

피동 정보를 통한 한국어 동사 어휘의미망 정제

이은령, 윤애선

부산대학교 한국어정보처리연구실

Abstract

To build a Korean wordnet, we translated semi-automatically the English wordnet PWN into Korean verbs. During this process, we found that some of translation errors are related to the arbitrariness of PWN's sense distinction in regard to the accusativity/inaccusativity of the same verb form in English. This study presents an empirically based method of remodeling the PWN for Korean wordnet and while revising the PWN's hierarchical structure, it shows the necessity of classifying the Korean passive verbs as semantically autonomous.

1. 서론

본 연구는 한국어 동사 어휘의미망 구축을 통해 얻어진 경험적 자료를 토대로 영어 워드넷(PWN)¹* 동사의 어휘의미 분류뿐만 아니라 다국어 어휘의미망 구축에 있어서 동사 어휘의 통사 및 의미속성에 따른 '의미 세분화(sense distinction)' 방법을 제시하는데 그 목적이 있다. 한국어 어휘의미망을 구축하는 한 방법으로 영어 워드넷의 동사 어휘의미망을 반 자동으로 한국어 동사로 대역하는 경우, 영어 동사의 특성 때문에 영어를 위한 어휘의미의 분류 방법이 목적언어(한국어)의 어휘의미망에서 그대로 적용될 수 없는 경우를 발견할 수 있다. 본 연구에서는 PWN을 근간으로 구축되는 한국어 동사 어휘의미망에서 한국어 대역어가 타동과 피동의 형태로 나타나는 경우를 분석하여 (1) 대역의 정제 기준과 오류의 수정, 그리고 (2)상이한 의미적 속성을 지닌 어휘의미를 의미망 내에서 각각 개별적인 노드로 분류해야 할 필요성을 제시하고자 한다. 한국어에서도 특히 <명사+하다>류의 동사에서 쉽게 찾아볼 수 있듯이 자동사와 타동사가 동일한 형태를 띠는 경우나 형태가 달라지는 경우 모두 어휘의미를 기술하는데 반영되어야 할 것이다. 특히 사동형이나 피동형이 형태소의 교착을 통한 파생이라는 점은 사동과 피동의 문제가 통사 단계에서의 기술보다는 어휘의 형태와 의미 단계에서 다루어져야 함을 보여주고 있다. 이와 같이 한국어에서 고려해야 하는 어휘의 형

¹ 프린스턴 대학의 심리학자 G Miller를 주축으로 만들어진 언어 자료로서 인간의 두뇌에 저장된 어휘 지식의 관계 구조를 모델링한 어휘의미망이다 1980년대 초기에 개발하기 시작하여 2005년 현재 21 버전까지 공개되었다 명사 어휘 79,689개, 동사 13,508개, 형용사 18,563개, 부사 3,664개의 동의어집합으로 구성되어 있다.

* 본 연구에서 PWN은 Princeton Wordnet을 말하며 wordnet이 각각 개별 언어의 어휘의미망을 지칭할 수 있기 때문에 프린스턴의 영어 워드넷을 특별히 구분하여 PWN이라고 부른다.

태적 특성에 따라 어휘의미의 분류 기준을 세우고 또한 한국어와 영어 및 다른 언어의 동사 어휘의미망 구축에 있어 유효한 분류 기준의 언어학적 근거를 마련하는 것은 한국어 어휘의미망 구축에 필수적인 요소이다.

본 논문의 구성으로 다음과 같다. 2장에서는 PWN의 자동과 타동의 이중적 통사 행태를 보이는 어휘 의미에 대한 분류 기준을 검토하고 그 특징과 문제점을 분석한다. 3장에서는 한국어 피동사가 형태, 통사적 특성에 의하여 하나의 독립된 의미로 분류되어야 할 언어학적 근거를 논의할 것이다. 4장에서는 영어 동사가 자동/타동의 의미를 구분하지 않을 때 한국어 대역어에서 나타난 유형과 오류를 중심으로 어휘의미의 구분과 어휘의미의 노드 위치 수정에 대한 해결방안을 제시하고자 한다.

2. PWN의 동사 어휘의미 분류

Princeton 대학에서 만들어진 대용량 어휘 및 지식 자료인 영어 워드넷(PWN)은 인지심리학의 심상어휘집(lexicon)에 대한 연구와 자연어의 전산적 처리에 의미의 이해의 필요성이 맞물려 구축된 어휘 지식의 네트워크이다. PWN은 품사별(명사, 동사, 형용사, 부사)로 어휘 의미를 구별하고 이를 계층적인 망으로 구성하여 어휘 의미간의 관계로 조직화 시켜 놓은 것이다. 어휘 의미를 관계라는 망으로 조직함으로써 PWN은 종래의 자연어 처리에서 해결할 수 없었던 구문 분석의 한계²를 극복할 수 있는 의미분석을 위한 자료이자 실제 검색과 교육 등의 다양하고 복잡한 기능을 수행해 낼 수 있는 언어자료이다. 또한 단일 언어의 어휘의미망인 PWN의 망구조는 다국어로의 확대를 용이하게 하여 유로워드넷이 구축될 수 있었는데 이는 다국어 교차 번역과 교차 언어간 정보 검색을 보조하는 언어자원으로 효율성을 인정받고 있다. 이렇게 자연어 처리 분야 및 여러 다양한 분야에서 어휘의미망과 나아가 다국어 어휘의미망에 대한 필요성이 대두되는 현 상황에서 현재 다목적 언어 처리를 위하여 한국어 어휘의미망(KorLex)이 구축 개발되고 있다. 이를 위하여 우리는 PWN 2.0을 반 자동 번역하여 명사 3만, 동사 5천 어휘의미에 대한 한국어 어휘의미망의 1차 버전의 완성 단계에 있다.

PWN에서는 동의관계에 있는 어휘들을 묶은 고유한 의미의 집합인 ‘동의어집합(synset)’이 기본 구성 요소이다. 다음은 PWN의 ‘freeze’라는 동사의 의미 10개 중 그 첫 번째 의미와 이 동의어집합이 갖는 여러 의미관계 및 상·하위 계층구조를 보여주고 있다.

change 1 (undergo a change, become different in essence, losing one's or its original nature
 ⇒ change state 1, turn 4 (undergo a transformation or a change of position or action)
 ⇒ freeze 1 (change to ice)

그림 1. PWN2.0 브라우저에 검색된 {freeze1}의 위계관계

² 종래의 어휘 사전으로 자연어의 구문 분석을 수행할 때 어휘가 갖는 여러 의미적 제약 조건을 반영할 수 없었으므로 통사적인 분석 만으로는 컴퓨터가 문장의 의미를 정확하게 파악해 낼 수 없는 한계를 갖고 있다

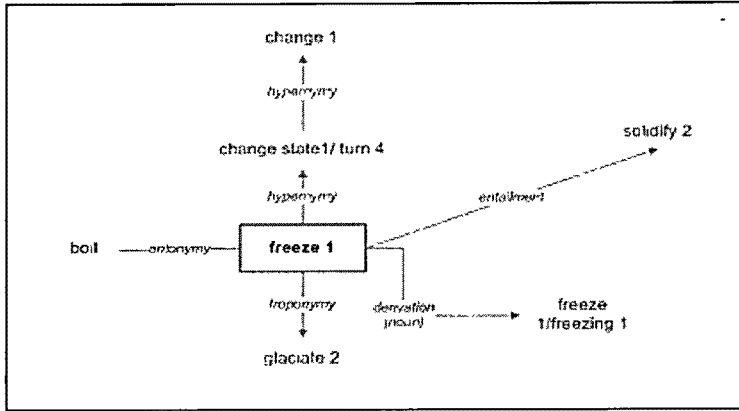


그림 2. PWN 동사 의미관계 구조

동의어집합의 의미에 대한 주석은 그림1에서처럼 원괄호 안에 뜻풀이와 예문으로 주어진다. 동사 동의어집합들은 상호간 상위어(hypernym), 하위어(troponym), 함의(entailment), 인과관계(cause), 반의어(antonym) 등의 의미관계를 갖는다. 또한 동의어집합을 구성하는 각 동사는 동의어집합의 의미에 해당하는 문형정보를 담고 있기 때문에 문형정보는 동의어집합의 의미와 그 구성 동사 어휘에 동시 의존적이다. 이는 동의관계에 있는 모든 동사가 반드시 동일한 통사적 행태를 갖지 않기 때문에 어휘에 개별적인 정보로 구성해 둔 것이었다.

Sense 11
 plow, plough, turn -- (to break and turn over earth especially with a plow, "Farmer Jones plowed his east field last week", "turn the earth in the Spring")
 *> Somebody ----s
 *> Somebody ----s something

그림 3. {plow, plough, turn}의 문형정보

또한 그림3에서와 같이 PWN의 문형정보는 동사가 갖는 통사구조를 나타내고 있고, 제한적이기는 하나 동사가 결합하는 명사(구) 논항의 의미자질을 somebody/something이라는 매우 기본적인 선택제약정보를 통하여 나타내고 있다. 비록 동사의 통사적 자질에 대한 정보의 완성도는 일반 언어학의 기술보다는 취약한 상태이지만 문형정보는 영어의 자동과 타동을 크게 대격(accusative)과 비대격(inaccusative)³이라는 논항 구조를 나타내는데 충분하다. 또한, 문형정보가 동사의 통사적 행태를 나타내기 때문에 동사의 의미를 파악하는 정확한 근거가 되기도 한다. 그러나 대격/비대격의 격전환(transitivity alternation⁴)을 하는 동사에 대한 의미분류의 일관성이 유지되지 않는 부분이 적지 않게 발견된다. 특히 대격과 비대격의 구조로 실

³ 일반적으로 영어 동사에서 자동(intransitivity)은 비대격과 능격(inchoative)의 용법으로, 타동(transitivity)은 대격과 사역의 용법으로 구분한다. 그러나 워드넷 동사의 문형으로 사역과 능격을 추론하기에는 부족하다. 그러나 뜻풀이 정보와 예문을 통해서 사역의 용법은 경우에 따라 추론이 가능하다.

⁴ 저자에 따라 'diathesis alternation', 'frame-alternation', 'transitivity alternation' 등의 용어를 쓴다.

현되는 동사를 구별하는 것은 자연어 문장 분석의 결과에 영향을 미치기 때문에 반드시 고려되어야 할 대상이다.

2.1 선행 연구

워드넷 동사의 의미 분류에 대한 연구는 동사의 통사적 속성과 어휘 의미의 상관관계의 문제가 주로 연구되었다. 그 이유는 크게 동사가 자연어 처리에서 필요한 문장 분석에서 중심 역할을 하기 때문이라고 볼 수 있다. 따라서 PWN의 동사의 어휘 의미 기술이나 분류보다는 통사적 정보에 대한 연구와 비판이 지배적이다. 특히 B. Levin의 Verb Class를 중심으로 영어 동사의 격 구조 및 격 구조 변환을 하는 동사의 기술에 대해 PWN동사 어휘 분류의 문제를 지적한다[Kohl et al. 1998, Zickus 1994].

Levin[1993]은 영어 동사에 대해서 유사한 격 교체 현상을 보이는 동사들을 그룹화하였다. 즉, 동일한 어휘 형태가 다양한 격 구조로 실현되는 경우 이는 어휘 의미의 기술에도 반영되어야 함을 보여 주었다고 할 수 있다. 예를 들어 ‘He broke the glass plate’와 ‘The glass plate broke’에서 동사 break는 각각 대격과 비대격으로 쓰였다. 여기서 Levin은 통사적 패턴 변화가 어휘 의미의 분류를 동반한다고 주장한다. 또한 격변화는 동사의 여러 다른 의미를 반영하고 있다고 하는 Jackendoff[1990]의 논의 역시 이러한 Levin의 동사 분류의 당위성을 예견하였다. 한편, 격 전환을 하는 하나의 동사는 논리적으로 다의적이라고 주장하는 Pustejovsky and Busa[1995]의 논의는 동일한 유형의 격 전환을 하는 동사간의 의미적 관련성을 설명하지는 못하지만 Levin의 동사 분류가 동사의 어휘적 중의성을 해소하는데 하나의 해결책으로 제시될 수 있음을 뒷받침하고 있다. 또한 Levin은 대격-비대격의 전환뿐만 아니라 여러 다양한 격 교체 현상을 보이는 동사들을 그룹화하여 통사적 행태를 동사의 의미기술에 직접적으로 적용한 VerbNet과 FrameNet 연구 개발에 폭넓게 적용되고 있다. 특히, Kohl et al[1998]에서는 EVCA 프로젝트에서 Levin의 Verb Class를 PWN동사에 사상(mapping)하였다. 이는 미시적인 어휘의 통사적 행태를 명세화하여 유형화 시킨 것을 동사의 의미 분류를 위해 적용하려는 시도로 평가된다. EVCA 프로젝트에서 수행한 PWN동사와의 매핑은 PWN 동의어집합을 구성하는 동사 어휘의 통사적 정보를 보충하고 따라서 전산적 처리를 효율적으로 만들고 PWN동사 어휘 의미 분류의 정합성과 데이터의 빈약함(sparseness) 정도를 비교적 정확히 평가할 수 있는 방법으로 적용될 수도 있다[Zickus 1994]. 따라서, 어휘 의미 기술을 위해 어휘가 갖는 통사적 속성과 의미적 특성의 관계를 보다 상세하게 구조화할 필요가 있다. 격 구조의 변화가 문제의 어휘 의미 변화를 수반한다면 어휘의미망에는 어휘 의미 구분을 야기하는 최소의 효율적인 통사적 정보가 활용되어야 할 것이다.

2.2 영어 워드넷의 대격-비대격 전환 동사의 의미 분류

영어 워드넷에서 대격과 비대격 교체를 하는 동사는 주로 verb.change⁵에 집중되어 있는

⁵ PWN에서는 동사의 의미 범주를 분야(domain) 별로 15개로 분류하여 각각 의미 부류에 맞는 사전편찬파일명

데 한국어로의 대역과정에서 이들 동사에 대한 비일관적인 의미 분류가 발견된다.

영어 워드넷과 같이 방대한 양의 어휘데이터의 수동구축과정에서는 오류와 일관성이 유지되지 못한 부분이 있게 마련이며 Fellbaum[1998]이 밝힌 바와 같이 ‘워드넷 동사 어휘는 어휘의미망이라는 계층구조에 더 치중하여 조직되어 있기 때문에 통사적 속성을 상세하게 반영하지 못하는 경우도 있다’⁶. 그러나 verb.change에서 발견할 수 있는 대격-비대격 교체 동사의 어휘 의미의 분류의 오류는 간과할 수 없다. 다음의 표에서는 PWN의 문형과 연관된 의미 분류의 방법을 3가지 유형으로 보여준다.

분류 유형	동의어집합	뜻풀이	문형	최 상위어	한국어 대역어
type I	dry1, dry out5	remove the moisture from and make dry	8_0(Somebody ---s something) 11_0(Something ---s something)	change1, alter1, modify11	말리다, 건조하다
	dry0, dry out0	become dry or drier	1_0(Something ---s)	change0	마르다, 건조되다
type II	powdenze1, pulverize1, pulvense1, powderse1	become powder or dust	(Something ---s)	change 0	가루가 되다
	Powdenze0, powderse0, pulvense0, pulvise 0	make into a powder by breaking up or cause to become dust	8_0(Somebody ---s something) 11_0(Something ---s something)		빻다, 가루로 만들다
type III	grill0	cook over a grill	1_0(Something ---s) 8_0(Somebody ---s something)	change 0	굽다, 굽히다

표 1 PWN의 문형정보에 따른 의미분류의 유형

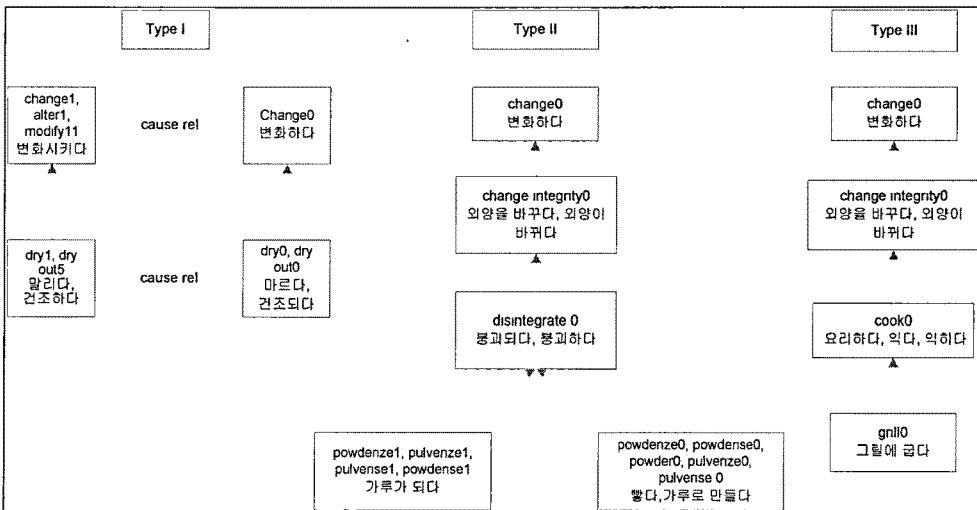


그림 4 표3의 동의어집합의 상위구조 (→ 상위어)

(lexicographer file name)을 붙이고 있다. 예를 들어 ‘변화’, ‘변형’ 등의 의미를 지니는 동사는 verb change에 분류되어 있다. 다음은 PWN에서 사용된 사전편찬명의 목록이다, verb body, verb change, verb cognition, verb communication, verb competition, verb consumption, verb contact, verb creation, verb emotion, verb motion, verb perception, verb possession, verb social, verb stative, verb weather

⁶ Fellbaum (1998), p.95

우선 type I에서 나타나는 것은 대격과 비대격의 짝을 이루는 동사의 동의어집합으로서 어휘 ‘dry, dry out’의 각각 상이한 두 개의 어휘 의미를 각각의 동의어집합으로 분류하여 개별적인 의미의 계층을 이루고 있다. Type II에서 나타나는 유형은 ‘powderise’동사와 동의어에 대한 대격과 비대격으로의 쓰임을 각각 개별적인 synset으로 구분해 놓았으나 이들은 공통의 상위어 및 최 상위어를 갖고 있고, 이 경우 최상위 노드의 개념은 비대격의 쓰임만을 갖는 동사의 집합이다. Type III의 경우는 ‘grill0’와 같이 대격과 비대격의 쓰임과 의미가 하나의 동의어집합에 공존하는 경우이다. 이는 보통 ‘make or become ___’의 뜻풀이로 이루어져 이중적 의미를 나타낸다. 여기서 verb.change(전체의 17%)의 경우만 예를 들어 보더라도, 대격-비대격의 통사적 속성과 함께 의미 구분이 이루어진 동사가 140개의 synset으로 나타나고, 대격과 비대격이라는 통사적 이질성에도 불구하고 동의어집합으로 의미를 따로 구분되지 않은 것이 117개가 있다.

2.3 대격과 비대격 동사의 비일관적 계층구조

그림 4에서 나타난 계층적 구조 관계에서 살펴보면, 우선 {dry0, dry out5}는 {change1, alter1, modify11}를 상위어로, {dry0, dry out0}는 {change 0}를 상위어로 두고 있다. 비대격의 쓰임을 갖는 동사의 동의어집합은 상위어 또한 비대격 동사의 집합을 갖고 있고, {dry0, dry out0}는 행위주(agent)가 의지를 갖고 대상에 대해 어떠한 행동을 하는 것으로 상위어가 설정되어 있다. 반면 type II와 III는 대격-비대격의 의미가 모두 비대격 동사의 동의어집합인 {change 0}를 상위어로 갖는다. 우리는 여기서 II와 III의 분류 방법에 일관성이 다소 떨어지고 있음을 발견할 수 있다.

우선 상위어에서 하위어로 의미적 속성은 승계(inheritance)되지만 통사 및 문법적 속성은 승계되지 않는다. 즉, 통사적 속성이 의미에 관련이 되어 있기는 하지만 이는 어휘에 개별적인 영향을 미치는 것으로서 synset 전체에 일괄적으로 적용이 되지 않는 것이다. 따라서 II와 III에서 최상위어의 통사적 행태는 하위어의 의미에 직접적인 영향을 주지 않는 것처럼 보인다. 그러나 타동성(transitivity)은 문법적 자질을 말하고 있지만 사실상 동사 lexicon의 의미에 전혀 영향을 주지 않는 것은 아니며 의미 속성과 관계된다. 따라서 우리는 이러한 비일관적 의미분류를 나누어 특히 II의 오류를 계층구조의 오류, III을 의미 세분화의 오류라고 부른다.)

2.4 대격-비대격 교체 동사의 인과관계 표시

위의 표1에서 도출된 유형 I은 자동과 타동의 의미를 각각 개별적인 의미로 분류하고 있다. 워드넷에서는 하나의 동형태 단어가 대격과 비대격으로 분리된 경우 이를 인과관계로 연결하고 있다. 워드넷의 인과관계는 타동에서 자동으로의 관계뿐만 아니라 사역(causative)에서 비대격 동사들을 의미적으로 연결하기도 한다. 대부분의 인과관계는 verb.change에서 나타나는데 이러한 관계로 연결된 동의어집합의 쌍은 70개뿐이다. 이는 비일관적인 의미세

분화가 가장 큰 이유로 보이며 향후 PWN에서도 표3에서 밝힌 유형 II와 III에 대한 동사 어휘의미의 세분화를 수행할 경우 인과 관계로 연결되는 동의어집합은 늘어날 것이다.

그러나 워드넷의 인과관계는 표1에서 나타난 유형 II와 III의 동의어집합의 경우에 대해서는 대격-비대격 동사간 인과관계를 설정해 놓지 않고 있다. 예를 들어 유형I에 분류된 {dry0, dry out5}와 {dry0, dry out0}사이에는 인과관계가 성립되어 있는 반면 유형 II에 속하는 {powderise 0}와 {powderise 1}과는 인과관계가 설정되어 있지 않다. 그러나 이러한 인과관계는 대격과 비대격, 사역과 비사역의 동사가 영어에서도 각각 독립적인 어휘 의미로 분류되어야 함을 반증하고 있다.

3. 피동사의 어휘통사적 특성과 대역어

2장에서 논의한 대격-비대격 동사 동의어집합에 대한 한국어 대역어를 찾아주는 과정에서 문제가 발견되었다. 워드넷이 동의어집합으로 설정한 의미가 유형 III과 같이 대격-비대격전환을 하는 동사로 구성된 경우 한국어에서는 대부분의 경우 (1)자동사와 타동사⁷ 또는 (2) 피동사 또는 피동형태(passive form)와 타동사로 대역되기 때문에 한국어 동사 어휘의미망을 구축하는데 있어 워드넷과는 다른 새로운 기준을 적용해야 할 필요성을 보이고 있다. 한국어에서는 자동, 타동, 사동, 피동사가 형태적 또는 통사적으로 다르기 때문에 상이한 의미자질을 갖기 때문이다. 따라서 유형 II에서 고찰된 계층구조의 오류에 대해 동사의 의미자질에 입각하여 한국어에서는 상위 구조를 수정할 필요가 있다.

3.1 피동사의 의미 분류의 필요성

우리는 여기서 한국어의 피동에 대한 선행 연구를 통하여 피동사를 어휘의미적으로 독립된 개체로 간주해야 하며 이를 영어와는 달리⁸ 한국어 동사 어휘의미망에 반영해야 할 필요성을 지적하고자 한다.

한국어에서 동사의 피동사는 능동형 동사 어휘에 교착된 접미사에 의해 실현되는데, 최현배[1971(1937)]에서는 다음의 세가지 유형의 형태소 교착으로 보았다. 우선 (1)접미사 ‘이, 히, 리, 기’ 등의 형태소에 의한 실현과 (2) <명사+하다>의 동사에 해당하는 ‘되’, ‘받’, ‘당하’를 통한 피동, 그리고 (3) 보조동사 ‘지다’에 의한 것이다.⁹ 이러한 형태소에 교착을 통한 피동사 생성은 매우 생산적이며 타동사는 원칙적으로 접미사 파생에 의해 어휘부에서 피동의 의미를 가질 수 있게 된다.

⁷ 본 연구에서는 한국어 동사에 대해서는 영어의 경우와 같이 대격-비대격이라는 용어를 쓰지 않기로 한다 이는 영어권 또는 유럽 언어권 언어학에서 의미하는 대격-비대격에 대한 정의가 한국어 동사의 논항적 특성을 기술하는 용어로는 적당하지 않다고 보기 때문이다. 우리는 목적어를 취하는가 취하지 않는가의 동사의 통사적 특성을 크게 타동사와 자동사라는 용어로 구분하기로 한다

⁸ 영어의 수동태는 통사의존적이며 ‘be+과거분사’의 형태로 동사가 변화한다고 하더라도 어휘의미망에서 이를 어휘의미적으로 독립된 개체로 구성할 수 없다.

⁹ 일반적으로 좁은 의미의 피동법은 (1)의 유형만을 언급하나 이정택[2001]에 따르면 현대 한국어에서 (2)의 유형을 가진 피동사가 가장 빈도가 높다 본 연구에서 피동사는 (1), (2), (3)의 유형에 의한 동사 어휘를 모두 의미한다

영어에서 수동태는 문법적으로 동사 어근의 굴절과 통사적 구조에 변형 및 새로운 요소의 추가(*be* 동사)를 요구하는 반면 한국어에서는 피동화가 특정 통사 구문에 수반되는 동사의 형태 변이라고 하기보다는 동사 어휘부에서 파생의 과정을 통해 실현된다. ‘경찰이 용의자를 잡았다’와 ‘용의자가 경찰에게 잡혔다’라는 두 문장에서는 영어의 수동문과 같이 능동태의 행위주와 대상이 피동문에서 각각 위치 변화를 받고 격변화¹⁰ 및 의미역이 전환이 되고는 있으나 이는 피동의 필요 충분조건은 아니다. 이는 첫째, 모든 타동사가 피동화 될 수 없다¹¹는 것과 능동-피동, 피동-능동의 구분이 항상 상호 대응되지 않는다는 현상에서 밝힐 수 있다[남기심 외 1985]. 또한 ‘영희가 철수에게 손을 잡혔다’라는 문장에서와 같이 피동사가 목적어를 가지는 경우¹²와 같은 예는 한국어 피동사가 단일한 통사 구조를 갖는 것이 아니라는 점을 보여준다. 한국어에서 피동사가 실현되는 문장 구문은 양정석[1995]에 따르면 크게 4가지 유형으로 상세하게는 10가지 유형의 구문 형태를 가질 수 있다¹³. 이 10개의 구문은 수동태에만 국한된 격 구조가 아니라 일반 자동, 타동, 사역동사 모두 이러한 구문에서 나타날 수 있어서 한국어에서의 피동은 영어의 수동태와는 달리 반드시 통사 구조에 의존적이지 않음을 볼 수 있다. 이와 더불어 한국어 타동사가 모두 피동 형태를 갖는 것도 아니라는 점에서, 한국어에서는 피동이라는 의미적 속성이 특정한 문장 구조에 의존하는 통사적 속성이 되기 보다는 어휘부에 의존하는 의미적 속성으로서 각각의 동사 의미에 따라 문장 구조가 선택된다고 볼 수 있겠다. 바로 이러한 점에서 영어와는 달리 한국어 동사 어휘의미망에서는 피동사를 독립적인 어휘화된 동사 개념으로 간주하여 동일한 어간을 둔 타동사와는 분리된 의미 분류를 할 필요성이 있음을 보여주고 있다.

3.2 한국어 대역어에서 오류 유형과 대역어 분포

본 절에서는 대역된 표 1에서 나타난 PWN 의미분류 유형 III에 대해서 한국어 대역어가 어떻게 나타나는지 검토한다. 이를 위하여 *verb.change*만을 선택해서 대역 오류 분석의 범위를 좁혀보겠다. 분석을 위하여 우리는 *verb.change*에서 대격과 비대격을 동시에 갖는 영어동사의 동의어집합을 추출하였다. 방법은 우선 비대격의 의미를 가능하게 하는 (1)의 문형 중 적어도 1개를 갖고 이와 동시에 대격이 실현되는 (2)의 문형을 1개 이상 갖는 동의어집합을 추출하였다.

- (1) Somebody__s
Something __s
- (2) Somebody ____s something

¹⁰ 능동문의 주격과 대격이 각각 피동문에서 여격과 주격으로 변화한다.

¹¹ 예를 들어, 주다, 받다, 얻다, 잃다, 참다, 돕다, 알다, 배우다, 바라다, 느끼다 등

¹² 고광주(2001)은 이러한 피동사를 타동사로 본다 이정식(1992)는 이 경우 ‘손을’을 엄밀한 의미에서 대격이 아니라 부분격이라고 주장한다. 자세한 논의는 고광주(2001)을 참고.

¹³ I [NP이 V], [NP이 NP에 V], [NP이 NP로 V], [NP이 NP이 V], [NP이 NP라고 V]

II [NP이 NP에게 V]

III [NP이 NP에게 NP를 V], [NP이 NP에 NP를 V]

IV [NP이 NP에 NP이 V], [NP이 NP에게 NP이 V]

Somebody ___s somebody

Something ___s somebody

Something ___s something

문형정보와 더불어 뜻풀이에 이중적인 의미정보가 내포되는지도 함께 고려하였다. 문형 정보가 대격과 비대격을 동시에 만족시키더라도 뜻풀이에서 이에 대한 의미가 모두 명백하게 설명되지 않은 경우도 있었다. 그러나 우리는 문형정보가 대격-비대격을 보충적으로 표현한다고 보고 뜻풀이 정보가 이중적인 의미를 갖지 않는다 하더라도 추출하였다. 또한 워드넷의 문형정보는 영영 사전(Merriam's Webster Unabridged)를 통해 확인하는 과정을 거쳤다. 여기서 추출된 동의어집합의 수는 verb.change에서 117개이고 그 중에서 연어(collocational phrase)의 형태의 어휘로 이루어진 2개의 동의어집합은 제외하였다. (총 115)

	1항	2항	3항	4항	
대역어휘형태의 결합적 분포	타동	자동	타동/피동	타동/자동	Total
Synset개수	50	15	7	43	115

표 2 영어 -한국어 자동, 타동, 피동의 대역어 분포

한국어 동사 어휘의 의미망 구축을 위해서 영-한 이중언어MRD(4만 단어)를 통한 반 자동 대역의 결과로 표2에서 나타난 것처럼 타동과 자동의 의미를 상세하게 반영하지 못함을 보여준다. 다음의 예는 이러한 문형을 반영하지 못한 오류의 예이다.

예1) {burst 1, burst 0}

Sentence frame : Something ----s

Somebody ----s something

Something ----s something

Gloss: break open or apart suddenly and forcefully

대역어: 터지다, 파열하다,

부족한 한국어 대역어: 터뜨리다. 파열시키다, 파열되다

예2) {steam 0}

Sentence frame: Something ----s

Somebody ----s something

Gloss: cook something by letting steam pass over it

대역어: 찌다

부족한 한국어 대역어: 찌지다

1)의 예에서는 현재 대역어가 자동의 의미만을 반영하고 있다. 따라서 이러한 경우는 타동의 의미를 갖는 타동사 ‘파열하다’, ‘터지다’가 추가되어야 한다. 예2는 현재 대역어가 타동의 의미만을 반영하고 있으나 이의 피동사 또한 추가 되어 대역되어야 할 것이다.

이렇게 워드넷에 주어진 문형정보는 모호한 뜻풀이로 인한 대역의 오류를 막고 대역어

오류 정제의 기준으로 삼을 수 있다. 위의 표2에서 나타난 대역어를 이러한 우리의 어휘의미망 정제 기준에 따라 오류 수정을 하여 다음의 표에서 보여주는 결과를 얻게 되었다.

	1항	2항	3항	4항	
대역어휘형태의 결합적 분포	타동	자동	타동/피동	타동/자동	Total
Synset개수	20	2	34	65	115

표 3 대역어 수정 결과표

4. 피동의 의미자질을 고려한 어휘의미의 분류와 의미망 구축

3장에서는 우리는 PWN이 동사의 문형정보로 제공하는 대격-비대격을 정제 기준으로 삼아 한국어 대역어의 오류를 수정하였다. 본 장에서는 그 결과인 표3의 자료에서 한국어에서 대역 유형을 통해 한국어에서 피동의 의미 자질을 구분해야 할 당위성을 검토하고 이를 토대로 어휘의미의 계층구조에 구성에 대한 방법을 제시하고자 한다.

4.1 피동/타동 어휘 의미의 구분

표3에서는 네 가지의 대역어 유형을 볼 수 있다. 1항과 2항은 한국어 대역어가 타동사 또는 자동사로만 나타나는 경우이다. 여기에는 두 가지 이유가 있는데 첫째, 영어와 병렬하여 다국어 어휘의미망을 구축하는 경우 보편적으로 나타나는 문제 중 하나로서 영어 어휘의 통사적 속성이 한국어에서도 평행하게 적용되지 않기 때문이다. 즉 영어의 타동개념이 항상 한국어에서는 동일한 의미자질을 갖는 타동사로서 대역되지 않는다. 반대로 영어의 자동의 의미자질을 갖는 동사가 한국어에서 반드시 영어의 의미자질을 그대로 대역되는 것을 아니다. 또한 영어의 자동/타동의 의미속성을 갖는 하나의 어휘개념이 한국어에서는 이중 하나의 의미속성을 갖는 어휘화된 개념으로밖에 대역될 수 없는 경우이다. 이것은 언어간 어휘개념의 비평행성과 어휘의 통사적 속성이 개별언어에 종속적이라는 것을 보여주는 예이다.

그러나 3항의 결과는 1, 2항의 결과와는 달리 개별언어에 종속적인 어휘의 통사적 속성과 어휘의미망 내의 동사 의미 분류의 문제를 병행하는 경우이다. 이는 영어 동사 어휘의미가 2개의 상이한 의미자질 자동/타동이 한국어에서 역시 2개의 상이한 의미자질을 갖는 어휘로 대역되지만 대역된 어휘가 다르고 그것이 표상하는 개념도 한국어에서는 상이하게 구분되어야 하기 때문이다. 즉 한국어에서는 타동의 의미속성을 갖는 어휘와 자동의 의미속성을 갖는 어휘를 하나의 동의어집합으로 분류할 수 없다는 것이다. 그 첫 번째 이유는 다음의 예를 통해 살펴 본다. 영어 {change2, alter 0, vary1}에 대한 한국어 대역어 ‘바꾸다’와 ‘바뀌다(또는 바꾸이다¹⁴)’는 다음의 문장으로 실현된다.

¹⁴ ‘바뀌다’의 본딤말

예3) 그는 이 문제에 대한 우리의 견해를 바꾸었다.

예4) 그에 의해 이 문제에 대한 우리의 견해가 바뀌었다.

여기서 ‘바꾸다’와 ‘바뀌다’가 동일한 사행에 대한 관점의 전환을 전제한다. 또한 동사의 명사 논항의 선택제약이 완전히 다르며 어휘의 통사적 속성 또한 다르다. 이는 마치 영어의 능동 구문과 수동 구문이 동일한 의미를 갖지 않는 것과 같다. 한국어에서는 영어에서 문장단위에서 통사적 구성으로 실현되는 수동의 의미가 한국어에서는 어휘부에서 실현되기 때문에 영어의 의미구조와는 매우 다르다고 할 수 있다.

두 번째 이유는 타동/피동을 하나의 동의어집합으로 묶어 주는 것은 어휘의미망의 기본 구성요소인 동의어집합의 정의를 위배한다는 것이다. 동의어집합은 유사한 어휘의미의 집합이며 이 어휘들은 적어도 하나 이상의 문맥에서 문장의 의미를 변화시키지 않고 교체 가능해야 한다. 이는 워드넷의 동의어집합의 정의 “동의어는 어떤 문맥에서 교체 가능하다 (WorldNet’s synonyms can be interchangeable in some context)”를 보다 구성적 원칙으로 우리가 재정의한 것이다. ‘바꾸다’와 ‘바뀌다’를 비롯하여 한국어의 타동사와 그 피동사는 어떠한 문맥에서도 교체될 수 없다. 결국 위의 두 가지 언어학적 근거에서 한국어에서 타동과 피동을 하나의 동의어집합이 아닌 두 개의 동의어집합으로 나누어야 하고 계층적 구조 또한 변형해야 한다는 결론에 이르게 한다.

4.2 동사 어휘의미망 내의 피동사의 위치

그렇다면 한국어 동사 어휘의미망에서 타동과 피동의 어휘의미를 어떻게 계층적 구조로 구축할 것인가의 문제가 남아있다. 한국어 동사 어휘의미망은 영어의 계층구조를 상당부분 수용하지만 기존에 구축된 계층구조와 최상위 및 상위노드의 의미와 통사적 속성, 상위어들이 갖는 인과관계, 함의 관계 및 명사와의 파생관계를 정확히 파악하여 한국어 동사어휘의 의미적인 승계 및 관계에 따른 위치를 찾고 구조를 재정비하는 방법을 모색해야 한다.

우리는 앞서 표1에서 워드넷의 verb.change의 대격-비대격 교체 동사의 의미 분류 유형을 보았다. 유형I에서 타동/자동에 따라 의미가 구분된 경우 최상위는 각각 자동과 타동의 의미속성을 반영하는 의미로 설정되어 있다. verb.change의 최상위 노드는 34개의 동의어집합으로 하나의 동사가 타동과 자동의 통사적 특성을 가짐과 동시에 의미 구분이 될 수 있는 경우 각각의 계층구조에서 짝을 이룰 수 있는데 대부분의 경우 {change 0}(비대격)-{change1, alter1, modify 11}(대격)의 최상위 노드의 쌍으로 귀속된다. 또한 여기서 새로 구조 변형을 하는 피동사와 타동사 간의 공유 어휘소(lexeme)의 존재를 표시하기 위하여 한국어 어휘의미망 내에 인과관계와 같이 피동-능동의 관계를 설정해 주어야 한다.

다음은 영어 {bake 0}¹⁵의 계층 구조를 나타낸 것이다.

¹⁵ 의미: cook and make edible by putting in a hot oven, "bake the potatoes"
문형정보. Something ----s,

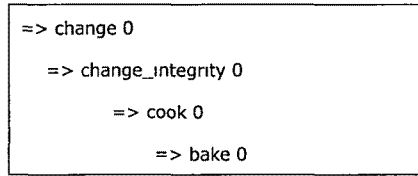


그림 5 PWN의 bake의 계층구조

현재 계층구조로는 ‘굽다’와 ‘굽히다’가 모두 {bake 0}의 대역어로 한국어 어휘의미망에서 하나의 동의어집합을 이루게 되나 4.1에서 논의한 바와 같이 이를 각각의 어휘의미로 세분화하고 이에 따라 독립적인 계층 구조를 구성하게 된다.

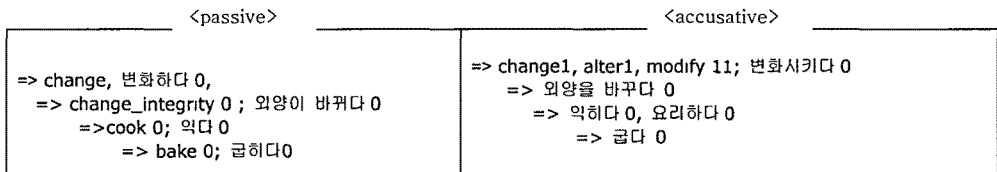


그림 6 피동과 타동사의 계층구조 분리

그림6에서는 피동과 타동의 동사 의미를 각각 분리하여 계층구조로 형성하였다. 왼쪽의 피동사 ‘굽히다’는 상위어 또한 ‘익다’라는 피동사를 갖게 되며 최상위는 ‘change 0’의 ‘변화하다’로 설정된다. 여기서 피동사 ‘굽히다’의 최 상위어 ‘변화하다’는 한국어에서 자동사이다. 한국어 동사 어휘의미망 내에서 피동사는 자동사의 의미 계층구조에 무리 없이 연결될 수 있고 피동사만을 위한 상위 계층구조는 만들지 않는다¹⁶.

‘굽다’는 최 상위어를 ‘change1, alter1, modify 11’의 대역어 ‘변화시키다’를 갖게 된다. 결과적으로 한국어에서는 ‘굽히다’와 ‘굽다’라는 새로운 계층 구조를 형성하게 되었다. 영어 ‘bake 0’의 상위어 ‘cook 0’와 ‘change integrity 0’ 또한 대격-비대격이 동시에 존재하는 동사의 집합이었으므로 새로운 계층 구조에서 모두 어휘 의미의 세분화가 일어나게 되었다. 그 결과 피동사 ‘굽히다 0’의 계층구조는 영어의 ‘bake 0’와 연결되지만 타동사 ‘굽다 0’의 계층구조는 최상위 노드인 ‘변화시키다’를 제외하고 영어와 직접적으로 연계되지는 않는다. 이러한 방법을 통하여 표3의 3항에 타동과 피동사가 함께 대역되어 있는 경우에 대해 의미 세분화와 동시에 상위 구조를 수정하였고 다음의 표 4는 그 결과를 보여준다.

타동과 피동에 따른 분류 ‘변화시키다’에 따른 분류
 타동과 피동에 따른 분류 ‘최상위노드’에 따른 분류

Somebody ----s something.
 Something ----s something

¹⁶ 한송화(2000)는 피동사를 자동사의 특수한 하위 범주로 구분할 수 있는 통사 및 의미적 속성이 없기 때문에 이를 자동사로 분류하고 있다.

수정 전		34	5	29
수정 후	34	34(+6)	34	34

표 4 피동 어휘의미의 계층 구조 수정 결과

표4에서 보는 바와 같이 수정 이전에는 타동과 피동사가 하나의 동의어집합에 공존하였다. 그러나 수정 후 각각 상이한 34개의 동의어집합으로 구분되었다. 특히 피동의 의미에서 본래 출발 동의어집합에서 보다 6개의 의미 노드가 더 추가 되었는데 이는 영어 워드넷이 표현하지 않은 한국어의 피동의 개념 표현을 위한 인공적인 의미노드 5개¹⁷와 결여 의미 노드¹⁸ 1개를 포함한다. 현재 PWN에서는 verb.change에서 일정 동사의 의미를 표현하기 위하여 인공적인 의미 노드들을 최상위 노드의 직접 하위에 두고 있으며 이는 대격과 비대격으로 혼용되고 있다. 이를 한국어에서는 타동과 피동으로 또는 타동과 자동으로 의미 세분화한 노드로 독립시켜 줄 필요가 있으며 본 연구에서는 새로이 분화되는 피동사의 개념을 최 상위어와 연결해 주기 위해 6개를 이용하였다.

또한 수정 전에는 대격-비대격의 쓰임을 모두 갖는 동사가 비대격 동사를 최상위로 갖거나(29) 또는 대격 동사를 최상위로 갖는 경우(5)가 혼재하였다. 이 경우 계층 구조의 수정을 위하여 각 34개의 동의어집합의 의미 계층 구조를 분석하여 수정하였다. 표4에서 나타난 바와 같이 한국어의 타동사가 표현하는 개념의 상위어는 {변화시키다}로 설정하고 피동사가 표현하는 개념의 상위어는 {변화하다}로 각각 이분화된 쌍으로 설정하였다. 또한 이렇게 분화된 타동과 피동사의 동의어집합에는 피동-타동의 태그를 설정하여 각각 독립적인 계층구조로 분화된 의미간의 관계를 연결해 주기 위하여 타동-피동의 태그를 연결하였다.

5. 결론과 향후 연구

지금까지 우리는 영어에서 빈번하게 고찰되는 대격-비대격 교체 동사가 비일관적인 의미 분류의 원인이 되고 있음을 보았다. 특히 한국어에서 타동과 피동으로 대역되는 현상을 근거로 하여 한국어 어휘의 형태, 통사적 속성을 고려한 어휘 의미의 구분이 필요함을 증명하였다. 어휘의미의 정확한 구분과 분류는 한국어 어휘의미망이 다국어 어휘의미망으로서 영어-한국어 기계번역과 정보 및 지식 검색에 적용되었을 때 대역의 정확성과 다국어 연동 및 한국어 의미분석(semantic parsing)에 효율성을 보장할 수 있을 것이다.

본 연구의 결과는 PWN의 비일관적인 분류가 개별 언어의 특성을 고려하여 수정되어야 하며 필요에 따라 계층 구조의 변형도 불가피 함을 보여준다는 점에서 PWN에 연동된 모든

¹⁷이들은 각각, {change_intgrnty}, {change_surface}, {change_state}, {change_by_reversal}, {change_magnitude}이다

¹⁸ PWN에는 {decay}(부식하다, 부식시키다)의 대격 쓰임과 의미는 있으나 비대격의 쓰임과 의미는 부재한다. 한국어에서는 '썩다, 부패하다'등 의 동사를 위해 '부식되다'의 상위 의미가 필요하다.

다국어 어휘의미망 구축에서 일반적 적용이 가능하다고 하겠다. 또한 접사의 파생으로 어휘 의미의 형성이 매우 생산적인 교착어와 PWN과의 사상에서 나타날 수 있는 문제는 동사에 있어서 매우 유형이 다양할 수 있다. 이러한 점에서 본 연구에서는 논의하지 않은 타동과 자동의 문제 또한 본 연구의 연장선상에서 다루어져야 할 문제이다. 이와 더불어 verb.change외의 동사 의미 부류에서도 한국어의 타동/피동의 의미 구분과 계층구조 변경의 방법 또한 계속적으로 연구되어야 할 것이다.

Index

PWN Verb Frames

- 1 Something ~----s
- 2 Somebody ~----s
- 3 It is ~----ing
- 4 Something is ~----ing PP
- 5 Something ~----s something Adjective/Noun
- 6 Something ~----s Adjective/Noun
- 7 Somebody ~----s Adjective
- 8 Somebody ~----s something
- 9 Somebody ~----s somebody
- 10 Something ~----s somebody
- 11 Something ~----s something
- 12 Something ~----s to somebody
- 13 Somebody ~----s on something
- 14 Somebody ~----s somebody something
- 15 Somebody ~----s something to somebody
- 16 Somebody ~----s something from somebody
- 17 Somebody ~----s somebody with something
- 18 Somebody ~----s somebody of something
- 19 Somebody ~----s something on somebody
- 20 Somebody ~----s somebody PP
- 21 Somebody ~----s something PP
- 22 Somebody ~----s PP
- 23 Somebody's (body part) ~----s
- 24 Somebody ~----s somebody to INFINITIVE
- 25 Somebody ~----s somebody INFINITIVE
- 26 Somebody ~----s that CLAUSE
- 27 Somebody ~----s to somebody
- 28 Somebody ~----s to INFINITIVE
- 29 Somebody ~----s whether INFINITIVE
- 30 Somebody ~----s somebody into V-ing something
- 31 Somebody ~----s something with something
- 32 Somebody ~----s INFINITIVE
- 33 Somebody ~----s VERB-ing
- 34 It ~----s that CLAUSE
- 35 Something ~----s INFINITIVE

References

- 고광주. 2001. 목적어 있는 피동문: 한국어의 목적어: 월인, 103-122
- 김중복. 2005. 한국어 피동구문의 제약기반이론적 접근: 한국어학 26권, 67-92

- 김현권. 2003. 어휘의미지식의 기술방법: *Proceedings in Korean Linguistic Society*
- 남기심, 고영근. 1985. 표준국어문법론: 탑 출판사
- 남승호. 2004. 한국어 술어의 의미 유형 연구: 다의 술어의 사건 구조와 논항 구조를 중심으로: *Proceedings of Korean Linguistic Society*. PDF file format
- 양정석. 1995. 국어 동사의 의미분석과 연결이론: 박이정
- 이은령, 윤애선. 2005. 워드넷 기반 한국어 동사 어휘의미망 정제: *Proceedings of Korean Linguistic Society*, 225-227
- 이은령 외. 2004. 다국어 어휘의미망 구축의 현황과 문제점 - PWN과 EWN을 중심으로: 프랑
스문화예술학회 12, 369-401
- 이정택. 2001. 피동성 표현에 관한 연구-‘되다, 받다, 당하다’를 대상으로: 한글 251, 143-166
- 한송화. 2000. 현대 국어 자동사 연구: 한국문화사
- Bilgin, O et al. 2003. Morphosemantic Relations In and Across WordNets- A Study Based on Turkish, *Proceedings of the Second International WordNet Conference*, 60-66
- Cruse, D.A. 1986. *Lexical Semantics*: Cambridge University Press
- Fellbaum, C. 1998. *WordNet- An Electronic Lexical Database*: MIT Press
- Gomez, F. 2003. Grounding the ontology on the semantic interpretation algorithm: *Proceedings of the Second International WordNet Conference*, 124-129
- Jackendoff, R. 1990., *Semantic structure*: MIT Press
- Karen T. Kohl et al. 1998. Representing Verb Alternations in WordNet: *WordNet- An Electronic Lexical Database*: MIT Press
- Levin, B. 1993. *English Verb Classes and Alternations-A Preliminary Investigation*: University of Chicago Press
- Pustejovsky, J. 1998. *The Generative Lexicon*: MIT Press
- Vossen, P. et al. 1998. Compatibility in Interpretation of Relations in EuroWordNet: *Computers and Humanities* 32, 153-184