

# 구문분석 말뭉치를 이용한 문법 관계의 선호 체언 어휘와 의미 유형 연구

홍정하<sup>+</sup>

고려대학교 민족문화연구원

kleist@korea.ac.kr

## A Treebank-Based Approach to Preferred Nominal Words in Grammatical Relations and their Semantic Types

Jungha Hong<sup>+</sup>

Institute of Korean Culture, Korea University

### 요약

이 논문은 각 문법 관계(grammatical relation)에서 선호되는 체언 어휘를 파악하고, 이 어휘들의 의미적 유형 및 그 위계를 파악하는 것이 목적이다. 이를 위해 80만 어절의 21세기 세종계획 구문분석 말뭉치에서 그 분포를 추출하고, 통계적 검증을 통해 각 문법 관계에서 선호되는 체언 어휘를 선별한다. 이 연구에서 관찰하는 문법 관계는 주어, 목적어, 용언수식어로 하며, 이들 문법 관계에서 선호되는 어휘 추출 대상 품사는 대명사, 고유명사, 일반명사로 한다.

한정성의 강도에 따라 주어 분포 경향이 나타나며, 이에 따라 대명사 > 고유명사 > 일반명사 순으로 주어 분포 경향이 나타난다. 그러나 일반적 예측과 다르게 한정성의 강도가 더 강한 것으로 알려진 대명사가 고유명사보다 목적어와 용언수식어에서 분포 경향이 더 강하여, 일반명사 > 대명사 > 고유명사의 순으로 분포 경향이 나타난다. 대명사, 고유명사, 일반명사는 공통적으로 주어에서는 사람 지시어, 목적어에서는 사물과 장소 지시어, 그리고 용언수식어에서는 시공간 표현이 선호되어 분포한다. 특히 대명사는 각 문법기능에서 인칭대명사의 경우 인칭에 따라, 그리고 지시대명사의 경우 원근칭에 따라 선호도의 차이를 보인다. 이러한 체언 어휘의 의미적 분포 특성은 문법 관계에 통사적 기능 외에도 의미적 경향이 반영된 것으로 고려될 수 있다.

### 1. 서론

[1]은 의미역과 관련한 인칭과 유생성(animacy)에 대한 구문적 유효성 위계를 다음과 같이 제시하였다.

(1) 가. 1/2인칭 대명사 > 3인칭 대명사 > 고유명사 >

사람 명사 > 유정성 명사 > 무정성 명사

나. 행위자역 > 대상역

(1-가)에서 상위의 위계일수록 (1-나)의 상위 위계인 행위자역으로 실현되는 경향이, (1-가)에서 하위의 위계일수록 (1-나)의 하위 위계인 대상역으로 실현되는 경향이 강하게 나타난다는 것이다.

이와 같은 위계를 토대로 문법 관계와 관련해서 [2]에서는 인칭 및 의미역의 위계를(2-3), [3]에서는 유생성 및 한정성의 위계를(4-5) 제시하였다.

(2) 가. 1/2인칭 주어 > 3인칭 주어

나. 3인칭 비주어 > 1/2인칭 비주어

(3) 가. 행위자역 주어 > 대상역 주어

나. 대상역 비주어 > 행위자역 비주어

(4) 가. 사람 주어 > 유정성 주어 > 무정성 주어

나. 무정성 목적어 > 유정성 목적어 > 사람 목적어  
(5) 가. 대명사 주어 > 고유명사 주어 > 한정성 주어 >

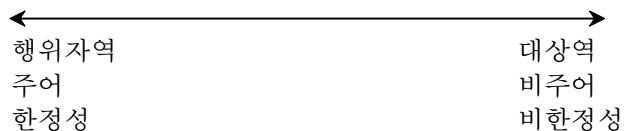
비한정성 주어

나. 비한정성 목적어 > 한정성 목적어 > 고유명사 목적어 > 대명사 목적어

(2) - (5)의 인칭, 의미역, 유생성 및 한정성 위계에서 모두 주어와 비주어 또는 주어와 목적어는 서로 상반된 위계를 보이고 있다.

(1) - (5)의 모든 위계를 정리하면, (1-가)의 위계 의미역, 문법 관계, 한정성의 척도를 (6)과 같이 적용한 것과 같다. (1-가)의 구문적 유효성 위계의 상위에 위치한 체언일수록 행위자역으로, 주어로, 그리고 한정성이 높은 논항으로 분포할 가능성이 높다.

(6) 1/2인칭 대명사 > ... > 무정성 명사



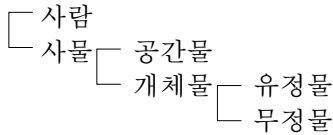
결국 (6)의 척도에서 주어는 행위자역으로 그리고 한정성이 있는 전형적인 문법관계로 분포할 가능성이 높다.

한편 (1-가)에 제시된 체언 분류는 지시성에 기반하

<sup>+</sup> HK 연구원

므로 실체명사에 국한되며, 이는 다시 대명사와 고유명사의 의미부류로 확대된다. (7)은 실체명사에 대한 존재론에 기반한 의미 분류이며, [4]에서는 각각의 부류가 지시성에 따라 대명사 및 고유명사와 대응될 수 있음을 제시하였다. 예를 들어 ‘사람’ 부류에는 인칭대명사와 사람을 지시하는 고유명사가, ‘공간물’ 부류에는 처소 대명사와 이에 해당되는 고유명사가 대응된다.

(7) 실체명사의 의미 분류



이 논문에서는 각 문법 관계에서 선호되는 체언 어휘를 파악하고, 이 어휘들의 의미적 유형 및 그 위계를 파악하는 것을 목적으로 한다. 이러한 접근은 (7)의 실체명사의 의미부류에 따라 구문적 유효성 위계의 적합성을 검증하는 연구가 될 것이다. 이를 위해 80만 어절의 21세기 세종계획 구문분석 말뭉치에서 그 분포를 추출하고, 통계적 검증을 통해 각 문법 관계에서 선호되는 체언 어휘를 선별한다. 이 연구에서 관찰하는 문법 관계는 주어, 목적어, 용언수식어로 하며, 이들 문법 관계에서 선호되는 어휘 추출 대상 품사는 대명사, 고유명사, 일반명사로 한다.

2. 연구 대상 자료의 처리와 특성

이 연구에서는 효율적인 구문적 분포 연구를 위해 대상 자료인 21세기 세종계획 구문분석 말뭉치(이하 SKT)에<sup>1)</sup> 부가적 기능 표지 부착을 하고 빈도 산출 대상 범위를 제한하였다. SKT에서는 구문적 노드에 구문표지(문장, 체언구, 용언구 등)와 기능표지(주어, 목적어 등)를 병기하였으나, ‘체언 + 체언’ 수식 구성이나 동격 체언구 구성의 경우에 해당 노드에 기능표지를 부착하지 않았다. 그러나 이 논문에서는 효율적인 통계적 처리를 위해 기능표지가 부착되지 않은 노드에 명시적 문법기능이 표지되지 않았음을 나타내는 무기능 표지를 새롭게 부착하여 처리하였다.

한편 SKT에서는 인용부호와 같이 쌍을 이루는 부호 내부에 두 개 이상의 문장이 포함되었을 경우 이 문장들을 독립적 문장으로 분리하고, 그 문장 자리에 “Q=2”와 같이 인용절에서 분리된 문장 수가 표시되었다. 그러나 본 연구의 빈도 산출 대상에서 분리된 인용절은 개별 문장으로 처리하되 문장 분리 표시는 빈도 산출 대상에서 제외하였다.

이와 같은 처리를 통해 산출된 SKT는 총 827,881 어절, 총 73,824 문장의 규모로 구축되어 있다. 문장당 평균 어절수는 11.214 어절, 문장 길이는 최소 1어절에서부터 최대 124어절로 구성되어 있고 문장 길이별 분포는 [그림-1]과 같다.

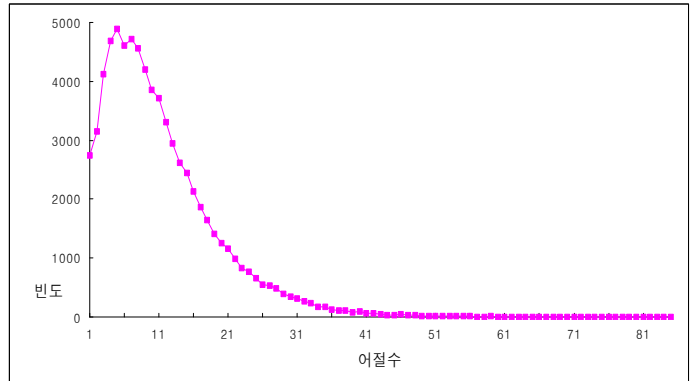


그림 1. 문장 길이 분포

표 1. 텍스트 유형별 분포

매체	장르	비율	매체	장르	비율
신문	일반	1.77 %	책	상상	25.38 %
	사회	2.40 %		체험	3.75 %
	문화	6.32 %		인문	9.46 %
	생활과학	5.60 %		사회	10.60 %
	사설	5.62 %		과학	3.83 %
	인문	6.33 %		생활	3.55 %
잡지	문화	3.37 %	전자출판	자연	0.77 %
	교육	8.75 %		생활	2.51 %

이 뿐만 아니라 SKT는 다양한 매체 및 장르에서 생산된 텍스트로 구성되어 있다. [표-1]은 SKT의 텍스트 유형별 분포로 생산 매체 및 장르에 따른 분포를 나타내고 있다. 이상과 같이 이 논문의 연구 대상 자료인 SKT는 문장 길이와 텍스트 유형 측면에서 다양성과 균형성이 확보되었다고 보여진다. 이는 문장 길이에 따른 다양한 구문적 특성이 반영되어 있을 뿐만 아니라, 다양한 텍스트 유형에 따른 일반적 구문 특성을 포착하기에 적절하다 하겠다.

3. 체언의 품사별 문법 관계적 분포 특성

체언의 문법 관계적 분포 특성을 파악하기 위해 (8)과 같이 하나의 체언으로 구성된 또는 하나의 체언에 문법 형태소가 결합되어 구성된 체언구만을 대상으로 하여 문법 관계별 분포를 추출하였다. 물론 SKT는 두 개 이상의 체언 또는 서로 다른 품사로 구성된 다양한 유형의 체언구도 상당수 포함하고 있지만, 이 논문은 개별 어휘의 문법 관계적 분포를 파악하기 위해 이러한 구성의 분포는 연구에서 제외하였다. 이는 문법 관계의 분포에서

1) 구문분석 구축 지침에 대해서는 [5] 참조.

다른 어휘에 의한 영향을 배제하기 위함이다.

(8) 관찰 대상 체언 어절의 구조

가. 단일 체언

나. 단일 체언 + 문법 형태소

체언 중 대명사, 고유명사, 일반명사의 문법 관계별 분포는 [표-2]와 같다. 동일한 체언이라 하더라도 [표-2]에서 제시된 바와 같이 품사에 따라 문법 관계별 분포가 상이하며, 대명사에서는 주어, 고유명사에서는 주어와 무기능표지가 두드러진 분포를 보이며, 일반명사는 주어, 목적어, 용언수식어, 무기능표지에서 비교적 고른 분포를 나타내고 있다.

표 2. 체언의 품사별 문법 관계 분포

기능표지	대명사	고유명사	일반명사
주어	44.67%	21.37%	19.23%
목적어	10.08%	4.79%	19.65%
용언수식어	16.72%	10.94%	23.79%
체언수식어	16.20%	14.72%	9.48%
무기능표지	11.20%	39.66%	21.27%
보어	0.45%	0.31%	1.42%
삽입어	-	0.08%	0.01%
접속어	0.39%	7.24%	4.91%
독립어	0.29%	0.90%	0.24%
합계	100.00%	100.00%	100.00%

[표-2]에서 이 논문의 관찰 문법 관계인 주어, 목적어, 용언수식어에 대해 살펴보면 문법 관계별 품사의 분포 위계는 다음과 같다.

(9) 가. 주어: 대명사 > 고유명사 > 일반명사

나. 목적어: 일반명사 > 대명사 > 고유명사

다. 용언수식어: 일반명사 > 대명사 > 고유명사

(9)의 위계는 주어에 대한 (5-가)의 위계 예측과 일치하지만, 목적어 및 용언수식어에 대한 (5-나)의 위계 예측과는 차이가 있다.

이 논문에서는 문법 관계에서 선호되는 어휘를 선별하기 위해  $\chi^2$  검증을 사용하여 문법 관계적 분포 특성이 동일 품사의 일반적 어휘들과 차이가 있는 어휘를 분류하였다.  $\chi^2 > 26.125(df = 8, \alpha = 0.001)$ 인 어휘가, 단, 대명사의 경우  $\chi^2 > 24.322(df = 7, \alpha = 0.001)$ 인 어휘가 여기에 해당되며,  $\chi^2$  값이 이 기준치 보다 클 경우 동일 품사의 일반적 어휘와는 상이한 분포적 경향을 나타내는 어휘로 볼 수 있다.

이렇게 선별된 어휘를 대상으로 각 기능 표지의 분포 양상을 관찰하기 위해 (10)의 공식을 이용하여 개별 어휘의 기능 표지별 관찰치와 기대치 차이에 대한 값을 산출하여 그 상관도를 측정하였다.

$$(10) \text{ 기능 표지별 상관도} = (\text{관찰치} - \text{기대치})^2 / \text{기대치}$$

(10)의 기능 표지별 상관도가 높을수록 특정 어휘가 해당 기능 표지에서 기대했던 분포보다 더 높거나 낮게 분포함을 의미한다. 이 연구에서는 관찰치가 기대치보다 적게 분포한 경우를 제외한 관측치가 기대치보다 높은 어휘를 통해 그 분포적 특성을 파악한다.

또한 구문분석 말뭉치의 분석 오류 가능성과 통계적 왜곡 가능성을 고려하여 이 상관도 측정에서 각 어휘의 기능 표지별 분포 빈도가 10회 이상인 어휘만을 고려하였다. 이 절에서는 체언의 문법 관계 표지별 분포에서 공통적으로 두루 분포하는 주요 문법 관계 표지인 주어, 목적어, 용언수식어에 대해서 그 분포적 특성을 제시한다. 또한 각 문법 관계 표지별 분포 목록의 수가 많은 경우 상위 30개 어휘만을 제시한다.

4. 대명사

신정보가 제시될 때 타동사 구문에서 주어 자리는 기피한다는 구어에 대한 [6]의 관찰과 문어에 대한 [7]의 관찰은 역으로 구정보를 전달하는 대명사의 전형적인 주어 실현 경향을 의미한다. 대명사는 [표-2]와 같이 주어에서 두드러진 분포를 보이는데, 그 이유는 대명사는 일반적으로 문맥 또는 상황에 의해서 환기되는 구정보인 지시체를 가리키기 때문이다.

표 3. 대명사 어휘의 문법 관계 표지별 분포

기능표지	대명사
주어	나/내(217.26), 이(123.96), 그것(90.78), 누구(46.72), 너/네(37.79), 아무(36.33), 그(23.32), 이거(10.22), 당신(7.29), 저것(3.14), 제/저(1.98), 그대(1.20), 자네(0.99), 그거(0.84), 그녀(0.24), 니(0.16)
목적어	그것(730.14), 무엇(378.99), 뭐(233.22), 이것(60.57), 이(40.65), 그거(25.18), 이거(9.12), 그곳(2.70), 너희(0.72), 이곳(0.59), 그녀(0.36)
용언수식어	거기(1764.91), 여기(1527.88), 어디(1122.99), 이(816.25), 그곳(141.04), 무엇(113.02), 언제(99.97), 이곳(92.86), 저기(49.73), 그쪽(28.25), 여러분(1.59)

\* ( ) 안의 수치는 개별 어휘의 기능표지별 상관도

[표-3]은 대명사 어휘 중 문법 관계 표지별 분포 경향이 두드러진 어휘와 그 상관도를 나타낸 것이다. 주어의 경우, 인칭대명사, 지시대명사, 의문대명사, 부정대명사의 분포를 보이지만 몇 가지 특징으로 정리할 수 있다. 첫째, 의문대명사 ‘누구’, 부정대명사 ‘아무’를 포함하여 사람을 지시하는 대명사의 분포 종류가 많다. 이는 (4-가)의 예측과 동일하다. 둘째, 지시대명사 중에는 장소를 지시하는 ‘거기, 여기, 그곳’과 같은 처소 대명사는 주어에서 두드러진 분포를 나타내지 않는다. 셋째, (2-가)의 예측과 마찬가지로 인칭대명사의 경우 1인칭에서 3인칭 순(나/내 > 너/네 > 그/그녀)으로, 지시대명사의 경우 근칭에서 원칭 순(이것/이거 > 그것/그거)으로 주

어 분포 상관도의 차이를 보인다.<sup>2)</sup> 이러한 주어 분포 상관도의 차이는 영어에서 두 단어를 나열할 때, “here and there”, “now and then” 처럼 화자를 중심으로 어순을 선택하는 원리인 ‘나 먼저 원리’(Me first principle)[9]와 유사하게 나타난다.

반면 (2-5)에서 예측된 것처럼 목적어는 주어의 분포와 대조를 보인다. 첫째, (4-나)에서 예측된 바와 같이 의문대명사 ‘무엇, 뭐’를 포함하여 사물과 장소를 지시하는 대명사의 사용이 인칭을 나타내는 대명사보다 두드러진다. 둘째, 신정보의 출현을 대신하는 의문대명사의 목적어 분포 상관도가 높다. 이는 [6]와 [7]의 관찰에 대한 또 하나의 증거이다. 셋째, 지시대명사에서 사물을 지시하는 대명사이든 장소를 지시하는 처소 대명사이든 상관없이 원칭이 근칭에 비해 목적어 분포 상관도가 높다.(그것/그거/그곳 > 이것/이거/이곳) 이것은 주어와 반대의 경우이며, 목적어에서는 화자와 거리가 먼 것에 대한 기술 경향이 강하다고 할 수 있다.

용언수식어의 경우, 장소와 관련된 어휘가 많이 분포하며, 목적어와 마찬가지로 원칭이 근칭에 비해 분포 상관도가 높다.(거기/그곳 > 여기/이곳) 또한 용언수식어로 사용된 의문대명사의 경우 장소(어디) > 사물(무엇) > 시간(언제)의 순서로 분포 상관도가 높다.

한편, 이러한 문법 관계별 분포적 차이와 어휘의 빈도순 목록과 비교해 보면 그 상관성이 커 보이지는 않는다. [표-4]은 대명사 상위 20개 어휘의 빈도순 목록과 그 빈도를 제시한 것이다.

표 4. 대명사 어휘의 빈도순 목록

순위	대명사	빈도	순위	대명사	빈도
1	그	5910	11	누구	728
2	나	4279	12	이것	644
3	우리	3819	13	뭐	620
4	그것	2118	14	어디	611
5	이	2044	15	당신	559
6	자기	1008	16	여기	534
7	그녀	989	17	거기	501
8	내	980	18	저	467
9	너	913	19	너희	256
10	무엇	766	20	네	183

[표-4]에서 인칭 대명사 분포가 많다는 것을 제외하면, [표-3]의 대명사 어휘의 문법 관계 표지별 분포와 상관성이 있어 보이지는 않는다. 물론 대명사의 문법 관계 분포에서 비교적 주어에 집중되어 분포된다는 점과 대명사 상위 20개 어휘의 빈도순 목록에서 인칭 대명사의 분포가 두드러진다는 점에서 [표-2]의 주어 분포에 인칭 대명사의 분포적 경향이 나타난다고 가정할 수도 있다. 그러나 인칭 대명사 분포가 목적어와 용언수식어에서 비

2) ‘이, 그’는 상황지시 및 문맥지시에 모두 사용되나, 지시대명사 ‘저’는 상황지시에만 사용되어 분포적 제약이 있다.[8] 그래서 ‘저것/저기’를 ‘이것/여기’, ‘그것/거기’와 분포적 강도를 직접적으로 비교하는 것은 어렵다.

교적 드문 이유를 설명할 수는 없다.

표 5. 대명사: 문법 관계와 빈도의 상관 분석 결과

기능 표지	주어	목적어	용언수식어
$\chi^2 > 24.322$ 의 대명사	0.17	-0.29	-0.414
	3	5	

[표-5]는  $\chi^2 > 24.322(df = 7, \alpha = 0.001)$ 인 대명사 전체에 대한 문법 기능별 분포 상관도와 각 어휘의 빈도 사이의 상관 분석 결과이다.  $\chi^2$  값이 유의미한 분포를 보이는 대명사는 주어, 목적어, 용언수식어에서 빈도와 낮은 상관성을 보인다. 이는 문법 관계에 따라 집중되어 분포하는 의미적 유형에 대명사 어휘의 빈도가 큰 영향을 미치지 못함을 의미한다.

### 5. 고유명사

고유명사는 환기된 지시체뿐만 아니라, 추론 가능한 지시체 및 새로운 지시체를 가리킬 수 있다는 점에서 대명사에 비해 광범위하게 사용된다. 그러나 고유명사는 일반적으로 고정된 지시체를 가리키므로 그 지시체가 명확하기 때문에 혼동의 가능성이 적다[10]. 그래서 [표-2]과 같이 대명사에 비해 고유명사는 무기능표지 및 접속어에 분산되어 분포되어 주어 분포의 강도가 떨어진 다. 그러나 신정보가 선호하는 목적어의 분포비율은 대명사에 비해서 현저하게 떨어진다.

[표-6]은 고유명사 어휘 중 문법 관계 표지별 분포 경향이 두드러진 어휘와 그 상관도를 나타낸 것이다. 주어에 두드러진 분포 경향을 나타내는 상위 30개의 고유명사는 모두 인명을 나타낸다. 이는 대명사의 주어 분포에서 사람을 지시하는 대명사가 주어로 많이 분포함과 일치하며, (4-가)의 예측과도 일치한다.

표 6. 고유명사 어휘의 문법 관계 표지별 분포

기능표지	고유명사
주어	경지(499.74), 현우(433.39), 페르츠(357.42), 조국(332.97), 혜자(301.38), 규혁(209.70), 영애(200.07), 태봉이(176.91), 두환(148.72), 승주(126.91), 수창이(102.91), 화숙이(91.39), 경십이(88.52), 세호(80.86), 사이드(80.85), 민이(71.80), 영재(62.53), 태호(60.35), 새롭이(53.62), 지은(52.09), 삼월이(45.57), 콘라드(45.41), 미셀(43.74), 페로(43.08), 일리히(40.38), 라라(34.36), 레닌(32.85), 황(30.38), 갈릴레오(30.02)
목적어	만다린(445.61), 영어(229.31), 햄릿(79.79), 페르츠(50.44), 세호(39.64), 조동팔(38.22), 예수(28.77), 민요섭(28.77), 영애(19.76), 집(17.10), 황(15.09), 소희(8.46), 북한(6.24), 인도(4.79), 현우(3.43), 승주(0.61), 경지(0.49), 두환(0.19), 규혁(0.04)
용언수식어	서울(227.56), 유럽(81.61), 한국(75.31), 부산(54.03), 미국(45.41), 빠리(39.85), 경주(38.00), 갈여울(37.77), 양동(35.49), 브라질(35.42), 연천(30.68), 인천(28.60), 파투산(28.01), 남한(27.22), 일본(22.53), 광주(17.43), 영어(15.35), 중국(8.78), 프랑스(8.78)

64), 에스파냐(6.81), 이탈리아(6.16), 이사벨라(5.79), 뉴튼(5.63), 영국(3.40), 로마(3.35), 인도(2.10), 김부식(1.70), 북한(1.28), 수창이(0.65)
---

\* ( ) 안의 수치는 개별 어휘의 기능표지별 상관도

고유명사의 목적어 분포는 ‘만다린, 영어’와 같은 언어명의 상관도가 높으며 인명, 지명의 순으로 상관도의 분포 차이를 보인다. 이는 대명사의 목적어 분포에서 사람과 장소가 주로 분포하는 것과 차이를 보인다. 그러나 고유명사의 다른 문법 관계 분포에 비해 상관도도 떨어지며, 분포하는 고유명사 어휘의 수 또한 현저한 차이가 난다. 이는 고유명사가 신정보의 전형적 위치인 목적어를 기피하는 경향을 나타낸다. 용언수식어는 고유명사와 마찬가지로 주로 지명을 나타내는 고유명사의 두드러진 분포를 보인다.

한편, 대명사의 상위 20개 어휘 빈도 목록에서 인칭대명사가 분포하는 비율이 높았음에 비해, 고유명사에서는 지명과, 인명이 분포의 주를 이룬다.<sup>3)</sup> [표-7]은 고유명사 상위 20개 어휘의 빈도순 목록과 그 빈도를 제시한 것이다.

표 7. 고유명사 어휘의 빈도순 목록

순위	고유명사	빈도	순위	고유명사	빈도
1	한국	1125	11	아하스	453
2	미국	823	12	페르츠	448
3	서울	776	13	프랑스	405
4	영국	694	14	네덜란드	394
5	경지	536	15	중국	372
6	현우	508	16	김	351
7	유럽	499	17	남	339
8	일본	499	18	박	339
9	인도	495	19	조국	322
10	세호	491	20	승주	321

이 고유명사 어휘의 빈도적 특성이 [표-6]의 문법 관계 표지별 분포에서 인명과 지명에 주로 국한된 이유를 반영하고 있다고 볼 수는 있겠지만, 각 문법 관계별 의미 유형 분포와는 직접적 연관성이 있어 보이지는 않는다. [표-8]은  $\chi^2 > 26.125(df = 7, \alpha = 0.001)$ 인 고유명사 전체에 대한 문법 기능별 분포 상관도와 각 어휘의 빈도 사이의 상관 분석 결과이다.  $\chi^2$  값이 유의미한 분포를 보이는 고유명사는 주어, 목적어에서 빈도와 지극히 낮은 상관성을 보인다.

표 8. 고유명사: 문법 관계와 빈도의 상관 분석 결과

기능 표지	주어	목적어	용언수식어
$\chi^2 > 26.125$ 의 고유명사	0.019	-0.029	0.342
[표-6] 어휘	0.939	0.269	0.701

3) SKT에서는 고유명사의 범위를 인명, 종족명, 지명, 건축물, 단체명, 창작물, 언어명으로 제한한다.

## 6. 일반명사

(5)의 한정성 위계에 있어서 일반명사는 주어 분포 경향에서는 최하위에, 목적어 분포 경향에서는 최상위에 위치한다.<sup>4)</sup> 일반명사는 [표-2]에서와 같이 주어, 목적어, 용언수식어 분포가 비교적 고르면서, 목적어와 용언수식어의 분포가 (5-나)의 예측과 마찬가지로 대명사와 고유명사에 비해 현저히 높게 나타난다. 이는 신정보가 제시될 때 타동사 구문에서 주어 자리는 기피한다는 [6]와 [7]의 연구 결과에서 일반명사의 신정보 전달 경향을 나타내는 것이기도 하다.

[표-9]은 일반명사 어휘 중 문법 관계 표지별 분포 경향이 두드러지게 나타나는 어휘와 그 상관도이다. [표-9]에 제시된 일반명사의 주어 분포 경향을 나타내는 어휘는 상관도가 높은 상위 30개 어휘이다. 여기에서도 대명사, 고유명사와 마찬가지로 주어는 주로 사람을 지시하는 일반명사가 분포한다.

표 9. 일반명사 어휘의 문법 관계 표지별 분포

기능표지	일반명사
주어	사람(2026.77), 일(424.33), 아이(419.76), 모두(404.55), 선생(342.72), 필요(302.64), 아버지(272.71), 생각(259.32), 어머니(255.24), 차이(225.63), 이유(189.94), 기분(181.52), 필자(180.02), 주인(152.95), 여자(151.27), 소리(133.39), 애(132.76), 남편(131.90), 가슴(126.20), 자신(124.80), 학생(117.62), 남자(115.85), 젊은이(110.30), 혼자(107.34), 친구(101.56), 아내(99.96), 두꺼비(98.83), 대부분(94.25), 키(93.37)
목적어	고개(758.81), 돈(538.67), 말(425.81), 영향(391.16), 애기(316.58), 예(312.58), 역할(287.94), 옷(227.86), 밥(207.02), 관심(198.34), 문(192.25), 몸(185.79), 물건(178.08), 술(173.07), 이야기(160.59), 담배(159.15), 손(158.18), 목숨(140.31), 모습(133.68), 공부(131.10), 머리(116.57), 팔(115.26), 눈(114.47), 정체(114.14), 소리(112.22), 어깨(102.65), 구슬(98.85), 깃(97.52), 흉내(95.00)
용언수식어	때(4270.66), 속(1687.51), 전(1159.23), 동안(1115.58), 앞(1082.29), 처음(1036.71), 뒤(1010.35), 후(941.81), 위(825.77), 지금(717.07), 안(669.57), 이제(609.06), 경우(598.42), 그때(559.22), 대신(494.46), 가운데(488.22), 동시(459.26), 사이(436.58), 이후(434.31), 다음(418.43), 한편(414.31), 결국(409.98), 최근(395.21), 아래(372.98), 열(361.60), 밖(352.77), 나중(334.98), 순간(282.48), 반면(251.00)

\* ( ) 안의 수치는 개별 어휘의 기능표지별 상관도

목적어에서는 대명사와 마찬가지로 사람을 지시하는 명사는 찾아 볼 수 없으며, 고개, 몸, 손, 머리, 팔, 눈, 어깨와 같은 신체어 및 ‘돈, 옷, 밥, 문, 물건, 술, 담배’와

4) 물론 일반명사의 한정성과 비한정성에 따라 차이가 있지만, 이 논문에서 관찰하는 분포는 이러한 맥락을 고려치 않으므로 일반명사의 한정성과 비한정성 구별에 관한 논의는 배제한다.

같은 사물을 지시하는 무정성의 실체 명사가 주로 분포한다.<sup>5)</sup> 이는 (4-나)의 예측대로 무정성 명사가 목적어 분포 경향을 나타낸다.

용언수식어에는 대명사, 고유명사와 마찬가지로 주로 시공간 표현이 주를 이루며, ‘한편, 결국, 반면’ 과 같은 부사성 명사는 이에 비해 상관도가 낮게 나타난다. 또한 비록 실체명사는 아니지만, 공간 관계 표현에서는 ‘전/앞 > 후/뒤, 위 > 아래, 속/안 > 밖’ 의 선호도 차이를 보이며, 시간 관계 표현에서는 ‘처음 > 다음/나중, 지금/이제 > 그때’ 의 선호도 차이를 보인다.

표 10. 일반명사 어휘의 빈도순 목록

순위	일반명사	빈도	순위	일반명사	빈도
1	사람	4343	11	문화	1320
2	말	4300	12	집	1190
3	때	3122	13	민족	1172
4	일	2343	14	눈	1146
5	생각	2173	15	시작	1127
6	세계	1588	16	아이	1127
7	사회	1526	17	영화	1057
8	자신	1474	18	나라	1049
9	문제	1401	19	인간	1031
10	속	1366	20	경우	1015

[표-10]은 일반명사 상위 20개 어휘의 빈도순 목록과 그 빈도를 제시한 것이다. 일반명사 상위 20개 어휘에는 대명사나 고유명사처럼 사람이나 장소에 관련된 의미 유형에 국한된 것이 아니라, 다양한 의미 유형으로 분포되어 있다. 그래서 대명사나 고유명사에 비해 일반명사의 문법 관계별 선호 의미 유형은 다양하지만, 대명사와 고유명사와 공통적인 의미 유형의 분포적 특성을 보인다. [표-11]은  $\chi^2 > 26.125(df = 7, \alpha = 0.001)$ 인 일반명사 전체에 대한 문법 기능별 분포 상관도와 각 어휘의 빈도 사이의 상관 분석 결과이다.  $\chi^2$  값이 유의미한 분포를 보이는 일반명사는 주어를 제외한 문법 관계에서 빈도와 낮은 상관성을 보인다.

표 11. 일반명사: 문법 관계와 빈도의 상관 분석 결과

기능 표지	주어	목적어	용언수식어
$\chi^2 > 26.125$ 의 일반명사	0.441	-0.540	0.254

## 7. 결론 및 남은 과제

이 논문에서는 각 문법 관계에서 선호되는 체언 어휘

5) 이 밖에도 주어에서는 ‘필요, 차이, 이유, 기분’ 과 같은 상태 명사가 주로 분포하며, 목적어에서는 ‘말, 얘기, 이야기’ 의 언어행위 명사, ‘영향, 역할, 공부’ 와 같은 사건 명사가 주로 분포를 보인다. 이에 대한 분류는 향후 과제로 남겨두기로 한다.

를 파악하고, 이 어휘들의 실체명사적 의미 유형을 파악하였다. 먼저 한정성이 높은 품사일수록 주어 분포 경향이 강했으나, 목적어와 용언수식어에서는 대명사가 고유명사에 비해 분포 경향이 약한 것으로 나타났다. 그리고 대명사, 고유명사, 일반명사는 공통적으로 주어에서는 사람 지시어, 목적어에서는 사물(특히 무정성 명사) 또는 장소 지시어, 그리고 용언수식어에서는 시공간 표현이 선호되어 분포한다.

특히 전형적인 구정보 및 한정성을 나타내는 대명사는 주어 분포에서 인칭대명사의 경우 1인칭에서 3인칭 순으로, 지시대명사의 경우 근칭에서 원칭 순으로 상관도의 차이를 보인다. 목적어에서는 신정보의 출현을 대신하는 의문대명사의 분포 상관도가 높고, 주어와 대조적으로 목적어와 용언수식어에서는 지시대명사에서 원칭이 근칭에 비해 분포 상관도가 높다. 또한 비록 실체명사는 아니지만, 일반명사의 공간 관계 표현에서는 ‘전/앞 > 후/뒤, 위 > 아래, 속/안 > 밖’ 의, 그리고 시간 관계 표현에서는 ‘처음 > 다음/나중, 지금/이제 > 그때’ 의 분포적 상관도 차이를 나타낸다.

이상의 분포적 경향은 문법 관계가 단순히 통사적 속성 뿐만 아니라, 인칭, 의미역, 유생성 및 한정성과 관련한 의미적 속성도 반영되어 있음을 나타낸다. 그러나 이 연구의 한계를 세 가지 점으로 정리할 수 있다. 첫째, 대명사를 제외한 고유명사와 일반명사의 분포에서는 상위 30개 어휘로 제한하여 그 의미적 유형을 분류하였다. 그러나 상위 30개 어휘는 각 문법 관계 분포적 상관도가 높은 어휘이긴 하지만 각 문법 관계에 선호되어 분포되는 어휘 의미 유형의 체계적인 분석 자료로 사용하기에는 한계가 있다고 할 수 있다. 둘째, 모든 품사에서 비교적 많은 분포를 차지하고 있는 무기능표지에 대한 분석을 이 논문에서는 제외하였으나, 그 범주에 속하는 분포도 문법 관계별 체언의 의미 유형 파악에 단서를 제공할 가능성이 있다. 마지막으로 일반명사의 의미 유형을 보다 세밀하게 관찰하기 위해서는 일반명사가 출현하는 구문 유형도 아울러 고려할 필요가 있다. 이러한 점들을 보완한다면, 문법관계에서 선호되는 체언의 의미적 유형에 관한 완성도 높은 분포 연구가 될 것이다.

## 참고문헌

[1] Silverstein, Michael, Hierarchy of Features and Ergativity, in R. M. W. Dixon(ed.), *Grammatical Categories in Australian Languages*, 112-171, Australian Institute of Aboriginal Studies, Canberra, 1976.  
 [2] Aissen, Judith, Markedness and Subject Choice in Optimality Theory, *Natural Language and Linguistic Theory* 17, 673-711, 1999.  
 [3] Aissen, Judith, Differential Object Marking: Iconicity vs. Economy, *Natural Language and Linguistic Theory* 21, 435-483, 2003.  
 [4] 최경봉, “명사의 의미 분류에 대하여”, *한국어학* 4권 가을호, 11-45, 1998.

- [5] 강범모, 김의수, “세종 구문분석 말뭉치를 위한 구문 분석 방법”, 코퍼스과 어휘 데이터베이스, 2004.
- [6] Du Bois, John, Discourse and grammar. Course material on discourse and grammar, LSA Summer Institute, 2001.
- [7] 정소우, 정명순, “담화유형별 선호 논항 구조”, 담화와 인지, 12-3, 121-136, 2005.
- [8] 김일웅, 우리말 대용어 연구, 부산대학교 박사학위 논문, 1982.
- [9] Cooper, W. E., & Ross, J. R.. World Order. In R. E. Grossman, L. J. San, & T. J.Vance(Eds.), *Papers from the parasession on functionalism*, 63-111, 1975.
- [10] 박철우, 한국어 정보구조에서의 화제와 초점, 서울대학교 박사학위 논문, 1998.