

구조적 학술용어사전 구축에 있어서  
인문사회분야 복합명사의 한자어 접미사  
형태분석에 의한 용어의 개념화에 관한 연구  
A Study on the Conceptualization of Terminology for  
Construction of Structured Glossary - A Morphological  
Analysis of Sino-Korean Suffixes of Compound Noun  
in Humanities & Social Sciences

박성은, 성균관대학교 대학원 문헌정보학과, pse3598@skku.edu  
신유정, 성균관대학교 대학원 문헌정보학과, shinyou02@skku.edu

Seong-Eun, Park, Library & Information Science, SungKyunkwan University

You-Jung, Shin, Library & Information Science, Sungkyundwan University

본 연구는 한자어 접미사를 가진 학술용어 복합명사의 형태를 분석하고, 한자어 접미사를 가진 복합명사의 범주화를 위한 알고리즘을 제시하는 것을 목적으로 한다. 연구의 배경은 구조적 학술용어사전의 구축과 관련하여 용어의 효율적인 개념화 방법론을 개발하는 것이다. 연구 목적을 이루기 위해 한자어 접미사를 가진 인문사회분야 복합명사의 형태를 분석하여, 출현 빈도가 10회 이상인 한자어 접미사를 각 분야별로 14개씩 추출하였다. 그리고 각 한자어 접미사가 배정된 개념범주와 개념범주 배정 빈도를 분석한 결과, 동음이의 한자어 접미사를 가진 경우 개념범주가 다양하게 배정될 수 있지만, 동음동형 한자어 접미사를 가진 경우에는 그렇지 않다는 특징을 파악하였다. 또한 특정 한자어 접미사가 하나의 개념범주에 배정된 빈도가 높고 그 외의 개념범주에 배정된 빈도는 현저히 낮을 경우 빈도가 높은 개념범주에 배정되는 것이 더 타당한 것으로 분석되었다.

## 1. 서론

구조적 학술용어사전이란 학술용어를 유사한 속성을 가진 개념 범주에 따라 분류하고, 그 범주가 가지는 속성을 체계화 한 다음, 그 속성에 따라 해당 학술 용어 하나하나의 의미를 정의한 사전을 말한다. 범주별로 구조화되는 속성들 간의 의미적 연관관계를 생성시켜 구조적 학술용어사전을 지식베이스로 발전시킬 경우 의미적 연관 검색이 가능하게 된다. 이러한 지식베이스는 연구를 수행하는데 있어 반드시 선행되어야 하는 연구자의 정보문제 해결, 즉

유사하거나 동일한 주제의 연구 성과물 사이에 나타나는 의미적 연관관계를 파악하는데 도움이 된다.

구조적 학술용어사전을 구축하기 위해서는 유사한 속성을 가지는 학술용어의 개념을 유형화하여 범주를 만드는 것이 필요하다. 학술용어의 범주화를 위한 선행 작업에서 가장 일반적으로 사용되는 처리방안의 하나는 형태소분석기를 통한 용어의 분석이다. 그러나 형태소분석 알고리즘은 형태소 분리 및 복합명사 분해 문제가 복합되어 분석 효율의 저하 및 분석 중의성 증가 문제가 발생하게 된다 (강승식

1998). 붙여 쓴 복합명사를 분해할 때 ‘대학생 선교회’가 ‘대학생’+‘선교회’ 또는 ‘대학’+‘생선’+‘교회’로 분해되는 경우 등이 그 사례이다 (윤보현 1995).

특히 인문사회분야의 학술용어의 경우 인명, 문헌명, 개념 등을 나타내는 고유명사와 복합명사들이 많아 개념 유형을 판단해 정확한 범주를 부여하기 위해서는 개체명 등의 전문용어와 복합명사에 대한 별도의 처리방안이 요구된다 (고영만 2012). 이와 관련하여 본 연구는 한자어 접미사를 가진 학술용어 복합명사의 형태를 분석하고, 한자어 접미사를 가진 복합명사의 범주화를 위한 알고리즘을 제시하는 것을 목적으로 한다. 본 연구의 결과는 구조적 학술용어사전의 효율적 구축을 위한 보조도로 활용될 수 있을 것이다. 다만 본 연구는 인문사회과학 분야의 용어만을 대상으로 하기 때문에 복합명사 한자어 접미사의 분석 결과를 전주제 분야로 일반화시키기에는 한계가 있다.

## 2. 이론적 배경 및 선행연구

### 2.1 이론적 배경

#### 2.1.1 개념범주

개념범주란 유사한 속성을 가지는 학술용어를 동일한 상위 개념으로 범주화한 것이며 구조적 학술용어사전에 포함된 개념들을 집단화하기 위한 값으로 토픽맵과 같은 온톨로지 구축 측면에서는 클래스에 해당한다. 동일한 개념범주에 속하는 모든 용어들은 동일한 개념 속성을 갖게 된다.

예를 들어 인문사회분야의 최상위 개념범주는 실체, 추상, 일반, 개체의 네 가지로 구분해 볼 수 있으며, 중위 범주, 하위 범주 등으로 세분해 볼 수 있다(<표 1> 참조; 정보관리연구소 2013).

<표 1> 인문사회분야 학술용어의 개념범주화 예

기본 범주	중위범주		기본 범주	중위범주	
실체 (A)	A01	인간	일반 (C)	C01	공간
	A02	조직집단		C02	시간/시대
	A03	자연물		C03	방법
	A04	인공물(전체/부분)		C04	언어/각국어
추상 (B)	B01	행위/활동	개체 (D)	D01	인명
	B02	변화(변동)		D02	저작물명
	B03	현상/이슈		D03	지명
	B04	관계/상호작용		D04	사건명
	B05	가상인공물		D05	기념물(문화재)
	B06	제도(체제)		D06	법률명
	B07	특성(성질)		D07	기관/단체명
	B08	심리(감정)		D08	상품명
	B09	이론(사상)			
	B10	형식(틀)			
	B11	기법(평가)			

#### 2.1.2 전문용어와 복합명사

전문용어(terminology)는 일반인들이 일상생활 속에서 사용하는 일반어와 대립되는 것으로, 사용자가 전문인들로 제한되어 있고 특수한 목적을 가지고 있는 언어를 말한다. 어휘·형태적인 측면에서 전문용어는 다음과 같은 특성을 지니고 있다. 첫째, 전문용어는 일반어에 비해 어휘특정성이 큰 용어이다. 둘째, 일반어는 불특정 다수에 의해 일상생활에서 비교적 자연스럽게 형성된 단어인 반면, 전문용어는 특정 분야의 전문가에 의해 생성된 용어이다. 셋째, 전문용어의 생성에는 한자어 접사가 매우 생산적으로 쓰인다. 한자어 접사는 어휘적 의미를 가지기 때문에 단어보다 간략한 형태로서 개념결합력을 이끌어 내기 때문이다 (조은경 2001).

그러나 전문용어의 한자어 접사의 경우 몇 가지 문제를 지닌다. 첫째, 한자어 접사의 결합범위는 전문용어의 띄어쓰기에 어려움을 초래한다. 둘째, 한자어 접사가 동일음으로 인한 중의성을 발생시키는 면이 있기도 하지만, 동일음의 한자어 접사는 학문분야에 따라 대체로 다른 경향을 보인다 (조은경 2001). 위와

같은 특성과 문제점으로 인해 전문용어의 분석은 연구의 목적에 맞게 적절한 분석단위를 고려할 필요성이 있다.

복합명사란 하나의 형태소로 이루어진 단어인 단일어와 구별되며, 둘 이상의 실질 형태소가 결합된 합성어와 실질 형태소와 접사가 결합되어 이루어진 파생어를 통칭한다. 우리나라는 오랫동안 한자를 표기문자로 사용해왔기 때문에 국어 어휘에서 한자어가 차지하는 비중이 매우 높다. 따라서 한자어 접사가 붙어 형성된 복합명사 또한 많은 비중을 차지한다. 또한 한국어에서는 복합명사가 다양한 구문적 유형으로 표현된다. 표면적으로 동일하게 표현된 복합명사일지라도 의미적으로는 서로 다르게 나타나거나 여러 개의 단위명사로 분리될 수 있다. 따라서 복합명사를 구성하는 단위명사간의 의미적 결합관계를 올바르게 파악하기 위해서는 복합명사를 그 형태 및 성분으로 분석하는 것이 중요하다 (남궁황, 이태영 2001).

그러나 다음과 같은 이유로 복합명사 분석에 어려움이 있다. 첫째, 한국어 복합명사는 띄어쓰기를 하므로 단위명사로 분해하는 것이 독일어, 일본어, 중국어 등에 비해 용이하다. 그러나 붙여 쓴 복합명사의 경우 단위명사로 분해하는데 어려움이 따른다. 둘째, 복합명사는 조합이 자유롭고 길이에 제한이 없어 그 분석에 어려움이 따른다. 더욱이 접사나 조사가 복합명사의 구성요소로 분석되는 경우 복합명사의 분석을 더욱 어렵게 한다.

## 2.2 선행연구

전문용어와 관련된 기존의 연구에서는 전문용어와 일상용어의 차이점을 밝히고 전문용어의 형태적 분석을 통한 조어법 및 용어의 특성을 밝히고자 하였다. 김정우, 김성원(2012)은 일상어와 구분되는 전문용어 조어법의 특징을 밝히고 곤충관련 용어를 중심으로 한 전문용어

의 분석사례를 제시했다. 조은경(2001)은 『표준국어대사전』의 전문어 목록을 대상으로 하여 전문용어의 어휘·형태적 특성에 대해 연구했다.

복합명사에 대한 기존 연구는 자연어 처리 및 정보검색의 효율을 향상시키기 위해 형태소 분석기를 이용한 복합명사의 분해와 합성과 관련된 연구가 대부분이었다. 강승식(1998)은 형태소 분석결과로 추정된 복합명사를 단위명사들로 분해하는 방법을 고안하여 최적후보를 선택하는 알고리즘을 제안하였다. 원형석, 박미화, 이근배(2000)는 명사구 색인과 복합명사 분할을 포함한 복합명사 처리를 위해 통합적 색인기법을 제안하였다. 남궁황, 이태영(2001)은 복합명사의 의미적 구조분석 통해 복합명사의 표층구조와 심층구조간의 의미관계를 기반으로 복합명사 분석 시 적용할 수 있는 의미규칙을 도출하였다. 김수남 등(1998)은 특정 단어의 의미정보에 따른 제약조건을 규칙베이스로 구축하고 이를 의미정보로 활용하여 한국어 복합명사를 분석하였다.

이 외에 김규환, 장보성, 남영준(2008)은 한국학술진흥재단(현 한국연구재단)의 용어데이터베이스에 수록되어 있는 전문용어를 대상으로 인문사회분야에서 사용되는 개념과 그 개념들 간의 관계를 정의하고, 이를 계층구조로 표현함으로써 인문사회분야 온톨로지를 구축하고자 하였다.

위에서 볼 수 있듯이 전문용어 복합명사를 분석한 연구는 아직 수행되지 않았다. 그러나 전문용어의 특성 상 복합명사의 비중이 높으며, 앞서 밝혔듯이 전문용어의 경우 한자어 접미사가 매우 생산적으로 작용하기 때문에 한자어 접미사를 가진 전문용어 복합명사의 형태분석은 중요한 의미를 지닌다. 본 연구의 분석 결과는 특정 한자어 접미사가 어떤 개념범주에 속하는지 파악함으로써 구조적 학술용어사전의 구축 시 개념범주를 효율적으로 부여할 수 있는 하나의 방법론으로 발전 될 수 있을 것이다.

### 3. 복합명사의 형태 및 성분분석

#### 3.1 대상용어 추출과정

한국연구재단 학술연구분야 분류표 기준 인문학과 사회과학 분야에서 2007년부터 2011년까지 발간된 KCI 등록 학술지는 인문학 500종, 사회과학 674종으로 총 1,174종이며 여기에 수록된 논문은 인문학 46,914편, 사회과학 83,263편으로 총 130,177편이다. 각 논문에 출현한 저자키워드는 총 474,645개이며 동일 형태 용어의 중복을 제거한 고유저자키워드 수는 총 75,279개이다(<표 2> 참조).

<표 2> 2007-2011년 발간 학술논문 수, 저자키워드 출현 수

	2007	2008	2009	2010	2011	합계
학술논문 수	20,351	24,587	27,570	29,678	27,991	130,177
저자키워드 출현 수 (고유저자키워드 수)	73,792	89,962	101,154	108,547	101,190	474,645 (75,279)

그 중 단일어, 숫자가 포함된 복합명사, 외국어 약어, 고유명사(인명)를 제외한 인문학(1,067개), 사회과학(2,548개) 용어를 분석대상으로 선정하였으며, 그 중 한자어 접미사가 포함된 용어는 각각 479개와 536개이다. 복합명사 특성 상 용어 자체에 띄어쓰기가 포함된 경우 띄어쓰기를 기준으로 자동 분리하였고, 붙여 쓴 복합명사의 경우 연구자가 직접 분리하였으며 이를 바탕으로 한자어 접미사가 포함된 용어를 추출하여 그를 분석하였다(<표 3> 참조).

<표 3> 대상용어 추출내용

	전체용어	분석대상이 된 복합명사	한자어 접미사가 포함된 용어
인문학	1,930	1,067	479
사회과학	2,925	2,548	536

#### 3.2 한자어 접미사가 포함된 용어 분석

한자어 접미사가 포함된 용어를 분석하여 학문분야별로 빈도수가 10회 이상인 상위 14개 한자어 접미사를 추출하였다. 인문학의 경우 전체 479개의 용어 중 325개(67.8%)의 용어가, 사회과학의 경우 전체 536개의 용어 중 445개(83.1%)의 용어가 14개의 한자어 접미사로 구성되어 있었다(<표 4> 참조).

<표 4> 한자어 접미사의 개념범주 및 출현 빈도

인문학						
no	한자어 접미사		배정된 개념범주		빈도수	
1	성	性	B03	1	62	
			B04	1		
			B07	35		
			B08	6		
			B09	16		
C04	3					
2	사	使	3	A01	2	44
				A02	1	
		史	27	A04	1	
				B02	20	
				B09	4	
				D02	1	
				E15	1	
		寺	4	C01	1	
				D05	2	
				D08	1	
師	2	A01				
死	1	B01				
詞	5	C04	4			
		D02	1			
辭	2	C04				
3	론	論	A02	1	37	
			B09	31		
			B10	2		
			C04	1		
			D02	2		
4	학	學	32	B09	28	33
				B10	1	
				B11	1	
				C04	1	
		D02	1			
諺	1	B11				
5	어	語	B05	1	28	
			C04	27		
6	기	器	3	A04	2	19
				C02	1	
		期	8	C02		
		氣	1	B09		
記	7	D02	6	1		
		D05	1			
7	화	化	B01	2	17	
			B02	1		
			B03	5		
			B08	1		
			B10	2		
			B11	3		
C04	3					
8	시	詩	B09	1	16	
			B10	15		
9	가	歌	B10	5	15	
			D02	10		
10	왕	王	D01		12	
11	군	君	1	D01		11
		軍	8	A02	4	
				D07	4	
郡	2	C01	1	1		
		D03	1			

12	법	法	B06 B09 B10 C04	2 2 6 1	11
13	묘	墓	A04 C01 D08	5 4 1	10
14	전	傳	D02		10

사회과학										
no	한자어 접미사		배정된 개념범주		빈도수					
1	성	性	B02	6	92					
			B03	10						
			B04	2						
			B06	10						
			B07	56						
			B08	2						
			B09	1						
			B10	3						
			B11	2						
			2	법		法	B01	2	91	
							B06	21		
B09	1									
B10	2									
B11	3									
C03	1									
D06	57									
E15	4									
3	권	權			2		76	78		
					B06					1
					B11					1
			B01	1						
			B04	1						
B06	70									
B07	3									
E15	1									
4	자	者	A01	26	31					
			A02	3						
			B07	1						
			E15	1						
			B01	3						
B07	17									
B08	1									
B11	2									
6	화	化	21	21	22					
			B02			4				
			B03			5				
			B06			5				
			B09			1				
			B11			1				
E15	1									
7	제	制	1	B08	18					
			B06							
8	율/률	率	B02	1	18					
			B03	7						
			B07	6						
			B11	1						
			C03	3						
9	력	力	B06	1	17					
			B07	15						
			B08	1						
			B08	1						
10	인	人	A01	13	15					
			B06	1						
			D07	1						
			D07	1						
11	회	會	A02	4	11					
			B01	1						
			D07	6						
			D07	6						
12	감	感	B07	3	10					
			B08	7						
13	사	事	2	B02	10					
			3							
			3							
			1							
			1							
14	죄	罪	1	B06	10					
			1							

\*'율/률(率)'은 '비율'의 뜻을 더하는 접미사로, 앞말에 따라 '율'과 '률'로 달리 쓰이기 때문에 용어 분석에서 같은 한자어 접미사로 판단하고 그 빈도를 통합하였다.

14개의 한자어 접미사를 분석한 결과 나타나는 두드러진 특징은 동음이의 한자어 접미사를 가진 경우 개념범주가 다양하게 배정될 수 있다는 점이다. 그 사례로 인문학의 사(使, 史, 寺, 師, 死, 詞, 辭), 기(器, 基, 氣, 記) 등을 들 수 있다. 반면 동음동형 한자어 접미사를 가진 경우는 특정 개념범주로 모을 수 있다. <표 5>와 같이 인문학에서 어(語), 시(時), 사회과학에서 자(者), 도(度), 력(力), 인(人), 감(感)의 경우 특정 개념범주의 빈도가 높고 그 외 개념범주의 빈도가 1, 2회로 현저히 낮을 경우 빈도가 높은 개념범주로 배정되는 것이 더 타당할 수 있다는 것을 고려해야 한다.

<표 5> 개념범주 배정 오류 및 추천 개념범주

	한자어 접미사	개념범주 배정 오류로 판단되는 용어	기존 배정 개념범주	추천 개념범주
인문학	어 語	민족어	B05	C04
	시 時	순수시	B09	B10
사회과학	자 者	인터넷서비스제공자	A02	A01
		기관투자자	A02	
		보험자	A02	
	도 度	비정규직근로자	B07	B07
		개인투자자	E15	
		교육요구도	B01	
력 力	학습준비도	B01	B07	
	학습참여도	B01		
	학습만족도	B08		
	가격민감도	B11		
인 人	실행도	B11	B07	
	증거력	B06		
인 人	주의력	B08	B07	
	준법지원인	B06		
감 感	과산관제인	D07	A01	
	쓰기효능감	B07		
	양육효능감	B07		
		탐효능감	B07	B08

구조적 학술용어사전 구축 시 개념범주 속성 및 개념속성 구축 지침서가 있기는 하지만, 이를 참고하더라도 구축자의 주관적인 판단이 반영되는 것을 배제할 수 없다. 따라서 <표 5>에서 제시한 것처럼 특정 한자어 접미사가 한 개념범주에 높은 빈도로 배정되었다는 것을 알

수 있다면 구축 시 구축자가 참고할 수 있을 것이다. 즉 여러 구축자가 동일한 한자어 접미사를 가진 용어를 구축할 때 일관된 개념범주 배정이 가능하도록 개념범주를 추천해줄 수 있다. 나아가 구축자간 용어 검수 시에도 일관된 개념범주 배정이 가능할 것이다.

#### 4. 결론

본 연구는 인문사회과학 분야 학술용어를 대상으로 한자어 접미사를 포함하는 복합명사의 형태를 분석하였다. 각 한자어 접미사가 배정된 개념범주와 개념범주 배정 빈도를 분석한 결과, 동음이의 한자어 접미사를 가진 경우 개념범주가 다양하게 배정될 수 있지만, 동음동형 한자어 접미사를 가진 경우에는 그렇지 않은 것으로 나타났다. 또한 특정 한자어 접미사가 하나의 개념범주에 배정된 빈도가 높고 그 외의 개념범주에 배정된 빈도는 현저히 낮은 경우 빈도가 높은 개념범주에 배정되는 것이 더 타당한 것으로 분석되었다. 본 연구의 결과는 구조적 학술용어사전 구축을 위한 개념범주 추천 알고리즘을 만드는 데 있어 기초가 되는 연구로 의의가 있다.

본 연구는 한자어 접미사를 포함한 복합명사만을 대상으로 하였으며, 한자어 접미사가 지닌 의미와 개념범주 부여의 연관 관계에 대해서만 파악하였다. 그러나 복합명사는 앞의 어절성분과 의미적으로 관련 있는 다중어절(multi-word term) 복합명사가 많이 있으므로 좌, 우 단위 명사간의 어의관계에 대한 정확한 파악 또한 필요하다. 따라서 향후 다중어절 복합명사의 단위명사 간 명확한 어의관계 파악에 대한 후속연구가 이루어질 경우, 구조적 학술용어사전 구축을 위한 개념범주 추천이 보다 정확하게 될 것이다.

#### 참고문헌

- 강승식. 1998. 한국어 복합명사 분해 알고리즘. 『정보과학회 논문지(B)』, 25(1): 172-182.
- 고영만. 2012. 2012년도 토대연구지원사업 연구계획서.
- 김규환, 장보성, 남영준. 2008. 전문용어 기반 인문사회분야 온톨로지 구축에 관한 연구. 『한국정보관리학회 학술대회 논문집』, 2008(-): 181-188.
- 김수남 외. 1998. 의미정보를 이용한 한국어 복합명사 분석. 『한국정보과학회 가을 학술발표 논문집』, 25(2): 195-197.
- 김정우, 김성원. 2012. 전문용어 조어법 분석의 한 사례-곤충 관련 용어를 중심으로-. 『인문사회과학 연구』, 13(2): 99-125.
- 김지영, 권혁철. 1992. 합성명사의 의미관계분석 시스템을 위한 지식베이스 구축 기법. 『한국정보과학회 추계 학술발표대회 논문집』, 19(2): 23-41.
- 남궁황, 이태영. 2001. 복합명사의 의미적 구조 분석에 관한 연구. 『한국정보관리학회 학술대회 논문집』, 8(-): 177-180.
- 원형석, 박미화, 이근배. 2000. 복합명사 분할과 명사구 합성을 이용한 통합 색인 기법. 『정보과학회논문지: 소프트웨어 및 응용』, 27(1): 84-95.
- 윤보현, 임희석, 임해창. 1995. 통계정보를 이용한 한국어 복합명사의 분석방법. 『한국정보과학회 봄 학술발표 논문집』, 22(1): 925-928.
- 정보관리연구소. 2013. 구조적 용어사전 구축 지침.
- 조은경. 2001. 전문용어의 어휘형태적 특성 연구. 석사학위논문, 연세대학교 대학원, 국어정보학 협동과정.