

결혼이주여성 건강 담론에 대한 텍스트마이닝 분석: 언론 보도를 중심으로*

강영신**·송도언***

초 록

결혼이주여성의 건강 문제는 개인의 건강 상태를 넘어 사회적 담론 속에서 구성되고 이해되는 측면이 있다. 특히 언론은 특정 집단의 건강 이슈가 사회적으로 어떠한 의미로 해석되고 재현되는지에 영향을 미치는 중요한 담론 형성의 장이지만, 결혼이주여성의 건강 관련 보도에 대한 체계적 분석은 상대적으로 부족한 실정이다. 이에 본 연구는 언론 텍스트 분석을 통해 결혼이주여성의 건강 관련 담론이 한국 사회에서 어떠한 방식으로 재현되고 구성되는지를 탐색하고자 하였다. 이를 위해 전국 단위 주요 언론 보도에서 '결혼이주여성' 및 '건강' 관련 키워드가 포함된 기사들을 수집하고, 텍스트마이닝 기법을 활용하여 핵심 단어의 빈도와 중요도(TF-IDF) 및 잠재적 주제 구조를 분석하였다. 분석 결과, '다문화가족', '지원', '가족', '프로그램', '교육' 등이 주요 키워드로 나타났으며, TF-IDF 분석에서는 '다문화가족', '건강검진', '무료', '보건소', '프로그램' 등이 상대적으로 높은 중요도를 보였다. 토픽모델링 결과, 결혼이주여성의 건강 관련 보도는 토픽1은 '임산부 지원', 토픽2는 '의료지원', 토픽3은 '가정폭력 피해', 토픽4는 '다문화가족 지원'의 네 가지 주요 담론 주제로 구조화되었다. 이러한 결과는 결혼이주여성의 건강 이슈가 지원과 보호, 위험과 취약성 중심의 담론 구조 속에서 재현되고 있음을 보여준다. 본 연구는 결혼이주여성의 건강 이슈에 대해 언론 담론에서 나타난 사회적 구성 방식을 조명하였다는 점에서 의의를 지니며, 연구의 한계와 향후 연구 방향을 함께 논의하였다.

주제어 : 결혼이주여성, 건강 이슈, 언론 담론, 텍스트마이닝, 토픽모델링

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2021S1A5C2A02089141)

** 제1저자: 전남대학교 심리학과 교수(lavieenrose@jnu.ac.kr)

*** 교신저자: 전남대학교 심리학과 박사과정수료(vv8813@nate.com)

I. 서론

1. 연구의 필요성

국내 국제결혼은 2000년대 이후 한국 사회의 중요한 혼인 형태 중 하나로 자리 잡았으며, 최근에는 감소세 이후 다시 증가하는 추세를 보이고 있다(통계청, 2024). 결혼이민자의 규모 역시 꾸준히 유지되거나 증가하는 경향을 보이며, 2024년 기준 국내 체류 결혼이민자는 181,436명으로 전년대비 3.7% 증가하였다(법무부, 2025). 이러한 변화는 결혼을 기반으로 한국에 입국한 결혼이민자의 규모의 확대와도 연결된다(여성가족부, 2022). 최근 전국다문화가족 실태조사에 따르면 결혼이민자의 80% 이상이 여성으로 나타났으며, 연령대별로는 30~40대가 60% 가량을 차지하는 것으로 보고되었다(여성가족부, 2025). 나아가 다문화가족 자녀의 고등교육기관(대학 등) 순취학률이 61.9%로 크게 상승하고, 월평균 소득 300만 원 이상 가구 비율이 65.8%로 역대 최고치를 기록하는 등 다문화가족의 정착 양상이 점차 안정화되는 경향도 확인된다.

이러한 통계적 지표들은 결혼이주여성이 한국 사회의 가족 형성과 돌봄, 자녀 양육 과정에서 중요한 역할을 담당하는 집단임을 보여준다. 더불어 2024년도 11월 기준 지방자치단체 외국인주민 현황 자료에 따르면, 결혼이민자의 체류기간은 10년 이상이 51.2%로 나타나 외국인 근로자(0.5%), 유학생(0.3%), 외국국적동포(7.7%)와 비교해 현저히 높은 수준을 보였다(행정안전부, 2025). 이러한 통계는 결혼이주여성이 다른 이주민 집단과 비교해 장기 체류 비율이 높은 정착 집단이라는 점에서 한국 사회의 가족 변화와 사회통합을 이해하는 데 중요한 연구 대상임을 시사한다.

특히 결혼이주여성의 문제를 고찰할 때 건강은 핵심적인 의제라고 볼 수 있다. 건강은 개인의 신체적·정신적 안녕에만 국한되지 않고, 일상생활 수행, 노동시장 참여, 가족 관계 유지, 자녀 양육, 지역사회 참여를 가능하게 하는 기초 조건이기 때문이다(Marmot et al., 2008; World Health Organization, 2008). 결혼이주여성의 건강이 위협받을 경우, 이는 개인의 삶의 질 저하에 그치지 않고, 배우자 관계, 자녀의 발달과 적응, 가족 기능, 나아가 지역사회 통합에까지 부정적인 영향을 미칠 수 있다(Kim, Kim & Shin, 2010). 다시 말해 결혼이주여성의 건강은 특정 개인의 사적 문제가 아니라 가족 단위의 안녕과 사회적 지속 가능성에 직결되는 사회적 과제라고 할 수 있다.

2. 이론적 배경 및 선행 연구 검토

결혼이주는 ‘결혼’과 ‘이주’라는 두 가지 중대한 생애사적 사건이 동시에 일어나는 사건이라는 점에서, 결혼이주여성은 다른 이주민 집단과 구별되는 특수한 조건에 놓여 있다. 새로운 국가로의 이주는 언어, 생활 방식, 가치 체계, 가족 관계, 사회 규범 등 일상 전반에 걸친 적응을 요구하며, 이는 단순한 거주지 이동을 넘어서 정체성과 삶의 구조를 재구성하는 과정이기도 하다. Berry and Sam(1997)의 문화적응 이론이 시사하듯, 이러한 적응 과정은 이주민에게 심리적 스트레스와 사회적 긴장을 유발할 가능성이 높다. 여기에 결혼이주여성은 일반적인 문화적응 스트레스에 더해 배우자 가족과의 관계 형성, 한국식 양육 방식에 대한 압력, 경제활동의 제약 등 추가적인 부담을 경험할 수 있다는 점에서 복합적 취약성에 놓이기 쉽다(임혁, 2010).

선행 연구에서도 국내 결혼이주여성이 문화적응 스트레스, 언어 장벽, 사회적 고립, 배우자 및 시가 가족과의 관계 갈등, 양육 부담 등 다양한 영역에서 취약성을 보인다는 점이 반복적으로 확인되어 왔다(e.g., 정기선·한지은, 2009; 현경자·김연수, 2012). 특히 한국어 능력의 제한은 단순한 일상적 의사소통의 어려움에 그치지 않고, 의료기관 이용 과정에서의 의사소통 장벽과 치료 이해의 저해로 이어질 수 있으며(박지경, 2014), 건강정보 접근 및 활용 능력을 약화시켜 만성질환 관리와 같은 적극적 건강관리에도 부정적인 영향을 미칠 가능성이 있다(Kim & Son, 2020). 여기에 외국인에 대한 사회적 편견과 차별, 불안정한 체류 지위, 취업 기회의 제한, 경제적 어려움 등이 복합적으로 작용하면서 결혼이주여성의 건강은 구조적으로 취약한 환경 속에서 유지되기 쉽다(김현숙·김옥녀, 2017; 류한수, 2016; 이안나·강영신, 2023; 조선주·오현경·민현주, 2017; 최윤정·성미혜, 2017)

결혼이주여성의 건강 문제는 개인의 삶에 국한되지 않고 가족 전체의 기능과 안녕에도 중대한 영향을 미친다. 결혼이주여성의 정신건강과 적응 수준은 배우자의 심리적 안정, 자녀의 정서·행동·학업 적응, 가족 관계의 질 등 여러 측면과 밀접하게 연결되어 있음이 보고되어 왔다(강하라, 2021; 박현선, 2014; 보금로·장영은, 2022; 한수연 외, 2017). 예컨대 결혼이주여성의 우울이나 스트레스 수준이 높을수록 양육 스트레스가 증가하고, 이는 다시 자녀의 심리적 어려움과 학교생활 부적응으로 이어질 가능성이 높아진다(권태연, 2020; 이형하, 2021). 따라서 결혼이주여성의 건강은 단순한 개인의 신체적·정신적 상태를 넘어 가족 단위의 건강, 돌봄의 질, 지역사회 통합에 영향을 미

치는 핵심 요인으로 이해될 필요가 있다.

이러한 문제는 개인의 적응 실패나 문화적 차이만으로 설명되기보다 사회 구조적 맥락 속에서 이해될 필요가 있다. Crenshaw(1989)의 교차성(intersectionality) 개념에 따르면, 특정 집단의 취약성은 단일 요인으로 환원될 수 없으며, 성별, 인종·민족성, 사회경제적 지위, 이주 지위 등 다양한 정체성이 교차하면서 복합적으로 형성된다. 결혼이주여성은 여성이라는 성별 정체성, 이주민이라는 법적·사회적 지위, 상대적으로 낮은 사회경제적 조건, 문화적 소수자로서의 위치 등이 동시에 교차하며 구조적 취약성을 경험하는 집단일 수 있다(Castañeda et al., 2015; Derose, Escarce & Lurie, 2007). 특히 한국 사회의 건강 자원 접근성, 의료 제도의 이용 가능성, 공중보건 정책에 내재된 암묵적 편향 등은 결혼이주여성의 건강 불평등을 더욱 심화시킬 가능성이 있다.

건강은 객관적·의학적 지표로만 존재하는 것이 아니라, 특정 사회적 맥락 속에서 의미가 부여되고 책임이 귀속되며 해결 방식이 규정되는 사회적 구성물이다(Conrad & Barker, 2010; WHO, 2008). 결혼이주여성의 경우, 건강 문제는 종종 문화적 차이, 언어 능력 부족, 개인의 적응 실패와 같은 개인적 요인으로 설명되거나, 보호와 관리의 대상으로 표상되는 경향이 있다(최나연·이병숙, 2018; Yi & Lee, 2018). 이러한 담론적 구성은 결혼이주여성을 능동적인 건강 주체라기보다 취약하고 의존적인 존재로 고정시키는 동시에, 건강 불평등을 야기하는 사회 구조적 요인에 대한 주목을 약화시킬 위험을 내포한다. 즉, 건강 문제가 개인화되거나 문화화될수록 사회적 책임과 정책적 개입의 필요성은 상대적으로 가려질 수 있다.

언론은 이러한 건강 담론이 형성되고 확산되는 핵심 사회적 장치이다(Seale, 2003). 언론이 결혼이주여성의 건강을 어떤 문제로 설정하고, 어떠한 원인과 해결 방식을 제시하는가는 대중의 인식뿐 아니라 정책 의제 설정과 자원 배분의 방향에도 실질적인 영향을 미친다. 언론이 특정 집단을 지나치게 취약한 존재로 묘사하거나, 반대로 문제를 개인의 책임으로 한정해 다룰 경우, 사회 구성원들의 인식과 정책적 대응 역시 왜곡될 수 있다(Niederdeppe, Gollust & Barry, 2014). 따라서 결혼이주여성의 건강 문제가 언론에서 어떻게 다뤄지는지는 단순한 정보 전달을 넘어 사회적 의미 구성 과정에 직접적으로 개입하는 중요한 과정이다.

3. 연구 공백 및 연구 목적

그럼에도 불구하고 국내 연구는 결혼이주여성의 건강 문제를 주로 개인 경험이나 실태조사, 변수 간 관계 분석에 초점을 맞추어 왔으며(강영신 외, 2025), 접근 방식이나 주제가 제한적으로 다루어지는 경향이 있다(최지희 외, 2025). 그러나 언론이라는 사회적 재현 장치를 통해 건강 문제가 어떻게 구성되고 유통되는지를 체계적으로 분석한 연구는 상대적으로 부족하다. 국내에서 건강 관련 이슈를 대상으로 언론 보도나 사회적 담론을 분석한 연구들이 일부 수행되어 왔다. 예를 들어, 건강 및 사회문제에 대한 언론 보도의 프레이밍 방식이나 보도 경향을 분석한 연구들은 특정 이슈가 사회적으로 어떻게 구성되고 의미화되는지를 보여주었으며(e.g., 안순태·이하나, 2016; 정의철, 2008), 최근에는 텍스트마이닝 기법을 활용하여 뉴스 기사나 온라인 텍스트를 기반으로 주요 키워드의 구조와 주제 분포를 탐색하는 연구들도 점차 증가하고 있다(e.g., 김광기·제갈정·최명일, 2021; 김연정·김현실, 2025; 김주영·윤소원, 2025; 양혜승, 2022). 그러나 이러한 연구들은 다양한 건강 및 사회적 이슈를 다루어 왔음에도 불구하고, 결혼이주여성과 같은 특정 사회적 집단의 건강 문제를 언론 담론 차원에서 텍스트마이닝 기법을 활용하여 분석한 연구는 제한적으로 이루어져 왔다.

이러한 연구 공백을 보완하기 위해 본 연구는 인터넷 뉴스 기사를 분석 대상으로 삼아 결혼이주여성의 건강 이슈가 어떠한 방식으로 문제화되고 재현되는지를 탐색하고자 한다. 인터넷 뉴스는 다양한 사회적 관점과 담론이 반영되는 주요 매체로서(최영·김병철, 2000), 학문적 연구가 상대적으로 부족한 주제에 대한 탐색적 분석에 유용한 자료원이다. 또한 최근 텍스트마이닝(text-mining) 기법은 대규모 비정형 텍스트 자료를 체계적으로 분석하여 핵심 키워드, 의미적 연결 구조, 잠재적 주제와 담론의 특성을 파악하는 데 효과적으로 활용되고 있다(Aggarwal & Zhai, 2012; DiMaggio, Nag & Blei, 2013; Feldman & Sanger, 2007). 건강과 같이 사회적·문화적 의미가 복합적으로 얹힌 주제의 경우, 텍스트마이닝은 언론 담론의 경향과 의미 구성 방식을 비교적 객관적으로 조망할 수 있는 유용한 방법론이다(Dhenia, 2022; Zolnoori et al., 2019).

이에 본 연구는 언론 텍스트를 분석 대상으로 삼아, 결혼이주여성의 건강 관련 담론이 한국 사회에서 어떠한 방식으로 재현되고 구성되는지를 규명하고자 한다. 구체적으로, 전국 단위의 주요 언론 보도에서 결혼이주여성 건강

관련 키워드가 포함된 기사들을 수집하여, 해당 텍스트에 나타난 건강 이슈의 담론적 특징을 탐색하고자 하였다. 이를 위해 텍스트마이닝 기법을 활용하여 핵심 단어의 빈도와 중요도, 잠재적 주제 구성을 종합적으로 분석하였다. 이러한 분석을 통해 결혼이주여성의 건강 이슈가 언론에서 어떤 방식으로 문제화 되고, 어떤 관점과 프레임으로 재현되는지 파악함으로써, 기존의 개인 중심 연구나 실태조사에서는 충분히 포착하지 못했던 결혼이주여성 건강 이슈의 담론적 특성을 조명하고자 한다. 이에 따라 본 연구는 다음과 같은 연구 문제를 설정하였다.

4. 연구 문제

1. 결혼이주여성 건강 관련 뉴스 기사에서 나타나는 주요 핵심어는 무엇인가?
2. 결혼이주여성 건강 이슈는 어떠한 잠재적 주요 주제(topic) 구조를 형성하는가?

II. 연구 방법

1. 분석 대상

결혼이주여성 건강에 관한 텍스트마이닝 분석을 수행하기 위해 네이버 뉴스 기사를 분석 대상으로 선정하였다. 네이버는 정치/경제, 의료/건강, 생활 정보 등 다양한 분야의 정보 탐색에 활용되는 국내 주요 포털사이트로, 2025년 기준 약 4,400만 명 이상의 월간 활성이용자를 보유한 대표적인 플랫폼이며(아이지에이웍스, 2025), 국내 검색엔진 시장 점유율 약 62.86%를 차지하는 국내 1위 검색엔진이다(연합뉴스, 2026). 특히 네이버 뉴스는 2025년 11월 기준 약 900여 개 이상의 제휴 언론사를 보유한 국내 최대 규모의 인터넷 뉴스 플랫폼이다(AP신문, 2025). 이는 다양한 언론사의 보도를 종합적으로 포함할 수 있다는 점에서 분석 자료의 폭을 확장하는 데 유용하다. 실제로 동일 이슈에 대한 키워드 분석을 하더라도 언론사에 따라 강조되는 키워드와 서술 방식이 상이하게 나타날 수 있다는 점을 고려할 때(감미아·송민, 2012), 네이버 뉴스 기사는 결혼이주여성 건강과 관련된 다양한 담론적 양상을 포괄적으로 탐

색할 수 있는 자료로서 적합하다.

2. 자료 수집

연구주제에 적합한 기사를 수집하기 위해 ‘결혼이주여성’과 ‘건강’을 검색어로 사용하였다. 또한, 연구주제와 연관성이 떨어지는 기사가 검색 결과에 포함되는 것을 방지하기 위해 ‘이주민’과 ‘이주노동자’를 제외 검색어로 설정하여 해당 단어를 포함한 기사는 검색 결과에서 제외되도록 하였다. 그러나 검색 필터에도 불구하고 검색 결과에 주제 연관성이 떨어지는 기사가 다수 포함되어, 웹 크롤링(web crawling)을 통해 기사 데이터를 일괄 추출하는 방식 대신, 연구자 1인이 개별 기사의 제목 및 본문을 직접 검토한 후 LISTLY 브라우저 확장 프로그램을 사용하여 각 기사의 텍스트 데이터를 추출하였다.

이 과정에서 결혼이주여성 건강과 직접적인 관련성이 없는 기사, 단순 언급에 그치는 기사, 행사 및 일반 보도 기사 등 기사 내 일부 내용에 결혼이주여성이 언급되었더라도 연구주제와 실질적 관련성이 낮은 경우는 분석에서 제외하였다. 예를 들어, 2024년 5월 15일 ‘여수 흥국사 봉축법요식 봉행... 불자 마음도 활짝’이라는 기사처럼 사실상 결혼이주여성 건강과는 무관하지만 기사 내의 특정 내용(예: 이주여성의 인터뷰)으로 인해 검색 결과에 노출된 경우는 결혼이주여성 건강과 직접적인 관련이 없어 제외하였다. 이처럼 기사 내용의 적합성을 확인하는 1차 자료 수집을 마친 후에는 다문화 및 이주 관련 연구 경험을 가진 박사 2인이 기사 내용을 재검토하여 자료의 적합성을 재확인하고 최종 분석 자료를 선정하였다.

자료 수집 기간은 2010년부터 2025년 10월 31일까지로 최근 약 15년으로 설정하였다. 기사 검색을 2010년부터 설정한 이유는 2010년 전후로 결혼이주여성의 건강 관련 연구와 논의가 점차 증가하고(김이선, 2010), 같은 시기 제1차 다문화가족정책 기본계획이 수립되어 결혼이주여성을 포함한 다문화가족에 대한 국가 차원의 정책적 개입이 체계화되기 시작하였기 때문이다(정현주, 2014). 이에 2010년은 관련 담론과 정책 환경이 동시에 가시화된 시점으로 판단하였다. 최종 수집된 자료의 연도별 현황은 <표 1>에 제시하였다.

〈표 1〉 자료 수집 현황

매체	네이버 뉴스 기사					
검색 기간	2010.01.01. ~ 2025.10.31.					
검색어	결혼이주여성 + 건강					
총 수집량	1,200건					
연도별 수집량	2010년	109건	2016년	63건	2022년	16건
	2011년	112건	2017년	68건	2023년	60건
	2012년	80건	2018년	59건	2024년	127건
	2013년	59건	2019년	112건	2025년	96건
	2014년	89건	2020년	35건		
	2015년	79건	2021년	36건		

3. 자료 분석

최종 수집된 자료에 대한 텍스트마이닝 분석에는 RStudio 2022.02.0+443 프로그램이 활용되었다. 형태소 분석에는 KoNLP(Korean Natural Language Processing) 패키지를 사용하였고, 한국어의 특성을 반영한 정교한 분석을 위해 K-ICT 빅데이터센터에서 배포한 약 93만 개의 어휘가 수록된 NIADic(NIA 사전)을 참조 사전으로 설정하였다(K-ICT 빅데이터센터, 2017).

분석 전, 기사 원문에 대한 두 단계의 전처리 작업이 이루어졌다. 먼저 텍스트 내의 영문, 한글 단모음/단자음, 숫자, 특수문자 등을 제거하는 문서 전처리 작업을 수행하였다. 첫 단계에서는 단일 명사를 추출하기 위해 단어를 결합·분리시키는 띄어쓰기 교정도 이루어졌는데, 대표적으로 ‘가정 폭력’, ‘의료 봉사’, ‘건강 교실’과 같은 단어는 모두 붙여서 하나의 단어로 인식하도록 하였다. 반대로, ‘가정폭력범죄’, ‘무료건강검진’, ‘정신건강증진’과 같은 단어는 ‘가정폭력 범죄’, ‘무료 건강검진’, ‘정신건강 증진’처럼 분리하여 각각의 명사를 따로 인식하도록 작업하였다.

1차 문서 전처리를 완료한 후 두 번째 단계로, 불용어 처리 및 단어 통합 전처리를 수행하였다. 먼저 불용어 처리에서는 기사 원문에 대한 형태소 분석을 거친 뒤 추출된 명사를 바탕으로 무의미 단어와 분석 목적에 부합하지 않는 고빈도 단어를 제거하였다. 예를 들어, 검색어로 사용된 ‘결혼이주여성’과 ‘건강’은 대부분의 기사에 공통적으로 포함되는 단어이므로 분석 결과에서 제외하였고, ‘이요’, ‘때문’, ‘나타났’과 같은 무의미 단어도 제거하였다. 이와 같은

단어가 추출되는 이유는 한국어의 교착어적 특성으로 인해 형태소 분석 과정에서 용언의 어간이 명사로 오분류되거나, 사전에 등재되지 않은 표현이 불완전하게 분리되기 때문이다. 이러한 단어들은 연구 주제와 직접적인 의미적 관련성이 낮고 분석 결과의 해석을 저해할 수 있으므로, 전처리 과정에서 불용어로 분류하여 제거하였다. 또한 전처리 과정에서 추출된 명사를 검토하여 동일하거나 유사한 의미를 지닌 단어는 하나의 대표어로 통합하였다. 본 연구에서는 별도의 유의어 사전을 구축하여 적용하기보다는 연구자가 추출 결과를 검토하는 과정을 통해 의미적으로 동일하거나 유사한 단어를 대표어로 정리하였다. 대표적으로 ‘다문화가정’은 ‘다문화가족’으로, ‘폭행’은 ‘폭력’으로 통합하였다.

두 차례의 전처리 작업을 마친 뒤 텍스트마이닝 분석을 위해 빈도분석과 토픽모델링이 실시되었다. 빈도분석의 경우, 명사 출현 빈도를 계산한 단순 빈도와 문서 내에서 단어의 중요도를 나타내는 TF-IDF(Term Frequency-Inverse Document Frequency) 가중치를 계산하여 결과에 제시하였다(Qaiser & Ali, 2018). 주요 키워드를 보다 직관적으로 확인할 수 있도록 빈도분석 결과를 워드클라우드로 시각화하였고, 단순 빈도와 TF-IDF를 따로 구분하여 제시하였다. TF-IDF 산출 시 문서(document)는 분석에 사용된 개별 텍스트 레코드(N=7,481)로 정의하였으며, 전체 레코드 집합을 코퍼스(corpus)로 사용하였다.

본 연구는 시기별 또는 세부 문서 간 비교보다는 결혼이주여성 건강 관련 언론 담론 전반에서 상대적으로 중요하게 나타나는 핵심 키워드를 탐색하는데 목적이 있었다. 이에 TF-IDF는 단순 빈도분석에서 높게 나타날 수 있는 일반적 단어의 영향을 줄이고, 개별 기사에서 상대적으로 중요하게 사용되는 단어의 중요도를 반영하기 위한 지표로 활용하였다. 또한 각 문서에서 산출된 TF-IDF 값을 동일 단어를 기준으로 전체 문서에 걸쳐 합산하여 집계하였다. 이는 전체 코퍼스에서 특정 단어가 갖는 누적적 중요도를 반영하기 위한 것으로, 개별 문서 수준의 중요도보다 전체 담론 수준에서 중요하게 다루어진 핵심 키워드를 식별하기 위해 단어별 TF-IDF 합산값을 활용하였다.

더불어, LDA(Latent Dirichlet Allocation) 토픽모델링을 통해 기사 내의 주요 키워드 간의 군집을 확인하였다. LDA 토픽모델링은 전체 문서 내에 잠재된 토픽을 식별하고, 각각의 토픽별 주요 키워드를 분석하는 대표적인 토픽모델링 기법이다(박준형·오효정, 2017). LDA 모델의 분석은 R의 topicmodels 패키지를 활용하였으며, 변분 기대최대화(Variational Expectation-Maximization, VEM) 알고리즘을 적용한 LDA() 함수를 사용하였다. topicmodels 패키지의

VEM 방식은 문서-토픽 분포와 토픽-단어 분포를 추정하며, 기본 설정에 따라 문서-토픽 사전분포 파라미터인 α (알파)와 토픽-단어 사전분포 파라미터인 β (베타)를 반복적 최적화 과정을 통해 동시 추정한다. 본 연구에서는 별도의 하이퍼파라미터를 임의로 지정하지 않고, 패키지의 기본 VEM 추정 설정을 적용하여 토픽 구조를 추정하였다.

또한, 본 연구에서는 모델의 통계적 타당성을 확보하기 위해 혼란도(perplexity)와 토픽 일관성(topic coherence) 지수를 산출하였으며, 일반적으로 혼란도가 낮고 일관성이 높은 수준에서 토픽의 해석 가능성이 극대화된다는 선행 연구의 접근을 기반으로 최종 토픽 수를 확정하였다(Gan & Qi, 2021; Sievert & Shirley, 2014). 분석 결과의 시각화에는 LDAvis 패키지를 활용하였는데, 이는 다차원 척도법(MDS)을 통해 토픽 간의 거리와 관계를 평면상에 구현할 뿐만 아니라, 각 토픽을 구성하는 키워드의 돌출도(saliency)와 연관성을 직관적으로 파악할 수 있도록 돕는다.

III. 연구결과

1. 빈도분석 결과

빈도분석 결과, 단순 빈도상 주요 키워드는 ‘다문화가족’, ‘지원’, ‘가족’, ‘프로그램’, ‘교육’, TF-IDF에서는 ‘다문화가족’, ‘건강검진’, ‘무료’, ‘보건소’, ‘프로그램’ 등이 차례로 도출되었다. 표 2는 단순 빈도와 TF-IDF 각각의 상위 20개 키워드를 정리한 것이고, 그림 1과 그림 2는 단순 빈도와 TF-IDF 각각의 상위 100개 키워드를 워드클라우드로 시각화한 것이다.

단순빈도와 TF-IDF 모두에서 20위 안에 포함된 공통 상위 키워드는 총 17개로, ‘다문화가족’, ‘건강검진’, ‘무료’, ‘보건소’, ‘프로그램’, ‘지원’, ‘출산’, ‘가족’, ‘베트남’, ‘가족폭력’, ‘임산부’, ‘교육’, ‘임신’, ‘남편’, ‘사업’, ‘국제결혼’, ‘다문화’가 확인되었다. 17개의 공통 상위 키워드는 기사 내에서 자주 언급되었을 뿐 아니라 단순 반복된 수치를 제외하고도 문서 내의 중요도가 높게 나타난 핵심어에 해당한다. 특히, 단순빈도와 TF-IDF 모두에서 5위 안에 포함된 ‘다문화가족’과 ‘프로그램’은 네이버 뉴스 기사에서 다루어지는 결혼이주여성 건강에 관한 주요 이슈와 담론을 가장 잘 대표하는 키워드라고 할 수 있다.

두 단어가 포함된 기사 원문을 살펴본 결과, 결혼이주여성의 건강 증진과

관련한 다양한 의료/건강 지원사업이나 다문화가족 복지지원 프로그램을 소개하는 기사가 전체 문서에서 많은 비중을 차지하고 있었다. 대표적인 기사 원문으로는 ‘전주시다문화가족지원센터는 … 정신적 스트레스를 호소하는 결혼이주여성 20명을 대상으로 … 마음챙김프로그램을 비대면 방식인 온라인으로 실시한다’, ‘결혼이주여성을 위한 영양교육 프로그램을 운영 …’, ‘보건소·다문화가족지원센터와 협력해 베트남 결혼이민자를 대상으로 프로그램을 운영 …’ 등이 있었다.

〈표 2〉 Top 20 키워드

순위	단순 빈도		TF-IDF	
	키워드	빈도수	키워드	TF-IDF
1	다문화가족	1715	다문화가족	268.965
2	지원	1098	건강검진	182.0808
3	가족	911	무료	178.5961
4	프로그램	891	보건소	157.2585
5	교육	749	프로그램	155.0055
6	베트남	727	지원	145.2944
7	출산	706	출산	140.6964
8	보건소	598	가족	132.7579
9	사업	545	베트남	129.6949
10	임신	544	가족폭력	120.943
11	무료	538	임산부	119.5879
12	남편	493	교육	119.4978
13	가족폭력	415	임신	115.3649
14	도움*	409	폭력	109.2578
15	임산부	401	남편	107.9491
16	건강검진	398	사업	99.86608
17	문제	383	국제결혼	97.34963
18	서비스	355	경찰	97.19843
19	다문화	354	다문화	89.8387
20	국제결혼	350	의료봉사	86.0502

* 굵은 이탤릭체로 된 키워드는 단순 빈도와 TF-IDF의 순위 차이로 인해 다르게 도출된 키워드임.

단순빈도와 TF-IDF의 키워드별 순위 차이로 인해 각각 3개의 단어가 서로 다르게 나타났는데, 단순빈도에서는 ‘도움’, ‘문제’, ‘서비스’, TF-IDF에서는 ‘폭력’, ‘경찰’, ‘의료봉사’가 확인되었다. 이러한 차이는 키워드의 보편성과 특수성을 동시에 보여준다. 단순빈도 상위권인 ‘도움’, ‘문제’, ‘서비스’는 관련 보도 전반에 걸쳐 넓게 나타나는 일반적인 주제어다. 반면 TF-IDF에서 ‘폭

력’, ‘경찰’, ‘의료봉사’가 새롭게 부각된 것은 이 단어들이 특정 기사군에서 핵심적으로 다루어지는 경향이 있음을 의미한다. 특히, ‘폭력’과 ‘경찰’이 상위 키워드로 나타난 결과는, 결혼이주여성 건강 이슈가 단순한 질병 예방을 넘어, 가정폭력 피해 대응 및 치안 서비스와 밀접하게 결합되어 보도되는 경향을 반영한다. 이러한 경향은 “양평署-다문화가족지원센터, 가정·성폭력 대응 협약 체결”, “가정폭력 피해 이주여성에게 ‘구원 손길’ 내민 의정부지검”, “경찰은 가정폭력 발생 시 신속 대응 체계 구축 및 군 다문화가족지원센터와의 유기적인 협력과 지원을 통해 … 다문화 가정이 건강한 가정으로 정착될 수 있도록 지원할 예정이다” 등의 기사에서도 확인되며, 이는 결혼이주여성의 건강 이슈가 언론 보도에서 보호 체계나 사법적 대응의 맥락 속에서 강조되는 경향이 있음을 보여준다.



[그림 1] 단순빈도 Top 100
워드클라우드

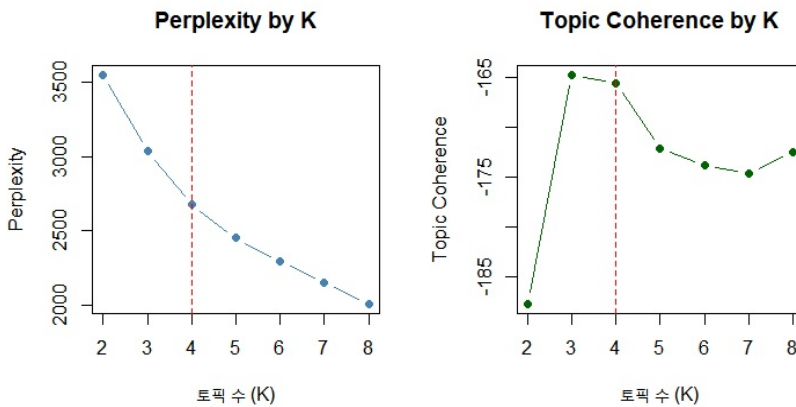


[그림 2] TF-IDF Top 100
워드클라우드

2. 토픽모델링 결과

결혼이주여성의 건강에 관한 기사를 바탕으로 토픽모델링을 실시한 결과, 주요 토픽은 총 4개가 확인되었다. 토픽 수 결정을 위해 $K=2\sim 8$ 범위에서 혼란도와 토픽 일관성을 산출한 결과, 혼란도는 $K=4(2674.08)$ 이후 감소폭이 현저히 둔화되었으며, 토픽 일관성은 $K=3(-164.70)$ 에서 가장 높았으나 $K=4(-165.58)$ 와의 차이가 0.88로 미미하였으며, $K=5(-172.15)$ 부터 급격히 하락하는 양상을 보였다. 이에 혼란도와 토픽 일관성 두 지표의 균형을 고려하여 최적 토픽 수로 $K=4$ 를 결정하였다. 그림 3에서 확인할 수 있듯이, 혼란도

곡선은 K=4 지점에서 기울기가 완만해지는 형태를 보이고, 토픽 일관성은 K=4를 전후로 하락하는 양상을 나타낸다. 모델의 주요 파라미터 자동 추정 결과, α 값은 0.076으로 확인되어 각 문서가 소수의 토픽에 집중되는 본 연구 데이터의 특성이 반영되었다. β 값은 단어별로 개별 추정되었으며, 0.00002~0.018 범위로 나타났다. 반복 횟수는 최대 1,000회로 설정된 가운데 Log-Likelihood 변화량이 수렴 기준값(0.0001) 이하로 떨어지는 시점에 자동 종료되었다.



[그림 3] 토픽 수에 따른 혼란도 및 토픽 일관성 지수 변화

최종 추출된 토픽의 주제어(토픽명)는 군집별 주요 키워드와 더불어 γ 값 상위 대표 문서에서 확인된 언론 담론을 분석하여 설정하였는데, 표 3에 제시된 바와 같이 각 토픽별 상위 문서의 γ 값은 0.997~0.999로 나타나, 해당 문서들이 특정 토픽의 특성을 잘 반영하는 것으로 확인되었다.

먼저 토픽1은 결혼이주여성의 임신 및 출산에 관한 보건 지원을 담고 있어 ‘임산부 지원’으로 명명하였다. 토픽2는 결혼이주여성을 포함한 이민자 전반의 의료서비스 접근 문제나 치료 및 검진 지원에 관한 언론 담론을 담고 있어 ‘의료지원’으로 명명하였다. 다음 토픽3은 결혼이주여성의 신체적 및 정신적 안전을 위협하는 가정 내의 폭력 문제에 관한 언론 담론이 포함되어 있어 ‘가정폭력 피해’로 명명하였다. 마지막으로, 토픽4는 결혼이주여성의 사회적 지원 및 통합에 관한 담론이 주를 이루고 있어 ‘다문화가족 지원’으로 명명하였다.

〈표 3〉 토픽별 γ 값 및 상위 대표 문서

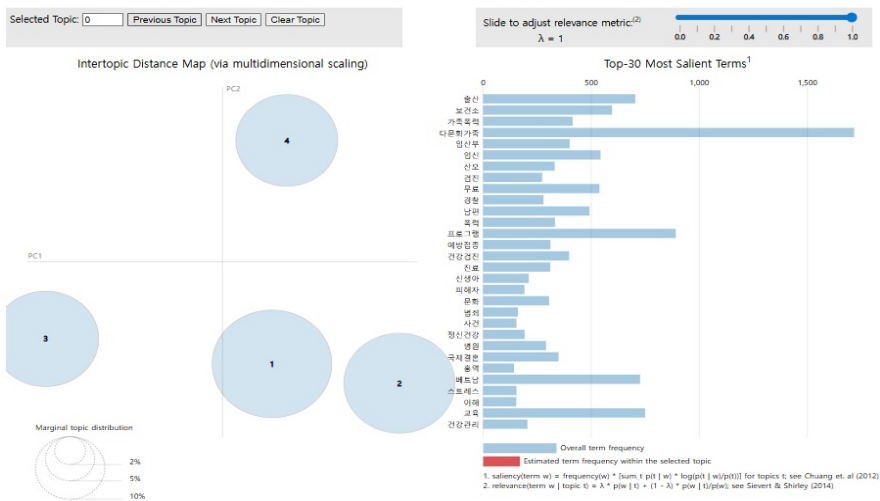
토픽번호	원자료 식별번호*	γ 값	기사내용 요약
1	655	0.999	결혼이주여성 산모돌보미 서비스
	1071	0.998	결혼이주여성 난소암 투병 지원
	1391	0.998	결혼이주여성 산모돌보미 서비스
	1263	0.998	결혼이주여성 산모돌보미 서비스
	800	0.998	결혼이주여성 산모돌보미 서비스
2	2286	0.999	시골병원의 실태, 결혼이주여성 등 많아
	1870	0.998	가톨릭근로자회관 이민자 지원 의료봉사상
	2334	0.998	전남의사회 '이주민 찾아가는 진료'
	2382	0.998	전남의사회 '이주민 찾아가는 진료'
	880	0.998	광주시이주민건강센터, 의료격차 해소 앞장
3	671	0.999	결혼이주여성, 남편 폭언 및 폭행 해방
	3036	0.998	범죄피해 이주여성 보호·지원 협의체
	11970	0.998	한국 남편에 의한 결혼이주여성 피살 사건
	497	0.997	국제결혼중개업 법률의 인권 규정 신설
	431	0.997	'베트남 아내 살해' 남성 징역 7년 선고
4	2078	1.00	제주국제가정문화원 다문화 도민의식 배양
	2542	0.999	다문화가정과 따뜻한 동행, 한국어 강습 등
	911	0.999	러브인 제주 사업, 다문화 지원
	1264	0.999	결혼이주여성 건강 보고서, 지원체계 고민
	80	0.999	의료통역 예약 앱, 의료취약계층 권리증진

* 원자료 식별번호는 데이터 관리 및 추적을 위해 부여된 고유 식별자이며, 기사 수(1,200건)나 문서 수(N=7,481)와는 무관함.

각각의 토픽명과 토픽 내 상위 10개 키워드를 정리하여 표 4에 제시하였고, 토픽 군집 간의 구조적 위치 및 관계를 LDAvis로 시각화하여 그림 4에 제시하였다. 그림 4의 LDAvis 시각화에서 각 원은 하나의 토픽을 의미하며, 원의 크기는 전체 코퍼스에서 해당 토픽이 차지하는 비중을 나타낸다. 전체 문서에서 각 토픽이 차지하는 비율은 토픽1 28.6%, 토픽2 21.8%, 토픽3 27.5%, 토픽4 22.2%로 나타나 네 토픽이 비교적 균등하게 분포된 것으로 확인되었다. 원 간의 거리가 멀수록 토픽 간 내용적 차별성이 크다는 것을 의미하며, 본 연구의 4개 토픽은 원들이 서로 겹치지 않고 분리되어 있어 각 토픽이 변별적인 주제를 형성하고 있음을 보여준다.

〈표 4〉 주요 토픽 및 Top 10 키워드

순위	토픽1 (28.6%)	토픽2 (21.8%)	토픽3 (27.5%)	토픽4 (22.2%)
	임산부 지원	의료지원	가정폭력 피해	다문화가족 지원
1	출산	무료	가족폭력	다문화가족
2	보건소	건강검진	남편	프로그램
3	임신	검진	베트남	가족
4	다문화가족	진료	폭력	교육
5	임산부	다문화가족	경찰	지원
6	지원	지원	국제결혼	베트남
7	산모	병원	가족	문화
8	예방접종	치료	피해자	다문화
9	사업	검사	지원	사회
10	교육	이주민	범죄	지역사회



[그림 4] LDAvis를 활용한 토픽 간 거리 및 분포 시각화

먼저 토픽1 ‘임산부 지원’은 주요 키워드로 ‘출산’, ‘보건소’, ‘임신’, ‘다문화가족’, ‘임산부’, ‘지원’, ‘산모’, ‘예방접종’, ‘사업’, ‘교육’이 확인되었다. 토픽 1은 결혼이주여성의 임신과 출산이라는 생애주기적 건강 이슈를 중심으로 구성되어 있다는 점에서, 의료지원이나 다문화가족 지원과 같이 보다 포괄적인 지원 체계를 다룬 다른 토픽들과 구별된다. 해당 토픽에서는 결혼이주여성의 임신·출산 과정에서 나타나는 건강 문제와 이에 대응하기 위한 정부, 지방자

치단체, 보건소 및 민간지원의 지원 활동이 주요하게 다루어졌다. 특히 산전·산후 관리, 예방접종, 건강교육, 출산 지원 서비스 등이 핵심 내용으로 나타나 결혼이주여성의 모성 건강 증진에 초점을 둔 담론이 형성되고 있음을 보여준다. 기사 원문을 살펴보면, “강북구 보건소가 결혼이주여성의 출산과 육아를 돕기 위해 나섰다”, “한림대학교강남성심병원과 서울시가 … 임신과 출산을 준비 중인 서울 서남부지역의 결혼이주여성을 초청해 건강한 출산을 지원하고 임신과 출산에 대한 건강지식과 올바른 육아 정보를 전달할 계획이다”와 같은 내용을 확인할 수 있었다.

다음 토픽2 ‘의료지원’의 경우, ‘무료’, ‘건강검진’, ‘검진’, ‘진료’, ‘다문화가족’, ‘지원’, ‘병원’, ‘치료’, ‘검사’, ‘이주민’이 주요 키워드로 나타났다. 해당 키워드를 포함하는 기사들은 주로 결혼이주여성을 위한 무료 건강검진과 같은 의료복지 서비스에 대한 내용들로 이루어졌는데, 결혼이주여성의 신체 건강에 대한 의료지원 사업이나 의료봉사를 다루는 글이 많았다. 예를 들어, 국민건강보험공단이나 병원, 보건소와 같은 기관에서 결혼이주여성이 대한민국 국민으로서 기본적인 건강권을 보장받을 수 있도록 지원하는 사회봉사나 복지지원을 소개하는 기사가 주를 이루었다. 대표적인 기사 원문으로는 “보건의료정보에 취약해 건강검진 기회가 적은 결혼이주여성들을 대상으로 … 부여군보건소 건강교육실에서 무료 건강검진을 실시한다”, “지난 4월 경북도와 도내 41개 병원 간 체결한 결혼이민여성 무료건강검진 업무협약”, “국민건강보험공단 … 의료봉사단은 … 인천시의료원 의료진과 함께 저소득 다문화 가정 100여 명에게 의료봉사활동을 펼쳤다” 등을 확인할 수 있었다.

토픽3 ‘가정폭력 피해’의 주요 키워드는 ‘가족폭력’, ‘남편’, ‘베트남’, ‘폭력’, ‘경찰’, ‘국제결혼’, ‘가족’, ‘피해자’, ‘지원’, ‘범죄’로 나타났다. 이러한 키워드 구성은 결혼이주여성의 건강 이슈가 언론 보도에서 가정폭력과 관련된 문제와 함께 다루어지는 경향이 있음을 보여준다. 특히 ‘2010년 베트남 신부 살해사건’과 ‘2019년 베트남 결혼이주여성 가정폭력 사건’과 같은 특정 사건이 관련 보도에서 반복적으로 언급되면서, 해당 토픽 형성에 영향을 미친 것으로 해석된다. 이는 언론보도에서 결혼이주여성의 건강 이슈가 개별 사건 중심의 폭력 사례와 결부되어 재현되는 양상이 존재함을 시사한다. 실제 기사에서는 “베트남인 아내를 상습적으로 무차별 폭행한 한국인 남편에 대한 …”, “가정폭력 피해 이주여성에게 ‘구원 손길’ 내민 의정부지검”, “국제결혼 연평균 1만 8000쌍…빈곤 여성 많아 잦은 폭력에도 하소연 못해” 등의 서술이 확인되었다. 이러한 보도 내용은 결혼이주여성의 건강 문제가 언론 담론에서 가족폭력, 범죄, 피해 및 지원과 관련된 맥락 속에서 구성되고 재현되는 경향이 있음을 보여준

다. 따라서 토픽3은 결혼이주여성의 건강 이슈가 언론 보도에서 가정폭력과 같은 사회적 위험과 연결된 문제로 의미화되는 담론적 특성을 반영하는 것으로 해석할 수 있다.

마지막으로 토픽4 ‘다문화가족 지원’의 주요 키워드는 ‘다문화가족’, ‘프로그램’, ‘가족’, ‘교육’, ‘지원’, ‘베트남’, ‘문화’, ‘다문화’, ‘사회’, ‘지역사회’가 확인되었다. 주요 키워드를 포함한 기사 내용을 확인한 결과, 건강가정·다문화가족지원센터나 다문화가정센터와 같은 기관을 중심으로 결혼이주여성을 포함한 다문화가족이 국내 생활에 적응하고 지역사회에 안정적으로 정착할 수 있도록 돕기 위한 다양한 지원 프로그램이 이루어지고 있음을 알 수 있었다. 주요 프로그램 내용은 영양, 사회적응, 출산, 건강검진, 정서 안정 등으로 매우 다양하였는데, 대체로 이주여성의 사회적응과 심리적·신체적 건강을 지원하고 돌보고자 하는 목적의 프로그램이 주를 이루고 있었다. 이와 관련한 기사 원문을 살펴보면, “전주시다문화가족지원센터는 코로나19 여파로 ... 정신적 스트레스를 호소하는 결혼이주여성 20명을 대상으로 ... ‘마음챙김프로그램’을 비대면 방식인 온라인으로 실시한다”, “대구 달성군가족센터, 결혼이주여성과 설맞이 프로그램 전개 ... 다양한 나라 결혼이주여성 한자리에 모여 모국에 대한 그리움 해소”, “우리다문화가정센터는 이주 결혼여성의 순조로운 출산을 위해 ‘HI~! 나의 아가’ 프로그램을 진행하고 있다 ...”, “우리다문화가정센터는 다문화 가정의 건강과 행복한 생활을 할 수 있도록 사회적응교육프로그램과 다문화가정 아동을 위한 프로그램을 운영하는 등 ...”과 같은 내용을 확인할 수 있었다.

IV. 논의

1. 주요 연구결과에 대한 논의

본 연구는 텍스트마이닝 분석을 통해 지난 15년간 한국 사회에서 결혼이주여성의 건강 이슈가 어떠한 방식으로 구성되고 논의되어 왔는지를 탐색하는 것을 목표로 하였다. 이를 위해 빈도분석, 토픽모델링을 주요 분석 기법으로 활용하여, 총 1,200건에 해당하는 뉴스 기사를 분석 대상으로 삼아 주요 키워드와 잠재적 주제 구조를 종합적으로 이해하고자 하였다. 분석 결과를 종합하면, 결혼이주여성 건강 담론은 크게 세 가지 특징적인 의미 구조 속에서 구성

되고 있는 것으로 나타났다.

첫째, 결혼이주여성의 건강 문제는 ‘다문화가족’, ‘가족’, ‘납편’, ‘임신’, ‘임산부’와 같은 키워드를 중심으로 가족 형성과 재생산의 맥락 속에서 이해되는 경향이 강하게 나타났다. 단순 빈도 분석과 TF-IDF 분석 모두에서 ‘다문화가족’이 가장 높은 빈도와 중요도를 보였다는 점과, 토픽모델링에서 ‘다문화가족 지원’이 주요 주제로 도출된 점은 결혼이주여성의 건강 문제가 개인의 건강 문제라기보다 다문화가족 정책 범주 속에서 이해되는 경향이 있음을 보여준다. 특히 ‘건강검진’, ‘보건소’, ‘프로그램’, ‘지원’, ‘출산’, ‘임신’, ‘임산부’ 등의 키워드가 상위에 위치한 결과는 결혼이주여성의 건강이 예방적 의료서비스 이용과 출산·양육의 맥락 속에서 주로 의미화되고 있음을 보여준다.

이러한 결과는 한국의 결혼이주 정책이 오랫동안 저출산 대응과 가족 형성의 맥락 속에서 추진되어 온 사회적 배경과도 관련이 있는 것으로 볼 수 있다(김경숙, 2014; Kim & Kilkey, 2018). 선행 연구에서도 결혼이주여성은 가족 형성과 돌봄의 핵심 주체로 기대되는 위치에 놓여 있으며, 이러한 사회적 기대가 여성의 건강 경험과 삶의 조건에 영향을 미친다고 보고한다(이봉숙, 2010; 최지희 외, 2025). 따라서 언론 담론에서 결혼이주여성의 건강은 개인의 건강권이나 삶의 질의 문제이면서도, 가족 형성과 유지에 필요한 요소로 더욱 강조되는 경향을 보이는 것으로 해석할 수 있다.

둘째, ‘건강검진’, ‘보건소’, ‘무료’, ‘프로그램’, ‘사업’, ‘서비스’, ‘의료봉사’와 같은 키워드가 높은 중요도를 보인 점과 토픽모델링에서도 ‘의료지원’이 주요 주제로 도출된 점은 결혼이주여성의 건강 문제가 주로 공공 의료서비스 접근성과 지원 프로그램 중심으로 재현되고 있음을 보여준다. 언론 보도에서 결혼이주여성의 건강은 질병 치료나 장기적인 건강관리의 관점보다는 예방적 건강검진이나 공공보건 서비스 이용과 같은 지원 체계의 맥락 속에서 언급되는 경우가 많았다. 특히 ‘무료’, ‘지원’, ‘프로그램’과 같은 표현이 반복적으로 등장하는 양상은 결혼이주여성의 건강 문제가 개인의 권리나 역량의 관점에서도 이해될 수 있음에도 불구하고, 언론 담론에서는 공공복지 차원의 지원 대상이라는 관점이 상대적으로 더 두드러지게 나타나고 있음을 보여준다. 즉, 언론 담론 구조 속에서 결혼이주여성은 자신의 건강을 능동적으로 관리하고 의사결정에 참여하는 주체라기보다 제도적 지원의 대상으로 재현되는 경향이 나타난다.

이러한 보도 경향은 결혼이주여성을 건강관리의 적극적 참여 주체로 바라보는 관점이 언론 담론에서 상대적으로 제한적으로 반영되어 왔음을 보여준다. 해외 일부 국가에서 포레 이주민을 건강 리더(cultural health navigator 또는 peer health navigator)로 양성하여 의료서비스 이용 과정에서 정보 제공

과 의사소통을 지원하도록 하는 사례(e.g., Said et al., 2025; Schuster et al., 2024)는 결혼이주여성을 지원의 대상으로 위치 짓는 담론과는 다른 관점이 존재할 수 있음을 보여준다. 이는 결혼이주여성의 건강 문제가 단순한 지원 대상의 문제로만 이해되기보다, 건강관리 과정에 참여하는 주체적 행위자의 관점에서도 재해석될 필요가 있음을 시사한다고 볼 수 있다.

셋째, ‘가족폭력’, ‘폭력’, ‘경찰’, ‘도움’, ‘문제’와 같은 키워드가 동시에 등장하는 양상은 결혼이주여성의 건강 문제가 의료적 차원을 넘어 가정폭력과 같은 사회적 위험과 연결된 문제로 재현되고 있음을 보여준다. 구체적으로, ‘가족폭력’, ‘납편’, ‘경찰’과 같은 단어들이 함께 등장하는 패턴은 결혼이주여성의 건강 문제가 친밀한 관계 내 폭력과 사회적 보호 문제와 밀접하게 연결되어 인식되고 있음을 시사한다. 이는 결혼이주여성의 건강이 단순한 신체적 건강 상태의 문제가 아니라 관계적 환경과 사회적 조건 속에서 형성되는 문제로 이해되고 있음을 보여준다. 또한 토픽모델링 결과에서 ‘가정폭력 피해’가 독립된 주제로 도출된 점은 결혼이주여성의 건강 문제가 의료적·신체적 차원에 국한되지 않으며, 사회적 관계와 제도적 조건 속에서 형성되는 구조적 위험요인과 긴밀히 연결되어 있음을 분명히 보여준다. 특히 언론 보도에서는 결혼이주여성이 언어 장벽, 체류 자격 의존성, 사회적 지지망의 제한 등 다양한 구조적 취약성과 관련된 맥락 속에서 재현되는 경향이 나타났다(김현숙, 2011; Choi & Byoun, 2014; Jang et al., 2013). 이러한 조건은 친밀한 관계 내 폭력 상황에서도 도움을 요청하거나 관계를 벗어나기 어렵게 만드는 요인으로 작용할 수 있다. 따라서 가정폭력 피해가 주요 담론 주제로 등장한 것은 결혼이주여성의 건강 문제가 단순한 의료 서비스 이용의 문제가 아니라 관계적 환경과 제도적 보호 수준과 깊이 맞물려 있음을 보여주는 결과로 해석될 수 있다.

종합하면, 언론은 결혼이주여성의 건강을 단순한 의료적 문제가 아니라 임신·출산, 가족관계, 의료지원, 가정폭력 및 공공 지원이 복합적으로 얽혀 있는 사회적 건강 이슈로 재현하고 있는 것으로 나타났다. 특히 건강 담론이 임신·출산과 가족 형성, 공공 지원 서비스, 폭력 피해 대응을 중심으로 전개된 점은, 결혼이주여성의 건강이 개인의 건강권 차원에 국한되기보다 가족·복지·정책의 맥락에서 주로 다루어지고 있음을 보여준다. 이와 같은 담론 구성은 결혼이주여성의 건강이 가족 내 역할 및 사회적 보호의 틀 속에서 의미화되고 있음을 시사하며, 나아가 이들의 건강 이슈가 젠더, 이주, 가족, 사회적 안전망이 교차하는 다차원적 맥락에서 논의될 필요성을 제기한다.

2. 연구의 의의 및 한계

본 연구는 결혼이주여성의 건강 문제를 사회적 담론의 차원에서 이해하기 위해 네이버 뉴스 기사를 분석 대상으로 삼아 텍스트마이닝 기법을 적용하였다. 비정형 텍스트를 분석 가능한 사회적 자료로 전환함으로써 다문화 연구, 보건 연구, 심리학적 접근을 연결하는 융합적 분석 틀을 제시하였다는 점에서 학문적 의미를 지닌다. 특히 언론 담론을 통해 결혼이주여성의 건강 문제가 한국 사회에서 어떠한 방식으로 의미화되고 재현되는지를 탐색했다는 점에서 결혼이주여성의 건강 문제를 이해하는 또 하나의 분석 관점을 제공한다는 의의를 지닌다.

그러나 이러한 의의에도 불구하고, 텍스트마이닝 기법을 활용한 본 연구 설계는 몇 가지 한계를 지닌다. 첫째, 네이버 뉴스 기사는 사회적 인식과 언론 프레임을 파악하는 데 적합한 자료이지만, 결혼이주여성 본인이 실제로 사용하는 언어, 감정, 경험을 직접적으로 반영하지는 못한다. 둘째, 언론사는 이슈의 선정, 제목 구성, 표현 방식에서 고유한 편향을 가질 수 있으며, 특정 사건 중심 보도나 시기별 집중 보도로 인해 담론이 왜곡될 가능성이 있다. 또한 담론 생산 주체(언론, 정부, 시민 단체 등)에 따른 차별적 재현 양상을 충분히 비교하거나 검토하지 못했다는 한계가 있다. 특히 본 연구는 네이버 뉴스 기사를 분석 단위로 설정함에 따라 개별 언론사의 유형이나 특성(신문/방송, 중앙/지방)에 따른 차이 혹은 기사 분야별 분포(정치, 사회, 경제 등)를 충분히 고려하지 못하였으며, 이는 담론 형성의 맥락적 다양성을 충분히 반영하지 못하는 한계로 이어질 수 있다. 또한 언론사별 보도 비중이나 특성을 체계적으로 검토하지 못하였기 때문에, 본 연구에서 도출된 담론 구조가 특정 유형의 언론 보도에 의해 상대적으로 강화된 결과인지 확인하기 어려울 수 있다. 셋째, 모든 텍스트마이닝 연구와 마찬가지로 본 연구 또한 특정 검색어를 중심으로 데이터를 수집함에 따라 담론의 범위를 제한할 수 있다는 한계가 있다.

넷째, 텍스트마이닝은 방대한 데이터를 효율적으로 분석할 수 있으나, 언어의 맥락적 의미나 담화적 뉘앙스를 충분히 해석하기는 어렵다. 이를테면, 형태소 분석 과정에서 의미가 왜곡되거나 정보가 손실될 가능성이 있고, LDA 토픽모델링은 통계적 확률 기반 모델로, 연구자의 해석이 개입될 수 있다. 또한 기사 작성자의 의도나 기사 생산 맥락(예: 정부 보도자료, NGO 보도 등)을 구분하지 못함으로써 심층적 의미 분석의 제약이 존재한다. 다섯째, 네이버 뉴스는 한국에서 가장 다양한 언론사가 유입되는 플랫폼이지만, 전체 사회적 담

론을 완전히 대표한다고 보기는 어렵다. 즉, 포털 알고리즘이 노출되는 기사 종류와 순서를 결정하며, 제휴되지 않은 언론사의 뉴스는 수집되지 않을 수 있고, 포털 뉴스의 댓글, SNS 확산 구조 등은 분석에서 제외된다. 따라서 분석 결과는 네이버 뉴스라는 플랫폼에 제한한 사회적 담론의 일부라는 점을 고려해야 한다.

여섯째, 본 연구의 분석 자료가 특정 시기 네이버 기사에 한정되어 있어, 다른 언론 플랫폼과의 해당 시기별 차이점을 비교하여 포괄적으로 반영하지 못하였다. 따라서 연구 결과를 모든 사회적 담론이나 일반 대중의 인식으로 일반화하기에는 신중함이 요구된다. 일곱째, 본 연구는 ‘결혼이주여성’과 ‘건강’이라는 명시적 키워드를 중심으로 자료를 수집하였기 때문에, 해당 단어가 직접적으로 포함되지 않더라도 결혼이주여성의 건강과 관련된 내용을 다루는 일부 기사들이 분석에서 제외되었을 가능성이 있다. 이는 키워드 기반 자료 수집 방식이 지니는 한계로 볼 수 있다.

여덟째, 본 연구는 2010년부터 2025년까지의 자료를 활용하였으나, 전체 기간에 걸쳐 형성된 결혼이주여성 건강 관련 담론의 구조를 탐색하는 데 초점을 두었다. 따라서 시기별 사회적 사건이나 정책 변화에 따른 담론의 변화 양상은 분석 범위에 포함하지 않았다. 이에 따라 본 연구 결과는 해당 기간 동안 형성된 전반적인 담론 구조를 보여준다는 의의가 있으나, 담론이 시간의 흐름에 따라 어떠한 방식으로 변화하였는지까지 설명하지는 못한다. 아홉째, 본 연구는 결혼이주여성 건강 관련 언론 담론 자체를 탐색하는 데 초점을 두었으며, 일반 여성 건강 담론, 이주민 건강 담론, 다문화 정책 담론 등과의 비교 분석은 수행하지 않았다. 따라서 본 연구에서 도출된 주요 키워드와 토픽 구조가 결혼이주여성 건강 담론에 상대적으로 특징적인 현상인지, 또는 건강·여성·이주 관련 담론 전반에 공통적으로 나타나는 특성인지는 본 연구만으로 판단하기 어렵다. 따라서 연구 결과의 해석은 결혼이주여성 건강 관련 언론 담론의 특성을 탐색적으로 제시하는 수준에서 이해될 필요가 있을 것이다.

3. 후속연구 제언

이러한 한계를 보완하기 위해 향후 연구에서는 다음과 같은 방향을 고려할 필요가 있다. 첫째, 결혼이주여성 당사자의 언어를 직접 반영하는 데이터로의 확장이 필요하다. 예를 들어, 결혼이주여성 온라인 커뮤니티 게시글이나, 블로그, 유튜브 등 사용자 생성 콘텐츠(user-generated content) 등은 결혼이주

여성의 실제 경험, 감정, 요구를 보다 직접적으로 반영하므로, 언론 담론과 당사자 담론을 비교 분석하는 연구도 가능할 것이다. 둘째, 기사 작성자가 정부, 언론, 시민단체인지에 따라 시각이 다름에도 현재 분석은 이를 구분하지 못하였다. 후속 연구에서는 데이터 전처리 단계에서 개체명 인식(named entity recognition)을 활용하여 발화 주체별 토픽 구조를 분리하여 분석하는 방법을 활용해 볼 수 있을 것이다. 또한 향후 연구에서는 언론사 유형(신문/방송, 중앙/지방 등) 및 언론사별 보도 비중을 함께 분석함으로써, 결혼이주여성 건강 담론이 어떠한 매체 환경 속에서 주로 형성되고 확산되는지를 구체적으로 비교·검토할 필요가 있다.

셋째, 본 연구는 전체 기간의 담론 구조를 탐색하는 데 초점을 두었기 때문에 시기별 변화 양상을 분석하지 못하였다. 따라서 향후 연구에서는 시계열 분석을 통해 결혼이주여성 건강 담론의 변화 과정을 보다 구체적으로 검토할 필요가 있다. 예를 들어, 다문화가족 정책의 변화, COVID-19 팬데믹, 주요 가정폭력 사건, 이주민 관련 사회적 이슈 등을 기준으로 시기를 구분하여 토픽의 등장 및 변화 양상을 비교하거나, 동적 토픽모델링(dynamic topic modeling)과 같은 방법을 활용하여 담론의 장기적 변화 과정을 분석할 수 있을 것이다. 이를 통해 결혼이주여성 건강에 대한 사회적 관심이 어떠한 계기와 맥락 속에서 변화해 왔는지를 보다 심층적으로 이해할 수 있을 것이다.

넷째, 텍스트마이닝은 패턴 탐색에는 강하지만, 의미와 맥락 해석에는 한계가 있다. 따라서 향후 연구는 다음과 같은 혼합방법 접근을 고려할 수 있다. 예를 들어, 토픽모델링을 통해 도출된 주제를 대상으로 한 질적 내용 분석을 수행하거나, 추가적으로 의미연결망(semantic network) 분석을 실시하여 핵심 키워드 간 관계를 탐색한 뒤 그 맥락적 의미를 해석하는 방식이 가능할 것이다. 이를 통해 정량적 패턴과 정성적 해석을 결합함으로써 보다 깊이 있는 결혼이주여성 건강 담론 연구가 이루어질 수 있을 것이다. 다섯째, 향후 연구에서는 일반 여성 건강 담론, 다문화 정책 담론 등과의 비교 분석을 수행할 필요가 있다. 이를 통해 본 연구에서 도출된 키워드와 토픽 구조가 결혼이주여성 건강 담론에 상대적으로 특징적인 현상인지, 혹은 보다 넓은 건강·이주 담론에서 공통적으로 나타나는 현상인지를 구분할 수 있을 것이다.

참고문헌

- 감미아·송민(2012). “텍스트 마이닝을 활용한 신문사에 따른 내용 및 논조 차이점 분석”. **지능정보연구**, 제18권 제3호, 53-77. <https://doi.org/10.13088/jiis.2012.18.3.053>
- 강영신·송도안·신예지·이안나(2025). 한국 내 이주민의 건강권 문제: 텍스트마이닝 기법을 활용하여 [글로벌디아스포라연구 총서]. 글로벌디아스포라연구소.
- 강하라(2021). “다문화 가족 어머니의 문화적응 스트레스가 부모 양육 행동을 매개로 초기 청소년 자녀의 학교생활적응에 미치는 종단적 영향”. **현대사회와 다문화**, 제11권 제2호, 127-162. <https://doi.org/10.35281/hmc.2021.11.2.127>
- 권태연(2020). “부모의 우울이 초등학교 자녀의 언어발달, 문제행동 및 학교적응에 미치는 종단적 영향”. **한국산학기술학회논문지**, 제21권 제1호, 338-348. <https://doi.org/10.5762/KAIS.2020.21.1.338>
- 김경숙(2014). “한국의 여성 결혼이주자정책: 상호문화주의적 조망과 함의”. **디지털융복합연구**, 제12권 제9호, 21-33. <https://doi.org/10.14400/JDC.2014.12.9.21>
- 김광기·제갈정·최명일(2021). “음주 관련 신문기사 제목에 나타난 의미 구성과 공중보건학적 함의: 의미연결망 분석의 적용”. **보건교육건강증진학회지**, 제38권 제3호, 53-63.
- 김연정·김현실(2025). “사회적 고립청년의 정신건강 관련 언론보도 담론 분석: 키워드 네트워크 분석”. **한국사회과학연구**, 제44권 제3호, 217-251.
- 김이선(2010). **2010년 한국이주동향 - 결혼이주자**. IOM이민정책연구원.
- 김주영·윤소원(2025). “다문화 이주민의 여가 관련 뉴스 담론 분석”. **한국여가레크리에이션학회지**, 제49권 제2호, 13-25.
- 김현숙·김옥녀(2017). “한국인의 배타적 태도에 대한 이주민 반응에 관한 연구: 결혼이주여성을 중심으로”. **다문화사회연구**, 제10권 제2호, 145-191. <https://doi.org/10.35281/hmc.2017.10.2.145>
- 류한수(2016). “결혼이주여성의 차별경험이 신체적, 정신적 건강에 미치는 영향”. **한국콘텐츠학회논문지**, 제16권 제8호, 345-356. <https://doi.org/10.5392/JKCA.2016.16.08.345>
- 박준형·오효정(2017). “국내 기록관리학 연구동향 분석을 위한 토픽모델링 기법 비교: LDA와 HDP를 중심으로”. **한국도서관·정보학회지**, 제48권 제4호,

- 235-258. <https://doi.org/10.16981/kliss.48.4.201712.235>
- 박지경(2014). “결혼이민자·귀화자의 의료기관 이용 실태에 관한 연구”. **디지털융복합연구**, 제12권 제1호, 491-500. <https://doi.org/10.14400/JDPM.2014.12.1.491>
- 박현선(2014). “결혼이주여성의 문화적응스트레스 및 양육스트레스가 자녀의 내재화문제에 미치는 중단 매개효과분석: 우울의 매개효과를 중심으로”. **학교사회복지**, 제27권, 353-376. <https://doi.org/10.22271/ssw.2014.27.353>
- 법무부(2025). **2024 출입국·외국인 정책 통계연보**. 법무부 출입국·외국인정책본부.
- 보금로·장영은(2022). “결혼이주여성의 문화적응 스트레스가 다문화 청소년 자녀의 우울에 미치는 영향”. **학습자중심교과교육연구**, 제22권 제19호, 837-848. <https://doi.org/10.22251/jlcci.2022.22.19.837>
- 아이지에이웍스(2025. 1.). 2025년 1월 모바일 앱 랜드마크 리포트. <https://www.mobileindex.com/>
- 안순태·이하나(2016). “자살예방을 위한 미디어 보도 방향: 건강신념모델을 통한 우울증 보도 내용분석”. **보건사회연구**, 제36권 제1호, 529-564.
- 양혜승(2022). “여성, 이주민, 노인 혐오성 댓글에 대한 텍스트 마이닝 분석: 네이버 범죄뉴스를 맥락으로”. **한국방송학보**, 제36권 제3호, 5-45.
- 여성가족부(2022). 2021년 전국다문화가족실태조사. 여성가족부. https://www.mogef.go.kr/mp/pcd/mp_pcd_s001d.do?bbtSn=704929&mid=plc503
- 여성가족부(2025). 2024년 전국다문화가족실태조사 연구. 여성가족부. https://www.mogef.go.kr/mp/pcd/mp_pcd_s001d.do?mid=plc503&bbtSn=705039
- 연합뉴스(2026. 1. 4). “네이버 검색 점유율 63%…3년 만에 60% 재돌파”. <https://www.yna.co.kr/view/AKR20260102129700017>
- 이봉숙(2010). “다문화가정 결혼이주여성의 건강 관련 경험”. **성인간호학회지**, 제22권 제5호, 477-487. <https://doi.org/10.7475/kjan.2010.22.5.477>
- 이안나·강영신(2023). “여성결혼이민자의 지각된 차별감이 우울에 미치는 영향: 사회적 지지의 조절효과”. **한국심리학회지: 여성**, 제28권 제1호, 71-91. <https://doi.org/10.18205/kpa.2023.28.1.004>
- 이형하(2021). “부모의 스트레스와 다문화청소년의 심리적 부적응, 학업중단다위험 간 중단적 인과관계: 자기회귀교차지연모형 적용”. **디지털융복합연구**,

- 제19권 제6호, 373-385. <https://doi.org/10.14400/JDC.2021.19.6.373>
- 임혁(2010). “여성결혼이민자의 정신건강 영향요인에 관한 연구”. **정신건강과 사회복지**, 제34권, 189-221. <https://doi.org/10.24301/MHSW.2010.04.34.189>
- 정기선·한지은(2009). “국제결혼이민자의 적응과 정신건강”. **한국인구학**, 제32권 제2호, 87-114. <https://doi.org/10.22741/kapa.2009.32.2.87>
- 정의철(2008). “에이즈 뉴스 프레임 분석: 비판적 헬스저널리즘 관점을 중심으로: 비판적 헬스저널리즘 관점을 중심으로”. **한국언론학보**, 제52권 제4호, 223-249.
- 정현주(2014). 다문화가족정책 기본계획. 행정안전부 국가기록원. <https://www.archives.go.kr/next/newsearch/listSubjectDescription.do?id=009210&sitePage>
- 조선주·오현경·민현주(2017). “결혼이민자의 사회적 차별경험이 한국생활만족도에 미치는 영향”. **여성연구**, 제92권 제1호, 109-140. <https://doi.org/10.33949/tws.2017..94.004>
- 최나연·이병숙(2018). “여성 결혼이민자의 건강불평등 경험”. **인문사회** 21, 제9권 제2호, 425-438. <https://doi.org/10.22143/HSS21.9.2.35>
- 최영·김병철(2000). “인터넷 신문의 상호작용성에 관한 연구: 국내 인터넷 신문의 상호작용 매커니즘에 대한 실증 분석”. **한국언론학보**, 제44권 제4호, 172-200. <https://doi.org/10.15833/KAME.2000.44.4.172>
- 최윤정·성미혜(2017). “결혼이주여성의 지각된 사회적 지지, 대처전략, 문화적응 태도가 사회통합에 미치는 영향”. **디지털융복합연구**, 제15권 제10호, 513-525. <https://doi.org/10.14400/JDC.2017.15.10.513>
- 최지희·박은자·전진아·이홍림·곽윤경·천희란·이나경(2025). “결혼이주여성의 건강 연구에 대한 주제범위 문헌고찰”. **보건사회연구**, 제45권 제3호, 357-382. <https://doi.org/10.15709/hswr.2025.45.3.357>
- 통계청(2024. 3. 19.). 2023년 혼인통계 [보도자료]. 통계청. https://mods.go.kr/board.es?act=view&bid=204&list_no=430883&mid=a20108110000
- 한수연·설염주·성춘향·하태희(2017). “결혼이주여성의 사회적지지, 문화적응스트레스, 자녀의 학교생활적응간의 관계”. **교육방법연구**, 제29권 제1호, 1-23. <https://doi.org/10.17927/tkjems.2017.29.1.1>
- 현경자·김연수(2012). “여성결혼이민자의 문화적응 스트레스와 지각된 사회적 지

- 지가 우울에 미치는 효과: 사회적 지지의 직접 및 조절 효과”. **한국심리학회지: 여성**, 제17권 제4호, 653-677. <https://doi.org/10.18205/kpa.2012.17.4.004>
- 행정안전부(2025). 2024 지방자치단체 외국인주민 현황(2024. 11. 1 기준). 행정안전부 사회통합지원과.
- AP신문(2025. 11. 18.). “2025년 네이버 뉴스 제휴 매체 리스트”. AP신문. <https://www.apnews.kr/news/articleView.html?idxno=3042387>
- Aggarwal, C. C., & Zhai, C. X.(2012). An introduction to text mining. In C. C. Aggarwal & C. X. Zhai (Eds.) *Mining text data* (pp. 1-10). Springer.
- Berry, J. W. & Sam, D. L.(1997). Acculturation and adaptation. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 3, pp. 291-326). Allyn & Bacon.
- Castañeda, H., Holmes, S. M., Madrigal, D. S., Young, M. E. D., Beyeler, N., & Quesada, J.(2015). Immigration as a social determinant of health. *Annual Review of Public Health*, 36, 375-392. <https://doi.org/10.1146/annurev-publhealth-032013-182419>
- Choi, G. Y., & Byoun, S. J.(2014). Domestic violence against migrant women in South Korea: Addressing the needs of a uniquely situated victim population in domestic violence policy. *International Social Work*, 57(6), 645-660. <https://doi.org/10.1177/0020872812448492>
- Conrad, P., & Barker, K. K.(2010). The social construction of illness: Key insights and policy implications. *Journal of Health and Social Behavior*, 51(Suppl), S67-S79. <https://doi.org/10.1177/0022146510383495>
- Crenshaw, K. W.(1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), 139-167.
- Derosé, K. P., Escarce, J. J., & Lurie, N.(2007). Immigrants and health care: Sources of vulnerability. *Health Affairs*, 26(5), 1258-1268.

- <https://doi.org/10.1377/hlthaff.26.5.1258>
- Dhenia, R. N. K.(2022). Text mining and social media analysis for mental health insights. *World Journal of Advanced Research and Reviews*, 15(3), 640-645.
<https://doi.org/10.30574/wjarr.2022.15.3.0892>
- DiMaggio, P., Nag, M., & Blei, D.(2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6), 570-606.
<https://doi.org/10.1016/j.poetic.2013.08.004>
- Feldman, R., & Sanger, J.(2007). *The text mining handbook: Advanced approaches to analyzing unstructured data*. Cambridge University Press.
- Gan, J., & Qi, Y.(2021). Selection of the optimal number of topics for LDA topic model-Taking patent policy analysis as an example. *Entropy*, 23(10), 1301. <https://doi.org/10.3390/e23101301>
- Jang, I. A., Kim, M. Y., Lee, S. R., Jeong, K. A., & Chung, H. W.(2013). Factors related to dysmenorrhea among Vietnamese and Vietnamese marriage immigrant women in South Korea. *Obstetrics & Gynecology Science*, 56(4), 242-248.
<https://doi.org/10.5468/ogs.2013.56.4.242>
- K-ICT Big Data Center[K-ICT 빅데이터센터](2017). *Korean Morphological Dictionary* [형태소사전].
<https://kbig.kr/portal/kbig/datacube/niadict.page>
- Kim, B., & Son, K.-B.(2020). Factors affecting the self-rated health of immigrant women married to native men and raising children in South Korea. *BMC Public Health*, 20, 1496.
<https://doi.org/10.1186/s12889-020-09605-z>
- Kim, G., & Kilkey, M.(2018). Marriage migration policy in South Korea: Social investment beyond the nation state. *International Migration*, 56(1), 23-38. <https://doi.org/10.1111/imig.12350>
- Kim, H. S.(2011). Marriage migration, gender, and family in South Korea. *The Journal of Korean Studies*, 16(2), 155-180.
<https://doi.org/10.1353/jks.2011.0017>

- Kim, Y. S., Kim, H. S., & Shin, Y. J.(2010). Marriage migration and health inequalities among migrant women in South Korea. *Asian Population Studies*, 6(2), 165-181.
<https://doi.org/10.1080/17441730.2010.494451>
- Marmot, M., Friel, S., Bell, R., Houweling, T., & Taylor, S.(2008). Closing the gap in a generation: Health equity through action on the social determinants of health. *The Lancet*, 372(9650), 1661-1669. [https://doi.org/10.1016/S0140-6736\(08\)61690-6](https://doi.org/10.1016/S0140-6736(08)61690-6)
- Niederdeppe, J., Gollust, S. E., & Barry, C. L.(2014). Media framing of health issues. *Annual Review of Public Health*, 35, 317-333.
<https://doi.org/10.1146/annurev-publhealth-032013-182503>
- 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.
<https://doi.org/10.5120/ijca2018917395>
- Said, S., Cvetanovska, N., Whicker, S. D., Stockman, K., Nguyen, D., Campbell, D., ... & Jessup, R. L.(2025). Attributes, skills and resources required for peer health navigator roles: A qualitative study of the perspective of patients, healthcare professionals and health navigators. *Patient Education and Counseling*, 138, 109201. <https://doi.org/10.1016/j.pec.2025.109201>
- Schuster, R. C., Wachter, K., McRae, K., McDaniel, A., Davis, O. I., Nizigiyimana, J., & Johnson-Agbakwu, C. E.(2024). "I promised them I would be there": A qualitative study of the changing roles of cultural health navigators who serve refugees during the COVID-19 pandemic. *Social Sciences & Humanities Open*, 10, 101002. <https://doi.org/10.1016/j.ssaho.2024.101002>
- Seale, C.(2003). Health and media: An overview. *Sociology of Health & Illness*, 25(6), 513-531.
<https://doi.org/10.1111/1467-9566.t01-1-00356>
- Sievert, C., & Shirley, K.(2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70). Association for Computational Linguistics.

<https://doi.org/10.3115/v1/W14-3110>

World Health Organization.(2008). *Closing the gap in a generation: Health equity through action on the social determinants of health*. <https://iris.who.int/handle/10665/43943>

Yi, J., & Lee, I.(2018). Factors affecting unmet healthcare needs of working married immigrant women in South Korea. *Journal of Korean Academy of Community Health Nursing*, 29(1), 41-53. <https://doi.org/10.12799/jkachn.2018.29.1.41>

Zolnoori, M., Huang, M., Patten, C. A., Balls-Berry, J. E., Goudarzvand, S., Brockman, T. A., ... & Yao, L.(2019). Mining news media for understanding public health concerns. *Journal of Clinical and Translational Science*, 5(1), e1. <https://doi.org/10.1017/cts.2019.434>

Abstract

A Text Mining Analysis of Health Discourse on Marriage Migrant Women: Focusing on Media Coverage

Young-Shin Kang*·Do Eon Song**

Health issues affecting marriage migrant women extend beyond individual conditions and are shaped within broader social discourses, with media serving as a key arena for interpreting and representing the health issues of specific social groups. Despite this significance, systematic analyses of media coverage on the health of marriage migrant women remain limited. This study explores how health-related discourses concerning marriage migrant women are represented and constructed within Korean media. News articles containing keywords related to “marriage migrant women” and “health” were retrieved from major nationwide media, and analyzed through text mining, including frequency analysis, TF-IDF weighting, and topic modeling. Frequency analysis showed that keywords such as multicultural families, support, family, programs, and education appeared most often, while TF-IDF analysis highlighted multicultural families, preventive health examinations, free of charge, public health centers, and programs as salient terms. Topic modeling identified four major themes: support for multicultural families, medical support, support for pregnant women, and domestic violence victimization, indicating that media portrayals were largely framed around support and protection as well as risk and vulnerability. By examining media discourse, this study contributes to understanding the social construction of health issues affecting marriage migrant women, with discussion on its implications, limitations, and future research directions.

Keywords: marriage migrant women, health issues, media discourse, text mining, topic modeling

* First Author: Professor, Department of Psychology, Chonnam National University

** Corresponding Author: Doctoral Candidate, Department of Psychology, Chonnam National University