

교육과정평가연구  
The Journal of Curriculum and Evaluation  
2025, Vol. 28, No. 1, pp.107~138  
DOI: <https://doi.org/10.29221/jce.2025.28.1.107>

## LDA 토픽 모델링을 통한 문해력에 대한 담론 연구: 뉴스 데이터를 중심으로

박윤경 (연천중학교 교사)\*

### 요약

코로나-19 팬데믹 이후 국내 뉴스 기사에 나타난 ‘문해력’이란 주제어 및 토픽과 관련한 사회적 담론을 분석하고 이를 교육적 시각에서 해석하고자 하였다. 본 연구는 2020년 1월 1일부터 2023년 11월 18일까지 빅카인즈 플랫폼을 활용하여 국내 54개 언론사의 뉴스 기사 데이터를 수집한 후, R 프로그램의 한글 텍스트 분석을 하기 위해 KoNLP, NIADic 패키지를 활용해 데이터를 전처리한 후 LDA 토픽 모델링 방법으로 텍스트 마이닝을 수행하였다. 분석 결과, ‘디지털 기술, 디지털 리터러시, 플랫폼, 데이터’(Topic 2)가 핵심적인 주제로 도출되었으며, ‘넷플릭스, 유튜브, 소셜 미디어(SNS), 콘텐츠’(Topic 1, 5)을 통해 디지털 자원의 활용의 중요성과 더불어 ‘교과서, EBS, 초등학교, 고등학교’(Topic 3, 5)을 통해 교육적 접근이 중요한 논점으로 부각되었다. 팬데믹 이후 디지털 환경의 급속한 확산은 복합양식 텍스트 읽기의 중요성을 증대시켰으며, 이를 통해 학습자의 문해력 향상과 비판적 사고력 계발이 필수적인 과제로 대두되었다. 본 연구는 이러한 분석 결과를 바탕으로, 디지털 교과서와 에듀테크를 활용한 맞춤형 학습 환경의 설계, 복합양식 텍스트 읽기 전략의 체계적 개발 및 적용의 중요성, 그리고 매체 교육의 강화를 통해 학습자의 디지털 문해력 역량을 제고할 수 있는 방안을 제안한다. 디지털 시대 청소년들에게 익숙한 복합양식 텍스트를 비판적으로 읽어내고 전략적으로 활용할 수 있도록 지원하고, 교과 문해력을 함양할 수 있도록 교육적 비계를 모색해야 할 것이다.

주제어: 복합양식 텍스트, 교과 문해력, LDA 토픽 모델링

\* 제1저자 및 교신저자, nikyta@korea.kr

## I. 서론

문해력에 대한 모두의 관심이 뜨겁다. 청소년 시기는 인지적·정서적 역량 등이 종합적으로 형성되는 결정적 시기(Critical period)라 할 수 있다. 문해력 함양이 청소년 시기에 이루어져야 성인이 되어서도 이를 활용해 직업이나 학업을 효율적으로 수행할 수 있기 때문이다. ‘학습자에게 더 많은 표현, 소통, 해석할 수 있는 문해력 자원(resource)은 무엇보다 학습자의 미래 생활을 책임질 강력한 힘(power)’이 될 수 있다(Janks, 2009). 2000년 전후 전 세계적으로 유행한 코로나-19 팬데믹(COVID-19 Pandemic) 사태로 전례 없는 교육의 변화가 있었다. 그간 대면 수업으로 진행되던 것이 감염의 우려로 인해 비대면 수업으로 전격 전환되는 시기가 우리나라의 경우 1년 6개월 이상 이어졌다. 이로 인해 청소년 학습자가 자택에서 비대면 교육 중심으로 공교육이 이루어져 ‘교육의 결손, 문해력 및 학업 역량 저하’ 등의 문제가 야기되었다.

코로나-19 팬데믹은 대면 수업의 어려움으로 인해 2020년부터 원격 수업이 중심이 되었으며, 이로 인해 ‘기초 학력 부진’, ‘국어·영어·수학 성취도 저하’, ‘문해력 저하’ 등의 문제가 대두되었다. Ro & Elizabeth(2021)는 팬데믹 이후 미국 고등교육에서 아시아계 학생들에 대한 인종차별 문제를 지적하며 ‘교육적 형평성(equity)’을 논의하였다. 이수진(2022)은 한국의 온라인 개학이 지속 가능한 교육을 실현하려는 노력으로 긍정적 영향을 주었음을 강조하였으며, 이정우 외(2021)는 온라인 개학이 디지털 교육 인프라 구축과 원격 수업 실효성을 검증하는 계기가 되었다고 평가하였다. World Bank(2020)는 팬데믹 이후 교육적 대응 방향으로 ‘대처하기(Coping)’, ‘연속성 있게 운영하기(Managing continuity)’, ‘개선과 발전(Improving and accelerating)’ 단계를 제안하며, 학업 결손 해소와 대면·원격 수업의 혼합 방안을 강조하였다. 한편 코로나19 이후 국가 별로 디지털 기술을 활용한 ‘원격 학습 솔루션(remote-learning solutions)’이 활성화되었다. 아르헨티나는 Seguimos Educando 프로그램을 통해 TV·라디오 방송과 가상 현실 체험을 제공하였고, 인도는 DIKSHA 포털과 e-Pathshala 애플리케이션 등 다양한 플랫폼을 통해 학습 콘텐츠를 제공하였다(World Bank, 2020). Schwab(2016)은 ‘4차 산업혁명’에서 디지털 혁명이 가져올 변화를 예측하며 자율주행차, 블록체인, 3D 프린팅 간 이식 등 기술적 혁신을 제시하였다. 물론 비대면 수업 중심의 학교 교육이 대면으로 교실에 등교할 수 없는 상황에서도 유연하게 활용될 수 있다는 새로운 가능성을 검증한 점은 의미가 있다.

본 연구에서는 이러한 코로나-19 팬데믹 사태 이후 ‘문해력에 대한 인식 변화’를 ‘뉴스 기사’를 중심으로 탐구하여 교육적 시사점을 탐색하고자 한다. 즉, ‘문해력’이라는 화두

(topic)와 관련하여 우리 사회에서 코로나-19 팬데믹 이후 어떠한 담론이 형성되어 있는가에 대해 고찰하고자 한다. ‘R 프로그램’을 활용해 LDA 토픽 모델링’으로 코로나-19 팬데믹 이후 ‘문해력’으로 대한 뉴스 기사를 검색하여 관련도가 높은 주제어를 도출하여 궁극적으로 ‘문해력과 관련해 어떤 교육적 지원이 필요’한지 제안하고자 한다.

본 연구의 연구 문제는 아래와 같다.

- 연구 문제 1 : 2020년 이후 뉴스 기사에서 ‘문해력’과 관련한 토픽(topic)과 주제어(keywords)는 무엇인가?
- 연구 문제 2 : LDA 토픽 모델링 분석 결과를 바탕으로 ‘사회 담론에서 모색한 학습자의 문해력 향상을 위한 교육적 지원 방향’은 무엇인가?

## II. 선행 연구(prior research)

### 1. 문해력 연구

청소년 시기에 발달시켜야 할 여러 역량(competencies) 중에 ‘학업 역량’을 빼놓을 수 없다. 그 학업 역량에 영향을 주는 문해력이 ‘교과 문해력(Disciplinary Literacy)’이다. 교과 문해력에 대한 연구는 국외에선 십여 년 전부터 활발히 연구되었다(Howell et al., 2021). 국내에서도 비슷한 시기에 이에 대한 연구가 시작되어 많은 연구자들이 관심 갖고 연구하는 주제가 되었다(김종운 외, 2018, 김영란, 2021, 이용준, 2021, 편지윤, 2021, 박수자, 2019, 이경화, 2019, 천경록, 2018b, 장성민, 2023, 신지현·최영환, 2006). 교과 문해력 교육의 방향은 해당 분야의 ‘전문가처럼 생각하고 읽는’ 능력을 기르는 것을 목표로 한다. 왜 많은 연구자들이 ‘교과 문해력의 함양’에 대한 탐구를 할까. 미국의 경우 교육과정에 ‘과학·기술 교과와 역사·사회 교과 문해력에 대한 교육 내용’이 제시되어 있다(CCSS, 2010). 미국 뿐 아니라 캐나다 온타리오주 교육과정(2022, 2013)에서도 이를 강조하고 있다. 교과 문해력의 대표적인 연구자는 미국의 ‘Shanahan & Shanahan(2008, 2012, 2014)’와 ‘Moje(1996, 2000, 2008)’다. 학생들의 문해력을 지원할 사회·문화·교육적 자원은 Janks(2009)가 지적한 것처럼 학습자가 소속된 지역 공동체의 배경에 따라 상이하여 격차가 발생할 수 있는데 교육 기관에서 제공하는 비계(scaffolding)를 통해 해소할 수 있다. 교과 문해력은 단순한 이론이나 방법적 지식이 아니다. 특정 학문에서 지식을 발전시키기 위한 ‘탐구의 시작과 과정 전체’를 포함한다. 교과 별로 최초로 작은 의문이나 질문을 제기하는 것에서 출발하여, 이에 대한 논의를 확

장하고 다양한 사례나 근거 자료를 축적하여 그 분야의 전문가들이 그 내용에 대해 숙의하고 질문하고 논리적 타당성을 확보하는 과정을 거치게 된다. 이는 언어, 사회, 과학 등 교과별로 매우 상이하다. Gee(1990)는 학문의 전문가(학자)들이 탐구하는 담화 공동체(discourse communities)에서 지식을 개발하기 위해 읽고, 쓰고, 추론하는 모든 실천을 교과 문해력으로 보았다. 다가오는 21세기에도 문해력의 힘(power of the literacy)은 여전히 유효할까. ‘21세기 글로벌 역량은 언어 문해력과 수리 문해력으로 최첨단 인공지능(AI) 로봇 기술이 익숙한 학습자에게도 여전히 중등 교육의 풀어야 할 숙제는 필수 문해력 교육(Boyd, 2021)’이 답이 아닐까.

문해력은 왜 학생들에게 유익할까. 일반적인 문해력은 ‘잘 읽고 쓰고, 자신의 생각을 잘 표현하는 역량’으로 본다면 이는 학생들이 성장하여 평생 활용할 수 있는 유용한 역량이란 점은 주지의 사실이다. 교과 문해력은 기본 문해력에서 한층 심화하여 학생들의 교과 학습에 유의미한 교육 효과를 가져올 수 있다. ‘Shanahan & Shanahan(2008, 2012, 2014)’에서도 ‘교과 문해력이 학생들의 성공적인 학업 수행이나 중등학교를 졸업하고 직업 세계에서 유용하게 기여할 수 있는 점’이 강조된다. 특히 사회 및 과학 교과 분야의 글은 학생들이 중등 교육과정부터 본격적으로 접하게 되므로 더욱 ‘교과 문해력 역량’이 중요하다. 교과 문해력은 해당 교과를 연구하는 ‘학자(disciplinarian)’의 관점이 무엇인지 탐색하는 것에서 출발한다. 모든 학문은 지식을 오랜 기간 동안 탐구하고 발전시켜 확고한 이론이나 원리, 개념으로 정립하는 과정을 갖는다.

‘Moje(1996, 2000, 2008, 2015)’는 고등학교의 화학 교사로 교실 내에서 수행할 수 있는 ‘문해력 탐구’를 탐구하여 ‘교과 문해력의 실천적 휴리스틱 내용’을 개발하였다. Moje & Speyer(2008)에서는 고등학교 과학 및 사회 교과의 도전적인 텍스트의 현실과 교사가 이해를 중재할 수 있는 방법을 밝혀 ‘과학이나 사회 교과’에 대한 교과 문해력 교육의 중요성을 시사했다고 볼 수 있다. Jenkins(2023)가 연구한 내용은 중요한 의미가 있다. 과학 교과에서도 ‘과학 교과의 문해력’이 교육적으로 의미있는 주제라는 것과 특히 중등 교육에서 ‘지식적 용어나 어휘’를 충분히 아는 것과 학생들의 과학적 역량 사이의 긍정적 인과 관계를 밝히고 있다. 특히 교과 간의 협업(cross curricula links)을 통해 ‘과학 교과 문해력 교육’의 신장을 강조한 점은 국어 교과에서 어떤 역할을 해야 할지 시사한다. Jenkins(2023)는 영어 교과의 읽기 전략 및 어휘 이해 학습이 과학 교과의 글 읽기 유창성(reading fluency)을 향상시켰다는 점을 중등 교사들의 협업을 통해 밝힌다. 물론 이는 굳이 박사 논문에서 증명하지 않아도 충분히 예상되는 인과관계이다. ‘과학 및 사회 교과 문해력 교육’의 방향 역시 이렇게 교과 간의 협업을 통해 실현이 가능할 수 있다.

Jenkins(2023)는 교사로서 학생들이 학습에 곤란을 지켜보며 ‘문해력 교육’의 필요성

을 절감하며 Shanahan and Shanahan(2008)이 강조한 ‘교과 문해력 교육’의 중요성을 깨닫고 연구를 시작하게 된다. Wellington and Osborne(2001:2)에서 주장된 “every science lesson is a language lesson”을 적극적으로 수용하는 모습에서 과학 교육에서 ‘언어의 학습’은 절대적임을 드러낸다. Wellington and Osborne(2001)에서 강조한 ‘학습의 전달(delivering learning)’로서 언어 교육은 과학 교육에 절대적으로 중요하다고 본 점 주목할 부분이다. 언어는 지식을 표현하고 담는 도구로서 이에 대한 연습과 학습 없이 어떤 학문의 ‘지식’도 성장시키기 어렵다 할 것이다. 그는 과학 교과의 문해력에 대한 중요성을 강조하면서도 연구자들 사이에 그에 대한 완벽한 개념 정의는 부재하다고 지적한다. ‘과학 문맥에서 문해력과 과학적 문해력(literacy in science and scientific literacy)’은 분명히 구분되어야 하고, ‘과학적 문해력’의 경우 Hurd(1969)이 오래전부터 정립한 개념은 ‘과학에 대한 이해와 그 활용’이라고 볼 수 있다고 설명한다. ‘과학이라는 학문 공동체’에서도 ‘과학적 문해력’의 범위에 대한 논의는 여전히 진행중이며 아직 명확히 합의된 바가 없다는 것을 알 수 있다. 본 연구의 주제인 ‘교과 문해력’ 역시 범위나 개념 설정의 어려움이 비슷한 사유에서 기인한다. 여기서 문해력 지도 교사의 역할은 ‘학문의 전문가처럼 글을 이해하고 활용하는 방법’을 시범보이고 학생들은 이를 접하고 실천해보는 것이 중요하다(Quigley, 2018, Jenkins, 2023:27에서 재인용)고 강조한다. 교사는 ‘언어적 동의, 반대 의견 표현, 근거로 활용(Quigley, 2018, Jenkins, 2023:28에서 재인용)’하는 학생들의 언어적 의사소통을 지원할 수 있다. 일반적 문해력의 과정에서 교과 특수한(discipline specific) 내용이 포함된 것이 교과 문해력의 교육 요소이다. 본 연구에서 주목하는 교과는 과학 교과와 사회 교과에 대한 것이다. 이를 연구한 결과들에서 공통적으로 발견된 내용은 아래와 같다.

<표 1> 과학 및 사회 교과의 학문 특수성(Burke & Kennedy, 2024)

교과	학문 특수성
과학	관찰이나 실험을 토대로 이론이나 개념의 정립, 내용 사이의 인과 관계 강조
사회	1차 및 2차 사료의 신빙성 강조, 내용의 편향성 판단

국내에도 교과 문해력에 대한 주요한 연구는 ‘김종윤 외(2018), 천경록(2018)’라 할 수 있다. 실제 고등학교 교과 교사를 대상으로 연구를 수행하여 ‘학문의 고유한 인식론, 용어 활용, 내용 지식, 담화 관습’을 ‘학문 문식성’으로 보고 국어 교과에서 이를 교육할 것을 강조한 김종윤 외(2018)의 연구는 본 연구의 방향과 맥을 같이 한다. 본 연구에서 활용하고 있는 ‘교과 문해력’이라는 용어 이전에 대부분의 연구에서 ‘학문 문식성(김종윤 외, 2018, 김영란, 2021, 이용준, 2021, 편지윤, 2021), 내용문식성(박수자, 2019), 교과 문식성(이경화, 2019), 내용교과 문식성(천경록, 2018b), 교과 문해력(장성민,

2023), 교과 문해력(신지현, 최영환, 2006)’으로 활용되었다. 국외에서 ‘교과 문해력’에 대한 연구가 오랫동안 많은 연구가 수행되었고, 최근 국내에서도 활발하게 청소년 학습자의 ‘문해력 함양’에 대한 관심이 모아지게 되었다. 교과의 특성을 반영하여 ‘읽고 쓰는 문해력 교육’이 강조된다. 김종윤 외(2018), 천경록(2018a, 2018b), 이경화(2019a, 2019b), 김영란(2021), 이용준(2021) 등의 연구는 각 교과의 고유한 특성에 맞는 읽기·쓰기 교육이 필요하다는 점을 제시하였다. 또한 실제 교육 현장의 상황을 분석하고 개선 방안을 모색하는 데 초점을 둔 천경록(2018a)의 교사 인식 조사, 김주환 외(2020)의 중등 작문 교수·학습 실태 조사, 장성민(2023)의 교과 문해력 검사 설계 연구 등도 이에 해당한다.

De La Paz(2005)는 2주간 중등 학생들에게 ‘역사적 탐구 전략(historical inquiry strategies)’을 활용해 ‘역사적 사고력과 논증력’을 함양하는 사회 교과 문해력 교육을 실시한다. SRSD(Graham, Harris, MacArthur, & Schwartz, 1998; Harris & Graham, 1996, De La Paz, 2005:140에서 재인용)를 사용하여 개념적으로 어려운 아이디어를 가르치고, 1차 및 2차 설명을 통해 추론하여 복잡한 역사적 사건에 대해 이해하고자 하였다. Wineburg(1991)에서 역사가들의 전문적 읽기 전략은 ‘출처 분석(Sourcing), 상호 비교(Corroboration), 맥락화(Contextualization)’로 학생들이 역사 텍스트를 읽을 때 ‘무조건적 정보 수용 태도’와는 명백한 차이가 있으므로, 역사라는 학문적 특수성을 고려한 읽기 교육을 수행할 필요가 있다고 설명한다. 여기서 글의 출처를 분석하고 상호 비교 분석하여 내용을 비판적 사고를 수행할 수 있는 능력이 필요하다고 보았다. 출처 분석은 저자와 출처(author and origin)를 분석하고, 상호 비교는 일관성과 신뢰성(consistency and reliability)을 역사가끼리 상호 비교하는 작업을 수행한다는 것이다. 맥락화는 역사적 지식에 기반하여 신뢰도(credibility)를 더 넓은 시공간적 맥락에서 보는 관점이다.

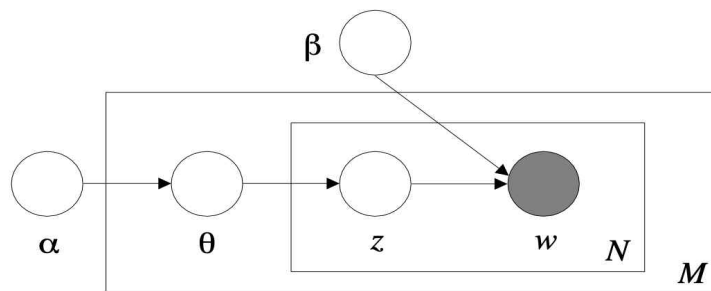
전 세계적으로 일반적인 읽기 방법으로 충분히 중등 교육에서 과학, 사회 교과의 글과 같은 전문화된 글을 소화하기에 어려움이 있다는 것을 밝힌 연구가 많다(Grigg et al., 2007; Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007, Shanahan and Shanahan, 2008:45에서 재인용). 중등 교육에서 ‘교과 문해력’을 탐구한 Shanahan and Shanahan(2008)은 주요 내용 교과 교사의 ‘글 이해 전략(comprehension strategy)’을 탐색한 연구는 이 분야에서 중요한 업적으로 널리 인정 받고 있다. 여기엔 본 연구가 밝히려는 ‘교과 문해력의 교육 방안’의 기초 토대가 되는 부분이 담겨 있다. 교과 문해력에는 교과 전문가들이 지식을 발전시키는 모든 과정이 포함된다. Shanahan and Shanahan(2008)은 사회, 과학, 수학 교과의 문해력에 대해 밝혔는데 중등 교육에서 학생들이 유용하게 활용할 수 있을 것으로 기대했기 때문일 것이다. 화학, 역사, 수학의 구체적이고 고도로 전문화된 학문적 독서 요구 사항을 보다 직접적이고 명확하게 다룰 수

있도록 ‘교과 문해력 중심’의 읽기 교육이 필요하다. 많은 인지 및 언어 전문가들은 ‘지식의 전문가’들이 학문을 발전시키는 과정에서 사용하는 구체적인 ‘언어를 활용한다’는 점을 강조하고 있다(Bazerman, 1998; Fang, 2004; Geisler, 1994; Halliday, 1998; Schleppegrell, 2004, Shanahan and Shanahan, 2008:47에서 재인용). 국제화 시대 문해력의 중요성(Carnevale, 1991), 교육과 소득의 관계(Arc, Phillips, & McKenzie, 2000; Barton & Jenkins, 1995), 문해력과 건강(Berkman, DeWalt, Pignone, Sheridan, Lohr, Lux, et al., 2004), 학업의 성공과 문해력(American College Testing, 2006) 중등 교육 이전에 초등 교육 단계에서 문해력 교육을 철저히 했다고 해도 과학, 사회 교과에 글을 이해할 수 있는 ‘독해 전략 습득’은 저절로 습득이 되지 않았다(Perle et al., 2005). 문해력에 대한 새로운 요구(Levy & Murnane, 2004)를 고려, 학문적 또는 기술적 문해력 사용에 내재된 고수준 기술과 능력이 요구되는 것이 현대 사회이다.

과학 교과와 역사 교과의 학문 탐색의 원리가 상이하다. 과학은 학문적 특성이 ‘실험·관찰 결과의 해석과 요소들의 인과 관계’를 중시하고, 역사는 ‘사료의 신뢰도와 다양한 관점으로 해석하는 것’에 초점을 둔다(Shanahan & Shanahan, 2008, 2012, 2014). 이처럼 각 교과에 글은 학문적 특수성이 반영되어 있고, 이를 학생들이 고려해 읽을 수 있도록 지도하는 것이 필요하다.

## 2. LDA 분석

텍스트 마이닝 중 하나인 LDA는 토픽 모델링으로 텍스트 내에 잠재된 토픽(주제)의 구조를 찾는다. Blei, et al.(2003:996)에서 LDA(Latent Dirichlet Allocation)의 원리는 ‘말뭉치(corpus)의 생성 확률 모델(generative probabilistic model)’이라고 설명한다. 기본적인 아이디어는 ‘문서에 잠재된 토픽(latent topics)에 대한 무작위 혼합으로 구성되고, 각 토픽은 단어의 분포라는 특징이 있다’는 점이다. 즉 문서는 토픽들이 각각의 확률 분포를 갖고 혼합되어 있고, 토픽은 단어들이 각각의 확률 분포를 갖고 혼합되어 있다고 볼 수 있고, 토픽들 사이엔 단어의 중복이 있을 수 있다. LDA는 문서에서 토픽이 갖는 비율과 토픽에서 단어가 갖는 비율을 곱하여 그 문서에서 토픽이 갖는 단어의 비율을 구하고, 이 과정을 1000회 이상 반복하여 연구자가 정한 토픽 개수와 토픽을 구성하는 단어의 비율을 산출하게 된다. 이 확률 분포에 따라 단어와 토픽을 산출하고 할당하는 기술적 작업은 ‘R 프로그램’으로 수행하게 된다.



[그림 1] LDA의 시각화 모델(Blei, et al. , 2003:997)

[그림 1] LDA의 시각화 모델(Graphical model)에서 사각형은 복제물(replicates)을 나타내는 “판(plates)”이다. 바깥쪽 판은 문서를 나타내고, 안쪽 판은 문서 내에서 반복되는 토픽과 단어의 선택을 나타낸다. [그림 1]에서 매개변수  $\alpha$ 와  $\beta$ 가 주어지면 토픽의 혼합은  $\theta$ ,  $N$ 개의 토픽 세트는  $z$ ,  $N$  단어 세트는  $w$ 로 이들의 결합 분포(joint distribution)를 나타낸다. LDA 모델에서  $M$ 은 문서 집합의 크기를 나타내는 문서의 총 개수이며,  $N$ 은 각 문서에서 단어의 총개수를 의미하여 문서의 길이를 나타낸다. 모수(Parameters)로는  $\alpha$ 와  $\beta$ 가 있으며,  $\alpha$ 는 각 문서의 주제 분포의 사전 분포를 정의하는 디리클레 분포의 모수로, 문서가 얼마나 다양한 주제를 포함할 수 있는지를 조정한다.  $\beta$ 는 주제당 단어 분포를 정의하는 디리클레 분포의 모수로, 특정 주제 내에서 어떤 단어가 중요한지를 결정한다. 변수(Variables)는  $\theta$ ,  $z$ ,  $w$ 가 있으며,  $\theta$ 는 문서별 주제 분포로 디리클레 분포에서 샘플링된 값이며, 문서가 어떤 주제들로 구성되어 있는지를 나타낸다.  $z$ 는 각 단어의 주제 할당을 나타내는 잠재 변수로, 문서 내의 각 단어가 어떤 주제에 속하는지를 표현한다.  $w$ 는 단어 관측치로, 문서에서 실제로 관측된 단어를 나타내며, 각 단어는 주제와 단어 분포에 의해 생성된다(Blei, 2013). 깁스 샘플링(Gibbs Sampling)은 복잡한 확률 분포에서 샘플을 생성하기 위해 각 변수를 하나씩 조건부 분포를 기반으로 샘플링하는 방법이다. 나머지 변수들의 값을 고정한 상태에서 한 변수씩 순차적으로 값을 업데이트하며, 이 과정을 반복해 전체 분포를 점진적으로 추정한다. 복잡한 계산을 단순화할 수 있어 베이지안 추론(Bayesian estimation)과 토픽 모델링 같은 분야에서 널리 사용된다. LDA 모델에서는 깁스 샘플링을 통해 각 단어의 주제 할당을 반복적으로 샘플링하며 문서의 주제 분포와 주제별 단어 분포를 추정한다. 김성연(2020)은 “토픽모델링을 활용한 SIAM Journal on Applied Mathematics의 연구 동향 분석”에서 깁스 샘플링을 적용하여 주요 연구 주제를 도출하며 LDA 모델을 활용했다. 남소희, 전수영(2019)은 “점별 상호정보량을 이용한 LDA 토픽모형 추론”에서 깁스 샘플링을 활용해 LDA 모델을 분석했다.



이 모델은 문서가 여러 주제의 혼합으로 구성되고, 각 주제는 단어에 대한 특정 분포를 가진다고 가정한다. 이를 통해 문서 집합의 잠재 구조를 파악하며, 새로운 문서에 대한 주제 분포를 추론한다. 교육 분야에서도 텍스트 마이닝 연구법은 유용하다. ‘학술 논문이나 초록, 강의계획서, 교육과정, 뉴스 기사’의 빅데이터에서 ‘주제어’를 중심으로 검색하여 ‘빅데이터를 활용한 교육의 필요성, 교과서나 교육과정 분석’ 등의 연구가 주효하다. 황홍섭(2019)은 ‘빅데이터, 사회과 교수·학습, 미래교육’을 검색어로 ‘포털 사이트와 학술 논문의 데이터’를 텍스트 마이닝으로 탐구하였다. 이를 통해 ‘빅데이터를 활용한 사회과 교수·학습 방안이 초복잡계(supercomplex system)의 성격을 가진 미래 사회와 사회과 교육과정의 핵심 역량을 신장시킬 수 있다’고 주장한 바 있다. 장성민(2019)은 강의계획서의 ‘주제어’를 중심으로 ‘빈도 분석’ 등의 텍스트 마이닝 연구를 수행하였다. 김선아 외(2015)는 텍스트 마이닝 방법을 활용해 미술과-국어과 교육과정에서 융복합 공통 키워드를 추출해 R 프로그램으로 빈도 및 연결 관계를 분석하였다. 이해인(2023)은 고등학교에 ‘인공지능 기초’ 과목이 도입되면서 8종 교과서의 핵심적 학습 요소를 추출하여 상위 20개의 개념을 정하여 교과서별 어떤 개념이 주로 수록되어 있는지 텍스트 마이닝 방법을 통해 탐구하였다. 최규리, 오윤정(2023)은 국내외 초등 환경교육 연구 동향을 텍스트 마이닝으로 탐색하여 초등과학 교육과정과 연계한 환경교육의 방안을 연구하였다. 이경진 외(2023)는 문화예술 교육 프로그램의 학교 교육과정 연계를 중심으로 전문가 견해를 텍스트 마이닝하여 학교 문화예술교육의 실천 방안을 모색했다. 조남옥, 조규락(2023)은 ‘창의적 체험활동’을 검색어로 학술 논문의 초록을 텍스트 마이닝하여 창의적 체험활동의 연구 동향을 탐색하였다. 노들(2021)은 ‘디지털 리터러시’라는 키워드를 중심으로 KCI에 등재된 논문 초록 데이터를 크롤링하여 텍스트 마이닝으로 국어교육에 주는 시사점을 탐색한 바 있다. 이우진, 백혜진(2022)은 ‘리터러시’와 ‘교육’을 검색어로 학술 논문을 분석하여 ‘디지털 리터러시’, ‘미디어 리터러시’, ‘초등학교’가 상위 빈도를 보였음을 밝혔고, 디지털 융복합 리터러시 교육의 필요성을 주장하였다. 박숙경(2023)은 학술지에 게재된 논문을 중심으로 ‘건강 문해력’을 키워드로 활용하여 토픽 모델링 기법을 활용해 텍스트 마이닝을 수행하였다. 박지우 외(2022)에서도 ‘TF-IDF, LDA 기반 텍스트 마이닝을 중심’으로 학습 연구 동향을 분석해 ‘인공지능·데이터 기반 개별화 학습’ 등의 주제어를 도출해 코로나 19 이후 디지털 기술 활용한 비대면 학습 환경을 강조한 바 있다. 방미현, 이영민(2020)에서도 TF-IDF를 통해 핵심 주제를 도출하여 청년 시대에 문제와 해결책을 고찰하였다.

텍스트 마이닝은 정형 및 비정형 텍스트(unstructured)로 구성된 빅데이터를 분석하고 시각화하여 주제나 의미를 탐구하는 연구 방법이다(추형석 외, 2023). 텍스트 마이닝 기법은 주로 ‘정책에 대한 인식’을 조사하여 정책적 개선안을 마련하거나 ‘특정 주제에

대하여 학술 논문 검색'하여 '연구 동향'을 분석할 때 사용된다. 교육학 분야에서도 '교육과정 텍스트를 분석'하여 개선안이나 시사점을 도출한다. 한편 텍스트 마이닝은 '정책적 결정'을 합리적으로 수행하기 위해 사람들의 인식이나 연구 동향이 조사되기도 한다. 유아 지원 정책을 위해 뉴스 기사의 댓글을 텍스트 마이닝한 연구(조수진, 2023)와 디지털 평생교육 연구 동향을 텍스트 마이닝하여 정책 인식을 연구한 사례(안현용, 2023)가 그것이다. 배정임(2023)은 '독서치로 연구의 동향'을 분석하기 위해 텍스트 마이닝과 네트워크 분석을 수행하였다.

### 3. 뉴스 빅데이터 연구

빅데이터는 '지식 정보의 홍수'라는 현대 사회에서 다양한 분야에서 유용하게 활용될 수 있는 '정보 자산(information assets)'이다. Gartner<sup>1)</sup>가 이야기하는 빅데이터의 개념<sup>2)</sup>은 '대용량(high volume), 초고속(high velocity), 높은 다양성(high variety)을 가진 정보 자산'으로 볼 수 있다. Manyika, et al.(2011)에서 빅데이터를 활용해 가치(value)를 창출하고 혁신적 잠재력을 탐색하여 더 나은 의사 결정을 내릴 수 있도록 빅데이터의 집계(aggregate), 조작(manipulate), 분석, 시각화하는 기술을 제안하였다.

빅데이터인 '뉴스 기사'를 활용한 연구도 활발하다. 특히 사회적으로 논의되고 있는 '코로나 19, 환경문제, 인공지능(AI) 윤리' 등 다양한 논제에 대한 비판적 시사점을 포함한 뉴스 빅데이터를 활용한 연구가 다수다. 나경식, 이지수(2018)에서도 빅카인즈를 활용해 뉴스 빅데이터를 연구하여 20년간의 메타데이터 및 연관어 분석 결과를 토대로 정보화 트렌드를 탐구한 바 있다. 김예지, 김갑수(2023)에서 4차 산업혁명 지능 정보 사회에서 빅데이터를 활용하여 초등학생을 위한 교수학습 모형을 탐색한다. 이현주(2023)는 텍스트 마이닝 방법을 활용해 20년간 한국에 체류하고 있는 '유학생'과 관련된 신문 기사를 탐구하여 언론 분석 사이트인 빅카인즈를 활용하여 당대 현실을 반영하고 있는 것을 확인한다. 김은정 외(2023)에서 뉴스 기사를 빅데이터로 활용하여 '뉴스가 사회적 이슈를 잘 반영하고 정보의 정확성 면에서 비교적 신뢰할 만하다'고 설명한다. 박지원(2023)은 LDA 토픽 모델링은 텍스트 형태의 본문에서 잠재적 의미구조를 발견하기 위한 통계적 알고리즘을 활용하여 청소년 정책을 키워드로 탐구하였다. 이찬중(2023)은 '코로나19(감염병 키워드)'를 중심으로 영상 플랫폼 유튜브(Youtube)에서 검색하여 키워드 네트워크를 분석하여 사회적 이슈를 탐구하였다. 특히 미래에 대한 예측이나 방향

1) "가트너 주식회사(Gartner, Inc.)는 미국의 정보 기술 연구 및 자문 회사로, 다국적 IT기업 및 각국의 정부기관 등을 주 고객으로 두고 있으며 설문 조사 부분의 높은 신뢰도로 공신력이 크다(황홍섭, 2019)."

2) <https://www.gartner.com/en/information-technology/glossary/big-data>

을 탐색하기 위해 급변하는 사회적 트렌드나 이슈를 신속하게 전달하는 매체의 뉴스 빅데이터가 연구 자료로서 유용하다고 설명한다(박대민, 2016, 나경식 등, 2018). 김수환, 류승완(2022)은 LDA 토픽모델링 방법을 활용해 ‘ESG와 문화예술의 관계’를 중심으로 뉴스 기사 데이터를 분석하여 숨겨진 의미구조와 패턴을 탐색하였다.

본 연구에서도 ‘뉴스 기사’라는 빅데이터를 연구 자료로 활용하여 ‘문해력’이란 검색어로 수집한 데이터를 텍스트 마이닝한 후 도출된 ‘주제어’를 ‘교육적 시각’에서 탐구하고자 한다.

### III. LDA 토픽 모델링 연구

#### 1. LDA 토픽 모델링 분석 과정

연구 방법은 <표 2>에 제시한 대로 LDA 토픽 모델링을 수행한다. 분석할 뉴스 기사 데이터는 언론 분석 사이트인 빅카인즈(Big Kinds, [www.bigkinds.or.kr](http://www.bigkinds.or.kr))를 활용하여 기사 원문에서 추출된 키워드 데이터를 크롤링(crawling)한다. 여기서 수집된 자료를 전처리하여 LDA 토픽 모델링 분석을 위하여 ‘R 프로그램’을 활용하고자 한다.

<표 2> R 프로그램 활용 연구 방법

연구 절차		단계별 연구 내용
1단계	키워드 설정 및 검토	‘문해력’ 유사어와 불용어 설정 기간: 2020.01.01.~2023.11.18
2단계	자료 수집 및 검토	자료 종류: 54개 신문, 방송사 뉴스 방법: 빅카인즈에서 크롤링
3단계	수집된 데이터 전처리(Pre-processing)	‘R 프로그램’의 KoNLP, NIADic 패키지(packages) 활용
4단계	LDA 토픽 모델링 분석	‘R 프로그램’의 LDA, tidytext, tm, ggplot, dplyr 패키지 활용

LDA 토픽 모델링은 빅데이터 속 잠재된 주제를 파악하고자 할 때 유용한 방법이다. ‘R 프로그램’ 활용을 위한 데이터 전처리는 KoNLP(Korean Natural Language Processing) 한글 자연어 처리 패키지와 형태소 분석 사전인 NIADic(National Institute for Japanese Language and Linguistics Dictionary)를 활용하였다. 데이터 전처리를 위해 분석에서 제외할 단어를 ‘tm 패키지’의 불용어(stopwords) 제거 함수를 활용하였다.

마이닝할 키워드는 ‘문해력’과 ‘문식성, 리터러시, literacy, 文識性, 文解力’을 유의어로 처리하였다. ‘R 프로그램’을 활용해 수집된 뉴스 기사의 키워드 중 가장 많은 빈도 (frequency)를 보이는 단어를 추출하였다. 뉴스 기사 검색과 ‘R 프로그램’에서 제외된 단어는 173개로 ‘도서관, 도서관과 관련된 행사, 도서 안내’와 ‘기업체의 문해력 프로그램 홍보’ 및 연구 주제와 무관한 특정 정치인과 관련된 기사이다. 빅카인즈(Big Kinds, [www.bigkinds.or.kr](http://www.bigkinds.or.kr))에서 수집된 뉴스 데이터는 검색 기간은 ‘2020.01.01.~2023.11.18’로 ‘코비드(Covid)-19 팬데믹 이후부터’이다. 수집한 뉴스 기사는 우리나라 모든 신문 방송 54개(경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보, 매일경제, 머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 헤럴드경제, 강원도민일보, 강원일보, 경기일보, 경남도민일보, 경남신문, 경상일보, 경인일보, 광주매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문, 무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 중도일보, 중부매일, 중부일보, 충북일보, 충청일보, 충청투데이, 한라일보, KBS, MBC, OBS, SBS, YTN, 디지털타임스, 전자신문) 회사에서 제공한 내용이다. 최초 수집된 뉴스 검색 결과는 6,364 건이다. 뉴스 기사 검색과 ‘R 프로그램’에서 제외된 단어는 173개로 ‘도서관, 도서관과 관련된 행사, 도서 안내’와 ‘기업체의 문해력 프로그램 홍보’ 및 연구 주제와 무관한 특정 정당, 정치인과 관련된 기사이다. 이 중에 연구 주제와 관련이 없는 내용은 제외 처리를 하니 최종적으로 739건의 기사가 수집되었다.

## 2. LDA 토픽 모델링 분석 결과

본격적인 토픽 모델링에 앞서 ‘R 프로그램’ 활용해 ‘문해력’을 중심으로 수집한 뉴스 기사 데이터에서 빈도수가 높은 단어를 <표 2>와 같이 추출하였다. 빈도수 1위는 ‘교육’으로 문해력과 가장 관련이 깊은 것을 알 수 있다. 2위는 ‘디지털’로 ‘포스트 코비드(Covid)-19 팬데믹’ 시대에 급격한 발전을 보이는 ‘정보 통신 기술’로 야기된 ‘지능 정보 사회상(Intelligent Information Society Phenomenon)’과 관련이 깊다. 3위는 ‘이해’로 ‘문해력’은 이해를 토대로 형성된다고 볼 수 있다.

## 가. 주제어 빈도 분석

<표 3> R 프로그램 활용 주제어 빈도(keywords frequency) 분석

순위	단어	빈도수	순위	단어	빈도수	순위	단어	빈도수
1	교육	1841	11	사람	622	21	학습	436
2	디지털	1227	12	시대	621	22	데이터	428
3	이해	944	13	학생	560	23	의미	412
4	능력	833	14	중요	546	24	변화	404
5	정보	830	15	국어	511	25	자신	403
6	사회	807	16	활용	501	26	과목	402
7	ai	799	17	수업	500	27	수학	402
8	생각	704	18	독서	469	28	학교	400
9	세대	686	19	영어	453	29	수준	391
10	미디어	666	20	사과	437	30	아이들/표현	388

## 나. 워드 클라우드 분석

단어 빈도 산출 결과를 토대로 ‘R 프로그램’을 활용해 ‘Wordcloud 패키지’를 활용해 워드 클라우드를 생성한 결과가 아래와 같다. 워드 클라우드는 글자의 크기로 문서에서 단어의 빈도나 중요도를 표현한 이미지를 의미한다. 아래 [그림 2]에 ‘문해력’과 관련하여 절실한 것이 ‘교육’이라는 점이 시각적으로 드러나고 있다.



[그림 2] 워드 클라우드

## 다. TF-IDF 분석

TF-IDF(Term Frequency-Inverse Document Frequency)는 ‘문서 집합 내에서 단어의 중요도를 정교하게 분석하는 텍스트 마이닝 방법’이다. R에서 ‘tidytext 패키지’의 ‘bind\_tf\_idf() 함수’를 활용해 분석한다. ‘문해력이란 주제어로 검색된 뉴스 기사’라는 문서 집합에서 추출된(extracted) 주요한 단어들을 통해 교육적 시사점을 분석해 보면 다음과 같다. 첫째, ‘교육’의 중요성을 알 수 있다. 가장 높은 TF-IDF 점수인 ‘33.47’을 보이고 있다. 문해력과 관련되어 ‘독서 교육, 문해력 교육, 공교육’ 등의 모든 교육이 중요하다. 둘째, ‘디지털’이 ‘28.08’로 두 번째 높은 TF-IDF 점수를 보이고 있어, ‘디지털 시대에 대한 문해력 논의가 활발함’을 보여준다. 셋째, ‘AI(인공지능)’이 ‘18.85’로 ‘문해력이 AI(인공지능) 기술의 활용과 관련성이 높음’으로 볼 수 있다. 특히 생성형 AI(인공지능)를 활용하여 ‘학습자의 독서 프로필 분석, 독서 추천, 개인별 수준별 독서 프로그램’ 등 문해력 및 독서 교육이 긍정적 역할을 할 수 있음을 시사한다. 넷째, ‘미디어’가 ‘18.84’로 미디어 리터러시(media literacy)가 중요한 교육적 화두임을 알 수 있고, 이는 전통적 책임기에서 확장된 ‘비판적 이해를 중심으로 다양한 미디어의 복합양식 텍스트 읽기’ 교육이 필요하다. 다섯째, ‘이해’가 ‘18.14’로 문해력과 관련이 깊은 것은 ‘이해’임을 드러낸다. ‘이해’는 독서 교육을 통해 함양되는 기능(skill)으로 문해력 함양엔 여전히 대단한 왕도가 있을까. 학습자가 꾸준히 평생 독자가 될 수 있도록 생성형 AI 교육 프로그램을 효율적으로 활용해 디지털 네이티브(Digital Natives, Prensky, 2001)라는 청소년의 학업 성취와 진로 개발을 교육적으로 지원해야 할 것이다.

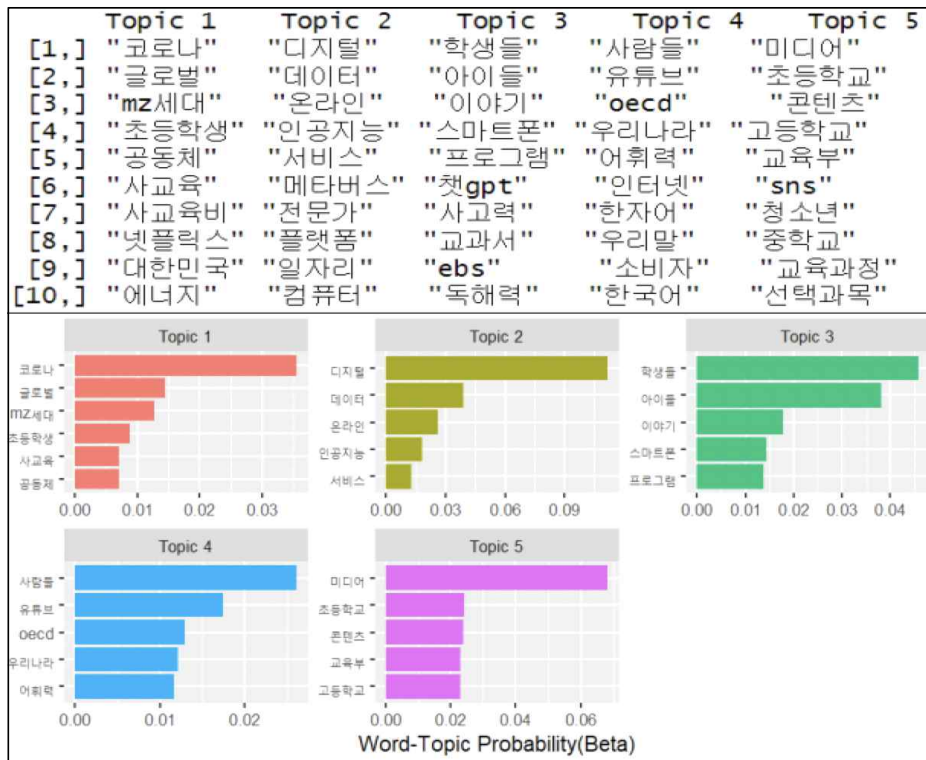
<표 4> TF-IDF 분석 결과(상위 10개)

순위(Rank)	주제어 (Keywords)	TF-IDF score
1	교육	33.47029297
2	디지털	28.08558246
3	AI	18.85798408
4	미디어	18.84816014
5	이해	18.14574721
6	정보	17.53267034
7	능력	17.0217847
8	국어	15.67253891
9	독서	15.16415013
10	사회	15.13465098

## 라. LDA 분석

‘R 프로그램’의 토픽 모델링(topic models) 패키지를 활용해 LDA를 분석한 결과를 <표 5>에 제시하였다. ‘Gibbs’ 방법을 활용하여 주제어는 각 10개씩 산출하며 토픽(k)은 5개로 설정하였고, 반복 횟수는 1000회로 설정하여 적절한 5개의 토픽의 수가 유의미한 해석을 수행하는데 적절하다고 보았다. Latent Dirichlet Allocation(LDA) 모델에서 토픽 수(k)와 반복 횟수는 모델의 성능과 해석 가능성에 핵심적인 요소로 작용한다(Röder et al., 2015:399). Griffiths & Steyvers(2004)는 Gibbs 샘플링을 통한 반복적 추론 과정이 사후 확률 분포에 수렴하는 데 중요한 역할을 한다고 강조하였다. 반복 횟수를 충분히 설정하면, 데이터에 대한 신뢰도 높은 추정을 가능하게 하며, LDA 모델이 안정적인 결과를 도출할 수 있다(Griffiths & Steyvers, 2004:5230). 실제로 연구에서는 Gibbs 샘플링의 반복 횟수를 1000회로 설정하여 결과가 수렴하는 과정을 설명하였고, 이는 LDA 모델링에서 일반적으로 권장되는 설정으로 받아들여지고 있다. Blei et al.(2003)은 LDA 모델의 구조와 수렴 과정을 기술하며, MCMC(Markov chain Monte Carlo) 활용이 토픽과 단어의 분포를 효율적으로 추론하는 데 효과적임을 보여주었다. 이들은 적절한 반복 횟수가 모델의 최적화를 보장한다고 설명하였다(Blei et al., 2003:1000). Gibbs 샘플링도 MCMC 기법 중 하나다. 본 연구에서는 선행 연구의 결과를 바탕으로 토픽 수를 5개로 설정하고, Gibbs 샘플링의 반복 횟수를 1000회로 설정하였다.

LDA 분석 결과 베타(beta)값의 경우 “beta: [1:5, 1:12454] -11.5 -9.2 -11.4 -11.5 -11.6”에서 보듯이 산출한 토픽과 단어간의 관련 확률을 나타내는 것으로 아래엔 5개의 토픽과 12,454개의 단어 간의 확률이 원데이터(raw data)로 산출되고 이를 지수(exponential) 함수로 변환시키면 단어가 갖는 토픽과의 관련성을 확률로 파악할 수 있다. 토픽별 상위의 단어를 베타값을 중심으로 ‘ggplot 패키지’를 사용하여 시각화한 것이 [그림 3] 토픽별 LDA 분석 결과 막대그래프이다. 감마(gamma)값은 “gamma: [1:739, 1:5] 0.5046 0.1569 0.1481 0.2222 0.0976”으로 문서와 토픽과의 관련성을 확률로 나타내고 총 739개의 문서에서 5개의 토픽간의 관련성을 나타낸다. 본 연구에서는 뉴스를 항목별로 분류해 탐구하지 않았으므로 감마값을 중심으로 분석하지 않았다.



[그림 3] 토픽별 LDA 분석 결과 - 토픽별 주제어와 막대그래프

아래 <표 5>에 [그림 3]에서 분석된 토픽별 LDA 분석 결과를 토대로 범주화하고 의미를 정리하였다. ‘문해력’은 ‘학업에 대한 결손’과 ‘디지털 기술의 발전’과 관련이 깊다. LDA 분석 결과 ‘Topic 1’ 그룹의 주제어는 ‘코로나, 글로벌, mz세대, 초등학생, 공동체, 사교육, 넷플릭스, 대한민국, 에너지’로 도출된 주제어의 의미는 ‘코비드(Covid)-19 팬데믹, 학생, 글로벌 OTT(Over The Top)’로 해석하였다. ‘Topic 2’ 그룹의 주제어는 ‘디지털, 데이터, 인공지능, 서비스, 메타버스, 전문가, 플랫폼, 일자리, 컴퓨터’로 도출된 주제어의 의미는 ‘디지털 기술, 디지털 리터러시’로 해석하였다. ‘Topic 3’ 그룹의 주제어는 ‘학생들, 아이들, 이야기, 스마트폰, 프로그램, 챗gpt, 사고력, 교과서, ebs, 독해력’으로 도출된 주제어의 의미는 ‘학생, 디지털 기기, 교육’으로 해석하였다. ‘Topic 4’ 그룹의 주제어는 ‘사람들, 유튜브, oecd, 우리나라, 어휘력, 인터넷, 우리말, 소비자, 한국어’로 도출된 주제어의 의미는 ‘언어, 인터넷, 사람’으로 해석하였다. ‘Topic 5’ 그룹의 주제어는 ‘미디어, 초등학교, 콘텐츠, 고등학교, 교육부, sns, 청소년, 중학교, 교육과정, 선택과목’으로 도출된 주제어의 의미는 ‘학교 교육, 청소년, 소셜 미디어’로 해석하였다.



&lt;표 5&gt; ‘문해력’ 중심의 토픽별 LDA 분석 결과

토픽 그룹의 이해		토픽 그룹의 주제어
코로나-19 팬데믹, 학생, 글로벌 OTT	Topic 1	코로나, 글로벌, mz세대, 초등학생, 공동체, 사교육, 넷플릭스, 대한민국, 에너지
디지털 기술, 디지털 리터러시	Topic 2	디지털, 데이터, 인공지능, 서비스, 메타버스, 전문가, 플랫폼, 일자리, 컴퓨터
학생, 디지털 기기, 교육	Topic 3	학생들, 아이들, 이야기, 스마트폰, 프로그램, 챗gpt, 사고력, 교과서, ebs, 독해력
언어, 인터넷, 사람	Topic 4	사람들, 유튜브, oecd, 우리나라, 어휘력, 인터넷, 우리말, 소비자, 한국어
학교 교육, 청소년, 소셜 미디어	Topic 5	미디어, 초등학교, 콘텐츠, 고등학교, 교육부, sns, 청소년, 중학교, 교육과정, 선택과목

### 3. 교육적 관점에서 바라본 토픽별 주제어

위 <표 5>의 ‘문해력 주제어 중심의 LDA 분석 결과’를 기반으로 다음과 같은 읽기 교육의 시사점을 정리해 보고자 한다.

#### 가. 디지털 리터러시 중심의 읽기 교육

‘디지털 기술, 디지털 리터러시, 플랫폼, 데이터’(Topic 2)는 학생들이 디지털 콘텐츠를 활용한 학습이 증가하고 있음을 보여준다. 이러한 강조된 디지털 환경에서 텍스트를 분석하고 신뢰할 수 있는 정보를 선별하는 ‘비판적 디지털 읽기 전략’이 중요하다. 구체적인 ‘읽기 전략’으로는 디지털 도구를 활용해 텍스트에서 핵심 내용을 요약하거나, 여러 소스의 정보를 비교·분석하는 활동을 생각해 볼 수 있다. 디지털 리터러시(digital literacy)는 단순히 첨단 디지털 기기나 프로그램을 능숙하게 활용하는 기술을 넘어서, 디지털 도구를 활용해 정보의 이해, 평가, 생산을 통해 학습자의 학습과 생활에 자유롭고 효율적으로 활용하여 ‘학업 성취와 삶의 풍요(well-being)’를 가져올 수 있는 방향으로 교육해야 한다. 동시에 학습자는 비판적 사고, 문제 해결 능력, 윤리적 책임을 모두 갖도록 지도한다(캐나다 브리티시 주, 2025). 디지털 리터러시는 Gilster(1997)에서 연구되어 캐나다 브리티시 주의 ‘디지털 리터러시와 인공지능(AI)의 사용(2025)’ 교육과정에서 인공지능(AI)에 대해 알고리즘 작동 방식, 사회 문화적 맥락에서 인공지능(AI)의 개념, 인공지능(AI)의 효과적인 학습적 활용법과 한계점, 인공지능(AI)을 활용시 윤리적 문제와 디지털 시민성 등을 강조한 교육 내용은 우리도 교육과정에 적극 도입할 필요가 있다. 또한 위 ‘Topic 2’의 주제어는 현대 디지털 환경에서 문해력이 정보의 평가와 활용을 중심으로 확장되어야 함을 시사한다. 디지털 기술의 발전이 정보 접근성을 높이는

동시에 정보의 양적 증가와 질적 편차를 초래하고 있음을 보여준다. 이에 따라, 디지털 리터러시의 중요성은 여러 학술 연구에서도 강조되고 있다. Gilster(1997)는 정보의 홍수 속에 허위 정보에 대한 비판적 이해를 기초로 하는 디지털 리터러시의 정의와 그 중요성을 명확히 하였고, Mackey & Jacobson(2011)은 디지털 도구를 활용한 정보의 평가 능력을 향상시키기 위한 교육적 접근이 필요함을 강조하였다. 아울러 디지털 시대에 소셜 미디어와 온라인 커뮤니티를 활용한 혁신적 협업과 소통을 강조하며 사용자가 필요한 정보를 비판적으로 판단하는 정보 리터러시 등을 복합적으로 융합한 메타 리터러시(Metaliteracy)를 소개하기도 한다. 이렇게 디지털 미디어에서 학습자가 학습과 진로 개발, 일상적 정보의 자유로운 생산자와 소비자로 활동할 수 있는 역량의 함양을 지원할 수 있는 교육이 디지털 리터러시 교육의 출발이다.

#### 나. 복합양식 텍스트 읽기 교육

‘넷플릭스, 유튜브, 소셜 미디어(SNS), 콘텐츠’(Topic 1, 5)는 영상 및 이미지 중심 매체가 학생들의 주요 정보 학습원임을 보여준다. 텍스트, 이미지, 영상이 결합된 형태의 ‘복합양식 텍스트 자료’를 활용해 다양한 형식의 정보를 분석하고 의미를 통합하는 교육이 필요하다. 구체적인 ‘읽기 전략’으로는 뉴스 기사와 동영상 자료를 결합한 읽기·토론 활동, 소셜 미디어 콘텐츠의 메시지를 비판적으로 분석하는 활동을 생각해 볼 수 있다. 영상, 텍스트, 이미지가 결합된 복합양식 텍스트가 현대인의 주요 정보 습득 경로임을 나타낸다. 복합양식 자료를 이해하고 분석하는 능력은 문해력의 새로운 구성 요소로 자리 잡고 있다. 학문적으로는 이러한 복합양식 텍스트가 독자의 이해에 미치는 영향을 실증적으로 연구하고, 이를 교육적 도구로 활용하는 방안을 모색할 필요가 있다. 교육적 가치가 높은 다양한 ‘복합양식 텍스트’는 문자 언어 중심 읽기와 달리 ‘보들읽기(Viewrauding, 서혁, 2023)’로 문자보다 ‘보여주는(showing) 특성이 크기(서혁 외, 2018ㄴ)’ 때문에 ‘학습자가 주도성을 갖고 읽어야 한다(서혁 외, 2018ㄱ)’.

Buckingham(2015)에서 학생들은 디지털 정보를 이해할 때 가져야 할 ‘비판적 사고(critical understanding)’에 주목해 그 정보는 사회 문화적 맥락 속에서 ‘재현(representation)’된 것으로 정보가 생산된 상업적 배경이나 편향성, 신뢰성을 평가하며 읽을 것을 강조하였다. 이어 디지털 매체에서 접하는 다양한 형식의 ‘복합양식 텍스트’의 이해와 생산을 강조하고 있어, 2022 국어과 교육과정에서 ‘매체’ 과목이 신설되어 ‘디지털 리터러시 교육’이 강조되고 있지만, 더욱 ‘복합양식 텍스트의 교육’을 강화할 필요가 있다. Buckingham(2015)은 디지털 정보를 이해할 때 비판적 사고와 함께 텍스트, 이미지, 영상을 결합한 방식으로 메시지를 분석하는 방법을 다음과 같이 제시한다. ‘학습자들이 매

체의 이면에 숨겨진 사회적, 정치적 맥락을 이해하고 더 나아가 창의적으로 활용할 수 있도록 지원이 필요하다. 디지털 매체는 특정한 가치관을 갖고 현실을 해석하고 선택적으로 제시할 수 있어 학습자가 정보의 출처나 제작자의 의도, 표현된 가치와 이념을 이해해야 한다. 디지털 매체를 다루는 도구적 능력을 넘어서는 정보를 평가하여 내가 필요한 지식으로 변환할 수 있는 고등 사고 능력이 요구된다(Buckingham, 2015).’ Buckingham (2015)은 “If you want to use television to teach somebody, you must first teach them how to use television(Umberto Eco, 1979).”을 제시하며 ‘디지털 매체 교육이 단순히 도구 활용 기능이 아닌 정보의 자유로운 활용 주체가 될 수 있도록 역량 있는 학습자의 교육’을 강조한다.

Rowse(2012)에서도 ‘복합양식 텍스트’ 중심의 디지털 텍스트를 비판적인 관점에서 학생들이 주도성을 갖고 협업할 수 있는 다양한 디지털 프로그램을 적극 활용해 ‘다양한 시각, 영상, 청각 등 복합양식 텍스트’를 이해하고 생산할 수 있도록 지도할 것을 설명하였다. Rowse(2013)은 ‘복합양식 텍스트(multimodal texts)’가 학생들의 읽기 능력에 미치는 영향에 대한 연구를 통해, 이러한 교육 방식이 학생들의 ‘비판적 사고 능력’을 강화한다고 주장하였다. 또한, ‘디자인 중심 사고(design epistemologies)’를 활용하여 학생들의 의미 생성 능력을 발전시키는 것이 중요함을 강조하였다. Rowse는 전통적인 인쇄물 중심 학습에서 벗어나 ‘디지털 및 상호작용적 학습 환경’으로의 전환을 제안하며, 텍스트, 이미지, 소리와 같은 다양한 모드 간의 ‘상호 연결성과 상호 의존성(interconnection and interdependence)’을 통해 복합양식 텍스트의 확장성을 강조하였다. 더불어, 학생들 간의 피드백과 상호작용을 통해 ‘협력적(collaborative) 학습 환경’을 조성하고, 복합양식 및 디지털 협업을 기반으로 한 ‘참여적(participatory) 프레임워크’ 학습의 중요성을 역설하였다. 이에 중등 교육에서도 학습자의 비판적 사고와 창의적 학습 역량에 대한 지원이 필요하다. 실시간 협업 도구(Google Docs, Padlet 등)를 활용해 공동 읽기 및 의견 교환 활동을 강화해야 한다. 한 가지 주제를 여러 자료를 통해 분석한 후 온라인 협업 도구에 결과를 정리하고 발표하는 활동을 생각해 볼 수 있다.

#### 다. 개별화 읽기 프로그램

‘학생들, 아이들, 사고력, 독해력’(Topic 3)은 학생 개인의 읽기 수준과 흥미를 고려한 맞춤형 읽기 교육의 필요성을 보여준다. 생성형 인공지능(AI) 기반 분석 도구를 활용해 개인의 읽기 성취도와 약점을 파악하고, 수준별 학습 자료를 제공해야 한다. 챗GPT와 같은 생성형 인공지능(AI) 도구를 활용해 ‘질문-답변 형태’로 독해력을 강화하는 활동이 효과적이다. 여기서 어떻게 질문을 던지느냐에 따라 다양한 답변이 제시되는데, 이

러한 프롬프트(prompt)를 효과적으로 제시하는 방법을 학습자가 잘 활용할 수 있는 ‘프롬프트 리터러시 교육(Fisher, Michael & Jacobs, Heidi Hayes, 2023, 이정미, 2023)’도 새롭게 도입할 필요가 있다. 프롬프트는 인공지능에게 질문을 던지는 기본 단위로서 라틴어 ‘promptus’에서 유래되었으며 ‘준비된, 즉각적 행동’의 의미가 있다. 컴퓨터에게 명령을 전달하는 입력어나 질문이라는 의미이다. 인공지능에게 어떤 ‘프롬프트’를 입력하느냐에 따라 매번 다른 출력값, 답변을 제시하기 때문에 ‘프롬프트 던지는 전략’도 학습하는 것이 필요하다. 학습자의 특성을 반영한 개인화된 학습 전략, 특히 생성형 인공지능(AI)을 활용한 맞춤형 교육 콘텐츠 개발이 중요하다. Shute & Zapata-Rivera(2010)는 개인화된 피드백을 제공하는 시스템이 학습자의 독해력 향상에 기여한다고 보고하였다. Eltahir & Babiker(2024)는 AI를 활용한 독해력 분석 도구가 학생들의 성취도를 향상시키는 데 도움이 된다고 주장하였다.

Rohit, et al.(2024)은 AI 기반 학습 시스템이 학생들의 개별적인 학습 요구를 충족시키는 데 효과적이라고 밝혔다. 아래 <표 6>에서 알 수 있는 AI 기반 학습 시스템의 장점은 ‘학습자의 학습 속도에 맞춰 개별화 교육을 할 수 있고, 다양한 복합양식 텍스트 교육 자원을 활용해 학습의 숙달이 충분히 될 때까지 지속적인 평가와 적응적 기술 활용’으로 ‘학습자가 주도성을 갖고 자기 점검’을 하며 학습할 수 있다.

<표 6> AI 기반 학습 시스템(Rohit, et al., 2024:814)

개인화 학습 시스템 (Personalized Learning System)	전통적 학습 시스템 (Traditional Learning System)
<ul style="list-style-type: none"> <li>• <b>개별화된 지도(Individualized Instruction):</b> 학생의 필요, 관심사, 학습 스타일에 맞게 수업과 활동 조정.</li> <li>• <b>유연한 속도(Flexible Pace):</b> 학생의 이해 속도에 따라 학습.</li> <li>• <b>다양한 자원(Resources):</b> 비디오, 기사, 시뮬레이션과 같은 다양한 자료 활용.</li> <li>• <b>지속적인 평가(Continuous Assessment):</b> 학습 지원을 제공하기 위해 지속적으로 평가.</li> <li>• <b>학습자 주도(Learner Agency):</b> 학생이 자신의 학습 경로를 선택할 수 있도록 지원, 동기와 참여 증대.</li> <li>• <b>실생활과의 연관성(Real-World Relevance):</b> 학생의 관심사 및 현실 세계와 연결.</li> <li>• <b>협력 학습(Collaborative Learning):</b> 학생들 간의 협력을 장려하여 서로 배울 수 있는 환경 제공.</li> <li>• <b>적응형 기술(Adaptive Technology):</b> 개별 학습 진행, 실시간 피드백 제공 기술 활용.</li> <li>• <b>숙달에 집중(Focus on Mastery):</b> 학습의 완전한 숙달에 초점.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>표준화된 교육과정:</b> 모든 학생에게 동일한 내용을 가르치는 교육과정.</li> <li>• <b>고정된 속도(Set Pace):</b> 모든 학생이 동일한 속도로 학습, 개별 이해도를 고려하지 않음.</li> <li>• <b>제한된 자원:</b> 주로 교과서와 강의에 의존하며 학습 자료가 제한적.</li> <li>• <b>주기적 평가(Periodic Assessment):</b> 시험과 같은 주기적 평가 사용, 즉각적인 피드백 부족.</li> <li>• <b>교사 중심(Teacher-Centered):</b> 학습 과정에서 교사의 역할이 주도적이며, 학생의 참여가 적음.</li> <li>• <b>일률적 접근(Uniform Approach):</b> 모든 학생이 동일한 방식으로 학습한다고 가정.</li> <li>• <b>제한된 협력(Limited Collaboration):</b> 학생들 간의 협력 기회 부족.</li> <li>• <b>최소한의 기술 통합(Minimal Technology Integration):</b> 콘텐츠 전달에 초점, 학습 개인화에는 활용하지 않음.</li> <li>• <b>완료에 집중:</b> 학습의 깊이는 이해보다는 과제 완료와 진도에 초점.</li> </ul>

## 라. 어휘 및 언어 중심 통합적 문해력 교육

‘어휘력, 우리말, 한국어, 독해력’(Topic 4)은 기초 문해력의 핵심 요소인 언어적 능력을 강화할 필요성을 시사한다. 다양한 주제의 읽기 자료를 활용해 학생들이 문맥 속에서 어휘의 의미를 학습하고, 이해를 확장하도록 지도해야 한다. 다양한 교과와 글을 읽으며 문맥에 따른 어휘의 의미 변화나 어감 차이를 학습하게 한다. 문해력의 핵심 요소로서 어휘 능력을 강화해야 함을 보여준다. 텍스트와 공동체 문맥 기반 어휘 학습이 독해력 향상에 미치는 영향을 고려하여, 다양한 교과에서 어휘 사용이 문해력과 학업 성취에 기여할 수 있는 교육적 지원이 필요하다. Snow(2010)는 어휘력과 독해력 간의 상관관계를 규명하였으며, Beck et al.(2013)는 어휘력 교육이 학생들의 독해력 향상에 중요한 역할을 한다고 밝혔다. 또한, Graves, M. F.(2016)는 어휘력 향상을 위한 교육적 접근이 학생들의 학업 성취에 직접적인 영향을 미친다고 주장하였다. 이러한 어휘 교육을 중심에 둔 통합적 문해력 교육 프로그램을 설계하고 실행해야 할 것이다.

## 마. 문해력 향상과 연계된 주제별 읽기 활동

‘글로벌, 공동체, 에너지’(Topic 1) 및 ‘oecd, 청소년, 교육과정’(Topic 4, 5)은 읽기 교육을 통해 사회적 이슈와 글로벌 주제에 대한 이해를 높일 수 있음을 보여준다. 환경, 사회 정의, 글로벌 문제 등 복합적이고 다차원적인 사회적 쟁점을 다룸으로써 학습자의 비판적 사고력과 공감 역량을 통합적으로 증진할 수 있는 교육적 설계가 요구된다. 교육적 관점에서는 환경, 다문화, 사회 정의와 같은 현대적이고 시의성 있는 사회적 쟁점을 중심으로 한 읽기 자료를 선정하고, 이를 학습자의 사회적 책임 의식, 비판적 사고력, 그리고 윤리적 판단 능력을 심화시키는 도구로 활용하는 전략이 필요하다. Murnane, et al.(2012)는 글로벌 이슈를 다룬 텍스트가 학습자의 사회적 공감 능력과 비판적 사고를 고양하는 데 효과적이라고 주장하였으며, Liu et al.(2020)은 환경 문제와 관련된 텍스트가 학습자의 환경적 책임감을 제고하는 데 기여함을 보고하였다. 또한, Willis(2017)는 사회적 이슈를 다룬 텍스트의 분석적 학습이 학습자의 사회적 참여 의식과 시민적 책무성을 강화하는 데 중요한 역할을 한다고 언급하였다. 예를 들어, 글로벌 OTT(Over The Top) 플랫폼에서 제공하는 다문화 관련 다큐멘터리를 보고 관련 텍스트를 읽고 토론하는 활동을 생각해 볼 수 있다.

## 바. 전통적 읽기와 디지털 읽기의 균형

‘교과서, ebs, 초등학교, 고등학교’(Topic 3, 5)는 여전히 전통적 읽기 교육이 중요하

다는 점을 시사한다. 디지털 콘텐츠 활용과 더불어 종이 교과서 및 학습 자료를 통해 깊이 있는 읽기와 사고를 병행해야 한다. 전통적 종이 교과서와 디지털 텍스트 읽기를 학습 목표와 글의 주제에 따라 병행하는 혼합 읽기 활동을 생각해 볼 수 있다. 전통적 읽기를 통해 학습자의 심층적 이해, 비판적 사고, 논리적 추론 능력을 강화하는 동시에, 디지털 읽기의 장점을 활용하여 학습 접근성, 흥미, 멀티태스킹(multi-tasking) 능력을 증진하는 균형 있는 교육적 틀을 제공할 수 있다. 전통적 읽기가 독자의 심층적 이해와 비판적 사고에 미치는 영향과 디지털 읽기 방식이 독자의 학습 접근성과 흥미를 높이는 방식 간의 균형을 탐구해야 한다. Greenfield(2014)는 전통적 읽기가 학습자의 심층적 사고와 인지적 숙고를 유도한다고 밝혔으며, Liu(2005)는 디지털 매체가 정보 처리와 학습 효율성에 긍정적인 영향을 미친다고 주장하였다. Carr(2020)는 디지털 환경이 학습자의 사고방식과 주의 지속성에 영향을 미칠 수 있음을 논의하며, 전통적 읽기와 디지털 읽기 간의 균형 있는 접근의 필요성을 강조하였다.

#### 사. 사회적 참여와 연계된 읽기 활동

‘공동체, 사람들, 소비자’(Topic 1, 4)는 읽기 활동이 사회적 맥락 속에서 실행될 수 있음을 나타낸다. 읽기 교육은 단순히 텍스트 해석 능력을 강화하는 것을 넘어, 학습자들이 사회적 문제를 인식하고, 비판적 사고를 통해 실질적인 행동으로 이어질 수 있도록 설계되어야 한다. 읽기 교육은 학습자의 문해력을 사회적 실천과 결합하여, 텍스트를 통해 비판적 사고와 공감 능력을 강화하고, 이를 기반으로 공동체와 연결되어 실질적인 행동 변화를 유도할 수 있는 방향으로 설계되어야 한다. 이러한 접근은 문해력이 학습자의 전반적 인지적, 정서적, 사회적 역량을 함양하여 궁극적으로 책임감 있는 시민을 양성할 수 있는 배경이 될 수 있다. Vygotsky(1978)의 사회적 상호작용 이론이 읽기 활동과 학습자 간 상호작용의 중요성을 강조하며, Green & Green(1997)은 읽기 활동이 지역사회 문제 해결과 사회적 참여를 촉진한다고 주장하였다. Schleicher(2019)는 읽기 교육이 학습자의 사회적 책임감과 시민 의식을 고양하는 데 중요한 역할을 한다고 밝혔다. 학생들이 읽기 활동을 통해 지역사회 문제 해결이나 사회적 참여로 연결할 수 있는 프로젝트 기반 학습이 필요하다. 환경 보호 관련 자료를 읽고 지역 캠페인을 기획해보는 활동을 생각해 볼 수 있다.

## IV. 결 론

교육의 시각에서 ‘문해력’은 ‘학습자의 행복한 학창 시절과 진로 개발에 성공한 성인이 되기 위한 핵심적 열쇠’라 할 수 있다. “독서는 ‘체득(體得)적 삶’을 살아가는 성찰적 존재로서의 인간이 ‘인간답게’ 살아가도록 하는 핵심 행위라는 점(서혁, 2023:10)”을 교육의 상황에서도 상기해야 한다. 본 연구에서 ‘뉴스 기사’라는 우리 사회의 담론을 탐색한 결과 ‘디지털 기술, 디지털 리터러시, 플랫폼, 데이터’(Topic 2)라는 것이 중요한 화두이며 다양한 ‘넷플릭스, 유튜브, 소셜 미디어(SNS), 콘텐츠’(Topic 1, 5) 자원을 활용한 ‘교과서, ebs, 초등학교, 고등학교’(Topic 3, 5)이 주요 주제어라는 점에 주목할 필요가 있다.

본 연구에서는 코비드-19 팬데믹 사태 이후 ‘문해력 중심으로 국내 뉴스 기사를 텍스트 마이닝’하여 ‘토픽별 주제어’를 도출하였다. 학업 역량의 신장을 위해 학생들이 평소 자주 접하는 사진, 영상처럼 ‘다양한 복합양식 텍스트 이해 전략’을 강조하였고, 에듀테크를 잘 활용할 수 있도록 다양한 디지털 학습 프로그램의 개발이 필요함을 제안하였다. 4차 산업혁명 이후 급속히 발전하고 있는 디지털 기술로 인해 학생들은 영상 텍스트를 일상적으로 접하고 있다. 학교 교육에서 학교에서 ‘개발된 다양한 복합양식 텍스트 읽기 전략을 지도하고 학생들이 연습할 수 있는 기회’를 제공하는 것이 필요하다. Schwab (2016)은 ‘4차 산업혁명’ 시대 디지털 혁명은 첨단 로봇공학(advanced robotics), 유전자 편집(gene editing) 등 물리적, 디지털, 생물학적 영역에서 혁신을 가져올 것으로 예상했다. 이렇게 권정숙(2014)이 꽤 오래전에 지적해온 디지털 기술이 나날이 급속도로 발전하면서 학습자가 접하게 되는 다양한 미디어 텍스트도 증대되면서 학교 교육에서도 이러한 변화에 맞는 ‘문해력’ 교육이 절실하다. ‘다양한 복합양식 텍스트 읽기 교육이 심층적으로 연구되어 현장 학교에서 실시되도록 교육 여건’을 마련하는 것이 필요하다. 청소년들이 코비드-19 팬데믹 사태 이전에도 인터넷의 정보 탐색이나 활용이 익숙했었는데, 코비드-19 팬데믹 사태 이후 언택트 교육이 보편화된 이후 더욱 강화되었다고 볼 수 있다. 학교 교육의 정규 교육과정에서 ‘매체 교육’ 영역을 신설하여 이를 활성화하려는 점은 고무적이다. 하지만 독립된 영역으로 교육하는 것은 ‘학습 내용의 전이나 유용성’ 면에서 한계가 있을 수 있다.

문해력 향상을 위해 학교 현장에서 노력해야 한다는 것은 재론의 여지가 없다. 다만 학생들의 흥미와 참여도를 고려하여 평소 학생들이 자주 접하는 영상, 시각 자료인 복합양식 텍스트 읽기 교육에 더 중점을 둘 필요가 있다. 이에 국어 교과에서 디지털 교과서를 도입하는 문제에 대해 긍정적인 면을 극대화할 수 있는 방안이 절실하다. 문해력 향상을 위해 학생 수가 급속히 감축되는 문제에 대한 대안의 하나로 하나의 교실에 교사

들이 2인 이상 참여하여 수준별 수업이 가능해진다면 어떨까 한다. 김은혜, 서혁(2022)이 ‘현대 사회에 필요한 비판적 읽기 교육을 위한 실천이란 성찰을 토대로한 의미 구성이자 현실 세계의 다양한 문제 해결을 수행하기 위한 사회적 행동’으로 제안한 점을 주목하여 실제적 교수학습이 되도록 교육 현장의 지원이 필요하다. 학습자의 자율적 학습 선택권을 존중하기 위해 하나의 교실에서 다양한 교육 활동이 이루어지는 것이 바람직하다. 읽기 교육은 전통적 텍스트 읽기와 디지털 기술 활용을 결합한 방식으로 발전해야 하며, 학생 개인의 흥미와 수준을 고려한 맞춤형 교육이 중요하다. 또한, 사회적 이슈와 연계된 읽기 활동을 통해 학생들의 비판적 사고력과 공감 능력을 함께 계발할 필요가 있다. 마지막으로 생성형 인공지능 기술과 인간이 함께 공존하는 시대, 학교 교육은 교과 문해력(Disciplinary Literacy)을 활용해 학생 스스로 지식을 깊이 탐구하고 비판적으로 사고하는 읽기 교육에 관심을 기울일 필요가 있다. 복합양식 텍스트를 ‘보들읽기(서혁, 2023)’처럼 실제적인 전략으로 능숙하게 읽는 독자는 ‘호모 루덴스(Homo Ludens, Huizinga, 1938)로서 상상의 기쁨과 호모 파베르(Homo Faber, Bergson, 1907)로서 노동의 기쁨을 향유할 수 있는 진정한(authentic) 지적 자유와 힘’을 갖게 되지 않을까.



## 참고문헌

- 교육부(2015). 국어과 교육과정(교육부 고시 제2015-74호 [별책 5]). 교육부.
- 교육부(2022). 국어과 교육과정(교육부 고시 제2022-33호 [별책 5]). 교육부.
- 교육부(2022). 제1차 기초학력 보장 종합계획.
- 강성국(2020). 코로나 19 대응 국가 수준 원격교육체제 진단 및 과제(IP2020-02). 한국교육개발원 연구보고서.
- 김성연(2020). 토픽모델링을 활용한 SIAM Journal on Applied Mathematics 의 연구 동향 분석. 한국산학기술학회 논문지, 21(7), 607-615. 한국산학기술학회.
- 김은정, 장석권, 이상용(2023). 토픽모델링과 시계열 회귀분석을 활용한 헬스케어 분야의 뉴스 빅데이터 분석. 경영정보학연구, 25(3), 163-177. 한국경영정보학회.
- 김은혜, 서혁(2022). 비판적 읽기 교육에서의 ‘실천’의 의미 연구. 독서연구, 65, 201-229.
- 김세영(2023). TF-IDF 분석과 토픽 모델링을 활용한 AI 기반 개별화 학습 국내외 연구동향 분석. 정보교육학회논문지, 27(4), 453-464. 한국정보교육학회.
- 김예지, 김갑수(2023). 초등학생을 위한 빅데이터 교수·학습 모형 개발 : 미래 역량을 중심으로. 정보교육학회논문지, 27(4), 465-475. 한국정보교육학회.
- 김선아, 이삼형, 김종우, 김시정, 김동성(2015). 융복합 교육을 위한 미술과-국어과 교육과정 분석 연구. 미술교육논총, 29(1), 1-24. 한국미술교육학회.
- 김수환, 류승완(2022). ESG와 문화예술의 관계 및 이슈 분석 : 뉴스 데이터를 활용한 LDA 토픽모델링 분석을 중심으로. 문화예술경영학연구, 15(2), 53-73. 한국문화예술경영학회.
- 김종윤, 변태진, 이해영(2018). 역사 및 물리 교사의 텍스트 읽기 양상 비교를 통한 학문 문식성 개념의 교육적 적용 방안 탐색, 교육과정평가연구 21(4), 43-72. 한국교육과정평가원.
- 김혜진(2021). 독도에 대한 초등학생의 인식 분석-심상도 분석 및 언어네트워크 분석을 중심으로. 사회과교육, 60(2), 115-134. 한국사회과교육연구학회.
- 나경식, 이지수(2018). 신문빅데이터를 바탕으로 본 국내 정보화의 경향과 도서관의 역할. 한국콘텐츠학회논문지. 제18권. 제9호, 14-33. 한국콘텐츠학회.
- 남소희, 전수영(2019). 점별 상호정보량을 이용한 LDA 토픽모형 추론. Journal of The Korean Data Analysis Society, 21(6), 2789-2800. 한국자료분석학회.
- 남유선, 김인숙(2015). 독일의 개방형 의사소통 시스템 “플랫폼” -독일 제4차 산업혁명을

- 중심으로. **독일언어문학** 제70권, 47-66. 한국독일언어문학회.
- 노들(2021). 텍스트 마이닝을 활용한 국내 디지털 리터러시 연구 동향 분석-국어교육에 주는 시사점을 중심으로. **국어교육**, 175, 199-238. 한국어교육학회.
- 방미현, 이영민.(2020). 20대 청년세대에 관한 연구 동향 분석. **한국콘텐츠학회논문지**, 20(7), 223-232. 한국콘텐츠학회.
- 박대민(2016). 장기시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성: 100만 건 기사의 정보원과 주제로 본 신문 26년, **한국언론학보**, 제60권, 제5호, 353-407. 한국언론학회.
- 박숙경(2023). 토픽 모델링을 활용한 건강문해력 국내연구동향 분석: 2012~2021년 학술지 게재 논문을 중심으로. **인문사회** 21, 14(3), 1185-1200.
- 박지원(2023). 소셜 빅데이터를 활용한 청소년정책 분석 : 텍스트 마이닝 분석을 중심으로. 중앙대학교 대학원 아동청소년학과 청소년전공 박사학위논문.
- 박지우, 박하나, 정성희, 임규연(2022). 텍스트 마이닝을 활용한 국내 개별화 학습 연구동향 분석: TF-IDF, LDA 기반 텍스트 마이닝을 중심으로. **교육방법연구**, 34(4), 711-739. 한국교육방법학회.
- 배경임(2023). 텍스트 마이닝과 네트워크 분석에 의한 독서치료 연구 동향의 비교 분석 연구. **차세대융합기술학회논문지**, 7(10), 1685-1696. 국제차세대융합기술학회.
- 서혁(2023). 독서 사회의 변화와 새로운 독서 교육 : 독서 환경의 변화와 교육적 대응을 중심으로. **독서연구**, 68(0), 9-33. 한국독서학회.
- 서혁, 변은지, 한지수, 편지윤(2018ㄱ). 복합양식 텍스트의 텍스트성 재개념화를 위한 시론. **학습자중심교과교육연구**, 18, 493-522. 학습자중심교과교육학회.
- 서혁, 한지수, 변은지, 편지윤(2018ㄴ). 복합양식 텍스트성으로서의 영상성의 의미 고찰. **독서연구**, 49, 125-160. 한국독서학회.
- 성태제(2017). 제4차 산업혁명시대의 인간상과 교육의 방향 및 제언. **교육학연구**, 55(2), 1-21. 한국교육학회.
- 손지현, 이해선(2021). 의미연결망 분석을 통한 국내 시각문화교육 연구 동향 분석. **미술교육논총**, 35(4), 69-100. 한국미술교육학회.
- 안현용(2023). 텍스트마이닝을 이용한 국내의 디지털 평생교육 연구 동향 분석. **평생교육학연구**, 29(3), 105-135. 한국평생교육학회.
- 이우진, 백혜진(2022). 키워드 네트워크 분석을 통한 리터러시 교육 연구 동향. **디지털융복합연구**, 20(5). 한국디지털정책학회.

- 이승석(2023). 텍스트 마이닝을 활용한 ‘혁신학교’에 대한 인식분석. 국내석사학위논문 한  
국교원대학교 교육대학원, 충청북도.
- 이정미(2023). ChatGPT, 생성형 AI 시대 도서관의 데이터 리터러시 교육에 대한 연구. 한  
국문헌정보학회지, 57(3), 303-323. 한국문헌정보학회.
- 이찬중(2023). 코로나19 유행시기에 따른 키워드 네트워크 분석. 국내석사학위논문 고려  
대학교 보건대학원, 서울
- 이찬승(2016). 미래사회 변화와 진로교육의 새로운 방향. 한국진로교육학회 학술대회지(),  
1-31.
- 이현주(2023). 텍스트 마이닝을 활용한 외국인 유학생 관련 신문 기사 분석. 새국어교육.  
(), 545-583. 한국국어교육학회.
- 이혜인(2023). 학습요소를 중심으로 ‘인공지능 기초’ 교과서 분석. 국내석사학위논문 성균  
관대학교 교육대학원, 서울
- 이경진, 최나영, 강주희(2023). 문화예술교육 프로그램과 학교교육과정 연계 활성화 방안  
탐색: 텍스트 마이닝을 활용하여. 예술교육연구, 21(3), 115-136. 한국예술교육학회.
- 이수진(2022). 코로나 시대 교육의 변화 양상 분석 및 과제 탐색. 국내박사학위논문 경남  
대학교 대학원, 경상남도.
- 이정우 외(2021). 2020 교육분야 코로나19 대응. 교육부. 『기술보고』, TR 2021-28.  
진천: 한국교육개발원.
- 장경숙 외(2020). 국외 COVID-19 학교 교육 대체 방안 사례 분석을 통한 후속 추진  
과제. 진천: 한국교육과정평가원.
- 장성민(2019). 국어교육과 “교과 논리 및 논술” 교과목의 교육 내용 체계화를 위한 기초  
연구-2009~2018년 강의계획서에 대한 텍스트 마이닝을 중심으로-. 작문연구,  
40(0), 7-43. 한국작문학회.
- 장미경(2022). 한국어 읽기 텍스트 난이도 연구의 동향과 과제. 돈암어문학, 41, 321-347.  
돈암어문학회.
- 조남옥, 조규락(2023). 텍스트 마이닝을 활용한 창의적 체험활동의 연구동향 분석. 한국교  
육학연구(구 안암교육학연구), 29(3), 205-224. 안암교육학회.
- 조수진(2023). 비정형 텍스트 빅데이터 분석을 통한 유아 지원 정책에 대한 연구 : 오피니  
언 마이닝 기법을 활용하여. 차세대융합기술학회논문지, 7(10), 1659-1672. 국제차  
세대융합기술학회.
- 최규리, 오윤정(2023). 텍스트 마이닝을 활용한 초등환경교육 연구 동향 탐색. 환경교육,

- 36(3), 191–206. 한국환경교육학회.
- 추형석, 조원영, 유재홍(2023). 언어 모델과 텍스트 마이닝을 활용한 인공지능 신뢰성 정책 이슈 분석 방법에 관한 연구. *기술혁신학회지*, 26(4), 572–593. 한국기술혁신학회.
- 황규호(2020). 포스트 코로나 시대 국가교육과정의 과제. *교육과정연구*, 38(4), 83–106. 한국교육과정학회.
- 황홍섭(2019). 빅데이터를 활용한 사회과 교수·학습 모형의 탐색. *사회과교육*, 58(1), 63–98. 한국사회과교육연구학회.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.
- Bergson, H. (1907). *L'évolution créatrice, paris*. Librairie Félix alcan.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Blei, D. M.(2013). *Probabilistic topic models*. 2013 Topic Modeling Workshop at NIPS.
- Boyd, T. (2021). Education reform in Ontario: Building capacity through collaboration. *Implementing Deeper Learning and 21st Century Education Reforms: Building an Education Renaissance After a Global Pandemic*, 39–58.
- Buckingham, D. (2015). Defining digital literacy—What do young people need to know about digital media?. *Nordic journal of digital literacy*, 10(Jubileumsnummer), 21–35.
- Carr, N. (2020). *The shallows: What the Internet is doing to our brains*. WW Norton & Company.
- Eltahir, M. E., & Babiker, F. M. E. (2024). The Influence of Artificial Intelligence Tools on Student Performance in e-Learning Environments: Case Study. *Electronic Journal of e-Learning*, 22(9), 91–110.
- Fisher, Michael & Jacobs, Heidi Hayes(2023). PROMPT LITERACY: A Key for AI-Based Learning. *Educational Leadership*. Jun2023, Vol. 80 Issue 9, 18–19.
- Gee, J. P. (1990). *Social linguistics and literacies: Ideology in discourses*. New York: Falmer.
- Gilster, P. (1997). *Digital literacy*. Seoul Heanaem.
- Government of British Columbia(2025). *Curriculum Connections: Digital Literacy and the Use of AI*. (C:/Users/USER/Downloads/curriculum-connections-digital-literacy-and-the-use-of-ai%20(3).pdf)

- Graves, M. F. (2016). *The vocabulary book: Learning and instruction*. Teachers College Press.
- G. Rani, T. Sharma and A. Sharma(2023), *Future Database Technologies for Big Data Analytics*, *International Conference on Intelligent Systems for Communication, IoT and Security* (ICISCOIS), Coimbatore, India, 349–354.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl 1), 5228–5235.
- Green, A., & Green, A. (1997). *Education, globalization and the nation state* (pp.130–186). Palgrave Macmillan UK.
- Greenfield, P. M. (2014). *Mind and media: The effects of television, video games, and computers*. Psychology Press.
- Hague, C. (2010). *Digital Literacy Across the Curriculum*.
- Howell, E., Barlow, W., & Dyches, J. (2021). *Disciplinary literacy: Successes and challenges of professional development*. Journal of Language & Literacy Education.
- Huizinga, J. (1938). *Homo Ludens: Proeve eener bepaling van het spel-element der cultuur*. Haarlem: Tjeenk Willink.
- Hurd, P. D. (1969). *New directions in teaching secondary school science*. Chicago : Rand McNall.
- Janks, H. (2009). *Literacy and power*. Routledge.
- Jenkins, C. (2023). *Literacy in science: exploring barriers to literacy in secondary school science and the impact of an in-class intervention* (Doctoral dissertation, University of Oxford).
- Kwon, Sang-mi, Jeong, Cheol Ja.(2023). *Public Perception of Interpreters in South Korea : Text Mining Social Media*. Interpreting and Translation Studies, 27(3), 1–25.
- Lent, R. C. (2015). *This is disciplinary literacy: Reading, writing, thinking, and doing... content area by content area*. Corwin Press.
- Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of documentation*, 61(6), 700–712.
- Liu, H., Wu, H., & Lyu, Z. (2020). Nonlinear resonance of FG multilayer beam-type nanocomposites: effects of graphene nanoplatelet-reinforcement and geometric imperfection. *Aerospace Science and Technology*, 98, 105702.

- 노현경(Ro, Hyun Kyoung), Ramon, Elizabeth.(2021). U.S. Higher Education Institutions: Challenges and Opportunities Since COVID-19. *비교교육연구*, 31(5), 61-85.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity. Big Data: The Next Frontier for Innovation, Competition & Productivity*, 1 - 143.
- Mackey, T. P., & Jacobson, T. E. (2011). Reframing information literacy as a metaliteracy. *College & research libraries*, 72(1), 62-78.
- Moje, E. B.(1996). "I teach students, not subjects": Teacher-student relationships as contexts for secondary-literacy, *Reading Research Quarterly*, 31(2), 172-195.
- Moje, E. B.(2000). *All the Stories That We Have : Adolescents' Insights About Literacy and Learning in Secondary School*, Newark, Del. : International Reading Association.
- Moje, E. B., & Speyer, J. (2008). *The reality of challenging texts in high school science and social studies : How teachers can mediate comprehension*.
- Moje, E. B.(2015). Doing and Teaching Disciplinary Literacy with Adolescent Learners: A Social and Cultural Enterprise, *Harvard Educational Review*, 85(2), 254-278.
- Murnane, R., Sawhill, I., & Snow, C. (2012). Literacy challenges for the twenty-first century: Introducing the issue. *The future of children*, 3-15.
- Ontario Ministry of Education(2022). *The Ontario Curriculum, Grades 9-12: Science*. ([https://assets-us-01.kc-usercontent.com/fbd574c4-da36-0066-a0c5-849ffb2de96e/a393c5f1-4619-4c34-877a-c22321d01e18/The%20Ontario%20Curriculum%20-%20Science\\_Grade%209\\_De-streamed%20Course\\_2022.pdf](https://assets-us-01.kc-usercontent.com/fbd574c4-da36-0066-a0c5-849ffb2de96e/a393c5f1-4619-4c34-877a-c22321d01e18/The%20Ontario%20Curriculum%20-%20Science_Grade%209_De-streamed%20Course_2022.pdf))
- 캐나다 온타리오주 사회 과학 및 인문학(Social Sciences and Humanities) 9-12학년의 교육과정(2013) (<https://www.edu.gov.on.ca/eng/curriculum/secondary/ssciences9to122013.pdf>)
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1 - 6.
- Kaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25-29.
- Rowell, J. (2012). Maureen Walsh, *Multimodal Literacy: Researching Classroom Practice*. *Journal of Early Childhood Literacy*, 12(2), 223-227.
- Rowell, J. (2013). *Working with multimodality: Rethinking literacy in a digital age*.

- Routledge.
- Röder, M., Both, A., & Hinneburg, A. (2015). *Exploring the space of topic coherence measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*, 399–408.
- Rohit, N., Aman, D., Jay, T., Smit, C., & Rina, S. (2024). AI Based Personalized Learning System. *International Journal of Trend in Scientific Research and Development*, 8(5), 811–816.
- Schwab, K. (2016). *The fourth industrial revolution*. New York, NY : Crown Business
- Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD Publishing.
- Shanahan, T., & Shanahan, C.(2008). Teaching Disciplinary Literacy to adolescents: Rethinking content–area literacy. *Harvard Educational Review*, 78(1), 40–59, 279.
- Shanahan, T., & Shanahan, C.(2012). What is Disciplinary Literacy and why does it?, *Topics In Language Disorders*, 32(1), 7–18.
- Shanahan, T., & Shanahan, C.(2014). Does Disciplinary Literacy Have A Place in Elementary School?, *The Reading Teacher*, 67(8), 636–639.
- Shute, V. J., & Zapata–Rivera, D. (2010). *Educational measurement and intelligent systems. of the International Encyclopedia of Education*. Oxford, UK: Elsevier Publishers, 4, 75–80.
- Shannon, P. (2014). *Reading poverty in America*. Routledge.
- Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *science*, 328(5977), 450–452.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.
- World Bank.(2020). *The COVID–19 Pandemic: Shocks to Education and Policy Responses*. World Bank, Washington, DC.
- Wineburg, S. S. (1990). *Historical problem–solving: A study of the cognitive processes used in the evaluation of documentary evidence*. Stanford University.
- Willis, P. (2017). *Learning to labour: How working class kids get working class jobs*. Routledge.

ABSTRACT

## A Discourse Study on Literacy Using LDA Topic Modeling: Focusing on News Data

Yoon-Kyoung Park

Teacher, Yeoncheon Middle School

This study aims to analyze the societal discourse surrounding the keyword ‘literacy’ and related topics in domestic news articles following the COVID-19 pandemic, interpreting the findings from an educational perspective. Utilizing the Big Kinds platform, we collected news article data from 54 domestic media outlets spanning from January 1, 2020, to November 18, 2023. The collected data was preprocessed using the KoNLP and NIADic packages in R to conduct a Korean text analysis, followed by text mining using the LDA topic modeling method.

The analysis identified “digital technology, digital literacy, platforms, and data”(Topic 2) as the central theme. Additionally, the significance of leveraging digital resources was underscored through topics such as “Netflix, YouTube, social media(SNS), and content” (Topics 1 & 5), while the necessity of educational approaches was highlighted through topics including “textbooks, EBS, elementary school, and high school”(Topics 3 & 5). The rapid expansion of the digital environment post-pandemic has heightened the importance of multimodal text reading, emphasizing the essential role of literacy enhancement and the development of critical thinking skills among learners.

Based on these findings, this study proposes strategies to strengthen digital literacy by designing personalized learning environments through digital textbooks and EdTech, systematically developing and applying multimodal text reading strategies, and reinforcing media education. It is imperative to support adolescents in critically analyzing and strategically utilizing multimodal texts that are familiar to them in the digital age while also fostering their Disciplinary Literacy, through well-structured educational scaffolding.

*Key Words:* Multimodal Texts, Disciplinary Literacy, LDA topic modeling