

# 의미연결망을 활용한 중등 학생 필자의 쓰기 특성 분석

권태현 충북대학교 국어교육과 조교수

- \* 이 논문은 제46회 2022 세계 한국어학자 대회(2022.7.1.)에서 발표한 것을 수정·보완한 것이다. 이 논문은 2021 학년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음.

- I. 서론
- II. 이론적 배경
- III. 연구 방법
- IV. 연구 결과
- V. 결론

## I. 서론

이 연구는 의미연결망 분석 방식을 통해 중등 학생 필자들이 작성한 글의 특성을 파악하는 것을 목적으로 한다. 이를 위해 동일한 쓰기 과제에 대한 중학생과 고등학생의 글들을 평가하여 쓰기 수준을 확인하고, 의미연결망 분석을 통해 쓰기 발달 단계별 글의 특성을 비교해 보고자 한다.

학생들의 글쓰기를 촉진하기 위해서는 학생 필자에게 부족한 쓰기 능력이 무엇인지 확인하여 적절한 교육적 중재가 이루어져야 한다. 작문을 지도하는 교사는 이를 위해 학생들의 쓰기 결과를 평가하고 글 속에서 작문 교육의 내용을 찾아낼 수 있어야 한다. 그런데 사실 학생들이 쓴 글의 세밀한 특성을 제대로 평가해 내기란 쉬운 일이 아니다. 글을 구성하고 있는 다양한 언어적 요소의 특성을 정확히 판단하려면 매우 많은 시간과 노력이 들 뿐 아니라 들인 노력에 비해 만족할 만한 성과를 얻기도 어렵다. 작문 교육적 차원에서 글을 평가할 때는 글의 표면적 특성에 대한 이해만이 아니라 그 글을 쓴 필자의 의미 구성 과정을 추론할 수 있어야 하기 때문이다.

또한 쓰기 평가의 과정이 매우 주관적이라는 점도 인식할 필요가 있다.

평가자들은 서로 다른 방식으로 학습자의 글을 읽고 의사 결정 과정을 거친다(박종임, 2012). 평가 기준을 사용하여 평가의 객관성을 담보하려 하지만 평가 준거에 대한 평가자 간 해석이 달라 일관성 측면에도 한계를 지니고 있다(백유진, 2020; 천해주·김영서·박소희·김주환, 2020). 작문 평가가 단지 학생 글의 수준을 구분하기 위한 것이 아니라 학생 글에 대한 탐색을 통해 적절한 피드백을 제공하는 과정이라는 점을 고려할 때, 작문 평가의 이러한 주관성과 편향성은 실효성 있는 작문 교육의 활성화를 가로막는 한계라고 볼 수 있다.

최근에는 이를 극복하기 위해 자연어 처리기술을 통해 글의 형식적 측면을 기계 채점하거나 학생 글이 지닌 양적 정보를 피드백하여 쓰기 평가의 주관성을 보완하는 방안이 논의되고 있다(이현준·박영민, 2019ㄱ; Attali, 2004; Foltz, et al., 2013; Page & Peterson, 1995). 학생 글에 포함된 단어나 어절 수, 구문의 복잡성 등을 측정하여 양적 정보로 제공하고 이를 쓰기 평가의 과정에서 객관적 지표로 활용하는 것이다. 그런데 이러한 양적 정보는 글의 표면적 특성을 이해하는 데에 도움이 될 수는 있으나, 글을 통해 필자의 의미 구성 과정을 추론하는 데에는 크게 기여하지 못할 수 있다. 글이란 여러 단어들로 구성되지만 단순히 이 단어들의 의미를 모은다고 해서 글의 의미를 파악할 수 있는 것은 아니기 때문이다. 역으로 글을 쓰는 필자들은 다양한 단어를 사용하지만 그 단어들의 기능이나 역할은 모두 다르다고 할 수 있다. 따라서 단지 학생들이 쓴 텍스트의 표면적 특성을 양적으로 제시하는 것으로는 한계가 있다. Hillock(1986)은 텍스트 특성의 단편적 정보로 도출되는 구문의 복잡성이나 유창성이 필자의 수준이나 등급을 구분하는 기준이 될 수 있는 가능성은 있지만 텍스트가 지니고 있는 의미 연결, 즉 구성적 측면과 분명한 관련성은 없다고 지적한 바 있다.

글은 그 글을 구성하는 다양한 요인들의 내적 결속력을 통해 의미를 구성해 나가는 특성을 지니고 있다(이현준·박영민, 2016). 따라서 글을 통해 필자의 의미 구성을 추론해 내기 위해서는 글을 구성하는 각 부분들이 어떤 연

결 관계를 맺고 있으며, 전체 구조 속에서 각 부분은 어떤 의미와 역할을 갖는지 등을 살펴볼 필요가 있다. 나아가 쓰기 발달 및 수행 수준에 따라 이러한 연결 관계에 어떠한 차이가 있는지를 탐색해 볼 필요가 있다.

이 연구는 이러한 글의 분절적이면서도 종합적 특성을 파악하기 위해 의미연결망(Semantic Network) 분석을 적용하고자 한다. 네트워크 분석 또는 연결망 분석은 하나의 집단을 구성하고 있는 구성 요소들의 간의 연결 관계에 초점을 둔다(Brown, 1952). 즉, 하나의 체계는 그 구성원 간의 관계에 의해 창조된다는 관점에 각 구성 요소들이 맺고 있는 ‘관계’와 그에 따른 ‘구조’에 주목한다(Martin & Wellman, 2011). 이러한 분석 방법을 텍스트 단위에 적용한 것이 텍스트 연결망 또는 의미연결망 분석 방법이다. 텍스트 구성의 기본 단위가 단어이므로 단어 연결망이라고도 부르는 이 방법은 단어들이 연결되는 관계와 구조의 특성을 비교, 분석하는 데 유용하다. 또한 전체 텍스트가 갖는 여러 정보의 관계를 직관적으로 이해할 수 있도록 시각화하여 제시할 수 있다는 장점이 있다(이현준·박영민, 2019ㄴ).

쓰기 특성과 관련된 선행 연구들은 일반적으로 학생 글에 대한 평가자의 직관적 판단에 의존해왔는데, 대량의 학생 글을 채점하는 과정에서 평가자의 엄격성 수준이나 평가 기준 해석이 달라지기 쉬워 실제 집단별로 쓰기 능력에 어떠한 차이가 있는지를 정확히 판단해 내기 어려웠다. 이 연구는 의미연결망 분석을 통해 쓰기 수행 수준별 글의 차이를 객관화, 가시화하는 방안을 마련하고자 한다. 더불어 학생 글이 어떠한 관계적 특성을 보이고 있는지, 전체 구조 속에서 텍스트의 각 부분이 담당하는 역할이 무엇인지를 파악함으로써 글을 통해 학생들의 쓰기 과정을 추론할 수 있는 방안을 탐색해 보고자 한다.

## II. 이론적 배경

### 1. 중등 학생 필자의 쓰기 능력 발달

본 연구는 동일한 쓰기 과제에 대한 중학생과 고등학생 글을 비교, 분석함으로써 중등 학생 필자가 산출한 글의 의미연결망에 차이가 있는지 분석하는 데에 그 목적이 있다. 이를 위해 우선 쓰기 발달을 다룬 기존 연구에서 중등 학생 필자의 쓰기 능력 발달을 어떻게 다루고 있는지 살펴보고자 한다.

쓰기 능력은 다양한 요인들이 영향을 미치는 복합적인 고등 사고 능력으로 그 발달의 과정 역시 복잡하고 장기간에 걸쳐 이루어진다(박영민·이재기·이수진·박종임·박찬홍, 2018). 필자의 특성이나 상황적 요소에 따른 차이를 인정하더라도 쓰기 능력 발달 연구는 쓰기 교육 전반에서 매우 중요한 함의를 갖는다. 그러나 쓰기 능력 발달 연구는 그 중요성에 비해 연구에 필요한 시간과 노력이 과도하여 충분한 연구가 축적되지는 못한 분야라고 할 수 있다.

쓰기 능력 발달과 관련하여 가장 영향력 있는 이론은 Bereiter(1980)의 연구이다. 인지주의에 기반을 둔 이 연구에서는 쓰기 능력이 서로 다른 쓰기 기능들로 구성되어 있고, 하나의 기능이 자동화되어 완전히 습득된 후에 보다 높은 수준의 기능으로 통합되고 다시 자동화되는 과정을 거친다고 보았다. 즉, 능숙한 쓰기란 여러 가지 쓰기 기능들이 서로 잘 통합되었을 때 가능한 것이라는 관점이다. 이러한 쓰기 기능의 숙달을 바탕으로 Bereiter는 쓰기 능력 발달을 연상적 쓰기 단계, 수행적 쓰기 단계, 의사소통적 쓰기 단계, 통합적 쓰기 단계, 인식적 쓰기 단계로 구분하였다. Bereiter의 연구에서 이 단계가 특정 학령과 직접 대응되지는 않았으나 머릿속에 떠오르는 생각을 그대로 옮겨 적는 연상적 쓰기 단계나 쓰기의 어법이나 규칙, 관습에 익숙해지는 수행적 쓰기 단계의 교육 내용이 주로 초등학교급과 관련된다고 볼 때,

중등 학생 필자의 글은 주로 수행적 쓰기 단계 이상의 발달 수준을 보인다고 추론해 볼 수 있다.

물론 이러한 단계 구분은 절대적인 것이 아니며 학습자의 개인차도 존재하겠으나 본 연구에서 중등 학생의 쓰기 발달과 관련하여 가장 주목한 것은 통합적 쓰기 단계이다. 통합적 쓰기(unified writing) 단계는 필자가 쓰기 과정에서 예상되는 독자의 입장을 고려하는 것과 동시에 자신이 독자가 되어 글을 비판적으로 점검하고 수정할 수 있는 단계를 의미한다. 이는 필자 자신의 비판적 읽기를 전제로 한 것인데, 이성영(2008)에 따르면 통합적 쓰기 단계는 읽기 발달 단계 중 비판적 읽기기(고등학교 10~12학년)<sup>1)</sup>와 관련된다. 기초 문식성이 어느 정도 완수되는 중학생 필자는 의사소통 목적과 예상 독자를 고려한 쓰기 기능이 완성되는 단계에 있다고 볼 수 있으며, 여기서 좀 더 나아가 고등학생 필자는 자신의 글쓰기를 비판적으로 점검하고 수정하는 상위 인지가 발달하는 시기라고 볼 수 있다. 이렇게 볼 때, 상위 인지를 통한 글의 수정과 완성도 제고는 중학생과 고등학생 필자의 글을 구분하는 하나의 준거로 작용할 수 있다.

이러한 이론적 고찰의 결과는 교육과정의 성취기준을 통해서도 확인이 가능하다. 2015 개정 교육과정에서는 중학교 1~3학년군과 고등학교 1학년의 쓰기 성취기준을 별도의 범주로 설정하고 각각의 교육 목표를 정해 두고 있다. <표 1>은 2015 개정 국어과 교육과정(교육부, 2015) 중학교군과 고등학교 1학년 쓰기 성취기준에서 유사한 쓰기 기능을 담은 성취기준을 비교한 것이다.

---

1) 이성영(2008)에서는 Chall(1996)의 읽기 발달 단계가 Bereiter(1980)의 쓰기 발달 단계와 일정하게 대응된다고 보았다. 이 연구에서는 Chall의 4단계에 해당하는 '다양한 관점을 읽는 단계'를 글 혹은 필자의 관점에서 벗어나 다양한 시각에서 글을 바라볼 수 있다는 점에서 비판적 읽기 단계로 부를 수 있다고 보았다.

〈표 1〉 중학교와 고등학교 쓰기 성취기준 비교(2015 개정)

중학교 1-3학년군	고등학교 1학년
(9국03-04)주장하는 내용에 맞게 타당한 근거를 들어 글을 쓴다.	(10국03-02)주제, 독자에 대한 분석을 바탕으로 타당한 근거를 들어 설득하는 글을 쓴다.
(9국03-05)자신의 삶과 경험을 바탕으로 하여 독자에게 감동이나 즐거움을 주는 글을 쓴다.	(10국03-03)자신의 경험과 성찰을 담아 정서를 표현하는 글을 쓴다.
(9국03-09)고쳐쓰기의 일반 원리를 고려하여 글을 고쳐 쓴다.	(10국03-04)쓰기 맥락을 고려하여 쓰기 과정을 점검·조정하며 글을 고쳐 쓴다.

상기 성취기준을 살펴보면 각 시기별로 쓰기 교육의 초점이 어디에 있는지 확인할 수 있다. 중학교 시기에는 쓰기의 맥락을 이해하고 다양한 유형의 글을 효과적으로 쓰는 방법을 익히는 데 초점이 있다면, 고등학교에서는 쓰기의 목적과 맥락을 고려하여 자신의 글을 점검하고 조정하면서 글을 쓰고 소통하는 능력에 중점을 두고 있다. 설득하는 글을 예로 들면 중학교에서는 [9국03-04]를 통해 주로 글의 내용적 측면에서의 타당성에 초점을 두었다면 고등학교에서는 [10국03-02]를 통해 보다 확장된 쓰기 맥락에 대한 분석을 기반으로 상황에 맞는 글쓰기를 강조하고 있다. 그리고 이러한 발달을 가능하게 하는 쓰기 능력으로 [10국03-04]를 제시하여 상위 인지를 통한 글 고쳐 쓰기를 제시하고 있다.

교육과정에서 설정한 성취기준은 실제 학생들이 학교 쓰기 교육을 통해 도달해야 하는 기준점과 관련되므로 이상적으로 보면 이들 성취기준이 제시하는 쓰기 특성이 해당 학령의 학생 글에 어느 정도 드러나야 한다고 볼 수 있다. 물론 학교 쓰기 교육의 실제성 정도나 학습자의 개인차 등이 존재하겠으나 교육과정에서 상정한 학년별 쓰기 교육의 목표 역시 중학생과 고등학생 글의 특성을 구분하는 하나의 준거로 기능할 수 있을 것이다.

그렇다면 상술한 쓰기 능력의 발달을 가정할 때, 실제 학생 글에서 그 차이는 어떠한 특성으로 드러날 수 있을까? 학생들이 산출한 텍스트에 그들의 쓰기 능력이 반영되어 있다고 가정한다면 쓰기 결과를 평가하는 내용 및



조직의 측면에서 일정한 발달의 양상이 드러날 것이다. 우선 내용적 측면에서는 보다 확장되고 적절한 쓰기 내용의 선정을 들 수 있다. 쓰기 발달 단계상 중학교 시기는 Bereiter(1980)의 의사소통적 쓰기 단계에 가깝고 가은아(2011)에 따르면 확장적 쓰기 단계에 해당한다. 이 시기는 예상 독자를 고려하여 다양한 내용을 확장적으로 구성할 수 있는 단계이므로 이를 전후하여 학생들의 글은 내용적 다양성이 보다 확대될 것이라 기대해 볼 수 있다. 가은아(2011)에서는 중학교 1~3학년과 고등학교 1학년까지를 확장적 쓰기 단계로 구분하고 이때 학생들의 관심사나 배경지식이 확대되면서 화제의 선택이 보다 다양해지고 초점이 명확해진다고 보았다. 이러한 발달 과정이 일반적이라면 중학생의 글에 비해 고등학생의 글에는 보다 다양한 화제가 등장하고 독자의 요구에 기반하여 가치 있는 정보의 선택이 이루어질 것으로 추측할 수 있다. 이러한 특성은 글에 드러나는 유창성 지표로도 확인될 수 있는데, 김영옥(2015)에 의하면 중학생 글에 비해 고등학생의 글은 글자 수, 단어 수, 문장 수, 조사 수, 화제 수 등 쓰기 유창성 평균값이 모두 높았으며, 이는 글의 전체 점수와 일정한 상관성을 보이고 있음을 확인하였다.

일반적으로 내용적 다양성은 글 조직의 체계성을 요구하게 된다. 간단한 내용은 단순한 조직으로도 충분히 전달이 가능하지만 글의 주제와 관련하여 보다 다양한 내용이 개입되면 이를 독자에게 전달하기 위해 능숙한 필자는 글의 조직을 체계화하게 되기 때문이다. 쓰기 발달의 측면에서도 학년이 올라갈수록 학생들의 글은 필자의 비판적 점검을 통해 보다 체계적인 구조를 취하는 경향이 있다. 가은아(2011)에서는 중학교 시기와 고등학교 1학년까지를 확장적 쓰기 단계, 고등학교 2, 3학년 시기를 통합적 쓰기 단계로 구분하고 고등학교 시기에 학생 글이 질적으로 큰 변화를 보인다고 주장하였다. 특히 고등학교 2, 3학년 시기는 그 이전 시기와 비교해서 글의 조직적 측면의 변화가 두드러지게 나타났다. 즉, 글의 조직에서 유기적이고 체계적인 글의 비중이 높아진 것인데 이는 학생들이 내용과 조직, 표현적인 부분을 총체적으로 통합하면서 자신의 글을 점검하고 수정할 수 있는 능력을 갖추

있기 때문인 것으로 보았다.

이러한 발달의 과정은 엄격하게 구분되기보다는 서로 중첩되거나 특정 단계를 건너뛰는 특성도 있는 것으로 알려져 있다. 더불어 글 유형에 따른 편차도 존재한다는 점에서 일반화하기에는 무리가 있다. 그러나 보편적인 수준에서 학년이 올라갈수록 글의 유창성이나 담화 지식, 비판적 문식성이 증가한다는 가정은 가능할 것이다. 본 연구에서는 선행 연구에서 발견한 이러한 차이가 실제 학생 글에서 도출된 의미연결망을 통해 어떻게 드러나는지를 확인해 보고자 한다.

## 2. 작문 연구에서 의미연결망 분석

의미연결망 분석은 네트워크 분석 방법의 하나로 텍스트마이닝을 기반으로 하는 연구 방법이다. 네트워크 또는 연결망 분석은 사람이나 집단, 사물이나 대상의 연결 관계를 통해 개별 구성 요인이 차지하는 역할이나 다른 요인들과의 관계를 설명하여 전체 구성 집단의 특성을 파악하고자 하는 분석 방법이다(김우주, 2015). 기존의 사회과학 연구가 개체가 지닌 속성(attribute)을 중심으로 가설을 세우고 검증하는 방식으로 연구를 발전시켜 왔다면, 연결망 분석에서는 이를 관계적 측면으로 전환한다(김용학·김영진, 2019). 이러한 전환의 근거에는 ‘부분의 합이 곧 전체와 같은가?’라는 질문이 자리하고 있다. 대개 개체의 속성을 이해한다고 해서 그 집단의 전체적 양상을 파악하기는 어렵다. 예를 들어 신규 교사의 교수 성향은 교사의 성별이나 경력 등과 같은 개인적 속성뿐만 아니라 그가 학교에서 마주하는 학생이나 교류하는 교사들과의 상호작용으로 더 많이 설명될 수 있다. 이처럼 연결망 분석은 개별 속성 중심의 연구 방법에서 파악하지 못한 개체 간의 ‘관계’와 이에 따른 ‘구조’에 대해 설명할 수 있다는 장점이 있다.

연결망 분석이 복잡한 관계를 좀 더 쉽게 설명할 수 있는 분석 도구로 각광받으면서, ‘개체’와 그 개체들로 구성된 ‘구조’를 분석하는 다양한 학문

분야로 연구의 범위가 확대되어 갔다. 글 역시 단어라는 개체로 구성된 구조물이라는 점에서 연결망 분석의 대상이 된다. 이 연구 방법은 텍스트 분석을 위해 의미 있는 단어를 선정하고 그 단어들의 동시 출현 관계와 빈도, 유사성 등을 분석하는 방법으로 시작되었다(Callon, et al., 1983). 최근에는 텍스트마이닝 기법의 발전으로 빅데이터나 대단위 말뭉치에서 특정 주제나 시기와 관련된 단어를 선정하여 그 변화 양상을 살펴보거나 단어 간의 의미 관계를 분석하는 방식으로 발전되었다. 텍스트를 의미연결망 모델로 변환하여 시각화하면 해당 텍스트를 빠르게 탐색하고 분석해서 한 눈에 파악할 수 있으며, 단어가 연결망에서 어떻게 상호 연결되고 맥락적으로 위치하는지를 알 수 있다(Drieger, 2013).

한 편의 글을 하나의 사회 집단으로 본다면 개별 단어 또는 그 단어들이 구성하는 메시지들이 해당 집단의 구성원이 된다고 볼 수 있다. 한 편의 글을 구성하는 하위 메시지들 간의 관계를 파악하면 전체 글의 의미와 구조적 특성을 확인해 볼 수 있다. 이러한 발상은 글 자체가 하나의 유기체라는 관점에 기반한 것이다. 한 편의 글이 단어와 단어 간의 연결 관계를 통해 의미를 구성해 간다는 전제하에 단어의 의미 관계를 통해 전체 글의 특성을 분석하는 것이 의미연결망 분석인 것이다.

글을 대상으로 하는 의미연결망 분석은 일반적으로 텍스트에 등장하는 단어를 동시 출현 네트워크(co-occurrence network)로 변환하여 빈도 분석하는 접근 방식을 취한다. 여기서 텍스트에 쓰인 단어의 연결 관계를 파악하는 핵심적인 방식은 인접행렬(adjacency matrix)을 활용하는 것이다. 인접행렬은  $(i, j)$  사이의 관계가 있고 없음을 1과 0으로 나타내는 행렬을 의미하는데 본래 어떤 한 지점과 다른 한 지점으로 이어지는 경로를 분석하고 이해하는 방법으로 쓰였다. 네트워크 분석에서 사용하는 인접행렬 역시 이러한 수학적 개념에 토대를 두고 있는데 각각의 점이 되는 개체를 ‘노드(node)’ 또는 ‘액터(actor)’로, 점과 점 사이의 경로를 ‘엣지(edge)’ 또는 ‘링크(link)’라고 부른다. 즉, 연결망 분석에서 인접행렬은 ‘노드와 노드의 연결 관계를 엣

지로 표현한 행렬'이라고 부를 수 있다. 이러한 노드의 관계는 연결의 방향성과 가중치의 유무에 따라 구분되는데, 일반적으로 특정한 방향성이 없는 단순 연결 관계는 양방향성이 모두 가능한 것으로 간주되며, 연결의 가중치란 두 노드의 연결 관계가 몇 번 나오는가를 의미한다(Laydesdorff & Welbers, 2011).

의미연결망 분석에서 노드는 텍스트를 구성하는 각 단어가 되며, 가중치란 단어 간 연결 빈도가 된다. 즉 단어와 단어의 연결 빈도가 높으면 가중치가 커지고 연결 빈도가 낮아지면 작아진다. 또한 텍스트에서 단어의 연결은 방향성을 설정하기 어렵기 때문에 가중치를 지닌 단순 연결 관계로 표현할 수밖에 없다. 요컨대 의미연결망 분석의 기초 자료가 되는 인접행렬은 텍스트에 등장하는 '단어의 연결 관계를 엣지로 표현한 행렬'이 된다. 이를 통해 전체 연결망의 구조나 관계의 특성, 개별 단어의 기능이나 위상 등을 파악할 수 있다.

의미연결망 분석에서는 이러한 인접행렬을 토대로 그래프를 그릴 수 있다. 즉, 의미연결망 분석에서 연결망 그래프는 인접행렬의 구축을 통해 만들어진 결과물에 해당한다. 일반적으로 중등 학생 필자가 작성하는 글은 수백, 수천 개의 단어로 구성되는데, 단어의 수가 증가하여 그 연결 관계가 복잡해지면 직관적인 파악이 어려워진다. 이러한 단점을 보완하여 연결 관계를 시각적으로 표현한 것이 의미연결망 그래프이다. 이 그래프에서는 단어와 단어 간의 연결 관계가 있으면 선으로 잇고, 가중치를 반영하여 선의 굵기를 달리하기도 한다. 최근에는 NetMiner와 Pajek, Gephi 등 다양한 프로그램이 등장하여 단어 연결 관계를 반영한 그래프 생성이 수월해졌다.

의미연결망 분석을 활용하면 글에서 사용된 단어의 의미를 토대로 단어 간 구조적 관계와 패턴을 분석하여 연결망 구조 속에 내포된 의미 체계를 발견할 수 있다(김효림·전익기, 2018). 특히 네트워크 그래프를 통해 선형적으로 구성된 글에서는 잘 발견되지 않던 특성들을 파악할 수 있으며 각 단어들의 연결 관계뿐만 아니라 전체 글의 단어 구조가 지닌 밀도와 군집성 등을

파악할 수 있게 해 준다는 점에서 의미연결망 분석은 작문 연구 분야에서의 활용 가치가 높다고 할 수 있다.

### III. 연구 방법

#### 1. 분석 대상

이 연구에서는 쓰기 발달 단계에 따른 학생 글의 특성을 분석하기 위해 2020년 9월~12월 간 수집된 동일 지역의 중학생과 고등학생 글 각 50편을 연구 대상으로 선정하였다. 중학생과 고등학생에게는 국어 수행 평가 시간에 동일한 쓰기 과제가 부여되었고, 별도의 자료는 제공하지 않았다. 학생들은 쓰기 과제인 ‘과학 연구를 위해 동물 실험은 정당하다’라는 논제에 대해 주장하는 글쓰기를 수행하였다. 학생들이 쓴 글은 일차적으로 분석을 위해 워드 프로그램에 입력하였다. 다만 본 연구는 의미연결망 분석을 목적으로 하고 있으므로 맞춤법이나 띄어쓰기 등 글에 나타난 표면적 오류는 입력 과정에서 수정하였다.

해당 글에 대한 평가는 경력 5년 이상의 현직 국어 교사 3인에게 의뢰하였으며, 평가 결과에 따라 전체 평균 점수를 기준으로 학교급별로 상, 하 수준으로 구분하였다. 학생 글에 대한 평가 기준은 박영민·가은아(2009)에서 활용된 평가 기준을 일부 수정하여 3개 범주(내용, 조직, 표현)에 대한 6점 척도의 평가 기준표를 활용하였다.

#### 2. 분석 도구

의미연결망 분석에 앞서 수집된 학생 글을 평가하고 채점 결과를 검증하기 위해 python 3.7의 scipy.stats 패키지를 이용하여 채점 결과에 대한 기

술 통계 및 평가자 간 신뢰도, 쓰기 평균 점수를 기준으로 한 상, 하 수준 구분의 타당성을 확인하였다.

학생들이 작성한 글에 대한 데이터 분석에도 python 3.7 버전을 사용하였다. 본 연구에서는 글에서 의미를 지닌 명사만을 추출하기 위해 KoNLPy 패키지 중 Mecab 클래스를 사용하였다. 명사 추출 과정에서는 글의 중심 의미에 초점을 두기 위해 일반명사(NNG)와 고유명사(NNP)만을 추출하였다. 수사나 대명사 등을 모두 포함할 경우 의미없는 연결 관계까지 모두 연결망에 잡혀 분석에 어려움이 있기 때문이다. 형태소 분석 후 명사를 추출하는 방식으로 연구를 진행하였기 때문에 일반적인 전처리 과정(숫자, 부호 제거, 불용어 제거 등)은 불필요하였다.

의미연결망 분석을 위한 프로그램에는 UCINET과 Gephi를 활용하였다. UCINET은 연결망 분석을 위해 가장 광범위하게 사용되고 있는 프로그램으로 UC Irvine 대학의 Linton Freeman 교수와 그 동료들에 의해 발전되어 왔다(김용학·김영진, 2019). 다만 UCINET에서 제공하는 NetDraw는 레이아웃이 제한적이며 대규모 네트워크를 처리하는 데 무리가 있어 연결망 시각화에는 Gephi 0.9.2를 활용하였다.

### 3. 분석 방법

본 연구에서는 중, 고등학생 쓰기의 전반적 특성을 살펴보기 위해 학생들의 개별 글을 기본적인 분석 대상으로 삼고, 쓰기 발달 단계별 특성을 살펴보기 위해 학교급, 수준별로 글을 통합하여 총 6개의 집단 텍스트(중학교 전체, 고등학교 전체, 중학교 상 집단, 중학교 하 집단, 고등학교 상 집단, 고등학교 하 집단)에 대한 분석도 진행하였다.

우선 학생 글을 의미연결망 분석이 가능한 데이터로 가공하기 위해 python 3.7을 활용하여 자연어처리하고 각 텍스트에서 명사를 추출하였다. 개별 글은 단어 수가 많지 않아 분석 대상이 되는 단어의 수를 제한할 필요가

없으나, 학생 집단별 텍스트는 단어의 수가 지나치게 많아져 이를 일정 수준으로 제한하였다. 의미망 분석에서 너무 많은 단어를 동시에 처리하면 연결망 프로그램의 허용 범위를 넘어서기도 하고 의미 없는 단어들까지 분석 대상에 포함되면, 연결 관계가 지나치게 복잡해져 관계 파악에 어려움이 따른다. 이 연구에서는 고등학교와 중학교 전체 집단의 텍스트에서 출현 빈도 12회 이상의 단어를, 각 학교급별 상 수준과 하 수준 텍스트에서는 출현 빈도 6회 이상의 단어들을 추출하였다. 이후 nltk 패키지를 활용하여 개별 글과 집단 텍스트에서 추출한 단어들의 bigram을 형성하고, 동시 출현 관계를 통해 인접행렬을 도출하였다. 이렇게 구성된 인접행렬의 각 단어는 연결망에서 노드가 되고 연결 관계의 수는 엣지가 된다(윤태일·이수안, 2018).

이렇게 선정된 단어들과 그 인접행렬의 분석을 위해 UCINET 6.0을 활용하여 각 텍스트가 지닌 관계적 속성을 분석하였다. 이 연구에서 의미연결망 분석은 학생들이 쓴 텍스트가 분석 대상이므로 일반적인 사회연결망 분석(Social Network Analysis)과는 연구 방식이 다소 다르다. 소셜네트워크 분석은 네트워크 수준(network level), 노드 수준(node level), 관계 수준(dyad level) 등 세 가지 수준에서 분석이 가능하다(Borgatti, et al., 2013). 일반적인 사회연결망 분석에서는 구성원의 역할이나 관계의 측면에 주목하므로 노드 수준이나 관계 수준의 분석에 초점이 있다. 그러나 글을 대상으로 하는 연결망 분석에서는 글의 주제가 부여되었으며, 이미 동시 출현 네트워크를 구성하는 과정에서 빈도수가 높은 단어들이 추출되었고 네트워크 내에서 어떤 단어가 강한 영향력을 가지는지 등은 어느 정도 파악이 가능하다. 따라서 텍스트 전체의 특성, 즉 각 연결망의 성격이나 연결망 간의 차이가 보다 중요한 분석 내용이 된다. 이를 위해 본 연구에서는 네트워크 전체의 연결 정도나 밀도, 이행성(transitivity)<sup>2)</sup> 등을 통해 네트워크의 연결성과 응

---

2) 이행성(transitivity)은 세 노드 간의 3자 관계를 바탕으로 하여 네트워크의 밀집도를 나타내는 지표이다. 연결망 전체에서 이행성이 가능한 3자 관계 대비 얼마나 많은 이전성을

집성을 확인하고자 한다. 또한 이 연구에서는 Gephi에서 제공하는 모듈러리티(modularity) 기능을 활용하여 각 연결망의 의미 군집을 탐색하였다. 모듈러리티는 네트워크 내 단어 군집의 밀도 차이를 이용하여 연결망의 하위 집단을 탐색하는 기법으로 이 연구에서는 각 텍스트의 의미 분화의 정도를 확인하기 위해 참고한 지표이다.

전체 연결망의 구조를 살펴본 후에는 각 연결망의 의미적 관련성을 살펴보기 위해 집단 중심성 수치를 확인하였다. 집단 중심성이란 연결망 분석에서 노드 수준의 중심성을 네트워크 전체 수준으로 확장한 개념으로 전체 연결망에서 각 노드의 중심성이 얼마나 다른지를 나타내는 지표이다. 이를 활용하면 각 단어가 지니는 의미적 연결성이나 중앙성, 매개성 등이 전체 연결망 수준에서 어느 정도 차이가 나는지를 확인할 수 있다. 이는 각 텍스트에서 단어들이 지니는 기능이나 역할의 차이를 드러내는 주는 것으로서, 이를 통해 각 글에서 중심 내용과 주변 내용의 의미 분화 정도를 살펴볼 수 있다.

마지막으로 쓰기 발달 단계에 따른 학생 글의 의미연결망 차이를 시각적으로 비교, 분석해 보기 위해 시각화 프로그램인 Gephi를 활용하였다. 네트워크 시각화를 지원하는 다양한 소프트웨어가 있으나 이 중에서도 Gephi는 기초적인 분석 기능 외에 탁월한 시각화 기능을 제공하여 최근 가장 활발히 쓰이고 있는 프로그램이다(곽기영, 2019). 이 연구에서는 분석 대상인 중학교와 고등학교 상, 하 집단 및 개별 글에 대한 연결망을 시각화하여 제시하였다. 연결망 시각화는 수치로 구성된 연결 관계를 가시화함으로써 전체 연결망이 어떤 구조를 지니고 있는지 알 수 있도록 해 준다. 또한 노드의 크기나 엣지의 굵기 등을 조절함으로써 구성원의 역할이나 연결 관계를 직관적으로 보여줄 수 있다. 이 연구에서는 연결망 패턴을 통해 유사한 수준의 학생 텍스트에서 집약적으로 드러나는 특성을 확인해 보고자 하였다.

---

갖춘 3자 관계가 존재하는지를 계산하여 산출한다(김용학·김영진, 2019).



## IV. 연구 결과

### 1. 학생 글 평가 결과

의미연결망에 대한 통계적 분석을 시행하기에 앞서 학생 글에 대한 평가 결과의 신뢰도 검증을 통해 글의 수준 구분이 타당한지 살펴보았다.

〈표 2〉 평가자 간 신뢰도

	평가자 수	항목 수	ICC	
			단일측도	평균측도
중학생 글	3	50	.856	.952
고등학생 글	3	50	.909	.981

글에 대한 평가는 경력 5년 이상의 현직 교사 3명이 중학생 글 50편과 고등학생 글 50편, 총 100편의 글을 대상으로 실시하였다. 평가 과정에서는 중, 고등학교급과 무관하게 동일한 채점 기준과 엄격성을 유지하여 중학생과 고등학생 글을 채점하도록 하였다. 평가 결과 평가자 3인의 평가자 간 신뢰도는 급내상관계수(ICC)의 평균측도가 모두 .90 이상( $p < 0.00$ )으로 나타나 평가자 간 매우 유사한 채점 경향을 보인 것으로 나타났다.

〈표 3〉 평가 결과의 기술 통계 및 t 검증 결과

학년	집단	N	M	SD	SE	t	df	p
중	상	19	29.21	1.69	.45	9.250	35	.00
	하	31	19.61	1.56	.20			
고	상	26	30.61	1.23	.35	9.140	38	.00
	하	24	23.59	1.30	.28			

수행 수준을 상, 하로 나누는 기준은 각 학년의 평균 점수를 기준으로 중학생 23.56, 고등학생 27.23으로 정하였다. 평가자들이 동일한 평가 엄격성을 유지했다고 했을 때, 중학생 집단과 고등학생 집단의 평균 차이는 3.67 점 정도로 나타나 통계적으로 유의했으나 예상보다 차이가 크지 않았다.

이와 함께 상, 하 수준에 따른 차이가 유의한지 살펴보기 위해 python의 scipy.stats 패키지를 활용하여 t검증을 실시하였다. 이때 두 집단의 등분산을 가정할 수 있는지 살펴보기 위해 Levene의 등분산 가정을 실시한 결과, 유의확률은 중학교 집단이 .406, 고등학생 집단이 .246으로 두 집단 모두 등분산이 가정됨을 확인하였다. 이에 따라 t검증을 실시한 결과 중학교의 경우, 상 집단의 평균이 29.21, 하 집단의 평균이 19.61로 나타나 통계적으로 유의한 차이가 있는 것으로 분석되었고, 고등학교의 경우 상 집단의 평균이 30.61, 하 집단의 평균이 23.59로 역시 통계적으로 유의한 차이가 있는 것으로 분석되었다. 또한 내용, 조직, 표현 등 평가 변인에 따른 평균 차이를 검증한 결과 역시 모두 유의한 차이가 있는 것으로 나타났다.

## 2. 의미연결망 크기 비교

학생들이 글을 쓸 때 사용하는 단어의 수는 연결망 전체의 크기를 결정한다. 일반적으로 연결망의 분석 단위인 노드는 형태소, 단어, 합성어, 구 등 다양하게 정할 수 있는데 이 연구에서는 글의 내용을 전개하는 기본 단위로서 일반 명사만을 대상으로 하였다. 명사 외에 동사나 형용사, 일반 부사까지도 내용어의 범주에 포함시키기도 하는데(배도용, 2014), 그러한 단어 모두를 분석 대상으로 삼을 경우 연결망의 연결 관계가 지나치게 복잡해져 텍스트의 주요 의미 구조를 분석해 내는 데 오히려 어려움이 따를 것으로 판단하였다.

그런데 글에서 반복적으로 등장하는 모든 명사들이 연결망의 노드로 표현되는 것은 아니다. 학생들이 반복적으로 쓰는 명사는 하나의 노드로 표현되고, 다만 노드의 크기나 다른 노드와의 연결선의 개수, 굵기 등을 통해 반

복 사용이 표현될 뿐이다. 따라서 학생 글에 등장하는 전체 단어 수와 함께 동일한 형태를 지닌 단어를 통합한 빈도 반영 명사 수를 확인해 볼 필요가 있다. <표 4>는 학교급 및 수행 수준에 따라 명사 수의 평균과 비율을 요약한 결과이다.

<표 4> 집단별 연결망 크기 비교

학년	집단	표본 크기	전체 명사 수		빈도 반영 명사 수		(a, b) 비율
			평균(a)	표준편차	평균(b)	표준편차	
고	전체	50	181.70	42.16	103.48	25.20	56.95%
	상	26	188.69	37.92	106.61	20.33	56.50%
	하	24	165.70	35.36	96.97	23.55	58.52%
중	전체	50	148.65	19.72	73.33	45.56	49.33%
	상	19	186.58	40.33	90.02	23.55	48.25%
	하	31	125.62	12.54	64.36	17.58	51.23%

이 연구에서 수집한 학생들의 글 각각에서 추출한 명사 수의 평균을 살펴보면 중학생 집단은 148.65, 고등학생 집단은 181.70으로 학년별로 유의한 차이가 있는 것으로 나타났다. 이를 다시 수준별로 살펴보면 중학생 상 집단은 평균 186.58개, 하 집단은 평균 125.62개의 명사를 사용했으며, 고등학생 상 집단은 평균 188.69개, 하 집단은 평균 165.70개의 명사를 사용한 것으로 나타났다. 상, 하 수준별 평균 차이를 보면 고등학교가 22.9개, 중학교가 60.9개로 거의 3배 가까이 차이가 난다. 이를 통해 볼 때, 중학교와 고등학교의 학교급별 차이보다는 각 학년의 상, 하 수준별 차이가 더 크다는 것을 알 수 있다. 특히 중학교 하 집단 글에서 사용된 평균 명사 수가 125개 정도에 그친 점으로 볼 때, 중학생 하 집단은 쓰기 과제에 대해 일정 분량 이상의 글을 생성해 내는 데 어려움이 있었던 것으로 보인다.

전체 명사 수가 학생들이 작성한 글의 길이와 관련된다면 빈도 반영 명

사 수는 학생들이 사용한 단어의 다양성과 연관된다. 전체 추출 명사 대비 동일 형태의 명사를 통합한 빈도 반영 명사 비율을 살펴보면 고등학교와 중학교가 각각 56.95%, 49.33%로 중학생에 비해 고등학생 글의 어휘 다양성이 전반적으로 더 높은 것으로 나타났다. 그러나 학교급 내의 상, 하 수준별 비율은 다소 상이한 결과를 보여준다. 고등학교 상, 하 집단이 각각 56.50%, 58.52%로, 중학교 상, 하 집단은 각각 48.25%, 51.23%로 나타나 중, 고등학교 모두에서 상 집단 비율이 하 집단보다 낮았다. 이러한 결과는 상 집단의 어휘 다양성이 하 집단보다 낮다는 것을 의미한다. 그러나 이는 상 집단이 전반적으로 하 집단에 비해 긴 글을 쓴다는 점과 관련지어 해석해야 한다. 즉, 상 집단은 상대적으로 적은 수의 단어를 사용하여 긴 글을 작성할 수 있는 능력이 있다고 해석할 수 있다. 다만, 다양한 길이의 글이 포함된 전체 집단을 기준으로 보면 고등학생의 어휘 다양성이 중학생보다 높음을 확인할 수 있다.

〈표 4〉를 통해 상 수준 학생들은 하 수준 학생들에 비해 더 많은 명사 즉 의미 단어를 사용하여 글을 쓴다는 점을 확인할 수 있다. 의미 단어의 수가 늘어난다는 것은 설득하는 글을 쓰는 과정에서 보다 많은 주장과 이유, 근거를 들고 관련 설명을 풍부하게 담아낸다는 것을 의미한다. 물론 단순히 동일한 내용을 반복해도 단어의 개수는 늘어나지만 빈도 반영 명사의 비율을 고려해 보면 그러한 반복성이 있다 하더라도 학년 및 수준별 편차는 크지 않을 것으로 보인다. 쓰기 과제나 표본 수에 제한이 있으므로 일반화하기는 어렵지만 이 연구에서 확인한 바에 따르면 글에서 사용한 의미 단어의 수는 중학교와 고등학교급의 차이에 비해 각 학교급 내 쓰기 수준별 차이가 더 심한 것으로 나타났다. 이는 기본적으로 중학교와 고등학교 모두에서 쓰기 수준이 낮은 집단을 대상으로 쓰기 유창성에 대한 교육적 처치가 필요함을 보여주는 결과라고 할 수 있다.

### 3. 의미연결망 구조 분석

텍스트에서 추출한 단어의 빈도는 연결망 전체의 규모를 결정하는 요인이나 이를 통해 연결망의 구조를 확인하기는 어렵다. 이 연구에서는 텍스트의 의미연결망 구조를 살펴보기 위해 각 연결망의 연결 정도와 밀도, 이행성의 평균을 살펴보았다. 이는 각 텍스트별로 도출해 낸 연결망을 분석 대상으로 하여 분석한 구조 지표의 수치를 평균한 것이다. <표 5>는 연결망의 구조 지표를 제시한 것이다.

<표 5> 의미연결망의 구조 지표

학년	집단	연결 정도 Average degree		밀도 density		이행성 transitivity		모듈래리티 modularity	
		평균	표준 편차	평균	표준 편차	평균	표준 편차	평균	표준 편차
고 등 학 교	전체	82.43	20.56	0.653	0.220	0.336	0.189	0.127(4.5)	0.045
	상	37.08	11.29	0.192	0.620	0.204	0.120	0.171(5.2)	0.020
	하	42.14	15.64	0.361	0.433	0.276	0.251	0.129(4.9)	0.034
중 학 교	전체	50.60	23.55	0.858	0.345	0.329	0.220	0.104(3.3)	0.012
	상	34.33	16.75	0.276	0.530	0.211	0.230	0.160(4.1)	0.024
	하	40.65	20.41	0.492	0.366	0.271	0.271	0.087(2.5)	0.018

연결망은 기본적으로 연결 여부에 따라 관측되므로, 연결망의 구성원들이 얼마나 긴밀하게 연결되었는지를 측정하고 이를 바탕으로 서로 다른 연결망의 구조적 특성을 비교할 수 있다. 평균 연결 정도는 연결망 내에 존재하는 총 연결 수를 연결망에 속해있는 총 노드 수로 나눈 값이다. 이는 텍스트의 각 단어들 간에 얼마나 많은 연결이 생성되었는지와 관련되는데, <표 5>를 보면 중학교에 비해 고등학교의 연결 정도가 유의하게 높음을 알 수 있다( $p>0.05$ ). 그러나 각 학교급별 상, 하 수준을 살펴보면 하 집단의 연결 정

도가 상 집단에 비해 더 높게 나타났다. 즉, 하 집단의 글에서 단어들 간의 연결성이 더 높았던 것인데, <표 4>에서 하 집단의 연결망 규모가 더 작다는 점을 고려할 때, 이는 하 집단의 학생들이 상 집단에 비해 더 짧은 글에서 같은 단어를 많이 씀으로써 단어 간 연결의 중복이 더 많이 이루어졌다고 해석할 수 있다. 이는 결과적으로 하 집단 학생들의 글이 상 집단에 비해 유사한 내용으로 전개됨을 의미한다.

밀도는 하나의 텍스트를 구성하는 모든 단어들 간에 완벽한 연결이 이루어질 수 있는 가능성에 비해 실제 연결되어 있는 비율과 관련된다. 그런데, 실제 연결망 내 단어들의 연결 관계는 연결망 전체의 규모에 영향을 받는다. <표 4>에서 확인했듯이, 상 집단의 글들이 하 집단보다 평균적으로 보다 많은 단어를 사용하므로 연결망에서 가능한 연결의 수가 자체가 늘어나게 된다. 따라서 연결망의 규모가 커질수록 밀도는 작아지는 경향이 있다(김용학·김영진, 2019). <표 5>를 보면 각 학교급별 상, 하 집단의 연결 정도와 밀도는 비례 관계를 알 수 있다. 즉, 상 집단은 더 큰 규모의 연결망을 형성하지만 글의 내용이 다양하여 단어들 간의 연결 정도는 더 낮고 그에 따라 연결망의 밀도도 낮다고 볼 수 있다. 반대로 하 집단의 연결 정도와 밀도가 높은 것은 유사한 단어들이 반복적으로 사용되어 글 내용의 다양성이 덜하다는 점을 드러낸다.

이러한 가설은 연결망의 군집성을 드러내는 지표를 통해서도 확인할 수 있다. 이행성(transitivity)과 모듈래리티(modularity)는 연결망의 군집화 정도를 나타내는 기능을 한다. 이행성이란 삼자관계(triad)를 기반으로 산출되는 지표로 해당 연결망이 무작위 연결망에 비해 얼마나 조밀한 구조를 갖는지를 보여준다. 연결망 내에 이행성을 충족하는 삼자관계가 많을수록 그 연결망은 ‘덩어리 구조’를 갖는 경향이 있으며, 노드들이 서로 연결된 하위 집단을 많이 갖게 된다(곽기영, 2019). 본 연구에서 각 학교급, 쓰기 수준별로 이행성은 유의한 차이가 나타나지 않았다. 반면 모듈래리티에서는 상, 하에서 유의한 차이가 발견되었다. 모듈래리티는 대규모 연결망 분석에서 집단을 탐색하는 데 쓰이는 기법 중 하나로, 다양한 학문 분야에서 광범위하게

쓰이는 군집화 방식이다(Newman, 2006). 모듈래리티란 연결망 내 밀도 차이를 통해 집단을 구분하며, 지표값이 높을수록 의미 분화의 정도가 크다. <표 5>를 보면 학교급별로는 중학교보다는 고등학교가, 수준별로는 하 집단에 비해 상 집단의 더 모듈래리티 수치가 높다. 괄호 안의 숫자는 모듈래리티의 방식으로 얻어진 연결망 내의 하위 집단의 수의 평균이다. 이를 통해 볼 때, 고등학생이 중학생에 비해 의미적으로 보다 다양하게 분화된 글을 쓴다고 추론할 수 있으며, 특히 고등학교 상 집단은 평균 5.2개의 집단이 형성되는 것으로 보아 학생들이 쓴 글 내에 다양한 근거와 사례가 포함되어 있음을 추론해 볼 수 있다.

#### 4. 집단 중심성 분석

집단 중심성이란 연결망 내의 개별 단어들이 중심성에 있어서 얼마나 다른지를 나타내는 지표이다. 연결망의 중심성을 개별 단어 수준에서 접근하면 해당 연결망에서 중요한 역할을 하는 단어들을 파악할 수 있으나 이를 통해 연결망 전체의 단어들이 갖는 중요도나 의미 분화의 정도와 같은 구조적 특성을 확인하기는 어렵다. 집단 중심성을 확인하면 각 중심성별로 단어들의 지위나 역할에서 얼마나 차이가 있는지를 통해 전체 연결망의 구조를 이해할 수 있다.

<표 6> 집단 중심성 분석

학년	집단	집단 연결 중심성 Degree centrality		집단 위세 중심성 Eigenvector centrality		집단 매개 중심성 Betweenness centrality	
		평균	표준 편차	평균	표준 편차	평균	표준 편차
고 등 학 교	전체	0.029	0.011	102.62	3.44	16.58	5.62
	상	0.021	0.012	102.68	2.53	25.16	3.21
	하	0.034	0.023	102.35	4.55	24.65	6.52

중 학 교	전체	0.041	0.028	108.05	4.75	22.99	7.52
	상	0.032	0.016	106.92	3.56	38.63	6.12
	하	0.044	0.037	108.20	5.66	28.31	7.99

〈표 6〉의 집단 연결 중심성의 경우, 고등학교보다는 중학교가, 상 집단에 비해서는 하 집단이 높음을 확인할 수 있다. 즉, 고등학교보다는 중학교 집단이, 상 집단보다는 하 집단의 글에서 개별 단어들의 연결성 차이가 더 크다는 것을 알 수 있다. 연결 중심성이란 노드와 노드가 연결된 정도로 이 수치가 크다면 해당 노드는 글 전체에서 다양한 단어들과 관련됨을 의미한다. 따라서 집단 연결 중심성이 크다는 것은 개별 단어들 간 의미적 중요도의 차이가 크다는 점을 드러낸다. 다만, 연결 중심성은 연결된 노드의 개수에만 초점이 있기 때문에 국지적이라는 한계가 있다. 이를 연결망 전체로 확장한 개념이 위세 중심성 또는 아이겐벡터중심성이다(길호연, 2020). 이는 연결된 노드의 개수뿐만 아니라 그 노드의 중요도를 함께 고려함으로써 연결망에서 어느 정도 영향력 있는 노드인지를 알려주는 기능을 한다. 위세 중심성에 초점을 두고 보면 고등학교 상, 하 수준은 큰 차이가 없으며 중학교의 경우는 하 수준의 글이 상 수준보다 중심성 수치가 유의하게 더 높았다. 이는 중학교 하 수준 글이 상 수준에 비해 글에 쓰인 단어의 영향력에서 큰 차이가 있음을 드러낸다. 앞서 〈표 4〉와 〈표 5〉를 통해 확인하였다시피 중학교 하 수준의 글들은 연결망의 규모가 작고 밀도가 높아 유사한 단어가 반복되거나 의미 분화의 정도가 약한 특성이 있다. 이를 집단 중심성 수치와 함께 비교해 보면, 이 집단의 글들은 의미 분화가 덜하여 글의 주제와 관련된 핵심 단어들과 그 외의 기능을 하는 단어들의 의미적 중요도의 차이가 크고, 글의 구조도 중심 내용과 세부 내용의 직접적 연결로 이루어진 단순한 구조일 것이라고 추론해 볼 수 있다.

한편, 집단 매개 중심성은 조금 다른 양상을 보였다. 개별 노드 수준에



서 매개 중심성이란 직접적으로 연결되지 않은 노드들 간의 관계를 통제 또는 중개하는 정도를 의미한다(곽기영, 2019). 이렇게 볼 때, 매개 중심성이 높은 단어는 텍스트 전체에서 의미 단위를 연결해주는 기능을 한다고 볼 수 있으며, 집단 매개 중심성이 높다는 것은 단어들 간의 매개 역할의 차이가 크다는 것을 말한다. <표 6>을 보면 고등학교 전체 집단에 비해서는 중학교 전체 집단의 매개 중심성이 높았으나 상, 하 수준에서는 상 집단이 하 집단에 비해 매개 중심성이 유의하게 높았다. 이는 상 집단 글에는 하 집단에 비해 의미 매개에 있어 중요한 역할을 하는 단어들이 더 많았다는 점을 드러낸다. 이를 위세 중심성 수치와 비교해서 해석해 보면, 각 학교급의 상위 집단은 하위 집단에 비해 개별 단어들의 의미적 영향력의 차이는 적었으나, 단어의 연결에 있어서 핵심적인 기능을 하는 단어들이 존재함을 말해준다. 이러한 차이는 텍스트 내의 의미의 분화 정도와 관련이 있다. 모듈래리티를 통해 확인한 결과 상위 집단의 글이 하위 집단에 비해 의미의 분화 정도가 더 높았는데, 이 경우 글에서 유사한 영향력을 지닌 단어의 개수가 늘어나 집단 위세중심성의 지표는 낮아지고(대등한 위세 중심성을 지닌 단어들이 늘어남에 따라), 각각의 의미 단위를 연결하는 단어들이 등장함에 따라 집단 매개중심성 지표가 높아진다고 추론해 볼 수 있다.

## 5. 의미연결망 패턴 분석

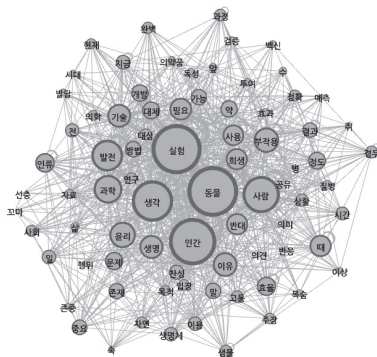
연결망의 연결 정도나 밀도, 중심성 지표 등을 통해 학생들의 글로 구성된 연결망의 대략적인 특성을 파악할 수는 있으나 실제 다양한 연결의 특성을 직관적으로 파악하기에는 무리가 있다. 따라서 이 연구에서는 집단별 연결망 그래프와 개별 글의 연결망 그래프를 확인해 보았다.

연결망 레이아웃은 Gephi에서 제공하는 스프링 임베딩 레이아웃인 ForceAtlas2를 사용하였고, 노드의 크기는 위세 중심성의 정도로, 색은 모듈래리티(Modularity)의 클래스로 표시하였다. 앞서 설명한 바와 같이 일종의

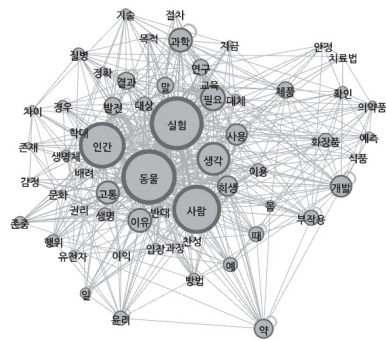
군집 분석이라고 볼 수 있는 모듈래리티를 활용하면 서로 의미 연결성이 높은 단어들의 집단을 파악함으로써 네트워크 전체의 의미 분화 양상을 살필 수 있다.

1) 집단별 의미연결망 패턴 비교

개별 글의 연결망 그래프를 확인하기에 앞서 쓰기 단계와 수준별로 집단별 그래프에 어떤 차이가 있는지 확인하고자 하였다. 집단별 그래프는 중학교와 고등학교 전체 집단과 학교급별 상, 하 수준의 글들을 하나로 통합하고, 전체 집단은 출현 빈도 12회 이상의 단어로, 상, 하 수준별 집단은 출현 빈도 6회 이상의 단어들을 추출하여 해당 단어들이 텍스트 내에서 지니는 관계적 속성을 시각화하였다. 전체 집단과 수준별 집단의 단어 빈도수를 다르게 설정한 이유는 전체 노드의 수를 일정하게 유지하여 그래프의 관계적 속성을 보다 잘 드러내기 위한 것이다.



고등학교 전체 집단 연결망  
node: 80, Edge: 979



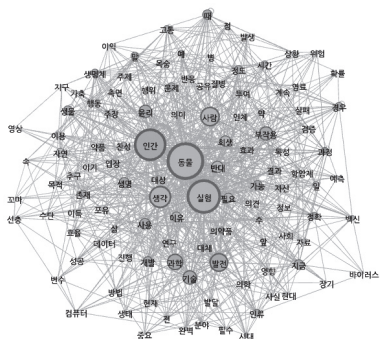
중학교 전체 집단 연결망  
node: 60, Edge: 517

〈그림 1〉 학교급별 의미연결망 패턴

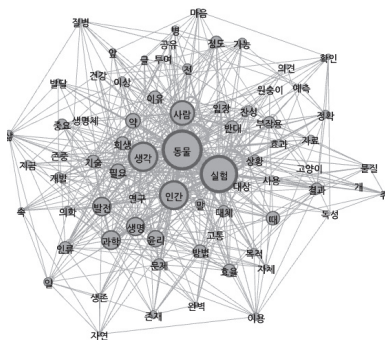
〈그림 1〉은 고등학교와 중학교 집단 전체의 의미연결망을 시각화한 것이다. 우선 중학교가 고등학교에 비해 연결선들이 중앙에 집중되어 있음

알 수 있다. 이는 중학생 글들이 중심 내용에 대한 집중도가 높고 그에 비해 세부 내용들의 연결성이 약함을 보여준다. 이는 연결선의 패턴과도 관련되는데, 고등학교는 중앙에서 주변으로 연결선들이 조밀하게 연결되어 점차 확장되어 나가는 구조를 보이는 데 반해, 중학교는 중앙과 그 외 지역의 밀도 차이가 크고 연결망의 우측으로 중심 내용과 관련성이 약한 별개의 의미 구조가 형성됨을 확인할 수 있다. 예를 들어 중학교 연결망에서 ‘제품’, ‘화장품’, ‘부작용’ 등의 단어를 제거할 경우, 연결망이 좌우로 와해될 가능성이 높다. 이는 의미 단위 간 연결성이 촘촘하지 못함을 보여주며, 텍스트의 응집성이 약하는 증거이다. 연결망에서 단어들 간의 거리가 멀다는 것은 직접적인 연결이 없다는 것을 의미함을 고려할 때, 중학교에서는 주제와 관련성이 약한 의미 구조가 발견됨을 확인할 수 있다.

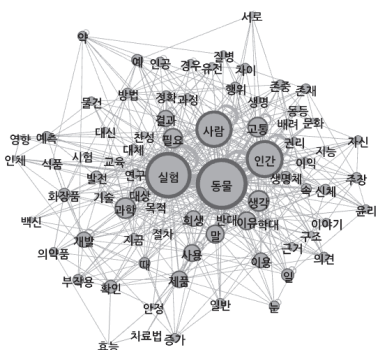
각 학교급별 상, 하 수준의 연결망의 차이에서도 유사한 특성을 발견할 수 있다. 먼저 <그림 2>의 고등학교 상, 하 수준 연결망을 살펴보면, 상 수준 텍스트는 하 수준에 비해 더 큰 연결망을 구성하고 있으며, 중앙에서 주변까지 매우 촘촘한 연결이 이루어져 있다. ‘동물’, ‘실험’, ‘사람’, ‘생각’ 등 중심성이 높은 단어들이 중앙의 고밀도 부분을 이루고 있으며 그 주변을 ‘기술’, ‘발전’, ‘윤리’, ‘생각’ 등의 단어들이 감싸는 구조가 드러난다. 이는 글의 논제와 관련된 주제어들과 핵심 논거들의 관계가 일차적으로 구성되고 이후 하위 근거와 관련된 단어들이 순차적으로 배열되는 구조임을 보여주는 것이다. 반면 하위 집단에서는 이러한 연결의 순차성이 잘 드러나지 않으며, 모듈래리티(노드의 색)를 통해 확인된 하위 집단 역시 좌우로 구분되어 의미 단위의 연결성이 약함을 알 수 있다. 이는 글의 핵심 주장이 하위 주장과 그에 따른 논거들로 체계화되지 못하고, 몇 가지 주요 화제들이 단순하게 나열되는 구조를 취하고 있음을 추론해 볼 수 있다.



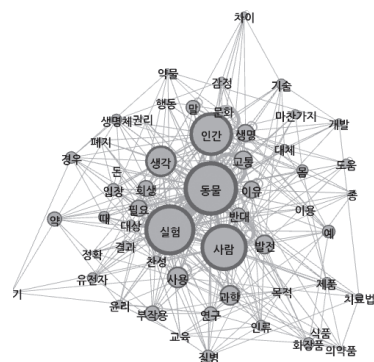
고등학교 상위 집단 연결망  
node: 115, edge: 991



고등학교 하위 집단 연결망  
node: 72, edge: 563



중학교 상위 집단 연결망  
node: 79, edge: 492



중학교 하위 집단 연결망  
node: 55, edge: 344

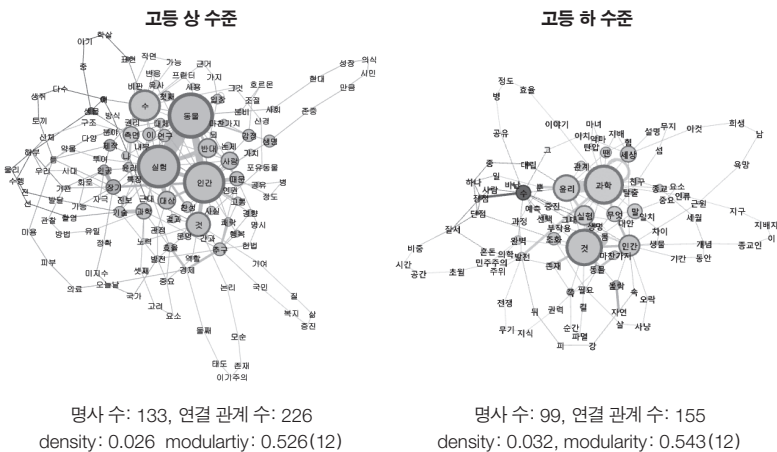
### 〈그림 2〉 쓰기 수준별 집단별 의미연결망

〈그림 2〉의 고등학교 집단별 그래프를 중학교 그래프와 비교해보면, 〈그림 1〉에서 확인했던 연결망 차이가 보다 명확히 드러난다. 즉, 〈그림 2〉의 중학교 상, 하 집단별 연결망은 고등학교 상, 하 집단 그래프에 비해 원형으로 퍼져나가는 구조를 덜 갖추고 있음을 확인할 수 있다. 그래도 상위 집단에서는 중앙부의 고밀도 핵심어에서 주변의 저밀도 단어로 약하게 뻗어나가는 패턴이 확인되나, 하위 집단은 고밀도 핵심부를 제외하면 나머지 부분의 연결 관계의 밀도가 높지 못하며 의미 단위 역시 크게 좌우로 구분된 형태를 취하고 있다.

특히 하위 집단에서는 연결망의 외곽에 튀어나온 노드가 여러 개 발견되는데 예를 들어 ‘차이’, ‘장기’, ‘의약품’ 등의 단어가 그러하다. 이 단어들은 특정한 글에서만 사용되었거나 다른 단어들과 밀접한 관련을 맺지 못하고 연결 관계가 약하게 사용된 단어들이라고 볼 수 있다. 상위 집단의 연결망에서 그러한 패턴을 보이는 노드가 적다는 점을 고려할 때, 이는 글을 구성할 때 관련 단어들이 글의 핵심 주제와 관련성이 높지 않은 방식으로 사용되었다고 볼 수 있다.

## 2) 개별 글의 의미연결망 패턴 비교

집단별 그래프를 통해 쓰기 발달 수준별로 전반적인 의미연결망의 패턴을 살펴보았다면 실제 학생 글의 개별적인 의미연결망 양상을 확인해 볼 필요가 있다. 본 연구에서는 지면 관계상 관계망의 차이가 잘 드러날 수 있도록 중학생과 고등학생 최상위 글과 하위 30% 글의 의미연결망을 시각화하여 비교하였다. 앞서 살펴본 집단별 그래프에서는 일정 빈도수 이상의 단어들만 추출하여 시각화하였으나, 개별 글 그래프는 단어 수가 많지 않아 고빈도 단어로 한정하지 않고 각 글을 구성하고 있는 단어 전체의 연결망을 시각화하였다. 먼저 <그림 3>은 고등학교 상 수준과 하 수준 학생 글의 의미연결망 시각화 결과이다.

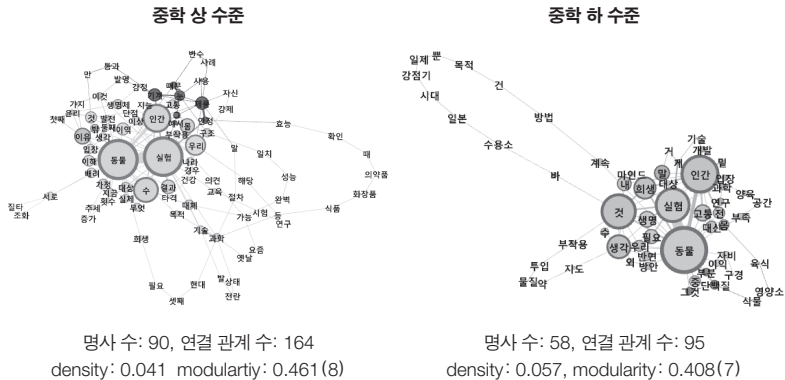


<그림 3> 고등학생 개별 글 의미연결망

앞서 통계적 지표를 통해 추론한 글의 구조가 연결망 그래프에 나타나지를 확인해 볼 수 있다. 먼저 상 수준의 글은 하 수준에 비해 노드의 수가 더 많고 전반적으로 노드 간 거리가 가깝다. 또한 중앙의 핵심어들을 중심으로 연결선이 집중되어 있기는 하지만 핵심부를 둘러싼 주변부에도 꽤 많은 연결 관계가 포진되어 있어 글의 핵심 내용과 세부 내용 사이를 연결하는 다양한 단어들이 존재함을 알 수 있다. 중앙부의 ‘동물’이나 ‘실험’, ‘인간’ 사이의 선이 굵게 표시된 것은 이 단어들 간의 동시 출현 빈도가 높음을 의미한다. 즉 이 단어들은 글 전체에서 반복적으로 등장하며 글의 주제나 필자의 입장을 드러내는 데 활용되었음을 알 수 있다. 그러나 이 단어들 외에도 ‘반대’, ‘사람’, ‘대상’, ‘장기’ 등의 단어들에도 굵은 선이 연결되어 있으며, 이들 단어를 통해 다양한 단어들이 중심부와 연결되고 있다. 이렇게 중심부에서 주변부로 연결이 확장되어 간다는 것은 글의 내용이 순차적으로 전개되고 내용 간 응집성이 높다는 것을 의미한다. 이러한 구조의 글은 의미 분화의 가능성도 높아지는데, 그림에서 다양한 색으로 표현된 모듈래티는 연결망의 의미 분화 수준이 높음을 드러낸다. 물론 상 수준과 하 수준 글의 모듈래티가 수치상 유사하며, 의미 범주의 개수도 동일하게 나타났으나 실제 상 수준 글에서는 중심성이 높은 단어들, 즉 노드의 크기가 큰 단어 수준에서 의미 분화가 일어나 독자가 체감하는 글 내용의 다양성은 하 수준에 비해 더 높다고 볼 수 있다.

반면 하 수준의 글은 상 수준에 비해 노드의 수도 적고 연결망의 규모도 작다. 특히 중앙의 핵심부를 제외하면 노드 간 연결 관계가 활발하지 않아, 전체적으로 연결망의 밀도 차이가 크다는 것을 알 수 있다. 앞선 구조 분석에서는 하 수준의 글이 상 수준에 비해 연결망의 연결 정도가 높고 밀도가 높은 것은 유사한 단어들이 반복되면서 단어 간 연결 정도가 높아지기 때문이라고 해석하였다. <그림 3>의 하 수준 그래프를 보면 연결망 내 대부분의 단어들이 중앙의 몇몇 핵심어들과 직접적인 연결 관계를 갖고 있어 핵심어의 연결 정도가 매우 높다는 것을 알 수 있다. 이는 앞서 살펴본 집단 중심

성의 차이로도 설명이 가능한데, 글에서 중요한 역할을 담당하는 단어들이 쓰기 과제와 관련된 몇몇 핵심어로 한정되어 있는 것이다. 외곽으로 길게 튀어나온 연결선은 다른 단어들과 긴밀한 연결을 맺지 못하는 개별적 의미 단위를 나타낸다. 하 수준 연결망에서는 상 수준에 비해 이러한 연결선의 수가 더 많은데, 이는 해당 단어들이 다른 단어들과 안정적인 관계를 유지하지 못함을 의미한다. 연결망 분석에서는 두 노드 간의 효율적 경로(최단경로)보다는 두 노드를 연결하는 연결 정도가 얼마나 많은지가 더 중요하다(곽기영, 2019). 이러한 양상은 글에서 여러 하위 내용들이 체계적으로 조직되지 못하고 주제와 직접 관련된 단편적 이유나 사례 등이 반복적으로 제시되고 있음을 드러낸다고 볼 수 있다.



〈그림 4〉 중학생 개별 글 의미연결망

〈그림 4〉의 중학생 개별 글 연결망은 전반적으로 〈그림 3〉의 고등학생 글 연결망에 비해 노드나 연결선의 수가 적게 형성되어 연결망의 규모가 작음을 알 수 있다. 특히 중앙의 핵심어 간의 연결 관계를 제외하면 연결선 자체가 드물게 생성되어 단어들 간의 연결이 활발하지 못함을 알 수 있다. 전반적인 단어 수가 적다 보니 중앙 핵심부의 연결성이 전체 연결망의 연결 정도와 밀도에 많은 영향을 끼치게 된다. 이는 〈표 5〉의 연결망의 구조 지표에



서 중학교 집단의 연결 정도와 밀도가 고등학교에 비해 높았던 점과 관련된다. 의미 분화를 나타내는 모듈래리티 역시 고등학교 집단에 비해 낮은 편이고 의미 범주의 수도 더 적게 형성되고 있음을 알 수 있다. 다만 상 수준의 그래프에는 외곽으로 뻗어가는 다양한 연결선이 존재하며 이들 연결선이 핵심부의 여러 노드와 연결되어 있는 데 반해 하 수준의 연결망에서는 이러한 연결선이 2~3개에 불과하다. 이는 중학교 하 수준의 글이 단일한 내용으로 전개되고 여러 하위 내용을 갖지 못함을 드러낸다.

### 6. 개별 글의 단어 중심성 분석

연결망의 패턴을 개별 단어의 중심성과 연관지어 보면 글의 특성을 보다 깊이 이해할 수 있다. <표 7>은 앞서 살펴본 고등학생 글의 나타난 주요 단어들의 중심성 수치이다.

<표 7> 고등학생 글에 사용된 단어의 중심성

	위세 중심성				매개 중심성			
	상 수준		하 수준		상 수준		하 수준	
1	실험	0.59	과학	0.58	인간	36.95	과학	36.83
2	동물	0.58	윤리	0.45	동물	33.16	인간	32.46
3	인간	0.30	실험	0.22	실험	23.92	윤리	22.03
4	반대	0.22	무엇	0.22	추구	15.98	발전	15.55
5	대상	0.20	인간	0.14	장기	14.15	세상	13.61
6	때문	0.10	세상	0.14	과학	9.52	무엇	10.67
7	찬성	0.09	발전	0.14	기술	9.10	실험	10.22
8	측면	0.09	관계	0.11	해부	8.71	조화	9.29
9	사람	0.09	조화	0.11	생물	8.37	개념	8.08



10	장기	0.07	종교	0.07	경제	7.99	동물	7.18
11	과학	0.07	친구	0.07	제작	6.74	완벽	6.83
12	기술	0.06	탄압	0.06	발전	6.73	장점	6.41
13	감정	0.06	악마	0.06	측면	6.68	필요	5.83
14	생명	0.06	이야기	0.06	때문	6.64	질서	5.73
15	인공	0.06	인류	0.06	인공	6.41	단점	5.59

〈표 7〉을 보면, 상 수준 글에서 글의 논제와 관련된 ‘동물’, ‘실험’, ‘인간’ 등을 제외하면 ‘반대’, ‘찬성’과 같은 입장과 관련된 단어, ‘때문’이나 ‘측면’ 등 어떠한 이유나 관점을 드러내는 단어의 위세 중심성이 높다는 점을 알 수 있다. 이는 이 글이 필자의 관점이나 주장의 이유를 전개하는 방식으로 구조화되어 있음을 보여준다. 매개 중심성에서 주목할 만한 점은 위세 중심성과 달리 ‘추구’의 순위가 높다는 점인데, 이는 특정한 입장과 관련하여 ‘추구’라는 단어가 다른 단어와의 매개 역할을 한다는 점을 드러낸다. 아래는 이러한 단어들이 사용된 글의 일부이다.

“나는 ‘과학 연구를 위해 동물실험을 해야 한다’라는 논제에 대해 반대의 입장이다. 이러한 입장에 대해서 세 가지의 근거를 들 수 있다. 사람은 고통은 피하고 쾌락을 추구하는 경향이 있다.”(고등 상 수준 글)

반면 하위 30%의 글에서는 최상위 글과 달리 위세 중심성에 필자의 입장이나 이유, 관점과 관련되는 단어의 순위가 높지 않음을 알 수 있다. 상 수준 글에서 볼 수 없는 단어로 ‘윤리’, ‘무엇’, ‘세상’, ‘발전’ 등이 있는데, 이들 단어는 논제와 관련된 필자의 논증 자체와 관련된 단어들이다. 즉, 하 수준 글의 위세 중심성에서는 필자의 주장, 관점을 독자에게 제시하는 어휘들이 중요한 역할을 하지 못하고 필자의 논증을 반복적으로 드러내는 단어들의 수치가 높음을 알 수 있다. 하 수준 글의 매개 중심성 역시 위세 중심성과

단어의 순위에 큰 차이가 없으나 순위가 낮은 단어 중에는 ‘장점’, ‘단점’과 같이 필자의 관점을 형성하는 데 활용된 단어들이 일부 나타남을 확인할 수 있다.

〈표 8〉 중학생 글에 사용된 단어의 중심성

	위세 중심성				매개 중심성			
	상 수준		하 수준		상 수준		하 수준	
1	동물	0.62	동물	0.59	실험	47.27	동물	56.76
2	실험	0.62	실험	0.57	동물	37.52	인간	30.63
3	인간	0.21	인간	0.27	인간	25.52	실험	23.07
4	결과	0.17	희생	0.18	과학	24.03	계속	16.13
5	기계	0.10	생명	0.17	제품	17.15	방법	16.13
6	우리	0.10	필요	0.16	기술	14.39	목적	16.13
7	이유	0.09	고통	0.12	우리	13.80	일제	16.13
8	안정	0.09	생각	0.12	안정	10.44	강점기	16.13
9	생각	0.09	몸	0.10	가능	6.61	시대	16.13
10	대상	0.08	대신	0.07	눈	6.36	일본	16.13
11	가정	0.08	우리	0.07	몸	5.72	수용소	16.13
12	몸	0.07	반면	0.06	기계	5.58	생각	12.70
13	이익	0.06	단백질	0.06	효능	5.49	단백질	8.49
14	이해	0.06	과학	0.06	확인	5.49	지도	6.83
15	배려	0.06	그것	0.06	때	5.49	약	6.83

〈표 8〉에서 중학생 글에 사용된 개별 단어들의 중심성을 살펴보면 상 수준 글에는 하 수준에서 볼 수 없는 ‘이유’, ‘가정’, ‘이해’와 같이 필자가 글의 내용을 전개하기 위해 사용한 단어들의 위세 중심성 수치가 높음을 알 수 있다. 이들 단어는 앞서 고등학생 상 수준의 글에서 살핀 바와 같이 필자가 자

신의 주장을 전개하거나 근거를 세우는 과정에서 사용된 단어들이다.

“첫째, 윤리적인 이유이다.”(중학 상 수준 글)

“만약 동물이 인간보다 지능이 높아 ‘인간 실험’을 한다고 가정해 보자.”(중학 상 수준 글)

“만약 그렇다면 우리는 동물들을 이해할 수 있을까?”(중학 상 수준 글)

이에 비해 하 수준의 글에는 그러한 기능을 하는 단어를 찾기 어렵고 상 수준에 없는 ‘희생’, ‘단백질’, ‘과학’ 등의 단어들은 모두 필자의 주장 내용과 직접 관련된 단어들이다. Bereiter & Scardamalia(1987)는 능숙한 필자와 미숙한 필자가 쓰기 과정에서 사용한 전략의 차이를 구체화하면서, 미숙한 필자는 화제에 대해 알고 있는 것을 단순히 나열하는 방식으로 쓰기에 접근하는 ‘지식 서술 모형’을 사용하고, 능숙한 필자는 지속적으로 문제를 해결하고 지식을 변형시키면서 글에 접근하는 ‘지식 변형 모형’을 활용한다고 하였다. <표 8>의 결과를 놓고 볼 때, 위세 중심성이 높은 단어들의 양상은 이러한 전략의 차이와도 관련이 있어 보인다. 즉, 상 수준의 글은 내용 전개 과정에서 다양한 단어를 통해 독자가 글의 내용을 받아들일 수 있도록 구체화한다고 볼 수 있다. 이는 필자가 ‘누구에게 무엇을 어떻게 말할 것인지’를 분석하고 글에 반영한 증거라고 볼 수 있다. 반면 하 수준의 글에서 중심성이 높은 단어들은 화제와 관련된 필자의 지식 자체와 관련된 단어들을 중심으로 구성되어 있다는 점을 확인할 수 있다. 이러한 차이는 단지 글에 쓰인 단어의 빈도 분석으로는 파악하기 어렵고 의미연결망 분석을 통해 중심성이 높은 단어를 확인함으로써 파악 가능한 정보라고 볼 수 있다.

매개 중심성의 경우 하 수준의 글은 쓰기 과제인 ‘동물’의 매개 중심성이 매우 높은 반면 다른 단어들의 중심성이 상 수준에 비해 눈에 띄게 낮은 수준이다. 이는 하 수준의 글에서 의미 매개 역할을 하는 단어가 적다는 것을 의미한다. 특히 하 수준의 글에는 매개 중심성 수치가 같은 단어들이 많

은데, 4위인 ‘계속’부터 11위인 ‘수용소’까지의 단어들이 모두 16.33으로 동일하다. 실제 이들 단어는 모두 <그림 4>에서 좌측 상단으로 길게 뻗어 간 연결선에 걸쳐 있는 단어들로서, 핵심부의 다른 단어들과는 밀접한 연결 관계를 맺지 못하고 별도의 의미 구조 내에서 서로를 매개하는 단어들임을 확인할 수 있다. 이는 하 수준의 글에서 매개 연결성이 높은 단어가 드물며, 글의 다양한 의미 구조를 연결하는 단어가 부재함을 드러내 준다.

## V. 결론

이 연구에서는 의미연결망 분석을 통해 쓰기 발달 과정에 있는 중등 학생 필자의 글 특성을 판별하고 이를 객관화함으로써 학생들의 쓰기 특성에 대해 판단할 수 있는 정보를 탐색하고자 하였다. 이는 크게 학생 글에 쓰인 개별 단어의 특성과 단어들 간의 관계적 특성으로 구분될 수 있다. 개별 단어 특성으로는 의미 연결망의 크기를 결정하는 단어 수와 각 단어의 중심성 수치 등이 있으며 단어 간 관계적 정보에는 의미연결망의 구조 지표, 집단 중심성 수치, 의미연결망의 패턴 등을 들 수 있다. 이러한 변인을 중심으로 학생 글 특성을 비교한 결과는 다음과 같다.

먼저 글의 의미를 구성하기 위해 활용한 단어(명사)의 수는 중학생보다는 고등학생 집단이, 하 수준보다는 상 수준이 많았으며 학교급 차이보다는 학교급 내의 쓰기 수준별 차이가 더 크게 나타났다. 어휘 다양성을 의미하는 빈도 반영 명사 수의 경우 고등학생 집단이 중학생 집단보다 유의하게 많았으나 상 수준이 하 수준에 비해서는 적은 양상을 보여주었는데 이는 쓰기 수준이 높은 학생들이 보다 적은 수의 단어로 더 긴 글을 쓰는 능력이 있음을 보여준다.

둘째, 단어들의 연결 관계를 통해 의미연결망의 구조를 확인해 본 결과,

전체적인 단어 간 연결 정도는 고등학생 집단이 중학생에 비해 높았으나 쓰기 수준별로 보면 상 수준이 하 수준에 비해 연결 정도와 밀도가 더 낮은 것으로 나타났다. 연결망의 규모를 함께 고려하면, 이러한 결과는 하 수준 학생들의 글에서 유사한 단어들이 반복적으로 사용되었음을 의미한다고 볼 수 있다. 또한 밀도 차이에 기반한 군집 방식인 모듈래리티 지표에 따르면 학교급과 쓰기 수행 수준이 높아질수록 글의 의미 분화의 정도가 커졌음을 확인하였다. 의미 분화 양상은 집단 중심성 수치와도 관련된다. 연결 중심성이나 위세 중심성의 경우, 전반적으로 고등학생보다는 중학생 집단이, 상 수준보다는 하 수준의 글에서 중심성 차이가 컸는데 이는 의미 분화가 덜한 글일수록 핵심어와 그 밖에 단어들 간 기능 차이가 크다는 것을 말해준다. 반대로 집단 매개 중심성은 상 수준이 하 수준에 비해 높았는데 이는 의미 분화 정도가 높은 상 수준 학생들의 글에서 의미 매개의 역할을 하는 단어들의 기능이 두드러졌음을 나타낸다.

셋째, 의미연결망 그래프의 패턴 분석 결과 학교급과 쓰기 수준에 높아질수록 중앙에서 주변으로 보다 촘촘한 연결을 보여주며 주제어와 핵심 논거, 주변 논거들이 순차적으로 배열되는 구조를 보여주었다. 반면 쓰기 수준이 낮은 글에서는 핵심부와 주변부의 밀도 차이가 크고 연결망의 전반적인 연결 수준이 고르지 않아 의미 군집이 크게 분할되거나, 다른 단어들과 긴밀한 관계를 맺지 못하고 돌출되는 연결선이 보이는 등 응집력이 약한 의미망 패턴이 나타남을 확인할 수 있었다.

넷째, 개별 글의 단어 중심성을 분석한 결과 상 수준의 글에서는 하 수준과 달리 쓰기 주제와 관련된 내용어뿐만 아니라 필자의 관점을 드러내거나 논리를 전개하는 데 활용된 단어들의 위세 중심성과 매개 중심성 수치가 높게 나타났다. 이는 능숙한 필자일수록 자신의 지식을 변형하여 독자가 글의 내용을 받아들일 수 있도록 유도하는 지식 변형적 글쓰기를 하는 경향과 관련된다고 볼 수 있다.

이 연구는 쓰기 발달의 양상을 다루고자 하였으나 표집의 수가 작아 일

반화에 한계가 있다. 이는 개별 글 각각에 대한 연결망을 모두 분석해 내는데 물리적 한계가 있었기 때문이다. 또한 분석 대상을 명사로 한정된 점 역시 연구의 제한점으로 볼 수 있다. 추후 보다 많은 텍스트를 대상으로, 명사 뿐만 아니라 동사나 형용사 등 다양한 내용어를 포함한 연결망 분석 연구가 이루어질 필요가 있다.

이러한 한계에도 불구하고 이 연구는 의미연결망을 활용하여 중등 학생 필자의 쓰기 특성을 실증적으로 분석하고자 했다는 점에서 의의가 있다. 특히 그간 국어교육에서 잘 다루지 않았던 모듈래리티 개념을 통해 의미 군집을 분석했다는 점, 집단 중심성 개념을 통해 개별 글뿐만 아니라 중, 고등학생 집단별 의미연결망 특성을 함께 분석했다는 점 등에서 의의를 찾을 수 있다.

학생 작문에 대한 의미연결망 연구는 선형적인 글 읽기에 기반한 쓰기 평가에서는 알아내기 어려웠던 글의 특성을 입체적, 가시적으로 드러내 줄 수 있다는 장점이 있다. 이런 의미에서 본 연구는 쓰기 평가의 평가자 신뢰도 보완 방안, 학생 글에 대한 자동 피드백 방안 연구 등에서 다양하게 활용될 수 있다. 향후 의미연결망 분석을 통해 도출되는 다양한 통계적 지표들을 통해 그동안 발견하기 어려웠던 글의 특성을 규명하고, 보다 효과적인 쓰기 피드백 방안이 마련되기를 기대한다.

\* 본 논문은 2022.07.29. 투고되었으며, 2022.08.07. 심사가 시작되어 2022.09.13. 심사가 종료되었음.

## 참고문헌

- 가은아(2011), 「쓰기 발달의 양상과 특성 연구」, 한국교원대학교 박사학위논문.
- 곽기영(2019), 『소셜 네트워크 분석』, 서울: 도서출판 청람.
- 교육부(2015), 『국어과 교육과정(제2015-74호)』, 세종: 교육부.
- 길호현(2020), 「읽기 지문 텍스트의 주제 중심성 유형 연구」, 『청람어문교육』 74, 39-70.
- 김용학·김영진(2019), 『사회 연결망 분석』, 서울: 박영사.
- 김영옥(2015), 「쓰기 유창성 요소와 중·고등학생 설명문 쓰기 능력의 상관 분석」, 한국교원대학교 석사학위논문.
- 김우주(2015), 『네트워크 중심성 이론』, 서울: 카오스북.
- 김희림·전익기(2018), 「텍스트 마이닝 기법을 활용한 여가활동 키워드 분석」, 『한국여가레크레이션학회지』 42(3), 59-69.
- 박영민·가은아(2009), 「중학생의 쓰기 동기 수준 및 성별에 따른 쓰기 수행의 차이」, 『국어교육학연구』 35, 241-269.
- 박영민·이재기·이수진·박종임·박찬홍(2018), 『작문교육론』, 서울: 역락.
- 박종임(2012), 「사고구조를 활용한 쓰기 평가 특성 분석」, 『작문연구』 16, 199-229.
- 배도용(2014), 「한국어 학습자 산출물에 나타난 어휘 풍요도 측정을 위해 해결해야 할 몇 가지 문제」, 『우리말연구』 36, 133-156.
- 백유진(2020), 「일관성 수준에 따른 논설문 평가 과정의 인지적 차이 분석」, 『작문연구』 44, 37-72.
- 윤태일·이수안(2018), 『파이썬으로 텍스트 분석하기』, 서울: 늘봄.
- 이성영(2008), 「읽기 발달 단계에 대한 연구」, 『국어교육』 127, 51-80.
- 이현준·박영민(2016), 「쓰기 수행 수준에 따른 학생 글의 연결망 특성 비교 분석」, 『학습자중심교과교육연구』 16(9), 1039-1067.
- 이현준·박영민(2019ㄱ), 「자연어처리를 활용한 텍스트 연구 분야의 비교를 통한 자동채점 변인 탐색」, 『작문연구』 41, 255-287.
- 이현준·박영민(2019ㄴ), 「텍스트 의미망 분석을 통한 학습부진 학생의 읽기 교육 방안」, 『청람어문교육』 70, 129-168.
- 천해주·김영서·박소희·김주환(2020), 「작문 평가 준거의 평가자 간 해석 차이에 대한 연구」, 『작문연구』 47, 109-144.
- Attali, Y. (2004), "Exploring the feedback and revision feature of criterion", *Journal of Second Language Writing* 14, 191-205.
- Bereiter, C. (1980), Development in writing, In L. W. Gregg & E. R. Steinberg(Eds.), *Cognitive processes in writing*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bereiter, C. & Scardamalia, M. (1987), *The psychology of written composition*, Hillsdale, NJ: Lawrence Erlbaum Associates.

- Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013), *Analyzing social networks*, Thousand Oaks, CA: SAGE Publication Limited.
- Brown, R. (1952), *Structure and function in primitive society : Essays and addresses*, glencoe, IL: Free Press.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983), "From translation to problematic networks: An introduction to co-word analysis", *Social science information*, 22(2), 191-235.
- Chall, J. S. (1996), *Stages of reading development*, (2nd Ed.), Fort Worth, TX: Harcourt Brace College Publishers.
- Drieger, P. (2013), "Semantic network analysis as a method for visual text analytics", *Procidia-Social and Behavioral Science* 79, 4-17.
- Foltz, P. W., Streeter, L. A., Lochbaum, K. E., & Landauer, T. K. (2013), "Implementation and applications of the intelligent essay assessor", In D. M. Shermis & J. Burstein(Eds.), *Handbook of automated essay evaluation*, New York, NY: Routledge.
- Hillocks, G. (1986), *Research on written composition*, Urbana, IL: National Council of Teachers of English.
- Laydesdorff, L. & Welbers, K. (2011), "The semantic mapping of words and co-words in context", *Journal of Informetrics* 5(3), 469-475.
- Martin, A. & Wellman, B. (2011), Social network analysis: An introduction, In J. Scott & P. J. Carrington(Eds.), *The SAGE handbook of social network analysis*, Thousand Oaks, CA: SAGE Publications.
- Newman, M. E. J. (2006), "Modularity and community structure in networks", *Proceedings of the National Academy of Science* 103(23), 8577-8582.
- Page, E. B. & Peterson, N. S. (1995), "The computer moves into essay grading: Updating the ancient test", *Phi delta kappan* 76(7), 561-565.



## 의미연결망을 활용한 중등 학생 필자의 쓰기 특성 분석

권태현

이 연구는 의미연결망 분석을 통해 중등 학생 필자의 글 특성을 분석하는 데 목적이 있다. 학생 글에 대한 개별 단어 특성 및 단어 관계적 특성을 분석한 결과는 다음과 같다.

먼저, 글에서 사용된 단어(명사)의 수는 중학생보다는 고등학생 집단이 많았으며, 어휘 다양성을 의미하는 빈도 반영 명사 수의 경우 고등학생 집단이 중학생 집단보다 많았으나 상 수준이 하 수준에 비해서는 적은 양상을 보여주었다. 이는 쓰기 수준이 높은 학생들이 보다 적은 수의 단어로 더 긴 글을 쓰는 능력이 있음을 보여준다.

둘째, 의미연결망의 구조를 확인해 본 결과, 전체적인 단어 간 연결 정도는 고등학생 집단이 중학생에 비해 높았으나 쓰기 수준별로 보면 상 수준이 하 수준에 비해 연결 정도와 밀도가 더 낮은 것으로 나타났다. 상 수준 글은 내용이 다양하여 단어 간 연결 정도 및 밀도가 낮은 것으로 볼 수 있다. 밀도 차이에 기반한 군집 방식인 모듈래티지 지표에서도 학교급과 쓰기 수행 수준이 높아질수록 글의 의미 분화 정도가 커졌음을 확인할 수 있다.

셋째, 의미연결망 그래프의 패턴 분석 결과 학교급과 쓰기 수준이 높아질수록 중앙에서 주변으로 보다 촘촘한 연결을 보여주었다. 반면 쓰기 수준이 낮은 글에서는 핵심부와 주변부의 밀도 차이가 크고 연결망의 전반적인 응집력이 약한 의미망 패턴이 나타남을 확인할 수 있었다.

넷째, 개별 글의 단어 중심성을 분석한 결과 상 수준의 글에서는 하 수준과 달리 쓰기 주제와 관련된 내용어뿐만 아니라 필자의 관점을 드러내거

나 논리를 전개하는 데 활용된 단어들의 중심성 수치가 높게 나타났다. 이는 능숙한 필자일수록 자신의 지식을 변형하여 독자가 글의 내용을 받아들일 수 있도록 유도하는 지식 변형적 글쓰기의 양상과 관련된다고 볼 수 있다.

핵심어 의미연결망, 중등 학생 필자, 쓰기 특성, 중심성, 모듈래리티

## ABSTRACT

# An Analysis of Writing Characteristics of Secondary Students Using Semantic Networks

Kwon Taehyun

This study aims to analyze the writing characteristics of middle school students through semantic network analysis. The study results are as follows. First, the number of words (nouns) used in the text was higher in the high school group than in the middle school group, and in the case of the number of frequency-reflecting nouns indicating lexical diversity, the high school group had more words than the middle school group, but less than the upper-level group. Thus, students with higher writing levels have the ability to write longer texts with fewer words. Second, as a result of checking the structure of the semantic network, the overall degree of connection between words was higher in the high school group than in the middle school group, but by writing level, the degree centrality and density were lower than those of the upper and lower levels. According to the modularity index, which is a clustering method based on the difference in density, it was confirmed that the degree of semantic differentiation of texts increased as the school and writing performance levels increased. Third, as a result of analyzing the pattern of the semantic network graph, the higher the school and writing levels, the denser the connection from the center to the periphery. Conversely, it was confirmed that the semantic network pattern appeared with a large difference in density between the core and the periphery and weak overall cohesiveness of the network in writing with a low writing level. Fourth, as a result of analyzing the word centrality of individual texts, the centrality of words used to reveal the author's standpoint or develop logic, as well as content words related to the writing topic, were high in the upper-level texts, unlike the lower-level texts. This can be seen to be related to the aspect of knowledge-transforming writing, in which the more proficient

the writer, the more his knowledge is transformed to induce the reader to accept the content of the text.

**KEYWORDS** Semantic Network Analysis, Secondary Student Writer, Writing Characteristics, Centrality, Modularity