

# 세종명사전자사전의 통사정보

홍재성<sup>U</sup> 고길수  
서울대학교 불어불문학과  
{cshong, vthm044}@snu.ac.kr

## Syntactic informations in the Sejong Electronic Dictionary of Korean Nouns

Chai-song Hong<sup>U</sup> Kil-Soo Ko  
Dept. of French Language and Literature, Seoul National University

### 요 약

세종전자사전은 한국어 어휘의 정보를 총체적으로 표상하는 전산어휘자료체이다. 세종전자사전은 궁극적으로 다양한 유형과 기능의 한국어 자동처리에의 활용을 목표로 한다.

세종체언전자사전은 최종적으로 5만여 항목을 대상으로 구축될 것이다. 세종명사전자사전은 전산적 효율성을 고려하여 명사 어휘의 정보를 8개의 하위정보구획과 50여개의 세부정보항목으로 구분하여 표상한다. 특히, 명사의 어휘·통사적인 결합에 관한 정보는 한국어 명사구와 문장의 자동생성에 직접 활용될 수 있는 정보이다. 명사는 수식어 요소 또는 조사와 결합하여 명사구를 형성하고, 동사 및 형용사와 결합하여 문장의 생성에 참여한다. 개별 명사들은 이 과정에서 다양한 제약적인 양상들을 보여주고 있으며, 세종명사전자사전은 이 정보들을 명시적으로 표상한다.

또한 명사는 기능동사와 결합하는 술어명사와 그렇지 않은 비술어명사로 구분이 된다. 술어명사가 기능동사와 결합하여 문장을 형성할 때, 명사와 그 논항들은 다양한 통사적 기능을 담당한다. 또한 술어명사는 논항과 결합하여 명사구를 형성한다. 그러나, 술어명사의 명사구형성과 기능동사의 결합은 불규칙적이고, 명사와 기능동사의 의미적 특성에 따라 다양한 제약이 발견된다. 이 정보들의 정밀한 표상은 개별 술어명사로부터 생성될 수 있는 가능한 명사구와 문장의 형태를 구체적으로 밝혀주게 된다.

세종명사전자사전의 어휘·통사적인 결합에 관한 정보들은 명사구 또는 문장의 자동생성의 정확성과 효율성을 높여줄 것이다.

## 1. 예비적 지적: 세종전자사전

### 1.1. 세종전자사전의 목표와 의의

세종전자사전(이하 세종사전으로 약칭함)은 현대 한국어를 모국어로 하는 화자가 활용할 수 있는 어휘정보를 총체적으로 표상하는 정밀한 전산 어휘자료체의 구성을 목표로 한다. 세종사전은 분석, 기술 및 표상에 있어, 언어학적 타당성, 심리적 실재성과 전산적 효율성을 최대한 조화시키면서 이 목표를 달성하고자 한다.

세종사전은 궁극적으로 다양한 유형과 기능의 전산 어휘자료체로 가공되어 한국어 자연언어처리를 위한 프로그램의 개선과 개발에 활용될 것이다. 그러나 또 한편으로 웹 서비스를 통해 한국어 화자와 외국인들이 참조하고 이들을 대상으로 하는 한국어 교육을 위한 언어사전으로 활용될 것이며, 세종사전에 표상되는 어휘지식은 한국어를 기반으로 하는 제반 연구의 기초자료로 제공되기도 할 것이다.

세종사전은 한국어를 기반으로 하는 모든 영역에서 다양한 기능을 수행하는 범영역성과 범기능성을 갖는다. 범용성으로 요약되는 세종사전의 특성은 기존의 인쇄사전 및 특수한 영역의 프로그램에 사용되는 전산 어휘자료체와 구분된다. 특히, 범용성은 전체 사전의 성격과 그 구성 및 표상 방법을 결정하는 핵심 요인으로 작용하여, 세종사전에는 아래와 같은 특성들이 추가적으로 나타나고 있다.

- 이론중립성
- 모듈적 구성
- 형식적 표상
- 연동성

### 1.2. 세종전자사전의 특징

첫째, 세종사전은 다양한 한국어 연구와 자연언어처리

프로그램 지원을 위한 이론중립적인 전산 어휘자료체이다. 세종사전은 이를 위해 말뭉치, 기존의 인쇄사전 및 연구자료 등의 풍부한 어휘자원을 활용한다. 이를 통해 내성적인 분석의 한계를 보완하여 언어사용의 현실성을 반영한 정밀하고 풍부한 어휘지식을 형식적이고 효율적으로 표상하고자 한다.

그러나, 인쇄사전과 같은 기존의 자원은 전산적 활용의 측면에서 정보의 정밀성과 체계성 등이 결여되어, 직접 활용하는 데에는 한계가 있다. 이에 세종사전은 최신의 언어학과 자연언어처리분야의 연구성과들을 도입하여, 한국어 어휘를 철저히 재분석한다. 어휘의 범주, 범주별 정보유형, 범주별 사전의 구조 등이 주요 쟁점들이며, 특히 세종사전은 범주별 어휘들의 통사정보에 중점을 두고 이를 정밀하게 분석하여 기술하고자 한다.

둘째, 세종사전은 사전의 구성과 정보 표상에 있어 하위모듈이 통합된 형태로 구성된다. 세종사전은 기본적인 문법범주와 형태정보만이 부착된 대규모 기초전자사전과 전체 어휘정보가 부착된 상세전자사전으로 나뉘고, 이들은 다시 개별 문법범주별 사전으로 분할되어 개발된다. 특히, 세종사전은 전자사전으로써의 효율성과 활용성의 극대화를 위해 기성 전자사전의 구성에서 미진하였던 조사, 어미, 관형사, 부사 등의 문법범주와 언어, 관용표현 등을 분리하여, 별도의 사전으로 개발한다.

모듈적 구성은 정보 표상에도 반영된다. 범주별 어휘들에 관한 정보들을 유형화하고, 별도의 구획으로 구분하여 표상한다. 이에 따라 범주별 개별 항목은 여러 하위구획으로 나뉘고, 각 하위구획은 다시 여러 세부정보항목으로 분할되는 위계적인 모듈의 형태를 갖는다<sup>1)</sup>.

세종사전은 모듈적 구성으로 개선이 용이하다. 세종사전은 언어학과 자연언어처리 분야의 발전에 따라 지속적으로 보완·개선되어야 하지만, 그 방대한 규모로 인해 이 작업은 만만치 않은 작업이 될 것이다. 그러나, 모듈적 구성으로 이 작업은 특정 모듈에만 국한되어 행해질 수 있다. 개선의 용이성으로 세종사전은 향후 예상되는 획기적인 인공지능 및 전산기술과 쉽게 접목될 수 있는 융통성을 보인다.

셋째, 세종사전은 즉각적인 전산 활용을 위해 형식적인 표상을 한다. 세종사전은 XML, SGML의 방식을 도입하여 전산학 분야의 데이터 처리 및 표상 방법을 따른다. 즉, 위의 방식이 정한 문법에 따라 개별 어휘 항목과 세부정보항목은 특정 표지를 통해 구분되고, 일정한 원칙에 따라 구체적인 정보들이 표상된다. 이를 통해 세종사전과 세부 정보들은 현 상태 그대로 전산 프로그램에 활용될 수 있으며, 다른 유형의 데이터베이스 형태로도 쉽게 변환될 수 있다.

마지막으로 세종사전은 타분야 및 위성사전과의 연동성을 추구한다. 한국어 자동처리 지원이 기본 목표인 세종사전은 다각적인 언어공학적 타당성의 검증과 평가를 추구한다. 이는 앞서 언급한 융통성과 맞물려 직접적으로 이 분야에 활용될 수 있는 전자사전의 개발을 가능

하게 할 것이다. 또한, 세종사전은 하나의 핵심 모듈이 되어, 고유명사, 전문용어, 방언, 북한어, 구어, 음성사전 등의 위성사전과의 연계 및 연동성을 추구하기 위해 필요한 정보들을 표상한다.

## 2. 세종체인전자사전

### 2.1. 세종체인전자사전의 특징

세종체인전자사전(이하 체언사전으로 약칭함)은 대표적인 하위범주 전자사전이며, 최종적으로 5만 항목의 상세전자사전과 12만 항목의 기초전자사전의 개발을 목표로 한다<sup>2)</sup>.

체언사전은 모듈적 구성에 따라 몇 개의 하위사전으로 분할된다. 한국어의 체언은 전통적으로 보통명사, 의존명사, 대명사, 수사어 범주로 구분되지만, 체언사전은 일반명사와 의존명사를 그 속성에 따라 다시 하위의 범주로 분할한다.

- (1) a. 문, 불, 나무, 책...
- b. 철문, 방문(房門), 문지기, 문지방, 문고리...
- (2) a. 참, 동안, 때, 가운데...
- b. 명, 분, 개, 마리, 잔, 키로그램...

(1)은 모두 보통명사에 속하지만 그 내적인 구조에 따라 단일형태와 복합형태로 구분된다. (2a)와 (2b)는 의존적인 속성을 공유하지만, 이들이 구성하는 명사구는 전혀 다른 양상을 보인다. 체언사전은 이를 고려하여 (1a)와 (1b)를 각각 단일어명사사전과 복합어명사사전으로 분할하여 기술하고<sup>3)</sup>, (2a)와 (2b)는 의존명사사전에 별도의 항목을 설정하여 필요한 정보들을 기술한다.

모듈적 특성은 체언사전의 사전구성에도 반영된다. 체언사전은 명사 어휘에 관한 정보를 유형에 따라 먼저 형태, 속어, 의미의 하위구획으로 구분하고, 형태와 의미 구획은 다시 세부구획으로 분할된다. 예를 들어, 형태구획은 (1)의 예와 같이 복합형태의 내적인 구조를 표상하는 구획과 표제어의 합성 및 파생에 관한 정보를 표상하는 구획으로 분할된다. 또한 (2)의 의존적 어휘의 명사구 형성에 관한 정보는 의미구획의 하위구획인 통사구획에 표상된다.

### 2.2. 세종명사사전의 구성

세종명사사전(이하 명사사전으로 약칭함)은 체언사전의 구조에 따라 8개의 하위구획과 50여개의 정보항목으로 구성된다<sup>4)</sup>. 명사사전의 세분화된 정보구조는 명사 어휘에 대한 정보를 정밀하게 분석하고 형식적으로 표상하기 위한 것이다.

예를 들어, 형태구획은 전술한 것과 같이 표제어의 내적인 형태에 관한 구획과 다른 어휘와의 형태적인 관계 및 결합에 관한 구획으로 구분된다. 두 구획은 개별 정보를 기술하는 여러 세부정보항목으로 구성된다. 형태적

결합정보는 결합하는 요소의 범주와 결합하는 양상에 따라 파생(der), 합성(comp), 첨어 및 혼성어(metc) 항목으로 분할되며, 결합된 어휘의 범주에 따라 명사(n), 동사(v), 형용사(a), 부사(av)로 구분되어 표상된다.

다음으로 어휘의미관계구획을 살펴보기로 하자. 명사 어휘들은 아래의 예에서와 같이 다양한 어휘의미관계를 갖는다.

- (3) a. 아버지/아빠/부친; 죽음/사망; 옷/의복...
- b. 삶/죽음; 고온/저온; 아버지/어머니...
- c. 개/동물; 사과/과일; 비둘기/새...
- d. 손가락/손; 수화기/전화(기); 입/얼굴...
- e. 영화/극장/배우; 화재/소방차/소방관...

(3a)가 적어도 하나 이상의 문맥에서 서로 대치되는 동의어 관계를 보여준다면, (3b)는 의미적 대립성이 있는 반의어 관계에 해당한다. (3c)는 의미 또는 개념의 관계망을 축으로 상·하위의 관계에 있는 명사들이며, 사과, 배, 딸기, 수박 등과 같이 동일한 상위어를 갖는 어휘들의 관계도 발견된다. (3d)의 명사들은 전체와 부분의 관계에 있는 것들이고, (3e)는 의미 또는 개념적으로 밀접한 관계를 갖지만 앞의 네 유형의 관계에는 속하지 않는 명사들이다. 명사사전은 다양한 어휘의미관계를 별도의 정보항목으로 구분하여 표상한다.

명사사전은 이와 같은 세분화된 정보표상으로 즉각적으로 한국어 자연언어처리에 활용될 수 있을 것이다. 특히 명사사전은 자연어 생성(production, generation)의 층위를 염두에 두고 정보를 표상한다. 기존의 자연언어처리 프로그램들은 대부분 입력되는 자연어 문장을 분석하고 해석하는 층위에 중점을 두고 있으며, 전자사전은 이러한 기능을 최대한 지원하는 수준에 머물러 있었다.

반면, 자연어 생성에 필요한 정보들은 개별 명사가 다른 요소들과 결합하여 구 또는 문장을 구성하는 정보이다. 이를 고려한 명사사전은 어휘·통사적 층위에서의 결합에 관한 풍부한 정보들을 수록한다. 예를 들어, 통사정보구획의 형용사언어항목을 살펴보자. 이 항목은 아래와 같이 명사의 의미적 강도를 표현하는 형용사 형태를 제시한다<sup>5)</sup>.

- (4) a. 새빨간 거짓말
- b. (높은+희박한+희미한) 가능성
- c. 큰 기대

그런데, 이러한 형용사의 구체적인 형태는 개별 명사에 따라 다르게 실현된다.

- (5) a. 피로가 극심하다
- b. 저항이 완강하다
- c. 조사가 철저하다

또한, 많은 한국어 형용사들은 서술적 용법과 관형적 용법이 모두 가능하지만, 형용사언어항목의 형용사들은 표제어 명사에 따라 하나의 용법으로만 사용되기도 한다.

- (6) a. \*거짓말이 새빨강다 (새빨간 거짓말)
- b. \*깊은 유서 (유서가 깊다)
- c. \*높은 악명 (악명이 높다)
- d. \*과도가 성나다 (성난 과도)

이와 같이 명사와 의미적 강도를 표현하는 형용사의 결합은 어휘계약적인 불규칙성을 보인다. 이들의 결합이 자동으로 생성되기 위해서는, 개별 명사에 대해 형용사의 형태와 그 용법에 대한 구체적인 정보가 반드시 필요하다. 이 정보들을 명시적으로 기술하는 항목이 명사사전의 형용사언어항목이다<sup>6)</sup>.

### 2.3. 세종명사사전의 의미구분

한국어 자동 생성의 관점은 다의어 분할과 의미구분의 문제에도 중요한 요인으로 작용한다. 즉, 의미변이를 보이고 각 의미가 서로 다른 어휘·통사적인 양상을 보이면, 이들은 다의어로 분할되어 기술된다. 구체적인 예를 통해 살펴보기로 하자.

- (7) a. 장미를 (심다+꺾다)
- b. 은행이 (끝나다+무너지다+친절하다)
- c. 질문을 (하다+이해하다)

(7a)는 <식물><sup>7)</sup>과 <꽃>을, (7b)는 <기관>, <건물>과 <인간>을, (7c)는 <언어행위>와 그 <내용>을 의미한다. 이들은 광범위한 어휘들에서 발견되는 규칙적 또는 체계적 다의성(regular or systemic polysemy)<sup>8)</sup>으로 분류된다. 많은 사전들은 규칙성을 기반으로 이들을 하나의 의미로 통합하여 표상한다<sup>9)</sup>.

그러나, 이 명사들을 자세히 살펴보면, 의미변이에 따라 어휘·통사적으로 다른 형태들이 발견된다.

- (8') a. 장미를 한 그루를 (심다+꺾다)
- b. 장미를 한 송이를 (\*심다+꺾다)

- (9) a. 은행이 업무가 (끝나다+무너지다+친절하다)
- b. 은행이 건물이 (\*끝나다+무너지다+친절하다)
- c. 은행이 직원들이 (\*끝나다+무너지다+친절하다)

- (10) a. 그는 나에게 질문을 (하다+이해하다)
- b. 그는 나의 질문을 (\*하다+이해하다)

이러한 의미변이와 그 행태의 차이는 온전한 한국어 문장 생성에 반드시 필요한 정보이다. 하나의 의미로 통합하여 표상하는 것은 의미변이와 이에 따른 행태의 차이가 전체 어휘에서 규칙적으로 나타날 때에만 가능

하다. 그러나, 이에 대한 적절한 판단을 내리기는 이론적으로나 실제적인 측면에서 매우 어려운 과제이다.

또한, 의미변이의 규칙성은 다양한 반례를 통해 쟁점화될 여지가 있다<sup>10)</sup>.

- (11) a. 개나리 한 (\*송이+그루)  
b. 국화 한 (송이+\*그루)
- (12) 정부가 (\*끝나다+\*무너지다+신중하다)<sup>11)</sup>
- (13) a. 그는 나에게 반박을 했다  
b. 그는 나의 반박을 ?\*이해했다<sup>12)</sup>

이에 명사사전은 어휘·통사상 행태의 차이가 명확히 구별되는 의미변이는 모두 다의어로 분할하고, 각 의미에 대해 구체적인 정보들을 표상하는 방법을 채택한다. 한편, 일부 명사는 의미의 관련성에도 불구하고, 다중범주어로 간주되어 타범주의 세종하위사전에 기술된다. 예를 들어, 많은 <용기> 명사들은 물질명사의 <양>을 표현하는 단위명사로 사용된다.

- (14) a. 잔을 (비우다+들다+깨다)  
b. (술+물+주스) 한 잔  
c. 한 잔의 (술+물+주스)

(14b-c)의 잔은 문법적으로 수사 및 물질명사와 결합하여 명사구를 형성하고, 의미적으로도 한정된 부류의 명사들과 결합한다. 이러한 행태는 분류사와 동일한 것으로 (14a)와는 구분된다. 따라서, 체언사전은 잔을 동형어인 다중범주어로 분할하여, 명사사전과 의존명사사전에 별도의 항목으로 처리한다. 이 원칙은 <시간>, <장소>, <양>을 나타내는 어휘들에도 적용되고 있다.

- (15) a. 오늘, 어제, 지금, 잠깐...  
b. 여기, 저기...  
c. 약간, 조금, 전부, 모두...

위 어휘들은 격조사와 결합하여 술어의 논항 역할을 담당할 수도 있고, 격조사 실현 없이 부가어로 사용될 수도 있다. 즉, 이 어휘들은 체언과 부사로서의 용법을 모두 보여준다. 세종사전은 이 어휘들을 명사사전과 세종부사전자사전에 모두 등재하여, 각각의 용법에 대응되는 다양한 속성을 정밀하게 기술한다.

### 3. 세종명사사전의 통사정보구획

#### 3.1. 통사정보구획의 구성

통사정보구획(이하 통사구획으로 약칭함)은 주로 표제어가 통사 단위를 형성하는 과정에서 다른 요소와 결합하는 관계를 표상한다.

명사 어휘가 단독으로 문장 내에 실현되어 통사적인 기능을 수행하는 것은 드문 경우로, 일반적으로는 적어도 하나 이상의 다른 요소와 결합한다. 이 정보는 특성에 따라 여러 하위유형의 정보로 구분될 수 있다. 우선 결합하는 요소의 범주와 분포 위치를 고려한 구별이 가능하다. 다음으로 결합 양상에 따라 어휘계약적인 것과 통사규칙을 따르는 자유로운 것으로 나눌 수 있다. 이때 공기하는 어휘요소가 명사의 의미에 의해 요구되는 것인지를 고려할 수 있다. 마지막으로 결합하여 형성된 단위에 따라, 온전한 통사적인 기능을 수행하는 통사 단위와 그렇지 않은 명사구로 분할될 수 있다.

명사사전은 이들을 고려하여 통사구획을 세 개의 하위구획으로 분할하여 정보항목들을 표상한다. 통사구획 A는 어휘계약적으로 결합하는 어휘요소들에 대한정보를, 통사구획 B는 연어정보 및 술어와 결합하여 통사단위를 형성하는 정보를, 통사구획 C는 한정사 및 수식어 요소에 대한 정보가 표상된다.

통사구획의 정밀한 결합정보는 한국어 자동 처리의 측면에서 결합하는 어휘요소들을 구체적으로 밝혀줌으로써 분석의 속도와 효율성을 높여줄 것이다. 특히, 어휘계약적인 결합정보와 실현 가능한 형태에 대한 정보는 기계번역, 문서요약 등과 같이 문장 생성을 요구하는 프로그램의 정확성을 향상시킬 것이다.

특히, 통사구획은 술어명사(predicate noun)의 정밀한 결합정보를 표상한다. 명사는 속성에 따라 술어명사와 비술어명사로 구분된다. 술어명사는 기능동사(support verb)와 결합하여 특징적인 단문을 구성한다. 이 과정에서 의미적으로 내포된 논항들이 통사적으로 실현된다. 또한, 이 문장은 통사적인 변형을 통해 술어명사를 머리어(head)로 하는 명사구로 환원되며, 논항들은 조사와 결합하여 수식어로 나타날 수 있다. 이런 점에서 술어명사의 문장과 명사구 구성에 대한 통사정보의 표상은 문장 생성과 같은 한국어 자동처리에 절실하게 요구된다.

통사구획은 전술한 바와 같이 세 개의 하위구획으로 나뉘며, 각 구획의 세부정보항목들은 아래와 같다.

- 통사구획A: 단위표현(cl)/조사결합계약(prt)/부사적 용법(av)/연결어미적 용법(ds)
- 통사구획B: 형용사자유구성(comb\_aj)/형용사연어(magn)/동사자유구성(comb\_v)/명사연어(comb\_n)/기능동사(sup\_v)/최대논항명사구(max\_n)/논항선택계약(sel\_res)
- 통사구획C: 한정사계약(pre\_d)/수식어명사(pre\_n)/수식어문장(pre\_s)

여기에서는 조사결합계약, 수식어문장의 두 항목만을 간략히 살펴보기로 한다.

#### 3.2. 조사결합계약항목

한국어 명사는 일반적으로 여러 유형의 조사들과 자유롭게 결합한다. 그러나, 일부의 명사들은 보조사를 제

외한 몇 개의 조사들하고만 결합하고 있다.

- (16) a. 암전을 (떨다+빠다)
- b. 자타가 (공인하다+인정하다)
- c. 입추의 여지

(16)의 명사들은 다른 조사들과는 쉽게 결합하지 않는다. 이러한 제한된 결합양상은 조사이외의 어휘요소들에서도 발견된다.

- (16') a. \*암전-(이+으로)
- b. \*자타-(를+의+로)
- c. \*입추-(가+를+로)
- (16'') a. ??암전을 (칭찬하다+비난하다+묘사하다)
- b. \*자타가 (칭찬하다+비난하다+주고 받다)
- c. \*입추의 (가능성+틈+실마리)

즉, 이 명사들은 특정 어휘와 결합하며, 통사적으로 제한된 기능만을 수행하는 것으로 분석된다. 이러한 제약적인 양상을 기술하는 항목이 조사제약항목이다. 이 항목은 결합이 가능한 구체적인 조사의 형태를 제시하며, 조사 이외의 어휘요소에 대한 정보<sup>13)</sup>를 추가적으로 기술한다.

### 3.3. 수식어문장항목

다음으로 수식어문장항목은 명사와 결합하는 관형절 형태의 수식어 문장에 관한 정보를 표상한다. 많은 경우, 수식어 문장은 의미적으로 명사와 밀접한 관련성을 갖고 있는 것으로 분석된다.

- (20) a. 나는 그가 뛰는 광경을 보았다
- b. 나는 자리에 앉으라는 명령을 듣지 못했다

(20a)의 수식어 문장은 의미적으로 명사의 내용을 밝혀주는 동격절에 해당하며, (20b)의 관형절은 <언어행위>에 속하는 술어명사의 보문이 실현된 것이다. 이와 같이 명사의 의미와 밀접하게 관련되는 수식어 문장은 명사에 따르는 시상제약과 양상제약을 보인다.

- (20') a. 그가 뛰-(는+\*ㄴ+\*르) 광경
- b. 자리에 앉-(으라는+\*는다는+\*느냐는+\*자는) 명령

- (21) a. 삼관왕에 오르는 수훈
- b. 건강이 어떠하냐는 질문

명사사건은 수식어 문장의 시상 및 양상제약과 제약이 발현되는 구체적인 형태를 목록화하여, 개별 정보들을 일정한 약호를 통해 표상하고 있다.<sup>14)</sup>

### 4. 술어명사와 통사정보구획

#### 4.1. 술어명사의 통사 속성: 단문구성

술어명사는 전술한 것과 같이 기능동사와 결합하여 특징적인 단문을 형성한다.

- (22) a. 그는 사과를 했다
- b. 그는 전화를 했다

이 때, 술어명사의 의미에 포함되어 있는 논항들이 통사 논항으로 자연스럽게 실현될 수 있다.

- (23) a. 그는 나에게 사과를 했다
- b. 그는 그 일에 대해 사과를 했다
- c. 그는 나에게 그 일에 대해 미안하다고 사과를 했다

- (24) a. 그는 나에게 전화를 했다
- b. 그는 나에게 아버지가 돌아가셨다고 전화를 했다

위의 두 예문은 술어명사 사과와 전화의 통사적인 논항들이 추가되는 과정을 보여준다. 논항의 양상은 술어명사와 결합하는 다양한 기능동사들을 고려할 때, 더욱 복잡하게 나타난다. 예를 들어, 위의 두 술어명사는 피동구문과 유사한 의미를 표현하는 받다와 결합하여 다음과 같은 문장을 형성한다.

- (25) a. 나는 그에게서 그 일에 대해 미안하다고 사과를 받았다
- b. 나는 그-(에게서+로부터) 아버지가 돌아가셨다고 전화를 받았다

(25)에서 논항들은 다른 통사 기능을 담당하고 있으며, (23)-(24)와는 논항들의 기능이 전환된 관계를 보인다. 이를 통해 기능동사가 술어명사와 그 논항들의 통사기능을 결정하고 있음을 알 수 있다.

일반적으로 하다는 '능동'과 '주동'을, 되다, 당하다, 받다는 피동을, 시키다는 사동의 의미를 표현한다. 그러나, 이들 기능동사들이 술어명사와 일정하게 결합하는 것은 아니다.

- (26) a. (사과+전화)-를 (하다+받다+\*당하다+시키다)
- b. (사과+전화)-가 ?되다

- (27) a. (연결+감소)-를 (하다+??당하다+\*받다+시키다)
- b. (연결+감소)-가 되다

또한, 술어명사에 따라 기능동사들 사이의 의미관계도 불규칙적인 양상을 보인다.

- (27') a. 그는 끊어진 선을 연결을 (했다+시켰다)

b. 인구가 지속적으로 (감소를 한다+감소가 된다)

(27'a)에서 하다는 시키다와, (27'b)에서 하다는 되다와 대체될 수 있다. 그런데, 기능동사의 교체에도 불구하고, (27)의 예문들에서는 문장 전체의 의미와 논항의 통사적인 기능에 아무런 변화가 일어나지 않고 있다.

또한 일부 명사는 의미에 따라 다른 수의 논항을 갖고, 기능동사와의 결합양상에서도 다른 행태를 보인다.

- (28) a. 남과 북이 통일-(을 하다가이 되다)
- b. 그 장군이 삼국을 통일을 (했다+시켰다)

(28)에서 명사 통일은 의미에 따라 두 구조의 구문을 구성한다. 이때, 두 구문은 모두 기능동사 하다와 결합하지만, 다른 기능동사와 교체되고 있다.

#### 4.2. 술어명사의 통사 속성: 명사구 구성

한편, 술어명사와 기능동사가 구성하는 단문은 통사적 변형을 통해 명사구로 치환된다. 이때, 술어명사는 머리어로서 명사구를 구성하는 핵심 성분으로 기능한다. 즉, 술어명사를 중심으로 그 의미적 논항들이 수식어 성분으로 실현되는 것이다. 일반적으로 기능동사구문을 구성 하였던 모든 논항들은 명사구에서도 실현된다.

- (28') a. 남과 북의 통일
- b. 그 장군의 삼국 통일

그러나, 사과나 전화와 같은 명사들의 경우에는 논항들이 부분적으로만 실현된다.

- (29) a. 나에 대한 그의 미안하다는 사과
- b. 그 일에 대한 그의 미안하다는 사과

- (30) a. (\*나에 대한+E) 그의 전화
- b. 아버지가 돌아가셨다는 그의 전화

(29)에서 다른 논항들이 동일한 조사로 실현되고 있어, 이들은 한 명사구에 동시에 나타나지 못한다. 반면, (30)에서 전화의 특정 논항은 다른 논항들과 동시에 출현하지 못하고, 아래의 예와 같이 단독으로만 나타난다.

- (30') 나의 전화

이것은 명사구의 조사 사용이 문장보다는 제한되어 있기 때문이다. 이로 인해 모든 논항이 실현되지 못하기도 하며, 의미적 관계가 다른 논항들이 동일한 조사와 결합하는 것이다.

술어명사의 논항은 의미와 형태의 층위에서도 복잡한 양상을 보인다. 형태적으로 대부분의 논항은 명사 또는 대명사의 형태로 실현된다. 그러나 앞의 예들에서와 같이 문장

의 형태가 논항으로 직접 나타나기도 한다. 이때, 문장은 일정한 어미들을 통해 실현되는데, 술어명사의 의미에 따라 어미의 결합이 제한된 예들이 발견된다.

앞서 언급하였던 명사 사과가 어미의 제약 현상을 잘 보여주고 있다. 전화와는 대조적으로 사과의 보문 논항은 특정 어미로만 실현이 가능하다(3.3. 수식어문장항목 참고).

- (31) a. 미안하-(다는+\*라는+\*자는+\*냐는) 사과
- b. 공부하-(ㄴ다는+라는+자는+냐는) 전화

이러한 제약은 술어명사의 의미와 밀접하게 연관된다. 즉, 명사 사과는 의미적으로 실제 일어난 사실을 전제하고 있으며, 이를 표현하는 보문 논항은 반드시 사실성을 표현하는 어미로만 실현된다. 이와 같은 논항에 의미적인 제약 현상은 명사 논항에서도 선택제약을 통해 잘 알려져 있다.

#### 4.3. 술어명사의 통사정보항목

술어명사가 단문과 명사구를 구성한다는 점에서 이를 이용한 한국어 문장과 명사구의 자동 생성이 가능할 것이다. 그러나, 앞서 살펴본 것과 같이 술어명사는 어휘적으로 제약된 매우 복잡한 통사적인 특성을 보여주고 있다. 한국어 자동 생성을 위해서는 이에 대한 정밀한 정보가 체계적으로 표상된 전산 어휘자료체가 필수적으로 요구된다.

명사사전은 이러한 필요성을 고려하여 다음의 세 정보항목을 통해 술어명사의 단문과 명사구 구성에 관한 정보를 기술한다.

- 기능동사(supv)
- 최대논항명사구(max\_n)
- 논항선택제약(sel\_res)

기능동사항목은 단문 구성에 관한 정보를, 최대논항명사구항목은 명사구 구성에 관한 정보를, 논항선택제약항목은 논항의 의미적 제약에 관한 정보를 제공한다.

먼저 기능동사항목은 명사와 결합하는 기능동사와 논항 형태에 관한 정보를 기술한다. 이 항목은 우선 흔하게 사용되는 9 개의 기능동사<sup>16)</sup>에 국한하여, 모든 논항이 실현된 문장의 구조를 모두 나열한다. 일부의 논항만이 실현되는 구조는 논항의 삭제 규칙이 적용되어 생성되는 것을 원칙으로 한다.

예를 들어, 명사 사과의 경우 아래의 두 구문이 모두 제시된다.

- (32) a. X-가 Y-에게 Z-에 대해 S라고 ~를 하다
- b. Y-가 X-에게서/로부터 Z-에 대해 S라고 ~를 받다

논항들은 동일한 기호로 표기되고, 이를 통해 두 구문 사이의 관계가 명시적으로 표상된다. 또한 명사 논항과

구별하기 위해 보문 논항은 S로 표기된다.

그러나, (27')와 (28')과 같이 동일한 논항구조를 갖고 같은 의미를 표현하는 기능동사들은 'I' 기호를 이용해 통합되어 기술된다.

(33) 연결 : X-가 Y-를 Z-에와 ~을 하다시키다

'I'는 기능동사들의 교체가능성을 명시적으로 제시한다. 이 기호는 한 논항의 조사 교체와 명사 및 보문 논항사이의 교체를 기술하는 데에도 사용되고 있다.

한편, 이 항목에 기술되는 모든 기능동사들에 대해 술어명사와 기능동사가 통합된 형태가 존재하는지의 여부를 관련동사항목(cor\_v)에 추가적으로 제시한다. 이 정보는 합성어 정보와 더불어 세종용언사전의 해당항목의 참조를 가능하게 해준다.

최대논항구조항목은 명사구의 구조를 기능동사항목과 동일한 방식에 따라 기술한다.

(32') a. Y-에 대한 X-의 S다는 ~  
b. Z-에 대한 X-의 S다는 ~

(33') X-에 의한 Y-와 Z-의 ~

즉, 이 항목은 최대논항이 실현된 모든 구조를 구체적으로 나열한다. 일부 논항만을 갖는 명사구도 논항 삭제 규칙에 의해 자동으로 생성될 수 있다. 그런데, 논항들이 부분적으로 실현된 일부의 명사구는 논항 삭제 규칙으로 설명되기 어렵다. 예를 들어, '다른 논항과 동일한 조사와 결합하고 동시에 실현되지 못하는 (30')의 명사구는 최대논항구조로부터 생성되지 않는다. 따라서 최대논항구조항목은 이와 같은 구조를 추가적으로 제시한다.

(34) a. X-의 S는 ~  
b. Y-의 ~

한편, 단문과 명사구를 구성하는 논항들은 두 항목에 걸쳐 일관되게 표기되는 것을 원칙으로 한다. 이를 통해 단문과 명사구는 규칙적으로 대응되며, 두 구조 사이의 자동생성이 가능하게 된다.

마지막으로 논항선택제약항목은 술어명사 논항으로 나타날 수 있는 명사들의 의미적인 속성을 제시한다. 이 항목은 원칙적으로 의미부류를 통해 제시하고 추가 정보로서 구체적인 명사 형태를 나열한다. 의미부류를 제시하기 어려운 경우, 구체적인 명사들만을 나열하는 것이 가능하다.

(35) 사과 : X=인간(친구);Y=인간(선배);Z=행위(거짓말)사건(잘못)

이 항목은 명사 논항에 대한 정보만을 표상하며, 보문 논항에 대한 제약은 기술하지 않는다. 보문 논항의 제약은

보문의 어미를 통해 제시될 수 있다. 이에 보문에 나타나는 어미의 목록과 표기법을 마련하여, 직접 기능동사항목과 최대논항구조항목에 표상한다.

### 5. 세종명사사전의 기술표본 예시

```

<superEntry>
● 전화17)
<entryn=1>
  <tolevel> @form=[전화] @pos=[nn] @see=[]
  <morph_a> @var[xs={};xd={};xx=] @abb=[] @lng=[]
  @str=[] @org=[si(電話)]</morph_a>
  <morph_b> @hom[E] @der=[(n):(v):(a):(av)] @comp=[(n)무선|유선|휴대|공중|시외|시내|국제|직통~::~기|료|선|비|국|망|박스|번호|벨|세|질|통|폭력|회선;(v)~하다|받다:(a):(av)]
  @metc=[(n):(v):(a):(av)] @img[xs={};xd={};xx={};xp={}]
  </morph_b>
  <froz> @idnp=[] @idna=[] @idnv=[] @idda=[] @prv=[]
  @idetc=[] </froz> </tolevel>
  <sense n=1> @eg=[그는 사무실에 ~를 여러 대를 설치했다]
  @trans=[telephone] @domain=[] @reg=[] @con=[] @curs=[C]
  @sem=[제조물]
  <lr> @syn=[~기통] @ant=[] @hyper=[] @hypo=[]
  @coord=[팩스|텔레비전] @holo=[] @mero=[수화기|다이얼|벨|선] @rel=[응답기|모뎀] </lr>
  <synt_a> @cl=[uni=대|개:grp=:div=:qnt=] @prt=[]
  @av=[] @ds=[] </synt_a>
  <synt_b> @comb_aj=[] @magn=[] @comb_v=[~를 놓다|설치하다|부수다|끊다|바꾸다|연결하다::~가|고장나다|잘|들리다]
  @comb_n=[] @supv=[: cor_v={}] @max_n=[] @sel_res=[]
  </synt_b>
  <synt_c> @pre_d=[] @pre_n=[] @pre_s=[] </synt_c>
</sense>
  <sense n=2> @eg=[그는 전화국에 새 ~를 신청했다]@trans=[telephone line] @domain=[] @reg=[] @con=[]
  @curs=[C] @sem=[시설물]
  <lr> @syn=[] @ant=[] @hyper=[] @hypo=[] @coord=[]
  @holo=[] @mero=[] @rel=[전화번호|전화국] </lr>
  <synt_a> @cl=[uni=개:grp=:div=:qnt=] @prt=[] @av=[]
  @ds=[] </synt_a>
  <synt_b> @comb_aj=[] @magn=[] @comb_v=[~를 신청하다|놓다|설치하다|끊다] @comb_n=[] @supv=[: cor_v= {}]
  @max_n=[] @sel_res=[] </synt_b>
  <synt_c> @pre_d=[] @pre_n=[] @pre_s=[] </synt_c>
</sense>
  <sense n=3> @eg=[그는 시골에 계신 부모님께 ~를 드렸다]
  @trans=[phone call] @domain=[] @reg=[] @con=[]
  @curs=[C|S] @sem=[행위]
  <lr> @syn=[통화] @ant=[] @hyper=[] @hypo=[]
  @coord=[편지] @holo=[] @mero=[] @rel=[] </lr>
  <synt_a> @cl=[uni=통|통화|번|차례:grp=:div=:qnt=]
  @prt=[] @av=[] @ds=[] </synt_a>
  <synt_b> @comb_aj=[~가|길다|짧다|단단하다]

```

@magn=[ ] @comb\_v=[~를 걷다|굽다|연결하다|놓다:~가 오다:~로 연락하다|알리다] @comb\_n=[안부|장난 ~::~~ 연락]  
 @supv=[X-가 Y-에게 ~를 하다:X-가 Y-와 ~를 하다:X-와 Y-가 ~를 하다:Y-가 X-에게(서)로부터 ~를 받다: cor\_v={~하다|받다}] @max\_n=[X-의 ~:Y-의 ~:X-와 Y (간|사이)의 ~]  
 @sel\_res=[X=인간:Y=인간] </synt\_b>  
 <synt\_c> @pre\_d=[ ] @pre\_n=[ ] @pre\_s=[S는 ~] </synt\_c> </sense> </entry> </superEntry>

## 6. 결론

본고는 세종사전을 전체적으로 개괄하고, 하위범주 전자사전인 명사사전의 구성과 특성을 살펴보았다. 세종사전은 다양한 기능의 한국어 자동 처리에 활용될 것을 염두에 두고 있다. 명사사전은 이러한 목표에 적합하도록 사전구조와 정보항목들을 구성한다. 명사사전은 전산적 효율성을 고려하여 명사 어휘의 통사정보를 정밀하게 포함한다. 특히, 명사사전은 어휘계약적이고 불규칙적인 결합에 관한 정보들을 중점적으로 기술한다. 명사사전은 정보들을 구체적으로 제시하고 있어 그 활용가능성이 높을 것으로 전망된다. 명사사전은 별도의 정보항목을 통해 술어명사가 구성하는 기능동사구문과 명사구에 대한 정보를 제시한다. 이 항목들은 술어명사를 통해 실현될 수 있는 단문과 명사구의 형태와 구조를 명시적으로 표상한다. 따라서 이 정보들은 한국어 문장 및 명사구의 자동 처리, 특히 자동 생성에 즉각적으로 활용될 수 있을 것이다.

## 미주

1) 세종사전은 최종적으로 범주별 하위사전들이 통합된 통합전자사전의 성격을 갖는다. 개별 하위사전들은 고유한 정보항목구조와 사전구축지침을 갖는다. 사전구축지침은 각 정보항목에 기술되는 정보들의 유형과 그 표상 방법이 상세하게 제시되어 있다(세종전자사전 2000년도 결과보고서의 하위사전별 사전구축지침 참조). 1997-2000 사이의 3년간 1단계의 연구가 끝난 현 시점에서는 하위사전의 정보항목구조와 사전구축지침을 보완하는 한편, 통합전자사전의 정보항목구조를 정비하고 이에 대한 사전구축지침을 마련하고 있다.

2) 체언사전의 최종규모는 단일어명사20,000 항목, 복합어명사30,000 항목(80,000항목), 의존명사, 대명사, 수사, 분류사 1,200 항목이다.

3) 이러한 구분은 사전 구축의 효율성을 위한 것이며, 최종적으로는 두 사전은 통합되어 세종명사사전을 구성하게 된다.

4) 세종명사사전의 하위구획과 정보항목의 구체적인 예는 4절의 표본을 참조할 것.

5) 형용사 이외에도 일부 동사와 명사가 명사와 결합하여 이러한 유형의 의미를 표현한다.

i) (굳은+불굴의) 의지/무거운 침묵/침예한 대립성난 파도  
 (일루+일말)-의 희망

세종명사사전은 이러한 동사와 명사를 형용사와 같이 형용사언어항목에 제시하며, 이들을 명사언어와 동사자어결합항목에 중복하여 표상한다.

6) 형용사언어항목에 기술되는 정보들은 언어사전을 통해 별도의 사전표제어로 제시된다.

7) <>는 의미부류를 나타낸다.

8) J. Apresjan 1972와 Cruse 2000을 참고.

9) 표준국어대사전과 연세한국어사전 참조.

10) 영어에서 규칙적 다의성으로 분석되는 동물명과 그 고기의 의미변이는 한국어에는 부분적으로 대응된다.

i) 인도인들은 (돼지+돼지고기)-를 먹는다

ii) (\*돼지+돼지고기) 한 근

이러한 점을 고려할 때, 영어와 한국어의 이개어사전은 이 명사들을 다의어로 분할하는 것이 적절하다.

11) 이 예문의 끝나다와 무너지다는 예문 (7)과 (9)의 의미에만 국한된다. 따라서 접부에서는 '정권'의 의미가 배제된다.

12) 이 문장이 수용될 수 있다면, 그것은 '그가 나의 반박하는 말의 내용이 아니라, 내가 반박하는 행위(의 동기)를 이해했다'는 의미로서일 것이다.

13) 이 정보는 일종의 언어정보로서, 결합하는 어휘요소의 범주에 따라 명사, 형용사, 동사항목에 중복하여 기술된다.

14) 결합가능구조에 대한 자세한 설명은 세종전자사전 2000년도 결과보고서의 의존명사사전의 사전구축지침을 참조.

15) 수식어문장계약항목의 형태 및 표상 방법은 기능동사항목과 최대방향명사구항목에 기술되는 보문 정보들과 통합되고 있다. 자세한 설명은 세종전자사전 2000년도 결과보고서의 명사사전의 사전구축지침을 참조.

16) 자세한 기능동사 목록은 2000년도 결과보고서의 단일어명사사전의 사전구축지침을 참조.

17) 이 명사는 올해 상세전자사전에서 작업 중에 있는 예이며, 실제의 기술에서는 4개의 명사 동형어로 구분되고 있다.

## 참고문헌

- [1]. Apresjan, J. D. 1972. Regular polysemy, *Linguistics* 124, pp 5-39.
- [2]. Cruse, D. A. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press.
- [3]. Hong, C.-S. & Park, M.-G. 2001. Developing a Large Scale Computational Lexical Database of contemporary Korean: SELK, *ICCPOL 2001*, 19th International Conference on Computer Processing of Oriental Languages.
- [4]. Schneider, J.-G. 1987. *Les predicats nominaux en français: les phrases simples à verbe support*, Librairie Droz.
- [5]. 홍재성, 1999. 기능동사구문 연구의 한 시각: 어휘적 접근, 인문논총 41집, 서울대학교 인문학연구소.
- [6]. 홍재성의, 2000. 2000년도 세종계획 전자사전개발분과 결과보고서, 문화관광부.
- [7]. 국립국어연구원편, 2000. 표준국어대사전, 두산동아.
- [8]. 연세대학교 언어정보개발연구원편, 2001. 연세한국어사전, 두산동아.