



## 한국인 대학생 에세이에 나타난 텍스트 응집력과 영어 능력 간의 관계 탐색

이영주 (한밭대학교)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: November 10, 2020  
Revised: April 28, 2021  
Accepted: May 25, 2021

Young-Ju Lee  
Professor, Hanbat National  
University  
[yjulee@hanbat.ac.kr](mailto:yjulee@hanbat.ac.kr)

### ABSTRACT

Lee, Young-Ju. (2021). The relationship between text cohesion features and English proficiency for Korean college students. *Korean Journal of English Language and Linguistics* 21, 435-449.

This study investigates the relationship between text cohesion features and English proficiency for Korean college students. That is, it examines how cohesion features can be used to distinguish among essays written by different English proficiency levels. The ICNALE (International Corpus Network of Asian Learners of English) corpus was employed in this study and 600 essays on two prompts (i.e., smoking and part-time jobs) written by Korean students were analyzed. The Tool for the Automatic Analysis of Cohesion (TAACO), the recently-developed program for automatic analysis of cohesion, was employed. Two statistical analyses were performed; a multivariate analysis of variance (MANOVA) and a stepwise discriminant function analysis. Results of the study showed that cohesion feature indices were significantly affected by English proficiency, implying that essays written by Korean students with different English proficiency levels can be differentiated in terms of various cohesion features. Results of a stepwise discriminant function analysis revealed that the best predictor for distinguishing three groups of English proficiency is pronoun density. High-level Korean students produced more cohesive essays than mid- or low-level students in that they used pronouns, overlapping arguments, and lemmas as a way for linking ideas across sentences. However, high-level students underused connectives, compared with low-level students. Implications of this study for English writing pedagogy are also discussed.

### KEYWORDS

text cohesion, the Tool for the Automatic Analysis of Cohesion (TAACO), L2 writing, writing pedagogy

## 1. 서론

### 1.1 연구의 필요성 및 목적

텍스트 응집력(text cohesion)은 쓰기 능력의 핵심 요소로 정보의 의미를 전달하기 위해서 사용되는 명시적인 언어적 장치이며, 텍스트의 이해도에 영향을 미치는 중요한 요인이다(Halliday and Hassan 1976, Witte and Faigley 1981). 전체 에세이의 의미는 개별 문장들이 모여진 총합의 의미에 있는 것이 아니라 개별 문장들이 어떻게 서로 연결되어 있는가에 달려 있다(Crossley and McNamara 2009). 즉, 똑같은 의미를 갖고 있더라도 문장이 어떻게 유기적으로 연결되어 있는가에 따라서 텍스트의 이해도는 달라질 수 있다. 최근의 쓰기 지도는 과정 중심의 작문 지도 접근법(a process-approach to writing instruction)을 따르고 있으며 문법, 글의 조직과 전개, 적절한 어휘의 사용을 강조하여 지도하고 있다. 하지만 대학의 영작문 수업에서 텍스트 응집력의 비중은 크지 않으며 실제로 한국 학생들의 에세이에서 텍스트 응집력이 영어 수준별로 어떻게 나타나 있는가에 대한 연구는 수행되지 못했다.

본 연구에서는 한국 대학생들이 작성한 에세이에 나타난 텍스트 응집력의 특성이 영어 수준별로 어떻게 다르게 나타나는지를 다양한 통계분석을 통해 살펴보고자 한다. 지금까지의 선행연구(Crossley, Kyle and McNamara 2015, Guo, Crossley and McNamara 2013)는 에세이의 점수를 가장 잘 예측한 텍스트의 응집력 지표가 무엇인지를 단계적 회귀분석(stepwise regression analysis)을 통해 살펴보았으나, 본 연구에서는 판별분석(discriminant function analysis)을 통해 응집력 지표와 영어 수준 간의 관계를 자세히 살펴보고자 한다.

텍스트 응집력은 국부적 응집력(local cohesion), 전체 텍스트 응집력(overall text cohesion), 총체적 응집력(global cohesion)의 세 개로 분류되는데 각각의 특성과 차이점을 살펴보면 다음과 같다(Crossley, Kyle and McNamara 2016). 첫째, 국부적 응집력은 문장 수준에서의 응집력을 의미하며, 예를 들면 however, therefore와 같은 연결사(connectives)를 활용하여 문장을 연결하는 것을 말한다. 연결사는 응집력의 명백한 단서를 제공하며, 독자로 하여금 문장 사이에 개념의 밀접한 관련이 있음을 알려준다. 둘째, 총체적 응집력은 단락 수준에서의 응집력을 의미하며, 예를 들면 단락 간에 특정 명사의 유의어(synonym)가 중복되는 것을 말한다. 단락 수준의 응집력은 표면적으로 드러나지 않기 때문에 총체적인 특성을 가지며, 단락 사이에 제시된 인과 관계나 의미적 유사성을 포함한다. 예를 들면, 한 단락에 제시된 단어나 구가 뒤에 나오는 단락에 반복되어 나오면, 의미적 유사성이 커지기 때문에 총체적 응집력이 높아지게 된다. 셋째, 전체 텍스트의 응집력은 특정한 응집력의 자질이 전체 텍스트에 걸쳐서 나오는 빈도를 말한다. 예를 들면 특정 단어가 에세이 전체에 반복적으로 나온 빈도를 말한다.

국부적 응집력, 총체적 응집력, 전체 텍스트 응집력의 차이와 예시를 정리하면 <표 1>과 같다. <표 1>의 예시에서 해당되는 응집력 지표는 이탤릭체로 제시되어 있다. 총체적 응집력은 실제로는 단락 간에 유사한 의미를 갖는 단어나 구가 반복되는 것을 말하나, 예시는 이해를 돕기 위해 문장 간에 반복된 단어가 간략히 제시되어 있다.

표 1. 국부적, 총체적, 전체 텍스트 응집력의 비교(Crossley, Kyle and McNamara 2016)

응집력의 분류	단위	지표의 예	예시
국부적	문장	연결사	<i>First, she was happy.</i>
총체적	단락	단어나 구의 의미적 유사성	<i>The animal was small.</i> <i>It was a cat.</i>
전체 텍스트	전체 에세이	특정 단어	<i>The big dog saw the big cat.</i>

1.2 TAACO 프로그램 소개

본 연구에서는 최근에 개발된 자동화된 텍스트 분석 프로그램인 TAACO(the Tool for the Automatic Analysis of Cohesion)를 활용하여 텍스트 응집력을 분석하였다. TAACO 프로그램이 개발되기 이전에는 대부분의 선행연구에서는 Coh-Metrix 프로그램을 활용하였다. Coh-Metrix 프로그램은 텍스트 응집력뿐 아니라 어휘의 정교화, 통사적 복잡성, 텍스트 구조를 자동적으로 분석할 수 있다. Coh-Metrix 프로그램은 25개의 텍스트 응집력만을 분석하며 인터넷 접속이 반드시 있어야 하고 한 번에 한 개의 에세이만 분석이 가능하다는 제약이 있다. 하지만 TAACO 프로그램은 168개의 텍스트 응집력을 분석할 수 있으며 한 번에 분석 가능한 에세이의 개수에 제한이 없고 인터넷의 접속도 필요로 하지 않는다는 장점을 갖는다(Crossley et al. 2015). TAACO 프로그램과 Coh-Metrix 프로그램을 비교하면 아래의 <표 2>와 같다.

표 2. TAACO 프로그램과 Coh-Metrix 프로그램의 비교

	TAACO 프로그램	Coh-Metrix 프로그램
분석 대상	텍스트 응집력	어휘의 정교화, 텍스트 응집력, 통사적 복잡성, 텍스트 구조
분석 가능한 응집력 지표 개수	168개	25개
분석 가능한 응집력의 종류	1) 국부적 2) 전체 텍스트 3) 총체적	1) 국부적 2) 전체 텍스트
인터넷 접속 필요 여부	접속 필요 없음	접속이 반드시 필요함
한 번에 분석 가능한 에세이 개수	개수 제한 없음	1개

Kyle이 개발한 현재의 가장 최신 버전인 TAACO 2.0 프로그램의 화면 예시는 <그림 1>과 같다.

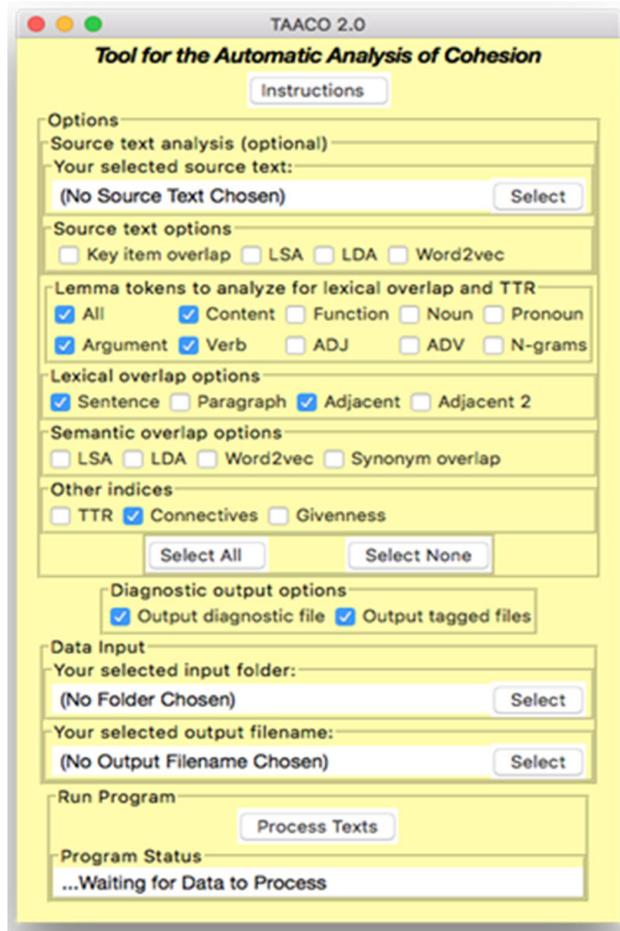


그림 1. TAACO 2.0 프로그램 화면 예시

TAACO 프로그램은 분석하고자 하는 응집력 지표를 모두 체크하여 txt 형태의 에세이 파일을 선택하면 분석 처리가 시작되고, 분석이 완료되면 응집력 지표 결과 파일뿐 아니라 진단파일도 제공된다.

### 1.3 선행 연구와의 차별화

본 연구에서 살펴보고자 하는 구체적인 연구 질문은 다음과 같다. 첫째, 한국 대학생이 작성한 에세이에서 텍스트 응집력의 특성이 영어 수준별로 통계적으로 유의미하게 다르게 나타나는가? 둘째, 상 수준, 중 수준, 하 수준의 세 개의 집단으로 분류할 수 있는 텍스트 응집력 지표에는 무엇이 있는가?

본 연구는 한국인 에세이 코퍼스의 영어 수준별 비교를 통해 텍스트 응집력에서 어떠한 차이점이 있는지를 살펴보고, 다음의 세 가지 측면에서 선행 연구와 차별화된다. 첫째, TAACO는 2015년에 개발된 텍스트 응집력 분석 도구로 국내에서 이를 활용하여 한국 학생들의 에세이를 분석한 연구가 거의 없다. 즉, 본 연구는 TAACO를 활용하여 한국 학생의 영어 수준별로 텍스트 응집력의 특성이 어떻게 다르게 나타나는가를 탐색하는 최초의 연구이다. 둘째, 본 연구는 국내의 선행

연구에서 교과서의 학년별 연계성 분석이라는 기존의 코퍼스 연구 주제에서 탈피하여, 한국인 에세이 코퍼스에서 영어 수준별로 응집력의 특성이 어떠한 차이를 보이는가를 살펴보고자 한다. 셋째, 본 연구에서는 보다 정교화된 통계분석 방법을 활용하여 영어 수준별로 텍스트의 응집력의 특성이 어떻게 다르게 나타나는가를 살펴보고자 한다. 지금까지 수행된 해외의 선행 연구는 에세이의 점수를 가장 잘 예측한 TAACO의 응집력 지표가 무엇인지를 단계적 회귀분석(stepwise regression analysis)을 통해 살펴보았으나, 본 연구에서는 판별분석(discriminant function analysis)을 통해 응집력 지표와 영어 수준 간의 관계를 자세히 살펴보고자 한다.

## 2. 선행 연구 검토 및 비교

텍스트의 응집력을 비교한 과거의 선행 연구는 언어적 요소와 수사적 구조에 대해 손으로 직접 세어서 계산했으며, 이러한 분석 결과는 계산상의 오류를 수반하거나 살펴보고자 하는 에세이 개수 혹은 언어 요소의 제약이 따를 수밖에 없다(Crossley and McNamara 2009, Reid 1992). Coh-Metrix의 자동화된 텍스트 분석 프로그램이 개발되면서 많은 연구가 Coh-Metrix 라는 프로그램을 활용하여 텍스트의 응집력을 분석하였으나, 지금까지 대부분의 선행 연구(Crossley and McNamara 2012, Guo, Crossley and McNamara 2013, Jung, Crossley and McNamara 2015)는 주로 해외에서 수행되었다.

국내의 경우 Coh-Metrix 프로그램을 활용하여 교과서의 학년 간 위계를 분석한 연구(김정렬 2014, 송해성 2013, 전문기, 임인재 2009)가 수행되었으나 실제로 에세이를 분석한 연구는 김소정과 전문기(2016)를 제외하면 많지 않은 실정이다. 김소정과 전문기(2016)는 초등학교 6학년 학생의 에세이 176개를 Coh-Metrix 프로그램을 활용하여 분석했으며, 연구 결과를 요약하면 다음과 같다. 첫째, 상위 수준의 초등학생은 많은 수의 단어를 사용하여 긴 문장을 썼다. 둘째, 단어 빈도수, 어휘의 다양성에 대해서는 영어 수준별로 통계적으로 유의미한 차이를 보이지 않았으며, 이는 상·중·하의 세 집단 간에 어휘력은 비슷하다는 것을 보여준다. 셋째, 세 집단 간에 읽기 가독성 지표에 대해 통계적으로 유의미한 차이를 보였는데, 이는 상위 수준의 초등학생이 작성한 에세이의 이해도가 가장 높다는 것을 보여준다. 넷째, 상위 수준의 초등학생은 하위 수준의 학생보다 더 많은 접속사와 대명사를 활용하여 에세이를 작성했다. 즉, 이는 상위 수준의 초등학생이 하위 수준 보다 더 응집력 있는 에세이를 작성했음을 시사한다.

Guo, Crossley와 McNamara(2013)는 어휘의 정교화(lexical sophistication), 통사적 복잡성(syntactic complexity), 응집력과 같은 언어자질이 TOEFL iBT 쓰기 과업의 에세이 점수를 예측할 수 있는지를 Coh-Metrix 프로그램을 활용하여 분석하였다. TOEFL iBT 쓰기 과업은 독립형 과업(independent task)과 통합형 과업(integrated task)으로 구분된다. 독립형 과업은 학생들로 하여금 주어진 주제에 대해 논리적으로 에세이를 작성하도록 하며, 통합형 과업은 특정 주제에 대해 서로 상반된 입장을 나타내는 듣기 자료와 읽기 지문이 제시된 후에 요약문을 작성하도록 한다. 연구 결과를 살펴보면 다음과 같다. 상의 반복(aspect repetition), 내용어 반복(content word overlap), 조건적 연결사(conditional connectives)와 같은 국부적 응집력 지표(indices of local cohesion)는 독립형 과업의 에세이 점수를 부정적으로 예측하는 것으로 나타났다. 반면에 듣기와 읽기 지문의 이해를 요구하는 통합형 과업은 국부적 응집력 지표가 에세이 점수를 정적으로 예측하는 것으로 나타났다. 이러한 연구결과는 TOEFL iBT의 독립형 과업과 통합형 과업이 서로 다

른 장르에 속한다는 것을 시사한다.

Crossley와 McNamara(2012)는 홍콩의 고등학생이 작문 시험으로 작성한 514개 에세이의 응집력과 어휘의 정교화를 Coh-Metrix를 활용하여 분석하였다. 이를 위해 514개의 에세이를 예비 분석용 에세이 세트(training set) 344개와 실제 분석용 에세이 세트(testing set) 170개의 두 개로 분류하였다. 주요 연구 결과를 살펴보면 다음과 같다. 첫째, 어휘의 다양성, 단어 친숙도(word familiarity), 내용어 빈도(content word frequency), 어휘의 의미성(word meaningfulness), 상의 반복(aspect repetition)의 5개 변인은 에세이 점수 변인의 21%를 차지하는 것으로 나타났다. 둘째, 상위 수준 학생의 경우 어휘의 반복이나, 상의 반복과 같은 응집력을 덜 사용하는 것으로 나타났다. 셋째, 어휘의 정교화의 경우, 상위 수준의 학생일수록 친숙하지 않고(less familiar), 구체적이지 않으며(less concrete), 이미지를 형성하기 어려운(less imageable) 어휘를 사용하는 것으로 나타났다.

최근에 텍스트 응집력만을 분석하는 도구인 TAACO 프로그램이 개발되었고 이를 활용하여 분석한 연구가 해외에서만 수행되었다. Crossley, Kyle과 McNamara(2016)는 한 학기 동안 텍스트 응집력이 어떻게 발전되는지를 TAACO 프로그램을 활용하여 분석하였다. 연구에 참여한 학생은 57명의 ESL 학생이며, 학기 초, 학기 중, 학기 말에 각각 30분 동안 에세이를 써서 제출했다. 연구결과 한 학기 동안 통계적으로 가장 유의미한 차이를 보인 응집력 지표는 단락 간의 명사 반복(adjacent overlap two paragraphs: noun)으로 나타났다. 회귀분석 결과 에세이 점수를 통계적으로 유의미하게 예측한 변인은 기능어와 연관된 응집력 지표로, 단락 간의 기능어 반복(adjacent overlap two paragraphs: function words), 문장 간의 기능어 반복(adjacent overlap two sentences: function words), 단락 간의 대명사 반복(adjacent overlap two paragraphs: pronouns)으로 나타났다.

지금까지 텍스트 응집력에 대한 선행 연구는 크게 두 가지로 요약될 수 있다. 첫째, 대부분의 선행 연구는 국내외 모두 Coh-Metrix 라는 프로그램을 활용하였다. 해외에서는 Coh-Metrix 프로그램을 활용하여 에세이의 응집력과 어휘의 정교화를 분석한 연구가 많으나, 국내에서는 에세이의 분석 보다는 영어 교과서의 학년 간 연계성을 분석한 연구가 많다. 둘째, 자동화된 텍스트 응집력 분석 프로그램인 TAACO가 개발되면서 이를 활용하여 ESL 학생을 대상으로 분석한 연구가 해외에서만 제한적으로 수행되었다. 따라서 본 연구는 TAACO 프로그램을 활용하여 한국 대학생의 에세이를 분석하고 응집력의 특성이 영어 수준별로 어떻게 다르게 나타나는가를 살펴본 후에, 영어 쓰기 교육의 제언을 도출하고자 하는 목적으로 수행되었다.

### 3. 연구 방법

#### 3.1 연구 대상

연구 대상은 ICNALE(International Corpus Network of Asian Learners of English) 코퍼스에서 추출한 한국 대학생 300명이 작성한 에세이 600개이다. 에세이 주제는 대학생이 파트타임 직업을 갖는 것이 왜 중요한지와 흡연이 레스토랑에서 완전히 금지되어야 하는지에 대해 자신의 생각을 한 단락으로 40분 동안 작문하는 것이다.

본 연구에 사용된 ICNALE 코퍼스의 한국 대학생 에세이에 대한 정보는 <표 3>에 제시되어

있다. 한국 학생 에세이 600개 중에서 파트타임 주제 에세이는 300개, 흡연 주제 에세이는 300개가 포함되어 있다. <표 3>에 제시된 것처럼 파트타임 주제와 흡연 주제 에세이의 평균 단어 수는 각각 225.05, 225.24로 비슷했으며 표준편차는 26.88, 29.14로 흡연 주제 에세이가 약간 크게 나타났다.

표 3. 한국 대학생 에세이 단어 수의 기술통계

작문 주제	최소값	최대값	평균	표준편차	범위
파트타임	181	334	225.05	26.88	153
흡연	181	325	225.24	29.14	144

한국 학생 300명 중에서 75명은 유럽 공동 표준기준인 CEFR(Common European Framework of Reference)의 A2, 149명은 B1, 76명은 B2 레벨에 속한다. CEFR의 A2 수준은 토익 성적의 545 미만, B1 수준은 550 이상 670 미만, B2 수준은 670 이상 800 미만에 속한다. 즉, 300명 학생을 영어 능숙도로 분류하면 75명은 하 수준, 149명은 중 수준, 76명은 상 수준에 속한다고 볼 수 있다.

### 3.2 자료분석

SPSS 25.0을 이용하여 다변량 분산분석과 단계적 판별분석(a stepwise discriminant function analysis)의 통계분석이 수행되었다. 상·중·하의 영어 수준별로 응집력 지표가 통계적으로 유의미하게 다르게 나타나는지를 살펴보기 위해 다변량 분산분석을 수행하였다. 또한 텍스트 응집력의 지표와 영어 수준 간의 관계를 파악하기 위해 판별분석을 수행하였다. 판별분석에서 독립변인은 응집력 지표이고, 종속변인은 CEFR로 분류한 상·중·하의 영어 능력 수준이다. 판별분석을 통해 통계적으로 유의미한 집단분류 능력이 있는 응집력 지표는 무엇이며 영어 수준별로 응집력 지표의 양상이 어떻게 다른지를 탐색해 볼 수 있다.

## 4. 연구 결과

### 4.1 MANOVA 분석

영어 수준별로 텍스트 응집력의 특성이 통계적으로 유의미하게 다르게 나타나는가를 알아보기 위해 다변량 분산분석을 수행하였으며, 결과는 <표 4>에 제시되어 있다. 34개의 텍스트 응집력 지표가 통계적으로 유의미한 것으로 나타났으며(Wilks' lambda=0.578, F=3.380, p<0.000) 이는 상·중·하의 세 개의 집단에 따라 응집력 지표의 평균벡터에 유의미한 차이가 있고 각각의 집단은 응집력 지표의 사용이 유의미하게 다르다는 것을 보여준다.

표 4. 다변량 분산분석 결과

	람다 값	F	자유도	유의도
영어 능숙도 집단	0.578	3.380	102	0.000

4.2 단계적 판별분석

한국 학생을 상·중·하의 세 개의 집단으로 유의미하게 분류하는 응집력 지표가 무엇인지를 알아보기 위해 단계적 판별분석이 수행되었고, 그 결과는 <표 5>와 같다. 상·중·하의 세 개의 영어 수준으로 구분하는 판별함수가 도출되었고, 도출된 판별함수는 통계적으로 유의미하였다. 판별함수 1에 의한 상 수준 학생의 집단중심점은 0.841, 중 수준 학생의 집단중심점은 -0.052, 하 수준 학생의 집단중심점은 -0.75로 나타났으며 영어 수준이 높아질수록 집단중심점이 높게 나타났다. 즉, 영어 수준이 높을수록 응집력 지표의 평균이 더 높다는 것을 알 수 있다. <표 5>에 제시된 구조행렬 계수는 응집력 지표의 상대적 중요도로 구조행렬 계수가 0.33 이상인 지표가 변량의 10% 이상을 차지하기 때문에 판별함수 해석에서 유의미하다(Tabachnik and Fidell 2007).

표 5. 단계적 판별분석 요약

판별함수		함수 1	함수 2
구조 행렬	대명사의 밀도	0.573*	0.218
	명사와 대명사의 중복이 다음 문장에 반복되는 문장의 개수	0.371*	0.062
	연결사	-0.357*	-0.208
	두 개의 문장에 반복되는 명사와 대명사의 개수	0.352*	0.35
	렘마의 빈도	0.332*	-0.171
집단 중심점	A2 (하 수준 집단)	-0.750	0.222
	B1 (중 수준 집단)	-0.052	-0.199
	B2 (상 수준 집단)	0.841	0.171
정준상관계수		0.494	0.195
아이겐 값		0.322	0.040
설명변량 (%)		89.1	10.9
람다값		0.727	0.962

첫 번째 판별함수에 의해 상 수준 학생을 하 수준과 중 수준 학생으로부터 구별 가능한 응집력 지표에는 5개가 있으며, 대명사의 밀도(pronoun\_density, 구조행렬 계수 = 0.573), 명사와 대명사의 중복이 다음 문장에 반복되는 문장의 개수(adjacent\_overlap\_2\_argument\_sent, 구조행렬 계수 = 0.371), 연결사(basic\_connectives, 구조행렬 계수 = -0.357), 두 개의 문장에 반복되는 명사와 대명사의 개수(adjacent\_overlap\_binary\_argument\_sent, 구조행렬 계수 = 0.352), 렘마의 빈도(lemma\_mattr, 구조행렬 계수 = 0.332)의 순으로 나타났다. 단계적 판별분석 결과 통계적으로 유의미한 5개의 응집력 지표 중에서 대명사의 밀도와 렘마의 빈도는 전체 텍스트 응집력(overall text cohesion), 명사와 대명사의 중복이 다음 문장에 반복되는 문장의 개수, 연결사, 두 개의 문장에 반복되는 명사와 대명사의 개수는 국부적 응집력(local cohesion)에 속한다.

판별함수를 영어 능숙도의 수준별로 도식화 하면 <그림 2>, <그림 3>, <그림 4>와 같다. 하 수준의 경우 판별함수 평균은 -0.75, 중 수준 학생의 판별함수 평균은 -0.05, 상 수준 학생의 판별함수의 평균은 0.841이고, <표 5>의 단계적 판별분석 요약에 제시된 집단 중심점과 같다. 상 수준 학생의 집단 중심점이 가장 높았는데 이는 상 수준 학생의 경우 전반적으로 응집력 지표를 다른 집단의 학생에 비해 많이 사용했음을 단적으로 보여준다.

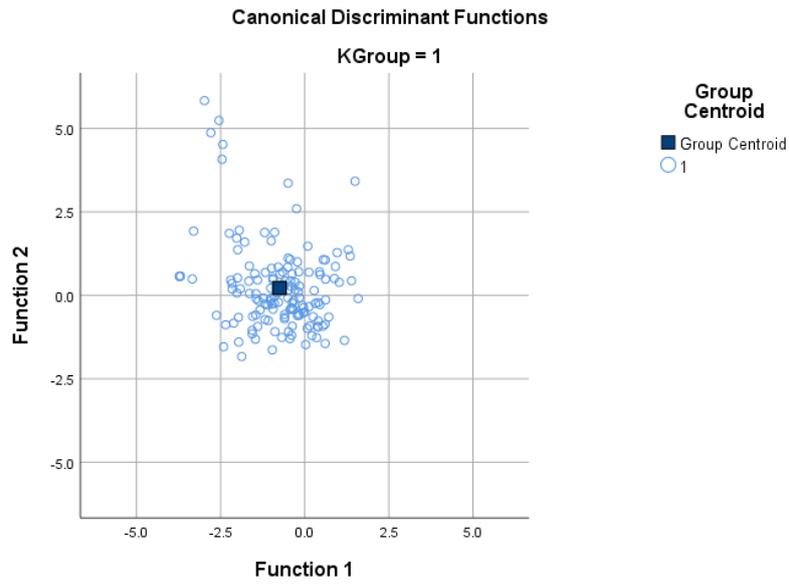


그림 2. 하 수준 학생의 판별함수

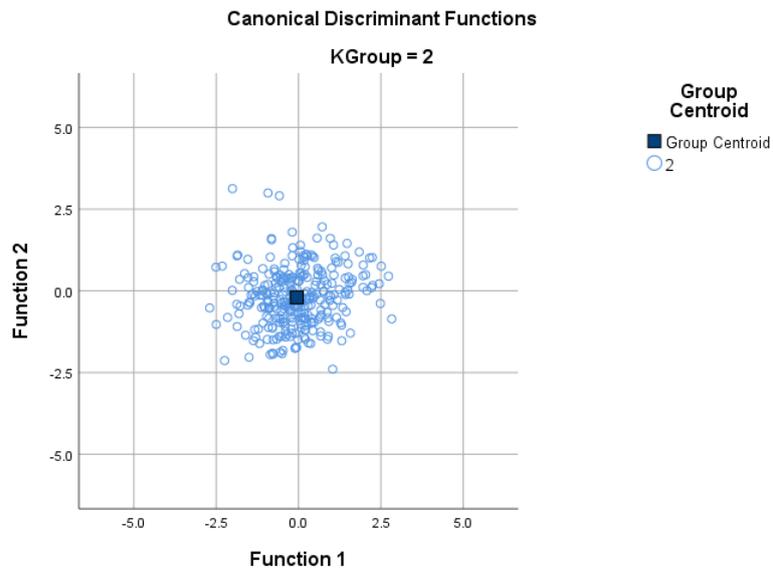


그림 3. 중 수준 학생의 판별함수

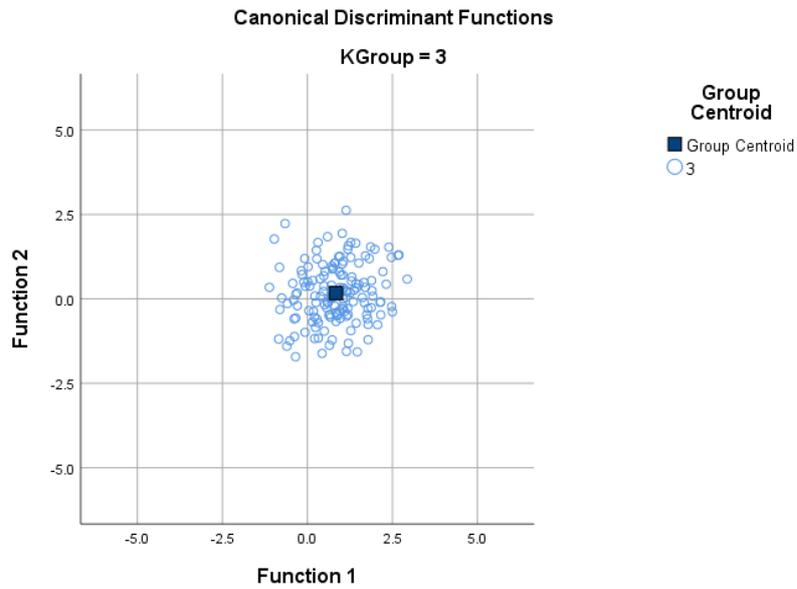


그림 4. 상 수준 학생의 판별함수

전체 학생의 판별함수는 <그림 5>와 같다. 판별함수 X축을 기준으로 상 수준은 오른쪽에, 중 수준은 가운데에, 하 수준은 왼쪽에 전반적으로 분포되어 있으며, 영어 능숙도에 따라 세 집단이 다르게 분포되어 있다.

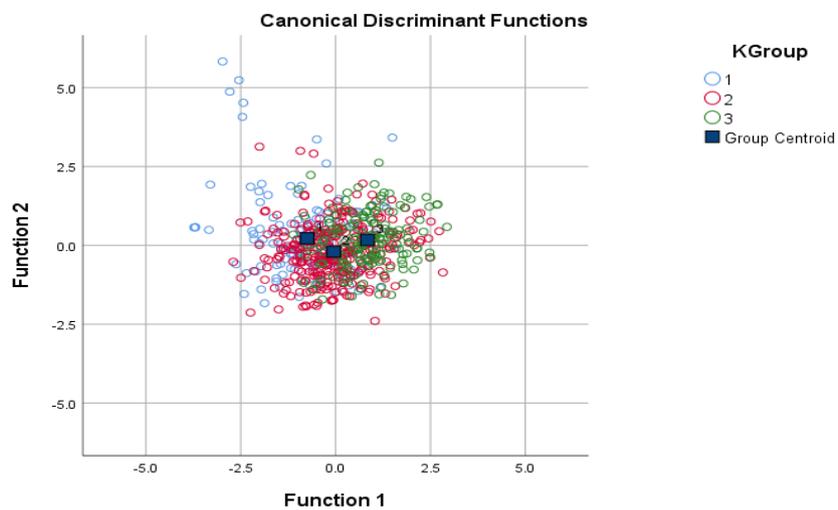


그림 5. 전체 학생의 판별함수

영어 수준별로 에세이를 유의미하게 구분하는 다섯 개 응집력 지표의 평균값은 <표 6>에 제시되어 있다.

표 6. 판별분석 결과 유의미한 5개 응집력 지표의 평균 비교

	하 수준	중 수준	상 수준
대명사의 밀도	0.0384	0.0460	0.0607
명사와 대명사의 중복이 다음 문장에 반복되는 문장의 개수	0.2984	0.3241	0.3632
연결사	0.0405	0.0381	0.0323
두 개의 문장에 반복되는 명사와 대명사의 개수	0.5795	0.5991	0.6876
렘마의 빈도	0.7136	0.7269	0.7371

대명사의 밀도의 경우 구조행렬 계수는 0.573이며, 상 수준 학생의 평균은 0.0607, 중 수준 학생의 평균은 0.0460, 하 수준 학생의 평균은 0.0384로 상 수준 학생의 평균이 가장 높게 나타났다. 즉, 상 수준 학생의 경우 대명사의 사용이 에세이 전반에 걸쳐 응집력 있게 나타난다는 것을 알 수 있다.

명사와 대명사의 중복이 다음 문장에 반복되는 문장의 개수의 경우 구조행렬 계수는 0.371로 두 번째로 높게 나타났으며, 상 수준 학생의 평균은 0.3632, 중 수준 학생의 평균은 0.3241, 하 수준 학생의 평균은 0.2984로 상 수준 학생의 평균이 가장 높게 나타났다. 즉, 상 수준 학생의 경우 명사와 대명사가 다음 문장에 반복되는 응집력의 특징을 갖는다는 것을 알 수 있다.

연결사의 경우 구조행렬 계수는 -0.357의 음수 값을 가지며, 상 수준 학생의 평균은 0.0323, 중 수준 학생의 평균은 0.0381, 하 수준 학생의 평균은 0.0405로 하 수준 학생의 평균이 가장 높게 나타났다. 즉, 하 수준 학생은 상 수준 학생에 비해 단순 연결사를 과잉 사용하였으며, 상 수준 학생은 단순 연결사를 많이 사용하지 않음을 알 수 있다.

두 개의 문장에 반복되는 명사와 대명사의 개수의 경우 구조행렬 계수는 0.352이며, 상 수준 학생의 평균은 0.6876, 중 수준 학생의 평균은 0.5991, 하 수준 학생의 평균은 0.5795로 상 수준 학생의 평균이 더 높게 나타났다. 즉, 상 수준 학생의 경우 두 개의 문장에 반복되는 명사와 대명사의 빈도가 높은 특징을 갖는다는 것을 알 수 있다.

렘마는 고유한 단어 형태로 예를 들면, sing, sings, singing의 렘마는 sing이 된다. 렘마의 빈도는 50 단어를 기준으로 한 평균 렘마의 빈도이다. 렘마의 빈도의 경우 구조행렬 계수는 0.332의 값을 가지며, 상 수준 학생의 평균은 0.7371, 중 수준 학생의 평균은 0.7269, 하 수준 학생의 평균은 0.7136으로 영어 수준이 높아질수록 근소한 차이로 높게 나타났다. 즉, 상 수준 학생의 경우 어휘 지식이 높아 고유한 단어를 사용하는 렘마의 빈도는 높게 나타나는 특징을 보임을 알 수 있다.

#### 4.3 에세이 분석

영어 수준별로 응집력 지표의 사용에 있어서 어떠한 차이가 있는지를 단계적 판별분석 결과에서 가장 유의미한 지표로 나타난 대명사의 밀도를 통해 살펴보고자 한다. 예시는 파트타임 주제에 대해 작문한 에세이이며 첫 번째 에세이는 상 수준 학생, 두 번째는 하 수준 학생이 작성한 것이다. 대명사의 밀도 지표의 값이 상 수준 학생은 0.141553, 하 수준 학생은 0.004651로, 상 수준 학생이 더 높게 나타났으며, 에세이 예시에서 명사와 이를 지칭하는 대명사는 볼드체로 표시되어 있다

I think that it is very important for **college student** to do some part time jobs. At this age, **they** are not just a baby of **their** parents anymore. **They** are not also just students

of the school anymore who can't do his or her own works alone. As **a college student**, **they** have to prepare for **their** independent life when graduating from the school. Therefore, **they** have to learn how to earn the money to sustain **their** lives themselves. (상 수준 학생 에세이, 대명사의 밀도 지표 값은 0.141553)

So, I want to suggest to college students that they have to get **a part-time job** to make money by themselves. **Part-time job** is good way to make money even though they are not graduated from their school. **Part-time job** can make people earn money not giving up their study. Also, there is another good point of doing **part-time job**. If college students do **the part-time job**, they can feel economical society prior to getting a job. (하 수준 학생 에세이, 대명사의 밀도 지표 값은 0.004651)

상 수준 학생은 명사 college students 대신에 대명사 they, their, themselves를 사용하여 응집력 있는 에세이를 완성한 반면에, 하 수준 학생은 대명사를 사용하기보다는 명사 part-time job 을 반복적으로 사용하는 특징을 보였다.

영어 수준별로 접속사의 사용에 있어서 어떠한 차이가 있는지를 에세이 예시를 통해 살펴보면 다음과 같다. 아래의 예시는 흡연 주제에 대해 작문한 에세이이며 첫 번째 에세이는 상 수준 학생, 두 번째는 하 수준 학생이 작성한 것이다. 접속사 지표의 값이 상 수준 학생은 0.003289, 하 수준 학생은 0.07451로, 하 수준 학생이 더 높게 나타났으며, 에세이 예시에서 접속사는 볼드체로 표시되어 있다.

Some people argue that we should not completely ban smoking at all the restaurants. They say that it violates the human freedom right. **However**, I disagree with this argument **because** if human right is important, we should also respect other people's right to be healthy as well as our own right of freedom. Above all, I oppose to the idea that we should allow people to smoke anywhere including public areas **because** it is the same as the behavior to make other people's health worse. (상 수준 학생 에세이, 접속사 지표 값은 0.003289)

Do you like smoking **and** where will you smoke? Like it **or** you are alone? Smoking is a bad ways not in home **but** outside if you do something yourself **and** your friends **or** others, you can smoke in WC **or** outside you are not live for yourself **but** others, you can do it in the place where have nobody if you can you should not smoking in somewhere if you should do it never **because** it is a bad habited health is important to protect your health someone think smoking is a way to relax yourself **but** in public it is harmful to others. (하 수준 학생 에세이, 접속사 지표 값은 0.07451)

상 수준 학생은 접속사를 효율적으로 사용하여 앞뒤 문장을 논리적으로 연결했는데, 예를 들면, 세 번째 문장에서 대조를 나타내는 however와 이유를 나타내는 because를 사용하여 앞뒤 문장을 응집력 있게 연결하였다. 이와는 대조적으로 하 수준 학생은 접속사를 과잉 사용(overuse) 하였

는데, and, or, but, because가 한 단락에서 10번 단순 반복하는 특징을 보였다. 이는 EFL 학습자의 경우 연결사를 과잉 사용 하는 경우가 많으며, 의미의 고려 없이 사용한 연결사는 에세이의 응집력을 높이는 데에 도움이 되지 않는다는 Hinkel(2001)의 연구 결과와 유사하다.

## 5. 결론

텍스트 응집력은 쓰기 능력의 핵심요소로 텍스트의 이해도에 영향을 미치는 중요한 요인이다. 텍스트의 응집력에 대한 선행 연구(Connor 1984)는 언어적 요소와 수사적 구조에 대해 손으로 직접 세어서 계산했으며, 에세이 개수나 언어 요소의 제약이 수반되어 연구 결과를 일반화하기에는 제한점이 많았다. 하지만 Coh-Metrix와 같은 텍스트 분석 프로그램이 개발되면서 해외의 많은 연구는(Crossley and McNamara 2012, Guo, Crossley and McNamara 2013, Jung, Crossley and McNamara 2015) 이를 활용하여 텍스트의 응집력을 분석하였다. 최근에 응집력만을 분석하는 TAACO 프로그램이 개발되었으며 이를 활용하여 응집력 지표(cohesion indices)가 영어 쓰기 능력과 어떠한 관련이 있는지를 보다 더 심층적으로 분석할 수 있게 되었다.

본 연구에서는 ICNALE 코퍼스에서 추출한 한국인 대학생 에세이 600개를 TAACO 프로그램을 활용하여 분석하였으며, 영어 수준별로 응집력의 특성이 어떠한 차이를 보이는가를 정교화된 통계 분석 방법을 활용하여 살펴보았다. 본 연구에서 살펴본 연구 질문은 두 개이며, 한국 대학생이 작성한 에세이에서 텍스트 응집력의 특성이 영어 수준별로 통계적으로 유의미하게 다르게 나타나는지와 상 수준, 중 수준, 하 수준의 세 개의 집단으로 분류할 수 있는 텍스트 응집력 지표에는 무엇이 있는가이다. 본 연구에서는 영어 수준별로 텍스트 응집력의 특성이 어떠한 양상으로 다르게 나타나는지를 다변량 분산분석과 단계적 판별분석을 통해 살펴보았다. 텍스트 응집력의 특성이 영어 수준별로 통계적으로 유의미하게 다르게 나타나는가를 알아보기 위해 다변량 분산분석을 수행하였다. 또한 판별분석을 통해 상·중·하의 영어 수준이 어떠한 텍스트 응집력 지표의 특징적 양상으로 설명될 수 있는지를 살펴보았다.

본 연구 결과를 요약하면 다음과 같다. 첫째, 다변량 분산분석 결과, 34개의 텍스트 응집력 지표가 상·중·하의 세 개의 집단에 따라 평균벡터에 유의미한 차이가 있고 각각의 집단은 응집력 지표의 사용이 유의미하게 다르게 나타났다. 둘째, 판별분석 결과, 상 수준 학생을 하 수준과 중 수준 학생으로부터 구별 가능한 응집력 지표에는 5개가 있으며, 대명사의 밀도, 명사와 대명사의 중복이 다음 문장에 반복되는 문장의 개수, 연결사, 두 개의 문장에 반복되는 명사와 대명사의 개수, 램마의 빈도의 순으로 나타났다.

상 수준 학생을 하 수준과 중 수준 학생으로부터 최대로 구별 가능했던 응집력 지표는 3인칭 대명사의 빈도를 전체 단어 수로 나눈 대명사의 밀도로, 상 수준 학생은 3인칭 대명사의 사용이 높다는 것을 알 수 있다. 이는 단락 간의 대명사 반복 지표가 에세이 점수를 통계적으로 유의미하게 예측했다는 Crossley et al.(2016)의 연구 결과와 비슷하다. 또한 원어민과 비원어민이 작성한 에세이로 구분하는 유의미한 응집력 지표가 대명사와 같은 주어진 정보를 활용하는 것이라는(Crossley and McNamara 2009) 선행연구 결과와도 유사하다. 연결사의 경우 구조행렬 계수의 값이 -0.357로 음수 값을 갖는데, 이는 상 수준 학생의 경우 연결사를 많이 사용하지 않았고, 영어 수준이 낮아질수록 단순 연결사 지표의 사용이 높다는 것을 나타낸다. 램마의 경우 영어 수준이 높아질수록 고유한 단어 형태인 램마의 빈도가 약간 높게 나타났다. 이는 상 수준 학생은 어휘

수준이 높아 같은 단어를 반복적으로 사용하기보다는 고유한 형태를 갖는 단어를 많이 사용했음을 알 수 있다.

본 연구 결과를 요약하면, CEFR의 B2 수준에 상응하는 상 수준 한국 학생의 경우, 대명사의 사용이 에세이 전반에 걸쳐 응집력있게 나타나고 명사와 대명사가 뒷 문장이나 두 개의 문장에 반복되는 양상을 보이며 어휘의 다양성으로 인해 렘마의 빈도는 높으나 단순 연결사를 많이 사용하지 않는 특징을 보였다. 따라서 중 수준과 하 수준 학생들의 경우 단순 연결사의 과잉 사용(overuse)을 지양하며 명사와 대명사의 효율적인 사용을 강화하여 응집력 있는 에세이를 작성하도록 지도할 필요가 있다.

본 연구 결과는 영어 쓰기 교육에서 다음과 같이 활용될 수 있다. 첫째, 영어 수준별로 텍스트 응집력의 특성이 어떻게 다르게 나타나는가를 고찰함으로써 언어습득과 영어 교육에서 결과가 활용될 수 있다. 즉, 교사는 하 수준 학생들의 경우 텍스트 응집력에서 상 수준 학생들과 어떻게 다르게 나타나는가에 대한 정보를 통해 하 수준 학생들을 위한 맞춤형 지도를 할 수 있다. 둘째, 본 연구 결과를 바탕으로 실효성 있고 획기적인 쓰기 교육의 자료를 제공할 수 있다. 하 수준 학생들에게 부족한 응집력의 특성인 대명사의 적절한 사용과 연속된 문장에 반복되는 명사와 대명사의 활용을 강화하는 쓰기 지도가 필요하다. 특히, 문장 간의 응집력을 강화할 수 있는 응집력 장치(Park 2005)에 대한 연습과 훈련의 기회가 마련되어야 하며, 이를 통해 에세이 점수 향상에 기여할 수 있을 것이다(Yang and Sun 2012).

## 참고 문헌

- 김소정·전문기(Kim, S-J. and M-G. Jeon). 2016. Coh-Metrix를 활용한 초등학교 6학년 영어학 습자의 영작문 분석 연구(An analysis study of English writing of elementary school 6th grade English language learners using Coh-Metrix). 《현대영어교육》(Modern English Education) 17(3), 263-287.
- 김정렬(Kim, J.). 2014. 초중등 영어교육 연계성 문제: 영어 교과서의 계열성을 중심으로 (Continuity problems of elementary and secondary English education: Sequence analysis of English textbooks). 《어학연구》(Language Research) 50(1), 161-184.
- 송해성(Song, 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.
- 전문기·임인재(Jeon, M-G. and I-J. Lim). 2009. 코메트릭스(Coh-metrix)를 이용한 중학교 1 학년 개정 영어 교과서의 코퍼스 언어학적 비교 분석(A Corpus-based analysis of middle school English 1 textbooks with Coh-Metrix). 《영어교육연구》(The Journal of English Education) 21(4), 265-292.
- Connor, U. 1984. A study of cohesion and coherence in ESL students' writing. *International Journal of Human Communication* 17, 301-316.
- Crossley, S. A. and D. S. McNamara. 2009. Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing* 18, 119-135.

- Crossley, S. A. and D. S. McNamara. 2012. Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading* 35, 115–135.
- Crossley, S. A., K. Kyle and D. S. McNamara. 2015. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods* 48(4), 1227–1237.
- Crossley, S. A., K. Kyle and D. S. McNamara. 2016. The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing* 32(2), 1–16.
- Guo, L., S. A. Crossley and D. S. McNamara. 2013. Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing* 18, 218–238.
- Halliday, M. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hinkel, E. 2001. Matters of cohesion in L2 academic texts. *Applied Linguistics* 12, 111–132.
- Jung, Y., S. A. Crossley and D. S. McNamara. 2015. Linguistic features in MELAB writing performances. *CaMLA Working Paper* 5, 1–17.
- Park, T. 2005. Putting cohesion in its place: Justification for teaching sentence cohesion in EFL writing classes. *English Language and Linguistics* 19, 201–214.
- Reid, J. R. 1992. A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing* 1, 79–107.
- Tabachnik, B. and L. Fidell. 2007. *Using multivariate statistics*. Boston: Pearson Education.
- Witte, S. P. and L. Faigley. 1981. Coherence, cohesion, and writing quality. *College Composition and Communication* 32, 189–204.
- Yang, W. and Y. Sun. 2012. The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education* 23, 31–48.

Examples in: English

Applicable Languages: English

Applicable Level: Secondary/Tertiary