

조사 상당 구성의 유형 분류*

- 의미 기능을 중심으로 -

문 병 열 (한남대)

< 목 차 >

- | | |
|---------------------|--------------------|
| 1. 서론 | 4. 조사 상당 구성의 유형 분류 |
| 2. 기본적 논의 | 5. 결론 |
| 3. 조사 상당의 구성의 군집 분석 | |

국문초록

본고는 한국어의 조사 상당 구성들을 좀 더 세밀하게 분류하고 이를 통해 조사 상당 구성의 전체 체계를 조망하는 것을 목적으로 하였다. 조사 상당 구성은 다수의 항목들이 매우 유사한 의미·기능을 보이되 미묘한 의미 차이를 지니며 공존하고 있다. 문법적 기능을 기준으로 하는 거시적인 분류 체계를 사용하고 있으나, 개별 구성들의 미묘한 의미 차이에 기반한 미시적인 분류 체계 역시 세워야 할 것이다.

따라서, 본고는 개별 조사 상당 구성들의 선행 요소 결합 양상을 추출하여 이를 기초하여 계층적 병합 군집 분석을 시도하였다. 그 결과 이들은 [인과 연쇄] 관련 구성과 [목적-수단] 관련 구성으로 크게 분류되었고 [인과 연쇄] 관련 구성들은 다시 [시간], [원인], [경로], [추론] 등으로 분류되었으며 [초점대상]과 관련되었다. [목적-수단] 관련 구성들은 [목적], [수단], [도구] 등으로 분류되었고 [일반 대상]과 관련되었다.

1. 서론

본고는 한국어의 조사 상당 구성들을 좀 더 세밀하게 분류하고 이를 통해 조사 상당 구성의 전체 체계를 조망하는 것을 목적으로 한다. 문병열(2014: 140)에서는 ‘둘 이상의 단어가 서로 긴밀하게 결합하여 하나의 기능(또는 의미)을 담당하는 구성’을 복합 구성(complex construction)이라 명명한 바 있다. 조사 상당 구성은 복합 구성의 하위 부류이며 [조사+용언 활용형]의 고정된 결합 구성으로 그 전체가 조사의 기능을 담당하는 구성을 말한다(문병열 2015: 27).¹⁾²⁾

《표준국어대사전》에 등재된 항목을 기준으로 할 때, 한국어의 조사는 대략 160여 개 정도이다.³⁾ 이들은 그 기능에 따라 각각 격조사, 보조사, 접속조사 등으로 구분되고 격조사는 다시 주격조사, 목적격조사, 관형격조사, 부사격조사 등으로 구분된다. 이러한 분류는 개별 조사들의 문법적 기능을 기준으로 한 것인데 이는 160여 개의 조사들의 전체 체계를 세밀하게 기술하

* 본 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5B5A02023195).

- 1) 본고는 문병열(2015)의 후속 연구 성격이 짙다. 따라서 ‘조사 상당 구성’의 기초적 논의(정의, 구성의 목록 확정, 거시적 분류 체계 등)는 문병열(2015)를 따른다.
- 2) 본고의 조사 상당 구성은 임근석(2005, 2006)의 ‘문법적 언어’ 중 ‘조사적 언어’(임근석 2009), 서경숙(2004)의 ‘조사 상당 구성’과 같은 개념이며, 이희자(1995)의 ‘형태적 언어’의 하위 유형에 해당된다.
- 3) 《표준국어대사전》에 등재된 조사의 수는 362개이다. 이에는 [옛말], [방언], [북한어] 등이 포함되어 있으며 이형태로 볼 수 있는 것들이 중복 등재되어 있다(문병열 2015: 1).

기에 부족해 보인다. 특히 부사격조사의 경우, 다수의 항목들이 매우 유사한 의미·기능을 보이되 미묘한 의미 차이를 지니며 공존하는 것을 쉽게 확인할 수 있다. 문법적 기능을 기준으로 하는 거시적인 조사 분류 체계에 더하여 개별 조사들의 미묘한 의미 차이에 기반한 미시적인 조사 분류 체계를 세워야 할 것이다.

어떤 언어 형식의 의미·기능을 기술하기 위해서는 그 형식의 분포를 살펴야 한다. 특정 형식의 분포를 면밀히 살펴보아야만 그 형식의 (문법적) 기능, 중심 의미 등을 파악할 수 있고 특히 우리의 언어 직관으로 파악할 수 없는 미묘한 의미까지 알아낼 수 있다. 미묘한 의미(또는 의미 차이)는 맥락의 영향을 많이 받게 되는데 특정 형식의 분포를 면밀히 기술하는 것은 결국 그 형식의 사용 맥락을 세밀하게 살피는 것이기 때문이다. 160여 개에 달하는 한국어 조사들이 미묘한 의미 차이를 보이며 공존하는 현상을 기술하기 위해서는 개별 조사들의 분포를 세밀하게 살펴보고 이에 기반하여 다양한 방식의 미시적 분류를 시도해야 할 것이다.

부사격조사 ‘-로부터’, ‘-에서’ 등은 각각 [-를 부터], [-에 서] 등의 통사적 구성이 문법화 과정을 겪어 하나의 조사로 굳어진 것들이다. [조사+용언 활용형]의 조사 상당 구성이 문법화 되어 조사로 굳어지는 현상은 현대 한국어에서도 활발히 진행되고 있다. ‘조사 상당 구성 > 조사’의 과정은 한국어 조사 체계 형성에 있어 매우 중요한 역할을 담당하고 있으며 특히 다양한 부사격조사를 형성하는 주요 통로라 할 수 있다. 그 결과 부사격조사는 매우 다양한 형태들이 미묘한 의미의 차이를 보이며 공존하는 현상을 보인다. 조사 체계 내에서 다수의 형태들이 미묘한 의미 차이를 보이며 공존하는 현상을 올바르게 이해하고 설명하기 위해서는 먼저 ‘조사 상당 구성’들이 보이는 이러한 현상을 이해·설명해야 할 것이다.

2. 기본적인 논의

2.1. 조사 상당 구성의 목록 및 분류

조사 상당 구성은 [조사+용언 활용형]의 고정된 결합 구성으로 그 전체가 조사의 기능을 담당하는 구성을 말한다(문병열 2015: 27). 다음과 같은 구성들이 이에 해당된다.

- (1) 가. 가장 많이 가입한 보험 상품에 대해 알아보자.⁴⁾
 - 나. 장기간에 걸친 정부의 직접 개입
 - 다. 그 형성 시기에 있어 마을마다 차이가 난다.
 - 라. 1등은 고사하고 중간도 못 가는 성적이다.
 - 마. 광한루원을 비롯해 만복사지, 춘향 테마파크 등을 만날 수 있다.

(1)의 밑줄 친 구성들은 모두 [조사+용언 활용형]의 통사적 구성으로 그 전체가 조사의 기능을 담당하고 있다. (1가)의 [-에 대해] 구성은 해당 문장에서 목적격조사와 같은 기능을 보인다. (1나)의 [-에 걸친] 구성은 관형격조사의 기능을 담당한다. (1라)의 [-에 있어] 구성은 부사격조사의 기능을, (1마)의 [-은 고사하고] 구성은 보조사의 기능을, (1마)의 [-을 비롯해] 구성은 접속조사의 기능을 각각 담당하고 있다.

(1)의 밑줄 친 구성들은 조사 상당 구성의 전형적인 예들이다. 조사 상당 구성의 전체 목록에 대해서는 연구자에 따라 차이를 보이는데 대략 30여 개에서 60여 개 정도로 파악하고 있다.⁵⁾ 본고에서는 문병열(2015: 61)의 논의

4) 익명의 심사자께서 “보험 상품(A/B) 알아보자”의 경우 ‘A’ 또는 ‘B’가 보조사인지 격조사인지 알 수 없으며 따라서 [-에 대해] 구성이 보조사 상당 구성인지 격조사 상당 구성인지 확정할 수 없다 지적해 주셨다. 그러나 [-에 대해] 구성은 보조사와 같이 다양한 분포를 보이지 않으며 ‘목적격’ 위치, ‘부사격’ 위치에만 실현된다. 따라서 [-에 대해] 구성은 목적격 조사 상당 구성 또는 부사격 조사 상당 구성으로 볼 수 있다.

5) Martin(1992)에서는 53개의 조사 상당 구성 목록을 제시하였고, 이희자(1995)에서

를 받아들여 다음의 조사 상당 구성들을 연구 대상으로 삼을 것이다.⁶⁾

(2) 조사 상당 구성의 목록

- [-에 걸쳐], [-에 걸친]; [-에 관하여], [-에 관한];
- [-에 대하여], [-에 대한]; [-에 따라], [-에 따른];
- [-에 반하여]. [-에 부처], [-에 부치는], [-에 비추어]
- [-에 의하여], [-에 의한]; [-에(게) 있어], [-에(게) 대고]
- [-에도 불구하고], [-로 말미암아], [-로 말미암은]; [-로 미루어]
- [-로 인하여], [-로 인한]; [-은 고사하고], [-을 가지고/갖고]
- [-을 기하여], [-을 놓고], [-을 두고], [-을 둘러싸고], [-을 둘러싼];
- [-을 보고], [-을 비롯하여], [-을 비롯한]; [-을 위하여], [-을 위한];
- [-을 전후하여], [-을 전후한]; [-을 통하여], [-을 통한];
- [-을 향하여], [-을 향한]

(2)의 조사 상당 구성들은 개별 구성의 문법적 기능에 따라 분류할 수 있다. (1)에서 조사 상당 구성의 전형적인 예들을 보이며 간략하게 그 분류를 보인 바 있다. 이를 다음과 같이 도식화 할 수 있을 것이다(문병열 2015: 69).

(3) 조사 상당 구성의 분류



격조사 상당 구성에는 목적격조사 상당 구성, 관형격조사 상당 구성, 부사

는 24개의 목록을 제시하였다. 서경숙(2004)에서는 59개의 목록을, 임근석(2009)에서는 61개의 목록을 각각 제시하였다.

6) 조사 상당 구성의 목록 확정과 관련된 논의는 문병열(2015: 51~61)을 참고할 수 있다.

격조사 상당 구성이 확인되며 주격조사 상당의 구성은 찾아볼 수 없었다. [-에 대해] 구성과 [-에 관해] 구성의 일부 용례가 목적격조사 상당 구성으로 확인되었으며 또한 보조사 상당 구성의 경우 [-은 고사하고]⁷⁾ 구성만 확인되었고 접속조사 상당 구성에는 [-을 비롯해] 구성만 확인되었다. 따라서 대부분의 조사 상당 구성들은 관형격조사 상당 구성과 부사격조사 상당 구성에 해당되었다.⁸⁾

앞서 언급한 바, (3)과 같은 거시적 체계는 조사 상당 구성의 전체 모습을 조망하기에 부족함이 있다. 특히 구성의 대부분이 관형격조사 상당 구성과 부사격조사 상당 구성에 속해 있으며 이들이 미묘한 의미 차이를 보이며 공존한다는 점에서 미시적 분류 체계를 세울 필요가 있을 것이다.⁹⁾ 이를 위해 개별 조사 상당 구성들의 분포를 확인하고 이를 통계 처리 하여 분류해야 할 것이다.

2.2. 문법소의 의미 기술을 위한 통계 기법

어떤 언어 형식의 의미를 세밀하게 기술하기 위해서는 그 분포를 정확히 살펴야 한다. 특정 언어 형식의 분포는 그 단위의 의미·기능을 파악할 수 있는 (간접적이지만 유일한) 근거이기 때문이다. 이러한 분포의 기술에는 통

7) [-은 고사하고] 구성은 보조사 ‘-은/컸/녕’과 그 기능과 분포가 매우 비슷하여 보조사 상당 구성으로 처리하였다.

8) [-에 대해] 구성과 [-에 관해] 구성의 일부 용례는 목적격조사의 기능을 보이고 있으나 대부분의 용례들은 부사격조사의 기능을 담당한다. [-을 비롯해] 구성의 경우도 용례의 일부는 접속 조사의 기능을 담당하고 다른 예들은 부사격조사의 기능을 담당한다.

익명의 심사자께서 [-에 대해] 구성은 그 분포가 목적격 조사와 다르기 때문에 부사격 조사 상당 구성으로 파악해야 함을 지적해 주셨다. 그러나 ‘X’와 ‘X 상당 구성’의 분포가 항상 일치하는 것은 아니며 [-에 대해] 구성을 부사격 조사 상당 구성으로 파악하면 해당 구성이 포함된 전체 문장은 목적어가 없는 문장이 된다. 따라서 본고에서는 [-에 대해] 구성의 목적격 조사 상당의 기능을 인정한다.

9) 문병열(2015)에서는 조사 상당 구성의 미묘한 의미 차이를 의미 지도 모형을 통해 파악한 바 있다.

계적 기법이 사용되어야 한다. 특정 언어 형식 A의 분포는 결국 그에 인접한 또 다른 형식 B, C, D 등을 모아 놓은 것인데 이때 A와 B, C, D 등의 결합 관계가 (통계적으로) 유의미한 것인지 판단해야 하기 때문이다.

이렇게 인접한 두 언어 형식의 결합 관계가 유의미한지 판단하는 통계 기법에는 여러 가지가 있으나 언어학에서는 주로 ‘상대 확률’, ‘T-test’, ‘Z-score’ 등이 사용되었다. 먼저, ‘상대 확률’에 대해 살펴보자. 상대 확률은 어떤 언어 형식 A가 말뭉치 전체에서 출현할 확률과 그 언어 형식 A가 또 다른 언어 형식 B와 공기할 확률을 비교한 값을 말한다. ‘언어 형식 A가 말뭉치 전체에서 출현할 확률’을 P1이라 하고 ‘그 언어 형식 A가 또 다른 언어 형식 B와 공기할 확률’을 P2라 할 때, ‘P2/P1’의 값이 클수록 A와 B의 결합 강도가 큰 것으로 볼 수 있다. 이를 식으로 나타내면 다음과 같다.

(4) 상대 확률

$$P2/P1 = \frac{P2 = (A와 B의 공기 빈도 / B의 전체 출현 빈도)}{P1 = (A의 전체 출현 빈도 / 말뭉치 전체의 크기)}$$

‘P1’은 전체 말뭉치에서 A라는 언어 형식이 출현한 빈도를 나타내는데 이것은 이 말뭉치 안에서 A가 출현할 일반적인 확률(기댓값)을 나타내고 P2는 A가 B와 함께 출현한 실제 값(관찰값)을 B의 전체 출현 빈도와 비교하여 확률로 나타낸 것이다.

다음으로 ‘T-검정’에 대해 살펴보겠다. T-검정의 계산식은 다음과 같다.

(5) T-test

$$t = \frac{O-E}{\sqrt{o}}$$

(5)의 식에서 ‘O’는 ‘검색 공간내의 관찰값’, ‘E’는 ‘기댓값’에 해당된다. 이 식은 언어학뿐만 아니라 사회과학에서도 일반적으로 많이 사용된다. 분모의

‘ \sqrt{o} ’는 관찰값의 보정을 위한 ‘표준 오차’라 할 수 있다. 관찰값이란 특정 표본을 대상으로 조사한 값이기 때문에 모집단의 실제 모습과는 얼마간 차이를 보일 수밖에 없으며 따라서 ‘표준오차’를 사용하여 이를 보정해야 한다. T-test는 모집단이 정규 분포를 따를 때 사용한다.

언어학에서 가장 많이 사용되는 통계 기법 중 하나는 ‘Z-score’이다. 다음은 Z-score를 찾는 검정식이다.

(6) Z-test

$$Z = \frac{O-E}{\sigma}$$

(5)의 식에서와 같이, ‘O’는 ‘검색 공간내의 관찰값’, ‘E’는 ‘기댓값’에 해당된다. 그리고 ‘ σ ’은 ‘모집단의 표준분산’을 의미한다. ‘ σ ’ 역시 관찰값의 보정을 위한 표준오차이다. “이 식은 원칙적으로 모집단의 평균과 분산을 정확히 알고 있는 경우에만 쓸 수 있다. 다만 실제 공기관계 연구 적용에 있어 식 적용의 결과가 대체로 언어 직관과 일치하며 내용이 크게 왜곡되지 않는다고 판단되어 여러 연구들에서 사용되고 있다(박병선 2005: 51).” 실제 언어 자료를 대상으로 Z-score를 구해 보면, 그 결과가 우리의 언어 직관에 크게 벗어나지 않는 것을 확인할 수 있다. 그러나 그 값이 일반적인 통계학에서 예상·통용되는 것에 비해 상당히 크게 나타난다. 이는 T-test의 경우도 마찬가지인데 따라서 T-test와 Z-score를 사용할 때는 각별한 주의가 필요하다. 또한 이들은 관찰값과 기댓값의 차이가 정규분포를 이룬다는 가정을 바탕으로 하고 있음을 염두에 두어야 할 것이다.¹⁰⁾

(4~6)의 통계 기법들은 결국 A라는 언어 형식이 출현할 것으로 예상되는 일반적인 확률(기댓값)과 A라는 형식이 B와 공기하여 나타난 실제 출현 빈도(관찰값) 또는 그 확률을 비교하여 A와 B의 결합 관계가 통계적으로 유의

10) Dunning(1993)에서는 이 가정이 실제와는 다른 양상을 보임을 지적한 바 있다. (박병선 2005: 50에서 재인용)

미한 것인지 판단하는 것이다. (4)에 비해, (5, 6)의 기법들은 통계적으로 좀 더 세련되고 정제되어 있다는 점에서 장점을 갖는다.¹¹⁾ 반면 여러 가지 가정에 기반하고 있다는 점, 그리고 그 가정들이 실제 언어 자료의 양상과 다를 수 있다는 점을 유의해야 할 것이다.

지금까지는 언어 형식 A와 B의 공기 관계가 유의미한 것인지 판단하는 통계 기법을 살펴보았다. 언어 형식 A가 문법 형식이라면 A는 비슷한 의미·기능을 갖는 일군의 문법 형식들과 하나의 체계를 이루고 있을 것이다. 만약 A가 조사 중 하나라면 이는 비슷한 의미·기능을 갖는 다양한 조사들과 하나의 체계를 이루고 있을 것이며 개별 조사들의 친소 관계를 파악하여 그 공통점과 차이점을 체계적으로 기술해야 할 것이다.

다수의 요소들을 그 친소 관계에 따라 체계적으로 묶어 보여주는 통계 기법으로 군집 분석이 있다. 예를 들어, 5명의 학생이 있고, 5개의 교과목이 있을 때, 이 5명의 학생들이 성적(또는 적성)이라는 면에서 서로 어떤 친소 관계를 보이는지 확인하여 이들을 여러 층위(또는 단계)에서 묶어 볼 수 있다. 다음의 표를 살펴보자.

(7) 5명 학생들의 과목 성적

	학생 A	학생 B	학생 C	학생 D	학생 E
국어	100	100	90	70	70
영어	100	100	90	70	60
수학	100	70	70	90	95
사회	95	80	70	90	80
과학	95	60	60	95	95

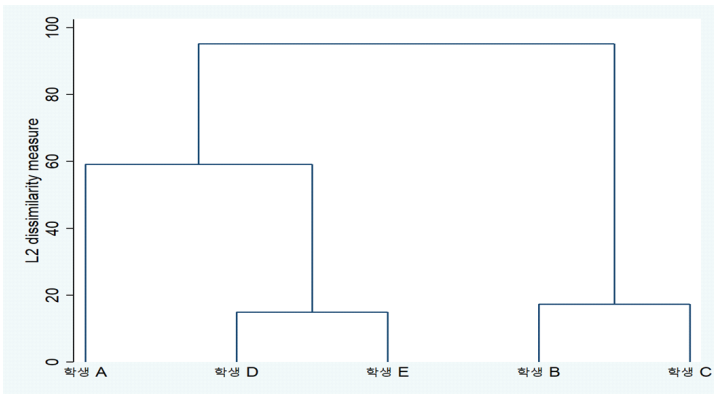
먼저 학생 B와 학생 C는 국어, 영어 점수가 다른 과목의 점수에 비해 높다는 점에서 같이 묶일 것이고, 학생 D와 학생 E는 수학, 과학 점수가 다른

11) 관찰값의 오차 가능성을 표준 오차로 보정한다는 점에서 그러하다.

과목에 비해 높기 때문에, 역시 같은 군집으로 분석될 것이다. 학생 A는 모든 과목에서 높은 점수를 보이고 있어 따로 분석된다. 따라서 우리는 학생 A는 모든 영역에서 뛰어난 군집으로 분류하고, 학생 B, C는 언어적 능력이 뛰어난 군집으로 분류하고 학생 D, E는 수학적 능력이 뛰어난 군집으로 분류할 수 있다. 이를 통해, 우리는 개별 학생들의 특성뿐 아니라 전체 학생들이 보이는 친소 관계의 체계를 확인할 수 있게 된다.

군집 분석에도 여러 가지 기법이 존재한다. 본고는 박진호(2003)에서 사용된 계층적 병합 분석(hierarchical agglomerative clustering)을 사용하며, 유클리드 거리(Euclidean distance)를 사용한다.¹²⁾ 즉, 군집을 몇 개로 나눌 것인지 미리 정하지 않고 유클리드 거리를 사용하여 가까운 거리에 있는 구성들을 먼저 군집으로 묶어 올라가는 방식을 사용한다. (7)의 예시를 대상으로 군집 분석을 하면 다음과 같은 결과가 도출된다.

(8) 5명 학생들의 군집 분석 결과



12) 군집 분석과 관련된 자세한 내용은 박진호(2003: 114~118)을 참고할 수 있다.

3. 조사 상당 구성의 군집 분석

본 장에서는 2.1.에서 제시한 40여 개의 조사 상당 구성들의 군집 분석을 시도한다. 조사 상당 구성의 분류와 관련하여 보조사 상당 구성, 접속조사 상당 구성의 구성원은 각각 1개뿐이며 목적격조사 상당 구성의 구성원은 2개뿐이다. 따라서 본 장에서는 관형격조사 상당 구성과 부사격조사 상당 구성의 두 부류만 군집 분석을 실시한다.

3.1. 조사 상당 구성의 결합 양상

본고의 분석 대상이 되는 말뭉치는 ‘21세기 세종계획 형태의미분석 말뭉치’이다. 대략 1,200만 어절의 규모이며 동음이의어가 구별되어 있다. 이를 대상으로 개별 조사 상당 구성들의 선행 요소를 추출하여 그 상대 확률과 Z-score를 구한다. 다음은 [-에 걸친] 구성의 경우이다.¹³⁾¹⁴⁾

13) 이 과정은 Python 프로그래밍 언어를 사용하여 추출하였다. 자세한 내용은 문병열(2015)를 참고할 수 있다. 상대 확률과 Z-score는 R-프로그래밍 언어를 사용하여 추출하였다. 다만 상대확률을 구하는 식의 입력값은 문병열(2015)와 다르다. (1)은 문병열(2015: 13)의 것이고 (2)는 본고의 방식이다. (2)의 방식이 [-에 걸친] 구성의 확률 분포를 보이기엔 더 적합하다.

- (1) $P1 = \text{‘전반’의 전체 출현 빈도}(351) / \text{말뭉치 전체의 어절 수}(12,691,945)$
 $P2 = \text{‘전반이 [-에 걸친] 구성 앞에 출현한 빈도}(23) / \text{‘전반’의 전체 출현빈도}(351)$
 $P2/P1 = \{(23/351)/(351/12,691,945)\} = 2,369.419$
- (2) $P1 = \text{‘전반’의 전체 출현 빈도}(351) / \text{명사 전체의 어절 수}(5,314,715)$
 $P2 = \text{‘전반이 [-에 걸친] 구성 앞에 출현한 빈도}(23) / \text{‘[-에 걸친]’의 전체 출현빈도}(169)$
 $P2/P1 = \{(23/169)/(351/5,314,715)\} = 592.275$

14) 본고에서는 조사 상당 구성의 의미 기능을 파악하기 위해 그 선행 요소만 추출하였다. 전체 분포를 확인하려면 후행 요소도 추출해야 할 것이나 본고에서는 다

(9) [-에 걸친] 구성의 선행 요소 결합 양상

명사	빈도 1	빈도	상대 확률	z-score	명사	빈도 1	빈도	상대 확률	z-score
차례_01/NNG	2028	29	129.2506	5.48035	행_01/NNG	84	1	107.6028	0.99155
전반_03/NNG	351	23	592.275	4.88421	연도_01/NNG	85	1	106.3368	0.99144
년_02/NNB	30216	54	1.91571	3.68496	주간_05/NNG	92	1	98.24599	0.99066
분야/NNG	2169	13	54.17345	3.57877	양면/NNG	93	1	97.18958	0.99055
장기간/NNG	127	7	498.1923	2.6563	코플/NNG	107	1	84.47319	0.989
평생/NNG	786	7	80.49671	2.62857	쪽_x02/NNB	19	1	56.41828	0.9831
시간_04/NNG	10154	8	7.12124	2.44794	목덜미/NNG	178	1	50.77883	0.98114
간_x16/NNG	1459	6	37.17052	2.39584	전반_04/NNG	181	1	49.93719	0.98081
시간/NNB	497	5	10.78418	2.03731	페이지/NNG	200	1	45.19315	0.9787
차/NNB	2445	8	3.50739	2.03563	생애/NNG	270	1	33.47641	0.97095
수차례/NNG	59	4	612.7885	2.00356	항목/NNG	323	1	27.98338	0.96508
수차_04/NNG	12	3	2259.658	1.73572	턱_01/NNG	361	1	25.03776	0.96088
다방면/NNG	21	3	1291.233	1.73514	수_04/NNB	43	1	24.92901	0.96069
부문_06/NNG	1053	3	25.75109	1.66905	광년/NNB	45	1	23.82105	0.95882
어깨_01/NNG	1489	3	18.21081	1.64113	일생_01/NNG	410	1	22.04544	0.95545
대_06/NNB	2385	6	2.69672	1.54882	업종_01/NNG	430	1	21.02007	0.95323
달_05/NNG	3018	3	8.98472	1.54321	마루_03/NNG	436	1	20.7308	0.95257
반세기/NNG	72	2	251.0731	1.41098	아들_01/NNG	537	1	16.83171	0.94139
방면_01/NNG	191	2	94.64535	1.40166	간_16/NNB	67	1	15.99921	0.93827
앞뒤/NNG	229	2	78.94001	1.39868	일대_03/NNG	583	1	15.50366	0.93629
항_08/NNG	353	2	51.21037	1.38896	끝_01/NNG	665	1	13.59193	0.92721
장기_10/NNG	462	2	39.12827	1.38042	범위/NNG	739	1	12.23089	0.91902
아프리카/NNP	599	2	30.17907	1.36968	마일/NNB	96	1	11.16612	0.91118
며칠/NNG	1266	2	14.27904	1.31741	감각_02/NNG	917	1	9.85674	0.89931
영역_03/NNG	1448	2	12.4843	1.30315	코_01/NNG	1015	1	8.90506	0.88845

음의 이유로 선행 요소만 파악할 것이다. 첫째, [조사+용언활용형]의 조사 상당 구성의 의미·기능은 구성 내의 ‘용언활용형’이 상당부분 담당한다. 예를 들어 [-로 인하여]의 의미가 [원인]인 것은 ‘인하여’ 때문이며, [-을 통한]이 [경로]인 것은 ‘통한’ 때문이다. 따라서 조사 상당 구성의 의미·기능은 후행 요소보다 선행 요소와 더 깊이 관련될 것이다. 선행 요소가 구성 내의 용언활용형의 논항인 경우가 대부분이기 때문이다. 둘째, [-에 걸쳐서]의 후행 요소인 ‘연구하다’를 추출하려면 충분한 크기의 구문 분석 말뭉치가 필요하지만 현재 구축된 구문 분석 말뭉치는 그 크기가 너무 작아서 이러한 연구에 적합하지 않다.

세대_02/NNG	1545	2	1170049	129554	세월_02/NNG	1118	1	8.08464	0.87705
개월/NNB	1601	4	2.67819	1.25729	소련/NNP	1120	1	8.07021	0.87683
기간_07/NNG	2210	2	8.17976	1.24343	동안_01/NNG	6899	2	2.62027	0.87597
세기_03/NNG	2675	2	6.75785	1.20699	한참/NNG	1332	1	6.78576	0.85335
부분_01/NNG	2980	2	6.0662	1.18309	독일/NNP	1738	1	5.20059	0.80839
회/NNB	1773	4	2.41838	1.17678	시기_04/NNG	1935	1	4.67113	0.78658
주_26/NNB	375	2	5.71705	1.16874	세계_02/NNG	8680	2	2.08263	0.7364
용문동/NNP	1	1	9038.631	1.00074	전체_01/NNG	3073	1	2.94131	0.66056
다기_01/NNG	2	1	4519.315	1.00063	밤_01/NNG	4278	1	2.11282	0.52713
최승로/NNP	2	1	4519.315	1.00063	머리_01/NNG	4697	1	1.92434	0.48074
광범위/NNG	3	1	3012.877	1.00052	간_10/NNB	4232	5	1.26648	0.47233
경상남도/NNP	4	1	2259.658	1.00041	과정_03/NNG	4798	1	1.88383	0.46955
누대_02/NNG	6	1	1506.438	1.00019	주일/NNB	633	1	1.69344	0.40972
소아시아/NNP	9	1	1004.292	0.99985	지역_03/NNG	5979	1	1.51173	0.33878
여/NNP	15	1	602.574	0.99919	사이_01/NNG	6601	1	1.36928	0.26991
팔걸이/NNG	17	1	531.6842	0.99897	몸_01/NNG	6929	1	1.30446	0.23359
달_05여/NNG	18	1	502.1462	0.99886	손_01/NNG	7562	1	1.19527	0.1635
일_07/NNG	18	1	502.1462	0.99886	초_03/NNB	1333	1	0.80416	-0.2436
단기간/NNG	61	1	148.1743	0.9941					

(9)에서 ‘빈도1’은 말뭉치 전체에서 해당 명사가 출현한 빈도이고 ‘빈도2’는 이 명사가 [-에 걸친] 앞에 출현한 빈도이다. 상대 확률과 Z-score는 각각 (4), (6)의 식을 사용하여 얻은 값이다. ‘전반_03/NNG’을 예로 들어보면 다음과 같다.

(10) 가. 상대확률식

$$P2/P1 = \frac{(\text{‘전반_03’과 [-에 걸친]의 공기 빈도} / \text{[-에 걸친]의 전체 출현 빈도})}{(\text{‘전반_03’의 전체 출현 빈도} / \text{명사 전체의 크기})}$$

나. '전반_03/NGG'의 상대확률

$$592.275 = \frac{(23 / 588)}{(351 / 5,314,715)}$$

(11) Z-score 수식¹⁵⁾¹⁶⁾

$$Z\text{-score} = \frac{\hat{P}(N_j|CFK_i, N) - \hat{P}(N_j|N)}{\sqrt{\frac{\hat{P}(N_j|CFK_i, N) \times [1 - \hat{P}(N_j|CFK_i, N)]}{\# \text{ of } CFK_i} + \frac{\hat{P}(N_j|N) \times [1 - \hat{P}(N_j|N)]}{\# \text{ of } N}}$$

(9)의 표를 통해, 우리는 [-에 걸친] 구성의 특성을 확인할 수 있다. [-에 걸친]에 선행하는 명사들이 주로 시간이나 횟수를 나타내는 것임을 알 수

15) CFK는 Complex-Functional Case-Marker의 약어이다. 복합 구성(Complex Construction)의 약어에 대해서는 문병열(2014)를 참고할 수 있다.

16) 문병열(2015)에서는 다음과 같은 Z-score 계산식을 사용하였다.

$$Z\text{-score} = \frac{\hat{P}(N_j|CFK_i, N) - \hat{P}(N_j|N)}{\sqrt{\frac{\hat{P}(N_j|N) \times [1 - \hat{P}(N_j|N)]}{\# \text{ of } N}}}$$

Z-score 식의 분모는 '표준오차'를 반영하는데, 보통의 경우 관찰값만을 추정치로 간주하여 이의 표준오차를 분모에 반영한다. 그러나 본고의 경우, 관찰값과 기대값 모두 추정치이기 때문에 이들 모두의 표준오차를 반영해야 한다. 따라서 본문 (11)과 같은 계산식을 사용하는 것이 더 적절하다.

기존 연구의 대부분이 각주의 방식으로 Z-score를 계산해 왔다. 그 결과 분모의 값이 너무 작아 상대적으로 Z-score의 값이 크게 산출되었다. 본고는 본문 (11)의 방식으로 계산식을 세밀하게 보정한바 그 결과값이 통계학적으로 받아들일 만한 수준으로 보정되었다. 통계학적으로 Z-score 값이 1.96을 넘으면 그 결합관계가 유의미한 것으로 판단할 수 있다. 그러나 각주의 방식으로 Z-score를 계산하게 되면 그 값이 100 이상으로 산출되는 경우도 빈번하다. 통계학의 관점에서 이러한 값들은 매우 이례적이며 이해하기 어려운 것이다.

있다. 용언 ‘걸치-’의 어휘 의미를 생각해 볼 때, 공간이나 장소를 나타내는 명사들과도 공기할 수 있을 것으로 보이지만 실제로는 그러한 공기 관계가 통계적으로 유의미하게 확인되지 않는다. 이를 통해, 우리는 [-에 걸친] 구성 이 이미 어느 정도 문법화 되었음을 알 수 있다.

또 다른 구성인 [-을 둘러싼]을 살펴보자. [-을 둘러싼] 구성의 선행 요소 결합 양상은 다음과 같다.

(12) [-을 둘러싼] 구성의 선행 요소 결합 양상

명사	빈도 1	빈도	상대 확률	z-score	명사	빈도 1	빈도	상대 확률	z-score
문제_06/NNG	14816	12	6.02882	2.91407	석작/NNG	12	1	620.2982	0.99909
한반도/NNP	697	6	64.07672	2.42146	검사_02제/NNG	13	1	572.583	0.99895
당_14/NNG	1718	4	17.3308	1.88989	뇌사_03/NNG	13	1	572.583	0.99895
정책_02/NNG	3453	4	8.62274	1.77301	이시은/NNP	13	1	572.583	0.99895
사건_01/NNG	4534	4	6.5669	1.70019	발효유/NNG	15	1	496.2386	0.99868
공사_02/NNG	908	3	24.59332	1.66512	분/NNG	16	1	465.2237	0.99855
여부_01/NNG	1040	3	21.47186	1.65486	공과_04/NNG	18	1	413.5321	0.99828
수입_02/NNG	1067	3	20.92852	1.65276	광해군/NNP	21	1	354.4561	0.99788
수입료/NNG	32	2	465.2237	1.41315	금서_03/NNG	23	1	323.6338	0.99761
사퇴_04/NNG	139	2	107.1019	1.40297	이라크전/NNG	23	1	323.6338	0.99761
공천_02/NNG	174	2	85.55837	1.39964	쟁점들/NNG	23	1	323.6338	0.99761
분배/NNG	204	2	72.97626	1.39679	계획표/NNG	24	1	310.1491	0.99747
쟁점/NNG	245	2	60.76391	1.39289	건의안/NNG	27	1	275.6881	0.99707
선정_07/NNG	332	2	44.84083	1.38461	국방비/NNG	27	1	275.6881	0.99707
사방_03/NNG	499	2	29.83398	1.36873	도심지/NNG	27	1	275.6881	0.99707
통치_03/NNG	597	2	24.93661	1.3594	증액/NNG	27	1	275.6881	0.99707
도입/NNG	598	2	24.89491	1.35931	월차_01/NNG	28	1	265.8421	0.99694
건축_01/NNG	681	2	21.86073	1.35141	내각_03제/NNG	29	1	256.6751	0.9968
구성_07/NNG	1122	2	13.26841	1.30946	편입학/NNG	29	1	256.6751	0.9968
핵/NNG	1199	2	12.41631	1.30213	수칙_01/NNG	30	1	248.1193	0.99667
평화_02/NNG	1398	2	10.6489	1.2832	동강/NNP	34	1	218.9288	0.99613
건설/NNG	1494	2	9.96463	1.27407	일부일차제/NNG	34	1	218.9288	0.99613
권력_02/NNG	1796	2	8.28906	1.24534	인양/NNG	36	1	206.7661	0.99586
평가_03/NNG	1857	2	8.01678	1.23953	거취_02/NNG	37	1	201.1778	0.99573
책임/NNG	2121	2	7.01893	1.21442	지참금/NNG	39	1	190.861	0.99546

개념/NNG	2158	2	6.89859	1.2109	변별력/NNG	40	1	186.0895	0.99532
성격_02/NNG	2482	2	5.99805	1.18008	후에/NNP	41	1	181.5507	0.99519
규모/NNG	2567	2	5.79944	1.17199	대우그룹/NNP	42	1	177.2281	0.99505
자신_01/NNG	12393	4	2.40251	1.17079	성곽/NNG	42	1	177.2281	0.99505
가치_06/NNG	2701	2	5.51172	1.15924	표절/NNG	42	1	177.2281	0.99505
문화_01/NNG	8154	3	2.73862	1.10189	노점_03/NNG	44	1	169.1722	0.99479
채굴권/NNG	3	1	2481.193	1.0003	의사당/NNG	44	1	169.1722	0.99479
확전론/NNG	3	1	2481.193	1.0003	산청_09/NNG	46	1	161.8169	0.99452
심장사/NNG	5	1	1488.716	1.00003	납품/NNG	47	1	158.374	0.99438
전자업계/NNG	5	1	1488.716	1.00003	수주_06/NNG	48	1	155.0746	0.99425
탁치/NNG	5	1	1488.716	1.00003	책임론/NNG	48	1	155.0746	0.99425
향권/NNG	5	1	1488.716	1.00003	특검제/NNG	48	1	155.0746	0.99425
찬탈/NNG	7	1	1063.368	0.99976	난이도/NNG	49	1	151.9098	0.99411
보전_04/NNG	8	1	930.4473	0.99963	들도끼/NNG	51	1	145.9525	0.99384
성배_02/NNG	8	1	930.4473	0.99963	예술론/NNG	52	1	143.1457	0.99371
국조_88권/NNG	9	1	827.0643	0.99949	위행찬/NNP	52	1	143.1457	0.99371
여행비/NNG	9	1	827.0643	0.99949	유로화/NNG	54	1	137.8441	0.99344
카지노업계/NNG	9	1	827.0643	0.99949	착용/NNG	55	1	135.3378	0.99331
수음_02/NNG	10	1	744.3578	0.99936	이하 생략				

(12)의 표 역시 [-을 둘러싼] 구성을 대상으로 (10, 11)과 같은 방식으로 그 상대 확률 및 Z-score 값을 얻은 것이다. 나머지 조사 상당 구성들 모두 같은 방식으로 그 선행 명사의 결합 양상을 구하였다.

3.2. 관형격조사 상당 구성의 군집 분석

본 절의 분석 대상이 되는 관형격조사 상당 구성의 목록은 다음과 같다.

(13) 관형격조사 상당 구성의 목록

[-에 걸친], [-에 관한], [-에 대한], [-에 따른], [-에 의한]
[-로 말미암은], [-로 인한] [-을 둘러싼], [-을 비롯한], [-을 위한]
[-을 진후한], [-을 통한], [-을 향한]

앞서, (9)를 통해 [-에 걸친] 구성의 선행 요소 결합 양상을 살펴보고 (12)를 통해 [-을 둘러싼] 구성의 선행 요소 결합 양상을 살펴보았다. (13)의 나머지 조사 상당 구성들 역시 (9, 12)와 같은 방식으로 선행 요소 결합 양상을 모두 정리하였다.

우리는 앞서 (7-8)을 통해, 5명의 학생들을 5개의 교과목 성적을 기준으로 군집 분석을 시도한 바 있다. 군집 분석을 수행하기 위해서는 분석 대상 자료가 (7)과 같은 표로 정리되어야 한다. 조사 상당 구성들 각각은 (7)의 개별 학생들에 해당되고 선행하는 체언들은 교과목에 해당된다. 그리고 Z-score는 과목 성적에 해당된다. 이를 반영하여, 관형격조사 상당 구성의 군집 분석을 위해 관련 자료를 정리하면 다음과 같다.

(14)

	[-에 걸친]	[-에 관한]	[-에 대한]	[-에 따른]	...
차례_01	5.48035	-45.04192	-10.69239	-45.04192	...
관계_05	-82.80601	3.81969	4.38661	-0.24534	...
개념	-46.46371	0.86598	2.7218	-0.07925	...
개정_02	-21.09591	1.08355	-2.23607	2.13853	...
⋮	⋮	⋮	⋮	⋮	...

(14)의 표에서 확인할 수 있듯이, Z-score는 양의 값(+)을 가질 수도 있고, 음의 값(-)을 가질 수도 있다. 이는 관련된 조사 상당 구성과 그 선행 요소의 결합 관계에 대해 중요한 정보를 제공해 준다.

어떤 언어 단위의 의미·기능을 파악하기 위해 그 선행 요소와의 결합 양상을 파악할 때에는 다음과 같은 구분이 필요하다.

(15) 결합 관계의 구분

가. 통계적으로 유의미한 결합 관계

- ㄱ. 양의 결합 관계: Z-score가 1.96 이상인 결합 관계
- ㄴ. 음의 결합 관계: Z-score가 -1.96 이하인 결합 관계
- 나. 통계적으로 무의미한 결합 관계(우연한 결합)
 - ㄱ. 양의 결합 관계: Z-score가 0 ~ 1.96 사이인 결합 관계
 - ㄴ. 음의 결합 관계: Z-score가 0 ~ -1.96 사이인 결합 관계

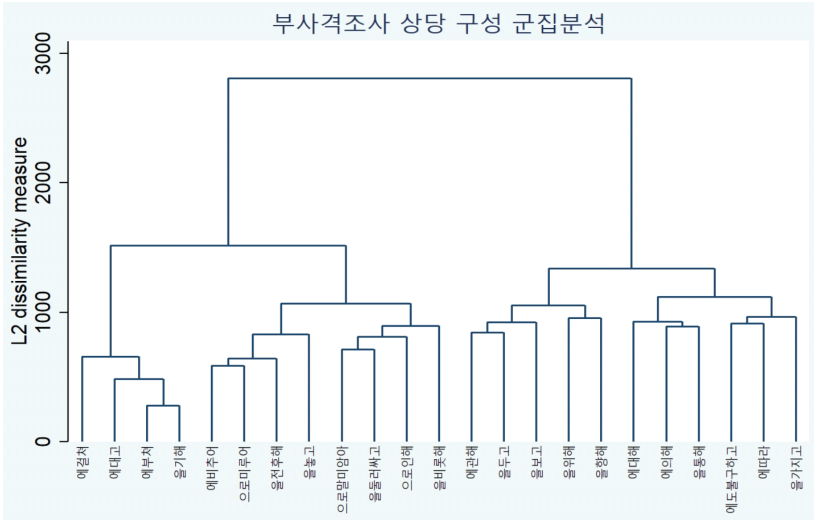
A와 B의 결합 관계를 Z-score로 환산했을 때, 그 값이 +1.96 이상이면 이들의 결합 관계는 통계적으로 유의미하다. 또한 그 값이 -1.96 이하인 경우도 마찬가지이다. 반면, 그 값이 -1.96과 +1.96의 사이라면 그 결합 관계는 통계적으로 무의미한 것이다. 예를 들어 [-에 따른] 구성과 ‘차례_01’의 결합 관계 값은 ‘-45.04192’이다. 이는 이 두 요소의 결합이 실현되지 않는 것이 통계적으로 매우 유의미하다는 것이다. 즉 ‘차례에 따른’과 같은 결합이 실현되지 않는 것은 우연한 것이 아니며 이 두 요소의 본질적 속성에 따른 것일 가능성이 높다는 것이다. ‘개정_02’와의 결합이 실현되는 것 역시 통계적으로 유의미한데 ‘개정에 따른’과 같은 결합의 실현은 이 두 요소의 본질적 의미와 관련된다는 것이다. 반면, ‘개념에 따른’이나 ‘개정에 따른’이 실현되지 않는 것(또는 실현되는 것)은 우연의 문제이다.

기존 논의에서는 (15가ㄱ)과 (15나ㄱ)의 구분을 중시하였다. 그러나 본고에서는 (15가ㄴ)역시 주목하고자 한다. 특정 조사 상당 구성이 특정 어휘와 (통계적으로 유의미하게) 결합하지 않는 것 역시, 해당 조사 상당 구성의 의미·기능을 파악하는 데 중요한 역할을 담당하기 때문이다.

(14)의 테이블을 자료로 하여 유클리드 거리(Euclidean distance) 값을 내고, 이를 계층적 병합 분석하면 다음과 같이 관형격조사 상당 구성의 군집 분석이 이루어진다.¹⁷⁾

17) 1,200만 어절 형태의미분석 말뭉치에서 조사 상당 구성과 그 선행 체인을 추출하는 과정은 Python 프로그래밍 언어를 사용하였고 그것의 상대 확률과 Z-score를 내는 과정은 R-프로그래밍 언어를 사용하였다. 그리고 군집 분석은 STATA 프로그래밍 언어를 사용하였다. 군집 분석과 관련된 통계 처리는 한남대학교 경제학부 전계형 교수께서 도움을 주셨다. 이에 甚深한 謝意를 표한다.

(18) 부사격조사 상당 구성의 군집 분석



4. 조사 상당 구성의 유형 분류

4.1. 관형격 조사 상당 구성의 유형 분류

(16)의 군집 분석 결과를 보면, 가장 먼저 [-에 대한] 구성과 [-에 관한] 구성이 묶이는 것을 알 수 있다. [-에 대한] 구성에 선행하는 요소들로는 {무정물}에 해당되는 ‘복제물, 중소기업, 사물’ 등이 있으며 {인간}에 해당되는 ‘관련자들, 타자, 타인’, {공간물}의 ‘업소, 식민지’, {추상}의 ‘이성, 대중문화’, {사건}의 ‘침해, 질문’, {현상}의 ‘질서, 자극’, {사건}의 ‘죽음, 실패’ 등 매우 다양한 명사들과 결합한다.¹⁸⁾ [-에 관한] 구성은 {추상}과 관련된 명사들

18) 명사의 부류 구분은 최경봉(1998: 81)을 따랐다. 명사의 하위 분류와 관련하여서는 이현우(1995), 정희정(2000), 이선웅(2004/2005), 신중진(2007), 이영제(2014) 등을 참고할 수 있다.

이 선행 요소로 가장 많이 나타났다. ‘전반, 기원, 본질, 문제, 신상, 철학’ 등이 그 예들이다. 또한 {사건}으로 ‘처벌, 운영, 형성, 보호’, {무정물}로 ‘부동산, 소프트웨어, 프로그램’ 등이 확인된다. 이 두 구성은 [대상]을 나타내려는 공통점을 지닌다. 다만, [-에 대한] 구성이 좀 더 보편적인 [대상]을 나타낸다면 [-에 관한] 구성은 좀 더 추상적인 [대상]을 나타내는 경향이 있다.

다음은 이 구성들의 구체적인 예이다.

- (19) 가. 불법 복제물에 대한 단속
 - 나. 타자에 대한 책임
 - 다. 대중문화에 대한 인식
 - 라. 죽음에 대한 공포
- (20) 가. 인간의 본질에 관한 담론
 - 나. 철학에 관한 성찰
 - 다. 폭력행위 등 처벌에 관한 법률
 - 라. 소프트웨어에 관한 토막글

다음으로, [-에 의한] 구성과 [-을 위한] 구성이 하나의 군집으로 묶인다. [-에 의한]의 경우, {사건}에 해당되는 ‘강요, 작용’, {추상}에 해당되는 ‘저작권, 미디어, 규정, 법률’ 등과 결합한다. [-에 의한] 구성은 [도구]·[수단]·[원인] 등의 기능이 확인된다. [-을 위한] 구성의 경우, {추상}의 ‘공익, 통일, 승리, 성공’ 등과 결합하였으며 {사건}의 ‘저지, 달성, 방지’이 확인된다. [-을 위한]의 경우 {인간}에 해당되는 ‘장애인, 청소년, 어린이들, 농민, 시민’ 등도 그 선행 명사로 확인되었다. 이 구성은 [목적]·[수혜자]의 기능을 담당하고 있다.

다음은 이 구성들의 구체적인 예이다.

- (21) 가. 강요에 의한 명예퇴직
 - 나. 법률에 의한 통제
- (22) 가. 청소년을 위한 음악회
 - 나. 시민을 위한 강좌

다음으로 [-을 둘러싼] 구성과 [-을 비롯한] 구성이 묶인다. [-을 둘러싼] 구성은 ‘문제, 정책, 사건’ 등의 {추상}을 나타내는 명사들, ‘나, 자신’ 등의 {인간} 명사들, ‘이, 그’ 등의 지시대명사 등과 결합된다. 주요 기능은 논의의 [초점대상]을 나타낸다. [-을 비롯한] 구성의 경우, {공간}에 해당되는 ‘소련, 일본, 미국, 한국’ 등 국가명이 주로 결합하였다. 이 구성은 [대표]의 기능을 갖는다. 이 두 구성과 가깝게 묶이는 것으로 [-을 향한] 구성이 있다. [-을 향한] 구성은 {추상}의 ‘미래, 통일’, {공간}의 ‘세계’, {인간}의 ‘자신’ 등과 결합하였다. 이 구성은 [방향]·[목적] 등의 기능을 갖는다.

다음은 이 구성들의 구체적인 예이다.

- (23) 가. 이를 둘러싼 여러 문제들
나. 정책을 둘러싼 갈등들
- (24) 가. 한국을 비롯한 동아시아 국가들
나. 달리기를 비롯한 다양한 운동 경기
- (25) 가. 미래를 향한 도전
나. 세계를 향한 비전

[-에 따른] 구성과 [-로 인한] 구성이 하나의 군집으로 분석된다. [-에 따른] 구성은 {사건}에 해당되는 ‘급등, 침체, 상승’ 등과 결합하고 {추상}에 해당되는 ‘양심, 원칙, 계획, 능력’ 등과 결합한다. [-로 인한] 구성 역시 {추상}의 ‘오염, 문제’, {사건}의 ‘침체, 상승, 사용’ 등과 결합한다. 이 구성들의 기능은 [원인]·[기준]으로 볼 수 있다. 이 군집과 가깝게 묶이는 구성으로 [-을 통한]이 있다. [-을 통한] 구성은 {추상}에 해당되는 ‘망, 네트워크, 채널, 통신’, {무정물}에 해당되는 ‘은행, 전화, 인터넷, 매체’ 등과 결합한다. 이 구성은 [경로]·[과정]의 기능을 보인다. 다음은 이 구성들의 구체적인 예이다.

- (26) 가. IT 산업 부진에 따른 내수경기 침체
나. 능력에 따른 분배
- (27) 가. 경기 침체로 인한 수요 감퇴

- 나. 물가 상승으로 인한 서민들의 생활난
 (28) 가. 네트워크를 통한 마케팅
 나. 전화를 통한 예약 관련 질문들

[-로 말미암은] 구성과 [-을 전후한] 구성이 하나의 군집으로 묶인다. [-로 말미암은] 구성은 대명사 ‘이, 그’, {현상}의 ‘현상, 태도’, {사건}의 ‘행위, 침해, 등장’ 등과 결합하였다. 이 구성의 기능은 [원인]이다. [-을 전후한] 구성은 특정 시기 또는 특정 사건을 나타내는 명사들이 결합한다. ‘가을’과 같은 일반적인 시기를 나타내는 명사, ‘아침’, ‘오늘’ 등 짧은 시간을 나타내는 명사, ‘한 달, 일 년’ 등 기간을 나타내는 명사들은 오지 않는다. 이 구성의 기능은 [시기]라 할 수 있다. 이 군집과 가까이 묶이는 구성으로 [-에 걸친]이 있다. [-에 걸친]은 주로 {시간}과 관련되는 ‘장기간, 평생, 시간, 차례, 차, 대, 번’ 등과 결합한다. 따라서 이 구성은 [시간적 범위]를 나타낸다 할 수 있다.

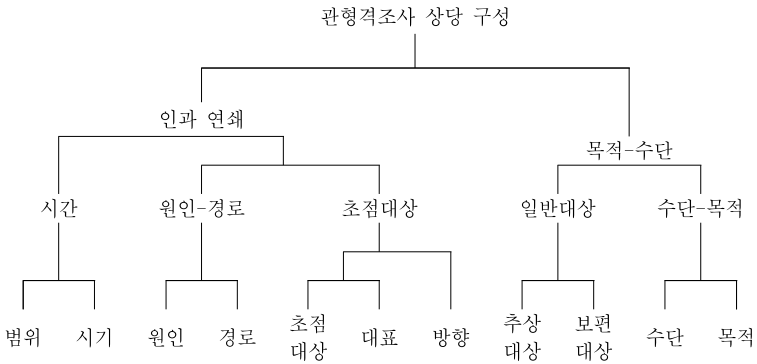
다음은 이 구성들의 구체적인 예이다.

- (29) 가. 선한 행위로 말미암은 축복
 나. 이로 말미암은 문제점
 (30) 가. 1940년을 전후한 조선의 언어 상황
 나. 설 명절을 전후한 선거법 안내
 (31) 가. 장기간에 걸친 가뭄
 나. 3번에 걸친 시도

위의 논의를 바탕으로 관형격조사 상당 구성의 유형을 분류하면 다음과 같다.

- (32) 관형격조사 상당 구성의 유형 분류¹⁹⁾

19) [-로 말미암은] 구성은 분류에서 제외되었다. 이 구성은 군집 분석에서 [-에 걸친], [-을 전후한] 등의 구성과 가까운 것으로 묶였으나 이는 언어학적으로 해석하기 어려울 듯하다. 자료의 한계, 통계 분석의 한계로 보인다.



(32)의 표를 통해, 우리는 [일반대상]과 [수단-목적]이 더 가깝게 묶이고 [초점대상]과 [원인-경로]가 더 가깝게 묶이는 것을 알 수 있다. 또한 [시간]이 [원인-경로]·[초점대상]과 더 가깝게 묶이는 것을 알 수 있다. 그리고 인과 연쇄(causal chain) 상의 [결과]가 관형격조사 상당 구성에서는 발견되지 않음을 알 수 있는데 이는 관형격조사 상당 구성의 수식을 받는 명사가 주로 [결과]를 나타내기 때문이다.

(32)의 표를 조금 더 살펴보겠다. 관형격 조사 상당 구성은 크게 두 부류로 나뉘는데, 그 하나는 [인과 연쇄]와 관련된 것으로 “[원인] - [경로] - ([결과])”의 큰 틀을 보인다. 이 틀에서 대상은 사건의 중요한 역할을 담당하는 [초점대상]으로 볼 수 있다. 또한 [시간]과 관련을 맺는다는 점에서 ‘변화’의 관점을 취하는 것으로 볼 수 있다.

관형격 조사 상당 구성의 다른 한 부류는 “[목적]-[수단]”의 틀을 보이며, 이에 관련된 대상은 [일반 대상]이다. [시간]과 관련되지 않는다는 점에서 ‘변화’의 관점을 취하지 않는 것으로 볼 수 있다.

4.2. 부사격조사 상당 구성의 유형 분류

(18)의 부사격조사 상당 구성의 군집 분석 결과를 살펴보겠다. 먼저, [-에

부처] 구성과 [-을 기해] 구성이 묶인다. [-에 부처] 구성은 ‘특정 시기’와 관련된 명사와 결합한다. ‘한글날, 식목일’ 등의 {시간} 명사나 ‘출간, 출판’ 등의 {사건} 명사와 결합한다. [-을 기해] 구성도 마찬가지이다. 특정한 날짜나 시간과 결합하거나 특정한 사건과 결합한다. 이 두 구성의 기능은 [기준 시기]라 할 수 있다. 이 두 구성과 가깝게 묶이는 것으로 [-에 대고] 구성이 있다. 이는 주로 {신체 부위}의 ‘뒤통수, 귀’ 또는 {공간}의 ‘허공, 안’ 등과 결합한다. 이 구성의 기능은 [접지점]·[방향]으로 볼 수 있다. [-에 걸쳐] 구성이 이들 군집과 결합하는데 이 구성은 {시간}의 ‘장기간, 세월, 세기, 기간’, {장소}의 ‘지역, 세계’ 등과 결합한다. 이 구성의 기능은 [범위]이다. 이 구성들은 [범위-시기]로 묶을 수 있다.

다음은 이 구성들의 구체적인 예이다.

- (33) 가. 한글날에 부처 쓰는 글
나. 식목일에 부처 삼가 씀
- (34) 가. 7. 17.을 기해 적대 행위를 중단할 것
나. 오후 1시를 기해 미세먼지 주의보 발령
- (35) 가. 뒤통수에 대고 큰 소리로 말했다.
나. 벽에 대고 소리라도 지르라.
- (36) 가. 오랜 세월에 걸쳐 만들어진 광경
나. 전 세계에 걸쳐 나타나는 현상이다.

[-에 비추어] 구성과 [-로 미루어] 구성이 하나로 묶인다. [-에 비추어] 구성은 주로 {추상}의 ‘경험, 상황, 현실, 사실’ 등의 명사와 결합한다. [-로 미루어] 구성도 마찬가지이다. {추상}의 ‘사실, 말, (~한)점’ 등과 결합한다. 이 구성들은 [의거점]·[기준]의 기능을 보이며 주로 [추론 기준]으로 기능한다. 다음으로 묶이는 [-을 전후해] 구성은 ‘금융위기, 폐전’ 등의 ‘특정 사건’을 변화의 기준으로 삼는 기능을 한다. [-을 놓고] 구성은 ‘투표 결과, 문제’ 등의 명사와 결합하는데 주로 ‘경쟁 사건에서 얻고자 하는’ [목표 대상]을 나타낸다. 이 구성들은 [추론-기준]으로 묶을 수 있다.

다음은 이 구성들의 구체적인 예이다.

- (37) 가. 경험에 비추어 판단했다.
나. 현실에 비추어 생각해 보자.
- (38) 가. 이 사실로 미루어 그 당시에 큰 폭발이 있었을 것이다.
나. 약기가 눈에 띄는 점으로 미루어 볼 때...
- (39) 가. 금융위기를 전후해 노동 구조의 변화가 일어났다.
나. 패전을 전후해 일본제국주의의 변화를 살펴보아야 한다.
- (40) 가. 투표 결과를 놓고 감론을박 하고 있다.
나. 이 문제를 놓고 고민하는 중이다.

[-로 말미암아] 구성과 [-로 인해] 구성은 모두 [원인]을 나타낸다. [-을 둘러싸고] 구성은 [초점 대상]을 나타내고 [-을 비롯해] 구성은 [대포]를 나타낸다. 이들은 모두 같은 구성의 관형격 조사 상당 구성과 유사한 체언들을 취하고 있다. 이들은 [원인-초점대상]으로 묶을 수 있을 것이다.

다음은 이 구성들의 구체적인 예이다.

- (41) 가. 한 사람의 범죄로 말미암아 사망이 들어왔나니...
나. 악성코드로 인해 발생하는 문제들
- (42) 가. 당신을 둘러싸고 있는 감정은 무엇인가?
나. 영국을 비롯해 전 세계에 살고 있는 스코틀랜드인들은...

[-에 관해], [-을 두고], [-을 보고] 구성들은 모두 [대상]을 나타내는 기능을 보인다. [-에 관해] 구성의 경우 [-에 관한] 구성과 마찬가지로 주로 추상적 대상을 나타내는 데 사용된다. [-을 보고] 구성의 경우는 발화 행위의 상대가 되는 대상을 나타내는 경우가 많았다. [-을 위해] 구성은 [-을 위한] 구성과 마찬가지로 [목적]을 나타낸다. [-을 향해] 구성의 경우 [목적]과 [방향]의 기능을 모두 나타낼 수 있다. [-에 관해], [-을 두고], [-을 보고] 구성을 [대상]으로 묶고, [-을 위해], [-을 향해] 구성을 [목적]으로 묶는다면 이 구성들 전체의 군집은 [대상-목적]으로 묶을 수 있을 것이다.

다음은 이 구성들의 구체적인 예이다.

- (43) 가. 불륜에 관해 다시 생각해 보기
 나. 그가 갑자기 사퇴한 것을 두고 안 좋은 소문이 떠돌았다.
 다. 철수가 날 보고 그렇게 말하더라고.
- (44) 가. 세상을 떠난 아빠가 아들을 위해 깜짝 생일 이벤트
 나. 내일을 향해 쫓라

[-에 의해] 구성은 [-에 의한] 구성과 마찬가지로 [수단]·[도구]·[원인] 등의 기능을 갖는다. [-을 통해] 구성은 [경로]·[방식]의 의미를 갖는다. 이 두 구성은 [수단]이라는 기능으로 하나로 묶을 수 있다. [-에 대해] 구성은 [-에 대한] 구성과 같이 보편적인 [대상]에 두루 사용된다. 이 군집은 [대상-수단]으로 묶을 수 있을 것이다.

다음은 이 구성들의 구체적인 예이다.

- (45) 가. 인간, 기술에 의해 멸망할 것인가, 신이 될 것인가.
 나. 100일 간의 거절을 통해 배운 것들
- (46) 가. 남대문에 대해 조금 더 알아본 사실들
 나. 인공지능에 대해 우리가 간과한 것들

[-에도 불구하고] 구성은 {현상}의 ‘호조, 부진’, {사건}의 ‘둔화, 증가’, {추상}의 ‘비판, 한계, 반대’ 등의 명사와 결합한다. [-에도 불구하고] 구성은 선행하는 명사가 나타내는 사태를 기준으로 그와 반대되는 또는 기대하지 않은 결과가 일어남을 나타내는 기능을 한다. [-에 따라] 구성은 [-에 따른] 구성과 마찬가지로 [의거점]·[기준] 등의 기능을 한다. 이 두 구성의 군집은 [기준]이라는 기능을 담당하고 있다. [-을 가지고] 구성은 [도구]·[수단]의 기능을 한다. 이들 구성들이 이루는 군집은 [기준-도구]라 할 수 있다.

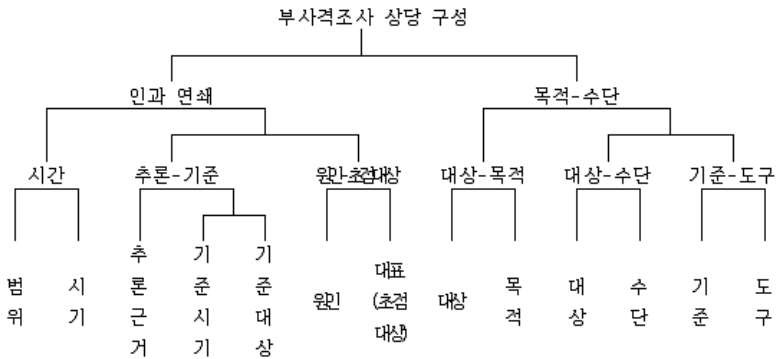
다음은 이 구성들의 구체적인 예이다.

- (47) 가. 고용지표 부진에도 불구하고 미 경기 상승 기조는 유효
 나. 이러한 한계에도 불구하고 기독교 교육의 새로운 기회가 되고 있다.

- (48) 가. 상황에 따라 다르다.
 나. 기분에 따라 색다르게 연출할 수 있다.
- (49) 가. 스마트 물질을 갖고 재미있게 놀아라.
 나. 망치를 갖고 나무를 두드렸다.

위의 논의를 바탕으로 부사격조사 상당 구성의 유형을 분류하면 다음과 같다.

(50) 부사격조사 상당 구성의 유형 분류



(49)의 부사격조사 상당 구성 유형 분류는 몇 가지 점에서 (31)의 관형격 조사 상당 구성의 유형 분류와 차이를 보인다. 우선, [대상]이 여럿으로 구별된다. 관형격조사 상당 구성의 경우 [일반대상-[수단-목적]]으로 군집을 이룬 반면, 부사격조사 상당 구성의 경우, [대상-수단], [대상-목적]이 구분된다. [기준] 역시 [추론-기준]과 [기준-도구]로 구분된다. 관형격 조사 상당 구성보다 좀 더 복잡한 양상을 확인할 수 있다.

부사격 상당 구성 역시 크게 두 부류로 구분된다. 첫째, “[원인] - [추론] - ([결과])”의 틀이 그것이다. 이는 관형격 조사 상당 구성의 “[원인] - [경로] - ([결과])”의 틀과 유사한데, 이에서 [경로]와 관련된 구성이 빠지고

[추론]과 관련된 구성이 더해진 것으로 보인다. [인과 연쇄]의 [초점대상]은 [원인]과 더 가깝게 묶여 있음을 알 수 있다. 이 부류 역시 [시간]과 관련되며 변화의 관점을 취하는 것으로 볼 수 있다.

부사격 상당 구성의 또 다른 부류는 “[목적] - [수단] - [도구]”의 틀로 볼 수 있을 것이다. 이 역시 관형격 조사 상당 구성의 “[목적]-[수단]”의 틀과 유사하다. 다만 [도구]와 관련된 구성이 좀더 세분되어 나타난다.²⁰⁾

5. 결론

본고는 한국어의 조사 상당 구성들을 좀 더 세밀하게 분류하고 이를 통해 조사 상당 구성의 전체 체계를 조망하는 것을 목적으로 논의를 진행하였다. 이를 위해, 40여 개의 조사 상당 구성들 각각의 선행 요소들을 말뭉치로부터 추출하여 그 선행 요소와의 결합 강도를 Z-score를 통해 확인하였다. 그리고 그 결과를 바탕으로 40여 개의 조사 상당 구성들을 군집 분석하여 통계 분석에 기반한 조사 상당 구성의 미시적 분류를 시도하였다.

통계 처리 기법에 있어서도 본고는 기존 논의의 단점들을 보완하였다. 첫째, 상대확률비를 내는 방식을 수정하여 조사 상당 구성과 관련된 확률 분포가 좀더 분명하게 드러나는 방식을 고안하였다. 둘째, Z-score 수식을 수정하여 관찰값뿐 아니라 기댓값에 대해서도 표준오차를 반영하도록 하였다. 이를 통해 수식의 분모의 값을 줄여 Z-score 값이 통계적으로 이해할 수 있는 수준으로 실현되도록 보정하였다. 셋째, Z-score의 음(-)의 값을 적극적으로 해석하여 관련 구성의 문법적 기능을 확인하는 데 적용하였다.

조사 상당 구성들의 군집 분석 결과 이들은 [인과 연쇄] 관련 구성과 [목

20) 전체적으로 관형격조사 상당 구성들보다 부사격조사 상당 구성들이 좀더 세분되는데 이는 구성의 전체 수가 관형격조사 상당 구성들보다 부사격조사 상당 구성들이 더 많기 때문인 것으로 볼 수 있다.

적-수단] 관련 구성으로 크게 구분되었다. [인과 연쇄] 관련 구성들은 [시간], [원인], [경로], [추론] 등으로 구성되며 변화의 관점을 취하며 [초점대상]과 관련된다. [목적-수단] 관련 구성들은 [목적], [수단], [도구] 등으로 구성되며 변화의 관점을 취하지 않으며 [일반 대상]과 관련되었다. [인과 연쇄]의 [결과]는 조사 상당 구성으로 실현되지 않았는데 이는 해당 구성의 수식을 받든 명사구가 주로 담당하였다.

군집 분석의 결과가 우리의 언어 직관과 일치하지 않는 경우가 더러 발생하였는데 사용 빈도가 낮은 구성들이 주로 이러한 현상을 보였다. 특히 문어 체에는 잘 사용되지 않는 구어적인 구성들이 그러했는데 이는 본고에서 사용한 말뭉치의 한계로 볼 수 있다. 말뭉치의 크기를 키우고 구어 말뭉치를 구축하여 보완하는 것이 근본적인 해결책으로 보이지만 통계 기법을 좀 더 정밀하게 구성하여 이를 보완하는 시도도 필요해 보인다.

참고 문헌

- 문병열(2014), 「어휘부 등재소와 복합 구성」, 『국어학』 제69집, 국어학회, 135-166.
- 문병열(2015), 「한국어 조사 상당 구성에 대한 연구」, 서울대학교 박사학위 논문.
- 박병선(2005), 『한국어 계량적 연구 방법론』, 도서출판 역락.
- 박진호(2003), 「한국어의 동사와 문법요소의 결합 양상」, 서울대학교 박사학위 논문.
- 서경숙(2004), 「현대국어 조사 상당어에 대한 연구」, 서울대 석사학위 논문.
- 신중진(2007), 『개화기국어의 명사 어휘 연구』, 태학사.
- 이선웅(2004/2005), 『국어 명사의 논항구조 연구』, 월인.
- 이영제(2014), 「한국어 기능명사 연구」, 고려대학교 박사학위논문.
- 이현우(1995), 「현대국어 명사구의 구조 연구」, 서울대학교 박사학위논문.
- 이희자(1995), 「현대 국어 관용구의 결합 관계 고찰」, 『대동문화연구』 30, 성균관대학교 대동문화연구소, 411-444.
- 임근석(2005), 「문법적 언어의 개념 정립을 위하여」, 『형태론』 7-2, 박이정, 277-301.
- 임근석(2006), 「한국어 언어 연구」, 서울대학교 박사학위 논문.
- 임근석(2009), 「문법적 언어와 한국어 교육 -조사적 언어를 중심으로-」, 『한국어교육』 20, 국제한국어교육학회, 161-184.
- 정희정(2000), 『한국어 명사 연구』, 한국문화사.
- 최경봉(1998), 『국어 명사의 의미 연구』, 태학사.
- Dunning, T.(1993), 'Accurate Method for the Statistics of Surprise and Coincidence', Computational Linguistics, 19(1), pp. 61-74.
- Martin, S. E.(1992), A Reference Grammar of Korean, Charles E. Tuttle Company, Inc.

【Abstracts】

A Study on Classification of constructions corresponding to particles in Korean

Moon Byeong-yeol

The purpose of this paper is to classify the constructions corresponding to particles(CCPs.) in Korean more closely and to view the entire system of the CCPs. Many of the items in the CCPs have very similar meanings and functions and coexist with subtle differences in meaning. Although a macroscopic classification system based on grammatical functions is used, a micro classification system based on the subtle difference in meaning of the individual constructions should also be established.

Therefore, this paper attempts to analyze the hierarchical agglomerative clustering based on the extraction of the binding patterns of the individual CCPs related items. As a result, they are largely classified into [causal chain] and [purpose-means] and the [causal chain] constructions were classified into [Time], [Cause], [Path], [Reason]. And the [Purpose - Means] constructions were classified as [purpose], [means], [tool].

Key words: constructions corresponding to particles, micro classification system, hierarchical agglomerative clustering

이 논문은 2017년 10월 1일에 투고되었으며, 2017년 11월 9일에 심사 완료되어 2017년 11월 10일에 게재가 확정되었음.