

Coh-Metrix 프로그램을 활용한 2015년 개정 교육과정 중학교 1학년 영어 교과서 비교 분석*

류지수 (건국대학교)

전문기 (건국대학교)

Ryu, Jisu and Moongee Jeon. 2020. A Coh-Metrix analysis of the first grade middle school English textbooks revised by the 2015 National Curriculum. *Korean Journal of English Language and Linguistics* 20, 89–121. This study aims to analyze the differences among seven first grade middle school English textbooks revised by the 2015 national curriculum using Coh-Metrix, a computer tool that has been widely used for analyzing various types of corpora including textbooks, student essays, and literature works. The Coh-Metrix measures selected for the current study included basic counts (number of words, number of sentences, mean sentence length, mean word length), word frequency, word information (imageability, concreteness, age of acquisition), lexical diversity (type-token ratio for content words), content words (number of nouns, number of verbs, number of adjectives, number of adverbs), personal pronouns (first person pronouns, second person pronouns, third person pronouns), connectives (causal connectives, additive connectives, temporal connectives), standard readability scores (Flesch Reading Ease score, Flesch-Kincaid Grade Level score), syntactic complexity (noun density score, number of words before main verbs), co-reference cohesion (argument overlap for adjacent sentences, argument overlap for all sentences), and semantic similarity (LSA cosine for adjacent sentences, LSA cosine for all sentences) measures. The findings of this study showed that there were differences among seven first grade middle school English textbooks for the majority of the Coh-Metrix measures including basic counts, word frequency, word information, lexical diversity, content words, personal pronouns, connectives, readability, syntactic complexity, co-reference cohesion, and semantic similarity. There were also wide discrepancies in the measures among chapters of each textbook. These results afford several pedagogical and psycholinguistic implications for the development of English textbooks for EFL learners.

Keywords: Coh-Metrix, corpus analysis, middle school English textbooks, reading materials

* 제1저자: 류지수, 교신저자: 전문기

1. 서론

우리 사회가 점점 더 국제화 및 다문화 사회로 접어들면서 국제 공용어인 영어의 중요성이 커지고 있다. 이에 따라 영어교육에 대한 관심도 날로 증가하고 있다. 세계 시장에서 국가의 경쟁력을 높이기 위해 영어교육의 효과적인 설계 및 실행이 핵심적인 과제일 것이다. 이때, 대한민국과 같은 영어를 외국어로 학습하는 EFL(English as a foreign language) 환경에서는 일상의 교육현장에서 사용되는 영어 교과서가 가장 일차적인 영어 학습 및 교육 매체로써 기능한다(권인숙 2004). 이 때문에, 지난 수 십 년 동안 영어 교과서에 대한 연구가 지속적으로 이루어져 왔다(강지영, 고경희 2015, 고광윤 2008, 권인숙 2002, 김경옥, 나경희 2017, 김정렬, 양지윤 2012, 배지영 2019, 전문기 2011, 전문기 2015, 전문기, 임인재 2009).

그동안의 영어 교과서에 대한 연구는 다방면으로 진행되어왔다. 영어의 네 가지 기능인 듣기 및 말하기(김은주 2001, 송혜성 2013, 임인재, 전문기 2013), 읽기(전문기, 임인재 2009, 조원향 2003), 쓰기(손미용, 이재근 2005, 전도희, 김선영 2015)에 대한 연구가 활발히 이루어졌다. 이에 더하여, 어휘(고광윤, 박정준 2007, 민수정, 정보현 2012)와 문법(문안나 2008, 유세진, 김효영 2015)에 대한 연구도 다수 찾아볼 수 있다. 또한, 다문화 시대의 흐름에 따라 교과서에 나타난 문화적 요소를 분석한 연구(강지영, 고경희 2015, 정은혁, 민주경 2005), 다른 국가의 교과서와의 비교 연구(김지아, 최문홍 2018) 또한 진행되었다. 특히 최근에는 2015 개정 교육과정이 도입되면서 새롭게 개발된 교과서에 대한 연구들도 이미 다수 보고되었다(권정혜, 임정완 2018, 배지영 2019, 이경남 2019, 이소영, 이현우 2019). EFL 환경에서 영어 교과서가 차지하는 상대적 중요성을 고려하면(전문기 2015), 앞으로도 영어 교과서에 대해 지속적으로 연구가 수행될 것으로 기대된다.

영어 교과서에 대한 연구를 방법론적 관점에서 살펴보자면, 초기 연구는 수작업으로 이루어진 질적 연구가 주를 이루었다(김은주 2001, 손미용, 이재근 2005, 송자경, 윤이정 2003, 조원향 2003). 이러한 연구들은 교과서 자료를 수집하고 분석하는 과정이 수작업으로 이루어졌기 때문에 수집된 자료의 양이 제한적이었고, 소수의 측정치를 기반으로 분석이 이루어졌기 때문에 분석 단위가 교과서의 특정 측면에 편중될 여지도 있었다(박정준 2007). 이에 대해, 전산언어학(Jurafsky and Martin 2008) 및 코퍼스 언어학(Meyer 2002) 분야에서는 기존의 연구들에 대한 기술적 대안을 제공하였다(e.g., Antconc, NLPTools, Range, VocabProfile, WordSmith Tools). 후속 연구들에서는 이러한 분석 도구들을 통해 무수히 많은 양의 교과서 자료들을 더욱 정확하고 신속하게 처리할 수 있었다(고광윤, 박정준 2007, 문안나 2008, 민수정, 정보현 2012, 유세진, 김효영 2015).

전산언어학과 코퍼스 언어학 분야에서 개발된 위와 같은 분석 도구들은 교과서 분석에서 기존 연구들이 안고 있던 여러 문제점(예를 들어, 분석 방법의 객관성, 정확성 및 신속성)들을 보완해줄 수 있어 영어 교과서 연구에서 널리 사용되었다. 그럼에도 불구하고, 이러한 도구들은 자연어 처리 기술의 한계로 인하여 분석의 범위가 어휘와 문장 수준으로 국한되었

고, 분석 가능한 측정치들 또한 어휘 및 문장 수, 평균 문장 길이와 같은 표층 수준의 측정치들에 대해서만 제공해주었다는 점에서 여전히 한계가 있다. 교과서에 포함된 텍스트의 다양한 특성을 더욱 체계적으로 분석하기 위해서는 표층 수준의 측정치에 더불어 심층 수준에서의 측정치 또한 분석해야 할 것이다. 실제로, 미국 멤피스대학교 지능형시스템연구소에서 개발된, 다양한 텍스트와 담화를 처리하기 위한 분석 도구인 Coh-Metrix는 어휘의 다양성, 통사 구조의 복잡성을 포함한 여러 표층 수준의 측정치들은 물론 텍스트 및 담화의 정합성(cohesion)과 응집성(coherence)과 같은 심층 수준의 측정치들까지 분석하여 제공한다(Graesser, Jeon, Cai, and McNamara 2008, Graesser, Jeon, Yan, and Cai 2007). Coh-Metrix가 제공하는 다양한 심리언어학적 측정치들을 통해 영어 교과서의 특성을 더욱 총체적이고 체계적으로 분석할 수 있게 되면서 Coh-Metrix를 이용한 다수의 영어 교과서 연구들이 보고되었다(김정렬 2014, 김정렬, 양지운 2012, 김지은, 전문기 2013, 배지영 2019, 송해성 2013, 이정민, 현태덕 2013, 임인재, 전문기 2013, 전문기 2011, 전문기 2015, 전문기, 임인재 2009). 특히, 일련의 연구자들은 Coh-Metrix의 여러 측정치를 이용하여 학년 간 연계성이 잘 확립되어 있는지(전문기 2015), 그리고 학년 내 교과서 출판사 간 지문 특성에 차이가 있는지를 조사한 연구들(이정민, 현태덕 2013, 전문기, 임인재 2009)이 수행되었다. 이러한 연구들은 Coh-Metrix에서 제공하는 여러 측정치에 대해 학년 간 연계성이 잘 확립되어 있지 않으며, 교과서 간에도 상당한 차이가 발견되었음을 밝히고 있다. 이를 통해, 교과서 설계 및 선정 시 더욱 체계적이고 객관적인 기준을 통하여 이루어질 것을 촉구하며, 그 기준으로 Coh-Metrix가 하나의 대안이 될 수 있음을 시사하고 있다.

최근 2015년 개정 영어 교과서가 개발됨에 따라 개정 영어 교과서에 대해서도 학년 간 연계성 및 학년 내 교과서 간 차이에 대해 조사해야 할 필요성이 대두되었다. 이에 본 연구는 먼저 2015 개정 영어 교과서에 수록된 읽기 지문을 발췌하여 Coh-Metrix가 제공하는 여러 언어학 및 심리언어학적 측정치를 이용하여 학년 내 교과서 간 차이에 대해 비교 분석해 보고자 한다. 본 연구에서는 개발이 완료된 중학교 1학년 영어 교과서로 국한하여 분석하였는데, 이는 고등학생에 비하여 중학생의 경우 다른 매체보다 교과서를 더욱 의존하여 학습할 것으로 예상되기 때문이다.

위에서 논의한 연구의 필요성에 기반하여 본 연구에서는 다음의 연구 목적을 달성하고자 한다.

첫째, 2015 개정 중학교 1학년 영어 교과서 7종에 대해 교과서 간 난이도 차이가 있는지 분석하는 것이다. 이를 위해 Coh-Metrix에서 제공하는 27개의 표층 및 심층 측정치를 이용하여 지문의 총체적인 특성들을 체계적으로 비교해보고자 한다.

둘째, 각 교과서 내 단원 간 난이도 차이에 대해 같은 방식으로 분석하는 것이다. 다시 말하면, Coh-Metrix 측정치에 기초해 단원 간 편차를 체계적으로 비교분석하고자 한다.

마지막으로, 이러한 결과를 토대로 차후 교과서 개발 및 학교 영어교육 현장에 유용하게 적용될 수 있는 교육적 함의점을 제공하고자 한다.

2. 이론적 배경

영어 교과서에 대한 연구는 영어의 네 기능에 대해 각각 다수의 연구가 이루어졌다. 듣기 및 말하기에 대한 연구로 김은주(2001), 송해성(2013), 임인재와 전문기(2013) 그리고 김영숙(2016) 등이 있으며, 읽기(김정렬, 양지윤 2012, 김지은, 전문기 2013, 전문기, 임인재 2009, 조원향 2003)에 대한 연구, 쓰기(손미용, 이재근 2005, 전도희, 김선영 2015)에 대한 연구들이 있다. 예를 들면, 김정렬과 양지윤(2012)은 초등학교 6학년 교과서와 중학교 1학년 교과서의 읽기 지문을 추출하여 학년 간 연계성이 잘 확립되었는지 Coh-Matrix를 이용하여 분석하였다. 담화 수준 및 여러 언어학적 측정치에 대해서는 근소한 차이만 발견되었으나 다른 여러 측정치에 대해서는 급격한 변화가 감지되어 학습자들에게 영어 학습 부담을 줄 수 있음이 시사되었다.

네 기능에 대한 연구에 더하여 어휘(고광윤, 박정준 2007, 권인숙 2004, 민수정, 정보현 2012)와 문법(문안나 2008, 유세진, 김효영 2015, 이지현, 이현우 2016)에 대한 연구 또한 진행되었다. 다문화 시대에 맞게 영어교육이 잘 설계되었는지를 조사하기 위해 교과서에 나타난 문화적 요소를 분석한 연구들(정은혁, 민주경 2005, 강지영, 고경희 2015), 다른 국가의 교과서와 우리나라 교과서를 비교한 연구(김지아, 최문홍 2018) 또한 찾아볼 수 있다.

특히 최근에는 2015년 개정 교육과정이 도입됨에 따라 새로 개발된 교과서에 대한 연구들도 이미 다수 보고되었다(권정혜, 임정완 2018, 배지영 2019, 이경남 2019, 이소영, 이현우 2019). 예를 들면, 이경남(2019)은 2015 개정 교육과정 기본어휘 목록과 고등학교 영어 교과서에 사용된 어휘 동사를 COCA 코퍼스의 핵심 학업 동사에서 추출한 고빈도 200 어휘 동사에 기초하여 분석하였다. 그 결과, 200개의 핵심 학업 동사의 35%만이 고등학교 영어 교과서에 사용된 것이 밝혀졌다. 이경남(2019)은 연구 결과를 통해 고등학교 학습자가 영어 학업 역량을 키우기 위해서는 영어 교과서에서 더 많은 학업 동사가 다양하게 사용되어야 한다고 지적하고 있다.

영어 교과서에 대한 연구를 방법론적 관점에서 살펴보면, 초기 연구들은 수작업으로 이루어진 질적 연구가 주를 이루었다(김은주 2001, 손미용, 이재근 2005, 송자경, 윤이정 2003, 조원향 2003). 예를 들면, 송자경과 윤이정(2003)은 초등학교 3, 4, 5, 6학년 교과서와 교사용 지도서, CD-ROM, 오디오 테이프에서 사용되는 지문들을 발췌하였다. 그리고 이를 제 7차 외국어 교육과정에 제시된 ‘의사소통 기능과 예시문’을 기준으로 하여 수작업으로 나누었고, 교과서에 나타난 의사소통 기능들이 단원별로, 그리고 학년별로 반복, 심화, 확대되는 정도를 분석하였다. 그 결과 초등학교 영어 교과서에 제시된 의사소통 기능들 중 일부 기능에 대해서는 체계적이지 않은 방식으로 제시가 되었으나 전반적으로는 적절한 반복, 심화, 확대를 통하여 체계적으로 구성되어 있음을 확인할 수 있었다. 최근에도, 주로 교과서의 문화적인 요소를 분석하기 위해 수작업을 통한 질적 분석한 연구들이 여럿 보고되고 있다(강순여, 홍경선 2016, 강지영, 고경희 2015, 권정혜, 임정완 2018, 허은선, 김혜정 2015).

교과서가 담고 있는 문화적 요소에 대한 분석과 같이 질적인 분석을 통해서만 밝혀낼 수 있는 연구들도 있지만, 기존의 여러 수작업을 통한 질적 분석은 전산언어학(Jurafsky and Martin 2008) 및 코퍼스 언어학(Meyer 2002) 분야에서 개발된 여러 자동화된 분석 도구를 이용하게 되면 훨씬 더 수월하게 분석이 가능하다. 더욱이, 기존의 수작업으로 이루어진 연구들이 안고 있던 여러 문제들(분석 과정의 정확성과 수집된 자료 양의 제한성 등)을 대폭 개선해줄 수 있었다. 이에 따라, 분석 도구를 활용한 영어 교과서 연구도 활발히 진행되었는데, 그러한 연구들을 2.1과 2.2에 나누어 제시한다.

2.1 코퍼스 언어학적 방법론에 기반한 영어교과서 분석 연구

전산언어학(Jurafsky and Martin 2008) 및 코퍼스 언어학(Meyer 2002) 분야에서는 기존의 수작업을 기반으로 한 연구들에 대한 기술적 대안을 제공하였다. 1960년대 초반 최초의 컴퓨터 코퍼스인 Brown Corpus가 개발된 이래, 현재까지 다양한 목적의 여러 코퍼스가 구축되어 이용 가능하다. 이와 더불어 Antconc, NLPTools, Range, VocabProfile, WordSmith Tools와 같은 분석 도구들이 개발되면서 방대한 양의 자료를 정교한 컴퓨터 분석 방법에 따라 처리할 수 있어 이러한 도구들을 사용한 영어 교과서 연구 결과의 객관성 및 신뢰성이 한층 더 증가하였다(고광운 2008, 고광운, 박정준 2007, 김미혜, 이성원 2018, 문안나 2008, 민수정, 정보현 2012, 유세진, 김효영 2015).

예를 들면, 민수정과 정보현(2012)은 고등학교 영어 교과서에 사용된 어휘에 대해서 조사하고 교육과정에 제시된 기준에 얼마나 성공적으로 부합하는지 알아보려고 하였다. 이를 위하여 WordSmith 5.0을 이용하여 교육과정 기본 어휘를 원어민 코퍼스인 COCA 코퍼스와 비교하여 그 타당성을 검증하고, 영어 교과서에 수록된 어휘들이 교육과정에 명시된 기준에 부합하는지를 살펴보았다. 그 결과, 약 90%의 기본 어휘 목록이 COCA의 고빈도 단어에 포함되어 있었다. 반면, 영어 교과서들은 대체로 학년이 올라감에 따라 교육과정 기본 어휘를 더 많은 비율 반영시키고 있었지만, 교육과정 지침 기준에는 훨씬 미달한 수치였다. 따라서 민수정과 정보현(2012)은 교과서 집필 시 더욱 체계적으로 교육과정 기준이 반영될 수 있도록 더 많은 주의가 필요하다고 지적하였다.

유세진과 김효영(2015)은 고등학교 영어 교과서 읽기 자료에 나타난 시제와 상의 사용이 실제 원어민의 사용과 유사한지를 조사하였다. 이를 위해 Antconc 분석 도구를 이용하여 원어민 코퍼스인 Ame06과 6종의 고등학교 영어 교과서를 선정하여 비교 분석하였다. 그 결과 한국 고등학교 영어 교과서는 시제와 상의 사용 빈도의 측면에서는 원어민의 실제 사용을 잘 반영해주고 있는 것으로 드러났다.

이처럼, 코퍼스 언어학적 방법론에 기반한 여러 분석 도구들은 교과서 분석에서의 신뢰성과 일반화 가능성을 더욱 향상시켜 주었다는 측면에서 교과서 분석에서 유용하게 사용되어 왔지만, 위 연구들에서 보듯 자연어 처리 기술의 한계로 인하여 분석의 단위가 어휘나 문장

수준에 대하여 한정적으로 이루어질 수밖에 없다는 점 때문에 영어 교과서의 다양한 측면에 대한 비교가 어렵다는 제약이 있다. 가령, 영어 교과서 분석 시 중요하게 살펴보아야 할 요인 중 하나인 이독성을 구성하는 것이 비단 어휘나 문장 수준에서의 표층 수준 측정치만은 아닐 것이다. 이에 더하여, 텍스트 및 담화 수준에서의 정합성(cohesion) 및 응집성(coherence)과 같은 심층 수준의 측정치 또한 이독성에 영향을 미칠 수 있다. 따라서 영어 교과서 분석의 타당도를 높이기 위해서는 텍스트의 더욱 다양한 측면을 분석하는 것이 필요할 것이다. 다음 장에서는 텍스트의 다양한 특성을 더욱 총체적으로 분석해주는 도구인 Coh-Metrix를 이용한 영어교과서 연구를 제시한다.

2.2 Coh-Metrix를 이용한 영어 교과서 연구

미국 멤피스대학교 지능형시스템연구소에서 개발된 Coh-Metrix는 여러 전산언어학적 알고리즘(Brill(1995)이 제안한 품사분석기, Landauer(2007)가 제안한 Latent Semantic Analysis 수학적 알고리즘 등)과 코퍼스언어학적 자료들(어휘집, 코퍼스 자료 등)이 내장되어 있어 어휘의 다양성, 통사 구조의 복잡성과 같은 표층 수준의 측정치들은 물론 텍스트 및 담화의 정합성(cohesion)과 응집성(coherence)과 같은 심층 수준의 측정치들까지 분석하여 제공한다(Graesser, Jeon, Yan and Cai 2007). Coh-Metrix가 제공하는 다양한 언어학 및 심리언어학적 측정치들을 통해 영어 교과서의 특성을 더욱 총체적이고 체계적으로 분석할 수 있게 되면서 영어 교과서 분석 연구에도 Coh-Metrix를 활용한 연구들이 다수 이루어졌다(김정렬 2014, 김정렬, 양지운 2012, 김지은, 전문기 2013, 배지영 2019, 송해성 2013, 이정민, 현대덕 2013, 임인재, 전문기 2013, 전문기 2011, 전문기 2015, 전문기, 임인재 2009).

예를 들면, 전문기(2015)는 Coh-Metrix를 활용하여 중학교 영어 교과서의 학년 간 연계성을 분석하였다. 이를 위해 각 학년에서 15종의 영어 교과서를 선별하여 코퍼스를 구축하였고 Coh-Metrix에서 제공하는 측정치 중 22개의 측정치에 대해 중 1, 2, 3학년 교과서를 비교하였다. 연구 결과, 단어 및 문장 수, 평균 문장 길이와 같은 기초 산출치와 주어부 밀도에 대해서는 학년 간 연계성이 적절히 통제가 되었지만, 그 외의 측정치에 대해서는 연계성이 잘 확립되지 않은 것으로 보고되었다. 이와 같은 결과는 이후 교과서 개발 시 Coh-Metrix에서 제시하는 여러 측정치들이 준거로써 활용될 수 있음을 보여주고 있다.

이와 유사하게, 배지영(2019)은 2015 개정 교육과정이 도입되면서 초등학교 6학년 영어 교과서와 중학교 1학년 영어 교과서와의 연계성을 조사하였다. 이를 위해 5종의 초등학교 6학년 영어 교과서와 11종의 중학교 1학년 영어 교과서를 선정하여 읽기 지문을 추출하였고, Coh-Metrix를 활용하여 23개의 측정치에 대해서 비교 분석하였다. 분석 결과, 이독성(readability)의 경우에는 초등학교 6학년 영어 교과서보다 중학교 1학년 영어 교과서가 더 높게 잘 통제가 되었지만, 다른 어휘, 통사적 측정치와 응집성 등에 대해서는 두 학년 간의

연계성이 체계적으로 잘 이루어지지 않은 것으로 드러났다. 전문기(2015)와 배지영(2019)의 결과는, 교과서 개발 및 선정 시 비단 어휘적 특성만 고려할 것이 아니라 다양한 측정치를 고려하는 것이 중요하다는 것을 시사하고 있다.

한편, 영어 교과서의 학년 간 연계성에 대한 분석에 앞서, 학년 내 교과서 간 특성에 차이가 있는지를 조사하는 것이 더욱 우선적인 과제일 것이다. 이러한 중요성에 주목하며 전문기와 임인재(2009)는 중학교 1학년 2009년 개정 영어 교과서(출판사) 간, 그리고 단원 간 비교 분석하였다. 이를 위해 8종의 중학교 1학년 영어 교과서를 추출하여 코퍼스를 구축하였고 12개의 Coh-Metrix 측정치를 통하여 비교하였다. 연구 결과, 기초 산출치, 단어빈도수, 읽기 난이도 지수, 통사적 복잡성, 어휘 다양성, 참조적 응집성, 그리고 의미적 응집성에 대해서 교과서 간 차이가 두드러졌다. 또한, 단원 간 차이도 발견되었다. 이러한 결과는 영어 교과서 설계 및 실행 시 더욱 체계적인 기준이 마련되어야 할 필요성을 강조하면서, 연구에서 사용된 Coh-Metrix가 제시하는 측정치들이 구체적인 지침이 될 수 있다고 제안한다.

마찬가지로 이정민과 현태덕(2013)도 교과서 간 특성 차이를 조사하였다. 구체적으로, 중학교 3학년 2011년 개정 영어 교과서 5종의 읽기 자료를 추출하여 Coh-Metrix가 제공하는 측정치 중 13개의 측정치에 대하여 비교 분석하였다. 또한, 각 챕터 간 비교 분석도 진행하였다. 분석 결과, 중학교 3학년 영어 교과서 5종 간, 그리고 챕터 간 여러 측정치들에 대하여 상당한 차이가 발견되었다. 이정민과 현태덕(2013)은 그들의 연구 결과를 통해 향후 교과서 개발, 검토 및 선정 시 더욱 체계적이고 객관적인 기준을 통하여 실행될 것을 촉구하고 있다.

앞선 연구들(이정민, 현태덕 2013, 전문기, 임인재 2009)에 이어 2015년 개정 교육과정이 도입됨에 따라 새롭게 개발된 영어 교과서에 대해서도 교과서 간 차이에 대해 조사해야 할 필요성이 대두되었다. 이에 본 연구는 2015년 개정 영어 교과서에 수록된 읽기 지문을 추출하여 Coh-Metrix가 제공하는 여러 언어학 및 심리언어학적 측정치를 이용하여 학년 내 교과서 간 차이에 대해 비교 분석해보고자 한다.

3. 연구방법

3.1 연구대상 교과서 코퍼스

본 연구에서는 2015년 개정 교육과정에 의해 개발되어 현재 중학교 1학년 수업 시간에 사용되고 있는 총 13종의 영어 교과서 중에서 7종이 선택되어 비교 분석되었다. 본 연구에서 사용된 교과서는 출판사에 따라 A 교과서(송미정 외 2017), B 교과서(김진완 외 2017), C 교과서(양현권 외 2017), D 교과서(최인철 외 2017), E 교과서(민찬규 외

2017), F 교과서(강용순 외 2017), G 교과서(이재영 외 2017)로 표기되었다. 각 교과서에 대한 자세한 정보는 참고문헌에 제시하였다.

3.2 코퍼스 구축과 교과서 분석 도구

본 연구에서는 학교 교육현장에서 사용되는 빈도와 출판사의 다양함을 고려하여 선택된 총 7종의 중학교 1학년 영어 교과서의 차이점을 분석하기 위해 최근 들어 다양한 유형의 텍스트 및 담화 특성을 분석하기 위해 광범위하게 사용되고 있는 언어분석 시스템인 Coh-Metrix 컴퓨터 프로그램이 활용되었다(전문기 2015, 전문기, 최윤희 2019, Graesser, Jeon, Yan and Cai 2007). 그리고 차후 학년 간 연계성 분석과 중고등 영어교육과정의 처음 시작점이 되는 가장 기본적인 중학교 1학년 영어교과서 분석의 필요성이 대두되어 먼저 중학교 1학년 영어교과서를 분석대상으로 선택하였다.

본 연구에서는 개인 PC에서 구동되는 Coh-Metrix 프로그램이 활용되었다. 개인 PC용 Coh-Metrix 프로그램은 시스템의 안정성과 신뢰성을 효율적으로 점검 및 평가하기 위해 비공개용으로 개발되었다. 개인 PC용 프로그램은 파일 확장자가 txt인 컴퓨터 파일만 처리할 수 있기 때문에 본 연구에서 분석된 영어 교과서 코퍼스 파일은 모두 txt 파일 형태로 저장되었다. 구체적으로 출판사별 비교 분석을 위해 각 출판사별로 총 7개의 영어 교과서 txt 파일이 생성되었다. 또한 각 교과서의 단원별 비교 분석을 위해 총 56개(A 교과서: 9개, B 교과서: 8개, C 교과서: 8개, D 교과서: 8개, E 교과서: 7개, F 교과서: 8개, G 교과서: 8개)의 txt 파일이 구축되었다. 따라서 본 연구에서는 총 63개의 txt 파일이 분석되었다. 본 연구에서는 읽기 자료만 분석 대상으로 제한하였다. 따라서 각각의 txt 파일은 읽기 자료만을 포함하였다. 본 연구에서는 Coh-Metrix가 제공하는 측정치 중에서 텍스트 난이도, 교과서 분석, 이해도, 이독성 연구(이혜수, 전문기 2019, 전문기, 최윤희 2019)에 많이 사용되는 총 27개가 선택되어 분석되었다. 이러한 측정치는 언어 이해도 및 텍스트 난이도에 중요한 영향을 미치는 언어적 및 심리언어적 요소들을 밝힌 선행연구들(Cirilo 1981, Haberlandt and Bingham 1978, Millis 1994, Murray 1997)에 기초해 선택되었다. 이러한 측정치 목록에는 기본측정치(평균 어휘 수, 평균 문장 수, 평균 문장 길이, 평균 어휘 길이), 어휘빈도수(대수변환 어휘빈도수), 어휘속성(심상성, 구체성, 습득나이), 어휘다양성(타입-토크 비율), 주요 품사(명사 수, 동사 수, 형용사 수, 부사 수), 인칭대명사(1인칭 대명사 수, 2인칭 대명사 수, 3인칭 대명사 수), 접속사(인과 접속사 수, 부가 접속사 수, 시간 접속사 수), 표준 읽기가독성 지표(FRE 점수, FKGL 점수), 통사구조 복잡성(명사구 밀도, 본동사 앞 단어수), 참조 정합성(인접 문장 간 논항 중복, 모든 문장 간 논항 중복), 의미적 유사성(인접 문장 간 LSA 코사인 값, 모든 문장 간 LSA 코사인 값) 측정치가 포함된다.

4. 연구결과 및 논의

4.1 출판사별 분석 결과

4.1.1 기본측정치 분석 결과

기본 측정치 분석 결과는 표 1에 제시되었다. 표 1에 제시된 것처럼, 7종에 사용된 평균 어휘 수와 문장 수는 각각 1,693개와 244개였다. 구체적으로 G 교과서와 D 교과서에 상대적으로 더 많은 어휘와 문장이 사용되었다. 반면에 E 교과서는 상대적으로 가장 적은 어휘와 문장으로 구성되었다.

표 1. 출판사별 기본텍스트 측정치와 전체 평균

기본측정치	A	B	C	D	E	F	G	평균
평균 어휘 수	1,666	1,526	1,531	1,938	1,269	1,874	2,049	1,693
평균 문장 수	253	220	208	288	185	263	291	244
평균 문장 길이	6.59	6.94	7.36	6.73	6.86	7.13	7.04	6.95
평균 어휘 길이	1.33	1.25	1.25	1.31	1.35	1.27	1.31	1.30

또한 평균 문장 길이 분석 결과(7종 평균 6.95)를 살펴보면, C 교과서와 F 교과서에 상대적으로 긴 문장들이 사용되었다. 반면에 A 교과서의 평균 문장 길이 점수가 가장 낮았다. 이러한 결과는 상대적으로 C 교과서와 F 교과서의 이해 난이도가 높음을 시사한다(Graesser, Jeon, Yan and Cai 2007).

마지막으로 평균 어휘 길이 분석 결과를 제시하면, E 교과서와 A 교과서에 상대적으로 길이가 긴 어휘들이 많이 사용되었다. 반면에 B 교과서와 C 교과서에 사용된 어휘들의 길이가 상대적으로 가장 짧았다. 이러한 결과는 B 교과서와 C 교과서에 상대적으로 이해 및 습득하는데 어려운 어휘들이 많이 사용되었음을 제시한다(Graesser, Jeon, Yan and Cai 2007).

4.1.2 어휘빈도수 분석 결과

어휘빈도수 분석 결과는 표 2에 제시되었다. Coh-Metrix 시스템이 제시하는 어휘빈도수 측정치는 네덜란드 어휘정보센터(Dutch Centre for Lexical Information)에서 만든 CELEX 어휘 데이터베이스(Database)에 의해 산출된다(Baayen, Piepenbrock and Gulikers 1995). 특히, 본 연구에서는 대수변환(logarithmic transformation)에 의해 계산된 어휘빈도수 분석 결과를 제시한다. 원어휘빈도수 점수를 대수변환하면 빈도의 분포가 정규분포에 근사해지기

때문에 본 연구에서는 대수변환된 어휘빈도수를 선택하여 분석하였다(Graesser, McNamara, Louwerse and Cai 2004). 표 2에 제시된 것처럼, E 교과서와 A 교과서에 상대적으로 저빈도 어휘(low frequency word)들이 많이 사용되었다. 이러한 결과는 어휘 길이 분석 결과와 일치한다. 어휘의 길이가 길수록 그 어휘가 저빈도일 가능성이 증가하는 경향이 있다(Graesser, McNamara, Louwerse and Cai 2004).

표 2. 출판사별 어휘빈도수 측정치와 전체 평균

어휘빈도 수	A	B	C	D	E	F	G	평균
대수변환 어휘빈도수 (내용어)	2.44	2.50	2.50	2.51	2.38	2.47	2.48	2.47

어휘빈도수 분석 결과는 E 교과서와 A 교과서에 사용된 어휘들을 처리하는데 상대적으로 어려움이 더 수반됨을 제시한다. 반면에 C 교과서와 D 교과서에는 고빈도 어휘(high frequency word)들이 많이 사용되었다. 이러한 결과는 C 교과서와 D 교과서의 이해 난이도가 상대적으로 낮을 수 있음을 시사한다.

4.1.3 어휘속성 분석 결과

어휘속성 분석 결과는 표 3에 제시되었다. Coh-Metrix 시스템이 제공하는 어휘속성 측정치는 다양한 어휘 속성 정보를 담고 있는 MRC Psycholinguistic 어휘 데이터베이스에 의해 계산된다(Coltheart, 1981). 이 데이터베이스에서 어휘속성 측정치 값의 범위 100에서 700사이의 점수로 제시된다. 심상성 점수와 구체성 점수는 높을수록 텍스트에 이해하기 쉬운 어휘들이 많이 사용되었음을 의미하고, 습득나이 측정치는 높을수록 상대적으로 텍스트에 높은 연령대에 습득되는 어휘들이 많이 사용되었음을 제시한다. 따라서 습득나이 점수가 높으면 텍스트의 이해 난이도가 증가할 수 있다(Graesser, Jeon, Yan and Cai 2007).

표 3. 출판사별 어휘속성 측정치와 전체 평균

어휘속성	A	B	C	D	E	F	G	평균
심상성	440	446	441	423	444	455	435	440
구체성	407	407	405	390	407	420	400	405
습득나이	286	262	255	269	283	252	276	269

먼저 심상성 분석 결과를 살펴보면 F 교과서의 점수가 상대적으로 가장 높고 D 교과서가 가장 낮다. 따라서 F 교과서에 상대적으로 이해하기 쉬운 어휘들이 많이 사용되고 있음을

알 수 있다. 반면에 이러한 결과는 D 교과서에 상대적으로 이해하기 어려운 어휘들이 사용되고 있음을 제시한다.

구체성 어휘 분석 결과를 살펴보면, F 교과서의 점수가 가장 높고 D 교과서가 가장 낮다. 이러한 결과는 심상성 분석 결과와 일치한다. 다시 말하면, 구체성 분석 결과는 F 교과서에 사용된 어휘들이 상대적으로 이해하기 쉽고 D 교과서에 사용된 어휘들이 상대적으로 어려움을 시사한다.

마지막으로 표 3에 제시된 습득나이 분석 결과에 의하면, A 교과서와 E 교과서의 점수가 상대적으로 다른 교과서들의 점수에 비해서 높았다. 다시 말하면 A 교과서와 E 교과서에 높은 연령대에 습득되는 어휘들이 많이 사용되었다. 따라서 A 교과서와 E 교과서의 이해 난이도가 증가할 수 있다(전문기, 최윤희 2019). 이러한 결과는 어휘빈도수 분석 결과와 일치한다. 일반적으로 저빈도 어휘들이 높은 연령대에 습득된다는 점을 고려하면 예상되는 결과라고 할 수 있겠다.

4.1.4 어휘다양성 분석 결과

어휘다양성 분석 결과는 표 4에 제시되었다. Coh-Metrix 프로그램이 제공하는 어휘다양성 지표는 타입-토큰 비율(type-token ratio) 값에 의해 표현된다(Graesser, McNamara, Louwerse and Cai 2004). 텍스트 내에서 타입은 개개 어휘들의 첫 출현 빈도 값(따라서 타입은 항상 1임)을 지칭하고 토큰은 개개 어휘들이 텍스트 내에서 반복되어 사용되는 빈도 값을 제시한다. 예를 들면, “교과서”라는 어휘가 텍스트 내에 100번 반복되어 사용되면 “교과서”의 타입은 1이고 토큰은 100이 된다. 타입-토큰 비율 값은 타입의 값을 토큰의 값으로 나눈 값을 지칭한다. 따라서 이 값이 높을수록 텍스트 내에 다양한 어휘들이 사용될 가능성이 증가한다(전문기, 최윤희 2019, Graesser, Jeon, Yan and Cai 2007).

표 4. 출판사별 어휘다양성 측정치와 전체 평균

어휘다양성	A	B	C	D	E	F	G	평균
타입-토큰 비율(내용어)	0.54	0.56	0.53	0.47	0.55	0.52	0.54	0.53

표 4에 제시된 어휘다양성 분석 결과를 살펴보면, 평균(0.53)을 기준으로 했을 때 D 교과서를 제외하고 대부분 교과서들의 어휘다양성 점수는 거의 유사하다(0.52에서 0.56사이). 따라서 본 연구에서 분석된 대부분의 교과서들에 사용된 어휘다양성 정도는 교과서(출판사) 간에 큰 편차가 없다고 할 수 있다. 다시 말하면 이러한 결과는 교과서 집필 시 어휘들이 교육인적자원부에서 제시한 기본 어휘 관련 지침에 따라 교과서 간에 비교적 잘 통제되어 사용되었음을 시사한다. 다만 D 교과서에서 동일한 어휘들이 상대적으로 많이 반복되어 사

용되고 있다. 따라서 어휘다양성 정도를 고려할 때, D 교과서의 이해 난이도가 다른 교과서들에 비해서 상대적으로 낮을 수 있겠다.

4.1.5 주요 품사 분석 결과

출판사별 주요 품사(명사, 동사, 형용사, 부사) 분석 결과는 표 5에 제시되었다. 주요 품사 측정치는 코퍼스 길이에 의한 가위 변인(confounding variable) 효과를 통제하기 위해 발생 점수(incidence score)로 변환되었다. 발생 점수는 비례 점수로 표현되는데 Coh-Metrix 시스템에서는 1,000을 기준으로 산출된다(Graesser, Jeon, Yan and Cai 2007). 예를 들면, 분석하고자 하는 코퍼스의 길이가 100(100개 어휘로 구성되어 있다는 의미)이고 그 코퍼스에 명사가 30번 나온다고 가정하면, 이때 그 명사의 발생 점수는 “100 : 30 = 1,000 : 발생 점수” 비례식을 통해 계산된다. 이 경우 그 명사의 발생 점수는 300이 된다.

먼저 명사수 분석 결과를 살펴보면, A 교과서와 G 교과서에 상대적으로 많은 명사가 사용되었고 D 교과서에 가장 적은 명사가 사용되었다. 반면에 B 교과서에 가장 많은 동사가 사용되었고 G 교과서에 가장 적은 동사가 사용되었다. 형용사 사용 빈도는 평균(75)을 고려했을 때 대체적으로 교과서 간에 큰 차이를 보이지 않았다. 다만 D 교과서에 큰 차이를 보이는 않았지만 가장 많은 형용사가 사용되었다. 마지막으로 부사 분석 결과를 살펴보면, A 교과서와 D 교과서에 상대적으로 많은 부사가 사용되었다.

표 5. 출판사별 주요 품사 측정치와 전체 평균

주요 품사	A	B	C	D	E	F	G	평균
명사 수	297	272	273	238	279	282	287	275
동사 수	134	144	137	135	137	140	122	136
형용사 수	70	75	73	81	78	74	73	75
부사 수	64	48	54	63	46	55	62	56

주요 품사 분석에서 주목할 점은 각 교과서들에 사용된 내용어(명사, 동사, 형용사, 부사)들의 구체적인 빈도 분포를 알 수 있다는 것이다. 특히 어휘빈도, 어휘속성(심상성, 구체성, 습득나이), 어휘다양성(타입-토큰 값) 측정치들에 대한 분석은 내용어들에 대한 결과들이기 때문에 주요 품사 분석 결과를 이러한 측정치들에 대한 분석 결과들과 연결시켜 해석할 수 있다. 먼저 A 교과서의 평균 어휘 길이 및 습득나이 측정치 값들은 가장 높았다. 따라서 A 교과서에 사용된 명사와 부사들이 타 교과서들에 비해 처리 및 습득이 어려울 수 있다는 점이 시사된다. 또한 A 교과서에 사용된 이러한 어휘들의 빈도수도 상대적으로 낮았음을 고려할 때 전반적으로 A 교과서에 어려운 명사와 부사들이 사용되었다고 할 수 있다. 반면

D 교과서에는 고빈도 어휘(형용사, 부사)들이 많이 사용되었고(표 2 참조), 타입-토큰 비율 분석 결과에서 제시된 것처럼(표 4 참조) 이러한 어휘들이 상대적으로 많이 반복되어 사용되었을 가능성이 높기 때문에 전반적으로 이해 난이도가 낮을 가능성이 암시된다.

4.1.6 인칭대명사 분석 결과

Coh-Metrix 시스템이 제공하는 인칭대명사에 대한 분석 결과는 표 6에 요약되었다. 표 6에 제시된 것처럼, 먼저 1인칭 대명사 분석 결과를 살펴보면, E 교과서와 G 교과서에 상대적으로 많은 1인칭 대명사가 사용되었다. 반면에 2인칭 대명사는 D 교과서에 가장 많이 사용되었고 3인칭 대명사는 F 교과서와 G 교과서에 많이 사용되었고 B 교과서와 E 교과서에 적게 사용되었다.

표 6. 출판사별 대명사 측정치와 전체 평균

인칭대명사	A	B	C	D	E	F	G	평균
1인칭 대명사수	27.0	30.2	26.2	27.4	37.1	19.0	34.9	28.8
2인칭 대명사수	5.4	11.5	10.5	15.8	13.4	8.3	7.6	10.3
3인칭 대명사수	18.0	12.1	15.0	17.0	12.6	21.1	19.1	16.4

일반적으로 텍스트 내에 대명사가 많으면 대명사가 지칭하는 대상을 식별해야 되기 때문에 이해 난이도가 상승한다(전문기 2011, Graeser, Jeon, Yan and Cai 2007). 따라서 E 교과서(1인칭, 2인칭 참조)와 G 교과서(1인칭, 3인칭 참조)의 이해 난이도가 증가할 가능성이 시사된다. 반면에 A 교과서와 G 교과서의 (2인칭 대명사 참조)의 이해 난이도는 감소할 가능성이 제시된다.

4.1.7 접속사 분석 결과

Coh-Metrix 시스템이 제공하는 접속사 측정치 중에서 본 연구에서 분석된 접속사는 표 7에 제시되었다. 일반적으로 접속사는 텍스트 이해 처리를 촉진하는 효과가 있다(Millis and Just 1994).

표 7. 출판사별 접속사 측정치와 전체 평균

접속사	A	B	C	D	E	F	G	평균
인과 접속사	22.8	17.7	24.2	26.3	26.8	20.8	21.0	22.8
부가 접속사	4.2	2.0	5.9	12.9	5.5	8.5	7.3	6.6
시간 접속사	17.4	30.8	28.7	19.6	27.6	21.9	34.2	25.7

특히 텍스트 내에 사용된 인과 접속사(예를 들면, because)는 텍스트의 인과성을 크게 촉진시킨다(Millis and Just 1994, Murray 1997). 표 7에 제시된 결과를 살펴보면, 큰 차이를 보이지는 않지만 D 교과서와 E 교과서에 상대적으로 많은 인과적 접속사가 사용되었다. E 교과서의 경우 평균 어휘 길이, 단어빈도, 습득나이 측정치로 인해 증가된 이해 난이도가 인과적 접속사로 인해 다소 낮아질 가능성이 시사된다.

부가 접속사 분석 결과, A 교과서와 B 교과서에 상대적으로 적은 부가 접속사(예를 들면, and, moreover)가 사용되었다. 반면에 D 교과서에 가장 많이 사용되었다. 부가 접속사는 인과 접속사와 다르게 텍스트의 이해도를 유의미하게 촉진하는 효과(Millis and Just 1994)는 없지만 문장의 길이에 영향을 미칠 수 있다. 전술한 바와 같이 A 교과서의 평균 문장 길이는 가장 짧았다. 시간 접속사(예를 들면, until)는 G 교과서에 가장 많이 사용되었고 A 교과서에 가장 적게 사용되었다. 전반적으로 접속사 분석 결과는 각 교과서 간에 큰 편차를 보여주었다.

4.1.8 표준 읽기가독성 지표 분석 결과

읽기가독성 분석 결과는 표 8에 요약되었다. 읽기가독성 지표와 관련하여 Coh-Metrix 시스템은 두 개의 측정치를 제공한다. 먼저 표 8에 제시된 Flesch Reading Ease(FRE) 점수는 0에서 100 사이의 점수로 계산되는데 이 점수가 높은 경우 텍스트의 이해 난이도는 낮아진다(전문기 2015, Graesser et al. 2004).

표 8. 출판사별 표준 읽기가독성 지표 측정치와 전체 평균

읽기가독성	A	B	C	D	E	F	G	평균
FRE 점수	88	94	94	89	85	92	89	90
FKGL 점수	2.7	1.8	2.0	2.5	3.1	2.2	2.6	2.4

반면에 Flesch-Kincaid Grade Level(FKGL) 점수는 1과 12 사이의 점수로 산출되는데 이 값이 클수록 텍스트의 이해 난이도는 증가한다. 특히 FKGL 점수는 미국 초중고 학년 등급(1학년 ~ 12학년)에 상응한다. 다시 말하면 어떤 텍스트의 FKGL 점수가 1이라고 가정하면 그 텍스트의 읽기 수준은 미국 초등학교 1학년에 해당한다고 볼 수 있다. 두 지표의 산출 공식을 살펴보면, 모두 단어 길이와 문장 길이에 의해 계산된다(Klare 1974-1975). 따라서 단어 길이와 문장 길이가 길면 이 두 지표에 의한 텍스트의 이해 난이도는 증가한다(전문기 2015).

표 8에 제시된 FRE 점수 분석 결과를 살펴보면, E 교과서와 A 교과서의 점수가 타 교과서들에 비해서 상대적으로 낮았다. 반면에 B 교과서와 C 교과서는 상대적으로 높았다. 따라서 E 교과서와 A 교과서의 읽기 난이도가 상대적으로 높고, B 교과서와 C 교과서는 낮다

고 할 수 있다. 이러한 결과는 FKGL 분석 결과와 유사하다. 표 8에 제시된 것처럼 E 교과서와 A 교과서의 FKGL 점수는 타 교과서에 비해 상대적으로 높았고 B 교과서와 C 교과서는 상대적으로 낮았다. 이러한 결과는 평균 어휘 길이 분석 결과와도 일치한다(표 1 평균 어휘 길이 결과 참조). 그러나 흥미롭게도 C 교과서의 평균 문장 길이는 가장 높았다. 이러한 이유는 평균 이상으로 C 교과서에 사용된 접속사(인과 접속사, 시간 접속사) 때문이라고 설명할 수 있다.

4.1.9 통사구조 복잡성 분석 결과

통사 구조 복잡성 측정치 분석 결과는 표 9에 제시되었다. 표 9에 제시된 것처럼 Coh-Metrix 시스템은 두 유형의 측정치를 제공한다. 먼저 명사구 밀도(noun density) 점수는 명사구 내에 있는 표제어(headnoun)를 수식하는 수식어의 수를 전체 명사구의 수로 나눈 값을 지칭한다(Graesser, Jeon, Yan and Cai 2007). 따라서 이 값이 클수록 통사적으로 복잡할 가능성이 증가한다. 본동사 앞 어휘수는 본동사 앞에 위치한 어휘들의 수를 나타낸다. 명사구 밀도 점수와 같이 이 값이 클수록 문장의 통사 구조 복잡성은 증가한다(Graesser, Jeon, Yan and Cai 2007).

표 9. 출판사별 문장 통사 구조 측정치와 전체 평균

통사구조	A	B	C	D	E	F	G	평균
명사구 밀도	0.72	0.71	0.69	0.59	0.59	0.66	0.65	0.66
본동사 앞 어휘수	1.58	1.60	1.44	1.58	1.52	1.76	1.64	1.59

먼저, 표 9에 제시된 명사구 밀도 점수 결과를 제시하면, A 교과서와 B 교과서의 점수가 타 교과서에 비해 상대적으로 높았고, D 교과서와 E 교과서의 점수는 평균(0.66)보다 낮았다. 반면에 본동사 앞 어휘수 결과를 살펴보면, F 교과서의 점수가 가장 높았다. 이러한 결과는 F 교과서에 상대적으로 긴 문장들이 사용되어 통사적 복잡성이 높음을 시사한다. 이 결과는 평균 문장 길이 분석 결과와 유사하다. F 교과서에 사용된 문장들은 실제로 평균보다 높았다. 요약하면 통사 구조 분석 결과 역시 교과서(출판사) 간에 큰 편차가 존재함이 제시되었다.

4.1.10 참조정합성 분석

참조정합성(co-reference cohesion)은 문장 사이에 논항(argument)이 반복되어 사용될 때 발생한다(Cirilo 1981, Haberlandt and Bingham 1978). 논항은 명사, 대명사, 명사구의

로 구성된다. 일반적으로 문장 사이에 논항이 반복되어 사용되면 참조정합성이 증가하여 결과적으로 텍스트 이해 처리가 용이해진다(Graesser, McNamara, Louwerse and Cai 2004, 전문기 2014, 2015, 전문기, 최윤희 2019). 다시 말하면, 텍스트 이해 난이도가 감소한다. 본 연구에서는 두 유형의 참조정합성 측정치를 분석하였다. 먼저 인접 문장 간 논항 중복 측정치인 지엽적 정합성(local cohesion) 결과를 살펴보면(표 10 참조), E 교과서와 F 교과서의 값이 타 교과서에 비해 상대적으로 크다. 반면에 B 교과서는 가장 낮았다. 이러한 결과는 E 교과서와 F 교과서의 이해 난이도가 논항 반복으로 인해 낮아질 가능성을 시사한다. 특히 평균어휘 길이, 습득나이, FRE 점수, FKGL 분석 결과에 기초했을 때, E 교과서의 이해 난이도는 높은 편이었는데 논항반복으로 인해 낮아질 가능성이 제시된다.

표 10. 출판사별 참조정합성 측정치와 전체 평균

참조정합성	A	B	C	D	E	F	G	평균
논항 중복 (인접 문장 간)	0.30	0.26	0.31	0.36	0.43	0.41	0.31	0.34
논항 중복 (모든 문장 간)	0.20	0.22	0.22	0.21	0.28	0.23	0.20	0.22

모든 문장 간 논항 중복 측정치인 전반적 정합성(global cohesion) 분석 결과를 제시하면, 표 10에 제시된 것처럼 E 교과서의 점수가 가장 높았다. 이러한 결과는 지엽적 정합성 분석 결과와 일치한다. 전반적으로 평균(0.22)과 비교했을 때 E 교과서를 제외하고 나머지 교과서들 간의 전반적 정합성은 대체적으로 유사하였다.

4.1.11 의미적 유사성 분석 결과

의미적 유사성(semantic similarity) 분석 결과는 표 11에 제시되었다. Coh-Metrix 시스템이 제공하는 의미적 유사성은 수학적 모형인 Latent Semantic Analysis(LSA)에 의해서 계산된다(Graesser, McNamara, Louwerse and Cai 2004). LSA는 어휘 간, 문장 간, 단락 간의 의미적 유사성을 산출하는데 이용된다(Landauer, Foltz and Laham 1998). 구체적으로 LSA의 산출 원리는 빅 데이터(big data)에 의해 생성된 2차원(어휘-문장)의 거대 행렬(matrix)을 SVD(Singular Value Decomposition)에 의해 작은 행렬로 분해하는 과정에서 생성된 의미 공간(semantic space) 상에서 어휘 간, 문장 간, 단락 간의 의미적 유사성을 도출하는 것이다. 이 공간 상에서 어휘 간, 문장, 단락 간의 의미적 유사성은 두 벡터(vector) 요소(말하자면, 어휘, 문장, 단락) 간의 코사인(cosine) 값으로 표현된다. 이 값이 클수록 두 요소 간의 의미적 유사성은 증가한다.

표 11에 제시된 의미적 유사성 분석 결과를 살펴보면, 먼저 인접 문장 간 LSA 값은 E

교과서와 D 교과서에서 상대적으로 높았다. 반면에 A 교과서와 G 교과서에서는 낮았다. 특히 E 교과서 분석 결과는 참조정합성 분석 결과와 일치한다. 따라서 LSA 값을 고려했을 때 E 교과서의 이해 난이도는 감소할 수 있음이 제시된다. 반면에 A 교과서의 LSA 값은 7종 교과서 중에서 가장 낮았다. A 교과서의 이해 난이도는 전반적으로 높은 편(평균 어휘 길이, 습득나이, FRE 점수, FKGL 점수, 명사구 밀도 점수 참조)이었는데 이러한 난이도가 인접 문장 간의 상대적으로 낮은 의미적 유사성에 의해 더 증가할 수 있겠다.

마지막으로 모든 문장 간 LSA 값을 살펴보면(표 11 참조), E 교과서가 가장 높았다. 이러한 결과는 참조정합성 분석 결과와 일치한다. 따라서 이러한 결과는 E 교과서의 이해 난이도가 이 측정치를 고려했을 때 낮아질 수 있음을 시사한다. E 교과서를 제외한 나머지 교과서들의 의미적 유사성 값은 평균(0.13)을 기준으로 했을 때 C 교과서의 값이 조금 높았지만 전반적으로 큰 편차를 보이지 않았다.

표 11. 출판사별 의미적 유사성 측정치와 전체 평균

의미정합성	A	B	C	D	E	F	G	평균
LSA 코사인 (인접 문장 간)	0.12	0.14	0.16	0.18	0.19	0.16	0.13	0.15
LSA 코사인 (모든 문장 간)	0.11	0.10	0.16	0.13	0.22	0.08	0.13	0.13

4.2 단원별 분석 결과

각 교과서의 단원별 분석 결과는 논문의 지면 제한 상 부록(부록 A ~ 부록 G 참조)에 제시하였다. 출판사별 분석에서처럼 단원별 분석에서도 기본측정치(평균 어휘 수, 평균 문장 수, 평균 문장 길이, 평균 어휘 길이), 어휘빈도수(대수변환 어휘빈도수), 어휘속성(심상성, 구체성, 습득나이), 어휘다양성(타입-토큰 비율), 주요 품사(명사 수, 동사 수, 형용사 수, 부사 수), 인칭대명사(1인칭 대명사 수, 2인칭 대명사 수, 3인칭 대명사 수), 접속사(인접 접속사 수, 부가 접속사 수, 시간 접속사 수), 표준 읽기가독성 지표(FRE 점수, FKGL 점수), 통사 구조 복잡성(명사구 밀도, 본동사 앞 단어수), 참조 정합성(인접 문장 간 논항 중복, 모든 문장 간 논항 중복), 의미적 유사성(인접 문장 간 LSA 코사인 값, 모든 문장 간 LSA 코사인 값) 측정치를 분석하였다.

부록에 제시된 것처럼 각 교과서의 단원별 Coh-Metrix 분석 결과를 간략히 제시해 보면, 예외 없이 각 교과서의 단원 간에 큰 편차가 존재하였다. 교육인적자원부에서 제시한 교과서 개발 지침이 있음에도 불구하고 단원 간에 이러한 편차가 발생한 이유는 각 출판사의 별도 출판 지침과 교과서 개발자의 개인적 특성 때문이라고 설명할 수 있다. 물론 교재 개발 시 학습지도 전략에 따라 단원별 이해 난이도 수준을 조정하는 것은 필요한 일이다. 그러나

부록에 제시된 분석 결과를 살펴보면 이러한 난이도 수준이 체계적으로 조정되었다고 보기 어렵다. 교과서 집필 시 집필 회의 등을 통해 개발자들이 이러한 편차 문제를 해결할 수도 있겠지만 현실적으로 객관적인 지표 없이 편차를 조정하는 데 한계가 있다. 본 연구에서 도출한 Coh-Metrix 분석 결과가 교과서 집필 시 각 단원별 이해 난이도 수준을 체계적으로 조정하는데 기여할 수 있을 것으로 기대한다.

5. 결론

세계화의 맥락 속에서 영어의 위상이 높아짐에 따라 영어교육에 대한 관심도 증대되었다. 영어 교육의 효과적인 설계와 실행을 위해서는 영어 교과서를 잘 구축하는 것이 핵심적인 과제일 것이다. 이는, 한국과 같은 EFL 환경에서 영어교육 시 실제 학교 현장에서 사용하게 되는 영어 교과서가 일차적인 학습 매체로 작용하고 있기 때문이다. 이전의 연구자들은 영어 교과서 연구의 중요성을 인식하며 교과서의 다양한 측면을 분석하였다. 하지만, 기존 많은 연구들은 주로 수작업을 통하여 이루어진 질적 연구이거나 어휘나 문장 수준에서의 분석만 가능한 분석 도구를 이용하여 분석의 범위가 제한적이었다. 그러나, 영어 교과서의 더욱 총체적인 분석을 위해서는 비단 어휘와 같은 표층 측정치뿐만 아니라 논항 중복, 심상성 등과 같은 심층 측정치들에 대해서도 분석을 해야 할 것이다. 현재 연구에서는 이러한 필요성에 주목하면서 교과서 지문 분석을 위해 표층 및 심층 측정치 모두를 포함하였다. 구체적으로 담화 분석을 위해 개발된 분석 도구인 Coh-Metrix를 활용하여 2015 개정 중학교 1학년 영어 교과서 7종의 읽기 지문을 추출한 뒤 7종 교과서 간, 그리고 각 교과서 내 단원 간 난이도를 비교 분석하였고, 분석을 위해 27개의 표층 및 심층 측정치를 사용하였다. 현재 연구의 분석 결과를 다음과 같이 요약할 수 있다.

먼저, 현재 연구에서 분석한 7종의 교과서들은 여러 측정치에 대해 많은 차이를 보였다. 구체적으로, 기본측정치, 어휘빈도수, 어휘속성, 주요 품사, 인칭대명사, 접속사, 읽기가독성, 통사구조, 그리고 참조정합성 및 의미정합성과 같은 지표에 대해 큰 편차가 존재하는 것을 수치로 확인할 수 있었다. 가령, A 교과서는 다른 교과서들과 비교해 보았을 때 표층 측정치(기본 측정치, 어휘빈도수, 어휘속성, 읽기가독성) 및 심층 측정치(참조정합성, 의미정합성)에 대해 상대적으로 높은 난이도를 보였다. 교과서 간 여러 측정치들에 대해 큰 차이가 있다는 점은 영어 교과서 출판사 및 단원 간 비교 분석한 기존 연구(이정민, 현대덕 2013, 전문기, 임인재 2009)에서도 보고되었다. 이러한 결과는 영어 교과서 읽기 자료를 개발할 때 국가교육과정에 대한 기준이 일괄적으로 적용되었다고 보기 어려울 것이다. 이 때문에, 학교 현장에서 교과서 선정 시 교사가 직접 지문의 여러 특성들을 비교 분석해보고 선정해야 하는 어려움이 있을 수 있다.

특히, 이러한 어려움은 일부 교과서들에 나타난, 서로 다른 측정치에 의하여 난이도를 조

절하려는 현상으로 인하여 더욱 심화될 수 있다. 예를 들면, E 교과서의 경우에는 표층 측정치에서는 상대적인 난이도가 높았지만 여러 심층 측정치에서 쉬운 것으로 드러난 반면, B 교과서는 표층 측정치에서는 상대적으로 난이도가 낮지만 심층 측정치에서는 상대적으로 더 어려운 것으로 나타났다. 이를 통해 전체적으로 난이도가 지나치게 쉽거나 어려워지는 것을 막아줄 수 있다. 이러한 결과들이 시사하는 바는 교과서 선정 시 지문의 특정 측면만을 통해 선정하기보다는 텍스트의 여러 측면을 더욱 총체적으로 파악해서 선정하는 것이 중요하다는 것이다. 한편으로는 교과서 간 읽기 난이도 수준이 다양하다는 점을 잘 활용한다면 학습자의 수준이나 필요에 맞추어 적절한 난이도의 교과서를 선택할 수 있다는 가능성을 제공해줄 수 있기 때문에, 더욱 객관적이고 종합적으로 교과서의 난이도를 파악해서 목적에 맞게 선정을 해야 할 것이다. 또한 이러한 교과서 간 편차에 대한 체계적인 기준과 원리가 확립되어 국가교육과정에 반영이 되기를 기대한다. 현재 연구에서는 Coh-Metrix와 같은 자동적으로 담화를 분석해주는 도구를 통해 이러한 부분이 가능하다는 것을 제안한다.

교과서 간 차이에 더해, 각 교과서 내 단원들 간에도 편차 또한 두드러졌다. 현재 연구에서 비교 분석한 7종 중학교 영어 교과서 각 단원들은 표층 및 심층 측정치 모두에 대해 큰 차이를 보였는데, 이때 특정한 방향성을 보이는 것이 아니라(가령, 단원이 증가함에 따라 읽기 난이도도 상승하는 방식과 같이) 불규칙적으로 편차가 크게 드러났다. 이 때문에 학교 현장에서 적용될 때, 난이도를 예측하기 어려워 학습자에게 부담을 줄 수 있을 것이라 예상된다. 따라서 교과서 단원 개발 시 학습 시기에 맞추어 난이도가 점층적으로 올라가는 방식으로 구성하게 된다면, Krashen(1985)의 입력 가설(input hypothesis)에 따라 학습자들에게 더욱 이해 가능한 입력(comprehensible input [i+1])을 순차적으로 제공해주어 학습을 촉진할 수 있을 것이다. 텍스트의 읽기 난이도를 결정하는 데 수많은 요인이 동시에 작용하기 때문에 개별 집필진들이 판단하기 어려울 수 있다. 이에 따라, 교과서 개발 시에 Coh-Metrix가 더욱 객관적이고 신뢰할 만한 도구로써 기능할 수 있기 때문에 활용될 것을 권고한다. 가령, 교과서 내 단원 개발 시에 집필진의 직관에 따라 난이도를 점층적으로 상승하는 방향으로 구성한 후, Coh-Metrix를 활용하여 여러 측정치에 대해서 그러한 방향성이 일관적으로 적용되고 있는지 확인할 수 있겠다. 이를 통하여, 어떠한 부분에 대해 수정해야 할지 더욱 구체적인 지침을 얻을 수 있을 것이다.

현재 연구의 제한점에 기반하여 후속 연구에 대한 제언을 하고자 한다. 먼저, 현재 연구에서는 영어 교과서 분석을 위해 중학교 1학년 영어 교과서 7종을 선별하였다. 그러나 후속 연구에서는 더 많은 출판사에서 발간된 교과서를 포함하여 분석한다면 학교 현장에서 교과서 선정에 있어서 더욱 유익한 정보를 제공해 줄 것이라고 본다.

또한, 본 연구에서는 2015 개정 중학교 1학년 교과서만으로 한정하여 분석하였는데, 후속 연구에서는 중학교 2, 3학년 영어 교과서까지 포함하여 학년 간 연계성을 분석해 본다면 중학교 학습자들이 영어 학습을 하는 데 무리가 없는 방식으로 연계성이 잘 형성이 되었는지 점검해 볼 수 있을 것이다. 나아가 중학교뿐만 아니라 초등학교, 고등학교 영어 교과서까지

분석한다면 한국의 초중고 영어 교과서의 설계 및 사용의 현 상태에 대한 보다 포괄적인 이해가 가능할 것이다.

마지막으로, 본 연구에서 영어 교과서를 분석하기 위해 사용한 측정치는 Coh-Metrix가 제공하는 다양한 측정치 중 27가지이다. 현재 연구와 유사하게 Coh-Metrix를 활용하여 영어 교과서의 출판사 간 분석을 진행한 기존 연구인 전문기, 임인재(2009)에서 사용한 12개, 그리고 이정민, 현태덕(2013)에서 사용된 13개의 측정치에 비하면 보다 확대된 수치이다. 후속 연구에서도 점차적으로 더욱 확대하여 텍스트의 더 많은 자질을 분석한다면 텍스트 개발, 평가, 선정 등에서 효과적으로 이용될 수 있을 것이라 기대한다.

마지막으로, 현재 연구에서는 교과서 간 분석을 위해 Coh-Metrix가 제공하는 양적 수치만을 기반으로 살펴보았다. 후속 연구에서는 이를 보완하여, 먼저 양적 분석을 통하여 교과서 간, 혹은 교과서 내 단원 간 언어적 특성의 차이를 객관적으로 발견한 뒤, 그에 상응하는 실제 교과서 지문을 비교하는 질적 분석을 더한다면 더욱 구체적으로 지문 간 특성 차이가 다각도에서 밝혀질 것으로 기대한다.

코퍼스 언어학 및 전산언어학 분야가 점점 더 발전함에 따라, 교과서 지문의 난이도를 측정하기 위해 사용된 기존의 수작업적 접근과 주로 어휘에서만 편중되었던 기존의 분석 도구들이 갖는 여러 한계점에 대해 많은 개선이 이루어져 왔다. 특히 현재 연구에서 사용한 웹 기반 자동 담화 분석 도구인 Coh-Metrix를 이용한다면 더욱 신속하고 신뢰할 수 있게 방대한 양의 텍스트에 대해 다양한 측면에 대한 총체적 분석이 가능하다. 현재 연구에서는 2015 개정 중학교 1학년 영어 교과서의 출판사 간 난이도 비교 분석을 위해 Coh-Metrix를 활용하여 다양한 표층 및 심층 측정치에 대해 큰 차이가 있음을 밝혀내었다. 이처럼, 향후 교과서 관련 연구를 포함하여 여러 텍스트 분석 연구에 Coh-Metrix가 광범위하게 활용되어 체계적인 점검 도구로써 텍스트 개발, 평가, 선정 등에 도움을 줄 수 있을 것이라 기대한다.

참고문헌

- 강순여 · 홍경선(Kang, S. and K. Hong). 2016. 2009개정 교육과정용 초등 5·6학년 영어교과서의 문화요소(Cultural content in 5th-& 6th-grade English textbooks on 2009 national curriculum). 《영어영문학》(*The Journal of English Language and Literature*) 21-1, 339-365.
- 강용순 · 김혜동 · Gorge Whitehead · 권혜연 · 구나현 · 한경 · 홍기만(Kang, Y., H. Kim, G. Whitehad, H. Kwon, N. Koo and K. Hong). 2018. *Middle school English 1*. 서울: 다락원.
- 강지영 · 고경희(Kang, J. and K. Ko). 2015. 문화 간 의사소통능력 신장을 위한 초등영어

- 교과서의 문화교육 자료 분석(Culture education materials of elementary English textbooks as a means to develop intercultural communicative competence). 《영어영문학》(*The Journal of English Language and Literature*) 20-3, 229-251.
- 고광윤(Goh, G.). 2008. 교과서 영어의 영어다움에 대한 코퍼스 기반 분석: 기능어의 쓰임을 중심으로(A corpus-based analysis of function words in Korean English textbooks). 《영어학연구》(*English Language and Linguistics*) 26, 1-19.
- 고광윤·박정준(Goh, G. and J. Park). 2007. 중학교 영어 교과서의 어휘적 연계성에 대한 코퍼스 바탕 연구(A corpus-based study of continuity in Korean middle school English textbooks). 《영어학연구》(*English Language and Linguistics*) 24, 27-45.
- 권인숙(Kwon, I.). 2002. 중학교 영어 교과서의 코퍼스 언어학적 어휘 비교 분석(A corpus-based lexical analysis of middle school English textbooks). 《영어교육》(*English Teaching*) 57-4, 409-444.
- 권인숙(Kwon, I.). 2004. 한국 중학교 6차 및 7차 교육과정 영어교과서의 코퍼스 언어학적 어휘 비교 분석(A corpus-based lexical analysis of middle school English textbooks of the 6th and the 7th national curriculum). 《외국어교육》(*Foreign Languages Education*) 11-1, 211-251.
- 권정혜·임정완(Kwon, J. and J. Lim). 2018. 2015 개정 교육과정에 따른 고등학교 공통영어교과서 문화 내용 분석(An analysis of the topics and cultural contents in high school English textbooks under the national curriculum revised in 2015). 《언어학》(*The Linguistic Association of Korea Journal*) 26-3, 117-137.
- 김경옥·나경희(Kim, G. and K. Rha). 2009 개정 영어과 교육과정에 따른 중학교 영어교과서의 쓰기활동유형 분석(An analysis of writing activities of middle school English textbooks based on the 2009 revised national curriculum). 《영어영문학연구》(*Journal of The English Language and Literature*) 59-4, 407-434.
- 김미혜·이성원(Kim, M. and Lee S.). EBS 자료와 CSAT 및 영어교과서 어휘 분석(Vocabulary analysis on EBS materials, CSAT, and English textbooks). 《예술인문사회 융합 멀티미디어 논문지》(*Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*) 8-4, 417-427.
- 김영숙(Kim, Y.). 2016. 2009개정 교육과정 기반 초등영어 교과서의 파닉스 지도 내용 및 방법 분석(An analysis of phonics activities in English textbooks for Korean elementary school students). 《초등영어교육》(*Primary English Education*) 22-4, 237-257.
- 김은주(Kim, E.). 2001. 제 7차 교육과정에 의한 중학교 1학년 영어교과서의 듣기·말하기 활동 분석 연구(A study on the listening and speaking activities in the middle school English textbooks). 《영어교육》(*English Education*) 56-2, 219-243.

- 김정렬(Kim, J.). 2014. 초중등 영어교육 연계성 문제: 영어교과서의 계열성을 중심으로 (Continuity problems of elementary and secondary English education: Sequence analysis of English textbooks). 《어학연구》(*Language Research*) 50-1, 161-184.
- 김정렬 · 양지윤(Kim, J. and J. Yang). 2012. Coh-Metrix를 통한 초중등 영어교과서 연계성 분석(An analysis of the continuity of elementary and middle school English textbooks using Coh-Metrix). 《영어교육》(*English Teaching*) 67-2, 319-341.
- 김지아 · 최문홍(Kim, J. and M. Choe). 2018. 한국과 일본의 중등학교 영어 교과서 비교 연구(A comparative analysis of English textbooks in Korea and Japan). 《교육문화연구》(*Journal of Education & Culture*) 24-3, 265-284.
- 김지은 · 전문기(Kim, J. and M. Jeon). 2013. 중학교 영어교과서 듣기자료의 Coh-Metrix 측정치에 대한 연계성 분석(A corpus-based analysis of the continuity of the listening materials in middle school English textbooks for Coh-Metrix measures). *Journal of the Korean Data Analysis Society* 15-4, 1987-2000.
- 김진완 · 황종배 · Judy Yin · 이윤희 · 신미경 · 조성옥 · 조현정(Kim, J., J. Hwang, J. Yin, Y. Lee, M. Shin, S. Cho and H. Cho). 2018. Middle school English 1. 서울: 비상교육.
- 문안나(Moon, A.). 2008. 코퍼스 기반 고등학교 영어 교과서의 동사 문법패턴 분석(A corpus-based analysis of grammar patterns of verbs in Korean high school English textbooks). 《언어연구》(*The Journal of Studies in Language*) 23-4, 659-681.
- 민수정 · 정보현(Min, S. and B. Chung). 2012. 고등학교 영어 교과서에 사용된 기본 어휘에 대한 코퍼스 기반 분석(A corpus-based analysis of vocabulary used in high school English textbooks in Korea). 《영어영문학》(*The Journal of English Language and Literature*) 17-1, 143-159.
- 민찬규 · 김윤규 · 정현성 · 이상기 · 최진희 · 박세란 · 염지선 · Walter Foreman(Min, C., Y. Kim, H. Jung, S. Lee, H. Choi, S. Park, J. Yeom and W. Foreman). 2018. *Middle school English 1*. 서울: 지학사.
- 배지영(Bae, J.). 2019. 2015 개정 교육과정이 적용된 초·중등 영어교과서의 읽기 지문 연계성 분석(A continuity analysis of the reading passages in elementary and middle school English textbooks with 2015 revised national curriculum). 《현대영어영문학》(*Modern Studies in English Language & Literature*) 63-2, 81-107.
- 손미영 · 이재근(Son, M. and J. Lee). 2005. 중학교 1학년 영어교과서의 쓰기 활동 자료 분석(An analysis of writing activity materials in the middle school English textbooks). 《영어어문교육》(*English Language & Literature Teaching*) 11-3, 139-164.
- 송미정 · 권진아 · 모윤숙 · 신정아 · 이수하 · 유현주 · 정지윤(Song, M., J. Kwon, Y. Mo, J.

- Shin, S. Lee, H. Yoo and J. Jung). 2018. *Middle school English 1*. 서울: YBM.
- 송자경 · 윤이정(Song J. and L. Yun). 2003. 초등학교 영어 교과서의 연계성 분석: 제7차 교육과정의 의사소통 기능문을 중심으로(An analysis of the sequencing and grading in elementary school English textbooks: Focused on communicative functions and their exponents in the 7th national curriculum). 《초등영어교육》(*Primary English Education*) 9-2, 149-188.
- 송해성(Sohng, H.). 2013. 고등학교 영어 교과서 듣기 과업의 학년별 연계성 연구(Exploring the connectedness of listening tasks according to grades in high school English textbooks). 《영어어문교육》(*English Language & Literature Teaching*) 19-4, 279-305.
- 양현권 · 이창수 · 김기택 · 최정윤 · 고아영(Yang, H., C. Lee, K. Kim, J. Cho and A. Ko). 2018. *Middle school English 1*. 서울: NE능률.
- 유세진 · 김효영(Yu, S. and H. Kim). 2015. 고등학교 영어교과서 읽기 자료에 나타난 시제와 상 분포(Distributions of Tense and Aspect in Readings from Korean high school textbooks). 《외국학연구》(*The Journal of Foreign Studies*) 34, 77-98.
- 이경남(Lee, K.). 2019. 2015 개정 교육과정 기본어휘 목록 및 고등학교 영어 교과서 어휘동사 사용 분석: COCA 핵심학업동사에 기초하여(An analysis of the lexical verbs in the guided word list of the 2015 revised English curriculum and high school English textbooks: Based on COCA core academic verbs). 《외국어교육》(*Foreign Languages Education*) 26-4, 113-133.
- 이소영 · 이현우(Lee, S. and H. Lee). 2019. 2015 개정 영어과 교육과정에 따른 중2 영어 교과서 어휘 분포 분석(An analysis of the distribution of vocabulary in middle school English 2 textbooks based on 2015 revised English national curriculum). 《중등영어교육》(*Secondary English Education*) 12-3, 71-93.
- 이재영 · 안병규 · 오준일 · 배태일 · 김순천 · 박성근 · 신수진(Lee, J., B. An, J. Oh, T. Pae, S. Kim, S. Park and S. Shin). 2018. *Middle school English 1*. 서울: (주)천재교육.
- 이정민 · 현대덕(Lee, J. and T. Hyun). 2013. 중학교 3학년 영어 교과서 읽기 자료 비교 분석(A comparative analysis of the reading materials in the middle school English textbooks for the 3rd year students). 《현대영어교육》(*Modern English Education*) 14-4, 263-286.
- 이지현 · 이현우(Lee, J. and H. Lee). 2016. 영어 I 교과서와 대학수학능력시험 영어 영역에 나타난 부정사의 분포에 대한 분석(Distributional analysis of infinitives in English I textbooks and the CSAT English sections). 《중등영어교육》(*Secondary English Education*) 9-2, 115-139.
- 이혜수 · 전문기(Lee, H. and M. Jeon). 2019. 『걸리버 여행기』와 『통 이야기』를 중심으로

- 로(A corpus stylistic analysis of Jonathan Swift's writing style in Gulliver's Travels and A Tale of a Tub). 《영어학》(*Korean Journal of English Language and Linguistics*) 19-1, 120-141.
- 임인재 · 전문기(Lim, I. and M. Jeon). 2013. 코퍼스에 기반한 중학교 1학년 개정 영어교과서 듣기자료 비교 분석(A corpus-based analysis of the listening materials in middle school English 1 textbooks). 《영어학》(*Korean Journal of English Language and Linguistics*) 13-1, 105-134.
- 전도희 · 김선영(Jeon, D. and S. Kim). 2015. 중등 영어 교과서의 쓰기활동 분석: 성취기준, 활동유형 및 장르분석을 중심으로(An analysis of writing activities in middle school English textbooks: Achievement objectives, activity types and genre). 《중등영어교육》(*Secondary English Education*) 8-1, 95-119.
- 전문기(Jeon, M.). 2011. Coh-Metrix를 이용한 중학교 1학년과 2학년 개정 영어교과서 읽기 자료의 코퍼스 언어학적 연계성 분석(A corpus-based analysis of the continuity of the reading materials in middle school English 1 and 2 textbooks with Coh-Metrix). 《언어과학연구》(*The Journal of Linguistic Science*) 56, 201-218.
- 전문기(Jeon, M.). 2015. 자동화된 언어분석 프로그램을 활용한 중학교 영어교과서의 학년 간 연계성 분석(An analysis of the continuity among middle school English textbooks with an automated language analysis program). 《현대영어교육》(*Modern English Education*) 16-1, 195-218.
- 전문기 · 임인재(Jeon, M. and I. Lim). 2009. 코메트릭스(Coh-Metrix)를 이용한 중학교 1학년 개정 영어 교과서의 코퍼스 언어학적 비교 분석(A corpus-based analysis of middle school English 1 textbooks with Coh-Metrix). 《영어교육연구》(*English Language Teaching*) 21-4, 265-292.
- 전문기 · 최윤희(Jeon, M. and Y. Choe). 2019. 코메트릭스를 활용한 한국 대학생들의 영어 논설문과 설명문 요약문 비교 분석(A Coh-Metrix analysis of Korean EFL learners' summary writings in the English argumentative and expository texts). 《영어학》(*Korean Journal of English Language and Linguistics*) 19-3, 539-559.
- 정은혁 · 민주경(Jung, E. and J. Min). 2005. 중학교 영어 문화교육 연구: 학습자와 교과서 분석 중심으로(Culture teaching in secondary English education with a focus on learners and textbooks). 《영어학연구》(*English Language and Linguistics*) 19, 37-63.
- 조원향(Cho, W.). 2003. 제 7차 교육과정에 따른 중학교 2학년 영어 교과서의 읽기 자료 분석(An analysis of the English materials of the middle school English textbooks for the 2nd year students based on the 7th curriculum). 《영미어문학》(*Studies in British and American language and Literature Association*) 19-1, 151-170.

- 최인철 · 박태자 · 서원화 · 홍우정 · 강유나 · 송해리 · 김지윤 · 이정하(Choi, I., T. Park, W. Seo, W. Hong, Y. Kang, H. Song, J. Kim, H. Lee and J. Park). 2018. *Middle school English 1*. 서울: 금성출판사.
- 허은선 · 김혜정(Heo, E. and H. Kim). 2015. 초등학교 6학년 영어 교과서 문화내용 분석 (Analysis of cultural contents in the 6th grade's elementary school English textbooks). 《영어영문학》(*The Mirae Journal of English Language and Literature*) 20-2, 247-267.
- Baayen, R. H., R. Piepenbrock and L. Gulikers. 1995. *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Brill, E. 1995. *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Cirilo, R. 1981. Referential coherence and text structure in story comprehension. *Journal of Verbal Learning and Verbal Behavior* 20, 358-367.
- Coltheart, M. 1981. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* 33, 497-505.
- Graesser, A. C., M. Jeon, Z. Cai and D. S. McNamara. 2008. Automatic analyses of language, discourse, and situation models. In J. Auracher and W. van Peer, eds., *New Beginnings in Literary Studies*, 72-88. Cambridge: Cambridge Scholars Publishing.
- Graesser, A. C., M. Jeon, Y. Yan and Z. Cai. 2007. Discourse cohesion in text and tutorial dialogue. *Information Design Journal* 15(3), 199-213.
- Graesser, A. C., D. S. McNamara, M. M. Louwerse and Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers* 36(2), 193-202.
- Haberlandt, K. and G. Bingham. 1978. Verbs contribute to the coherence of brief narrative: Reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior* 17, 419-425.
- Jeon, M. 2011. A corpus-based analysis of the continuity of the reading materials in middle school English 1 and 2 textbooks with Coh-Metrix. *The Journal of Linguistic Science*, 56, 201-218. (in Korean).
- Jurafsky, D. and 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.
- Klare, G. R. 1974-1975. Assessing readability. *Reading Research Quarterly* 10, 62-102.

- Krashen, S. D. 1985. *The Input Hypothesis: Issues and Implications*. Addison-Wesley Longman Ltd.
- Landauer, T. K. 2007. LSA as a theory of meaning, In T. K. Landauer, D. S. McNamara and W. Kintsch, eds., *Handbook of Latent Semantic Analysis*, 3-34. Mahwah, NJ: Erlbaum.
- Landauer, T. K., P. W. Foltz and D. Laham. 1998. Introduction to latent semantic analysis. *Discourse Processes* 25, 259-284.
- Meyer, C. F. 2002. *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Millis, K. and M. Just. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language* 33, 128-147.
- Murray, J. (1997) Connective and narrative text: The role of continuity. *Memory & Cognition* 25(2), 227-236.

예시언어(Examples in): 영어(English)
적용가능 언어(Applicable Languages): 영어(English)
적용가능 수준(Applicable Level): 모든 수준(All)

류지수(Ryu, Jisu), 초빙연구원(Visiting Scholar)
다언어다문화연구소(The Konkuk Research Institute for Multilingualism and Multiculturalism)
건국대학교(Konkuk University)
서울시(Seoul 05029, Korea)
광진구 능동로 120(120 Neungdong-ro, Gwangjin-gu)
Tel: 02) 450-3356
E-mail: ghi03169@gmail.com

전문기(Jeon, Moongee), 교수(Professor)
영어영문학과(Dept. of English), 건국대학교(Konkuk University)
서울시(Seoul 05029, Korea)
광진구 능동로 120(120 Neungdong-ro, Gwangjin-gu)
Tel: 02) 450-3356
E-mail: mjeon1@konkuk.ac.kr

논문 접수(Received): 2020년 2월 5일(February 5, 2020)
논문 수정(Revised): 2020년 3월 10일(March 10, 2020)
게재 확정(Accepted): 2020년 3월 20일(March 20, 2020)

부록 A

A 교과서 단원별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	8단원	9단원	평균
어휘수	148	157	184	194	175	182	229	200	197	185
문장수	28	26	23	30	29	27	28	33	29	28
문장길이	5.29	6.04	8.00	6.47	6.03	6.74	8.18	6.06	6.79	6.62
어휘길이	1.31	1.27	1.33	1.26	1.38	1.37	1.33	1.22	1.50	1.33
어휘빈도	2.56	2.39	2.46	2.48	2.40	2.51	2.39	2.43	2.33	2.44
심상성	421	433	470	431	461	435	425	431	456	440
구체성	379	386	436	401	443	398	388	407	419	406
습득나이	263	313	291	274	279	275	280	313	281	285
타입토큰	0.65	0.71	0.64	0.70	0.67	0.70	0.79	0.57	0.74	0.68
명사수	358	318	359	253	234	253	314	285	310	298
동사수	128	127	87	186	114	126	144	135	152	133
형용사수	81	76	92	52	69	49	48	95	71	70
부사수	54	51	33	93	69	82	70	70	51	64
1인칭대명사	77.8	35.1	46.2	12.9	2.9	30.3	37.2	7.5	5.1	28.3
2인칭대명사	13.5	9.6	2.7	7.7	0.0	16.5	0.0	2.5	0.0	5.8
3인칭대명사	0.0	51.0	0.0	33.5	31.5	8.3	21.9	5.0	12.7	18.2
인과접속사	6.8	6.4	21.7	10.3	22.9	16.5	39.3	55.0	15.2	21.6
부가접속사	6.8	0.0	10.9	5.2	5.7	5.5	4.4	0.0	0.0	4.3
시간접속사	13.5	6.4	21.7	15.5	17.1	33.0	17.5	20.0	10.2	17.2
FRE	91	93	87	94	84	84	86	97	73	88
FKGL	1.94	1.73	3.18	1.78	3.08	3.25	3.32	1.17	4.72	2.69
명사구밀도	0.72	0.56	0.84	0.62	0.78	0.64	0.65	0.92	0.82	0.73
본동사앞	1.07	1.39	1.52	1.77	1.72	1.26	1.68	1.88	1.79	1.56
논항중복 (로컬)	0.41	0.52	0.32	0.21	0.14	0.15	0.37	0.31	0.36	0.31
논항중복 (글로벌)	0.32	0.22	0.34	0.20	0.15	0.15	0.27	0.22	0.21	0.23
LSA 코사인 (로컬)	0.17	0.09	0.17	0.11	0.12	0.08	0.12	0.16	0.16	0.13
LSA 코사인 (글로벌)	0.00	0.00	0.00	0.14	0.11	0.16	0.06	0.08	-0.01	0.06

부록 B

B 교과서 단위별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	8단원	평균
어휘수	130	169	239	172	220	176	172	248	191
문장수	23	37	39	23	28	21	21	28	28
문장길이	5.65	4.57	6.13	7.48	7.86	8.38	8.19	8.86	7.14
어휘길이	1.38	1.30	1.12	1.33	1.22	1.26	1.26	1.22	1.26
어휘빈도	2.51	2.59	2.22	2.41	2.74	2.41	2.51	2.62	2.50
심상성	400	424	486	459	405	489	453	436	444
구체성	367	375	460	417	365	451	414	396	405
습득나이	254	283	227	262	287	254	325	258	269
타입토큰	0.80	0.74	0.51	0.86	0.60	0.77	0.74	0.70	0.71
명사수	262	284	314	326	223	267	238	262	272
동사수	131	172	167	122	123	142	157	137	144
형용사수	100	59	42	87	100	68	99	60	77
부사수	69	59	21	41	59	51	64	36	50
1인칭대명사	57.8	5.9	18.8	0.0	0.0	42.7	61.1	60.5	30.9
2인칭대명사	11.6	29.6	14.7	23.3	2.3	2.8	0.0	10.1	11.8
3인칭대명사	11.6	11.8	29.3	0.0	18.2	0.0	11.6	8.1	11.3
인과접속사	15.4	23.7	29.3	5.8	4.5	45.5	11.6	8.1	18.0
부가접속사	0.0	5.9	0.0	0.0	0.0	0.0	0.0	8.1	1.7
시간접속사	23.1	29.6	33.5	11.6	13.6	62.5	40.7	32.3	30.9
FRE	85	92	100	87	96	92	92	95	92
FKGL	2.86	1.56	0.03	2.97	1.85	2.56	2.43	2.24	2.06
명사구밀도	0.72	0.48	0.69	0.86	0.96	0.60	0.73	0.61	0.71
본동사앞	0.96	0.87	1.44	1.57	2.00	1.71	1.48	2.93	1.62
논항중복 (로컬)	0.23	0.08	0.16	0.14	0.26	0.55	0.55	0.37	0.29
논항중복 (글로벌)	0.18	0.09	0.20	0.14	0.27	0.51	0.54	0.24	0.27
LSA 코사인 (로컬)	0.11	0.11	0.09	0.14	0.20	0.28	0.13	0.15	0.15
LSA 코사인 (글로벌)	0.00	0.00	0.08	0.00	0.18	0.00	0.00	0.07	0.04

부록 C

C 교과서 단원별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	8단원	평균
어휘수	143	148	195	186	191	188	272	208	191
문장수	30	22	29	27	21	21	35	23	26
문장길이	4.77	6.73	6.72	6.89	9.10	8.95	7.77	9.04	7.50
어휘길이	1.31	1.33	1.16	1.20	1.23	1.29	1.22	1.27	1.25
어휘빈도	2.53	2.40	2.51	2.15	2.25	2.81	2.61	2.68	2.49
심상성	447	421	369	489	462	406	464	434	437
구체성	382	375	337	465	455	366	430	388	400
습득나이	275	276	234	251	251	277	230	263	257
타입토큰	0.58	0.74	0.73	0.77	0.69	0.75	0.58	0.64	0.68
명사수	350	297	185	290	304	234	261	293	277
동사수	84	122	164	161	120	144	173	96	133
형용사수	77	74	62	32	63	90	96	77	71
부사수	35	41	87	38	37	74	59	48	52
1인칭대명사	77.0	33.8	64.2	0.0	0.0	18.6	12.9	21.7	28.5
2인칭대명사	7.0	20.3	12.8	16.1	2.6	10.6	7.4	9.6	10.8
3인칭대명사	0.0	0.0	20.5	8.1	10.5	42.6	20.2	9.6	13.9
인과접속사	0.0	27.0	41.0	10.8	26.2	31.9	22.1	28.8	23.5
부가접속사	7.0	0.0	10.3	5.4	0.0	10.6	11.0	0.0	5.5
시간접속사	7.0	20.3	51.3	10.8	5.2	69.1	33.1	24.0	27.6
FRE	91	87	100	98	94	88	96	90	93
FKGL	1.70	2.74	0.71	1.30	2.41	3.16	1.80	2.97	2.10
명사구밀도	0.73	0.67	0.29	0.74	0.98	0.60	0.81	0.72	0.69
본동사앞	1.20	0.64	1.14	0.48	2.52	1.43	2.17	1.96	1.44
논항중복 (로컬)	0.48	0.33	0.29	0.12	0.40	0.20	0.32	0.41	0.32
논항중복 (글로벌)	0.33	0.27	0.34	0.07	0.27	0.21	0.29	0.25	0.25
LSA 코사인 (로컬)	0.22	0.19	0.09	0.19	0.21	0.09	0.18	0.12	0.16
LSA 코사인 (글로벌)	0.26	0.08	0.07	0.00	0.00	0.19	0.15	0.00	0.09

부록 D

D 교과서 단원별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	8단원	평균
어휘수	207	222	212	205	230	290	273	299	242
문장수	39	43	32	27	33	34	34	46	36
문장길이	5.31	5.16	6.63	7.59	6.97	8.53	8.03	6.50	6.84
어휘길이	1.28	1.38	1.29	1.26	1.29	1.37	1.32	1.30	1.31
어휘빈도	2.72	2.67	2.57	2.51	2.24	2.39	2.50	2.54	2.52
심상성	409	407	423	406	457	434	410	436	423
구체성	353	382	381	368	435	393	377	421	389
습득나이	270	294	257	294	252	300	272	209	268
타입토큰	0.59	0.64	0.65	0.63	0.68	0.71	0.64	0.61	0.64
명사수	300	243	245	200	204	276	216	221	238
동사수	169	131	132	156	126	107	92	174	136
형용사수	68	90	80	59	109	93	77	70	81
부사수	48	59	85	88	52	52	48	77	64
1인칭대명사	24.2	33.8	44.9	9.8	19.6	50.1	7.3	26.8	27.0
2인칭대명사	9.7	9.0	2.4	58.6	28.3	1.7	14.7	10.0	16.8
3인칭대명사	50.8	15.8	14.2	0.0	0.0	10.4	25.7	20.1	17.1
인과접속사	9.7	40.5	23.6	43.9	26.1	27.6	22.0	20.1	26.7
부가접속사	33.8	4.5	18.9	4.9	4.3	0.0	14.7	23.4	13.1
시간접속사	4.8	22.5	28.3	19.5	30.4	31.0	7.3	13.4	19.7
FRE	93	85	91	92	91	82	87	90	89
FKGL	1.58	2.74	2.19	2.28	2.32	3.89	3.14	2.26	2.55
명사구밀도	0.41	0.60	0.59	0.74	0.59	0.70	0.63	0.51	0.59
본동사앞	1.13	1.30	1.53	2.33	1.70	1.35	1.91	1.65	1.61
논항중복 (로컬)	0.42	0.17	0.26	0.42	0.28	0.73	0.39	0.36	0.38
논항중복 (글로벌)	0.22	0.10	0.15	0.34	0.24	0.47	0.22	0.19	0.24
LSA 코사인 (로컬)	0.22	0.08	0.18	0.18	0.17	0.22	0.22	0.18	0.18
LSA 코사인 (글로벌)	0.00	0.08	0.04	0.00	0.00	0.00	0.10	0.18	0.05

부록 E

E 교과서 단위별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	평균
어휘수	140	138	175	179	189	200	248	181
문장수	23	25	32	25	25	27	28	26
문장길이	6.09	5.52	5.47	7.16	7.56	7.41	8.86	6.87
어휘길이	1.31	1.39	1.25	1.32	1.42	1.35	1.40	1.35
어휘빈도	2.43	2.29	2.46	2.34	2.37	2.32	2.43	2.38
심상성	425	496	438	439	425	445	447	445
구체성	369	456	412	410	393	413	400	408
습득나이	328	228	244	291	302	284	295	282
타입토큰	0.75	0.76	0.62	0.73	0.65	0.68	0.72	0.70
명사수	314	283	183	307	296	280	290	279
동사수	107	188	114	156	138	155	113	139
형용사수	43	22	109	101	90	65	93	74
부사수	29	22	80	39	32	75	40	45
1인칭대명사	89.4	72.5	45.8	53.1	2.6	5.0	22.2	41.5
2인칭대명사	3.6	14.5	28.6	0.0	5.3	7.5	28.3	12.5
3인칭대명사	0.0	10.9	14.3	16.8	21.2	22.5	2.0	12.5
인과접속사	0.0	50.7	62.9	33.5	10.6	5.0	28.2	27.3
부가접속사	0.0	0.0	5.7	5.6	5.3	5.0	12.1	4.8
시간접속사	14.3	50.7	40.0	27.9	5.3	40.0	20.2	28.3
FRE	90	84	95	88	79	85	79	86
FKGL	2.21	2.98	1.31	2.76	4.15	3.23	4.42	3.01
명사구밀도	0.57	0.56	0.49	0.59	0.45	0.78	0.67	0.59
본동사앞	1.61	1.20	0.91	1.72	1.80	1.78	1.75	1.54
논항중복 (로컬)	0.68	0.29	0.19	0.46	0.63	0.35	0.59	0.46
논항중복 (글로벌)	0.57	0.19	0.17	0.44	0.38	0.24	0.30	0.33
LSA 코사인 (로컬)	0.17	0.21	0.16	0.11	0.35	0.20	0.19	0.20
LSA 코사인 (글로벌)	0.22	0.18	0.22	0.00	0.23	0.00	0.00	0.12

부록 F

F 교과서 단원별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	8단원	평균
어휘수	193	223	202	235	249	228	285	259	234
문장수	30	36	29	32	34	31	36	35	33
문장길이	6.43	6.19	6.97	7.34	7.32	7.36	7.92	7.40	7.12
어휘길이	1.23	1.19	1.22	1.30	1.27	1.29	1.26	1.37	1.26
어휘빈도	2.55	2.43	2.58	2.44	2.50	2.44	2.46	2.41	2.47
심상성	457	443	450	459	459	463	449	458	455
구체성	419	412	433	423	423	406	422	424	420
습득나이	260	314	208	244	243	256	234	270	254
타입토큰	0.73	0.61	0.62	0.75	0.69	0.75	0.58	0.72	0.68
명사수	337	287	277	264	241	250	337	263	282
동사수	88	152	89	162	129	167	172	139	137
형용사수	78	67	69	111	88	44	49	89	74
부사수	62	58	89	60	52	53	46	35	57
1인칭대명사	64.8	2.2	0.0	34.1	22.1	0.0	12.3	21.3	19.6
2인칭대명사	2.6	15.7	17.3	0.0	6.0	15.4	10.5	0.0	8.4
3인칭대명사	0.0	40.4	37.2	8.5	6.0	57.1	14.0	9.7	21.6
인과접속사	10.4	13.5	34.7	12.8	12.0	26.3	31.6	23.2	20.5
부가접속사	5.2	4.5	14.9	12.8	0.0	4.4	10.5	15.4	8.5
시간접속사	25.9	26.9	9.9	12.8	16.1	30.7	21.1	30.9	21.8
FRE	96	100	97	89	92	91	93	84	93
FKGL	1.41	0.90	1.50	2.64	2.24	2.44	2.32	3.43	2.11
명사구밀도	0.70	0.86	0.61	0.74	0.53	0.59	0.75	0.58	0.67
본동사앞	1.53	1.11	1.35	1.34	2.21	1.58	2.72	2.06	1.74
논항중복 (로컬)	0.52	0.49	0.46	0.32	0.49	0.43	0.37	0.27	0.42
논항중복 (글로벌)	0.28	0.35	0.24	0.16	0.38	0.20	0.28	0.20	0.26
LSA 코사인 (로컬)	0.25	0.14	0.18	0.12	0.14	0.27	0.14	0.11	0.17
LSA 코사인 (글로벌)	0.18	-0.01	0.00	0.07	0.00	0.00	0.08	0.00	0.04

부록 G

G 교과서 단원별 분석 결과

	1단원	2단원	3단원	4단원	5단원	6단원	7단원	8단원	평균
어휘수	216	212	199	240	268	278	273	363	256
문장수	44	44	23	32	37	34	31	46	36
문장길이	4.91	4.82	8.65	7.50	7.24	8.18	8.81	7.89	7.25
어휘길이	1.26	1.19	1.46	1.40	1.38	1.39	1.25	1.22	1.32
어휘빈도	2.45	2.43	2.53	2.46	2.48	2.52	2.51	2.47	2.48
심상성	452	452	424	442	439	430	417	432	436
구체성	408	417	391	417	404	386	391	391	401
습득나이	265	241	326	260	272	292	266	280	275
타입토큰	0.71	0.63	0.77	0.76	0.80	0.78	0.81	0.76	0.75
명사수	282	203	342	342	295	295	278	267	288
동사수	97	179	101	113	131	119	125	118	123
형용사수	102	33	85	71	67	90	77	61	73
부사수	51	113	50	38	78	40	62	66	62
1인칭대명사	53.3	75.5	7.5	8.3	28.0	43.2	22.0	41.4	34.9
2인칭대명사	20.9	0.0	0.0	10.4	0.0	10.8	3.7	12.4	7.3
3인칭대명사	0.0	14.2	47.8	16.7	14.9	1.8	51.3	11.0	19.7
인과접속사	4.6	0.0	20.1	33.3	22.4	21.6	25.6	30.3	19.7
부가접속사	13.9	4.7	5.0	20.8	3.7	3.6	3.7	5.5	7.6
시간접속사	32.4	9.4	30.2	37.5	63.4	32.4	29.3	33.1	33.5
FRE	96	100	75	80	83	81	92	96	88
FKGL	1.13	0.37	4.98	3.90	3.48	3.98	2.63	1.82	2.79
명사구밀도	0.60	0.38	0.83	0.64	0.65	0.76	0.60	0.68	0.64
본동사앞	0.89	1.55	1.70	1.66	1.68	1.65	2.68	1.67	1.68
논항중복 (로컬)	0.26	0.21	0.32	0.29	0.22	0.52	0.43	0.33	0.32
논항중복 (글로벌)	0.14	0.24	0.19	0.16	0.14	0.38	0.22	0.27	0.22
LSA 코사인 (로컬)	0.14	0.14	0.19	0.15	0.09	0.12	0.11	0.15	0.13
LSA 코사인 (글로벌)	0.17	0.09	0.00	0.03	0.00	0.05	0.00	0.06	0.05