

국립국어원 2019-01-24

발간등록번호

11-1371028-000777-01

# 어휘의미 분석 말뭉치 구축

연구 책임자  
이도길



국립국어원



## 제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 '어휘의미 분석 말뭉치 구축'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업기간: 2019년 7월 ~ 2020년 1월

2020년 1월 8일

연구책임자: 이도길(고려대학교)

연구 기관	고려대학교 산학협력단 서울대학교 산학협력단 성신여자대학교 산학협력단
연구 책임자	이도길
공동 연구원	김일환, 박진호, 송상헌, 유현조, 윤태진, 이규범, 이영제, 정성훈, 정연주, 최운호



## 어휘의미 분석 말뭉치 구축

이 사업은 인공 지능의 발전을 위한 우리말 기초 자원으로서 고품질의 한국어 어휘의미 분석 말뭉치를 구축하고, 어휘의미 분석 말뭉치 구축을 위한 표준적인 지침을 개발하는 데 주요 목적이 있다.

사업의 범위는 크게 두 부분으로 나눌 수 있다. 첫째는 어휘의미 분석 말뭉치 구축 지침의 수립으로, 동형이의어를 분별한 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 다의어를 분별하는 수준에 맞게 수정하고 상세한 예를 통하여 구체화하였다. 둘째는 어휘의미 분석 말뭉치 구축으로, 새로 마련된 어휘의미 분석 말뭉치 구축 지침을 바탕으로 총 300만 어절 규모(문어 200만 어절, 구어 100만 어절)의 어휘의미 분석 말뭉치를 구축하였다.

어휘의미 분석 말뭉치 구축 지침을 수립하기 위하여 <21세기 세종계획>의 어휘의미 분석 말뭉치 구축 지침을 기반으로 삼되 문제점을 보완하였다. 이 과정에서 어휘의미 분석 질의응답 게시판을 운영하여 자주 질문되는 문제를 해결할 수 있는 설명과 사례를 지침에 추가하였다. 지침 보완의 방향성과 주요 보완 사항은 다음과 같다.

- ① 사전 미등재 유형의 구분을 위한 어휘의미 표지 신설
  - <우리말샘>에 형태가 등재되지 않은 경우 ‘777’
  - <우리말샘>에 형태는 등재되어 있으나 해당 의미가 없는 경우 ‘888’
  - 말뭉치 원어절의 오타, 탈자 등 표기 오류인 경우 ‘999’
- ② 분석 대상 품사의 차이에 따른 세부 지침의 신설
  - 고유명사와 관련하여 동명이인 등은 의미 미등재어로 다루었다.
  - 수사는 형태 분석된 품사와 일치하는 의미 번호를 부여하였다.
  - 분석 대상 어휘에 알파벳과 숫자가 있는 경우 이를 한글 표기로 전환한 형태가 <우리말샘>에 등재되었는가를 살폈다.
  - 분석 대상 어휘에 기호가 있는 경우 이를 생략한 형태가 <우리말샘>에 등재되었는가를 살폈다.
- ③ 기존 규정에 대한 상세 기술

- 의미를 분석할 때 전후 맥락을 최대한 활용하는 것을 원칙으로 하였다.
- 갈래 뜻 사이에 구분이 어려운 경우 예문을 검토하여 의미를 확정하였다.
- 한쪽에만 예문이 있는 경우 예문이 있는 갈래 뜻으로 분석하였다.
- 양쪽 모두 예문이 없는 경우 포괄적인 의미를 선택하였다.

본 사업에서 구축한 어휘의미 분석 말뭉치의 총 규모는 300만 어절로, 기사문으로 구성된 200만 어절의 문어 말뭉치와 뉴스, 인터뷰, 일상 대화 등으로 이루어진 100만 어절의 구어 말뭉치로 구성되었다. 어휘의미 분석 말뭉치의 기반이 되는 형태 분석 말뭉치는 ‘형태 분석 말뭉치 구축(2019년)’ 사업의 결과물을 어휘 분석용 말뭉치 수준으로 변환시킨 것으로, 형태 분석 말뭉치 구축 과정에서 분석된 접사 중 일부를 어근에 재통합시킨 말뭉치이다.

어휘의미 분석 말뭉치 구축은 ‘어휘 분석용 말뭉치 변화 → 분석 도구(워크벤치) 구현 → 작업 교육 → 자동 분석 → 분석 결과 수정 → 최종 결과물 산출’의 순으로 이루어졌다.

구축 작업은 어휘의미 분석 말뭉치 구축에 최적화된 워크벤치를 개발·활용하였다. 작업자와 검수자는 워크벤치에서 자신에게 배분된 어휘를 확인하고 어휘의미 번호를 부여하였다. 어휘의미 번호 입력은 드롭다운 선택 방식을 택하여 입력 오류를 원천적으로 차단하였다.

본 사업에서는 말뭉치 구축의 편의를 도모하고 정확성을 높이기 위하여 지도학습 기반의 고성능 어휘의미 분석기(서울대 어휘의미 분석기)를 사용하였다. 서울대 어휘의미 분석기는 <우리말샘>의 의미 식별번호에 대응하는 방식으로 어휘의미를 분석하며 작업 초기 실사용 테스트에서 87.05%의 정확도를 보였다.

분석 결과 수정은 3단계로 이루어졌다. 1단계는 작업자가 자동 분석 결과를 수정하는 단계이고, 2단계는 자동 분석 결과와 작업자 수정 결과를 비교하며 검수자가 형태 분석 결과를 검수하는 단계이다. 3단계는 상위 작업자 그룹이 어휘 분석 과정에서 신고된 형태 분석의 오류를 검토하고 수정하여 의미 번호를 확정하거나 어휘 재분배를 통하여 의미 분석을 재요청하는 단계이다. 이러한 과정을 거쳐 만들어진 최종 결과물은 JSON 형식으로 변환하여 산출하였다.

# 차례

제1장 서론 .....	1
1. 사업의 목적 .....	1
2. 사업의 범위 .....	3
2.1. 어휘의미 분석 말뭉치 구축 지침 수립 .....	3
2.2. 어휘의미 분석 말뭉치 구축 .....	4
제2장 어휘의미 분석 말뭉치의 구축 .....	5
1. 어휘의미 분석 말뭉치의 구성 .....	5
1.1. 분석 말뭉치의 규모 .....	5
1.2. 분석 대상 어휘의 규모 .....	7
2. 어휘의미 분석 말뭉치 구축 절차 .....	9
2.1. 어휘 분석용 말뭉치 변환 .....	9
2.2. 분석 도구(워크벤치) 구현 .....	20
2.3. 작업 교육 .....	29
2.4. 자동 분석 .....	29
2.5. 분석 결과 수정 .....	30
2.6. 최종 결과물 산출 .....	31
제3장 어휘의미 분석 말뭉치 구축 지침 수립 .....	36
1. 지침 수립 과정 .....	36
1.1. <21세기 세종계획> 형태 분석 말뭉치 구축 지침 검토 .....	36
1.2. 지침의 보완 방향 .....	39
2. 어휘의미 분석 말뭉치 구축 지침 .....	56

제4장 결론 ..... 68

<ABSTRACT> ..... 70

부록 1: 어휘의미 분석 말뭉치(문어)의 예시 ..... 73

부록 2: 어휘의미 분석 말뭉치(구어)의 예시 ..... 92



# 제1장 서론

## 1. 사업의 목적

- 어휘의미 분석 말뭉치 구축 지침 수립
- 어휘의미 분석 말뭉치(300만 어절) 구축

이 사업은 인공 지능의 발전을 위한 우리말 기초 자원으로서 고품질의 한국어 어휘의미 분석 말뭉치를 구축하고, 어휘의미 분석 말뭉치 구축을 위한 표준적인 지침을 개발하는 데 주요 목적이 있다. 이 사업을 통해 구축된 어휘의미 분석 말뭉치는 학계뿐 아니라 산업계에서 인공 지능 개발 등을 위한 토대 자원으로 활용될 것이다.

한국어 처리 인공 지능의 발전을 위해서는 높은 정확성과 일관성을 갖춘 대규모의 어휘의미 분석 말뭉치가 요구된다. 동형이의어로 인해 발생하는 어휘의 중의성 문제는 자연어 처리에 있어서 반드시 해결해야 할 중요한 문제 중 하나로 이를 적절히 해소하기 위해서는 어휘의 의미를 정교하게 식별하고 분류해낼 수 있는 어휘의미 분석 시스템이 필수적이다. 어휘의미 분석 말뭉치는 이러한 의미 분석 시스템 개발의 가장 기본적인 학습 자원이라는 점에서 이를 확장하고 정교화하는 작업은 국어 인공 지능 기술 발전의 필수적인 부분이라 할 수 있다.

그러나 2007년 배포된 <21세기 세종계획>을 마지막으로 공공 자원으로서의 어휘의미 분석 말뭉치는 전혀 구축되지 못하였다. 그로 인해 최근 인공 지능 개발이 요구하는 학습 자원의 양에 비해 사용 가능한 자원의 규모는 매우 부족한 형편이다. 게다가 오늘날과 <21세기 세종계획>이 구축된 시점 사이에는 시간적 간극이 있어, 그 사이 새로 유입된 신어나 어휘의미의 변화를 인공 지능 기술 개발에 반영할 수 없었다.

이에 본 사업에서는 2009~2017년에 생산된 300만 어절 규모의 원시 자원으로부터 어휘의미 분석 말뭉치를 구축하였으며, 이를 통해 공공 자원으로서의 대규모 어휘의미 분석 말뭉치 구축의 기반을 다시금 마련하였다.

한편 고성능 자동 번역과 언어를 통한 지능적 사물 제어 등 고급 인공 지능 시스템은 사용 맥락에 따른 어휘의미 차이를 구별할 수 있어야 한다. 이에 본 사업에서는 동형이

의어뿐만 아니라 다의어 구별 수준까지 분석을 정교화하였고, 그에 따라 <21세기 세종 계획>의 말뭉치 구축 지침을 수정·보완한 새로운 표준지침을 마련하였다. 분석 지침의 기준 사전으로는 <우리말샘>을 택하였다. <표 1>에서 확인할 수 있듯이 <우리말샘>은 기본적으로 의미 식별번호를 다의 수준에 부여하고 있고, 사용 맥락에 따른 다의 또한 풍부하게 제공하고 있어 이를 기준으로 한 본 사업의 어휘의미 분석 말뭉치는 의미 분석 자원의 고품질화에 기여할 것으로 기대된다.

표준국어대사전	우리말샘
<p>고구마 「명사」</p> <p>「1」 『식물』 메꽃과의 여러해살이풀. 줄기는 덩굴이 되어 땅 위로 뻗으며 꽃은 보통 피지 않으나 때로 연한 붉은빛의 꽃이 나팔 모양으로 피기도 한다. 땅속뿌리는 식용하거나 공업용으로 쓰고 잎과 줄기도 나물로 식용한다. 북아메리카가 원산지로 따뜻한 지방에서 재배된다. (Ipomoea batatas)</p> <p>「2」 ‘「1」’의 덩이뿌리. 흔히 길쭉한 타원형으로 녹말이 많아 식용하며 공업용으로도 쓴다. ≍감서, 감저, 남감저, 단감자, 저우.</p>	<p>고구마 「001」 「명사」 『식물』 메꽃과의 여러해살이풀. 줄기는 덩굴이 되어 땅 위로 뻗으며 꽃은 보통 피지 않으나 때로 연한 붉은빛의 꽃이 나팔 모양으로 피기도 한다. 땅속뿌리는 식용하거나 공업용으로 쓰고 잎과 줄기도 나물로 식용한다. 북아메리카가 원산지로 따뜻한 지방에서 재배된다.(Ipomoea batatas)</p> <p>고구마 「002」 「명사」 ‘고구마’의 덩이뿌리. 흔히 길쭉한 타원형으로 녹말이 많아 식용하며 공업용으로도 쓴다.</p> <p>고구마 「003」 「명사」 유통성이 없어 답답하게 구는 사람이나 일이 뜻대로 되지 않아 답답한 상황을 고구마를 먹고 목이 메는 것에 비유하여 이르는 말.</p>

<표 1> 표제어 ‘고구마’에 대한 <표준국어대사전>과 <우리말샘> 비교

이에 본 사업에서는 <21세기 세종계획>의 말뭉치 구축 지침을 수정·보완한 표준적인 지침을 마련하여 그것을 바탕으로 분석의 일관성을 높인 300만 어절 규모의 어휘의미 분석 말뭉치를 구축하였으며, 이를 통해 공공 자원으로서의 대규모 어휘의미 분석 말뭉치 구축의 기반을 다시금 마련하였다.

## 2. 사업의 범위

사업의 범위는 크게 두 부분으로 나눌 수 있다. 첫째는 어휘의미 분석 말뭉치 구축 지침 수립으로, 동형이의어를 구별하기 위한 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 다의어 구별에 맞게 수정하고 상세한 예를 통하여 구체화하였다. 둘째는 어휘의미 분석 말뭉치 구축으로, 새로 마련된 어휘의미 분석 말뭉치 구축 지침을 바탕으로 총 300만 어절 규모(문어 200만 어절, 구어 100만 어절)의 어휘의미 분석 말뭉치를 구축하였다.

### 2.1. 어휘의미 분석 말뭉치 구축 지침 수립

어휘의미 분석 말뭉치 구축 지침을 수립하기 위하여, <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 기반으로 삼아 문제점을 보완하고, <우리말샘>의 의미 체계에 따른 어휘의미 표준 지침을 개발하였다.

<21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침은 국내의 다른 어휘의미 분석 말뭉치 구축 과정에서 두루 활용되고 있을 만큼 표준적인 지위를 확보하고 있다. 그러나 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침은 <표준국어대사전>의 의미 체계를 따르며, 분석 대상 어휘를 동형이의어 수준으로 의미를 분석하였다. 또한 지침이 적용되는 다양한 상황에서의 예시가 거의 없어 지침에 따라 분석되는 의미의 일관성에 문제가 나타날 우려가 있다.

이에 본 사업에서는 <우리말샘>의 의미 체계를 따름으로써 다의어 수준의 의미 분석이 가능한 지침을 구축하였다. 또한 형태와 의미 차원을 구분하여 형태 미등재어와 의미 미등재어를 위한 별도의 어휘의미 표지를 각각 마련하고, 말뭉치의 오타 및 탈자 등을 표시하기 위한 표지를 설정하여 표기 오류를 구분하였다. 또한 분석 지침의 항목별

설명을 좀 더 상세히 제시하고 다양한 사례를 포함함으로써 실제 구축 과정에서 도출되는 오류와 비일관성을 줄이고자 하였다. 어휘의미 말뭉치 구축 과정에서 어휘의미 분석 질의응답 게시판을 운영하며 수집된 다양한 사례를 지침에 추가하였다.

## 2.2. 어휘의미 분석 말뭉치 구축

수립된 어휘의미 분석 말뭉치 구축 지침을 바탕으로 300만 어절 규모의 어휘의미 분석 말뭉치를 구축하였다. 말뭉치는 문어 200만 어절, 구어 100만 어절로 구성되어 있으며 동일한 형태 분석 말뭉치 구축 지침으로 문어와 구어를 아울러 분석할 수 있음을 확인하였다.

말뭉치 구축의 편의를 도모하고 정확성을 높이기 위하여 높은 분석 정확률을 갖춘 어휘의미 분석기를 작업에 활용하였다. 고려대 어휘의미 분석기는 1,500만 어절 규모의 <세종 형태의미 분석 말뭉치>로부터 어휘의미 분석 결과를 학습하여 개발한 것이다. 서울대 어휘의미 분석기는 <우리말샘>의 단어 명사 346,047개, 다의 명사 센스 185,087개를 대상으로 용례 9만 문장, 88만 어절 규모의 학습 데이터 구축하고, 이를 LSTM 기반의 신경망 모델을 통해 학습한 것으로, <우리말샘>의 의미 식별번호에 대응이 가능하다.

한편으로 어휘의미 분석 말뭉치 구축에 최적화된 워크벤치를 개발하였다. 워크벤치에서는 서울대 어휘의미 분석기의 어휘의미 분석 결과를 보여 주고 드롭다운 선택 방식을 통해 입력 오류를 원천적으로 차단하였다. 또한 동일 어절에 대한 어휘의미 분석기의 분석 결과와 작업자의 분석 결과를 비교하고 분석 결과가 일치하지 않는 부분에 바탕색을 다르게 표시하여 검수 작업의 효율을 높였다. 이처럼 동일한 분석 대상 어휘를 작업자와 검수자가 한 차례씩 의미 분석하여 직관에 의한 오분석을 최소화하였다.

이를 바탕으로 원시 말뭉치로부터 형태 분석된 말뭉치에 나타나는 체언을 대상으로 <우리말샘>의 의미 번호를 부착하는 작업을 하였다. 다만 형태 분석 과정에서 분리된 접사 중 일부를 앞말과 뒷말에 통합하는 과정을 거쳤다. 또한 <우리말샘>에 해당 형태가 없는 경우 어휘의미 표지 '777'을, 형태는 있되 해당 의미가 없는 경우 어휘의미 표지 '888'을 부여하여 사전 미등재 유형을 형태와 의미 차원으로 구분하고, 말뭉치 원어절의 오타, 탈자 등에 어휘의미 표지 '999'를 부여하여 표기 오류를 고려하였다.

## 제2장 어휘의미 분석 말뭉치의 구축

### 1. 어휘의미 분석 말뭉치의 구성

#### 1.1. 분석 말뭉치의 규모

이 사업을 통해 구축된 어휘의미 분석 말뭉치의 총 규모는 300만 어절(문어 200만 어절, 구어 100만 어절)이다. 어휘의미 분석 말뭉치의 기반이 되는 원시 말뭉치는 2004년 이후 생산된 현대 국어 자료를 수집한 ‘2018년 국어 말뭉치 연구 및 구축’ 사업 결과물이다. 그중 문어 말뭉치는 전체가 2009~2017년에 작성된 기사문에서 발췌되었으며, 여러 신문사의 다양한 주제로 구성되었다. <표 2>는 문어 말뭉치의 규모를 신문사별 정리한 것이고, <표 3>은 주제별로 정리한 것이다.

	2009년	2010년	2011년	2012년	2013년	2014년	2015년	2016년	2017년	합계
동아일보	63,255	65,395	68,339	99,455	70,978	102,223	65,855	96,257	65,658	697,415
조선일보	94,124	98,183	94,644	3,886	107,244	4,181	98,291	392	105,934	606,879
한겨레	72,357	72,244	72,649	85,397	73,356	87,460	73,967	86,092	72,233	695,755
합계	229,736	235,822	235,632	188,738	251,578	193,864	238,113	182,741	243,825	2,000,049

<표 2> 문어 말뭉치(200만 어절)의 구성 (신문사별)

	동아일보	조선일보	한겨레	합계	비율
경제	96,237	45,603	123,893	265,733	13.3%
과학	10,406	358	15,995	26,759	1.3%
국제	45,338	56,231	82,917	184,486	9.2%
기획	57,833	3,600	0	61,433	3.1%
문화	86,670	79,006	77,602	243,278	12.2%
사람들	7,000	33,042	2,209	42,251	2.1%
사회	102,134	105,932	170,881	378,947	18.9%
스포츠	66,340	56,537	46,321	169,198	8.5%
오피니언	60,667	46,126	28,815	135,608	6.8%
정치	141,546	124,396	130,317	396,259	19.8%
지역	23,244	56,048	16,805	96,097	4.8%
합계	697,415	606,879	695,755	2,000,049	100%

<표 3> 문어 말뭉치(200만 어절)의 구성 (기사 주제별)

구어 말뭉치는 공적 독백과 공적 대화, 사적 대화로 구성되었다. 공적 독백에 해당하는 자료에는 뉴스, 강연, 학술 발표 자료가 있고, 공적 대화에 속하는 자료에는 방송 대화와 인터뷰 자료가 있으며, 사적 대화 자료에는 2인의 일상 대화 자료가 포함되었다. 이를 정리하면 <표 4>와 같다.

분류	자료명	비고	어절 수	비율
공적 독백	EBS 정오 뉴스	2018년	24,756	2.5%
	강연	학술발표회(한글대장경의 구성과 번역의 문제) 학술특강(사회언어학 이후의 사회언어학) 학술발표회(언어 연구와 문법) 학술발표회(언어 연구와 정보, 복합지식) 학술발표회(학습용 기본 명사 언어 빈도 사전)	48,665	4.9%
	제1~5회 일송학술대회	2009년~2013년	62,153	6.2%
공적 대화	EBS 초대석		203,015	20.3%
	김어준의 뉴스공장	TBS	130,031	13%
	뜨거운 사이다	온스타일, 2017년	68,891	6.9%
	인터뷰 전사	정현중, 윤여순, 정바비 인터뷰 (2014년)	19,485	1.9%
	최고의 요리 비결	EBS	150,395	15%
	팟캐스트	프로파일러 배상훈의 CRIME (2016~2018년) 풀어 듣는 문화 이야기 (2016~2017년) 서늘한 마음썰 (2018년)	60,215	6%
사적 대화	2인 일상대화	2018년	233,854	23.3%
합계			1,001,460	100%

<표 4> 구어 말뭉치(100만 어절)의 구성

## 1.2. 분석 대상 어휘의 규모

어휘의미 분석 말뭉치의 분석 대상 품사는 기능어(function words)를 제외한 내용어(content words) 중 체언을 중심으로 선정하였다. 다시 말해 형태 분석된 어휘 중 일반

명사(NNG), 고유명사(NNP), 의존명사(NNB), 대명사(NP), 수사(NR), 어근(XR)이 대상 범주가 된다. 문어 말뭉치와 구어 말뭉치에서 출현한 체언류의 빈도를 대상 범주별로 정리하면 <표 5>와 같다.

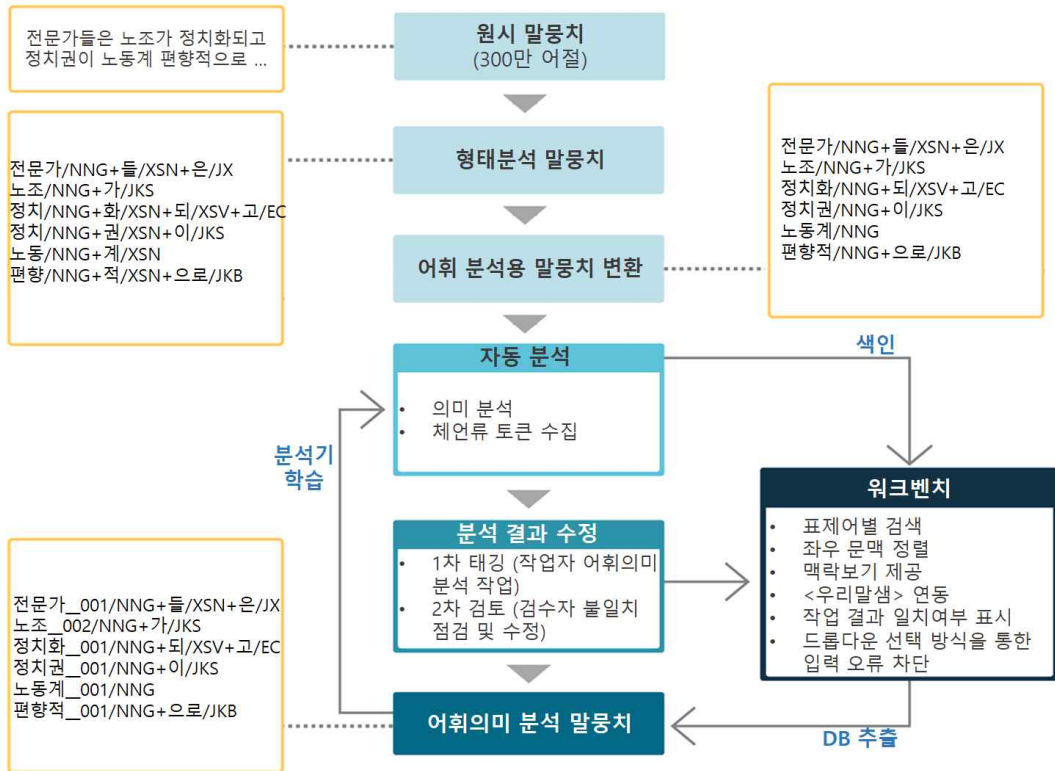
	문어 말뭉치	구어 말뭉치	합계	비율
일반명사(NNG)	1,247,972	315,388	1,563,360	73.05%
고유명사(NNP)	201,848	20,799	222,647	10.40%
의존명사(NNB)	181,241	63,406	244,647	11.43%
대명사(NP)	22,901	53,485	76,386	3.57%
수사(NR)	18,349	14,224	32,573	1.52%
어근(XR)	375	174	549	0.03%
합계	1,672,686	467,476	2,140,162	100%

<표 5> 어휘의미 분석 말뭉치의 분석 규모



## 2. 어휘의미 분석 말뭉치 구축 절차

어휘의미 분석 지침을 수립한 후 원시 말뭉치로부터 어휘의미 분석 말뭉치가 구축되는 과정을 정리하면 <그림 1>과 같다.



<그림 1> 어휘의미 말뭉치 구축 과정

### 2.1. 어휘 분석용 말뭉치 변환

어휘의미 분석 말뭉치는 “형태 분석 말뭉치 구축(2019년)” 결과를 바탕으로 하여 형태 분석 말뭉치의 원어절과 분석 결과 모두를 가급적 수정하지 않은 채 체언을 중심으로 구축되었다. 그러나 어휘의미의 기본 단위를 <우리말샘>에 등재된 단어로 설정하였기 때문에 생산성이 높은 접사를 분석한 “형태 분석 말뭉치 구축(2019년)”의 결과를 어

휘의미 분석에 적합하게 변환하는 과정이 요구되었다.

### 2.1.1. 통합 접사의 선정

어휘의미 분석 말뭉치를 구축과 관련하여 “형태 분석 말뭉치 구축(2019년)” 결과를 어휘 분석용 말뭉치 수준으로 변환하는 과정에서 접사의 통합 여부가 주요 쟁점으로 대두되었다. “형태 분석 말뭉치 구축(2019년)”에서 수립한 형태 분석 지침에서 분석 대상으로 삼은 접사의 목록과 접사 처리(통합) 여부를 정리하면 다음과 같다.

● 체언접두사			
가(假)	가건물	소(小)	소강당
고(高)	고물가	신(新)	신정당
과(過)	과보호	왕(王)	왕족발
구(舊)	구소련	재(再)	재충전
날	날음식	저(低)	저임금
노(老)	노부부	제(第)	제13차
대(大)	대전배	준(準)	준전시
말	말아들	초(超)	초만원
맨	맨몸	최(最)	최고급
무(無)	무의식	친(親)	친러시아
미(未)	미완성	탈(脫)	탈냉전
반(反)	반독재	폐(廢)	폐광산
범(汎)	범세계	푯	푯살구
부(不)	부도덕	피(被)	피고소인
불(不)	불합리	한	한가운데
비(非)	비논리	헛	헛고생
생(生)	생김치		

<표 6> “형태 분석 말뭉치 구축(2019년)”의 분석 대상 접두사 목록

<표 6>에 제시된 것처럼 형태 분석 말뭉치 구축 과정에서 어근과 분리하여 분석한 체언접두사는 총 33개이다. 이들의 경우 어근과 결합한 형태가 어휘의미의 기본 단위인 <우리말샘>에 등재된 단어이기 때문에 접사 처리(통합) 과정이 필요하다. 다만 체언접두사의 통합 과정에서 어근의 품사가 자동으로 승계되지 않고 전후 맥락에 따라 품사가 달라지는 것에 유의하여 변동된 품사로 재분석해야 한다.

<표 7>은 형태 분석 말뭉치 구축 지침에서 분석 대상으로 삼은 명사파생접미사이다. 이들은 체언접두사와 달리 앞말과의 결합형이 <우리말샘>에 등재될 만한 한 단어를 이루지 못하는 경우가 있다.

● 명사파생접미사			
가(哥)	김가	분지(分之)	삼분지 일
가(價)	매매가	빨	조카빨
가량	1시간가량, 다섯 명가량	산(産)	중국산
간(間)	한 달간	상(上)	역사상
경(頃)	두 시경	생1(生)	갑자생
계(界)	교육계	생2(生)	견습생
계(系)	몽고계	성(性)	인간성
광(狂)	메모광	시(視)	영웅시
권(券)	만 원권	씩	만원씩
권(圈)	운동권	어치	만원어치
권(權)	참정권	여(餘)	삼십여
기(氣)	기름기	용(用)	전쟁용
계	10분계	율(率)	출산율
꿀	십 원꿀	장이	간판장이
꾼	노름꾼	쟁이	심술쟁이
끼리	전우끼리	적(的)	사상적
네	동이네	정(整)	일만 원정
님	선생님	제(制)	봉건제

당(當)	한 사람당	질	서방질
대(臺)	억대	짜리	백 원짜리
택(宅)	청주택	째1	이틀째
들	우리들	째2	옹기째
들이	1ㄹ들이	썸	내일썸
론(論)	비평론	층(層)	선수층
류(類)	자연류	치(值)	기대치
률(率)	경쟁률	치레	인사치레
리(裡)	비밀리	투성이	먼지투성이
발(發)	서울발	풍(風)	복고풍
배기	열 살배기	하(下)	지배하
별(別)	가구별	형(型)	기본형
부(附)	12일부	형(形)	계란형
분(分)	3분의 일	화(化)	도구화

<표 7> “형태 분석 말뭉치 구축(2019년)”의 분석 대상 명사파생접미사 목록

명사파생접미사 중 앞말과의 결합형이 <우리말샘>에 등재될 만한 한 단어를 이루지 못하는 경우는 크게 두 가지로 구분된다. 첫째는 주로 구와 결합하여 앞말과 합쳤을 때 사전에 등재될 만한 한 단어를 이루지 못하는 경우이다. 주로 통사적 접사가 이에 해당한다.

(1) 주로 구와 결합하여 등재어를 이루지 못하는 경우

- 가량 다섯 명가량, 10%가량, 30세가량
- 간(間) 한 달간, 삼십 일간,
- 경(頃) 두 시경, 오전 9시경, 16세기경
- 계 10분계, 이달 말계, 서울역계
- 꿀 십 원꿀, 100원꿀, 한 명꿀, 열 개꿀
- 끼리 전우끼리, 우리끼리, 자기들끼리
- 당(當) 한 사람당, 40명당
- 대(臺) 만 원대, 백삼십만 원대, 수천억대

들	우리들, 사람들, 그들, 사건들
들이	1L들이, 한 말들이, 1리터들이
발(發)	3월 12일발, 열 시발
배기	두 살배기, 다섯 살배기
분지	삼분지 일
생1(生)	1960년 1월 1일생, 이십 년생
씩	두 사람씩, 열 그릇씩, 며칠씩, 하나씩
어치	한 푼어치, 천 원어치, 얼마어치
여(餘)	이십여 년, 백여 개, 십오 년여의
정(整)	일만 원정
짜리	백 원짜리, 한 뺨짜리, 방 두 개짜리
째1	이틀째, 며칠째, 다섯 달째, 두 잔째
쫘	내일쫘, 얼마쫘, 12월 20일쫘
하	독재 정권하, 식민 지배하, 분단 체제하

(1)에 제시된 명사파생접미사 중 ‘당(當)’, ‘발(發)’, ‘생1(生)’은 어근과의 결합형 일부가 <우리말샘>에 등재되어 있다. 그러나 복합어 자체의 의미 특성이 크게 나타나지 않기 때문에 접사 결합형에 의미 번호를 부여하는 것과 앞말에 의미 번호를 부여하는 것이 별반 다르지 않기 때문에 접사 처리(통합)에서 제외하였다.

다만 ‘세 살, 다섯 살’ 등 수량을 나타내는 구와 결합하는 ‘-배기’는 접사 처리(통합) 과정에서 제외하나, ‘나이배기, 알배기, 공짜배기, 대짜배기, 진짜배기’ 등 <우리말샘>에 등재된 어휘는 결합형으로 처리한다. 또한 ‘-째1’은 일반적으로 접사 처리(통합) 과정에서 제외하나, 의존명사 ‘번’과 결합하는 ‘-째1’은 앞말과의 결합형 ‘번째’를 의존명사로 처리했다.

둘째는 주로 고유명사와 결합하여 일반명사처럼 쓰이나, 결합형이 사전에 등재되어 있지 않는 경우로, 다음과 같은 명사파생접미사가 이에 속한다.

(2) 주로 고유명사와 결합하여 등재어를 이루지 못하는 경우

가(哥)	김가, 이가, 박가
계(系)	독일계, 중국계, 몽고계

네       철수네, 아저씨네, 김 서방네  
 택(宅)   안성택, 광주택, 상주택, 철수택, 참봉택

(2)에 제시된 ‘가(哥)’, ‘계(系)’, ‘네’, ‘택(宅)’은 주로 고유명사와 결합하나, 결합형이 <우리말샘>에 등재되어 있지 않아 특정 의미 번호를 부여하기가 어렵다. 이처럼 의미 번호를 부여하는 데에 효용성이 떨어지는 명사파생접미사는 접사 처리(통합)에서 제외하고자 한다.

이상의 내용을 바탕으로 접사 처리(통합) 과정을 통해 앞말과 결합하여 어휘의미를 분석하는 명사파생접미사는 <표 8>과 같다.

● 접사 처리(통합) 대상 명사파생접미사 목록			
가(價)	매매가	생2(生)	견습생
계(界)	교육계	성(性)	인간성
광(狂)	메모광	시(視)	영웅시
권(券)	만 원권	용(用)	전쟁용
권(圈)	운동권	율(率)	출산율
권(權)	참정권	장이	간판장이
기(氣)	기름기	쟁이	심술쟁이
꾼	노름꾼	적(的)	사상적
네	동이네	제(制)	봉건제
님	선생님	질	서방질
론(論)	비평론	짜2	웅기짜
류(類)	자연류	층(層)	선수층
률(率)	경쟁률	치(值)	기대치
리(裡)	비밀리	치레	인사치레
별(別)	가구별	투성이	먼지투성이
부(附)	12일부	풍(風)	복고풍
분(分)	3분의 일	형(型)	기본형

빨	조카빨	형(形)	계란형
산(産)	중국산	화(化)	도구화
상(上)	역사상		

<표 8> 접사 처리(통합) 대상 명사파생접미사 목록

<표 9>는 형태 분석 말뭉치 구축 지침에서 분석 대상으로 삼은 동사파생접미사와 형용사파생접미사이다. 이들은 어기 또는 어근에 붙어서 그것을 동사나 형용사로 만들어 주는 기능을 갖는 접미사이다. 이들 역시 접사의 의미가 파생어의 의미에 기여하는 바가 거의 없다. 그러므로 어휘의미 분석 말뭉치의 향후 활용성 등을 감안했을 때 결합형에 어휘의미 번호를 부여하는 것보다 어기 또는 어근의 의미를 분석하는 것이 보다 유의미하다.

● 동사파생접미사	
당하	아군이 <u>공격당하는</u> 데에는 이유가 있다.
되	아침식사가 이미 <u>준비되어</u> 있었다.
시키	오늘 강아지를 <u>운동시키려고</u> 공원에 나갔다.
하	외국에서 <u>공부하는</u> 일이 쉬운 것은 아니다.
받	몇몇은 집세 인상을 <u>강요받았다</u> .
● 형용사파생접미사	
답	사람이 <u>사람답게</u> 행동해야 사람이지,
되	<u>거짓된</u> 말은 들통 나기 마련이다.
롭	어려운 일일수록 <u>슬기롭게</u> 대처하라.
스럽	그녀의 <u>사랑스러운</u> 표정을 보거라.
하	<u>건강한</u> 신체에 건강한 정신이 깃든다.

<표 9> “형태 분석 말뭉치 구축(2019년)”의 분석 대상 용언화 접미사 목록

위에서 말한 바와 같이 어휘의미 분석 말뭉치를 구축하는 과정에서 접사 처리(통합)하는 것이 원칙이나, 당해의 구축 말뭉치는 서술성 명사류에 용언을 파생시키는 접미사

가 결합된 경우에는 접사 처리(통합) 과정을 거치지 않고 어근(어기)에 해당하는 부분에 어휘의미의 표지를 부여하였다.

### 2.1.2. 접사 처리(통합) 과정

어휘의미 분석에 앞서 어휘 분석용 말뭉치를 마련하기 위해 2.1.1.에서 언급한 통합 대상 접사를 앞말 혹은 뒷말과 결합하는 접사 처리 과정을 진행하였다. 다시 말해서 형태 분석 말뭉치에서 분리되었던 접두사와 <표 8>에 제시된 명사파생접미사를 어근에 다시 결합시켰다. 다만 접사 처리(통합) 대상이 되는 접사일지라도 이 과정에서 따옴표나 괄호, 띄어쓰기 등으로 어근과 직접적인 결합이 불가능한 경우 접사 처리(통합) 과정에서 제외하였다.

(3) 따옴표나 괄호, 숫자 등으로 인해 접사 처리(통합) 과정에서 제외된 경우

가. 친(親)시장적 친/XPN+ (/SS+ 親/SH+ )/SS+ 시장적/NNG

가'. 친동생과 친동생/NNG+ 과/JC

나. (재)재협상은 (/SS+ 재/XPN+ )/SS+ 재협상/NNG+ 은/JX

나'. 재발견의 재발견/NNG+ 의/JKG

다. 비(非)수술적 비/XPN+ (/SS+ 非/SH+ )/SS+ 수술적/NNG

다'. 비공개로 비공개/NNG+ 로/JKB

라. 노(老)신사가 노/XPN+ (/SS+ 老/SH+ )/SS+ 신사/NNG+ 가/JKS

라'. 노부부의 노부부/NNG+ 의/JKG

마. 제2차 제/XPN+ 2/SN+ 차/NNB

마'. 제철공화국 제철/NR+ 공화국/NNG

(3)에 나타난 접두사 ‘친-, 재-, 비-, 노-’는 어휘의미 분석을 위한 말뭉치에서는 어근과 재결합해야 하나, 괄호에 의해 뒷말과 직접적인 결합이 불가능하다. ‘제-’는 역시 숫자와 같은 기호와 결합하여 복합어를 구성하는 것이 불가능하다. 이러한 경우 어근과 통합하는 접사 처리 과정 없이 체언을 대상으로 어휘의미를 분석하였다.

이 밖에 접사 처리(통합) 과정에서 유의할 점은 어근의 품사가 자동으로 승계되지 않는 경우가 존재한다는 것이다. 접두사는 일반적으로 어근의 품사를 바꾸지 않으나 다음



과 같이 어근에 체언접두사가 결합하여 부사가 되는 경우가 있다. 따라서 접사 통합형의 품사를 결정할 때에는 반드시 문장에서의 쓰임을 고려해야 한다.

(4) 체언접두사 결합형에서 뒷말의 품사가 자동으로 승계되지 않는 경우

가. 대/XPN+ 부분/NNG → 대부분/MAG

예. 그의 말은 대부분 거짓말이었다.

가'. 대/XPN+ 부분/NNG → 대부분/NNG

예. 그는 수입의 대부분을 저축한다.

나. 무/XPN+ 작정/NNG → 무작정/MAG

예. 계획도 없이 무작정 발 가는 대로 서울로 올라왔다.

나'. 무/XPN+ 작정/NNG → 무작정/NNG

예. 무작정으로 집을 나온 것이었다.

다. 무/XPN+ 조건/NNG → 무조건/MAG

예. 그는 이유를 듣지도 않고 무조건 화부터 냈다.

다'. 무/XPN+ 조건/NNG → 무조건/NNG

예. 무조건에 가까운 거부 반응을 일으키곤 했다.

라. 한/XPN+ 바탕/NNG → 한바탕/MAG

예. 이웃집에서 삼대독자가 없어졌다고 한바탕 난리가 났다.

라'. 한/XPN+ 바탕/NNG → 한바탕/NNG

예. 한바탕의 곡성이 울렸다.

명사파생접미사를 앞말과 통합시킬 때에도 결합형의 품사가 앞말의 품사를 승계하지 않는 경우가 있다. 접사 처리(통합) 과정에서 일반명사 또는 의존명사가 명사파생접미사와 결합할 때 어근의 품사가 유지되는 것이 일반적이나, 일부 복합어에서 품사가 바뀌는 경우가 일부 나타났다.

(5) 일반명사와 명사파생접미사 결합형에서 앞말의 품사가 승계되지 않는 경우

가. 비교/NNG+ 적/XSN → 비교적/MAG

예. 그의 방은 비교적 깨끗했다.

가'. 비교/NNG+ 적/XSN → 비교적/NNG

예. 비교적인 관점

나. 사실/NNG+ 상/XSN → 사실상/MAG

예. 이번 계획은 사실상 실패로 돌아갔다.

나. 사실/NNG+ 상/XSN → 사실상/NNG

예. 사실상의 주인이나 다름없었다.

다. 송두리/NNG+ 썩/XSN → 송두리썩/MAG

예. 재산을 송두리썩 다 써 버리다.

(6) 의존명사와 명사과생접미사 결합형에서 앞말의 품사가 승계되지 않는 경우

가. 내/NNB+ 적/XSN → 내적/NNG

예. 경기 내적인 이유는 재경기 사유가 될 수 없다.

나. 외/NNB+ 적/XSN → 외적/NNG

예. 작품 외적인 일이 그의 예술을 가릴 순 없다.

이와 달리 고유명사가 명사과생접미사와 결합할 때에는 어근의 품사가 승계되지 않는 것이 일반적이어서 결합형은 대부분 일반명사로 나타났다.

(7) 고유명사와 명사과생접미사 결합형에서 앞말의 품사가 승계되지 않는 경우

가. 중국/NNP+ 산/XSN → 중국산/NNG

예. 중국산 배추를 추가 수입하였다.

나. 동해/NNP+ 상/XSN → 동해상/NNG

예. 동해상에서 오징어잡이 어선 ‘천왕호’를 탔다.

다. 유럽/NNP+ 용/XSN → 유럽용/NNG

예. 미국이나 유럽용 광고에서 연예인을 쓴 적이 없다.

라. 한국/NNP+ 적/XSN → 한국적/NNG

예. 전주는 가장 한국적인 도시이다.

마. 한국/NNP+ 형/XSN → 한국형/NNG

예. 한국형 원자로의 탄생은 전적으로 여러분 손에 달렸다.

바. 중국/NNP+ 화/XSN → 중국화/NNG

예. 카자흐스탄의 급속한 중국화는 불 보듯 뻔하다.

이러한 점에 유의하여 “형태 분석 말뭉치 구축(2019년)” 결과를 어휘 분석용 말뭉치 수준으로 변환하였다. <표 10>은 접사 처리(통합) 과정을 거쳐 만들어진 어휘 분석용 말뭉치의 일부이다.

형태 분석 말뭉치	전문가들은	전문가/NNG+ 들/XSN+ 은/JX
	노조가	노조/NNG+ 가/JKS
	정치화되고	정치/NNG+ 화/XSN+ 되/XSV+ 고/EC
	정치권이	정치/NNG+ 권/XSN+ 이/JKS
	노동계	노동/NNG+ 계/XSN
	편향적으로	편향/NNG+ 적/XSN+ 으로/JKB
	가고	가/VV+ 고/EC
	있어	있/VX+ 어/EC
	우려스럽다고	우려/NNG+ 슥/XSA+ 다고/EC
	덧붙였다.	덧붙이/VV+ 었/EP+ 다/EF+ ./SF



어휘 분석용 말뭉치	전문가들은	전문가/NNG+ 들/XSN+ 은/JX
	노조가	노조/NNG+ 가/JKS
	정치화되고	<b>정치화/NNG+</b> 되/XSV+ 고/EC
	정치권이	<b>정치권/NNG+</b> 이/JKS
	노동계	<b>노동계/NNG</b>
	편향적으로	<b>편향적/NNG+</b> 으로/JKB
	가고	가/VV+ 고/EC
	있어	있/VX+ 어/EC
	우려스럽다고	우려/NNG+ 슥/XSA+ 다고/EC
	덧붙였다.	덧붙이/VV+ 었/EP+ 다/EF+ ./SF

<표 10> 어휘 분석용 말뭉치 변환의 예

형태 분석 말뭉치 구축 지침에 의거하여 형태 분석 말뭉치는 접사 ‘들/XSN, 화/XSN, 되/XSV, 권/XSN, 계/XSN, 적/XSN, 슥/XSA’을 분석하였다. 그러나 접사 처리(통합)

과정을 통해 통사적 접사 ‘들/XSN’과 용언화 접미사 ‘되/XSV, 스럽/XSA’를 제외한 ‘화/XSN, 권/XSN, 계/XSN, 적/XSN’을 앞말과 결합하여 형태 분석 말뭉치를 어휘 분석용 말뭉치로 변환하였다.

## 2.2. 분석 도구(워크벤치) 구현

어휘의미 분석 말뭉치 구축 과정에서 어휘의미 분석 오류를 효율적으로 수정하고 작업 진도를 모니터링할 수 있는 분석 도구가 필요하다. 본 사업에서는 어휘의미 분석 오류를 효율적으로 수정하고 작업 진도를 관리하기 위해 웹 기반의 워크벤치를 구축, 활용하였다. 웹 기반의 워크벤치는 작업자들의 동시 접속과 다중 작업 수행을 돕고, 작업자에게 균등한 분량의 작업을 배분해 준다. 자신에게 배분된 어휘만을 불러오기 때문에 웹상의 작업 처리 속도를 일정하게 유지할 수 있으며 분석 결과를 취합하는 과정에서 발생하는 오류를 차단할 수 있다. <우리말샘>과 연동하여 사전을 쉽게 참조할 수 있게 하는 등 효율적인 수정 작업을 지원하였다.

<그림 2>는 워크벤치 첫 화면이다. 작업자 등 모든 사용자는 워크벤치 홈 화면을 통하여 최신 지침과 작업 시 유의 사항을 확인할 수 있다.



<그림 2> 워크벤치 홈 화면

<그림 3>의 작업 현황을 통해 모든 워크벤치 사용자는 전체 작업 현황과 자신의 작업 현황 등을 확인할 수 있다. 좌측 원그래프에 제시된 전체 작업에 대한 검수 비율을 통해

운영자는 향후 일정 등을 계획할 수 있고, 우측 중간에 있는 원그래프를 통해 완료된 작업에 대한 검수 비율을 확인하면서 진행률이 낮은 부분에 필요한 조치를 취할 수 있다.

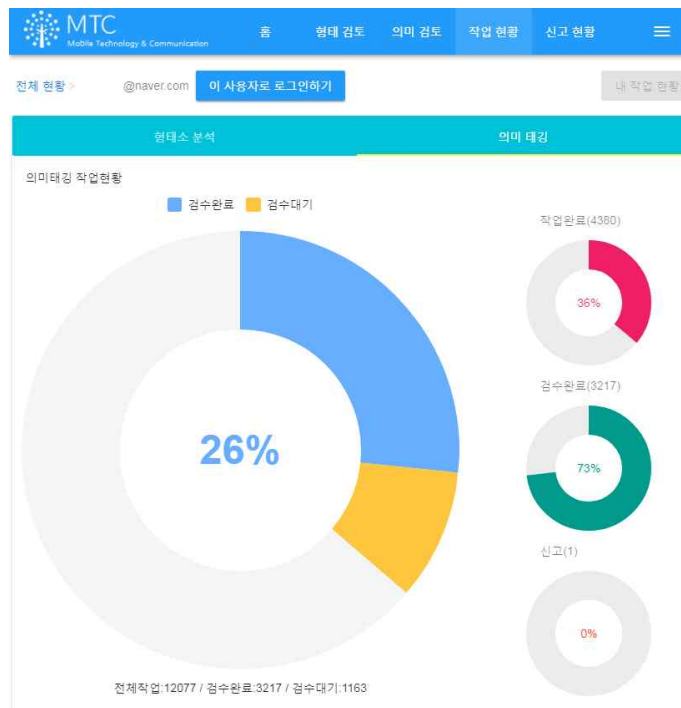


<그림 3> 워크벤치의 전체 작업 현황 화면

또한 운영자는 작업 현황 화면 하단에서 개별 사용자의 아이디를 확인할 수 있고, 해당 아이디를 선택하여 개별 사용자의 작업 현황 및 검수 현황을 살펴볼 수 있다. 전체 작업 현황과 동일하게 운영자는 해당 작업자 개인의 작업 완료 상태와 검수 비율 등을 확인함으로써 작업을 독려할 수 있다. <그림 4>는 워크벤치 개별 사용자를 보여 주고, <그림 5>는 개별 사용자의 작업 현황을 보여 준다.

MTC Mobile Technology & Communication			
홈	형태 검토	의미 검토	작업 현황
작업자 목록		작업자 검색	
userID	이름	소속	역할
@snu.ac.kr	조	서울대	작업자
@snu.ac.kr	황	서울대	작업자
@korea.ac.kr	홍	고려대	검수자
@naver.com	홍	고려대	작업자
i@gmail.com	최	성신여대	검수자

<그림 4> 워크벤치의 사용자 목록 화면



<그림 5> 워크벤치의 개별 사용자 작업 현황 화면

워크벤치는 사용자의 역할에 따라 최적화된 기능을 활용할 수 있도록 구성되었다. 따라서 작업자와 검수자, 운영자는 자신의 역할에 따라 활용할 수 있는 기능이 서로 다르다. 우선 작업자는 자신이 의미 분석해야 할 어휘의 형태와 해당 형태의 작업 수를 <그

림 6>과 같은 화면을 통하여 확인할 수 있다. 또한 작업자는 작업 선택 화면을 통해 대기, 진행, 완료 등 자신의 어휘별 작업 상태를 확인할 수 있고, '명사 검색' 등을 통해 작업 어휘를 선택할 수 있다. '명사, 품사, 총 작업 수, 제출 작업 수, 작업상태' 등 항목별로 정렬 기능을 제공하여 작업 어휘 선택의 편의성을 높였다.

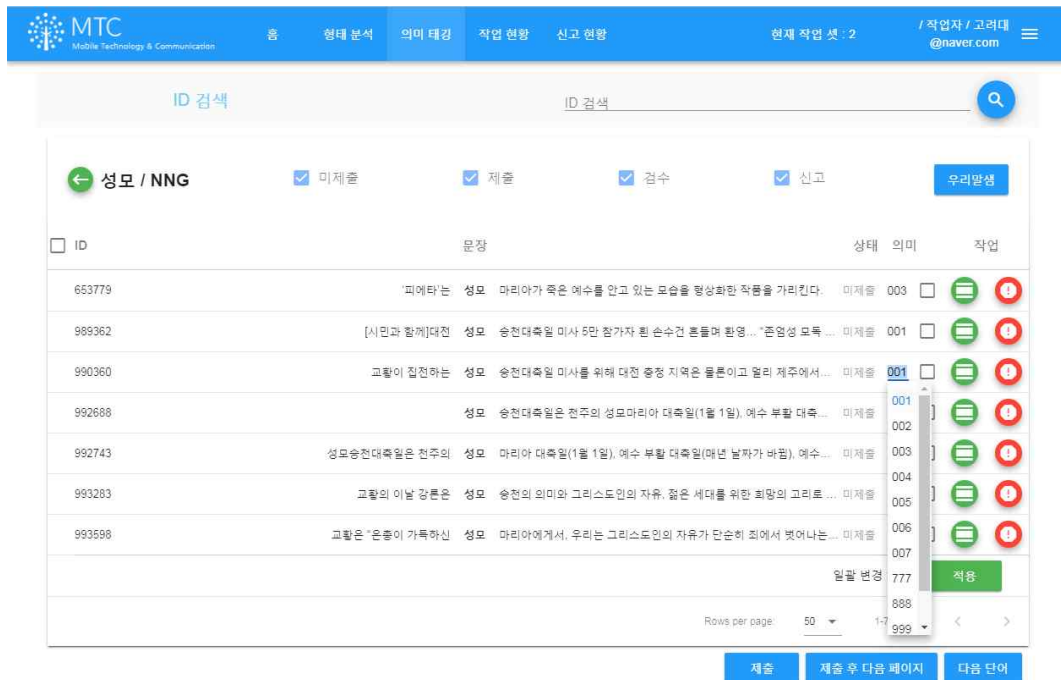
명사	품사	↓총 작업 수	제출 작업 수	작업상태
관계자	NNG	1590	1590	완료
조	NR	991	0	대기
인사	NNG	833	0	대기
운량	NNG	484	0	대기
집	NNG	437	0	대기
경기	NNP	341	341	완료
보고	NNG	319	2	진행
갈등	NNG	284	284	완료
이익	NNG	249	0	대기
개국	NNB	223	223	완료
명어	NNP	214	0	대기
지구	NNG	198	0	대기
일기	NNG	168	0	대기

<그림 6> 작업자의 작업 선택 화면

작업자가 작업 선택 화면에서 작업 어휘를 선택하면 <그림 7>과 같은 작업 화면으로 연결된다. 작업 화면에는 동일한 형태의 작업 어휘가 포함된 문장이 제시되어 있고 의미 부분에 자동 분석 결과를 디폴트값으로 제시하여 작업자가 어휘의미 분석기의 분석 결과를 참고할 수 있도록 하였다. 작업자는 작업 화면에 주어진 앞뒤 문맥과 자동 분석 결과를 바탕으로 작업 어휘의 의미를 분석하고 의미 번호를 부여하게 된다.

분석 결과는 드롭다운 선택 방식으로 입력하도록 제한함으로써 분석 결과 수정 시의 도치 않은 형식 오류가 발생하는 것을 원천적으로 차단했다. 드롭다운 메뉴의 선택항역시 <우리말샘>에 등재된 의미 번호와 함께 지침에 의해 규정된 어휘의미 표지 777,

888, 999과 NA로 한정하였다. 예를 들어 <그림 7>의 ‘성모’는 <우리말샘>에 001부터 007까지 등재되어 있기 때문에 드롭다운 방식을 통해 001~007과 777, 888, 999, NA 만 선택 가능하다. 이처럼 불필요한 의미 번호를 제거함으로써 오류의 발생 가능성을 최소화하였다. 여기서 ‘NA’는 형태 분석 오류로 인해 의미 번호를 부여할 수 없을 때 임시로 사용하는 표지로, 해당 표지가 붙은 어휘들은 형태 분석을 수정한 후 다시 의미 번호를 부여할 수 있도록 작업자에게 재분배하였다.



<그림 7> 작업자의 작업 화면

어휘의미 분석 시 작업자는 워크벤치에서 제공하는 다양한 기능을 활용할 수 있다. 우선 우측 상단에 있는 ‘우리말샘’을 통해 <우리말샘>에 접속하여 해당 어휘의 사전적 의미를 확인할 수 있다. 그리고 화면에 주어진 앞뒤 문맥만으로 의미 분석이 불가능할 경우 ‘맥락보기’를 통해 더 많은 문맥을 확인할 수 있다. 작업 중 문제가 발생할 경우 ‘신고하기’를 통해 이슈를 생성하고 운영자 그룹과 소통할 수 있다.



새로 고침 명사 선택 명사 검색

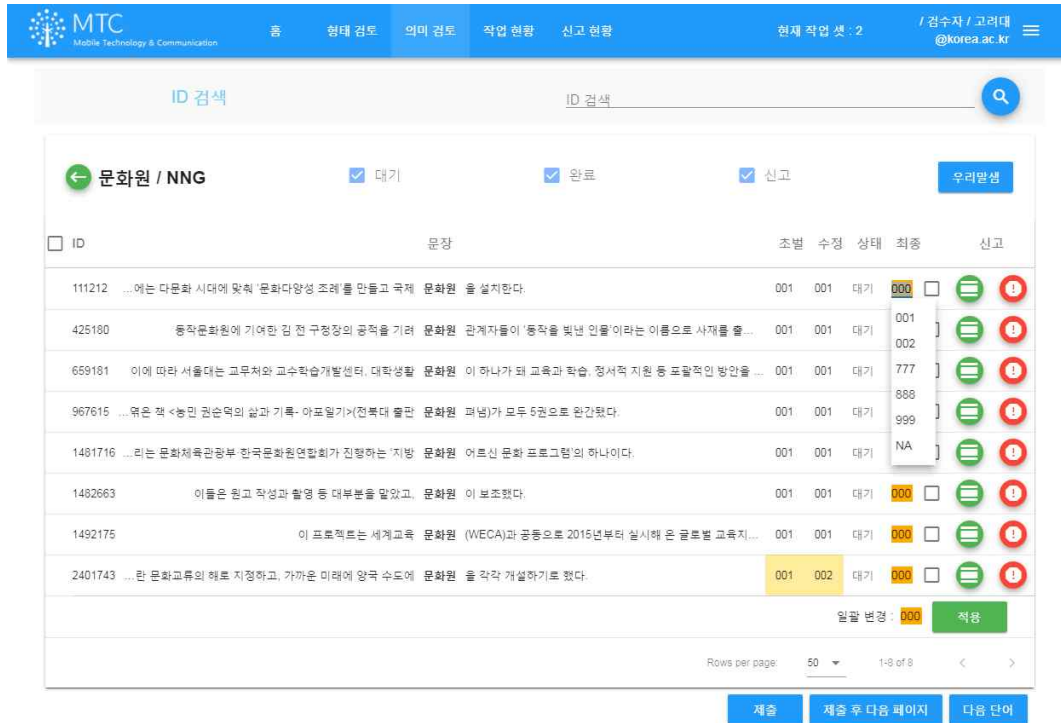
명사	품사	작업 수↑	검토 상태	수정A 할당	수정B 할당	수정A 상태	수정B 상태	수정 일치율
CMG제약	NNP	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
G6	NNP	1	대기	quss135@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
LS-Nikko동계현	NNP	1	대기	quss135@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
R사우나클럽	NNG	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
Swift	NNG	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
가구전	NNG	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
가래	NNP	1	작업미완료	quss135@naver.com	[단일 할당 작업]	대기	[단일 할당 작업]	[단일 할당 작업]
가리	NNP	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
가정용품	NNG	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]
가족장	NNG	1	작업미완료	quss135@naver.com	[단일 할당 작업]	대기	[단일 할당 작업]	[단일 할당 작업]
가조오	NNP	1	작업미완료	quss135@naver.com	[단일 할당 작업]	대기	[단일 할당 작업]	[단일 할당 작업]
가중	NNG	1	작업미완료	quss135@naver.com	[단일 할당 작업]	대기	[단일 할당 작업]	[단일 할당 작업]
가치세법상	NNG	1	완료	0120ejin@naver.com	[단일 할당 작업]	완료	[단일 할당 작업]	[단일 할당 작업]

<그림 8> 검수자의 검수 선택 화면

검수자는 <그림 8>과 같은 검수 선택 화면을 통해 자신에게 분배된 검수 어휘를 확인할 수 있고, ‘명사 검색’ 등을 통해 검수 어휘를 선택할 수 있다. ‘검토 상태’에 제시된 대기, 진행, 완료, 작업미완료를 통해 어휘별 검수 상태를 확인할 수 있으며, 작업자를 가리키는 수정A의 상태를 통해 어휘별 작업자의 작업 상태를 점검할 수 있다. 이를 통해 검수자는 작업자를 독려하고 작업 속도를 조절할 수 있다. 작업자의 작업 선택 화면과 마찬가지로 ‘명사, 품사, 작업 수, 검토 상태, 수정A 상태’ 등 항목별로 정렬 기능을 제공하여 검수 어휘 선택의 편의성을 높였다.

검수자가 검수 선택 화면에서 검수 어휘를 선택하면 <그림 9>와 같은 검수 화면이 나타난다. 검수 화면에는 동일한 형태의 검수 어휘가 포함된 문장이 제시되어 있다는 점에서 작업 화면과 유사하나, ‘초벌’ 의미와 ‘수정’ 의미를 함께 제공한다는 점에서 차이를 보인다. ‘초벌’은 어휘의 의미 분석기의 자동 분석 결과를 가리키고, ‘수정’은 작업자의 분석 결과를 나타낸다. 그리고 이 둘의 분석 결과가 다를 경우 검수자가 주의를 기울일

수 있도록 바탕색을 달리하여 표시하였다. 검수자는 검수 화면에 주어진 앞뒤 문맥과 자동 분석 결과 및 작업자의 분석 결과를 바탕으로 어휘의 의미 번호를 최종 선택한다.

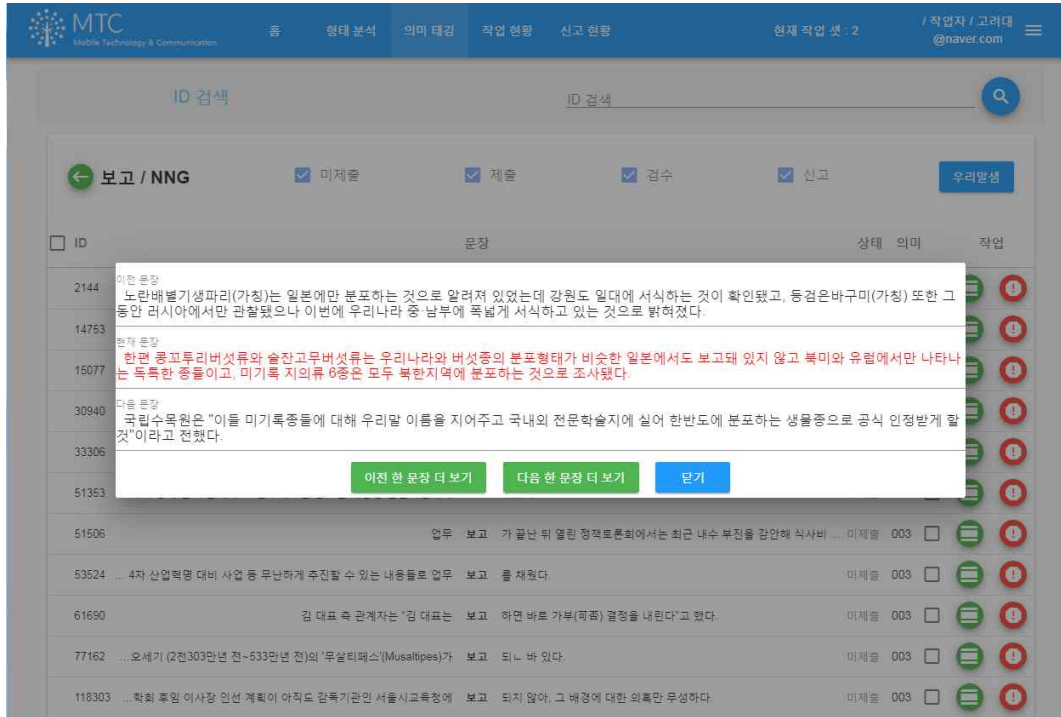


<그림 9> 검수자의 검수 선택 화면

작업자의 작업 화면과 마찬가지로 검수 화면에 다양한 기능을 제공하였다. 우선 검수 결과를 드롭다운 선택 방식으로 입력하도록 제한하였고, 드롭다운 메뉴의 선택항 역시 <우리말샘>을 바탕으로 한정하였다. <그림 9>의 '문화원'의 경우 <우리말샘>에 두 가지 의미로 등재되어 있기 때문에 의미 번호 '001', '002'와 어휘의미 표시 '777, 888, 999'와 형태 분석 오류를 의미하는 'NA'만 선택 가능하도록 구성되었다. 이 역시 검수 하는 과정에서 의도치 않은 형식 오류의 발생을 원천적으로 차단하기 위한 것이다.

이 밖에 검수자는 우측 상단에 있는 '우리말샘'을 통해 <우리말샘>에 접속하여 해당 어휘의 사전적 의미를 확인할 수 있고, '맥락보기'를 통해 화면에 주어진 문장 이상의 문 맥을 확인할 수 있다. 검수 중 문제가 발생할 경우 '신고하기'를 통해 이슈를 생성하고

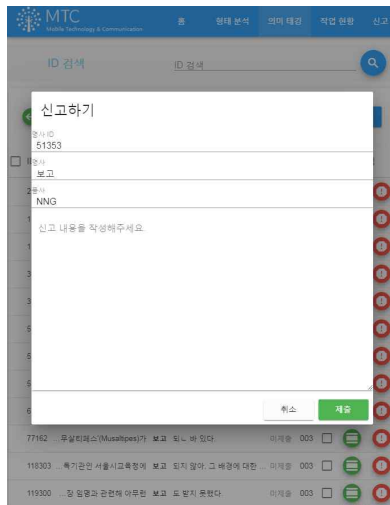
운영자 그룹과 소통할 수 있다.



<그림 10> ‘맥락보기’ 화면

<그림 10>은 ‘맥락보기’를 선택했을 때 나타나는 화면이다. 작업 또는 검수 과정에서 분석 대상 어휘가 포함된 문장 이상의 맥락을 살펴야 할 때 작업자와 검수자는 ‘맥락보기’를 통하여 선택 어휘가 포함된 문장의 앞뒤 문장을 확인할 수 있다. 그럼에도 불구하고 어휘의미 분석이 불가능할 경우 ‘이전 한 문장 더 보기’ 또는 ‘다음 한 문장 더 보기’를 통하여 맥락을 추가적으로 확인할 수 있다.

어휘의미 분석 및 검수 과정에서 발생하는 이슈를 운영자 그룹과 공유하기 위해 느낌표 모양의 ‘신고하기’ 기능을 제공한다. <그림 11>은 ‘신고하기’를 선택했을 때 나타나는 화면이다.



<그림 11> ‘신고하기’ 화면



<그림 12> 워크bench의 신고 현황 화면

작업자와 검수자는 ‘신고하기’를 통해 어휘의미 분석의 이슈뿐만 아니라 형태 오분석

을 신고할 수 있다. 이는 “형태 분석 말뭉치 구축(2019년)”을 통해 만들어진 형태 분석 말뭉치의 오류율을 낮추는 역할을 한다. 운영자는 <그림 12>와 같은 신고 현황 화면을 통해 어휘의미 분석 및 검수 과정에서 생성된 이슈를 파악하고 피드백을 제공한다. 운영자는 신고 현황에서 확인된 내용을 바탕으로 어휘의미 분석 말뭉치 구축 지침을 지속적으로 수정하고 보완하였다.

### 2.3. 작업 교육

본격적으로 어휘의미 분석 말뭉치를 구축하기에 앞서 어휘의미 분석 말뭉치 구축 지침과 분석 도구인 워크벤치 사용에 대한 교육이 이루어져야 한다. 아울러 비밀 유지와 자료 보안, 문서 보안 등과 관련한 보안 교육도 필요하다. 이에 사업 전체 참여자를 대상으로 보안 교육 및 형태 분석 말뭉치 구축 지침, 워크벤치 사용에 대한 교육을 실시하였다.

특히 어휘의미 분석 말뭉치 구축 지침에 대한 교육은 1회의 교육만으로는 불충분하며 수시로 개별 작업자의 작업 수행에 대한 피드백이 이루어져야 한다. 본 사업은 작업자가 어휘의미 분석 작업을 수행한 후 어휘의미 분석 결과를 검수자가 검토하는 방식으로 이루어졌다. 그 과정에서 개별 작업자가 빈번히 만들어 내는 오류를 검수자가 파악하여 지속적으로 피드백하였다. 또한 전체 분석 결과의 오류와 일관성을 검토하는 최종 검수 단계에서 작업자들이 공통적으로 만들어 내는 오류를 공유하여 지침 및 교육 자료를 보완하였다.

### 2.4. 자동 분석

어휘의미 분석 말뭉치 구축 과정에서 작업자는 자동 어휘의미 분석 결과를 토대로 오류를 수정하게 된다. 따라서 높은 분석 정확률을 갖춘 자동 어휘의미 분석 도구를 사용하는 것이 작업의 효율을 높이기 위한 관건이 된다. 이에 본 사업에서는 지도학습 기반의 고성능 어휘의미 분석기를 사용해 초벌 분석 결과를 산출하고 이를 수정 단계로 전달하였다.

자동 분석에 사용한 서울대 어휘의미 분석기는 <우리말샘>의 단어 명사 346,047개와 단어 명사 센스 185,087개를 대상으로 용례 9만 문장, 88만 어절 규모의 학습 데이

터를 구축하고, 이를 LSTM 기반의 신경망 모델을 통해 학습한 고성능 분석기다. <우리말샘>의 의미 식별번호에 대응하는 방식으로 어휘의미를 분석하며 작업 초기 실사용 테스트에서 87.05%의 정확도를 보였다.

자동 분석기는 별도의 API를 통해 워크벤치와 결합되었다. 작업자나 검수자가 워크벤치를 통해 작업 화면에 진입하면 자동으로 작업대상 어휘가 자동분석기로 전송되며 이후 분석기에 의한 초벌 분석 결과가 거의 실시간으로 작업 화면에 전달된다. 모든 데이터를 한번에 분석해 놓고 수정 작업을 진행하는 것이 아닌, 실시간으로 분석을 받아오는 방식을 택한 이유는 분석기의 성능이 지속적으로 개선되었기 때문이다. 정해진 스케줄에 따라 중간 작업 결과물을 학습 자료로 사용하여 자동 분석기의 성능을 향상시켰고 이를 통해 작업자와 검수자의 분석 작업을 도왔다.

## 2.5. 분석 결과 수정

어휘의미 분석기의 자동 분석 결과는 웹 기반 워크벤치의 작업자 화면(<그림 7>)에서 ‘의미’ 부분에 기본값으로 설정되어 있으며, 작업자는 자동 분석 결과를 수정하는 방식으로 어휘의미 분석 말뭉치 구축 작업을 진행하였다.

어휘의미에 대한 분석 결과 수정은 다음과 같이 3단계로 이루어졌다.

### ○ 1단계: 작업자의 자동 분석 결과 수정

어휘의미 분석 대상 어휘에 대한 자동 분석 결과가 작업 화면에서 의미의 기본값으로 제시된다. 이를 토대로 작업자는 어휘의미에 대한 자동 분석 결과의 수정 작업을 진행한다.

### ○ 2단계: 분석기의 자동 분석 결과와 작업자의 수정 결과에 대한 검수자의 검수

검수 화면에서 분석기의 자동 분석 결과는 ‘초벌 의미’로, 작업자의 수정 결과는 ‘수정 의미’로 제시된다. 검수자는 작업자의 수정 결과를 검토하면서, 자동 분석 결과와 불일치하는 경우에 주의하여 어휘의미 분석 결과를 수정한다.

### ○ 3단계: 전체 작업 결과물에 대한 최종 검수

어휘의미 분석 결과 검수가 끝나면, 공동연구원을 중심으로 한 상위 그룹이 전체 어휘의미 분석 과정에서 신고된 형태 분석의 오류를 검토하고 수정한다. 수정 사항이 없을 경우 작업자와 검수자에게 의미 번호 확정을 요청하고, 형태를 재분석

할 경우 추가된 어휘 항목을 재분배하여 1단계와 2단계의 분석 결과 수정 절차를 진행한다.

본 사업에서는 위와 같은 자동 분석과 3단계의 분석 결과의 수정 공정을 하나의 회기로 묶어 총 5회기를 진행하였다.

	분석 말뭉치 구축량	
	회기당	누적
문어 시범	문어: 10만	문어: 10만
문어 1차	문어: 90만	문어: 100만
문어 2차	문어: 100만	문어: 200만
구어 시범	구어: 10만	문어: 200만 구어: 10만
구어	구어: 90만	문어: 200만 구어: 100만

<표 11> 회기별 어휘의미 말뭉치 구축량

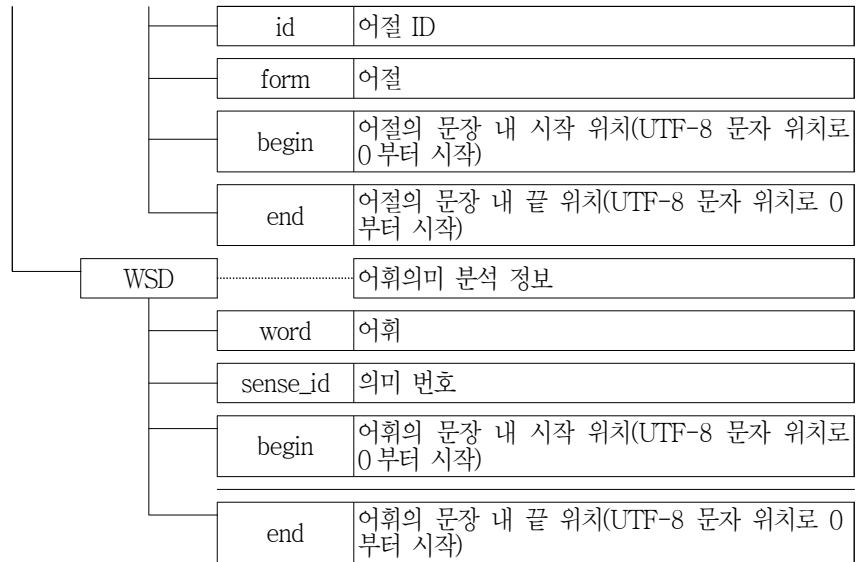
## 2.6. 최종 결과물 산출

어휘의미 분석 말뭉치 구축의 최종 단계는 어휘의미를 분석하고 오류를 수정한 결과를 JSON 형식으로 변환하여 최종 결과물을 산출하는 것이다.

<표 11>은 JSON 형식의 기본 구조이고, <표 12>는 JSON 형식으로 전환한 최종 결과물의 일부이다.

1 수준	2 수준	3 수준	4 수준	설명
	id			파일 ID
	metadata			파일의 메타 정보
		title		파일 제목
		author		작성자, 게시자
		publisher		출판사, 신문사
		year		출판년도
		note		부가 설명(샘플링 방식 등 기타 정보)
	document			문서 정보
		id		문서 ID
		metadata		문서의 메타 정보
			title	문서 제목
			author	작성자, 게시자
			publisher	출판사, 신문사
			url	URL 주소
			date	작성일시, 게시일시
			category	분류
			annotation_level	분석 층위
			note	부가 설명(구어 사용 맥락 정보, 샘플링 방식 등 기타 정보)
		sentence		문장
			id	문장 ID
			form	문장 정보
			word	어절 정보





<표 12> JSON 형식의 기본 구조

```

{
  "id": "WRAK00031",
  "metadata": {
    "title": "문어 200만 배포용",
    "author": "홍길동",
    "publisher": "국립국어원",
    "year": "2019",
    "note": "부분 추출 - 임의추출"
  },
  "document": [
    {
      "id": "WRAK00031-1",
      "metadata": {
        "title": "설 고속도로 언제 막히나",
        "author": "김철수",
        "publisher": "국어일보",
        "url": "https://www.korean.go.kr/",
        "date": "20190125",
        "category": "사회",
        "annotation_level": [ "어휘의미 분석" ],
        "note": "문서 부가 설명"
      },
      "sentence": [
        {
          "id": "WRAK00031-1.1",
          "form": "짧은 귀성 연휴, 작년보다 정체 심할듯",
          "word": [
            { "id": 1, "form": "짧은", "begin": 0, "end": 2 },
            { "id": 2, "form": "귀성", "begin": 3, "end": 5 },
            { "id": 3, "form": "연휴,", "begin": 6, "end": 9 },
            { "id": 4, "form": "작년보다", "begin": 10, "end": 14 },

```

```
{ "id": 5, "form": "정체", "begin": 15, "end": 17 },
  { "id": 6, "form": "심할듯", "begin": 18, "end": 21 }
],
"WSD": [
  { "word": "귀성", "sense_id": 4, "begin": 3, "end": 5 },
  { "word": "연휴", "sense_id": 1, "begin": 6, "end": 8 },
  { "word": "작년", "sense_id": 1, "begin": 10, "end": 12 },
  { "word": "정체", "sense_id": 6, "begin": 15, "end": 17 },
  { "word": "듯", "sense_id": 4, "begin": 21, "end": 22 }
]
}
]
}
]
```

<표 13> JSON 형식의 예시

## 제3장 어휘의미 분석 말뭉치 구축 지침 수립

### 1. 지침 수립 과정

지침 수립 과정은 <21세기 세종계획> 형태의미 분석 말뭉치 구축 지침의 미비점을 검토하여 이를 보완하고, 어휘의미 말뭉치 구축 과정에서 나타나는 실례를 추가하는 단계로 나뉜다.

<21세기 세종계획>의 어휘의미 분석 말뭉치 구축 지침은 국내의 다른 어휘의미 분석 말뭉치 구축 과정에서도 활용되고 있을 만큼 표준적인 지위를 확보하고 있다. 그러나 <표준국어대사전>의 의미 체계를 바탕으로 ‘동형이의어’ 수준의 의미 분석을 하기 때문에 <우리말샘>의 의미 체계를 바탕으로 ‘다의’ 수준을 고려한 어휘의미 분석 말뭉치를 구축하는 데에 그대로 적용하기에는 어려움이 있다.

이에 본 사업에서는 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 기반으로 하되, 일부 내용을 수정하고 항목별 설명을 상세화하며 다양한 사례를 포함함으로써 실제 구축 과정에서 도출되는 오류를 줄이고자 하였다.

#### 1.1. <21세기 세종계획> 형태 분석 말뭉치 구축 지침 검토

본 사업에서 어휘의미 분석 말뭉치를 구축하는 데 있어 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 그대로 적용하기에는 의미 분석의 수준과 분석 대상 품사의 차이 등의 문제가 발생한다. 따라서 <21세기 세종계획>의 형태의미 분석 말뭉치에는 보완이 필요한 부분이 포함되어 있는데, 주요 보완 사항은 다음과 같다.

- 의미 분석의 수준에 따른 사전 미등재 유형의 구분
- 분석 대상 품사의 차이에 따른 세부 지침의 부재
- 불필요한 규정과 모호한 지침 규정

### 1.1.1. 의미 분석의 수준에 따른 사전 미등재 유형의 구분

<21세기 세종계획> 형태의미 분석 말뭉치 구축 지침을 보면, 아래와 같은 내용이 제시되어 있음을 확인할 수 있다. 이를 통해 <21세기 세종계획>에서 구축한 형태의미 분석 말뭉치가 <표준국어대사전>의 의미 체계에 따라 동형이의어(원문에서는 ‘동음이의어’로 기술되어 있음.)를 구별하는 수준에서 의미 분석이 이루어졌음을 확인할 수 있다. 그 결과 사전에 형태가 등재되어 있지 않은 어휘에 부여하는 어휘의미 표지만 설정되어 있다.

#### 1. 기본 원칙

나. 동음이의어 분석 말뭉치의 구축 시, 형태분석 말뭉치를 어휘 분석 말뭉치 수준으로 변환하여 구축할 수 있다. 형태분석 말뭉치를 대상으로 품사 내 동음이의어 중의성을 해결한 말뭉치를 형태의미분석 말뭉치라 한다면, 접사 처리 과정을 거쳐 만들어진 어휘 분석 말뭉치를 대상으로 동음이의어 중의성을 해결한 말뭉치는 단어 의미 분석 말뭉치로 부를 수 있다. 당해 연도에 구축한 형태의미분석 말뭉치는 이 중 형태의미분석 말뭉치에 속한다.

#### 2. 세부 사항

나. 동음이의어 목록은 국립국어원의 <표준국어대사전>을 기준으로 한다.

마. 사전에 등재되지 않은 어휘가 나타날 경우, 어깨번호는 ‘88’으로 한다.

예) 12일자 '길'난에 보도된 '교도소 가는 [부정]의 : 부정\_88/NNG + '/SS + 의 /JKG] 인태군 씨에게 전해 달라며 박찬수...

반의어인 ‘모정’(母情)이라는 단어는 사전에 존재하지만 위의 ‘부정’(父情)이라는 단어는 사전에 등재된 어휘가 아니다. 이 경우에는 어깨번호를 ‘88’로 붙이고 추후 해결 방안을 모색한다.

### 1.1.2. 분석 대상 품사의 차이에 따른 세부 지침의 부재

<21세기 세종계획> 형태의미 분석 말뭉치는 일반명사, 의존명사, 동사, 형용사, 관형사, 일반부사에 해당하는 어휘에 대해 의미를 분석하고 해당 의미 번호를 부여하였다. 분석 대상으로 설정된 품사에 대해서도 분석 지침이 세밀하지 못한 채 <표준국어대사전>의 나타난 표제어 번호를 부여하는 것으로 기술되어 있다. 이상의 내용을 <21세기 세종계획> 형태의미 분석 말뭉치 구축 지침에서 확인할 수 있다.

## 2. 세부 사항

가. 분석 대상 품사는 기능어(function words)를 제외한 내용어(content words)를 중심으로 선정하였다. 대상 범주는 다음과 같다. : 일반명사(NNG), 의존명사(NNB), 동사(VV), 형용사(VA), 관형사(MM), 일반부사(MAG)

다. 말뭉치 정보 부착 방법은 해당 어절의 분석 결과에 ‘\_’과 두자리 어깨번호를 부착하는 방식을 사용한다. 어깨번호는 국립국어원 <표준국어대사전>에 나타난 표제어 번호와 일치한다.

예) 5BT\_B4920255840 비비/MAG 결의했다. 천(千)의 다리, 천의 팔이 [비비 : 비비\_02/MAG] 꼬인 이 매듭을. 재갈 물린

index

대상형태소

앞 5어절

어절

분석

뒤 5어절

### 1.1.3. 불필요한 규정과 모호한 지침 규정

<21세기 세종계획> 형태의미 분석 말뭉치 구축 지침의 ‘1. 기본 원칙’을 보면, “접사 처리 과정을 거쳐 만들어진 어휘 분석 말뭉치를 대상으로”라는 표현이 나타나지만 ‘접사 처리 과정’에 대한 설명이 부재하다. 그리고 형태분석 말뭉치의 분석 오류를 위한 어깨번호 ‘99’를 설정하고 있지만, 형태 재분석을 통해 실제 형태의미 분석 말뭉치에는 해당 기호가 나타나지 않기 때문에 지침에서 해당 표지를 규정할 필요가 없다. 또한 중의성과 관련하여 ‘@’을 지정하였으나 이는 분석 도구의 한계로 인한 것이지 말뭉치 자체의 문제가 아니기 때문에 ‘99’와 마찬가지로 별도의 규정으로 존재할 필요가 없다.

## 1. 기본 원칙

나. 동음이의어 분석 말뭉치의 구축 시, 형태분석 말뭉치를 어휘 분석 말뭉치 수준으로 변환하여 구축할 수 있다. 형태분석 말뭉치를 대상으로 품사 내 동음이의어 중의성을 해결한 말뭉치를 형태이미분석 말뭉치라 한다면, 접사 처리 과정을 거쳐 만들어진 어휘 분석 말뭉치를 대상으로 동음이의어 중의성을 해결한 말뭉치는 단어 의미 분석 말뭉치로 부를 수 있다. 당해 연도에 구축한 형태이미분석 말뭉치는 이 중 형태이미분석 말뭉치에 속한다.

## 2. 세부 사항

바. 형태분석 말뭉치의 분석 오류가 발견되면, 어깨번호를 '99'로 하고 이후 후처리 단계에서 형태분석 오류를 수정한다. 형태분석 오류 수정은 당해연도 형태분석 지침에 따른다.

사. 좌우 5어절의 환경만으로는 동음이의어 중의성 해결이 어려운 경우, 또는 기타 오류로 인해 추후 처리가 필요한 경우에는, 용례의 맨 앞에 @를 붙이고, 최종적인 말뭉치 수정 단계에서 이를 찾아 해결한다.

### 1.2. 지침의 보완 방향

앞에서 언급한 바와 같이 본 사업은 '다의(多意, sense)'를 구별하는 수준의 어휘의미 분석 말뭉치를 구축하는 것을 목표로 한다. 따라서 '동형이의어'를 구별하기 위한 <21세기 세종계획>의 형태이미 분석 말뭉치 구축 지침을 그대로 적용하기에는 문제가 있다. 여기에서는 1.1.에서 언급한 보완 사항을 중심으로 본 사업팀에서 마련한 세부 규정을 언급하고자 한다.

### 1.2.1. 사전 미등재 유형의 구분을 위한 어휘의미 표지 신설

본 사업에서 구축하는 어휘의미 분석 말뭉치는 <우리말샘>의 의미 체계를 따르기 때문에 다의어 수준의 의미 분석이 이루어진다. 그러나 기존 <21세기 세종계획>의 지침은 의미 미등재에 대한 고려가 불가능하기 때문에 형태 미등재어에만 ‘88’을 부여하였다. 그 결과 다음과 같은 어휘에 부여할 수 있는 어휘의미 표지가 존재하지 않았다.

#### (8) <우리말샘>의 의미 미등재어 1

- 가. 집을 구하려면 부동산을 찾아가 보는 것이 편리하다.
- 나. 국가 대표 상비군들이 출전한다.
- 다. 형태소 분석기를 돌려서 결과가 나오면
- 라. 인간이 서로 존중하는 상호성의 원칙도 심어줘야 한다.
- 마. 창업 아이템이 나온다.
- 바. 맥스 미디어의 시대
- 사. 대부분이 여행사의 마라톤 패키지 상품으로 홍콩을 찾았다.
- 아. 이 사전연명의료의향서가 어떤 형식을 갖추고 있는지 준비했습니다.
- 자. 밀어내기 볼넷으로 5-6 역전을 허용하고 말았다.
- 차. 하나의 경제 벨트로

(8)에 제시된 어휘들은 <우리말샘>에 그 형태가 등재되어 있으나, <우리말샘>에 기술된 의미로 해석하기에는 무리가 있다. ‘부동산’의 경우 <우리말샘>에 표제어로 등재되어 있으나, 현재 사전상의 의미는 “움직여 옮길 수 없는 재산”, 즉 ‘동산’(動産)의 반대 의미를 뜻한다. 그런데 위 예문에 쓰인 ‘부동산’의 의미는 “공인 중개 사무소”이다. ‘상비군’ 역시 <우리말샘>에 “국가 비상사태에 항상 대비할 수 있도록 편성된 군대”라는 용어로 등재되어 있으나, 해당 의미로 (8나)를 해석하기에는 어려움이 존재한다. ‘상비군’이라는 군사 분야의 전문 용어가 다른 분야에 사용되면서 의미 변화가 발생했지만 <우리말샘>은 아직 그 의미를 수용하지 못한 것으로 판단된다.

그 외에도 공학 일반 분야의 전문 용어로 <우리말샘>에 등재된 ‘분석기’는 “음파나 전파 따위의 모양이나 주파수를 분석하는 데에 쓰는 기구”를 뜻한다. 그러나 해당 어휘가 다른 분야에 쓰이면서 주파수가 아닌 다른 대상을 분석하는 기구 정도를 의미하는



것으로 사용되었다. ‘상호성’ 역시 경제 분야에서 “상대국이 자기 나라를 어떻게 취급하느냐에 따라 수출입품의 제한, 관세, 기업 활동과 금융의 자유화 따위에 대한 결정이 달라지는 성질”을 뜻하는 말이나 (8라)처럼 일상생활에서 국가 간의 관계와 무관하게 사용되고 있다.

<우리말샘>에 “패션과 관계된 의류 및 액세서리류의 품목” 또는 “한 단위로 다루어지는 데이터의 집합”으로 등재된 ‘아이템’은 각각 공예와 정보·통신 분야에서 사용되는 전문 용어이나, (8마)처럼 해당 전문 분야와 무관하게 사용되곤 한다. ‘매스’ 또한 <우리말샘>에서는 “부피를 가진 하나의 덩어리로 느껴지는 물체나 인체의 부분”이라는 미술 분야의 전문 용어로 등재되어 있으나 (8바) ‘매스 미디어’의 ‘매스’를 풀이하기에는 무리가 있다.

<우리말샘>에 따르면 ‘패키지’는 영상 분야에 사용되는 전문 용어로 “시나리오 작가와 인기 배우, 인기 소설과 인기 배우 등의 결합”을 의미한다. 그러나 위 예문에서 “일괄 상품 혹은 상품의 결합” 정도의 의미로 파악되는 ‘패키지’는 <우리말샘>에 등재된 어휘와 의미적 관련성을 찾을 수도 있으나 서비스업이라는 다른 분야에서 사용되는 용어이기 때문에 해당 어휘의 의미가 등재되지 않은 것으로 판단할 수 있다. ‘의향서’ 역시 <우리말샘>에 “계약에 앞서 참여할 의사를 표시하는 서류”라는 경제 분야의 의미로 한정되어 있어 (8아)에 제시된 ‘의향서’는 의미 미등재어로 판단된다.

동일한 영역에서 사용됨에도 불구하고 <우리말샘>에 등재된 의미로 보기 어려운 것들이 존재한다. (8자)의 ‘밀어내기’의 경우 경제 분야에서 사용되는 “대기업과 중소기업 또는 본사와 대리점 따위의 관계에서 주도권을 가지고 있는 측이 주도권이 없는 측에게 상품이나 매출에 대한 책임 따위를 떠맡기는 일” 말고도, “씨름에서 상대방을 씨름판 밖으로 밀어 내어 점수를 따는 일”이라는 체육 분야의 용어로 등재되어 있다. 하지만 (8자)에 제시된 ‘밀어내기’는 야구 경기에서 사용되는 용어로 같은 체육 분야에서 사용되는 용어가 등재되어 있음에도 그 의미가 통하지 않기 때문에 의미 미등재어로 처리할 수밖에 없다.

‘벨트’는 위에서 언급한 것들과 다른 양상을 보인다. 다시 말해서 “옷의 허리 부분에 둘러매는 띠” 또는 “두 개의 바퀴에 걸어 동력을 전하는 띠 모양의 물건”을 뜻하는 일상 용어이지만 (8차)처럼 경제라는 특수한 영역에서 사용되면서 의미 변화를 겪은 것으로 판단되어 의미 미등재어로 처리하였다.

이상의 어휘들은 그 형태가 <우리말샘>에 특정 분야의 전문 용어로 등재되어 있으나

다른 영역에서 사용되는 과정에서 의미 변화를 겪은 것들이다. 반대로 일상 용어가 전문 분야에서 사용되는 과정에서 의미가 변한 것들이다. 그럼에도 불구하고 <우리말샘>에서는 변화된 의미를 반영하지 못하였다. 향후 이러한 의미의 등재 가능성을 고려하면 이들 어휘에 별도의 어휘의미 표지를 부여할 필요가 있다. 따라서 본 사업에서는 888을 부여하여 의미 미등재어를 명시하였다.

(8)과 같은 유형 외에도 의미 미등재어에는 (9)와 같은 경우가 있다. 이들은 대개 어종의 차이 또는 원어 표기의 측면에서 구별될 수 있으나 동일한 한글 표기가 <우리말샘>에 등재되었다는 점에서 의미 미등재어로 분석된 예들이다.

#### (9) <우리말샘>의 의미 미등재어 2

가. 코리아 드림을 꿈꾸다.

나. 단국 대학교 컴퓨터 공학

다. 금융감독원의 제재를 받았다.

‘드림’은 “매달아서 길게 늘이는 물건” 등의 의미로 <우리말샘>에 등재되어 있으나, “꿈”에 해당하는 의미가 등재되어 있지 않다. ‘단국’ 역시 “맛이 단 국물” 등으로 <우리말샘>에 기술되어 있으나 ‘단국대’의 의미는 나타나지 않는다. 어종은 다르지만 동일한 한글 형태가 <우리말샘>에 등재되었다는 점에서 본 사업에서는 이를 의미 미등재어로 분석하였다. 끝으로 “감독하는 직무를 맡은 사람”을 의미하는 ‘감독원(監督員)’이 <우리말샘>에 등재되어 있으나 “감독 기관”에 해당하는 의미가 제시되어 있지 않으므로 (9다)의 ‘감독원(監督院)’ 역시 의미 미등재어에 속한다.

우리가 사용하는 어휘 중에는 <우리말샘>에 형태가 등재되지 않은 것도 있으나 이처럼 의미가 기술되지 않은 것도 존재한다. 따라서 이 둘을 구분하여 어휘의미 표지를 부여하는 것은 정밀한 어휘의미 분석 말뭉치를 구축하는 과정에서 선행되어야 하는 작업이다. 이에 본 사업에서는 의미 미등재어에 어휘의미 표지 ‘888’을 부여했다. 이와 달리 말뭉치에 출현하는 형태 미등재어에 대해서는 ‘777’을 부여하여 형태와 의미 차원의 미등재를 구분하였다. 형태 미등재어의 예는 다음과 같다.

#### (10) <우리말샘>의 형태 미등재어 1

가. 나는 된장을 끓을 때, 홍고추도 넣고 청고추도 넣는다.

나. 전공에 관련된 텍스트북을 구비하였다.

다. 요즘처럼 날씨가 더운 때에는 공캉스가 최고야.

(10)에 제시된 ‘청고추’은 ‘홍고추’와 대비되는 “푸른 빛깔의 고추”를 뜻하나, <우리말샘>에 그 형태가 등재되어 있지 않다. “교과서”를 뜻하는 ‘텍스트북’과 “공항에서 보내는 바캉스”를 의미하는 ‘공캉스’는 일상생활에서 자주 사용되지만 <우리말샘>에 등재되지 않은 어휘이다. 특히 ‘청고추’의 경우 ‘청고추채’가 <우리말샘>에 등재되어 있고, 뜻풀이에 ‘청고추’가 사용되고 있다는 점을 통하여 향후 사전 등재 가능성을 짐작할 수 있다.

본 사업에서는 형태 미등재어를 위한 어휘의미 표지 ‘777’과 의미 미등재어를 위한 어휘의미 표지 ‘888’ 외에 어휘의미 표지 ‘999’를 설정하였다. <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침에서 형태 분석의 오류를 표시하기 위해 ‘99’를 사용한 것과 달리 본 사업에서는 말뭉치 원어절의 오타, 탈자 등 표기 오류를 표시하기 위한 수단으로 사용하였다. 형태 분석의 경우 어휘의미 분석 과정에서 NA를 부여하고 즉각적으로 수정하여 형태 분석의 오류를 제거하였다.

(11) 어휘의미 표지 ‘999’의 예

가. 고짚을 부리다.

‘999’ 표지를 부여받은 어휘는 ‘777’, ‘888’과는 달리 사전 등재 가능성이 없기 때문에 별도로 구분하여 다룰 필요가 있다. 다만 <우리말샘>에 등재된 표준어에 대한 일반적인 비표준어형과의 구분이 필요하다. (12)와 같은 비표준어형은 향후 <우리말샘>에 등재될 가능성이 있기 때문에 경우에 따라 ‘777’ 또는 ‘888’을 부여하였다.

(12) 비표준어형의 예

가. 요즘 같은 경우는 다 씨씨티비가 있기 때문에

나. 대한민국이 이 오이씨디에 가입한 것은 천구백구십육년입니다.

‘씨씨티비’와 ‘오이씨디’는 각각 <우리말샘>에 등재된 ‘시시<sup>^</sup>티브이(CCTV)’와 ‘오이시디(OECD)’를 구어적으로 말할 때 나타날 수 있는 형태이다. 이들은 향후 사전 등재

가능성을 염두에 두고 오타 등의 표기 오류로 다루기보다는 비표준어형으로 다루어야 한다. (12)에 제시된 예는 전부 형태 미등재어이기 때문에 ‘777’을 부여했다.

이상의 논의를 반영하여 본 사업에서 마련한 어휘의미 분석 말뭉치 구축 지침에서는 세 가지 어휘의미 표지를 설정하고 각각을 다음과 같이 규정하였다.

<p>■ &lt;우리말샘&gt;에 동일한 한글 배열로 이루어진 형태가 등재되지 않은 어휘의 경우, 어휘의미 번호는 ‘777’로 한다.</p>
<p>[예시] 나는 된장을 끓을 때, <u>홍고추도</u> 넣고 <u>청고추도</u> 넣는다. [청고추__777/NNG+ 도/JX]</p> <p>전공에 관련된 <u>텍스트북</u>을 구비하였다. [텍스트북__777/NNG+ 을/JKO]</p> <p>요즘처럼 날씨가 더운 때에는 <u>공강스가</u> 최고야. [공강스__777/NNG+ 가/JKS]</p>
<p>● 말뭉치에 나타난 준말에 대응하는 본말이 &lt;우리말샘&gt;에 등재되어 있더라도 준말의 형태는 미등재어이기 때문에 어휘의미 번호 ‘777’을 부여한다.</p>
<p>[예시] 오송역에서 <u>교원대로</u> 향했다. [교원대__777/NNG+ 로/JKB]</p> <p>금융투자협회 회장 선거가 여의도 <u>금투협회</u>에서 열렸다. [금투협회__777/NNG+ 에서/JKB]</p>
<p>■ &lt;우리말샘&gt;에 동일한 한글 배열로 이루어진 형태가 등재되어 있으나 해당 어휘의 의미가 제시되어 있지 않은 경우, 어휘의미 번호는 ‘888’로 한다.</p>
<p>[예시] 집을 구하려면 <u>부동산</u>을 찾아가 보는 것이 편리하다. [부동산__888/NNG+ 을/JKO]</p> <p>코리아 <u>드림</u>을 꿈꾸다. [드림__888/NNG+ 을/JKO]</p> <p><u>금융감독원</u>의 제재를 받았다. [금융__001/NNG+ 감독원__888/NNG+ 의/JKG]</p>
<p>● 특정 분야의 전문 용어로 &lt;우리말샘&gt;에 등재된 어휘가 다른 분야에서 쓰이거나 일상 용어로 사용된 경우 어휘의미 번호 ‘888’을 부여한다.</p>

[예시] 대부분이 여행사의 마라톤 <u>패키지</u> 상품으로 <u>홍콩</u> 을 찾았다.	[패키지__888/NNG]
골프채의 헤드와 <u>샤프트</u> 모두 나무를 깎아 만들기 시작했다.	[샤프트__004/NNG]
<p>■ 말뭉치 원어절의 오타, 탈자 등의 오류로 의미 분석이 불가능한 경우, 어휘의미 번호는 '999'로 한다.</p>	
[예시] <u>고깃을</u> 부리다.	[고깃__999/NNG+ 을/JKO]
<p>● 단, &lt;우리말샘&gt;에 등재된 표준어에 대한 일반적인 비표준어형으로 판단되는 경우, 향후 등재 가능성을 고려하여 '777'을 부여한다.</p>	
[예시] 가카__777/NNG : '각하'의 비표준어형	
가라테__777/NNG : '가라테'의 비표준어형	

### 1.2.2. 분석 대상 품사의 차이에 따른 세부 지침의 신설

본 사업에서 구축한 어휘의미 분석 말뭉치는 내용어 중 체언, 즉 일반명사, 의존명사, 고유명사, 대명사, 수사, 어근을 대상으로 어휘의미 번호를 부여했다. 이는 <21세기 세종계획>에 의한 형태의미 분석 말뭉치에서 다루는 품사 중에 동사와 형용사, 관형사, 일반부사가 제외되고 고유명사와 대명사, 수사, 어근이 분석 대상에 추가되었다. 따라서 분석 대상에 새로 추가된 품사, 특히 고유명사와 수사에 대한 세부 지침을 마련하였다.

#### ① 고유명사

고유명사는 대상을 서로 변별하기 위하여 특정한 사물에 붙인 이름으로 지시 대상만 가질 뿐이다. 다시 말해서 사전에 기술된 고유명사의 뜻풀이는 의미가 아닌 지시 대상의 속성을 가진다. 따라서 <우리말샘>에 등재된 일반명사가 고유명사로 쓰이거나, 고유명사가 일반명사로 사용된 경우 고유명사의 대상성에 주목하여 의미 미등재어로 분석하였다.

(13) 고유명사의 대상성

가. 새누리당 저희 당 민주당 그 다음에 국민의당 이렇게 네 개잖아요.

나. 한겨레를 펼쳐 보다.

다. 영화 '변호인'을 보기 위해 극장으로 몰려갔다.

라. 파격적 구도의 고려 수월관음도가 일본에서 발견되었다.

<우리말샘>에 제시된 '민주당'은 1970년 12월 15일에 해체된 “1967년 4월 29일 구 민주당계 인사와 신민당 이탈자가 중심이 되어, 구 민주당의 정신을 계승하고 순수 야당을 재건한다는 취지로 창당한 정당”으로 규정되어 있다. 하지만 (13가)의 ‘민주당’은 오늘날 존재하는 ‘더불어 민주당’을 뜻하기 때문에 서로 다른 지시 대상을 가진다. 따라서 이를 의미 미등재어로 판단하였다.

(13나, 다)는 일반명사가 고유명사로 쓰인 경우이다. ‘한겨레’는 “큰 겨레라는 뜻으로, 우리 겨레를 이르는 말”이지만 (13나)에서는 신문사를 지시하는 이름으로 사용되었다. ‘변호인’ 역시 “형사 소송에서, 피의자나 피고인의 이익을 보호하는 보조자로서 변호를 담당하는 사람”을 뜻하는 일반명사이지만 (13다)에서는 영화 제목으로 쓰였다. ‘변호인’의 경우 일반명사의 의미와 영화 내용이 유사성을 가지지만 고유명사는 의미가 아닌 지시 대상을 가지기 때문에 의미 미등재어로 분류하였다.

(13라)는 고유명사가 일반명사로 사용된 예이다. <우리말샘>에 따르면 ‘수월관음도’는 보물 제1286호로 지정된 “경기도 용인시 용인 대학교에 소장된 고려 시대의 불화”를 가리킨다. 그러나 (13라)의 ‘수월관음도’는 <우리말샘>에서 지시하는 대상이 아닌 일본에서 발견된 고려 불화를 의미한다. 이는 ‘수월관음도’라는 고유명사가 “보물 제1286호”라는 지시 대상을 상실하고, “관세음보살의 서른세 가지 모습 중 수월관음이 그려진 그림”이라는 의미를 가지기 때문에 의미 미등재어로 분석하였다.

고유명사 중 대표적인 것은 인명과 지명이다. 이 경우 <우리말샘>을 참고하여 최대한 어휘의미를 부여하는 것을 원칙으로 하되, 동일 형태가 없는 인명이나 지명은 형태 미등재어로 처리하고, 동명이인이나 동명이물 등은 의미 미등재어로 처리하였다. 이러한 경우 인명의 ‘성(姓)’이 문제가 되는데 해당 인물의 본관을 일일이 확인할 수 없기 때문에 부득이하게 국가통계포털에 제시된 “성씨·본관별 인구(5인 이상)-전국”을 참고하여 대표적인 성의 의미 번호를 통일하였다. 이상의 내용과 관련된 규정은 다음과 같다.

● <우리말샘>에 등재된 일반명사가 고유명사로 쓰이거나, 고유명사가 일반명사로 사용된 경우 고유명사의 대상성에 주목하여 어휘의미 번호 '888'을 부여한다.

[예시] 한겨레를 펼쳐 보다. [한겨레\_\_888/NNP+ 를/JKO]

영화 '변호인'을 보기 위해 극장으로 몰려갔다.

[ '/SS+ 변호인\_\_888/NNP+ 'SS+ 을/JKO]

과격적 구도의 고려 수월관음도가 일본에서 발견되었다.

[수월관음도\_\_888/NNG+ 가/JKS]

● 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.

[예시] 이순신\_\_002/NNP : 충무공 김옥균\_\_001/NNP : 갑신정변

아인슈타인\_\_001/NNP : 상대성이론 비틀즈\_\_001/NNP : 록 그룹

김종서\_\_001/NNP : 조선 전기의 충신 김종서\_\_888/NNP : 록 가수

김대중\_\_777/NNP 노무현\_\_777/NNP

이명박\_\_777/NNP 박근혜\_\_777/NNP

● '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.

[예시] 정 시인은 주로 자연의 정취를 주제로 서정시를 쓴다. [정\_\_054/NNP]

붙임. 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	철__013	심__015	유__022	주__012	하__007
노__013	범__006	성__006	양__017	윤__001	지__010	호__016
너__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>을 참조하여 최대한 어휘의미를 부여하는 것을 원칙으로 하되, 한자 또는 영문 표기의 일치 여부와 관계없이 해당 표기가 사전에 등재되어 있지 않은 것은 어휘의미 번호 ‘777’을 부여하고, 등재되어 있는 것은 어휘의미 번호 ‘888’을 부여한다.

[예시] 서울__002/NNP	송파구__777/NNP
캘리포니아__777/NNP	런던__003/NNP
런던__888/NNP : 캐나다 온타리오주	분당__888/NNP : 경기 성남
가수리__002/NNP : 경북 영천	가수리__003/NNP : 경남 의령
가수리__888/NNP : 강원 정선	가수리__888/NNP : 전남 화순

## ② 수사

수사는 간혹 수관형사 또는 명사와의 구별이 어려울 때가 있다. 그로 인해 형태 분석에서는 특이한 형식을 가진 예만을 수관형사로 취급하고, 그 밖의 것들은 모두 수사로 분석하였다. 이에 어휘의미 분석에서도 수사로 처리된 어휘에 대해 수사의 의미 번호를 부여하여 품사와 어휘의미 번호를 일치시켰다.

● 수사는 때로 수사와 수관형사, 수사와 명사의 구별이 애매한 경우가 있다. 이에 형태 분석에서는 특이한 형식을 가진 예만을 수관형사로 취급하고, 그 밖의 것들은



모두 수사로 분석하였다. 이에 어휘의미 분석에서도 수사로 처리된 어휘에 대해 수사의 의미 번호를 부여하여 품사와 어휘의미 번호를 일치하게 한다.

[예시] 우리는 <u>하나로</u> 뭉쳤다.	[하나__001/NR+ 로/JKB]
만점자가 주위에 <u>한둘이</u> 아니다.	[한둘__001/NR+ 이/JKC]
상대에게 <u>몇</u> 차례 결정적 기회를 내줬다.	[몇__001/NR]
지금도 <u>몇몇</u> 분야는 두각을 나타내고 있다.	[몇몇__001/NR]
농사지은 것으로 <u>일년</u> 내내 밥상을 차렸다.	[일__001/NR+ 년002/NNB]

### ③ 기타

말뭉치의 분석 대상 어휘가 간혹 <우리말샘>에 ‘^’을 포함한 구로 등재된 경우가 있다. 이는 경우에 따라 한 단어로 인식되기 때문에 어휘의미 분석 과정에서 해당 구 등재어의 의미 번호를 부여하였다.

#### (14) 분석 대상 어휘가 ‘^’을 포함한 구로 등재된 경우

가. 국가인권위원회는 학습권 보장의 필요성을 제기했다.

나. 이 건물은 평소 템플스테이 숙소로 쓰여 왔다.

나. 뜻밖에 템플 스테이

(14가)의 ‘국가인권위원회’는 <우리말샘>에 ‘국가^인권^위원회’로 등재되어 있고, 형태 분석 말뭉치 구축 지침에 의거하여 하나의 어휘로 분석된다. 이 경우 ‘국가인권위원회’ 전체에 의미 번호 ‘001’을 부여하였다. (14나)의 ‘템플스테이’ 형태 분석 말뭉치에서 하나의 어휘로 분석되어 있다. 이 역시 <우리말샘>에 ‘템플^스테이’로 등재되어 있기 때문에 의미 번호 ‘001’로 분석하였다. 단 (14나)처럼 ‘템플, 스테이’가 어절로 분리되어 말뭉치에 나타난 경우 ‘템플’과 ‘스테이’에 각각 분석하여 어휘의미 번호를 부여하였다.

말뭉치의 분석 대상 어휘가 <우리말샘>에 북한어, 방언 등으로 등재되어 있으면 해당 어휘의미 번호를 부여하였다. 다만 분석 대상 어휘의 품사와 방언의 품사가 불일치할 경우 품사 일치 여부를 우선 적용하여 해당 어휘의미 번호를 부여하지 않았다.

(15) 분석 대상 어휘가 방언일 경우

가. 며늘아, 차례 음식은 대행에 맡기렴.

나. 것도 그렇고

‘며늘’은 ‘며느리’의 경북, 중국 흑룡강성 방언으로 <우리말샘>에 단독으로 등재되어 있고, (15가)의 의미가 “며느리”를 뜻하기 때문에 해당 어휘에 의미 번호 ‘001’을 부여하였다. 이와 달리 (15나)의 ‘것’은 ‘그것’을 구어적으로 이르는 말로, 그 형태와 의미가 <우리말샘>에 “‘그것’의 제주 방언”으로 등재되어 어휘와 유사성을 가진다. 그러나 (15나)의 ‘것’은 품사가 대명사이고, <우리말샘> 등재어의 품사는 명사이기 때문에 품사와 어휘의미 번호 일치를 고려하여 ‘것’에 어휘의미 번호 ‘888’을 부여하였다.

분석 대상 어휘가 숫자나 알파벳을 포함하는 경우가 있다. 이 경우 분석 대상 어휘에 나타나는 숫자나 알파벳을 한글로 바꿔 쓴 형태를 <우리말샘>에 검색하여 등재 유무에 따라 해당 어휘의미 번호를 부여하였다.

(16) 분석 대상 어휘가 숫자나 알파벳을 포함한 경우

가. 모두 한국신기록을 세우며 대회 3관왕이 됐다.

나. U턴 문제 등 우려하는 문제가 불거지지 않도록 하겠다.

형태 분석 말뭉치 구축 지침에 의거하여 (16)처럼 ‘3관왕’과 ‘U턴’이 하나의 어휘로 분석되었다. 이 경우 숫자나 알파벳을 한글 표기로 전환하여 ‘삼관왕’과 ‘유턴’을 <우리말샘> 검색하여 해당 의미 번호를 부여하였다. 실제로 <우리말샘>에는 ‘삼관왕’이 두 가지 갈래 뜻으로, ‘유턴’은 단독형으로 등재되어 있다.

분석 대상 어휘 내부에 기호가 포함된 경우도 존재한다. 이 경우에는 기호를 생략한 형태가 <우리말샘>에 등재되어 있는가를 확인하였다.

(17) 분석 대상 어휘가 기호를 포함한 경우

가. 건물의 냉·난방비를 줄여준다.

나. 국·공립대 중 1인당 장학금 수혜율 3위를 기록하고 있다.

‘냉·난방비’와 ‘국·공립대’ 역시 형태 분석 말뭉치 구축 지침에 따라 더 이상 분석되지

않는다. 분석 대상 어휘에서 기호를 생략한 형태인 ‘냉난방비’와 ‘국공립대’가 <우리말샘>에 단독으로 등재되어 있으므로 각각 의미 번호 ‘001’로 분석하였다. 이상의 내용을 바탕으로 마련된 지침은 다음과 같다.

<p>● 말뭉치의 분석 대상 어휘가 &lt;우리말샘&gt;에 구로 등재되어 있으면 분리하지 않고 전체에 해당 어휘의미 번호를 부여한다. 단, 말뭉치에서 분리된 두 단어가 합쳐져서 &lt;우리말샘&gt;에 구로 등재되어 있더라도 합하지 않고 각각에 해당하는 어휘의미 번호를 부여한다.</p>
<p>[예시] 국가인권위원회는 학습권 보장의 필요성을 제기했다. [국가인권위원회__001/NNP+ 는/JX]  이 건물은 평소 템플스테이 숙소로 쓰여 왔다. [템플스테이__001/NNG]  뜻밖에 템플 스테이 [템플__888/NNG, 스테이__888/NNG]</p>
<p>● 말뭉치의 분석 대상 어휘가 &lt;우리말샘&gt;에 북한어, 방언 등으로 등재되어 있으면 해당 어휘의미 번호를 부여한다.</p>
<p>[예시] 며늘아, 차례 음식은 대행에 맡기렴. [며늘__001/NNG+ 아/JKV+ ,/SP]  ※며늘001 : ‘며느리’의 방언(경북, 중국 흑룡강성).</p>
<p>● 말뭉치의 분석 대상 어휘가 숫자나 알파벳을 포함한 경우, 해당 숫자나 알파벳을 한글로 옮겨 쓴 형태가 &lt;우리말샘&gt;에 등재되어 있으면 해당 어휘의미 번호를 부여한다.</p>
<p>[예시] 모두 한국신기록을 세우며 대회 3관왕이 됐다. [3관왕__002/NNG+ 이/JKC]  U턴 문제 등 우려하는 문제가 불거지지 않도록 하겠다. [U턴__001/NNG]</p>
<p>● 말뭉치의 분석 대상 어휘가 기호를 포함한 경우, 기호를 생략한 형태가 &lt;우리말샘&gt;에 등재되어 있으면 해당 어휘의미 번호를 부여한다.</p>
<p>[예시] 건물의 냉·난방비를 줄여준다. [냉·난방비__001/NNG+ 를/JKO]  국·공립대 중 1인당 장학금 수혜율 3위를 기록하고 있다. [국·공립대__001/NNG]</p>

### 1.2.3. 기존 규정에 대한 상세 기술

<21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침은 규정이 간략하게 기술되어 있는 편이다. 앞에서 언급한 바와 같이 접사 처리 방법에 대한 설명이 없어 어근에 결합 혹은 분리해야 할 접사 목록을 확인할 수 없다. 이 지침으로 어휘의미를 분석할 경우 접사 결합형에서 분석 대상 어휘 자체의 형태가 달라지게 된다. 따라서 본 사업에서 마련한 어휘의미 분석 말뭉치 구축 지침에는 접사 처리(통합) 과정에서 제외된 접사 목록을 제시함으로써 어휘의미 분석용 말뭉치의 어휘 형태의 통일성을 담보하였다.

의미 분석과 관련하여 <21세기 세종계획>의 형태의미 말뭉치 구축 지침은 좌우 5어 절 환경에서 대상 어휘의 의미를 분석하고 <표준국어대사전>의 어깨번호를 부착한다고 간략히 기술되어 있다. 이와 달리 본 사업은 미세한 차이로 인해 나누어진 의미를 반영하여 분석하기 때문에 지침에 <우리말샘>에 제시된 갈래 뜻 사이에 구분이 불분명할 경우 의미 번호를 선택하는 몇 가지 기준을 마련함으로써 분석의 통일성을 확보하였다.

#### (18) 맥락에 의한 의미 번호 선택

- 가. 허무주의에 대한 반발
- 나. 고동색 갈색
- 다. 분위기

대상 어휘의 의미를 분석하는 가장 기본적인 방법은 맥락에 적절한 의미 번호를 선택하는 것이다. (18가)에 제시된 ‘허무주의’의 경우 각각 ‘사회 일반’, ‘정치’, ‘철학’ 분야의 전문 용어로 등재되어 있다. 그런데 분야에 따른 세 갈래의 뜻의 구분이 불분명하여 하나의 의미 번호를 선택하기 어렵다. 이 경우 문장의 앞뒤 맥락에 나타나는 ‘니힐리즘’이라는 표현에 주목하여 ‘니힐리즘’의 유의어로 등재되어 있는 ‘003’의 의미 번호를 부여하였다.

‘고동색’은 <우리말샘>에 “001 검붉은 빛을 띤 누런색” 또는 “002 붉은빛이 조금 나는 짙은 갈색”으로 등재되어 있는데, 그 의미를 구분하는 것이 어렵다. 예문을 활용하는 방안이 있으나 두 갈래 뜻 모두에 예문조차 주어지지 않기 때문에 맥락에 최대한 의존할 수밖에 없다. 이 경우 뒤이어 ‘갈색’이라는 표현이 나타난 것을 참고하여 뜻풀이에 ‘갈색’이 포함되어 있는 ‘002’의 의미 번호를 부여하였다.

‘분위기’의 여러 갈래 뜻 중 “002 그 자리나 장면에서 느껴지는 기분”과 “003 주위를 둘러싸고 있는 상황이나 환경”은 의미 구분에 있어서 모호한 부분이 있다. 이 경우 앞뒤 맥락을 통해서 당장의 상황으로 파악되면 의미 번호 ‘002’로, 그렇지 않으면 ‘003’으로 분석하였다.

(19) 예문 검토를 통한 의미 번호 선택

가. 후보가 되면 심층 인성평가를 거친다.

나. 자신의 삶이 나아지는 것에 익숙하게 경험을 하고 살았어요.

다. 얼굴이 예쁘다.

라. 가정 생계를 순전히 남자의 수입에만 의존해가지고는

‘인성’은 <우리말샘>에 “001 사람의 성품”과 “002 각 개인이 가지는 사고와 태도 및 행동 특성”이라는 의미로 제시되어 있는데, 둘 사이에 구분이 불분명하다. 이러한 경우 예문을 통해 ‘인성’의 의미 번호를 선택할 수 있다. 전자에 제시된 예문 ‘인성 교육, 인성이 착하다, 올바른 인성을 기르다’과 후자에 제시된 예문 ‘인성 개발, 인성 검사’을 비교하여 (19가)에 제시된 ‘인성’에 어휘의미 번호 ‘002’를 부여하였다.

‘경험’은 <우리말샘>에 “001 자신이 실제로 해 보거나 겪어 봄. 또는 거기서 얻은 지식이나 기능”과 “002 객관적 대상에 대한 감각이나 지각 작용에 의하여 깨닫게 되는 내용”으로 풀이되어 있다. 이 경우 후자의 예문 ‘경험과 직관’보다는 전자의 예문 ‘그동안의 경험을 살려서 한번 잘 해 보시오.’에 가까운 의미로 판단하였다.

‘얼굴’은 <우리말샘>에 “001 눈, 코, 입이 있는 머리의 앞면” 또는 “002 머리 앞면의 전체적 윤곽이나 생김새”를 의미한다. 이 두 의미가 유사하기 때문에 ‘얼굴’ 역시 예문을 근거로 의미 번호를 결정해야 한다. 이 경우 전자의 예문 ‘복면으로 얼굴을 가리다. 그녀는 얼굴에 로션을 발랐다.’보다는 후자의 예문 ‘얼굴이 참 곱다. 어른들은 그녀의 얼굴이 복스럽다며 매우 좋아하셨다.’가 적절한 것으로 판단되어 (19다)에 제시된 ‘얼굴’에 어휘의미 번호 ‘002’를 부여하였다.

‘수입’은 <우리말샘>에 “001 돈이나 물품 따위를 거두어들임. 또는 그 돈이나 물품” 또는 “002 개인, 국가, 단체 따위가 법에 어긋나지 아니하게 얻어 들이는 일정액의 금액”으로 정의되어 있다. 이 경우 역시 전자의 제시된 예문 “쥐꼬리만 한 내 수입으로 다섯 식구가 생활하고 있다.”를 참고하여 의미 번호 ‘001’을 선택하였다.

(20) 예문 유무에 따른 의미 번호 선택

- 가. 매서운 강풍이 문풍지를 뒤흔들었다.
- 나. 다람쥐는 없는 거 같아
- 다. 의지
- 라. 나는 슬픔에 잠겼다.
- 마. 서울메트로가 운영하는 지하철을 타야한다.
- 사. 가지 꽃나무란
- 아. 체제

<우리말샘>에 등재된 ‘강풍’의 다양한 의미 중 ‘003’과 ‘006’은 “세계 부는 바람”으로 의미가 동일하다. 그러나 ‘003’에는 예문이 없고 ‘006’에만 예문이 있기 때문에 (20가)에 제시된 ‘강풍’에 어휘의미 번호 ‘006’을 부여하였다. ‘다람쥐, 의지, 슬픔, 지하철, 꽃나무, 체제’ 등도 <우리말샘>에 두 가지 이상의 유사한 의미로 정의되어 있으나, 예문이 하나의 뜻풀이에만 제시되어 있다. 이 경우 예문이 제시된 의미 번호를 선택하여 ‘다람쥐\_\_001/NNG, 의지\_\_007/NNG, 슬픔\_\_001/NNG, 지하철\_\_001/NNG, 꽃나무\_\_001/NNG, 체제\_\_003/NNG’로 분석하였다.

(21) 포괄적인 의미 번호 선택

- 가. 창밖으로 시원한 소나무숲이 들어온다.
- 나. 근처 중국식당에서 피로연을 즐겼다.
- 다. 국어 기본법이 제정되었다.
- 라. 고려시대 배가 모두 사차에 걸치는 밭길 작업을 마쳤습니다.

<우리말샘>에 등재된 ‘소나무’의 다양한 의미 중 ‘001’은 “소나무과의 모든 식물을 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 (21가)에 제시된 ‘소나무’에 어휘의미 번호로 부여하였다.

<우리말샘>에 등재된 ‘중국’ 역시 다양한 의미 중 ‘001’은 삼황오제 시대부터 지금의 중화 인민 공화국을 가리키는 “아시아 동부에 있는 나라”로, “아시아 동북부에 있는 인

민 공화국”을 의미하는 ‘002’를 포함하고 있다. 두 갈래 뜻 모두에 예문이 제시되지 않아 한쪽을 선택하기 어려우므로 포괄적인 의미를 나타내는 ‘001’을 (21나)에 제시된 ‘중국’에 어휘의미 번호로 부여하였다.

<우리말샘>에 등재된 ‘기본법’의 다양한 의미 중 “002 다른 여러 가지 법의 기본이 되는 법”과 “003 일반적으로 국가의 기본 조직을 규정하는 법. 헌법 따위를 이른다.”는 모두 예문이 제시되어 있어 한쪽의 의미 번호를 선택하기 어려우므로 포괄적인 의미를 나타내는 ‘002’을 (21다)에 제시된 ‘기본법’의 어휘의미 번호로 부여하였다.

(21라)의 ‘작업’은 (21)의 나머지 예와 조금 다르다. <우리말샘>에 등재된 ‘작업’의 다양한 의미 중 “001 일을 함. 또는 그 일”과 “002 일정한 목적과 계획 아래 하는 일”는 유사한 예문을 가지고 있어 예문으로 의미를 구분하는 것이 어렵다. 이 경우 예문이 없는 유형과 동일하게 포괄적인 의미를 선택하여, (21라)의 ‘작업’에는 의미 번호 ‘001’을 부여하였다.

## 2. 어휘의미 분석 말뭉치 구축 지침

이 장에서는 본 사업에서 어휘의미 분석 말뭉치를 구축하기 위하여 적용한 지침의 전모를 보이고자 한다. 지침의 구성은 다음과 같다.

1. 기본 원칙
  - 가. 분석대상
  - 나. 분석원리
  - 다. 분석원칙
2. 어휘의미 분석 표지
3. 세부 지침
  - 가. 일반명사(NNG)
  - 나. 고유명사(NNP)
  - 다. 수사(NR)
  - 라. 기타



### 가. 분석대상

어휘의미 분석 말뭉치의 분석 대상이 되는 형태 분석 말뭉치의 원문(원어절과 분석 결과 모두)은 가급적 수정하지 않은 채, 당해 분석 대상 품사는 기능어(function words)를 제외한 내용어(content words) 중 체언을 중심으로 선정하였다. 그중 대상 범주는 다음과 같다.

일반명사(NNG), 의존명사(NNB), 고유명사(NNP), 대명사(NP), 수사(NR), 어근(XR)

단, 형태 분석 말뭉치에서 발견되는 분석 오류에 대해서는 이를 별도로 기록한 후 형태 분석 말뭉치의 개선에 반영될 수 있도록 한다.

### 나. 분석원리

어휘의미 분석 말뭉치의 구축 시, 형태 분석 말뭉치를 어휘 분석 말뭉치 수준으로 변환하여 구축하는 것을 원칙으로 한다. 즉, 형태 분석 말뭉치에서 접사 처리(통합) 과정을 거쳐 만들어진 어휘 분석 말뭉치를 대상으로 어휘의미(갈래 뜻, sense)를 구별하여 표지(번호)를 부착한다. 단, 통사적 접사를 포함하여 복합어 자체의 의미 특성을 크게 나타내지 않아 어휘의미의 표지를 부여하는 것에 효용성이 떨어지는 접사는 접사 처리(통합) 과정에서 제외한다. 또한 당해의 구축 말뭉치는 향후 활용성을 감안하여 서술성 명사류에 용언을 파생시키는 접미사가 결합된 파생 용언의 경우에는 접사 처리(통합) 과정을 거치지 않고 어근(어기)에 해당하는 부분에 어휘의미의 표지를 부여한다.

붙임. 접사 처리(통합) 과정에서 제외된 접사 목록

명사과생접미사				동사과생접미사	
가(哥)	김가	들이	10들이	당하	공격당하다
가량	1시간가량	발(發)	서울발	되	준비되다
간(間)	한 달간	배기*	나이배기	받	강요받다
경(頃)	두 시경	분지(分之)	삼분지 일	시키	운동시키다
계(系)	몽고계	생1(生)	갑자생	하	공부하다
계	10분계	씩	만원씩		
꿀	십 원꿀	어치	만원어치	형용사과생접미사	
끼리	진우끼리	여(餘)	삼십여	답	사람답다
네	동이네	정(整)	일만 원정	되	거짓되다
당(當)	한 사람당	짜리	백 원짜리	롭	슬기롭다
대(臺)	억대	째1**	이틀째	스럽	사랑스럽다
댁(宅)	청주댁	썸	내일썸	하	건강하다
들	우리들	하(下)	지배하		

\* '-배기'는 '세 살, 다섯 살' 등 수량을 나타내는 구와 결합하는 경우 접사 처리(통합)에서 제외하나, '나이배기, 알배기, 공짜배기, 대짜배기, 진짜배기'는 결합형으로 처리함.

\*\* '-째1'은 접사 처리(통합)에서 제외하나, 의존명사 '번째'는 결합형으로 처리함.

\*\*\* 이 밖에 숫자나 문장부호, 띄어쓰기 등으로 어근과 접사가 분리된 경우 접사 처리(통합)에서 제외함. (예. 제2차, 고(高)지대를, 4분의 1 등)

[예시] 로이터통신의 서울발 기사였다.

[서울발\_\_001/NNG] (×) [서울\_\_002/NNG+ 발/XSN] (○)

주중에는 공부하고 주말에는 야구하자.

[공부하\_\_001/VV+ 고/EC] (×) [공부\_\_001/NNG+ 하/XSV+ 고/EC] (○)

## 다. 분석원칙

분석 대상 후보가 되는 모든 어형에 어휘의미 번호를 부여한다.

### 2. 어휘의미 분석 표지

가. 어휘의미의 목록과 체계는 국립국어원의 <우리말샘>을 기준으로 한다.

나. 말뭉치 정보 부착 방법은 해당 어절의 분석 결과에 ‘\_\_’과 세 자리의 어휘의미 번호 ‘○○○’를 부착하는 방식을 사용한다. 의미 번호는 국립국어원 <우리말샘>에 나타난 표제어 등재 번호와 일치한다.

---

[예시] 돈이 꽃이 되면 <u>기적</u> 같은 일을 일으킨다.	[기적__003/NNG]
서울 종로구청 앞에 첫 <u>번째</u> 소녀상을 세웠다.	[번째__001/NNB]
<u>나는</u> 지금 하는 일을 좋아서 하고 있는가.	[나__003/NP+ 는/JX]

---

<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.

---

[예시] 후보가 되면 심층 <u>인성평가</u> 를 거친다.	[인성__002/NNG+ 평가__001/NNG+ 를/JKO]
-----------------------------------	-----------------------------------

※<우리말샘>에 등재된 ‘인성’의 다양한 의미 중 “001 사람의 성품”과 “002 각 개인이 가지는 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.

---

매서운 <u>강풍</u> 이 문풍지를 뒤흔들었다.	[강풍__006/NNG+ 이/JKS]
-----------------------------	----------------------

※<우리말샘>에 등재된 ‘강풍’의 다양한 의미 중 ‘003’과 ‘006’은 “세계 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.

[예시] 창밖으로 시원한 소나무숲이 들어온다.

[소나무\_\_001/NNG+ 숲\_\_001/NNG+ 이/JKS]

※<우리말샘>에 등재된 ‘소나무’의 다양한 의미 중 ‘001’은 “소나무과와 모든 식물을 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 ‘소나무’에 어휘의미 번호로 부여한다.

근처 중국식당에서 피로연을 즐겼다.

[중국\_\_001/NNP+ 식당\_\_002/NNG+ 에서/JKB]

※<우리말샘>에 등재된 ‘중국’의 다양한 의미 중 ‘001’은 삼황오제 시대부터 지금의 중화 인민 공화국을 가리키는 “아시아 동부에 있는 나라”로, “아시아 동북부에 있는 인민 공화국”을 의미하는 ‘002’를 포함하고 있다. 두 갈래 뜻 모두에 예문이 제시되지 않아 한쪽을 선택하기 어려우므로 포괄적인 의미를 나타내는 ‘001’을 위 문장에 제시된 ‘중국’에 어휘의미 번호로 부여한다.

다. <우리말샘>에 동일한 한글 배열로 이루어진 형태가 등재되지 않은 어휘의 경우, 어휘의미 번호는 ‘777’로 한다.

[예시] 나는 된장을 끓을 때, 홍고추도 넣고 청고추도 넣는다. [청고추\_\_777/NNG+ 도/JX]

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

말뭉치에 나타난 준말에 대응하는 본말이 <우리말샘>에 등재되어 있더라도 준말의 형태는 미등재어이기 때문에 어휘의미 번호 ‘777’을 부여한다.

[예시] 오송역에서 교원대로 향했다. [교원대\_\_777/NNG+ 로/JKB]

금융투자협회 회장 선거가 여의도 금투협회에서 열렸다.

[금투협회\_\_777/NNG+ 에서/JKB]

※‘교원대’와 ‘금투협회’의 본말인 ‘한국^교원대’와 ‘한국^금융^투자^협회’가 <우리말샘>에 등재되어 있더라도 준말의 형태는 미등재어에 해당하기 때문에 어휘의미 번호 ‘777’을 부여한다.

라. <우리말샘>에 동일한 한글 배열로 이루어진 형태가 등재되어 있으나 해당 어휘의 의미가 제시되어 있지 않은 경우, 어휘의미 번호는 ‘888’로 한다.

---

[예시] 집을 구하려면 부동산을 찾아가 보는 것이 편리하다. [부동산\_\_888/NNG+ 을/JKO]

※<우리말샘>에 ‘부동산’이라는 표제어가 등재되어 있는데, 현재 사전상의 의미는 ‘동산(動産)의 반의어인 ‘부동산’(不動産)이다. 그런데 위 예문에 쓰인 ‘부동산’의 의미는 “움직여 옮길 수 없는 재산”이 아니라 “공인 중개 사무소”이다. 따라서 그 형태는 등재되어 있지만 해당 의미가 제시되어 있지 않은 ‘부동산’에 어휘의미 번호 ‘888’을 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.

금융감독원의 제재를 받았다. [금융\_\_001/NNG+ 감독원\_\_888/NNG+ 의/JKG]

※<우리말샘>에 “감독하는 직무를 맡은 사람”을 의미하는 ‘감독원(監督員)’이 등재되어 있으나 “감독 기관”에 해당하는 의미가 제시되어 있지 않으므로 ‘감독원(監督院)’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.

---

[예시] 대부분이 여행사의 마라톤 패키지 상품으로 홍콩을 찾았다. [패키지\_\_888/NNG]  
※<우리말샘>에 제시된 ‘패키지’는 『영상』 분야에 사용되는 전문 용어로 “시나리오 작가와 인기 배우, 인기 소설과 인기 배우 등의 결합”을 의미한다. 위 예문에 쓰인 ‘패키지’ 역시 “일괄 상품 혹은 상품의 결합” 정도의 의미로 <우리말샘>에 등재된 어휘와 의미적 관련성을 찾을 수 있으나 『서비스업』이라는 다른 분야에서 사용되는 용어이기 때문에 해당 어휘의 의미가 등재되지 않은 것으로 판단하여 어휘의미 번호 ‘888’을 부여한다.

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 예문에 쓰인 ‘샤프트’는 “골프채의 긴 막대 부분”을 가리키기 때문에 <우리말샘>에 등재된 어휘와 의미적 관련성을 찾을 수 있다. 비록 구체적인 운동 종목은 다르나 『체육』이라는 동일 분야에서 사용되는 용어이기 때문에 <우리말샘>에 등재된 어휘의미 번호 ‘004’를 부여한다.

---

<우리말샘>에 등재된 일반명사가 고유명사로 쓰이거나, 고유명사가 일반명사로 사용된 경우 고유명사의 대상성에 주목하여 어휘의미 번호 ‘888’을 부여한다.

---

[예시] 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 대상성에 주목하여 어휘의미 번호 ‘888’을 부여한다.

영화 ‘변호인’을 보기 위해 극장으로 몰려갔다.  
[‘/SS+ 변호인\_\_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/NNG+ 가/JKS]  
※<우리말샘>에 등재된 ‘수월관음도’는 용인대학교에 소장된 보물 제1286호를 가리킨다. 그러나 위 예문처럼 문화재명 ‘수월관음도’가 맥락상 일반명사로 쓰일 경우 어휘의미 번호 ‘888’을 부여한다.

---

마. 말뭉치 원어절의 오타, 탈자 등의 오류로 의미 분석이 불가능한 경우, 어휘의미 번호는 ‘999’로 한다.

[예시] 고짚을 부리다.

[고짚\_\_999/NNG+ 을/JKO]

※‘고짚’은 ‘고집(NNG)’의 오타가 분명해 보이거나 원문을 수정할 수 없으므로 ‘고집’으로 수정하지 않고 어휘의미 번호 ‘999’를 부여한다.

단, <우리말샘>에 등재된 표준어에 대한 일반적인 비표준어형으로 판단되는 경우, 향후 등재 가능성을 고려하여 ‘777’을 부여한다.

[예시] 가카\_\_777/NNG : ‘각하’의 비표준어형

가라테\_\_777/NNG : ‘가라테’의 비표준어형

※‘가카’와 ‘가라테’는 각각 ‘각하’와 ‘가라테’의 발음에 이끌려 적은 형태로, 단순 오타보다는 비표준어형의 사용으로 판단되므로 어휘의미 번호 ‘777’을 부여한다.

### 3

## 세부 지침

### 가 일반명사(NNG)

수를 셀 수 있는 단위사로 쓰일 수 있는 명사들은 기본 뜻풀이와 셀 때의 뜻풀이가 서로 상이하므로, 그에 맞는 어휘의미 번호를 주어야 한다.

[예시] 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.

[예시] 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> 등재  
 어 ‘대법원’의 포함되는 의미로 다를 수 있다.

## 나 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 있  
 는 것은 어휘의미 번호 ‘888’을 부여한다.

[예시] 이순신__002/NNP : 충무공	김옥균__001/NNP : 갑신정변
아인슈타인__001/NNP : 상대성이론	비틀즈__001/NNP : 록 그룹
김종서__001/NNP : 조선 전기의 충신	김종서__888/NNP : 록 가수
김대중__777/NNP	노무현__777/NNP
이명박__777/NNP	박근혜__777/NNP

‘성’(性)의 경우에는 ‘우리나라 성씨 순위, 성씨별 인구밀도 순위’ 등의 자료를 참고  
 하여 사전에 등재되어 있는 같은 ‘성’(性) 중 대표적인 성의 의미 번호로 통일한다.

[예시] 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.

붙임. 둘 이상 등재된 성(性)의 대표 의미 번호



강__004	도__015	빈__003	수__008*	용__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
녀__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>을 참조하여 최대한 어휘의미를 부여하는 것을 원칙으로 하되, 한자 또는 영문 표기의 일치 여부와 관계없이 해당 표기가 사전에 등재되어 있지 않은 것은 어휘의미 번호 ‘777’을 부여하고, 등재되어 있는 것은 어휘의미 번호 ‘888’을 부여한다.

[예시] 서울__002/NNP	송파구__777/NNP
캘리포니아__777/NNP	런던__003/NNP
런던__888/NNP : 캐나다 온타리오주	분당__888/NNP : 경기 성남
가수리__002/NNP : 경북 영천	가수리__003/NNP : 경남 의령
가수리__888/NNP : 강원 정선	가수리__888/NNP : 전남 화순

## 다 수사(NR)

수사는 때로 수사와 수관형사, 수사와 명사의 구별이 애매한 경우가 있다. 이에 형태 분석에서는 특이한 형식을 가진 예만을 수관형사로 취급하고, 그 밖의 것들은 모두 수사로 분석하였다. 이에 어휘의미 분석에서도 수사로 처리된 어휘에 대해 수사의 의미 번호를 부여하여 품사와 어휘의미 번호를 일치하게 한다.

[예시] 우리는 <u>하나</u> 로 뭉쳤다.	[하나__001/NR+ 로/JKB]
만점자가 주위에 <u>한둘</u> 이 아니다.	[한둘__001/NR+ 이/JKC]
상대에게 <u>몇</u> 차례 결정적 기회를 내줬다.	[몇__001/NR]
지금도 <u>몇몇</u> 분야는 두각을 나타내고 있다.	[몇몇__001/NR]
농사지은 것으로 <u>일년</u> 내내 밥상을 차렸다.	[일__001/NR+ 년002/NNB]

## 라 기타

말뭉치의 분석 대상 어휘가 <우리말샘>에 구로 등재되어 있으면 분리하지 않고 전체에 해당 어휘의미 번호를 부여한다. 단, 말뭉치에서 분리된 두 단어가 합쳐져서 <우리말샘>에 구로 등재되어 있더라도 합하지 않고 각각에 해당하는 어휘의미 번호를 부여한다.

[예시] 국가인권위원회는 학습권 보장의 필요성을 제기했다.

[국가인권위원회\_\_001/NNP+ 는/JX]

※‘국가인권위원회’가 <우리말샘>에 ‘국가^인권^위원회’로 등재되어 있으므로 ‘국가인권위원회\_\_001/NNP’처럼 전체에 어휘의미 번호를 부여한다. 단, ‘국가, 인권, 위원회’가 어절로 분리되어 말뭉치에 나타난 경우 ‘국가\_\_001/NNG, 인권\_\_001/NNG, 위원회\_\_001/NNG’처럼 각 요소의 어휘의미 번호를 부여한다.

이 건물은 평소 템플스테이 숙소로 쓰여 왔다. [템플스테이\_\_001/NNG]

뜻밖에 템플 스테이 [템플\_\_888/NNG, 스테이\_\_888/NNG]

말뭉치의 분석 대상 어휘가 <우리말샘>에 북한어, 방언 등으로 등재되어 있으면 해당 어휘의미 번호를 부여한다.

[예시] 며늘아, 차례 음식은 대행에 맡기렴. [며늘\_\_001/NNG+ 아/JKV+ ,/SP]

※며늘001 : ‘며느리’의 방언(경북, 중국 흑룡강성).

말뭉치의 분석 대상 어휘가 숫자나 알파벳을 포함한 경우, 해당 숫자나 알파벳을 한글로 옮겨 쓴 형태가 <우리말샘>에 등재되어 있으면 해당 어휘의미 번호를 부여한다.

---

[예시] 모두 한국신기록을 세우며 대회 3관왕이 됐다. [3관왕\_\_002/NNG+ 이/JKC]  
U턴 문제 등 우려하는 문제가 불거지지 않도록 하겠다. [U턴\_\_001/NNG]

---

말뭉치의 분석 대상 어휘가 기호를 포함한 경우, 기호를 생략한 형태가 <우리말샘>에 등재되어 있으면 해당 어휘의미 번호를 부여한다.

---

[예시] 건물의 냉·난방비를 줄여준다. [냉·난방비\_\_001/NNG+ 를/JKO]  
국·공립대 중 1인당 장학금 수혜율 3위를 기록하고 있다. [국·공립대\_\_001/NNG]

---

## 제4장 결론

이 사업은 인공 지능의 발전을 위한 우리말 기초 자원으로서 고품질의 한국어 어휘의미 분석 말뭉치를 구축하고, 어휘의미 분석 말뭉치 구축을 위한 표준적인 지침을 개발하는 데 주요 목적이 있다.

사업의 범위는 크게 두 부분으로 나눌 수 있다. 첫째는 어휘의미 분석 말뭉치 구축 지침의 수립으로, 동형이의어를 분별한 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 다의어를 분별하는 수준에 맞게 수정하고 상세한 예를 통하여 구체화하였다. 둘째는 어휘의미 분석 말뭉치 구축으로, 새로 마련된 어휘의미 분석 말뭉치 구축 지침을 바탕으로 총 300만 어절 규모(문어 200만 어절, 구어 100만 어절)의 어휘의미 분석 말뭉치를 구축하였다.

지침 수립 과정은 <21세기 세종계획> 형태의미 분석 말뭉치 구축 지침의 미비점을 검토하여 이를 보완하고, 어휘의미 말뭉치 구축 과정에서 나타나는 실례를 추가하는 단계로 나뉜다. 본 사업에서 구축한 어휘의미 분석 말뭉치와 달리 <21세기 세종계획>의 형태의미 분석 말뭉치는 ‘다의(多意, sense)’를 분별하지 않기 때문에 지침에 이에 대한 표지가 부재하였다. 또한 분석 대상 품사의 차이로 인해 고유명사 등에 대한 세부 지침이 마련되지 않았고, 지침의 기술이 전반적으로 상세하지 않아 분석의 통일성을 담보하기 어려웠다.

본 사업에서는 <21세기 세종계획>의 형태의미 분석 말뭉치 구축 지침을 기반으로 하되 일부 내용을 수정하고 항목별 설명을 상세화하며 다양한 사례를 포함하는 방향으로 새로운 지침을 구축하였다. 우선 사전 미등재 유형의 구분을 위한 어휘의미 표지를 신설하여 <우리말샘>에 형태가 등재되어 있지 않은 경우 어휘의미 표지 ‘777’을, 형태는 있으나 해당 의미가 없는 경우 어휘의미 표지 ‘888’을 부여하였다. 더불어 말뭉치 원어절의 오타, 탈자 등 표기 오류를 위한 어휘의미 표지 ‘999’를 설정하였다.

다음으로 분석 대상 품사의 차이로 <21세기 세종계획>의 형태의미 말뭉치 구축 지침에서 고려되지 않았던 고유명사와 수사의 의미 분석과 관련된 지침을 수립하였다. 고유명사의 경우 <우리말샘>을 참조하여 최대한 어휘의미를 부여하는 것을 원칙으로 하되, 동명이인이나 동명이물 등을 의미 미등재어로 다루었다. 수사의 경우 형태 분석된 품사에 일치하는 의미 번호를 부여하였다. 품사와는 별개로 알파벳이나 숫자, 기호가 포함된

어휘에 대한 분석 방안도 마련하였다. 알파벳이나 숫자는 한글 표기로 전환하고, 기호는 생략하는 방식으로 <우리말샘>에 등재 여부를 판명하였다.

마지막으로 간략한 지침 내용을 상세화하였다. 대상 어휘의 의미를 분석할 때 전후 맥락을 최대한 활용하는 것을 원칙으로 하되, 갈래 뜻 사이에 구분이 어려운 경우 예문을 최대한 검토하여 의미를 확정하였다. 이때 한쪽에만 예문이 있는 경우 예문이 있는 갈래 뜻으로 분석하였고, 양쪽 모두 예문이 없으면 포괄적인 의미를 선택하였다.

본 사업에서 구축한 어휘의미 분석 말뭉치의 총 규모는 300만 어절로, 기사문으로 구성된 200만 어절의 문어 말뭉치와 뉴스, 인터뷰, 일상 대화 등으로 이루어진 100만 어절의 구어 말뭉치로 구성되었다. 어휘의미 분석 말뭉치의 기반이 되는 형태 분석 말뭉치는 ‘형태 분석 말뭉치 구축(2019년)’ 사업의 결과물을 어휘 분석용 말뭉치 수준으로 변환시킨 것으로, 형태 분석 말뭉치 구축 과정에서 분석된 접사 중 일부를 어근에 재통합시킨 말뭉치이다.

어휘의미 분석 말뭉치 구축은 ‘어휘 분석용 말뭉치 변화 → 분석 도구(워크벤치) 구현 → 작업 교육 → 자동 분석 → 분석 결과 수정 → 최종 결과물 산출’의 순으로 이루어졌으며, 구축 작업은 어휘의미 분석 말뭉치 구축에 최적화된 워크벤치를 개발·활용하였다. 작업자와 검수자는 워크벤치에서 자신에게 배분된 어휘를 확인하고 어휘의미 번호를 부여하였다. 어휘의미 번호 입력은 드롭다운 선택 방식을 택하여 입력 오류를 원천적으로 차단하였다.

본 사업에서는 말뭉치 구축의 편의를 도모하고 정확성을 높이기 위하여 지도학습 기반의 고성능 어휘의미 분석기(서울대 어휘의미 분석기)를 사용하였다. 서울대 어휘의미 분석기는 <우리말샘>의 의미 식별번호에 대응하는 방식으로 어휘의미를 분석하며 작업 초기 실사용 테스트에서 87.05%의 정확도를 보였다.

분석 결과 수정은 3단계로 이루어졌다. 1단계는 작업자가 자동 분석 결과를 수정하는 단계이고, 2단계는 자동 분석 결과와 작업자 수정 결과를 비교하며 검수자가 형태 분석 결과를 검수하는 단계이다. 3단계는 상위 작업자 그룹이 어휘 분석 과정에서 신고된 형태 분석의 오류를 검토하고 수정하여 의미 번호를 확정하거나 어휘 재분배를 통하여 의미 분석을 재요청하는 단계이다. 이러한 과정을 거쳐 만들어진 최종 결과물은 JSON 형식으로 변환하여 산출하였다.

<ABSTRACT>

## The Construction of a Disambiguated Tagged Corpus

This initiative aims to construct a high quality of Korean "Disambiguated Tagged Corpus" as a fundamental asset to develop AI (Artificial Intelligence) and to invent a standard guideline for the construction of "Disambiguated Tagged Corpus".

This initiative could be divided into the main 2 parts. Firstly, establishment of construction guideline to "Disambiguated Tagged Corpus", construction guideline on "Constructing Sejong Morph Sense Analysis Corpora" was modified up to a distinctive level for polysemy and specified with detailed instances.

Secondly, construction of "Disambiguated Tagged Corpus" containing totally 3 million of word-segment (2 million of written word-segment, 1 million of verbal word-segment)

The construction guideline of "Constructing Sejong Morph Sense Analysis Corpora" was utilized as a basement for construction and improvement of guidelines to "Disambiguated Tagged Corpus"

During this process, answers and examples (to FAQs) are added to guideline via the operation of the Q&A bulletin board on "Disambiguated Tagged".

Main concept and details of supplementation are as follows.

① Newly establishment of "meaning of word" to classify non-listed type in dictionary.

- In case, a form is not listed in *UrimalSaem*: '777'
- In case, a form is listed in *UrimalSaem*, but no appointed meaning: '888'
- In case of notation error in word-segment's corpus such as typos, deserters.

② Newly establishment of detailed guideline upon difference of "part of

speech” to be analyzed.

- Homonym was classified as ‘888’ regarding proper noun.
- As for numeral, number (with meaning) was assigned matched with “part of speech” to be analyzed.
- In case of alphabet and number within scope of analysis, these descriptions were checked whether it’s listed or not within *Urimalsaem* as a from of “Transcription to Korean”.
- In case of sign within scope of vocabulary to be analysis, it was checked whether it’s listed or not within *Urimalsaem* as a “Deleted Form”.

③ Adding a detailed description to existing guideline

- Maximum utilization of context was considered as a principle in meaning analysis.
- Meaning was fixed by consideration of example sentence in case, difficult to divide a meaning between senses.
- As for single sided example sentence, this was classified into sense which contains an example sentence.
- As for both sided non example sentences, inclusive sense was considered.

The volume of “Disambiguated Tagged Corpus” is total 3 million of word-segment scope being constructed upon this initiative, this volume is consisting of both 2 million of written word-segment corpus (on description) and 1 million of verbal word-segment corpus (on news, interview, daily conversation, etc.).

The “POS Tagged Korean Corpus” as the base of “Disambiguated Tagged Corpus” is the result of transcription to the level of “corpus” being used for vocabulary analysis with the output of “The Construction of POS Tagged Korean Corpus(2019)”, and it’s a corpus re-unified some analyzed-affix to root during

the process of “The Construction of POS Tagged Korean Corpus”.

The construction of “Disambiguated Tagged Corpus” was processed as follows.

Change of corpus for vocabulary analysis → Invention of analysis tool(work-bench) → Training to overhaul construction → Automatic Analysis → Correction of analysis result → Generation of final output

As for construction affairs, work-bench was developed and utilized which mostly customized to the construction of “Disambiguated Tagged Corpus”. Each operator and checker confirmed, and classified word respectively assigned them including numbering for meaning of words. Numbering and input to “meaning of word” were fundamentally blocked to input-error using the drop-down input approach.

Under this initiative, “supervised learning based” high-performing vocabulary analyzer (Seoul Nat’l Univ vocabulary analyzer) was adopted to seek convenience and accuracy in corpus construction. This analyzer showed an 87.05% accuracy in initially actual execution test with an analysis of vocabulary meaning by corresponding to identification number of Urimalsaem.

Upon analysis, the correction process was executed by 3 steps. 1st step is the correction phase to automatic analysis result by operator and 2nd step is the phase of confirming by checker upon comparison between automatic analysis result and correction by operator and finally 3rd step is the phase of fixation to numbers(with meaning upon checking and correction to error found by upper-level operator during vocabulary analysis process) and is the phase of reordering for meaning analysis by vocabulary re-assign. The final output made by this process was converted into JSON type.



## 부록 1: 어휘의미 분석 말뭉치(문어)의 예시

```
[ {
  "id": "",
  "metadata": {
    "title": "동아일보, 조선일보, 한겨레 2009~2017년 기사",
    "author": "동아일보, 조선일보, 한겨레",
    "publisher": "동아일보사, 조선일보사, 한겨레",
    "year": "2009~2017",
    "note": ""
  },
  "document": [
    {
      "id": "NWRW1800000036-0225",
      "metadata": {
        "title": "NULL",
        "author": "NULL",
        "publisher": "NULL",
        "url": "NULL",
        "date": "20120505",
        "category": "사회",
        "annotation_level": [
          "형태분석",
          "어휘의미 분석"
        ],
        "note": "NULL"
      },
      "sentence": [
        {
          "id": "NWRW1800000036-0225.1",
          "form": "우탄이는 왜 쇼를 거부했나",
          "word": [
            {"id": 1, "form": "우탄이는", "begin": 0, "end": 4},
            {"id": 2, "form": "왜", "begin": 5, "end": 6},
            {"id": 3, "form": "쇼를", "begin": 7, "end": 9},
            {"id": 4, "form": "거부했나", "begin": 10, "end": 14}
          ]
        }
      ]
    }
  ]
}
```

```

"morpheme": [
  { "id": 1, "form": "우탄이", "label": "NNP", "word_id": 1, "position": 1},
  { "id": 2, "form": "는", "label": "JX", "word_id": 1, "position": 2},
  { "id": 3, "form": "왜", "label": "MAG", "word_id": 2, "position": 1},
  { "id": 4, "form": "쇼", "label": "NNG", "word_id": 3, "position": 1},
  { "id": 5, "form": "를", "label": "JKO", "word_id": 3, "position": 2},
  { "id": 6, "form": "거부", "label": "NNG", "word_id": 4, "position": 1},
  { "id": 7, "form": "하", "label": "XSV", "word_id": 4, "position": 2},
  { "id": 8, "form": "았", "label": "EP", "word_id": 4, "position": 3},
  { "id": 9, "form": "나", "label": "EF", "word_id": 4, "position": 4}
],
"WSD": [
  { "word": "우탄이", "sense_id": 777, "begin": 0, "end": 3},
  { "word": "쇼", "sense_id": 3, "begin": 7, "end": 8},
  { "word": "거부", "sense_id": 3, "begin": 10, "end": 12}
]
},
{
  "id": "NNRW1800000036-0225.2",
  "form": "[한겨레] [토요판] 커버스토리",
  "word": [
    { "id": 1, "form": "[한겨레]", "begin": 0, "end": 5},
    { "id": 2, "form": "[토요판]", "begin": 6, "end": 11},
    { "id": 3, "form": "커버스토리", "begin": 12, "end": 17}
  ],
  "morpheme": [
    { "id": 1, "form": "[", "label": "SS", "word_id": 1, "position": 1},
    { "id": 2, "form": "한겨레", "label": "NNP", "word_id": 1, "position": 2},
    { "id": 3, "form": "]", "label": "SS", "word_id": 1, "position": 3},
    { "id": 4, "form": "[", "label": "SS", "word_id": 2, "position": 1},
    { "id": 5, "form": "토요판", "label": "NNG", "word_id": 2, "position": 2},
    { "id": 6, "form": "]", "label": "SS", "word_id": 2, "position": 3},
    { "id": 7, "form": "커버스토리", "label": "NNG", "word_id": 3, "position": 1}
  ],
  "WSD": [
    { "word": "한겨레", "sense_id": 777, "begin": 1, "end": 4},
    { "word": "토요판", "sense_id": 777, "begin": 7, "end": 10},
  ]
}

```

```

    { "word": "커버스토리", "sense_id": 1, "begin": 12, "end": 17}
  ]
},
{
  "id": "NWRW1800000036-0225.3",
  "form": "오랑우탄 '우탄이'(20살 추정 수컷)는 요즘 좁은 우리 안에서만 지낸다.",
  "word": [
    { "id": 1, "form": "오랑우탄", "begin": 0, "end": 4},
    { "id": 2, "form": "'우탄이'(20살", "begin": 5, "end": 14},
    { "id": 3, "form": "추정", "begin": 15, "end": 17},
    { "id": 4, "form": "수컷)는", "begin": 18, "end": 22},
    { "id": 5, "form": "요즘", "begin": 23, "end": 25},
    { "id": 6, "form": "좁은", "begin": 26, "end": 28},
    { "id": 7, "form": "우리", "begin": 29, "end": 31},
    { "id": 8, "form": "안에서만", "begin": 32, "end": 36},
    { "id": 9, "form": "지낸다.", "begin": 37, "end": 41}
  ],
  "morpheme": [
    { "id": 1, "form": "오랑우탄", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "'", "label": "SS", "word_id": 2, "position": 1},
    { "id": 3, "form": "우탄이", "label": "NNP", "word_id": 2, "position": 2},
    { "id": 4, "form": "'", "label": "SS", "word_id": 2, "position": 3},
    { "id": 5, "form": "(", "label": "SS", "word_id": 2, "position": 4},
    { "id": 6, "form": "20", "label": "SN", "word_id": 2, "position": 5},
    { "id": 7, "form": "살", "label": "NNB", "word_id": 2, "position": 6},
    { "id": 8, "form": "추정", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 9, "form": "수컷", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 10, "form": ")", "label": "SS", "word_id": 4, "position": 2},
    { "id": 11, "form": "는", "label": "JX", "word_id": 4, "position": 3},
    { "id": 12, "form": "요즘", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 13, "form": "좁", "label": "VA", "word_id": 6, "position": 1},
    { "id": 14, "form": "은", "label": "ETM", "word_id": 6, "position": 2},
    { "id": 15, "form": "우리", "label": "NNG", "word_id": 7, "position": 1},
    { "id": 16, "form": "안", "label": "NNG", "word_id": 8, "position": 1},
    { "id": 17, "form": "에서", "label": "JKB", "word_id": 8, "position": 2},
    { "id": 18, "form": "만", "label": "JX", "word_id": 8, "position": 3},
    { "id": 19, "form": "지내", "label": "VV", "word_id": 9, "position": 1},

```

```

    { "id": 20, "form": "ㄴ다", "label": "EF", "word_id": 9, "position": 2},
    { "id": 21, "form": ".", "label": "SF", "word_id": 9, "position": 3}
  ],
  "WSD": [
    { "word": "오랑우탄", "sense_id": 1, "begin": 0, "end": 4},
    { "word": "우탄이", "sense_id": 777, "begin": 6, "end": 9},
    { "word": "살", "sense_id": 18, "begin": 13, "end": 14},
    { "word": "추정", "sense_id": 2, "begin": 15, "end": 17},
    { "word": "수컷", "sense_id": 1, "begin": 18, "end": 20},
    { "word": "요즘", "sense_id": 1, "begin": 23, "end": 25},
    { "word": "우리", "sense_id": 1, "begin": 29, "end": 31},
    { "word": "안", "sense_id": 1, "begin": 32, "end": 33}
  ]
},
{
  "id": "NWRW1800000036-0225.4",
  "form": "우탄이는 한때 최고의 동물 스타였다.",
  "word": [
    { "id": 1, "form": "우탄이는", "begin": 0, "end": 4},
    { "id": 2, "form": "한때", "begin": 5, "end": 7},
    { "id": 3, "form": "최고의", "begin": 8, "end": 11},
    { "id": 4, "form": "동물", "begin": 12, "end": 14},
    { "id": 5, "form": "스타였다.", "begin": 15, "end": 20}
  ],
  "morpheme": [
    { "id": 1, "form": "우탄이", "label": "NNP", "word_id": 1, "position": 1},
    { "id": 2, "form": "는", "label": "JX", "word_id": 1, "position": 2},
    { "id": 3, "form": "한때", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 4, "form": "최고", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 5, "form": "의", "label": "JKG", "word_id": 3, "position": 2},
    { "id": 6, "form": "동물", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 7, "form": "스타", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 8, "form": "이", "label": "VCP", "word_id": 5, "position": 2},
    { "id": 9, "form": "였", "label": "EP", "word_id": 5, "position": 3},
    { "id": 10, "form": "다", "label": "EF", "word_id": 5, "position": 4},
    { "id": 11, "form": ".", "label": "SF", "word_id": 5, "position": 5}
  ],

```

```

"WSD": [
  { "word": "우탄이", "sense_id": 777, "begin": 0, "end": 3},
  { "word": "한때", "sense_id": 1, "begin": 5, "end": 7},
  { "word": "최고", "sense_id": 3, "begin": 8, "end": 10},
  { "word": "동물", "sense_id": 2, "begin": 12, "end": 14},
  { "word": "스타", "sense_id": 1, "begin": 15, "end": 17}
]
},
{
  "id": "NWRW1800000036-0225.5",
  "form": "에스비에스의 과 한국방송 Wu003c주주클럽Wu003e에서 사람 옷을 입고 사람 행동을 따라하자 어린이들은 열광했다.",
  "word": [
    { "id": 1, "form": "에스비에스의", "begin": 0, "end": 6},
    { "id": 2, "form": "과", "begin": 7, "end": 8},
    { "id": 3, "form": "한국방송", "begin": 9, "end": 13},
    { "id": 4, "form": "Wu003c주주클럽Wu003e에서", "begin": 14, "end": 22},
    { "id": 5, "form": "사람", "begin": 23, "end": 25},
    { "id": 6, "form": "옷을", "begin": 26, "end": 28},
    { "id": 7, "form": "입고", "begin": 29, "end": 31},
    { "id": 8, "form": "사람", "begin": 32, "end": 34},
    { "id": 9, "form": "행동을", "begin": 35, "end": 38},
    { "id": 10, "form": "따라하자", "begin": 39, "end": 43},
    { "id": 11, "form": "어린이들은", "begin": 44, "end": 49},
    { "id": 12, "form": "열광했다.", "begin": 50, "end": 55}
  ],
  "morpheme": [
    { "id": 1, "form": "에스비에스", "label": "NNP", "word_id": 1, "position": 1},
    { "id": 2, "form": "의", "label": "JKG", "word_id": 1, "position": 2},
    { "id": 3, "form": "과", "label": "NA", "word_id": 2, "position": 1},
    { "id": 4, "form": "한국방송", "label": "NNP", "word_id": 3, "position": 1},
    { "id": 5, "form": "Wu003c", "label": "SS", "word_id": 4, "position": 1},
    { "id": 6, "form": "주주클럽", "label": "NNG", "word_id": 4, "position": 2},
    { "id": 7, "form": "Wu003e", "label": "SS", "word_id": 4, "position": 3},
    { "id": 8, "form": "에서", "label": "JKB", "word_id": 4, "position": 4},
    { "id": 9, "form": "사람", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 10, "form": "옷", "label": "NNG", "word_id": 6, "position": 1},

```

```

    { "id": 11, "form": "을", "label": "JKO", "word_id": 6, "position": 2},
    { "id": 12, "form": "입", "label": "VV", "word_id": 7, "position": 1},
    { "id": 13, "form": "고", "label": "EC", "word_id": 7, "position": 2},
    { "id": 14, "form": "사람", "label": "NNG", "word_id": 8, "position": 1},
    { "id": 15, "form": "행동", "label": "NNG", "word_id": 9, "position": 1},
    { "id": 16, "form": "을", "label": "JKO", "word_id": 9, "position": 2},
    { "id": 17, "form": "따르", "label": "VV", "word_id": 10, "position": 1},
    { "id": 18, "form": "아", "label": "EC", "word_id": 10, "position": 2},
    { "id": 19, "form": "하", "label": "VV", "word_id": 10, "position": 3},
    { "id": 20, "form": "자", "label": "EC", "word_id": 10, "position": 4},
    { "id": 21, "form": "어린이", "label": "NNG", "word_id": 11, "position": 1},
    { "id": 22, "form": "들", "label": "XSN", "word_id": 11, "position": 2},
    { "id": 23, "form": "은", "label": "JX", "word_id": 11, "position": 3},
    { "id": 24, "form": "열광", "label": "NNG", "word_id": 12, "position": 1},
    { "id": 25, "form": "하", "label": "XSV", "word_id": 12, "position": 2},
    { "id": 26, "form": "았", "label": "EP", "word_id": 12, "position": 3},
    { "id": 27, "form": "다", "label": "EF", "word_id": 12, "position": 4},
    { "id": 28, "form": ".", "label": "SF", "word_id": 12, "position": 5}
  ],
  "WSD": [
    { "word": "에스비에스", "sense_id": 888, "begin": 0, "end": 5},
    { "word": "한국방송", "sense_id": 1, "begin": 9, "end": 13},
    { "word": "주주클럽", "sense_id": 777, "begin": 15, "end": 19},
    { "word": "사람", "sense_id": 1, "begin": 23, "end": 25},
    { "word": "옷", "sense_id": 1, "begin": 26, "end": 27},
    { "word": "사람", "sense_id": 1, "begin": 32, "end": 34},
    { "word": "행동", "sense_id": 1, "begin": 35, "end": 37},
    { "word": "어린이", "sense_id": 1, "begin": 44, "end": 47},
    { "word": "열광", "sense_id": 4, "begin": 50, "end": 52}
  ]
},
{
  "id": "NWRW1800000036-0225.6",
  "form": "하지만 우탄이는 언젠가부터 흥미를 잃고 짜증을 냈다.",
  "word": [
    { "id": 1, "form": "하지만", "begin": 0, "end": 3},
    { "id": 2, "form": "우탄이는", "begin": 4, "end": 8},

```

```

    { "id": 3, "form": "언젠가부터", "begin": 9, "end": 14},
    { "id": 4, "form": "흥미를", "begin": 15, "end": 18},
    { "id": 5, "form": "잃고", "begin": 19, "end": 21},
    { "id": 6, "form": "짜증을", "begin": 22, "end": 25},
    { "id": 7, "form": "냈다.", "begin": 26, "end": 29}
  ],
  "morpheme": [
    { "id": 1, "form": "하지만", "label": "MAJ", "word_id": 1, "position": 1},
    { "id": 2, "form": "우탄이", "label": "NNP", "word_id": 2, "position": 1},
    { "id": 3, "form": "는", "label": "JX", "word_id": 2, "position": 2},
    { "id": 4, "form": "언젠가", "label": "MAG", "word_id": 3, "position": 1},
    { "id": 5, "form": "부터", "label": "JX", "word_id": 3, "position": 2},
    { "id": 6, "form": "흥미", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 7, "form": "를", "label": "JKO", "word_id": 4, "position": 2},
    { "id": 8, "form": "잃", "label": "VV", "word_id": 5, "position": 1},
    { "id": 9, "form": "고", "label": "EC", "word_id": 5, "position": 2},
    { "id": 10, "form": "짜증", "label": "NNG", "word_id": 6, "position": 1},
    { "id": 11, "form": "을", "label": "JKO", "word_id": 6, "position": 2},
    { "id": 12, "form": "내", "label": "VV", "word_id": 7, "position": 1},
    { "id": 13, "form": "였", "label": "EP", "word_id": 7, "position": 2},
    { "id": 14, "form": "다", "label": "EF", "word_id": 7, "position": 3},
    { "id": 15, "form": ".", "label": "SF", "word_id": 7, "position": 4}
  ],
  "WSD": [
    { "word": "우탄이", "sense_id": 777, "begin": 4, "end": 7},
    { "word": "흥미", "sense_id": 1, "begin": 15, "end": 17},
    { "word": "짜증", "sense_id": 1, "begin": 22, "end": 24}
  ]
},
{
  "id": "NWRW1800000036-0225.7",
  "form": "작고 귀엽던 몸집은 커졌고 사람들도 우탄이를 예전만큼 찾지 않았다.",
  "word": [
    { "id": 1, "form": "작고", "begin": 0, "end": 2},
    { "id": 2, "form": "귀엽던", "begin": 3, "end": 6},
    { "id": 3, "form": "몸집은", "begin": 7, "end": 10},
    { "id": 4, "form": "커졌고", "begin": 11, "end": 14},

```

```

    { "id": 5, "form": "사람들도", "begin": 15, "end": 19},
    { "id": 6, "form": "우탄이를", "begin": 20, "end": 24},
    { "id": 7, "form": "예전만큼", "begin": 25, "end": 29},
    { "id": 8, "form": "찾지", "begin": 30, "end": 32},
    { "id": 9, "form": "않았다.", "begin": 33, "end": 37}
  ],
  "morpheme": [
    { "id": 1, "form": "작", "label": "VA", "word_id": 1, "position": 1},
    { "id": 2, "form": "고", "label": "EC", "word_id": 1, "position": 2},
    { "id": 3, "form": "귀엽", "label": "VA", "word_id": 2, "position": 1},
    { "id": 4, "form": "던", "label": "ETM", "word_id": 2, "position": 2},
    { "id": 5, "form": "몸집", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 6, "form": "은", "label": "JX", "word_id": 3, "position": 2},
    { "id": 7, "form": "커지", "label": "VV", "word_id": 4, "position": 1},
    { "id": 8, "form": "었", "label": "EP", "word_id": 4, "position": 2},
    { "id": 9, "form": "고", "label": "EC", "word_id": 4, "position": 3},
    { "id": 10, "form": "사람", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 11, "form": "들", "label": "XSN", "word_id": 5, "position": 2},
    { "id": 12, "form": "도", "label": "JX", "word_id": 5, "position": 3},
    { "id": 13, "form": "우탄이", "label": "NNP", "word_id": 6, "position": 1},
    { "id": 14, "form": "를", "label": "JKO", "word_id": 6, "position": 2},
    { "id": 15, "form": "예전", "label": "NNG", "word_id": 7, "position": 1},
    { "id": 16, "form": "만큼", "label": "JKB", "word_id": 7, "position": 2},
    { "id": 17, "form": "찾", "label": "VV", "word_id": 8, "position": 1},
    { "id": 18, "form": "지", "label": "EC", "word_id": 8, "position": 2},
    { "id": 19, "form": "않", "label": "VX", "word_id": 9, "position": 1},
    { "id": 20, "form": "았", "label": "EP", "word_id": 9, "position": 2},
    { "id": 21, "form": "다", "label": "EF", "word_id": 9, "position": 3},
    { "id": 22, "form": ".", "label": "SF", "word_id": 9, "position": 4}
  ],
  "WSD": [
    { "word": "몸집", "sense_id": 1, "begin": 7, "end": 9},
    { "word": "사람", "sense_id": 1, "begin": 15, "end": 17},
    { "word": "우탄이", "sense_id": 777, "begin": 20, "end": 23},
    { "word": "예전", "sense_id": 1, "begin": 25, "end": 27}
  ]
},

```



```

{
  "id": "NWRW1800000036-0225.8",
  "form": "한국은 동남아시아에 이은 동물쇼의 천국이다.",
  "word": [
    { "id": 1, "form": "한국은", "begin": 0, "end": 3},
    { "id": 2, "form": "동남아시아에", "begin": 4, "end": 10},
    { "id": 3, "form": "이은", "begin": 11, "end": 13},
    { "id": 4, "form": "동물쇼의", "begin": 14, "end": 18},
    { "id": 5, "form": "천국이다.", "begin": 19, "end": 24}
  ],
  "morpheme": [
    { "id": 1, "form": "한국", "label": "NNP", "word_id": 1, "position": 1},
    { "id": 2, "form": "은", "label": "JX", "word_id": 1, "position": 2},
    { "id": 3, "form": "동남아시아", "label": "NNP", "word_id": 2, "position": 1},
    { "id": 4, "form": "에", "label": "JKB", "word_id": 2, "position": 2},
    { "id": 5, "form": "잇", "label": "VV", "word_id": 3, "position": 1},
    { "id": 6, "form": "은", "label": "ETM", "word_id": 3, "position": 2},
    { "id": 7, "form": "동물", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 8, "form": "쇼", "label": "NNG", "word_id": 4, "position": 2},
    { "id": 9, "form": "의", "label": "JKG", "word_id": 4, "position": 3},
    { "id": 10, "form": "천국", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 11, "form": "이", "label": "VCP", "word_id": 5, "position": 2},
    { "id": 12, "form": "다", "label": "EF", "word_id": 5, "position": 3},
    { "id": 13, "form": ".", "label": "SF", "word_id": 5, "position": 4}
  ],
  "WSD": [
    { "word": "한국", "sense_id": 7, "begin": 0, "end": 2},
    { "word": "동남아시아", "sense_id": 1, "begin": 4, "end": 9},
    { "word": "동물", "sense_id": 2, "begin": 14, "end": 16},
    { "word": "쇼", "sense_id": 3, "begin": 16, "end": 17},
    { "word": "천국", "sense_id": 2, "begin": 19, "end": 21}
  ]
},
{

```

```

  "id": "NWRW1800000036-0225.9",
  "form": "한국동물원수족관협회(KAZA)에 등록된 19개 회원사 가운데 어린이대공원(서울 능
동), 삼성에버랜드동물원(경기도 용인시), 퍼시픽랜드(제주도 서귀포시), 제주동물원(경기도 고양시), 한화

```

63시티(씨월드·서울 여의도) 등 5개 회원사에서 동물쇼를 벌이고 있다.",

```
"word": [  
  { "id": 1, "form": "한국동물원수족관협회(KAZA)에", "begin": 0, "end": 17},  
  { "id": 2, "form": "등록된", "begin": 18, "end": 21},  
  { "id": 3, "form": "19개", "begin": 22, "end": 25},  
  { "id": 4, "form": "회원사", "begin": 26, "end": 29},  
  { "id": 5, "form": "가운데", "begin": 30, "end": 33},  
  { "id": 6, "form": "어린이대공원(서울", "begin": 34, "end": 43},  
  { "id": 7, "form": "능동)", "begin": 44, "end": 48},  
  { "id": 8, "form": "삼성에버랜드동물원(경기도", "begin": 49, "end": 62},  
  { "id": 9, "form": "용인시)", "begin": 63, "end": 68},  
  { "id": 10, "form": "퍼시픽랜드(제주도", "begin": 69, "end": 78},  
  { "id": 11, "form": "서귀포시)", "begin": 79, "end": 85},  
  { "id": 12, "form": "주주동물원(경기도", "begin": 86, "end": 95},  
  { "id": 13, "form": "고양시)", "begin": 96, "end": 101},  
  { "id": 14, "form": "한화", "begin": 102, "end": 104},  
  { "id": 15, "form": "63시티(씨월드·서울", "begin": 105, "end": 116},  
  { "id": 16, "form": "여의도)", "begin": 117, "end": 121},  
  { "id": 17, "form": "등", "begin": 122, "end": 123},  
  { "id": 18, "form": "5개", "begin": 124, "end": 126},  
  { "id": 19, "form": "회원사에서", "begin": 127, "end": 132},  
  { "id": 20, "form": "동물쇼를", "begin": 133, "end": 137},  
  { "id": 21, "form": "벌이고", "begin": 138, "end": 141},  
  { "id": 22, "form": "있다.", "begin": 142, "end": 145}  
],
```

```
"morpheme": [  
  { "id": 1, "form": "한국동물원수족관협회", "label": "NNP", "word_id": 1, "position": 1},  
  { "id": 2, "form": "(", "label": "SS", "word_id": 1, "position": 2},  
  { "id": 3, "form": "KAZA", "label": "SL", "word_id": 1, "position": 3},  
  { "id": 4, "form": ")", "label": "SS", "word_id": 1, "position": 4},  
  { "id": 5, "form": "에", "label": "JKB", "word_id": 1, "position": 5},  
  { "id": 6, "form": "등록", "label": "NNG", "word_id": 2, "position": 1},  
  { "id": 7, "form": "되", "label": "XSV", "word_id": 2, "position": 2},  
  { "id": 8, "form": "ㄴ", "label": "ETM", "word_id": 2, "position": 3},  
  { "id": 9, "form": "19", "label": "SN", "word_id": 3, "position": 1},  
  { "id": 10, "form": "개", "label": "NNB", "word_id": 3, "position": 2},  
  { "id": 11, "form": "회원사", "label": "NNG", "word_id": 4, "position": 1},
```

{ "id": 12, "form": "카운테", "label": "NNG", "word\_id": 5, "position": 1},  
 { "id": 13, "form": "어린이대공원", "label": "NNP", "word\_id": 6, "position": 1},  
 { "id": 14, "form": "(", "label": "SS", "word\_id": 6, "position": 2},  
 { "id": 15, "form": "서울", "label": "NNP", "word\_id": 6, "position": 3},  
 { "id": 16, "form": "능동", "label": "NNP", "word\_id": 7, "position": 1},  
 { "id": 17, "form": ")", "label": "SS", "word\_id": 7, "position": 2},  
 { "id": 18, "form": ",", "label": "SP", "word\_id": 7, "position": 3},  
 { "id": 19, "form": "삼성에버랜드동물원", "label": "NNP", "word\_id": 8, "position": 1},  
 { "id": 20, "form": "(", "label": "SS", "word\_id": 8, "position": 2},  
 { "id": 21, "form": "경기도", "label": "NNP", "word\_id": 8, "position": 3},  
 { "id": 22, "form": "용인시", "label": "NNP", "word\_id": 9, "position": 1},  
 { "id": 23, "form": ")", "label": "SS", "word\_id": 9, "position": 2},  
 { "id": 24, "form": ",", "label": "SP", "word\_id": 9, "position": 3},  
 { "id": 25, "form": "퍼시픽랜드", "label": "NNP", "word\_id": 10, "position": 1},  
 { "id": 26, "form": "(", "label": "SS", "word\_id": 10, "position": 2},  
 { "id": 27, "form": "제주도", "label": "NNP", "word\_id": 10, "position": 3},  
 { "id": 28, "form": "서귀포시", "label": "NNP", "word\_id": 11, "position": 1},  
 { "id": 29, "form": ")", "label": "SS", "word\_id": 11, "position": 2},  
 { "id": 30, "form": ",", "label": "SP", "word\_id": 11, "position": 3},  
 { "id": 31, "form": "주주동물원", "label": "NNP", "word\_id": 12, "position": 1},  
 { "id": 32, "form": "(", "label": "SS", "word\_id": 12, "position": 2},  
 { "id": 33, "form": "경기도", "label": "NNP", "word\_id": 12, "position": 3},  
 { "id": 34, "form": "고양시", "label": "NNP", "word\_id": 13, "position": 1},  
 { "id": 35, "form": ")", "label": "SS", "word\_id": 13, "position": 2},  
 { "id": 36, "form": ",", "label": "SP", "word\_id": 13, "position": 3},  
 { "id": 37, "form": "한화", "label": "NNP", "word\_id": 14, "position": 1},  
 { "id": 38, "form": "63시티", "label": "NNP", "word\_id": 15, "position": 1},  
 { "id": 39, "form": "(", "label": "SS", "word\_id": 15, "position": 2},  
 { "id": 40, "form": "씨월드", "label": "NNP", "word\_id": 15, "position": 3},  
 { "id": 41, "form": ".", "label": "SP", "word\_id": 15, "position": 4},  
 { "id": 42, "form": "서울", "label": "NNP", "word\_id": 15, "position": 5},  
 { "id": 43, "form": "여의도", "label": "NNP", "word\_id": 16, "position": 1},  
 { "id": 44, "form": ")", "label": "SS", "word\_id": 16, "position": 2},  
 { "id": 45, "form": "등", "label": "NNB", "word\_id": 17, "position": 1},  
 { "id": 46, "form": "5", "label": "SN", "word\_id": 18, "position": 1},  
 { "id": 47, "form": "개", "label": "NNB", "word\_id": 18, "position": 2},  
 { "id": 48, "form": "회원사", "label": "NNG", "word\_id": 19, "position": 1},

```

    { "id": 49, "form": "에서", "label": "JKB", "word_id": 19, "position": 2},
    { "id": 50, "form": "동물", "label": "NNG", "word_id": 20, "position": 1},
    { "id": 51, "form": "쇼", "label": "NNG", "word_id": 20, "position": 2},
    { "id": 52, "form": "를", "label": "JKO", "word_id": 20, "position": 3},
    { "id": 53, "form": "별이", "label": "VV", "word_id": 21, "position": 1},
    { "id": 54, "form": "고", "label": "EC", "word_id": 21, "position": 2},
    { "id": 55, "form": "있", "label": "VX", "word_id": 22, "position": 1},
    { "id": 56, "form": "다", "label": "EF", "word_id": 22, "position": 2},
    { "id": 57, "form": ".", "label": "SF", "word_id": 22, "position": 3}
  ],
  "WSD": [
    { "word": "한국동물원수족관협회", "sense_id": 777, "begin": 0, "end": 10},
    { "word": "등록", "sense_id": 1, "begin": 18, "end": 20},
    { "word": "개", "sense_id": 13, "begin": 24, "end": 25},
    { "word": "회원사", "sense_id": 1, "begin": 26, "end": 29},
    { "word": "가운데", "sense_id": 3, "begin": 30, "end": 33},
    { "word": "어린이대공원", "sense_id": 1, "begin": 34, "end": 40},
    { "word": "서울", "sense_id": 2, "begin": 41, "end": 43},
    { "word": "능동", "sense_id": 888, "begin": 44, "end": 46},
    { "word": "삼성에버랜드동물원", "sense_id": 777, "begin": 49, "end": 58},
    { "word": "경기도", "sense_id": 1, "begin": 59, "end": 62},
    { "word": "용인시", "sense_id": 1, "begin": 63, "end": 66},
    { "word": "퍼시픽랜드", "sense_id": 777, "begin": 69, "end": 74},
    { "word": "제주도", "sense_id": 1, "begin": 75, "end": 78},
    { "word": "서귀포시", "sense_id": 1, "begin": 79, "end": 83},
    { "word": "주주동물원", "sense_id": 777, "begin": 86, "end": 91},
    { "word": "경기도", "sense_id": 1, "begin": 92, "end": 95},
    { "word": "고양시", "sense_id": 1, "begin": 96, "end": 99},
    { "word": "한화", "sense_id": 888, "begin": 102, "end": 104},
    { "word": "63시티", "sense_id": 777, "begin": 105, "end": 109},
    { "word": "씨월드", "sense_id": 777, "begin": 110, "end": 113},
    { "word": "서울", "sense_id": 2, "begin": 114, "end": 116},
    { "word": "여의도", "sense_id": 1, "begin": 117, "end": 120},
    { "word": "등", "sense_id": 11, "begin": 122, "end": 123},
    { "word": "개", "sense_id": 13, "begin": 125, "end": 126},
    { "word": "회원사", "sense_id": 1, "begin": 127, "end": 130},
    { "word": "동물", "sense_id": 2, "begin": 133, "end": 135},
  ]

```

```

    { "word": "쇼", "sense_id": 3, "begin": 135, "end": 136}
  ]
},
{
  "id": "NWRW1800000036-0225.10",
  "form": "어린이대공원에서 하루 5~7번 원숭이·물개·백로·황조롱이 등 15종이 출연하는 멀
티쇼가 열린다.",
  "word": [
    { "id": 1, "form": "어린이대공원에서", "begin": 0, "end": 9},
    { "id": 2, "form": "하루", "begin": 10, "end": 12},
    { "id": 3, "form": "5~7번", "begin": 13, "end": 17},
    { "id": 4, "form": "원숭이·물개·백로·황조롱이", "begin": 18, "end": 32},
    { "id": 5, "form": "등", "begin": 33, "end": 34},
    { "id": 6, "form": "15종이", "begin": 35, "end": 39},
    { "id": 7, "form": "출연하는", "begin": 40, "end": 44},
    { "id": 8, "form": "멀티쇼가", "begin": 45, "end": 49},
    { "id": 9, "form": "열린다.", "begin": 50, "end": 54}
  ],
  "morpheme": [
    { "id": 1, "form": "어린이대공원", "label": "NNP", "word_id": 1, "position": 1},
    { "id": 2, "form": "에서", "label": "JKB", "word_id": 1, "position": 2},
    { "id": 3, "form": "는", "label": "JX", "word_id": 1, "position": 3},
    { "id": 4, "form": "하루", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 5, "form": "5", "label": "SN", "word_id": 3, "position": 1},
    { "id": 6, "form": "~", "label": "SO", "word_id": 3, "position": 2},
    { "id": 7, "form": "7", "label": "SN", "word_id": 3, "position": 3},
    { "id": 8, "form": "번", "label": "NNB", "word_id": 3, "position": 4},
    { "id": 9, "form": "원숭이", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 10, "form": ".", "label": "SP", "word_id": 4, "position": 2},
    { "id": 11, "form": "물개", "label": "NNG", "word_id": 4, "position": 3},
    { "id": 12, "form": ".", "label": "SP", "word_id": 4, "position": 4},
    { "id": 13, "form": "백로", "label": "NNG", "word_id": 4, "position": 5},
    { "id": 14, "form": ".", "label": "SP", "word_id": 4, "position": 6},
    { "id": 15, "form": "황조롱이", "label": "NNG", "word_id": 4, "position": 7},
    { "id": 16, "form": "등", "label": "NNB", "word_id": 5, "position": 1},
    { "id": 17, "form": "15", "label": "SN", "word_id": 6, "position": 1},
    { "id": 18, "form": "종", "label": "NNG", "word_id": 6, "position": 2},

```

```

    { "id": 19, "form": "이", "label": "JKS", "word_id": 6, "position": 3},
    { "id": 20, "form": "출연", "label": "NNG", "word_id": 7, "position": 1},
    { "id": 21, "form": "하", "label": "XSV", "word_id": 7, "position": 2},
    { "id": 22, "form": "는", "label": "ETM", "word_id": 7, "position": 3},
    { "id": 23, "form": "멀티쇼", "label": "NNG", "word_id": 8, "position": 1},
    { "id": 24, "form": "가", "label": "JKS", "word_id": 8, "position": 2},
    { "id": 25, "form": "열리", "label": "VV", "word_id": 9, "position": 1},
    { "id": 26, "form": "니다", "label": "EF", "word_id": 9, "position": 2},
    { "id": 27, "form": ".", "label": "SF", "word_id": 9, "position": 3}
  ],

```

```

  "WSD": [
    { "word": "어린이대공원", "sense_id": 1, "begin": 0, "end": 6},
    { "word": "하루", "sense_id": 2, "begin": 10, "end": 12},
    { "word": "번", "sense_id": 6, "begin": 16, "end": 17},
    { "word": "원숭이", "sense_id": 1, "begin": 18, "end": 21},
    { "word": "물개", "sense_id": 1, "begin": 22, "end": 24},
    { "word": "백로", "sense_id": 3, "begin": 25, "end": 27},
    { "word": "황조롱이", "sense_id": 1, "begin": 28, "end": 32},
    { "word": "등", "sense_id": 10, "begin": 33, "end": 34},
    { "word": "중", "sense_id": 15, "begin": 37, "end": 38},
    { "word": "출연", "sense_id": 3, "begin": 40, "end": 42},
    { "word": "멀티쇼", "sense_id": 777, "begin": 45, "end": 48}
  ]

```

```

},
{

```

```

  "id": "NWRW1800000036-0225.11",

```

"form": "삼성에버랜드동물원에서는 물개쇼와 공작 등 조류 7종 500마리가 출연하는 새공연, 오랑우탄·흰손긴팔원숭이 등 41종 140여마리가 출연하는 멀티쇼 등 3개의 공연이 벌어진다.",

```

  "word": [
    { "id": 1, "form": "삼성에버랜드동물원에서는", "begin": 0, "end": 12},
    { "id": 2, "form": "물개쇼와", "begin": 13, "end": 17},
    { "id": 3, "form": "공작", "begin": 18, "end": 20},
    { "id": 4, "form": "등", "begin": 21, "end": 22},
    { "id": 5, "form": "조류", "begin": 23, "end": 25},
    { "id": 6, "form": "7종", "begin": 26, "end": 28},
    { "id": 7, "form": "500마리가", "begin": 29, "end": 35},
    { "id": 8, "form": "출연하는", "begin": 36, "end": 40},
  ]

```

```

    { "id": 9, "form": "새공연", "begin": 41, "end": 45},
    { "id": 10, "form": "오랑우탄·흰손긴팔원숭이", "begin": 46, "end": 58},
    { "id": 11, "form": "등", "begin": 59, "end": 60},
    { "id": 12, "form": "41종", "begin": 61, "end": 64},
    { "id": 13, "form": "140여마리가", "begin": 65, "end": 72},
    { "id": 14, "form": "출연하는", "begin": 73, "end": 77},
    { "id": 15, "form": "멀티쇼", "begin": 78, "end": 81},
    { "id": 16, "form": "등", "begin": 82, "end": 83},
    { "id": 17, "form": "3개의", "begin": 84, "end": 87},
    { "id": 18, "form": "공연이", "begin": 88, "end": 91},
    { "id": 19, "form": "벌어진다.", "begin": 92, "end": 97}
  ],
  "morpheme": [
    { "id": 1, "form": "삼성에버랜드동물원", "label": "NNP", "word_id": 1, "position": 1},
    { "id": 2, "form": "에서", "label": "JKB", "word_id": 1, "position": 2},
    { "id": 3, "form": "는", "label": "JX", "word_id": 1, "position": 3},
    { "id": 4, "form": "물개", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 5, "form": "쇼", "label": "NNG", "word_id": 2, "position": 2},
    { "id": 6, "form": "와", "label": "JC", "word_id": 2, "position": 3},
    { "id": 7, "form": "공작", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 8, "form": "등", "label": "NNB", "word_id": 4, "position": 1},
    { "id": 9, "form": "조류", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 10, "form": "7", "label": "SN", "word_id": 6, "position": 1},
    { "id": 11, "form": "종", "label": "NNG", "word_id": 6, "position": 2},
    { "id": 12, "form": "500", "label": "SN", "word_id": 7, "position": 1},
    { "id": 13, "form": "마리", "label": "NNB", "word_id": 7, "position": 2},
    { "id": 14, "form": "가", "label": "JKS", "word_id": 7, "position": 3},
    { "id": 15, "form": "출연", "label": "NNG", "word_id": 8, "position": 1},
    { "id": 16, "form": "하", "label": "XSV", "word_id": 8, "position": 2},
    { "id": 17, "form": "는", "label": "ETM", "word_id": 8, "position": 3},
    { "id": 18, "form": "새", "label": "NNG", "word_id": 9, "position": 1},
    { "id": 19, "form": "공연", "label": "NNG", "word_id": 9, "position": 2},
    { "id": 20, "form": ",", "label": "SP", "word_id": 9, "position": 3},
    { "id": 21, "form": "오랑우탄", "label": "NNG", "word_id": 10, "position": 1},
    { "id": 22, "form": ".", "label": "SP", "word_id": 10, "position": 2},
    { "id": 23, "form": "흰손긴팔원숭이", "label": "NNG", "word_id": 10, "position": 3},
    { "id": 24, "form": "등", "label": "NNB", "word_id": 11, "position": 1},

```

```

    { "id": 25, "form": "41", "label": "SN", "word_id": 12, "position": 1},
    { "id": 26, "form": "중", "label": "NNG", "word_id": 12, "position": 2},
    { "id": 27, "form": "140", "label": "SN", "word_id": 13, "position": 1},
    { "id": 28, "form": "여", "label": "XSN", "word_id": 13, "position": 2},
    { "id": 29, "form": "마리", "label": "NNB", "word_id": 13, "position": 3},
    { "id": 30, "form": "가", "label": "JKS", "word_id": 13, "position": 4},
    { "id": 31, "form": "출연", "label": "NNG", "word_id": 14, "position": 1},
    { "id": 32, "form": "하", "label": "XSV", "word_id": 14, "position": 2},
    { "id": 33, "form": "는", "label": "ETM", "word_id": 14, "position": 3},
    { "id": 34, "form": "멀티쇼", "label": "NNG", "word_id": 15, "position": 1},
    { "id": 35, "form": "등", "label": "NNB", "word_id": 16, "position": 1},
    { "id": 36, "form": "3", "label": "SN", "word_id": 17, "position": 1},
    { "id": 37, "form": "개", "label": "NNB", "word_id": 17, "position": 2},
    { "id": 38, "form": "의", "label": "JKG", "word_id": 17, "position": 3},
    { "id": 39, "form": "공연", "label": "NNG", "word_id": 18, "position": 1},
    { "id": 40, "form": "이", "label": "JKS", "word_id": 18, "position": 2},
    { "id": 41, "form": "벌어지", "label": "VV", "word_id": 19, "position": 1},
    { "id": 42, "form": "나다", "label": "EF", "word_id": 19, "position": 2},
    { "id": 43, "form": ".", "label": "SF", "word_id": 19, "position": 3}
  ],
  "WSD": [

```

```

    { "word": "삼성에버랜드동물원", "sense_id": 777, "begin": 0, "end": 9},
    { "word": "물개", "sense_id": 1, "begin": 13, "end": 15},
    { "word": "쇼", "sense_id": 3, "begin": 15, "end": 16},
    { "word": "공작", "sense_id": 7, "begin": 18, "end": 20},
    { "word": "등", "sense_id": 10, "begin": 21, "end": 22},
    { "word": "조류", "sense_id": 1, "begin": 23, "end": 25},
    { "word": "중", "sense_id": 15, "begin": 27, "end": 28},
    { "word": "마리", "sense_id": 1, "begin": 32, "end": 34},
    { "word": "출연", "sense_id": 3, "begin": 36, "end": 38},
    { "word": "새", "sense_id": 6, "begin": 41, "end": 42},
    { "word": "공연", "sense_id": 2, "begin": 42, "end": 44},
    { "word": "오랑우탄", "sense_id": 1, "begin": 46, "end": 50},
    { "word": "흰손긴팔원숭이", "sense_id": 777, "begin": 51, "end": 58},
    { "word": "등", "sense_id": 10, "begin": 59, "end": 60},
    { "word": "중", "sense_id": 15, "begin": 63, "end": 64},
    { "word": "마리", "sense_id": 1, "begin": 69, "end": 71},

```



```

    { "word": "출연", "sense_id": 3, "begin": 73, "end": 75},
    { "word": "멀티쇼", "sense_id": 777, "begin": 78, "end": 81},
    { "word": "등", "sense_id": 11, "begin": 82, "end": 83},
    { "word": "개", "sense_id": 13, "begin": 85, "end": 86},
    { "word": "공연", "sense_id": 2, "begin": 88, "end": 90}
  ]
},
{
  "id": "NWRW1800000036-0225.12",
  "form": "제주도 서귀포시 점보빌리지의 코끼리쇼 등 협회 비회원사가 벌이는 쇼도 3개다.",
  "word": [
    { "id": 1, "form": "제주도", "begin": 0, "end": 3},
    { "id": 2, "form": "서귀포시", "begin": 4, "end": 8},
    { "id": 3, "form": "점보빌리지의", "begin": 9, "end": 15},
    { "id": 4, "form": "코끼리쇼", "begin": 16, "end": 20},
    { "id": 5, "form": "등", "begin": 21, "end": 22},
    { "id": 6, "form": "협회", "begin": 23, "end": 25},
    { "id": 7, "form": "비회원사가", "begin": 26, "end": 31},
    { "id": 8, "form": "벌이는", "begin": 32, "end": 35},
    { "id": 9, "form": "쇼도", "begin": 36, "end": 38},
    { "id": 10, "form": "3개다.", "begin": 39, "end": 43}
  ],
  "morpheme": [
    { "id": 1, "form": "제주도", "label": "NNP", "word_id": 1, "position": 1},
    { "id": 2, "form": "서귀포시", "label": "NNP", "word_id": 2, "position": 1},
    { "id": 3, "form": "점보빌리지", "label": "NNP", "word_id": 3, "position": 1},
    { "id": 4, "form": "의", "label": "JKG", "word_id": 3, "position": 2},
    { "id": 5, "form": "코끼리", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 6, "form": "쇼", "label": "NNG", "word_id": 4, "position": 2},
    { "id": 7, "form": "등", "label": "NNB", "word_id": 5, "position": 1},
    { "id": 8, "form": "협회", "label": "NNG", "word_id": 6, "position": 1},
    { "id": 9, "form": "비회원사", "label": "NNG", "word_id": 7, "position": 1},
    { "id": 10, "form": "가", "label": "JKS", "word_id": 7, "position": 2},
    { "id": 11, "form": "벌이", "label": "VV", "word_id": 8, "position": 1},
    { "id": 12, "form": "는", "label": "ETM", "word_id": 8, "position": 2},
    { "id": 13, "form": "쇼", "label": "NNG", "word_id": 9, "position": 1},
    { "id": 14, "form": "도", "label": "JX", "word_id": 9, "position": 2},
  ]
}

```

```

    { "id": 15, "form": "3", "label": "SN", "word_id": 10, "position": 1},
    { "id": 16, "form": "개", "label": "NNB", "word_id": 10, "position": 2},
    { "id": 17, "form": "이", "label": "VCP", "word_id": 10, "position": 3},
    { "id": 18, "form": "다", "label": "EF", "word_id": 10, "position": 4},
    { "id": 19, "form": ".", "label": "SF", "word_id": 10, "position": 5}
  ],

```

```

"WSD": [
  { "word": "제주도", "sense_id": 1, "begin": 0, "end": 3},
  { "word": "서귀포시", "sense_id": 1, "begin": 4, "end": 8},
  { "word": "점보빌리지", "sense_id": 777, "begin": 9, "end": 14},
  { "word": "코끼리", "sense_id": 1, "begin": 16, "end": 19},
  { "word": "쇼", "sense_id": 3, "begin": 19, "end": 20},
  { "word": "등", "sense_id": 10, "begin": 21, "end": 22},
  { "word": "협회", "sense_id": 1, "begin": 23, "end": 25},
  { "word": "비회원사", "sense_id": 777, "begin": 26, "end": 30},
  { "word": "쇼", "sense_id": 3, "begin": 36, "end": 37},
  { "word": "개", "sense_id": 13, "begin": 40, "end": 41}
]

```

```

},
{

```

```

  "id": "NWRW1800000036-0225.13",

```

```

  "form": "동물쇼는 동물학대 측면이 조명되면서, 세계적으로 ‘나쁜 관광’ 논란에 휩싸이고 있
다.",

```

```

"word": [
  { "id": 1, "form": "동물쇼는", "begin": 0, "end": 4},
  { "id": 2, "form": "동물학대", "begin": 5, "end": 9},
  { "id": 3, "form": "측면이", "begin": 10, "end": 13},
  { "id": 4, "form": "조명되면서,", "begin": 14, "end": 20},
  { "id": 5, "form": "세계적으로", "begin": 21, "end": 26},
  { "id": 6, "form": "'나쁜", "begin": 27, "end": 30},
  { "id": 7, "form": "관광'", "begin": 31, "end": 34},
  { "id": 8, "form": "논란에", "begin": 35, "end": 38},
  { "id": 9, "form": "휩싸이고", "begin": 39, "end": 43},
  { "id": 10, "form": "있다.", "begin": 44, "end": 47}
]

```

```

],

```

```

"morpheme": [

```

```

  { "id": 1, "form": "동물", "label": "NNG", "word_id": 1, "position": 1},

```

```

    { "id": 2, "form": "쇼", "label": "NNG", "word_id": 1, "position": 2},
    { "id": 3, "form": "늑", "label": "JX", "word_id": 1, "position": 3},
    { "id": 4, "form": "동물", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 5, "form": "학대", "label": "NNG", "word_id": 2, "position": 2},
    { "id": 6, "form": "측면", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 7, "form": "이", "label": "JKS", "word_id": 3, "position": 2},
    { "id": 8, "form": "조명", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 9, "form": "되", "label": "XSV", "word_id": 4, "position": 2},
    { "id": 10, "form": "면서", "label": "EC", "word_id": 4, "position": 3},
    { "id": 11, "form": ",", "label": "SP", "word_id": 4, "position": 4},
    { "id": 12, "form": "세계적", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 13, "form": "으로", "label": "JKB", "word_id": 5, "position": 2},
    { "id": 14, "form": "", "label": "SS", "word_id": 6, "position": 1},
    { "id": 15, "form": "나쁘", "label": "VA", "word_id": 6, "position": 2},
    { "id": 16, "form": "ㄴ", "label": "ETM", "word_id": 6, "position": 3},
    { "id": 17, "form": "관광", "label": "NNG", "word_id": 7, "position": 1},
    { "id": 18, "form": "", "label": "SS", "word_id": 7, "position": 2},
    { "id": 19, "form": "논란", "label": "NNG", "word_id": 8, "position": 1},
    { "id": 20, "form": "에", "label": "JKB", "word_id": 8, "position": 2},
    { "id": 21, "form": "휩싸이", "label": "VV", "word_id": 9, "position": 1},
    { "id": 22, "form": "고", "label": "EC", "word_id": 9, "position": 2},
    { "id": 23, "form": "있", "label": "VX", "word_id": 10, "position": 1},
    { "id": 24, "form": "다", "label": "EF", "word_id": 10, "position": 2},
    { "id": 25, "form": ".", "label": "SF", "word_id": 10, "position": 3}
  ],
  "WSD": [
    { "word": "동물", "sense_id": 2, "begin": 0, "end": 2},
    { "word": "쇼", "sense_id": 3, "begin": 2, "end": 3},
    { "word": "동물", "sense_id": 2, "begin": 5, "end": 7},
    { "word": "학대", "sense_id": 1, "begin": 7, "end": 9},
    { "word": "측면", "sense_id": 2, "begin": 10, "end": 12},
    { "word": "조명", "sense_id": 7, "begin": 14, "end": 16},
    { "word": "세계적", "sense_id": 1, "begin": 21, "end": 24},
    { "word": "관광", "sense_id": 3, "begin": 31, "end": 33},
    { "word": "논란", "sense_id": 1, "begin": 35, "end": 37}
  ]
},

```

## 부록 2: 어휘의미 분석 말뭉치(구어)의 예시

```
[ {
  "id": "SRBK00123",
  "metadata": {
    "title": "최고의 요리비결 2",
    "author": "윤형빈, 임미자",
    "publisher": "EBS",
    "year": "2014",
    "note": ""
  },
  "document": [
    {
      "id": "SRBK00123",
      "metadata": {
        "title": "NULL",
        "author": "NULL",
        "publisher": "NULL",
        "url": "NULL",
        "date": "",
        "category": "방송대화",
        "annotation_level": [
          "형태분석",
          "어휘의미 분석"
        ],
        "note": "NULL"
      },
      "sentence": [
        {
          "id": "SRBK00123.1",
          "form": "안녕하세요 최고의 요리비결의 윤형빈입니다.",
          "word": [
            { "id": 1, "form": "안녕하세요", "begin": 0, "end": 5},
            { "id": 2, "form": "최고의", "begin": 6, "end": 9},
            { "id": 3, "form": "요리비결의", "begin": 10, "end": 15},
            { "id": 4, "form": "윤형빈입니다.", "begin": 16, "end": 23}
          ]
        }
      ]
    }
  ]
}
```

```

"morpheme": [
  { "id": 1, "form": "안녕", "label": "NNG", "word_id": 1, "position": 1},
  { "id": 2, "form": "하", "label": "XSA", "word_id": 1, "position": 2},
  { "id": 3, "form": "시", "label": "EP", "word_id": 1, "position": 3},
  { "id": 4, "form": "어요", "label": "EF", "word_id": 1, "position": 4},
  { "id": 5, "form": "최고", "label": "NNG", "word_id": 2, "position": 1},
  { "id": 6, "form": "의", "label": "JKG", "word_id": 2, "position": 2},
  { "id": 7, "form": "요리", "label": "NNG", "word_id": 3, "position": 1},
  { "id": 8, "form": "비결", "label": "NNG", "word_id": 3, "position": 2},
  { "id": 9, "form": "의", "label": "JKG", "word_id": 3, "position": 3},
  { "id": 10, "form": "윤형빈", "label": "NNP", "word_id": 4, "position": 1},
  { "id": 11, "form": "이", "label": "VCP", "word_id": 4, "position": 2},
  { "id": 12, "form": "보니다", "label": "EF", "word_id": 4, "position": 3},
  { "id": 13, "form": ".", "label": "SF", "word_id": 4, "position": 4}
],
"WSD": [
  { "word": "안녕", "sense_id": 1, "begin": 0, "end": 2},
  { "word": "최고", "sense_id": 3, "begin": 6, "end": 8},
  { "word": "요리", "sense_id": 6, "begin": 10, "end": 12},
  { "word": "비결", "sense_id": 3, "begin": 12, "end": 14},
  { "word": "윤형빈", "sense_id": 777, "begin": 16, "end": 19}
]
},
{
  "id": "SRBK00123.2",
  "form": "추석 음식을 준비하다 보문요 손은 많이 가는데 이게 또 티는 안 나는 그런 요리들이
있죠.",
  "word": [
    { "id": 1, "form": "추석", "begin": 0, "end": 2},
    { "id": 2, "form": "음식을", "begin": 3, "end": 6},
    { "id": 3, "form": "준비하다", "begin": 7, "end": 11},
    { "id": 4, "form": "보문요", "begin": 12, "end": 15},
    { "id": 5, "form": "손은", "begin": 16, "end": 18},
    { "id": 6, "form": "많이", "begin": 19, "end": 21},
    { "id": 7, "form": "가는데", "begin": 22, "end": 25},
    { "id": 8, "form": "이게", "begin": 26, "end": 28},
    { "id": 9, "form": "또", "begin": 29, "end": 30},
  ]
}

```

```

    { "id": 10, "form": "티는", "begin": 31, "end": 33},
    { "id": 11, "form": "안", "begin": 34, "end": 35},
    { "id": 12, "form": "나는", "begin": 36, "end": 38},
    { "id": 13, "form": "그런", "begin": 39, "end": 41},
    { "id": 14, "form": "요리들이", "begin": 42, "end": 46},
    { "id": 15, "form": "있죠.", "begin": 47, "end": 50}
  ],
  "morpheme": [
    { "id": 1, "form": "추석", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "음식", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 3, "form": "을", "label": "JKO", "word_id": 2, "position": 2},
    { "id": 4, "form": "준비", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 5, "form": "하", "label": "XSV", "word_id": 3, "position": 2},
    { "id": 6, "form": "다", "label": "EC", "word_id": 3, "position": 3},
    { "id": 7, "form": "보", "label": "VX", "word_id": 4, "position": 1},
    { "id": 8, "form": "믄", "label": "EC", "word_id": 4, "position": 2},
    { "id": 9, "form": "요", "label": "JX", "word_id": 4, "position": 3},
    { "id": 10, "form": "손", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 11, "form": "은", "label": "JX", "word_id": 5, "position": 2},
    { "id": 12, "form": "많이", "label": "MAG", "word_id": 6, "position": 1},
    { "id": 13, "form": "가", "label": "VV", "word_id": 7, "position": 1},
    { "id": 14, "form": "는데", "label": "EC", "word_id": 7, "position": 2},
    { "id": 15, "form": "이거", "label": "NP", "word_id": 8, "position": 1},
    { "id": 16, "form": "이", "label": "JKS", "word_id": 8, "position": 2},
    { "id": 17, "form": "또", "label": "MAG", "word_id": 9, "position": 1},
    { "id": 18, "form": "티", "label": "NNG", "word_id": 10, "position": 1},
    { "id": 19, "form": "는", "label": "JX", "word_id": 10, "position": 2},
    { "id": 20, "form": "안", "label": "MAG", "word_id": 11, "position": 1},
    { "id": 21, "form": "나", "label": "VV", "word_id": 12, "position": 1},
    { "id": 22, "form": "는", "label": "ETM", "word_id": 12, "position": 2},
    { "id": 23, "form": "그런", "label": "MMD", "word_id": 13, "position": 1},
    { "id": 24, "form": "요리", "label": "NNG", "word_id": 14, "position": 1},
    { "id": 25, "form": "들", "label": "XSN", "word_id": 14, "position": 2},
    { "id": 26, "form": "이", "label": "JKS", "word_id": 14, "position": 3},
    { "id": 27, "form": "있", "label": "VA", "word_id": 15, "position": 1},
    { "id": 28, "form": "죠", "label": "EF", "word_id": 15, "position": 2},
    { "id": 29, "form": ".", "label": "SF", "word_id": 15, "position": 3}
  ]
}

```

```

],
"WSD": [
  { "word": "추석", "sense_id": 1, "begin": 0, "end": 2},
  { "word": "음식", "sense_id": 1, "begin": 3, "end": 5},
  { "word": "준비", "sense_id": 1, "begin": 7, "end": 9},
  { "word": "손", "sense_id": 4, "begin": 16, "end": 17},
  { "word": "이거", "sense_id": 1, "begin": 26, "end": 28},
  { "word": "티", "sense_id": 3, "begin": 31, "end": 32},
  { "word": "요리", "sense_id": 6, "begin": 42, "end": 44}
]
},
{
  "id": "SRBK00123.3",
  "form": "그 중의 하나가 이~ 전 요리가 아닌가 싶는데요.",
  "word": [
    { "id": 1, "form": "그", "begin": 0, "end": 1},
    { "id": 2, "form": "중의", "begin": 2, "end": 4},
    { "id": 3, "form": "하나가", "begin": 5, "end": 8},
    { "id": 4, "form": "이~", "begin": 9, "end": 11},
    { "id": 5, "form": "전", "begin": 12, "end": 13},
    { "id": 6, "form": "요리가", "begin": 14, "end": 17},
    { "id": 7, "form": "아닌가", "begin": 18, "end": 21},
    { "id": 8, "form": "싶는데요.", "begin": 22, "end": 27}
  ],
  "morpheme": [
    { "id": 1, "form": "그", "label": "MMD", "word_id": 1, "position": 1},
    { "id": 2, "form": "중", "label": "NNB", "word_id": 2, "position": 1},
    { "id": 3, "form": "의", "label": "JKG", "word_id": 2, "position": 2},
    { "id": 4, "form": "하나", "label": "NR", "word_id": 3, "position": 1},
    { "id": 5, "form": "가", "label": "JKS", "word_id": 3, "position": 2},
    { "id": 6, "form": "이", "label": "IC", "word_id": 4, "position": 1},
    { "id": 7, "form": "전", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 8, "form": "요리", "label": "NNG", "word_id": 6, "position": 1},
    { "id": 9, "form": "가", "label": "JKC", "word_id": 6, "position": 2},
    { "id": 10, "form": "아니", "label": "VCN", "word_id": 7, "position": 1},
    { "id": 11, "form": "ㄴ가", "label": "EF", "word_id": 7, "position": 2},
    { "id": 12, "form": "싶", "label": "VX", "word_id": 8, "position": 1},

```

```

    { "id": 13, "form": "은테", "label": "EC", "word_id": 8, "position": 2},
    { "id": 14, "form": "요", "label": "JX", "word_id": 8, "position": 3},
    { "id": 15, "form": ".", "label": "SF", "word_id": 8, "position": 4}
  ],
  "WSD": [
    { "word": "중", "sense_id": 9, "begin": 2, "end": 3},
    { "word": "하나", "sense_id": 1, "begin": 5, "end": 7},
    { "word": "전", "sense_id": 23, "begin": 12, "end": 13},
    { "word": "요리", "sense_id": 6, "begin": 14, "end": 16}
  ]
},
{
  "id": "SRBK00123.4",
  "form": "어~ 전 요리를 어~ 되도록이면 좀 더 쉽게 하면서",
  "word": [
    { "id": 1, "form": "어~", "begin": 0, "end": 2},
    { "id": 2, "form": "전", "begin": 3, "end": 4},
    { "id": 3, "form": "요리를", "begin": 5, "end": 8},
    { "id": 4, "form": "어~", "begin": 9, "end": 11},
    { "id": 5, "form": "되도록이면", "begin": 12, "end": 17},
    { "id": 6, "form": "좀", "begin": 18, "end": 19},
    { "id": 7, "form": "더", "begin": 20, "end": 21},
    { "id": 8, "form": "쉽게", "begin": 22, "end": 24},
    { "id": 9, "form": "하면서", "begin": 25, "end": 28}
  ],
  "morpheme": [
    { "id": 1, "form": "어", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": "전", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 3, "form": "요리", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 4, "form": "를", "label": "JKO", "word_id": 3, "position": 2},
    { "id": 5, "form": "어", "label": "IC", "word_id": 4, "position": 1},
    { "id": 6, "form": "되도록", "label": "MAG", "word_id": 5, "position": 1},
    { "id": 7, "form": "이", "label": "VCP", "word_id": 5, "position": 2},
    { "id": 8, "form": "면", "label": "EC", "word_id": 5, "position": 3},
    { "id": 9, "form": "좀", "label": "MAG", "word_id": 6, "position": 1},
    { "id": 10, "form": "더", "label": "MAG", "word_id": 7, "position": 1},
    { "id": 11, "form": "쉽", "label": "VA", "word_id": 8, "position": 1},

```



```

    { "id": 12, "form": "계", "label": "EC", "word_id": 8, "position": 2},
    { "id": 13, "form": "하", "label": "VX", "word_id": 9, "position": 1},
    { "id": 14, "form": "면서", "label": "EC", "word_id": 9, "position": 2}
  ],
  "WSD": [
    { "word": "전", "sense_id": 23, "begin": 3, "end": 4},
    { "word": "요리", "sense_id": 6, "begin": 5, "end": 7}
  ]
},
{
  "id": "SRBK00123.5",
  "form": "맛까지 안 놓친다면 또 그거보다 좋은 게 어~ 뭐가 있을까 싶기도 한데요.",
  "word": [
    { "id": 1, "form": "맛까지", "begin": 0, "end": 3},
    { "id": 2, "form": "안", "begin": 4, "end": 5},
    { "id": 3, "form": "놓친다면", "begin": 6, "end": 10},
    { "id": 4, "form": "또", "begin": 11, "end": 12},
    { "id": 5, "form": "그거보다", "begin": 13, "end": 17},
    { "id": 6, "form": "좋은", "begin": 18, "end": 20},
    { "id": 7, "form": "계", "begin": 21, "end": 22},
    { "id": 8, "form": "어~", "begin": 23, "end": 25},
    { "id": 9, "form": "뭐가", "begin": 26, "end": 28},
    { "id": 10, "form": "있을까", "begin": 29, "end": 32},
    { "id": 11, "form": "싶기도", "begin": 33, "end": 36},
    { "id": 12, "form": "한데요.", "begin": 37, "end": 41}
  ],
  "morpheme": [
    { "id": 1, "form": "맛", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "까지", "label": "JX", "word_id": 1, "position": 2},
    { "id": 3, "form": "안", "label": "MAG", "word_id": 2, "position": 1},
    { "id": 4, "form": "놓치", "label": "VV", "word_id": 3, "position": 1},
    { "id": 5, "form": "ㄴ다면", "label": "EC", "word_id": 3, "position": 2},
    { "id": 6, "form": "또", "label": "MAG", "word_id": 4, "position": 1},
    { "id": 7, "form": "그거", "label": "NP", "word_id": 5, "position": 1},
    { "id": 8, "form": "보다", "label": "JKB", "word_id": 5, "position": 2},
    { "id": 9, "form": "좋", "label": "VA", "word_id": 6, "position": 1},
    { "id": 10, "form": "은", "label": "ETM", "word_id": 6, "position": 2},

```

```

    { "id": 11, "form": "거", "label": "NNB", "word_id": 7, "position": 1},
    { "id": 12, "form": "이", "label": "JKS", "word_id": 7, "position": 2},
    { "id": 13, "form": "어", "label": "IC", "word_id": 8, "position": 1},
    { "id": 14, "form": "뭐", "label": "NP", "word_id": 9, "position": 1},
    { "id": 15, "form": "가", "label": "JKS", "word_id": 9, "position": 2},
    { "id": 16, "form": "있", "label": "VA", "word_id": 10, "position": 1},
    { "id": 17, "form": "을까", "label": "EF", "word_id": 10, "position": 2},
    { "id": 18, "form": "싶", "label": "VX", "word_id": 11, "position": 1},
    { "id": 19, "form": "기", "label": "ETN", "word_id": 11, "position": 2},
    { "id": 20, "form": "도", "label": "JX", "word_id": 11, "position": 3},
    { "id": 21, "form": "하", "label": "VX", "word_id": 12, "position": 1},
    { "id": 22, "form": "ㄴ테", "label": "EC", "word_id": 12, "position": 2},
    { "id": 23, "form": "요", "label": "JX", "word_id": 12, "position": 3},
    { "id": 24, "form": ".", "label": "SF", "word_id": 12, "position": 4}
  ],
  "WSD": [
    { "word": "맞", "sense_id": 1, "begin": 0, "end": 1},
    { "word": "그거", "sense_id": 1, "begin": 13, "end": 15},
    { "word": "거", "sense_id": 1, "begin": 21, "end": 22},
    { "word": "뭐", "sense_id": 2, "begin": 26, "end": 27}
  ]
},
{
  "id": "SRBK00123.6",
  "form": "임미자 선생님 요걸 또 쉽구 간편하구 맛있게 만드는 방법이 있따구오?",
  "word": [
    { "id": 1, "form": "임미자", "begin": 0, "end": 3},
    { "id": 2, "form": "선생님", "begin": 4, "end": 7},
    { "id": 3, "form": "요걸", "begin": 8, "end": 10},
    { "id": 4, "form": "또", "begin": 11, "end": 12},
    { "id": 5, "form": "쉽구", "begin": 13, "end": 15},
    { "id": 6, "form": "간편하구", "begin": 16, "end": 20},
    { "id": 7, "form": "맛있게", "begin": 21, "end": 24},
    { "id": 8, "form": "만드는", "begin": 25, "end": 28},
    { "id": 9, "form": "방법이", "begin": 29, "end": 32},
    { "id": 10, "form": "있따구오?", "begin": 33, "end": 38}
  ],

```

```

"morpheme": [
  { "id": 1, "form": "임미자", "label": "NNP", "word_id": 1, "position": 1},
  { "id": 2, "form": "선생님", "label": "NNG", "word_id": 2, "position": 1},
  { "id": 3, "form": "요거", "label": "NP", "word_id": 3, "position": 1},
  { "id": 4, "form": "르", "label": "JKO", "word_id": 3, "position": 2},
  { "id": 5, "form": "또", "label": "MAG", "word_id": 4, "position": 1},
  { "id": 6, "form": "쉽", "label": "VA", "word_id": 5, "position": 1},
  { "id": 7, "form": "구", "label": "EC", "word_id": 5, "position": 2},
  { "id": 8, "form": "간편하", "label": "VA", "word_id": 6, "position": 1},
  { "id": 9, "form": "구", "label": "EC", "word_id": 6, "position": 2},
  { "id": 10, "form": "맞있", "label": "VA", "word_id": 7, "position": 1},
  { "id": 11, "form": "계", "label": "EC", "word_id": 7, "position": 2},
  { "id": 12, "form": "만들", "label": "VV", "word_id": 8, "position": 1},
  { "id": 13, "form": "는", "label": "ETM", "word_id": 8, "position": 2},
  { "id": 14, "form": "방법", "label": "NNG", "word_id": 9, "position": 1},
  { "id": 15, "form": "이", "label": "JKS", "word_id": 9, "position": 2},
  { "id": 16, "form": "있", "label": "VA", "word_id": 10, "position": 1},
  { "id": 17, "form": "다구", "label": "EF", "word_id": 10, "position": 2},
  { "id": 18, "form": "오", "label": "JX", "word_id": 10, "position": 3},
  { "id": 19, "form": "?", "label": "SF", "word_id": 10, "position": 4}
],
"WSD": [
  { "word": "임미자", "sense_id": 777, "begin": 0, "end": 3},
  { "word": "선생님", "sense_id": 2, "begin": 4, "end": 7},
  { "word": "요거", "sense_id": 1, "begin": 8, "end": 10},
  { "word": "방법", "sense_id": 1, "begin": 29, "end": 31}
]
},
{
  "id": "SRBK00123.7",
  "form": "네.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
  ]
}

```

```

    ],
    "WSD": null
  },
  {
    "id": "SRBK00123.8",
    "form": "네.",
    "word": [
      { "id": 1, "form": "네.", "begin": 0, "end": 2}
    ],
    "morpheme": [
      { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
      { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
    ],
    "WSD": null
  },
  {
    "id": "SRBK00123.9",
    "form": "아무리 추석 하면은 명절 때는 전이랑 적이 있어야 되잖아요.",
    "word": [
      { "id": 1, "form": "아무리", "begin": 0, "end": 3},
      { "id": 2, "form": "추석", "begin": 4, "end": 6},
      { "id": 3, "form": "하면은", "begin": 7, "end": 10},
      { "id": 4, "form": "명절", "begin": 11, "end": 13},
      { "id": 5, "form": "때는", "begin": 14, "end": 16},
      { "id": 6, "form": "전이랑", "begin": 17, "end": 20},
      { "id": 7, "form": "적이", "begin": 21, "end": 23},
      { "id": 8, "form": "있어야", "begin": 24, "end": 27},
      { "id": 9, "form": "되잖아요.", "begin": 28, "end": 33}
    ],
    "morpheme": [
      { "id": 1, "form": "아무리", "label": "MAG", "word_id": 1, "position": 1},
      { "id": 2, "form": "추석", "label": "NNG", "word_id": 2, "position": 1},
      { "id": 3, "form": "하", "label": "VV", "word_id": 3, "position": 1},
      { "id": 4, "form": "면", "label": "EC", "word_id": 3, "position": 2},
      { "id": 5, "form": "은", "label": "JX", "word_id": 3, "position": 3},
      { "id": 6, "form": "명절", "label": "NNG", "word_id": 4, "position": 1},
      { "id": 7, "form": "때", "label": "NNG", "word_id": 5, "position": 1},

```

```

    { "id": 8, "form": "는", "label": "JX", "word_id": 5, "position": 2},
    { "id": 9, "form": "전", "label": "NNG", "word_id": 6, "position": 1},
    { "id": 10, "form": "이랑", "label": "JC", "word_id": 6, "position": 2},
    { "id": 11, "form": "적", "label": "NNG", "word_id": 7, "position": 1},
    { "id": 12, "form": "이", "label": "JKS", "word_id": 7, "position": 2},
    { "id": 13, "form": "있", "label": "VA", "word_id": 8, "position": 1},
    { "id": 14, "form": "어야", "label": "EC", "word_id": 8, "position": 2},
    { "id": 15, "form": "되", "label": "VV", "word_id": 9, "position": 1},
    { "id": 16, "form": "잖아요", "label": "EF", "word_id": 9, "position": 2},
    { "id": 17, "form": ".", "label": "SF", "word_id": 9, "position": 3}
  ],
  "WSD": [
    { "word": "추석", "sense_id": 1, "begin": 4, "end": 6},
    { "word": "명절", "sense_id": 1, "begin": 11, "end": 13},
    { "word": "때", "sense_id": 6, "begin": 14, "end": 15},
    { "word": "전", "sense_id": 23, "begin": 17, "end": 18},
    { "word": "적", "sense_id": 7, "begin": 21, "end": 22}
  ]
},
{
  "id": "SRBK00123.10",
  "form": "그릉쵸.",
  "word": [
    { "id": 1, "form": "그릉쵸.", "begin": 0, "end": 4}
  ],
  "morpheme": [
    { "id": 1, "form": "그릉", "label": "VA", "word_id": 1, "position": 1},
    { "id": 2, "form": "쵸", "label": "EF", "word_id": 1, "position": 2},
    { "id": 3, "form": ".", "label": "SF", "word_id": 1, "position": 3}
  ],
  "WSD": null
},
{
  "id": "SRBK00123.11",
  "form": "그래서 오늘 누름적과 동태전을 만들어 보겠습니다.",
  "word": [
    { "id": 1, "form": "그래서", "begin": 0, "end": 3},

```

```

    { "id": 2, "form": "오늘", "begin": 4, "end": 6},
    { "id": 3, "form": "누름적과", "begin": 7, "end": 11},
    { "id": 4, "form": "동태전을", "begin": 12, "end": 16},
    { "id": 5, "form": "만들어", "begin": 17, "end": 20},
    { "id": 6, "form": "보겠습니다.", "begin": 21, "end": 27}
  ],
  "morpheme": [
    { "id": 1, "form": "그래서", "label": "MAJ", "word_id": 1, "position": 1},
    { "id": 2, "form": "오늘", "label": "MAG", "word_id": 2, "position": 1},
    { "id": 3, "form": "누름적", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 4, "form": "과", "label": "JC", "word_id": 3, "position": 2},
    { "id": 5, "form": "동태전", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 6, "form": "을", "label": "JKO", "word_id": 4, "position": 2},
    { "id": 7, "form": "만들", "label": "VV", "word_id": 5, "position": 1},
    { "id": 8, "form": "어", "label": "EC", "word_id": 5, "position": 2},
    { "id": 9, "form": "보", "label": "VX", "word_id": 6, "position": 1},
    { "id": 10, "form": "겠", "label": "EP", "word_id": 6, "position": 2},
    { "id": 11, "form": "습니다", "label": "EF", "word_id": 6, "position": 3},
    { "id": 12, "form": ".", "label": "SF", "word_id": 6, "position": 4}
  ],
  "WSD": [
    { "word": "누름적", "sense_id": 1, "begin": 7, "end": 10},
    { "word": "동태전", "sense_id": 1, "begin": 12, "end": 15}
  ]
},
{
  "id": "SRBK00123.12",
  "form": "자 선생님의 어~ 숨겨진 비결.",
  "word": [
    { "id": 1, "form": "자", "begin": 0, "end": 1},
    { "id": 2, "form": "선생님의", "begin": 2, "end": 6},
    { "id": 3, "form": "어~", "begin": 7, "end": 9},
    { "id": 4, "form": "숨겨진", "begin": 10, "end": 13},
    { "id": 5, "form": "비결.", "begin": 14, "end": 17}
  ],
  "morpheme": [
    { "id": 1, "form": "자", "label": "IC", "word_id": 1, "position": 1},

```

```

    { "id": 2, "form": "선생님", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 3, "form": "의", "label": "JKG", "word_id": 2, "position": 2},
    { "id": 4, "form": "어", "label": "IC", "word_id": 3, "position": 1},
    { "id": 5, "form": "숨기", "label": "VV", "word_id": 4, "position": 1},
    { "id": 6, "form": "어", "label": "EC", "word_id": 4, "position": 2},
    { "id": 7, "form": "지", "label": "VX", "word_id": 4, "position": 3},
    { "id": 8, "form": "ㄴ", "label": "ETM", "word_id": 4, "position": 4},
    { "id": 9, "form": "비결", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 10, "form": ".", "label": "SF", "word_id": 5, "position": 2}
  ],
  "WSD": [
    { "word": "선생님", "sense_id": 2, "begin": 2, "end": 5},
    { "word": "비결", "sense_id": 3, "begin": 14, "end": 16}
  ]
},
{
  "id": "SRBK00123.13",
  "form": "전의 정석. 네.",
  "word": [
    { "id": 1, "form": "전의", "begin": 0, "end": 2},
    { "id": 2, "form": "정석.", "begin": 3, "end": 6},
    { "id": 3, "form": "네.", "begin": 7, "end": 9}
  ],
  "morpheme": [
    { "id": 1, "form": "전", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "의", "label": "JKG", "word_id": 1, "position": 2},
    { "id": 3, "form": "정석", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 4, "form": ".", "label": "SF", "word_id": 2, "position": 2},
    { "id": 5, "form": "네", "label": "IC", "word_id": 3, "position": 1},
    { "id": 6, "form": ".", "label": "SF", "word_id": 3, "position": 2}
  ],
  "WSD": [
    { "word": "전", "sense_id": 23, "begin": 0, "end": 1},
    { "word": "정석", "sense_id": 1, "begin": 3, "end": 5}
  ]
},
{

```

```

    "id": "SRBK00123.14",
    "form": "",
    "word": null,
    "morpheme": null,
    "WSD": null
  },
  {
    "id": "SRBK00123.15",
    "form": "나갑니다 네. ",
    "word": [
      { "id": 1, "form": "나갑니다", "begin": 0, "end": 4},
      { "id": 2, "form": "네.", "begin": 5, "end": 7}
    ],
    "morpheme": [
      { "id": 1, "form": "나가", "label": "VV", "word_id": 1, "position": 1},
      { "id": 2, "form": "버니다", "label": "EF", "word_id": 1, "position": 2},
      { "id": 3, "form": "네", "label": "IC", "word_id": 2, "position": 1},
      { "id": 4, "form": ".", "label": "SF", "word_id": 2, "position": 2}
    ],
    "WSD": null
  },
  {
    "id": "SRBK00123.16",
    "form": "누름적과 동태전입니다.",
    "word": [
      { "id": 1, "form": "누름적과", "begin": 0, "end": 4},
      { "id": 2, "form": "동태전입니다.", "begin": 5, "end": 12}
    ],
    "morpheme": [
      { "id": 1, "form": "누름적", "label": "NNG", "word_id": 1, "position": 1},
      { "id": 2, "form": "과", "label": "JC", "word_id": 1, "position": 2},
      { "id": 3, "form": "동태전", "label": "NNG", "word_id": 2, "position": 1},
      { "id": 4, "form": "이", "label": "VCP", "word_id": 2, "position": 2},
      { "id": 5, "form": "버니다", "label": "EF", "word_id": 2, "position": 3},
      { "id": 6, "form": ".", "label": "SF", "word_id": 2, "position": 4}
    ],
    "WSD": [

```



```

    { "word": "누름적", "sense_id": 1, "begin": 0, "end": 3},
    { "word": "동태전", "sense_id": 1, "begin": 5, "end": 8}
  ]
},
{
  "id": "SRBK00123.17",
  "form": "재료 나와주세요.",
  "word": [
    { "id": 1, "form": "재료", "begin": 0, "end": 2},
    { "id": 2, "form": "나와주세요.", "begin": 3, "end": 9}
  ],
  "morpheme": [
    { "id": 1, "form": "재료", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "나오", "label": "VV", "word_id": 2, "position": 1},
    { "id": 3, "form": "아", "label": "EC", "word_id": 2, "position": 2},
    { "id": 4, "form": "주", "label": "VX", "word_id": 2, "position": 3},
    { "id": 5, "form": "시", "label": "EP", "word_id": 2, "position": 4},
    { "id": 6, "form": "어요", "label": "EF", "word_id": 2, "position": 5},
    { "id": 7, "form": ".", "label": "SF", "word_id": 2, "position": 6}
  ],
  "WSD": [
    { "word": "재료", "sense_id": 1, "begin": 0, "end": 2}
  ]
},
{
  "id": "SRBK00123.18",
  "form": "자 선생님 그러면 이~ 누름적은 양념 먼저 만들어야 된다고요?",
  "word": [
    { "id": 1, "form": "자", "begin": 0, "end": 1},
    { "id": 2, "form": "선생님", "begin": 2, "end": 5},
    { "id": 3, "form": "그러면", "begin": 6, "end": 9},
    { "id": 4, "form": "이~", "begin": 10, "end": 12},
    { "id": 5, "form": "누름적은", "begin": 13, "end": 17},
    { "id": 6, "form": "양념", "begin": 18, "end": 20},
    { "id": 7, "form": "먼저", "begin": 21, "end": 23},
    { "id": 8, "form": "만들어야", "begin": 24, "end": 28},
    { "id": 9, "form": "된다고요?", "begin": 29, "end": 34}
  ]
}

```

```

],
"morpheme": [
  { "id": 1, "form": "자", "label": "IC", "word_id": 1, "position": 1},
  { "id": 2, "form": "선생님", "label": "NNG", "word_id": 2, "position": 1},
  { "id": 3, "form": "그러면", "label": "MAJ", "word_id": 3, "position": 1},
  { "id": 4, "form": "이", "label": "IC", "word_id": 4, "position": 1},
  { "id": 5, "form": "누름적", "label": "NNG", "word_id": 5, "position": 1},
  { "id": 6, "form": "은", "label": "JX", "word_id": 5, "position": 2},
  { "id": 7, "form": "양념", "label": "NNG", "word_id": 6, "position": 1},
  { "id": 8, "form": "먼저", "label": "MAG", "word_id": 7, "position": 1},
  { "id": 9, "form": "만들", "label": "VV", "word_id": 8, "position": 1},
  { "id": 10, "form": "어야", "label": "EC", "word_id": 8, "position": 2},
  { "id": 11, "form": "되", "label": "VV", "word_id": 9, "position": 1},
  { "id": 12, "form": "ㄴ 다구", "label": "EF", "word_id": 9, "position": 2},
  { "id": 13, "form": "요", "label": "JX", "word_id": 9, "position": 3},
  { "id": 14, "form": "?", "label": "SF", "word_id": 9, "position": 4}
],
"WSD": [
  { "word": "선생님", "sense_id": 2, "begin": 2, "end": 5},
  { "word": "누름적", "sense_id": 1, "begin": 13, "end": 16},
  { "word": "양념", "sense_id": 1, "begin": 18, "end": 20}
]
},
{
  "id": "SRBK00123.19",
  "form": "네. 양념 만들어서 인제",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2},
    { "id": 2, "form": "양념", "begin": 3, "end": 5},
    { "id": 3, "form": "만들어서", "begin": 6, "end": 10},
    { "id": 4, "form": "인제", "begin": 11, "end": 13}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2},
    { "id": 3, "form": "양념", "label": "NNG", "word_id": 2, "position": 1},
    { "id": 4, "form": "만들", "label": "VV", "word_id": 3, "position": 1},

```

```

    { "id": 5, "form": "어서", "label": "EC", "word_id": 3, "position": 2},
    { "id": 6, "form": "인제", "label": "IC", "word_id": 4, "position": 1}
  ],
  "WSD": [
    { "word": "양념", "sense_id": 1, "begin": 3, "end": 5}
  ]
},
{
  "id": "SRBK00123.20",
  "form": "네.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
  ],
  "WSD": null
},
{
  "id": "SRBK00123.21",
  "form": "톡톡톡톡 뚜드리는 고기에다 양념을 해 볼까요?",
  "word": [
    { "id": 1, "form": "톡톡톡톡", "begin": 0, "end": 4},
    { "id": 2, "form": "뚜드리는", "begin": 5, "end": 9},
    { "id": 3, "form": "고기에다", "begin": 10, "end": 14},
    { "id": 4, "form": "양념을", "begin": 15, "end": 18},
    { "id": 5, "form": "해", "begin": 19, "end": 20},
    { "id": 6, "form": "볼까요?", "begin": 21, "end": 25}
  ],
  "morpheme": [
    { "id": 1, "form": "톡톡", "label": "MAG", "word_id": 1, "position": 1},
    { "id": 2, "form": "톡톡", "label": "MAG", "word_id": 1, "position": 2},
    { "id": 3, "form": "뚜드리", "label": "VV", "word_id": 2, "position": 1},
    { "id": 4, "form": "는", "label": "ETM", "word_id": 2, "position": 2},
    { "id": 5, "form": "고기", "label": "NNG", "word_id": 3, "position": 1},
    { "id": 6, "form": "에다", "label": "JKB", "word_id": 3, "position": 2},

```

```

    { "id": 7, "form": "양념", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 8, "form": "을", "label": "JKO", "word_id": 4, "position": 2},
    { "id": 9, "form": "하", "label": "VV", "word_id": 5, "position": 1},
    { "id": 10, "form": "아", "label": "EC", "word_id": 5, "position": 2},
    { "id": 11, "form": "보", "label": "VX", "word_id": 6, "position": 1},
    { "id": 12, "form": "르까", "label": "EF", "word_id": 6, "position": 2},
    { "id": 13, "form": "요", "label": "JX", "word_id": 6, "position": 3},
    { "id": 14, "form": "?", "label": "SF", "word_id": 6, "position": 4}
  ],
  "WSD": [
    { "word": "고기", "sense_id": 1, "begin": 10, "end": 12},
    { "word": "양념", "sense_id": 1, "begin": 15, "end": 17}
  ]
},
{
  "id": "SRBK00123.22",
  "form": "네. 자.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2},
    { "id": 2, "form": "자.", "begin": 3, "end": 5}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2},
    { "id": 3, "form": "자", "label": "IC", "word_id": 2, "position": 1},
    { "id": 4, "form": ".", "label": "SF", "word_id": 2, "position": 2}
  ],
  "WSD": null
},
{
  "id": "SRBK00123.23",
  "form": "그러면 보통 인제 쇠고기 양념이니깐",
  "word": [
    { "id": 1, "form": "그러면", "begin": 0, "end": 3},
    { "id": 2, "form": "보통", "begin": 4, "end": 6},
    { "id": 3, "form": "인제", "begin": 7, "end": 9},
    { "id": 4, "form": "쇠고기", "begin": 10, "end": 13},

```

```

    { "id": 5, "form": "양념이니깐", "begin": 14, "end": 19}
  ],
  "morpheme": [
    { "id": 1, "form": "그러면", "label": "MAJ", "word_id": 1, "position": 1},
    { "id": 2, "form": "보통", "label": "MAG", "word_id": 2, "position": 1},
    { "id": 3, "form": "인제", "label": "IC", "word_id": 3, "position": 1},
    { "id": 4, "form": "최고기", "label": "NNG", "word_id": 4, "position": 1},
    { "id": 5, "form": "양념", "label": "NNG", "word_id": 5, "position": 1},
    { "id": 6, "form": "이", "label": "VCP", "word_id": 5, "position": 2},
    { "id": 7, "form": "니깐", "label": "EC", "word_id": 5, "position": 3}
  ],
  "WSD": [
    { "word": "최고기", "sense_id": 1, "begin": 10, "end": 13},
    { "word": "양념", "sense_id": 1, "begin": 14, "end": 16}
  ]
},
{
  "id": "SRBK00123.24",
  "form": "네.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
  ],
  "WSD": null
},
{
  "id": "SRBK00123.25",
  "form": "인제 간장 짝은 양념이에요.",
  "word": [
    { "id": 1, "form": "인제", "begin": 0, "end": 2},
    { "id": 2, "form": "간장", "begin": 3, "end": 5},
    { "id": 3, "form": " 짝은", "begin": 6, "end": 8},
    { "id": 4, "form": " 양념이에요.", "begin": 9, "end": 15}
  ],

```

```

"morpheme": [
  { "id": 1, "form": "인제", "label": "MAG", "word_id": 1, "position": 1},
  { "id": 2, "form": "간장", "label": "NNG", "word_id": 2, "position": 1},
  { "id": 3, "form": "꺾은", "label": "MMN", "word_id": 3, "position": 1},
  { "id": 4, "form": "양념", "label": "NNG", "word_id": 4, "position": 1},
  { "id": 5, "form": "이", "label": "VCP", "word_id": 4, "position": 2},
  { "id": 6, "form": "예요", "label": "EF", "word_id": 4, "position": 3},
  { "id": 7, "form": ".", "label": "SF", "word_id": 4, "position": 4}
],
"WSD": [
  { "word": "간장", "sense_id": 1, "begin": 3, "end": 5},
  { "word": "양념", "sense_id": 1, "begin": 9, "end": 11}
]
},
{
  "id": "SRBK00123.26",
  "form": "네.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
  ],
  "WSD": null
},
{
  "id": "SRBK00123.27",
  "form": "예. 간장하구요 설탕 또 요렇게 넣으시고.",
  "word": [
    { "id": 1, "form": "예.", "begin": 0, "end": 2},
    { "id": 2, "form": "간장하구요", "begin": 3, "end": 8},
    { "id": 3, "form": "설탕", "begin": 9, "end": 11},
    { "id": 4, "form": "또", "begin": 12, "end": 13},
    { "id": 5, "form": "요렇게", "begin": 14, "end": 17},
    { "id": 6, "form": "넣으시고.", "begin": 18, "end": 23}
  ],

```

```

"morpheme": [
  { "id": 1, "form": "예", "label": "IC", "word_id": 1, "position": 1},
  { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2},
  { "id": 3, "form": "간장", "label": "NNG", "word_id": 2, "position": 1},
  { "id": 4, "form": "하구", "label": "JC", "word_id": 2, "position": 2},
  { "id": 5, "form": "요", "label": "JX", "word_id": 2, "position": 3},
  { "id": 6, "form": "설탕", "label": "NNG", "word_id": 3, "position": 1},
  { "id": 7, "form": "또", "label": "MAG", "word_id": 4, "position": 1},
  { "id": 8, "form": "요랑", "label": "VA", "word_id": 5, "position": 1},
  { "id": 9, "form": "계", "label": "EC", "word_id": 5, "position": 2},
  { "id": 10, "form": "넣", "label": "VV", "word_id": 6, "position": 1},
  { "id": 11, "form": "으시", "label": "EP", "word_id": 6, "position": 2},
  { "id": 12, "form": "고", "label": "EC", "word_id": 6, "position": 3},
  { "id": 13, "form": ".", "label": "SF", "word_id": 6, "position": 4}
],
"WSD": [
  { "word": "간장", "sense_id": 1, "begin": 3, "end": 5},
  { "word": "설탕", "sense_id": 1, "begin": 9, "end": 11}
]
},
{
  "id": "SRBK00123.28",
  "form": "네.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
  ],
  "WSD": null
},
{
  "id": "SRBK00123.29",
  "form": "그답에 다진 마늘",
  "word": [
    { "id": 1, "form": "그답에", "begin": 0, "end": 3},

```

```

    { "id": 2, "form": "다진", "begin": 4, "end": 6},
    { "id": 3, "form": "마늘", "begin": 7, "end": 9}
  ],
  "morpheme": [
    { "id": 1, "form": "그답", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "에", "label": "JKB", "word_id": 1, "position": 2},
    { "id": 3, "form": "다지", "label": "VV", "word_id": 2, "position": 1},
    { "id": 4, "form": "ㄴ", "label": "ETM", "word_id": 2, "position": 2},
    { "id": 5, "form": "마늘", "label": "NNG", "word_id": 3, "position": 1}
  ],
  "WSD": [
    { "word": "그답", "sense_id": 1, "begin": 0, "end": 2},
    { "word": "마늘", "sense_id": 1, "begin": 7, "end": 9}
  ]
},
{
  "id": "SRBK00123.30",
  "form": "마늘이 들어갑니다.",
  "word": [
    { "id": 1, "form": "마늘이", "begin": 0, "end": 3},
    { "id": 2, "form": "들어갑니다.", "begin": 4, "end": 10}
  ],
  "morpheme": [
    { "id": 1, "form": "마늘", "label": "NNG", "word_id": 1, "position": 1},
    { "id": 2, "form": "이", "label": "JKS", "word_id": 1, "position": 2},
    { "id": 3, "form": "들어가", "label": "VV", "word_id": 2, "position": 1},
    { "id": 4, "form": "버니다", "label": "EF", "word_id": 2, "position": 2},
    { "id": 5, "form": ".", "label": "SF", "word_id": 2, "position": 3}
  ],
  "WSD": [
    { "word": "마늘", "sense_id": 1, "begin": 0, "end": 2}
  ]
},
{
  "id": "SRBK00123.31",
  "form": "다진 파",
  "word": [

```



```

    { "id": 1, "form": "다진", "begin": 0, "end": 2},
    { "id": 2, "form": "파", "begin": 3, "end": 4}
  ],
  "morpheme": [
    { "id": 1, "form": "다지", "label": "VV", "word_id": 1, "position": 1},
    { "id": 2, "form": "ㄴ", "label": "ETM", "word_id": 1, "position": 2},
    { "id": 3, "form": "파", "label": "NNG", "word_id": 2, "position": 1}
  ],
  "WSD": [
    { "word": "파", "sense_id": 1, "begin": 3, "end": 4}
  ]
},
{
  "id": "SRBK00123.32",
  "form": "네.",
  "word": [
    { "id": 1, "form": "네.", "begin": 0, "end": 2}
  ],
  "morpheme": [
    { "id": 1, "form": "네", "label": "IC", "word_id": 1, "position": 1},
    { "id": 2, "form": ".", "label": "SF", "word_id": 1, "position": 2}
  ],
  "WSD": null
},

```

연구 기관: 고려대학교 산학협력단 · 서울대학교 산학협력단 · 성신여자대학교 산학협력단

<연구진>	
연구 책임자	이도길(고려대학교)
공동 연구원	김일환(성신여자대학교)
	박진호(서울대학교)
	송상현(고려대학교)
	유현조(서울대학교)
	윤태진(성신여자대학교)
	이규범(고려대학교)
	이영제(고려대학교)
	정성훈(목포대학교)
	정연주(홍익대학교)
	최운호(목포대학교)
연구 보조원	강수연(고려대학교)
	김다미(서울대학교)
	김송희(서울대학교)
	김연우(고려대학교)
	김연희(경희대학교)
	김희숙(서울대학교)
	박종우(성신여자대학교)
	백인영(서울대학교)
	서신애(성신여자대학교)
	신현규(서울대학교)
	이상진(서강대학교)
	이세은(고려대학교)
이순옥(서울대학교)	

	이영경(고려대학교)
	이윤경(서울대학교)
	이종진(성신여자대학교)
	이준희(동국대학교)
	정슬아(성신여자대학교)
	정우현(서울대학교)
	최선지(고려대학교)
	최종원(서울대학교)
	최준호(서울대학교)
	홍승혜(고려대학교)

---

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2020년 1월 8일

발행일: 2020년 1월 8일

인 쇄: 현대문화사

---

※ 이 책은 국립국어원의 용역비로 수행한 ‘어휘의미 분석 말뭉치 구축’  
사업의 결과물을 발간한 것입니다.

국립국어원

2019  
|  
01  
|  
24

# 어휘의미 분석 말뭉치 구축

국립국어원

