

# 2018학년도 대학수학능력시험과 모의평가 영어 영역 어휘 분석\*

이현우 \*\*

## || 차례 ||

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

### [국문초록]

본 논문은 문항 난이도에 있어서 상당한 차이를 보인 2018학년도 대학수학능력시험과 같은 해 6월과 9월의 모의평가 영어 영역이 어휘 정교성과 밀집성에 있어서 유의미한 차이가 있는지를 카이제곱 검정을 통해 밝히고 세 시험의 어휘 다양성을 비교하고자 한다. 카이제곱 검정 결과, 등급화된 고급 단어 토권의 분포와 내용어 토권의 비율에 있어서 세 시험이 통계적으로 유의미한 차이를 보여주고 있다. 사후 분석 결과, 이 차이의 원인은 9월 모의평가에서의 고빈도 단어의 과다 사용과 저빈도 단어의 과소 사용에 있었다. 하지만 이 결과는 9월 모평이 두 시험보다 더 어려웠다는 사실과 모순되며, 따라서 세 시험의 문항 난이도가 어휘 정교성과 관련이 없었음을 시사한다. 비록 어휘 밀집성에 있어서 세 시험이 유의미한 차이를 보였지만, 이들 시험의 총 토권 수가 달라, 이를 문항 난이도의 차이를 야기한 요인으로 보기 어렵다. 어휘 다양성 측정치의 미비한 차이 또한 문항 난이도 설명에 큰 도움이 되지 못한다.

**주제어:** 대학수학능력시험(모의평가) 영어 영역, 문항 난이도, 어휘 정교성, 어휘 밀집성, 어휘 다양성, 단어 타입-토권 비율, 카이제곱 검정

\* 이 논문은 인하대학교의 지원에 의하여 연구되었음.

\*\* 인하대학교 영어교육과 교수: hylee@inha.ac.kr

196 국제언어문학 제41호(2018.12.)

## 1. 서론

2002년 9월에 처음으로 대학수학능력시험 모의평가가 시행된 후, 2004학년도부터 매년 11월의 대학수학능력시험(본수능)에 앞서 6월 및 9월에 전국적인 규모의 모의평가(모평)가 실시되고 있다. 대학수학능력시험 20년사(273면)는 수능 모평의 목적을 ‘수능과 같은 출제 방식에 따라 행하여지는 모의평가를 통하여 수험생에게 그 해에 치를 예정인 수능의 경향과 수준에 대비할 수 있는 기회를 주고, 또한 출제진에게는 그 해 응시자의 수준을 이해하게 하여 수능에서의 적당한 난이도 조절을 할 수 있는 기회를 제공’하는 것으로 밝히고 있다. 이러한 목적으로 시행되는 수능 모평은 타당도와 신뢰도를 비롯한 여러 면에서 본수능과 동질적인 평가로 여겨지고 있으며, 대학수학능력시험 20년사에서 언급한 바와 같이, 영어 영역의 문제가 정형화된 2005학년도 이후부터는 본수능과 수능 모평 영어 영역은 더욱 동질적인 시험으로 간주되어 왔다.

그럼에도 불구하고, 절대 평가로 치러진 2018학년도 본수능과 모평 영어 영역의 채점 결과에 따르면,<sup>1)</sup> 521,938명이 응시한 6월 모평은 90점 이상의 1등급이 42,183명으로 8.08%, 60점 이상의 4등급 이상이 311,197명으로 59.62%이고, 513,498명이 응시한 9월 모평은 1등급이 27,695명으로 5.39%, 4등급 이상이 278,228명으로 54.17%이고, 528,064명이 응시한 본수능은 1등급이 52,983명으로 10.03%, 4등급 이상이 385,885명으로 73.08%이다. 이러한 결과로 볼 때, 9월 모평 시험이 다른 두 시험에 비해 눈에 띄게 어려운 시험이라고 할 수 있다.

본수능 및 모평 영어 영역과 같은 다지선다형 표준화 시험의 문항 난이도에 영향을 주는 요인에 대한 연구가 지속적으로 이뤄져 왔다. 특히 시험의 난

1) 더 상세한 채점 결과는 2018학년도 대학수학능력시험 6월 모의평가 채점 결과 보도자료, 2018학년도 대학수학능력시험 9월 모의평가 채점 결과 보도자료, 그리고 2018학년도 대학수학능력시험 채점 결과 보도자료를 참고 바람.

이도에 영향을 요인으로 텍스트의 추상성(Paivio, 1986), 주제에 대한 익숙함 정도(Freedle과 Kostin, 1993), 단락의 길이(Yano, Long과 Ross, 1994), 문장 복잡도(Klare, 1974-1975; Leow, 1993), 정보처리 단위(Freedle과 Kostin, 1993), 지시 대상의 빈도 (Abrahamsen과 Shelton, 1989; Freedle과 Kostin, 1993), 어휘 수준(Klare, 1984; Graves, 1986; Freedle과 Kostin, 1993; Ozuru, Rowe, O'Reilly와 McNamara, 2008), 단서의 위치 (Kieras, 1985; Hare, Rabinowitz와 Schieble, 1989; Freedle과 Kostin, 1993), 오답지의 매력도(Drum, Calfee와 Cook, 1981; 장경숙, 2004) 등이 연구되어 왔다.

Hudson(2007)이 언급한 바와 같이, 어휘는 현대 심리학 및 교육학 연구의 역사를 통틀어 언어 이해에 있어 가장 중요한 요인으로 여겨져 왔다. Drum, Calfee와 Cook(1981)은 어휘 변수가 텍스트의 통사적 특성과 관련된 변수보다도 더 영향이 크다는 결과를 내놓았다. 특히, Qian(1999, 2002)과 Read(2000)는 통상 독자의 어휘 지식이 독해의 측정치와 상관관계가 높게 나와 흔히 글의 이해에 대한 유일한 예측 요인으로 여겨진다는 것을 보여 주었다. 또한 Ozuru, Rowe, O'Reilly와 McNamara(2008)는 7-9학년생의 Gates-MacGinitie 독해 시험에서 텍스트의 특성, 특히 어휘 난이도가 문항 난이도에 가장 중요한 영향을 미쳤다고 보고하고 있다.

문항 난이도와 관련된 어휘 요소로 단어의 음절의 길이, 저빈도와 고빈도의 단어 사용(Nation, 2001), 내용어와 기능어의 사용(Davey, 1988) 등이 연구되어 왔다. 최근의 제1 언어(L1)와 제2 언어(L2) 어휘 연구는 어휘 풍부성 (lexical richness)의 측정에 많은 관심을 쏟고 있다. 어휘 풍부성의 측정은 구어나 문어의 텍스트에서 얼마나 많은 단어가 사용되었는지를 측정하는 것인데, 측정 방식에 따라서 어휘 풍부성의 다른 면을 측정하게 된다. 어휘 정교성(sophistication)은 텍스트에 사용된 고급 단어의 비율을 뜻하고, 어휘 밀집성(density)은 텍스트에서 사용된 내용어의 비율을 뜻하고, 어휘 다양성(diversity)은 텍스트에 사용된 단어 타입과 토큰의 비율을 뜻한다(Read, 2000; Daller, Milton과 Treffers-Daller, 2007).<sup>2)</sup>

이상의 논의는 2018학년도 본수능 및 6월 모평과 9월 모평의 문항 난이도

## 198 국제언어문학 제41호(2018.12.)

의 차이가 이 세 시험에 사용된 어휘 풍부성의 세 측면(즉, 어휘 정교성, 밀집성과 다양성)의 차이로부터 기인하는 것인지 아닌지를 따져볼 필요성을 제기한다. 따라서 본 연구는 다음의 연구 문제에 대한 답을 구하고자 한다.

첫째, 2018학년도 본수능 및 6월과 9월 모평은 시험에 사용된 어휘의 정교성에 있어서 차이가 있는가?

둘째, 2018학년도 본수능 및 6월과 9월 모평은 시험에 사용된 어휘의 밀집성에 있어서 차이가 있는가?

셋째, 2018학년도 본수능 및 6월과 9월 모평은 어휘의 다양성에 있어서 어느 정도의 차이가 있는가?

## 2. 이론적 배경 및 선행 연구

### 2.1 이론적 배경

어휘 풍부성과 관련된 논의에서 가장 기본이 되는 것은 '단어'의 의미이다. 어떤 정의 하에서 '단어'를 사용하느냐에 따라 연구 결과가 달라질 수 있기 때문에 처음부터 명확하고 엄밀하게 정의할 필요가 있다. 최근의 응용 언어학 연구에서 '단어'는 토큰(token), 타입(type), 레마(lemma) 그리고 군(family)으로 사용된다(Read, 2000; Nation, 2001). 오른쪽으로 갈수록, 이들 용어는 구체적인 것으로부터 더욱 일반적인 것을 의미한다.

토큰은 특정한 텍스트에서 사용되는 모든 단어로 중복적으로 사용되는 단어도 다른 형태로 간주한다. 반면에, 타입은 특정한 텍스트에서 사용되는 유일무이한 단어 형태를 의미한다. 따라서 중복적으로 사용되는 동일한 단어들은 하나의 단어 타입으로 간주된다. 레마는 문법적으로 관련된 한 집단의 단

2) 이 세 개념 외에도 내용에 국한된 것만 다르지만 어휘 다양성과 같은 어휘 별이성과 한 집단에서 한 사람만에 의해서 사용된 단어의 비율인 어휘 개별성이 있다.

어를 대표하므로, 한 단어의 모든 굴절형(즉, transform, transforms, transformed, transforming)을 포함한다. 한 단어의 군은 굴절형뿐만 아니라 규칙적인 파생으로 만들어진 형태(즉, transformer, transformation, transformable, transformative, transformational, transformationalism)도 포함한다.<sup>3)</sup>

텍스트에 사용된 고급 단어(기술적인 용어 또는 저빈도 단어)의 비율이 높을수록 어휘 정교성이 높아진다. Meara와 Bell(2001)의 아래 예문을 통해, 어휘 정교성에 대한 개념을 좀 더 명확히 할 수 있다.

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

비록 (1)-(3) 문장의 단어 수가 같지만, (1)의 단어들이 (2)와 (3)의 단어들보다 더 쉽고 고빈도 단어들이다. 다른 모든 조건이 같다면, (2)나 (3)의 문장을 포함하는 텍스트가 (1)의 문장을 포함하는 텍스트보다 어휘 정교성이 더 높다고 할 수 있다.

어휘 밀집성은 텍스트에서 차지하는 내용어(명사, 동사, 형용사, 대부분의 부사 등)의 비율을 뜻한다. 높은 비율의 내용어를 포함하고 있는 텍스트가 높은 비율의 기능어(대명사, 조동사, 전치사, 일부 부사, 접속사, 간투사, 수사)를 포함하고 있는 텍스트보다 정보가 더 조밀하게 포장되었다.

Ure(1971)가 어휘 밀집성이라는 개념을 도입한 이후, 내용어를 무엇으로 보는지에 관한 논란이 있었다. Ure(1971)는 구 동사 turn up을 하나의 내용어 turn과 하나의 기능어 up의 결합으로 보았고, Halliday(1985)는 turn up 전체를 하나의 내용어로 취급했다. Halliday(1985)는 또한 전치사와 always나 perhaps와 같은 서법 부사는 내용어와 기능어의 중간 상태의 범주에 속한다고 보았다. 본 논문에서는 Ure(1971)를 따라 구동사, 전치사 그리고 서법 부사를 every, some, no와 같은 양화사와 마찬가지로 기능어로 취급하였다.

3) 본 논문에서는 단어의 타입과 토큰만이 사용된다.

## 200 국제언어문학 제41호(2018.12.)

텍스트에 다른 단어가 더 많이 사용되고 같은 단어의 반복이 적을수록 어휘 다양성이 높아진다. 이러한 어휘 다양성을 계량하는 가장 전통적인 방법은 Templin(1957)이 제안한 텍스트의 총 토큰 수 대비 단어 타입의 수의 비율, 타입-토큰 비율(type-token ratio, TTR)이다. 하지만 TTR 계량은 텍스트에 새로 추가되는 타입의 수가 점점 줄어드는 문제를 여전히 안고 있다. 이는 더 긴 텍스트가 더 짧은 텍스트보다 일반적으로 더 낮은 TTR 값을 갖게 된다는 것을 함축한다. 이러한 문제를 해결하기 위한 하나의 방안은 Guiraud(1954)가 고안한 Guiraud 지수(GI)를 사용하는 것이다. GI는  $V/\sqrt{N}$ 으로 표기되는 바와 같이 타입을 토큰의 제곱근으로 나눈 값이다. 이렇게 얻은 GI는 길이가 다른 텍스트의 비교에서도 유용하다.

Malvern, Richards, Chipere와 Durán(2004)이 지적했듯이, 어휘 정교성은 어휘 다양성과 밀접한 관계를 가지고 있다. 긴 텍스트에서 다양성은 다른 단어를 추가해야만 증가되는데, 다른 단어가 더 많이 추가될수록 더 고급 단어를 사용하기 때문이다. 이런 점에 착안하여, Daller, van Hout와 Treffers-Daller(2003)는 Advanced Guiraud(AG)를 제안하였다. AG는 고급 단어 타입을 토큰의 제곱근으로 나눈 값( $V_{adv}/\sqrt{N}$ )이다. Daller(2010)는 GI가 1,000개부터 100,000개 사이의 토큰을 가진 텍스트에서 안정적인 결과를 내놓기 때문에 어휘 풍부성의 타당한 계량이지만, 빈도가 가장 높은 2,000(아마도 1,000이 더 좋음) 단어를 배제하는 것이 더 좋다고 주장한다.

## 2.2 선행 연구

본수능의 문항 난이도 요인에 관한 연구는 그동안 지속적으로 이뤄져 왔다(장경숙, 2004; Jin과 Park, 2004; Chon과 Shin, 2010). 본수능 외국어(영어) 영역의 점수 관련 요인 연구에서 성윤미(2004)는 학습자의 문법 지식과 어휘력 변인이 영어 점수와 상관관계가 있다고 주장한 반면, 장경숙(2004)은 2002-3의 2개 학년도의 본수능 외국어(영어) 영역 읽기 문항을 토대로 문항 난이도와 실제 정답율 간의 상관관계에서 어휘 및 숙어 수준이 별 영향을 미치고 못한다고 분석하였다. 김낙복(2008)은 2001-8학년도의 대학수능능력

시험 외국어 영역 텍스트에 나타난 TTR, 고빈도 단어, 언어 관계를 중점적으로 다루면서 효과적인 영어 어휘 교육의 시사점을 얻고자 하였다.

이상의 선행 연구는 본수능 영어 영역 문항의 난이도를 예측하기 위한 모형 개발에 초점이 맞춰져 있거나 영어 어휘 교육의 시사점을 얻기 위한 목적으로 보이지만, 본수능 영어 영역 문항에서 어휘들이 어떻게 사용되었는지 통계적으로 밝히지 못하고 있으며 TTR의 간헐적인 언급 외에는 어휘 정교성, 밀집성, 다양성과 관련된 통계학적인 분석이 결여되어 있다. 이러한 선행 연구의 결과를 보완하기 위해, 본 연구는 이소영과 이현우(2017)와 이현우(2017)의 방법론을 따라 본수능 및 모평 영어 영역에서 사용된 단어를 어휘 풍부성의 세 측면에서 통계적으로 분석한 결과를 제시한다.

### 3. 연구 방법 및 절차

#### 3.1. 분석 대상

앞서 언급한 대로, 본 연구의 분석 대상은 2018학년도 본수능과 6월 모평 및 9월 모평 영어 영역이다. 이 세 시험에 사용된 단어에 대한 개략적인 정보는 아래 <표1>과 같다.

표 1 분석 대상의 단어에 대한 정보

분석 대상	단어 타입	단어 토큰
6월 모평	1,917	6,503
9월 모평	1,795	6,352
본수능	1,925	6,648

<표1>의 단어는 듣기와 읽기 시험에 사용된 단어를 의미한다. 9월 모평에

#### 202 국제언어문학 제41호(2018.12.)

서 가장 적은 단어 타입과 토큰이 사용되었고, 6월 모평과 본수능은 사용된 단어 타입과 토큰에 있어서 별 차이가 없다. 단어 토큰에 있어서, 9월 모평은 6월 모평보다는 151개, 본수능보다는 296개 적다.

#### 3.2. 단어의 등급화

어휘 정교성, 다양성, 그리고 밀집성에 관한 연구 문제를 해결하려면 우선 단어의 등급화가 필요하다. 2009 개정 영어과 교육과정의 기본 어휘를 기반으로 한 단어 등급화는 다음과 같다.<sup>4)</sup>

그룹 1: 교육과정의 기본 어휘 목록에 있는 단어로서 초등학교에서의 사용을 권장하는 단어이거나 Nation(2012)의 BNC/COCA 1st 1000 또는 2nd 1000 단어 목록에 있는 단어(1,881개)

그룹 2: 그룹 1에 속하지 않는 기본 어휘 목록에 있는 단어 또는 기본 어휘가 아닌 단어(1,024개+a개)

2018학년도 본수능과 6월 모평 및 9월 모평에 사용된 등급화된 단어의 분포는 <표2>와 같다.<sup>5)</sup>

표 2 분석 대상의 등급화된 단어에 대한 정보

분석 대상	등급	단어 타입	단어 토큰
6월 모평	1	1,278	5,564
	2	524	750
9월 모평	1	1,260	5,565

4) 본 논문에서는 2,000 단어 균을 실고 있는 West(1953)의 *The General Service List*를 따라, 2,000 단어 균을 그룹 1과 2의 분기점으로 삼았다.

5) <표2>의 등급화에는 교육과정의 기본 어휘로 취급하지 않는 외래어와 이름과 같은 고유 명사, 계절이나 요일과 같은 일련의 단어들은 배제되었다.

	2	442	607
본수능	1	1,266	5,682
	2	537	742

<표2>에서 보듯이, 기본 어휘에 속하는 그룹 1의 단어 타입의 빈도는 최대 편차가 18개로 세 시험이 매우 비슷하나, 그룹 2의 경우는 최대 편차가 95개에 이르는 빈도의 차이를 보이고 있다. 단어 토큰의 빈도도 단어 타입의 빈도와 유사한 양상을 보인다. 그룹 1에서는 최대 편차가 118개이지만 그룹 2에서는 143개나 되어 비유상 큰 차이를 보인다.

### 3.3. 연구 가설

어휘 정교성과 밀집성과 관련된 연구 문제의 해결을 위해 다음과 같이 대립 가설을 설정한다.

#### 대립 가설 1

2018학년도 본수능 및 6월과 9월 모평은 등급화된 단어 토큰의 빈도와 관계가 있다.

#### 대립 가설 2

2018학년도 본수능 및 6월과 9월 모평은 내용어와 기능어 토큰의 빈도와 관계가 있다.

어휘 다양성과 관련된 연구 문제의 해결을 위해 시험별로 사용된 총 단어 타입과 토큰의 비율과 그룹 2의 단어 타입과 토큰의 비율을 비교하고자 한다.

### 3.4. 분석 도구 및 방법

2018학년도 본수능과 6월과 9월 모평에서 등급화된 단어가 어떠한 빈도로 사용되었는지 알아보기 위해, Heatley와 Nation(2006)의 Range32H를 사용

## 204 국제언어문학 제41호(2018.12.)

한다. 이 소프트웨어는 각 시험에서 등급화된 단어 군, 타입, 토큰의 빈도를 보여준다.

Range32H의 작동을 위해 앞서 논한 두 개의 등급화된 단어 그룹과 외래어 그룹 및 고유 명사 등의 그룹으로 구성된 네 개의 베이스워드 목록이 사용된다. 베이스워드 목록은 표제어와 표제어로부터 굴절이나 교육과정의 기본 어휘 목록에서 허용되는 파생에 의해 만들어진 단어가 포함된다. 예를 들어, 베이스워드 목록에는 표제어 nation뿐만 아니라 national, nationally, nationality, unnational, nationalize, nationalizing, nationalized, nationalizes, nationalization, nationalizer 등이 포함된다.

등급화된 단어 토큰과 내용어와 기능어의 단어 토큰의 빈도를 조사한 후, 이들 빈도와 2018학년도 본수능 및 6월과 9월 모평 간에 관계가 있는지를 따지기 위해 SPSS 23.0 for Windows를 활용하여 카이제곱 검정을 한다. 두 변수 간에 유의미한 관계가 있는 경우, Bonferroni의 보정법에 의한 사후 분석을 한다.

## 4. 연구 결과 및 논의

### 4.1. 대립 가설 1 검증

2018학년도 본수능 및 6월과 9월 모평이 등급화된 단어 토큰의 사용에 있어서 차이가 있다는 대립 가설 1을 검증하기 위해, 교차분석을 한 결과는 아래 <표3>과 같다.

표 3 등급화된 단어 토큰에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	15.210 <sup>a</sup>	2	.000

유효 케이스 수	18,910		
----------	--------	--	--

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 685.09임.

유의확률이 .000으로 유의수준인 .05보다 작다. 따라서 2018학년도 본능 및 6월과 9월 모평은 등급화된 단어 토큰의 사용에 있어서 차이가 있다는 대립 가설 1을 수용할 수밖에 없다.

이와 같이 대립 가설 1의 수용에 결정적으로 기여한 차이가 어디에 있는지 조사하고자 Bonferroni 방법에 의한 사후 분석을 하였다. 그 결과는 <표 4>와 <표 5>와 같다.

표 4 6월과 9월 모평의 등급화된 단어 토큰에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	13.456a	1	.000
유효 케이스 수	12,486		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 670.78임.

표 5 9월 모평과 본능의 등급화된 단어 토큰에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	9.689a	1	.002
유효 케이스 수	12,596		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 661.01임.

<표 4>에서 보듯이, 6월과 9월 모평에 사용된 그룹 1과 2의 단어 토큰의 빈도에 대한 카이제곱 검정 결과, 유의확률이 .000였다. Bonferroni 보정법에 의해, 시험 간 비교에 의한 다중 검정 회수 3을 이 p값에 곱해도 .000으로 .05보다 작다. 따라서 그룹 1과 2의 단어 토큰의 빈도가 6월 모평과 9월 모평에 따라서 유의미하게 달라졌음을 알 수 있다. <표 5>에서는 유의확률이

206 국제언어문학 제41호(2018.12.)

.002였다. Bonferroni 보정법에 의해, 이 p값에 3을 곱해도 .006으로 .05보다 작다.6) 따라서 그룹 1과 2의 단어 토큰의 빈도가 9월 모평과 본능에 따라서 유의미하게 달라졌음을 알 수 있다. 결국, 6월 모평과 9월 모평 사이의 차이와 9월 모평과 본능 사이의 차이가 대립 가설 1의 수용에 결정적인 역할을 하였다.

4.2. 대립 가설 2 검증

2018학년도 본능 및 6월과 9월 모평이 내용어와 기능어 토큰의 빈도와 관계가 있다는 대립 가설 2의 검증을 위해 각 시험에 사용된 내용어와 기능어의 총 토큰 빈도를 아래와 같이 조사했다.

표 6 분석 대상의 내용어와 기능어 토큰 빈도

분석 대상	내용어	기능어	합계
6월 모평	3,033	3,470	6,503
9월 모평	3,101	3,251	6,352
본능	3,232	3,416	6,648

<표 6>의 정보를 토대로 교차분석을 한 결과는 아래 <표 7>과 같다.

표 7 내용어와 기능어 토큰에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	7.533a	2	.023
유효 케이스 수	19,512		

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

6) 6월 모평과 본능 사이의 보정된 p값은 1.695이다.

3,050.45임.

유의확률이 .023으로 유의수준인 .05보다 작다. 따라서 2018학년도 본수능 및 6월과 9월 모평은 내용어와 기능어 토큰의 사용에 있어서 차이가 있다는 대립 가설 2를 수용할 수밖에 없다.

대립 가설 2의 수용에 결정적으로 기여한 차이가 어디에 있는지 조사하고자 Bonferroni 방법에 의한 사후 분석을 하였다. 그 결과는 <표8>과 같다.

표 8 6월과 9월 모평의 내용어와 기능어 토큰에 대한 교차분석

	값	자유도	근사 유의확률(양측 검정)
Pearson 카이제곱	6.117a	1	.013
유효 케이스 수	12,885		

a. 0셀(0.0%)은 5보다 작은 기대빈도를 가지는 셀임. 최소 기대빈도는 3,030.97임.

6월과 9월 모평에 사용된 내용어와 기능어 토큰의 빈도에 대한 카이제곱 검정 결과, 유의확률이 .013였다. Bonferroni 보정법에 의해, 이 p값에 다중 검정 회수 3을 곱해도 .039로 .05보다 작다. 내용어와 기능어 토큰의 빈도가 6월 모평과 9월 모평에 따라서 유의미하게 달라졌음을 알 수 있다.<sup>7)</sup> 6월과 9월 모평 사이의 차이가 대립 가설 2의 수용에 결정적인 역할을 하였다.

### 4.3 단어 타입-토큰의 비교

어휘 다양성을 어떤 방식으로 구하는 간에 핵심은 단어의 토큰 수 대비 타입의 수의 비율이다. 하지만 한 텍스트의 단어 토큰의 집합은 타입의 집합을 포함하기 때문에, 토큰과 타입을 독립 변수로 하는 카이제곱 검정은 불가능하

7) 6월 모평과 본수능 사이의 보정된 p값은 .069이고, 9월 모평과 본수능 사이의 보정된 p값은 2.451이다.

## 208 국제언어문학 제41호(2018.12.)

다.

이에 따라 본 논문에서는 어휘 다양성을 계량하는 가장 일반적인 두 방법인 타입-토큰 비율(TTR)과 Guiraud 지수(GI)를 사용하여, 2018학년도 본수능 및 6월과 9월 모평의 어휘 다양성을 비교하고자 한다.

세 시험의 TTR과 GI의 산출을 위해 각 시험에서 사용된 총 단어 타입 및 토큰 수 그리고 등급화된 그룹 2에 속하는 단어 타입의 수를 알아야 하는데, 이들에 관한 정보는 <표1>과 <표2>에 나와 있다. TTR과 GI 산출 공식에 따라 얻은 값은 다음과 같다.

표 9 어휘 다양성 계량

분석 대상	TTR	GI	총 토큰
6월 모평	.295	23.772	6,503
9월 모평	.283	22.522	6,352
본수능	.290	23.609	6,648

3.1에서 언급했듯이, 단어의 총 토큰 수에 있어서, 9월 모평은 6월 모평보다는 151개, 본수능보다는 296개 적다. 296개의 편차는 본수능 총 토큰 수의 4.45%에 해당하기 때문에 TTR과 GI 둘 다 유효한 계량치를 보여준다고 여겨진다. 미세한 차이이지만, 어휘 다양성에 있어서, 6월 모평이 가장 높고, 그 다음이 본수능이고, 마지막이 9월 모평의 순이다.

### 4.4. 논의

이전의 논의는 2018학년도 본수능 및 6월과 9월 모평이 등급화된 단어 토큰의 사용에 있어서 차이가 있다는 대립 가설 1이 수용되었고, 이는 6월과 9월 모평의 차이 그리고 9월 모평과 본수능의 차이에 기인한다는 것을 보여주었다. 이는 9월 모평이 다른 두 시험과 다른 점이 있음을 의미하며, 대립 가설 1의 수용은 어휘 정교성과 관련하여 세 시험이 유의미한 차이를 보인다는

것을 가리킨다.

그렇다면 어느 시험이 어느 다른 시험보다 어휘 정교성에 있어서 차이가 보이는가에 대한 답을 해야 한다. 2장의 이론적 배경에서 한 텍스트의 AG 값이 클수록 고급(저빈도) 단어가 더 많이 사용되었다는 것을 뜻한다고 했다. Daller(2010)의 제안대로 고급 단어를 그룹 2 또는 그 이상의 단어로 정하고 세 시험의 AG 값(Vadv/N)을 구해보면, 6월 모평이 6.498, 9월 모평이 5.546, 본수능이 6.586이 된다. 대립 가설 1의 수용은 결국 이 세 AG 값의 차이가 유의미한 차이라는 것을 말하는 것이다. 결론적으로, 본수능의 텍스트가 어휘 정교성에서 가장 높고 9월 모평이 가장 떨어진다고 할 수 있다.

어휘 정교성에 관한 이러한 결과는 9월 모평이 다른 두 시험보다 문항 난이도가 높았다는 사실과 모순이 된다. 본 연구의 결과가 맞는 것이라고 한다면, 세 시험의 문항 난이도는 어휘 정교성과 무관한 것으로 결론을 내릴 수 있겠다.

2018학년도 본수능 및 6월과 9월 모평이 내용어와 기능어 토큰의 빈도와 관계가 있다는 대립 가설 2가 수용되었고, 이는 6월과 9월 모평의 차이에 기인한다는 사실은 이 두 시험의 정보 포장이 유의미한 차이를 보인다는 것을 가리킨다.

그렇다면 어느 시험의 정보 포장이 다른 시험보다 더 조밀한가? 6월 모평에서는 총 6,503개의 토큰 중에서 46.6%에 해당하는 3,033개의 내용어가 사용되었고, 9월 모평에서는 총 6,352개의 토큰 중에서 48.8%에 해당하는 3,101개의 내용어가 사용된 것으로 보아, 9월 모평이 6월 모평보다 정보 포장이 세밀하게 되었다고 할 수 있다.

그럼에도 불구하고, 본수능에서 총 6,648개의 토큰 중에서 48.6%에 해당하는 3,232개의 내용어가 사용되었다는 점과 9월 모평에서 사용된 총 토큰의 수가 본수능보다 296개나 적다는 점에서 어휘 밀집성의 유의미한 차이가 이들 세 시험의 문항 난이도에 영향을 끼쳤다고 말하기는 어려워 보인다.

바로 앞 절의 <표9>에서 보듯이, 2018학년도 본수능 및 6월과 9월 모평에서의 TTR과 GI 값은 차이가 미세하다. TTR 값의 차이가 .005부터 .012 그리고 GI 값의 차이가 .163부터 1.25까지인 것으로 보아 이들 세 시험이 어휘

## 210 국제언어문학 제41호(2018.12.)

다양성에 있어서 별 차이가 없다고 할 수 있겠다.

결론적으로, 어휘 정교성에 있어서는 본수능이, 어휘 밀집성에 있어서는 9월 모평이, 그리고 어휘 다양성에 있어서는 6월 모평이 가장 높게 나타났다. 어휘 다양성에 관한 세 시험의 차이가 통계적으로 유의미한지 알 수 없고 어휘 밀집성보다는 어휘 정교성이 문항 난이도에 영향을 더 준다는 점에서, 2018학년도 본수능 및 6월과 9월 모평의 문항 난이도 차이는 어휘와 관련된 요인보다는 텍스트의 추상성, 단락의 길이, 주제에 대한 익숙함 정도, 문장의 복잡도 등과 같은 다른 요인에 기인한 것으로 여겨진다.

## 5. 결론

본 논문은 문항 난이도에 있어서 상당한 차이를 보인 2018학년도 본수능과 6월과 9월 모평이 어휘 정교성과 밀집성에 있어서 유의미한 차이가 있는지를 카이제곱 검정을 통해 밝히고 세 시험의 어휘 다양성을 비교하고자 했다. 카이제곱 검정 결과, 등급화된 단어 토큰의 분포(어휘 정교성)와 내용어 토큰의 비율(어휘 밀집성)에 있어서 세 시험이 통계적으로 유의미한 차이가 있는 것으로 나타났다.

사후 분석 결과, 이 차이의 원인은 9월 모평에서의 고빈도 단어의 과다 사용과 저빈도 단어의 과소 사용에 있었음이 밝혀졌다. 하지만 이 결과는 9월 모평이 두 시험보다 더 어려웠다는 사실과 모순되며, 따라서 세 시험의 문항 난이도를 어휘 정교성과 관련 지을 수가 없음을 시사한다. 비록 어휘 밀집성에 있어서 세 시험이 유의미한 차이를 보였지만, 이들 시험의 총 토큰 수가 달라, 이를 문항 난이도의 차이를 야기한 요인으로 보기 어렵다. 어휘 다양성 측정치의 미미한 차이 또한 문항 난이도 설명에 큰 도움이 되지 못함을 밝혔다.

본 연구에서 얻은 결과는 몇 가지 시사점을 던져 준다. 첫째, 본수능과 모평 영어 시험이 어휘에 있어서 동일한 시험이 되기 위해서는 이들 시험의 AG

값( $Vadv/\sqrt{N}$ )이 통계적으로 유의미한 차이가 나지 않도록 해야 한다. 1,000 또는 2,000 단어 수준 이상의 단어를 AG 값이 같은 범위 내에서 사용함으로써 이들 시험이 어휘적으로 더욱 안정적인 시험이 되도록 노력해야 한다. 둘째, 동일한 어휘 정교성이 확보되면, 어휘 밀집성과 다양성도 자연스럽게 확보될 수 있겠지만, 가급적 세 시험의 밀집성과 다양성도 거의 같도록 하는 것이 필요하다. 셋째, 어휘 요인 외에도 시험의 난이도에 영향을 주는 다른 요인들도 신중하게 고려하여 이들 시험이 더욱 안정적으로 시행되도록 하는 것이 필요하다.

212 국제언어문학 제41호(2018.12.)

#### [Abstract]

#### Vocabulary Analysis of the English Section of the 2017 CSAT and CSAT Simulations

Hyoen woo-Lee

Given the English Section of the 2017 College Scholastic Ability Test (CSAT) and its June- and September-simulations substantially differ in item difficulty, this paper addresses the question of whether the three tests differ significantly with respect to lexical sophistication (the distribution of advanced words) and lexical density (the portion of content words), comparing lexical diversity (the type-token ratio) within the tests. Chi-square tests show that for lexical sophistication and density, there is a significant difference inasmuch as word tokens are concerned. The post-hoc analyses show that these differences resulted from the fact that the 2017 CAST September- simulation used more high frequency words and less low frequency ones than the other tests. This finding, however, contradicts the fact that the CSAT September-simulation was most difficult, suggesting that the item difficulty of the three tests could not be associated with lexical sophistication. Though the tests significantly differ in lexical density, this fact cannot be taken as a factor that caused them to differ in item difficulty, nor can small differences of measures of lexical diversity.

**Key words:** CSAT (simulation) English Section, Item difficulty, Lexical sophistication, Lexical diversity, Lexical density, Type-token ratio, Chi-square test

## 참고문헌

## □ 단행본

- 교육과학기술부, 『영어과 교육과정. 교육과학기술부 고시 제 2011-361호 [별책 14]』, 교육과학기술부, 2011.
- 한국교육과정평가원, 『대학수학능력시험 20년사』, 한국교육과정평가원, 서울, 2014.
- Daller, H., Milton, J., & Treffers-Daller, J., (eds.), *Modelling and assessing vocabulary knowledge*, Cambridge University Press, Cambridge, 2007.
- Halliday, M. A. K., *Spoken and written language*, Deakin University Press, Geelong, Victoria, 1985.
- Hudson, T., *Teaching second language reading*, Oxford University Press, Oxford, 2007.
- Guiraud, P., *Les caractères statistiques du vocabulaire: Essai de méthodologie*, Presses Universitaires de France, Paris, 1954.
- Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P., *Lexical diversity and language development: quantification and assessment*, Palgrave Macmillan, Houndmills etc., 2004.
- Nation, I. S. P., *Learning vocabulary in another language*, Cambridge University Press, Cambridge, 2001.
- Paivio, A., *Mental representations*, Oxford University Press, New York, 1986.
- Read, J., *Assessing vocabulary*, Cambridge University Press, Cambridge, 2000.
- Templin, M. C., *Certain language skills in children: Their development and inter-relationships*, University of Minnesota Press, Minneapolis, 1957.
- West, M., *A general service list of English words*, Longman, Green and Co, London, 1953.

## □ 논문 및 기타 자료

- 김낙복, 「대학수학능력시험 외국어(영어)영역의 코퍼스 언어학적 어휘 비교 분석」, 『영어어문교육』, 14(4), 2008, 201-20면.
- 성윤미, 「대학수학능력시험 외국어(영어) 영역에 영향을 미치는 요인들」, 『영어어문교육』, 9(2), 2004, 213-41면.
- 이소영, 이현우, 「중3 영어 교과서에 나타나는 영어 어휘의 분포」, 『중등영어교육』, 10(4), 2017, 121-42면.
- 이현우, 「중3 영어 교과서와 학업성취도평가 시험에 나타나는 부정사의 분포」, 『국제

## 214 국제언어문학 제41호(2018.12.)

- 언어문학』, 36, 2017, 229-60면.
- 장경숙, 「대학수학능력시험 외국어(영어)영역 읽기 난이도 예측 모형 개발. 외국어 교육」, 11(1), 2004, 111-30면.
- 한국교육과정평가원, 「2018학년도 대학수학능력시험 6월 모의평가 채점 결과 보도자료」, 한국교육과정평가원, 서울, 2017.
- 한국교육과정평가원, 「2018학년도 대학수학능력시험 9월 모의평가 채점 결과 보도자료」, 한국교육과정평가원, 서울, 2017.
- 한국교육과정평가원, 「2018학년도 대학수학능력시험 채점 결과 보도자료」, 한국교육과정평가원, 서울, 2017.
- Abrahamsen, E. P., & Shelton, K. C., Reading comprehension in adolescents with learning disabilities: Semantic and syntactic effects, *Journal of Learning Disabilities*, 22(9), 1989, 569-72면.
- Chon, Y., & Shin, T., Item difficulty predictors of a multiple-choice reading test, *English Teaching*, 65(4), 2010, 257-82면.
- Daller, H., van Hout, R., & Treffers-Daller, J., Lexical richness in the spontaneous speech of bilinguals, *Applied Linguistics*, 24(2), 2003, 197-222면.
- 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/>.
- Davey, B., Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers, *Journal of Experimental Education*, 56(2), 1988, 67-76면.
- Drum, P. A., Calfee, R. C., & Cook, L. K., The effects of sentence structure variables on performance in reading comprehension tests, *Reading Research Quarterly*, 16, 1981, 486-514면.
- Freedle, R., & Kostin, I., The prediction of TOEFL reading item difficulty: implications for construct validity, *Language Testing*, 10(2), 1993, 133-67면.
- Graves, M. F., Vocabulary learning and instruction, *Review of Research in Education*, 13, 1986, 49-89면.
- Hare, V., Rabinowitz, M., & Schieble, K., Text effects on main idea comprehension, *Reading Research Quarterly*, 24, 1989, 72-88면.
- Heatley, A., & Nation, I. S. P., Range32H (computer software), Victoria University of Wellington, Wellington, New Zealand, 2006.
- Jin, K.-A., & Park, C., The prediction of English item difficulty in college scholastic ability test, *English Teaching*, 59(1), 2004, 267-78면.
- Kieras, D. E., Thematic processes in prose, in B. Britton & J. Black (eds.),

- Understanding expository text*, Erlbaum, Hillsdale, NJ, 1985, 11-64면.
- Klare, G. R., Assessing readability, *Reading Research Quarterly*, 10(1), 1974-1975, 62-102면.
- Klare, G. R., Readability, in P. D. Pearson (ed.), *Handbook of reading research*, Longman, New York, 1984, 681-744면.
- Leow, R., To simplify or not to simplify: A look at intake, *Studies in Second Language Acquisition*, 15(3), 1993, 333-56면.
- Meara, P., & Bell, H., P\_Lex: a simple and effective way of describing the lexical characteristics of short L2 texts, *Prospect* 16, 2001, 5-19면.
- Nation, I. S. P., The BNC/COCA headword lists, Retrieved September 15, 2018 from <https://www.victoria.ac.nz/lals/about/staff/paul-nation/#vocab-lists>, 2012.
- Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S., Where's the difficulty in standardized reading tests: The passage or the question?, *Behavior Research Methods*, 40(4), 2008, 1001-15면.
- Qian, D. D., Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension, *The Canadian Modern Language Review*, 56(2), 1999, 282-308면.
- Qian, D. D., Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective, *Language Learning*, 52(3), 2002, 513-36면.
- 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 1969, Cambridge University Press, Cambridge, 1971, 443-52면.
- Yano, Y., Long, M. H., & Ross, S., The effect of simplified and elaborated texts on foreign language reading comprehension, *Language Learning*, 44(2), 1994, 189-219면.

이 논문은 2018년 11월 15일 접수되어 12월 06일까지 심사받아 12월 17일 게재 확정됨.