술어의 의미번호(sense\_id)는 의미역 분석 초별 말뭉치 생성 시에는 임의값(1)으로 초기화한 뒤, 후처리 단계에서 어휘의미 분석 말뭉치 결과와 연동하여 보정하였다. 이를 위해 어휘의미 결과의 word\_id\_display를 정규화하여 의미역 predicate\_word\_id와 매칭하고, 일치하는 경우 해당 어휘의미 말뭉치의 의미번호를 의미역 분석 말뭉치에 재부착하였다. 의미번호는 선행 0을 유지한 문자열(예: ‘003’) 형태로 저장하여 말뭉치 스키마의 형식적 일관성을 확보하였다.

위 말뭉치 또한 작업자에게 배포되어 수정 및 보완이 이루어졌으며, 검수 작업까지 반영된 말뭉치는 초별 말뭉치 성능 향상에 일부 활용되었다. 주 차별로 누적된 말뭉치를 바탕으로 앞선 분석 단위와의 일관성 확인 과정도 계속해서 진행되었기에 최종적으로 형태 분석, 어휘의미 분석, 구문 분석, 의미역 분석 및 무형 대용어 복원 총 5개의 층위에서 일관된 말뭉치를 구축하게 되었다.

## 2.2. 의미역 분석 작업 중 확인된 문제점 및 개선 방향

본 사업에서는 어휘의미 분석과 구문 분석이 완료된 말뭉치에 의미역 분석을 진행하는 방식으로 다층위 구조 분석 말뭉치를 구축하였다. 이 과정에서 어휘의미 분석 결과를 바탕으로 <우리말샘>의 다의어 번호와 그 격들을 확인하여 필수 의미역 분석에 도움을 받을 수 있었으며 구문 분석 결과를 통해 각 서술어와 함께 실현된 논항과 부가어의 정보를 확인하여 그에 맞는 의미역 정보를 부여할 수 있었다.

이러한 각각의 말뭉치 분석 정보의 연계는 도구 개발을 통해 가능해졌다. 구문 분석 결과 확인된 의존 관계를 통해 서술어의 논항과 부가어 정보를 확인할 수 있었고 이는 도구에서 논항과 부가어의 시작 위치값부터 종료 위치값 산출을 통해 논항 및 부가어의

범위(span)가 자동으로 의미역 분석 정보에 반영되는 효과를 가져왔다. 이를 통해 단일 의미역 분석 과정에서는 각각의 논항과 부가어의 범위를 작업자가 하나하나 설정하고 입력해야 했던 번거로운 과정을 생략할 수 있었고 기계 처리로 인한 작업자 오류를 감소시켜 의미역 분석 결과의 정확도를 향상시킬 수 있었다.

그런데 다층위 구조 분석 말뭉치 구축을 위한 각각의 말뭉치 연계는 이러한 이점만 가지는 것은 아니었는데, 의미역 분석 과정에서 <우리말샘>에서 제공하고 있는 격틀 정보와 실제 말뭉치에서 확인되는 격틀이 달리 실현된 경우는 여러 단계의 말뭉치 구축 과정에서 문제를 야기하였다. ‘어휘의미 분석 작업 중 확인된 문제점 및 개선 방향’에서 언급한 ‘기대하다’의 격틀 문제가 그 대표적인 경우라 할 수 있다. ‘인터넷에 50% 세일이라고 떴다’에서 확인되는 [-가 -에 -라고 뜨다] 역시 <우리말샘>의 격틀 정보에서는 확인할 수 없는데 상당수의 서술어가 ‘-라고’를 통해 인용절을 그 논항으로 사용함을 말뭉치 분석 결과 확인할 수 있었다. 이렇게 <우리말샘>의 격틀 정보에 제시되지 않은 어휘들은 추후 <우리말샘>이 보완되는 과정을 통해 <우리말샘>의 등재어로 반영될 것으로 보인다. 그러나 충분한 격틀 정보를 확인할 수 있는 사전의 부재는 어휘의미 분석에 더하여 의미역 분석과 이어지는 무형 대용어 복원까지도 어렵게 만드는 주요 원인으로 파악된다. 기존의 ‘의미역 기술 모형 연구 및 시범 구축’ 사업을 통해 제안된 의미역 기술 모형은 의미역 분석의 기초 자료로서 활용될 수 있을 것으로 기대하였으나 그 규모가 크지 않아 실제 말뭉치 분석의 기초 자료로 사용되기에는 제한이 많아 대규모 의미역 사전의 구축이 시급함을 확인할 수 있었다.

또한 본 사업팀에서는 동사(VV), 형용사(VA)로 이루어진 용언구(VP)의 서술어만을 대상으로 의미역 분석을 수행하였는데 앞으로의 의미역 분석 말뭉치 구축 과정에서는 계사 구문이나 논항을 거느리는 용언구 이외의 대상들에 대해서도 의미역 분석 작업이 이루어질 필요가 있다. 이러한 작업은 형태 분석과 구문 분석 정보가 기초 자료로 활용되면 충분히 가능할 것으로 기대되는데, 각각의 지배-의존 관계 분석 결과를 바탕으로 의미역을 분석하고 그 결과를 사전에 반영하여 격틀 정보를 구성할 경우 실제 한국어 사용 모습에 가까운 사전을 구축하는 것이 가능할 것으로 기대된다.



## 제 7 장

무형 대용어 복원  
말뭉치 구축



# 1. 무형 대용어 복원 말뭉치 구축 지침 개선

## 1.1. 기존 지침 보완 사항

무형 대용어 복원 말뭉치는 ‘2024년 구문 분석 및 무형 대용어 복원 말뭉치 연구’ 사업의 결과 보고서를 바탕으로 수정이 필요한 사항을 파악하여 지침 개선 작업을 진행하였다.

‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상대화 말뭉치 2024’를 대상으로 원문을 정제한 후 형태 분석과 어휘의미 분석, 구문 분석을 진행한 말뭉치를 대상으로 서술어의 생략된 논항을 주변 문맥에서 탐지하여 복원하는 작업을 진행하였다.

‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서는 2020년에 구축된 국립국어원 구어 무형 대용어 복원 말뭉치의 선행어와 서술어 간의 문장 거리 통계치 산출 결과, 평균 문장 거리가 -13.8094로 확인됨에 따라 구어의 복원 범위를 분석 대상 서술어가 포함된 문장의 앞 15 문장 / 뒤 2 문장으로 설정하면서도 문어의 경우는 이전 지침들에서 제시한 ‘하나의 문서 전체’라는 범위를 그대로 유지하였다.

그러나 본 사업의 대상 말뭉치 중 문어 말뭉치는 ‘국립국어원 신문 말뭉치 2024’라는 점을 고려하면 무형 대용어의 동일 지시적 성분은 해당 주제의 신문 기사 내에서만 확인할 수 있다는 것을 짐작할 수 있다. 따라서 문어의 복원 범위를 방대한 분량의 해당 문서 전체로 삼는 것은 분석 결과의 정확성을 올리기보다는 오히려 작업의 효율성만을 떨어트리는 결과를 초래하게 된다. 이에 본 사업에서는 문어 말뭉치도 구어 말뭉치와 동일하게 분석 대상 서술어가 포함된 문장의 앞 15 문장 / 뒤 2 문장으로 복원 범위를 설정하여 작업의 효율성을 추구하고 불필요한 탐색으로 인한 복원 정확성 저하 가능성을 방지하였다.

변경 전	변경 후
문어: 하나의 문서 구어: 앞 15문장 / 뒤 2문장	문어/구어: 앞 15문장 / 뒤 2문장

<표 25> 무형 대용어 복원의 선행어 탐색 범위

또한 ‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서 나타나는 용어 사용의 혼란을 정리하여 지침의 가독성을 높였다. 이는 지난 사업 지침에서 확인되는 일부 혼란상이 용어 사용의 문제에서 기인한다는 판단에 따른 것이다. 즉 ‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서 사용되고 있는 ‘복원하는 표현’은 복원 이전의 대상으로도, 복원이 완료된



복원 이후의 대상으로도 해석이 가능하며, ‘선행어’, ‘선행어 후보’, ‘후행하는 문장에서 선행어를 찾아’ 등과 같은 표현 역시 그 의미를 정확히 파악하기 어려워 용어 사용의 혼란스러운 상황을 가중시키고 있다는 것이다.

이에 사용되는 용어를 정리하여 무형 대용어 복원의 기준이 되는 서술어는 ‘복원 기준 서술어’로, 문장에서 서술어가 요구하는 필수 성분 중 생략된 성분은 ‘무형 대용어’로, 문장에서 실현되지 않은 필수 성분과 지시하는 대상이 동일한 성분은 ‘동일 지시적 성분’, 선행 문장 혹은 후행 문장을 참고하여 격틀에 맞게 복원한 성분은 ‘복원 성분’으로 용어를 정리하고 지침에서 일관되게 사용하여 용어 사용의 혼란에서 일어날 수 있는 문제점을 해소하였다.

	변경 전	변경 후
용어	(복원) 대상 서술어, 무형 대용어 복원 대상, 탐색 대상	복원 기준 서술어
	무형 대용어, 생략어	무형 대용어(=생략 성분)
	선행어, 선행어 후보	동일 지시적 성분
	무형 대용어, 선행어	복원 성분

<표 26> 무형 대용어 복원 관련 용어 정비

용어 사용의 혼란상은 ‘2024년 구문 분석 및 무형 대용어 복원 말뭉치 연구 사업’의 결과물 JSON 파일에서도 확인되는데 기존 JSON 파일의 데이터 구조에서 ‘antecedent’의 하위 항목 중 ‘type’은 복원 성분의 정보이고, ‘sentence\_id’, ‘begin’, ‘end’는 동일 지시적 성분의 정보로 확인된다. ‘form’은 명사구일 경우(조사를 포함하지 않은 경우)에는 복원 성분 및 동일 지시적 성분에 공통적인 정보이고, 비지시적 대명사일 경우(조사를 포함한 경우)에는 복원 성분의 정보가 되어 용어 사용의 층위가 일정하게 정리되지 않은 모습을 보인다. 이에 복원 성분 관련 정보와 동일 지시적 성분 관련 정보를 구분하기 위해 JSON 파일 구조를 아래와 같이 수정하여 체계의 혼란상을 정리하고 작업 결과물의 가독성을 높일 수 있었다.

	변경 전	변경 후
항목	선행어 관련 정보: 형태, 문장 성분 유형, 문장 번호, 문자열 시작 번호, 문자열 끝 번호	복원 성분 관련 정보: 복원 성분의 형태, 문장 성분 유형
		동일 지시적 성분 관련 정보: 동일 지시적 성분의 형태, 문장 번호, 문자열 시작 번호, 문자열 끝 번호
예시	<pre>{   "predicate": {     "form": "싸늘했다",     "sentence_id": "NIXX2400000000.123.45.6",     "begin": 0,     "end": 4   },   "antecedent": [     {       "form": "공기",       "type": "subject",       "sentence_id": "NIXX2400000000.123.45.5",       "begin": 8,       "end": 11     }   ] },</pre>	<pre>{   "predicate": {     "form": "싸늘했다",     "sentence_id": "NIXX2400000000.123.45.6",     "begin": 0,     "end": 4   },   "restored": [     {       "form": "공기가",       "type": "subject"     }   ]   "antecedent": [     {       "form": "공기를",       "sentence_id": "NIXX2400000000.123.45.5",       "begin": 8,       "end": 11     }   ] }</pre>

<표 27> 무형 대용어 복원 말뭉치 JSON 구조

복원 대상 서술어에 대해서도 개선이 진행되었다. ‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서는 복원 기준 서술어 어절에 기호나 보충 설명이 있어 서술어가 여러 개 포함되는 경우 그 형태와 관계없이 각각의 서술어를 모두 분석하였다. 그러나 층위별 말뭉치의 통합을 고려한다면 각각의 분석 단위와 내용이 일관되게 유지시킬 필요가 있다.

본 사업은 형태, 어휘의미, 구문, 의미역 분석 말뭉치를 각각 구성하여 구조 분석 말뭉치를 구성하는 것과 무형 대용어 복원 말뭉치를 구축하는 두 가지 분리된 과제를 수행하였다. 그러나 두 과제의 대상 말뭉치가 동일하므로 형태, 어휘의미, 구문, 의미역 분석 말뭉치와 함께 하나의 통합된 말뭉치를 구성하고 말뭉치 구축의 일관성을 확보하였다. 이에 따라 구문 분석 결과와 의미역 분석 결과를 기초로 삼아 무형 대용어 복원을 진행하여 훨씬 효율적인 작업이 이루어질 수 있었다.

따라서 본 사업에서는 이전 단계의 구조 분석 결과를 바탕으로 무형 대용어 복원 대상을 선정하였으며 이에 따라 기호를 통해 하나의 어절에 여러 서술어가 포함된 경우 이전 단계인 형태 분석과 구문 분석, 어휘의미 분석 결과를 고려하여 일관된 분석 방향을 유지하였다. 일례로 ‘악수, 격려하며’와 같이 연결 부호로 통합된 서술어는 형태 분석 단계에서 ‘악수’는 NNG, ‘격려하며’는 VV로 분석되고 이에 따라 이어지는 구문 분석에서 ‘악수’는 NP, ‘격려하며’는 VP로 분석되는 점을 고려한다면 다층위 구조 분석 말뭉치 구성을 위해서는 ‘격려하다’에 대해서만 무형 대용어 복원이 이루어지는 것이 가능해진다. 만약 ‘상실(20% 이상)했다면’과 같이 하나의 서술어가 쓰였지만 어문 규범에 따라 두 어절로 분리될 수밖에 없는 경우에는 구문 분석 단계에서 NP와 NP\_AJT로 분석된다는 점을 고려하면 무형 대용어 복원의 대상에서 제외하는 것이 타당성을 얻게 된다.

	변경 전	변경 후
지침	복원 기준 서술어: 악수하다, 격려하다	복원 기준 서술어: 격려하다
예문	선수들을 보자 그는 일일이 (그) (선수들) <b>악수.</b> (그) (선수들) <b>격려하며</b> 기념사진까지 찍었다.	선수들을 보자 그는 일일이 <b>악수.</b> (그) (선수들) <b>격려하며</b> 기념사진까지 찍었다.
지침	복원 기준 서술어: 상실하다	복원 기준 서술어 없음
예문	산불로 사업용 자산 등을 (누군가) <b>상실(20% 이상)했다면</b> 소득세·법인세를 공제하고 사업상 어려움이 있다면 연말까지 세무조사 등도 연기한다.	산불로 사업용 자산 등을 <b>상실(20% 이상)했다면</b> 소득세·법인세를 공제하고 사업상 어려움이 있다면 연말까지 세무조사 등도 연기한다.

<표 28> 서술성 명사 나열 구성과 서술어 분리 구성의 처리

‘2024년 무형 대용어 복원 말뭉치 구축 지침’은 ‘2019년 주격 무형 대용어 복원 말뭉치 구축 지침’과 ‘2020년 목적격 무형 대용어 복원 말뭉치 구축 지침’을 바탕으로 그 일부를 개정하여 사용하였는데, 무형 대용어의 동일 지시적 성분(선행어)이 구로 제시된 경우에는 그 마지막 어절만을 동일 지시적 성분으로 지정하여 무형 대용어를 복원하는 방식은 그대로 유지된 것이 확인된다.

그러나 이러한 방식을 따르게 되면 이름과 호칭어 및 직책, 직위명만을 복원하게 되어 언어학적 지식에 맞지 않는 분석이 이루어지게 되며, 한 문장 안에 같은 직책의 여러 인물이 등장할 경우 생략된 성분이 누구를 지시하는지 불분명해지는 문제가 발생하게 된다. 이는 생략된 구의 핵심 마지막 어절만을 복원하면서 발생하는 문제로 생략된 전체 어절 복원을 통해 생략된 정보의 온전한 파악이 가능해질 것으로 판단하여 본 사업에서는 구

전체를 복원 대상 범위로 선정하여 생략된 정보의 온전한 파악이 가능하도록 지침을 개선하였다.

	변경 전	변경 후
지침	구의 마지막 어절만 복원	구 전체를 복원
예문	(씨) 서울에 살고 있는 김철수 씨	(김철수 씨가) 서울에 살고 있는 김철수 씨
	(영희) 서울에 살고 있는 철수와 영희	(철수와 영희가) 서울에 살고 있는 철수와 영희
	김 의원은 이 의원과의 단일화를 추진하고 있으며 5일 (의원) 마지막 협상에 나선다.	김 의원은 이 의원과의 단일화를 추진하고 있으며 5일 (김 의원이) 마지막 협상에 나선다.

<표 29> 복원 대상 범위

비지시적 대명사에 대해서도 수정 필요성이 확인되었다. ‘2019년 주격 무형 대용어 복원 말뭉치 구축 지침’에서는 주어에 대한 비지시적 대명사로 조사가 없는 ‘누군가, 무언가’를 분석에 사용하였고 ‘020년 목적격 무형 대용어 복원 말뭉치 구축 지침’에서는 목적어에 대한 비지시적 대명사로 조사가 결합된 ‘누구를, 무엇을’을 분석에 사용하였는데, ‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서 이들을 하나의 체계로 묶어 제시하면서 조사의 결합 여부 관련 일관성을 잃게 되는 문제점이 확인되었다.

생략된 성분이 조사가 결합된 논항임을 고려한다면 복원되는 성분 역시 서술어의 문형에 따라 조사가 결합된 형태가 되어야 한다. ‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서도 이러한 점을 고려하여 보어나 부사어에 대한 비지시적 대명사에 조사가 결합된 형태를 사용하고 있다. 따라서 주어의 비지시적 대명사에도 조사가 결합된 ‘누군가가, 무언가가’를 사용하여 무형 대용어 복원을 진행하는 것이 형식의 일관성을 확보할 수 있다고 판단하여 각 복원 성분은 조사가 결합된 형태로 복원하는 것을 원칙으로 삼아 무형 대용어 복원을 진행하였다.

생략어의 문장 성분	변경 전	변경 후
주어	누군가	누군가가
예문	민간건설사가 분양하는 판교신도시의 마지막 중대형 물량으로 (누군가) 채권입찰제는 (어딘가에) <u>적용하지</u> 않는다.	민간건설사가 분양하는 판교신도시의 마지막 중대형 물량으로 (누군가가) 채권입찰제는 (어딘가에) <u>적용하지</u> 않는다.

목적어	누구를	누군가를
예문	내가 (누구를) <b>보러</b> 갈게.	내가 (누군가를) <b>보러</b> 갈게.
보어	누군가가, 무언가가	누군가가, 무언가가
부사어	어딘가에, 어딘가에서, 누군가에게서/무언가에게서, 누군가로/무언가로, 누군가와/무언가와, 누군가보다/무언가보다	어딘가에, 어딘가에서, 누군가에게서/무언가에게서, 누군가로/무언가로, 누군가와/무언가와, 누군가보다/무언가보다

<표 30> 비지시적 대명사를 이용한 무형 대용어 복원

복원되는 형태에 있어서도 기존 지침에서 보이던 조사 결합의 혼란상을 정리하였다. ‘2024년 무형 대용어 복원 말뭉치 구축 지침’은 복원 성분의 일관성이 유지되지 않고 있음을 알 수 있다. 지침의 설명과 일부 제시된 예문에서는 조사가 결합되지 않은 형태로 복원 성분을 제시하고 있지만 조사가 결합한 형태도 일부 존재하며, 위에서 제시된 비지시적 대명사는 조사가 결합된 형태로 복원 성분을 사용하고 있어 지침 내 일관성이 지켜지지 않고 있었다. 또한 결과물 형식에서도 조사가 결합한 형태와 조사가 결합되지 않은 형태가 모두 복원 성분에 드러나고 있어 지침과 결과물의 일관성 또한 지켜지지 않고 있는 문제가 확인되었다.

이에 본 사업에서는 비지시적 대명사의 형태와 같이 모든 복원 성분을 조사 결합 형태로 유지하는 것을 통하여 지침 내 일관성의 문제, 지침과 결과물의 일관성 문제를 모두 해결할 수 있었다.

복원 성분의 문장성분	복원 성분의 실현 형태
주어	NPOI/NP가
목적어	NP을/NP를
보어	NPOI/NP가
부사어	NP에, NP에서, NP에게서, NP로/NP으로, NP와/NP과, NP보다

<표 31> 복원 성분의 문장성분과 실현 형태

	변경 전	변경 후
지침	조사 미결합 / 결합 형태 공존	조사 결합 형태로 분석
예문	(김철수) 서울에 <b>살고</b> 있는 김철수	(김철수가) 서울에 <b>살고</b> 있는 김철수

	내가 (김철수) 아는 김철수	내가 (김철수를) 아는 김철수
--	-----------------	------------------

<표 32> 복원 성분의 격표지 표시

‘2024년 무형 대용어 복원 말뭉치 구축 지침’은 주격 중출 구문에 대한 별도의 세부 지침이 마련되어 있지 않아 ‘코끼리가 코가 길다’와 같은 주격 중출 구문의 필수 성분이 생략된 경우 생략된 어느 한 성분의 복원만 가능하였다.

본 사업의 의미역 분석 지침에서는 주격 중출 구문의 두 논항을 ‘THM-PRT(부분)’, ‘THM-FTR(속성)’으로 다루고 있으므로 이와 연계된 분석을 위하여 관련 내용을 지침에 추가하여 주격 중출 구문의 두 논항 중 어느 하나가 생략되었더라도 생략된 논항을 분석할 수 있는 방안을 마련하여 보다 정밀하고 세밀한 구조의 분석이 가능하게 하였다.

	변경 전	변경 후
지침	생략된 주어를 복원(별도의 지침 없음)	주격 중출 구문의 NP1과 NP2를 모두 복원
예문	코가 <b>긴</b> 코끼리는 코를 손처럼 사용한다.	(코끼리가) 코가 <b>긴</b> 코끼리는 코를 손처럼 사용한다.
	돈이 <b>많은</b> 철수는 친구들에게 밥을 잘 산다.	(철수가) 돈이 <b>많은</b> 철수는 친구들에게 밥을 잘 산다.

<표 33> 이중주어문의 무형 대용어 복원

또한 ‘2024년 무형 대용어 복원 말뭉치 구축 지침’은 의미 부류가 맞지 않으나 조사 ‘을/를’이 실현된 경우에 목적격조사가 실현되어 서술어의 격틀이 충족되었다고 판단하여 목적어를 따로 복원하지 않았다.

그러나 아래 예문 ‘다섯 시간을 걸었다’의 ‘다섯 시간을’은 구문 분석 결과 ‘부사어(AJT)’로 분석되며 의미역 분석 결과 ‘수량(M-QNT)’의 의미역을 부여받는 대상이 되므로 목적어로 다루면 분석의 일관성을 잃게 되는 문제가 발생한다.

이에 이전 단계에서의 한국어 구조 분석 결과를 바탕으로, 조사 ‘을/를’이 실현되었으나 의미 부류가 맞지 않는 요소를 부사어로 다루고 본래의 목적어 논항을 복원하여 통합 분석 말뭉치와의 일관된 분석을 지향하였다.

	변경 전	변경 후
지침	목적어가 실현되어 복원 안 함	자동사적 용법이 있어 목적어를 복원 안 함
예문	나는 학교까지 다섯 시간을(OBJ) 걸었다.	나는 학교까지 다섯 시간을(AJT) 걸었다.
지침	목적어가 실현되어 복원 안 함	등산을
예문	나는 작년부터 등산을 다니기 시작했다. 작년부터 올해까지 5번을(OBJ) 다녔다.	나는 작년부터 등산을 다니기 시작했다. 작년부터 올해까지 (등산을) 5번을(AJT) 다녔다.

<표 34> 수량 표현의 무형 대용어 복원

‘2024년 무형 대용어 복원 말뭉치 구축 지침’에서는 직접 인용절에서 화자 및 청자에 해당하는 성분이 생략된 경우 정해진 범위 내에서 동일 지시적 성분을 찾아 복원하며, 정해진 범위 내에 동일 지시적 성분이 없을 경우 비지시적 성분을 사용해 복원하도록 하였다.

그러나 아래와 같은 예문에서 생략된 성분은 동일 지시적 성분을 앞뒤 문장에서 찾을 수 없을 뿐, 그 대상이 화자(혹은 청자)로 분명하게 규정된다는 점에서 비지시적 대명사로 처리하기에 문제가 있었다. 이에 ‘화자’, ‘청자’를 복원 표현에 추가하여 정해진 범위 내에서 화자/청자에 대한 동일 지시적 성분을 찾을 수 없는 경우에도 직접 인용절의 분석 정확성을 높였다.

	변경 전	변경 후
지침	비지시적 대명사	화자, 청자
예문	철수가 “(누군가) 밥 <b>먹는</b> 중이다”라며 곧 집에 가겠다고 한다	철수가 “(화자가) 밥 <b>먹는</b> 중이다”라며 곧 집에 가겠다고 한다
	김철수는 “(팬들이) (누군가에게) 많은 사랑을 <b>주는</b> 팬들에게 감사하다.”라고 말했다.	김철수는 “(팬들이) (화자에게) 많은 사랑을 <b>주</b> 는 팬들에게 감사하다.”라고 말했다.

<표 35> ‘화자’ 및 ‘청자’를 이용한 무형 대용어 복원

기존 지침에서 관형절의 피수식어가 의존명사 ‘것’으로 실현된 경우 ‘것’으로 복원하던 것도 의존명사 ‘것’ 앞과 뒤에 지시 관형사와 조사를 결합하여 ‘그것을’과 같이 복원하는 방식으로 개선하여 복원된 형태의 이질감을 해소하고 가독성을 높였다.

	변경 전	변경 후
--	------	------

지침	것	그것을
예문	이 빵집은 [(누군가) 쌀가루로 (것) 만든] 것을 주로 판다.	이 빵집은 [(누군가가) 쌀가루로 (그것을) 만든] 것을 주로 판다.

<표 36> 의존명사 구성의 무형 대용어 복원

## 1.2. 무형 대용어 복원 말뭉치 구축 지침

무형 대용어 복원 말뭉치 구축 지침의 목차는 아래와 같다. 지침 세부 내용은 [붙임 5] 참조.

### 무형 대용어 복원 말뭉치 구축 지침

#### 목차

1. 총론
  - 1.1. 지침의 목적
  - 1.2. 용어 정의
  - 1.3. 기본 원칙
2. 무형 대용어 복원의 원리
  - 2.1. 복원 기준 서술어 확인
    - 2.1.1. 복원 기준 서술어
    - 2.1.2. 복원 제외 서술어
  - 2.2. 필수 성분 복원 절차
    - 2.2.1. 생략 성분의 확인
    - 2.2.2. 생략 성분과 동일 지시적인 성분 탐지
    - 2.2.3. 필수 성분 복원
      - 2.2.3.1. 복원 단위
      - 2.2.3.2. 복원 표현
3. 무형 대용어 복원의 실제
  - 3.1. 주어 복원
    - 3.1.1. 주어 복원 방법
    - 3.1.2. 주어 복원 시 주의 사항



## 3.2. 목적어 복원

### 3.2.1. 목적어 복원 방법

### 3.2.2. 목적어 복원 시 주의 사항

## 3.3. 보어 복원

### 3.3.1. 보어 복원 방법

### 3.3.2. 보어 복원 시 주의 사항

## 3.4. 필수 부사어 복원

### 3.4.1. 부사어 복원 방법

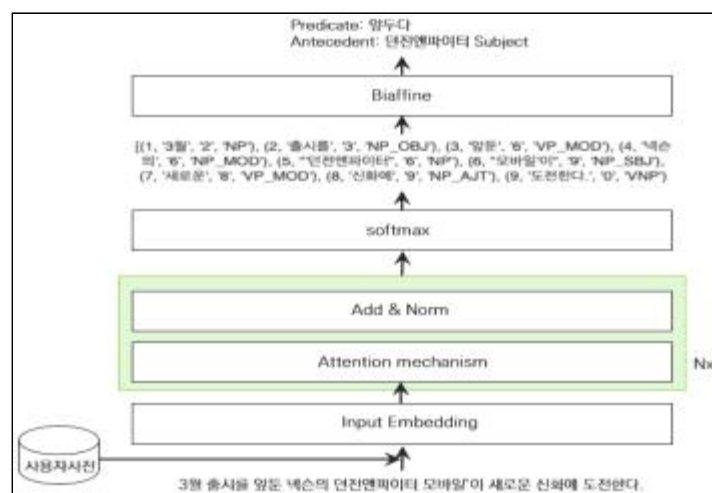
### 3.4.2. 필수 부사어 복원 시 주의 사항

## 2. 무형 대용어 복원 말뭉치 구축 절차

### 2.1. 무형 대용어 복원 초벌 작업

무형 대용어 복원 초벌 말뭉치는 기존에 구축된 형태소, 구문 분석 정보를 활용하고, 대규모 언어모델(LLM)을 결합한 하이브리드 방식으로 생성하였다. 먼저 문장 단위로 술어와 논항 후보를 자동 추출하여 LLM의 입력 범위를 제한하고, 이후 LLM이 예측한 복원 결과를 다시 말뭉치 구조에 맞게 정규화·그룹핑하여 각 문장에 무형 대용어 복원 정보를 부착하는 절차로 진행되었다.

무형 대용어 복원 말뭉치 초벌에 사용된 모델의 구조는 아래 [그림 17]과 같다.



[그림 17] 무형 대용어 복원 말뭉치 구축 모델 구조

### 2.1.1. 문장별 무형 대용어 복원 후보 생성

무형 대용어 복원 말뭉치 초벌에 입력된 자료는 형태소, 어휘의미, 구문, 의미역 총 4개의 층위로 쌓인 한국어 구조 분석 말뭉치 JSON으로, 각 문서는 document-sentence 계층 아래에 어절 정보(word), 형태소 분석 결과(morph), 구문 분석 결과(DP), 의미역 분석 결과(SRL)를 포함한다. 먼저 문서 내 모든 문장에 대해 문장 ID와 표면 문자열을 매핑하고, 구문 분석 결과를 이용해 무형 대용어 복원 후보를 추출한다.

구체적으로, 각 문장의 의존 구문 트리에서 VP, VNP 계열의 노드를 서술어(predicate) 후보로 선정하고, 각 노드의 라벨에 포함된 기능 표지를 이용하여 다음과 같이 논항 후보를 구분한다. SBJ가 포함된 노드는 주어 후보, OBJ가 포함된 노드는 목적어 후보, CMP 또는 VNP, VNP\_MOD로 표시된 노드는 보어 후보로 수집한다. 수식적 명사구인 NP\_AJT는 필수적 부사어의 후보로 간주하여 부사어 목록에 포함한다. 여기에 더해 형태소 분석 결과에서 부사격 조사 JKB가 부착된 어절을 탐지하거나, 표면형이 ‘에게서, 에서, 에게, 으로/로, 와/과, 에, 보다, 라고’ 등과 같이 부사격 조사로 끝나는 어절을 추가로 탐색하여 부사어 후보를 보강한다.

이렇게 수집된 후보 집합은 중복을 제거하되 최초 출현 순서를 보존하여, 문장별로 서술어 목록과 주어·목적어·보어·부사어 후보 목록을 묶은 무형 대용어 복원 후보 사전으로 정리된다. 이 정보는 이후 LLM이 복원해야 할 대상과 탐색 범위를 명시적으로 제시하는 데에 사용되었다.

### 2.1.2. LLM 기반 무형 대용어 복원 예측

문장별 무형 대용어 복원 후보가 구축되면, 각 문장에 대해 문장 ID와 문장 표면 문자열의 쌍과 함께 무형 대용어 복원 후보 사전을 LLM 기반 태거(LLM\_Tagger, task="za")에 입력한다. LLM은 주어진 문맥과 술어·논항 후보 목록을 바탕으로, 문서 내에서 생략된 주어, 목적어, 보어, 필수적 부사어가 존재하는지 판별하고, 해당 생략어가 지시하는 선행 성분(선행어)의 표면 형태와 문장 위치를 예측한다.

LLM의 출력은 개별 예측 단위마다 “서술어 표현, 복원할 역할 유형(예: subject/object/complement/adjunct), 선행어 표현, 서술어가 속한 문장 ID, 선행어가 속한 문장 ID” 형태로 반환되며, 이는 후속 정규화 단계에서 말뭉치 JSON 구조에 맞는 무형 대용어 복원(ZA) 항목으로 변환된다.

### 2.1.3. 복원 결과 정규화 및 ZA 구조 부착

LLM이 산출한 예측 결과는 그대로 사용하지 않고, 실제 말뭉치 내 어절 정보와의 정합성을 확보하기 위해 별도의 정규화·그룹핑 절차를 거친다. 먼저 각 예측 항목에 대해 서술어와 선행어로 제시된 문자열에서 인용부호 및 말미의 문장부호를 제거하고, 해당 문자열이 속한 문장의 어절 목록과 대조하여 실제 말뭉치에 존재하는 어절 형태로 매핑한다. 이 과정에서 문자열이 어절과 정확히 일치하지 않는 경우에는 말뭉치 어절에서 문장부호를 제거한 형태와 비교하거나, 공백으로 분리된 어절들 가운데 가장 적절한 항목을 선택하는 방식으로 근사 매칭을 수행한다.

이렇게 정규화된 결과는 서술어 표현과 서술어 문장 ID를 기준으로 묶어, 하나의 서술어에 대해 하나의 ZA 항목을 구성한다. 각 ZA 항목은 predicate와 antecedent로 이루어지며, predicate에는 서술어의 표층형과 문장 ID를, antecedent에는 복원 대상 역할 유형(주어, 목적어, 보어, 부사어), 선행어의 표층형, 선행어가 속한 문장 ID를 포함한 객체 목록이 저장된다. 동일 서술어·동일 선행어·동일 역할 조합이 중복으로 생성되는 경우에는 한 번만 유지하여, 의미적으로 중복되는 복원 결과가 과다하게 저장되지 않도록 하였다.

마지막으로 이렇게 구성된 ZA 항목들을 문장 ID를 기준으로 다시 재배열하여, 각 문장 객체에 ZA 필드로 부착함으로써 문장별로 서술어와 그에 대응하는 무형 대용어 복원 정보가 연결된 초별 ZA 말뭉치를 완성한다. 이후 단계에서는 이 초별 ZA 말뭉치를 기반으로 구축 작업 및 전문가 검수를 수행하여 복원 결과의 정확성과 일관성을 보정하였다.

위 모델을 활용한 초별 말뭉치의 출력값 예시는 아래와 같다.

```
>>> from nltkor.tagimport EspressoTagger
>>> tagger = EspressoTagger(task='za')
Downloading Espresso5 model...
>>> sent=" 3월 출시를 앞둔 넥슨의 던전애파이터모바일'이 새로운 신화에 도전한다 "
>>> print(tagger.tag(sent))
Predicate: 앞두다
Antecedent: 던전애파이터모바일, Subject
```

[그림 18] 무형 대용어 말뭉치 초별 출력값 예시

무형 대용어 복원 결과, <표 37>과 같이 문어 말뭉치에서 31,712개, 구어 말뭉치에서는 18,535개의 서술어에 대한 무형 대용어 복원이 이루어졌다.

말뭉치 구분	무형 대용어 복원 서술어 수
문어	31,712
구어	18,535

<표 37> 문어/구어 무형 대용어 복원 서술어 수량

무형 대용어 복원 말뭉치 결과를 확인한 결과, 문서 내에서 선행어가 복원된 비율은 문어가 74%, 구어가 75%이고 비지시 대명사로 복원된 비율은 문어가 약 26%, 구어가 25%이다. 문어와 구어 말뭉치 모두 선행어가 복원된 비율이 더 높았으며, 두 말뭉치 간 선행어 복원 차이도 크지 않은 결과를 보였다.

말뭉치 구분	전체 선행어	선행어 비율	비지시 대명사 복원 비율
문어	37,960	0.74	0.26
구어	24,863	0.75	0.25

<표 38> 선행어 및 비지시 대명사 복원 비율

선행어 및 비지시적 대명사 복원 비율 추출은 구축 말뭉치 json 파일 내 무형 대용어 복원(ZA) 하단 선행어 수량 추출을 우선하였다. 선행어는 앞 15개 문장, 뒤 2개 문장에서 찾은 form이 기입된 형태와 기입되지 않은 형태로 구성되어 있다. 선행어 복원 비율은 form이 기입되어 있는 수량을 바탕으로 추출하였으며 비지시적 대명사 복원 비율은 form이 #로 입력되어 있는 대상을 산정하여 추출하였다. 추출된 복원 비율은 <표 40>과 같다.

유형(type)	문어		구어	
	빈도	비율	빈도	비율
주어(subject)	22,321	58.878%	13,203	53.105%
목적어(object)	2,638	6.959%	3,532	14.206%
보어(complement)	1	0.003%	17	0.068%

부사어(adjunct)	3,134	8.267%	1,878	7.554%
<b>합계</b>	<b>28,094</b>	<b>74%</b>	<b>18,630</b>	<b>75%</b>

<표 39> 선행어 유형별 복원 비율

유형별로 복원된 선행어 빈도 및 비율의 경우 문어와 구어 말뭉치 모두 주어의 복원 비중이 가장 높았고, 문어 말뭉치는 부사어가 두 번째로 많은 복원 비율을 보였으나 구어는 목적어가 그다음으로 높은 비율을 보였다.

유형(type)	문어		구어	
	빈도	비율	빈도	비율
주어(subject)	4,534	11.960%	3,634	14.617%
목적어(object)	877	2.313%	893	3.592%
보어(complement)	12	0.032%	20	0.080%
부사어(adjunct)	4,392	11.586%	1,685	6.777%
<b>합계</b>	<b>9,815</b>	<b>26%</b>	<b>6,232</b>	<b>25%</b>

<표 40> 비지시 대명사 유형별 복원 비율

비지시 대명사로 복원된 선행어 비율은 위 [표 40]과 같다. 복원된 결과를 보면 문어의 경우 주어의 복원 비율이 가장 높았으며 부사어의 경우 구어 말뭉치 부사어 비율의 약 3배 가량 높은 비율을 보였다. 구어에서는 주어의 복원 비율이 가장 높았다. 이외 목적어와 보어의 복원 비율은 문어와 구어 모두 유사한 비율을 보였다.

복원된 비지시적 보어, 필수적 부사어 대명사 목록은 아래 [표 41], [표 42]와 같다. 문어와 구어 말뭉치 분석 결과 발견된 목록이 상이하여 두 개의 표로 추출하여 비교할 수 있도록 하였다.

대분류	목록
필수적 부사어	#
	나와
	누구에게
	누군가가
	누군가로
	누군가보다
	누군가에
	누군가에게
	누군가에게서

대분류	목록
필수적 부사어	너에게
	누군가로
	누군가보다
	누군가에게
	누군가에게서
	누군가에서
	누군가와
	누군가한테
	무언가로

	누군가와
	무언가
	무언가가
	무언가라고
	무언가로
	무언가를
	무언가보다
	무언가에
	무언가에게
	무언가에게서
	무언가에서
	무언가와
	무엇이라고
	무엇인가에
	소식통에서
	어딘가로
	어딘가로부터
	어딘가를
	어딘가에
	어딘가에서
	어딘가와
	어딘에게
	청자가
	청자에게
	키움에서
	화자에
	화자에게
	화자와
보어	무언가가

<표 41> 비지시 보어, 필수적 부사어  
대명사 목록(문어)

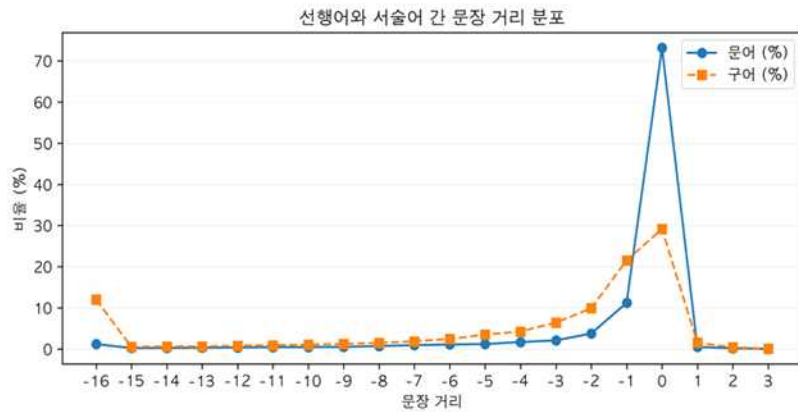
	무언가를
	무언가보다
	무언가에
	무언가에게
	무언가에게서
	무언가에서
	무언가와
	무엇과
	무엇으로
	무엇인가로
	어디에서
	어딘가로
	어딘가보다
	어딘가에
	어딘가에
	어딘가에게
	어딘가에서
	어떻게
	언젠가로
	집에
	청자에게
	청자와
	호텔에
	화자에게
	화자에게서
	화자와
보어	누군가가
	무언가가
	무엇이

<표 42> 비지시 보어, 필수적 부사어  
대명사 목록(구어)

마지막으로 복원된 선행어와 서술어 간 문장 거리 분포 및 비율은 아래 [표 43]과 같고, 분포는 [그림 19]와 같다. 금년도 구축 무형 대용어 복원 지침상 선행어 탐색 거리를 문어와 구어 모두 앞 15 문장, 뒤 2 문장으로 한정하였다. 그 결과, 문어와 구어 모두 서술어가 있는 문장에서 가장 높은 복원 비율을 보였다. 그러나 문어는 앞 3 문장 외 복원된 비율이 높았으며, 이외 거리에서는 고른 분포를 보였다. 구어는 특히 16 이하 누적 거리에 대해 높은 비율을 보였는데, 구어 말뭉치 특성상 잦은 문장 분리로 인한 결과임을 확인할 수 있다.

문장 거리	문어		구어	
	빈도	비율 (%)	빈도	비율 (%)
-16 이하 누적	565	1.16	3,217	12.04
-15	98	0.20	128	0.48
-14	113	0.23	167	0.63
-13	139	0.29	157	0.59
-12	167	0.34	213	0.80
-11	210	0.43	237	0.89
-10	213	0.44	277	1.04
-9	233	0.48	325	1.22
-8	341	0.70	388	1.45
-7	440	0.91	496	1.86
-6	520	1.07	646	2.42
-5	569	1.17	927	3.47
-4	806	1.66	1,129	4.23
-3	1,005	2.07	1,710	6.40
-2	1,813	3.74	2,637	9.87
-1	5,433	11.20	5,748	21.52
0	35,527	73.24	7,785	29.14
1	234	0.48	408	1.53
2	77	0.16	103	0.39
3 이상 누적	5	0.01	14	0.05
합계	48,508	100.00	26,712	100.00

<표 43> 선행어와 서술어 간 문장 거리 비율



[그림 19] 선행어와 서술어 간 문장 거리 분포

## 2.2. 무형 대용어 복원 작업 중 확인된 문제점 및 개선 방향

무형 대용어 복원은 문장 내 서술어의 격틀과 실현된 논항을 확인하고 비실현된 논항과 동일 지시적인 성분을 정해진 문서의 범위 내에서 탐색하여 복원하는 과정을 거쳐 무형

대용어 복원을 완료한다.

이 과정에서 문제가 되는 것은 확인된 동일 지시적 성분을 통해 논항을 복원할 때 그 범위(Span)를 어디까지 설정해야 하는가의 문제이다. 가령 ‘서울시에 거주하는 A씨는 판매 사원을 통해 계약서에 서명하고 보험 가입을 마쳤다.’라는 문장에서 ‘서명하다’의 주어로 복원되는 대상은 ‘서울시에 거주하는 A씨’인지 ‘A씨’인지가 복원 과정에서 문제가 되는 것이다. 이러한 문제는 ‘우리 학교가 3년 만에 참가했지만 예선에서 탈락한 그 대회’와 같이 수식하는 말이 길고 수식받는 대상인 ‘대회’와 직접적인 관련성이 떨어질수록 그 판단이 어려워진다. 이러한 판단의 어려움은 개별 작업자의 몫이 되고 구축된 말뭉치의 일관된 분석을 제한하게 된다.

따라서 무형 대용어의 스패 문제는 무형 대용어 복원 과정 이전에 충분한 논의를 통해 합의가 이루어져야 하는 부분이라고 할 수 있다. 말뭉치에 실현된 문장에 따라 무형 대용어의 스패가 달라질 수 있으므로 특정 어절 수로 제한하거나 수식받는 어절만으로 복원 대상을 한정할 경우 복원 결과의 가독성이 떨어지는 문제를 야기할 수 있다.

구조 분석 결과를 참조하되 동일지시적 성분의 탐색 범위 내에서 가장 짧은 대상을 선택하여 복원하는 방안이나 최초에 등장하는 형태로 복원하는 방안도 하나의 해결책이 될 수 있을 것이다.





## 제 8 장

# 다층위 구조 분석 말뭉치 구축



## 1. 다층위 구조 분석 말뭉치 구축 지침 작성

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업의 목적은 각 층위의 한국어 분석 말뭉치를 구축한 후 각 분석 결과가 하나로 통합된 다층위 구조 분석 말뭉치를 구축하는 것이다. 이러한 목적에 따라 지침도 각 층위별 지침들 사이의 통일성을 확보하고 각각의 분석 층위 간 관계가 유기적으로 드러날 수 있도록 단계별 한국어 분석 지침을 하나로 통합하여 통합 지침을 구성하였다.

다층위 구조 분석 지침의 구성은 한국어 구조 분석에 필요한 용어 정의로 시작하여 각각의 분석 단계가 이전 단계의 분석 결과를 참고하여 진행될 수 있도록 구성하였으며, 또 이전 단계의 분석은 이후 분석 단계의 작업을 고려하여 그 토대를 구성할 수 있도록 유기적인 체계를 구성하였다.

형태 분석 단계에서 불필요한 단어 내부 분석을 지양하여 이어지는 어휘의미 분석과 구문 분석 과정에서의 불필요한 분석 소요를 줄였으며 의미역 분석과 무형 대용어 복원도 구문 분석 결과를 토대로 가시적으로 드러나는 ‘논항-서술어 관계’와 ‘수식-피수식어 관계’를 통해 분석 및 복원 대상을 확인할 수 있어 효율적인 작업을 진행할 수 있었다.

## 2. 다층위 구조 분석 말뭉치 구축 지침

다층위 구조 분석 말뭉치 구축 지침의 목차는 아래와 같다. 지침 세부 내용은 [붙임 6] 참조.

### 다층위 구조 분석 말뭉치 구축 지침

#### 목차

#### 1. 총론

##### 1.1. 지침의 목적

##### 1.2. 용어 정의

##### 1.3. 기본 원칙

#### 2. 형태 분석 표지 부착의 원리

##### 2.1. 형태 분석 대상

##### 2.2. 형태 분석 표지의 집합

## 2.3. 형태 분석의 실제

## 3. 어휘의미 분석 표지 부착의 원리

### 3.1. 어휘의미 분석 대상

### 3.2. 어휘의미 분석 표지의 집합

### 3.3. 어휘의미 분석의 실제

## 4. 구문 분석 표지 부착의 원리

### 4.1. 구문 분석 대상

### 4.2. 구문 분석 표지의 집합

#### 4.2.1. 구문 표지의 부착

#### 4.2.2. 기능 표지의 부착

### 4.3. 분석의 실제

## 5. 의미역 분석 표지 부착의 원리

### 5.1. 의미역 분석 대상

### 5.2. 의미역 분석 표지의 집합

#### 5.2.1. 필수 의미역 표지의 부착

#### 5.2.2. 부가 의미역 표지의 부착

### 5.3. 의미역 분석의 실제

## 6. 무형 대용어 복원의 원리

### 6.1. 무형 대용어 복원 대상

### 6.2. 무형 대용어 복원 형태의 집합

#### 6.2.1. 기본 복원 형태

#### 6.2.2. 비지시적 대명사의 복원 형태

### 6.3. 무형 대용어 복원의 실제

### 3. 다층위 구조 분석 말뭉치 구축의 의의와 확인된 문제점

본 사업의 목적은 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’를 대상으로 형태 분석 말뭉치와 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치, 무형 대용어 복원 말뭉치를 각각 구축하고 이들을 하나의 말뭉치로 통합하여 다층위 구조 분석 말뭉치를 구축하는 것이다. 따라서 각각의 말뭉치의 분석 정보가 상호 연동될 수 있도록 구조 분석 단계마다 이전과 이후 단계에 이루어질 분석들을 고려하여 분석 대상과 분석 내용을 한정하였다.

그 대표적인 작업들이 고유명(NNP)의 범위를 확대 적용하여 어휘의미 분석과 구문 분석에서의 불필요한 분석 과정을 줄이는 것이었다. <우리말샘>은 기존 사전과 달리 노래나 책의 제목, 법률명 등 복합명사나 구로 이루어진 고유명이 등재되어 있어 각각의 등재어는 그 내부 구조 정보를 <우리말샘> 사전을 통하여 확인할 수 있다. 따라서 어휘의미 말뭉치 구축 단계에서 각각의 분석 대상을 <우리말샘>의 다의어 정보와 연결해 주면 등재어의 내부 구성 정보는 말뭉치를 통해 밝히지 않아도 <우리말샘>을 통해 충분히 확인이 가능해진다. 오히려 이러한 등재어의 내부 구성 정보를 각각의 말뭉치에서 분석하게 되면 분석 대상의 증가로 인해 말뭉치의 구조가 복잡해지고 작업 난이도가 어려워져 결과물의 정확도가 낮아지는 결과를 초래할 수 있다.

이렇게 한 어절로 통합된 고유명들은 <우리말샘>에 등재되어 있지 않은 경우 어휘의미 777이나 888을 부여받게 되는데, 이들은 차후 <우리말샘>의 등재어 후보가 되어 사전 정보의 양적 증가를 위한 기초 자료로 활용될 수 있다. 어휘의미 777을 부여받은 대상들만을 한정하고 검토하는 작업을 통해 <우리말샘>의 새로운 등재어를 확보하는 절차를 간소화할 수 있으며 어휘의미 888을 부여받은 대상들만을 한정하여 기존 <우리말샘> 등재어의 새로운 어휘의미와 격률 정보를 보완할 수 있게 되는 것이다.

다층위 구조 분석 말뭉치는 언어학 연구에도 그 쓰임이 활발할 것으로 예상되는데 하나의 어절이 가지고 있는 여러 단계의 구조 정보를 통해 한국어의 다채로운 쓰임 양상의 포착이 가능해질 것으로 예상된다.

구문 분석 과정에서 부사어를 필수 부사어와 수의적 부사어로 구분하는 작업이 이루어지지 못한 점은 아쉬운 부분이라 할 수 있는데, 구문 분석 과정에서 필수 부사어와 수의적 부사어의 구분이 이루어진다면 이후 진행되는 의미역 분석 단계에서 필수 부사어와 수의

부사어를 구분하는 작업이 불필요해지므로 분석 난이도를 낮출 수 있으며 필수 부사어에 부가 의미역이 부여되는 오류 또한 사전에 방지할 수 있게 될 것이다.

이러한 구분은 무형 대용어 복원 과정에서도 유용하게 쓰일 수 있다. 무형 대용어는 문장에서 생략된 논항을 복원하는 작업이므로 이전 단계인 어휘의미 분석 과정에서 확인된 어휘의미와 격틀 정보를 통해 논항 목록을 확정하고 구문 분석 결과를 통해 실현된 논항과 생략된 논항을 파악할 수 있어 말뭉치에 나타난 특정 부사어가 논항인지 아닌지를 판별해야 하는 불필요한 과정을 생략할 수 있게 된다.

따라서 차후 이루어지는 말뭉치 구축 사업에서는 각각의 말뭉치 구축뿐만 아니라 각 단계의 한국어 구조 분석 결과가 연동될 수 있는 다층위 구조 분석 말뭉치 구축을 염두에 두어 각각의 지침 수립과 말뭉치 구축에 힘쓸 필요가 있다.



## 제 9 장

# 납품 자료의 품질 관리



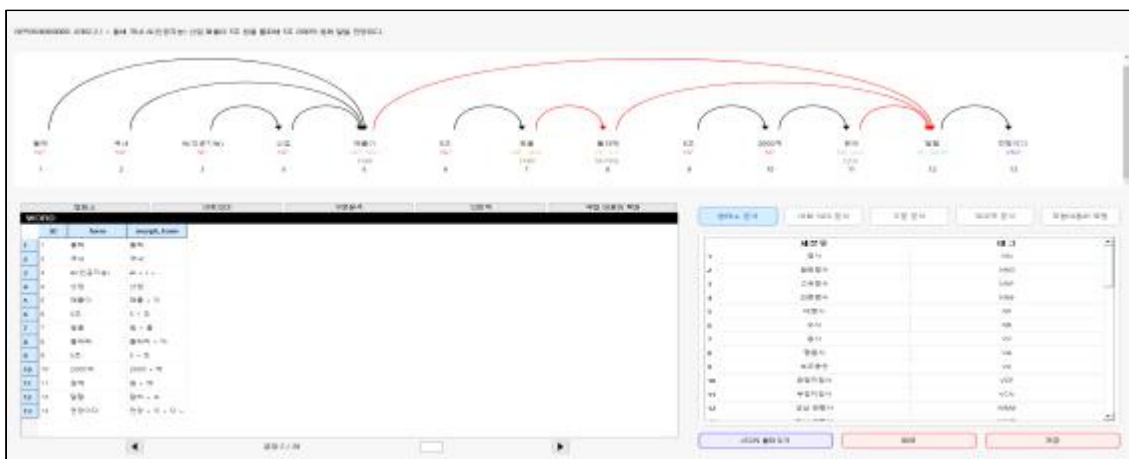
# 1. 시스템 개발

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’에서는 형태소 분석-어휘의미 분석-구문 분석-의미역 분석-무형 대용어 복원의 5종 말뭉치를 단일 환경에서 구축·검수·관리할 수 있는 통합 워크벤치를 개발하였다.

본 시스템은 데이터 입력부터 구조 분석, 오류 수정, 결과 저장 및 JSON 변환까지의 전 과정을 하나의 흐름으로 연결하여 구축 효율성과 품질을 동시에 확보하도록 설계되었다.

## 1.1. 워크벤치

본 사업에서 사용된 워크벤치는 형태 분석, 어휘의미 분석, 구문 분석, 의미역 분석, 무형 대용어 분석 등 총 5종의 말뭉치를 구축하고 앞선 분석 결과를 순차·상호 참조하여 각각의 분석 층위를 쌓아 다층위 구조 분석 말뭉치를 구축할 수 있는 통합 분석 도구이다. 이에 개별 분석 말뭉치 구축에 용이한 도구를 개발하되, 각 분석 단계에서 오류가 발견될 경우 이전 층위의 작업 내용을 즉시 수정할 수 있는 기능을 갖춘 통합 분석 도구를 개발하고 보완, 관리하였다. 통합 분석 도구의 기본 화면 구성은 아래 [그림 20]과 같다.



[그림 20] 통합 분석 도구 전체 화면

한국어 통합 구축 도구 최상단에는 한국어 구조 분석 대상 문장이 나타난다. 5종의 말뭉치 구축 작업의 기본 단위는 한 문장으로 설정되어 있다. 구조 분석 대상 문장 하단으로 말뭉치 구축 결과물을 한눈에 볼 수 있는 뷰어 창이 존재하는데 가로와 세로 모두 스크롤 기능이 있어 긴 문장도 무리 없이 작업이 가능하다. 뷰어 창에는 BRAT(Brat Rapid Annotation Tool)을 활용하여 지배소-의존소의 의존 관계 및 구문 표지 그리고

의미역 표지 등을 각 word 단위에 부착한 결과를 보이도록 제공하여 각각의 분석 층위 시에도 앞선 층위 결과를 참고하는 데에 도움을 준다. 구축 결괏값은 하단에 위치한 표에 자동으로 기입되어 분석 결과의 층위 쌓기 간 발생한 오류를 즉각적으로 확인하고 수정할 수 있다. 구축할 말뭉치는 표 상단에 위치한 박스 탭을 통해 선택할 수 있다. 작업자는 구축할 말뭉치를 좌측 표 상단과 우측 가이드라인 화면 상단에서 모두 선택한 후 작업을 진행한다.

말뭉치 구축 작업 창은 각 작업에 참조할 수 있는 정보가 입력된 표와 직접 수정이 가능한 표로 구성되어 있다. 말뭉치 분석 작업의 가이드라인 뷰어 창은 작업 중 가이드라인 내용을 참고할 수 있도록 각 작업별 태그 세트, 우리말샘 검색 창, 원문 등이 표시되어 효율적인 작업을 가능하게 한다. 그 하단으로 작업과 관련된 부가적인 기능 버튼들이 위치해 있는데, 작업 문서 열기는 'json 불러오기' 버튼을, 작업자와 검수자 간 의견 교환은 '메모' 버튼으로, 작업이 완료되었을 때 '저장' 버튼을 누르면 json 문서에 결괏값이 자동으로 반영된다.

#### (1) 형태소 분석 작업

형태소 분석은 한국어 구조 분석 말뭉치 구축의 기초가 되는 절차로, 이후 구축되는 어휘의미 분석(WSD), 구문 분석(DP), 의미역 분석(SRL), 무형 대용어 복원(ZA)의 정확도를 좌우하는 핵심 단계이다. 본 사업에서는 작업자와 검수자가 형태소 초벌 결과를 검토하고 표지를 부착하는 과정에서 필요시 분석 단위를 수정할 수 있도록, 형태소 분석 도구에 행 추가·삭제 기능을 도입하였다. 해당 기능은 word ID 열에서의 마우스 우클릭으로만 사용 가능하도록 제한하여 형태소 분석 단위를 보완하거나 분할/통합해야 하는 경우 이외엔 활용되지 않도록 하였다.

추가된 행에 형태소 정보(form 또는 pos\_tag)가 입력되지 않은 상태에서는 저장을 제한하여, 불완전한 형태소 구조가 말뭉치에 반영되는 것을 방지하였다. 또한, 형태소 분석값 수정 시에는 morph\_form을 구성하는 토큰 수와 형태소 표지(pos\_tag)의 수가 일치하지 않는 오류가 발생하지 않도록 도구 내부에 자동 검증 기능을 추가하였다. 작업자가 두 항목의 수량을 일치하지 않게 입력하는 경우, 도구는 각 분석값의 토큰 수를 계산하여 기존 값으로 복원하거나 오류 경고를 제공함으로써 형태소-표지 단위 간 불일치로 인한 구축 오류를 방지하였다. 형태소 분석 도구의 기능은 통합 분석 도구에도 동일하게 적용되어, 상위 분석 단계에서 형태소 정보를 재검토하거나 수정해야 하는 경우에도 동일한



방식으로 작업을 수행할 수 있다. 이를 통해 이후 구축될 어휘의미, 구문 분석, 의미역 분석으로 이어지는 계층 간 분석 결과의 정합성을 유지하고, 작업자가 자연스럽게 오류를 탐지·수정할 수 있는 작업 환경을 제공하였다.

ID	form	morph. form	pos. tag
1	출매	출매	NNG
2	국내	국내	NNG
3	AI(인공지능)	AI + ( +	NNG + SS +
4	산업	산업	NNG
5	매출	매출 + 이	NNG + KS
6	매출액	매출 + 액	NN + NR

[그림 21] 형태소 분석 행 추가/삭제 기능

ID	form	morph. form	pos. tag
1	출매		
2	국내		
3	AI(인공지능)		
4	산업		
5	매출액	매출 + 액	NNG +

[그림 22] 형태소 분석 토큰 수 자동 검수 기능

## (2) 어휘의미 분석 작업

어휘의미 분석 작업은 형태소 분석을 기반으로 각 단어의 문맥적 의미를 결정하는 단계로, 먼저, 가이드라인 화면에 <우리말샘> 검색 기능을 추가하여 구축자가 의미 후보를 빠르게 검색하고 입력할 수 있도록 하였다. 이를 통해 다의어·동음이의어의 의미 후보를 확인하는 과정이 간소화되었으며, 의미번호 선택의 근거를 명확하게 확보할 수 있었다.

또한, 분석값 입력 표에는 word 목록과 형태소 분석 결과를 표시하여, 구축자가 WSD 판단에 필요한 정보를 즉시 확인할 수 있도록 하였으며, 작업 편의성과 형식 오류 방지를 위해 sense\_id 자동 정규화 기능을 도구에 구현하였다. 이는 작업자가 ‘1, 2, 11’과 같이 정수 형태로 입력하더라도, 저장 시점에 시스템이 자동으로 ‘001, 002, 011’과 같이 세 자리 문자열 형태로 변환하여 저장되는 기능이다. 이를 통해 작업자의 작업 효율을 높이고 입력 오류를 방지할 수 있었다.



[그림 23] 어휘의미 분석 가이드라인 <우리말샘>

### (3) 구문 분석 작업

앞서 구축된 말뭉치를 도구로 열고, 구문 분석 작업을 선택하면 화면 우측 가이드라인 영역에 구문 태그(지배소 유형)와 기능 태그(SBJ, OBJ, CMP 등)가 등장하고, 분석값 입력 표에는 word 목록과 형태소 분석 결과와 구문 분석 말뭉치 초벌 결과가 자동으로 제시된다. 작업자는 이를 참고하여 지배소(head)와 의존소(dependent), 그리고 기능 태그를 직관적으로 입력할 수 있다. 특히 본 사업에서는 구문 표지 입력의 정확성과 속도를 높이기 위해 태그 부착 방안을 간소화하였다. 작업자는 구분 분석 표지(DP\_label)를 입력할 셀을 선택한 뒤 가이드라인에 제시된 구문 태그 또는 기능 태그를 더블 클릭하면 해당 태그가 자동으로 기입되며, 두 태그를 연이어 선택할 경우 필요한 구분자(\_)도 자동 삽입되도록 설계하였다. 복잡한 표지를 직접 입력하지 않아도 되어 입력 오류를 크게 줄였으며, 저장 시점에는 입력된 DP label이 뷰어 영역의 각 word 하단에 자동 표시되어 부착 결과를 즉각 확인할 수 있으며, head 값까지 입력하면 지배-의존 관계가 화살표로 시각화되어 작업자가 작업 결과를 제출 이전에 재검토할 수 있도록 하였다.

HPFW040000001\_47462.2.2 : 2021년 이혼 신청료 41.9% 설정한 결과다.

형태소		어휘코드		구문분석		의미역		주요 의존어 목록	
ID	form	morph,form	pos,tag	WSD Form	Sense_ID	head	DP Label		
1	2021년	2021 + 년	SN + NNB	년	002	2	NP		
2	이혼	이혼	NNB	이혼	003	5	NP_Alt		
3	신청료	신청료	NNB	신청료	001	4	NP		
4	41.9%	41.9 + %	SN + SW			5	NP		
5	설정함	설정함 + L	VV + ETH	설정함	002	6	VP_MOD		
6	결과다	결과 + 이 + 다	NNB + VCP + EF	결과 + 이	004 + 008	0	VNP		

[그림 24] 구문 분석 작업 결과 화면

### (4) 의미역 분석 작업

분석값 표와 가이드라인 상단 ‘의미역 분석’ 버튼을 선택하면 의미역 분석 가이드라인에는 ‘논항의 의미역 표지’와 ‘부가어의 의미역 표지’가 표시되고 작업 화면에는 앞서 구축된 모든 층위의 결과값이 표시된다. 작업자는 작업 화면에서 의미역 분석 작업에 필요한 분석

단위를 제외한 불필요한 단위는 표 위 해당 분석명이 기입된 버튼을 눌러 숨길 수 있다.

작업자는 의미역 초별 말뭉치를 바탕으로 논항 및 부가어의 범위(SRL span), 의미역 표지(label), 술어의 원형(predicate lemma) 값을 확인하고 서술어와 논항 및 부가어의 의미 관계를 파악하여 값을 수정한다. 논항 범위는 해당 논항을 구성하는 word\_id를 쉼표로 구분하여 직접 입력하는 방식으로 수정하며, 의미역 표지는 특정 행을 선택한 뒤 가이드라인에 제시된 의미역 표지(ARG, ARGM 등)를 더블 클릭하면 해당 표지가 자동으로 입력되는 방식으로 진행하여 표기 오류를 방지하였다. 또한, 술어의 원형은 ‘word id/lemma’ 형식으로 간략하게 기입할 수 있도록 하였다. 의미역 분석 결과도 구문 분석과 동일하게 ‘저장’ 버튼을 누르면 입력된 정보가 즉시 뷰어 창에 반영되어 span 및 lemma 값 누락 등의 오류를 쉽게 점검할 수 있게 하였다.



[그림 25] 의미역 분석 작업 방법

#### (5) 무형 대용어 복원 작업

‘무형 대용어 복원’ 작업을 선택하면, 현재까지 구축된 결과물이 뷰어 창에 표시되고 작업자가 문장 간 문맥을 파악하고 복원 가능한 선행어를 탐색할 수 있도록 가이드라인 창에는 ‘문서 원문’이 표시된다. 작업 화면에는 word 목록과 형태소 분석 결과가 함께 제공되어, 복원 대상이 되는 서술어의 위치를 쉽게 확인할 수 있다. 작업자는 초별 말뭉치 및 구문 분석 정보를 참고하여 무형 대용어가 생략된 서술어(predicate)를 탐색하고, 문서 내 등장하는 동일 지시 성분 후보를 검토한 뒤 적절한 선행어를 입력한다. 이후 복원 성분을 직접 입력하고 복원 문장 성분(주어, 목적어, 보어, 부사어)을 선택한다.

선행어 입력은 테이블의 ant\_sen\_id와 ant\_word\_id 칸에 해당 문장의 번호와 어절 ID를 기입하는 방식으로 이루어진다. 문장 번호 입력 후 ENTER를 누르면 해당 문장 전체가 팝업 형태로 제시되고, 작업자는 이 중 선행어를 선택하여 자동으로 값을 입력할 수 있다. 동일 지시 성분이 여러 개인 경우에는 복원 후보 목록이 제시되며, 작업자가 적절한 선행어를 선택하면 기호 ‘+’로 이어진 형태가 선행어(ant\_form) 셀에 자동 입력된다. 선행어가 없는 경우에는 ‘선행어 없음’을 선택하면 ‘#’이 자동 입력된다. restored\_form과 restored\_type은 작업자가 직접 기입하는 셀로, form에는 복원 성분의 형태를, type에는 주어·목적어·보어·부사어 중 하나를 한글로 기입한다. type은 문서 저장 시 자동으로 영문으로 변경되도록 저장 옵션을 추가하여 작업자의 기입 오류를 최소화했다.

끝으로 하나의 서술어에 대한 복원 성분이 2개 이상인 경우에는 선행어 입력 셀에 있는 기존 작성 값 뒤, 기호 ‘/’를 입력한 후 문장 번호를 기입하는 방법으로 진행하였다.



[그림 26] 무형 대용어 복원 작업 방법

## 1.2. 도구 개선 방향

구축 작업 중 작업자 및 검수자의 입력 오류를 줄이고, 층위별 작업이 누적됨에 따라 발생하는 작업 피로도와 정합성 문제를 최소화하기 위해 한국어 통합 구조 분석 도구는 여러 기능적 개선을 거쳤다. 초기 버전에서도 입력 제약과 기본 오류 방지 기능이 제공되었으나, 실제 작업 과정에서 빈번히 요구되었던 기능을 중심으로 도구가 개선·고도화되었다.

먼저, 의미역 분석 및 무형 대용어 복원 단계에서 앞선 층위를 수정하는 경우가 자주

발생함에 따라, 수정 이력을 문장 단위로 기록할 수 있는 ‘메모 기능’을 추가하였다. 작업자가 특정 문장의 앞서 구축된 분석 단위를 수정하면 해당 문장에 메모를 남기도록 했으며, 검수자는 구축 말뭉치를 열었을 때 메모 버튼의 색상만으로 메모의 유무를 즉시 파악할 수 있도록 설계하였다. 이를 통해 작업 흐름의 투명성이 향상되었으며, 검수 과정에서 발생하던 누락·재검토 등의 부담을 줄였다.

둘째, 입력값 수정 후 이를 저장하지 않고 넘어가는 문제를 방지하기 위해 ‘테이블 값 수정 시 저장 버튼 자동 활성화 기능’을 도입하였다. 작업자가 테이블의 값을 조금이라도 변경하면 저장 버튼이 즉시 활성화되고 색상이 변화하며, 저장 완료 시 다시 기본 상태로 복귀한다. 이러한 시각적 피드백 체계는 층위별 입력량이 많은 말뭉치 구축 환경에서 실수로 인한 데이터 손실을 방지하는 데 효과적이었다.

셋째, 구어 말뭉치처럼 문장 수가 매우 많은 문서에서는 특정 문장을 다시 확인하거나 복원 지점을 찾아야 하는 일이 빈번하게 발생하였다. 이를 해결하기 위해 ‘특정 문장으로 즉시 이동 기능’을 추가하여, 문서 상단의 이동 입력창에서 문장 번호를 입력하면 해당 문장으로 바로 이동할 수 있도록 하였다. 이는 긴 문서에서 스크롤 탐색으로 발생하던 작업 시간을 크게 줄여 작업 효율을 현저히 개선하였다.

한국어 구조 통합 도구 수정 내역			
순번	구분	수정사항	작업 도구 버전
1	통합 도구	저장 오류	한국어 구조 분석_v.1.5 (25.06.05.)
2	형태소 분석	morph 기준 '+' 기호로 구분	한국어 구조 분석_v.1.5 (25.06.05.)
3	어휘의미 분석	sense_id 입력 시 자동으로 000 세자리로 변환	한국어 구조 분석_v.1.5 (25.06.05.)
4	구문 분석	Head값이 vp인 경우, 뷰어에 빨간색 화살표로 표시	한국어 구조 분석_v.1.6 (25.06.10.)
5	구문 분석	DP_Label 입력 방안 간소화	한국어 구조 분석_v.1.6 (25.06.10.)
6	구문 분석	Head 값이 교차될 경우 오류인 셀 표시	한국어 구조 분석_v.1.7 (25.06.12.)
7	통합 도구	셀 값 삭제 'delete' 버튼으로 고정	한국어 구조 분석_v.1.7 (25.06.12.)
8	무형 대응어 복원	가이드라인 창에 원문(문서 단위) 추가	한국어 구조 분석_v.1.8 (25.06.17.)

9	무형 대용어 복원	type 값 콤보박스로 추가	한국어 구조 분석_v.1.8 (25.06.17.)
10	통합 도구	원하는 문장으로 즉시 이동 기능 추가	한국어 구조 분석_v.2.1 (25.07.24.)
11	어휘의미 분석	우리말샘 홈페이지 추가	한국어 구조 분석_v.2.1 (25.07.24.)
12	통합 도구	테이블 값 수정 시, 저장 버튼 색 변경으로 표시	한국어 구조 분석_v.2.4 (25.07.31.)
13	통합 도구	의미역 분석 값 저장 오류	한국어 구조 분석_v.2.5 (25.08.01.)
14	무형 대용어 복원	선행어 추천 값 입력 행 추가	한국어 구조 분석_v.2.6 (25.08.12.)
15	의미역 분석	SRL Form 작성 방안 간소화	한국어 구조 분석_v.2.6 (25.08.12.)
16	의미역 분석	SRL Predicate form 삭제 후 lemma 복원	한국어 구조 분석_v.2.6 (25.08.12.)
17	의미역 분석	논항의 맨 마지막 word id 행에 값이 입력되는 구조로 변경	한국어 구조 분석_v.2.6 (25.08.12.)
18	의미역 분석	v2.6 수정 사항 오류 검토 및 테스트 완료 도구로 변경	한국어 구조 분석_v.2.7 (25.08.14.)
19	통합 도구	저장 오류 수정	한국어 구조 분석_v.2.8 (25.08.18.)
20	의미역 분석	의미역 태그 수정(필수 18개, 부가 23개)	한국어 구조 분석_v.3.0 (25.08.21.)
21	무형 대용어 복원	1. 선행어가 없는 경우 #으로 표기 2. restored_type 작업 도구 내 한글로 변경 3. 작업 중 문장을 가이드라인 창에 다른 색으로 표시	한국어 구조 분석_v.3.0 (25.08.21.)
22	신규 기능 추가	의미역, 무형 대용어 복원 작업 중 앞서 구축된 말뭉치 수정 시 내역을 기입할 메모 기능 추가	한국어 구조 분석_v.3.0 (25.08.21.)
23	의미역 분석	의미역 초벌 말뭉치 외 문서도 도구에서 확인 및 수정 가능	한국어 구조 분석_v.3.1 (25.08.22.)
24	의미역 분석	1. 뷰어 창 내 화살표 위치 수정 2. 특수문자가 포함된 경우 뷰어 창에 표시되지 않던 현상 수정	한국어 구조 분석_v.3.2 (25.08.26.)
25	의미역 분석	1. 부가의의미역 추가 (M-NEG)	한국어 구조 분석

		2. 메모 기능 색상 표시 추가	_v.3.4(25.09.12.)
26	의미역 분석	1. 형태소 분석 결과에 VX가 포함된 어절이 화살표의 끝점을 검은색으로 2. 뷰어에서 보이는 그림의 기본 크기 수정	한국어 구조 분석 _v.3.5(25.09.18.)
27	내부 오류	내부 오류 일부 수정	한국어 구조 분석 _v.3.6(25.09.18.)
28	통합 도구	뷰어의 가로 스크롤바 추가	한국어 구조 분석 _v.3.7(25.09.29.)
29	무형 대응어 복원	복원 값 중복 입력 시, 다른 값도 함께 삭제되는 오류 수정	한국어 구조 분석 _v.3.8(25.10.02.)
30	통합 도구	1) 행 추가, 삭제 시 어절 번호, SRL Span 값 자동 정렬 2) 1번 수정으로 인해, form이 변경된 복원어를 포함하고 있는 문장의 word form과 ant_form를 비교 후 일치하지 않는 경우, ant_word_id셀이 빨간색으로 자동 변경 3) 행 삭제 시, 삭제 행의 윗 행이 삭제 행의 id를 head값으로 가지고 있었다면 그 head값을 삭제 행의 'head-1'값으로 대체 4) DP Label의 값이 VP로 시작하는 경우 해당 셀의 배경색을 노란색으로 표시 5) 같은 ant_sen_id를 중복으로 입력할 경우 다른 값이 복사되어 들어가던 현상 수정	한국어 구조 분석 _v.4.0(25.10.17.)
31	무형 대응어 복원	ant_form에 침표(.)가 포함된 경우 내부 코드상으로 구분자로 인식하여 사라지는 현상으로 인한 오류 수정	한국어 구조 분석 _v.4.1(25.10.23.)
32	통합 도구	1) 형태소 표시 수정 불가 오류 수정 2) 무형 대응어 restored_form 입력시 구분자 양옆 공백으로 기준 변경	한국어 구조 분석 _v.4.2(25.11.18.)

<표 44> 한국어 통합 분석 도구 개선 내역

이 외에도 층위별 구축 특성을 반영한 다양한 기능이 추가되었다. 추가 및 수정된 내역은 위 <표 44>와 같다. 이와 같은 기능 개선은 궁극적으로 작업자의 부담을 줄이고 검수자의 판별을 용이하게 하여, 전체 말뭉치의 품질과 층위 간 정합성을 동시에 높이는 효과를 가져왔다.

### 1.3. 기타 도구 개발 내역

본 사업에서는 층위별 구축 도구의 안정성과 작업 효율을 향상시키기 위해, 정규 기능 외에도 여러 보조 도구를 개발하였다. 구축 과정에서는 도구 업데이트 및 수정 요청이 빈번하게 발생할 수 있으며, 이러한 변경이 작업 일정에 영향을 미치지 않도록 작업 지연 최소화 전략을 함께 마련하였다.

먼저, 작업자가 1차로 구축한 말뭉치를 엑셀 형식으로 변환하여 전반적인 분석값을 검토하고 수정한 뒤, 해당 결과를 구축 말뭉치에 반영하는 XLXS-JSON 기능을 추가한 ‘말뭉치 관리용 관리자 도구(admin)’를 개발하였다. 관리자 도구(admin)는 작업물을 관리하는 전문가 1인만 사용할 수 있도록 제한하여 말뭉치 관리 및 반영에 오류가 발생하지 않도록 했다. 이를 통해 한국어 구조 통합 분석 도구 내부 수정이 필요한 기간에도 구축 작업자는 엑셀 파일 내용을 검토하는 등 지속적으로 작업을 진행할 수 있는 환경을 구축하였다.

이외 말뭉치 검수 및 반영 과정에서 요구되는 구조적 점검(JSON 스키마 검증, 누락 필드 탐지, word\_id 일관성 확인, MP·WSD·DP 간 교차 오류 점검 등)을 자동화하기 위해 관리자 도구(admin)에 기능을 추가하였다. 최종적으로 사용 가능한 도구는 아래 <표 45>와 같다.

기능	설명
일괄 검수 기능	형태소·구문·의미역·무형 대용어 복원 등 각 층위의 JSON 결과를 자동 검사하여 오류를 탐지.
엑셀 변환 및 반영 기능	작업자가 제출한 XLSX 결과를 JSON 구조로 재변환하여 도구에 일괄 반영.
구어/문어 말뭉치 원문 비교 기능	원문 손실 또는 정제 오류 탐지.
WSD 및 DP 추출·반영 기능	특정 층위만 선별적으로 재처리할 수 있도록 분리된 기능 제공.
오류 검출 리포트 생성	누락 표지, word_id 불일치, label 오류 등 자동 리포트 제공.

<표 45> 말뭉치 관리용 관리자 도구(admin) 기능

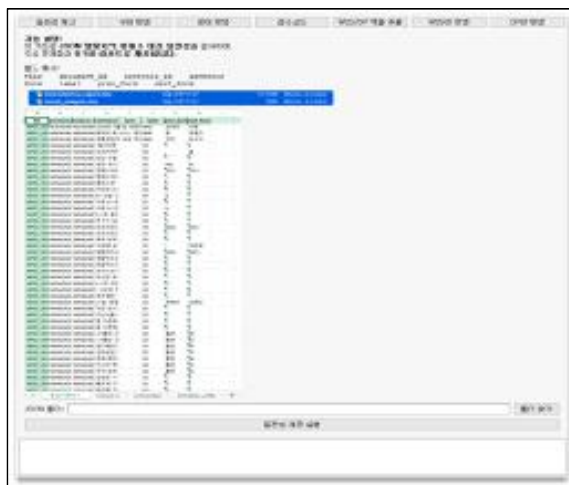
이러한 기능은 개별 문서 형태로 구성된 작업 말뭉치(json)의 일반적인 작업 오류나 일괄 수정이 필요한 표지 등을 수정 및 반영할 때 즉시 엑셀 파일을 생성하여 검수자의 수정 및 검수를 거친 뒤 json 말뭉치에 반영하는 방식으로 설계되었으며, 관리자가 대규모 말뭉치 파일을 효율적으로 검수·반영할 수 있게 하였다. 특히 사업 후반에는, 말뭉치 규모가 커짐에 따라 구축 결과의 일관성과 정확성을 유지하기 위해 “최종 반영 여부 검수 기능”이 추가되었으며, 이는 JSON과 XLSX 간 상호 검증을 통해 반영 누락을 방지하는 역할을



수행하였다.

말뭉치 관리용 관리자 도구의 경우 실시간으로 기능을 추가하기 어렵다는 제한이 있어, 의미역 분석 및 무형 대용어 복원 작업의 경우 검수용 파일 생성 기능을 웹앱으로 구축하였다. 웹앱은 'streamlit'으로, 의미역 분석(SRL)의 불필요 값 삭제, 표지 변경, 검수용 엑셀 파일 생성 기능을 추가하여 관리자에게 전달되었다. 의미역 분석과 무형 대용어 복원 값은 앞서 구축된 한국어 구조 분석 단위와 복잡하게 얽혀 있어, 검수용 엑셀 파일은 도구 수정 시 참고용 파일로만 활용되었다. 추후 최종 말뭉치 수정 시에는 다수의 의미역 분석 표지 수정만 말뭉치에 일괄 반영되도록 기능을 추가하였다.

이와 같은 다양한 보조 도구 개발은 한국어 구조 통합 도구의 안정성을 강화하고 구축-검수-반영의 전체 주기를 효율화하여, 결과적으로 말뭉치 구축의 품질과 생산성을 크게 향상시켰다.



[그림 27] 말뭉치 관리용 관리자 도구(admin) 화면



[그림 28] 관리자용 웹앱 화면

## 1.4. 기타 정보 추출 문서 목록

그밖에 말뭉치 정제, 구축 및 검수 과정에서 필요하여 별도로 추출한 파일과 문서의 목록은 아래 <표 46>과 같다.

추출 유형	내용
매핑테이블	- 사업 대상 원시 말뭉치와 정제된 작업 말뭉치 간 참조 관계

	<p>확인을 위한 파일</p> <ul style="list-style-type: none"> <li>- 파일 id와 문장 id 단위 매핑테이블, 문장 전체 혹은 부분 사용 id 매핑테이블, 부분 문장 id 부여 매핑테이블로 구성됨</li> </ul>
우리말샘 등재 단어 목록	<ul style="list-style-type: none"> <li>- 복합 명사, 고유명사 형식 수정 비교 대상으로 사용됨</li> <li>- 약 38만개의 단어 목록 파일</li> </ul>
우리말샘-구축 말뭉치 매칭 목록	<ul style="list-style-type: none"> <li>- 우리말샘 등재 단어 목록을 샷갓표(^) 포함 단어와 띄어쓰기 포함 단어로 분리하여 분석 말뭉치 단어와 비교한 결과를 추출함</li> </ul>
형태소 분석 수동 수정 대상 목록	<ul style="list-style-type: none"> <li>- 형태소 초벌 말뭉치 내부 VV, VA 분리 분석 오류 목록 생성 후 일괄 수정</li> <li>- 형태소 분석 말뭉치 초벌에 사용된 모델의 내부 구조도 변경</li> </ul>
구문 분석 오류 및 VP to VP_AJT 변경 목록	<ul style="list-style-type: none"> <li>- 구문 분석 초벌 말뭉치 내 금년도 구문 분석 표지 이외 라벨 부착 오류 발견(EEVB, XV, XNVB 등)</li> <li>- 작업 중 다수 발견된 일괄 VP_AJT 목록(4주차, 약 500건)</li> <li>- 초벌 말뭉치 후처리에 위 오류들을 검토 및 수정하는 과정 추가</li> </ul>
의미역 모형 말뭉치	<ul style="list-style-type: none"> <li>- 국립국어원 의미역 분석 말뭉치(NIKL_CF_v1.0_XML)의 3,660개 파일 대상</li> <li>- word/sense_num/constituent/arg_num/arg_role/arg_josa 순서로 엑셀 파일 추출 후 구축 및 검수 작업에 참고용으로 활용함</li> </ul>
word 오류 목록	<ul style="list-style-type: none"> <li>- 원문의 word 단위</li> </ul>
일관성 검사 결과	<ol style="list-style-type: none"> <li>1. 형태소 분석(morph_mismatch) <ul style="list-style-type: none"> <li>- word form은 동일하나, 형태소 분석값이 다른 어절 추출 목록</li> </ul> </li> <li>2. 의미역 분석(srl_form_mismatch) <ul style="list-style-type: none"> <li>- srl_span 값을 가진 word id를 합친 form과 srl_form이 다른 문장 추출 목록</li> </ul> </li> <li>3. 무형 대용어 복원(za_form_mismatch) <ul style="list-style-type: none"> <li>- za_form이 word와 다른 문장 추출 목록</li> </ul> </li> </ol>
form-original_form 비교 목록	<ul style="list-style-type: none"> <li>- delete -&gt; form과 original_form을 비교 결과, form에 _ 혹은 단어가 추가된 경우</li> <li>- insert -&gt; form과 original_form을 비교 결과, original_form 일부가 삭제된 경우(임의 삭제 내용인지 확인 필요)</li> <li>- replace -&gt; form과 original_form을 비교했을 때 같은 자리에 다른 값이 존재할 경우(임의 수정 내용 혹은 오류인지 확인 필요)</li> </ul>
의미역 label 누락 및 불일치 목록	<ul style="list-style-type: none"> <li>- srl label 값이 누락된 문장 목록</li> </ul>
의미역 predicate lemma 검사 결과	<ul style="list-style-type: none"> <li>- SRL predicate lemma 값에 임의로 '+다'를 넣어, expected_lemma_form을 생성하고 구축 말뭉치의 predicate lemma form과 비교하여 일치하지 않는 경우를 추출한 목록</li> </ul>

<표 46> 기타 정보 추출 문서 목록

## 2. 말뭉치 성능 평가

품질 검증 단계는 ‘품질 검수’와 ‘품질 검증’ 과정으로 진행한다. 전문가에 의한 품질 검수를 거친 작업 결과는 전문 구축 인력에 의해 구축된 정답 말뭉치와의 비교를 통해 성능 검증 과정을 거쳐 F1 0.95~0.99 이상의 평균 성능 점수를 나타내는 최종 구축 품질을 확보했다.

### 2.1. 평가 방안 - 참조 기반 평가

본 사업의 구축 말뭉치는 총 30만 어절 규모로 구축 말뭉치의 품질 평가를 위해 이 중 10% 이상(약 3만 어절)을 무작위 추출하여 전문가가 직접 검수를 진행하였다. 최종 평균 성능 0.95~0.99 달성을 목표로 한다.

문어 말뭉치	
형태소 분석	0.99
어휘의미 분석	0.99
구문 분석	0.97
의미역 분석	0.98
무형 대용어 복원	0.99

구어 말뭉치	
형태소 분석	0.99
어휘의미 분석	0.99
구문 분석	0.95
의미역 분석	0.98
무형 대용어 복원	0.98

<표 47> 문어 말뭉치 성능 평가(F1) 결과    <표 48> 구어 말뭉치 성능 평가(F1) 결과

### 2.2. 평가 방안 - 다층 분석 말뭉치 간 일관성 검증 방안

형태소 분석, 어휘의미 분석, 구문 분석, 의미역 분석 및 무형 대용어 복원으로 진행해 가는 과정에서 서로 다른 분석 층(layer)의 분석 결과가 서로 논리적으로 맞지 않거나, 불일치하여 생기는 문제를 찾고 수정하기 위해 다음과 같이 다층 분석 말뭉치 간 일관성 검증을 진행하였다.

1단계는 규칙 기반(Rule-based) 일관성 검증과 오류 패턴 사전 활용 오류 자동 탐지로 진행되었다. 앞선 관련 연구들을 분석하여 각 분석 계층의 결과 간 관계에서 나타나는 대표적인 오류 패턴을 미리 규칙으로 정의한 뒤, 규칙과 어긋난 사례를 자동 탐지하여 오류로 분류, 이를 관리자가 확인하여 수정을 진행하였다. 말뭉치 구축 과정에서 발견된 오류 패턴을 데이터베이스(DB)화하여 사전 형태로 관리하고, 이를 바탕으로 말뭉치의

일관성을 검증하였다. 작업 중 애자일(Agile) 방식으로 짧은 주기마다 주기적으로 오류를 탐지하고 수정을 진행함으로써 작업 말뭉치 간 일관성을 유지하였다.

2단계는 자동 정합성 평가와 자동 일관성 평가이다. 형태소 분석과 구문 분석 및 의미역 분석의 정보를 결합한 의존 관계 구조(Dependency Tree)를 생성하여 각 층위 간 관계가 논리적으로 적절히 연결되는지를 자동적으로 평가하였다. 전문가가 미리 정합성을 판단한 데이터(정답 말뭉치)를 구축하고 이를 기반으로 딥러닝 분류 모델(BERT, RoBERTa 등 Transformer 기반)을 학습하여, 각 분석 층위 간 관계가 적절한지 자동으로 판별하는 모델을 구축하고 정확한 사례와 오류 사례를 충분히 학습시켜 최종 구축된 데이터에 대해서는 자동 일관성 평가를 진행하였으며 평가 결과는 아래 [표 49]의 목록으로 추출하여 검수팀의 최종 검수에 활용되었다.

구분	내용
원문 정제	<ol style="list-style-type: none"> <li>1. 띄어쓰기 반영 내역.xlsx <ul style="list-style-type: none"> <li>- 문어와 구어 말뭉치와 우리말샘 매칭 목록 중 띄어쓰기 포함 단어 반영 목록</li> </ul> </li> <li>2. 삿갓표 반영 내역.xlsx <ul style="list-style-type: none"> <li>- 문어와 구어 말뭉치와 우리말샘 매칭 목록 중 삿갓표 포함 단어 반영 목록</li> </ul> </li> <li>3. 우리말샘 매칭 목록.xlsx <ul style="list-style-type: none"> <li>- 우리말샘 사전과 매칭된 단어 전체 목록</li> </ul> </li> <li>4. 구어/문어_form_original_form 비교 <ul style="list-style-type: none"> <li>- delete : form과 original_form을 비교 결과, form에 _ 혹은 단어가 추가된 경우</li> <li>- insert : form과 original_form을 비교 결과, original_form 일부가 삭제된 경우</li> <li>- replace : form과 original_form을 비교했을 때 같은 자리에 다른 값이 존재할 경우</li> </ul> </li> </ol>
한국어 구조 분석	<ol style="list-style-type: none"> <li>1. 형태소 분석(morph_mismatch) <ul style="list-style-type: none"> <li>- word form을 기준으로, 어절 단위 형태소 분석값이 상이한 경우 추출 목록</li> </ul> </li> <li>2. 구문 분석(dp_mismatch) <ul style="list-style-type: none"> <li>- dp_label 오류, 교차 오류, head 오류 추출 목록</li> </ul> </li> <li>3. 의미역 분석(srl_form_mismatch) <ul style="list-style-type: none"> <li>- srl_form이 sentence &gt; word_id를 합친 word와 다른 경우 추출 목록</li> </ul> </li> <li>4. 의미역 분석(srl_predicate_lemma) <ul style="list-style-type: none"> <li>- srl_predicate의 lemma 값이 form과 맞지 않는 경우 추출 목록</li> </ul> </li> </ol>
무형 대용어 복원	<ol style="list-style-type: none"> <li>1. 무형 대용어 복원(za_form_mismatch) <ul style="list-style-type: none"> <li>- za_form이 word와 다른 경우를 추출한 목록</li> </ul> </li> </ol>

<표 49> 일관성 검사 목록 및 내용

## 2.3. 품질 관리 및 검증 체계

본 사업의 품질 관리는 자동 검증과 인간 검수의 이중 구조로 설계하였다. 먼저 구축 완료된 각 층위 말뭉치에 대해 형식 검증 스크립트를 적용하여 JSON 스키마 위반, ID 불일치, 누락 필드 등 구조적 오류를 점검하고, 층위 간 일관성(예: 구문 분석의 논항 구조와 의미역 분석의 논항 구조 일치 여부, 무형 대용어 복원 결과와 주어·목적어 태그의 정합성 등)을 자동 확인하였다. 이후 층위별 책임 연구원이 무작위 표본과 문제 사례를 중심으로 정밀 검수를 수행하여 분석 지침에 대한 준수 여부와 해석상의 일관성을 추가로 점검하였다.

다층위 구조 분석 말뭉치의 경우 형태소-어휘의미-구문-의미역-무형 대용어 복원 정보가 단일 문장 ID를 공유하므로, 층위 간 상호 제약을 활용한 검증이 가능하다. 예를 들어, 의미역 분석에서 AGT로 태깅된 성분이 구문 분석에서 SBJ로 일관되게 분석되었는지, 무형 대용어로 복원된 주어가 해당 서술어의 논항 구조 및 의미역 체계와 충돌하지 않는지 등을 자동 스크립트로 점검하였다. 또한 문어·구어 자료 각각에 대해 일정 비율의 문장을 표본 추출하여 무형 대용어 복원 정확도를 수동 평가하고, 오류 유형을 분류함으로써 향후 복원 규칙 및 도구 개편의 근거를 마련하였다.

작업 형식 오류 목록은 중간 납품 결과물을 대상으로 한 국립국어원의 검토 의견을 수합하였다. 검토 의견을 바탕으로 수정한 내용이며, 수정 중 발견된 기타 오류 또한 기계적 추출-검수자 확인 및 수정-기계적 검토 과정을 거쳤다.

## 2.4. 기타 작업 형식 오류 목록

국립국어원에 제출한 중간 납품물에 대한 검토 의견을 바탕으로 수행한 말뭉치 형식 수정 내역을 정리하였다. 1차 수정에서는 파일명, json 형식, 데이터 구조 등 기본적인 형식 수정과 JSON 구조 전반의 일관성을 확보하는 데 중점을 두었으며, 2차 수정에서는 원문 정제 오류, 형태소·어휘의미·구문·의미역·무형 대용어 복원 층위별 정합성 문제를 세부적으로 검토하여 정비하였다.

일자	내용
25.08.25.	1. 파일명 형식 수정 - NX00, SX00 구성이 되도록 수정(예: NXMP2502509081.json) 2. json 형식 수정

	<ul style="list-style-type: none"> <li>- word_id에서 datatype이 str, int 혼용을 모두 int로 통일</li> <li>- '250827_한국어 구조 json 형식 v1.0' 작성 및 전달</li> </ul> <ol style="list-style-type: none"> <li>1. 불필요한 정보 미제거 <ul style="list-style-type: none"> <li>- 신문 말뭉치 내부, 형태소 분석 초벌 결과('expresso_compounded_norm') 제거</li> </ul> </li> <li>3. paragraph 누락 <ul style="list-style-type: none"> <li>- 구조 분석 말뭉치 중 신문 말뭉치 일부에서 누락된 paragraph 단락 일괄 수정</li> </ul> </li> <li>4. 형태소 분석값 누락 <ul style="list-style-type: none"> <li>- 형태소 분석 중 값이 누락된 목록을 일괄 추출 후 검수팀 공유</li> </ul> </li> <li>5. 어휘의미 분석 데이터 구조 수정 <ul style="list-style-type: none"> <li>- 19년도 결과물 참고하여, pos값 추가</li> <li>- sense_id 문자열 형식("001", "002") 그대로 유지</li> </ul> </li> <li>6. 단어 내부 '_' 부착 기준 검수 <ul style="list-style-type: none"> <li>- 우리말샘 등재 단어 목록 재추출 후 작업 대상 말뭉치와 비교 검수</li> </ul> </li> <li>7. 구문분석 최하단 말미 head 값 수정 <ul style="list-style-type: none"> <li>- "head: -1"로 일괄 수정</li> </ul> </li> </ol>
25.10.22.	<ol style="list-style-type: none"> <li>1. 원문 정제 <ul style="list-style-type: none"> <li>- 원문 정제 중 단어의 잘림 현상 및 중복 기재 오류 목록 추출 및 검수</li> <li>- 우리말샘 등재 단어와 말뭉치 내 단어 비교 목록 추출</li> </ul> </li> <li>2. 의미역 분석 말뭉치 json 형식 <ul style="list-style-type: none"> <li>- word_id라는 동일 key에 대해 list와 int가 혼재하는 형식을 list로 통일</li> <li>- predicate의 경우 단일한 데이터로, list 밖으로 형식 수정</li> </ul> </li> <li>3. 의미역 분석 말뭉치 <ul style="list-style-type: none"> <li>- form과 lemma가 일치하지 않는 오류, lemma word_id와 form을 비교 결과 추출</li> <li>- label 누락 및 오타, predicate 및 lemma 누락 목록 추출 후 검수</li> </ul> </li> <li>4. 무형 대용어 복원 말뭉치 <ul style="list-style-type: none"> <li>- ZA 층위의 경우 한국어 구조 분석 말뭉치 4종과 동일하게 sentence 내부에 들어있도록 json 형식 수정. (251208_한국어 구조 json 형식 v1.7.hwp)</li> </ul> </li> </ol>

<표 50> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 형식 오류

### 3. 작업자 관리 및 의사소통

#### 3.1. 작업자 관리

##### 3.1.1. 작업자 모집

한국어 구조 분석 말뭉치 구축 사업은 층위별 한국어 구조 분석 말뭉치 5종(형태 분석 말뭉치, 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치와 무형 대용어 복원 말뭉치) 구축을 목표로 한다. 작업 지연을 최소화하기 위해 작업 시작 전 작업자 교육을 실시하였다. 한국어 구조 분석 작업에 참여한 작업자의 선정 기준과 작업 내용은 다음과 같다.

사업 내용	<ul style="list-style-type: none"> <li>국립국어원 신문 말뭉치 및 일상 대화 말뭉치를 대상으로 하여 형태 분석, 어휘의미 분석, 구문 분석, 의미역 분석 및 무형 대용어 복원 말뭉치 구축</li> </ul>
지원 자격	<ul style="list-style-type: none"> <li>국어학 관련 석박사 과정생, 수료생, 졸업생</li> </ul>
작업 내용	<ul style="list-style-type: none"> <li>A팀: 문어 또는 구어 문장에 대한 구문 분석 및 무형 대용어 복원</li> <li>B팀: 문어 또는 구어 문장에 대한 어휘의미 분석 및 의미역 분석</li> </ul>

<표 51> 한국어 구조 분석 말뭉치 작업자 선정 기준

한국어 구조 분석 말뭉치 구축을 위한 작업자는 국어학 관련 석·박사 과정생, 수료생, 졸업생을 대상으로 모집하였다. 모집 공고 단계에서는 한국어 구조 분석 말뭉치 구축 사업의 지원 자격과 주요 업무, 요구 역량을 명확히 안내하였다. 팀 구성은 작업 내용, 경력, 이전 유사 사업 경험을 종합 검토하여 이루어졌으며, 지원자가 희망하는 팀이 있거나 경험 이력이 두 팀 중 하나의 성격과 명확히 부합하는 경우 해당 팀으로 우선 배치하였다. 최종 선발은 성실히 참여 가능한 지원자를 중심으로 확정되었으며, 그 결과 총 22명의 작업자가 A팀과 B팀으로 구분되어 선발되었다.

작업자	부합 조건
A01	국어국문학과 박사졸업
A02	국어국문학과 석사과정

A03	국어국문학과 박사졸업
A04	국어국문학과 석사과정
A05	국어국문학과 석사과정
A06	국어국문학과 박사수료
A07	국어국문학과 석사과정
A08	국어국문학과 석사과정
A09	국어국문학과 석사과정
A10	국어국문학과 박사수료
A11	국어국문학과 석사과정
A12	국어국문학과 박사졸업
B01	국어국문학과 석사졸업
B02	국어국문학과 박사수료
B03	국어국문학과 박사과정
B04	한국어교육학과 박사졸업
B05	국어국문학과 박사과정
B06	국어국문학과 박사수료
B07	국어국문학과 박사수료
B08	국어국문학과 석사과정
B09	국어국문학과 석사과정
B10	국어국문학과 석사졸업

<표 52> 한국어 구조 분석 말뭉치 작업자 명단

### 3.1.2. 작업자 교육

각 작업을 본격적으로 시작하기 전, 작업자들이 문어 및 구어 문장에 대한 구문 분석, 어휘의미 분석, 의미역 분석, 무형 대용어 복원 과업을 정확히 이해하고, 구축 도구의 기본 사용 방법을 숙지할 수 있도록 작업 시작 전 교육을 실시하였다. 지침 기반 교육은 지침 작성을 담당한 한림대 전문가가 주도하였으며, 전체 사업의 수행 일정은 한림대 관리자가 안내하였다. 구축 작업 도구의 사용 방법에 대한 교육은 도구를 개발·운영한 데이터리 관리팀이 담당하여 진행하였다. 관리자는 주기적으로 작업물의 성능을 점검하고, 성능이 미흡한 경우에는 즉시 재교육을 진행하였고, 지침, 도구의 변경 사항에 대해서는 수시로 공지하여 작업자들이 항상 업데이트된 도구를 사용하여 지침의 내용을 반영할 수 있도록 했다. 형태 분석, 구문 분석, 어휘의미 분석, 의미역 분석, 무형 대용어 복원 작업의 구체적 교육 일정은 아래와 같다.



No.	일자	교육 내용
1	2025.04.29.	• 형태 분석 지침 및 도구 교육
2	2025.06.02.	• 어휘의미 분석 지침 및 도구 교육
3	2025.06.04.	• 구문 분석 지침 및 도구 교육
4	2025.06.09.	• 어휘의미 분석 지침 교육(2차)
5	2025.06.11.	• 구문 분석 지침 교육(2차)
6	2025.08.18.	• 무형 대용어 복원 지침 교육
7	2025.08.19.	• 의미역 분석 지침 교육
8	2025.09.01.	• 의미역 분석 지침 교육(2차)
9	2025.09.02.	• 무형 대용어 복원 지침 교육(2차)

<표 53> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 교육 일정

층위별 한국어 구조 분석 말뭉치 4종 및 무형 대용어 복원 말뭉치 구축 1차 교육에서는 지침에 대한 전반적인 설명과 도구 사용 방법을 다루었고, 구글 폼을 이용한 실습을 통해 작업자들의 작업 이해도를 확인했다.

2차 교육에서는 작업 진행 이후 실습 경험을 기반으로 설계되었다. 즉, 초기 지침 교육에서 다루었던 내용을 단순히 반복하는 것이 아니라, 실제 작업 단계에서 나타난 개별 질의와 해석 차이 등을 중심으로 구성하였다. 특히, 1차 교육 후 말뭉치 구축을 진행하는 과정에서 수집된 질의응답 사례를 예시 자료로 활용함으로써, 해석에 대한 혼란을 줄이고 일관적인 적용 방향을 마련하고자 했다. 결과적으로, 2차 교육은 실제 작업 단계에서 접한 문장을 함께 논의함으로써, 해석 차이를 상호 점검하고 기준을 조정하여 정확성을 확보하고 품질을 안정적으로 유지하는 데 목적을 두었다.

### 3.1.3. 작업자별 정확률 제고

구문 분석과 어휘의미 분석이 종료된 후, 의미역 분석 및 무형 대용어 복원으로 전환되는 시점에서, 검수자 분석 결과와 대조하여 작업자 분석의 정확률을 확인하는 과정을 4주간 운영하였다. 정확률 확인은 작업자가 분석한 결과와 그 내용을 점검한 검수자가

산출한 결과를 비교하는 방식으로 설계하였다. 특히, 의미역과 무형 대용어는 분석 범위나 내용이 작업자별로 달라지는 경우가 빈번하다는 점을 고려하여, 결과 비교를 통해 오류를 확인하고 판단 기준의 균형을 맞추는 과정 자체에 평가 목적을 두었다. 이러한 비교 평가는 작업자 개인의 수행 역량을 서열화하기 위한 목적이 아니라, 이해도 격차를 조기에 탐지하고, 이후 구축 단계의 안정성과 정확도를 확보하기 위한 방향으로 진행되었다. 결과적으로 정확률 확인은 해석 편차를 줄이고, 작업자별 오류 개선을 판단과 개인 피드백 제공의 근거를 마련하는 체계적 점검 절차로 운영되었다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
file_name	doc_id	sent_id	sentence	word_id	word	morph	WSD Form	head	DP Label	SRL Span	SRL Label	predicate	Loan from	label from	le Lamma	IL Span	VL Label	Uccate Lam	작중
SR02_SDR'SDRW240XSDRW240X	너는	요즘	01	너는	너/NP + 001/001	5 NP_SBI	1	MAG	5/줄기다	1	AGT	5/줄기다	0	X	0	0	X		
SR02_SDR'SDRW240XSDRW240X	너는	요즘	02	요즘	요즘/NING 요즘/001	5 NP_AIT	2	M-TMP	5/줄기다	2	M-TMP	5/줄기다	0	0	0	0	0	0	0
SR02_SDR'SDRW240XSDRW240X	너는	요즘	03	어떤	어떤/MMD	4 DP													
SR02_SDR'SDRW240XSDRW240X	너는	요즘	04	취미	취미/NING 취미/007	5 NP_OBI	3, 4	THM	5/줄기다	3, 4	THM	5/줄기다	0	0	0	0	0	0	0
SR02_SDR'SDRW240XSDRW240X	너는	요즘	05	즐거	즐거/VV + 즐거/002	6 VP													
SR02_SDR'SDRW240XSDRW240X	너는	요즘	06	있어	있/VX + 002/021	0 VP													
SR02_SDR'SDRW240XSDRW240X	너는	전파	07	어	어/VC	2 IP													
SR02_SDR'SDRW240XSDRW240X	너는	전파	08	나는	나/NP + 001/003	6 NP_SBI	1, 2	THM	5/있다	1, 2	LOC	6/있다	0	X	0	0	X		X
SR02_SDR'SDRW240XSDRW240X	너는	전파	09	전파	전파/NING 전파/001	4 NP_AIT	3	M-MOD	4/오래되다	3	M-MOD	4/오래되다	0	0	0	0	0	0	0
SR02_SDR'SDRW240XSDRW240X	너는	전파	10	오래	오래/VV 오래되/00	5 VP_MOD													
SR02_SDR'SDRW240XSDRW240X	너는	전파	11	취미	취미/NING 취미/007	6 NP_SBI	3, 4, 5	FTR	6/있다	3, 4, 5	THM	6/있다	0	X	0	0	X		X

[그림 29] 작업자에게 전달된 오류 리포트 화면

위의 내용은 실제 작업자에게 전달된 오류 리포트이다. 작업자는 본인의 정확률을 점검하는 동시에, 검수자가 수정한 주석을 확인하며 오류를 조정할 수 있었고, 이 과정에서 같은 실수를 줄이기 위한 기준을 스스로 정비하였다. 또한 검수자는 작업 과정에서 반복적으로 확인된 오류 유형을 작업자에게 정리하여 전달할 수 있었다.

연번	작업자 번호	검수 결과			정확률
		O	X	합	
1	04	1404	356	1760	80%
2	05	1638	411	2049	80%
3	06	1590	352	1942	82%
4	08	1394	583	1977	71%
5	09	1556	210	1766	88%
6	10	1112	950	2062	54%
7	11	1277	537	1814	70%
8	12	1413	615	2028	70%
9	13	1380	483	1863	74%
10	14	1577	427	2004	79%

[그림 30] 정확률 대조 확인 화면

연번	작업자 번호	정확률			
		1주차	2주차	3주차	4주차
1	04	88%	93%	91%	80%
2	05	77%	80%	91%	80%
3	06	77%	76%	91%	82%
4	08	58%	82%	81%	71%
5	09	81%	90%	90%	88%
6	10	9%	37%	43%	54%
7	11	64%	65%	60%	70%
8	12	76%	67%	51%	70%
9	13	89%	91%	87%	74%
10	14	-	83%	85%	79%

[그림 31] 의미역 분석의 주차별 정확률 비교 결과

주 차별 정확률 비교 결과, 초기 구어 말뭉치 구축 단계에서는 비교적 높은 정확률을 보였으나, 문어 말뭉치로 전환된 4주 차에는 정확률이 다소 낮아지는 경향이 확인되었다. 결국 데이터의 특성이 정확도에 직접적인 영향을 미친다는 점을 실증적으로 확인할 수 있는 의미 있는 결과였다. 이러한 비교 점검 과정에서 작업자들은 본인의 분석 결과와 검수 결과, 그리고 이를 기반으로 산출된 정확률을 확인하여 분석 판단 기준을 정비하고 일관성을 강화하였다. 또한 정확률이 낮은 작업자의 경우 개별 추가 교육이 진행되었으며, 이를 통해 과업의 이해도를 보완하고 정확률을 향상시킬 수 있었다. 아래는 의미역 분석과 무형 대용어 복원 개별 교육을 진행한 내역이다.

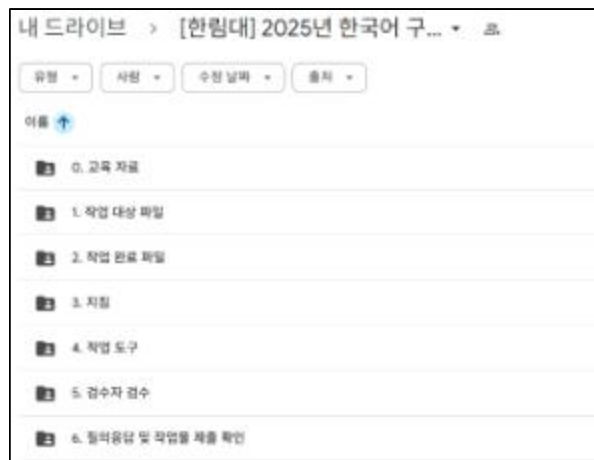
No.	일자	교육 내용
1	2025.09.10.	• 의미역 분석 추가 교육 1차
2	2025.09.19.	• 무형 대용어 복원 추가 교육
3	2025.09.29.	• 의미역 분석 추가 교육 2차

<표 54> 의미역 분석 및 무형 대용어 추가 교육 일정

## 3.2. 구글 드라이브 운영

### 3.2.1. 작업 지침 및 교육 자료 저장

한국어 구조 분석 사업의 작업 지침, 교육 자료, 작업물 배포와 수합은 접근 편의성 확보를 목적으로 구글 드라이브로 운영되었다. 자료 저장은 지정된 폴더에 작업 자료를 업로드하는 방식으로 이루어졌으며, 사업 일정 공지, 질의응답 공유, 구축 도구 안내, 최종 파일 수합 또한 구글 드라이브를 통해 통합적으로 진행되었다. 구체적인 저장 구조는 아래 그림과 같다.



[그림 32] 구글 드라이브 화면

### 3.2.2. 작업 관련 이슈 관리

작업 관련 회의는 매주 정기적으로 진행되었으며, 회의에서 진행 상황과 검수 내용 등을 논의하였다. 검수 논의 과정에서 작업자에게 즉시 안내가 필요한 내용이 도출된 경우, 일정 지연을 방지하고 작업 기준의 일관성을 확보하기 위해 단체 채팅방에 즉각적으로 공지하였다. 채팅방은 작업 일정 안내, 긴급 공지 등 업무 소통 전반에 활용되었으며, 이를 통해 작업자 간 실시간 소통을 보장하고, 정보 전달의 신속성과 작업 일정 준수를 확보하였다.

또한 작업 과정에서 발생한 질문과 추가 확인이 필요한 사항은 질의응답 시트에 기록

되어 관리하였다. 작업자가 질의응답 시트에 질문을 등록하면, 지침을 담당하는 한림대 지침 담당자가 응답하는 방식으로 운영되었다.

작업 도구 개선이 필요한 경우에는 도구 수정 시트를 별도로 운영하여 수정 요청 사항을 등록하였다. 도구 개발과 업데이트는 데이터리에서 수행하였으며, 시트에 등록된 내용을 확인한 뒤 도구 수정을 진행하였다. 수정된 도구는 구글 드라이브에 업로드하였고, 이를 공지함으로써 최신 작업 도구 환경에서 실시간 구축 작업을 지속할 수 있도록 하였다.

2. 형태 분석 질의응답
3. 구문 분석 질의응답(A형)
4. 어휘의미 분석 질의응답(B형)
5. 우형대응어 복원 질의응답(A형)
6. 의미역 분석 질의응답(B형)
7. 원문 수정 및 작업 도구 관련 질의응답(A, B형)

[그림 33] 질의응답 시트 관리

일차	일련번호	질문자	질문 내용	답변	추가 설명	답변자	답변일
0	5.28	김도연	1. 나는 2. 무엇을 3. 할지 4. 할지	이 문장에서 '나는' 가 주어, '무엇을' 가 목적어, '할지' 가 목적어, '할지' 가 목적어	이 문장에서 '나는' 가 주어, '무엇을' 가 목적어, '할지' 가 목적어, '할지' 가 목적어	김도연	5.28
1	6.7	김도연	1. 목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	김도연	6.7
2	6.7	김도연	1. 목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	김도연	6.7
3	6.7	김도연	1. 목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	목적어 NP_OBJ 2. 주어 VP 3. 목적어 VP 4. 목적어 NP_O	김도연	6.7

[그림 34] 질의응답 시트 내용

### 3.3. 검수 관리

한림대학교는 지침 정비와 일관성 점검을 위해 정기적인 검수 회의를 진행하였다. 다음 표는 주요 회의 일자를 정리한 것이다.

No.	일자	주요 내용
1	2025.04.29.	• 원문 정제 및 형태 분석 말뭉치 구축 일정 조율
2	2025.05.08.	• 기존 형태 분석 말뭉치 구축 지침 검토
3	2025.05.13.	• 형태 분석 작업 진행 상황 보고 및 문제점 논의
4	2025.05.27.	• 1주차 형태 분석 작업 결과 검토
5	2025.06.04.	• 2주차 형태 분석 작업 결과 개인별 보고
6	2025.06.10.	• 어휘의미, 구문 실습 결과 검토
7	2025.06.17.	• 어휘의미, 구문 2차 교육 검토
8	2025.06.24.	• 구문 분석 및 어휘의미 분석 말뭉치 구축 진행 상황 검토
9	2025.07.01.	• 검수자 회의 진행을 통한 지침 보완
10	2025.07.08.	• 검수 결과 문제 작업자 및 개인 피드백 대상자 선별
11	2025.07.15.	• 검수자 회의를 통해 파악된 빈출 오류를 통한 지침 보완
12	2025.07.22.	• 의미역 분석 지침 초안 검토 및 수정 사항 논의
13	2025.07.28.	• 구문 분석 및 어휘의미 분석 말뭉치 구축 종료 상황 검토
14	2025.08.06.	• 무형 대응어 복원 말뭉치 구축 지침 초안 검토 및 수정 사항 논의
15	2025.08.12.	• 의미역 분석 및 무형 대응어 복원 교육 준비
16	2025.08.26.	• 의미역 분석 및 무형 대응어 지침 초안 수정 사항 논의
17	2025.09.03.	• 의미역 분석 및 무형 대응어 지침 수정 사항 논의
18	2025.09.10.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유 및 주요 문제점 처리 방안 논의
19	2025.09.17.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유 및 주요 문제점 처리 방안 논의
20	2025.09.24.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유 및 주요 문제점 처리 방안 논의
21	2025.10.01.	• 지침 수정 필요 사항 검토 및 작업자 주의사항 전달
22	2025.10.15.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유
23	2025.10.22.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유
24	2025.10.29.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유
25	2025.11.05.	• 의미역, 무형 대응어 구축 말뭉치 검수 결과 공유, 이후 일정 확인
26	2025.11.12.	• 무형 대응어 구축 말뭉치 검수 결과 공유
27	2025.11.19.	• 통합 검수 수정 방향 논의
28	2025.11.25.	• 통합 검수 수정 방향 논의

<표 55> 주요 회의 일자

매주 진행된 정규 회의에서는 작업자 질의를 바탕으로 추가로 필요한 사항을 정리했고, 검수 과정에서 쟁점이 되었던 분석의 차이를 논의하며 지침을 보완하였다. 이 회의 결과는 작업 기준을 재정비하고 지침을 수정·보완하는 데에 반영되었다.

## 3.4. 파일 관리

### 3.4.1. 주 차별 구축 경과

한국어 구조 분석 말뭉치 4종과 무형 대용어 말뭉치 구축 작업은 사업 착수 후 원문 선별과 정제 과정을 거쳐, 5월 12일부터 11월 18일까지 진행되었다. 먼저 형태 분석 말뭉치를 구축 완료하였고, 이어 동일 말뭉치를 대상으로 어휘의미 분석과 구문 분석 말뭉치를 구축하였다. 이후 의미역 분석과 무형 대용어 복원 단계가 순차적으로 진행되었다. 구축 전 과정은 1주 단위로 운영되었으며, 주차별 목표 어절 수를 작업자에게 배정하여 진행률을 점검하였다. 다음 표에서는 주차별 구축 어절 수, 누적 어절 수, 전체 구축 규모를 정리하여 제시한 것이다.

□ 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 내역

(단위: 어절)

작업 내용	차수	작업 일자	구축량	누적량
형태 분석	1	2025.05.12.~2025.05.18.	64,948	64,948
	2	2025.05.19.~2025.05.25.	133,384	198,332
	3	2025.05.26.~2025.06.01.	138,193	<b>336,525</b>
어휘의미 및 구문 분석	1	2025.06.04.~2025.06.09.	28,453	28,453
	2	2025.06.09.~2025.06.16.	50,163	78,616
	3	2025.06.16.~2025.06.23.	45,142	123,758
	4	2025.06.23.~2025.06.30.	35,503	159,261
	5	2025.06.30.~2025.07.07.	35,545	194,806
	6	2025.07.07.~2025.07.14.	38,022	232,828
	7	2025.07.14.~2025.07.21.	56,602	289,430
	8	2025.07.21.~2025.07.28.	47,095	<b>336,525</b>
의미역 분석	1	2025.08.26.~2025.09.01.	37,890	37,890
	2	2025.09.01.~2025.09.08.	33,576	71,466
	3	2025.09.08.~2025.09.15.	35,119	106,585
	4	2025.09.15.~2025.09.22.	43,944	150,529
	5	2025.09.22.~2025.09.29.	41,535	192,064
	6	2025.09.29.~2025.10.06.	40,297	232,361
	7	2025.10.06.~2025.10.13.	29,190	261,551
	8	2025.10.13.~2025.10.20.	41,474	303,025
	9	2025.10.20.~2025.10.27.	33,500	<b>336,525</b>

무형 대용어 복원	1	2025.08.26.~2025.09.01.	14,513	14,513
	2	2025.09.01.~2025.09.08.	-	14,513
	3	2025.09.09.~2025.09.16.	23,377	37,890
	4	2025.09.16.~2025.09.23.	33,576	71,466
	5	2025.09.23.~2025.09.30.	35,119	106,585
	6	2025.09.30.~2025.10.07.	43,944	150,529
	7	2025.10.07.~2025.10.14.	41,535	192,064
	8	2025.10.14.~2025.10.21.	40,297	232,361
	9	2025.10.21.~2025.10.28.	29,190	261,551
	10	2025.10.28.~2025.11.04.	41,474	303,025
	11	2025.11.04.~2025.11.11.	33,500	<b>336,525</b>

<표 56> 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 내역

어휘의미 분석, 구문 분석, 의미역 분석, 무형 대용어 복원 작업의 1주 차에는 구축 목표 어절 수를 다른 주차보다 낮게 배정하여 도구 사용 방법을 익히고, 작업 지침 숙지에 필요한 시간을 확보하였다.

어휘의미 분석과 구문 분석 말뭉치 구축 작업은 동일한 기간 동안 병렬로 진행되었다. 반면 의미역 분석과 무형 대용어 복원 작업은 의미역 검수가 완료된 작업물을 기반으로 무형 대용어 복원 작업을 진행하고자 수행 주차를 달리 운영하였다. 무형 대용어 복원 작업에 의미역 분석 결과를 참고하여 작업의 난도를 낮추고, 효율성과 정확성을 향상시키기 위한 목적으로 설계한 것이다. 이 과정에서 무형 대용어 복원 2주 차에는 작업자 간 교차 검수를 실시하여 작업자의 결과물을 상호 검토하였다. 이를 통해 오류 유형을 조기에 확인하고, 작업자 간 작업 기준을 정렬함으로써 품질 향상에 도움이 되었다.

사업 수행 기간 중 추석 연휴 일정 등으로 인해 주 차별 목표 어절 수를 일부 조정된 경우가 있었지만, 초기 목표량 설계에 보정 가능한 여유치를 두어 일정 지연 위험을 차단하였다. 또한 작업자의 숙련도를 고려하여 사업 진행 후반부로 갈수록 주 차별 목표 어절 수를 점진적으로 확대 배정하여 구축 속도와 품질 안정성을 확보하고자 하였다.

이러한 운영 방식을 통해, 사업 전 과정에서 말뭉치 구축의 일관성과 정확성을 확보하였고, 구축된 말뭉치 결과물은 당초 계획 일정에 맞추어 차질 없이 최종 제출하였다.

### 3.3.2. 중간 납품 경과

중간 납품은 사업 착수 시 수립된 일정에 따라 운영하였다. 제안 요청서에서는 사업 착수 후 분석 대상 원문 자료는 6주 이내, 4개월 시점에는 층위별 분석 말뭉치 4종 50%



와 무형 대용어 복원 말뭉치 15% 이상, 6개월 시점에는 층위별 분석 말뭉치 4종 80% 이상과 무형 대용어 복원 말뭉치 40% 이상의 구축 결과물을 제출하도록 명시되었다. 그러나 이 사업은 ‘형태 분석 → 어휘의미 분석·구문 분석 → 의미역 분석·무형 대용어 복원’의 순차적 흐름을 전제로 하며, 선행 분석 완료 이후에만 후속 단계 수행이 가능한 구조를 가진다. 특히 구문 분석 말뭉치 구축이 완료되어야 의미역 분석과 무형 대용어 복원이 가능하다. 이에 따라, 사업 변경 신청을 통해 승인을 거쳐 납품 목표를 조정하였다. 4개월 시점의 목표는 형태·어휘의미·구문 분석 말뭉치의 100% 제출로 조정하였고, 6개월 시점의 납품 목표는 형태·어휘의미·구문 분석 각 100%, 의미역 분석 50%, 무형 대용어 복원 50%를 제출하여 전체 공정률 80%까지 중간 납품하는 방식으로 조정하였다. 조정된 목표는 당초 일정 틀에서 벗어나지 않는 범위 내에서 관리하였고, 아래의 일정에 따라 중간 납품을 수행하였다.

No.	일자	교육 내용
1	2025.06.04.	• 분석 대상 원문 자료 제출
2	2025.08.22.	• 형태·어휘·구문 100%
3	2025.10.22.	• 형태·어휘·구문 100% • 무형 대용어, 의미역 50%

<표 57> 중간 납품 일정

### 3.4.3. 말뭉치 구축 결과

한국어 구조 분석 및 무형 대용어 복원 말뭉치 연구 결과는 구축과 검수, 통합 검수를 거쳐 완성하였다.

층위별 구조 분석 말뭉치는 작업이 종료된 다음 주부터 검수 작업을 진행하였다. 1주차 구축 어절 수에는 검수자가 작업한 어절 또한 포함되어 있는데, 이를 통해 검수 기준을 실질적으로 점검하고, 이후 검수 단계에서 발생할 수 있는 오류를 선제적으로 최소화하고자 하였다. 형태·어휘의미·구문 분석 및 의미역 분석, 무형 대용어 복원 말뭉치 검수 작업은 아래와 같이 진행되었다.

(단위: 어절)

작업 내용	차수	작업 일자	작업량	검수량
형태 분석	1	2025.05.12.~2025.05.18.	64,948	64,948
	2	2025.05.19.~2025.05.25.	133,384	133,384
	3	2025.05.26.~2025.06.01.	138,193	139,907
어휘의미 및 구문 분석	1	2025.06.04.~2025.06.09.	28,453	-
	2	2025.06.09.~2025.06.16.	50,163	28,453
	3	2025.06.16.~2025.06.23.	45,142	50,163
	4	2025.06.23.~2025.06.30.	35,503	45,142
	5	2025.06.30.~2025.07.07.	35,545	35,503
	6	2025.07.07.~2025.07.14.	38,022	35,545
	7	2025.07.14.~2025.07.21.	56,602	38,022
	8	2025.07.21.~2025.07.28.	47,095	56,602
	9	2025.07.28.~2025.08.05.	-	47,095
의미역 분석	1	2025.08.26.~2025.09.01.	37,890	-
	2	2025.09.01.~2025.09.08.	33,576	37,890
	3	2025.09.08.~2025.09.15.	35,119	33,576
	4	2025.09.15.~2025.09.22.	43,944	35,119
	5	2025.09.22.~2025.09.29.	41,535	43,944
	6	2025.09.29.~2025.10.06.	40,297	41,535
	7	2025.10.06.~2025.10.13.	29,190	40,297
	8	2025.10.13.~2025.10.20.	41,474	29,190
	9	2025.10.20.~2025.10.27.	33,500	41,474
	10	2025.10.27.~2025.11.03.	-	33,500
무형 대용어 복원	1	2025.08.26.~2025.09.01.	14,513	-
	2	2025.09.01.~2025.09.08.	-	14,513
	3	2025.09.09.~2025.09.16.	23,377	-
	4	2025.09.16.~2025.09.23.	33,576	23,377
	5	2025.09.23.~2025.09.30.	35,119	33,576
	6	2025.09.30.~2025.10.07.	43,540	35,119
	7	2025.10.07.~2025.10.14.	41,535	43,540
	8	2025.10.14.~2025.10.21.	40,297	29,030
	9	2025.10.21.~2025.10.28.	29,190	29,733
	10	2025.10.28.~2025.11.04.	41,878	29,068
	11	2025.11.04.~2025.11.11.	33,500	65,069
	12	2025.11.11.~2025.11.18.	-	33,500

&lt;표 58&gt; 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 결과

검수는 1주 단위로 진행되어, 11월 18일 종료되었다. 어휘의미 분석 검수는 엑셀을, 구문 분석 검수는 엑셀과 작업 도구를 함께 활용하여 수행되었다. 형태 분석, 의미역

No.	작업일	I			J	K			L	M			N	O			P	Q			R	S			T	U			V	W			X	Y			Z	AA			AB	AC			AD	AE			AF	AG			AH	AI			AJ	AK			AL	AM			AN	AO			AP	AQ			AR	AS			AT	AU			AV	AW			AX	AY			AZ	BA			BB	BC			BD	BE			BF	BG			BH	BI			BJ	BK			BL	BM			BN	BO			BP	BQ			BR	BS			BT	BU			BV	BW			BX	BY			BZ	CA			CB	CC			CD	CE			CE	CF			CF	CG			CG	CH			CH	CI			CI	CJ			CJ	CK			CK	CL			CL	CM			CM	CN			CN	CO			CO	CP			CP	CQ			CQ	CR			CR	CS			CS	CT			CT	CU			CU	CV			CV	CW			CW	CX			CX	CY			CY	CA			CA	CB			CB	CC			CC	CD			CD	CE			CE	CF			CF	CG			CG	CH			CH	CI			CI	CJ			CJ	CK			CK	CL			CL	CM			CM	CN			CN	CO			CO	CP			CP	CQ			CQ	CR			CR	CS			CS	CT			CT	CU			CU	CV			CV	CW			CW	CX			CX	CY			CY	CA			CA
-----	-----	---	--	--	---	---	--	--	---	---	--	--	---	---	--	--	---	---	--	--	---	---	--	--	---	---	--	--	---	---	--	--	---	---	--	--	---	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----	----	--	--	----

어휘의미 분석의 경우, 동일 어휘 항목을 모아 비교하면 오류를 명확히 확인할 수 있다는 특성이 존재하므로, 엑셀에서 어휘 기준 정렬을 적용하여 검수자가 동일 어휘를 확인할 수 있도록 운영하였다. 한편, 구문 분석은 문장 단위 검토가 필수적인 층위이므로 검수 작업 또한 문장 단위로 배포하고 도구에서 검수하도록 하였다. 구문 분석 표지는 엑셀에서, 지배 의존 관계는 구축 도구에서 각각 명시적으로 확인할 수 있다는 장점이 존재하기에, 두 장점을 분리하여 교차 적용하는 방식으로 검수가 이루어졌다. 이에 따라, 각 단계의 검수는 말뭉치 분석 층위별 특성에 기반하여 일관된 기준 아래 진행하였다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
연번	File Name	File ID	파일명	내용	종류	종류	종류	종류	종류	종류	종류	종류	종류	종류	종류	종류
1030	30R04000001770.jpg	30R04000001770.1	30R04000001770.1.1.178	한글 파일 열람한 기록 파일 저장	5	연계	7	NP_A01	5	LOC	7	기타				
1087	30R04000001770.jpg	30R04000001770.1	30R04000001770.1.1.287	기타에 연대하여 목출된 그 비공인 발문은 소문과 같고 있다.	2	연계	10	NP_A01	2	LOC	2	기타				
1283	30R04000001788.jpg	30R04000001788.1	30R04000001788.1.1.41	인수준은 현재까지 어느 지역에도 실려 있지 않다	5	기타	6	NP_A01	5	LOC	5	기타				
1350	30R04000001788.jpg	30R04000001788.1	30R04000001788.1.1.81	연세에서 이제 여가 있어 그 때로 찍었는데	1	연세	8	NP_A01	1	LOC	8	기타				
3116	30R04000001682.jpg	30R04000001682.1	30R04000001682.1.1.18	북쪽에서 가장 가까운 길은 가는 길로 빠져도 되고	3	기타	4	NP_A01	3	LOC	4	기타				
3413	30R04000001682.jpg	30R04000001682.1	30R04000001682.1.1.88	북쪽에서 가장 가까운 길은 가는 길로 빠져도 되고	4	기타	3	NP_A01	4	LOC	3	기타				
3608	30R04000001682.jpg	30R04000001682.1	30R04000001682.1.1.119	북쪽에서 가장 가까운 길은 가는 길로 빠져도 되고	1	연세	8	NP_A01	1	LOC	8	기타				
3642	30R04000001682.jpg	30R04000001682.1	30R04000001682.1.1.118	북쪽에서 가장 가까운 길은 가는 길로 빠져도 되고	8	기타	3	NP_A01	8	LOC	3	기타				
3863	30R04000001682.jpg	30R04000001682.1	30R04000001682.1.1.124	연세에서 가는 길로 빠져도 되고	2	연세	8	NP_A01	2	LOC	8	기타				
3908	30R04000001682.jpg	30R04000001682.1	30R04000001682.1.1.125	북쪽에서 가장 가까운 길은 가는 길로 빠져도 되고	1	연세	8	NP_A01	1	LOC	8	기타				

이후 데이터리에서 진행한 일관성 검사 결과를 토대로 작업 도구에서의 수정이 이루어졌으며, 수정 사항을 반영하여 최종 납품 결과물을 산출하였다.



## 제 10 장

### 결론



# 1. 요약

본 사업은 ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’를 대상으로 문어 20만 어절 및 구어 10만 어절로 구성된 총 30만 어절 이상 규모의 형태 분석 말뭉치와 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치, 무형 대용어 복원 말뭉치를 구축하고 이들을 통합한 다층위 구조 분석 말뭉치를 구축하였다.

말뭉치 구축 작업에 선행하여 기존의 한국어 구조 분석 지침들을 연구하고 분석하여 다층위 구조 분석 말뭉치 구축이 가능한 단계로 각각의 지침을 수정하고 보완하였으며 이를 통합하여 다층위 구조 분석 말뭉치 구축 지침을 수립하였다.

또한 정확도 높은 분석을 위하여 원문 정제 단계를 통해 맞춤법과 띄어쓰기 등 어문 규범에 맞지 않는 표현들과 말더듬 표현이나 머뭇거림 표현 등 분석에 불필요한 부분들을 삭제하고 정리하여 원활한 다층위 구조 분석 말뭉치 구축이 가능한 환경을 조성하였다.

이러한 과정을 통해 형태 분석 말뭉치와 어휘의미 분석 말뭉치, 구문 분석 말뭉치, 의미역 분석 말뭉치, 무형 대용어 복원 말뭉치를 구축하였으며 각각의 말뭉치가 통합된 336,525 어절의 다층위 구조 분석 말뭉치를 구축하였다.

무형 대용어 복원 말뭉치는 문어 말뭉치에서 31,714개, 구어 말뭉치에서 18,536개 서술어를 대상으로 구축하였고 문어 말뭉치에서는 28,093개의 선행어와 9,815개의 비지시 대명사를, 구어 말뭉치에서는 18,630개의 선행어와 6,232개의 비지시 대명사를 복원하였다.

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업은 다음과 같은 특징 및 장점을 지닌다.

첫째, 본 사업을 통해 체계적인 다층위 말뭉치를 구축하였다. 본 사업팀이 가지고 있는 풍부한 말뭉치 구축 경험을 바탕으로 효율적인 말뭉치 구축 체계를 마련하고 기계 분석과 인간 작업자 분석을 병행하여 고품질 말뭉치를 구축하였다.

둘째, 본 사업을 통해 통합 지침을 작성하였다. 본 사업팀만의 다양한 분석 지침 작성 경험을 바탕으로 작업자들의 효율적인 작업이 가능하도록 통합 지침을 작성하고 기존 공개된 각 층위별 말뭉치 구축 지침에 대한 면밀한 연구와 분석을 바탕으로 층위별 말뭉치 분석 지침 사이의 일관성을 확보하였다.

셋째, 효율적인 품질 보증 체계를 수립하였다. 본 사업팀은 5년간의 국립국어원 과제를 성공적으로 수행한 경험을 가지고 있다. 성공적인 과제 수행 노하우를 바탕으로 효율적인 사업이 진행될 수 있도록 작업 도구를 구축하여 작업 일관성을 확보하고 한국어 관련 석박사 이상의 학력 및 한국어 구조 분석 사업 참여 경험이 있는 숙련된 작업자를 확보하고

전문적인 교육을 통해 고품질 말뭉치를 구축하였다.

## 2. 말뭉치 구축의 의의

‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업은 국내 표준화 및 참조 기반 자료가 될 수 있는 정밀한 언어 정보 부착 말뭉치를 분석하여 한국어 말뭉치 자료의 가치와 활용도를 제고하며 인공지능 언어 모델의 한국어 처리 기술 고도화 및 인공지능 평가 자료가 되는 말뭉치 및 관련 지침들을 구축하고 배포하여 관련 연구 사업에 도움이 될 수 있도록 하는 데에 그 의의가 있다.

정밀한 언어 정보 부착 말뭉치는 인공지능 개발을 위한 학습 자료로 활용될 수 있으며 정확하고 자연스러운 한국어 문장 생성 시스템 개발에 이바지하고 한국어의 형태, 구문, 의미 사이의 관련성에 대한 연구 자료로도 활용되어 인공지능이 도입된 여러 산업 분야와 한국어 연구 분야에 다양하게 활용될 수 있을 것이다.

또한 본 사업을 통해 마련된 다층위 구조 분석 말뭉치 구축 지침은 기존 연구 성과를 비판적인 검토 과정을 통해 보완하고 개선하는 과정을 통해 마련한 것으로 관련 연구 분야의 국내 표준화 자료로 활용될 수 있으며 유사 말뭉치 구축 시의 참조 기반 자료로도 활용될 수 있다. 다층위 통합 분석 말뭉치를 통해 확인할 수 있는 각 층위별 분석 정보들은 한국어 문장 구조의 특성을 보여주는 실증적인 연구 자료로 활용될 수 있다. 다층위 구조 분석 말뭉치에 포함된 한국어 문장의 형태, 어휘의미, 구문, 의미역 및 생략 정보는 ‘형태 정보와 의미역 정보’, ‘구문 정보와 의미역 정보’ 등 여러 방식으로 교차 검색할 수 있어 연구자들이 자신의 연구에 필요한 정보만을 취사선택할 수 있는 활용도 높은 연구 환경을 제공할 것이다.

다층위 구조 분석 말뭉치 구축을 위해 개발된 구축 도구 또한 향후 이루어질 유사 말뭉치 구축 사업 진행 시에도 활용할 수 있는 확장성을 지니고 있다. 또한 본 사업을 통해 마련된 다층위 구조 분석 말뭉치 구축 지침과 작업자 교육 프로그램을 통해 전문적인 한국어 분석 작업자를 양성할 수 있는 하나의 완성된 시스템으로 활용될 수 있다.

본 사업을 통해 구축된 말뭉치와 설명 자료가 민간에 공개되고 이후 추가적인 말뭉치 구축 과정을 통해 대량의 통합 말뭉치가 구축된다면 한국어 말뭉치의 활용도를 제고하게 될 것이며 인공지능 산업과 빅데이터 산업 등 관련 분야 발전의 토대를 형성하게 될 것이다.

### 3. 제언

본 사업은 다층위 구조 분석 말뭉치 시범 구축을 위한 연구의 성격이 짙다. 따라서 대량의 말뭉치를 구축하기보다는 30만 어절 규모의 다층위 구조 분석 말뭉치를 구축하면서 그 과정에서 발생할 수 있는 여러 문제점들을 고민하고 나름의 해결책들을 지침에 반영하는 것에 모든 노력을 투자하였다. 다층위 구조 분석 말뭉치가 그 효용을 가지기 위해서는 그 무엇보다 말뭉치 규모의 양적인 확대가 필요하다. 따라서 향후 이루어질 말뭉치 구축 사업에서는 본 사업을 통해 마련된 말뭉치 구축 지침과 도구, 작업자 관리 시스템을 최대한 그대로 활용하면서 일관된 작업 진행을 통해 대규모의 다층위 구조 분석 말뭉치 구축의 효율성을 추구해야 할 것이다.

다만 어휘의미 분석과 의미역 분석, 무형 대용어 분석은 사전의 격틀에 의존하고 있는 만큼 말뭉치 구축 과정에서 참조할 수 있는 신뢰도 높은 사전을 마련하는 작업이 선행되어야 할 것이다. 논항과 부가어의 판단, 맥락과 환경에 따라 달라지는 어휘의미의 판별은 작업자 개인이 판단하기에는 전체 결과물의 일관성 확보에 어려움이 있다. 말뭉치 구축 사업을 통해 확보된 자료들을 통해 사전을 수정 보완하는 작업이 함께 이루어진다면 사전에 기반한 보다 정교한 말뭉치 구축 작업이 이루어질 것이라 예측된다.

# 부록

## [붙임1] 형태 분석 말뭉치 구축 지침

### 2025년 형태 분석 말뭉치 구축 지침

#### 1. 총론

##### 1.1. 지침의 목적

이 지침은 한국어 문장을 형태 분석하여 형태 분석 말뭉치 구축 방안을 제시하는 것을 목적으로 한다. 한국어 문장의 각 어절을 형태 단위로 분리하고 각 형태에 형태 표지를 부착하는 것이 세부 목적이다. 형태 분석은 원시 말뭉치를 대상으로 수행한다.

- 본 지침의 어절 분석 표지는 '21세기 세종계획'의 '형태 분석 말뭉치 구축 지침'과 '2020 형태분석 말뭉치 구축 지침'을 토대로 작성된 것이다.
- 어절 분석 표지는 큰 틀은 21세기 세종계획의 어절 분석 표지를 따르고, 품사 태그의 경우는 TTA의 분석 표지를 참고하였다. 그리고 세종 말뭉치의 문어, 구어 분석 표지를 통합한 것이다.
- 원문을 정제하고 가공하여 사용하게 됨에 따라 말뭉치 오류 및 예외 처리에 관련된 지침을 대폭 축소하였다.
- 본 지침은 이후 어휘의미 분석, 구문 분석 등 단계적인 분석을 할 수 있도록 고려하였다.

##### 1.2. 용어 정의

- 형태: 하나의 형태소가 환경에 따라 달리 실현된 모습
  - 대표형으로 되돌리지 않고 최대한 각각의 형태를 분석한다.
- 단어: 하나의 어절을 구성하는 자립 형식으로, <우리말샘> 사전의 기본 등재 단위
  - 고유명사는 그 구성에 관계없이 하나의 단어 자격으로 분리한다.
  - 단어의 내부 분석은 지양한다.
- 어절: 문장을 구성하고 있는 문장 성분의 최소 단위로서 띄어쓰기의 단위
  - 한글맞춤법의 띄어쓰기 규정에 따라 어절을 분리하되 <우리말샘>의 등재어는 구나 문장도 하나의 어절로 분리한다.



- 품사: 단어를 형태, 기능, 의미 등 기준에 따라 나눈 갈래
  - 학교 문법의 품사 분류 체계를 따르되 ‘이다’와 ‘아니다’는 용언의 하위에 두어 지정사로 분석한다.

### 1.3. 기본 원칙

- 분석 대상
  - 형태 분석은 하나의 어절을 분석 대상으로 한다.
- 분석 원리
  - 본 분석은 ‘형태소’ 차원이 아닌 ‘형태’ 차원의 분석이므로 이형태를 최대한 반영한다.
- 분석 원칙
  - 원시 말뭉치는 한글맞춤법과 <우리말샘>의 표제어를 기준으로 정제하여 사용한다.
  - <우리말샘>의 표제어를 기준으로 어절을 분리하고 형태 분석 표지를 부착한다.
  - 최대한 단어 단위로 분석하고, 접사는 꼭 필요한 경우에 한하여 분석한다.
  - 지시 대상이 하나라면 구나 절로 표현된 대상도 하나의 단어로 취급한다.

## 2. 형태 분석의 원리

### 2.1. 형태 분리의 원칙

- 본 형태 분석 지침은 형태 분석 결과와 어휘의미 분석 결과, 구문 분석 결과, 의미역 분석 결과를 층위별로 쌓은 다층위 구조 분석 말뭉치 구축을 위해 마련되었다. 따라서 형태 분석은 다음 단계의 어휘의미 분석과 구문 분석에 필요한 정보만을 제공할 수 있도록 단어 단위 분석을 지향하여 체언과 조사를 분리하고, 어간과 어미를 분리한다.
- 꼭 필요한 경우가 아니면 단어 내부 분석은 지양한다. 단어 내부 구성 정보는 사전의 형태 정보를 참고할 수 있다.
- <우리말샘>의 등재어를 기준으로 형태 분석 대상을 분리한다.
  - <우리말샘>의 등재어는 붙어서 한 단위로 분석하고 미등재어는 한글맞춤법에 맞게 어절을 분리하여 분석한다.  
[예] 군인연금 (사전 O), 직장인 연금 (사전 X)
  - <우리말샘>의 등재어 중, '^'로 표시된 것은 띄어 쓰지 않고 한 어절로 붙이고, 띄어 쓰고 있는 것은 '\_'를 이용하여 붙여 한 단위로 분석한다.  
[예] 연금의현가계수 (사전: 연금의^현가^계수)  
[예] 경제학\_원리 (사전: 경제학 원리)
- '반(反)-, 대(對)-' 등 체언접두사는 따로 분리하지 않고 어근과 통합한 단위로 분리한다.
- '-하다, -되다' 등 동사파생접미사는 따로 분리하지 않고 어근과 통합한 단위로 분리한다.
- 단, '업로드(upload)하다', '처리하다' 등 어근이 괄호나 기호로 분리된 경우에는 '-하다'를 접미사로 분리하여 분석한다.
- '-적(的), -성(性), -화(化)' 등 명사파생접미사는 따로 분리하지 않고 어근과 통합한 단위로 분리한다.
- 합성어는 내부 어근을 분리하지 않고 하나의 단어로 분석한다.
- 지시 대상이 하나인 구나 절도 하나의 단어로 취급하여 내부 구성 성분을 분리하지 않는다.

## 2.2. 형태 분석 표지 부착

- 형태 분석 표지

대분류	소분류	세분류	태그
체언	명사	일반명사	NNG
		고유명사	NNP
		의존명사	NNB
	대명사	대명사	NP
	수사	수사	NR
용언	동사	동사	VV
	형용사	형용사	VA
	지정사	긍정지정사	VCP
		부정지정사	VCN
	보조용언	보조용언	VX
수식언	관형사	성상 관형사	MMA
		지시 관형사	MMD
		수 관형사	MMN
	부사	일반부사	MAG
		접속부사	MAJ
독립언	감탄사	감탄사	IC
관계언	격조사	주격조사	JKS
		보격조사	JKC
		관형격조사	JKG
		목적격조사	JKO
		부사격조사	JKB
		호격조사	JKV
		인용격조사	JKQ
	보조사	보조사	JX
	접속조사	접속조사	JC
의존형태	어미	선어말어미	EP
		종결어미	EF
		연결어미	EC
		명사형전성어미	ETN
		관형형전성어미	ETM
	접두사	체언접두사	XPN
	접미사	명사파생접미사	XSN
		동사파생접미사	XSV
		형용사파생접미사	XSA
	어근	어근	XR

기호	일반기호	마침표, 물음표, 느낌표	SF
		쉼표, 가운뎃점, 콜론, 빗금	SP
		따옴표, 괄호표, 줄표	SS
		줄임표	SE
		붙임표(물결)	SO
		기타 기호	SW
	외국 문자	외국 문자	SL
	한자	한자	SH
	숫자	숫자	SN
	기타	분석불능범주	NA

### 3. 형태 분석의 실제

#### 3.1. 체언

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

##### 1) 접두사의 결합형

- (1) ‘반(反)-, 대(對)-’ 등 체언접두사는 따로 분리하지 않고 어근과 통합한 형식 전체를 명사로 분석한다. 이때 어근의 성격에 따라 고유명사 혹은 일반명사로 분석한다.

원어절	형태 분석 어절
반독재	반독재/NNG
대미	대미/NNP

- (2) 접두사가 괄호와 한자로 함께 표기된 경우에는 각각을 분리하여 분석한다.

반(反)독재	반/XPN + (/SS + 反/SH + )/SS + 독재/NNG
대(對)미	대/XPN + (/SS + 對/SH + )/SS + 미/NNP

##### 2) 접미사의 결합형

- (1) ‘-적(的), -성(性), -화(化)’ 등 명사파생접미사는 따로 분리하지 않고 어근과 통합한 형식 전체를 명사로 분석한다. <우리말샘>에는 명사파생접미사 ‘-적(的)’이 결합된 형태를 관형사와 명사로 다루고 있으나 어근의 성격을 고려하여 명사(NNG)로만 분석한다.

문화적	문화적/NNG
정확성	정확성/NNG
기계화	기계화/NNG

- (2) ‘비교적’, ‘가급적’ 등 명사 혹은 부사로 쓰일 수 있는 경우에는 문맥을 살펴 알맞은 품사로 분석한다.

예문	원어절	형태 분석 어절
비교적인 관점에서 바라본다.	비교적인	비교적/NNG + 이/VCP + L/ETM
그의 방은 비교적 깨끗했다.	비교적	비교적/MAG
가급적이면 빨리 가자.	가급적이면	가급적/NNG + 이/VCP+ 면/EC
가급적 도와주시기 바랍니다.	가급적	가급적/MAG

### 3.1.1. 명사

명사는 사물의 이름을 나타내는 품사이다. 명사의 형태 분석 표지는 일반명사, 고유명사, 의존명사로 세분한다.

#### 3.1.1.1. 일반명사(NNG)

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

##### 1) 일반명사로 분석할 수 있는 단어

###### (1) <우리말샘>의 명사 표제어

원어절	형태 분석 어절
국어연구	국어/NNG+연구/NNG

###### (2) <우리말샘>에 접미사로 등재되지 않은 1음절 한자어

서울초등학교 줄	줄/NNG
----------	-------

###### (3) 한자성어

백척간두(百尺竿頭)	백척간두/NNG+(/SS+百尺竿頭/SH+)/SS
------------	----------------------------

###### (4) 기타 다른 품사로 분석될 수 없는 단위

표기상 한글과 기타 문자(부호나 숫자, 외국 문자)가 섞여 있거나 숫자나 외국 문자로만 표기되어 있으나 대응되는 한글 표제어가 <우리말샘>에 등재되어 있으면, 한글 표제어와 이형태 관계로 파악하여 하나의 단어로 통합하여 분석한다.

원어절	사전 등재어	형태 분석 어절
3.1절	삼일절	3.1절/NNG
6.25	육이오	6.25/NNG
1루수	일루수	1루수/NNG
5관왕	오관왕	5관왕/NNG
병커C유	병커시유	병커C유/NNG

표기상 한글과 기타 문자(부호나 숫자, 외국 문자)가 섞여 있으나 <우리말샘>에 한글로 등재되어 있지 않은 경우에는 각각을 분석한다.

16개교	16/SN+개/NNB+교/NNG
------	-------------------

## 2) 명사 상당어의 분석

- (1) 동사의 활용형이 따옴표 없이 문장 속에서 명사처럼 기능하는 경우는 원래 품사대로 분석한다.

예문	원어절	형태 분석 어절
어디 가느냐가 그의 물음이었다.	가느냐가	가/VV+느냐/EF+가/JKS

- (2) 따옴표를 가진 성분이나 그에 준하는 대상도 명사처럼 기능할 수 있으나, 원래 품사대로 분석한다.

그것은 “는”이 아니라 “를”이다.	“는”이	“/SS+는/JX+”/SS+이/JKC
A: 미안. B: 미안이지 많이 뭐니?	많이	않/NA+이/JKS

\* 따옴표는 없지만 상대방의 오타를 지적하며 오타 부분에 해당하는 ‘않’을 따와 명사처럼 사용한 경우이다. 이때의 ‘않’은 그 자체로 품사를 가지지 않는 말이므로 분석 불능 범주(NA)를 부여한다.

- (3) 부사 뒤에 격조사가 쓰이는 것도 의미론적인 따옴의 효과에 의하여 부사가 명사적인 용법을 가지는 것이므로 분석은 ‘부사’로 한다.

기름을 짹 채우려면 가득을 누르세요.	가득을	가득/MAG+을/JK0
----------------------	-----	--------------

### 3.1.1.2. 고유명사(NNP)

고유명사는 특정한 사물에 붙여진 이름으로, 기본적으로 최하의어에 속하는 대상을 서로 변별하기 위하여 붙인 이름이며 원칙적으로 지시 대상만 가질 뿐 의미 내용은 가지지 않는다.

#### 1) 고유명사로 분석할 수 있는 단어

- (1) 전체 구나 절이 하나의 지시 대상을 가지는 경우 하나의 단어로 분석한다.

단, ‘^’이 포함된 <우리말샘> 등재어는 한 어절로 붙이고, 띄어쓰기가 포함된 경우는 한 어절로 붙이되 ‘\_’를 이용해 경계를 구분한다.

부정^청탁^및^금품^등^수수의^금지에^관한^법률 (법률명)	부정청탁및금품등수수의금지에관한법률/NNP
바람과 함께 사라지다 (영화 제목)	바람과_함께_사라지다/NNP

- (2) 성 또는 이름 뒤에 같이 쓰이는 호칭어나 직책명은 분리해서 분석하나, <우리말샘>에 하나의 표제어로 등재되어 있으면 전체를 하나의 단어로 분석한다. 단, 특정 단체에서만 사용하는 직책명은 고유명으로 분석한다.

남수군	남수/NNP+군/NNB
최치원옹	최치원/NNP+옹/NNB
정과장	정/NNP+과장/NNG
정 패키지세종사업팀장	정/NNP+패키지세종사업팀장/NNP

최선생	최/NNP+선생/NNG
김씨	김씨/NNP
세종대왕	세종대왕/NNP
광개토대왕	광개토대왕/NNP
이태백	이태백/NNP

(3) 사람 이름의 뒤에 붙는 접미사 ‘이’는 이름과 함께 하나의 단어로 분석한다.

진현이가	진현이/NNP+가/JKS
------	---------------

(4) 특정 종족, 언어, 지명, 지역명, 국가, 건축물 및 시설물 등 다른 것들과 구별하여 부르기 위한 이름은 하나의 단어로 분석한다.

알타이족	알타이족/NNP
피그미족	피그미족/NNP
이집트인	이집트인/NNP
아제르바이잔인	아제르바이잔인/NNP
한국어	한국어/NNP
일본어	일본어/NNP
영어	영어/NNP
알타이어	알타이어/NNP
카스피해	카스피해/NNP
템즈강	템즈강/NNP
북한산	북한산/NNP
가자지구	가자지구/NNP
경기지역	경기지역/NNP
경인공업지대	경인공업지대/NNP
대한민국	대한민국/NNP
러시아연방	러시아연방/NNP
미얀마연방공화국	미얀마연방공화국/NNP
서울지하철	서울지하철/NNP
인천공항	인천공항/NNP
테헤란로	테헤란로/NNP

단, 어느 지역의 지하철에나 존재하는 ‘1호선, 2호선’ 등은 특정성이 낮으므로 고유명사로 보지 않는다.

1호선	1/SN+호/NNB+선/XSN
-----	------------------

맥락에 따라 고유명사와 일반명사로 달리 쓰이는 경우가 있으므로 주의한다.

서초구 문화예술공원	문화예술공원	문화예술공원/NNP
------------	--------	------------



(특정 공원의 이름일 때)		
우리 구에 문화예술공원을 설립합시다. (공원의 종류를 나타낼 때)	문화예술공원	문화예술공원/NNG

- (5) 회사나 기관, 단체의 이름은 외국 문자와 한글이 함께 쓰인 경우, 한글 표기 없이 외국 문자만 쓰인 경우 모두 고유명사로 분석한다.

삼성	삼성/NNP
삼성그룹	삼성그룹/NNP
LG전자	LG전자/NNP
현대자동차서비스	현대자동차서비스/NNP
국어학회	국어학회/NNP
한국사학회	한국사학회/NNP
프로야구선수협회	프로야구선수협회/NNP
헌법재판소	헌법재판소/NNP
국립국어원	국립국어원/NNP
여성가족부	여성가족부/NNP
LG에서	LG/NNP+에서/JKB

- 지명 등의 고유명사를 포함하고 있는 정부기관의 명칭 또한 하나의 고유명사로 분석한다. <우리말샘>에 구로 등재된 단위라고 하여도 그 통합형을 고유명사로 분석한다.

서울고등법원	서울고등법원/NNP
서울시경찰서	서울시경찰서/NNP
서대문구치소	서대문구치소/NNP
춘천지방법원원주지원	춘천지방법원원주지원/NNP

- (6) 특정 회사의 상품명과 브랜드명이 하나의 고유명사로 쓰이는 경우와 회사명과 상품명을 분리할 수 있는 경우를 구분하여 분석한다.

원어절	형태 분석 어절	비고
칠성사이다	칠성/NNP+사이다/NNP	회사명+상품명
신한SOL	신한/NNP+SOL/NNP	회사명+상품명
삼성갤럭시S8	삼성/NNP+갤럭시S8/NNP	회사명+상품명
농심새우깡	농심/NNP+새우깡/NNP	회사명+상품명
e편한세상아파트	e편한세상아파트/NNP	브랜드명+종류명
하나에스라인적금	하나에스라인적금/NNP	회사명+상품명+종류명

- (7) 고유명사가 축약된 형태(준말)로 쓰일 경우 본디말과 함께 준말도 인정하여 축약된 형태 그대로를 고유명사로 분석한다.

육사	육사/NNP
----	--------

고대	고대/NNP
자민련	자민련/NNP
서울고법	서울고법/NNP

#### (8) 단체장명의 처리

지명/단체명에 ‘-장(長)’, ‘수(守)’ 등의 요소가 결합하여 단체장명이 만들어진 경우 해당 요소를 앞말에 붙여 분석한다. 단체장명은 일반명사임에 유의한다.

가평군수	가평군수/NNG
서울시장	서울시장NNG
헌법재판소장	헌법재판소장/NNG
비대위원장	비대위원장/NNG

단, 접사가 아니라 ‘소장, 원장, 지사’와 같은 명사가 결합하여 단체장명이 만들어지는 경우에는 분리하여 각각을 분석한다.

헌법재판소소장	헌법재판소/NNG+소장/NNG
민족문화연구원원장	민족문화연구원/NNG+원장/NNG
서울시의원	서울시/NNP+의원/NNG
제주도지사	제주/NNP+도지사/NNG

#### (9) 원문에서 따옴표로 묶인 대상의 처리

대회나 축제, 회의 등의 행사와 제목으로 쓰인 회의 및 발표의 주제가 원문에서 연도, 횟수와 함께 따옴표 안에 표시된 경우 고유명으로 파악하여 그 내부를 분석하지 않는다.

‘2023_BTS_FESTA’(공연, 축제명)	/SS+2023_BTS_FESTA/NNP+/SS
‘2023년도_제1회_경상북도_지방공무원_공개경쟁_임용시험’(시험명)	/SS+2023년도_제1회_경상북도_지방공무원_공개경쟁_임용시험/NNP+/SS
‘챗GPT_변화와_기획의_시대’(발표 주제명)	/SS+챗GPT_변화와_기획의_시대/NNP+/SS

### 3.1.1.3. 의존명사(NNB)

의존명사는 자립해서 쓰일 수 없는 명사로, 수식 성분을 반드시 동반해야 한다. 의존명사와 일반명사의 구분은 <우리말샘>에 따른다.

#### 1) 의존명사와 일반명사의 구분

(1) ‘연대, 연도’는 ‘년대, 년도’와 달리 일반명사이다.

연도별로 정리된 자료	연도별로	연도/NNG+별/XSN+로/JKB
몇 년도에 일어난 일	년도에	년도/NNB+에 JKB

(2) ‘월, 연, 일, 주, 달러, 원’ 등은 독립되어 쓰일 경우에는 모두 일반명사의 자격을 가지므로 의존명사

와 일반명사를 구분하여 분석해야 한다.

나는 월 30만 원을 받는다.	월	월/NNG
달러의 가치는	달러의	달러/NNG+의/JKG
시간당 만 원을 받는다.	시간당	시간/NNG+당/XSN

‘원’은 <우리말샘>에 의존명사로만 올라 있지만, 수식 성분 없이 독립되어 쓰인 경우에는 일반명사로 분석한다. 다른 유사 경우도 이에 따라 처리한다.

1달러를 원으로 환산하면	원으로	원/NNG+으로/JKB
---------------	-----	--------------

## 2) 단위를 나타내는 표현

(1) 길이, 무게, 수효, 시간 따위의 수량을 수치로 나타내는 단위들 중 ‘미터, 그램, 리터’ 등은 의존명사(NNB)로, 외국어로 된 ‘m, g, l’ 등은 기호(SW)로 분석한다.

(2) 일반명사가 단위적인 용법으로 쓰인 경우에는 의존명사가 아니므로 주의한다.

한 사람이 교실로 들어왔다.	사람이	사람/NNG+이/JKS
자장면 한 그릇만 주세요.	그릇만	그릇/NNG+만/JX

## 3) ‘것’과 구어형 ‘거’의 분석

‘거’의 형태를 그대로 인정하여 분석한다.

공부할 거를 준비해 왔니?	거를	거/NNB+를/JKO
공부할 걸 가져왔니?	걸	거/NNB+ㄹ/JKO
연습할 건 있니?	건	거/NNB+ㄴ/JX
먹을 게 모자라다.	게	거/NNB+이/JKS

## 3.1.2. 대명사(NP)

대명사는 그 자체로는 자신의 본유적 지시물을 가지지 않은 채, 다만 사람이나 사물 등 어떤 대상을 간접적으로 지시하는 품사이다. 단, 동일한 대명사가 방언이나 고어의 이형태를 가진 경우에는 이들도 대명사로 같이 분석한다.

### 1) 인칭 대명사

#### (1) 1인칭 대명사

[예] 나, 내, 우리, 저, 제, 저희

(2) 2인칭 대명사

[예] 너, 네, 그대, 당신, 댁

(3) 기타 대명사

[예] 이이, 이분, 그이, 그분, 저이, 저분, 아무, 아무개, 누구, 무엇, 뭐, 뭐시기, 어디, 언제, 자기, 개, 재, 애, 이것, 저것, 그것, 이거, 저거, 그거, 여기, 저기, 거기, 이곳, 그곳, 저곳, 어디, 모(某), 모모(某某)

2) 대명사와 관형사의 두 가지 분석이 가능한 단어

(1) '모(某)'는 관형사와 대명사로 분석될 수 있으므로 주의를 요한다.

모 기업체	모	모/MMD
김 모씨	모씨	모/NP+씨/NNB

(2) '모모(某某)'도 위와 같이 분석될 수 있다.

모모가 말했다.	모모가	모모/NP+가/JKS
모모 기관의 조사를 마쳤다.	모모	모모/MMD

3) 대명사의 이형태 분석

(1) '이것, 그것, 저것; 이거, 그거, 저거'는 분석하지 않고 대명사로 인정한다. '~거'의 경우, '~거'의 형태를 그대로 인정하여 분석한다.

난 저거를 먹을래.	저거를	저거/NP+를/JKO
나는 여태 그걸 믿어 왔단다.	그걸	그거/NP+를/JKO

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

이제부터는 내가 명령을 내린다.	내가	내/NP+가/JKS
이제부터는 내 명령을 따라라.	내	나/NP+의/JKG
내게 전자우편으로 알려 다오.	내게	나/NP+에게/JKB
어제 네게 보낸 선물이 잘못되었다.	네게	너/NP+에게/JKB
문제가 있다면 제게 말씀해 주세요.	제게	저/NP+에게/JKB
누가 전화를 하는지 보고해라.	누가	누구/NP+가/JKS
뉘 집 얘기가 이렇게 울고 있는 거야?	뉘	누구/NP+의/JKG
도대체 뭐가 문제라는 거야?	뭐가	뭐/NP+가/JKS

(3) '제'의 경우, '제/NP+가/JKS'를 제외하고는 모두 '저/NP+의/JKG'로 분석한다.

제가 갈 것입니다.	제가	제/NP+가/JKS
철수는 제 잘못을 안다.	제	저/NP+의/JKG

제 무게를 못 견디다.	제	저/NP+의/JKG
--------------	---	------------

### 3.1.3. 수사(NR)

수사는 사물의 수량이나 차례를 나타내는 품사를 말한다.

#### 1) 수사의 종류

##### (1) 양수사

[예] 하나, 둘, 셋, 넷, 다섯, 열, 스물, 서른, 마흔, 천  
백한둘, 두서넛, 서넛, 너넛, 네다섯  
일, 이, 삼, 사, 오, 십, 백, 천, 만, 억, 조  
기십, 기백, 기천, 수십, 수백, 수천, 수만

##### (2) 서수사

[예] 첫째, 둘째, 셋째, 넷째, ..., 열째, 열한째, ..., 스물한째, ...

#### 2) 수사와 다른 품사의 구분

(1) '하나'는 <우리말샘>에 그 품사가 명사와 수사로 되어 있지만 모두 수사로만 분석한다.

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

(2) 수사과 수 관형사의 구별이 애매한 경우가 있다. 다음과 같이 특이한 형식을 가진 예만을 수 관형사로 취급하고, 그 밖의 것들은 모두 수사로 분석한다. 사전에 수사과 관형사 동일 형태로 등재된 '몇'과 '몇몇' 역시 모두 수사로 분석한다. 이는 <우리말샘>의 품사 처리와는 다른 방식임에 유의한다.

[예] 한, 한두, 한두어, 두, 두어, 두세, 두서너, 세, 석, 서, 서너, 네, 너, 닉

(3) 값이 없는 수를 나타내는 '영', '공'은 사전에 명사로 올라 있으나 수사로 분석한다.

(4) 외국어로 된 말인 '제로, 원, 투, 쓰리' 등은 우리말 단위 명사와 잘 결합하지 않는 등 일반적인 수사와 다른 성격을 지니므로 일반명사로 분석한다.

#### 3) 수사의 분리와 통합

(1) 숫자와 수사가 함께 쓰여 금액을 나타내는 말은 각각을 분리하여 분석한다.

1억	1/SN+억/NR
----	-----------

(2) 금액이나 소수 등 숫자와 기호가 결합된 형태는 하나의 단위로 통합하여 분석한다.

1,500원	1,500/SN+원/NNB
--------	----------------

(3) 복수의 수사가 한 어절 내에 나타날 때에는 전체를 통합해서 분석한다.

오천삼십사	오천삼십사/NR
공일공일이삼사오륙칠팔	공일공일이삼사오륙칠팔/NR

[참고] ‘일일구’, ‘일일이’ 등 <우리말샘>에 명사로 올라 있는 것은 명사로 분석한다.

(4) 순서를 나타내는 ‘제일, 제이’ 등은 접두사 ‘제-’와 수사의 결합으로 분석한다.

제일	제/XPN+일/NR
제이	제/XPN+이/NR

(5) 순서를 나타내는 ‘첫째, 둘째’ 등에 포함된 접미사 ‘-째’는 분석하지 않는다.

첫째	첫째/NR
둘째	둘째/NR

## 3.2. 용언

용언은 동사, 형용사, 지정사를 가리킨다. 용언 범주에서는 분석 대상이 본용언일 경우에만 동사와 형용사로 구분하여 표시하고, 보조용언은 보조동사와 보조형용사를 구분하지 않고 ‘VX’로 분석한다.

또한 학교 문법에서 서술격조사로 다루는 ‘이다’는 조사의 범주에 넣지 않고 용언의 하위범주인 ‘지정사’로 분석한다. 지정사는 다시 긍정지정사(VCP)와 부정지정사(VCN)로 세분된다.

‘영하다’, ‘수처하다’ 등 <우리말샘>에 등재되지 않은 용언은 그 활용형을 살펴 동사(VV) 혹은 형용사(VA)로 구분하여 분석한다.

### 3.2.1. 동사(VV)

동사는 사물의 움직임이나 작용을 나타내는 용언을 말한다. 동사는 일반적으로 목적어의 필요성 여부에 따라 자동사, 타동사로 나누기도 하지만 세분하지 않고 모두 ‘VV’로 표시한다.

#### 1) 동사의 종류

##### (1) 자동사

- : 오다, 가다, 웃다, 불다, 피다, 쏘다, ...
- : 보이다, 먹히다, 쫓기다, 들리다, ...

(2) 타동사

: 먹다, 듣다, 읽다, 입다, 부르다, ...  
: 앉히다, 눕히다, 남기다, 웃기다, ...  
: 먹이다, 읽히다, 지우다, 맡기다, ...

2) 동사 판단의 근거

(1) 평서형에 어미 ‘-ㄴ다/는다’가 결합된다.

[예] 간다, 온다, 먹는다, 듣는다

(2) 명령형 ‘-어라’, 청유형 ‘-자’, ‘-읍시다’ 등이 결합할 수 있다.

[예] 먹어라 (명령형)

먹자 / 먹읍시다 (청유형)

### 3.2.2. 형용사(VA)

형용사는 사물의 성질이나 상태를 나타내는 용언을 가리킨다.

1) 형용사의 종류

예쁘다, 착하다, 크다, 작다, 따뜻하다, 많다 등

2) 형용사 판단의 근거

(1) 평서형에 어미 ‘-다’가 결합된다.

[예] 크다, 작다, 예쁘다, 착하다 등

(2) 명령형 ‘-어라’, 청유형 ‘-자’, ‘-읍시다’ 등이 결합할 수 없다.

[예] \*착해라 (명령형)

\*착하자 / \*착합시다. (청유형)

3) 형용사와 동사의 구분

(1) 사전에 형용사로 등재된 단어가 동사와 같은 활용을 보일 때가 있다. 그러나 그럴 때에도 사전을 따라 형용사(VA)로 분석한다.

[예] 현실과 동떨어지는 문제가 있다. [동떨어지/VA+는/ETM]

→ ‘동떨어지다’는 <우리말샘>에 형용사로 등재되어 있다. 위의 예에서는 관형형 어미 ‘-는’ 과 결합하여 동사와 같은 활용 양상을 보여 주고 있으나, 본 지침에서는 이러한 경우에도 <우리말샘>의 품사를 따라 형용사로 분석한다.

(2) ‘못하다’는 <우리말샘>에 보조용언, 형용사, 동사 모두로 등재되어 있다. 따라서 용법에 맞게 품사를

구별하여 분석해야 한다.

[예] 노래를 못한다.	[못하/VV+ㄴ 다/EF+./SF]
[예] 맛이 예전만 못하다.	[못하/VA+다/EF+./SF]
못해도 열 명은 올 것이다.	[못하/VA+아도/EC]
[예] 밥을 먹지 못한다.	[못하/VX+ㄴ 다/EF+./SF]
웁지 못하다.	[못하/VX+다/EF+./SF]
보다 못해 간섭을 했다.	[못하/VX+아/EC]

#### 4) 형용사 '있다'와 동사 '있다'의 구분

'있다'는 동사 용법과 형용사 용법을 모두 가지고 있다. '있다'는 대개의 경우 형용사로 쓰이는 것으로 보아, '있다'가 동사로 쓰였다는 적극적인 증거가 있을 때에만 동사로 분석하고 나머지 경우는 형용사로 분석한다.

'-고 있다', '-어 있다' 형으로 쓰여 앞의 사태가 진행/지속되고 있음을 나타내거나 앞의 사태가 끝나고 그 결과가 유지되고 있음을 나타낸다면 그때의 '있다'는 보조용언(VX)임에 유의한다.

##### (1) 형용사 '있다'의 특징

- ① '존재하다', '누가 어떤 자격으로 있다'의 뜻을 갖는 것은 형용사이다.

[예] 신이 있다 / 날지 못하는 새도 있다  
그는 지금 대기업의 과장으로 있다.

- ② '수 있다', '바 있다', '적이 있다' 구성의 '있다'도 모두 형용사이다.

- ③ '~에 있어서' 구성의 '있다'도 형용사이다.

[예] 인간에게 있어서 중요한 것은 사랑이다.

- ④ 내포문에서 '있다'가 쓰였을 때에는, 종결형으로 바꾸었을 때 '있다'를 사용하여 표현할 수 있는 경우면 모두 형용사로 판단한다.

[예] 서울광장에서 있었던 콘서트가 그 예이다.  
→ '서울광장에서 콘서트가 있다.'로 바꾸어도 문장이 성립하므로 형용사로 판단한다.

##### (2) 동사 '있다'의 특징

- ① '얼마의 시간이 경과하다'의 뜻일 때에는 동사이다.

[예] 10분 있다 만나자. / 얼마 안 있어 기다리던 시간이 왔다.

- ② '-고 싶다', '-려(고) 하다'는 주로 동사와 결합하므로, 이와 결합한 '있다'는 동사로 판단한다.



[예] 나도 그 자리에 있고 싶다. 나도 여기 있으려고 한다.

- ③ ‘있은 지’, ‘있은 후’ 등에서 나타나는 ‘있은’의 ‘있-’은 동사로 판단한다. ‘-은’이 결합하여 과거를 나타내는 것이 동사의 특성이기도 하고, ‘-니 지’, ‘-니 후’ 등도 주로 동사와 결합하여 쓰이기 때문이다.

[예] 그 일이 있는 지 수일이 지났다.

### 3.2.3. 지정사

지정사는 학교 문법의 서술격 조사로 다루고 있으나 용언과 같이 활용한다는 특성을 중시하여 지정사로 품사를 분류하고 학교 문법의 ‘이다’는 긍정지정사로, ‘아니다’는 부정지정사로 하위 구분한다. 일반적으로 ‘아니다’는 형용사로 다루어지기도 하나, ‘이다’의 부정형이라는 점을 중시하여 ‘부정지정사’로 다룬다.

#### 3.2.3.1. 긍정지정사(VCP)

- 1) 긍정지정사 ‘이/VCP’를 복원해야 하는 경우

- (1) 체언에 어미가 직접 연결된 경우

철수는 훌륭한 교사다.	교사이다.	교사/NNG+이/VCP+다/EF+./SF
--------------	-------	------------------------

- (2) 조사에 어미가 직접 연결된 경우

우리가 그를 본 것은 서울에서다.	서울에서다.	서울/NNP+에서/JKB+이/VCP+다/EF+./SF
--------------------	--------	-------------------------------

- (3) ‘-였다’

그 당시 나는 아이였다.	아이였다.	아이/NNG+이/VCP+였/EP+다/EF+./SF
---------------	-------	-----------------------------

- (4) 어미 ‘-라도, -라서’

나이가 어린 자라도 존중해 주어야 한다.	자라도	자/NNB+이/VCP+라도/EC
너는 부자라서 우릴 이해하지 못할 것이다.	부자라서	부자/NNG+이/VCP+라서/EC

- (5) 참고로 피인용문 뒤에서 ‘하-’가 생략된 채 쓰인 ‘-며’, ‘-는’ 등은 ‘하-’의 복원 없이 형태 표지를 부여한다.

얼마나 친절하냐?”며	친절하냐?”며	친절하/VA+냐/EF+?/SF+”/SS+며/EC
-------------	---------	----------------------------

얼마나 친절하냐?"는	친절하냐?"는	친절하/VA+냐/EF+?/SF+"/SS+는/ETM
-------------	---------	-----------------------------

- 따옴표에 이어지는 형태이므로 직접인용으로 볼 때 ‘-라며’와 ‘-라는’의 본래 형태는 ‘(이)라고 하며’, ‘(이)라고 하는’이다. ‘하-’ 없이 인용절에 연결어미나 전성어미가 직접 결합한다고 보기 어렵지만 이때 ‘하-’를 복원하면 분석 대상 어절을 두 어절로 분석해야 하는 문제가 발생하므로 ‘하-’의 복원 없이 피인용문에 ‘-며’나 ‘-는’ 등이 직접 결합하는 것으로 분석한다.(3.5.1.7. 참고)

### 3.2.3.2. 부정지정사(VCN)

긍정지정사 ‘이다’의 부정어로 쓰이는 ‘아니다’는 부정지정사(VCN)로 분석한다.

서술어 ‘되다, 아니다’의 두 번째 논항을 보어로 분석함에 따라 부정지정사(VCN)는 보격조사(JKC)가 결합된 논항과 더불어 분석한다.

철수는 범인이 아니다.	범인이 아니다.	범인/NNG+이/JKC 아니/VCN+다/EF+./SF
그건 사실이 아니야.	사실이 아니야.	사실/NNG+이/JKC 아니/VCN+야/EF+./SF
그 문제의 정답은 2번이 아니고 4번이다.	2번이 아니고	2/SN+번/NNB+이/JKC 아니/VCN+고/EC

### 3.2.4. 보조용언(VX)

보조용언은 보조동사와 보조형용사로 하위 구분하지 않고 보조용언(VX)로 분석한다.

#### 1) 보조용언 분석 원칙

- (1) 보조용언의 후보는 <우리말샘>에 그 쓰임이 제시되어 있어야 한다.
- (2) 보조용언 앞에는 반드시 다른 용언(본 용언)이 위치해 있어야 한다.
- (3) 보조용언이 동시에 두 개 이상이 연결되어 나타날 수도 있다.
- (4) ‘-어하다’, ‘-어지다’ 결합형처럼 본용언과 보조용언의 결합형이 <우리말샘>에 하나의 어휘로 등재되어 있으면 보조용언을 따로 분석하지 않고 전체를 하나의 용언으로 처리한다.

아이를 예뻐하고	예뻐하고	예뻐하/VV+고/EC
눈이 동그래졌다.	동그래졌다.	동그래지/VV+었/EP+다/EF+./SF

#### 2) 보조용언의 예

- (1) <우리말샘>에 등재된 보조용언의 예는 다음과 같다.

보조용언	예문	원어절	형태 분석 어절
가다	책을 다 읽어 간다.	간다.	가/VX+ㄴ 다/EF+./SF

가지다	일을 그렇게 해 가지고는 기일을 맞출 수 없다.	가지고는	가지/VX+고는/EC
계시다	손님께서 와 계십니다.	계십니다.	계시/VX+ㅁ니다/EF+./SF
나가다	정책을 추진해 나가는 과정에서 문제가 생겼다.	나가는	나가/VX+는/ETM
나다	일을 마치고 나니 상쾌하다.	나니	나/VX+니/EC
내다	힘들겠지만 잘 견뎌 내야 한다.	내야	내/VX+아야/EC
놓다	약속을 잡아 놓고 출장을 가다니	놓고	놓/VX+고/EC
달다	이번 시험 문제의 정답을 알려 다오.	다오.	달/VX+오/EF+./SF
대다	자꾸 졸라 대는 통에 그만 허락해 주고 말았다.	대는	대/VX+는/ETM
두다	남겨 둔 돈도 이제 바닥이 났다.	둔	두/VX+ㄴ/ETM
드리다	염려를 끼쳐 드려 송구하옵니다.	드려	드리/VX+어/EC
들다	도무지 내 말은 믿으려 들지 않는다.	들지	들/VX+지/EC
말다	어렵더라도 희망을 잃지 말아야 한다.	말아야	말/VX+아야/EC
먹다	나는 오늘도 약속을 잊어 먹었다.	먹었다.	먹/VX+었/EP+다/EF+./SF
못하다	그 참상을 차마 보지는 못할 것이다.	못할	못하/VX+ㄹ/ETM
버리다	음식이 다 타 버렸다.	버렸다.	버리/VX+었/EP+다/EF+./SF
보다	이제는 새벽이 오는가 보다.	보다.	보/VX+다/EF+./SF
빠지다	썩어 빠진 생선을 사오다니	빠진	빠지/VX+ㄴ/ETM
싶다	너를 보고 싶다.	싶다.	싶/VX+다/EF+./SF
쌍다	꼬치꼬치 물어 쌍는 통에 정신이 없었다.	쌍는	쌍/VX+는/ETM
아니하다	일이 순리대로 풀리지 아니했다.	아니했다.	아니하/VX+았/EP+다/EF+./SF
않다	시간이 지나도 기차는 오지 않았다.	않았다.	않/VX+았/EP+다/EF+./SF
오다	날이 밝아 온다.	온다.	오/VX+ㄴ 다/EF+./SF
있다	그녀는 검정 옷을 입고 있었다.	있었다.	있/VX+었/EP+다/EF+./SF
주다	아버지는 아기에게 동화책을 읽어 주었다.	주었다.	주/VX+었/EP+다/EF+./SF
치우다	다섯 명이 10인분의 식사를 먹어 치웠다.	치웠다.	치우/VX+었/EP+다/EF+./SF
터지다	끓인 지 오래 되어서 라면이 불어 터졌다.	터졌다.	터지/VX+었/EP+다/EF+./SF
하다	나귀를 쉬게 하는 것이 좋겠다.	하는	하/VX+는/ETM

(2) <우리말샘>에 등재된 보조용언이 준말 형태로 나타나는 경우, 그 준말 형태도 보조용언으로 분석한다.

하고픈 거	하고픈	하/VV+고/EC+프/VX+ㄴ/ETM
벌써 가는갑다.	가는갑다.	가/VV+는가/EF+ㅁ/VX+다/EF+./SF

### 3) 보조용언과 의존명사 구성의 구분

‘양하다’, ‘체하다’, ‘척하다’, ‘듯하다’, ‘법하다’, ‘뻔하다’ 등은 <우리말샘>에서 보조용언으로 취급되고 있으나, 이들 앞에는 항상 관형어가 온다는 분포적인 특성을 중시하여 ‘의존명사+접사’ 또는 ‘의존명사+

보조용언'으로 분석한다.

아이인 양한다.	양한다.	양/NNB+하/XSV+L 다/EF./SF
그런 듯싶다.	듯싶다.	듯/NNB+싶/VX+다/EF+./SF

### 3.3. 수식언

#### 3.3.1. 관형사

관형사는 체언 앞에서 그것을 꾸미는 품사를 말한다. 관형사는 성상 관형사(MMA), 지시 관형사(MMD), 수 관형사(MMN)로 세분한다.

새 책	새	새/MMA
이 사람	이	이/MMD
한 가정	한	한/MMN

##### 3.3.1.1. 성상 관형사(MMA)

체언의 성질이나 상태를 나타내는 관형사는 성상 관형사로 분석한다.

[예] 새, 헌, 옛, 순(純), 구(舊), 주(主), 약(約), 양대(兩大), 만(滿) 10세, 단(單), 총(總)

##### 1) 관형사의 구분

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

##### ① 관형사, 명사 통용

전 학기에 장학금을 받았다.	전	전/MMD
그 사람을 전에 본 적이 있다.	전	전/NNG

##### ② 관형사, 부사 통용

단 세 명에서 그 일을 꾸몄다.	단	단/MMA
단, 그 일은 해서는 안 된다.	단,	단/MAJ+./SP

##### ③ 관형사, 명사, 부사 통용

이내 마음을 어찌 알리요.	이내	이내/MMD
아침 들판에 이내가 끼었다.	이내	이내/NNG
그는 이내 떠나갔다.	이내	이내/MAG

(2) '성상, 지시, 수' 중 어느 한 쪽으로 보기 힘든 관형사는 모두 '성상 관형사'로 분석한다.

### 3.3.1.2. 지시 관형사(MMD)

‘이, 그, 저’와 같이 발화 현장이나 문장 밖에 존재하는 대상을 가리키는 관형사를 지시 관형사로 분석한다.

‘어느, 무슨, 웬’과 같이 정해지지 않은 것을 대신하는 관형사, ‘이내 신세’의 ‘이내’와 같이 인칭 의미를 나타내는 관형사도 지시 관형사로 분석한다.

이 밖에 ‘귀(貴)’와 ‘본(本)’은 청자 측과 화자 측을 지시하고 ‘동(同)’은 공간을, ‘현(現)’과 ‘전(前)’은 시간을 지시한다는 점에서 지시 관형사로 볼 수 있다.

[예] 이, 그, 저, 요, 고, 조, 이런, 그런, 저런, 다른, 타(他), 어느, 무슨, 웬, 어떤, 아무, 아무런, 귀(貴), 본(本), 동(同), 현(現), 전(前), 모(某), 그까짓, 각(各), 매(每), 오른, 왼

### 3.3.1.3. 수 관형사(MMN)

체언 앞에서 사물의 수량이나 차례를 나타내는 관형사를 수 관형사로 분석한다. 단 ‘다섯, 여섯’ 등 수사과 수 관형사의 형태가 동일한 경우에는 수사로 분석한다.

[예] 한, 두, 세/서/석, 네/너/넉, 다섯, 여섯, 스무, 한두, 두세, 서너, 두서너, 일이, 이삼, 삼사, 여러, 모든, 온, 온갖, 갓은, 전(全), 첫, 양(兩)

- 1) 복수의 수사와 수 관형사가 한 어절 내에 나타날 때에는 전체를 통합해서 수 관형사로 분석한다.

스물한	스물한/MMN
십수	십수/MMN

- 2) ‘한’은 맥락을 살펴 수 관형사 또는 성상 관형사로 분석한다.

예문	원어절	형태 분석 어절	의미
책 한 권	한	한/MMN	하나
한 마을에 효자가 살고 있었다.	한	한/MMN	어떤
동생과 나는 한 이불을 덮고 잔다.	한	한/MMN	같은
한 20분쯤 걸었다.	한	한/MMA	대략

- 3) 수사와 관형사의 품사 통용은 인정하지 않는다. 따라서 다음과 같이 ‘다섯’은 모든 환경에서 중의성 없이 ‘수사’로만 분석된다.

다섯이 먹기에 충분하다.	다섯	다섯/NR+이/JKS
다섯 명이 앉아 있었다.	다섯	다섯/NR

### 3.3.2. 부사

부사는 주로 용언을 꾸며서 그 뜻을 더 세밀하고 분명하게 해 주는 품사를 말한다. 여기서는 부사를 세분하지 않고, 일반부사와 접속부사로만 나누기로 한다.

아주 바쁘다.	아주	아주/MAG
일이 늘었다. 그래서 바쁘다.	그래서	그래서/MAJ

#### 3.3.2.1. 일반부사(MAG)

일반부사는 성상부사, 지시부사, 부정부사로 나누는 경우도 있으나 모두 일반부사(MAG)로 분석한다.

공이 잘 구른다.	잘	잘/MAG
이리 가까이 오게.	이리	이리/MAG
날씨가 안 춥다.	안	안/MAG

한 문장 안에 여러 일반부사가 사용된 경우 각각의 일반부사를 분리하여 분석한다.

저리 잘 안 먹는 아이는 처음 본다.	저리 잘 안	저리/MAG 잘/MAG 안/MAG
----------------------	--------------	--------------------------

#### 1) 일반부사와 다른 품사와의 구분

- (1) 일반부사는 종종 일반명사와 동일 형태를 띠고 있어 구분이 어려운 경우가 있다. 이들은 뒤에 조사가 결합하느냐의 여부와 문맥에서 후행 명사를 수식하느냐의 여부에 따라 부사와 명사로 분석될 수 있다.

너의 진짜 속셈이 무엇인지 말해 봐라.	진짜	진짜/NNG
그 수학 문제는 진짜 어려웠다.	진짜	진짜/MAG
지금에 공부하기 딱 좋은 때이다.	지금에	지금/NNG+이/JKS
나는 지금 막 집에 도착했다.	지금	지금/MAG

- (2) 부사적인 용법을 가졌음에도 불구하고 일반부사가 아닌 일반명사로만 <우리말샘>에 등재되어 있는 단어는 오로지 일반명사로만 분석한다.

[예] 구석구석, 여기저기, 오랫동안, 이곳저곳, 좌우간, 처음, 최근

- (3) 일반부사로 분석하기 쉬운 활용상의 불완전동사인 ‘덜달아, 더불어’는 모두 동사로 옳게 분석해야 함에 주의한다.

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

- (4) '명사+없이'는 원칙적으로 '일반명사+없이/MAG'로 태깅하지만, 아래와 같이 하나의 단어로 굳어져 사전에 등재된 경우는 '없이' 통합형 자체를 하나의 일반부사로 분석한다.

[예] 관계없이, 그지없이, 꾸밈없이, 끊임없이, 난데없이, 남김없이 등

### 3.3.2.2. 접속부사(MAJ)

접속부사는 앞 문장의 의미를 뒷 문장에 이어 주면서 꾸미는 부사로, <우리말샘>에 등재된 접속부사만 접속부사(MAJ)로 분석한다.

[예] 그리고, 그러나, 그러므로, 그러면, ...

#### 1) 접속 부사와 용언의 활용형과의 구분

- (1) 접속부사는 종종 용언의 활용형으로도 쓰일 수 있으므로 주의한다.

그래서 마지막에는 조심하라고 했지?	그래서	그래서/MAJ
상황이 그래서 영희가 결석을 했구나.	그래서	그렇/VA+어서/EC

- (2) '그래도'는 용언의 활용형일 수도 있고 접속부사일 수도 있다. 동사 '그러다'의 활용형인지 형용사 '그렇다'의 활용형인지가 불분명하고 두 가지 해석이 모두 가능할 때는 형용사 '그렇다'의 활용형으로 판단한다.

그래도 그러면 안 되지.	그래도	그래도/MAJ
(누가) 그래도	그래도	그러/VV+어도/EC
(상황이) 그래도	그래도	그렇/VA+어도/EC

- (3) '그리고 나서'의 분석

그리고 나서	그리고	그리하/VV+고/EC
--------	-----	-------------

- (4) '그런데도'의 분석

그런데도	그런데도	그렇/VA+ㄴ 데/EC+도/JX
------	------	-------------------

- (5) '그러면서'의 분석

그러면서	그러면서	그러/VV + 면서/EC
------	------	---------------

#### 2) 접속부사와 일반부사의 구분

'곧, 즉, 또, 오히려, 및, 또는, 혹은' 등 <우리말샘>에 접속부사로 설명되지 않은 말들은 일반부사(MAG)로 분석한다.

월요일 또는 수요일	또는	또는/MAG
------------	----	--------

### 3) 접속부사와 감탄사의 구분

상대방의 말에 맞장구칠 때 쓰이는 '그러니까(요)'도 접속부사로 처리한다. 접속부사로부터 용법의 변화가 발생한 것으로 보이는 사례도 있으나 뒷말이 생략된 것으로 볼 수 있는 사례도 있음을 고려한 것이다.

A: 날씨가 너무 추워졌어요. B: 그러니까요.	그러니까요.	그러니까/MAJ+요/JX
-------------------------------	--------	---------------

## 3.4. 독립언

### 3.4.1. 감탄사(IC)

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

[예] 그럼, 야호, 어머, 앓, 아, 예, 그래, 아니(요), 글썄, 참, 참나, 참내, 아이구, 와아, 오호, 세상에

#### 1) 감탄사와 체언의 구분

- (1) <우리말샘>에 명사로 등재된 단어가 단독으로 쓰여 감탄사와 같은 용법을 보일 때가 있지만 그런 경우에도 <우리말샘>의 품사를 따라 명사로 분석한다.

진짜? 대박!	대박/NNG+!/SF
---------	-------------

- (2) <우리말샘>에 감탄사로 올라 있는 단어가 조사나 지정사 앞에서 쓰인 경우, 의미론적인 따옴의 효과가 있는 표현으로 볼 수 있으므로 감탄사로 분석한다. 다만 감탄사로서의 의미와 떨어진 채 조사나 지정사 앞에서 쓰인다면 명사로 분석한다.

화이팅이에요.	화이팅/IC+이/VCP+에요/EF+./SF
---------	-------------------------

\* 감탄사가 지정사 앞에 쓰인 경우이다. 이때에도 감탄사로 분석한다.

요즘 컨디션이 메롱이에요.	메롱이에요.	메롱/NNG+이/VCP+에요/EF+./SF
----------------	--------	-------------------------

\* '메롱'은 <우리말샘>에 놀림의 감탄사로 올라 있으나, 이 예에서는 놀림의 의미에서 떨어진 채 지정사 앞에서 쓰였다. 이런 경우에는 <우리말샘>의 품사와 달리 명사로 분석한다.

- (3) '뭐'는 문맥에 따라 대명사와 감탄사의 두 가지 쓰임이 있다.

뭔지도 모른 채	뭔지도	뭐/NP+이/VCP+L 지/EF+도/JX
신문에 뭐 대단한 특종이라도 실렸습니까?	뭐	뭐/IC

#### 2) 감탄사와 관형사의 구분

- (1) <우리말샘>에 관형사로 등재된 단어가 용언의 활용형이나 부사 앞에 쓰여 관형사의 수식 대상이 불분명한 경우에는 감탄사로 분석한다.



그 아주 그냥 어떻게 보면	그/IC+아주/MAG
----------------	-------------

- (2) <우리말샘>에 관형사로 등재된 단어가 고유명, 총칭적 표현, 대명사, 서술성 명사 등의 앞에 쓰여 관형사가 이어지는 대상의 범위를 한정하는 기능을 하지 못하 경우에도 감탄사로 분석한다.

그 용산 그쪽에서 사는데	그/IC+용산/NNP
---------------	-------------

### 3) 감탄사와 부사의 구분

- (1) 사람이 입으로 직접 내는 소리를 대상으로 하되, 흥내를 내는 의도가 없는 것과 본능적인 놀람이나 느낌을 나타내는 것을 대상으로 한다.

야호! 드디어 정상이다.	야호!	야호/IC+!/SF
---------------	-----	------------

- (2) 또한 감탄사와 혼동되는 부사로서 음성상징어류의 부사어가 있는데, 이는 감탄사가 아닌 일반부사로 분석한다.

쿨럭쿨럭 기침을 했다.	쿨럭쿨럭	쿨럭쿨럭/MAG
--------------	------	----------

- (3) 동물의 울음소리 등은 감탄사가 아니라 일반부사로 분석한다.

검둥이는 멍멍 짖으며 수풀 속으로 뛰어들어갔다.	멍멍	멍멍/MAG
----------------------------	----	--------

### 4) 감탄사와 동사의 구분

‘맞아’는 문맥에 따라 ‘어떤 기억이나 생각이 갑자기 떠올랐을 때 하는 말’은 감탄사로, ‘그렇다’ 또는 ‘옳다’의 뜻을 나타내는 말은 동사로 쓰이므로 주의하여 분석한다.

맞아, 내일 아버지 제사지.	맞아,	맞아/IC,/SP
맞아, 나도 그렇게 할 거야.	맞아,	맞/VV+아/EF+/,SP

### 5) 기타 감탄사로 분석할 수 있는 대상

- (1) 욕이나 욕설을 나타내는 말은 전체를 감탄사로 분석한다.

빌어먹을!	빌어먹을/IC+!/SF
-------	--------------

- (2) 한 어절이 비정상적으로 늘어나거나 비정상적으로 늘어난 것에 다른 기호가 개입되었을 경우 분석 불능범주(NA)로 분석한다.

그러어엄	그러어엄/NA
으 ~ 어 ~ 이	으 ~ 어 ~ 이/NA

- (3) 구어에서 나타나는 담화 표지는 <우리말샘>을 참고로 하여 감탄사 표지(IC)를 부여한다. 물결표는 분석하지 않는다.

저~	저/IC
음~	음/IC

저기~	저기/IC
어~	어/IC
그~	그/IC

### 3.5. 관계언

조사는 주로 체언과 결합하여 다른 말과의 문법적 관계를 나타내거나, 특별한 뜻을 더해 주는 품사를 말한다. 조사의 전체 목록은 <우리말샘>을 따르는 것을 원칙으로 한다. 조사는 크게 격조사, 보조사, 접속 조사로 나뉘는데, 그 구분 역시 <우리말샘>을 따른다.

#### 1) 조사 결합형의 처리

한국어는 조사가 여러 개 결합하는 경우가 많은데, 조사 결합형은 아래와 같은 방식으로 세분 여부를 결정한다.

(1) 조사 결합형이 <우리말샘>에 등재되어 있지 않으면 각 조사를 분리하여 분석한다.

예문	원어절	형태 분석 어절	비고
부산에서도 대형 사고가 있었다.	부산에서도	부산/NNP+에서/JKB+도/JX	‘에서도’ 미등재
그녀와의 약속이 갑자기 잡혔다.	그녀와의	그녀/NP+와/JKB+의/JKG	‘와의’ 미등재

(2) 조사 결합형이 <우리말샘>에 등재되어 있으면, 사전의 뜻풀이를 참고하여 결합형 자체에 ‘격조사’나 ‘보조사’라고 풀이되어 있으면 더 분석하지 않고 하나의 조사로 둔다.

원어절	형태 분석 어절	사전 뜻풀이
에다가	에다가/JKB	일정한 위치를 나타내는 격조사. 격조사 ‘에’에 보조사 ‘다가’가 결합한 말이다.

(3) 만약 조사 결합형이 <우리말샘>에 등재되어 있는데 ‘어떤 조사와 어떤 조사가 결합한 말’로만 풀이되어 있으면 두 개의 조사로 분리하여 분석한다.

원어절	형태 분석 어절	사전 뜻풀이
에는	에/JKB+는/JX	부사격 조사 ‘에’에 보조사 ‘는’이 결합한 말. 강조와 대조의 뜻을 나타내는 조사이다.

#### 3.5.1. 격조사(JK)

격조사는 체언과 다른 성분 간의 일정한 문법 관계를 나타내는 조사이다.

### 3.5.1.1. 주격조사(JKS)

선행 체언으로 하여금 주어가 되게 하는 조사이다.

#### (1) 주격조사의 예

이/가	산이 보인다.	산이	산/NNG+이/JKS
	우리 둘이 갈게.	둘이	둘/NR+이/JKS
께서	선생님께서 오신다.	선생님께서	선생님/NNG+께서/JKS
(이)서	둘이서 그 일을 꾸몄다고?	둘이서	둘/NR+이서/JKS
	혼자서 그 일을 꾸몄다고?	혼자서	혼자/NNG+서/JKS
께오서	부대장님께오서	부대장님께오서	부대장/NNG+님/XSN+께오서/JKS
께옵서	황제께옵서 드나드신다.	황제께옵서	황제/NNG+께옵서/JKS

(2) '이서'의 경우, <우리말샘>에서는 '이'를 접미사로, '서'를 주격조사로 보고 있으나 여기에서는 '이서' 전체를 주격조사로 본다.

(3) 앞말을 지정하여 강조하는 뜻을 나타내는 보조사 '이/가'는 주격조사 '이/가'와 형태가 같지만 보조사로 분석한다.

예쁘지가 않다.	예쁘지가	예쁘/V+A+지/EC+가/JX
----------	------	------------------

### 3.5.1.2. 보격조사(JKC)

선행 체언으로 하여금 서술어 '되다, 아니다'의 보어가 되게 하는 조사이다. '되다, 아니다' 앞의 주어가 아닌 요소에 결합한 '이/가'는 보격조사로 분석한다.

이/가	얼음이 물이 되었다.	물이	물/NNG+이/JKC
	철수는 범인이 아니다.	범인이	범인/NNG+이/JKC

### 3.5.1.3. 목적격조사(JKO)

선행 체언으로 하여금 목적어가 되게 하는 조사이다.

ㄹ/을/를	나는 바람소리를 들었다.	바람소리를	바람소리/NNG+를/JKO
-------	---------------	-------	----------------

강조하는 뜻을 나타내는 보조사 'ㄹ/을/를'은 목적격조사 'ㄹ/을/를'과 형태가 같지만 보조사로 분석한다.

ㄹ/을/를	밥을 먹질 않는다.	먹질	먹/VV+지/EC+ㄹ/JX
-------	------------	----	----------------

### 3.5.1.4. 관형격조사(JKG)

선행 체언으로 하여금 관형어가 되게 하는 조사이다.

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

구어에서 '의'가 '에'로 발음되어 '에'로 전사한 경우에도 관형격조사(JKG)로 분석한다.

에	우리에 문제가 바로 그거야.	우리에	우리/NP+에/JKG
---	-----------------	-----	-------------

### 3.5.1.5. 부사격조사(JKB)

선행 체언으로 하여금 부사어가 되게 하는 조사이다.

(으)로	망치로 못을 박아야지.	망치로	망치/NNG+로/JKB
(으)로서	장관으로서 책임을 다해야 한다.	장관으로서	장관/NNG+으로서/JKB
(으)로써	돌로써 지붕을 만든다고?	돌로써	돌/NNG+로써/JKB
같이	바보같이 웃고 다닌다.	바보같이	바보/NNG+같이/JKB
더러	나더러 이것도 하라고 한다.	나더러	나/NP+더러/JKB
랑	너랑 많이 닮았다.	너랑	너/NP+랑/JKB
(으)로부터	TV로부터 받는 영향력이 너무 크다.	TV로부터	TV/SL+로부터/JKB
마냥	기영이마냥 놀 수만은 없다.	기영이마냥	기영이/NNP+마냥/JKB
마따나	네 말마따나 나도 그래야 한다.	말마따나	말/NNG+마따나/JKB
만큼	눈물만큼 콧물도 흐른다니까.	눈물만큼	눈물/NNG+만큼/JKB
보고	영자보고 놀자고 좀 해라.	영자보고	영자/NNP+보고/JKB
보다	직관보다는 논리가 동원돼야 한다.	직관보다는	직관/NNG+보다/JKB+는/JX
에	나는 너에 대해 아무것도 모른다.	너에	너/NP+에/JKB
에게	너에게 말하기 싫다.	너에게	너/NP+에게/JKB
에게서	나는 철수에게서 그 말을 들었다.	철수에게서	철수/NNP+에게서/JKB
에서	집에서 학교까지 너무 멀다.	집에서	집/NNG+에서/JKB
에서부터	연구소에서부터 가게까지는 너무 멀다.	연구소에서부터	연구소/NNG+에서부터/JKB
와/과	경미와 함께 다닌다면,	경미와	경미/NNP+와/JKB
처럼	사람처럼 행동하는 동물이 있다.	사람처럼	사람/NNG+처럼/JKB
하고	그 일하고 관련된 사람은 아무도 없다.	일하고	일/NNG+하고/JKB

### 3.5.1.6. 호격조사(JKV)

주로 사람을 가리키는 체언 뒤에 연결되어 그것으로 하여금 부름의 대상이 되게 하는 조사이다.

아	호동아! 이제 그만 일어나거라.	호동아!	호동/NNP+아/JKV+!/SF
야	철수야! 밥 먹어라.	철수야!	철수/NNP+야/JKV+!/SF

여	주여, 우리에게 힘을 주소서.	주여,	주/NNG+여/JKV+./SP
(이)시여	신이시여! 우리를 저버리지 마소서.	신이시여!	신/NNG+이시여/JKV+!/SF

호격조사와 어말어미는 구분해서 분석해야 한다.

저기 오는 것이 철수야.	철수야.	철수/NNP+이/VCP+야/EF+./SF
---------------	------	------------------------

### 3.5.1.7. 인용격조사(JKQ)

피인용문을 동사에 대한 부사적 성분으로 도입하는 조사이다.

라고	“문제가 심각하다.”라고 보고했다.	심각하다.”라고	심각하/VA+다/EF+./SF+”/SS+라고/JKQ
이라고	가게에는 “휴가 중”이라고 메모가 붙어 있었다.	중”이라고	중/NNB+”/SS+이라고/JKQ
라	“밥 먹었어.”라는 말에 안심했다.	먹었어.”라는	먹/VV+었/EP+어/EF+./SF+”/SS+라/JKQ+는/ETM
이라	문에는 “개 조심”이라 쓰여 있었다.	조심”이라	조심/NNG+”/SS+이라/JKQ
고	그는 “이제 가도 좋다.”고 말했다.	좋다.”고	좋/VA+다/EF+./SF+”/SS+고/JKQ

(1) 인용격조사에는 ‘(이)라고’, ‘(이)라’, ‘고’가 있다. 인용격조사는 연결어미와 구별하기 어려운 경우가 있으므로 주의한다.

철수는 영희가 좋다고 말했다.	좋다고	좋/VA+다/EF+고/JKQ
철수는 자기가 학생이라고 말했다.	학생이라고	학생/NNG+이/VCP+라/EF+고/JKQ
철수는 “다음 주에 놀러 가도 좋다”고 말하였다.	좋다”고	좋/VA+다/EF+”/SS+고/JKQ
“집에 간다.”라는 말에 놀랐다.	간다.”라는	가/VV+L 다/EF+./SF+”/SS+라/JKQ+는/ETM
“밥을 먹고”라면서	먹고”라면서	먹/VV+고/EC+”/SS+라/JKQ+면서/EC

\* 따옴표 안에 문장부호가 없고 따옴표 뒤 ‘(이)라고’ 대신 ‘고’가 쓰인 사례는 신문 텍스트에서 많이 발견된다.

(2) 직접인용에서 따옴표가 생략된 경우에도 직접 인용으로 분석한다.

집에 간다라고 했다.	간다라고	가/VV+L 다/EF+라고/JKQ
집에 간다라는 말	간다라는	가/VV+L 다/EF+라/JKQ+는/ETM

\* 인용격조사 ‘(이)라고’에 쓰인 ‘이’는 긍정지정사(VCP)가 아닌 인용격조사의 일부이다. 따라서 긍정지정사 ‘이’를 복원하지 않음에 주의해야 한다.

(3) 명사 뒤에 따옴표와 ‘라고/이라고’, ‘라/이라’가 이어지는 경우에도 따옴표 뒤의 ‘라고/이라고’, ‘라/이라’를 인용격 조사로 분석한다.

“그것이 우리의 목표”라고 말했다.	목표”라고	목표/NNG+”/SS+라고/JKQ
“바보”라며	“바보”라며	”/SS+바보/NNG+”/SS+라/JKQ+며/EC

- (4) 간접인용은 종결어미와 인용격 조사를 분리하여 분석하고, 종결어미가 ‘-라’로 실현된 경우 긍정지정사를 복원한다.

친구는 술이 좋다고 말했다.	좋다고	좋/VA+다/EF+고/JKQ
팀장은 휴가 중이라고 결정을 미뤘다.	중이라고	중/NNB+이/VCP+라/EF+고/JKQ
철수는 자기는 반대라고 말했다.	반대라고	반대/NNG+이/VCP+라/EF+고/JKQ
아내는 술을 전혀 못 마시냐고 물었다.	마시냐고	마시/VV+냐/EF+고/JKQ
엄마는 아이에게 빨리 밥을 먹으라고 했다.	먹으라고	먹/VV+으라/EF+고/JKQ
아이는 소풍을 가자고 떼를 쓴다.	가자고	가/VV+자/EF+고/JKQ

- (5) ‘-라며’와 ‘-라는’과 같이 인용격조사를 포함한 ‘고 하’가 축약된 형태는 종결어미 뒤에 연결어미 또는 관형사형전성어미가 바로 결합된 것으로 분석한다.

친구는 음식이 맛있다며 계속 먹었다.	맛있다며	맛있/VA+다/EF+며/EC
음식이 맛있다는 말에 모두 놀랐다.	맛있다는	맛있/VA+다/EF+는/ETM
철수는 여기가 자기 학교라며 웃었다.	학교라며	학교/NNG+이/VCP+라/EF+며/EC
철수는 그곳이 자기 학교라는 말을 안 했다.	학교라는	학교/NNG+이/VCP+라/EF+는/ETM
술을 전혀 못 마시냐며 음료를 권했다.	마시냐며	마시/VV+냐/EF+며/EC
술을 전혀 못 마시냐는 질문을 종종 듣는다.	마시냐는	마시/VV+냐/EF+는/ETM
엄마는 빨리 밥을 먹으라며 재촉했다.	먹으라며	먹/VV+으라/EF+며/EC
아이는 빨리 밥을 먹으라는 말도 듣지 않았다.	먹으라는	먹/VV+으라/EF+는/ETM
아이는 소풍을 가자며 떼를 쓴다.	가자며	가/VV+자/EF+며/EC
소풍을 가자는 말이 반가웠다.	가자는	가/VV+자/EF+는/ETM

### 3.5.2. 보조사(JX)

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

#### 1) 보조사의 예

그려	참 잘 먹습니다그려.	먹습니다그려	먹 / VV + 습 니 다 / EF + 그 려 / JX+./SF
까지(꺼정/까장)	너까지 나에게 이럴 줄이야.	너까지	너/NP+까지/JX
깨나	너도 사람깨나 울렸겠구나.	사람깨나	사람/NNG+깨나/JX
(이)나	너나 가라!	너나	너/NP+나/JX
(이)나마	빵이나마 먹어라.	빵이나마	빵/NNG+이나마/JX

ㄴ/은/는	이 종이는 어제 사 온 것이다.	종이는	종이/NNG+는/JX
ㄴ 커녕/은커녕 /는커녕	돈은커녕 먹을 쌀도 없다.	돈은커녕	돈/NNG+은커녕/JX
다	그 물건을 거기다 놓아라.	거기다	거기/NP+다/JX
다가	책상을 어디다가 둘까요?	어디다가	어디/NP+다가/JX
대로(대루)	너는 너대로 살아라.	너대로	너/NP+대로/JX
따라	오늘따라 택시도 안 잡힌다.	오늘따라	오늘/NNG+따라/JX
도/두	강아지도 주인은 알아본다.	강아지도	강아지/NNG+도/JX
(이)란	코알라란 호주에 사는 초식동물이다.	코알라란	코알라/NNG+란/JX
만	인간은 빵만으로 살 수 없다.	빵만으로	빵/NNG+만/JX+으로/JKB
밖에	그래 봐야 죽기밖에 더 하랴.	죽기밖에	죽/VV+기/ETN+밖에/JX
부터/부텀	우선 노인부터 태워라.	노인부터	노인/NNG+부터/JX
뿐	가진 건 고작 집 한 채뿐.	채뿐	채/NNB+뿐/JX
(이)야	그가 인간성이야 그만이지.	인간성이야	인간성/NNG+이야/JX
요	나는요 그림을요 예쁘게 그립니다.	나는요	나/NP+는/JX+요/JX
조차	이젠 집조차 빼앗기는	집조차	집/NNG+조차/JX
치고	값싼 물건치고 쓸 만하다.	물건치고	물건/NNG+치고/JX

## 2) 보조사 '요'의 분석

(1) 구어에서 받침 있는 말 뒤에서 '요' 대신 쓰이는 '이요'는 보조사로 분석한다.

A: 넌 머 먹을래? B: 전 밥이요.	밥이요.	밥/NNG+이요/JX+./SF
-----------------------	------	------------------

(2) '종결어미+요(보조사)'는 <우리말샘>에 '어미'로 등재되어 있는 '어요, 지요, 래요' 등을 제외하고 모두 원래의 범주인 종결어미와 보조사로 분리하여 분석한다.

우리 집에 갈까요?	갈까요?	가/VV+ㄹ까/EF+요/JX+?/SF
어디서 저녁 먹나요?	먹나요?	먹/VV+나/EF+요/JX+?/SF
빨리 공부해야지요.	공부해야지요.	공부하/VV+아야지/EF+요/JX+./SF

철수는 이제 집에 간대요.	간대요.	가/VV+ㄴ대/EF+요/JX+./SF
----------------	------	----------------------

\* <우리말샘>에 '-ㄴ대요'가 등재되어 있지만, 이 예에서처럼 인용의 의미를 가지며 '-ㄴ다 고 해요'가 줄어든 말인 경우에는 '어미'가 아니라 '줄어든 말'로서 올라 있다. <우리말샘>에 '어미'로 등재된 경우가 아니므로 종결어미와 보조사로 분리하여 분석한다.

선생님, 철수는 장난친대요.	장난친대요.	장난치/VV+ㄴ대요/EF+./SF
-----------------	--------	--------------------

\* 이 예에서는 '-ㄴ대요'가 인용의 의미 없이 남에게 일러바치거나 남을 놀리는 의미로 사용되었다. 이러한 용법의 '-ㄴ대요'는 <우리말샘>에 '어미'로 등재되어 있으므로, '-ㄴ대요' 전체를 종결어미로 처리한다.

(3) '비종결어미+요(보조사)'는 통합하지 않고 각각 분석해 준다.

제가 몸이 좀 아파서요 지각을 했어요.	아파서요.	아프/VA+아서/EC+요/JX+./SF
내가요, 왜요?	내가요, 왜요?	내/NP+가/JKS+요/JX, 왜/MAG+요/JX+?/SF

### 3) '이'형 조사와 지정사 '이다'의 활용형의 구분

(1) '이-' 뒤에 '-시-'나 '-었-' 등의 선어말어미가 결합할 수 있으면 그 뒤의 요소는 어미이다. '이-' 뒤에 선어말어미가 결합할 수 없으면 전체가 '이'를 포함하는 조사이다.

(2) '체언+이-'의 주어를 상정할 수 있으면 그 뒤의 요소는 어미이다. '체언+이-'의 주어를 상정할 수 없으면 전체가 '이'를 포함하는 조사이다.

학생이라도 지원할 수 있습니다.	학생이라도	학생/NNG+이/VCP+라도/EC
-------------------	-------	--------------------

cf) 학생이시라도 지원하실 수 있습니다.

cf) [철수가 학생이라도] 지원할 수 있습니다.

사람이 부족하니 선생님이라도 빨리 오세요.	선생님이라도	선생님/NNG+이라도/JX
-------------------------	--------	----------------

cf)\*선생님이시라도 빨리 오세요.

cf) \*[당신이 선생님이라도] 빨리 오세요.

### 4) 보조사와 형태가 유사한 격조사, 어미와의 구분

(1) 다음과 같이 보조사와 격조사, 어미의 형태가 동일한 경우 분석에 주의해야 한다.

(이)란	많은 사람들이 하루 종일 자는 동물이 코알라란 사실을 모른다.	코알라란	코알라/NNG+이/VCP+란/ETM
(이)란	코알라란 매우 귀여운 동물이다.	코알라란	코알라/NNG+란/JX
(이)나	밥이나 빵을 먹도록 해라.	밥이나	밥/NNG+이나/JC
(이)나	밥이나 먹자.	밥이나	밥/NNG+이나/JX
나	그가 비록 열심히 하나 능력은 부족하다.	하나	하/VV+나/EC
나	어제 내가 술을 마셨나?	마셨나?	마시/VV+었/EP+나/EF+?/SF
(이)야	철수야 당연히 그 일을 할 수 있지.	철수야	철수/NNP+야/JX
야	내가 좋아하는 것은 철수야.	철수야.	철수/NNP+이/VCP+야/EF+./SF
야	철수야! 부르는 소리	철수야!	철수/NNP+야/JKV+!/SF
(이)요	밥을 먹다가요	먹다가요	먹/VV+다가/EC+요/JX
이요	밥이요 빵이요	밥이요	밥/NNG+이/VCP+요/EC

### 3.5.3. 접속조사(JC)

접속조사는 두 단어를 같은 자격으로 이어 주는 구실을 하는 조사를 말한다.



### 1) 접속조사의 예

와	그 아주머니는 딸기와 사과를 샀다.	딸기와	딸기/NNG+와/JC
과	그 기계는 사람과 컴퓨터를 구별하지 못한다.	사람과	사람/NNG+과/JC
나	사과나 배는 모두 몸에 좋은 과일이다.	사과나	사과/NNG+나/JC
랑	머루랑 다래랑 먹으며 청산에 살고 싶어라.	머루랑	머루/NNG+랑/JC
하고	이번 준비물로 칼하고 연필을 샀다.	칼하고	칼/NNG+하고/JC

### 2) 접속조사와 부사격조사의 구분

체언과 체언 사이에서 두 체언을 이어주는 요소는 접속조사이고, 그 외의 경우에는 부사격조사이다.

철수와 영희가 왔다.	철수와	철수/NNP+와/JC
철수와 같이 놀았다.	철수와	철수/NNP+와/JKB

## 3.6. 의존형태

### 3.6.1. 어미

#### 3.6.1.1. 선어말어미(EP)

선어말어미는 용언이 활용할 때, 어간과 어말어미 사이에 결합하여 높임법이나 시제, 양태를 나타내는 문법적인 요소이다.

#### 1) 선어말어미의 예

-겠-	그 일은 내일 처리하겠다.	처리하겠다.	처리하/VV+겠/EP+다/EF+./SF
-(으)시 -	선생님께서 손수 만드신	만드신	만들/VV+시/EP+L/ETM
-사오-	당신을 믿사오니 힘내세요.	믿사오니	믿/VV+사오/EP+니/EC
-옵-	어머님께 선물을 바치옵고	바치옵고	바치/VV+옵/EP+고/EC
-았/었-	우리가 먹었던 음식에 문제가 있다.	먹었던	먹/VV+었/EP+던/ETM
-았었/ 었었-	거기는 우리가 전에 갔었던 곳이야.	갔었던	가/VV+았었/EP+던/ETM

\* '-오리다', '-사오리다', '-더라' 등 <우리말샘>에 하나의 어미로 등재되어 있는 경우는 그 내부를 분석하지 않는다.

#### 2) 선어말어미 분석

(1) 어간 '하-' 뒤에 과거 시제 선어말어미가 결합하여 '했'의 형태로 나타나거나 '하였'의 형태로 나타

날 수 있는데, 본 분석에서는 이 경우 ‘-였-’ 형태를 인정하지 않고 모두 ‘-았-’으로 분석한다. 이외에 ‘아/어’를 포함하고 있는 모든 어미 역시, ‘하-’ 뒤에 나타난 경우에는 ‘아 X’형으로 분석한다.

(2) 다음의 선어말어미는 그 어간이 생략되었을 경우에 어간을 복원해 준다.

-겠-	이것은 그대로 두어야겠다.	두 어 야 겠다.	두/VV+어야/EC+하/VX+겠/EP+다/EF+./SF
-았/었-	철수가 그것을 가져오랬다.	가 져 오 랐다.	가져오/VV+라/EF+하/VV+았/EP+다/EF+./SF
-시-	선생님께서 가자시오.	가자시오.	가/VV+자/EF+하/VV+시/EP+오/EF+./SF

(3) 여러 개의 어미가 결합한 준말은 그 안에 분석 대상 선어말어미가 들어 있는 경우에 한해서만 복원한다.

간다는	가/VV+ㄴ 다/EF+는/ETM
간뎠어.	가/VV+ㄴ 다/EF+하/VV+았/EP+어/EF+./SF

### 3.6.1.2. 종결어미(EF)

종결어미는 용언의 어간이나 선어말어미 뒤에 연결되어 용언의 형식을 완성시키는 어미로서 한 문장을 끝맺는 역할을 한다.

#### 1) 종결어미의 예

-게	그만한 돈이 있으면 좋게.	좋게.	좋/VA+게/EF+./SF
-ㄴ 가	이것이 무엇인가?	무엇인가?	무엇/NP+이/VCP+ㄴ 가/EF+?/SF
-ㄴ 걸	이제 시작인걸.	시작인걸.	시작/NNG+이/VCP+ㄴ 걸/EF+./SF
-ㄴ 다	이건 말도 안 된다.	된다.	되/VV+ㄴ 다/EF+./SF
-나	자네 그리로 가나?	가나?	가/VV+나/EF+?/SF
-냐	키가 얼마나 크냐?	크냐?	크/VA+냐/EF+?/SF
-네	정말 큰일 났네!	났네!	나/VV+았/EP+네/EF+!/SF
-는걸	그는 벌써 갔는걸.	갔는걸.	가/VV+았/EP+는걸/EF+./SF
-는구나	앞이 잘 안 보이는구나.	보이는구나.	보이/VV+는구나/EF+./SF
-는구려	잘도 먹는구려.	먹는구려.	먹/VV+는구려/EF+./SF
-는구먼	공부를 잘하는구먼.	잘하는구먼.	잘하/VV+는구먼/EF+./SF
-는다	아이가 글을 잘 읽는다.	읽는다.	읽/VV+는다/EF+./SF
-다	그게 사실이다.	사실이다.	사실/NNG+이/VCP+다/EF+./SF
-ㄹ 게	그렇게 할게.	할게.	하/VV+ㄹ 게/EF+./SF
-ㄴ니까	이제야 옵니까?	옵니까?	오/VV+ㄴ니까/EF+?/SF

-ㅂ니다	이렇게 합니다.	합니다.	하/VV+ㅂ니다/EF+./SF
-습니까	그래도 되겠습니까?	되겠습니까?	되/VV+겠/EP+습니까/EF+?/SF
-습니다	정말 재미있습니다.	재미있습니다.	재미있/VA+습니다/EF+./SF
-ㅂ시다	다시 만납시다.	만납시다.	만나/VV+ㅂ시다/EF+./SF
-ㅂ시오	서둘러 주십시오.	주십시오.	주/VX+시/EP+ㅂ시오/EF+./SF
-으냐	물이 얼마나 깊은냐?	깊으냐?	깊/VA+으냐/EF+?/SF
-은가	그것이 좋은가?	좋은가?	좋/VA+은가/EF+?/SF
-오/으오/소	물이 깨끗하오.	깨끗하오.	깨끗하/VA+오/EF+./SF
-ㅂ디다/습디다	참 좋은 곳입디다.	곳입디다.	곳/NNB+이/VCP+ㅂ디다/EF+./SF
-거든	나는 이것이 좋거든!	좋거든!	좋/VA+거든/EF+!/SF
-ㄴ결/은결	힘이 꽤 센결.	센결.	세/VA+ㄴ결/EF+./SF
-ㄹ결/을결	모른다고 할결.	할결.	하/VV+ㄹ결/EF+./SF
-ㄹ까	이제 밥을 할까?	할까?	하/VV+ㄹ까/EF+?/SF
-다오	그가 가지고 있다오.	있다오.	있/VX+다오/EF+./SF
-다네	일을 망쳤다네.	망쳤다네.	망치/VV+었/EP+다네/EF+./SF
-다구	돈이 많다구?	많다구?	많/VA+다구/EF+?/SF
-다니까	돈이 없다니까!	없다니까!	없/VA+다니까/EF+!/SF
-냐고/느냐고	그가 누구냐고?	누구냐고?	누구/NP+이/VCP+냐고/EF+?/SF
-도다	꽃이 아름답도다.	아름답도다.	아름답/VA+도다/EF+./SF
-다니	그가 책을 읽다니!	읽다니!	읽/VV+다니/EF+!/SF
-는가	같이 가겠는가?	가겠는가?	가/VV+겠/EP+는가/EF+?/SF
-ㅂ디까/습디까	보기에 좋습디까?	좋습디까?	좋/VA+습디까/EF+?/SF
-다면서	술은 싫다면서?	싫다면서?	싫/VA+다면서/EF+?/SF
-다나	그도 가겠다나.	가겠다나.	가/VV+겠/EP+다나/EF+./SF
-렴/으렴	맘대로 해 보렴.	보렴.	보/VX+렴/EF+./SF
-려무나	책이나 읽으려무나.	읽으려무나.	읽/VV+으려무나/EF+./SF
-라	그의 이름은 길동이라.	길동이라.	길동/NNG+이/VCP+라/EF+./SF
-라니까	그 사람이 아니라니까.	아니라니까.	아니/VCN+라니까/EF+./SF
-세	일이나 하세.	하세.	하/VV+세/EF+./SF
-자꾸나	약속을 좀 미루자꾸나.	미루자꾸나.	미루/VV+자꾸나/EF+./SF
-자니까	그만 따지자니까.	따지자니까.	따지/VV+자니까/EF+./SF
-아/어/야	밥 먹어!	먹어!	먹/VV+어/EF+!/SF

-ㅁ세/음세	그날 꼭 음세.	음세.	오/VV+ㅁ세/EF+./SF
-단다	애들이 다쳤단다.	다쳤단다.	다치/VV+었/EP+단다/EF+./SF
-더라고	아까 보니 철수가 집에 가 더라고.	가더라고.	가/VV+더라고/EF+./SF

## 2) 종결어미의 분석

- (1) '종결어미+요(보조사)'는 <우리말샘>에 '어미'로 등재되어 있는 '어요, 지요, 래요' 등을 제외하고 모두 종결어미와 보조사로 분리하여 분석한다.

말씀대로 했는걸요.	했는걸요.	하/VV+았/EP+는걸/EF+요/JX+./SF
------------	-------	---------------------------

- (2) '-세요'는 다음과 같이 선어말어미를 분석한다.

어서 출근하세요.	출근하세요.	출근하/VV+시/EP+어요/EF+./SF
-----------	--------	------------------------

- (3) '-조'는 축약형을 그대로 태깅한다. 단, 종결어미 '-어야지'와 '요'가 결합하여 '-어야죠' 형식으로 나왔을 때는 '-지'와 '요'를 분리한다.

어서 출근하죠.	출근하죠.	출근하/VV+죠/EF+./SF
어서 출근해야죠.	출근해야죠.	출근하/VV+아야지/EF+요/JX+./SF

- (4) "앞의 사실을 청자가 이미 알고 있음"을 나타내는 '잖'은 어말어미와 결합하여 표지를 부여한다. 단, '하' 생략과 함께 '지 않'이 줄어들어서 나타난 '잖' 형은 '하'와 함께 '지 않'을 복원하여 분석한다.

저 오늘 일찍 일어났잖아요.	일어났잖아요.	일어나/VV+았/EP+잖아요/EF+./SF
제가 갔잖습니까.	갔잖습니까.	가/VV+았/EP+잖습니까/EF+./SF
녹록잖은 일이다.	일이다.	녹록하/VA+지/EC+않/VX+은/ETM

- (5) '-려고'는 <우리말샘>에서 의심과 반문의 용법으로만 종결어미 자격을 갖는 것으로 등재되어 있다. 하지만 아래와 같이 뒤에 생략된 말 없이 주어의 의도만을 밝히며 문말에서 쓰이는 '-려고'는 종결어미 용법으로 볼 수 있으므로 종결어미로 분석한다.

나는 오늘 집에 일찍 가려고.	가려고.	가/VV+려고/EF+./SF
------------------	------	-----------------

- \* '-려고' 뒤에 '생각하다' 정도의 동사가 생략되어 있다. 이때는 주어의 의도가 무엇인지만을 밝히며 문말에서 쓰인 '-려고'로 볼 수 있으며, 종결어미로 분석한다.

나 요즘 매일 운동해. 살 빼려고.	빼려고.	빼/VV+려고/EC+./SF
---------------------	------	-----------------

- \* '-려고' 뒤에 '생각하다'가 생략된 것이 아니며 '운동하다'와 같이 주어의 의도를 실현하기 위한 행동을 나타내는 말이 생략되어 있다. 이때는 '-려고'가 연결어미로 쓰인 것이다.

### 3.6.1.3. 연결어미(EC)

연결어미는 용언의 어간이나 선어말어미 뒤에 결합하여 문장을 종결시키지 못하고 뒤에 오는 절을 연결시켜 주는 어미를 말한다.

#### 1) 연결어미의 예

-거나	누가 오거나 알은 체 할 것 없다.	오거나	오/VV+거나/EC
-거든	거기 가거든 김 사장이 있는지 보아라.	가거든	가/VV+거든/EC
-건대	내가 보건대, 네 말이 옳다.	보건대	보/VV+건대/EC
-건마는	말렸건마는 아직도 축축하다.	말렸건마는	말리/VV+었/EP+건마는/EC
-게	개를 굶게 하지 마라.	굶게	굶/VV+게/EC
-고	일을 하고 밥을 먹자.	하고	하/VV+고/EC
-곤	숙제한 것도 빌려가곤 한다.	빌려가곤	빌리/VV+어/EC+가/VV+곤/EC
-기에	늦게라도 왔기에 용서해 주었다.	왔기에	오/VV+았/EP+기에/EC
-ㄴ 다기에	잠시 천다기에 승낙했다.	천다기에	쉬/VV+ㄴ 다기에/EC
-ㄴ 다손/다손	입다손 치더라도 구박하지 말자.	입다손	입/VA+다손/EC
-ㄴ 들/는들	간다 한들 아주 갈까?	한들	하/VV+ㄴ 들/EC
-ㄴ 족	배가 고평족 속이 쓰리다.	고평족	고프/VA+ㄴ 족/EC
-ㄴ 지라/는지라	눈이 온지라 길이 미끄럽다.	온지라	오/VV+ㄴ 지라/EC
-나	눈이 오나 비가 오나 같다.	오나	오/VV+나/EC
-나마	맛이 좋지 못하나마 많이 드십시오.	못하나마	못하/VX+나마/EC
-는다기에	빵을 먹는다기에 주었다.	먹는다기에	먹/VV+는다기에/EC
-니까	너를 보니까 좋다.	보니까	보/VV+니까/EC
-다가	자랑하다가 망신당했다.	자랑하다가	자랑하/VV+다가/EC
-다기에	꽃이 예쁘다기에 보러 왔소.	예쁘다기에	예쁘/VA+다기에/EC
-대도	시간이 있대도 만나 주질 않는다.	있대도	있/VA+대도/EC
-더라도	가더라도 꼭 돌아와라.	가더라도	가/VV+더라도/EC
-던들	진작 알았던들 방법을 취했지.	알았던들	알/VV+았/EP+던들/EC
-든지	외모가 어떠하든지 무슨 상관인가?	어떠하든지	어떠하/VA+든지/EC
-ㄹ 뿐더러	비가 올뿐더러 바람도 분다.	올뿐더러	오/VV+ㄹ 뿐더러/EC
-ㄹ 수록	높이 올라갈수록 춥다.	올라갈수록	올라가/VV+ㄹ 수록/EC
-ㄹ 지	비가 얼마나 올지 천둥이 다 친다.	올지	오/VV+ㄹ 지/EC
-ㄹ 지라도	이길지라도 명예롭지는 않다.	이길지라도	이기/VV+ㄹ 지라도/EC

-르지언정	죽을지언정 그 일은 못하겠다.	죽을지언정	죽/VV+을지언정/EC
-락	자락 깨락 잠을 설쳤다.	자락	자/VV+락/EC
-랍시고	그는 반장이랍시고 거드름만 피운다.	반장이랍시고	반장/NNG+이/VCP+랍시고/EC
-려니와	비용도 문제려니와 일꾼도 문제다.	문제려니와	문제/NNG+이/VCP+려니와/EC
-련마는	보면 반가우련마는 볼 수가 없네.	반가우련마는	반갑/VA+으련마는/EC
-며	남편은 친절하며 부인은 인정이 많다.	친절하며	친절하/VA+며/EC
-면	지옥이 존재하면 만원일 것이다.	존재하면	존재/NNG+하/XSV+면/EC
-면서	푸르면서 검은 물빛	푸르면서	푸르/VA+면서/EC
-므로	비가 오므로 가지 않겠다.	오므로	오/VV+므로/EC
-아/어	입을 막아 버렸다.	막아	막/VV+아/EC
-아도/어도	암만 봐도 모르겠다.	봐도	보/VV+아도/EC
-아서/어서	덧을 놓아서 꿩을 잡았다.	놓아고	놓/VV+아서/EC
-아야	이 일은 잘해야 한다.	잘해야	잘/VV+아야/EC
-으나	밥을 먹으나 마나이다.	먹으나	먹/VV+으나/EC
-으나마	맛은 없으나마 많이 드세요.	없으나마	없/VA+으나마/EC
-자마자	집에 오자마자 씻었다.	오자마자	오/VV+자마자/EC
-지	밥을 먹지 못했다.	먹지	먹/VV+지/EC
-지마는	비가 오지마는 가야 한다.	오지마는	오/VV+지마는/EC

## 2) 연결어미와 종결어미의 구분

(1) 연결어미와 종결어미의 형태가 동일한 경우 문맥을 통해 형태분석을 결정한다.

너는 내가 왔는데 기쁘지도 않니?	왔는데	오/VV+았/EP+는데/EC
철수가 있는 데가 어디지?	있는 데가	있/VA+는/ETM+데/NNB+가/JKS
다들 만족하는지 아무런 불평이 없다.	만족하는지	만족하/VV+는지/EC
다들 만족하는지는 모르겠다.	만족하는지는	만족하/VV+는지/EF+는/JX
너를 만난 지도 꽤 오래구나.	만난 지도	만나/VV+ㄴ/ETM+지/NNB+도/JX

(2) 통사적 구성에 나타나는 ‘-음직’은 ‘음직/EC’로 분석한다. 그러나 ‘바람직하다, 먹음직하다’ 등 사전에 등재되어 있는 단어의 내부에서 확인되는 ‘-음직’은 더 이상 분석할 수 없다는 것에 유의한다.

철수라면 외국에 갔음직 하다.	갔음직 하다.	가/VV+았/EP+음직/EC 하/VA+다/EF+./SF
어른답고 믿음직하게 행동해라.	믿음직하게	믿음직하/VA+게/EC
그것 참 먹음직스럽다.	먹음직스럽다.	먹음직스럽/VA+다/EF+./SF

그것은 매우 바람직한 일이다.	바람직한	바람직하/VA+L/ETM
------------------	------	---------------

(3) '-을까', '-는가, -은가'는 언제나 종결어미임에 유의한다.

(4) '누군가', '어딘가' 등은 '누구/NP+이/VCP+L 가/EF', '어디/NP+이/VCP+L 가/EF' 등으로 분석한다.

(5) '-을지', '-는지, -은지'는 연결어미 용법과 종결어미 용법을 모두 갖는데, 뒤에 조사가 오거나 '모르다'의 목적어 자리에서 쓰이는 경우 종결어미 용법에 해당한다는 점에 유의한다.

#### 3.6.1.4. 명사형전성어미(ETN)

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

1) 명사형전성어미의 예

-기	그 일은 정말 중요하기 때문이다.	중요하기	중요하/VA+기/ETN
-ㄴ/은	장사는 신용을 얻음이 제일이다.	얻음이	얻/VV+음/ETN+이/JKS

2) 명사형전성어미의 분석

(1) ㅁ 불규칙 용언 어간에 명사형전성어미 '-음'이 결합한 경우 '-음'이 아닌 '-ㄴ'으로 분석 한다. ㅂ 불규칙 용언 어간에 결합하는 '-음'은 '음/ETN'으로 분석한다.

아니꼬움을 견디지 못하고	아니꼬움을	아니꼬ㅁ/VA+ㄴ/ETN
김철수 지음	지음	짓/VV+음/ETN

(2) '-음, -기'가 붙은 말이 단순히 명사형이냐 아니면 굳어진 명사이냐 하는 것은 물론 문맥에 따라 결정되어야 하지만 먼저 그것이 사전에 등재되어 있느냐의 여부를 살펴보아야 한다.

책을 읽기가 어렵다.	읽기가	읽/VV+기/ETN+가/JKS
읽기 교육이 문제가 된다.	읽기	읽기/NNG

#### 3.6.1.5. 관형사형전성어미(ETM)

관형사형전성어미는 용언의 성격을 임시로 바꾸어 다른 문장 속에서 관형사적인 역할을 하게 하는 어미이다.

1) 관형사형전성어미의 예

-ㄴ/은	어제 먹은 빵에 이상이 있었다.	먹은	먹/VV+은/ETM
-는	잃어버린 물건을 찾는 일은 어렵다.	찾는	찾/VV+는/ETM
-던	이제까지 미루던 일을 오늘 해치웠다.	미루던	미루/VV+던/ETM
-ㄹ/을	나에게는 아직 처리할 일이 있다.	처리할	처리하/VV+ㄹ/ETM

## 2) 관형사형전성어미의 분석

ㄴ 불규칙 용언 어간에 관형사형 전성 어미가 결합한 경우 ‘-은, -을’이 아닌 ‘-ㄴ, -ㄹ’로 분석한다. ㄴ 불규칙 용언 어간에 결합하는 ‘-은, -을’은 ‘은/ETM, 을/ETM’으로 분석한다.

그녀의 고운 얼굴	고운	곱/VA+ㄴ/ETM
꽃밭은 매우 아름다울 것이다.	아름다울	아름답/VA+ㄹ/ETM
집을 지을 거야.	지을	짓/VV+을/ETM

## 3.6.2. 접사

접사는 복합어에서 어근에 결합하여 형식적 의미를 갖는 요소를 말한다.

본 지침의 목적은 다층위 한국어 구조 분석 말뭉치를 구축하기 위한 기본 작업으로 형태 분석 말뭉치를 구축하는 것이다. 따라서 <우리말샘>에 등재된 표제어의 내부 구조에 대해서는 가급적 분석을 하지 않는다. 표제어의 접사(접두사, 접미사) 결합 정보는 <우리말샘> 사전의 형태 정보로 제공된다.

### 1) 접사의 분리 원칙

- (1) <우리말샘> 미등재어가 하나의 이름으로 쓰이는 경우 혹은 접사 분리 시 어근이 남는다면 접사를 분리하지 않는다.

나는 아침형 인간이 아니라 밤형 인간이다.	밤형	밤/NNG+형/XSN
-------------------------	----	-------------

- (2) 복합어가 <우리말샘> 미등재어인 경우에는 <우리말샘> 등재 어휘를 참조하여 분석 여부를 결정한다.

최대형 (미등재어)	최대형	최대형/NNG
------------	-----	---------

\* 사전 등재어인 ‘최소형’을 참고하여 처리할 수 있다.

대의원회 (미등재어)	대의원회	대의원회/NNG
-------------	------	----------

\* 사전 등재어인 ‘평의원회’를 참고하여 처리할 수 있다.

- (3) 숫자, 로마자 등 기타 기호에 접사가 결합한 것은 한글 형태의 <우리말샘> 등재어의 쓰임을 검토하



여 분리 여부를 판단한다.

3분의 일	3분의 일	3/NR+분/XSN+의/JKG 일/NR
-------	-------	--------------------------

### 3.6.2.1. 체언접두사(XPN)

체언접두사가 결합한 파생어는 대체로 사전에 등재되므로 분석의 대상이 아니다.

#### 1) 체언접두사 결합형의 분석

- (1) 체언접두사는 따로 분리하지 않고 어근과 통합한 단위로 분리한다. <우리말샘>에 등재되지 않은 접두파생어도 이 기준에 따라 하나의 단위로 분석한다. 체언접두사 결합형은 어근의 형태 분석 결과에 따라 일반명사 혹은 고유명사로 분석한다.

반독재	반독재/NNG
대미	대미/NNP

- (2) <우리말샘>에 하나의 어휘로 등재되어 있지만 괄호가 어절 중간에 개입한 경우에는 구성요소 각각을 분리하여 분석한다.

대중국	대중국	대중국/NNP
대(對)중국	대(對)중국	대/XSN+(/SS+對/SH)/SS+중국/NNP
반독재	반독재	반독재/NNG
반(反)독재	반(反)독재	반/XPN + (/SS + 反/SH + )/SS + 독재/NNG

### 3.6.2.2. 명사파생접미사(XSN)

명사파생접미사는 명사나 다른 어근에 결합하여 그것이 명사의 기능을 수행할 수 있도록 만들어 주는 의존 형태이다.

#### 1) 명사파생접미사 결합형의 분석

- (1) <우리말샘>에 등재된 어휘는 명사파생접미사를 따로 떼어 분석하지 않는다.

사상적 기반	사상적	사상적/NNG
전쟁용 핵무기	전쟁용	전쟁용/NNG
교육계에 몸담아	교육계에	교육계/NNG+에/JKB

- (2) <우리말샘>에 등재되지 않는 구성에서는 명사파생접미사를 따로 떼어 분석한다.

십 원꼴	십 원꼴	십/NR 원/NNB+꼴/XSN
만 원씩	만 원씩	만/NR 원/NNB+씩/XSN
백 원짜리	백 원짜리	백/NR 원/NNB+짜리/XSN

- (3) ‘비교적’, ‘가급적’ 등 명사 혹은 부사로 쓰일 수 있는 경우를 제외하고, <우리말샘>에 관형사와 명사로 다루고 있는 명사파생접미사 ‘-적(的)’이 결합된 형태는 어근의 성격을 고려하여 명사(NNG)로만 분석한다.

예문	원어절	형태 분석 어절
문화적 차이를 실감하다.	문화적	문화적/NNG
비교적인 관점에서 바라본다.	비교적인	비교적/NNG + 이/VCP + L/ETM
그의 방은 비교적 깨끗했다.	비교적	비교적/MAG
가급적이면 빨리 가자.	가급적이면	가급적/NNG + 이/VCP+ 면/EC
가급적 도와주시기 바랍니다.	가급적	가급적/MAG

## 2) 명사와 같은 형태를 지닌 명사파생접미사의 분석

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

사람들이 우리 집에 왔다.	사람들이	사람/NNG+들/XSN+이/JKS
그들은 밥을 먹고들 싶었다.	먹고들	먹/VV+고/EC+들/XSN

- (2) ‘-님’은 문맥에 따라, ‘임’의 의미로 쓰인 경우 일반명사로, 사람의 ‘이름’이나 ‘성’ 뒤에서 쓰인 경우 의존명사로, 그 밖의 경우 명사파생접미사로 분석한다.

님과 이별하다.	님과	님/NNG+과/JKB
김철수님께서 오셨습니다.	김철수님께서	김철수/NNP+님/NNB+께서/JKS
과장님이 부르십니다.	과장님이	과장/NNG+님/XSN+이/JKS

- (3) <우리말샘>에 접미사로 등재된 1음절 한자어는 명사파생접미사(XSN)으로, 등재되지 않은 1음절 한자어는 일반명사(NNG)로 분석한다.

서울지하철 1호선	1호선	1/SN+호/NNB+선/XSN
서울초등학교 졸	졸	졸/NNG

### 3.6.2.3. 동사파생접미사(XSV)

동사파생접미사는 어근 또는 어근에 붙어서 그것을 동사로 만들어 주는 기능을 갖는 접미사로, ‘-당하다, -되다, -시키다, -하다, -받다’ 정도가 생산성을 가지고 쓰인다.

#### 1) 동사파생접미사 결합형의 분석

(1) 동사파생접미사는 어근과 묶어서 한 단위로 분석한다.

-하다	외국에서 공부하는 일이 쉬운 것은 아니다.	공부하는	공부하/VV+는/ETM
-되다	아침식사가 이미 준비되어 있었다.	준비되어	준비되/VV+어/EC
-받다	몇몇은 집세 인상을 강요받았다.	강요받았다.	강요받/VV+았/EP+다/EF+./SF
-당하다	아군이 공격당하는 데에는 이유가 있다.	공격당하는	공격당하/VV+는/ETM
-시키다	오늘 강아지를 운동시키려고 공원에 나갔다.	운동시키려고	운동시키/VV+려고/EC

(2) 동사파생접미사 결합형이 <우리말샘>에 등재되어 있지 않은 경우에도 어근과 결합하여 하나의 동사로 분석한다.

-하다	팀장은 팀원의 휴가 계획을 컨펌했다.	컨펌했다	컨펌하/VV+았/EP+다/EF
-되다	수업 자료가 업로드되어 제공된다.	업로드되어	업로드되/VV+어/EC

(3) 괄호와 함께 외국 문자나 한자로 어근의 원어 정보가 제공된 경우 동사파생접미사를 어근과 분리하여 분석한다.

스터디(study)하여	스터디/NNG+(/SS+study/NNG+)/SS+하/XSV+아/EC
업로드(upload)되어	업로드/NNG+(/SS+upload/NNG+)/SS+되/XSV+어/EC

(4) <우리말샘> 등재어가 따옴표로 어근과 접사가 분리된 경우에도 어근과 접사를 분리하여 분석한다.

그 여자는 자신의 다이어트 비법을 ‘공유’했다.	/SS+공유/NNG+/SS+하/XSV+았/EP+다/EF+./SF
----------------------------	-------------------------------------

### 3.6.2.4. 형용사파생접미사(XSA)

형용사파생접미사는 어기나 어근에 붙어서 그것을 형용사로 파생시키는 접미사로, ‘-답다, -되다, -롭다, -스럽다, -하다’ 정도가 생산성을 가지고 쓰인다.

#### 1) 형용사파생접미사 결합형의 분석

(1) 형용사파생접미사는 어근과 묶어서 한 단위로 분석한다.

-하다	건강한 신체에 건강한 정신이 깃든다.	건강한	건강하/VA+L/ETM
-되다	거짓된 말은 들통나기 마련이다.	거짓된	거짓되/VA+L/ETM
-답다	사람이 사람답게 행동해야지.	사람답게	사람답/VA+게/EC

-롭다	어려운일일수록 슬기롭게 대처하라.	슬기롭게	슬기롭/VA+게/EC
-스럽다	그녀의 사랑스러운 표정을 보거라.	사랑스러운	사랑스럽/VA+ㄴ/ETM

### 3.6.3. 어근(XR)

어근은 복합어에서 실질적인 의미를 가지는 요소이다.

#### 1) 어근과 어근 결합형의 분석

- (1) <우리말샘>에 등재되어 있는 파생어와 합성어는 하나의 단어로 분석한다. 합성어가 북한어나 방언으로 등재되어 있어도 분석하려는 말과 의미가 동일하다면 표준어와 동일하게 처리한다.

사전 표기	원어절	형태 분석 어절
정치-권력	정치권력	정치권력/NNG

- (2) <우리말샘>에 구로 등재되어 있는 말(A^B)은 어근을 분리하지 않고 하나의 단어로 분석하는 것을 원칙으로 한다.

사전 표기	원어절	형태 분석 어절
학생^운동	학생운동	학생운동/NNG

- (3) <우리말샘>에 합성어로 등재되어 있지 않은 표제어를 더 작은 요소로 분리했을 때 어근이 남거나 품사를 부여하기 어려운 요소가 남는다면, 해당 요소를 분리하지 않고 앞말 또는 뒷말과 결합하여 형태 표지를 부여한다.

당정청	당정청	당정청/NNG
-----	-----	---------

#### 2) 어근의 분석

- (1) 파생된 용언에서 보조사 등의 삽입에 의해 어근과 접사가 분리되는 경우, 분리된 각각을 어근(XR)과 접사로 분석한다.

날씨가 따듯도 하다.	따듯도 하다.	따듯/XR+도/JX 하/VA+다/EF+./SF
-------------	---------	---------------------------

- (2) 형용사의 어근 분리 시 어근 뒤에 오는 '하다'에는 형용사(VA) 표지를 부여한다. 아래와 같은 경우도 '하다'를 형용사(VA)로 본다.

영희는 키가 철수만 하다.	철수만 하다.	철수/NNP+만/JX 하/VA+다/EF+./SF
비가 올 듯도 하다.	듯도 하다.	듯/NNB+도/JX 하/VA+다/EF+./SF

- (3) 신문 기사의 제목 등에서 어근으로 문장이 끝나는 경우에도 어근 표지를 할당할 수 있다. 단 명사로의 쓰임이 가능하면 일반명사(NNG)로 분석한다.

대회 3일 차 분위기 무난	무난	무난/XR
주민 대부분 공원 조성에 만족	만족	만족/NNG

## 3.7. 기호

### 3.7.1. 일반기호

#### 3.7.1.1. 마침표, 물음표, 느낌표(SF)

- 1) 문장 끝에서 쓰인 마침표(.)와 물음표(?), 느낌표(!)는 SF 로 분석한다. 따옴표 안 인용문의 문장 끝에서 쓰인 경우도 마찬가지이다.

아이가 들판하다.	들판하다.	들판하/VA+다/EF+./SF
언제 와?	와?	오/VV+아/EF+?/SF
철수야!	철수야!	철수/NNP+야/JKV+!/SF
“아이가 착하다. 또 들판하다.”라고 말했다.	들판하다.”라고	들판하/VA+다/EF+./SF+”/SS+라고/JKQ

- 2) 아래와 같이 ‘장, 절’, ‘월, 일’ 등의 한글을 대신해 숫자가 쓰인 경우에는 위치와 관계 없이 SF 로 분석한다.

1. 서론	1.	1/SN+./SF
1.1. 연구 목적	1.1.	1/SN+./SF+1/SN+./SF
3.14. (3월 14일)	3.14.	3/SN+./SF+14/SN
www.an.com	www.an.com	www/SL+./SF+an/SL+./SF+com/SL

#### 3.7.1.2. 쉼표, 가운뎃점, 콜론, 빗금(SP)

- 1) 쉼표(.), 가운뎃점(.), 콜론(:), 빗금(/)은 SP 로 분석한다.

강아지, 고양이 등	강아지,	강아지/NNG+./SP
철수·영희	철수·영희	철수/NNP+./SP+영희/NNP
장소: 서울시 강서구	장소:	장소/NNG+:/SP
날짜: 5/4	5/4	5/SN+//SP+4/SN

- 2) 말줄임표 대신 쉼표가 쓰인 경우에는 각각의 쉼표를 따로따로 SP 로 처리한다. 말줄임표로서 마침표를 여러 개 찍는 것은 어문 규범에 부합하는 것인데 반해, 쉼표를 여러 개 찍는 것은 어문 규범에서

인정하는 말줄임표가 아님을 고려한 것이다.

안 기빠요,,,	기빠요,,,	기쁘/V/A+어요/EF+./SP+./SP+./SP
----------	--------	-----------------------------

3) 가운뎃점 대신 마침표가 쓰인 경우 기호의 의미를 고려하여 다르게 분석한다.

6.25 (사전: 육이오)	6.25	6.25/NNG
몸무게가 6.25kg 이다.	6.25kg 이다.	6.25/SN+kg/SW+이/VCP+다/EF+./SF
교회의 실천 덕목은 믿음.소망.사랑이다.	믿음.소망.사랑이다.	믿음/NNG+./SP+소망/NNG+./SP+사랑/NNG+이/VCP+다/EF+./SF

### 3.7.1.3. 따옴표, 괄호, 줄표(SS)

1) 따옴표(“)와 괄호(()), 줄표(-)는 SS 로 분석한다.

“철수야” 하고 친구가 불렀다.	“철수야”	/SS+철수/NNP+야/JKV+“/SS
이진호(35), 한국 마라톤 금메달	이진호(35),	이진호/NNP+(/SS+35/SN+)/SS+./SF
전투기 이름은 ‘KF-’ 뒤에 숫자를 더한다	‘KF-’	/SS+KF/SL+-/SS+’/SS

2) <우리말샘>에 하나의 어휘로 등재되어 있지만 괄호가 어절 중간에 쓰였을 경우 각각의 단위를 분리하여 분석한다. 단 고유명사 내부에 괄호가 쓰여 괄호를 분리 분석하면 고유명사 분석이 어려울 경우에는 괄호를 분리하지 않고 전체 구성을 고유명사로 분석한다.

5인(人)승	5/SN+인/NNB+(/SS+人/SH+)/SS+승/NNG
대(對)중국	대/XSN+(/SS+對/SH)/SS+중국/NNP
대중(中)국	대중(中)국/NNP
마이크로소프트(microsoft)사	마이크로소프트/NNP+(/SS+microsoft/NNP+)/SS+사/NNG

### 3.7.1.4. 줄임표(SE)

말줄임표 대신 마침표가 쓰인 경우, 마침표의 개수에 관계없이 모두 묶어 SE로 처리한다.

그리고....	그리고/MAJ+..../SE
---------	-----------------

### 3.7.1.5. 붙임표(물결)(SO)

기호를 뺀 말이 사전에 한 단어로 등재되어 있는 어휘라면 붙임표가 어절 중간에 개입했더라도 전체를 통합하여 표지를 부여한다. <우리말샘>에 등재될 수 없는 어휘라면 구성요소 각각을 분리하여 분석한다.

의~리	의~리/NNG
-----	---------

사이~소	사/VV+이~소/EF
------	-------------

\* 사이~소 (사전: 어미 ‘-이소’)

### 3.7.1.6. 기타 기호(SW)

- 1) 길이, 무게, 수효, 시간 따위의 수량을 수치로 나타내는 단위들 중 ‘미터, 그램, 리터’ 등은 의존명사(NNB)로, 한글이 아닌 ‘m, g, l’ 등은 기타 기호(SW)로 분석한다. ‘제곱미터, 퍼센트 포인트’ 등 사전에 한 단어로 올라 있는 단위 명사를 기호로 나타낸 것도 아래와 같이 하나의 기호로 분석한다.

5m	5/SN+m/SW
2%p	2/SN+%p/SW

- 2) 한글이 원이나 괄호 속에 들어가 한글을 따로 분리할 수 없는 경우에는 SW로 처리한다.

(주)	(주)/SW
(ㄱ)	(ㄱ)/SW
㉔	㉔/SW
(주)	(/SS+주/NNG+)/SS

\* (주)와 같이 괄호와 한글을 분리할 수 있는 경우에는 각각을 따로 분석한다.

### 3.7.2. 외국 문자(SL)

- 1) 한자(SH)를 제외한 외국 문자(로마자, 가나 등)가 한국어를 표현하기 위한 것이 아니라 외국어를 표현하거나 기호로 쓰인 경우에는 SL로 처리하고 로마숫자 I, II, III 등도 SL로 처리한다.

전자를 A, 후자를 B라고 하자.	A,	A/SL+./SP
I.서론	I.	I/SL+./SF
가게 open 시간은 5시입니다.	open	open/NNG
Busan just arrived	Busan just arrived	Busan/NNP+just/SL+arrived/SL

- 2) 외국 문자가 원이나 괄호 속에 들어가 기호를 따로 분리할 수 없는 경우에는 기호도 SL로 처리한다.

(a)	(a)/SL
Ⓐ	Ⓐ/SL
(a)	(/SS+a/SL+)/SS

\* 이와 같이 괄호와 외국 문자를 분리할 수 있는 경우에는 각각을 따로 분석한다.

- 3) 아래와 같이 어떤 표현의 구체적인 내용을 숨기려는 의도로, 또는 구어 전사 시 말이 정확히 들리지

않아 로마자 X 나 O 표시를 사용하는 경우가 있다. 그런 경우 아래와 같이 처리한다.

이런 버르장머리 없는 X	X	X/SL
XXXX 이론	XXXX	XXXX/SL

\* 이처럼 한 어절 전체가 X로 되어 있는 경우, 전체를 묶어 SL로 처리한다.

※ 만약 로마자가 아닌 x(곱셈표)나 △ 등이 쓰였으면 SW로 처리한다.

어쩌라는 거야 씨 X	씨 X	씨 X/NA
XX 스의 이론	XX 스의	XX 스/NA+의/JKG
XXX 의 이론	XXX 의	XXX/NA+의/JKG

### 3.7.3. 한자(SH)

1) 한글이 아닌 한자는 SH 로 처리한다. 한자가 원이나 괄호 속에 들어가 기호를 따로 분리할 수 없는 경우에는 기호도 SH 로 처리한다.

㉠	㉠/SH
(五)	(五)/SH
(五)	(/SS+五/SH+)/SS

\* (五)와 같이 괄호와 한자를 분리할 수 있는 경우에는 각각을 따로 분석한다.

2) 한국어를 표기하기 위해 한자가 쓰인 경우에는 의미를 고려하여 형태를 분석한다.

中國 정부에 항의했다.	中國/NNP
--------------	--------

### 3.7.4. 숫자(SN)

1) 아라비아 숫자(0, 1, 2 등) 및 아라비아 숫자가 원이나 괄호 속에 들어가 기호를 따로 분리할 수 없는 경우에는 기호를 SN 으로 처리한다.

① 조리법	①	①/SN
(1) 조리법	(1)	(1)/SN
(1) 조리법	(1)	(/SS+1/SN+)/SS

→ 이와 같이 괄호와 숫자를 분리할 수 있는 경우에는 각각을 따로 분석한다.

2) <우리말샘>에 한글로 등재되어 있는 어휘는 숫자로 바꾸어 표현되어 있어도 하나의 단어로 통합하여 분석하고, <우리말샘>에 한글로 등재되어 있지 않고 열린 집합을 형성할 경우에는 각각을 나누어 숫



자(SN)로 분석한다.

3분의 1	3분의 1	3/SN+분/XSN+익/JKG 1/SN
4대강	4대강	4/SN+대/XSN+강/NNG
5관왕(사전: 오관왕)	5관왕	5관왕/NNG
오관왕(사전: 오관왕)	오관왕	오관왕/NNG
11관왕	11관왕	11/SN+관/NNG+왕/NNG

## 3.8. 기타

### 3.8.1. 분석불능범주

#### 3.8.1.1. 분석불능범주(NA)

분석이 어렵고 품사를 특정할 수 없는 경우에 한하여 분석불능범주(NA)로 처리한다.

- 1) 분석하기 어려운 방언형의 경우 분석불능범주로 처리한다.

담배가 쏘매턴게 하마 자라서 빼나?	쏘매턴게	쏘매턴게/NA
---------------------	------	---------

- 2) 아래와 같이 외국어의 '한 문장'이 한글로 전사되어 나타난 경우, 각 어절을 내부 분석 없이 NA로 처리한다.

렛츠고	렛츠고	렛츠고/NA
익스큐즈 미	익스큐즈 미	익스큐즈/NA 미/NA
아이 러브 유	아이 러브 유	아이/NA 러브/NA 유/NA
굿!	굿!	굿/NA+!/SF
곤니치와	곤니치와	곤니치와/NA
니하오	니하오	니하오/NA
해피버스데이 투 유	해피버스데이 투 유	해피버스데이/NA 투/NA 유/NA

## 4. 구어 분석에 따른 주의사항

구어 자료의 형태 분석 방법은 기본적으로 문어 자료의 형태 분석 방법과 동일하다. 다만 구어에서 나타나는 준말과 형태 변이 현상을 되도록 분석에 반영하기 위해 아래의 지침을 별도로 마련하였다.

### 4.1. 구어에서 나타나는 준말과 형태 변이 현상의 처리

#### 1) 하나의 요소 내부에서 형태 변이가 일어난 경우

아래와 같이 하나의 형태 표지가 붙는 단위에 구어의 음성적 특성이 반영되어 형태 변이가 일어났을 때는, 원어절의 형태를 바꾸지 않고 표준형에 비추어 형태 표지를 부여한다.

건(<그건) 어렵지 않아요	건	거/NP+ㄴ/JX
것두(<그것도) 좋은데	것두	것/NP+두/JX
늦을까 봐 날라서 왔어.	날라서	날르/VV+아서/EC
이걸로	이걸로	이거/NP+ㄹ로/JKB
좋으까?	좋으까?	좋/VV+으까/EF+?/SF
여기 앉어	앉어	앉/VV+어/EF
그렇게 하더라도	하더라도	하/VV+더라도/EC
학교 간대더라	간대더라	가/VV+ㄴ 대더라/EF
할런지 모르겠다	할런지	하/VV+ㄹ런지/EF
갈 것 같애	같애	같/VV+애/EF
가기를 바래	바래	바라/VV+아/EF

\* ‘애’가 이처럼 표기상으로 분리되어 드러난 경우에만 ‘애’로 분석한다. ‘가기를 바래’, ‘나만 나무래’에서처럼 ‘애’가 표기상으로 분리되어 드러나지 않은 경우에는 모음조화에 따라 ‘아’로 분석한다.

#### 2) 본래 둘 이상으로 분석되어야 하는 요소인데 축약되어 형태 분리가 어려워진 경우

##### (1) 용언 어간과 어미의 결합형인 경우

사전에 한 단어로 올라 있는 말이 아니어서 용언 어간과 어미로 분석해야 하는 말이 다음과 같이 축약되어 전사된 경우가 있다. 이때는 아래와 같이 구어의 변이 형태를 그대로 인정하면서 어간과 어미를 분리하여 분석한다. 형태상 분리가 어려움에도 어간과 어미를 분리하도록 한 것은, 절을 꾸리는 데 있어서 용언의 역할이 중요한 만큼 용언 어간의 모습을 드러낼 필요가 있기 때문이다.

##### ① 형용사 ‘이렇-’, ‘그렇-’, ‘저렇-’, ‘어떻-’류의 변이 형태

일케	일/VV+게/EC
----	-----------

이케	일/VA+게/EC
이르케	이룅/VA+게/EC
요러케	요렁/VA+게/EC
요렇게	요렁/VA+게/EC
요케	욕/VA+게/EC
요로케	요룅/VA+게/EC
그르케	그룅/VA+게/EC
그런케	그렁/VA+게/EC

- \* 어미가 ‘게’로 잘못 전사되었지만 ‘ㅎ’과 ‘게’가 만나 ‘케’가 된 것으로 보아야 하므로 ‘게/EC’로 분석한다.

그러치	그러치/IC,
그러치	그렁/VA+지/EF

- \* 사전에 ‘그렇지’가 “틀림없이 그렇다는 뜻으로 하는 말”로서 감탄사로 올라 있다. 이에 따라, 일어난 사태에 대한 만족을 표시하며 혼잣말로 쓰이는 ‘그러치, 그룅치, 그치’는 더 분석하지 않고 감탄사로 보아야 한다. 이때 ‘그렇지’로 전사되어야 할 것이 ‘그러치’로 전사되었다. 하지만 위에서 본 ‘요렇게’의 경우와 달리 내부 형태 분석을 하는 상황이 아니므로 원문의 표기를 그대로 따른다.
- \* 그런데 구어 말뭉치에서 나타나는 ‘그렇지’는 대부분, 일어난 사태에 대한 만족을 표시하는 혼잣말로 쓰이는 것이 아니라 상대방에게 동의를 표시하는 말로 쓰인다. 이 경우에는 어간 ‘그렁-’과 어미 ‘-지’로 분리하여 분석해야 한다. 즉 ‘그렁/VA+지/EF’로 분석해야 함에 유의한다.

그치	그치/IC,
그치	글/VA+지/EF

- \* 이 역시, 만족을 표시하는 감탄사로 쓰인 경우에는 ‘그치/IC’로, 상대방에게 동의를 표시하는 말로 쓰인 경우에는 ‘글/VA+지/EF’로 분석한다.

그찮아	그찮아	글/VA+잖아/EF
그치 않습니까?	그치	글/VA+지/EC
그룅치 않습니까?	그룅치	그룅/VA+지/EC

- \* 이때는 ‘그치’, ‘그룅치’가 감탄사로 쓰인 것이 아니다. 따라서 위와 같이 용언 어간과 어미로 분석해야 한다.
- \* ‘그룅치’는 ‘그룅지’로 전사되어야 할 것이 잘못 전사된 것이다. 잘못 전사된 부분에서 형태 분석이 이루어지므로, 위에서 본 ‘요렇게’의 경우와 마찬가지로 ‘치/EC’가 아니라 ‘지/EC’로 분석한다.

어뜯케	어뜯/VA+게/EC
어뜨케	어뜯하/VV+아/EF

- \* ‘어뜯케’는 ‘어뜯게’로 전사되어야 할 것이 잘못 전사된 것이다. 잘못 전사된 부분에서 형태 분석이 이루어지므로, 위에서 본 ‘요렇게, 그룅치’의 경우와 마찬가지로 ‘케/EC’가 아니라 ‘게/EC’로 분석한다.

- \* ‘어뜨케’는 ‘어떡해’의 변이형이다. 사전에 ‘어떡하다’가 등재되어 있어 ‘어떡해’를 ‘어떡하/VV+아/EF’로 분석하는 것을 참고하여 ‘어뜨하/VV+아/EF’로 분석한다.
- \* 물론, ‘어뜨케 됐어?’에서처럼 부사어로 쓰인 것은 ‘어똥/VA+게/EC’가 될 것이다.

## ② 그 외 용언의 변이 형태

따르케	따름/VA+게/EC
다르케	다름/VA+게/EC
요만하케	요만함/VA+게/EC

- \* ‘다르다’, ‘요만하다’의 변이 형태가 나타났다. 역시 변이 형태를 그대로 인정하여 용언 어간과 어미를 분리한다. ‘따르+케’, ‘요만하+케’로 분석될 수도 있을 것이나 가능한 한 용언 어간 쪽에서 변이 형태를 인정하기로 한다.

## ③ 두 어절 이상에 해당하는 용언 어간+어미의 경우

어케요	얼/VA+게/EC+하/VV+아요/EF
이케서	일/VA+게/EC+하/VV+아서/EC

- \* ‘어케요’는 두 어절에 해당하는 ‘어떻게 해요’가 줄어들었고, 그 속에 용언 어간과 어미가 있다. 용언 어간 ‘하’가 생략되었는데, 이때 ‘하’를 복원하지 않으면 어절 속에 동사의 어간이 없는 셈이 되므로 ‘하’를 복원해야 한다. ‘하’ 앞의 ‘어케’는 ‘얼/VA+게/EC’로 분석한다.
- \* ‘이케서’ 역시 두 어절에 해당하는 ‘이렇게 해서’가 줄어들었다. 용언 어간 ‘하’를 복원하지 않으면 어절 속에 동사의 어간이 없는 셈이 되므로 복원해야 한다. ‘하’ 앞의 ‘이케’는 ‘일/VA+게/EC’로 분석한다.

왜케	왜케/MAG
왈케	왈케/MAG

- \* 두 어절에 해당하는 ‘왜 이렇게’가 줄어들었다. ‘왜케’, ‘왈케’가 절에서 서술어로 쓰이는 일은 없으므로, 이 경우에는 예외적으로 더 분석하지 않고 ‘왜케, 왈케’를 일반부사로 처리한다.

## ④ ‘이케~’와 같이 원문에 물결표 표시가 있는 것은 물결표를 제외하고 IC로 분석한다.

- ⑤ ‘X하-’ 형태에서 ‘하’가 아예 생략되거나 ‘ㅎ’만 남은 아래와 같은 경우에는 ‘하’의 형태를 완전하게 복원한다. ‘하’를 복원하지 않으면 어절 속에 용언의 어간이 없는 셈이 되므로 복원하지 않을 수 없다.

나중에 논의토록 하자.	논의토록	논의/NNG+하/XSV+도록/EC
생각지 못한 일	생각지	생각/NNG+하/XSV+지/EC

## (2) 용언 어간과 어미의 결합형이 아닌 경우

용언 어간과 어미의 결합형이 아니라면, 아래와 같이 형태 분리가 어려운 구어의 축약형을 더 분석하지 않고 하나의 단어로 인정하는 방안을 취하기로 한다. 형태 분리가 어려운 경우란, 형태 분리를 했을 때 적어도 한 요소가 사전 미등재어이고 그 요소가 다른 환경에서는 나타나지 않는 경우를 말한다. 형

태 표지는 해당 단어의 문장 성분을 고려하여 부여한다.(예: 부사어→ 부사)

내비뒤	내비뒤/VV+어/EF
넙뒤	넙뒤/VV+어/EF

- \* '내비뒤'의 '뒤-'는 분석 가능하지만 '내비'가 사전 미등재어이다. 그리고 '내비'는 '뒤-' 앞 외의 다른 환경에서는 거의 나타나지 않는다. 이에 따라 '내비뒤-' 전체를 동사로 처리한다.

여따(<여기에다가) 놔.	여따	여따/MAG
---------------	----	--------

- \* '여'는 등재되어 있지만 '따'가 미등재어이다. 이 '따'는 '여따, 거따, 저따' 외에서는 보기 어렵다. 이에 따라 '여따' 전체를 한 단어로 처리한다. 문장 속에서 부사어로 쓰이므로 일반부사로 처리한다.

언놈이(<어느 놈이) 그래?	언놈이	언놈/NP+이/JKS
-----------------	-----	-------------

- \* '놈'은 분석 가능하지만 '언'이 사전 미등재어이다. 그리고 '언'이 '놈' 앞 외의 다른 환경에서는 나타나지 않는다.

얼다 대고	얼다	얼다/MAG
-------	----	--------

- \* '-다'는 분석 가능하지만 '얼'이 사전 미등재어이다. 그리고 '얼'이 '-다' 앞 외의 다른 환경에서는 나타나지 않는다.

클났다.	클났다.	클나/VV+았/EP+다/EF+./SF
------	------	----------------------

- \* '클'이 사전 미등재어이고 '나다' 외의 다른 환경에서 나타나지 않는다.

어서(<어디서) 그래?	어서	어/NA+서/ JKB
--------------	----	-------------

- \* '어'가 사전 미등재어이고 '서' 외의 다른 환경에서 나타나지 않는다.

짱난다	짱난다	짱나/VV+ㄴ 다/EF
-----	-----	--------------

- \* '짱'이 사전 미등재어이고 '나다' 외의 환경에서 나타나지 않는다.

#### <주의사항>

- ① '이리로, 그리로, 저리로, 요리로, 고리로, 조리로'뿐 아니라 '일로, 글로, 절로, 올로, 골로, 졸로'가 <우리말샘>에 부사로 등재되어 있으므로 MAG 로 분석해야 함에 유의한다.

- ② 아래와 같이 같은 모음이 겹치면서 축약된 경우에는 본래 형태를 복원한다.

어땀어요.	어땀어요.	어디/NP+있/VA+어요/EF+./SF
-------	-------	-----------------------

③ 비표준적인 준말 활용형

아래와 같이 비표준적인 준말 활용형이 나타난 경우, 용언 어간은 표준형으로 복원하되 어미에서 매개모음 ‘으’를 빼고 분석한다.

여기다 논(<놓은) 거야.	논	놓/VV+ㄴ/ETM
여기다 노셨던(<놓으셨던) 거야.	노셨던	놓/VV+시/EP+었/EP+던/ETM
찌시더니(<찜으시더니)	찌시더니	찜/VV+시/EP+더니/EC
아이를 낳면은(<낳으면은)	나면은	낳/VV+면/EC+은/JX

④ 삼중모음

모음 ‘ㄱ’과 ‘ㅅ’가 축약되어 삼중모음 발음이 나타난 경우, ‘사귀’어요’와 같이 ‘로’ 표시되어 있다. 형태 분석 시에 이 ‘는’ 반영하지 않는다.

바뀌었었어요.	바뀌었었어요.	바뀌/VV+었었/EP+어요/EF+./SF
---------	---------	------------------------

⑤ 관형격조사 ‘에’

관형격조사 ‘의’의 발음을 ‘에’로 전사한 경우가 있다. 이 경우 ‘에/JKG’로 형태 표지를 붙인다. ‘에’가 부사격조사인지 관형격조사인지 판단이 어려운 경우에는 부사격조사(JKB)로 형태 표지를 붙인다.

나에 생각	나에	나/NP+에/JKG
-------	----	------------

⑦ 지정사 ‘이다’

구어에서 이중모음이 단모음으로 발음되는 현상이 자주 일어나 그 결과 ‘예’가 ‘에’로 발음되고, 아래와 같이 지정사 ‘이다’가 생략된 것으로 보이는 현상이 있다. 이때는 문법적으로 지정사가 있으나 단지 이중모음이 단모음으로 발음된 것으로 보아 지정사를 복원한다.

이렇게 얘기할 거예요.	거예요.	거/NNB+이/VCP+예요/EF+./SF
--------------	------	------------------------

⑧ 구어에서 나타나는 사전 미등재 요소

사전에 등재되지 않은 문법 요소 ‘-르랑’, ‘-르동’은 사전에 등재된 연결어미 ‘-르락’을 참고 하여 연결어미로 분석한다.

이해가 갈랑말랑 하길래	갈랑말랑	가/VV+르랑/EC+말/VV+르랑/EC
--------------	------	-----------------------

⑨ 의성의태어를 구성하는 요소가 여러 번 반복되어 나오는 경우의 처리는 아래와 같다.

지글지글지글지글	지글지글/MAG+지글지글/MAG
지글지글지글	지글지글지글/MAG

\* 사전에는 ‘지글지글’이 한 단어로 올라 있고, ‘지글’은 어근에 해당한다. 위의 경우 단어 ‘지글지글’이 두 번 연

달아 나온 것으로 분석할 수 있으므로 두 단어로 나누어 처리한다.

- \* '지글지글'을 한 단어로 처리하면 어근에 해당하는 '지글'이 남는다. 이런 경우에는 어근을 앞말에 붙여서 '지글지글지글' 전체를 하나의 단어로 처리한다.

## 4.2. 구어 전사 시 이용된 마크업과 표지의 처리

### 가) 물결표(~)

구어 전사 시에 이용되어 머뭇거림을 나타내는 담화 표지에 ~(물결표)가 붙어 있다. 이 경우 ~는 분석에서 제외하고 ~ 앞에 있는 말에 IC 를 부여한다.

아~	아	아/IC
그~	그	그/IC
뭐~	뭐	뭐/IC

단, 아래와 같이 머뭇거림을 나타내는 담화 표지가 아닌 것에 ~(물결표)가 붙어 있는 경우가 있다. 그런 경우는 물결표를 넣지 않아야 할 곳에 넣은 전사 오류에 해당하므로, 물결표를 분석에서 제외하고, 남은 요소에 형태 표지를 부여한다.

국호를~을~	국호를을	국호/NNG+를/JKO+을/JKO
--------	------	--------------------

### 나) 마크업 기호

<trunc>, </trunc> 등의 마크업 기호는 한 어절로 두고 형태 표지를 부여하지 않는다. 단, 아래의 주의 사항에 유의한다.

#### <주의 사항>

- ① <note> </note> 마크업의 경우에는 마크업 기호와 그 안의 내용을 모두 한 줄에 보여주 고, 어떠한 표지도 부여하지 않는다.

[예] <note>배경 화면 잠깐 나옴</note> → 한 어절로 두고 분석하지 않음.

- ② 사람 이름, 주소 등 개인 정보 보호를 위한 마크업은 다음과 같은 방식으로 형태 표지를 부여한다.

<anon type="name" n="1"/>가	name1/NNP+가/JKS
<anon type="name"/>가	name/NNP+가/JKS
<anon type="address" n="2"/>은	address2/NNP+은/JX

구어 전사 지침상, 일반 대화에서 대화자들 및 관련인의 개인 정보가 드러난 경우에는 개인 정보 보호를 위하여 해당 정보를 위와 같이 마크업으로 가리도록 하였다. 그런데 전사 실수로 이름 등의 개인 정보가 마크업 없이 노출된 경우가 있다. 이때에는 해당 개인 정보를 '000'과 같이 고치고 SL 표지를

부여하고, 향후 보완 방법을 모색할 수 있도록 한다.

지현이가 그러는데	지현이가	000/SL+가/JKS
-----------	------	--------------

③ 마크업으로 인해 사람 이름에서 분리된 접미사 ‘-이’에는 NA 를 부여한다.

<anon type="name" n="5"/>이 말 고 하나가 더 있니?	이	이/NA
---	---	------

\* ‘말고’ 앞에는 주격조사가 올 수 없다. 이때 ‘말고’ 앞에 나온 ‘이’는 ‘영속이’에서 볼 수 있는 접사 ‘이’로 판단 가능하다. 본 지침의 비분석 접사가 분리되어 나온 경우이므로 이러한 ‘이’는 ‘이/NA’로 처리한다.

다) <trunc> </trunc> 사이의 요소

<trunc> </trunc> 사이에 표시되어 있는 끊어진 어절(단어가 불완전하게 발화된 경우)에는 NA(분석 불능범주) 표지를 부여한다.

미국과 <trunc>같</trunc> 같은	같 같은	같/NA 같/VA+은/EP
-------------------------	------	----------------

단, <trunc> </trunc> 사이에 있는 요소를 제외할 경우 앞뒤의 말이 이어지지 않는 경우라면, <trunc> </trunc> 사이에 있는 요소이더라도 형태 표지를 부여한다.

또 <trunc>문의하</trunc> 하기도 했습 니다.	문의하 하기도	문의/NNG+하/XSV 하/XSV+기 /ETN+도/JX
------------------------------------	---------	-----------------------------------

라) <unclear> </unclear> 사이의 요소

전사 시 잘 들리지 않은 부분은 <unclear> </unclear>로 표시되어 있다. 가능한 한 각 요소에 맞는 형태 표지를 부여하고, 형태 표지를 부여할 수 없는 경우에는 NA(분석불능범주)를 부여한다.

(1) 정확히 들리지 않았으나 x 표시 없이 전사된 경우 가능한 한 각 요소에 맞는 형태 표지를 부여한다.

<unclear>더 힘들어</unclear> 더	더 힘들어	더/MAG 힘들/VA+어/EF
있<unclear>어요</unclear>→ 있	있	있/VA
있어<unclear>요</unclear>→ 있어	있어요	있/VA+어/EF 요/JX

(2) 일부 음절이 들리지 않은 경우

일부 음절이 들리지 않은 경우에는 해당 음절이 x로 표시되어 있다. 이때는 x가 포함된 단어 부분에 NA(분석불능범주)를 부여한다.

<unclear>xx스의</unclear> 이론을	xx스의 이론을	xx스/NA+의/JKG 이론/NNG+을 /JKO
-----------------------------	----------	-------------------------------



- (3) <unclear> </unclear> 마크업으로 인해 표지를 주기 어려운 음절이 발생하는 경우 아래와 같이 <unclear> 마크업으로 인해 단어가 분리되어 표지를 주기 어려운 음절이 발생하는 경우에는, 각 음절에 NA(분석 불능 범주)를 부여한다.

임시정 <unclear>부</unclear>→ 임시정	임시정 부	임시/NNG+정/NA 부/NA
-------------------------------	-------	------------------

### 4.3. 전사 오류 및 해석 불능 어절의 처리

#### 가) 탈자로 인해 형태 표지 부여가 어려운 경우

아래와 같이 전사 과정에서 탈자가 발생하였거나 혹은 발화 실수로 과도한 생략이 일어나 형태 표지 부여가 어려워지는 경우가 있다. 누락된 내용이 분명하게 판단 가능할 경우 해당 내용을 복원하여 분석하고, 누락된 내용의 판단이 모호한 경우에는 해당 요소에 NA(분석 불능 범주)를 부여한다.

하지 못는(<못하는)	못하는	못하/VX+는/ETM
하지 못는(< ?)	못는	못/VA+는/ETM

#### 나) 잉여적인 요소가 덧붙은 경우

아래와 같이 전사 과정에서 첨자가 발생하였거나 혹은 발화 실수로 잉여적인 형태가 덧붙은 경우가 있다. 잉여적인 요소가 분명하면 내용에 맞게 원문을 수정하여 분석하고, 만약 원문 수정이 어렵다면 NA(분석 불능 범주)를 부여한다.

국호를을	국호를	국호/NNG+를/JKO
됐습니다다.	됐습니다.	되/VV+었/EP+습니다/EF+./SF

#### 다) 띄어쓰기 오류로 인해 형태 표지 부여가 어려운 경우

‘그럴걸’이 ‘그럴 걸’로, ‘뒤치락거리다’가 ‘뒤치락 거리다’로 띄어쓰기와 함께 전사된 경우가 있다. 규범에 맞게 띄어쓰기를 수정하여 해당 어절의 형태를 분석하고 원문 수정과 형태 표지 부여가 어려운 요소에는 NA(분석불능범주)를 부여한다.

뒤치락 거리고	뒤치락거리고	뒤치락거리/VV+고/EC
그럴 걸	그럴걸	그러/VV+ㄹ걸/EF
집에 갔는 지 모르겠다	갔는지	가/VV+았/EP+는지/EC

#### 라) 어절의 의미 파악이 어려운 경우

어절의 의미를 파악하기 어려운 경우, 의미는 불분명하더라도 아래와 같이 해당 어절을 이루는 요소의 문법적 지위를 확정할 수 있다면 그에 따라 최대한 형태 표지를 부여한다. 문법적 지위를 확정하기 어렵거나 형태 표지를 부여하기 어려운 경우에 한해 NA를 부여한다.

조선 용 어~ 왕조 실의 역대 왕들의 왕릉	실의	실/NA+의/JKG
-------------------------	----	------------

- \* 이때 '실의'의 의미를 정확히 파악하기가 어려우나, 맥락상 '조선 왕실의'에서 '실의'가 분리된 것으로 보인다. 이에 따라 실/NA+의/JKG로 형태 표지를 부여한다.

아직까지 충청에 민심이 복마진입니다.	복마진입니다.	복마진/NNG+이/VCP+ㅂ니다/EF+./SF
----------------------	---------	---------------------------

- \* 이때 '복마진'의 의미를 정확히 파악하기가 어려우나 지정사 앞에 나타나는 문법적 특성으로 보아 일반명사로 판단할 수 있다. 이에 따라 복마진/NNG로 표지를 부여한다.

위의 두 방식을 적용하여도 문법적 지위를 확정하기 어려운 요소가 있다면 해당 요소에 NA를 부여한다.

마) 한 어절 내에 마크업이 포함된 경우

다음과 같이 한 어절 내에 마크업이 포함된 경우에는, 마크업을 제외하고 남은 부분만을 지침에 따라 분석한다.

<trunc>아쉬</trunc>아쉬움이	아쉬 아쉬움이	아쉬/NA+아쉬움/NNG+이/JKS
-----------------------	---------	---------------------

## [붙임2] 어휘의미 분석 말뭉치 지침

### 어휘의미 분석 말뭉치 구축 지침

#### 1. 총론

##### 1.1. 지침의 목적

이 지침은 한국어 문장에 사용된 어휘의 의미를 분석하여 어휘의미 분석 말뭉치를 구축하는 방안을 제시하는 것을 목적으로 한다. 형태 분석된 한국어 문장의 체언류와 용언류 어절을 대상으로 각 형태에 어휘의미 표지를 부착하는 것이 세부 목적이다. 어휘의미 분석은 형태 분석 말뭉치를 대상으로 수행한다.

- 본 지침의 어절 분석 표지는 '21세기 세종계획'의 '형태의미 분석 말뭉치 구축 지침'과 '2020 어휘 분석 말뭉치 구축 지침'을 토대로 작성된 것이다.
- 원문을 정제하고 가공하여 사용하게 됨에 따라 말뭉치 오류 및 예외 처리에 관련된 지침을 대폭 축소하였다.
- 본 지침은 단계적인 분석을 할 수 있도록 고려하였다.

##### 1.2. 용어 정의

- 어휘의미(sense): 갈래뜻

##### 1.3. 기본 원칙

- 어휘의미 분석은 형태 분석 말뭉치의 체언과 용언을 분석 대상으로 한다. 구체적인 대상 범주는 다음과 같다.
  - 일반명사(NNG), 의존명사(NNB), 고유명사(NNP), 대명사(NP), 수사(NR), 어근(XR), 동사(VV), 형용사(VA), 보조용언(VX), 긍정지정사(VCP), 부정지정사(VCN)
- 형태 분석 말뭉치에서 분석 오류가 발견될 경우 이를 올바르게 수정하여 어휘의미 분석을 실시한 후 이를 별도로 기록한다.

## 2. 어휘의미 분석의 원리

### 2.1. 어휘의미 분석 대상

- 어휘의미 분석은 어휘의미(갈래뜻, sense)를 구별하여 다의어 수준의 어휘의미를 분석하는 것이다.
- 어휘의미 분석은 형태 분석 단계에서 체언 및 용언으로 분석된 것을 대상으로 한다.
- 원어절에서는 실현되지 않은 형태이더라도 형태 분석 단계에서 복원된 형태이면 이를 어휘의미 분석의 대상으로 삼는다.

예문	형태 및 어휘의미 분석 어절
이어 “끝까지 머리 붙잡고 늘어졌어야 하는데 돈이나 뜯자고 생각하는 거냐. 세상 정말 무섭다. 제3자가 보기엔 잘못된 부분도 분명히 있을 거라는 생각이 든다”면서도 “단언컨대 사기꾼 가족이 아니다”며 억울해 했다.	거__001/NNB+이__888/VC+냐/EF+./SF

- <2025 한국어 구조 분석 말뭉치>의 형태 분석 단계에서는 단어 단위 분석을 수행하여 접두사나 접미사가 통합된 단어 단위로 형태 분석 표지를 부착하였다. 어휘의미 분석도 이 단위에 따라 표지를 부착한다.

유형	형태 분석 어절	형태 및 어휘의미 분석 어절
접두사 통합형	반독재/NNG	반독재__001/NNG
명사 파생 접미사 통합형	문화적/NNG	문화적__001/NNG
동사 파생 접미사 통합형	공부하/VV	공부하__001/VV
형용사 파생 접미사 통합형	건강하/VA	건강하__002/VA

### 2.2. 어휘의미 분석 표지 부착

- 어휘의미 분석은 국립국어원에서 구축한 <우리말샘>의 다의어 번호를 이용하여 어휘의미를 구분한다.

<p>국어(國語) [구거 ㉠]</p> <p>* 국어 「001」 「명사」 한 나라의 국민이 쓰는 말.</p> <p>* 국어 「002」 「명사」 우리나라의 언어. '한국어'를 우리나라 사람이 이르는 말이다.</p>
---

- 어휘의미 분석 결과에 따라 ‘\_\_’과 세 자리의 어휘의미 번호 ‘○○○’를 부착한다.

예문	원어절	형태 및 어휘의미 분석 어절
아버지는 학교에서 국어를 가르치신다.	국어	국어__002/NNG

- <우리말샘>은 등재 어휘가 꾸준히 증가하고 있는 동적 사전이므로 어휘의미 분석의 기준 시점을 제시할 필요가 있다. 따라서 2025년 1분기의 <우리말샘> 등재 어휘를 기준으로 어휘의미를 분석한다.

## 2.2.1. 등재어: 등재 번호

- <우리말샘>에 해당 어휘와 해당 의미가 등재되어 있는 경우 표제어 등재 번호를 부착한다.

돈이 꽃이 되면 기적 같은 일을 일으킨다.	기적	기적__003/NNG
영화는 이순신 장군의 한산도대첩을 다루고 있다.	이순신	이순신__002/NNP
서울 종로구청 앞에 첫 번째 소녀상을 세웠다.	번째	번째__001/NNB
나는 지금 하는 일을 좋아서 하고 있는가.	나는	나__003/NP+는/JX
동생과 나는 서울에 둘만 살고 있다.	둘만	둘__00/NNB+만/JX
원나라의 몰락과 홍건적의 침입을 겪으며	겪으며	겪__001/VV+으며/EC
어떠한 정보도 제공하지 않고 있다.	어떠한	어떠하__001/VA+L /ETM
공간을 이동하는 ‘순간이동’ 마술이다.	마술이다	마술__002/NNG+이__004/VCP+다/EF+./SF
미신은 종교가 아니다.	아니다	아니__001/VCN+다/EF+./SF
책을 다 읽어 간다.	간다	가__034/VX+L 다/EF+./SF

- <우리말샘>에 제시된 갈래 뜻 사이에 구분이 불분명할 경우에는 <우리말샘>의 예문을 최대한 검토하여 확정한다. 이때 한쪽이 예문이 없는 것이라면 예문이 있는 쪽의 갈래 뜻을 선택한다.

예문	형태 및 어휘의미 분석 어절	설명
후보가 되면 심층 인성 평가를 거친다.	인성__002/NNG	<우리말샘>에 등재된 ‘인성’의 다양한 의미 중 “001 사람의 성품”과 “002 각 개인이 가 지는 사고와 태도 및 행동 특성”은 뜻 사이에 구분이 불분명하다. 이때 전자의 예문 ‘인성 교육, 인성이 착하다, 올바른 인성을 기르다’와 후자의 예문 ‘인성 개발, 인성 검사’을 검토하여 위 문장에 제시된 ‘인성’에 어휘의미 번호 ‘002’를 부여한다.
제주도가 세계환경수도 조성에 찬걸음을 내딛었다.	내딛__002/VV+었/EP+다/EF+./SF	<우리말샘>에 등재된 ‘내딛다’는 “001 【…을】 ‘내디디다’의 준말”과 “002 【…을】 ‘내디디다’의 준말”로 동일하게 풀이되어 있다. 이때 전자의 예문 ‘길이 험하여 발을 내딛기가 힘들다.’를 통하여 001은 장소의 이동과 관련된 것으로 파악할 수 있고, 후자의 예문 ‘새로운 삶에 첫발을 내딛는 결혼.’을 통하여 002는 새로운 범위 안에 처음 들어섬으로 분석할 수 있다. 이러한 예문 검토를 통하여 제시된 ‘내딛다’에 어휘의미 번호 ‘002’를 부여한다.
중국과 얼어붙은 관계를 녹이려는 분명한 신호를 보냈다.	녹이__005/VV+려는/ETM	<우리말샘>에 등재된 ‘녹이다’의 의미 중 “005 【…을】 감정을 누그러지게 하다. ‘녹다’의 사동사”와 “006 【…을】 감정을 누그러지게 하다. ‘녹다’의

		사동사”은 동일하게 풀이되어 있다. 이때 전자의 예문 ‘미움을 녹일 수 있는 것은 서로의 처지를 이해하고 아끼며 보살 펴 주는 마음밖에는 없다.’를 통하여 ‘녹이다005’는 “감정이 누그러지다”에 해당하는 ‘녹다 004’의 사동사로 파악할 수 있고, 후자의 예문 ‘요염한 여자가 남자를 녹이다.’를 통하여 ‘녹이다006’은 “어떤 대상에 몹시 반하거나 홀리다”에 해당하는 ‘녹다011’의 사동사로 분석할 수 있다. 이러한 예문 검토를 통하여 위 문장에 제시된 ‘녹이다’에 어휘의미 번호 ‘005’를 부여한다.
매서운 강풍이 문풍지를 뒤흔들었다.	강풍__006/NNG+이/JKS	<우리말샘>에 등재된 ‘강풍’의 다양한 의미 중 ‘003’과 ‘006’은 “세게 부는 바람”으로 의미가 동일하다. 그러나 ‘003’에는 예문이 없고 ‘006’에만 예문이 있기 때문에 위 문장에 제시된 ‘강풍’에 어휘의미 번호 ‘006’을 부여한다.

- 단, 양쪽 모두 예문이 없다면 포괄적인 갈래 뜻을 선택한다.

창밖으로 시원한 단풍나무 숲이 들어온다.	단풍나무__001/NNG	<우리말샘>에 등재된 ‘단풍나무’의 다양한 의미 중 ‘001’은 “단풍나뭇과의 나무를 통틀어 이르는 말”로, 학명이 함께 제시된 ‘002’의 의미를 포함하고 있다. 이때 예문을 검토하여 의미를 확정하는 것이 원칙이나, 두 갈래 뜻 모두 예문을 제시하지 않아 한쪽을 선택하기 어렵다. 이 경우 포괄적인 의미를 나타내는 ‘001’을 위 문장에 제시된 ‘단풍나무’에 어휘의미 번호로 부여한다.
------------------------	---------------	--

## 2.2.2. 어휘 미등재: 777

- <우리말샘>에 동일한 한글 배열로 이루어진 형태가 등재되지 않은 어휘의 경우, 어휘의미 번호는 ‘777’로 한다.

전공에 관련된 텍스트북을 구비하였다.	텍스트북__777/NNG+을/JKO	
요즘처럼 날씨가 더운 때에는 공강스가 최고야.	공강스__777/NNG+가/JKS	공강스 : 공항에서 보내는 바캉스의 의미
군데군데 시멘트가 덧발려 있었고,	덧발리__777/VV+어/EC	
팽창기가 달콤고소한 튀밥만 안겨준 것은 아니었다.	달콤고소하__777/VA+L/ETM	

- 말뭉치에 나타난 준말에 대응하는 본말이 <우리말샘>에 등재되어 있더라도 준말의 형태는 미등재 어이기 때문에 어휘의미 번호 ‘777’을 부여한다.

오송역에서 교원대로 향했	교원대__777/NNP+로/JKB	‘교원대’의 본말인 ‘한국교원대’가 <우리말샘>에
---------------	--------------------	-----------------------------

다.		등재되어 있더라도 준말의 형태는 미등재어에 해당하기 때문에 어휘의미 번호 '777'을 부여한다.
금융투자협회 회장 선거가 여의도 금투협회에서 열렸다.	금투협회__777/NNP+에서/JKB	금투협회'의 본말인 '한국금융투자협회'가 <우리말샘>에 등재되어 있더라도 준말의 형태는 미등재어에 해당하기 때문에 어휘의미 번호 '777'을 부여한다.

### 2.2.3. 의미 미등재: 888

- <우리말샘>에 동일한 한글 배열로 이루어진 형태가 등재되어 있으나 해당 어휘의 의미가 제시되어 있지 않은 경우, 어휘의미 번호는 '888'로 한다.

골프채의 헤드와 샤프트 모두 나무를 깎아 만들기 시작했다.	샤프트__004/NNG	<우리말샘>에 제시된 '샤프트'는 체육 분야에 사용되는 전문 용어로 "배드민턴 라켓의 긴 막대 부분"을 의미한다. 위 문장에 쓰인 '샤프트'는 "골프채의 긴 막대 부분"을 가리키기 때문에 <우리말샘>의 등재 어휘와 의미적 관련성을 찾을 수 있다. 비록 구체적인 운동 종목은 다르나 체육이라는 동일 분야에서 사용되는 용어이기 때문에 <우리말샘>에 등재된 어휘의미 번호 '004'를 부여한다.
코리안 드림을 꿈꾸다.	드림__888/NNG+을/JKO	<우리말샘>에 "매달아서 길게 늘이는 물건" 등의 의미로 '드림'이 등재되어 있으나 "꿈"에 해당하는 의미가 제시되어 있지 않으므로 '드림(dream)'에 어휘의미 번호 '888'을 부여한다.
그는 신작에 다양한 제주 설화를 끌어들이었다.	끌어들이다__888/VV+있/EP+다/EF+./SF	<우리말샘>에 따르면 '끌어들이다'는 "남을 권하거나 꺾어서 자기편이 되게 하다"를 의미하나, 위 문장에 사용된 '끌어들이다'는 "인용하다"에 가까운 의미로 판단된다.
우완 투수들이 공을 던질 때 왼쪽 어깨가 달혀 있어야 한다는 게 원칙이지만,	달히__888/VV+어/EC	<우리말샘>에 '달히다'는 "001 열린 문짝, 뚜껑, 서랍 따위가 도로 제자리로 가 막히다. '달다'의 피동사", "002 하루의 영업이 끝나다. '달다'의 피동사", "003 굳게 다물어지다. '달다'의 피동사"의 의미로 등재되어 있다. 그러나 위 문장은 투수가 공을 던질 때 타자에게 가슴이 보이지 않게끔 어깨가 벌어지지 않아야 한다는 것을 의미하는데, 이러한 의미의 '달히다'는 <우리말샘>에 등재되어 있지 않다. 따라서 위 문장에 사용된 '달히다'에 어휘의미 번호 '888'을 부여한다.

- 특정 분야의 전문 용어로 <우리말샘>에 등재된 어휘가 다른 분야에서 쓰이거나 일상 용어로 사용된 경우 어휘의미 번호 '888'을 부여한다.

대부분이 여행사의 마라톤 패키지 상품으로 홍콩	패키지__888/NNG	<우리말샘>에 제시된 '패키지'는 영상 분야에 사용되는 전문 용어로 "시나리오 작가와 인기 배우, 인
---------------------------	--------------	--

을 찾았다.		기 소설과 인기 배우 등의 결합”을 의미한다. 위 문장에 쓰인 ‘패키지’ 역시 “일괄 상품 혹은 상품의 결합” 정도의 의미로 <우리말샘>의 등재 어휘와 의미적 관련성을 찾을 수 있으나 서비스업이라는 다른 분야에서 사용되는 용어이기 때문에 해당 어휘의 의미가 등재되지 않은 것으로 판단하여 어휘 의미 번호 ‘888’을 부여한다.
--------	--	--

- <우리말샘>에 등재된 일반명사가 고유명사로 쓰이거나, 고유명사가 일반명사로 사용된 경우 고유명사의 대상성을 고려하여 어휘의미 번호 ‘888’을 부여한다.

한겨레를 펼쳐 보다.	한겨레__888/NNP+를/JKO	매체명 ‘한겨레’의 경우 <우리말샘>에 등재된 ‘한겨레’와 의미적 유사성을 가질 수도 있으나 그러한지를 일일이 판단하고 확인하는 것이 불가능하기 때문에 고유명사의 대상성을 고려하여 어휘의미 번호 ‘888’을 부여한다.
영화 ‘변호인’을 보기 위해 극장으로 몰려갔다.	’ / S S + 변 호 인 __888/NNP+’/SS+을/JKO	일반명사에서 파생된 고유명사의 경우 <우리말샘>에 등재되어 있더라도 대상성을 고려하여 어휘의미 번호 ‘888’을 부여한다.
파격적 구도의 고려 수월관음도가 일본에서 발견되었다.	수월관음도__888/NGG+가 /JKS	<우리말샘>에 등재된 ‘수월관음도’는 용인대학교에 소장된 보물 제1286호를 가리킨다. 그러나 위 문장처럼 문화재명 ‘수월관음도’가 맥락상 일반명사로 쓰일 경우 어휘의미 번호 ‘888’을 부여한다.

## 2.2.4. 분석 불가: 999

- <우리말샘> 미등재어 중 ‘777’, ‘888’로 어휘의미를 부여할 수 없는 대상에는 어휘의미 번호 ‘999’를 부여한다.

# 3. 어휘의미 분석의 실제

## 3.1. 체언

### 3.1.1. 명사(NN)

#### 3.1.1.1. 일반명사(NNG)

- 수를 셀 수 있는 단위로 쓰일 수 있는 명사들은 기본 뜻풀이와 셀 때의 뜻풀이가 상이하므로, 그에 맞는 어휘의미 번호를 부여한다.



예문	형태 및 어휘의미 분석 어절
골목은 폭이 좁아 마주 오는 사람과 스칠 듯했다.	사람__001/NNG+과/JKB
동갑내기인 두 사람은 중학교 시절부터 친구 사이였다.	사람__010/NNG+은/JX

- 정부와 국가기관 관련 어휘는 뜻풀이에 ‘우리나라’로 한정되어 있더라도 다른 나라에 유사한 기관이 존재하는 것이 인정되므로 동일한 어휘의미 번호를 부여한다.

연방 결혼법이 미국 대법원에서 위헌 판결이 났다.	대법원__001/NNG+에서/JKB	<우리말샘>에 등재된 ‘대법원’은 “우리나라의 최고 법원”으로 정의되어 있으나, 미국 등 다른 나라의 최고 법원 역시 대법원으로 번역되므로 <우리말샘> 등재어 ‘대법원’의 포함되는 의미로 다룰 수 있다.
-----------------------------	---------------------	---

### 3.1.1.2. 고유명사(NNP)

- 인명의 경우 <우리말샘>을 참조하여 최대한 어휘의미를 부여하는 것을 원칙으로 하되, 사전에 등재되어 있지 않은 것(외국인의 성, 이름 포함)은 어휘의미 번호 ‘777’을 부여하고, 한자 또는 영문 표기의 일치 여부와 관계없이 해당 표기가 등재되어 있는 것은 어휘의미 번호 ‘888’을 부여한다.

원어절	형태 및 어휘의미 분석 어절	비고
이순신	이순신__002/NNP	충무공
김옥균	김옥균__001/NNP	갑신정변
아인슈타인	아인슈타인__001/NNP	상대성이론
비틀즈	비틀즈__001/NNP	록 그룹
김종서	김종서__001/NNP	조선 전기의 충신
김종서	김종서__888/NNP	록 가수
김대중	김대중__777/NNP	역대 대통령
노무현	노무현__777/NNP	역대 대통령
이명박	이명박__777/NNP	역대 대통령
박근혜	박근혜__777/NNP	역대 대통령

- ‘성’(性)의 경우에는 ‘우리나라 성씨 순위, 성씨별 인구밀도 순위’ 등의 자료를 참고하여 사전에 등재되어 있는 같은 ‘성’(性) 중 대표적인 성의 의미 번호로 통일한다.

정 시인은 주로 자연의 정취를 주제로 서정시를 쓴다.	정__054/NNP	<우리말샘>에는 나주, 창원 등 10여 본이 현존하는 ‘정(丁)’과 경주, 동래 등 120여 본이 현존하는 ‘정(鄭)’이 등재되어 있는데, 성씨별 인구수 등을 고려하여 ‘정(鄭)’의 의미 번호 ‘054’를 부여한다.
-------------------------------	------------	--

- 둘 이상 등재된 성(性)의 대표 의미 번호

강__004	도__015	빈__003	수__008	웅__006	장__028	채__026
경__036	돈__012	사__011	순__007	우__015	전__007	천__002
공__012	마__006	서__013	승__009	운__005	정__054	초__018
구__013	모__029	석__007	시__009	원__003	조__023	추__001
국__008	반__019	선__012	신__003	위__033	종__020	편__003
기__010	방__003	설__013	심__015	유__022	주__012	하__007
노__013	범__006	성__006	양__017	윤__001	지__010	호__016
뇌__005	변__003	소__025	여__018	이__023	진__026	
단__017	봉__010	송__001	연__003	임__010	창__007	

\* 국가통계포털 “성씨·본관별 인구(5인 이상)-전국”을 참고함.(인구조사표에는 ‘수’(隋)의 인구가 많으나 <우리말샘>에 미등재되어 있어 그 다음 인구수를 보이는 ‘수’(水)의 의미 번호를 부여함.)

- 지명의 경우 <우리말샘> 등재어의 한자 또는 영문 표기의 일치 여부를 살펴, 해당 표기가 사전에 등재되어 있지 않은 것은 ‘777’을, 등재되어 있는 것은 ‘888’을 부여한다.

서울	서울__002/NNP	
캘리포니아	캘리포니아__777/NNP	
런던	런던__003/NNP	영국 잉글랜드
런던	런던__888/NNP	캐나다 온타리오주

### 3.1.1.3. 의존명사(NNB)

- 의존명사는 의미에 따라, 혹은 선행하고 후행하는 요소가 무엇인지에 따라 구분되므로 <우리말샘>에 등재된 내용을 살펴 각각에 맞는 어휘의미 번호를 부여한다.

나아갈 바를 밝히다.	바__004/NNB+를/JKO
이왕 산 중턱까지 온 바에 꼭대기까지 올라갑시다.	바__005/NNB+에/JKB
우리의 강한 의지를 내외에 천명하는 바이다.	바__006/NNB+이/VCP+다/EF+./SF

### 3.1.2. 대명사(NP)

- 동일한 형태의 대명사가 다른 의미로 쓰이는 경우가 있으므로, 분석 대상 어휘가 쓰인 문장의 구조를 파악하여 <우리말샘>의 뜻풀이와 일치하는 대명사의 어휘의미 번호를 부여해야 한다.

그에게 편지 썼니?	그__001/NP+에게/JKB	<우리말샘>에 대명사 ‘그’는 ‘말하는 이와 듣는 이가 아닌 사람을 가리키는 삼인칭 대명사’인 ‘001’의 의미와 ‘앞에서 이미 이야기하였거나 듣는 이가 생각하고 있는 대상을 가리키는 지시 대명사’인 ‘002’의 의미로 구분되어 있다.
그와 같은 사실.	그__002/NP+와/JKB	

### 3.1.3. 수사(NR)

- 수사는 때로 수사와 수 관형사, 수사와 명사의 구별이 애매한 경우가 있다. 이에 형태 분석에서는 특이한 형식을 가진 예만을 수 관형사로 취급하고, 그 밖의 것들은 모두 수사로 분석하였다. 이에 어휘의미 분석에서도 수사로 처리된 어휘에 대해 수사의 의미 번호를 부여하여 품사와 어휘의미 번호를 일치하게 한다.

우리는 하나로 뭉쳤다.	하나__001/NR+로/JKB
만점자가 주위에 한둘이 아니다.	한둘__001/NR+이/JKC
상대에게 몇 차례 결정적 기회를 내줬다.	몇__001/NR
지금도 몇몇 분야는 두각을 나타내고 있다.	몇몇__001/NR
농사지은 것으로 일 년 내내 밥상을 차렸다.	일__018/NR

## 3.2. 용언

- 하나의 형태가 동사와 형용사, 보조용언 모두로 쓰일 수 있을 때, 형태 분석에 근거하여 의미 번호를 부여한다. 단, 형태 분석에 오류가 있을 경우 형태 분석 결과를 올바르게 수정하고 이를 별도로 기록하여 통합 분석 말뭉치의 개선에 반영한다.

1시간 가량 조용히 있다가 갑자기 일어났다.	있__001/VV+다가/EC	<우리말샘>에 ‘있다’는 동사(001~004), 형용사(005~021), 보조용언(022~023)으로 등재되어 있다. 이 경우 형태 분석 결과와 동일한 품사에 해당하는 의미 번호를 분석 대상 어휘에 부여한다.
현재 갈등에도 분명 해결책이 있을 것이다.	있__006/VA+을/ETM	
서울에 살고 있는 동생	있__023/VX+는/ETM	

### 3.2.1. 동사(VV)

- 뜻을풀이를 중심으로 갈래뜻의 차이를 반영하여 <우리말샘>에 등재된 의미를 최대한 부여한다. 다만 뜻을풀이만으로 명확하지 않은 경우 예문의 쓰임을 보고 의미 번호 부여 가능성을 판단한다.

선수단이 직접 닭과 돼지를 기르기도 한다.	기르__001/VV+기/ETN+도/JX	<우리말샘>에 “001 【…을】 동식물을 보살펴 자라게 하다”, “002 【…을】 아이를 보살펴 키우다”, “003 【…을】 사람을 가르쳐 키우다” 등으로 등재되어 있는 ‘기르다’는 보살펴 자라게 하는 대상에 따라 구분되며, 물리적인 돌봄과 정신적인 가르침 등으로 구분된다. 이처럼 <우리말샘>에 나타나는 의미 차를 반영하여 분석 대상 어휘의 의미 번호를 부여 한다.
아이를 낳고 기를 수 있는 사회 환경을 만들어야 한다.	기르__002/VV+ㄹ/ETM	
광복 이후에는 광주국악원을 만들어 후배를 길렀다.	기르__003/VV+었/EP+다/EF+./SF	
체체파리는 것을 먹여 새끼를 뱃속에서 기르는데 평생 그 수는 6마리에 그친다.	기르__002/VV+는데/EC	위에서 살핀 바와 같이 ‘기르다’는 대상이 동식물인 경우(001)와 아이인 경우(002)로 구분된다. 위 문장에 나타나는 보살핌의 대상은 체체파리의 새끼, 즉 동식물에 속하지만 주체인 체체파리 입장에서는 ‘아이’에 해당하므로 <우리말샘>에

		등재된 의미 “【…을】 아이를 보살펴 키우다”의 어휘의미 번호 ‘002’를 부여한다.
--	--	---

- 분석 대상 어휘가 쓰인 문장의 구조를 파악하여 해당 용언이 요구하는 문장 성분에 따라 어휘의미 번호를 부여한다. 특히 갈래뜻 사이에 구분이 불분명한 경우 문장 구조에 따라 알맞은 어휘의미 번호를 부여할 수 있다. 이때 전후 맥락을 파악하여 생략된 문장 성분을 복원하여 분석한다.

이 팀은 종합예술을 만든다는 게 강점이다.	만들__011/VV+L 다는/ETM	<우리말샘>에 13개의 의미로 등재되어 있는 ‘만들다’는 3개의 문장 유형으로 구분된다. 대 다수는 목적격 조사 {을}을 요구하는 문장으로, 위의 첫 문장이 이에 해당한다. 따라서 동 일한 문장 유형의 의미 번호(001~011) 중 ‘영화나 드라마 따위를 제작하다’를 의미하는 011을 부여한다. 선행 용언에 연결어미 {-게}가 결합한 두 번째 문장 역시 동일한 문장 구조를 가지는 013을, {으로}가 결합된 부사어가 나타난 세 번째 문장은 {을}과 {으로}를 동시에 요구하는 012를 각각 부여한다. 이때 전후 맥락을 통하여 세 번째 문장에 생략된 목적어 “특정인물”을 복원하여 분석한다.
그간 정부가 밝혀온 원칙을 무색하게 만들었다.	만들__013/VV+었/EP+다/EF+./SF	
게시판의 열독자로 만들 수 있다.	만들__012/VV+ㄹ/ETM	
위층 사람들이 아래층에 내려왔다.	내려오__001//VV+았/EP+다/EF+./SF	<우리말샘>에 등재된 ‘내려오다’의 의미 중 “001 【…에】 【…으로】 높은 곳에서 낮은 곳으로 또는 위에서 아래로 가다”와 “008 【…을】 높은 곳에서 낮은 곳으로 위치를 옮기다”는 갈래뜻 사이에 구분이 불분명하다. 이러한 경우 문장 구조를 통해 어휘의미 번호를 부여할 수 있는데, ‘001’은 주로 도착 지점과 함께 쓰이고, ‘008’은 주로 내려오는 행위가 이루어지는 장소와 함께 나타난다. <우리말샘>에서는 ‘001’과 관련하여 조사 ‘에’와 ‘으로’를 제시하고 있으나 도착점과 관련된 ‘까지’도 ‘001’에 포함될 수 있다. ‘008’의 경우 조사 ‘을’뿐만 아니라 행동이 이루어지고 있는 장소와 관련된 ‘에서’도 포함될 수 있다.
산 밑으로 흘러 내려오는 물을 받아 온다.	내려오__001/VV+는/ETM	
정상을 거쳐 성판악휴게소까지 내려오는 등산코스	내려오__001/VV+는/ETM	
간신히 산을 내려온 이들의 사연을 앞다퉈 보도했다.	내려오__008/VV+L/ETM	
집 잘 지어놔오니 산에서 내려오면 들려 달라.	내려오__008/VV+면/EC	
흥기를 든 20대 남성이 계단에서 내려오고 있었다.	내려오__008/VV+고/EC	

- 다만 뜻풀이에 ‘따위’ 등이 사용되어 대상이 확장될 가능성이 있을 경우에는 해당 의미 번호를 부여한다. 이때 대상의 확장 가능성은 <우리말샘> 어휘지도에 제시된 반의어와 유의어를 통해 확인할 수 있다.

산 정상 부근의 얼음이 급속히 녹아든 것이 원인이다.	녹아들__001/VV+L/ETM	<우리말샘>에 등재된 ‘녹아들다’는 “001 【…에】 다른 물질에 스며들거나 녹아 들어가다”와 “002 【…에】 사상이나 문화 따위가 섞여 어울리다”로 구분된다. 따라서 얼음이 녹아 든 것은 ‘001’이 되고, 영향이 녹아든 것은 ‘002’가 된다. ‘삼성화재의 배구’의 경우 팀 분위기 혹은 경기 방식을 가리키는데, 이는 ‘사상이나 문화 따위’에 속하는 것으로 해당 분석 대상 어휘에 어휘의미 번호 ‘002’를 부여한다.
현대 클래식의 영향이 두루 녹아들었다.	녹아들__002/VV+었/EP+다/EF+./SF	
그는 삼성화재의 배구에 녹아들지 못하고 있다.	녹아들__002/VV+지/EC	
시나몬 향이 청량하면서도 은은한 향기를 자아낸다.	자아내__003/VV+L 다/EF+./SF	위 문장의 ‘향기를 자아낸다’는 “향기가 저절로 생겨난다” 정도의 의미인데, 이는 <우리말샘>의 등재 의미 “003 【…을】 어떤 감정이나 생각, 웃

		음, 눈물 따위가 저절로 생기거나 나오도록 일으켜 내다”와 관련된 것으로 판단된다. <우리말샘>에 따르면 ‘자아내다003’의 대상은 ‘감정이나 생각, 웃음, 눈물 따위’인데 여기에 ‘향기’가 포함될 가능성이 있기 때문에 위 문장의 ‘자아내다’에 어휘의미 번호 ‘003’을 부여한다.
인터넷세대 당기는 복간본 시집의 매력	당기__888/VV+는/ETM	위 문장에 사용된 ‘당기다’는 <우리말샘>의 등재 의미 “003 【…을】 물건 따위를 힘을 주어 자기 쪽이나 일정한 방향으로 가까이 오게 하다”와의 관련성을 생각해 볼 수 있다. 이때 뜻풀이에 제시된 대상의 범주인 ‘물건 따위’에 ‘인터넷세대’가 포함될 수 있는가를 판단해야 한다. 여기서 활용할 수 있는 것이 ‘당기다003’의 반의어 ‘밀다’이다. 위 문장에 사용된 ‘당기다’는 ‘밀다’와 반의어 관계를 형성하지 않으므로 ‘003’보다는 등재되지 않은 새로운 의미로 판단된다. 실제로 위 문장의 ‘당기다’는 “유혹하다” 정도의 의미로 판단되므로 어휘의미 번호 ‘888’을 부여한다.
새로운 민원이 5건 넘게 들어와 방법을 모색하는 중이다.	들어오__888/VV+아/EC	위 문장에 사용된 ‘들어오다’는 <우리말샘>의 등재 의미 “002 【…에】 【…으로】 수입 따위가 생기다”와의 관련성을 생각해 볼 수 있다. 이때 뜻풀이에 제시된 대상의 범주인 ‘수입 따위’에 ‘민원’이 포함될 수 있는가를 판단해야 한다. 여기서 활용할 수 있는 것이 ‘들어오다002’의 반의어 ‘나가다’이다. 위 문장에 사용된 ‘들어오다’는 ‘나가다’와 반의어 관계를 형성하지 않으므로 ‘002’보다는 등재되지 않은 새로운 의미로 판단된다. 실제로 위 문장의 ‘들어오다’는 “접수되다” 정도의 의미로 판단되므로 어휘의미 번호 ‘888’을 부여한다.

- 분석 대상 용언이 선행하는 체언구와 함께 어우러져 비유적 표현의 일부로 사용된 경우, 비유적 표현 전체 의미의 서술어 부분에 해당하는 의미가 등재되지 않은 경우 ‘888’을 부여한다. 이때 비유적 표현의 전체 의미는 <우리말샘> ‘속담·관용구’를 참고한다.

여당과 야당이 손을 잡다.	잡__888/VV+다/EF+./SF	위 문장에 쓰인 ‘잡다’는 명사구 ‘손을’과 결합하여 “서로 힘을 합쳐 협력하다” 정도를 의미한다. 그러나 <우리말샘>에 등재된 ‘잡다’의 다양한 의미 중 “서로 힘을 합쳐 협력하다”의 서술구인 ‘협력하다’에 해당하는 의미가 없으므로 ‘잡다’에 어휘의미 번호 ‘888’을 부여한다.
부모들은 억장이 무너지는 기분이라고 입을 모았다.	모으__888/VV+았/EP+다/EF+./SF	위 문장에 쓰인 ‘모으다’는 흔히 ‘입’과 함께 쓰여 “여러 사람이 같은 의견을 말하다”를 의미하는데, <우리말샘>에 등재된 ‘모으다’의 다양한 의미 중 “여러 사람이 같은 의견을 말하다”의 서술구인 ‘의견을 말하다’ 정도에 해당하는 의미가 없으므로 ‘모으다’에 어휘의미 번호 ‘888’을 부여한다.
사업 아이템이 줄줄이 열매를 맺었다.	맺__888/VV+었/EP+다/EF+./SF	위 문장에 쓰인 ‘맺다’는 ‘열매’와 함께 쓰여 “노력한 일의 성과가 나타나다”, 즉 “성과를 내다”를 의미하는데, <우리말샘>에 등재된 ‘맺다’의 다양한 의미 중 “노력한 일의 성과를 내다”의

		서술구인 ‘성과를 내다’ 정도에 해당하는 의미가 없으므로 ‘맷다’에 어휘의미 번호 ‘888’을 부여한다.
--	--	--

- 말뭉치에 나타난 분석 대상 용언이 <우리말샘> 뜻풀이에서 제시하고 있는 대상 이외의 체언류와 함께 사용된 경우, 어휘의미 번호 ‘888’을 부여한다.

구시가지지를 갈아엎고 고층 아파트를 올렸다.	갈아엎__001/VV+고/EC	<우리말샘>에 등재된 ‘갈아엎다’는 “【…을】 땅을 갈아서 흙을 뒤집어엎다”처럼 흙을 대상으로 풀이되어 있다. ‘구시가지지를 갈아엎고’의 경우 특정 지역의 땅을 갈아서 흙을 뒤집어 엎은 것과 관련되나, ‘판을 갈아엎는’의 경우 실제 땅과 흙이 아니라 판, 즉 분위기 정도를 뒤집어엎은 것으로 분석된다. 이처럼 대상이 땅, 흙과 관련된 ‘갈아엎다’에는 어휘의미 번호 ‘001’을 부여하고, 그 밖에 대상과 함께 쓰인 ‘갈아엎다’에는 어휘의미 번호 ‘888’을 부여한다.
판을 완전히 갈아엎는 인사다.	갈아엎__888/VV+는/ETM	
대체 투입된 버스로 갈아타 인명피해는 없었다.	갈아타__001/VV+아/EC	<우리말샘>에 등재된 ‘갈아타다’는 “【…으로】【…을】 타고 가던 것에서 내려 다른 것으로 바꾸어 타다”로 등재되어 있는데, 예문을 살펴보면 ‘버스, 지하철, 비행기, 말’ 등 탈것을 대상으로 한다. 그러나 실제 쓰임에서는 ‘대출, 정당, 펀드’ 등 탈것과 무관한 것을 대상으로 하며 “옮기다” 정도를 의미한다. 이처럼 <우리말샘>에 등재되지 않은 ‘갈아타다’에는 어휘의미 번호 ‘888’을 부여한다.
저금리 대출로 갈아탈 수 있다	갈아타__888/VV+ㄹ/ETM	

### 3.2.2. 형용사(VA)

- 뜻을풀이를 중심으로 <우리말샘>에 등재된 의미를 최대한 부여한다. 다만 뜻풀이만으로 명확하지 않은 경우 예문의 쓰임을 보고 의미 번호를 부여한다.

취득세 감면 시한이 다가오면서 가파른 상승곡선을 그렸다.	가파른__001/VA+ㄴ/ETM	<우리말샘>에 따르면 ‘가파르다’는 “001 산이나 길이 몹시 기울어져 있다”를 의미한다. 위 문장의 경우 ‘산’이나 ‘길’이 아닌 ‘주택 거래량’의 상태를 가리키는데, 뜻풀이만으로는 거래량의 모습과 산의 모습을 나타내는 ‘가파르다’가 동일한 의미인지 파악하기 어렵다. 하지만 ‘001’의 예문 ‘이에 따라 휘발유 소비량이 가파른 상승세를 보이고 있습니다.」 MBC 뉴스데스크 1998년 3월 》’를 통해 거래량이 몹시 기울어진 모습도 ‘001’에 해당하는 것으로 판단할 수 있다. 따라서 위 문장에 사용된 ‘가파르다’에 어휘의미 번호 ‘001’을 부여한다.
---------------------------------	-------------------	--

- 문장의 구조로 어휘의미 번호를 확정하기 어려운 경우, 분석 대상 어휘와 공기하는 체언류의 의미 부류에 근거하여 어휘의미를 분석한다.

마술사는 아무 일도 없었다는 듯이 무대 위로 걸어 올라왔다	없__003/VA+였/EP+다/EF+는/ETM	<우리말샘>에 ‘없다’는 16개의 의미로 등재되어 있다. 이중 15개가 형용사와 관련되는데 특별한
----------------------------------	---------------------------	--

다.		부사격 조사를 취하지 않는 유형과 {에}나 {에게} 등을 요구하는 유형으로 구분된다. 위의 네 문장은 요구하는 문장 성분 측면에서 동일하기 때문에 문장의 구조로 의미를 분석하기 어렵다. 이러한 경우 <우리말샘>에 뜻풀이를 최대한 활용하여 어휘의미를 분석한다. 즉 첫 번째 문장은 마술사가 아무 일도 생겨 나타나지 않은 듯이 행동하고 있으므로 “003 어떤 일이나 현상이나 증상 따위가 생겨 나타나지 않은 상태이다”로 분석 가능하고, 두 번째 문장은 ‘별로’라는 부사를 통해 “004 어떤 것이 많지 않은 상태이다”로 분석할 수 있다. 그러나 무언가가 존재하지 않은 상태를 나타내는 세 번째 문장과 네 번째 문장은 ‘없다’와 공기하는 체언류의 의미 부류에 근거하여 의미 번호를 부여한다. <우리말샘>에 따르면 001은 ‘사람, 동물, 물체’ 등 구체적인 대상과 함께 나타나는 데 반해, 002는 ‘사실이나 현상’처럼 추상적인 것과 주로 어울린다. 이를 바탕으로 위의 세 번째 문장과 네 번째 문장에 사용된 ‘없다’의 의미를 구별할 수 있다.
나라간 이견이 큰 현안에서는 진전이 별로 없었다고 해석했다.	없__004/VA+었/EP+다/EF+고/JKQ	
구속 담당 장관이 없다는 이유로	없__001/VA+다/EF+는/ETM	
처벌 가치가 없는 것으로 판단했다.	없__002/VA+는/ETM	

### 3.2.3. 지정사(VC)

#### 3.2.3.1. 긍정지정사(VCP)

- 긍정지정사(VCP) ‘이다’는 <우리말샘>에 조사 ‘이다’로 등재된 어휘의미 번호를 부여한다. ‘이다’는 <우리말샘>에 결합하는 성분에 따라 구분되어 ‘004~008’의 서술격 조사로 등재되어 있다. 선행하는 성분이 체언인지 접미사 ‘-적’이 결합된 형태인지, 혹은 부사나 연결어미인지를 고려하여 알맞은 어휘의미를 부여한다.

철수는 자기가 학생이라고 말했다.	학생__002/NNG+이__004/VCP+라/EF+고/JKQ
세계적인 스타들이 펼치는 축구의 향연	세계적__001/NNG+이__005/VCP+L/ETM
그의 강점을 살린 영리한 선택이었다.	선택__001/NNG+이__006/VCP+었/EP+다/EF+./SF
불법 감금 기간이 상당히 오래이고	오래/MAG+이__007/VCP+고/EC
이사를 생각한 것은 집이 좁게 느껴서다.	느끼__006/VV+어서/EC+이__008/VCP+다/EF+./SF

#### 3.2.3.2. 부정지정사(VCN)

- 부정지정사(VCN) ‘아니다’는 <우리말샘>에 형용사 ‘아니다’로 등재된 어휘의미 번호를 부여한다.

봄이 와서 꽃이 피는 것이 아니라	아니__001/VCN+라/EC
선점 효과를 막으려 하는 게 아니냐는 얘기가.	아니__002/VCN+냐/EF+는/ETM

### 3.2.4. 보조용언(VX)

- 보조동사와 보조형용사는 모두 보조용언(VX)으로 형태 분석되어 있다. 보조용언(VX)은 <우리말샘>에 등재된 보조동사와 보조형용사 목록을 참고하여 일치하는 의미의 어휘의미 번호를 부여한다.

일을 마치고 나니 기분이 상쾌해졌다.	나__032/VX+니/EC
비가 오는가 싶어 빨래를 걷었다.	싶__002/VX+어/EC

## 4. 주의사항

### 4.1. 북한어 및 방언인 경우

- 말뭉치의 분석 대상 어휘가 <우리말샘>에 북한어, 방언 등으로 등재되어 있으면 해당 어휘의미 번호를 부여한다.

며느라, 차례 음식은 대행에 맡기렴.	며느__001/NNG+아/JKV+/,SP	며느001 : '며느리'의 방언(경북, 중국 흑룡강성).
애님니께~~~~	애 니 __ 0 0 1 / V C N + 님 니 께 /EF+~/SO+~/SO+~/SO+~/SO	애니다001 : '아니다'의 방언(경북).
아빠 뭘 뽀수고	뽀수__001/VV+고/EC	뽀수다001 : '부수다'의 방언(강원, 경기, 경상, 전라, 제주, 충청).
국수 한그릇 후루룩	한/MMN+그릇__001/NNG	그릇001 : '그릇'의 방언(강원, 경상, 전남, 충남).
멀미땀시 이따 다시올게요	멀미__001/NNG+땀시__001/NNB	"때문에(전북)"의 방언형으로 등재된 '땀시'는 품사 정보가 제시되지 않아 어떤 품사로든 연결이 가능하다. 따라서 위 문장에 사용된 '땀시'에 어휘의미 번호 '001'을 부여한다.

### 4.2. 외래어인 경우

- <우리말샘>에 등재된 외래어 중에는 규범 표기가 미확정된 것들이 많다. 이처럼 외국어로 된 단어의 경우 표준형을 결정하기가 어렵다. 따라서 어휘의미 분석 대상 어휘가 외국어로 된 단어인 경우 해당 어휘의 사전 등재 유무에 따라 어휘의미 번호를 부여한다.

그냥 타지마할 사진 찍었다고 해도 믿겠어요	타지마할__001/NNP	
키자니아 엄청 자주 가던데	키자니아__777/NNP	
에픽호이 타블라	에픽호이__777/NNP	
난 크리스토프도 종던덩...친구같	크리스토프__888/NNP+도/JX	



아서		
냉 가니쉬랑 다 들어있어서 요리하기 간편해요~~~	가니쉬__777/NNG+랑/JC	<우리말샘>에 '가니시'가 등재되어 있으나 규범 표기가 미확정이기 때문에 이를 표준형으로 보기 어렵다. 이에 '가니쉬' 역시 가능한 표기로 판단하여 어휘의미 번호 '777'을 부여한다.
룸싸롱 못가서 그렇습니다	룸싸롱__777/NNG	<우리말샘>에 '룸살롱'이 규범적 표기로 등재되어 있으나 다른 외래어와 동일하게 다양한 발음의 가능성을 인정하여 어휘의미 번호 '777'을 부여한다.

### 4.3. 숫자나 알파벳, 한자가 포함된 경우

- 숫자나 알파벳, 한자가 포함된 단어는 <우리말샘>에 해당 숫자나 알파벳을 한글로 옮겨 쓴 형태가 등재되어 있으면 해당 어휘의미 번호를 부여한다.

모두 한국신기록을 세우며 대회 3관왕이 됐다.	3관왕__002/NNG+이/JKC
U턴 문제 등 우려하는 문제가 불거지지 않도록 하겠다.	U턴__001/NNG
많은 사람들이 3.1절 기념식에 참석했다.	3.1절__001/NNP
A.I로봇이 생산 라인에 투입되었다.	A.I로봇__001/NNG+이/JKS
韓中 수교 30주년 기념 가요제가 열렸다.	韓中__007/NNP

### 4.4. 기호가 포함된 경우

- <우리말샘>에 가운뎃점을 생략한 형태가 등재되어 있으면 해당 어휘의미 번호를 부여한다.

건물의 냉·난방비를 줄여준다.	냉·난방비__001/NNG+를/JKO
국·공립대 중 1인당 장학금 수혜율 3위를 기록하고 있다.	국·공립대__001/NNG

- <우리말샘> 등재어가 괄호나 따옴표로 어근과 접사가 분리된 경우에는 어근과 접사를 분리하여 각각에 맞는 어휘의미 번호를 부여한다.

스터디(study)하여	스터디__001/NNG+(/SS+study__001/NNG+)/SS+하/XSV+아/EC
그 여자는 자신의 다이어트 비법을 '공유'했다.	'/SS+공유__002/NNG+ '/SS+하/XSV+았/EP+다/EF+./SF

### 4.5. 문자 모양의 유사성에 기반한 경우

- 문자 모양의 유사성에 기반하여 형태를 변형한 단어는 오표기로 판단하지 않고 <우리말샘> 등재

유무에 따라 해당 어휘의미 번호를 부여한다.

땡땡이는 산책 안 시킴 힘들 어지죠	땡땡이__003/NNG+는/JX	‘땡땡이’는 ‘땡땡이’를 문자 모양의 유사성에 기반하여 변형한 단어이다. <우리말샘>에 ‘땡땡이’가 등재되어 있으므로 해당 의미번호인 ‘003’을 부여한다.
팔도네넬핀 이것도 좋아요	팔도네넬핀__777/NNP	‘팔도네넬핀’는 ‘팔도비빔면’을 문자 모양의 유사성에 기반하여 변형한 단어이다. <우리말샘>에 해당 형태가 등재되어 있지 않으므로 어휘의미 번호 ‘777’을 부여한다.

## 4.6. 구 등재어인 경우

- 어휘의미 분석은 형태 분석 말뭉치의 분석 결과를 바탕으로 한다. 즉 형태 분석 지침에 의거하여 구성 요소가 분리되어 분석된 경우에는 각 구성 요소에 어휘의미 번호를 부여하고, 하나로 분석된 어휘에는 하나의 어휘의미 번호를 부여한다. 고유명사는 형태 분석 지침에 따라 하나의 어절로 분석되어 있다. 형태 분석 결과 구로 이루어진 고유명사의 어절이 분리되어 있는 경우 이를 하나의 어절로 통합하고 <우리말샘>을 참고하여 해당하는 어휘의미 번호를 부여한다.

국사편찬위원회 위원장으 로써 올해의 계획이 또 있으실 것 같거든요?	국 사 편 찬 위 원 회 __777/NNG	
국가인권위원회는 학습권 보장의 필요성을 제기했 다.	국 가 인 권 위 원 회 __001/NNP+는/JX	‘국가인권위원회’는 형태 분석 지침에 따라 하나의 고유명사로 분석되기 때문에 <우리말샘>에 구로 등재된 ‘국가^인권^위원회’의 어휘의미 번호를 ‘국가인권위원회 001/NNP’처럼 분석 대상 어휘 전체에 하나의 어휘의미 번호를 부여한다.
이 건물은 평소 템플스테이 숙소로 쓰여 왔다.	템플스테이__001/NNG	

## 4.7. <우리말샘> 등재어가 여러 개인 경우

- 로마자 약자로 이루어진 단어의 <우리말샘> 등재어가 여러 개인 경우, 문자의 이름으로 이루어진 등재어의 어휘의미 번호를 우선적으로 부여하고 그러한 형태가 등재되어 있지 않은 경우에는 단어로 읽는 등재어의 어휘의미 번호를 부여한다.

FIFA/NNP	피파, 에프아이에프에이, 국제축구연맹	에프아이에프에이__001
IMF/NNP	아이엠에프, 국제통화기 금	아이엠에프__001

## 4.8. 참여자 제안 정보인 경우

- <우리말샘> ‘참여자 제안 정보’에 제시된 단어는 어휘의미 분석에 활용하지 않는다.

아삿추는 진리쵸	아삿추__777/NNG+는/JX	'아삿추'가 <우리말샘> '참여자 제안 정보'에 "'아이스티에 삿' 추가'의 줄임말"로 제안되었으나, 이를 고려하지 않고 어휘의미 번호 '777'을 부여한다.
안먹는게 중요하쵸 행님	행님__777/NNG	'행님'이 <우리말샘> '참여자 제안 정보'에 "'깡패집단 사이에서 '형님'을 부르는 말"로 제안되었으나, 이를 고려하지 않고 어휘의미 번호 '777'을 부여한다.

## 4.9. 맥락이 불완전한 경우

- 맥락이 불완전하여 의미 분석이 곤란한 단어는 어휘의미 번호 '001'을 부여한다.

고물	고물__001/NNG	위의 '고물'은 끝말잇기 상황에서 출현한 단어로 맥락상 그 의미 파악이 불가능하다. 이 경우 기본 의미로 판단되는 '001'을 부여한다.
----	-------------	--

## [붙임3] 구문 분석 말뭉치 구축 지침

### 구문 분석 말뭉치 구축 지침

## 1. 총론

### 1.1. 지침의 목적

이 지침은 한국어 문장을 구문 분석하여 의존 구문 분석 말뭉치를 구축하는 방안을 제시하는 것을 목적으로 한다. 한국어 문장의 각 어절에 구문 표지와 기능 표지를 부착하고 어절들 간의 지배-의존 관계를 표시하는 것이 세부 목적이다. 구문 분석은 형태 분석 말뭉치를 대상으로 수행한다.

- 본 지침은 ‘2024년 구문 및 무형 대용어 복원 말뭉치 구축 지침’을 토대로 작성한 것이다.
- 본 지침은 이전 지침과의 내용적 일관성을 최대한 유지하되 층위별 말뭉치 구축을 염두에 두고 지침을 개정·보완하였다.
- 본 사업에서는 원문을 정제하고 가공하여 사용하게 됨에 따라 이전 지침에 있던 말뭉치 오류 및 예외 처리에 관련된 사항은 대폭 축소하였다.
- 본 지침은 이전 지침들과 마찬가지로 한국정보통신기술협회(TTA) 및 한국전자통신연구원(ETRI)의 기준을 준수하되 언어학적으로 정합적이지 않은 내용은 수정하였다.
- 본 지침은 이후 의미역 분석, 무형 대용어 복원 등 단계적인 분석을 할 수 있도록 고려하였다.

### 1.2. 용어 정의

- 구문 표지: 단어, 구, 절의 통사적 성격을 규정한 표지 표지로 어절 앞부분의 형태 정보를 바탕으로 부여한다.
- 기능 표지: 해당 어절이 문장 내의 다른 어절과 가지는 문법적 관계를 규정한 표지로 어절 뒷부분의 형태 정보를 바탕으로 부여한다.
- 지배-의존 관계: 문장 내 지배소와 의존소 사이의 관계. 다음 세 가지로 분류된다.
  - 논항-서술어 관계
  - 수식어-피수식어 관계
  - 접속 관계

### 1.3. 기본 원칙

- (1) 자연언어처리를 위한 일관성 유지와 효율성 제고에 초점을 두되, 일반 언어학적 관점에서도 크게 벗어나지 않도록 한다.
- (2) 문장의 표층 구조를 중시하여 분석한다.
- (3) 의존 관계 분석의 기본 단위로 어절을 사용한다.
- (4) 지배소 후위 원칙에 따라 각 어절의 지배소는 자신보다 뒤에 위치하도록 분석한다.
- (5) 각 어절은 1개의 지배소를 가진다. (Single-Head Constraint)
- (6) 각 어절 및 지배소쌍은 서로 교차하지 않는다. (Projective Constraint)
- (7) 보어와 부가어를 구분하되 보어의 범위를 엄격히 제한한다.
- (8) 원칙적으로 접속과 내포를 구별하지 않으며, 접속절은 모두 부사절로 분석한다. (다만, 명사구 접속은 인정한다.)
- (9) 하나의 성분이 모절과 내포절 모두에 관련되어 있으면 내포절의 유형에 따라 해당 주어의 지배소를 결정한다.
- (10) 하나의 어절이 여러 가지 형태로 이루어진 경우 어절 앞부분의 형태 정보를 통해 해당 어절의 구문 표지를 결정하고 어절 뒷부분의 형태 정보를 통해 해당 어절의 기능 표지를 결정한다.

### 2. 구문 분석 표지 부착

## 2.1. 구문 분석 표지 집합

### 2.1.1. 구문 표지 집합

구문 표지	의미	예시
NP	체언 (명사, 대명사, 수사)	-비가(NP) 와서 우산을(NP) 샀다. -공신의 딸을(NP) 부인으로(NP) 삼았다. -그는(NP) 셋을(NP) 세었다.
VP	용언 (동사, 형용사, 보조용언)	-비가 와서(VP) 우산을 샀다(VP). -피부가 몹시 건조하다(VP).
VNP	긍정 지정사구 (명사+이다)	-이게 우리집이야(VNP). -할머니는 걱정이셨다(VNP).
DP	관형사구	-벌써 새(DP) 학기가 되었다. -한(DP) 마흔(DP) 살쯤 되어 보인다.
AP	부사구	-전화 잠깐(AP) 써도 될까요? -피부가 몹시(AP) 건조하다.
IP	감탄사구(호칭 및 대답 등의 표현)	-아이고!(IP) 이 일을 어쩔까. -선생님! (IP) 질문이 있습니다.
X	의사 구 (pseudo phrase, 조사 단독 어절 또는 기호 등)	-(五)(X)
L	부호 (왼쪽 괄호 및 따옴표)	-당신의 선택은?( (L) )
R	부호 (오른쪽 괄호 및 따옴표)	-당신의 선택은? ( ) (R)

### 2.1.2. 기능 표지 집합

기능 표지	의미	예시
SBJ	주어	-비가(NP_SBJ) 와서 우산을 샀다. -그는(NP_SBJ) 셋을 세었다.
OBJ	목적어	-비가 와서 우산을(NP_OBJ) 샀다. -그는 셋을(NP_OBJ) 세었다.
CMP	보어	-나는 학생이(NP_CMP) 아니다. -얼음이 물이(NP_CMP) 되었다.
MOD	관형어 (체언 수식어)	-그(DP) 시계가 필요하다. -낮익은(VP_MOD) 자동차 한 대가 내려왔다. -어제 본(VP_MOD) 영화는 무척 재미있었다.
AJT	부사어 (용언 수식어)	-나는 서울에(NP_AJT) 산다. -과일을 칼로(NP_AJT) 잘랐다. -나는 싫다고(VP_AJT) 말했다.
CNJ	접속어	-철수와(NP_CNJ) 영희는 친구이다. -개와(NP_CNJ) 고양이의 관계.

## 2.2. 구문 분석 표지의 부착

- 한국어의 한 어절은 대개 어휘 요소와 문법 요소의 결합으로 이루어져 있다. 구문 표지는 어휘 요

소에 기반하여 결정하고, 기능 표지는 문법 요소에 기반하여 결정한다.

### 2.2.1. 구문 표지의 부착

- 구문 표지에는 NP, VP, VNP, DP, AP, IP, X, L, R이 있다. 구문 표지는 해당 어절의 범주적 특성을 표시하는 것으로서 어휘 요소에 기반하여 결정한다.
- 어휘 요소의 형태 분석 표지와 구문 표지의 전형적인 대응관계를 보이면 다음과 같다.

어휘 요소의 형태 분석 표지	구문 표지
NNG(일반명사) NNP(고유명사) NNB(의존명사) NP(대명사) NR(수사) XR(어근)	NP(체언)
VV(동사) VA(형용사) VCN(부정지정사) VX(보조용언)	VP(용언)
VCP(긍정지정사)	VNP(긍정 지정사구)
MMA(성상 관형사) MMD(지시 관형사) MMN(수 관형사)	DP(관형사구)
MAG(일반부사) MAJ(접속부사)	AP(부사구)
IC(감탄사)	IP(감탄사구)
SP(침표, 가운데점, 콜론, 빗금) SO(붙임표) SL(외국 문자) SH(한자)	X(의사 구)
SS(따옴표, 괄호표가 단독으로 쓰이는 경우)	L(왼쪽 부호)
SS(따옴표, 괄호표가 단독으로 쓰이는 경우)	R(오른쪽 부호)

### 2.2.2. 기능 표지의 부착

- 기능 표지에는 SBJ, OBJ, CMP, MOD, AJT, CNJ가 있다. 기능 표지는 해당 어절의 문장 내에서의 기능을 표시하는 것으로서 문법 요소에 기반하여 결정한다. 문법 요소의 형태 분석 표지와 기능 표지의 전형적인 대응관계를 보이면 다음과 같다.

문법 요소의 형태 분석 표지	기능 표지
JKS(주격 조사)	SBJ(주어)
JKO(목적격 조사)	OBJ(목적어)
JKC(보격 조사)	CMP(보어)
JKG(관형격 조사) ETM(관형사형 어미)	MOD(관형어)
JKB(부사격 조사)	AJT(부사어)

JC(접속 조사)	CNJ(접속어)
-----------	----------

- 관형사(MMA, MMD, MMN)는 문장 안에서 수식 대상을 한정하는 관형어의 기능을 가지며 일반부사(MAG)는 문장 안에서 수식받는 용언의 내용을 한정하는 부사어의 기능을 가진다. 분석 대상 어절의 형태가 관형사(MMA, MMD, MMN) 또는 일반부사(MAG)로만 분석될 경우에는 구문표지가 DP(관형사구) 또는 AP(부사구)로 분석되어 기능 표지 MOD(관형어)와 AJT(부사어)는 불필요하다.
- 그러나 하나의 어절이 여러 가지 형태로 이루어지고 관형사(MMA, MMD, MMN) 또는 일반부사(MAG)가 어절 뒷부분에 위치하는 경우에는 '어절 뒷부분의 형태 정보를 통해 해당 어절의 기능 표지를 결정한다'는 기본 원칙에 따라 기능 표지 MOD(관형어)와 AJT(부사어) 부여가 가능해짐에 주의해야 한다.

어휘 요소의 형태 분석 표지	기능 표지
MMA(성상 관형사) MMD(지시 관형사) MMN(수 관형사)	MOD(관형어)
MAG(일반 부사)	AJT(부사어)

- JX(보조사)와 같이 여러 기능 표지를 가질 수 있는 경우 해당 어절의 지배소를 먼저 찾고 지배소와 의존소의 관계에 기반하여 기능 표지를 결정한다.

### 3. 지배-의존 관계 분석의 원리

- 지배-의존 관계 설정은 해당 어절이 문장 내의 어떤 어절에 의존하는지를 표시하는 것이다. 즉, 해당 어절의 지배소를 어절 번호로 나타낸다.
- 지배-의존 관계 분석은 다음 세 가지 유형으로 나눌 수 있다.
  - (1) 논항-서술어 관계 분석(→)
  - (2) 수식어-피수식어 관계 분석(→)
  - (3) 접속 관계 분석(→)

예) 추운 겨울이 가고 따뜻한 봄이 왔다.

1	추운	춌/VA+L/ETM	VP MOD	2	
2	겨울이	겨울/NNG+이/JKS	NP SBJ	3	
3	가고	가/VV+고/EC	VP AJT	6	
4	따뜻한	따뜻하/VA+L/ETM	VP MOD	5	
5	봄이	봄/NNG+이/JKS	NP SBJ	6	
6	왔다.	오/VV+았/EP+다/EF+./SF	VP	0	



### 3.1. 논항-서술어 관계 분석

#### 3.1.1. 서술어의 성립

- 지배-의존 관계 분석에서 논항-서술어 관계를 분석하기 위해서는 우선 문장 내에서 서술어를 찾아야 한다. 서술어로 삼는 것은 VP라는 구문 표지를 가진 어절 중 VV, VA의 형태 분석 표지를 가진 것과 VNP라는 구문 표지를 가진 어절, 즉 VCP라는 형태 분석 표지를 가진 것이다.

어휘 요소의 형태 분석 표지	구문 표지
VV(동사) VA(형용사) VCN(부정지정사)	VP(용언)
VCP(긍정지정사)	VNP(긍정 지정사구)
문법 요소의 형태 분석 표지	구문 표지
XSV(동사 파생 접미사) XSA(형용사 파생 접미사)	VP(용언)

- 문형별 분석 방안
- 논항-서술어 관계는 사전의 문형 정보를 참고하여 분석한다. <우리말샘>에는 주어 논항에 대한 정보가 없지만, 모든 서술어는 주어를 가지는 것으로 간주한다. 사전의 문형 정보에 기반하여 각 논항은 서술어에 의존하는 것으로 분석한다.

NP1-이 VP

NP1-이 NP2-을 VP

NP1-이 NP2-이 VP (주격-보격)

NP1-이 NP2-에 VP

NP1-이 NP2-에게 VP

NP1-이 NP2-에서 VP

NP1-이 NP2-로 VP

NP1-이 NP2-이 VP (주격-주격)

NP1-이 NP2-에게 NP3-을 VP

NP1-이 NP2-을 NP3-로 VP

...



(1) NP1-이 VP

예) 배경은 이렇다.

1	배경은	배경 은	NNG+JX	NP SBJ	2		
2	이렇다.	이렇 다 .	VA+EF+SF	VP	0		

(2) NP1-이 NP2-을 VP

예) 해당 분야 선배들이 채용 정보와 상담을 책임진다.

1	해당	해당	NNG	NP	2		
2	분야	분야	NNG	NP	3		
3	선배들이	선배 들 이	NNG+XSN+JKS	NP SBJ	7		
4	채용	채용	NNG	NP	5		
5	정보와	정보 와	NNG+JC	NP CNJ	6		
6	상담을	상담 을	NNG+JKO	NP OBJ	7		
7	책임진다.	책임지 니 다 .	VV+EF+SF	VP	0		

(3) NP1-이 NP2-이 VP (주격-보격)

예) 선이란 게 앉아 있기만 하는 게 아닙니다.

1	선이란	선 이 란	NNG+VCP+ETM	VP MOD	2		
2	게	거 이	NNB+JKS	NP SBJ	7		
3	앉아	안 아	VV+EC	VP	4		
4	있기만	있 기 만	VX+ETN+JX	VP	5		
5	하는	하 는	VX+ETM	VP MOD	6		
6	게	거 이	NNB+JKC	NP_CMP	7		
7	아닙니다.	아니 브 니 다 .	VCN+EF+SF	VP	0		

(4) NP1-이 NP2-에 VP

예) 여기엔 이유가 있었다.

1	여기엔	여기 에	NP+JKB+JX	NP AJT	3		
2	이유가	이유 가	NNG+JKS	NP SBJ	3		
3	있었다.	있 었 다 .	VA+EP+EF+SF	VP	0		

(5) NP1-이 NP2-에게 VP

예) 해마다 미국 대사에게 초청장이 갔지만 매년 거절당했다.

1	해마다	해마다	MAG	AP	5		
---	-----	-----	-----	----	---	--	--

2	미국	미국	NNP	NP	3	
3	대사에게	대사에게	NNG+JKB	NP AJT	5	
4	초청장이	초청장 이	NNG+JKS	NP SBJ	5	
5	갔지만	가 았 지만	VV+EP+EC	VP AJT	7	
6	매번	매번	MAG	AP	7	
7	거절당했다.	거절당하 았 다 .	VV+EP+EF+SF	VP	0	

(6) NP1-이 NP2-에서 VP

예) 남미 칠레에서 33인의 기적이 일어났다.

1	남미	남미	NNP	NP	2	
2	칠레에서	칠레에서	NNP+JKB	NP AJT	5	
3	33인의	33 인 의	SN+NNB+JKG	NP MOD	4	
4	기적이	기적 이	NNG+JKS	NP SBJ	5	
5	일어났다.	일어나 았 다 .	VV+EP+EF+SF	VP	0	

(7) NP1-이 NP2-로 VP

예) 경남에선 처음 열리는 이 박람회는 경남도 내 의료기관, 보건의료단체, 전문병원, 건강생활 체험 등의 프로그램으로 구성된다.

1	경남에선	경남에서	NNP+JKB+JX	NP AJT	3	
2	처음	처음	NNG	NP AJT	3	
3	열리는	열리 는	VV+ETM	VP MOD	5	
4	이	이	MMD	DP	5	
5	박람회는	박람회 는	NNG+JX	NP SBJ	15	
6	경남도	경남도	NNP	NP	7	
7	내	내	NNB	NP	8	
8	의료기관	의료기관	NNG+SP	NP	12	
9	보건의료단체	보건의료 단체	NNG+NNG+SP	NP	12	
10	전문병원	전문병원	NNG+SP	NP	12	
11	건강생활	건강생활	NNG	NP	12	
12	체험	체험	NNG	NP	13	
13	등의	등 의	NNB+JKG	NP MOD	14	
14	프로그램으로	프로그램 으로	NNG+JKB	NP AJT	15	
15	구성된다.	구성되 나 다 .	VV+EF+SF	VP	0	

(8) NP1-이 NP2-이 VP (주격-주격)

- 서술어가 감정형용사인 경우 한 문장에 두 개의 주어가 출현하는 경우가 있다. 두 개의 주어 모두 서술어에 의존하는 것으로 분석한다.

예) 이모(31)씨는 외출이 무섭다.

1	이모(31)씨는	이 모 ( 31 ) 씨 는	NNP+NP+SS+SN+SS+NNB+JX	NP SBJ	3	
2	외출이	외출 이	NNG+JKS	NP SBJ	3	
3	무섭다.	무섭 다 .	VA+EF+SF	VP	0	

- 서술어가 형용사인 경우 한 문장에 두 개의 주어가 출현하는 경우가 있다. 두 개의 주어 모두 서술어에 의존하는 것으로 분석한다. 두 개의 명사구 사이에 전체-부분, 대상-속성 등의 의미 관계가 성립하는데 그렇다고 하여 앞의 주어가 뒤의 주어에 의존하는 것으로 분석하지 않도록 주의한다. 통사적으로는 두 명사구가 모두 서술어에 의존하는 것이다.

예) 그래서 대검찰청 중앙수사부 수사팀은 고민이 크다.

1	그래서	그래서	MAJ	AP	6	
2	대검찰청	대검찰청	NNG	NP	3	
3	중앙수사부	중앙수사부	NNG	NP	4	
4	수사팀은	수사팀 은	NNG+JX	NP SBJ	6	
5	고민이	고민 이	NNG+JKS	NP SBJ	6	
6	크다.	크 다 .	VA+EF+SF	VP	0	

(9) NP1-이 NP2-에게 NP3-을 VP

예) 엄 군이 속한 팀은 전교생에게 큰 웃음을 주고 특기상을 받았다.

1	엄	엄	NNP	NP	2		
2	군이	군 이	NNB+JKS	NP SBJ	3		
3	속하	속하	VV+ETM	VP MOD	4		
4	팀이	팀 이	NNG+JKS	NP SBJ	8		
5	전교생에게	전교생 에게	NNG+JKB	NP AJT	8		
6	큰	크	VA+ETM	VP MOD	7		
7	웃음을	웃음 을	NNG+JKO	NP OBJ	8		
8	주고	주 고	VV+EC	VP AJT	10		
9	특기상을	특기 상 을	NNG+NNG+JKO	NP OBJ	10		
10	받았다.	받 았 다 .	VV+EP+EF+SF	VP	0		

(10) NP1-이 NP2-을 NP3-로 VP

예) 중국은 미국에 대한 수출을 성장 동력으로 삼고 있다.

1	중국	중국	NNP+JX	NP SBJ	7		
2	미국에	미국 에	NNP+JKB	NP AJT	3		
3	대한	대하	VV+ETM	VP MOD	4		
4	수출을	수출 을	NNG+JKO	NP OBJ	7		
5	성장을	성장	NNG	NP	6		
6	동력으로	동력 으로	NNG+JKB	NP AJT	7		
7	삼고	삼 고	VV+EC	VP	8		
8	있다.	있 다 .	VX+EF+SF	VP	0		

[주의] 보조용언의 처리

- 보조용언은 VP라는 구문 표지를 가지지만 논항-서술어 관계를 이루지는 않는다.

어휘 요소의 형태 분석 표지	구문 표지
VX(보조동사)	VP(용언)

- 서술어에 '본용언 + 보조 용언' 구성이 사용된 경우 본용언은 보조용언에 의존한다.

예) 멜라닌은 물에는 용해되지 않는다.

1	멜라닌은	NP SBJ	3		
2	물에는	NP AJT	3		
3	용해되지	VP	4		
4	않는다.	VP	0		

- 의존명사를 포함한 우연적 구성이 의사 보조 용언으로 사용된 경우 '본용언 + 보조용언' 구성과 같이 분석한다.

1	그는	NP SBJ	2		
2	입어날	VP MOD	3		
3	수	NP SBJ	4		
4	없었다.	VA	0		

## 3.1.2. 필수 논항의 성립

### 3.1.2.1. 주어의 성립

- 주어는 구문 표지 NP(체언)와 기능 표지 SBJ(주어)를 가진 어절이다. JKS(주격 조사)나 JX(보조사)라는 형태 분석 표지를 가지는 것이 일반적이지만 조사가 생략되어 나타나지 않는 경우도 있다.

(1) 명사류+JKS: NP\_SBJ

예) '펜과 붓'이 함께한 여행...잃어버린 나'를 찾다

1	'펜과	'펜 과	SS+NNG+JC	NP CNJ	2
2	붓'이	붓' 이	NNG+SS+JKS	NP SBJ	3
3	함께한	함께 하 L	MAG+XSV+ETM	VP MOD	4
4	여행...	여행 ...	NNG+SE	NP AJT	7
5	'잃어버린	' 잃어버리 L	SS+VV+ETM	VP MOD	6
6	나'를	나' 를	NP+SS+JKO	NP OBJ	7
7	찾다	찾 다	VV+EF	VP	0

(2) 명사류+JX: NP\_SBJ

예) 하지만 은행 문턱은 더 높아졌다.

1	하지만	하지만	MAJ	AP	5
2	은행	은행	NNG	NP	3
3	문턱은	문턱 은	NNG+JX	NP SBJ	5
4	더	더	MAG	AP	5
5	높아졌다.	높아지 었 다 .	VV+EP+EF+SF	VP	0

(3) 명사류: NP\_SBJ

- 조사가 생략되어 나타나지 않은 경우라도 구문 표지 NP(체언)을 가진 어절이 서술어가 나타내는 행위의 주체, 상태/성질 혹은 정체 발힘의 대상으로 쓰이면 기능 표지 SBJ(주어)를 가진다.

예) 쉬는 시간엔 '방콕'... 리듬 깨질까봐 낮잠도 못자

1	쉬는	쉬 는	VV+ETM	NP MOD	2
2	시간엔	시간 에 L	NNG+JKB+JX	NP AJT	3
3	'방콕'...	' 방콕 ' ...	SS+NNG+SS+SE	NP	8
4	리듬	리듬	NNG	NP SBJ	5
5	깨질까봐	깨지 르까 보 아	VV+EF+VX+EC	VP AJT	8
6	낮잠도	낮잠 도	NNG+JX	NP OBJ	8
7	못	못	MAG	AP	8
8	자	자 아	VV+EF	VP	0

[주의] 보조사적 쓰임을 보이는 '이/가'

- JKS(주격조사)와 형태가 동일하지만 주어를 나타내지 않는 JX(보조사)가 결합되어 있으면 기능 표지 SBJ(주어)를 가지지 않는다.

### 3.1.2.2. 목적어의 성립

- 목적어는 구문 표지 NP(체언)와 기능 표지 OBJ(목적어)를 가진 어절이다. 구문 표지 VP(용언)를 가진 어절 중 형태 분석 표지 VV를 가진 서술어에 선행한다. JKO(목적격 조사)나 JX(보조사)라는 형태 분석 표지를 가지는 것이 일반적이지만 조사가 생략되어 나타나지 않는 경우도 있다.

(1) 명사류+JKO: NP\_OBJ

예) 3월에는 한 달 동안 1000km를 달렸다.

1	3월에는	3월 에 는	NNG+JKB+JX	NP_AJT	6
2	하	하	MMN	DP	3
3	달	달	NNB	NP	4
4	동안	동안	NNG	NP_AJT	6
5	1000km를	1000 km 를	SN+SW+JKO	NP_OBJ	6
6	달렸다.	달리 었 다 .	VV+EP+EF+SF	VP	0

(2) 명사류+JX: NP\_OBJ

예) “대부분 시간을 방에서 보내지만 낮잠도 못 자요.”

1	“대부분이	“ 대부분 이	SS+NNG+JKS	NP_SBJ	4
2	시간을	시간 을	NNG+JKO	NP_OBJ	4
3	방에서	방 에서	NNG+JKB	NP_AJT	4
4	보내지만	보내 지 만	VV+EC+JX	VP_AJT	7
5	낮잠도	낮잠 도	NNG+JX	NP_OBJ	7
6	못	못	MAG	AP	7
7	자요.”	자 아요 . ”	VV+EF+SF+SS	VP	0

(3) 명사류: NP\_OBJ

- 조사가 생략되어 나타나지 않은 경우라도 구문 표지 NP(체언)을 가진 어절이 타동사 서술어가 나타내는 동작이나 작용의 대상으로 쓰이면 기능 표지 OBJ(목적어)를 가진다.

예) 국회의원 3명 중 2명 불황에도 재산 불려

1	국회의원	국회의원	NNG	NP	2
2	3명	3 명	SN+NNB	NP	3
3	중	중	NNB	NP	4
4	2명	2 명	SN+NNB	NP_SBJ	7
5	불황에도	불황 에 도	NNG+JKB+JX	NP_AJT	7
6	재산	재산	NNG	NP_OBJ	7
7	불려	불리 어	VV+EF	VP	0

[주의] 보조사적 쓰임을 보이는 ‘을/를’

- JKO(목적격 조사)와 형태가 동일하지만 목적어를 나타내지 않는 JX(보조사)가 결합되어 있으면 기능 표지 OBJ(목적어)를 가지지 않는다.
- 

### 3.1.2.3. 보어의 성립

- 보어는 구문 표지 NP(체언)와 기능 표지 CMP(보어)를 가진 어절로, 서술어 '되/VV'나 '아니/VCN'에 선행한다. 형태 분석 표지 JKC(보격 조사)를 가지는 것이 일반적이지만 조사가 생략되어 나타나지 않는 경우도 있다.

(1) 명사류+JKC: NP\_CMP

예) 김 지사가 가장 강력한 대안이 될 것이다.

1	김	김	NNP	NP	2
2	지사가	지사 가	NNG+JKS	NP_SBJ	6
3	가장	가장	MAG	AP	4
4	강력한	강력하 L	VA+ETM	VP_MOD	5
5	대안이	대안 이	NNG+JKC	NP_CMP	6
6	될	되 르	VV+ETM	VP_MOD	7
7	것이다.	것 이 다 .	NNB+VCP+EF+SF	VNP	0

(2) 명사류+JX: NP\_CMP

(3) 명사류: NP\_CMP

- 조사가 생략되어 나타나지 않은 경우라도 서술어의 형태 분석 표지가 VCN(부정지정사)이거나 VV(동사) 중 '되다'로 나타나면 기능 표지 CMP(보어)를 가진다.

예) 이것은 항아리에 보관 중인 7년, 10년 된 묵은 된장을 이용한 전통 건강요법이다.

1	이것은	이것은	NP+JX	NP_SBJ	12
2	항아리에	항아리 에	NNG+JKB	NP_AJT	3
3	보관	보관	NNG	NP	4
4	중인	중 이	NNB+VCP+ETM	VNP_MOD	9
5	7년	7 년	SN+NNB+SP	NP	6
6	10년	10 년	SN+NNB	NP_CMP	7
7	된	되	VV+ETM	VP_MOD	9
8	묵은	묵은	VV+ETM	VP_MOD	9
9	된장을	된장 을	NNG+JKO	NP_OBJ	10
10	이용한	이용하	VV+ETM	VP_MOD	12
11	전통	전통	NNG	NP	12
12	건강요법이다.	건강 요법 이 다 .	NNG+NNG+VCP+EF+SF	VNP	0

### 3.1.2.4. 필수적 부사어의 성립

- 구문 표지 NP(체언)와 기능 표지 AJT(부사어)를 가진 어절 중에는 문장을 이루기 위해 필수적으로 요구되는 것들이 있다. <우리말샘>을 참고하여 서술어의 문형 정보에 부사어가 포함되어 있으면 이들은 필수 부사어로 분석한다.
- 필수적 부사어를 가지는 주요 서술어에는 삼다, 같다, 다니다, 놓다, 넣다, 두다, 주다, 던지다, 다가서다 등이 있다.

예) 지엠대우는 이번 새 모델을 역전의 발판으로 삼을 계획이다.

1	지엠대우는	지엠대우 는	NNP+JX	NP_SBJ	7
2	이번	이번	NNG	NP	4
3	새	새	MMA	DP	4
4	모델을	모델 을	NNG+JKO	NP_OBJ	7
5	역전의	역전 의	NNG+JKB	NP_MOD	6
6	발판으로	발판 으로	NNG+JKB	NP_AJT	7
7	삼을	삼 을	VV+ETM	VP_MOD	8
8	계획이다.	계획 이 다 .	NNG+VCP+EF+SF	VNP	0

## 3.2. 수식어-피수식어 관계 분석

### 3.2.1. 관형어의 성립

(1) 관형사: DP

예) 비교적 많이 알려진 이 두 종류의 커피 직업군을 제치고 최근 가장 큰 주목을 받고 있는 게 '커피'다.

1	비교적	비교 적	NNG+XSN	NP_AJT	3
2	많이	많이	MAG	AP	3
3	알려진	알려지	VV+ETM	VP_MOD	8
4	이	이	MMD	DP	6
5	두	두	MMN	DP	6

6	종류의	종류 의	NNG+JKG	NP MOD	8
7	커피	커피	NNG	NP	8
8	직업군을	직업 군 을	NNG+JKO	NP OBJ	9
9	제치고	제치 고	VV+EC	VP AJT	14
10	최근	최근	NNG	NP AJT	14
11	가장	가장	MAG	AP	12
12	크	크	VA+ETM	VP MOD	13
13	주목을	주목 을	NNG+JKO	NP OBJ	14
14	받고	받 고	VV+EC	VP	15
15	있는	있 는	VX+ETM	VP MOD	16
16	게	거 이	NNB+JKS	NP SBJ	17
17	'커피'다.	'커피' 이 다 .	SS+NNG+SS+VCP+EF+SF	VNP	0

## (2) 명사+의: NP\_MOD

예) 작년 가격의 5% 수준인 셈이다.

1	작년	작년	NNG	NP	2
2	가격의	가격 의	NNG+JKG	NP MOD	3
3	5%	5%	SN+SW	NP	4
4	수준인	수준 이	NNG+VCP+ETM	VNP MOD	5
5	셈이다.	셈 이 다 .	NNB+VCP+EF+SF	VNP	0

- 단, 관형격조사 '의'가 나타나지 않은 경우에는 기능표지 MOD를 부여하지 않는다.

## (3) 관형절: VP\_MOD

예) 겉으로는 허름한 시골집처럼 꾸며놓고 내부에는 특급 호텔 시설을 갖춘 곳도 있다고 한다.

1	겉으로는	겉 으 로 는	NNG+JKB+JX	NP AJT	4
2	허름한	허름하	VA+ETM	VP MOD	3
3	시골집처럼	시골집 처럼	NNG+JKB	NP AJT	4
4	꾸며	꾸미 어	VV+EC	VP	5
5	놓고	놓 고	VX+EC	VP AJT	10
6	내부에는	내부 에 는	NNG+JKB+JX	NP AJT	10
7	특급	특급	NNG	NP	8
8	호텔	호텔	NNG	NP	9
9	시설을	시설 을	NNG+JKO	NP OBJ	10
10	갖춘	갖추	VV+ETM	VP MOD	11
11	곳도	곳 도	NNG+JX	NP SBJ	12
12	있다고	있 다고	VA+EC	VP AJT	13
13	한다.	하	VV+EF+SF	VP	0

## 3.2.2. 수의적 부사어의 성립

### (1) 부사: AP

예) 이미 경부고속철도 서울~대구 구간 개통으로 비행기 승객이 확 줄었는데, 다시 운행시간이 22분 더 단축되면서 승객 감소가 불가피할 것으로 전망된 탓이다.

1	이미	이미	MAG	AP	9
2	경부고속철도	경부고속철도	NNP	NP	5
3	서울~대구	서울 ~ 대구	NNP+SO+NNP	NP	4
4	구간	구간	NNG	NP	5
5	개통으로	개통 으 로	NNG+JKB	NP AJT	9
6	비행기	비행기	NNG	NP	7
7	승객이	승객 이	NNG+JKS	NP SBJ	9
8	확	확	MAG	AP	9
9	줄었는데,	줄 었 는데 ,	VV+EP+EC+SP	VP AJT	15
10	다시	다시	MAG	AP	15
11	운행	운행	NNG	NP	12
12	시간이	시간 이	NNG+JKS	NP SBJ	15
13	22분	22 분	SN+NNB	NP SBJ	15



14	더	더	MAG	AP	15
15	단축되면서	단축되 면서	VV+EC	VP AJT	18
16	승객	승객	NNG	NP	17
17	감소가	감소 가	NNG+JKS	NP SBJ	18
18	불가피할	불가피하 르	VA+ETM	VP MOD	19
19	것으로	것 으로	NNB+JKB	NP AJT	20
20	전망된	전망되 니	VV+ETM	VP MOD	21
21	탓이다.	탓 이 다 .	NNG+VCP+EF+SF	VNP	0

(2) 명사: NP\_AJT

예) 이씨는 작년 5월 인천의 한 병원에서 아르바이트하면서 마약류를 훔쳐 투약했다가 적발돼 벌금 1000만원을 선고받은 적이 있다.

1	이씨는	이 씨 는	NNP+NNB+JX	NP SBJ	15
2	작년	작년	NNG	NP	3
3	5월	5 월	SN+NNB	NP AJT	15
4	인천의	인천 의	NNP+JKB	NP MOD	6
5	한	한	MMN	DP	6
6	병원에서	병원 에서	NNG+JKB	NP AJT	7
7	아르바이트하면서	아르바이트하 면서	VV+EC	VP AJT	9
8	마약류를	마약류 를	NNG+JKO	NP OBJ	9
9	훔쳐	훔치 어	VV+EC	VP AJT	10
10	투약했다가	투약하 았 다가	VV+EP+EC	VP AJT	11
11	적발돼	적발되 어	VV+EC	VP AJT	15
12	벌금	벌금	NNG	NP	13
13	1000만	1000 만	SN+NR	NP	14
14	원을	원 을	NNB+JKO	NP OBJ	15
15	선고받은	선고받 은	VV+ETM	VP MOD	16
16	적이	적 이	NNB+JKS	NP SBJ	17
17	있다.	있 다 .	VA+EF+SF	VP	0

(3) 명사+부사격 조사: NP\_AJT

예) 최 부회장은 1심에서 무죄를 선고받았으나, 공소장 변경으로 그의 역할이 부각되면 불리해진다 는 얘기가 나온다.

1	최	최	NNP	NP	2
2	부회장은	부회장 은	NNG+JX	NP SBJ	5
3	1심에서	1심 에서	NNG+JKB	NP AJT	5
4	무죄를	무죄 를	NNG+JKO	NP OBJ	5
5	선고받았으나,	선고받 았 으나 ,	VV+EP+EC+SP	VP AJT	10
6	공소장변경으로	공소장변경 으로	NNG+JKB	NP AJT	10
7	그의	그 의	NP+JKB	NP MOD	8
8	역할이	역할 이	NNG+JKS	NP SBJ	9
9	부각되면	부각되 면	VV+EC	VP AJT	10
10	불리해진다는	불리해지 니 다 는	VV+EF+ETM	VP MOD	11
11	얘기가	얘기 가	NNG+JKS	NP SBJ	12
12	나온다.	나오 니 다 .	VV+EF+SF	VP	0

(4) 명사구: NP\_AJT

예) 환경부 등에 문의한 결과, 이 개구리는 우리나라 서남부 지역에서만 서식하는 희귀종인 금개구 리인 것으로 확인됐다.

1	환경부	환경부	NNG	NP	2
2	등에	등 에	NNB+JKB	NP AJT	3
3	문의한	문의하 니	VV+ETM	VP MOD	4
4	결과,	결과 ,	NNG+SP	NP AJT	14
5	이	이	MMD	DP	6
6	개구리는	개구리 는	NNG+JX	NP SBJ	14
7	우리나라	우리나라	NNG	NP	9
8	서남부	서남부	NNG	NP	9
9	지역에서만	지역 에서 만	NNG+JKB+JX	NP AJT	10
10	서식하는	서식하 는	VV+ETM	VP MOD	11

11	히귀종인	히귀종 이 L	NNG+VCP+ETM	VNP MOD	12
12	금개구리인	금개구리 이 L	NNG+VCP+ETM	VNP MOD	13
13	것으로	것 으로	NNB+JKB	NP AJT	14
14	확인됐다.	확인되 었 다 .	VV+EP+EF+SF	VP	0

(5) 부사절: VP\_AJT

예) 그런데 이날 평가원이 발표한 점수를 보면, 같은 만점자라도 영어 A형을 선택한 학생의 표준점수가 더 높게 나타났다.

1	그런데	그런데	MAJ	AP	16
2	이날	이날	NNG	NP AJT	4
3	평가원이	평가원 이	NNG+JKS	NP SBJ	4
4	발표한	발표하 L	VV+ETM	VP MOD	5
5	점수를	점수 를	NNG+JKO	NP OBJ	6
6	보면,	보 면 ,	VV+EC+SP	VP AJT	16
7	같은	같 은 ,	VA+ETM	VP MOD	8
8	만점자라도	만점자 이 라도	NNG+VCP+EC	NP AJT	16
9	영어	영어	NNP	NP	10
10	A형을	A 형 을	SL+XSN+JKO	NP OBJ	11
11	선택한	선택하 L	VV+ETM	VP MOD	12
12	학생의	학생 의	NNG+JKG	NP MOD	13
13	표준점수가	표준점수 가	NNG+JKS	NP SBJ	16
14	더	더	MAG	AP	15
15	높게	높 게	VA+EC	VP AJT	16
16	나타났다.	나타나 았 다 .	VV+EP+EF+SF	VP	0

→ 부사절이 서술어의 필수 논항이거나 서술어의 수식어일 때에만 VP\_AJT로 분석한다. 부사절이 후행하는 절에 의존할 때에는 VP로 분석한다.

### 3.3. 접속 관계 분석

#### 3.3.1. 명사구 접속

- 명사구와 명사구가 접속되면 선행 명사구가 후행 명사구에 의존하는 것으로 분석한다.
- 기능표지 CNJ(접속어)는 형태분석 표지 JC(접속조사)에 대해서만 부여하고 접속조사가 쓰이지 않은 경우에는 기능표지를 부여하지 않는다.

(1) 명사+접속조사 명사: NP\_CNJ

예) 이 아파트는 대우건설과 서해종합건설이 공동 시공하며 123~337m<sup>2</sup> 규모다.

1	이	이	MMD	DP	2
2	아파트는	아파트 는	NNG+JX	NP SBJ	8
3	대우건설과	대우건설 과	NNP+JC	NP_CNJ	4
4	서해종합건설이	서해종합건설 이	NNP+JKS	NP SBJ	6
5	공동	공동	NNG	NP	6
6	시공하며	시공하 며	VV+EC	VP AJT	8
7	123~337m <sup>2</sup>	123 ~ 337 m <sup>2</sup>	SN+SO+SN+SW	NP	8
8	규모다.	규모 이 다 .	NNG+VCP+EF+SF	VNP	0

(2) 명사, 명사: NP

예) 이달 27일 하이닉스, 한국전력의 종가가 기준 가격이다.

1	이달	이달	NNG	NP	2
---	----	----	-----	----	---

2	27일	27 일	SN+NNB	NP	5
3	하이닉스	하이닉스	NNP+SP	NP	4
4	한국전력의	한국전력 의	NNP+JKG	NP MOD	5
5	총가가	총가 가	NNG+JKS	NP SBJ	6
6	기준가격이다.	기준가격 이 다 .	NNG+VCP+EF+SF	VNP	0

### 3.3.2. 절 접속

- 접속문에서 선행절은 부사어로 처리한다. 의미적으로 대등하게 접속된 것이라고 하더라도 통사적으로는 선행절이 후행절에 의존하는 것으로 분석한다. 선행절은 VP\_AJT로, 후행절은 VP로 분석한다.  
예) 아쉬움은 컸지만 눈물은 나오지 않았습니다.

1	아쉬움은	아쉬움 은	NNG+JX	NP SBJ	2
2	컸지만	크 었 지만	VA+EP+EC	VP AJT	4
3	눈물은	눈물 은	NNG+JX	NP SBJ	4
4	나오지	나오 지	VV+EC	VP	5
5	않았습니다.	않 았 습니다 .	VX+EP+EF+SF	VP	0

## 4. 지배-의존 관계 분석의 실제

### 4.1. 문장 유형별 의존 관계 분석

- 의존 관계 설정을 위한 문장 유형 구분은 일반 언어학 통사론의 기준을 따라 서술어가 한 개인 단문과 서술어가 두 개 이상인 복문으로 문장을 분류하여 분석한다.
- 기본 원칙 (8)에 따라 접속문의 선행절은 후행절에 의존하는 것으로 분석한다.
- 포유문은 명사절 포유문, 관형사절 포유문, 부사절 포유문의 세 가지 유형으로 분류하여 분석한다.

#### 4.1.1. 단문

- 단문에서 관형어를 제외한 요소는 모두 서술어에 의존한다.

예) 멜라닌은 사람의 피부색을 결정한다.

1	멜라닌은	NP SBJ	4	
2	사람의	NP MOD	3	
3	피부색을	NP OBJ	4	
4	결정한다.	VP	0	

## 4.1.2. 접속문

### 4.1.2.1. 일반적인 접속문

- 선행절의 서술어는 후행절의 서술어에 의존하는 것으로 분석한다.

예) 나는 과자를, 동생은 빵을 먹었다.

1	나는	NP_SBJ	3	
2	과자를	NP_OBJ	3	
3	먹고	VP_AJT	6	←←
4	동생은	NP_SBJ	6	
5	빵을	NP_OBJ	6	
6	먹었다.	VP	0	←←

### 4.1.2.2. 선행절에 서술어가 실현되지 않은 경우

- 선행절에 서술어가 실현되지 않은 경우 내포절의 서술어와 관련 있는 성분은 각각 모절의 서술어에 의존하는 것으로 분석한다.

예) 나는 과자를, 동생은 빵을 먹었다.

1	나는	NP_SBJ	5	
2	과자를,	NP_OBJ	5	
3	동생은	NP_SBJ	5	
4	빵을	NP_OBJ	5	
5	먹었다.	VP	0	←←

예) 티셔츠는 2002년 1500만 장, 2006년 1000만 장이나 팔려 나갔다.

1	티셔츠는	NP_SBJ	9	
2	2002년	NP_AJT	8	
3	1500만	NP	4	
4	장,	NP_SBJ	8	←
5	2006년	NP_AJT	8	
6	1000만	NP	7	
7	장이나	NP_SBJ	8	←
8	팔려	VP	9	←←
9	나갔다.	VP	0	←

## 4.1.3. 포유문

### 4.1.2.1. 명사절 포유문

- 모절과 명사절의 주어가 다른 경우, 모절의 주어는 모절의 서술어에, 명사절의 주어는 명사절의 서술어에 의존하는 것으로 분석한다.

예) 우리는 시민의 관전 태도도 그만큼 성숙했음을 잊지 말아야 한다.

1	우리는	NP_SBJ	7	
2	시민의	NP_MOD	4	
3	관전	NP	4	
4	태도도	NP_SBJ	6	←
5	그만큼	AP	6	

6	성숙했음을	VP OBJ	7		
7	일지	VP	8		
8	말아야	VP	9		
9	한다.	VP	0		

- 모절과 명사절에 공통적인 요소가 있는 경우 결합한 조사를 살펴 지배소를 판단하되, 동일지시적인 요소가 문두에 보조사 '은/는'으로 실현된 경우 모절의 성분으로 분석한다.

예) 그는 목표를 이루었음에 매우 감사했다.

1	그는	NP SBJ	5		
2	목표를	NP OBJ	3		
3	이루었음에	VP AJT	5		
4	매우	AP	5		
5	감사했다.	VP	0		

#### 4.1.2.2. 관형사절 포유문

- 모절과 관형사절의 주어가 다른 경우, 모절의 주어는 모절의 서술어에, 관형사절의 주어는 관형사절의 서술어에 의존하는 것으로 분석한다.

예) 내가 좋아하는 꽃은 들국화이다.

1	내가	NP SBJ	2		
2	좋아하는	VP MOD	3		
3	꽃은	NP SBJ	4		
4	들국화이다.	VNP	0		

예) 내가 사진을 좋아하는 사실을 친구들은 다 안다.

1	내가	NP SBJ	3		
2	사진을	NP OBJ	3		
3	좋아하는	VP MOD	4		
4	사실을	NP OBJ	7		
5	친구들은	NP SBJ	7		
6	다	AP	7		
7	안다.	VP	0		

- 모절과 관형사절에 공통적인 요소가 있는 경우 결합한 조사를 살펴 지배소를 판단하되, 동일지시적인 요소가 문두에 보조사 '은/는'으로 실현된 경우 모절의 성분으로 분석한다.

예) 세포벽은 식물 세포의 가장 바깥층을 에워싸고 있는 약간 두꺼운 막이다.

1	세포벽은	NP SBJ	9		
2	식물세포의	NP MOD	4		
3	가장	AP	4		
4	바깥층을	NP OBJ	5		
5	에워싸고	VP	6		
6	있는	VP MOD	9		
7	약간	AP	8		
8	두꺼운	VP MOD	9		
9	막이다.	VNP	0		

예) 멜라닌은 사람의 피부색을 결정하는 주요 요소이다.

1	멜라닌은	NP SBJ	6		
2	사람의	NP MOD	3		
3	피부색을	NP OBJ	4		
4	결정하는	VP MOD	6		
5	주요	NP	6		
6	요소이다.	VNP	0		

예) 이 사료는 가축들이 먹는 음식이다.

1	이	DP	2		
2	사료는	NP_SBJ	5		
3	가축들이	NP_SBJ	4		
4	먹는	VP_MOD	5		
5	음식이다.	VNP	0		

#### 4.1.2.3. 부사절 포유문

- 모절과 부사절의 주어가 다른 경우, 모절의 주어는 모절의 서술어에, 부사절의 주어는 부사절의 서술어에 의존하는 것으로 분석한다.
- 부사절이 서술어가 필수적으로 요구하는 성분이거나 서술어를 수식하는 부가어인 경우에는 VP\_AJT로 분석한다.
- 인용절도 부사절로 분석한다.

예) 그가 돌아왔다고 그녀가 알려 줬어.

1	그가	NP_SBJ	2		
2	돌아왔다고	VP_AJT	4		
3	그녀가	NP_SBJ	4		
4	알려	VP	5		
5	줬어.	VP	0		

- 부사절이 후행하는 절과 접속된 경우에도 VP\_AJT로 분석한다.

예) 그가 돌아와 줘서 우리는 정말 고마웠어.

1	그가	NP_SBJ	2		
2	돌아와	VP	3		
3	줘서	VP_AJT	6		
4	우리는	NP_SBJ	6		
5	정말	AP	6		
6	고마웠어.	VP	0		

- 모절과 부사절에 공통적인 요소가 있는 경우 결합한 조사를 살펴 지배소를 판단하되, 동일지시적인 요소가 문두에 보조사 '은/는'으로 실현된 경우 모절의 성분으로 분석한다.

예) 멜라닌은 자외선을 차단해서 자외선으로부터 피부를 보호한다.

1	멜라닌은	NP_SBJ	6		
2	자외선을	NP_OBJ	3		
3	차단해서	VP_AJT	6		
4	자외선으로부터	NP_AJT	6		
5	피부를	NP_OBJ	6		
6	보호한다.	VP	0		

예) 위원장은 "아직 준비가 미흡하다"며 "빠른 시일 안에 준비를 마칠겠다"고 강조했다.

1	위원장은	NP_SBJ	10		
2	"아직	AP	4		
3	준비가	NP_SBJ	4		
4	미흡하다"며	VP_AJT	10		
5	빠른	VP_MOD	6		
6	시일	NP	7		
7	안에	NP_AJT	9		
8	준비를	NP_OBJ	9		
9	마칠겠다"고	VP_AJT	10		

10	강조했다.	VP	0	
----	-------	----	---	--

- 주제어가 모절과 내포절의 논항으로 해석될 수 있으나 문장성분이 서로 다르게 파악되는 경우에는 후행절의 문장성분을 기준으로 해당 기능표지를 부여함.

예) '서치미'는 매주 월, 화 밤 12시에 U+tv를 통해 공개되며, KBS2에서는 매주 목요일 밤 9시 45분에 만나 볼 수 있다.

1	'서치미'는	NP OBJ	16	
2	매주	AP	4	
3	월,	NP	4	
4	화	NP	5	
5	밤	NP	6	
6	12시에	NP AJT	9	
7	U+tv를	NP OBJ	8	
8	통해	VP AJT	9	
9	공개되며,	VP AJT	16	
10	KBS2에서는	NP AJT	16	
11	매주	AP	12	
12	목요일	NP	13	
13	밤	NP	15	
14	9시	NP	15	
15	45분에	NP AJT	16	
16	만나	VP	17	
17	볼	VP MOD	18	
18	수	NP	19	
19	있다.	VP	0	

- 신문 기사 형식에서 주제어에 이어지는 시간, 장소 표현은 맥락을 파악하여 의존 관계를 분석하고, 맥락을 살펴도 의존 관계 파악이 어려운 경우에는 후행절 서술어에 의존하도록 분석한다.

예) 키움은 24일 서울 고척스카이돔에서 '2025 신한은행 SOL Bank KBO리그' KIA와의 경기를 9-6으로 승리하며 주중 3연전 첫 경기를 승리로 장식했다.

1	키움은	NP SBJ	19	
2	24일	NP AJT	19	
3	서울	NP	4	
4	고척스카이돔에서	NP AJT	19	
5	'2025	NP	6	
6	신한은행	NP	7	
7	SOL	NP	8	
8	Bank	NP	9	
9	KBO리그'	NP	11	
10	KIA와의	NP MOD	11	
11	경기를	NP OBJ	13	
12	9-6으로	NP AJT	13	
13	승리하며	VP AJT	19	
14	주중	NP	15	
15	3연전	NP	17	
16	첫	DP	17	
17	경기를	NP OBJ	19	
18	승리로	NP AJT	19	
19	장식했다.	VP	0	

- 인용절은 부사절로 분석한다. 인용절은 인용동사나 사유동사가 필수적으로 요구하는 성분이므로 VP\_AJT로 분석한다.

예) 그녀는 자기가 그 일을 했다고 말했다.

1	그녀는	NP SBJ	6	
2	자기가	NP SBJ	5	
3	그	DP	4	
4	일을	NP OBJ	5	
5	했다고	VP AJT	6	
6	말했다.	VP	0	

예) 칼라일은 “셰익스피어를 인도와도 바꾸지 않겠다.”라고 말했다.

1	칼라일은	NP SBJ	6	
2	“셰익스피어를	NP OBJ	4	
3	인도와도	NP AJT	4	
4	바꾸지	VP	5	
5	않겠다.”라고	VP AJT	6	
6	말했다.	VP	0	

예) 그녀는 “철수가 나를 믿기 때문이다”고 말했다.

1	그녀는	NP SBJ	6	
2	“철수가	NP SBJ	4	
3	나를	NP OBJ	4	
4	믿기	VP	5	
5	때문이다”고	VNP AJT	6	
6	말했다.	VP	0	

- ‘~’라며, ~’라고’로 연결된 인용절 포함 문장에서 ‘~’라며, ~’라고’ 등은 모두 후행절 서술어에 의존하도록 분석한다. ‘~’라며’를 후행하는 ‘~’라고’ 어절에 의존하도록 분석하지 않음에 주의한다.

예) 그는 “민주당은 이 대표가 구속될까 노심초사하며 지난해 8월 16일 이후 하루도 쉬지 않고 국회를 열고 있다”라며 “1월 임시회가 다음달 7일까지 열린다면 장장 176일간 국회가 열리는 것이다. 민주당은 이 대표에 대한 체포 동의서가 국회에 보내지면 즉각 부결시킬 태세”라고 지적했다.

1	그는	NP SBJ	38	
2	“민주당은	NP SBJ	6	
3	이	NP	4	
4	대표가	NP SBJ	5	
5	구속될까	VP	6	
6	노심초사하며	VP AJT	15	
7	지난해	NP AJT	9	
8	8월	NP	9	
9	16일	NP	10	
10	이후	NP AJT	15	
11	하루도	NP OBJ	12	
12	쉬지	VP	13	
13	않고	VP AJT	15	
14	국회를	NP OBJ	15	
15	열고	VP	16	
16	있다”라며	VP AJT	38	
17	“1월	NP	18	
18	임시회가	NP SBJ	21	
19	다음달	NP	20	
20	7일까지	NP AJT	21	
21	열린다면	VP AJT	25	
22	장장	DP	23	
23	176일간	NP AJT	24	
24	국회가	NP SBJ	25	
25	열리는	VP	26	
26	것이다.	VNP	37	
27	민주당은	NP SBJ	36	
28	이	NP	29	
29	대표에	NP AJT	30	
30	대한	VP MOD	32	
31	체포	NP	32	
32	동의서가	NP SBJ	34	



33	국회에	NP AJT	34	
34	보내지면	VP AJT	36	
35	조각	AP	36	
36	부결시킬	VP MOD	37	
37	태세"라고	NP AJT	38	
38	지적했다.	VP	0	

- 인용절 안의 따옴표 안에 여러 문장이 나열된 경우에는, 1~n-1번째 문장들은 모두 n번째 문장에 의존하도록 분석한다.

예) 이 매체는 “토틀넘은 새 출발이 필요하다. 지금 상황에서 손흥민을 팔아 현금화해 나폴리 스타일의 스쿼드를 꾸리는 것보다 좋은 선택은 없다”고 강조했다.

1	이	DP	2	
2	매체는	NP SBJ	20	
3	“토틀넘은	NP SBJ	6	
4	새	DP	5	
5	출발이	NP SBJ	6	
6	필요하다.	VP	19	
7	지금	NP AJT	8	
8	상황에서	NP AJT	19	
9	손흥민을	NP OBJ	10	
10	팔아	VP AJT	11	
11	현금화해	VP AJT	15	
12	나폴리	NP	13	
13	스타일의	NP MOD	14	
14	스쿼드를	NP OBJ	15	
15	꾸리는	VP MOD	16	
16	것보다	NP AJT	17	
17	좋은	VP MOD	18	
18	선택은	NP SBJ	19	
19	없다”고	VP AJT	20	
20	강조했다.	VP	0	

- ‘-다’며’는 ‘-다고 말하며’가 줄어든 것으로 보고 일반적인 부사절과 동일하게 분석한다.

예) 서 판사는 “도망칠 우려가 있다”며 구속영장을 발부했다.

1	서	NP	2	
2	판사는	NP SBJ	7	
3	“도망칠	VP MOD	4	
4	우려가	NP SBJ	5	
5	있다”며	VP AJT	7	
6	구속영장을	NP OBJ	7	
7	발부했다.	VP	0	

## 4.2. 구 구성별 의존 관계 분석

### 4.2.1. 명사구

#### 4.2.1.1. 관형어+명사+명사

- 관형어가 명사구를 수식할 때, 명사구 중 관형어의 지배소는 관형어가 의미적으로 수식하는 어휘를 지배소로 분석한다.

- 여러 개의 명사들이 특별한 수식 관계 없이 나열되어 있는 경우에는, 핵어 명사를 제외한 수식 명사 들을 모두 기능 표지 없이 NP로 처리하고, 각각 바로 다음 어절(오른쪽 명사)에 의존하는 것으로 분석한다.

예) 무분별한 포획 문제

1	무분별한	VP MOD	2	
2	포획	NP	3	
3	문제	NP	0	

예) 고려의 충신 정몽주

1	고려의	NP MOD	2	
2	충신	NP	3	
3	정몽주	NP	0	

예) 조선의 제3대 임금

1	조선의	NP MOD	3	
2	제3대	NP	3	
3	임금	NP	0	

예) 노엄 촘스키의 저서

1	노엄	NP	2	
2	촘스키의	NP MOD	3	
3	저서	NP	0	

예) 결명자의 한자 뜻

1	결명자의	NP MOD	3	
2	한자	NP	3	
3	뜻	NP	0	

예) 역사상 가장 위대한 천재 중 하나

1	역사상	NP	3	
2	가장	AP	3	
3	위대한	VP MOD	4	
4	천재	NP	5	
5	중	NP	6	
6	하나	NP	0	

예) 프랑스 파리 루브르 박물관

1	프랑스	NP	2	
2	파리	NP	3	
3	루브르	NP	4	
4	박물관	NP	0	

- 명사구가 여러 어절로 구성되어 지배소 후보가 여러 개인 경우에는 보다 인접한 어절을 지배소로 삼는다.

예) 부산시, 수도권 기업 대상 최대 규모 유치 설명회 개최

1	부산시,	NP	9	
2	수도권	NP	3	
3	기업	NP	4	
4	대상	NP	8	
5	최대	NP	6	
6	규모	NP	8	
7	유치	NP	8	
8	설명회	NP	9	
9	개최	NP	0	

→ '대상'의 경우 '설명회' 및 '개최'와 모두 연결 가능한 경우인데, 더 가까운 요소에 의존하는 것으로

로 처리한다.

예) 서울에 살고 있는 시인 김철수

1	서울에	NP AJT	2		
2	살고	VP	3		
3	있는	VP MOD	4		
4	시인	NP	5		
5	김철수	NP	0		

→ 위 예문은 '[서울에 살고 있는 시인] [김철수]'로도 분석 가능하고 '[서울에 살고 있는] [시인 김철수]'로도 분석 가능하다. 이 경우 '있는'을 더 가까운 요소인 '시인'과 연결한다.

예) 한국 등반대가 히말라야 정복

1	한국	NP	2		
2	등반대가	NP SBJ	4		
3	히말라야	NP OBJ	4		
4	정복	NP	0		

→ '히말라야'에는 조사가 없지만, '등반대가'에 조사가 실현되어 있으므로 명사절로 본다. 따라서 '히말라야'는 절 내부의 어절 중 조사가 생략된 경우로 보고 기능 태그를 부여한다.

- 'A의 [B와 C]'와 같이 관형어의 수식을 받는 대상이 명사구 접속 구성인 경우에는 관형어 성분 'A의'는 접속된 명사구의 핵인 C에 의존하도록 분석한다.

예) 또 금융·비즈니스 세션 발표자로 무대에 서는 오세현 SK텔레콤 부사장은 국내 웹3.0 및 블록체인의 서비스의 현재와 비전을 참가자와 공유할 예정이다.

1	또	AP	18		
2	금융·비즈니스	NP	3		
3	세션	NP	4		
4	발표자로	NP AJT	6		
5	무대에	NP AJT	6		
6	서는	VP MOD	9		
7	오세현	NP	9		
8	SK텔레콤	NP	9		
9	부사장은	NP SBJ	18		
10	국내	NP	11		
11	웹3.0	NP	14		
12	및	AP	14		
13	블록체인	NP	14		
14	서비스의	NP MOD	16		
15	현재와	NP CNJ	16		
16	비전을	NP OBJ	18		
17	참가자와	NP AJT	18		
18	공유할	VP MOD	19		
19	예정이다.	VNP	0		

#### 4.2.1.2. 수사 및 수 관형사를 포함한 명사구

- '한, 두, 세, 한두, 서너, 스무' 등 수 관형사는 관형사구(DP)로 처리하고 후행하는 단위성 명사에 연결한다.

예) 한 명

1	한	DP	2		
---	---	----	---	--	--

2	명	NP	0	◀
---	---	----	---	---

예) 서너 개

1	서너	DP	2	◀
2	개	NP	0	◀

- ‘일, 이, 삼, 십, 백, 천, 만’ 등 수사 및 관형사의 통용어로 등재되어 있는 경우, 후행하는 단위성 명사가 있으면 관형사구(DP)로, 없으면 명사구(NP)로 처리한다.

예) 이 년의 기다림

1	이	DP	2	◀
2	년	NP MOD	3	◀
3	기다림	NP	0	◀

예) 사과 하나 가격

1	사과	NP	2	◀
2	하나	NP	3	◀
3	가격	NP	0	◀

예) 아들 둘과 딸 셋

1	아들	NP	2	◀
2	둘과	NP CNJ	4	◀
3	딸	NP	4	◀
4	셋	NP	0	◀

예) 수 미터

1	수	DP	2	◀
2	미터	NP	0	◀

예) 몇 살

1	몇	DP	2	◀
2	살	NP	0	◀

#### 4.2.1.3. 명사구 접속 구성

- 복수 개의 명사구가 접속 또는 나열된 경우, 가장 마지막 명사구에 의존하도록 분석한다.
- 형태분석 표지 JC(접속조사)에 대해서만 기능표지 CNJ(접속어)를 부여한다.
- 접속조사로 인정하는 것은 〈우리말샘〉 기준으로 12개이다. (고031/이고005, 과021/와011, 나 021/이나010, 니006/이니002, 다010/이다009, 량007/이량008, 며001/이며001, 면015/이면005, 예027, 이략002, 하고008, 하며001)
- 이들 조사가 아닌 경우(‘(이)라든지’, ‘(이)라든가’ 등)는 접속으로 보지 않는다.
- ‘및’, ‘또는’, ‘그리고’ 등에 의해 명사구가 접속 또는 나열된 경우, 이들 접속부사는 후행하는 명사구에 의존하는 것으로 분석한다. 그리고 이때 구문 표지만 부착하고, 기능 표지는 부착하지 않는다.

예) 비단 피부뿐만이 아니라 털, 눈, 귀, 심지어 뇌에도 존재한다.

1	비단	AP	3	
2	피부뿐만이	NP CMP	3	
3	아니라	VP AJT	9	
4	털,	NP	8	
5	눈,	NP	8	
6	귀,	NP	8	
7	심지어	AP	8	
8	뇌에도	NP AJT	9	
9	존재한다.	VP	0	

예) 매리너스 협곡과 극관이 존재한다.

1	매리너스	NP	2	
2	협곡과	NP CNJ	3	
3	극관이	NP SBJ	4	
4	존재한다.	VP	0	

예) 주로 왕과 왕세자의 강론 및 정책토론을 주관했다.

1	주로	AP	7	
2	왕과	NP CNJ	3	
3	왕세자의	NP MOD	6	
4	강론	NP	6	
5	및	AP	6	
6	정책토론을	NP OBJ	7	
7	주관했다.	VP	0	

- 조사 '와/과'로 나타나는 명사구는 'N과 N' 순서인 경우에 명사구 접속으로, 'N이 N과'와 같은 순서인 경우에는 부사어로 처리한다.

예) 철수와 영희가 만났다.

1	철수와	NP CNJ	2	
2	영희가	NP SBJ	3	
3	만났다.	VP	0	

예) 철수가 영희와 헤어졌다.

1	철수가	NP SBJ	3	
2	영희와	NP AJT	3	
3	헤어졌다.	VP	0	

- 조사 '(이)라든지', '(이)라든지'는 나열 기능을 하는 조사로, '명사구+(이)라든지/라든지 명사구+(이)라든지/라든지 하는' 등의 구성으로 자주 나타난다. 이때의 '명사구+(이)라든지/라든지' 성분은 각각 NP\_AJT로 서술어에 연결한다.

예)) 그는 돈이라든지 명예라든지 하는 것에 연연해하지 않는다.

1	그는	NP SBJ	6	
2	돈이라든지	NP AJT	4	
3	명예라든지	NP AJT	4	
4	하는	VP MOD	5	
5	것에	NP AJT	6	
6	연연해하지	VP	7	
7	않는다.	VP	0	

## 4.2.2. 용언구

### 4.2.2.1. 본용언 + 본용언 구성

- 본용언이 연속적으로 나타날 경우, 두 용언과 공통적으로 관계있는 문장 성분들은 앞에 위치하는 서술어에 의존하는 것으로 분석한다.

예) 해구보다는 폭이 넓고 얇다.

1	해구보다는	NP AJT	3	
2	폭이	NP SBJ	3	
3	넓고	VP AJT	4	
4	얇다.	VP	0	

### 4.2.2.2. 본용언 + 보조 용언 구성

- 본용언과 보조 용언이 연속하여 나올 때에는 본용언의 논항들이 본용언에 의존하는 것으로 분석하고, 본용언은 보조 용언에 의존하는 것으로 분석한다.
- 보조 용언 구성이 두 개 이상 연속될 때에는 '본용언 → 보조 용언1 → 보조 용언2'의 순으로 의존하는 것으로 분석한다.

예) 멜라닌은 물에는 용해되지 않는다.

1	멜라닌은	NP SBJ	3	
2	물에는	NP AJT	3	
3	용해되지	VP	4	
4	않는다.	VP	0	

- 절과 절이 연결될 때에는 선행절이 후행절의 용언에 연결된다. 후행절이 본용언+보조 용언 구성을 이루는 경우 선행절이 후행절의 본용언에 의존하는 것으로 분석한다.

예) 갈등과 속내가 공개돼 화제를 모으고 있다.

1	갈등과	NP CNJ	2	
2	속내가	NP SBJ	3	
3	공개돼	VP AJT	5	
4	화제를	NP OBJ	5	
5	모으고	NP	6	
6	있다.	VP	0	

- 접속부사는 보조 용언이 아닌 본용언에 연결한다. 본용언+보조용언 구성과 유사한 기능을 하는 의존명사 구성의 경우에도 동일하게 적용한다.

예) 그러나 이번 분기는 생각보다 순조롭게 진행되고 있었다.

1	그러나	AP	6	
2	이번	NP	3	
3	분기는	NP SBJ	6	
4	생각보다	NP AJT	5	
5	순조롭게	VP AJT	6	
6	진행되고	VP	7	
7	있었다.	VP	0	

#### 4.2.2.3. 의존명사 구성

- 주 서술어 다음에 보조 용언은 아니지만 서법을 나타내는 의존명사가 포함된 구성이 오는 경우, 해당 서술어와 의존명사 구성을 별개의 단위로 처리하여 분석하고 의존 관계를 연결한다.
- 한국전자통신기술협회TTA에서 제시하는 의존명사 구성의 목록: ‘-ㄴ 수/리(가) 있다/없다’, ‘-은/을 것/터/뿐/따름/모양/지경/참/중/노릇/예정/길이다’, ‘-을 만/법/듯하다’, ‘-는 말이다’, ‘-은/을 듯(도)하다’, ‘-은/을 것 같다’, ‘-을 것(을/걸) 그랬다’, ‘-고 해서’, ‘-거나/-든지/-고/-다가... 하다/그러다[아우름 기능]’ 등.
- 의존명사에 ‘이다’, ‘하다’ 등이 결합해 있는 경우에는 그 자체를 각각 VNP, VP 등으로 처리한다.

예) 그는 일어날 수 없었다.

1	그는	NP SBJ	2	
2	일어날	VP MOD	3	
3	수	NP SBJ	4	
4	없었다.	VP	0	

예) 곧 비가 올 것 같다.

1	곧	AP	3	
2	비가	NP SBJ	3	
3	올	VP MOD	4	
4	것	NP	5	
5	같다.	VP	0	

예) 나는 곧 밥을 먹을 것이다.

1	나는	NP SBJ	4	
2	곧	AP	4	
3	밥을	NP OBJ	4	
4	먹을	VP MOD	5	
5	것이다.	VNP	0	

- 문장의 주어와 의존명사의 관계는 해당 의존명사가 내용 명사로 대체 가능한 경우와 불가능한 경우로 구분하여 분석한다. 즉, 의존명사가 주어와 공지시(co-reference)되는 경우에는 주어가 의존명사 어절에 의존하는 것으로 분석한다. 예컨대, ‘-ㄴ 것이다’에서 ‘것’을 다른 명사로 대체할 수 있는 경우 의존명사 어절에 의존하는 것으로 분석한다.

예) 이 사료는 가축들이 먹는 것이다.

1	이	DP	2	
2	사료는	NP SBJ	5	
3	가축들이	NP SBJ	4	
4	먹는	VP MOD	5	
5	것이다.	VNP	0	

예) 내가 놀란 것은 철수가 천재라는 것이다.

1	내가	NP SBJ	2		
2	놀란	VP MOD	3		
3	것은	NP SBJ	6		
4	철수가	NP SBJ	5		
5	천재라는	VNP MOD	6		
6	것이다.	VNP	0		

#### 4.2.2.4. 'NP 중이다' 구성

- 'NP 중이다' 구성에서 NP의 논항이 되는 성분은 '중이다'가 아닌 해당 NP에 의존한다.

예) 그들은 다른 방안을 검토 중이다.

1	그들은	NP SBJ	4		
2	다른	DP	3		
3	방안을	NP OBJ	4		
4	검토	NP	5		
5	중이다.	VNP	0		

#### 4.2.3. 부호 포함 어절

- 괄호와 괄호 안의 어절이 '부호의 띄어쓰기 원칙'에 따라 선행 어절과 분절되지 않고 한 어절을 이루는 경우 괄호 밖의 성분과 괄호 안의 성분을 각각 해당 구성의 핵에 의존하도록 아래 예와 같이 분석한다.

예) 올해 UDC는 온라인과 오프라인의 장점을 모두 수용한 하이브리드 형태(온·오프라인 동시 운영)로 진행된다.

1	올해	NP	12		
2	UDC는	NP SBJ	12		
3	온라인과	NP CNJ	4		
4	오프라인의	NP MOD	5		
5	장점을	NP OBJ	7		
6	모두	AP	7		
7	수용한	VP MOD	9		
8	하이브리드	NP	9		
9	형태(온·오프라인	NP	11		
10	동시	NP	11		
11	운영)로	NP AJT	12		
12	진행된다.	VP	0		

예) 과일(사과, 배 등)의 등급

1	과일(사과,	NP	2		
2	배	NP	3		
3	등)의	NP MOD	4		
4	등급	NP	0		

예) 철수의 버릇: 다리 꼬기, 이갈기

1	철수의	NP MOD	2		
2	버릇:	NP	5		
3	다리	NP	4		
4	꼬기,	NP	5		
5	이갈기	NP	0		



- 단 하나 이상의 서술어가 가운뎃점(·)으로 나열되며 어근과 접사가 분리된 경우에는 형태 분석 결과에 따라 구문 표지와 기능 표지를 부여한다.

예) 조작된 뉴스를 생산·유통시키는 것이 언론의\_자유라고 외치는 세력

1	조작된	VP MOD	2		
2	뉴스를	NP OBJ	3		
3	생산·유통시키는	NP MOD	4		
4	것이	NP SBJ	5		
5	언론의 자유라고	VNP AJT	6		
6	외치는	VP MOD	7		
7	세력	NP	0		

- 삽입구가 복수의 어절로 구성되어 있을 경우, 기호로 결합된 복합 형태소는 선행 성분을 기준으로 구문 태그를 결정하고, 후행 성분을 기준으로 기능 태그를 결정한다.

예) 전문위원의 임기는 3년을 보장한다(1차에 한하여 연임 가능).

1	전문위원의	NP MOD	2		
2	임기는	NP SBJ	4		
3	3년을	NP OBJ	4		
4	보장한다(1차에	VP AJT	5		
5	한하여	VP	7		
6	연임	NP SBJ	7		
7	가능).	NP	0		

#### 4.2.4. 외국 문자/외국어 포함 어절

- 외국 문자, 숫자를 비롯한 기능을 알 수 없는 미등재어의 구문 태그는 NP로 분석한다.

예) “닥쳐(Shut up)!”

1	“닥쳐(Shut	VP	2		
2	up)!”	NP	0		

예) 아이 러브 유

1	아이	NP	2		
2	러브	NP	3		
3	유	NP	0		

예) I love you

1	I	NP	2		
2	love	NP	3		
3	you	NP	0		

## 4.3. 세부 유형별 분석 방법

### 4.3.1. 보조사적 용법의 '이/가', '을/를'

- 본용언에 화용적 기능을 가지는 조사 {이/가, 을/를}이 붙은 경우는 기능 표지를 부착하지 않는다. 즉, 명사형(-음, -기) 및 일부 연결어미(-ㄴ가, -ㄴ지, -ㄹ까, -ㄹ지 등)를 제외한 용언의 활용형에 붙은 조사 {이/가, 을/를}은 주어나 목적으로 분석하지 않는다.

예) 철수가 밥을 이틀내 먹지를 않았다.

1	철수가	NP SBJ	4	
2	밥을	NP OBJ	4	
3	이틀내	NP AJT	4	
4	먹지 <sup>를</sup>	VP	5	
5	않았다.	VP	0	

예) 그 산은 그리 높지가 않다.

1	그	DP	2	
2	산은	NP SBJ	4	
3	그리	AP	4	
4	높지 <sup>가</sup>	VP	5	
5	않다.	VP	0	

- '-기 바라다', '-기 시작하다' 등의 '-기' 절을 요구하는 서술어의 경우 뒤에 격 조사가 붙는 경우와 붙지 않는 경우 모두 기능 태그를 부착하여 VP\_OBJ로 태깅한다.

예) 나는 네가 빨리 오기를 바란다.

1	나는	NP SBJ	5	
2	네가	NP SBJ	4	
3	빨리	AP	4	
4	오기 <sup>를</sup>	VP OBJ	5	
5	바란다.	VP	0	

- '-ㄴ가, -ㄴ지, -ㄹ까, -ㄹ지' 등이 결합된 용언 활용형이 논항으로 쓰이는 경우는 해당 기능 태그를 부착한다.

예) 철수는 뭐가 더 좋은가 생각했다.

1	철수는	NP SBJ	5	
2	뭐가	NP SBJ	4	
3	더	AP	4	
4	좋은가	VP OBJ	5	
5	생각했다.	VP	0	

예) 철수가 공부를 잘 하는지 몰랐다.

1	철수가	NP SBJ	4	
2	공부를	NP OBJ	4	
3	잘	AP	4	
4	하는지	VP OBJ	5	
5	몰랐다.	VP	0	

예) 조코비치가 출전할지 전 세계의 이목이 쏠리고 있다.

1	조코비치가	NP SBJ	2		
2	출전할지	VP AJT	6		
3	전	DP	4		
4	세계의	NP MOD	5		
5	이목이	NP SBJ	6		
6	쏠리고	VP	7		
7	있다.	VP	0		

#### 4.3.2. 부사성 명사구

- ‘~즈음’, ‘~쯤’, ‘~인 만큼’, ‘~ 결과’, ‘~하는 대신’ 등 부사성 명사나 명사구가 조사 없이 쓰여 문장 안에서 부사어로 기능하는 경우 기능표지 AJT를 부착한다.

예) 광복절 즈음 독립기념관을 찾았다.

1	광복절	NP	2		
2	즈음	NP AJT	4		
3	독립기념관을	NP OBJ	4		
4	찾았다.	VP	0		

예) 문화재 정책을 총괄하는 만큼 문화재가 유산으로 명칭이 바뀌면 이를 따라야 한다.

1	문화재	NP	2		
2	정책을	NP OBJ	3		
3	총괄하는	NP MOD	4		
4	만큼	NP AJT	10		
5	문화재가	NP SBJ	8		
6	유산으로	NP AJT	8		
7	명칭이	NP SBJ	8		
8	바뀌면	VP AJT	10		
9	이를	NP OBJ	10		
10	따라야	VP	11		
11	한다.	VP	0		

- 그러나 ‘~즈음’, ‘~쯤’ 뒤에 다른 격 조사가 쓰이는 경우 격 조사를 고려하여 주석한다.

예) 다섯 명 즈음의 학생이 길에 서 있었다.

1	다섯	DP	2		
2	명	NP	3		
3	즈음의	NP MOD	4		
4	학생이	NP SBJ	6		
5	길에	NP AJT	6		
6	서	VP	7		
7	있었다.	VP	0		

#### 4.3.3. 격 조사가 결합한 수량 표현

- 주격조사/목적격조사와 용언 사이에 시간이나 거리 등 수량이나 단위를 나타내는 명사구가 쓰인 경우에는 주어나 목적어가 아닌 부사어로 분석한다.

예) 기억에 남는 게 하나 있는데

1	기억에	NP AJT	2		
2	남는	VP MOD	3		
3	게	NP SBJ	5		
4	하나	NP AJT	5		
5	있는데	VP AJT	0		

예) 잘 키워 가지고 이파리를 하나 만들어 냈어.

1	잘	AP	2		
2	키워	VP	3		
3	가지고	VP AJT	6		
4	이파리를	NP OBJ	6		
5	하나	NP AJT	6		
6	만들어	VP	7		
7	냈어.	VP	0		

- 용언 앞에 주격조사/목적격조사가 결합하지 않은 명사구와 시간이나 거리 등 수량이나 단위를 나타내는 명사구가 연속되어 쓰인 경우에는 그 전체가 주어나 목적으로 분석될 수 있어 주의가 필요하다.

예) 고등학교 1학년 때 읽은 책 하나가 기억이 나는데

1	고등학교	NP	2		
2	1학년	NP	3		
3	때	NP AJT	4		
4	읽은	VP MOD	5		
5	책	NP	6		
6	하나가	NP SBJ	8		
7	기억이	NP SBJ	8		
8	나는데	VP AJT	0		

예) 젊은 애들은 그냥 폰 하나 가지고 별별의 걸 다 하는데

1	젊은	VP MOD	2		
2	애들은	NP SBJ	10		
3	그냥	AP	10		
4	폰	NP	5		
5	하나	NP OBJ	6		
6	가지고	VP AJT	10		
7	별별의	NP MOD	8		
8	걸	NP OBJ	10		
9	다	AP	10		
10	하는데	VP AJT	0		

- 용언 앞에 명사구가 나오지 않고 수사만 쓰인 경우에는 문장 내에서의 기능에 따라 기능 표지를 부여한다.

예) 하나 쓰면 하나 남는단 소린 거 같아요.

1	하나	NP OBJ	2		
2	쓰면	VP AJT	4		
3	하나	NP SBJ	4		
4	남는단	VP MOD	5		
5	소린	VNP_MOD	6		
6	거	NP	7		
7	같아요.	VP	0		

#### 4.3.4. 품사와 문장 성분이 일치하지 않는 경우

- 품사로는 부사나 활용형 등이 쓰였지만 문장 내에서 인용이 된 것처럼 쓰인 경우에는 해당 문장 성분을 기준으로 기능 태그를 부여하되, 구문 태그 분석은 단어의 본래 품사를 기준으로 한다. 다만, 형태소의 일부가 잘린 경우 미등재어로 보아 NP를 부여한다.

예) '거꾸로'는 '거꾸'와 '로'로 분석할 수 있을까?

1	'거꾸로'는	AP OBJ	4	
2	'거꾸'와	NP CNJ	3	
3	'로'로	NP AJT	4	
4	분석할	VP MOD	5	
5	조	NP SBJ	6	
6	있을까?	VP	0	

#### 4.3.5. 서술어적 용법의 '~으로' 명사구

- '~으로'가 의미적으로 '~인데'와 같이 사용되었다면, 예외적으로 서술어로 인정한다. 즉, 주어 논항을 가질 수 있다.

예) 그녀는 귀화한 한국인으로 그녀의 아버지가 대사관에서 일한다.

1	그녀는	NP SBJ	3	
2	귀화한	VP MOD	3	
3	한국인으로	NP AJT	7	
4	그녀의	NP MOD	5	
5	아버지가	NP SBJ	7	
6	대사관에서	NP AJT	7	
7	일한다.	VP	0	

#### 4.3.6. 부사 '없이', '같이', '달리', '함께'

- '없이', '같이', '달리', '함께'와 같은 부사는 서술어와 마찬가지로 논항을 취할 수 있다. '없이' 앞에 오는 '이' 명사구는 '없이'의 주어로, '같이, 달리, 함께' 앞에 오는 '와' 명사구는 이들 부사의 부사어로 분석한다.

예) 철수는 아무 생각도 없이 길을 나섰다.

1	철수는	NP SBJ	6	
2	아무	DP	3	
3	생각도	NP SBJ	4	
4	없이	AP	6	
5	길을	NP OBJ	6	
6	나섰다.	VP	0	

예) 예상한 바와 같이 주가가 크게 떨어졌다.

1	예상한	VP MOD	2	
2	바와	NP AJT	3	
3	같이	AP	6	

4	추가	가	NP SBJ	6	
5	크게		VP AJT	6	
6	떨어졌다.		VP	0	

#### 4.3.7. 연결된 부사구 ‘~부터 ~까지’, ‘~에서 ~으로’

- ‘~부터’, ‘~까지’ (또는 ‘~에서’, ‘~로’)와 같은 부사구는 각각을 지배소에 연결한다.

예) 그는 작년부터 지금까지 열심히 일했다.

1	그는		NP SBJ	5	
2	작년	부터	NP AJT	5	
3	지금	까지	NP AJT	5	
4	열심히		AP	5	
5	일했다.		VP	0	

예) 부산에서 서울로 가는 표 한 장 있나요?

1	부산	에서	NP AJT	3	
2	서울	로	NP AJT	3	
3	가는		VP MOD	4	
4	표		NP	6	
5	한		DP	6	
6	장		NP SBJ	7	
7	있나요?		VP	0	

- 그러나 만일 ‘~부터 ~까지를’, ‘~에서 ~로의’ 등과 같이 ‘~부터’, ‘~까지’(또는 ‘~에서’, ‘~로’) 부사구가 하나의 단위로 묶이는 경우 선행 성분을 후행 성분의 부사어(AJT)로 연결한 후, 후행 성분을 지배소에 연결한다.

예) 평균 90점부터 100점까지를 모두 금상으로 처리한다.

1	평균		NP	2	
2	90점	부터	NP AJT	3	
3	100점	까지	NP OBJ	6	
4	모두		AP	6	
5	금상으로		NP AJT	6	
6	처리한다.		VP	0	

예) 이것은 바로 개인주의에서 단체주의로의 전환을 의미했다.

1	이것은		NP SBJ	6	
2	바로		AP	6	
3	개인주의	에서	NP AJT	4	
4	단체주의	로의	NP MOD	5	
5	전환을		NP OBJ	6	
6	의미했다.		VP	0	

#### 4.3.8. 장형 사동 구문

- ‘-게 하다’의 장형 사동 구문은 다른 보조 용언 구성과 문장 구조가 다르다.
- ‘A가 B가 V-게 하다’와 ‘A가 B를I에게 V-게 하다’는 문장 구조가 다르므로 이를 분석에 반영한다.

즉, 'B가'는 내포절 성분으로 'B를'에게'는 모절 성분으로 분석한다.

예) 그는 철수가 집에 가게 하였다.

1	그는	NP SBJ	5	
2	철수가	NP SBJ	4	
3	집에	NP AJT	4	
4	가게	VP	5	
5	하였다.	VP	0	

예) 그는 철수를 집에 가게 하였다.

1	그는	NP SBJ	5	
2	철수를	NP OBJ	5	
3	집에	NP AJT	4	
4	가게	VP	5	
5	하였다.	VP	0	

#### 4.3.9. 명사-부사 통용어

- 명사-부사 통용어 중 뒤에 격 조사가 붙지 않고 서술어에 의존하는 경우('오늘' 등) AP로 처리한다.

예) 오늘 할 일

1	오늘	AP	2	
2	할	VP MOD	3	
3	일	NP	0	

- 본래 부사이지만 체언을 수식하는 용법으로 쓰이는 경우에는('가장', '아주', '바로' 등) 이들의 품사적 지위를 기준으로 하여 AP로 처리한다.

예) 친구네 집은 우리 집 바로 뒤에 있어.

1	친구네	NP	2	
2	집은	NP SBJ	7	
3	우리	NP	4	
4	집	NP	6	
5	바로	AP	6	
6	뒤에	NP AJT	7	
7	있어.	VP	0	

#### 4.3.10. 서술어 생략 문장

- 내포문의 서술어가 생략되었다고 판단되는 경우 내포문의 표층에 나타난 마지막 어절에 선행하는 성분들을 의존하게 하고 마지막 어절을 모문의 서술어에 의존하게 한다.
- 내포문의 서술어가 생략된 경우 형태 기준(조사)으로 분석한다.

예) 영희는 한 달을 조건으로 매장을 임시로 설치하였다.

1	영희는	NP SBJ	7	
---	-----	--------	---	--

2	하	DP	3		
3	답을	NP OBJ	4		
4	조거으로	NP AJT	7		
5	매장을	NP OBJ	7		
6	임시로	NP AJT	7		
7	설치하였다.	VP	0		

예) 나이는 여섯 살, 이름은 철수예요.

1	나이는	NP SBJ	3		
2	여섯	DP	3		
3	살,	NP	5		
4	이름은	NP SBJ	5		
5	철수예요.	VNP	0		

→ 위의 경우 형태를 기준으로 ‘살,’을 NP로 주석하여 후행절 서술어에 연결하도록 한다.

- 모문의 서술어가 생략된 경우 형태(조사) 기준으로 분석한다.

예) 과거 하계올림픽의 정식종목으로 맞는 것은?

1	과거	NP	2		
2	하계올림픽의	NP MOD	3		
3	정식종목으로	NP AJT	4		
4	맞는	VP MOD	5		
5	것은?	NP SBJ	0		

예) 밥을.

1	밥을.	NP OBJ	0		
---	-----	--------	---	--	--

예) 철수가 밥을.

1	철수가	NP SBJ	2		
2	밥을.	NP OBJ	0		

- 신문 제목 등에서 ‘하다, 되다’ 등의 서술어가 생략되어 나오는 경우에는 선행하는 명사구에 구문표지만 부여하고 기능표지는 부여하지 않는다.

예) 카카오, 다음뉴스 채팅형 댓글 ‘타임톡’ 베타버전 개시

1	카카오,	NP	7		
2	다음뉴스	NP	4		
3	채팅형	NP	4		
4	댓글	NP	5		
5	‘타임톡’	NP	7		
6	베타버전	NP	7		
7	개시	NP	8		



## 4.4. 구어 구문 분석을 위한 추가 지침

### 4.4.1. 단어, 구 단위

#### 4.4.1.1. 감탄사구

- 본 지침의 ‘감탄사’는 <우리말샘>에서 품사가 감탄사로 되어 있는 것(응답, 감탄 등)과 머뭇거림, 발화의 시작, 화제 전환 등에 쓰이는 이른바 ‘담화 표지’를 포함하는 개념을 뜻한다. 이들 감탄사에 대해서는 형태 분석 결과와 관계없이 IP(감탄사구) 표지를 부여하고 최상위 서술어에 의존하도록 분석한다.

##### 4.4.1.1.1. 관형사, 부사, 대명사와 감탄사의 구별

- 관형사나 부사, 대명사인지 감탄사인지가 모호한 경우에는 감탄사(IP)로 처리한다. 특히, 관형사나 부사 또는 대명사로 해석하여 의존소와 의존 관계를 설정하였을 때 교차 의존 금지 제약을 어기게 되는 경우에는 <우리말샘>의 품사에 따른 형태 분석 결과와 관계없이 감탄사구(IP)로 처리하여 가장 가까운 서술어에 의존하도록 분석한다.

예) 전원주택의 그 어떤 프라이빗한 그런 느낌을 나는 되게 느끼고 싶거든.

1	전원주택의	NP MOD	6	
2	그	IP	3	
3	어떤	IP	6	
4	프라이빗한	VP MOD	6	
5	그런	DP	6	
6	느낌을	NP OBJ	9	
7	나는	NP SBJ	9	
8	되게	AP	9	
9	느끼고	VP	10	
10	싶거든.	VP AJT	0	

예) 근데 요즘 사회적으로 아파트에서 개가 짖거나 이제 사람을 물거나 하는 문제가 많이 발생하더라고.

1	근데	IP	13	
2	요즘	NP	13	
3	사회적으로	NP AJT	13	
4	아파트에서	NP AJT	6	
5	개가	NP SBJ	6	
6	짖거나	VP AJT	10	
7	이제	IP	9	
8	사람을	NP OBJ	9	
9	물거나	VP AJT	10	
10	하는	VP MOD	11	
11	문제가	NP SBJ	13	
12	많이	AP	13	
13	발생하더라고.	VP	0	

→ ‘짖거나’가 ‘하는’에 의존하는데, ‘이제’가 ‘발생하더라고’에 의존할 경우 교차 의존 금지 제약을 위반하게 된다. 따라서 이때 ‘이제’는 감탄사로 보아 IP로 처리하고 바로 다음 서술어인 ‘물거나’에

의존하도록 한다.

예) 그래 가지고 그거 조퇴를 하고 집에서 하루 쉬면은 괜찮은 줄 알고

1	그래	AP	2	
2	가지고	VP AJT	5	
3	그거	IP	5	
4	조퇴를	NP OBJ	5	
5	하고	VP AJT	8	
6	집에서	NP AJT	8	
7	하루	NP	8	
8	쉬면은	VP AJT	9	
9	괜찮은	VP MOD	10	
10	줄	NP OBJ	11	
11	알고	VP	0	

예) 나이를 먹게 되면 이게 그 혈압이나 고혈압이 이렇게 많이 오는 거 같아요.

1	나이를	NP OBJ	2	
2	먹게	VP	3	
3	되면	VP AJT	10	
4	이게	IP	10	
5	그	DP	7	
6	혈압이나	NP CNJ	7	
7	고혈압이	NP SBJ	10	
8	이렇게	VP AJT	10	
9	많이	AP	10	
10	오는	VP MOD	11	
11	거	NP	12	
12	같아요.	VP	0	

#### 4.4.1.1.2. 감탄사의 주석

- 감탄사는 감정 감탄사, 머뭇거림, 발화의 시작, 화제 전환, 부름, 대답 등으로 유형이 다양하므로, 유형에 따라 구별하여 분석한다.
- 감정 감탄사, 부름과 대답, 발화의 시작, 화제 전환 등의 경우는 최상위 서술어에 의존하도록 분석한다.

예) 와 진짜 엄청나게 오래 살았네.

1	와	IP	5	
2	진짜	AP	3	
3	엄청나게	VP AJT	4	
4	오래	AP	5	
5	살았네.	VP	0	

예) 와우 평균 수명이 5년에서 6년인데 10년 동안 산 거면 정말 오래 살긴 했다.

1	와우	IP	13	
2	평균	NP	3	
3	수명이	NP SBJ	5	
4	5년에서	NP AJT	5	
5	6년인데	VP AJT	8	
6	10년	NP	7	
7	동안	NP OBJ	8	
8	산	VP MOD	9	
9	거면	VNP AJT	13	
10	정말	AP	11	
11	오래	AP	12	
12	살긴	VP	13	

13	했다.	VP	0	
----	-----	----	---	--

예) 애들아 그래서 좀 있다가 우리 집에서 술을 먹기로 했잖아?

1	애들아	IP	9	
2	그래서	AP	9	
3	좀	AP	4	
4	있다가	VP AJT	8	
5	우리	NP	6	
6	집에서	NP AJT	8	
7	술을	NP OBJ	8	
8	먹기로	VP AJT	9	
9	했잖아?	VP	0	

예) 응 암을 예방한다고 들었어.

1	응	IP	4	
2	암을	NP OBJ	3	
3	예방한다고	NP AJT	4	
4	들었어.	VP	0	

예) 예 저도 보고 있었어요, 해리포터.

1	예	IP	3	
2	저도	NP SBJ	3	
3	보고	VP	4	
4	있었어요,	VP	5	
5	해리포터	NP	0	

예) 아 언니 것이 있어?

1	아	IP	4	
2	언니	NP MOD	3	
3	것이	NP SBJ	4	
4	있어?	VP	0	

- 머뭇거림의 감탄사는 하나의 절 안에서도 여러 번 반복되어 쓰일 수 있으므로 바로 다음 어절에 의존하는 것으로 분석한다.

예) 이제 그걸 알고 그 미리 그 음식이나 운동으로 그 대처를 잘 하면은

1	이제	IP	3	
2	그걸	NP OBJ	3	
3	알고	VP AJT	11	
4	그	IP	5	
5	미리	AP	11	
6	그	IP	7	
7	음식이나	NP CNJ	8	
8	운동으로	NP AJT	11	
9	그	IP	10	
10	대처를	NP OBJ	11	
11	잘하면은	VP	0	

예) 탕수육 뭐 중식 이런 식 이런 쪽으로 시키는 것 같아요.

1	탕수육	NP	7	
2	뭐	IP	3	
3	중식	NP	7	
4	이런	DP	5	
5	식	NP	7	
6	이런	DP	7	
7	쪽으로	NP AJT	8	
8	시키는	VP MOD	9	
9	것	NP	10	
10	같아요.	VP	0	

예) 저기 신한여행 갔을 때

1	저기	IP	2	
2	신한여행	NP	3	
3	갔을	VP MOD	4	
4	때	NP	0	

- 연결할 성분이 없는 경우에는 다음 어절에 의존하는 것으로 처리한다.

예) 응 근력은

1	응	IP	2	
2	근력은	NP	0	

- ‘아니에요, 아닙니다’ 등과 같이 감탄사로 등재되어 있지는 않지만 응답에 사용되는 경우, 또는 ‘그렇지, 맞아’ 등과 같이 용언 활용형이 감탄사로 등재되어 있기는 하지만 그 감탄사의 용법이 응답에 쓰이는 것이 아닌 경우 ‘VP’로 처리한다.

예) 아니에요 배 안 고파요.

1	아니에요	VP	4	
2	배	NP SBJ	4	
3	안	AP	4	
4	고파요.	VP	0	

예) 맞아 나도 요즘 유산균을 안 먹어서 그런지 소화가 잘 안되더라고.

1	맞아	VP	10	
2	나도	NP SBJ	6	
3	요즘	NP AJT	6	
4	유산균을	NP OBJ	6	
5	안	AP	6	
6	먹어서	VP AJT	7	
7	그런지	VP AJT	10	
8	소화가	NP SBJ	10	
9	잘	AP	10	
10	안되더라고.	VP	0	

- 호칭어는 IP로 최상위 지배소에 의존하도록 처리한다. 또한 호칭어에 선행하는 수식 요소들은 일반적인 구문 분석 지침에 따라 처리한다.

예) name1 선배님 혹시 가장 최근에 읽으신 책이 어떤 책인가요?

1	name1	NP	2	
2	선배님	IP	9	
3	혹시	AP	9	
4	가장	AP	5	
5	최근에	NP AJT	6	
6	읽으신	VP MOD	7	
7	책이	NP SBJ	9	
8	어떤	DP	9	
9	책인가요?	VNP	0	

예) 선생님 그전에 그 젤라토 아이스크림 맛있다고 했던 거 기억하고 계시죠.

1	선생님	IP	10	
2	그전에	NP AJT	7	
3	그	DP	5	
4	젤라토	NP	5	
5	아이스크림	NP SBJ	6	



예) 할인 혜택이 가능한 메뉴는 커피 메뉴를 비롯해 에이드류, 스무디류, 디저트류 등 전 메뉴가 해당된다.

1	할인	NP	2		
2	혜택이	NP_SBJ	3	◀	
3	가능한	VP_MOD	4	◀	
4	메뉴는	NP_AJT	14	◀	
5	커피	NP	6	◀	
6	메뉴를	NP_OBJ	7	◀	
7	비롯해	VP_AJT	14	◀	
8	에이드류,	NP	10	◀	
9	스무디류,	NP	10	◀	
10	디저트류	NP	11	◀	
11	등	NP	13	◀	
12	전	DP	13	◀	
13	메뉴가	NP_SBJ	14	◀	
14	해당된다.	VP	0	◀	

## 4.4.2. 절, 문장 단위

### 4.4.2.1. 후보층 구문

- 후보층 구문이 사용된 경우에는 모문과 후보층 구문은 접속(부사절 연결)과 같이 처리한다. 모문의 최상위 지배소는 모문의 서술어이며, 모문의 서술어를 후보층 구문에 연결한다.

예) 애들아, 어제 만나기로 했잖아, 우리 집에서.

1	애들아,	IP	4		
2	어제	NP	4		
3	만나기로	VP_AJT	4		
4	했잖아,	VP	6	◀	
5	우리	NP	6	◀	
6	집에서.	NP_AJT	0	◀	

→ '했잖아.'가 모문의 최상위 지배소이다. '했잖아.'를 '집에서.'에 연결한다.

- 후보층 구문은 그 지배소가 앞에 있지만, 지배소 후위 원칙에 따라 의존 관계는 후보층 구문에 표시한다.

예) 그런데 안 한대요 본인은.

1	그런데	AP	3		
2	안	AP	3		
3	한대요	VP	4	◀	
4	본인은.	NP_SBJ	0	◀	

예) 혹시 봤나요 주말지?

1	혹시	AP	2		
2	봤나요	VP	3	◀	
3	주말지?	NP_OBJ	0	◀	

예) 짝꿍이 됐어, 여름부터.

1	짜꿍이	NP CMP	2	
2	뒀어.	VP	3	
3	여름부터.	NP AJT	0	

- 후보층 구문이 두 어절 이상의 복합 표현으로 나타난 경우 일반적인 구문 분석 지침에 따라 처리하고 후보층 구문에 선행하는 문장의 최상위 지배소는 후보층 구문의 최상위 지배소에 연결한다.

예) 근데 그 스키니가 몸매가 너무 적나라하게 다 보여, 자기 단점들이.

1	근데	AP	8	
2	그	AP	3	
3	스키니가	NP SBJ	8	
4	몸매가	NP SBJ	8	
5	너무	AP	6	
6	적나라하게	VP AJT	8	
7	다	AP	8	
8	보여,	VP	10	
9	자기	NP	10	
10	단점들이	NP SBJ	0	

예) 그래서 손톱만 봐도 그 사람들은 알더라고 그 사람의 직업을.

1	그래서	AP	6	
2	손톱만	NP OBJ	3	
3	봐도	VP AJT	6	
4	그	DP	5	
5	사람들은	NP SBJ	6	
6	알더라고	VP	9	
7	그	DP	8	
8	사람의	NP MOD	9	
9	직업을.	NP OBJ	0	

예) 만질 때 안 좋은 느낌이 드는 거야 그 네일아트의 화학 약품이.

1	만질	VP MOD	2	
2	때	NP AJT	6	
3	안	AP	4	
4	좋은	VP MOD	5	
5	느낌이	NP SBJ	6	
6	드는	VP MOD	7	
7	거야	VNP	11	
8	그	DP	11	
9	네일아트의	NP MOD	11	
10	화학	NP	11	
11	약품이.	NP SBJ	0	

#### 4.4.2.2. 구조적 중의성이 있는 명사구

- 첨표(.)나 '또는', '및', '이나', '하고' 등의 외현적인 표지 없이 명사구가 나열된 경우에는 기능표지 CNJ를 부여하지 않고 구문표지만 부여한다.

예) 나는 연어초밥 광어초밥 이런 건 기본적으로 다 좋아하고

1	나는	NP SBJ	8	
2	연어초밥	NP	3	
3	광어초밥	NP	5	
4	이런	DP	5	
5	건	NP OBJ	8	

6	기본적으로	NP AJT	8	
7	다	AP	8	
8	좋아하고	VP	0	

예) 중식뷔페면 사실 짜장면 탕수육 짬뽕 이런 거면 1시간만 먹어도 배부르지 않나?

1	중식뷔페면	VNP AJT	10	
2	사실	AP	10	
3	짜장면	NP	5	
4	탕수육	NP	5	
5	짬뽕	NP	7	
6	이런	DP	7	
7	거면	VNP AJT	10	
8	1시간만	AP AJT	9	
9	먹어도	VP AJT	10	
10	배부르지	VP	11	
11	않나?	VP	0	

#### 4.4.2.3. 주격/목적격 조사 결합형이 절 표지로 기능하는 경우

- 열린 분열문 구성의 경우 분열문과 형식적·의미적 자질을 상당 부분 공유하지만 ‘것’에 후행하는 내용이 독립된 절의 성격을 가지고 주절 서술어에 계사가 나타나지 않는다는 점에서 차이를 보인다. ‘것’을 포함하는 선행 요소는 격조사를 기준하여 논항으로 분석할 수 없으므로 배경의 의미를 나타내는 절 단위에 준하는 것으로 보아 부사어로 처리하고 모문의 최상위 지배소에 연결한다.

예) 맞아 그런 게 좀 확실히 있어 보이는 게 유럽 애들이 그런 걸 되게 잘하잖아.

1	맞아	IP	14	
2	그런	DP	3	
3	게	NP SBJ	6	
4	좀	AP	6	
5	확실히	AP	6	
6	있어	VP	7	
7	보이는	VP MOD	8	
8	게	NP AJT	14	
9	유럽	NP	10	
10	애들이	NP SBJ	14	
11	그런	DP	12	
12	걸	NP OBJ	14	
13	되게	AP	14	
14	잘하잖아.	VP	0	

예) 근데 그 당시에다 나는 좀 별로 보기가 안 좋은 게 사람들이 이제 응 보는 거 노는 거 좋은데 이제 사람들 막 뭉쳐 있으니까 그게 좀 불편하더라고 나는.

1	근데	IP	26	
2	그	DP	3	
3	당시에다	NP AJT	26	
4	나는	NP SBJ	9	
5	좀	AP	9	
6	별로	AP	9	
7	보기가	NP SBJ	9	
8	안	AP	9	
9	좋은	VP MOD	10	
10	게	NP AJT	26	
11	사람들이	NP SBJ	18	
12	이제	IP	18	
13	응	IP	18	
14	보는	VP MOD	15	
15	거	NP	17	
16	노는	VP MOD	17	






#### 4.4.2.5. 서술어가 나타나지 않은 경우

- 발화 수정, 발화 중단, 말차례 빼앗김 등의 이유로 서술어가 나타나지 않은 경우 제일 마지막 어절이 최상위 지배소가 된다.

예) 공대는 별로 관심

1	공	대	는	NP	3	
2	별	로		AP	3	
3	관	심		NP	0	

→ '관심'이 최상위 지배소이다.

## [붙임4] 의미역 분석 말뭉치 구축 지침

### 의미역 분석 말뭉치 구축 지침

#### 1. 총론

##### 1.1. 지침의 목적

이 지침은 한국어 문장의 논항과 부가어의 의미역을 분석하여 의미역 분석 말뭉치를 구축하는 방안을 제시하는 것을 목적으로 한다. 의미역 분석은 구문 분석 말뭉치를 대상으로 수행한다.

- 본 지침은 '2019년 의미역 말뭉치 구축 지침'과 '2019년 의미역 기술 모형 지침'을 토대로 작성한 것이다.
- 본 지침은 이전 지침과의 내용적 일관성을 최대한 유지하되 층위별 말뭉치 구축을 염두에 두고 지침을 개정·보완하였다.
- 본 지침은 이전 지침들과 마찬가지로 한국정보통신기술협회(TTA) 및 한국전자통신연구원(ETRI)의 기준을 따르되 언어학적으로 정합적이지 않은 내용은 수정하였다.

##### 1.2. 용어 정의

- 의미역(意味役, thematic role, theta-role, semantic role): 논항 혹은 부가어가 서술어에 대해 가지는 의미적 역할
- 논항(論項, argument): 서술어가 의미적으로 요구하는 필수적인 성분
- 부가어(附加語, adjunct): 서술어와 의미적으로 관련이 있으나 수의적인 성분

##### 1.3. 기본 원칙

- 의미역 분석은 구문 분석 결과를 바탕으로 수행한다.
- 서술어를 중심으로 논항과 부가어를 파악하고 의미역 표지를 부착한다.
- 의미역 표지는 논항 혹은 부가어를 이루는 구나 절의 마지막 어절에 부착한다.
- 논항의 의미역 분석은 서술어가 이루는 절 내에 실현된 성분에 대해서만 수행한다.
- 하나의 어절은 하나의 의미역만을 가진다.
- 필수 의미역은 하나의 절 안에서 한 번만 출현한다.

- 부가 의미역은 하나의 절 안에서 중복해서 출현할 수 있다.
- 필수 의미역과 부가 의미역의 표지는 번호형이 아닌 표지형으로 부착한다.

## 2. 의미역 분석 대상 및 범위

### 2.1. 서술어의 성립

#### 2.1.1. 대상 서술어

- 동사(VV), 형용사(VA)로 이루어진 용언구(VP)의 서술어를 기준으로 의미역 분석을 수행한다.
- 단일한 서술어만 대상 서술어로 삼는다.
- 띄어쓰기로 이루어진 연속 동사구로서 사전에 등재되지 않은 경우, 구문 분석의 결과에 따라 분석한다. 따라서 선행 용언과 후행 용언 각각을 모두 대상 서술어로 취급할 수도 있고 둘 중 하나만 분석할 수도 있다.
- 서술절을 인정하지 않는다. 주격중출문도 평판적인 구조로 파악한다.

- 철수가(THM) 머리가(PRT) 좋다.

- 연어 구성, 관용 표현의 경우 전체를 하나의 서술어로 보지 않고 구성 내의 서술어를 대상 서술어로 취급한다. 또한 중복역이 발생할 경우 이를 허용한다.

- 근로 태도가 철수의 승진을(THM) 발목(을)(THM) 잡았다. [관용 표현: 발목(을) 잡다]  
 - 입학 사정관 제도를(THM) 도입을(THM) 하다. [어근 분리 현상도 연어 구성으로 처리]  
 - 술이가 사과를(THM) 반을(M-QNT) 먹었다. [관용 표현이 없으므로 THM 을 중복 할당 불가]

- 어근 분리로 인해 서술어의 논항으로 볼 수 없는 성분이 있다면, 어근이 분리되지 않은 형태를 상정하여 격틀 구조를 살핀 후 해당 논항에 의미역을 부여할 수 있다.

- 아들이(AGT) 아버지에게(GOL) 말을(THM) 한다.  
 → 어근이 분리된 '하다001'은 'X에게'라는 논항을 요구하지 않는다. 그런데 어근이 분리되지 않은 형태의 '말하다'는 'X에게' 논항을 필요로 하므로, 이를 서술어로 삼아 '아버지에게'에 'GOL'이라는 의미역을 부여할 수 있다.

- 본용언 + 보조 용언 구성에서는 본용언을 대상 서술어로 취급한다. '-어 주다' 구문의 경우 보조 용언에 의해 격틀이 달라지지만, 본용언이 'N에게' 논항을 요구하는 것으로 분석한다.

- 철수가 책을 읽다. cf. 철수가 '영희에게(GOL)' 책을 읽어 주다.

- 사전에 등재되지 않아 '777'의 번호를 부여받은 서술어도 대상 서술어로 취급한다. 다만 격틀은 아래와 같은 방식을 따른다.
- 'N 하다'에 대응하는 'N-받다/-되다/-당하다'의 경우 능동문을 만든 후에 필수 의미역 여부를 판단한다.  
가. 치료하다(X가 Y를 치료하다) → 치료받다(Y가 [X에게: 부가 의미역] 치료받다)  
나. 배치하다(X가 Y를 Z에 배치하다) → 배치받다 (Y가 [X에 의해: 부가 의미역] [Z으로:필수 의미역] 배치받다)
- 'N 하다'에 대응되는 'N 시키다'는 사동사 역할을 하는 경우와 동의어 역할을 하는 경우로 구분한 뒤, 필수 의미역 여부를 판단한다.  
가. 복직하다(X가 Y에 복직하다) → 복직시키다([Z가:필수 의미역] X를 Y에 복직시키다)  
나. 연결하다(X가 Y를 Z에 연결하다) → 연결시키다(X가 Y를 Z에 연결시키다)처럼 동의어의 경우에는 N하단의 격틀을 참고한다.
- 'V 어지다'는 'V'의 피동문을 만들고 'V 어하다'는 'V'의 타동문을 만든 후에 필수 의미역 여부를 판단한다.  
가. 만들다(X가 Y를 만들다) → 만들어지다(X가 만들어지다)  
나. 힘들다(X가 힘들다) → 힘들어하다 (X가 Y를 힘들어하다)
- 그 외에 미등재어의 경우에는 맥락을 살펴 의미가 최대한 유사한 서술어의 격틀을 참고하여 의미역을 분석한다.

- 프레시하다777 → 신선하다003

- 단, 띄어쓰기 실수로 인해 하나의 서술어가 둘 이상의 어절로 나뉜 경우에는 원 어절을 수정한 후 의미역 분석을 수행한다.

가. 상실했 다 → '상실했다'로 분석한다.  
나. 먹고 살다 → 합성어로 사전에 등재되어 있으므로 '먹고살다'로 분석한다.  
다. 먹여살리다 → 사전에 등재되지 않은 연속 동사구이므로 '먹여 살리다'로 원문을 수정한다.

## 2.1.2. 제외 서술어

- 명사+'이다'로 이루어진 긍정지정사구(VNP)는 의미역 분석 대상에서 제외한다.
- 따라서 -ㄴ/ㄹ 것/터/뿐/따름/모양/지경/참/중/노릇/예정/전망/계획/처지/상태/이유/생각/길+이다 등의 구성은 분석하지 않는다.
- 서술성 명사도 분석 대상에서 제외한다.
- 따라서 신문 기사 헤드라인 등에서 자주 관찰되는 서술성 명사를 분석하지 않도록 주의한다.
- 가운뎃점, 쉼표 또는 괄호에 병기되어 나타난 서술어의 경우 다음과 같이 분석한다.

가. 체결·공포하다. → 구문 분석 결과 VP 로 분석되는 것은 ‘공포하다’이므로 이를 서술어로 삼음.  
 나. 체결, 공포하다. → 구문 분석 결과 VP 로 분석되는 ‘공포하다’이므로 이를 서술어로 삼음.  
 다. 체결(공포)하다. → 구문 분석의 결과 VP 로 분석되지 않으므로 전체를 작업 대상에서 제외.

- 가운데맺점, 침표가 없거나 어근이 나열된 서술어의 경우 다음과 같이 분석한다.

가. 폐점 혹은 이전하다. → ‘폐점’은 구문 분석 결과 NP 로 분석되므로 분석 대상 서술어로 삼지 않음. ‘이전하다’에 의존하는 ‘폐점’, ‘혹은’에 대해서는 의미역을 분석하지 않음.  
 나. 지속 확대하다. → ‘지속’은 구문 분석 결과 NP 로 분석되므로 분석 대상 서술어로 삼지 않음. ‘지속’은 ‘확대하다’를 수식하는 부가어로 분석함.  
 다. 무상 수거하다. → ‘무상’은 구문 분석 결과 NP 로 분석되므로 분석 대상 서술어로 삼지 않음. ‘무상’은 ‘수거하다’를 수식하는 부가어로 분석함.

- 단일 서술어가 따옴표 사이에 있을 경우, 이는 분석 대상으로 한다.

- 올면서 시작했으나 마지막에는 ‘웃었다’. → ‘웃었다’를 분석한다.

- 본용언+보조 용언 구성에서 보조 용언은 분석 대상에서 제외한다.
- 붙여쓰기로 처리된 본용언+보조 용언의 구성의 경우에는 원문을 수정한 후 보조 용언을 분석 대상에서 제외한다. 단, ‘-어지다’, ‘-어하다’는 띄어 써야 하는 본용언+보조 용언 구성이 아니므로 통합된 형태에 대한 의미역을 분석한다.
- 보조용언 목록은 <우리말샘>에 등재된 보조용언 목록을 따른다.

-어 가다	-지 말다	-ㄴ 뻔하다	-어 져하다
-어 가지다	-고(야) 말다	-게 생기다	-어 주다
-어 나가다	-어 말다	-ㄴ 성부르다	-어 죽다
-어 나다	-어 먹다	-ㄴ/ㄹ 성싶다	-ㄴ/음 직하다
-고 나다	-지 못하다	-ㄴ/ㄹ 성하다	-은/는 척하다
-어 내다	-다(가) 못하다	-고 싶다	-은/는 체하다
-어 놓다	-어 버릇하다	-ㄴ까 싶다	-어 치우다
-어 대다	-어 버리다	-었으면 싶다	-어 터지다
-어 두다	-ㄴ 법하다	-어 쌓다	-게 하다
-어 드리다	-어 보다	-지 않다	-었으면 하다
-려고/-기로/-자고 들다	-고 보다	-고 앓다	-어야 하다
-고 들다	-다(가) 보다	-ㄴ/는 양하다	-으려(고)/고자 하다
-어 들다	-ㄴ가/나 보다	-어 오다	-기는/기도/기나 하다
-ㄴ/ㄹ 듯싶다	-ㄴ까 보다	-어 있다	-고 하다
-ㄴ/ㄹ 듯하다	-다 보다	-고 있다	-고는/곤 하다
-ㄴ 만하다	-어 빠지다	-어 재끼다	

- 의사 보조용언 구성은 의미역 분석 서술어에서 제외한다.

-ㄴ 수 있다/없다	-ㄴ/ㄹ 예정이다
-ㄴ 리(가) 있다/없다	-ㄴ/ㄹ 참이다

-ㄴ/ㄹ 것이다	-ㄴ/ㄹ 터이다
-ㄴ/ㄹ 길이다	-ㄴ/ㄹ 것 같다
-ㄴ/ㄹ 노릇이다	-ㄹ 것(을/걸) 그랬다
-ㄴ/는 말이다	-어서는 안 되다
-ㄴ/ㄹ 모양이다	-든지 하다
-ㄴ/ㄹ 뿐이다	N-이지 않다/못하다/말다
VV-기는/기나/기도 VV	VA-기는/기나/기도 VA

- 주로 부사형, 관형사형으로 쓰여 문장에서 독립절을 구성하기 어려운 서술어는 의미역 분석 서술어에서 제외한다.
  - 부사형

가지다007	~를 가지고[=로]	밀가루를 가지고 만든 떡
각설하다003	각설하고	자, 각설하고, 어디 당신 계획이나 들어봅시다.
같다004	같으면, 같아서, 같으니	옛날 같으면
같다006	같으면, 같아서, 같으니	마음 같아서는
같다008	같으면, 같아서, 같으니	바보 같으니라고.
거두절미하다002	~을 거두절미하고	거두절미하고 용건만 말씀드리겠습니다.
걸치다002	~에 걸쳐	열 시간에 걸쳐 회의가 진행됐다.
관련되다001	~와/에 관련되어	적지 않은 사람들이 이번 사건과 관련돼 있다.
관련하다001	~와/에 관련해	교육 개혁안과 관련해 논평을 했다.
관하다002	~에 관해	다음은 여성의 사회적 지위에 관하여 토론하도록 하겠습니다.
기하다005	~을 기해	오늘 새벽 4시를 기해 폭풍 주의보가 발효되었다.
까닥하다002	까닥하면/까닥하다가는[추측]	까닥하다가는 당신 목숨까지 위태로울 수 있습니다.
깍하다001	깍해야	있는 돈을 다 모아도 깍해야 3천 원이다.
내놓다012	내놓고	내놓고 잘난 체하다.
까놓다001	까놓고 말해	어디 톡 까놓고 말을 해 봐!
넉넉잡다001	넉넉잡고/넉넉잡아	넉넉잡고 사흘이면 일이 끝난다.
놓다005	~을 놓고[=에 대해]	동문회에서 학교 이전 문제를 놓고 의견이 분분했다.
대다002	~에 대고	하늘에 대고 하소연을 했다.
대하다004	~에 대해	전통문화에 대한 관심
더볼다001~003	~와 더불어	이웃과 더불어 사는 세상
두다014	~을 두고	머리를 북쪽으로 두고 자면 빨리 죽는다는 말이 있다.
두르다011	둘러 말하다	자꾸 둘러 말하지 말고 직설적으로 말해라.
둘러싸다003	~를 둘러싸고[=에 대해]	이 문제를 둘러싸고 의견이 분분하다.
뒤잇다001	~에 뒤이어	축가에 뒤이어 축사가 있겠습니다.
들다013	[특정 시기] 들어/들자	올해 들어 해외여행자 수가 부쩍 늘었다.
따르다004	~을 따라	그는 아버지의 뜻을 따라서 법대에 진학했다.
따르다008	~에 따른	사용 목적에 따른 분류.

말다006	말고	너 말고 영희
말하다007	말하자면/ 소위, 다시 말해	말하자면 그것은 커다란 충격이었고 환멸이었다.
말하다008	~로 말하면	힘센 걸로 말하면 우리 아버지 따라갈 사람이 없다.
말미암다001	~로 말미암아	난데없는 구렁이의 출현으로 말미암아 우리 집은 삽시에 엉망진창이 되어 버렸다.
머지않다001	머지않아	머지않아 소식이 올 것이다.
못하다003	못해도 얼마	잡은 고기가 못해도 열 마리는 되겠지.
무작정하다001~002	무작정하고	무작정한 낙관주의.
무조건하다001	무조건하고	의기에 감동되어 무조건하고 곡식과 피륙을 바치는 사람들도 많았다.
밤새다001	밤새도록	밤새도록 일하다.
보다025	~를 보고[=에게/에 대해]	너를 보고 하는 말이 아니야.
불구하다002	불구하고	몸살에도 불구하고 출근하다.
비롯하다003	비롯해	이장을 비롯한 마을 사람들.
비하다002	비해	그는 사진에 비해서 실물이 훨씬 더 좋은 인상을 풍겼다.
비하다003	비하건대	낭비가 심한 그에게 돈을 주는 것은, 비하건대 밑 빠진 독에 물 붓기다.
아니다001	아니라/아니면	철수 아니라 영희/철수 아니면 영희
아무렇다002	아무렇게나	옷을 아무렇게나 벗어 놓았다.
암만하다001~002	암만해도	암만해도 알 수가 없다.
어째서001	어찌하여서, 어째서	너희들은 어째서 만나기만 하면 싸우냐?
얼추잡다001	얼추잡아	지각한 놈들만도 얼추잡아 여남은 명은 되었다.
위시하다002	~을 위시하여	그를 위시하여 여러 사람이 그 일에 참여했다.
위하다003	~을 위해	시장 조사를 위해 현지 출장을 갔다.
의하다001	~에 의하면/의해	노동에 의해 소득을 얻었다.
인하다001	~로 인해	부주의로 인한 사고
잇다004	이어/이어서	개회사에 이어 회장의 인사 말씀이 있겠습니다.
잇달다001~002	잇달아	유세장에 유권자들이 잇달아 몰려들었다.
잇따르다001~002	잇따라	잇따른 범죄 사건 때문에 밤길을 다니기가 두렵다.
연달다001~002	연달아	차가 연달아 온다.
연잇다001	연이어	영화 상영이 끝나자 사람이 연이어 극장 문을 나온다.
있다019	~에/에게 있어	인간에게 있어서 가장 중요한 것은 사랑이다.
자시다002	~고 자시고	누군들 그 경황에 누굴 알아보고 자시고 할 사이가 있었겠나.
작다007	작게는	전쟁은 작게는 개인의 생활에서 크게는 국가의 운명까지도 좌우한다.
잘못하다004	잘못하다가는	잘못하다가는 남편의 생명도 부지하지 못할 모양이다.
잘못하다005	잘못하면	눈치를 보니 자칫 잘못하면 큰 싸움이 날 것 같아



		나는 그 둘의 중재에 나섰다.
잘하다007	잘하면	잘하면 올해도 풍년이 들겠다.
잘하다008	잘해서/잘해야[추측]	잘해야 열 사람 중에서 한두 사람쯤 살아날 수 있 다는데.
줄잡다002	줄잡아	이번 출장은 줄잡아 한 달은 걸릴 것 같다.
즈음하다001	즈음하여/즈음해서	본 여단은 출동에 즈음해서 미증의 곤란을 겪고 있 습니다
참고삼다002	참고삼아	이건 어디까지나 참고삼아 말씀드리는 겁니다.
추후하다001	추후하여	자세한 의견은 추후하여 말씀드리겠습니다.
치다047	~로 치면/친다면	종갓집 형은 촌수로 치면 팔촌 형이고 연배로 치면 20여 세 위였다.
크다011	크게	교육을 크게 보면 국가와 사회를 위해 시키는 것이 고 작게 보면 자기 자신을 위해 하는 것이다.
통하다011~014	~을 통해	비상구를 통해 빠져나간다.
폐일언하다001	~을 폐일언하고	폐일언하고 당장 시작합시다.
하다019	~로 해서[=로]	우리는 부산으로 해서 일본에 갈 생각이다.
하다021	하면	물 하면 역시 우리나라 물이 최고다.
하다022	하고	내 친구는 이번에 서울하고도 강남에 집을 샀다.
하다030	하면	그는 컸다 하면 홈런이다.
하다032	하고	보초는 “손 들어!” 하고 크게 외쳤다.
하다033	하니/하면/하여/해서	꾸준히 연습하여라. 하면 어느 순간에 실력이 늘어 있을 것이다.
하다888	~만 해도[가정해도/따져봐 도]	철수만 해도 그 일을 잘한다.
향하다002~004	~을 향해	그의 모든 관심과 열정은 영화를 향해 있다.
허실삼다001	~을 허실삼아	행여 무엇이나 입에 맞는 떡이 있을까 하는 생각으 로 허실삼아 가 보았다.

○ 관형사형

갈데없다001	갈데없는	갈데없는 처지/신세
같다005	같은	말 같은 말
걸치다002	~에 걸친	이틀에 걸친 전투는 끝이 났다.
관련되다001	~와/에 관련된	마약과 관련된 범죄
관련하다001	~와/에 관련한	교육 개혁안과 관련한 논평
관하다001	~에 관한	그 문제에 관한 한 우리는 한 치도 양보할 수 없습 니다.
대하다004	~에 대한	신탁 통치안에 대한 우리 민족의 반대 운동은 전국 적이었다.
둘러싸다003	~를 둘러싼[=에 대한]	강도 사건을 둘러싼 사람들의 반응은 다양했다.
뒤잇다001	~에 뒤이은	브랜드 확장에 뒤이은 공격적 마케팅
따르다004/008	~에 따른	사용 목적에 따른 분류
말미암다001	~로 말미암은	달균이가 돈을 벌어 오겠다고 절 굴을 뜯는 것도 그

		로 말미암은 가난 때문이었다.
불구하다002	불구한	재개봉임에도 불구한 폭발적 인기
비롯하다003	비롯한	이장을 비롯한 마을 사람들.
비하다002	~에 비한	여름철, 남들에 비한 과한 땀
아니다001	아닌	철수 아닌 영희
아무려하다002	아무려한	그는 꿀 먹은 벙어리처럼 아무려한 대답도 못 하고 있었다.
어떠하다001	어떠한	어떠한 돈으로 이 물건을 샀느냐?
여하하다001	여하한	여하한 어려움도 견딜 수 있다.
어쩌다002	어쩐	어쩐 일로 전화하셨소?
위시하다002	~을 위시하여	황제는 대답 대신, 우발산과 변약유를 위시한 자기 사람들을 불러 모았다.
위하다003	~을 위한	시장 조사를 위한 현지 출장
의하다001	~에 의한	노동에 의한 소득
인하다001	~로 인한	부주의로 인한 사고
잇다004	이은	이 일은 업무 협약에 이은 일이다.
잇달다001~002	잇단	잇단 범죄 사건
잇따르다001/002	잇따른	잇따른 범죄 사건
연잇다001	연이은	하늘과 연이은 수평선
지나가다003	지나가는	지나가는 말
즈음하다	즈음한	어린이날 즈음한 성명서 발표
하다024	다고 하는	거짓말을 했다고 하는 증거
향하다002~004	~을 향한	그는 아까부터 그녀에게 시선을 향한 채 꿈쩍도 하지 않고 있다.
현행하다001	현행하는	세무 공무원이 현행하는 조세법을 몰라서야 말이 되는가.

- 해당 의미 번호가 아닌 다의어는 의미역 분석 대상이 됨을 유의한다. 예컨대 ‘대(對)하다’의 의미 중 배제 대상이 되는 것은 ‘004 【…에】 ((‘대한’, ‘대하여’ 꼴로 쓰여)) 대상이나 상대로 삼다.’만이다.

대-하다(對하다)
대-하다 「002」 「동사」 【…을】 【(…과) …을】 ((‘…과’가 나타나지 않을 때는 여럿임을 뜻하는 말이 주어로 온다))마주 향하여 있다.
대-하다 「003」 「동사」 【…에/에게 -게】 【…을 …으로】 【…을 -게】 ((‘…으로’나 ‘-게’ 성분은 ‘…처럼, -은/을 듯이’ 따위의 부사어나 ‘-이/히’ 부사로 대체될 수 있다))어떤 태도로 상대하다.
대-하다 「004」 「동사」 【…에】 ((‘대한’, ‘대하여’ 꼴로 쓰여))대상이나 상대로 삼다.
대-하다 「005」 「동사」 【…을】 작품 따위를 직접 읽거나 감상하다.

- 또한 이들 목록이 의미역을 요구하지 않기는 하지만, 이 구성의 범위 전체는 서술어에 대한 부가 의미역으로 얼마든지 분석될 수 있으므로 이를 혼동하지 않도록 한다.

가. <b>특검 연장에 의해(M-INS)</b> 성난 민심이 안정화되었다.
나. 우리는 <b>그 문제에 대해(M-EXT)</b> 대책을 논의했다.
다. <b>증거가 확인되지 않음에 따라(M-PRE)</b> 가해자는 불기소되었다.

- 감탄사처럼 쓰인 서술어도 분석 대상에서 제외한다.
  - ‘그래’, ‘응’, ‘맞아’, ‘아니야’, ‘좋아’처럼 상대방의 말에 동의하는 의미로 쓰인 경우, 혹은 단독으로 나타나거나 강조 부사와 동반하여 나타난 경우에는 분석 대상으로 삼지 않는다.
  - 그밖에 문두에서 쓰이는 ‘있지/있잖아’, ‘가만있자’나 ‘죽어라/죽자 (하고)’의 경우에도 분석 대상에서 제외한다.
  - 서술어가 아닌 ‘이제’, ‘뭐’처럼 실질적으로 어휘의미를 담지 않은 말버릇 등의 간투사 역시 의미역 분석 대상으로 삼지 않는다.

## 2.2. 논항의 성립

- 서술어가 필수적으로 요구하는 성분을 논항으로 분석한다.
- 논항에는 주어(NP\_SBJ), 목적어(NP\_OBJ), 보어(NP\_CMP), 필수적 부사어(NP\_AJT, VP\_AJT)가 포함된다.
- 서술어의 논항 여부는 <우리말샘>의 격틀을 참고하여 결정한다.
- 서술어가 다의어인 경우, 의미에 따라 격틀이 달라지는 경우가 있으므로 어휘의미분석 결과를 참고하여 해당 의미의 격틀을 참고한다.

보다  
 보다001 【…을】 을 눈으로 대상의 존재나 형태적 특징을 알다.  
 ...  
 보다043 【…을】 【…으로】 어떤 상황이나 사실로 미루어 짐작하거나 판단하다.  
 ...  
 보다027 【(…과)】 【…을】 ((‘…과’가 나타나지 않을 때는 여럿임을 뜻하는 말이 주어로 온다)) 사람을 만나다.  
 보다028 【…을 …으로】 【…을 -게】 【…을 -고】 【…으로】 【-고】 ((‘…으로’나 ‘-게’ 대신에 평가를 뜻하는 다른 부사어가 쓰이기도 한다)) 대상을 어떠한다고 평가하다.

- ‘명사’ 또는 ‘명사(구)+조사’의 형식으로 이루어진 것만 서술어의 논항으로 취급한다. 단, 발화동사와 사유동사의 경우 인용절을 논항(CNT)으로 취급한다. ‘굴다, 생기다, 보이다; 생각하다, 여기다’ 등의 서술어의 경우 ‘-게’ 부사절을 예외적으로 논항(MNR; FNS)으로 취급한다.
- 인용절은 길이를 따지지 않고 전체를 논항으로 설정하여 분석한다.
- 서술어가 ‘용언+명사형 어미 ‘-ㅁ’으로 이루어진 경우 논항이 ‘명사(구)+의’ 형태로 실현되어 있으면 의미역 분석에서 제외한다.

- 전투의 급박함 속에서도 소대장은 침착하게 행동했다.  
 → ‘전투’는 ‘급박하다’와 의미적으로 관련이 있으나 ‘전투의’가 ‘급박함’ 전체를 수식하므로 ‘전투의’는 ‘급박하다’가 이끄는 절 내의 성분이 아닌 것으로 보아 ‘급박하다’의 논항으로 분석하지 않는다.  
 - 전투가 급박함에도 소대장은 침착하게 행동했다.  
 → ‘전투가’가 ‘급박하다’의 논항이 되므로 분석 대상에 포함한다.

- 관형절의 경우 논항의 범위 전체를 대상으로 삼되, 절 내에서 실제 실현된 논항만 분석 대상으로 삼는다.

가. 영희는 **[[비싼] 화장품(THM)]** 좋아한다.

→ 서술어 '비싼'의 논항이 '화장품'일 수도 있음. 그러나 같은 절 내에 실현되지 않으므로 의미역 표지를 부여하지 않음. 다만 '비싼 화장품' 전체는 '좋아하다'의 논항임.

나. 영희는 **[[가격이(FTR) 비싼] 화장품(THM)]** 좋아한다.

→ 서술어 '비싼'의 논항인 '가격이'가 같은 절 내에 실현되었으므로 '비싼'을 서술어로 삼아 '가격이'에 의미역을 부여함. 또한 '가격이 비싼 화장품' 전체는 '좋아하다'의 논항임.

- 관형절인 유형별 처리 예시는 다음과 같다.

가. 관계 관형절

- 영희는 [가격이 비싼 화장품](THM) 좋아한다.

나. 관형절 + NP의 + 핵어 명사

- 영희는 [해가 빨리 뜨는 여름의 아침을](THM) 좋아한다.

다. '와/과' 접속 명사구가 포함된 관형절

- 영희는 [가격이 비싼 화장품과 예쁜 옷을](THM) 좋아한다.

라. 문장으로 환원되지 않는 관형절

- 영희는 [어린 시절을](THM) 회상했다.

- 영희는 [아기가 자고 있는 동안을](THM) 활용했다.

마. '것' 명사절

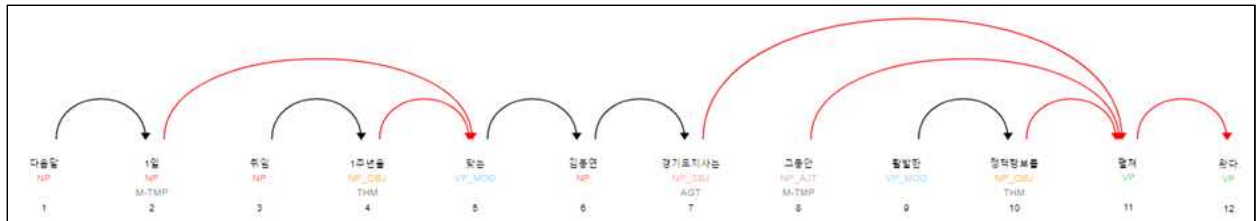
- 영희는 [아기가 잠든 것을](THM) 몰랐다.

## 2.3. 부가어의 성립

- 서술어에 의존하는 성분 중에서 필수적인 성분이 아닌 것을 부가어로 분석한다.
- 부가어에는 수의적인 부사어(NP\_AJT)와 부사절(VP\_AJT)이 포함된다.
- 구문 분석에서 접속문의 선행절도 부사절로 취급하므로 이들도 부가어로 분석한다.
- 절과 절 사이의 의미 관계도 후행절 서술어에 대한 선행절의 의미역으로 파악한다.
- '아주', '정성껏' 등 일반 부사어의 형태도 부가어로 분석한다.

## 2.4. 의미역 분석 대상 범위 설정

- 서술어에 의존하는 어절과 그 어절에 의존하는 어절들의 연쇄를 모두 포함하여 범위(span)을 결정한다.
- 다음의 예를 참고한다.



- 예문: 다음달 1일 취임 1주년을 맞는 김동연 경기도지사는 그동안 활발한 정책 행보를 펼쳐왔다.
- 우리의 분석 대상이 되는 서술어는 '5/맞는', '11/펼쳐'이다. Lemma 는 각각 '5/맞다'와 '11/펼치다'이다.
- 우선 서술어 '5/맞다'의 THM 의 범위는 단순히 끝점인 '[4]1주년을'이 아니다. '[4]1주년을'에 의존하는 '[3]취임'을 포함해야 한다.
- 따라서 서술어 '5/맞다'의 THM 의 범위는 다음과 같다.
  - 범위(span): [3, 4] 취임 1주년을
  - 의미역: THM
  - 서술어: 5/맞다
- 서술어 '5/맞다'의 M-TMP 의 범위는 단순히 끝점인 '[2]1일'이 아니다. '[2]1일'에 의존하는 '[1]다음달'을 포함해야 한다.
- 따라서 서술어 '5/맞다'의 M-TMP 의 범위는 다음과 같다.
  - 범위(span): [1, 2] 다음달 1일
  - 의미역: M-TMP
  - 서술어: 5/맞다
- 그리고 서술어 '11/펼치다'의 AGT 의 범위는 단순히 끝점인 '[7]경기도지사는'이 아니다.
- '[7]경기도지사는'에 의존하는 '[6]김동연'과 그것에 의존하는 '[5]맞는', 그리고 '[5]맞는'에 의존하는 '[4]1주년을', '[4]1주년을'에 의존하는 '[3]취임'을, 또한 '[5]맞는'에 의존하는 '[2]1일', 마지막으로 '[2]1일'에 의존하는 '[1]다음달'까지 모두 포함하는 것이어야 한다.
- 따라서 서술어 '11/펼치다'의 AGT 의 범위는 다음과 같다.
  - 범위(span): [1, 2, 3, 4, 5, 6, 7] 다음달 1일 취임 1주년을 맞는 김동연 경기도지사는
  - 의미역: AGT
  - 서술어: 11/펼치다
- 서술어 '11/펼치다'의 THM 의 범위는 단순히 끝점인 [10]'정책행보를'이 아니다.
- '[10]정책행보를'에 의존하는 '[9]활발한'을 포함하는 것이어야 한다.
- 따라서 서술어 '11/펼치다'의 THM 의 범위는 다음과 같다.

→ 범위(span): [9, 10] 활발한 정책행보를

의미역: THM

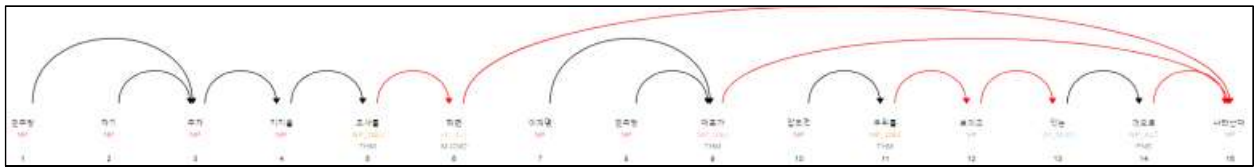
서술어: 11/펼치다

- '11/펼치다'는 오직 서술어 '11/펼치다'에만 의존하므로 가령, '11/펼치다'가 '11/펼치다'의 '11/펼치다'와 같은 '11/펼치다'에 의존한다고 보아서는 안 된다.
- 따라서 서술어 '11/펼치다'의 M-TMP의 범위는 다음과 같다.

→ 범위(span): [8] 그동안

의미역: M-TMP

서술어: 11/펼치다



- 예문: 민주당 차기 주자 지지율 조사를 하면 이재명 민주당 대표가 압도적 우위를 보이고 있는 것으로 나타난다.
- 우리의 분석 대상이 되는 서술어는 '6/하면', '12/보이고', '15/나타난다'이다. Lemma는 각각 '6/하다', '12/보이다', '15/나타나다'이다. '13/있는'은 보조 용언이므로 우리의 분석 대상이 아니다.
- 우선, 서술어 '6/하다'의 THM의 범위는 단순히 끝점인 '5/조사를'이 아니라 그에 의존하는 어절들의 연쇄를 모두 포함한 '1, 2, 3, 4, 5/민주당 차기 주자 지지율 조사를'이다.
- 여기서 서술어 '6/하면'이 이끄는 절은 '15/나타난다'에 의존하는 부사절 M-CND이다.
- '15/나타난다'에 의존하는 선행 부사절의 범위는 단순히 끝점인 '6/하면'이 아니라 해당 부사절 전체를 모두 포함하는 것이어야 한다.
- 따라서 서술어 '15/나타난다'의 선행 부사절 M-CND의 범위는 다음과 같다.

→ 범위(span): [1, 2, 3, 4, 5, 6] 민주당 차기 주자 지지율 조사를 하면

의미역: M-CND

서술어: 15/나타나다

- 서술어 '15/나타나다'의 THM의 범위는 '7, 8, 9/이재명 민주당 대표가'이다.
- 서술어 '12/보이고'의 THM의 범위는 '10, 11/압도적 우위를'이다.
- 한편 서술어 '15/나타나다'의 FNS는 단순히 끝점인 '14/것으로'가 아니다. '10, 11, 12, 13/압도적 우위를 보이고 있는'이 '14/것으로'에 의존하기 때문이다. '13/있는'은 그 자체로는 논항을 부여하지 않지만 그 앞의 '10, 11, 12/압도적 우위를 보이고'에 연쇄하여 '14/것으로'에 의존한다.
- 따라서 서술어 '15/나타나다'의 FNS의 범위는 다음과 같다.

→ 범위(span): [10, 11, 12, 13, 14] 압도적 우위를 보이고 있는 것으로

의미역: FNS

서술어: 15/나타나다

### 3. 의미역 표지 집합

#### 3.1. 필수 의미역

- 필수 의미역 표지는 논항의 의미역을 표시하는 것이다.

<필수 의미역 표지 집합>

연번	약호	영문명	국문명	예
1	AGT	Agent	행동주	내가 저녁을 만들었다. 철수가 영희에게 빵을 먹였다.(사동문의 주어) 선생님께서 우리를 부르신다. 정부에서 일본측에 항의하였다.
2	EXP	Experiencer	경험주	철수는 영희를 보았다.(지각적 경험 주체) 철수는 영희에 대해 잘 안다.(인지적 경험 주체) 철수는 영희가 무서웠다.(감정적 경험 주체)
3	MAG	Mental Agent	심리행위주	(의도성이 느껴지는 심리동사 서술어의 주체) 철수는 영희를 좋아한다. 철수는 영희를 무서워한다. cf. 철수는(EXP) 영희가 무서웠다. (심리형용사)
4	EFF	Effector	영향주	강풍이 온 산을 덮었다. 철수가 전봇대에 부딪혔다.
5	THM	Theme	대상	철수가 아주 침착하다.(형용사문의 주어) 감자에 싹이 났다.(자동사문의 주어) 도둑이 경찰에게 잡혔다.(피동문의 주어) 철수는 영희를 좋아한다.(타동사문의 목적어)
6	COM	Companion	동반주	철수는 영희와 싸웠다. (주어 참여자와 동일 지위) 철수는 모래를 물과 섞었다. (목적어 참여자와 동일 지위)
7	GOL	Goal	도착점	철수는 학교에 갔다. 철수가 자리에 앉았다. 철수가 영희에게 빵을 먹였다. (사동문의 피사동주)
8	LOC	Location	장소	철수는 방에 있다. (구체적 공간) 철수는 곤경에 처해 있다.(추상적 공간)
9	SRC	Source	출발점	정부가 옥수수를 미국으로부터 수입했다. 이 기단은 서쪽에서 왔다.
10	CNT	Content	내용	그는 자기가 천재라고 말했다. 나는 국민과 언제나 함께하겠다 연설했다.
11	INS	Instrument	도구	이 물질은 3개의 원소로 이루어졌다. 창고가 책으로 가득하다.
12	CRT	Criterion	기준	철수의 실력은 영희보다 낮다. 이 일은 내 능력에 부친다.
13	FNS	Final State	결과상태	항아리의 물이 포도주가/포도주로 되었다. 우리 업체는 인삼을 당의정으로 가공한다. 선생님은 철수를 사위로 삼았다.
14	DIR	Direction	방향	철수는 학교로 갔다. 동아시아에서 중앙아시아로 이주했다.

15	QLF	Qualification	자격	그는 <u>과학자</u> 로 유명하다. 그는 <u>군</u> 에서 <u>군의학관</u> 으로 복무했다.
16	MNR	Manner	방식	철수는 <u>어리석게</u> 군다. 철수는 <u>아이처럼</u> 군다. 영희가 <u>예쁘게</u> 생겼다. 영희가 <u>착하게</u> 보인다.
17	PRT	Part	부분	(주격중출문에서의 NP2) 철수가 <u>키가</u> 크다. (형용사문) 영희는 <u>가슴이</u> 두근거렸다. (자동사문)
18	FTR	Feature	속성	(주격중출문에서의 NP2) 영희가 <u>심성이</u> 곱다. 이 체육관은 <u>규모가</u> 크다



## 3.2. 부가 의미역

- 부가 의미역 표지는 부가어의 의미역을 표시하는 것이다.

<부가 의미역 표지 집합>

연번	약호	영문명	국문명	예문
1	M-TMP	Temporal	시간	이 지역에는 봄에(M-TMP) 안개가 정말 많이 낀다.
2	M-LOC	Location	장소	봄이 되자 공원에(M-LOC) 꽃이 만발했다.
3	M-DIR	Direction	방향	달이 서쪽으로(M-DIR) 기울었다.
4	M-ROU	Route	경로	우리는 목적지까지 <u>지름길로</u> (M-ROU) 가기로 했다.
5	M-INS	Instrument	도구	아이들은 색연필로(M-INS) 알록달록 그림을 칠했다.
6	M-MNR	Manner	방식	철수는 아들을 <u>정성으로/정성껏</u> (M-MNR) 간호했다.
7	M-QNT	Quantity	수량	질문 <u>하나</u> (M-QNT) 합시다.
8	M-DGR	Degree	정도	그들은 아름다운 경치에 <u>흠뻑</u> (M-DGR) 도취되었다.
9	M-EXT	Extent	범위	이 작품은 <u>그의 소설들 가운데</u> (M-EXT) 가장 걸출하다.
10	M-COM	Comitative	공동	철수는 <u>영희와</u> (M-COM) 극장에 갔다.
11	M-CRT	Criterion	기준	오늘이 <u>어제보다</u> (M-CRT) 더 시원하다.
12	M-QLF	Qualification	자격	그는 삼남삼녀 가운데 <u>막내로</u> (M-QLF) 태어났다.
13	M-CAU	Cause	원인	굴이 <u>설익어서</u> (M-CAU) 시다.
14	M-PRP	Purpose	목적	나는 형의 심부름으로 <u>커피를 사러</u> (M-PRP) 카페에 갔다.
15	M-CND	Condition	조건	아이들은 <u>마트에 가면</u> (M-CND) 장난감을 사 달라고 떼 쓴다.
16	M-CNC	Concession	양보	어머니는 <u>아무리 늦게 주무시더라도</u> (M-CNC) 항상 새벽 5시에 일어나신다.
17	M-PRE	Preceding	선행	설날에 <u>온 가족이 모여</u> (M-PRE) 만두를 빚었다.
18	M-BCK	Background	배경	<u>소설책을 샀는데</u> (M-BCK) 아주 재미있더라.
19	M-LST	List	나열	언니는 학교에 <u>가고</u> (M-LST) 부모님은 출근하셨다.
20	M-CTR	Contrast	대조	이 음식은 생긴 건 우습지만(M-CTR) 맛은 있다.
21	M-SEL	Selective	선택	도서관에는 <u>철수가 가든지</u> (M-SEL) 영희가 갈 것이다.
22	M-NEG	Negation	부정	철수는 술을 <u>안</u> (M-NEG) 마신다.
23	M-MOD	Modal	양태	<u>다행히</u> (M-MOD) 철수가 제시간에 나타났다.
24	M-CNJ	Conjunction	접속	<u>또한</u> (M-CNJ) 정부는 경영진에게도 강력히 경고했다.

## 4. 필수 의미역 분석의 실제

### 4.1. AGT (Agent, 행동주)

#### [정의]

- 서술어가 동작이나 행위를 표현할 때 자신의 의지에 의해 능동적으로 행위를 하는 주체.

#### [예시]

- (1) 다음은 주어가 의도성을 가지고 행위의 주체가 되는 의미역으로 '이/가', '께서'로 실현되는 성분이다. 그리고 보조사 '은/는'으로 실현된 주어도 행위의 주체이고 의도성을 가지면 AGT로 분석한다.

가. 내가(AGT) 그 빵을 먹었다.  
나. 선생님께서(AGT) 주변에게 칠판을 닦으라고 하셨다.  
다. 영희는(AGT) 학교에서 아이들을 가르친다.

- (2) 단체 명사 + '에서' 성분도 AGT에 해당한다.

가. 시에서(AGT) 생태 공원을 조성했다.  
나. 교수단체에서(AGT) 시국 선언을 검토했다.  
다. 질병관리청에서(AGT) 감염병에 대한 역학 조사를 시작했다.

- (3) 사동문의 사동주도 AGT에 해당한다.

가. 철수가(AGT) 영희에게 빵을 먹였다.  
나. 형은(AGT) 장난감을 뺏어 동생을 울렸다.

#### [주의]

- (1) 장형 사동문은 복문이므로, 사동주는 주절의 AGT이고, 피사동주는 내포절의 AGT이다.

가. 철수는(AGT) [영희가(AGT) 집에(GOL) 오게] 했다.  
나. 철수는(AGT) [영희가(AGT) 빵을(THM) 먹게] 했다.

- (2) 단형 사동문의 사동주는 AGT로 분석하고, 피사동주는 GOL로 분석한다.

가. 철수는(AGT) 지프차에(GOL) 가족들을(THM) 태웠다.  
나. 철수가(AGT) 영희에게(GOL) 빵을(THM) 먹였다.

- (3) 의인화 등 비유적 표현으로 AGT에 의도성을 가질 수 없는 무정물이 오더라도 작업자의 판단에 따라 유정물에 준한다고 보면 AGT를 부여할 수 있다.

가. 달은(AGT) 내게 어서 오라는 듯이 손짓했다.

나. 골목길에서 갑자기 오토바이가(AGT) 튀어나왔다.  
 다. 바람은(AGT) 노래하고, 꽃은(AGT) 춤을 추었다.

- (4) 형용사문의 주어와 작용을 나타내는 동사 서술어의 주어를 AGT로 분석하지 않도록 주의한다. 이러한 서술어의 주어는 의도성을 갖추지 못하므로 THM을 부여하여야 한다. [지침 6.1.의 (1-2)를 참고한다.]

가. 사람이(THM) 많다.  
 나. 사람이(THM) 많아졌다.  
 다. 철수가(THM) 맡은 일에 충실하다.  
 라. 주전자의 물이(THM) 끓는다.

## 4.2. EXP (Experiencer, 경험주)

### [정의]

- 서술어가 감정적, 심리적, 지각적, 인지적 경험을 나타낼 때 그러한 경험을 하는 주체.

### [예시]

- (1) 인지, 지각, 감정의 주체가 되는 논항의 의미역이 이에 해당한다. 주로 '이/가', 보조사 '은/는'으로 실현된다.

가. 철수는(EXP) 영희가 무서웠다.(감정적/심리적 경험의 주체)  
 나. 철수는(EXP) 영희를 보았다. (지각적 경험의 주체)  
 다. 철수는(EXP) 영희를 잘 안다. (인지적 경험의 주체)

- (2) 의도성이 없는 심리 형용사의 주체가 되는 논항의 의미역을 EXP로 처리한다.

가. 철수는(EXP) 영희가 좋다.  
 나. 철수는(EXP) 영희가 반가웠다.  
 다. 그는(EXP) 할아버지의 죽음이 슬펐다.

### [주의]

- (1) 맥락을 살펴 주체의 의도성이 확인되지 않을 때는 EXP, 의도성이 확인될 때는 MAG로 분석한다.  
 심리 형용사는 EXP로 분석하고, 심리 동사는 MAG로 분석한다.

가. 많은 행인들이(EXP) 그 광경을 보았다. → 의도치 않게 우연히 본 경우  
 나. 많은 행인들이(MAG) 그 광경을 보았다. → 의도성을 가지고 자리에 멈춰서 본 경우  
 다. 철수는(EXP) 오늘 오전에 총소리를 들었다. → 의도치 않게 우연히 들은 경우  
 라. 철수는(MAG) 그 연설을 경청했다. → 의도성을 가지고 들은 경우

- (2) 단, 주체의 의도성 여부만으로 EXP와 MAG의 판별이 어려울 경우 EXP로 분석한다.

가. 철수는(EXP) 오늘 오전에 총소리를 '우연히' 들었다. → 의도성 판별에 따라 EXP로 분석.  
 나. 철수는(EXP) 뉴스 소리를 들었다. → 의도성 판별이 어려울 경우 EXP로 분석.

다. 철수는(MAG) 뉴스 소리를 '유심히' 들었다. → 의도성 판별에 따라 MAG로 분석.

- (3) 일부 동사가 주체의 내적 상태를 나타내는 것으로 보아 임의적으로 EXP를 주지 않도록 한다. 또한 '경험, 시도'를 나타내는 일부 보조 용언과 결합한 것에 이끌려 EXP로 분석하지 않도록 주의한다.

가. 나는(AGT) 터져 나오는 울음을 겨우 참았다.  
나. 나는(AGT) 선생님의 의견에 동의했다.  
다. 나는(AGT) 술한 좌절을 무수히 경험했다.  
라. 나도(AGT) 해외 여행을 여러 번 해 봤다.

### 4.3. MAG (Mental agent, 심리행위주)

#### [정의]

- 심리 동사, 지각 동사 등의 서술어가 심리적 경험을 나타낼 때 그러한 경험을 하는 주체.

#### [예시]

- 심리동사와 심리형용사에 '어하다'가 결합된 서술어의 주체를 MAG로 처리한다.

가. 철수는(MAG) 영화를 좋아했다.  
나. 철수는(MAG) 영화를 반가워했다.  
다. 그는(MAG) 할아버지의 죽음을 슬퍼했다.

#### [주의]

- (1) 맥락을 살펴 주체의 의도성이 확인되지 않을 때는 EXP, 의도성이 확인될 때는 MAG로 분석한다.  
심리 형용사는 EXP로 분석하고, 심리 동사는 MAG로 분석한다.

가. 많은 행인들이(EXP) 그 광경을 보았다. → 의도치 않게 우연히 본 경우  
나. 많은 행인들이(MAG) 그 광경을 보았다. → 의도성을 가지고 자리에 멈춰서 본 경우  
다. 철수는(EXP) 오늘 오전에 총소리를 들었다. → 의도치 않게 우연히 들은 경우  
라. 철수는(MAG) 그 연설을 경청했다. → 의도성을 가지고 들은 경우

- (2) 단, 주체의 의도성 여부만으로 EXP와 MAG의 판별이 어려울 경우 EXP로 분석한다.

가. 철수는(EXP) 오늘 오전에 총소리를 '우연히' 들었다. → 의도성 판별에 따라 EXP로 분석.  
나. 철수는(EXP) 뉴스 소리를 들었다. → 의도성 판별이 어려울 경우 EXP로 분석.  
다. 철수는(MAG) 뉴스 소리를 '유심히' 들었다. → 의도성 판별에 따라 MAG로 분석.

### 4.4. EFF (Effector, 영향주)

#### [정의]

- 서술어가 나타내는 사건(행위, 상태)을 비의도적으로 유발하는 주체.

[예시]

- 비의도적인 자연 현상, 무정물이 영향주로 많이 쓰인다. 주로 '에'로 실현되지만 '이/가'로 실현되기도 한다.

가. 철수가 전봇대에(EFF) 부딪혔다.  
나. 나는 철수에게(EFF) 한눈에 반했다.  
다. 태풍이(EFF) 한반도에 더운 공기를 몰고 왔다.  
라. 신학문이(EFF) 조선 여성들을 새롭게 개화했다.

[주의]

- (1) 인물의 경우에 한해, 맥락을 살펴 주체의 의도성이 확인되지 않을 때는 EFF, 의도성이 확인될 때는 AGT로 분석한다.

가. 철수가(EFF) 나를 실망시켰다.  
→ 의도성이 없는 경우이면 '나는 철수에게(EFF) 실망했다'와 같이 처리.  
나. 철수가(AGT) 나를 실망시켰다.  
→ 의도성을 가진 경우 AGT로 처리.

- (2) 비유적인 표현에 한해, 의도성을 가지지 않는 주체여서 EFF가 아닌 THM으로 판단되는 것들이 있다. 그런데 동일 절 내에서 이미 분명한 THM이 있다면 이 의도성을 가지지 않는 주체를 EFF로 분석한다.

가. 자갈이(EFF) 상당한 면적을(THM) 점하고 있다.  
나. 정수기가(EFF) 수돗물을(THM) 1급수로 여과한다.

## 4.5. THM (Theme, 대상)

[정의]

- 행위나 과정의 영향을 받지만 그 과정을 지배하지는 못하는 참여자. 즉 위치, 상태가 바뀌거나 주어진 위치나 상태에 있는 참여자.

[예시]

- (1) 형용사문에서 어떤 위치, 상태에 놓여 있거나 위치, 상태 변화를 겪는 참여자.

가. 철수는(THM) 아주 영리하다.  
나. 백화점 옷이(THM) 매우 비싸다.  
다. 플라스틱 타는 냄새가(THM) 독하다.

- (2) 자동사문에서 어떤 위치, 상태에 놓여 있거나 위치, 상태 변화를 겪는 참여자.

가. 벽지에 곰팡이가(THM) 피다.  
나. 가상 화폐 가격이(THM) 연일 폭등했다.  
다. 나무토막이(THM) 하류로 떠내려왔다.

- (3) 피동문에서 어떤 위치, 상태에 놓여 있거나 위치, 상태 변화를 겪는 참여자. 피동문은 아니지만, 피동의 의미를 갖는 자동사문의 주어도 THM으로 분석한다.

가. 도둑이(THM) 경찰에게 잡혔다.  
 나. 노인용 실버폰이(THM) 개발되었다.  
 다. 신축 아파트가(THM) 비싼 값에 거래되었다.  
 라. 철수가(THM) 영희에게 맞았다.

- (4) 타동사문에서 행위의 대상이 되는 참여자. 주로 목적어에 해당하는 성분이다.

가. 철수는 영희를(THM) 끌어안았다.  
 나. 아버지가 할머니를(THM) 도와드렸다.  
 다. 형사는 차분히 범인의 얼굴을(THM) 찾았다.

## 4.6. COM (Companion, 동반주)

### [정의]

- 독립적으로 역할을 하기보다는 주로 행위주나 대상을 보조하여 그것과 같은 역할을 하는 참여자.

### [예시]

- (1) '와/과'와 결합하여 주어와 함께 서술어의 행위를 하는 참여자를 COM으로 분석한다.

가. 철수는 영희와(COM) 싸웠다.  
 나. 난류가 한류와(COM) 만나는 곳이다.

- (2) '와/과'와 결합하여 목적어와 함께 서술어의 행위를 하는 참여자도 COM으로 분석한다.

가. 철수는 모래를 물과(COM) 섞었다.  
 나. 철수는 영희를 정희와(COM) 혼동했다.

### [주의]

- (1) 서술어의 의미가 상호적이지 않은 경우는 부가어로 처리한다.

가. 철수는 영희와(M-COM) 극장에 갔다.  
 나. 철수는 영희를 희정이와(M-COM) 집에서 내보냈다.

- (2) 어순을 살펴서 논항이 아닌 것을 논항으로 처리하는 일이 없도록 한다.

- 철수는 물과 모래를(THM) 섞었다.

- (3) '같이', '함께'가 동반의 의미가 있기는 하나 참여자 혹은 대상(명사구)가 없이 쓰인 경우까지도 COM으로 처리하지 않도록 주의한다. 동반주는 어디까지나 주어와 관련한 참여자 혹은 목적어와 관련한 대상에 국한해 부여하는 의미역이다.

- 그들은 어려서부터 **같이(M-MNR)** 싸웠다.

## 4.7. GOL (Goal, 도착점)

### [정의]

- 서술어가 표현하는 사건이 물리적 이동을 포함하고 있을 경우 그 끝점, 추상적인 행위나 태도의 의미를 포함하고 있을 경우 그 지향점

### [예시]

- (1) 이동동사의 논항인 경우가 가장 전형이다. ‘에’, ‘에게’ 등이 결합된 조사 결합 구성을 GOL로 분석한다.

가. 철수가 **학교에(GOL)** 갔다.  
나. 나는 **창가에(GOL)** 화분을 놓았다.  
다. 철수는 고향에 계신 **부모님께(GOL)** 편지했다.  
라. 주인은 **내게(GOL)** 음식을 권했다.  
마. 아버지는 **딸에게(GOL)** 외출을 금지시켰다.  
바. 정부는 **피해자들에 대해(GOL)** 진상 규명을 약속했다.

- (2) 발화동사의 청자도 GOL로 분석한다.

가. 철수는 **영희에게(GOL)** 미안하다고 말했다.  
나. 철수가 **영희에게(GOL)** 사실을 말했다.  
다. 철수가 **영희에게(GOL)** 속삭였다.

### [주의]

- (1) 구체적/추상적인 변화가 방향성을 가질 때는 DIR로 분석해야 하므로 주의해야 한다. 즉, ‘(으)로’가 결합된 조사 결합 구성은 GOL이 아닌, DIR로 분석한다.

가. 아버지는 아이를 **서울에(GOL)** 보냈다.  
나. 아버지는 아들을 **서울로(DIR)** 보냈다.

- (2) ‘주다’류와 같은 수혜 동사 구문에서, 특히 ‘이/가’ 조사 결합 구성을 다른 논항으로 분석하지 않도록 주의한다.

가. 철수가 **영희에게(GOL)** 선물을 주었다.  
나. **영희가(GOL)** 철수에게서(SRC) 선물을(THM) 받았다.

- (3) ‘에’ 조사 결합 구성에서 서술어만으로 GOL과 LOC를 구별하기 어려운 경우에는 ‘로’ 조사 결합 구성으로 바꾸어 수용성이 떨어지면 LOC으로 처리하고, 수용성에 문제가 없으면 GOL로 처리한다. 그러나 이 판별 방식이 ‘로’ 조사 결합 구성을 GOL으로 처리해도 좋다는 의미는 아니다. ‘로’ 조사 결

합 구성이 방향성을 가지면 DIR로 처리한다.

가. 철수는 학교에/학교로 서류를 보냈다. → '로'가 성립하므로 '학교에'는 GOL임.  
나. 나는 강물에/강물로 돌을 던졌다. → '로'가 성립하므로 '강물에'는 GOL임.  
다. 하늘에/하늘로 별이 반짝인다. → '로'가 성립할 수 없으므로 '하늘에'는 LOC임.  
라. 나는 벽에/벽으로 페인트를 칠했다. → '로'가 성립할 수 없으므로 '벽에'는 LOC임.  
마. 그의 눈길에/그의 눈길로 애정이 담겼다. → '로'가 성립할 수 없으므로 '그의 눈길에'는 LOC임.

## 4.8. LOC (Location, 장소)

### [정의]

- 서술어가 나타내는 사건이 일어나는 공간적, 추상적 위치를 가리키는 논항.

### [예시]

- (1) 다음과 같이 '에, 에서'가 결합된 조사 결합 구성이 공간적 위치로 해석될 경우 LOC로 분석한다.

가. 광혜원은 서울 재동에(LOC) 설립되었다.  
나. 형사는 은밀하게 범죄 현장에(LOC) 잠입했다.  
다. 영화는 지금 친정에(LOC) 있다.

- (2) 다음과 같이 '에, 에서'가 결합된 조사 결합 구성이 추상적 위치로 해석될 경우에도 LOC로 분석한다.

가. 철수는 곤경에(LOC) 처해 있다.  
나. 팬들의 관심사는 가수의 은퇴 여부에(LOC) 모였다.  
다. 백화점이 파격 할인에(LOC) 돌입했다.

### [주의]

- (1) 장소어, 집단어+'에서' 구성은 장소어, 집단어+'이/가' 구성으로 변환해 보아서 수용성이 떨어지면 LOC로 분석하고, 주어일 경우에는 AGT로 분석한다.

가. 한국에서(LOC) 김치를 잘 먹는다.  
cf. \*한국이 김치를 잘 먹는다.  
나. 구청에서(AGT) 식당에 대한 위생 점검에 나섰다.  
cf. 구청이 식당에 대한 위생 점검에 나섰다.

- (2) 다음처럼 논항이 아니라 부가어인 경우와 혼동하지 않도록 한다.

가. 눈이 내리자 공원에(M-LOC) 길이 열었다.  
나. 우리는 올해 집에서(M-LOC) 설을 쇠었다.



## 4.9. SRC (Source, 출발점)

### [정의]

- 서술어가 표현하는 사건이 위치, 상태 변화를 나타낼 때 그 변화의 출발점.

### [예시]

- (1) 서술어가 이동의 의미를 포함하고 있는 경우, '로부터'가 결합된 조사 결합 구성을 SRC로 분석한다.

가. 이 기단은 서쪽으로부터(SRC) 왔다. (자동사문)  
나. 정부가 옥수수를 미국으로부터(SRC) 수입했다. (타동사문)

- (2) 서술어가 변화의 의미를 포함하고 있는 경우, '에서', '에게서'가 결합된 조사 결합 구성을 SRC로 분석한다.

가. 이 물질은 고체에서(SRC) 액체로 변했다. (자동사문)  
나. 영희가 철수에게서(SRC) 선물을 받았다. (타동사문)

### [주의]

- '~에서 ~까지' 구성이 서술어의 논항으로 나타나는 경우, 서술어의 어휘의미 번호를 '888'로 바꾼 후 '~에서'와 '~까지'를 각각 SRC와 GOL로 분석한다.

- 그럼 잠실에서(SRC) 구리까지(GOL) 지하철이 연결이 되지. (연결되다888)

## 4.10. CNT(Content, 내용)

### [정의]

- 발화 동사, 사유 동사 등이 지시하는 내용. 대부분 절 논항으로 구성되어 있다.

### [예시]

- (1) 직접 인용의 '(이)라고', 간접 인용의 '고'와 인용절 전체를 CNT로 분석한다.

가. 언니는 우리에게 "어서 집으로 돌아와!"라고(CNT) 소리쳤다.  
나. 철수는 영희에게 미안하다고(CNT) 말했다.  
다. 당선자는 언제나 국민의 편에 서겠다(CNT) 다짐했다.

- (2) 인용 표지로 쓰이는 '하고'와 인용절 전체를 CNT로 분석한다.

가. 포탄이 터지는 소리가 "쿵!"하고(CNT) 울렸다.  
나. 영희는 "불이야!"하고(CNT) 외쳤다.  
→ '하고'를 포함한 인용절을 CNT로 분석하되 이때의 '하고'(하다032)는 대상 서술어로 삼지 않는다. 따라서 "쿵!", "불이야!"에 의미역을 부여하지 않는다.

- (3) ‘그렇게 말하다’라는 의미의 ‘그러다’가 발화 동사 대신 쓰인 경우에도 인용절을 CNT로 분석한다.

가. 선생님이 이거 주라고(CNT) 그랬다.  
나. 영희가 주변을 잘 살펴 보라고(CNT) 그랬어.

#### [주의]

- (1) 서술어 ‘하다’가 ‘이르거나 말하다’의 용법으로 쓰인 경우 그 내용을 CNT로 분석한다.

가. 어떤 철학자는 “시간은 금이다.”라고(CNT) 했다.  
나. 친구가 영화 구경 가자고(CNT) 했다.  
다. 그가 거짓말을 했다고(CNT) 하는 증거가 있다.  
라. 철수가 “거짓말 치지 마!”(CNT) 했다.

- (2) 특정 서술어가 인용절을 요구함에도 불구하고 격틀 정보로 표시되지 않은 다음과 같은 경우에는 인용절 전체를 CNT로 분석한다.

가. 그녀가 남긴 쪽지에는 처가에 있겠다고(CNT) 적혀 있었다.  
나. 학교 정문에 “축! 입학!”이라고(CNT) 쓰인 플래카드가 걸렸다.  
다. 화면에 “GPS를 수신합니다.”라고(CNT) 떴다.

- (3) 특정 서술어가 단독으로는 인용절을 요구하지 않지만 그 서술어가 쓰인 구성이 인용절을 요구하는 경우 인용절 전체를 CNT로 분석한다.

- 그래서 아내가 예술가형이 맞았구나라고(CNT) 생각이 들더라고요.  
→ 어근이 분리된 ‘생각이 들다’의 ‘들다’는 인용절을 요구하지 않는다. 이런 경우 지침 2.1.1.에 따라 어근이 분리되지 않은 ‘생각들다’를 상정하여 격틀 구조를 살핀 후 인용절 논항에 CNT를 부여할 수 있다. (‘생각들다’는 사전에 등재되어 있지 않으므로 지침 2.1.1.에 따라 이와 유사한 서술어인 ‘생각되다’를 고려할 수 있다. ‘생각되다’는 인용절을 요구한다.)

- (4) ‘-는가 하면’ 구성의 하다056의 선행 성분도 CNT로 분석한다.

ㄱ. 6시 땡 하고 가야 된다라고 하는 성향이 있는가(CNT) 하면은  
ㄴ. 규제를 완화하고 계약금 비율을 낮추는가(CNT) 하면

- (5) CNT의 범위를 임의적으로 확대하거나 CNT가 아닌 것들을 CNT로 분석하지 않도록 주의한다.

ㄱ. 서울에 홍길동이라고 있다. → 논항이 아니므로 의미역을 부여하지 않는다.  
ㄴ. 알바 같은 일을 하려고(M-PRP) 생각을 하거든.  
ㄷ. 해외 여행 매출이 27배 급증한 것으로(FNS) 집계됐다.  
ㄹ. 신인이라고 해서(M-CAU) 무작정 감싸줘서는 안 된다.[‘생각하다’로 해석되는 하다023]  
cf) 그는 자기가 신인이라고(CNT) 했다. 그러나 신인이라고(M-CAU) 무작정 감싸줘서는 안 된다.

- (6) 여러 개의 문장이 인용될 때에는 다음과 같이 처리한다.

첫째, 둘 이상의 인용절이 분리되어 제시되는 경우

ㄱ. 스웨덴 잼버리 참가자 토르 윌러스(16. 남) 대원은 “대학 기숙사가 정말 쾌적하고 좋다”고(CNT) 하며(M-LST) “한국에서의 좋은 추억을 남길 수 있을 것 같아 기대된다”고(CNT) [말했다.]

ㄴ. 스웨덴 잼버리 참가자 토르 윌러스(16. 남) 대원은 “대학 기숙사가 정말 쾌적하고 좋다”며(M-LST) “한국에서의 좋은 추억을 남길 수 있을 것 같아 기대된다”고(CNT) [말했다.]

→ ㄱ.은 기존의 지침대로 처리하면 된다. 다만 ㄴ.의 선행절은 ‘좋다고 하며’가 축약(인용동사 ‘하-’가 생략)된 것으로 보아 선행하는 인용절은 M-LST로, 후행하는 인용절을 CNT로 처리한다.

둘째, 둘 이상의 인용절이 각각의 따옴표로 나열된 경우

ㄷ. 이를 두고 ‘자속하는 것 맞냐’, ‘아직도 정신을 못 차렸다’ 등 팬들의 비난이 이어졌고 이후 종아요를 취소했다.

[선행절이 후행절의 서술어 ‘차리다’에 의존함.]

→ 선행절은 후행절의 서술어인 ‘차리다’의 논항도 부가어도 아니다. 이처럼 단순히 문장이 나열된 경우에 서술어와 관련되지 않으면 의미역을 분석하지 않는다. [‘맞냐’ 행에 대한 의미역을 분석하지 않는다는 뜻이다.]

셋째, 둘 이상의 인용절이 하나의 따옴표 안에서 나열된 경우

ㄹ. 이영수 대행은 “[1]오늘 경기까지만 하고 그만하겠다고 구단에 미리 이야기를 했다. [2]오늘이 마지막 경기라고 생각한다. [3]선수들은 아직 모른다”라고(CNT) 설명했다.

→ [1]의 서술어 ‘했다’와 [2]의 서술어 ‘생각한다’는 [3]의 서술어 ‘모른다’에 의존하게 분석되어 있다. 구문 분석 지침에 따르면 모든 어절은 다른 어절에 의존하게 되어 있기 때문이다. 그러나 실제로는 [1]과 [2]는 [3]의 서술어 ‘모른다’의 논항도 부가어도 아니므로 이에 대한 의미역은 분석하지 않는다. [‘했다’ 행과 ‘생각한다’ 행에 대한 의미역을 분석하지 않는다는 뜻이다.]

## 4.11. INS(Instrument, 도구)

### [정의]

- 서술어가 나타내는 사건에 사용되는 물리적 도구나 사물, 재료의 역할을 하는 논항. 방법, 방식, 도구 등 추상적인 경우도 포함한다.

### [예시]

- (1) 다음과 같이 ‘로’가 결합된 조사 결합 구성은 INS로 분석한다. 이때, ‘~을 이용하여’로 치환하여 의미가 통하는지 여부가 분석에 도움이 된다.

가. 이 물질은 **3개의 원소로(INS)** 이루어졌다.  
나. 그는 소포들을 **노끈으로(INS)** 감았다.  
다. 나는 부패를 막기 위해 실험체를 **알코올로(INS)** 처리했다.

- (2) 어떤 사태가 성립되는 데 도구의 역할을 하는 참여자도 INS로 분석한다.

- 대합실이 **사람들로(INS)** 가득하다.  
cf. 대합실에 **사람들이(THM)** 가득하다.

[주의]

- 다음처럼 논항이 아니라 부가어인 경우와 혼동하지 않도록 한다.

가. 철수는 연필로(M-INS) 초상화를 그렸다.  
나. 나는 인터넷 뱅킹으로(M-INS) 아파트 관리비를 이체했다.

## 4.12. CRT(Criterion, 기준)

[정의]

- 서술어에 대해 평가되는 대상 혹은 판단하는 대상을 비교하는 논항.

[예시]

- (1) 서술어가 비교의 의미를 가지는 문장의 '와' 조사 결합 구성일 경우 CRT로 분석한다.

가. 철수는 영희와(CRT) 비슷하다.  
나. 철수는 영희와(CRT) 성격이 다르다.

- (2) 비교의 의미를 가지는 '이/가', '에/에서', '를' 조사 결합 구성과 보조사 '만, 보다, 처럼, 만큼' 등의 조사 결합 구성을 CRT로 분석한다.

가. 이 수박은 10kg이(CRT) 나간다.  
나. 우리나라의 기술 수준은 다른 나라에(CRT) 앞선다.  
다. 김 후보의 지지율은 1%에(CRT) 불과하다.  
라. 철수는 영희만(CRT) 못하다.  
마. 철수가 영희보다(CRT) 낫다.

[주의]

- 다음처럼 논항이 아니라 부가어인 경우와 혼동하지 않도록 한다.

가. 철수는 영희보다(M-CRT) 똑똑하다.  
나. 고속버스가 기차만큼(M-CRT) 빠르다.

## 4.13. FNS(Final state, 결과상태)

[정의]

- 서술어가 (인물의 자격 또는) 물질의 성질이나 용도의 변화를 나타낼 때 변화의 결과를 나타내는 논항

[예시]

- (1) 다음과 같이 '로'가 결합된 조사 결합 구성이 대부분이지만, '이/가' 조사 결합도 가능하다.

가. 향아리의 물이 포도주가(FNS) 되었다.

cf. 향아리의 물이 포도주로(FNS) 되었다.  
 나. 감독이 골키퍼를 김병지로(FNS) 바꾸었다.  
 다. 우리 업체는 인삼을 당의정으로(FNS) 가공한다.  
 라. 이 후보가 타 후보에 비해 높은 지지율을 보일 것으로(FNS) 예상됐다/조사됐다.  
 마. 영업 손실이(THM) 최저를(FNS) 기록했다.

(2) 인물의 자격이라 하더라도 그것이 변화의 종착점인 경우 FNS로 분석한다.

가. 철수는 투표를 통해 반장으로(FNS) 선출되었다.  
 나. 그는 각고의 노력 끝에 부사장으로(FNS) 승진했다.  
 다. 영희는 철수를 자신의 후임으로(FNS) 밀었다.

#### [주의]

- 창조동사 구문의 목적어 논항은 FNS 가 아니라 THM 으로 분석한다.

- 그는 밀가루로(INS) 빵을(THM) 만들었다.  
 cf. 그는 밀가루를(THM) 빵으로(FNS) 만들었다.

## 4.14. DIR(Direction, 방향)

#### [정의]

- 어떤 사태가 일어날 때 그 사태가 향하고 있는 방향으로 해석되는 논항.

#### [예시]

(1) 서술어가 이동의 의미를 가질 때 방향에 해당하는 '로' 조사 결합 구성을 DIR로 분석한다.

가. 경찰차가 집 앞으로(DIR) 지나갔다.  
 나. 동생은 터미널을 떠나 서울로(DIR) 출발했다.  
 다. 눈을 뜨자마자 금세 학교로(DIR) 달려갔다.

(2) '오른쪽', '왼쪽', '위쪽', '아래쪽', '앞으로', '뒤로', '동서남북'과 같이 방향을 의미하는 명사구나 명사 '방향'을 사용한 구성도 DIR로 분석한다.

가. 그녀의 시선이 아래쪽으로(DIR) 향했다.  
 나. 그는 갑자기 오른쪽으로(DIR) 돌았다.  
 다. 철새는 겨울이면 남쪽으로(DIR) 이동한다.

#### [주의]

(1) 구체적/추상적인 변화가 방향성을 가질 때는 DIR로 분석해야 하므로 주의해야 한다. '로'가 결합된 조사 결합 구성은 GOL이 아닌, DIR로 분석한다.

가. 아버지는 아이를 서울에(GOL) 보냈다.  
 나. 아버지는 아들을 서울로(DIR) 보냈다.

(2) 부가어를 논항으로 분석하지 않도록 주의한다.

가. 철수는 **북쪽으로(M-DIR)** 노를 저었다.  
나. 철수는 운동장을 **좌우로(M-DIR)** 뛰었다.

## 4.15. QLF(Qualification, 자격)

### [정의]

- 서술어가 나타내는 대상에 대하여 인물의 자격(신분/지위)이나 사물의 자격으로 해석되는 논항.

### [예시]

- (1) '로(서)'가 결합된 조사 결합 구성이 인물의 자격이나 사물의 자격으로 해석되는 논항인 경우 QLF로 분석한다.

가. 아인슈타인은 **과학자로(QLF)** 유명하다.  
나. 그는 한인 사회에서 **소식통으로(QLF)** 통한다.  
다. 철수는 군에서 **군의학관으로(QLF)** 복무했다.

### [주의]

- (1) 변화의 종착점인 경우 QLF가 아닌 FNS로 분석한다.

가. 철수는 **회장으로(FNS)** 선출되었다.  
나. 선생님은 영화를 **머느리로(FNS)** 삼았다.

- (2) 부가어를 논항으로 분석하지 않도록 주의한다. 다음 예문은 부가어 자격에 해당한다.

가. 철수는 오늘 미팅에 **자문위원으로(M-QLF)** 참석했다.  
나. 그 가수는 콘서트에 **특별 게스트로(M-QLF)** 출연했다.

## 4.16. MNR(Manner, 방식)

### [정의]

- 어떤 사태가 일어나는 방식을 나타내는 논항.

### [예시]

- (1) MNR을 나타내는 성분은 대개 부사어로 실현되지만 '굴다', '생기다', '보이다'와 같은 서술어에서는 논항으로 분석되는데, 이를 MNR로 분석한다. '로' 조사 결합 구성 외에 '명사+처럼/같이'와 같은 구성으로도 실현된다.

가. 철수는 **아이같이(MNR)** 군다.  
나. 영화는 **이국적으로(MNR)** 생겼다.  
다. 철수의 외모는 **40대처럼(MNR)** 보인다.

- (2) ‘굴다’, ‘보이다’, ‘생각하다’, ‘여기다’ 등의 서술어는 ‘-게’ 부사절을 취할 수 있고, 이를 MNR로 분석한다.

- 철수는 어리석게(MNR) 군다.

- (3) 그러나 ‘굴다’와 비슷한 ‘삼다’류 서술어는 FNS로 분석한다.

- 선생님은 철수를 사위로(FNS) 삼았다.

## 4.17. PRT(Part, 부분)

### [정의]

- 주격중출문에서 경험주의 신체 일부를 나타내는 논항.

### [예시]

- (1) 형용사를 서술어로 하는 주격중출문에서 첫 번째 명사구의 신체 일부, 감각 또는 감정으로 쓰인 두 번째 명사구는 PRT로 분석한다.

가. 철수는 키가(PRT) 크다.  
나. 영희는 피부가(PRT) 가려웠다.  
다. 현우는 마음이(PRT) 언짢았다.

- (2) 자동사를 서술어로 하는 주격중출문에서 첫 번째 명사구의 신체 일부, 신체 현상으로 쓰인 두 번째 명사구는 PRT로 분석한다.

가. 철수는 가슴이(PRT) 뒹다. (신체 부분의 변화)  
나. 영희는 목이(PRT) 메었다. (신체 부분의 변화)  
다. 선생님이 눈빛이(PRT) 빛났다. (기색의 변화)

- (3) 주격중출문에서 두 번째 명사구가 ‘신체 일부’인 PRT라면 첫 번째 명사구는 THM로 분석한다.

- 철수는(THM) 키가(PRT) 크다.

- (4) 주격중출문에서 두 번째 명사구가 ‘감각 또는 감정’인 PRT라면 첫 번째 명사구는 EXP로 분석한다.

가. 영희는(EXP) 피부가(PRT) 가려웠다.  
나. 현우는(EXP) 마음이(PRT) 언짢았다.

## 4.18. FTR(Feature, 속성)

### [정의]

- 주격중출문에서 대상의 속성을 나타내는 논항.

#### [예시]

- (1) 형용사를 서술어로 하는 주격중출문에서 첫 번째 명사구의 속성으로 쓰인 두 번째 명사구는 FTR로 분석한다.

가. 이 신제품은 가격이(FTR) 저렴하다.  
 나. 이번 시험은 범위가(FTR) 많다.  
 다. 이 자세는 난이도가(FTR) 적당하다.  
 라. 깃털은 무게가(FTR) 가볍다.  
 마. 이 활동은 봉사의 의미가(FTR) 있다/없다.

- (2) 주격중출문에서 두 번째 명사구가 FTR인 경우 첫 번째 명사구는 THM으로 분석한다.

가. 철수는(THM) 성격이(FTR) 모질다.  
 나. 이 운동화는(THM) 색깔이(FTR) 빨강다.  
 다. 이 체육관은(THM) 규모가(FTR) 크다.

## 5. 부가 의미역 분석의 실제

### 5.1. M-TMP (Temporal, 시간)

#### [정의]

- 서술어나 사건과 관계된 시간을 나타내는 부가어를 TMP 로 분석한다. 즉 명확한 날짜, 시기, 시대 혹은 시점, 기간, 빈도의 의미를 가지는 부가어를 TMP 로 분석한다.

#### [예시]

- (1) ‘당시, 현재’와 같이 시점을 나타내는 시간부사, ‘오래, 잠깐’과 같이 기간을 나타내는 시간부사, ‘가끔, 종종’과 같이 빈도를 나타내는 시간부사는 M-TMP로 분석한다.

가. 그는 현재(M-TMP) 군대에 입대했다. [시점]  
 나. 당시(M-TMP) 나에게서 여러 가지 불만찬 책이 많았다. [시점]  
 다. 병원에 감기 환자가 많아 내 차례를 오래(M-TMP) 기다렸다. [기간]  
 라. 나는 서울에 간 김에 이모 집에 잠깐(M-TMP) 들렀다. [기간]  
 마. 나는 힘든 일이 있으면 가끔(M-TMP) 술을 마시곤 한다. [빈도]  
 바. 나는 종종(M-TMP) 집 앞 공원으로 산책을 나간다. [빈도]

- (2) ‘에, 부터, 까지’ 등이 결합된 조사 결합 구성이 이에 해당된다.

가. 해안가라 이 지역에는 봄에(M-TMP) 안개가 정말 많이 낀다.  
 나. 1970년에(M-TMP) 우리 대학은 언어연구소를 개소하였다.  
 다. 나는 설날에(M-TMP) 웃어른들께 세배했다.



라. 지금부터(M-TMP) 시험을 시작하겠습니다.  
마. 1618년부터 1648년까지(M-TMP) 종교전쟁이 벌어졌다.

(3) '때, 직후, 시기, 동안' 등이 결합된 명사구도 M-TMP로 분석한다.

가. 나는 버스 탈 때(M-TMP) 지갑을 두고 왔다는 것을 알았다.  
나. 일제강점기 때(M-TMP) 우리 민족은 교육으로부터 철저히 소외되었다.  
다. 사고 직후(M-TMP) 목격자부터 찾아야 한다.  
라. 나는 너를 알게 된 10년 동안(M-TMP) 행복했다.  
마. 파편에 맞은 상처가 여름 내내(M-TMP) 진물을 내며 끓었다.

(4) '-자'와 같이 시간 표현 연결어미가 사용된 부사절도 M-TMP로 분석한다.

가. 봄이 되자(M-TMP) 공원에 꽃이 만발했다.  
나. 사장이 나오자(M-TMP) 직원들이 한쪽으로 비켜섰다.  
다. 일이 해결되자(M-TMP) 사장은 직원들에게 밀린 월급을 몰아주었다.

#### [주의]

- '-(으)면'이 사용된 조건 표현 부사절은 M-CND 로 분석한다.

가. 봄이 오면(M-CND) 꽃이 핀다.  
나. 우리 가족은 추석이면(M-CND) 항상 할머니 댁에 간다.

## 5.2. M-LOC (Location, 장소)

#### [정의]

- 서술어가 나타내는 사건이 일어나는 공간을 가리키는 부가어를 M-LOC 로 분석한다.

#### [예시]

(1) 다음과 같이 '에, 에서'가 결합된 조사 결합 구성은 M-LOC로 분석한다. 서술어의 의미에 이동성이 없어야 한다.

가. 봄이 되자 공원에(M-LOC) 꽃이 만발했다.  
나. 우리는 올해 집에서(M-LOC) 설을 쇠었다.

단, '에, 에서'가 결합한 구성은 논항일 가능성이 있으므로 논항을 부가어로 분석하지 않도록 주의한다. 다음 예문은 처소 논항이 LOC로 정의된 필수 논항이다.

가. 광혜원은 서울 재동에(LOC) 설립되었다.  
나. 강우규는 지란에(LOC) 동광 학교를 세웠다.  
다. 하이든은 현악 4중주 제2악장에(LOC) 국가를 사용했다.

- (2) 장소어, 집단어+‘에서’ 구성은 장소어, 집단어+‘이/가’ 구성으로 변환해 보아서 수용성이 떨어지면 M-LOC로 분석한다.

가. 터키에서(M-LOC) 페즈를 자주 쓴다.  
cf. \*터키가 페즈를 자주 쓴다.  
나. 로마에서(M-LOC) 호민관을 선출했다.  
cf. \*로마가 호민관을 선출했다.  
다. 영어권에서는(M-LOC) 사람들이 스테인리스강을 스테인리스스틸이라고 표기한다.  
cf. \*영어권이 사람들이 스테인리스강을 스테인리스스틸이라고 표기한다.

단, ‘장소어, 집단어+-이/가’ 구성으로 변환하여 주어로 판별이 되면 AGT로 분석한다.

- 환경부에서(AGT) 지난 달 현지 조사 때 뒤늦게 이 사실을 알았다.  
cf. 환경부가 지난 달 현지 조사 때 뒤늦게 이 사실을 알았다.

- (3) ‘장소어+-에서’ 구성이 SRC로 의심될 때에는 ‘-에서’를 ‘-(으)로부터’로 대체해 보아 불가능하면 M-LOC로 분석한다.

가. 헨리 헬릿 데일이 심장 조직에서(M-LOC) 아세틸콜린을 처음 발견하였다.  
cf. \*헨리 헬릿 데일이 심장 조직으로부터 아세틸콜린을 처음 발견하였다.  
나. 우리는 세상을 산 위에서(M-LOC) 본다.  
cf. \*우리는 세상을 산 위로부터 본다.

- (4) ‘N마다’ 구성에서 N이 장소명사인 경우 M-LOC로 분석한다.

- 폭설로 마을이 고립되어 가게마다(LOC) 생필품이 동났다.

단, ‘N마다’ 구성이 시간 표현을 나타낼 경우는 M-TMP로 분석한다.

- 그녀는 아침마다(TMP) 커피를 마시기 위해 물을 끓인다.

## [주의]

- (1) ‘중에서’ 구성은 LOC가 아니라, 범위로 해석하여 M-EXT(범위)로 분석한다.

가. 경찰은 관계자 중에서(M-EXT) 최초 유포자를 색출했다.  
나. 그 영화는 지금껏 내가 본 영화 중에서(M-EXT) 가장 그로테스크하다.

- (2) 구체적인 시기를 포함하더라도, 그 뒤에 ‘명사구+에서’가 나오면 M-TMP가 아닌 M-LOC로 분석한다.

- 2002년 대선에서(M-LOC) 노 후보가 당선되었다.

- (3) 서술어가 나타내는 ‘사건이 일어나는 공간’을 가리키는 경우에는 M-LOC로 분석하고, 서술어가 나타내는 ‘사건의 범위’로 해석되는 경우에는 M-EXT로 분석해야 함을 주의한다. (더 자세한 예는 5.9의 (4)를 참고한다.)

가. 그는 각종 선거에서(M-EXT) 반대 정당의 텃밭인 곳에서(M-LOC) 승리했다.  
 나. 지난 대선에서(M-EXT) 노 후보가 아무 연고도 없는 곳에서(M-LOC) 승리했다.

### 5.3. M-DIR (Direction, 방향)

#### [정의]

- 술어가 나타내는 행위의 방향으로 해석되는 부가어를 DIR 로 분석한다.

#### [예시]

- (1) ‘로’가 결합된 조사 결합 구성을 M-DIR로 분석한다.

가. 모든 동물들은 노폐물을 몸 밖으로(M-DIR) 배설한다.  
 나. 그녀의 뒷모습이 안개 속으로(M-DIR) 스러졌다.

- (2) ‘오른쪽’, ‘왼쪽’, ‘위쪽’, ‘아래쪽’, ‘앞으로’, ‘뒤로’, ‘동서남북’과 같이 방향을 의미하는 명사구나 명사 ‘방향’을 사용한 구성도 M-DIR로 분석한다.

가. 달이 서쪽으로(M-DIR) 기울었다.  
 나. 딸이 이불을 사방으로(M-DIR) 펼쳤다.  
 다. 경찰은 범인의 팔을 반대 방향으로(M-DIR) 꺾었다.  
 라. 아버지는 오른쪽 방향으로(M-DIR) 누우셨다.

- (3) ‘을 향해’를 사용한 구성도 M-DIR로 분석한다.

가. 뒤차는 앞차를 향해(M-DIR) 경적을 울렸다.  
 나. 그는 부모님이 계신 곳을 향해(M-DIR) 소리쳤다.

#### [주의]

- 서술어가 이동의 의미를 가지는 경우에는 ‘방향’이 논항일 가능성이 크므로, 논항을 부가어로 분석하지 않도록 주의한다.

가. 경찰차가 집 앞으로 지나갔다.  
 나. 동생은 터미널을 떠나 서울로 출발했다.  
 다. 나는 눈을 뜨자마자 금세 학교로 달려갔다.

## 5.4. M-ROU (Route, 경로)

### [정의]

- 서술어가 나타내는 이동의 행위로 주체가 거쳐 가는 중간 장소를 가리키는 부가어를 M-ROU 로 분석한다.

### [예시]

- (1) ‘로’가 결합한 구성이 다음과 같이 이동의 부가어로 나타나는 경우 M-ROU로 분석한다.

가. 우리는 목적지까지 직림길로(M-ROU) 가기로 했다.  
나. 이론차가 큰길로(M-ROU) 다니면 위험하다.

- (2) ‘-을 거쳐, ~을 따라’와 같은 구성으로 경로로 해석되는 절은 M-ROU로 분석한다.

가. 그는 일본을 거쳐(M-ROU) 상해로 망명하였다.  
나. 그 배는 미시시피강을 따라(M-ROU) 내려간다.

### [주의]

- ‘로’가 결합한 구성은 논항일 가능성이 있으므로 논항을 부가어로 분석하지 않도록 주의한다.

가. 이 항공기는 중국 영공으로(DIR) 지나간다.  
나. 경찰차 한 대가 집 앞으로(DIR) 지나갔다.

## 5.5. M-INS (Instrument, 도구)

### [정의]

- 서술어가 나타내는 사건에 사용되는 물리적 도구나 사물, 재료를 가리키는 부가어를 M-INS 로 분석한다. 방법, 방식, 도구 등이 추상적인 경우도 M-INS 로 분석한다.

### [예시]

- (1) 다음과 같이 ‘로’가 결합된 조사 결합 구성은 M-INS로 분석한다. 이때, ‘~을 이용하여’로 치환하여 의미가 통하는지의 여부가 분석에 도움이 된다.

가. 그는 시금치를 끓는 물로(M-INS) 살짝 데쳤다.  
나. 아이들은 색연필로(M-INS) 알록달록 그림을 칠했다.  
라. 철이 엄마와 철이는 수화로(M-INS) 소통한다.  
마. 나는 인터넷 뱅킹으로(M-INS) 아파트 관리비를 이체했다.

- (2) ‘-을 통해’와 같은 구성도 M-INS로 분석한다.

가. 빗물은 배수로를 통해(M-INS) 배수된다.  
 나. 정부의 꾸준한 노력을 통해(M-INS) 보육 정책이 내실화되고 있다.

(3) ‘-에 의하여’와 같은 구성도 M-INS로 분석한다.

가. 특검 연장에 의해(M-INS) 성난 민심이 안정화되었다.  
 나. 명문대 출신들이 블라인드 채용에 의해(M-INS) 기업 인사자들에게 역차별되었다.  
 다. 사기꾼의 손에 의해(M-INS) 조잡한 위조품이 수입 명품으로 둔갑했다.

(4) 피동문의 행위 주체로 ‘-에 의하여’를 사용하는 경우도 M-INS로 분석한다.

가. 야생마가 훈련사에 의해(M-INS) 순치되었다.  
 나. 그 바위는 부족민들에 의해(M-INS) 영화되었다.

#### [주의]

- 사물이나 재료이지만 도구로 해석되지 않는 경우가 있다. 결과 상태와 유사하지만 논항이 아닌 아래와 같은 경우는 자격으로 이해하여 M-QLF로 분석한다.

- 용의자의 계좌에서 거액이 현금으로(M-QLF) 입출금되었다.

## 5.6. M-MNR (Manner, 방식)

#### [정의]

- 서술어가 나타내는 행위의 방식으로 해석되는 부가어를 M-MNR로 분석한다.

#### [예시]

(1) 행위의 방식으로 해석되는 성상부사는 M-MNR로 분석한다.

가. 그녀는 상처받은 내 마음을 천천히(M-MNR) 위무하였다.  
 나. 철수는 아들을 정성껏(M-MNR) 간호했다.  
 다. 각 지방자치단체에서는 수세 시에 납세를 철저히(M-MNR) 검수한다.  
 라. 아들은 숨김없이(M-MNR) 자신의 계획을 전부 다 까냈다.  
 마. 우리는 얼어 죽지 않기 위해 조난당한 산속에서 서로의 허리를 꽂(M-MNR) 껴안았다.  
 바. 아이들은 가재를 잡기 위해 물속의 바위를 살짝(M-MNR) 들추었다.  
 사. 그녀는 인생을 깊이(M-MNR) 사색했다.  
 아. 학생들이 선생님의 강의를 열심히(M-MNR) 경청하고 있다.

단, 빈도부사의 경우 M-TMP로 분석한다.

가. 엄마는 아이 볼에 자주(M-TMP) 뽀뽀했다.  
 나. 선생님의 조언은 항상(M-TMP) 명쾌하여 고민이 생길 때면 찾아뵙게 된다.

(2) 연결어미 ‘-게’가 쓰인 부사절이 서술어가 나타내는 방식을 나타내면 M-MNR로 분석한다.

가. 원앙 한 쌍이 한가롭게(M-MNR) 호수에서 노닐고 있다.  
나. 진행자가 자신의 새 프로그램에 손흥민 선수를 어렵게(M-MNR) 섭외했다.  
다. 그녀는 1층 점포를 청년 사업자에게 싸게(M-MNR) 세놓았다.  
라. 정치권이 금권과 은밀하게(M-MNR) 야합했다.

(3) ‘-ㄴ 채’로 실현된 부사성 의존명사구도 M-MNR로 분석한다.

가. 종우는 꿈과 현실을 구분하지 못한 채(M-MNR) 주위 사람들에게 자신이 살던 고시원의 사람들이 살인마였다고 거짓말을 했다.  
나. 언니는 눈물을 머금은 채로(M-MNR) 가방 속의 편지를 가만히 보았다.

(4) 의성어 및 의태어도 M-MNR로 분석한다.

가. 요거는 몇 초만 타닥(M-MNR) 돌리면 된다.  
나. 예전에는 급하게 후루룩후루룩(M-MNR) 먹었는데

(5) 예시를 보이는 ‘~ 등’ 구성은 M-MNR로 분석한다. 단, 이러한 구성은 관형어로 기능하는 경우도 있고, 부사어로 기능하는 경우도 있으므로 분석에 유의한다.

가. 구내식당에서는 내부 설치된 모니터에 잔반 처리 관련 영상을 틀어놓는 등(M-MNR) 잔반 줄이기에 힘쓰고 있다.  
나. 식사 인원을 미리 파악하는 시스템을 도입하는 등(M-MNR) 잔반을 획기적으로 줄였다.  
다. 보궐이사는 후보자 결정 후 임명 동의서 등 구비서류와 결격사유 확인을 동시에 진행한다.  
→ ‘가-나’는 부사어로 기능하므로 M-MNR로 처리하고, ‘다’는 후행하는 명사구를 수식하는 관형어로 해석되므로 의미역의 분석 대상으로 삼지 않음.

#### [주의]

(1) ‘로, 로써’의 구성이 추상적인 방식으로 해석되어도 M-MNR로 분석하지 않고 M-INS로 분석한다.

가. 철이 엄마와 철이는 수화로(M-INS) 소통한다.  
나. 나는 인터넷 뱅킹으로(M-INS) 아파트 관리비를 이체했다.  
다. 동생은 자수성가에 성공함으로써(M-INS) 그때까지의 고정 관념을 몽땅 다 깨부셨다.  
라. 오늘 나는 마지막 인사를 전함으로써(M-INS) 그와의 끈질긴 인연을 끝맺었다.

(2) ‘로서’의 구성이 자격의 의미로 해석되면 M-MNR로 분석하지 않고 M-QLF로 분석한다.

가. 영화의 주제가로(M-QLF) 그 곡이 선곡되었다.  
나. 그 영화는 연극으로(M-QLF) 재탄생되었다.  
다. 그는 삼남삼녀 가운데 막내로(M-QLF) 태어났다.

(3) ‘보다’, ‘에 비해’, ‘와 달리’와 같은 구성은 다른 대상과의 비교의 의미로 사용되므로 M-MNR로 분석하지 않고 M-CRT로 분석한다.

가. 그는 실제 나이보다(M-CRT) 훨씬 앳되었다.

나. 예상과 달리(M-CRT) 미국에서 온 언니가 한국에 머문 시간은 겨우 열흘 남짓하였다.

(4) '아서/어서'와 같은 구성이 선행(계기)로 해석되면, M-MNR로 분석하지 않고 M-PRE로 분석한다.

가. 이장은 바둑판을 펼쳐서(M-PRE) 그 위에 바둑알을 쏟아부었다.  
나. 그 경리직원은 장부를 조작해서(M-PRE) 회삿돈을 생활비에 써먹었다.  
다. 나는 창호지에 구멍을 뚫어서(M-PRE) 방 안의 사람들을 엿보았다.

(5) '-을 거쳐'와 같은 구성으로 경로로 해석되는 절은 M-MNR로 분석하지 않고 M-ROU로 분석한다.

가. 그는 일본을 거쳐(M-ROU) 상해로 망명하였다.  
나. 그 배는 미시시피강을 따라(M-ROU) 내려간다.

## 5.7. M-QNT (Quantity, 수량)

### [정의]

- 크기, 높이, 양, 수 등 구체적인 수치를 나타내는 부가어를 M-QNT로 분석한다.

### [예시]

(1) 수사나 수 관형사 등이 쓰인 부가어를 M-QNT로 분석한다.

가. 질문 하나(M-QNT) 합시다.  
나. 그는 평소에 책을 한 권도(M-QNT) 읽지 않는다.  
다. 이 복싱 선수는 세계 선수권 대회를 세 번이나(M-QNT) 연패했다.

(2) 수량 표현+'정도/이상/이내' 등의 구성도 M-QNT로 분석한다.

가. 코스닥 지수가 7% 이상(M-QNT) 하락했다.  
나. 5년 이내에 암호 화폐는 3배 정도(M-QNT) 상승 가능성이 있다.

(3) 경기 결과인 점수를 나타내는 경우에도 M-QNT로 분석한다.

가. 기아가 엘지에 2:0으로(M-QNT) 승리했다.  
나. 엘지가 기아에 0:2로(M-QNT) 패배했다.

## 5.8. M-DGR (Degree, 정도)

### [정의]

- 서술어가 나타내는 사건이나 상태의 정도나 수준으로 해석되는 부가어를 M-DGR 로 분석한다.

### [예시]

- (1) '가장, 극히, 더욱, 많이, 매우, 모두, 썩, 아주, 완전히, 잔뜩, 최고로, 크게, 대단히' 와 같은 정도부사가 사용된 경우를 M-DGR로 분석한다.

가. 그들은 아름다운 경치에 **흠뻑(M-DGR)** 도취되었다.  
나. 그녀는 전국에 있는 대부분의 올레길을 **모두(M-DGR)** 답파했다.  
다. 가족들 간의 재산 싸움은 **대단히(M-DGR)** 불미스럽다.

- (2) '-르 정도로' 등의 구성도 M-DGR로 분석한다.

가. 웬만한 반지는 맞지 않을 정도로(M-DGR) 김 선생의 손가락은 굵다.  
나. 학생들의 의견이 하나로 수렴하기 힘들 정도로(M-DGR) 다양했다.

### [주의]

- (1) 구체적인 수치를 나타내는 숫자, 수사, 수 관형사가 쓰인 '양'이나 '수'와 관련된 성분은 M-DGR로 분석하지 않고 M-QNT로 분석한다.

가. 그 사고에서 무고한 사람이 **10명이나(M-QNT)** 압사되었다.  
나. 오늘 코스닥 지수가 615포인트에서 570포인트로 **7% 이상(M-QNT)** 떨어졌다.

- (2) 서술어가 나타내는 사건의 범위와 관련된 성분은 DGR로 분석하지 않고 EXT로 분석한다.

가. 언니는 내가 아는 사람들 중에서(M-EXT) 가장 너그럽다.  
나. 이 작품은 그의 소설들 가운데(M-EXT) 가장 걸출하다.

## 5.9. M-EXT (Extent, 범위)

### [정의]

- 서술어가 나타내는 사건의 범위로 해석되는 부가어를 M-EXT 로 분석한다.

### [예시]

- (1) '~ 중에', '~ 중에서', '~ 가운데', '~ 대상으로' 등이 결합된 구성을 M-EXT로 분석한다.

가. 여러 옷감 중에(M-EXT) 이 옷감의 빛깔이 가장 옅다.  
나. 언니는 내가 아는 사람들 중에서(M-EXT) 가장 너그럽다.  
다. 이 작품은 그의 소설들 가운데(M-EXT) 가장 걸출하다.



라. 여론 조사가 성인 남녀 1800명을 대상으로(M-EXT) 실시되었다.

- (2) 대상으로 해석되지 않는 '에 대하여' 구성도 M-EXT로 분석한다.

가. 우리는 그 문제에 대해(M-EXT) 대책을 논의했다.

나. 보험사는 자연재해로 인한 사고에 대해(M-EXT) 보험자의 보험금 지급을 면책했다.

다. 학생들은 교사의 체벌에 대하여(M-EXT) 눈을 부릅떴다.

#### [주의]

- (1) 범위가 시간이나 기간을 나타내는 경우에는 M-TMP로 분석한다.

가. 3달 이상(M-TMP) 휴대전화 요금이 미납되어 휴대전화를 사용할 수 없다.

나. 우리 어머니는 첫 직장에서 30년이나(M-TMP) 근무하셨다.

- (2) 구체적인 수치를 나타내는 숫자, 수사, 수 관형사가 쓰인 '양'이나 '수'와 관련된 성분은 M-EXT로 분석하지 않고 M-QNT로 분석한다.

가. 그 사고에서 무고한 사람이 10명이나(M-QNT) 압사되었다.

나. 오늘 코스닥 지수가 615포인트에서 570포인트로 7% 이상(M-QNT) 떨어졌다.

- (3) 서술어가 나타내는 사건이나 상태의 '정도'나 '수준'과 관련된 성분은 M-EXT로 분석하지 않고 M-DGR로 분석한다.

가. 오늘 성수의 차림새가 아주(M-DGR) 멋지다.

나. 선수들은 잘못된 판정이 몹시(M-DGR) 분했다.

- (4) 서술어가 나타내는 '사건의 범위'로 해석되는 경우에는 M-EXT로 분석하고, 서술어가 나타내는 '사건이 일어나는 공간'을 가리키는 경우에는 M-LOC로 분석한다.

가. 해외 출장 일정에서(M-EXT) 미국과 일본의 주요 도시를 방문했다.

나. 정치지도자로 누가 더 좋은가라는 질문에(M-EXT) 이 후보가 가장 높은 것으로 조사됐다.

다. 우리 동아리에서(M-EXT) 철수가 농구를 제일 잘한다.

라. 우리 반에서(M-EXT) 영희가 제일 예쁘다.

마. 그는 각종 선거에서(M-EXT) 반대 정당의 텃밭인 곳에서(M-LOC) 승리했다.

바. 지난 대선에서(M-EXT) 노 후보가 아무 연고도 없는 곳에서(M-LOC) 승리했다.

→ 특히 '마-바'의 M-EXT와 M-LOC를 혼동하지 않도록 주의한다.

- (5) '~의 경우' 구성은 특정 사실에 대한 진술이라면 M-EXT로 분석하고, 미래 사태에 대한 가정이라면 M-CND로 분석한다.

가. 퇴비의 경우(M-EXT) 사용 7일 전에 관할 지자체에 신고하기로 규정되어 있다.

나. 진통이 30분 이상 계속될 경우(M-CND) 약물 복용을 중단해야 한다.

- (6) 다음의 '~ 가운데' 구성을 M-BCK가 아닌 M-EXT로 분석하지 않도록 주의한다.

가. 영화가 더딘 흥행세를 보이고 있는 가운데(M-BCK) 손익 분기점을 넘을 것으로 보인다.  
 나. 애국가가 올려 퍼지는 가운데(M-BCK) 태극기가 게양되었다.

## 5.10. M-COM (Comitative, 공동)

### [정의]

- 주어 혹은 목적어 참여자와 같은 지위를 갖는 참여자로 해석되는 부가어를 M-COM 으로 분석한다.

### [예시]

- (1) ‘와/과’와 결합하여 주어와 동반하여 서술어의 행위를 하는 대상을 M-COM으로 분석한다.

가. 철수는 영희와(M-COM) 극장에 갔다.  
 나. 철수는 거기에 친구와(M-COM) 갔다.  
 다. 철수는 영희와(M-COM) 장난감을 가지고 놀았다.

- (2) 목적어 참여자와 같은 지위를 갖는 참여자도 M-COM으로 분석한다.

- 철수는 영희를 기영이와(M-COM) 집에 데리고 갔다.

- (3) ‘-와 함께, -와 같이’ 구성도 M-COM으로 분석한다.

가. 왕은 왕비가 아끼던 물건들을 그녀와 함께(M-COM) 부장했다.  
 나. 병사들은 유서를 머리카락과 함께(M-COM) 봉입했다.

### [주의]

- (1) 서술어가 상호적인 의미를 가지는 경우에는 서술어의 논항으로 분석한다.

가. 나는 오래간만에 고향 친구와 이야기를 나누었다.  
 나. 빗길에 자전거가 자동차와 부딪쳤다.

- (2) ‘같이’, ‘함께’가 동반의 의미가 있기는 하나 참여자 혹은 대상(명사구)이 없이 쓰인 경우까지 COM으로 처리하지 않도록 주의한다. 동반주는 어디까지나 주어와 관련한 참여자 혹은 목적어와 관련한 대상에 국한해 부여하는 의미역이다.

가. 어제 저녁을 같이(M-MNR) 먹은 사람이 누구지?  
 나. 어제 저녁을 누구와 같이/함께(M-COM) 먹었니?

## 5.11. M-CRT (Criterion, 기준)

### [정의]

- 서술어에 대해 평가되는 대상 혹은 판단하는 대상을 비교하는 부가어를 M-CRT 로 분석한다.

### [예시]

- (1) '보다, 처럼, 만큼' 등이 결합된 조사 결합 구성을 M-CRT로 분석한다.

가. 오늘이 어제보다(M-CRT) 더 시원하다.  
나. 철수는 영희보다(M-CRT) 덜 크다.  
다. 고속버스가 기차만큼(M-CRT) 빠르다.  
라. 나는 엄마처럼(M-CRT) 걷는다.

- (2) '-에 비해(서)'의 구성도 M-CRT로 분석한다.

가. 강철에 비해서(M-CRT) 적다.  
나. 수미는 얼굴에 비해서(M-CRT) 눈이 크다.

- (3) '-다기보다는', '-느니보다는' 등의 구성도 CRT로 분석한다.

가. 내가 크다기보다는(M-CRT) 네가 작은 것이다.  
나. 내가 가느니보다는(M-CRT) 네가 가는 것이 더 낫다.

### [주의]

- 서술어가 평가 혹은 판단의 의미를 가지는 경우에는 서술어의 논항으로 분석한다.

가. 이 일은 내 능력에(CRT) 부친다.  
나. 철수가 영희만(CRT) 못하다.

## 5.12. M-QLF (Qualification, 자격)

### [정의]

- 서술어가 나타내는 대상에 대하여 인물의 자격이나 사물의 자격으로 해석되는 부가어를 M-QLF 로 분석한다.

### [예시]

- (1) '로(서)'가 결합된 조사 결합 구성이 인물의 자격이나 사물의 자격으로 해석되는 경우 M-QLF로 분석한다.

가. 그는 삼남삼녀 가운데 막내로(M-QLF) 태어났다.  
나. 그 자료는 중간 매개체로서(M-QLF) 역할을 충분히 해 냈다.

(2) '로'가 결합된 조사 결합 구성이 사물의 상태를 나타내는 경우도 M-QLF로 분석한다.

- 용의자의 계좌에서 거액이 현금으로(M-QLF) 입출금되었다.

#### [주의]

(1) M-QLF와 M-MNR을 판단하기 어려울 수 있다. '무엇에 해당하는 것이 이것이다' 정도로 파악되면 M-QLF로 처리한다.

가. 책은 중고로(M-QLF) 사는 것이 좋다.  
나. 오래된 집은 두고 지금 집은 전세로(M-QLF) 산다.

(2) '로'가 결합된 조사 결합 구성이 '배경'으로 해석될 수 있는 경우도 있으므로 서술어와의 관계를 살펴 분석하도록 한다.

가. 3번 우익수로(M-QLF) 선발로(M-QLF) 출전했다. → 각각 '출전하다'의 자격이 맞음.  
나. 21세기 대중음악계의 대표적 인물로(M-BCK) 빌보드에서 뽑은 2010년대 가장 성공한 음악가로(M-QLF) 선풍적인 인기를 얻고 있다. → 전자는 '~인데(배경)'으로 해석.

## 5.13. M-CAU (Cause, 원인)

#### [정의]

- 서술어가 나타내는 사건을 일으킨 원인이나 이유로 해석되는 부가어를 M-CAU로 분석한다.

#### [예시]

(1) 원인이나 이유를 나타내는 '에', '로' 등이 결합된 조사 결합 구성을 M-CAU로 분석한다.

가. 그는 심사위원의 혹평에(M-CAU) 얼굴을 붉혔다.  
나. 강풍으로(M-CAU) 나무가 쓰러졌다.

(2) '~ 때문에', '-기 때문에' 구성을 M-CAU로 분석한다.

가. 그는 감기 때문에(M-CAU) 결석했다.  
나. 티 나지 않게 선동하는 것이 목적이기 때문에(M-CAU) 가짜 뉴스에 사용되는 수법들은 너무나도 교묘하다.

(3) '-아서', '-니까', '-느라', '-더니' 등이 결합된 부사절을 M-CAU로 분석한다.

가. 굴이 설익어서(M-CAU) 시다.  
나. 연우가 자료 정리하느라(M-CAU) 제일 수고했다.  
다. 장난감을 안 사 줬더니(M-CAU) 아이의 표정이 시무룩하다.

(4) ‘-아서인지’ ‘-아서 그런지’ 등의 구성도 M-CAU로 분석한다.

가. 숙원이 해소돼서인지(M-CAU) 사람들이 밤새도록 신나게 어깨춤을 논다.  
나. 물을 많이 넣었는지(M-CAU) 죽이 물었다.  
다. 애교가 넘쳐서 그런지(M-CAU) 언니는 형부를 너무 귀여워한다.

(5) 원인으로 해석되는 ‘-고’도 M-CAU로 분석한다.

가. 그 소식을 듣고(M-CAU) 마음이 요동쳤다.  
나. 나는 약을 맞고(M-CAU) 근육이 이완되어 일어설 수가 없었다.

#### [주의]

(1) 앞서 일어난 일을 나타내는 ‘-아서’는 M-CAU로 분석하지 않고 M-PRE으로 분석한다.

가. 이장은 바둑판을 펼쳐서(M-PRE) 그 위에 바둑알을 쏟아부었다.  
나. 그 경리직원은 장부를 조작해서(M-PRE) 회삿돈을 생활비에 써먹었다.  
다. 나는 창호지에 구멍을 뚫어서(M-PRE) 방 안의 사람들을 엿보았다.

(2) 선행과 원인으로 뚜렷하게 구별하기 어려운 ‘-아서’의 경우 선행절과 후행절의 행위주가 일치하면 M-PRE로 분석하고, 선행절과 후행절의 행위주가 일치하지 않으면 M-CAU로 분석한다.

가. 나는 일을 해서(M-PRE) 돈을 벌다.  
나. 내가 일을 해서(M-CAU) 가족들이 살아간다.

## 5.14. M-PRP (Purpose, 목적)

#### [정의]

- 서술어의 주체가 가리키는 목적이나 목표, 의도로 해석되는 부가어를 M-PRP 로 분석한다.

#### [예시]

(1) ‘으로’나 ‘에서’가 결합한 구성이 목적의 의미를 가지면 M-PRP로 분석한다.

가. 취재에 따르면 검찰은 낙마시킬 목적으로(M-PRP) 임명도 되기 전에 법무부 장관 내정자를 내 사했다.  
나. 그는 돈을 벌 목적으로(M-PRP) 다른 회사에 회사 기밀을 밀고하였다.  
다. 아기가 맘마를 받아먹겠다는 일념으로(M-PRP) 귀여운 얼굴을 하고는 입을 넓적한다.  
라. 한 영화사는 자랑하려는 의도에서(M-PRP) 대작을 제작했다.

(2) '-기 위하여'가 쓰인 구성은 M-PRP로 분석한다.

가. 독재에 저항하기 위하여(M-PRP) 학생들은 노동자들과 단합했다.  
나. 민희는 외국 회사에 들어가기 위해(M-PRP) 외국어를 독학했다.

(3) 다음과 같은 연결어미가 쓰인 구성은 모두 M-PRP로 분석한다.

1) -러

가. 그는 물을 마시러(M-PRP) 부엌에 여러 번 들락거렸다.  
나. 나는 형의 심부름으로 커피를 사러(M-PRP) 카페에 갔다.

2) -려고

가. 그녀는 자신에게 맞는지 확인하려고(M-PRP) 바지를 몸에 대보았다.  
나. 모든 직원들이 납품 기일을 맞추려고(M-PRP) 밤샘했다.

3) -고자

가. '백이'와 '숙제'는 절개를 지키고자(M-PRP) 수양산에 거처했다.  
나. 이모 부부는 아이들에게 작은 도움이라도 되고자(M-PRP) 소정의 금액을 복지원에 매월 기탁하  
신다.

4) -겠다고

가. 강아지가 개껌을 주워 먹겠다고(M-PRP) 귀여운 얼굴을 하고는 납작거렸다.  
나. 아기가 맘마를 받아먹겠다고(M-PRP) 입을 오물거렸다.

5) -도록

가. 엄마는 놀고 싶은 대로 놀도록(M-PRP) 아이를 놓아두었다.  
나. 아내는 내가 무안하지 않도록(M-PRP) 내 유머를 들을 때마다 박수를 치면서 깔깔거린다.

#### [주의]

- '-느라' 구성의 경우는 원인으로 해석되므로 M-PRP로 분석하지 않고 M-CAU로 분석한다.

- 에디는 무엇인가를 만드느라(M-CAU) 며칠 동안 두문불출했다.

## 5.15. M-CND (Condition, 조건)

### [정의]

- 서술어가 나타내는 사건의 조건으로 해석되는 부가어를 M-CND 로 분석한다.

### [예시]

- (1) '-면', '-다면', '-ㄴ수록' 등이 결합된 부사절을 M-CND로 분석한다.

가. 아이들은 마트에 가면(M-CND) 장난감을 사 달라고 떼쓴다.  
나. 양심이 있다면(M-CND) 검찰은 억지 수사를 그만해라.  
다. 상류로 갈수록(M-CND) 개울물이 얕아졌다.

- (2) 시간표현과 결합된 '-면'도 CND로 분석한다.

가. 아침이면(M-CND) 버스가 학생들로 바글거린다.  
나. 아버지는 부지런하셔서 아침 6시면(M-CND) 눈뜨신다.  
다. 명절이 되면(M-CND) 동네 아이들은 이곳에 모두 모여 신나게 널뛰었다.

- (3) 연결 어미가 아닌 'ㄴ/ㄹ + 경우'와 같은 표현들이 조건으로 해석될 경우에도 CND로 분석한다.

- 비가 올 경우에는(M-CND) 경기를 연기한다.

### [주의]

- '~의 경우' 구성이 미래 사태에 대한 가정이라면 M-CND 로 분석하고, 특정 사실에 대한 진술이라면 M-EXT 로 분석한다.

가. 진통이 30분 이상 계속될 경우(M-CND) 약물 복용을 중단해야 한다.  
나. 퇴비의 경우(M-EXT) 사용 7일 전에 관할 지자체에 신고하기로 규정되어 있다

## 5.16. M-CNC (Concession, 양보)

### [정의]

- 두 절이 연결되어 있을 때 선행절이 후행절에 대한 양보로 해석되는 경우, 선행하는 부사절을 M-CNC 로 분석한다.

### [예시]

- '-아도', '-더라도', '-ㄴ지라도', '-ㄴ들' 등이 결합된 부사절을 M-CNC 로 분석한다.

가. 언니는 갑자기 모함을 당했어도(M-CNC) 놀라울 만큼 냉정했다.

나. 어머니는 아무리 늦게 주무시더라도(M-CNC) 항상 새벽 5시에 일어나신다.  
 다. 부자는 아닐지라도(M-CNC) 문 선생님의 출신은 매우 귀하다.  
 라. 내가 곧장 출발한들(M-CNC) 버스를 탈 수 없을 것이다.

#### [주의]

\* ‘-면’, ‘-다면’ 등이 결합된 성분은 M-CNC로 분석하지 않고 M-CND로 분석한다.

가. 아이들은 마트에 가면(M-CND) 장난감 때문에 떼쓴다.  
 나. 양심이 있다면(M-CND) 검찰은 억지 수사를 그만해라.

## 5.17. M-PRE (Preceding, 선행)

#### [정의]

- 선행하는 서술어가 후행하는 서술어에 대한 앞선 사건을 나타내는 경우, 두 절 사이의 사건이 계기적으로 일어나는 경우, 선행하는 부사절을 PRE로 분석한다.

#### [예시]

(1) 연결어미 ‘-아’가 결합된 부사절을 M-PRE로 분석한다.

가. 설날에 온 가족이 모여(M-PRE) 만두를 빚었다.  
 나. 길고양이가 마당에 들어와(M-PRE) 꾸벅꾸벅 존다.  
 다. 아버지는 새벽같이 일어나(M-PRE) 산책을 하신다.

(2) 계기적으로 해석되는 ‘-아서’도 M-PRE로 분석한다.

가. 이장은 바둑판을 펼쳐서(M-PRE) 그 위에 바둑알을 쏟아부었다.  
 나. 어부는 고기를 잡아서(M-PRE) 시장에 팔았다.  
 다. 어머니는 오랜만에 친구들을 만나서(M-PRE) 술 한 잔을 하셨다.

(3) ‘에 따라’ 구성도 M-PRE로 분석한다.

가. 구청은 주민들의 요청에 따라(M-PRE) 개천을 복개했다.  
 나. 증거가 확인되지 않음에 따라(M-PRE) 가해자는 불기소되었다.  
 다. 이모의 유언에 따라(M-PRE) 소정의 금액이 복지원에 매월 기탁되었다.

#### [주의]

(1) 원인으로 해석되는 ‘-아서’는 M-PRE로 분석하지 않고 M-CAU로 분석한다.

가. 그는 불안해서(M-CAU) 쪽잠을 잤다.  
 나. 나는 점심을 급하게 먹어서(M-CAU) 체했다.  
 다. 마늘을 씹어서(M-CAU) 혀가 몹시 아리다.



- (2) 선행과 원인으로 뚜렷하게 구별하기 어려운 ‘-아서’의 경우 선행절과 후행절의 행위주가 일치하면 M-PRE로 분석하고, 선행절과 후행절의 행위주가 일치하지 않으면 M-CAU로 분석한다.

가. 나는 일을 해서(M-PRE) 돈을 벌다.  
나. 내가 일을 해서(M-CAU) 가족들이 살아간다.

- (3) ‘에 따르면’ 구성은 M-PRE로 분석하지 않고 M-CND로 분석한다.

가. 취재에 따르면(M-CND) 임명도 되기 전에 법무부 장관 내정자는 내사됐다.  
나. 아침 뉴스에 따르면(M-CND) 북미 회담 개최에 주가가 급반등하였다.  
다. 물리 법칙에 따르면(M-CND) 물은 열이 가해졌을 때 수증기가 되어 기화한다.

## 5.18. M-BCK (Background, 배경)

### [정의]

- 사건이 일어난 배경을 나타내는 부사절을 M-BCK 으로 분석한다.

### [예시]

- (1) ‘-는데’가 결합하여 배경을 나타내는 부사절을 M-BCK으로 분석한다.

가. 소설책을 샀는데(M-BCK) 아주 재미있더라.  
나. 유명한 식당인데(M-BCK) 거기에 갈래?  
다. 한국어를 배우는데(M-BCK) 정말 공부하는 게 재미있다.

- (2) ‘-다가, -으니’ 등이 결합되어 배경을 나타내는 부사절도 M-BCK으로 분석한다.

가. 어제 홍대에 갔다가(M-BCK) 친구를 만났다.  
나. 그가 주말마다 가는 곳이 있으니(M-BCK) 그곳은 작은 자원봉사 단체였다.

## 5.19. M-LST (List, 나열)

### [정의]

- 두 절이 연결되어 있을 때, 나열의 의미를 나타내는 선행하는 부사절을 M-LST 로 분석한다.

### [예시]

- (1) 대등적 의미를 나타내는 ‘-고’가 결합된 부사절을 M-LST로 분석한다.

가. 여름에는 비가 내리고(M-LST) 겨울에는 눈이 내린다.  
나. 언니는 학교에 가고(M-LST) 부모님은 출근하셨다.

(2) ‘-며’, ‘-면서’가 결합된 부사절을 M-LST로 분석한다.

가. 그 집 아이는 공부도 잘하며(M-LST) 운동도 잘한다.  
나. 저 바다 위로 갈매기 떼가 큰 소리를 내며(M-LST) 끼룩댄다.  
다. 우리는 축배를 들기 위해 구호를 외치며(M-LST) 잔을 내들었다.  
라. 거품을 일으키며(M-LST) 술이 괴고 있다.  
마. 지원자들은 합속하면서(M-LST) 친밀하게 교류했다.

(3) ‘-지’가 결합된 부사절이 대등적으로 접속하는 경우도 M-LST로 분석한다.

가. 이견 수필이지(M-LST) 논문이 아니다.  
나. 콩이 나지(M-LST) 팔이 나지 않는다.  
다. 건강은 건강할 때 지켜야지(M-LST) 나중에 후회한다.

## 5.20. M-CTR (Contrast, 대조)

[정의]

- 두 절이 연결되어 있을 때, 대조의 의미를 나타내는 선행하는 부사절을 M-CTR로 분석한다.

[예시]

(1) ‘-나’가 결합된 부사절을 M-CTR로 분석한다.

- 오늘 날씨는 눈은 내리나(M-CTR) 쌓이지는 않겠습니다.

(2) ‘-지만’이 결합된 부사절을 M-CTR로 분석한다.

가. 그는 어제는 안 왔지만(M-CTR) 오늘은 왔다.  
나. 이 음식은 생긴 건 우습지만(M-CTR) 맛은 있다.

[주의]

- 하나의 연결어미가 여러 가지 의미를 가지는 경우 주의하여 분석한다.

- 새 상품은 있는데(M-BCK) 새 상품은 가격이 높다.

## 5.21. M-SEL (Selective, 선택)

[정의]

- 두 절이 연결되어 있을 때, 선택의 의미로 해석되는 부사절을 M-SEL로 분석한다.

[예시]

- (1) ‘-거나, -든지’가 결합된 부사절은 M-SEL로 분석한다.

가. 우리가 그곳에 머무는 내내 눈이 왔거나(M-SEL) 비가 왔다.  
나. 도서관에는 철수가 가든지(M-SEL) 영화가 갈 것이다.

- (2) ‘-거나 -거나’, ‘-든지 -든지’와 같은 반복구성은 그 전체를 M-SEL로 분석한다.

가. 주말엔 보통 텔레비전을 보거나 컴퓨터 게임을 하거나(M-SEL) 해요.  
나. 철수가 가든지 영화가 가든지(M-SEL) 마음대로 해라.

## 5.22. M-NEG(Negation, 부정)

[정의]

- 절이 나타내는 사태를 부정하는 부정 부사를 M-NEG로 분석한다.

[예시]

- ‘안/아니’, ‘못’ 등의 부사어를 M-NEG로 분석한다.

가. 오늘은 비가 안(M-NEG) 온다.  
나. 형은 편지 한 장 아니(M-NEG) 남겼다.  
다. 철수는 술을 못(M-NEG) 마신다.

[주의]

- (1) ‘전혀, 결코, 도무지, 도저히’ 등의 부정 극어를 M-NEG로 분석하지 않도록 주의한다.

가. 결코(M-MOD) 나는 그 일을 하지 않았다.  
나. 전혀(M-MOD) 예상하지 못한 일이 발생했다.

- (2) ‘-지 않다’, ‘-지 못하다’, ‘-지 말다’와 같은 장형 부정문의 보조 용언을 M-NEG로 분석하지 않도록 주의한다.

가. 나는 밥을 먹지 않는다. (보조 용언이므로 대상 서술어가 아님.)  
나. 그런 태도는 옳지 못하다. (보조 용언이므로 대상 서술어가 아님.)  
다. 이곳에 쓰레기를 함부로 버리지 마시오. (보조 용언이므로 대상 서술어가 아님.)

## 5.23. M-MOD (Modal, 양태)

### [정의]

- 절이 나타내는 사태에 대한 화자의 태도를 나타내는 양태 부사를 M-MOD 로 분석한다.

### [예시]

- (1) ‘확실히, 분명히, 아마도, 어쩌면’과 같이 명제의 확실성에 대한 화자의 태도를 나타내는 부사어나 ‘보아하니, 듣자하니’와 같이 정보의 출처를 나타내는 부사어를 M-MOD로 분석한다.

가. **분명히(M-MOD)** 내가 열쇠를 여기에 뒀어.  
나. **아마(M-MOD)** 그는 지금쯤 집에 도착했겠다.  
다. **듣자하니(M-MOD)**, 요즘 고민이 많다면서?

- (2) ‘다행히도, 불행히도, 놀랍게도, 고맙게도; 운 좋게도, 친절하게도’와 같이 사건이나 참여자에 대한 화자의 평가를 나타내는 부사어를 M-MOD로 분석한다.

가. **다행히(M-MOD)** 철수가 제시간에 나타났다.  
나. **친절하게도(M-MOD)** 그분들이 나에게 먼저 말을 걸어주었다.

- (3) ‘요컨대, 예컨대, 이른바, 이를테면’과 같이 발화행위와 관련되는 부사어와 ‘사실, 솔직히, 실은; 제발, 부디, 아무쪼록, 원컨대, 바라건대’와 같은 화자 지향 부사어나 청자 지향 부사어를 M-MOD로 분석한다.

가. **요컨대(M-MOD)** 이번 사건에 대한 재조사가 필요하다고 할 수 있다.  
나. **솔직히(M-MOD)** 무슨 말인지 하나도 모르겠어요.  
다. **제발(M-MOD)** 한 골만 더 넣어다오.

## 5.24. M-CNJ(Conjunction, 접속)

### [정의]

- 선행절과 후행절 사이의 의미적 관계(첨가, 열거, 요약, 병렬, 결과/결론, 대조/양보, 전환 등)를 나타내는 접속 부사를 M-CNJ 로 분석한다.

### [예시]

- (1) ‘그리고, 그러나, 그래서, 그러면, 그러므로, 그러니까, 그러다가’와 같이 접속 기능을 하는 ‘그’ 계열의 부사어나 그밖에 ‘곧, 또, 또한, 왜냐하면, 즉, 하지만’과 같이 절과 절을 이어 주는 접속 부사를 M-CNJ로 분석한다.

가. 그는 자리에서 일어났다. **그리고(M-CNJ)** 창문을 열었다.  
나. 모두가 축구를 했다. **그런데(M-CNJ)** 나는 자리에 앉아 있었다.  
다. 배가 아팠다. **하지만(M-CNJ)** 밥을 먹었다.

(2) 접속 부사가 절과 절의 경계가 아닌 절 내에 있다 하더라도 M-CNJ로 분석한다.

가. 네가 좋다면 나 **또한(M-CNJ)** 좋다.  
나. 나의 임무는 **즉(M-CNJ)** 그 집에 무사히 편지를 전달하는 것이다.

(3) 단일 어휘 체언 형태의 '결국', '첫째', '둘째', '우선', '다음' 등의 부사어 역시 M-CNJ로 분석한다.

가. 그는 결코 좌절하지 않았다. **결국(M-CNJ)** 그는 성공했다.  
나. **첫째(M-CNJ)**, 손잡이의 안전핀을 뽑는다. **둘째(M-CNJ)**, 바람을 등지고 선다.  
다. **우선(M-CNJ)**, 아이의 호흡을 살핀다. **다음(M-CNJ)**, 아이의 기도를 확보한다.

#### [주의]

- 단어와 단어를 이어 주는 접속 부사로 쓰인 경우 M-CNJ로 분석하지 않는다.

가. 나 그리고 너  
나. 원서 교부 및 접수  
다. 월요일 또는 수요일

## 6. 주의 사항

### 6.1. 의미역 분석 표지 부착 시 주의 사항

(1) 의미역은 특정 조사와 공기하는 경우가 많지만, 그렇지 않은 경우도 많으므로 주의가 필요하다. 언제나 해당 논항이나 부가어가 서술어에 대해 가지는 의미에 기반하여 의미역 표지를 결정하도록 한다.

가. 술이는 **꽃을(THM)** 꺾었다.  
나. **꽃이(THM)** 예쁘다.  
다. 여기 **꽃도(THM)** 팔아요?  
라. 이 집 **꽃(THM)** 얼마나 싱싱해요?

(2) 동일 문장 성분이라도 서술어의 성격에 따라 의미역이 달라지므로 이 또한 주의가 필요하다.

가. **철수는(AGT)** 영희를 안았다.  
나. **철수는(EXP)** 영희가 좋다.  
다. **철수는(MAG)** 영희를 좋아한다.

(3) 서술어가 동일하더라도 격 교체에 의해 의미역이 달라지는 경우도 있으므로 이 역시도 주의가 필요하다.

가. 술이는 **학교에(GOL)** 갔다.  
나. 술이는 **학교로(DIR)** 갔다.

다. 술이는 **학교를(THM)** 갔다.

- (4) 구어에서 발화 실수나 삽입구 등 서술어의 논항이 될 수 없다고 판단되는 대상은 작업 대상에서 제외한다.

가. 그런 말 그렇게 크게 사고를 친 것도 없고  
나. 집에 집이 엄격해서  
다. 배고픔 물론 막걸리가 배가 부른 것도 있었지만  
→ 작업 대상에서 제외한다.

## 6.2. 구문 분석 결과 검토 시 주의 사항

- (1) 구문 분석에서 기능 표지가 잘못되어 있으면 이를 수정한 후 의미역을 분석한다.

가. **파리에(NP\_AJT)** 에펠탑이 있다. (옳은 분석이므로 수정할 필요 없이 LOC로 분석.)  
나. **파리는(NP\_AJT)** 에펠탑이 있다. (옳은 분석이므로 수정할 필요 없이 LOC로 분석.)  
다. **파리는(NP\_SBJ)** 에펠탑이 있다. (잘못된 분석이므로 NP\_AJT로 수정한 후 LOC로 분석.)

- (2) 구어에서 격 표지와 서술어의 격틀이 맞지 않는 경우 서술어를 중심으로 논항 여부를 판단하고, 부가어에 대해서는 작업자의 판단하에 의미상 가장 가까운 표지를 부착한다.

- **저의 학창 시절은(M-BCK)** 초등학교 중학교 고등학교 그리고 이제 지금은 **대학교를(THM)** 다니고  
있고요.  
→ 서술어 '다니다'의 격틀 구조는 'X가(AGT) Y에(GOL)' 또는 'X가(AGT) Y를(THM)'이다.  
→ '저의 학창 시절은'은 보조사 '은'이 붙어서 NP\_SBJ로 처리되어 있지만 '다니다'가 요구하는 논항이  
아니다.  
따라서 여기서는 의미상 '배경'에 가깝다고 보아 M-BCK로 분석한다.

- (3) 구문 분석의 지침에 따르면, 모든 어절은 반드시 다른 성분에 의존해야 한다. 따라서 문장이 분절된  
경우, 특정 서술어의 논항이라고 볼 수 없는 성분들이 형식적으로 연결되어 있는 경우가 있다. 또  
논항이라고 하기보다는 '주제'에 해당하여, 서술어의 특정 논항이 중복 할당되는 것으로 보이는 경  
우도 있다. 이런 경우에는 작업자가 판단하에 작업 대상에서 제외한다.

가. **비바리움은** 사실 **내가(AGT)** 파충류를 키우다 보니까  
→ 서술어 '키우다'의 격틀 구조는 'X가(AGT) Y를(THM)'이다.  
→ 이때 '비바리움은'과 '내가' 모두 NP\_SBJ로 처리되어 있지만 이 모두에 대해 AGT를 줄 수는 없다.  
'비바리움'은 '주제부' 그 뒤에 오는 내용들은 '서술부'에 해당할 수도 있고, 뒤에 이어지는 문장의  
서술어에 부착되어야 하는 주어일 수도 있다.  
따라서 여기서는 '비바리움은'은 분석 대상에서 제외한다.

나. **웹툰을 본 사람들은 [약간 드라마가 망했다고 보는 사람들이](THM)** 많더라고.  
→ 서술어 '많다'의 격틀 구조는 'X가(THM) 많다'이다.  
→ 이때 '웹툰을 본 사람들은'과 '약간 드라마가 망했다고 보는 사람들이' 모두에 THM을 줄 수는 없다.  
'웹툰을 본 사람들은'은 '주제부' 그 뒤에 오는 내용들은 '서술부'에 해당한다.  
따라서 여기서는 '웹툰을 본 사람들은'은 분석 대상에서 제외한다.

- (4) 구문 분석 지침에 따르면, 명사구가 나열된 경우에는 선행 명사구가 마지막 명사구에 의존하는 것으로 되어 있다. 만약 명사구 각각이 서술어에 의존하는 것으로 분석되었다면, 선행 명사구가 마지막 명사구에 의존하도록 구문을 수정하여 작업한다.

- 그거 말고는 스릴러 공포 영화 같은 거 공포 드라마들 좋아해.  
 → 서술어 '좋아하다'의 지배를 받는 명사구가 '스릴러 공포 영화 같은 거', '공포 드라마들'로 되어 있다면, 선행하는 명사구가 '좋아하다'에 의존하는 것이 아니라 마지막 명사구에 의존하게끔 선행 명사구의 head 정보를 수정한다.  
 → 그거 말고는 [스릴러 공포 영화 같은 거 공포 드라마들]THM 좋아해.  
 이렇게 처리함으로써 THM의 중복역을 피할 수 있다.

단, 다음과 같은 경우에는 head 정보의 수정 없이 하나의 서술어에 대해 중복역을 줄 수 있다.

가. 이 강아지를(THM) 태평양(THM) 내가 키우면서  
 나. 담양에(LOC) 슬로우 시티 주변에(LOC) 큰 한옥이 있었는데  
 다. 서울에서 광주로(DIR) 광주 터미널로(DIR) 가선  
 → 구문 분석 지침에 따르면 조사가 결합한 명사구가 조사가 결합한 명사구에 의존할 수 없다. 따라서 구문을 수정할 수 없으므로 이런 경우에는 중복역을 허용한다.

- (5) 구문 분석 지침에 따르면, 모든 어절은 다른 성분에 의존한다. 따라서 어순이 도치된 문장인 경우 선행절의 서술어가 후행절의 서술어에 의존하도록 구문이 분석되어 있다. 그러나 이에 대한 의미역은 분석하지 않는다.

가. 조금 피곤하고 좀 쉬는 날이 많이 없었어요. 그래서 워라벨이 좋지 않아 가지고  
 → 선행절의 서술어 '없다'가 후행절의 서술어 '좋다'에 의존하고 있지만 이를 '좋다'의 논항 내지 부가어로 분석하지 않는다.  
 나. 뭐 안경도 구매할 수 있더라고요. 후기를 잘 써 주면  
 → 선행절의 서술어 '있다'가 후행절의 '쓰다'에 의존하고 있지만 이를 '쓰다'의 논항 내지 부가어로 분석하지 않는다.

## 6.3. 논항 여부 판별 및 격틀 구조 분석 시 주의 사항

- (1) 서술어에 대한 격틀 구조는 <우리말샘>의 어휘의미 번호에 제시된 대로 따라야 한다. 그러므로 절대로 자의적으로 격틀 구조를 상정해서는 안 되며, 논항으로 처리해야 할 것들을 부가어로 처리하지 않도록 주의한다.

가. 파운드리 시장에서(M-EXT) 타워의 시장 점유율은 1.3%로 미미하다. (LOC로 처리하면 안 된다.)  
 나. 샌디에이고는 샌프란시스코자이언츠와(M-COM) 원정경기를 치렀다. (COM으로 처리하면 안 된다.)

- (2) 이와는 반대로 <우리말샘>의 격틀 구조에 따라 부가어로 처리해야 할 것들을 논항으로 처리하지 않도록 주의한다.

가. 그러니까 총인원이 한 170명 정도(FNS) 된 거야.  
 ('되다'의 논항이므로 M-QNT로 처리하면 안 된다.)  
 나. 거의 1400에(CRT) 가까운 거금을 모았다.  
 ('가깝다'의 논항이므로 M-QNT로 처리하면 안 된다.)

다. 의정부 시청팀은 **박성현·문현웅·정재원 선수로(INS)** 구성되었다.  
 ('구성되다'의 논항이므로 M-INS로 처리하면 안 된다.)  
 라. 유가는 **실적예(GOL)** 상당한 영향을 미친다.  
 ('미치다'의 논항이므로 M-LOC로 처리하면 안 된다.)

(3) <우리말샘>의 격틀 구조를 살폈다고 하더라도 예문에 부합하지 않는 것을 논항으로 처리하지 않도록 주의한다.

가. 문제 행동을 **영상으로(M-INS)** 기록했다.  
 ('기록하다001'의 격틀 구조가 [X를 Y로 기록하다]인 것은 맞지만, 예문은 '역사는 그를 위대한 스승으로 기록했다.'처럼 '~으로'는 QLF 내지 FNS에 해당하므로 가.의 '영상으로'는 M-INS이다.)  
 나. 원작 웹툰이 **시즌3로(M-QLF)** 돌아온다.  
 ('돌아오다001'의 격틀 구조가 [X가 Y로 돌아오다]인 것은 맞지만, 예문은 '아들이 고향으로 돌아왔다.'처럼 '~으로'는 DIR에 해당하므로 나.의 '시즌3로'는 M-QLF이다)

(4) 어휘의미 번호가 표시하는 격틀 구조에 부합하지 않음에도 분명히 논항으로 판단되는 것들이 있다면 어휘의미 번호가 틀렸을 가능성이 크다. 그럴 경우에는 어휘의미 번호를 수정하여 부가어를 논항으로 처리한다.

가. 나는 거기는 **사진을 찍으러 오는 데로(FNS)** 알고 있다.  
 → '알다001'[X가 Y을 알다]가 아닌 '알다011'[X가 Y을 Z으로 알다.]로 어휘의미 번호를 수정한 후 부가어(M-LOC)가 아닌 논항(FNS)으로 처리한다.  
 나. 직접 **신은 사람들에게(GOL)** 들어 보면  
 → '듣다001'[X가 Y을 듣다]가 아닌 '듣다005'[X가 Y에게 Z을 듣다]로 어휘 번호를 수정한 후 부가어(M-EXT)가 아닌 논항(GOL)로 처리한다.  
 다. **가평에(LOC)** 큰 카페가 있다.  
 → '있다001'[X가 있다]가 아닌 '있다011'[X가 Y에 있다]로 어휘 번호를 수정한 후 부가어(M-LOC)가 아닌 논항(LOC)로 처리한다.

(5) 어휘의미 번호가 틀리지 않았더라도 논항이라고 판단되는 경우에는 어휘의미 번호를 888로 수정하여 부가어를 논항으로 처리한다.

가. 연구팀은 **전기 신호를 차단할 수 있는 것으로(FNS)** 확인했다.  
 → '확인하다001'[X가 Y을 확인하다]로 FNS를 요구하지는 않지만, FNS로 판단되므로 확인하다888로 수정한 후 부가어(M-MNR)가 아닌 논항(FNS)으로 처리한다.  
 나. 정부는 **재건축 단지의 고민이 해소될 것으로(FNS)** 기대했다.  
 → '기대하다001'[X가 Y을/기를 기대하다]로 FNS를 요구하지는 않지만, FNS로 판단되므로 기대하다888로 수정한 후 부가어(M-MNR)가 아닌 논항(FNS)으로 처리한다.

## 6.4. 부사절 의미역 부착 시 주의 사항

(1) 명사구가 부사어로 기능하는 경우 적절한 조사를 넣어보거나 해당 구성을 어미로 교체해 본 후 서술어와의 의미 관계에 따라 의미역을 부착하도록 한다.

가. **신고 포상제를 실시한 결과(M-CAU)** 쓰레기 불법 투기가 줄어들었다. (실시한 결과로, 실시해서)



나. 검사를 한 결과(M-BCK) 유해 성분은 검출되지 않았다고 밝혔다. (검사를 했는데)  
다. 장화홍련전의 인물 설정을 차용(M-LST) 통과한 복수 이야기를 노랫말에 담아냈다. (차용하여)

- (2) 명사+조사 구성이 후행절에 대한 선행절로 기능하는 경우 해당 명사의 어휘적 성격을 살펴 의미역을 부착하도록 한다.

가. 주말 체험 영농을 목적으로(M-PRP) 취득한 농지는 제외된다.  
나. 산사태 위험을 이유로(M-CAU) 입산객들의 통행을 금지했다.  
다. 경찰은 범죄자들을 구속하는 동시에(M-LST) 자금의 출처를 파악하고 있다.  
라. 봉쇄를 중심으로(M-MNR) 하는 중국의 코로나 방역 대책이 논의되었다.

## [붙임5] 무형 대용어 복원 말뭉치 구축 지침

### 무형 대용어 복원 말뭉치 구축 지침

#### 1. 총론

##### 1.1. 지침의 목적

이 지침은 한국어 문장에서 생략된 성분을 복원한 무형 대용어 복원 말뭉치를 구축하는 방안을 기술하는 것을 목적으로 한다. 서술어가 요구하는 필수 성분이 생략되어 있을 때 이를 복원하는 것이다. 무형 대용어 복원은 구문 분석 말뭉치와 어휘의미 분석 말뭉치를 대상으로 수행한다.

- 본 지침은 2019년 주격 무형 대용어 복원 지침, 2020년 목적격 무형 대용어 복원 지침, 2024년 무형 대용어 복원 말뭉치 구축 지침을 통합하여 작성한 것이다. 2019년에는 주어(SBJ)를, 2020년에는 목적어(OBJ)를, 2024년에는 주어(SBJ), 목적어(OBJ), 보어(CMP), 필수 부사어(AJT)를 복원한 말뭉치를 구축한 바 있다.
- 본 지침은 이전 지침과의 내용적 일관성을 최대한 유지하되 층위별 말뭉치 구축을 염두에 두고 지침을 개정·보완하였다.

##### 1.2. 용어 정의

- 서술어: 한 문장에서 주어의 움직임, 상태, 성질 따위를 서술하는 말. 서술어는 주로 동사, 형용사, 서술격 조사로 나타나지만, 이 사업에서는 본용언으로 기능하는 동사와 형용사에 대해서만 무형 대용어 복원을 함.
- 복원 기준 서술어: 무형 대용어 복원의 기준이 되는 서술어로, 본용언으로 쓰인 동사와 형용사.
- 무형 대용어: 문장에서 서술어가 요구하는 필수 성분 중 생략된 성분. (= 생략 성분)
- 동일지시적 성분: 실제 문장에서 실현되지 않은 필수 성분과 지시하는 대상이 동일한 성분. 무형 대용어가 될 후보적인 성분. 대체로 선행 문장에서 찾을 수 있으며, 드물게 후행 문장에서 찾을 수 있는 경우도 있음.
- 복원 성분: 선행 문장 혹은 후행 문장을 참고하여 격틀에 맞게 복원한 성분.

### 1.3. 기본 원칙

- (1) 자연어 처리를 위한 일관성 유지와 효율성 제고에 초점을 두되, 일반 언어학적 관점에서 크게 벗어나지 않도록 생략된 성분을 복원한다.
- (2) 의존 구문 분석 말뭉치 구축을 위한 지침과 최대한 일관성을 유지하며, 두 층위에서 기술적으로 다를 수밖에 없는 부분은 지침에 명시한다.
- (3) 무형 대용어 복원은 '무형 대용어 복원 대상 서술어 탐지 → 동일지시적 성분 탐색 → 적절한 무형 대용어 복원'으로 진행된다.

## 2. 무형 대용어 복원의 원리

### 2.1. 복원 기준 서술어 확인

#### 2.1.1. 복원 기준 서술어

- 문장 내 모든 본용언은 복원 기준 서술어이다. 본용언으로 쓰인 동사와 형용사가 이에 해당된다.(보조용언은 복원 기준 서술어에서 제외한다.)
- 연속해서 나타난 서술어가 사전에 등재되지 않은 경우, 구문 분석의 결과를 따라 선행 용언과 후행 용언 각각을 모두 대상 서술어로 취급한다.
  - 자동차가 로터리를 돌아 나갔다.  
→ '돌아 나가다'는 <우리말샘>에 등재되어 있지 않고 구문 분석 결과 각각 VP 로 분석되기에 각각 대상 서술어로 취급하여 '돌다'와 '나가다'의 무형 대용어를 모두 복원한다.
  - 학생이 교무실에서 성적표를 챙겨 갔다.  
→ '챙겨 가다'는 <우리말샘>에 등재되어 있지 않고 구문 분석 결과 각각 VP 로 분석되기에 각각 대상 서술어로 취급하여 '챙기다'와 '가다'의 무형 대용어를 모두 복원한다.
- 가운뎃점, 쉼표, 괄호에 병기 등의 방법으로 서술어가 여러개가 나타나는 경우 구문 분석의 결과에 따라 분석한다.
  - 선수들을 보자 그는 일일이 악수·격려하며 기념사진까지 찍었다.  
→ 위의 서술어에서 구문 분석의 결과로 '격려하다'만 VP 로 분석되기에 '격려하다'의 격틀인 'X가 Y를 격려하다'로 복원한다.
  - 선수들을 보자 그는 일일이 악수, 격려하며 기념사진까지 찍었다.  
→ 위의 서술어에서 구문 분석의 결과로 '격려하다'만 VP 로 분석되기에 '격려하다'의 격틀인 'X가 Y를 격려하다'로 복원한다.

- 사전에 등재되지 않아 '777'의 번호를 부여받은 서술어도 복원 기준 서술어로 취급한다. 다만 격틀은 아래와 같은 방식을 따른다.
  - 'N 하다'에 대응하는 'N 받다/되다/당하다', 'V 다'에 대응되는 'V 어지다'의 경우 능동문의 격틀을 고려하여 판단한다.
    - 암에 걸려 치료받은 환자 중 완치된 사람들은 주치의를 믿고 긍정적인 생각을 가졌던 사람이 많다.
      - '치료하다'(X가 Y를 치료하다)를 고려하여 '치료받다'의 격틀을 'Y가 X에게 치료받다'로 분석. 피동문에서 '에게' 논항은 필수적이지 않기에 'Y가'만 복원함.
    - 그는 보병으로 배치받았다.
      - '배치하다'(X가 Y를 Z에 배치하다)를 고려하여 '배치받다'의 격틀을 'Y가 X에 의해 Z에 배치받다'로 분석. 피동문에서 '에 의해' 논항은 필수적이지 않기에 'Y가 Z에'를 복원함.
  - 'N 하다'에 대응되는 'N 시키다'의 경우 사동문과 'N 하다'와 동의어인 경우로 구분하여 복원한다.
    - 회사는 4명의 노동자를 복직시켰다.
      - 이때의 '복직시키다'는 사동의 의미를 지니고 있기에 격틀을 'X가 Y를 Z에 복직시키다'로 분석하여 복원함.
    - 작업자가 호스를 연결시켰다.
      - 이때의 '연결시키다'는 사동의 의미보다는 '연결하다'의 동의어인 것에 해당하기에 격틀을 '연결하다'의 격틀과 동일하게 'X를 Y에 연결시키다'로 분석하여 복원함.
  - 그 외 미등재어의 경우 맥락을 고려하여 의미가 유사한 서술어의 격틀을 고려하여 복원하는 것을 원칙으로 한다.
    - 골프는 홀릭되는 매력이 있다.
      - '홀릭되다'의 의미는 '중독되다'와 유사하다고 판단하여 격틀을 '중독되다'의 격틀인 'X가 Y에 홀릭되다'로 분석하여 복원함.
  - 띄어쓰기 오류로 인해, 하나의 서술어가 둘 이상으로 나뉘는 경우 원어절을 수정한다.
    - 산불로 사업용 자산 등을 상실했다면 소득세·법인세를 공제하고 사업상 어려움이 있다면 연말까지 세무조사 등도 연기한다.
      - '상실하다'로 수정 후 복원
    - 요즘은 먹고 사는 게 힘들다.
      - '먹고살다'로 수정 후 복원

## 2.1.2. 복원 제외 서술어

- 보조 용언은 복원 기준 서술어에서 제외한다. 보조 용언 목록은 <우리말샘>에 등재된 보조용언 목록

록을 따른다.

-어 가다	-지 말다	-ㄹ 뻔하다	-어 젖히다
-어 가지다	-고(야) 말다	-게 생기다	-어 주다
-어 나가다	-어 말다	-ㄹ 성부르다	-어 죽다
-어 나다	-어 먹다	-ㄴ/ㄹ 성싶다	-ㄴ/음 직하다
-고 나다	-지 못하다	-ㄴ/ㄹ 성하다	-은/는 척하다
-어 내다	-다(가) 못하다	-고 싶다	-은/는 체하다
-어 놓다	-어 버릇하다	-ㄹ까 싶다	-어 치우다
-어 대다	-어 버리다	-었으면 싶다	-어 터지다
-어 두다	-ㄹ 법하다	-어 쌓다	*-게 하다
-어 드리다	-어 보다	-지 않다	-었으면 하다
-려고/-기로/-자고 들다	-고 보다	-고 앓다	-어야 하다
-고 들다	-다(가) 보다	-ㄴ/는 양하다	-으려(고)/고자 하다
-어 들다	-ㄴ 가/나 보다	-어 오다	-기는/기도/기나 하다
-ㄴ/ㄹ 듯싶다	-을까 보다	-어 있다	-고 하다
-ㄴ/ㄹ 듯하다	-다 보다	-고 있다	-고는/곤 하다
-ㄹ 만하다	-어 빠지다	-어 재끼다	-기는/기도/기나 하다

→ ‘-게 하다’는 보조 용언이므로 분석 대상에서 제외되어야 하지만 구문 분석에서 사동주의 지배소가 ‘하다’로 분석되므로 주어가 생략된 경우에 한하여 사동주만을 복원한다.

- 본용언, 보조용언 구분에 유의해야 한다. 용언 두 개가 연속해 나타날 때 선행 용언 뒤에 ‘서’를 붙일 수 있으면 후행 용언을 본용언으로 판단하고, ‘서’를 붙일 수 없으면 후행 용언을 본용언이 아닌 것으로 판단한다.

○ 친척들을 모아 놓고 자신의 재산을 전부 나눠 주었다.

→ ‘나눠서 주다’가 성립 가능하니 이때의 ‘주다’는 본용언이기에 복원 대상 서술어이다.

- 의사 보조용언 구성은 복원 기준 서술어에서 제외한다.

-ㄹ 수 있다/없다	-ㄴ/ㄹ 예정이다
-ㄹ 리(가) 있다/없다	-ㄴ/ㄹ 참이다
-ㄴ/ㄹ 것이다	-ㄴ/ㄹ 터이다
-ㄴ/ㄹ 길이다	-ㄴ/ㄹ 것 같다
-ㄴ/ㄹ 노릇이다	-ㄹ 것(을/걸) 그랬다
-ㄴ/는 말이다	-어서는 안 되다
-ㄴ/ㄹ 모양이다	-든지 하다
-ㄴ/ㄹ 뿐이다	

- 주로 부사형, 관형사형으로 쓰여 문장에서 독립절을 구성하기 어려운 술어는 복원 기준 서술어에서 제외한다.

- 부사형

가지다007 가지다008	~를 가지고[=로]	밀가루를 가지고 만든 떡 한 가지 일을 가지고 너무 오래 끌지 마라.
각설하다003	각설하고	자, 각설하고, 어디 당신 계획이나 들어봅시다.
같다004	같으면, 같아서, 같으니	옛날 같으면
같다006	같으면, 같아서, 같으니	마음 같아서는
같다008	같으면, 같아서, 같으니	바보 같으니라고
거두절미하다002	~을 거두절미하고	거두절미하고 용건만 말씀드리겠습니다.
걸치다002	~에 걸쳐	열 시간에 걸쳐 회의가 진행됐다.
관련되다001	~와/에 관련되어	적지 않은 사람들이 이번 사건과 관련돼 있다.
관련하다001	~와/에 관련해	교육 개혁안과 관련해 논평을 했다.
관하다002	~에 관해	다음은 여성의 사회적 지위에 관하여 토론하도록 하겠습니다.
그렇게001	그렇게	왜 그렇게 성을 내니?
기하다005	~을 기해	오늘 새벽 4시를 기해 폭풍 주의보가 발효되었다.
까닥하다002	까닥하면/까닥하다가는[추측]	까닥하다가는 당신 목숨까지 위태로울 수 있습니다.
깍하다001	깍해야	있는 돈을 다 모아도 깍해야 3천 원이다.
내놓다012	내놓고	내놓고 잘난 체하다.
까놓다001	까놓고 말해	어디 특 까놓고 말을 해 봐!
넉넉잡다001	넉넉잡고/넉넉잡아	넉넉잡고 사흘이면 일이 끝난다.
놓다005	~을 놓고[=에 대해]	동문회에서 학교 이전 문제를 놓고 의견이 분분했다.
대다002	~에 대고	하늘에 대고 하소연을 했다.
대하다004	~에 대해	전통문화에 대한 관심
더불다001~003	~와 더불어	이웃과 더불어 사는 세상
두다014	~을 두고	머리를 북쪽으로 두고 자면 빨리 죽는다는 말이 있다.
두르다011	둘러 말하다	자꾸 둘러 말하지 말고 직설적으로 말해라.
둘러싸다003	~를 둘러싸고[=에 대해]	이 문제를 둘러싸고 의견이 분분하다.
뒤잇다001	~에 뒤이어	축가에 뒤이어 축사가 있겠습니다.
들다013	[특정 시기] 들어/들자	올해 들어 해외여행자 수가 부쩍 늘었다.
따르다004	~을 따라	그는 아버지의 뜻을 따라서 법대에 진학했다.
따르다008	~에 따른	사용 목적에 따른 분류.
말다006	말고	너 말고 영희
말하다007	말하자면/ 소위, 다시 말해	말하자면 그것은 커다란 충격이었고 환멸이었다.
말하다008	~로 말하면	힘센 걸로 말하면 우리 아버지 따라갈 사람이 없다.
말미암다001	~로 말미암아	난데없는 구렁이의 출현으로 말미암아 우리 집은 삼시에 엉망진창이 되어 버렸다.
머지않다001	머지않아	머지않아 소식이 올 것이다.
못하다003	못해도 얼마	잡은 고기가 못해도 열 마리는 되겠지.
무작정하다001~002	무작정하고	무작정한 낙관주의.
무조건하다001	무조건하고	의기에 감동되어 무조건하고 곡식과 피륙을 바치는 사람들도 많았다.

밤새다001	밤새도록	밤새도록 일하다.
보다025	~를 보고[=에게/에 대해]	너를 보고 하는 말이 아니야.
불구하다002	불구하고	몸살에도 불구하고 출근하다.
비롯하다003	비롯해	이장을 비롯한 마을 사람들.
비하다002	비해	그는 사진에 비해서 실물이 훨씬 더 좋은 인상을 풍겼다.
비하다003	비하건대	낭비가 심한 그에게 돈을 주는 것은, 비하건대 밑 빠진 독에 물 붓기다.
시작하다001	~기 시작하다	차가 움직이기 시작했다.
아니다001	아니라/아니면	철수 아니라 영희/철수 아니면 영희
아무렇다002	아무렇게나	옷을 아무렇게나 벗어 놓았다.
암만하다001~002	암만해도	암만해도 알 수가 없다.
어째서001	어찌하여서, 어째서	너희들은 어째서 만나기만 하면 싸우냐?
얼추잡다001	얼추잡아	지각한 놈들만도 얼추잡아 여남은 명은 되었다.
오다015	[시간 표현] (에) 와서	지금 와서 계획이 취소되었다고 연락을 받았다.
위시하다002	~을 위시하여	그를 위시하여 여러 사람이 그 일에 참여했다.
위하다003	~을 위해	시장 조사를 위해 현지 출장을 갔다.
의하다001	~에 의하면/의해	노동에 의해 소득을 얻었다.
이렇게001	이렇게	그는 웃으며 이렇게 말했다.
인하다001	~로 인해	부주의로 인한 사고
잇다004	이어/이어서	개회사에 이어 회장의 인사 말씀이 있겠습니다.
잇달다001~002	잇달아	유세장에 유권자들이 잇달아 몰려들었다.
잇따르다001~002	잇따라	잇따른 범죄 사건 때문에 밤길을 다니기가 두렵다.
연달다001~002	연달아	차가 연달아 온다.
연잇다001	연이어	영화 상영이 끝나자 사람이 연이어 극장 문을 나온다.
있다019	~에/에게 있어	인간에게 있어서 가장 중요한 것은 사랑이다.
자시다002	~고 자시고	누군들 그 경황에 누굴 알아보고 자시고 할 사이가 있었겠나.
작다007	작게는	전쟁은 작게는 개인의 생활에서 크게는 국가의 운명까지도 좌우한다.
잘못하다004	잘못하다가는	잘못하다가는 남편의 생명도 부지하지 못할 모양이다.
잘못하다005	잘못하면	눈치를 보니 자칫 잘못하면 큰 싸움이 날 것 같아 나는 그 둘의 중재에 나섰다.
잘하다007	잘하면	잘하면 올해도 풍년이 들겠다.
잘하다008	잘해서/잘해야[추측]	잘해야 열 사람 중에서 한두 사람쯤 살아날 수 있는데.
저렇게001	저렇게	뭘 믿고 저렇게 큰소리를 탕탕 치는 건고?
줄잡다002	줄잡아	이번 출장은 줄잡아 한 달은 걸릴 것 같다.
즈음하다001	즈음하여/즈음해서	본 여단은 출동에 즈음해서 미증의 곤란을 겪고 있습니다.

참고삼다002	참고삼아	이건 어디까지나 참고삼아 말씀드리는 겁니다.
추후하다001	추후하여	자세한 의견은 추후하여 말씀드리겠습니다.
치다047	~로 치면/친다면	종갓집 형은 촌수로 치면 팔촌 형이고 연배로 치면 20여 세 위였다.
크다011	크게	교육을 크게 보면 국가와 사회를 위해 시키는 것이고 작게 보면 자기 자신을 위해 하는 것이다.
통하다011~014	~을 통해	비상구를 통해 빠져나간다.
폐일언하다001	~을 폐일언하고	폐일언하고 당장 시작합시다.
하다019	~로 해서[=로]	우리는 부산으로 해서 일본에 갈 생각이다.
하다021	하면	물 하면 역시 우리나라 물이 최고다.
하다022	하고	내 친구는 이번에 서울 하고도 강남에 집을 샀다.
하다030	하면	그는 쳇다 하면 홈런이다.
하다033	하니/하면/하여/해서	꾸준히 연습하여라. 하면 어느 순간에 실력이 늘어 있을 것이다.
하다888	~만 해도[가정해도/따져봐도]	철수만 해도 그 일을 잘한다.
향하다002~004	~을 향해	그의 모든 관심과 열정은 영화를 향해 있다.
허실삼다001	~을 허실삼아	행여 무엇이나 입에 맞는 떡이 있을까 하는 생각으로 허실삼아 가 보았다.

• 관형사형

갈데없다001	갈데없는	갈데없는 처지/신세
같다005	같은	말 같은 말
같다888	~와 같은 경우	저와 같은 경우 인터넷으로 정보를 찾습니다.
걸치다002	~에 걸친	이틀에 걸친 전투는 끝이 났다.
관련되다001	~와/에 관련된	마약과 관련된 범죄
관련하다001	~와/에 관련한	교육 개혁안과 관련한 논평
관하다001	~에 관한	그 문제에 관한 한 우리는 한 치도 양보할 수 없습니다.
그러하다002	그러한	별일 없는 그냥 그러한 일상
대하다004	~에 대한	신탁 통치안에 대한 우리 민족의 반대 운동은 전국적이었다.
둘러싸다003	~를 둘러싼[=에 대한]	강도 사건을 둘러싼 사람들의 반응은 다양했다.
뒤잇다001	~에 뒤이은	브랜드 확장에 뒤이은 공격적 마케팅
따르다004/008	~에 따른	사용 목적에 따른 분류.
말미암다001	~로 말미암은	달균이가 돈을 벌어 오겠다고 절 골을 뜯는 것도 그로 말미암은 가난 때문이었다.
불구하다002	불구한	재개봉임에도 불구하고 폭발적 인기
비롯하다003	비롯한	이장을 비롯한 마을 사람들.
비하다002	~에 비한	여름철, 남들에 비한 과한 땀
아니다001	아닌	철수 아닌 영희
아무려하다002	아무려한	그는 꿀 먹은 벙어리처럼 아무려한 대답도 못 하고 있었다.
어떠하다001	어떠한	어떠한 돈으로 이 물건을 샀느냐?



여하하다001	여하한	여하한 어려움도 견딜 수 있다.
어쩌다002	어쩐	어쩐 일로 전화하셨소?
오다019	오는	오는 일요일이 내 생일이다.
위시하다002	~을 위시하여	황제는 대답 대신, 우발산과 변약유를 위시한 자기 사람들을 불러 모았다
위하다003	~을 위한	시장 조사를 위한 현지 출장
의하다001	~에 의한	노동에 의한 소득
이러하다001	이러한	이러한 경우
인하다001	~로 인한	부주의로 인한 사고
잇다004	이은	이 일은 업무 협약에 이은 일이다.
잇달다001~002	잇단	잇단 범죄 사건.
잇따르다001/002	잇따른	잇따른 범죄 사건.
연잇다001	연이은	하늘과 연이은 수평선.
저러하다001	저러한	저러한 주장은 사회적으로 큰 물의를 일으킬 수도 있다.
지나다888	지난	지난 10에 만난 사람
지나가다003	지나가는	지나가는 말
즈음하다	즈음한	어린이날 즈음한 성명서 발표
하다024	다고 하는	거짓말을 했다고 하는 증거
향하다002~004	~을 향한	그는 아까부터 그녀에게 시선을 향한 채 꿈쩍도 하지 않고 있다.
현행하다001	현행하는	세무 공무원이 현행하는 조세법을 몰라서야 말이 되 는가.

- 해당 의미번호가 아닌 다의어는 복원 대상이 됨을 유의한다. 예컨대 ‘대(對)하다’의 의미 중 배제 대상이 되는 것은 ‘004 【…에】 ((‘대한’, ‘대하여’ 꼴로 쓰여)) 대상이나 상대로 삼다.’만이다.

대-하다(對하다)
대-하다 「002」 「동사」 【…을】 【(…과) …을】 ((‘…과’가 나타나지 않을 때는 여럿임을 뜻하는 말이 주어로 온다)) 마주 향하여 있다.
대-하다 「003」 「동사」 【…에/에게 -게】 【…을 …으로】 【…을 -게】 ((‘…으로’나 ‘-게’ 성분은 ‘…처럼, -은/을 뜻이’ 따위의 부사어나 ‘-이/히’ 부사로 대체될 수 있다)) 어떤 태도로 상대하다.
대-하다 「004」 「동사」 【…에】 ((‘대한’, ‘대하여’ 꼴로 쓰여)) 대상이나 상대로 삼다.
대-하다 「005」 「동사」 【…을】 작품 따위를 직접 읽거나 감상하다.

- 단독으로 나타나거나 강조부사만 동반하여, 감탄사처럼 쓰이는 서술어는 복원 기준 서술어에서 제외한다.
  - 상대의 말 또는 자신의 말을 부정하는 의미만을 지닌 ‘아니다’
    - “너 그 말 거짓말이지?” “아니예요. 저는 거짓말 안 해요.” (단독 실현, 상대의 말 부정)
    - “철수가 그렇게 말한 거 맞다니까요.” “진짜 아니예요.” (강조부사 실현, 상대의 말 부정)
    - 공급이 수요를 못. 아니지. 수요가 공급을 못 따라가. (단독 실현, 자신의 말 부정)
  - 상대의 말에 동의하는 의미로 쓰이는 ‘맞다’, ‘좋다’ 등

- “그거 연예인들이 이야기하는 방송이잖아.” “맞아.” (단독 실현, 상대의 말에 동의)
  - “그 사람 실력이 어마어마한 것 같아요.” “완전 맞아요.” (강조부사 실현, 상대의 말에 동의)
  - “우리 좀 있다가 밥 먹으러 갈까?” “좋지.” (단독 실현, 상대의 말에 동의)
  - “우리 피자 먹으러 갈까?” “너무 좋아.” (강조부사 실현, 상대의 말에 동의)
- 문두에 나타난 ‘있지/있잖아’, ‘가만있자/가만있어라’, ‘어디보자’, ‘죽어라/죽자’, ‘아니’ 등
  - 선행 맥락 전체를 지시하는 ‘그렇다’, ‘그러다’는 복원 기준 서술어에서 제외한다.
    - 요즘 젊은 층은 SNS 를 많이 사용합니다. 그렇기 때문에 SNS 마케팅에 주목해야 합니다.
    - 그런 룸을 방문하는 사람은 푸드 바이어 아미르(알렉스 울프)뿐이다. 그러던 어느 날 낯선 이들이 숲에 들이닥쳐 룸을 폭행한 후 소중한 돼지를 훔쳐가는 사건이 발생한다.
  - 책 제목, 노래 제목, 프로그램명 등과 같은 고유명에 포함된 서술어는 복원 기준 서술어에서 제외한다.
    - 놀면 뭐하니(프로그램명)
    - 지금은 맛고 그때는 틀리다(영화 제목)
  - 서술성 명사는 복원 대상 서술어에서 제외한다.

## 2.2. 필수 성분 복원 절차

### 2.2.1. 생략 성분의 확인

- 문장 내에서 구문 분석 결과 VP 로 태깅된 서술어를 찾아, 해당 서술어의 필수 성분 중 생략된 성분이 있는지 확인한다. 생략된 성분 중 주어(SBJ), 목적어(OBJ), 보어(CMP), 필수 부사어(AJT)가 복원 대상이 된다.
- 서술어가 필수적으로 요구하는 성분의 확인은 <우리말샘>에서 제시한 격틀을 기반으로 한다. 단, 주어는 필수적인 성분이기 때문에 <우리말샘>에 격틀이 제시되어 있지 않지만, 언제나 복원 대상인 것으로 상정한다.
- 서술어가 여러 격틀을 지닌 다의어인 경우, 어휘의미 분석 결과에 기반하여 격틀을 선택한다.
- 하나의 서술어가 여러 개의 격틀을 가질 때 어떤 격틀을 바탕으로 생략 성분을 복원할 것인지 결정한다. 이때, 동일지시적인 성분과 같은 성분으로의 복원을 우선한다.
  - 부모님은 방학이면 나를 시골로 보내곤 했다. 그렇게 나를 보내고 당신들끼리 해결해야 하

는 일들을 했다는 것을 그때는 몰랐다.

→ '보내다'는 'X가 Y를 Z에 보내다'와 'X가 Y를 Z로 보내다'가 가능하다. 이때 동일지시적인 성분과 같은 성분으로 복원을 우선하는 것을 고려하여 '시골로'로 복원한다.

- 격 교체 현상을 인정하여, 아래와 같은 경우는 격들이 충족된 것으로 보아 복원하지 않는다.

- ① 가-로: 하위 팀들이 좋은 성적을 거둘 것이/것으로 예상된다
- ② 가-를: 철수가 손이/손을 잡혔다
- ③ 가-에/에게: 천장이/천장에 손이 닿는다
- ④ 가-에서/에게서: 아이가 잠이/잠에서 깬다
- ⑤ 가-와: 나는 그 자리에 있는 모두가/모두와 서먹서먹했다
- ⑥ 가-의: 코끼리가/코끼리의 코가 길다
- ⑦ 를-게-도록: 아이들에게 청소를/청소하게/청소하도록 시켰다
- ⑧ 를-고: 인터뷰에서 관계자는 보다 나은 개혁안을 준비하고 있음을/준비하고 있다고 밝혔다
- ⑨ 를-로: 그는 죽기를/죽기로 각오했다
- ⑩ 를-보다: 금년 소출은 평년작을/평년작보다 밀돌았다
- ⑪ 를-에/에게: 그가 학교를/학교에 갔다, 어머니가 아이를/아이에게 선물을 주었다
- ⑫ 를-에서/에게서: 그가 공원을/공원에서 거닐었다
- ⑬ 를-와: 그가 철수를/철수와 만났다
- ⑭ 를 와-와 를: 내가 쓴 문장을 다른 사람의 문장과/내가 쓴 문장과 다른 사람의 문장을 건주었다
- ⑮ 를-의: 영수는 말을/말의 등을 두들겼다
- ⑯ 에/에게-에서/에게서-로: 열심히 공부한 결과가 성적에/성적에서/성적으로 나타났다
- ⑰ 에/에게-에서/에게서-보다: 철수는 영희에게/영희보다 뒤진다, 그는 앞서가는 대열에서/대열보다 조금 뒤져 있었다
- ⑱ 에/에게-에서/에게서-와: 우리 집은 학교에/학교에서/학교와 가깝다
- ⑲ 에/에게-로-게(도록, 려고, 라고): 좋은 사람이 되기에/되는 쪽으로/되게 애쓴다
- ⑳ 에/에게 가-가 로: 창틀에 먼지가/창틀이 먼지로 가득하다
- ㉑ 에/에게 를-를 로: 청소부들이 건물벽에 페인트를/건물벽을 페인트로 칠했다
- ㉒ 로-와: 몰래 일본으로/일본과 밀무역했다
- ㉓ 로-로서: 그것은 교사로/교사로서 할 일은 아니다

## 2.2.2. 생략 성분과 동일지시적인 성분 탐지

- 복원 기준 서술어가 속한 문장 내에 생략된 성분과 일치하는 동일지시적 성분이 존재하면 해당 성분으로 우선적으로 복원한다.
- 복원 기준 서술어가 속한 문장 내에 생략된 성분과 일치하는 동일지시적 성분이 존재하지 않으면

선행 문장에 위치한 성분으로 복원한다.

- 복원 기준 서술어가 속한 문장이나 선행 문장에 생략된 성분과 일치하는 동일지시적 성분이 존재하지 않으면 후행 문장에 위치한 성분으로 복원한다.
- 선행 문장이나 후행 문장에서 논항이나 부가어로 실현된 명사구를 동일지시적 성분으로 삼되, 경우에 따라서 VNP 에 포함된 명사구도 동일지시적 성분으로 선택할 수 있다.
  - 저한테는 참 신기했던 경험이었던 거 같습니다.  
→ ‘신기하다’의 주어로 ‘경험이’를 복원하고, 관계 관형절의 피수식어인 ‘경험’을 포함한 ‘경험이었던’을 동일지시적 성분으로 삼음.
- 생략된 성분을 복원할 수 있는 표현이 복원 범위 내에 존재하지 않으면 비지시적 대명사로 복원한다. 이때 복원 범위는 분석 대상 서술어가 포함된 앞 15문장 / 뒤 2문장 이내로 규정한다.

## 2.2.3. 필수 성분 복원

### 2.2.3.1. 복원 단위

- 명사 혹은 명사구에 조사가 결합된 형태로 복원한다.

생략어의 문장 성분	생략어 복원 형태
주어	NP이/NP가
목적어	NP을/NP를
보어	NP이/NP가
부사어	NP에, NP에게, NP에서, NP에게서, NP로/NP으로, NP와/NP과, NP보다

- 명사, 대명사, 수사는 모두 복원될 수 있다.
- 문장에서 독립적으로 쓰일 수 없는 의존명사도 복원될 수 있다. 단, ‘것’의 경우 선행하는 관형절 서술어의 논항이 되는 ‘것’과 명사화소 ‘것’을 구분하여야 한다. 의존명사 ‘것’ 앞에 지시 관형사를 넣어 보아 자연스러우면 서술어의 논항인 ‘것’으로, 부자연스러우면 명사화소인 ‘것’으로 구분할 수 있다. 서술어의 논항인 ‘것’을 수식하는 관형절 내의 생략 성분에 대해서는 ‘그것’으로 복원한다.
  - 이 빵집은 [(누군가가) 쌀가루로 (그것을) 만든] 것을 주로 판다.  
→ ‘만든 그 것’ 자연스러우니 복원 가능
  - [철수가 집에 간] 것을 알아.  
→ ‘집에 간 그 것’이 부자연스러우니 복원 불가능
  - 철수는 [먹는] 걸 너무 좋아한다.  
→ ‘철수가 먹는 그 거’가 부자연스러우니 복원 불가능
- 명사구에서 특정 명사만을 따로 떼어 복원 후보로 삼지 않는다. 특히, 직책/직위명만을 복원 성분으로 선택할 수 없음을 주의해야 한다.
  - 대통령은 김철수 의원을 불러서 대변인직을 맡겼다.

→ 대통령은 김철수 의원을 불러서 (대통령이) 대변인직을 (의원에게) 맡겼다. (X)  
 대통령은 김철수 의원을 불러서 (대통령이) 대변인직을 (김철수 의원에게) 맡겼다. (O)

### 2.2.3.2. 복원 표현

- 동일지시적 성분의 인칭을 그대로 복원한다.
- 인용절에서 화자/청자에 해당하는 성분을 복원해야 하는 경우, 인칭까지 고려하여 화자는 1인칭 대명사 '나', '저', '우리' 등 화자 자신을 가리키는 표현으로, 청자는 2인칭 대명사나 '형', '누나', '선생님', '감독님', '팬분들' 등 화자가 청자를 부르고 가리킬 때 사용하는 표현으로 복원해야 한다.
- 다만, 정해진 범위 내에서 동일지시적 성분을 찾을 수 없는 경우, 복원 성분으로 '화자', '청자'를 설정할 수 있다.
  - 페이커는 "팬분들이 저에게 많은 응원을 보내주신다. 많은 응원을 주시는 팬분들께 감사드린다."라고 소감을 말했다.
    - '주시는'은 '에게' 성분을 필요로 하는데, 이때 응원을 받는 실질적 대상은 '페이커'이기도 하지만, 인칭을 고려하여 '저'로 복원함.
  - "너, 정체가 뭐야?"라고 채근하던 경찰이 갑자기 "밥은 먹고 다니냐"라고 묻는 갑작스러운 상황에 모두가 당황했다.
    - '먹고'는 '이/가' 성분이 필요로 하기에, 인칭을 고려하여 '너'로 복원함.
  - 철수가 "밥 먹는 중이다"라며 곧 집에 가겠다고 한다.
    - '먹는' 행위 실질적으로 수행하는 것이 '철수'이기는 하지만 인칭이 맞지 않고, 범위 내에 적절한 대상 성분이 없기에 '화자'로 복원
- 문장, 문서 내에 복원 후보 성분이 없으면 아래와 같은 비지시적 대명사로 복원한다.

생략어의 문장 성분	비지시적 대명사
주어	누군가가, 무언가가
목적어	누군가를, 무언가를
보어	누군가가, 무언가가
부사어	어딘가에, 어딘가에서, 누군가에게, 무언가에게, 누군가에게서, 무언가에게서, 누군가로, 무언가로, 누군가와, 무언가와, 누군가보다, 무언가보다

- 시간은 장소를 나타내는 논항과 동일하게 '무언가로'로 복원한다.

- 민간 건설사가 분양하는 판교신도시의 마지막 중대형 물량으로 [(누군가가) 채권입찰제는 (어딘가에) 적용하지] 않는다.

### 3. 무형 대용어 복원의 실제

#### 3.1. 주어 복원

##### 3.1.1. 주어 복원 방법

- 서술어의 주어가 문장 내에 실현되어 있는지 확인한다. 문장 내에 실현되어 있지 않으면 일차적으로 선행 문장에서 복원 후보를 찾고, 이차적으로 후행 문장에서 복원 후보를 찾아 생략어를 복원한다. 문서 내에 적절한 복원 후보가 없으면 ‘누군가가’ 또는 ‘무언가가’로 복원한다.
- 단문에서의 주어 복원
  - 철수가 수석으로 대학에 입학했다. (철수가) 참 똑똑하다.
  - (내가) 그에게 연락했다. 나는 사과를 꼭 받아야만 했다.
- 관형사절에서의 주어 복원
  - 그는 [(그가) 마라톤 풀코스를 11번 완주한] 달리기 애호가이다.
  - 모스테파이는 2010년 무렵 경범죄를 저질렀다. [그 후 (모스테파이가) 급진적인 벨기에 출신 이슬람 지도자를 만나] [(모스테파이가) 과격 이슬람주의에 빠진] 것으로 알려졌다.
  - 물론 [(그분들이) 반려동물을 키우시는] 분들은 가족이다 하고 공감을 하지만 그렇지 않은 분들도 아직 많더라고요.  
→ 동일지시적 성분에 따라 ‘분들이’로 복원해야 하지만 ‘것’과 마찬가지로 지시 관형사 ‘그’를 넣어 복원한다.
- 명사절에서의 주어 복원
  - 그녀는 [(그녀가) 그를 사랑함]을 깨달았다.
  - [(김 대표가) 위험한 길을 선택함]에 따른 책임은 당사자인 김 대표가 지기로 했다.
- 인용절에서의 주어 복원
  - 그녀가[(그녀가) 그 일을 했다고] 스스로 말했다.
  - 박 시장은 [“(나는) 그를 만난 적도 없다”며] 모든 의혹을 부인했다.

### 3.1.2. 주어 복원 시 주의 사항

- 서술어가 감정 형용사이거나, NP1과 NP2가 전체-부분 관계인 경우 주격 중출 구문으로 분석하여 복원한다.
  - 행복한 결혼 생활을 하고 있다.  
→ ‘행복하다’는 감정 형용사이기에 주격 중출 구문으로 분석하여 ‘(누군가가) (결혼 생활이) 행복하다’로 복원한다.
  - 코가 긴 코끼리는 코를 손으로 사용한다.  
→ ‘코’와 ‘코끼리’는 전체-부분 관계이기 때문에 주격 중출 구문으로 분석하여 ‘(코끼리가)가 길다’로 복원한다.
- ‘만나다, 싸우다, 다투다, 충돌하다’와 같은 상호동사에 ‘서로’가 쓰인 경우 생략된 주어만을 복원하고 ‘~와/과’에 해당하는 성분은 복원하지 않는다.
  - 귀농인들은 서로 소통하고 더불어 살아갈 수 있는 마을 환경을 조성하였다.  
→ 동일지시적 성분이 복수인 ‘귀농인들’이므로 ‘~와/과’ 논항 없이 ‘(귀농인들이) 서로 소통하고’로 복원한다.

## 3.2. 목적어 복원

### 3.2.1. 목적어 복원 방법

- 서술어의 목적어가 문장 내에 실현되어 있는지 확인한다. 문장 내에 실현되어 있지 않으면 일차적으로 선행 문장에서 복원 후보를 찾고, 이차적으로 후행 문장에서 복원 후보를 찾아 생략어를 복원한다. 문서 내에 적절한 복원 후보가 없으면 ‘누군가를’ 또는 ‘무언가를’로 복원한다.
- 단문에서의 목적어 복원
  - 나는 범인이 누구인지 모른다. 철수가 (범인을) 안다.
  - 김 후보는 더 이상 (민심잡기를) 미룰 수가 없었다. 곧장 민심잡기에 돌입했다.
- 관형사절에서의 목적어 복원
  - [다빈치가 프랑스로 (모나리자를) 가져간] 모나리자는 현재 루브르 박물관에 전시되어 있다.
  - [은행이 (자영업자 대출총액 추산치를) 제시한] 자영업자 대출총액 추산치는 480조 2천억 원이다.
  - 이 빵집은 [쌀가루로 (것을) 만든] 것을 주로 판다.

### 3.2.2. 목적어 복원 시 주의 사항

- 의미 부류가 맞지 않더라도 생략된 성분 및 서술어와 의미적으로 관련성이 높은 성분이 구문 분석의 결과 목적어로 분석되어 격틀이 충족되어 있더라도 본래의 목적어를 복원한다.
  - 나는 학교까지 다섯 시간을(AJT) 걸었다.
    - ‘다섯 시간을’은 서술어 ‘걸다’와 관련성이 높고 조사 ‘을’을 동반하고 있지만, 구문 분석 결과 부사어(AJT)로 분석되어 있다. ‘걸다’는 자동사적 용법과 타동사적 용법을 모두 가지므로 이 경우에는 자동사적 용법으로 쓰인 것으로 보아 격틀이 충족된 것으로 판단한다.
  - 나는 작년부터 등산을 다니기 시작했다. 작년부터 올해까지 5번(AJT) 다녔다.
    - ‘5번’은 서술어 ‘다니다’와 관련성이 높지만, 구문 분석 결과 부사어(AJT)로 분석되어 있다. 이 경우 ‘다니다’의 격틀인 ‘X가 Y에 다니다’, ‘X가 Y를 다니다’를 충족하지 않았으므로 ‘다니다’에 부합하는 목적어를 적절하게 찾아 복원해야 한다.
- 자타양용동사는 문맥에 따라 타동사적 기능을 결정한다.
  - 앞에 가던 택시가 갑자기 감속하면서 뒤따르던 차들이 급정거했다.
    - ‘감속하다’의 격틀은 ‘X가 (Y를) 감속하다’인 자타양용동사이며, 위의 문장의 경우 맥락상 적절한 목적어 복원이 어렵기에 복원하지 않는다.
  - 비가 오는 날에는 감속하여 운행해야 한다.
    - ‘감속하다’의 격틀은 ‘X가 (Y를) 감속하다’인 자타양용동사이기에, 목적어 복원이 필수적이지 않지만, 위의 문장을 주어만 복원할 경우 문장이 모호해지기에 목적어도 필수적으로 복원해야 한다.
- <우리말샘> 사전에 둘 이상의 격틀이 제시되어 있는 경우, 제시된 둘 이상의 격틀 중 목적격 이외의 격틀이 실현되어 있으면 목적어를 복원하지 않는다.
  - 첫 질문은 "왜 이 영화에 천착하느냐"는 것이다.
    - ‘천착하다’의 격틀은 ‘X가 Y를 천착하다’와 ‘X가 Y에 천착하다’이다. 해당 예문은 ‘이 영화에’가 실현되어 있으니 이를 고려해 후자 격틀로 분석하여 ‘X가’만 복원한다.
- 오타 등으로 인해 맥락상 자동사인 것이 타동사로 나타났을 때, 형식은 타동사이지만 맥락 및 문장 의미를 고려하여 자동사로 분석하여 목적어를 복원하지 않는다.
  - 시의회 본연의 역할이 축소하면 행정의 나태와 비효율이 늘 수밖에 없다.
    - 이때의 ‘축소하다’는 맥락상 ‘축소되다’인 것으로 판단되기에 ‘축소되다’의 격틀인 ‘X가 축소되다’로 분석하여 목적어를 분석하지 않는다.
- 서술어의 필수 부사어인 ‘-게, -도록, -려고’가 실현되어 있으면 목적어 복원 기준 서술어에서 제외한다.



- 우리는 합리와 불합리를 모두 지키도록 강요받고 있다.  
→ ‘지키도록’은 VP 로 분석되지만 ‘강요받다’의 필수 부사어에 해당하기에 목적어를 복원하지 않는다.

### 3.3. 보어 복원

#### 3.3.1. 보어 복원 방법

- ‘되다’나 ‘아니다’의 보어가 문장 내에 실현되어 있는지 확인한다. 문장 내에 실현되어 있지 않으면 일차적으로 선행 문장에서 복원 후보를 찾고, 이차적으로 후행 문장에서 복원 후보를 찾아 생략어를 복원한다. 문서 내에 적절한 복원 후보가 없으면 ‘누군가가’ 또는 ‘무언가가’로 복원한다.
- 단문에서의 보어 복원
  - 철수는 대학생이 됐다. 영호도 (대학생이) 됐다.
  - 그는 정말 (범인이) 아니었다. 그를 범인으로 볼 여지가 전혀 없었다.
- 부사절에서의 보어 복원
  - [알기 쉬운 우리말 이름은 (무언가가) 안 되고] 굳이 어려운 영어 이름을 붙여야 국제도시의 위상에 맞는다는 걸까.

#### 3.3.2. 보어 복원시 주의 사항

- 보격조사와 주격조사의 형태가 동일함에 주의하여 보어의 생략 여부를 판단한다. 일반적으로 보어는 잘 생략되지 않으므로, 생략된 성분은 보어보다는 주어일 가능성이 크다. 특히 ‘되다’가 관형사절 내포문에 쓰일 경우 주로 생략되는 성분은 주어이다.
  - 대학생이 된 철수가 우리 앞에 나타났다.  
→ 철수가(SBJ) 대학생이(CMP) 되었다.
  - 3라운드까지 합계 10언더파 200타로 공동 선두가 된 앤서니\_김은 마지막 날 파란색을 입겠다고 공언했다.  
→ 앤서니\_김이(SBJ) 공동 선두가(CMP) 되었다.
- 명사구 또는 절을 접속하는 기능을 하는 ‘X 아니면 Y’의 ‘아니면’은 무형 대용어 탐색 대상 서술어에서 제외한다.
  - 철수 아니면 영희가 갈 것 같은데.
  - ‘모비우스’는 DCEU ‘수어사이드 스쿼드’의 길을 갈지 아니면 안티 히어로계 ‘어벤져스’의 길을 갈지 그 향방을 가늠해 볼 수 있는 작품이다.

### 3.4. 필수 부사어 복원

#### 3.4.1. 부사어 복원 방법

- 문장 내에서 본용언으로 쓰인 동사와 형용사를 <우리말샘>에서 검색하여 격틀을 확인한다. ‘에, 에서, 에게서, 로, 와, 보다’ 등의 부사격조사가 결합될 수 있는 필수 부사어가 문장 내에 실현되어 있지 않으면 일차적으로 선행 문장에서 복원 후보를 찾고, 이차적으로 후행 문장에서 복원 후보를 찾아 필수 부사어를 복원한다. 문서 내에 적절한 복원 후보가 없으면 ‘어딘가에, 어딘가에서, 누군가에게서/무언가에게서, 누군가로/무언가로, 누군가와/무언가와, 누군가보다/무언가보다’ 중 하나로 복원한다.
- 단문에서의 필수 부사어 복원
  - 나는 2019년부터 서울에 산다. 벌써 5년째 (서울에) 살고 있다.
  - 혹시나 하는 마음에 주머니에 손을 넣어 보았다. 하루 종일 찾던 지갑이 (주머니에서) 나왔다.
  - 나는 천 원짜리 지폐를 (동전으로) 바꾸고 싶었다. 아저씨에게 동전이 있냐고 물었다.
  - 철수는 작년에 (그녀와) 결혼했다. 그녀와 직장에서 만났다고 한다.
  - 그녀는 질을 (양보다) 우선시한다. 양이 많은지 적은지는 크게 고려하지 않는다.
- 내포절에서의 부사어 복원
  - 신고를 받은 소방서는 46m 짜리 고가사다리차를 동원했지만 [물줄기가 (어딘가에) 닿지] 않아 무용지물이었다.
  - 매점 음식이 비싸다는 얘기를 듣고 조금이라도 비용을 줄이려고 집에서 김밥을 싸 갔지만 입구 검사에서 적발돼 갖고 [(어딘가로) 들어갈] 수 없었던 것.
  - 박 대표도 이날 기자간담회에서 “과거 김 원내대표가 내무부 차관을 할 때 내가 민주당 대변인으로 있으면서 [(내가) (김 대표와) 가까워졌다]”며 “최근에도 함께 식사를 하며 잘 해보자는 덕담을 나누기도 했다”고 말했다.
  - 대부분 “[돈이 (무언가보다) 낫다]”고들 하지만 그냥 돈 봉투만 줄 수도 없는 일.

#### 3.4.2. 필수 부사어 복원시 주의 사항

- <우리말샘>에 격틀로 제시된 부사어만을 필수 부사어로 인정한다.
- 필수 부사어를 논항으로 갖는 서술어가 관형사절 내포문에 쓰일 경우 필수 부사어는 잘 생략되지 않는다는 점에 유의한다. 주로 생략되는 성분은 주어이다.

- 양주시에 사는 J 씨는 2004년 상조업체 방문판매 사원을 통해 월 6만 원씩 120회 납부 조건의 10년 만기 상품 4개 계좌에 가입했다.  
→ J씨가(SBJ) 양주시에(AJT) 살다.
- 사회적 기업은 사회적 성과를 주된 목적으로 삼는 기업이다.  
→ 기업이(SBJ) 사회적 성과를(OBJ) 주된 목적으로(AJT) 삼다.
- ‘에서’는 주격조사, 부사격조사 기능을 모두 가지고 있다. 구문 분석 단계에서 ‘에서’ 성분이 주어로 분석되었는지 부사어로 분석되었는지에 따라 생략된 성분을 판단하여 복원해야 한다.
  - 이번 대회는 우리 학교에서(SBJ) 우승을 차지했다.  
→ ‘우리 학교에서’는 ‘에서’가 실현되어 있으나 소위 단체 주어로 기능하기 때문에 구문 분석 단계에서도 이를 고려하여 SBJ 로 분석하고 있다. 그렇기에 ‘차지하다’의 주어가 이미 실현되어 있는 것으로 보아 복원하지 않는다.
  - 이 물건은 학교에서(AJT) 가져왔어.  
→ ‘학교에서’는 물건을 가져온 처소이기에 ‘가져오다’의 주어를 복원해야 한다.
  - ‘보이다’가 본용언으로 쓰이면 필수적 부사어 복원 대상 서술어에 해당한다. 그러나 ‘-어 보이다’로 쓰인 경우에는 복원하지 않는 것에 주의해야 한다.

보이다

보이다002 ((‘...으로’나 ‘-게’ 대신에 평가를 뜻하는 다른 부사어가 쓰이거나 ‘-아/어 보이다’ 구성으로 쓰이기도 한다)) 대상이 평가되다. ‘보다’의 피동사

## [붙임6] 다층위 구조 분석 말뭉치 구축 지침

### 다층위 구조 분석 말뭉치 구축 지침

#### 1. 총론

##### 1.1. 지침의 목적

이 지침은 한국어 문장의 형태 분석 결과와 어휘의미 분석 결과, 구문 분석 결과, 의미역 분석 결과를 층위별로 쌓은 다층위 구조 분석 말뭉치를 구축하는 방안을 제시하는 것을 목적으로 한다.

##### 1.2. 용어 정의

- 형태: 하나의 형태소가 환경에 따라 달리 실현된 모습
- 단어: 하나의 어절을 구성하는 자립 형식으로, <우리말샘> 사전의 기본 등재 단위
- 어절: 문장을 구성하고 있는 문장 성분의 최소 단위로서 띄어쓰기의 단위
- 품사: 단어를 형태, 기능, 의미 등 기준에 따라 나눈 갈래
- 어휘의미(sense): 갈래 뜻
- 구문 표지: 단어, 구, 절의 통사적 성격을 규정한 표지 표지로 어절 앞부분의 형태 정보를 바탕으로 부여한다.
- 기능 표지: 해당 어절이 문장 내의 다른 어절과 가지는 문법적 관계를 규정한 표지로 어절 뒷부분의 형태 정보를 바탕으로 부여한다.
- 지배-의존 관계: 문장 내 지배소와 의존소 사이의 관계. 다음 세 가지로 분류된다.
  - 논항-서술어 관계
  - 수식어-피수식어 관계
  - 접속 관계
- 논항(論項, argument): 서술어가 의미적으로 요구하는 필수적인 성분.
- 부가어(附加語, adjunct): 서술어와 의미적으로 관련이 있으나 수의적인 성분.
- 의미역(意味役, thematic role, theta-role, semantic role): 논항 혹은 부가어가 서술어에 대해 가지는 의미적 역할
- 무형 대용어: 문장에서 서술어가 요구하는 필수 성분 중 생략된 성분.(= 생략 성분)
- 동일지시적 성분: 실제 문장에서 실현되지 않은 필수 성분과 지시하는 대상이 동일한 성분.
- 복원 성분: 선행 문장 혹은 후행 문장을 참고하여 격틀에 맞게 복원한 성분.

### 1.3. 기본 원칙

- ‘국립국어원 신문 말뭉치 2024’와 ‘국립국어원 일상 대화 말뭉치 2024’를 대상으로 한국어 문장의 각 어절을 형태 단위로 분리하고 각 형태에 형태 표지를 부착한다.
- 형태 분석된 한국어 문장의 체언류와 용언류 어절을 대상으로 각 형태에 어휘의미 표지를 부착한다.
- 형태 분석된 한국어 문장의 각 어절에 구문 표지와 기능 표지를 부착하고 어절들 간의 지배-의존 관계를 표시한다.
- 구문 분석된 각 용언구의 지배-의존 관계와 용언구를 구성하는 동사(VV), 형용사(VA)의 어휘의미에 따른 격틀 정보를 바탕으로 각 서술어의 논항과 부가어에 대해 의미역 정보를 부착한다.
- 추가적으로 구문 분석된 각 용언구의 지배-의존 관계와 의미역 정보를 바탕으로 각 서술어의 생략된 논항을 복원한다.

## 2. 형태 분석 표지 부착의 원리

### 2.1. 형태 분석 대상

국립국어원 <우리말샘>의 등재어를 기준으로 형태 분석 대상을 분리하며 다음 단계의 어휘의미 분석과 구문 분석에 필요한 정보만을 제공할 수 있도록 단어 단위 분석을 지향하고 단어 내부 분석은 지양한다. 이를 위하여 ‘반(反)-, 대(對)-’ 등 체언접두사, ‘-적(的), -성(性), -화(化)’ 등 명사파생접미사, ‘-하다, -되다’ 등 동사파생접미사는 따로 분리하지 않고 어근과 통합한 단위로 분리한다. 반면 지시 대상이 하나인 구나 절도 하나의 단어로 취급하여 내부 구성 성분을 분리하지 않으며 합성어 또한 하나의 단어로 분석한다.

### 2.2. 형태 분석 표지의 집합

대분류	소분류	세분류	태그
체언	명사	일반명사	NNG
		고유명사	NNP
		의존명사	NNB
	대명사	대명사	NP
	수사	수사	NR
용언	동사	동사	VV
	형용사	형용사	VA
	지정사	긍정지정사	VCP
		부정지정사	VCN

	보조용언	보조용언	VX
수식언	관형사	성상 관형사	MMA
		지시 관형사	MMD
		수 관형사	MMN
	부사	일반부사	MAG
		접속부사	MAJ
독립언	감탄사	감탄사	IC
관계언	격조사	주격조사	JKS
		보격조사	JKC
		관형격조사	JKG
		목적격조사	JKO
		부사격조사	JKB
		호격조사	JKV
		인용격조사	JKQ
	보조사	보조사	JX
	접속조사	접속조사	JC
의존형태	어미	선어말어미	EP
		종결어미	EF
		연결어미	EC
		명사형전성어미	ETN
		관형형전성어미	ETM
	접두사	체언접두사	XPN
	접미사	명사파생접미사	XSN
		동사파생접미사	XSV
		형용사파생접미사	XSA
	어근	어근	XR
기호	일반기호	마침표, 물음표, 느낌표	SF
		쉼표, 가운뎃점, 콜론, 빗금	SP
		따옴표, 괄호표, 줄표	SS
		출임표	SE
		붙임표(물결)	SO
		기타기호	SW
	외국문자	외국문자	SL
	한자	한자	SH
	숫자	숫자	SN
기타	분석불능범주	분석불능범주	NA

## 2.3. 형태 분석의 실제

형태 분석 관련 자세한 사항은 ‘2025년 한국어 구조 분석 및 무형 대응어 복원 말뭉치 구축 연구’ 사업의 형태 분석 지침을 참고한다.

### 구어 분석의 예

SDRW2400001443.1.1. 지금 엄마가 말하는 거는 거의 그 옛날 코미디 프로 말하는 거 같은데.		
word_id	word	morph
1	지금	지금/MAG
2	엄마가	엄마/NNG + 가/JKS
3	말하는	말하/VV + 는/ETM
4	거는	거/NNB + 는/JX
5	거의	거의/MAG
6	그	그/MMD
7	옛날	옛날/NNG
8	코미디	코미디/NNG
9	프로	프로/NNG
10	말하는	말하/VV + 는/ETM
11	거	거/NNB
12	같은데.	같/VA + 은데/EF + ./SF

### 문어 분석의 예

NPRW2400000002.151.3.2. 지난해 국제 원자재 가격이 급등한 배경이었던 러시아·우크라이나전쟁 양상이 여전히 안갯속이고, 중국의 ‘제로코로나’ 정책 중단이 경제에 미칠 영향도 불확실성을 키우고 있다.		
word_id	word	morph
1	지난해	지난해/NNG
2	국제	국제/NNG
3	원자재	원자재/NNG
4	가격이	가격/NNG + 이/JKS
5	급등한	급등하/VV + ㄴ/ETM
6	배경이었던	배경/NNG + 이/VCP + 었/EP + 던/ETM
7	러시아·우크라이나전쟁	러시아·우크라이나전쟁/NNP
8	양상이	양상/NNG + 이/JKS
9	여전히	여전히/MAG
10	안갯속이고,	안갯속/NNG + 이/VCP + 고/EC + ./SP
11	중국의	중국/NNP + 의/JKG
12	‘제로코로나’	’/SS + 제로코로나/NNP + ’/SS
13	정책	정책/NNG

14	중단이	중단/NNG + 이/JKS
15	경제에	경제/NNG + 예/JKB
16	미칠	미치/VV + ㄹ/ETM
17	영향도	영향/NNG + 도/JX
18	불확실성을	불확실성/NNG + 을/JKO
19	키우고	키우/VV + 고/EC
20	있다.	있/VX + 다/EF + ./SF

### 3. 어휘의미 분석 표지 부착의 원리

#### 3.1. 어휘의미 분석 대상

국립국어원 <우리말샘> 등재어의 다의어 번호를 기준으로 형태 분석 말뭉치의 체언(NNG, NNP, NNB, NP, NR) 및 용언(VV, VA, VCP, VCN, VX)에 세 자리의 어휘의미 번호를 부착한다.

<우리말샘>은 등재 어휘가 꾸준히 증가하고 있는 동적 사전이므로 2025년 1분기의 <우리말샘> 등재 어휘를 기준으로 어휘의미를 분석한다.

#### 3.2. 어휘의미 분석 표지의 집합

구분	설명	표지
<우리말샘> 등재어	해당 어휘와 해당 의미가 등재되어 있는 경우	000
	해당 어휘는 등재되어 있으나 해당 의미가 등재되어 있지 않은 경우	888
<우리말샘> 미등재어	해당 어휘가 등재되어 있지 않은 경우	777
	기타 어휘의미 번호 '777'을 부여할 수 없는 경우	999

#### 3.3. 어휘의미 분석의 실제

어휘의미 분석 관련 자세한 사항은 '2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구' 사업의 어휘의미 분석 지침을 참고한다.

구어 분석의 예

SDRW2400001443.1.1. 지금 엄마가 말하는 거는 거의 그 옛날 코미디 프로 말하는 거 같은데.			
word_id	word	morph	WSD Form
1	지금	지금/MAG	



2	엄마가	엄마/NNG + 가/JKS	엄마/001
3	말하는	말하/VV + 는/ETM	말하/001
4	거는	거/NNB + 는/JX	거/001
5	거의	거의/MAG	
6	그	그/MMD	
7	옛날	옛날/NNG	옛날/001
8	코미디	코미디/NNG	코미디/001
9	프로	프로/NNG	프로/005
10	말하는	말하/VV + 는/ETM	말하/001
11	거	거/NNB	거/001
12	같은데.	같/VA + 은데/EF + ./SF	같/009

#### 문어 분석의 예

NPRW2400000002.151.3.2. 지난해 국제 원자재 가격이 급등한 배경이었던 러시아·우크라이나전쟁 양상이 여전히 안갯속이고, 중국의 ‘제로코로나’ 정책 중단이 경제에 미칠 영향도 불확실성을 키우고 있다.			
word_id	word	morph	WSD Form
1	지난해	지난해/NNG	지난해/001
2	국제	국제/NNG	국제/004
3	원자재	원자재/NNG	원자재/001
4	가격이	가격/NNG + 이/JKS	가격/003
5	급등한	급등하/VV + ㄴ/ETM	급등하/001
6	배경이었던	배경/NNG + 이/VCP + 었/EP + 던/ETM	배경/002 + 이/004
7	러시아·우크라이나전쟁	러시아·우크라이나전쟁/NNP	러시아·우크라이나전쟁/777
8	양상이	양상/NNG + 이/JKS	양상/008
9	여전히	여전히/MAG	
10	안갯속이고,	안갯속/NNG + 이/VCP + 고/EC + ./SP	안갯속/001 + 이/006
11	중국의	중국/NNP + 의/JKG	중국/002
12	‘제로코로나’	‘/SS + 제로코로나/NNP + ‘/SS	제로코로나/002
13	정책	정책/NNG	정책/003
14	중단이	중단/NNG + 이/JKS	중단/007
15	경제에	경제/NNG + 예/JKB	경제/005
16	미칠	미치/VV + ㄹ/ETM	미치/006
17	영향도	영향/NNG + 도/JX	영향/004
18	불확실성을	불확실성/NNG + 을/JKO	불확실성/001
19	키우고	키우/VV + 고/EC	키우/005
20	있다.	있/VX + 다/EF + ./SF	있/023

## 4. 구문 분석 표지 부착의 원리

### 4.1. 구문 분석 대상

형태 분석된 한국어 문장의 각 어절을 대상으로 구문 표지와 기능 표지를 부착하고 지배-의존 관계를 분석한다.

각 어절 앞부분의 형태 정보를 통해 해당 어절의 구문 표지를 결정하고 어절 뒷부분의 형태 정보를 통해 해당 어절의 기능 표지를 결정하며, 지배-의존 관계는 논항-서술어 관계, 수식어-피수식어 관계, 접속 관계를 구분하여 문장 내 지배소와 의존소 사이의 관계를 표시한다.

### 4.2. 구문 분석 표지의 집합

#### 4.2.1. 구문 표지의 부착

- 구문 표지에는 NP, VP, VNP, DP, AP, IP, X, L, R 이 있다. 구문 표지는 해당 어절의 범주적 특성을 표시하는 것으로서 어휘 요소에 기반하여 결정한다.
- 어휘 요소의 형태 분석 표지와 구문 표지의 전형적인 대응관계를 보이면 다음과 같다.

어휘 요소의 형태 분석 표지	구문 표지
NNG(일반명사) NNP(고유명사) NNB(의존명사) NP(대명사) NR(수사) XR(어근)	NP(체언)
VV(동사) VA(형용사) VCN(부정지정사) VX(보조용언)	VP(용언)
VCP(긍정지정사)	VNP(긍정 지정사구)
MMA(성상 관형사) MMD(지시 관형사) MMN(수 관형사)	DP(관형사구)
MAG(일반부사) MAJ(접속부사)	AP(부사구)
IC(감탄사)	IP(감탄사구)
SP(첨표, 가운뎃점, 콜론, 빗금) SO(붙임표)	X(의사 구)

SL(외국 문자)	
SH(한자)	
SS(따옴표, 괄호표가 단독으로 쓰이는 경우)	L(왼쪽 부호)
SS(따옴표, 괄호표가 단독으로 쓰이는 경우)	R(오른쪽 부호)

#### 4.2.2. 기능 표지의 부착

- 기능 표지에는 SBJ, OBJ, CMP, MOD, AJT, CNJ 가 있다. 기능 표지는 해당 어절의 문장 내에서의 기능을 표시하는 것으로서 문법 요소에 기반하여 결정한다.
- 문법 요소의 형태 분석 표지와 기능 표지의 전형적인 대응 관계를 보이면 다음과 같다.

문법 요소의 형태 분석 표지	기능 표지
JKS(주격 조사)	SBJ(주어)
JKO(목적격 조사)	OBJ(목적어)
JKC(보격 조사)	CMP(보어)
JKG(관형격 조사) ETM(관형사형 어미)	MOD(관형어)
JKB(부사격 조사)	AJT(부사어)
JC(접속 조사)	CNJ(접속어)

- 관형사(MMA, MMD, MMN)와 일반부사(MAG)는 문장 안에서 수식 대상을 한정하는 기능을 가진다. 분석 대상 어절의 형태가 관형사(MMA, MMD, MMN) 또는 일반부사(MAG)로만 분석될 경우에는 구문 표지가 DP(관형사구) 또는 AP(부사구)로 분석되어 기능 표지 MOD(관형어)와 AJT(부사어)는 불필요하지만 하나의 어절이 여러 가지 형태로 이루어진 경우 기능 표지 MOD(관형어)와 AJT(부사어)를 부착할 수 있다.

어휘 요소의 형태 분석 표지	기능 표지
MMA(성상 관형사) MMD(지시 관형사) MMN(수 관형사)	MOD(관형어)
MAG(일반 부사)	AJT(부사어)

### 4.3. 분석의 실제

구문 분석 관련 자세한 사항은 ‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업의 구문 분석 지침을 참고한다.

구어 분석의 예

SDRW2400001443.1.1. 지금 엄마가 말하는 거는 거의 그 옛날 코미디 프로 말하는 거 같은데.					
word_id	word	morph	WSD Form	head	DP Label

1	지금	지금/MAG		3	AP
2	엄마가	엄마/NNG + 가/JKS	엄마/001	3	NP_SBJ
3	말하는	말하/VV + 는/ETM	말하/001	4	VP_MOD
4	거는	거/NNB + 는/JX	거/001	10	NP_SBJ
5	거의	거의/MAG		10	AP
6	그	그/MMD		9	DP
7	옛날	옛날/NNG	옛날/001	9	NP
8	코미디	코미디/NNG	코미디/001	9	NP
9	프로	프로/NNG	프로/005	10	NP_OBJ
10	말하는	말하/VV + 는/ETM	말하/001	11	VP_MOD
11	거	거/NNB	거/001	12	NP
12	같은데.	같/VA + 은데/EF + ./SF	같/009	-1	VP

#### 문어 분석의 예

NPRW2400000002.151.3.2. 지난해 국제 원자재 가격이 급등한 배경이었던 러시아·우크라이나전쟁 양상이 여전히 안갯속이고, 중국의 ‘제로코로나’ 정책 중단이 경제에 미칠 영향도 불확실성을 키우고 있다.					
word_id	word	morph	WSD Form	head	DP Label
1	지난해	지난해/NNG	지난해/001	5	NP_AJT
2	국제	국제/NNG	국제/004	3	NP
3	원자재	원자재/NNG	원자재/001	4	NP
4	가격이	가격/NNG + 이/JKS	가격/003	5	NP_SBJ
5	급등한	급등하/VV + ㄴ/ETM	급등하/001	6	VP_MOD
6	배경이었던	배경/NNG + 이/VCP + 었/EP + 던/ETM	배경/002 + 이/004	7	VNP_MOD
7	러시아·우크라이나전쟁	러시아·우크라이나전쟁 /NNP	러시아·우크라이나전쟁/777	8	NP
8	양상이	양상/NNG + 이/JKS	양상/008	10	NP_SBJ
9	여전히	여전히/MAG		10	AP
10	안갯속이고,	안갯속/NNG + 이/VCP + 고/EC + ./SP	안갯속/001 + 이/006	19	VNP_AJT
11	중국의	중국/NNP + 의/JKG	중국/002	12	NP_MOD
12	‘제로코로나’	/SS + 제로코로나/NNP + /SS	제로코로나/002	13	NP
13	정책	정책/NNG	정책/003	14	NP
14	중단이	중단/NNG + 이/JKS	중단/007	16	NP_SBJ
15	경제에	경제/NNG + 예/JKB	경제/005	16	NP_AJT
16	미칠	미치/VV + ㄴ/ETM	미치/006	17	VP_MOD
17	영향도	영향/NNG + 도/JX	영향/004	19	NP_SBJ
18	불확실성을	불확실성/NNG + 을/JKO	불확실성/001	19	NP_OBJ
19	키우고	키우/VV + 고/EC	키우/005	20	VP

## 5. 의미역 분석 표지 부착의 원리

### 5.1. 의미역 분석 대상

형태 분석 결과 동사(VV), 형용사(VA)로 분석된 서술어의 어휘의미에 따른 격틀 정보와 구문 분석 결과로 확인된 논항-서술어 관계, 수식어-피수식어 관계, 접속 관계를 바탕으로 각 서술어의 논항과 부가어에 대해 의미역 정보를 부착한다.

### 5.2. 의미역 분석 표지의 집합

#### 5.2.1. 필수 의미역 표지의 부착

연번	약호	영문명	국문명	예
1	AGT	Agent	행동주	내가 저녁을 만들었다. 철수가 영희에게 빵을 먹였다.(사동문의 주어)
2	EXP	Experiencer	경험주	철수는 영희를 보았다.(지각적 경험 주체) 철수는 영희에 대해 잘 안다.(인지적 경험 주체)
3	MAG	Mental Agent	심리행위주	철수는 영희가 무서웠다.(감정적 경험 주체) (의도성이 느껴지는 심리동사 서술어의 주체)
4	EFF	Effector	영향주	철수는 영희를 좋아한다. 강풍이 온 산을 덮었다.
5	THM	Theme	대상	철수가 전봇대에 부딪혔다. 철수가 아주 침착하다.(형용사문의 주어) 감자에 싹이 났다.(자동사문의 주어) 도둑이 경찰에게 잡혔다.(피동문의 주어)
6	COM	Companion	동반주	철수는 영희를 좋아한다.(타동사문의 목적어) 철수는 영희와 싸웠다.(주어 참여자와 동일 지위)
7	GOL	Goal	도착점	철수는 모래를 물과 섞었다.(목적어 참여자와 동일 지위) 철수는 학교에 갔다.
8	LOC	Location	장소	철수가 영희에게 빵을 먹였다.(사동문의 피사동주) 철수는 방에 있다. (구체적 공간)
9	SRC	Source	출발점	철수는 공경에 처해 있다.(추상적 공간) 정부가 옥수수를 미국으로부터 수입했다.
10	CNT	Content	내용	이 기단은 서쪽에서 왔다. 그는 자기가 천재라고 말했다.
11	INS	Instrument	도구	나는 국민과 언제나 함께하겠다 연설했다. 이 물질은 3개의 원자로 이루어졌다.

12	CRT	Criterion	기준	참고가 책으로 가득하다. 철수의 실력은 영희보다 낮다. 이 일은 내 능력에 부친다.
13	FNS	Final State	결과상태	항아리의 물이 포도주가/포도주로 되었다. 선생님은 철수를 사위로 삼았다.
14	DIR	Direction	방향	철수는 학교로 갔다. 동아시아에서 중앙아시아로 이주했다.
15	QLF	Qualification	자격	그는 과학자로 유명하다. 그는 군에서 군의관으로 복무했다.
16	MNR	Manner	방식	철수는 어리석게 군다. 영희가 착하게 보인다.
17	PRT	Part	부분	(주격중출문에서의 NP2) 철수가 키가 크다. (형용사문) 영희는 가슴이 두근거렸다. (자동사문)
18	FTR	Feature	속성	(주격중출문에서의 NP2) 영희가 심성이 곱다.

### 5.2.2. 부가의미역 표지의 부착

연번	약호	영문명	국문명	예문
1	M-TMP	Temporal	시간	이 지역에는 봄에(TMP) 안개가 정말 많이 낀다.
2	M-LOC	Location	장소	봄이 되자 공원에(LOC) 꽃이 만발했다.
3	M-DIR	Direction	방향	달이 서쪽으로(DIR) 기울었다.
4	M-ROU	Route	경로	우리는 목적지까지 지름길로(ROU) 가기로 했다.
5	M-INS	Instrument	도구	아이들은 색연필로(INS) 알록달록 그림을 칠했다.
6	M-MNR	Manner	방식	철수는 아들을 정성으로/정성껏(MNR) 간호했다.
7	M-QNT	Quantity	수량	질문 하나(QNT) 합시다.
8	M-DGR	Degree	정도	그들은 아름다운 경치에 흠뻑(DGR) 도취되었다.
9	M-EXT	Extent	범위	이 작품은 그의 소설들 가운데(EXT) 가장 걸출하다.
10	M-COM	Comitative	공동	철수는 영희와(COM) 극장에 갔다.
11	M-CRT	Criterion	기준	오늘이 어제보다(CRT) 더 시원하다.
12	M-QLF	Qualification	자격	그는 삼남삼녀 가운데 막내로(QLF) 태어났다.
13	M-CAU	Cause	원인	굴이 설익어서(CAU) 시다.
14	M-PRP	Purpose	목적	나는 형의 심부름으로 커피를 사러(PRP) 카페에 갔다.
15	M-CND	Condition	조건	아이들은 마트에 가면(CND) 장난감을 사 달라고 떼쓴다.
16	M-CNC	Concession	양보	어머니는 아무리 늦게 주무시더라도(CNC) 항상 새벽 5시에 일어나신다.
17	M-PRE	Preceding	선행	설날에 온 가족이 모여(PRE) 만두를 빚었다.
18	M-BCK	Background	배경	소설책을 샀는데(BCK) 아주 재미있더라.
19	M-LST	List	나열	언니는 학교에 가고(LST) 부모님은 출근하셨다.
20	M-CTR	Contrast	대조	이 음식은 생긴 건 우습지만(CTR) 맛은 있다.
21	M-SEL	Selective	선택	도서관에는 철수가 가든지(SEL) 영희가 갈 것이다.

22	M-NEG	Negation	부정	철수는 술을 안(NEG) 마신다.
23	M-MOD	Modal	양태	다행히(MOD) 철수가 제 시간에 나타났다.
24	M-CNJ	Conjunction	접속	또한(CNJ) 정부는 경영진에게도 강력히 경고했다.

### 5.3. 의미역 분석의 실제

의미역 분석 관련 자세한 사항은 ‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업의 의미역 분석 지침을 참고한다.

#### 구어 분석의 예

SDRW2400001443.1.1. 지금 엄마가 말하는 거는 거의 그 옛날 코미디 프로 말하는 거 같은데.								
word_id	word	morph	WSD Form	head	DP Label	S R L Span	S R L Label	S R L Predicate Lamma
1	지금	지금/MAG		3	AP	1	M-TMP	3/말하다
2	엄마가	엄마/NNG + 가/JKS	엄마/001	3	NP_SBJ	2	AGT	3/말하다
3	말하는	말하/VV + 는/ETM	말하/001	4	VP_MOD			
4	거는	거/NNB + 는/JX	거/001	10	NP_SBJ	1, 2, 3, 4	THM	10/말하다
5	거의	거의/MAG		10	AP	5	M-DGR	10/말하다
6	그	그/MMD		9	DP			
7	옛날	옛날/NNG	옛날/001	9	NP			
8	코미디	코미디/NNG	코미디/001	9	NP			
9	프로	프로/NNG	프로/005	10	NP_OBJ	6, 7, 8, 9	THM	10/말하다
10	말하는	말하/VV + 는/ETM	말하/001	11	VP_MOD			
11	거	거/NNB	거/001	12	NP			
12	같은데.	같/VA + 은데/EF + ./SF	같/009	-1	VP			

#### 문어 분석의 예

NPRW2400000002.151.3.2. 지난해 국제 원자재 가격이 급등한 배경이었던 러시아-우크라이나전쟁 양상이 여전히 안갯속이고, 중국의 ‘제로코로나’ 정책 중단이 경제에 미칠 영향도 불확실성을 키우고 있다.								
word_id	word	morph	WSD Form	head	DP Label	S R L Span	S R L Label	S R L Predicate Lamma
1	지난해	지난해/NNG	지난해/001	5	NP_AJT	1	M-TMP	5/급등하다

2	국제	국제/NNG	국제/004	3	NP	
3	원자재	원자재/NNG	원자재/001	4	NP	
4	가격이	가격/NNG + 이 /JKS	가격/003	5	NP_SBJ	2, 3, 4 THM 5/급등하다
5	급등한	급등하/VV + L /ETM	급등하/001	6	VP_MOD	
6	배경이었던	배경/NNG + 이 /VCP + 었/EP + 던/ETM	배경/002 + 이 /004	7	VNP_MOD	
7	러시아·우크라이나전쟁	러시아·우크라이나전쟁/NNP	러시아·우크라이나전쟁/777	8	NP	
8	양상이	양상/NNG + 이 /JKS	양상/008	10	NP_SBJ	
9	여전히	여전히/MAG		10	AP	
10	안갯속이고,	안갯속/NNG + 이/VCP + 고 /EC + /SP	안갯속/001 + 이/006	19	VNP_AJT	1, 2, 3, 4, 5, 6, 7, 8, 9, 10 M-LST 19/키우다
11	중국의	중국/NNP + 의 /JKG	중국/002	12	NP_MOD	
12	‘제로코로나’	‘/SS + 제로코로나/NNP + ‘/SS	제로코로나/002	13	NP	
13	정책	정책/NNG	정책/003	14	NP	
14	중단이	중단/NNG + 이 /JKS	중단/007	16	NP_SBJ	1 1 , 1 2 , 1 3 , 14 EFF 16/미치다
15	경제에	경제/NNG + 예 /JKB	경제/005	16	NP_AJT	15 GOL 16/미치다
16	미칠	미치/VV + 르 /ETM	미치/006	17	VP_MOD	
17	영향도	영향/NNG + 도 /JX	영향/004	19	NP_SBJ	1 1 , 1 2 , 1 3 , 1 4 , 1 5 , 1 6 , 17 EFF 19/키우다
18	불확실성을	불확실성/NNG + 을/JKO	불확실성/001	19	NP_OBJ	18 THM 19/키우다
19	키우고	키우/VV + 고 /EC	키우/005	20	VP	



20	있다.	있/VX + 다/EF + ./SF	있/023	-1	VP	
----	-----	-----------------------	-------	----	----	--

## 6. 무형 대용어 복원의 원리

### 6.1. 무형 대용어 복원 대상

형태 분석 결과 동사(VV), 형용사(VA)의 어휘의미에 따른 격틀 정보를 바탕으로 구문 분석 단계에서 논항-서술어 관계, 의미역 분석 단계에서 필수 의미역으로 분석되지 않은 대상을 찾아 문장에서 실현되지 않은 논항을 복원한다. 단 격 교체 현상에 의해 <우리말샘>의 격틀 정보와 다른 형태의 논항이 실현된 경우에는 격틀이 충족된 것으로 보아 생략 성분을 복원하지 않는다.

무형 대용어 복원은 '무형 대용어 복원 대상 서술어 탐지 → 동일지시적 성분 탐색 → 적절한 무형 대용어 복원'으로 진행되며, 분석 대상 서술어가 포함된 앞 15문장 / 뒤 2문장 이내로 복원 범위를 한정하여 동일지시적인 성분을 탐색한다.

### 6.2. 무형 대용어 복원 형태의 집합

#### 6.2.1. 기본 복원 형태

무형 대용어의 기본 복원 단위는 명사 혹은 명사구에 조사가 결합된 형태이며, 문장에서 독립적으로 쓰일 수 없는 의존명사는 그 앞에 지시 관형사 '그'를 넣어 자연스러운 논항의 형태로 복원한다.

생략어의 문장 성분	생략어 복원 형태
주어	NP이/NP가
목적어	NP을/NP를
보어	NP이/NP가
부사어	NP에, NP에게, NP에서, NP에게서, NP로/NP으로, NP와/NP과, NP보다

#### 6.2.2. 비지시적 대명사의 복원 형태

문서 내 복원 범위 안에서 동일지시적인 성분이 탐지되지 않는 경우에는 아래와 같은 비지시적 대명사를 이용하여 무형 대용어를 복원한다.

생략어의 문장 성분	비지시적 대명사
주어	누군가가, 무언가가

목적어	누군가를, 무언가를
보어	누군가가, 무언가가
부사어	어딘가에, 어딘가에서, 누군가에게, 무언가에게, 누군가에게서, 무언가에게서, 누군가로, 무언가로, 누군가와, 무언가와, 누군가보다, 무언가보다

### 6.3. 무형 대용어 복원의 실제

무형 대용어 복원 관련 자세한 사항은 ‘2025년 한국어 구조 분석 및 무형 대용어 복원 말뭉치 구축 연구’ 사업의 무형 대용어 복원 지침을 참고한다.

구어 분석의 예

SDRW2400001443.1.1. 지금 엄마가 말하는 거는 거의 그 옛날 코미디 프로 말하는 거 같은데.													
word_id	word	morph	WSD Form	head	DP Label	SRL Span	SRL Label	SRL Predicate Lamma	ant_sen_id	ant_word_id	ant_form	restored_form	restored_type
1	지금	지 금 /MAG		3	AP	1	M-TMP	3 / 말하다					
2	엄 마 가	/NNG + 가 /JKS	엄 마 /001	3	NP_SBJ	2	AGT	3 / 말하다					
3	말 하 는	/VV + 는 /ETM	말 하 /001	4	VP_MODAL				# / 1.44	# / 4	# / 거는	청 자 에 게 / 그 겔	adjunct / object
4	거는	/NNB + 는 /JX	거 /001	10	NP_SBJ	1, 2, 3, 4	THM	10/말하다					
5	거의	거 의 /MAG		10	AP	5	M-DGR	10/말하다					
6	그	그		9	DP								

		/MMD											
7	옛날	옛날 /NNG	옛날 /001	9	NP								
8	코미디	코미디 /NNG	코미디 /001	9	NP								
9	프로	프로 /NNG	프로 /005	10	NP_O BJ	6, 7, 8, 9	THM	10/말 하다					
10	말하는	말하 /VV + 는 /ETM	말하 /001	11	VP_M OD				1.44 / #	2 / #	엄마 가 / #	엄마 가 / 청 자에게	subje ct / adjun ct
11	거	거 /NNB	거 /001	12	NP								
12	같은데.	같/ VA + 은 데/EF + ./SF	같 /009	-1	VP								

#### 문어 분석의 예

NPRW2400000002.151.3.2. 지난해 국제 원자재 가격이 급등한 배경이었던 러시아·우크라이나전쟁 양상이 여전히 안 갓속이고, 중국의 '제로코로나' 정책 중단이 경제에 미칠 영향도 불확실성을 키우고 있다.													
word _id	word	morph	W S D Form	head	D P Label	S R L S p an	S R L Label	S R L Predicat e Lamma	ant _s en_id	ant _wo rd_i d	ant_ for m	rest ore d_f orm	resto red_t ype
1	지 난 해	지 난 해/NNG	지 난 해 /001	5	NP_A JT	1	M-TMP	5/급 등하 다					
2	국 제	국 제/NNG	국 제 /004	3	NP								
3	원 자 재	원 자 재/NNG	원 자 재 /001	4	NP								
4	가 격 이	가 격/NNG + 이/JKS	가 격 /003	5	NP_S BJ	2 , 3 , 4	THM	5/급 등하 다					
5	급 등 한	급 등하/VV + L/ETM	급 등하 /001	6	VP_M OD								
6	배 경 이 었 던	배 경/NNG + 이/VCP + 었/EP + 던 /ETM	배 경 /002 + 이/004	7	VNP_ MOD								
7	러 시 아·우 크 라 이 나 전 쟁	러 시 아·우 크 라 이 나 전 쟁	러 시 아·우 크 라	8	NP								

	크 라 이 나 전 쟁	/NNP	이 나 전 쟁/777						
8	양 상 이	양상/NNG + 이/JKS	양 상 /008	10	NP_S BJ				
9	여 전 히	여전히/MAG		10	AP				
10	안 갓 속 이 고,	안갓속/NNG + 이/VCP + 고/EC + /SP	안 갓 속 /001 + 이/006	19	VNP_ AJT	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	M-LST	19/키우다	
11	중 국 의	중국/NNP + 의/JKG	중 국 /002	12	NP_M OD				
12	'제 로 코 로 나'	'/SS + 제로 코로나/NNP + '/SS	제 로 코 로 나 /002	13	NP				
13	정 책	정책/NNG	정 책 /003	14	NP				
14	중 단 이	중단/NNG + 이/JKS	중 단 /007	16	NP_S BJ	11, 12, 13, 14	EFF	16/미치다	
15	경 제 에	경제/NNG + 에/JKB	경 제 /005	16	NP_A JT	15	GOL	16/미치다	
16	미 칠	미치/VV + 르/ETM	미 치 /006	17	VP_M OD				3.2 17 영향 영 향 do 을 ob je ct
17	영 향 도	영향/NNG + 도/JX	영 향 /004	19	NP_S BJ	11, 12, 13, 14, 15, 16, 17	EFF	19/키우다	
18	불 확 실 성 을	불 확 실 성 /NNG + 을 /JKO	불 확 실 성/001	19	NP_O BJ	18	THM	19/키우다	
19	키 우 고	키우/VV + 고/EC	키 우 /005	20	VP				3.2 7 러시 러시 sub 아 · 아 · ject

									우크 라이 나전 쟁	우크 라이 나전 쟁이
20	있다.	있/VX + 다 /EF + ./SF	있/023	-1	VP					

## [붙임7] JSON 형식

### 1) 형태소 분석

데이터 구조	데이터 타입	내용
id	str	NXMP25일련번호8자리(문어) SXMP25일련번호8자리(구어)
metadata	obj	메타데이터
title	str	국립국어원 말뭉치 추출 id
creator	str	국립국어원
distributor	str	국립국어원
year	str	말뭉치 구축 연도(2025)
category	str	신문 > 인터넷 기반 신문, 지역 종합지, 전문지, 전국 종합지/ 구어 > 공적 대화, 사적 대화, 일상 대화
annotation_level	arr[str]	형태 분석
sampling	str	본문 전체
document	arr[obj]	문서
id	str	추출 문서 id
metadata	obj	추출 메타데이터와 동일
title	str	문서 제목/담화 유형
author	str	기자/말화자
publisher	str	신문사/수집 방식
date	str	작성 일자/수집 일자
topic	str	기사 분야/대화 주제
sentence	arr[obj]	문장 정보
id	str	추출 문장 id
form	str	문장 문자열 (사업팀 수정 문장)
original_form	str	원시 말뭉치 문장 정보
word	arr	문장을 구성하는 어절
id	int	어절에 부여한 id(1부터)
form	str	개별 어절 문자열
begin	int	문자 시작 번호(0부터)
end	int	문자 끝 번호
MP	arr[obj]	형태소 분석 정보
id	int	형태소 id
form	str	형태소
label	str	형태소 품사 태그
word_id	int	형태소가 포함된 어절 id
positon	int	형태소의 어절 내 위치

## 2) 어휘의미 분석

데이터 구조	데이터 타입	내용
id	str	NXLS25일련번호8자리(문어) SXLS25일련번호8자리(구어)
metadata	obj	메타데이터
title	str	국립국어원 말뭉치 추출 id
creator	str	국립국어원
distributor	str	국립국어원
year	str	말뭉치 구축 연도(2025)
category	str	신문 > 인터넷 기반 신문, 지역 종합지, 전문지, 전국 종합지/ 구어 > 공적 대화, 사적 대화, 일상 대화
annotation_level	arr[str]	어휘의미 분석
sampling	str	본문 전체
document	arr[obj]	문서
id	str	추출 문서 id
metadata	obj	추출 메타데이터와 동일
title	str	문서 제목/답화 유형
author	str	기자/발화자
publisher	str	신문사/수집 방식
date	str	작성 일자/수집 일자
topic	str	기사 분야/대화 주제
sentence	arr[obj]	문장 정보
id	str	추출 문장 id
form	str	문장 문자열 (사업팀 수정 문장)
original_form	str	원시 말뭉치 문장 정보
word	arr	문장을 구성하는 어절
id	int	어절에 부여한 id(1부터)
form	str	개별 어절 문자열
begin	int	문자 시작 번호(0부터)
end	int	문자 끝 번호
WSD	arr[obj]	어휘의미 정보
form	str	어휘의미 form
sense_id	str	어휘의미 번호
pos	str	형태소 품사 태그
begin	int	어휘의 문장 내 시작 위치
end	int	어휘의 문장 내 끝 위치
word_id	int	대상 어휘 어절 번호

### 3) 구문 분석

데이터 구조	데이터 타입	내용
id	str	NXDP25일련번호8자리(문어) SXDP25일련번호8자리(구어)
metadata	obj	메타데이터
title	str	국립국어원 말뭉치 추출 id
creator	str	국립국어원
distributor	str	국립국어원
year	str	말뭉치 구축 연도(2025)
category	str	신문 > 인터넷 기반 신문, 지역 종합지, 전문지, 전국 종합지/ 구어 > 공적 대화, 사적 대화, 일상 대화
annotation_level	arr[str]	구문 분석
sampling	str	본문 전체
document	arr[obj]	문서
id	str	추출 문서 id
metadata	obj	추출 메타데이터와 동일
title	str	문서 제목/담화 유형
author	str	기자/발화자
publisher	str	신문사/수집 방식
date	str	작성 일자/수집 일자
topic	str	기사 분야/대화 주제
sentence	arr[obj]	문장 정보
id	str	추출 문장 id
form	str	문장 문자열 (사업팀 수정 문장)
original_form	str	원시 말뭉치 문장 정보
word	arr	문장을 구성하는 어절
id	int	어절에 부여한 id(1부터)
form	str	
begin	int	문자 시작 번호(0부터)
end	int	문자 끝 번호
DP	arr[obj]	구문 분석 정보
word_id	int	word에 기재한 id
word_form	str	word에 기재한 form
head	int	지배소 id
label	str	구문 태그_기능 태그
dependent	arr[int]	의존소 id



#### 4) 의미역 분석

데이터 구조		데이터 타입	내용
id		str	NXSR25일련번호8자리(문어) SXSR25일련번호8자리(구어)
metadata		obj	메타데이터
title		str	국립국어원 말뭉치 추출 id
creator		str	국립국어원
distributor		str	국립국어원
year		str	말뭉치 구축 연도(2025)
category		str	신문 > 인터넷 기반 신문, 지역 종합지, 전문지, 전국 종합지/ 구어 > 공적 대화, 사적 대화, 일상 대화
annotation_level		arr[str]	의미역 분석
sampling		str	본문 전체
document		arr[obj]	문서
id		str	추출 문서 id
metadata		obj	추출 메타데이터와 동일
title		str	문서 제목/답화 유형
author		str	기자/발화자
publisher		str	신문사/수집 방식
date		str	작성 일자/수집 일자
topic		str	기사 분야/대화 주제
sentence		arr[obj]	문장 정보
id		str	추출 문장 id
form		str	문장 문자열 (사업팀 수정 문 장)
original_form		str	원시 말뭉치 문장 정보
word		arr	문장을 구성하는 어절
id		int	어절에 부여한 id(1부터)
form		str	개별 어절 문자열
begin		int	문자 시작 번호(0부터)
end		int	문자 끝 번호
SRL		arr[obj]	의미역 분석 정보
predicate		arr[obj]	서술어 정보
form		str	서술어
lemma		str	서술어의 표제어
sense_id		str	서술어의 의미 번호
word_id		int	서술어의 word id
begin		int	서술어의 문장 내 시작 위치
end		int	서술어의 문장 내 끝 위치
argument		arr[obj]	논항 정보
form		str	논항
label		str	필수 의미역 태그 ex. LOC
begin		int	논항의 문장 내 시작 위치
end		int	논항의 문장 내 끝 위치
word_id		arr[int]	word의 id

adjunct	arr[obj]	부가어 정보
form	str	부가어
label	str	부가 의미역 태그 ex. M-LOC
begin	int	부가어의 문장 내 시작 위치
end	int	부가어의 문장 내 끝 위치
word_id	arr[int]	word의 id

## 5) 무형 대용어 복원

데이터 구조	데이터 타입	내용
id	str	NXZA25일련번호8자리(문어) SXZA25일련번호8자리(구어)
metadata	obj	메타데이터
title	str	국립국어원 말뭉치 추출 id
creator	str	국립국어원
distributor	str	국립국어원
year	str	말뭉치 구축 연도(2025)
category	str	신문 > 인터넷 기반 신문, 지역 종합지, 전문지, 전국 종합지/ 구어 > 공적 대화, 사적 대화, 일상 대화
annotation _level	arr[str]	무형 대용어 복원
sampling	str	본문 전체
document	arr[obj]	문서
id	str	추출 문서 id
metadata	obj	추출 메타데이터와 동일
title	str	문서 제목/담화 유형
author	str	기자/발화자
publisher	str	신문사/수집 방식
date	str	작성 일자/수집 일자
topic	str	기사 분야/대화 주제
sentence	arr[obj]	문장 정보
id	str	추출 문장 id
form	str	문장 문자열 (사업팀 수정 문장)
original _form	str	원시 말뭉치 문장 정보
word	arr	문장을 구성하는 어절
id	int	어절에 부여한 id(1부터)
form	str	개별 어절 문자열
begin	int	문자 시작 번호(0부터)
end	int	문자 끝 번호
ZA	arr[obj]	무형 대용어 복원 정보
predicate	obj	서술어 정보
form	str	서술어 문자열
sentence_id	str	서술어를 포함한 문장 id
word_id	int	서술어의 word id
begin	int	word 시작 번호(0부터)
end	int	word 끝 번호
ellipsis	arr[obj]	생략어 정보

restored	obj	복원 성분 정보
form	str	복원 성분 문자열
type	str	복원 성문의 문장 성분 유형 (주어, 목적어, 보어 등)
antecedent	obj	동일 지시적 성분 정보
form	str	동일 지시적 성분 문자열
sentence_id	str	동일 지시적 성분을 포함한 sentence id 없는 경우 null
begin	int	문자 시작 번호(0부터) 비지시대명사인 경우 null
end	int	문자 끝 번호 비지시대명사인 경우 null

## 6) 한국어 다층위 구조 분석

데이터 구조	데이터 타입	내용
id	str	NXML25일련번호8자리(문어) SXML25일련번호8자리(구어)
metadata	obj	메타데이터
title	str	국립국어원 말뭉치 추출 id
creator	str	국립국어원
distributor	str	국립국어원
year	str	말뭉치 구축 연도(2025)
category	str	신문 > 인터넷 기반 신문, 지역 종합지, 전문지, 전국 종합지/ 구어 > 공식 대화, 사적 대화, 일상 대화
annotation_level	arr[str]	형태 분석, 어휘의미 분석, 구문 분석, 의미역 분석, 무형 대용어 복원
sampling	str	본문 전체
document	arr[obj]	문서
id	str	추출 문서 id
metadata	obj	추출 메타데이터와 동일
title	str	문서 제목/답화 유형
author	str	기자/발화자
publisher	str	신문사/수집 방식
date	str	작성 일자/수집 일자
topic	str	기사 분야/대화 주제
sentence	arr[obj]	문장 정보
id	str	추출 문장 id
form	str	문장 문자열 (사업팀 수정 문장)
original_form	str	원시 말뭉치 문장 정보
word	arr	문장을 구성하는 어절
id	int	어절에 부여한 id(1부터)
form	str	개별 어절 문자열
begin	int	문자 시작 번호(0부터)
end	int	문자 끝 번호
MP	arr[obj]	형태 분석 정보
id	int	형태소 id
form	str	형태소
label	str	형태소 품사 태그
word_id	int	형태소가 포함된 어절 id
positon	int	형태소의 어절 내 위치
WSD	arr[obj]	어휘의미 분석 정보
form	str	어휘의미 form
sense_id	str	어휘의미 번호
pos	str	형태소 품사 태그
begin	int	어휘의 문장 내 시작 위치
end	int	어휘의 문장 내 끝 위치
word_id	int	대상 어휘 어절 번호
DP	arr[obj]	구문 분석 정보
word_id	int	word에 기재한 id

	word_form	str	word에 기재한 form
	head	int	지배소 id
	label	str	구문 태그_기능 태그
	dependent	arr[int]	의존소 id
SRL		arr[obj]	의미역 분석 정보
	predicate	arr[obj]	서술어 정보
	form	str	서술어
	lemma	str	서술어의 표제어
	sense_id	str	서술어의 의미 번호
	word_id	int	서술어의 word id
	begin	int	서술어의 문장 내 시작 위치
	end	int	서술어의 문장 내 끝 위치
	argument	arr[obj]	논항 정보
	form	str	논항
	label	str	필수 의미역 태그
	begin	int	논항의 문장 내 시작 위치
	end	int	논항의 문장 내 끝 위치
	word_id	arr[int]	word에 기재한 id
	adjunct	arr[obj]	부가어 정보
	form	str	부가어
	label	str	부가 의미역 태그
	begin	int	부가어의 문장 내 시작 위치
	end	int	부가어의 문장 내 끝 위치
	word_id	arr[int]	word에 기재한 id
ZA		arr[obj]	무형 대용어 복원 정보
	predicate	obj	서술어 정보
	form	str	서술어 문자열
	sentence_id	str	서술어를 포함한 문장 id
	word_id	int	서술어의 word id
	begin	int	word 시작 번호(0부터)
	end	int	word 끝 번호
	ellipsis	arr[obj]	생략어 정보
	restored	obj	복원 성분 정보
	form	str	복원 성분 문자열
	type	str	복원 성문의 문장 성분 유형 (주어, 목적어, 보어 등)
	antecedent	obj	동일 지시적 성분 정보
	form	str	동일 지시적 성분 문자열
	sentence_id	str	동일 지시적 성분을 포함한 sentence id 없는 경우 null
	begin	int	문자 시작 번호(0부터) 비지시대명사인 경우 null
	end	int	문자 끝 번호 비지시대명사인 경우 null

## [붙임8] 대조 목록(원문 수정)

### 1) 전문어나 고유명사 중 띄어서 표기해야 하는 것(표제어에 공백 포함)

JSON_form	Label	우리말샘 form
장기_복무	NNG	장기 복무
정식_사과	NNG	정식 사과
합동_점검	NNG	합동 점검
처우_개선	NNG	처우 개선
여행_상품	NNG	여행 상품
장학_재단	NNG	장학 재단
문화_체험	NNG	문화 체험
예비_창업자	NNG	예비 창업자
폭탄_발언	NNG	폭탄 발언
체험_프로그램	NNG	체험 프로그램
공식_홈페이지	NNG	공식 홈페이지
주연_배우	NNG	주연 배우
국제_대회	NNG	국제 대회
신년_인사회	NNG	신년 인사회
신년_하례식	NNG	신년 하례식
대표_팀	NNG	대표 팀
빅_맨	NNG	빅 맨
거점_센터	NNG	거점 센터
최고_출력	NNG	최고 출력
다음_달	NNG	다음 달
언론의자유	NNG	언론의 자유
반지의_제왕	NNP	반지의 제왕
대면_수업	NNG	대면 수업
조선_시대	NNG	조선 시대
건강_증진	NNG	건강 증진
시사_교양	NNG	시사 교양
커플_매니저	NNG	커플 매니저
직장_생활	NNG	직장 생활
법정_공방	NNG	법정 공방
설_연휴	NNG	설 연휴
도서_지역	NNG	도서 지역
무순위_청약	NNG	무순위 청약
예비_입주자	NNG	예비 입주자
무작위_추첨	NNG	무작위 추첨

벼랑_끝	NNG	벼랑 끝
보도_자료	NNG	보도 자료
최고_수준	NNG	최고 수준
상승_추세	NNG	상승 추세
선택_폭	NNG	선택 폭
극_중	NNG	극 중
얼음_공주	NNG	얼음 공주
첫_방송	NNG	첫 방송
단체_관광객	NNG	단체 관광객
남자_친구	NNG	남자 친구
입술_화장	NNG	입술 화장
눈_화장	NNG	눈 화장
소속_팀	NNG	소속 팀
손_키스	NNG	손 키스
짧은_순간	NNG	짧은 순간

2) 전문어나 고유명사 중 띄어쓰는 것이 원칙이나 붙여 쓰는 것도 허용한 것(표제어에 ^ 포함)

JSON_form	Label	우리말샘 form
진상규명	NNG	진상^규명
군복무	NNG	군^복무
후생사업	NNG	후생^사업
헌법기관	NNG	헌법^기관
출석요구서	NNG	출석^요구서
경제지표	NNG	경제^지표
대규모집회	NNG	대규모^집회
간호인력	NNG	간호^인력
방송활동	NNG	방송^활동
관람등급	NNG	관람^등급
디지털콘텐츠	NNG	디지털^콘텐츠
고강도방역	NNG	고강도^방역
제로코로나	NNG	제로^코로나
기술창업	NNG	기술^창업
시장진입	NNG	시장^진입
국가정체성	NNG	국가^정체성
파리올림픽	NNP	파리^올림픽
도핑테스트	NNG	도핑^테스트
축구선수	NNG	축구^선수



조정기능	NNG	조정^기능
소셜미디어	NNG	소셜^미디어
사업자등록	NNG	사업자^등록
내부구조	NNG	내부^구조
사업자등록번호	NNG	사업자^등록^번호
개인명의	NNG	개인^명의
연속방송	NNG	연속^방송
오일머니	NNG	오일^머니
테니스선수	NNG	테니스^선수
신흥시장	NNG	신흥^시장
권력구조	NNG	권력^구조
바이오헬스산업	NNG	바이오헬스^산업
안전보장	NNG	안전^보장
입법취지	NNG	입법^취지
정신적고통	NNG	정신적^고통
대한민국국민	NNG	대한민국^국민
예능프로그램	NNG	예능^프로그램
스위치수비	NNG	스위치^수비
레버리지투자	NNG	레버리지^투자
시장규모	NNG	시장^규모
산업구조	NNG	산업^구조
연구개발	NNG	연구^개발
기술경쟁력	NNG	기술^경쟁력
삶의질	NNG	삶의^질
성능평가	NNG	성능^평가
품목분류	NNG	품목^분류
무역정책	NNG	무역^정책
공동연구	NNG	공동^연구
정지상태	NNG	정지^상태
미니앨범	NNG	미니^앨범

### 3) 숫자, 영문이 포함된 단어 및 구성

word form	pos	우리말샘 form
112	NNP	일일이
119	NNP	일일구
2030	NNP	이공삼공
3040	NNP	삼공사공
10월	NNP	시월
119안전센터	NNG	일일구^안전^센터

11월	NNG	십일월
12월	NNG	십이월
16강	NNG	십육강
16강전	NNG	십육강전
19금	NNG	십구금
1군	NNG	일군
1기신도시	NNG	일기^신도시
1대1	NNG	일대일
1루	NNG	일루
1루수	NNG	일루수
1박	NNG	일박
1분기	NNG	일분기
1사	NNG	일사
1심	NNG	일심
1월	NNG	일월
1주년	NNG	일주년
1주일	NNG	일주일
1차선도로	NNG	일차선^도로
1회성	NNG	일회성
20대	NNG	이십 대
2관왕	NNG	이관왕
2군	NNG	이군
2루	NNG	이루
2루수	NNG	이루수
2루타	NNG	이루타
2분기	NNG	이분기
2사	NNG	이사
2심	NNG	이심
2월	NNG	이월
2인자	NNG	이인자
2차전지	NNG	이차^전지
3.1절	NNG	삼일절
3·1운동	NNG	삼일^운동
3040_세대	NNG	삼공사공 세대
30대	NNG	삼십 대
3구삼진	NNG	삼구^삼진
3국	NNG	삼국
3급	NNG	삼급
3남	NNG	삼남
3루	NNG	삼루

3루수	NNG	삼루수
3분기	NNG	삼분기
3선	NNG	삼선
3성장군	NNG	삼성^장군
3월	NNG	삼월
3인방	NNG	삼인방
3인칭	NNG	삼인칭
3점슛	NNG	삼^점^슛
3중고	NNG	삼중고
3차원적	NNG	삼차원적
3파전	NNG	삼파전
40대	NNG	사십 대
4가독감백신	NNG	사가^독감^백신
4강	NNG	사강
4강전	NNG	사강전
4분기	NNG	사분기
4원소	NNG	사원소
4월	NNG	사월
4차산업혁명	NNG	사차^산업^혁명
50대	NNG	오십 대
5월	NNG	오월
5지선다형	NNG	오지선다형
6월	NNG	유월
7월	NNG	칠월
8강	NNG	팔강
8강전	NNG	팔강전
8월	NNG	팔월
9월	NNG	구월
A/S	NNG	에이에스
A4용지	NNG	에이포 용지
AfDB	NNG	에이에프디비
AI챗봇	NNG	에이아이^챗봇
A급	NNG	에이급
A매치	NNG	에이^매치
B2B	NNG	비투비
B2C	NNG	비투시
B2G	NNG	비투지
BUSAN	NNG	부산
B형	NNG	비형
B형간염바이러스	NNG	비형^간염^바이러스

DéjàVu	NNG	데자뷔
D램	NNG	디램
e-마켓플레이스	NNG	이마켓플레이스
e스포츠	NNG	이^스포츠
e-쿠폰	NNG	이쿠폰
F1	NNG	에프원
FA컵	NNG	에프에이^컵
G20	NNG	지^이십
G7	NNG	지세븐
IoT	NNG	아이오티

## 참고문헌

- 국립국어원(2019가), “의미역기술 모형 연구 및 시범 구축”, 국립국어원 연구용역 결과보고서(연구책임: 이선웅).
- 국립국어원(2019나), “형태 분석 말뭉치 구축”, 국립국어원 연구용역 결과보고서 (연구책임: 김일환).
- 국립국어원(2020가), “2019 의미역분석 말뭉치 구축”, 국립국어원 연구용역 결과 보고서(연구책임: 임성모).
- 국립국어원(2020나), “2020년 어휘의미 말뭉치 연구 분석 사업 ”, 국립국어원 연구용역 결과보고서(연구책임: 김일환).
- 국립국어원(2020다), “구문 및 무형 대용어 복원 말뭉치 연구 분석”, 국립국어원 연구용역 결과보고서(연구책임: 곽용진).
- 국립국어원(2020라), “어휘의미 분석 말뭉치 구축”, 국립국어원 연구용역 결과보고서(연구책임: 이도길).
- 국립국어원(2020마), “의미역기술 모형 연구 분석”, 국립국어원 연구용역 결과보고서(연구책임: 이선웅).
- 국립국어원(2020바), “주격 무형 대용어 복원 말뭉치 구축”, 국립국어원 연구용역 결과보고서(연구책임: 곽용진).
- 국립국어원(2024가), “2024년 구문 분석 및 무형 대용어 복원 말뭉치 연구”, 국립국어원 연구용역 결과보고서(연구책임: 이도길).
- 국립국어원(2024나), “2024년 신문 기사 원문 자료 수집 및 정제”, 국립국어원 연구용역 결과보고서(연구책임: 윤종웅)
- 국립국어원(2024다), “2024년 일상 대화 말뭉치 구축”, 국립국어원 연구용역 결과 보고서(연구책임: 이용주).
- 김종복(2004), 한국어 정보화와 구문 분석, 월인.
- 박진호(2004), “의존 문법에 기반한 한국어 구문분석기의 설계와 구현”, 한국어 정보화와 구문분석, 월인, pp. 287-323.
- 신서인(2019), “한국어 의미역 말뭉치 구축을 위한 의미역 표지 연구”, 한국어 의

- 미학 66, 한국어의미학회, pp. 175-204.
- 신서인(2022), “의미역표지 부착을 위한 부가어의 의미역 연구”, 언어사실과 관점 57, 연세대학교 언어정보연구원, 131-168.
- 신서인·채숙희(2025), “한국어 인용동사구문 분석 방안 연구: 형태 및 구문 분석 말뭉치를 중심으로”, 언어와 정보사회 56, 서강대학교 언어정보연구소, 147-182.
- 양명희(1998), 현대국어 대용어에 대한 연구, 태학사.
- 유현경(2005), “부사절을 필수적으로 요구하는 구문에 대한 연구”, 한국어학 29, 한국어학회, pp. 159-186.
- 이홍식(2020), “부가어의 의미역에 대하여”, 한국어와 문화 27, 숙명여자대학교 한국어문화연구소, pp. 5-29.
- 임홍빈·이홍식 외(2002), 한국어 구문분석 방법론, 한국문화사.
- 진송미·오진영·차정원(2025), Espresso5: 딥러닝을 이용한 경량화 한국어 분석기, 한국정보과학회 학술발표논문집, pp. 2039-2041.

<기획·연구>

국립국어원 이현주 언어정보과장

국립국어원 이유원 학예연구사

국립국어원 김예영 연구원

<연구 참여자>

연구 책임자 신서인(한림대학교 산학협력단)

공동 연구원 차정원(데이터리)

연구 보조원 이한범, 이수현, 박진현, 전상호, 김나현,

이준희, 정민교 (한림대학교 산학협력단)

최윤진, 이민희, 서민지, 박수연, 유승빈,

전광훈 (데이터리)

---

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2025년 12월 16일

발행일: 2025년 12월 16일

인 쇄: 보림 S&P

---

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘2025년 한국어 구조 분석 및  
무형 대용어 복원 말뭉치 구축 연구’ 사업의 결과물을 발간한 것입니다.

