

국립국어원 2009-01-30

발 간 등 록 번 호

11-1371028-000118-01

# 어휘의미 체계 기반 입체적 국어사전 확장

국립국어원 2009-01-30

발 간 등 록 번 호

11-1371028-000118-01

# 어휘의미 체계 기반 입체적 국어사전 확장

연구책임자 : 옥철영(울산대학교)



# 제 출 문

국립국어원장 귀하

“어휘의미 체계 기반 입체적 국어사전 확장”에 관하여 귀 원와 체결한 연구 용역 계약에 의하여 연구최종보고서를 작성하여 제출합니다.

2009년 11월 20일

울산대학교 산학협력단

연구책임자 : 옥철영(울산대학교)

공동연구원 : 김혜영(경남대학교)

보 조 원 : 배영준(울산대학교)

신준철(울산대학교)

이용훈(울산대학교)

정성린(울산대학교)

김윤정(울산대학교)

최효식(울산대학교)

조희산(울산대학교)

정용호(울산대학교)

김현수(경남대학교)

# 요 약 문

## 1. 사업명

“어휘의미 체계 기반 입체적 국어사전 확장”

## 2. 연구의 목적

- 다의어 수준 명사 10,000여 개를 추가하고 어휘의미 관계(상위어, 하위어, 동의어, 반의어) 설정
- 다의어 수준 동사, 형용사 7,000여 개를 추가하고 명사 어휘의미 체계 기반으로 논항정보 설정
- 다의어 수준 명사, 동사, 형용사 뜻풀이 및 용례에 대한 형태/의미 주석 부착
- 다의어 수준 동사, 형용사를 의미 부류별로 군집화
- 다의어 수준 동사, 형용사를 의미 부류별로 군집화하여 제시하는 브라우저 설계 및 구현

## 3. 연구의 필요성

- 국어정보화 산업의 활성화를 위하여 표준국어대사전의 기계가독성을 확보할 필요가 있음.
- 한국어 어휘의미 정보를 체계적이고 종합적으로 파악하기 위한 어휘의미 체계 기반 국어사전 구축
- 국립국어원에서 구축한 어휘의미 체계 기반 국어사전(2008년)의 지속적인 확장 필요

## 4. 연구내용 및 범위

- 추가 표제어 추출
  - 2008년도 작업에서 50,000개의 기본어휘를 선정하기 위해 표준국어대사

전 뜻풀이와 세종말뭉치에서 추출했던 기본어휘는 127,659개로, 이들 중 50,328개를 기본어휘로 선정하였다.

- 2009년 연구과제에서 새로 추가할 명사 10,000개, 용언 7,000개의 어휘는 2008년에 최종 선정되지 않은 어휘와 세종전자사전을 참조하여 선별하되 2008년 기본어휘 선정 원칙을 준수한다.
- 다의어 수준 표제어 간 어휘의미 관계 설정
  - 새로 명사가 추가됨에 따라, 추가된 명사 표제어 간 의미 관계(상위어/하위어, 동의어/반의어) 제시
  - 2008년 구축한 동사와 형용사의 논항정보를 새로 추가된 명사 어휘망을 반영하여 수정하며, 새로 추가된 용언에 대해서도 동일한 기준의 논항정보 구축
  - 새로운 용언/명사가 추가됨에 따라, 기 구축된 부사와 의미적으로 결합할 수 있는 동사와 형용사 제시
- 추가 선정된 다의어 수준 표제어의 형태·의미 주석 부착
  - 추가된 표제어의 뜻풀이에 대한 형태·의미 주석 부착
  - 2009년까지 선정된 표제어를 대상으로 표준국어대사전에 제시된 뜻풀이의 용례에 대한 형태·의미 주석 부착
- 다의어 수준 동사, 형용사를 의미 부류별로 군집화
- 다의어 수준 동사, 형용사를 의미 부류별로 군집화하여 제시하는 브라우저 설계 및 구현

## 5. 연구결과의 활용

- 품사 및 다의어 수준의 의미 주석된 표준국어대사전의 뜻풀이 및 용례(최소 70,000)를 확보함으로써, 표준국어대사전에서 다의어 수준의 다양한 어휘 정보를 추출할 수 있음.
- 웹 브라우저를 이용하여 어휘의 다양한 의미 관계(상위어/하위어/동의어/반의어, 하위범주 정보 및 부사 등)를 입체적으로 파악할 수 있음.

### 핵심어 :

어휘의미망, 입체적 국어사전, 의미 관계, 상위의, 하위어, 반의어, 동의어, 하위범주 정보, 다의어 태깅

## **Abstract**

### **1. Research Project Title**

“Expanding Multi-dimensional Korean Dictionary based on Word Network”

### **2. Objective of Research Project**

- Addition of 10,000 nouns in polysemy level and construction of semantic relations(hypernym, hyponym, synonym, and antonym).
- Addition of 7,000 predicate in polysemy level and subcategorization mapped on Noun's Word Network.
- Part of Speech and polysemy tagging of definition and example sentences in Standard Korean Great Dictionary(STGD).
- Semantic clusterization of predicates.
- Design and implementation of browser for semantic clusters.

### **3. Necessities of Research Project**

- In order to vitalize industry of Korean information, it is necessary to ensure machine readability of STGD.
- In order to grasp Korean lexical semantic information multidimensionally and systematically, it is necessary to construct Korean dictionary based on lexical semantic system.
- Expanding continuously the Korean dictionary based on lexical semantic system constructed by the National Institute of the Korean Language (2008).

### **4. Content and Scope**

- Selection of Additional words
  - 50,328 words in polysemy level were selected in 2008 among 127,659

words extracted from 21<sup>th</sup> Sejong corpus and definitions in the STGD which were tagged in POS and polysemy level.

- Additional 10,000 nouns and 7,000 predicates are selected from the remained words in 2008 and the Sejong Electronic Dictionary on the criteria of words selection in 2008.
- Construction of semantic relations in polysemy level
  - Construction of semantic relations(hypernym, hyponym, synonym, and antonym) for additional nouns.
  - Modification of subcategories of predicates selected in 2008 and subcategorization for the additional predicates on the same criteria.
  - Modification of semantic relation between adverbs and predicates.
- POS and polysemy tagging
  - Definitions in STGD of the additional words.
  - Examples in STGD of all selected definitions.
- Semantic clustering in polysemy level for all selected predicates.
- Design and implementation of WEB browser for semantic clusters.

## **5. Applications**

- Extraction of various semantic information from POS and polysemy tagged definition and examples.
- Multi-dimensional grasp of various semantic relations (hypernym, hyponym, synonym, and antonym, subcategory, adverbs) using Web browser.

### **Keywords :**

Lexical Semantic Network, Multi-dimensional Korean Dictionary, Semantic Relations, Hypernym, Hyponym, Synonym, Antonym, Subcategorization, POS and Polysemy Tagging

## 차 례

I. 추가 어휘 선정 .....	1
1. 추가 어휘 선정 지침 .....	1
2. 최종 선정/추가 어휘 .....	3
2.1. 입체적 국어사전 구축을 위한 선정 어휘 .....	3
2.2. 입체적 국어사전 구축 과정에서 추가된 어휘 .....	3
II. 표준국어대사전 뜻풀이 및 용례 태깅 .....	5
1. 표준국어대사전 비교 .....	5
2. 뜻풀이 및 용례 태깅 .....	6
2.1. 품사 태깅 원칙 .....	6
2.2. 의미 태깅 원칙 .....	8
2.3. 동형이의어/다의어 의미 태깅 도구 .....	11
2.4. 태깅 오류 직접 수정 도구 .....	13
III. 명사 어휘망 구축 .....	15
1. 명사 어휘망 .....	15
2. 선정 어휘의 U-WIN 매핑 .....	16
IV. 용언의 하위범주 정보 구축 .....	19
1. 하위범주 정보 .....	19
1.1. 논항정보 구축 대상 .....	19
1.2. 논항정보 구축 절차 .....	19
2. 용언의 하위범주 정보 구축 과정 .....	20
2.1. 용언의 하위범주 정보 추출 .....	20
2.2. 하위범주 정보 구축 도구 .....	24
2.3. 하위범주 정보 연결 과정 .....	25
3. 하위범주 정보구축 지침 .....	27
4. 구축 작업 중에 부딪치는 난점들 .....	38
4.1. 전문어 부류의 처리 문제 .....	38
4.2. 용례 가능성의 확대 문제 .....	39
4.3. 의미의 정확한 연결 문제 .....	40

V. 부사의 결합관계 정보 구축 .....	43
1. 부사의 결합관계 구축 과정 .....	43
1.1. 부사의 결합관계 추출 .....	43
1.2. 부사의 결합관계 구축 도구 .....	45
1.3. 부사의 결합관계 연결 과정 .....	46
2. 부사의 결합관계 구축 지침 .....	49
VI. 용언 의미 군집화 .....	53
1. 필요성 및 관련 연구 .....	53
2. 사전 뜻풀이를 이용한 의미 군집화 .....	54
2.1. 용언의 사전 뜻풀이 .....	54
2.2. 유의어 세트 추출 .....	54
3. 세종 <사태> 의미부류를 활용한 의미 군집 .....	58
3.1. <사태> 의미 부류 .....	58
3.2. 선정 용언의 <사태> 의미부류로의 매핑 .....	60
3.3. 용언 의미군집의 계층화 및 브라우저 .....	63
VII. 입체적 국어사전 브라우저 .....	65
1. 어휘망 편집 브라우저 .....	65
1.1. 어휘망 편집 브라우저 구성 요소 .....	65
1.2. 어휘 관계 설정 .....	67
1.3. 어휘 추가/수정/삭제 .....	69
2. 입체적 국어사전 브라우저 .....	72
2.1. 입체적 국어사전 브라우저 구성 요소 .....	72
2.2. 입체적 국어사전 브라우저 기능 .....	73
3. 브라우저의 데이터베이스 설계서 .....	85
3.1. 테이블 종류 및 구조 .....	85
3.2. 필드 간의 관계 .....	89
참고문헌 .....	91
부록 A. 선정 어휘 관리 목록 .....	93
부록 B. 품사/동형이의어 동시 태깅 시스템 사용설명서 .....	98

## <표 차례>

<표 I-1> 어휘선정 원칙에 따른 품사별 선정 어휘 수 .....	3
<표 I-2> 입체적 국어사전에 포함된 품사별 추가 어휘 수 .....	4
<표 II-1> <시디용표준>과 <웹용표준>의 뜻풀이 차원에서의 비교 .....	5
<표 II-2> '먹다'의 표준국어대사전 뜻풀이 .....	8
<표 III-1> U-WIN상에서의 기본어휘의 계층구조 .....	16
<표 IV-1> 구축 대상 용언 .....	19
<표 IV-2> 구축 작업 순서 .....	20
<표 IV-3> '먹다'의 뜻풀이 및 용례(표준국어대사전) .....	20
<표 IV-4> 다의어 태깅된 뜻풀이 및 용례에서 추출한 '갈다'의 하위범주 정보 .....	22
<표 V-1> 뜻풀이와 용례에서 추출한 {가급적, 가까스로}와 결합하는 부사 및 용언 .....	44
<표 VI-1> {먹다}를 유의어로 가지는 용언 .....	56
<표 VI-2> 유의어 {먹다}의 상위 유의어 연결 .....	57
<표 VI-3> 최상위 유의어 군집 예 .....	58
<표 VI-4> 세종전자사전(용언) .....	60
<표 VII-1> DB 내 각종 테이블 목록 .....	85
<표 VII-2> Relation 테이블 구조 .....	86
<표 VII-3> ERelation 테이블 구조 .....	86
<표 VII-4> rel-XXX 테이블 구조 .....	86
<표 VII-5> RName 테이블 구조 .....	87

## <그림 차례>

<그림 II-1> 의미 태깅 대상 어휘를 가진 모든 뜻풀이 태깅 도구 .....	11
<그림 II-2> 의미 태깅 대상 어휘를 가진 모든 용례 태깅 도구 .....	13
<그림 II-3> 태깅 오류 직접 수정 도구 .....	14
<그림 III-1> 입체적 국어사전에서의 명사 어휘망 .....	15
<그림 IV-1> 하위범주 구축 도구 .....	25
<그림 IV-2> 하위범주 정보 연결 과정 .....	26
<그림 IV-3> 입체적 국어사전에서 '먹다_02_01_01'에 연결된 논항정보 .....	27
<그림 IV-4> 간소화된 문형정보 .....	28
<그림 IV-5> {밥}의 최소상계노드 .....	29
<그림 IV-6> '먹다_02_01_01'의 용례 .....	30
<그림 IV-7> 논항 명사들의 최소상계노드 명사 .....	30
<그림 IV-8> 논항정보 연결 상태 .....	31
<그림 IV-9> '먹다_02_01_01'의 의미 .....	31
<그림 IV-10> '약_07_00_01'의 동위노드 명사들 .....	32
<그림 IV-11> 상하 관계의 {음식, 밥} .....	32
<그림 IV-12> 최종 연결 상태 .....	33
<그림 IV-13> '금하다_00_00_02'의 용례 .....	33
<그림 IV-14> 같은 줄기의 상위노드 .....	33
<그림 IV-15> '구체화되다_00_01_02'의 의미 .....	34
<그림 IV-16> 상하 관계에 있는 두 {진출} .....	34
<그림 IV-17> '진출_02_00_01'의 의미 .....	35
<그림 IV-18> '갈다_03_00_01'의 의미 .....	35
<그림 IV-19> 다른 줄기의 {경운기} .....	35
<그림 IV-20> '기동하다'의 의미 .....	36
<그림 IV-21> {경운기, 전동기}의 상위노드 .....	36
<그림 IV-22> '까먹다_00_00_01'의 용례 .....	37
<그림 IV-23> {도시락}의 위치 .....	37
<그림 IV-24> 최상위노드 목록 .....	38
<그림 IV-25> 최소상계노드 하위어 중 목적격으로 결합되지 못하는 어휘관계 설정 ..	41
<그림 V-1> 부사의 경합관계 구축 도구 .....	46
<그림 V-2> 부사의 결합관계 연결 과정 .....	47

<그림 V-3> 입체적 국어사전 브라우저에서 부사의 결합관계 .....	48
<그림 VI-1> 세종 <사태> 의미부류 .....	59
<그림 VI-2> <표준> 용언과 세종전자사전(용언)의 의미 매핑 도구 .....	61
<그림 VI-3> <표준> 용언과 세종전자사전(용언)의 의미 매핑 결과 .....	62
<그림 VI-4> 용언 의미 군집의 세종 <사태> 의미부류로의 계층화 .....	63
<그림 VI-5> 입체적 국어사전 브라우저에서의 용언 의미 군집 .....	64
<그림 VII-1> 어휘망 편집 브라우저의 구성 요소 .....	65
<그림 VII-2> ‘식용하다’의 검색 결과 .....	67
<그림 VII-3> 공통으로 적용되는 어휘관계 설정 .....	68
<그림 VII-4> ‘혀’와 ‘싸하다’의 주어 관계 설정 .....	69
<그림 VII-5> 어휘를 어휘망에 추가하기 위한 과정 .....	69
<그림 VII-6> 어휘망 내에 ‘가계’ 위치 확인 .....	70
<그림 VII-7> 선정된 어휘 목록 추가 .....	70
<그림 VII-8> 선정된 어휘 목록 수정 .....	71
<그림 VII-9> 선정된 어휘 목록 삭제 .....	71
<그림 VII-10> 입체용 국어사전 브라우저 구성 요소 .....	73
<그림 VII-11> 설정 정보 변경 .....	74
<그림 VII-12> 표제어 검색 .....	75
<그림 VII-13> 검색어 목록에 추가된 검색어 .....	75
<그림 VII-14> 표제어 목록 .....	76
<그림 VII-15> 선택된 어휘(예, ‘독가스’)에 대한 의미 관계 정보 .....	77
<그림 VII-16> 어휘 ‘가스’에 마우스를 올려 뜻풀이 및 용례 보기 .....	78
<그림 VII-17> 상위어 따라가기 .....	79
<그림 VII-18> 30개 이상인 하위어 보기 .....	80
<그림 VII-19> 양방향 추적 .....	81
<그림 VII-20> 최소상계노드의 용언 하위범주 정보 상속 .....	83
<그림 VII-21> 용언 하위범주 정보를 상속 받지 않는 경우 .....	84
<그림 VII-22> 테이블과 필드간의 관계 .....	89

# I. 추가 어휘 선정

## 1. 추가 어휘 선정 지침

본 연구는 “한국어 어휘의미 체계 기반 입체적 기본어휘 사전 구축” 연구과제(2008년)<sup>1)</sup>에서 선정한 다의어 수준 5만 항목에 10,000개의 명사와 7,000개의 용언을 추가 선정하여, 전체 최소 70,000개의 어휘에 대해 명사 어휘망의 구축, 용언의 하위범주 정보를 명사 어휘망의 최소상계노드로 설정하고, 부사와 의미적으로 결합하는 부사 및 용언을 설정하는 것을 목적으로 한다.

2009년에 추가한 명사 10,000개 용언 7,000개의 선정은 2008년 5만 어휘 선정 후 남은 어휘와 세종전자사전을 대상으로, 2008년 설정한 어휘 선정 세부 지침에 따라 선정하였으며 그 세부 내용은 다음과 같다.

- (1) 2008년 기본어휘 선정을 위해 고려했던 다의어 어휘들 중에서 미선정한 어휘를 대상으로 한다.
  - 명 사 : 43,974개
  - 동 사 : 9,235개
  - 형용사 : 2,242개
- (2) 품사정보에 ‘비표준어, 방언, 북한어, 옛말’이 포함되어 있는 어휘는 제외한다.
  - 명 사 : 38,382개
  - 동 사 : 8,297개
  - 형용사 : 1,824개
- (3) (1), (2)번의 과정을 거친 후 선정된 어휘들을 기본 대상으로 추가어휘 선정을 한다.
- (4) 세종사전과 비교하여 세종사전에 등재되어 있는 어휘를 우선 추가어휘에 포함시킨다.
  - 명 사 : 8,884개
  - 동 사 : 5,092개

1) 기본어휘 선정 방법 및 구체적인 선정 지침은 『한국어 어휘의미 체계 기반 입체적 기본어휘 사전 구축 연구』(옥철영2008) 참조

- 형용사 : 1,029개
- (5) '명사, 동사, 형용사'의 부족한 부분은 세종에서의 어휘빈도를 중심으로 추가하도록 한다.
- 명 사(빈도 10이상) : 11,583개
- 동 사(빈도 3이상) : 5,930개
- 형용사(빈도 2이상) : 1,431개

<명사 선정 세부지침>

- (1) 고유명사(인명)는 제외한다.
- (2) 이전 말이나 비속어 등도 '비표준어, 방언, 북한어, 옛말'이 아닌 한 포함한다.
- (3) 뜻을풀이를 살핀 후 '옛말'인 경우 '세종빈도, 2002빈도, 김광해, 연세'에 모두 등재되어 있는 경우에만 추가어휘로 선정한다.

<예> 해관 1 2 : 조선 고종 20년(1883)에 항구에 설치한 관아.

해관 1 3 : 중국 청나라 때에 개항장에 설치한 세관.

해관 1 1 : 항구에 설치한 관문.

상사람 0 0 : 조선 중기 이후에 '평민'을 이르던 말.

- (4) '명사'의 경우 품사정보에 '어근'만 있는 경우는 제외한다.
- (5) 'A(으)로 순화'로 되어 있는 항목은 검토 후 해당 어휘는 제외하고, 'A'에 포함되어 있는 어휘를 추가어휘목록으로 선정한다.

<동사, 형용사 선정 세부지침>

- (1) 용언의 경우 '~의 준말, ~의 원말'에 포함되어 있는 항목 검토 후 추가어휘로 선정한다.
- (2) '~의 사동사, ~의 피동사'에 포함되어 있는 항목은 검토 후 추가 고려한다.(O)
- (3) 뜻을풀이를 살핀 후 '옛말'인 경우 '세종빈도, 2002빈도, 김광해, 연세'에 모두 등재되어 있는 경우에만 추가어휘로 선정한다.
- (4) '~(으)로 순화'로 되어 있는 항목은 검토 후 해당 어휘는 제외하고, '~'에 포함되어 있는 어휘를 추가어휘목록으로 선정한다.

## 2. 최종 선정/추가 어휘

### 2.1. 입체적 국어사전 구축을 위한 선정 어휘

다음 <표 I-1>은 2009년까지 선정한 품사별 기본어휘 수이다.

<표 I-1> 어휘선정 원칙에 따른 품사별 선정 어휘 수

	명사	동사	형용사	부사	합계
2008년 선정	34,384	9,986	3,314	2,640	50,324
2009년 추가	10,035	5,578	1,406	0	17,019
합계	44,419	15,564	4,720	2,640	67,343
비율	66%	23%	7%	4%	100%

위 표에서 2009년에 품사별로 명사 10,087개, 동사 5,593개, 형용사 1,414개가 선정되었으나, <웹용 표준국어대사전>에 등재되지 않은 명사 52개, 동사 15개, 형용사 8개는 제외하였다.

### 2.2. 입체적 국어사전 구축 과정에서 추가된 어휘

입체적 국어사전은

- 명사 어휘망(상위어, 하위어, 반의어, 동의어)
- 용언의 하위범주 정보(<표준>의 문형 기준)
- 부사-부사, 부사-용언 결합정보를 구축하는 작업이다.

명사 어휘망은 울산대학교 사용자어휘지능망(U-WIN)을 기본으로 하여 선정된 명사를 U-WIN에 매핑시켜 상위어, 하위어를 구축하며, <표준>에서 제시하는 반의어(기호 ↔)와 동의어(뜻풀이가 =로 시작)를 추출한다. 용언의 하위범주 정보는 다의어 의미 주석부착(이하, 태깅)된 용례에서 문형의 격에 해당하는 전형적인 명사를 추출하고 이 명사를 포함하는 명사 어휘망에서의 “최소상계노드”를 설정하는 작업이다. 또한, 부사-부사, 부사-용언의 결합정보도 다의어 의미 태깅된 용례에서 결합 부사 및 용언을 추출하여 구축한다.

따라서 2009년까지 선정된 어휘를 기본어휘로 하여 입체적 국어사전을 구축하는 과정에서 명사의 경우 U-WIN의 계층구조에서 선정된 어휘의 상위어가 미선정된 경우 그 상위어를 추가하였으며, 용언과 부사의 경우 미선정된 어휘가 용례에서 추출될 수 있으며 이러한 미선정 용언과 부사는 입체적 국어사전에 최종 포함하였으며, 그 내용은 다음 <표 I-2>와 같다.

<표 I-2> 입체적 국어사전에 포함된 품사별 추가 어휘 수

	구분	선정어휘	추가어휘	합계
명사	상하위어	44,419	2008년 (2,214) 2009년 726	45,145
	동의어	(2,870)	18,318	21,188
	반의어	(809)	613	1,422
	최소상계노드	(4,021)	196	4,217
동사		15,564	171	15,735
형용사		4,720	16	4,736
부사		2,640	16	2,656
합계	선정어휘	67,343	20,056	87,399

※ ( )는 선정 명사 44,419개에 이미 포함된 것임.

## II. 표준국어대사전 뜻풀이 및 용례 태깅

### 1. 표준국어대사전 비교

새로이 개정된 <웹용 표준국어대사전>(이하 <웹용표준>)은 기본어휘 선정 연구가 추진되던 당시(2008년 초) 일반에 아직 공개(2008년 10월 공개)되지 않은 상태여서, 본 연구과제는 2002년에 발표한 <시디용 표준국어대사전>(이하 <시디용표준>)을 대상으로 우선 품사 및 다의어 태깅을 하고, 다의어 태깅한 결과의 빈도를 이용하여 기본어휘를 선정하였다.

그러나, 본 연구과제의 결과물은 새로이 개정된 <웹용표준>에 등재된 어휘를 대상으로 최종 기본어휘를 선정하고 선정된 기본어휘의 뜻풀이를 품사 및 다의어 태깅하여야 하므로, 우선 작업 대상으로 삼은 <시디용표준>과 새로이 개정된 <웹용표준>을 뜻풀이 차원에서 비교하여, 뜻풀이가 다른 것은 <웹용표준>에 맞게 뜻풀이 품사/의미 태깅했다.

아래 <표 II-1>은 두 사전을 뜻풀이 차원에서 비교한 결과이다.

<표 II-1> <시디용표준>과 <웹용표준>의 뜻풀이 차원에서의 비교

	<시디용표준>	<웹용표준>
동형어 개수	423,027	424,027
동형이의어 개수	507,096	513,479
다의어 개수	587,833	590,195
뜻풀이 완전일치	425,324	
유의어(≒) 제거 후 일치	44,271	
부호 변환((1)↔「1」) 후 일치	47,029	
띄어쓰기 제거 후 일치	3,470	
일치되는 뜻풀이 합계	520,094	
뜻풀이 불일치 개수	67,729	70,101
각 사전에서만 등재된 표제어	8,078	9,561

## 2. 뜻풀이 및 용례 태깅

### 2.1. 품사 태깅 원칙

뜻풀이 품사 태깅은 21세기 세종계획 결과물인 지능형 형태소 분석기(고려대)를 이용하여 자동 태깅한 후, 태깅 오류를 수작업으로 수정하였다. 품사 태깅은 다음의 원칙에 의해 수행하였다.

#### (1) 어절의 띄어쓰기는 뜻풀이의 띄어쓰기를 우선

어절의 띄어쓰기는 뜻풀이의 띄어쓰기된 것을 우선으로 한다. 그러나, 복합명사(“AB”)의 경우 의미 태깅 과정에서 어절 A나 B가 표제어로 등재되어 있지 않고 AB가 등재된 경우는 붙여 쓰기를 하여 품사 및 의미 태깅하였다. 예를 들어, 현재 표준국어대사전에서 ‘현가’는 現價, 絃歌, 懸枷, 顯加의 4개의 표제어로, ‘현가장치’는 懸架裝置로 등재되어 있다. 그러나 뜻풀이에서는 ‘현가 장치’로 띄어쓰기되어 있고 의미적으로 ‘懸架裝置’가 맞을 경우 “현가장치/NNG”로 붙여 쓰기를 하여 품사 및 의미 태깅하였다.

#### (2) 품사는 등재된 표제어 우선 : 어근(XR)의 품사 태깅

어근 ‘A’와 용언 ‘A하다’가 등재되어 있으며, ‘A’의 뜻풀이가 “‘A하다’의 어근”된 경우는 ‘A하다’로 품사 및 의미 태깅하였다. 예를 들어, 명사 ‘가공<sup>3</sup>(可覿)’의 뜻풀이는 “가공하다<sup>0</sup>의 어근”으로, 형용사 ‘가공하다<sup>3</sup>’의 뜻풀이는 “두려워하거나 놀랄 만하다”로 각각 등재되어 있다. 이 때, “가공할 무기”에서 ‘가공할’은 “가공하다\_003000/VA+ㄷ/ETM”으로 품사 및 의미 태깅하였다.

그러나, “간략화한 형식”에서 ‘간략<sup>2</sup>(簡略)’은 “간략하다의 어근”으로 “간략화한”은 “간략\_020000/NNG+화/XSN+하/XSV+ㄴ/ETM”로 어근 ‘간략’의 품사 태거를 XR이 아닌 NNG로 태깅하였다. 그렇지만, 어근 데이터만을 추출해야 할 경우 의미 태거번호를 이용하여 별도로 추출할 수 있다.

등재된 표제어의 품사를 우선하는 원칙으로, 부사 ‘가공스레’의 뜻풀이는 “가공스럽다”로 등재되어 있어 “가공스레/MAG”와 같이 부사로 태깅하였다.

#### (3) 합성명사의 경우도 원칙 (2)에 따라서, 표제어로 등재된 합성명사인 경우는 등재된 표제어의 품사로 태깅하였으며, 접사가 붙은 합성명사의 경우는 분리하여 품사 및 의미 태깅하였다. 예를 들어, ‘상대적’은 [명사][관형사]로 등

채되어 있어 “상대적/NNG”로 태깅하였으며, “대대적인 환영을”에서 ‘대대적인’은 “대대\_\_888888/NNG+적/XSN+이/VCP+ㄴ/ETM”으로 태깅하였다.

#### (4) 고유명사(NNP)의 태깅

등재된 어휘의 품사가 [명사]인 경우라도 전문용어 필드가 {고유명사}, {지명}, {인명}, {책명}인 것은 모두 고유명사(NNP)로 태깅하였다. 예를 들어, ‘홍성<sup>1</sup>(洪城)’은 전문용어 필드에 {지명}으로 등재되어 “홍성\_\_010000/NNP”로 태깅하였다.

반면, 전문용어 필드에 {인명}으로 등재되어 있지 않지만 명확히 사람 이름인 경우도 고유명사로 태깅하였다. 예를 들어 ‘예수<sup>6</sup>’은 “기독교의 창시자”로 전문용어 필드에는 {기독교}만으로 되어 있지만 “예수\_\_060000/NNP”로 태깅하였다.

#### (5) 의존명사(NNB)의 태깅

등재된 어휘의 품사가 [명사][의존]의 경우 수사나 관형사 다음에 사용되어 “단위”를 나타내는 경우는 의존명사(NNB)로 태깅하였다. 예를 들어, “삼 일 동안 계속”에서 ‘일’은 “일\_\_007002/NNB”로 태깅하였다.

그렇지만, “단위”를 나타내는 경우라도 품사가 [의존]이 아닌 경우는 모두 NNG로 태깅하였다. 예를 들어, ‘그릇’의 경우 금성출판사 국어사전에는 “그릇에 담긴 것의 분량을 세는 단위”로 의존명사로 등재되어 있으나, 표준국어대사전에는 [명사]로만 등재되어 있어 모두 NNG로 태깅하였다.

#### (6) 수사(NR)와 관형사(MM)의 태깅

등재된 어휘의 품사가 [수사][관형사]로 등재되어 있는 경우 사용 형태에 따라서 태깅하였다. 예를 들어 ‘서른’의 경우 [수사][관형사]로 등재되어 있으며, “서른이다”는 “서른/NR+이/VCP+다/EF”로, “서른 살”은 “서른/MM 살/NNB”로 태깅하였다.

그러나, 다의어로 [수사]와 [관형사]로 별도로 각각 등재되어 있는 경우, 예를 들어, ‘사<sup>11</sup>(四)’은 ‘사<sup>11</sup>①’은 [수사]로 “삼에 일을 더한 수”, ‘사<sup>11</sup>②’은 [관형사]로 “그 수량이 넷임을 나타내는 말”, ‘사<sup>11</sup>③’은 [관형사]로 “그 순서가 네 번째임을 나타내는 말”로 “사십오”는 태깅 작업자 관점에 따라 수사 혹은 관형사로 다르게 태깅되어 있을 수 있다.

## 2.2. 의미 태깅 원칙

품사 태깅된 뜻풀이의 다의어 수준의 의미 태깅은 다음의 원칙에 의해 수행하였다.

### (1) 다의어/동형이의어의 의미 태깅

뜻풀이에 사용된 모든 명사 및 용언(보조용언 포함)에 대해 의미 태깅하였다. 현재 의미 태깅된 단어가 접사인 경우도 일부 의미 태깅되어 있다.

### (2) 의미 태깅 번호

표준국어대사전의 동형이의어/다의어는 다음과 같은 의미 태깅 번호를 부착하였다.

- 동형이의어 번호 : 표준국어대사전의 동형이의어 어깨번호와 동일한 번호를 두 자리로 부여하였다.
- 문형 번호 : 용언의 경우 용언의 문형에 따라 별도의 문형번호가 부여되어 있으며 <웹용표준>의 문형번호에 맞게 두 자리로 부여하였다.
- 다의어 번호 : 표준국어대사전의 뜻풀이에 기술된 문형별 번호 순서대로 두 자리로 부여하였다.

예를 들어, <표 II-2>는 ‘먹다’의 동형이의어, 문형번호 및 다의어 번호를 나타내고 있으며, 각 뜻풀이 품사 및 의미 태깅한 결과를 보이고 있다. ‘먹다’는 두 개의 동형이의어로 등재되어 있으며, ‘먹다02’의 경우 세 가지의 문형으로 구분되어 있으며 각 문형에 따라 다의번호가 부여되어 있다. 용언의 하위범주 정보는 <표준>에 등재된 문형을 기준으로 구축하였다.

<표 II-2> ‘먹다’의 표준국어대사전 뜻풀이

동형 이의어 번호	문형 번호 (문형)	다의 번호	뜻풀이
			품사 태깅 및 의미 태깅한 뜻풀이
01	00 (…을)	00	귀나 코가 막혀서 제 기능을 하지 못하게 되다. 귀__010001/NNG+나/JX 코__010001/NNG+가/JKS 막히다__000300/VV+어서/EC 자기__040001/NNG+의/JKG 기능__010000/NNG+을/JKO 하다__010107/VV+지/EC 못하다__000300/VX+게/EC 되다__010304/VV+다/EF+./SF
			음식 따위를 입을 통하여 배 속에 들여보내다. 음식__000001/NNG 따위__000001/NNB+를/JKO 입__000001/NNG+을/JKO 통하다__000702/VV+아/EC 배__010001/NNG 속__010002/NNG+에/JKB 들여보내다__000101/VV+다/EF+./SF
02	01 …을	02	담배나 아편 따위를 피우다.

			담배__000003/NNG+나/JX 아편__010000/NNG 따위__000001/NNB+를/JKO 피우다__010005/VV+다/EF+./SF
		03	연기나 가스 따위를 들이마시다.
			연기__090000/NNG+나/JX 가스__000001/NNG 따위__000001/NNB+를/JKO 들이마시다__020000/VV+다/EF+./SF
		04	어떤 마음이나 감정을 품다.
			어떤__000003/MM 마음__010002/NNG+이나/JX 감정__060000/NNG+을/JKO 품다__010200/VV+다/EF+./SF
		05	일정한 나이에 이르거나 나이를 더하다.
			일정하다__000004/VA+ㄴ/ETM 나이__010000/NNG+에/JKB 이르다__010001/VV+거나/EC 나이__010000/NNG+를/JKO 더하다__000201/VV+다/EF+./SF
		06	욕, 핀잔 따위를 듣거나 당하다.
			욕__020001/NNG+./SP 핀잔__010000/NNG 따위__000001/NNB+를/JKO 듣다__010102/VV+거나/EC 당하다__010100/VV+다/EF+./SF
		07	(속되게) 뇌물을 받아 가지다.
			(/SS+속되다__000001/VA+게/EC+)/SS 뇌물/NNG+을/JKO 받다__010101/VV+아/EC 가지다__000102/VV+다/EF+./SF
		08	수익이나 이문을 차지하여 가지다.
			수익__010002/NNG+이나/JX 이문__060001/NNG+을/JKO 차지하다__010001/VV+아/EC 가지다__000102/VV+다/EF+./SF
		09	물이나 습기 따위를 빨아들이다.
			물__010001/NNG+이나/JX 습기__020000/NNG 따위__000001/NNB+를/JKO 빨아들이다__000001/VV+다/EF+./SF
		10	어떤 등급을 차지하거나 점수를 따다.
			어떤__000003/MM 등급__000101/NNG+을/JKO 차지하다__010001/VV+거나/EC 점수__060001/NNG+를/JKO 따다__010305/VV+다/EF+./SF
		11	구기 경기에서, 점수를 잃다.
			구기__110000/NNG 경기__110000/NNG+에서/JKB+./SP 점수__060001/NNG+를/JKO 잃다__000111/VV+다/EF+./SF
		12	(속되게) 여자의 정조를 유린하다.
			(/SS+속되다__000001/VA+게/EC+)/SS 여자__020001/NNG+의/JKG 정조__090002/NNG+를/JKO 유린하다/VV+다/EF+./SF
		13	매 따위를 맞다.
			매__010000/NNG 따위__000001/NNB+를/JKO 맞다__030001/VV+다/EF+./SF
		14	남의 재물을 다루거나 맡은 사람이 그 재물을 부당하게 자기의 것으로 만들다.
			남__010001/NNG+의/JKG 재물__030001/NNG+을/JKO 다루다__010101/VV+거나/EC 맡다__010102/VV+은/ETM 사람__000001/NNG+이/JKS 그__010102/MM 재물__030001/NNG+을/JKO 부당하다/VA+게/EC 자기__040001/NNG+의/JKG 것__010001/NNB+으로/JKB 만들다__000200/VV+다/EF+./SF
	02 ...에	01	날이 있는 도구가 소재를 깎거나 자르거나 갈거나 하는 작용을 하다.
			날__020000/NNG+이/JKS 있다__010405/VA+는/ETM 도구__100001/NNG+가/JKS 소재__050001/NNG+를/JKO 깎다__000101/VV+거나/EC 자르다__010101/VV+거나/EC 갈다__020101/VV+거나/EC

			하다_011001/VV+는/ETM 작용_010001/NNG+을/JKO 하다_010101/VV+다/EF+./SF
		02	바르는 물질이 배어들거나 고루 퍼지다. 바르다_010200/VV+는/ETM 물질_020001/NNG+이/JKS 배어들다_000001/VV+거나/EC 고루_010002/MAG 퍼지다_000201/VV+다/EF+./SF
		02	벌레, 균 따위가 파 들어가거나 퍼지다. 벌레_010001/NNG+./SP 균_030001/NNG 따위_000001/NNB+가/JKS 파다_010101/VV+아/EC 들어가다_010101/VV+거나/EC 퍼지다_000201/VV+다/EF+./SF
		04	돈이나 물자 따위가 들거나 쓰이다. 돈_010001/NNG+이나/JX 물자_020000/NNG 따위_000001/NNB+가/JKS 들다_010201/VV+거나/EC 쓰이다_030101/VV+다/EF+./SF
	03	00	앞말이 뜻하는 행동을 강조하는 말. 주로 그 행동이나 그 행동과 관련된 상황이 마음에 들지 않을 때 쓴다.
			앞말_010001/NNG+이/JKS 뜻하다_000200/VV+는/ETM 행동_000001/NNG+을/JKO 강조하다/VV+는/ETM 말_010004/NNG+./SF 주로_010000/MAG 그_010102/MM 행동_000001/NNG+이나/JX 그_010102/MM 행동_000001/NNG+과/JKB 관련되다/VV+ㄴ/ETM 상황_020000/NNG+이/JKS 마음_010004/NNG+에/JKB 들다_010206/VV+지/EC 않다_000200/VX+을/ETM 때_010005/NNG 쓰다_030303/VV+ㄴ다/EF+./SF

(3) 태깅 대상 명사의 의미가 뜻풀이에 없는 경우 : 복합명사

2.1.의 품사 태깅 원칙(1)에서와 같이 복합명사(“AB”)의 경우 의미 태깅 과정에서 어절 A의 의미가 표제어로 등재되어 있지 않고 AB가 등재된 경우는 붙여 쓰기를 하여 품사 및 의미 태깅하였다. 예를 들어, 현재 표준국어대사전에서 ‘현가’는 現價, 絃歌, 懸枷, 顯加의 4개의 표제어로, ‘현가장치’는 懸架裝置로 등재되어 있다. 그러나 뜻풀이에서는 ‘현가 장치’로 띄어쓰기되어 있고 의미적으로 ‘懸架裝置’가 맞을 경우 “현가장치/NNG”로 붙여 쓰기를 하여 품사 및 의미 태깅하였다.

(4) 태깅 대상 명사의 의미가 뜻풀이에 없는 경우 : 대등어의 준말

사전 뜻풀이에서 어떤 어휘를 설명하면서 연속된 대등어인 경우 줄여 쓰는 경우가 있다. 예를 들어, “동물, 식물”을 “동·식물”로, “육군, 해군 및 공군”을 “육·해·공군”으로 쓴다. 여기서 특수문자 ‘·’는 품사 태깅 과정에서 ‘,’로 변환되어 “동/NNG+./SS 식물/NNG” 혹은 “육/NNG+./SS 해/NNG+./SS 공군/NNG”으로 품사 태깅되어 있으며, 이를 의미 태깅하는 과정에 ‘동’, ‘육’, ‘해’의 적합한 의미가 발견되지 않는다. 이럴 경우 “동물/NNG”, “육군/NNG”, “해군/NNG”로 변환하여 의미 태깅하였다.

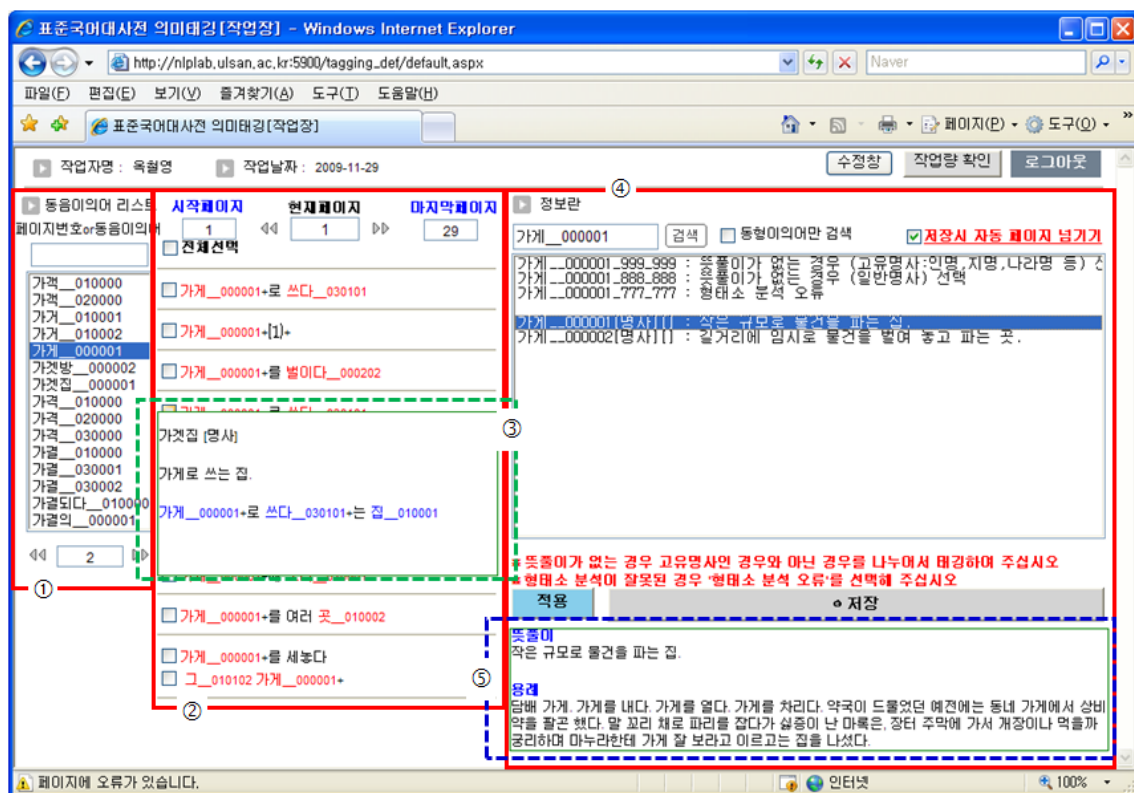
(5) 태깅 대상 명사의 의미가 뜻풀이에 없는 경우 : 누락 의미

태깅 대상 명사가 동형이의어/다의어이면서 문맥에 맞는 해당 의미가 없는 경우, 고유명사의 경우는 “999999”로 일반명사의 경우는 “888888”로 의미 태깅하였다. 고유명사의 경우는 주로 사람 이름에 대한 뜻풀이에서 ‘자’, ‘호’의 경우가 대부분이며, 일반 명사의 경우는 사전 편찬 과정에 누락된 의미이거나 표제어일 경우가 대부분이다. 특히 1 음절 한자에 대해 등재되지 않은 것이 많았다.

## 2.3. 동형이의어/다의어 의미 태깅 도구

### 2.3.1. 뜻풀이 의미 태깅 도구

<그림 II-1>은 다의어(동형이의어 포함) 태깅 대상 개별 어휘가 뜻풀이에 나타나는 모든 어휘를 대상으로 이를 태깅하는 도구의 화면이다.



<그림 II-1> 의미 태깅 대상 어휘를 가진 모든 뜻풀이 태깅 도구

<그림 II-1>의 박스 ①은 의미 태깅 대상 동형이의어/다의어(용언 포함)를 15개씩 리스트하고 있으며, 태깅 대상 동형이의어/다의어를 선택하면 선택된 동형이의어/다의어를 포함하는 뜻풀이를 10개 단위의 페이지로 출력하는 박스 ②가 나타난다. 그림에 나타난 박스 ①은 태깅이 완료된 상태의 태깅 대상 개별 어휘가 나타난 상황이다.

박스 ②의 상단에는 의미 태깅 대상 동형이의어/다의어를 포함하는 뜻풀이를 10개 단위의 페이지 개수를 보여준다. 뜻풀이 하나에 커서를 놓으면, 표제어, 뜻풀이, 품사, 원어, 전문용어, 의미 태깅된 결과를 보여 주는 박스 ③이 돌출창(팝업창)으로 나타난다. 실제 박스 ③에는 표제어 “가겟집”의 뜻풀이에 태깅 대상 어휘 ‘가게’가 있음을 확인할 수 있다.

박스 ①의 태깅 대상 동형이의어/다의어를 선택하면(예, ‘가게’), 박스 ④에 대상 동형이의어/다의어의 모든 의미를 의미 순(어깨번호\_문형\_다의번호)으로 출력하여 문맥에 맞는 의미를 선택할 수 있도록 하며, 해당 의미의 뜻풀이, 용법 및 용례가 박스 ⑤에 나타난다.

박스 ②에서 태깅할 뜻풀이 옆의 체크 박스를 체크하여 선택하고 박스 ④에서 해당 의미를 선택하여 적용 버튼을 누르거나 더블클릭하면 해당 의미 번호(예, “가게\_000001”)로 의미 태깅되면서 파란색으로 바뀌다. 이렇게 의미 태깅이 완료된 후 저장 버튼을 누르면 의미 태깅된 뜻풀이 데이터베이스에 저장된다.

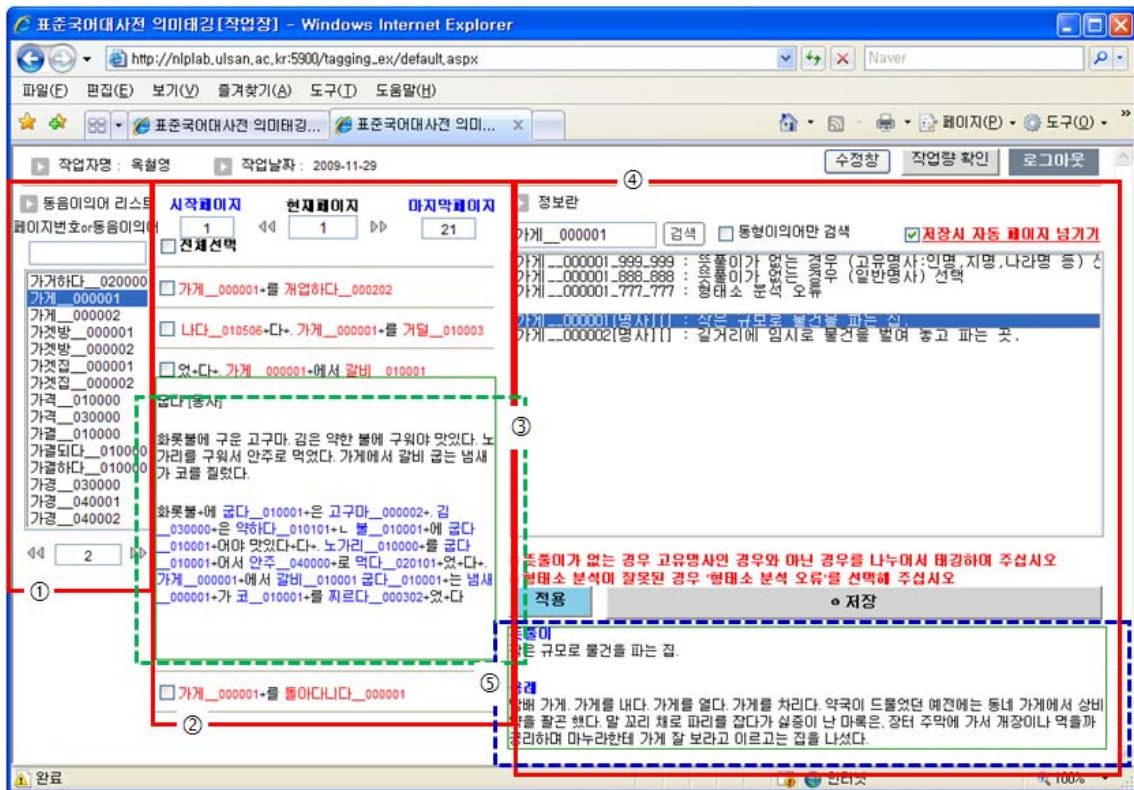
전체 뜻풀이를 대상으로 의미 태깅 대상 어휘는 다음과 같다.

- 다의어 수준 개별 어휘 수 : 107,306개
- 다의어 수준 개별 어휘의 총 수 : 3,610,106개

### 2.3.2. 용례 의미 태깅 도구

<그림 II-2>는 다의어(동형이의어 포함) 태깅 대상 개별 어휘가 용례에 나타나는 모든 어휘를 대상으로 이를 태깅하는 도구의 화면으로, 2.3.1과 태깅하는 방법은 같다.

박스 ③에는 표제어 ‘굽다’의 용례에 태깅 대상 어휘 ‘가게’가 나타남을 확인할 수 있다.



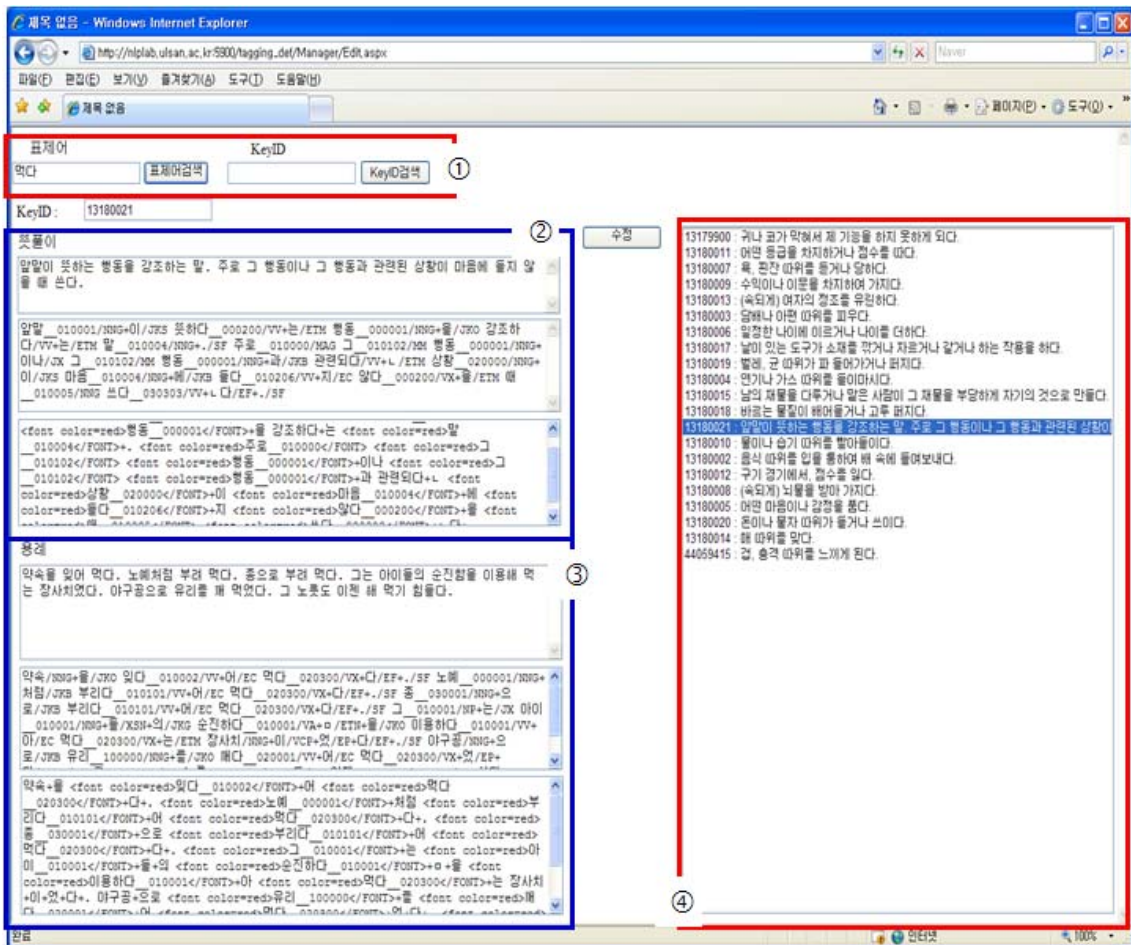
<그림 II-2> 의미 태깅 대상 어휘를 가진 모든 용례 태깅 도구

전체 용례에 나타나는 의미 태깅 대상 어휘는 다음과 같다.

- 다의어 수준 개별 어휘 수 : 78,083개
- 다의어 수준 개별 어휘의 총 수 : 1,121,759개

## 2.4. 태깅 오류 직접 수정 도구

<그림 II-3>은 <그림 II-1> 혹은 <그림 II-2>의 의미 태깅 작업 도중 발견된 품사 및 의미 태깅 오류를 직접 수정할 수 있는 도구의 화면으로, <그림 II-1> 과 <그림 II-2>의 오른쪽 상단의 수정창 버튼을 누르면 실행된다.



<그림 II-3> 태깅 오류 직접 수정 도구

<그림 II-3>의 박스 ①에서 수정할 표제어를 입력하면, 입력한 표제어의 뜻풀이 전체가 박스 ④에 나타난다. 박스 ④에서 수정할 뜻풀이를 선택하면 박스 ② ~ ③에 현재 태깅된 상태가 나타난다. 박스 ①에서 KeyID는 데이터베이스에서 각 표제어를 관리하는 번호이며, 직접 KeyID를 입력하여 수정할 뜻풀이를 선택할 수도 있다.

박스 ②는 뜻풀이에 대한 태깅 결과이며, 박스 ③은 용례에 대한 태깅 결과이다. 품사 혹은 의미 태깅된 결과를 수정할 경우 각 박스를 직접 수정한 후, 수정 버튼을 누르면 수정된다.

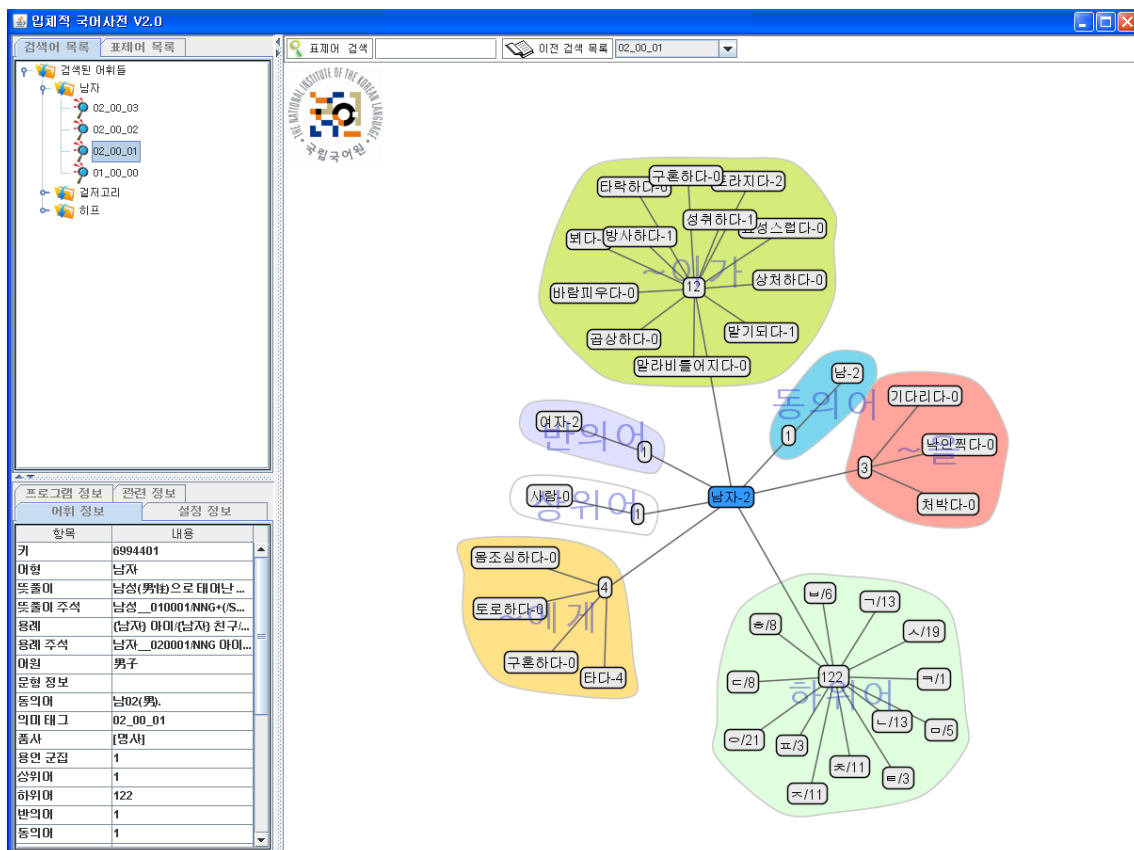
### III. 명사 어휘망 구축

#### 1. 명사 어휘망

본 연구과제에서 구축할 명사 어휘망(상위어, 하위어)은 기본적으로 울산대에서 구축한 30만 어휘의 U-WIN<sup>2)</sup>을 기반으로 하여 구축하였다.

선정 명사 어휘에 대한 동의어는 뜻풀이가 '='로 시작되는 일반 동의관계만을 <표준>에서 추출하였고, 반의어는 뜻풀이에 '↔'로 표시된 반의어만을 추출하였다.

이렇게 구축된 상위어, 하위어, 동의어 및 반의어는 <그림 III-1>과 같이 입체적 국어사전 브라우저를 통해 검색된다.



<그림 III-1> 입체적 국어사전에서의 명사 어휘망

2) U-WIN의 자세한 구조 및 구축 원리는 최호섭(2007)을 참조.

## 2. 선정 어휘의 U-WIN 매핑

본 연구과제에서 구축할 명사 어휘망은 기본적으로 선정된 기본어휘를 반영하는 어휘망을 구축한다. 그러나 선정된 기본어휘만으로 어휘망을 구축할 경우 비균형적인 어휘망이 구축될 가능성이 높고, 차후 어휘망의 확장을 고려할 경우 전체 어휘망의 틀이 매번 바뀌어야 하는 문제가 있다. 따라서 이번 연구에서 구축할 명사 어휘망은 울산대에서 구축한 30만 어휘의 U-WIN을 기반으로 하여 구축한다.

그렇지만 이 경우도, 선정된 기본어휘(특히 웹용 표준국어대사전에 새로이 추가된 어휘)가 U-WIN에 포함되지 않을 경우 이는 적절히 반영하여 구축한다.

구축한 입체적 국어사전 브라우저에서는 선정된 기본어휘들 간에 반의어 및 동의어 관계가 있을 경우뿐만 아니라, 선정되지 않은 반의어와 동의어도 포함하였다.

다음 <표 III-1>은 선정된 기본어휘를 U-WIN에 매핑한 결과이다.

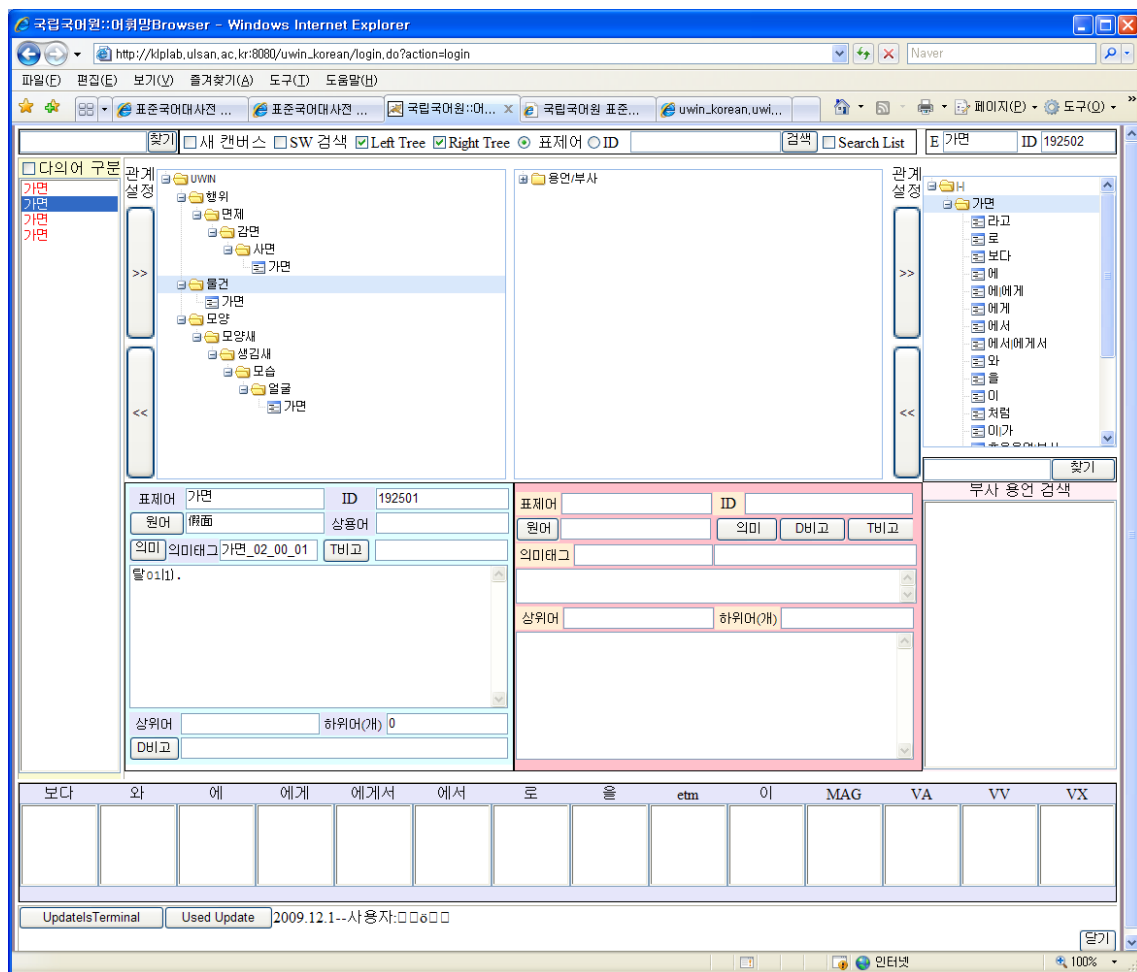
<표 III-1> U-WIN상에서의 기본어휘의 계층구조

기본어휘	원어	어깨 번호	다의 번호	레 벨	U-WIN 계층구조
가급적	可及的	0	1		
가까이		0	4	4	1공간_2곳_3부근
가랑잎		1	0	9	1물건_2물체_3물질_4조직_5기관_6영양기관(X)_7잎_8나뭇잎
가면	假面	2	1	5	1모양_2모양새_3생김새_4모습_5얼굴
가면	假面	2	2	2	1물건
가는귀		0	1	9	1물건_2물체_3물질_4조직_5기관_6감각기관(X)_7청각기(X)_8귀
거세	去勢	5	1	3	1행위_2제거
거세	去勢	5	2		1행위_2수탈_3갈취_4강탈(X)
광고주	廣告主				
사령부	司令部			4	1집단_2조직_3본부
입안	立案	2	2	7	1종류_2종목_3항목_4사항_5안건_6입안
신	神	9	3	8	1존재_2실재_3존재_4존재_5신_6유일신_7하느님
소출	所出	2		5	1정도_2수량_3양_4수확량

수재	水災	2		10	1작용_2영향_3해_4손해_5손해_6손실_7피해_8재해_9기상재해(X)
감명	感銘			11	1인지_2마음_3감정_4기분_5느낌_6흥감(X)_7감동_8감응_9감심(X)_10감격
경칩	驚蟄			7	1시간_2시점_3때_4시기(X)_5계절_6절기
살모사	殺母蛇			9	1생물_2동물_3후생동물(X)_4척추동물_5파충류_6뱀_7독사_8살무사
주위	周圍	2	1	7	1범위_2부분_3끝_4가장자리_5테두리_6둘레
엑스레이	X-ray		1	6	1모양_2현상_3과동_4전자기파_5광선
단어	單語			4	1기호_2언어_3말
최상급	最上級		1	6	1과정_2단계_3등급_4급_5상급

<표 III-1>에서 ‘가급적’은 <시디용표준>에 등재되었으나 U-WIN에 연결되지 않은 어휘이며, ‘광고주’는 <웹용표준>에 새로이 등재된 어휘이다. 또한 ‘가는귀’의 U-WIN 계층구조(“1물건\_2물체\_3물질\_4조직\_5기관\_6감각기관(X)\_7청각기(X)\_8귀”)에서 “감각기관, 청각기”는 U-WIN에는 연결된 어휘이나 기본어휘로 선정 기준(전문용어, 복합어 등)에 따라 선정되지 않은 어휘(2,214개)이다. 이렇게 U-WIN에 연결된 어휘가 선정되지 않은 경우는 기본어휘로 추가하였다.

<표 III-1>에서 ‘가면(假面)’은 기본어휘 선정 과정에서 두 개의 의미가 선정되었으며, 각 의미별로의 계층구조는 다음 <그림 III-2>와 같이 웹용 어휘망 편집 브라우저에서 확인할 수 있다.



<그림 III-2> '가면'의 계층구조(어휘망 구축 도구)

## IV. 용언의 하위범주 정보 구축

### 1. 하위범주 정보

#### 1.1. 논항정보 구축 대상

2009년에 구축하는 용언의 하위범주 정보 구축은 2008년 ~ 2009년에 선정된 용언을 대상으로 한다. 2008년에 세종전자사전을 이용하여 구축한 용언의 하위범주 정보(2008년 최종보고서 참조)에서 발견되는 오류를 수정하기 위해, 2009년 논항정보 구축 대상에 포함하여 다시 검증하였다.

용언의 하위범주 정보 구축은 동사와 형용사의 논항정보를 구축하는 작업으로, 논항정보는 2008년에 구축되고 2009년에 확장 구축된 명사 어휘망을 기본으로 하여 연결한다. 논항(argument)은 문장의 의미를 구성하는 데에 참여하는 명사를 가리킨다. 따라서 부사는 논항정보 연결 작업에서는 제외되고 별도의 작업(V장 참조)으로 구축하였다.

<표 IV-1> 구축 대상 용언

어휘	2008년 선정 용언		2009년 선정 용언	
품사	동사	형용사	동사	형용사
개수	9,986	3,314	5,749	1,422
합계	13,300		7,171	
총합	20,471			

#### 1.2. 논항정보 구축 절차

2009년 용언의 하위범주 정보는 다음의 절차에 따라 구축하였다.

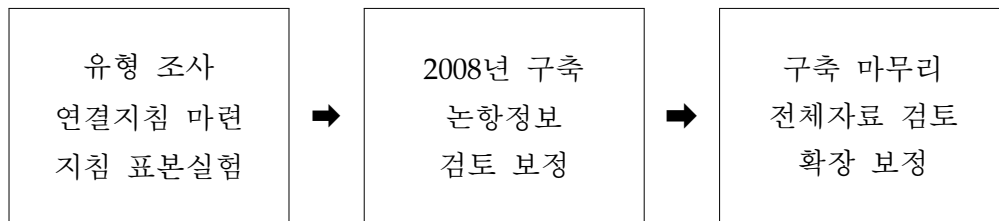
- ① 구축 용언 리스트 확인
- ② 명사 어휘망 노드 검토
- ③ 표본 추출 실험

- ④ 구축 지침 설정
- ⑤ 재검증과 검증
- ⑥ 마무리 점검 및 보정이다.

2008년 ~ 2009년 선정 용언 전체의 하위범주 정보를 구축하기 위해서는 일관된 지침이 필요하다. <표 IV-2>의 구축 과정에서 지침 마련은 중요한 선행 작업이다.

논항정보 구축 지침에 의거하여, 실제 용언의 하위범주 정보 구축 작업이 이루어졌다. 구축 지침은 2008년 ~ 2009년에 선정된 전체 용언에 공통 적용하였다. 2008년에 선정된 용언의 논항정보는 새로 마련된 지침에 맞게 재검증하여 다시 연결하고, 2009년에 선정된 용언의 논항정보도 동일한 지침으로 연결하였다. 보정 작업도 지침에 의거하여 실행되었다.

<표 IV-2> 구축 작업 순서



## 2. 용언의 하위범주 정보 구축 과정

### 2.1. 용언의 하위범주 정보 추출

하위범주 정보는 <표준>에서 기술된 문형정보를 기준으로 하여 구축하였다. 예를 들어 '먹다'의 의미별 문형정보 및 뜻풀이, 용례는 다음 <표 IV-3>과 같다.

<표 IV-3> '먹다'의 뜻풀이 및 용례(표준국어대사전)

표제어	품사	문형	뜻풀이
			용례
먹다01	동사	(...을)	<p>귀나 코가 막혀서 제 기능을 하지 못하게 되다.</p> <p>코 <b>먹은</b> 소리를 내다/이 사람은 귀가 <b>먹어서</b> 잘 못 들으니까 큰 소리로 말씀하셔야 돼요. ॥ 귀를 <b>먹었는지</b> 아무리 불러도 그냥 지나가더라.</p>
먹다02	동사	...을	<p>「1」 음식 따위를 입을 통하여 배 속에 들여보내다.</p> <p>밥을 <b>먹다</b>/술을 <b>먹다</b>/약을 <b>먹다</b>/물을 <b>먹다</b>/음식을 배불리 <b>먹다</b>/닭이 모이를 <b>먹다</b>/몸이 약해진 누나는 보약을 몇 차례나 <b>먹어도</b> 늘 골골거렸다.</p> <p>「2」 담배나 아편 따위를 피우다.</p> <p>담배를 <b>먹다</b>/아편을 <b>먹다</b>/정전이 야단 통에 두어 시간이나 담배를 끊었다가 <b>먹는</b> 담배라 머리가 아뜩한 것 같았다.《이광수, 흙》</p> <p>「3」 연기나 가스 따위를 들이마시다.</p> <p>연탄가스를 <b>먹다</b>/탄내를 <b>먹다</b>.</p> <p>「4」 어떤 마음이나 감정을 품다.</p> <p>양심을 <b>먹고</b> 투서를 하다/한번 <b>먹은</b> 마음이 변하지 않도록 하자./나는 마음을 독하게 <b>먹고</b> 그녀를 외면하였다.</p> <p>「5」 일정한 나이에 이르거나 나이를 더하다.</p> <p>네 살 <b>먹은</b> 아이/나이를 <b>먹다</b>/내년이면 삼십을 <b>먹는</b>구나.</p> <p>「6」 욕, 편견 따위를 듣거나 당하다.</p> <p>하루 종일 욕만 되게 <b>먹었네</b>./그래도 그는 속없는 소리를 하다가 가끔 편견을 <b>먹는</b> 것이었다.《이기영, 봄》</p> <p>「7」 (속되게) 뇌물을 받아 가지다.</p> <p>뇌물을 <b>먹다</b>/뇌물을 <b>먹고</b> 탈세를 눈감아 주다.</p> <p>「8」 수익이나 이문을 차지하여 가지다.</p> <p>남은 이익은 모두 내가 <b>먹어라</b>./시세가 마침 좋은 것 같아서 쌀을 붙였던 것인데 천 원을 <b>먹기는</b> 고사하고 본전 육백 원이 다 달아난 판이니...《채만식, 타류》</p> <p>「9」 물이나 습기 따위를 빨아들이다.</p> <p>기름 <b>먹은</b> 종이/김이 습기를 <b>먹어</b> 눅눅해졌다./숨이 물을 <b>먹어</b> 무겁다.</p> <p>「10」 어떤 등급을 차지하거나 점수를 따다.</p> <p>1등을 <b>먹다</b>/우승을 <b>먹다</b>/100점을 <b>먹다</b>/체육 대회에서 우리 반이 일 등을 <b>먹었다</b>.</p> <p>「11」 구기 경기에서, 점수를 잃다.</p> <p>상대편에게 먼저 한 골을 <b>먹었다</b>.</p> <p>「12」 (속되게) 여자의 정조를 유린하다.</p> <p>그는 벌써 여러 여자를 <b>먹었다</b>.</p> <p>「13」 매 따위를 맞다.</p> <p>상대의 쎈 주먹을 한 방 <b>먹고</b> 나가떨어졌다.</p> <p>「14」 남의 재물을 다루거나 맡은 사람이 그 재물을 부당하게 자기의 것으로 만들다.</p> <p>경리 직원이 회사의 공금을 <b>먹었다</b>.</p>
		...에	<p>「1」 날이 있는 도구가 소재를 깎거나 자르거나 갈거나 하는 작용을 하다.</p> <p>이 고기에는 칼이 잘 <b>먹지</b> 않는다./대패가 잘 <b>먹는다</b>.</p> <p>「2」 바르는 물질이 배어들거나 고루 퍼지다.</p> <p>옷감에 풀이 잘 <b>먹어야</b> 다림질하기가 좋다./얼굴에 화장이 잘 <b>먹지</b> 않고 들뜬다.</p> <p>「3」 벌레, 균 따위가 파 들어가거나 퍼지다.</p> <p>사과에 벌레가 많이 <b>먹었다</b>./옷에 좀이 <b>먹어</b> 못 입게 되었다./얼굴에 버짐이 <b>먹다</b>.</p> <p>「4」 돈이나 물자 따위가 들거나 쓰이다.</p>

			공사에 철근이 생각보다 많이 먹어 걱정이다./낡은 집 수리에는 자칫 새로 짓는 것보다 비용이 더 먹을 수 있다.
--	--	--	--

용언의 문형에 대한 전형적인 하위범주 정보는 기본적으로 각 뜻풀이의 용례에 기술되어 있으며, 본 과제에서는 II 장과 같이 다의어 수준까지 의미 태깅된 뜻풀이 및 용례에서 각 용언의 하위범주 정보 추출하였다.

뜻풀이/용례의 문장이 “.....(1)..... [용언\_1] .....(2)..... [용언\_2]”와 같을 경우,

- [용언\_1]의 문형의 하위범주 정보는 .....(1).....에서
- [용언\_2]의 문형의 하위범주 정보는 .....(2).....에서 추출하였다.

예를 들어, 뜻풀이 및 용례에서 사용된 ‘갈다’의 경우 다음 <표 IV-4>와 같은 하위범주 정보를 추출할 수 있다.

<표 IV-4> 다의어 태깅된 뜻풀이 및 용례에서 추출한 ‘갈다’의 하위범주 정보

용언	용언keyID	명사	명사keyID	빈도	격조사	비고	용언	용언keyID	명사	명사keyID	빈도	격조사	비고
갈다	702001	건전지	1632800	1	을	0	갈다	702001	불	17976001	1	으로	0
갈다	702001	공기	3213401	5	을	1	갈다	702001	불씨	18105001	1	을	0
갈다	702001	관	3582800	1	을	0	갈다	702001	성	21318200	2	을	0
갈다	702001	구두창	4120300	1	을	0	갈다	702001	술	23024200	1	을	0
갈다	702001	굽	4705804	1	을	0	갈다	702001	스토리	23211100	1	을	0
갈다	702001	굽창	4717200	1	을	0	갈다	702001	시트	23620101	1	을	0
갈다	702001	금상	5352700	1	을	0	갈다	702001	신발창	23803101	1	을	0
갈다	702001	기저귀	5826100	2	으로	1	갈다	702001	심지	24173701	1	을	0
갈다	702001	기저귀	5826100	1	을	0	갈다	702001	앞니	25421300	2	을	0
갈다	702001	깃	6062201	1	을	0	갈다	702001	애자	25530700	1	을	0
갈다	702001	깃털	6072100	1	을	0	갈다	702001	얼음주머니	26304300	1	을	0
갈다	702001	깎창	6151300	1	을	0	갈다	702001	연탄	27005100	2	을	1
갈다	702001	깍데기	6247102	1	을	0	갈다	702001	연탄불	27005700	2	을	1
갈다	702001	나무못	6632000	1	을	0	갈다	702001	유리	29535300	2	을	0
갈다	702001	넌빤지	7387800	1	을	0	갈다	702001	유리창	29545900	1	을	0
갈다	702001	돌	10059304	1	을	0	갈다	702001	이	30317701	1	을	0
갈다	702001	뒤창	10734600	1	을	0	갈다	702001	이빨	30494800	1	을	0
갈다	702001	마른자리	12459200	3	으로	0	갈다	702001	이영	30594900	1	을	0
갈다	702001	말굽	12757601	1	을	0	갈다	702001	자리	31627201	1	을	0
갈다	702001	물	14451801	9	을	1	갈다	702001	전구	32874600	3	을	1
갈다	702001	물건	14463101	1	을	0	갈다	702001	전등	32953500	1	을	0
갈다	702001	물고기	14471100	1	을	0	갈다	702001	전지	33205800	1	을	0
갈다	702001	미끼	14699501	1	을	0	갈다	702001	전지약	33206800	1	을	0
갈다	702001	밀창	15022801	1	을	0	갈다	702001	젓니	33858500	2	을	0
갈다	702001	바퀴	15227501	1	을	1	갈다	702001	진자리	36097602	3	을	0
갈다	702001	배넛니	15955100	1	을	0	갈다	702001	잡쌀	36788902	1	을	0
갈다	702001	번	16335101	1	을	1	갈다	702001	찾길	36791402	1	을	0
갈다	702001	병	16804101	2	을	1	갈다	702001	창	36793202	1	을	0
갈다	702001	부속품	17596100	2	을	0	갈다	702001	창유리	36835900	1	을	0
갈다	702001	부품	17751200	1	을	0	갈다	702001	책상보	36927900	1	을	0

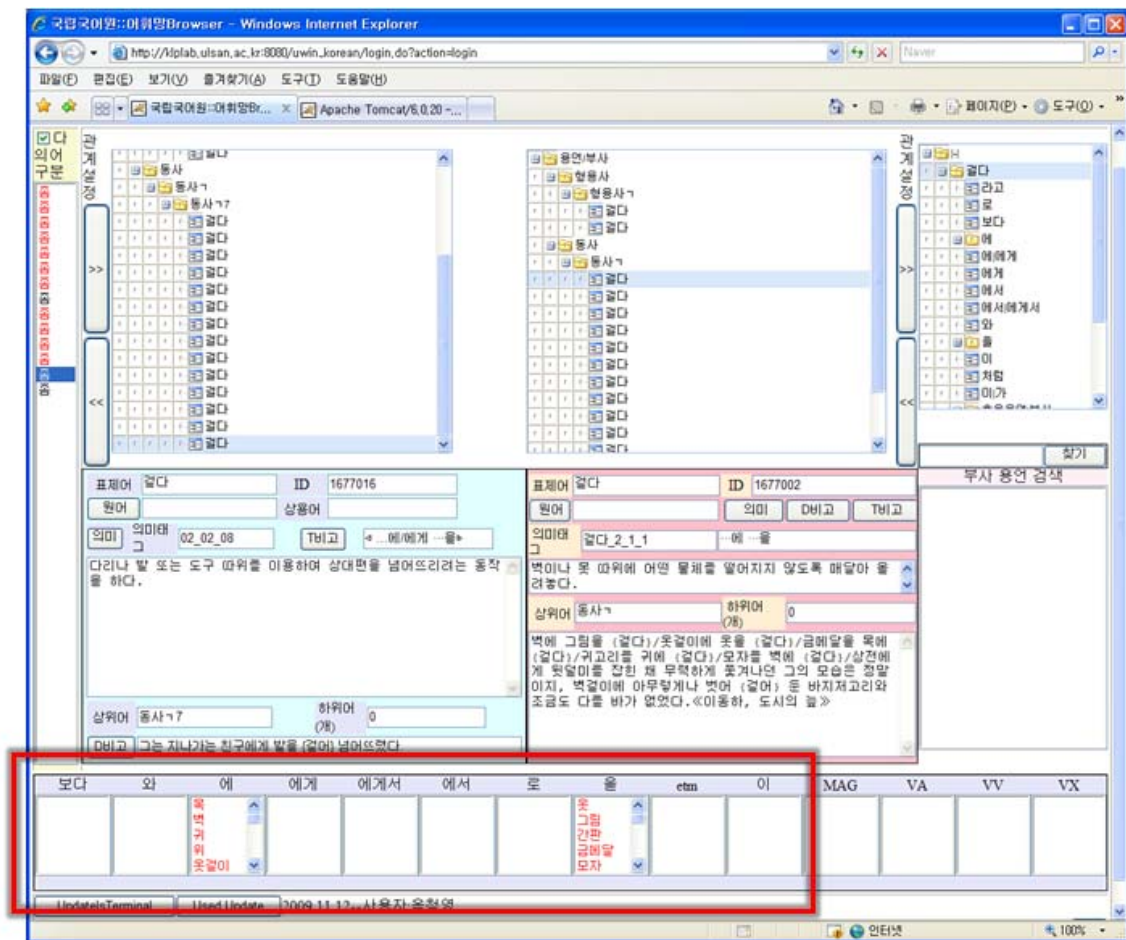
갈다	702001	콘택트렌즈	38749800	1	으로	0	갈다	702102	숫돌	23069900	1	을	0
갈다	702001	타이어	39078401	1	을	1	갈다	702102	안쪽	25130201	1	을	0
갈다	702001	탄	39125902	1	을	0	갈다	702102	암석	25305200	1	을	0
갈다	702001	탄불	39140300	1	을	0	갈다	702102	양면	25823701	1	을	0
갈다	702001	탄창	39171900	1	을	0	갈다	702102	양편	25943000	1	을	0
갈다	702001	탄환	39179202	1	을	1	갈다	702102	연마재	26867700	1	을	0
갈다	702001	털	39393701	3	을	0	갈다	702102	연장	26966801	1	을	0
갈다	702001	패킹	40276801	1	을	0	갈다	702102	옥	27856500	2	을	0
갈다	702001	퓨즈	40985700	1	을	0	갈다	702102	옥돌	27870100	2	을	1
갈다	702001	피	41074800	1	을	0	갈다	702102	옥석	27885901	1	을	0
갈다	702001	활자	43239700	1	을	0	갈다	702102	유리	29535300	1	을	0
갈다	702002	관원	3677100	1	을	0	갈다	702102	인쇄판	30933800	1	을	0
갈다	702002	면도날	13358302	1	을	0	갈다	702102	인조석	30981301	1	을	0
갈다	702002	벼슬	16645201	4	을	0	갈다	702102	철봉	37221501	1	을	0
갈다	702002	벼슬아치	16645800	3	을	0	갈다	702102	칼	38567200	14	을	1
갈다	702002	상대	20166601	1	을	0	갈다	702102	판재	40156801	1	을	0
갈다	702002	인물	30889501	1	으로	1	갈다	702102	평판	40511801	1	을	0
갈다	702002	임금	31277900	1	을	0	갈다	702102	표면	40755601	4	을	0
갈다	702002	임원	31307600	1	을	1	갈다	702102	행실	42038000	1	을	0
갈다	702002	전문가	32998300	1	으로	1	갈다	702103	가루	145200	1	을	0
갈다	702002	직무	35917400	1	을	0	갈다	702103	감자	867202	2	을	0
갈다	702002	책임자	36932500	2	을	1	갈다	702103	개암	1237901	1	을	0
갈다	702002	처	36942100	1	을	1	갈다	702103	겉면	1813400	1	을	0
갈다	702002	친구	38346601	1	을	1	갈다	702103	결석	2006400	1	을	0
갈다	702102	가공품	29200	1	을	0	갈다	702103	고기	2554201	4	을	1
갈다	702102	거울	1495301	1	을	0	갈다	702103	곡식	3006100	5	을	0
갈다	702102	겉면	1813400	2	을	0	갈다	702103	파일	3543300	2	을	0
갈다	702102	고체	2921300	1	을	0	갈다	702103	국물	4425201	1	을	0
갈다	702102	공예품	3365100	1	을	0	갈다	702103	귀비탕	4896600	1	을	0
갈다	702102	공작물	3386901	2	을	0	갈다	702103	깨	6172100	1	을	0
갈다	702102	굽	4705803	1	을	0	갈다	702103	껍데기	6247101	2	을	0
갈다	702102	기억자낫	5810000	1	을	0	갈다	702103	껍질	6250001	1	을	0
갈다	702102	날	6869300	5	을	1	갈다	702103	나무	6621301	3	을	0
갈다	702102	날붙이	6898800	1	을	0	갈다	702103	나무	6621302	1	을	0
갈다	702102	낫	7055400	4	을	1	갈다	702103	난알	6867101	4	을	0
갈다	702102	대검	9018400	1	을	0	갈다	702103	난알	6867102	1	을	0
갈다	702102	대팻날	9374900	1	을	0	갈다	702103	녹두	7717900	8	을	1
갈다	702102	덕	9487801	1	을	0	갈다	702103	논밭	7776500	2	을	0
갈다	702102	도	9647002	1	을	0	갈다	702103	닭고기	8805100	1	을	0
갈다	702102	도끼날	9681001	1	을	0	갈다	702103	돌	10059301	2	을	0
갈다	702102	돌	10059301	6	을	0	갈다	702103	동부	10279801	1	을	0
갈다	702102	돌레	10645401	1	을	0	갈다	702103	두부콩	10580200	3	을	0
갈다	702102	메스	13279301	1	을	0	갈다	702103	땅바닥	11167401	1	을	0
갈다	702102	면	13350602	3	을	0	갈다	702103	땅콩	11176101	1	을	0
갈다	702102	면도칼	13358600	2	을	0	갈다	702103	매밀	13270602	3	을	0
갈다	702102	무예	14186200	1	을	0	갈다	702103	무	13960101	1	을	0
갈다	702102	반대쪽	15357700	1	을	0	갈다	702103	무	13960102	2	을	1
갈다	702102	배달말	15960100	1	을	0	갈다	702103	무리풀	14075600	1	을	0
갈다	702102	부엌칼	17633300	1	을	1	갈다	702103	물건	14463101	2	을	0
갈다	702102	북부	17800400	1	을	0	갈다	702103	물고기	14471100	2	을	0
갈다	702102	손가락	22144900	1	을	1	갈다	702103	물질	14592703	1	을	0
갈다	702102	쇠	22331301	1	을	0	갈다	702103	밀	14941100	1	을	0
갈다	702102	쇠붙이	22359401	2	을	0	갈다	702103	밤	15672900	2	을	0
갈다	702102	쇠붙이	22359402	4	을	0	갈다	702103	밭	15914501	1	을	0

갈다	702103	백반	16138400	1	을	0	갈다	702103	약물	25706900	1	을	0
갈다	702103	벼	16627601	5	을	0	갈다	702103	약재	25737600	4	을	0
갈다	702103	보리쌀	16952400	2	을	0	갈다	702103	얼음	26299101	2	을	0
갈다	702103	빼	18851001	1	을	0	갈다	702103	연근	26827900	1	을	0
갈다	702103	뿌리	18899801	2	을	0	갈다	702103	연자매	26963100	1	을	0
갈다	702103	사과	19036800	1	을	1	갈다	702103	열매	27073200	3	을	0
갈다	702103	사항	19523300	1	을	0	갈다	702103	엿기름	27242200	2	을	0
갈다	702103	살	19821705	3	을	0	갈다	702103	오렌지	27619400	1	을	0
갈다	702103	생강	20560802	1	을	0	갈다	702103	옥수수	27890802	3	을	0
갈다	702103	섬유질	21293401	1	을	0	갈다	702103	은	29985600	1	을	0
갈다	702103	수수쌀	22638300	1	을	0	갈다	702103	잡곡	32108700	1	을	0
갈다	702103	쌀	24279101	2	을	0	갈다	702103	장목수수	32252400	1	을	0
갈다	702103	쌀	24279102	4	을	0	갈다	702103	조각	34116001	1	을	0
갈다	702103	씨	24454001	1	을	0	갈다	702103	주사	34874500	1	을	0
갈다	702103	알약	25237301	1	을	0	갈다	702103	차	36541902	1	을	0
갈다	702103	야채	25678302	2	을	0	갈다	702103	참쌀	36788901	1	을	0
갈다	702103	약	25690601	3	을	0	갈다	702103	채소	36893400	1	을	0

<표 IV-4>에서 용언keyID 및 명사keyID는 U-WIN 데이터베이스 관리번호이며, 각각 구축 도구에서 해당 어휘로 직접 매핑한다. 빈도는 용언의 논항으로 나타난 명사의 빈도수이며, 격조사는 하위범주의 격이다. 비고는 용언의 용례에서 명사의 논항이 출현한 경우는 '1'로, 뜻풀이나 용언이 아닌 다른 품사의 용례에서 출현한 경우는 '0'이다. 이는 이번 하위범주 정보 구축 작업을 용례에서 나타난 논항만을 우선 처리하기 위해 구분하였다.

## 2.2. 하위범주 정보 구축 도구

뜻풀이 및 용례에서 추출한 각 용언 문형의 하위범주 정보를 명사 어휘망의 최소상계노드와 연결하기 위한 구축 도구는 다음 <그림 IV-1>과 같다.

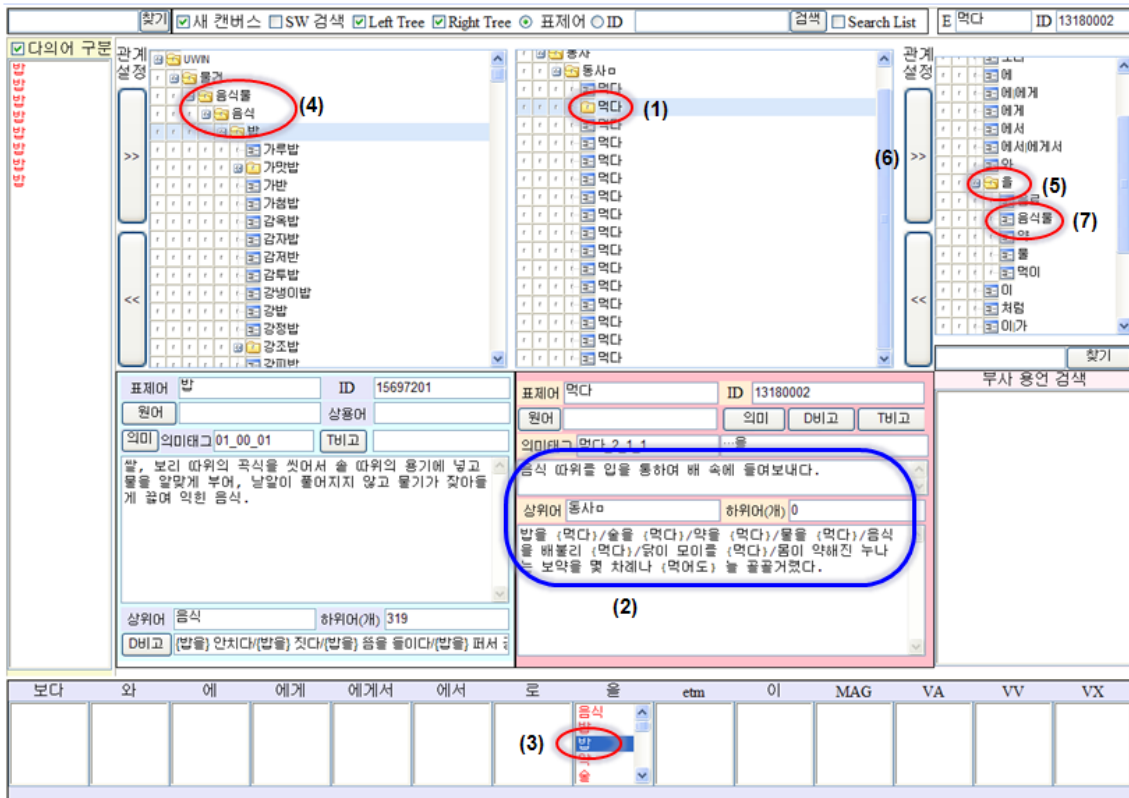


<그림 IV-1> 하위범주 구축 도구

<그림 IV-1>의 아래 쪽(붉은색 박스)에 추출된 하위범주 정보가 나타나며, 개별 하위범주는 직접 명사 어휘망과 연결되어 있다.

## 2.3. 하위범주 정보 연결 과정

용언의 하위범주 정보 구축은 어휘망 편집 브라우저를 통해 이루어진다. 이 어휘망 편집 브라우저에서 용언의 하위범주 정보 연결 과정은, 용언의 용례와 의미에서 논항 후보를 추출하고 명사 어휘망에서 구축 지침에 따라 최소상계노드의 논항을 선택하여 관계 트리에다 넣는 것이다. 논항 후보의 다의어 수가 많을 경우에는 어휘망 편집 브라우저에서 다의어 구분을 사용한다.



<그림 IV-2> 하위범주 정보 연결 과정

- ① 검색 창에서 '먹다'를 입력하면 선정된 다의어 '먹다'가 출력되고, 이 다의어 중에서 특정 용언(예, '먹다\_02\_01\_01')을 선택한다.
- ② 선택된 용언의 뜻풀이와 용례를 확인한다.
- ③ 용언의 하위범주 정보(예, '밥')를 더블클릭하면, 명사 어휘망에서 {밥}의 계층구조 및 상위어/하위어를 확인한다. 여기서 뜻풀이와 용례에서 추출한 논항정보 중에 용언의 용례에서 추출한 논항정보는 붉은색으로 표기되며, 뜻풀이나 다른 품사의 용례에서 추출한 논항정보는 검은색으로 표기된다. 이번 연구과제에서는 붉은색으로 표기된 논항정보를 우선 처리하였다
- ④ '먹다\_02\_01\_01'과 의미적으로 목적격("을")으로 연결될 수 있는 최소상계 노드를 결정한다.
- ⑤ 구축하고자 하는 하위범주(결합관계, 예, "을")를 선택한다.
- ⑥ 관계설정 버튼을 누르면 ⑦과 같이 선택된 명사의 최소상계노드가 용언의 하위범주 관계로 설정된다.

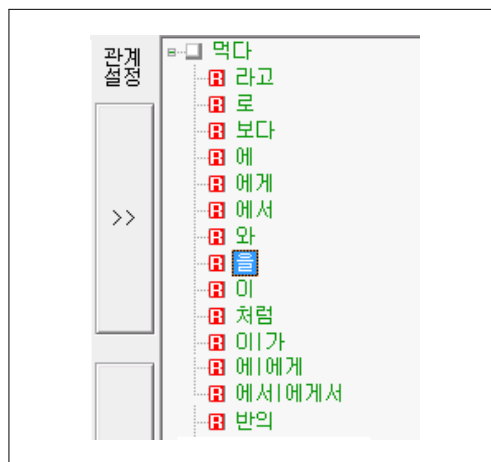
위 입체적 국어사전 브라우저에서 동사 ‘먹다\_02\_01\_01’을 중심으로 ‘먹다\_02\_01\_01’의 목적어(“을”)로 올 수 있는 명사들을 명사 어휘망에서의 최소 상계노드인 {먹이, 음식물, 물, 음료, 약}이 연결되어 있으며, ‘먹다\_02\_01\_01’와 결합하는 부사 {너끈히, 너무, 나날이, 남김없이, 냉큼, 냉큼냉큼, 날로} 등이 구축되어 있음을 확인할 수 있다.

일관성 있는 하위범주 정보 구축을 위해서 다음과 같은 지침을 마련하였다.

□□ 용언의 논항정보 구축 지침은 2008년 지침을 기본으로 삼는다.

(1-1) 용언 어휘망의 문형정보는 <표준>의 뜻풀이에 따른다.

2008년의 문형정보가 <표준>의 뜻풀이었으므로, 2009년 문형정보도 역시 그 뜻풀이에 따른다. 다만, 2009년은 편집기에서 문형정보 수를 <그림 IV-4>와 같이 간소화하였다. “-라고, -로, -보다, -에, -에게, -에서, -와, -을, -처럼, -이, -처럼, -이/가, -에/에게, -에서/에게서”로 간소화하였다. 이것은 구축 정보의 명확성을 보다 높이기 위한 것이다.

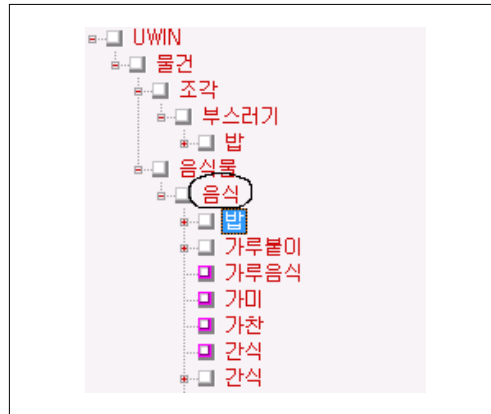


<그림 IV-4> 간소화된 문형정보

문형정보를 간소화하면 복잡 난해한 논항정보 연결 과정을 해결할 수 있다. 예를 들어, 문형정보를 간소화하기 전에는 {-로}가 지금처럼 하나가 아니라 “-로/대로, -로/라고, -로/에, -로/에/에게, -로/에/에게/에게로, -로/에게, -로/에게로, -로/에게로/에, -로/을” 등과 같이 여러 형태로 구성되어 있어서 논항정보 연결의 혼선을 빚을 수 있었다.

(1-2) 문형정보 해당 명사는 명사 어휘망의 최소상계노드와 연결한다.

기본적으로 2009년 용언의 논항정보 구축 작업에서도 최소상계노드를 기준으로 삼는다. 이것은 2008년 작업에서 기준이던 최소상계노드를 따르는 것이다. 이 기준은 국립국어원 2008년 정책연구용역제안서와 최종보고서에 근거한다. “용언 어휘망의 ‘먹다\_\_02\_01\_01’에는 문형정보 {-을}에 해당하는 목적어들을 명사 어휘망 최소상계노드와 연결해야 한다.”



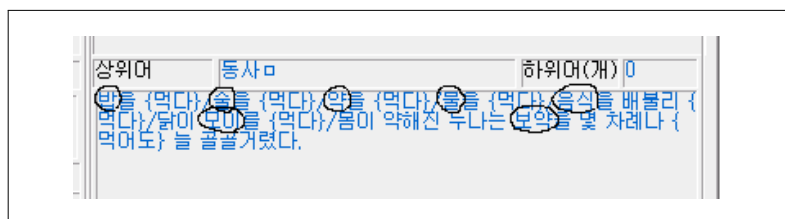
<그림 IV-5> {밥}의 최소상계노드

위의 사례처럼, '먹다\_\_02\_01\_01'은 "{밥}을 먹다"에서 {밥}을 바로 논항정보로 처리하지 않고 {밥}의 최소상계노드 명사인 {음식}을 논항정보로 처리한다. '먹다\_\_02\_01\_01'은 "{밥, 국수, 반찬, 빵}을 먹다"로 쓰인다. 작업에서는 이들 명사를 하나씩 논항으로 연결하기보다 모두를 아우르는 최소 범위의 상위노드 명사로서의 {음식}을 논항정보로 처리한다.

□□ 용언의 논항정보는 각 용언의 용례와 의미를 참조하여 구축한다.

(2-1) 용언의 논항정보는 용례에 있는 논항을 기본정보로 삼는다.

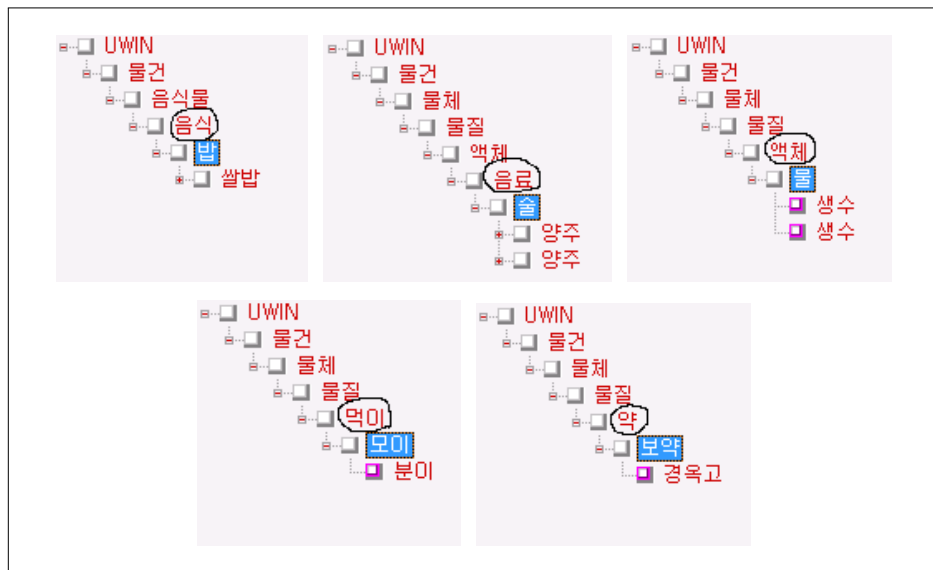
이 지침은 어휘망 편집 브라우저에 용언의 용례 정보가 들어있으므로 이것을 충분히 활용하고자 하는 입장에 준거한다. 용례에서 논항을 추출하는 것은 논항정보 구축 작업 자체의 경제적 효율성과도 관계가 있다. <그림 IV-6>의 '먹다\_\_02\_01\_01' 용례에서 '먹다'의 문형정보 {~을}과 결합된 논항 명사 {밥, 술, 약, 물, 음식, 모이, 보약}을 추출할 수 있다.



<그림 IV-6> '먹다\_\_02\_01\_01'의 용례

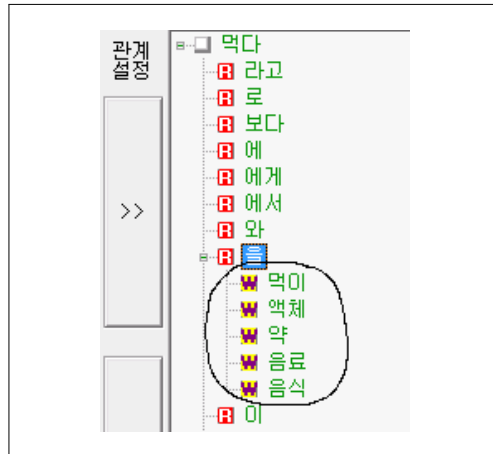
(2-2) 각 논항의 정보는 명사 어휘망에서 상위노드 명사로 연결한다.

용례에서 추출한 명사들은 어휘망에서 바로 가져와서 용언의 논항정보로 넣는 것이 아니라, 지침 (1-2)의 최소상계노드 연결 기준에 따라 각각의 상위노드 명사를 선택하여 논항정보로 연결한다. 명사 {밥, 술, 약, 물, 음식, 모이, 보약}은 최소상계노드 {음식, 음료, 액체, 먹이, 약}으로 논항 설정되어 동사 ‘먹다\_02\_01\_01’의 논항정보로 처리된다.



<그림 IV-7> 논항 명사들의 최소상계노드 명사

지침 (2-2)까지의 작업 결과, ‘먹다\_02\_01\_01’의 논항정보 연결 상태를 살펴보면 <그림 IV-8>과 같다. 용례 정보가 풍부하더라도, 논항정보가 풍부해지는 것은 아니다. 논항이 지침 (2-2)에 따라 연결될 때는 논항 명사들의 상위노드 명사가 이들을 아우르면 논항정보 개수가 축소된다. 따라서 관계 트리에 보이는 논항정보 양의 다소와는 관련성이 적다.

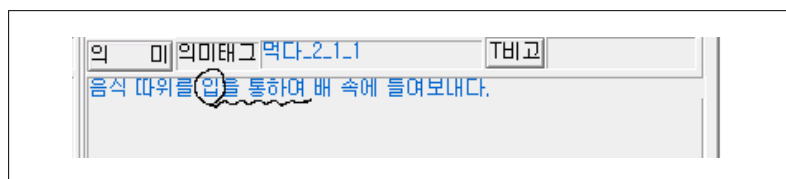


<그림 IV-8> 논항정보 연결 상태

□□ 의미정보상 최소상계노드 연결이 부적합하면 노드를 조정한다.

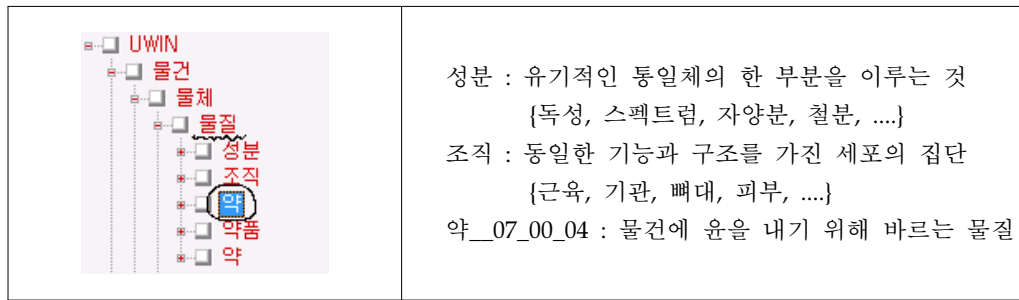
(3-1) 상위노드가 논항의 동위노드를 아우르지 못하면 자기노드로 연결한다.

지침 (2-1)의 <그림 IV-6>에 나오는 논항 명사들 중에서 “{약}을 먹다”의 ‘약\_07\_00\_01’은 최소상계노드로 바로 연결할 수 없다. 만일 그대로 연결하면 ‘먹다\_02\_01\_01’의 의미에 위배되는 논항정보가 된다. ‘약\_07\_00\_01’의 상위노드 명사는 “입을 통하여” 먹을 수 없는 약까지 포함하기 때문이다. 그래서 이 지침에 따라 ‘약\_07\_00\_01’은 자기노드로 연결한다.



<그림 IV-9> ‘먹다\_02\_01\_01’의 의미

‘약\_07\_00\_01’의 동위노드 명사들은 의미연관성 면에서 부적합하다. 그것의 동위노드에는 ‘먹다\_02\_01\_01’이 “입을 통하여” 들여보낼 수 없는 것을 포함하고 있는 {성분, 조직}뿐만 아니라 전혀 먹을 수 없는 물질인 ‘약\_07\_00\_04’도 있다. 상위노드로 연결하면 맞지 않다. 이런 경우가 더러 있으니, 논항정보를 연결할 때 지침 (3-1)로도 동시에 살펴야 한다.



<그림 IV-10> '약\_\_07\_00\_01'의 동위노드 명사들

(3-2) 한 용언의 두 논항이 상하 관계에 있으면 상계 상위노드로 연결한다.

한 용언의 두 논항이 지침 (2-2)에 따라 최소상계노드로 처리되어도, 그 결과가 상하 관계에 놓여있으면 지침 (3-2)로 조정해야 한다. 앞서 '먹다\_\_02\_01\_01'의 논항 {음식, 밥}은 상위노드 {음식물}로 처리될 수 있겠으나, 이것은 일차적인 것이다. <그림 IV-11>처럼 {음식, 밥}은 상하 관계에 있어서 {밥}은 {음식}보다 더 상위 단계의 {음식물}로 최종 연결한다.



<그림 IV-11> 상하 관계의 {음식, 밥}

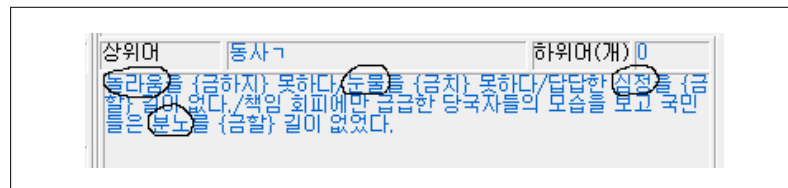
그러면 앞선 <그림 IV-8>의 논항정보 연결 상태는 지침 (3-2)의 적용으로 <그림 IV-12>와 같은 최종 연결 상태로 수정된다. 용언의 논항정보를 연결할 때는 하나의 지침만으로 결정되는 것이 아니다. 한 용언의 논항들 내에서도 다양한 형태의 상하 관계가 나타나기도 하므로, 다수의 지침으로 동시에 살펴봐야 한다. 그때 이런 수정 과정이 발생된다.



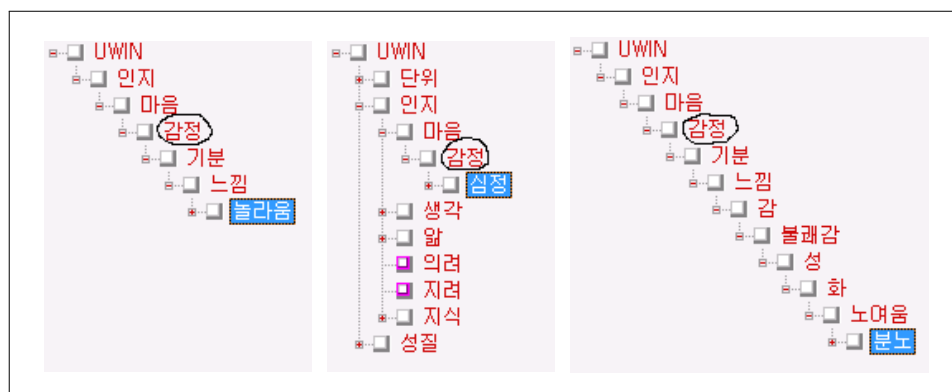
<그림 IV-12> 최종 연결 상태

(3-3) 논항들이 유사 의미장에 같은 줄기이면 공통 상위노드로 연결한다.

한 용언의 논항들이 유사 의미장에 속하고 최소상계노드를 설정할 때도 결국 같은 줄기의 하위어이면 공통 상위노드로 연결한다. ‘금하다\_\_00\_00\_02’에서 {~을}과 결합하는 논항 {놀라움, 심정, 분노}는 최소상계에 있는 노드가 각각 <그림 IV-14>와 같이 {느낌, 감정, 노여움}이지만 상위노드 {감정}의 한 줄기에 있으므로 {감정}을 공통 상위어로 처리한다.



<그림 IV-13> ‘금하다\_\_00\_00\_02’의 용례



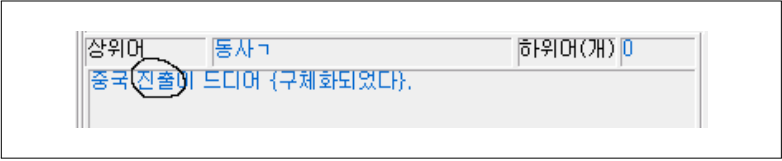
<그림 IV-14> 같은 줄기의 상위노드

지침 (3-3)은 (3-2)에서 확대되는 논항들을 취급한다. 대개는 지침 (3-2)로

처리되는 편이지만, 유사 의미장의 논항들 수가 둘 이상으로 늘어날 때에 처리하기 유용한 지침이다. 만일 추출했을 때는 유사 의미장의 논항들이지만 같은 줄기에 속해 있지 않으면, 각각의 상위노드를 선택해야 하고 제일 높은 공통 상위노드의 명사를 선택할 수 없다.

(3-4) 두 논항 형태가 같고 상하 관계이면 의미가 넓은 상위노드로 연결한다.

‘구체화되다\_\_00\_01\_02’의 용례에서 문형정보 {-이/가}와 결합하는 명사 {진출}을 추출할 수 있다. 이것을 명사 어휘망에서 살펴보면 상하 관계의 두 논항 형태로 나타난다. 이들이 의미상으로 확연히 구분되는 것이면 논항정보를 연결하기가 쉽지만, 그렇지 않을 경우 의미가 넓은 상위노드 쪽을 선택해야 한다. 전체 용언에서 이 같은 경우가 많다.

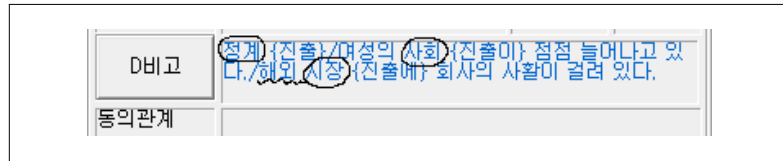


<그림 IV-15> ‘구체화되다\_\_00\_01\_02’의 의미

	<p>진행 : 앞으로 향하여 나아감.  진출__02_00_01 : 어떤 방향으로 활동 범위를 넓혀 나아감.  전진 : 앞으로 나아감. ≒ 진행  진출__02_00_02 : 앞으로 나아감.</p>
--	--

<그림 IV-16> 상하 관계에 있는 두 {진출}

두 논항 중에 상위의 것을 선택하면 의미가 넓은 상위모드이다. <그림 IV-16>과 같이 {진출\_\_02\_00\_02}보다 상위의 {진출\_\_02\_00\_01}을 선택하여 그것의 최소상계노드를 논항정보로 연결한다. <그림 IV-17>처럼 {진출\_\_02\_00\_01}의 의미정보를 보면 적합한 정도를 알 수 있다. “활동 범위를 넓혀 나가는” 진출의 의미가 ‘구체화되다\_\_00\_01\_02’에 더 적합해 보인다.

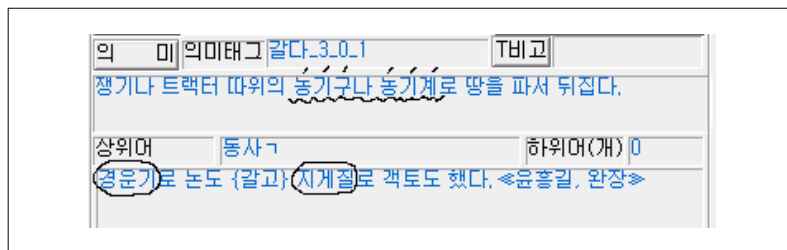


<그림 IV-17> '진출\_02\_00\_01'의 의미

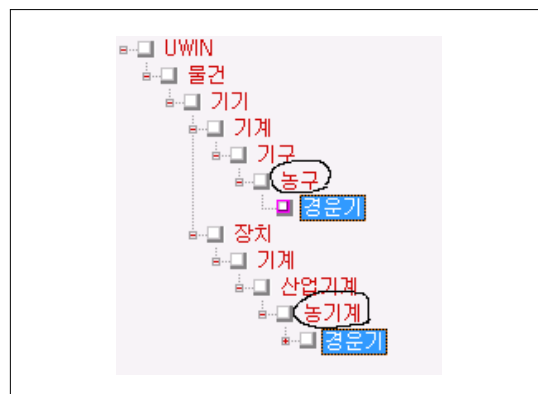
□□ 논항이 여러 노드에 걸쳐 나타나면 의미정보로 노드를 설정한다.

(4-1) 논항의 두 동형이 의미정보에 포함되면 복수 상위노드로 설정한다.

'갈다\_03\_00\_01'의 용례에서 {~로}와 결합된 {경운기}가 나오는데, 명사 어휘망에서 {경운기}를 찾으려면 두 개의 {경운기}가 나타난다. 지침 (3-3)이나 (3-4)에 예처럼 한 줄기에 소속된 것이 아니므로 <그림 IV-18>처럼 의미정보를 살펴볼 필요가 있다. '갈다\_03\_00\_01'의 의미정보에는 "농기구나 농기계"로 나오므로 각각의 상위노드를 결정할 수 있다.



<그림 IV-18> '갈다\_03\_00\_01'의 의미

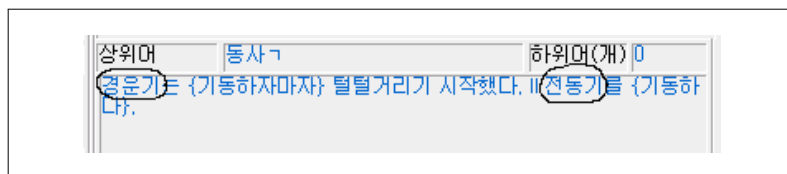


<그림 IV-19> 다른 줄기의 {경운기}

지침 (4-1)은 ‘갈다\_03\_00\_01’의 경우처럼 두 {경운기} 중에 어느 한쪽을 논항정보로 골라 관례적으로 처리하는 문제를 해결한다. ‘갈다\_03\_00\_01’의 의미정보는 “손으로 가는 {농구}로서의 {경운기}”와 “동력을 사용하는 {농기계}로서의 {경운기}”는 각각 포함하므로, 다른 줄기인 것을 감안하여 복수 상위노드를 설정하는 것이다. 의미적 요소가 중요하다.

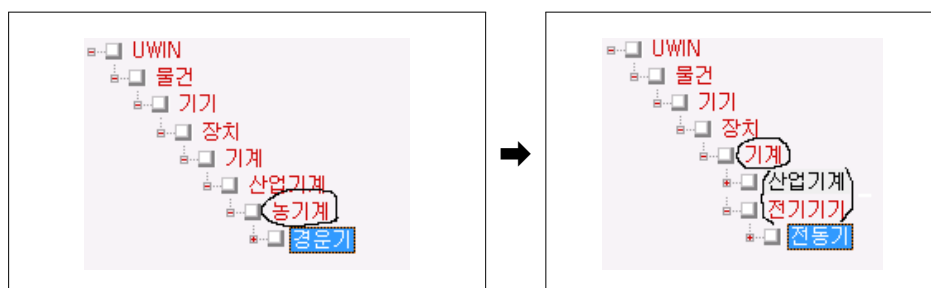
(4-2) 논항이 동형이라도 의미정보가 다르면 단수 상위노드로 설정한다.

‘기동하다’는 논항에 {경운기}가 들어있지만, 지침 (4-1)이 적용되는 ‘갈다\_03\_00\_01’과 {경운기}와 달리 의미정보의 제약을 받는다. 이때의 경운기는 <그림 IV-20>와 같이 “{농구}로서의 {경운기}”가 아니라 “동력을 사용하는 {경운기}”를 의미하기 때문이다. 그러면 복수 상위노드의 것을 설정할 수 없고 의미에 맞는 단수 상위노드의 것을 선택해야 한다.



<그림 IV-20> ‘기동하다’의 의미

여기에서 단수 상위노드로 처리된 ‘기동하다’의 논항 {경운기}는 한 용언 내의 다른 논항 {전동기}와의 관계에 의해, 지침 (4-2)에 이어 (3-2)의 적용을 받는다. {경운기}는 {전동기}와 한 줄기에 소속되어 있으므로, {농기계<산업기계}로서의 {경운기}와 {전기기기}로서의 {전동기}는 {기계}를 공통 상위노드로 받고, {기계}가 최종 논항정보이다.

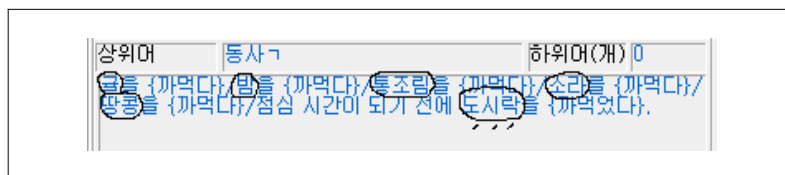


<그림 IV-21> {경운기, 전동기}의 상위노드

□□ 논향이 상위노드를 설정하지 못하는 상태이면 제자리를 유지한다.

(5-1) 논향의 상위노드를 설정하면 그 의미가 부적합할 때 자기노드에 둔다.

최소상계노드를 준수하는 지침 (2-2)에 의해 대부분의 논향은 상위노드를 설정하지만, 그렇게 하면 용언 본래의 의미에 부적합한 경우도 있다. ‘까먹다\_\_00\_00\_01’에서 {도시락}은 상위노드의 {그릇}을 설정하면 의미상 {까먹다}에 적합하지 않다. “{도시락}을 까먹다”는 통용되어도 “{그릇}을 까먹다”는 어색하다. 그래서 {도시락}을 그대로 연결한다.



<그림 IV-22> ‘까먹다\_\_00\_00\_01’의 용례



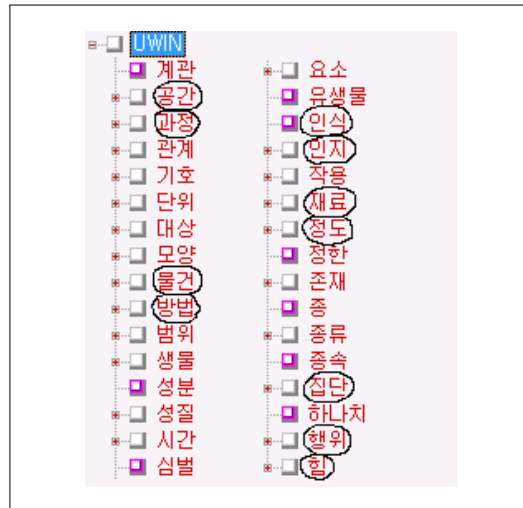
<그림 IV-23> {도시락}의 위치

(5-2) 논향이 최상위노드일 때 그 의미가 포괄적이더라도 자기노드에 둔다.

상위노드를 설정하다 보면 최상위노드까지 올라가는 경우가 있다. 최소상계노드로 결정하여 최상위노드 논향이 되거나, 최상위노드 논향 자체가 자기노드여서 더 이상 올리지 못하는 경우도 있다. 논향이 최상위노드이면 최소상계노드의 역할 면에서 적절하지는 않다. 최상위노드의 논향 자체가 너무 큰 덩어리여서 의미상 포괄적인 면도 크다.

하지만 최상위노드 논향들에 어떤 제약을 가하는 것은 일단 주의할 필요가 있다. 앞선 지침들로 논향정보를 구축하는 작업의 일관성과 관계있기 때문이다. 지침을 따로 부가해도 다른 지침의 적용 형평성에 영향을 끼칠 수 있다.

의미론적 논의가 필요하다. 그래서 최상위노드 논항은 의미가 포괄적이더라도 자기노드에 두는 지침을 적용한다.



<그림 IV-24> 최상위노드 목록

<그림 IV-24>는 명사 어휘망의 최상위노드에 있는 명사 리스트이다. 동그라미로 표시한 것은 논항정보를 연결할 때 최상위노드 중에서도 출현 빈도가 높은 항목들이다. 위의 {공간, 과정, 물건, 방법, 인식, 인지, 재료, 정도, 집단, 행위, 힘} 중에서도 특히 {물건, 방법, 행위, 힘} 등은 가장 빈번하게 상위노드의 논항정보로 연결되고 있는 것들이다.

## 4. 구축 작업 중에 부딪치는 난점들

### 4.1. 전문어 부류의 처리 문제

최소상계노드 기준은 용언의 하위범주 정보 구축 작업에서 가장 기본적인 기준으로 작용한다. 2009년 구축 작업에서의 최소상계노드 기준도 2008년 구축 작업에 사용된 이 기준의 맥락을 잇는 것이다. 앞선 구축 지침의 설명에서 지침 (1-2)로도 언급했다. 최소상계노드는 작업상 유용하기도 하고 판단을 모호하게 만드는 경우도 종종 있었다.

구축 작업은 제한된 기준 하에서 언어정보를 처리하는 것이므로, 구축하는 모든 자료에서 명확한 처리 결과를 요구하기는 힘들다. 이것은 언어정보 특유의 다양성이나 복잡성 때문이다. 어떤 일정한 기준에 의해 언어정보를 처리해도, 좀더 다른 면으로 접근하면 처리 결과가 정확하지 않을 수 있기 때문이다. 일종의 언어처리 판단 문제이다.

‘좌정하다\_00\_00\_01’의 경우 용례에 “영감이 아랫목에 좌정하다. 용상에 국왕이 좌정하다”가 나온다. 이때 문형정보 {~에}와 결합된 논항 명사 {아랫목, 용상}은 최소상계노드 기준을 적용하면 상위노드의 {방바닥, 평상}이 {좌정하다}의 논항정보로 연결된다. {좌정하다}의 의미는 “자리 잡아 앉음”이므로, 이 논항정보의 적절성에는 그다지 문제가 없다.

그런데 이때의 {좌정하다}는 일반어의 용례와 전문어의 용례를 포함하고 있는 점에서 기준 적용이나 논항정보 결정을 재고하게 만든다. “{용상}에 {국왕}이 좌정하다”는 별도 논항정보로 설정되어 있지 않으면, {좌정하다}를 입체국어사전 웹에서 검색할 때 발견하기 어려워진다. 일반어 논항 문치 안에 섞여 일부러 찾아내기 전엔 알 수가 없다.

그래서 이런 경우에 일반어와 전문어의 두 양상을 보여주기 위해 내어놓는 방식으로 논항정보를 연결하였다. 따라서 일반적인 {방바닥, 평상}과 전문적인 {용상}을 함께 보이도록 논항정보로 설정하였다. 그 결과 관계 리스트의 {~에}에 {방바닥, 평상, 용상}이 같이 들어있다. 이런 궁중 관련 어휘들이 제법 있기 때문에 작업에서 유의하는 부분이다.

## 4.2. 용례 가능성의 확대 문제

대부분의 경우 최소상계노드 기준에 의해 어떤 용언의 논항 명사의 상위노드를 결정하면 일반성이 확대되기 마련이다. “{의자}에 앉힌다”는 것은 상위노드의 “{결상}에 앉힌다”는 것으로 논항정보 처리해도 무리는 없다. 경우에 따라서 상위노드로 가는 자체가 입체국어사전 웹상으로 검색할 때는 더 풍부한 논항정보를 접할 수 있도록 만들기도 한다.

예를 들어, ‘죽여주다\_00\_00\_01’의 경우 용례에 “충치가 죽여주게 아프다”가 나온다. ‘죽여주다’의 의미에서처럼 “몹시 고통을 당하여 못 견디게 하다”

의 상태는 {충치}에 국한되지 않는다. {충치}의 최소상계노드의 상위어인 {증세}로 수많은 통증 대상의 것을 ‘죽여주다’의 논항정보로 볼 수가 있다. 이런 역할에 최소상계노드 기준이 작용하는 것이다.

문제는 “{증세}가 죽여주다” 자체가 용례로 가능하다는 점이다. 이것은 최소상계노드의 허용치를 올리는 언어직관을 작용하게 한다. 직관하는 범위가 넓어지면 {증세}의 상위어인 {상태}를 논항정보로 연결하는 것도 가능해진다. 사실 “{상태}가 죽여주다”도 용례로써 충분하다. “{집}을 짓다”의 상위 “{건물}을 짓다”도 ‘상위노드의 {건축물}로 이끈다.

이렇게 직관이 상위로 자꾸 올라가다 보면, 작업상 최소상계노드 기준 자체가 모호해지는 현상이 발생한다. 논항정보 매핑 작업자의 언어직관에 따라 용례 가능성의 범위가 상하 조절될 수 있기 때문이다. 그래서 2009년 구축 작업에서는 어휘망 편집 브라우저에 수록된 용언의 용례에 준거하여 최소상계노드 기준을 살피는 이유가 여기에 있다.

#### 4.3. 의미의 정확한 연결 문제

구체적인 의미를 가진 용언은 의미적 판별이 명확하므로 작업자가 해당 논항을 찾아 바로 연결할 수 있다. 그런데 추상적인 의미를 가진 용언은 용례 정보만으로는 다의어 속에서 직접 관련된 논항을 찾기가 쉽지 않을 때도 허다하다. 이때도 작업자의 언어직관이 많이 관여하게 된다. 다만, 보편적인 언어직관을 따르는 쪽에 강하게 움직인다.

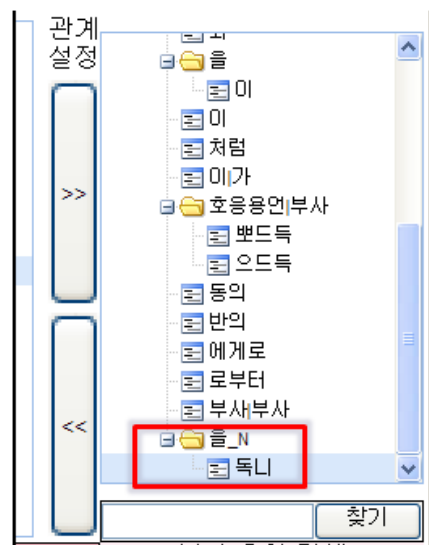
예를 들어, ‘졸이다\_\_00\_00\_02’의 경우 용례에 “마음을 졸이다. 가슴을 졸이다. 애를 졸이다”가 나온다. 문형정보 {~을}과 결합된 명사 {마음, 가슴, 애}는 언뜻 보기엔 바로 논항정보를 연결하게끔 보인다. 하지만 {가슴, 애}는 구체성을 띄므로 금방 연결하기 쉬워도 {마음}은 많은 ‘마음’ 다의어 속에서 구체적인 논항을 찾아내어야 상위어가 결정된다.

이 같은 경우에 가장 극심한 논항은 {일, 힘, 말, 형편} 등과 관련된 것이다. 용례에 {일}이 나오면 다의어 속에서 용언과의 보다 정확한 연결을 필요로 하는 논항정보를 찾는 문제가 불거진다. 작업자의 입장에서 다의어 중에 어느 {일}이 더 정확한 연결일 것인가를 가늠하는 문제이다. 이때는 용언의

의미정보도 도움이 되지 않는 경우가 많다.

그래서 최대한 {일} 자체의 의미정보를 활용하여 용언의 논항정보로 맞추려고 하는데, {일}의 정확성이 떨어지면 논항정보로 결정한 상위노드 논항 자체의 정확성도 떨어지게 된다. 이런 문제의 근본적인 해결을 위해서는 용례로 참고하는 국어사전의 자료가 보완되어야 한다. 현재의 문학작품 자료 위주가 아닌 논리적 자료의 수록도 필요하다.

또한, 용언의 하위범주 정보로 명사 어휘망의 최소상계노드로 설정했을 경우, 최소상계노드의 특정 하위어가 어휘 용법상 해당 용언과 결합하지 못하는 경우도 있다. 이번 연구과제에서는 이러한 경우까지를 모두 고려하여 용언의 하위범주 정보를 구축하지는 못하였으나, 용언의 결합관계에 이러한 예외 경우를 처리할 수 있는 관계("을\_N")를 <그림 IV-25>와 같이 설계하여 두었다. 이러한 예외 경우의 설정은 차후 연구과제로 남겨두었다.



<그림 IV-25> 최소상계노드 하위어 중  
목적격으로 결합되지 못하는 어휘 관계 설정

빈 페이지 임

## V. 부사의 결합관계 정보 구축

### 1. 부사의 결합관계 구축 과정

#### 1.1. 부사의 결합관계 추출

부사는 용언을 비롯한 문장의 여러 성분(체언, 관형사, 부사 등)을 수식하는 성분부사와, 문장 전체를 수식하는 문장 부사로 나뉜다. 또한 부사는 형태 구성의 관점에서도 품사가 (1) 원래 부사인 것, (2) 파생부사, (3) 합성부사, (4) 조사가 결합한 형태, (5) 어미가 결합한 형태 등 다양한 형태가 있다.

- (1) 품사가 부사 : 매우, 좀, 아주, 썩, ...
- (2) 파생부사 : 자연히, 없이, 같이, 많이, 달리, 높이, 빨리, ...
- (3) 합성부사 : 밤낮, 요즈음, 잘못, 좀더, ...
- (4) 조사가 결합한 형태 : 각중에, 순식간에, 정말로, 별도로, 되는데로, 그나마, 그만큼, ...
- (5) 어미가 결합한 형태 : 가볍게, 격렬하게, 되도록, 어쩌면, 끝없이, 각설하고, 못해도, ...

또한, 특정 부사(예, ‘결코’)는 의미적으로 부정형태 혹은 보조용언과도 결합하는 등 다양한 결합 형태를 보인다.

본 과제에서는 2008년에 선정된 2,640개의 부사에 대해, 부사-부사, 부사-용언과 의미적으로 결합하는 성분부사에 한해서 결합관계를 구축하였다.

용언의 하위범주 정보 추출과 마찬가지로 다의어 수준까지 의미 태깅된 뜻풀이 및 용례에서 각 부사의 결합관계를 추출하였다.

다의어 태깅된 뜻풀이/용례의 문장이 “.....(1)..... [부사\_1][부사\_2] .....(2)..... [용언\_1][용언\_2]”와 같을 경우,

- [부사\_1]의 결합관계는 [부사\_2], [용언\_1], [용언\_2]을
- [부사\_2]의 결합관계는 [용언\_1], [용언\_2]을 추출할 수 있다.

예를 들어, 다음 표는 ‘가급적’과 ‘가까스로’에 대해 뜻풀이와 용례에서 추

출된 결합관계의 부서, 용언이다.

<표 V-1> 뜻풀이와 용례에서 추출한 {가급적, 가까스로}와 결합하는 부서 및 용언

부서	부사keyID	호응부사/용언	결합부사/용언 keyID	품사	빈도	비고
가급적	40802	많이	12743400	MAG	1	1
가급적	40802	안	25001300	MAG	1	0
가급적	40802	노력하다	7590002	VV	2	0
가급적	40802	도와주다	9845200	VV	1	1
가급적	40802	되다	10449612	VV	1	0
가급적	40802	먹다	13180005	VV	1	0
가급적	40802	바라다	15123901	VV	1	1
가급적	40802	복원하다	17212404	VV	2	0
가급적	40802	빠르다	18762403	VA	1	1
가급적	40802	상대되다	20166607	VV	1	0
가급적	40802	않다	25168402	VX	1	0
가급적	40802	적다	32732400	VA	2	0
가급적	40802	크다	38859502	VA	1	0
가급적	40802	해치다	41962301	VV	1	0
가급적	40802	획득하다	43477804	VV	1	0
가까스로	42701	꿀꺽꿀꺽	6439802	MAG	1	1
가까스로	42701	가다	68941	VX	1	0
가까스로	42701	가혹하다	473402	VA	1	0
가까스로	42701	건지다	1645403	VV	1	0
가까스로	42701	건다	1665902	VV	1	0
가까스로	42701	걸치다	1702006	VV	1	0
가까스로	42701	계속하다	2437313	VV	1	0
가까스로	42701	구하다	4053503	VV	1	1
가까스로	42701	내다	7130823	VV	1	0
가까스로	42701	내다	7130833	VX	1	0
가까스로	42701	내뿜다	7278800	VV	1	0
가까스로	42701	넉히다	7405601	VV	1	0
가까스로	42701	넘기다	7408003	VV	2	0
가까스로	42701	넘다	7409803	VV	1	0
가까스로	42701	놓다	7903233	VX	1	0
가까스로	42701	누다	8107100	VV	1	0
가까스로	42701	대하다	9000305	VV	1	0
가까스로	42701	더부러지다	9462800	VV	1	0
가까스로	42701	데리다	9598801	VV	1	0
가까스로	42701	들다	10852829	VV	1	1
가까스로	42701	들어서다	10877602	VV	2	0
가까스로	42701	듯하다	10915804	VX	1	0
가까스로	42701	떠받들다	11214502	VV	1	0
가까스로	42701	뿜다	11316904	VV	1	0
가까스로	42701	뜨다	11338701	VV	1	0
가까스로	42701	만나다	12646603	VV	1	0
가까스로	42701	만들다	12663402	VV	1	0
가까스로	42701	말리다	12785701	VV	1	0
가까스로	42701	말하다	12748116	VV	2	0
가까스로	42701	매달리다	12978703	VV	1	0
가까스로	42701	멀다	13218901	VA	1	0
가까스로	42701	면하다	13350406	VV	1	0
가까스로	42701	묻다	14449601	VV	1	0
가까스로	42701	밀치다	14991400	VV	1	0
가까스로	42701	벗어나다	16530402	VV	1	0
가까스로	42701	벗어나다	16530404	VV	1	0
가까스로	42701	빌다	18678803	VV	1	0
가까스로	42701	삼키다	20082703	VV	1	1
가까스로	42701	성공하다	21329204	VV	1	0
가까스로	42701	세우다	21597912	VV	1	0
가까스로	42701	쉬다	23104901	VV	1	0

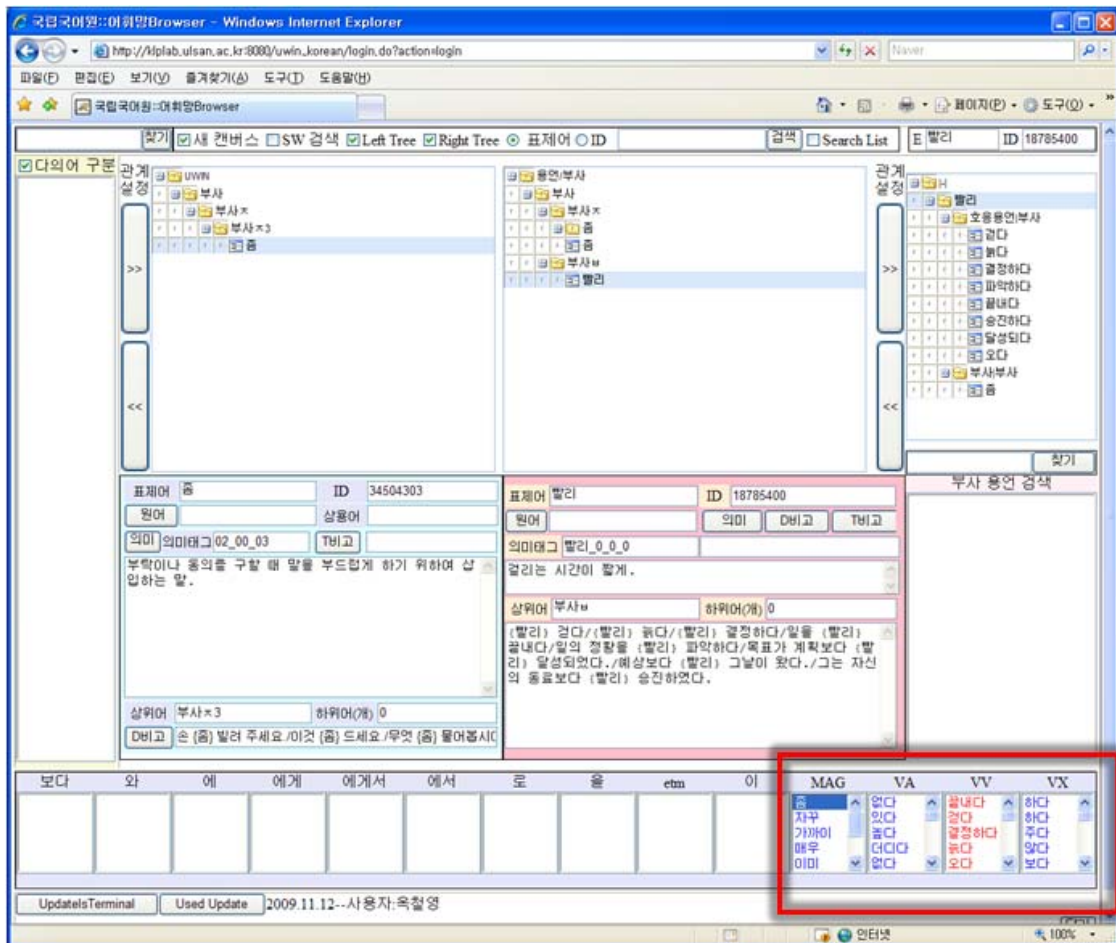
가까스로	42701	안돈하다	25038210	VV	1	0
가까스로	42701	어렵다	26019302	VA	1	0
가까스로	42701	열다	27055309	VV	1	0
가까스로	42701	열리다	27068902	VV	6	0
가까스로	42701	오다	27580113	VV	1	0
가까스로	42701	오다	27580131	VX	1	0
가까스로	42701	이렇다	30411900	VA	2	0
가까스로	42701	이루다	30420403	VV	1	0
가까스로	42701	일부러스럽다	31133800	VA	1	0
가까스로	42701	일어나다	31175712	VV	1	0
가까스로	42701	일어서다	31175902	VV	1	0
가까스로	42701	잇따르다	31422303	VV	2	0
가까스로	42701	있다	31425326	VX	1	0
가까스로	42701	있다	31425327	VX	1	0
가까스로	42701	취어짜다	35398605	VV	1	0
가까스로	42701	짚다	36344301	VV	1	0
가까스로	42701	참다	36731001	VV	6	1
가까스로	42701	참다	36731002	VV	2	0
가까스로	42701	참다	36731003	VV	1	0
가까스로	42701	찾아가다	36857501	VV	1	0
가까스로	42701	피하다	41075407	VV	1	0
가까스로	42701	하다	41273902	VV	2	0
가까스로	42701	하다	41273940	VX	1	0
가까스로	42702	가라앉다	99106	VV	1	0
가까스로	42702	건지다	1645403	VV	1	0
가까스로	42702	구조되다	4317502	VV	1	0
가까스로	42702	나다	6579740	VX	1	0
가까스로	42702	나타나다	6727302	VV	1	0
가까스로	42702	내전하다	7256906	VV	1	0
가까스로	42702	놓다	7903233	VX	1	0
가까스로	42702	돌아오다	10120802	VV	1	0
가까스로	42702	되다	10449620	VV	2	0
가까스로	42702	드나들다	10781104	VV	1	0
가까스로	42702	맞추다	12954412	VV	1	1
가까스로	42702	면궁하다	13355602	VV	1	0
가까스로	42702	면하다	13350405	VV	1	0
가까스로	42702	모면하다	13595204	VV	1	0
가까스로	42702	물리치다	14516901	VV	1	0
가까스로	42702	붙잡다	18245703	VV	1	0
가까스로	42702	비다	18379702	VV	1	0
가까스로	42702	비집다	18574801	VV	1	0

<표 V-1>에서 부사keyID 및 결합부사/용언keyID는 U-WIN 데이터베이스 관리번호이며, 각각 구축 도구에서 해당 어휘로 직접 매핑한다. 품사는 결합되는 어휘의 품사(부사(MAG), 동사(VV), 형용사(VA), 보조용언(VX))이며, 빈도는 부사와 함께 나타난 부사 및 용언의 빈도수이다. 비고에서는 부사의 용례에서 출현한 경우는 '1'로, 뜻풀이나 부사가 아닌 다른 품사의 용례에서 출현한 경우는 '0'이다. 이는 이번 부사 결합관계 정보 구축 작업을 용례에서 나타난 경우를 우선 처리하기 위해 구분하였다.

## 1.2. 부사의 결합관계 구축 도구

뜻풀이 및 용례에서 추출한 각 부사의 결합관계를 구축하기 위한 구축 도

구는 다음 그림과 같다.

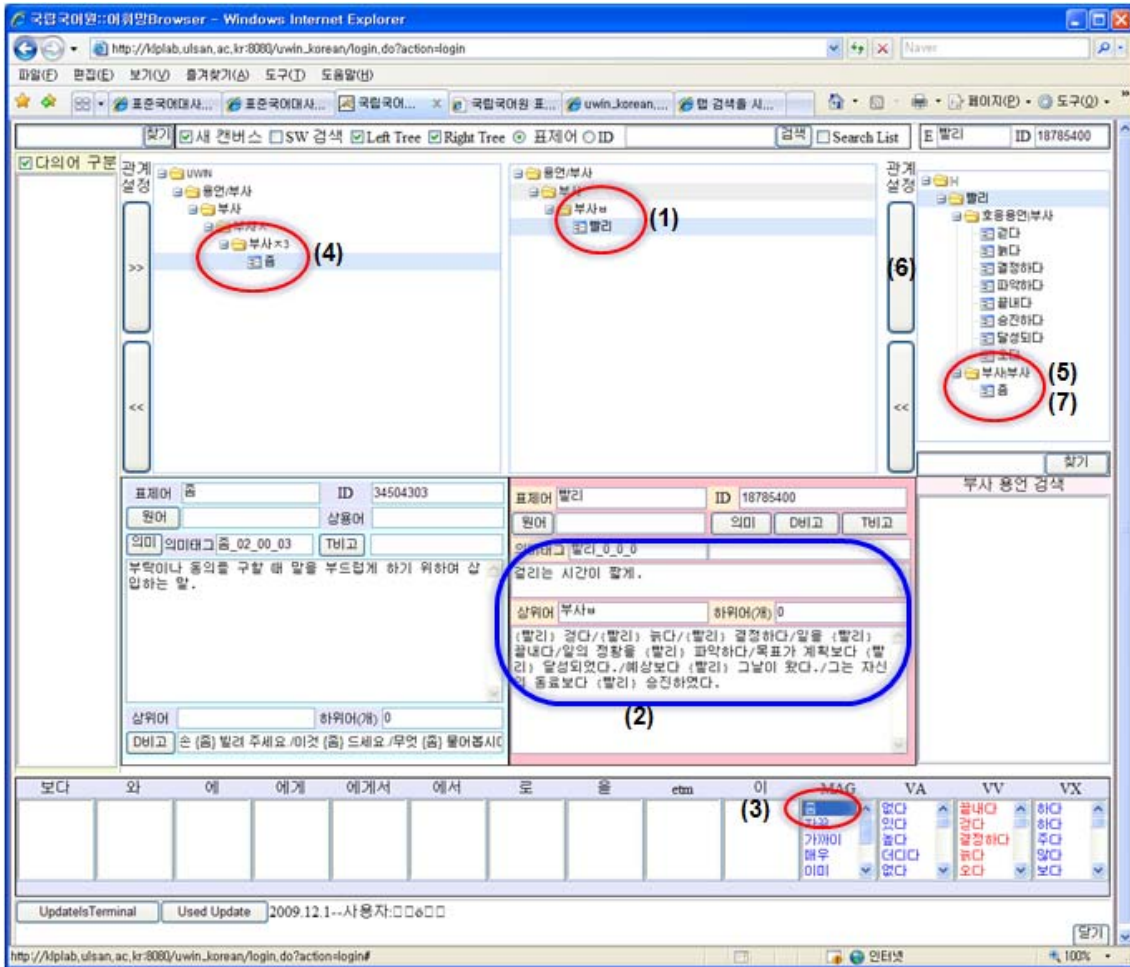


<그림 V-1> 부사의 결합관계 구축 도구

위 그림의 아래 쪽(붉은색 박스)에서 뜻풀이/용례에서 추출된, 특정 부사와 결합할 수 있는 부사, 용언(형용사, 동사, 보조용언)을 나타내며, U-WIN과 연결되어 있다.

### 1.3. 부사의 결합관계 연결 과정

부사의 부사와 용언과의 결합관계 구축은 어휘망 편집 브라우저를 통해 이루어진다. 이 어휘망 편집 브라우저에서 용언과 부사의 용례에서 추출한 결합관계 정보를 부사의 문법적인 수식 지침(V.2 장)에 따라 호응용언|부사 혹은 부사|부사 결합관계 트리에다 넣는 것이다.

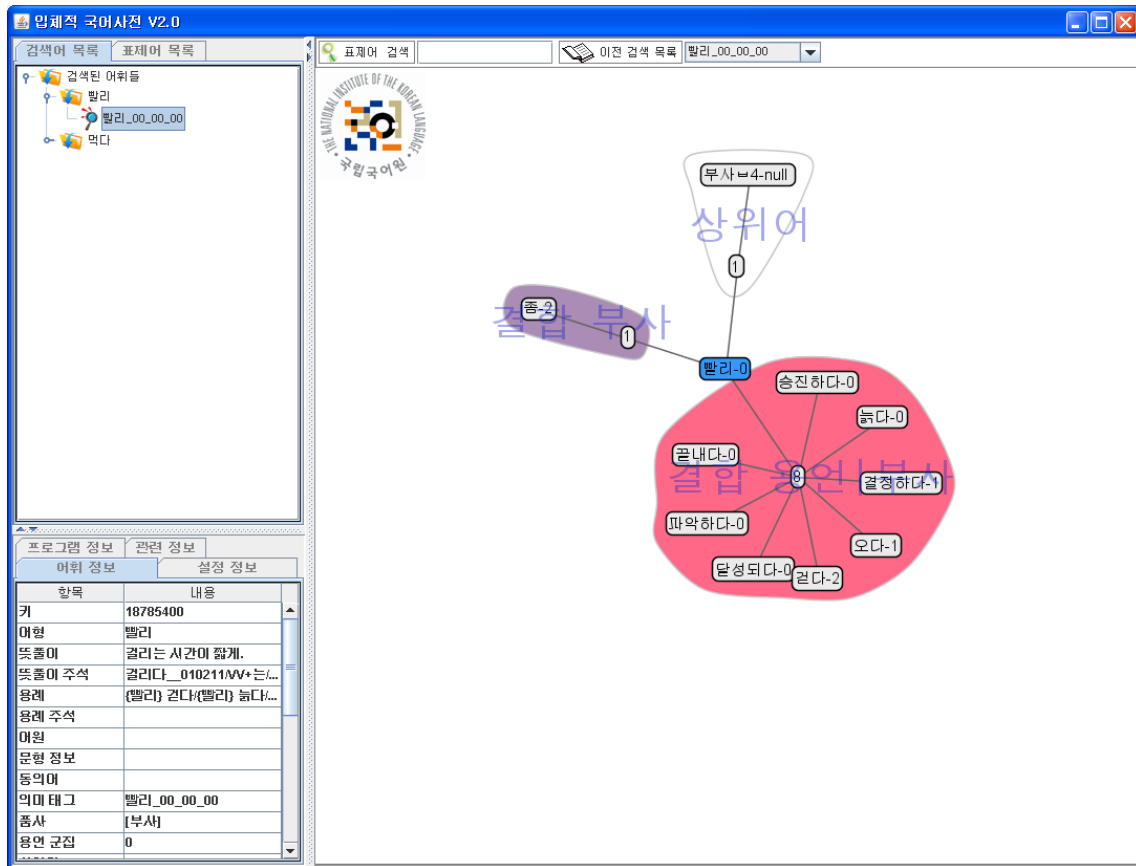


<그림 V-2> 부사의 결합관계 연결 과정

- ① 검색 창에서 선정부사(예, '빨리')를 입력하면 선정된 다의어 '빨리'가 출력되고, 이 다의어 중에서 하나를 선택한다.
- ② 선택된 부사의 뜻풀이와 용례를 통해 여러 다양한 용례를 확인한다.
- ③ 용례에서 추출된 부사\_부사의 결합관계인 '쥬'를 더블클릭한다. 여기서 뜻풀이와 용례에서 추출한 부사 결합관계 정보에서 부사와 용언의 용례에서 추출한 결합관계 정보는 붉은색으로 표기되며, 뜻풀이와 다른 품사의 용례에서 추출한 결합관계 정보는 검은색으로 표기된다. 이번 과제에서는 붉은색으로 표기된 어휘를 우선 처리하였다
- ④ U-WIN의 '용언/부사'에서 '쥬'의 뜻풀이와 용례(D비고)를 확인할 수 있다.
- ⑤ '빨리'와 '쥬'이 의미적으로 부사\_부사 결합관계가 성립되면 부사\_부사 결합관계를 선택한다.

- ⑥ 관계설정 버튼을 누르면 ⑦과 같이 선택된 부서 '쑤'이 '빨리'와 부서\_부사 결합관계로 설정된다.

이렇게 설정된 부서\_부사 혹은 부서\_용언의 결합관계 정보는 다음 입체적 국어사전에서 {결합 부서}, {결합 용언|부사}로 확인할 수 있다.



<그림 V-3> 입체적 국어사전 브라우저에서 부사의 결합관계

위 입체적 국어사전 브라우저에서 부서 '빨리\_00\_00\_00'를 중심으로 부서\_부사 결합관계인 {쑤}과 부서\_용언 결합관계인 {결다, 늡다, 달성되다, 파악하다, 끝내다, 결정하다, 승진하다, 오다} 등이 구축되어 있음을 확인할 수 있다.

## 2. 부사의 결합관계 구축 지침

<표준>에 등재된 용례를 기반으로 다음의 지침에 따라 {부사-용언, 부사-부사} 결합관계 정보를 구축하였다. <표준>에서 용례를 확인할 수 없는 경우나 문장부사 등 결합관계를 밝힐 수 없는 경우들은 처리하지 않도록 한다. 이는 임의로 용례를 만들거나 다른 말뭉치를 통해서 확인을 할 수도 있겠지만, 정확성을 기하기 위함이며 다음의 미처리 기준을 설정하였다.

#### □□ 사전에서 용례를 확인할 수 없는 경우 따로 처리하지 않는다.

- 가득가득\_\_01\_00\_03 : 냄새나 빛 따위가 공간에 매우 널리 퍼져 있는 상태. ≡가득가득히 「3」.
- 가득가득\_\_01\_00\_04 : 감정이나 정서, 생각 따위가 매우 많거나 강한 모양. ≡가득가득히 「4」.
- 살랑살랑\_\_01\_00\_02 : 물이 끓어오르며 이리저리 자꾸 움직이는 모양.
- 하늘하늘\_\_00\_00\_03 : 어디에 매인 데 없이 멋대로 한가하게 놀고 지내는 모양.

#### □□ 문장 부사는 따로 처리하지 않는다.

- 그러나\_\_00\_00\_00 : 앞의 내용과 뒤의 내용이 상반될 때 쓰는 접속 부사.
- ¶ 우리는 열심히 손을 흔들었다. 그러나 선수 중 아무도 돌아보는 사람이 없었다./아내는 조용히 그러나 단호하게 말했다./성기는 잠자코 밥숟가락을 들었다. 그러나 밥은 반도 먹지 않고 상을 물려 버렸다. <<김동리, 역마>>/근심스러운 소식, 듣기 싫은 소식, 그러나 또한 심중팔구는 반드시 나올 소식을 그들은 겁먹은 마음으로 기다리고 있는 것이었다.<<김동인, 운현궁의 봄>>

#### □□ 'N+이다' 서술어는 처리하지 않는다.

- 사사건건\_\_00\_00\_00 : 해당되는 모든 일마다. 또는 매사에. ≡건건사사 [II]·사사건건이.
- ¶ 동생은 사사건건 말썽이다./엄마는 내가 하는 일에 사사건건 간섭

이다./사사건건 걸고넘어지니 나보고 어찌란 말이야?/내버려진 춘복  
이를 거두어 기른 정이 애뜻한 공배로서는, 춘복이한테 하찮은 일도  
대수롭게 넘기지 못하고 사사건건 토를 달고 나서게 되는 것이었다.  
《최명희, 혼불》

- 어느새\_00\_00\_00 : 어느 틈에 벌써.

¶ 입학한 지가 어제 같은데 어느새 졸업이다./코흘리개 어린애가 어  
느새 자라서 청년이 되었구먼./한라산은 비구름에 허리까지 잠뎀 잠  
겨 있는데 감질나게 찢끔거리던 가랑비는 어느새 그쳐 있었다.《현기  
영, 변방에 우짖는 새》

□□ ‘용언+-이’의 형태 중 부사로 처리하기 힘든 경우는 처리하지 않는다.

- 상관없이\_00\_00\_01 : ⇒ 상관없다 「1」.

¶ 그는 몹시 피곤한 것 같았고 옷차림하곤 상관없이 초라하고 헐벗  
은 것처럼 보였다.《박완서, 도시의 흉년》

- 아랑곳없이000 : 어떤 일에 참견을 하거나 관심을 둘 필요가 없이.

¶ 들몰댁이 역전에 이르렀을 때는 서늘한 날씨는 아랑곳없이 전신은  
땀으로 젖어 있었다.《조정래, 태백산맥》/갈증으로 입이 탄 사병들은  
죽고 사는 것은 아랑곳없이 서로 미친 듯 물을 찾았다.《홍성원, 육  
이오》

□□ ‘구 구성, 절’을 수식하는 경우는 처리하지 않는다.

- 설령\_01\_00\_00 : ((뒤에 오는 ‘-다 하더라도’ 따위와 함께 쓰여)) 가정해  
서 말하여. 주로 부정적인 뜻을 가진 문장에 쓴다. ≒설사02(設  
使)·설약(設若)·설혹(設或)·억혹(抑或)·유혹01(猶或).

¶ 저들이 설령 우리를 이곳에서 내보내 준다 해도 아주 좋아주지는  
않을 것이다./그런 일은 없겠지만 설령 이번 일이 안된다고 하더라  
도 너무 실망은 하지 마라./설령 그가 그것을 훔쳤다 할지라도 너는  
그를 믿어 주어야 한다.

□□ 사전에서의 용법을 살핀 후 처리할 수 없는 것은 처리하지 않는다.

- 마치\_\_03\_00\_00 : ((흔히 ‘처럼’, ‘듯’, ‘듯이’ 따위가 붙은 단어나 ‘같다’, ‘양하다’ 따위와 함께 쓰여)) 거의 비슷하게.

¶ 마치 선녀처럼 고운 얼굴/반장은 마치 자기가 담임 선생님인 듯이 아이들에게 이래라저래라 했다./그녀의 목소리는 마치 천상에서 울리는 음악 소리 같다./전방에는 마치 사원 비슷한 높은 누각이 당당하게 막아서 있었다.《홍성원, 육이오》/창선은 마치 사막에 고개만 파 묻은 타조 모양으로 혼자 눈을 감고 아무것도 보지 않았다.《나도향, 뉘우치려 할 때》

- 어찌\_\_00\_00\_04 : ((감탄 표현의 ‘-니지’, ‘-는지’ 따위와 함께 쓰여))동작의 강도나 상태의 정도가 대단함을 나타낸다.

¶ 그 꽃을 보는 순간 어찌 예쁘던지 녀를 잃고 바라보았다.

□□ 용례가 잘못되어 부사로 보기 어려운 것들은 처리하지 않는다.

- 까르륵\_\_00\_00\_02 : 여자나 아이들이 자지러지게 웃는 소리. 또는 그 모양.

¶ 누가 무슨 말을 했는지 옆방에 한 패가 와글와글 떠들더니, 까르륵, 여자도 섞인 웃음소리가 터진다.《최명희, 혼불》

- 으하하\_\_00\_00\_00 : 입을 크게 벌리며 거리낌 없이 크게 웃는 소리. 또는 그 모양.

¶ 으하하! 저 녀석 보게.

빈 페이지 임

## VI. 용언 의미 군집화

### 1. 필요성 및 관련 연구

한국어 어휘의미 정보를 다각적이고 종합적으로 파악하기 위해서는 어휘의 의미 체계(명사 어휘망, 용언의 하위범주 정보, 부사-부사 및 부사-용언의 결합관계)를 구축하여야 한다.

명사는 사전 뜻풀이의 종차와 유개념을 이용하여 상위어, 하위어의 어휘개념 관계를 설정할 수 있으며, 이를 이용하여 명사 어휘망을 구축할 수 있다. 반면, 용언은 용언의 통사적 특성과 문형 논항의 하위범주 정보를 통해 용언의 개별 의미에 대한 의미정보를 구축할 수 있다. 그러나 용언의 경우 명사 어휘망과는 달리, 상위어, 하위어와 같은 계층 구조를 구축하는데 어려움이 있다.

영어 워드넷(WordNet)에서도 형용사는 “직접적인” 반의어와 “간접적인” 반의어 사이의 차이를 반영한 것으로, 반의 혹은 대립의 요소와 “의미적으로 유사한” 것으로 의미 군집을 형성하고 있다. 또한, 동사는 “양식 관계(troponymy)”와 “어휘적 함의 관계(lexical entailment)”를 이용하여 배열되어 있으며, “양식 관계”는 명사의 동의어 집합(synset) 사이에 유지되는 상하위 관계와 유사한 계층구조를 구성하나 네 단계를 넘는 경우가 거의 없다.

반면, 세종전자사전에서는 표제항 어휘 및 구절의 의미 표상과 논항 선택 제약 기술의 정확성 및 일관성을 제고하기 위해 의미 부류 체계를 구축하였다. 최상위 부류로는 <구체물>, <집단>, <장소>, <추상적대상>, <사태> 등 5개의 부류가 설정되었으며 이중 처음 4개는 논항명사의 의미 부류이고, 나머지 <사태> 부류는 술어명사의 의미 부류이다. 용언의 의미 부류 체계는 <사태> 부류를 용언의 의미 표상에 맞게끔 확장 보완하는 방식으로 구축되었다.

본 과제에서는 다분화된 용언의 의미를 용언의 뜻풀이를 이용하여 군집화하고 세종전자사전의 <사태> 부류로의 매핑을 시도하여, 세종전자사전에 등재되지 않은 선정 용언의 의미범주를 결정한다. 또한 이렇게 결정된 의미범주를 품사/동형이의어 동시 태깅 시스템(부록 B)에서 활용한다. 더 나아가 영어 워드넷의 용언과 비교하는 기초 자료를 확보하고자 한다.

## 2. 사전 뜻풀이를 이용한 의미 군집화

### 2.1. 용언의 사전 뜻풀이

국어사전에서 용언의 뜻풀이는 대체로 관련된 유의 관계 혹은 함의 관계로 기술되어 있다. 예를 들어 형용사 ‘맑다’는 <표준>에서 다음과 같이 정의되어 있다.

- ① 잡스럽고 탁한 것이 섞이지 아니하여 환하고 깨끗하다.
- ② 구름이나 안개가 끼지 아니하여 날씨가 깨끗하다.
- ③ 소리 따위가 가볍고 또랑또랑하여 듣기에 상쾌하다.
- ④ 정신이 초롱초롱하고 또렷하다.
- ⑤ 살림이 넉넉하지 못하고 박하다.

또한, 다음과 같은 용언들이 “먹다02\_①(음식 따위를 입을 통하여 배 속에 들여보내다)”의 의미를 가지고 있으며, 용언별로 약간의 의미 차이(양식 관계 혹은 함의 관계)는 있지만 {먹다}로 정의되어 있다.

- 간식하다\_01 : 간식\_02 (끼니와 끼니 사이에 음식을 먹음.)
- 갈겨먹다② : 남의 음식을 빼앗아 먹다.
- 거머먹다 : 이것저것 욕심스럽게 급히 건어 먹다.
- 걸치다\_03 : 음식을 아무렇게나 대충 먹다.
- 걸터먹다 : 이것저것 닥치는 대로 휘몰아 먹다.
- 곹먹다① : 곹절로 먹다.
- 그러먹다 : 한곳에 모아서 먹다.
- 나누다\_03① : 음식 따위를 함께 먹거나 갈라 먹다.

위의 예에서 보듯이 용언들은 사전 뜻풀이에서 적합한 유의어 세트(set)를 추출하여 의미 군집이 가능하다.

### 2.2. 유의어 세트 추출

### 2.2.1. 유의어 세트

이 연구과제에서는 다의어 의미 태깅된 용언의 뜻풀이에서 용언의 유의어 세트는 각 뜻풀이의 마지막 본용언으로 설정하였다. 예를 들어, ‘떡다02’의 다의어 태깅된 뜻풀이는 다음과 같다.

- (1) 음식\_\_000001/NNG 따위\_\_000001/NNB+를/JKO 입\_\_000001/NNG+을/JKO 통하다\_\_000702/VV+아/EC 배\_\_010001/NNG 속\_\_010002/NNG+에/JKB 들여보내다\_\_000101/VV+다/EF+./SF
- (2) 담배\_\_000003/NNG+나/JX 아편\_\_010000/NNG 따위\_\_000001/NNB+를/JKO 피우다\_\_010005/VV+다/EF+./SF
- (3) 연기\_\_090000/NNG+나/JX 가스\_\_000001/NNG 따위\_\_000001/NNB+를/JKO 들이마시다\_\_020000/VV+다/EF+./SF
- (4) 어떤\_\_000003/MM 마음\_\_010002/NNG+이나/JX 감정\_\_060000/NNG+을/JKO 풀다\_\_010200/VV+다/EF+./SF
- (5) 일정하다\_\_000004/VA+ㄴ/ETM 나이\_\_010000/NNG+에/JKB 이르다\_\_010001/VV+거나/EC 나이\_\_010000/NNG+를/JKO 더하다\_\_000201/VV+다/EF+./SF
- (6) 욕\_\_020001/NNG+./SP 편잔\_\_010000/NNG 따위\_\_000001/NNB+를/JKO 들다\_\_010102/VV+거나/EC 당하다\_\_010100/VV+다/EF+./SF
- (7) (/SS+속되다\_\_000001/VA+게/EC+)/SS 뇌물/NNG+을/JKO 받다\_\_010101/VV+아/EC 가지다\_\_000102/VV+다/EF+./SF
- (8) 수익\_\_010002/NNG+이나/JX 이문\_\_060001/NNG+을/JKO 차지하다\_\_010001/VV+아/EC 가지다\_\_000102/VV+다/EF+./SF
- (9) 물\_\_010001/NNG+이나/JX 습기\_\_020000/NNG 따위\_\_000001/NNB+를/JKO 빨아들이다\_\_000001/VV+다/EF+./SF
- (10) 어떤\_\_000003/MM 등급\_\_000101/NNG+을/JKO 차지하다\_\_010001/VV+거나/EC 점수\_\_060001/NNG+를/JKO 따다\_\_010305/VV+다/EF+./SF
- (11) 구기\_\_110000/NNG 경기\_\_110000/NNG+에서/JKB+./SP 점수\_\_060001/NNG+를/JK O 잃다\_\_000111/VV+다/EF+./SF
- (12) (/SS+속되다\_\_000001/VA+게/EC+)/SS 여자\_\_020001/NNG+의/JKG 정조\_\_090002/NNG+를/JKO 유린하다\_\_010101/VV+다/EF+./SF
- (13) 매\_\_010000/NNG 따위\_\_000001/NNB+를/JKO 맞다\_\_030001/VV+다/EF+./SF
- (14) 남\_\_010001/NNG+의/JKG 채물\_\_030001/NNG+을/JKO 다루다\_\_010101/VV+거나/EC C 말다\_\_010102/VV+은/ETM 사람\_\_000001/NNG+이/JKS 그\_\_010102/MM 채물\_\_030001/NNG+을/JKO 부당하다/VA+게/EC 자기\_\_040001/NNG+의/JKG 것\_\_010001/NNB+으로/JKB 만들다\_\_000200/VV+다/EF+./SF

이상과 같이 용언 의미별로 유의어 세트를 정의하고 뜻풀이의 동일 유의

어를 가지는 용언을 추출하면 다음 <표 VI-1>과 같다.

<표 VI-1> {먹다}를 유의어로 가지는 용언

유의어	용언 수	뜻풀이(유의어)
		뜻풀이에 유의어를 가진 용언
먹다__010000	1	귀나 코가 막혀서 제 기능을 하지 못하게 되다. 가는귀먹다
먹다__020101	292	음식 따위를 입을 통하여 배 속에 들여보내다. 가리다__030006 간식하다__010000 갈겨먹다__000002 감식하다__010000 감어하다 강다짐하다__000001 강반하다 개걸하다 개소하다__020002 거머먹다 건담하다 건식하다__010001 건식하다__010002 건식하다__030000 걸치다__000300 걸터먹다 검식하다 ....
먹다__020104	14	어떤 마음이나 감정을 품다. 각자위심하다 견어잡다__000002 결의하다__010000 깃집하다 뜻하다__000101 샘내다 생각하다__000104 식겁하다 앙구다__000500 용의하다__000001 입심하다 작심하다 함독하다__000002 회심하다__010002
먹다__020105	3	일정한 나이에 이르거나 나이를 더하다. 가령하다 증년하다 헛먹다
먹다__020106	6	육, 편잔 따위를 듣거나 당하다. 무사득방하다 연결먹다__000001 할딱거리다__000003 할딱할딱하다__000003 혈떡거리다__000003 혈떡혈떡하다__000003
먹다__020109	2	물이나 습기 따위를 빨아들이다. 서상하다__020000 음서하다__010000
먹다__020201	1	날이 있는 도구가 소재를 깎거나 자르거나 갈거나 하는 작용을 하다. 엇먹다__010002
먹다__020203	1	벌레, 균 따위가 껍 들어가거나 퍼지다. 썩음들다

위 표에서 용언 수는 뜻풀이의 마지막 본용언이 개별 유의어로 정의된 용언의 개수이다.

## 2.2.2. 유의어의 연결 구성

용언의 유의어도 용언으로 정의되기 때문에, 이러한 유의어들의 상위 유의어의 연결을 만들 수 있다. 다음 <표 VI-2>는 유의어 {먹다}의 상위 유의어 연결을 보이고 있다.

<표 VI-2> 유의어 {먹다}의 상위 유의어 연결

유의어	상위 유의어 연결
먹다__010000	되다__010304 → 다하다__000200 → 끝내다__000001 → mk_끝나다__000001
먹다__020101	들여보내다__000101 → 들어가다__010101 → 가다__010101 → 이동하다__000001 → 옮기다__000101 → mk_옮다__000101
먹다__020104	*품다__010200 → 가지다__000200
먹다__020105	더하다__000201 → 많다 → *넘다__010203 → 지나다__000302
먹다__020106	당하다__010100 → 받다__010103 → *입다__010002 → 당하다__010501
먹다__020109	빨아들이다__000001 → 흡수하다__020005 → 거두어들이다__000102 → 모으다__000101 → 합치다 → 이르다__020200 → 말하다__000300 → 보이다__010101 → psv_보다__010101
먹다__020201	하다__010101 → 이루다__010101 → 만들다__000300 → *되다__010202 → 이루어지다__000200
먹다__020203	퍼지다__000201 → 미치다__020200 → 가하다__010102 → 끼치다__020001 → *입다__010002 → 당하다__010501

상위 유의어 추출 과정에서 뜻풀이가 사동형, 피동형, 순환 정의된 경우가 많았으며, 위 표에서 {mk\_끝나다}는 “끝내다(‘끝나다’의 사동사)”, {psv\_보다}는 “보이다(‘보다’의 피동사)“이며, {\*입다\_\_010002}는 {당하다\_\_010501}의 유의어로 순환적으로 정의되어 있음을 뜻한다.

상위 유의어 세트의 연결을 완성하면, “먹다\_\_020104(어떤 마음이나 감정을 품다.)”의 최상단 유의어 {가지다\_\_000200(생각, 태도, 사상 따위를 마음에 품다.))처럼 전체적인 의미를 유지한다. 그렇지만, 초기 유의어와 최상단 유의어의 의미가 상당히 다른 연결이 만들어질 수 있다. 이는 다의어 태깅 오류로 인한 경우거나, “먹다\_\_010000”의 뜻풀이 “귀나 코가 막혀서 제 기능을 하지 못하게 되다.”와 같이 부정형을 고려하지 않아 발생될 수 있다.

이렇게 연결한 최상위 유의어 세트는

- 일반용언 133개
- 사동사(mk\_A) : 202개
- 피동사(psv\_A) : 228개로 구성되었다.

본 과제에서는 최상위의 유의어 세트를 용언의 1차 의미군집으로 보았으며, 각 의미군집에 속하는 용언의 수는 다음 <표 VI-3>과 같다.

<표 VI-3> 최상위 유의어 군집 예

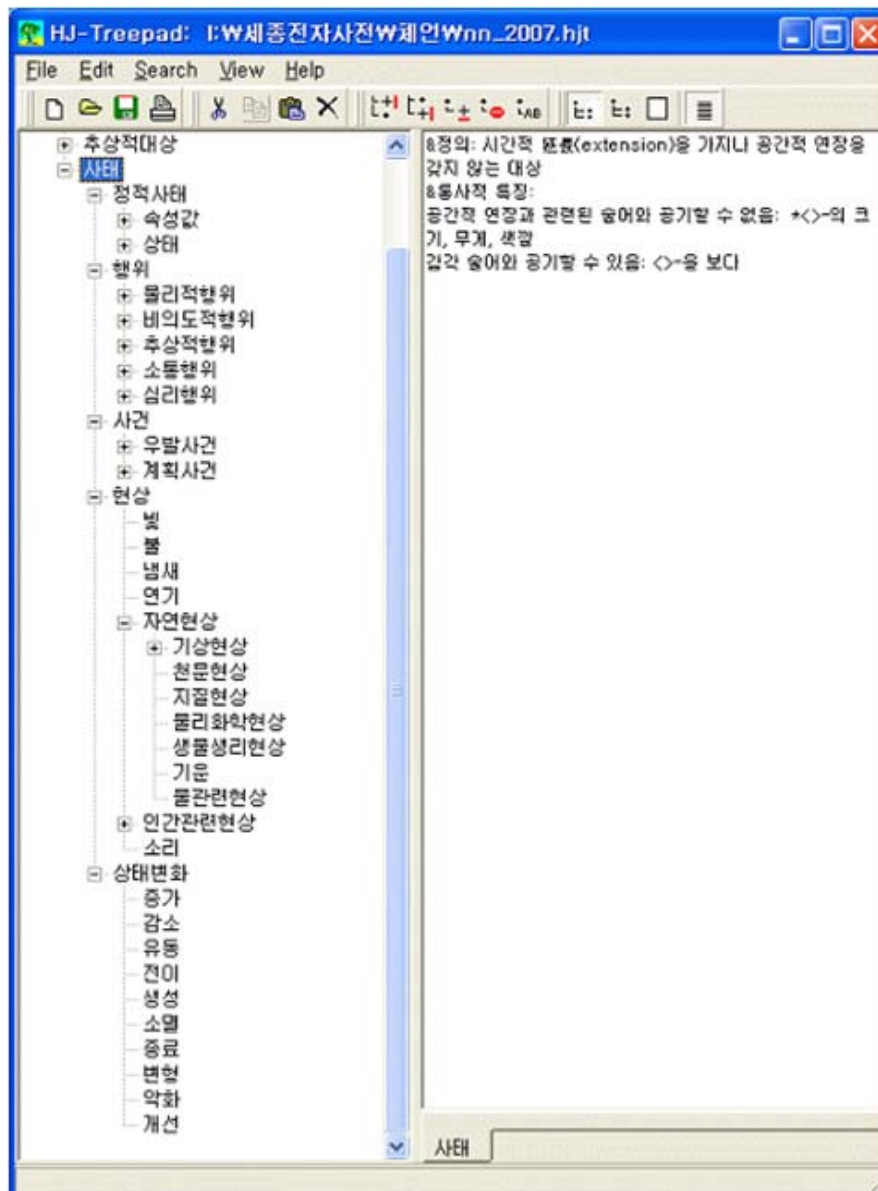
최상위 유의어 세트	포함 용언 수	유의어 뜻풀이
mk_나다__001021 (내다)	3	기풍, 멋 따위가 더 나아지다.
mk_나다__001022 (내다)	336	어떤 작용에 따른 효과, 결과 따위의 현상이 이루어져 나타나다.
mk_날다__001001 (날리다)	1	공중에 떠서 어떤 위치에서 다른 위치로 움직이다.
mk_남다__001001 (남기다)	2	다 쓰지 않거나 정해진 수준에 이르지 않아 나머지가 있게 되다.
psv_꺾다__000007 (꺾이다)	3	경기나 싸움 따위에서 상대를 이기다.
psv_꺾다__000008 (꺾이다)	1	방향을 바꾸어 돌리다.
내뺨치다__000001	1	세차게 뺨치다.
다르다__001001	97	비교가 되는 두 대상이 서로 같지 아니하다.
다투다__000002	4	승부나 우열을 겨루다.
돌려주다__000001	3	빌리거나 뺏거나 받거나 한 것을 주인에게 도로 주거나 갚다.
돌아다니다__000001	32	여기저기 여러 곳으로 다니다.
....		

### 3. 세종 <사태> 의미부류를 활용한 의미 군집

#### 3.1. <사태> 의미 부류

세종전자사전의 의미정보 소분과에서는 의미 부류 체계를 활용한 표제항 어휘 및 구절의 의미 표상과 논항 선택 제약 기술의 정확성 및 일관성을 제고하기 위해 명사 의미 부류 체계를 구축하였다. 세종 명사 의미 부류 체계는 최상위 정점 노드로 묶어진 최상위 부류들에 대해 단계적으로 의미 영역을 분화하여 구축된 의미 부류들의 위계적 체계로, 최상위 부류로는 <구체물>, <집단>, <장소>, <추상적대상>, <사태> 등 5개의 부류가 설정되었으며 이중 처음 4개는 논항명사의 의미 부류이고, 나머지 <사태> 부류는 술어명사의 의미 부류이다.

<사태> 부류는 술어가 표상할 수 있는 의미 영역들을 세분하는 <정적사태>, <행위>, <사건>, <현상>, <상태변화> 등의 의미 부류들과, 이들로부터 분할된 보다 세밀한 의미 영역을 지닌 다양한 층위의 하위 부류(212개)들을 포함하고 있다.



<그림 VI-1> 세종 <사태> 의미부류

### 3.2. 선정 용언의 <사태> 의미부류로의 매핑

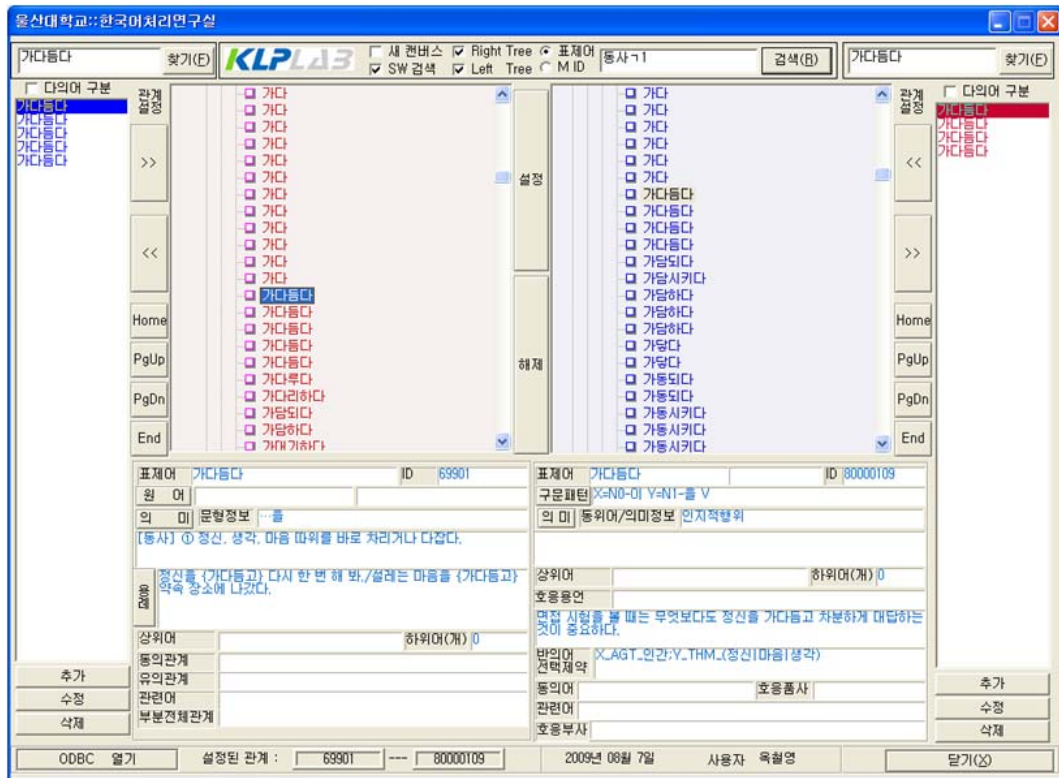
세종전자사전(용언)은 다음 <표 IV-4>와 같이 의미 프레임(frame)별로 의미정보, 구문패턴, 선택제약, 호응부사, 용례 등의 다양한 정보를 포함하고 있으며, 의미정보는 세종 <사태> 의미부류명이다. 세종 <사태> 의미부류는 이번 과제의 의미군집화 및 계층화의 기초 자료로 활용될 수 있다.

<표 VI-4> 세종전자사전(용언)

	C	D	E	F	G	H	I	J	K	L	M
1	sense/Frame_Subsen	표제어	어개 번호	다의 번호	원어	의미정보	구 문 패 턴	선택제약	용언과 호응하는 부사	용례	
106	72	105	가다	2		단독행위	X=N0-이 Y=N2-에 [에게 로]에게로	NY_GOL_장소(태백산 극장):X_AGT_인간 단체:Npr1_행위(사냥 소풍)		그들은 태백	
107	76	109	가다듬다	1	1	인지적행위	X=N0-이 Y=N1-을 V	X_AGT_인간:Y_THM_(정신 마음 생각)		면접 시험을	
108	77	110	가다듬다	1	2	개선행위	X=N0-이 Y=N1-을 V	Y_THM_(목적 목소리 호출):X_AGT_인간		동요 경연대	
109	78	111	가다듬다	1	3	개선행위	X=N0-이 Y=N1-을 V	Y_THM_(표정 자세 옷매무새):X_AGT_인간		그녀는 화장	
110	79	112	가다듬다	1	4	개선행위	X=N0-이 Y=N1-을 V	Y_THM_(전열 대열):X_AGT_인간 인간집단		우리 소대는	
111	80	113	가다듬다	1	1	파동적행위	X=N0-이 Y=N1-을 V	X_THM_인간 인간집단:Y_LOC_단체 유파		세무공무원	
112	81	114	가다듬다	1	1	시도저해의	X=N0-이 Y=N2-에 V=N1-을 V	X_AGT_이개이가진다:Y_THM_이개이가진다:7 LOC_다체 오피 해의 소려를 구			

세종전자사전(용언)의 의미정보를 본 과제의 의미군집으로 활용하기 위해, 우선 2009년까지 선정된 용언(20,284개)을 의미별로 세종전자사전(용언)과 의미 매핑 과정을 거쳤다. 세종전자사전은 뜻풀이는 없으나 용례를 가지고 있어 이 용례와 <표준>의 용언 뜻풀이와 용례를 비교하여 의미 매핑이 가능하다.

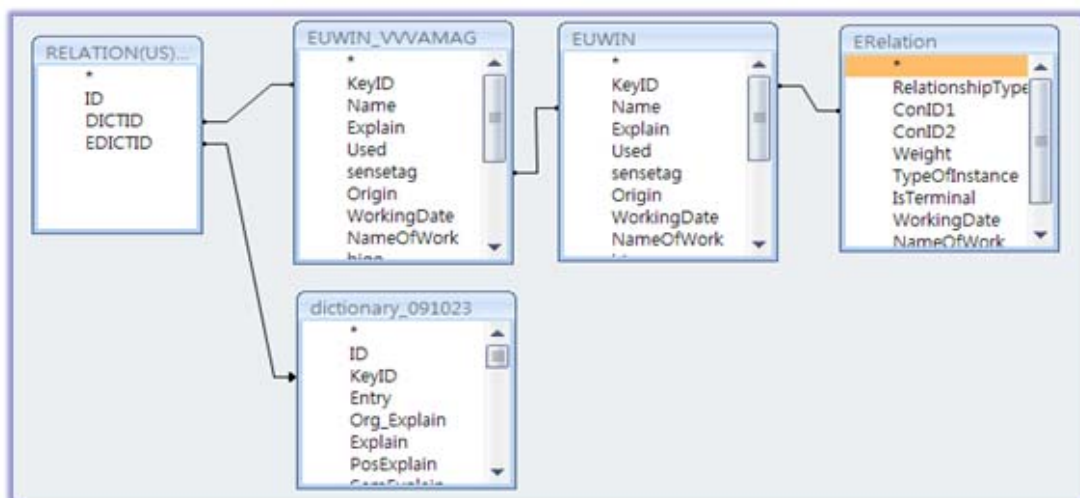
이러한 의미 매핑을 위해 다음 그림과 같은 도구를 개발하여 의미 매핑을 시도하였다.



<그림 VI-2> <표준> 용언과 세종전자사전(용언)의 의미 매핑 도구

이렇게 의미 매핑된 결과는 다음과 같다.

- 직접 의미 매핑되는 용언 : 14,262개 (70%)
- 직접 의미 매핑되지 않은 용언 : 6,022개 (30%)



EUWIN_VV	EUWIN_V	sensetag	EUWIN.k	EUWIN_1.Na	dictio	Entry	Homo.F	bigo	example
80000440	가하다	증가행위	50500000	결과행위	1704	가하다	01_01_01	원금에 높은 이자를 가해 원금에 이자를 가해서 값이	
80000441	가하다	결과행위	49900000	단독행위	1704	가하다	01_01_01	이 상태에서 속력을 더 가 원금에 이자를 가해서 값이	
80000442	가하다	방향성행위	52900000	물리적행위	1706	가하다	01_01_03	이익단체들은 자신들의 강대국은 우리나라에 더 단	
70031232	가하다	평가속성값	46600000	속성값	1807	가하다	02_00_01	그 정도의 행동은 가할 것 큰 공을 세웠으니 큰 상을	
70031233	가하다	평가속성값	46600000	속성값	1807	가하다	02_00_01	새 정부의 민주화 노력을 큰 공을 세웠으니 큰 상을	
70031234	가하다	평가속성값	46600000	속성값	1807	가하다	02_00_01	이 문제로 말하자면 논증 큰 공을 세웠으니 큰 상을	
80000001	가감되다	피동적행위	54900000	비의도적행위	5604	가감되다	00_00_00	국악 합주에서는 향피리 실적에 따라 월급의 액수	
80000002	가감되다	피동적행위	54900000	비의도적행위	5604	가감되다	00_00_00	전체 비용에서 인건비가 실적에 따라 월급의 액수	
80000003	가감하다	결과행위	49900000	단독행위	5606	가감하다	01_00_00	'소요산'이라는 처방에 중 체력에 맞추어 운동량을	
80000006	가격하다	방향성행위	52900000	물리적행위	10202	가격하다	00_00_00	날아오던 축구공이 영희 주먹으로 얼굴을 가격하다	
80000005	가격하다	방향성행위	52900000	물리적행위	10202	가격하다	00_00_00	미국이 신무기로 아프가 주먹으로 얼굴을 가격하다	
80000008	가결되다	피동적행위	54900000	비의도적행위	15902	가결되다	01_00_00	의원 대다수의 찬성으로 안전이 만장일치로 가결되	
80000015	가결하다	종료행위	50200000	결과행위	15904	가결하다	01_00_00	국회가 그 안건을 만장일치로 의안을 만장일치로 가결하	
80000016	가결하다	종료행위	50200000	결과행위	15904	가결하다	01_00_00	국회는 이번 회기에 동태 의안을 만장일치로 가결하	

<그림 VI-3> <표준> 용언과 세종전자사전(용언)의 의미 매핑 결과

<그림 VI-3>의 의미 매핑 결과를 보면 <표준>의 {가하다\_01\_01\_01, 가하다\_02\_00\_01, 가감되다\_00\_00\_00, 가격하다\_00\_00\_00, ...} 등은 세종전자사전에서 두 가지의 의미로 분화하여 정의되어 있다.

직접 의미 매핑되지 않은 용언의 의미부류는 다음 과정을 거쳐 결정하였다.

- (1) 매핑되지 않은 용언의 뜻풀이 유의어가 의미 매핑된 경우는 유의어의 의미범주를 갖도록 했다. 이렇게 유의어의 의미 매핑이 되는 경우는 2,791개이다.

가공하다\_030000 : 두려워하다\_000001/VV+거나/EC 놀라다\_000001/VV+르 /ETM 만하다\_000001/VX+다/EF+./SF {외향적심리행위}

- (2) 2.2.2.절에서 1차 의미군집을 했던 유의어의 연결 구성에서 가장 가까운 유의어의 의미범주로 매핑한다(2,956개).

- 힘쓰다\_000200 ⇨ 공헌하다\_000001 ⇨ 이바지하다\_010001 ⇨ 되다\_010602 ⇨ 놓이다\_000201 ⇨ psv\_놓다\_010201 {긍정적결과행위}  
 - 가무스름하다 ⇨ 감다\_040000 ⇨ 질다\_020101 ⇨ 강하다\_020102 ⇨ 높다\_000003 ⇨ \*있다\_010301 ⇨ 존재하다\_000001 {색속성값}

- 유의어 연결 구성에서 3단계 이상의 유의어와 의미 매핑된 경우는 뜻풀이와 용례를 통해 검증 작업을 거쳐 확정하였다.

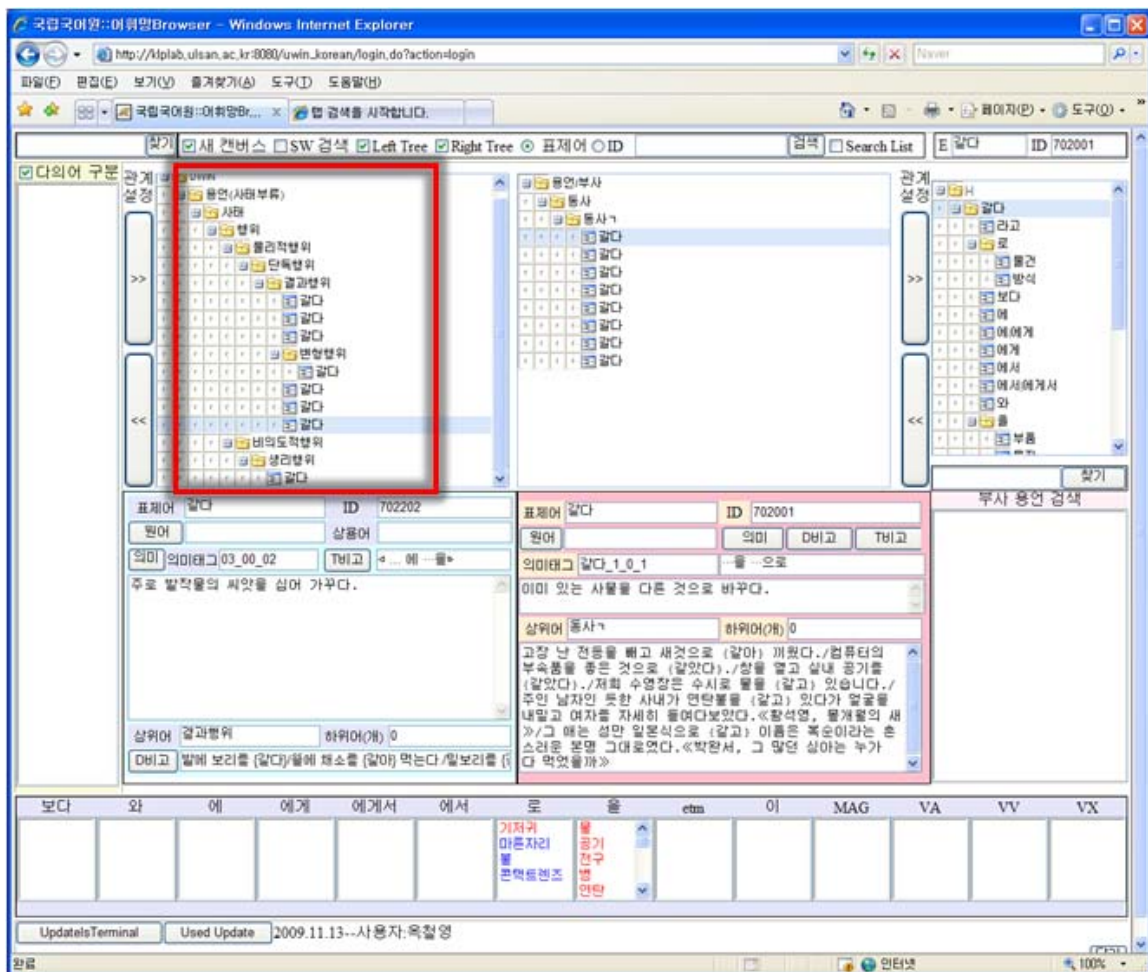
- (3) 2.2.절의 뜻풀이 유의어를 추출하는 과정에서 해당 용언의 뜻풀이가 명사를 참조하는 경우 뜻풀이의 이중 참조까지만 하였기 때문에, 다음과 같이 뜻풀이가 삼중으로 정의된 경우(275개)는 뜻풀이와 용례를

통해 검증 작업을 거쳐 확정하였다.

- 고전하다 : ⇒ 고전09 (股戰)
- 고전09 : ⇒ 고율02 (股慄)
- 고율02 : 무섭다\_\_000100/VA+어서/EC 다리\_\_010001/NNG+가/JKS 떨리  
다\_\_010001/VV+ㅁ/ETN+./SF {피동적행위}

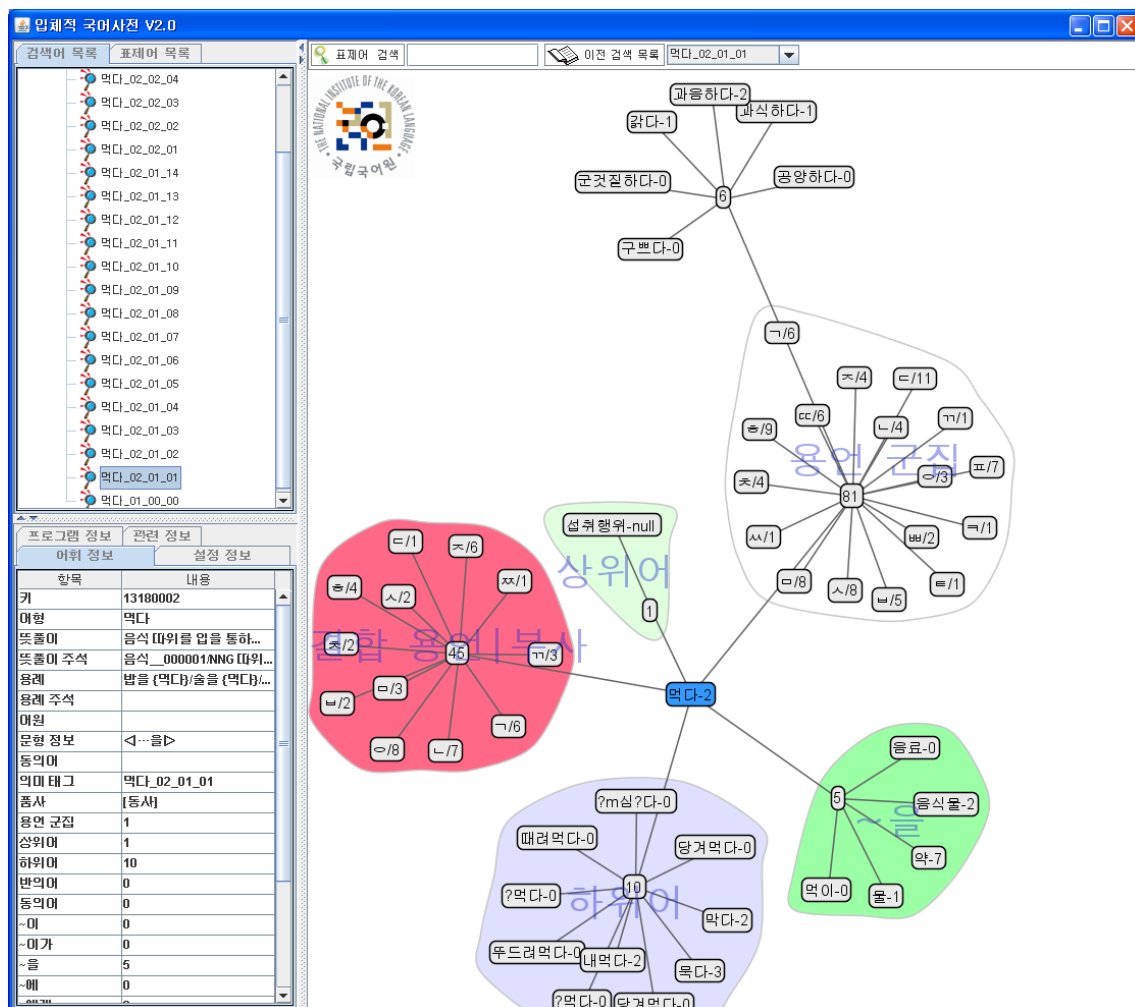
### 3.3. 용언 의미군집의 계층화 및 브라우저

의미부류가 결정된 선정된 용언(20,284개)의 군집화 및 계층화를 위해 다음 그림과 같이 어휘망 편집도구에서 세종 <사태>의미부류와 같은 계층구조가 나타나도록 설계하였다.



<그림 VI-4> 용언 의미 군집의 세종 <사태> 의미부류로의 계층화

또한 입체적 국어사전에서도 군집된 용언들이 나타나도록 하였다. 아래 그림에서 초성이 ‘ㄱ’의 {값다, 과식하다, 과음하다, 공양하다, 군것질하다, 구쁘다} 등이 “먹다\_02\_01\_01(음식 따위를 입을 통하여 배 속에 들여보내다.)”와 의미 군집되었음을 확인할 수 있다. 여기서 ‘먹다\_02\_01\_01’의 상위어 노드 {섭취행위}는 세종 <사태> 의미부류명이다.



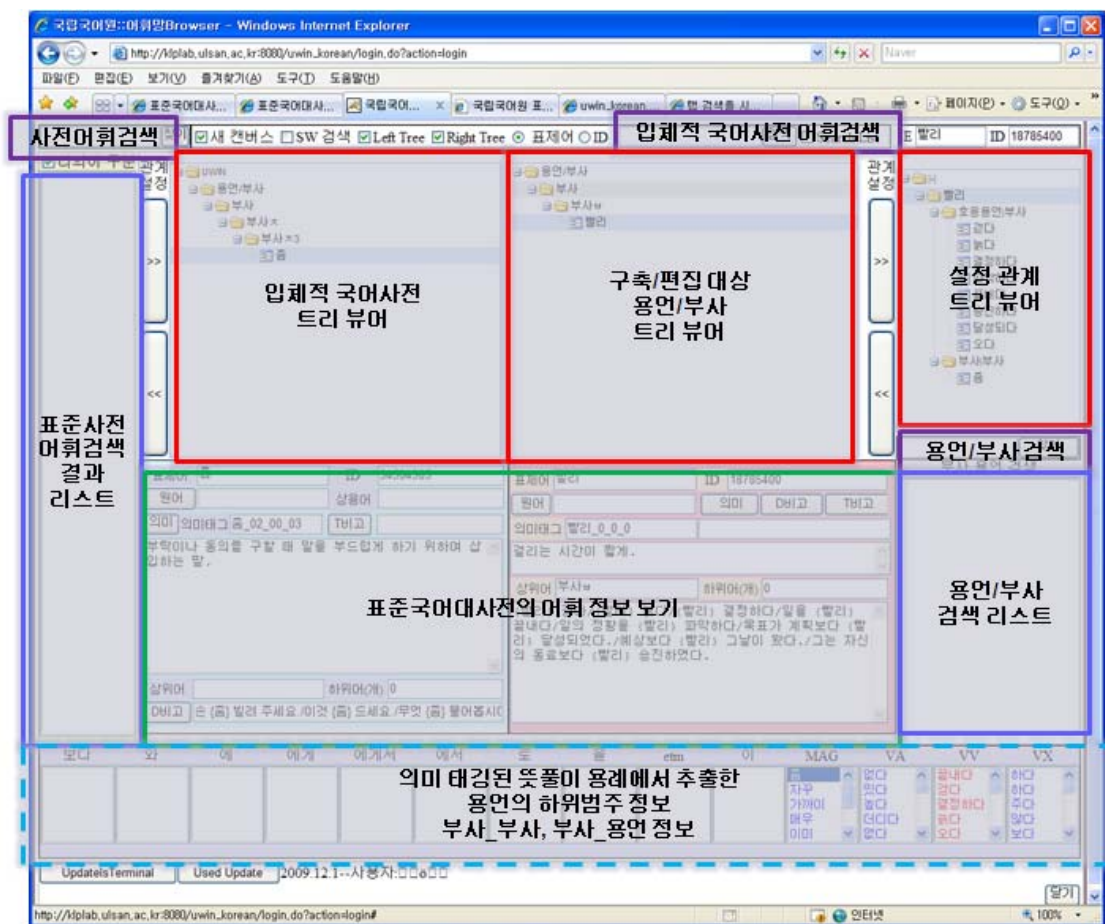
<그림 VI-5> 입체적 국어사전 브라우저에서의 용언 의미 군집

## VII. 입체적 국어사전 브라우저

### 1. 어휘망 편집 브라우저

#### 1.1. 어휘망 편집 브라우저 구성 요소

<그림 VII-1>은 현재 구현된 어휘망 편집 브라우저의 구성 요소이다. 구축한 어휘망 편집 브라우저는 입체적 국어사전 내의 어휘들에 대한 여러 정보를 수정/편집할 수 있는 관리용으로 2009년에 웹용으로 개발되었다.



<그림 VII-1> 어휘망 편집 브라우저의 구성 요소

어휘망 편집 브라우저는 3개의 검색창(표준국어대사전 어휘검색, 입체적 국어사전 어휘검색, 용언·부사검색), 2개의 리스트(표준국어대사전 어휘검색 결과,

용언·부사검색 결과), 3개의 트리 뷰어(입체적 국어사전, 편집대상 용언·부사, 설정관계), 선택용어 표준국어대사전의 어휘 정보 보기 창, 하위정보화 및 부사 관계 설정 대상 어휘 창으로 구성된다.

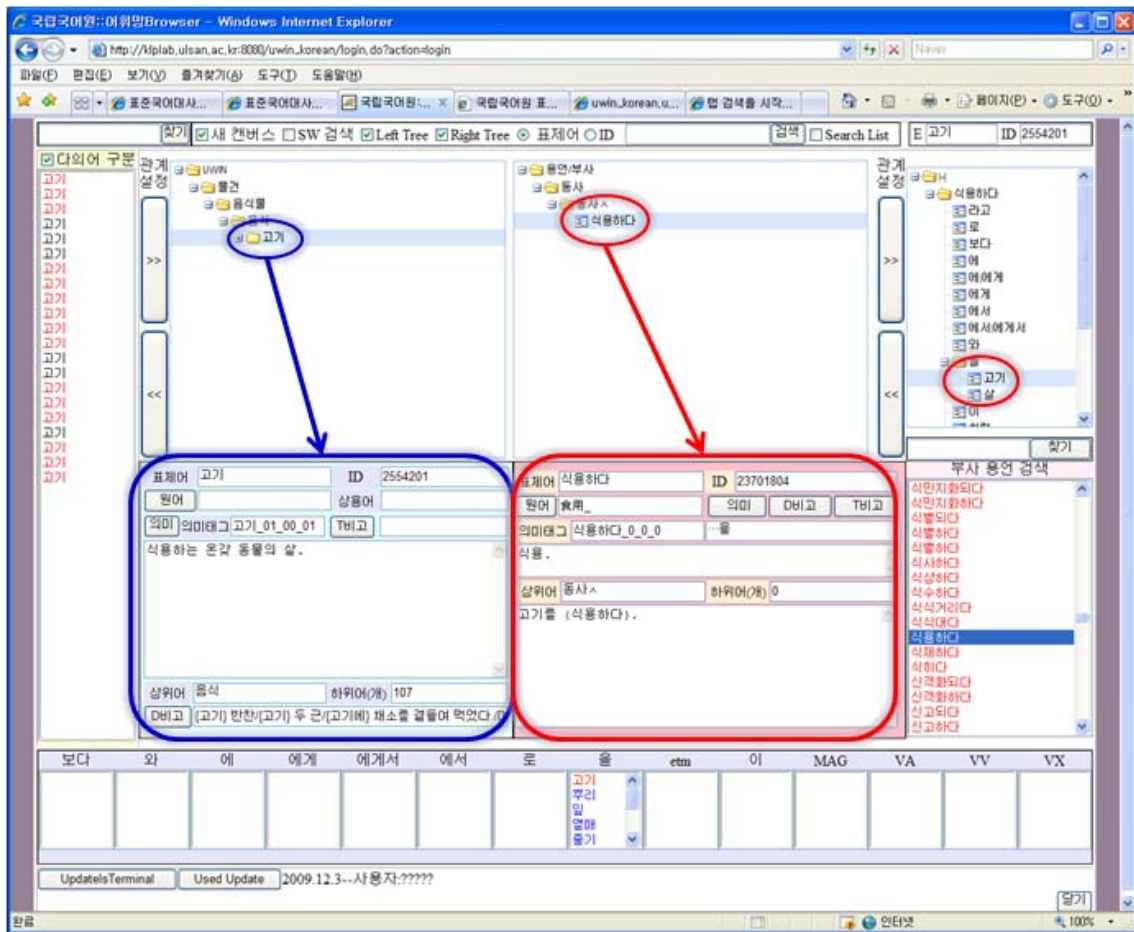
**표준국어대사전 어휘검색 결과 리스트**는 표준국어대사전에 등재된 어휘를 보여주며, 입체적국어사전의 명사 어휘망에 포함된 어휘면 붉은색으로 나타나고, 포함되지 않은 어휘면 검은색으로 나타난다. **용언·부사 검색 리스트**는 하위범주 구축 대상 용언과 부사\_부사 및 부사\_용언 결합관계를 설정하기 위한 선정된 전체 용언·부사를 나타내며, U-WIN 용언 어휘망에 포함된 어휘면 빨간색으로 포함되지 않은 단어이면 검은색으로 나타난다.

**입체적 국어사전 트리 뷰어**는 구축된 명사 어휘망을 상하위 계층구조를 보여준다. **편집대상 용언·부사 트리 뷰어**는 구축된 용언(형용사, 동사)의 구조를 나타내며, 여기서 선택된 용언인 경우는 세종 <사태> 의미부류 계층이 왼쪽 입체적 국어사전 트리 뷰어에 나타난다. **설정관계 트리 뷰어**는 특정 명사, 용언 또는 부사와의 문형에 따른 하위범주 정보, 부사 결합관계 등의 관계를 보여준다.

**표준국어대사전의 어휘 정보 보기 창**은 선택된 어휘에 대한 데이터베이스 내의 관리번호, 표준국어대사전 내의 원어, 뜻풀이, 의미태그, 상위어/하위어, 반의어/동의어, 용례 등의 정보를 제공한다.

**용언의 하위범주 정보 및 부사 결합관계 정보 창**은 다의어 태깅된 뜻풀이 및 용례에서 추출한 논항정보 및 부사\_부사, 부사\_용언 결합관계를 나타내며, 빨간색의 어휘는 용언 및 부사의 용례에서 추출한 어휘이며 검은색은 기타 다른 품사의 용례 혹은 전체 뜻풀이에서 추출한 어휘이다.

‘식용하다’의 검색 결과는 다음 <그림 VII-2>와 같다. 검색은 **용언·부사 검색** 혹은 **입체적 국어사전 검색창**에서 검색해도 된다. 현재 목적격으로 {고기, 살}이 설정되어 있으며 설정된 관계에서 {고기}를 더블클릭하면 명사 어휘망에서의 {고기}의 계층구조와 <표준>에서의 여러 정보를 나타낸다.



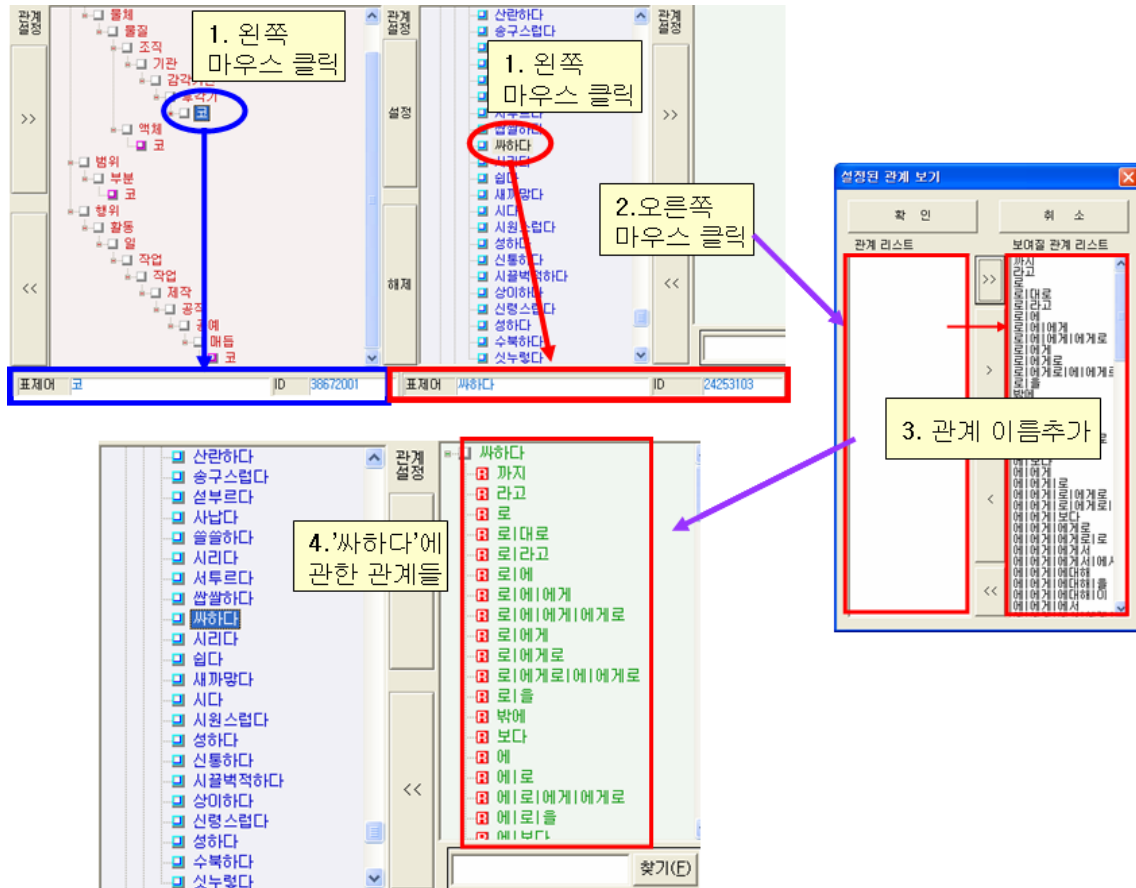
<그림 VII-2> '식용하다'의 검색 결과

## 1.2. 어휘 관계 설정

어휘망 편집 브라우저에서 모든 어휘에 적용되는, 설정된 어휘관계를 보거나 새로운 어휘관계를 추가하기 위해서는 다음과 같은 과정이 필요하다. (<그림 VII-3> 참조)

- ① 우선 하나의 어휘를 선택한다.
- ② 명사를 중심으로 한 어휘의 관계(명사를 주어, 목적어, 부사격으로 가지는 용언)를 보고 싶은 경우 **입체적 국어사전 트리**에서 어휘를 선택한 후 오른쪽 마우스를 누르고, 용언을 중심으로 한 어휘의 관계(용언의 하위범주 정보)를 보고 싶은 경우 용언트리에서 어휘를 선택한 후 오른쪽 마우스를 누른다.

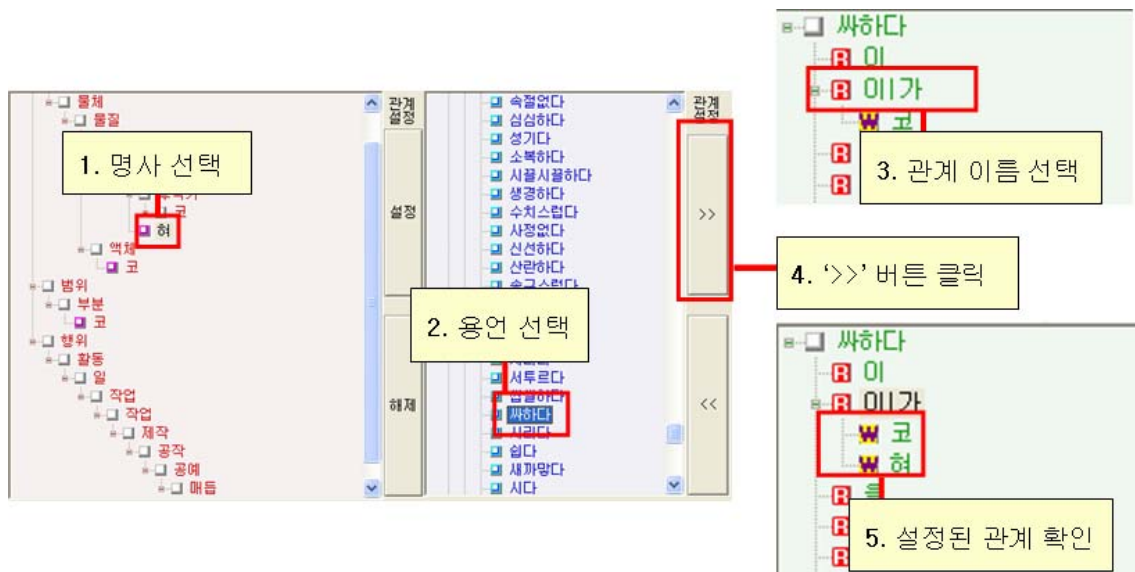
- ③ 그러면 보고 싶은 관계 이름을 선택할 수 있는 윈도우창이 나타나고, 관계 이름 선택 후 '확인' 버튼을 누르면 선택된 관계 이름들이 '관계 트리'에 나타난다.



<그림 VII-3> 공통으로 적용되는 어휘관계 설정

또한 특정 명사와 용언, 용언과 부사의 어휘관계를 설정하기 위해서 <그림 VII-4>와 같은 과정이 필요하다.

- ① 우선 특정 명사, 용언, 설정될 관계를 선택한다. 예를 들어 용언 '싸하다'와 명사 '혀'를 술주관계로 설정하려면, 용언 '싸하다'와 명사 '혀'와 관계 이름 '술주관계'를 선택한다.
- ② 3가지의 요소가 선택된 후 '>>' 버튼을 누르면 관계트리의 술주관계 노드에 '혀'가 단말 노드로 들어가게 된다.

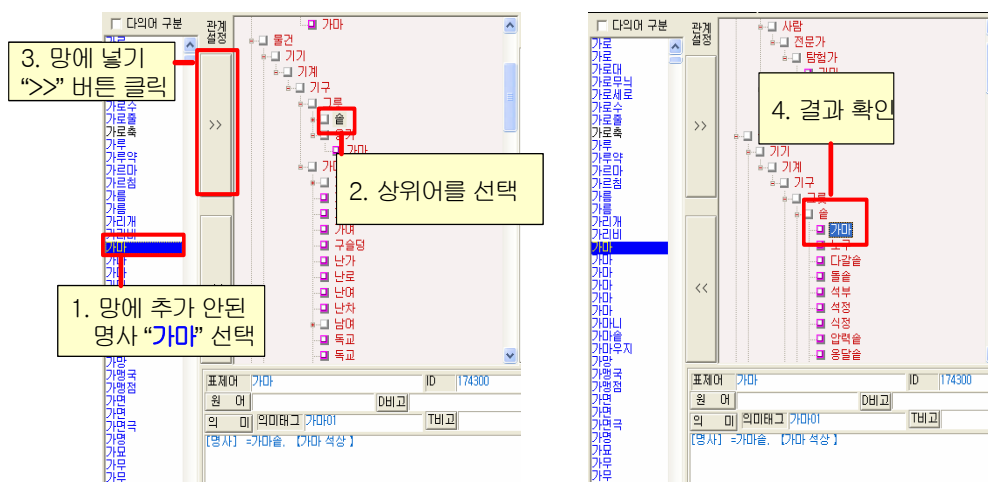


<그림 VII-4> '혀' 와 '싸하다'의 주어 관계 설정

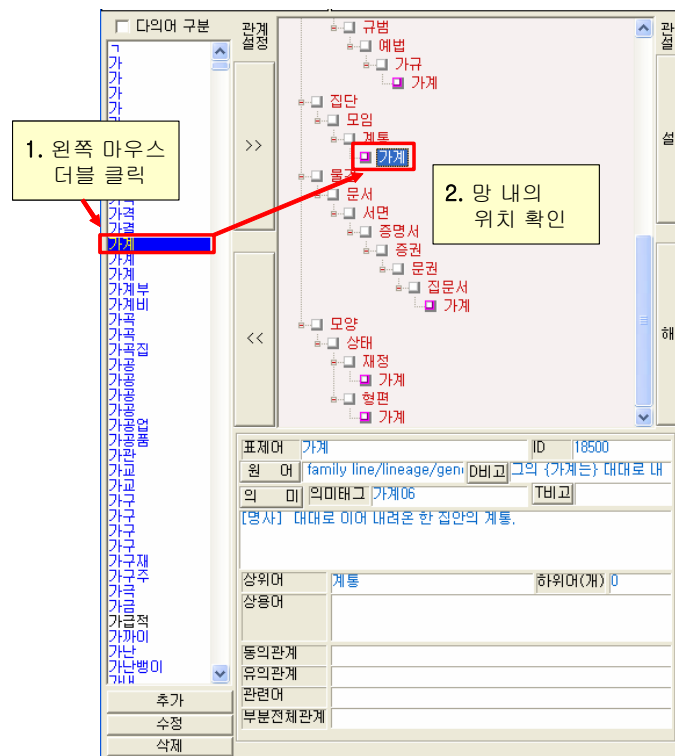
### 1.3. 어휘 추가/수정/삭제

<그림 VII-5>는 어휘망에 용어를 추가하기 위한 과정을 순서대로 보여준다. 명사 어휘망에 등록되지 않은 '가마'라는 용어를 추가하기 위해 뜻풀이를 확인한 후 상위어가 될 수 있는 용어를 선택한다. '술'이라는 상위어를 선택한 후 '>>' 버튼을 누르면 '가마'가 '술'의 하위어로 추가된다.

선택된 어휘가 어휘망 내에 어느 위치에 있는지 확인하려면, 하나의 어휘를 아래와 같이 더블클릭하면 된다.

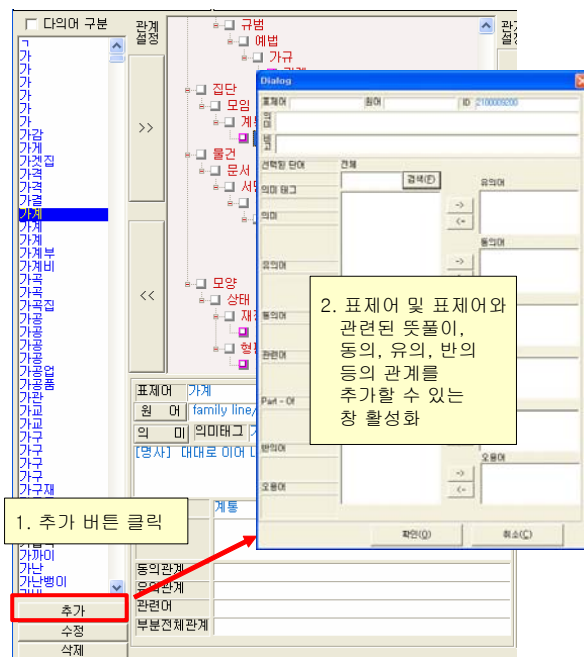


<그림 VII-5> 어휘를 어휘망에 추가하기 위한 과정

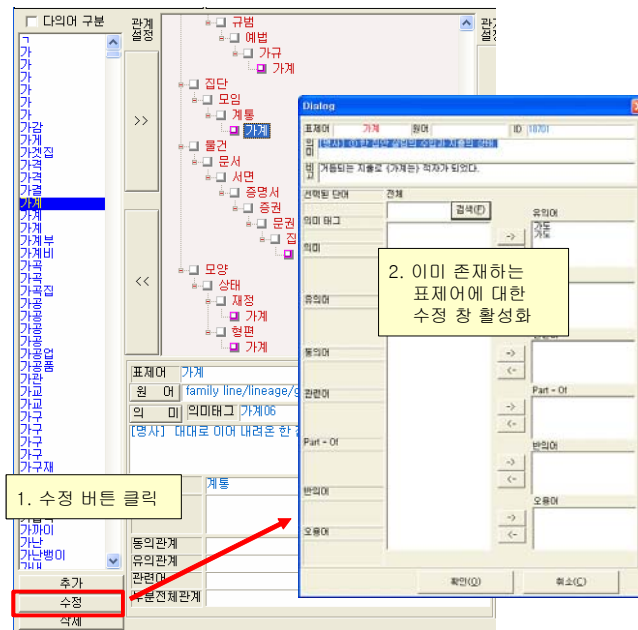


<그림 VII-6> 어휘망 내에 '가계' 위치 확인

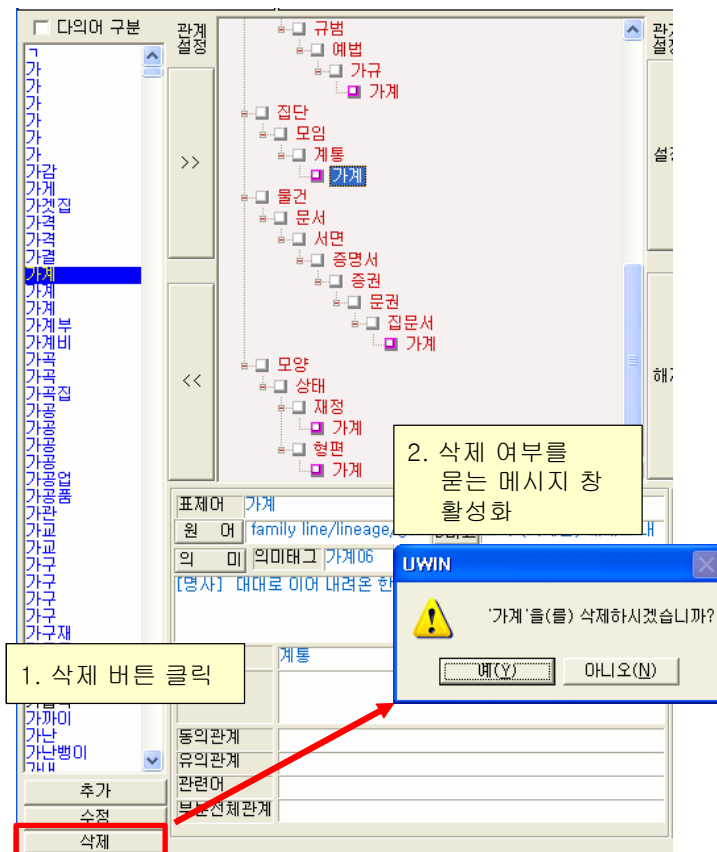
선정된 어휘 목록 추가, 수정, 삭제는 다음 <그림 VII-7>, <그림 VII-8>, <그림 VII-9>와 같다.



<그림 VII-7> 선정된 어휘 목록 추가



<그림 VII-8> 선정된 어휘 목록 수정



<그림 VII-9> 선정된 어휘 목록 삭제

## 2. 입체적 국어사전 브라우저<sup>3)</sup>

### 2.1. 입체적 국어사전 브라우저 구성 요소

<그림 VII-10>은 입체적 국어사전 브라우저로, 검색창, 선정어휘 및 검색어 리스트, 용어정보 보기, 설정된 관계 확인을 위한 동적 뷰어로 구성된다.

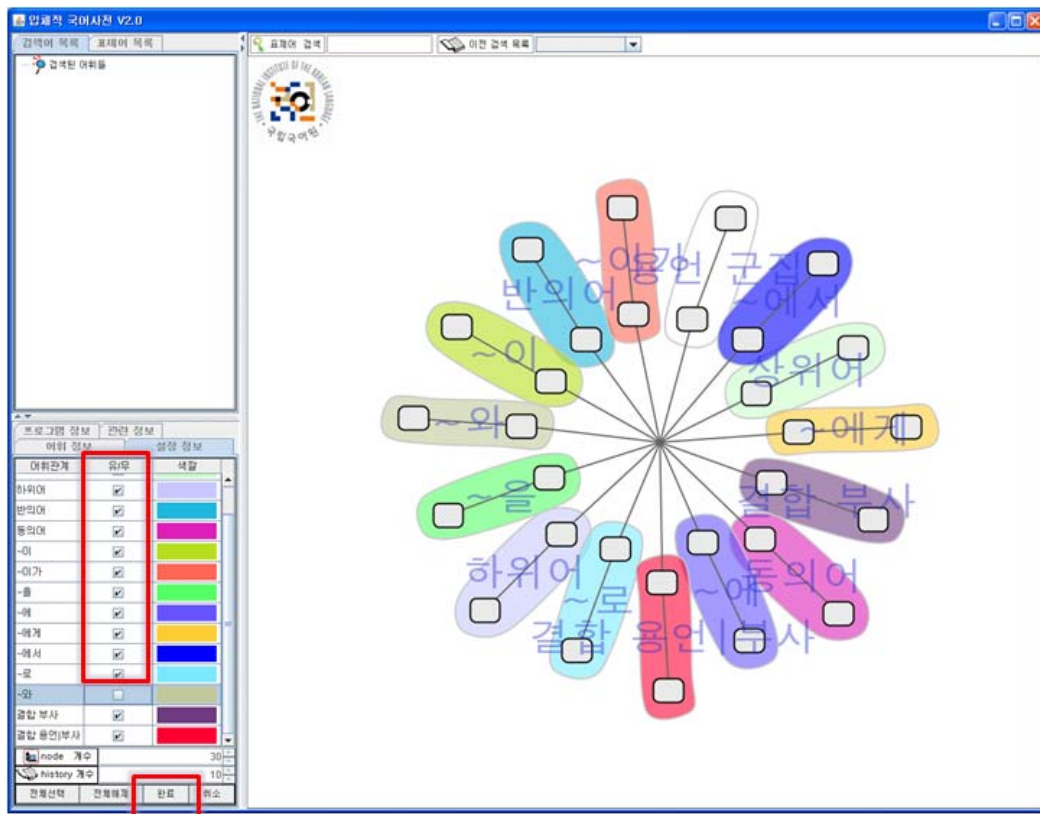
- 검색창은 명사, 용언, 부사 등의 입체적 국어사전에 포함된 어휘의 검색이 가능하며, 검색어 저장(히스토리) 기능이 있어 이전에 찾았던 어휘를 저장해 두어 차후에 손쉽게 찾게 해준다.
- 검색어목록 및 표제어목록에서는 선정 어휘와 검색된 어휘를 리스트 형식으로 볼 수 있으며, 더블클릭을 통해 설정된 관계를 볼 수 있다.
- 설정된 관계 확인을 위한 동적 뷰어는 검색된 용어의 설정된 관계를 동적인 트리 형식으로 보여준다. 마우스 스크롤을 통해 전체 크기를 조절할 수 있고, 끌어놓기를 통해 보고 싶은 부분을 원하는 위치에 가도록 할 수 있다. 특정 용어를 더블 클릭하면 해당 용어의 설정된 관계가 동적 뷰어에 나타나게 된다.
- 어휘정보 보기 창은 특정 어휘가 선택되었을 때 어휘의 ID, 원어, 뜻풀이, 의미태그, 용례 등의 상세 정보를 제공한다.

---

3) [http://124.137.201.197:8080/kooklip\\_v2.jnlp](http://124.137.201.197:8080/kooklip_v2.jnlp)



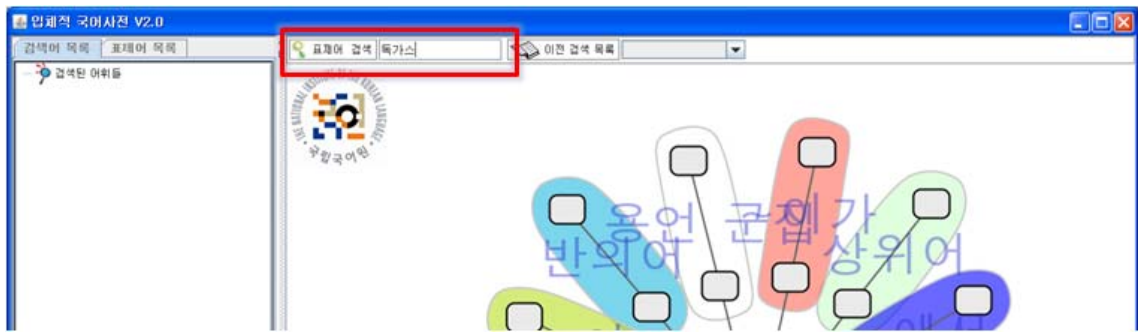
아래의 그림은 {용언군집, 상위어, 하위어, 동의어, 반의어, ~이, ~이|가, ~을, ~에, ~에게, ~에서, ~로, 결합부사, 결합용언|부사}의 관계를 선택한 후 완료 단추를 눌러 설정하는 방법을 보여준다. 기본적으로 최대 노드의 개수는 30개, 이전 검색 목록의 개수는 10개로 설정되어 있다.



<그림 VII-11> 설정 정보 변경

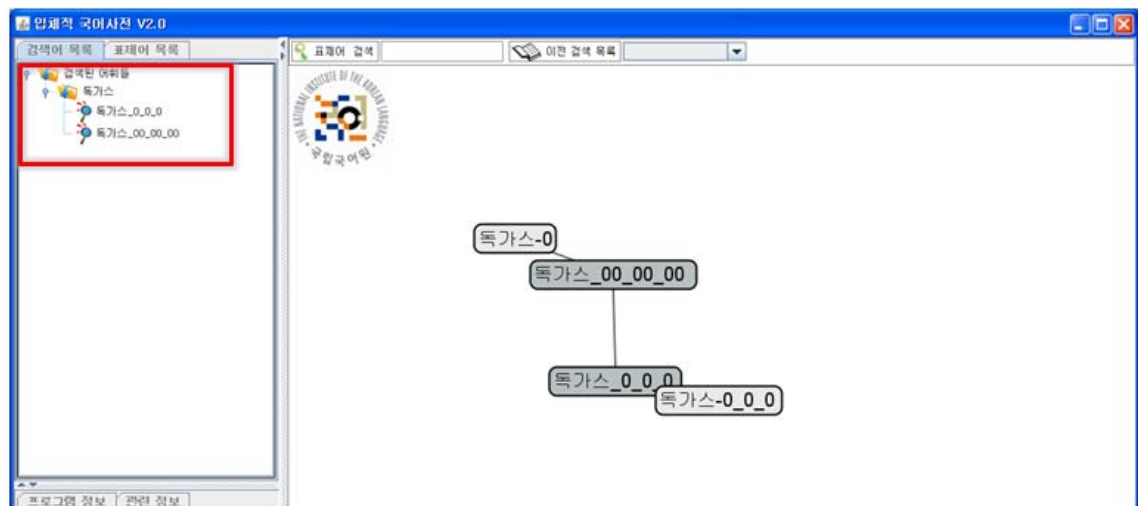
## 2.2.2. 검색어 입력 및 검색어 보기

사용자가 검색하고 싶은 어휘를 검색창에 입력한 뒤 <엔터(Enter)> 키를 누르면 입력된 어휘에 대한 상세정보가 출력된다. 검색창에는 “표제어 검색” 문구가 있다. 아래의 <그림 VII-12>는 어휘 ‘독가스’를 검색창에 입력한 것이다.



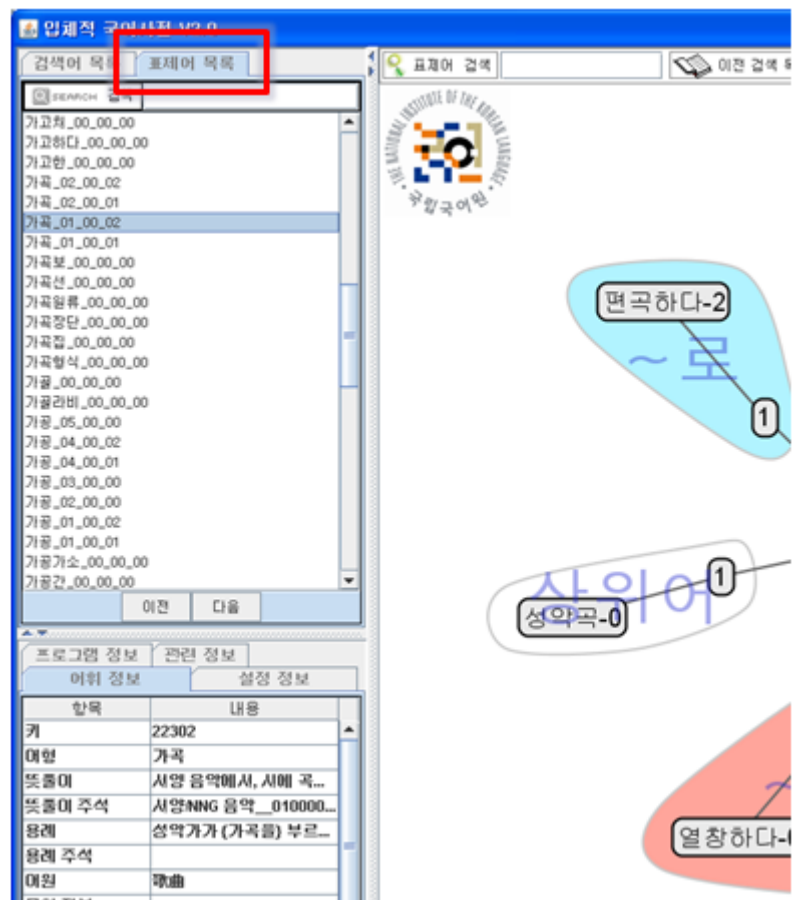
<그림 VII-12> 표제어 검색

아래의 <그림 VII-13>은 ‘독가스’를 검색한 후 검색어가 “검색어 목록”에 추가된 모습을 보여준다. 해당 어휘의 정보를 보고 싶을 경우 단말 노드의 어휘를 두 번 클릭하면 된다.



<그림 VII-13> 검색어 목록에 추가된 검색어

“검색어 목록” 탭 오른쪽에 “표제어 목록” 탭이 있다. “표제어 목록” 탭은 <그림 VII-14>와 같이 입체적 국어사전에 등록된 어휘 전체를 오름차순으로 정렬하여 보여준다. “검색”창에 어휘를 입력하면 아래의 리스트에서 그 어휘 위치로 이동하게 된다. 그리고 검색어 목록 트리와 마찬가지로 어휘의 정보를 보고 싶을 경우 어휘를 두 번 클릭하면 된다.



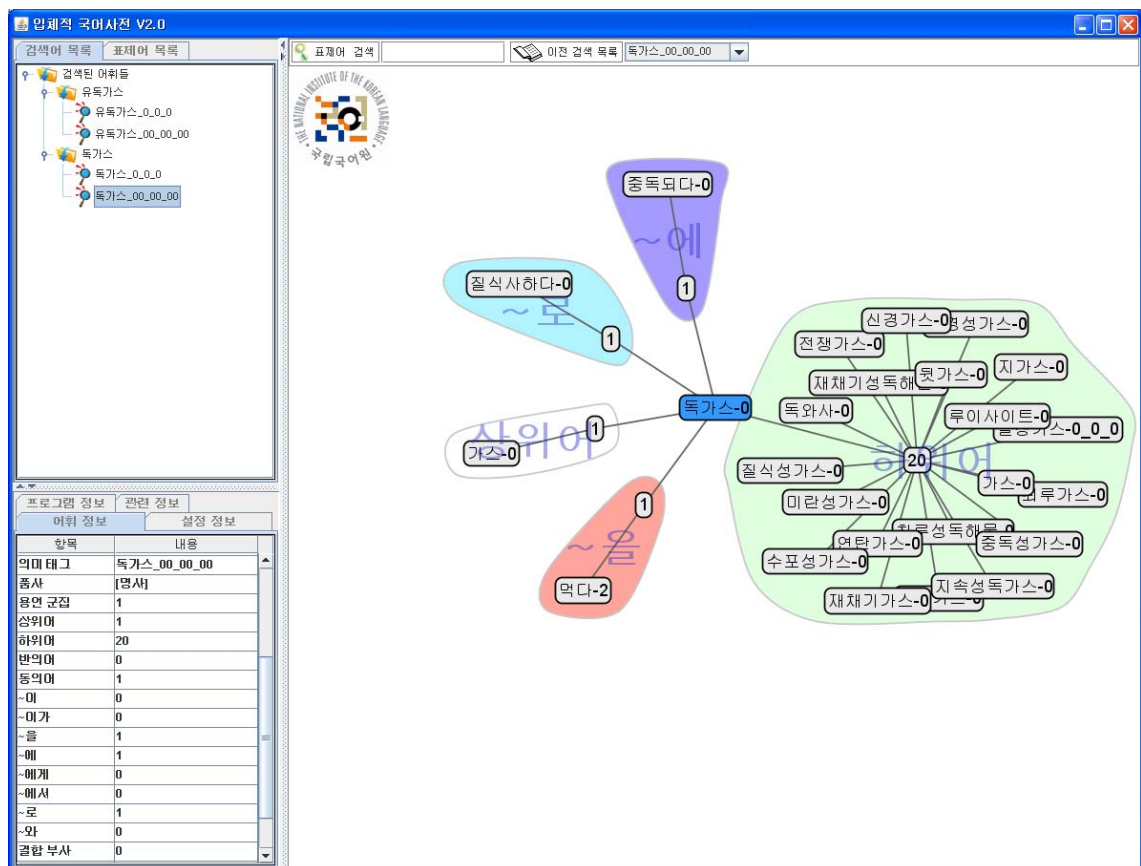
<그림 VII-14> 표제어 목록

### 2.2.3 검색 어휘 정보 보기

왼쪽 상단 트리에 나타난 {독가스\_00\_00\_00}를 두 번 클릭하면 <그림 VII-15>와 같이 오른쪽 창에 {독가스\_00\_00\_00}가 가진 의미 관계들 즉, 상위어, 하위어, 동의어, ~을, ~로, ~에 등이 나타난다. 여기서 {~을, ~로, ~에}는 V장의 용언 하위범주 정보 구축 시에 {독가스\_00\_00\_00}를 최소상계 노드로 하는 용언들의 집합이다. 즉, 동사 ‘중독되다’의 부사격 “에”의 하위범주 정보로 {독가스\_00\_00\_00}가 최소상계노드로 설정되어 있음을 나타낸다.

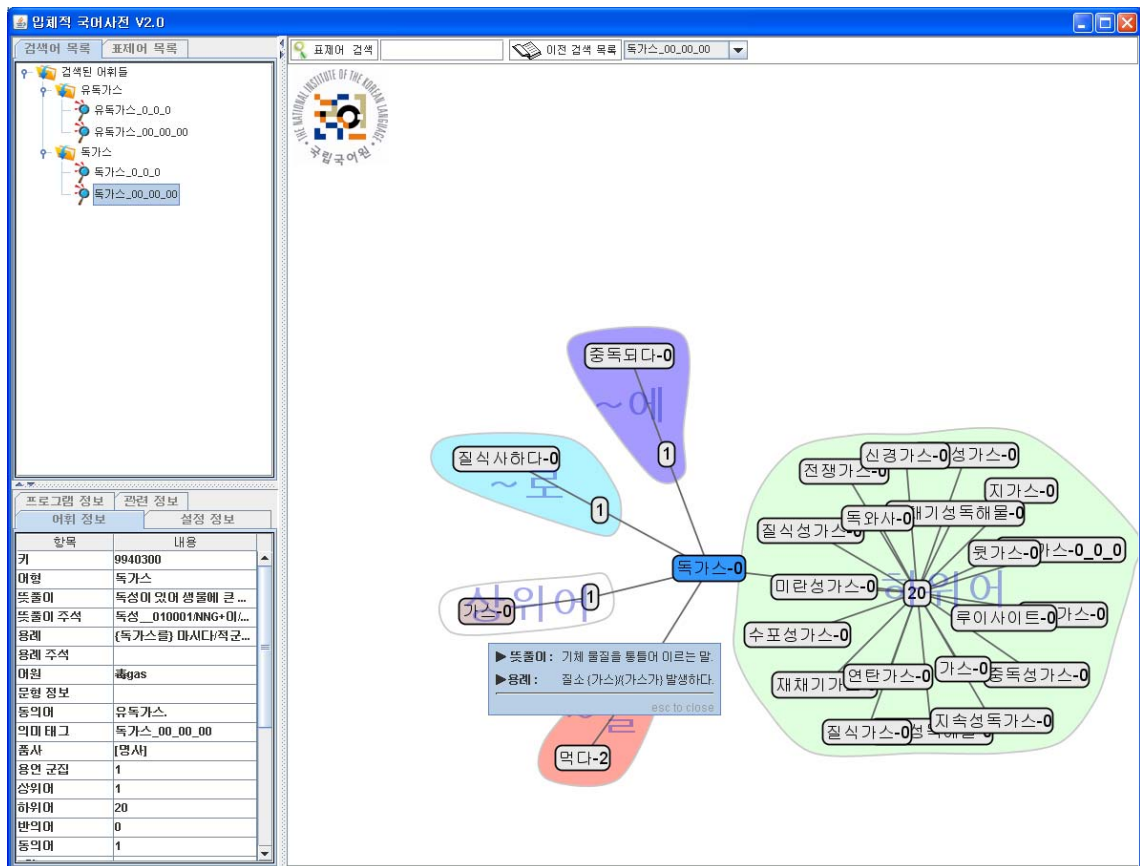
동적 뷰어 창에 나타난 어휘들을 클릭하면 왼쪽 중앙에 있는 ‘어휘 정보’에서 선택된 어휘의 정보를 볼 수 있다. 표제어, 한자, 원어, 상위어, 하위어, 동의어, 반의어, 결합용언|부사, ~이(주격) 관계, ~을(목적격) 관계, ~에(부사

격) 관계, ~로(부사격) 관계, ~에게(부사격) 관계, 뜻풀이, 형태소 분석과 의미 분석된 뜻풀이, 의미태그, 문형정보, 동의어목록, 분야 정보를 볼 수 있다. 특정 항목의 내용이 길어 다 볼 수가 없을 경우, 마우스를 그 항목에 올려놓으면 모든 내용을 볼 수가 있다. (<그림 VII-16> 참조)



<그림 VII-15> 선택된 어휘(예, '독가스')에 대한 의미 관계 정보

<그림 VII-16>은 왼쪽 상단에 있는 트리 상의 어휘에 마우스를 올려놓았을 때 그 어휘의 뜻풀이 정보를 보여주는 모습이다. 왼쪽의 트리뿐 아니라 오른쪽 창에서 나타난 어휘에 마우스를 올려놓아도 마찬가지로 뜻풀이 정보를 보여준다.

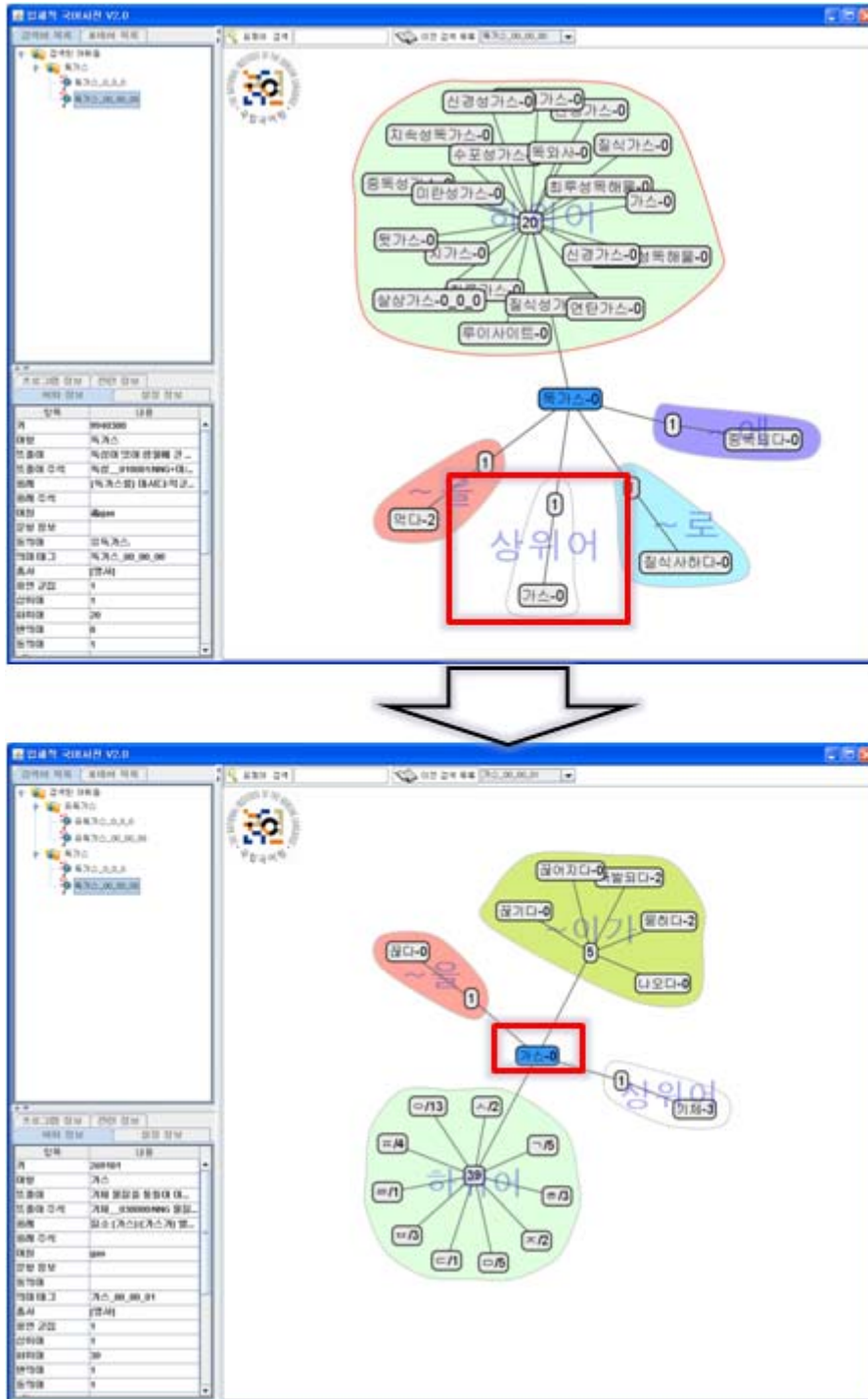


<그림 VII-16> 어휘 '가스'에 마우스를 올려 뜻풀이 및 용례 보기

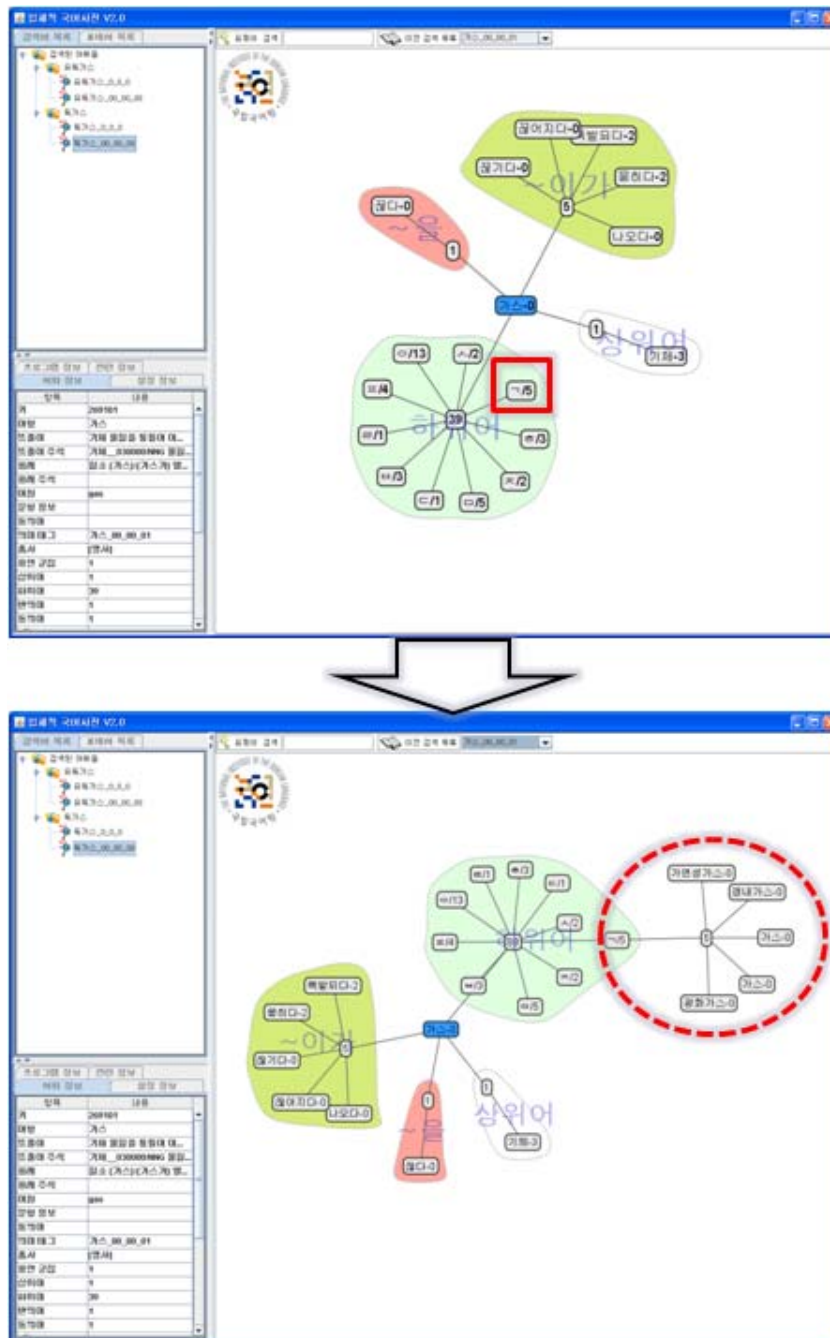
## 2.2.4 상위어, 하위어 따라가기

명사 어휘망은 상위어, 하위어를 기반으로 구축되었으며, 입체적 국어사전 브라우저에서 상위어, 하위어를 1단계씩 따라 갈 수 있으며, 특정 어휘의 상위어를 계속 따라가다 보면 최상위어에 도달하게 된다.

아래 <그림 VII-17>은 {독가스}의 상위어 {가스}를 따라간 후의 모습을 보여준다. 처음에는 '독가스'가 중심에 있지만 상위어를 두 번 클릭한 후에는 {가스}가 동적 뷰어의 중심에 자리하며, {가스}의 하위어(39개)와 {가스}를 최소상계노드로 하는 용언들의 집합이 나타난다. {가스}의 하위어가 39개로 가독성을 높이기 위하여 하위어들은 초성별로 묶여 있으며(예, 노드 {ㄱ/5}는 초성 'ㄱ'인 하위어가 5개 있음을 나타낸다. (<그림 VII-18> 참조)



<그림 VII-17> 상위어 따라가기

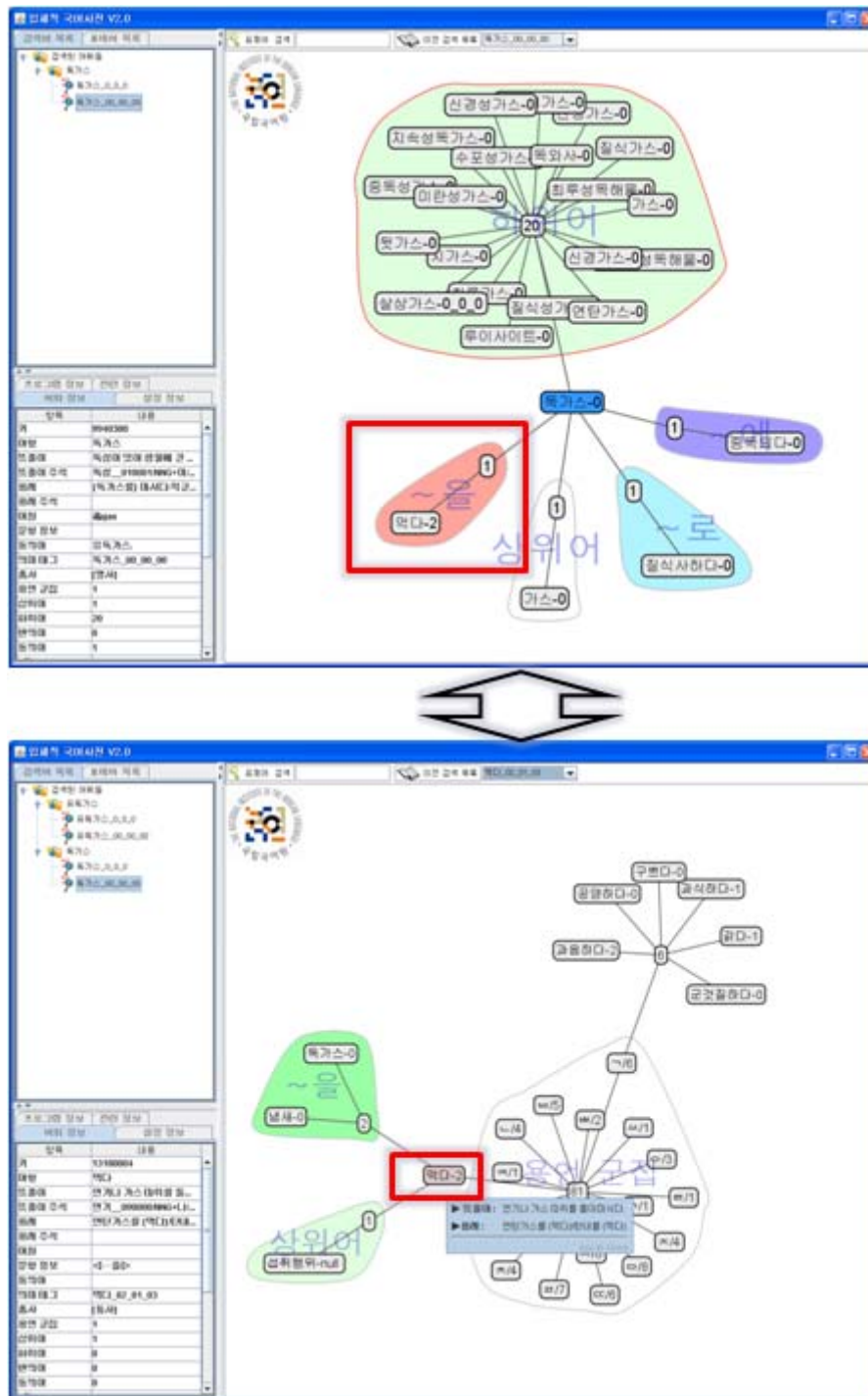


<그림 VII-18> 30개 이상인 하위어 보기

## 2.2.5. 양 방향 추적

입체적 국어사전의 동적 뷰어는 양 방향(용언에서 명사, 명사에서 용언, 용언에서 부사, 부사에서 용언) 추적이 가능하며, 특정 어휘에 대해 설정된 다양한

의미 관계를 편리하게 찾아볼 수 있다.



<그림 VII-19> 양방향 추적

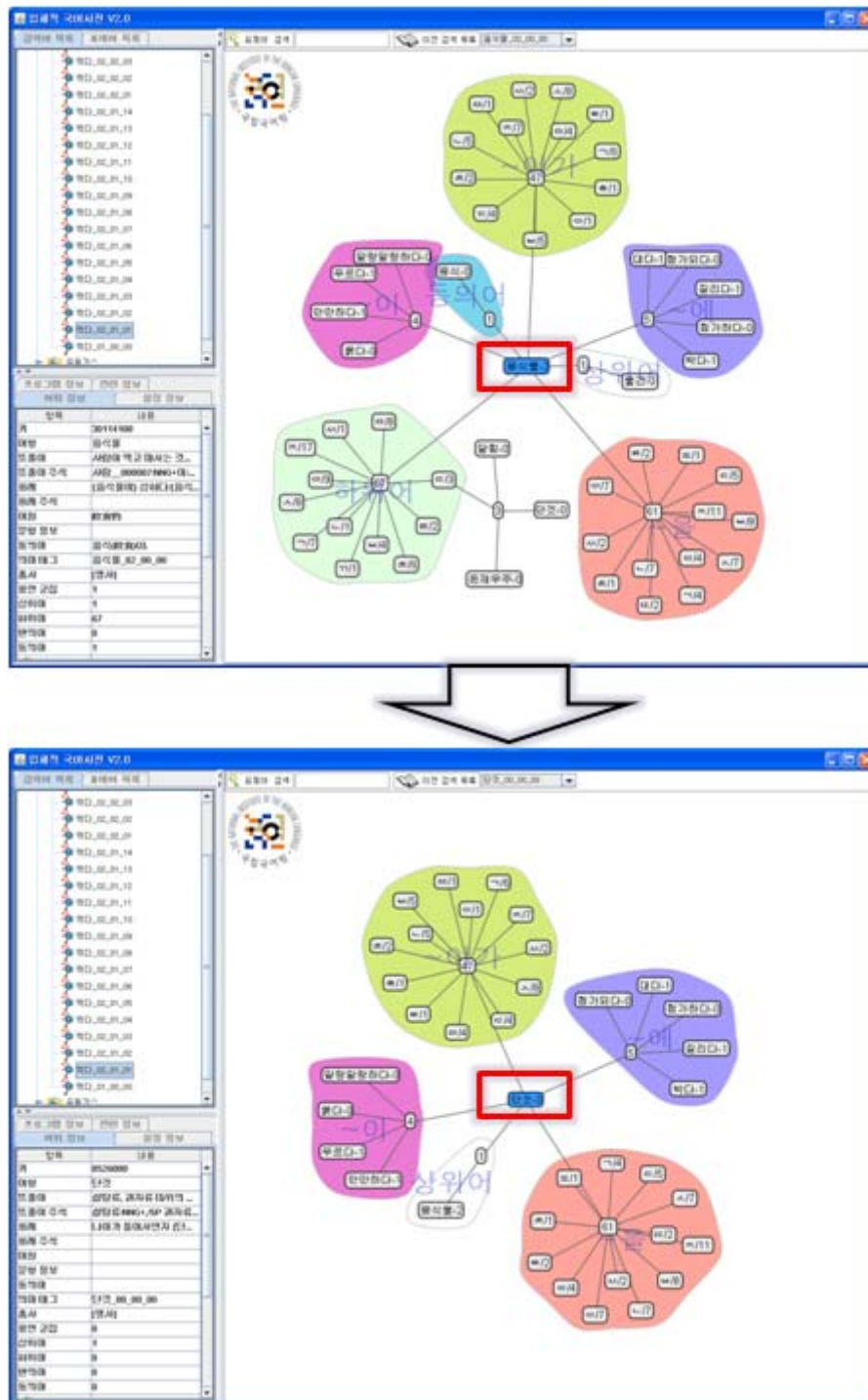
<그림 VII-19>에서 최소상계노드 {독가스\_00\_00\_00}를 목적격으로 가지는

동사 {먹다\_02\_01\_03(연기나 가스 따위를 들이마시다)}를 더블클릭하면 동사 {먹다\_02\_01\_03}을 중심으로 {먹다\_02\_01\_03}의 최소상계노드 {독가스, 냄새}와 {먹다\_02\_01\_03}의 의미(섭취행위)로 군집화된 용언군집이 나타난다. 만약 {먹다\_02\_01\_03}가 다른 하위범주 정보와 부사\_용언 결합관계를 가지면 이들도 함께 나타난다. 용언의 상위어로 나타나는 {섭취행위} 노드는 세종 <사태> 의미부류이며, {섭취행위}로 군집화된 용언 중에서 초성 ‘ㄱ’인 동사는 {공양하다, 값다, 구쁘다, 군건질하다, 과식하다, 과음하다}이다.

## 2.2.6. 최소상계노드의 용언 하위범주 정보 상속

모든 명사를 대상으로 용언의 하위범주 정보는 구축하는 것은 현실적으로 어려움이 많다. 세종전자사전의 경우도 어휘의 의미표상 시에 적정술어를 효과적으로 활용되도록 명사 의미 부류 체계를 확립하고, 용언의 논항의 선택제약을 명사 의미 부류 체계에 따라 제공하고 있다(<표 VI-4> 참조).

입체적 국어사전의 용언의 하위범주 정보는 명사 어휘망에서 해당 논항의 의미를 포함하는 최소상계노드로 구축하였다. 따라서 하위범주 정보에 해당되는 최소상계노드의 하위어도 최소상계노드를 하위범주 정보로 가지는 용언을 상속 받아야 한다.

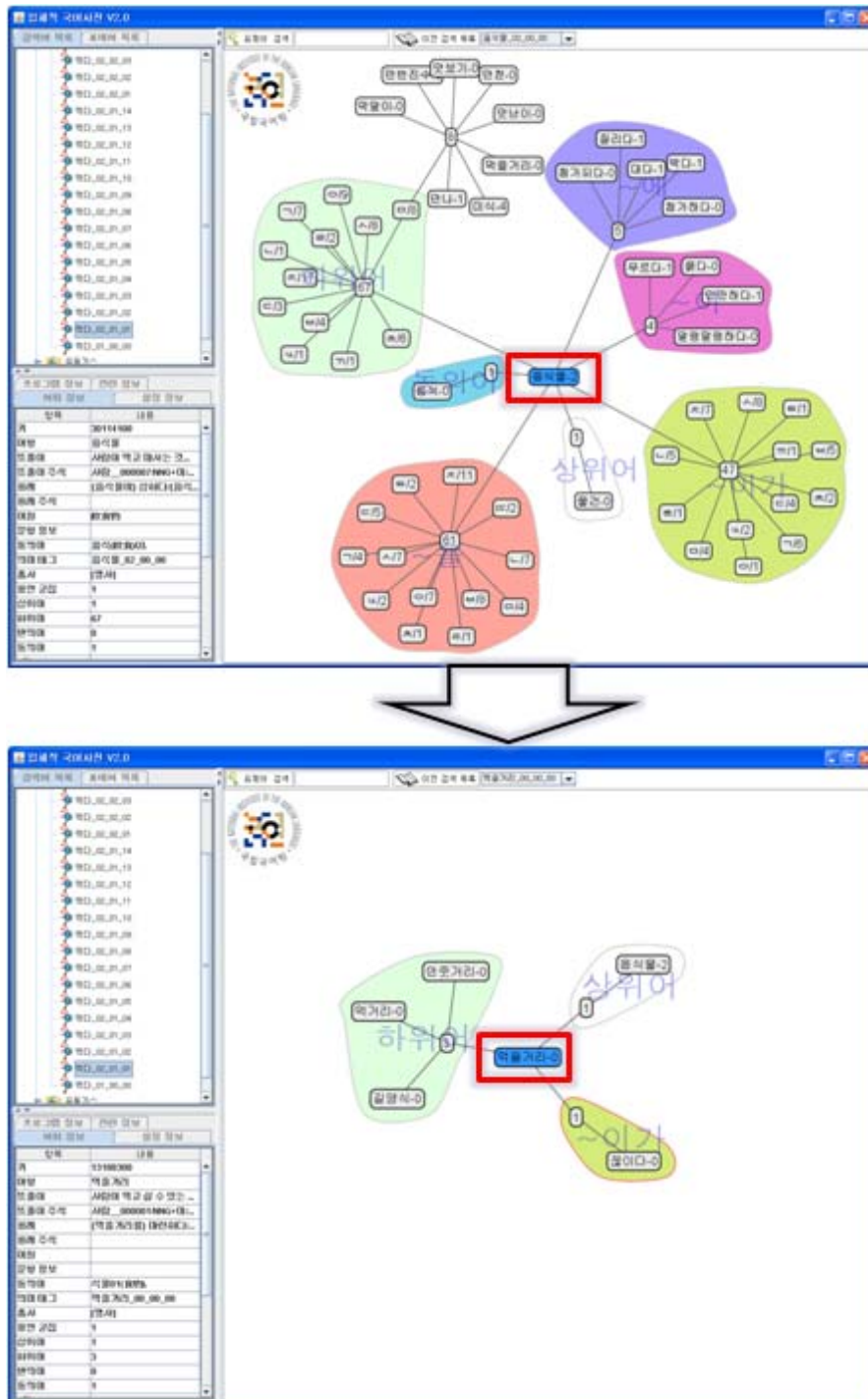


<그림 VII-20> 최소상계노드의 용언 하위범주 정보 상속

<그림 VII-20>은 용언의 하위범주 정보가 상속되는 현상을 보이고 있다. {음식물}은 {~이, ~이|가, ~에, ~을}의 하위범주 정보로 최소상계노드이며, {음식물}의 하위어인 {단것}을 더블클릭하면, {음식물}을 하위범주로 가지는 {~이, ~이|

가, ~에, ~을}의 용언이 그대로 상속되어 나타남을 확인할 수 있다.

이때 만약 하위어가 다른 하위범주 정보로 설정된 것이 있을 경우는 <그림 VII-21>처럼 하위범주 정보를 상속받지 않고 자신과 연결된 하위범주 정보만을 나타낸다.



<그림 VII-21> 용언 하위범주 정보를 상속 받지 않는 경우

<위 그림 VII-21>에서 {음식물}의 하위어인 {먹을거리}는 동사 {끓이다}의 {~이|가}의 하위범주 정보로 설정되어 있기 때문에, {음식물}이 {~이, ~이|가, ~에, ~을}의 하위범주로 가지는 용언들에 상속되지 않는다.

이러한 하위범주 정보의 상속은 명사와 용언 간에만 발생되고 부사나 다른 의미 관계(반의어, 동의어)에는 영향을 미치지 않는다.

### 3. 브라우저의 데이터베이스 설계서

#### 3.1. 테이블 종류 및 구조

웹 브라우징을 위한 데이터베이스 내의 각종 테이블은 목록은 <표 VII-1>과 같다.

<표 VII-1> DB 내 각종 테이블 목록

테이블 이름	필드 개수	테이블 설명
UWIN	12	선정된 기본어휘의 표제어, 뜻풀이, 뜻풀이의 형태소/의미태깅 결과, 어깨번호, 다의어번호, 문형번호, 용례, 원어정보, 문형정보, 동의어 목록을 보여주기 위한 테이블. 명사, 용언, 부사를 포함하고 있다. 어휘관계 웹용 뷰어와 편집도구의 왼쪽 트리를 보여주기 위한 정보들을 담고 있음.
EUWIN	12	선정된 기본어휘의 표제어, 뜻풀이, 뜻풀이의 형태소/의미태깅 결과, 어깨번호, 다의어번호, 문형번호, 용례, 원어정보, 문형정보, 동의어 목록을 보여주기 위한 테이블. 용언, 부사를 포함하고 있다. 어휘관계 편집도구의 중앙 트리를 보여주기 위한 정보들을 담고 있음.
Relation	9	선정된 기본어휘의 상하위 관계, 동의 관계를 설정하기 위한 테이블. 관계를 설정한 사람의 이름, 날짜, (설정된 이유)를 기록함.
ERelation	9	선정된 기본어휘의 상하위 관계, 동의 관계를 설정하기 위한 테이블. 관계를 설정한 사람의 이름, 날짜, (설정된 이유)를 포함.
RName	3	어휘간의 특정 관계 이름을 관리하기 위한 테이블. 어휘간의 관계가 한 방향 또는 양 방향으로 적용 되는 지 여부를 포함..
rel-XXX	9	선정된 기본어휘의 특정관계를 설정하기 위한 테이블. 테이블 이름의 'XXX' 부분에 특정 관계 이름이 들어감. 예를 들어 술목관계를 나타내는 '~을'에 대한 관계 테이블 이름은 'rel-을'이 됨. RName의 레코드 수만큼 rel-XXX 테이블이 생성됨. 특정 관계를 설정한 사람의 이름, 날짜, (설정된 이유)를 기록함.

각 테이블에 대한 상세 구조는 <표 VII-2> ~ <표 VII-5>와 같다.

<표 VII-2> Relation 테이블 구조

Relation : 테이블					
필드 이름		데이터 형식	설명		
RelationshipType		숫자	BT(상의어): 100, USE(동의어): 200		
ConID1		숫자	자신의 아이디		
ConID2		숫자	자신과 관계를 가지는 개념의 아이디		
Weight		숫자	두 개념 사이의 의미적 관련정도(이부분은 일단 신경쓰지 말기 바람)		
TypeOfInstance		숫자	인스턴스의 유형(이부분은 일단 신경쓰지 말기 바람)		
IsTerminal		숫자	단말노드인지체크(0:중간노드; 1:단말노드)		
WorkingDate		텍스트	최근 작업 일자		
NameOfWork		텍스트	최근 작업자 성명		

RelationshipType		ConID1		ConID2	
필드 크기	정수(Long)	필드 크기	정수(Long)	필드 크기	정수(Long)
형식		형식		형식	
소수 자릿수	자동	소수 자릿수	자동	소수 자릿수	자동
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값		기본값		기본값	
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	예	필수	예	필수	예

Weight		TypeOfInstance		IsTerminal	
필드 크기	실수(Double)	필드 크기	정수(Long)	필드 크기	정수(Long)
형식		형식		형식	
소수 자릿수	자동	소수 자릿수	자동	소수 자릿수	자동
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값	1	기본값	0	기본값	0
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	예	필수	예	필수	예

WorkingDate		NameOfWork		Bigo	
필드 크기	50	필드 크기	50	필드 크기	255
형식		형식		형식	
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값		기본값		기본값	
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	아니오	필수	아니오	필수	아니오
빈 문자열 허용	예	빈 문자열 허용	예	빈 문자열 허용	예
인덱스	아니오	인덱스	아니오	인덱스	아니오
유니코드 압축	아니오	유니코드 압축	아니오	유니코드 압축	아니오
IME 모드	한글	IME 모드	한글	IME 모드	한글
IME Sentence 모드	없음	IME Sentence 모드	없음	IME Sentence 모드	없음

<표 VII-3> ERelation 테이블 구조

ERelation : 테이블			
필드 이름	데이터 형식	설명	
RelationshipType	숫자	BT(상의어): 100, USE(동의어): 200	
ConID1	숫자	자신의 아이디	
ConID2	숫자	자신과 관계를 가지는 개념의 아이디	
Weight	숫자	두 개념 사이의 의미적 관련정도(이부분은 일단 신경쓰지 말기 바람)	
TypeOfInstance	숫자	인스턴스의 유형(이부분은 일단 신경쓰지 말기 바람)	
IsTerminal	숫자	단말노드인지체크(0:중간노드; 1:단말노드)	
WorkingDate	텍스트	최근 작업 일자	
NameOfWork	텍스트	최근 작업자 설명	

RelationshipType		ConID1		ConID2	
필드 크기	정수(Long)	필드 크기	정수(Long)	필드 크기	정수(Long)
형식		형식		형식	
소수 자릿수	자동	소수 자릿수	자동	소수 자릿수	자동
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값		기본값		기본값	
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	

Weight		TypeOfInstance		IsTerminal	
필드 크기	실수(Double)	필드 크기	정수(Long)	필드 크기	정수(Long)
형식		형식		형식	
소수 자릿수	자동	소수 자릿수	자동	소수 자릿수	자동
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값	1	기본값	0	기본값	0
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	

WorkingDate		NameOfWork		Bigo	
필드 크기	50	필드 크기	50	필드 크기	255
형식		형식		형식	
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값		기본값		기본값	
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	아니오	필수	아니오	필수	아니오
빈 문자열 허용	예	빈 문자열 허용	예	빈 문자열 허용	예
인덱스	아니오	인덱스	아니오	인덱스	아니오
유니코드 압축	아니오	유니코드 압축	아니오	유니코드 압축	아니오
IME 모드	한글	IME 모드	한글	IME 모드	한글
IME Sentence 모드	없음	IME Sentence 모드	없음	IME Sentence 모드	없음

<표 VII-4> rel-XXX 테이블 구조

rel-을 : 테이블

필드 이름	데이터 형식	설명
RelationshipType	숫자	(정상 관계):1000, (예외 관계): 2000
ConID1	숫자	자신의 아이디
ConID2	숫자	자신과 관계를 가지는 개념의 아이디
Weight	숫자	두 개념 사이의 의미적 관련정도(이부분은 일단 신경쓰지 말기 바람)
TypeOfInstance	숫자	인스턴스의 유형(이부분은 일단 신경쓰지 말기 바람)
IsTerminal	숫자	단말노드인지체크(0:중간노드; 1:단말노드)
WorkingDate	텍스트	최근 작업 일자
NameOfWork	텍스트	최근 작업자 성명

RelationshipType		ConID1		ConID2	
필드 크기	정수(Long)	필드 크기	정수(Long)	필드 크기	정수(Long)
형식		형식		형식	
소수 자릿수	자동	소수 자릿수	자동	소수 자릿수	자동
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값		기본값		기본값	
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	예	필수	예	필수	예

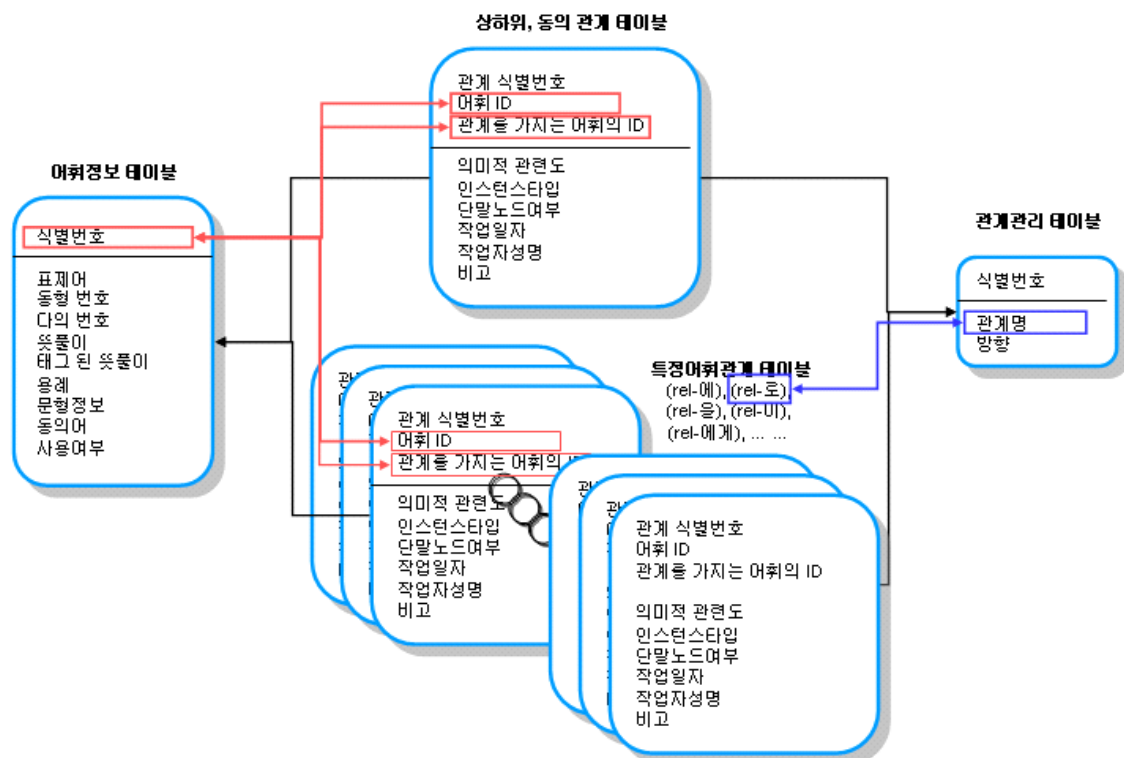
Weight		TypeOfInstance		IsTerminal	
필드 크기	실수(Double)	필드 크기	정수(Long)	필드 크기	정수(Long)
형식		형식		형식	
소수 자릿수	자동	소수 자릿수	자동	소수 자릿수	자동
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값	1	기본값	0	기본값	0
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	예	필수	예	필수	예

WorkingDate		NameOfWork		Bigo	
필드 크기	50	필드 크기	50	필드 크기	255
형식		형식		형식	
입력 마스크		입력 마스크		입력 마스크	
캡션		캡션		캡션	
기본값		기본값		기본값	
유효성 검사 규칙		유효성 검사 규칙		유효성 검사 규칙	
유효성 검사 텍스트		유효성 검사 텍스트		유효성 검사 텍스트	
필수	아니오	필수	아니오	필수	아니오
빈 문자열 허용	예	빈 문자열 허용	예	빈 문자열 허용	예
인덱스	아니오	인덱스	아니오	인덱스	아니오
유니코드 압축	아니오	유니코드 압축	아니오	유니코드 압축	아니오
IME 모드	한글	IME 모드	한글	IME 모드	한글
IME Sentence 모드	없음	IME Sentence 모드	없음	IME Sentence 모드	없음

<표 VII-5> RName 테이블 구조

RName : 테이블			
필드 이름		데이터 형식	설명
ID		Relation	Dual
필드 크기	정수(Long)	필드 크기 50	필드 크기 정수(Long)
형식		형식	형식
소수 자릿수	자동	입력 마스크	소수 자릿수 자동
입력 마스크		캡션	입력 마스크
캡션		기본값	캡션
기본값		유효성 검사 규칙	기본값 0
유효성 검사 규칙		유효성 검사 텍스트	유효성 검사 규칙
유효성 검사 텍스트		필수	유효성 검사 텍스트
필수	예	빈 문자열 허용	
		인덱스	예(중복 불가능)
		유니코드 압축	예
		IME 모드	한글
		IME Sentence 모드	없음

### 3.2. 필드 간의 관계



<그림 VII-22> 테이블과 필드간의 관계

상하위 테이블(Relation), 동의 관계 테이블(ERelation) 또는 특정어휘관계 테이블(rel-을, rel-이, rel-에서, rel-로, 등)의 어휘ID(ConID1), 관계를 가지는 어휘의 ID(ConID2)는 어휘정보 테이블(UWIN 테이블)의 식별번호(KeyID)를 기반으로 매핑된다.

관계관리 테이블(RName)의 관계명은 특정어휘관계 테이블의 테이블명과 일치하며, 관계관리 테이블(RName)의 개수만큼 특정어휘관계 테이블 개수가 만들어진다.

## 참고 문헌

- 권혁승(2003), “영어의 기초어휘 연구”, 『새국어생활』 제13권 3호, 국립국어원.
- 김광해(1993), 『국어어휘론개설』, 서울: 집문당.
- 김광해(2003), “기초어휘의 개념과 중요성”, 『새국어생활』 제13권 3호, 국립국어원.
- 김은영(2004), “국어 어휘의 계층적 의미관계에 대한 고찰”, 『한국언어문학』 40, 한국언어문학회.
- 김한샘(2003), “한국 현대 소설의 어휘 조사 연구”, 국립국어연구원.
- 서상규(1998), 『현대 한국어의 어휘 빈도』, 연세대학교 언어정보개발연구원 내부 보고서, 연세대학교 언어정보개발연구원.
- 서정국(1968), “국민학교 『국어』 1어휘의 통계분석적 연구”, 논문집.
- 유현경·남길임(2008), 『한국어 사전 편찬학 개론』, 도서출판 역락.
- 윤애선 외(2009), “한국어 어휘의미망 KorLex 1.5 구축”, 한국정보과학회 논문지:소프트웨어 및 응용, 2009년 1월.
- 이충우(1992), “국어교육용 어휘 연구”, 서울대학교 대학원 박사학위 논문.
- 임지룡(1991), “국어의 기초어휘에 대한 연구”, 『국어교육연구』 23집, 경북대학교 국어교육과.
- 이성현(2007), “사전 편찬에 있어서의 어휘의미망의 역할과 기능”, 『한국어 어휘의미망 구축과 사전편찬 학술회의』, 국립국어원.
- 최호섭·옥철영(2002), “한국어 의미망 구축과 활용”, 『한국어학』 17, 한국어학회
- 최호섭·옥철영(2004b), “UOU 온톨로지 구축 원리”, 『한국어 시소러스 연구』 (한유석·설근수), 한국문화사.
- 최호섭 외(2006a), “온톨로지 구축 방법과 사례”, 『한국정보과학회지』 제24권 제4호, 한국정보과학회.
- 최호섭 외(2006b), “대규모 우리말 어휘지능망 구축 방법”, 『제609돌 한글학회 전국국어학술대회 발표 자료집』.
- 최호섭(2007), “대규모 사용자 어휘지능망 구축과 활용”, 울산대학교 대학원 컴퓨터정보통신공학부 박사학위논문.
- 한정한·도원영(2005), “한국어 동사 의미망 구축을 위한 어휘의미관계 유형”, 『한국어학』 28, 한국어학회.
- 한정한 외(2004), 『유로워드넷』, 한국문화사.

Cruse,D.A.(1986), *Lexical Semantics*, Cambridge University Press.

Ogden,C.K.(1930), *Basic English: A General Introduction*, London:Kegan Paul.

West, M.P.(1953), *A General Service List of English Words*, London: Longman.

<자료>

김광해(2003 ㄱ), 『등급별 국어교육용 어휘』, 서울: 박이정.

김한샘(2005), “현대 국어 사용 빈도 조사2”, 국립국어연구원.

연세대학교 언어정보연구원 편(1998), 『연세한국어사전』, 두산동아.

옥철영(2008), 『한국어 어휘의미 체계 기반 입체적 기본어휘 사전 구축 연구』, 국립국어연구원.

조남호(2002), “현대 국어 사용 빈도 조사: 한국어 학습용 어휘선정을 위한 기초 조사”, 국립국어연구원.

21세기 세종계획 전자사전 개발 보고서(2001~2007), 국립국어원.

21세기 세종계획 홈페이지, [www.korean.go.kr/sejong](http://www.korean.go.kr/sejong)

KorLex 홈페이지, <http://korlex.cs.pusan.ac.kr/>

## 부록 A. 선정 어휘 관리 목록

1. 명 사 (64,272개)
2. 동 사 (15,735개)
3. 형용사 ( 4,736개)
4. 부 사 ( 2,656개)

모든 선정 어휘 목록은 CD로 제출

# 부록 A-1. 선정 어휘 관리 목록 (명사 샘플)

I	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
UWIN ID	선정연도	CD	사종의미	표제어	sense tan	뜻풀이(WEB)	8	종사/의미	타당 뜻풀이	한자원어	용례	용례타당	다어 사용	주어	간접	연체	U-WIN 계층	0	R	S
2545	1809300	2008	B		결남이	00.00.00	통제 결면의 넓이.	물체	..000001/NNG 결면/NNG+		무나 배우의 결대.	무	..020001/NNG+나/JK	7	1	1	4 X	1절도 2크기 3절이	결면적, 표	
2546	1810200	2009.게	A		결대	01.00.00	부상자의 결쪽에 붙은 줄기	부상자	..000001/NNG+의/JKB 결		사람들이 결발한 한	사람	..000002/NNG+					1인지 20음		속이음
2547	1812700	2009.반	E		결마음	00.00.00	결조로만 드러나는 진실되지	결	..010002/NNG+으로/JKB+		결것을 내다. 결악이	결것	..000001/NNG+	1	15	X	OK	17이로 29이로 3말		속말
2548	1812900	2009.반	A		결말	00.00.00	마음으로는 그렇지 않으면서	마음	..010002/NNG+으로/JKB+		결면들 코팅한 광고	결면	..000001/NNG+를/JKO 코팅	702	16	2	X	1모양 2장면 3단 4면	외면	
2549	1813200	2008	A		결면	00.00.00	결속 없이 결조로만 꾸리는	결속	..010001/NNG 없이		결면들 코팅한 광고	결면	..000001/NNG+를/JKO 코팅					1절도 2크기 3절이	결면이, 표	
2550	1813400	2008	B		결면적	00.00.00	결조로 드러나 보이는 면.	결	..010001/NNG+으로/JKB 드	面										
2551	1814400	2009.동	A		결면적	00.00.00	결남이.	결남이	..000001/NNG+/SF	面積										
2552	1815500	2008	B		결모습	00.00.00	결조로 드러나 보이는 모습.	결	..010001/NNG+으로/JKB 드		사람들 결모습만 보	사람	..000003/NNG+	23	55	10	X	1모양 2모양새 3장면 4모습 외용		
2553	1815600	2008	B	모양	결모양	00.00.00	결조로 보이는 모양.	결	..010001/NNG+으로/JKB 보	模樣	결모양은 평평한데	결모양	..000001/NNG+은/JKB 많	87	21	9	X	1모양 2모양새 3장면 4모습 외용		
2554	1818700	2008	A	모양	결보기	00.00.00	결조로 드러나 보이는 모양	결	..010002/NNG+으로/JKB 드		그는 결보기에는 돈	그	..010001/NP+는/JK	58	79	13	X	1모양 2모양새	외양	
2555	1819400	2009.동	A		결보기운	00.00.00	시운통.	시운통	..000001/NNG+/SF	運										
2556	1819900	2008	OK		결보리	00.00.01	결조를 벗기지 아니한 보리.	결조	..000001/NNG+를/JKO 벗		결보리를 키로 까들	결보리	..000001/NNG+	12	5	4	X	1생물 2식물 3종자 4두해살이 풀	시운통	
2557	1819900	2008	A		결보리	00.00.02	보리.	보리	..000001/NNG+를/JKO 벗											
2558	1820400	2008	B		결봉	00.00.01	봉의 결면.	봉투	..020000/NNG+의/JKB 결	封	편지 결봉에 주소를	편지	..020000/NNG 결	28	21	5	X	1봉위 2봉부 3봉지 4봉지 5봉지	결봉투, 피	
2559	1820400	2008	B		결봉	00.00.02	편지가 갖지 따위를 싸서 붙	편지	..020000/NNG+나/JKB 장	封	결봉을 뜯다. 외국서	결봉	..000002/NNG+	1	21	5	X	1물건 2물체 3주머니 4봉지 5	결봉투, 외	
2560	1820400	2008	B		결봉	00.00.03	편지를 봉투에 넣는 뒤에 다	편지	..020000/NNG+를/JKO 봉	封										
2561	1820500	2009.동	C		결봉투	00.00.01	결봉(1).	결봉	..000001/NNG+/SF	封套										
2562	1820500	2009.동	C		결봉투	00.00.02	결봉(2).	결봉	..000002/NNG+/SF	封套										
2563	1820500	2009.동	C		결봉투	00.00.03	결봉(3).	결봉	..000001/NNG+/SF	封套										
2564	1821000	2009.동	A		결배대	00.00.00	외골격.	외골격	..000001/NNG+/SF											
2565	1821200	2009.반	A		결상	01.00.00	옷에 싸이지 않고 늘 결에	옷	..010001/NNG+에/JKB 사이		자전거와 부딪쳤으나	자전거	..000001/NNG+와/JKB 부							속상
2566	1823100	2008	B		결싸식물	00.00.00	밀기가 싸방 안에 있지 않고	밀기	..000001/NNG+가/JKB 싸방	植物	겨울이라 두물한 결	겨울	..000001/NNG+이/JP+	31	53	8	3	OK	1생물 2식물 3고등식물 4종자	나지식물
2567	1824200	2008	B		결웃	00.00.00	결..010001/NNG+에/JKB 입다	결	..010001/NNG+에/JKB 입다		겨울이라 두물한 결	겨울	..000001/NNG+이/JP+					1물건 2의로 3웃	외의, 표의	
2568	1825900	2008	A		결장	02.00.01	여러 장으로 겹쳐 있는 종이	여러	..000001/NNG+의/JKB 겹	張										
2569	1825900	2008	A		결장	02.00.02	책의 표지.	책	..010002/NNG+의/JKB 표지	張	이 책의 초간본은 결	이	..050002/NM 책	16	5	1	X	1물건 2물체		속장
2570	1826100	2009.반	A		결저고리	00.00.01	저고리를 끼입을 때 맨 결에	저고리	..010001/NNG+를/JKO 끼입											속저고리
2571	1826200	2009.반	A		결작삼	00.00.00	저삼을 끼입을 때 맨 결에	저삼	..000001/NNG+를/JKO 끼입		무 결질이, 결질이	무	..020001/NNG 결질	7	6	2	X	1물건 2의로 3웃 4웃 5웃 5작삼	생질이, 엄	
2572	1826300	2008	B		결결삼	00.00.00	배수, 상수, 두 따위를 줄여	배수	..010000/NNG+/SF 상수		고구마 겉쪽을 갈로	고구마	..000002/NNG					1물건 2의로 3웃 4웃 5웃 5작삼	표면	
2573	1827500	2009.동	C		결쪽	00.00.00	표면(1).	표면	..000001/NNG+/SF		결치레 앞. 결치레	결치레	..000002/NNG 말	19	30	4	X	1공간 2구 3방향 4배쪽	표면	
2574	1828500	2008	B		결치레	00.00.00	같은 보기 공개 꾸미어 드러	결	..010002/NNG+만/JKB 보다		결치레 앞. 결치레	결치레	..000001/NNG 말					1물건 2의로 3웃 4웃 5웃 5작삼	표면	
2575	1828700	2009.반	A		결치마	00.00.01	치마를 끼입을 때 맨 결에	치마	..010001/NNG+를/JKO 끼											속치마
2576	1828700	2009.반	A		결치마	00.00.02	전통 혼례 때에, 신부가 치	전통	..060000/NNG 혼례											속치마
2577	1828800	2009	A		결치장	00.00.00	결조로 보기 공개 꾸임. 도	결	..010002/NNG+으로/JKB 보	治粧	이무리 결치장도 안	이무리	..000001/NNG 결치	1	4	X	X	1모양 2모양새 3꾸임새	속치마	
2578	1829100	2009.동	A		결켜	00.00.00	표용.	표용	..000001/NNG+/SF											속치마
2579	1830400	2008	OK		결계	01.00.00	심각함의 감각류를 통틀어	심각함	..000001/NNG+의/JKB 감각											속치마
2580	1830900	2009	A		결계	06.00.00	가타(伽陀).	가타	..000001/NNG+/SS+伽	偈										속치마
2581	1832300	2009	A		결거품	00.00.01	사람이나 동물이 몹시 괴롭	사람	..000001/NNG+이/나/JKB		그는 입에 거거품을	그	..010001/NP+는/JK	4	1	OK	OK	1물건 2의로 3웃 3웃 4구질 5물	표용	
2582	1837200	2009	OK		결근	00.00.00	남방의 생물체 또는 1개의	남방	..000001/NNG+의/JKB 생물체/NNG	?Benom										
2583	1839900	2008	A		결역지	00.00.01	계의 동역지.	계	..010000/NNG+의/JKB 동역											
2584	1843200	2009	A		결트마름	00.00.00	회백색의 금속 원소. 부스	회백색	..000001/NNG+의/JKB 금속	?Germaniu										
2585	1847200	2008	C	군대	결렬라	00.00.01	유격대(1).	유격대	..000001/NNG+/SF	?guerilla										유격대, 유

※ 선정연도 : 2009-계(선정어휘의 상위어가 미선정된 경우 추가)

2009\_반(선정어휘의 반의어가 미선정된 경우 추가)

※ CD:WEB : OK:E(선정어휘의 뜻풀이가 <웹용표준>과 다른 경우 <웹용표준>에 맞게 수정, 2009\_계/2009\_반/2009\_동은 수정하지 않았음)

부록 A-2. 선정 어휘 관리 목록 (동사 샘플)

UNIV ID	연도	신경	강도	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1																											
68	99109	2008	A	상태변화	가라앉다	00_02_05	떠들썩하던 것이 조용해	떠들썩하다	소란하던 장내 복위기	소란하다/VA+된/ETM	17	518	64	2	OK												
69	99110	2008	A	소멸	가라앉다	00_02_06	세차가 일어나던 것이 뜰 세차다_030001/VA+		적의 격렬한 공격이 가	적_130001/NGG+	4	518	64	2	OK												
70	99111	2008	A	악화	가라앉다	00_02_07	성하던 것이 활기를 잃어	성하다_030001/VA+	그녀가 적 가라앉은 후	그녀/NP+가/KS 착	4	518	64	2	OK												
71	99201	2008	OK	사동적행위	가라앉하다	00_01_00	'가라앉다[1](이)의 사용사: /SS+가라앉다		짜개기를 바닥에 가라	짜개기	78	153	18	2	OK	<...을>...에											
72	99203	2008	OK	사동적행위	가라앉하다	00_02_01	'가라앉다[2](이)의 사용사: /SS+가라앉다		홍분을 가라앉히고 내	홍분_000001/NGG+	49	153	18	2	OK	<...을>											
73	99204	2008	OK	사동적행위	가라앉하다	00_02_02	'가라앉다[2](이)의 사용사: /SS+가라앉다		사회자는 근거점을 몇	사회자/NGG+는/IX 큰	3	153	18	2	OK	<...을>											
74	99205	2008	OK	사동적행위	가라앉하다	00_02_03	'가라앉다[2](이)의 사용사: /SS+가라앉다		생산품 중에서 불량품	생산품/NGG 중	76	149	22	2	OK	<...을>											
75	122701	2008	A	결과행위	가라앉다	00_00_01	여러 가운데서 일정한 것	여럿/NGG 가운데	용의자들 중에서 진범	용의자/NGG+를/IX 중	19	149	22	2	OK	<...에서> ...											
76	122702	2008	A	인지적행위	가라앉다	00_00_02	진리나 가치 따위를 밝혀	진리_000003/NGG+	외국인 중에는 동양인	외국인_000001/NGG	10	9	4	2	X	<...을>											
77	123301	2008	OK	사각적행위	가라앉다	00_00_01	모양 따위를 구별하여 알	모양_020001/NGG 따	하나라를 배우면 얻을 일	하나라_000100/NR+	1	9	4	2	X	<...을>											
78	123302	2009	OK	인지적행위	가라앉다	00_00_02	진리나 가치 따위를 판단	진리_000001/NGG+			3	7	4	2	OK	<...을>...에											
79	130200	2009	A	배치	가라앉는다	00_00_00	가라앉려	놓다			1	9	4	2	X	<...을>											
80	130301	2008	A	피동적행위	가라앉는다	00_00_01	'가라앉는다'의 피동사: /SS+가라앉는다/VV+		겨울장은 마루 한중간	겨울장/NNP+은/IX 따	19	55	3	2	OK	<...에>											
81	130302	2008	A	피동적행위	가라앉는다	00_00_02	일의 진행이나 해결을 어	일_010002/NGG+	우리의 앞길에는 적지	우리_030001/NP+	2	55	3	2	OK	<...에>											
82	130701	2009	A	결과행위	가라앉다	00_00_01	가라앉고 또는 옆쪽으로 놓	가라_010000/NGG+	침대에 가라앉혀 잠을	침대_020000/NGG+	1	5	3	2	X	<...에>											
83	130702	2008	OK	몸짓의동작	가라앉는다	00_00_02	바닥에 기대앉게 놓다	또 바닥_010001/NGG+	길 위에 가라앉은 사람	길_010001/NGG 위	5	5	3	2	X	<...에>											
84	130900	2008	A	결과행위	가라앉다	00_00_00	'말하다[1]를 예스럽게 이	/SS+말하다	적장이 들을 보고 대장	적장_060000/NGG+	5	14	2	2	OK												
85	133501	2008	A	상태	가라앉다	00_00_01	알을 가라앉혀	말_000002/NGG+	강물을 가로막아 담을	강물/NGG+를/VO 가	35	348	41	2	OK	<...을>											
86	133502	2008	A	제약	가라앉다	00_00_02	말이나 행동, 일 따위를	말_010001/NGG+이	그가 그녀의 말을 가로	그_010001/NP+	14	348	41	2	OK	<...을>											
87	133503	2008	A	상태	가라앉다	00_00_03	알이 보이지 않도록 가리	알_000002/NGG+	집은 안개가 그의 시야	집_020104/VA+	4	348	41	2	OK	<...을>											
88	133701	2009	C	상태	가라앉다	00_00_01	'가라앉다(이)'의 피동사: /SS+가라앉다		그는 사람들에게 가로	그_010001/NP+는/IX	1	19	2	4	OK	<...에게>											
89	133703	2009	C	상태	가라앉다	00_00_03	'가라앉다(이)'의 피동사: /SS+가라앉다		일뒤 산들에 시야가 가	일뒤_000001/NGG 산	3	19	2	4	OK	<...에게>											
90	134001	2009	A	결과행위	가라앉다	00_00_01	남의 할 일을 가로채서 말	남_010001/NGG+	왜 네가 가로챘아서 전	왜_020000/MAG 네	4	3	1	2	X	OK	<...을>										
91	767002	2008	A	과거행위	감아먹다	00_00_02	소중한 사물이나 시간 따	소중하다/VA+~ /ETM	평은 그의 여성이 어엿	평_010001/NGG+	1	32	6	3	OK	<...을>											
92	767303	2009	B	반복적행위	감자거리다	00_00_01	날카롭고 뾰족한 끝으로	날카롭다_000001/VA+	강아지가 안으로 들어	강아지	4	2	2	2	X	X	<...을>										
93	767304	2009	B	반복적행위	감자거리다	00_00_02	되는데로 들이나 그림 따	되다_010202/VV+는다																			
94	771611	2008	C	감각행위	감각하다	02_00_01	감각[02(이)]	감각	너무 추워서 네 손이 얼	너무_010000/MAG 얼	7	10	1	4	OK	<...을>											
95	778700	2009	A	현상	감겨들다	00_00_00	가까이로 바짝 감기어 오	가까이_000200/NGG+	차가운 겨울 바람이 불	차가운_000001/VA+	0	7	4	2	X	<...에>											
96	778807	2008	C	외향적심리향감적하다	감격하다	00_00_01	감격(이)	감격	상감해서도 이 가혹하	상감_020000/NGG+	6	89	11	3	OK	<...에>											
97	778808	2008	C	외향적심리향감적하다	감격하다	00_00_02	감격(이)	감격	그는 사람들이 자신들	그_010001/NP+는/IX	5	89	11	3	OK	<...에>											
98	787902	2009	A	피동적행위	감금되다	00_00_00	감금	감금/NGG+ /SF	감옥에 감금되다	감옥_020001/NGG+	0	28	3	2	X	OK	<...에>										
99	1E+06	2008	C	결과행위	개방하다	03_00_02	개방[04(이)]	개방	외국에 문호를 개방하	외국_020000/NGG+	17	162	52	3	OK	<...에게>											
100	1E+06	2009	A	피동적행위	개방화되다	00_00_00	개방화	개방화/NGG+ /SF	사회가 개방화되다	사회_070003/NGG+	0	5	1	2	X												
101	1E+06	2009	C	과거행위	개방하다	00_00_01	개방[01(이)]	개방	전지가 개방하다	전지_010001/NGG+	4	7	2	4	OK												
102	1E+06	2008	C	결과행위	개방되다	00_00_01	개방[02(이)]	개방	담인지 봉투가 개봉되	담인지/NGG 봉투	0	78	4	2	X	OK											
103	1E+06	2008	C	결과행위	개방하다	02_00_01	개방[02(이)]	개방	관치를 개봉하다	서류 관치_020000/NGG+	1	105	4	2	X	OK	<...을>										
104	1E+06	2008	C	가시행위	개방하다	02_00_02	개방[02(이)]	개방	그 영화는 여제 개봉했	그_010102/MM 영화	4	105	4	2	X	OK	<...을>										
105	1E+06	2008	A	피동적행위	개방되다	01_00_00	개방[01]	개방	무역 수지가 개선되다	무역_020002/NGG 수	6	233	59	2	X	OK	<...을>										

※ 선정 연도 : 2009-1(부사와 결합하는 동사가 미선정된 경우 추가)

부록 A-3. 선정 어휘 관리 목록 (형용사 샘플)

UWIN ID	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	CD: 선정 WEB	번	번	번	표제어	sense tag	뜻풀이(WEB)	뜻풀이	태그	태그	태그	다의 비도	비도	비도	비도	비도	비도	결합부사 비도	이	S	T	어	어	어
200	990602	2008	OKE	높은 정도	강대하다	01_00_00	강대05.	강	국적이 강대하다. 세력 강력한 군대. 강력한 대 강력하다	국력_020000/NNG+	18	17	X	4	OK					힘, 사회 조직, 계획, 광채, 연상, 힘				
201	1000803	2008	OKE	높은 정도	강해하다	00_00_01	강해02.	강	강렬한 빛. 강렬한 천이 강렬하다	강렬_020101/VA+	124	1049	129	3	OK					조치, 계획, 광채, 연상, 힘				
202	1002002	2008	A	속성값	강성하다	02_00_00	강하고 세차다.	강	강렬한 빛. 강렬한 천이 강렬하다	강렬_020101/VA+	64	447	57	4	OK					조치, 계획, 광채, 연상, 힘				
203	1028502	2008	B	상태	강성하다	00_00_00	힘이 강하고 번성하다.	강	강렬한 빛. 강렬한 천이 강렬하다	강렬_020101/VA+	6	8	X	4	OK					조치, 계획, 광채, 연상, 힘				
204	1056602	2008	A	신체상태	강인하다	03_00_00	억제하고 질기다.	억제하다_000002/VA+	강인한 인상. 강인한 체 강인하다	강인_010102/MM 애	9	101	13	2	OK					연상, 힘, 정 성, 질				
205	1078802	2008	B	인성속성	강직하다	02_00_00	마음이 조곤하고 곧다.	마음_010002/NNG+	강직한 성품. 그는 정열 강직하다	강직_010102/MM 애	32	27	2	4	OK					연상, 힘, 정 성, 질				
206	1090602	2009	B	부정적인	강파르다	00_00_02	성질이 까다롭고 괴팍	성질_000001/NNG+	그에는 성미가 강파르 강파르하다	성질_020101/VA+	2	2	X	X	X					연상, 힘, 정 성, 질				
207	1091602	2008	A	높은 정도	강박하다	00_00_00	성격이 까다롭고 고집	성격_020001/NNG+	서성구는 평소에도 아 강박하다	서성_020101/VA+	67	5	363	1	1					연상, 힘, 정 성, 질				
208	957704	2008	A	속성값	강하다	02_01_01	물리적인 힘이 세다.	물리적	강한 손짐. 아내의 아이 강하다	강_020102/VA+	1051	5	363	1	1					연상, 힘, 정 성, 질				
209	957705	2008	A	속성값	강하다	02_01_02	수준이나 정도가 높다.	수준_000001/NNG+	강한 군대. 강한 물만. 강하다	강_020102/VA+	1	115	15	4	OK					연상, 힘, 정 성, 질				
210	6611304	2008	A	속성값	나쁜하다	00_00_02	힘이 없어 보드럽다.	힘_010001/NNG+	나쁜 게 나쁘다. 머리가 나쁘다	나쁜_010000/NNG+	453	1957	256	1	OK					표현, 사람, 동작, 가치				
211	6676002	2008	A	평가속성	나쁘다	01_01_01	출지 아니하다.	출지_010101/VA+	나쁜 짓. 거짓말하는 것 나쁘다	나쁜_010102/VA+	142	1957	256	1	OK					표현, 사람, 동작, 가치				
212	6676003	2008	A	평가속성	나쁘다	01_01_02	출지 아니하다.	출지_010001/VA+	나쁜 짓. 거짓말하는 것 나쁘다	나쁜_010102/VA+	142	1957	256	1	OK					표현, 사람, 동작, 가치				
213	6676004	2008	A	평가속성	나쁘다	01_02_00	간간 따위에 해롭다.	간간_030000/NNG 따	나쁜 게 나쁘다. 머리가 나쁘다	나쁜_000003/NNG+	17	1957	256	1	OK					왕성, 부위 등반				
214	6676006	2008	A	평가속성	나쁘다	01_03_01	어떤 일을 하기에 시기	어떤_000003/MM 일	나쁜 게 나쁘다. 머리가 나쁘다	나쁜_010000/NNG+	2	1957	256	1	OK					왕성, 부위 등반				
215	6676007	2008	A	높은 정도	나쁘다	01_03_02	어떤 일을 하기에 싫지	어떤_000003/MM 일	나쁜 게 나쁘다. 머리가 나쁘다	나쁜_010000/NNG+	2	1957	256	1	OK					왕성, 부위 등반				
216	6698403	2008	A	부정적인	나약하다	00_00_01	나약.	나약/NNG+/SF	나약한 마음. 왕은 쫌 나약하다	나약_050001/MM 집	14	114	15	4	OK					마음				
217	6698404	2008	A	속성값	나약하다	00_00_02	몸이 가늘고 약하다.	몸_010001/NNG+	나약한 마음. 왕은 쫌 나약하다	나약_000001/NNG+	5	114	15	4	OK					몸, 생김새				
218	6698601	2009	A	속성값	나자막하	00_00_01	나이가 어리다.	나이_010000/NNG+	나이에 턴어나오는 아 출지/NNG+에/JK8 턴어	나이/NNG+에/JK8 턴어	0	6	X	X	X					사람				
219	6720207	2008	A	낮은 정도	나자막하	00_00_01	워치가 꽤 나자막하.	워치_010001/NNG+	우리 동네에는 나자막 우리 동네에는 나자막	우리_030003/NP 등	21	102	14	3	OK					건축, 땅				
220	8336602	2009	A	높은 정도	다단하다	00_00_00	일이 갈레나 가닥이 많	일_010005/NNG+	우리 동네에는 나자막 우리 동네에는 나자막	우리_030003/NP 등	1	2	X	X	X					과제				
221	8342600	2008	A	상할값	다되다	00_00_00	완전히 그르진 상태에	완전히/MAG 그르진	다된 집안. 소 선생도 다디단 사탕. 시장해서	다되다/VA+~/_ETM. 집 다디단	10	13	X	X	X					사회				
222	8348401	2009	A	맞속성값	다되다	00_00_01	매우 많다.	매우_010000/MAG 많	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	1	14	X	X	X					음식				
223	8354801	2009	A	상태	다들다	00_00_01	때나 찌꺼기 따위가 있	때_020001/NNG+	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	1	1	X	X	X					의료, 부위				
224	8354802	2009	A	부정적인	다들다	00_00_02	연행이 순수하지 못하	연행/NNG+o/JKS 순	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	16	1	X	X	OK					연행				
225	8363701	2008	A	관계속성	다르다	01_00_01	비교가 되는 두 대상이	비교_010001/NNG+	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	1921	6301	1132	1	OK					부위, 모습				
226	8363702	2008	A	평가속성	다르다	01_00_02	비교의 것보다 두드러	비교_000100/NNG+	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	8	6301	1132	1	OK					승비, 감각				
227	8368800	2008	A	관계속성	다들다	00_00_00	견주어 보아 겹거나 비	견주어/VV+o/JEC 보다	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	19	453	59	3	OK					각기, 당연				
228	14E+07	2008	A	상태	무고하다	04_00_00	아무런 잘못이나 허를	아무런/MM 잘못	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	13	48	8	4	2					사람				
229	14E+07	2008	A	부정적인	무관심하	00_00_00	무관심	무관심/NNG+/SF	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	12	149	15	X	OK					사람, 시선, 활동				
230	14E+07	2008	B	속성값	무관하다	03_00_01	관계나 상관이 없다.	관계_050001/NNG+	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	27	590	53	4	OK					활동, 시선, 기색, 사람				
231	14E+07	2008	OKE	높은 정도	무구하다	00_00_01	무구06*1.	무	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	4	12	1	X	OK					행성, 코				
232	14E+07	2008	A	속성값	무공무진	00_00_00	무공무진.	무공무진/NNG+/SF	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	2	52	7	4	OK					기색, 사람				
233	14E+07	2008	A	상태	무공하다	00_00_00	무공.	무공/NNG+/SF	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	8	65	16	4	OK					기색, 사람				
234	14E+07	2009	A	속성값	무근하다	00_00_01	뿌리가 없다.	뿌리_000001/NNG+	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	1	3	X	X	X					식물				
235	14E+07	2009	B	속성값	무근하다	00_00_02	근거가 없다.	근거_000001/NNG+	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	2	3	X	X	X					활동, 시선, 기색, 사람				
236	14E+07	2008	A	신체상태	무기력하	00_00_00	무기력.	무기력/NNG+/SF	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	9	126	10	4	OK					기색, 사람				
237	14E+07	2008	A	속성값	무난하다	00_00_01	별로 어려움이 없다.	별로_010001/MAG 어	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	4	92	10	4	OK					기색, 사람				
238	15E+07	2008	A	속성값	반반하다	01_00_01	구김살이나 울물불를	구김살	다디단 사탕. 시장해서 다디단	다되다/VA+~/_ETM. 집 다디단	35	31	3	X	OK					기색, 사람				

※ 선정 연도 : 2009-1(부사와 결합하는 형용사가 미선정된 경우 추가)

부록 A-4. 선정 어휘 관리 목록 (부사 샘플)

UWINID	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	선정 년도	CDW EB	세종의미부류		표제어	sense_tag	뜻풀이일부	용사/의미 태깅 뜻풀이	한자원어	용례	용례태깅	다의 빈도	세종 빈도	국립 2002	연세 사전	결합부사	결합용언	
1																		
239	5337102	2008	OK E	시간부사/순서	금방	01_00_02	방금[01][1]2.	방금	今方	금방 비가 올 것처럼 하	금방_010002/MAG 비	11	1066	142	1	2	가속	오다,후회하다,저다,까먹다
240	5361600	2008	OK E	시간부사/순서	금세	01_00_00	지금 바로, '금세'에 가 붙	지금 바로 금_003001/NNNG 바		소문이 금세 퍼졌다. 약	소문_020000/NNNG+	2	345	56	X	2		퍼지다,나타나다,달아오르
241	5450704	2008	OK E	양태부사/방식	급격히	00_00_00	급격하다01.	급격하다_001000/VA+		사공이 급격히 감소하	급_0420001/NNNG+	60	282	34	3	OK	감소하다,진행되다,떨어지	
242	5454400	2008	OK E	급이하	급이하	00_00_00	마지막에 가서는.	마지막/NNNG+예/KB 가	及其也	급기야 일이 내가 강당	급기야/MAG 일	0	207	31	X	OK	감속하다	
243	5469004	2008	OK E	양태부사/방식	급속히	00_00_00	급속하다.	급속스럽다/VA+		급속히 변화하다. 양국	급속히/MAG 변화하	57	210	33	3	OK	변화하다,증가하다,잡히다	
244	5479102	2008	OK E	양태부사/방식	급작스레	00_00_00	급작스럽다.	급작스럽다/VA+		급속한 자가 돌단로도	급속하다/VV+된/ETM	6	13	X	X	OK	멈추다,서다,퍼붓다,당하다	
245	5448812	2008	OK E	양태부사/방식	급히	00_00_01	급하다1.	급하다_000001/VA+		급히 꼭 할 일이 있다.	급히_000001/MAG 꼭	59	383	40	X	OK	하다,필요하다,쓰다	
246	5448813	2008	OK E	양태부사/방식	급히	00_00_02	급하다2.	급하다_000002/VA+		급히 뛰어간다. 그렇게	급히_000002/MAG 뛰	87	383	40	X	OK	뛰어간다,뛰다	
247	5448814	2008	OK E	양태부사/방식	급히	00_00_03	급하다6.	급하다_000006/VA+		명함 해결은 바다 속이	명함/NNP 해함/NNNG+	7	383	40	X	OK	경사지다	
248	5448815	2008	OK E	양태부사/방식	급히	00_00_04	급하다7.	급하다_000007/VA+		강물이 금히 흐르다.	강물/NNNG+이/KB 급히	48	383	40	X	OK	흐르다	
249	5570300	2008	OK E	문장부사/양상부	기꺼이	00_00_00	마음속으로 은근히 기쁘	마음속/NNNG+으로/KB		기꺼이 받아들이다. 기	기꺼이/MAG 받아들이	17	189	26	3	OK	함께	
250	5570600	2008	OK E	양태부사/방식	기껏	00_00_00	힘이나 정도가 미치지	힘_001001/NNNG+이		기껏 서지를 해 놓았다	기껏/MAG 서지	14	167	21	3	OK	하다,놓다,위르하다	
251	5570701	2008	OK E	기껏해야	기껏해야	00_00_01	아무리 한다고 해야.	아무리_000001/MAG		내가 하는 일이 기껏해	네_010000/NP+가/KB	7	141	23	3	OK	만큼다	
252	5756802	2008	OK E	기실	기실	03_01_00	실제에 있어서.	실제_002001/NNNG+	其實	기실 알고 보면 그 사람	기실_030100/MAG 알	1	74	6	X	OK	알다,보다,단순하다,명백하	
253	5769001	2008	OK E	문장부사/양상부	기어이	00_00_01	기어코	기어코	期於	저들의 속셈이 무엇인지	기_040002/NP+	9	202	28	2	OK	밝히다,말다,가다,싫다	
254	5769002	2008	OK E	문장부사/양상부	기어이	00_00_02	기어코	기어코	期於	하늘이 잔뜩 흐리더니	하늘_010001/NNNG+	1	202	28	2	OK	오다,나가다,저지르다,말다,	
255	5769601	2008	OK E	문장부사/양상부	기어코	00_00_01	어떠한 일이 있더라도	어떠하다/VA+/_/ETM	期於	기어코 해내고야 말겠다	기어코_000001/MAG	5	124	9	3	OK	해내다,말다,값다,부리다,하	
256	5769602	2008	OK E	문장부사/양상부	기어코	00_00_02	결국에 가서는 는기어이	결국_000001/NNNG+	期於	어머님이 그렇게 말했는	어머님_000001/NNNG+	1	124	9	3	OK	부리치다,떠나다,타프르다,	
257	5794402	2008	OK E	양태부사/방식	기왕	00_02_00	기왕에	기왕에/MAG+/SF	既往	기왕 말이 나왔으니까	기왕_000200/MAG 말	19	79	7	X	2	나오다,하다,놓다	
258	5795000	2008	OK E	시간부사/양상부	기왕에	00_00_00	이미 그렇게 된 바에 는	이미_001000/MAG 그	既往	기왕에 왔으니까 만나나	기왕에/MAG 오다	0	60	8	X	X	오다,시작하다,붓다,되다	
259	5795100	2008	OK E	기왕이면	기왕이면	00_00_00	아차피 그렇게 된 바에	아차피/MAG 그렇다	既往	기왕이면 죽수 대신에	기왕이면/MAG 죽수	0	22	1	X	OK	먹다,하다	
260	5799200	2008	OK E	양태부사/상징	기우뚱	00_00_00	물체가 한쪽으로 약간	물체_000001/NNNG+		심한 파도로 배가 한쪽	심하다/VA+/_/ETM 파	19	15	1	X	OK	기울다,쏠리다	
261	5805800	2008	OK E	기운	기운	00_00_00	무엇을 보려고 고개나	무엇_000002/NNP+		찬식이가 부엌을 기운	찬식/NNP+이/XSN+	4	10	X	3	OK	틀어다보다,하다	
262	5805901	2008	OK E	기운기운	기운기운	00_00_01	남의 것을 탐내는 마음	남_001001/NNNG+		미군들도 이쪽을 호기심	미군/NNNG+를/XSN+	3	7	2	X	OK	넘겨다보다,틀어다보다	
263	5805902	2008	OK E	기운기운	기운기운	00_00_02	무엇을 보려고 고개나	무엇_000002/NP+				2	7	2	X	OK		
264	5893900	2008	OK E	양태부사/양상부	기릴코	00_00_00	반드시.	반드시/MAG+/SF	期必	이번 계약은 기릴코 성	이번_010000/NNNG 계	1	47	9	4	OK		틀어오다, 내다
265	5926304	2008	OK E	양태부사/양상부	기릴코	00_00_00	긴급하다.	긴급하다/VA+		정보를 받고 긴급하고	전보_040000/NNNG+	3	23	3	4	OK		내려가다,저리하다,소집하
266	5941204	2008	OK E	양태부사/방식	긴급히	00_00_00	긴급하다.	긴급하다/VA+		관계를 긴밀히 하다. 양	관계_050001/NNNG+	8	45	1	4	OK		하다,협조하다,취하다
267	5923307	2008	OK E	양태부사/방식	긴밀히	00_00_01	긴하다1.	긴하다_000001/VA+		긴히 볼 데가 있어서 그	긴히_000001/MAG 쏘	0	17	3	X	OK	쏘다	
268	5967401	2008	OK E	양태부사/방식	길길이	00_00_01	성이 나서 팔팔 뛰는 모	성_001000/NNNG+		길길이 늘었다. 길길이	길길이_000001/MAG	0	44	2	4	OK		늘뛰다,지르다,퍼붓다
269	5987000	2008	OK E	시간부사/지속상	길이	02_00_00	오랜 세월이 지나도록.	오랜/MM 세월		길이 보전하다. 이들을	길이_020000/MAG 보	18	16	5	4	2		보전하다,빛내다,남다
270	5987100	2008	OK E	시간부사/지속상	길이길이	00_00_00	아주 오래도록.	아주_001001/MAG 오		길이길이 간직하다. 길	길이길이/MAG 간직하	4	13	1	4	OK		간직하다,이어지다,전하다,
271	5993400	2008	OK E	시간부사/지속상	길이길이	00_00_00	어쩔이 다 조금 긴 모양.	여럿/NNNG+이/KB 다		떡을 길쭉길쭉 썰다. 마	떡_010001/NNNG+	4	4	X	X	X		썰다,시원스럽다
272	6073802	2008	OK E	양태부사/방식	길쭉히	00_00_00	길쭉하다1.	길쭉하다		산 길쭉이 들어간다. 모	산_010001/NNNG 길쭉	55	289	37	3	OK		틀어가다,내리다,쓰다,묻다,
273	6074801	2008	OK E	양태부사/방식	깊이	02_00_01	수준이 높게. 또는 정도	수준_000001/NNNG+		양속 깊이 묻다. 구멍이	양속/NNNG 깊이	171	1124	138	2	2		묻다,파다,감추다,몰려쓰다
274	6074802	2008	OK E	양태부사/방식	깊이	02_00_02	결에서 속까지의 거리가	결_001001/NNNG+예		깊이 생각하다.	깊이_020002/MAG 생	67	1124	138	2	2		생각하다

※ 선정연도 : 2009-1(부사-부사, 부사-용언과 결합하는 부사가 미선정된 경우 추가)

## 부록 B. 품사/동형이의어 동시 태깅 시스템 사용설명서

### 제 1 장 소프트웨어 소개

#### 1.1 소프트웨어 명칭 및 필요성

##### 1.1.1 소프트웨어 명칭

“세종 말뭉치 검증 및 원시 말뭉치 품사/동형이의어 동시 태깅 시스템”

##### 1.1.2 필요성

- 세종형태의미말뭉치는 21세기 세종계획으로 만들어진 형태 및 동형이의어 분석 말뭉치로 현재 339개의 세종형태의미말뭉치가 제공되고 있으며, 한국어정보처리의 다양한 분야에서 이를 활용하고 있음.
- 그러나, 이 말뭉치는 술어명사에 대해 단일한 형태의 품사 분석 결과 (예, “독점한”의 경우 독점/NNG+하/XSV+ㄴ/ETM)만 제공하고 있어서 술명사를 용언으로 분석하고자 할 경우 이를 재가공해야 하는 번거로움이 있음.
- 또한, 다수의 품사 분석 오류가 포함되어 있으며, 명사와 일부 주요 용언에 대해서만 동형이의어 분석되어 있음. 예를 들어 “새/NNB”, “새/MM”은 각각 “새\_\_05/NNB”, “새\_\_06/MM”으로 태깅되어야 함.
- 따라서, 세종말뭉치의 활용도를 높이기 위해서는 기존 세종말뭉치의 태깅결과를 검증하고 동형이의어 태깅되지 않은 어휘를 태깅할 수 있는 검증용 태깅 시스템이 필요하며, 더불어 아직 태깅되지 않은 원시말뭉치를 품사/동형이의어 태깅할 수 있는 시스템이 필요함.

##### 1.1.3 기대 효과

- 품사/동형이의어 동시 태깅 시스템은 기 제공된 세종형태의미말뭉치를 학습하여 만들어져 세종말뭉치 내의 오류 분석 어절의 결과를 포함하고 있음.

- 앞으로 이 도구를 이용하여 전체 세종말뭉치가 모두 정확하게 태깅된 후 이를 다시 학습하면 보다 정확한 결과를 제시할 수 있을 것으로 판단됨.

## 1.2 소프트웨어 사용 환경

- Windows 95/98/NT/XP/VISTA/7

## 1.3 소프트웨어 소개 내용

- 세종말뭉치와 울산대 한국어 처리 연구실 태깅 결과(품사 태그 SET 동일) 비교
- 문장 단위의 태깅 비교 결과 다른 문장에 대해
  - 다르게 분석된 어절에 대해 비교/수정
  - 직접 입력 가능
  - 수정 결과 로그 파일 기록
- 형태의미태깅 세종말뭉치는 문장 단위로 재구성하였으며 술어명사에 대해서 다음과 같이 두 가지 Version을 제공한다.  
예를 들어 “가늠하다”의 경우
  - 가늠/NNG+하/XSV+다/EF
  - 가늠/VV+다/EF    의 두 가지 형식의 태깅 결과 비교 가능.
- 문장 단위로 재구성한 세종말뭉치는
  - 어근(XR)은 NNG로 태깅되어 있음,
  - 본용언+보조용언/본용언 ⇒ 복합용언 (표준국어대사전에 등재된 경우)
  - 접두사+명사, 명사+접미사 ⇒ 명사 (표준국어대사전에 등재된 경우)
  - 관형사+명사, 명사+명사 ⇒ 명사 (표준국어대사전에 등재된 경우)로 통합하였다.
  - 1,900원, 1.5%과 같이 숫자 내에 “, .” 등의 기호가 있는 경우는 1,900/SN+원/NNB 혹은 1.5/SN+% /SW로 전처리하였다.

- 동형이의어 태깅 결과는 울산대 어휘지능망(U-WIN)과 연동되어 <표준국어대사전>을 통해 확인하지 않더라도 직관적으로 동형이의어 태깅 결과를 확인할 수 있도록 하였으며, 용언의 경우는 세종 <사태> 의미 부류로 매핑하였음.

## 제 2 장 소프트웨어 설치

### 2.1 소프트웨어의 구성

- 실행 파일
  - WinHlx\_1.exe : 실행에 필요한 library 정적 link version
  - WinHlx\_2.exe : 실행에 필요한 library 동적 link version
- 실행 환경 설정 파일
  - Hlxcfg.txt (3.1.1. 참조)
- 사전 파일
  - 분석 유형에 따른 기본 사전 및 태깅용 학습 DB
  - DIC\_세종\_통합(폴더), DIC\_세종\_분리(폴더)로 각각 분리되어 있음.  
(3.1.1의 실행환경 설정에 따라 접근하는 사전 폴더가 다름)
  - 기본사전 : CDIC.txt(체언), VDIC.txt(용언), Joeo.txt(조사/어미 사전)
  - 태깅용 학습 DB
    - ① ndic.dic ndic.idx NDIC.txt (고유명사 사전 및 DB)
    - ② biAA.dic biAA.idx biAA.txt (bigram 학습 DB)
    - ③ biAD.idx biAD.txt (앞 어절과 뒤 어절의 첫 번째 형태소 학습 DB)
    - ④ biLD.idx biLD.txt (앞 어절의 마지막 형태소와 뒤 어절의 첫 번째 형태소 학습 DB)
    - ⑤ unigram.dic unigram.idx unigram.txt (unigram의 학습 DB)
    - ⑥ high.idx high.txt (동형이의어 상위개념어 매핑을 위한 DB)
    - ⑦ patScore.idx patScore.txt (어절의 분석 품사 패턴 학습 DB)
- 문장단위로 재구성한 형태의미 태깅말뭉치 폴더 (임의의 위치)
  - 의미분석\_어휘분리\_sent : 술어명사 “명사+하” 형태로 분리

- 의미분석\_어휘통합\_sent : 술어명사 하나의 용언으로 통합
- 원시말뭉치 : 테스트용 원시말뭉치

## 2.2 소프트웨어 실행 폴더 구조

- 임의의 폴더(예를 들어 \Tagger)를 생성한 후, \Tagger 폴더에 압축파일을 푼다. 압축파일을 풀면,
  - DIC\_세종\_통합 폴더 (통합으로 분석할 때 필요한 사전 파일)
  - DIC\_세종\_분리 폴더 (분리로 분석할 때 필요한 사전 파일)
  - WinHlx 폴더 (두 가지 종류의 실행 파일)
  - Hlxcfg.txt 파일이 생성
- WinHlx 폴더로 이동
- WinHlx-1.exe 실행 파일을 실행 폴더

## 제 3 장 소프트웨어 사용법

### 3.1 사용 준비

#### 3.1.1 실행 환경 설정 (config 파일)

- 술어명사는 다음 두 가지로 분석할 수 있음. (예, “독점한”)
  - 독점/NNG+하/XSV+ㄴ/ETM
  - 독점하/VV+ㄴ/ETM
- 위 두 가지 분석 방법은 Hlxcfg.txt을 수정하여 설정

```

;-----
;어휘사전
;-----
; 세종 통합 태깅 (가능하/VV)
;HLX_DIR                ..\DIC_세종_통합\
;WRD_DIR                ..\DIC_세종_통합\
;TAG_DIR                ..\DIC_세종_통합\
;JOEO_INF               ..\DIC_세종_통합\Joeo.txt
;
; 세종 분리 태깅 (가능/NNG+하/XSV)
HLX_DIR                ..\DIC_세종_분리\
WRD_DIR                ..\DIC_세종_분리\
TAG_DIR                ..\DIC_세종_분리\
JOEO_INF               ..\DIC_세종_통합\Joeo.txt
;

```

- 현재 config 파일은 술어명사를 “명사+하”와 같이 분석할 경우로 설정되었음.
- 세종말뭉치를 불러 올 때 해당 폴더(의미분석\_어휘분리\_sent 폴더)의 파일을 불러와야 함.

○ 한자를 한글로 변환하여 분석할 경우는 Hlxcfg.txt을 수정하여 설정

```

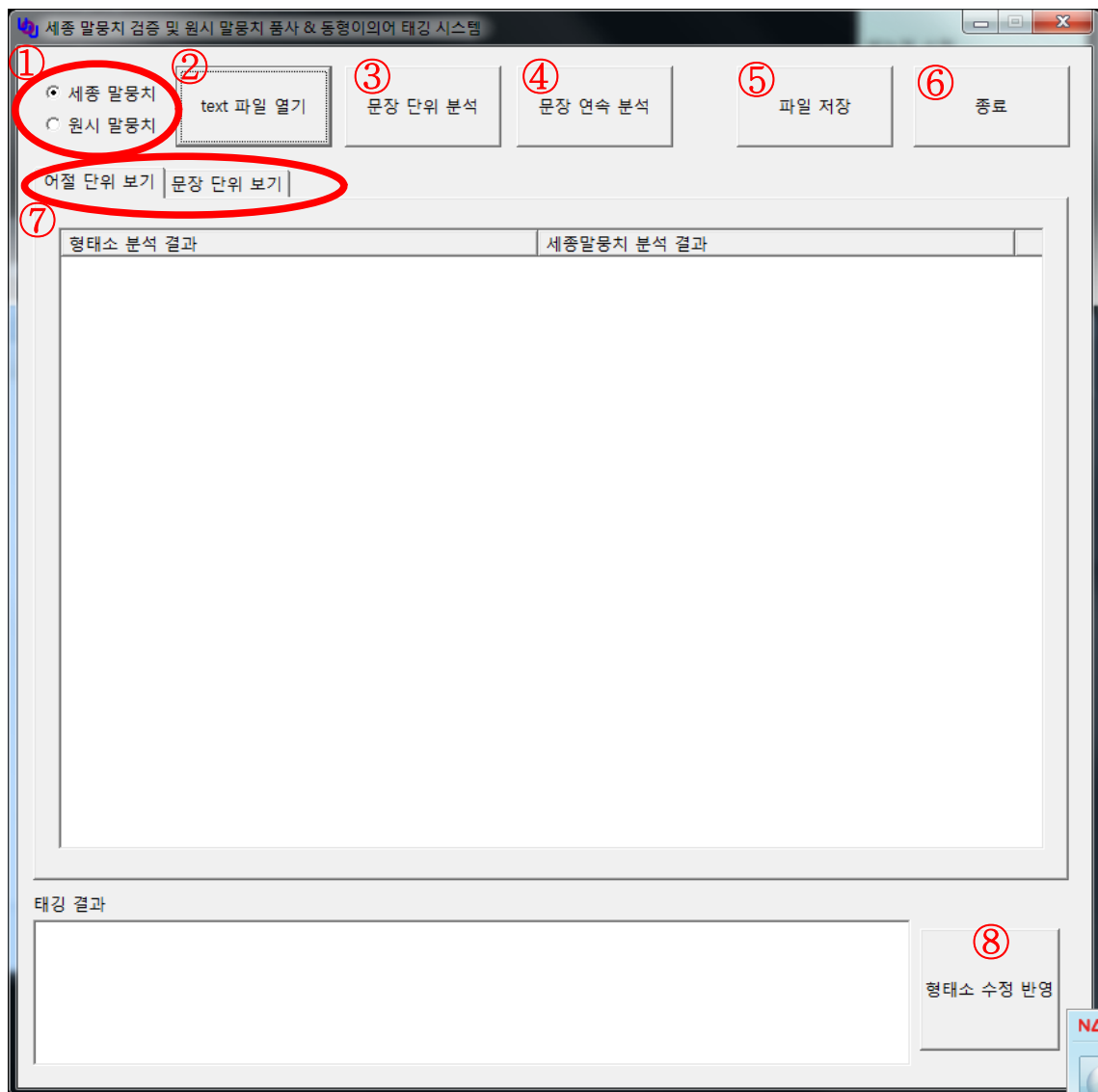
;=====
; 분석 조건
;=====
;
; (1) 한자어 처리(한자 변환 여부 1: 변환, 0:변환안함)
HANJA_TRN              0
;-----

```

## 3.2 세종말뭉치 검증

### 3.2.1 초기 실행 화면

○ 실행 파일 (WinHlx\_1.exe 혹은 WinHlx\_2.exe) 실행합니다.

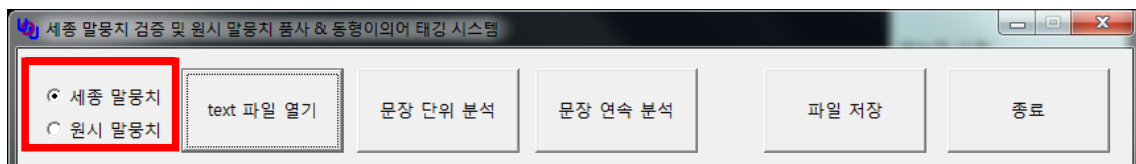


- ① 분석 말뚝지를 선택합니다. 기본 선택은 세종 말뚝치입니다. 파일을 열기 전에 선택합니다.
- ② 파일을 열어서 화면에 출력하는 버튼입니다. 분석말뚝치에 따라 text파일을 선택합니다.
- ③ 파일을 읽어온 후 한 문장 단위로 차례로 분석하는 버튼입니다.
- ④ 세종 말뚝치 검증 시 분석 결과가 다를 때까지 문장을 연속으로 분석하는 버튼입니다.
- ⑤ 파일을 저장합니다.
- ⑥ 프로그램을 종료합니다.
- ⑦ 문장 단위 또는 어절 단위로 보는 것을 선택합니다. 어절 단위는 형태

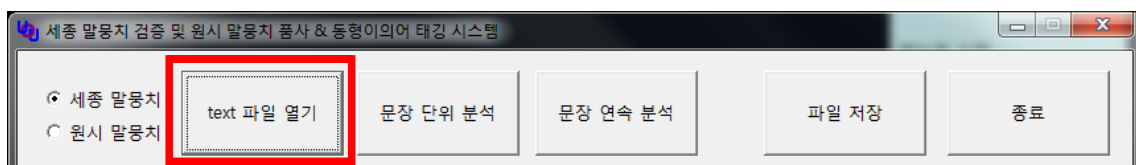
소 분석 결과입니다.

- ⑧ 틀린 태깅 결과의 경우 수정하고 나서 결과를 반영합니다. 수정했을 때만 반영합니다.

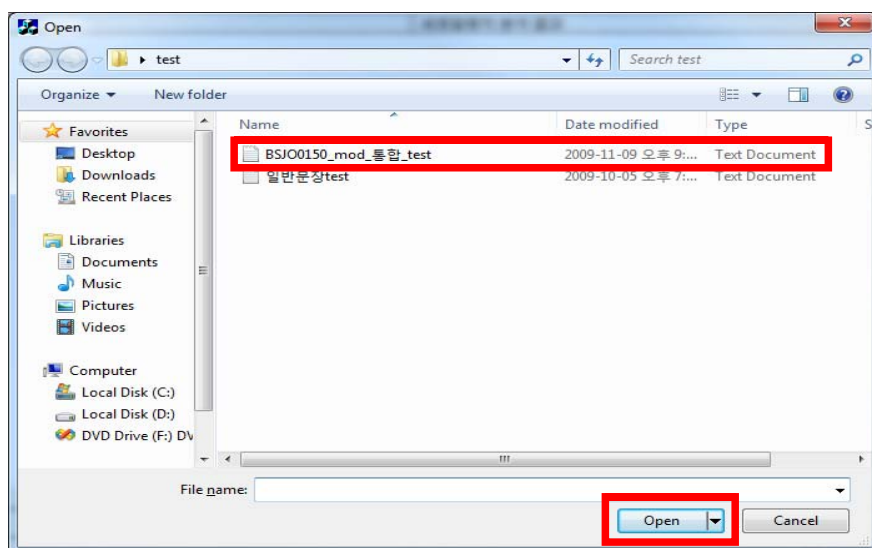
### 3.2.2 세종 말뭉치 검증



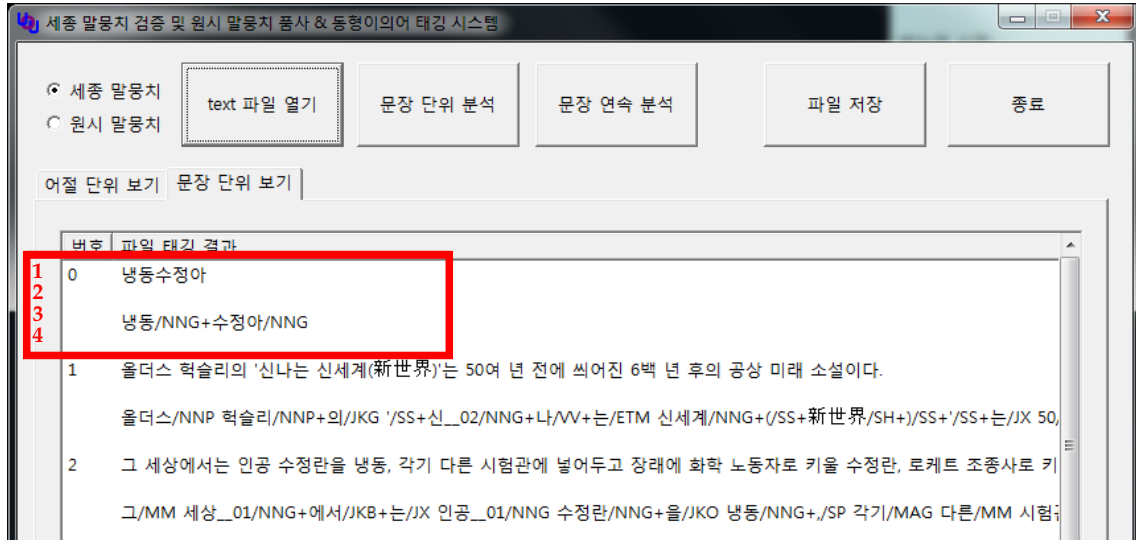
- ① 파일을 열기 전에 말뭉치 종류를 선택합니다. 기본은 <세종 말뭉치>입니다.
- ② <세종 말뭉치>를 선택하고, 세종 말뭉치 파일을 열어서 검증을 할 수 있습니다.



- ③ <text 파일 열기> 버튼을 클릭합니다. 다음과 같은 창이 나타납니다.

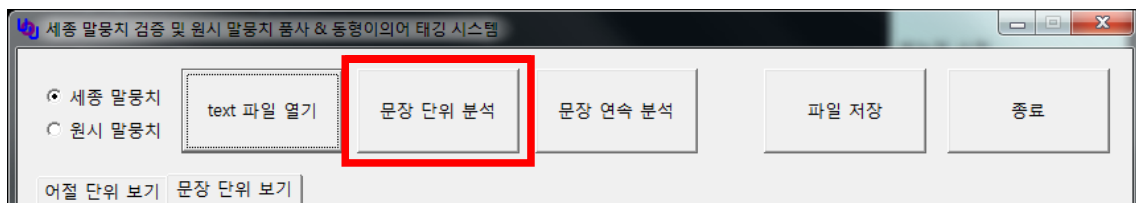


- ④ 여기서 열고자 하는 파일을 더블 클릭 또는 선택하고 <열기>를 클릭합니다.

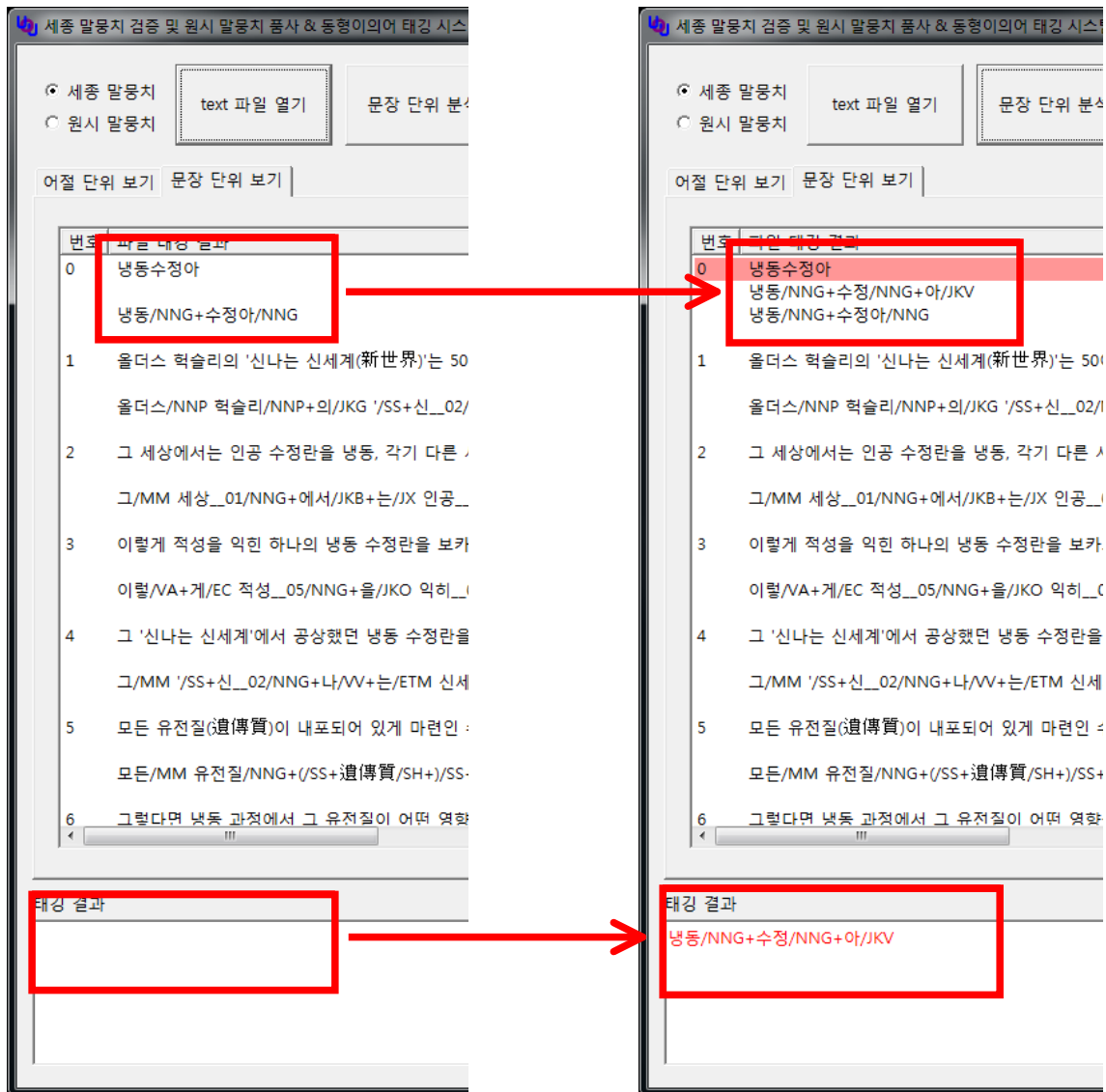


- ⑤ 세종 말뚝치 파일을 열었을 경우 화면입니다.  
화면에 나타난 순서의 의미는 다음과 같습니다.
- 1 냉동수정아 - 일반 문장입니다.
  - 2 빈칸 - 울산대 태깅 시스템이 분석한 결과 위치
  - 3 냉동/NNG+수정아/NNG -세종 분석 결과
  - 4 빈칸 - 문장을 구분하기 위한 빈칸입니다.

### 3.2.3 분석 및 비교



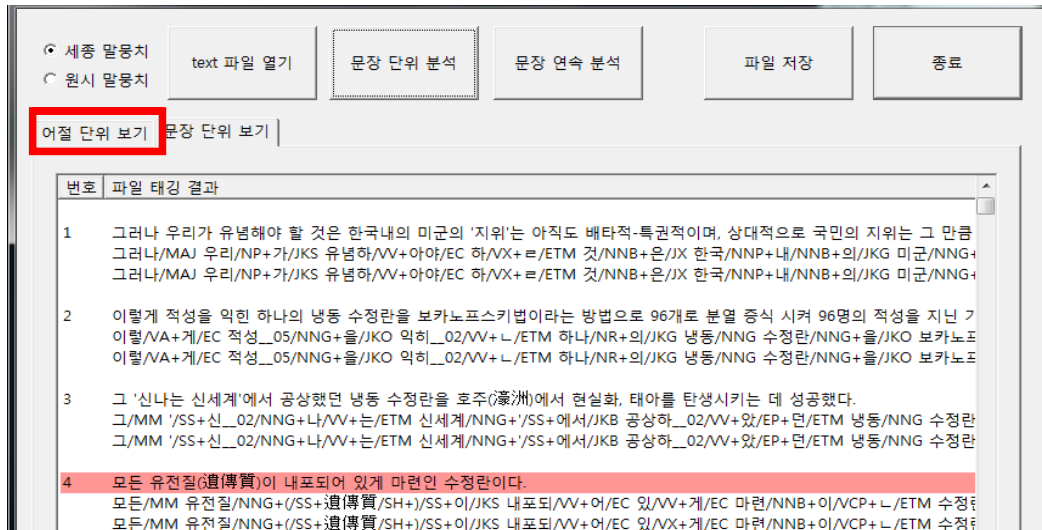
- ① 파일을 열었으면, 다음으로 <문장 단위 분석> 버튼을 클릭합니다.  
- <문장 단위 분석> 버튼은 시작 문장부터 차례대로 한 문장 단위로 분석, 비교를 합니다.



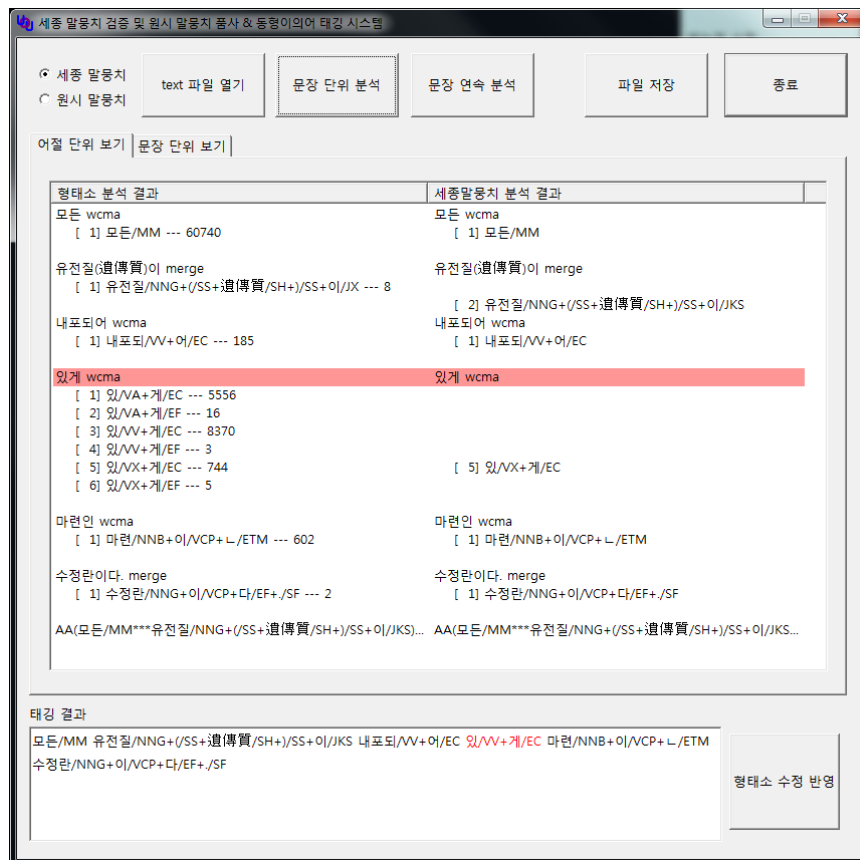
- 가운데 빈칸에 태깅 결과(냉동/NNG+수정/NNG+아/JKV)가 출력된 것을 볼 수 있습니다.
- ② 이 태깅 결과와 아래의 세종 태깅 결과를 비교하여 다르면 위 그림과 같이 붉은색이 표시되어 두 태깅 결과가 서로 다르다는 것을 확인할 수 있습니다. 아래의 [태깅 결과]에도 태깅 결과가 출력되었습니다. 만일 두 태깅 결과가 서로 같다면 이와 같이 초록색이 표시됩니다.

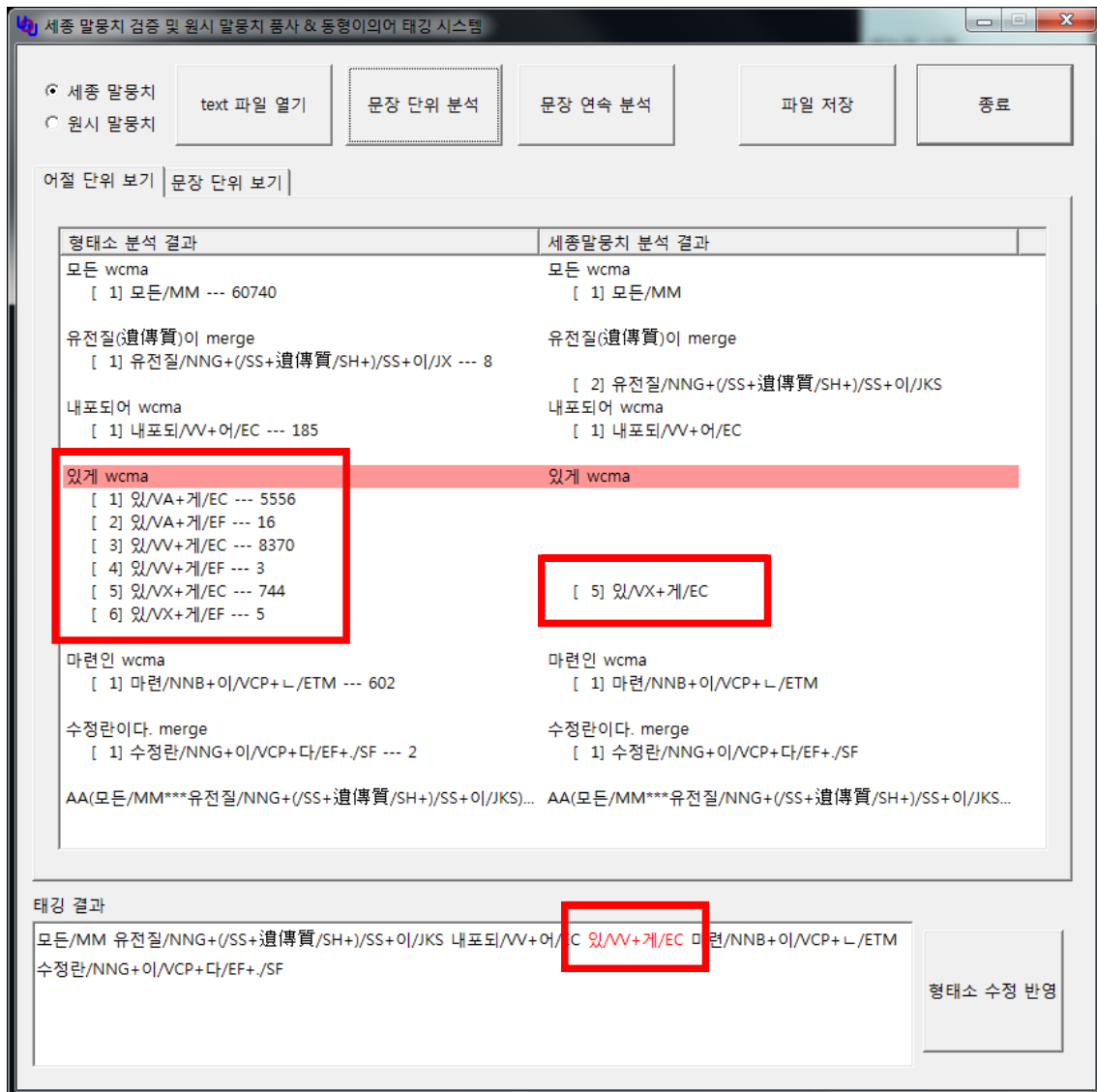
<p>출력 결과</p> <p>출력 결과</p>	<p>출력 결과</p> <p>출력 결과</p>
---------------------------	---------------------------

### 3.2.4 태깅 결과가 다른 경우 검증 및 수정



- ① 두 결과가 틀렸다고 붉은색으로 표시가 되면, <어절 단위 보기> 탭을 클릭하면, 결과가 다른 어절에 한해 붉은색이 표시되어 확인할 수 있습니다.





- ② “있게 wcma”에 대하여, 왼쪽 [ 1]~[ 6]은 울산대 태깅 시스템에서 분석한 형태소들입니다.
- 오른쪽 “[ 5] 있/VX+게/EC”는 세종의 분석입니다. 울산대의 분석 결과와 같은 줄에 위치합니다.
  - 아래쪽의 태깅 결과를 보면 “있/VV+게/EC”가 선택되어져 다르다는 것을 알 수 있습니다.
- ③ 여기서 “[ 5]”의 분석으로 수정하려면 그 줄을 마우스로 <더블클릭>합니다.

있게 wcma	있게 wcma
[ 1] 있/VA+계/EC --- 5556	
[ 2] 있/VA+계/EF --- 16	
[ 3] 있/VV+계/EC --- 8370	
[ 4] 있/VV+계/EF --- 3	
[ 5] 있/VX+계/EC --- 744	[ 5] 있/VX+계/EC
[ 6] 있/VX+계/EF --- 5	
마련인 wcma	마련인 wcma
[ 1] 마련/NNB+이/VCP+L/ETM --- 602	[ 1] 마련/NNB+이/VCP+L/ETM
수정란이다. merge	수정란이다. merge
[ 1] 수정란/NNG+이/VCP+다/EF+./SF --- 2	[ 1] 수정란/NNG+이/VCP+다/EF+./SF
AA(모든/MM***유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS...	AA(모든/MM***유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS...

태깅 결과
모든/MM 유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS 내포되/VV+어 EC <b>있/VX+계/EC</b> 마련/NNB+이/VCP+L/ETM 수정란/NNG+이/VCP+다/EF+./SF

형태소 수정 반영

- 더블 클릭 후 화면입니다.
- 태깅결과 창에서 “있/VV+계/EC”가 “있/VX+계/EC”로 수정된 것을 확인할 수 있습니다.

④ 만약 세종 분석과 울산대 분석 내용 중에 알맞은 형태소 분석 결과가 없다면, 형태소 분석 중에 바꾸고 싶은 분석 내용에서 <마우스 오른쪽 클릭>을 합니다.

있게 wcma

[ 1] 있/VA+계/EC --- 5556

[ 2] 있/VA+계/EF --- 16

[ 3] **있/VV+계/EC --- 8370**

[ 4] 있/VV+계/EF --- 3

[ 5] 있/VX+계/EC --- 744

[ 6] 있/VX+계/EF --- 5

→

있게 wcma

[ 1] 있/VA+계/EC --- 5556

[ 2] 있/VA+계/EF --- 16

[ 3] **있/VX+계/EC --- 8370**

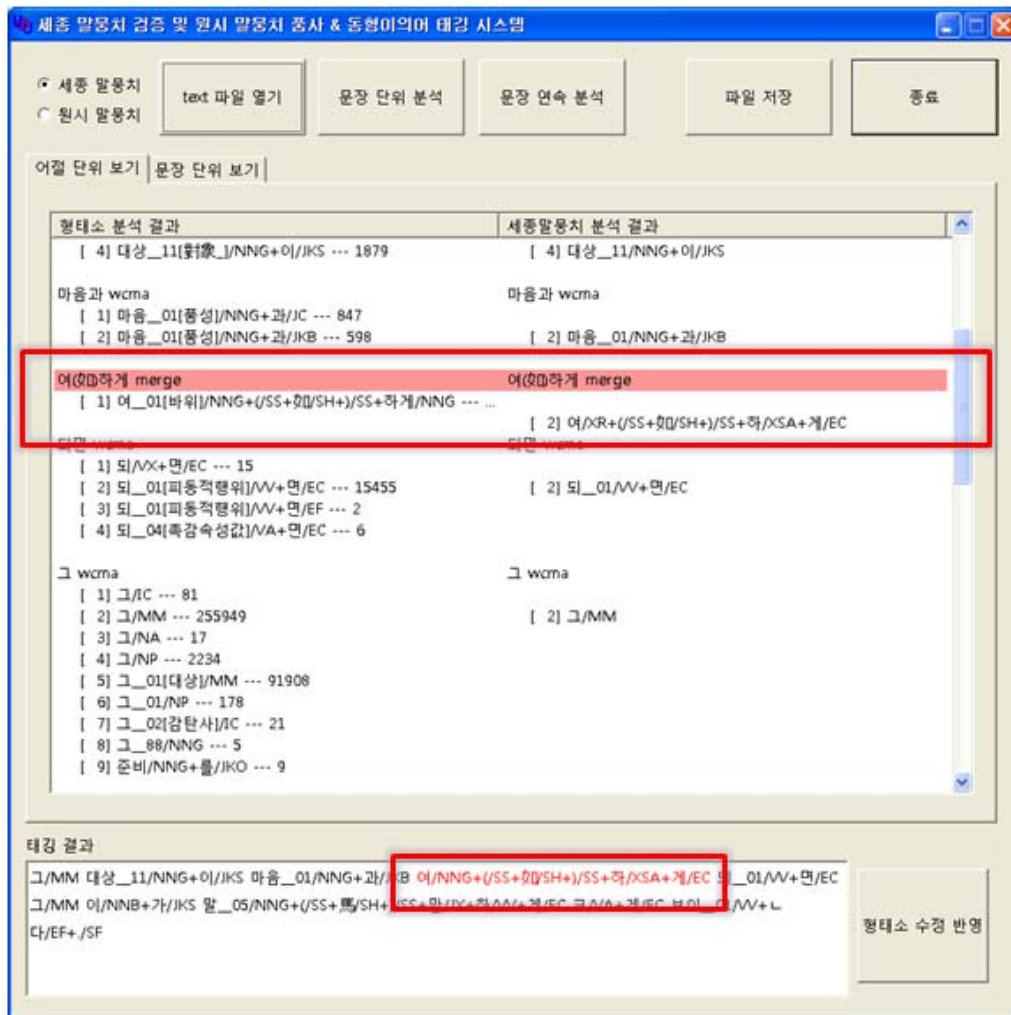
[ 4] 있/VV+계/EF --- 3

[ 5] 있/VX+계/EC --- 744

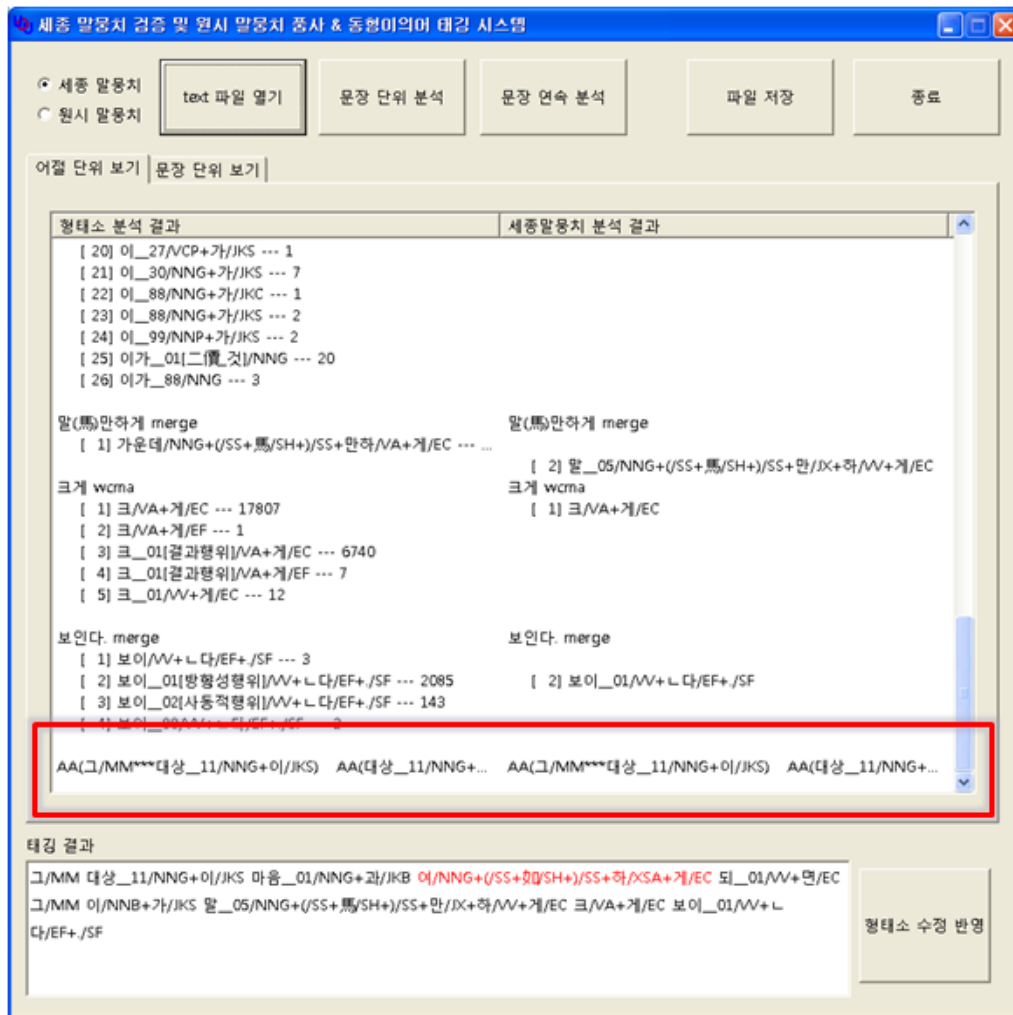
[ 6] 있/VX+계/EF --- 5

- 마우스 오른쪽 클릭 후 화면이며, 직접 내용을 수정할 수 있습니다.
- 수정 후 [Enter]키를 누르면 내용이 바뀌고, 더블클릭을 하면 결과를 바꿀 수 있습니다.

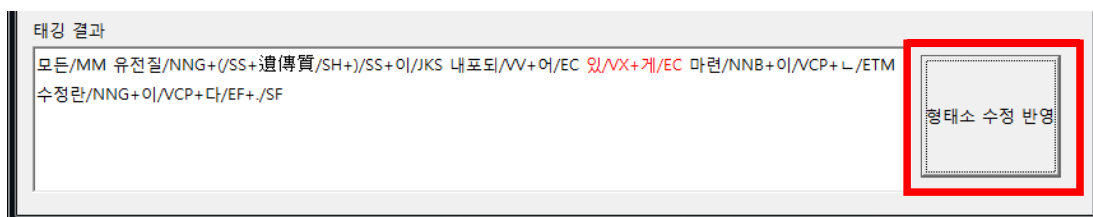
⑤ [태깅 결과]의 어절 분석 결과가 울산대 형태소 분석 결과와 세종 분석 결과에도 없는 경우



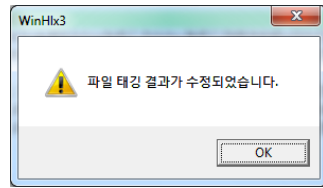
- 위 경우와 같이 [태깅결과] 창에 좌측 “형태소 분석 결과”와 “세종 말뚝치 분석 결과”에도 없는 다른 분석 결과가 나오는 경우는 울산 대 태깅 시스템이 내부 규칙에 따라 결정한 경우입니다. 이는 주로 어절이 한자나 특수 문자를 포함한 경우입니다(merge로 표시).
- 위 경우는 아래 그림의 창의 맨 마지막 줄(수직 스크롤 바를 움직임)을 보면, 해당 어절의 앞뒤 두 어절의 분석결정이 "AA/AD/LD" 이면 [태깅 결과] 창의 것이 맞는 것으로 결정합니다.
- 그렇지 않은 경우 사용자가 판단하여 수정하십시오.



- ⑥ 수정이 완료되었으면, [태깅 결과] 옆에 있는 <형태소 수정 반영> 버튼을 클릭합니다.



- 결과가 정상적으로 수정되면 다음과 같은 창이 나타납니다.



OK버튼을 클릭합니다.

4 모든 유전질(遺傳質)이 내포되어 있게 마련인 수정란이다.

모든/MM 유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS 내포되/VV+어/EC 있/VX+게/EC 마련/NNB+이/VCP+L/ETM 수정란  
모든/MM 유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS 내포되/VV+어/EC 있/VX+게/EC 마련/NNB+이/VCP+L/ETM 수정란

- 수정 후 비교 결과가 같으면 초록색으로 표시되는 것을 확인할 수 있습니다.

⑥ 만약 수정을 해야 하는 문장을 지나쳤다면 그 문장에 마우스를 <더블 클릭>합니다.

1 그러나 우리가 유념해야 할 것은 한국내의 미군의 '지위'는 아직도 배타적-특권적이며, 상대적으로 국민의 지위는 그 만큼  
그러나/MAJ 우리/NP+가/JKS 유념하/VV+아야/EC 하/VX+ㄹ/ETM 것/NNB+은/JX 한국/NNP+내/NNB+의/JKG 미군/NNG+  
그러나/MAJ 우리/NP+가/JKS 유념하/VV+아야/EC 하/VX+ㄹ/ETM 것/NNB+은/JX 한국/NNP+내/NNB+의/JKG 미군/NNG+

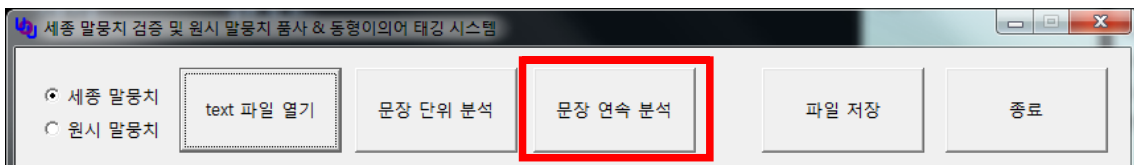
2 이렇게 적성을 익힌 하나의 냉동 수정란을 보카노프스키법이라는 방법으로 96개로 분열 증식시켜 96명의 적성을 지닌 가  
이혈/VA+게/EC 적성\_05/NNG+을/JKO 익히\_\_02/VV+L/ETM 하나/NR+의/JKG 냉동/NNG 수정란/NNG+을/JKO 보카노프  
이혈/VA+게/EC 적성\_05/NNG+을/JKO 익히\_\_02/VV+L/ETM 하나/NR+의/JKG 냉동/NNG 수정란/NNG+을/JKO 보카노프

3 그 '신나는 신세계'에서 공상했던 냉동 수정란을 호주(濠洲)에서 현실화, 태아를 탄생시키는 데 성공했다.  
그/MM 'SS+신\_02/NNG+나/VV+는/ETM 신세계/NNG+'SS+에서/JKB 공상하\_\_02/VV+았/EP+던/ETM 냉동/NNG 수정란  
그/MM 'SS+신\_02/NNG+나/VV+는/ETM 신세계/NNG+'SS+에서/JKB 공상하\_\_02/VV+았/EP+던/ETM 냉동/NNG 수정란

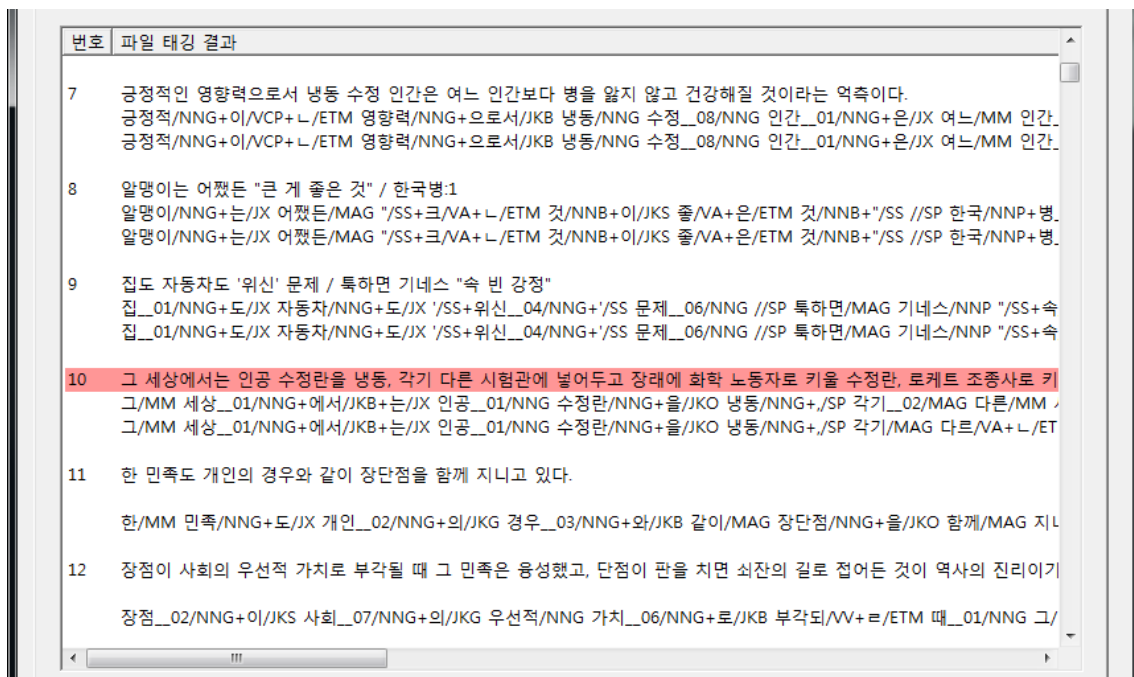
4 모든 유전질(遺傳質)이 내포되어 있게 마련인 수정란이다.  
모든/MM 유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS 내포되/VV+어/EC 있/VX+게/EC 마련/NNB+이/VCP+L/ETM 수정란  
모든/MM 유전질/NNG+(/SS+遺傳質/SH+)/SS+이/JKS 내포되/VV+어/EC 있/VX+게/EC 마련/NNB+이/VCP+L/ETM 수정란

- 그 문장에 대하여 형태소 분석 보기, 수정을 할 수 있습니다.  
- 수정을 다 하면 다시 <문장 단위 분석>을 클릭하면 하던 검증을 이어 할 수 있습니다.

### 3.2.5 문장 연속 분석

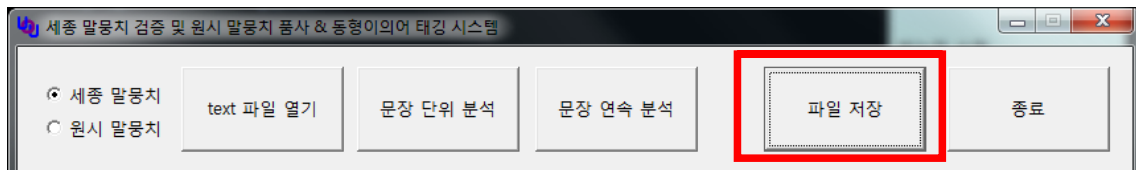


- <문장 연속 분석> 버튼은 다른 분석 결과가 나올 때까지 분석하는 기능입니다.
- 세종말뭉치를 검증할 때 태깅 결과를 비교하여 같다면 넘어가고 결과가 서로 다르다면 그곳에 멈추는 버튼입니다.
- 현재 문장 번호가 5번인데 여기서 문장 연속 분석 버튼을 클릭해 보겠습니다.

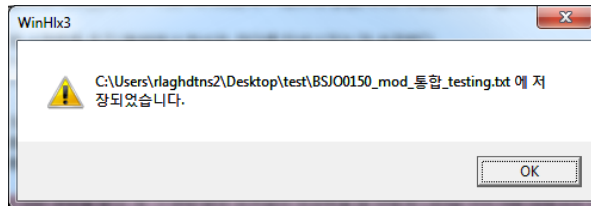


- 버튼을 클릭한 후 문장 번호가 10번에 위치하는 것을 확인할 수 있습니다.
- 6번부터 9번까지는 비교 결과가 같기 때문에 자동으로 넘어간 것입니다.

### 3.2.6 저장



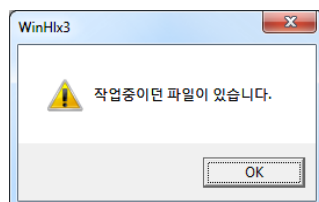
- ① 저장하려면 <파일 저장> 버튼을 클릭합니다.



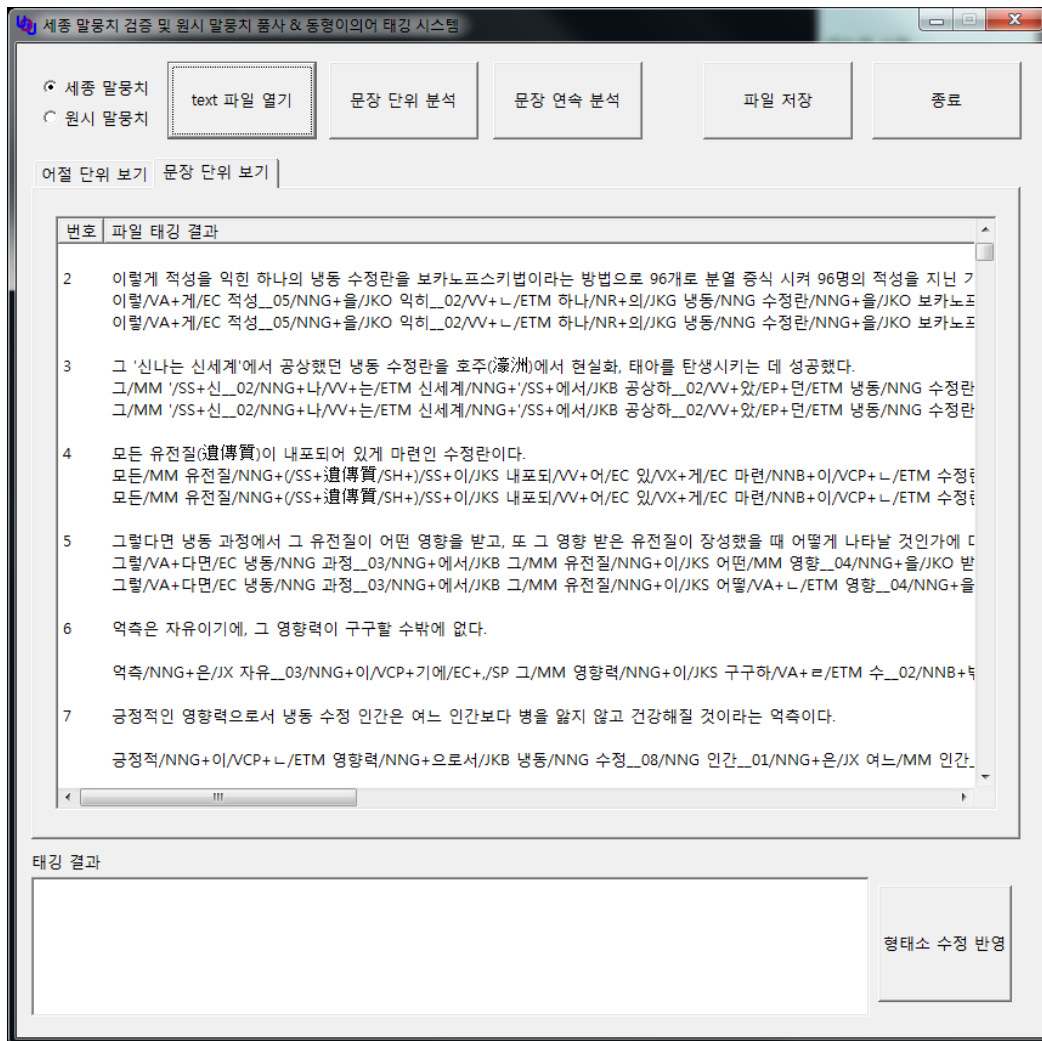
OK버튼을 클릭합니다.

- ② 작업 도중에 저장을 하게 되면 “파일이름ing.txt”에 저장합니다.  
작업을 마치고 저장을 하면 “파일이름ing.txt”는 지워지고, “파일이름  
end.txt”에 저장합니다.  
수정된 형태소의 경우에는 “파일이름edit.txt”에 문장 번호와 수정된 어  
절 결과가 저장합니다.

- ③ 작업 도중에 저장을 하고 다시 그 파일을 불러와 보겠습니다.

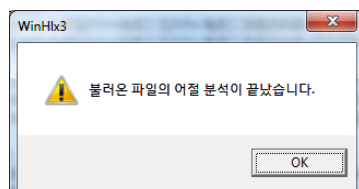


OK버튼을 클릭합니다.



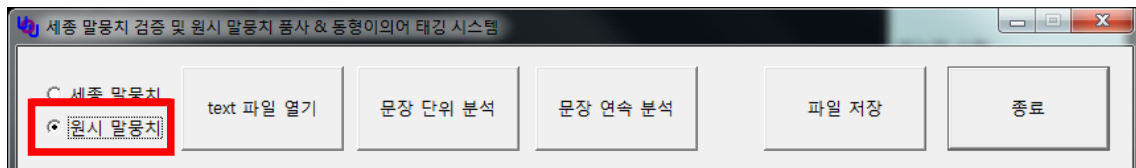
- 위와 같이 작업한 곳까지 저장되어져서 불러왔습니다.
- <문장 단위 분석> 버튼을 클릭하면 계속 이어서 작업을 할 수 있습니다.

④ 파일을 끝까지 분석한 경우에는 아래와 같이 메시지 창이 뜹니다.

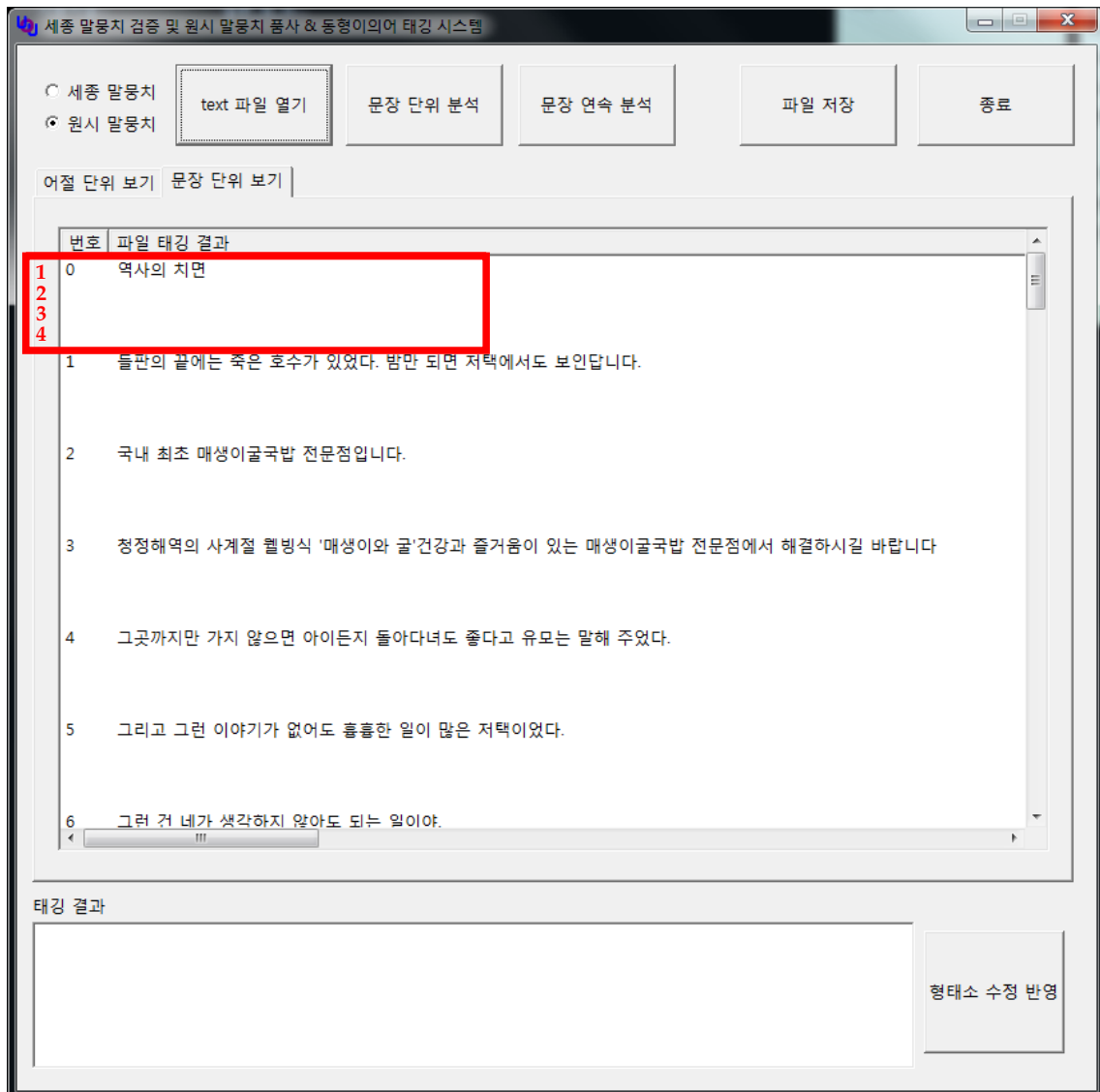


OK버튼을 클릭합니다.

### 3.2 원시말뭉치 분석



- ① 먼저 <원시 말뭉치>를 선택합니다.
- ② 다음으로 <text 파일 열기>를 선택합니다. 여기까지는 세종말뭉치와 같습니다.



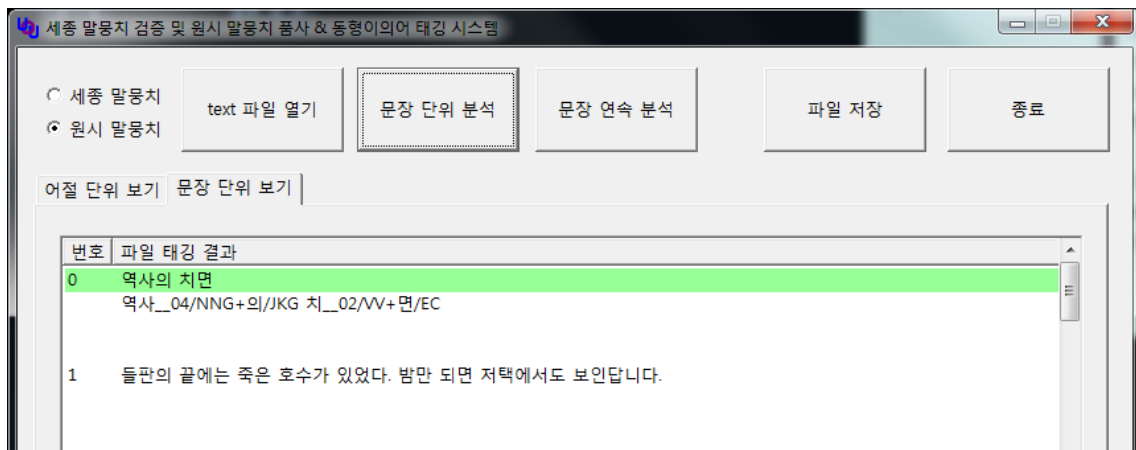
③ 원시 말뭉치를 열면 위와 같이 화면이 나타납니다.

화면에 나타난 순서의 의미는 다음과 같습니다.

- 1 역사의 치면 - 일반 문장입니다.
- 2 빈칸 - 울산대 태깅 시스템이 분석한 결과 위치
- 3 빈칸 - 문장을 구분하기 위한 빈칸입니다.
- 4 빈칸 - 문장을 구분하기 위한 빈칸입니다.

④ 이제 <문장 단위 분석> 버튼을 클릭합니다.

- 원시 말뭉치 분석할 때 <문장 연속 분석> 버튼은 <문장 단위 분석>과 같습니다.
- 세종 말뭉치처럼 비교 결과가 없기 때문에 한 문장만 분석합니다.



- 아래쪽에 태깅 결과가 들어간 것을 확인할 수 있습니다.
- 여기서 초록색으로 나타난 것은 현재 선택된 문장을 알려주는 의미입니다.

⑤ <어절 단위 보기> 탭을 선택합니다.



- 왼쪽의 [형태소 분석 결과]는 선택하거나 수정할 수 있는 형태소 분석들입니다.
- 오른쪽의 [세종말뚝치 분석 결과]는 울산대 태깅 시스템이 분석한 형태소 분석 결과 중 태깅 시스템이 결정한 형태소 분석 결과를 나타냅니다.

⑥ 그 외의 선택 및 수정, 저장은 세종말뚝치의 방법과 동일합니다.

연구책임자: 옥철영  
서울대학교 컴퓨터공학과 대학원 졸업(공학박사)  
현 울산대학교 컴퓨터정보통신공학부  
e-mail : okcy@ulsan.ac.kr

## 어휘의미 체계 기반 입체적 국어사전 확장

---

발행처 : 국립국어원  
서울시 강서구 방화3동 827  
전화 02-2669-9775  
인쇄일 : 2009년 11월 20일  
발행일 : 2009년 11월 22일  
인쇄처 : 울산대학교 발간실  
전화 052-259-2043

---

THE NATIONAL INSTITUTE OF KOREAN LANGUAGE



국립국어원  
2009-01-30

어휘의미 체계 기반 입체적 국어사전 확장



국립국어원