
국어 어휘 의미망 구축의 개념과 사전 편찬*

옥철영** • 울산대학교 컴퓨터정보통신공학부 교수

1. 머리말

컴퓨터가 한국어를 처리하기 위해서는 여러 언어 정보 및 자원이 필요하지만, 무엇보다 국어사전이 가장 기본적이고 기초적인 자원이다. 국어 사전에는 표제어에 대한 품사, 뜻풀이(상위·하위 개념, 부분·전체 개념), 의미 자질, 용례, 동의어, 유의어, 반의어, 전문 용어, 특수어, 관련어, 방언, 어원, 관용구(idiom) 등 다양한 정보가 포함되어 있다. 그러나 이러한 국어사전은 인간 가독형 사전으로, 인간과 같은 언어 처리 능력을 갖지 못한 현재의 컴퓨터가 국어사전을 바로 해독할 수는 없다. 또한 데이터베이스화된 전자 사전도 어휘의 검색 측면에서 인쇄 사전과는 다른 인터페이스(interface)를 제공할 뿐 결국은 인간에 의해 해독된다.

현재까지 한국어 정보 처리에서는 국어사전의 다양한 정보 중 표제어,

* 이 글은 최호섭 연구원(한국과학기술정보연구원 정보기술개발단, hschoe@kisti.re.kr)과 공동으로 작성하였음.

** okcy@ulsan.ac.kr

품사 정보, 약간의 의미 정보(동의어, 유의, 관용구어 등) 정도만을 활용하여 정보 검색 시스템, 철차 오류 교정 시스템, 형태소 태깅 시스템, 기계 번역 시스템 등이 개발되어 왔다. 그러나 이러한 시스템들은 엄밀한 의미에서 아직 의미 처리를 하지 못하고 있다. 이것은 국어사전의 뜻풀이에 내재된 다양한 의미 정보를 컴퓨터가 바로 해독할 수 있는 정도의 언어 처리 기술이 개발되지 못했을 뿐만 아니라, 컴퓨터가 바로 활용할 수 있는 형태로 제공되지 못하기 때문이다.

어휘 의미망은 일종의 지식 베이스(knowledge base)로, 사전 내의 다양한 의미 정보를 컴퓨터가 처리할 수 있도록 체계화시킨 다양한 의미 관계 데이터베이스이다. 컴퓨터가 인간과 대화할 수 있는 정도의 고부가 가치적인 언어 처리가 가능하게 하기 위해서는 이러한 어휘 의미망이 필수적인 요소이다.

본고에서는 그동안 울산대학교 한국어처리연구실에서 구축한 국어사전에 기반한 한국어 어휘 의미망을 소개하고, 이러한 어휘 의미망이 사전 편찬에 어떻게 활용될 수 있을지 살펴보고자 한다.

2. 어휘 의미망의 구축 현황

최근 들어서는 자연 언어의 기본적인 어휘적 의미, 구문적 의미, 담화적 의미를 바탕으로 행위나 현상, 상태 등에 담긴 의미론적·개념론적 특성을 포함하고 있는 의미적 언어 자원(semantic language resource)에 대한 연구가 활발하다. 사전(dictionary&encyclopedia)을 비롯하여 의미 주석 말뭉치(sense-tagged corpus), 어휘분류, 시소러스, 어휘 의미망, 온톨로지 등이 대표적인 의미적 언어 자원이라 할 수 있는데, 이 중 어휘 의미망은 최근 모든 분야에서 집중적으로 연구되고 있는 대상이다. 특히 어휘 의미망과 관련된 논의는 언어학과 자연 언어 처리 분야를 중심으로 대규모 단일어 어휘 의미망, 다국어 어휘 의미망을 구축하기 위한 방법과 이를 바탕으로 실제 구축에 대한 연구 개발이 최근 몇 년간 다양하게 진행되어 있다.

국외에서는 어휘 데이터베이스, 시소러스, 어휘 의미망, 온톨로지와 같

은 어휘 기반의 의미적 언어 자원이 체계적으로 장기간 연구 개발되고 있는데, WordNet, Euro WordNet, EDR, CYC, UMLS 등이 그 대표적인 예라 할 수 있다. 특히 미국 프린스턴대학 인지과학연구실에서 1985년부터 현재까지 연구 개발 중인 워드넷(WordNet)은 국제적으로 가장 많이 알려진 어휘 데이터베이스이자 어휘 의미망의 표준 구축 사례로 인정받고 있다. 나아가 이 연구 사례는 네덜란드어, 이탈리아어, 스페인어, 독일어, 프랑스어, 에스토니아어, 체코어 등 유럽 언어에 대한 통합형 어휘 의미망인 Euro WordNet이라는 대규모 프로젝트로까지 이어졌으며, 최근에는 이를 바탕으로 러시아어, 중국어, 한국어 등까지 확대하여 범용적 어휘 의미망 구조를 형성하기 위한 많은 연구 개발이 진행되고 있다. 이러한 국외의 일련의 연구 개발 사례는 실험적인 구축에 그치는 것이 아니라 언어 교육, 언어 처리, 기계 번역, 정보 검색, 경영 정보 시스템, 의료 정보 시스템, 시맨틱 웹(Semantic Web) 등 다양한 분야에서 핵심적인 언어 자원으로 활용되고 있는 실정이다.

국내에서도 최근 몇 년 동안 이러한 의미적 언어 자원에 대한 연구 개발이 활발히 진행되고 있다. 그러나 학문적·기술적·산업적 경쟁력을 가진 의미적 언어 자원에 대한 연구 개발은 아직 미흡한 상태라 할 수 있다. 이는 이론적 기반 문제, 기술적 문제, 장기간의 연구 개발 지원 미비 등 다양한 연구 개발 구축 환경의 어려움으로 인해 실험적 수준으로 의미적 언어 자원 구축이 진행되고 있기 때문이다. 이러한 국내의 연구 개발 환경에도 불구하고 국내의 몇몇 연구소, 대학 연구실, 업체 등을 중심으로 대규모의 의미적 언어 자원이 연구 개발되고 있는데, 한국전자통신연구원(ETRI)의 어휘 개념망, 국립국어원(21세기 세종계획)의 대규모 의미 주석 말뭉치 및 어휘 의미 부류, 한국과학기술원의 NTT 어휘 대계를 이용한 한·중·일 다국어 어휘 의미망(CorNet), 부산대학교의 워드넷(WordNet) 영한 번역 결과와 한국어 실정에 맞게 수정한 어휘 의미망(KorLex), 울산대학교의 사용자 어휘 지능망(U-WIN) 등이 그 예라 할 수 있다.

미국, 유럽, 일본 등 기술 선진국에서는 이미 자국 언어에 대한 대량의 어휘 의미망을 비롯한 기타 의미적 언어 자원이 구축되었을 뿐만 아니라 타국어에 대한 기술력까지도 확보하고 있는 상태라는 것을 알 수 있다.

그러므로 국내에서도 기술 선진국과 같은 의미적 언어 자원 연구에 대한 학문적·기술적 향상을 비롯한 연구 및 기술 경쟁력을 갖추기 위해서는 한국인의 지식 체계를 담고 있는 대규모의 어휘 의미망 형태의 의미적 언어 자원이 시급히 구축되어야 한다.

3. 사용자 어휘 지능망(User-Word Intelligent Network, U-WIN)¹⁾ 소개

3.1. U-WIN 개발 목표

울산대학교 한국어처리연구실에서 2002년부터 개발 중인 U-WIN은 한국어의 공통적이고 개별적인 속성을 바탕으로 한국인의 보편적인 인지 체계와 개념 관계를 파악하여 이를 어휘의 의미적·개념적 네트워크로 형성한 어휘 의미망이다. 이러한 U-WIN은 언어학을 비롯하여 한국어 정보 처리, 정보 검색, 기계 번역, 시맨틱 웹 등 다양한 분야에 이용될 수 있는 대규모 어휘 지식 베이스를 목표로 하고 있다. 현재 언어 교육용 시스템, 자동 어휘 학습 시스템, 복합 명사 자동 생성 및 뜻풀이 생성 기술, 전문 분야별 개념 체계 자동 생성 기술, 온톨로지 기반 의미적 주석(ontology-based semantic annotation) 기술과 관련 있는 단어 중의성 해소(word sense disambiguation)와 의미 주석(semantic tagging) 기술, 정보 검색에서의 의미적 질의 확장 등 다양한 기술에서 활용되고 있다. 나아가 U-WIN 영어 버전을 구축 중에 있어서 조만간 한국어를 중심으로 한 다국어 U-WIN이 개발될 계획이며 워드넷(WordNet)과의 사상 구조(mapping structure)도 기대할 수 있게 되었다.

U-WIN은 다의어 수준의 어휘 의미망으로 현재 30만 여 어휘가 구축된 상태이다. 그 구축 대상도 한국어 어휘 전체(모든 품사 및 언어 단위)로서, 핵심적 대상으로는 명사, 동사, 형용사이며, 부수적 대상으로는 부사, 관형사, 대명사, 감탄사, 조사, 수사, 의존 명사 등이며, 기타 정보적

1) 보다 자세한 내용은 최호섭 외(2006b) 참조.

대상으로는 북한어, 방언, 옛말, 전문 용어, 고유 명사, 어근, 어미 등으로 한국어 어휘 전체를 대상으로 연구 개발 중이다. U-WIN은 현재 1단계 ‘기본 기술 연구 및 내부 구조 구축’(2002~2003), 2단계 ‘U-WIN의 세부 구축’(2004~2005), 3단계 ‘U-WIN 검증 및 확장 구축 및 응용 기술, 영어 버전 1단계’(2006~2007)로 나누어 장기간 연구 개발 계획을 수립하여 연구 개발이 진행되고 있다.

3.2. 동형 이의어 및 다의어 처리

동형 이의어와 다의어에 대한 어휘망과 관련된 논의는 없는 실정인데, 2) 표준화된 어휘망 관리, 상이하게 구축된 어휘망 간의 통합 및 비교, 다국어 어휘망 구축에서의 의미 표지 관리 방법 등 이들과 관련된 기술 방법과 처리 원칙이 연구되어야 할 것이다. 이러한 문제 때문에, U-WIN에서는 기초 자원 중 『표준국어대사전』에 기술된 동형 이의어와 다의어에 대한 기술 방식을 그대로 반영하되, 추가적으로 기술되는 어휘나 의미에 대해서는 나름대로 원칙으로 기술하고자 하였다.³⁾ 또한 동형 이의어와 다의어를 관리할 수 있는 개별 식별자(identifier)를 할당하여 관리하였다.

먼저 동형 이의어는 국어사전에서 기술되고 있는 어휘의 어깨번호(예를 들어 배¹, 사과¹ 등)를 이용하여 의미 표지(sense tag)를 ‘배_1, 사과_1’과 같은 형태로 기술하고, 동형 이의어가 아닌 경우에는 의미 표지를 부착하지 않았다.

2) 동형 이의어(동음이의어)와 다의어에 대한 연구는 다수 있었으나, 더욱 심도 있게 연구되어야 할 대상으로 생각된다. 나아가 사전 편찬에서의 처리 지침, 동형 이의어와 다의어에 대한 의미 표지 기술 방법 및 의미 기술 순서 등에 대해서도 논의되어야 할 것이다. 이는 언어 처리적 관점에서 이들에 대한 표준화된 의미 표지(sense tag)는 의미 분석 기술에서 반드시 요구되는 사항이기도 하다.

3) 예를 들어 『표준국어대사전』에서는 ‘지지(地支)=십이지(十二支)’의 하나인 ‘진(辰)’에 대한 의미가 누락되어 있다. 그리하여 ‘진_04’의 3가지 의미에 이어 4번째 의미를 추가함과 동시에 동일한 의미 태그를 부여하고, 개별적인 식별자를 할당하였다. 또한 ‘기록(記錄)_2’와 같이 그 의미가 “① 주로 후일에 남길 목적으로 어떤 사실을 적음. 또는 그런 글.”이라고 기술되어 있는데, U-WIN에서는 용례와 말뭉치에서 사용 실태를 분석하여 ‘기록_2’를 ‘표현’이라는 어휘 개념과 ‘글(대상물)’이라는 어휘 개념으로 분리시켜 ‘지지’와 마찬가지로 설정하였다. 동형 이의어에 대해서는 새로운 어휘가 추가될 경우 마지막 동형 이의어 의미 표지의 다음 번호를 할당하여 관리한다.

다음으로 다의어는 하나의 어휘가 가지는 다의적 특성으로 인한 의미적 포괄성을 해소하기 위하여, 다의어의 뜻풀이 하나하나를 개별적인 어휘로 구분하였다. 그리하여 어휘와 뜻풀이의 관계를 1:n의 관계가 아닌 1:1의 관계로 설정하여 관리하였다. 이러한 다의어의 개별 어휘화는 의미적 계층 구조의 세밀성과 더불어, 동의 관계, 유의 관계 등과 같은 의미 관계 설정에서도 중요한 역할을 담당한다.

3.3. 의미 관계와 개념 관계

U-WIN에서의 의미 관계는 어휘 의미론에서 다루고 있는 상하 관계, 동의 관계, 유의 관계, 반의 관계, 부분 전체 관계, 함의 관계⁴⁾ 등을 사용하는데, 사전에서 추출될 수 있는 각종 어휘 정보를 통해 이러한 반자동으로 구축함과 동시에, 기초 자원들을 활용하여 문장의 표면 구조를 중심으로 하여 어휘의 계열(paradigmatic)과 통합(syntagmatic) 관계를 분석하여 어휘의 의미 관계를 설정에 활용하였다.

3.3.1. 동의 관계, 유의 관계, 반의 관계

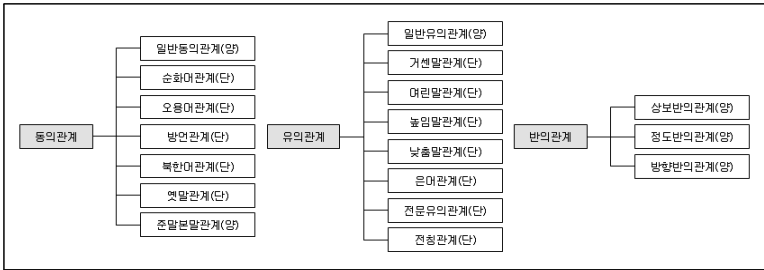
U-WIN에서는 <그림 1>과 같이 동의 관계, 유의 관계, 반의 관계를 세밀화하여, 일반적인 어휘 의미망의 기본 어휘 관계와는 다르게 한국어의 특징 학습 및 활용적 측면을 강화할 수 있는 어휘 관계를 설정하였다.⁵⁾

3.3.2. 상하 관계, 부분 전체 관계

상하 관계는 넓은 의미에서 부분 전체 관계를 포함하는 개념이다. 어휘 의미망 연구에서는 부분 전체 관계로 어휘 전체에 대한 계층적 구조를 형성시키지 못하므로 상하 관계로 설정하지 않는다. U-WIN에서도 상하 관계와 부분 전체 관계를 구분하여 구축하고 있다.

4) 함의 관계는 현재 U-WIN에서 고려하지 않은 관계이다. 관련 연구 및 용언 어휘 의미망의 2단계 구축 작업이 진행되면 함의 관계에 대한 필요성 유무가 결정될 듯하다.

5) U-WIN에서는 일반적으로 국어사전에서 기술하고 있지 않는 ‘서술성 명사+용언화 접사’류 용언에 대한 동의·유의·반의 관계를 서술성 명사 정보를 이용하여 자동으로 연결시켰다.



<그림 1> U-WIN에서의 동의·유의·반의 관계의 세밀화(괄호는 방향성을 나타낸 것)

U-WIN에서는 용례 기반의 인식 구조 및 특정 관련성에 의해 수집된 부류의 집합 구조 형태의 ‘의미 분류적 구조’에서 탈피하여, 엄밀한 의미에서의 상하 관계를 중심으로 층위를 형성한 어휘 집합체를 ‘계층적 구조’로 설정한다. 이러한 계층적 구조는 상하위 층위(또는 노드)가 의미적으로 밀접한 연관성을 가짐과 동시에, 보편적인 개념화(추상화) 과정으로 인식되는 구조이며, 문장의 표면 구조에서 어휘의 계열 관계와 통합 관계 분석을 통해 어느 정도 증명될 수 있는 구조라 할 수 있다.

이러한 상하 관계에 의한 계층 구조 형성의 기본 원칙을 정리하면 다음과 같다.

- ① Is_A, Kind_Of 관계만을 상하 관계로 설정한다.
- ② Top-Down/Bottom-Up 방식 모두를 이용한다.
- ③ 상하 관계는 어휘의 본질적 의미를 통해 연결된다.
- ④ 상하 관계 검증 방법을 적절히 이용한다.
- ⑤ 하위어의 정보를 상위어가 가지도록 한다.
- ⑥ 하위어는 상위어의 속성을 상속받는다(제약 조건이 있을 수 있다).
- ⑦ 먼저 어휘의 형태적 측면(한자, 형태소 분석 정보)을 통해 상위어를 검토하고, 다음으로 각종 사전의 뜻을 통해 상위어를 검토하고, 마지막으로 말뭉치와 각종 문서를 통해 용례를 검토한다. 이들은 모두 단계적으로 분석되어야 할 대상이다.
- ⑧ 전문 용어는 그 특수성을 고려하여 분야별 전문 용어 분류 체계를

최대한 반영하되, 여러 분류 체계가 있을 경우에는 일반적인 상하 관계 체계를 따른다.

- ⑨ 동의 관계와 유의 관계 중 양방향 어휘 관계를 제외한 나머지 관계는 상하 관계에 포함시킨다.
- ⑩ 계층성을 가지는 용언 어휘 의미망과 부사 어휘 의미망은 별도의 원칙을 설정한다.6)

U-WIN에서의 부분 전체 관계는 어휘의 형태적 측면을 고려한 관계 설정을 우선시하고, 특정 물건이나 기관의 세밀한 부분 전체 관계는 개념 관계에서 다루도록 하였다. 그리하여 아래와 같은 세 종류의 경우에만 기본 의미 관계의 부분 전체 관계로 설정하였다.7)

- 특정 어휘 개수를 표현한 어휘: 사육신(死六臣), 십계(十界), 십이지(十二支), 이십사절기(二十四節氣)……
- 한자 합성어 중 의미 파생이 없는 어휘: 춘추(春秋), 하동(夏冬), 가관(歌管), 승패(勝敗)……
- 한글 합성어 중 의미 파생이 없는 어휘: 봄여름, 봄가을……

U-WIN에서의 개념 관계는 <표 1>과 같이 의미 관계의 세부적이지자 확장적인 의미적 속성을 이용한 관계를 말한다. 이는 기본 개념 관계8)와 확

-
- 6) 용언 어휘 의미망과 부사 어휘 의미망은 현재 구축 중에 있는 작업으로, 용언 어휘 의미망은 형용사의 경우 형용사의 일반적인 분류인 성상, 정도, 수량, 지시 형용사로 구분하여 구축하고, 동사의 경우 자동사, 타동사, 자타동사로 분류하여 용례와 구문 관계 등을 중심으로 구축하고 있다. 부사 어휘 의미망은 성상, 정도, 의태, 의성, 문장, 시간 부사 등으로 분류하고, 용례를 중심으로 하되 구문 관계를 적절히 이용하여 구축 중이다. 이러한 일련의 작업은 본 연구팀이 보유 중인 의미 주석 뜻풀이 말뭉치를 통해 일차적으로 자동 구축된 용언과 부사 계층 구조를 중심으로 진행되고 있다.
 - 7) 부분 전체 관계와 관련된 연구는 좀 더 세밀하게 연구할 필요가 있다. 그리하여 이 부분 전체 관계의 세밀한 부분은 개념 관계에서 다루어 면밀한 분석 작업이 이후 관계 설정이 가능하리라 생각된다.
 - 8) U-WIN에서의 기본 개념 관계는 '21세기 세종계획'의 전자 사전 개발 본과의 연구 결과로 도출된 개념 관계를 이용한 것이다. 이 개념 관계는 의미 관계의 확장이라 할 수 있다. 예를 들어 구성원 계, 구성 요소 관계, 생산 관계 등이 있다.

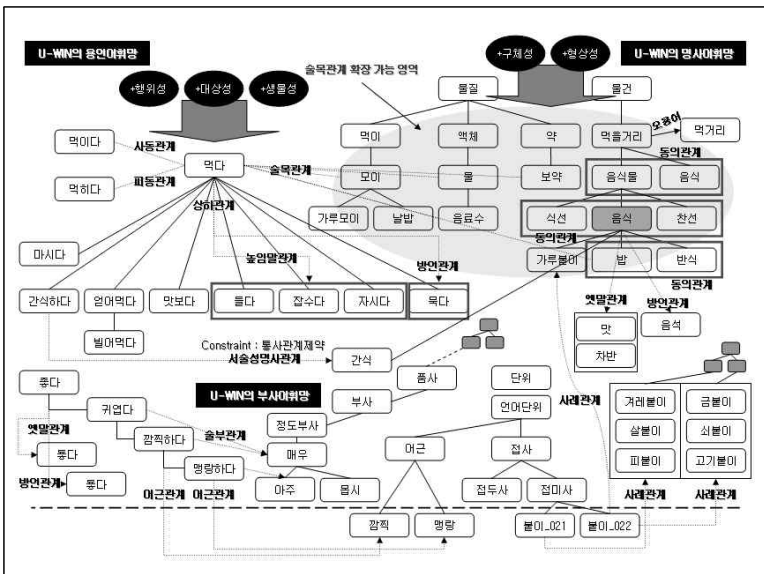
<표 1> U-WIN에서의 개념 관계

기본 개념 관계	포함 관계, 재료 관계, 인술 관계, 운용 관계, 제어 관계, 소유 관계, 상징 관계, 이름 관계, 구성원 관계, 구성 요소 관계, 발명 관계, 측정 관계, 분류 관계
확장 개념 관계	연락 관계, 발견물 관계, 위치 관계, 관리 관계, 성질 관계, 발표·출판 관계, 연구 관계, 이론·학설 관계, 활성화 관계, 분포 관계, 기능 관계, 증상 관계, 반응 물질 관계, 현상 관계.....

장적 개념 관계로 분리하여, 일반적인 개념 관계와 특정 분야 중심의 개념 관계를 구분하고자 하였다. 이러한 개념 관계는 어휘 의미망 전체의 공통적인 어휘 관계라기보다는, 어떠한 공통적인 특징을 가진 특정 어휘(노드) 집합이나 특정 분야와 관련된 어휘 집합에서 발생하는 어휘 관계이다.

3.4. 내부 구조

구축된 U-WIN의 내부 구조는 <그림 2>와 같다.



<그림 2> U-WIN 내부 구조의 도식적 표현

4. 어휘 의미망의 사전 편찬에의 활용

4.1. 현재의 국어사전의 문제점

현재의 국어사전은 대체로 다음과 같은 문제점이 있다.

(1) 국어사전은 누구를 대상으로 만들어졌는가? 현재의 국어사전에 등재된 표제어에 대해 어휘의 중요도 혹은 어휘 학습 수준에 따른 기본 어휘에 대한 정보가 제공되지 않고 있다. 물론 어린이용 국어사전 등이 있으나, 우리가 흔히 접하는 영어사전에는 중학교 기본 어휘(‘+’), 고등학교 기본 어휘(‘**’), 대학 교양 정도의 어휘(‘*’)에 대한 정보를 제공하고 있다. 그동안 국립국어원에서도 한국어 학습용 어휘 선정을 위해 세종 21세기 말뭉치에 기반한 국어 사용 빈도 조사(조남호, 2002; 김한샘, 2003; 김한샘, 2005)가 있었으나, 아직은 기초 어휘·기본 어휘에 대한 합의가 없어 이러한 조사 결과가 사전에 반영되지 못하고 있다.

(2) 뜻풀이에 사용된 어휘가 정제된 것인가? 기본적으로 뜻풀이에 사용된 어휘는 사전 표제어로 등재되어야 하지만 그렇지 못한 경우도 다수 발견되고 있다. 또한, 기본 어휘가 정의되지 않음으로써, 어휘의 뜻풀이에서 사용되는 어휘 수준에 대한 어떠한 지침이 없이 정의되고 있다. 예를 들어 고등학교 기본 어휘는 최소한 중학교 및 고등학교의 기본 어휘로 정의되어야 하는데 현실적으로 그렇게 정의되지 않는다. 이러한 사정은 교과서에서 사용되는 어휘들에서도 발견되며, 특히 전문 용어는 아무런 기준 없이 사용되고 있는 실정이다. 사전 편찬 작업은 다수의 인력에 의해 오랜 기간 진행되기 때문에 이러한 작업은 사전 편찬에 앞서 선행되어야 할 과제이다.

(3) 뜻풀이 서술 방법에 원칙이 있는가? 동일한 차원의 어휘를 설명하는 방법이 통일되지 않고 편찬자에 따라서 그 방법이 다르거나, 정보가 누락되는 경우가 있다. 예를 들어, ‘십이지(十二支)’ 각각이 『표준국어대

사전』에 다음과 같이 정의되어 있다.

자: 십이지(十二支)의 첫째. 쥐를 상징한다.
축: 십이지(十二支)의 둘째. 소를 상징한다.
인: 십이지의 셋째. 범을 상징한다.
묘: 십이지의 넷째. 토끼를 상징한다.
진: “정의되어 있지 않음.”
사: 십이지의 여섯째. 뱀을 상징한다.
오: 십이지의 일곱째.
미: 십이지의 여덟째. 양을 상징한다.
신: 십이지의 아홉째. 원숭이를 상징한다.
유: 십이지(十二支)의 열째.
술: 십이지의 열한째. 개를 상징한다.
해: 십이지의 열두째. 돼지를 상징한다.

여기서 ‘진’은 정의되어 있지 않으며, ‘오’와 ‘유’는 여느 항목과 다르게 정의되어 있다.

또한 ‘위’와 ‘아래’의 뜻풀이 말은 다음과 같이 정의되어 있다.

1. 어떤 기준보다 더 높은 쪽. 또는 사물의 중간 부분보다 더 높은 쪽.
2. 길고 높은 것의 꼭대기나 그쪽에 가까운 곳.
3. 어떤 사물의 거주이나 바닥의 표면.
4. 신분, 지위, 연령, 등급, 정도 따위에서 어떠한 것보다 더 높거나 나은 쪽.
5. 글 따위에서, 앞에서 밝힌 내용.
6. 강 따위의 물이 흘러가는 반대 방향이나 부분.
7. 시간적 순서가 앞에 오는 것.
8. 어떤 일이나 조건 따위에 의하여 특징지어지는 테두리나 범위.
9. 어떤 것의 바깥이나 이외.

1. 어떤 기준보다 낮은 위치.
2. 신분, 연령, 지위, 정도 따위에서 어떠한 것보다 낮은 쪽.
3. 조건, 영향 따위가 미치는 범위.
4. 글 따위에서, 뒤에 오는 내용.
5. ‘음부07(陰部)’를 완곡하게 이르는 말.

‘위’와 ‘아래’의 다의적인 의미의 분화가 달라 다의적인 의미의 개수와 중요도가 다를 수 있겠으나 첫 번째 의미의 뜻풀이의 상위 개념어가 ‘쪽’과 ‘위치’로 서로 다르며, ‘위’와 ‘아래’가 반대어라면 다의적인 서술 순서를 일치시킬 필요가 있다.

(4) **표제어를 정의하는 개념 체계(어휘 의미망)가 있는가?** 일반적으로 표제어(명사류)는 유개념과 종차로 정의된다. 여기서 유개념은 표제어가 가진 개념의 상위(상하위 관계) 혹은 전체(부분-전체 관계) 개념어이다. 그러나 한국어 어휘에 대한 이러한 체계적이고 포괄적인 개념 체계가 존재하지 않으므로 해서(혹은 개념 체계가 존재한다 하더라도 이를 활용하지 않으므로 해서) 순환 정의 및 동일 수준의 개념어를 다른 개념 범주로 정의하는 경우가 많다. 이러한 체계적이고 포괄적인 개념 체계 구축 작업은 사전 편찬 작업이 다수의 인력에 의해 병렬적으로 진행되기 때문에 사전 편찬에 앞서 선행되어야 할 것이다.

(5) **표제어 등재 원칙이 있는가?** 현재 국어사전에는 사전 규모에 따라서 50,000 어휘에서 400,000 어휘가 등재되어 있다. 그러나 어휘(전문 용어, 복합 명사) 등재 원칙이 없음으로 해서 균형적인 구성을 갖추지 못하고 있다. 사전의 등재된 어휘 수가 막연히 많다고 해서 좋은 사전이라고는 말할 수 없다. 현재의 『표준국어대사전』은 약 400,000어휘를 등재한 대규모 사전이지만 백과사전적으로 표제어가 등재되어 있다. 어휘의 사용 빈도 등이 고려할 수 있는 등재 원칙이지만, 이러한 사용 빈도도 동형 이

의어별, 다의어의 의미별로 자세히 조사되어야 할 것이다.

이러한 사전의 문제점들은 대체로 사전 편찬 전에 필요한 기준이나 원칙이 설정되어 있지 않고, 설사 이러한 원칙이 있다 하더라도 장기간 여러 편찬자에 의해 통일적인 관점이나 도구 없이 작업이 이루어진 결과이다. 따라서, 어휘 의미망과 사전 편찬에 필요한 도구가 있다면 이러한 여러 문제(일관성, 완전성, 대칭성)들은 상당히 해소될 수 있을 것이라 판단된다. 이 장에서는 그동안 울산대학교 한국어처리연구실에서 구축한 한국어 어휘 의미망인 U-WIN을 국어사전 편찬에 어떻게 활용할 수 있을지를 살펴보고자 한다.

4.2. 품사 및 다의어 수준 의미 태그 부착 뜻풀이로부터 기초 어휘 추출

울산대학교 한국어처리연구실에서는 동형 이의어 태깅 시스템을 구현하기 위하여, 약 15만여 개의 표제어로 구성된 중·소규모의 국어사전(금성사전)의 뜻풀이 전체(약 100만 어절)를 대상으로 품사 태그 및 동형 이의어 수준의 의미(sense) 태그를 부착한 뜻풀이 말뭉치를 구축(2002년~2004년)하고, 동형 이의어와 함께 사용된 체언(일반명사)과 용언(동사, 형용사)의 공기 빈도를 동형 이의어 분별에 필요한 의미 정보로 추출하였다. 또한 현재는 U-WIN의 용언 어휘 의미망 구축과 구축된 동형 이의어 태깅 시스템을 U-WIN과 연계하고 다의어 태깅 시스템 구축을 위해 『표준국어대사전』의 뜻풀이 말을 다의어 수준에서 태깅하고 있다.

품사 및 다의어 수준의 의미 태그가 부착된 뜻풀이로부터 한국어 처리 및 국어사전 편찬에 필요한 다음과 같은 여러 형태의 통계 자료를 추출할 수 있다.

- ① 어절별 품사 태그 빈도
- ② 어근별 품사 빈도
- ③ 동형 이의어별 사용 빈도
- ④ 다의어별 사용 빈도

Mm	1	2	3	4	5	6	7	8	9	10	11	12	NUM
공통어													1
공통어													1
공통어													1
공통어													2
공통어													22
공통어													5
공통어													4
공통어													4
공통어													39
공통어													1
공통어													4
공통어													2
공통어													1
공통어													1
공통어													10
공통어													4
공통어													274
공통어													16
공통어													81
공통어													12
공통어													2
공통어													1
공통어													20
공통어													4
공통어													2
공통어													7
공통어													1
공통어													2
공통어													13
공통어													1
공통어													1
공통어													14
공통어													3
공통어													2
공통어													3
공통어													2
공통어													1
공통어													4
공통어													2
공통어													5
공통어													1
공통어													4
공통어													1
공통어													3
공통어													2
공통어													2
공통어													1
공통어													3
공통어													1

<그림 3> 사전 뜻풀이에 나타난 어근별 품사 빈도 예

Mm	Tag	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	24	31	NUM
공통어	MAG																									
가	ETR	4410																								
가	NNB																									
가	NNV																									
가	WV																									
가	XSN																									
가	SPN																									

<그림 4> 사전 뜻풀이에 나타난 동형 이의어별 사용 빈도 예

앞의 <그림 3>과 <그림 4>는 어근별 품사 빈도 및 동형 이의어별 사용 빈도를 보이고 있다.

사전 편찬을 위한 기초 어휘 및 기본 어휘에 대한 통계 자료는 다양한 분야의 균형 있는 말뭉치에서도 추출되어야겠지만, 적어도 현재까지 편찬된 국어사전의 태깅된 뜻풀이 말에서 추출한 정보들을 이용하여서도 사전 뜻풀이 말에서 필요한 어휘 및 동형 이의어별·의미별로 빈도에 따른 중요도 등을 결정할 수 있을 것이다.

4.3. U-WIN 구축 관리 도구를 이용한 통합적 사전 편찬 환경 제공⁹⁾

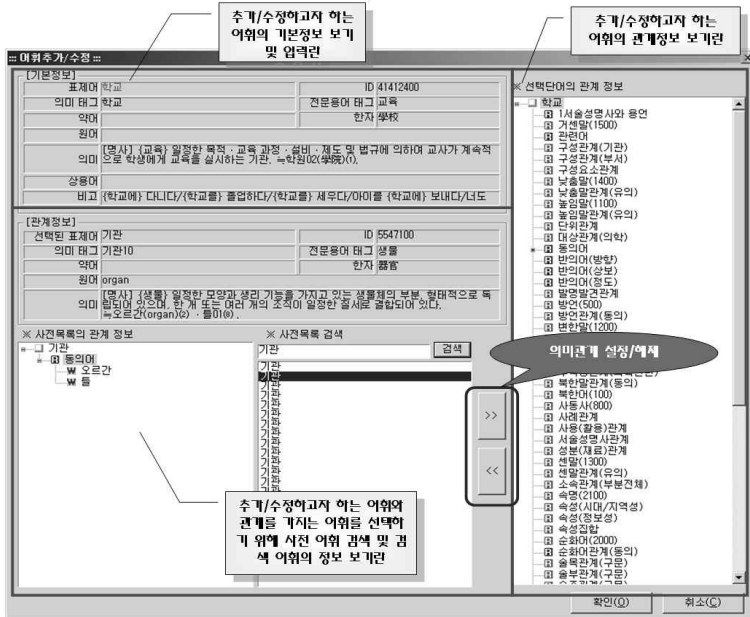
구축 관리 도구는 어휘 추가·수정·삭제 등의 어휘 관리 기능, 어휘망 및 사전 검색 기능, 관계 추가·수정·삭제 등의 관계 관리 기능, 작업 관리 및 환경 설정 기능 등의 기본적인 기능을 비롯하여, 자질 관리 기능, 용례 검색 기능, 패턴 기반의 자동 추천 기능, 전문 분야별 어휘 지능망 자동 생성 기능 등의 U-WIN을 구축하는 데 있어서 좀 더 세밀화하고 효율적인 기능을 지원해야 한다.

<그림 5>는 U-WIN 구축 관리 도구의 실행 화면으로서, ①번 영역은 U-WIN의 상하 관계를 트리(tree) 구조로 검색하는 곳이며, 검색 옵션으로 어휘·식별자(ID) 검색 설정, 새 캔버스 설정, 다의어 구분 검색 설정이 있다. ②번 영역은 리스트 형태의 사전 검색을 하는 곳으로, 검색 옵션은 어휘·뜻풀이 검색 설정이 있으며 패턴 검색을 지원한다. ③번 영역은 리스트에 나타난 사전 검색 결과를 선택한 다음 ①번 영역의 트리에서 선택된 어휘의 하위어로 상하 관계를 설정하거나, 트리에서 선택된 어휘의 상하 관계를 해제하는 곳이다. ④번 영역은 U-WIN의 사전 데이터베이스에 어휘를 추가·수정·삭제하는 곳이며, ⑤번 영역은 상하 관계 이외의 기타 관계 설정 및 표시, 용례 검색, 관계 관리, 자질 관리, DSN 설정 등의 환경 설정을 지원하는 곳으로 탭(tab) 방식으로 구성되어 있다.

9) 보다 자세한 내용은 최호섭(2007) 참조.



<그림 5> U-WIN 구축 관리 도구



<그림 6> U-WIN 구축 관리 도구를 이용한 어휘 수정 화면

다음에서 U-WIN 구축 관리 도구를 기능별로 하나씩 살펴본다.

4.3.1. 어휘 관리(추가·수정·삭제) 기능

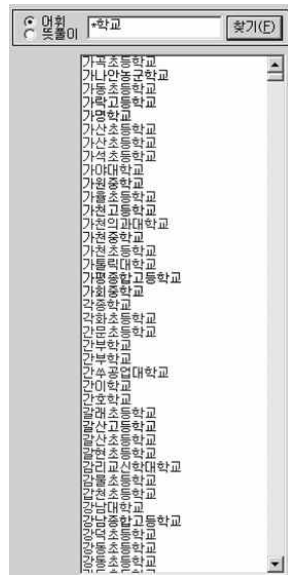
U-WIN 구축 관리 도구는 새로운 어휘를 추가하거나, 기존의 어휘 정보를 수정·삭제하는 어휘 관리 기능을 제공한다. 어휘의 기본적인 정보 및 관계 정보를 관리하는 것은 어휘망을 구성하는 데 있어 가장 기본이 되는 기능이다.

<그림 6>은 어휘 ‘학교’의 수정 화면이다. 표제어·의미 태그·약어·한자·의미 등의 어휘의 기본적인 정보를 보여 주는 부분과 관계 정보를 보여 주는 부분이 있다. 그리고 사전 DB에서 다른 어휘를 검색하여 의미 관계를 설정하거나 해제할 수 있으며, 이러한 의미 관계 설정 및 해제 기능은 메인 화면에서 실행하는 것과 동일한 과정으로 이루어지도록 구성하였다.

4.3.2. 어휘망 및 사전 검색 기능

U-WIN은 많은 어휘와 다양한 관계를 포함하고 있어, 단순한 어휘 검색만으로는 U-WIN 구축자의 어려움이 뒤따른다. 세부적으로 U-WIN을 구축하는 과정에서 검색 기능은 기존의 상하 관계 내에서의 어휘 검색이나, 상하 관계 정보를 제외한 사전 검색 등이 필요하다.

구축 관리 도구에서는 U-WIN의 가장 기본적인 상하 관계는 트리 검색을 통해, 그리고 사전 검색은 리스트 검색을 통해 이루어지며, 트리 및 리스트 검색은 유기적으로 연동되어 있다. 트리 검색은 최상위 노드(U-WIN)를 중심으로 해당 어휘의 상위어 리스트를 확인할 수 있으며, 검색 대상에 따라 표제어(어휘) 검색과



<그림 7> 패턴 검색의 예 (*'학교')

어휘 식별자(ID) 검색으로 구분한다. 특히 어휘 식별자를 통한 검색 방법은 특정 어휘만의 상하 관계 정보를 확인할 수 있도록 하였다. 리스트 검색은 사전 테이블의 검색 필드에 따라 표제어 검색과 뜻풀이 검색으로 구분하며, 검색 방법에 따라서는 단순 스트링 매칭 검색과 패턴 검색이 있다. 패턴 검색은 와일드카드(*)를 사용하여 ‘*검색어*’, ‘*검색어’, ‘검색어*’의 세 가지 검색어 형태를 제공한다. <그림 7>에서는 ‘*학교’의 검색어를 통한 리스트 패턴 검색 결과를 확인할 수 있으며, U-WIN에서의 사용 여부에 따라 리스트의 검색 결과가 파란색(사용 중), 검은색(미사용)¹⁰⁾으로 표시된다.

그리고 리스트 검색 결과에서 해당 하위의 더블 클릭(click)을 통해 좌측 트리 검색과 연동되며, 이때 ‘다의어 구분’ 체크를 통해 어휘 식별자 검색 방법 및 동일한 결과를 확인할 수 있다.

4.3.3. 다양한 관계 관리 기능

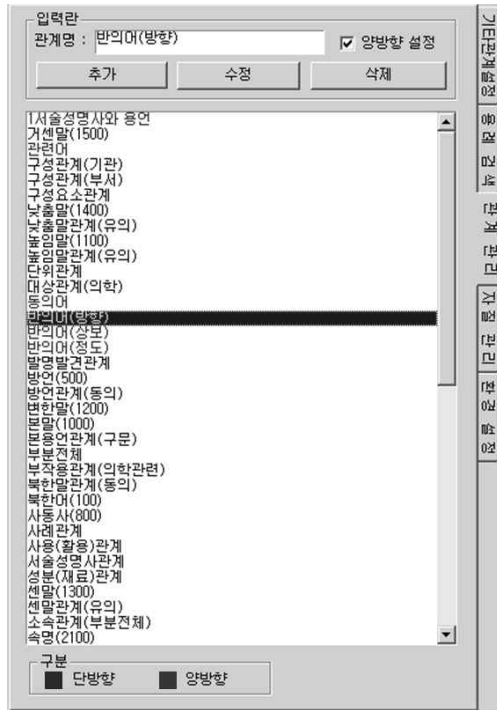
U-WIN에서 가장 중요한 요소는 관계 정보로서, 작업자가 얼마나 의미 있는 관계를 정의하였는가는 어휘망의 품질을 좌우한다. 현재 U-WIN에서는 상하 관계를 중심으로 동의 관계·유의 관계·반의 관계·부분 전체 관계 등의 기본적인 의미 관계뿐만 아니라 형태 관계, 술주 관계·술목 관계·술부 관계 등의 구문 관계 등 다양한 관계를 정의하고 있다.

이러한 각각의 관계들을 추가·수정·삭제하거나, 각 관계의 방향성을 설정해야 한다. <그림 8>은 ‘관계 관리’ 부분으로 상단에 관계를 추가하거나 수정·삭제할 수 있으며, U-WIN에 설정된 관계들이 나열되어 있다. 이때, 관계의 방향성에 따라 파란색(양방향), 검은색(단방향)으로 이루어져 있으며, ‘관계 관리’ 탭(tab)에서 수정된 사항은 U-WIN에 즉시 반영되어 나타난다.

4.3.4. 용례 검색 기능

어휘의 정확한 의미를 파악하기 위해서는 뜻풀이·원어·한자·전문

10) 리스트의 검색 결과가 회색인 경우는 하위어를 가지지만 상위어는 가지고 있지 않은 임시 저장 상태를 나타낸다.



<그림 8> 관계 관리 탭(tab)

분야 등의 사전 정보를 활용하는 방법뿐만 아니라, 말뭉치에서의 어휘의 사용 실태를 살펴보는 방법이 있다. 어휘망을 구축할 때에는 기본적으로 전자의 방법을 권장하지만, 이를 통해 어휘의 의미를 파악하기 어려운 경우에는 후자의 방법이 필요하다.

U-WIN의 구축 관리 도구는 후자의 방법을 지원하기 위해, <그림 9>와 같이 세종 350만 어절의 말뭉치뿐만 아니라 일반 텍스트 파일 형태의 말뭉치 파일에서 말뭉치상에서의 사용 실태를 확인할 수 있다. 용례 검색에 사용되는 말뭉치는 원시 말뭉치와 품사 주석 말뭉치, 의미 주석 말뭉치를 지원하여, 좀 더 정확한 용례 검색 결과를 제공하였다. 또한 U-WIN을 구축할 때 어휘 간의 관계를 수작업으로 추출해야 하지만, 용례 검색 기능을 활용한 패턴 기반 반자동 구축 방법을 이용할 수도 있게 하였다.

	전문어	중심어	우문어	원어
	9월 9일 저녁밥을	먹다가	그릇을 깬습...	3.₩1. VC++₩1. 프로젝트(연구)...
량배로 머기던 남지하고 단들이 마주 앉아 밥을	먹다나?	먹다가	밥줄이 상에...	3.₩1. VC++₩1. 프로젝트(연구)...
민수는 밥줄을 흘릴까 봐 조심조심 밥을	먹다가도	먹다가도	불쑥불쑥 망...	3.₩1. VC++₩1. 프로젝트(연구)...
효철은 길을 가다가도 밥을	얻어먹다가	얻어먹다가	굴참나무 밑...	3.₩1. VC++₩1. 프로젝트(연구)...
거지마이는 이 집 저 집 기웃거리며 먹을 것을	먹다가	먹다가	터진 것이 ...	3.₩1. VC++₩1. 프로젝트(연구)...
국회까지 기금이라는 옆주머니 풀단지를 빨마	먹다가	먹다가	입에 안 맞...	3.₩1. VC++₩1. 프로젝트(연구)...
미는 놀머지게 하품을 하고 먹여 주는 대로 받아	먹다	먹다	보면 밥줄이...	3.₩1. VC++₩1. 프로젝트(연구)...
뿔여 놓은 원장을 풀리나 마나 쪽쪽 떠다가 비벼	먹다가	먹다가	삼정에 남아...	3.₩1. VC++₩1. 프로젝트(연구)...
승씨는 미국을 출항하고 갈비구이를	먹다	먹다	남긴 음식...	3.₩1. VC++₩1. 프로젝트(연구)...
당이나 가게들을 이용한 손님들이 커피나 빵 등				

<그림 9> 원서 말뭉치에서 '먹다'의 용례 검색 결과

4.3.5. 자질 관리 기능

어휘의 의미를 밝히기 위해서는 일반적으로 두 가지 방법이 있다. 첫째는 다른 어휘와의 관계를 규정하는 것으로, 다양한 어휘 관계를 확인할 수 있다. 둘째는 어휘의 개별적이고 독자적인 관계를 규정하는 것으로, 성분 분석(componential analysis) 방법을 통해 이루어질 수 있다. 성분 분석은 한 어휘의 의미를 개별적인 의미 조각의 집합체로 보고, 어휘의 의미를 이루는 의미 성분을 분석하여 의미를 밝히는 방법이다.

U-WIN의 견고한 구축을 위해서는 의미 성분으로 여길 수 있는 자질(feature)을 명확하게 정의하여, 자질 기반의 구축 방법을 병행해야 한다. 자질은 양성·중성·음성의 세 가지 종류를 가지는 것과, 양성·음성의 두 가지 종류를 가지는 것으로 구분할 수 있으며, 자질들 간에도 상하 관계, 반의 관계 등의 관계를 정의할 수 있다.

4.3.6. 패턴 기반의 자동 추천 기능

한 어휘가 다른 어휘와 다양한 관계를 기반으로 유기적으로 연계되어 있는 어휘망을 구축할 때에는, 상하 관계뿐만 아니라 여러 관계에 해당하는 어휘를 수동으로 찾아야 하는 어려움이 있다.

U-WIN은 『표준국어대사전』을 비롯한 다수의 사전을 기초 자료로 하여 어휘 사전 데이터베이스가 구성되었으므로, 하나의 어휘 의미를 판별하는 데 표제어, 한자, 약어, 원어, 뜻풀이, 용례 등의 여러 가지 정보를 이용할 수 있다. 특히 뜻풀이와 용례는 어휘 간의 관계를 설정하는 데 있어서 기반이 되는 자원이므로, 뜻풀이와 용례를 분석하여 해당 어휘의 다양한 관계를 자동으로 추출할 수 있어야 한다.

구축 지원 도구의 용례 검색 기능을 통해 어휘가 뜻풀이와 용례, 일반 말뭉치에서 나타나는 형태를 살펴보면, 특정 형태 및 구문적인 요소와 함께 다른 어휘가 연관되어 있는 경우가 발생한다. 이때 빈번히 발생하는 패턴을 해당 관계와 매칭시켜, 말뭉치에서 패턴 매칭을 통해 해당 어휘의 관계 정보를 자동으로 추출할 수 있다. 이러한 기능은 정의된 패턴의 세밀성, 어휘 간의 관계 추출을 위한 패턴 매칭의 정확성 등을 통해 다양한 관계 정보를 반자동으로 구축할 수 있으므로 구축의 효율성을 향상시킨다.

술주 관계·술목 관계·술부 관계 등의 구문 관계가 패턴 기반 자동 추천 기능의 대표적인 예이다. 예를 들어 ‘~을(를), ~이(가), ~(에)서, ~9(으)로’ 등의 조사를 중심으로 구문 패턴을 정의하여 U-WIN의 구문 관계와 매칭시킬 수 있다. 그런 다음 U-WIN을 구축할 때, 해당 어휘와 구문 관계를 맺는 어휘를 말뭉치를 통해 자동으로 추천할 수 있다.

5. 결론: 어휘 의미망을 이용한 체계적인 국어사전 편찬을 위한 제안

울산대학교 한국어처리연구실은 다의어 수준의 의미 중의성 분별시스템 개발과 보다 정밀한 자질(feature) 추출 및 명사 및 용언의 구문적·의미적 자질(feature) 기반 U-WIN 구축을 위해 2006년부터 기존의 『표준국어대사전』 전체를 대상으로 품사 태깅 및 다의어 수준의 의미 태깅 작업을 진행하고 있다. 뜻풀이의 품사 태깅 작업은 완료되었으며 품사 태깅의 정확률은 약 95%이다. 뜻풀이의 다의어 수준의 의미 태깅 작업은 2007년 8월 말까지 1단계 작업 완료를 목표로 작업 중이며, 현재 70% 정도 진행되고 있다. 이러한 작업이 완료되면 앞서 중소기업의 국어사전을 대상으로 진행했던 결과보다 광범위하고 정밀한 결과를 추출할 수 있을 것이며, 이러한 결과물들은 현재 국립국어원에서 추진 중인 한국어 어휘 의미망 구축과 이를 이용한 국어사전 편찬에 활용될 수 있을 것으로 기대한다.

국내의 어휘 의미망 관련 연구 개발은 얼마 전까지는 이론적 수준에 그친 것이 사실이다. 그러나 최근 몇 년 사이에 어휘 의미망 관련 국내 연구

결과물이 어느 정도의 성과를 보임과 동시에, 구축된 일부 자원들이 공개됨으로써 많은 관련 분야에서 어휘 의미망에 대한 연구가 급속도로 확대되고 있는 추세이다. 따라서 앞으로 보다 세밀한 작업을 위한 체계적인 구축과 더불어, 검증·평가에 대한 사항도 고려해야 할 것으로 생각한다.

나아가 국내 어휘 의미망 연구의 국제적 수준 및 경쟁력 강화를 위해서는 이론적 기반, 공개용 구축 사례, 검증 및 평가, 활용 사례 등을 체계적으로 논의할 수 있는 제 학문적 연구자망에 대한 논의가 있어야 할 것이다. 또한 장기간이 소요되는 어휘 의미망 연구 개발과 같은 핵심적인 기반 연구 개발에 대한 정부와 학계의 적극적인 지원과 관심도 있어야 할 것이다.

| 참고 문헌 |

- 김은영(2004), ‘국어 어휘의 계층적 의미 관계에 대한 고찰’, “한국언어문학” 40, 한국언어문학회.
- 김준수(2005), ‘의미정보와 시소러스를 이용한 한국어 어휘 중의성 해소 모델’, 울산대학교 대학원 컴퓨터정보통신공학부 박사학위논문.
- 김한샘(2003), “한국 현대 소설의 어휘 조사 연구”, 국립국어연구원.
- 김한샘(2005), “현대 국어 사용 빈도 조사 2”, 국립국어연구원.
- 김현권·김병욱(2003), ‘어휘 의미 지식 표상의 방법’, “한글” 262, 한글학회.
- 도원영·이봉원·최경봉·한정한, ‘온톨로지에 기반한 한국어 동사 의미망 구축 시고’, “한국어학” 24, 한국어학회.
- 박만규(2002), ‘다의어의 의미 분할과 의미 부류’, “한글” 257, 한글학회.
- 백지원·정연경(2005), ‘지식조직체계의 용어관계 유형에 관한 연구’, “한국문헌정보학회지” 제39권 제4호, 한국문헌정보학회.

- 이동혁(2004), ‘의미 관계의 저장과 기능에 대하여’, “한글” 263, 한글학회.
- 이은령 · 윤애선(2005), ‘피동 정보를 통한 한국어 동사 어휘 의미망 정제’, “한국어학” 28, 한국어학회.
- 이정식(2003), “다의어 발생론”, 도서출판 역락.
- 정영미(1997), ‘지식 구조론’, “현대정보관리학총서” 28, 한국도서관협회.
- 조남호(2002), “현대 국어 사용 빈도 조사: 한국어 학습용 어휘 선정을 위한 기초 조사”, 국립국어연구원.
- 지식 정보 처리와 온톨로지(KIPONTO) 워크숍 발표 자료집(2003~2005).
- 최경봉(2001), ‘지식 기반 구축을 위한 어휘의 의미 분류’, “담화와 인지” 제8권 2호, 담화인지언어학회.
- 최경봉 · 도원영(2005), ‘한국어 동사 의미망 구축을 위한 상위 온톨로지 구성에 관한 연구’, “한국어학” 28, 한국어학회.
- 최기선 외(2005), “다국어 어휘 의미망(CoreNet)”, 한국과학기술원 전문용어언어공학센터.
- 최호섭 · 옥철영(2002), ‘한국어 의미망 구축과 활용’, “한국어학” 17, 한국어학회.
- 최호섭 · 임지희 · 옥철영(2003), “한국어 사전과 의미망을 이용한 복합 명사 분석과 생성”, 2003년도 한국인지과학회 춘계 학술대회 발표 자료집.
- 최호섭 · 옥철영(2004a), ‘정보 검색 시스템과 온톨로지’, “한국정보과학회지” 제22권 제4호, 한국정보과학회.
- 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 어휘 지능망 구축 방법”, 제609돌 한글학회 전국 국어 학술 대회.
- 최호섭(2007), “대규모 사용자 어휘 지능망 구축과 활용”, 울산대학교 대학원 컴퓨터정보통신공학부 박사학위논문.

- 필옥덕(2004), “현대 한국어 동사 의미 결합 관계 연구”, 도서출판 역락.
- 한정한 · 도원영(2005), ‘한국어 동사 의미망 구축을 위한 어휘 의미 관계 유형’, “한국어학” 28, 한국어학회.
- 황순희 · 윤애선(2005), ‘의미 자질을 고려한 명사 어휘 의미망 구축(1)’, “한국어학” 29, 한국어학회.
- Cruse, D. A.(1986), “Lexical Semantics,” Cambridge University Press.
- Fellbaum, C.(1998), “WordNet: An Electronic Lexical Database,” The MIT Press.
- Murphy M. L.(2003), “Semantic Relation and the Lexicon,” Cambridge University Press.
- Saint-Dizier, P., Viegas, E.(1995), “Computational Lexical Semantics,” Cambridge University Press.
- Sharifian, F.(2002). “Processing Hyponymy in L1 and L2,” *Journal of Psycholinguistic Research*, 31(4).
- Vossen, P.(1998). “EuroWordNet: A Multilingual Database with Lexical Semantic Network,” The Kluwer Academic Publishers.