



## 영어교과서 코퍼스 분석 기반 2022 개정 영어과 교육과정 기본 어휘의 학년별 수준 구분

Seoyong Choi · Kitaek Kim · Sun-Young Oh (Seoul National University) · Min-Chang Sung (Gyeongin National University of Education)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: December 9, 2025

Revised: March 23, 2026

Accepted: April 12, 2026

Choi, Seoyong (First author)  
Doctoral Student, Department of English Language Education  
Seoul National University  
Email: [solideogloria0614@snu.ac.kr](mailto:solideogloria0614@snu.ac.kr)

Kim, Kitaek (Corresponding author)  
Professor, Department of English Language Education & Learning Sciences Research Institute  
Seoul National University  
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Seoul, Korea  
Tel: +82-2-880-7680  
Email: [kitaek@snu.ac.kr](mailto:kitaek@snu.ac.kr)

Oh, Sun-Young  
Professor, Department of English Language Education & Learning Sciences Research Institute  
Seoul National University  
Email: [sunoh@snu.ac.kr](mailto:sunoh@snu.ac.kr)

Sung, Min-Chang  
Associate Professor, Department of English Education  
Gyeongin National University of Education  
Email: [mcsung@ginue.ac.kr](mailto:mcsung@ginue.ac.kr)

\* This work was supported by the Learning Sciences Research Institute at Seoul National University (0767-20250011).

### ABSTRACT

Choi, Seoyong, Kitaek Kim, Sun-Young Oh and Min-Chang Sung. 2026. Grade-level classification of the basic vocabulary in the 2022 revised English curriculum: A corpus-based analysis of English textbooks. *Korean Journal of English Language and Linguistics* 26, 546-563.

This study aimed to classify the basic vocabulary list specified in South Korea's 2022 revised English curriculum into detailed grade levels through a corpus-based analysis of English textbooks used nationwide. To this end, three key criteria—range, type frequency, and token frequency—were established and comprehensively applied to determine the grade levels of 2,879 lexical items. By integrating multiple corpus-based measures, this study proposes a systematic approach to assigning vocabulary items to grade levels within a national curriculum. The results indicate that these three measures can serve as useful indicators for English vocabulary profiling. The analysis yielded 98 words for Grade 3, 83 for Grade 4, 196 for Grade 5, 157 for Grade 6, 390 for the first year of middle school, 357 for the second year, 344 for the third year, 514 for the first year of high school, and 740 for the second year and above. Based on these findings, the study discusses pedagogical implications for establishing grade-appropriate vocabulary benchmarks and for informing principled vocabulary selection in English textbook development.

### KEYWORDS

2022 revised national curriculum of English, basic vocabulary list, English Vocabulary Profile, corpus analysis

## 1. 서론

여러 제2언어습득 이론을 통해 오랫동안 강조되어왔듯(Krashen 1982, Long 1983), 제2언어 입력의 수준은 학습자가 이해할 수 있는 정도의 적절한 수준이 되어야 한다. EFL 환경인 우리나라에서는 초·중·고등학교 학생들이 주로 영어 교과서를 중심으로 영어를 학습한다는 점에서 특정 학년의 영어 교과서는 해당 학년의 평균적인 학습자가 이해할 수 있는 입력 수준을 반영한다고 볼 수 있다. 즉, 학년별 영어 교과서의 언어 수준은 해당 학년 학습자에게 적절한 영어 입력 수준의 기준 중 하나가 될 수 있다.

영어 교과서의 언어 수준을 측정할 때 핵심이 되는 요소는 문장 내 어휘와 문법 수준이다. 텍스트를 구성하는 언어적 기본 단위는 문장이며, 이 문장의 난이도를 결정하는 핵심 요소가 어휘와 문법이기 때문이다. 이러한 관점에서 어휘 및 문법 수준은 학습자의 언어 발달 수준이나 텍스트의 언어적 난이도를 평가하는 데 널리 활용되어 왔다(Sung and Kim 2022). 특히 비교적 최근에는 대규모 영어 학습자 코퍼스 또는 영어 교재 코퍼스를 기반으로 어휘 및 문법 수준에 대한 정교한 연구가 활발히 이루어지고 있으며, 이를 통해 보다 객관적인 분석이 가능해지고 있다. 대표적으로, English Profile 프로젝트(Hawkins and Filipović 2012, O’Keeffe and Mark 2017)는 영어 어휘 및 문법을 CEFR 여섯 단계의 수준으로 구분하고 있으며 CEFR-J(Tono 2019)는 10여 개의 수준으로 더욱 정교하게 구분하고 있다. 우리나라 교육과정에서도 별표 자료를 통해 어휘와 문법 항목을 학교급별 3단계 수준으로 구분하고 있으며, English Grammar Profile(O’Keeffe and Mark 2017)과 같은 참조 자료를 활용하여 그 타당성을 확보하고 있다(주형미 외 2022, Kim et al. 2025).

본 연구는 우리나라 초·중·고등학교 영어 교과서의 학년별 어휘 수준을 측정하고 3단계의 어휘 수준을 학년별(초3~고2) 9단계의 어휘 수준으로 세분화하는 것을 목표로 한다. 현행 2022 개정 영어과 교육과정의 별표 자료에는 3,000개의 어휘군이 기본 어휘로 제시되어 있으며, 각 어휘의 권장 수준은 초등학교, 중학교 및 고등학교 공통과목, 고등학교 선택 과목의 세 단계로 분류되어 있다(교육부 2022). 그러나 이러한 3단계 구분은 초등학교 3학년 이후 약 10년에 걸친 영어 학습 과정에서 학년 간 세부적인 언어 수준 차이를 충분히 반영하지 못하는 한계가 있다(최서용 외 2025).

지금까지 국내에서 수행된 교육과정 어휘 분석 연구는 교육과정 어휘 기준 설정(이문복, 신동광 2015, 이현우 2023, 최원경 2025), 교육과정 및 교과서 어휘의 적절성 분석(장세영 외 2024, 황필아 2019), 교육과정 어휘를 활용한 다독·다청 자료 탐색(김기택, 성민창 2019, 성민창, 김기택 2021a, 2021b, 성민창 외 2023) 등 다양한 주제를 탐구하여 왔다. 하지만 초·중·고 전 학년을 대상으로 교육과정 어휘 수준을 세분화한 연구는 아직 이루어지지 않았다. 이는 학년별 어휘 수준을 구분할 수 있는 명확한 방법론이 부재했고, 전 학년의 영어 교과서를 아우르는 코퍼스 데이터가 구축되지 못하였기 때문으로 보인다.

최근 최서용 외(2025)는 2015 개정 영어과 교과서 코퍼스를 대상으로 English Profile(Hawkins and Filipović 2012)의 수준 구분 방법론을 창의적으로 적용하여, 영어교육과정의 학교급별 3단계 문법 수준을 학년별 9단계(초3~고2)로 세분화하는 연구를 수행하였다. 그러나 교육과정 어휘 수준을 학년 단위로 체계적으로 세분화한 어휘 연구는 아직 이루어지지 않았다. 이러한 학년 단위의 세분화가 학습자의 영어 수준에 적합한 입력을 제공하기 위한 참고 자료로 활용될 수 있다는

측면에서, 본 연구는 교육과정의 3단계 어휘 수준을 학년별 9단계(초3~고2)로 세분화하고자 한다. 다만, 문법 항목은 상대적으로 한정된 구조적 단위로 구성되어 ‘최초 등장 시점’이 수준 판정의 주요 지표로 기능할 수 있는 반면, 어휘는 그 수와 종류가 훨씬 방대하고 분포 양상이 이질적이어서 ‘사용 빈도, 범위, 분산’ 등 복수의 분포 지표를 종합적으로 고려할 필요가 있다. 이를 위해 기존의 어휘 수준 구분 방법론을 탐색하고 이를 창의적으로 적용한다. 본 연구는 최서용 외(2025)에서 사용한 문법 수준 세분화와는 차별화된 방법을 사용했는데, 구체적으로 코퍼스 언어학의 주요 어휘 지표에 기반한 범위 기준, 유형 빈도 기준, 출현 빈도 기준 등 3가지 기준을 종합하여 어휘별 최종 학년 수준을 결정하였다. 그러나 2022 개정 교육과정에 기반한 교과서가 아직 전 학년에 걸쳐 충분히 확보되지 않은 상황으로 인해, 본 연구는 2015 개정 교과서를 분석 자료로 활용하였으며, 이에 따라 연구 결과 해석에 일부 제한이 있다.

## 2. 이론적 배경

### 2.1 영어 어휘 수준 구분 연구의 필요성

학습자 언어 수준에 적합한 언어 입력 수준의 중요성은 제2언어습득 이론에서 오랫동안 논의되어 왔다. 대표적으로 Krashen(1982)은 제2언어학습자가 자신의 현재 언어 능력( $i$ )보다 약간 상위 수준( $i+1$ )의 이해가능한 입력(comprehensible input)에 노출될 때 제2언어습득이 촉진된다고 주장하였다. Long(1983)은 상호작용 과정에서 발생하는 의미 협상(negotiation of meaning)을 통해 입력이 학습자 수준에 맞게 조정되는 과정을 설명하며, 학습자 수준에 적합한 언어 입력의 중요성을 강조하였다. Pienemann(1998)은 학습자의 현재 수준을 크게 초과하는 복잡한 언어 항목은 심리언어적으로 처리하기 어렵고, 따라서 학습에도 효과적이지 않다고 본다는 그의 처리가능성 이론(Processability Theory)을 통해, 학습자 언어 수준에 적합한 입력의 중요성을 시사하고 있다.

학습자 언어 수준에 적합한 언어 입력 수준을 결정함에 있어서, 어휘 정보는 핵심적인 역할을 수행한다. 많은 연구들이 영어 학습자가 주어진 영어 입력 자료를 이해하기 위해서는, 자료가 제시하는 어휘의 상당 부분을 알고 있어야 함을 지적하였다(Hirsh and Nation 1992, Laufer 1989, Liu and Nation 1985, Nation 2013, Webb and Rodgers 2009). 즉, 학습자의 어휘 수준이 이해가능한 입력의 범위와 밀접히 연관된다. 예를 들어, 일반적인 영어 소설에 사용된 어휘의 95%를 이해하기 위해서는 약 4,000개 이상의 어휘 지식을 갖출 필요가 있다(Nation 2006).

언어 수준 책정에서 이러한 어휘의 중요성에 입각하여, 학습자 수준에 따라 어휘의 수준을 구분하는 대규모 연구가 수행되어 왔다. 대표적으로 English Profile 프로젝트는 학습자의 실제 영어 사용 데이터를 바탕으로 어휘 수준을 유럽공통어참조기준(CEFR)의 여섯 개의 등급(A1, A2, B1, B2, C1, C2)으로 분류하고 있다. 이러한 분류는 케임브리지 학습자 말뭉치 등의 방대한 자료를 통해 학습자들이 해당 단어를 적절히 사용하기 시작하는 평균적인 수준을 산출한 결과이다. English Profile 프로젝트가 전 세계 영어 학습자를 대상으로 하고 있다면, CEFR-J는 일본의 영어 학습자를 대상으로 언어 수준을 구분하고 있다(Tono 2019). CEFR이 6개의 수준으로 구분된 것과는 달리, CEFR-J는 A1, A2, B1, B2의 네 단계를 Pre-A1, A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2와 같이

10개의 단계로 더 세분화한 것이 특징적이다. 이는 일본의 영어학습 환경을 반영한 것으로, 대부분의 일본 영어 학습자들이 CEFR A~B 단계에 분포하고 있다는 점에서 A~B 단계에 집중하였으며, 또한 학년이 올라감에 따라 작은 영어 수준의 향상도 포착할 수 있도록 단계를 세분화하였다.

우리나라 맥락에서는 언어적 수준 구분이 국가 영어교육과정 주도로 이루어지고 있다. 이는 국가 영어교육과정 별표 항목에서 찾아볼 수 있는데, 어휘 및 문법 수준과 관련하여 대체로 학교급(초, 중, 고)을 기준으로 세 등급으로 제시되어 있다. 이러한 세 등급 수준과 관련한 국내 연구로는 CEFR 수준과의 비교 연구(최정은 외 2022), 학교급별 어휘 수준을 반영한 다독 및 다청 교재 탐색(김기택, 성민창 2019, 성민창, 김기택 2021a, 2021b, 성민창 외 2023, 이운 2016)에 대한 연구 등이 있으며, CEFR-J와 같이 수준을 세분화하는 연구는 English Profile 프로젝트의 문법 수준 분류 방법론을 수정 적용하여 세 등급의 문법 수준을 학년별(초3~고2) 9개 수준으로 구분한 최서용 외(2025)의 연구에 불과하다. 다음 장에서는 국가 영어교육과정 어휘 수준을 분석한 방법론을 제시하고, 교과서 코퍼스를 활용하여 그 수준을 세분화하는 방법론을 논의한다.

## 2.2 2022 개정 영어과 교육과정 기본 어휘 수준 구분

본 연구는 초·중·고등학교의 전체 교과서 텍스트를 확보할 수 있는 2015 개정 기반 영어 교과서를 분석의 대상으로 한다. 이는 2022 개정 영어 교육과정에 기반한 교과서는 현 시점에서 모든 학년의 텍스트가 공개되어 있지 않기 때문이다. 그럼에도 불구하고 2022 개정 영어과 교육과정 기본 어휘를 분석의 기준점으로 사용함으로써 향후 2022 개정 기반 영어 교과서에 대한 어휘 분석의 틀을 마련하고자 하였다. 이를 위해 2022 개정 영어과 교육과정에 제시된 ‘기본 어휘 지침’에 대한 이해가 필수적이다. 이 지침에는 영어과의 성취 기준 달성에 필요한 3,000개의 기본 어휘를 제시하고 있으며, 이러한 기본 어휘는 초등학교에서 사용이 권장되는 어휘군(800개), 중학교와 고등학교 공통과목에서 사용이 권장되는 어휘군(1,200개), 그 외의 과목에서 사용이 권장되는 어휘군(1,000개)으로 구분된다.

2022 개정 영어과 교육과정의 3,000개의 기본 어휘는 2015 개정 영어과 교육과정의 3,000개의 기본 어휘를 바탕으로 일부 어휘를 교체하는 방식으로 진행되었기 때문에(이현우 2023), 먼저 2015 개정 교육과정이 어떻게 각 어휘군의 어휘 수준을 결정했는지 살펴볼 필요가 있다. 2015 개정 교육과정 기본 어휘 목록은 8개의 저명한 해외 코퍼스를 기반으로, 친숙도, 범위, 빈도수 등을 고려하여 3,000개의 기본 어휘 목록이 선정되었다(이문복, 신동광 2015). 1차 선정 단계에서는 원어민의 어휘 사용 양상을 고려하여, ‘빈도수>범위>알파벳 순서’의 기준을 사용하여 3,500개의 어휘를 선정하였고, 2차 선정 단계에서는 초, 중, 고 영어교사 40명의 어휘 친숙도에 대한 의견과 초등 권장 어휘 추천 목록을 바탕으로 초등(800개), 중등 일반 선택(1800개), 진로·전문교과I(400개)와 같은 3단계의 권장 어휘 수준을 선정하였다. 마지막 단계에서는 영어 교사들의 최종검토 과정을 거쳐 일부 어휘를 교체하거나 삭제하여 최종 어휘를 확정하였다. 2022 개정 영어과 교육과정에서는 이러한 2015 교육과정의 어휘군을 기초로 수정 작업이 이루어졌다(이현우 2023). 먼저, 초등 수준의 어휘군은 2015 개정 영어과 교육과정에서 초등 수준으로 권장한 어휘군 800개 중 BNC/COCA Level 6 word family lists(Nation 2017)과 New General

Service List(Browne et al. 2013) 내에서 1,000번째 순위 내의 어휘(589개) 또는 초등학생에게 친숙한 소재 관련 어휘(82개)를 포함하였다. 다음으로, 2015 개정 교육과정에서 초등 수준으로 권장된 어휘군 중 중학생에게 더 적합한 어휘(102개)와 중학생에게 더 적합한 근원어(source word) 57개를 선정하였다. 또한 2015 개정 교육과정에서 그 외 기본 어휘 중에서 BNC/COCA Level 6 word family lists와 BNC 14,000 word lists(Nation 2006)에서 2,000번째 순위 내에 있는 어휘 1,009개와 중학교 영어와 공통 영어에 더 적합해 보이는 어휘 19개를 포함하였으며, 새롭게 13개의 어휘를 추가하였다. 마지막으로, 가장 높은 수준의 어휘군은 기존 2015 개정 교육과정에서는 초등 권장 수준이었던 어휘 1개(academy), 고등학생에게 적합한 근원어 10개 그리고 새롭게 추가된 어휘 4개를 포함하였다. 정리하자면, 2015 개정 교육과정과 2022 개정 교육과정에서 제시된 기본 어휘 목록은 공통적으로 3개 수준의 권장 어휘를 제시하며, 학교군 별로 구체적으로 사용 가능한 어휘 수를 규정하고 있다.

그러나 이러한 3단계의 어휘 수준 구별은 어휘를 CEFR 6단계 수준(A1~C2)으로 구분한 English Vocabulary Profile보다도, 혹은 CEFR A와 B 수준만을 10개로 세분화한 CEFR-J에 비교해서도 그 범위가 너무 넓어 보인다. 특히, 세 단계의 수준은 학생들의 학년별 언어적 어휘 발달을 포착하는데 충분하지 않은데, 예를 들어 현재의 영어교육과정 어휘 수준에 따르면 초등학교 3학년과 6학년은 같은 초등학교급에 포함되어, 두 학년 간 학생의 어휘 발달을 포착하는 것이 쉽지 않다. 현행 교육과정의 포괄적인 어휘 구분은 교육 및 연구 전반에 걸쳐 실질적인 제약을 초래한다. 우선, 학년별 표준 로드맵의 부재는 교과서 간 어휘 배치와 난이도의 불일치를 심화시켜, 학습 기회의 형평성을 저해하고 평가 타당도와 학교급 간 수직적 연계성을 약화시킬 수 있다. 또한, 정교한 분류 체계의 부족은 학습자의 어휘 발달 수준이나 교재의 적절성을 객관적으로 검증하는데 방법론적 한계로 작용한다. 이러한 제약을 극복하기 위해서는 현행 체제를 넘어 교육과정 어휘 수준을 보다 정교하게 세분화할 필요가 있다.

교육과정에서 제시된 기본 어휘 목록의 세부 학년별 어휘 수준을 결정하는 방법론과 관련하여 앞서 언급한 English Vocabulary Profile과 CEFR-J를 참고할 수 있다. 두 방법론은 어휘 수준 결정을 위해 활용한 자료에서 차이를 보이는데, 먼저, English Vocabulary Profile은 케임브리지 학습자 말뭉치를 활용하여, 특정 CEFR 수준의 학습자가 해당 어휘를 사용할 수 있는지를 그 주요 기준으로 삼고 있는 반면, CEFR-J는 영어 교과서 등 학습자가 사용하는 교재의 어휘 수준을 주요 기준으로 활용하였다(Tono 2019). 본 연구는 수준별 한국인 영어 학습자 코퍼스가 충분하지 않다는 측면에서, CEFR-J의 방법론을 따라 영어 교과서 코퍼스를 바탕으로 학년별 적정 어휘를 결정하고자 한다. 특히 EFL 환경인 우리나라에서는 초·중·고등학교 학생들이 주로 영어 교과서를 중심으로 영어를 학습한다는 점에서 특정 학년의 교과서에 포함된 영어 수준은 해당 학년 평균 학습자가 이해할 수 있는 입력 수준을 반영한다고 볼 수 있으며, 따라서 학년별 영어 교과서의 언어 수준은 해당 학년 학습자에게 적절한 영어 입력의 기준이 될 수 있다. 다음 장에서는 이러한 교과서 코퍼스를 통해 어휘 수준을 결정하는 구체적인 방법론을 다룬다.

### 2.3 교과서 코퍼스 어휘 수준 분석 방법론

어휘 선정과 관련하여 대표적인 기준은 빈도(frequency)와 범위(range)이다(Biber et al. 1999). 빈도는 크게 구현 빈도(token frequency)와 유형 빈도(type frequency)로 구분이 되는데, 먼저 구현

빈도는 얼마나 자주 등장하는지에 대한 개념으로, 대체로 최상위 빈도 단어(예: *the, is, have*)는 대부분의 영어 텍스트에 등장하며 학습 초기부터 배우게 된다. 예를 들어, 출현 빈도가 높은 상위 약 2,000개 유형 어휘가 대체로 텍스트에 제시된 모든 출현 어휘의 80%를 차지한다는 연구(Nation 2006)들은 이러한 고빈도 어휘를 먼저 학습하는 것이 학습 효율 측면에서 유리하다는 점을 뒷받침한다. 특히 읽기 이해를 위해 필요한 어휘 규모에 대한 연구들에 따르면, 독자들이 텍스트를 도움 없이도 스스로 이해할 수 있으려면 출현 어휘의 약 90%~98%를 알고 있어야 한다(Hirsh and Nation 1992, Laufer 1989, Liu and Nation 1985, Nation 2013, Webb and Rodgers 2009). 이러한 결과는 고빈도 어휘 습득의 중요성을 시사하며, 이들 어휘가 학습의 초기 단계부터 체계적으로 제시되어야 할 필요성을 뒷받침한다. 대표적으로 *General Service List(GSL)*나 *New GSL* 등은 빈도를 기반으로 기본 학습 어휘를 채택하고 있으며, 출현 빈도가 높은 어휘를 먼저 배우고 낮은 어휘는 나중에 배우는 단계적 학습이 이루어지고 있다.

반면 유형 빈도는 어휘 가짓수에 해당하는 데, 하나의 대표형이 여러 형태의 변이형으로 활용되는 현상과 관련이 있다. 예를 들어, 대표형 *ACT*가 *act, acting, action, actor*의 변이형으로 활용되었다면, 해당 대표형의 유형 빈도는 4로 측정된다. 이와 같은 유형 빈도가 학습자 수준별 어휘 분석에 시사하는 바는 어휘의 생산성과 분산으로, 유형 빈도가 높은 어휘, 즉 다양한 형태로 사용된 어휘는 해당 텍스트에서 생산성이 높고 여러 문맥에 고르게 분포하는 주요 어휘로 분석할 수 있다(Gries 2015).

범위는 특정 어휘가 코퍼스 내 얼마나 다양한 자료에 고르게 분포되어 있는지, 혹은 일부 자료에만 국한되어 쓰이는지를 측정한다. 어휘 선정과 관련하여, 범위는 빈도를 보완하는 지표로서 널리 활용되어 오고 있는데, 일례로, Coxhead(2000)는 학술 어휘 목록(*Academic Word List*) 구축 과정에서 학문 분야와 전공 주제로 분석 어휘가 출현하는 코퍼스의 범위를 반영하였다. 구체적으로, 학술 말뭉치의 4개 주요 분야 각각에서 최소 10회 이상 출현하는 어휘, 전체 28개 전공 주제 중 15개 이상에서 나타나는 어휘만을 목록에 포함하였다.

이러한 빈도 및 범위는 어휘 수준을 결정하는 데 있어 중요한 지표로 활용되었다. 대표적으로, CEFR-J의 수준별 어휘 선정은 한국, 중국, 대만 등에서 활용되는 영어 교과서를 대상으로 각 교과서의 CEFR 수준을 분석한 후, 각 교과서에서 등장하는 어휘의 출현 빈도와 각 국가에서 공통적으로 나타나는 어휘를 중심으로 CEFR-J의 어휘를 10개의 수준으로 구분하였다(Tono 2019). 우리나라 맥락에서도 이러한 빈도 및 범위는 어휘 선정에 있어 중요하게 고려되고 있다. 일례로, 장경숙 외(2011, p. 215)의 연구에서는 영어교육과정 기본 어휘 목록 개정 작업에서 “K1~20 Words 빈도수 레벨 중 K20 Words보다는 K1 Words쪽의 어휘를 우선적으로 선택하여 기본 어휘 목록에 포함시키는 것을 고려해야 한다”며 고빈도 어휘의 우선 사용을 강조하고 있으며, “기존 검인정 교과서에 공통적으로 나타나는 어휘를 우선 선정”을 제안하며 어휘 선정에 있어 범위의 중요성을 명시하고 있다. 본 연구 역시 학년별 교과서에 제시된 어휘의 빈도 및 범위를 바탕으로 그 수준을 세분화하고자 한다. 보다 구체적으로, 선행연구를 바탕으로 범위, 유형 빈도, 출현 빈도 기준을 마련하여 이를 적용한 결과를 제시한다. 또한, 해당 결과의 타당성을 탐구하기 위해, (1) 범위, 유형 빈도, 출현 빈도 기준을 각각 적용하여 도출한 학년별 어휘 목록이 서로 어느 정도로 일치하는지 살펴보고, (2) 위 3가지 기준을 종합적으로 고려한 어휘별 학년 수준의 결과를 도출한다.

### 3. 연구 방법

#### 3.1 전 학년 영어 교과서 코퍼스 데이터 구축

본 연구는 우리나라 영어과 교육과정의 3,000개 기본 어휘들이 실제 영어 교육 현장에서 사용되는 영어 교과서에서 어느 정도 빈도를 갖고 등장하는지 분석하였다. 이를 위해, 가장 먼저 초·중·고등학교 전 학년 교과서 텍스트를 코퍼스로 구축하였다(조규희, 이동환 2025, Kim and Oh 2020). 여기서 초등학교는 읽기와 듣기 텍스트를, 중·고등학교는 읽기 텍스트로 코퍼스를 구축하였다. 이는 초등학교에서는 읽기와 듣기 텍스트에서 어휘 항목이 균등하게 사용되지만, 중·고등학교에서는 읽기 텍스트 중심으로 다양한 어휘가 사용되고 학습된다는 사실을 고려한 것이다.

앞서 언급한 바와 같이, 본 연구는 초·중·고등학교의 전체 교과서 텍스트를 확보할 수 있는 2015 개정 기반 영어 교과서를 분석의 대상으로 삼았다. 구체적으로, 초등학교 영어(각 학년별 5종), 중학교 영어(13종), 고등학교 영어(11종), 고등학교 영어 I(10종), 고등학교 영어 II(9종), 고등학교 영어 독해와 작문(5종)을 분석 대상으로 하였다. 나아가 2015 개정 교육과정과 2022 개정 교육과정은 기본 어휘 목록의 규모와 내용에서 공통점이 많기 때문에, 2022 개정 교육과정 기본 어휘의 학년별 수준 분석에 2015 개정 교육과정의 교과서를 활용할 수 있을 것이라고 판단하였다. 2015 개정 초·중·고등학교 영어 교과서를 #LancsBox(ver 5.5.0.)를 활용하여 분석한 결과, 총 9개 학년도에 걸쳐 본 연구에서 분석된 총 어휘의 규모는 40,718개였으며, 이들의 분포는 표 1과 같다.<sup>1</sup>

표 1. 영어 교과서 코퍼스에서 분석된 어휘 규모

학년	분석된 어휘 수
초등학교 3학년	396
초등학교 4학년	593
초등학교 5학년	1,376
초등학교 6학년	1,542
중학교 1학년	2,975
중학교 2학년	3,849
중학교 3학년	4,626
고등학교 1학년	6,721
고등학교 2학년 이상	18,640
분석된 총 어휘 수	40,718

<sup>1</sup> 현재 표 1에서 확인할 수 있는 분석된 단어의 수는 각 세부 학년별로 동일한 어휘의 중복이 허용된 숫자이다. 또한 분석된 총 단어의 수는 기본형(lemma)과 그 변이형을 별도의 어휘로 간주하여 산출한 결과이다.

### 3.2 학년별 기본 어휘 활용 양상 분석

2022 개정 교육과정에 근거한 3,000개의 기본 어휘 목록을 알파벳순으로 정리하고 초등학교 권장 어휘(800개), 중학교와 고등학교 공통과목 권장 어휘(1,200개), 나머지 어휘(1,000개) 등과 같이 권장 수준을 표시한 어휘 분석표를 작성하였다. 이후 랭카스터 대학교(Lancaster University)에서 개발된 언어 데이터 및 코퍼스 분석 소프트웨어인 #LancsBox(ver 5.5.0.)을 활용하여 기구축한 학년별 영어 교과서 코퍼스에서 등장한 3,000개의 기본 어휘별 정보를 산출하였다(Brezina et al. 2020). 구체적으로, 모든 어휘는 소문자로 변환하고 절대 빈도수, 상대 빈도수<sup>2</sup>, 코퍼스 간 범위 등에 대한 어휘 정보를 산출하였다. 이후, 교육과정 평가원의 단어 수 검증 웹에서 제공하는 기본 어휘별 변이형 목록을 고려하여, 변이형 단어들도 기본 어휘 목록 3,000개에 맞춰 그 빈도수를 합산하였다. 예를 들어, ‘happy’라는 단어는 ‘unhappy, happier, happily’ 등 11개의 변이형 목록이 제공되고 있는데, 이 모든 단어들의 빈도수는 ‘happy’라는 단어에 통합되어 계산되도록 하였다. 한편, 이 코퍼스 프로그램은 특정 어휘의 절대 빈도수뿐 아니라 상대 빈도수 정보도 함께 제시하는데, 본 연구는 세부 학년별 영어 교과서의 지문 길이와 사용된 어휘 수의 차이가 있음을 고려하여 각 어휘의 상대 빈도수를 도출하여 각 어휘별 상대 빈도 분석표를 완성하였다. 그림 1은 이와 같은 과정을 통해 완성된 학년별 영어 교과서에 나타난 각 어휘별 상대 빈도 분석 결과를 보여준다. 첫 번째와 두 번째 열은 1번부터 3,000번까지의 각 어휘에 부여된 번호와 알파벳순으로 나열된 어휘를 보여준다. 그리고 첫 번째 행은 본 연구에서 분석한 각 세부 학년별 교과서 코퍼스를 나타낸다. 마지막으로, 가운데 제시된 각 숫자는 세부 학년별 교과서에 등장한 상대 빈도수를 의미한다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	번호	기본 어휘	구분	초3	초4	초5	초6	중1	중2	중3	고_1	고_영어1	고_영어2	고_영독작
2	1	a	초등	25708.53	12746.39	21815.48	30589.98	27935.04	26758.1	29007.66	27488.88	28174.24	26733.9	29760.24
3	2	abandon	고급	0	0	0	0	0	0	0	29.46	37.17	53.07	93.29
4	3	able	중고	0	0	0	0	0	295.82	847.08	1031.19	873.46	1026.17	1368.27
5	4	aboard	고급	0	0	0	0	0	0	20.66	29.46	0	17.69	0
6	5	abort	고급	0	0	0	0	0	0	0	0	0	0	0
7	6	abound	고급	0	0	0	0	0	0	0	14.73	74.34	88.46	0
8	7	about	초등	774.35	1182.65	3629.41	5511.37	3575.69	3576.71	3842.9	3447.16	3270.89	2972.4	2829.87
9	8	above	초등	0	0	0	0	37.25	161.36	82.64	206.24	204.43	194.62	248.78
10	9	abroad	고급	0	0	0	0	0	80.68	0	29.46	0	0	93.29
11	10	absent	고급	0	0	0	0	0	0	0	0	0	17.69	0
12	11	absolute	중고	0	0	0	0	0	26.89	0	29.46	74.33	53.08	93.29
13	12	absorb	고급	0	0	0	0	0	161.35	41.32	162.04	55.74	70.77	62.19
14	13	abstract	고급	0	0	0	0	0	0	0	14.73	130.09	0	31.1
15	14	absurd	고급	0	0	0	0	0	0	0	0	55.74	0	0
16	15	abuse	고급	0	0	0	0	0	0	0	0	55.75	17.69	0
17	16	academy	고급	0	0	0	0	0	26.89	0	58.92	92.92	88.46	31.1
18	17	accelerate	고급	0	0	0	0	0	0	0	0	18.58	35.38	62.19
19	18	accent	중고	0	0	0	0	0	0	0	0	0	0	0
20	19	accept	중고	0	0	0	0	0	0	41.32	117.85	148.67	336.16	62.2
21	20	access	중고	0	0	0	0	0	0	0	29.46	167.26	141.54	93.29

그림 1. 2022 개정 영어과 교육과정 기본 어휘의 상대 빈도 분석

### 3.3 기본 어휘의 적정 학년 수준 결정 방법

본 연구는 2022 개정 영어과 교육과정의 기본 어휘 3,000개의 적정 학년 수준을 배정하기

<sup>2</sup> 본 연구에서는 상대 빈도수를 구하기 위해 절대 빈도수를 100만 단어 기준으로 표준화하는 방식을 사용하였다.

위하여, 코퍼스 언어학의 주요 어휘 지표인 범위, 유형 빈도, 출현 빈도를 반영하였다. 세 기준은 어휘 수준의 서로 다른 측면을 포착한다. 범위 기준은 특정 어휘가 동일 학년의 복수 교과서에 얼마나 널리 분포하는지를, 유형 빈도 기준은 해당 어휘가 변이형을 포함하여 얼마나 생산적으로 사용되는지를, 출현 빈도 기준은 특정 학년 이후 사용량이 얼마나 뚜렷하게 증가하는지를 반영한다. 따라서 본 연구는 단순한 최초 등장 시점만으로 수준을 판정하는 것이 아니라, 분포의 확산, 생산성, 사용량 증가라는 세 측면을 종합하여 각 어휘의 적정 학년 수준을 판단하였다.

가. [범위 기준] 범위 기준을 도입하여 수준을 구분한 Tono(2019)의 방법론을 참고하여, 본 연구에서도 범위 기준을 활용하였다. 범위 기준은 어휘 연구에서 중요하게 활용되는 지표로서, 2015 영어과 교육과정의 기본 어휘 목록 개발(이문복, 신동광 2015)과 최근 텍스트 마이닝 연구에서 자주 사용되는 TF-IDF 분석 기법(오영교 2025)이 대표적인 예이다. Tono(2019)에서는 같은 CEFR 수준의 한국, 중국, 대만 교과서에 모두 등장한 단어, 즉 국가별 교재를 범위의 기준으로 삼았으나, 본 연구는 우리나라 교과서만을 분석대상으로 하기에 ‘출판사 별 교과서’를 범위의 기준으로 삼았다. Tono(2019)가 같은 CEFR 수준의 모든 국가별 교재에서 등장하는 어휘를 선정한 반면, 본 연구에서 분석한 우리나라 교과서의 경우 초등은 5종, 중등은 10종 이상의 다양한 출판사로 구성되어 있다는 점에서, 같은 학년의 모든 출판사 대신 ‘복수’의 출판사에서 등장하는 어휘를 범위 기준으로 정하였다. 이는 범위 기준을 출현 교과서 3종 이상으로 할 경우, 이를 만족시키는 특정 어휘의 숫자가 급격하게 줄어들어 세부 학년 수준을 결정하는 데 어려움이 있었기 때문이다.<sup>3</sup> 따라서, 본 연구는 특정 어휘의 출현 교과서 누적 수치가 2종 이상이 되었을 경우, 해당 학년으로 그 수준을 결정하였다. 즉, 특정 어휘가 최초 등장한 학년이 아닌, 2종 이상의 출판사에서 어휘가 확산적으로 사용된 시점을 어휘 수준 판단의 근거로 삼았다. 예를 들어, 특정 어휘가 초등학교 3학년 A 출판사에서 사용되고, 4학년 A 출판사에서 사용되고, 5학년 B 출판사에서 사용되었다면, 5학년에 누적 수치가 2종이 되고, 따라서 해당 어휘는 범위 기준으로는 초등학교 5학년 어휘로 결정된다.

나. [유형 빈도 기준] 유형 빈도 기준을 도입하여 어휘 수준을 구분한 Gries(2015)의 방법론을 참고하여, 본 연구에서도 유형 빈도 기준을 활용하였다. 유형 빈도 기준은 코퍼스 언어학을 비롯한 다양한 연구 분야(예: 언어 습득)에서 중요한 분석 기준으로 사용되기 때문에 본 연구에서도 주요한 학년 선정 기준 중 하나로 사용하였다. 본 연구는 출현 유형 빈도가 2 이상으로 누적되는 경우, 해당 학년으로 그 수준을 결정하였다. 구체적으로, 유형 빈도 기준은 교육과정 평가원의 변이형 목록을 기준으로 특정 학년 코퍼스에 등장한 고유한 변이형의 종류가 2개 이상이 되는 시점을 측정하는 것이다. 다시 말해, 특정 어휘가 1가지

<sup>3</sup> 2015 개정 중학교 교과서 코퍼스를 분석한 결과, 동일 학년에서 교과서에 동시에 등장하는 어휘의 수는 상당히 제한적인 것으로 나타났다. 이는 출판사별로 학년별 어휘 수준에 대한 명시적 기준보다는 학교급 단위의 어휘 기준을 바탕으로 교과서가 개발된 데에서 비롯된 결과로 해석할 수 있다. 만약 5종 이상과 같은 보다 엄격한 기준을 적용할 경우, 해당 기준을 충족하는 어휘 수가 크게 감소하여 범위 기준을 적용할 수 있는 어휘의 분류 가능성이 현저히 제한되는 문제가 발생하였다. 아울러 초등과 중등에서 서로 다른 범위 기준을 적용하는 방안도 검토하였으나, 인접 학년 간 기준 단절로 인한 왜곡 가능성을 고려하여 “2종 이상”이라는 동일 기준을 유지하였다.

유형만으로 최초 등장한 학년이 아니라, 서로 다른 유형으로 누적하여 2번 이상 등장하기 시작했던 학년을 해당 어휘의 학년 수준으로 판단하였다. 이는 범위 기준과 마찬가지로, 출현 유형 빈도가 3 이상으로 누적되는 어휘가 충분치 않아 적절한 유형 빈도 기준을 2 이상으로 결정한 것이다. 예를 들어, 특정 동사가 중학교 1학년에서는 대표형으로만 사용되고, 중학교 2학년에서는 과거형으로 사용되었다면, 2학년에서 유형 빈도가 2로 누적되고, 따라서 해당 어휘는 유형 빈도 기준으로는 중학교 2학년 어휘로 결정된다. 참고로, 교육과정의 변이형 목록은 대표형(예: develop)이 파생 변이형(예: development)을 포함한다. 따라서 본 연구의 유형 빈도 분석은 코퍼스 연구에서 자주 활용되는 어휘군(word family) 분석에 해당한다(Milliner and Pinchbeck 2025).

다. [출현 빈도 기준] 특정 어휘의 상대 빈도수가 직전 학년 수준에 비해 급격하게 늘어날 경우, 해당 학년을 어휘 수준으로 판정한다. 증가 추세의 판단 기준으로, 최서용 외(2025)의 연구에서 문법 항목의 학년 수준 구별을 위해 제안된 3-1 규칙을 어휘 항목의 수준 구분에도 동일하게 적용한다. 3-1 규칙은 English Profile 연구에서 문법항목의 CEFR 수준을 결정하기 위해 도입한 10-1 규칙을 수정 적용한 것이다(Hawkins and Filipović 2012).<sup>4</sup> 본 연구에서도 교과서 코퍼스에서 특정 어휘의 학년별 상대 빈도수를 고려하기 때문에 동일한 방법론을 사용하는 것이 유효하다고 판단하였다. 3-1 규칙은 특정한 어휘의 상대 빈도수가 직전 학년에 비해 최소 3배 이상 증가한 경우를 고려한다. 만약 3배 이상 증가하는 경우가 없을 경우, 특정 어휘가 최초 등장한 학년을 해당 어휘 수준으로 판정한다. 예를 들어, 특정 어휘가 중학교 3학년까지 거의 등장하지 않다가, 고등학교 1학년에 직전 학년인 중학교 3학년에 비해 상대 빈도수가 3배 이상 급격히 증가한 경우, 해당 어휘는 출현 빈도 기준으로 고등학교 1학년 어휘로 결정된다. 만약 또 다른 특정 어휘가 초등학교 5학년에 최초 등장했으나, 고등학교 수준까지 상대 빈도수가 3배 이상 급격한 증가를 보이지 않을 경우에는 최초 등장한 초등학교 5학년 수준으로 결정된다. 이와 같이 출현 빈도로 개별 어휘와 특정 코퍼스(예: 5학년 교과서) 간의 관련성을 파악하는 방식은 코퍼스 언어학의 기본 접근법에 해당하며, 이에 입각한 대표적인 분석 기법으로 키워드 분석(keyword analysis)이 있다(Rayson and Potts 2021).

위에서 제시한 3가지 기준(범위, 유형 빈도, 출현 빈도)이 어휘의 학년 수준을 결정하는 기준으로서 얼마나 타당한지 살펴보기 위하여 3가지 기준 사이의 관련성과 일치도를 분석하였다. 이를 위하여, 각 기준을 적용하여 도출한 어휘별 학년 수준 간 Pearson 상관분석을 실시하고, Intraclass Correlation Coefficient(ICC)를 산출하였다. 특히, ICC는 다양한 평가자가 동일한 대상을 판단할 때의 신뢰도를 계량적으로 검토하는 데 적합한 분석 도구로, 측정 지표 간의 판단

<sup>4</sup> 최서용 외(2025)는 우리나라 교과서 코퍼스의 분량으로는 Hawkins와 Filipović(2012)의 10-1규칙 적용이 어렵다는 점을 확인하였다. 따라서, 2-1 규칙, 3-1 규칙, 4-1 규칙, 5-1 규칙 등 다양한 기준을 적용하여 문법 항목의 수준을 분석하였고, 최종적으로 3-1 규칙을 적용하였다. 본 연구 역시, 2-1 규칙, 4-1 규칙, 5-1 규칙 등 다양한 규칙을 어휘 수준 결정을 위한 분석에 적용하였고, 그 결과 3-1 규칙이 어휘 수준을 위한 분석에서도 가장 안정적이었기에 3-1 규칙을 적용하였다.

일치도(consistency)와 정확도(agreement)를 동시에 고려할 수 있다는 점에서 3가지 기준의 타당성을 판단하는 데 적합하다(Shrout and Fleiss 1979).

이후, 3가지 기준을 종합하여 어휘별 최종 학년 수준을 결정하였다. 3가지 기준을 통합적으로 고려하는 이유는 1가지 기준만으로는 어휘의 학년 수준을 결정하지 못하는 경우가 존재하기 때문이다. 예를 들어, 범위 기준을 적용했을 때 1종의 교과서에만 등장하는 특정 어휘는 최종 학년을 결정하지 못한다. 또한, 유형 빈도 기준에서 1가지 유형(예를 들어, 대표형으로만 사용되는 어휘)으로만 사용된 어휘의 경우에는 이 기준만으로 최종 학년을 결정할 수 없다. 따라서, 1가지 기준만으로는 놓칠 수 있는 어휘를 다른 기준이 보완할 수 있도록 3가지 기준을 종합적으로 사용하였다. 구체적으로, 3가지 기준을 바탕으로 이상치(outlier)를 회피할 수 있는 ‘중간값(median)’을 활용하였다. 예를 들어, 특정 어휘가 범위 기준에서는 초6, 유형 빈도 기준에서는 중1, 출현 빈도 기준에서는 고2 이상으로 결정되었다면, 출현 빈도 기준(고2 이상)은 회피할 필요가 있었다. 먼저, 최종 학년 수준을 나타내는 중간값을 구하기 위하여 각 학년을 숫자로 치환하였다(초3=1, 초4=2, ..., 고1=8, 고2 이상=9). 다음으로, 3가지 기준에 의해 산출된 학년 수준의 중간값을 구하였다. 예를 들어, 특정한 어휘가 범위 기준에서 고2 이상(9), 유형 빈도 기준에서 고1(8), 출현 빈도 기준에서 중2(6)으로 판단되었다면, 최종 학년 수준은 중간값이 8인 ‘고1’ 수준으로 결정하였다. 그러나, 3가지 기준을 모두 갖고 있지 않은 어휘의 경우, 중간값을 구하기 어렵기 때문에, 교육과정의 권장 수준 등을 고려하여 질적으로 검토한 후 최종 학년 수준을 결정하였다.<sup>5</sup>

#### 4. 연구 결과 및 논의

2022 개정 영어과 교육과정 기본 어휘 3,000개 중 2,879개 어휘가 2015 개정 교과서 코퍼스에서 사용된 것으로 분석되어, 본 연구는 이들 어휘에 대한 세부 학년 수준을 보고한다. 참고로 해당 코퍼스에서 사용되지 않은 121개의 어휘는 표 2와 같다. 121개의 미사용 어휘는 본 연구의 분석 코퍼스에 포함되지 않은 진로 선택 및 융합 선택 과목, 그리고 중, 고등학교의 듣기 코퍼스 등에서 출현할 것으로 예상된다. 또한 2015 개정 교육과정과 2022 개정 교육과정에서 각각 제시하고 있는 기본 어휘 목록의 차이에서도 비롯된 것으로 보인다.<sup>6</sup>

<sup>5</sup> 3가지 기준을 모두 갖고 있지 않은 어휘의 경우, 다음과 같은 방식으로 질적 분석을 수행하였다. 첫째, 2022 개정 교육과정에서 권장하고 있는 학교급 수준에 더 근접한 학년 수준을 반영하였다. 예를 들어, 2가지 기준에서 초등학교 3학년과 초등학교 4학년 수준으로 판별되고, 교육과정의 권장 수준이 중학교인 특정 어휘가 있을 때, 중학교 수준과 더 근접한 초등학교 4학년을 최종 학년 수준으로 결정하였다. 둘째, 선행연구에서 범위 기준과 유형 빈도 기준이 출현 빈도 기준(3-1 규칙)에 비해 폭넓게 활용되었다는 점을 바탕으로, 범위 기준 및 유형 빈도 기준의 결과값을 출현 빈도 기준보다 우선적으로 적용하였다.

<sup>6</sup> 2022 개정 영어과 교육과정의 기본 어휘 목록은 2015 개정 목록을 기반으로 일부 어휘가 교체·보완된 것으로, 이현우(2023)에 따르면 기존 126개 어휘가 제외되고 일부 어휘가 신규로 추가된 것으로 보고된다. 예를 들어, ‘twenty’ 관련 어휘는 2015 개정 영어과 기본 어휘 목록에는 포함되어 있지 않았으나, 2022 개정 목록에서는 새롭게 추가된 항목에 해당한다. 다만 표 2에 제시된 어휘는 이러한 교육과정 간 차이뿐 아니라, 본 연구에서 포함하지 못한 일부 코퍼스(진로 및 융합 선택 과목, 듣기 자료 등)의 영향으로 나타났을 가능성도 존재한다.

표 2. 영어 교과서 코퍼스에서 미사용된 어휘 목록

기본 어휘 수준	미사용된 어휘 수	어휘 예시
초등 권장 수준 어휘 (800 단어)	5	<i>gum, recreation, twenty-one, twenty-two, twenty-three</i>
중·고등 공통과목 권장 수준 어휘 (1,200 단어)	14	<i>accent, catalogue, cinema, cottage, disc, disgust, essay, hint, licence, livingroom, loan, pardon, pub, villa</i>
고등 선택과목 권장 수준 어휘 (1,000 단어)	102	<i>abort, aesthetic, agenda, breed, censor, certificate, chamber, chorus, clause, coincide, communist, compatible, compound</i>

#### 4.1 범위, 유형 빈도, 출현 빈도 기준의 일치도

표 3은 범위, 유형 빈도, 출현 빈도 기준에 의해 산출된 세부 학년이 모두 일치한 어휘 항목의 수를 보여준다. 먼저, 본 연구에서 분석한 2,879개의 어휘 중 3가지 기준이 모두 산출된 어휘는 2,252개로 확인되었다. 이중에서 세부 학년 수준이 3가지 기준 모두에서 완전히 일치하는 어휘는 총 797개로 약 35.4%의 어휘가 이에 해당되었다. 또한 3가지 기준에서 동일한 학교급으로 판단된 어휘, 즉 모든 기준에서 학교급(초등, 중등, 고등)이 일치한 어휘는 총 1,361개로 약 60.4%의 어휘가 이에 해당되었다. 이는 서로 다르게 정의된 각 기준으로 세부 학년을 결정하였음에도 불구하고, 3가지 기준이 각각 결정한 세부 어휘 항목 수준이 상당히 일치함을 시사한다.

표 3. 세 가지 기준에서 동일한 학년 수준으로 분류된 어휘 항목 수

학년	어휘 수
초등학교 3학년	15
초등학교 4학년	9
초등학교 5학년	37
초등학교 6학년	23
중학교 1학년	131
중학교 2학년	95
중학교 3학년	103
고등학교 1학년	159
고등학교 2학년 이상	225
총 합계	797

이와 더불어, 본 연구는 범위, 유형 빈도, 출현 빈도 기준을 각각 적용하여 도출한 학년별 어휘 목록은 서로 어느 정도로 일치하는지를 확인하기 위해 3가지 기준 간의 상관성과 일관성을 다각도로 분석하였다. 먼저, 3가지 기준 간 상관관계를 파악하기 위해 Pearson 상관분석을 실시한 결과, 범위 기준과 유형 빈도 기준 간에는 강한 상관관계가 나타났으며( $r = .757, p < .001$ ), 범위 기준과 출현 빈도 기준 간( $r = .635, p < .001$ )과 유형 빈도 기준과 출현 빈도 기준 간( $r = .559, p$

< .001)에는 중간 수준의 상관관계가 확인되었다. 출현 빈도 기준은 3-1 규칙을 만족하지 않을 시 최초에 그 어휘가 등장한 학년을 해당 수준으로 간주했기 때문에, 모든 어휘에 학년 수준을 부여할 수 있다는 장점을 가지지만, 앞선 2가지 기준(범위 기준과 유형 빈도 기준)과의 상관도에는 다소 부정적인 영향을 미친 것으로 판단된다.

아울러, 3가지 기준(범위, 유형 빈도, 출현 빈도)을 통해 도출된 어휘별 세부 학년 수준 간의 일치성과 측정 신뢰도를 검토하기 위해 Intra-class Correlation Coefficient(ICC)를 산출하였다(Shrout and Fleiss 1979). 본 연구에서는 각 어휘 항목을 동일한 평가 대상으로 간주하고, 3가지 기준을 고정된 평가자(fixed raters)로 설정하여 측정값의 일관성을 평가할 수 있는 ICC(3, k) 모델을 적용하였다. 그 결과, 3가지 기준(범위, 유형 빈도, 출현 빈도)이 동일 어휘에 대해 부여한 세부 학년 결과 사이의 일치도를 평가한 결과, 세 지표 간 측정값은 매우 높은 수준의 일관성과 신뢰도를 보여주었다(ICC(3, k) = .859,  $F = 7.12$ ,  $p < .001$ ). 이는 세 기준이 어휘의 학년 수준을 서로 다른 측면에서 포착하면서도 전반적으로 높은 내적 정합성을 보인다는 점을 시사한다. 따라서 본 연구는 세 기준을 상호 보완적으로 활용하여 어휘의 세부 학년 수준을 결정하였다. 다만 이러한 결과는 기준들 간의 내적 일관성을 보여주는 것이며, 교육 현장에서의 내용 타당도는 후속 연구에서 교사 및 교재 개발자 검토를 통해 추가로 검증될 필요가 있다.

#### 4.2 범위, 유형 빈도, 출현 빈도 기준을 반영한 학년별 어휘 목록 도출

위에서 보고한 바와 같이, 본 연구에서 분석한 2,879개의 어휘 중 797개의 어휘는 각 기준이 산출한 학년 수준이 완전히 일치하였으며, 1,455개의 어휘는 이상치(outlier)를 제거하는 중간값(median)을 적용하여 학년별 어휘 수준을 결정하였다. 나머지 627개의 어휘는 1개 또는 2개의 기준에 의한 학년 수준 결과값만 존재하여 본 연구자들의 질적인 검토(각주 5 참조)와 합의를 통해 최종적인 학년별 어휘 수준을 판단하였다. 본 연구에서 도출한 학년 수준을 기존 교육과정의 학교급 분류와 비교한 결과, 분석 대상 2,879개 어휘 중 2,241개(약 78%)가 동일한 학교급으로 분류되어 전반적으로 높은 수준의 일치도를 보였다. 표 4는 이와 같은 과정을 통해 도출된 학년별 어휘 목록 중 일부를 알파벳 순으로 제시하고 있으며, 모든 어휘 목록 및 이를 범주화한 파일은 별도의 사이트로 제공한다.<sup>7</sup>

표 4에서 확인할 수 있듯이, 분석된 2,879개 어휘 중 초등학교 어휘는 534개, 중학교 어휘는 1,091개, 고등학교 어휘는 1,254개로 각 학교급별 수준이 결정되었다. 먼저, 초등학교 어휘 중 초등학교 3학년 어휘는 98개, 4학년은 83개, 5학년은 196개, 6학년은 157개의 학년별 수준으로 구분되었다. 다음으로 중학교 어휘 중 중학교 1학년 어휘는 390개, 2학년은 357개, 3학년은 344개로 제안하였다. 마지막으로, 고등학교 어휘 중 고등학교 1학년의 어휘는 514개, 고등학교 2학년 이상의 어휘는 740개로 결정되었다. 이는 학교급이 높아질수록 학습해야 할 어휘량이 증가하고 있음을 보여주고 있다. 특히 학교급이 전환되는 시기인 초등학교 6학년 수준 어휘 수가 157개에서 중학교 1학년 수준 어휘 390개로 급격하게 증가하여 초등학교에 비해 중학교에서 어휘 학습 부담이 커질 수 있음을 보여준다. 또한 고등학교 학생들 역시 중학교에 비해 상대적으로 큰 어휘 학습 부담을 경험할 수 있음을 시사한다.

<sup>7</sup> [https://osf.io/jyu2h/overview?view\\_only=f00ae15aab1c4994b51d73621f38681f](https://osf.io/jyu2h/overview?view_only=f00ae15aab1c4994b51d73621f38681f)

표 4. 세 가지 기준에 기반한 기본 어휘 목록의 세부 학년 수준

학년	어휘 수	예시
초등학교 3학년	98	<i>about, banana, cat, dance, fish</i>
초등학교 4학년	83	<i>apple, busy, cook, friend, late</i>
초등학교 5학년	196	<i>bake, bird, camp, cousin, fall</i>
초등학교 6학년	157	<i>answer, break, carry, elevate, festival</i>
중학교 1학년	390	<i>aid, bend, champion, coast, count</i>
중학교 2학년	357	<i>able, apply, bump, chew, continue</i>
중학교 3학년	344	<i>accept, appreciate, belong, consider, describe</i>
고등학교 1학년	514	<i>advantage, arrange, branch, civil, constant</i>
고등학교 2학년 이상	740	<i>abound, assault, bundle, compensate, declare</i>
총 어휘 수	2,879	

## 5. 결론 및 제언

본 연구는 2022 개정 영어과 교육과정에서 제시한 3,000개의 기본 어휘 목록을 중심으로, 2015 개정 초·중·고등학교 영어 교과서에 실제로 나타난 어휘 양상을 분석하여 각 어휘의 적절한 세부 학년 수준을 탐구하였다. 이를 위해 3가지 핵심 기준인 범위(range), 유형 빈도(type frequency), 출현 빈도(token frequency)를 설정하고, 이 기준들을 종합적으로 활용하여 각 어휘 항목별 적정 학년 수준을 최종적으로 도출하였다.

첫 번째 연구 질문과 관련하여, 본 연구는 피어슨 상관관계와 ICC 결과를 분석하였다. 그 결과, 3가지 기준(범위, 유형 빈도, 출현 빈도 기준)이 각각 도출한 세부 학년 결과 간에 유의미한 상관관계가 있음을 확인하였다. 예를 들어, 범위와 유형 빈도 간에는 강한 상관관계( $r = .757, p < .001$ ), 범위와 출현 빈도 간에는 중간 이상( $r = .635, p < .001$ )의 상관관계가 있음을 확인하였다. 또한, 3가지 기준을 적용하여 얻어낸 세부 학년 결과는 높은 수준의 신뢰도와 일관성( $ICC(3, k) = .859, F = 7.12, p < .001$ )을 보여주었다. 이는 본 연구에서 제안한 3가지 기준이 어휘 항목의 세부 학년 수준 결정에 있어 상호 보완적이면서도 안정적인 지표가 될 수 있음을 시사한다.

두 번째 연구 질문과 관련하여, 본 연구는 3가지 기준(범위, 유형 빈도, 출현 빈도 기준)에 의해 산출된 결과를 종합적으로 고려하여, 총 2,879개의 개별 어휘에 대한 최종 세부 학년 수준을 제시하였다. 특히, 3가지 기준에 의한 결과가 완전히 일치하지 않을 경우, 중간값(median)의 개념을 사용하여 이상치(outlier)를 제거하는 세부 학년 결정 과정을 거쳤다. 또한 질적 검토를 추가하여 어휘 항목의 학년 수준 결정에 신뢰도를 제고하였으며, 표 4에서 각 학년별 어휘 수와 그 예시 어휘 목록을 확인할 수 있다. 특히, 초등학교에서 중학교로, 그리고 중학교에서 고등학교로 전환되는 시점에 학습 어휘량이 큰 폭으로 증가하는 양상이 확인되었는데, 이는 학교급 간 어휘량 증가가 보다 점진적으로 이루어질 수 있도록 상위 수준으로 분류된 일부 어휘를 하향 조정하는 등 어휘 수준 배치에 대한 재검토의 필요성을 시사한다.

본 연구는 2022 개정 영어과 교육과정의 기본 어휘 항목 3,000개를 실제 교과서 사용 맥락

속에서 세부 학년 수준을 결정하려는 첫 시도라는 점에서 그 의의를 갖는다. 특히, 기존의 어휘 수준 연구들이 교육과정상의 학교급 범주에만 머물렀던 데 비해, 본 연구는 체계적인 3가지 기준(범위, 유형 빈도, 출현 빈도 기준)과 중간값의 개념을 활용하여 세부적으로 적정한 학년별 어휘를 제시하였다. 이러한 개별 어휘 항목에 대한 세부 학년 수준 정보는 기초학력 수준 측면에서 어휘 교육의 최소 성취 기준 설정, 교과서 개발 시 학년별 어휘 구성의 적절성 판단, 학습자 어휘 발달 연구 및 진단 평가 기준 마련 등 다양한 활용 가능성을 제공한다.

그러나 본 연구는 몇 가지 제한점을 갖는다. 첫째, 본 연구의 어휘 수준 결정에 활용한 3가지 기준은 연구자의 판단에 의존할 수밖에 없었기 때문에, 선정 기준의 임의성 측면에서 한계를 갖는다. 그러나 이는 어휘 수준을 결정하는 선행 연구 전반에서 발견할 수 있는 동일한 제한점이다. 둘째, 출현 빈도 기준에서 3-1 규칙을 적용하지 못하는 경우 최초 등장 학년 규칙을 적용한 방식은 범위 기준과 유형 기준과의 각각의 상관관계에 부정적인 영향을 미쳤을 뿐 아니라, 저빈도 어휘의 오분류 가능성을 야기하는 한계를 갖는다. 이러한 한계점은 향후 연구에서 보완되어야 한다. 일례로, 일정 수 이상의 총 빈도수를 보이는 어휘에만 3-1 규칙을 적용하고, 이보다 낮은 총 빈도수를 지닌 어휘는 해당 규칙을 적용하지 않고 질적 검토를 통해 판단하는 방법을 고려해볼 수 있다. 셋째, English Vocabulary Profile과 달리 본 연구는 단어의 개별 의미 수준까지 분리하여 분석하지 못했다는 점에서 한계를 갖는다. 이는 향후 연구를 통해 보완되어야 할 과제로 남는다. 넷째, 본 연구가 제시한 세부 학년 수준의 교육 현장 타당성은 후속 연구에서 현직 교사와 교재 개발자에 의한 내용 타당도 검토를 통해 추가로 확인될 필요가 있다. 마지막으로, 본 연구는 2022 개정 영어과 교육과정의 기본 어휘를 준거 목록으로 삼았으나, 실제 사용 양상 분석에는 2015 개정 영어 교과서 코퍼스를 활용하였다. 따라서 본 연구의 결과는 2022 개정 기본 어휘의 세부 학년 수준을 직접 확인한 것이라기보다, 현시점에서 확보 가능한 전 학년 교과서 자료를 바탕으로 그 수준을 추정한 결과로 이해할 필요가 있다. 그러나 본 연구에서 제안한 방법론은 향후 2022 개정 교육과정 기반 교과서에서도 적용 가능하다는 점은 주목할 만하다. 요약하면, 본 연구의 결과는 세부 학년별 어휘 수준 결정에 있어 유용한 방법론과 실증적 근거를 제시했다는 점에서 의의를 갖는다.

## 참고 문헌

- 교육부(Ministry of Education). 2022. 『2022 개정 영어과 교육과정』 (2022 Revised National Curriculum for English). 세종: 교육부(Sejong: Ministry of Education).
- 김기택, 성민창(Kim, K. and M. Sung). 2019. 초등영어 학습자 어휘 수준에 적합한 영어 애니메이션 영화 탐색(Exploring English-language animated movies lexically appropriate for Korean elementary school students). <<영어학>>(Korean Journal of English Language and Linguistics) 19(4), 837-857.
- 성민창, 김기택(Sung, M. and K. Kim). 2021a. 2015 개정 영어과 교육과정 초등 수준 어휘 목록을 활용한 영어 애니메이션 영화의 어휘 사용 분석(Analysis of the vocabulary level in English-language animated movies based on the vocabulary list in the Korean national curriculum for English). <<영어학>>(Korean Journal of English Language and Linguistics) 21, 87-109.

- 성민창, 김기택(Sung, M. and K. Kim). 2021b. 우리나라 초등영어교육에 적합한 팝송 탐색: 언어 수준과 내용 분석(Pop songs for primary English education in South Korea: An analysis of language level and content). <<영어학>>(Korean Journal of English Language and Linguistics) 21, 450-471.
- 성민창, 김은형, 김기택(Sung, M., E.-H. Kim and K. Kim). 2023. 초등영어 어휘 수준에 적합한 영어 아동문학 작품 탐색: 코퍼스 기반 연구(Children's English novels lexically appropriate for Korean elementary school students: A corpus-based study). <<어학연구>>(Language Research) 59(1), 43-68.
- 오영교(Oh, Y.). 2025. 인공지능 디지털 영어교육의 현황과 전망: 자연어처리 기반 텍스트 분석(Issues and prospects of AI digital English education: Text analysis based on natural language processing). <<영어학>>(Korean Journal of English Language and Linguistics) 25, 330-366.
- 이문복, 신동광(Lee, M.-B. and D.-K. Shin). 2015. 2015 영어과 교육과정 기본 어휘 목록 개발(Development of the Korean basic English word list of the 2015 revised national curriculum of English). <<영어교과교육>>(Journal of the Korea English Education Society) 14(4), 115-134.
- 이윤(Lee, Y.). 2016. 초등영어 교과서 읽기 텍스트의 어휘 난이도(Word difficulty of reading text in elementary school English textbooks). <<외국어교육>>(Foreign Languages Education) 23(1), 221-242.
- 이현우(Lee, H.). 2023. 2022 개정 영어과 교육과정 기본 어휘 목록: 개발과 적용(The basic vocabulary list of the 2022 revised national curriculum of English: Development and application). <<영어학>>(Korean Journal of English Language and Linguistics) 23, 59-88.
- 장경숙, 정규태, 이병천(Chang, K., K. Jung and B. Lee). 2011. 2009 교육과정 개정에 따른 영어과 기본 어휘 목록 및 지침 개발을 위한 기초 연구(A preliminary study on English basic vocabulary list and guidelines for the 2009 Revised National Curriculum). <<현대영어교육>>(Modern English Education) 12(2), 190-218.
- 장세영, 손소혜, 심지혜, 박예은, 이동환(Jang, S., S. Son, J. Sim, Y. Park and D.-H. Lee). 2024. 영어과 교육과정의 어휘 목록과 규정 분석: 초등영어를 위한 주제적 분류와 시사점(An analysis of the English curriculum vocabulary lists and regulations: A thematic clustering for primary English education). <<초등영어교육>>(Primary English Education) 30(3), 71-95.
- 조규희, 이동환(Jo, K. and D.-H. Lee). 2025. 머신러닝 활용 초등영어 교과서 텍스트 군집화 및 그림책 텍스트 분류 연구(A machine-learning approach to clustering primary English textbook dialogues and classifying supplementary picture-book texts). <<초등영어교육>>(Primary English Education) 31(2), 37-59.
- 주형미, 김미경, 김성혜, 배주경 외(Joo, H. M., M. Kim, S. Kim, J.-K. Pae, et al.). 2022. 『2022 개정 영어과 교육과정 시안(최종안) 개발 연구』 (A Study on the Development of the 2022 Revised National Curriculum of English: Part 2). 세종: 교육부(Sejong: Ministry of Education).
- 최서용, 김기택, 오선영(Choi, S., K. Kim and S.-Y. Oh). 2025. 2015 개정 영어과 교육과정 기본 영어 교과서 문법 항목의 학년별 수준 구분(Grade-level classification of grammar items in English textbooks based on the 2015 revised English curriculum). <<영어교육>>(English Teaching) 80(2), 29-55.
- 최원경(Choi, W.). 2025. 2022 개정 영어과 교육과정 기본 어휘 목록의 적절성 분석: 접사 규정을 중심으로(Analysis of the appropriateness of basic vocabulary list of 2022 revised English curriculum: Focused on regulations related to affixes). <<초등영어교육>>(Primary English Education) 31(1), 5-24.
- 최정은, 이병민, 오선영, 소영순(Choi, J., B. Lee, S.-Y. Oh and Y. So). 2022. 2015 개정 영어과 교육과정의

- 성취 수준과 대학수학능력 영어 평가 수준의 일관성 문제(Consistency between the targeted achievement levels of the 2015 revised national curriculum of English and the assessment levels of the College Scholastic Ability Test of English in South Korea). <<한국영어평가>>(English Language Assessment) 17(1), 11-43.
- 황필아(Hwang, P.-A.). 2019. A study on the adequacy of the level of vocabulary and meanings in the primary English textbooks revised in 2015(2015 개정 초등영어 교과서에 나타난 어휘 및 의미 단계의 적정성 연구). <<영어어문교육>>(English Language & Literature Teaching) 25(4), 153-172.
- Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan and R. Quirk. 1999. *Longman Grammar of Spoken and Written English*. New York: Longman.
- Brezina, V., P. Weill-Tessier and T. McEnery. 2020. #LancsBox v.5.x [Software]. Lancaster University. Available Online at <http://corpora.lancs.ac.uk/lancsbox>
- Browne, C., B. Culligan and J. Phillips. 2013. *New General Service List: The Most Important Words for Second Language Learners of English*. Available Online at <http://www.newgeneralservicelist.org>
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2), 213-238.
- Gries, S. Th. 2015. Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics* 16(1), 93-117.
- Hawkins, J. A. and L. Filipović. 2012. *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework (Vol. 1)*. Cambridge, England: Cambridge University Press.
- Hirsh, D. and I. S. P. Nation. 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8(2), 689-696.
- Kim, J., D. Isbell and K. Kim. 2025. The influence of CEFR in South Korea's 2022 national English curriculum: A focus on grammar. *English Today* (Online First), 1-6.
- Kim, Y. and S. Oh. 2020. A corpus-based analysis of collocations in Korean middle and high school English textbooks. *Language Research* 56(3), 437-461.
- Krashen, S. 1982. *Principles and Practices in Second Language Acquisition*. New York: Pergamon.
- Laufer, B. 1989. What percentage of text-lexis is essential for comprehension? In C. Lauren and M. Nordman, eds., *Special Language: From Humans Thinking to Thinking Machines*, 316-323. Clevedon, England: Multilingual Matters.
- Liu, N. and I. S. P. Nation. 1985. Factors affecting guessing vocabulary in context. *RELC Journal* 16, 33-42.
- Long, M. H. 1983. Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics* 4(2), 126-141.
- Milliner, B. and G. Pinchbeck. 2025. Challenging lexical coverage conventions: Evaluating the vocabulary demands of family-genre film and television. *Research Methods in Applied Linguistics* 4(3), 100230.
- Nation, I. S. P. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63(1), 59-82.
- Nation, I. S. P. 2013. Vocabulary size in a second language. In C. A. Chapelle, ed., *The Encyclopedia of Applied Linguistics*, 6674-6677. Hoboken, NJ: Wiley-Blackwell.
- Nation, I. S. P. 2017. *The BNC/COCA Level 6 Word Family Lists (Version 1.0.0)*. Available Online at <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- O'Keeffe, A. and G. Mark. 2017. The English Grammar Profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics* 22(4), 457-489.
- Pienemann, M. 1998. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

- Rayson, P. and A. Potts. 2021. Analysing keyword lists. In M. Paquot and S. Th. Gries, eds., *A Practical Handbook of Corpus Linguistics*, 119-139. Cham, Switzerland: Springer International Publishing.
- Shrout, P. E. and J. L. Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2), 420-428.
- Sung, M. and H. Kim. 2022. Effects of verb-construction association on second language constructional generalizations in production and comprehension. *Second Language Research* 38(2), 233-257.
- Tono, Y. 2019. Coming full circle—From CEFR to CEFR-J and back. *CEFR Journal—Research and Practice* 1(1), 5-17.
- Webb, S. and M. P. Rodgers. 2009. The lexical coverage of movies. *Applied Linguistics* 30(3), 407-427.

예시 언어(Examples in): English

적용 가능 언어(Applicable Languages): English

적용 가능 수준(Applicable Level): Elementary, Secondary