

말뭉치 기반 한국어교육용 N+N 복합 구성 선정의 쟁점 연구

권민수 · 강현화

— 차례 —

- | | |
|----------------------|------------------|
| 1. 서론 | 4. 교육용 목록 선정의 쟁점 |
| 2. 선행 연구 검토 | 5. 결론 |
| 3. 말뭉치 및 분석
범위 설정 | |

〈버리〉

본 연구는 한국어 모어 화자의 말뭉치에 나타난 N+N 구성을 활용하여 한국어교육용 명사구 복합 구성을 선정하는 과정에서 논의되어야 할 쟁점에 대해 고찰하는 것을 목적으로 한다. 이를 위해 국립국어원 모두의말뭉치를 활용하여 여러 하위 장르의 말뭉치에서 추출한 N+N 구성의 활용성을 검토하고, 설정된 분석 범위 내 항목을 교육 목적의 어휘로 선정할 때 고려할 사항을 논의하였다. 먼저 문어, 구어성 문어, 구어 속성의 하위 말뭉치에서 검색되는 N+N 구성의 특성을 살펴본 결과 일상대화, 구어, 신문 장르 말뭉치를 활용하는 것이 합당하다는 판단을 내렸다. 다음으로, 세 말뭉치의 범위 내에서 추출된 N+N 복합 구성을 후보로 취하여 그중 교육용 목록에서 제외해야 할 유형과 교육용 구성으로 선정 시 고려해야 하는 쟁점을 유형별로 논의하였다. 제외 기준으로는 결합의 우연성, 명칭의 고유성, 의미적 투명성, 학습 필요성 및 적합성을 다루고, 선정 항목에 대한 쟁점으로는 단위 설정, 구성 요소의 특성, 의미 관계 유형, 교육 활용성을 주제로 논하였다.

주제어: 말뭉치 기반 연구, 명사구, 복합 구성, 어휘 교육, 어휘 목록 선정

1. 서론

한국어에는 ‘고속 도로’, ‘교육 문법’과 같이 특정 명사끼리 자주 함께 어울려 하나의 명사와 같은 개념을 지시하면서도 그 의미가 각 명사 의미의 온전한 결합만으로는 설명되지 않는 경우가 있다. 이렇듯 자주 결합하는 명사 복합 형식은 비단 한국어에서만이 아니라 다른 언어에도 자주 나타나서, 언어 교육에서는 이러한 다단어 구성에 대한 관심이 높다. 언어 간 대당어는 복합 구성이 한 단어로 번역이 이루어지거나 구성 요소의 품사가 달라지기도 하다. ‘고속 도로’는 영어의 ‘expressway’에 해당되는데, ‘형용사+명사’ 구조로 구성 요소의 품사에 차이를 보인다. 이렇게 어휘의 조합이 언어마다 달리 나타난다는 사실은, 학습자들에게는 학습의 부담과 어려움을 유발하는 요소로 작용할 수 있다. 특정 어휘의 결합 여부와 결합으로 발생하는 의미를 학습해야 하기 때문이다.

그간의 연구에서 두 개 이상의 어휘가 어울려 하나의 구성으로 나타나는 단위를 복합구 또는 복합 구성이라 일컫는데, 강현화(2015)에서는 언어 교육에서 다룰 만한 복합구의 유형으로 다음의 세 가지를 설명하였다.

(1) 강현화(2015)의 복합구의 세 유형

- 가. 사전에 등재되어 있는 구 표제어와 유사 범주에 해당하나 아직 미등재된 구
- 나. 구로서 가지는 의미와 별도로 독립된 단어 수준으로 사용되는 구
- 다. 단순히 공기 빈도가 높아서 한 덩어리로 인식되는 구

(1가)에서는 사전 등재어와 유사한 수준의 구성, (1나)에서는 별도의 특정 의미를 가지는 구성을 복합구로 보고 있다. 이에 따라 그 자체로 하나의 단위로서의 지위를 가질 정도로 하나의 개념처럼 사용되는 구성을 일컬어 복합구(또는 복합 구성)라고 함을 알 수 있다. 이렇게 하나의 단위로서 나타나는 구성은 공기 빈도가 매우 높아서 하나의 덩어리로 인식된다는 (1다)의 기술과도 맞닿아 있으며, 이에 하나의 확장된 어휘로서의 구 단위 결합을 복합 구성으로 볼 수 있다. 같은 명사 연속 조합이라도 한국어에서 ‘남자 선배’와 ‘남자 친구’는 그 결합 강도와 의미의 측면에서 전혀 다른 결합 양상을 보이며, 일부 학습자에게는 ‘아침 식사, 저녁 식사’ 등의 한국어 명사구가 자신의 모국어에서 한 어절의 단어 단위(예를 들어, 영어로 breakfast, dinner)로 인식될 수 있다. 아침에 먹는 밥을 칭할 때 한국어에서는 단일어 ‘아침’, 합성어 ‘아침밥’, 구 구성 ‘아침 식사’의 세 가지 단위로 나타나는 가운데 독일어 ‘Frühstück’이나 일본어 ‘朝ごはん’은 두 단어가 한 단위로 결합한 합성어로 나타난다. 이러한 점에서 명사구 복합 구성은 언어마다 달리 나타남을 볼 수 있다. 이러한 명사구 복합 구성은 일상적으로 접하는 다양한 텍스트에서 두루 나타나기 때문에 그 형태·통사적 특성과 의미 관계를 아울러 교육적으로 주목할 만한 단위가 된다.

N+N 구성은 하나의 개념을 지시하거나 하나의 단위로 자주 사용되는 유의미한 어휘 항목임에도 불구하고 그동안 기초적인 한국어의 교육용 목록 선정 대상으로는 조명받은 바가 적다. 학습자들의 학습 부담과 필요성을 고려하면 기본적으로 알아야 하는 명사구 항목의 목록 선정의 기준과 추출 과정에 대한 고민이 이루어져야 한다.

이에 본 연구에서는 유의미한 표현 단위로서의 명사구 복합구의 특성에 주목하여 교육용 목록의 선정 과정에서 논의될 필요가 있

는 쟁점을 구체적으로 고찰해 보고자 한다. 특히 본 논의에서는 한국어 모어 화자들의 말뭉치를 자료로 하여 학습 과정에서 기초적으로 다루어져야 할 N+N 복합 구성의 선정 과정과 기준에 관해 논의하고자 한다. 이미 사전이나 교재 등에서 선별 및 정제되어 제시되는 구 단위 표현에 한정하지 않고, 말뭉치를 기반으로 분석한다면 보다 실제적인 맥락에서 일상적으로 빈번하게 사용되는 단위들을 파악할 수 있을 것이다.

이에, 한국어 모어 화자들이 생산한 여러 장르의 말뭉치를 넓게 검토하여, 한국어 N+N 구성의 유용성과 사용 맥락의 다양성을 아우르는 적합한 대상 말뭉치를 선별하여 목록을 구성하고자 한다. 다음으로, 대상 목록에 출현한 N+N 항목들을 하나씩 검토하며 교육용 목록의 항목으로서의 활용 가능성을 점검하고자 한다. 이 과정에서의 선정의 쟁점을 살펴보는 일은 학습 목적에 타당하고 유의미한 N+N 복합 구성을 선정하여 교육용 기본 목록을 구축하는 데 필수적이라고 여겨진다. 그간의 연구들은 목록을 제시하는 데에 중점을 두고, 목록 선정의 과정과 쟁점에 대해서는 상대적으로 논의가 적었음에 주목하고, 본고는 선정의 쟁점에 대해 집중적으로 논의하고자 한다.

2. 선행 연구 검토

명사로 이루어진 복합 구성은 둘 이상의 명사가 결합하여 하나의 단위로 사용된다는 점에서 합성명사와의 관련성 또한 살펴보게 한다. 합성명사와 명사구는 형성 방식이 동일하거나 유사한 경우가 많다(김성환 2004). 박소영, 무아흠(2024)에서는 전통적으로 형태적 구성인 합성명사와 통사적 구성인 구의 구분 기준으로 통사론적 비

분리성(inseparability), 의미 해석의 비합성성(non-compositionality) 등이 거론되었으며, 합성명사의 중요한 특징으로 병치된 명사 의미의 단순 합을 뛰어넘는 특수한 의미로 해석될 수 있다는 점을 언급하였다. 이찬영(2024)에서는 구성 요소의 의미로써 그 전체 의미를 직접적으로 도출해 낼 수 없는 경우가 많은 합성명사의 관계적 의미를 구 단위에서도 구체적으로 살펴야 함을 언급하였다.

이처럼 결합된 전체 단위와 그것을 이루는 각 구성 요소의 관계는 단어 단위를 넘어 명사 연속 구성의 경우에도 확장하여 살펴볼 만하다. 어휘 단위를 어휘적 단어와 어휘적 구성으로 구분한 오규환(2016)에서는 어휘적 구성 안에 복합 구성(complex construction)을 분류하였다. 해당 연구에서는 이러한 복합 구성이 분리 가능성을 가지며, 의미 합성성의 준수 여부 및 확장 가능성이 경우에 따라 다른 것으로 보았다. 명사가 연속되어 하나의 단위가 된 구성은 형태적으로는 단순 병치되어 보일지라도 두 명사 결합 의미의 합성성과 확장 가능성 면에서 다양하게 드러날 수 있기에 단순 합성으로서의 의미를 뛰어넘는 상세한 분석이 요구된다고 하겠다.

한국어교육 분야에서는 교육적 목적을 가지고 정제된 사전이나 교재 등에 나타난 어휘를 기반으로 N+N 구성이 다루어진 측면이 있다. 손연정, 강현화(2023)에서는 한국어 학습사전인 한국어기초사전에서의 구 단위 표제어 선정 방안을 논의하였다. 한국어 교육 분야에서 ‘표현 문형’으로 불리는 문법적 구 단위, 의미적 은유를 가지는 관용어 구 단위, 담화 표지적 구 단위와 더불어 어휘적 구 단위의 유형을 분류하였다. 어휘적 구 단위에는 의미적 합성성이 약하여 A+B=C가 되는 ‘가을 야구’, ‘시간 싸움’ 등이 있고, 사용의 범위 및 빈도 변화를 보인 ‘인터넷 예매’, ‘문자 메시지’, ‘천연 조미료’ 등이, 신조어 ‘코로나 세대’, ‘줌 회의’, ‘N의 N’ 꼴인 ‘국군의 날’, ‘삶의 질’ 등이 포함되었다. 임상은, 강현화(2017)에서는 한국

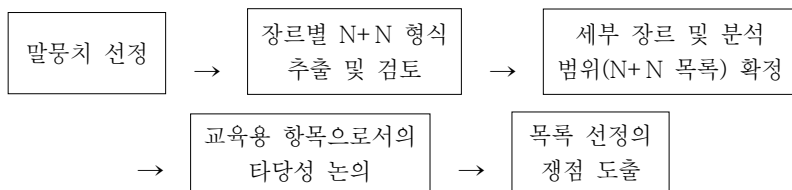
어 명사 복합 구성의 유형을 의미 관계, 의미적 투명성의 기준에 따라 분류하였다. 이들 연구는 기존 한국어 교육 자료에서 나타난 명사 복합 구성을 유형화하여 밝혔다는 점에서 의미가 있으나, 표현과 장르의 다양성이 반영된 실제 발화된 언어 자료를 바탕으로 하지 않은 한계가 있다.

이론적 접근의 연구 외에도 한국어 교육에의 구체적인 활용될 가능성까지 염두에 둔 말뭉치의 장르와 빈도 기반의 연구가 이루어졌다. 강현화(2016)에서는 말뭉치에서의 빈도는 실제 언어 사용에서의 입력을 의미하며, 이는 실제 학습자들에게 노출될 가능성을 의미한다고 하였다. 강현화, 권민수(2025)에서는 세 가지 장르의 말뭉치 자료에서 N+N 구성 중 N1과 N2가 각각 생산성이 높은 경우를 추출하여 장르별 특성이 있음을 보였다. 강현화(2025)에서는 기존 사전과 한국어교육용 교재의 구 단위 목록을 추출한 뒤 코퍼스 분석으로 기존 목록의 빈도를 점검하며 ‘N+N’ 구성 선정에 대한 기초 분석을 진행한 바 있다. 분석 결과, 고빈도 순위의 목록이라도 장르 간의 중복도는 낮게 나타나거나, 빈도는 높지 않은 편이지만 장르 간 중복도는 매우 높은 항목들도 발견된다고 하였다. 해당 연구에서는 고빈도임에도 의미적 합성성이 크다면 교육 목록으로 타당하지 않을 수 있으며, 한국어 교육과정과의 주제 관련성도 중요하다고 지적하였다. 또한, 구교빈, 강현화(2025)는 한국어 학습 사전에서의 효율적인 명사구 표제어 선정을 위해 말뭉치 자료를 활용하였는데, 사전 간 중복도, 말뭉치의 빈도 분석, 전문가 평정을 통해 한국어교육과 관련되는 사전 등재어 삭제 및 추가 목록을 제안하였다. 이상의 연구는 말뭉치에서 추출한 명사구의 빈도를 절대적 기준으로 취하기는 어려우며 명사 2개가 하나의 단위를 이룰 때 구성 요소 각각의 특성도 살펴볼 필요가 있음을 시사한다. 그러나 이들 연구 모두는 구체적인 목록의 제시나 목록

선정의 쟁점을 다루지 못한 한계를 보인다.

한편, 그간 어휘 목록을 선정한 다수의 연구들은 교재나 사전 등제어를 기준으로 삼은 경우가 많다. 권우진(2022)에서는 <표준국어대사전>의 전체 표제어 중 식감 관련 형용사를 추출해 총 505개 어휘 목록을 선정하고 그 의미적, 조어적 특징을 분석하였다. 기존의 교육용 어휘 목록의 비교 연구도 있는데, 이유경(2021a)에서는 <한국어 학습용 기초 어휘 선정>, <한국어교육 어휘 내용 개발>, <국제 통용 한국어 표준 교육과정>의 명사 어휘를 비교하여 목록 간 차이를 분석하고, 이러한 차이는 선정 방식의 차이와 관련이 있다고 보았다. 이러한 지적을 통해 어휘 목록의 사용 목적에 따라 목록이 가변적일 수 있으며, 목적에 부합하는 기준의 제시와 정비가 필요함을 알 수 있다. 이유경(2021b)에서는 국가 주도로 이루어진 어휘 목록 선정 연구를 성과와 보완 방식을 중심으로 검토하면서, 1990년대에서 2000년대 초반 시점의 말뭉치는 시의성이 떨어진다는 점과 어휘 목록 선정 시 다양해진 학습자 변인을 반영하지 못함을 논하였다. 이는 어휘 목록을 개발할 때 최신 언어 사용 자료와 활용의 주체를 고려한 접근이 필요함을 시사한다.

이상의 선행 연구를 종합하면 N+N 복합 구성은 그 구성과 단위에 주목하여 형식적, 의미적 측면의 결합 관계를 살펴볼 가치가 있으며, 실제성 있는 자료를 기반으로 한 선정, 그리고 목표에 부합한 선정의 논의가 필요함을 확인하였다. 이러한 시사점을 종합하여 본 연구는 다음과 같은 방식과 절차를 통해 목록 선정의 쟁점에 대해 논의하고자 한다.



3. 말뭉치 및 분석 범위 설정

3.1. 말뭉치 및 범위 설정 방법

본고에서는 교육용 복합 구성 목록을 추출하는 데 활용할 말뭉치 자료로 국립국어원 ‘모두의 말뭉치’를 대상으로 삼고자 한다. 현행 제공되는 말뭉치 중 장르와 규모가 가장 큰 말뭉치이기 때문이다. 본 연구에서는 한국어 모어 화자의 일상적인 언어 사용에서 관찰되는 N+N 복합 구성을 추출하기 위해 한국어 문어, 구어, 구어성 문어 유형의 특징적인 하위 장르를 선정하였다. 구체적으로는 문어 영역의 하위 장르로 문어, 신문, 비출판물, 온라인 게시 자료 말뭉치를, 구어성 문어 영역의 하위 장르로 온라인 대화, 메신저 말뭉치를, 구어 영역의 하위 장르로 일상대화, 구어 말뭉치를 대상으로 삼았다. 먼저 이들 말뭉치 각각에서 형태를 기준으로 명사 두 개가 연속된 N+N 단위만을 추출하였다. 말뭉치별로 추출된 N+N 구성의 전체 개수와 장르의 특징은 <표 1>에서 확인할 수 있다.

장르 유형		말뭉치	N+N	비고	
문어	문어	저작물 10,045건	344,824	책, 잡지, 보고서	
	신문 2022	신문 기사 978,342건	510,339		
	비출판물	파일 10,753개	1,083	시, 동시, 일기, 편지, 소설, 감상문	
	온라인 게시 자료	305,174건	47,818	온라인 카페 등 게시판 및 누리소통망	
구어성 문어	온라인 대화	74,665건	11,961		
	메신저	3,836건	1,878		
구어	사적	일상대화 2021	4,143건	3,261	
	공적	구어	파일 25,696개	105,060	공적 독백, 공적 대화, 준구어

<표 1> 말뭉치 장르별 추출 N+N 구성

N+N 항목들의 분석 범위 설정은 다음의 절차로 진행하였다. 먼저, 각 장르 내에서 추출된 N+N 항목을 빈도순으로 나열하고, 교육적 유용성 및 필요성을 고빈도 출현 항목부터 검토하였다. 즉, 출현 빈도를 우선적 기준으로 나열하되 장르별 유용성의 밀도를 살폈다¹⁾. 의미 없이 우연히 나열된 N+N이 어느 지점부터 많아지는지, 혹은 저빈도이어도 유의미한 구성은 어느 지점까지 발견되는지를 검토하였다. 이러한 검토는 단순히 빈도에만 의존하지 않고, 복합 구성으로서의 가능성이 있는 목록(의미적 투명성이 약하고

1) 유용성이란 말뭉치의 일정 범위에서 유효하며 유의미한 복합 구성을 의미한다. 가령 빈도 순으로 나열한 N+N 항목들 중 의미적 투명성이 낮고 결합 관계 빈도가 긴밀한 것이 더 유용하다고 보았다. 예를 들면 ‘고속도로, 가을 야구’는 ‘가정 자녀, 저녁 시간’보다 유용하다.

한 개념을 지시하는 구성)을 파악하기 위함이다. 이러한 과정을 통해 선정된 장르별로 유의미한 목록을 정리하였다.

3.2. 말뭉치 및 범위 설정 결과

말뭉치는 구어, 구어성 문어, 문어의 세 가지 유형으로 구분하고 각각의 하위 장르별 말뭉치를 검토하였다. 검토 과정에 따른 결과는 다음과 같다. 첫째, 구어 장르로는 일상대화(DIALOGUE, 2021)와 구어(SPOKEN)를 살펴보았다. 일상대화 말뭉치에서는 빈도 70 이하에서 제외할 항목들이 꽤 발견되기 시작한 한편, 빈도 70 이하이지만 ‘전기 요금’, ‘인간 관계’, ‘개인 정보’, ‘사실 관계’(26회), ‘기말 고사, 회전 초밥’(24회), ‘육아 휴직, 물가 상승, 특수 학교’(23회)와 같은 유의미한 항목들이 발견되었다. 이에 해당 장르를 유용하다고 보아 분석 범위에 포함하였다. 구어(SPOKEN) 말뭉치는 방송 등에서 사용할 법한 N+N형 전문 어휘를 다수 포함하고 있었고, ‘인증 샷’, ‘티비 토론’, ‘반대 의견’ 등 장르를 대표하는 항목도 출현하였다. 500회 나타난 ‘택배 기사’를 기점으로 그 이하의 빈도 목록은 대체로 단순한 자유 결합의 구로 판명되었다.

둘째, 구어성 문어 성격의 하위 장르로는 온라인 대화(Online Text Message, 2021)와 메신저(Messenger)를 살펴보았다. 온라인 대화 말뭉치의 경우 상대적으로 한국어의 이해 및 학습에 기본적인 도움이 될 것으로 보이는 항목이 적었다. 해당 말뭉치에서 추출한 전체 고빈도 N+N 구성의 상위 0.1% 내에 빈번히 나타난 ‘가을 옷’, ‘오늘 점심’ 등 단순 나열된 구성, ‘당근마켓’, ‘아이패드’와 같은 특정 서비스나 사물의 명칭, 또는 ‘긋긋’, ‘알겠’, ‘꿀팁’, ‘조아조양’처럼 매체 특성을 반영하는 신어 구성은 교육용 복합 구성으로서의 유용성이 떨어진다. 해당 말뭉치에서 추출된 N+N

후보군은 총 11,961개로 그 규모가 작지 않음에도 불구하고, 자료에서 추출한 N+N 항목들은 한국어 교육용 어휘로서의 활용성이 충분치 않다고 판단되어 분석 대상 범위 내 포함이 필수적이지 않았다. 메신저 말뭉치는 1,878개로 그 규모가 매우 작았는데, 유의미하게 활용될 만한 메신저 장르만의 특별한 구성은 눈에 띄지 않았다. ‘카페 알바, 홈 트레이닝, 생일 파티’(21회), ‘근무 시간, 기차 여행, 다이어트 도시락, 기분 전환, 전업 주부, 문화 센터’(20회), ‘맛집 투어, 할인 쿠폰, 안마 의자, 혈액 순환’(19회), ‘인공 눈물’(18회) 등이 나타났는데, 이러한 목록은 일상대화 말뭉치에서도 출현 가능성이 있는 것들이다²⁾. 이에 규모가 크다는 점과 목록 포함성이 있다는 점에서, 일상대화 말뭉치를 분석 대상으로 삼았다.

셋째, 문어 장르를 검토하였다. 문어 장르는 큰 규모에 비해, 유용한 항목이 밀도 있게 나타나지 않았다. 신문(Newspaper, 2022), 온라인 게시 자료(Online Posting Materials, 2022), 문어(Written), 비출판물(Non-Publication) 각각에서 N+N 복합 구성 후보군을 살펴본 결과, 신문 말뭉치를 분석 범위에 포함하는 것이 적절해 보였다. 다만, 신문 말뭉치에는 정치, 경제 관련 유용한 구 단위가 많으나 한국어교육 목표의 우선순위에 들기 어려운 전문 용어가 다수를 이루어 선별 과정이 필요해 보였다. 이에 고빈도이면서 유의미한 N+N 구성의 범위를 한정하여 분석 대상에 포함하였다. 온라인 게시 자료의 경우 우연히 인접한 명사들로 구성되어 복합 구성의 자격에 미치지 못하는 항목도 많이 드러났으며, ‘N 추천’, ‘N 맛집’, ‘N 맘’ 등 대체로 사용 영역과 연령층이 한정되어 있는 성격의

2) 실제 일상대화 말뭉치에서는 ‘근무 시간(23회), 기분 전환(24회), 할인 쿠폰(26회), 생일 파티(27회), 혈액 순환(39회), 기차 여행(72회)’이 발견되었으며, 나머지 ‘홈 트레이닝’, ‘다이어트 도시락’ 등은 교육적 필요성 측면에서 우선되지 않으므로 나타나지 않아도 무방하다.

표현들이 발견되었다. 비출판물 말뭉치는 시, 동시, 소설, 일기 등으로 구성되어 추출된 어휘가 학습용으로는 필수적이지 않다고 판단하였다. 한편, 문어 말뭉치에는 공적 방송, 강의 담화 등을 포함하는 구어 말뭉치와 중복되는 항목들도 다수 있어서, 문어와 구어에 모두 사용된 항목에 주목할 수 있었다.

이상의 검토 결과를 바탕으로, N+N 복합 구성 예비 목록 추출을 위한 분석 범위에 포함한 말뭉치 하위 장르는 일상대화(2021), 구어, 신문(2022)이다. 이들 말뭉치에서 추출한 N+N 후보 목록의 규모에는 상당한 차이가 있었으므로, 각 장르 말뭉치에서 추출한 N+N의 출현 빈도를 기준으로 하되, 장르별로 유의미한 구성은 부가적으로 포함하여 범위를 한정하였다. 따라서 빈도 기준은 위에서 살펴본 바와 같이 기본 빈도 500 내외로 포함되는 N+N 구성의 유용성을 살펴, 각 장르의 자료에서 복합 구성으로서의 유용성이 밀도 있게 나타나도록 설정하였다. 구체적으로 대상 범위를 조합해 본 결과, 적합하다고 판단된 말뭉치 장르별 N+N 구성을 보이면 <표 2>와 같다.

말뭉치 장르	포함 범위	N+N 개수 (개)
일상대화 2021	전체	3,261
신문 2022	빈도 619 이상	11,560
구어	빈도 500 이상	1,222
전체		16,043
중복 제거 후		14,463

<표 2> 분석 범위 선정 결과

각 장르의 분석 범위 내 N+N 항목에서 중복되는 항목을 제거한 결과, 총 14,463개의 N+N 형식의 분석 대상이 설정되었다. 다음

장에서는 한국어교육용 복합 구성으로서의 적합성과 필요성을 기준으로 이들 항목에 대해 제외 구성과 유의미한 구성을 분류하고 목록 선정 시의 쟁점을 논의하겠다.

4. 교육용 목록 선정의 쟁점

4.1. 제외 기준에 대한 논의

무엇을 한국어교육용 목록으로 삼을 것인가는 교육의 목표와 연계된다. 본고는 범용적 한국어교육에 초점을 두면서 구체적인 목록 선정에서 고려해야 할 요인들을 차례로 고찰하고자 했다.

우선, 대상 목록은 N-gram을 기준으로 단순히 빈도순으로 추출한 목록이므로 정제 작업이 필수적이다. 이에 단순 결합 목록과 한국어의 기본 명사구 복합 구성으로서의 교육적 함의가 부족한 구성들을 제외할 필요가 있다.³⁾

다음으로, 제한된 교육과정을 고려할 때, 모든 목록을 필수 목록으로 제공할 수 없다는 점에서 이해와 산출에 필수적인 구성에 우선적으로 초점을 두었다. (2)는 교육용 ‘복합 구성’의 단위에서 제외해도 좋다고 판단한 구체적인 항목의 유형이다.

(2) N+N 복합 구성 목록 제외 기준 및 예시

가. 특정 인명, 구체적 명칭: (예) 문 대통령, 국 힘, 토스 증권, 텍스, 코스트 코, 미디어 오늘, 샤오 미, 페이스 북

나. 당시 부서명 또는 관직명, 혹은 지명으로 사용되던 단위나 그 일부: (예)부 장관, 국토 교통, 국토 부, 복지 부

3) 이에 유의미한 한 개념을 이루지 못하고 우연한 공기로 단순 인접한 어절들은 정제하였다.: (예) 힘 대선, 티 콘, 런 장관, 링 공간, 타 차

- 다. 의미적 투명성이 다소 높아 단순히 두 명사의 명제적 의미의 결합으로 쉽게 이해할 수 있는 경우: (예) 사업 추진, 감염 경로, 아파트 값, 외국인 근로자, 외국인 유학생
- 라. 학계 및 정치, 사회, 경제, 과학 등 각 분야의 전문 용어로서 사용 범위가 제한적이라고 판단한 경우: (예) 원내 대표, 투자 증권, 가상 자산, 심근 경색, 전담 병상, 발광 다이오드, 배임 혐의, 헤지 펀드, 사전 청약, 추가 경쟁
- 마. 비속어 또는 차별 표현에 해당하여 교육적 적합성이 고려되는 경우: (예) 발달 장애인, 진상 손님, 노래 주점
- 바. 고빈도 구 단위가 축약되어 신어처럼 사용되는 경우: (예)취준, 학 폭

(2가)는 인명이나 지명이 사용된 것으로 특정성이 있다. (2나)의 부서명의 경우 주기적으로 바뀌므로 주로 참고 어휘로만 사용된다고 판단하였다. (2다)는 의미가 투명하여 두 명사의 명제적 의미 결합만으로 이해할 수 있는 구성으로, 별도의 목록을 제공할 필요는 없다고 보았다. 다만, 한 단위로 자주 사용되더라도 (2바)의 ‘취준’, ‘학폭’처럼 음절이 축약되는 경우에는 포함하지 않았다. ‘학교 폭력’과 같은 축약 이전의 단위를 목록으로 삼는 게 더 적절하다고 판단하였다. (2라)은 다소 전문적인 구성으로 학문 목적 등의 학습자에게는 필요할 수 있으나, 일반 학습용 목록에서는 제외해도 좋다고 판단하였다. 또한 실제적인 교육적 활용을 염두에 둔다면 (2마)의 비속어 또는 차별 표현 등은 우선적으로 목표 어휘로 삼기는 어렵다.

본고는 교육의 지향점에서 벗어난 비속어 및 차별 표현, 그리고 이미 한 단어 단위로 굳어진 축약형 신어는 교육용 복합 구성 후보에서 제외할 수 있다고 보았다. 명사 두 개의 연쇄 구성 중, 공기우연성이 높거나 투명성이 지극히 높은 경우에는 하나의 단위로 기능하지 못하여 단순 나열에 가깝고 복합 구성으로 보기 어렵기

때문이다. 한편 복합 구성으로는 포함할 수 있지만 특정 학습자에게만 제한된다면 제외하고자 했다.

4.2. 교육용 목록 선정의 고려 사항

4.1에서 찾은 우선적으로 제외 가능한 항목들 외에, 교육용 목록으로 제시할 한국어 N+N 복합 구성에 대한 구체적인 선정 기준을 마련하기 위해서는 다음의 몇 가지 쟁점이 있을 것이다.

(3) N+N 복합 구성 목록 선정의 쟁점

- 가. 어떤 형식의 단위까지 N+N 복합 구성으로 볼 것인가?
- 나. 복합 구성 내 어떤 구성 요소를 포함하는가?
- 다. 두 명사는 어떠한 의미 관계로 결합되는가?
- 라. 교육용 목록으로서 어떠한 활용성을 가지는가?

이들은 각각 복합 구성의 단위 설정, 구성 요소, 의미 관계 및 교육 활용 방향에 대한 논의들이다.⁴⁾

4.2.1. N+N 단위 설정에 관한 쟁점

어떠한 형식의 단위까지 N+N 복합 구성으로 설정할 수 있을지에 관한 쟁점은 아래와 같이 정리할 수 있다.

(4) N+N 단위 설정에 관한 쟁점

- 가. 사전 등재 단위 여부
- 나. 어휘의 확장 가능성

우선, 사전의 등재어를 기준으로 삼을 것이냐의 문제이다. 기존

4) 쟁점에 대한 판단은 약 10여 명의 한국어교육 어휘 전문가의 인터뷰를 통해 본고의 제안에 대한 타당성에 관한 자문을 받았다.

사전에 한 단어로 등재된 항목은 복합 구성이 아니라 합성어이다. 예를 들어 <고려대사전>에는 ‘공동체의식’, ‘문화상품권’, ‘놀이기구’, ‘웨딩홀’, ‘프로게이머’ 등이, <표준국어대사전>에는 ‘구조대원’, ‘놀이공원’, ‘자연재해’, ‘대중교통’, ‘인간관계’, ‘식물인간’ 등은 모두 한 단어이다. 그런데 사전 등재 단위는 시간이 지나면서 새롭게 등재되거나 바뀌기도 하고, 유사한 구성이 등재가 되기도 하고 되지 않기도 하여 일관성이 부족하여⁵⁾ 사전 등재 여부를 절대적 기준으로 삼아 판단을 하기에는 무리가 있다. 다만, 등재어 중에서, 두 명사의 결합으로 하나의 긴밀한 단위를 이루는 비합성적 구성을 보이는 N+N 구성은 목록에 포함할 수 있을 것이다.

다음으로, 분석 대상 목록 중에는 형식이 고정적이지 않고 계열어를 이루어 확장이 가능한 경우가 있다. 예를 들어, ‘과학 교육’이 포함되었으나 유사한 교과목인 ‘국어 교육’, ‘수학 교육’ 등은 포함되지 않았다. 마찬가지로 ‘컴퓨터 공학’과 ‘이공 대’, ‘한의 대’는 나타났으나 다른 학문 분야의 명칭이나 대학의 명칭은 나타나지 않았다. 하지만 특정 어휘군에 소속된 구성은 하나를 이해할 때 관련 계열어를 함께 학습하는 것이 도움이 되며, 목록에 없다고 이들을 누락시키기도 어렵다. 이에 계열어를 이루는 구성을 추가로 포함하되, 빈도나 중요도를 산정하여 교육용 목록을 선별할 수 있을 것이다.

마지막으로, ‘N+N+N’ 구성의 문제이다. 다시 ‘N+(N+N)’으로 분석이 가능한 경우가 다수이나, 세 개 명사가 나란히 병렬되어 복합 구성을 이루는 경우도 있다. 가령 ‘대학 수학’은 그 자체로 사용 빈도가 높은 한 단위가 되지 못할 수 있으나 ‘대학 수학 능력

5) 일례로 최근 표준국어대사전에서 2024년 1월 개정 고시된 ‘입안’류(입안, 입숙, 코안, 콧속 등) 단어는 기존에 일반어, 전문어 구분에 따라 띄어쓰기 유무를 달리 했던 것을 항상 붙여 쓰는 한 단어로 수정하였다.

시험'이라는 'N+N+N+N' 구성이 되면 한국어 사용 맥락에서 상당히 고빈도의 중요한 구성이 된다. 이에 세 개 이상의 명사가 복합적으로 하나의 중요한 개념을 이루는 경우도 주목할 필요가 있다. N+N 구성을 기본으로 하되 한국어에서 유의미하게 사용되는 복합 구성으로 기능한다면, 확장된 연쇄 형식 또한 목록에 포함할 수 있을 것이다.

4.2.2. N+N 구성 요소에 관한 쟁점

N+N을 구성하는 각각의 어휘 요소들이 가지는 특성을 한국어 교육용 목록에 활용할 수 있을지도 검토가 필요하다.

(5) N+N 구성 요소에 관한 쟁점

- 가. 구성 항목의 난이도
- 나. 외래어 포함 가능성
- 다. 신어적 어휘의 용인 가능성
- 라. 구성 요소 어휘의 생산성

먼저, 기존 사전에 등재되어 있는 복합 구성이지만, 구성 요소 중에 난이도가 높은 어휘를 포함하는 구성을 교육용 복합 구성으로 포함할 것이냐의 논의이다. 즉, 구성 요소인 N이 한국어 기초 어휘 목록에 포함되지 않는 경우이다. 복합 구성으로서의 개념적 단위에만 주목한다면 포함해야겠지만, 한국어 교육에의 활용도를 고려한다면 이들을 제외해야 한다는 입장도 생각해 볼 수 있다.⁶⁾ 수준이 높은 어휘가 포함된 항목의 경우 학습자의 학습 수준을 넘

6) 다만 '블루오션'이 <고려대> 사전에 한 단어로 등재되어 있으나, 관련어인 '레드 오션'은 국제 통용의 어휘 목록에는 포함되지 않았다. 하지만 일상에서 고빈도로 사용되는 구성이며, 대응어가 포함된다면 수용할 수도 있을 것이다.

어서거나, 전문어를 포함하여 교육 목적의 주제와는 멀어질 가능성이 있기 때문이다. 이는 목록 선정 과정에서 말뭉치를 기반으로 한 실제 어휘 추출 후 다시 교재나 학습사전, 또는 국제 통용 한국어 표준 교육과정이나 기초어휘와 같은 정제된 자료를 통한 선별이 이루어지는 것이 바람직하겠다.

다음으로, 구성 요소에 외래어를 포함할 것이냐도 쟁점이 된다. 예를 들어 ‘빅 데이터(big data)’, ‘포토 카드(photo card)’, ‘온라인 플랫폼(online platform)’와 같이 외래어의 원어 그대로를 한국식 표기로 옮겨와 사용되는 경우이다. 이들이 일상에서 고빈도로 사용된다면 이를 포함해야 할 것이다. 학습자들은 외래어의 철자, 차용 시의 형태 등에서 어려움을 겪을 수 있기 때문이다. 또한 구성 요소의 일부만 외래어인 경우도 있는데, 팝(pop) 음악, 방송 댄스(dance), 커버(cover) 곡 등이 있다. 이들을 ‘팝 노래’, ‘커버 노래’ 등으로 대치한다면 실제 사용성이 낮다는 점에서, 이들도 수용할 만하다. 한편 ‘계란 후라이’, ‘티브이 프로’처럼 원어로 파악하기 힘든 한국식 외래어가 고빈도 구성이라면 이들 구성도 포함할 필요가 있다.

또한, 신어적 성격을 띠는 항목으로는 ‘템플 스테이, 바디 프로필, 갑 질, 뉴 노멀, 직캠, 브이 로그, 손 절, 불멍, 광팬’ 등이 있다. 신어는 일반 어휘로 정착되어 계속 사용되는 확률이 매우 낮고 대다수가 만들어져 쓰이다가 빠르게 사라지기 때문에 이들을 포함할 것이냐도 문제이다. 신어 학습에 흥미를 느끼는 학습자들이 많고 대체로 일상에서 보다 쉽게 자주 접할 수 있으므로 교육 내용에 반영하는 것이 좋다는 의견이 존재하는 한편, 변화 가능성이 큰 어휘를 모든 학습자들이 목표로 하는 교육용 목록에 포함해야 하는지에 대해서는 여전히 이견이 존재한다. 이는 사적 담화에서 주로 사용되며 빠르게 사라질 가능성이 보다 높은 유행어는 제

외하고, 사회 일반에서 두루 사용되며 공인성이 있는 뉴스 등에서 사용되며 일상생활에서 고빈도로 노출된다면 포함해야 할 것이다.

마지막으로, 생산성이 높은 N을 포함하는 항목에 대한 논의도 필요하다. 분석 목록에는 ‘N 시설, N 버스, N 사항, N 결과’ 등의 패턴이 많았는데, 이러한 구성들은 특정 의미에 관련된 어휘군 학습이 가능하다는 점에서 포함할 필요가 있다. 다만, 생산된 모든 목록이 아닌 개별 구 항목별 출현 빈도, 결합 강도, 실제적 사용 가능성 등의 기준으로 목록 내에서 다시 선별하는 작업을 거쳐야 할 것이다.

이밖에도 일부 목록은 한국 사회의 사회, 문화적 지식 및 맥락을 이해해야만 한 단위로서 이해할 수 있는 구성도 있다. ‘마을 버스’, ‘고속 버스’, ‘노인 시설’, ‘안전 시설’, ‘유흥 시설’ 등은 생산성이 높은 ‘N 버스’, ‘N 시설’의 형식을 취하면서도 단순 N의 의미 합성으로는 전체 의미 파악이 어렵기 때문에 이들도 포함할 필요가 있다고 하겠다.

4.2.3. 명사의 의미 관계 유형에 관한 쟁점

N+N 복합 구성은 그것을 이루는 각 요소들이 상호적으로 가지는 긴밀한 의미 관계를 바탕으로 성립된다. 이에 연속된 명사가 하나의 개념을 이루는 N+N 복합 구성을 살필 때 두 명사 사이의 의미 관계 양상에 관한 검토도 필요하다. 어떠한 의미 관계 유형이 우연한 명사의 단순 나열을 넘어서 교육용 복합 구성으로서 흔히 발견될 수 있는지에 관한 종합적인 검토가 요구된다.

(6) 의미 관계 유형에 관한 쟁점

- 가. 연속된 명사 의미의 합성성
- 나. 결합 관계의 의존성

의미 관계 유형에 대하여서는 하나의 복합 구성을 이루는 두 명사 의미의 합성성의 정도, 그리고 결합 관계의 종속성을 중심으로 하여 쟁점을 논할 수 있다.

첫째, 의미의 합성성에 대한 쟁점이다. 의미의 합성성을 기준으로 한 유형은 아래와 같이 구분할 수 있다.⁷⁾

(7) 두 명사 의미의 합성성을 기준으로 한 유형

- 가. 비합성적: 하루 아침, 극단적 선택, 우물 안, 밥 값
- 나. 일부 합성적: 문자 폭탄, 장롱 면허, 주말 부부, 인공 지능
- 다. 합성적: 휴대 전화, 생활 방식, 화상 통화, 지문 인식

이들은 각 명사의 의미 자질에 기대어, N1, N2 모두의 의미가 투명하지 않은 경우, 둘 중 하나의 의미가 투명하지 않고 다른 하나의 의미는 명제적 의미 그대로 이해되는 경우, 두 명사 모두의 명제적 의미가 유지되는 경우로 구분된 것이다.

우선, (7가)의 비합성적인 경우는 두 명사가 하나의 비유적 개념 또는 상황을 나타내는 관용적인 표현이 많이 해당된다. ‘하루 아침’은 어떤 사태가 발생한 기간이 꼭 하루 이거나 아침 사이가 아니더라도 그만큼 빠르게 급변했다는 의미를 가지고 있으므로 합성성이 약하다. 명제적 의미의 극단적인 선택에는 여러 가지 극한 상황에서의 결정들을 나타낼 수 있으나 한국어 사용 공동체에서

7) 두 명사 의미가 얼마만큼 합성적인가에 대해서는, N1과 N2의 의미가 매우 투명하게 그대로 반영되는 경우부터 N1과 N2 본래의 명제적 의미와는 상당히 멀어져서 두 명사로써 아예 다른 의미의 한 개념을 지칭하는 경우가 있다. 전자에 가까운 것을 합성적인 구성, 후자에 가까운 것을 비합성적인 구성이라 볼 수 있으나, 엄밀히 합성성의 여부로 구분하기는 어려우며, 스펙트럼상의 합성성 정도 차이로 이해할 수 있을 것이다.

‘극단적 선택’은 스스로 목숨을 끊는 행위를 지칭하므로 이 역시 비합성적이며, 나머지 예시들 또한 N1과 N2의 명제적 의미의 조합으로는 이해할 수 없는 복합 구성이다.

다음으로 (7나)의 일부만 합성적인 경우는 다른 한쪽의 의미가 비유적으로 확장되었거나(문자 폭탄, 장롱 면허), 하나의 결합 단위 일 때 특별히 한쪽이 가지게 되는 지시 대상의 외연에 관한 추가적인 배경 지식이 필요한 경우(주말 부부, 인공 지능)를 찾을 수 있다.

마지막으로 (7다)의 합성적인 경우는 학습자에게는 가장 이해하기 쉬울 것이다. 세 가지 유형 중 합성성을 가진 경우에는 두 명사의 명제적 의미의 조합만으로도 이해되는 구성이므로, 이들 유형을 목록에서 배제할 수 있을 것이다.⁸⁾ 본고는 한국어교육에의 적용을 고려한다면 합성성이 있더라도 고빈도로 사용되는 표현 단위로서의 유용성을 고려하여 포함하기로 한다. 이는 앞선 절에서 논의한 언어 간 대조의 측면에서 볼 때, 동일한 구성 요소로 번역되지 않거나 두 구성 요소의 결합 정보 자체가 학습자들에게는 유의미할 수 있기 때문이다.

둘째, 계열적 결합 관계를 기준으로 의미 관계를 본다면 N1과 N2 사이에서 어느 한쪽에 의존하여 계열 관계를 이루는 경우와 대등적인 경우로 다시 나누어 볼 수 있다. 의존성이 드러나는 경우는 두 명사 중 하나의 생산성이 강해서 다양한 계열어를 이룰 수 있다는 점이 특징이다.

(8) 계열적 결합 관계를 기준으로 한 유형

- 가. N1에 의존적 : 평생 학습, 평생 직장, 반려 동물, 반려 식물
- 나. N2에 의존적 : 요구 사항, 문의 사항, 편의 시설, 안전 시설

8) 자문 중 소수는 합성적 의미의 결합을 보이는 단위가 복합 구성으로서 인정받아야 하는지 의문을 제기하였다.

다. 대등적 : 예방 접종, 확진 판정, 신문 기사

(8가)는 N1에 의존하여 ‘평생 N’, ‘반려 N’과 같이, (8나)는 N2에 의존하여 ‘N 사항’, ‘N 시설’과 같이 그 복합 구성을 일반화할 수 있다. 그에 비해 계열어를 이루는 경향이 적거나 발견되지 않는 대등적 관계를 보이는 경우의 예로는 (8다)와 같이 ‘예방 접종’, ‘확진 판정’, ‘신문 기사’ 등이 발견된다. 이는 각 복합 구성 속에서 어느 한쪽의 생산성이 두드러지게 크지 않고, N1과 N2 사이의 결합 강도가 강한 것으로 볼 수 있다.

의존성의 방향을 기준으로 의미 관계 유형을 살펴보면 어휘가 확장되는 모습을 볼 수 있다. N1에 오는 ‘평생’, ‘반려’, N2에 오는 ‘사항’, ‘시설’은 그 자체만으로 사용되지 않거나 선후행 명사와 결합할 때 특정 의미를 가진다. 그렇기에 다른 명사와 결합한 단위로서 이해되고 학습될 필요가 있다. 이러한 구성의 학습은 학습자들의 확장된 어휘 학습을 촉진할 수 있는 장점이 있을 것이다. 또한 결합 유형에 대한 지식은 결합 오류를 예방하는 효과도 있을 것이다.

4.2.4. 교육 활용 방향에 관한 쟁점

다음은 교육용 목록임을 감안했을 때 추후 교육 목적의 활용 가능성에 관한 문제이다.

(9) 교육 활용 방향에 관한 쟁점

가. 언어문화 교수에의 활용성

나. 학습 목적별 교수에의 활용성

우선, 특정 목적에 따라 유용하게 활용할 수 있는 어휘군들을

교육에 활용할 수 있을 것이다. 목록에는 한국 문화와 역사에 관련되는 항목들이 발견되며, 이들은 한국의 역사와 사회 배경을 파악할 수 있는 어휘들이다.

범주	N+N 예시
한국 문화 일반	배달 앱, 노래 연습장, 골목 상권, 무형 문화재, 새벽 배송, 군 복무, 재난 문자, 대체 공휴일, 김치 냉장고, 젠더 갈등, 퍼스널 컬러, 셀프 웨딩, 감정 노동 등
북한 관련	이탈 주민, 대북 정책, 주한 미군, 이산가족 상봉 등
한국사 관련	만세 운동, 독립 운동가, 일제 강점기, 참전 용사, 독립 유공자, 독립 운동 등

<표 4> 한국 문화 관련 N+N 구성

과도하게 전문적이거나 사용 맥락이 지엽적이지 않은 경우 사회 문화적 맥락의 중요성을 고려하여 N+N 목록으로 포함할 수 있을 것이다. 다만 문화 일반, 역사 등의 세부 주제에 해당하는 복합 구성의 범위와 유형의 엄격한 구분 기준에 대해서는 교육 현장에서의 목적과 학습 수준 등에 따른 선별이 필요하다.

다음으로, 목록에는 특정 목적의 학습자에게 필요하고 유용한 항목들이 발견된다.

범주	N+N 예시
비즈니스 및 직업	근무 환경, 직업 훈련, 실업 급여, 노동 시간, 계절 근로자, 근로 계약, 인사 제도, 시장 경쟁력 등
다문화	정부 지원금, 결혼 이민자, 다자녀 가구 등
학문	학술 대회, 전문 대학원, 선택 과목, 박사 학위, 특수 목적, 박사 과정, 석사 학위, 지도 교수, 학술 연구 등

<표 5> 한국어 학습 목적 관련 N+N 구성

이들은 구체적인 대상을 염두에 둔 한국어교육을 위한 복합 구성을 선정하는 과정에서는 학습 목적과의 관련에 따라 다시 선별할 수 있을 것이다.

이상으로 유형에 따라 쟁점을 나누어 살펴보았다. 4.2절에서 제시한 추가 작업(빈도 및 범위 산정 등)의 자료들은 선행 연구의 결과물이나 본고의 말뭉치 분석 자료를 활용할 수 있다는 점에서 구체적인 목록 확정의 결과를 도출할 수 있을 것이다. 본고에서 논의한 확장 가능성, 생산성, 의미 관계의 의존성, 의미의 합성성 등의 쟁점은 서로 상관성을 가짐에 따라 추후 목록을 실제 선정하는 작업에서는 선정 기준의 복합성을 두루 고려하여야 할 것이다.

5. 결론

이상으로 한국어 어휘 교육 내용 중 하나로 N+N으로 구성된 단위의 중요성을 살피고, 보다 실제성을 가지는 말뭉치 자료를 통해 유의미한 N+N 복합 구성을 선정하는 과정에서의 쟁점을 논의하였다. 3장에서는 국립국어원 모두의말뭉치를 기반으로 문어, 구어성 문어, 구어 속성을 두루 아우르는 일상대화, 구어, 신문 장르의 말뭉치 내 N+N 복합 구성의 활용성을 검토하며 분석에 포함할 N+N 항목의 범위를 설정하였다. 4장에서는 추출된 N+N 구성을 낱말이 살펴봄여 교육용 목록으로의 포함 여부를 결정할 때 고려되어야 할 쟁점을 다루었다. 교육용 목적에 부합하지 않아 일차적으로 제외해야 할 기준으로는 결합의 우연성, 명칭의 고유성, 의미적 투명성, 학습 필요성 및 적합성이 논의되었다. 선정 항목과 관련된 쟁점으로는 단위 설정, 구성 요소의 특성, 의미 관계 유형,

교육 활용성을 중심으로 쟁점을 정리하며 교육용 어휘로서의 적합성과 확장 가능성을 살피고자 하였다.⁹⁾

분석 결과, 말뭉치 자료를 참조하여 교육적 활용을 위한 어휘 목록을 추출하고 교육용 목록으로 선정하는 작업에는 다양한 쟁점이 존재함을 확인할 수 있었다. 그간의 논의들은 N+N 구성의 특수한 형식과 의미 관계에만 주목하였거나, 빈도에만 근거한 양적 논의의 연구에만 주목한 경향이 있었다. 이에 본고는 한국어 교육에의 적용성의 측면에서 목록 선정의 쟁점을 논의함으로 해서, 보다 실제적인 한국어교육용 목록의 선정과 교육적 활용에 도움이 되고자 하였다.

〈참고 문헌〉

- 강현화. 2015. 「한국어교육용 복합어의 조어 단위 연구-한국어교육 자료를 중심으로-」, 언어와문화 11(3). 한국언어문화교육학회. 45~72쪽.
- 강현화. 2016. 「의존(성) 명사를 포함하는 한국어교육 문법항목 연구 -교육용 문법항목 선정의 쟁점을 중심으로 -」, 언어사실과 관점 38. 연세대학교 언어정보연구원. 63~86쪽.
- 강현화. 2025. 「한국어교육용 사전 등재어 'N+N' 구성 선정에 대한 기초 연구 - 사전 분석과 코퍼스 추출 구성을 바탕으로 -」, 언어사실과 관점 65. 연세대학교 언어정보연구원. 5~31쪽.
- 강현화, 권민수. 2025. 「'N+N' 구성의 결합 생산성 분석 연구」, 어문론총 103. 한국문학언어학회. 7~34쪽.
- 구교빈, 강현화. 2025. 「한국어 학습 사전 명사구 표제어 선정 연구 - 'N+N' 구성을 중심으로 -」, 어문논집 104. 민족어문학회.

9) 본고는 목록 추출에 목적을 두지 않았으며, 목록 선정에의 쟁점을 논의하고자 하였다. 구체적인 목록은 다른 지면을 통해 논의될 것이다.

255~284쪽.

- 권우진. 2022. 「식감형용사 어후목록 개발 연구」, 한국언어문화 78. 한국언어문화학회. 5~31쪽.
- 김성환. 2004. 「‘N+N’형 합성명사의 형성과 구조」, 형태론 6(1). 형태론. 67~88쪽.
- 박소영, 무아흠. 2024. 「한국어 합성명사 형성 원리 재고 - ‘N+N’형 합성명사를 중심으로 -」, 국어학 110. 국어학회. 125~162쪽.
- 손연정, 강현화. 2023. 「『한국어기초사전』의 구 단위 표제어 선정 방안 연구」, 한국사전학 41. 한국사전학회. 148~170쪽.
- 오규환. 2016. 「한국어 어휘 단위 형성과 변화 연구」, 서울대학교 박사학위 논문
- 이유경. 2021a. 「한국어 교육용 어휘 목록 선정 연구의 성과와 과제」, 한국언어문화학 18(2). 국제한국언어문화학회. 247~278쪽.
- 이유경. 2021b. 「한국어 교육용 어휘 목록의 비교 연구 - 명사 목록을 중심으로 -」, 돈암어문학 40. 돈암어문화회. 285~308쪽.
- 이찬영. 2024. 「합성명사에 대한 인지언어학적 의미 분석」, 이화어문논집 62. 이화어문학회. 201~233쪽.
- 임상은, 강현화. 2017. 「한국어 어휘 교육을 위한 명사 복합 구성 유형 연구」, 문법 교육. 한국문법교육학회. 271~296쪽.

제1저자: 권민수

주소: [03722] 서울특별시 서대문구 연세로 50

소속·직위: 연세대학교 국어국문학과 박사 수료

누리편지: iamangie99@naver.com

교신저자: 강현화

주소: [03722] 서울특별시 서대문구 연세로 50

소속·직위: 연세대학교 국어국문학과 교수

누리편지: khang@yonsei.ac.kr

<Abstract>

Corpus-based Analysis on the Issues in Selecting N+N Complex Compositions for Korean Language Education

Kwon Minsoo, Khang Hyounhwa

The purpose of this study is to examine the issues that should be considered when selecting N+N complex compositions specifically tailored for Korean language education. A corpus analysis was conducted to examine the issues. For this, we examined the frequency of N+N constructions across various subgenres in the National Institute of Korean Language's "Modu Corpus". First, we agreed to utilize N+N constructions in the subgenre corpora of everyday conversation, spoken language, and newspapers as the primary resources. And then, we selected the N+N compound constructions from these three corpora as candidates. The exclusion criteria addressed the randomness of the combination, the distinctiveness of the name, semantic transparency, and the necessity and suitability for learning. The selection of items was discussed based on the criteria of unit definition, component characteristics, types of semantic relationships, and educational applicability.

*** Key words:** Corpus-based analysis, Noun Phrase, Complex composition, Vocabulary education, Selection of vocabulary list

한글 제87권 제2호 (2026. 6.)

〈논문 받은 날: 2026. 4. 10.〉

〈심사한 날: 2026. 4. 24.~5. 8.〉

〈신기로 한 날: 2026. 5. 29.〉