언어 자료를 활용한 한국어 복합명사 구조 분석

김동성
(고려대학교)

Kim, Dong-Sung. 2011. Analysis of Korean Complex Nominals, Using Linguistic Data. The Linguistic Association of Korea Journal. 19(3): 129-150. This paper introduces an analysis on the structure of Korean complex nominals. The analysis attracts both theoretical linguistics and language processing related areas, such as information retrieval and speech synthesis. Our approach has three stages. First, we identify endocentric data and exocentric data, using human intuition. Since exocentric data does not have the internal structure, we do not consider exocentric data. Second, we do the bracketing experiment on the endocentric data for representing a hierarchical structure of constituent parts, using statistical collocation measurements based on the 10 million Sejong corpus. The last stage is composed of several processes to figure out head-modifier or predicate-argument relations, using argument structure and selection restriction specified in the Sejong electronic dictionary. Our method is based on not only the corpus-based materials but also linguistic knowledge (with intuition-based judgement). The importance of our approach is to show how to use language resources to utilize linguistic knowledges in analyzing linguistic data.

Key Words: Morphology, Collocation, Complex Nominals, Computational Linguistics, Corpus Statistics, Computational Morphology, Endocentricity/Exocentricity, Bracketing

1. 머리말

Spencer (1993)의 본격에 따르면 영어에서 세 개의 명사가 연쇄적으로 만들어내는 복합 명사 구조의 경우에 (1)은 괄호매김(Bracketing)을 어떻게 하느냐에 따라서 구조적으로 (2a) 나 (2b)로 분석될 수 있다.

(1) student film society
(2) a. [student [film society]] (=film society for students)  
b. [[student film] society] (=society for student films)  

한국어 복합명사 구조의 경우도 (2a, b)와 같은 구조적 종의성이 발견된다. 예를 들어서 '한국 교육사연구소'라는 복합명사 구조는 (3a)나 (3b)의 구조 분석 중 하나가 가능하다. 

(3) a. [[한국 교육사] 연구소]  
(=한국 교육사들의 역사를 연구하는 연구소)  
b. [한국 [교육사 연구소]]  
(=한국을 포함한 세계의 교육사를 연구하는 연구소로 한국에 위치함)

복합명사 구조는 먼저 내심구조(endocentric) 합성 방식인지, 외심구조(exocentric) 합성 방식인지 구분한다. 외심구조는 동적 합성어를 포함하는데, 내심구조의 경우에는 각 구성 요소들 간 관계성이 문장성분 사이에서 관찰되는 관계와의 유사성을 분석할 수 있지만, 외심구조 합성 방식은 이러한 구조 분석이 불가능하다. Spencer (1993)에 따르면 이러한 내심 구조 복합 구조 분석에는 형태론적 정보 외에에도 동사론적 정보도 필요하다. 복합 구조와 함께 형태에 관계되는 부분과 분석된다.

복합명사 구조 분석의 중요성을 살펴보면 다음과 같다. 첫째로는 복수구 합성의 생산성 (Levi 1978)과 연관된 측면이다. 복수구 합성은 생산성이 높은 형태론적 과정으로 만약 언어 학적으로 예측이 가능하지 않다면 새로운 구조가 생성될 때마다 새로운 구조에 대한 분석이 있어야 하므로 효용적이지 않다. 따라서 합성구조 생산성에 대한 언어학적 분석에 대한 예측이 가능해야 하며, 이러한 복수 합성구조에 대한 연구는 언어 이론적 측면에서 형태론적 합성어 분석의 측면에서 주요하다(Spencer 1993; Selkirk 1982; Lieber 1983; Di Scullio and Williams 1987; Sprott 1985; Roepr 1988).

둘째로는 정보패턴을 위한 목적인 셀리어를 작성할 때, 목적어 작업에서 복합명사구의 분석은 중요하다. 예를 들어서 '강남 성심 병원'과 '강북 성심 병원'은 '성심 병원'이라는 셀리어로 통일할 수가 있으므로, [강남 (강북) [성심 병원]]의 구조로 분석할 수가 있다. 이와 같이 복합명사 구조의 분석은 정보 처리의 관점에서 필수적이다.

마지막으로 음성합성이나 음운론 연구의 경우에 빠져 넘기는 운율의 단위를 결정하는 주요 요소이다. 빠져 넘기려는 경우에 '리더리 스포 감독'의 구조는 [[리더리 스포] 감독]이지 [리