

소셜 빅데이터에 나타난 국제결혼 및 다문화: 토픽모델 분석

Topic Modeling of Social Big Data Related to International Marriage and Multiculturalism in Korea

장유민¹(Yumin Jang) <https://orcid.org/0000-0002-0805-7015> 이재림^{2*}(Jaerim Lee) <https://orcid.org/0000-0002-3328-9117>
그레이스 정³(Grace H. Chung) <https://orcid.org/0000-0002-0091-1854>

¹Dept. of Child Development and Family Studies, Seoul National University, Graduate student

²Dept. of Child Development and Family Studies and the Research Institute of Human Ecology,
Seoul National University, Associate Professor

³Dept. of Child Development and Family Studies and the Research Institute of Human Ecology,
Seoul National University, Professor

<Abstract>

The purpose of this study is to identify key terms and latent topics in social media posts, which mentioned “*kuk-jeh-gyul-hon*” (international marriage) or “*da-moon-hwa*” (multicultural). Our data were 7,570 web-crawled posts collected from the multiple platforms of online communities, forums, and blogs in Korea. Using topic modeling with latent Dirichlet allocation, we identified and named 24 latent topics. We then developed a conceptual diagram to better understand the structure of identified topics. Our diagram displays nine themes under three domains. The first domain included three themes, which were getting married, marital life, and divorce. The second domain, micro-level factors, highlighted family members, gender and sexuality, and economic factors. Macro-level factors, the third domain, showed that the state and society, public policies, and social norms were important in online posts related to international marriage and multiculturalism. Our results show that negative perceptions towards international marriage and multiculturalism were prevalent online, and these negative attitudes were related to the stereotypical image of international marriages involving bride purchasing. We also found that the scope of identified topics in this study covered only limited aspects of intimate partnerships and marital life. Further, our findings suggest that prejudice against multicultural families may originate from the issues of economy, gender, and sexuality, rather than racism.

▲주제어(Keywords): 이주여성(migrant women), 다문화가족(multicultural families), 온라인 카페(online communities), 블로그(blog), 소셜 빅데이터 분석(social big data analysis)

• 본 연구는 2020년 가정과삶의질학회 추계학술대회에서 포스터발표한 내용을 수정, 보완한 것임.

* Corresponding Author: Jaerim Lee, Department of Child Development and Family Studies, Seoul National University, 1 Gwanak-ro Gwanak-gu Seoul 08826, Rep. of Korea. Tel: +82-2-880-6853, E-mail: jrlee@snu.ac.kr
[Received] September 16, 2021; [Revised] November 2, 2021; [Accepted] December 21, 2021

I. 서론

한국사회는 2000년대부터 본격적으로 이주노동자와 결혼이민자가 증가하면서 다문화사회에 접어들었다. 한국의 다문화사회는 농촌지역 남성의 결혼난 해결을 위한 동남아 여성의 결혼이주와 동남아 출신 저임금 노동자유입을 주축으로 했다는 점에서, 투자 또는 고학력 취업이민, 난민 등을 중심으로 형성된 서구의 다문화사회와는 차이가 있다(이인희, 황경아, 2013; Brat, 2021). 특히 한국사회의 다문화가족 정책 및 다문화 관련 담론은 주로 여성 결혼이민자를 중심으로 형성되었다(김혜순, 2017).

다문화사회에 접어들었지만, 이들 결혼이민자와 다문화가족에 대한 편견과 차별은 여전하다. 2018년 전국다문화가족실태조사 결과, 결혼이민자와 귀화자 중 30.9%가 지난 1년간 한국에서 생활하면서 외국 출신이라는 이유로 차별받은 경험이 있었다(여성가족부, 2019b). 특히 경제적으로 비교열위에 있는 동남아시아 출신 여성, 그리고 이들과 한국 남성의 결혼으로 형성된 다문화가족에 대한 편견이 존재한다(박종도, 2019). 한국사회의 1세대 결혼이민자들에 대한 사회적 편견은 그들의 자녀에게도 재생산되어, 학교 등 공적 공간에서 차별과 배제의 대상이 되기도 한다(박종도, 2019). 이러한 문제의식을 바탕으로 정부와 시민단체 등에서 다문화가족에 대한 편견을 해소하기 위한 다문화인식개선사업을 실시하고 있다(김혜순, 2017).

그러나 2018년 국민다문화수용성조사 결과에 따르면, 일반 국민의 다문화수용성은 100점 만점에 52.81점으로 중간 수준이었으며, 2015년 대비 1.14점 하락하였다(여성가족부, 2019a). 다문화수용성의 구성요소 중 고정관념 및 차별이나 거부·회피정서 관련 수용성은 다른 구성요소에 비하여 점수가 높고, 매년 상승하는 경향을 보인다. 마찬가지로 일반적이고 보편적인 사회적 바람직성을 측정하는 세계시민행동지수는 증가하는 경향이다. 반면, 교류행동지수는 2012년에 비해 2015년 약간 상승했다가 2018년에는 다시 2012년보다도 낮아졌고, 이주민에 대한 일방적 동화기대는 날로 증가하고 있다. 즉, 일반적 인식 측면에서의 다문화수용성은 지속적으로 상승하고 있으나, 이주민과 직접 교류를 하는 등 실질적 통합 측면에서는 수용성이 오히려 저하되는 경향을 보인다(여성가족부, 2019a). 이러한 다문화수용성의 일반적 인식 측면과 실질적 통합 측면의 괴리는 사회적 바람직성의 영향을 받는 설문조사의 결과와 일반 국민의 내밀한 실제 인식에는 차이가 존재

할 수 있음을 시사한다. 따라서 국제결혼이나 다문화와 관련된 일반인들의 인식을 탐색하기 위해서는, 온라인 공간과 같이 익명성이 보장되어서 사회적 바람직성의 영향을 적게 받는 곳에서 표출된 의견을 살펴볼 필요가 있다.

소셜 빅데이터는 국제결혼 및 다문화와 관련된 한국사회의 실제 담론과 대중 인식을 탐색하기에 유용한 자료이다. 소셜 빅데이터란 일반 대중이 온라인 카페, 블로그, 소셜 네트워크 서비스(SNS) 등의 온라인 채널에서 생성한 대규모 데이터로 개인의 경험과 생각, 감정 등을 담고 있는 자료를 의미한다(박선영, 이재림, 2021; Ishikawa, 2015). 소셜 빅데이터는 대용량의 데이터로, 제한된 표본을 대상으로 하는 기존 연구방식의 한계를 극복할 수 있다(최은수, 최인철, 최종안, 2019). 또한, 설문조사 자료가 미리 구성된 변수와 문항이 있는 고정된 양식을 통해 얻은 자료인 것(조은별, 민지연, 박수원, 2020)과 달리, 소셜 빅데이터는 개인이 온라인에서 자유롭게 솔직한 생각과 감성을 표현함에 따라 일상적인 맥락에서 형성된 자연언어 텍스트이기 때문에 진실성과 진정성이 확보된다(송길영, 2012; Kitchin & McArdle, 2016). 따라서 사회적 바람직성의 영향을 받을 수 있는 설문조사 자료보다 실제에 가깝게 사회적 현상과 대중 인식을 파악할 수 있으며, 연구자가 미처 생각하지 못한 담론의 영역을 확인할 수 있다(안명숙, 2018).

따라서 소셜 빅데이터는 주로 여론, 동향, 담론(이병호, 김가연, 2019; 오승욱, 한진욱, 김민수, 2020), 대중적 인식, 이슈, 트렌드 연구에 사용된다. 소셜 빅데이터 수집 채널 중에서도 온라인 카페, 블로그, 게시판의 자료는 대중 인식과 담론 등을 파악하기에 적절한 자료이다. 온라인 카페는 유사성을 공유한 회원들 간 자유로운 논의 및 의견교환, 집단적 지식생산과 공유, 관계 형성의 장으로(박선영, 이재림, 2021), 한국에서 온라인 문화의 중요한 비중을 차지하고 있다. 블로그는 사용자 개인 또는 집단의 관심사에 따라 일기, 칼럼, 기사 등 자유로운 형태의 기록이 가능한 홈페이지이다. 온라인 게시판 중 네이버 지식인과 다음 TIP은, 실제적인 정보를 구하기 위해 질문과 답변을 올리는 플랫폼으로 대중의 관심과 인식을 반영한다(김주아, 이경숙, 이재원, 2019). 한편, 온라인 게시판 중 디시인사이드 갤러리, 보배드림 등은 수평성과 익명성을 특징으로 한다(이현아, 류석진, 2013). 이와 같은 온라인 커뮤니티는 온라인 콘텐츠의 생산·유통·확산의 핵심 공간으로서, 새로운 대안 미디어와 공론의 장으로 기능하고 있다.

본 연구의 목적은 ‘국제결혼’ 또는 ‘다문화’를 언급한 온

라인 카페, 블로그, 게시판의 글을 분석하여 소셜 빅데이터에 나타난 국제결혼과 다문화 관련 핵심단어 및 토픽을 분석하는 것이다. ‘국제결혼’은 국제결혼, 결혼이주, 결혼 이민 등을 포괄하는 주제어이며, ‘다문화’는 다문화사회, 다문화가족, 다문화가정, 다문화가정자녀 등을 포괄한다. 본 연구는 선행연구가 주로 다문화가족 당사자(Chung & Yoo, 2013), 서비스 기관 및 인력(박홍식, 2013; 소라미, 2009), 전국민 대상(여성가족부, 2019a) 조사를 통해 인식을 살펴본 것과는 달리 소셜 빅데이터를 활용한 연구라는 차별성과 의의를 갖는다. 본 연구에서 사용할 토픽모델 분석 방법은 통계추론모형을 기반으로 토픽을 도출하고 명명하여 텍스트를 객관적이고 명확하게 분석할 수 있는 빅데이터 분석기법 중 하나이다. 본 연구의 결과를 토대로 다문화화가 심화되고 있지만 여전히 차별과 인식 개선이 주요 의제인 한국 사회에서, 국제결혼과 다문화에 대한 대중의 실제 인식을 파악하고 실천적 제언을 도출할 수 있을 것으로 기대된다. 본 연구의 연구문제는 다음과 같다.

- 연구문제 1. 국제결혼 및 다문화 관련 온라인 카페, 블로그, 게시판 게시글에 자주 등장하는 단어는 무엇인가?
- 연구문제 2. 국제결혼 및 다문화 관련 온라인 카페, 블로그, 게시판 게시글의 주요 토픽은 무엇인가?

II. 선행연구 고찰

1. 국제결혼과 다문화 연구동향

국제결혼과 다문화 관련 국내 연구는 2005년을 기점으로 급격하게 증가했으며(김미희, 2007), 주로 국제결혼이주여성을 중심으로 이루어졌다. 이는 취업이민 및 난민을 중심의 서구의 다문화 연구와는 차이가 있다(이인희, 황경아, 2013; Brat, 2021). 결혼이민은 한국과 대만에서만 별도의 지원정책대상이며, 서구에서는 국민의 배우자라는 입국자격으로만 존재한다. 즉, 국제결혼이주여성은 한국 등 동아시아의 특수성을 반영한 정책집단이다(김혜순, 2014; 설동훈, 강기정, 이병하, 2011). 따라서 국외에서도 한국과 일본, 대만을 중심으로 한 아시아의 다문화와 국제결혼 및 결혼이주의 특수성에 관한 연구가 진행되었다(Dickey, 2014; Jones, 2012; Jones & Shen, 2008; Toyota, 2008).

국내에서 국제결혼과 다문화와 관련된 대표적인 연구주

제는 국제결혼이주여성의 정체성(강미옥, 이정애, 최은경, 2019; 박신규, 2008), 인권(이하윤, 2005; 조홍석, 2020), 언어(김선숙 외, 2020), 문화차이와 문화적응(김정현, 2015; 엄동문, 2019), 심리·사회적 적응(박동숙 임혜영, 김은경, 2019; Chung & Lim, 2016), 사회적 인식 및 차별(류한수, 2016), 사회지지망(김경미, 2012; 김혜선, 2019) 등이다. 국제결혼 자체에 관한 선행연구의 주제는 국제결혼중개 관련 법과 권리 등이 있다(조홍석, 2020; 최현태, 2011). 국제결혼을 통해 형성된 다문화가족에 대한 연구는 주로 부부 갈등 등 결혼생활의 어려움(김오남, 2006), 가정폭력(류은영, 조숙정, 2021), 결혼이주여성의 배우자(정선아, 최수찬, 2017)와 다문화가족 자녀의 심리·사회적 적응(이지연, 그레이스 정, 2019b; 허청아, 그레이스 정, 2018), 자녀교육(김승희, 2011; 배성찬, 편성진, 2013)을 다루었다. 최근에는 다문화가정 자녀의 이중민족사회화에 대한 논의도 진행되고 있다(이지연, 그레이스 정, 2019a; Chung, Lee, & Gonzales-Backen, 2020).

다문화사회 형성 과정과 맥락이 유사하면서도, 한국보다 앞서 다문화 현상을 경험한 일본과 대만 연구는 다음과 같다. 일본은 국적이나 민족이 다른 사람들이 대등한 관계를 바탕으로 서로 문화적인 차이를 인정하면서 지역 사회 구성원으로서 함께 살아가는 다문화공생주의를 기조로 한 정책을 실시하고 있으며(김경희, 2021; 천호성, 이정희, 2014), 보다 통합적인 차원에서 다문화사회와 다문화교육(Tsuneyoshi, Okano, & Boocock, 2010), 다문화가정의 자녀양육(Koide et al., 2019; Kuramoto, Koide, Yoshida, & Ogawa, 2017)이 연구되었다. 대만의 다문화 시민교육 연구(정용교, 2016), 한국과 대만의 다문화정책 비교(이종영, 정준호, 2012), 다문화수용성 비교(윤인진, 팽정, 2021) 연구도 이루어졌다. 하지만 일본은 지역사회 중심의 다문화 정책이 개선되고 있으며, 대만은 원래 다종족사회였기에(정용교, 2016) 한국과 같은 맥락으로 보기에는 어려움이 있다.

서구에서는 1900년대 우편주문신부(mail-order bride)가 아시아의 국제결혼이주여성과 비슷한 맥락에서 연구되었다(Kelly, 2001; Kusel, 2014). 국제결혼이주여성의 권리와 젠더에 관한 논의가 서구에서 직접적으로 시작되기도 했다(Piper, 1997). 하지만 최근에는 다문화주의와 난민, 인종과 종교의 다양성 정책 위주로 연구되고 있다(Brat, 2021; Igarashi, 2019). 이에, 세부 맥락의 차이는 있지만, 한국의 다문화수용성 논의의 기반으로 서구에서 개발된 다문화수용성 개념을 활용하고 있다(Bennett, 1986;

Berry & Kalin, 1995).

한편, 우리나라 정부에서는 국제결혼 및 다문화와 관련된 실태조사와 정책연구를 수행하고 있다. 3년마다 실시하는 여성가족부의 다문화가족실태조사가 대표적이다. 관련하여 사회통합지표(문유경, 전기택, 2011; 설동훈, 윤홍식, 2008), 한국형 다문화수용성 진단도구(여성가족부, 2012) 등도 개발되었다. 일반인들의 다문화 수용성에 대한 연구도 활발하게 이루어지고 있으며(김갑순, 윤상기, 2019; 문성은, 방은령, 2014), 다문화 담론과 미디어 보도프레임에 관한 연구(이인희, 황경아, 2013; 오은주, 고진호, 2021)도 진행되고 있다. 다문화가족을 위한 가족 정책 전달체계인 다문화가족지원센터에 관한 연구(박윤정, 2016), 다문화 부모교육 프로그램에 관한 연구(조은영, 이신동, 2020)도 존재한다.

이상과 같은 국내 선행연구는 결혼이주여성 등 다문화가족 당사자의 경험과 어려움을 파악했다는 의의가 있다. 국제결혼과 다문화에 대한 대중의 인식이나 태도를 살펴본 계량적 연구도 상당수 이루어졌고, 관련 지표를 개발하려는 노력도 성과를 거두었다. 그러나 연구결과가 다소 피상적이며 사회적 바람직성의 영향으로 솔직한 응답을 확보했는지에 대한 의문이 남는다. 사회조사 연구의 특성상 연구자가 선택한 변수나 문항 이외의 내용을 고려하기 어렵다는 제한점도 있다. 한편, 국외의 상황에 관심을 둔 연구들은 한국사회의 맥락을 온전히 반영하지 못한다. 이러한 연구동향을 토대로 보았을 때, 한국사회의 구성원들이 온라인에서 자연스럽게 생성한 텍스트 형태의 소셜 빅데이터에 등장하는 핵심단어와 토픽을 살펴보는 것은 선행연구의 한계를 극복하는 데 도움이 될 것이다.

2. 국제결혼과 다문화 관련 빅데이터 연구동향

빅데이터 활용 사회과학 연구가 증가하면서 국제결혼과 다문화 분야에서도 학술자료, 소셜 빅데이터, 언론기사를 분석한 연구가 최근에 몇 편 발표되었다. 학술자료에 빅데이터 분석기법을 적용하여 동향을 파악한 국내 연구는 박종도(2019), 강진구와 이기성(2018)의 연구가 있다. 박종도(2019)는 ‘다문화’를 키워드로 한 논문 5,345편의 제목, 키워드, 초록 텍스트에 대해 토픽모델 분석을 실시하여 학교 다문화교육, 결혼이주여성, 문화교육 및 다양성, 문화정체성과 민족주의, 외국인 사회통합정책이라는 토픽을 발견하였다. 강진구와 이기성(2018)은 한국문학 분야 국내학술지 논문 중 다문화와 관련이 있는 307편에 대해 토픽모델

링을 실시하였다. 발견된 토픽은 다문화사회나 현상을 반영한 문학작품에 대한 분석, 다문화교육과 한국어교육, 국내의 이주자들의 정체성이었다.

다문화와 관련하여 소셜 빅데이터를 분석한 연구는 안명숙(2018)의 연구와 학술자료와 인터넷자료를 동시에 분석한 신영진, 이광석(2019)의 연구가 있다. 안명숙(2018)은 블로그, 카페, 페이스북, 트위터, 유튜브에서 ‘결혼이주여성’을 키워드로 자료를 수집해 네트워크 분석을 실시하였다. 그 결과, 결혼이주여성들은 문화의 차이, 이주민이면서 여성으로서 경험하는 이중차별 등의 어려움을 겪고 있음을 파악하였고, 이들의 인권을 위한 지원과 대중매체의 역할을 강조하였다. 신영진과 이광석(2019)은 여성 결혼이민자 지원정책 관련 분야의 연구논문, 연구보고서, 정부 발간백서, 통계 자료에 더해, 트위터 자료를 대상으로 추세 분석과 텍스트마이닝을 실시하였다. 트위터 자료를 분석한 결과, 결혼이주여성이 자녀출산 시 국적을 취득하도록 하는 법안 발의에 반대하는 여론을 통해 결혼이주민의 사회통합에 대한 배타적 견해를 확인하였다.

언론기사를 분석한 선행연구로는 양정은과 노법래(2020)의 연구가 있다. 이 연구에서는 2000년 이후 다문화 관련 언론기사문 약 8만 5천 건을 수집하여 12개의 토픽을 도출하였다. 분석 결과, 다문화주의 정책이 형성될 여건이 갖추어졌으나 제도적 확장이 일어나지 못했고, 여성의 결혼이민을 통해 형성된 다문화가족 이외의 이주민 집단과 관련한 의제는 명맥만 유지하는 것으로 나타났다(양정은, 노법래, 2020).

국외에서는 빅데이터를 활용한 이민 연구의 주제로 다양성, 다문화주의, 건강, 노동 시장, 사회통합 등이 등장했다(Ashton, Bhattacharyya, Galatsanou, Ogoe, & Wilkinson, 2016). 하지만, 여기서 빅데이터는 국가 데이터 등을 주로 다루고 있어 빅데이터의 대용량(volume), 속도(velocity), 다양성(variety)의 주요 특징 중, 대용량이라는 특성만 반영한다는 한계가 있다(Kitchin & McArdle, 2016). 그 외 빅데이터를 통해 이민 동향을 파악한 연구(Sirbu et al., 2021; Wanner, 2021), 트위터의 iMood를 활용하여 미국 이민개혁 정책과 관련된 정서를 알아본 연구(Chung & Zeng, 2013) 등이 있다. 다문화와 관련해서는 밀라노의 다문화화 정도를 파악하기 위해 트위터에서의 다언어 사용을 분석한 연구가 있다(Arnaboldi et al., 2016).

종합하면, 국제결혼 및 다문화 관련 빅데이터 연구는 매우 초보적인 수준이다. 양적으로도 연구가 부족하며, 질적으로도 제한점이 많다. 빅데이터 분석기법을 활용하였지

만, 분석자료는 빅데이터라고 보기 어려운 경우도 적지 않다. 특히 국내와 국외 모두 소셜 빅데이터 활용 사례가 적어서, 국제결혼 및 다문화에 대한 대중의 여과 없는 인식이나 태도에 대해 충분한 지식이 축적되지 않은 상태이다. 방법론적으로도 토픽모델 분석 시 너무 적은 수의 토픽을 추출하여 복잡한 현상을 지나치게 단순하게 설명했다는 아쉬움이 있다. 본 연구에서는 학자나 언론이 아닌 대중의 인식이 반영된 소셜 빅데이터를 통해 국제결혼과 다문화 담론을 살펴봄으로써 한국 사회 다문화 담론에 대한 이해에 기여하고자 한다.

III. 연구방법

1. 데이터 수집

본 연구는 ‘텍스트마이닝을 활용한 결혼 관련 소셜 빅데이터 분석 연구’(연구책임자 이재림)라는 연구과제에서 구축한 빅데이터 중 ‘국제결혼’과 ‘다문화’라는 단어가 포함된 온라인 카페, 블로그, 게시판의 글을 분석대상으로 하였다. 해당 연구과제의 자료는 결혼, 비혼, 이혼, 동거 등 커플 파트너십과 관련된 21개 단어¹⁾를 검색어로 하여, 해당 단어가 제목 또는 원글에 포함된 웹문서를 수집한 것이다. 데이터 수집기간은 2017년 7월부터 2019년 6월까지 3개년이었으며, SK텔레콤 스마트인사이트에 의뢰하여 다양한 온라인 사이트의 원글과 댓글을 웹크롤링하였다. 웹크롤링(web crawling)이란 웹사이트에서 특정 정보를 추출하는 자동화 프로그램의 작업을 의미한다. 자료의 성격이 흐려지지 않도록 광고 기사, 연예인 결혼이나 이혼 보도, 드라마 속 파트너십 등과 관련된 단어는 불용어(不用語)로 처리하였다.

데이터 수집기간은, 크롤링 시점인 2019년 여름을 기준으로 주어진 연구비 내에서 최대로 자료를 수집할 수 있는 기간인 최근 3년으로 설정되었다. 또한 2017년부터 2019년은 국제결혼과 다문화와 관련된 사회문화적 맥락 측면에서도 중요한 시기였다. 2017년은 미국의 트럼프 정부가 출범한 시기이며, 국제적으로 이민자 관련 폭동 및 테러와 사건, 난민의 배당과 정착 문제 등이 경기침체와 맞물리며 반 이민정서와 자국중심주의가 확장되는 대외적 변화가 있었다(김혜순, 2017; 장주영, 박민정, 김기태, 2019). 국내

에서도 다문화에 대한 회의, 경계, 우려, 역차별, 반다문화 정서가 부상했다(김혜순, 2017). 2018년에는 제주 예멘난민사태가 발생하기도 했다. 국내 정책적으로는 다문화인식개선사업의 재고와 주류화가 강조되었던 시기로, 2017년에 제2차 다문화가족정책(2013-2017)이 완료되고 2018년에 제3차 다문화가족정책(2018-2022)이 시작되었다. 제2차 다문화가족정책에서는 다문화수용성 정책에 중점을 두었고, 제3차 다문화가족정책은 결혼이민자와 귀화자의 초기 적응지원에서 장기 정착화에 따른 정책으로 변화했으며, 결혼이민자의 인권보호와 사회적 다문화수용성 증진을 주요 과제로 하였다. 한편, 데이터 수집 종료 시점인 2019년은 외국인과의 혼인 건수 및 비중이 감소하기 직전, 증가세를 유지하던 시점이다. 2017년에서 2019년은 외국인과의 혼인 건수가 2만 건대를 유지하면서 해마다 증가하였으며, 전체 혼인 중 외국인과의 혼인 비율도 각각 7.9%, 8.8%, 9.9%로 꾸준히 증가하다가, 2020년부터 1만 건대로 크게 감소하였다(통계청, 2021).

본 연구에서는 해당 연구과제의 자료 중 ‘국제결혼’ 또는 ‘다문화’가 언급된 온라인 카페, 블로그, 게시판의 웹문서를 추출하여 분석하였다. 본 연구의 최종 분석자료는 카페 게시글 3,604건, 블로그 게시글 1,856건, 온라인 게시판 게시글 2,110건으로 총 7,570건이었다. 본 연구의 자료는 결혼 등 커플 파트너십 관련 소셜 빅데이터에서 추출했기 때문에, 다문화 전반에 대한 것이라기보다 파트너십의 맥락에서의 국제결혼과 다문화에 대한 것이라고 볼 수 있다. 댓글은 텍스트의 길이가 짧아 토픽모델링에 적합하지 않은 경우가 많아서 제외하고, 원글만 사용하였다. 자료 수집된 사이트의 유형은 온라인 카페(네이버 카페, 다음 카페), 블로그(네이버 블로그, 다음 블로그, 티스토리, 이글루스), 온라인 게시판(네이버 지식인, 다음 TIP, 디시인사이드 갤러리, 세티즌, 네이버톡, youtube, MLB_PARK_BULLPAN, 인벤 게시판, 보배드림, 웃긴 대학)이었다.

2. 데이터 전처리

자연언어, 즉 일상적인 텍스트 형태의 소셜 빅데이터는 비정형상태이기 때문에 계량적으로 분석하기 위해서는 정형데이터로 변환하는 전처리를 거친다. 본 연구에서도 웹크롤링으로 추출된 비정형상태의 텍스트를 정형상태인 문서-단어행렬(document-term matrix)로 변환하여 분석

1) ‘결혼’, ‘혼인’, ‘기혼’, ‘초혼’, ‘재혼’, ‘신혼’, ‘부부 생활’, ‘부부 관계’, ‘혼전 동거’, ‘동거 생활’, ‘동거 관계’, ‘비혼 동거’, ‘겨론’, ‘시집장가’, ‘이혼’, ‘출혼’, ‘비혼’, ‘미혼’, ‘독신’, ‘싱글 생활’, ‘싱글 라이프’

하였다. 행렬의 형태로 자료를 정리할 때는 크게 두 가지 방식을 사용할 수 있다. 하나는 웹문서에 등장한 모든 단어를 활용하는 것이고, 다른 하나는 연구주제에 관한 온톨로지를 먼저 개발하고 온톨로지에 포함된 단어로 행렬을 구성하는 것이다. 본 연구에서는 후자의 방식을 사용하였다.

온톨로지란 추상적 개념이나 현상을 이해하기 위해 관련성 있는 구체적인 단어들의 구조적 관계를 정리한 체계이다. 소셜 빅데이터 수집에 앞서 주제와 관련된 개념과 단어를 구조적으로 정리하는 온톨로지를 개발함으로써 효율적으로 자료를 탐색, 분석, 이해할 수 있다(Studer, Benjamins, & Fensel, 1998). 본 연구에서는 소셜 빅데이터 연구를 목적으로 개발된 커플 파트너십 온톨로지(이진경, 이재림, 박선영, 2020)를 활용하였다. 이 온톨로지는 연애, 결혼, 결혼생활, 비혼과 동거, 이혼과 재혼 등 ‘파트너십 차원’과 일상생활, 미시적 요인, 거시적 요인의 ‘파트너십을 둘러싼 차원’으로 구성되어 있다. 각 분류체계는 학술 용어 등 전문용어와 더불어 신조어, 비속어, 은어와 신생 이슈 관련 단어까지 포함하고 있다. 본 연구에서는 온톨로지에 포함된 약 4천 개의 단어를 행으로 하고 7,570건의 문서를 열로 하는 문서-단어행렬 형태의 자료를 분석하였다. 행렬에서 각 칸의 값은 각 문서에 온톨로지의 단어가 등장한 횟수였다.

다음으로, 문서-단어행렬에서 동음이의어 및 불필요어를 삭제하고 유사어를 통합하였다. 단수-복수형, 축약형, 연음형 등 여러 형태로 표현된 유사어들을 대표적인 한 단어로 통합하였다. 예를 들어 ‘아이’와 ‘아이들’은 ‘아이’로, ‘힘드’, ‘힘들’, ‘힘들다’, ‘힘듦’, ‘힘듬’은 ‘힘드’으로 통합하였다.

3. 데이터 분석

전처리된 소셜 빅데이터를 활용하여 빈도 분석과 토픽 모델 분석을 실시하였다. 통계 프로그램은 R version 3.6.3을 사용하였다. 첫 번째 연구문제를 탐색하기 위하여 단어별로 단순 출현빈도 TF(term-frequency)를 산출하였다. TF는 특정 단어가 분석대상 자료에서 문서 내에 얼마나 빈번하게 등장하는지를 나타내는 값이다. 두 번째 연구문제를 위해서는 토픽모델 분석(topic modeling)을 실시하였다. 토픽모델 분석은 대용량의 문서 데이터에 잠재된 주제(topic)가 무엇인지 추출하는 방법이다. 토픽모델 분석은 빅데이터 분석기법 중 사전정보를 입력하지 않는 비

지도 방법(unsupervised method)으로, 연구자의 주관을 최소화하며 객관적이고 정량적인 통계적 방법으로 해당 문서가 포함하고 있는 토픽을 도출할 수 있다(김형지, 김선우, 김성태, 2018). 따라서 구조화되지 않은 숨겨진 주제를 밝힐 수 있다는 장점이 있다(추주희, 오승원, 2020).

본 연구에서는 토픽모델을 구현하는 여러 알고리즘 중 가장 널리 사용되는 잠재적 디리클레 할당(latent Dirichlet allocation; LDA)을 사용하였다(Blei, 2012). LDA는 각 토픽별 단어 수 분포를 기반으로 주어진 문서에서 기 단어수를 분석하여 해당 문서가 어떤 주제들을 함께 다루고 있는지 예측하는 방식이다. 토픽모델 분석에서는 연구자가 토픽의 수 설정을 변경하면서 최적의 개수를 찾는다. 본 연구에서도 토픽의 수를 점진적으로 늘려가며 분석결과를 심도있게 살펴본 후, 각 토픽이 차별적이고 해석가능한 정보를 제공한 토픽의 수가 24개인 모형을 선정하였다. 분석을 위해 R에서 ‘topicmodels’ 패키지를 활용하였다.

이어서 토픽별로 각 핵심단어가 출현할 확률을 나타내는 베타값을 고려하여 연구자가 토픽명을 부여하고 수차례 수정하여 최종 토픽명을 결정하였다. 이 과정에서 해당 토픽 속 핵심단어들의 의미적 연관성을 살펴보았고 해당 주제 관련 지식, 연구문제, 게시글의 내용 등을 바탕으로 토픽명을 설정하였다. 마지막으로 24개 토픽을 대주제로 묶고 도식화하여 상위 수준의 설명을 시도하였다.

IV. 연구결과

1. 국제결혼과 다문화 관련 게시글의 핵심단어

첫 번째 연구문제인 국제결혼 및 다문화 관련 온라인 카페, 블로그, 게시판 게시글에 자주 등장하는 단어는 무엇인지 살펴보기 위해 단어별 단순 출현빈도인 TF 분석을 실시한 결과는 <표 1>과 같다. 본 연구의 주제어인 ‘국제결혼(27,247회)’은 1위, ‘다문화(813회)’는 50위로 나타났다. 2위는 ‘베트남(14,494회)’이었다. 1위 ‘국제결혼’은 3만 회에 가까운 빈도로 언급되었고 ‘베트남’도 1만 회를 넘어서 3위인 ‘여자(8,395회)’와도 격차가 있었다. 베트남의 경우 26위인 ‘베트남여성(1,288회)’과 합하면 총 15,782회 출현하였다. 이는 데이터 수집기간의 시작점인 2017년부터, 외국출신 아내의 국적 중 ‘베트남(27.7%)’의 비중이 가장 컸던 것(통계청, 2017)과 관련이 있다. 베트남 이외에도 37위에 ‘캄보디아(1,148회)’, 40위에 ‘필리핀(1,116회)’, 51위에 ‘중국(800회)’이 등장하는 등, 동남아시아와

표 1. Term Frequency(TF) 기준 빈출 단어 및 빈도

순위	단어	TF	순위	단어	TF	순위	단어	TF	순위	단어	TF
1	국제결혼	27,247	26	베트남여성	1,288	51	중국	800	76	부부	559
2	베트남	14,494	27	이혼	1,268	52	성격	796	77	필요	555
3	여자	8,395	28	맞선	1,258	53	좋아	785	78	행복한	551
4	결혼	6,269	29	어머니	1,234	54	한국인	783	79	어려운	551
5	한국	4,282	30	몸무게	1,223	55	준비	776	80	변명	530
6	신부	4,110	31	아내	1,216	56	연락	768	81	기업	526
7	아가씨	3,473	32	만남	1,208	57	자녀	726	82	조건	516
8	나이	3,378	33	관심	1,201	58	행복	722	83	힘든	516
9	사진	2,955	34	현재	1,195	59	얼굴	717	84	여행	506
10	남자	2,927	35	남편	1,179	60	가장	708	85	서로	502
11	함께	2,380	36	재혼	1,152	61	결혼식	674	86	외모	496
12	부모	2,177	37	캄보디아	1,148	62	같이	672	87	시간	495
13	전화	2,167	38	배우자	1,139	63	원하	669	88	약속	489
14	상담	1,918	39	아버지	1,130	64	사랑	651	89	노력	487
15	소개	1,758	40	필리핀	1,116	65	신랑	649	90	회사	481
16	학력	1,674	41	문제	1,116	66	우리나라	644	91	카페	481
17	직업	1,628	42	아이	945	67	가족관계	642	92	동의	478
18	혼인신고	1,483	43	외국인	914	68	비자	610	93	이해	477
19	이야기	1,421	44	고졸	888	69	거짓말	595	94	혼자	470
20	초혼	1,411	45	혼인	879	70	인연	582	95	예쁜	467
21	사람	1,411	46	비용	847	71	고맙다	582	96	이유	462
22	없다	1,389	47	고등학교	846	72	회사원	576	97	결혼생활	442
23	가족	1,330	48	인생	820	73	정보	574	98	희망	439
24	카카오톡	1,317	49	가정	814	74	본인	571	99	가사	428
25	좋은	1,311	50	다문화	813	75	서울	563	100	막내	426

아시아 국가가 자주 언급되었다.

한편, 3위에 ‘여자(8,395회)’가 등장한 이후로, 6위 ‘신부(4,110회)’, 7위 ‘아가씨(3,473회)’, 29위 ‘어머니(1,234회)’, 31위 ‘아내(1,216회)’ 등이 등장했는데, 이는 각각 10위 ‘남자(2,927회)’, 65위 ‘신랑(649회)’, 39위 ‘아버지(1,130회)’, 35위 ‘남편(1,179회)’와 대응되어 여성이 남성보다 자주 언급되는 것을 확인할 수 있다. 이는 국제결혼과 다문화 관련 담론이 여성을 중심으로 이루어지고 있음을 의미하며, 국제결혼에서 남편이 외국인인 경우보다 아내가 외국인인 경우가 많은 성차(통계청, 2020)를 반영하는 것으로 보인다.

국제결혼당사자의 혼인상태에 관련된 단어인 7위 ‘아가씨(3,473회)’, 20위 ‘초혼(1,411회)’, 27위 ‘이혼(1,268회)’, 36위 ‘재혼(1,152회)’ 등도 등장했다. 다음으로는 외모와 관련된 단어가 9위 ‘사진(2,955회)’, 30위 ‘몸무게(1,223회)’, 59위 ‘얼굴(717회)’, 86위 ‘외모(496회)’, 95위 ‘예쁜(467회)’ 등으로 자주 언급되었다. 그 외 사회경제적 측면의 결혼조건과

관련된 8위 ‘나이(3,378회)’, 16위 ‘학력(1,674회)’, 17위 ‘직업(1,628회)’, 44위 ‘고졸(888회)’ 등의 단어가 등장했다. 또한, 만남과 소통의 방식, 중개비용 등을 의미하는 13위 ‘전화(2,167회)’, 14위 ‘상담(1,918회)’, 15위 ‘소개(1,758회)’, 28위 ‘맞선(1,258회)’, 32위 ‘만남(1,208회)’, 46위 ‘비용(847회)’, 24위 ‘카톡(1,317회)’ 등도 자주 언급되었다. 마지막으로 25위 ‘좋은(1,311회)’, 53위 ‘좋아(785회)’, 58위 ‘행복(722회)’, 64위 ‘사랑(651회)’, 71위 ‘고맙다(582회)’, 78위 ‘행복한(551회)’, 98위 ‘희망(439회)’ 등 긍정적인 단어가 자주 등장했다.

2. 국제결혼과 다문화 관련 게시글 토픽 분석

두 번째 연구문제인 국제결혼 및 다문화 관련 온라인 카페, 블로그, 게시판 게시글의 주요 토픽이 무엇인지 살펴보기 위해 토픽모델 분석을 실시한 결과는 <표 2>와 같다. 토픽 수를 하나씩 늘려가면서 검토한 결과, 각 토픽의

해석가능성과 의미가 가장 적절한 토픽 수 24개 모형을 최종모형으로 선택하였다. 이후 각 토픽에 속할 확률이 높음을 의미하는 베타값이 큰 단어들을 활용하여 토픽을 명명하였다.

먼저, 토픽 1(T1)은 ‘혼인신고’, ‘비자’ 등 국제결혼을 위해 법적으로 필요한 절차와 관련된 단어의 베타값이 높아서 국제결혼 법적 절차라고 명명하였다. 토픽 2(T2)는 다문화가족에 대한 ‘사회’적 ‘지원’ 및 ‘정책’에 관한 것으로, ‘자녀’, ‘양육’과 ‘교육’, ‘문화’, ‘차별’, ‘이해’ 등 지원이 필요한 다양한 영역이 키워드로 나타났다. 다문화가족에 대한 지원은 국가적 차원에서 ‘인구’ ‘문제’와도 관련이 있었으며, ‘변화’하는 ‘새로운’ 다문화 가족에 대한 이해와 지원의 필요성이 나타났다. 이에 토픽 2는 다문화가족 지원이라고 명명하였다.

토픽 3과 토픽 4는 비슷한 핵심단어들을 포함하지만, 세부적인 내용에는 차이가 있었다. 토픽 3(T3)에서는 ‘나이’, ‘학력’, ‘부모’, ‘직업’, ‘초혼’ 여부 등 다양한 조건이 등장했다. 그 중 ‘나이’의 베타값이 다른 단어에 비해 확연히 컸다. 그 외에도 ‘학력’과 관련된 단어가 ‘고졸’, ‘대졸’로 세분되어 나타났으며, ‘초혼’, ‘재혼’, ‘전남편’ 등 혼인 이력과 관련된 단어가 여러 개 등장하였다. ‘부모’는 ‘형제자매’, ‘부양’과 함께 등장하여 부양할 가족의 유무 등도 중요한 조건임을 유추할 수 있다. 이를 토대로 토픽 3은 배우자의 조건 1(나이, 학력, 혼인이력, 부양가족 등)이라고 명명하였다. 한편, 토픽 4(T4)에서는 토픽 3과 유사하게 조건 관련 단어들이 등장했지만 ‘조건’이라는 단어가 직접적으로 등장한 것이 특징이다. 그중에서도 ‘나이’, ‘젊은’, ‘50대’라는 나이와 관련된 단어, ‘외모’, ‘예쁜’과 같이 외모와 관련된 단어가 중복해서 나타났다. ‘남자’의 베타값이 다른 단어들에 비해 확연히 큰데, 이는 남성이 원하는 상대의 조건 또는 반대로 남성의 조건으로 해석가능하다. 따라서 토픽 4는 배우자의 조건 2(나이와 외모)라고 명명하였다.

토픽 5(T5)에서는 ‘소득’, ‘재산’, ‘대출’, ‘주택’, ‘능력’ 등 경제적 요건과 관련된 단어가 주로 등장하였다. ‘직장’과 ‘지역’도 경제적 요인과 연결해서 설명할 수 있으므로 토픽 5는 경제적 요인이라고 명명하였다. 토픽 6(T6)에서는 외모, 초혼 여부, 학력 등 다양한 조건이 나타났다. 그 중 ‘사진’이 다른 단어에 비해 베타값이 확연히 크고, ‘몸무게’, ‘얼굴’ 등이 상위 단어로 나타나 토픽 6은 외모라고 명명하였다.

토픽 7(T7)에서는 ‘결혼’, ‘결혼식’에 있어 ‘맞선’, ‘비용’, ‘결정’, ‘계약’, ‘성사’, ‘선택’, ‘상담’, ‘성혼’ 등의 단어가 등장

했다. 즉, 연애결혼보다는 결혼중매업체를 통한 결혼에 관한 토픽으로 볼 수 있으므로 토픽 7은 결혼중매업체를 통한 결혼이라고 명명했다. 토픽 8(T8)에서는 베타값이 확연히 큰 ‘국제결혼’을 중심으로 ‘카페’, ‘맛선’, ‘여행’, ‘연락’ 등 국제결혼 당사자들의 만남과 연락과 관련된 단어가 핵심이었다. 이에 토픽 8은 국제결혼 시 연락/만남이라고 명명하였다. 토픽 9(T9)에 등장한 단어들은 전반적으로 인물에 대한 ‘소개’와 관련된 단어로 이해할 수 있으며, 수집된 게시글의 제목 중 “국제결혼 프로젝트. 러시아여성 한국 배우자 찾기”가 있는 것으로 보아 결혼중매업체의 인물소개라고 명명하였다. 토픽 10(T10)에는 ‘한국’, ‘태국’, ‘일본’, ‘미국’ 등 다양한 국가가 등장했으며, ‘한국’과 ‘한국인’, ‘외국’과 ‘외국인’이 대비되어 나타났다. 맥락을 파악하기 위해 수집된 게시글들을 따로 확인한 결과, 한국의 외국인, 다문화 트렌드를 언급하는 게시글들로, 토픽 10은 한국의 외국인이라고 명명하였다.

토픽 11(T11)에서는 ‘상담’, ‘전화’, ‘만남’, ‘기업’, ‘출발’ 등 결혼중매업체와 관련된 단어, ‘직업’, ‘학력’, ‘재혼’, ‘미혼’ 등 조건과 관련된 단어가 등장하였다. 다른 토픽과 다른 점은 ‘캄보디아’라는 국가가 키워드로 등장한 것이다. 캄보디아는 2008년 캄보디아 여성과 한국인 남성의 결혼을 인신매매로 인식해 법적으로 금지한 이력이 있으며, 2011년에는 캄보디아 여성과 결혼하려는 외국인 남성에게 소득 기준과 나이 등의 자격 기준을 둔 점에서 다른 국가와 차별성을 갖는다. 이에 토픽 11은 캄보디아 상대와의 만남이라고 명명하였다.

토픽 12, 13, 14는 베트남여성과 관련된 토픽이다. 먼저, 토픽 12(T12)에서 ‘베트남’의 베타값이 확연히 크며 ‘베트남’과 함께 ‘나이’, ‘사진’, ‘몸무게’, ‘귀엽다’, ‘딸이’, ‘미혼모’, ‘전남편’, ‘날씬’, ‘초혼’ 등이 베타값 상위단어로 나타났다. 이에 토픽 12는 베트남여성의 특징이라고 명명하였다. 토픽 13(T13)은 핵심단어인 ‘신부’, ‘베트남’, ‘관심’의 조합을 토대로 베트남신부에 대한 관심이라고 명명하였다. 그 외 단어들도 베트남 신부 관련 정보나 특성을 나타내는 것으로 보인다. 토픽 14(T14)는 ‘베트남’, ‘국제결혼’, ‘베트남 여성’ 등 핵심단어의 조합으로 베트남 여성과의 국제결혼이라고 명명하였다.

토픽 15(T15)에서는 ‘아가씨’의 베타값이 다른 핵심단어에 비해 확연히 커서, ‘아가씨’, 즉 미혼여성과의 만남에 관한 토픽으로 이해할 수 있다. 이에 토픽 15는 아가씨와의 만남이라고 명명하였다. 토픽 16(T16)은 베타값 상위 단어인 ‘여자’와 ‘한국남성’은 베타값 상위 3번째 단어로 등장

표 2. 토픽별 주제명 및 핵심단어

번호	주제명	상위 20개 단어
T1	국제결혼 법적 절차	한국, 혼인신고, 필리핀, 준비, 비자, 여자친구, 한국인, 와이프, 필요, 사람, 현재, 문제, 상황, 불법, 정부, 나라, 어려운, 도와, 교육, 임신
T2	다문화가족 지원	가족 , 지원, 사회, 자녀, 교육, 다문화, 변화, 새로운, 사업, 문화, 동거, 인구, 문제, 이해, 정책, 가정, 차별, 양육, 생활, 국가
T3	배우자의 조건1 (나이, 학력, 혼인이력, 부양가족 등)	나이 , 학력, 부모, 직업, 상담, 초혼, 고졸, 재혼, 소개, 원하, 카카오톡, 회사원, 전남편, 신랑, 형제자매, 취미, 대출, 부양, 본인, 결혼
T4	배우자의 조건2 (나이와 외모)	남자 , 결혼, 나이, 조건, 외모, 원하, 어려운, 힘든, 예쁜, 좋다, 좋은, 사람, 없다, 젊은, 잘하, 50대, 괜찮, 연애, 자신, 수입
T5	경제적 요인	배우자, 소득, 재산, 같이, 자녀, 필요, 공무원, 도움, 대출, 최대, 현재, 충족, 직장, 외국인, 본인, 능력, 인정, 지역, 주택, 대한민국
T6	외모	사진 , 초혼, 몸무게, 얼굴, 가족관계, 카카오톡, 서울, 약속, 고졸, 직업, 둘째, 부모, 첫째, 막내, 학교, 증졸, 본인, 개인, 생일, 셋째
T7	결혼중매업체를 통한 결혼	결혼, 결혼식, 맞선, 비용, 신랑, 결정, 계약, 같이, 인터넷, 성사, 선택, 상담, 커플, 신혼여행, 동의, 없다, 성혼, 이유, 책임, 시간
T8	국제결혼 시 연락/만남	국제결혼 , 카페, 맞선, 여행, 좋아, 연락, 정보, 유튜브, 이상형, 언니, 필요, 귀엽다, 남동생, 여동생, 예쁜, 취미, 영화, 걱정, 좋아요, 매력
T9	결혼중매업체의 인물소개	어머니, 아버지, 소개, 성격, 희망, 연락, 고등학교, 동의, 현재, 관심, 러시아, 오빠, 취미, 아름다운, 성혼, 남동생, 언니, 종교, 초혼, 좋아
T10	한국의 외국인	한국, 외국인, 중국, 한국인, 다문화, 태국, 일본, 가장, 국가, 미국, 외국, 대한민국, 나라, 출신, 문화, 출신, 남녀, 해외, 경기도, 동남아
T11	캄보디아 상대와의 만남	상담, 캄보디아, 전화, 만남, 직업, 학력, 기업, 재혼, 함께, 가사, 미혼, 회사, 행복, 출발, 가을, 겨울, 평생, 기대, 건강, 최선
T12	베트남여성의 특징	베트남, 함께, 나이, 이야기, 비용, 부모, 사진, 몸무게, 귀엽다, 전화, 중학교, 정보, 딸이, 위로, 상담, 미혼모, 전남편, 낯설, 깔끔, 초혼
T13	베트남신부에 대한 관심	신부, 베트남, 이야기, 관심, 전화, 함께, 사진, 몸무게, 부모, 나이, 학력, 장녀, 막내, 위로, 고졸, 예쁜, 전남편, 딸이, 대학, 무남독녀
T14	베트남여성과의 국제결혼	베트남, 국제결혼, 베트남 여성, 소개, 정보, 노력, 딸아이, 젊은, 대학교, 혼인상태, 장녀, 착각, 똑똑한, 비용, 본인, 못할, 가족, 4년제, 괜찮, 발전
T15	아가씨와의 만남	아가씨 , 인생, 현재, 거짓말, 인연, 고맙다, 만남, 좋은, 변명, 함께, 고등학교, 카카오톡, 발달, 전화, 화사, 베트남여성, 남자, 회사원, 몸매, 성격
T16	맞선 상대여성에 대한 호감	여자 , 한국남성, 맞선, 착한, 인터넷, 본인, 좋고, 동남아, 화장, 호감, 자신감, 과거, 성향, 우리나라, 싫어, 현모양처, 진정, 사진, 얼굴, 가능성
T17	가정/결혼에 대한 긍정감성	좋은, 행복한, 가정, 사랑, 행복, 노력, 공부, 서로, 가장, 아름다운, 최선, 중요, 예쁜, 성공, 착한, 도와, 미소, 경험, 배려, 시간
T18	결혼에 대한 부정적 태도	결혼, 사람, 혼자, 한국남자, 없다, 한국여자, 연애, 피해자, 동남아, 한남, 이유, 제대, 문제, 가난, 집안, 도망, 매매혼, 현실, 농촌, 상대
T19	국제결혼에 대한 인식	국제결혼 , 쉽지, 선입견, 괜찮, 성혼, 대학교, 인테리어, 힐링, 예쁜, 생활, 소통, 직원, 금수저, 광주, 여인, 가깝, 좋네, 커플, 신혼생활, 이모님
T20	국제결혼이 매매혼이라는 인식	국제결혼 , 경매 , 자신감, 사랑해, 적응, 걱정없, 원하, 직원, 외동, 성숙, 추석, 일베충, 박탈, 공부, 외국, 시절, 생존, 선물, 미소, 좋아
T21	결혼생활 문제	아내, 남편, 문제, 우리나라, 결혼생활, 이해, 없다, 사람, 잘못, 힘든, 상황, 이유, 도움, 가장, 쉽게, 어려운, 걱정, 적응, 혼자, 잘하
T22	이혼	이혼, 혼인, 부부, 출산, 자녀, 인정, 사기, 임신, 경력, 소송, 피해, 가출, 부인, 위자료, 민법, 이혼율, 위장결혼, 결혼생활, 양육, 부부관계
T23	자녀	아이, 없다, 엄마, 사람, 아기, 친구, 좋아, 힘든, 아빠, 시간, 바람, 동생, 애기, 혼자, 남자친구, 태어, 안되, 식당, 포기, 걱정
T24	고부관계	부모, 함께, 아들, 며느리, 사주, 시어머니, 생활, 사람, 독신, 재혼, 같이, 사랑, 음식, 시집, 관계, 결혼, 집안, 가정, 가족, 여행

주: 상위 20개 단어는 베타값 순이며, 진하게 표시된 단어는 베타값이 확연히 큰 경우임

하는 ‘맞선’의 당사자로 볼 수 있으며, ‘착한’, ‘좋고’, ‘호감’이라는 단어를 통해 맞선 상대에 대한 호감에 대한 토픽임을 유추할 수 있다. 특히 ‘여자’가 베타값이 가장 높고, ‘현모양처’라는 단어가 등장하여 토픽 16은 맞선 상대 여성에 대한 호감이라고 토픽명을 정하였다.

토픽 17(T17)은 ‘좋은’, ‘행복한’, ‘사랑’, ‘행복’, ‘아름다운’, ‘예쁜’, ‘배려’ 등 가정 및 결혼에 대한 긍정적인 형용사들이 주로 등장하여서 가정/결혼에 대한 긍정감성이라고 명명하였다. 반면, 토픽 18(T18)은 결혼에 대해 전반적으로 부정적인 태도가 드러났다. ‘한국남자’와 함께 ‘한남’이라는 단어가 등장해 한국 남성에 대한 부정적인 인식이 나타나며, 이는 한국사회의 젠더갈등과도 관련이 있다. 결혼과 대비되어 ‘혼자’라는 단어가 상위에 등장하였다. ‘동남아’, ‘농촌’과 관련된 국제결혼에 대해서는 ‘가난’, ‘도망’, ‘매매혼’ 등의 부정적 편견이 나타났다. 이를 종합하여 토픽 18은 결혼에 대한 부정적 태도라고 명명하였다.

토픽 19(T19)는 ‘국제결혼’의 베타값이 다른 단어에 비해 확연히 커, 국제결혼에 관한 토픽이며, ‘쉽지’, ‘선입견’, ‘괜찮’ 등의 단어를 통해 국제결혼에 대한 다양한 인식에 관한 토픽이라고 볼 수 있다. 이에 토픽 19는 국제결혼에 대한 인식이라고 명명하였다. 토픽 20(T20)은 ‘국제결혼’과 ‘경매’의 베타값이 확연히 크게 나타나 국제결혼은 경매, 즉 매매혼이라는 인식을 반영하는 것으로 해석하였다. 따라서 국제결혼이 매매혼이라는 인식으로 토픽명을 정하였다.

결혼생활과 관련해서, 토픽 21(T21)에서는 ‘아내’와 ‘남편’이 핵심단어로 등장하여 부부의 결혼생활에 관한 토픽임을 알 수 있다. 더불어 ‘문제’, ‘잘못’, ‘힘든’, ‘어려운’, ‘걱정’ 등 부정적 단어가 주로 나타나 토픽 21은 부부관계 문제라고 명명하였다. ‘이해’, ‘적응’, ‘쉽게’, ‘잘하’ 등의 단어는 긍정적인 단어이지만, ‘없다’, ‘어려운’ 등과 함께 부정적 맥락에서 사용되었을 가능성이 있다. 토픽 22(T22)는 ‘이혼’, ‘사기’, ‘소송’, ‘피해’, ‘위자료’, ‘민법’, ‘이혼율’, ‘위장결혼’ 등 이혼 및 파경 관련 단어가 주로 등장하여 이혼이라고 명명하였다.

관계적 측면에서, 토픽 23(T23)은 반복해서 등장하는 단어인 ‘아이’, ‘아기’, ‘애기’와 ‘엄마’, ‘아빠’ 같이 부모의 정체성을 나타내는 단어를 토대로 자녀라고 명명하였다. 토픽 24(T24)는 ‘아들’, ‘며느리’, ‘시어머니’, ‘시집’ 등의 단어를 통해 가족관계 중에서도 고부관계에 집중된 토픽임을 확인하고 고부관계라고 명명하였다. 이 토픽에서는 ‘함께’, ‘같이’ ‘생활’하는 경우가 많아 ‘관계’가 중요하며 ‘집안’ 단위의 결합인 결혼에서 ‘사랑’이 중요하고, ‘시집’살이, ‘음

식’과 관련된 이슈가 중요한 것으로 나타났다.

한편, 토픽 중에는 본 연구의 주제어인 ‘다문화’, ‘국제결혼’가 베타값 기준 핵심단어에 포함된 토픽과 포함되지 않은 토픽이 있었다. 주제어가 직접 등장하지 않더라도 ‘동남아’, ‘베트남’, ‘러시아’ 등 국가명, ‘비자’ 등 국제결혼임을 유추할 수 있는 단어들이 등장할 경우 국제결혼과 다문화 맥락에서 토픽명을 설정하고 해석하였다. 그러나 배우자의 조건(T3, T4), 경제적 요인(T5), 외모(T6), 가정/결혼에 대한 긍정감성(T17), 결혼생활 문제(T21), 이혼(T22), 자녀(T23), 고부관계(T24) 토픽은 ‘국제결혼’이나 ‘다문화’가 해당 게시글에 등장하기는 했지만 토픽별 핵심단어로 등장하지 않아 국제결혼이나 다문화에 국한되지 않고 일반적인 파트너십에 대한 토픽으로 이해된다. 즉, 본 연구에서 발견한 24개의 토픽 중 13개는 국제결혼 및 다문화와 관련된 토픽이며 9개는 일반적인 파트너십에 대한 토픽이라고 볼 수 있다.

다음으로, 24개 토픽의 내용과 구조를 체계적으로 이해하기 위해 상위범주로 토픽을 분류하고 범주 간의 관계를 도식화한 결과는 <그림 1>과 같다. 도식화는 본 연구의 소셜 빅데이터 수집의 기반이 된 커플 파트너십 온톨로지(이진경 외, 2020)의 구조인 파트너십의 차원(연애, 결혼, 결혼생활, 비혼, 동거, 이혼, 재혼, 사별)과 파트너십을 둘러싼 차원(일상생활, 미시적 요인, 거시적 요인)의 대부분, 중분류에 본 연구의 24개 토픽을 적용하여 작업하였다. 그 결과, 커플 파트너십 온톨로지의 ‘파트너십의 차원’ 대부분에서 결혼, 결혼생활, 이혼이 본 연구의 토픽에서도 중요하게 나타났다. 한편, 온톨로지에서 파트너십을 둘러싼 대부분으로 중요하게 다른 ‘일상생활’관련 토픽은 본 연구에서 등장하지 않았다. 온톨로지의 ‘미시적 요인’ 중에서는 가족, 젠더·섹슈얼리티, 경제적 요인이, ‘거시적 요인’ 중에서는 국가·사회, 정책, 규범이 본 연구에서 부각되었다.

온톨로지의 중분류별로 살펴보면, ‘결혼’ 분류에서는 배우자 선택, 결혼시장, 국제결혼이 강조되었다. ‘결혼생활’ 분류에서는 갈등만 등장하였고, 결혼생활의 다른 다양한 측면과 관련된 토픽은 나타나지 않았다. ‘가족’ 분류에서는 자녀와 고부관계가, ‘젠더·섹슈얼리티’ 분류에서는 아가씨 또는 베트남, 캄보디아 여성의 특징이나 그들과의 만남에 대한 토픽이 포함되었다. ‘국가·사회’ 분류에서는 베트남, 캄보디아, 한국의 외국인, ‘정책’ 분류에서는 국제결혼 시 법적 절차, 다문화가족 지원이, ‘규범’ 분류에는 결혼/가정에 대한 긍정적, 부정적 태도가 본 연

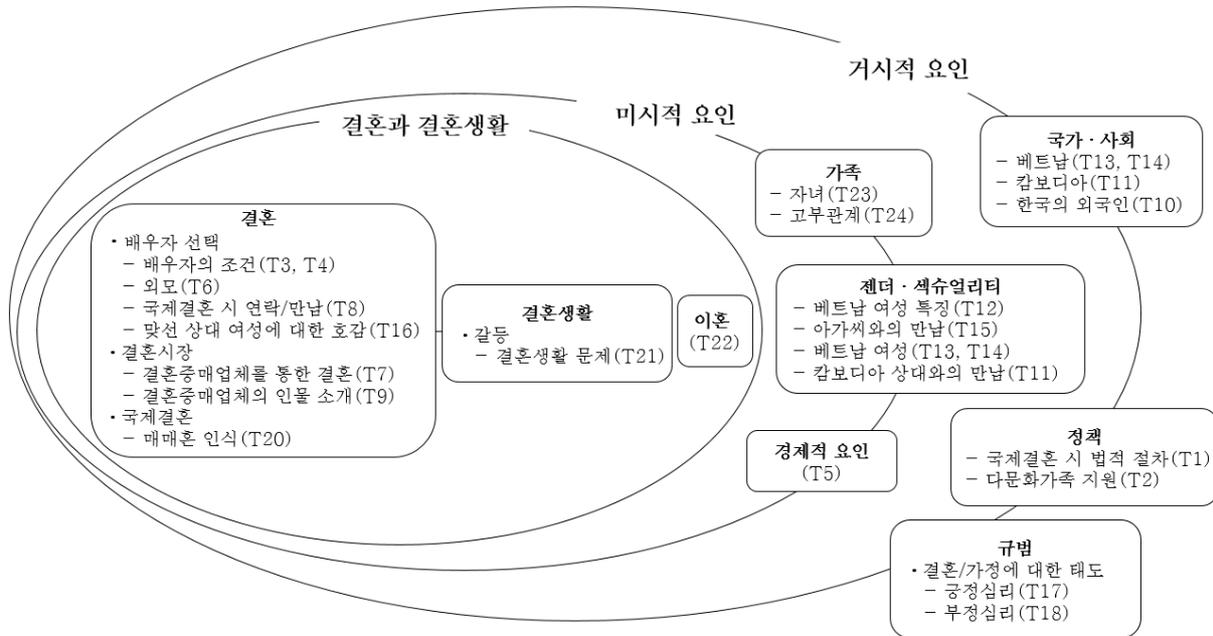


그림 1. 토픽의 의미구조 도식화: 커플 파트너십 온톨로지(이진경 외, 2020) 적용

주: 괄호 안은 토픽 번호임

구의 토픽에서 부각되었다.

V. 결론 및 논의

본 연구에서는 온라인 카페, 블로그, 게시판에 게시된 국제결혼과 다문화 관련 글에 높은 빈도로 등장하는 핵심 단어를 파악하고, 토픽모델 분석을 통해 국제결혼과 다문화 관련 주요 주제를 도출하였다. 이를 위해 국내 대표적인 온라인 포털인 네이버와 다음의 카페 및 블로그, 이외의 다양한 온라인 게시판에 2017년부터 2019년까지 3년간 게시된 글 중 ‘국제결혼’ 또는 ‘다문화’를 포함한 문서 7,570건을 분석하였다. 연구문제별로 연구결과를 요약하고 논의하면 다음과 같다.

첫 번째 연구문제인 국제결혼 및 다문화 관련 온라인 카페, 블로그, 게시판의 게시글에 자주 등장한 단어를 분석한 결과, ‘국제결혼’의 빈도가 가장 높았고, ‘다문화’의 빈도는 상대적으로 낮았다. 이러한 결과는 결혼 등 파트너십 맥락에서의 온라인의 논의가 다문화가 아닌 국제결혼을 중심으로 형성되어 있음을 의미한다. 또한, ‘베트남’이 2위로 나타나 최근 결혼이주여성의 국적 중 베트남의 비중이 가장 큰 현상을 반영하며, 이외에도 캄보디아, 필리핀, 중국 등 국가명이 등장하여 국제결혼이 주로 동남아시아를 비롯한 아시아 출신 여성과의 결혼으로 논의되고 있

음을 알 수 있다. 한편, 국제결혼과 다문화 관련 담론이 남성보다는 여성 위주로 진행되고 있었으며, 국제결혼 당사자의 조건으로 외모, 초혼 여부, 나이, 학력, 직업, 경제력 등이 중요함을 파악할 수 있었다. 또한 만남과 소통의 방식, 중개와 비용에 대한 단어의 언급빈도가 높았다.

두 번째 연구문제인 국제결혼 및 다문화 관련 온라인 카페, 블로그, 게시판의 게시글에 내재된 주요 토픽을 분석한 결과 24개의 토픽이 도출되었다. 각 토픽은 국제결혼 법적 절차(T1), 다문화가족 지원(T2), 배우자의 조건 1(나이, 학력, 혼인이력, 부양가족 등) (T3), 배우자의 조건 2(나이와 외모) (T4), 경제적 요인(T5), 외모(T6), 결혼중매업체를 통한 결혼(T7), 국제결혼 시 연락/만남(T8), 결혼중매업체의 인물 소개(T9), 한국의 외국인(T10), 캄보디아 상대와의 만남(T11), 베트남여성의 특징(T12), 베트남신부에 대한 관심(T13), 베트남여성과의 국제결혼(T14), 아가씨와의 만남(T15), 맞선 상대여성에 대한 호감(T16), 가정/결혼에 대한 긍정감성(T17), 결혼에 대한 부정적 태도(T18), 국제결혼에 대한 인식(T19), 국제결혼이 매매혼이라는 인식(T20), 결혼생활 문제(T21), 이혼(T22), 자녀(T23), 고부관계(T24)라고 명명하였다.

이어서, 24개 토픽을 범주화하고 친밀한 파트너십을 전반적으로 아우르는 커플 파트너십 온톨로지(이진경 외, 2020)의 구조와 비교한 결과 다음과 같은 차이가 있었다.

첫째, 본 연구에서는 파트너십 차원에서 연애에 관한 토픽이 거의 등장하지 않았다. 국제결혼과 다문화가족 형성 시 애정적 측면과 연애과정보다는 배우자 선택조건 등 조건적 만남이 강조되는 경향이 있었다. 결혼생활 관련해서는 일상적인 토픽이 나타나지 않았고, 부부갈등과 이혼과 같은 부정적 측면이 두드러졌다. 또한 일반적으로 최근 온라인 게시물에서 두드러지는 비혼, 동거, 재혼, 사별 같은 다양한 파트너십 형태, 라이프스타일과 여가, 미디어, 소비, 반려동물, 가사노동, 종교 등 다양한 일상생활 영역에 대한 토픽이 본 연구에서는 등장하지 않은 점도 주목할 만하다. 이러한 연구결과는 온라인 채널에서 국제결혼과 다문화에 대한 논의가 조건을 중심으로 한 배우자선택이나 결혼생활의 부정적 측면에 집중되어 있고, 연애나 결혼생활의 다양한 측면, 일상생활 영역에 대한 일반적이고 다차원적인 논의는 활발하지 않음을 의미한다. 즉, 일반적인 커플 파트너십과 비교할 때, 국제결혼과 다문화에 대한 관심과 이해, 담론과 인식이 부분적·제한적임을 알 수 있다.

둘째, 미시적 요인 중에서는 외국인 여성과 관련된 젠더 및 섹슈얼리티 이슈가 뚜렷하게 드러났다. 현재 파트너십과 관련된 우리나라 소셜미디어 게시글은 페미니즘 성격이 강한 반면(이재림, 이진경, 2020), 본 연구에서는 페미니즘 성격이 거의 나타나지 않았고, 외국인 여성을 결혼상대로 대상화하는 내용이 부각되었다. ‘몸무게’, ‘얼굴’ 등 외모 관련 내용이 별도의 토픽으로 나타날 만큼 비중이 컸으며, ‘아가씨’, ‘초혼’ 등 처녀성, 순결성과 관련된 토픽과 핵심단어가 등장했다. 이를 통해 중개업에 의한 국제결혼에서 나타나는 여성의 대상화, 몸 상품화, 순결이데올로기 강조 양상(소라미, 2009)이 온라인에도 자리잡고 있음을 확인할 수 있었다.

나아가 국제결혼이 매매혼이라는 부정적 인식은 미시적 영역에서 나타난 젠더와 섹슈얼리티 이슈와 연결되어 있었다. 국제결혼과 관련하여 소득, 재산, 직장, 주택 등 경제적 요인이 토픽으로 나타났으며, 중개기업 등을 통해 조건 위주로 진행되는 양상이 두드러졌다. 이는 국제결혼이 매매혼이라는 인식과 밀접한 관련이 있다. 매매혼 성격의 국제결혼은 금전적 거래의 형식을 띠고 있어 남녀간 평등한 결혼이 이루어지지 않는다는 점에서 가부장 권력의 작동을 반영하며(권창규, 2017), 평등 가족을 지향하는 최근 한국사회의 트렌드에 역행하는 문화적 퇴행으로 인식된다(이수자, 2004). 관련하여 국제결혼과 다문화 관련 부정적 인식 토픽의 핵심단어로 가부장적인 한국남성을 부정적으로 일컫는 은어인 ‘한남’이 등장하였다. 데이

터 수집기간인 2017년에서 2019년은 한국사회 젠더갈등이 심화되고 표면화된 시기였으며, 실제로 2019년 1월 청와대 국민청원 게시판에 ‘농촌 총각 국제결혼 지원금(매매혼 장려금) 세금 지원 폐지하세요’라는 청원이 올라왔다. 비슷한 내용이 2017년에도 게시되었다. 해당 시기 여성단체와 시민단체가 2006년부터 시작된 미혼 남성 국제결혼 지원제도를 ‘매매혼 장려금’이라고 표현하기 시작했으며, 이에 대한 비판적 여론이 국민청원으로까지 등장한 것이다. 이로 인해 지방자치단체의 국제결혼 지원 정책 규모가 점차 축소되고 있는 상황은 국제결혼이 매매혼이라는 부정적 인식과 한국사회 젠더갈등이 융합되는 양상을 보여주는 사례이다.

따라서 과거 한국사회의 다문화가족에 대한 편견과 부정적 인식은 인종 및 민족적 차원, 단일민족주의로 인한 것으로 이해되었다면, 현재는 단순히 민족주의적 이유가 아니라 경제적 요인과 젠더, 섹슈얼리티 요인으로 인해 증폭되고 있다고 설명할 수 있다. 즉, 토픽모델 분석에서 나타난 한국사회 젠더갈등과 관련된 단어들은 국제결혼과 다문화 관련 이슈를 젠더 등 변화하는 한국사회의 맥락에서 이해해야 함을 보여준다.

셋째, 커플 파트너십 온톨로지에 포함된 사회관계(친구 및 동료, 공동체, 네트워크, 연대 등)에 대한 토픽이 본 연구에서는 등장하지 않았다. 이러한 결과는 2018년 전국다문화가족실태조사(여성가족부, 2019b)의 결과와도 맥을 같이 한다. 이 조사에서 한국에 거주한 기간이 늘면서 한국생활 적응력은 높아진 반면, 도움을 받거나 의논할 상대가 없다고 답한 비율도 함께 늘어났다. 또한 자녀교육 관련 등 사회적 관계가 없다고 응답한 비율도 2015년 대비 증가하는 등 사회관계망 부족이 심화된 것으로 나타났다. 다양한 사회집단과 교류하고 접촉하는 것이 결혼이주 여성의 차별 경험과 주관적 건강 악화의 메커니즘에서 보호요인으로 작용한다는 연구결과도 있는 만큼(Cheon & Chung, 2016), 결혼이주여성과 다문화가정의 사회관계망에 대한 연구와 지원이 필요하다.

마지막으로, 거시적 요인에서는 국가와 사회, 정책, 규범의 영역이 등장했다. 이를 통해 국가와 사회의 규범과 정책이 국제결혼과 다문화 관련 토픽 전반에 영향을 미치고, 또 영향을 받는 것을 확인할 수 있다. 또한 한국이 다문화사회로 발전하는 과정에서, 다문화가족 정책이나 다문화가족지원센터를 통한 서비스 제공 등 국가 주도의 대응 정책이 큰 역할을 한 현실도 반영되었을 것으로 보인다. 한편, 인식, 외모, 결단, 고정관념, 상호작용 등과 관련

된 토픽도 비중 있게 등장하였다. 인식에서는 긍정심리와 부정심리가 모두 나타나 고정관념과 달리 국제결혼과 다문화 관련 긍정심리도 주요하게 나타났다. 후속연구에서는 국제결혼과 다문화와 관련된 긍정심리와 부정심리의 내면을 좀 더 심층적으로 살펴볼 필요가 있다.

이상과 같은 본 연구의 결과는 국제결혼, 결혼이주여성, 다문화가족에 대한 부정적인 인식을 해소하기 위한 정책과 교육의 방향성에 시사하는 바가 크다. 특히, 국제결혼을 매매혼이라고 바라보는 일반인들의 인식이 선행연구에서는 간접적으로 나타났던 것에 비해, 본 연구에서는 직접적으로 확인되어 관련 개입의 경험적 근거로 활용될 수 있다. 2020년에 여성가족부는 교육부, 국방부 등 관계부처와 함께 다문화수용성 제고, 인권보호 강화, 균등한 기회보장 및 포용사회환경 조성, 사각지대 없는 복지서비스 제공의 4대 지원방안을 골자로 하는 다문화가족 포용대책을 발표하였다. 다문화인구는 꾸준히 증가하는데 비해, 다문화가족이 일상에서 겪는 차별은 여전하다는 지적에 따른 것이다. 본 연구의 결과는 이러한 정책의 필요성을 전반적으로 지지하며 해당 정책을 강화하는 근거로 활용될 수 있다.

또한 본 연구의 결과를 토대로 한국사회 구성원의 국제결혼 및 다문화 관련 인식을 연구할 때, 단일민족주의뿐 아니라, 국제결혼이 매매혼이라는 인식이나 젠더 이슈, 인권에 관한 문항 등을 포함하는 것이 효과적일 것이라고 제안할 수 있다. 현재 다문화수용성 및 인식과 관련된 국내외 연구는 국민 정체성, 단일민족주의, 귀속적 기준, 일자리 등 외국 이주민에 대한 위협인식만을 포함하는데(여성가족부, 2019a), 젠더 및 인권을 고려한다면 변화하는 한국 사회의 모습을 포착할 수 있을 것이다. 이외에도 본 연구를 통해 국제결혼, 결혼이주여성, 다문화가족에 대한 일반인들의 인식과 논의가 부정적 측면에 집중되지 않도록 다양한 파트너십 형태, 라이프스타일, 여가, 미디어, 소비, 반려동물, 가사노동, 종교 등 다양한 일상생활 영역과 긍정적 측면을 조명할 필요가 있음을 알 수 있다. 특히 사회관계망에 대한 연구 및 개입의 필요성이 나타났다.

본 연구의 제한점과 후속연구에 대한 제언은 다음과 같다. 첫째, 본 연구에서 분석한 소셜 빅데이터는 커플 파트너십 온톨로지에 기반하여 정형화되었기 때문에, 다문화와 국제결혼과 관련된 특수한 단어 중 일부는 포함되지 않았을 가능성이 있다. 추후 국제결혼 및 다문화 관련 소셜 빅데이터 연구를 목적으로 한 전용 온톨로지를 개발한다면 보다 풍부한 단어를 포함할 수 있을 것이다. 둘째, 본 연

구에서는 3년간의 온라인 게시물을 수집했기 때문에 장기간에 걸친 트렌드나 시계열적 변화는 알기 어렵다. 셋째, 확률모형에 기반한 토픽모델 분석을 실시한 것은 본 연구의 강점이지만 감성분석 등 다른 빅데이터 분석기법의 적용 가능성도 열려 있다. 감성분석은 문서에 나타난 감성단어를 활용하여 긍정, 부정, 중립 등의 감정과 태도를 분석하는 방법으로 해당 문서가 내포하는 감정이 긍정적인지 부정적인지를 파악할 수 있다. 감성분석을 실시한다면 어떠한 맥락에서 긍정적 또는 부정적 뉘앙스의 단어가 사용되었는지 구체적인 양상을 확인할 수 있을 것이다.

본 연구는 다문화가족에 대한 편견과 차별, 인식 개선이 사회적 의제인 한국 사회에서, 사회구성원의 솔직한 생각과 인식을 반영한 소셜 빅데이터 자료를 활용했으며 토픽모델 분석이라는 진일보한 연구방법을 사용하였다. 이를 통해 사회적 바람직성의 문제가 발생할 가능성이 있는 전통적인 조사방법으로 확인하기 어려운 국제결혼과 다문화에 대한 제한적, 부정적 인식을 경험적으로 드러냈다는 점에서 본 연구의 의의가 있다. 다문화 가정 자녀들이 청소년기와 청년기에 접어드는 등 한국 사회가 다문화 사회 성숙기에 진입한만큼, 새로운 세대의 담론장인 온라인상의 소셜 빅데이터 분석 연구는 앞으로 더욱 중요해질 것으로 기대된다.

참고문헌

- 강미옥, 이정애, 최은경(2019). 베트남 결혼이주 여성의 정체성, 자본, 이데올로기. *다문화와 평화*, 13(2), 107-126. <http://dx.doi.org/10.22446/mpisk.2019.13.2.006>
- 강진구(2011). 국제결혼담론의 계보학적 연구 : '사진결혼'을 중심으로. *아시아여성연구*, 50(2), 211-236.
- 강진구, 이기성(2018). 한국문학의 다문화담론에 대한 메타분석(meta-analysis). *우리문학연구*, 59, 7-28. <http://dx.doi.org/10.20864/skl.2018.07.59.7>
- 권창규(2017). 가부장 권력과 화폐 권력의 결탁과 경합: 김유정 소설을 중심으로. *여성문학연구*, 42(1), 159-184.
- 김갑순, 윤상기(2019). 예비유아교사의 인권의식, 다문화 수용성, 다문화 교수효능감이 다문화 교육역량에 미치는 영향. *학습자중심교과교육학회지*, 19(14), 137-157. <http://dx.doi.org/10.22251/jlcci.2019.19.14.137>
- 김경미(2012). 여성결혼이민자의 사회연결망과 한국생활 만족도: 중국, 베트남, 일본 출신을 중심으로. *한국인구학*, 35(2), 185-208.

- 김정희(2021). 일본 다문화공생 정책의 평가와 과제-공생은 어떻게 가능한가-. *日本語文學*, 1(90), 223-248. <http://dx.doi.org/10.18704/kjill.2021.9.90.223>
- 김미희(2007). 결혼 이민 여성의 결혼에 대한 인식: 베트남 이민여성에 대한 사례 연구. *교육연구논총*, 28(2), 183-207.
- 김선숙, 고대영, 김혜원, 조요셉, 최성은, 민기채(2020). 결혼 이주여성의 한국어 능력이 생활만족도와 우울에 미치는 영향: 주관적 계층의식의 매개효과 검증. *사회과학연구*, 31(3), 59-78. <http://dx.doi.org/10.16881/jss.2020.07.31.3.59>
- 김승희(2011). 다문화가족의 자녀교육 문제에 관한 질적 연구. *한국가정관리학회지*, 29(4), 17-33.
- 김오남(2006). 국제결혼 이주여성의 부부갈등 결정요인 연구. *가족과 문화*, 18(3), 63-106.
- 김정현(2015). 도시형 다문화가정 결혼이주여성의 문화적응과 식생활행태 분석. *한국가정과교육학회지*, 27(4), 173-189.
- 김주아, 이경숙, 이제원(2019). 텍스트 마이닝을 활용한 영재 교육에 대한 대중 인식 분석. *교육과학연구*, 50(4), 275-299. <http://dx.doi.org/10.15854/jes.2019.12.50.4.275>
- 김형지, 김선우, 김성태(2018). 스마트폰 중독에 관한 미디어 보도 토픽 모델링 분석: 2010년~2018년까지 정부 정책 비교를 중심으로. *방송통신연구*, 104, 38-69.
- 김혜선(2019). 베트남 결혼이주여성의 사회적 네트워크와 연계적 사회자본. *가족과 문화*, 31(3), 139-178. <http://dx.doi.org/10.21478/family.31.3.201909.005>
- 김혜순(2014). 결혼이민여성의 이혼과 '다문화정책': 관료적 확장에 따른 가족정책과 여성정책의 몰이민적·물성적 결합. *한국사회학*, 48(1), 299-344. <http://dx.doi.org/10.21562/kjs.2014.02.48.1.299>
- 김혜순(2017). 국내외 정책 환경을 감안한 다문화가족정책 조정방안: 다문화인식개선사업 재고와 주류화. *입법과 정책*, 9(1), 31-55.
- 류은영, 조숙정(2021). 다문화가정 가정폭력 피해 경험 실태 분석과 함의. *Crisisonomy*, 17(2), 145-159. <https://doi.org/10.14251/crisisonomy.2021.17.2.145>
- 류한수(2016). 결혼이주여성의 차별경험이 신체적, 정신적 건강에 미치는 영향. *한국콘텐츠학회논문지*, 16(8), 345-356. <http://dx.doi.org/10.5392/JKCA.2016.16.08.345>
- 문성은, 방은령(2014). 유아교사의 다문화수용성이 다문화교육 실체에 미치는 영향: 다문화 교수효능감의 매개효과를 중심으로. *한국자치행정학보*, 28(2), 285-313. <http://dx.doi.org/10.18398/kjlgas.2014.28.2.285>
- 문유경, 전기택(2011). *남녀 결혼이민자 사회통합지표 개발 연구*. 서울: 한국여성정책연구원.
- 민무숙, 김이선, 주유선, 이정연(2013). 결혼이주여성의 사회적 관계 양상을 통해 본 사회통합의 과제. *여성연구*, 85(2), 5-43. <http://dx.doi.org/10.33949/tws.2013..2.001>
- 박동숙, 임해영, 김은경(2019). 결혼이주여성의 적응 경험에 관한 연구. *한국가족복지학*, 65, 5-41. <http://dx.doi.org/10.16975/kjfs.2019..65.001>
- 박선영, 이재립(2021). 소셜 빅데이터로 알아본 코로나19와 가족생활: 토픽모델 접근. *한국콘텐츠학회논문지*, 21(3), 282-300. <http://dx.doi.org/10.5392/JKCA.2021.21.03.282>
- 박신규(2008). 국제결혼이주여성의 정체성 및 주체성의 사회적 위치성에 따른 변화: 구미 지역의 국제결혼이주여성의 생애사 분석을 중심으로. *한국지역지리학회지*, 14(1), 40-53.
- 박윤정(2016). 다문화가족지원센터의 기능 및 사례. *한중미래연구*, 7, 61-87.
- 박종도(2019). 토픽모델링을 활용한 다문화 연구의 이슈 추적 연구. *한국문화정보학회지*, 53(3), 273-289.
- 박홍식(2013). 결혼중개업의 관리에 관한 법률 해석과 국제결혼중개업의 공익화 방안에 대한 연구. *법학연구*, 24(2), 95-139.
- 배성찬, 편성진(2013). 다문화자녀 교육·복지 서비스 관리를 위한 통합관리 시스템개발에 관한 연구. *한국자치행정학보*, 27(1), 293-318. <http://dx.doi.org/10.18398/kjlgas.2013.27.1.293>
- 설동훈, 강기정, 이병하(2011). *다문화가족 사회통합관련 정책과제 개발 및 발전방안 연구: 우리나라와 외국의 정책현황 및 추진체계 비교분석을 중심으로*. 서울: 여성가족부.
- 설동훈, 윤홍식(2008). 여성결혼이민자의 사회경제적 적응과 복지정책의 과제: 출신국가와 거주 지역에 따른 상이성을 중심으로. *사회보장연구*, 24(2), 109-134.
- 소라미(2009). 결혼 이주여성의 인권 실태와 한국 법제도 현황에 대한 검토. *법학논총*, 16(2), 1-32.
- 송길영(2012). Social big data mining service 활용의 예. *한국 IT서비스학회 학술대회논문집*, 161-187.
- 신영진, 이광석(2019). 여성결혼이민자지원정책의 현황진단 및 개선방안. *한국범죄정보연구*, 5(2), 63-84. <https://doi.org/10.33563/KSCIA.2019.5.2.4>
- 안명숙(2018). 빅데이터를 활용한 다문화 핵심단어 및 네트워크 분석. *융복합지식학회논문지*, 6(2), 67-76.
- 양경은, 노법래(2020). 한국 다문화 담론 구조와 그 시계열적

- 변동 - 언론 기사문 텍스트 마이닝 분석을 중심으로. *한국사회복지학*, 72(3), 33-58.
- 양순미, 정현숙(2006). 농촌 국제결혼부부의 결혼생활 적응과 만족에 대한 영향 요인. *한국가족관계학회지*, 11(3), 223-252.
- 여성가족부(2012). *국민 다문화수용성 조사 연구*. 서울: 여성가족부.
- 여성가족부(2019a). 2018 다문화수용성조사. 서울: 여성가족부.
- 여성가족부(2019b). 2018 전국다문화가족실태조사. 서울: 여성가족부.
- 염동문(2019). 결혼이주여성의 문화적응 스트레스에 대한 중단적 변화양상과 변화양상별 영향요인. *복지상담교육연구*, 8(2), 263-282. <http://dx.doi.org/10.20497/jwce.2019.8.2.263>
- 오승욱, 한진욱, 김민수(2020). 소셜 빅데이터를 활용한 참여스포츠 담론 분석. *한국스포츠산업경영학회지*, 25(2), 18-30. <http://dx.doi.org/10.31308/KSSM.25.2.2>
- 오은주, 고진호(2021). 한국 이주 다문화 담론기술과 분석방법론 연구. *인문사회* 21, 12(3), 1645-1657. <http://dx.doi.org/10.22143/HSS21.12.3.115>
- 윤인진, 팽정(2021). 한국인과 대만인의 다문화 수용성 : 국민 정체성과 국가자긍심의 효과를 중심으로. *아시아연구*, 64(1), 245-288. <http://dx.doi.org/10.31930/JAS.2021.03.64.1.245>
- 이병호, 김가연(2019). 소셜 빅데이터로 본 노인 관련 담론의 특징과 변화. *사회복지정책*, 46(3), 179-201. <http://dx.doi.org/10.15855/swp.2019.46.3.179>
- 이상노, 이현심(2014). 결혼이주여성의 한국생활 적응. *한국지역사회복지학*, 49, 105-136. <http://dx.doi.org/10.15300/jcw.2014.49.2.105>
- 이수자(2004). 이주여성 디아스포라 : 국제성별분업, 문화혼성성, 타자화와 섹슈얼리티. *한국사회학*, 38(2), 189-219.
- 이인희, 황경아(2013). 다문화 관련 미디어 보도 프레임 연구에 대한 메타분석. *다문화사회연구* 6(2), 83-108. <http://dx.doi.org/10.15685/jms.2013.08.6.2.83>
- 이재림, 이진경(2020, 11월). 비혼 관련 온라인 뉴스, 블로그, 카페, 트위터 게시글의 토픽모델 분석. *한국가족관계학회 2020년 추계학술대회 자료집*, 60-70.
- 이종열, 정준호(2012). 한국과 대만의 다문화사회 정책 특성과 유형에 관한 비교분석: 추진전략과 기능을 중심으로. *한국비교정부학보*, 16(3), 157-186. <http://dx.doi.org/10.18397/kcgr.2012.16.3.157>
- 이지연, 그레이스 정(2019a). 베트남 출신 결혼이주여성의 자녀양육 경험: 위계의 교차 위에 놓여있는 이중민족사회. *가족과 문화*, 31(2), 182-227. <http://dx.doi.org/10.21478/family.31.2.201906.006>
- 이지연, 그레이스 정(2019b). 잠재프로파일 분석을 통한 청소년기 다문화가정 자녀의 심리적 적응 유형 분류. *다문화사회연구*, 12(2), 85-114. <http://dx.doi.org/10.14431/jms.2019.06.12.2.85>
- 이진경, 이재림, 박선영(2020). 소셜 빅데이터 연구를 위한 커플 파트너십 온톨로지 개발. *가정과삶의질연구*, 38(4), 1-21. <http://doi.org/10.7466/JFBL.2020.38.4.1>
- 이하윤(2005). 인신매매 형태의 국제결혼에 대한 고찰. *공의과인권*, 2(1), 99-128.
- 이현아, 류석진(2013). 온라인 커뮤니티의 사회적 자본과 제도: 디시인사이드 바람의 화원/바람의 나라 갤러리를 중심으로. *정보사회와 미디어*, 27, 25-55.
- 장주영, 박민정, 김기태(2019). 난민에 대한 국민의 인지도 연구. 경기: 이민정책연구원.
- 정선아, 최수찬(2017). 다문화가정의 한국인 남성배우자의 문화적응태도, 부부갈등, 직무몰입에 관한 연구. *사회복지 실천과 연구*, 14(2), 39-62.
- 정용교(2016). 다문화사회의 성공적 정착과 실제: 대만의 다문화 시민교육을 중심으로. *동아인문학*, 37, 817-843.
- 조은별, 민지연, 박수원(2020). 학술연구와 소셜 빅데이터의 창의성에 대한 인식 비교 : 텍스트마이닝을 활용한 주제 분석을 중심으로. *창의력교육연구*, 20(3), 47-67. <http://dx.doi.org/10.36358/JCE.2020.20.3.47>
- 조은영, 이신동(2020). 다문화 부모교육 프로그램 효과에 대한 메타분석. *학습자중심교과교육연구*, 20(7), 377-396. <http://dx.doi.org/10.22251/jlcci.2020.20.7.377>
- 조흥석(2020). 국제결혼중개업법과 결혼이주여성의 인권보호. *법과 정책연구*, 20(4), 131-160.
- 천호성, 이정희(2014). 일본 다문화정책의 정책기조와 특징-다문화교육과 거버넌스 체제를 중심으로 -. *사회과교육*, 53(3), 15-29.
- 최은수, 최인철, 최종안(2019). 대한민국 행복 지도: 소셜미디어를 활용한 대한민국 지역별 행복 비교. *한국심리학회지: 일반*, 38(4), 647-667. <http://dx.doi.org/10.22257/kjp.2019.12.38.4.647>
- 최현태(2011). 국제결혼중개계약 관련 법률의 문제점과 해결 방안. *법과 정책연구*, 11(1), 91-114.
- 추주희, 오승원(2020). 토픽모델링으로 본 한국 가족연구 동향 분석: 2000년대 이후 국내 학술지를 중심으로. *한국가족관계학회지*, 24(4), 3-23. <http://dx.doi.org/10.21321/jfr.24.4.3>
- 통계청(2017). 2017년 다문화 인구동태 통계. 대전: 통계청.

- 통계청(2020). 2019년 혼인·이혼 통계. 대전: 통계청.
- 통계청(2021). 2020년 혼인·이혼 통계. 대전: 통계청.
- 허청아, 그레이스 정(2018). 다문화 청소년의 차별경험, 스트레스, 자아존중감의 잠재프로파일 분류 및 영향요인 검증. *한국가정관리학회지*, 36(4), 177-190. <http://dx.doi.org/10.7466/JKHMA.2018.36.4.177>
- Arnaboldi, M., Brambilla, M., Cassottana, B., Ciuccarelli, P., Ripamonti, D., Vantini, S., & Volonterio, R. (2016, April). *Studying multicultural diversity of cities and neighborhoods through social media language detection*. Tenth International AAAI Conference on Web and Social Media, Cologne, Germany.
- Ashton, W., Bhattacharyya, P., Galatsanou, E., Ogoe, S., & Wilkinson, L. (2016). *Emerging uses of big data in immigration research*. Canada: Rural Development Institute, Brandon University.
- Bennett, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10(2), 179-196. [https://doi.org/10.1016/0147-1767\(86\)90005-2](https://doi.org/10.1016/0147-1767(86)90005-2)
- Berry, J. W., & Kalin, R. (1995). Multicultural and ethnic attitudes in Canada: An overview of the 1991 national survey. *Canadian Journal of Behavioural Science*, 27(3), 301-320. <https://doi.org/10.1037/0008-400X.27.3.301>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. <https://doi.org/10.1145/2133806.2133826>
- Brat, M. (2021). Mindfulness of inclusivity and belonging: Taking refuge in a multicultural community. *Western Journal of Communication*, 1-8. <https://doi.org/10.1080/10570314.2021.1949031>
- Cheon, Y. M., & Chung, G. H. (2016). The role of social contact in the experience of discrimination: Implications for subjective health among marriage migrant women in South Korea. *Asian Social Work and Policy Review*, 10(1), 34-49. <https://doi.org/10.1111/aswp.12073>
- Chung, G. H. & Lim, J. (2016). Marriage immigrant mothers' experience of perceived discrimination, maternal depression, parenting behaviors, and adolescent psychological adjustment among multicultural families in South Korea. *Journal of Child and Family Studies*, 25(9), 2894-2903. <https://doi.org/10.1007/s10826-016-0445-2>
- Chung, G. H. & Yoo, J. P. (2013). Using the multicultural family support centers and adjustment among inter-ethnic and interracial families in South Korea. *Family Relations*, 62(1), 241-253. <https://doi.org/10.1111/j.1741-3729.2012.00754.x>
- Chung, G. H., Lee, J., & Gonzales-Backen, M. (2020). Context of biethnic acceptance, biethnic identity affirmation, and life satisfaction among biethnic adolescents in South Korea. *Journal of Adolescence*, 80, 242-253. <https://doi.org/10.1016/j.adolescence.2020.03.004>
- Chung, W., & Zeng, D. (2013, January). *iMood: discovering US immigration reform sentiment*. 23rd Workshop on Information Technology and Systems: Leveraging Big Data Analytics for Societal Benefits, Milan, Italy.
- Dickey, R. J. (2014). "Multicultural families" in Asia: Examining the past, analyzing the present, projecting future policies. *한국사회과학연구*, 33(2), 181-207. <http://dx.doi.org/10.18284/jss.2014.12.33.2.181>
- Igarashi, A. (2019). Till multiculturalism do us part: Multicultural policies and the national identification of immigrants in European countries. *Social Science Research*, 77, 88-100. <https://doi.org/10.1016/j.ssresearch.2018.10.005>
- Ishikawa, H. (2015). *Social big data mining*. Florida: CRC Press.
- Jones, G. W. (2012). *International marriage in Asia: What do we know, and what do we need to know?*. Singapore: Asia Research Institute.
- Jones, G., & Shen, H. H. (2008). International marriage in East and Southeast Asia: Trends and research emphases. *Citizenship Studies*, 12(1), 9-25. <https://doi.org/10.1080/13621020701794091>
- Kelly, L. (2001). Marriage for sale: The mail-order bride industry and the changing value of marriage. *Journal of Gender, Race and Justice*, 5, 175-196.
- Kitchin, R., & McArdle, G. (2016). What makes big Data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 1-10. <https://doi.org/10.1177/2053951716631130>
- Koide, T., Yoshida, T., Ogawa, E., Kuramoto, M., Homma, J. E., & Naruse, M. (2019). Uncovering inner dilemmas experienced by parents of multicultural families in Japan. *Journal of Comparative Family Studies*, 50(3), 261-284. <https://doi.org/10.3138/jcfs.50.3.004>
- Kuramoto, M., Koide, T., Yoshida, T., & Ogawa, E. (2017). Raising multicultural children in Japan: A mixed me-

- thods examination of parent-child-society dynamics. *Journal of Intercultural Communication Research*, 46(4), 360-384. <https://doi.org/10.1080/17475759.2017.1355839>
- Kusel, V. I. (2014). Gender disparity, domestic abuse, and the mail-order bride industry. *Albany Government Law Review*, 7, 166-186.
- Piper, N. (1997). International marriage in Japan: 'Race' and 'gender' perspectives. *Gender, Place and Culture: A Journal of Feminist Geography*, 4(3), 321-338. <https://doi.org/10.1080/09663699725378>
- Sercu, L. (2004). Assessing intercultural competence: A framework for systematic test development in foreign language education and beyond. *Intercultural Education*, 15(1), 73-89. <https://doi.org/10.1080/1467598042000190004>
- Sirbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., ... & Sharma, R. (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11, 341-360. <https://doi.org/10.1007/s41060-020-00213-5>
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161-197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Toyota, M. (2008). Editorial introduction: International marriage, rights and the state in East and Southeast Asia. *Citizenship Studies*, 12(1), 1-7. <https://doi.org/10.1080/13621020701794083>
- Tsuneyoshi, R., Okano, K. H., & Boocock, S. (Eds.). (2010). *Minorities and Education in Multicultural Japan: An interactive perspective*. New York: Routledge.
- Wanner, P. (2021). How well can we estimate immigration trends using Google data?. *Quality & Quantity*, 55, 1181-1202. <https://doi.org/10.1007/s11135-020-01047-w>