

# 복합명사의 의미적 구조분석에 관한 연구

## A study on the Semantic Structure Analysis of Korean Compound Nouns

남궁황, 이태영, 전북대학교 문헌정보학과

NamGung Hwang, Lee Tae-Yeong

Dept. of Library and Information Science, Chonbuk National University

복합명사의 효율적인 분석은 정보검색 시스템의 성능 향상에 중요한 영향을 미친다. 한국어에서는 복합명사가 다양한 구문적 유형으로 표현된다. 표면적으로 동일하게 표현된 복합명사일지라도 의미적으로는 서로 다르게 나타내거나 여러 개의 단위명사로 분리될 수가 있다. 이러한 문제를 해결하기 위해서는 복합명사를 구성하는 단위명사간의 의미적 결합관계를 올바르게 파악하는 것이 무엇보다도 중요하다. 본 논문에서는 의미적 구문관점에서 복합명사의 표층구조와 심층구조간의 의미관계를 기반으로 복합명사 분석시 적용할 수 있는 의미규칙을 도출하는데 있다.

### 1. 서론

정보검색시스템의 성능은 텍스트에서 표현되고 있는 주제를 얼마나 정확하게 분석하여 색인어로 추출하느냐에 따라 크게 좌우된다. 일반적으로 텍스트 내에서 색인어 대상이 될 수 있는 단어는 주로 명사에 국한되는데, 이는 한국어의 특성상 중요한 의미를 나타낼 수 있는 언어 성분이 명사이며, 텍스트 검색시에는 대부분 명사를 이용하여 검색하기 때문이다.

한국어의 명사는 구조적으로 단순명사, 파생명사, 복합명사, 명사구 등 여러 형태적 특징을 지니고 있다. 이로 인해 한국어에서는 파생명사인지 복합명사인지의 식별은 물론 단순명사와 복합명사를 구분하는데 애매한 경우가 있다. 또한 복합명사는 일본어나 중국어와는 달리 한국어에서는 어절단위로 띄어쓰는 것을 원칙으로 하나 붙여 쓸 수도 있다. 특히, 붙여 쓴 복합명사는 단위명사로 분해시 중의적 분할이 발생되고 한 단어가 문맥에 따라 명사가 되기도 하고 용언이 될 수도 있다. 이러한 명사의 특성 및 기능을 고려하여 텍스트 내에서 올바른 색인어를 선별하기 위해서는 자연언어를 처리할 때 명사나명사구, 복합명사 등을 언어학적 관점에서 분석하고, 그 문맥에서 나타내고 있는 바대로 해석해서 그 단어의 핵심의

미를 파악하는 일이 무엇보다도 중요하다.

현재 가장 일반적이며 기초적인 언어학적 분석 방법은 형태소 분석기법을 기반으로 한 복합명사 분석이나 부분적인 구문분석이며, 이에 대한 많은 연구와 새로운 기법들이 개발되어 정보검색분야에서 검색의 효율성을 높이는데 활용되고 있다. 그러나 보다 발전된 형태의 정보검색을 위해서는 문장구조를 파악하는 구문분석과 의미분석을 거친 색인시스템이 요구된다. 이러한 관점에서 한국어에서는 복합명사와 관련된 언어 현상이 매우 다양하게 나타나기 때문에 아직도 복합명사의 처리기법에 대해 개선해야 할 여지가 많이 있다.

본 논문에서는 한 단어가 명사임을 식별하고 복합명사와 명사구를 본래의 쓰임새 대로 분석하는 기존 형태소 분석기법들에 대해 살펴보고, 언어학적 측면에서 복합명사의 단어구조와 그 형성의 실체를 의미적으로 분석함으로써 복합명사 분석시 발생하는 단위명사의 중의적 문제나 모호성 문제 등을 해결할 수 있는 의미규칙을 도출하고자 한다.

### 2. 복합명사분석의 필요성 및 문제점

#### 2.1 복합명사 분석의 필요성

복합명사 분석에는 크게 두 가지 방식으로 구

분할 수 있다.

첫째는 표면적으로 동일한 유형으로 나타나지는 않지만, 동일한 의미로 인식될 수 있는 복합명사를 추출하는 방식이다. 이 방식은 검색의 정확율을 향상시키는데 중요한 영향을 미친다. 이러한 유형의 예를 들면 다음과 같다.

- 예) 정보검색시스템  
 정보의 검색  
 정보를 효율적으로 검색하는...  
 정보를 검색하고 색인하는.... 등

상기의 예는 모두 정보검색이라는 복합명사의 의미를 함축하고 있는 구문이다. 문맥상 동일한 의미가 각기 다른 형태로 표현되고 있으나 문헌검색시 적합한 문헌으로 인식되기 위해서는 모두 '정보검색'이라는 복합명사로 색인어를 추출해야 한다.

둘째는 한국어 특성상 붙어 쓴 복합명사에 대해 올바른 단위명사로 분해하는 작업이다. 이는 재현율을 향상시키는데 기여한다. 예를 들면, 문헌에서 '전략'에 관련된 내용도 다루면서 '전략전술'이라는 복합명사는 기술되어 있는 반면, 실제 '전략'이라는 단위명사는 표현되지 않는 경우에 적합문헌을 검색해 주기 위해서는 복합명사를 단위명사로 분리해 주는 작업이 필요하다.

## 2.2 복합명사 분석시 문제점

### 2.2.1 복합명사의 인식 문제

색인어로 선정될 수 있는 명사는 단순히 형태소를 분석한 결과에 따라 직접 명사로 파악되는 단어뿐만 아니라 명시적으로 복합명사의 형태로 나타나지는 않았지만, 복합명사의 의미를 내포할 경우 이를 인식할 수 있어야 한다.

### 2.2.2 복합명사의 진위판단 문제

텍스트 내의 다양한 어절가운데 분해대상 어절이 복합명사인지의 여부를 판단할 수 있는 능력이 있어야 한다. 예를 들면 어절 '공부하고는'은 용언인데 복합명사에 포함된 어절로 식별될 수 있기 때문에 이러한 부적절한 어절이 선정되지 않도록 해야 한다.

### 2.2.3 복합명사와 조사/어미분리 문제

복합명사와 연결된 조사/어미 사이에서 서로 음절이 중첩되어 분리위치를 찾기가 어렵고, 이를 확실히 분리하면 복합명사 분해가 실패하는 원인이 된다. 예를 들어 어절 '사회민주주의'에서 2음절 명사인 주의의 끝음절 '의'와 조사 '의'가 중첩되지만 이를 조사로 인식하여 분리함으로써 복합명사 분해를 실패하게 된다.

### 2.2.4 복합명사의 중의적 분해문제

단위명사로의 분해과정에서 복합명사의 분해 후보는 두 종류 이상 존재할 수 있으므로 이들 가운데 최적의 후보라고 판단되는 단위명사를 선택해야 한다.

### 2.2.5 미등록어가 포함된 복합명사 분해

명사사전에 실제 사용되어지는 단위명사를 전부 수록하는 것이 이상적이지만, 현실적으로 인명이나 지명, 회사명, 기관명, 신조어 등과 같은 단어는 미등록되는 경우가 많아서 복합명사를 분해하는데 실패하는 원인이 된다.

## 2.3 기존 연구

복합명사 분석에 대한 기존 연구는 크게 복합명사 구성의 음절패턴을 정형화하여 이를 적용하는 기법과 통계적 정보를 이용한 기법 그리고 이 두 가지를 혼용한 기법으로 나눌 수 있다.

복합명사 음절 패턴기법은 복합명사의 길이에 따라 음절별로 구성 패턴을 미리 순서적으로 정한 다음 이에 맞게 명사를 분해하는 방식이다. 통계적 기법은 먼저 복합명사의 구성패턴을 적용하여 단위명사를 분해한 후 이 과정에서 단위명사의 중의적 분해나 의미의 모호성이 발생할 경우 통계 정보를 이용하는 방법이며, 여기에서 사용된 통계 정보는 단위명사가 복합명사 내에서 중심어로는 몇 번 사용되었으며 수식어로는 몇 번 사용되었는가 하는 단순한 빈도수 정보를 말한다. 최근들어 길이가 긴 복합명사, 중의적 분해, 미등록 단위명사 등을 처리할 수 있는 다양한 기법들이 제안되고 있다. 말뭉치로부터 획득한 음절간 상호정보를 바탕으로 분리점을 찾고 이 점을 기준으로 주어진 복합명사를 이등분함으로써 복합명사

의 길이에 상관없이 분석할 수 있는 방식, 복합명사에서 가능한 분해후보를 생성하고, 분해 후보에 대해 가중치를 부여함으로써 최적 후보를 선택하는 방식, 통계정보와 선호규칙을 이용하여 복합명사를 단위명사로 분해하는 방식 등이 제시되었다.

궁극적으로 복합명사 분석 알고리즘은 단위명사의 중의적인 문제를 해결함과 동시에 미등록어가 포함된 복합명사와 그렇지 않은 복합명사, 그리고 길이가 긴 복합명사에 대해서 일관성있게 적용될 수 있도록 구축되어야 할 것이다.

### 3. 복합명사의 의미론적 분석

#### 3.1 동작명사와 복합명사의 관계

한국어의 언어학적 특징 중 하나는 동사성 명사가 다수 존재한다는 것이다. 동작명사의 개념은 '하다', '되다' 등이 결합되어 용언의 역할을 할 수 있는 명사를 말한다. 예를 들면, 정보검색에서 검색이라는 명사는 '하다'가 붙어서 '검색하다'가 되어 동사로 활용이 가능한 명사이기 때문에 동작성 명사에 해당되고, 정보라는 명사는 '하다' 또는 '되다'가 결합되어서 쓰일 수 없는 명사이므로 비동작성 명사이다.

동작성 명사로 분류되는 명사가 복합명사 내의 어느 위치에 어떻게 나타나느냐에 따라 복합명사의 의미구조가 달라진다. 복합명사의 구조자체가 통사적(syntactic) 구조라기 보다는 의미적(semantic) 구조를 형성하고 있으므로 그 명사 자체가 동작성 명사의 여부에 따라 큰영향을 미친다. 동작명사가 포함된 복합명사는 '주어-술어', '목적어-술어', '부사어-술어' 등 여러 가지 관계를 이루고 있다. 예컨대, 공동분담방식의 복합명사는 '공동으로 분담한다'로 표현할 수 있으므로 '부사-술어' 관계를 이룬다.

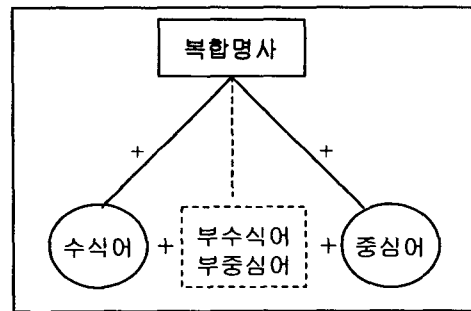
#### 3.2 복합명사의 의미적 관계

복합명사는 단위명사들이 일정한 구문론적인 의미관계로 결합된 것이다. 단위명사들이 결합되어 복합명사화된 후에는 이것들이 한 덩이로 융합되어 한 구문단위를 이루지만, 이것들은 본래의 구문론적 관계로 이어진 것이라고 할 수 있다. 즉, 복합명사를 구성하는 단위명사들은 의미적으

로 상호간에 긴밀한 관련성을 유지하고 있는 것이다. 따라서 복합명사 분석시에는 좌, 우 단위명사간의 어의관계를 정확히 파악하는 것이 필요하다.

복합명사의 형태구조는 '손발'과 같이 두 단위명사 성분이 대등한 관계로 결합된 것을 제외하고는, 일반적으로 수식어(specifier)와 중심어(head)를 갖는 종속관계의 구조를 형성하고 있다. 그리고 복합명사가 여러 개의 단위명사로 이루어져 있을 경우 수식어와 중심어의 중간에 나오는 단위명사는 부수식어나 부중심어로 구분할 수 있다. (그림 1)은 복합명사의 형태적 구조를 나타내고 있다.

여기에서 '+'의 의미는 복합명사의 구성요소가 한 번이상 출현해야 한다는 것을 표시한 것이다.



(그림 1) 복합명사의 형태적 구조

복합명사의 의미구조는 주로 특수의미의 복합명사와 변형생성적 복합명사로 구분할 수 있으며, 특수의미의 복합명사는 ① 독립된 단어로서의 의미적 결합을 할 수 없는 은유적 복합명사 ② 속어적 복합명사 ③ 독립된 단어의 의미자질과 접사형 명사가 결합된 접사형 복합명사 등의 유형들이 이에 해당된다. 이러한 복합명사는 사전적 복합명사로서 단위명사가 결합되어 새로운 언어적 의미를 갖게 됨으로 구조분석이 불가능하다. 변형생성적 복합명사는 표층구조의 의미가 심층구조의 의미를 축약하여 내포하고 있기 때문에 복합명사의 표층구조는 구조분석을 통해 문장의 변형단계를 거쳐 심층구조에서 도출할 수 있다. 심층구조에서 얻을 수 있는 표층구조의 의미관계는 '원인', '소유', '생산', '도구', '동격', '장소', '목적', '기원', '화제(話題)', '유사(類似)' 등으로 분류할 수 있다.

이상과 같이 복합명사가 갖는 의미적 구조관계를 정확히 파악하여, 복합명사 분석시 발생될 수 있는 단위명사의 중의적 분해나 모호성 문제 또는 문장 내에서 독립된 단위명사간의 의미적 결합에 의한 복합명사 인식 등의 문제를 처리해야 한다. 따라서 이러한 의미구조 이론을 복합명사 분석에 활용하기 위한 의미규칙을 제시하면 다음과 같다.

<의미규칙 1> 특수어미 복합명사인지 변형생성적 복합명사인지를 구분한다.

특수어미의 복합명사는 한 덩이로 융합되어 굳어진 한 단어로서 명사사전에 가능한한 등록하여 분석이 불가능한 구조는 미리 제거한다. 예) 몸집, 고투자울, 심장마비 등

<의미규칙 2> 변형생성적 복합명사의 형태구조를 수식어(부수식어)와 중심어(부중심어)로 나눈다.

한국어에서는 선행성분이 한정기능(수식어)을 나타내고 마지막 성분이 피한정어(중심어)가 되는 것이 보통인데, 영어에서는 수식어가 뒤에 오는 것도 있지만 한국어에서는 역수식 복합명사는 거의 존재하지 않는다.

<의미규칙 3> 복합명사의 단위명사가 동작명사인지의 여부를 확인한다.

동작명사를 중심으로 해서 '주어-술어', '목적어-술어', '부사어-술어', '관형형 용언'의 관계를 형성한다.

<의미규칙 4> 복합명사의 표층구조를 심층구조의 문장으로 변환한다.

복합명사의 표층구조는 심층구조의 의미를 축약하여 내포하고 있으므로 하나의 문장으로 전환이 가능하다. 반대로 심층구조의 문장은 표층구조의 복합명사로 축약할수 있다.

예) 골프장사업자 ↔ 그 사업자는 골프장을 운영하고 있다.

<의미규칙 5> 심층구조의 문장에서 복합명사의 수식어(부수식어)와 중심어(부중심어)의 어의 관계를 파악한다.

예) 원인, 소유, 생산, 도구, 동격, 장소, 목적, 기원, 화제(話題), 유사(類似)

#### 4. 결론 및 향후과제

명사 중심의 색인어 선정에서 중요한 문제가

되는 것이 복합명사이다. 이는 복합명사가 띄어쓰기가 자유롭고, 동일한 의미를 가지는 복합명사가 표면적으로 여러 가지 다른 유형으로 표현될 수 있고, 단위명사가 복합명사로 결합되면서 구문적 의미관계를 내포하기 때문이다.

본 논문에서는 자동색인어 추출시 중요한 문제로 대두되고 있는 복합명사의 구조가 단순히 단위명사간의 결합관계로만 구성된 것이 아니라, 이들 간에는 의미적으로 긴밀한 함축적 관계가 있다는 가정하에서 복합명사의 표층구조를 심층구조로 분석하여 올바른 단위명사를 선별하거나 독립된 단위명사가 상호 결합할 수 있도록 의미규칙을 5가지 제시하였다. 이 규칙은 복합명사에서 단위명사로의 분해과정 또는 단위명사간의 결합관계에 동시 적용할 수 있다는 장점이 있다. 향후 이 의미규칙을 토대로 해서 정보검색분야에 이를 적용할 수 있는 알고리즘 및 구문분석기를 개발하고 이에 대한 실험, 평가가 이루어 져야 할 것이다.

#### 참고문헌

1. 강승식. 1998. 한국어 복합명사 분해 알고리즘. 정보과학회논문지(B), 25(1) : 172 ~182.
2. 백용학. 1982. 복합명사의 구조와 그 의미에 관한 연구. 동아대학교 동아논총, 제19집 : 85~112.
3. 심광섭. 1997. 합성된 상호 정보를 이용한 복합명사 분리. 정보과학회논문지(B), 24(11) : 1307~1317.
4. 윤보현 외. 1997. 통계 정보와 선호규칙을 이용한 한국어 복합명사의 분해. 정보과학회논문지(B), 24(8) : 900~909.
5. 양성현. 1997. 복합명사 분석을 통한 자동 색인 및 정보검색 시스템의 구현. 연세대학교 대학원 석사학위논문.
6. Kenneth Church and Patrick Hanks.1990. Word association norms, mutual information, and lexicography. Computation Linguistics, 16(1) : 22~29.
7. Juntae Yoon, Mansuk Song. 1997. Yet another compound noun analysis using co-occurrence relation. In Proceedings of NLPRS 97.