

빅데이터를 통해 살펴본 다문화가정에 대한 사회적 인식 탐색

- LDA 토픽모델링 분석을 중심으로 -

이강훈* · 정수정**

투고일: 2023.08.03. 심사일: 2023.11.17. 게재확정일: 2023.12.01.

목 차

I. 서 론	III. 연구결과
II. 연구방법	IV. 논의 및 결론

요약 본 연구의 목적은 빅데이터를 활용하여 다문화가정에 대한 사회적 인식을 살펴봄으로써 좀 더 효과적인 다문화가정 지원을 위한 기초자료를 제공하는데 있다. 이러한 목적을 달성하기 위하여 TEXTOM에서 제공하고 있는 빅데이터에 기반해 ‘다문화가정’이라는 키워드를 중심으로 rawDATA를 수집 및 분석하였다. 데이터 수집대상은 1차·2차 데이터 정제과정을 거친 데이터 중에 빈도 상위 200개 키워드를 핵심키워드로 선정하고, 이를 1-mode matrix dataset로 변환한 후 키워드 네트워크 분석을 실시하였다. 또한 알고리즘을 바탕으로 머신러닝 분석을 통해 토픽이 자동적으로 도출 가능한 LDA(latent dirichlet allocation) 토픽모델링 분석을 실시하였다. 연구결과, 다문화가정에 대한 사회적 인식은 이해관계 집단, 이해관계자, 정책 및 행정, 사회적 반응, 학문적 관심 등에 관한 내용들을 포함하고 있는 것으로 나타났다. 또한 다문화가정에 대한 토픽모델링 분석 결과, 3개의 토픽(지역 맞춤형 다양한 다문화가정 지원 혜택에 대한 요구, 사회 맞춤형 다문화가정 지원 프로그램에 대한 요구, 가족 중심 다문화가정 지원 방안 마련을 위한 연구에 대한 요구)이 도출되었다. 본 결과를 바탕으로 다문화가정 지원을 위한 다양한 고려사항들을 확인할 수 있었을 뿐만 아니라 효과적으로 다문화가정 지원 방안 마련을 위한 의미 있는 시사점을 제공하였다.

주제어: 빅데이터, 다문화가정, 사회적 인식, LDA 토픽모델링 분석, 키워드 네트워크 분석

* 제1저자, 교육빅데이터연구소 소장, E-mail: darkengal@nate.com

** 교신저자, 신라대학교 사회복지학과 조교수, E-mail: jsj@silla.ac.kr

I. 서 론

콩고에서 온 조나단과 파트리샤 남매가 최근 TV 공중파 방송에서 큰 인기를 얻고 있다. 조나단과 파트리샤 남매는 2013년 가족들과 KBS 인간극장에 출연하며 처음 얼굴을 알렸고, 그 후 MBC 라디오스타·전지적 참견시점 등 각종 예능프로그램에 출연하여 만능크리에이터의 면모를 보여주며 시청자들에게 많은 인기를 얻고 있다. 특히 조나단은 현재 2023년 8월 기준으로 유튜브 채널구독자가 80만을 넘기고 있어 예능대세의 면모를 거침없이 선보이고 있다. 최근 우리나라의 경우 외국인 출입국자 수가 크게 늘어났을 뿐만 아니라 다문화가정이 크게 증가하였다. 이에 따라 조나단과 파트리샤 남매처럼 우리와 다른 존재처럼 느껴졌던 다문화가정의 구성원들이 한국사회에 조화로운 공존하는 것에 많은 호감과 사회적 관심을 보여주고 있다. 이처럼 다문화라는 요소가 유튜브라는 플랫폼을 통해 SNS의 주된 소비자인 청소년을 만나게 되면서 다문화에 대한 사회적 관심이 높아지는 추세이다.

우리나라의 인구구조에도 변화가 일고 있다. 최근 이루어진 2022년 인구주택총조사 결과를 살펴보면 한국의 총 인구는 51,692,272명으로 전년 대비 45,799명이 감소하여 2년 연속 감소세를 보이는 상황인 반면, 한국에서 3개월 이상 거주한 외국인이 1,752,346명으로 전년대비 102,379명이나 증가한 것으로 나타났다(통계청, 2022). 이러한 변화의 결과로 이제는 감소한 내국인 수를 외국인으로 채워 나가야하는 사회적 구조로 변화하고 있음을 의미한다. 한국의 초저출생과 인구감소가 어제오늘의 문제가 아닌 만큼 내국인의 감소와 외국인의 증가 추이는 앞으로 더욱 심화 될 가능성이 높다. 즉, 한국은 다양한 국적의 인종이 함께 사는 다문화 사회라는 것이 이미 기정사실화되었음을 의미한다. 특히 한국에 거주하는 외국인의 증가에 따라 내국인과 외국인이 가정을 이루어 나가는 다문화 가구 역시 매년 증가하고 있다는 점에서도 다문화가정이 한국사회에 잘 적응하여 살아갈 수 있도록 그들에게 필요한 국가적 지원이 무엇인지를 살펴볼 필요가 있다.

이러한 다문화가정을 효과적으로 지원하기 위해서는 다문화가정의 특성을 파악할 필요가 있다. 다문화가정의 특성을 살펴보기 위해 2022년 인구주택총조사를 살펴보면 다문화가정은 399,396가구로 전년에 비해 14,177가구(3.7%) 증가하였을 뿐만 아니라, 매년 꾸준히 증가하고 있음을 알 수 있다. 특히 다문화 혼인을 한 외국인 및 귀화자

아내의 출신 국적 비중을 살펴보면 중국이 23.9%로 가장 많았으며, 그 다음으로 베트남(13.5%), 태국(11.4%)의 순서로 나타났다(통계청, 2022). 이는 한국과 이웃 나라인 아시아 국가 출신의 이주배경 여성이 많은 비중을 차지하고 있음을 알 수 있다. 동시에 다문화 혼인을 선택한 외국인 및 귀화자 남편의 출신 국적은 미국이 9.4%, 중국이 8.5%, 베트남이 3.2% 순으로 나타났다(통계청, 2021). 이는 다문화 남편과 다문화 아내의 출신 국적의 비중이 서로 다르다는 점도 확인할 수 있다. 이러한 사회적 변화에 맞춰 최근 지역별로 다문화 및 탈북 학생을 위한 대입 설명회가 개최되거나 다문화 학생에게 찾아가는 SW·AI 교육프로그램을 운영하는 등 다문화가정의 자녀, 즉 다문화가정의 청소년에 대한 지원이 다각도로 이루어지고 있다.

다문화가정에 대한 선행연구들을 살펴보면 다문화가정에 지원되는 전반적 정책에 대한 연구(권오경, 김남희, 김혜빈, 2022; 장안리, 박미화, 2022), 다문화가정의 자녀와 가족의 언어교육에 대한 정책과 프로그램을 분석한 연구(김선희, 성수나, 남청미, 2023; 서영선, 우인혜, 2022; 서월아, 2022), 다문화가정 자녀의 문화적응에 대한 정책에 대한 연구(박지나, 2023; 이수민, 문찬주, 2022; 이채연, 김태림, 유은, 2022), 다문화가정 자녀의 진로지원 정책에 관한 연구(유두호, 최정윤, 2022; 이수민, 문찬주, 2022) 등이 주로 이루어진 것으로 나타났다.

그러나 다양한 특성을 가진 다문화가정을 지원하기 위해서는 다문화가정의 유형별로 맞춤형 지원이 중요함에도 불구하고 최근까지 이루어진 다문화가정을 지원하기 위한 정책과 연구들 대부분이 다문화가정을 하나의 정형화된 유형으로 보는 경우가 대부분이었다(이강훈, 2016a, 2016b, 2019a, 2019b, 2020a, 2020b, 2021, 2022a, 이강훈, 황해익, 2016). 이에 본 연구에서는 다문화가정 지원 정책을 마련하기에 앞서 다문화가정에 대하여 사회적 담론이 어떻게 형성 및 구조화되어 있는지를 살펴볼 필요가 있었다. 이는 다문화가정을 지원하기 위한 다양한 고려사항들을 확인할 수 있었을 뿐만 아니라 효과적으로 다문화가정을 지원하기 위한 의미 있는 시사점들을 제공할 것으로 기대된다.

사회적 담론을 살펴보기 위한 방법이 다양하게 있겠으나, 최근 들어 다양한 분야에서 빅데이터를 활용한 사회적 담론 분석을 다수 수행하고 있으며(황해익, 김선희, 이강훈, 2018; Kim, Kim, Lee, & Lee, 2022), 사회과학 분야에서도 다양한 주제의 사회적 담론 분석 연구들이 빅데이터를 활용하여 활발히 진행되고 있다(김병만, 이강훈, 2020, 2022; 김병만, 이강훈, 윤정진, 2022; 이강훈, 2022b; 이강훈, 황해익, 2020a, 2020b;

Kim & Lee, 2020). 빅데이터는 전통적인 문서자료뿐만 아니라 ICT 기술에 기반한 디지털 환경에서 생성되는 다양한 종류의 방대한 자료를 의미하며(정혜영, 김병만, 2022), 이처럼 빅데이터를 활용하여 다문화가정에 대한 방대한 자료를 분석하는 것은 다문화가정을 지원하기 위한 정책 마련의 방향성을 탐색하는 데에 의미 있는 시사점 제공이 가능할 것이다.

빅데이터는 자료의 크기가 방대할 뿐만 아니라 규모(volume), 다양성(variety), 진실성(veracity), 가치(value)와 함께 복잡성(complexity)을 의미하는 5V+1C라는 특성을 지니고 있으며(서원석, 이강훈, 김석우, 2019; 이강훈, 2022b; 이강훈, 제희선, 2018), 정형·비정형의 데이터가 혼재되어 있는 집합체이기에 전통적인 분석방법으로는 데이터를 수집하고 정제하여 분석하는 것은 쉽지 않다(박리나, 황해익, 이강훈, 2019; 이강훈, 2019c; Kim & Lee, 2021). 하지만 최근 ICT 기술이 발달함에 따라 기존에는 분석하기 어려웠던 정형·비정형의 빅데이터에 대한 분석들이 더욱 활발히 진행되고 있다(남문희 외, 2022). 이러한 빅데이터를 분석하는 데에도 다양한 방법이 있겠으나, 키워드로 생성되는 비정형 데이터를 활용할 경우 키워드들 간의 관계를 살펴보고자 키워드 네트워크 분석이 많이 활용되고 있을 뿐만 아니라(정혜영, 김병만, 2022; 황해익, 이강훈, 제희선, 2018; Borgatti, Everett, & Johnson, 2018; Freeman, 2017; Kim, Youn, Jeong, & Lee, 2018; Scott, 2017), 특정 키워드를 중심으로 형성된 지식구조가 어떠한 양상을 보이는지를 살펴보는 데에도 효과적이기 때문에 주요 토픽에 대한 사회적 담론을 살펴볼 때에도 의미 있는 정보 제공이 가능하다.

또한 주제어를 분석하기 위한 대표적인 분석방법으로 토픽모델링 분석이 있는데, 토픽모델링 분석에는 크게 WSC(word-level semantic clustering)방법과 LDA(latent dirichlet allocation)방법이 있다(문영주, 2020a, 2020b). WSC 방식은 문서 내 포함된 키워드들을 바탕으로 토픽을 도출하는 방식이며, LDA 방식은 알고리즘을 활용한 머신러닝 분석을 실시함으로써 주요 토픽을 자동으로 도출하는 방식이다(김병만, 이강훈, 2023). 사회과학 분야에서는 토픽모델링 분석을 수행한 선행연구(문영주, 2020a, 2020b, 김병만, 이강훈, 2023; 배유나, 이강훈, 2022)를 살펴본 결과 대체로 알고리즘이 활용된 머신러닝 분석을 통해 토픽이 도출되는 LDA 방식을 채택하고 있었으며, 이에 본 연구에서도 LDA 토픽모델링 분석을 통해 다문화가정에 대한 사회적 인식을 탐색해보고자 한다.

따라서 본 연구에서 빅데이터를 활용한 키워드 네트워크 분석과 LDA 토픽모델링

분석 등을 통하여 다문화가정에 대한 사회적 인식을 살펴봄으로써 다문화가정을 지원하기 위한 방안 마련의 기초자료를 제공하고자 한다. 본 연구에서 설정한 목적 달성을 위한 연구문제는 다음과 같다.

- 연구문제 1. 다문화가정 관련 키워드에 대한 빈도분석 결과는 어떠한가?
 연구문제 2. 다문화가정에 대한 키워드 네트워크 분석 결과는 어떠한가?
 연구문제 3. 다문화가정에 대한 LDA 토픽모델링 분석 결과는 어떠한가?

II. 연구방법

1. 데이터 수집

본 연구에서 활용된 데이터는 TEXTOM(2023)에서 제공하고 있는 웹데이터로 국내·외 주요 포털사이트 중 3곳(NAVER, DAUM, GOOGLE)에서 ‘다문화가정’이라는 키워드를 중심으로 하여 rawDATA를 수집하였다. 다문화가정에 대한 동향을 살펴보기 위해 여성가족부(2019, 2022)에서 실시하는 전국다문화가족실태조사는 5년을 주기로 결과보고가 이루어지기 때문에 2019년 5월에 공개된 2018년 전국다문화가족실태조사를 끝으로 최근까지 공개된 실태조사 결과보고는 전무한 실정이다. 이에 데이터 수집기간은 전국다문화가족실태조사 결과보고가 이루어지지 않은 2019년부터 2023년까지 최근 5년간으로 하였다. 수집된 데이터는 웹 1,329건, 뉴스 2,056건, 블로그 584건, 카페 315건, 지식IN 991건, 전문정보 1,010건, 페이스북 273건, 전체 6,558건으로 대규모 데이터이다(이강훈, 2019c, 이수상, 2013).

〈표 1〉 수집된 데이터

Keyword	웹	뉴스	블로그	카페	지식IN	전문정보	페이스북	전체
다문화가정	1,329	2,056	584	315	991	1,010	273	6,558

2. 데이터 분석도구

본 연구에서는 데이터의 수집 및 정제를 위해 ㈜The IMC의 빅데이터 분석 솔루션

인 TEXTOM(2023)을 활용하였으며, 다문화가정 관련 키워드 간의 관계성 분석을 위해 키워드 네트워크 분석 Tool인 UCINET 6.645 for Windows(Freeman, 2017)와 네트워크 시각화에 적합한 Tool인 NetDraw 2.161을 함께 활용하여 분석하였다(Borgatti et al., 2018).

3. 데이터 처리

수집한 데이터는 명확한 해석과 더불어 이해도를 높이기 위하여 1차·2차 데이터 정제 등을 실시하였다(김병만, 이강훈, 2020; Kim & Lee, 2020, 2021). 우선 1차 정제 과정에서는 수집된 rawDATA 상에서 고유명사와 복합명사를 결과값에 그대로 반영하는 Espresso K 방식을 채택해 명사형 키워드만 추출하는 텍스트 마이닝(text mining)을 실시한 후(배유나, 이강훈, 2022), 한 음절로 된 키워드, 기호, 한자, 숫자 등 의미부여가 어려운 키워드들은 TEXTOM(2023)에서 제공하는 ‘바로 편집하기’를 활용하여 삭제 처리하였다(이강훈, 황해익, 2020a, 2020b; Kim et al, 2018). 이때 다양한 수집채널에서 공유하기 혹은 퍼가기 기능을 통하여 중복 또는 재생산된 데이터의 경우 동일 자료라고 할지라도 사회적인 관심이나 요구도가 높아 주요하게 다루어지고 있는 정보로 해석될 수 있기 때문에 전문정보의 이중 노출 등을 제외한 모든 데이터를 그대로 활용하였다(김병만, 이강훈, 2023; 박리나 외, 2019; 이강훈, 2022b). 1차 정제 예시는 아래 표 2와 같다.

2차 정제과정에서는 키워드를 치환, 병합, 그리고 삭제하는 과정을 실시하였다(이강훈, 2022b). 먼저, 유의어와 동의어는 대표적인 키워드로 치환하였으며, 오타자는 교정하였고 적절하지 않은 띄어쓰기의 경우에는 TEXTOM(2023)의 ‘바로 편집하기’를 활용하여 하나의 키워드로 병합하는 작업을 수행하였다(이강훈, 제희선, 2018, 정혜영, 김병만, 2022; Hwang & Lee, 2017). 이때 검색어로 활용한 ‘다문화가정’은 모든 rawDATA에서 공통으로 출현하는 키워드로 네트워크의 구조적 속성 및 중심성 분석 결과에 적절한 의미부여가 어려울 것으로 고려되어 제거하였다(김병만, 2019; 이강훈, 2019c). 2차 정제 예시를 살펴보면 아래 표 3과 같다.

〈표 2〉 1차 정제의 예시

자료유형	1차 정제 예시
웹페이지	다문화 가정 많다고 공립이나 병설유치원 꺼리는 학부모 많나요? => 다문화, 가정, 공립, 병설유치원, 학부모
블로그	다문화가정 학부모의 자녀 교육 역량 강화를 통해 다문화학생의 학교생활 적응을 지원하기 위해 오는 3월 10일(금)까지 ‘2023. 다문화가정 학부모 교육’을... => 다문화가정, 학부모, 자녀, 교육, 역량, 강화, 다문화학생, 학교생활, 적응, 지원
뉴스	이번 프로그램은 다문화 가족 자녀가 영·유아기부터 자연스럽게 이중언어로 소통할 수 있는 환경을 조성해 자녀의 정체성 확립과 글로벌 인재로 성장할 수 있도록 지원하는... => 프로그램, 다문화, 가족, 자녀, 영·유아기, 이중언어, 소통, 환경, 조성, 정체성, 확립, 글로벌, 인재, 성장, 지원
카페	다자녀보다 다문화 가정 혜택이 많나요? => 다자녀, 다문화, 가정, 혜택
지식IN	다문화가정 뒤통수대려진 혜택같은것있나요? 다누리 말고 다른 사이트있나요? 다문화가정 혜택중 쓸만한거없나요? 수세미만드는거 이런거만있네요 => 다문화가정, 혜택, 다누리, 사이트, 수세미
전문정보	다문화가정 이주여성의 여가 만족이 생활만족과 심리적 안녕감에 미치는 영향 => 다문화가정, 이주여성, 여가, 만족, 생활만족, 심리적안녕감, 영향

〈표 3〉 2차 정제의 예시

정제유형	키워드	2차 정제 예시
치환과 삭제	다문화가정	다문화가정 / 다문화 + 가정 / 다문화가족 / 다문화 + 가족
치환	한국	한국 / 대한민국
	아버지	아버지 / 아빠
	어머니	어머니 / 엄마
병합	프로그램	프로그램 / 프로 + 램
	이중언어	이중언어 / 이중언 + 어

전반적으로 본 연구에서 수행한 1차·2차 데이터 정제가 적절하게 진행되었는지 검토하기 위해 삼각측정법(triangulation)을 활용하였으며, 구체적으로 연구자 간 검토, 연구자료 간 검토, 외부 전문가와 연구자 간의 검토 과정 등을 수행하였다(김병만, 이강훈, 2023; 황해익, 이강훈, 제희선, 2018). 이때 외부 전문가의 경우 빅데이터를 통해 키워드 네트워크 분석과 LDA 토픽모델링 분석 연구를 수행한 경험이 있는 전문가 3인을 구성하여 내용타당도를 검증하였으며, 구체적으로 자료를 검토하는 과정에서 일치하지 않는 의견에 대해서는 내용타당도 전문가 3인의 반복적인 정제할 키워드에 대한 의견 조정 과정을 통해 키워드를 정제 및 재정제하는 과정을 거쳤다. 내용타당도 검증 결과 1차·2차 데이터 정제 과정이 적절한 것으로 사료되어 최종 정제된 데이터

를 바탕으로 키워드 네트워크 분석과 LDA 토픽모델링 분석을 실시하였다.

4. 자료 분석

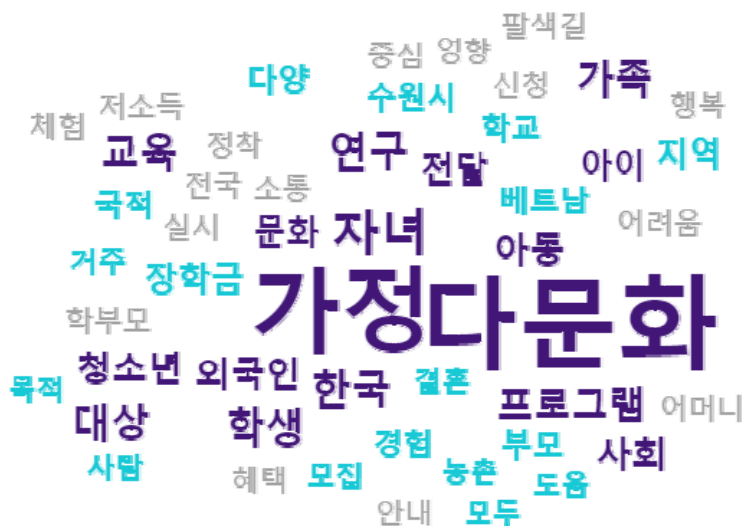
다문화가정에 대한 전반적인 사회적 인식이 어떠한지를 살펴보기 위해 TEXTOM(2023)에서 제공하는 rawDATA를 수집하였으며, 텍스트 마이닝을 통해 1차·2차 데이터 정제과정을 실시하였다(김병만, 이강훈, 2023; 이수상, 2013). 자료분석의 절차를 구체적으로 살펴보면 다음과 같다. 첫째, 1차·2차 데이터 정제를 거친 데이터를 활용하여 빈도분석을 실시하였고, 상위 200개의 빈도 키워드를 중심으로 200×200 1-mode 메트릭스를 작성하였다(이강훈, 제희선, 2018; Borgatti et al., 2008). 둘째, 다문화가정에 대한 사회적 인식을 분석하고자 키워드 네트워크 분석을 실시하였고, 다문화가정에 대한 거시적인 네트워크의 구조적 속성을 살펴보고자 노드(node), 평균연결거리(average distance), 밀도(density), 컴포넌트 수(components), 평균연결강도(average degree), 네트워크 집중도(network centralization), 지름(diameter)을 검토하였다(서원석 외, 2019; 이강훈, 2019c; 이강훈, 황해익, 2020a, 2020b). 또한 전체 네트워크 밀도의 통계적 유의성 검증을 위해 부트스트래핑 기법을 활용해 단일표본 t-test를 실시하였고(곽기영, 2017; 김병만, 이강훈, 2020, 2022; 서정아, 오익근, 2017), 미시적 네트워크의 구조적 속성들을 확인하고자 연결중심성, 매개중심성, 근접중심성, 위세중심성 등을 중심으로 중심성 분석을 진행하였다(김병만, 2019; 이강훈, 2019c; 이강훈, 제희선, 2018). 셋째, 다문화가정과 관련된 주요 토픽이 무엇인지를 살펴보기 위해 알고리즘이 적용된 머신러닝 분석을 통하여 토픽을 자동적으로 도출하는 LDA 기법의 토픽모델링 분석을 실시하였다(김병만, 이강훈, 2023; 문영주, 2020a, 2020b; 배유나, 이강훈, 2022). 마지막으로 네트워크 분석 결과에 대한 이해도를 높이고자 워드 클라우드 및 NetDraw를 활용해 시각화하였다(김병만, 이강훈, 2022; 김용학, 김영진, 2016; Kim et al., 2018).

Ⅲ. 연구결과

1. 다문화가정 관련 키워드에 대한 빈도분석

본 연구에서 빅데이터를 통해 다문화가정 관련 키워드들을 수집한 결과 전체

23,239개의 키워드가 추출되었으며, 1차·2차 데이터 정제를 거친 데이터들 중에서 상위 200개의 빈도 키워드를 최종적으로 분석대상으로 선정하였다. 선정된 키워드들 중 상위 50개 빈도 키워드를 중심으로 하여 워드 클라우드와 빈도분석 결과를 그림으로 제시하면 아래 그림 1, 표 4와 같다.



〈그림 1〉다문화가정 관련 키워드에 대한 워드 클라우드

〈표 4〉 다문화가정 관련 키워드에 대한 빈도분석(상위 50개의 키워드)

Rank	Keyword	N	Rank	Keyword	N	Rank	Keyword	N	Rank	Keyword	N	Rank	Keyword	N
1	다문화	4,520	11	사회	728	21	다양	419	31	결혼	279	41	안내	248
2	가정	4,013	12	청소년	705	22	국적	348	32	사람	275	42	겨주	246
3	자녀	1,496	13	외국인	700	22	학교	348	33	어려움	274	43	목적	245
4	한국	980	14	아동	674	24	베트남	339	33	행복	274	44	중심	243
5	학생	963	15	전달	628	25	수원시	319	35	모집	272	45	정착	240
6	연구	934	16	문화	552	26	경험	306	36	신청	271	46	농촌	236
7	교육	848	17	아이	514	27	어머니	281	37	실시	270	47	전국	230
8	대상	846	18	장학금	507	27	혜택	281	38	학부모	269	48	팔색길	225
9	가족	831	19	지역	500	27	영향	281	39	소통	265	48	모두	225
10	프로그램	743	20	부모	477	30	체험	280	40	저소득	264	50	도움	220

다문화가정 관련 키워드에 대한 빈도분석을 실시한 결과 다문화(4,520)가 가장 높게

나타났으며, 그 다음으로 가정(4,013), 자녀(1,496), 한국(980), 학생(963), 연구(934), 교육(848), 대상(846), 가족(831), 프로그램(743) 순으로 나타났다. 다문화가정 관련 키워드들 중 이해관계자 집단 관련 요인으로는 가정(4,013), 한국(980), 사회(728), 학교(348), 베트남(339), 수원시(319), 농촌(236), 전국(230), 대학(209), 중국(188) 등의 키워드들이 나타났으며, 이해관계자 관련 요인으로는 자녀(1,496), 학생(962), 가족(831), 청소년(705), 외국인(700), 아동(674), 아이(514), 부모(477), 어머니(281), 학부모(269) 등의 키워드들이 나타났다. 정책 및 행정 관련 요인으로는 다문화(4,520), 교육(848), 프로그램(743), 장학금(507), 지역(500), 국적(348), 체험(280), 행복(274), 모집(272), 신청(271) 등의 키워드들이 나타났으며, 사회적 반응 관련 요인으로는 전달(628), 문화(552), 다양(419), 경험(306), 혜택(281), 결혼(279), 사람(275), 어려움(274), 의미(207), 희망(205) 등의 키워드들이 나타났다. 마지막으로 학문적 관심 관련 요인으로는 연구(934), 대상(846), 영향(281), 국내(192), 증가(183), 향상(182), 관계(158), 분석(141), 사례(140), 연계(122) 등의 키워드들이 나타났다.

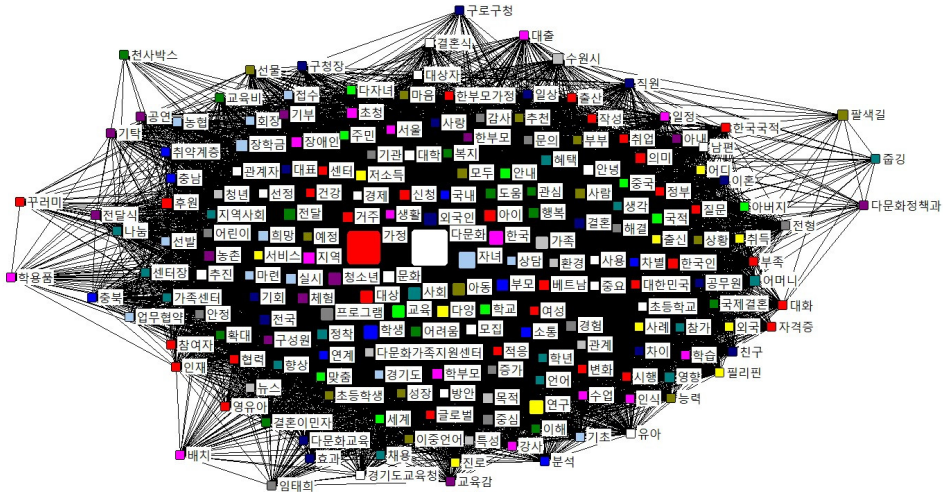
2. 다문화가정에 대한 키워드 네트워크 분석

1) 다문화가정에 대한 거시적 수준의 전체 네트워크 분석

다문화가정에 대한 전체적인 네트워크의 구조적 속성을 살펴보면 노드는 200, 밀도는 .588, 평균연결강도는 117.080, 평균연결거리는 1.412, 컴포넌트 수는 1, 지름은 2, 네트워크 집중도는 2.440%로 나타났다. 선정된 키워드들 중 상위 200개 빈도 키워드를 중심으로 시각화하면 아래 그림 2와 같다.

2) 다문화가정에 대한 전체 네트워크의 통계적 유의성 검정

다문화가정에 대한 네트워크 전체의 통계적 유의성 검정을 위하여 모집단의 네트워크 밀도를 0으로 가정한 표본에서 추정되어진 밀도가 우연에 의한 값인지를 검증하는 부트스트래핑(bootstrapping) 기법을 활용하였다. 다문화가정 네트워크 밀도에 대한 통계적 유의성을 검증한 결과 네트워크 데이터의 표집분포 평균(M)은 11.2836이었으며, 표준오차(SE)는 1.8589, $Z=6.0158$ 로 나타났다. 이는 다문화가정에 대한 네트워크 통계량이 $|Z\text{-score}|$ 보다 클 확률이 0.0002이므로, 95% 신뢰구간에서 영가설을 기각하였으



〈그림 2〉다문화가정에 대한 전체 네트워크의 구조

로 본 연구에서 활용한 다문화가정에 대한 전체 네트워크를 활용하여 분석결과를 해석하는 것은 적절하겠다(곽기영, 2017; 김병만, 이강훈, 2023; 배유나, 이강훈, 2022; 서정아, 오익근, 2017).

3) 다문화가정에 대한 미시적 수준의 전체 네트워크 분석

다문화가정에 대한 미시적 수준에서의 네트워크 속성을 확인하고자 중심성 분석을 실시하였고, 상위 10개 핵심 키워드를 중심으로 중심성 분석 결과를 제시하면 아래 표 5와 같다.

〈표 5〉 다문화가정에 대한 중심성 분석 (상위 10개 핵심 키워드)

Rank	NrmDgree	Closeness	nBetweenness	Eigenvector
1	다문화 .2572	자녀 .995	자녀 1.100	다문화 .617
2	가정 .2345	가정 .992	가족 .957	가정 .604
3	자녀 .881	다문화 .990	가정 .870	자녀 .175
4	한국 .602	가족 .980	다문화 .828	한국 .147
5	학생 .567	한국 .975	한국 .644	교육 .109
6	가족 .516	대상 .975	대상 .639	학생 .106
7	교육 .468	문화 .970	문화 .615	가족 .105
8	대상 .433	교육 .967	행복 .609	사회 .105
9	사회 .391	사회 .965	지역 .599	외국인 .103
10	외국인 .384	다양 .965	교육 .582	문화 .100

연결중심성 분석의 결과를 살펴보면 키워드들 간 가장 많은 관계를 맺고 있는 키워드는 다문화(2.572)로 나타났으며, 그 다음으로 가정(2.345), 자녀(.881), 한국(.602), 학생(.567), 가족(.516), 교육(.468), 대상(.433), 사회(.391), 외국인(.384) 순으로 나타났다.

근접중심성 분석의 결과를 살펴보면 다른 키워드들과 밀접한 관계를 형성한 키워드는 자녀(.995)로 나타났으며, 그 다음으로 가정(.992), 다문화(.990), 가족(.980), 한국(.975), 대상(.975), 문화(.970), 교육(.967), 사회(.965), 다양(.965) 순으로 나타났다.

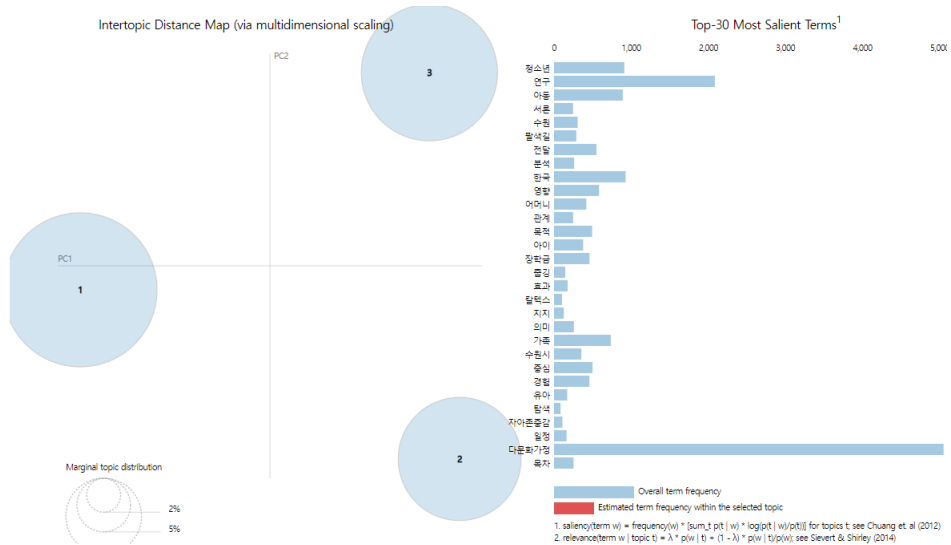
매개중심성 분석의 결과를 살펴보면 키워드들 간의 관계에서 중요 매개자 역할을 하는 키워드는 자녀(1.100)로 나타났으며, 그 다음으로 가족(.957), 가정(.870), 다문화(.828), 한국(.644), 대상(.639), 문화(.609), 지역(.599), 교육(.582) 순으로 나타났다.

위세중심성 분석의 결과를 살펴보면 중심성이 높은 키워드들과 긴밀하게 연결된 키워드는 다문화(.617)로 나타났으며, 그 다음으로 가정(.604), 자녀(.175), 한국(.147), 교육(.109), 학생(.106), 가족(.105), 사회(.105), 외국인(.103), 문화(.100) 순으로 나타났다.

이러한 결과는 버즈량을 기준으로 가중치는 부여하는 빈도분석 뿐만 아니라 연결강도, 매개강도, 근접강도, 위세강도 등을 기준으로 가중치를 부여한 중심성 분석에서도 다문화가정을 지원하기 위한 프로그램들을 개발하고 운영해 나가는 것이 중요함을 시사한다. 또한 이러한 다문화가정 지원 프로그램은 다문화가정 출신 자녀뿐만 아니라 가족 전반을 대상으로 하는 프로그램이 필요한 것으로 나타났으며, 특히 다문화가정의 행복한 정착을 위한 다문화교육 뿐만 아니라 장학금과 같은 외국인을 위한 지역 맞춤형으로 다양한 다문화가정 지원 혜택도 함께 제공될 필요가 있는 것으로 나타났다.

3. 다문화가정에 대한 LDA 토픽모델링 분석

다문화가정에 대한 LDA 토픽모델링 분석 결과 다문화가정 관련 토픽에 포함된 키워드들 중에 가장 두드러진 30개 키워드는 청소년, 연구, 아동, 서론, 수원, 팔색길, 전달, 분석, 한국, 영향, 어머니, 관계, 목적, 아이, 장학금, 줄길, 효과, 칼텍스, 지지, 의미, 가족, 수원시, 중심, 경험, 유아, 탐색, 자아존중감, 일정, 다문화가정, 목차 등으로 나타났다. 토픽별 도출되어진 키워드에 기반해 의미를 정리한 후 각 토픽의 주제를 명명하였으며, 다문화교육에 대한 LDA 토픽모델링 분석 결과는 아래 그림 3과 같다.



〈그림 3〉다문화가정에 대한 LDA 토픽모델링

다문화교육에 대한 토픽별 주제 및 키워드를 살펴보면 아래 표 6과 같다.

〈표 6〉 다문화가정에 대한 토픽별 키워드들

No	Topic	%	Keyword
1	지역 맞춤형 다양한 다문화가정 지원 혜택에 대한 요구	20.4	다문화, 가정, 자녀, 연구, 가족, 학생, 장학금, 대상, 경험, 수원, 프로그램, 사회, 팔색길, 수원시. 전달, 부모, 의미, 지역
2	사회 맞춤형 다문화가정 지원 프로그램에 대한 요구	14.6	다문화, 가정, 전달, 교육, 아동, 가족, 아이, 대상, 중심, 학생, 자녀, 청소년, 연구, 프로그램, 사회, 효과
3	가족 중심 다문화가정 지원 방안 마련을 위한 연구에 대한 요구	22.3	다문화, 연구, 가정, 청소년, 한국, 아동, 자녀, 영향, 사회, 외국인, 어머니, 목적, 교육, 프로그램, 분석, 중심, 부모, 관계

토픽1은 다문화, 가정, 자녀, 연구, 가족, 학생, 장학금, 대상, 경험, 수원, 프로그램, 사회, 팔색길, 수원시. 전달, 부모, 의미, 지역을 포함하는 것으로 나타났다. 이러한 키워드들은 장학금과 같은 다문화가정에 대한 재정적 지원뿐만 아니라 지역 맞춤형으로 다양한 다문화가정 지원 혜택에 대한 요구 담론과 관련된 내용을 다루는 것으로 나타났으므로 토픽1은 ‘지역 맞춤형 다문화가정 지원 혜택에 대한 요구’(20.4%)로 명명하였다.

토픽2는 다문화, 가정, 전달, 교육, 아동, 가족, 아이, 대상, 중심, 학생, 자녀, 청소년, 연구, 프로그램, 사회, 효과를 포함하는 것으로 나타났다. 이러한 키워드들은 획일적인

다문화가정 지원 프로그램의 운영이 아니라 사회 맞춤형 다문화가정 지원 프로그램을 개발 및 운영에 대한 요구 담론의 내용을 다루고 있으므로 토픽2는 ‘사회 맞춤형 다문화가정 지원 프로그램에 대한 요구’(14.6%)로 명명하였다.

토픽3은 다문화, 연구, 가정, 청소년, 아동, 자녀, 영향, 사회, 외국인, 어머니, 목적, 교육, 프로그램, 분석, 중심, 부모, 관계를 포함하는 것으로 나타났다. 이러한 키워드들은 다문화가정의 부모나 자녀 등 특정 개인에 초점을 맞추기 보다는 가족 중심의 다문화가정 지원 방안 마련을 위한 연구에 대한 요구 담론의 내용을 다루고 있으므로 토픽3은 ‘가족 중심 다문화가정 지원 방안 마련을 위한 연구에 대한 요구’(22.3%)로 명명하였다.

IV. 논의 및 결론

본 연구에서는 빅데이터를 활용하여 다문화가정에 대한 사회적 인식이 어떠한지를 살펴봄으로써 다문화가정에 대한 효과적 지원을 위한 기초자료를 제공하고자 하였다. 이를 위해 TEXTOM에서 제공하는 빅데이터에 기반하여 키워드 ‘다문화가정’을 중심으로 rawDATA를 수집하였으며, 1차·2차 데이터 정제과정을 완료한 데이터에서 빈도 상위 200개의 키워드를 핵심키워드로 선정하여 키워드 네트워크 분석을 실시하였다. 또한 공출현 빈도와 버즈량이 높은 키워드를 중심으로 LDA 토픽모델링 분석을 실시하였다. 본 연구결과를 바탕으로 연구문제에 대한 논의는 다음과 같다.

첫째, 다문화가정 관련 키워드에 대한 빈도분석을 실시한 결과 다문화(4,520)가 가장 높게 나타났으며, 그 다음으로 가정(4,013), 자녀(1,496), 한국(980), 학생(963), 연구(934), 교육(848), 대상(846), 가족(831), 프로그램(743) 순으로 나타났다. 이러한 결과는 다문화가정과 관련하여 단순히 다문화가정 지원 프로그램의 운영뿐만 아니라 다양한 다문화가정 지원 정책이 필요하다는 사회적 인식이 드러난 결과로 볼 수 있다. 특히 다문화가정 지원 프로그램은 다문화가정의 행복한 정착을 위한 체험 중심의 다문화교육 등 다문화가정 출신 자녀뿐만 아니라 가족 전반을 대상으로 하는 프로그램의 개발 및 운영이 필요하다. 그리고 장학금과 같은 다양한 다문화가정 지원 혜택도 함께 제공될 필요가 있음을 시사한다.

둘째, 다문화가정과 관련된 키워드들을 중심으로 빈도분석 및 키워드 네트워크 분석을 실시한 결과 다문화가정 관련 요인으로는 이해관계자 집단, 이해관계자, 정책 및 행정, 사회적 반응, 학문적인 관심 등에 관한 내용들을 포함하는 것으로 나타났다. 이를 좀 더 구체적으로 살펴보면 이해관계자 집단 관련 요인으로는 가정, 한국, 사회, 학교, 베트남, 수원시, 농촌, 전국, 대학, 중국 등의 키워드들이 나타났으며, 이해관계자 관련 요인으로는 자녀, 학생, 가족, 청소년, 외국인, 아동, 아이, 부모, 어머니, 학부모 등의 키워드들이 나타났다. 한편, 정책 및 행정 관련 요인으로는 다문화, 교육, 프로그램, 장학금 지역, 국적, 체험, 행복, 모집, 신청 등의 키워드들이 도출되었으며, 사회적 반응과 관련 요인으로는 전달, 문화, 다양, 경험, 혜택, 결혼, 사람, 어려움, 의미, 희망 등의 키워드들이 나타났다. 마지막으로 학문적 관심에 대한 요인으로는 연구, 대상, 영향, 국내, 증가, 향상, 관계, 분석, 사례, 연계 등의 키워드들이 나타났다. 이러한 결과는 다문화가정을 효과적으로 지원하기 위해서는 다문화가정과 관련된 다양한 이해관계자 집단과 이해관계자들의 요구도를 다각적으로 경청할 필요가 있다고 본 선행연구들(박새빈, 최화림, 박연진, 2023; 이은혜, 장인실, 2021; 이현아, 2021)의 주장과 같은 맥락으로 다문화가정 지원을 위한 정책을 마련하고 운영 할 시 정책의 혜택을 보는 사용자 중심의 정책 마련이 이루어지길 기대한다. 특히 다문화가정은 결혼 이후 서로 다른 문화로 인해 다양한 어려움을 겪는다는 점에서 서로 다른 문화를 이해하고 행복한 결혼 생활을 유지할 수 있도록 희망을 줄 수 있는 중장기적인 정책마련이 이루어져야 할 것이다.

셋째, 다문화가정에 대한 키워드 네트워크 분석을 거시적 측면에서 고찰한 결과, 다문화가정에 대한 전체 네트워크의 구조적 속성의 경우 노드는 200, 밀도는 .588, 평균 연결강도는 117.080, 평균연결거리는 1.412, 컴포넌트 수는 1, 지름은 2, 네트워크 집중도는 2.440%로 나타났다. 부트스트래핑 기법을 활용하여 다문화가정에 대한 네트워크 밀도의 통계적 유의성을 검증한 결과 네트워크 데이터의 표집분포 평균(M)은 11.2836이었으며, 표준오차(SE)는 1.8589, $Z=6.0158$ 로 나타났다. 이와 같은 결과는 분석대상으로 선정된 다문화가정 관련 키워드들이 특정 키워드에 편중되기보다는 다양한 키워드들이 상호 긴밀하게 연결되어 있음을 보여준다(곽기영, 2017; 서정아, 오익근, 2017). 특히 다문화가정 관련 키워드들 간 관계성을 검토한 결과 네트워크의 집중도는 낮고, 밀도는 높은 것으로 나타났다. 한편, 미시적 측면에서 다문화가정에 대한 전체 네트워크의 중심성을 검토한 결과, 다문화, 가정, 자녀, 한국, 학생, 가족, 교육, 대상, 사회,

외국인, 문화, 다양, 행복, 지역 등의 키워드들은 중심성이 높은 것으로 나타났다. 이러한 결과는 다문화가정 관련 키워드들이 상호 밀접하게 연결되어 있음을 나타냄과 동시에 키워드들 간의 관계성이 원활하게 이루어지고 있음을 알 수 있다. 즉, 다문화가정 관련 키워드들은 각각의 키워드들이 서로 분절되어 있다기보다는 매우 다양하게 논의되고 있으며, 다양하게 정의 및 언급되고 있음을 보여준다(김병만, 이강훈, 2023). 따라서 다문화가정 지원 정책을 마련하기 위한 방향성을 설정하거나 다문화가정을 대상으로 하는 프로그램을 개발과 운영하는 데 있어서 본 연구에서 나타난 키워드들을 주요하게 다룰 필요가 있을 것으로 사료된다.

넷째, 다문화가정과 관련된 키워드를 중심으로 LDA 토픽모델링을 실시한 결과, 다문화가정 관련 토픽으로는 지역 맞춤형 다문화가정 지원 혜택에 대한 요구, 사회 맞춤형 다문화가정 지원 프로그램에 대한 요구, 가족 중심 다문화가정 지원 방안 마련을 위한 연구의 필요에 대한 담론으로 나타났다. 이를 구체적으로 살펴보면 토픽1은 장학금과 같은 다문화가정에 대한 재정적 지원뿐만 아니라 지역 맞춤형으로 다양한 다문화가정 지원 혜택에 대한 요구 담론과 밀접하게 관련이 있는 키워드들이 주요 키워드들로 나타났다. 특히 다문화, 가정, 자녀, 연구, 가족, 학생, 장학금, 대상, 경험, 수원, 프로그램, 사회, 팔색길, 수원시, 전달, 부모, 의미, 지역 등의 역 맞춤형 다문화가정 지원 혜택에 대한 요구와 관련된 키워드들이 눈에 띄었다. 이러한 결과는 다문화가정 관련 선행연구들(원숙경, 2017; 유두호, 최정운, 2022; 장인실, 우라미, 2022; 최청락, 박금식, 김혜정, 김성순, 2021)에서 언급한 바와 같이 지역 특성을 고려하여 지원 정책을 마련해야 한다는 주장을 지지하는 것으로 지역 맞춤형 다문화가정 지원 혜택에 대한 요구 담론이 다문화가정 지원 정책 개발 및 운영에도 영향을 미칠 것으로 사료된다.

또한 토픽2는 획일적인 다문화가정 지원 프로그램의 운영이 아니라 사회 맞춤형 다문화가정 지원 프로그램을 개발 및 운영에 대한 요구 담론과 밀접 관련이 있는 키워드들이 주요 키워드로 나타났다. 특히 다문화, 가정, 전달, 교육, 아동, 가족, 아이, 대상, 중심, 학생, 자녀, 청소년, 연구, 프로그램, 사회, 효과 등의 사회 맞춤형 다문화가정 지원 프로그램에 대한 요구와 관련된 키워드들이 눈에 띄었다. 이러한 결과는 효과적인 다문화가정 지원 정책 마련을 위해서는 다문화가정의 특성과 요구도를 바르게 파악하고 이를 다문화가정 지원 정책에 적극 반영하여 수요자 맞춤형으로 다문화가정 지원 프로그램의 개발이 필요하다고 본 선행연구(이병화, 김민영, 2022; 전경숙, 2016)

의 주장과 맥을 같이 하는 것으로, 단순히 보여주기식 다문화가정 지원 프로그램의 개발 및 운영을 넘어서 각 사회적 특성을 충분히 고려한 사회 맞춤형 다문화가정 지원 프로그램을 개발하고 운영할 필요가 있다는 것을 시사한다.

토픽3은 다문화가정의 부모나 자녀 등 특정 개인에 초점을 맞추기 보다는 가족 중심의 다문화가정 지원 방안 마련을 위한 연구에 대한 요구 담론과 밀접 관련이 있는 키워드들이 주요 키워드들로 나타났다. 특히 다문화, 연구, 가정, 청소년, 아동, 자녀, 영향, 사회, 외국인, 어머니, 목적, 교육, 프로그램, 분석, 중심, 부모, 관계 등의 가족 중심 다문화가정 지원 방안 마련을 위한 연구에 대한 요구와 관련된 키워드들이 눈에 띄었다. 이러한 결과는 외국에서 결혼을 목적으로 이주한 결혼이주여성과 그의 자녀인 한 개인에 초점을 두기 보다는 양가부모와 친인척 등 다양한 가족 구성원들 모두가 행복을 꿈꿀 수 있는 가족 중심의 다문화가정 지원 정책 마련을 위한 다양한 연구들이 필요하다고 본 선행연구(서월아, 2022; 최미희, 2022)의 주장을 지지하는 것으로, 유치원과 어린이집 그리고 초·중·고등학교에 재학 중인 다문화가정 출신 자녀뿐만 아니라 가족 전반을 대상으로 하는 가족 중심 다문화가정 지원 방안에 대한 다각적 연구들이 이루어질 필요가 있음을 시사한다.

이러한 논의들을 바탕으로 후속연구를 위한 제언을 다음과 같이 하고자 한다.

첫째, 본 연구에서 웹데이터를 활용해 사회적 인식을 분석하고 있어 효과적인 다문화가정 지원을 위한 기초자료를 제공하였다는 측면에서 의의가 있지만, 다문화가정에 지원되고 있는 정책에 대한 연구동향을 살펴보기 어렵다는 점에서는 한계가 있다. 따라서 후속연구에서는 토픽모델링분석 등을 통해 다문화가정 지원정책에 대한 연구 동향 분석들이 이루어지길 기대한다.

둘째, 본 연구의 결과에서 도출된 바와 같이 다문화가구 인구 100만시대를 맞이하여 행복한 가정생활을 유지할 수 있도록 단기보다는 장기적인 다문화가정 지원정책마련이 필요한 것으로 나타났다. 따라서 후속 연구에서는 다문화가정 지원을 위한 수요자 맞춤형 증장기적 정책 마련을 위한 연구가 다각적으로 이루어지길 기대한다.

셋째, 본 연구에서는 빅데이터라는 2차 자료를 통해 분석함으로써 다문화가정 지원 정책에 대한 방향성은 탐색할 수 있었던 반면, 실천적·정책적 접근 차원에서의 구체적인 방안이 나오지 못하였다는 한계가 있다. 따라서 후속연구에는 현재 우리나라에 거주하는 다문화가정을 대상으로 한 설문조사를 통해 다문화가정에 실질적으로 도움

을 줄 수 있는 다문화가정 지원 정책이 무엇인지에 대한 조사가 이루어질 기대한다.

참고문헌

- 곽기영 (2017). **소셜 네트워크분석**. 서울: 도서출판 청량.
- 권오경, 김남희, 김혜빈 (2022). 다문화가족지원정책 기본계획의 통시적 고찰. **다문화사회와 교육연구**, 10, 31-64.
- 김병만 (2019). 빅데이터를 통해 바라본 유보통합의 사회적 인식 네트워크 분석. **교육혁신연구**, 29(2), 17-39.
- 김병만, 이강훈 (2020). 복잡계에 근거한 공동선 증진 유아교육재정 정책 방향 탐색 사회적 담론 및 경쟁 담론 분석을 중심으로. **학습자중심교과교육연구**, 20(10), 407-438.
- 김병만, 이강훈 (2022). 키워드 네트워크 분석을 통해 살펴본 유아안전교육에 대한 역대 정부 별 연구동향 분석: KCI 학술지를 중심으로. **학습자중심교과교육연구**, 22(5), 583-609.
- 김병만, 이강훈 (2023). 빅데이터를 통해 살펴본 유아안전교육에 대한 사회적 인식 탐색: LDA 토픽모델링 분석을 중심으로. **유아교육연구**, 43(1), 5-29.
- 김병만, 이강훈, 윤정진 (2022). 키워드 네트워크 분석을 통해 살펴본 『교육과 문화』 연구동향 분석. **교육과 문화**, 4(2), 1-38.
- 김선희, 성수나, 남정미 (2023). 다문화가정 자녀의 이중언어 활용 지원 프로그램에 관한 연구. **마인드 교육**, 2(1), 1-27.
- 김용학, 김영진 (2016). **사회 연결망 분석**(4판). 서울: 박영사.
- 남문희, 김혜옥, 김요나, 김현정, 남효연, 박효진, 이강훈 (2022). **간호정보와 빅데이터**. 서울: 고문사.
- 문영주 (2020a). 토픽모델링과 언어 네트워크 분석을 활용한 교장공모제 관련 연구동향 분석. **교육문화연구**, 26(1), 217-242.
- 문영주 (2020b). 토픽모델링을 이용한 하브루타 연구동향 분석. **학습자중심교과교육연구**, 20(4), 1149-1175.
- 박리나, 황해익, 이강훈 (2019). 유아기 인성교육에 대한 키워드 네트워크 분석: 빅데이터를 활용하여. **열린유아교육연구**, 24(4), 171-197.
- 박새빈, 최화림, 박연진(2023). 어린이급식관리지원센터, 농촌지역 다문화 어린이들의 올바른 식습관 형성을 위한 기초 연구. 2023 한국지역사회생활고학회 춘계학술대회 자료

- 집. **지역사회의 인구변화에 따른 지역 생활 연구의 과제**(2023.05.12.). 132.
- 박지나 (2023). 다문화가정 자녀의 문화적응스트레스와 이중문화수용태도가 국가정체성에 미치는 구조적 관계: 다문화가족지원정책 경험 여부를 중심으로. **다문화교육연구**, 16(1), 165-190.
- 배유나, 이강훈 (2022). 빅데이터를 통해 살펴본 앙트러프러너십 교육에 대한 사회적 인식 탐색: LDA 토픽모델링 분석을 중심으로. **한국산학기술학회논문지**, 23(2), 377-388.
- 서영선, 우인혜 (2022). 중도입국 · 외국인가정 청소년 한국어교육에 관한 학부모의 경험 분석. **한국언어문화학**, 19(3), 73-106.
- 서원석, 이강훈, 김석우 (2019). 유아교육재정에 대한 키워드 네트워크 분석: 빅데이터를 중심으로. **학습자중심교과교육연구**, 19(12), 97-124.
- 서월아 (2022). 한국 다문화 가족을 위한 한국어 교육에 관한 연구. **인문사회** 21, 13(5), 3721-3732.
- 서정아, 오익근 (2017). 사회연결망 분석을 활용한 대구의 관광지 이미지 분석: 온라인 빅데이터를 중심으로. **한국콘텐츠학회논문지**, 17(7), 443-454.
- 여성가족부 (2019). **2018년 전국다문화가족실태조사 결과 연구보고서**. 서울: 여성가족부.
- 여성가족부 (2022). **2021년 전국다문화가족실태조사 결과보고서**. 서울: 여성가족부.
- 원숙경 (2017). 다문화가정 소통권 확보를 위한 미디어교육정책에 관한 연구: 부산지역 다문화미디어교육종사자 심층인터뷰를 중심으로. **지역과 커뮤니케이션**, 21(4), 37-63.
- 유두호, 최정운 (2022). 진로준비와 진로지원정책이 다문화가정 자녀의 진로결정에 미치는 영향 연구: 지역규모별 분석을 중심으로. **GRI 연구논총**, 24(1), 211-234.
- 이강훈 (2016a). 교사 다문화 태도와 유아 다문화 인식의 관계에서 교사 다문화 교수효능감의 매개효과. **보육학회지**, 16(1), 21-51.
- 이강훈 (2016b). 유아의 다문화 인식 증진을 위한 유아교사의 다문화 관련 변인들 간의 구조적 관계분석. **열린유아교육연구**, 21(3), 223-251.
- 이강훈 (2019a). 다문화가정 부모의 부모효능감과 사회적지지, 아동의 삶의 만족도 간의 관계에 대한 종단연구. **아동교육**, 28(4), 247-273.
- 이강훈 (2019b). 유아교사의 다문화 태도, 다문화교수역량, 다문화 교수효능감 간의 구조관계 분석. **열린유아교육연구**, 24(5), 269-292.
- 이강훈 (2019c). 빅데이터를 통해 살펴본 유아기 놀이성 증진 방향 탐색. **교육혁신연구**, 29(3), 257-289.
- 이강훈 (2020a). 유아의 다문화 인식에 영향을 미치는 유아교사의 다문화 태도, 다문화 교수

- 효능감, 어머니의 다문화 인식 간의 구조관계 분석. **교육과 문화**, 2(1), 13-38.
- 이강훈 (2020b). 유아교사의 다문화 연수경험 및 해외 체류경험이 다문화 교수 역량에 미치는 영향. **교육과 문화**, 2(2), 1-29.
- 이강훈 (2021). 어머니와 교사의 다문화 관련 변인들이 유아의 다문화 인식에 미치는 영향: 위계적 선형모형(HLM)분석을 중심으로. **교육과 문화**, 3(2), 1-32.
- 이강훈 (2022a). 교사의 다문화 태도 및 교수효능감과 어머니의 다문화 인식 및 문화성향이 유아의 다문화 인식에 미치는 영향. **교육과 문화**, 4(1), 5-30.
- 이강훈 (2022b). 빅데이터를 활용한 유아교사의 행복한 삶에 대한 인식 탐색: 키워드 네트워크 분석을 중심으로. **구성주의유아교육연구**, 9(1), 103-136.
- 이강훈, 제희선 (2018). 빅데이터를 통해 살펴본 유아창의인성교육에 대한 키워드 네트워크 분석. **미래유아교육학회지**, 25(4), 121-144.
- 이강훈, 황혜익 (2016). 유아교사용 다문화교수역량 척도 개발 연구. **아동교육**, 25(4), 267-292.
- 이강훈, 황혜익(2020a). 빅데이터를 통해 살펴본 누리과정 평가 담론 분석. **미래유아교육학회지**, 27(4), 77-109.
- 이강훈, 황혜익 (2020b). 빅데이터에 기초한 ‘2019 개정 누리과정 평가’ 키워드 네트워크 분석. **열린유아교육연구**, 25(4), 79-109.
- 이병화, 김민영 (2022). 다문화 장애아동 가족의 생활실태와 지원 과제. 2022년도 한국가족사회복지학회 춘계 학술대회 자료집. 인구감소시대, 사회적 돌봄의 현재와 미래를 논하다(2022.05.27.). 88-111.
- 이수민, 문찬주 (2022). 다문화 학부모의 문화적응 유형이 중학생의 학교생활 및 진로결정성에 미치는 영향. **교육문화연구**, 28(4), 415-438.
- 이수상 (2013). **네트워크 분석 방법론**. 서울: 논형.
- 이은혜, 장인실 (2021). 코로나시대 다문화가정 밀집지역 초등학교의 원격수업 운영 연구. **다문화교육연구**, 14(1), 71-93.
- 이채연, 김태림, 유은 (2022). 다문화 학생의 학업지속과 문화적응을 위한 맞춤 멘토링 서비스 제안. **Journal of Integrated Design Research**, 21(3) 135-152.
- 이현아 (2021). 지역사회 아버지대상 프로그램의 참여 효과 분석: 건강가정·다문화가족지원센터 참여 아버지에 대한 FGI조사. 2021 가정학분야 춘계공동학술대회 자료집. 제4차 건강가정기본계획에 대한 발전적 논의(2021.05.29.). 238.
- 장안리, 박미화 (2022). 고려인 가족단위 이주자들 중심의 지원정책 연구: 함박마을을 중심으로. **다문화사회연구**, 15(2), 63-108.

- 장인실, 우라미 (2022). 다문화학생 밀집지역의 다문화교육 정책에 대한 비판적 고찰: S교육 지원청을 중심으로. **다문화교육연구**, 15(3), 95-112.
- 전경숙 (2016). 중도입국청소년의 한국사회 적응실태와 정책적 지원방안 고찰. **다문화아동 청소년연구**, 1(2), 3-25.
- 정혜영, 김병만 (2022). 빅데이터로 바라본 저출산의 사회적 인식 네트워크 분석: 유아교육 · 보육정책을 중심으로. **한국영유아보육학**, 132, 77-103.
- 최미희 (2022). **다문화가족의 재구조화에 관한 질적 사례연구**. 대구가톨릭대학교 대학원 다문화학과 박사학위논문.
- 최청락, 박금식, 김혜정, 김성순 (2021). **연구보고서 2021-05, 부산지역 다문화가족 자녀 적응실태 및 지원방안**. 부산: 부산여성가족개발원.
- 통계청 (2021). **2021년 다문화 인구동태 통계 결과**. 대전: 통계청.
- 통계청 (2022). **2022년 인구주택총조사**. 대전: 통계청.
- 황해익, 김선형, 이강훈 (2018). 네트워크 분석을 통한 어머니의 행복 개념 연구. **미래유아교육학회지**, 25(2), 45-66.
- 황해익, 이강훈, 제희선 (2018). 빅데이터를 통해 살펴본 유아의 자아존중감에 대한 네트워크 분석. **유아교육연구**, 38(4), 213-235.
- Borgatti, S. P., Everett, M. G. & Johnson, J. C. (2018). *Analyzing social networks*. Thousand Oaks, CA: Sage Publications.
- Freeman. L. C. (2017). *Research methods in social network analysis*. London, England: Routledge.
- Hwang, H. I. & Lee, K. H. (2017). A network analysis of early childhood teachers' happiness based on big data. *International Journal of Korea Open Association for Early Childhood Education*, 3, 73-90.
- Kim, B. M., Kim, J. I., Lee, Y. W. & Lee, K. H. (2022). A study on exploring direction for future education for the common good based on big data. *Journal of Convergence for Information Technology*, 12(2), 37-46.
- Kim, B. M. & Lee, K. H. (2020). Keyword network analysis on the integrated research trends of early childhood education and childcare. *International Journal of Innovation, Creativity and Change*, 13(3), 595-607.
- Kim, B. M. & Lee, K. H. (2021). A study on exploring directions for early childhood safety education based on big data: Focus on semantic network analysis. *Review*

of International Geographical Education, 11(2), 636-647.

Kim, B. M., Youn, J. J., Jeong, S. J. & Lee, K. H. (2018). Keyword network analysis for creativity and personality of early childhood teachers: Focus on big data. *Indian Journal of Public Health Research & Development*, 9(11), 850-856.

Scott, J. (2017). *Social network analysis*. Thousand Oaks, CA: Sage Publications.

Textom (2023). 빅데이터 분석 솔루션 텍스툼(Textom). <http://www.textom.co.kr/home/main/main.php>에서 2023년 6월 18일 인출.

An Exploratory Study on Social Perception of Multicultural Families based on Big Data: Focus on LDA Topic Modeling Analysis

KangHoon Lee(Director, Education Big Data Research),

SuJeong Jeong(Assistant Professor, Silla University)

Abstract The purpose of this study is to provide basic data on effective support for multicultural families by examining the social perception of multicultural families through big data. To achieve this purpose, raw data was collected with ‘multicultural families’ keywords based on the big data provided by TEXTOM. The data collection process included the 1st and 2nd data cleansing processes and converted them into a 1-mode matrix dataset to perform keyword network analysis. In addition, LDA(latent dirichlet allocation) topic modeling analysis, which automatically derives topics through machine learning analysis using algorithms, was performed. Results: The social perception of multicultural families was found to include information on stakeholder group, stakeholder, policy & administration, social response and academic interest. Also, as a result of the topic modeling analysis on multicultural families, three topics (request for regionally customized multicultural family support benefits, request for a socially customized multicultural family support program, and request for research to develop family-centered support measures for multicultural families) were derived. Based on these results, not only was it possible to identify various considerations for supporting multicultural families, but it is also expected to provide meaningful implications for the preparation of measures for effective multicultural family support.

Key words: Big data, Multicultural families, Social perception, LDA topic modeling analysis, Keyword network analysis