

사전 뜻풀이말에서 구축한 한국어 명사 의미계층구조

A Semantic Hierarchy of Korean Nouns using the Definitions of Words in a Dictionary

조 평 옥 안 미 정 옥 철 영 이 수 동
(Pyeong-Ok Jo) (Mi-Jeong An) (Cheol-Yung Ock) (Soo-Dong Lee)

요 약 한국어 명사들을 의미별로 분류하여 계층화시킨 '한국어 명사 의미계층 구조'는, 한국어 문장을 처리할 때 의미 정보를 제공할 수 있는 매우 중요한 것들 중의 하나이다. 본 논문에서는, 국어 사전의 명사에 대한 뜻풀이말을 이용하여 bottom-up 방식으로 '한국어 명사 의미 계층 구조'를 구축하였다.

본 논문에서 구축한 '한국어 명사 의미계층 구조'는, 트리(tree)가 43개, 중간 노드(node)가 2,443개, 단말 노드(terminal node)가 10,347개이며, 깊이(depth)가 17인 하나의 포리스트(forest)를 이룬다. 이것의 제 1, 2 계층(level 1, 2)에서의 분류 형태는 top-down 방식에 의한 기존의 분류들과 매우 다른 모습인 반면에, 제 3 이하의 계층들은 상당히 객관적인 형태를 이루고 있다.

주제어: 상의어, 바로 위 상의어, 동형이의어, 다의어, 형클어진 의미계층구조

Abstract A Semantic Hierarchy of Korean Nouns(SHKN) where Korean nouns are semantically and hierarchically classified is one of the most important things that can provide semantic information concerned with processing the Korean sentences. In this paper, SHKN is constructed in bottom-up method, by making use of the definition of a noun in the Korean dictionary.

SHKN constructed in this paper is a forest, which consist of 43 trees and 2,443 non-terminal nodes and 10,347 terminal nodes, depth of which is 17. The classes of level 1, 2 of SHKN is quite different from the existing structure classified in top-down method, but the lower classes than level 2 is very objective.

1. 서론

컴퓨터를 이용한 한국어 문장의 각 분석 단계에서, 자연 언어의 중의성으로 말미암은 애매성들이 많이 나타난다. 이것들을 효과적으로 해소하기 위한 하나의 방법으로서, 관련 어절들 사이의 의미 호응 여부를 컴퓨터가 판단하게 하는 것을 들 수 있다. 이것은 문장 내에서 서로 의미 호응을 이룰 수 있는 낱말들의 짝을 나타내는 정보를 구축하여, 이 정보를 컴퓨터가 이용할 수 있게 함으로써 가능하다.

이런 정보를 구축하기 위해서는 한국어 명사들을 의미별로 분류하여 계층화시킨 '한국어 명사 의미계층 구조'가 필요하고, 따라서 '한국어 명사 의미계층 구조'의 구축은 시급한 과제라고 할 수 있다.

본 논문에서는, 사전에 기술된 뜻풀이말의 의미에 의해서 각 명사들의 바로 위 상의어(上義語, hypernym)¹⁾를 정해줌으로써, 기존의 하향식 한국어 명사 분류 연구들과는 달리 연구자의 주관에 개입되지 않은 비교적 객관적인 '한국어 명사 의미계층 구조 (A Semantic Hierarchy of Korean Nouns)'가 자동으로 bottom-up 방식으로 구축되게 하였다.

* 울산대학교 전자계산학과
Depat. of Computer Science, University Of Ulsan
울산광역시 남구 무거2동 산 29번지, 680-749
Tel : 052-259-2222
Fax : 052-259-1687
E-mail : okcy@uou.ulsan.ac.kr

1) 상의어들 중에서 가장 아래 계층의 상의어

2. 관련 연구

지금까지 자연언어 정보 처리를 위하여 다양한 의미계층 구조 표현 방법이 제시되었지만 그것들의 장단점 때문에 그 방법들은 많은 논란을 불러 일으켜 왔는데,^(23,26,27,28) 그것들 중에서 1990년 미국의 Princeton 대학에서 만들어져 1993년에 개정판이 나온 영어 WordNet은 실용화되어 자연언어 처리에 많은 기여를 하고 있다. 이 WordNet은 언어 심리학 이론들이 반영되어 설계된 어휘 참조 체계(lexical reference system)로서 영어의 명사, 동사, 형용사 등을 동의어 집합(synonym set, synset)들로 나누어 각 집합들이 어휘 개념을 나타내게 하였는데, 명사에 대해서는 약 57,000개의 단어들을 대략 48,800개의 synset들로 나누고 있다.⁽²³⁾

국내에서는 최근에 영어 WordNet의 어휘 개념을 이용하여 한국어 명사 어휘 개념을 한 개의 단어가 아닌 동의어 집합으로서 표현한 연구⁽²⁰⁾가 이루어졌다. 그런데 언어와 사고(思考) 사이에는 밀접한 관계가 있으므로⁽²⁵⁾ 영어 어휘 개념과 한국어 어휘 개념 사이에는 다소의 차이가 있을 수 있는데,⁽²⁰⁾은 이와 관련된 문제점을 지니고 있다고 할 수 있다.

[21]과 [22]에서도 한국어 명사 분류를 시도하였는데, [21]은 소수(85 개)의 명사를 가지고 명사와 동사 사이의 구분 관계를 이용하여 명사 분류를 실험한 연구로서 명사의 상·하의어 관계에 의한 계층적 분류 연구는 아니다. [22]는 연구자의 주관에 의해서 미리 정해진 기준에 따라 상위(上位) 영역을 먼저 정한 다음 그것의 하위(下位) 영역을 정해 가는 top-down 방식에 의해서 한국어 명사 개념 분류 체계를 제시한 연구로서, 분류 체계에 대한 객관성이 결여될 수 있으며 또한 부분적으로 상위 쪽의 분류 체계만 제시했을 뿐 전체 명사를 대상으로 한 구체적인 분류 체계는 제시하지 않았다.

3. 한국어 명사 의미 계층 구조 구축

본 논문에서는, 다음과 같은 과정을 통하여 '한국어 명사 의미계층 구조'를 구축하였으며, 이중 첫 번째 과정에서 세 번째 과정까지의 작업은 연구자의 수작업으로 이루어졌으며 네 번째 작업은 자동으로 수행되었다.

첫째, 대상(對象) 명사들을 선정.

둘째, 명사들의 '바로 위 상의어' 결정

셋째, 명사들의 바로 위 상의어 정보 데이터(data)

작성.

넷째, 의미 계층 구조 구성.

3.1 대상(對象) 명사 선정

'한국어 명사 의미계층 구조' 구축 작업 대상 명사를 선정하는 것은 구축된 '한국어 명사 의미계층 구조'의 활용도(活用度)에 있어서 대단히 중요하므로 모든 한국어 명사를 가지고 의미 계층 구조를 구축하는 것이 가장 바람직하다. 그러나 모든 한국어 명사를 대상으로 하여 작업하는 데에는 많은 어려움이 따른다. 따라서 본 논문에서는, 초등 학교 교과서들에 나오는 낱말 및 일상 생활에서 자주 쓰이는 낱말들을 싣고 있는 초등 학생용 국어 사전 [5]와 [7]에 실린 명사들 중에서, 다음의 두 가지 원칙에 맞는 것들을 일차(一次)로 '한국어 명사 의미계층 구조' 구축을 위한 명사로 선정하였다.

(가) '-하다'형 자동사, '-하다'형 타동사, '-스럽다'형 형용사, '-하다'형 형용사 등이 되지 않는 순수 명사들 중에서 보통 명사만을 선정한다.

(나) 복합 명사 및 외래어는 제외한다.

원칙 (가)에 의해서 '-하다'형 동사 및 형용사가 작업 대상 명사에서 제외됨으로써 이것들이 지니고 있는 동작성과 상태성의 의미 범주가 '한국어 명사 의미계층 구조'에서 빠지게 되는 문제점을 우려할 수 있으나, 위의 두 원칙들에 의해서 일차로 선정된 순수 보통 명사들 중에 동작성 및 상태성의 명사들이 포함되기 때문에 이런 문제점은 해결될 수 있다.

일차 선정에서 제외된 명사들 중에는 일차로 선정된 명사들의 '바로 위 상의어'인 것들이 많이 들어 있는데, 이 것들은 선정된 명사들의 바로 위 상의어를 결정하는 과정에서 추가로 선정하여 포함시켜 나갔다. 이렇게 하여 본 논문에서는 결과적으로 총 12,833 개의 명사를 '한국어 명사 의미계층 구조' 구축 대상으로 선정하였다.

3.2 명사들의 바로 위 상의어 결정

선정된 각 명사들의 '바로 위 상의어'를 사전의 명사에 대한 뜻풀이말을 이용하여 결정하였는데, 사전(辭典)의 뜻풀이말들은 여러 가지 형태로 이루어져 있으며, [표 3.1]에서 처럼, 같은 명사의 것일지라도 사전에 따라서 다른 경우도 많다.

3.2.1 같은 명사의 뜻풀이말이 사전에 따라 다른 경우

같은 명사의 뜻풀이말이 사전에 따라서 다른 경우에는, 그것들 중에서 해당 명사의 '바로 위 상의어'로서 가장 타당한 핵심어를 포함하고 있는 것을 뜻풀이말로 채택하였다. 예를 들면, [표 3.1]의 경우에는 '반나절'의 상의어로서 '반'보다는 '동안'이 더욱 타당하므로 '한나절의 반쯤 되는 동안'을 '반나절'의 뜻풀이말로 채택하였다.

[표 3.1] 뜻풀이말 비교표

표제어	뜻풀이말	핵심어	상의어	사 전
반나절	한나절의 반쯤 되는 동안.	동안	동안	'동아새국어사전', 동아출판사, 1994
	한나절의 반.	반		'뉴에이스국어사전', 금성교과서주식회사, 1991

3.2.2 뜻풀이말의 형태에 따른 상의어 결정

명사의 뜻풀이말들은 그 형태가 다양한데, 본 논문에서는 사전 [5]와 [7]의 뜻풀이말의 형태를 조사·분석하여 다음과 같이 11 가지의 형태로 분류하여 각 형태에 따라 '바로 위 상의어' 결정 방식을 달리 정하였다.

- 형태1 : 맨 끝에 핵심어가 있는 형태.
- 형태2 : 2 개의 핵심어 A, B에 의해서 'A와 B'의 꼴로 되어 있는 형태.
- 형태3 : 두 개의 핵심어 A, B에 의해서 'A 또는 B'인 형태.
- 형태4 : 핵심어가 맨 앞에 있는 형태.
- 형태5 : 핵심어 A에 대하여, 'A의 비유', 'A를 ~는 말', 'A의 속된 말', 'A의 예사스러운 말' 등으로 되어 있는 형태.
- 형태6 : 핵심어 'A'에 대하여 'A의 일종' 또는 '~ A의 하나'인 형태.
- 형태7 : '~의 부분'인 형태.
- 형태8 : 'A의 준말' 또는 'A의 본디말'인 형태.
- 형태9 : 뜻풀이말이 표제어의 동의어 또는 유의어로 이루어져 있는 형태.
- 형태10 : 뜻풀이말 중에 의미상 핵심어가 전혀 들어 있지 않고 어떤 형태나 상태, 정도 등을 나타내고 있는 형태.

형태11 : 'A와 B의 총칭'의 꼴로 되어 있는 형태.

위의 각 형태들에 대하여 '바로 위 상의어'를 결정한 방식들은 아래와 같다.

형태1 : 그 말의 맨 끝에 있는 핵심어를 표제어의 '바로 위 상의어'로 정한다.

(보기) 「가락국수 길게 뽑은 국수」에서 「가락국수 < 국수 >」²⁾.

형태2 : 두 핵심어 A, B의 상의어들 중에서 처음으로 일치하는 것을 해당 표제어의 '바로 위 상의어'로 정한다.

(보기) 「가로세로 가로와 세로」에서 「가로 < 방향 >」이고 「세로 < 방향 >」이므로, 「가로세로 < 방향 >」.

그런데 이 경우에, 뜻풀이말의 두 핵심어 A, B가 아래의 보기처럼 같은 상의어를 갖지 않는 경우도 있다.

(보기) 「강약 강함과 약함」에서 강함과 약함의 공통 상의어를 찾을 수 없다.

이런 뜻풀이말을 갖는 표제어로는 [표 3.2]에서 열거한 것과 같은 것들이 있으며, 이런 표제어들은 작업 대상에서 제외시켰다.

[표 3.2] 상의어를 정하기가 곤란한 말들

강약(強弱)	계책(計策)	공과(功過)
권익(權益)	난이(難易)	목석(木石)
방책(方策)	방침(方針)	복리(福利)
빈부(貧富)	서화(書畵)	성쇠(盛衰)
성패(成敗)	손익(損益)	승부(勝負)
성패(勝敗)	신의(信義)	심신(心身)
여부(與否)	우열(優劣)	유서(由緒)
이해(利害)	인과(因果)	인의(仁義)
장단(長短)	재색(才色)	재질(才質)
정리(情理)	진퇴(進退)	천지(天地)
충의(忠義)	충절(忠節)	태갈(態-)
표리(表裏)	풍상(風霜)	풍파(風波)
피골(皮骨)	학덕(學德)	학술(學術)
학예(學藝)	화복(禍福)

2) 'A < B'는 표제어 A의 바로 위 상의어가 B임을 나타내는 기호임.

형태3 : 핵심어 A, B를 모두 해당 표제어의 '바로 위 상의어'로 정하여, 해당 표제어가 두 개의 '바로 위 상의어'를 갖도록 한다.

(보기) 「간사 모임이나 단체의 중심이 되어 일을 처리하는 직책 또는 그 사람」에서 「간사<직책, 사람」.

형태4 : 뜻풀이말의 맨 앞에 나오는 핵심어를 표제어의 '바로 위 상의어'로 정한다.

(보기) 「머리카락 머리털의 날개」에서 「머리카락<머리털」.

형태5 : 핵심어 A를 표제어의 '바로 위 상의어'로 정한다.

- (보기) ① 「어린이 어린이를 높여서 부르는 말」에서 「어린이<아이」.
- ② 「가시 미운 사람의 비유」에서 「가시<사람」.
- ③ 「양복짜리 양복쟁이」의 속된 말」에서 「양복짜리<양복쟁이」.

형태6 : 핵심어 A를 표제어의 '바로 위 상의어'로 정한다.

- (보기) ① 「산토끼 산·숲 속에 사는 토끼의 일종」에서 「산토끼<토끼」.
- ② 「부분 전체를 몇 개로 나눈 것의 하나」에서 「부분<것」.

형태7 : '~ 부분'의 꼴로 '바로 위 상의어'를 정한다.

- (보기) ① 「가슴 몸의 앞쪽, 배와 몸 사이의 부분」에서 「가슴<신체 부분」.
- ② 「줄기 식물의 가장 중심이 되는 부분」에서 「줄기<식물 부분」.
- ③ 「웃소매 웃옷의 두 팔을 꿰는 부분」에서 「웃소매<웃 부분」.
- ④ 「처마 지붕이 도리 밖으로 길게 내민 부분」에서 「처마<지붕 부분」.

형태8 : A의 '바로 위 상의어'를 표제어의 '바로 위

상의어'로 정한다.

(보기) 「가문비 가문비나무의 준말」에서 「가문비나무<나무」이므로 「가문비<나무」.

형태9 : 뜻풀이말의 '바로 위 상의어'를 표제어의 '바로 위 상의어'로 정한다.

(보기) 「괘 손」에서 「손<사람」이므로 「괘<사람」.

형태10 : '형태'나 '상태', '정도'를 '바로 위 상의어'로 정한다.

- (보기) ① 「고가 땅 위로 높이 가로질러 걸침」에서 「고가<형태」.
- ② 「괴로움 몸이나 마음이 아프거나 편안하지 못함」에서 「괴로움<상태」.
- ③ 「최강 가장 강함」에서 「최강<정도」.

형태11 : 바로 위 상의어를 정하지 않고 표 '..'로써 바로 위 상의어를 나타내어 이 경우의 표제어는 최상의어가 되게 한다.

(보기) 「사물 일과 물건의 총칭」에서 「사물<..」

그런데 신체 부위를 나타내는 말들의 경우에는 위의 방식들에 따라 그것들의 '바로 위 상의어'를 정하면 의미상으로 어울리지 않게 된다. 그래서 신체 부위를 나타내는 말들의 바로 위 상의어는 그것들의 뜻풀이말에 관계없이 '신체 부위'를 바로 위 상의어로 정하였다(예 : 「손등 손의 바깥 쪽」에서 「손등<신체 부위」). 그리고, [17]에서 보통명사를 분류한 '자연물, 인조물, 동안(時間), 일안(空間), 일(精神), 짓(行動)'들 중에서 '물건'을 상의어로 갖는 '자연물'과 '인조물'을 제외한 나머지의 항목들에 해당하는 '동안', '공간', '정신', '짓' 등에 대해서는 이것들이 최상의어가 되게 하기 위해서, 앞의 형태1에서 형태10까지의 방식들을 적용시키지 않고 형태11의 경우와 같이 처리하여 바로 위 상의어를 아래의 (보기)와 같이 정하였다.

(보기) '공간<...', '동안<...',
'정신<...', '짓<...'

3.2.3 다의어의 상의어 결정

명사가 다의어(多義語, polysemy)인 경우에는 그것의 분화된 뜻들에 따라서 두 가지 이상의 바로 위 상의어를 가질 수도 있는데, 그것의 분화된 뜻들은 사전에 따라서 종류와 가지 수가 다르게 나타난다. 본 논문에서는 다의어가 지닌 여러 가지의 분화된 뜻들 중에서 일상적으로 많이 쓰이는 것들 위주로 뜻풀이를 하고 있는, 두 개의 초등학교용 국어사전 [5]와 [7]에 나오는 뜻풀이말을 주로 이용하여 아래의 두 가지 방식으로 다의어인 명사들의 바로 위 상의어를 결정하였다.

첫째, 각 뜻풀이말들이 나타내는 '바로 위 상의어'가 다를 때에는 그 다른 상의어들을 모두, 표제어의 상의어로 택한다.

(보기) 「물컵이 ①물컹한 물건. ②몸이나 의지가 몹시 약한 사람.」에서 물컵이<물건, 사람

둘째, 뜻풀이말이 다른데도 그것들이 나타내는 바로 위 상의어가 같을 때에는 아래의 보기와 같이, 뜻풀이말들에 공통인 '바로 위 상의어'로써 표제어의 바로 위 상의어로 삼는다.

(보기) 「기공(氣孔) ①곤충류의 몸뚱이 옆에 있어서 숨을 쉬는 구멍. ②식물의 잎의 뒷면에 있는 무수한 작은 구멍. 공기 중의 이산화탄소를 빨아들이고 산소와 수분을 밖으로 내보내는 구실을 함.」에서 뜻풀이말 ①과 ②에 의한 '바로 위 상의어'가 모두 '구멍'이므로 「기공<구멍」으로 정한다.

3.3 바로 위 상의어 정보 데이터(data) 작성.

3.2 장에 의해서 결정된, 각 명사들의 바로 위 상의어를 나타내는 정보 데이터(data)를 다음과 같은 네 단계의 과정을 거쳐서 작성하였다.

- 첫째 단계에서는, 해당 명사의 바로 위 상의어가 :
- i) 하나 뿐이면, 「[표제어]{바로 위 상의어...}」인 형태로,
 - ii) 두 개 이상이면, 「[표제어]{바로 위 상의어1..}{바로 위 상의어2..}...{바로 위 상의어n..}」인

형태로

해당 데이터들을 작성하였다. 단, 표제어의 바로 위 상의어가 '표제어<...'인 꼴로 정해진 것의 정보 데이터는 「[표제어]{...}」인 형태로 작성하였다.

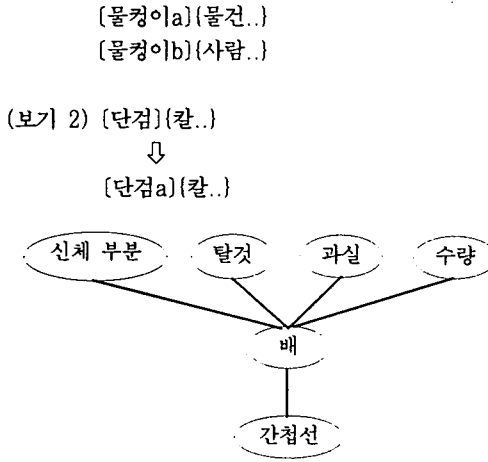
- (보기) ① 간첩선<배
↓
[간첩선]{배...}
② 물컵이<물건, 사람
↓
[물컵이]{물건..}{사람..}
③ 공간<...
↓
[공간]{...}

둘째 단계에서는, 첫 단계에서 만들어진 데이터들 중에서 「[표제어]{바로 위 상의어1..}{바로 위 상의어2..}...{바로 위 상의어n..}」의 꼴인 것을 각 상의어에 표제어를 반복하여 대응시켜서, 「[표제어]{바로 위 상의어1..}」, 「[표제어]{바로 위 상의어2..}」, ..., 「[표제어]{바로 위 상의어n..}」과 같이 바로 위 상의어 개수만큼의 데이터로 분리시켰다.

- (보기) [물컵이]{물건..}{사람..}
↓
[물컵이]{물건..}
[물컵이]{사람..}

셋째 단계에서는, 표제어가 같으면서도 바로 위 상의어가 다른 동형이의어(同形異義語, homonym) 및 다의어들의 데이터들에 대하여 서로 다른 의미의 표제어임을 나타내기 위하여, 표제어가 같은 데이터들끼리 모아서 각 데이터의 표제어를 아래의 (보기 1)처럼 알파벳 소문자로써 차례대로 태깅(tagging)을 하였다. 이때 상의어가 하나뿐인 데이터의 표제어에 대해서도 아래의 (보기 2)와 같이 문자 'a'로써 태깅을 하였다. 만약 동형이의어 및 다의어들을 태깅 없이 그것들의 상의어와 연결하면 [그림 3.1]과 같은 명확하지 못한 쓸모 없는 '헝클어진 의미 계층구조(sense-tangled hierarchy)'가 생성되게 된다[23].

- (보기 1) [물컵이]{물건..}
[물컵이]{사람..}
↓



{그림 3.1} 형클어진 계층 구조

네 번째 단계에서는, 아직까지 명확하지 못한 각 데이터들의 표제어 의미를 명확하게 하기 위하여 현재까지 작성되어진 정보 데이터를 이용하여 각 데이터들의 바로 위 상의어를 a, b, c, ... 등의 문자로써 아래 (보기)와 같이 태깅을 하였다. 단, 명사형 전성어미(轉成語尾)로 끝나는 말(예: 흐름), 의존 명사(예: 바), '-하다'형 용언(예: 활동) 그리고 '사물' 등과 같이 사전의 뜻풀이말에 의해서 바로 위 상의어를 정할 수 없어서 상의어 정보 데이터 형태가 '{표제어}{...}'인 것(예: {사물}{...})들은 제외시켰다.

(보기) '{간첩선a}{배...}'의 '배'는 여러 의미의 것들 중에서 탈것인 '배'이므로, {표 3.3}과 같은 내용으로 이미 작성되어 있는 '배'의 바로 위 상의어에 관한 정보 데이터를 이용하여 '배'를 'b'로써 태깅을 하여, '{간첩선a}{배...}'를 '{간첩선a}{배b...}'로 바꾼다.

{표 3.3} 상의어 태깅 이전의 데이터

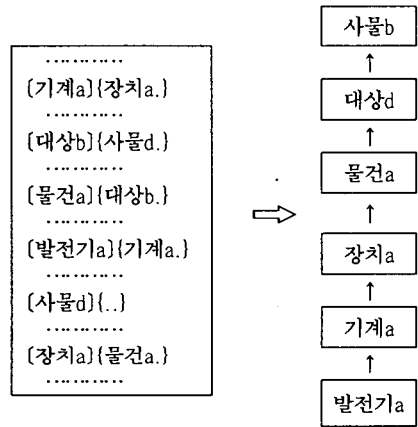
{배a}{신체 부분..}
{배b}{탈것..}
{배c}{과실..}
{배d}{수량..}

3.4 한국어 명사 의미 계층 구조 구성

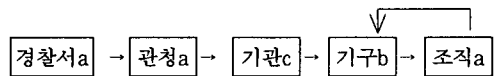
3.3 장까지의 작업 과정을 통하여 작성된 '상의어 정보 데이터'를 이용하여, 각 명사들을 프로그램에 의해서 자동으로 자신의 바로 위 상의어에 {그림 3.2}

처럼 연결시킴으로서, '한국어 명사 의미계층구조'가 bottom-up 방식으로 저절로 구성되게 하였다.

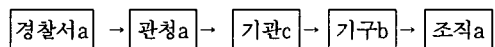
그런데 모든 명사들을 그것의 바로 위 상의어와 연결시키고 나면 두 가지의 연결 형태가 나타나는데, 하나는 {그림 3.2}과 같이 상의어를 갖지 않는 명사에서 연결이 끝나는 형태이고, 다른 하나는 {그림 3.3}처럼 이미 연결된 명사, 즉 자신의 하의어에 연결되는 순환 연결(loop) 형태이다. 이 중에서 첫째 경우는 그대로 두고, 둘째 경우에는, 각 명사들이 자신의 바로 위 상의어에 연결되는 과정에서, {그림 3.4}처럼 이미 연결된 명사가 상의어로 다시 나타나는 곳에서 자동으로 더 이상 연결되지 않게 하였다.



{그림 3.2} Loop가 없는 연결 형태



{그림 3.3} Loop가 있는 연결 형태



{그림 3.4} Loop를 제거한 형태

3.5 구축된 '한국어 명사 의미 계층 구조'

이상의 과정을 거쳐, 본 논문에서 구축된 '한국어 명사 의미 계층 구조'는 :

- 트리(tree) 수 : 43
- 중간 노드(node) 수 : 2,443
- 단말 노드(terminal node) 수 : 10,347
- 깊이(depth) : 17

인 하나의 포리스트(forest)이다.

‘한국어 명사 의미 계층 구조’의 포리스트를 이루고 있는 43 개의 트리들을 살펴보면 [표 3.4]와 같다. 이 트리들의 최상위 노드(root node)들을 이루고 있는 말들은 그 기능에 따라 [표 3.5]와 같이 분류되는데, 상위 계층 용어들의 ‘일반화로 인한 개념 불명확성 경향⁽²⁴⁾’ 때문에 그 개념이 명확하지 않은 명사형 전성 어미로 끝나는 말들이 다수 포함되어 있음을 볼 수 있다.

[표 3.4] 트리 일람표

차 례	최상위 노드	노드 수	단말노드 수	깊이	차 례	최상위 노드	노드 수	단말노드 수	깊이
1	갈음a	2	1	2	23	어려움a	5	3	3
2	것a	3,054	2,404	10	24	어울림a	32	24	6
3	고마움a	10	5	4	25	없음a	18	17	2
4	공간a	1,305	1,065	10	26	움음a	2	1	2
5	관련a	208	168	9	27	요소b	39	33	6
6	구별a	2	1	2	28	움직임a	45	41	4
7	기호a	115	96	8	29	이별a	2	1	2
8	나눔a	95	71	11	30	일b	12	6	5
9	다름a	8	6	3	31	일컬음a	43	34	5
10	동안a	215	185	6	32	작용a	66	56	7
11	때a	260	221	7	33	정신a	289	242	7
12	바a	3	1	3	34	존재a	34	23	5
13	방법a	74	57	4	35	좋음a	4	2	3
14	방향a	61	50	4	36	것a	657	523	7
15	보임a	3	1	3	37	층b	38	23	6
16	부상c	2	1	2	38	탈b	28	17	7
17	사물b	4,772	3,906	17	39	플d b	792	636	12
18	수고a	8	5	4	40	한계b	8	5	4
19	수고로움a	2	1	2	41	행사a	86	71	5
20	알림a	3	1	3	42	활동a	58	43	6
21	양b	363	294	9	43	흐름a	8	5	3
22	어긋남a	2	11	2	44				

[표 3.5] 트리의 루트 노드 분류표

종 류	개수	루트 노드
순수명사	16	공간, 기호, 동안, 때, 방법, 방향, 사물, 양, 요소, 일, 정신, 것, 층, 탈, 플, 한계
의존명사	2	것, 바
~하다 동사형 명사	9	관련, 구별, 부상, 수고, 이별, 작용, 존재, 행사, 활동
명사형 전성어미로 끝나는 말	16	갈음, 고마움, 나눔, 다름, 보임, 수고로움, 알림, 어긋남, 어려움, 어울림, 없음, 움음, 움직임, 일컬음, 좋음, 흐름

[표 3.6] ‘한국어 명사 의미계층구조’의 예

사물b→대상d→물건a→유형물a→물체a→생물a→동물a→말a→명마
 사물b→대상d→물건a→유형물a→물체a→생물a→동물a→말a→백마a
 사물b→대상d→물건a→유형물a→물체a→생물a→동물a→말a→애마a
 사물b→대상d→물건a→유형물a→물체a→생물a→동물a→말a→조랑말a
 사물b→대상d→물건a→유형물a→물체a→생물a→동물a→말a→궁영이a
 사물b→대상d→물건a→유형물a→물체a→생물a→동물a→말a→독벌레a
 사물b→대상d→물건a→유형물a→물체a→고체a→결정a→과망산칼륨a
 사물b→대상d→물건a→유형물a→물체a→고체a→결정a→백반a
 사물b→대상d→물건a→유형물a→물체a→고체a→결정a→일음a→고드름a
 사물b→대상d→물건a→유형물a→물체a→고체a→결정a→일음a→살얼음a
 사물b→대상d→물건a→유형물a→물체a→고체a→결정a→일음a→성에a

영어 ‘WordNet’에서 설정한 명사의 각 기본 개념⁽²³⁾에 해당하는 노드를 본 논문의 ‘한국어 명사 의미 계층 구조’에서 찾아내어, ‘WordNet’의 기본 개념들을 그것에 해당된 노드의 최상위 노드에 대응시켜, 영어 ‘WordNet’의 명사 기본 개념 목록과 ‘한국어 명사 의미계층 구조’의 최상위 계층을 비교하면 그 관계가 아래 [그림 3.5]와 같이 나타난다. 이 그림으로부터 ‘한국어 명사 의미계층구조’의 최상위 노드들 중 일부는 그 의미 범주가 지나치게 넓게 설정되었고 또 일부는 지나치게 좁게 설정되었으며, 상위 계층 용어들의 불명확한 개념들로 인하여 의미 범주로서 적합하지 않은 것들이 최상위 노드들로 다수 설정되어 있음을 알 수 있다. 그러나 의미 영역의 겹침 현상은 나타나지 않았으며 이것은 본 논문의 ‘한국어 명사 의미계층 구조’가 비교적 객관적으로 구축되었음을 나타낸다고 볼 수 있다.

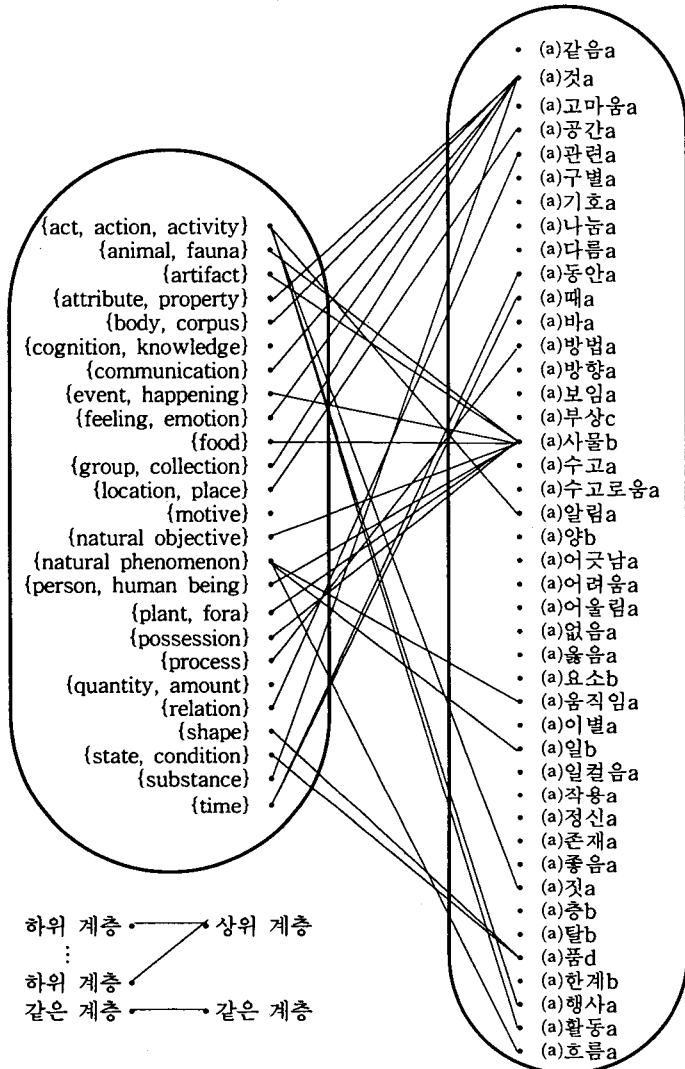
4. 결론 및 문제점

본 논문에서, 국어사전의 명사에 대한 뜻풀이말을 이용하여 구축한 ‘한국어 명사 의미계층구조’는 대체로 객관적인 분류 형태를, 특히 제 3 및 그 아래 계층에서는 매우 객관적인 분류 형태를 나타내고 있지만 다음과 같은 문제점들도 지니고 있다.

첫째, 최상위 노드(root node)들의 수가 지나치게 많으며, 제 1, 2 계층(level)의 분류 형태가 국어학에서의 일반적인 명사 분류 형태와 매우 다르다.

둘째, 사전에 따라서 뜻풀이말이 다르거나 뜻풀이말이 지닌 순환성(循環性)³⁾ 때문에 바로 위 상의어를 명확하게 정할 수 없는 명사들의 바로 위 상의어는 자의적(恣意的)으로 판단하여 결정하였는데, 이런 결정은 연구하는 사람과 이용된 사전(辭典)에 따라서 다를 수가 있으므로 국어 사전의 뜻풀이말을 이용하

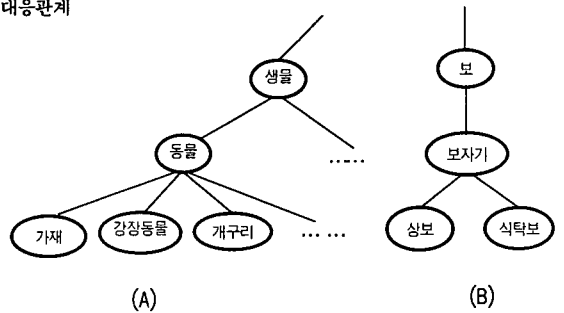
3) 두 명사 A, B의 뜻풀이말이 ‘A는 B이다.’ ‘B는 A이다.’인 꼴로 되어 있는 것.



(그림 3.5) WordNet의 기본 개념과 SHKN의 최상위 노드 사이의 대응관계

여 구축한 '한국어 명사 의미계층구조'가 달라질 수 있다.

셋째, '한국어 명사 의미계층구조'에서 몇몇 명사들은 너무 높게 부착되거나 너무 낮게 부착됨으로써 [그림 4.1] (A)의 '가재', '강장동물', '개구리' 등과 같이 분류 계층이 서로 달라야 할 명사들이 같은 계층으로 분류되거나, [그림 4.1] (B)의 '보', '보자기' 들처럼 같은 계층으로 분류되어야 할 명사들이 다른 계층으로 분류됨으로써 트리의 깊이(depth)가 비합리



(그림 4.1) 잘못된 계층 분류

적으로 되는 현상이 나타났다. 이런 점들은 [24] 및 [30]에서도 지적되고 있다.

사전으로부터 생성된 상위 계층의 용어들은 더욱 일반화되어 개념이 명확해지지 않는 경향이 있으며, 더구나 그런 용어들 중 몇몇의 것에 대해서는 그것들의 개념을 정확하고 간단하게 지시할 용어가 없다. 첫째와 둘째 번의 문제점들은 바로 이런 현상 때문에 생겨나는 문제점들로서 범주 변환(convert categories) 즉 그 어떤 특정 단어와도 일치하지 않는 개념들의 도입으로써 해결이 가능해질 수 있다^[24]. [표 3.2]에서 예를 든 것과 같은 한자(漢字) 합성어들에 대한 의미 영역도 이 범주 변환에 의해서 정의될 수 있을 것이다.

세 번째 문제점은 사전의 명사 뜻풀이말의 비일관성 때문에 생겨나는 것으로서, 관련 분야의 분류학을 참고로 한 보완 연구 및 유의어 처리, 그리고 국어사전의 뜻풀이말에 들어있는 중심어(즉, 표제어의 바로 위 상의어)의 수식어으로써 정의한 의미 소성(意味索性, semantic feature)에 의한 재분류 등에 의해서 개선될 수 있다^[29].

위와 같은 방법으로 문제점들이 해결된 '한국어 명사 의미계층구조'가 자연 언어 처리 과정에서 나타나는 한국어의 애매성을 해결하는 데 활용될 수 있게 하기 위해서는, '한국어 명사 의미계층구조'의 각 의미 영역과 용언과의 의미 호응 점검 작업 및 의미 호응에 의한 대응 관계를 정의하는 작업이 이루어져야 할 것이다.

참 고 문 헌

- [1] 금성출판사 사서부, "뉴에이스 국어사전", 금성출판사, 1991년.
- [2] 김영택, "자연언어처리", 교학사, 1994
- [3] 남영신, "우리말 분류 사전", 한강문화사, 1987.
- [4] 대중서관 편집국, "국어대사전", 도서출판 대중서관, 1979년.
- [5] 동아출판사 편집국, "동아 국민학교 새국어사전", 동아출판사, 1994년.
- [6] 동아출판사 편집국, "동아 새 국어사전", 동아출판사, 1994년.
- [7] 민중서림 편집국, "국민학교 민중 새국어사전", 민중서림, 1994년.
- [8] 박영자, 송만석, "자연언어 처리를 위한 한국어 동사·명사의 개념 분류", 1992년도 제4회 한글 및 한국어 정보처리 학술발표 논문집
- [9] 박용수, "겨레말 갈래 큰사전", 서울대학교출판부, 1973.
- [10] 신기철, 신용철, "새 우리말 큰사전", 삼성출판사, 1980년.
- [11] 심재기, 이기용, 이정민, "意味論序說", 집문당, 1994
- [12] 이승녕, "대국어사전", 현문사, 1983년.
- [13] 이행순, "세계백과대사전", 교육출판사, 1981년.
- [14] 이희승, "국어대사전", 민중서림, 1988년.
- [15] 정보·전자 연구회, "자연언어처리입문", 대광서림, 1993년.
- [16] 천시권, 김종택, "국어의미론", 형설출판사, 1977년.
- [17] 최현배, "우리말본", 정음사, 1975, pp.212.
- [18] 한국교육개발원, "중학교 국어 3-1", 대한교과서주식회사, 1991.
- [19] 한글학회, "우리말 큰 사전", 어문각, 1991년.
- [20] 문유진, "한국어 명사를 위한 WordNet의 설계와 구현", 한국어정보과학회, 1996.
- [21] 정연수, 조정미, 김길창, "개념분류를 이용한 한국어 명사 분류", 한글 및 한국어 정보처리 학술대회, 1995.
- [22] 박영자, 송만석, "자연언어 처리를 위한 한국어 동·명사의 개념 분류", 한글 및 한국어 정보처리 학술대회, 1992.
- [23] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller, "Introduction to WordNet: An On-line Lexical Database" in Five Papers on Wordnet, CSL report, Cognitive Science Laboratory, Princeton University, 1993.
- [24] N. Ide and J. Véronis, "Knowledge Extraction from Machine-Readable Dictionaries: An Evaluation", Lecture Notes in Computer Science, Vol.898, pp.19-34, 1995.
- [25] Chomsky N., "Language and Mind", Harcourt Brace Jovanoovich, Inc., 1972.
- [26] Gutrie, L. and Bruce, R. "Genus Disambiguation: A Study in Weighted Preference", Proc. of COLING-92, pp.1187-1191. Aug. 1992.
- [27] Hernet, P., "KASSYS: A Definition Acquisition System in Natural Language", Proc. of COLING-94, pp.263-267, Aug. 1994.

- [28] Montemagni, S. and Vanderwende, L., "Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries", Proc. of COLING-92, pp.546-552, Aug. 1992.
- [29] Pyeong-Ok Cho, Cheol-Yeong Ock and etc., "A Korean Noun semantic Hierarchy Based on Semantic Features", Proceedings of the Eighteenth International conference on Computer Processing of Oriental Languages, Vol.1, 1999.
- [30] N. Ide and J. Véronis, "Refining taxonomies extracted from machine-readable dictionaries", Offord university Press.