

텍스트 마이닝을 활용한 중·고등학생의 쓰기 인식 연구 —은유 분석을 중심으로

이슬기 한국교원대학교 국어교육과 박사과정

- I. 머리말
- II. 이론적 배경
- III. 연구 방법 및 절차
- IV. 연구 결과 및 논의
- V. 맺음말

I. 머리말

인식이란 대상에 대해 경험적이고 이론적으로 파악하고자 하는 의식 활동으로, 학습자의 인식은 단순히 내면적 양상으로 형성된 결과라기보다 학습자가 속한 환경 속에서 형성되고 잠재된 형태로 볼 수 있다(장영란, 2000). 쓰기 연구자들은 학습자의 쓰기 인식에 지속적인 관심을 보여 왔는데 쓰기 인식은 쓰기 행위에 대한 학습자 스스로의 인식인바, 학습자의 쓰기 인식을 파악한다는 것은 학습자 개인만이 아니라 학습자를 둘러싼 쓰기 환경¹⁾에 대한 이해를 높이는 과정이기 때문이다. 학습자에 대한 정확한 이해는 적절한 교육 처치를 가능케 하기 때문에 연구자들은 대상과 내용의 범주를 달리해 가며 학생들의 쓰기 인식에 대해 분석해 왔다.

대표적으로 초·중·고등학생(김정자, 2014; 박혜영, 2016; 장은섭, 2014; 장은주·박영민, 2013)과 대학 신입생을 대상으로 한 연구(한금윤·최선경,

1) 여기에서 언급한 '환경'은 매우 광범위한 의미로 교수자, 또래집단 등의 사회적 관계 및 교실 같은 물리적 환경 등 쓰기에 영향을 미치는 제반을 총칭하는 뜻으로 사용하였다.

2016) 외에도 교육대학교 학생들(김혜연, 2015; 정혜승·서수현, 2012)이나 이공계열 학생들(최윤희, 2017)을 대상으로 피험자를 한정하여 논의한 연구들이 있다. 이런 연구들은 인지적, 정서적 측면에서 쓰기 인식을 탐구하여 학생들의 글쓰기에 대한 이해와 요구 등을 고찰했다는 점에서 의의가 있다. 그러나 이상의 연구들은 선다형의 폐쇄형 설문 형태로 이루어졌기 때문에 응답의 제한으로 인해 쓰기에 대한 다양하고 심층적인 인식은 살펴보기 어려웠다.

최근 들어 학생들의 쓰기 인식을 개방형 설문의 형태로 분석하려는 시도가 이루어졌다(변경가, 2017; 서수현, 2016; 장성민, 2016). 전술한 연구의 대부분은 은유 분석의 형태를 활용하여 쓰기에 대해 평소 지니고 있던 생각을 자유롭게 기술할 수 있도록 했다. 이러한 질적 연구들이 은유 분석에 주목한 이유는, 은유가 복잡한 개념에 대해 간결성을 제공하고 낯선 대상을 친숙하게 표현할 수 있는 인지 전략이나 학습이나 경험을 통해 형성된 추상적이고 복잡한 개념을 이해하고 설명하는 강력한 수단이 되기 때문이다(정혜승, 2005). 그러나 이 경우, 분석의 어려움으로 응답 대상자를 제한하여 그 결과를 일반화하는데 한계가 있었고 연구자에 따라 달라지는 범주 설정에 대해 객관성 확보의 문제가 지속적으로 제기되었다. 또한 연구자가 설문을 분석할 때, 우선 은유 표현을 판별해서 범주화한 후 범주에 따라 이유 기술문을 분석하는, 뒤바뀐 절차를 따르고 있어 인식의 정확한 도출이 어렵다는 단점이 있었다.²⁾ 일반적으로 학생들의 은유 설문은 학습자의 응답이 자신이 지

2) 기존의 분석 절차를 따를 때의 문제점 중 하나는 은유 표현이 동일하다면 모두 동일한 이유 범주로 구분한다는 점이다. 하지만 실제 학생들의 응답을 분석해 보면 은유 표현은 동일하나 다른 범주를 이유로 제시한 경우가 다수 발견된다. 이는 은유의 의미란 한 사물에서 연상된 의미가 다른 사물에서 연상된 의미와 관련하여 일부 속성은 부각시키고 나머지 속성은 감춰짐으로 생겨난다는 이성만(2000)의 견해와 관련지어 생각해 볼 수 있다. 즉, 학생들은 글쓰기를 동일한 사물로 표현하였더라도 그 사물이 지니고 있는 속성 중 각기 다른 부분을 부각시켜 응답한다는 것이다. 이는 은유 표현 자체의 범주화가 아닌 이유 기술문의 범주화가 필요함을 역설하는 예라 하겠다. 이에 대해서는 뒷부분에서 자세히 논할 것이다.

니고 있는 쓰기에 대한 인식을 은유로 표상화 한 후 그렇게 표현한 이유를 기술하는 일련의 방식을 따르도록 설계되어 있다. 그렇기 때문에 은유를 분석하는 연구자는 이 과정을 거꾸로 되짚어가야 학습자가 지닌 쓰기 인식에 대해 정확히 추정할 수 있는 것이다.

이 연구에서는 기존 연구가 지닌 한계점을 보완하기 위해 은유 분석에서 텍스트 마이닝의 방법 중 군집화를 활용한 자동 범주화를 제안하려고 한다. 이 연구는 먼저 텍스트 마이닝의 방법으로 이유 기술문에서 핵심 키워드를 추출하여 군집화하고, 범주화된 인식에 따라 각기 은유 표현이 어떻게 드러나는지를 살핀다는 점에서 선행 연구들과 차별성이 있다. 본고에서 제시한 분석 과정이 논리적 정합성을 얻는다면, 은유 분석을 활용한 인식 연구에 새로운 방법론을 제공할 수 있을 것이다. 이 연구에서는 다음과 같이 연구 문제를 설정하였다.

첫째, 쓰기에 대한 중·고등학생의 이유 기술문은 군집화에 의해 어떻게 범주화되며 그에 따른 은유 표현은 무엇인가?

둘째, 각 범주는 쓰기에 대한 중·고등학생의 인식을 어떻게 설명하는가?

셋째, 집단에 따라 쓰기 인식의 차이점이 존재하는가?

II. 이론적 배경

1. 은유 분석(Metaphor Analysis)

전통적으로 은유는 ‘A는 B이다’ 같은 구조를 지닌 간접적 언어 행위로 설명된다(김용직, 1992; 이수정, 2011). 은유의 어원인 그리스어 metapherein은 ‘너머’의 의미인 meta와 ‘가져가다’라는 의미의 pherein의 합성어이다(권

낙원, 2008). 이는 특정 어휘에 포함된 의미가 유사성이 없는 다른 어휘로 이동하면서 다른 어휘가 앞선 어휘의 의미를 가지는 일종의 ‘의미론적 전이’로 이해할 수 있다. 이때의 전이는 동일한 의미의 재진술이 아닌, 두 어휘가 독립성을 유지하면서 각각 의미의 일부가 상호 작용함으로써 얻게 되는 새로운 의미이다(남민우, 2011).

1980년대 이후 대두된 인지언어학 분야에서는 은유를 인지와 행동을 구조적으로 연결하는 도구로 간주한다. 이러한 인지적 관점은 은유를 단순한 언어 자체의 문제가 아니라, 사고와 인지의 문제로 접근한다(Lakoff & Johnson, 1980). 은유는 개념의 속성이자 대상에 대해 사고하고 그 대상을 개념화하는 방식인 것이다. 어떤 대상을 지각했다는 것은 ‘무엇으로’ 지각했는다는 의미인데, 이는 관습적인 사고 체계에 의한 인식으로 여겨지는데, 한 영역의 경험을 다른 영역의 경험으로 투사하여 지각함을 의미한다.

은유 분석은 인식의 기제로서의 은유를 강조하는 Lakoff & Johnson(1980)의 개념 은유 이론(Conceptual metaphor theory)을 근간으로 한다. 은유 분석은 근원영역(source domain)과 목표영역(target domain) 간의 유사성을 분석하여 은유 표현이 함축하고 있는 의미를 이해함으로써, 그 표현을 창출한 개인의 인지 구조를 파악하는 탐구 방법이다. Lakoff & Johnson은 낯설고 추상적인 개념(목표 영역)을 익숙하고 구체적인 개념(근원 영역)에 투사하여 표현하고 이해하는 것을 인간의 기본 능력으로 간주하면서, 은유를 인간의 경험에 대한 해석과 의미 부여를 가능케 하는 지각의 틀로 이해했다.

연구 방법으로서 은유 분석이란 연구 참여자들의 언어적 은유(linguistic metaphor)를 수집하고 범주화하여 특정 주제에 대한 연구 참여자들의 신념이나 사고 유형을 이해하는 데 통찰력을 제공하는 분석적 접근이다(Armstrong et al., 2011; Cameron & Low, 1999). 이는 인간의 학습, 경험을 통해 형성된 개념 체계와 이해 체계를 해석하고 설명하는 도구가 될 수 있다는 점에서 그 중요성을 인정받고 있다. 최근 들어 국내에서는 은유 분석이

교사 교육 분야 및 여러 교과에서 교수자들과 학생들의 신념과 태도, 인식을 조사하는 방법으로 활용되고 있다(김진호·김상미, 2014; 박혜영, 2016; 연희정·김은주·송주은, 2015; 유솔아, 2011; 윤초희, 2013; 이경언, 2015; 이은영·박선영, 2014; 장성민, 2015).

은유 분석은 신뢰성과 객관성의 확보가 필수적이다. 이를 위해 많은 연구자들은 조금씩 상이한 은유 분석의 방법론을 제시해 왔다. 특히 교육 연구 분야에서는 Schmitt(2005)와 Andriessen(2006), Saban(2006, 2010), Gök & Erdorğam(2010), Armstrong et al.(2011)의 방법론이 살펴볼 만하다. 각 연구자들이 제시한 분석 절차 및 단계를 정리하면 <표 1>과 같다.

<표 1> 연구자에 따른 은유 분석 절차 및 단계

연구자	절차 및 단계
Schmitt, Andriessen	① 은유 분석 대상 선정 ② 연구 목적 설정 및 표집방법 선택 ③ 준거 설정 후 은유적 표현 선별 ④ 범주화 ⑤ 동일 범주 내 은유적 표현들의 특성 파악 및 범주명 명명 ⑥ 통계 분석
Saban	① 코딩 및 제거 ② 대표 은유 선정 및 전사 ③ 분류 및 범주화 ④ 내부 신뢰도 평정 ⑤ 양적 분석
Gök & Erdorğam	① 은유 판별 ② 은유 범주화 ③ 범주 개발 ④ 타당성 및 신뢰성 확보 ⑤ 양적 통계 분석
Armstrong et al.	① 은유 표현 수집 ② 목표영역과 근원영역 파악 ③ 근원영역의 특성 파악 ④ 근원영역과 인식 대상의 관련성 분석 ⑤ 은유 유형 추론 ⑥ 범주명 명명 ⑦ 숨겨진 특징 확인 ⑧ 주제 확인

일반적으로 은유 분석을 위한 절차에는 일정한 순서가 존재한다. 그 중 주목할 부분은 대부분의 은유 분석 과정에서 실질적인 분석을 실행하기 전에 연구자의 유형범주화가 선행된다는 점이다. 이연선(2015)의 지적처럼, 이 단계에서 주관적 해석이 개입할 가능성이 존재하며, 바로 이 지점이 은유 분석에서 가장 주의해야 할 대목 중 하나이다. 연구 목적이나 연구자의 주관에 따라 분석 절차는 다소 차이가 날 수 있으나 유의한 연구 결과를 얻기 위해서는 절차의 타당성에 대한 비판적 검증이 필요하다.

2. 텍스트 분석의 주요 기법

(1) 군집화(Clustering)

군집화는 문서 내에 존재하는 텍스트 간의 유사성을 기반으로 문서를 분할하여 문서가 지니고 있는 내적 정보를 자동으로 발견하는 비지도학습(unsupervised learning) 기법이다(Andrews & Fox, 2007). 군집화는 문서 분류 정보를 통해 학습이 이루어지는 문서 분류(document classification)와 달리, 사전정보 없이 유사한 특징을 갖고 있는 문서들을 그룹화 해주기 때문에 분석에 있어 연구자의 주관성을 축소할 수 있다. 또한, 많은 양의 문서를 동질 집단으로 군집화하기 때문에 정보손실을 최소화하는 자료의 단순화 과정이라고 할 수 있다. 문서는 그 자체만으로는 군집화 분석이 어렵기 때문에 일반적으로 문서로부터 단어를 추출한 후, 문서 사이의 유사도를 계산해 문서-단어행렬(document-term matrix)을 만들고, 이를 활용하여 군집화를 수행한다.

군집분석은 계층적(hierarchical) 군집분석과 비계층적(nonhierarchical) 군집분석으로 나눌 수 있다. 계층적 군집분석은 각각의 개체를 하나의 군집으로 보고 이를 순차적으로 병합하거나, 전체 개체를 하나로 간주하고 분할해나가는 방법이다. 이에 반해, 비계층적 군집분석은 미리 군집 수 k 를 규정한 뒤 k 개의 군집에서 시작하여 다른 군집의 중심(centroid)과 모든 개

체 사이의 유클리드 거리(euclidean distance)를 계산한 후 가장 가까운 중심을 갖는 군집에 개체들을 할당하여 최종적인 군집을 형성하는 방법이다. 이 방법에는 중심을 정하는 것에 따라 K-평균법(K-means method)과 K-대표개체법(K-medoids method), K-중위수법(K-median method) 등의 알고리즘이 있다(정민지, 2017). 그 중 가장 일반적인 것은 K-평균 알고리즘인데, n 개의 입력 데이터를 k 개의 군집으로 분할하는 방법이다(Rasmussen, 1992). k 개를 선택하는 방법에는 임의의 선택, 데이터 집합 순서로 선택, 거리 최적화 방법 등이 있다. 이 중 거리 최적화 방법은 초기값을 구하는 시간이 오래 걸리는 단점이 있지만 항상 동일한 결과를 보이고 군집화 성능 면에서도 좋은 결과를 보인다는 장점이 있어 가장 통용되는 방법이다.

(2) 감정 분석

감정이란 내적 또는 외적 현상에 대해 일어나는 직접적 반응으로, 의식의 가담 없이 발현되는 정신적 상태로 정의된다(Rohracher, 1976). 감정은 기분을 느끼는 주체가 어떠한 자극을 인지함으로 발생하며 이러한 인지는 주체에 의해 표현된다. 그 중 언어는 감정을 가장 정교한 방식으로 표현하고 전달할 수 있는 수단인바, 감정을 표현하는 언어의 형태를 감정 단어라 한다(홍종선·정연주, 2009).

한국어 감정 단어 목록에 관한 연구는 다수 존재하는데(손선주 외, 2012; 이준웅 외, 2008) 대부분 전처리 과정을 거쳐 추출한 핵심어, 강조어, 수식어 등을 사용하여 감정 사전을 구축하는 일련의 과정을 거친다. 상술했다면, 문서에서 추출한 어휘는 감정 신뢰도에 따라 극성을 판별한 뒤, 감정 사전을 구축하게 된다. 감정 신뢰도란 어휘의 극성(긍정 및 부정) 판단에 신뢰성을 주는 값으로, 신뢰도가 p 이상이면 완전 극성으로 볼 수 있다. 반면 신뢰도가 q 이하이면 해당 어휘만으로는 감정어의 극성이 판단이 불가능한 경우를 나타낸다. 이때 p 와 q 값은 해당 문서에 따라 달라진다. 어휘의 신뢰도 값이 p 이상이면 어휘의 긍정도와 부정도의 비교만으로도 감정어의 극성 판별이 가능

하다. 하지만 신뢰도가 q 이하의 값을 갖게 되면 감정어의 극성 판별이 무의미하기 때문에 해당 어휘는 감정어 사전에 포함시키지 않는다. 신뢰도 값이 q 이상이면 p 이하인 경우에는 문장에 따라 긍정어 또는 부정어가 될 수 있다.³⁾ 따라서 이러한 어휘는 해당 어휘의 연관어와 강조어에 따라 감정어의 극성을 결정하게 된다(이슬기, 2017; 최지은 외, 2016). 이렇게 구축된 사전을 통해 문서의 감정 분석이 가능한데, 감정 사전에 포함된 감정어 목록 뿐 아니라 해당 어휘의 가중치를 가져와 계산이 이루어진다.

III. 연구 방법 및 절차

1. 연구 대상 및 분석 도구

이 연구에서는 중·고등학생들의 쓰기 인식을 살펴보기 위해 중학교 3학년 학생들과 고등학교 2학년 학생들을 연구 대상으로 선정하였다. 중학생의 경우, 은유에 대한 이론뿐만 아니라 다양한 표현을 구사할 수 있는 능력까지 갖추어야 했으므로 3학년 학생을 대상으로 삼았고, 고등학생은 중학생과 발달의 차이가 존재하면서 작문 수업 시간을 활용할 수 있는 2학년을 대상으로 했다. 설문 작성 후 자료 정제 과정에서 은유적 표현이 아닌 응답들을 분석에서 제외하고 나니 집단 간 표본 수에 다소 차이가 났다. <표 2>는 최종 분석이 이루어진 학생들의 정보를 정리한 것이다.

3) 단어의 신뢰도만으로 극성의 판별이 어려울 때, 'Combination 함수'를 이용한다. 이 함수는 해당 어휘를 입력 값으로 input하여 기 구축된 단어와 연관어의 조합을 통해 극성을 판별한 뒤 결과를 output하는 기능을 한다.

〈표 2〉 연구 대상자

	고등학교	중학교	합계
남학생	561	523	1,084
여학생	799	482	1,281
계	1,360	1,005	2,365

학생들의 응답 내용을 분석하기 위해 이 연구에서는 엑소브레인 한국어 언어 분석 툴킷 Ver 2.0을 사용하여 형태소 분석을 한 후, json으로 변환하여 사용하였다. 컴퓨터 프로그램 언어는 json의 파싱이 가능한 Java를 사용하여 어휘 추출과 기계학습을 실행하였다. 또한, 분석 항목의 클러스터링(clustering) 결과와 은유 표현을 워드클라우드(wordcloud)로 제시하기 위해 시각화에 용이한 컴퓨터 프로그램 언어 R을 사용하였다. 자동으로 클러스터링된 자료들은 군집화가 적절히 이루어졌는지 검증할 필요가 있었다. 이를 위해 형태소 분석기인 AntConc Ver. 3.5.0을 사용하여 핵심 키워드를 중심으로 문장의 내용을 살펴보았다.

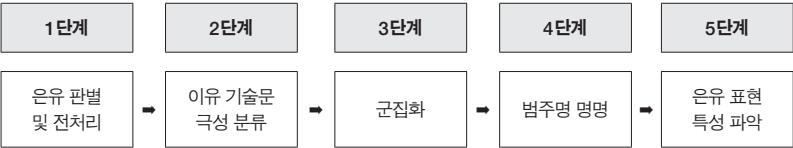
2. 분석 절차 및 방법

학생들의 쓰기 인식을 알아보기 위해 학생들의 동의를 얻은 후 설문 조사를 실시하였다. 설문 문항은 개방형으로 “나에게 글쓰기는 ~이다.”와 같은 은유 표현과 “왜냐하면 ~.”의 이유 기술문을 함께 제시하여 은유 표현에 대한 이유를 설명하도록 하였다. 학생들이 작성한 설문의 응답은 분석을 위해 은유 표현과 이유를 각각 항목으로 나누어 엑셀로 입력하였다. 잘 정제된 데이터일수록 형태소 분석의 성능을 높일 수 있다. 이 연구의 목적은 학생들의 쓰기 행위 자체보다 인식에 대한 탐구가 우선이었으므로, 띄어쓰기와 표기 오류는 최대한 수정하여 입력하였다.

입력한 자료의 텍스트 분석을 위해 전처리를 실시하였다. 필요한 어휘

의 태깅을 위해 Visual Studio를 사용하여 코드를 편집한 후, Java와 연동 가능한 프로그램 언어 Scala를 사용하여 어휘 추출을 실행하였다. 이 연구에서는 기존 연구들과 달리 이유 기술문을 우선적으로 분석하였다. <표 3>은 이 연구에서 수행한 은유 분석의 절차이다.

<표 3> 분석 절차



학생들이 제시한 이유 기술문은 쓰기를 정서적 측면에서 바라보는 응답과 쓰기의 인지적 측면을 강조한 응답으로 대별 되었다. 정서적 관점에서 인식한 기술문에는 쓰기에 대한 긍정적 반응과 부정적 반응이 드러난 반면, 인지 관점의 기술문은 설명적 성격이 강했다. 따라서 정서적 인식을 보인 기술문을 극성(긍정, 부정)으로 인지적 관점을 견지하며 설명적 서술이 되어있는 기술문을 무극성으로 나누는 감정 분석을 실시하였다. 긍정 어휘는 선행 연구의 감정 어휘(김은영, 2005, 2004; 박인조·민경환, 2005; 안신호 외, 1993)에서 중복 출현한 어휘를 선정하였고, 추가로 세종 형태분석 말뭉치에서 고빈도 감정 어휘를 추출한 후, 표준 국어 대사전의 유의어를 참조하여 확정하였다.⁴⁾ 부정 어휘 역시 동일한 방법으로 추출한 후, 이미 선정된 긍정 어휘에 ‘~지 않다’의 부정 표현을 추가⁵⁾하여 목록을 확정하였다. 부정 표현에 ‘~

4) 본고에서 사용한 어휘 목록은 ‘감정 단어’ 보다 ‘감정 어구’가 더욱 적확하다. 이는 말뭉치 언어학(corpus linguistics)의 기본 전제인, 감정 어휘의 범주가 ‘구’로 확장될 필요가 있다는 의견과 일맥상통 한다. 그간 논의되어 온 한국어의 의사소통 단위는 ‘의미 단위(semantic unit)’이며, 의미란 상당 부분이 ‘구’의 형태로 교환되기 때문이다(남길임, 2016). 본고에서는 아직까지 ‘어휘’의 의미가 통용되는 범주를 고려하여 의미 구분을 다소 범박하게 제시하였다.

5) 예를 들면, 기 선정된 긍정 어휘인 ‘기쁘다’, ‘즐겁다’의 경우, 대응쌍 ‘기쁘지 않다’, ‘즐겁

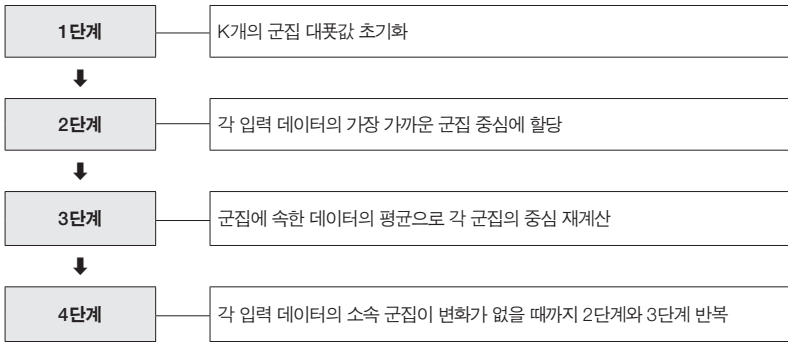
지 않다'만을 추가하여 작업을 실시한 이유는 긴 부정은 합성어나 파생어의 공기 제약에서 짧은 부정에 비해 비교적 자유롭기 때문이다(고영근·구본관, 2011). 또한 감정용언의 다수가 형용사로 되어있어 '못' 부정은 상대적으로 기능 부담량이 적기 때문에 이에 대한 추가 작업은 별도로 실시하지 않았다. 먼저 학습 데이터를 500개 설정하여 긍정에 1, 부정에 -1, 무극성에 0 값을 부여한 후, 미리 설정한 테스트 데이터 100개에 파일럿 테스트를 실시하였다. 그 결과, 예측률 0.848로 다소 부정확한 결과값을 얻게 되었다. 왜냐하면 극성 값이 강하나 긍정문과 부정문에 동시에 사용되는 어휘가 있어 무극성 어휘와 혼돈이 발생했기 때문이다. 이에 설계를 수정하여 극성에 1, 무극성에 -1 값을 부여하여 극성과 무극성을 먼저 분류한 후, 극성 문서만을 모아 긍정에 1, 부정에 -1의 값으로 재분류하는 2단계 방법을 사용하였다. 역시 100개의 자료로 준거 타당도(criterion-related validity)를 테스트한 결과 0.942로 성능이 더 좋아짐을 확인할 수 있었다.⁶⁾

무극성 문서들은 문서에서 추출한 키워드를 기반으로 군집분석을 실시하였다. 군집분석에서 유사도 행렬을 계산하기 위해 컴퓨터 프로그램 R의 dist 함수를 사용하여 유클리드 거리를 계산하였다. 사용한 알고리즘은 SCS(Simple Cluster Seeking)⁷⁾ 방법을 이용하여 거리를 최적화하는 K-평균 알고리즘을 사용했다.

지 않다'는 부정 어휘로 선정하였다.

- 6) 1단계 결과값은 0.954이고 2단계 결과값은 0.987로 긍정-부정의 구분에 더 정확한 예측력을 보였다.
- 7) SCS 알고리즘을 간단히 설명하면 다음과 같다. 1단계: 첫 군집의 중심을 처음 입력 데이터로 초기화($C_1 = x_1$) 한다. 2단계: $j=2, \dots, N$ 에 대해, 모든 군집의 중심 C_k 에 대하여 $\|x_j - C_k\|$ 이 p 이면 x_j 를 새로운 군집의 중심으로 설정한다(p : 임계치). 이때, K 개 군집 중심이 초기화되면 종료한다. 3단계: 모든 입력 집합에 대한 계산 후 K 보다 적은 군집의 중심이 생성되면, p 를 감소하고 1단계와 2단계를 반복한다(지태창 외, 2007).

〈표 4〉 K-평균 알고리즘

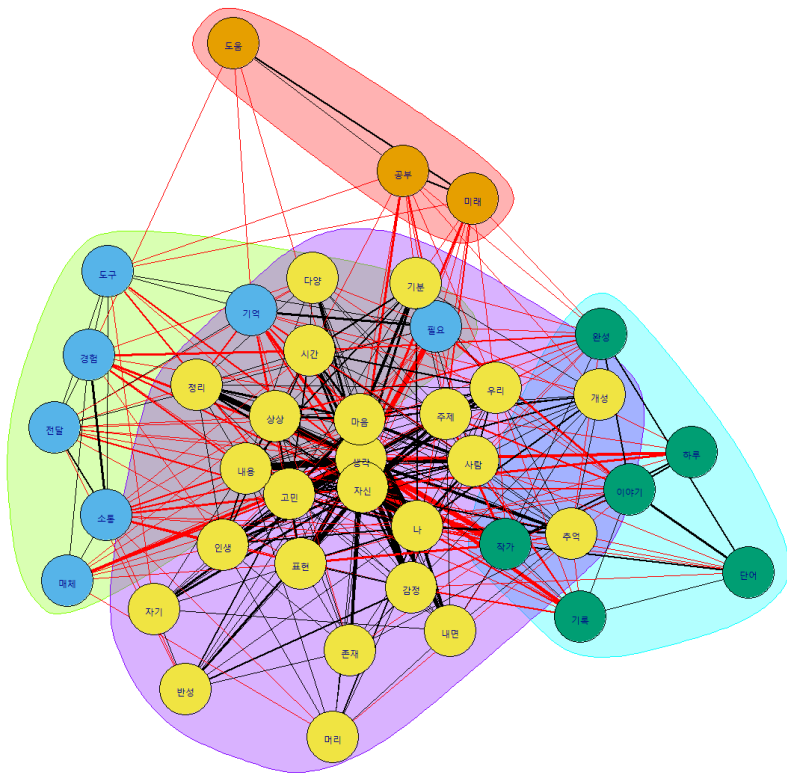


〈표 4〉는 K-평균 알고리즘의 수행 방법이다. 군집 분석의 결과를 직관적으로 살피기 위해 후속 작업으로 igraph의 walktrap 알고리즘을 사용하였다.⁸⁾ 거리 최적화 방법으로 군집화한 결과 〈그림 1〉과 같이 4개의 집단으로 자동 분류됨을 확인할 수 있었다.

문서에 포함된 키워드 중에는 두 가지 군집에 동시에 포함되는 어휘들이 존재했다. 이는 은유분석 자체가 지닌 복합적이고 중층적인 성격 때문이기도 하며, 앞서 실시한 감정 분석에서 극성성이 약한 어휘가 존재하는 것과 유사한 이유 때문이기도 하다. 집단의 이름을 명명하기 위해서는 다른 범주와 중첩되지 않으면서 집단의 특징을 잘 드러내는 단어들을 중심으로 작업을 진행할 필요가 있었다.

첫 번째 군집은 ‘자기’, ‘반성’, ‘표현’, ‘나’, ‘감정’, ‘내면’, ‘머리’, ‘생각’의 키워드를 통해 ‘자기 지향적 사고’로 명명하였다. 두 번째 군집은 ‘소통’, ‘전달’, ‘매체’, ‘경험’의 키워드가 포함되어 있어 ‘타자 지향적 사고’로, 세 번째 군집은 ‘도움’, ‘공부’, ‘미래’라는 키워드를 근간으로 하여 ‘도구적 사고’라고

8) 군집 추출은 igraph에서 제공하는 walktrap, edge betweenness, multi-level 알고리즘을 주로 이용한다. 대체로 link 기반인 edge betweenness는 느리고 단점이 많은 반면, flow 기반인 walktrap, multi-level은 품질이 높아 최근 다양한 연구에서 사용되고 있다 (Blondel et al., 2008).



〈그림 1〉 이유 서술문의 무극성 군집 분석 결과

명명하였다. 마지막 군집은 ‘완성’, ‘단어’, ‘기록’, ‘이야기’로 쓰기 행위 본질에 관한 어휘들이 중심을 이루고 있었으므로 ‘본질적 사고’로 이름을 붙였다. 키워드로 군집화한 분류가 실제 이유 기술문의 특징을 반영하는지에 대해 검증하기 위해 AntConc Ver. 3.5.0을 이용해 해당 키워드를 중심으로 추출해 보았다. 〈그림 2〉는 ‘타자 지향적 사고’ 범주의 이유 기술문에서 ‘전달’을 키워드로 하여 추출한 예시 자료이다. Concordance Hits는 ‘전달’이 사용된 문서의 개수이며, File 아래 숫자는 이유 기술문에서 해당 단어를 키워드로 사용한 학생들의 글 번호이다.

〈그림 2〉의 이유 기술문을 살펴보면, ‘전달’은 학생들이 타인과의 관계

Concordance Hits 96			File
Hit	KWIC		
79	유일한 수단 중 하나인 것 같다. 다른 사람에게 내 생각과 마음을 전달할 수 있고, 인터넷 사용이 발달된 현대에 글쓰기는 더 와닿는		_0136.txt
80	렵기 때문이다. 그리고 자신의 생각을 글로 표현하여 말하기 힘든 것을 전달할 수도 있다. 또 글은 수업 시간에 많이 쓰여 잘 쓰면 칭찬받는다.		_0170.txt
81	기를 하면서 신선한 생각이 될 수도 있고 내 생각들을 간략하게 글쓰기로 전달할 수도 있는 것이라고 생각한다.		_0481.txt
82	글쓰기를 통해 속에 담아 두었던 말이나 생각을 전달할 수 있기 때문이다. 직접 얼굴을 마주하고 말하기 어려운 말들은		_0565.txt
83	수 있기 때문이다. 직접 얼굴을 마주하고 말하기 어려운 말들은 글로써 전달할 수 있다. 나 자신도 앞에서 못했던 말들을 편지로 써서 전달		_0565.txt
84	전달할 수 있다. 나 자신도 앞에서 하지 못했던 말들을 편지로 써서 전달할 수 있다. 어렵듯 글을 통해서 우리는 얼굴을 직접 마주하지 않아도		_0570.txt
85	가장 많이 보편화되어 있고 우리가 많이 듣는 노래도 가사에서 감정을 전달하는 것을 보면 글이 가장 소통하기 좋은 방법인 것 같다. 글을 쓰는		_0578.txt
86	고민하게 된다. 따라서 옛날에 새를 통해 사람들에게 알려고자 한 것은 전달하는 것이 나의 글쓰기와 유사한 것 같다.		_0579.txt
87	기도 하고 힘들지만 누군가가 나의 배구니, 나의 글쓰기를 누군가에게 전달할 때 내 배구니로 인해 감동을 받을 수 있다면 글쓰는 것이 나에게		_0636.txt
88	우리는 서로 이야기하거나 소식을 전달할 때 말을 사용한다. 하지만 나는 이야기를 귀로 듣거나 말로 전달할 때 말을 사용한다.		_0636.txt
89	움을 전달할 때 말을 사용한다. 하지만 나는 이야기를 귀로 듣거나 말로 전달할 때 말을 사용한다. 하지만 나는 이야기를 귀로 듣거나 말로 전달		_0636.txt
90	하는 것보다 문자화 시켜서 글로 전하는 것이 더 편하다. 글로 얘기하면 전달할 때 한번 더 생각할 수 있고 둘을 때도 내용을 정리하기 편하기		_0649.txt
91	내가 글을 쓰게되면 내 생각을 더 조리있게 표현할 수 있고 말로 전달하는 것보다 내 생각을 더 잘 전달할 수 있어서 글쓰기는 나에게 발		_0649.txt
92	조리있게 표현할 수 있고 말로 전달하는 것보다 내 생각을 더 잘 전달할 수 있어서 글쓰기는 나에게 발표이다.		_0661.txt
93	누군가가 자신의 이야기를 전달하기 위해 글을 쓰기도 하고 우리가 읽은 책이나 글들이 글쓰기로		_0755.txt
94	하지만 절로는 표현하기가 어렵다 글을 쓰는 것이 어떻게 나의 마음을 전달할 수 있다고 할지 고민하는 것처럼 많이 생각해야 하기 때문이다.		_0864.txt
95	하면서 글쓰기는 항상 함께한다 내가 표현하기 어려운 의라도 글쓰기를 전달할 수 있다고 인생에서 빠질 수 없는 것이다 또 사회 생활에서도 글		_0935.txt
96	은 사람들 앞에서 말하지 못한 것들을 솔직하게 표현하거나 사람들에게 전달할 수단이 되거나 사람들에게 내세워지 못한 나의 솔직한 심정을		

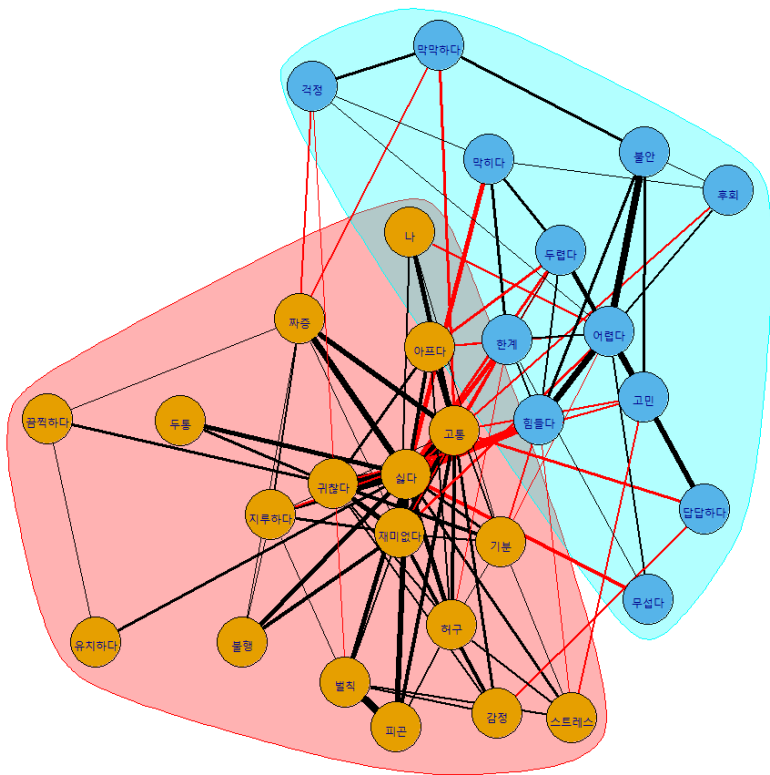
〈그림 2〉 학생 글 코퍼스 분석 결과

속에서 쓰기의 의미를 찾으려는 맥락에서 주로 사용되었음을 알 수 있다. 이 점은 쓰기를 ‘타자 지향적 사고’의 측면에서 인지하고 있는 것이라 해석 가능하다. 즉, 군집화된 어휘로부터 학생들의 쓰기 인식이 추론 가능함을 보여 주는 것이다.

긍정과 부정의 극성 문서도 집단의 세분화가 가능한지 알아보기 위해 군집 분석을 실시하였다. 극성 문서는 긍정과 부정이 대응을 이루고 있었기 때문에, 부정 문서를 기준으로 분석을 실시하여 명명하였다.

분석 결과 극성 문서는 두 개의 범주로 군집화 되었다. ‘싫다’, ‘귀찮다’, ‘피곤’, ‘지루하다’ 등의 키워드를 중심으로 한 범주를 ‘태도(부정)’로, ‘어렵다’, ‘막막하다’, ‘불안’ 등의 키워드를 중심으로 한 범주를 ‘효능감(부정)’으로 명명하였다. 긍정 문서 역시 두 개의 범주로 군집화 가능했는데 ‘태도(긍정)’에는 ‘즐겁다’, ‘재미있다’, ‘놀이’의 키워드가, ‘효능감(긍정)’에는 ‘칭찬’, ‘뿌듯하다’, ‘자신감’ 등의 키워드가 포함되었다.

군집화 결과에 따라 쓰기 인식의 범주를 인지 영역과 정서 영역으로 구분하고, 인지 영역에는 ‘자기 지향적 사고’, ‘타자 지향적 사고’, ‘도구적 사



〈그림 3〉 이유 서술문의 극성(부정) 군집 분석 결과

고, ‘본질적 사고’를, 정서 영역에는 ‘태도_긍정’, ‘태도_부정’, ‘효능감_긍정’, ‘효능감_부정’⁹⁾을 항목으로 설정한 후, 교과 교육 전문가 2인에게 타당도를 검증받았다. 이렇게 군집화된 어휘를 바탕으로 500개의 학습 데이터와 100개의 테스트 데이터 문서를 분류한 후, 단어-문서 행렬에 가중치를 적용하여 범주별 사전을 재구축하였다. 완성된 사전을 바탕으로 2,365개의 이유 서

9) 효능감은 ‘높다/낮다’의 정도로 표현하는 것이 바람직하다. 다만, 본고에서는 태도와의 서술상의 통일성을 고려하여 높은 효능감을 ‘효능감-긍정’으로 낮은 효능감은 ‘효능감-부정’으로 범주화하였다.

술문을 분류하였다.

IV. 연구 결과 및 논의

1. 쓰기에 대한 학생들의 인식 범주와 은유 양상

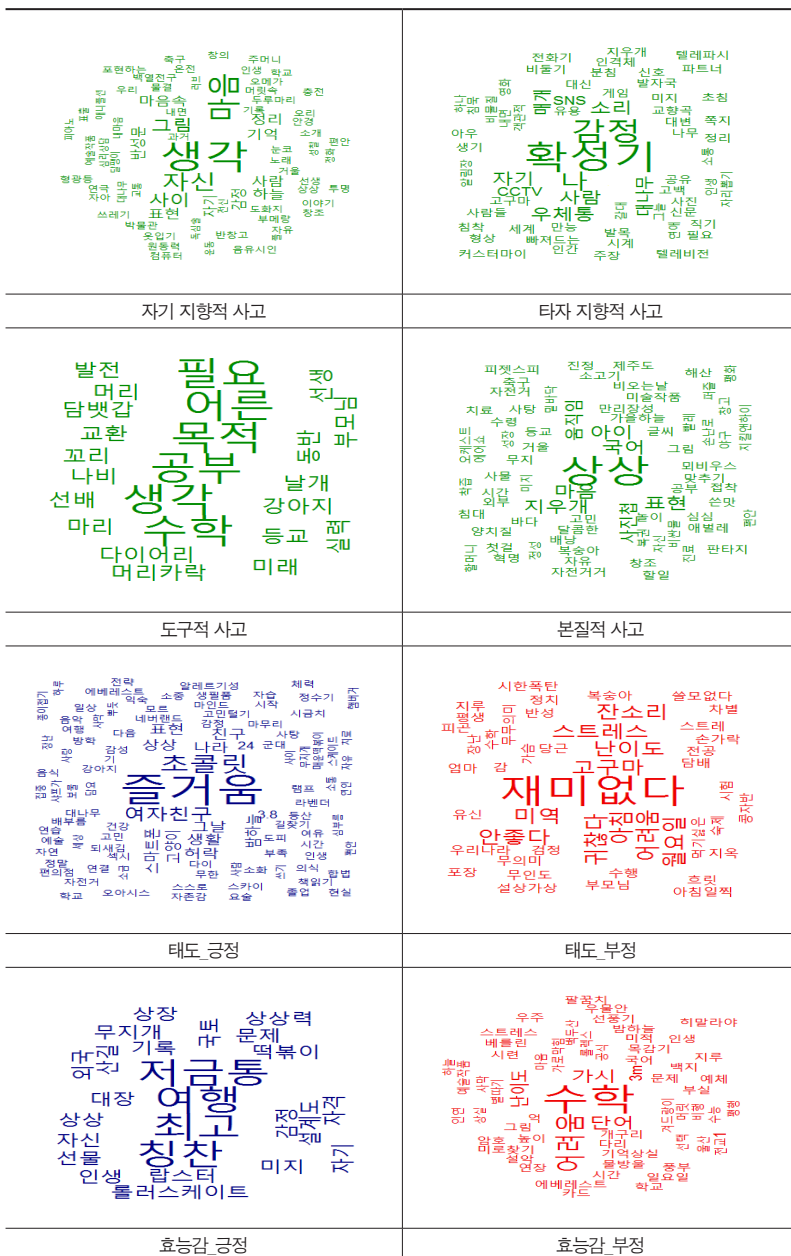
군집화 방법에 따라 분류된 중·고등학생의 쓰기에 대한 인식을 정리하면 <표 5>와 같다. 설문 결과 쓰기를 행위 자체의 본질적 측면에서 인식하고 있는 학생들이 가장 높은 비율로 분포한 반면, 쓰기에 대해 긍정적 효능감을 중심으로 인식하고 있는 학생들은 가장 적은 비율이었다.

<표 5> 쓰기 인식에 대한 범주별 빈도

인지				정서			
1,392명(58.86%)				973명(41.14%)			
자기 지향적 사고	타자 지향적 사고	도구적 사고	본질적 사고	태도_긍정	태도_부정	효능감_긍정	효능감_부정
387명 (16.37%)	165명 (6.98%)	133명 (5.62%)	707명 (29.89%)	308명 (13.02%)	313명 (13.24%)	84명 (3.55%)	268명 (11.33%)

학생들이 쓰기에 대한 인식은 범주에 따라 각기 다른 은유 표현들로 표상되었다. <표 6>은 학생들이 활용한 은유적 표현을 범주별로 워드 클라우드 형태로 제시한 것이다. 워드 클라우드는 도출된 어휘들을 빈도에 따라 크기를 다르게 표현하는 태그 클라우드 기법으로 문서의 주요 정보를 직관적으로 파악할 수 있게 한다는 장점이 있다.

〈표 6〉 범주별 은유 양상



범주에 따라 사용된 은유 표현은 비교적 큰 차이를 보였다. 특히 대표성을 지닌 몇몇 은유 표현들은 목표영역과의 거리가 밀접하여 이유 기술문을 따로 분석하지 않더라도 용이한 해석이 가능했다. 반면, 다른 범주에 있으나 동일한 은유 표현으로 제시된 응답들이 있어 해석에 주의를 요한다.

〈1167〉 레고 블록: 레고 블록을 선택해 쌓는 모양에 따라 상대방에게 다르게 보이게 만들 수 있다. 문장을 쓰더라도 어떤 단어를 선택하여 어떤 구조로 쓰느냐에 따라 상대방에게 전달할 내용과 뉘앙스가 천차만별 달라진다.

〈1540〉 레고 블록: 레고 블록은 한 개 한 개가 모여 큰 작품이 되듯이 글쓰기도 한 문장, 한 문장 정성스럽게 쌓여 만들어지기 때문이다. 글쓰기를 하면서 내용을 고민하고 맞춤법을 생각하면서 작성하듯 레고 블록도 설명서를 보고 고민하면서 맞추기 때문에 글쓰기와 같다. 완성된 글은 완성된 레고이다. 완성된 레고를 보기 위해서는 열심히 생각하는 과정이 필요한데 글쓰기도 마찬가지라고 생각한다.

두 학생은 글쓰기를 ‘레고 블록’이라는 동일한 은유 표현으로 사용하였다. 그러나 두 학생이 인식한 ‘글쓰기’와 ‘레고 블록’ 사이의 유사성에는 간극이 존재했다. 〈1167〉번 학생이 ‘타자 지향적 사고’의 일환으로 유사성을 인식한데 반해 〈1540〉번 학생은 쓰기 행위 자체, 즉 ‘본질적 사고’의 범주 내에서 목표 영역과 근원 영역 사이의 유사성을 찾았기 때문이다. 이러한 사례는 ‘레고 블록’ 외에도 ‘수학’, ‘생각’, ‘여행’ 등의 표현에서도 발견할 수 있었다. 은유가 가치판단으로 기능하고 은유 분석을 통해 그 표현을 해석 가능한 은유로 수용하고자 한다면 은유 표현 자체의 범주화가 아닌, 이유 서술문 분석이 선행되어야 함을 지적할 수 있는 대목인 것이다.

(1) 자기 지향적 사고

학생들의 일부(16.37%)는 쓰기를 자신을 되돌아보고 자신의 목소리를 드러내는 수단으로 인식하고 있었다. 자아 정체성 확립의 장면에서 쓰기는 내면을 정돈하는 도구로 작용하며 정체성을 창조하는 행위로 간주 된다. 이를 통해 학생들은 구체화된 성찰의 기회를 갖게 되고 주체의 역할을 강조하게 되는 것이다.

〈236〉 마음의 소리: 나의 마음속에서 들리는 소리를 내가 조합해서 창조해내는 일이기 때문이다. 내가 무엇을 원하는지는 내면세계에 있다. 글쓰기는 내 속에서 울리는 자아의 소리를 받아 적는 것이다.

〈1868〉 반성문: 글을 쓰면서 나를 돌아보게 만드는 자아 성찰의 기회를 가질 수 있기 때문이다. 반성문을 쓸 때 내 잘못을 돌아보는 것처럼 글을 쓰면서 후회되는 점을 반성하고 더 나아질 방도를 찾아 나의 마음이 성숙해질 수 있다.

〈1718〉 실 풀기: 나의 머릿속에 있던 이야기들을 하나로 정리해 적는 것이기 때문이다. 정리한다는 것은 실 풀기가 않을까 싶다.

쓰기를 통한 자기 지향적 사고는 개인적인 자기반성에서부터 인지적 측면에 해당하는 생각의 심화까지 폭넓은 의미로 사용되었다. 이 범주에서 사용된 표현의 의미는 성찰의 외연적 확장이라는 점에서, ‘타자 지향적 사고’와 대별 된다. 타자 지향적 사고는 특정한 대상이 존재하고 뚜렷한 목적과 그에 부합한 전형적 형식을 지니고 있다면, 이 범주의 표현은 상대가 부재하며, 지극히 개인적이기 때문이다. Britton과 그의 동료들(1975)의 견해처럼 소통이 양식(form)에 속하는 것이라면 자기표현은 방식(mode)에 근접하다고 볼 수 있다.

〈628〉 하늘하늘한 웃: 내 마음과 감정 생각을 예쁘고 아름답게 표현할 수 있기 때문이다. 글쓰기는 하늘거리는 웃처럼 틀에 박히지 않고 자유롭게 내 생각을

적을 수 있기 때문이다.

〈2235〉 목소리: 눈물이 날 때 습관처럼 글을 썼다. 어쩔 수 없이 화를 뱃듯 글을 썼다. 그래서 내가 써 내려간 모든 텍스트는 가장 비밀스러운 것들이다. 글쓰기란 내가 가진 가장 솔직한 목소리다.

(2) 타자 지향적 사고

쓰기를 타자 지향적 사고의 관점으로 바라본다는 것은 쓰기를 개념적 측면에서 인식하고 있다는 것을 의미한다. 타자 지향적 사고의 범주에서 쓰기를 인지한 학생들(6.98%)은 전달의 도구를 근원영역으로 하여 쓰기와 연관 지으려 했다. 타자 지향적 사고 영역에서는 소통이 지니고 있는 상호작용과 전달의 특성이 강해서인지 ‘우체통’, ‘확성기’, ‘SNS’ 같은 다소 상투적인 은유 표현들이 주로 나타났고 이유 기술문에서도 유사한 내용이 반복 제시되었다. 타자 지향적 사고에는 뚜렷한 대상이 존재한다. 학생들은 대체로 타자를 주변인으로 한정했으나 범주를 일반화하여 제시하거나 사회적으로 확장하여 의미를 창조하려는 노력도 찾아볼 수 있었다. 쓰기 과정에서 어떤 대상을 상정한다는 것은 학생들이 예상 독자에 대한 인식을 내면화하고 있음을 보여주는 결과이다. 이러한 결과는 학생들이 이미 예상 독자에 관련한 인식 자원을 지니고 있으며, 쓰기에서 독자 고려의 중요성에 대해 파악하고 있다는 선행 논의의 결과(정혜승·서수현, 2009)를 뒷받침한다.

〈980〉 대화: 말하는 사람과 듣는 사람이 상호작용을 하는 대화처럼, 글쓰기는 작가와 독자가 서로 소통하는 대화이다.

〈1154〉 신문고: 신문고처럼 뒷사람에게 전하고 싶은 내용을 알릴 수도 있고 그런 방식으로 사회와 나라를 바꿀 수도 있기 때문이다.

(3) 도구적 사고

일부 학생들(5.62%)은 쓰기를 자신이 긍정적 상태로 나아갈 수 있도록

도움을 주는 행위로 인식하고 있었다. 쓰기를 개인의 발전을 위한 도구로 응답한 학생들은 쓰기의 효용적 측면을 강조하고 있었다. 이러한 범주로 응답한 학생들의 이유 기술문을 분석해 본 결과, 학생들은 크게 두 측면의 발전을 상정하고 있었다. 한 측면은 쓰기가 자신의 생각을 심화·확대한다는 인지적 성장에 초점을 두고 있는 응답이었고, 다른 측면은 쓰기를 수단으로 하여 학업 성취나 성공적인 취업 같은 가시적 성과를 이루기 위한 능력 향상에 초점을 둔 것이었다. 능력 향상을 강조한 응답은 쓰기의 기대 효과에 대한 선행 연구에서 ‘직업을 얻기 위해서’와 ‘좋은 성적을 얻기 위해서’라는 응답 비율이 높았던 결과와도 유사한 측면이 있다(Shell et al., 1995).

〈589〉 성장: 글을 씀으로써 생각의 폭을 넓혀나갈 수 있을 것이라고 생각된다. 나의 생각을 정리하며 나를 더욱 발전시킬 수 있을 것 같다. 이로 인해 결과적으로 글쓰기는 나를 성장하도록 만들어주는 원동력이 될 수 있을 것이다.

〈1079〉 자소서 연습: 나중에 대학교나 취업을 위해 자기소개서를 쓸 때 글쓰기를 하면서 연습했던 내용을 이용하면 더 잘 쓸 수 있을 것이라고 생각한다. 글쓰기를 연습하면 내가 원하는 곳을 들어가게 될 것이다.

(4) 본질적 사고

학생 응답의 가장 많은 비율(29.89%)을 차지한 본질적 사고라는 범주는 다른 범주에 비해 특징이 가시적이지 않고, 다소 범박한 측면이 있다. 본질적 사고로 분류된 문서들의 특징을 살피기 위해 해당 학생의 글들을 살펴 보았다.

〈738〉 목재: 글쓰기는 목재처럼 하나의, 또는 여러 가지의 주제를 변화시켜내는 것이기 때문이다. 주제 자체는 다듬어지기 전의 목재에 불과하지만 글을 쓰는 사람의 손에 의해 글로 변해가는 과정에서 목재가, 책상과 의자가 되어가는 것이라고 생각한다.

〈1203〉 요리: 요리할 때 레시피를 알고 냉장고에서 그에 맞는 재료를 찾아 만들어야 맛있는 요리를 완성할 수 있듯이, 글쓰기도 마찬가지로 목적에 맞게 언어와 문장을 내 머릿속에서 찾아 써야 하기 때문이다.

〈2006〉 점, 선, 면: 글쓰기가 완성되는 것이 점이 모여 선이 되고 선이 모여 면이 되는 느낌이 들기 때문이다.

본질적 사고로 분류된 이유 기술문은 쓰기에 2차적 목적을 별도로 상정하지 않고, 일련의 쓰기 행위 그 자체에 초점을 두고 작성된 내용이 대부분이었다. 이 범주에서는 한편의 글을 완성하는 것 자체가 유의미한 일로, 학생들이 생각하는 쓰기란 완성된 글을 미리 상정하고 그것을 만들어가는 과정으로 인지하고 있었다. 이는 쓰기 인식을 다룬 기존의 논의들에서 본질적 사고를 텍스트 생산을 위한 행위로 간주하고 도판, 건축 등에 비유하거나 인지주의 측면으로 접근하여 문자 언어로 나타내는 과정으로 해석한 것과 유사한 측면이 있다(변경가, 2017; 정혜승·서수현, 2010). 다만 이 연구에서는 해당 범주에서 쓰기 수행의 기능, 장점, 노동으로써의 쓰기 수행 등 쓰기 자체에 대한 본질적 측면을 보다 폭넓게 살펴볼 수 있었다.

〈112〉 운동: 글쓰기를 하면 반복적으로 글씨를 써야 해서 손 운동이 되기 때문이다.

〈520〉 모두의 상상: 가장 쉬운 방법으로 상상을 이루는 방법이기 때문이다. 글 속에서는 누구나 원하는 곳에 갈 수도 있고 하고 싶은 것도 할 수 있다.

〈729〉 꽃: 꽃에게 물을 잘 주고 꾸준히 관리하면 빛을 발하지만 꽃에 관심을 두지 않고 잘 관리하지 않으면 꽃이 금방 시드는 것처럼 글쓰기도 잘 쓰고 꾸준히 쓰면 빛을 발하는 글이 될 수 있고 관심을 가지지 않고 꾸준히 쓰지도 않으면 글쓰기 능력이 퇴화되기 때문이다.

(5) 태도

태도란 어떤 대상이나 문제에 대한 일관적인 긍정적이거나 부정적 경향을 의미한다(Eagly & Chaiken, 1993). 동일 맥락에서 쓰기 태도는 쓰기와 관련하여 일관적으로 호의나 비호의적인 감정을 불러일으키는 학습된 성향으로 정의될 수 있다(윤준채, 2008).

쓰기 태도에 영향을 미치는 요인으로는 개인적 요인, 환경 요인, 타인의 기대에 대한 신념 등을 들 수 있다(오택환, 2009). 분석 결과, 학생들이 제시한 긍정적 태도는 대부분 경험과 목적의식, 결과에 대한 신념을 중심으로 이루어진 개인적 요인에 의해 형성되었음을 알 수 있었다. 또한 긍정적 태도 범주에서는 다른 범주에 비해 다양한 비유적 표현을 찾아볼 수 있었는데 학생들이 긍정적 인식을 자신이 좋아하는 사물에 비유해 개성적인 은유 표현들을 살펴볼 수 있었다.

〈35〉 껌: 껌을 씹으면 집중이 잘 되고 스트레스가 해소된다. 그런 것처럼 글쓰기를 하면 막히는 것이 있거나 스트레스를 받을 때 해결이 되고 글쓰기에 집중이 되어 스트레스가 풀리기도 한다.

〈2202〉 팝콘: 완성된 맛있는 팝콘에 계속 손이 가는 것처럼 글쓰기에도 계속 흥미를 느끼기 때문이다.

〈2320〉 부채: 더운 날 부채로 부치면 시원하고 상쾌한 느낌이 드는 것처럼 글을 쓰면 마음이 편안해지고 기분이 좋아진다.

한편, 부정적 태도에 자주 사용된 ‘재미없다, 귀찮다, 좋지 않다’ 등의 표현은 대부분 관형어의 형태로 제시되어 있었다. ‘재미없는 숙제’, ‘귀찮은 월요일’, ‘몸에 안 좋은 담배’의 형식으로 사용되었는데, 숙제나 월요일, 담배처럼 통상적으로 부정적 이미지를 지니고 있는 명사에 덧붙여져 부정적 태도를 강조하고 있었다.

〈19〉 괴로운 고통: 글쓰기를 할 때 손으로 쓴다면 팔이 아프고 쓰다 보면 글이 뻑뻑하게 써져서 눈이 아프기도 하고 컴퓨터로 쓴다면 창작의 고통 때문에 머리가 아프기 때문이다.

〈1142〉 귀찮은 일: 재미도 없고 감동도 없고 얻어가는 느낌조차 없는 의미 없는 일이라고 생각한다.

글쓰기에 대한 부정적 태도는 학생들이 쓰기 경험을 통해 형성된 인지적, 감정적 불만족의 발현으로 볼 수 있다. 부정적 태도를 지닌 학생들이 부정의 정도를 강하게 제시하고 있다는 점에서 특히 교육의 필요성이 강조되는데, 강한 부정적 태도는 쓰기 행동을 회피하는 결과를 초래하게 되고,¹⁰⁾ 쓰기로 인한 만족을 경험하지 못한 학생은 성인이 된 이후로도 부정적 태도를 견지하게 될 가능성이 높기 때문이다.

(6) 효능감

쓰기 효능감은 Bandura(1986)가 제안한 ‘자기 효능감(self-efficacy)’에서 개념을 도출할 수 있다. 자기 효능감은 일종의 자기 능력에 대한 신념 체계로 과제 수행에 영향을 미친다. 또한 효능감은 성공적인 학업 성취를 예측하거나 보장하는 강력한 표지로 작용하기 때문에 교육에 있어 매우 중요한 요소라 할 수 있다(박영민·김승희, 2007). ‘쓰기 효능감(writing self-efficacy)’은 쓰기에 대한 자신의 평가, 혹은 쓰기 과제 수행 능력에 대한 자신의 판단을 의미한다(Pajares, 1996). 즉, 쓰기 수행에서 스스로 느끼는 자신감의 정도가 쓰기 효능감이다.

10) 국어교육학 사전에서는 태도를 정의적 속성만으로 강조하지 않고, 인지적, 행동적 속성의 결합으로 규정하고 있다(서울대학교 국어교육연구소, 1999). 쓰기에 대한 흥미와 선호의 여부가 정의적 속성이라면, 쓰기 활동에 즐거이 참여할 의사 여부는 행동적 속성에 해당된다. 이 점으로 미루어볼 때, 국어교육학에서도 태도를 감정의 차원에서만이 아니라 행동의 측면으로까지 확대 연결하여 해석하고 있음을 알 수 있다.

긍정적 효능감의 은유 표현으로 ‘칭찬’이 빈번하게 나타난 점은 긍정적 태도 범주에서 ‘즐거움’이 고빈도로 제시된 점과 비교해 볼 만하다. 즐거움은 내적 발현이 가능하지만, 칭찬은 외적 보상의 한 형태이기 때문이다. 쓰기 자신감은 외부의 긍정적 자극에 의해 강화될 수 있고 이런 자신감은 긍정적 효능감으로 이어진다.

〈1840〉 상: 어렸을 때 글짓기와 관련된 대회에 나가서 상을 받아 뿌듯했다. 그 기억 때문에 글짓기 대회를 많이 나가게 되었고 더 많은 상을 받게 되었다.

〈2303〉 저금통: 초등학교 때부터 계속해서 글을 써왔다. 부족했던 그때의 실력부터 차곡차곡 동전처럼 쌓아둔 결과로 지금의 실력을 만든 것이기 때문이다. 저금통에 모아놓은 동전으로 여러 가지를 살 수 있는 것처럼 앞으로가 기대된다.

반면, 부정적 효능감은 자신감의 결여와 관련되어 있다. 이유 기술문을 분석한 결과, 부정적 효능감을 표현한 학생 중에는 단순히 쓰기의 어려움을 호소하는 학생보다 ‘잘 쓰고 싶다’는 심리가 내재 되어있는 학생들이 더 많이 존재했다.

〈198〉 가시밭: 가시밭은 가시가 넓게 깔려 있어 들어가기가 두렵고 그 사이를 걸으면 아파 밭을 통과하기가 어렵다. 글쓰기는 나에게 잘 맞지 않고 어렵다. 글을 쓰라고 하면 막상 두렵고 막막하다. 그래서 글을 잘 쓰는 친구나 지인들이 부럽고 나는 재능이 없다고 생각을 한다.

〈2323〉 깊은 숲: 글을 쓰는 일은 너무 어렵고 한 번도 글을 잘 써본 적이 없기 때문이다. 물론 나도 나의 생각을 잘 펼치고 싶지만 조금 생각하기도 어렵고 막막하다. 그래서 글을 쓰는 일은 너무나도 어려워 생각해도 나오지 않는 깊은 숲이라고 하는 것이다.

쓰기 효능감이 낮은 학습자는 쓰기 학습 장면을 회피하려는 심리 기제

가 작동하여 학습에 소극적인 모습을 보일 것이고, 이러한 수동적 자세는 쓰기 능력의 발달을 지연시켜 쓰기 성취의 실패로 이어질 가능성이 높다. 쓰기에서 성공적 경험을 해 보지 못한 학습자는 쓰기에 대해 자신감이 결여되어 효능감이 더 낮아지는 악순환에 빠지게 될 것이다. 또한 쓰기에 대한 성취 욕구가 강할수록 그것이 충족되지 못할 때에도 부정적 효능감이 커지게 된다. 따라서 부정적 효능감을 지닌 학생들에게는 자신감을 심어줄 수 있는 정 의적 측면의 보상뿐만 아니라 ‘잘 쓰고 싶은’ 쓰기 욕구를 충족시켜 줄 수 있는 교육적 차원의 처치가 제공되어야 할 것이다.

2. 쓰기 인식의 집단별 특징

집단 간 쓰기 인식의 차이를 살펴보고자 학교급 및 성별로 집단을 분류한 후, 각 집단의 범주별 분포를 결과 값으로 산출하였다. 집단 간 분포 비율이 상이했는데, 차이의 유의성을 확인하기 위해 <표 7>과 같이 상관 분석을 실시하였다.

<표 7> 집단별 상관관계

		중등 남학생	중등 여학생	고등 남학생	고등 여학생
중등 남학생	Pearson 상관계수	1	.810*	.804*	.716*
중등 여학생	Pearson 상관계수		1	.908**	.893**
고등 남학생	Pearson 상관계수			1	.919**
고등 여학생	Pearson 상관계수				1

다른 집단에 비해 중등 남학생의 상관계수가 다소 낮았으나, 모든 집단에 유의한 차이가 존재하는 것으로 나타났다. <표 8>은 집단별 범주 분포를 비율로 나타낸 것이고, <그림 4>는 집단 간 차이를 직관적으로 비교하기 위해 <표 8>을 방사형 그래프로 도식화한 것이다.

〈표 8〉 집단 간 인식의 범주별 분포

범주	고등 여학생	고등 남학생	중등 여학생	중등 남학생
자기 지향적 사고	21.03	16.76	13.90	11.09
타자 지향적 사고	6.51	9.98	6.43	4.97
도구적 사고	3.88	9.45	3.94	5.74
본질적 사고	30.16	26.92	34.65	28.30
태도_긍정	11.26	11.76	18.46	12.05
태도_부정	9.26	12.83	10.17	22.56
효능감_긍정	3.38	2.14	3.11	5.74
효능감_부정	14.52	10.16	9.34	9.56

(단위: %)



〈그림 4〉 집단 간 인식의 범주별 분포 그래프

중학교 남학생의 경우, 다른 집단에 비해 태도 측면에서 부정적 인식이 높게 나타남을 볼 수 있다. 중학교 여학생이 긍정적 태도 지수가 높은 점과는 대별 되는데, 쓰기 태도에 있어 성별의 차이가 존재함을 확인할 수 있다. 중학교 남학생의 높은 부정적 태도는 쓰기 교육에 시사점을 제공한다. 부정적 태도는 쓰기 행위를 유발하고 지속시키는 데 방해가 되는 주요 원인이기

때문이다. 이러한 결과는 중학교 남학생의 수업 상황에서 인지적 측면의 강조만으로는 교육적 효과가 미미할 것임을 방증하는 동시에 정의적 측면을 중시한 교육의 필요성을 강조한다.

여학생은 조금 다른 양상을 보이는데, 중학교 여학생이 높은 비율의 긍정적 태도를 보이는데 반해, 고등학생의 결과에서는 긍정적 태도의 응답이 감소했다. 그러나 이러한 감소가 고등학생의 쓰기 태도가 부정적으로 변모했음을 의미하지는 않는다. 왜냐하면 이 연구는 동일한 항목을 대상으로 척도를 수치화한 방식을 따르지 않았기 때문이다. 즉, 여학생의 긍정적 태도 비율 감소는 부정적 태도로의 이양을 의미하는 것이 아니라 쓰기 인식에서 태도적 측면보다 다른 측면을 더 고려하였기 때문에 나타난 변화로 받아들여야 할 것이다. 비록 적은 수치의 변화이지만 여중생에 비해 여고생의 부정적 태도 비율도 감소했다는 것 역시 이러한 해석에 근거를 제공한다.

한편, 고등학교 여학생이 부정적 효능감 범주에서 높은 비율을 나타낸 점은 주목할 필요가 있다. 이러한 현상은 고등학교 시기에 쓰기 효능감이 떨어진다고 밝힌 박영민·최숙기(2009)의 연구 결과와도 상통한다. 효능감 발달 연구에서 부정적 효능감은 실질적인 쓰기 능력이 향상되었다 하더라도 학년이 올라갈수록 높아지는 과제 난도에 따른 자신감 결여와 스스로에 대한 엄격성의 증대로 인한 결과로 생각할 수 있다. 효능감은 성취와 유관하며 능력에 대한 신념으로 동기의 주된 요인이 된다(최건아, 2011). 고등학교 여학생이 보다 적극적으로 쓰기에 참여할 수 있도록 하기 위해서는 부정적 효능감을 줄여줄 수 있는 긍정적 보상의 마련이 선행되어야 한다.

집단별 특징에서 마지막으로 눈여겨볼 부분은 고등학교 남학생들이 쓰기를 도구로 간주한다는 점이다. 또한 이들은 다른 집단에 비해 높은 비율로 쓰기를 타자 지향적 사고의 수단으로 받아들이고 있었다. 이러한 결과를 통해 고등학교 남학생은 쓰기를 유목적적 행위로 인식하고 그 필요성에 대해 인지하는 경향이 큼을 알 수 있다. 덧붙여 고등학교 남학생들이 생각하는 쓰기란 학습 작문으로써의 성격이 강하게 작용함을 짐작케 한다. 고등학교 남

학생이 쓰기의 필요성을 인식하고 있는 만큼, 그 필요를 충족시켜줄 쓰기 전략에 대한 학습과 실제적 쓰기 능력 향상을 위한 수업 방안이 강구되어야 할 것이다.

V. 맺음말

이 연구에서는 중·고등학생들을 대상으로 쓰기에 대한 은유 표현과 이 유 기술문을 수집한 후, 텍스트 마이닝 방법을 활용하여 학생들이 지닌 쓰기 인식을 분석하였다. 먼저 전처리 과정을 거친 이 유 기술문에서 키워드를 추출하여 감정 분석을 통해 극성(긍정, 부정)과 무극성으로 분류하였다. 분류된 키워드는 극성 별로 군집화를 실시하여 자동으로 범주화하였는데 무극성은 4개의 범주로, 긍정과 부정은 각각 2개의 범주로 군집화됨을 확인할 수 있었다. 핵심 키워드들을 중심으로 범주명을 명명한 결과 중·고등학생들의 쓰기에 대한 인식은 자기 지향적 사고, 타자 지향적 사고, 도구적 사고, 본질적 사고, 태도_긍정, 태도_부정, 효능감_긍정, 효능감_부정의 여덟 가지로 범주화되었다.

은유 분석 결과를 바탕으로 학생들의 쓰기에 대한 인식의 특성을 요약하면 다음과 같다. 첫째, 25%가 넘는 학생들이 쓰기를 자기 지향적 사고와 타자 지향적 사고, 도구적 사고로 인식했다. 이러한 관점에서 쓰기를 인식한 학생들은 대부분 쓰기를 목적 달성을 위한 수단으로 간주하고 있었고, 동일 맥락에서 쓰기의 필요성을 인지하고 있었다. 둘째, 30%에 해당하는 학생들이 쓰기 행위 자체에 초점을 두어 본질적 측면에서 바라보며 인지 주의적 차원으로 접근하고 있었다. 이 학생들은 쓰기를 완결된 글을 지향하는 일련의 행위 과정이라 인식했는데 이는 텍스트 생산과 직결되는 쓰기의 기능 자체에 방점을 둔 결과로 추정할 수 있다. 셋째, 쓰기의 효능감에서 긍정적 반응

(3.55%)에 비해 부정적 반응(11.33%)을 나타낸 학생의 비율이 높았다. 낮은 효능감은 자신감 결여와 밀접한데 이러한 결과는 성공하지 못한 쓰기 경험의 반복에서 비롯된 것으로 짐작 가능하다. 부정적 효능감을 보인 학생 중에는 글을 잘 쓰고 싶은 성취 욕구를 함께 표현하고 있는 경우가 다수 있었기 때문이다. 넷째, 집단에 따라 은유 범주별로 유의한 인식의 차이가 존재했다. 중학교 남학생에서는 부정적 태도가, 고등학교 여학생에서는 부정적 효능감이 많이 나타났다. 고등학교 남학생은 소통과 필요성 측면에서 쓰기를 인식한 응답이 많았다.

이 연구에서는 텍스트 마이닝 기법인 군집화로 자동 분류한 범주에 의해 학생들의 쓰기 인식에 대한 은유 표현을 분석하였다. 이 연구는 은유 분석이 지니고 있던 자의적 범주화 문제를 해소하고 분석의 합리적 절차를 따라 쓰기에 대한 학생들의 인식을 객관적으로 지표화 하였다는 점, 은유 분석의 결과에 대해 집단별 비교를 통하여 학교급과 성별에 따른 쓰기 인식의 차이를 실증적으로 도출했다는 점에서 의의를 찾을 수 있다.

이 연구의 결과가 보다 유의미해지기 위해서는 다음의 내용들을 보완해야 할 것이다. 우선 본고에서는 학생들의 쓰기 인식만을 조사하여 제시하였을 뿐, 학생들이 지닌 인식이 실제 글에 어떻게 투영되는지는 살펴보지 못했다. 또한, 클러스터링 과정에서 K-평균법만을 사용했기 때문에 다양한 알고리즘의 적용에 따른 결과를 비교하지 못하였다. 후속 연구에서 이러한 한계점들이 충족된다면 최적화된 알고리즘 개발이 가능할 것이고, 텍스트 마이닝을 활용한 은유 분석이 쓰기 인식을 연구하는 심도 있고 객관적인, 새로운 분석 방법으로 자리 잡을 수 있을 것이다. 또한 이 분석 방법이 영역을 확장해 다양한 국어교육 분야에서 폭넓게 활용될 수 있는 연구 방법이 될 수 있을 것이라 기대해 본다.

* 본 논문은 2018. 4. 3. 투고되었으며, 2018. 5. 8. 심사가 시작되어 2018. 6. 1. 심사가 종료되었음.

참고문헌

- 고영근·구본관(2011), 『우리말 문법론』, 서울: 집문당.
- 권낙원(2008), 「은유분석을 통한 교사상 분석」, 『학습자중심교과교육연구』 8(2), 21-47.
- 김용직(1992), 『문학개설』, 서울: 박영사.
- 김은영(2004), 「감정동사 유의어의 의미 연구」, 『한국어 의미학』 14, 121-147
- 김은영(2005), 「현대 국어 감정동사의 범위와 의미 특성에 대한 연구」, 『한국어 의미학』 16, 99-124.
- 김정자(2014), 「학생의 쓰기 경험에 대한 연구: 학생들은 어떤 쓰기 경험을 하며 필자로 성장하는가」, 『한국초등국어교육』 55, 127-150.
- 김진호·김상미(2014), 「예비초등교사의 수학교과서에 대한 은유 분석」, 『수학교육』 53(1), 147-162.
- 김혜연(2015), 「예비 초등 교사의 글쓰기 요구 인식: 정체성 인식과의 관련성을 중심으로」, 『국어교육연구』 36, 425-462.
- 남길임(2016), 「비유 언어에 대한 인지언어학적 접근: “어휘”에서 “담화”까지」, 『어문논총』 68, 457-462.
- 남민우(2011), 「은유 교육의 목표와 내용 연구」, 『문학교육학』 36, 257-285.
- 박영민·김승희(2007), 「쓰기 효능감 및 성별 차이가 중학생의 쓰기 수행에 미치는 효과」, 『국어교육학연구』 28, 327-359.
- 박영민·최숙기(2009), 「우리나라 학생들의 쓰기 효능감 발달 연구」, 『새국어교육』 82, 95-125.
- 박인조·민경환(2005), 「한국어 감정단어의 목록 작성과 차원 탐색」, 『한국심리학회지 사회 및 성격』 19(1), 109-129.
- 박혜영(2016), 「국어교사의 쓰기 장르 인식에 관한 은유 분석」, 『학습자중심교과교육연구』 16(5), 403-423.
- 변경가(2017), 「초·중등 학습자의 쓰기 개념화 양상 연구-은유 분석을 중심으로」, 『작문연구』 32, 213-328.
- 서수현(2016), 「은유 분석을 활용한 대학생 필자의 ‘협동 작문’에 대한 인식 연구」, 『대학작문』 18, 303-329.
- 서울대학교국어교육연구소(1999), 『국어교육학 사전』, 대교출판.
- 손전주·박미숙·박지은·손진훈(2012), 「한국어 감정표현단어의 추출과 범주화」, 『감성과학』 15(1), 105-120.
- 안신호·이승혜·권오식(1993), 「정서의 구조」, 『한국심리학회지 사회 및 성격』 7(1), 107-123.
- 연희정·김은주·송주은(2015), 「유아교육기관의 ‘교실’과 ‘바깥놀이터’에 대한 유아교사의 은유분석」, 『학습자중심교과교육연구』 15(7), 673-699.
- 오택환(2009), 「고등학생 필자의 쓰기 태도 발달에 관한 연구」, 『국어교육연구』 45, 125-148.
- 유솔아(2011), 「은유 분석을 통해 본 다문화교육에 대한 초등교사들의 인식」, 『초등교육연구』

- 24(3), 117-145.
- 윤준채(2011), 「한국판 초등학생 읽기 태도 검사 도구의 타당도 검증」, 『한국초등국어교육』 37, 223-245.
- 윤초희(2013), 「은유(metaphor)를 통해 본 대학생의 글쓰기 신념」, 『교육심리연구』 27(2), 367-394.
- 이경언(2015), 「수학 은유 분석을 통한 대학생들의 학교 수학에 대한 인식과 선호도 조사」, 『수학교육논문집』 29(1), 51-72.
- 이성만(2000), 「인지과정을 통한 논평의 주제구조 분석」, 『독일어문학』 11, 443-468.
- 이수정(2011), 「은유의 의미를 이해한다는 것: 은유로서의 예술 개념을 중심으로」, 『한국예술연구』 3, 5-34.
- 이슬기(2017), 「작문 평가에서 텍스트 마이닝의 활용 가능성 탐색」, 『작문연구』 35, 99-131.
- 이연선(2015), 「은유와 은유분석 고찰을 통한 교육연구에서의 은유분석 역할탐구」, 『유아교육학논집』 19(1), 253-275.
- 이은영·박선영(2011), 「은유를 통해 본 예비교사의 과학과 과학자에 대한 인식」, 『유아교육학논집』 18(5), 247-270.
- 이준웅·손현주·나은경(2008), 「정서 단어 분류를 통한 정서의 구성 차원 및 위계적 범주에 관한 연구」, 『한국언론학회』 52(1), 85-116.
- 장성민(2015), 「은유 분석을 통한 고등학생의 발표 장르에 대한 인식 조사」, 『화법연구』 30, 325-360.
- 장성민(2016), 「예비 초등교사의 글쓰기 협력학습에 대한 인식: 동료지도의 작문 워크숍에 대한 은유 분석을 중심으로」, 『작문연구』 30, 253-287.
- 장영란(2000), 「그리스 비극과 철학에 나타난 자연법 사상: 소포클레스의 『안티고네』와 헤겔의 인륜 개념을 중심으로」, 『인문학 연구』 5, 150-169.
- 장은섭(2014), 「초,중,고등학생들의 쓰기에 대한 인식 조사 연구」, 『새국어교육』 99, 173-208.
- 장은주·박영민(2013), 「쓰기 중요도·선호·난도·습관에 대한 학생들의 반응 분석」, 『교육과정평가연구』 16(3), 73-97.
- 정민지(2017), 「텍스트 마이닝에서 근접성 데이터의 군집화 기법 연구」, 부산대학교 석사학위논문.
- 정혜승(2005), 「은유의 기능과 국어교육적 함의」, 『국어교육』 118, 181-219.
- 정혜승·서수현(2009), 「중등학생 필자의 독자 인식 양상」, 『청람어문교육』 39, 213-250.
- 정혜승·서수현(2010), 「중등학생 필자의 독자 고려 전략」, 『청람어문교육』 41, 291-312.
- 정혜승·서수현(2012), 「교육대학교 학생들의 작문 능력 조사 연구」, 『한국교육연구』 29(4), 321-346.
- 지태창·이현진·이일병(2007), 「인공지능: 온라인 문서 군집화에서 군집 수 결정 방법」, 『정보처리학회논문지B』 14(7), 513-522.
- 최건아(2011), 「고등학생의 쓰기 효능감 및 쓰기 결과 기대, 글 점수 간의 상관관계 분석」, 『국어교육학연구』 40, 569-591.

- 최윤희(2017), 「이공계열 학생의 글쓰기 인식에 대한 연구: K대학을 중심으로」, 『교양교육연구』 11(1), 581-608.
- 최지은·유혜진·유다빈·김나라·김윤희(2016), 「리뷰 감정 분석을 통한 전자상거래 상품 분석 및 평가 시스템 설계」, 『정보과학회 컴퓨팅의 실제 논문지』 22(5), 209-217.
- 한금윤·최선경(2016), 「대학 신입생의 글쓰기에 대한 생각과 글쓰기 교육에 대한 인식 탐색」, 『교양교육연구』 10(3), 549-576.
- 홍중선·정연주(2009), 「감정동사의 범주 규정과 유형 분류」, 『한국어학』 45, 387-420.
- Andrews, N. O., & Fox, E. A. (2007), "Recent developments in document clustering", *Technical Report 7*(35), 2-25.
- Andriessen, D. (2006), "On the metaphorical nature of intellectual capital: A textual analysis", *Journal of Intellectual Capital* 7(1), 93-110.
- Armstrong, S. L., Davis, H. S., & Paulson, E. J. (2011), "The subjectivity problem: Improving triangulation approaches in metaphor analysis studies", *International Journal of Qualitative Methods* 10(2), 151-163.
- Bandura, A. (1986), "The explanatory and predictive scope of self-efficacy theory", *Journal of Social and Clinical Psychology* 4, 359-373.
- Blondel, V. D., Guillaume, J. L., & Lambiotte, R. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment* 2008(9), 1-12.
- Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975), *The development of writing abilities (11-18)*, London: Macmillan Education Ltd.
- Cameron, L. & Low, G. (1999), "Identifying and describing metaphor in spoken discourse data, In L.", *Researching and Applying Metaphor*, 105-132.
- Eagly, A. H. & Chaiken, S. (1993), *The psychology of attitudes*, Orlando, FL: Harcourt Brace Jovanovich College.
- Gök, B. & Erdorğam, T. (2010), "Investigation of pre-service teachers' perceptions about concept of technology through metaphor analysis", *The Turkish Online Journal of Educational Technology* 9(2), 145-160.
- Lakoff, G. & Johnson, M. (1980), *Metaphor we lived by*, Chicago, IL: University of Chicago Press.
- Pajares, F. (1996), "Self-efficacy beliefs and the writing performance of entering high school students", *Psychology in the Schools* 33, 163-175.
- Rasmussen, E. N. (1992), "Clustering algorithms", *Information Retrieval: Data Structures & Algorithms*, 419-442.
- Rohracher, H. (1976), "Regulatory processes in psychic life", *Encyclopedic Handbook of Medical Psychology*, 460.
- Saban, A. (2010), "Prospective teacher's metaphorical conceptualization of learner",

Teaching and Teacher Education 26, 290-305.

Saban, A., Kocgeker, B. N., & Saban, A. (2006), "An investigation of concept of teacher among prospective teachers through metaphor analysis", *Theory & Practice* 6, 509-522.

Schmitt, R. (2005), "Systematic metaphor analysis as a method of qualitative research", *The Qualitative Report* 10(2), 358-394.

Shell, D. F., Colvin, C., & Bruning, R. H. (1995), "Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences", *Journal of Educational Psychology* 87(3), 386-398.

텍스트 마이닝을 활용한 중·고등학생의 쓰기 인식 연구 —은유 분석을 중심으로—

이슬기

이 연구는 텍스트 마이닝을 활용하여 한국 중·고등학생들의 쓰기 인식을 살펴보는 데 목적이 있다. 이를 위해 이 연구에서는 학생들을 대상으로 쓰기에 대한 은유 표현과 이유 기술문을 수집한 후, 텍스트 마이닝 방법을 활용하여 학생들이 지닌 쓰기 인식을 분석하였다. 먼저 이유 기술문에서 키워드를 추출하여 감정 분석을 실시하여 극성(긍정, 부정)과 무극성으로 분류하였다. 분류된 키워드는 극성 별로 군집화를 통해 자동 범주화하였는데, 무극성은 4개의 범주로, 긍정과 부정은 각각 2개의 범주로 군집화됨을 확인할 수 있었다. 핵심 키워드들을 중심으로 범주 명을 명명한 결과 중·고등학생들의 쓰기에 대한 인식은 타자 지향적 사고, 도구적 사고, 자기 지향적 사고, 본질적 사고, 태도_긍정, 태도_부정, 효능감_긍정, 효능감_부정의 여덟 가지로 범주화되었다. 각 범주에 따라 어떤 은유적 표현들이 있는지를 직관적으로 살펴보기 위해 범주별로 은유 표현들을 워드 클라우드로 제시하였다. 동일한 은유 표현이 서로 다른 범주에 속하는 예도 있었으나, 대체로 은유 표현들은 범주별로 상이한 차이를 나타냈다. 이유 기술문의 범주와 은유 표현의 분석을 통해 중·고등학교 학생들이 지니고 있는 쓰기 인식에 대해 추정 가능했다. 이 연구를 통해 은유 분석을 활용한 인식 연구에서 텍스트 마이닝이 새로운 방법론으로 자리 잡을 수 있음을 확인할 수 있었다.

핵심어 텍스트 마이닝, 은유 분석, 쓰기 인식, 군집화, 감정 분석, 워드 클라우드, 코퍼스 분석

Middle and High School Students' Awareness of Writing Using Text Mining —Focus on Metaphor Analysis

Lee Seulki

The objective of this study is to examine Korean middle and high school students' awareness of writing using text mining. Samples of metaphorical expressions and descriptive writing of reason were obtained from students used to analyze students' awareness of writing. First, keywords were extracted from descriptive writing of reason and a sentiment analysis was conducted to classify the keywords into polar (positive, negative) and nonpolar words. The classified keywords are automatically categorized through clustering by polarity: nonpolar keywords are clustered into four categories, and positive and negative into two categories each. As a result of naming the categories with a focus on the main keywords, middle and high school students' awareness of writing was categorized into eight types: 1) self-oriented recognition, 2) other-oriented recognition, 3) intrinsic recognition, 4) instrumental recognition, 5) attitude_ positive, 6) attitude_ negative, 7) efficacy_ positive, and 8) efficacy_ negative. To intuitively examine the metaphorical expressions in each category, this study presented the metaphorical expressions in word clouds. In some cases, the same metaphorical expressions belonged to different categories, but generally the metaphorical expressions showed differences among the categories. By analyzing the categories of descriptive writing and metaphorical expressions, it was possible to estimate middle and high school students' awareness of writing. This study verified that text mining can be used as a new methodology in awareness studies using metaphor analysis.

KEYWORDS Text Mining, Metaphor Analysis, Writing Awareness, Clustering, Emotional Analysis, Word Cloud, AntConc