

2021학년도 대학수학능력시험 영어 영역과 영어 I-II 교과서의 어휘 비교*

이현우** · 류혜경***

- 차례 -

1. 서론
2. 이론적 배경
3. 연구 방법 및 절차
4. 연구 결과 및 논의
5. 결론

* 이 논문은 인하대학교의 지원에 의하여 연구되었음.

** 제1 저자, 인하대학교 영어교육과 교수, hylee@inha.ac.kr

*** 교신 저자, 대구대학교 영어영문학과 교수, hryoo2002@daegu.ac.kr

[국문초록]

본 논문은 2021학년도 대학수학능력시험 영어 영역과 이 시험의 바탕이 되는 *영어 I-II* 교과서가 어휘 정교성, 어휘 밀도와 어휘 다양성에 있어서 유의차가 있는지의 문제를 다뤘다. 이를 위해 여섯 개의 귀무가설을 설정하고 이를 검정하였다. 검정 결과, 구어 자료가든지 문어 자료가든지 간에 자료 간에 등급화된 단어 출현형의 빈도와 내용어의 빈도가 자료에 따라 다르다는 것이 밝혀졌다. 또한 자료들의 STTR 값 비교와 길이를 같게 한 자료를 대상으로 한 카이제곱 검정을 통해 몇몇 자료 간에 유의한 어휘 다양성의 차가 존재함을 확인하였다. 한 마디로, 어떤 교과서의 구어 또는 문어 자료는 어휘 정교성, 어휘 밀도, 또는 어휘 다양성이 2021학년도 수능 영어 시험 또는 다른 교과서 자료보다 유의하게 높거나 낮게 나왔다.

주제어: 2015 개정 영어과 교육과정, 대학수학능력시험 영어 영역, *영어 I-II* 교과서, 어휘 정교성, 어휘 밀도, 어휘 다양성

1. 서론

대학수학능력시험은 고등학교 3학년 학생들에게 대학 입학에 위한 가장 중요하고 어려운 절차 중의 하나라고 할 수 있다. 대학 입시에서 차지하는 이 시험의 중요성과 학생과 학부모, 그리고 사회 전반으로부터의 높은 관심을 고려할 때 시험 유효성을 포함하는, 대학수학능력시험의 전반적인 평가 내용과 척도에 관한 지속적인 연구가 필요하다. 특히, 대학수학능력시험 영어 영역(이후 수능 영어로 칭함)은 사교육 의존으로 인한 과중한 경제적 부담과 관련하여 사회 전반에 많은 논란을 일으켜 왔고, 수능 영어의 전반적인 난이도와 유효성에 대한 비판이 많이 있었다. 이에 대한 반응으로 교육부는 2018학년도부터 수능 영어를 절대평가로 전환했다. 이러한 정책적 변화에 따라 수능 영어와 이 시험의 출제 범위가 되는 *영어 I* 과 *영어 II* 교과서와의 관계를 규명할 필요성이 대두되었다.

한국교육과정평가원은 「2021학년도 대학수학능력시험 시행기본계획」에서 수능 영어는 2015 개정 영어과 교육과정의 성취 기준 달성 정도와 대학 수학에 필요한 영어 사용 능력의 측정을 목적으로 고등학교 교육과정의 영어 교과인 *영어 I* 과 *영어 II*를 바탕으로 다양한 소재의 지문과 자료를 활용하고, 교육 과정에 포함되는 기본 어휘와 시험 과목 수준의 어휘 가운데 사용 빈도가 높은 것을 사용하여 출제한다고 명시하고 있다. 더 구체적으로 말해, 2021학년도 수능 영어는 2015 개정 영어과 교육과정을 기반으로 제작된 2019, 2020년판 *영어 I* 과 *영어 II* 교과서를 바탕으로 한다.

2015 개정 영어과 교육과정은 *영어 I* 과 *영어 II* 교과서에서 사용할 수 있는 어휘 수를 각각 1,800과 2,500 낱말 수 이내로 한정하고 있다. 여기에서 ‘낱말’은 단어 군을 지칭하며 ‘이내’는 사용 가능한 총 어휘 수의 5% 범위 내에서 사용할 수 있다는 것을 뜻한다. 주목할 점은 교육과정이 각 교과서가 권장 어휘 수의 80% 이상을 기본 어휘 목록에서 사용할 것을 권장하고 있다는 것이다. 이에 근거하여 교과서에 포함될 수 있는 기본 어휘가 아닌 어휘의 수는 *영어 I* 은 최대 378 단어 군, *영어 II*는 최대 525 단어 군이다. 중요한 것은 학생 대부

분이 동일한 필진의 *영어 I* 과 *영어 II* 교과서로 공부하는데, 이들 *영어 I* 과 *영어 II* 교과서의 합본(이후 *영어 I-II* 교과서로 칭함)에 포함된 어휘를 익히면 어휘 면에서 수능 영어를 충분히 대비할 수 있는지의 문제이다. 다시 말해, *영어 I-II* 교과서에서 사용된 어휘(기본 어휘와 기본 어휘 외의 어휘) 수준이 이들 교과서를 바탕으로 한다고 표방한 수능 영어의 어휘(기본 어휘와 기본 어휘 외의 어휘) 수준과 동일한지를 따지는 것이 그 무엇보다도 필요해 보인다.

이 문제를 해결하기 위해서 본 연구는 다음 귀무가설을 검정하고자 한다.

귀무가설 1: 2021 수능 영어와 *영어 I-II* 교과서의 구어 자료는 어휘 정교성에 있어서 동일하다.

귀무가설 2: 2021 수능 영어와 *영어 I-II* 교과서의 문어 자료는 어휘 정교성에 있어서 동일하다.

귀무가설 3: 2021 수능 영어와 *영어 I-II* 교과서의 구어 자료는 어휘 밀도에 있어서 동일하다.

귀무가설 4: 2021 수능 영어와 *영어 I-II* 교과서의 문어 자료는 어휘 밀도에 있어서 동일하다.

귀무가설 5: 2021 수능 영어와 *영어 I-II* 교과서의 구어 자료는 어휘 다양성에 있어서 동일하다.

귀무가설 6: 2021 수능 영어와 *영어 I-II* 교과서의 문어 자료는 어휘 다양성에 있어서 동일하다.

2. 이론적 배경

2.1. 어휘 풍부성의 척도

코퍼스 언어학 기반 어휘 연구에서 가장 기본이 되는 것 중의 하나는 단

어를 어떻게 정의하는지에 관한 것이다. 기존의 응용언어학 연구에서 단어의 개념으로 널리 알려진 것은 출현형(token), 유형(type), 기본형(lemma) 그리고 군(family)이다(Read 2000, Nation 2001). 단어 출현형은 하나의 텍스트에 쓰인 모든 단어의 존재를 뜻한다. 이 개념에서는 한 단어가 반복적으로 쓰일 때 반복적인 형태 각각 별개의 출현형으로 여겨진다. 단어 출현형과 달리, 단어 유형은 다른 것과 구분되는 유일무이의 단어 형태를 의미한다. 이 개념에서는 한 단어가 아무리 반복해서 쓰였다고 하더라도 하나의 단어 유형으로 간주된다. 하나의 단어 유형이 복수의 출현형으로 쓰인 경우이다. 단어 기본형은 문법적으로 관련된 일단의 단어를 대표한다. 예를 들어, 표제어(headword) *speak*은 동사 *speak*과 이 동사의 모든 굴절형 *speaks*, *spoke*, *spoken*, *speaking*을 대표한다. 단어 군은 기본형이 포함하는 표제어의 굴절형 뿐만 아니라 파생형(*speaker*, *speakers*, *unspeakable*, *unspeakably*, *unspoken* 등)도 포함한다.

최근의 코퍼스 언어학 기반 어휘 연구에서 어휘 풍부성을 어떻게 측정할지가 많은 연구자의 관심을 끌고 있다. 어휘 풍부성이란 하나의 텍스트에서 단어가 얼마나 많이 사용되었는가를 의미한다. 측정하는 방식에 따라서 어휘 풍부성은 대략 어휘 정교성, 어휘 밀도, 어휘 다양성으로 세분된다. 어휘 정교성은 텍스트에 나오는 고급 단어의 비율을 뜻한다. 어휘 밀도는 텍스트에 나오는 총 단어 대비 내용어의 비율을 뜻한다. 어휘 다양성은 텍스트에 쓰인 유일무이의 단어 형태(즉, 단어 유형)가 얼마나 다양한지를 따진다(Read 2000, Daller, Milton and Treffers-Daller, 2007).

어휘 정교성을 설명하기 위해, Meara and Bell(2001)은 다음 예문을 제시하고 있다.

- (1) The man saw the woman.
- (2) The bishop observed the actress.
- (3) The magistrate sentenced the burglar.

위 (1)-(3)의 문장은 단어 출현형과 유형의 수에 있어서 같다. 하지만 (1)의 단어가 (2)의 단어보다, (2)의 단어가 (3)의 단어보다, 더 고빈도로 더 쉽다. 다른 모든 조건이 같다면, (2)의 문장이 포함된 글이 (1)의 문장이 포함된 글보다, 그리고 (3)의 문장이 포함된 글이 (2)의 문장이 포함된 글보다, 더 어렵다. 이런 결과는 어휘 정교성의 차이로 야기된 결과이다.

어휘 밀도는 명사, 동사, 형용사, 부사에 해당하는 내용어가 텍스트에서 차지하는 비율을 뜻한다. 대명사, 조동사, 전치사, 일부 부사, 접속사, 수사 등의 기능어보다 내용어의 비율이 높을수록 텍스트의 정보가 더욱더 조밀하게 포장되었다고 여겨진다. 대부분은 내용어와 기능어를 쉽게 구분할 수 있으나 몇몇 경우는 구분이 그리 간단치 않다. 본 논문에서는 Ure(1971)를 따라, *every, some, no*와 같은 양화사, 서법 부사, 구동사의 소사(particle) 그리고 전치사를 기능어로 취급한다.

텍스트에서 새 단어가 더 많이 사용되고 단어의 반복이 적을수록 어휘 다양성이 더 높아진다. 이러한 어휘 다양성을 계량하는 가장 전통적인 척도는 텍스트의 단어 유형의 수를 단어 출현형 총수로 나눈 Templin(1957)의 유형-출현형 비율(type-token ratio, TTR)이다. 하지만 TTR은 텍스트가 길어질수록 새로 추가되는 단어 유형의 수가 점점 줄어들기 때문에, 일반적으로 더 긴 텍스트가 더 짧은 텍스트보다 TTR 값이 낮게 나오는 문제점을 안고 있다. 이런 문제점 해결을 위해, Guiraud(1954)는 Guiraud 지수(GI)를 고안했다. GI는 단어 유형의 수를 단어 출현형 수 대신 이의 제곱근으로 나눈 값이다. 길이가 다른 텍스트의 어휘 다양성 비교에서 GI가 일반적으로 유용하나 본 연구에서처럼 길이의 차이가 있는 경우에는 신뢰할 수 없으므로 본 연구에서는 Scott(2016)의 Standardised TTR(STTR)을 이용한다. STTR은 분석할 텍스트를 임의의 단어 출현형 수(일반적으로 1,000)대로 나눈 후 분할된 각 부분의 TTR로 평균을 구한 값을 뜻한다.

Malvern et al(2004)에 따르면, 어휘 다양성은 어휘 정교성과 관계가 밀접하다. 텍스트의 어휘 다양성이 높아지려면 텍스트에 새 단어를 추가해야 하는데,

이렇게 추가되는 단어가 많이 있을수록 고급 단어가 더 많이 사용되기 때문이다.

Daller, van Hout and Treffers-Daller(2003)는 어휘 정교성을 계량하는 척도로 Advanced Guiraud(AG)를 제안하였다. AG는 고급 단어 유형의 수를 단어 출현형의 수의 제곱근으로 나눈 값이다. 다음 장에서 보듯이, 본 연구에서는 고빈도 2K 또는 1K 단어 수준 이상의 단어를 고급 단어로 간주한다.

2.2. 선행 연구

코퍼스 언어학의 기법을 활용하여, 영어 교과서나 수능 영어와 같은 자료를 상호 비교·분석하는 연구가 주목을 받고 있다. 이들 연구는 대체로 이들 자료 사이에 어휘량의 차이가 존재하며 이러한 차이의 원인과 교육적 시사점을 논하고 있다.

오준일(2014)은 2009 개정 영어과 교육과정을 기반으로 한 4종의 *영어II* 교과서와 수능-EBS 연계 교재인 『수능완성』의 읽기 지문에 사용된 어휘 수준을 분석하였다. 분석 결과, 이들 교과서보다 『수능완성』에서 저빈도 고급 어휘와 학술 어휘의 사용 비율이 더 높았다. 오준일(2014)은 EBS-수능 연계 교재를 원활하게 이해하기 위해서는 교과서보다 20% 정도 더 많은 단어 군을 알아야 한다고 주장했지만, 이 주장은 수능 영어의 바탕이 되는 *영어I* 과 *영어II* 교과서 대신 *영어II* 교과서만 비교 대상으로 하였기에 설득력이 떨어진다.

김제우와 이동주(2017)는 코퍼스 언어학적 분석을 통해 2009 개정 영어과 교육과정에 따른 *영어 I-II* 교과서, 수능 영어 및 모의고사, 수능-EBS 연계 교재에 나오는 어휘를 분석하였다. 이들은 수능-EBS 연계 교재의 어휘 수준이 교과서와 수능 영어의 어휘 수준보다 높으며, 수능-EBS 연계정책 시행 이후의 수능 영어가 연계 이전의 수능 영어보다 어휘 수준이 높아졌다는 점을 지적하고 있다. 하지만 *영어 I-II* 각 교과서와 수능 영어의 어휘 수준을 직접적으로 비교하는 분석이 없는 문제점을 안고 있다.

노윤오(2019)는 이소영과 이현우(2017)의 단어 등급화 방법론에 따라 영어 교과서와 수능 영어의 어휘를 다양한 방식으로 분석하였다. 먼저, *영어 I* 교과서와 2017학년도와 2018학년도 수능 영어의 어휘를, *영어 II* 교과서와 2017학년도와 2018학년도 수능 영어의 어휘를, 마지막으로 *영어 I-II* 교과서와 2017학년도와 2018학년도 수능 영어의 어휘를 비교하였다. 세 층위 각각에서 교과서 간에 그리고 교과서와 수능 영어 간에 등급화된 단어 출현형이나 유형의 분포에 있어서 유의차가 있음을 확인하였다. 이러한 결과에도 불구하고, 노윤오(2019)는 어휘 분석을 어휘 정교성에만 국한하여 수행했다는 한계점을 안고 있다.

이현우(2019)는 *영어 I-II* 교과서의 읽기 본문과 2018학년도와 2019학년도 수능 영어 읽기 지문을 어휘 정교성, 어휘 밀도, 어휘 다양성의 측면에서 비교하였다. 분석 결과, 어휘 정교성과 어휘 밀도와 관련하여, 교과서 간에 그리고 교과서와 수능 영어와도 유의차가 있었다. 어떤 교과서는 어휘 정교성이나 어휘 밀도에 있어서 수능 영어보다 높았고 어떤 교과서는 낮았다. 전반적으로 수능 영어보다 영어 교과서에서 어휘가 더 다양하게 사용된 것으로 나왔으나, 이는 어휘 다양성의 척도로 GI를 사용했고 교과서가 수능 영어보다 길이가 압도적으로 길었기 때문으로 여겨진다.

3. 연구 방법 및 절차

3.1. 분석 대상

1장에서 언급한 대로, 본 연구의 분석 대상은 2021학년도 수능 영어와 이 시험이 토대로 한 9종 영어 교과서이다. 어휘의 사용이 구어와 문어 텍스트에 따라 다르므로, 이들 시험과 교과서를 구어와 문어 자료로 구분한다. 수능 시험의 구어 자료는 이 시험의 듣기와 (간접) 말하기 문항의 대화와 영어 선지로 되

어 있고, 문어 자료는 읽기와 (간접) 쓰기 문항의 지문과 영어 선지로 되어 있다. 반면에, 교과서의 구어 자료는 대표 저자가 같은 영어 I-II 듣기와 말하기 부분의 대화로 구성되어 있고, 문어 자료는 영어 I-II의 읽기 본문으로 구성되어 있다. 이러한 읽기 본문에는 각 단원에서 ‘More Reading’이라는 이름으로 추가로 제공하는 읽기 본문이나 Special Lesson 형태로 추가로 제공하는 읽기 본문을 포함한다.

분석 대상 자료를 구어와 문어와 구분하여 비교한 결과는 아래 <표1>과 같다.

<표1> 분석 대상 자료의 단어 제원

구분	분석 대상	단어 출현형	단어 유형
구 어	2021 수능 영어	2,322	741
	교과서 A	12,521	2,075
	교과서 B	7,368	1,558
	교과서 C	5,464	1,258
	교과서 D	7,141	1,726
	교과서 E	5,088	1,280
	교과서 F	4,846	1,177
	교과서 G	8,189	1,717
	교과서 H	4,999	1,222
	교과서 I	6,150	1,375
	구어 전체	64,088	5,503
문 어	2021 수능 영어	4,189	1,470
	교과서 A	16,485	3,107
	교과서 B	11,162	2,648
	교과서 C	11,224	2,451
	교과서 D	14,647	3,240
	교과서 E	11,773	2,583
	교과서 F	16,622	3,489
	교과서 G	14,662	3,112
	교과서 H	13,864	2,727
	교과서 I	9,018	2,249
	문어 전체	123,646	10,159

2021 수능 영어 구어 자료에서는 2,322개의 단어 출현형과 741개의 단어 유형이 사용되었고, 문어 자료에서는 4,189개의 단어 출현형과 1,470개의 단어 유형이 사용되었다. *영어 I-II* 교과서 구어 자료에서는 최대 12,521개와 최소 4,846개(평균 6,863개)의 단어 출현형과 최대 2,075개와 최소 1,177개(평균 1,488개)의 단어 유형이 사용되었다. 교과서 문어 자료에서는 최대 16,622개와 최소 9,018개(평균 13,273개)의 단어 출현형과 최대 3,489개와 최소 2,249개(평균 2,845개)의 단어 유형이 사용되었다.

3.2. 단어의 등급화

어휘 정교성의 문제를 해결하려면 먼저 단어를 등급화할 필요가 있다. 이를 위해 먼저 2015 개정 영어과 교육과정의 기본 어휘를 살펴볼 필요가 있다. 기본 어휘는 모두 3,000개의 단어 군으로 되어 있는데, 이 가운데 800개는 초등 과정에서 사용하기를 권장하고 400개는 진로 선택 및 전문 교과 I 에서 사용하기를 권장하고 있다. 즉, 기본 어휘는 ‘초등’, ‘일반’, ‘전문’의 3등급으로 나뉘어 있다고 볼 수 있다. 하지만 ‘일반’ 등급에 1,800개의 단어 군이 밀집되어 있어 교육과정의 기본 어휘 등급화는 적절한 것으로 보기에 어렵다. 더군다나, 2015 개정 영어과 교육과정은 생활 주변에서 흔히 사용하고 있는 외래어의 근원어(source word) 200개 가운데 학년별로 50개씩 사용하는 것이 가능하며 이는 새로운 어휘로 간주하지 않고 있다. 하지만 정채관과 권혁승(2017)에 따르면, 이들 200개의 근원어의 굴절형(예를 들어, *computer*의 복수형 *computers*)과 파생형(예를 들어, *event*의 형용사형 *eventful*)은 교육과정의 기본 어휘 외의 어휘로 취급되고 있다. 기본 어휘 외의 어휘는 보통 저빈도의 고급 어휘로 학습용이성(learnability)이 낮은 어휘이다. 빈도와 학습용이성이 높다는 점에서 200개 근원어의 굴절형과 파생형을 단순히 기본 어휘 외의 어휘로 다루는 것은 문제가 있다.

반면에, 저빈도 고급 단어의 사용 척도로 AG를 주장하면서, Daller(2010)는

빈도가 가장 높은 단어 군 2K 또는 1K 이상 수준의 단어를 고급 단어로 정의하는 것이 좋다는 의견을 개진하고 있다. Adolphs and Schmitt(2003:431)에 따르면, 5백만 단어의 자연스러운 대화로 된 코퍼스 Cambridge and Nottingham Corpus of Discourse in English에서 최빈도 990 단어 군이 91.52%의 포괄률을 보인다. Nation(2006:79)은 1억 단어 출현형의 문어 텍스트(90%)와 구어 텍스트(10%)로 된 British National Corpus로부터 추출된 14개의 1K 단어 군 목록 중 첫 번째 목록의 단어 군이 고유명사와 함께 일반적인 문어 텍스트의 78-81%를 그리고 구어 텍스트의 약 85%를 포괄한다고 주장한다.

바로 위의 연구를 종합하면, 구어 텍스트에서의 고급 단어의 기준을 최빈도 1K로 하지만 문어 텍스트의 경우는 최빈도 2K가 적절해 보인다.¹ 2015 개정 영어과 교육과정의 기본 어휘 구분 토대에 이러한 점을 반영하여 단어 등급화를 하면 아래와 같다.

〈구어 텍스트용〉

그룹 1: 2015 개정 영어과 교육과정의 기본 어휘 중 ‘초등’ 등급의 800개 단어 군과 교육과정에서 정한 200개의 외래어 근원어의 굴절형과 파생형을 대표하는 표제어 193개(총 993개의 단어 군)

그룹 2: 그 외의 단어 군

〈문어 텍스트용〉

그룹 1: 구어 텍스트용 그룹 1에 속하는 993개의 단어 군과 2015 개정 영어과 교육과정의 기본 어휘 중 ‘일반’ 등급의 단어 군이면서 Brezina와 Gablasova(2015)의 *The New General Service List*의 최빈도 2,000 단어인 772개의 단어 군(총 1,715 단어 군)

그룹 2: 그 외의 단어 군

1 West(1953)의 *A General Service List of English Words* 또한 2,000 단어 군으로 되어 있다는 점을 참고 바람.

구어 텍스트용 단어 등급을 본 연구의 구어 자료에 적용한 결과는 <표2>와 같다.²

<표2> 구어 자료의 등급별 단어 제원

분석 대상	등급	단어 출현형	단어 유형
2021 수능 영어	그룹 1	1,847(83.5%)	428(62.8%)
	그룹 2	364(16.5%)	254(37.2%)
교과서 A	그룹 1	10,020(83.4%)	928(49.6%)
	그룹 2	1,998(16.6%)	942(50.4%)
교과서 B	그룹 1	5,977(83.8%)	797(55.2%)
	그룹 2	1,157(16.2%)	647(44.8%)
교과서 C	그룹 1	4,420(84.6%)	641(57.0%)
	그룹 2	807(15.4%)	484(43.0%)
교과서 D	그룹 1	5,461(80.8%)	780(51.2%)
	그룹 2	1,298(19.2%)	743(48.8%)
교과서 E	그룹 1	4,021(82.1%)	652(55.4%)
	그룹 2	874(17.9%)	525(44.6%)
교과서 F	그룹 1	3,851(83.2%)	609(56.9%)
	그룹 2	775(16.8%)	462(43.1%)
교과서 G	그룹 1	6,442(82.0%)	810(51.9%)
	그룹 2	1,413(18.0%)	751(48.1%)
교과서 H	그룹 1	4,029(83.6%)	650(57.5%)
	그룹 2	790(16.4%)	480(42.5%)
교과서 I 교과서	그룹 1	4,861(83.6%)	666(55.4%)
	그룹 2	954(16.4%)	536(44.6%)
전체	그룹 1	50,930(83.0%)	1,788(37.5%)
	그룹 2	10,429(17.0%)	2,980(62.5%)

2021 수능 영어 구어 자료에서는 1,847개의 그룹 1 단어 출현형(83.5%)과 364개의 그룹 2 단어 출현형(16.5%)이 사용되었다. 영어 I-II 교과서 구어 자료에서는 최대 10,021개, 최소 3,851개(평균 5,454개)의 단어 출현형과 최대 1,998개, 최소 775개(평균 1,118개)의 그룹 2 단어 출현형이 사용되었다. 비록 등급화된 단어 출현형의 빈도가 2021 수능 영어 구어 자료와 영어 I-II 교과서 구어 자료 간에 그리고 교과서 구어 자료 간에 매우 큰 편차를 보이지만, 그

룹 1과 2의 단어 출현형의 비율은 거의 동일하다. 2021 수능 영어 구어 자료에서는 428개의 그룹 1 단어 유형(62.8%)과 254개의 그룹 2 단어 유형(37.2%)이 사용되었다. 교과서 구어 자료에서는 최대 928개, 최소 609개(평균 726개)의 단어 유형과 최대 942개, 최소 462개(평균 619개)의 그룹 2 단어 유형이 사용되었다. 그룹 1과 2의 단어 유형의 빈도뿐만 아니라 비율도 2021 수능 영어 구어 자료와 영어 I-II 교과서 구어 자료 간에 그리고 교과서 구어 자료 간에 다소 큰 편차를 보인다. 다른 자료와 달리 교과서 A에서는 그룹 2의 단어 유형이 그룹 1의 단어 유형보다 더 많이 사용되었다.

문어 텍스트용 단어 등급을 문어 자료에 적용한 결과는 <표3>과 같다.

<표3> 문어 자료의 등급별 단어 체원

분석 대상	등급	단어 출현형	단어 유형
2021 수능 영어	그룹 1	3,683(90.7%)	1,053(75.4%)
	그룹 2	377(9.3%)	344(24.6%)
교과서 A	그룹 1	14,814(93.7%)	2,124(73.9%)
	그룹 2	998(6.3%)	749(26.1%)
교과서 B	그룹 1	9,795(91.3%)	1,790(73.2%)
	그룹 2	931(8.7%)	655(26.8%)
교과서 C	그룹 1	10,012(93.0%)	1,727(76.2%)
	그룹 2	755(7.0%)	538(23.8%)
교과서 D	그룹 1	12,846(91.7%)	2,151(71.7%)
	그룹 2	1,164(8.3%)	851(28.3%)
교과서 E	그룹 1	10,372(92.2%)	1779(73.9%)
	그룹 2	874(7.8%)	627(26.1%)
교과서 F	그룹 1	14,577(92.4%)	2323(72.6%)
	그룹 2	1,203(7.6%)	877(27.4%)
교과서 G	그룹 1	12,590(90.5%)	2016(71.2%)
	그룹 2	1,326(9.5%)	815(28.8%)
교과서 H	그룹 1	12,446(93.5%)	1854(74.6%)
	그룹 2	864(6.5%)	632(25.4%)
교과서 I	그룹 1	7,708(91.1%)	1548(75.7%)
	그룹 2	755(8.9%)	497(24.3%)

전체	그룹 1	107,079(90.7%)	4,865(54.7%)
	그룹 2	11,011(9.3%)	4,026(45.3%)

2021 수능 영어 문어 자료에서는 3,683개의 그룹 1 단어 출현형(90.7%)과 377개의 그룹 2 단어 출현형(9.3%)이 사용되었다. *영어 I-II* 교과서 문어 자료에서는 최대 14,814개, 최소 7,708개(평균 11,684개)의 단어 출현형과 최대 1,326개, 최소 755개(평균 986개)의 그룹 2 단어 출현형이 사용되었다. 비록 등급화된 단어 출현형의 빈도가 2021 수능 영어 문어 자료와 *영어 I-II* 교과서 구어 자료 간에 그리고 교과서 구어 자료 간에 매우 큰 편차를 보이지만, 그룹 1과 2의 단어 출현형의 비율은 거의 동일하다. 2021 수능 영어 문어 자료에서는 1,053개의 그룹 1 단어 유형(75.4%)과 344개의 그룹 2 단어 유형(24.6%)이 사용되었다. 교과서 문어 자료에서는 최대 2,323개, 최소 1,548개(평균 1,924개)의 단어 유형과 최대 877개, 최소 497개(평균 693개)의 그룹 2 단어 유형이 사용되었다. 그룹 1과 2의 단어 유형의 빈도가 2021 수능 영어 구어 자료와 *영어 I-II* 교과서 구어 자료 간에 그리고 교과서 구어 자료 간에 큰 편차를 보이지만, 그룹 1과 2의 단어 유형의 비율은 편차가 줄었다(71.2%~76.2%).

3.3. 분석 도구 및 방법

2021 수능 영어와 *영어 I-II* 교과서에서 그룹 1과 2로 등급화된 단어가 어떤 빈도로 사용되었는지 알아보기 위해, Heatley, Nation과 Coxhead(2002)의 Range32와 Scott(2016)의 WordSmith Tools 7을 사용한다. 이들 소프트웨어를 상호보정하는 방식으로 이용하면 각 자료에서 쓰인 그룹 1과 2의 단어 군, 타입, 토큰의 빈도를 정확히 알 수 있다.

Range32의 작동을 위해 그룹 1과 2의 단어 그룹과 교육과정의 외래어 근원어 그룹 그리고 고유명사와 단위명 등으로 이뤄진 그룹으로 구성된 네 개의 베이스워드 목록이 필요하다. 베이스워드 목록은 표제어와 표제어의 굴절형

과 파생형을 포함한다. 예를 들어, 표제어 *nation*뿐만 아니라 *nations, national, nationally, nationality, international, unnational, nationalize, nationalizing, nationalized, nationalizes, nationalization, nationalizer* 등이 포함된다.

그룹 1과 2의 단어 출현형과 내용어와 기능어의 단어 출현형의 빈도를 조사한 후, 이들 빈도에 관해 2021 수능 영어와 영어 I-II 교과서가 동일하지를 따지기 위해 SPSS 26.0 for Windows를 활용하여 카이제곱 검정을 한다. 두 변수 간에 유의한 관계가 있는 경우, Bonferroni의 보정법에 따라 사후분석을 한다.

4. 연구 결과 및 논의

4.1. 귀무가설 1과 2의 검정

귀무가설 1과 2의 검정은 2021 수능 영어와 영어 I-II 교과서가 어휘 정교성, 즉 그룹 1과 2의 단어 사용에 있어서 동일한지를 따지는 것이다. 귀무가설 1을 검정하기 위해, <표2>를 토대로 교차분석을 한 결과는 아래 <표4>와 같다.

<표4> 구어 자료에 사용된 등급화된 단어 출현형에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	47.926 ^a	9	.000
유효 케이스 수	61,359		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 375.83임.

유의 확률이 .000에 가까우므로 유의 수준 .05보다 작다. 따라서 2021 수능 영어와 영어 I-II 교과서의 구어 자료가 그룹 1과 2의 단어 사용에 있어서 동일하다는 귀무가설 1을 기각할 수밖에 없다.

귀무가설 1의 기각에 결정적으로 이바지한 쌍을 조사하고자 Bonferroni 방법에

따라 사후분석을 하였다. 그 결과는 <그림1>과 같다.

<그림1> 귀무가설 1의 기각에 결정적인 차이를 보인 구어 자료 쌍

	2021수	교A	교B	교C	교D	교E	교F	교G	교H	교I
2021수										
교A										
교B										
교C										
교D										
교E										
교F										
교G										
교H										
교I										

<그림1>에서 음영 처리된 자료 쌍은 카이제곱 검정 유의 확률이 .001 이하이다. 교과서 쌍 A-D, B-D, C-D, C-G, H-D와 I-D는 .000이고 교과서 쌍 C-E와 F-D는 .001이다. 이들 유의 확률 값에 Bonferroni 보정법의 다중 검정 회수 45를 곱해도 .045 이하로 유의 수준 .05보다 작다. 흥미롭게도, 교과서 H와 I는 각각의 유의 확률 .945와 .946에서 보듯이 2021 수능 영어와 그룹 1과 2의 단어 사용에 있어서 거의 동일하다.

수능 영어와 영어 I-II 교과서의 문어 자료가 그룹 1과 2의 단어 사용에 있어서 동일하다는 귀무가설 2를 검정하기 위해, <표3>를 토대로 교차분석을 한 결과는 <표5>와 같다.

<표5> 문어 자료에 사용된 등급화된 단어 출현형에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	167,824 ^a	9	.000
유효 케이스 수	118,090		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 378.56임.

유의 확률이 .000에 가까우므로 유의 수준 .05보다 작다. 따라서 2021 수능

영어와 영어 I-II 교과서의 문어 자료가 어휘 정교성에 있어서 동일하다는 귀무가설 2를 기각할 수밖에 없다.

귀무가설 2의 기각에 결정적으로 이바지한 쌍을 조사하고자 Bonferroni 방법에 따라 사후분석을 하였다. 그 결과는 <그림2>와 같다.

<그림2> 귀무가설 2의 기각에 결정적인 차이를 보인 문어 자료 쌍

	2021수	교A	교B	교C	교D	교E	교F	교G	교H	교I
2021수										
교A										
교B										
교C										
교D										
교E										
교F										
교G										
교H										
교I										

<그림1>에서 음영 처리된 자료 쌍은 카이제곱 검정 유의 확률이 .001 이하이다. 교과서 A-2021 수능 영어, 교과서 C-2021 수능 영어와 교과서 H-2021 수능 영어 그리고 교과서 쌍 A-B, A-D, A-E, A-F, A-G, A-I, C-B, H-B, C-D, C-F, C-G, C-I, D-G, H-D, E-G, H-E, H-F, H-G, H-I는 .000이고 교과서 쌍 F-G는 .001이다. 이들 유의 확률 값에 Bonferroni 보정법의 다중 검정 회수 45를 곱해도 .045 이하로 유의 수준 .05보다 작다. 흥미롭게도, 교과서 F와 I는 유의 확률 1.000에서 보듯이 그룹 1과 2의 단어 사용에 있어서 거의 동일하다.

4.2. 귀무가설 3과 4의 검정

2021 수능 영어와 영어 I-II 교과서가 어휘 밀도에 있어서 동일한지를 따지기 위해서는 이들 자료에 내용어와 기능어의 출현형이 얼마나 많이 사용되었

는지를 알아야 한다. 이들 자료에 사용된 내용어와 기능어의 출현형 빈도는 아래 <표6>과 같다.

<표6> 분석 대상 자료에 사용된 내용어와 기능어의 출현형 빈도

구분	분석 대상	내용어	기능어
구어	2021 수능 영어	1,037(44.7%)	1,285(55.3%)
	교과서 A	5,327(42.5%)	7,194(57.5%)
	교과서 B	3,186(43.2%)	4,182(56.8%)
	교과서 C	2,389(43.7%)	3,075(56.3%)
	교과서 D	3,235(45.3%)	3,906(54.7%)
	교과서 E	2,245(44.1%)	2,843(55.9%)
	교과서 F	2,182(45.0%)	2,664(55.0%)
	교과서 G	3,696(45.1%)	4,493(54.9%)
	교과서 H	2,270(45.4%)	2,729(54.6%)
	교과서 I	2,793(45.4%)	3,357(54.6%)
	구어 전체	28,360(44.3%)	35,728(55.7%)
문어	2021 수능 영어	2,115(50.5%)	2,074(49.5%)
	교과서 A	7,906(48.0%)	8,579(52.0%)
	교과서 B	5,571(49.9%)	5,591(50.1%)
	교과서 C	5,424(48.3%)	5,800(51.7%)
	교과서 D	7,272(49.7%)	7,375(50.3%)
	교과서 E	5,746(48.8%)	6,027(51.2%)
	교과서 F	8,562(51.5%)	8,060(48.5%)
	교과서 G	7,492(51.1%)	7,170(48.9%)
	교과서 H	6,561(47.3%)	7,303(52.7%)
	교과서 I	4,515(50.0%)	4,503(50.0%)
	문어 전체	61,164(49.5%)	62,482(50.5%)

2021 수능 영어 구어 자료에서는 1,037개의 내용어와 1,285개의 기능어가 사용되어 어휘 밀도가 .447이다. 영어 I-II 교과서 구어 자료에서는 평균적으로 3,036개의 내용어와 3,827개의 기능어가 사용되어 평균 어휘 밀도가 .442이다. 전체 구어 자료의 어휘 밀도가 .443이지만 특정 구어 자료 간에 상당한 어휘 밀도의 편차가 존재한다. 문어 자료도 구어 자료와 비슷한 양상을 보인다. 2021 수능 영어 문어 자료에서는 2,115개의 내용어와 2,074개의 기능어가 사

용되어 어휘 밀도가 .505이다. 영어 I-II 교과서 문어 자료에서는 평균적으로 59,049개의 내용어와 60,408개의 기능어가 사용되어 평균 어휘 밀도가 .494이다. 전체 구어 자료의 어휘 밀도가 .445이지만 특정 문어 자료 간에 상당한 어휘 밀도의 편차가 존재한다.

귀무가설 3을 검정하기 위해, <표6>의 구어 관련 자료를 토대로 교차분석을 한 결과는 <표7>과 같다.

<표7> 구어 자료의 내용어와 기능어 출현형에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	31.694 ^a	9	.000
유효 케이스 수	64,088		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 1,027.52임.

유의 확률이 .000에 가까우므로 유의 수준 .05보다 작다. 따라서 2021 수능 영어와 영어 I-II 교과서의 구어 자료가 어휘 밀도, 즉 내용어와 기능어의 출현형 빈도에 있어서 동일하다는 귀무가설 3을 기각할 수밖에 없다.

귀무가설 3의 기각에 결정적으로 이바지한 자료 쌍을 찾고자 Bonferroni 방법에 따라 사후분석을 하였다. 카이제곱 검정 유의 확률이 .000인 자료 쌍은 교과서 A-D, A-G와 A-I이고 유의 확률이 .001인 쌍은 교과서 A-H이다. 이들 4개 쌍이 귀무가설 3의 기각에 결정적인 영향을 미쳤다. Fisher의 정확검정 유의 확률이 .913부터 1.000이라는 사실에서 보듯이 교과서 D, H, I는 어휘 밀도에 관해서 거의 동일하다.

귀무가설 4를 검정하기 위해, <표6>의 문어 관련 자료를 토대로 교차분석을 한 결과는 <표8>과 같다.

<표8> 구어 자료의 내용어와 기능어 출현형에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	95.861 ^a	9	.000

유효 케이스 수	123,646		
----------	---------	--	--

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 2,072.17임.

유의 확률이 .000에 가까우므로 유의 수준 .05보다 작다. 따라서 2021 수능 영어와 영어 I-II 교과서의 문어 자료가 어휘 밀도에 있어서 동일하다는 귀무 가설 4를 기각할 수밖에 없다.

귀무가설 4의 기각에 결정적으로 이바지한 자료 쌍을 찾고자 Bonferroni 방법에 따라 사후분석을 하였다. 카이제곱 검정 유의 확률이 .000인 자료 쌍은 2021 수능 영어-교과서 H와 교과서 쌍 A-F, A-G, C-F, C-G, E-F, E-G, H-B, H-D, H-F, H-G, H-I이다.

4.3. 귀무가설 5와 6의 검정

2장에서 설명했듯이, 어휘 다양성을 측정하는 전통적인 척도는 TTR과 이의 문제점을 보완한 GI 그리고 STTR이 있다. 이들 척도에 따라, <표1>을 토대로 분석 대상 자료의 어휘 다양성을 계량하면 <표9>와 같다.

<표9> 분석 대상 자료의 어휘 다양성 계량치

구분	분석 대상	TTR	GI	STTR/1000
구 어	2021 수능 시험	.319	15,378	.417
	교과서 A	.166	18,544	.402
	교과서 B	.211	18,151	.410
	교과서 C	.230	17,019	.421
	교과서 D	.242	20,425	.450
	교과서 E	.252	17,945	.435
	교과서 F	.243	16,908	.404
	교과서 G	.210	18,974	.417
	교과서 H	.244	17,283	.422
	교과서 I	.224	17,533	.421
	구어 전체	.086	21,738	.418
	2021 수능 시험	.351	22,712	.506
	교과서 A	.188	24,199	.470

	교과서 B	.237	25,064	.480
문	교과서 C	.218	23,135	.470
	교과서 D	.221	26,771	.490
	교과서 E	.219	23,806	.486
	교과서 F	.210	27,062	.476
어	교과서 G	.212	25,701	.471
	교과서 H	.197	23,160	.467
	교과서 I	.249	23,683	.480
	문어 전체	.082	28,891	.477

TTR이 텍스트의 길이에 큰 영향을 받으므로 다른 자료에 비해 길이가 매우 짧은 2021 수능 영어 구어 자료(.319)와 매우 긴 교과서 A(.166)가 TTR상 어휘 다양성이 가장 높고 가장 낮게 나오는 것은 당연하다. 하지만 텍스트 길이의 영향을 보정한 GI와 1,000 단어 출현형 대비 단어 유형의 비율의 평균인 STTR을 보면 상황이 달라진다. GI 값으로는 교과서 D가 20.425로 어휘 다양성이 가장 높은 것으로 나오고, 수능 영어는 15.378로 가장 낮은 것으로 나온다. 또한 어휘 다양성의 이 세 척도 중 가장 믿을 수 있는 STTR을 보면, 수능 영어는 .417로 중간에서 조금 못 미치는 수준으로 나온다. 이에 반해, 교과서 A는 .402로 STTR 값이 가장 낮다. GI 값이 가장 큰 교과서 D가 STTR도 가장 크고, 두 번째로 작은 GI 값을 가진 교과서 F가 STTR에서도 두 번째로 값이 작다. 이런 차이는 교과서 A, D, F가 어휘 다양성에 있어서 유의한 차를 보일 가능성이 높다.

구어 자료와 마찬가지로, 2021 수능 영어 문어 자료(.319)가 다른 자료에 비해 길이가 매우 짧지만 교과서 A는 매우 길다. 따라서 TTR상으로 전자(.351)가 어휘 다양성이 가장 높게 나오고 후자(.188)가 가장 낮게 나온다. 전자는 보정이 제대로 안 되어 GI상으로는 어휘 다양성이 가장 낮게 나오지만(22.712), STTR로는 값이 .506으로 가장 높다. 2021 수능 영어 문어 자료를 제외하면, 가장 작은 GI 값을 가진 교과서 C가 STTR 값에서 두 번째로 작은 값을 가진다. 가장 작은 STTR 값(.467)을 가진 교과서 H는 GI 값이 세 번째로 작다. 교과서 C와 H는 2021 수능 영어 문어 자료뿐만 아니라 STTR 값이 상대적으로 큰 교과

서 D와 E와도 어휘 다양성에 있어서 유의한 차를 보일 가능성이 크다.

STTR나 GI의 값의 차가 아무리 크다고 하더라도 이 차이가 통계적 유의차 인지는 말할 수 없다. 따라서 2021 수능 영어 구어 및 문어 자료와 같은 코퍼스를 인위적으로 만들어 카이제곱 검정을 통해 위에서 논한 어휘 다양성의 계량치 차가 유의한 것인지를 확인해 보고자 한다.

먼저 *영어 I* 교과서의 앞부분과 뒷부분 그리고 *영어 II* 교과서의 앞부분과 뒷부분에서 거의 균등하게 텍스트를 추출하여 구어 교과서 자료는 단어 출현형이 2,322개가 되도록 하고 문어 교과서 자료는 4,189개가 되도록 한다. 이들 자료에서 한 번만 사용되는 고어(hapax legamenon)를 포함하여 첫 번째로 사용된 단어의 수와 그 외 반복적으로 사용된 단어의 수를 조사하면 <표10>과 같다.

<표10> 크기가 동일한 자료에서의 단어 출현형 양상

구분	분석 대상	첫 번째로 사용된 단어	그 외의 단어
구 어	2021 수능 영어	741(31.9%)	1,581(68.1%)
	교과서 A	660(28.4%)	1,662(71.6%)
	교과서 B	715(30.8%)	1,607(69.2%)
	교과서 C	709(30.5%)	1,613(69.5%)
	교과서 D	737(31.7%)	1,585(68.3%)
	교과서 E	791(34.1%)	1,531(65.9%)
	교과서 F	729(31.4%)	1,593(68.6%)
	교과서 G	748(32.2%)	1,574(67.8%)
	교과서 H	739(31.8%)	1,583(68.2%)
	교과서 I	694(29.9%)	1,628(70.1%)
	소계	7,263(31.3%)	15,957(68.7%)
문	2021 수능 영어	1,470(35.1%)	2,719(64.9%)
	교과서 A	3,107(18.8%)	13,378(81.2%)
	교과서 B	2,648(23.7%)	8,514(76.3%)
	교과서 C	2,451(21.8%)	8,773(78.2%)
	교과서 D	3,240(22.1%)	11,407(77.9%)
	교과서 E	2,583(21.9%)	9,190(78.1%)
	교과서 F	3,489(21.0%)	13,133(79.0%)

어	교과서 G	3,112(21.2%)	11,550(78.8%)
	교과서 H	2,727(19.7%)	11,137(80.3%)
	교과서 I	2,249(24.9%)	6,769(75.1%)
	소계	27,076(21.9%)	96,570(78.1%)

〈표10〉에서 텍스트에서 처음으로 사용된 단어 수는 단어 유형의 수와 같으며 그 외 반복적으로 사용된 단어 수는 단어 출현형 총수에서 단어 유형의 수를 뺀 수와 같다.

귀무가설 5를 간접적으로 검증하기 위해 〈표10〉의 구어 자료를 토대로 교차분석을 한 결과는 각각 〈표11〉과 같다.

〈표11〉 구어 자료의 어휘 다양성에 관한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	22.083 ^a	9	.009
유효 케이스 수	23,220		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 726.30임.

유의 확률이 .009이므로 유의 수준 .05보다 작다. 따라서 2021 수능 영어와 영어 I-II 교과서의 구어 자료가 일정 부분의 어휘 다양성에 있어서 동일하다는 귀무가설 5를 기각할 수밖에 없다.

사후분석의 일환으로 수정된 잔차를 보니, 교과서 A와 E의 수정된 잔차가 둘 다 1 ± 3.11 이다. 〈표10〉에서 보듯이, TTR 값이 A는 .284이고 E는 .341이다. 이 둘의 유의차가 귀무가설 5의 기각으로 귀결되었다고 볼 수 있다.

귀무가설 6을 간접적으로 검증하기 위해 〈표10〉의 문어 자료를 토대로 교차분석을 한 결과는 각각 〈표12〉와 같다.

〈표12〉 문어 자료의 어휘 다양성에 관한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	639.184 ^a	9	.000

유효 케이스 수	123,646		
----------	---------	--	--

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 2,072.17임.

유의 확률이 .000에 가까우므로 유의 수준 .05보다 작다. 따라서 2021 수능 영어와 *영어 I-II* 교과서의 문어 자료가 일정 부분의 어휘 다양성에 있어서 동일하다는 귀무가설 6을 기각할 수밖에 없다.

수정된 잔차가 1 ± 2.01 이상인 자료는 2021 수능 영어(1 ± 6.01), 교과서 F(1 ± 4.11), A(1 ± 3.91), D(1 ± 2.21) 그리고 H(1 ± 3.31)이다. <표10>에서 보듯이, 앞의 두 자료는 TTR 값이 크고 뒤의 세 자료는 작다. 이러한 유의차가 귀무가설 6의 기각으로 귀결되었다고 볼 수 있다.

4.4. 논의

이제까지의 귀무가설 검정은 구어 자료이든지 문어 자료이든지 간에 2021 수능 영어와 *영어 I-II* 교과서가 어휘 정교성과 어휘 밀도 그리고 어휘 다양성에 있어서 유의차가 있다는 점을 보여주었다.

등급화된 그룹 1과 2의 단어의 사용과 관련하여, 어떤 *영어 I-II* 교과서 구어 자료도 수능 영어 구어 자료와 유의차가 없었지만, 몇몇 교과서 간에는 유의차가 있었다. 유의차가 있는 교과서 쌍인 A-D, B-D, C-D, F-D, H-D, I-D, C-E, C-G에서 그룹 1과 2의 단어가 어느 정도의 비율로 사용되었는지를 수정된 잔차와 비교하여 살펴보면, 교과서 D에서 저빈도 고급 단어가 매우 높은 비율로 사용되었지만, 교과서 C에서는 반대로 고빈도 기본 단어가 매우 높은 비율로 사용되었음을 알 수 있다. 교과서 G도 저빈도 고급 단어의 비율이 유의하게 높다고 할 수 있다.

어휘 정교성에 관한 자료 간의 유의차는 문어 자료에서 더욱 심하게 나타난다. 먼저, 교과서 A, C, H는 수능 영어에 비해 그룹 2의 단어 사용 비율이 유의하게 낮다. 유의차가 있는 교과서 쌍인 A-B, A-D, A-E, A-F, A-G, A-I, C-B,

C-D, C-F, C-G, C-I, H-B, H-D, H-E, H-F, H-G, H-I, D-G, E-G, F-G와 관련된 수정된 잔차를 살펴보면, 교과서 A, C, H에서 고빈도 기본 단어인 그룹 1 단어가 유의하게 높은 비율로 사용되었지만, 교과서 G에서 저빈도 고급 단어가 유의하게 높은 비율로 사용되었다는 것을 알 수 있다. 교과서 B와 F도 저빈도 고급 단어의 비율이 유의하게 높다. 이런 결과는 8종의 영어 I-II 교과서와 2개 년도의 수능 영어를 비교한 이현우(2019)의 연구 결과와 부합한다. 이 연구에서도 교과서와 수능 영어 간에 그리고 교과서 간에 어휘 정교성의 유의차가 상당히 존재한다는 것이 밝혀졌다.

구어 자료의 어휘 밀도에 관한 귀무가설 3의 기각에 따른 사후분석을 통해 교과서 쌍 A-D, A-G, A-H와 A-I가 유의하게 다르다는 것이 밝혀졌다. 관련된 수정된 잔차와 비교하면, 교과서 A가 다른 구어 자료에 비해 어휘 밀도가 유의하게 낮음을 알 수 있다. 교과서 D, H, I는 어휘 밀도가 다른 자료보다 유의하게 높지는 않지만 상당히 높다고 할 수 있다. 어휘 밀도에 관한 자료 간의 유의차가 문어 자료에서 좀 더 심하다. 유의차가 있는 교과서 쌍 A-F, A-G, C-F, C-G, H-B, H-D, H-F, H-G, H-I, E-F, E-G와 관련된 수정된 잔차와 비교하면, 교과서 A, C, H는 어휘 밀도가 다른 자료에 비해 유의하게 낮고 교과서 F와 G는 유의하게 높다. 특히, 교과서 H는 유일하게 2021 수능 영어와 유의차가 있는 것으로 나왔다.

앞서 설명했듯이, 자료 간에 어휘 다양성의 유의차가 있는지를 입증하는 것은 매우 어렵다. 비록 STTR이 어휘 다양성을 측정하는 믿을 만한 척도라고 하더라도 STTR 값만을 보고 두 자료 간에 유의차가 있는지 없는지를 결정할 수 없기 때문이다. 본 연구에서는 이런 점을 보완하기 위해 교과서의 자료를 수능 영어의 자료와 길이를 같게 한 후 카이제곱 검정을 하여 어휘 다양성의 유의차를 간접적으로 조사하였다. 이 결과에 TTR, GI 그리고 STTR 값을 더해 종합적으로 고려할 때, 구어 자료 중에서 교과서 A의 어휘 다양성이 유의하게 낮다고 확신할 수 있다. 교과서 F는 유의차가 있을 정도는 아니지만 어휘 다양성이 상대적으로 낮고, 교과서 D와 E는 이와 반대로 유의차가 있을 정도는 아니지만

상대적으로 높다고 할 수 있다. 같은 논리로, 문어 자료 중에서 2021 수능 영어의 어휘 다양성이 유의하게 높고 교과서 H의 어휘 다양성이 유의하게 낮다고 할 수 있다. 교과서 D, E, F는 어휘 다양성이 상대적으로 높은 편이고, 교과서 A와 C는 상대적으로 낮은 편이라고 할 수 있다.

5. 결론

2015 개정 영어과 교육과정에 기초한 수능 영어가 2020년에 처음으로 시행되었다. 이에 따라 이 시험의 출제 범위인 *영어 I-II* 교과서와 이 시험이 어휘 풍부성의 주요 세 측면인 어휘 정교성, 어휘 밀도와 어휘 다양성에 있어서 유의차가 있는지의 문제를 따져보았다. 어휘의 사용이 구어와 문어에 따라 달라지므로 자료의 비교는 구어와 문어로 구분하여서 하였다. 구어와 문어의 구분에 따라 자료를 비교하다 보니 여섯 개의 귀무가설이 설정되었고 이의 검정으로 연구가 수행되었다.

검정 결과, 모든 귀무가설이 기각되었다. 구어 자료이든지 문어 자료이든지 간에 자료 간에 어휘 정교성의 유의차가 있었다. 즉, 등급화된 단어 출현형의 빈도에 관한 교차분석 결과, 등급화된 단어 출현형의 빈도가 자료에 따라 다르다는 것이 밝혀졌다. 자료 간에는 어휘 밀도의 유의차도 존재했다. 내용어와 기능어의 출현형 빈도에 관한 교차분석 결과, 내용어의 빈도가 자료에 따라 다르게 나왔다. 통계적 유의차를 확정적으로 말하기가 어렵지만, 어휘 다양성도 거의 같은 결과를 보여주었다. 자료들의 GI와 STTR 값의 비교와 인위적으로 길이를 같게 한 자료를 대상으로 한 교차분석의 결과 몇몇 자료 간에 유의미한 어휘 다양성의 차가 존재함을 확인하였다.³

사후분석 결과, 등급화된 단어의 빈도에 있어서 유의차가 있는 교과서 구

3 인위적으로 길이를 같게 한 다른 자료로 카이제곱 검정을 하여도 같은 결과가 나올 가능성이 크므로 이처럼 결론을 내리는 데 문제가 없어 보인다.

어 자료 쌍과 교과서 문어 자료 쌍이 상당수 있었으며, 문어 자료의 경우에는 2021 수능 영어와 유의차가 있는 교과서들도 있었다. 내용어의 빈도에 관한 양상도 등급화된 단어의 빈도 양상과 비슷하다. 유의차가 있는 쌍은 한 교과서 문어 자료와 수능 영어 쌍을 빼고는 모두 교과서 자료 쌍들이었다. 구어 자료보다 문어 자료에서 어휘 밀도의 편차가 더 심하게 나타났다. 어휘 다양성 측정의 내재적인 문제점 때문에 어휘 다양성의 유의차가 있는 자료 쌍을 특정하기 쉽지 않지만, 몇몇 구어 및 문어 자료는 다른 자료에 비해 어휘 다양성이 유의하게 높거나 낮다고 할 수 있다.

이상의 결과를 종합하면, 어떤 교과서의 구어 또는 문어 자료는 어휘 정교성, 어휘 밀도, 또는 어휘 다양성에 있어서 2021 수능 영어보다 유의하게 높거나 낮았다. 이런 유의차가 한두 면에 국한되는 것이라면 문제가 없겠으나 여러 면에서 일관되게 유의하게 높거나 유의하게 낮으면 문제가 된다. 이런 경우에는 이들 영어 I-II 교과서가 수능 영어의 바탕 역할을 할 수 없기 때문이다. 학생들이 어떤 교과서를 갖고 공부하는지에 따라서 학생들의 수능 영어 대비가 효율적일 수 있고 비효율적일 수 있다면 이는 결코 소홀히 할 수 없는 문제라고 생각된다. 향후 영어 I-II 교과서를 제작하거나 수능 영어를 출제할 때 이들 교과서와 수능 영어의 어휘 연계성에 관한 보다 체계적이고 세밀한 연구가 선행되어야 할 것으로 여겨진다. 예를 들어, 교육과정에 수능 영어에서 사용할 수 있는 어휘 수를 2,800 낱말로 한정하고 이들 어휘의 80% 이상을 기본 어휘 목록에서 사용하도록 정하면 영어 I-II 교과서와 수능 영어는 어휘의 연계성이 한층 높아질 것이다.

[ABSTRACT]

Vocabulary Analysis of the English Section of the 2021 CSAT and the
High School English I and II Textbooks

Lee, HyunOo(Inha University),
Ryoo, HyeKyung(Daegu University)

This paper is concerned with the question of whether the English Section of the College Scholastic Ability Test (CSAT) for the 2021 academic year and the *High School English I and II* textbooks are homogeneous with respect to lexical sophistication, lexical density and lexical diversity. To seek an answer, six null hypotheses were formulated and tested. Chi-square tests on graded word tokens and content vs. function words show that the *High School English I-II* textbooks differ significantly not only from each other but also from the CAST English section, whether they were spoken or written materials. The same result obtained in regard to lexical diversity. The STTR values of the materials, combined with the Chi-square tests on the abridged materials of equal size, show that some of them were significantly higher or lower in lexical density.

Key words: 2015 Revised National Curriculum of English, English Section of CSAT, Lexical sophistication, Lexical diversity, Lexical density

[참고문헌]

□ 기본자료

권혁승, 박은성, 김현정, 오건석, 강신자, 윤소영, High School English I, 서울: (주) 동아출판, 2018.

권혁승, 박은성, 김현정, 오건석, 강신자, 윤소영, High School English II, 서울: (주) 동아출판, 2019.

김길중, Putlack Michael Aaron, 임정원, 장진화, 김건우, 김나현 외, High School English I, 서울: (주) 다락원, 2018.

김길중, Putlack Michael Aaron, 임정원, 장진화, 김건우, 김나현 외, High School English II, 서울: (주) 다락원, 2019.

김성곤, 윤진호, 전형주, 서정환, 이후고, 김윤자 외, High School English I, 서울: (주) 능률교육, 2019.

김성곤, 윤진호, 구은영, 전형주, 서정환, 이후고 외, High School English II, 서울: (주) 엔이능률, 2019.

민찬규, 정현성, 이상기, 김윤규, 나우철, 안효선 외, High School English I, 서울: (주) 지학사, 2018.

민찬규, 정현성, 이상기, 김윤규, 나우철, 안효선 외, High School English II, 서울: (주) 지학사, 2019.

박준언, 윤병우, 김선형, 최선영, 최지영, 김진수 외, High School English I, 서울: (주) 와이비엠, 2018.

박준언, 윤병우, 김선형, 최선영, 최지영, 김진수 외, High School English II, 서울: (주) 와이비엠, 2019.

이재영, 문안나, 오영일, 조수경, 이윤정, 김정현 외, High School English I, 서울: (주) 천재교육, 2018.

이재영, 문안나, 오영일, 조수경, 이윤정, 김정현 외, High School English II, 서울: (주) 천재교육, 2019.

- 최인철, 박리리, 김근영, 최수하, 김주혜, 손지혜 외, High School English I, 서울: (주) 금성출판사, 2018.
- 최인철, 서원화, 이윤경, 김주혜, 전예지, 라문선 외, High School English II, 서울: (주) 금성출판사, 2019.
- 한상호, 정은귀, 김예리, 김재란, 이보희, 이혜은, High School English I, 서울: (주) 와이비엠텍스트, 2018.
- 한상호, 정은귀, 김예리, 김재란, 이보희, 이혜은, High School English II, 서울: (주) 와이비엠텍스트, 2019.
- 홍민표, 안현기, 박연미, 김정태, 장현욱, 신정섭 외, High School English I, 서울: (주) 비상교육, 2018.
- 홍민표, 안현기, Pak Richard, O'Flaherty David Desmond, 조금희, High School English II, 서울: (주) 비상교육, 2019.

□ 단행본 및 논문

- 교육부, 『영어과 교육과정. 교육부 고시 제 2015-74호 [별책 14]』, 교육부, 2015.
- 김제우 · 이동주, 「코퍼스기반 고등학교 어휘 사용 연구」, 『언어과학연구』 80, 2017, 57-74면.
- 노윤오, 「2009 개정교육과정 고등학교 영어 교과서 어휘 사용 양상에 관한 연구」, 『인하대학교 박사학위 논문』, 2019.
- 오준일, 「고등학교 영어II와 EBS 수능 연계 교재의 지문에 사용된 어휘 수준에 관한 연구」, 『영어교과교육』 13, 2014, 115-32면.
- 이소영 · 이현우, 「중3 영어 교과서에 나타나는 영어 어휘의 분포」, 『중등영어교육』 10(4), 2017, 121-42면.
- 이현우, 「중3 영어 교과서와 학업성취도평가 시험에 나타나는 부정사의 분포」, 『국제언어문학』 36, 2017, 229-60면.
- 이현우, 「영어 I-II 교과서와 대학수학능력시험 영어 영역 읽기 파트에 나타난 등급화된 단어의 분포」, 『중등영어교육』 12(2), 2019, 79-99면.

정채관 · 권혁승, 「2015 개정 영어과 교육과정에 따른 기본 어휘 및 외래어 목록 변
화형 연구: 코퍼스적 접근」, 『외국어교육연구』 31, 2017, 175-201면.

한국교육과정평가원, 「2021학년도 대학수학능력시험 시행기본계획」, 한국교육과
정평가원, 서울, 2020.

Adolphs, S. & N. Schmitt, Lexical coverage of spoken discourse, *Applied
Linguistics* 24(4), 2003, pp.425-438.

Brezina, V. & D. Gablasova, Is there a core general vocabulary? Introducing the
New General Service List, *Applied Linguistics* 36(1), 2015, pp.1-22.

Daller, H., Milton, J., & Treffers-Daller, J., (eds.), Modelling and Assessing
Vocabulary Knowledge, Cambridge University Press, Cambridge, 2007.

Daller, H., van Hout, R. & J. Treffers-Daller, Lexical richness in the spontaneous
speech of bilinguals, *Applied Linguistics* 24(2), 2003, pp.197-222.

Daller, M, Guiraud's index of lexical richness, PP presentation, University of West
England, Bristol, 2010. Eprint available at <http://eprints.uwe.ac.uk/11902/>.

Guiraud, P., Les Caractères Statistiques du Vocabulaire: Essai de Méthodologie,
Presses Universitaires de France, Paris, 1954.

Heatley, A., Nation, I. S. P., & Coxhead, A, RANGE and FREQUENCY programs
(computer software), Victoria University of Wellington, Wellington, New
Zealand, 2002.

Malvern, D. D., Richards, B. J., Chipere, N, & Durán, P., Lexical Diversity
and Language Development: Quantification and Assessment, Palgrave
Macmillan, Houndmills etc., 2004.

Meara, P. & H. Bell, P-Lex: a simple and effective way of describing the lexical
characteristics of short L2 texts, *Prospect* 16, 2001, pp.5-19.

Nation, I. S. P., How large a vocabulary is needed for reading and listening, *The
Canadian Modern Language Review* 63(1), 2006, pp.59-82.

Nation, I. S. P., Learning Vocabulary in Another Language, Cambridge University

- Press, Cambridge, 2001.
- Nation, I. S. P., The BNC/COCA headword lists, Retrieved September 15, 2018 from <https://www.victoria.ac.nz/lals/about/staff/paul-na-tion#vocab-lists>, 2012.
- Read, J., *Assessing Vocabulary*, Cambridge University Press, Cambridge, 2000.
- Scott, M., *WordSmith Tools version 7 (lexical analysis software)*, Stroud, 2016.
- Templin, M. C., *Certain Language Skills in Children: Their Development and Interrelationships*, University of Minnesota Press, Minneapolis, 1957.
- Ure, J., Lexical density and register differentiation, In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics: Selected papers of the Second International Congress of Applied Linguistics*, Cambridge University Press, 1971, pp.443-52.
- West, M., *A General Service List of English Words*, Longman, Green and Co, London, 1953.

접수일: 2021. 04. 01 총평일: 2021. 04. 28 게재확정: 2021. 04. 29