

코퍼스 기반 한국어 문학 텍스트 일반 명사 어휘 선정 및 분석

-숙달도 단계 등급과 의미 범주를 중심으로-

신민섭*

-차 례-

1. 서론
2. 선행 연구
3. 연구 대상 및 방법
 - 3.1. 연구 대상
 - 3.2. 연구 방법
4. 한국어 문학 텍스트 일반 명사 어휘 분석
 - 4.1. 숙달도 단계 등급별 일반 명사 분류
 - 4.2. 의미 범주별 일반 명사 분류
 - 4.3. 종합적 논의
5. 결론

* 연세대학교 국어국문학과 박사과정.

[국문초록]

본 연구의 목적은 다양한 한국어 문학 텍스트에 출현하는 고빈도 일반 명사 어휘의 양상을 살펴보는 것이다. 이를 위하여 약 154만 어절 규모의 한국어 문학 코퍼스를 구축하였다. 연구 대상으로는 해당 코퍼스에서 출현 빈도 100회 이상의 일반 명사 어휘 954개를 선정하였다. 선정한 어휘를 대상으로 숙달도 단계 등급과 의미 범주에 따라 분류하였다. 분류는 국립국어원의 『한국어기초사전』 등재 표제어 목록을 참조하여 이루어졌다. 숙달도 단계 등급에 따른 분류 결과 초급 어휘 500개(52.41%), 중급 346개(36.27%), 고급 67개(7.02%), 등재 표제어지만 등급이 부여되지 않은 어휘 41개(4.30%)였다. 의미 범주에 따른 분류 결과 총 954개 일반 명사 어휘 중 725개(76%) 어휘는 『한국어기초사전』에서 의미 범주별로 분류되어 있었다. 229개(24%)는 등재 표제어지만 분류되어 있지 않아서 사전의 의미에 따라 분류하였다. 본 연구가 제시한 어휘 목록과 분석 결과는 한국어 문학 텍스트의 일반 명사 어휘 양상을 파악할 수 있는 자료가 된다. 또한 문학 관련 학습 자료 개발과 사전 편찬 및 보완 과정에 활용될 수 있다.

주제어 : 한국어교육, 한국어 문학, 코퍼스, 일반 명사, 한국어기초사전, 어휘 등급, 의미 범주

1. 서론

문학 텍스트를 텍스트 유형의 관점으로 살펴보면 문학의 여러 갈래에 따라 한국어 모어 화자의 문어 자료 또는 준구어 자료로 분류할 수 있으며 문학 텍스트에 나타나는 어휘는 일상생활에서 주로 사용되는 기본어휘부터 비유와 상징 등 문학의 특수성을 반영하는 어휘까지 폭넓게 나타나는 특성이 있다.¹⁾ 이러한 텍스트 유형과 어휘적 특성을 지닌 문학 텍스트를 한국어 학습자가 충분히 이해하려면 문학 텍스트 전반에 출현하는 개별 어휘의 의미를 우선적으로 파악해야 할 필요가 있다.²⁾

한국어 문학 텍스트와 한국어 교육에 대한 논의는 다양한 주제로 진행되었다. 기존의 논의는 한국어 교육을 위한 문학 작품 선정 및 교수·학습 방안, 문학을 활용한 한국어 어휘 교육, 특정 문학 작품에 나타나는 어휘 분석 등을 중심으로 이루어졌고 이론적인 측면에서 접근하는 연구가 비교적 많이 실시되었다(윤여탁, 1999; 황인교, 2001; 신윤경, 2008; 박배식, 2009; 주재환, 2014; 장연분홍, 2021 외). 특히 문학 텍스트의 어휘 관련 연구들에서 연구 대상으로 선정한 문학 텍스트는 주로 학교 국어 교육에서 비중 있게 다루어지거나 많은 한국어 교재에 중복으로 수록된 문학 작품에 그쳐 문학 텍스트 어휘 측면에서 다양성이 다소 결여된다고 지적할 수 있으며 전반적인 문학 텍스트 어휘의 양상을 확인하기 어렵다는 한계가 있다.³⁾ 또한 많은 한국어 교육 현장에서

-
- 1) 본 연구가 언급한 기본어휘는 일상생활 각 영역에서의 필요와 사회 구성원으로서 정상적인 기본 생활을 하는 데 필요하다고 간주되는 어휘를 의미한다(이충우, 1994: 33-34면).
 - 2) 박현진 외(2013: 133-134면)는 한국어 학습자가 문학 텍스트를 이해하는 데 실패하게 만드는 요인으로 총 여섯 가지를 제시하였는데 그 중 첫 번째와 두 번째 요인이 어휘와 관련되어 있었다. 해당되는 이해 실패 요인은 각각 모르는 어휘에 대한 잘못된 추측과 아는 어휘에 대한 확대 해석이었다. 이는 문학 텍스트에 나타나는 어휘가 지닌 상징적인 의미를 추측하는 것에 앞서 어휘의 사전적 의미를 먼저 앎으로써 모르는 어휘가 문학 텍스트 이해에 미치는 영향의 비중을 최소화해야 함을 보여준다.
 - 3) 동일한 문학 텍스트에 대한 어휘 관련 연구가 두 편 이상 이루어진 대표적인 작품으로는 황순원의 <소나기>(주재환, 2014; 박영미, 2019; 장연분홍, 2021)가 있다.

도 사용하는 교재에 수록되었거나 학습 자료로 선정한 문학 텍스트에 따라 해당 텍스트에 제시된 어휘만을 다루게 되는 경우가 많다. 이러한 제한적인 어휘의 제시는 한국어 학습자가 일상 속에서 또는 한국어능력시험(TOPIK: Test of Proficiency in Korean)과 같은 각종 평가에서 새롭게 접하는 낯선 문학 텍스트를 이해하는 데 어려움을 겪거나 실패하게 되는 요인으로 작용할 가능성이 있으므로 한국어 문학 텍스트 전반에 출현하는 특징적인 어휘 목록을 작성하여 학습자에게 제시할 필요성이 있다. 따라서 본 연구의 목적은 다양한 한국어 문학 텍스트에 출현하는 고빈도 일반 명사 어휘의 양상을 살펴봄으로써 어휘 목록과 분석 결과를 제시하는 데 있다. 이는 한국어 문학 텍스트의 일반 명사 어휘 양상을 파악할 수 있는 자료가 될 수 있으며 문학 관련 학습 자료 개발과 사전 편찬 및 보완 과정에 활용될 수 있다.

2. 선행 연구

본 연구가 논의하고자 하는 주제의 선행 연구 검토를 위하여 코퍼스를 기반으로 한 어휘의 계량적 분석 관련 연구, 문학을 활용한 한국어 교육 연구, 한국어 학습자의 문학 텍스트 이해에 관련된 연구, 특정 문학 작품에 나타나는 어휘 분석 연구로 나누어 기존의 논의들을 살펴보았다.

먼저 코퍼스를 기반으로 한 어휘의 계량적 분석에 관련된 연구로는 한영균(2006; 2007)의 논의를 살펴보았다. 한영균(2006)은 코퍼스를 계량적으로 분석하여 얻어지는 어휘의 빈도 정보 활용에서의 문제점을 지적하고 실제 빈도 조사 결과 사례를 제시하면서 이러한 문제를 보완하기 위한 방법을 구체화하였다. 한영균(2006)은 어휘 교육 및 학습을 위한 자료를 개발하는 데 코퍼스를 활용하는 경우 일부의 자료에서 표본을 추출하거나 빈도 조사 기준을 분명하게 세우지 않으면 처리 방식에 따라 계량 결과가 달라진다는 점을 언급하면서 계량적 분석의 방향을 모색하였고 어휘 목록의 편차를 줄이는 방안을 제안하였다. 또한 한영균(2007)은 코퍼스 분석 결과인 어휘 통계 정보의 중요성을 확인하

면서 어휘의 계량적 분석 결과가 어휘 교육과 학습을 위한 기초 자료 개발에 유용하게 활용될 수 있음을 논의하였다. 본 연구는 한영균(2006; 2007)이 코퍼스 기반 어휘 연구 방법론과 그 활용에 대하여 다룬 내용을 참고하여 한국어 문학 텍스트 일반 명사 어휘를 선정하고 분류하고자 한다.

문학을 활용한 한국어 교육 연구와 한국어 학습자의 문학 텍스트 이해에 관련된 연구로는 윤여탁(1999), 황인교(2001), 박배식(2009), 신윤경(2014), 박현진 외(2013)를 살펴보았다. 윤여탁(1999)과 황인교(2001)는 한국어 교육에서의 문학 교육의 실태와 한계를 지적하면서 체계화된 문학 교육의 방안과 개선 방향에 대하여 이론적 측면에서 논의하였다. 윤여탁(1999)은 한국어 학습자의 학습 목표가 다양화됨에 따라 한국 문학의 교육적 필요성과 한국 문학 학습을 위한 방법론이 요구된다는 점을 강조하면서 문학 작품을 통한 한국어 교육 방법의 이론적 기반과 실재를 검토하였다. 또한 문학이 언어를 학습하는 중요한 학습 자료가 될 수 있고 효과적인 한국어의 이해와 표현 학습을 수행할 수 있으므로 문학을 활용하여 언어 능력을 기르는 교육이 이루어져야 함을 주장하면서 문학에 나타난 어휘와 언어적 표현을 이해하는 학습의 필요성을 언급하였고 이를 한국어 교육 과정과 교재 구성의 문제와 함께 논의하였다. 황인교(2001)는 외국어로서의 한국 문학 교육의 개념을 정리하고 문학을 활용한 한국어 학습에서의 한계와 해결 방안에 대하여 논의하면서 학습자에게 문학 작품의 어휘를 어떻게 선별하고 제시해야 하는가에 대하여 언급하였고 문학 교육을 언어 모델, 문화 모델, 개인 성장 모델로 나누어 세 가지 이론적 틀을 제시하면서 체계적 실천 방향을 모색하였다. 황인교(2001)는 문학 작품에 쓰이는 특별한 어휘의 과도한 출현이 일정 수준 이상의 언어 능력을 갖춘 학습자에게만 문학 교육이 가능하다는 주장의 근거가 되고 있음을 지적하면서 이를 해결하려면 자료와 교수법 측면에서의 노력이 필요하고 그 예로는 자료 자체를 쉽게 고치거나 모국어 번역을 제시하는 방법, 작품을 어휘 수준별로 분류하여 제시하는 방법, 문맥을 활용하여 자연스럽게 어휘를 익히도록 유도하는 방법이 있음을 주장하였다. 즉 학습자가 스스로 문학 작품의 어휘를 이해하고 익힐 수 있는 단계에 이르려면 학습자 수

준에 적절한 문학 어휘를 파악하고 선별하여 학습자에게 제시할 수 있어야 한다고 보았다.

박배식(2009)은 한국어 문학 교육의 방향을 제시하고 문학의 장르 가운데 시 문학을 중심으로 문법, 어휘, 문화, 정서 교육의 실행 방안에 대하여 논의하였다. 박배식(2009)은 시를 활용한 어휘 교육에 관련된 논의에서 어휘에 대한 바른 이해를 강조하면서 시어로 쓰이는 어휘의 대부분이 비유적 의미로 확장되어 이해되지만 반드시 어휘의 중심 의미에 대한 이해가 선행되어야 한다고 하였다. 신윤경(2014)은 한국어 교육에서 문학 텍스트 읽기를 위한 어휘 이해 방안에 대하여 어휘 주석을 중심으로 논의하였다. 신윤경(2014)은 문학 텍스트에 쓰인 어휘의 의미를 제대로 이해하는 것의 중요성을 강조하면서 문학 교재의 어휘 주석에 문화 배경과 맥락 이해를 위한 설명을 추가해야 한다고 주장하였다. 그러나 읽기 자료 또는 문학 교재에 수록된 텍스트에 출현하는 어휘의 시사적 의미, 고유문화적 의미, 맥락적 의미를 주석에 함께 제시하여도 해당 텍스트에 한정된 의미일 수 있고 어휘의 중심 의미를 명확히 알고 있지 않으면 오히려 학습자에게 혼동을 줄 수도 있다는 점이 문제가 될 수 있으며 한국어 교육 현장을 벗어난 상황에서 접하게 되는 넓은 범위로서의 문학 텍스트를 이해하는 데 근본적으로 필요한 것은 문학 텍스트 전반에 주로 쓰이는 어휘를 파악하고 이를 학습하는 것이라고 여겨진다. 박현진 외(2013)는 이러한 맥락에서 학습자가 문학 텍스트를 이해하는 데 어려움을 겪는 요인과 실패 요인이 무엇인지 밝히기 위한 질적 연구를 수행하였는데 결과적으로 어휘와 관련된 요인의 비중이 높게 나타남을 확인할 수 있었다. 반면 맥락, 배경 지식, 문학적 표현으로 인한 이해 실패는 어휘에서 비롯된 이해 실패보다 비교적 낮은 수준으로 나타났다. 문학을 활용한 한국어 교육 연구와 문학 텍스트 이해에 관련된 선행 연구들을 검토한 결과 이상의 연구를 포함한 대부분의 논의가 이론적인 측면에서 이루어졌고 실제 문학 텍스트의 어휘를 분석하는 논의는 다소 부족한 것으로 확인되었다. 또한 문학 텍스트와 관련된 한국어 어휘 교육 연구의 측면에서도 분석을 통한 구체적인 자료를 제시하기보다 원론적인 교육 방안과 개선 방향을 제안하는 정도에 그치기 때문에 이러한 부분을 검토하고 보완할 필요가 있다고 여

겨진다.

특정 문학 작품에 나타나는 어휘 분석 연구로는 주재환(2014)을 살펴보았다. 주재환(2014)은 황순원 작가의 <소나기>를 대상으로 어휘의 난이도를 측정하고 분류하였으며 이를 토대로 수준별 읽기 교재 개발의 방향을 검토하였다. 주재환(2014)은 텍스트에 나타난 어휘의 품사별 사용 양상을 살펴보고 국립국어원이 2011년에 발표한 ‘한국어 학습용 어휘 목록’과 대조하여 어휘 난이도와 빈도를 분석하였다. 분석 결과를 토대로 <소나기>에 사용된 어휘의 난이도가 매우 높은 편이라고 판단하였으며 수준별 읽기 교재 개발을 위해서는 텍스트에 등장하는 어휘의 등급별 비율과 빈도에 대한 구체적인 분석이 이루어져야 함을 제언하였다. 주재환(2014)은 실제 문학 텍스트에 출현한 어휘의 난이도와 빈도를 분석하였다는 점에서 본 연구의 방법과 유사하지만 특정 문학 작품만을 대상으로 삼아 연구를 진행하였다는 점에서 차이가 있다. 따라서 본 연구는 앞서 살펴본 선행 연구들의 논의를 반영하여 다양한 문학 텍스트로 구성된 코퍼스를 구축한 뒤 기존의 연구를 보완하는 계량적 분석의 연구를 진행하고자 한다.

3. 연구 대상 및 방법

3.1. 연구 대상

본 연구의 연구 대상은 연구를 위하여 연구자가 직접 구축한 총 1,540,776어절 규모의 한국어 문학 코퍼스이다. 한국어 문학 코퍼스는 한국 현대 문학⁴⁾을 위주로 구성하였고 문학의 장르에 따라 시, 소설, 수필, 희곡으로 나누어 구성하였다.⁵⁾

4) 고전 문학을 제외하고 현대 문학을 위주로 코퍼스를 구축한 까닭은 고전 문학의 경우 수집할 수 있는 절대적인 텍스트 수가 적을 뿐만 아니라 수집 가능한 대부분의 자료가 원본을 현대 국어에 맞게 개작한 자료여서 동일한 문학 작품을 보더라도 편집자에 따라 어휘나 표현이 상이한 것을 확인할 수 있었고 비교적 쉬운 어휘와 현 시대에 주로 쓰이는 어휘로 수정되어 있어서 현대 문학 텍스트에 출현하는 어휘만으로도 충분히 보완이 가능할 것이라고 판단하였기 때문이다.

한국어 문학 코퍼스를 구축하기 위한 절차를 간략하게 소개하면 먼저 텍스트 자료를 수집할 문학 작품을 선정하였다. 문학 작품은 일차적으로 한국어 교육용 문학 작품을 선별한 연구(박배식, 2009; 윤여탁 외, 2014; 장연분홍, 2021 외)가 제시하는 목록에 포함된 문학 작품을 선정하였고 이차적으로는 초·중·고등학생 대상 필독 문학, 대학수학능력시험 출제 문학, 대학생 대상 필독 문학의 목록을 참고하여 선정하였다.⁶⁾ 또한 국내 온라인 서점의 분야별 도서 목록을 참고하여 최근까지 출판된 문학 작품을 추가적으로 선정하였다.⁷⁾ 선정한 문학 작품들은 코퍼스 구축 기준에 따라 모두 한국 현대 문학으로 분류되는 작품에 해당되었고 서로 중복되지 않도록 하였다. 문학 작품 선정 이후 각각의 문학 작품을 텍스트 자료로 수집하였고 이를 장르별로 나누어 형태 분석과 문맥색인 주석이 포함된 형식으로 한국어 문학 코퍼스를 구축하였다.⁸⁾ 이러한 절차에 따라 본 연구의 대상으로 구축한 약 154만 어절 규모의 한국어 문학 코퍼스를 표로 정리하면 다음과 같다.

<표 1> 한국어 문학 코퍼스 구성 및 규모

장르		포함된 문학 작품(일부 예시)	어절 수
한국 현	시	가난한 사랑 노래(신경림), 나룻배와 행인(한용운), 모란이 피기까지는(김영랑)	127,392
	소	병신과 머저리(이청준), 술 권하는 사회(현진건),	498,805

- 5) 문학의 장르를 나누는 기준은 일반적으로 알려진 한국 문학의 갈래를 따랐으며 코퍼스 구축에 활용한 문학 텍스트는 모두 원본을 기준으로 수집하였다. 코퍼스 규모에 있어서 어느 특정 장르에 편중되지 않도록 최대한 균형적인 코퍼스를 구축하고자 하였으나 수집한 자료에 한계가 있었고 특히 시 문학 텍스트의 경우 장르의 특성상 하나의 작품에 포함된 어절 수가 비교적 적기 때문에 이를 반영하여 최대한 많은 수의 작품을 포함시켰음에도 불구하고 다른 장르의 코퍼스 규모와 다소 차이가 발생하였다.
- 6) 필독 문학에 해당되는 문학 작품 목록은 단편집 또는 모음집으로 출판된 서적들의 목차를 토대로 정리하였고 출제 문학의 목록은 다양한 문학 참고서에 수록된 문학 작품 목록을 토대로 파악하였다.
- 7) 본 연구는 국내 온라인 서점으로 교보문고 누리집을 활용하였다.
(<http://www.kyobobook.co.kr>).
- 8) 형태 분석에는 울산대학교 한국어처리연구실에서 개발한 ‘UTagger 2.2.1’을 활용하였고 문맥색인 주석에는 ‘UTagger_PostProcessor’를 활용하였다.

대 문 학	설	엄마를 부탁해(신경숙)	
	수 필	거리의 악사(박경리), 꼴찌에게 보내는 갈채(박완서), 낙엽을 태우면서(이효석)	441,068
	회 곡	국물 있사옵니다(이근삼), 그림자 재판(오테영), 만만한 인생(이대영)	473,511
총 규모			1,540,776

3.2. 연구 방법

본 연구는 약 154만 어절 규모의 한국어 문학 코퍼스를 기반으로 한국어 문학 텍스트에 출현한 일반 명사 어휘의 양상을 살펴보기 위하여 어휘 빈도 정보를 활용하였다. 어휘 빈도 정보를 파악하기 위한 빈도 추출 도구로는 ‘Counter’를 활용하였다. 일반 명사 어휘의 양상은 출현 빈도를 기준으로 살펴보았는데 전체 코퍼스에서 일반 명사로 분류되는 어휘는 총 34,810종이었고 이를 빈도순으로 나열하여 한국어 문학 텍스트의 일반 명사 어휘 출현 양상을 확인하였다.⁹⁾

일반 명사 어휘 전체를 다루기에는 연구의 범위에서 한계가 있으므로 출현 빈도 100회 이상인 954개 어휘를 추출하여 한국어 문학 코퍼스에 포함된 고빈도 일반 명사 어휘 목록을 작성하였고 이 목록을 분석을 위한 최종 어휘 목록으로 채택하였다.

다음으로는 고빈도 일반 명사 어휘 목록에 포함된 어휘를 숙달도 단계 등급과 의미 범주에 따라 분류하였다. 숙달도 단계 등급과 의미 범주에 따른 분류는 국립국어원 누리집의 『한국어기초사전』 등재 표제어 목록을 참조하였다.

본 연구가 참조한 『한국어기초사전』은 한국어 학습자의 학습 요구에 부응하는 사전과 다국어 대역의 기반이 되는 기초 사전을 목표로 구축되어 약 5만 어휘의 표제어가 등재되어 있다(강현화 외, 2012: 7-9). 또한 사전 편찬을 위한 학습자의 요구 분석, 관련 자료의 데이터베이

9) 한국어 문학 코퍼스에서 가장 고빈도로 출현한 일반 명사 어휘는 ‘사람’으로 총 9,695회 출현하였다.

스, 효율적 편찬 도구 개발이 체계적으로 이루어졌으며 학습자의 수요에 기초한 접근성을 고려하여 표제어를 선정하였다(강현화 외, 2015: 40-44). 『한국어기초사전』은 기존의 여러 한국어 사전이 제시하는 단어에 대한 일반적인 정보 외에 한국어 학습자를 위한 단위별 표제어, 체계화된 예문, 어휘 등급, 주제 및 상황 범주, 의미 범주, 다중 매체 정보 등을 함께 제시한다는 점에서 기존의 한국어 사전들과 차별화된다.¹⁰⁾ 본 연구는 『한국어기초사전』의 이러한 특성을 고려하여 숙달도 단계 등급 및 의미 범주별 어휘 분류 방법으로 활용하였다.

『한국어기초사전』의 사전 통계 및 현황¹¹⁾을 살펴보면 단어, 구, 문법·표현의 모든 단위를 포함한 총 표제어는 51,957개이고 그 중 단어 단위의 표제어가 50,041개이며 명사 어휘는 29,088개 포함되어 있다. 본 연구가 활용할 표제어의 숙달도 단계별 어휘 등급과 의미 범주 현황은 다음과 같다. 먼저 숙달도 단계별 어휘 등급은 초급, 중급, 고급, 등급 없음으로 구분되며 각각의 표제어 수는 초급 1,942개, 중급 3,982개, 고급 5,106개, 등급 없음 40,927개이다.¹²⁾ 의미 범주는 대분류와 소분류로 나누어지며 총 14개의 대분류 범주와 각각의 대분류에 해당되는 소분류 범주로 구성되어 있다. 소분류 범주는 총 139개로 대분류에 따라 최소 6개에서 최대 19개의 소분류 범주가 하위 범주로 포함되어 있다.¹³⁾

10) 『한국어기초사전』의 구축 방안 및 편찬 원리에 대한 세부적인 내용은 강현화 외(2012; 2015)에 보다 자세히 소개되어 있다.

11) 『한국어기초사전』의 구조와 관련된 내용은 강현화 외(2012; 2015)에 제시되어 있지만 최근까지의 수정 사항을 반영하고 있지는 않다. 따라서 본 연구는 최근의 사전 구조인 국립국어원 누리집의 『한국어기초사전』 사전 통계를 추가적으로 살펴보고 그 결과를 토대로 사전 통계 및 현황을 제시하였다.

12) 『한국어기초사전』 사전 통계에 따르면 『한국어기초사전』이 제시하는 숙달도 단계별 어휘 등급은 2012-2015년 사이에 국립국어원이 발표한 ‘한국어 교육 어휘 내용 개발 1-4단계’ 보고서 결과를 반영한 어휘 숙달도 단계이다. 숙달도 단계별 어휘 등급은 단어, 구, 문법·표현의 모든 단위를 포함한 총 표제어 51,957개에 등급을 부여한 것이므로 등급별 표제어 합계도 이와 같다.

13) 14개의 대분류 범주와 각 대분류 범주에 포함되는 소분류 범주의 수를 제시하면 다음과 같다. 1. 인간(16개), 2. 삶(12개), 3. 식생활(10개), 4. 의생활(8개), 5. 주생활(8개), 6. 사회생활(16개), 7. 경제생활(6개), 8. 교육(8개), 9. 종교(7개), 10. 문화(9개), 11. 정치와 행정(7개), 12. 자연(6개), 13. 동식물(7개), 14.

본 연구는 이러한 연구 방법에 따라 연구 대상으로 선정한 954개 어휘와 『한국어기초사전』 등재 표제어를 비교하여 각 어휘의 숙달도 단계 등급과 의미 범주를 파악하고 분석하였다. 어휘의 숙달도 단계 등급은 『한국어기초사전』을 참조하여 초급, 중급, 고급의 세 등급으로 분류하였고 『한국어기초사전』에서 ‘등급 없음’에 해당되는 어휘이거나 등재되어 있지 않은 어휘는 별도로 제시하였다. 의미 범주도 마찬가지로 『한국어기초사전』을 참조하여 분류하였고 『한국어기초사전』에서 분류되지 않은 ‘분류 없음’ 어휘는 『한국어기초사전』의 의미 범주 분류 기준에 따라 직접 해당되는 의미 범주에 추가하고 제시하였다.

4. 한국어 문학 텍스트 일반 명사 어휘 분석

4.1. 숙달도 단계 등급별 일반 명사 분류

한국어 문학 텍스트의 고빈도 일반 명사 어휘 목록을 숙달도 단계 등급에 따라 분류한 결과 초급 500개, 중급 346개, 고급 67개로 분류되었다. 또한 『한국어기초사전』에 등급이 제시되지 않은 어휘는 41개로 나타났고 미등재 어휘로 분류되는 어휘는 없었다.

숙달도 단계 등급에서 초급으로 분류된 어휘는 총 500개로 한국어 문학 텍스트의 고빈도 일반 명사 어휘 전체의 52.41%를 차지하는 것으로 나타나 절반 이상의 어휘가 초급 단계 어휘에 해당된다는 것을 확인할 수 있었다. 문학 텍스트에 나타나는 일반 명사 어휘를 고빈도순으로 나열하고 그 중 출현 빈도 100회 이상인 어휘를 살펴보았을 때 초급 어휘로 500개가 분류되었다는 것은 『한국어기초사전』 전체의 초급 어휘 표제어 총 1,942개의 25.75%가 이 범위에 분포한다는 것을 보여준다. 이는 『한국어기초사전』에서 초급 단계 표제어로 등재된 어휘 가운데 비교적 많은 수의 어휘가 한국어 문학 텍스트의 고빈도 일반 명사 어휘를 구성하고 있음을 보여주며 실제 코퍼스를 기반으로 어휘 빈도

정보를 확인하였을 때 초급 단계에 해당되는 어휘가 고빈도로 출현한다는 양상을 보여준다. 초급으로 분류된 고빈도 일반 명사 어휘 500개를 제시하면 다음과 같다.

<표 2> 초급 단계 등급 어휘 목록14)

한국어 문학 텍스트 고빈도 일반 명사 어휘 - 초급(500개)
사람(9695), 말01(8529), 때01(6758), 일01(5885), 집01(4213), 소리01(4138), 속01(4028), 여자02(3694), 눈01(3637), 아이01(3201), 자신01(3128), 얼굴01(3089), 앞(2838), 손01(2766), 생각01(2700), 위01(2533), 몸01(2428), 어머니01(2374), 뒤01(2268), 곳01(2241), 마음01(2229), 전07(2097), 안01(2049), 날01(2018), 아버지(1983), 모습01(1958), 시간04(1954), 길01(1822), 방07(1803), 아내01(1737), 엄마(1734), 남자01(1712), 그때(1609), 머리01(1574), 세상01(1530), 돈01(1515), 물01(1515), 이야기(1512), 이름(1510), 친구02(1490), 입(1470), 정도11(1440), 동안01(1399), 자리01(1392), 지금02(1345), 고개01(1339), 사이01(1334), 처음(1327), 오늘(1321), 후08(1299), 하늘01(1279), 밤01(1277), 사실04(1256), 나라01(1253), 가슴01(1210), 목소리(1200), 문04(1192), 할머니(1170), 열(1155), 남편01(1140), 사랑01(1094), 밖(1067), 문제06(1035), 마을01(1024), 선생님(1018), 다음01(1014), 아빠(1009), 차06(1004), 바람01(985), 끝01(982), 얘기(965), 아침(958), 힘01(952), 땅01(947), 혼자01(947), 이상04(937), 모양02(933), 책01(927), 아저씨(913), 나무01(907), 산01(901), 꿈01(881), 술01(881), 이제01(855), 아래01(852), 학교(851), 옷01(830), 애02(828), 전화07(827), 바다(825), 잠01(820), 세계02(802), 이번01(796), 남01(795), 할아버지(787), 눈물01(786), 발01(774), 아들(770), 밀01(750), 이유04(738), 노래01(731), 꽃01(714), 조금01(713), 기본01(698), 불01(696), 나이01(693), 하루01(665), 새03(653), 귀01(641), 대학01(639), 느낌(631), 비01(630), 생활(630), 딸01(616), 얼마(614), 영화01(614), 해01(614), 도시03(610), 담배(602), 뜻(601), 학생(596), 다리01(587), 글(583), 웃음(580), 회사04(580), 오빠(577), 어른01(562), 형01(558), 어깨01(557), 기억02(556), 자식01(556), 목01(555), 그날(552), 피02(550), 배02(548), 마지막(547), 사진06(540), 냄새(530), 밥01(527), 아기01(526), 주인01(515), 저녁(513), 그림01(492), 역사03(492), 소설03(488), 옛날(487), 대답(485), 의미02(476), 팔01(476), 건물03(474), 동생01(467), 벽05(467), 반05(465), 모두01(460), 동네(458), 고향02(457), 왕04(453), 말씀(448), 결혼(447), 거리01(446), 주위02(437), 돌02(435), 필요(432), 걱정(427), 물건(423), 며칠(421), 공부01(419), 신문10(417), 문화01(414), 교수06(410), 층01(409), 사장15(408), 편지02(404), 화05(403), 강01(401), 방법(400), 키01(399), 노인

14) 속달도 단계 등급별 어휘 목록은 빈도순으로 나열한 것이며 괄호 안의 숫자는 출현 빈도를 의미한다. 어휘의 동형어 번호는 『표준국어대사전』의 동형어 번호와 일치한다.

01(393), 컴퓨터(393), 고양이(392), 겨울(390), 요즘(389), 새벽01(382), 가게(380), 가족 01(379), 가운데(375), 관심01(375), 침대02(371), 버스02(370), 입술(369), 바닥01(365), 배01(363), 이후02(362), 잔02(361), 청년(359), 아파트(358), 여름01(353), 손님(352), 언 니(352), 개03(350), 손가락(347), 기자04(343), 내용02(339), 병원02(339), 가을01(337), 구두01(337), 별01(337), 섬03(333), 부부03(332), 관계05(331), 나중01(330), 허리 01(328), 자연01(326), 걸음(324), 내일(323), 분위기(322), 동시02(321), 봄01(321), 학년 (321), 코01(315), 선배(313), 겨울01(312), 창문(312), 부인01(310), 마당(308), 식구 01(308), 약속(306), 호랑이(306), 음악01(303), 부분01(302), 병04(300), 종이01(299), 주변04(299), 시장04(297), 여성01(296), 소03(295), 아주머니(293), 음식(292), 의사 03(291), 질문(290), 의사12(289), 무릎(288), 여행02(288), 칼01(288), 맛01(285), 등 01(283), 책상01(280), 상처02(279), 과거02(278), 인사02(278), 기회03(277), 아무것 (275), 운동02(273), 소식04(270), 달05(267), 부엌(264), 계단04(262), 땀01(262), 결과 02(261), 고기01(260), 대부분(259), 울음(257), 줄01(255), 택시(252), 행동(251), 비행기 (248), 시골(247), 동물(242), 만약(242), 희망(242), 오후02(241), 혀01(234), 방향 01(233), 호텔(232), 기쁨(230), 닭(230), 슬픔(229), 춤01(229), 화장실(228), 구름 01(227), 비밀(226), 이마01(226), 능력02(225), 준비(225), 전채01(224), 이때(223), 자유 03(223), 주머니(223), 누나01(221), 도로07(221), 지방04(221), 최고02(221), 어제 01(220), 진짜(218), 값(216), 눈04(216), 바깥(215), 부장06(215), 가방01(214), 대화 06(214), 식당(213), 식사03(213), 그릇01(212), 근처(211), 아가씨(211), 커피(210), 평소 (210), 우리나라(209), 중심01(209), 안개(208), 교육(207), 도움(207), 여기저기(206), 속 도01(205), 보통(204), 아줌마(202), 장소05(200), 행복02(200), 살01(199), 우산01(199), 게임(198), 태도03(197), 거리07(194), 사고12(194), 열차02(193), 이웃(193), 풍경 01(193), 비디오(192), 성격02(192), 신랑02(192), 시험03(191), 사무실(190), 미래 02(189), 엉덩이(188), 햇빛(188), 남쪽(187), 어린이01(187), 입구02(187), 스스로(186), 거짓말(185), 다음날(185), 배03(185), 텔레비전(184), 경찰04(183), 시계01(183), 짐 01(183), 동쪽(182), 나머지(181), 쌀(179), 짝01(179), 계획01(178), 낮(177), 부모님 (175), 자동차(175), 색03(174), 왼쪽(174), 선물03(173), 트럭(172), 경험(171), 점심 (171), 환자03(171), 반08(170), 폐지(167), 공장02(166), 방문02(166), 날씨01(164), 프 로그램(164), 맥주(163), 여학생(163), 직원03(161), 기차01(160), 병05(160), 손바닥 (159), 방송01(157), 이불01(157), 뱀(156), 시내03(156), 지하(156), 군인(155), 설명 (155), 시작01(155), 약07(155), 그동안(154), 언어01(154), 오른쪽(153), 직장05(152), 고 등학교(151), 목적03(151), 제목02(151), 중간01(151), 머리카락(150), 놀이01(149), 바닷 가(149), 빵01(149), 색깔(149), 이사13(149), 중앙01(149), 바지01(148), 서쪽(148), 상자 08(146), 오랫동안(146), 교회02(145), 물고기(145), 버릇01(145), 애인02(145), 이전 03(144), 일01(144), 회의03(144), 사업04(143), 오랜만(143), 반대02(141), 북쪽(141), 연락02(141), 형제01(141), 신발(140), 예술(140), 최근(140), 덕분(139), 소주04(139), 안 경03(139), 잘못(139), 약간(138), 직업(138), 유리07(137), 광고02(136), 외국02(136), 일 부02(136), 카페(136), 재미01(135), 단어(134), 이해05(134), 숫자(133), 만일01(132), 수

업03(132), 과일01(131), 농담01(131), 위치01(131), 현재02(131), 방학(130), 사흘(130), 노력01(129), 공원03(128), 선수05(128), 크기(127), 다리02(126), 부탁(126), 인형 01(126), 긴장(125), 중학교(123), 거실02(122), 아까(122), 공항02(121), 구경01(121), 남 녀(121), 모자07(121), 역11(121), 접시(120), 도서관(119), 얼음01(119), 후배05(119), 일 주일(117), 수23(116), 어젯밤(116), 올해(116), 공01(115), 이틀01(114), 소개02(113), 대 학생(112), 신호01(112), 호수07(112), 글씨(111), 엘리베이터(111), 자전거(111), 정원 06(111), 제일04(111), 극장(110), 기간01(110), 부자06(110), 코끼리(110), 생일02(108), 카메라(108), 물론01(107), 시민(107), 서류02(106), 카드(106), 신부10(104), 어머님 (104), 박물관(103), 장미05(103), 건강03(102), 건너편(102), 운동장(102), 치마01(102), 높이01(101), 종류02(101), 큰소리01(101), 강아지(100), 뉴스(100), 일요일(100), 작년 (100)

『한국어기초사전』의 등급별 표제어 수를 보면 초급에서 고급으로 숙 달도가 올라갈수록 어휘의 수가 증가하는 반면 한국어 문학 텍스트의 고빈도 일반 명사 어휘 양상은 중급으로 분류된 어휘의 수가 초급으로 분류된 어휘의 수보다 다소 적은 경향을 보였다. 중급 단계에 해당되는 어휘의 수는 총 346개이며 이는 전체 954개의 고빈도 일반 명사 어휘 에서 36.27% 비율을 차지하는 어휘가 숙달도 중급에 해당된다는 것을 의미한다. 이를 『한국어기초사전』 전체 중급 어휘 표제어 수에 비교하 면 총 3,982개의 중급 단계 표제어 중 346개 어휘가 한국어 문학 텍스 트의 고빈도 일반 명사 어휘 목록에 포함되어 있음을 확인할 수 있고 분포로는 8.69%의 중급 단계 표제어가 이에 해당되는 것을 볼 수 있 다.

중급으로 분류된 고빈도 일반 명사 어휘 346개를 제시하면 다음과 같다.

<표 3> 중급 단계 등급 어휘 목록

한국어 문학 텍스트 고빈도 일반 명사 어휘 - 중급(346개)
인간01(1106), 순간03(1069), 표정02(1033), 삶(963), 정신12(800), 소녀02(651), 빛(648), 사내01(629), 기사10(597), 시대02(596), 결국(591), 사회07(582), 숨 01(565), 시절01(565), 시선03(564), 점10(503), 죽음01(496), 어둠(488), 경우 02(485), 한참(484), 상황02(480), 전쟁(465), 박사01(450), 사건01(439), 대신 03(434), 당시02(430), 시09(425), 상태01(418), 존재(412), 소년01(406), 숲

01(404), 스님(397), 생명(383), 미소05(378), 마찬가지(377), 여인01(372), 날개
 01(357), 인생01(357), 현실02(357), 공간05(356), 새끼02(352), 고통(351), 시인
 10(348), 짓01(346), 결01(344), 세월02(336), 눈치(332), 지원02(331), 집안
 01(331), 눈빛01(326), 감정06(320), 오토바이(317), 작가01(316), 한마디(315),
 바위01(310), 창07(302), 흙01(300), 소문02(298), 의식03(297), 문학01(295), 작
 품01(289), 온몸(288), 영혼02(283), 원래01(279), 눈길01(278), 당장02(278), 한
 쪽(276), 쥐02(275), 언덕(272), 목숨(268), 틈01(268), 강물(265), 싸움(263), 공
 주01(262), 지역02(261), 한숨02(261), 부모01(260), 비명02(260), 형태(260), 처
 녀(259), 대통령(258), 차이(257), 국가01(256), 햇살(255), 신경04(254), 불빛
 (253), 겁02(251), 무덤(249), 예14(242), 눈앞(240), 화면05(240), 인물(237), 풀
 02(236), 젊은이(234), 절02(231), 민족(230), 잠자리02(228), 국민(226), 귀신
 01(225), 파도(225), 골목01(223), 너머(223), 꼬리01(222), 자세02(222), 정보
 06(220), 구멍(219), 연구03(217), 통일02(217), 구석01(216), 기운01(216), 세기
 03(216), 상대04(215), 용09(214), 때01(213), 운명01(211), 만남(207), 여유
 (204), 불교(203), 대상11(202), 부처01(201), 천사04(201), 기술01(199), 알
 01(198), 면05(197), 그림자(195), 욕망(195), 죄03(195), 공포08(194), 향기
 01(193), 지구04(189), 현장03(189), 작업01(188), 짐승(188), 주먹(187), 평생
 (187), 평화01(187), 사정07(186), 정부08(185), 까닭(184), 흔적(184), 바보(183),
 열매01(183), 공기06(182), 이미지(182), 태양01(182), 한동안(182), 전설
 04(180), 똥(179), 사방02(179), 침묵(178), 변화(176), 용기02(176), 지붕(176),
 매표(173), 방식01(173), 기계06(171), 마디01(171), 먼지01(170), 군대02(168),
 음성02(168), 복도04(167), 출신(167), 한때(167), 운동자(166), 소나무(166), 동료
 (163), 연기09(163), 가족01(162), 농사01(162), 턱01(161), 대문03(160), 장면
 04(160), 장사01(160), 가치04(159), 사나이(159), 두려움(158), 입장04(158), 상
 상07(157), 마루03(156), 역할(156), 추억(156), 인류01(155), 기도03(154), 불꽃
 01(154), 나비03(153), 반응(153), 정치03(153), 비닐(152), 뿌리(152), 가능성
 (151), 다행(151), 머릿속(150), 탑01(150), 진실02(149), 침01(149), 현상
 04(149), 문명03(148), 부15(148), 육체03(148), 자료03(148), 표현(148), 감각
 02(147), 고생(147), 본래(147), 조각01(147), 우주02(146), 은03(145), 행위
 (145), 교장03(144), 품01(143), 기대02(142), 명령01(142), 심장02(142), 조사
 27(142), 활동02(141), 신01(140), 한가운데(140), 강의02(139), 마음속(139), 유
 리창01(139), 지옥(139), 충격02(139), 분노(138), 인연03(138), 주인공(138), 호
 기심(137), 과학(136), 새01(136), 소원04(136), 환상02(135), 양18(134), 조건
 02(134), 기록02(133), 큰일01(133), 종일01(132), 발자국(131), 의지06(131), 심
 정01(130), 아픔(130), 유적08(130), 제도01(129), 권력02(128), 법01(128), 장난
 (128), 칸01(128), 가정06(127), 말투(127), 빠(127), 털(127), 감동02(126), 움직
 임(126), 정17(126), 개인02(125), 곳곳(125), 여관03(125), 뺨(124), 현대
 01(124), 웬일(123), 종교(123), 광경(122), 오늘날(122), 욕02(122), 지식02(122),

거지01(121), 끼01(121), 의견01(121), 일상04(121), 주민(121), 형님(121), 환경01(121), 경제04(120), 동화06(119), 상03(119), 바늘(118), 감독01(117), 어린애(117), 일종03(117), 사막03(116), 순식간(116), 탁자01(116), 구15(115), 농부01(115), 신화04(115), 한05(115), 흥분(115), 뉘모습(114), 동작03(113), 주제04(113), 학자01(113), 한국인(113), 실제02(112), 일행01(112), 하느님(112), 형편01(112), 농민(111), 머느리(111), 솜씨(111), 수준(111), 욕심(111), 천장02(111), 껌질(110), 전통05(110), 계곡01(109), 담01(109), 독자04(109), 머리칼(109), 양쪽(109), 원인02(109), 연구소(108), 모래01(107), 봉지06(107), 울음소리(107), 공사02(106), 달빛(106), 산업(106), 현관01(106), 가지01(105), 책임(105), 총03(105), 꿈03(104), 무대05(104), 물결(104), 부대07(104), 씨앗(104), 주장03(104), 발걸음(103), 정신적(103), 회장06(103), 그리움(102), 맘01(102), 배경01(102), 통09(102), 난리02(101), 발길(101), 안방02(101), 영향04(101), 결론02(100), 구조08(100), 그늘(100), 담당(100), 실력02(100), 원숭이(100), 전국03(100), 커튼(100), 피해01(100), 해방05(100)

숙달도 단계 등급에서 고급으로 분류되는 일반 명사 어휘의 수는 초급과 중급에 비해 현저히 적게 나타났다. 고급 단계의 어휘는 총 67개로 전체 954개의 고빈도 일반 명사 어휘에서 7.02% 정도에 그침을 확인할 수 있었다. 또한 『한국어기초사전』에서 고급 어휘 표제어로 분류된 5,106개의 어휘 중 1.31%만이 고빈도 일반 명사 어휘 목록에 포함되었음을 확인하였다.

이러한 양상은 앞서 언급하였듯이 『한국어기초사전』의 등급별 표제어 수와 고빈도 일반 명사 어휘의 숙달도 단계 등급별 수가 반비례 관계에 있는 것에서 기인한다고 판단할 수 있다. 그러나 사전 등재 표제어와 계량적으로 선정한 어휘의 등급별 어휘 수 차이를 단순히 일차원적인 관점에서 개수와 분포에 따라 비교할 수는 없다. 이는 한국어 학습자를 위한 기초 사전의 표제어 선정 역시 대규모의 코퍼스를 토대로 이루어진 결과이며 선정 과정에서 숙달도에 따른 일정 수준의 어휘 규모를 고려하였다 해도 최종 표제어 선정 결과에 큰 영향을 주었다고 판단하기는 어렵기 때문이다. 따라서 『한국어기초사전』의 등급별 표제어 수와 고빈도 일반 명사 어휘의 숙달도 단계 등급별 수에 나타나는 차이는 본 연구가 설정한 출현 빈도 100회 이상의 범위보다 출현 빈도를 낮추어 연구 범위를 확장하면 할수록 줄어들 것이라고 예측할 수 있다.

또한 『한국어기초사전』의 등급별 표제어 수는 일반 명사와 더불어 다른 품사의 어휘도 함께 포함되어 분류된 결과이므로 이러한 점을 염두에 둔다면 둘의 차이를 수치만 놓고 비교할 수 없다는 것을 알 수 있다. 즉 본 연구가 제시한 『한국어기초사전』과 고빈도 일반 명사 어휘 목록의 숙달도 단계 등급별 비교는 한국어 문학 텍스트에 고빈도로 출현한 일반 명사 어휘들의 숙달도에 따른 양상을 파악하기 위한 자료라고 볼 수 있다.

숙달도 단계 등급별 분류에서 고급에 해당되는 67개 일반 명사 어휘를 제시하면 다음과 같다.

<표 4> 고급 단계 등급 어휘 목록

한국어 문학 텍스트 고빈도 일반 명사 어휘 - 고급(67개)
신09(574), 엄지01(378), 군05(343), 자체02(297), 차례01(295), 탓(287), 발01(266), 기11(253), 참새01(235), 장군04(231), 형사02(224), 새택(211), 허공(209), 선09(206), 풀01(198), 과정01(195), 장29(184), 도깨비(183), 예전01(176), 노릇01(174), 총각01(169), 종09(168), 파04(163), 시체02(161), 어미01(159), 의05(159), 고함01(155), 아비01(154), 중년01(149), 꼬마(148), 터01(142), 가01(140), 황금02(139), 잠자리01(138), 생02(137), 성06(135), 황제02(135), 논01(134), 골짜기01(129), 양반03(127), 까치01(126), 의원06(125), 괴물(123), 전투(120), 여사02(119), 백성(118), 처지(118), 들판01(116), 금02(114), 혼02(114), 지상01(113), 신음01(112), 파04(112), 철학(110), 인02(108), 마누라01(107), 당12(106), 무리01(106), 시기04(106), 귀족01(105), 차림(105), 피리01(104), 광장02(102), 국도04(101), 대07(101), 검03(100), 수화기(100)

한국어 문학 텍스트의 고빈도 일반 명사 어휘 목록 가운데 『한국어기초사전』에 등재된 표제어지만 ‘등급 없음’으로 분류되어 어휘 등급을 파악하지 못한 어휘는 총 41개이며 전체 954개 어휘에서 4.30%를 차지하는 것으로 나타났다. 이는 향후 한국어 학습자를 위한 사전을 개발·편찬하거나 한국어 문학 교수·학습 자료를 개발할 때 참고할 수 있는 목록으로 활용될 수 있다.¹⁵⁾

15) 한국어 문학 텍스트 일반 명사 어휘 목록에서 ‘등급 없음’으로 분류된 41개 어휘를 숙달도 단계 등급별로 분류하기 위해서는 연구자의 주관적인 직관보다 객관적이고 타당한 분류 기준이 반드시 마련되어야 하고 검증 가능한 단계들을 거

<표 5> 등급이 제시되지 않은 어휘 목록

한국어 문학 텍스트 고빈도 일반 명사 어휘 - 등급 없음(41개)
하나(1254), 혜성01(509), 선생01(425), 파리01(359), 서로01(305), 을01(276), 병사02(272), 동자승(269), 원시인(258), 성07(257), 한번(244), 임금01(207), 서방02(182), 태자(175), 영감01(161), 섹스(160), 에밀레종(158), 골04(141), 사태05(137), 제국02(137), 무아(136), 전생02(134), 이빨(126), 국민학교(125), 감나무(120), 갈매기(119), 강우02(117), 하나님(113), 명배우(112), 사모님(111), 계집애(110), 밥상(109), 무명03(108), 종03(108), 사령관(107), 순사02(104), 왕비(104), 가슴속(101), 난쟁이(101), 사과나무(100), 은하(100)

4.2. 의미 범주별 일반 명사 분류

본 연구의 의미 범주별 고빈도 일반 명사 어휘 분류는 『한국어기초사전』에 등재된 표제어 의미 범주 분류를 참조하였다. 한국어 문학 코퍼스를 기반으로 선정한 고빈도 일반 명사 어휘 총 954개를 14개의 대범주와 139개의 소범주로 분류하였고 분류 과정에서 어휘가 포함되지 않는 범주는 제외하였다.¹⁶⁾ 또한 전체 954개 어휘 가운데 『한국어기초사전』에서 범주별로 분류되지 않은 ‘분류 없음’ 어휘가 총 229개임을 확인하였는데 해당 어휘는 어휘의 사전적 의미를 파악하여 직접 해당되는 의미 범주에 추가하고 그 양상을 살폈다.

고빈도 일반 명사 어휘를 『한국어기초사전』을 참조하여 의미 범주에 따라 분류한 결과는 다음과 같다.¹⁷⁾ 먼저 대범주 ‘인간’으로 분류된 어휘는 총 149개이며 이를 ‘인간’의 하위 범주인 14개 소범주에 따라 분

쳐야하므로 본 연구는 해당 어휘들을 초·중·고급으로 분류하지 않고 제시만 하였다.

16) 분류 과정에서 제외된 범주는 인간(2개, 신체 변화, 신체에 가하는 행위), 식생활(1개, 채소), 의생활(4개, 옷감, 의생활 관련 장소, 의복 착용 행위, 미용 행위), 주생활(2개, 주거 형태, 주거 상태), 사회 생활(2개, 사회 행사, 사회 생활 상태), 경제 생활(2개, 경제 산물, 경제 상태), 교육(1개, 학문 용어), 종교(1개, 종교 활동 도구), 정치와 행정(2개, 공공 기관, 정치 및 치안 상태), 자연(1개, 재해), 동식물(2개, 동식물 행위, 동물 소리), 개념(5개, 성질, 온도, 빈도, 지시, 접속)이다.

17) 분류가 이루어지지 않은 어휘 229개를 제외한 725개의 분류 결과이다.

류하면 ‘신체 부위’, ‘사람의 종류’, ‘감정’에 해당되는 어휘가 비교적 많이 포함되는 것을 확인하였다. 대범주 ‘삶’으로 분류된 어휘는 총 62개가 있었다. ‘삶’의 소범주에 따라 살펴보면 ‘친족 관계’의 의미를 지닌 어휘가 29개로 전체 어휘의 절반 가까운 수를 차지하였다.

생활과 관련된 대범주는 ‘식생활’, ‘의생활’, ‘주생활’, ‘사회 생활’, ‘경제 생활’로 구분되었다. 대범주 ‘식생활’로는 총 21개의 어휘가 분류되었고 소범주별로 해당되는 어휘 수에 큰 차이가 발견되지 않았다. 대범주 ‘의생활’ 관련 어휘는 총 12개로 소범주에 따라 ‘모자, 신발, 장신구’, ‘옷 종류’, ‘옷의 부분’, ‘의복 착용 상태’로 각각 어휘가 분류되었으며 대부분의 어휘가 ‘모자, 신발, 장신구’에 분포되어 있음을 확인할 수 있었다. 대범주 ‘주생활’에 해당되는 어휘는 총 45개이며 ‘방07’, ‘마당’, ‘마루03’ 등 ‘주택 구성’의 의미를 지닌 어휘가 비교적 많이 분류됨을 볼 수 있었다. 대범주 ‘사회 생활’로 분류된 어휘는 총 100개로 나타났다. 이를 ‘사회 생활’의 14개 소범주별로 구분하여 살펴본 결과 ‘언어 행위’, ‘사회 활동’, ‘교통 수단’, ‘인간관계’의 의미로 분류되는 어휘가 절반 이상을 차지하였다. 대범주 ‘경제 생활’은 총 14개 어휘가 포함되었고 소범주 ‘경제 행위 장소’, ‘경제 행위 주체’, ‘경제 수단’, ‘경제 행위’에 따라 비교적 균등하게 분류되었음을 확인할 수 있었다.

‘교육’, ‘종교’, ‘문화’, ‘정치와 행정’ 관련 어휘는 총 77개로 각각의 대범주에 따라 살펴보면 다음과 같다. 대범주 ‘교육’은 총 27개의 어휘로 구성되었고 ‘교수 학습 주체’, ‘교수 학습 행위’, ‘전공과 교과목’, ‘교육 기관’, ‘학교 시설’, ‘학습 관련 사물’, ‘학문 행위’의 소범주에 따라 일반적인 분포 양상을 보였다. 대범주 ‘종교’로 분류된 어휘는 총 10개였으며 소범주별로 살펴보면 ‘신앙 대상’ 3개, ‘종교 유형’, ‘종교어’로 각 2개, ‘종교 활동 장소’, ‘종교인’, ‘종교 행위’로 각 1개 어휘가 분류됨을 확인하였다. 대범주 ‘문화’에 해당되는 어휘는 총 20개로 나타났고 9개 소범주 ‘예술’, ‘전통 문화’, ‘문화 활동 주체’, ‘문학’, ‘문화 생활 장소’, ‘음악’, ‘미술’, ‘대중 문화’, ‘문화 활동’으로 각각 분류되었다. 대범주 ‘정치와 행정’으로 분류된 어휘는 총 20개가 있었다. 소범주를 기준으로 세부적인 분류를 실시한 결과 ‘정치 및 행정 주체’, ‘사법 및 치안 주체’, ‘사법 및 치안 행위’, ‘정치 및 행정 행위’, ‘무기’로 분류되었

고 ‘정치 및 행정 주체’ 소범주로 분류된 어휘가 가장 많았다.

‘자연’과 ‘동식물’에 해당되는 의미를 지닌 어휘는 총 71개가 있었다. 대범주 ‘자연’으로는 총 37개의 어휘가 분류되었다. 이를 5개 소범주에 따라 분류한 결과 ‘지형’, ‘기상 및 기후’, ‘천체’의 의미를 지닌 어휘가 비교적 많이 나타났다. 대범주 ‘동식물’로 분류된 어휘는 총 34개가 있었고 소범주로는 ‘동물류’에 해당되는 의미를 지닌 어휘가 가장 많이 나타남을 파악하였다.

마지막 대범주인 ‘개념’으로는 총 174개 어휘가 분류되었고 14개 소범주에 따라 세부적으로 분류되었다. 대범주 ‘개념’의 경우 많은 어휘가 분류된 상위 두 소범주 ‘시간’과 ‘위치 및 방향’에 해당되는 어휘가 과반수를 차지하는 양상을 보였다.

이상의 의미 범주별 일반 명사 어휘 분류 결과를 표로 정리하면 다음과 같다.

<표 6> 의미 범주가 제시된 일반 명사 어휘 분류 결과 - 725개 어휘

구분	한국어 문학 텍스트 일반 명사 어휘
대범주	인간(149개, 20.55%)
신체 부위 (37)	눈01, 얼굴01, 손01, 몸01, 머리01, 입, 고개01, 가슴01, 발01, 귀01, 다리01, 어깨01, 목01, 팔01, 엄지01, 입술, 배01, 손가락, 허리01, 코01, 무릎, 온몸, 등01, 혀01, 이마01, 살01, 엉덩이, 주먹, 눈동자, 턱01, 손바닥, 머리카락, 육체03, 품01, 털, 뺨, 머리칼
사람의 종류 (36)	사람, 여자02, 아이01, 남자01, 인간01, 아저씨, 애02, 소녀02, 사내01, 어른01, 아기01, 소년01, 노인01, 여인01, 청년, 여성01, 아주머니, 처녀, 인물, 젊은이, 새댁, 아가씨, 아줌마, 어린이01, 바보, 총각01, 사나이, 꼬마, 주인공, 남녀, 주민, 여사02, 어린애, 한국인, 부자06, 독자04
감정 (27)	사랑01, 기분01, 느낌, 걱정, 화05, 관심01, 감정06, 겁02, 희망, 기쁨, 슬픔, 행복02, 욕망, 공포08, 두려움, 분노, 호기심, 재미01, 심정01, 감동02, 정17, 긴장, 한05, 흥분, 욕심, 책임, 그리움
인지 행위 (10)	생각01, 마음01, 뜻, 기억02, 계획01, 경험, 가치04, 상상07, 충격02, 이해05
생리 현상	목소리, 눈물01, 숨01, 땀01, 한숨02, 음성02, 침01, 울음소리

감각	소리01, 정신12, 냄새, 고통, 의식03, 신경04, 감각02
신체 행위	웃음, 미소05, 비명02, 울음, 고함01, 움직임
능력	기11, 능력02, 가능성, 숨쉴, 실력02
태도	자세02, 태도03, 용기02, 의지06
신체 내부 구성	피02, 장29, 심장02, 뼈
체력 상태	힘01, 건강03
용모	표경02
성격	성격02
소리	신음01
대범주	삶(62개, 8.55%)
친족 관계 (29)	어머니01, 아버지, 아내01, 엄마, 할머니, 남편01, 아빠, 할아버지, 아들, 딸01, 오빠, 형01, 자식01, 동생01, 가족01, 언니, 부부03, 부인01, 식구01, 부모01, 누나01, 부모님, 어머니01, 아버지01, 형제01, 형님, 며느리, 마누라01, 어머니님
가족 행사	결혼, 차례01, 신랑02, 생일02, 신부10
여가 도구	게임, 놀이01, 인형01, 공01, 피리01
여가 활동	노래01, 여행02, 운동02, 풍경01, 구경01
일상 행위	꿈01, 잠01, 걸음, 버릇01
병과 증상	병04, 상처02, 기운01, 아픔
삶의 상태	자유03, 중년01, 생02, 처지
여가 시설	공원03, 운동장
삶의 행위	죽음01
치료 행위	환자03
치료 시설	병원02
약품류	약07
대범주	식생활(21개, 2.90%)
음료	물01, 술01, 커피, 맥주, 소주04
음식	담배, 밥01, 음식
식생활 관련 장소	부엌, 식당, 카페
조리 도구	칼01, 그릇01, 접시
과일	배03, 과일01
식재료	고기01, 얼음01
곡류	쌀
식사 및 조리 행위	식사03
맛	맛01

대범주	의생활(12개, 1.66%)
모자, 신발, 장신구	구두01, 가방01, 우산01, 신01, 신발, 안경03, 모자07
옷 종류	옷01, 바지01, 치마01
옷의 부분	주머니
의복 착용 상태	차림
대범주	주생활(45개, 6.21%)
주택 구성 (23)	방07, 문04, 벽05, 층01, 바닥01, 창문, 마당, 창07, 계단04, 화장실, 지붕, 방문02, 대문03, 마루03, 지하, 유리창01, 거실02, 엘리베이터, 정원06, 천장02, 담01, 현관01, 안방02
생활 용품 (12)	컴퓨터, 침대02, 거울01, 의자03, 책상01, 텔레비전, 시계01, 이불01, 탁자01, 카메라, 수화기, 커튼
가사 행위	불01, 빵01, 유리07
건물 종류	집01, 건물03, 아파트
주거 지역	마을01, 동네, 시내03
주거 행위	이사13
대범주	사회 생활(100개, 13.79%)
언어 행위 (20)	말01, 이야기, 얘기, 글, 대답, 의미02, 말슴, 소문02, 질문, 대화06, 거짓말, 침묵, 설명, 표현, 명령01, 농담01, 욕02, 의견01, 주장03, 큰소리01
사회 활동 (15)	관계05, 약속, 인사02, 싸움, 예14, 절02, 도움, 만남, 노릇01, 선물03, 연락02, 활동02, 덕분, 부탁, 소개02
교통 수단 (11)	차06, 배02, 버스02, 오토바이, 택시, 비행기, 열차02, 자동차, 트럭, 기차01, 자전거
인간관계	친구02, 남01, 선배, 이웃, 동료, 애인02, 인연03, 후배05, 일행
교통 이용 장소	길01, 거리01, 입구02, 다리02, 공항02, 역11, 건너편, 국도04
직위	사장15, 공주01, 부장06, 직원03, 교장03, 양반03, 귀족01, 회장
매체	책01, 기사10, 신문10, 비디오, 프로그램, 방송01, 뉴스
직업	기자04, 의사12, 직업, 선수05, 농부01
직장 생활	지원02, 사무실, 작업01, 회의03, 서류02
소통 수단	전화07, 편지02, 언어01, 글씨
통신 행위	소식04, 광고02, 신호01
말	한마디, 단어
직장	회사04, 직장05
교통 이용 행위	사고12

대범주	경제 생활(14개, 1.93%)
경제 행위 장소	가게, 시장04, 밭01, 공장02, 논01
경제 수단	돈01, 값, 카드
경제 행위	장사01, 사업04, 경제04
경제 행위 주체	주인01, 손님, 농민
대범주	교육(27개, 3.72%)
교수 학습 주체	선생님, 학생, 교수06, 여학생, 학자01, 대학생, 박사01
교수 학습 행위	공부01, 교육, 과정01, 시험03, 강의02, 수업03
전공과 교과목	역사03, 과04, 과학, 철학
교육 기관	학교, 대학01, 고등학교, 중학교
학교 시설	반08, 도서관, 연구소
학문 행위	연구03
학습 관련 사물	종이01, 자료03
대범주	종교(10개, 1.38%)
신앙 대상	신09, 부처01, 하느님
종교어	천사04, 지옥
종교 유형	불교, 종교
종교인	스님
종교 행위	기도03
종교 활동 장소	교회02
대범주	문화(20개, 2.76%)
예술	사진06, 작품01, 춤01, 예술
전통 문화	전설04, 유적08, 신화04, 전통05
문학	소설03, 시09, 문학01
문화 활동 주체	시인10, 작가01, 감독01
문화 생활 장소	극장, 박물관
대중 문화	영화01
문화 활동	문화01
미술	그림01
음악	음악01
대범주	정치와 행정(20개, 2.76%)
정치 및 행정 주체	왕04, 대통령, 국가01, 국민, 정부08, 황제02, 의원06, 시민, 당
사법 및 치안 주체	군05, 장군04, 형사02, 경찰04, 군인
사법 및 치안	죄03, 군대02, 법01

행위	
정치 및 행정 행위	정치03, 권력02
무기	총03
대범주	자연(37개, 5.10%)
지형 (15)	땅01, 산01, 바다, 숲01, 강01, 섬03, 언덕, 강물, 바닷가, 가01, 골짜기01, 들01, 사막03, 호수07, 계곡01
기상 및 기후	바람01, 비01, 햇살, 구름01, 파도, 눈04, 안개, 날씨01
천체	하늘01, 해01, 별01, 지구04, 햇빛, 태양01, 우주02, 달빛
지표면 사물	돌02, 바위01, 흙01, 모래01
자원	은03, 황금02
대범주	동식물(34개, 4.69%)
동물류 (19)	새03, 고양이, 개03, 호랑이, 소03, 쥐02, 동물, 참새01, 닭, 용09, 짐승, 돼지, 뱀, 물고기, 까치01, 코끼리, 곰03, 강아지, 원숭이
식물의 부분	꽃01, 열매01, 뿌리, 잎01, 가지01, 씨앗
식물류	나무01, 풀02, 소나무, 장미05
동물의 부분	날개01, 꼬리01, 가죽01
곤충류	잠자리02, 나비03
대범주	개념(174개, 24.00%)
시간 (66)	때01, 뒤01, 그때, 동안01, 지금02, 오늘, 밤01, 순간03, 아침, 이제01, 이번01, 하루01, 시대02, 결국, 시절01, 그날, 저녁, 옛날, 한참, 당시02, 겨울, 요즘, 새벽01, 여름01, 가을01, 세월02, 내일, 동시02, 봄01, 학년, 과거02, 당장02, 기회03, 틈01, 오후02, 이때, 어제01, 세기03, 평소, 미래02, 평생, 다음날, 한동안, 낮, 예전01, 점심, 한때, 그동안, 오랫동안, 오랜만, 최근, 종일01, 현재02, 사흘, 현대01, 아까, 오늘날, 일주일, 순식간, 어젯밤, 올해, 이틀01, 기간01, 시기04, 일요일, 작년
위치 및 방향 (36)	속01, 앞, 위01, 안01, 자리01, 사이01, 옆, 밖, 아래01, 밑01, 주위02, 가운데, 결01, 주변04, 한쪽, 방향01, 바깥, 근처, 중심01, 여기저기, 거리07, 남쪽, 동쪽, 사방02, 왼쪽, 오른쪽, 중간01, 중앙01, 서쪽, 터01, 반대02, 북쪽, 한가운데, 위치01, 양쪽, 광장02
지역 (16)	곳01, 세상01, 나라01, 세계02, 도시03, 고향02, 공간05, 지역02, 시골, 지방04, 구석01, 우리나라, 장소05, 외국02, 곳곳, 전국03
세는 말	날01, 시간04, 잔02, 달05, 공기06, 짝01, 마디01, 종09, 병05,

(13)	상자08, 끼01, 인02, 봉지06
정도	정도11, 이상04, 부분01, 차이, 최고02, 약간, 일부02, 수준, 제일04
순서	전07, 처음, 다음01, 끝01, 마지막, 이후02, 나중01, 시작01, 이전03
모양	모습01, 모양02, 키01, 풀01, 크기, 금02, 높이01
양	조금01, 반05, 대부분, 전체01, 나머지, 양18
수	모두01, 숫자, 수23
의문	얼마, 며칠, 아무것
밝기	빛, 어둠
색깔	색03, 색깔
속도	속도01
인칭	자신01
합계	725개, 100%

한국어 문학 텍스트의 고빈도 일반 명사 어휘 954개 중 『한국어기초사전』에서 의미 범주별로 분류된 결과를 찾을 수 있는 어휘는 총 725개로 앞서 제시한 분류 결과와 같다. 그러나 이를 제외한 229개의 ‘분류 없음’ 어휘는 『한국어기초사전』에 등재된 표제어지만 범주에 따른 분류가 이루어지지 않았음을 확인하였다.¹⁸⁾ 이에 따라 본 연구는 229개 어휘의 사전적 의미를 파악한 뒤 직접 해당되는 대범주와 소범주로 분류하여 그 양상을 살펴보았다. 229개 어휘의 의미 범주별 분류 결과와 양상을 제시하면 다음과 같다.

먼저 대범주 ‘인간’으로 총 49개의 어휘가 분류되었다. ‘인간’에 포함된 기존의 소범주는 16개인 반면 ‘분류 없음’에 해당되는 어휘를 분류하면서 어휘가 포함되지 않는 소범주를 제외하자 총 10개의 소범주가 남게 되었다. ‘인간’에서 가장 많은 어휘가 분류된 소범주는 ‘인지 행위’임을 확인하였다. 대범주 ‘삶’으로는 총 25개 어휘가 분류되었고 기존의 소범주 12개에서 8개가 제외되어 ‘삶의 상태’, ‘삶의 행위’, ‘일상 행위’, ‘친족 관계’ 4개의 소범주 각각에 개별 어휘가 분류되었다. 특징적

18) 『한국어기초사전』에 등재된 표제어지만 범주별로 분류되지 않은 어휘 229개에는 숙달도 단계 등급별 어휘 분류에서 어휘 등급을 파악하지 못한 41개 어휘가 전부 포함되어 있다.

으로는 ‘삶의 상태’ 소범주에 해당되는 어휘가 대범주 ‘삶’으로 분류된 전체 어휘 수의 절반을 약간 넘는 것을 확인할 수 있었다.

어휘를 분류하는 과정에서 ‘식생활’과 ‘의생활’ 대범주에는 단 하나의 어휘도 분류되지 않는 결과가 나타나 제외하였고 ‘주생활’의 의미 범주로는 총 15개의 어휘가 분류되었다. ‘주생활’의 경우 기존에 8개 소범주가 포함되어 있었지만 분류를 하는 과정에서 해당되는 어휘가 없는 4개의 소범주를 제외하여 ‘생활 용품’, ‘건물 종류’, ‘주거 지역’, ‘주택 구성’의 범주만 남게 되었다. 또한 ‘사회 생활’과 ‘경제 생활’로 분류된 어휘는 각각 22개와 4개가 있었고 특히 ‘사회 생활’은 기존의 16개 소범주에서 5개로 대폭 축소되는 양상을 보였다.

‘교육’, ‘종교’, ‘문화’, ‘정치와 행정’ 범주로는 각각 4개, 8개, 5개, 11개 어휘가 분류되었다. 특히 ‘정치와 행정’의 대범주로 분류된 어휘들을 살펴보면 ‘임금01’, ‘태자’, ‘왕비’, ‘순사02’, ‘검03’과 같이 시대를 반영하는 어휘가 특징적으로 나타남을 확인할 수 있다.

마지막 대범주인 ‘개념’으로 분류된 어휘의 수는 총 68개로 『한국어 기초사전』에 의미 범주가 제시되지 않은 229개 일반 명사 어휘 중 가장 많은 수의 어휘가 분류되었다. 이를 소범주에 따라 세부적으로 살펴보면 소범주 ‘성질’과 ‘모양’으로 분류된 어휘의 수가 각각 18개, 17개로 ‘개념’에 해당되는 68개 어휘의 과반수를 차지하고 있음을 확인할 수 있었다. 이상의 분류 결과를 표로 정리하여 제시하면 다음과 같다.

<표 7> 의미 범주가 제시되지 않은 일반 명사 어휘 분류 결과 - 229개 어휘

구분	한국어 문학 텍스트 일반 명사 어휘
대범주	인간(49개, 21.40%)
인지 행위 (15)	사실04, 사건01, 존재, 눈치, 분위기, 눈길01, 비밀, 운명01, 추억, 다행, 머릿속, 마음속, 환상, 지식02, 가슴속
사람의 종류	원시인, 시체02, 영감01, 거지01, 계집애, 난쟁이
태도	경우02, 준비, 변화, 목적03, 기대02, 소원04
능력	방법, 기술01, 방식01, 조건02, 노력01
신체 행위	짓01, 행동, 동작03, 발걸음, 발길
감각	시선03, 향기01, 이미지, 정신적
감정	필요, 눈빛01, 잘못, 맘01

신체 내부 구성	영혼02, 혼02
체력 상태	목숨
생리 현상	똥
대범주	삶(25개, 10.92%)
삶의 상태 (13)	삶, 나이01, 생명, 인생01, 여유, 평화01, 짐01, 출신, 고생, 무아, 형편01, 피해01, 해방05
삶의 행위	일01, 생활, 큰일01, 웬일
일상 행위	섹스, 행위, 장난, 일상04, 공사02
친족 관계	집안01, 민족, 서방02
대범주	주생활(15개, 6.55%)
생활 용품	물건, 매01, 비닐, 상03, 바늘, 밥상, 통09
건물 종류	성07, 호텔, 탑01, 여관03
주거 지역	가정06, 부대07
주택 구성	복도04, 잠자리01
대범주	사회 생활(22개, 9.61%)
언어 행위	문제06, 이유04, 탓, 사정07, 까닭, 말투, 물론01
사회 활동	사회07, 대신03, 입장04, 역할, 반응, 담당
인간관계	서로01, 상대04, 대상11, 파04
소통 수단	화면05, 정보06, 기록02
직위	사모님, 종03
대범주	경제 생활(4개, 1.75%)
경제 행위	농사01, 산업
경제 수단	부15
경제 행위 주체	기계06
대범주	교육(4개, 1.75%)
교수 학습 주체	선생01
교수 학습 행위	방학
교육 기관	국민학교
학문 행위	조사27
대범주	종교(8개, 3.49%)
종교어	귀신01, 도깨비, 에밀레종, 전생02, 괴물, 무덤
종교인	동자승
신앙 대상	하나님
대범주	문화(5개, 2.18%)
문화 활동 주체	명배우

문화 생활 장소	무대05
전통 문화	대07
문학	동화06
예술	연기09
대범주	정치와 행정(11개, 4.80%)
정치 및 행정 주체	임금01, 태자, 백성, 왕비
사법 및 치안 주체	병사02, 사령관, 순사02
정치 및 행정 행위	전쟁, 전투
사법 및 치안 행위	제도01
무기	검03
대범주	자연(9개, 3.93%)
지형	자연01, 골04, 환경01, 지상01, 물결
기상 및 기후	강우02, 그늘
천체	혜성01, 은하
대범주	동식물(9개, 3.93%)
동물류	새끼02, 인류01, 갈매기
식물류	감나무, 사과나무
동물의 부분	알01, 이빨
식물의 부분	껍질
곤충류	파리01
대범주	개념(68개, 29.69%)
성질 (18)	상황02, 상태01, 마찬가지로, 내용02, 자체02, 통일02, 선09, 의 05, 진실02, 문명03, 사태05, 성06, 실제02, 원인02, 배경01, 난리02, 영향04, 종류02
모양 (17)	점01, 형태, 구멍, 떼01, 면05, 그림자, 흔적, 먼지01, 장면04, 현상04, 조각01, 발자국, 칸01, 광경, 뒷모습, 무리01, 구조08
위치 및 방향	눈앞, 골목01, 너머, 도로07, 허공, 현장03, 구15
순서	원래01, 을01, 결과02, 본래, 결론
인칭	혼자01, 스스로, 개인02, 무명03
정도	보통, 새01, 진짜, 일종03
세는 말	하나, 줄01, 한번
지시	이름, 제목02, 주제04
밝기	불빛, 불꽃01
시간	후08, 현실02
의문	만약, 만일01

지역	제국02
합계	229개, 100%

4.3. 종합적 논의

본 연구의 분석 과정에서는 연구 대상인 한국어 문학 코퍼스를 기반으로 고빈도 일반 명사 어휘 총 954개의 목록을 작성하여 숙달도 단계 등급과 의미 범주에 따라 분류하였다. 분류 방법과 기준으로는 『한국어 기초사전』의 등재 표제어 목록을 참조하였다.

숙달도 단계 등급별 일반 명사 어휘 분류 결과 초급에 해당되는 어휘의 수는 총 500개로 전체 고빈도 일반 명사 어휘 목록의 52.41%를 차지하고 있었고 『한국어기초사전』의 초급 단계 표제어 1,942개에서 25.75%를 포함하고 있었다. 이는 약 154만 어절 규모의 한국어 문학 코퍼스에서 추출한 고빈도 일반 명사 어휘 목록 내에 『한국어기초사전』 초급 어휘 25.75%가 반영되어 있음을 의미한다. 중급에 해당되는 어휘는 총 346개로 36.27%를 차지하여 초급 단계에 비해 다소 적은 수의 어휘가 분류되었지만 『한국어기초사전』의 중급 단계 표제어 수와 비교하면 8.69%에 그쳤다. 또한 고급으로 분류된 어휘는 총 67개로 고빈도 일반 명사 어휘 목록의 7.02%를 차지하고 있었고 『한국어기초사전』의 고급 단계 표제어 전체에서 1.31%에 해당되는 어휘만이 반영되어 있었다. 분류 결과를 살펴보면 고빈도 일반 명사 어휘 목록에서 가장 많은 분포를 차지하는 등급은 초급이었으며 중급과 고급의 순서로 점점 어휘의 수가 적어지는 것을 확인할 수 있다. 이는 『한국어기초사전』의 숙달도별 표제어를 얼마만큼 반영하고 있는지에 대한 수치에서도 동일하게 나타났는데 앞서 숙달도에 따른 분류 결과에서 지적하였듯이 이러한 차이는 본 연구가 한국어 문학 텍스트의 일반 명사 어휘 954개를 출현 빈도 기준으로 선정하였기 때문에 나타난 것으로 판단된다. 일반적으로 고빈도 어휘에 초급 어휘가 대다수 집중되어 있다는 점과 『한국어기초사전』에 표제어로 등재된 명사 어휘가 총 29,088개이고 본 연구의 한국어 문학 코퍼스를 구성하는 명사 어휘는 총 34,810

개임을 고려한다면 일반 명사 어휘 목록 선정의 범위를 확장시켜 비교할 어휘 수가 많아질수록 초급으로 분류되는 어휘에 비해 중급과 고급에 해당되는 어휘 수가 증가할 것으로 예상되고 이에 따라 한국어 문학 텍스트 일반 명사 어휘 목록에 대한 『한국어기초사전』의 숙달도별 표제어 반영 비율도 중·고급 어휘에서 높아질 것으로 판단된다. 그리고 954개의 고빈도 일반 명사 어휘 가운데 『한국어기초사전』에서 숙달도 등급이 부여되지 않은 어휘는 총 41개로 4.30%를 차지하고 있었는데 해당 어휘에 대해서는 향후 타당한 분류 기준과 검증 단계를 통한 분류 작업이 이루어질 필요가 있다고 판단하여 그 목록을 제시하였다.

의미 범주별 일반 명사 어휘 분류 결과 총 954개 어휘 중 725개 어휘(76%)가 『한국어기초사전』에서 의미 범주에 따라 분류되어 있음을 확인하였고 분류가 되어 있지 않은 나머지 229개 어휘(24%)는 ‘분류 없음’으로 구분하였다. 본 연구는 725개 어휘와 229개 어휘를 나누어 의미 범주별 양상을 파악하였다. 먼저 725개 일반 명사 어휘의 분류 결과를 살펴보면 대범주 ‘개념’, ‘인간’, ‘사회 생활’로 분류되는 의미를 지닌 어휘가 각각 100개 이상이었고 세 범주로 분류된 어휘의 합이 423개로 58.34%를 차지하고 있었다. 이는 한국어 문학 텍스트에 고빈도로 출현하는 일반 명사 어휘를 의미 범주에 따라 분류하였을 때 ‘시간’, ‘위치 및 방향’, ‘지역’ 등에 해당되는 ‘개념’ 관련 어휘, ‘신체 부위’, ‘사람의 종류’, ‘감정’ 등에 해당되는 ‘인간’ 관련 어휘, 그리고 ‘언어 행위’, ‘사회 활동’ 등을 비롯한 ‘사회 생활’ 관련 어휘가 전체 어휘 목록에서 절반 이상을 차지하고 있다는 것을 의미한다. 따라서 한국어 학습자에게 문학 관련 고빈도 일반 명사 어휘를 제시한다면 이러한 점을 고려할 수 있을 것이다. 반면 ‘종교’, ‘의생활’, ‘경제 생활’ 관련 어휘는 상대적으로 적게 분류되어 세 범주의 어휘 분포가 총 4.97%에 그쳤다. 이러한 분류 결과를 통하여 해당 의미 범주에 포함되는 일반 명사 어휘들이 한국어 문학 텍스트에는 고빈도로 출현하지 않는다는 것을 파악할 수 있고 이러한 사실을 하나의 정보로 활용할 수 있을 것이다. 다음으로 『한국어기초사전』에 표제어로 등재되어 있지만 의미 범주별 분류가 이루어지지 않은 229개 어휘를 대범주와 소범주로 각각 분류하여 결과를 살펴보았다. 분류 결과 ‘개념’의 대범주로 가장 많은 어휘가 분류되었

고 ‘인간’의 의미 범주가 뒤를 이었다. 상위 두 대범주로 분류된 어휘 수는 총 117개로 전체 229개 어휘에서 51.09%의 분포를 보였다. 반면 ‘식생활’과 ‘의생활’의 대범주로 분류된 어휘는 없었고 ‘경제 생활’과 ‘교육’의 대범주로 각각 4개의 어휘가 분류되어 가장 적은 어휘가 포함된 범주로 판정되었다. 이를 토대로 ‘분류 없음’의 229개 어휘를 의미 범주에 따라 분류한 결과와 『한국어기초사전』의 의미 범주별 분류를 참조한 725개 어휘의 분류 결과가 대체로 유사한 양상을 보이고 있음을 판단할 수 있다. 따라서 본 연구가 제시하는 한국어 문학 텍스트 고빈도 일반 명사 어휘의 의미 범주별 분류 결과를 활용하고자 한다면 725개 어휘의 분류 결과와 229개 어휘의 분류 결과를 통합하여 전체 954개 어휘의 의미 범주별 양상을 파악하는 것도 충분히 가능하리라 여겨진다.

5. 결론

본 연구는 한국어 문학 텍스트 전반에 출현하는 고빈도 일반 명사 어휘를 선정하고 숙달도 단계와 의미 범주를 중심으로 분석하였다. 한국어 교육 분야에서의 문학 관련 기존 논의는 주로 이론적인 측면에서 교수·학습 방안과 관련하여 이루어졌으며 문학 텍스트의 어휘를 선정하거나 분석하는 논의는 특정 문학 작품을 대상으로 한정된 텍스트와 제한적인 어휘를 위주로 다루었다. 이에 따라 본 연구는 보다 다양한 문학 텍스트에서 도출되는 실제성 있는 결과를 제시하고자 코퍼스를 기반으로 계량적인 접근법을 취하여 연구를 진행하였다.

연구 대상은 약 154만 어절 규모의 한국어 문학 코퍼스이며 해당 코퍼스의 어휘 빈도 정보를 활용하여 출현 빈도 100회 이상의 고빈도 일반 명사 어휘 총 954개를 선정한 뒤 어휘 목록을 작성하였다. 선정한 어휘의 숙달도 단계 등급과 의미 범주에 따른 분류는 『한국어기초사전』에 등재된 표제어 목록을 참조하였다.

한국어 문학 텍스트의 고빈도 일반 명사 어휘를 숙달도 단계 등급별로 분류한 결과 초급 어휘 500개, 중급 어휘 346개, 고급 어휘 67개와

분류되지 않은 어휘 41개로 어휘 목록이 구성되어 있음을 파악하였다. 이를 『한국어기초사전』의 숙달도별 표제어 수와 비교하여 그 양상을 살펴보았는데 초급 어휘는 25.75%, 중급 어휘는 8.69%, 고급 어휘는 1.31% 반영되어 있었다. 그러나 본 연구가 선정한 어휘들의 범위는 한국어 문학 코퍼스에 출현하는 전체 일반 명사 어휘에서 출현 빈도 100회 이상의 고빈도 어휘만을 포함하고 있으며 일반적으로 고빈도 어휘에는 초급 어휘가 집중된다. 그리고 『한국어기초사전』의 등급별 표제어는 일반 명사뿐만 아니라 다른 품사의 어휘도 포함되어 분류된 결과이다. 이러한 특성들을 근거로 선정 어휘의 범위를 확장할수록 숙달도 단계 등급별 어휘 수와 『한국어기초사전』 등급별 표제어 반영률에 차이가 있을 것을 예측하였다.

의미 범주에 따른 분류 결과 총 954개 일반 명사 어휘의 24%에 해당하는 229개 어휘가 『한국어기초사전』에서 의미 범주별로 분류되어 있지 않음을 확인할 수 있었다. 이를 토대로 725개 어휘 목록과 229개 어휘 목록을 나누어 725개의 어휘는 『한국어기초사전』을 참조하여 의미 범주별로 분류하여 제시하였고 나머지 229개는 사전적 의미를 파악하여 의미 범주별로 분류하였다. 이에 따라 결과적으로 두 어휘 목록의 분류 결과가 유사한 양상을 보인다고 판단하였다.

본 연구는 기존의 연구 범위를 발전시키기 위하여 코퍼스를 기반으로 한국어 문학 텍스트의 고빈도 일반 명사 어휘 목록을 작성하고 제시하였으며 해당 어휘 목록을 『한국어기초사전』의 표제어 목록과 비교하여 살펴보았다는 점에서 기존의 연구를 보완한다. 그러나 숙달도 단계 등급별 일반 명사 어휘 분류에서 『한국어기초사전』에 등급이 제시되지 않은 41개 어휘를 분류할 객관적인 기준과 검증 단계를 마련하지 못하여 해당 어휘들을 초·중·고급으로 분류하지 않고 단순히 제시하는 데 그쳤다는 한계가 있다. 또한 분석 결과를 실제로 활용할 수 있는 구체적인 방안까지 다루지 못하였다는 한계가 있다. 향후 연구에서는 이러한 한계를 보완하여 보다 유의미한 연구가 이루어지기를 기대한다.

본 연구가 제시한 어휘 목록과 분류 및 분석 결과는 한국어 문학 텍스트에 고빈도로 출현하는 일반 명사 어휘 양상을 파악할 수 있는 기초 자료가 될 수 있고 한국어 교육 현장에서 문학 관련 학습 자료 및 교수

자료를 개발하는 데 활용될 수 있다. 또한 더 나아가 한국어 학습자를 위한 사전을 개발하거나 편찬하는 과정 또는 기존의 사전을 보완하는 과정에서 참고할 수 있는 자료가 되기를 바란다.

[ABSTRACT]

Selection and analysis of common nouns based on Korean literature corpus

-Focused on vocabulary level and semantic category-

Shin Minsub(Yonsei University)

This research aims to analyze the aspect of usage of high-frequency common nouns which appear in overall Korean literature corpus. As part of this study, first, this study constructs a Korean literature corpus with approximately 1.54 million word phrases. Second, as the subjects of analysis, 954 common nouns are extracted in accordance with their frequency which is more than 100. Next, these nouns are classified with their vocabulary level and semantic categories. The categories refer to the list of headwords registered in Korean Basic Dictionary by National Institute of Korean Language. As a result of classification with the level, basic level accounts for 52.41%(500 nouns), intermediate level represents 36.27%(346 nouns), advanced level shows 7.02%(67 nouns), and 4.30%(41 nouns), though they are registered in the dictionary, cannot be identified by their level. With regard to the result of semantic classification, 76% of nouns among 954 nouns, which are a total of 725 nouns are able to be classified according to Korean Basic Dictionary. The rest of 229 nouns are classified with their own semantics as they are not included certain semantic category of the dictionary although they are registered headwords. The analysis of this research suggests the material which enables to identify the aspect of common nouns in Korean literature corpus. Furthermore, the results are

able to be applied to develop learning material involved in literature and compile and revise Korean dictionaries.

Key words : Korean language education, Korean literature, corpus, common noun, Korean Basic Dictionary, vocabulary level, semantic category

[참고문헌]

■기본자료

교보문고 (<http://www.kyobobook.co.kr>)

국립국어원 『표준국어대사전』 (<http://stdict.korean.go.kr>)

국립국어원 『한국어기초사전』 (<http://krdict.korean.go.kr>)

■단행본

이충우, 『한국어 교육용 어휘 연구』, 국학자료원, 1994.

■논문

강현화·원미진, 「한국어학습자를 위한 『한국어기초사전』 구축 방안 연구」, 『한국사전학』 제20권, 한국사전학회, 2012, 7-30면.

강현화·원미진, 「언어학습자를 위한 『한국어기초사전』 편찬의 원리와 실제」, 『민족문화 연구』 제67권, 고려대학교 민족문화연구원, 2015, 39-63면.

박배식, 「한국어 문학교육의 실행 방안」, 『현대문학이론연구』 제37권, 현대문학이론학회, 2009, 311-336면.

박영미, 「〈소나기〉를 활용한 단편소설 한국어 교재 개발」, 『인문과 예술』 제7권, 인문 예술학회, 2019, 127-147면.

박현진·니콜라프라스키니, 「한국어 학습자의 문학 텍스트 이해 실패 요인 분석」, 『이중 언어학』 제51권, 이중언어학회, 2013, 111-149면.

신윤경, 「한국어 교육을 위한 문학 텍스트 연구: 문학 텍스트 선정기준과 교수 방법을 중심으로」, 고려대학교 대학원 박사학위논문, 2008.

신윤경, 「한국문학 읽기를 위한 어휘 주석 연구 -사회·문화적 배경지식 및 맥락 이해 추가 설명을 중심으로-」, 『한국언어문화학』 제11권 1호, 국제한국언어문화학회, 2014, 43-62면.

윤여탁, 「문학을 활용한 한국어 교육 방법」, 『국어교육연구』 제6권, 서울대학교 국어교육연구소, 1999, 239-256면.

윤여탁·유영미·박은숙, 「한국어교육에서 한국문학 정전 목록 선정 연구 -한국, 미국, 중국을 중심으로-」, 『국어교육연구』 제34권, 서울대학교 국어교육연구소, 2014, 0-29면.

장연봉홍, 「한국어 교육을 위한 문학 작품 선정 기준 연구 -코퍼스 분석에 기반한 명사의 어휘 난이도를 중심으로-」, 『한국언어문화학』 제116권, 한국언어문화학회, 2021, 107-131면.

- 주재환, 「한국어 수준별 읽기 교재 개발을 위한 문학 텍스트 어휘 분석 연구」, 『언어와 문화』 제10권 1호, 한국언어문화교육학회, 2014, 254-273면.
- 한영균, 「한국어 어휘 교육·학습 자료 개발을 위한 계량적 분석의 한 방향」, 『어문학』 제94권, 한국어문학회, 2006, 119-146면.
- 한영균, 「한국어 어휘 교육과 어휘 통계 정보」, 『한국어 교육』 제18권 3호, 국제한국어 교육학회, 2007, 249-272면.
- 황인교, 「외국어로서의 한국문학교육의 가능태」, 『외국어로서의 한국어교육』 제26권, 연세대학교언어연구교육원 한국어학당, 2001, 409-434면.