

Coh-Metrix를 활용한 초등학교 6학년 영어학습자의 영작문 분석 연구*

김소정**

건국대학교

전문기

건국대학교

Kim, Sojung & Jeon, Moongee. (2016). An analysis study of English writing of elementary school 6th grade English language learners using Coh-Metrix. *Modern English Education*, 17(3), 263-287.

The purpose of this study is to analyze English writings collected from 6th grade students in a Korean elementary school using Coh-Metrix. Coh-Metrix is a computer tool that has been widely used to study English corpus with various psycholinguistic measures. A total of 176 English writing samples were collected from Korean elementary school 6th grade students. Specifically, the English writing corpus consisted of three groups classified by the 6th grade students' English proficiency performance assessment tests. The findings of our study indicated that there were statistically significant differences among the three groups for number of words, number of sentences, Flesch Reading Ease score, Flesch-Kincaid Grade Level score, additive connective, pronoun ratio, and the third person pronoun measures. On the other hand, there were not statistically significant differences among the groups for mean sentence length, word frequency, syntactic complexity, type-token ratio, co-referential cohesion, semantic cohesion, casual connective, temporal connective, the first person pronoun, the second person pronoun, and word information measures. Implications for English writing education for elementary school students are discussed.

[English writing analysis/corpus/Coh-Metrix/computer tool/
영작문 분석/코퍼스/코메트릭스/컴퓨터 도구]

* 이 논문은 2014학년도 건국대학교의 연구년교원 지원에 의하여 연구되었음.

이 논문은 제1저자 김소정의 석사 학위논문 자료를 바탕으로 확대 논의한 것임.

** 제1저자: 김소정, 교신저자: 전문기

I. 서론

2009년 개정 교육과정에 따르면 영어 글쓰기 교육은 초등학교 5학년 과정부터 시작되는데 기초 교육과정이 지나 6학년이 되면 본격적으로 감사 편지 등과 같은 영어 실용문과 친숙한 주제로 영어 글쓰기를 작성하는 교육이 수행된다(교육과학기술부, 2011). 이러한 영어 글쓰기 교육이 초등학교 교육과정을 통해 실행되는 이유는 영어교육의 중요성이 강조되면서 자연스럽게 영어 환경에 노출되는 빈도가 증가됨에 따라 영어를 읽고 듣는 능력은 비교적 잘 학습이 되고 있지만 쓰고 말하는 능력은 상대적으로 잘 학습이 되고 있지 않기 때문이다. 다시 말하면 영어로 쓰고 말하는 언어 표현 및 산출 능력은 읽고 듣는 언어 수용 및 이해 능력보다 습득하기가 어렵다(이인호, 배주경, 김성혜, 2015; 전영주, 2010). 그럼에도 불구하고 언어로 자기의 생각과 의견들을 잘 표현하는 능력이 사회 곳곳에서 점점 더 요구됨에 따라 글쓰기 및 말하기 능력의 중요성이 더 부각 되고 있다. 특히 국제화 및 다문화 시기를 살아가면서 영어로 자기의 생각들을 표현하는 능력은 매우 중요하다고 하겠다.

이러한 맥락 속에서 우리나라 초등학교 교육과정에 영어 교육이 도입되기 시작했고, 특히 초등학교 고학년의 영어 교육 과정(5, 6학년 과정) 중 쓰기 및 말하기 능력에 기초한 표현 기능의 비중이 점차 커지고 있다. 그러나 2014년에 전국적으로 실시되었던 국가 수준의 학업 성취도 평가 결과를 살펴보면, 흥미롭게도 중학교 3학년과 고등학교 2학년의 영어 쓰기 평가 영역의 수행점수가 다른 평가 영역들(말하기, 듣기, 읽기)의 점수들에 비해 현저하게 낮았다. 구체적으로 두 학년 모두 쓰기 평가 영역의 평균 점수가 가장 높은 점수를 보였던 듣기 평가 영역의 점수보다 약 20점 가량 낮았다. 이러한 결과는 현 영어 교육 과정에서 영어 글쓰기에 대한 교육이 더욱 강화되어야 함을 시사한다(이인호, 배주경, 김성혜, 2015).

이처럼 영어 글쓰기 교육의 필요성이 대두되면서 영어 교육 시 교과서에 대한 의존도가 높은 우리나라와 같은 EFL(English as a Foreign Language) 영어 교육 환경 속에서 영어 교과서 쓰기 영역에 대한 분석을 시도한 연구들이 있었다. 예를 들면, 영어 교과서 내에 포함된 쓰기 활동을 체계적으로 분석하거나(나임정, 김태은, 2014; 유지연, 나경희, 2013) 쓰기 활동을 중심으로 학년 간의 연계성을 분석한 연구(김유주, 2009) 등이 있었다.

또한 본 연구의 주제와 관련하여 EFL 영어학습자가 작성한 영어 쓰기 자료를 분석한 연구들이 있었다(김영민, 배영직, 2007; 김지나, 진송화, 2015; 김혜리, 정지숙, 2008; 김혜리, 김소중, 2010; 노경진, 김신혜, 2012; 송다진, 김태은, 2016; 송주영, 이동한, 2006; 우연순, 2013; 이동한, 2004; 정동빈, 강시경, 2008). 예를 들면, 초등학생이 작성한 영어 일기 쓰기 자료에 사용된 어휘들을 사용하는 능력을 분석한 연구(김영민, 배영직, 2007), 초등학생을 대상으로 대화식 저널쓰기가 쓰기 양상에 미치는 영향을 분석한 연구(김혜리, 정지숙, 2008), 초등학생을 대

상으로 문학 텍스트가 영어 쓰기 활동에 미치는 영향을 분석한 연구(김혜리, 김소중, 2010), 초등학생들이 작성한 영어 쓰기에 나타난 오류들을 분석한 연구(송주연, 이동한, 2006; 이동한, 2004), 영어 동화를 통한 저널 쓰기 활동이 쓰기 능력에 미치는 영향을 분석한 연구(노경진, 김신혜, 2012) 등이 있다. 이처럼 영어 쓰기 활동 자료들은 다양한 측면에서 분석되었다.

또 다른 측면에서 코퍼스언어학(Lindquist, 2009; Meyer, 2002)과 전산언어학(Allen, 1995; Jurafsky & Martin, 2008)을 접목시킨 융복합적 연구방법이 다양한 유형의 영어 텍스트 및 영어 쓰기 자료를 분석하는데 활발하게 활용되고 있다. 이러한 컴퓨터과학에 기반한 텍스트 분석 알고리즘과 대용량의 코퍼스 자료가 영어 교육 분야에 광범위하게 적용되는 이유는 방대한 크기의 코퍼스 자료를 텍스트 분석에 최적화된 컴퓨터 알고리즘을 적용해 분석함으로써 분석 과정의 효율성을 증대시키고 분석 결과의 객관성을 증대시킬 수 있기 때문이다(전문기, 2014, 2015).

본 연구는 텍스트 분석 도구인 Coh-Metrix 시스템(Graesser, Jeon, Cai, & McNamara, 2008)을 활용하여 초등학교 6학년 영어학습자의 영어 글쓰기 자료를 분석하였다. 기존의 초등학교 영어 글쓰기 분석 연구들은 대부분 영어 글쓰기에 나타난 어휘들의 특성(김영민, 배영직, 2007)과 문법 오류 유형(송주영, 이동한, 2006; 이동한, 2004)을 분석하거나 대화식 저널 쓰기와 같은 쓰기 활동이 쓰기 양상에 미치는 효과(김혜리, 정지숙, 2008) 등을 분석하였다. 다시 말하면, 초등학교 영어학습자가 작성한 영어 글쓰기 자료의 언어적 특성을 광범위한 유형의 심리언어학적 측정치(어휘, 문장, 텍스트 수준 측정치)를 제공하는 Coh-Metrix와 같은 도구를 활용하여 객관적으로 분석한 연구들은 상대적으로 부재하였다. 따라서 본 연구는 Coh-Metrix를 활용하여 초등학교 영어학습자가 작성한 영어 글쓰기 자료의 언어적 특성을 보다 객관적이고 체계적으로 분석하고자 하였다.

본 연구는 Coh-Metrix 시스템이 제공하는 광범위한 유형의 측정치들 중에서 텍스트 및 영작문 자료의 정합성(cohesion)에 직간접적으로 중요한 영향을 미치는 총 21개의 언어적 및 심리언어적 측정치들을 선택하여 영어 쓰기 자료를 분석하는데 활용하였다. 구체적으로 본 연구에 이용된 Coh-Metrix 측정치는 기초 산출치(단어수, 문장수, 평균문장길이), 단어빈도수, 읽기 가독성 지표 점수(Flesch Reading Ease 점수, Flesch-Kincaid Grade Level 점수), 통사적 복잡성(본동사 앞 단어수, 명사구 밀도), 어휘다양성(타입-토큰 비율), 공통참조 정합성 측정치(논항 반복 비율, LSA 측정치), 접속사 정보 측정치(인과적 접속사, 부가적 접속사, 시간적 접속사), 대명사 정보 측정치(대명사 비율, 1인칭 대명사, 2인칭 대명사, 3인칭 대명사), 어휘 정보 측정치(심상성, 구체성, 습득나이)로 구성되었다.

요약하면, 본 연구는 초등학교 6학년 EFL 영어학습자들이 작성한 영작문 자료를 분석하기 위해 다음과 같은 연구 목표를 설정하였다.

- 1) 본 연구는 초등학교 6학년 영어학습자들이 작성한 영어 글쓰기 자료를 영어 수행평가 점수에 따라 세 집단으로 분류하여 Coh-Metrix 각 측정치에 대해 각 집단 간에 통계적으로 유의미한 차이가 있는지를 분석하고자 하였다.
- 2) 본 연구에서 얻어진 결과가 시사하는 교육적 함의점들을 제시하고자 하였다.

II. 이론적 배경

1. EFL 영어학습자 영어 쓰기 자료 분석 연구

영어 쓰기 교육의 중요성과 필요성이 대두되면서 EFL 영어학습자가 작성한 영어 쓰기 자료를 분석한 다양한 연구들이 있었다(김지나, 진송화, 2015; 김혜리, 경지숙, 2008; 김혜리, 김소중, 2010; 노경진, 김신혜, 2012; 송다진, 김태은, 2016; 송주영, 이동한, 2006; 우연순, 2013; 이동한, 2004; 정동빈, 강시경, 2008).

우연순(2013)은 초등학교 6학년 영어 학습자들의 글쓰기 능력에 영향을 미치는 피드백 유형(직접 수정 피드백, 간접 수정 피드백)의 효과를 분석하였다. 구체적으로 초등학교 6학년 영어 학습자들이 작성한 영어 글쓰기 자료를 국가 성취도 평가 점수에 기반해 세 집단(상위 집단, 중위 집단, 하위 집단)으로 분류하여 집단 간의 차이점을 분석하였다. 연구 결과, 교사가 오류가 있는 부분을 직접 수정하여 피드백을 준 경우(직접 수정 피드백 유형)에는 상위 집단과 하위 집단의 바른 교정률에 차이가 없었다. 반면에 교사가 오류가 있는 부분에 대해 집적적으로 수정하지 않고 위치 표시와 힌트 주기와 같은 간접적 피드백을 준 경우에는 하위 집단의 교정률이 상위 집단의 교정률에 비해 매우 낮았다. 이러한 결과는 영어 학습자들의 수준에 따라 서로 다른 피드백 전략을 사용해야 함을 시사한다.

김지나와 진송화(2015)는 대학부설 영재교육원의 초등학교 영재들(4, 5, 6학년 대상)을 대상으로 그들이 작성한 영어 글쓰기 자료(에세이 자료)에 나타난 원어민 교사의 피드백 유형(의미 중심 피드백, 형태기반 피드백)에 대해 분석하였다. 그들의 연구에서 의미중심 피드백은 간접 피드백으로 분류되었고 형태기반 피드백은 직접 피드백으로 구분되었다. 분석 결과, 에세이 평가 점수가 매우 높은 학생들의 에세이에는 형태기반 피드백이 제시되었다. 그러나 에세이 점수가 매우 높은 학생들을 제외한 다른 학생들의 에세이에는 의미 중심 피드백과 형태기반 피드백이 모두 제시되었다. 그들의 연구 결과, 이러한 학생들을 중 일부 학생들만 간접 피드백 유형인 의미 중심 피드백에 따라 올바르게 교정을 하였다. 이러한 결과 또한 EFL 영어 학습자들의 수준에 따라 적절한 피드백을 제시해야 함을 시사한다.

영어 쓰기 자료에 나타난 오류들을 분석한 연구들도 있었다. 예를 들면, 송주

영과 이동한(2006)은 초등학교 5학년 영어 학습자가 작성한 영어 일기 자료에 나타난 오류들의 유형을 분석하였다. 구체적으로 그들은 초등학교 학생들이 작성한 일기 자료들을 초등학교 영어 능력 검증시험(PELT) 점수에 따라 세 개의 집단(상 집단, 중 집단, 하 집단)으로 분류하였다. 그들은 영어 일기 자료에 나타난 오류들을 오류의 형태(문장 단위, 주어, 동사, 기타 오류)와 오류의 원인(회피에 의한 오류, 모국어 전이에 의한 오류, 과잉일반화에 의한 오류)에 따라 분류하였다. 그들의 연구 결과, 상 집단에 속한 학생들의 일기 자료에는 어순 오류가 가장 많이 발견되었다. 반면에 중 집단에 속한 학생들의 일기 자료에는 단어 조합 문장 오류가 많이 발견되었다. 그리고 하 집단에 속한 학생들의 일기 자료에는 한글 혼용문장 오류가 빈번하게 발견되었다. 이러한 결과는 학생들의 수준에 따른 영어 쓰기 교육의 필요성을 시사한다.

김영민과 배영직(2007)은 초등학교 5, 6학년 영어 학습자들이 작성한 일기 자료에 나타난 어휘의 양을 분석하였다. 구체적으로 그들은 초등학교 학생들이 자유롭게 작성한 영어 일기 자료를 교사가 평가한 영어 수행 평가 점수에 의해 세 개의 집단(상 집단, 중 집단, 하 집단)으로 분류하였다. 그들의 연구 결과, 상 집단에 속한 일기 자료에 사용된 어휘량이 하 집단에 속한 어휘량보다 높았다. 이러한 결과는 하위 집단에 속한 학생들을 위한 별도의 어휘 학습 전략이 필요함을 시사한다.

송다진과 김태은(2016)은 초등학교 6학년 영어 학습자들이 자유롭게 작성한 영어 쓰기 자료들에 나타난 오류들의 유형을 분석하였다. 구체적으로 그들은 초등학교 6학년 영어 학습자들이 작성한 쓰기 자료들을 1학기 수행평가 및 상시평가의 점수를 바탕으로 세 개의 집단(상위 집단, 중위 집단, 하위 집단)으로 구분하였다. 그들의 연구 결과, 상위 집단 학생들의 영어 쓰기 자료에는 구두점의 오류가 다른 집단들에 비해 상대적으로 적게 발견되었다. 그러나 상위 집단의 쓰기 자료에는 중위 집단에 비해 전치가 오류가 근소하게 높게 나타났다. 이러한 결과는 영어 학습자들의 수준에 따라 영어 쓰기 전략을 다르게 해야 함을 제시한다.

지금까지 기술한 선행 연구들은 전술한 바와 같이 영어 글쓰기 자료에 나타난 어휘 특성과 오류를 분석하거나 피드백 유형이 영어 글쓰기에 미치는 효과 등에 대해 분석하였다. 이러한 선행 연구들의 결과는 영어 글쓰기 자료의 언어적 특성을 다양한 유형의 언어적 측정치에 기반해 객관적이고 체계적으로 분석할 필요성을 시사한다. 본 연구는 초등학교 6학년 학생들이 작성한 영어 글쓰기 자료를 영어 학습자의 수준에 따라 분류한 후 광범위한 유형의 언어적 측정치를 제공하는 Coh-Metrix 시스템을 활용하여 객관적이고 체계적으로 분석하고자 하였다.

2. 코퍼스 기반 연구

최근 들어 영어교육, 심리언어학, 언어학 분야 등에 코퍼스가 광범위하게 활용되고 있다. 이러한 이유는 코퍼스가 분석 결과의 일반화 정도를 증대시키고 컴퓨터 프로그램을 활용해 자료를 분석함으로써 자료 분석의 정확성과 객관성을 높일 수 있기 때문이다(전문기, 2014, 2015). 본 연구의 주제와 관련하여 영어 에세이 및 영어 글쓰기 자료들에 대한 코퍼스를 구축하여 코퍼스 분석 프로그램을 활용하여 분석한 연구들이 있었다.

예를 들면, 신동광, 배주경, 송민영(2014)은 고등학교 학생들이 작성한 국가 영어능력 평가시험 글쓰기 답안지 중 1,119개를 발췌하여 영어 글쓰기 코퍼스를 구축하였다. 이 영어 글쓰기 코퍼스는 약 7만개의 어휘들을 포함하였다. 그들은 용례분석 프로그램을 활용하여 이러한 글쓰기 코퍼스에 나타난 영어 단어 조합 오류들의 빈도를 비교 분석하였다. 구체적으로 단어 선택 오류, 전치사 선택 오류, L1-L2 번역 오류의 빈도를 비교 분석하였다. 그들의 연구 결과, 단어 선택 오류가 가장 빈번하게 발생하였다.

최연정과 유원호(2012)는 대학교 815명의 신입생들을 대상으로 구축된 영어 에세이 코퍼스에 나타난 영어 오류들을 분석하였다. 구체적으로 그들은 V-ing 사용 오류들의 빈도를 분석하였다. 그들이 구축한 코퍼스는 총 200,983개의 어휘들로 구성되어 있었다. 그들은 용례분석 프로그램인 MonoConc Pro를 활용하여 V-ing 사용 오류들의 빈도 분포를 분석하였다. 영어 에세이 코퍼스 분석 결과, 현재분사 형태로 사용된 V-ing 용례들 중 11.1%의 오류가 발견되었고 동명사의 형태로 사용된 V-ing 용례들 중 5.3%의 오류가 발견되었다. 이러한 연구 결과들은 전산언어학과 코퍼스언어학에 기반한 다양한 유형의 용례분석 프로그램들이 영어 에세이 오류 분석 등에 효과적으로 이용될 수 있음을 보여주고 있다.

한편으로 영어 코퍼스 자료 분석에 유용하게 활용될 수 있는 컴퓨터 프로그램을 개발한 연구자들이 있었다. 예를 들면, 김지은과 이공주(2007)는 중학교 학생들이 작성한 영어 에세이들에 나타난 문법 오류들을 자동으로 검색 및 평가하는 컴퓨터 시스템을 개발하였다. 이 시스템은 영어 학습 현장에 유용하게 활용될 수 있다. 이 시스템의 장점은 교사가 학습시키고자 하는 문법적인 문장들을 시스템에 직접 입력할 수 있다는 점이다. 이 시스템은 입력된 문장들의 구조를 분석하여 주요 문법 구문들의 구조를 자동으로 도식화 한 후 학습자가 입력한 문장들과 비교하여 문법성을 평가한다. 다시 말하면 이 문법 분석 시스템은 학습자가 입력한 문장들을 시스템이 도식화한 문법 구조와 비교하여 학습자가 작성한 문장들의 문법성을 자동으로 분석해 학습자에게 적절한 피드백을 제시하기 때문에 영어 에세이에 나타난 오류들을 체계적으로 학습하는데 유용하게 활용될 수 있다.

김동성, 김상철, 채희락(2008)은 문법성과 어휘 응집성을 기반으로 하여 영어

에세이를 자동으로 분석하는데 이용될 수 있는 컴퓨터 시스템을 개발하였다. 구체적으로 그들은 문법성과 어휘 간의 의미적 연결 정도를 자동으로 분석하는 시스템을 개발하여 영어 에세이 평가에 적용하였다. 그들이 개발한 시스템의 자동 분석 결과를 수동 분석 결과와 비교 분석해 본 결과 통계적으로 유의미한 차이가 없었다. 이러한 결과는 컴퓨터 시스템을 영어 에세이 분석 및 평가에 효과적으로 적용할 수 있는 가능성이 점진적으로 증대되고 있음을 시사한다.

본 연구에서 적용된 컴퓨터 분석 도구인 Coh-Metrix 시스템(Graesser, Jeon, Cai, & McNamara, 2008; Graesser, McNamara, Louwerse, & Cai, 2004)은 전술한 김동성, 김상철, 채희락(2008)이 개발한 영어 에세이 분석 프로그램보다 광범위한 유형의 언어적 측정치들을 제공한다. 따라서 영어 글쓰기의 언어적 특성을 보다 체계적으로 분석하는데 유용하게 활용될 수 있을 것으로 기대한다.

III. 연구 방법

1. 연구 대상과 범위

본 연구를 위해 경기도 성남시 분당구에 위치한 C초등학교 초등학교 6학년 186명을 대상으로 영작문 자료들을 수집하여 제2언어 영어 학습자 코퍼스를 구축하였다. 본 연구의 주저자는 연구 당시 C초등학교 영어회화강사로 일하면서 수업시간을 통해 영작문 코퍼스 자료를 수집하였다. 구체적으로 본 연구에 참여한 초등학교 학습자는 1학기 쓰기 수행평가를 대비할 목적으로 수업시간 중 약 20분 동안 “Things I Like To Do”라는 주제로 영작문을 자유롭게 작성하였다. 단, 연구 대상이 초등학생임을 고려할 때 아직 영작문 쓰기 능력이 잘 발달되어 있지 않고 집중력 등의 저하로 영작문 작업에 소홀할 가능성이 대두되어 가능한 한 많은 문장을 사용해서 영작문을 작성하도록 지시하였다.

본 연구의 주된 목적은 초등학생들이 수업 시간을 통해 작성한 영작문 자료를 학습자의 수준에 따라 비교 분석하는 것이기 때문에 1학기에 실시된 영어 수행평가 총점을 기반으로 하여 본 연구에 참여한 초등학생들을 세 집단(상위 집단, 중위 집단, 하위 집단)으로 분류하였다. 영어 학습자의 수준에 따라 초등학교 영어 글쓰기 자료를 분석한 연구들(김영민, 배영직, 2007; 송다진, 김태은, 2016)에서 영어 수행 평가 점수가 집단 분류에 활용되었던 점을 근거로 본 연구는 영어 수행 평가 점수를 집단 분류에 이용하였다. 이 과정에서 결석 등의 이유로 수행평가의 점수가 없을 경우 분석대상에서 제외시켰다.

영어 수행평가 문항은 영어교과서 교사용 지도서에 포함된 단원평가 문제들을 활용하여 쓰기, 읽기, 말하기, 듣기 영역으로 구성되었다. 각 영역의 만점은 10점이었다. 듣기와 읽기 영역은 10개의 단답식 문제로 10점 만점으로 평가되었고 말하기 영역은 사진에 제시된 인물에 대해 1분간 묘사하는 것에 의해 수

업 중 일부분으로 본 연구의 주저자에 의해 10점 만점으로 평가되었다. 쓰기는 일기 형식의 글쓰기를 통해 본 연구의 주저자에 의해 10점 만점으로 평가되었다.

본 연구에 참여한 연구 대상자의 수행평가 점수 범위는 24점에서 40점 사이였다. 구체적으로 상위 집단의 학생수는 66명으로 이들의 수행평가 점수 범위는 36점에서 40점이었고 하위 집단의 학생수는 57명으로 이들의 수행평가 점수 범위는 24점에서 31점이었다. 그리고 중위 집단의 학생수는 53명으로 이들의 수행평가 점수 범위는 32점에서 35점이었다. 본 연구에서 분석 대상을 세 집단으로 분류한 이유는 연구 대상의 영어 수행평가 수준을 세분화하여 각 수준별 영작문의 언어적 특성을 좀 더 세부적으로 밝히기 위해서였다. 집단을 상위 집단과 하위 집단으로만 구분할 경우 중간 수준의 집단에 속한 학생들이 상위 혹은 하위 집단에 포함되어 두 집단의 차이가 희석될 수 때문에 본 연구에서는 연구 대상을 세 집단으로 분류하여 상위 집단과 하위 집단의 차이를 좀 더 명확하게 하였다. 최종적으로 176개의 영작문 표본들이 분석대상에 포함되었다. 구체적으로 상위 집단에는 66개, 중위 집단에는 53개, 하위 집단에는 57개의 영작문 자료들이 포함되었다.

2. 영작문 코퍼스 분석 도구: Coh-Metrix 시스템

웹 기반 영작문 및 텍스트 분석 컴퓨터 프로그램인 Coh-Metrix는 영작문 및 텍스트 코퍼스 자료를 광범위한 언어적 및 심리언어적 측정치들에 기초해 심층적으로 분석하기 위해 개발되었다(Graesser, McNamara, Louwerse, & Cai, 2004; Graesser, Jeon, Cai, & McNamara, 2008). Coh-Metrix 시스템은 어휘 특성들을 분석하는데 적용되는 어휘 데이터베이스와 다양한 텍스트 장르 코퍼스에 기초해 구축된 준거(norm)를 포함한다. 또한 텍스트 내에 포함된 문장들의 문장 구조와 어휘 품사 정보를 분석하는데 필요한 구문분석기 및 품사분석기를 포함한다. Coh-Metrix는 잠재의미분석(Latent Semantic Analysis) 통계기법을 활용해 텍스트 내에 포함된 문장들 간의 의미적 유사성을 분석하는데 이용되기도 한다(전문기, 2014, 2015; 전문기, 임인재, 2009).

그림 1에 제시된 것처럼, 데스크톱 컴퓨터용 Coh-Metrix는 대용량의 코퍼스 자료들을 시스템에 탑재하여 분석할 수 있다.



그림 1 데스크톱 컴퓨터용 Coh-Metrix 인터페이스 캡처 화면

그림 1의 왼쪽 창에는 분석 대상인 코퍼스 자료에 대한 분석 파일 수 및 크기 정보와 분석 진행 상황이 제시된다. 그림 1의 오른쪽 창에는 분석 중인 코퍼스 파일의 내용이 제시된다. 따라서 사용자는 실시간으로 진행되는 코퍼스 자료 분석 진행 상황과 분석 중인 파일에 대한 정보를 시각적으로 확인할 수 있다. 자료 분석 결과는 그림 2에 제시된 것처럼 마이크로소프트 오피스 엑셀 파일 형태로 제공된다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R		
1	[File]	READRW	READNS	READNP	READAS1	READAP1	READAS1	NOUNs	VERBs	ADJs	ADVs	LexDens1	GERUNd1	RF1	AGLSPSV	CRFBA1	CRFBAu	CRFBS		
2	[File]	READRW	READNS	READNP	READAS1	READAP1	READAS1	NOUNs	VERBs	ADJs	ADVs	LexDens1	GERUNd1	RF1	AGLSPSV	CRFBA1	CRFBAu	CRFBS		
3	[File]	READRW	READNS	READNP	READAS1	READAP1	READAS1	NOUNs	No	VERBs	ADJs	ADVs	LexDens1	GERUNd1	RF1	Infintr	AGLSPSV	CRFBA1	CRFBAu	CRFBS
4	D\X	54	7	1	7.714	7	1.222	333.334	111.112	37.037	10.519	0.5	9.269	45.343	0	0.667	0.476	0.1		
5	D\X	52	7	1	7.429	7	1.365	192.308	115.385	76.923	76.923	0.538	28.875	19.25	0	0.667	0.714	0.1		
6	D\X	43	7	1	6.143	7	1.419	209.302	139.535	93.023	93.023	0.465	11.64	34.919	11.64	0.333	0.238	0.1		
7	D\X	47	7	1	6.714	7	1.191	276.595	191.489	43.85	0	0.468	21.296	10.649	0	1	1	0.1		
8	D\X	39	7	1	5.571	7	1.282	179.487	179.487	35.641	51.282	0.564	0	77	0	0.667	0.714	0.1		
9	D\X	76	10	1	7.6	10	1.382	184.211	65.79	52.632	65.789	0.632	6.586	26.342	0	0.889	0.822	0.1		
10	D\X	44	8	1	5.5	8	1.205	204.545	181.818	0	0	0.614	0	91	0	1	1	0.1		
11	D\X	68	9	1	7.566	9	1.265	161.764	89.236	73.529	14.706	0.602	0	36.891	0	0.625	0.722	0.3		
12	D\X	34	1	1	34	1	1.118	147.059	176.471	0	0	0.676	14.721	58.882	0	0	0	0.1		
13	D\X	80	13	1	6.154	13	1.463	425	112.5	50	50	0.363	12.513	0	0	0.333	0.293	0.3		
14	D\X	183	12	1	15.25	12	1.508	256.831	65.573	81.967	76.503	0.519	13.675	8.205	0	0.636	0.708	0.1		
15	D\X	88	11	1	8	11	1.239	216.909	126	56.918	56.918	0.545	11.375	45.5	0	0.7	0.627	0.1		
16	D\X	85	9	1	9.444	9	1.376	329.412	70.589	35.294	35.294	0.529	17.665	5.888	0	0.5	0.444	0.1		
17	D\X	106	11	1	9.636	11	1.377	226.416	84.906	37.736	56.604	0.594	4.722	28.33	0	0.7	0.727	0.1		
18	D\X	80	6	1	13.333	6	1.15	212.5	112.5	12.5	50	0.613	12.513	37.538	0	1	0.8	0.1		
19	D\X	63	5	1	12.6	5	1.206	236.095	126.994	15.873	15.873	0.603	0	55.611	0	0.75	0.7	0.1		
20	D\X	29	6	1	4.833	6	1.069	310.345	103.448	0	34.483	0.552	0	0	0	0.4	0.467	0.1		
21	D\X	38	7	1	5.429	7	1.158	184.211	210.527	0	0	0.606	13.171	92.197	0	1	1	0.1		
22	D\X	36	7	1	5.143	7	1.306	250	83.334	27.778	111.111	0.528	27.806	19.903	0	1	1	0.1		
23	D\X	45	7	1	6.429	7	1.111	177.778	88.889	22.222	44.444	0.667	0	66.733	0	1	1	0.1		
24	D\X	58	8	1	7.25	8	1.448	362.059	68.964	51.724	51.724	0.466	8.629	8.629	0	0.714	0.679	0.1		
25	D\X	82	10	1	8.2	10	1.451	329.268	146.341	60.976	48.78	0.415	12.207	12.207	0	0.667	0.533	0.4		
26	D\X	38	7	1	5.429	7	1.211	184.211	210.527	0	0	0.605	13.171	92.197	0	1	1	0.1		
27	D\X	40	7	1	5.714	7	1.375	250	175	0	0	0.575	25.025	62.563	0	1	1	0.1		
28	D\X	28	5	1	5.6	5	1.393	392.857	71.429	71.429	0	0.464	0	0	0	0.75	0.9	0.1		
29	D\X	46	7	1	6.571	7	1.229	239.13	173.913	0	43.478	0.543	10.58	65.283	0	0.667	0.714	0.1		
30	D\X	34	5	1	6.8	5	1.382	352.941	147.06	58.824	0	0.441	29.441	0	0	0.5	0.4	0.1		
31	D\X	49	5	1	9.8	5	1.265	346.939	40.816	20.408	20.408	0.571	0	20.429	0	0.5	0.6	0.1		
32	D\X	70	9	1	7.778	9	1.386	300.001	71.429	14.286	114.286	0.5	0	14.3	0	0.75	0.778	0.1		

그림 2 Coh-Metrix 측정지 엑셀 파일 캡처 화면

그림 1에 제시된 것처럼, 데스크톱 컴퓨터용 Coh-Metrix는 광범위한 유형의

측정치들을 제공한다. 본 연구에서는 이러한 측정치들 중 영작문 및 텍스트의 이해도와 난이도에 직간접적으로 영향을 미치는 총 21개의 측정치들을 선택하여 본 연구의 영작문 분석에 활용하였다. 본 연구에 적용된 측정치들에 대해 설명하면 다음과 같다.

1) 기초 산출치

기초 산출치에는 텍스트 내에 포함된 단어수, 문장수, 그리고 평균문장길이(mean sentence length)가 포함된다. 평균문장길이는 문장길이에 의한 효과를 통제하고자 텍스트 내에 포함된 문장들의 수를 텍스트 내에 포함된 단어들의 수로 나누어 산출된다(전문기, 2015). 일반적으로 이러한 측정치들은 영작문의 이해도와 난이도를 평가하는 기초적인 점수들로 활용될 수 있다. 점수들이 높을수록 텍스트의 난이도는 증가하는 경향이 있다(전문기, 2015).

2) 단어빈도수

단어빈도수는 텍스트 내에 포함된 개별 단어들이 텍스트 내에서 얼마나 자주 반복되어 사용되는지를 제시하는 측정치다. Coh-Metrix는 CELEX 어휘 데이터베이스(Baayen, Piepenbrock, & Gulikers, 1995)를 이용하여 단어빈도수를 계산해 낸다(Graesser et al., 2004). 일반적으로 단어빈도수가 높은 단어들이 단어빈도수가 낮은 단어들에 비해 우리에게 친숙하게 느껴지기 때문에 빈도가 높은 단어들이 많이 포함된 텍스트의 이해도가 높아지는 경향이 있다(전문기, 2014, 2015; 전문기, 임인재, 2009). 따라서 단어빈도수는 영작문의 이해도 및 난이도를 평가하는데 유용하게 활용될 수 있다.

3) 읽기 가독성 지표 점수

Coh-Metrix는 읽기 가독성 지표 점수로 기능할 수 있는 두 개의 읽기 가독성 지표 점수들을 제공한다(Graesser et al., 2004). Coh-Metrix는 Flesch Reading Ease(FRE) 점수와 Flesch-Kincaid Grade Level(FKGL) 점수를 제공한다. FRE점수는 0에서 100 사이의 수치로 제시되는데 점수가 높을수록 텍스트의 이해 난이도는 감소한다. FKGL은 1에서 12사이의 점수로 제시되는데 점수가 높을수록 이해 난이도는 증가한다. 따라서 이러한 읽기 가독성 지표 점수들은 영어 텍스트의 이해도와 난이도를 측정하는 직접적인 지표 점수들로 이용된다(전문기, 2014, 2015).

4) 통사적 복잡성

Coh-Metrix 시스템은 구문분석기와 품사분석기를 이용하여 영어 텍스트 내에 포함된 문장들의 통사적 복잡성을 계산한다(Graesser et al., 2004). Coh-Metrix는 텍스트 내에 포함된 문장들의 통사적 복잡성 분석에 이용될 수 있는 명사구 밀도(noun phrase density) 점수와 본동사 앞 단어수를 제공한다. 명사구 밀도 점수는 명사구 내에 포함된 중심 명사(head noun)를 수식하는 수식어들의 수를 그 명사구 내에 포함된 전체 단어들의 수로 나눈 값을 지칭하고 본동사 앞 단어수는 본동사 앞에 위치한 단어들의 수를 지칭하기 때문에 이러한 측정치들의 점수들이 높은 텍스트는 낮은 텍스트에 비해 이해 난이도가 증가한다(전문기, 2014, 2015; 전문기, 임인재, 2009). 따라서 Coh-Metrix가 제공하는 이러한 통사적 복잡성 측정치 점수들은 본 연구에서 분석된 영작문에 포함된 문장들의 통사적 복잡성을 분석하는데 활용되었다.

5) 어휘다양성

Coh-Metrix는 영어교육 및 심리언어학 분야 등에서 텍스트 내에 포함된 어휘들의 다양성을 분석하는데 유용하게 활용되는 타입-토큰 비율(type-token ratio, TTR) 점수를 이용하여 어휘다양성 측정치를 산출한다. 일반적으로 TTR 점수가 높은 텍스트가 낮은 텍스트에 비해 더 다양한 단어들을 포함하고 있을 가능성이 높기 때문에 TTR 점수는 영작문 및 텍스트의 이해 난이도를 분석하는데 적용된다(전문기, 2014, 2015).

6) 정합성 측정치

텍스트의 정합성(cohesion)은 텍스트의 이해도와 난이도에 영향을 미치는 중요한 심리언어학적 개념이다(McNamara, Louwerse, McCarthy, & Graesser, 2010). 텍스트의 정합성은 인접된 문장들에 공통으로 사용된 논항들(명사, 대명사, 명사구)을 매개로 하여 형성될 수 있다(Graesser, McNamara, Louwerse, & Cai, 2004). 예를 들면, 아래 첫 번째 문장과 두 번째 문장을 살펴보면 어휘 'banana'가 공통으로 사용되고 있다. 심리언어학적으로 두 개의 문장에 특정 어휘가 공통으로 반복되어 사용되면 텍스트의 정합성이 높아지고 텍스트의 이해도가 증가한다(Kintsch & Van Dijk, 1978). 따라서 논항 반복 비율은 텍스트의 정합성을 측정하는 지표 점수가 될 수 있다.

I like a banana. The banana is really sweet.

Coh-Metrix는 인접된 문장 쌍들에 대한 논항반복 측정치를 계산하여 텍스트

의 정합성 분석에 활용될 수 있는 측정치를 제공한다. 논항반복 측정치의 점수가 높은 텍스트는 낮은 텍스트에 비해서 이해 난이도가 감소한다(전문기, 2014, 2015; 전문기, 임인재, 2009). Coh-Metrix는 또한 텍스트 내에 포함된 서로 인접된 문장들 간의 의미적 유사성을 측정하는데 이용되는 수학적 알고리즘인 LSA(Landauer, Foltz, & Laham, 1998)를 이용하여 텍스트의 정합성을 측정한다. 논항반복 측정치와 같이 LSA 측정치의 점수가 높은 텍스트는 낮은 텍스트에 비해서 이해 난이도가 감소한다(전문기, 2014, 2015; 전문기, 임인재, 2009). 따라서 이러한 정합성 측정치들은 영작문의 이해도 및 난이도를 분석하는데 사용된다.

7) 접속사 정보

텍스트 내에 포함된 다양한 유형의 접속사들은 텍스트 내에 포함된 문장들의 연결을 명확하게 해 주는 역할을 하기 때문에 텍스트의 이해도와 난이도에 중요한 영향을 미치는 언어적 요소들이다(Caron, Micko, & Thuring, 1988; Graesser, Jeon, Yan, & Cai, 2007; Graesser et al., 2004; Halliday & Hasan, 1976; Millis & Just, 1994; Murray, 1997; Segal, Duchan, & Scott, 1991). Coh-Metrix는 텍스트 내에 포함된 인과적 접속사(therefore, because, so, thus 등), 부가적 접속사(and, however, but 등), 시간적 접속사(until, when, before, after, while 등) 측정치 점수들을 제공한다.

8) 대명사 정보

텍스트 내에 포함된 대명사들은 텍스트의 이해 및 난이도에 영향을 미치는 중요한 언어적 요소들이다(전문기, 2014, 2015; 전문기, 임인재, 2009). 텍스트 내에 대명사들이 적절하게 사용되면 텍스트의 이해도가 높아질 가능성이 높기 때문에 대명사는 텍스트의 이해도와 난이도를 분석하는데 이용될 수 있다(Wolf, Gibson, & Desmet, 2004).

9) 어휘 정보

Coh-Metrix는 MRC Psycholinguistics Database(Coltheart, 1981)를 이용하여 텍스트 내에 포함된 단어들의 속성을 분석하는데 유용하게 활용될 수 있는 측정치들을 제공한다. Coh-Metrix는 MRC 데이터베이스를 활용하여 텍스트 내에 포함된 단어들의 구체성(concreteness), 심상성(imageability), 습득나이(age of acquisition)를 분석하는데 적용될 수 있는 측정치들을 제공한다(Graesser et al., 2004). Coh-Metrix가 제공하는 이러한 측정치들의 점수들은 100에서 700사이의 점수로 제시된다. 구체성과 심상성 측정치들의 점수들이 높은 텍스트는 낮은 텍스트에 비해 구체적이고 쉽게 이미지화 할 수 있는 단어들을 상대적으로 많이 포함하고 있을 가능성이 높기 때문에 텍스트의 이해 난이도가 감소한다(전문기, 2014,

2015). 반면에 습득나이 측정치의 점수가 높은 텍스트는 낮은 텍스트에 비해서 상대적으로 높은 연령대에서 학습되는 단어들을 많이 포함하고 있을 가능성이 높기 때문에 텍스트의 이해 난이도는 증가한다(전문기, 2014, 2015).

3. 영작문 코퍼스 구축과 통계분석 방법

본 연구에 사용된 제2언어 영어 학습자 영작문 코퍼스를 구축하기 위해 학생들이 펜을 이용해 작성한 영작문들을 컴퓨터를 활용해 각각 American Standard Code for Information Interchange(ASCII) 형식의 텍스트 파일들로 전산화하였다. ASCII 파일 형식을 사용한 이유는 이 파일 형식이 여전히 코퍼스를 전산화하는데 많이 사용되고 있고(Meyer, 2002) 본 연구에서 이용한 데스크톱 컴퓨터용 Coh-Metrix도 ASCII 파일 형식의 텍스트 파일들을 분석하는데 최적화되어 있기 때문이다. 전산화 과정 중에 학생들이 한글로 쓴 TV 프로그램 이름이나 지명과 같은 고유명사는 영어로 번역하여 전산화하였다.

이러한 전산화 과정을 통해 총 176개의 텍스트 파일들이 생성되었다. 구체적으로 상위 학습 집단으로부터 66개, 중위 학습 집단으로부터 53개, 하위 학습 집단으로부터 57개의 텍스트 파일들이 생성되었다.

영작문 코퍼스 전산화 작업이 완료된 후 전산화된 텍스트 파일들을 Coh-Metrix에 탑재하여 분석하였다. Coh-Metrix가 제공하는 측정치들 중에서 영작문의 이해도 및 난이도를 분석하는데 직간접적으로 영향을 미칠 수 있는 측정치들(전문기, 2014, 2015)을 선택하여 본 연구의 영작문 분석에 이용하였다. 본 연구에 총 21개의 Coh-Metrix 측정치들이 이용되었다. 이러한 측정치들은 기초 산출치(단어수, 문장수, 평균문장길이), 단어빈도수, 읽기 가독성 지표 점수(Flesch Reading Ease 점수, Flesch-Kincaid Grade Level 점수), 통사적 복잡성(본동사 앞 단어수, 명사구 밀도), 어휘다양성(타입-토크 비율), 공통참조 정합성 측정치(논항 반복 비율, LSA 측정치), 접속사 정보 측정치(인과적 접속사, 부가적 접속사, 시간적 접속사), 대명사 정보 측정치(대명사 비율, 1인칭 대명사, 2인칭 대명사, 3인칭 대명사), 어휘 정보 측정치(심상성, 구체성, 습득나이)를 포함하였다.

본 연구에서는 초등학교 6학년 초등학생들이 작성한 영작문을 세 집단(하위 학습 집단, 중위 학습 집단, 상위 학습 집단)으로 구분하여 Coh-Metrix 각 측정치에 대한 세 집단들 간의 차이점을 통계분석을 통해 분석하였다. 본 연구에는 피험자간 디자인(between-subjects design)이 사용되었다. 본 연구에 사용된 독립변수는 집단유형(하위 학습 집단, 중위 학습 집단, 상위 학습 집단)이었고 종속변수는 Coh-Metrix 각 측정치였다.

본 연구에서는 세 집단들 간의 차이점을 분석하기 위해 일원분산분석을 실시하였다. 그리고 세 집단 간에 통계적으로 유의미한 차이가 있을 경우 구체적으로 어느 집단들 간에 차이가 있는지를 추가적으로 검증하기 위해 Tukey 사후분석을 실시하였다. 본 연구는 유의수준 $p = .05$ 에서 통계적 검증을 실시하였다.

IV. 연구 결과 및 논의

본 연구의 통계분석 결과인 일원분산분석 결과는 표 1에 제시되었고 Tukey 사후 분석 결과는 표 2에 제시되었다. 표 1에 제시된 것처럼, 단어수에 대해서 세 집단들 간에 통계적으로 유의미한 차이가 나타났다($F(2, 173) = 17.455, p < .05$). 다시 말하면, 하위 학습 집단에서 상위 학습 집단으로 갈수록 단어수가 증가하는 경향이 있었다. Tukey 사후분석 결과, 하위 집단과 중위 집단, 상위 집단과 중위 집단, 상위 집단과 하위 집단 간 차이들이 모두 통계적으로 유의미하였다(표 2 참조). 또한 문장수에 대해서도 세 집단들 간에 통계적으로 유의미한 차이가 나타났다($F(2, 173) = 11.779, p < .05$). 표 2에 제시된 Tukey 사후 분석 결과를 살펴보면, 하위 학습 집단과 중위 학습 집단, 상위 학습 집단과 하위 학습 집단 간 차이들이 통계적으로 유의미하였다(표 2 참조). 요약하면, 상위 집단의 학생들이 정해진 시간 내에 평균적으로 가장 긴 영작문을 작성했음을 알 수 있다. 이러한 결과는 본 연구에 참여한 C 초등학교 상위 집단에 속한 학생들의 영어 글쓰기 표현 능력(단어수, 문장수 측정치에 제한함)이 상대적으로 높다고 볼 수 있다. 김영민과 배영직(2007)의 연구 결과에 의하면, 상위 집단의 초등학교 영어학습자(5, 6학년 대상)의 어휘량이 하위 집단의 초등학교 영어학습자의 어휘량보다 높았다. 본 연구의 결과 상위 집단의 학생들이 사용한 단어수가 가장 높았다. 이러한 결과는 김영민과 배영직(2007)의 연구 결과를 지지한다고 하겠다.

표 1
각 집단별 Coh-Metrix 측정치 평균

	하위 (n = 57)	중위 (n = 53)	상위 (n = 66)
기초 산출치			
단어수*	47	59	70
문장수*	6	7	8
평균문장길이	8.1	8.3	8.7
단어빈도수	3.14	3.14	3.17
읽기가독성지표			
FRE*	91	87	85
FKGL*	2.6	3.1	3.5
통사적 복잡성			
본동사 앞 단어수	1.7	2	2.2
명사구밀도	.49	.49	.49
어휘다양성			
타입-토크비율	.76	.72	.77
정합성 측정치			
논항반복비율	.71	.75	.76
LSA 측정치	.26	.25	.25
접속사 정보			
인과적 접속사	25.7	21.1	23.1
부가적 접속사*	2.7	5.8	8.5
시간적 접속사	5	5.8	5.4
대명사 정보			
대명사 비율*	.42	.48	.44
1인칭 대명사	71.5	76.3	68.6
2인칭 대명사	1.1	1.3	.9
3인칭 대명사*	0	0.3	1.7
어휘 정보			
심상성	447	435	437
구체성	401	387	393
습득나이	255	254	256

* $p < .05$

표 1에 제시된 것처럼 본 연구 대상인 C초등학교 학생들의 영작문은 평균적으로 6개에서 8개의 문장들로 구성되어 있었다. 본 연구에 참여한 연구대상자가 초등학생인 점과 연구대상자의 지역과 교육 환경을 고려했을 때 평균 이상의 통계치가 산출되었다고 판단된다. 또한 문장 단위에서 분석되는 표준가독성 지표, 통사적 복잡성, 정합성과 같은 Coh-Metrix 측정치(Graesser et al., 2007)는 이론적으로 최소 2개 이상이면 가능하기 때문에 본 연구에서 분석된 영작문에 대해 신뢰롭고 안정적인 Coh-Metrix 측정치가 산출되었다고 판단된다.

표 1에 의하면, 평균문장길이 측정치에 대해서는 세 집단들 간에 통계적으로 유의미한 차이가 발견되지 않았다($F(2, 173) = .559, p = .573$). 이러한 결과는 상위 학습 집단의 학생들이 상대적으로 높은 영어 글쓰기 표현 능력(평균문장 길이 측정치에 제한함)을 가지고 있지만 영작문을 작성할 때는 다른 집단의 학생들과 평균적으로 비슷한 길이의 문장들을 사용하고 있음을 제시한다. 다시 말하

면 다른 집단의 학생들과 유사한 통사 구조를 지닌 문장들을 사용하여 영작문을 작성하고 있음을 제시한다.

표 1에 제시된 것처럼, 흥미롭게도 단어빈도수에 대해서는 세 집단들 간에 통계적으로 유의미한 차이가 나타나지 않았다($F(2, 173) = .501, p = .607$). 상위 학습 집단의 학생들이 상대적으로 더 많은 저빈도 단어들을 사용하여 영작문을 작성할 것으로 기대했으나 세 학습 집단들의 학생들은 빈도가 유사한 단어들을 사용하여 영작문을 작성하였다. 이러한 결과는 본 연구에 참여한 C초등학교 학생들이 수업 중에 동일한 영어 교과서를 이용하여 영어를 학습했기 때문에 어휘 수준에서 큰 차이를 보이지 않았을 것이라고 해석된다.

표 2
Tukey를 이용한 집단 별 사후분석 결과표

	하위 vs. 중위	상위 vs. 중위	상위 vs. 하위
기초 산출치			
단어수	.007*	.026*	.000*
문장수	.029*	.096	.000*
평균문장길이	.993	.793	.554
단어빈도수	.998	.698	.645
표준가독성지표			
FRE	.133	.598	.008*
FKGL	.245	.506	.014*
통사적 복잡성			
본동사 앞 단어수	.315	.864	.101
명사구밀도	.991	.999	.994
어휘다양성			
타입-토크비율	.190	.053	.852
정합성 측정치			
논항반복비율	.795	.959	.601
LSA 측정치	.723	.996	.750
접속사 정보			
인과적 접속사	.698	.927	.883
부가적 접속사	.327	.379	.013*
시간적 접속사	.930	.986	.974
대명사 정보			
대명사 비율	.015*	.175	.487
1인칭 대명사	.449	.114	.723
2인칭 대명사	.927	.764	.944
3인칭 대명사	.823	.044*	.007*
어휘 정보			
심상성	.231	.982	.271
구체성	.151	.721	.464
습득나이	.992	.982	.998

* $p < .05$

읽기 가독성 지표 분석 결과를 살펴보면, 먼저 FRE 점수에 대해 세 집단들 간에 통계적으로 유의미한 차이가 발견되었다($F(2, 173) = 4.587, p < .05$). Tukey

사후분석 결과 상위 학습 집단과 하위 학습 집단 간에 통계적으로 유의미한 차이가 발견되었다(표 2 참조). FKGL 점수에 대해서도 세 집단들 간에 통계적으로 유의미한 차이가 있었다($F(2, 173) = 4.036, p < .05$). FRE 분석 결과와 동일하게 Tukey 사후분석 결과 상위 학습 집단과 하위 학습 집단 간에 통계적으로 유의미한 차이가 발견되었다(표 2 참조). 이러한 결과는 상위 학습 집단의 학생들이 작성한 영작문의 이해 난이도가 하위 학습 집단의 학생들보다 높다고 볼 수 있다. 그러나 이러한 이해 난이도의 차이가 어휘 수준이 아닌 문장의 구조에서 기인된다고 생각해 볼 수 있다. 왜냐하면 FRE 점수와 FKGL 점수는 단어길이와 문장길이에 의해서 영향을 받기 때문이다. 흥미롭게도 표 2에 제시된 것처럼 평균문장길이 차이는 두 집단 간에 통계적으로 유의미한 차이가 없었기 때문에 상위 학습 집단의 학생들이 좀 더 긴 단어를 사용하여 영작문을 작성했다고 볼 수 있다. 일반적으로 학생들이 짧은 영어 단어들보다 긴 영어 단어들을 학습하는데 더 어려움을 느낄 수 있기 때문(Klare, 1974-1975)에 상위 학습 집단 학생들의 어휘학습 능력이 상대적으로 높다고 볼 수 있겠다.

표 1에 제시된 것처럼, 통사적 복잡성 분석 결과를 살펴보면 본동사 앞 단어수와 명사구 밀도 점수에 대해 통계적으로 유의미한 차이가 발견되지 않았다(각각 $F(2, 173) = 2.236, p = .110, F(2, 173) = .009, p = .991$). 이러한 결과는 평균문장길이 분석 결과와 유사하게 본 연구에 참여한 C초등학교 학생들이 대체적으로 비슷한 정도의 통사 구조를 지닌 문장들을 사용하여 영작문을 작성함을 시사한다. 표 1에 제시된 것처럼, 세 학습 집단들의 학생들은 본동사 앞에 평균 1.7에서 2.2개의 단어들과 명사구 내에 평균 1개 정도의 수식어를 사용하여 영작문을 작성하는 경향이 있었다.

어휘다양성 지표인 타입-토크 비율 분석에 대해서는 세 학습 집단들 간에 통계적으로 유의미한 차이가 발견되지 않았다($F(2, 173) = 2.907, p = .057$). 이러한 결과는 단어빈도수 측정치 분석 결과에서 논의한 것처럼 본 연구에 참여한 C초등학교 학생들이 주로 학교에서 동일한 영어 교과서를 통해 영어를 학습하고 영어 학습 시간도 길지 않기 때문에 어휘 학습 수준에서 큰 차이를 보이지 않아 수업 시간 등을 통해 배운 한정된 단어들을 사용하여 영작문을 작성했음을 시사한다. 김영민과 배영직(2007)은 상위 집단의 초등학교 학생들(5, 6학년 대상)의 일기 쓰기 자료가 하위 집단의 쓰기 자료보다 더 많은 어휘들을 포함하고 있음을 제시했다. 그들의 연구 결과와 어휘다양성 측정 지표인 타입-토크 비율을 분석한 본 연구의 결과를 비교해 보면 본 연구의 상위 집단 학생들의 어휘량은 하위 집단 학생들의 어휘량보다 높았지만 사용한 어휘들의 다양성 정도는 큰 차이가 나지 않았음을 시사한다.

표 1에 제시된 것처럼 정합성 측정치인 인접된 문장들 간의 논항반복비율과 LSA 점수에 대해서도 세 학습 집단들 간에 통계적으로 유의미한 차이가 나타나지 않았다(각각 $F(2, 173) = .481, p = .619, F(2, 173) = .370, p = .691$). 상위 학습 집단 영작문의 정합성이 높을 것이라 기대와 달리 본 연구에 참여한 C초등학

교 학생들은 정합성 수준이 비슷한 영작문을 작성하였다.

접속사 정보 측정치 분석 결과를 살펴보면 인과적 접속사와 시간적 접속사에 대해서는 학습 집단들 간에 통계적으로 유의미한 차이가 발견되지 않았다(각각 $F(2, 173) = .330, p = .719, F(2, 173) = .066, p = .936$). 그러나 흥미롭게도 부가적 접속사에 대해서는 학습 집단들 간에 통계적으로 유의미한 차이가 있었다($F(2, 173) = 4.118, p < .05$). 부가적 접속사에 대한 Tukey 사후분석 결과 상위 학습 집단과 하위 학습 집단 간에 통계적으로 유의미한 차이가 있었다(표 2 참조). 이러한 결과는 본 연구에 참여한 C초등학교 상위 학습 집단의 학생들이 하위 학습 집단의 학생들보다 영작문을 작성할 때 and, however, but 등과 같은 부가적 접속사를 사용하여 문장들을 연결하려는 경향이 높다고 볼 수 있다. 김영민과 배영직(2007)의 연구 결과에 의하면, 그들의 연구에 참여한 초등학교 학생들(5, 6학년 대상)은 영어 일기를 작성할 때 and와 but과 같은 부가적 접속사를 빈번하게 사용하는 경향이 있었다. 그들은 부가적 접속사의 사용 빈도를 분석할 때 학생들의 수준을 고려하지 않았다. 본 연구의 결과가 그들의 부가적 접속사 빈도 분석 결과를 좀 더 세부적으로 보완할 수 있으리라고 기대한다. 표 2에 제시된 것처럼 본 연구에 참여하였던 C초등학교 학생들은 영작문을 작성할 때 인과적 접속사(because, therefore 등)를 많이 사용하였다.

표 2에 제시된 대명사 정보 분석 결과를 살펴보면 먼저 전체 대명사 비율에 대해서는 세 학습 집단들 간에 통계적으로 유의미한 차이가 있었다($F(2, 173) = 4.027, p < .05$). Tukey 사후분석 결과 흥미롭게도 중위 집단과 하위 집단 간에 통계적으로 유의미한 차이가 발견되었다(표 2 참조). 그러나 상위 집단과 중위 집단 간, 상위 집단과 하위 집단 간에는 통계적으로 유의미한 차이가 없었다(표 2 참조). 이러한 결과는 본 연구에 참여한 C초등학교 하위 집단의 학생들이 대명사를 활용하는 능력이 상대적으로 낮을 가능성을 제시한다. 계속해서 1인칭 대명사와 2인칭 대명사에 대해서는 세 집단들 간에 통계적으로 유의미한 차이가 발견되지 않았다(각각 $F(2, 173) = 2.026, p = .135, F(2, 173) = .244, p = .783$). 그러나 3인칭 대명사에 대해서는 세 집단들 간에 통계적으로 유의미한 차이가 있었다($F(2, 173) = 5.422, p < .05$). Tukey 사후분석 결과 상위 학습 집단과 중위 집단 간, 상위 집단과 하위 집단 간에 통계적으로 유의미한 차이가 있었다(표 2 참조). 이러한 결과들은 본 연구에 참여한 C초등학교 하위 집단의 학생들이 전체적으로 대명사를 활용하는 능력이 낮을 가능성과 특히 3인칭 대명사를 활용하는 능력이 상대적으로 낮을 가능성을 암시한다. 김영민과 배영직(2007)의 어휘 빈도 분석 결과에 의하면, 그들의 연구에 참여한 대부분의 초등학교 학생들(5, 6학년 대상)은 일기를 작성할 때 1인칭 대명사와 2인칭 대명사를 빈번하게 사용하는 경향이 있었다. 3인칭 대명사는 상대적으로 빈번하게 사용되지 않았다. 이러한 연구 결과는 본 연구의 연구 결과와 유사하며 특히 3인칭 대명사는 상위 집단의 학생들이 주로 사용함을 시사한다고 하겠다.

마지막으로 어휘 분석 결과를 살펴보면 심상성 측정치, 구체성 측정치, 습득

나이 측정치에 대해 세 학습 집단들 간에 통계적으로 유의미한 차이가 발견되지 않았다(각각 $F(2, 173) = 1.691, p = .187, F(2, 173) = 1.784, p = .171, F(2, 173) = .017, p = .983$). 이러한 결과는 단어빈도수, 어휘다양성 측정치 분석결과에서 논의한 것처럼 본 연구에 참여한 C초등학교 학생들이 학교 현장에서 주로 영어 교과서를 통해 영어를 학습하기 때문에 심상성, 구체성, 습득나이 정도가 비슷한 단어들을 사용하여 영작문을 작성했을 가능성을 제시한다.

V. 결론

본 연구는 영작문 및 텍스트 분석에 유용하게 활용될 수 있는 텍스트 분석 시스템인 Coh-Metrix가 제공하는 다양한 유형의 언어학적 측정치들에 기초하여 C초등학교 6학년 학생들이 작성한 영작문 코퍼스를 수준별로 분류하여 분석하였다.

본 연구의 결과를 요약하면 다음과 같다. 첫째, 본 연구의 결과 상위 학습 집단의 C초등학교 학생들은 상대적으로 많은 단어들과 문장들을 사용하여 가장 긴 영작문을 작성하였다. 이러한 결과는 단어수와 문장수에 기초했을 때 본 연구에 참여했던 상위 학습 집단 학생들의 영어 작문 표현 능력이 다른 집단들(중위 학습 집단, 하위 학습 집단)의 학생들보다 높을 가능성을 제시한다. 그러나 평균문장길이에 대해서는 세 집단들 간에 통계적으로 유의미한 차이가 발견되지 않았다.

둘째, 단어빈도수, 어휘다양성(타입-토큰 비율) 측정치, 어휘 정보 측정치(심상성, 구체성, 습득나이)에 대해서는 세 학습 집단들 간에 통계적으로 유의미한 차이를 보이지 않았다. 다시 말해서 세 학습 집단들의 학생들은 서로 비슷한 정도의 어휘력을 가지고 있다고 할 수 있다.

셋째, 읽기 가독성 지표 점수(FRE 점수, FKGL 점수)에 대해 세 학습 집단들 간에 통계적으로 유의미한 차이가 있었다. 다시 말하면 상위 학습 집단의 학생들이 작성한 영작문의 이해 난이도가 가장 높았다. 각 집단 간 이러한 읽기 난이도의 차이는 문장 길이보다는 단어 길이에 의한 것이라고 해석할 수 있다. 왜냐하면 FRE 점수와 FKGL 점수는 단어 길이와 문장 길이에 의해 크게 영향을 받는데 세 학습 집단들은 평균문장길이에 대해 통계적으로 유의미한 차이를 보이지 않았기 때문이다. 또한 통사적 복잡성 측정치의 분석 결과도 이러한 사실을 뒷받침해 준다. 다시 말하면 문장의 통사적 복잡 정도를 나타내는 본동사 앞 단어수와 명사구 밀도 점수들에 대해 세 집단들은 통계적으로 유의미한 차이를 보이지 않았다. 요약하면, 세 집단들의 학생들은 문장 구성 능력 있어서는 비슷하지만 어휘 습득력에 있어서는 차이를 보인다고 할 수 있다. 다시 말하면 이러한 결과는 상위 학습 집단의 학생들이 상대적으로 길이가 긴 어휘들을 더 많이 학습하고 있음을 제시한다.

마지막으로 C초등학교 상위 학습 집단의 학생들은 상대적으로 더 많은 부가적 접속사와 3인칭 대명사를 사용하여 영작문을 작성하는 경향이 있었다. 이러한 결과는 하위 집단 학생들이 상대적으로 부가적 접속사와 3인칭 대명사를 사용하는 능력이 낮을 가능성을 제시한다.

요약하면, 학교 현장에서 교사들이 동일한 영어교과서를 이용하여 학습함에도 불구하고 단어수, 문장수, FRE 점수, FKGL 점수, 부가적 접속사, 대명사 비율, 3인칭 대명사 측정치에 대해 세 학습 집단 간에 통계적으로 유의미한 차이를 보였다. 이러한 결과는 본 연구에서 고려되지는 않았지만 사교육 효과가 결과에 영향을 주었을 수도 있을 것이라 생각해 볼 수 있다. 다시 말하면 상위 학습 집단의 학생들이 사교육에 상대적으로 더 많이 노출되었을 가능성을 배제할 수 없다. 따라서 학교 현장에서 사교육에 의한 효과를 막기 위해서 교사들이 하위 학생들을 대상으로 수업 시간 외에 별도로 방과후 수업 등을 통해 개별적으로 영작문 작성 지도를 수행할 수 있겠다. 물론, 차후 영작문 작성 능력에 미치는 사교육의 효과를 밝히는 연구가 선행되어야 할 것이다.

본 연구가 제시한 연구결과 외에 본 연구에 내포된 제한점을 제시하면 다음과 같다. 먼저 본 연구는 C초등학교 6학년만을 대상으로 영작문 코퍼스를 구축하여 Coh-Metrix 시스템을 활용하여 분석하였다. 따라서 본 연구의 결과는 일차적으로 C초등학교 6학년 초등학생들에게만 일반화될 수 있겠다. 초등학교 3학년부터 영어 교육이 실시됨을 고려할 때 차후 연구에서는 연구대상을 확대하여 여러 초등학교 3학년, 4학년, 5학년에 대한 영작문 코퍼스도 구축하여 영어 작문 능력의 발달 과정을 체계적으로 연구해 볼 수 있을 것으로 기대한다. 또한 본 연구는 특정 지역의 초등학생들만 대상으로 영작문 코퍼스 자료를 구축하여 그 자료를 분석하였기 때문에 지역 차이에 의한 효과가 차후 연구에서 밝혀지기를 기대한다. 그리고 본 연구는 Coh-Metrix 시스템 측정치 결과와 양적 통계분석 결과에 기초하여 초등학교 영작문 자료를 분석하였다. 따라서 차후 연구에서는 영작문을 질적으로 평가하여 양적 연구 방법에 기초한 본 연구의 결과를 보완할 수 있을 것으로 생각된다. 예를 들면, 접속사나 대명사 등과 같은 특정 언어적 요소가 어떠한 문맥에서 사용되는지 등에 대한 분석을 질적 연구방법을 통해 보완할 수 있으리라고 기대한다. 마지막으로 본 연구에서는 영어 수행평가 점수를 활용하여 집단을 분류하였다. 영어 수행평가 점수를 활용하여 집단을 분류한 연구들(김영민, 배영직, 2007; 송다진, 김태은, 2016)이 있지만 영어 수행평가 점수에 의해 학생들의 절대적인 영어 능력을 평가하는 데는 제한점이 따른다. 차후 연구에서는 초등학생의 영어 능력을 절대적으로 평가할 수 있는 일종의 영어공인시험 등이 연구 및 개발되어 학생들의 수준을 평가하는데 활용되기를 기대한다. 또한 EFL 영어 학습자가 작성한 영어 글쓰기 자료를 몇 명의 교사들이 직접 평가하여 평가자 간에 신뢰도를 확립한 후 그 평가 점수를 집단의 수준을 분류하는데 활용하는 것도 대안이 될 수 있겠다.

VI. 제언

본 연구의 연구 결과들에 기초하여 영어 교육 현장에 본 연구의 결과들이 활용될 수 있는 방안들을 제시하면 다음과 같다. 첫째, 본 연구의 결과 C초등학교 상위 학습 집단들의 학생들이 상대적으로 많은 단어들과 문장들을 활용하여 영작문을 작성하는 경향이 시사되었다. 따라서 학교 교육 현장에서 교사들이 하위 학생들의 영어 표현 능력을 향상시키기 위해 Coh-Metrix와 같은 텍스트 분석 시스템을 적극적으로 활용하여 학생들의 영작문 표현 능력을 객관적으로 분석하여 그 결과를 학습 전략에 제시할 수 있을 것으로 기대한다. 부연하면 학교 현장에서 방과후 수업 등을 통해 하위 학생들을 독려하여 가능한 한 더 많은 단어들과 문장들을 사용하여 영작문을 작성하게 교육함으로써 그들의 영어 표현 능력을 향상시킬 수 있을 것으로 기대한다. 본 연구에서는 데스크톱 컴퓨터용 Coh-Metrix가 활용되었지만 학교 현장에서 교사는 웹용 Coh-Metrix를 활용할 수 있을 것으로 기대한다. Coh-Metrix 홈페이지(<http://cohmetrix.com/>)에 접속한 후 사용자 등록을 하면 누구든지 Coh-Metrix를 무료로 활용할 수 있다. 이용 방법도 매우 간단하기 때문에 쉽게 Coh-Metrix 사용방법을 학습할 수 있을 것으로 기대한다. 교사는 웹용 Coh-Metrix를 활용하여 학생이 작성한 영작문을 분석함으로써 해당 학생에게 신속하고 구체적인 피드백을 제공하여 결과적으로 영어 쓰기 능력을 향상시킬 수 있다.

둘째, 본 연구의 결과 C초등학교 세 학습 집단들의 학생들은 서로 비슷한 정도의 어휘력을 가지고 있을 가능성이 높다고 시사되었다(단어빈도수, 어휘다양성, 어휘 정보 측정치 참조). 이러한 결과는 초등학교 현장에서 실시되는 영어 교육이 주로 어휘력 향상에 초점이 맞추어져 있음을 암시한다고 하겠다. 본 연구에서는 C초등학교 6학년 학생들이 작성한 영작문을 분석하였다. 초등학교 6학년 학생들은 졸업 후 중학교 영어 교과 과정을 대비해야 하기 때문에 초등학교 현장에서 실시되는 영어 교육이 어휘력 학습뿐만 아니라 문장 구조의 학습도 적극적으로 반영하기를 기대한다. 예를 들면 영작문 작성 수업 시간에 Coh-Metrix와 같은 시스템을 활용하여 학생들이 작성한 영작문에 나타난 문장들의 특성을 신속하게 분석하여 적절한 피드백을 제공함으로써 학생들의 영작문 작성 능력을 향상시킬 수 있을 것으로 기대한다. 학교 현장에서 교사가 많은 학생들이 작성한 영작문에 반영된 문장들의 특성을 컴퓨터 시스템 도움 없이 일일이 분석하는 것은 상당한 시간을 필요로 하는 일이다. 또한 사람은 평가 시 인지적 편견에서 자유로울 수 없기 때문에 Coh-Metrix와 같은 텍스트 분석 시스템이 효과적으로 영어 학습 보조 도구로 활용될 수 있을 것으로 기대한다(전문기, 2015).

셋째, 본 연구의 읽기 가독성 지표 점수(FRE 점수, FKGL 점수) 결과 상위 집단 학생들이 상대적으로 길이가 긴 어휘들을 더 많이 학습하고 있을 가능성이 제시되었다. 따라서 학교 현장에서 하위 학생들도 이러한 어휘들을 자연스럽게

학습할 수 있도록 효율적인 학습 전략이 수립되기를 기대한다. 구체적으로 수업 시간이나 방과후 학습 시 학생들에게 학습해야 될 단어들을 반복적으로 제시함으로써 어휘 학습 능력을 향상시킬 수 있다. 또한 학습 대상이 초등학생인 점을 고려할 때 흥미와 집중을 유발시킬 수 있도록 단어 만들기 및 단어 맞추기 게임 등을 통해 효과적으로 어휘력을 향상시킬 수 있을 것으로 기대한다.

넷째, 본 연구 결과 C초등학교 하위 학생들은 부가적 접속사와 3인칭 대명사를 사용하는 능력이 낮을 수 있다는 가능성이 제시되었다. 따라서 영어 수업 시간이나 영어 활동 시간 등을 통해 하위 학생들도 이러한 접속사와 대명사를 적절하게 사용하는 방법을 배울 수 있는 학습 방법이 제시될 필요가 있다. 예를 들면, 방과후 수업 시간에 해당 학생들에게 다양한 유형의 접속사와 문장들을 제시한 후 접속사를 이용하여 문장들을 연결시키게 하거나 접속사가 올바르게 사용되지 않은 문장 예들을 제시하여 평가하게 함으로써 접속사 개념을 학습시킬 수 있다. 같은 방법으로 대명사가 올바르게 사용되지 않은 문장 예들을 제시하여 평가하게 하거나 특정 명사를 대신하는 대명사를 올바른 대명사를 선택하게 함으로써 대명사 사용방법을 학습할 수 있을 것으로 기대한다.

마지막으로, Coh-Metrix 프로그램 활용 측면에서 본 연구의 의의점을 제시할 수 있다. 학교 현장에서 교사는 다수의 학생들을 상대로 영어 글쓰기 자료를 평가해야 하기 때문에 학생들에게 구체적인 피드백을 제공하는데 제한점이 발생한다. Coh-Metrix와 같은 프로그램을 활용하면 다양한 언어적 측정치들에 기반해 많은 양의 영어 글쓰기 자료를 신속하고 정확하게 분석할 수 있기 때문에 영어 학습자들에게 구체적이고 신뢰로운 피드백을 제공할 수 있을 것으로 기대한다. 또한 영어 글쓰기 평가 시 교사의 주관적인 판단의 영향을 완전히 배제할 수 없기 때문에 Coh-Metrix와 같은 컴퓨터 프로그램이 영어 글쓰기 평가에 활용된다면 평가의 객관성을 높일 수 있을 것으로 기대한다.

참고문헌

- 교육과학기술부. (2011). *영어과 교육과정*. 서울: 교육과학기술부.
- 김동성, 김상철, 채희락. (2008). 문법성과 어휘 응집성 기반의 영어 작문 평가 시스템. *인지과학*, 19(3), 223-255.
- 김영민, 배영직. (2007). 영어 일기쓰기에 나타난 초등학생의 어휘 활용 능력 분석. *현대영어영문학*, 51(4), 95-119.
- 김유주. (2009). 중학교 영어 교과서 쓰기 영역의 연계성 분석. *현대영어영문학*, 10, 217-224.
- 김지나, 진송화. (2015). 초등 영재들을 위한 영어 글쓰기 지도와 효과적인 첨삭 방법에 대한 사례연구. *언어학*, 23(3), 153-171.
- 김지은, 이공주. (2007). 중학생 영작문 실력 향상을 위한 자동 문법 채점 시스템

- 템 구축. *한국콘텐츠학회*, 7(5), 36-46.
- 김혜리, 경지숙. (2008). 대화식 저널쓰기를 기반으로 하는 초등영어 문자교육 증진 방안에 대한 연구. *영어교육연구*, 20(3), 169-198.
- 김혜리, 김소중. (2010). 문항텍스트 기반 반응저널 쓰기를 통한 초등영어쓰기에 대한 연구. *Studies in English Education*, 15(1), 196-220.
- 나임정, 김태은. (2014). 5·6학년 초등 영어 교과서의 쓰기 활동 및 학습자 인식 분석. *영어어문교육*, 20(3), 231-256.
- 노경진, 김신혜. (2012). 영어동화에 기반한 리딩 저널쓰기 활동이 초등학생의 영어 쓰기능력에 미치는 영향. *초등영어교육*, 18(3), 97-122.
- 송다진, 김태은. (2016). 초등 학습자의 영어 쓰기에 나타난 문법 오류 분석. *현대영어교육*, 17(1), 135-157.
- 송주영, 이동한. (2006). 초등학생의 영어 일기 쓰기에서 나타난 오류 분석. *영상 영어교육*, 7(2), 187-207.
- 신동광, 배주경, 송민경. (2014). 영어 쓰기 답안에 나타난 한국 고등학생들의 단어 조합 오류 유형 분석. *Studies in English Education*, 19(2), 261-282.
- 우연순. (2013). 초등영어 학습자의 수준에 따른 쓰기 피드백의 효과 연구. *초등 영어교육*, 19(2), 131-154.
- 유지연, 나경희. (2013). 2009 개정 교육과정에 의한 고등학교 영어교과서 쓰기 활동 분석. *언어학연구*, 28, 199-221.
- 이동한. (2004). 초등학습자의 영어 쓰기에서 나타난 오류 분석에 관한 사례연구. *영어교육연구*, 16(1), 157-182.
- 이인호, 배주경, 김성혜. (2015). *2014 국가수준 학업성취도 평가 결과 분석(영어)*. 서울: 한국교육과정평가원.
- 전문기. (2014). 고등학교 개정 영어교과서 읽기 자료의 연계성 분석. *Journal of the Korean Data Analysis Society*, 16(2), 925-938.
- 전문기. (2015). 자동화된 언어분석 프로그램을 활용한 중학교 영어교과서의 학년 간 연계성 분석. *현대영어교육*, 16(1), 195-218.
- 전문기, 임인재. (2009). 코메트릭스를 이용한 중학교 1학년 개정 영어 교과서의 코퍼스 언어학적 비교 분석. *영어교육연구*, 21(4), 265-292.
- 전영주. (2010). 예비 영어교사를 위한 영어 쓰기능력 향상 방안. *영어교과교육*, 9(3), 251-273.
- 정동빈, 강시경. (2008). 블렌디드 러닝을 활용한 자기 주도적 초등 영어쓰기교수-학습효과. *언어연구*, 23(4), 577-599.
- 최연정, 유원호. (2012). 한국 대학 신입생들의 영작문 코퍼스에 나타난 V-ing 오류분석. *영어교육*, 67(4), 311-335.
- Allen, J. (1995). *Natural language understanding*. Redwood City, CA: Benjamin Cummings.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-

- ROM). Philadelphia: University of Pennsylvania.
- Caron, J., Micko, H., & Thuring, M. (1988). Conjunctions and the recall of composite sentence. *Journal of Memory and Language*, 27(3), 309-323.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33(4), 497-505.
- Graesser, A. C., Jeon, M., Cai, Z., & McNamara, D. S. (2008). Automatic analyses of language, discourse, and situation models. In J. Auracher & W. van Peer (Eds.), *New beginnings in literary studies* (pp. 72-88). Cambridge: Cambridge Scholars Publishing.
- 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-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36(2), 193-202.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- 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.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Klare, G. R. (1974-1975). Assessing readability. *Reading Research Quarterly*, 10(1), 62-102.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25(2-3), 259-284.
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4), 292-330.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Millis, K., & Just, M. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1), 128-147.
- Murray, J. (1997). Connective and narrative text: The role of continuity. *Memory & Cognition*, 25(2), 227-236.
- Segal, E., Duchan, J., & Scott, P. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes*, 14(1), 27-54.

Wolf, F., Gibson, E., & Desmet, T. (2004). Coherence and pronoun resolution. *Language and Cognitive Processes, 19*(6), 665-675.

예시언어(Examples in): English
적용가능 언어(Applicable Languages): English
적용가능 수준(Applicable Levels): Elementary

김소정
건국대학교 영어영문학과
143-701 서울시 광진구 능동로 120
Tel: (02) 450-3335/ H.P.: 010-7238-9975
Email: sjk7100@gmail.com

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

Received 27 June 2016
Revised 9 August 2016
Accepted 19 August 2016