

자동화된 언어분석 프로그램을 활용한 중학교 영어교과서의 학년 간 연계성 분석*

전문기

건국대학교

Jeon, Moongee. (2015). An analysis of the continuity among middle school English textbooks with an automated language analysis program. *Modern English Education*, 16(1), 195-218.

The main purpose of this study was to analyze the continuity of the reading scripts in the Korean middle school English 1, 2, 3 textbooks using an automated language analysis program, Coh-Metrix. Coh-Metrix is a computer program which can be applied to evaluate English textbooks based on many linguistic and psycholinguistic measures. The Coh-Metrix measures analyzed for this study include basic counts (the number of words, the number of sentences, average sentence length), word frequency, standard readability scores (Flesch reading ease score, Flesch-Kincaid grade level), syntactic complexity (subject density, noun density), lexical diversity (type-token ratio), co-reference cohesion, semantic cohesion, pronouns, and connectives. In general, the findings of this study showed that the continuity among the middle school English 1, 2, 3 textbooks was controlled properly for the basic counts and the subject density score. However, the continuity among the middle school English 1, 2, 3 textbooks was not controlled suitably for the other measures. The findings of this study embrace valuable implications for the development of English textbooks.

[English textbook/continuity/corpus/corpus analysis/computer tool/
영어교과서/연계성/코퍼스/코퍼스 분석/컴퓨터 도구]

I. 서론

중등영어교육 현장에서 영어교과서는 학습자가 영어를 효율적으로 학습하는데 일차적으로 요구되는 중요한 학습 매개체이다(성일호, 2014; 전문기, 2011; 전문기, 임인재, 2009). 이러한 맥락에서, 성공적인 교육과정의 실현을 위해 영어

* 이 논문은 2012년도 건국대학교 학술진흥연구비 지원에 의한 논문임.

교과서의 역할은 매우 크다고 볼 수 있다(김지은, 전문기, 2013).

영어교과서의 중요성에 대한 인식 속에서 영어교과서에 대한 연구는 다양한 각도에서 수행되었다. 영어교과서의 읽기 자료를 분석한 연구(김민정, 송희심, 2007; 조원향, 2003), 듣기 및 말하기 활동 자료를 분석한 연구(김은주, 2001), 쓰기 활동 자료를 분석한 연구(손미용, 이재근, 2005), 영어교과서에 포함된 문화 교육에 대한 연구(정은혁, 민주경, 2005), 연계성을 분석한 연구(이정민, 김해동, 2007) 등이 있다. 그러나 제한된 자료 분석과 객관적 분석 도구의 부재로 인해 이러한 연구들의 분석 결과를 일반화하여 확대 해석하는데 문제점과 제한점이 시사되었다(고광운, 2008; 고광운, 박정준, 2007; 전문기, 2014).

대용량의 코퍼스 자료와 컴퓨터과학 분야(Jurafsky & Martin, 2008)에서 개발된 다양한 언어분석 도구들을 영어교과서 연구에 활용하면서 선행연구들에서 야기된 제한점과 문제점을 해결할 수 있는 가능성이 제시되어 왔다. 구체적으로 코퍼스언어학에서 소개된 다양한 코퍼스 구축 및 분석 방법들(Lindquist, 2009; McEnergy & Wilson, 2007; Meyer, 2002)과 전산언어학에서 개발된 다양한 컴퓨터 분석 도구들을 바탕으로 대용량의 영어교과서 구축과 객관적인 영어교과서 분석이 가능해졌다(고광운, 2008; 고광운, 박정준, 2007; 권인숙, 2002; 권인숙, 2004; 김정렬, 천윤희, 2008). 코퍼스에 기반한 이러한 영어교과서 분석 연구들은 방대한 크기의 교과서 코퍼스 자료와 컴퓨터 프로그램을 기반으로 하여 자료 분석의 객관성과 신뢰성을 증가시켰다는 점에서 의의를 찾을 수 있다. 그러나 이러한 코퍼스 기반 연구들에 활용된 컴퓨터 분석 도구들은 단어수, 문장수, 평균문장길이, 타입-토큰 비율(type-token ratio) 등과 같은 표층수준의 측정치만을 제공한다는 점에서 제한점이 대두되었다(김지은, 전문기, 2013; 전문기, 2011; 전문기, 2014).

최근 들어, 이러한 기존 코퍼스 기반 영어교과서 분석 연구들에서 제기된 제한점을 해결하기 위한 대안으로 Coh-Metrix(Graesser, M. Jeon, Yan, & Cai, 2007; Graesser, McNamara, Louwerse, & Cai, 2004)와 같은 자동화된 컴퓨터 언어분석 프로그램이 영어교과서 연구에 활발히 이용되고 있다(김정렬, 2014; 김정렬, 양지윤, 2012; 김지은, 전문기, 2013; 성일호, 2014; 송해성, 2013; 전문기, 2011; 전문기, 2014; 전문기, 임인재, 2009; 이승환, 2013; 이정민, 현태덕, 2013; 이화자, 2011; 임인재, 전문기, 2013). Coh-Metrix는 미국 멤피스대학교(The University of Memphis)의 지능형시스템연구소(Institute for Intelligent Systems)에서 개발한 교재 및 텍스트 분석 컴퓨터 시스템이다(Graesser, Jeon, Yan, & Cai, 2007; Graesser, et al., 2004). Coh-Metrix는 표층수준의 측정치뿐만 아니라 텍스트의 정합성, 통사적 구조, 어휘정보, 접속사, 대명사 등과 같은 보다 다양하고 광범위한 유형의 언어학 및 심리언어학적 측정치를 제공한다.

본 연구는 개정 교육과정의 중학교 영어교과서의 연계성을 광범위한 유형의 언어학 및 심리언어적 측정치에 기반해 체계적이고 객관적으로 분석하기 위해 자동화된 언어분석 컴퓨터 프로그램인 Coh-Metrix를 활용하였다. 구체적으로 중

학교 1, 2, 3학년 영어교과서의 연계성을 총체적으로 분석하기 위해 각 교과서에 대한 학년별 영어교과서 코퍼스 자료를 구축하여 개인 PC용 Coh-Metrix를 이용하여 분석하였다.

II. 이론적 배경

지구 사회가 점차 글로벌화되면서 국가 구성원 간 의사소통의 수단으로 영어의 중요성이 부각되고 있다(전문기, 2014). 이러한 점을 고려하면 학교 영어 교육현장에서 성공적인 교육과정의 실현을 통해 경쟁력 있는 글로벌 인재를 양성하기 위해 영어 교육은 매우 중요하다고 볼 수 있다(임인재, 전문기, 2013). 특히, 우리나라와 같은 EFL(English as a Foreign Language) 영어교육 환경에서 효과적인 영어 교육을 위해 영어교과서는 매우 중요한 학습 매개체라고 할 수 있다(권인숙, 2002; 김민정, 송희심, 2007). 이러한 맥락 속에서 영어교과서에 대한 연구가 영어교육 연구의 중요한 핵심부분을 차지하고 있다고 볼 수 있겠다(손미용, 이재근, 2005; 송자경, 윤이정, 2003).

예를 들면, 손미용과 이재근(2005)은 13종의 중학교 1학년 영어교과서 내에 포함된 쓰기 활동 관련 자료를 체계적으로 연구하였다. 손미용과 이재근의 연구 결과, 중학교 1학년 영어교과서 쓰기 활동 자료는 대부분 통제 작문과 유도 작문 활동으로 구성되어 있었다. 이러한 연구결과는 통제 및 유도와 같은 특정 유형의 작문이 학습자의 학습 동기를 감소시킬 가능성이 있기 때문에 다양한 유형의 작문 활동이 더 개발되어야 함을 제시한다.

본 연구의 주목적인 영어교과서의 연계성 주제와 관련하여, 김은주(2001)는 중학교 1학년 영어교과서에 포함된 듣기 지문 자료를 분석하였다. 김은주의 연구 결과, 중학교 영어교과서에 포함된 듣기 활동은 교육과정에서 제시한 학습 목표를 대체로 잘 제시하고 있었다. 그러나 자유롭게 반응할 수 있는 듣기 활동의 비율이 상대적으로 낮아서 교사가 학습자의 자발적인 반응을 유도할 수 있는 학습 전략이 요구됨이 제시되었다.

다른 한편, 이정민과 김해동(2007)은 고등학교 영어회화 교과서와 대학수학능력시험 영어영역 듣기 자료 사이의 연계성을 분석하였는데 분석 결과, 영어회화 교과서와 대학수학능력시험 듣기자료 간의 연계성이 잘 통제되지 않았음이 제시되었다. 송자경과 윤이정(2003)은 초등학교 영어교과서의 연계성을 분석하였는데 그들의 분석 결과, 초등학교 영어교과서의 연계성이 의사소통의 기능적 측면에서 잘 통제되지 않았음이 보고되었다. 요약하면, 이러한 연구결과들은 영어교과서의 학년 간 난이도와 이해도를 체계적으로 분석하기 위해 영어교과서의 학년 간 연계성에 대한 연구가 중요함을 시사한다고 하겠다. 또한, 이러한 영어교과서의 학년 간 연계성에 연구들은 제한된 영어교과서 자료와 컴퓨터 기반 언어분석 도구와 같은 객관적 분석 도구의 부재로 인해 연구 결과를 일반화

하는데 문제점과 제한점이 야기되었고 자연스럽게 코퍼스 기반 영어교과서 연구들을 활성화시켰다(고광윤, 2008; 권인숙, 2002; 권인숙, 2004).

1. 용어색인 프로그램 기반 영어교과서 분석 연구

앞서 제시된 영어교과서 연구들에서 제시된 문제점과 제한점을 극복하기 위한 대안으로 코퍼스언어학에서 제시된 다양한 텍스트 분석 방법과 컴퓨터과학에서 제시된 언어분석 도구들을 영어교과서 연구에 적용한 연구들이 수행되었다(고광윤, 2008; 고광윤, 박정준, 2007; 권인숙, 2002; 권인숙, 2004; 김정렬, 천윤희, 2008; 문안나, 2008).

고광윤(2008)은 고등학교 영어독해 교과서에 수록된 관계대명사, 인칭대명사, 전치사, 접속사, 조동사와 같은 기능어가 원어민들의 영어와 얼마나 유사하게 사용되는지를 분석하였다. 구체적으로 영어독해 교과서에 대한 코퍼스 자료를 구축하여 코퍼스 분석 프로그램인 WordSmith Tools 4.0을 이용해 분석하였다. 분석 결과, 고등학교 영어독해 교과서에 포함된 기능어는 원어민의 영어와 다소 다르게 사용됨이 관찰되었지만 영어 학습에 미치는 영향은 크지 않음이 제시되었다.

문안나(2008)는 고등학교 영어교과서에 포함된 동사가 원어민의 영어 특성을 얼마나 잘 반영하는지를 분석하였다. 구체적으로 고등학교 영어교과서 35종에 대한 코퍼스 자료를 구축한 후, WordSmith Tools 4.0의 용어 색인(concordance) 기능을 이용해 특정 동사의 발생빈도를 분석하였다. 그리고 이 동사들에 대한 발생빈도를 BNC(British National Corpus) 코퍼스를 통해 다시 산출하여 두 발생빈도를 비교 분석하였다. BNC 코퍼스와 비교 분석 결과, 고등학교 영어교과서에 포함된 몇몇 동사들은 BNC 코퍼스와 비교했을 때 발생빈도가 너무 높거나 낮았다. 이러한 연구 결과는 영어교과서의 진정성이 고려되어 영어교과서가 개발되어야 함을 시사한다.

김정렬과 천윤희(2008)는 초등학교 6학년 영어교과서와 중학교 1학년 영어교과서의 연계성을 연구하였다. 구체적으로 김정렬과 천윤희는 각 학년에 대한 코퍼스 자료를 구축하여 NLPTools를 이용해 두 학년 간의 연계성을 비교 분석하였다. 그들은 어절수, 어휘수, 신출 어휘수, 어휘빈도수, 문장의 통사구조를 비교 분석하였다. 분석 결과, 초등학교 6학년 영어교과서에 비해 중학교 1학년 영어교과서에서 어휘와 신출어휘가 급격하게 증가됨이 관찰되었다. 또한 초등학교 6학년 영어교과서에 비해 중학교 1학년 영어교과서에서 문장의 통사적 복잡성 역시 급격하게 증가됨이 보고되었다. 이러한 결과는 중학교 영어교과서 개발 시 초등학교 영어교과서의 수준을 고려해야 함을 시사한다.

WordSmith Tools와 NLP Tools과 같은 용어색인 프로그램에 기반한 이러한 영어교과서 코퍼스 분석 연구들은 객관화된 측정치를 제공하는 컴퓨터 분석 도구와 기존의 선행연구들에 비해서 상대적으로 큰 크기의 코퍼스를 이용해 좀 더

객관적이고 일반화의 가능성이 높은 결과를 제공했다는 점에서 의의를 찾을 수 있겠다. 그러나 이러한 연구들에 적용된 용어색인 프로그램들은 주로 단어 및 특정 표현을 검색하는 용어색인 기능이 주기능이기 때문에 다양한 언어학 및 심리언어학적 측정치에 기초해 영어 교재를 분석하는데 적용하기에는 한계점이 수반된다(전문기, 2011; 전문기, 2014). 이러한 한계점을 극복할 수 있는 대안으로 최근에 자동화된 언어분석 프로그램인 Coh-Metrix를 활용한 다양한 영어교과서 분석 연구들이 수행되었다(김정렬, 2014; 김정렬, 양지윤, 2012; 김지은, 전문기, 2013; 성일호, 2014; 송해성, 2013; 전문기, 2011; 전문기, 2014; 전문기, 임인재, 2010; 정미경, 김남국, 2014; 이승환, 2013; 이정민, 현대덕, 2013; 이화자, 2011; 임인재, 전문기, 2013).

2. 자동화된 언어분석 프로그램 기반 영어교과서 분석 연구

본 연구의 주제와 관련하여, 김정렬(2014)은 최근 자동화된 언어분석 프로그램인 Coh-Metrix를 활용하여 초등학교 4, 5, 6학년 영어교과서와 중학교 1학년 영어교과서 사이의 연계성을 분석하였다. 김정렬의 연구 결과, 초등학교 4, 5, 6학년 영어교과서와 중학교 1학년 영어교과서의 간의 연계성은 단어 수와 길이, 어휘다양성 측정치, 단어빈도수, 습득나이 측정치, 대명사 전체(1인칭, 2인칭, 3인칭 모두 포함) 측정치에 대해 전반적으로 연계성이 비교적 잘 성립되어 있음이 시사되었다. 그러나, 구체성, 심상성과 같은 어휘측정치, 동일지시 응집성(co-reference cohesion), 의미적 정합성(semantic cohesion), 문장수, 문장길이, 읽기가독성 지표 점수, 통사적 복잡성 측정치에 대해서는 연계성이 잘 성립되지 않았음이 제시되었다. 이러한 결과는 영어교과서 집필 시 초등학교와 중학교 간의 연계성이 고려되어야함을 시사한다.

김지은과 전문기(2013)는 Coh-Metrix를 활용하여 중학교 영어교과서 듣기 지문 자료의 연계성을 분석하였다. 그들의 연구 결과, 중학교 영어교과서 듣기 자료의 연계성은 단어수, 문장수, 단어빈도수, 읽기가독성지표, 주어구 밀도(본동사 앞 단어수) 측정치에 대해서 잘 성립되어 있음이 제시되었다. 반면 명사구 밀도, 어휘다양성, 동일지시 정합성, 의미적 정합성 측정치에 대해서는 전반적으로 연계성이 잘 통제되지 않았음이 보고되었다. 이러한 연구들은 영어교육 학습현장에서 효과적인 영어학습을 위해 필요한 영어교과서 개발 시 학년 간의 연계성이 잘 고려되어 영어교과서가 집필되어야 함을 시사한다.

본 연구의 주제와 관련하여 다른 측면에서, 이화자(2011)는 Coh-Metrix를 활용하여 고등학교 영어교과서 읽기지문 자료에 대한 코퍼스를 구축하여 영어교과서의 자연스러움에 대해 분석하였다. 이화자는 외국 영어서적 자료에 대한 코퍼스를 구축하여 고등학교 영어교과서와 비교 분석하였다. 분석 결과, 고등학교 영어교과서의 동일지시 정합성과 의미적 정합성 측정치는 외국 영어서적의 측정치와 유사하였다. 그러나, 고등학교 영어교과서에 대한 기초산출치, 단어빈

도수, 표준가독성 점수, 통사적 복잡성, 어휘다양성 측정치 점수는 외국 영어서적의 측정치 점수와 다름이 보고되었다. 이러한 결과는 학습자가 영어 원어민의 자연스러운 표현을 배울 수 있도록 영어교재 집필 시 외국 영어서적과 차이가 나는 기초산출치, 단어빈도수, 표준가독성 점수, 통사적 복잡성, 어휘다양성과 같은 측정치들을 고려해야 함을 제시한다.

전문기(2011)는 Coh-Metrix를 활용하여 중학교 1학년과 2학년 영어교과서 사이의 연계성을 분석하였다. 구체적으로 중학교 1학년 영어교과서 19종과 중학교 2학년 영어교과서 19종에 대한 코퍼스를 구축하여 Coh-Metrix를 이용해 연계성을 분석하였다. 분석 결과, 두 학년 간의 연계성은 기초산출치, 표준 읽기 가독성지표 점수, 주어구 밀도(본동사 앞 단어수), 접속사, 심상성, 구체성, 습득 나이 점수에 대해 잘 통제되어 있음이 보고되었다. 그러나, 단어빈도수, 명사구 밀도, 동일지시 및 의미적 정합성, 대명사 측정치에 대해서는 두 학년 간의 연계성이 잘 성립되어 있지 않음이 제시되었다. 본 연구는 전문기(2011)의 연구를 확장하여 중학교 1, 2, 3학년 영어교과서 간의 총체적인 연계성을 자동화된 언어 분석 프로그램인 개인 PC용 Coh-Metrix를 활용해 분석하였다. 두 연구는 Coh-Metrix를 활용하여 2009년에 개정된 중학교 영어교과서의 학년 간 연계성을 총 22개의 Coh-Metrix 측정치에 대해 분석했다는 점에서 공통점이 있다. 그러나, 본 연구에서는 중학교 3학년 영어교과서 15종에 대한 코퍼스 자료를 구축하여 중학교 1, 2, 3학년 영어교과서의 학년 간 연계성을 총체적으로 분석하였기 때문에 이전 영어교과서 분석 연구와 차이점이 있다. 따라서 본 연구에서는 중학교 1학년 영어교과서 25종 중 15종, 중학교 2학년 영어교과서 19종 중 15종, 중학교 영어교과서 15종 중 15종에 대해 총 45개의 코퍼스 파일을 구축하여 학년 간 연계성 분석을 수행하였다. 각 학년 별로 총 15개의 영어교과서 코퍼스 파일을 구축한 이유는 중학교 3학년 영어교과서가 총 15개의 출판사에 의해서만 출간되었기 때문이었다. 따라서 본 연구에서는 출판사에 의한 연계성 효과를 통제하기 위해 동일 출판사에 의해서만 출간된 영어교과서만을 분석 대상으로 선정하였다.

III. 자동화된 언어분석 프로그램: Coh-Metrix

본 연구에 적용된 자동화된 언어분석 프로그램인 Coh-Metrix는 미국 뎀피스 대학교 지능형시스템연구소에서 영어교재 및 텍스트의 난이도와 이해도를 평가하기 위해 개발되었다(Graesser, et al., 2004). Coh-Metrix는 WordSmith Tools나 NLPTools와 같은 용어색인 기능 기반 프로그램과 달리 광범위한 유형의 언어학 및 심리언어학 측정치에 기초해 영어교재 및 텍스트의 난이도와 이해도를 분석한다(Graesser, et al., 2007; Graesser, et al., 2004). Coh-Metrix는 광범위한 유형의 언어학 및 심리언어학적 측정치를 산출하기 위해 전산언어학에서 제시된 다양

한 컴퓨터 알고리즘을 포함한다. 구체적으로 텍스트 내의 문장의 통사 구조와 어휘 품사정보를 처리하기 위해 자동화된 통사구문분석기와 품사분석기(Brill, 1995)를 이용한다(전문기, 2014). 그리고 인접한 문장 사이의 의미적 관계를 분석하기 위해 LSA(Latent Semantic Analysis) 알고리즘을 이용한다(Landauer, Foltz, & Laham, 1998). 이러한 시스템 외에 다양한 어휘집(lexicons), 코퍼스 준거 및 다양한 유형의 전산언어학적 알고리즘(Jurafsky & Martin, 2008)을 포함한다(전문기, 2011; 전문기, 2014).

개인 PC용 Coh-Metrix는 여러 언어심리학자들의 실험을 통해 검증된 텍스트의 이해도와 정합성(cohesion)에 영향을 준다고 밝혀진 광범위한 유형의 언어학 및 심리언어학적 측정치(Graesser, et al., 2007; Graesser, et al., 2004; Halliday & Hasan, 1976)를 제공하는데 본 연구에서는 중학교 1, 2, 3학년 영어교과서의 총체적인 연계성을 분석하기 위해 총 22개의 측정치를 선택하였다. 이러한 측정치는 기초산출치(단어수, 문장수, 평균문장길이), 단어빈도수, 표준가독성지표(Flesch Reading Ease score, Flesch-Kincaid Grade Level), 통사적 복잡성(주어부 밀도, 명사구 밀도), 어휘다양성(타입-토큰 비율), 동일지시 정합성(인접된 문장 간의 논항 반복 비율), 의미적 정합성(LSA 측정치), 접속사(인과적, 부가적, 시간적 접속사), 대명사 정보(전체 대명사 비율, 1인칭, 2인칭, 3인칭 대명사), 어휘정보(심상성, 구체성, 습득나이)를 포함한다.

1. 기초산출치

Coh-Metrix 시스템의 기초산출치는 단어수, 문장수, 평균문장길이를 포함한다. 기초산출치는 문장의 이해도와 난이도에 영향을 미치는데 구체적으로 단어수와 문장수가 많고 문장의 길이가 길수록 텍스트의 이해도와 난이도는 증가한다(전문기, 2011; 전문기, 2014). 따라서 본 연구에서는 중학교 1, 2, 3학년 영어교과서의 학년 간 연계성을 분석하기 위해 기초산출치를 분석하였다.

2. 단어빈도수

단어빈도수는 문장의 이해도와 난이도에 영향을 미치는데 고빈도 단어보다 저빈도 단어를 많이 포함하고 있는 텍스트가 이해하기 어려운 경향이 있다(전문기, 2011; 전문기, 2014). Coh-Metrix는 네덜란드 어휘정보센터에서 제공하는 CELEX 어휘 데이터베이스(Baayen, Piepenbrock, & Gulikers, 1995)를 이용해 단어빈도수를 산출한다. 본 연구에서는 일종의 로그함수에 의해서 대수변환된 단어빈도수를 이용하였다. 이러한 이유는 대수변환된 단어빈도수 측정치가 읽기 난이도를 측정하는데 더 적합하다고 알려져 있고, 큰 수치로 나타나는 단어빈도수를 로그함수 특성 상 작은 수치로 변환하여 제시함으로써 정보 가독성을 높이기 위해서였다(전문기, 2011). 본 연구에서는 중학교, 1, 2, 3학년 영어교과서의

연계성을 분석하기 위해 교재의 읽기 난이도에 영향을 줄 수 있는 단어빈도수를 이용하였다.

3. 표준가독성 읽기 지표

표준가독성 읽기 지표는 교육학 분야 등에서 교재나 텍스트의 읽기 난이도를 평가하기 위해 많이 사용된다(전문기, 2011). Coh-Metrix는 Flesch Reading Ease(이하 FRE) 점수와 Flesch-Kincaid Grade Level(이하 FKGL) 점수를 제공한다. FRE 점수는 0과 100사이의 점수로 표현되는데 이 점수가 높은 텍스트는 쉽게 이해된다. 반면 FKGL 점수는 0과 12사이의 점수로 표현되는데 이 점수가 높은 텍스트는 이해하는데 어렵다. FKGL의 점수는 미국 초중등 학년 등급(1학년-12학년)을 나타낸다. 본 연구에서는 중학교 영어교과서 1, 2, 3학년 간 연계성을 분석하기 위해 표준가독성 지표를 이용하였다.

4. 통사적 복잡성

텍스트 내 문장의 통사적 복잡성은 텍스트의 이해도와 난이도에 영향을 미친다. 일반적으로 문장의 통사적 구조가 복잡한 텍스트가 이해하는데 어렵다(전문기, 2011). Coh-Metrix의 통사적 복잡성 측정치는 주어부 밀도와 명사구 밀도로 구성된다. 주어부 밀도 점수는 본동사 앞 단어수를 지칭한다. 명사구 밀도 점수는 특정 명사구 내의 수식어의 수를 그 명사구 내의 전체 단어수로 나눈 값을 나타낸다. 이러한 통사적 복잡성 측정치는 Coh-Metrix의 통사구문분석기와 품사분석기에 의해서 산출된다(Brill, 1995).

5. 어휘다양성 측정치

어휘다양성 측정치는 타입-토큰 비율을 나타낸다. 타입은 텍스트 내에 포함된 개개의 어휘를 지칭한다(Graesser, et al., 2004). 따라서 타입 값은 항상 1이다. 그리고 토큰은 그 개개의 어휘가 텍스트 내에서 반복되어 나타나는 빈도수를 지칭한다. 만약 텍스트 내에 ‘코퍼스’라는 단어가 10번 반복되어 나타나면 ‘코퍼스’의 타입 값은 1이고 토큰 값은 10이 된다. 따라서 타입-토큰 비율은 0.1이다. 일반적으로 이 타입-토큰 비율의 값이 클수록 텍스트 내에 다양한 어휘가 포함되어 있을 가능성이 높아지므로 텍스트의 이해도는 증가한다고 볼 수 있다(전문기, 2011; 전문기, 2014). 본 연구에서는 중학교 1, 2, 3학년 영어교과서의 연계성을 분석하기 위해 어휘다양성 측정치를 활용하였다.

6. 접합성(cohesion) 측정치

텍스트의 정합성은 텍스트의 이해도에 영향을 미치는 중요한 요소이다(Halliday & Hasan, 1976; Haviland & Clark, 1974). 구체적으로 텍스트의 정합성이 높을수록 그 텍스트를 이해하는데 용이한 경향이 있다(Haviland & Clark, 1974). Coh-Metrix의 정합성 측정치는 동일지시 정합성과 의미적 정합성 측정치로 구성된다. 동일지시 정합성은 측정치는 인접한 문장 사이의 논항반복 비율을 나타낸다. 일반적으로 문장 사이에 논항이 반복되어 위치할 때 그 문장 사이의 정합성은 증가한다(Haviland & Hasan, 1974). 의미적 정합성 측정치는 인접한 문장 사이의 LSA 코사인(Latent Semantic Analysis cosine) 값을 나타낸다. LSA는 두 텍스트 구성 요소(단어, 문장) 사이의 의미적 유사성을 산출하는데 이용되는 수학적 알고리즘이다(Landauer, Foltz, & Laham, 1998). 요약하면 높은 정합성의 텍스트가 이해하는데 용이하다고 볼 수 있다(Haviland & Clark, 1974). 본 연구에서는 교재의 심층적 특성을 반영하는 정합성 측정치에 기초해 중학교 1, 2, 3학년 영어교과서의 학년 간 연계성을 분석하였다.

7. 접속사(connective) 측정치

접속사는 문장의 이해도에 영향을 미친다. 일반적으로 접속사는 문장의 이해도를 촉진하는 경향이 있다(Halliday & Hasan, 1976). Coh-Metrix는 인과적(therefore, because 등), 부가적(and, moreover 등), 시간적(until, before 등) 접속사 측정치를 제공한다. 접속사 측정치는 발생점수(incidence score)로 산출된다(Graesser, McNamara, Louwerse, & Cai, 2004). 발생점수는 텍스트의 길이에 의한 효과를 통제하기 위한 값으로 1,000 단어마다 나타나는 특정 어휘의 비례 점수이다. 본 연구에서는 중학교 1, 2, 3학년 영어교과서의 연계성을 체계적으로 분석하기 위해 텍스트의 이해도와 난이도에 영향을 줄 수 있는 접속사 측정치를 이용하였다.

8. 대명사(pronoun) 측정치

텍스트 내의 대명사는 문장의 이해도와 난이도에 영향을 미친다(Halliday & Hasan, 1976). 대명사는 일반적으로 선행 문장의 정보를 지칭하기 때문에 텍스트의 이해도를 촉진시킬 수 있다. 그러나 텍스트 내에 너무 많은 대명사가 있을 경우 한정된 작업 기억(working memory) 내에서 그 대명사의 지칭 대상을 일일이 탐색해야 하기 때문에 텍스트의 이해 난이도가 증가할 수 있다(전문기, 2011). Coh-Metrix는 1인칭 대명사(i.e., I, we), 2인칭 대명사(i.e., you), 3인칭 대명사(i.e., it, she, he, they)에 대한 발생점수를 제공한다. 그리고 전체 대명사에 대한 비율 점수를 제공하는데 이 점수는 텍스트 내에 포함된 전체 대명사의 수를 그

텍스트 내의 전체 명사구의 수로 나눈 값을 나타낸다(Graesser, et al., 2004). 본 연구에서는 텍스트의 난이도에 영향을 줄 수 있는 대명사 측정치를 이용하여 중학교 1, 2, 3학년 영어교과서의 학년 간 연계성을 분석하였다.

9. 어휘 측정치

텍스트 내의 각각의 어휘 특성 정보는 텍스트의 이해도에 영향을 미친다. Coh-Metrix는 MRC Psycholinguistic 데이터베이스(Coltheart, 1981)를 이용해 어휘의 심상성(imageability), 구체성, 습득나이 점수를 산출한다. 심상성, 구체성, 습득나이 어휘 점수는 100과 700사이의 점수의 나타나는데 심상성과 구체성의 경우는 수치가 높을수록 이해하기 쉬운 어휘를 지칭한다. 반면 습득나이 수치는 수치가 높을수록 높은 연령대에서 학습하는 단어를 나타내기 때문에 상대적으로 어려운 어휘를 지칭한다(Graesser, McNamara, Louwerse, & Cai, 2004).

IV. 연구방법

1. 연계성 분석 대상 중학교 영어교과서

본 연구에서는 중학교 1, 2, 3학년 영어교과서의 총체적인 연계성을 분석하기 위해 총 45종의 영어교과서를 선택하였다. 구체적으로 2009년에 개정된 중학교 1학년 영어교과서 25종 중 15종, 중학교 2학년 영어교과서 19종 중 15종, 그리고 중학교 3학년 영어교과서 15종 중 15종을 분석대상으로 선택하였다. 중학교 1학년과 2학년 영어교과서에서 각각 15종을 선택한 이유는 각 학년의 교과서 저자와 출판사를 일치시킴으로써 분석의 신뢰성과 타당성을 높이기 위해서였다. 본 연구에서 분석된 총 45종의 영어교과서 목록은 표 1에 제시되었다.

표 1

분석대상 중학교 영어교과서 목록

중학교 영어교과서	출판사
Middle School English 1, 2, 3	고려출판
Middle School English 1, 2, 3	교학사
Middle School English 1, 2, 3	금성출판사
Middle School English 1, 2, 3	능률교육
Middle School English 1, 2, 3	대교
Middle School English 1, 2, 3	미래엔컬처그룹
Middle School English 1, 2, 3	비상교육
Middle School English 1, 2, 3	에듀왕
Middle School English 1, 2, 3	좋은책신사고
Middle School English 1, 2, 3	지학사
Middle School English 1, 2, 3	천재교육
Middle School English 1, 2, 3	천재교육
Middle School English 1, 2, 3	천재교육
Middle School English 1, 2, 3	두산동아
Middle School English 1, 2, 3	두산동아

2. 중학교 영어교과서 코퍼스 구축과 분석 도구

본 연구에서는 자동화된 언어분석 프로그램인 개인 PC용 Coh-Metrix를 활용하여 중학교 영어교과서의 연계성을 분석하기 위해 총 45종의 영어교과서 파일을 이용하였다. 개인 PC용 Coh-Metrix는 유니코드 형식의 텍스트 파일(txt 파일)만 처리가능 하므로 본 연구에서는 개인 PC용 Coh-Metrix를 이용하기 위해 45종의 영어교과서 각각에 대한 총 45개의 텍스트 파일을 생성하였다. 이렇게 형성된 텍스트 파일을 개인 PC용 Coh-Metrix를 이용하여 분석하였다. 본 연구에서는 총 22개의 Coh-Metrix 측정치를 분석하였는데 이러한 측정치는 기초산출치(단어수, 문장수, 평균문장길이), 단어빈도수, 표준가독성지표(Flesch Reading Ease score, Flesch-Kincaid Grade Level), 통사적 복잡성(주어부 밀도, 명사구 밀도), 어휘다양성(타입-토큰 비율), 동일지시 정합성(인접된 문장 간의 논항 반복 비율), 의미적 정합성(LSA 측정치), 접속사(인과적, 부가적, 시간적 접속사), 대명사 정보(전체 대명사 비율, 1인칭, 2인칭, 3인칭 대명사), 어휘정보(심상성, 구체성, 습득나이)를 포함하였다.

V. 연구결과

1. 기초산출치

기초산출치에 대한 일원분산분석 통계분석 결과는 표 2에 제시되었다. 표 2에 제시된 것처럼 중학교 1, 2, 3학년 영어교과서는 단어수, 문장수, 평균문장길

이에 대해 학년 간에 통계적으로 유의미한 차이를 보였다. 전체적으로 학년이 올라갈수록 단어수와 문장수가 증가하고 문장길이가 증가하는 경향을 보였다. 구체적으로 어느 학년 간에 유의미한 차이가 있는지를 검증하기 위해 Tukey 사후검정을 실시하였다. 사후검정 결과는 표 3에 제시되었다.

표 2

기초산출치 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
단어수	1928(233)	3118(945)	5266(761)	84.313	.000***
문장수	259(37)	337(107)	489(75)	32.944	.000***
평균문장길이	7.5(.83)	9.4(1.36)	10.8(.91)	37.164	.000***

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 3에 제시된 것처럼 단어수, 문장수, 평균문장길이에 대해 각 학년 간에 모두 통계적으로 유의미한 차이를 보였다. 이러한 결과는 중학교 1, 2, 3학년 영어 교과서가 단어수, 문장수, 평균문장길이에 대해 연계성이 잘 성립되어 있음을 제시한다. 이러한 점은 영어교과서 집필자들이 단어수, 문장수, 문장길이와 같은 측정치를 고려하면서 학년에 따른 영어교과서의 난이도를 비교적 잘 통제하였음을 제시한다.

표 3

기초산출치 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
단어수	$p = .000***$	$p = .000***$	$p = .000***$
문장수	$p = .025*$	$p = .000***$	$p = .000***$
평균문장길이	$p = .000***$	$p = .000***$	$p = .002**$

* $p < .05$, ** $p < .01$, *** $p < .0001$

2. 단어빈도수

단어빈도수에 대한 연계성 분석 결과는 표 4에 제시되었다. 표 4에 제시된 것처럼 단어빈도수에 대해 전체 학년 간에 통계적으로 유의미한 차이를 보였다. Tukey 사후검정 결과는 표 5에 제시되었다. 표에 제시된 단어빈도수 수치는 전술한 것처럼 로그함수에 의해 대수변환된 측정치를 의미한다. 이러한 대수변환된 측정치는 읽기난이도를 측정하는데 더 적합하다고 알려져 있기 때문에 본 연구에서는 대수변환된 단어빈도수를 연계성 분석에 이용하였다(전문기, 2011).

표 4

단어빈도수 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
단어빈도수	2.51(.05)	2.52(.05)	2.45(.05)	6.853	.003**

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 5의 결과를 살펴보면, 중학교 1학년과 2학년간에는 통계적으로 유의미한 차이를 보이지 않았다. 그러나 중학교 2학년과 3학년 사이에는 통계적으로 유의미한 차이를 보였다. 이러한 결과는 중학교 1학년과 2학년 사이의 연계성이 단어빈도수에 대한 적절하게 잘 통제되지 않았음을 나타내며 차후 교과서 집필 시 단어빈도수가 고려되어 연계성이 통제되어야 함을 시사한다. 영어교과서 개발 시 개개 단어의 단어빈도수를 고려하면서 교과서를 집필하는 것은 어려운 일이다. Coh-Metrix와 같은 자동화된 언어분석 프로그램이 교재 집필 시 교재의 난이도를 수시로 측정하는데 활용되기를 기대한다.

표 5

단어빈도수 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
단어빈도수(내용어)	$p = .907$	$p = .013^*$	$p = .004^{**}$

* $p < .05$, ** $p < .01$, *** $p < .0001$

3. 표준가독성(standard readability) 읽기 지표

표준가독성 지표 점수에 대한 통계분석 결과는 표 6에 제시되었다. 표 6에 제시된 것처럼 중학교 1, 2, 3학년 영어교과서는 FRE 점수와 FKGL 점수에 대해 전체 학년 간에 통계적으로 유의미한 차이를 보였다.

표 6

표준가독성지표 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
FRE	88(2.5)	85(3.9)	80(3.8)	22.960	.000***
FKGL	2.8(.5)	3.7(.8)	4.9(.7)	32.245	.000***

* $p < .05$, ** $p < .01$, *** $p < .0001$

표준가독성 지표 점수에 대한 Tukey 사후검정 결과는 표 7에 제시되었다. 표 7에 제시된 것처럼, 중학교 1, 2, 3학년 영어교과서는 FKGL 점수에 대해 각 학년 간에 유의미한 차이를 보였다. 이러한 결과는 학년이 올라갈수록 읽기 난이도가 증가하여 연계성이 잘 성립되어 있음을 제시한다. 그러나 중학교 2, 3학년 영어교과서는 FRE 점수에 대해 통계적으로 유의미한 차이를 보였지만 중학교

1, 2학년 간에는 통계적으로 거의 유의미한 차이를 보였다. 이러한 결과는 전반적으로 학년이 올라갈수록 읽기 난이도가 증가함을 시사한다고 하겠다. 일반적으로 표준가독성 지표 측정치는 단어길이에 문장길이에 민감하다. 다시 말해서 단어와 문장이 길이가 길어지면 FKGL 점수는 높아지고 FRE 점수는 낮아지는 경향이 있다. 일반적으로 단어와 문장의 길이가 길어지면 읽기난이도는 증가하는 경향이 있다. 그러나, 단어와 문장의 길이가 길다고 항상 교재의 읽기난이도가 증가하는 것은 아닐 수 있다. 따라서 영어교과서의 전체적인 이해도와 난이도를 평가하기 위해서는 아래 제시된 정합성, 접속사, 대명사와 같은 다른 측정치들을 총체적으로 함께 고려해야 할 것이다.

표 7

표준가독성지표 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
FRE	$p = .056$	$p = .000***$	$p = .000***$
FKGL	$p = .003**$	$p = .000***$	$p = .000***$

* $p < .05$, ** $p < .01$, *** $p < .0001$

4. 통사적 복잡성

통사적 복잡성 측정치에 대한 분석 결과는 표 8에 제시되었다. 표 8에 제시된 것처럼 중학교 1, 2, 3학년 영어교과서는 주어부 밀도와 명사구 밀도 점수에 대해 전체 학년 간에 통계적으로 유의미한 차이를 보였다. Tukey 사후검정 결과는 표 9에 제시되었다.

표 8

통사적 복잡성 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
주어부밀도	1.6(.21)	2.2(.29)	2.6(.26)	47.181	.000***
명사구밀도	.66(.05)	.66(.06)	.71(.06)	3.921	.027*

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 9

통사적 복잡성 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
주어부밀도	$p = .000***$	$p = .000***$	$p = .001**$
명사구밀도	$p = .993$	$p = .057$	$p = .045*$

* $p < .05$, ** $p < .01$, *** $p < .0001$

Tukey 사후검정 결과 표 9에 제시된 것처럼, 중학교 1, 2, 3학년 영어교과서는

주어부 밀도 점수에 대해 각 학년 간에 통계적으로 유의미한 차이를 보였다. 구체적으로 학년이 올라갈수록 주어부가 통사적으로 복잡해지는 경향을 보였는데 이러한 결과는 이 측정치에 대해 중학교 1, 2, 3학년 영어교과서의 연계성이 잘 통제되었음을 제시한다. 반면 명사구 밀도 점수는 중학교 2, 3학년 간에는 통계적으로 유의미한 차이를 보였지만 중학교 1, 2학년 간에는 유의미한 차이를 보이지 않았다. 이러한 결과는 연계성 고려 시 문장 전체의 구조뿐만 아니라 좀 더 작은 단위인 명사구의 구조도 통제하여야 함을 시사한다. 명사구의 구조와 같은 통사적 복잡성 측정치를 의식하면서 교재를 집필하는 것은 어려운 일이다. 교재 집필 시 Coh-Metrix가 문장의 통사구조를 측정하는데 활발하게 이용되기를 기대한다.

5. 어휘다양성 측정치

어휘다양성 분석 결과는 표 10에 제시되었다. 표 10에 제시된 것처럼 중학교 1, 2, 3학년 영어교과서는 타입-토큰비율 값에 대해 전체 학년 간에 통계적으로 유의미한 차이를 보였다. 사후검정 결과는 표 11에 제시되었다.

표 11에 제시된 결과, 중학교 1, 2학년 영어교과서는 타입-토큰비율 값에 대해 통계적으로 유의미한 차이를 보였다. 그러나 2, 3학년 간에는 유의미한 차이를 보이지 않았다.

어휘다양성의 측정치인 타입-토큰 비율 값은 클수록 특정 텍스트가 다양한 어휘를 포함할 가능성이 증가함을 나타내는데 흥미롭게도 중학교 1학년 영어교과서의 타입-토큰 비율 값이 가장 높았다. 이러한 결과는 중학교 1학년 영어교과서에 상대적으로 학습해야 할 많은 어휘가 제시되어 학습자의 학습 동기를 저하시킬 가능성이 있음을 시사한다. 따라서 차후 교재 집필 시 각 학년 간에 학습해야 할 어휘의 수를 점차적으로 증가시킬 필요가 있겠다.

표 10

어휘다양성 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
타입-토큰비율	.51(.04)	.46(0.7)	.43(.04)	9.559	.000***

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 11

어휘다양성 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
타입-토큰비율	$p = .024^*$	$p = .000^{***}$	$p = .267$

* $p < .05$, ** $p < .01$, *** $p < .0001$

6. 정합성 측정치

정합성 측정치에 대한 분석 결과는 표 12에 제시되었다. 표 12에 제시된 것처럼 중학교 영어교과서 1, 2, 3학년 영어교과서는 동일지시 및 의미적 정합성 측정치에 전체 학년 간에 통계적으로 유의미한 차이를 나타냈다. Tukey 사후검정 결과는 표 13에 제시되었다.

표 12
정합성 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
논항반복비율	.37(.05)	.42(.04)	.45(.04)	15.653	.000***
LSA 측정치	.16(.02)	.19(.03)	.19(.02)	8.322	.001**

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 13에 제시된 것처럼, 중학교 1, 2학년 영어교과서는 두 정합성 측정치에 대해 통계적으로 유의미한 차이를 보였다. 그러나 중학교 2, 3학년 영어교과서는 두 정합성 측정치에 대해 유의미한 차이를 보이지 않았다. 이러한 결과는 차후 교재 집필 시 학년에 따라 문장 간의 정합성이 좀 더 엄격하게 통제되어야 함을 제시한다. 심리언어학적으로 인접된 문장 사이의 정합성이 증가하면 교재의 이해도가 촉진되면서 교재의 난이도가 감소하는 경향이 있다. 따라서 교재 집필 시 교재의 목적과 유형에 따라 정합성과 같은 측정치를 잘 통제하는 것이 중요하다. 예를 들면 교재의 난이도를 높이기 위해서는 인접된 문장 간의 정합성을 낮게 조작하는 것이 필요하다. 문장 사이의 정합성이 낮으면 독자는 추론과 같은 전략을 활용하여 교재를 이해해야 하기 때문이다. 따라서 교재 개발 시 학년이 올라갈수록 일반적으로 문장 간의 연결이 잘 되도록 교재를 집필하는 것도 중요하지만 교재의 읽기 난이도를 조작하기 위해 필요할 경우 의식적으로 인접된 문장 사이의 정합성을 낮추어야 할 필요도 있을 것이라 생각된다. 그러나, 문장의 정합성과 같은 심리언어학적 측정치를 계산하면서 교재를 집필하는 것은 어려운 일이다. Coh-Metrix와 같은 자동화된 도구가 교재 집필 시 문장의 심층구조를 파악하는데 활용되기를 기대한다.

표 13
정합성 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
논항반복비율	$p = .003^{**}$	$p = .000^{***}$	$p = .124$
LSA 측정치	$p = .006^{**}$	$p = .001^{**}$	$p = .860$

* $p < .05$, ** $p < .01$, *** $p < .0001$

7. 접속사 측정치

접속사 측정치는 표 14에 제시되었다. 표 14에 제시된 것처럼 중학교 1, 2, 3학년 영어교과서는 인과적, 부가적, 전체 접속사 측정치에 대해 학년 간에 통계적으로 유의미한 차이를 보였다. 그러나 시간적 접속사 측정치에 대해서는 학년 간에 유의미한 차이를 보이지 않았다. Tukey 사후검정 결과는 표 15에 제시되었다. 먼저 중학교 1, 2학년 영어교과서는 인과적 접속사에 대해 통계적으로 유의미한 차이를 보여주었다. 그러나 중학교 2, 3학년 간에는 유의미한 차이를 보이지 않았다. 반면 중학교 1, 2학년 영어교과서는 부가적 접속사에 대해 통계적으로 거의(marginally) 유의미한 차이를 보였다. 그러나 중학교 2, 3학년 간에는 유의미한 차이를 보이지 않았다. 반면 시간적 접속사에 대해서는 학년 간에 통계적으로 유의미한 차이가 발견되지 않았다. 접속사 전체에 대해서는 중학교 1, 2학년 영어교과서 간에 통계적으로 유의미한 차이가 발견되었고 중학교 2, 3학년 영어교과서 간에는 유의미한 차이가 발견되지 않았다. 이러한 결과는 교과서 집필 시 접속사가 체계적으로 통제되지 않았음을 제시하며 차후 교과서 개발 시 접속사 항목이 고려되어야 함을 시사한다고 하겠다. 접속사는 일반적으로 문장의 연결을 촉진하는 효과가 있기 때문에 교재의 전체적인 정합성을 통제하는데 활용될 수 있다. 접속사로 연결된 문장의 통사구조는 접속사로 인해 복잡해 질 수 있지만 교재 내에 접속사가 많을 경우 전체적인 이해도는 증가할 수 있다. 따라서 교재의 학년 간 연계성을 고려하면서 교재의 난이도를 점진적으로 증가시키기 위해서는 접속사의 수를 효과적으로 잘 통제해야 할 필요가 있겠다.

표 14
접속사 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
인과적 접속사	20.2(4.7)	25.2(4.2)	27.1(3.3)	11.299	.000***
부가적 접속사	8.0(1.7)	9.5(2.0)	9.7(1.6)	4.248	.021*
시간적 접속사	23.2(6.1)	21.2(4.0)	20.7(2.9)	1.357	.268
접속사 전체	61.7(8.5)	79.0(9.2)	84.2(6.5)	31.507	.000***

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 15
접속사 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
인과적 접속사	$p = .005^{**}$	$p = .000^{***}$	$p = .407$
부가적 접속사	$p = .056$	$p = .030^{*}$	$p = .960$
시간적 접속사	$p = .430$	$p = .277$	$p = .951$
접속사 전체	$p = .000^{***}$	$p = .000^{***}$	$p = .195$

* $p < .05$, ** $p < .01$, *** $p < .0001$

8. 대명사 측정치

대명사 분석 결과는 표 16에 제시되었다. 표 16에 제시된 것처럼 중학교 1, 2, 3학년 영어교과서는 대명사 비율과 1인칭 대명사 측정치에 대해 통계적으로 유의미한 차이를 나타냈었다. 그러나 2인칭, 3인칭 대명사에 대해서는 전체 학년 간에 유의미한 차이를 보이지 않았다. Tukey 사후검정 결과는 표 17에 제시되었다. 표 17에 제시된 것처럼 중학교 1, 2 학년 영어교과서는 대명사 비율과 1인칭 대명사 측정치에 대해 통계적으로 유의미한 차이를 보이지 않았다. 그러나 중학교 2, 3 학년 간에는 유의미한 차이가 발견되었다. 그리고 2인칭, 3인칭 대명사에 대해서는 각 학년 간에 통계적으로 유의미한 차이가 발견되지 않았다. 이러한 결과는 전체적으로 교재 개발 시 대명사의 사용이 잘 통제되지 않았음을 제시한다고 하겠다. 차후 교재 개발 시 다양한 유형의 대명사가 전체적으로 고려되어 반영되기를 기대한다. 대명사는 일반적으로 교재의 전체적인 정합성을 증가시키는 경향이 있다. 대명사가 교재 내에 적절하게 사용될 경우 교재의 정합성은 증가하는 경향이 있다. 따라서 대명사의 사용에 따른 교재의 학년 간 연계성을 조작하기 위해 Coh-Metrix와 같은 자동화된 도구가 활발하게 이용되기를 기대한다.

표 16
대명사 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
대명사비율	.38(.04)	.38(.05)	.33(.05)	6.733	.003**
1인칭	27.5(5.0)	22.3(8.1)	16.1(5.2)	12.478	000***
2인칭	12.6(3.0)	13.8(4.5)	11.7(5.2)	.867	.428
3인칭	16.1(4.1)	17.0(5.9)	15.4(4.0)	.460	.634

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 17
대명사 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
대명사비율	$p = .941$	$p = .012^*$	$p = .005^{**}$
1인칭	$p = .071$	$p = .000^{***}$	$p = .025^*$
2인칭	$p = .753$	$p = .824$	$p = .395$
3인칭	$p = .848$	$p = .912$	$p = .608$

* $p < .05$, ** $p < .01$, *** $p < .0001$

9. 어휘 측정치

단어의 심사성, 구체성, 습득나이에 대한 어휘 측정치 분석 결과는 표 18에 제시되었다.

표 18

어휘정보 분석 결과

	1학년(n=15)	2학년(n=15)	3학년(n=15)	F	p
심상성	438(8.4)	427(8.8)	422(9.4)	13.164	.000***
구체성	401(9.5)	393(10.0)	390(10.6)	4.545	.016*
습득나이	265(8.4)	279(9.3)	303(11.3)	57.040	.000***

* $p < .05$, ** $p < .01$, *** $p < .0001$

표 18에 제시된 것처럼, 중학교 1, 2, 3학년 영어 교과서는 심상성, 구체성, 습득나이 측정치에 대해 전체 학년 간에 통계적으로 유의미한 차이를 보여주었다. Tukey 사후검정 결과는 표 19에 제시되었다.

표 19에 제시된 것처럼, 중학교 1, 2학년 영어교과서는 심상성에 대해 통계적으로 유의미한 차이를 보였다. 그러나 중학교 2, 3학년 간에는 유의미한 차이가 발견되지 않았다. 반면 구체성에 대해서는 중학교 1, 2학년 및 중학교 2, 3학년 간에 유의미한 차이가 발견되지 않았다. 그러나 습득나이에 대해서는 중학교 1, 2학년 및 중학교 2, 3학년 영어교과서 간에 통계적으로 유의미한 차이가 발견되어 이 측정치에 대해 연계성이 잘 성립되어 있음이 시사되었다. 전체적으로 교재 개발 시 심상성, 구체성, 습득나이와 같은 어휘 자체의 속성 정보가 반영되기를 기대한다. 교재 개발 시 개개 단어의 심상성, 구체성, 습득나이와 같은 속성을 고려하기는 어려운 일이다. 교재 개발 시 Coh-Metrix와 같은 언어분석 프로그램이 개개 단어의 속성을 측정하는데 유용하게 활용되기를 기대한다.

표 19

어휘정보 사후 검정 분석 결과

	1학년 vs. 2학년	1학년 vs. 3학년	2학년 vs. 3학년
심상성	$p = .003^{**}$	$p = .000^{***}$	$p = .311$
구체성	$p = .129$	$p = .014^*$	$p = .596$
습득나이	$p = .001^{**}$	$p = .000^{***}$	$p = .000^{***}$

* $p < .05$, ** $p < .01$, *** $p < .0001$

VI. 결론 및 제언

본 연구에서는 자동화된 언어분석 프로그램인 개인 PC용 Coh-Metrix를 활용하여 중학교 1, 2, 3학년 영어교과서 전체의 총체적인 연계성을 분석하였다. 본 연구에서는 총 22개의 Coh-Metrix 측정치를 교과서 연계성 분석에 적용하였다. 이러한 측정치는 기초산출치(단어수, 문장수, 평균문장길이), 단어빈도수, 표준가독성 읽기지표(FRE, FKGL), 통사적 복잡성(주어부 밀도, 명사구 밀도), 어휘다양성(타입-토큰 비율), 동일지시 정합성(인접된 문장 간의 논항 반복 비율), 의

미적 정합성(LSA cosine 측정치), 접속사(인과적, 부가적, 시간적 접속사), 대명사 정보(전체 대명사 비율, 1인칭, 2인칭, 3인칭 대명사), 어휘정보(심상성, 구체성, 습득나이)를 포함하였다.

본 연구의 결과 중학교 1, 2, 3학년 영어교과서의 학년 간 연계성은 기초산출치인 단어수, 문장수, 평균문장길이에 대해 잘 통제되어 있음이 제시되었다. 또한 통사적 복잡성 측정치인 주어부 밀도 점수에 대해서도 각 학년 간에 연계성이 잘 성립되어 있음이 제시되었다. 그러나 부가적 및 시간적 접속사, 2인칭 및 3인칭 대명사, 그리고 구체성 측정치에 대해서는 각 학년 간에 통계적으로 유의미한 차이가 발견되지 않아 이러한 측정치들에 대해 연계성이 잘 성립되지 않았음이 제시되었다. 또한 어휘다양성 측정치에 대해서는 중학교 1, 2학년 간에 유의미한 차이가 발견되었지만 오히려 중학교 1학년 영어교과서에 다양한 어휘가 더 많이 포함되어 연계성이 올바르게 잘 조작되지 않았음이 제시되었다. 나머지 다른 측정치들에 대해서는 부분적으로만 연계성이 확립되어 있음이 제시되었다. 구체적으로 중학교 1, 2학년 영어교과서는 동일지시 및 의미적 정합성, 인과적 접속사, 접속사 전체, 심상성 측정치에 대해서 통계적으로 유의미한 차이를 보였고 중학교 2, 3학년 영어교과서는 단어빈도수, FRE, 명사구밀도, 대명사 비율, 1인칭 대명사 측정치에 대해서 통계적으로 유의미한 차이를 보였다.

요약하면, 본 연구의 결과는 중학교 1, 2, 3학년 영어교과서 간의 연계성이 기초산출치와 주어부 밀도 점수를 제외하고 본 연구에서 분석된 다양한 다른 측정치에 대해 잘 성립되어 있지 않음을 보여주었다. 따라서 차후 교재 개발 시 본 연구에서 분석된 여러 유형의 언어학 및 심리언어학적 측정치가 교재의 이해도와 난이도를 평가하는데 하나의 표준적인 준거로 활용되기를 기대한다.

부연하면, 본 연구의 결과는 영어교과서 개발 시 그 동안 표준화된 준거의 부족으로 인해 영어교과서의 학년 간 연계성이 잘 고려되지 않았음을 시사한다. 따라서 Coh-Metrix와 같은 도구를 활용하여 영어교과서 개발 시 활용될 수 있을 것으로 기대되는 표준화된 준거를 구축하는 것이 필요할 것이라 생각된다. 구체적으로, 현재 초중등 학년에서 사용되는 영어교재 전체에 대한 방대한 크기의 학년 별 코퍼스 자료를 구축하여 Coh-Metrix를 통해 분석함으로써 학년 별 표준화된 준거 점수를 구축할 수 있을 것으로 기대한다. 이러한 학년 별 준거 점수에 기초해 개발 중인 영어교재의 난이도를 상대적으로 평가할 수 있을 것으로 기대한다.

예를 들면, 교재 내에 포함된 인접된 문장 사이의 정합성과 같은 심층수준의 측정치들은 교재 개발 시 Coh-Metrix와 같은 자동화된 언어분석 프로그램을 활용하지 않는다면 통제하는데 어려움이 있다. 따라서 교재 개발 시 수시로 집필된 지문 자료들을 Coh-Metrix를 통해 분석한 후 이러한 결과 자료를 차후 대용량의 코퍼스에 기반해 구축된 표준화된 학년 별 준거와 비교함으로써 집필 중인 교재의 상대적인 난이도를 평가할 수 있을 것으로 기대한다. 구체적으로 교재 개발 시 인접된 문장 사이의 정합성을 통제하기 위해 의식적으로 문장 간의

논항(말하자면, 명사, 대명사) 반복 비율을 조작할 수 있을 것으로 기대한다. 또한 개개 단어에 대한 심상성과 같은 어휘적 측정치나 문장 내에 포함된 대명사나 접속사와 같은 요소들도 텍스트의 이해도에 영향을 미치기 때문에 차후 교재 개발 시 Coh-Metrix를 통해 분석되어 반영되기를 기대한다.

본 연구는 중학교 영어교과서의 학년 간 연계성을 평가하는데 Coh-Metrix와 같은 언어분석 프로그램이 유용하게 활용될 수 있음을 제시하였다. 그러나 이러한 본 연구의 의의점과 시사점 외에 본 연구가 포함한 제한점에 대해서 기술하면 다음과 같다. 먼저 본 연구에서는 중학교 1, 2, 3학년 영어교과서에 포함된 읽기지문만을 분석의 대상으로 선정하여 분석하였다. 따라서 차후 연구에서는 영어교과서에 포함된 다양한 유형의 지문들이 Coh-Metrix와 같은 자동화된 프로그램을 통해 분석되기를 희망한다. 그리고, 본 연구에 적용된 Coh-Metrix는 양적인 측정치만을 제공하기 때문에 영어교과서의 질적인 부분에 대한 평가가 별도로 필요하리라고 생각된다. 예를 들면 영어교과서의 난이도에 영향을 미치는 요소들을 분석하기 위해 인터뷰나 설문지 등을 통해 자료를 수집한 후, 이들 자료를 기초로 영어교과서의 난이도에 대한 질적인 분석을 시도할 수 있을 것으로 기대한다. 이러한 질적 분석 결과를 Coh-Metrix의 양적 분석 결과와 비교함으로써 영어교과서의 학년 간 연계성에 대한 통합적 분석을 시도할 수 있을 것으로 기대한다. Coh-Metrix와 같은 자동화된 언어분석 프로그램이 앞으로 영어교육, 영어교재 개발, 교재 난이도 측정, 영어에세이 평가 등에 활발하게 이용되기를 기대한다.

참고문헌

- 고광윤. (2008). 교과서 영어의 영어다움에 대한 코퍼스 기반 분석: 기능어의 쓰임을 중심으로. *영어학연구*, 26, 1-19.
- 고광윤, 박정준. (2007). 중학교 영어 교과서의 어휘적 연계성에 대한 코퍼스 바탕 연구. *영어학연구*, 24, 27-45.
- 권인숙. (2002). 중학교 영어 교과서의 코퍼스 언어학적 어휘 비교 분석. *영어교육*, 57(4), 409-444.
- 권인숙. (2004). 한국 중학교 6차 및 7차 교육과정 영어교과서의 코퍼스 언어학적 어휘 비교 분석. *Foreign Languages Education*, 11(1), 211-251.
- 김민정, 송희심. (2007). 중·고등학교 영어 교과서에 나타난 읽기요소 분석. *영어학*, 7(2), 211-238.
- 김은주. (2001). 제7차 교육과정에 의한 중학교 1학년 영어교과서의 듣기, 말하기 활동 분석 연구. *영어교육*, 56(2), 219-243.
- 김정렬. (2014). 초중등 영어교육 연계성 문제: 영어교과서의 계열성을 중심으로. *어학연구*, 50(1), 161-184.

- 김정렬, 천윤희. (2008). 코퍼스 언어학적 분석을 통한 초·중등 영어 교과서의 연계성 연구. *영어교육*, 63(2), 361-394.
- 김정렬, 양지윤. (2012). Coh-Metrix를 통한 초·중등 영어교과서 연계성 분석. *영어교육*, 67(2), 319-341.
- 김지은, 전문기. (2013). 중학교 영어교과서 듣기자료의 Coh-Metrix 측정치에 대한 연계성 분석. *Journal of the Korean Data Analysis Society*, 15(4), 1987-2000.
- 문안나. (2008). 코퍼스 기반 고등학교 영어 교과서의 동사 문법패턴 분석. *언어연구*, 24(4), 659-681.
- 성일호. (2014). 이독성공식과 Coh-Metrix를 활용한 우리나라 고등학교 영어교과서 이독성 분석. *영어영문학연구*, 40(4), 299-320.
- 손미용, 이재근. (2005). 중학교 1학년 영어교과서의 쓰기 활동 자료 분석. *영어어문교육*, 11(3), 139-164.
- 송자경, 윤이정. (2003). 초등학교 영어 교과서의 연계성 분석: 제 7차 교육과정의 의사소통 기능문을 중심으로. *초등영어교육*, 9(2), 149-188.
- 송해성. (2013). 고등학교 영어 교과서 듣기 과업의 학년별 연계성 연구. *영어어문교육*, 19(4), 279-305.
- 이승환. (2013). 코메트릭스(Coh-Metrix)를 이용한 고등학교 영어1과 영어 2 교과서 문어 텍스트의 코퍼스 언어학적 난이도 분석. *외국어교육연구*, 27(2), 131-148.
- 이정민, 김해동. (2007). 영어 회화 교과서에 대한 학습자·교사 반응과 대학수학능력시험과의 연계성. *영어교육연구*, 19(2), 277-305.
- 이정민, 현대덕. (2013). 중학교 3학년 영어 교과서 읽기 자료 비교 분석. *현대영어교육*, 14(4), 263-286.
- 이화자. (2011). 고등학교 영어 교과서 읽기 지문 텍스트의 자연스러움에 대한 연구. *Foreign Languages Education*, 18(1), 121-144.
- 임인재, 전문기. (2013). 코퍼스에 기반한 중학교 1학년 개정 영어교과서 듣기자료 비교 분석. *영어학*, 13(1), 105-134.
- 전문기. (2011). Coh-Metrix를 이용한 중학교 1학년과 2학년 개정 영어교과서 읽기 자료의 코퍼스 언어학적 연계성 분석. *언어과학연구*, 56(1), 201-218.
- 전문기. (2014). 고등학교 개정 영어교과서 읽기 자료의 연계성 분석. *Journal of the Korean Data Analysis Society*, 16(2), 925-938.
- 전문기, 임인재. (2009). 코메트릭스(Coh-Metrix)를 이용한 중학교 1학년 개정 영어 교과서의 코퍼스 언어학적 비교 분석. *영어교육연구*, 21(4), 265-292.
- 전문기, 임인재. (2010). 고등학교 영어 교과서 읽기 자료의 코퍼스 언어학적 비교 분석. *Foreign Languages Education*, 17(1), 209-233.
- 정미경, 김남국. (2014). Coh-Metrix를 통한 한국인 EFL 학습자와 영어 모국어 화자의 논설문 코퍼스에 나타난 언어적 특성 분석. *언어학 연구*, 33, 373-395.

- 정은혁, 민주경. (2005). 중학교 영어 문화교육 연구: 학습자와 교과서 분석 중심으로. *영어학연구*, 19, 37-63.
- 조원향. (2003). 제7차 교육과정에 따른 중학교 2학년 영어 교과서의 읽기 자료 분석. *영미어문학연구*, 19(1), 151-169.
- Baayen, R. H., Piepenbrock, R., & Gulikers., L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21, 543-566.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33, 497-505.
- Graesser, A. C., Jeon, M., Yan, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, 15(3), 199-213.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Matrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193-202.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Haviland, S., & Clark, H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512-521.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.
- McEnery, T., & Wilson, A. (2007). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

예시언어(Examples in): English

적용가능 언어(Applicable Languages): English

적용가능 수준(Applicable Levels): Secondary

전문기
건국대학교 영어영문학과
143-701 서울시 광진구 능동로 120
Tel: (02) 450-3356/ H.P.: 010-7671-1165
Email: mjeon1@konkuk.ac.kr

Received 20 December 2014

Revised 29 January 2015

Accepted 15 February 2015