

뉴스 빅데이터를 활용한 코로나19 언론보도 분석

:토픽모델링 분석을 중심으로

COVID-19 News Analysis Using News Big Data

: Focusing on Topic Modeling Analysis

김태종

한국청소년정책연구원 청소년정책분석평가센터

Tae-Jong Kim(k2boy3@naver.com)

요약

본 연구의 목적은 최근 확산되고 있는 코로나19의 뉴스 빅데이터를 활용하여 언론을 통해 사회적으로 형성되고 있는 주요 의제가 무엇이며 어떻게 변화하는지 파악해, 추후 언론보도의 방향성을 제안하는 것이다. 이를 위해 2019년 12월 31일부터 2020년 3월 11일까지 보도된 47,816건의 뉴스 빅데이터를 감염병 위기경보 4단계(관심-주의-경계-심각)를 기준으로 4개 시기로 구분하여 토픽모델링 분석을 실시해, 총 20개의 토픽을 도출하였다.

토픽 모델링 분석 결과를 토대로, 본 연구에서는 다음 사항들을 제안하였다. 첫째, '불안', '공포' 등의 자극적인 표현을 자제하고 중립적이고 객관적인 보도용어 사용이 필요하다. 둘째, 단순 사건보도식 뉴스제작을 탈피하여, 더욱 심층적이고 맥락적인 뉴스 제작이 요구된다. 셋째, 감염병 관련 상황별 세부 위키커뮤니케이션 매뉴얼 마련이 필요하다. 넷째, 시민주도의 위기 극복노력을 중점적으로 조명하는 보도가 필요하다. 본 연구는 코로나19 뉴스 빅데이터를 토픽모델링 분석방법을 활용하여 분석한 최초의 논문이라는 학술적 의의와 국가 위키커뮤니케이션 정책개발의 기초자료로 활용될 수 있는 정책적 의의를 가진다.

■ 중심어 : | 코로나19 | 감염병 | 뉴스 빅데이터 | 언론보도 분석 | 토픽모델링 분석 |

Abstract

The purpose of this study is to find out what the main agenda of social formation is and how it changes through the media by utilizing the news big data of COVID-19 which is spreading recently, and to suggest the direction of future reporting. In order to achieve the purpose of the research, 47,816 cases of news big data reported from December 31, 2019 to March 11, 2020 were divided into four periods based on the fourth stage of the crisis warning for infectious diseases, and a total of 20 topics were derived.

Based on the results of the Topic Modeling analysis, this study proposed the following. First, it is necessary to refrain from provocative expressions such as "anxiety" and "fear" and use neutral and objective reporting terms. Second, more in-depth and contextual news production is required, breaking away from simple event news production. Third, it is necessary to prepare detailed crisis communication manuals for each situation related to infectious diseases. Fourth, we need reports that focus on citizens-led efforts to overcome the crisis. This research has the academic significance that it is the first paper to analyze news big data on COVID-19 using the Topic Modeling Analysis method, and the policy significance that can be used as the basis for developing national crisis communication policy.

■ keyword : | COVID-19 | Infectious Disease | News Big Data | News Analysis | Topic Modeling Analysis |

I. 서론

본 연구의 목적은 코로나바이러스감염증-19(이하 '코로나19') 관련 국내 뉴스 빅데이터를 토픽모델링 분석방법을 활용하여 분석함으로써, 감염병 위기경보 단계별로 나타난 코로나19 언론보도의 주요 키워드를 식별하고 토픽을 도출하며 토픽의 변화 양상을 파악하는 데에 있다.

2020년 1월 20일 국내에서 코로나19에 감염된 첫 확진자가 발생한 이후, 44일 만에 누적 확진자가 5,000명을 넘었다[1][2]. 2020년 4월 21일 0시 기준 코로나19 국내 총 누적 확진자수는 10,683명, 사망자 수는 237명이며, 국외에서는 중국, 미국, 스페인, 이탈리아 등 215개국에서 2,408,855명의 확진자가 발생하여 168,491명이 사망했다[3].

코로나 바이러스는 동물 및 사람에게 감염될 수 있는 바이러스로서, 현재까지 사람에게 중증 폐렴을 일으키는 유형으로는 사스(SARS-CoV), 메르스(MERS-CoV), 코로나19(COVID-19)가 발견되었다. 코로나19는 비말, 접촉을 통해 전파되며, 발열, 권태감, 기침, 호흡곤란, 폐렴, 급성호흡곤란증후군 등 경증에서 중증까지 다양한 호흡기감염증상이 나타난다. 코로나19는 고령, 면역기능이 저하된 환자, 기저질환을 가진 환자에게 중증 또는 사망을 초래할 수 있으며, 코로나19에 대한 백신은 아직 개발되지 않았다[4].

코로나19 확산에 따라 정부는 2020년 2월 23일 감염병 위기경보 단계를 '경계' 단계에서 최고 단계인 '심각' 단계로 격상했다. 정부는 중앙재난안전대책본부를 설치하여 중앙정부 및 지방자치단체의 지원 체계를 한층 강화하고, 코로나19의 해외 유입 차단, 환자 발견 및 접촉자 격리 등의 봉쇄정책, 지역사회 확산 차단 및 최소화를 위한 조치를 추진했다[5].

세계보건기구(WHO: World Health Organization)는 2020년 3월 11일 코로나19에 대해 전염병 경보 최고 단계인 '세계적 대유행', 즉 '팬데믹(Pandemic)'을 선언했다. 세계보건기구는 코로나19의 전 세계적 확산 사태가 단순한 공중보건의 위기가 아니라 정치·경제·사회문화 등 모든 분야에 걸쳐 영향을 미치는 심각한 위기이며, 이를 극복하기 위한 전세계 국가의 적극적인 대응을 촉구했다[6].

이러한 감염병 확산 시기에 언론의 뉴스 보도량은 크게 증가하며, 언론은 감염병 위협의 경고자이자 의제 설정자, 위험인식의 확산자이자 국민 의견의 전달자 등의 역할을 수행한다[7]. 한편, 언론은 지나친 속보 경쟁으로 인한 부정확한 보도, 공포를 조장하는 자극적인 보도, 불필요한 피해자 신원 노출 및 혐오 확산 보도를 한다는 지적을 받기도 한다[8]. 따라서, 언론이 감염병과 관련하여 어떠한 의제를 설정하여 어떻게 보도하는지 객관적인 데이터를 바탕으로 분석하는 연구가 필요하다.

현재까지 이루어진 코로나19 관련 국내 연구(KCI 등재지 기준)로는, 최예린(2020)의 코로나19에 대한 보육교사의 정서적 경험과 달라진 어린이집 일상에 관한 연구, 문현철(2020)의 해외 감염병재난 대응체계의 개선방안에 관한 연구, 김영국(2020)의 개정데이터 3법과 보험업의 과제(디지털 헬스케어 서비스 활성화를 중심으로), 유연실(2020)의 중국 근현대 의료사 연구의 새로운 흐름과 동향 등이 있으나[9-12], 코로나19 관련 언론보도 현상을 분석한 연구는 아직 이루어지지 않았다.

기존 감염병 관련 언론보도를 분석한 국내 연구(KCI 등재지 기준)에서는, 2009년 신종 플루부터 2015년 메르스에 이르기까지 국내에서 확산된 감염병 언론보도를 대상으로 빈도분석, 내용분석, 프레임분석, 의미연결망분석, 감성분석 등 다양한 분석방법이 사용되었다[13-19]. 최근에는 감염병 관련 뉴스 빅데이터와 텍스트마이닝 기법을 적용한 토픽모델링 분석방법을 활용하여 분석의 신뢰도와 타당도를 높이는 연구가 주목을 받고 있다[20].

따라서, 본 연구에서는 뉴스 빅데이터와 토픽모델링 분석방법을 활용하여 코로나19 관련 언론보도에 나타난 주요 키워드 및 토픽, 토픽의 변화를 시기(감염병 위기경보 단계)별로 파악하기 위해 다음과 같이 연구문제를 설정하였다.

〈연구문제1〉 국내 코로나19 언론보도에 나타난 시기별 주요 키워드는 무엇인가?

〈연구문제2〉 국내 코로나19 언론보도에 나타난 시기별 토픽은 무엇이며, 어떻게 변화하는가?

II. 연구 방법

1. 자료 수집

분석 대상 언론매체는 [표 1]과 같이 10개 신문사와 3개 방송사 등 총 13개 매체이다. 신문사 선정기준은 발행부수 순이며, 방송사는 지상파로 선정했다.

표 1. 분석 대상 언론매체

유형	매체수	매체
신문사	10	조선일보, 중앙일보, 동아일보, 한겨레, 경향신문, 한국일보, 서울신문, 세계일보, 국민일보, 문화일보
방송사	3	KBS, MBC, SBS

분석기간은 [그림 1]과 같이 중국 우한에서 코로나19 발생 상황이 최초로 국내에서 보도된 2019년 12월 31일부터 2020년 3월 11일까지 총 72일이다. 정부의 감염병 위기경보 단계 격상과 함께 보도건수가 급증하는 현상이 나타남에 따라, 감염병 위기경보 단계를 기준으로 4개 시기로 구분하여 분석을 실시하였다. 제1기는 감염병 위기경보 ‘관심’ 단계인 2019년 12월 31일부터 2020년 1월 19일까지, 제2기는 ‘주의’ 단계인 2020년 1월 20일부터 1월 26일까지, 제3기는 ‘경계’ 단계인 2020년 1월 27일부터 2월 22일까지, 제4기는 ‘심각’ 단계인 2020년 2월 23일부터 코로나19 보도량이 크게 감소한 3월 11일까지이다.

자료수집을 위한 검색어는 ‘코로나19’, ‘신종 코로나’, ‘우한 폐렴’으로 선정하였다. 자료수집은 한국언론진흥

재단에서 운영하는 빅카인즈(BIGKinds)를 활용하였다. 분석 대상 뉴스 건수는 [표 2]와 같으며, 최초 수집된 자료는 총 50,263건으로, 이 중 중복되어 보도된 뉴스와 단순 행사안내 및 인물동정 등의 뉴스 2,447건을 제외하여 최종 47,816건의 뉴스를 대상으로 분석하였다. 본 연구에서는 데이터 전처리와 형태소 추출이 완료된 정제 데이터를 빅카인즈로부터 제공받아 분석을 실시하였다.

표 2. 분석 대상 뉴스 건수(건)

단계	제1기	제2기	제3기	제4기	계
최초 수집 뉴스	316	1,002	18,947	29,998	50,263
제외 뉴스	26	63	971	1,387	2,447
분석 대상 뉴스	290	939	17,976	28,611	47,816

2. 분석 방법

본 연구에서는 코로나19 관련 뉴스 빅데이터 텍스트에 내재된 키워드와 토픽을 추출하기 위해, 컴퓨터 기반 텍스트 분석방법인 토픽모델링 중 최근 널리 쓰이고 있는 LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당) 기법을 사용하였다[21][22].

LDA기법은 문헌들을 구성하는 키워드들을 출현확률 및 분포에 따라 군집화하여 토픽을 추론하는 통계적 텍스트 처리 기법이다[23]. 본 연구에서는 NetMiner 프로그램을 활용하여 뉴스 텍스트에서 추출된 키워드를 매트릭스로 생성해 LDA 분석($\alpha=0.1$, $\beta=0.01$)을 실시하였다[24].

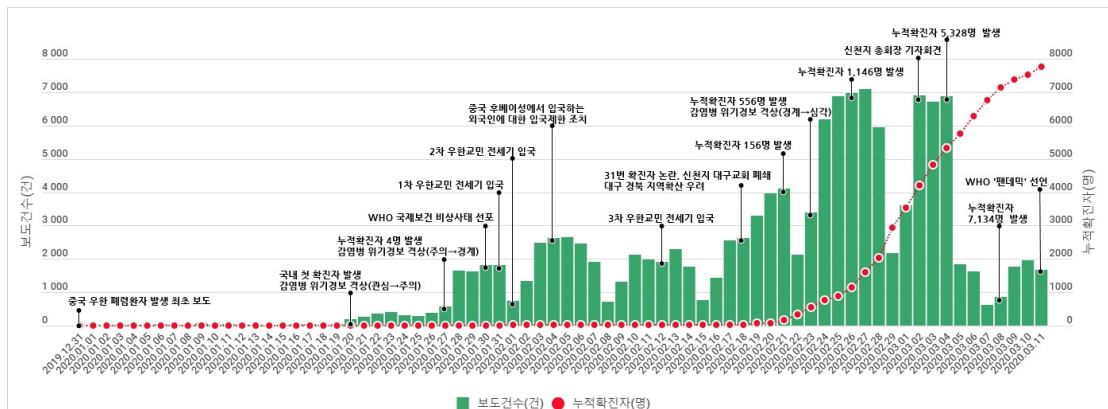


그림 1. 일자별 보도건수 및 누적확진자 수

* 자료 출처: 질병관리본부 홈페이지(www.cdc.go.kr) 및 빅카인즈 홈페이지(www.bigkinds.or.kr)에서 재구성

토픽의 수를 결정하기 위해, 샘플링 반복횟수를 1,000회로 설정하여 토픽의 수를 2개에서 30개로 변경하면서 토픽모델링을 실시한 결과, 각 시기별로 5개의 토픽을 추출했을 경우 주요 키워드의 중복이 최소화되고 범주별로 해석이 용이하게 분류되어, 토픽수를 5개로 설정해 분석을 수행했다. 토픽모델링 결과, 토픽별로 뉴스들이 분류되었으며, 토픽별 출현확률이 높은 순으로 키워드가 도출되었다. 이후 도출된 주요 키워드와 뉴스원문을 비교하여 토픽명을 부여하였다.

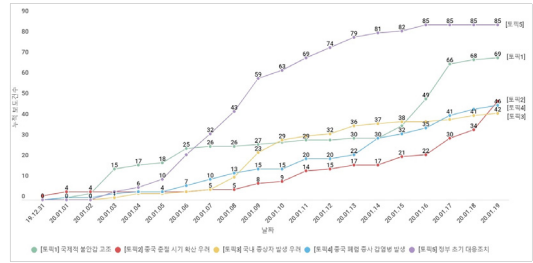


그림 2. 제1기 토픽 변화

III. 연구 결과

1. 제1기 '관심' 단계 분석 결과

제1기 감염병 위기경보 '관심' 단계인 2019년 12월 31일부터 2020년 1월 19일까지 보도된 코로나19 관련 뉴스 290건을 분석한 결과, 토픽별 주요 키워드와 토픽은 [표 3], 토픽의 변화는 [그림 2]와 같이 나타났다.

[토픽1]은 세계보건기구, 불안감, 감염자, 감염, 전염 등의 키워드가 도출됨에 따라, 토픽명을 '국제적 불안감 고조'로 부여했으며, 관련 뉴스건수는 69건(23.8%)이다.

[토픽2]는 감염자, 춘절, 대이동, 확진, 환자 등의 키워드가 도출됨에 따라, 토픽명을 '중국 춘절 시기 확산 우려'로 부여했으며, 관련 뉴스건수는 48건(16.5%)이다.

[토픽3]은 증상자, 당국, 질병관리본부, 감염병, 사람 등의 키워드가 도출됨에 따라, 토픽명을 '국내 증상자 발생 우려'로 부여했으며, 관련 뉴스건수는 42건(14.5%)이다.

표 3. 제1기 토픽 및 주요 키워드

토픽명	주요 키워드	출현비율	주요 뉴스 제목	뉴스건수(건)	토픽 비중(%)
[토픽1] 국제적 불안감 고조	세계보건기구	0.022	[세계보건기구] "중국 폐렴 대응 긴급위원회 소집할 수도"	69	23.8
	불안감	0.021	중국, 홍콩, 대만, 사스 [불안감] 에 긴장 고조		
	감염자	0.02	태국 이어 일본서도 '우한 폐렴' [감염자] 확인		
	감염	0.016	'사스' 공포 부른 中 폐렴 집단 [감염] 주변국도 초비상		
[토픽2] 중국 춘절 시기 확산 우려	전염	0.015	일본서도 '우한 폐렴' 확인 사람 간 [전염] 가능성도	48	16.5
	감염자	0.027	영연구진, "중국 신종 코로나바이러스 [감염자] 1700명 넘을듯"		
	춘절	0.025	중국 '우한 폐렴' 2번째 사망자 발생 [춘절] 앞두고 '전염 확산' 우려		
	대이동	0.021	중 '우한 폐렴 확진' 62명 달해 춘제 [대이동] 시작에 확산 우려		
	확진	0.013	中 '우한 폐렴' 하루새 17명 [확진] '춘제' 대규모 확산 우려		
[토픽3] 국내 증상자 발생 우려	환자	0.009	중국 '우한 폐렴' [환자] 17명 또 발생 춘제 대이동 비상	42	14.5
	증상자	0.042	국내서 '중국 우한 원인불명 폐렴' [증상자] 1명 발생		
	당국	0.039	中 춘제 앞두고 '우한 폐렴' 확산 보건 [당국] 비상		
	질병관리본부	0.036	[질병관리본부] "우한 방문 중국인 여성 1명 격리치료 중"		
	감염병	0.029	국경 없는 [감염병] 매년 오염지역서 수백만명 한국 입국		
[토픽4] 중국 폐렴 증상 감염병 발생	사람	0.016	중원인불명 폐렴 증상 여성 호전 [사람] 간 전파 배제 못한다"	46	15.9
	폐렴	0.057	중국 '원인 모를 [폐렴] 확산 '사스 공포' 퍼지자 시장 폐쇄		
	환자	0.054	中 원인 불명 폐렴 [환자] 44명으로 급속히 늘어 11명 증대		
	후베이	0.037	중국 [후베이] 성 우한에서 발생한 원인불명 폐렴		
	당국	0.027	中 우한시 [당국] "원인불명 집단폐렴, 사스 아니다"		
[토픽5] 정부 초기 대응조치	발생	0.026	"中 신종 폐렴으로 최초 사망자 [발생] "	85	29.3
	질병관리본부	0.026	중국서 폐렴환자 집단발생, [질병관리본부] 24시간 대응체계 가동		
	입국자	0.017	中 우한서 원인불명 폐렴 집단 발생 일본 [입국자] 검역 강화"		
	조사	0.017	中 미스터리 폐렴 증상 국내서도 발생... 36세 중국여성 [조사] 중		
	대응	0.013	중국발 코로나바이러스 전파 우려, 일본 "지역사회 [대응] 강화"		
계	검사법	0.012	신종 코로나는 사스(SARS)와 유사 일본 "1달 내 [검사법] 개발"	290	100

[토픽4]는 폐렴, 환자, 후베이, 당국, 발생 등의 키워드가 도출됨에 따라, 토픽명을 '중국 폐렴 증상 감염병 발생'으로 부여했으며, 관련 뉴스건수는 46건(15.9%)이다.

[토픽5]는 질병관리본부, 입국자, 조사, 대응, 검사법 등의 키워드가 도출됨에 따라, 토픽명을 '정부 초기 대응조치'로 부여했으며, 관련 뉴스건수는 85건(29.3%)이다.

특히, 질병관리본부가 24시간 대응체계를 가동하기 시작한 1월 5일부터 [토픽5] '정부 초기 대응조치' 관련 뉴스가 다른 토픽들에 비해 크게 증가하면서 국가 주도의 코로나19 대응활동이 언론의 주목을 받았다.

2. 제2기 '주의' 단계 분석 결과

제2기 감염병 위기경보 '주의' 단계인 2020년 1월 20일부터 2020년 1월 26일까지 보도된 코로나19 관련 뉴스 939건을 분석한 결과, 토픽별 주요 키워드와 토픽은 [표 4], 토픽의 변화는 [그림 3]과 같이 나타났다.

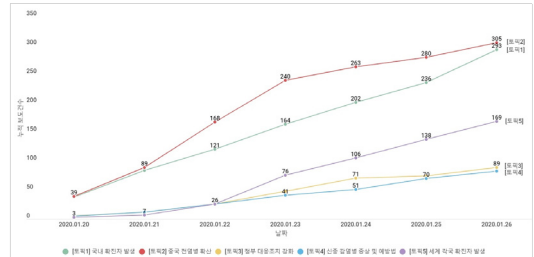


그림 3. 제2기 토픽 변화

[토픽1]은 인천공항, 유증상자, 접촉자, 감염병, 환자 등의 키워드가 도출됨에 따라, 토픽명을 '국내 확진자 발생'으로 부여했으며, 관련 뉴스건수는 293건(31.2%)이다.

[토픽2]는 전염병, 상하이, 광둥, 시장, 사람 등의 키워드가 도출됨에 따라, 토픽명을 '중국 전염병 확산'으로 부여했으며, 관련 뉴스건수는 305건(32.5%)이다.

[토픽3]은 보건당국, 환자, 격리, 치료, 폐렴 등의 키워드가 도출됨에 따라, 토픽명을 '정부 대응조치 강화'로 부여했으며, 관련 뉴스건수는 89건(9.5%)이다.

[토픽4]는 환자, 신종, 감염, 예방, 발생 등의 키워드

표 4. 제2기 토픽 및 주요 키워드

토픽명	주요 키워드	출현비율	주요 뉴스 제목	뉴스건수(건)	토픽 비중(%)
[토픽1] 국내 확진자 발생	인천공항	0.029	국내 첫 우한 폐렴 확진 35세 中여성 [인천공항] 서 바로 격리	293	31.2
	유증상자	0.026	국내 '우한 폐렴' [유증상자] 5명, 1명은 확진에 4명은 검사 중		
	접촉자	0.019	우한 폐렴 국내 첫 환자 [접촉자] 모두 감시체계 진행 문제는 '전염력'		
	감염병	0.019	국내 첫 확진 [감염병] 위기 경보 '주의'		
	환자	0.017	국내 '우한 폐렴' 두 번째 확진 [환자] 발생		
[토픽2] 중국 전염병 확산	전염병	0.031	'우한 폐렴' 확진 60명, 사망 2명 또 늘어 '사스급 [전염병] 우려	305	32.5
	상하이	0.014	中 '우한 폐렴' [상하이] 서도 첫 환자 발생 확진자 218명으로 급증		
	광둥	0.014	베이징 [광둥] 상하이까지 '우한폐렴' 중국서 급속도 확산		
	시장	0.014	中 뒤덮은 폐렴 바이러스, "우한 해산물 [시장] 서 팔린 '박쥐'에서 확산"		
	사람	0.013	우한 폐렴, [사람] 간 전파 확인 中 대확산 1~2주가 고비		
[토픽3] 정부 대응조치 강화	보건당국	0.044	외교부, 우한 여행 자제 격상, [보건당국] '24시간 방역' 체제로	89	9.5
	환자	0.028	국내서 첫 우한 폐렴 [환자] 확인 위기경보 '주의' 격상		
	격리	0.024	국내서 '우한 폐렴' 확진자 확인... 보건 당국 [격리] 검사		
	치료	0.019	일본 "늦게 발견하면 치명률 높아져" 조기발견 [치료] 당부		
	폐렴	0.017	일본 "중국 본토 전체가 오염 감시대상" 우한 [폐렴] 조치 확대		
[토픽4] 신종 감염병 증상 및 예방법	환자	0.025	우한폐렴 [환자] '감기'인줄 뚝 찢어지는 고통"	83	8.8
	신종	0.021	[신종] 코로나바이러스 '우한 폐렴' 증상? "발열 오한 감기와 비슷"		
	감염	0.016	급속 확산 중인 중국 '우한 폐렴' [감염] 여부 및 예방 수칙은		
	예방	0.011	'우한 폐렴'에 정부 24시간 대응체계 [예방] 손 씻기' 가장 중요		
	발생	0.011	한국인 첫 확진자 [발생] 전파력과 예방 수칙		
[토픽5] 세계 각국 확진자 발생	미국	0.024	[미국] 도 '우한 폐렴' 방역망 뚫렸다 금융시장까지 공포감 확산	169	18.0
	일본	0.022	'열 나네요' [일본] 관광 中 여성 '우한 폐렴' 3번째 감염자		
	프랑스	0.015	[프랑스] 호주서도 '우한 폐렴' 확진자 전세계 확산세		
	봉쇄	0.013	뒤늦은 [봉쇄] 골든타인 놓쳤다 전 세계로 '확산'		
	태국	0.013	'우한 폐렴' 동남아시아로 확산 [태국] 확진자 5명 말레이는 3명		
계				939	100

가 도출됨에 따라, 토픽명을 '신종 감염병 증상 및 예방법'으로 부여했으며, 관련 뉴스건수는 83건(8.8%)이다.

[토픽5]는 미국, 일본, 프랑스, 봉쇄, 태국 등의 키워드가 도출됨에 따라, 토픽명을 '세계 각국 확진자 발생'으로 부여했으며, 관련 뉴스건수는 169건(18.0%)이다.

특히, 2020년 1월 20일 국내에서 첫 확진자가 발생하면서 [토픽1] '국내 확진자 발생' 관련 뉴스가 증가했고, 이와 함께 [토픽2] '중국 전염병 확산' 관련 보도량이 더욱 가파른 상승 추세를 보였다.

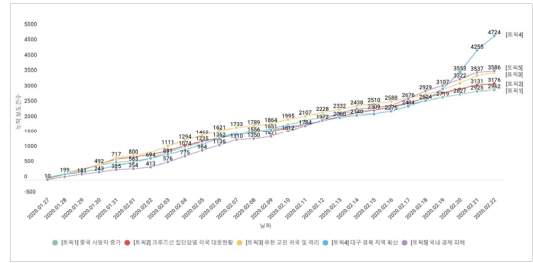


그림 4. 제3기 토픽 변화

3. 제3기 '경제' 단계 분석 결과

제3기 감염병 위기경보 '경제' 단계인 2020년 1월 27일부터 2020년 2월 22일까지 보도된 코로나19 관련 뉴스 17,976건을 분석한 결과, 토픽별 주요 키워드와 토픽은 [표 5], 토픽의 변화는 [그림 4]와 같이 나타났다.

[토픽1]은 후베이, 우한, 사망자, 신종, 사스 등의 키워드가 도출됨에 따라, 토픽명을 '중국 사망자 증가'로 부여했으며, 관련 뉴스건수는 2,962건(16.5%)이다.

[토픽2]는 일본, 미국, 한국, 크루즈, 전세계 등의 키워드가 도출됨에 따라, 토픽명을 '크루즈선 집단감염 각국 대응현황'으로 부여했으며, 관련 뉴스건수는 3,176건(17.7%)이다.

[토픽3]은 교민, 아산, 진천, 주민, 격리 등의 키워드가 도출됨에 따라, 토픽명을 '우한 교민 귀국 및 격리'로 부여했으며, 관련 뉴스건수는 3,528건(19.6%)이다.

표 5. 제3기 토픽 및 주요 키워드

토픽명	주요 키워드	출현비율	주요 뉴스 제목	뉴스건수(건)	토픽 비중(%)
[토픽1] 중국 사망자 증가	후베이	0.032	'우한 폐렴'으로 중국 [후베이]성서만 100명 사망	2,962	16.5
	우한	0.028	발병서 입원까지 10일 사이 중증 사망 [우한]의 비극		
	사망자	0.021	중코로나19 [사망자] 2000명 육박, 확진 7만 2000명 넘어		
	신종	0.017	중국서 [신종] 코로나' 800명 이상 사망		
	사스	0.015	중국 내 신종 코로나 사망자 수, [사스] 넘어섰다		
[토픽2] 크루즈선 집단감염 각국 대응현황	일본	0.033	[일본] 정부 '집단 감염' 크루즈선 대응 '갈팡질팡'	3,176	17.7
	미국	0.02	신종코로나로 日 크루즈 고립 [미국] 국적 승객 300명 전세계로 귀국		
	한국	0.019	日 크루즈선 [한국] 국적 7명 오늘 정부 수송기로 귀국		
	크루즈	0.016	[크루즈] 재양에도 日정부, 영국 배라며 책임 회피		
[토픽3] 우한 교민 귀국 및 격리	전세기	0.012	캐나다 홍콩 대만도 [전세기] 보내 日 크루즈 탑승자 철수 지원	3,528	19.6
	교민	0.009	30, 31일 우한에 전세기 [교민] 700명 국내 이송		
	아산	0.008	정부, 우한 교민 [아산] 진천 격리수용 결정		
	진천	0.007	격리 지역 번복한 정부 [진천] 아산은 봉쇄 시위		
	주민	0.006	마음 돌린 진천·아산 [주민], "우한 교민 2주간 편하게 생활하셨으면"		
[토픽4] 대구 경북 지역 확산	격리	0.005	우한 교민 2주 [격리] 해제 366명 1차 퇴소	4,724	26.3
	대구	0.027	코로나 청정지역 [대구]서 첫 확진자 31번 환자는 61세 여성		
	신천지	0.013	[신천지]교회 동시 예배자만 1000명 31번 환자 동선따라 코로나 공포 확산		
	경북	0.01	229명 확진자 중 대구 [경북] 확진자 200명 오늘-내일 확산 중대기로		
	교회	0.009	신천지 대구 [교회]서 예배한 신도 200여명, 전국에 흩어졌다		
[토픽5] 국내 경제 피해	지역	0.009	[지역]사회 확산 우려 현실로 대구경북 등 무더기 확진	3,586	19.9
	경제	0.009	신종 코로나 '경제' 타격 우려 현실로 GDP 1%대 성장 전망도		
	기업	0.007	중소[기업] 34% 코로나 타격, 서비스업 '손님 줄었다' 울상		
	시장	0.006	공포에 전염된 금융[시장], 코스피 코스닥 3% 급락		
	소상공인	0.006	중기부, 중소기업 [소상공인] 신종 코로나 피해에 금융지원 나서		
업체	0.006	'우한 폐렴' 피해 [업체]에 법인세 최대 9개월 연장 세무조사도 멈춘다	17,976	100	

[토픽4]는 대구, 신천지, 경북, 교회, 지역 등의 키워드가 도출됨에 따라, 토픽명을 '대구-경북지역 확산'으로 부여했으며, 관련 뉴스건수는 4,724건(26.3%)이다.

[토픽5]는 경제, 기업, 시장, 소상공인, 업체 등의 키워드가 도출됨에 따라, 토픽명을 '국내 경제 피해'로 부여했으며, 관련 뉴스건수는 3,586건(19.9%)이다.

대부분의 토픽들이 유사한 수준에서 보도되다가, 2020년 2월 18일 31번 확진자의 동선과 관련된 논란이 발생하면서, [토픽4] '대구-경북지역 확산' 뉴스가 신천지 대구교회 신도들의 확진과 지역사회 확산에 대한 우려 내용을 포함하며 증가했다.

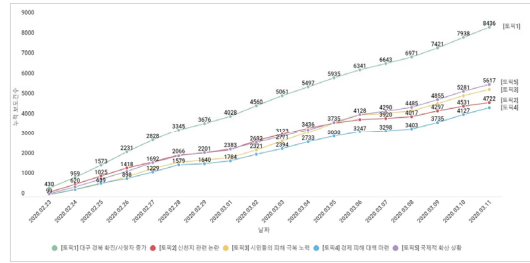


그림 5. 제4기 토픽 변화

4. 제4기 '심각' 단계 분석 결과

제4기 감염병 위기경보 '심각' 단계인 2020년 2월 23일부터 2020년 3월 11일까지 보도된 코로나19 관련 뉴스 28,611건을 분석한 결과, 토픽별 주요 키워드와 토픽은 [표 6], 토픽의 변화는 [그림 5]와 같이 나타났다.

[토픽1]은 격리, 경북, 확진자, 사망자, 접촉자 등의 키워드가 도출됨에 따라, 토픽명을 '대구-경북지역 확진/사망자 증가'로 부여했으며, 관련 뉴스건수는 8,436건(29.5%)이다.

[토픽2]는 교회, 신도, 이만희, 총회장, 기자 등의 키워드가 도출됨에 따라, 토픽명을 '신천지 관련 논란'으로 부여했으며, 관련 뉴스건수는 4,722건(16.5%)이다.

[토픽3]은 지역, 의료진, 시민, 지원, 확산 등의 키워드가 도출됨에 따라, 토픽명을 '시민들의 피해 극복 노력'으로 부여했으며, 관련 뉴스건수는 5,369건(18.8%)이다.

표 6. 제4기 토픽 및 주요 키워드

토픽명	주요 키워드	출현비율	주요 뉴스 제목	뉴스건수(건)	토픽 비중(%)
[토픽1] 대구-경북지역 확진/사망자 증가	격리	0.02	코로나19로 대구 아파트 첫 코호트 [격리] "46명 확진, 출입금지"	8,436	29.5
	경북	0.018	코로나19 [경북] 확진환자 1000명 육박 7일 0시 기준 990명		
	확진자	0.017	이 시각 대구 경북 상황 [확진자] 5,677명		
	사망자	0.009	대구 45번째 [사망자] 발생 국내 사망자수 66명으로 늘어		
	접촉자	0.008	대구 시내버스도 뚫렸다 이젠 밀집 [접촉자] 확인도 어려워		
[토픽2] 신천지 관련 논란	교회	0.011	"신천지 대구 [교회] 유증상자 1193명 중 80% 이상 확진"	4,722	16.5
	신도	0.01	"신천지 [신도] 21만2000명 명단 확보했다 진단 검사 진행"		
	이만희	0.007	"신도 수 숨겨 역학조사 방해" 신천지 교주 [이만희] 검찰 고발		
	총회장	0.005	檢, '감염병예방법' 위반 신천지 이만희 [총회장] 수사 착수		
	기자	0.005	신천지 이만희 "국민 여러분께 사죄" [기자] 회견		
[토픽3] 시민들의 피해 극복 노력	지역	0.009	인천 [지역] '착한 임대료 물결' 열흘 만에 367개 점포 동참	5,369	18.8
	의료진	0.007	'우리가 대구다' 전국서 [의료진] 물자 지원 쇄도		
	시민	0.006	품질 대란 속 "마스크 나눔 해요" 팔 걷어붙인 [시민]들		
	지원	0.006	대구 취약주민 [지원] 나선 세월호 가족 대구 시민도 "감사"		
	확산	0.005	"모두 힘내세요" 코로나19 [확산]에 통달집 사정도, 시민도 '기부 행렬'		
[토픽4] 경제 피해 대책 마련	소상공인	0.01	정부, 20조 풀어 코로나 충격 중기 [소상공인] 지원	4,467	15.6
	금융	0.008	정부, 코로나 대응에 30조원 이상 투입 재정 [금융] 세제에 韓韓까지 동원		
	시장	0.008	"기관, [시장] 안전판 역할해야" 코로나 19 금융상황 점검회의 열려		
	기업	0.007	우한 코로나 피해 중소 [기업] 금융지원 늘린다		
	경제	0.006	무디스 "한국 코로나19 추경, [경제]에 도움될 것"		
[토픽5] 국제적 확산 상황	미국	0.022	[미국]서 첫 감염경로 불명 환자 지역사회 전파 우려	5,617	19.6
	일본	0.019	[일본] 코로나19 확진 하루새 58명 늘어 1278명		
	이탈리아	0.013	[이탈리아] 누적 확진자 1만명 넘었다 하루 새 사망자 168명		
	사망자	0.01	독일서 코로나19 첫 [사망자] 발생 89세 여성 등 2명		
	감염자	0.007	전 세계 코로나19 [감염자] 10만명 돌파 발생국도 90곳 넘어		
계				28,611	100

[토픽4]는 소상공인, 금융, 시장, 기업, 경제 등의 키워드가 도출됨에 따라, 토픽명을 '경제 피해 대책 마련'으로 부여했으며, 관련 뉴스건수는 4,467건(15.6%)이다.

[토픽5]는 미국, 일본, 이탈리아, 사망자, 감염자 등의 키워드가 도출됨에 따라, 토픽명을 '국제적 확산 상황'으로 부여했으며, 관련 뉴스건수는 5,617건(19.6%)이다.

대구·경북지역에서 코로나19 확진자와 사망자가 크게 늘면서 [토픽1] '대구·경북지역 확진/사망자' 관련 보도량이 다른 토픽들의 보도량에 비해 약 2배 가까이 증가하는 추세를 보였다.

IV. 결론

본 연구의 목적은 최근 확산되고 있는 코로나19의 뉴스 빅데이터를 활용하여 언론을 통해 사회적으로 형성되고 있는 주요 의제가 무엇이며 어떻게 변화하는지 파악하여, 추후 언론보도의 방향성을 제안하는 것이다. 이를 위해 2019년 12월 31일부터 2020년 3월 11일까지 보도된 47,816건의 뉴스 빅데이터를 감염병 위기경보 4단계를 기준으로 4개 시기로 구분하여 토픽모델링 분석을 실시하였으며, 시기별로 다음과 같은 토픽이 도출되었다.

제1기에는 ①국제적 불안감 고조 ②중국 춘절 시기 확산 우려 ③국내 증상자 발생 우려 ④중국 폐렴 증상 감염병 발생 ⑤정부 초기 대응조치, 제2기에는 ①국내 확진자 발생 ②중국 전염병 확산 ③정부 대응조치 강화 ④신종 감염병 증상 및 예방법 ⑤세계 각국 확진자 발생, 제3기에는 ①중국 사망자 증가 ②크루즈선 집단감염 각국 대응현황 ③우한 교민 귀국 및 격리 ④대구·경북지역 확산 ⑤국내 경제 피해, 제4기에는 ①대구·경북지역 확진/사망자 증가 ②신천지 관련 논란 ③시민들의 피해 극복 노력 ④경제 피해 대책 마련 ⑤국제적 확산 상황 등의 토픽이 나타났다.

토픽의 변화 양상을 살펴보면, 제1기에는 한국 정부의 대응조치 관련 보도량이 초기에 크게 증가하다가, 제2기에는 중국의 전염병 확산과 국내 확진자 발생 상황이 중점적으로 보도되었다. 제3기에는 신천지와 관련한 대구·경북지역의 확산 상황이 언론의 주목을 받았으

며, 이러한 현상이 제4기에 심화되어 대구·경북지역 확진/사망자 관련 보도가 급증하는 모습을 보였다.

이번 코로나19와 관련된 언론보도 토픽들은 과거 감염병 보도양상과 비교하여 유사점과 차이점을 동시에 보이면서 다음과 같은 시사점이 나타났다[25].

첫째, 제1기에 도출된 5개의 토픽 중 3개의 토픽(국제적 불안감 고조, 중국 춘절 시기 확산 우려, 국내 증상자 발생 우려) 관련 뉴스에서 '불안', '공포', '우려' 등의 자극적인 용어가 다수 사용되었다. 이는 언론이 뉴스 수용자들의 부정적인 감정을 발화시켜 뉴스 주목도를 높이기 위해 사용했던 보도관행이다. 이러한 표현은 국민에게 불확실한 공포감을 유발시켜 감염병에 대한 정확한 이해와 대응에 혼란을 야기시킬 수 있으므로, 자극적인 표현을 자제하고 중립적이고 객관적인 보도 용어 사용이 필요하다.

둘째, 제1기부터 제4기에 도출된 20개의 토픽 중 15개의 토픽이 국내·외 코로나19 발생 및 확산 상황을 단순 기술하는 스트레이트 중심의 뉴스들로 구성되어 있다. 이에 반해 대응조치, 예방법, 극복 노력 등의 심층적인 맥락화된 뉴스로 구성된 토픽은 5개(정부 초기 대응조치, 정부 대응조치 강화, 신종 감염병 증상 및 예방법, 시민들의 피해 극복 노력, 경제 피해 대책 마련)에 불과했다. 즉, 감염자 발생과 피해상황, 사망자 수와 같은 사후 결과를 중점적으로 보도하는 반면, 감염병의 원인과 통제방안, 정부 대응의 적절성 검증, 시민들의 실천방안 보도 등의 예방 저널리즘은 상대적으로 부족했다. 따라서 감염병에 대한 단순 사건보도식 뉴스제작을 탈피하여, 더욱 심층적이고 맥락적으로 분석하여 위기를 극복할 수 있는 세부 실천방안을 제시하는 뉴스 제작이 요구된다.

셋째, 과거 신종플루·메르스 시기의 언론보도에서 중점적으로 나타나지 않았던 현상으로, 감염병 확산 초기 정부의 적극적인 대응조치가 언론의 주목을 받았다. 언론사의 정치적 성향에 따라 정부의 대응조치에 긍정적 또는 부정적으로 보도가 이루어지긴 했으나, 정부가 과거 감염병 상황시 지적받았던 '늑장 대응', '정보 비공개' 등의 문제점을 개선하여 좀 더 적극적·공개적으로 위기 커뮤니케이션을 실시한 것으로 나타났다. 신종 감염병의 특성상 예방과 방역, 검사와 치료, 해외 국가 및 국

내 지자체와의 협력 등 국가 차원의 컨트롤 타워 역할이 매우 중요하므로, 현재 코로나19에서 이루어졌던 위기커뮤니케이션 활동을 바탕으로, 더욱 세부적인 상황별 커뮤니케이션 매뉴얼을 마련하는 것이 필요하다.

넷째, 과거와 달리 이번 코로나19 관련 언론보도에서는, 시민들의 피해 극복 노력 관련 토픽이 중요한 비중으로 보도되었다. 특히, 감염병 위기경보단계가 '심각'으로 상향되고, 누적확진자가 500여명에서 5,000여명으로 급증하는 심각한 상황 속에서도, 시민들이 자발적으로 의료진을 지원하고 기부하며, '마스크 나눔', '착한 임대료' 등의 캠페인을 펼치는 등의 활동이 뉴스에서 부각되었다. 이러한 언론보도는 '시민이 주도하는 방역 활동'을 촉구하는 것은 물론, 위기극복과 사회적 공동체성 회복에 긍정적인 영향을 미칠 것이므로, 추후에도 이러한 시민들의 위기극복 노력을 중점적으로 조명하는 보도가 필요하다.

본 연구는 코로나19 언론보도를 분석한 최초의 논문으로서, 언론이 코로나19와 관련하여 어떠한 의제를 설정하여 보도하는지 뉴스 빅데이터와 토픽모델링 분석 기법을 활용하여 실증적·체계적으로 분석한 연구라는 학술적 의의를 가진다. 또한, 언론의 공포 소구 보도 관행과 단순 사건 보도식 뉴스제작 관행을 객관적 데이터를 통해 드러냄으로써, 국가 위기커뮤니케이션 정책 개발 시 '감염병 보도준칙' 마련과 허위정보 유포에 따른 '인포데믹(Infordemic)' 확산 방지 방안 수립, 감염병 위기경보 단계별 정부 공보조치 매트릭스 작성에 기초 자료로 활용될 수 있는 정책적 의의를 가진다.

본 연구의 한계 및 제언사항으로, 첫째, 분석기간이 총 72일에 한정되어, 전 시기에 나타난 코로나19 관련 토픽을 파악하는 데에 한계가 있다. 따라서, 코로나19 상황이 종료되는 시점까지의 뉴스를 추가하여 분석한다면 더욱 정확한 토픽을 파악할 수 있을 것이다. 둘째, 코로나19와 관련된 토픽은 뉴스뿐만 아니라 각계 각층에서 생산하는 다양한 텍스트를 통해서도 파악할 수 있다. 후속 연구에서는 뉴스와 함께, 정부 보도자료, SNS, 논문초록 등을 추가하여 분석을 수행할 것을 제언한다.

참고 문헌

- [1] 질병관리본부, “검역단계에서 해외유입 신종코로나바이러스 확진환자 확인, 감염병 위기 경보를 ‘주의’단계로 상향, 대응,” 질병관리본부 보도자료, 2020.1.20.
- [2] 질병관리본부, “코로나바이러스감염증-19 국내 발생 현황,” 질병관리본부 보도자료, 2020.3.4.
- [3] 질병관리본부, “코로나바이러스감염증-19 국내 발생 현황,” 질병관리본부 보도자료, 2020.4.21.
- [4] 김남순, “코로나바이러스감염증-19 현황과 과제,” 보건복지 Issue&Focus, 제373호, pp.1-12, 2020.
- [5] 질병관리본부, “코로나바이러스감염증-19 범정부대책회의 브리핑,” 질병관리본부 보도자료, 2020.2.23.
- [6] 옥철, “WHO, 코로나19 팬데믹 선언,” 연합뉴스, 2020.3.12.
- [7] 강아영, “감염병 보도 무엇보다 정확해야... 가이드라인 논의 필요,” 한국기자협회뉴스, 2020.2.14.
- [8] 김여라, “감염병 보도 규제의 현황 및 개선 방안,” 국회입법조사처 이슈와 논점, 제1665호, pp.1-4, 2020.
- [9] 최예린, “코로나 19에 대한 보육교사의 정서적 경험과 달라진 어린이집 일상에 관한 연구,” 한국유아교육연구, 제22권, 제1호, pp.253-279, 2020.
- [10] 문현철, “해외 감염병재난 대응체계의 개선 방안에 관한 연구,” 인문사회21, 제11권, 제1호, pp.155-170, 2020.
- [11] 김영국, “개정데이터 3법과 보험업의 과제(디지털 헬스케어 서비스 활성화를 중심으로),” 보험법연구, 제14권, 제1호, pp.495-524, 2020.
- [12] 유연실, “중국 근현대 의료사 연구의 새로운 흐름과 동향,” 역사학연구, 제77권, pp.1-49, 2020.
- [13] 주영기,유명순, “신문TV뉴스의 신종 출몰형 질환 및 만성질환 보도 패턴 분석,” 한국언론학보, 제54권, 제2호, pp.363-381, 2010.
- [14] 주영기,유명순, “한국 언론의 신종플루 보도 연구,” 한국언론학보, 제55권, 제5호, pp.30-54, 2011.
- [15] 차동필, “신종 조류독감 보도에 대한 제3자 효과 연구,” 홍보학연구, 제18권, 제1호, pp.188-211, 2014.
- [16] 진나영, 정정주, “한국 메르스 사태에 대한 국내의 언론보도 의미망 비교연구,” 언론과학연구, 제18권, 제2호, pp.222-262, 2016.
- [17] 박건숙, “전염병 보도 프레임 연구,” 한국어의미학, 제52권, pp.1-29, 2016.

- [18] 김영옥, 함승경, “보도 내러티브와 문화적 편향 그리고 사회자본이 낙인에 미치는 영향(메르스 사태를 중심으로 한 분석),” *Crissonomy*, 제13권, 제1호, pp.63-84, 2017.
- [19] 김병철, “메르스 보도가 메르스 정보 검색행위에 미치는 영향,” *커뮤니케이션학연구*, 제27권, 제1호, pp.207-224, 2019.
- [20] 조재희, 조인호, “2018 메르스 해외 재유입에 대한 주요 온라인 이슈 탐색,” *한국디지털콘텐츠학회논문지*, 제20권, 제5호, pp.1051-1060, 2019.
- [21] 정원준, 김대욱, 윤호영, 이형민, 박진우, 김동성, 손영곤, 전홍식, 천용석, 정유미, 박종구, *빅데이터의 분석방법과 활용*, 학지사, 2020.
- [22] 김태중, 박상욱, “뉴스 빅데이터를 활용한 평생교육 토픽 분석,” *평생교육학연구*, 제25권, 제3호, pp.29-63, 2019.
- [23] D. M. Blei and J. D. Lafferty, “Topic Models,” *Text Mining: Classification, clustering, and application*, Vol.10, No.71, pp.1-34, 2009.
- [24] W. Zhao, J. Chen, and W. Zen, “Best Practices in Building Topic Models with LDA for Mining Regulatory Textual Documents,” *CDER 9Th November*, 2015.
- [25] 권호천, “메르스 사태에 대한 신문보도의 의미연결망 분석,” *의료커뮤니케이션*, 제11권, 제1호, pp.63-80, 2016.

저 자 소 개

김 태 중(Tae-Jong Kim)

정회원



- 1999년 2월 : 경북대학교 신문방송학과(문학사)
- 2012년 2월 : 연세대학교 언론학과(언론학석사)
- 2019년 8월 : 공주대학교 교육학과(교육학박사)
- 2020년 4월 ~ 현재 : 한국청소년

정책연구원 부연구위원

〈관심분야〉 : 뉴스 빅데이터 분석, 미디어 교육, 평생교육