

의미자질을 고려한 명사 어휘의미망 구축¹⁾

- 남/여성 자질을 중심으로

황순희, 윤애선 (부산대학교)

1. 들어가는 말

자연어를 이해하는 시스템 개발을 위해서는 무엇보다도 인간이 지닌 지식 베이스(knowledge base)가 요구되는데, 이러한 지식 베이스를 전산학에 도입하여 보다 체계적인 성과로 나타난 것이 온톨로지(ontology)와 시소러스(thesaurus)이다. 이들은 개념 또는 의미 체계를 인공적 계층구조로 표상하도록 형성되는 것이 일반적이며, 어떤 영역에서 공통적으로 사용되는 의미 또는 개념들의 계층적 집합을 의미하게 된다. 최근 들어 이와 아울러 주목받고 있는 것은 어휘의미망(lexical semantic network)이다. 자연어 처리의 가장 기본단계인 형태소 분석, 구문 분석 등이 해당 분야에서 성과를 거두고 있다면, 보다 많은 변수가 존재하여 접근이 난해한 의미 분석은 많은 연구 과제를 남기고 있다. 이러한 상황에서, 어휘의미망 구축과 그 분석은 의미 분석을 위한 한 가지 대안으로 고려될 수 있다. 또한 어휘의미망은 인간의 두뇌에 어휘 지식이 저장되는 구조인 심상어휘집(mental lexicon)에서 착안하여 이와 유사한 구조로 의미 개념을 개별 언어의 어휘를 이용하여 데이터 베이스화 하였다는 특징이 있다. 그 중 가장 대표적이며 다른 언어로의 대역 작업을 통해서도 그 활용도가 높은 것으로 프린스턴(Princeton) 대학의 워드넷(WordNet, 이하 PWN)이다.

특정 개별어의 어휘의미망을 구축하는 방식은 한 개별어에만 고유한 어휘의미망을 독자적으로 구축하는 직접 구축 방식과 다른 언어로 이미 구축된 어휘의미망을 대역하여 구축하는 간접 구축 방식²⁾으로 구분된다. 먼저, 직접구축 방법은 단일어 사전(monolingual dictionary)의 정의를문을 참고로 하고, 핵심어(headword)를 중심으로 상하위 계층관계를 설정해 가는 상향식(bottom-up) 방법과 해당 전문분야 전문가의 전문지식을 이용한 하향식(top-down) 방법으

¹⁾ 직접 구축 방식의 대표적인 예로는 영어를 대상으로 한 워드넷(WordNet)이 있다. PWN은 프린스턴(Princeton) 대학의 G. Miller 중심으로 1985년부터 구축되기 시작한 어휘의미망으로 이를 바탕으로 여러 개 언어를 대상으로 한 어휘의미망 구축이 시도되고 있다. 또한 일본어와 중국어를 각각 대상으로 한 NTT와 하우넷(HowNet)도 직접 구축 방식의 대표적인 사례이다. 중국의 하우넷은 PWN처럼, 단일어인 중국어를 대상으로 구축된 어휘의미망이며 중국어 20개 품사, 대용량 코퍼스를 기반으로 빈도수를 고려한 8만 개의 어휘를 대상으로 한다. 기본 구조 역시 PWN의 계층구조와는 달리 그래프(graph) 구조를 갖는 특징이 있다.

²⁾ 간접 구축 방식의 대표적인 예로는 PWN의 대역어 작업을 기초로 구축된 유로워드넷(EuroWordNet), 발칸넷(Balkanet) 등이 있다. 유로 워드넷은 EU의 8개국어인 영어, 이탈리아어, 스페인어, 네덜란드어, 독일어, 프랑스어, 에스토니아어, 체코어로 구성되어 있으며, 발칸넷은 발칸 반도 인접 8개국어인 불가리아어, 체코어, 그리스어, 루마니아어, 터키어, 세르비아어로 되어 있다. 이밖에도 NTT의 시소러스를 대역하여 한국어, 일본어, 중국어를 연동하는 KAIST의 코어넷(CoreNet) 등이 있다.

로 양분된다. 상향식 구축 방식의 경우, 계층관계 설정에 비교적 객관성과 일관성을 유지할 수 있다는 점과 구축시간이 상대적으로 짧게 소요된다는 장점이 있다. 반면, 사전의 정의문이 항상 만족스러운 정보를 주는 것이 아니다. 정의문 자체가 순환적일 경우, 또는 사전적 정의가 정확하지 않은 경우가 단점으로 지적된다. 한편, 하향식 구축방식의 경우 해당 전문가의 전문지식을 이용하여 구축되므로, 보다 세분화된 계층관계 설정이 가능하다는 큰 장점이 있지만 전문가라 하더라도 계층관계 설정 시, 자의성과 비일관성이 개입될 여지가 크고, 구축에 긴 시간이 요구된다는 점이 큰 단점으로 지적될 수 있다. 간접구축 방법은 기구축된 어휘의미망을 대역하여 구축하는 방법이다. 가령, PWN, NTT, HowNet 등을 기반으로 대역과정을 통해 한국어 어휘의미망을 구축하는 방법이 그 한 예이다. 간접구축 방식의 1단계는 PWN의 표제어를 영한 이중어 기계가독형 사전(Machine Readable Dictionary)을 이용하여 자동으로 사상(mapping)하는 것이다. 자동사상은 방대한 어휘의미 관계를 비교적 신속하게 처리할 수 있다는 장점이 있으나, 본질적으로 잘 알려진 동형이의어 구분, 의미 세분화, 언어간 의미 및 개념 차이 등 많은 문제점을 안고 있어, 자동 사상의 오류여부 확인 및 선택된 대역어의 적합성 여부는 반드시 정제³와 검증⁴을 거쳐야 한다. 또한 간접 구축 방법의 큰 어려움 중 하나는 두 언어간 의미 계층구조가 대부분의 경우 정확히 일치하지 않는다는 점이다. 따라서 이러한 문제점 극복을 위해, 각각의 개별어에 존재하는 의미 관계 구조 및 어휘 세분화 등의 문제도 다루어져야 한다.

본 연구는 간접구축 방식으로 한국어 어휘의미망(KorLex, 이하 KL) 구축 시, 신셋(synset)⁴을 보다 효과적으로 기술하고 구축할 수 있는 준거 설정을 위해 시작되었다. 이를 위해, 명사⁵의 특정 부류 신셋을 구축함에 있어 관여적 기준으로 판단되는 의미자질(semantic feature) 개념을 도입한다. 의미론에서 항상 논의되는 ‘의미자질⁶’ 중 하나인 남/여성 자질이 어휘의미망 구축에 관여적 기준이 될 가능성은 다음 두 가지 이유에서 비롯된다. 첫째, 어휘의미망 구축에서 선결되어야 할 과제 중 하나는 보다 완전하고 정교한 온톨로지⁷의 설정이나, 이를 개별어의 어휘의미망으로 구축할 경우 온톨로지는 ‘어휘화’ 과정에서 경우에 따라 수

³ 간접구축된 어휘의미망의 정제와 검증 방법에 관해서는 황순희, 윤애선 (2005)을 참고하라.

⁴ 개별어에서 ‘어휘’는 구체적으로 처리(processing)가 가능한 단위이지만, ‘동음이의 관계’나 ‘다의 관계’ 등 하나 이상의 어휘-의미를 가질 수 있다. 어휘의미망을 구성하는 기본 단위는 어휘의미이며, 동의 관계를 갖는 어휘의미 집합을 동의어 집합(synonym set: synset)으로 부른다. 본 논문에서도 PWN의 용어를 그대로 빌어 신셋이라 지칭하기로 한다.

⁵ 본 연구의 대상인 영어 PWN 2.0 version의 명사는 총 79,689 개 신셋(synset, 동의어집합)과 141,690 개 어휘로 구성되어 있으며, 2005년 3월 발표된 2.1 version 은 명사 81,426 개 신셋과 117,097 개 어휘로 구성되었다. 특기할 만한 것은 2.0 version 과 달리 2.1 의 모든 명사는 최상위어로 {entity} 가 설정되어 있다

⁶ 논의 중인 ‘남/여성 자질’ 이외에, 의미론이 제공하는 ‘생물/무생물’, ‘인간/동물’ 등의 전통적인 의미자질에서 착안하여 ‘존칭관계’, ‘언어사용역(language register)’, ‘수사학적 방법’ 등을 의미자질로 삼아 어휘의미망을 구축하는 한 가지 기준으로 삼을 수도 있다

⁷ 온톨로지(ontology)란 어떤 영역에서 공통적으로 사용되는 개념들의 집합을 의미한다. 최근 들어, 온톨로지에 관한 관심이 높아진 이유는 시맨틱 웹(semantic web)의 등장과 지식 체계화의 한 방법으로 대두되면서인데, 자연어 처리에도 이를 잘 활용하면 형태/구문 분석 이상의 처리 단계 및 결과를 기대할 수 있다. 반면, 어휘의미망(lexico-semantic network)이란 어휘화된 개념들을 상의(hyperonymy), 하의(hyponymy), 전의(holonymy), 분의(meronymy), 함의(entailment) 등의 상관적 관계 용어(RT related term)를 이용하여, 상하위 계층관계(hierarchical structure)가 고려된 망 구조로 연결시킨 것이다

정, 변경이 불가피하다. 따라서, 어휘의미망 구축 시 개별어에 관여적일 수 있는 준거의 마련이 선행되어야 한다. 둘째, PWN을 기반으로 KL을 구축하는 것은 1차적으로 PWN 신셋의 의미 및 계층구조를 여과 없이 수용함을 의미하지만, 영어와 한국어의 의미계층구조가 항상 정확히 일치하지 않음을 감안해야 한다. 이 경우 의미 개념에 포함된 남/여성 자질의 고려하면, 보다 적절한 어휘의미망 구축이 가능하다. PWN의 {businessperson 1⁸}, '사업가'를 예로 들어보자. PWN은 이 개념의 하위 개념을 각각 {businessman 1}과 {businesswoman 1}로 분화시켰다. 그러나, 이 계층구조를 여과 없이 수용할 경우 KL에는 '남성 사업가', '여성 사업가'라는 개념이 '사업가'의 하위 개념으로 구축되는 부적합한 관계가 형성될 것이다. 이것은 한국어에 '사업가'는 오랫동안 남성에게만 국한된 직업 개념이었고, '여성 사업가'가 등장하면서 '사업가'로부터 '여성 사업가' 개념 및 어휘가 파생되었다고 보는 것이 타당하기 때문이다. 따라서, KL 구축 시 '남성 사업가' 개념을 독립시키지 않고, '여성 사업가'를 '사업가'의 직속 하위 개념으로 연결하는 것이 보다 적절하다.

실제로 개념들 간의 계층구조(hierarchical structure)는 PWN의 근간이 되어 다각도로 논의의 대상이 되어왔지만, 개별 신셋들의 의미 자질과 관련된 의미적 접근은 이루어진 적이 없다. 본 연구는 특정 개념(의미 또는 신셋)을 의미론적으로 접근할 수 있는 기본적인 방법 중 하나인 구성적 어휘 의미론(componential lexical semantics)적 접근 방식에서 착안하였다. 즉 의미 분석을 위해서 특정 신셋의 의미를 그 구성 요소인 의미 자질을 기준으로 분석하고 그 자질의 유무를 따져 분류, 접근해보고자 하는 것이다.

본 논문은 다음과 같이 구성된다. 먼저, 2장에서는 남/여성 자질이 포함된 개념 구조의 유형을 고찰하고, 영어와 한국어의 어휘화 형성 규칙을 살펴본다. 또한 본 연구의 대상과 범위를 제시한다. 3장은 PWN의 남/여성 자질이 포함된 신셋들을 분석하고, 문제점을 분석한다. 4장은 KL 구축 시, 관련 신셋들에 남/여성 자질을 준거로 적용하여 구축하는 방식을 소개한다.

2. 남/여성 의미 개념의 구성방식

본 장에서는 남/여성 의미 자질이 포함된 개념 또는 의미 단위의 계층화, 남/여성 분화 문제와 이와 관련된 어휘화 형성 규칙에 관해 살펴본다. 이러한 개념을 계층 구조화하는 방식에는, 선험적(a priori)으로 몇 가지가 가능하다.⁹ 또한 고려해야 할 사항은 남/여성 자질이 포함된 개념의 구조가 영어, 한국어 등 상이한 개별어에서 각각 다르게 구현된다는 점이다. 즉 이들 개념이 어떻게 어휘화되는가를 어휘 형성 규칙을 통해 살펴본다. 2.2에서는 남/여

⁸ 이 숫자는 PWN의 신셋을 구성하는 표제어의 의미분화(sense differentiation)를 위한 번호로, 동일 어휘가 상이한 신셋에 사용됨에 따라 번호가 부여된다. 논고에서는 편의상 PWN의 표제어는 { }으로, 의미자질은 []을 이용 [+male] 등으로 표기한다

⁹ 물론 개념 구조를 바탕으로 한 어휘화 문제는 해당 개념에 관한 유형화 방식과 사회 언어학적 요인, 화용적 요인 등이 복합적으로 반영된 문제이기도 하다

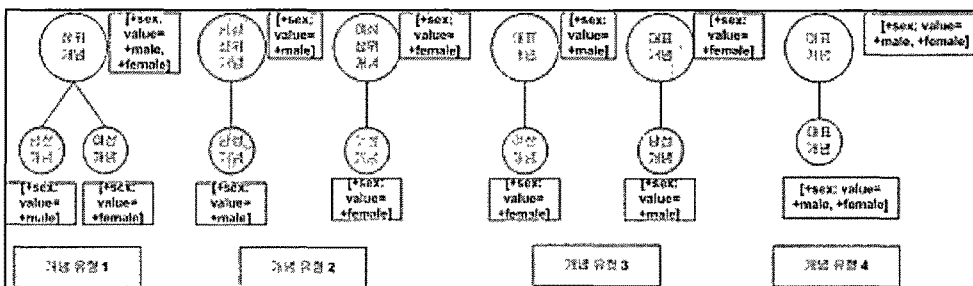
성 개념이 포함된 신셋들과 관련하여, 본 연구의 대상 및 범위를 논의한다.

2.1 기본 개념구조의 유형

먼저 자질의 승계 문제를 고려해야 한다. PWN의 기본이 되는 상하위 관계 및 계층구조는 자질승계 체계(inheritance systems)¹⁰에 근간한다. 계층구조는 하나의 상위어로부터 출발하여 개별적인 의미자질이 하향 승계 되면서 의미구조를 형성한다. 따라서, 특정 상위어는 총체적이고 보편적인 의미자질을 하위어에 그대로 물려주고(상위어 자질), 하위어는 승계 받은 의미자질과 자신과 직속 상위어를 구별해 줄 자질(구별자, distinguishers)을 적어도 하나 이상 추가하여 갖는다. 신셋 {mermaid 1}은 직속 상위어 {imaginary being 1, imaginary creature 1}의 자질들을 승계 받고, 상위어와의 구별자인 ‘half woman and half fish; lives in the sea’ 라는 의미자질을 소유한다.

그런데, 자질 승계와 관련하여 직접구축 방식에서 잦은 오류를 범하는데, 신셋의 의미가 항상 명시적이지 않으며, 비일관적으로 제시될 수 있다는 점이다. {tree}의 의미는 ‘a **tall perennial woody** plant having a main **trunk** and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms’ 인데, ‘tall, perennial, woody, and has a trunk’ 등은 하위어들과 공유할 {tree}를 설명하는 의미 자질이다. 문제는 여기서 {tree}를 정의하는 여타 의미 속성, 가령 ‘having roots, cellulose walls’ 등은 명시되지 않았고, 이러한 정보는 개인 스스로 복원해야 한다. 이처럼 상위어 정의에 하위어와 공유하는 의미 자질이 명시되지 않는다면 하위어와 상관 관계 설정에 문제가 생길 수 있다.

남/여성 자질을 포함한 개념 또는 의미 단위들을 계층구조화 함에 있어, 어휘화 형성 규칙 또는 사회적, 문화적 요인이 배제된 상태에서 이를 상정하면 다음 네 가지 유형이 가능하다.



<그림1. 남/여성 자질이 포함된 개념 구조에 관한 기본도식>

¹⁰ 논의 중인 계층구조의 표상(hierarchical representations)은 정보검색 시스템(information retrieval systems) 구축에서 널리 사용되는 것으로, Touretzky(1986)는 이러한 관계를 “승계 체계(inheritance systems)”라고 명명한다(Miller 1990 8)에서 재인용)

① 개념 유형 1: 상위 개념 => 남/여성 개념 양분화

[+sex: value = +male, +female] => [+sex: value = +male], [+sex: value = +female]

첫 번째 개념 유형 (유형1)은 [+male, +female]을 공유한 상위 개념으로부터 남/여성 개념의 분화가 가능한 유형이다. 이 경우 남/여성 자질 개념의 의미화는 독립적으로 이루어지고, 상위 개념에 공통된 의미자질이 승계된다. 이 유형은 남/여성 개념과 관련된 가장 단순한 형태로, 예를 들어 {자식}은 [+male, +female]을 공유하며, {아들}[+male]과 {딸}[+female]로 분화된다.

② 개념 유형 2: 상위 개념 => 동일 자질 승계

[+sex: value = +male] => [+sex: value = +male]

[+sex: value = +female] => [+sex: value = +female]

유형2는 남/여성 상위 개념의 자질이 하위 개념에 동일하게 승계된다. 이때 하위 개념은 상위 개념의 세분화된 개념이며, 상위 개념 각각은 보다 상위인 일정 개념에 의해 상호 연관성을 갖는다. 가령, {독신남}은 상위 개념인 {남성}의 자질을, {독신여성}은 {여성}의 자질을 승계 받는다.

③ 개념 유형 3: 상위 개념(대표 개념) => 성((姓) 자질 전환, 승계

1- [+sex: value = +male] => [+sex: value = +female]

2- [+sex: value = +female] => [+sex: value = +male]

유형3의 상위 개념은 남/여성 개념을 모두 포괄하는 대표 개념이다. 이 경우, 처음에는 대표 개념 외의 독립된 다른 성의 개념은 상정할 필요가 없다. 그러나, 일정 기간을 거쳐 사회적, 문화적 요인으로 인해 대표 개념으로부터 여성 또는 남성 개념 분화¹¹의 필요성이 생겨나 이들이 분화된다. 이 경우 상위 개념의 남/여성 자질은 하위 개념에서 전환, 승계된다. 여기서 성 자질이 전환된다 함은 {의사}, {변호사}, {교사} 등은 대표형 [+male] 자질로부터 화용적 이유에서 여성개념이 분화된다. 즉, {의사}, {변호사}, {교사} 등은 {여의사}, {여성 변호사}, {여교사}로 분화된다. 반면, 대표 개념이 [+female] 자질인 {간호사}, {부모} 등은 [+male] {남자 간호사}, {남자 부모}로 전환될 수 있다.

¹¹ 개념 유형 3의 대부분의 경우, 대표개념은 '남성 개념'이고, '여성 개념'이 이로부터 분화된다. 그렇더라도, 대표 개념이 '여성 개념'이고, '남성 개념'이 분화된 경우도 있다

④ 개념 유형 4: 상위 개념(대표 개념) => 대표개념

[+sex: value = +male, +female] => [+sex: value= +male, +female]

[+male, +female]을 공유한 상위 개념으로부터 남/여성 개념 분화가 불필요하다고 판단되는 경우이다. 즉, 남/여성 개념을 분화하지 않고, 이 둘의 포괄 가능한 대표 개념을 유지한다. {조상}, {귀족} 등은 남/여성 개념으로의 분화가 불필요하므로, 이들 대표 개념을 유지하게 된다.

2.2 어휘화 형성 규칙

개별어에서는 전술된 기본 개념 구조와 아울러 고려해야 할 점은, 남/여성 자질을 지닌 의미의 개념화가 어휘화 규칙과도 밀접한 관련이 있어 보인다는 점이다. 하나의 개념이 개별어에서 어휘화되는 과정은 일반화 될 수 없는 개별어에만 국한된 고유한 언어 규칙에 속한다. 즉 남/여성 자질 개념이 실현되는 방식은 개별어마다 다르다. 가령, 프랑스어의 자연성(sex)¹²과 관련하여, 남/여성 개념은 주어, 목적어, 속사 등 문법적 구성 성분에 그대로 반영되어 실현될 뿐 아니라, 명사는 물론이고 형용사 등의 특정 품사를 통해서도 직접 실현된다. 한편 영어에는 남/여성 개념이 프랑스어 보다는 다소 제한적으로 반영되어, 실현된다. 문법 성분을 고려하면 여성형 대명사의 사용을 들 수 있고, 특정 품사와 관련하여서는 여성 개념이 반영된 독립된 어휘를 단독으로 사용하는 경우와 여성 표시 접사를 첨가하는 것이 전형적인 방법이다. 반면 프랑스어 및 영어에 비해, 한국어에는 남/여성 개념이 문법적으로 반영되는 경우가 상대적으로 드물다. 문장 내 구성성분이나 형용사 등의 특정 품사를 통해서도 실현되는 경우가 없다. 한국어에서 여성 개념이 실현되는 유일한 방법은 여성형 명사를 사용하는 것이다.

이제 일반적으로 사용되는 남/여성 명사 형성 규칙을 논의하기로 한다.

영어의 남/여성 명사가 형성되는 규칙은 다음 세 가지로 분류된다. 첫째, 남성명사에 여성 표시 접미사를 붙이는 것으로 prince ↔ princess, hero ↔ heroine, waiter ↔ waitress 등이 이 규칙에 의해 형성되었다. 두 번째 규칙은 하나의 공통된 어휘를 중심으로 성(gender)을 나타내는 이형의 어휘가 첨가된다. boy-friend ↔ girl-friend, man-servant ↔ maid-servant 등이 이 분류에 속한다. 마지막 규칙은 남/여성 명사가 전혀 다른 이형의 어휘를 단독으로 사용하는 것이다. son ↔ daughter, man ↔ woman, brother ↔ sister, husband ↔ wife 등의 쌍이 이에 속한다.

¹² 주지하는 바와 같이 프랑스어에는 문법성과 자연성이 따로 존재하며, 본고의 논의의 대상은 자연성에만 국한된다.

한국어¹³의 경우, 크게 두 가지 규칙이 있다. 첫째, 사장 ↔ 여사장, 교사 ↔ 여교사, 의사 ↔ 여의사 등과 같이 여성임을 나타내는 접두사 또는 접미사를 첨가하는 방법이다. 접두사의 경우 '여'의 이형태(allomorpe)¹⁴인 '여자', '여류', '여성' 등도 어휘에 따라 첨가될 수 있다. 어휘에 따라 여성 접두사를 첨가하는 방식에 특별한 규칙은 없으나, 가령 '여류'는 어떤 전문적인 일에 능숙한 여자를 의미하여 '여류 소설가', '여류 비행사', '여류 문인', '여류 수필가', '여류 시인', '여류 화가' 등을 허용한다. 또한 '여', '여자'는 여성의 일반직종 또는 신분과 관련된 어휘에 첨가되어, '여종업원', '여자 종업원', '여백작', '여자 백작' 등이 가능하고, 둘 중 하나만 허용하는 어휘도 있다. 가령 '여자 거지'는 가능하나, '여거지'는 허용되지 않는다. '여성'의 어휘 첨가에도 일반화된 규칙은 없으나 여성의 전문직종과 관련된 어휘에 첨가되는 경향을 보인다. '여성 국회의원', '여성 최고경영자' 등이 그 예이며, '여자 국회의원', '여자 변호사', '여의사' 등도 함께 허용된다. 접미사를 첨가한 예로는 소년 ↔ 소녀, 손자 ↔ 손녀, 동정녀, 약혼녀, 유부녀, 이혼녀 등이 있다. 두 번째 규칙은 형태적 유사성이 전혀 없는 어휘를 단독으로 사용하는 것인데, 아들 ↔ 딸, 아버지 ↔ 어머니, 삼촌 ↔ 이모, 총각 ↔ 처녀 등이 이 유형에 속한다.

2.3 연구 대상 및 범위

이제 본 연구의 대상 및 범위를 언급한다. PWN 명사는 총 79,689 개의 신셋으로 구성되어 있으며, 이들은 총 25¹⁵개의 세부 분야로 구분된다. PWN 명사는 9개의 최상위어(root)들로부터 의미자질의 상속(lexical inheritance) 유무에 따라 최고 10~12 단계(depth)를 거쳐 말단 노드(leaf node)로 연결된다. 본 연구의 직접적 대상은 인물명사(noun.person) 내에 속한 신셋들 중 남/여성 자질을 지닌 신셋들로 국한된다. 인물명사는 총 10,666 개의 신셋이 든 방대한 의미 집합으로, 신셋들의 의미적 특성을 살펴보면 크게 ①특정 인물과 관련된 '고유명사', ②개인의 '직업명', '신분명', ③'한 개인의 특성에 대해 기술하는' 의미(예: {geezzer 1, bloke 1} '괴상한 노인', {ball-buster 1, ball-breaker 1} '까다로운 여자' 등)의 신셋들로 구성됨을 알 수 있다. 또한, 인물명사에 속하는 신셋들은 특징적으로 의미론에서 전통적으로 논의되는 남/여성 의미자질과 깊은 상관성을 보인다. 그리고, 특별히 고유명사의 경우, 남/여성자질이 글로스(gloss)¹⁶ 내에 명시되어 있더라도 본 연구대상에서는 제외하기로 한다. 그 이유는 예를 들어, 물의 요정인 {nymph 1} '님프'¹⁷에는 분명 여성 자질 [+female] 이 명시돼 있으나, 이에 대한

¹³ 한국어 인물명사는 사람 일반, 친척, 직업으로 분류되거나, 고유명사(proper noun), 친족명사(kinship noun), 직업명사(occupational noun)로 구분된다 임지룡 (1991), 히데키 (2002)

¹⁴ 여성 자질 개념을 명시화하는 접사인 '여', '여자', '여성', '여류'의 분포(distribution)의 문제는 본 논문에서는 논외로 한다

¹⁵ 25개 세부 의미 분야는 다음과 같다. {act, animal, artifact, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time}

¹⁶ PWN 브라우저 상 신셋의 구조는 다음과 같다 가령, 신셋 {child, kid}은 이 신셋의 설명(gloss)에 해당하는 'a human offspring (son or daughter) of any age, "they had three children"; "they were able to send their kids to college" '와 상위, 하위, 분의, 전의, 파생 등의 의미관계로 기술된다

¹⁷ '님프'에 대한 글로스(gloss)는 "(classical mythology) a minor nature goddess usually depicted as a beautiful maiden, "the ancient Greeks believed that nymphs inhabited forests and bodies of water" 으로 여성적 자질이 명시되어 있다

남성형을 상징할 필요가 없기 때문이다.

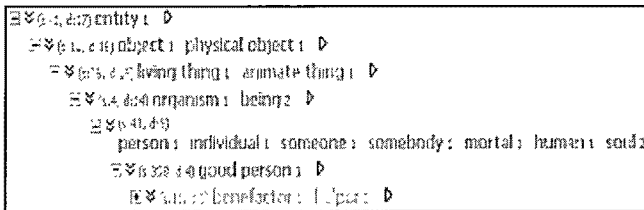
본 연구는 전체 10,666개 중 특정 인물명(인명 또는 신(神)의 이름)인 고유명사를 제외한 1,334개(전체 대비 12.51%) 신셋을 1차 연구 대상으로 삼았다. 이들을 다시, 남성 자질만을 가진 신셋 296개, 여성 자질만을 가진 신셋 360개, 남/여성 자질을 공유한 신셋 23¹⁸개와 남/여성 자질이 명시되지 않은 신셋 655개로 분류할 수 있었다. 이들 중, KL 구축 시 실제 적용대상으로 삼은 것은 남/여성 자질이 명시되지 않은 신셋¹⁹ 655개를 제외한 679개 신셋이다.²⁰

3. PWN의 남/여성 자질 구현 방식

본 3장에서는 2장에서 살펴본 남/여성 자질 개념의 구조가 PWN에서 어떻게 실현되는지 검토한다. 이를 위해 PWN에서 추출한 남/여성 자질이 포함된 신셋을 검토 분석하고, 의미구조의 특징에 따른 유형화를 시도한다. 아울러, 신셋의 분석 결과를 토대로 PWN 남/여성 자질의 구현방식에 관한 비판의 논거를 제시하여 KL 구축 시 중요한 증거 자료로 삼고자 한다.

3.1 신셋의 구조

다음은 남/여성 자질이 포함된 신셋의 구조를 나타낸 그림이다.



<그림2. 'a person who~' 로 설명되는 [+male]을 지닌 신셋>

또한, PWN에서 남/여성 의미자질을 지닌 신셋의 특징은 다음 두 가지로 요약된다.

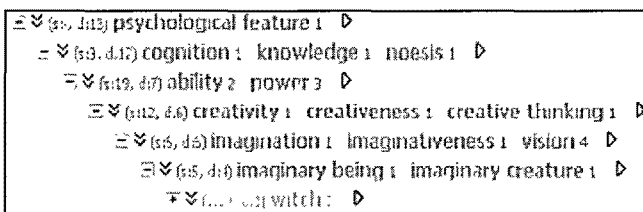
¹⁸ {adolescent 1, strpling 1, teenager 1}, {parent 1}, {godparent 1}, {stepchild 1} 등의 신셋이 이에 속한다.
¹⁹ 논의 중인 655개 신셋에는 남/여성 자질이 명시되지 않았고, 대표적인 단독 표제어로만 표현되었다 또한 이들은 개인의 직업 또는 신분과 관련된 의미가 주를 이룬다. 예를 들면, {professor 1}, {college student 1}, {radio announcer 1}, {dressmaker's model 1}, {lawyer 1}, {doctor 1}, {crewman 1}, {chef 1}, {protagonist 2}, {academician 2}, {teacher 1} 등이 포함되 있는데 이들은 KL 구축에서 남/여성형으로 분화 가능한 의미들이다. 또한 이 655개 신셋 중 남/여성 분화가 요구되지 않는 상당수 신셋이 있는데, 이들에 관해 본 연구에서는 분화하지 않는다 {dental technician 1} '치과 기공사', {caster 1} '주물공', {bonesetter} '접골사' 등이 그 예이다.
²⁰ PWN의 관련 신셋에 대한 대역어 이외에 KL 구축 시 허용하는 한국어 남/여성 자질 관련 어휘 풀(pool)은 다음으로 구성된다. ① 표준 사전, ② 산업자원부 구축 직업명 사전, ③ 대용량 말뭉치, ④ 사전편찬자(언어학 박사 3명, 학사 이상 4명)의 직관

① 신셋 설명(gloss)의 표현방식

② 계층구조의 공통성

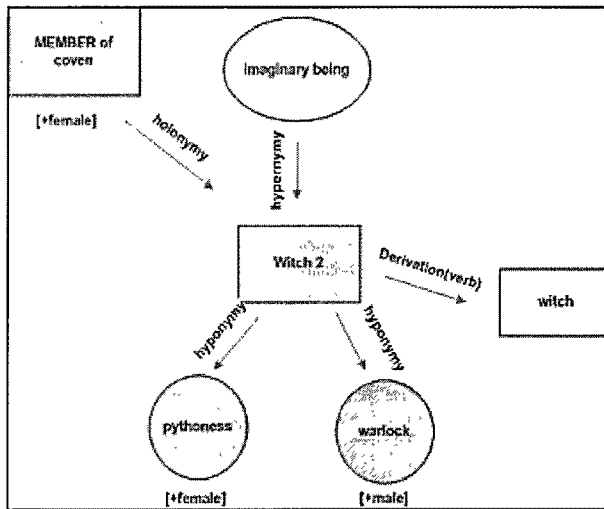
첫째, 남성 [+male] 자질을 지닌 신셋의 경우 'a man~', 'a male~', 'a boy~'로 설명(gloss)되거나, 많은 경우 'a person~', 'someone~', '~any person', 'any of ~', 'one who~' 등으로 설명된다. 따라서, 해당 신셋의 의미가 [+male], [+female] 자질을 모두 갖는 대표형의 의미인지 또는 [+male] 자질하고만 관련이 있는지 불분명한 경우가 있다. 위에 제시된 {benefactor 1}도 이에 해당하는데, 이에 관해서는 다음 두 가지를 유추할 수 있다. 첫째, {benefactor 1}에 [+male] 자질만을 가진 노드가 따로 분리되어 있지 않다면 {benefactor 1}은 남/여성을 대표하는 대표형 일 것이다. 둘째, 만일 {benefactor 1}의 반의²¹ 개념인 [+female] 자질을 포함한 개념이 독립적으로 존재한다면(e.g. benefactress), {benefactor 1}의 신셋은 [+male] 자질만을 지닌 단독형일 뿐만 아니라 남/여성을 포괄하는 대표형이기도 하다. 반면, 여성[+female] 자질을 지닌 신셋에는 'a woman~', 'a female~', 'a girl~' 등의 '여성'임이 명시적으로 나타난 설명이 주어지거나, 여성에게만 국한된 직업명(e.g., {stewardess 1} 'a woman steward on an airplane', {magdlen 1} 'a reformed prostitute')이 들어 있다.

둘째, 남/여성 자질이 든 신셋의 계층구조를 검토하면 대부분의 신셋이 최상위 노드 {entity 1}로부터 {object 1, physical object 1}, {living thing 1, animate thing 1}, {organism 1, being 2}, {person 1, individual 1, someone 1, somebody 1, mortal 1, human 1, soul 2}까지의 노드를 공유하며, 이 노드 이후 의미 분화가 일어난다. 반면, 이와는 다른 의미 계층구조를 갖는 신셋들은 소수로 확인되며, 이들의 최상위어는 주로 {psychological feature 1}, {animal, animate being 1, beast 1, brute 2}이며, '가상의 인물' 개념을 가진 신셋들이다.



<그림3. 최상위어가 {psychological feature 1}인 여성 명사의 계층구조>

²¹ 반의(antonymy) 관계란 형태상으로 서로 다른 두 어휘가 의미적인 면에서 반의어적 대조를 이루는 경우를 의미한다 또한 반의 관계는 성격상 여러 가지로 분류가 가능한데 Lyons(1977: 270-290), '남/여성'의 반의관계는 동시에 동일한 진리가를 가질 수 없는 모순대조(contradictory) 관계에 있다 가령, benefactor-benefactress는 모순 대조관계에 있는 반의어인데, PWN에는 일부 신셋만을 반의 관계로 설정했고 이들 신셋의 반의 관계는 설정되지 않았다.

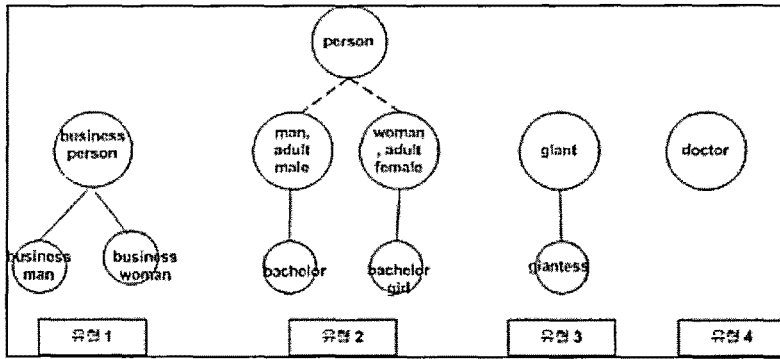


<그림4. witch 2 의 의미관계 구조>

{witch 2}의 의미관계는 위 그림으로 구조화된다. {witch 2}는 직속 상위어(immediate hypernym)로 {imaginary being 1}을 가지며, 하위어로 {pythoness 5}와 {warlock 1}을 갖는다. 여기서 흥미로운 사실은 {witch 2}의 의미가 ‘a being (usually female) imagined to have special powers derived from the devil’로 여성 자질을 가지고 있는데, 이와 모순되게 하위어인 {warlock 1}은 ‘a male witch or demon’ 의미로 남성자질을 갖는다는 사실이다. 즉, 상위어가 지닌 여성자질이 승계되지 않았다. 이는 PWN에서 상위어가 갖는 의미자질을 하위어는 승계 받는다는 대원칙에 위배되는 사항이기도 하다. 이밖에 {witch 2}는 {coven 1}을 전의어로 가지며, 동사 {witch}와도 파생관계에 있다.

3.2 신셋의 유형

PWN의 남/여성 자질이 포함된 신셋들은 다음 네 가지 유형으로 분류된다.



<그림 5. PWN의 남/여성 자질이 포함된 신셋의 유형화>

① 첫 번째 유형(유형1)은 남/여성 명사형이 독립적으로 존재하며, 공통된 자질을 부여 받은 상위어의 직속 하위어가 되는 구조이다. 특히 이 유형에 속한 신셋의 표제어는 어휘 형성 규칙면에서 볼 때, 대표형에 접사가 추가되는 단순한 방식이 아니라, 주로 이형태의 어휘가 합성된 형태가 된다. {merman 2}-{mermaid 1}(남자인어-여자인어), {incubus 1}-{succubus 1, succuba 1}(남자몽마-여자몽마), {foster-father 1}-{foster-mother 2}(양아버지-양어머니) 등이 이 유형에 속한다.

② 두 번째 유형(유형2)은, 관련된 남/여성 명사형이 각기 다른 계보로부터 내려온 유형이다. {bachelor 1, unmarried man 1} - {bachelor girl 1}(독신남-독신녀), {grass widower 1, divorced man 1} - {divorcee 1, grass widow 1}(이혼남 또는 별거중인 남자-이혼녀 또는 별거중인 여자), {housefather 1} - {housemother 1}(사감-여사감) 등이 이 유형에 속한다. 그런데, 여기서 계보가 다르다 함은 noun.person에 속하는 대부분의 신셋들이 {person 1, individual 1, someone 1, somebody 1, mortal 1, human 1, soul 2}을 공유한다는 의미에서, 최상위 노드를 비교하는 것은 무의미하고 {person 1}의 하위 노드를 비교하는 것만을 고려한다. 따라서, 상이한 계보에서 분화되었다는 것은 해당 남/여성 명사가 직속 상위어(부모 노드)나 또는 그 윗 단계 상위어(조부모 노드)를 고려하여 한번 이상의 분화가 더 일어난 경우를 의미한다. {bachelor 1}는 상위어로 {man 1, adult male 1}을, {bachelor girl 1}은 상위어로 {woman 1, adult female 1}을 갖고, 그보다 상위어인 {person 1, individual 1, someone 1, somebody 1, mortal 1, human 1, soul 1}을 상위 개념으로 공유한다.

또 다른 예인 {housefather 1}과 {housemother 1}은 흥미로운 계층구조를 갖는다. {housefather 1}는 {person 1}→{male 2}→{man 1}→{housefather 1}로, {housemother 1}는 {preserver 3}→{defender 1}→{chaperon 1}→{housemother 1}로 연결되어서, 표면적으로는 이 둘 간에 상호 관련성이 없어 보인다. 이러한 경우도, 서로 상이한 계보로부터 분화된 경우라 할 수 있는데 이 둘의 gloss를 비교해 보면 의미의 상호 관계성이 확인된다.

{housefather 1} : a man in charge of children in an institution

{housemother 1} : a woman employed as a chaperon in a residence for young people

③ 세 번째 유형(유형3)은 남/여성 명사의 대표형이 상위어로 존재하며, 이로부터 여성형이 분화된 경우이다. 이 유형은 당연값(default) 남성 자질이 여성 자질로 교체되었다고 할 수 있다. 가령, {giant 6}(거인)는 여성형 {giantess 1}와, {ogre 2}(식인거인)은 여성형 {ogress 1}와, {Jew 1, Hebrew 2, Israelite 2} - {Jewess 1} 등이 상하위어 관계에 있다. 그런데, 여기서 상위 개념인 {giant}, {ogre}, {Jew 1} 등은 남성형이라고만 단언할 수 없다. {Jew}의 글로스를 보면 ‘a person belonging to the worldwide group claiming descent from Jacob (or converted to it) and connected by cultural or religious ties’ 로 [+male]만을 가지는 것이 아니기 때문이다. 반면 {Jewess 1}는 ‘a woman who is a Jew’ 의 뜻으로 [+female]이라는 의미자질을 갖는다.

④ 네 번째 유형(유형4)은 PWN에서 확인된 독특한 유형으로 하나의 개념이 단독으로 남/여성의 의미를 모두 포괄하는 유형인데, 이 유형은 다시 두 가지로 구분된다.

④-① 유형은, {professor 1}, {doctor 1}, {lawyer 1} 등으로 대표되는 경우로 이들 어휘에 상응되는 여성개념인 ‘여의사’, ‘여교수’, ‘여자 변호사’ 등의 특수 직업군과 관련된 여성 개념이 가능하나 독립적으로 존재하지 않는다. 즉, PWN은 이들 개념에 대해서 남/여성 분화를 시키지 않았다. 그러나, 이것은 실제 언어 생활에 비춰볼 때 이의의 소지가 있다.

④-② 유형은 {lover 3}, {spouse 1} 등과 같이 역시 남/여성 자질을 공유한 신셋들이다. {lover 3} ‘a significant other to whom you are not related by marriage’. {spouse 1, partner 1, married person 1, mate 4, better half 1} ‘a person's partner in marriage’, {body servant 1} ‘a valet or personal maid’ 등으로 남/여성 자질을 공유한다. 또 다른 예는 {bondsmen 2, bondswoman 1} ‘someone who signs a bond as surety for someone else’으로, 남/여성형의 분화된 독립 신셋이 존재하나 이 둘을 단일 신셋에 넣어 남/여성 자질 공유의 독립된 신셋을 생성하고 있다.

이상과 같이 파악된 네 가지 유형에 속하는 신셋의 분포율은 다음과 같다.

	유형 1	유형 2	유형 3	대립되는 성(性)부재	유형 4	합계
[+ male}을 지닌 신셋	57	65	55	119		296
[+ female}을 지닌 신셋	64	72	110	114		360
[+ male, + female]	-	-	-	-	23	23

공유한 신셋						
합계	121	137	165	233	23	679
분포율 % 전체(679 개)	17.82	20.17	24.30	33.82	3.38	100

<표1. PWN 남/여성 자질명사의 유형별 분포율>

3.3 비판의 논거들

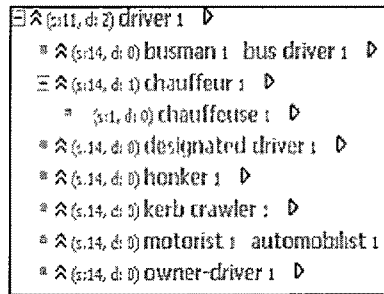
PWN 남/여성 자질이 포함된 신셋들의 분석 결과, 그 구축방식에 관해 다음과 같은 비판을 가할 수 있다.

① 개념 분화 적용 시 일관성 결여

PWN에는 남/여성형으로 개념적 분화가 가능한 모든 신셋에 대해 일관성 있는 기준으로 남/여성형이 분화되지 않았다. 이 문제는 '개념 미분화'와 '개념 과분화', 두 가지 양상으로 나타난다.

개념 미분화란 특정 어휘개념에 대해서 의미 분화가 전혀 이루어지지 않음을 의미한다. 반면, 어떤 어휘 개념은 인접 노드에서조차 남/여성 의미 분화가 세부적으로 이루어졌고, 더 나아가 경우에 따라서는 지나친 의미 세분화(fine-grained)가 이루어졌다. 후자는 개념 과분화이다.

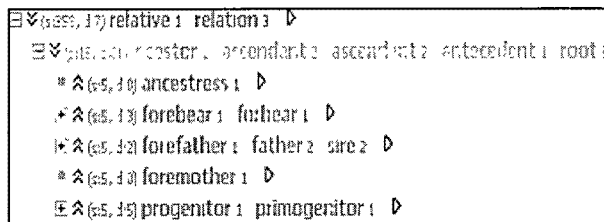
상위어 {driver 1}의 일부 하위어들을 보면, 남/여성 개념 분화 시 일관성이 결여되어 있음을 쉽게 알 수 있다. 우선 {chauffeur 1}은 'a man paid to drive a privately owned car' 로, 여성형 {chauffeuse 1} 'a woman chauffeur'가 하위어로 분화되어 있다. 반면, {busman 1, bus driver 1} 'someone who drives a bus' 도 분명 '여성 버스기사'가 존재 할 수 있으나 이에 대해 의미 분화를 시키지 않았고, 이와 자매 관계에 있는 {owner-driver 1} 'a motorist who owns the car that he/she drives'과 {motorist 1, automobilist 1} 'someone who drives (or travels in) an automobile' 역시 개념상 남/여성에 관한 분화가 가능하나 신셋을 분리하지 않았다.



<그림6. 개념 분화 적용에 일관성이 결여된 신셋들²²⁾>

② 계층성의 일관성 결여

전술한 바와 같이 PWN에는 남/여성 명사가 형제-자매 (sibling) 관계로 된 유형 1, 상이한 계보로부터 내려온 유형 2, 부모-자식(parent-child) 관계로 된 유형 3, 그리고 하나의 단독 어휘를 사용하는 유형 4에 따라 남/여성 명사형이 선정된 것으로 확인된다. 그러나, 이들 남/여성 개념구조를 세분화하는데 필요했을 준거는 어느 곳에도 명시적으로 표현되지 않았으며, 이점은 남/여성 개념의 구조화에 큰 걸림돌이 될 수 있다.

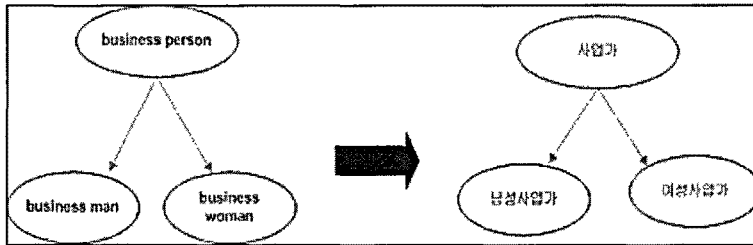


<그림7. 개념 세분화의 기준 문제와 관련된 신셋들>

‘선조’, ‘조상’ 이라는 의미의 {ancestor 1}과 {ancestress 1}, {forefather 1}과 {foremother 1}은 각각 남/여성 자질을 포함한 반의어이나, 흥미롭게도 {ancestor 1}과 {ancestress 1}은 부모-자식 관계로, {forefather 1}과 {foremother 1}은 형제-자매 관계로 설정되어 있다. 이처럼, 동일한 의미 개념을 세분화하는데, 일관된 기준이 적용되지 않았다. 그런데, 일관성이 결여된 기준이 적용된 PWN의 개념구조를 KL 구축 시, 아무런 수정 없이 그대로 수용할 경우 KL 한국어 어휘의미망에도 오류가 따르게 될 것이다. {businessman 1}, {businesswoman 1}의 형제-자매 구조를 KL에 그대로 적용, 구축할 경우 다음과 같은 구조가 성립되지만, 이것은 한국어에 ‘남성사업가’라는 어휘가 사용되지 않는다는 점과, ‘여성사업가’는 ‘사업가’라는 대표 개념

²²⁾ 관련 신셋들 중 남/여성으로 개념 분화가 가능한 것으로 판단되나, 미분화된 경우는 [+male]만 지닌 신셋 중 43개(14.86%), [+female]만 지닌 신셋 중 66개(18.61%)로 확인된다

에서 분화된 개념이자 어휘라는 사실이 제대로 반영될 수 없다는 오류로 직결된다.



<그림8. 부적절한 KL 구축의 예>

남/여성 개념 구조와 관련하여, PWN이 반의 관계(antonym)로 연결하고 있는 것은 {man 1}-{woman 1}, {father 1}-{mother 1}, {son 1}-{daughter 1}, {brother 1}-{sister 1}, {parent 1}-{child 2}, {boy 1}-{girl 2}, {uncle 1}-{aunt 1} 등의 극히 기본적인 어휘에 국한한다. 반면, {businessman 1}-{businesswoman 1}, {foster-father 1}-{foster-mother 1}, {bachelor 1}-{bachelor girl 1}에는 남/여성 자질이 포함되어 있고 전통적 의미론적 관점에서 반의어로 판단되는 대부분의 개념들은 반의 관계로 설정되지 않았다.

4. 남/여성 의미자질을 고려한 KL 구축

본 장에서는 앞서 검토한 PWN의 남/여성 자질이 포함된 신셋의 어휘의미망을 기반으로, KL 구축 방법에 대해 고찰한다. 특별히 영어와 한국어의 의미계층구조가 정확히 일치하지 않음을 감안하여, 출발은 PWN 의미 구조에서 하위 한국어의 특성을 반영할 수 있는 방식으로 구축해야 한다.

4.1. KL 구축의 준거

해당 신셋의 KL 구축에는 적절한 준거가 요구되는데, 다음으로 요약된다.

첫째, 먼저 PWN 계층구조 수용을 원칙으로 하되, 남/여성 관련 신셋 및 직속 상위어의 남/여성 자질 '유무'를 우선적으로 확인한다. PWN 계층구조 수용 문제는 이미 존재하는 노드의 수용 또는 수정 후 수용, 그리고 부재한 개념의 노드 생성 또는 제거라는 매우 복잡한 문제로 연결된다. 본 연구는 우선 이미 존재한 노드의 수용 또는 수정 후 수용만으로 제한한다. 먼저 {Jew} - {Jewess}, {유대인} - {유대인 여자} 등과 같이 개념상 '부모-자식' 노드 또는 '형제-자매' 노드로 분화가 모두 가능한 경우 PWN의 계층구조를 수정 없이 수용한다. 또한 기존의 노드를 수정, KL 을 구축하는 경우는 다음과 같다. PWN의 {beggar}는 [+male], [+female]을 공유한 신셋으로 남/여성 분화는 {beggarman}과 {beggarwoman}으로 일어나며, 이

둘이 형제-자매 노드에 연결된다. 그러나, KL 구축 시 이를 {거지}에서 {여자거지}가 분화된 형태의 부모-자식 노드로 수정 구축하는 것이 타당하다. 마지막으로, PWN에 {soul brother} 'a fellow Black man'는 존재하나, 대응되는 여성형인 {soul sister}는 부재하다. 보다 정확한 KL 구축에는 여성형 노드의 생성이 불가피하지만, 이는 추후 연구과제로 삼는다. 본 연구에서는 분화 불필요로 처리한다. 유사한 예로 {homeboy 0} 'a male friend from your neighborhood or hometown' (고향친구), {shaver 1} 'an adult male who shaves (면도사)'의 경우 {고향 여자친구}, {여자 면도사} 등이 가능하다.

둘째, 또한 상위어를 중심으로 남/여성형 분화 시 동일 자질의 '직속 승계' 또는 '변환 승계' 여부를 고려해야 한다. {차장}, {여(자)차장}의 경우 여성형의 남성형으로부터의 분화와 형제-자매 노드노드의 분화, 모두 가능한 것으로 보이는데 이에 대한 근거가 확실치 않은 경우 PWN 구조 수용, 구축한다. {assistant}는 [+male, +female]를 공유하고, 여성형 {girl Friday}를 형제-자매 노드로 갖는다. 이 경우 KL은 {보조자}, {여성 보조자}로 구축하며 {여성 보조자}는 {보조자}의 2개의 자질 중 하나만 승계된다.

셋째, 상위어 신셋의 남/여성 자질이 확인되면, 적정 어휘의 선택과 화용적 문맥에서 남/여성형 분화의 필요성을 고려해야 한다. 가령, 개념상 남/여성형의 분화에 무리가 없으나, 화용적 차원의 분화 불필요에 따라 사용되지 않는 개념들이 많기 때문이다. PWN을 참고로 한 {편곡자}, {투우사}, {감정사} 등은 남/여성형 분화가 불필요하다. 또한 PWN 에 {선조}, {여자선조?}, {우상승배자}, {여성 우상승배자?}, {동포}, {여자동포?}, {친척}, {여자친척?} 등으로 분화된 개념도 남/여성형의 분화가 불필요한 경우도 보고, 대표형만으로 구축한다. 이를 요약하면, 분화 불필요한 경우는 다음으로 요약된다.

- 1) 남성 또는 여성의 한 가지 의미만 갖는 신셋: {테너}, {환관}, {임산부}, {미망인} 등.
- 2) 한국어 개념상 분화 불필요한 경우: {선조}, {조상} 등.
- 3) PWN 구조에 분화가 안 된 노드로 KL에서는 노드 추가가 되어 할 경우

4.2. 구축 유형

위의 준거를 근거로 남/여성형 KL을 구축하면 다음 남/여성형 분화, 동일자질 승계, 동일자질 전환 승계 등의 네 가지 유형으로 분류된다. 이것은 각각 상위어의 남/여성 자질 공유 여부와 한 가지 자질만 갖는지 여부에 의해 분류되었다.

4.2.1. 남/여성 자질 공유한 상위어 => 남/여성형 구축

[+sex: value = +male, +female] => [+sex: value = +male], [+sex: value = +female]

먼저 관련 신셋의 상위어가 [+male], [+female]을 공유하고, 남/여성 분화의 필요성이 확인되면 이를 각각 구축한다. {자녀}에서 {아들} - {딸}, {약혼자}²³에서 {약혼자} - {약혼녀}, {수양자식}에서 {수양아들} - {수양딸}, {부모}에서 {아버지} - {어머니} 등이 구축된다.

4.2.2. 단일 남/여성 자질의 상위어 => 동일 자질 승계, 세분화된 의미 구성

[+sex: value = +male] => [+sex: value = +male]

[+sex: value = +female] => [+sex: value = +female]

상위 개념의 남/여성 자질이 하위어에 직속, 승계되는 유형으로 의미 세분화가 일어나면 발견된다. 가령 {전업남편}은 상위어로 {남편}의 자질을, {골프 미망인}²⁴은 {아내}의 자질을 승계 받아 구축한다. {카운터 테너}는 {알토}, {전쟁미망인}은 {미망인}, {섭정 여왕}은 {여왕}의 자질을 각각 승계 받아 구축한다. 또한 보다 세분화된 의미 표현의 경우 적정 어휘는 사전 및 그 활용도를 감안, 선정한다.

4.2.3. 단일 남/여성 자질의 상위어 => 동일 자질, 전환 승계, 남/여성형 분화

[+sex: value = +male] => [+sex: value = +female]

[+sex: value = +female] => [+sex: value = +male]

이 유형은 단일 대표형에서 남/여성형이 분화된다. 또한 여기서 대표형의 남/여성 자질의 승계는 전환 승계되는 것만을 의미한다. {후원자} - {여성 후원자}, {법정상속인} - {여자법정상속인}, {교사} - {여교사}, {배우} - {여(자)배우} 등은 단일 대표형에서 여성형이 분화된 예이다. 이 경우 대표형 대신 '남성' 또는 '남자'라는 접두사의 사용이 허용되며, {남성 후원자}, {남성(자)법정상속인}, {남자 교사}, {남자 배우} 등이 무난히 사용된다. 반면, 대표형으로부터 여성형이 분화된 또 다른 유형이 있다.

{감독} - {여(자)감독}, {가정교사} - {여자 가정교사}, {사업가} - {여성(자)사업가}, {국회의원} - {여성(자)국회의원}, {의사} - {여(자)의사}, {교수} - {여교수}, {변호사} - {여(성)(자)변호사}, {사장} - {여(자)사장}, {최고경영자} - {여성 최고경영자}, {아나운서} - {여자 아나운서} 등은 대표형에 '남성' 또는 '남자'의 접두사가 사용될 수 없다. {남자의사}, {남자교수}, {남자변호사}, {남자최고경영자} 등의 어휘가 매우 제한된 특정 문맥을 제외하고 그 쓰임이 상당히 어색하다. 접두사 '남성' 또는 '남자'의 첨가가 허용되지 않는 것은 {의사}, {교수},

²³ 한국어의 약혼자는 '~와 약혼한 사람'과 '~와 약혼한 남자'의 두 가지 의미로 사용된다.

²⁴ {전업남편}과 {골프미망인}은 각각 PWN의 {house husband 1} 'a husband who keeps house while his wife earns the family income', {golf widow 1} 'a wife who is left alone much of the time because her husband is playing golf' 에 해당한다

{변호사} 등이 여전히 관련 직업군을 대표하는 대표형으로 사용, 인식되고 있음을 보여준다.

4.2.4. 단일 남/여성 자질의 상위어 => 남/여성형 분화 불필요

[+sex: value = +male] => ∅

[+sex: value = +female] => ∅

남/여성형 분화가 불필요하거나 불가능한 유형이다. PWN의 {남자인어} - {여자인어}, {남자배심장} - {여자배심장}, {남자시조} - {여자시조} 등은 KL에 여성 개념의 분화 및 어휘화가 불필요하다. {남자인어}는 한국어에 개념 자체가 부재한 예이고, {남자시조} - {여자시조}도 상당히 어색한 예이다. 따라서, {거인}, {인어}, {조상} 등은 남/여성형 분화가 불필요하다.

이상의 결과를 신셋의 분포율로 정리하면 다음과 같다.

유형	[+ male]과 [+ female] 공유한 상위어	[+ male] 또는 [+ female] 지닌 상위어			합계
		동일 자질 승계	동일 자질 전환 승계	분화불필요	
남/여성 구축방법	남/여성 분화				
[+ male]을 지닌 신셋	-	104	59	133	296
[+ female]을 지닌 신셋	-	147	86	127	360
[+ male, + female] 공유한 신셋	23	-	-	-	
합계		251	145	260	679

<표2. KL 남/여성형 구축의 결과>

위의 표는 관련 신셋을 KL로 구축한 결과를 요약하였다. 이를 위해 관련 신셋들을 상위어의 남/여성 자질 공유 여부와 남/여성 한 가지 자질 소유 여부를 중심으로 1차 분류하였고, 이들을 남/여성형 분화와 관련하여 KL로 구축해 보았다. 이들은 남/여성형 분화, 자질 승계, 자질 전환 승계 기준에 따라 위의 유형으로 분포한다.

5. 끝맺는 말

본 연구는 의미론이 제공하는 의미적 자질중 하나인 남/여성 자질 [+male, female]을 고려

하여, 보다 정확하고 정교한 KL 구축의 준거로 삼고자 시작되었다. 이를 위해 먼저 KL 구축의 기반이 될 PWN의 남/여성 자질이 포함된 신셋들의 의미계층 구조 분석 및 유형화를 시도하였다. 이 과정에서 PWN내 관련 신셋들의 구성 방식에 일관성이 결여되어 있고, 관여적인 어떤 원칙이나 규칙도 적용되지 않았음을 확인할 수 있었다. 또한 KL 구축 시 PWN이 제공하는 신셋의 의미와 계층 구조는 일부 수용하되, 한국어 화용적 상황과 부합되지 않는 어휘의미 구조는 적절히 수정하여 KL 구축의 준거로 삼아야 함에 주목하였다.

무엇보다 남/여성 자질이 영어-한국어에서 반드시 동일하게 실현되지 않으며, 이것은 의미적, 어휘적, 화용적 이유에서 비롯한다. 따라서, 이 점을 감안하여 보다 정교한 어휘의미망 구축을 위한 심화된 준거 마련과 활용이 검토되어야 하며, 향후 한국어 어휘의미망 추가 및 확장에 중요한 변수가 될 수 있다. 보다 관여적인 준거의 설정 및 개발은 향후 진행될 핵심 연구과제 중 하나이기도 하다.

참고 문헌

- 김권양. 1999. “한국어 WordNet의 분석과 방향”, 경일대학교 논문집, vol.16, No.6, 793-807.
- 김현권. 2000. “EuroWordNet의 구성 원리와 설계”, 언어학, vol. 27, n. 1, 한국언어학회, 145-177.
- 오장근. 2002. 유로워드넷 기반의 어휘 데이터베이스 활용을 위한 한국어-독일어 ILI 대응 방법론 연구, 독어학 6권, 한국독어학회.
- 문유진. 1996. 의미론적 어휘 개념에 기반한 한국어 명사 Wordnet의 설계와 초록, 서울대학교 컴퓨터 공학과 박사 논문.
- 이은령, 윤애선, 권혁철. 2004. “워드넷(WordNet 2.0) 커뮤니케이션 명사의 결여 의미관계”, 한국인지과학회 춘계학술대회, 한국인지과학회, 177-183.
- 이은령, 황순희, 윤애선. 2004. “다국어 어휘의미망 구축의 현황과 문제점-PWN과 EWN을 중심으로”, 프랑스문화예술연구회 제12집(6권3호), 프랑스 문화예술학회, 369-401.
- 이창기, 이근배. 1999. “Wordnet을 이용한 한국어 시소러스 자동 구축”, 제11회 한글 및 한국정보처리학회, 한국정보과학회, 156-161.
- 임지룡. 1989. “국어 분류 어휘집의 체계와 상관성”. 국어학 19, 395-425, 서울, 국어학회.
- 최호섭, 옥철영. 2002. “한국어 의미망 구축과 활용”. 한국어학 17, 301-329, 서울, 한국어학회.
- 최호철, 한정환. 2003. “다국어 기계번역을 위한 중간언어 모형과 방법론 연구[1]”, 컴퓨터와 인문학 8, 고려대학교 민족문화연구원.
- 황순희, 이은령, 윤애선. 2004. “불한 전문텍스트 번역에서 의미해석의 문제”, 프랑스어문교육 제18집, 한국프랑스어문교육학회, 315-334.

- 황순희, 윤애선. 2005. '워드넷 기반 한국어 명사 어휘의미망의 정제', 한국인지과학회 춘계 학술대회, 한국인지과학회, 267-272.
- 히데키. 2002. 한국어 어휘와 문법의 상관구조, 서울, 태학사.
- Alonge, A. and Lönneker, B. 2003. "The Heart of problem: How shall we represent metaphors in wordnet" Global Wordnet Conference 2004 Proceedings, Brno, 1.
- 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, Tzizgov Chark, 1997. <http://xxx.lanl.gov/abs/cmp-lg/9709003>
- Balkova, V et al. 2003. "The Russian wordnet", Global Wordnet Conference 2004 Proceedings, Brno, 31-38.
- Black, W. et El- Kateb, S. 2003. "English-Arabic Prototype WordNet", Global Wordnet Conference 2004 Proceedings, Brno, 67-74.
- Bilgin, O et al. 2003. "Morphosemantic Relations In and Across Wordnets", Global Wordnet Conference 2004 Proceedings, Brno, 60-66.
- Bentivogli, L., Bocco, A. and Pianta, E. 2003. "ArchiWordNet: Intergrating Wordnet with domain-specific knowledge", Global Wordnet Conference 2004 Proceedings, Brno, 34-46.
- 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, 91-96.
- 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, 97-105.
- Cruse, D. A. 1986. Lexical semantics, Cambridge, Cambridge University Press.
- Evens M. W. (eds.) 1988. Relational models of the lexicon, Cambridge, Cambridge University Press.
- Fellbaum, C. (eds.) 1998. WordNet - An Electronic Lexical Database, Cambridge, MIT Press.
- Gomez, F. 2003. "Grounding the ontology on the semantic interpretation algorithm", in Proceedings of the second international WordNet conference 2004, Brno, 124-129.
- Lyons, J. 1977. Semantics I, Cambridge, New York, Cambridge University Press.
- Miháلتz, M. et Prózéký, G. 2003. "Results and Evaluation of Hungarian Nominal WordNet v1.0" Global Wordnet Conference 2004 Proceedings, Brno, 175-180.
- 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>.

Nirenburg, S and V. Raskin. 2004. *Ontological semantics*, Cambridge, MIT Press.

Smrž, P. 2004. "Quality Control for WordNet Development", Global Wordnet Conference 2004 Proceedings, Brno, 206-212.

Vossen, P. 1998. *EuroWordNet: A multilingual database with lexical semantic networks*. Dordrecht, Kluwer Academic Press [유로워드넷, (한정 한 외 공역), 서울, 한국 문화사.]

Vossen, P. 2004. *EuroWordNet General Document*, <http://www.hum.uva.nl/~ewn>.

참고 웹사이트

EuroWordNet Home Page; <http://www.hum.uva.nl/~ewn>

HowNet Home Page; <http://www.keenage.com>

Princeton Univ. WordNet Home Page; <http://www.cogsci.princeton.edu/~wn/>

The Global Wordnet Association Home Page; <http://www.globalwordnet.org/>

Electronic Dictionary Research Institut Home Page; <http://www2.crl.go.jp/kk/e416/EDR/>