

# 워드넷 기반 한국어 명사 어휘의미망의 정제

황순희, 윤애선, 부산대학교 한국어 정보처리 연구실, 불어불문학과  
soonheehwang, asyoon@pusan.ac.kr

## Refinement of KorLex based on WordNet

Soonhee HWANG, Aesun YOON, Korean Language Processing Lab, Department of French  
- Pusan National University  
soonheehwang, asyoon@pusan.ac.kr

### 요 약

최근 들어 온톨로지(ontology), 시소러스(thesaurus) 등과 함께 주목받고 있는 Princeton 대학의 워드넷(WordNet, 이하 PWN)은 자연어 처리(NLP)와 관련하여 대안을 제시할 수 있는 어휘의미망(lexico-semantic network)이다. 또한 PWN을 기반으로 상이한 개별어 어휘의미망 구축이 여러 차례 시도되었고, 현재도 진행 중이다. 본 연구는 간접 구축 방식에 의한 어휘의미망 구축 시 요구되는 정제(refinement) 방식들을 검토하고, 이를 한국어 명사 어휘의미망(KL)에 적용하여 정확도 검증 방법의 한 대안으로 제시하였다. 또한 보다 정교한 정제 방법의 모색과 고찰은 향후 과제로 삼고자 한다.

#### 1. 서론

온톨로지(ontology)와 시소러스(thesaurus)는 자연어 이해를 위한 시스템 개발을 위해, 인간의 지식 베이스(knowledge base)와 전산학 방법론이 도입되어 체계적인 성과로 나타난 것이다. 이와 아울러 최근 주목받고 있는 어휘의미망(lexico-semantic network)은 자연어처리(NLP) 분야 중 그 접근과 분석이 상대적으로 난해한 의미 분석에 활용될 수 있다는 특징이 있다. 또한, 어휘의미망은 인간의 두뇌에 어휘 지식이 저장되는 구조인 심상어휘집(mental lexicon)에서 착안, 이와 유사한 구조로 의미 개념을 데이터 베이스화 한 결과물로, 어휘화된 개념들을 의미 계층관계로 구조화하였다. 본 연구는 프린스턴(Princeton) 대학의 WordNet(이하 PWN)기반으로 한국어 명사의 어휘의미망(KorLex, 이하 KL)을 구축하면서 발견되는 여러 유형의 문제점 해결을 위해, 자동정제와 수동정제를 위한 정제 준거를 제시하는데 그 목적이 있다. KL 구축 방법은 PWN 기반의 대역과정을 통한 간접구축 방식으로, 먼저 PWN의 표제어를 영한 이중어 기계 가독형(MRD)의 동종 이개국어 사전(homogeneous bilingual dictionary)을 활용, PWN 표제어와 한국어 간 1차 자동 사상(mapping)한다. 이어, 매칭된 대역어의 적합성 여부 및 정확성 검증을 위해 자동정제(automatic refinement) 및 전문가에 의한 수동정제(manual refinement)를 거친다. KL 구축 시 큰 어려움 중 하나는 PWN과의 의미 계층구조가 대부분의 경우 정확히 일치하지 않는다는 점이다. 본 연구의 목적은 다음과 같다. 첫째, 간접 구축 방식에 의한 어휘의미망 구축 시 정제 방법을 고안, 실제 검증하여 구축 결과물의 정확도를 높이는 것이다. 둘째, 보다 효과적인 자동 정제 방법의 개발 및 연구는 KL 뿐만 아니라 향후 그 구축이 완성될 개별어 어휘의미망에 보편

적으로 적용될 가능성이 크므로 연구가 갖는 의의는 크다.

본 논문은 다음과 같이 구성된다. 2장에서는 간접 구축 방식을 통한 어휘의미망 구축 시 정제 방법과 관련된 선행 연구들을 고찰하고, 본 연구의 대상 및 범위를 간단히 언급한다. 3.1.에서는 KL 구축 시 특별히 정제가 요구되는 문제가 되는 신셋 들을 유형화한다. 3.2. KL 구축 시 이용 가능한 정제 방법과 그 특성을 언급한다. 본 연구에서는 MultiWordNet에서 고안 개발한 정제 방법 17 가지를 간단히 살펴보고, 이중 활용도가 높으며 KL에 적용되기에 보편성이 있는 몇 가지를 실제로 적용하였으며, KL 구축 시 독자적으로 개발한 정제 방법도 추가하였다. 결론 및 향후 연구로, 보다 개발될 자동 정제 방법을 간단히 언급하였다.

#### 2. 관련 연구

##### 2.1. 연구 배경

PWN은 Pinceton 대학의 G. Miller를 중심으로 1985년부터 구축되기 시작한 어휘의미망으로 이를 기반으로 여러 개별어를 대상으로 한 어휘의미망 구축이 시도되었다. 그 대표적인 예로 PWN의 대역어 작업을 기초로 구축된 유로워드넷(EuroWordNet, 이하 EWN), 발카넷(Balkanet), MultiWordNet(이하 MWN) 등이 있다. EWN은 EU 8개국어인 영어, 이탈리아어, 스페인어, 네덜란드어, 독일어, 프랑스어, 에스토니아어, 체코어로 구성되었고, 발카넷은 불가리아어, 체코어, 그리스어, 루마니아어, 터키어, 세르비아어 등 6개 국어로 구성되었다. 이중 MWN은 간접구축에 의한 어휘의미망 구축 시 정제와 관련된 연구를 병행한 것에 주목할 만하다. MWN은 EWN 구축 project의

세부 연구로 현재도 진행 중이며, 이태리어 어휘의미망 1차 버전에는 28,000개 신셋과 37,000개 단어가 선정되었다. 특별히 MWN은 PWN 뿐만 아니라 다국어 워드넷과의 연계성 및 호환성이 우수하다는 평가를 받는다<sup>3)</sup>. 상이한 개별어로 서로 다른 어휘의미망이 구축될 경우 체계적이고 일관된 어휘망이 구성되기 어려운데, 이것은 언어간 이질성뿐만 아니라 상당 부분 상이한 구축 준거에서 비롯된다. 이 점에서 MWN은 PWN 구축의 준거와 기준을 일관되게 적용함으로써 PWN과의 연계성 및 호환성을 보장한다. 또한, MWN 구축 연구자들은 간접구축방식을 이용한 어휘망의 자동 정제방법<sup>4)</sup>에 관해 비교적 자세히 논의하였는데, 그 17 가지는 다음과 같다. PWN의 단의성, 다의성 여부와 PWN 표제어 개수, 스페인어 대역어 개수 등을 이용한 부류적 방법(8가지), PWN 표제어대 동일 스페인 대역어 관계를 이용한 구조적 방법(4가지), 개념간 거리(conceptual distance)를 이용한 방법(3가지)과 태그 등을 이용한 방법(2가지)이다. 본 연구에서는 KLK 구축에 필요한 정제 방법으로 MWN의 정제 방법도 일부 적용하여 검증하고자 한다.

## 2.2. 연구대상

PWN 대역을 통한 KL 구축에 앞서, PWN의 핵심적 의미계층구조를 검토하였다. PWN의 명사는 79,689 개의 신셋(synsets, 동의어집합, {}표시)으로 구성되었으며, 이들 중 무작위 샘플링(random sampling) 결과 영-한 대역 시문제점 유형은 크게 영-한 양국어간 개념의 부정합, 계층구조 불일치, 한국어 대역어 선정문제 등으로 도출된다.

① 영-한 양국어 간 개념 부정합<sup>5)</sup>: PWN 내 동의어 집합이 KL에 부재하거나 일부만 존재하는 경우

② 계층구조 불일치: PWN의 특정 신셋의 계층구조를 KL로 구축했을 경우 수정 또는 재구성성이 필요한 경우

③ 한국어 대역어 선정 문제:

1. 사전에 등재된 어휘로 대역어 불가능한 경우, 복합 명사 또는 수식 명사구 형태의 조어를 허용한다.

2. 위의 두 경우가 아닌 경우, 음차 또는 의미대역을 허용한다.

- 음차: 가령, 신셋 {Blighty}의 대역어로 {블라이트}를 선정

- 의미대역: 가령, 신셋 {loose trousers}의 대역어로 {보호용 땃바지}를 선정

## 3. KorLex 구축

KL은 PWN을 1차 자동 사상한 결과로 구축되지만, 이것이 완전하고 정교한 의미망이라고 할 수는 없다. 따라

- 3) [16]에서 Vossen은 MWN의 다국어간 호환성 및 연계성의 우수성을 주장한다.
- 4) 자동정제 방법에 관해서는 [12], [16]을 참고할 것.
- 5) 계층구조 부정합과 계층구조 불일치의 문제는 [3]에서 논의하였다.

서, 전체 신셋을 대상으로 한 정제가 반드시 요구된다.

## 3.1. 문제가 되는 신셋 유형

PWN의 1차 대역 결과물에 관한 검토는 다각도로 진행된다. 이것은 대역에 사용된 MRD가 지닌 의미구조의 체계성, 일관성과 직결되는 문제로, 정제 시 문제가 되는 신셋 유형들에 관한 검토가 선행되어야 한다. 문제가 되는 신셋의 유형은 영-한 의미장(semantic field)의 차이로 인한 어휘의미 규정의 문제, PWN 신셋의 의미파악 난해 또는 중의성, PWN 자체의 계층 구조오류, PWN의 상이한 신셋에 KL 동일 대역어 할당 등의 문제로 요약될 수 있다.

① 첫째, PWN 신셋 의미와 KL 대역어 의미 간 의미장의 차이의 문제가 제기되는데, 다음 세 가지 유형으로 요약된다.

a) PWN 신셋의 의미와 KL 대역어 의미가 부분적으로 일치(PWN 신셋 ∩ KL 대역어 의미)

b) PWN 신셋 의미가 KL 대역어 의미보다 큰 경우 (PWN 신셋 > KL 대역어 의미)

PWN의 {flowage}는 '범람하는 동작'과 '범람한 상태, 결과'까지를 모두 포함하는 의미이나, KL 대역어의 {범람}에는 '범람하는 동작' 의미만 들어있다. 이 경우 사전에 {범람}의 '상태, 결과'의 의미를 추가한다.

c) PWN 신셋 의미가 KL 대역어 의미보다 작은 경우 (PWN 신셋 < KL 대역어 의미)

PWN의 {virginity}는 'the condition or quality of being a virgin(미혼 여성의 상태)'의 의미로, 'virginity'의 자질을 갖는 명사 (virgin)은 독립된 신셋으로 분화되었다. 반면 KL의 {처녀}는 '미혼 여성의 상태'라는 자질과 그런 자질을 지닌 개인의 '처녀'까지 모두 포함한다.

## ② PWN 신셋의 정확한 의미파악 난해

PWN 신셋 자체의 의미 파악 난해 또는 중의성 문제는 부적합한 KL 표제어 선정으로 직결된다. PWN의 {pleaser}, {bombshell} 등은 사전이 추천하는 표제어만으로 적절한 의미가 전달되지 않으며, 의미 계층구조를 통한 상하위어, 자매 관계어 등을 참고로 적절한 의미 규정이 요구된다.

{pleaser}는 신셋의 의미가 지나치게 일반적이고, {bombshell}은 한국어 의미 개념 자체가 결여되어 있다. 앞에서 제시한 {loose trousers}는 정확한 의미파악에 충분한 정보를 제공하지 못한다. 이런 경우 전문가에 의한 수동 정제 및 적절한 대역어 선정이 요구된다.

## ③ PWN 자체의 구조오류

PWN의 {fibrous tissue}는 '동물 섬유조직'을 모두 포괄하며(lattice 구조), 계층구조상 '동물의 섬유조직', '식

```

{entity}
  {object} : physical object
  {living thing} : animate thing
  {organism} : being
    - {person} : individual : someone : somebody : mortal
    - {entertainer}
    * {pleaser}
    * {bombshell}
    
```

<그림 7> 의미 파악이 어려운 신셋, {pleaser}와 {bombshell} 물의 섬유 조직'의 상이한 두 개의 상위노드로 연결된다.

```

{entity}
  {thing}
    {part}
      {body part}
        {tissue}
          {animal tissue}
            {fibrous tissue}
  {entity}
    {object} : physical object
      {natural object}
        {plant part} : plant structure
          {plant tissue}
            {fibrous tissue}
    
```

<그림 8> PWN 자체 오류 신셋

1차 자동 사상에서 추천된 대역어는 {섬유조직}이나, 이것은 '식물'에만 국한된 어휘로, lattice 구조임을 감안, 정제시 {섬유조직}, {동물조직}을 표제어로 함께 넣어야 한다.

④ PWN의 상이한 신셋에 KL 동일 대역어 할당

```

{entity}
  {location}
    {region}
      {geographical area} : geographic area : geographical region : region
      {wilderness} : wild
      {waste} : waste : wasteland
      * {heath} : heathland
    
```

<그림 9> 동일 대역어가 할당된 신셋

PWN의 {barren, waste, wasteland}, {heath, heathland}, {badlands}에 추천된 대역어는 {황무지}인데,

KL의 동일 대역어가 상이한 노드에 동시에 할당된다. 따라서, 이들 대역어의 적합성 및 노드 간 상관성 여부는 수동 검증되어야 한다.

3.2. KL 정제 방법과 특성

앞에서 우리는 KL의 1차 사상 이후 정제의 필요성을 간단히 논의하였다. KL정제에는 자동 정제와 수동 정제 방법이 가능하다. 자동 정제 방법은 일관된 기준을 통한 정제로 비교적 객관성과 체계성을 유지할 수 있고 정제시간이 상대적으로 짧게 소요된다는 장점이 있다. 수동 정제는 해당 분야 전문가에 의한 정제 및 검증으로 보다 세분화된 의미 파악과 계층관계 설정이 가능하다는 장점이 있으나, 전문가의 검증에 자의성과 비일관성이 개입될 여지가 크고, 정제에 많은 시간이 요구된다는 단점을 피할 수 없다. 본 연구는 전체 신셋을 대상으로 자동 정제 및 수동 정제를 동시에 시행하였다.

3.2.1. 부류적 방법(class methods)<sup>6)</sup>

부류 방법은 PWN 표제어의 단일성(monosemic), 다의성(polysemic) 여부, PWN 표제어 개수, KL 대역어 개수 등을 이용하여 개별적인 신셋들의 좌표를 확인하고 각각의 대역어 매칭 관계의 적합성 여부를 확인한다. 본 연구에서는 대역어 전체 신셋 중 무작위 10%에 대해 검증하였다.

PWN 표제어 단일성/다의성 여부	유형	PWN 표제어 개수	KL 대역어 개수
monosemic	①	1	1
	②	1	N (N>1)
	③	M (M >1)	1
	④	M	N
polysemic	⑤	1	1
	⑥	1	N (N>1)
	⑦	M	1
	⑧	M	N

<표1. 부류적 방법의 기준>

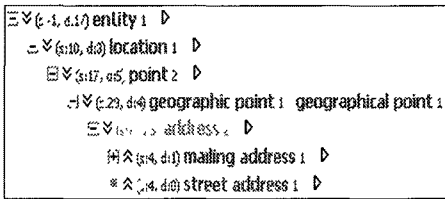
가령, PWN의 신체 명사(noun.body) 중 {cremains}는 단일어로, 추천된 KL 대역어는 1개, {유골}이다(① 유형). 또한 {colon}은 다의어이며, PWN 표제어는 1개, 추천된 KL 대역어는 {결장}, {장록 창자}의 2개 이다(⑥ 유형). 이 방법은 전체 신셋의 좌표를 확인, KL 대역어와의 관계를 검증하는 가장 기본적인 검증 방법에 속한다.

6) 부류적 방법과 구조적 방법은 가장 기본적인 정제 방법으로 간주되며, MWN의 연구를 참고하지 않더라도 고안할 수 있는 보편적인 방법이다.

3.2.2. 구조적 방법(structural methods)

구조적 정제 방법은 PWN 신셋에 대해 중복적으로 출현하는 KL의 동일 대역어를 추출, 이들의 적합성 여부를 검증한다. PWN의 상이한 노드에 대해 KL 동일 대역어가 매칭된 경우는 크게 부모-자식(parent-child) 노드, 형제-자매(brother-sister) 노드, 사촌 노드(distant hyperonymy), 기타 노드 등에서 발견된다(2번 이상 출현 가능).

a) 부모-자식 노드 간 동일 대역어 출현



<그림 10> 부모-자식 노드 간 동일 대역어

신셋 {address}와 {street address}은 KL의 동일 대역어 {주소}로 매칭된다. 이 둘은 부모-자식 노드에 위치하는데, 이 경우 의미상 뚜렷한 차이가 없는 경우 동일 대역어가 확정되며, 노드 병합이 일어난다.

b) 형제-자매 노드 간 동일 대역어

신셋 {sheikdom, sheikhdom}과 {emirate}은 KL의 동일 대역어 {토후국}으로 매칭된다. 이 둘은 형제-자매 노드에 위치하는데, 노드 병합이 가능하다.

c) 사촌 노드 간 동일 대역어

{mishap}, {accident}, {calamity}와 {trial}은 인접 노드에 위치하고, KL의 동일 대역어 {재난}으로 매칭된다. 이중 {accident}과 {trial}은 사촌 노드에 위치한다.

d) 기타 관계에 동일 대역어

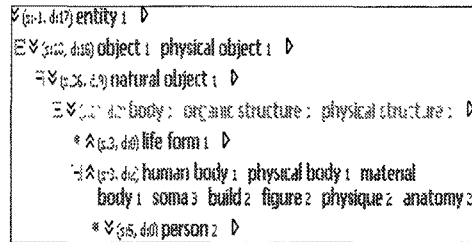
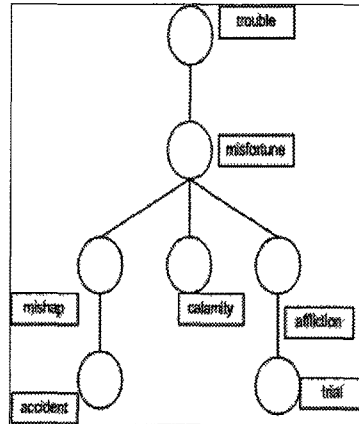
d)-1. 조부모-손자 노드 간 동일 대역어

신셋 {body, organic structure, physical structure}과 {person}은 KL의 동일 대역어 {몸}으로 매칭된다. 이 둘은 조부모-손자 노드 간 노드에 위치한다.

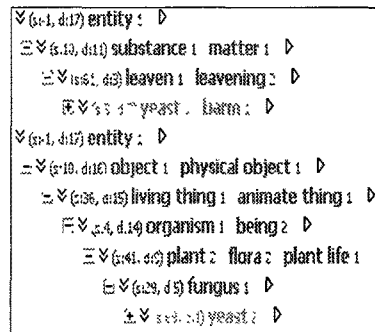
<그림 5> 사촌 노드 간 동일 대역어

<그림 6> 조부모-손자 노드 간 동일 대역어

d)-2. 상이한 계보에 동일 대역어



신셋 {yeast}와 {leaven, leavening}은 KL의 동일 대역어 {효모}로 매칭된다. 이 둘은 서로 다른 계보에서 내려온 노드에 위치한다. 이상과 같이 상이한 PWN 신셋에 대해 KL의 동일 대역어가 할당한 경우를 추출, 이들에 관해 검증한다.



<그림 7> 상이한 계보 노드 간 동일 대역어

3.2.3. 개념간 거리 이용법(conceptual distance methods, 이하 CD)

개념간 거리는 어휘들 간 의미의 밀접성(closeness in meaning)을 결정하는 지표가 되는 것으로, “계층적 의미망에서 개념들을 연결하는 최단 경로”로 정의 된다<sup>7)</sup>. 두 개

넙간 거리는 다음 공식으로 산출된다.

$$Dist(w1, w2) = \min \sum \frac{1}{depth(Ck)}$$

$$C1i \in W1 \quad Ck \in$$

$$C2j \in W2 \quad path(C1i, C2j)$$

<표1> 개념간 거리 측정법

CD를 구하려면, 먼저 비교의 대상이 되는 두 어휘가 정해져야 한다. 첫째, 사전의 동일 정의문 내에 사용된 두 어휘를 비교하여 이들 간 CD를 구할 수 있다. {혼란}의 사전적 정의는 '뒤죽박죽이 되어 어지럽고 질서가 없음'으로, 정의문에 사용된 '뒤죽박죽'과 '질서'가 각각 비교될 수 있다. 둘째, 사전의 동일 정의문내에 사용된 핵심어와 종류(genus) 관련 어휘 간 비교이다. {학생}은 '학교에 다니면서 공부하는 사람'인데, 핵심어인 '공부'와 '사람'의 CD를 구할 수 있다. 셋째, 이중어 사전에서 한영 번역(한국어=> 영어)이 여럿인 경우 이들 번역의 중심어들을 서로 비교할 수 있다.

3.2.4. 기 구축된 전문 용어 사전과 KL 표제어 매칭

PWN에는 전문용어(컴퓨터, 기계, 전자, 의학 등 과학 전분야)와 비전문용어가 별 다른 구분 없이 사용되고 있으나, 전문용어 시소러스와 KL 대역어간 매칭관계 검증도 자동정제의 한 방법이 될 수 있다. 그 첫 번째 방법은, PWN의 KL 대역어 전체를 기 구축된 전문용어 시소러스와 미칭시켜 그 일치 관계를 확인한다. 둘째, PWN 표제어 내에 대문자가 포함되고, 최하위 노드(terminal node)에 출현하는 어휘를 대상으로 전문용어와의 매칭 관계를 확인할 수 있다. 전문용어의 상당 수가 세분화된 의미로 어휘의미망에서 최하위 노드에 위치할 가능성이 높으며, 대문자로 시작하거나 포함될 확률이 높기 때문이다.

3.2.5. PWN 명사 중 타품사와 파생 관계(derivation)를 맺는 신셋 들의 표제어 비교

이 방법은 PWN 명사 중 타품사와의 파생관계가 있는 신셋 들을 대상으로 이들의 의미적 상관관계를 확인하고, KL 대역어 매칭의 적합성 여부를 판단하는 방법이다. 동사 {irrupt, intrude}와 파생 관계에 있는 {irruption}은 {난입}으로 대역될 수 있다. 이 경우 {쳐들어오다}와 {난입}은 형태적 관련성이 없지만, 이런 유형의 신셋들을 추출, 의미적 상관성을 검토하여 KL 대역어가 적절히 매칭되었는가 확인할 수 있다.

4. 결론 및 향후 연구

7) CD를 구하는 방법은 [7]을 참고할 것.

본 본문에서는 간접 구축 방식에 의한 어휘의미망 구축 시 반드시 요구되는 정제 방법에 관해 논의하였다. 먼저, 이를 위해 PWN 전체의 의미 계층구조를 검토, 분석하고 무작위 샘플링을 통해 KL로 구축할 경우 예상되는 문제점들을 유형화 하였다. 보다 정교한 어휘망 구축을 위해 정제가 요구되는데, EWN 구축 project의 일환으로 시작 된 MWN은 정제의 방법론적 측면에서 주목할 만하다. MWN이 제시한 17 가지 방법은 현재도 보충, 심화되고 있으며 인구어(Indo-European Language)가 아닌 개별어에 적용될 수 있는 보편성을 갖고 있다.

본 연구에서는 KL 구축 시 적합한 정제 방법 모색을 논의하여 이를 적용하여 실제 데이터를 검증하였다. 첫째, 부류적 방법과 구조적 방법은 가장 기본적인 정제 방법으로, MWN의 연구를 참고하지 않더라도 고안할 수 있는 보편적인 방법이다. 부류적 방법은 출발어의 단의성과 다의성, 대역어 개수 등을 지표로 대역어 전체 데이터의 좌표를 정확히 파악할 수 있는 장점이 있다. 다만, 구축된 어휘망의 추가 및 확장이 있는 경우 전체 데이터의 좌표는 유동적이어서 이를 감안한 검증이 요구된다. 한편, 구조적 방법은 대역어 어휘망에서 동일 대역어가 출현하는 구조 및 의미를 검토, 검증한다. 셋째, 개념간 거리 측정법은 통계에 기반한 검증법으로 체계적이고 일관된 MRD의 활용이 수반되어야 한다. 넷째, 기 구축된 전문 용어 시소러스의 활용은 KL 구축 검증에 역시 활용될 수 있고, 타품사와의 매칭관계 또한 검증의 한 방법으로 활용가능하다. 특히 후자는 복잡한 변수가 많이 작용하므로 추후 방법론 연구의 대상으로 삼고자 한다.

Acknowledgements

본 연구는 한국과학재단 국가지정연구실 사업 M10203000028 -02J0000-01510 지원으로 수행되었음.

참고문헌

[1] 문유진 (1996), 의미론적 어휘 개념에 기반한 한국어 명사 Wordnet의 설계와 초록, 서울대학교 컴퓨터 공학과 박사 논문.  
 [2] 이은령, 윤애선, 권혁철 (2004), "워드넷(WordNet 2. 0) 커뮤니티케이션 명사의 결어 의미관계", 한국인지과학회 춘계학술대회, 한국인지과학회, pp. 177-183.  
 [3] 이은령, 황순희, 윤애선 (2004), "다국어 어휘의미망 구축의 현황과 문제점-PWN과 EWN을 중심으로", 프랑스문화예술연구회 제12집(6권3호), 프랑스 문화예술학회, pp.369-401.  
 [4] 이창기, 이근배 (1999), "Wordnet을 이용한 한국어 시소러스 자동 구축", 제11회 한글 및 한국정보처리학술대회, 한국정보과학회, pp. 156-161.  
 [5] 최호섭·옥철영 (2002), "한국어 의미망 구축과 활용", 한국어학 17, pp. 301-329, 서울, 한국어학회.  
 [6] 황순희, 이은령, 윤애선 (2004), "불한 전문텍스트 번역에서 의미해석의 문제", 프랑스어문교육 제18집, 한국프랑스어문교육학회, pp. 315-334.

- [7] Atserias, J. et al. (1997), "Combining multiple methods for the automatic construction of multilingual WordNets." In: Proceedings of the International Conference on Recent Advances in Natural Language Processing, Tzigov Chark, <http://xxx.lanl.gov/abs/cmp-lg/9709003>
- [8] Bilgin, O et al. (2003), "Morphosemantic Relations In and Across Wordnets", Global Wordnet Conference 2004 Proceedings, Brno, pp.60-66.
- [9] Bentivogli, L., Bocco, A. and Pianta, E. (2003), "ArchWordNet: Intergrating Wordnet with domain-specific knowledge", Global Wordnet Conference 2004 Proceedings, Brno, pp. 34-46.
- [10] Choi, K-S and Bae, H-S (2003), "Procedures and Problems in Korean-Chinese-Japanese Wordnet with Shared Semantic Hierarchy", in Global Wordnet Conference 2003 Proceedings, Brno, pp. 91-96.
- [11] Clough, P. and Stevenson, M. (2003), "Evaluating the contribution of Eurowordnet and word sense disambiguation to cross-language information retrieval", Global Wordnet Conference 2003 Proceedings, Brno, pp. 97-105.
- [12] Farreres, H. et al. (2002), Semiautomatic creation of taxonomies, In: Proceedings of the Coling 2002 Workshop "SemaNet'02: Building and Using Semantic Networks", Taipei, August 2002.  
<http://www.cs.ust.hk/~htc/semanet02/pdf/farreres.pdf>
- [13] Fellbaum, C. (eds.) (1998), WordNet - An Electronic Lexical Database, Cambridge, MIT Press.
- [14] Gomez, F. (2003), "Grounding the ontology on the semantic interpretation algorithm", in Proceedings of the second international WordNet conference 2004, Brno, pp. 124-129.
- [15] Miller, G. et al., (1990), "Introduction to WordNet: an on-line lexical database." International Journal of Lexicography 3 (4), pp.235-244.  
<ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>.
- [16] Pianta, E. et al, (2002), "MultiWordNet : Developing an aligned multilingual database", Proceedings of the 1st International WordNet Conference, pp.293-302.
- [17] Smrž, P. (2004), "Quality Control for WordNet Development", Global Wordnet Conference 2004 Proceedings, Brno, pp.206-212.
- [18] Vossen, P. (1998), EuroWordNet : A multilingual database with lexical semantic networks. Dordrecht, Kluwer Academic Press [유로워드넷. (한정한 외 공역), 서울, 한국 문화사.]
- [20] Vossen, P. (2004), EuroWordNet General Document, <http://www.hum.uva.nl/~ewn>.