

토픽 모델링과 로지스틱 회귀분석을 활용한 한국 사회의 혐오 담론 탐구 및 정책 결정 과정과의 상관성 분석

이지인¹, 우병득², 이민구³, 박세윤⁴

요 약

한국 사회의 혐오 담론은 어떻게 변화되어 왔는가? 또한 혐오 담론이 실질적으로 정책 결정 과정에 미치는 영향력은 어떠한가? 다양한 선행연구들이 혐오에 관한 학문적 이해의 폭을 넓혔음에도 불구하고, 혐오 담론의 변화 과정과 혐오 담론이 정책 결정 과정에 미치는 영향력에 대한 이해는 여전히 부족하다. 본 논문은 2000년부터 2021년까지 한국 사회의 혐오 담론 지형을 실증적으로 파악하고 혐오 담론이 정책 결정 과정에 미치는 영향력을 살펴본다. 본 연구에서는 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 기법을 활용하여 뉴스 빅데이터를 분석한다. 한국의 혐오 담론을 살펴본 결과 정치와 관련된 혐오는 지속적으로 높은 관심을 받아왔지만, 성소수자, 이민자, 남성, 플랫폼 등의 주제들은 특정 사건이 발생함에 따라 주목도가 상승하는 모습을 보였다. 이에 더해, 로지스틱 회귀분석 모형을 통해 혐오 담론이 정책 결정 과정에 미치는 영향력을 분석함으로써 남녀갈등에 관한 혐오 기사의 수가 증가할수록 국회의원들의 여성 대상 폭력과 관련한 법안의 발의 확률이 감소한다는 점을 발견하였다. 또한, 성별과 혐오 기사 수와의 교차항을 활용하여 여성 의원들이 남성 의원들에 비해서 남녀갈등에 관한 혐오 기사의 수에 영향을 덜 받는다는 점을 검증하였다.

주요어: 혐오 담론, 텍스트 분석, 잠재 디리클레 할당, 정책 결정 과정, 로지스틱 회귀분석.

1. 서론

한국 사회의 혐오 담론은 어떻게 변화되어 왔는가? 또한 혐오 담론이 실질적으로 정책 결정 과정에 미치는 영향력은 어떠한가? 현재 한국 사회에서 혐오(Hatred)는 오늘날 한국 사회에서 폭넓은 사회문화적 맥락을 포함하고 있으며, 자신과 다른 타인에 대한 거부부터 성별, 인종, 민족, 종교 등 특정 집단에 대한 비방과 모욕을 포함한 복합적인 개념으로 정의된다(Son, 2018).

혐오는 비단 우리나라에서만 관찰되는 현상이 아니다. 2000년대 들어 전 세계적으로 혐오의 부정적

¹37673 경상북도 포항시 남구 청암로 77, 포항공과대학교 융합대학원 소셜데이터사이언스전공 석사과정.
Email: jiinlee0017@postech.ac.kr

²(교신저자) 37673 경상북도 포항시 남구 청암로 77, 포항공과대학교 사회문화데이터사이언스연구소 박사후연구원.
Email: byungdeukwoo@postech.ac.kr

³337673 경상북도 포항시 남구 청암로 77, 포항공과대학교 융합대학원 소셜데이터사이언스전공 석사과정.
Email: mglee815@postech.ac.kr

⁴37673 경상북도 포항시 남구 청암로 77, 포항공과대학교 융합대학원 소셜데이터사이언스전공 석사과정.
Email: seyoonpark@postech.ac.kr

영향력에 대한 학문적 관심이 비약적으로 증가했다. Harel, Jameson, Maoz(2020)는 질적연구를 통해 페이스북(Facebook)과 같은 소셜 미디어가 미국 사회에서 혐오 담론의 창출에 미치는 영향력을 강조하였으며, Bilewicz et al.(2017)는 실험적 연구방법을 통해 폴란드의 사례를 분석하여 사회 지배 지향성이 높은 사람일수록 혐오 표현에 대한 수용성이 높다는 점을 발견했다. 이외에도 학생이 가지고 있는 혐오에 대한 감정이 면학 분위기에 미치는 영향력 등과 같은 다양한 연구들이 진행되고 있다(Lee et al., 2022).

이러한 선행연구들이 혐오에 관한 학문적 이해의 폭을 넓혔음에도 불구하고, 혐오 담론의 변화 과정에 대한 학문적 이해는 아직 부족하며, 혐오 담론이 국민들의 실생활과 밀접한 연관이 있는 정책 결정 과정에 어떠한 영향을 미치는지에 대한 분석 또한 활발히 진행되지 않았다. 격화되고 있는 혐오 문제와 정책 결과가 국민들에게 미치는 영향을 고려하였을 때, 이러한 연구의 부재는 예상 밖이다. 따라서 본 연구는 빅카인즈가 제공하는 뉴스 빅데이터를 활용하여 한국 사회의 혐오 담론 지형을 파악하고자 한다. 또한, 남녀갈등과 관련한 혐오를 중심으로 로지스틱 회귀분석 모형을 활용하여 혐오 담론이 정책 결정 과정에 미치는 영향력을 분석하려 한다.

2. 한국 사회의 혐오 담론 분석

2.1 분석 방법

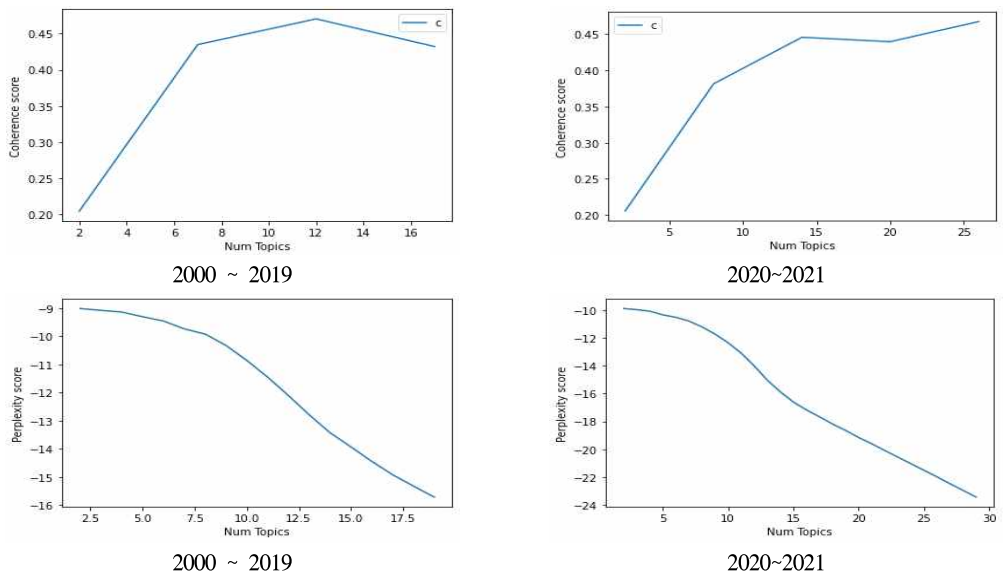
본 연구는 한국언론진흥재단의 분석 시스템인 빅카인즈(Bigkinds)에서 제공하는 뉴스 빅데이터를 활용하여 한국의 혐오 담론 지형을 파악하고자 한다. 이를 위해 빅카인즈가 뉴스 빅데이터를 누적하기 시작한 1990년 1월 1일부터 2021년 12월 31일까지, 빅카인즈에 ‘혐오’를 검색 쿼리(query)로 입력하여 매칭되는 총 10,493개의 기사 데이터를 수집하였으며 본격적으로 혐오 담론이 언론에서 대두되기 시작한 2000년대 이후의 데이터를 분석에 활용하였다. 빅카인즈는 기사별 상위 키워드를 텍스트랭크 모형(TextRank Model)을 통해 추출한다. 텍스트랭크 모형은 두 단어 사이의 동시 출현 관계를 이용하여 문서를 설명하는 특성 단어를 추출하는 자연어처리 모형이다(Li, Zhao, 2016).

본 논문에서는 분석 대상으로 기사의 본문을 대표하는 기사별 상위 10개의 키워드들을 설정하고 LDA(Latent Dirichlet Allocation) 분석을 진행하여 한국 사회의 혐오 담론 지형을 파악한다. LDA 분석은 다양한 사회과학 연구 분야에서 대량의 말뭉치(corpus)의 주제를 요약하거나 시기에 따른 주제의 변화를 살펴보는 방법으로 사용되고 있다(Mun, 2021; Suh, 2020). 이를테면, Lee, Park, Cho(2021)는 미국 대통령 선거에서 당선된 “조 바이든(Joseph Robinette Biden)”이 현지 언론에서 어떻게 보도되고 있는지를 LDA에 기반한 주제 모형으로 해석하였다.

LDA 분석 기법을 사용할 때 연구자는 대량의 문서에 잠재되어 있는 토픽의 수를 예측해야 한다. 토픽의 개수를 정확히 파악하고 있지 못할 경우, 주로 혼잡도(Perplexity)나 응집성(Coherence) 지수를 통해 최적의 토픽 개수를 찾는다(Kang, Lim, 2020). 혼잡도란 어떠한 단어 집합이 주어졌을 때, 모델이 다음 단어를 예측할 수 있는 정도를 의미하며, 응집성이란 각 클러스터(cluster)들의 평균적인 응집력을 의미한다(Mimno et al., 2011). 혼잡도는 모델의 분류 성능에 집중하여 결과를 얼마나 정확하게 예측하는지 판단하는 지표임에 반해 응집성 지수는 하나의 클러스터가 얼마나 잘 나뉘었는지를 평가하는 지표다. 본 연구에서는 연구자가 LDA 분석 결과를 통하여 기사에 나타나는 혐오 이슈를 탐색하는 것이 목적이므로 혼잡도보다는 토픽의 의미론적 일관성을 측정하는 응집성 지수를 내재된 토픽의 개수를 선정하는 기준으로 사용했다.

본 연구는 토픽 모델링을 통해 2000년부터 2020년까지 한국 언론에서 대두된 혐오 주제들을 폭넓은 시각에서 살펴보고자 하였다. Kang, Lim(2021)은 검색어의 주제 유형을 분석하기 위한 토픽 모델링을 진행하기에 앞서 코로나19와 관련된 질의가 특정 시기에 급증한다는 점에서 범용적 검색 질의 주제에 대비해 대표성이 떨어진다는 것으로 판단, 코로나19를 제외하고 토픽 모델링을 진행하였다. 2020년부터 등장한 코로나19 이슈는 큰 쟁점이 되며 특정 시기 내 작성된 언론 기사에 많은 영향력을 끼쳤다. Son(2020)은 코로나19 이슈가 공중 보건뿐 아니라 정치, 경제, 사회, 문화를 포괄하는 전 분야에 영향을 미쳤으며, 언론 영역 역시 코로나19로 인해 큰 변화를 겪고 있음을 시사하였다. 이렇듯 특정 시기에 코로나19가 미친 영향력과 선행연구를 고려하였을 때 코로나가 영향을 미치기 전과 후의 뉴스 기사를 분리하여 토픽 모델링을 진행할 필요가 있음을 판단하였다. 2000년부터 2021년까지의 뉴스 기사 전체를 분석 대상으로 선정하였을 때 코로나19라는 주제가 2019년 이전의 데이터에도 영향을 미칠 수 있음을 고려하여, 토픽 모델링의 대상은 코로나19의 영향을 받기 전인 2000년부터 2019년까지의 기사와 영향을 받은 이후인 2020년 이후로 분할되었다. 본 연구는 코로나19 기점으로 나뉜 두 데이터 그룹에 대해 각각 응집성과 혼잡도 지수를 구한 뒤 응집성 지수를 기반으로 LDA 분석을 시행하였다.

Figure 1은 코로나19 발생 이전과 이후 각각 데이터에 대해 응집성 지수와 혼잡도 지수를 구한 결과이다. 응집성 지수를 통해서는 토픽의 개수를 산정할 수 있었으나, 혼잡도 지수에서는 유의미한 결과를 발견할 수 없었다. 앞서 언급했듯 의미론적 일관성의 파악에 응집성 지수가 더 적절한 수치인 점에 더해 혼잡도 지수의 계산 결과가 유의미하게 나오지 않았기 때문에, 본 연구에서는 응집도 지수만을 근거로 LDA에서 도출해낼 토픽의 개수를 산정하였다. 그 결과 코로나19 이전의 시기는 14개의 토픽, 코로나19 이후의 시기는 12개의 토픽으로 분류하여 LDA 분석을 진행하였다.



Note: The coherence and perplexity scores are computed through the Gensim package in Python.

Figure 1. Coherence and perplexity scores for data before and after Covid19

Table 1. Topics of hate-related news and keywords before and after Covid19

2000-2019		
Topics	Topic number	Keywords
Politics	0	Busan, Daegu, Gyeonggi-do, stakeholder, the public power
	1	Voters, anti-Korean, voters, Hate, Realmeter
	4	Liberty Korea Party, Democratic Party, Politicians, The Blue House, swear word
	7	petition, bulletin board, The Blue House, Ferry Sewol, press conference
	12	Seoul, Seoul City, Korea, Incheon, Seoul Plaza
	13	Civil servants, consumers, networks, the Ministry of Environment, Yongin
Foreign conflicts	6	Japan, China, Ministry of Justice, Japanese, Community
	10	United States, United Kingdom, Germany, France, Europe
Immigrants	3	Korea, Foreigners, Workers, Human Rights Commission, Jobs
Male	5	Victims, Seoul, Gangnam Station, Sexual Violence, toilet
LGBT	2	Homosexuality, minorities, sexual minorities, homosexuals, queers
Hazardous Facilities	11	Residents, hate facilities, treatment plants, incinerators, committees
Japan	9	Saenuri Party, Japan, Japanese Military Sexual Slavery, Abe, Korea and Japan
Platforms	8	Online, malicious comments, netizen, Instagram, SNS
2020-2021		
Politics	0	GEF, Political Party, NSEDPD, Ministry of Gender Equality and Family, Gyeongnam
	6	Democratic Party, Yoon Seok-yeol, Democratic Party of Korea, Lee Jae-myung, politician
	11	By-election, Young People, Incheon, Floor Leader, Boys' Club
Gender Conflict	5	Online, The Blue House, misandry, petition, gs25
	9	Victims, sexual violence, sexual harassment, women, perpetrators
LGBT	3	Korea, Sim Sang Jung, queer, promotional material, identity
	7	Sex minorities, minorities, prohibitions, committees, chatbots
	8	Sex minorities, people, Ansan, homosexuality, transgender
AI	10	ai, luda, user, ScatterLab, United States
Covid19	1	Daegu, Residents, Shincheonji, Chungbuk, Gyeongbuk
	2	COVID-19, the U.S., COVID-19, China, Korea
	4	U.S., Korea, Asian, Covid-19, Victim

2.2 분석 결과

Table 1은 혐오 관련 기사에 대해 LDA 토픽 모델링을 수행한 결과다. 코로나19 전과 후의 데이터로부터 LDA를 통해 적정 수의 토픽과 각 토픽을 설명하는 상위 10개의 키워드를 각각 도출하였다. 연구자는 LDA를 통해서 각 토픽에 해당하는 단어의 분포와 단어별 토픽에 대한 기여도를 파악할 수 있다. LDA를 사용하는 연구들은 보편적으로 토픽별 단어 중 상위 5개 혹은 상위 10개를 사용하여 토픽을 해석하며, 사용되는 키워드의 수는 해석력에 따라 다양해질 수 있다(Omar et al., 2015). 이에 따라, 본 연구에서는 상위 10개의 키워드를 사용해 주제를 도출하였다. 상위 키워드를 기준으로 유사한 내용을 지닌 토픽들을 그룹화하여 코로나19 이전의 경우 8개, 이후의 경우 5개의 혐오 주제로 분류하였다. LDA를 포함한 토픽 모델링 기법을 활용하는 대부분의 선행 연구들은 연구자의 주관에 따라 토픽들을 분류한다(Hingmire et al., 2013; Kang, Lim, 2020; Porter, 2018).

본 연구에서도 역시 토픽 모델링을 활용하는 기존 연구의 학문적 관습에 따라서 각 혐오 주제를 연구자의 판단에 따라 명명하였다. 각 토픽의 키워드는 독립적으로 존재하는 것이 아니라 서로 연관

되어 맥락을 형성하고, 그 맥락에 따라 연구자는 토픽명을 정할 수 있다. 이를테면, Japan이라는 토픽의 ‘Saenuri Party, Japan, Japanese Military Sexual Slavery, Abe, Korea and Japan’이라는 키워드를 통해 위안부 문제 혹은 정당 관련된 일본 혐오 문제를 살펴볼 수 있다. Gender Conflict 토픽의 ‘Victims, sexual violence, sexual harassment, women, perpetrators’이라는 키워드는 성범죄에 따라 형성된 성별 갈등을 대표하는 단어들이며, Covid19 토픽의 ‘U.S., Korea, Asian, Covid-19, Victim’ 키워드는 코로나19 사태 이후 한국인 또는 동양인에 관한 혐오가 증가하는 현상을 설명한다. ‘Victim’이라는 단어는 Gender Conflict와 Covid19 주제에 동시에 등장하였지만, 각 주제에서 높은 기여도를 나타낸 다른 키워드가 무엇인지에 따라 상이하게 해석될 수 있다.

Table 1은 각 토픽을 설명하는 상위 10개의 키워드 중 토픽을 더 명료하게 표현할 수 있다고 판단되는 키워드 5개를 연구자가 선별한 것이다. LDA 분석을 통해 살펴본 코로나19 이전의 한국 사회의 혐오 이슈로는 정치, 외국의 혐오 문제, 이민자, 남성, LGBT, 혐오시설, 일본, 플랫폼상 혐오가 있었으며, 코로나19 이후로는 정치, 성별 갈등, LGBT, 인공지능, 코로나와 관련된 혐오 이슈들이 있었다.

Table 2. Random samples for topics of hate-related news

2000 ~ 2019	
Topics	Title
Politics	“Is President Moon Sane” Crossing the Line at Gwanghwamun Rally
	President Moon’s approval rating is 48.5 percent, man in their 20s being 29.4 percent lower
	Focusing on National Assembly Inspection/Three questions about the Government
Foreign conflicts	U.S. public broadcaster shakes by right-wing’s ‘hidden camera trap’
	WP: “Even the North elite are unhappy with the regime, and cynicism prevails.”
	Clement’s non-running, Sealing’s running. Legendary pitchers’ mixed political moves
Immigrants	Migrant Women appeals to Human Rights Commission for Mayor of Iksan’s discriminatory speech
	Korean University Students Killed by Group Assault During Russia Language Training
	Misunderstandings and truths about Yemeni refugees, such as ‘potential sex offenders’ and “IS terrorists”
Male	“Because of jealousy!” Webtoon writer’s confession about alleged sexual assault
	About 500 Women Participating for Demonstrations denouncing the violent punishment of drug sex crimes at Hyeonha Station
	While some driving “Seoul City University’s hidden camera incident” as self-fabrication, the Police investigated other crimes after booking the criminal
LGBT	“I wanted to let you know that there are sexual minorities nearby.”
	The Queer Festival “Opposition to sexual minority hate to total minority hate”
	Kim Cho Kwang-soo officially announced his first “Homosexual Marriage” in Korea, so no more hate!
Hazardous Facilities	A lot of tension for the relocation of hazardous facilities
	Yanggu-gun held a contest for a candidate site for a rural waste disposal site for hate facilities
	Hate facilities concentrated near IC, ‘Tie of jade’
Japan	Ministry of Culture, Sports and Tourism officially asks IOC to ban the use of the Rising Sun Flag at Tokyo Olympics
	Professor of Ewha Womans University’s being controversial for her remarks, “There’s no way that you shouldn’t be a Japanese Military Sexual Slavery,”
	Justice for the Comfort Women asserts immediate discipline for the students saying that “Japanese Military Sexual Slavery is a kind of a prostitution.”
Platforms	Public broadcasting and shamanistic drama(cultural walk)
	Kakao “Abolishes Entertainment News Comments” by Prioritizing Public Interest Value Over Company Revenue
	What Causes the End of the Online Community

Table 2. Random samples for topics of hate-related news(continued)

2020 ~ 2021	
Topics	Title
Politics	[Night Focus] "Suspicion of sexual misconduct" ahead of local elections
	"Jeong Jin-woong should be excluded from the job." spreading voices inside the prosecution
	[D-9 in the general election] The slip of the tongue means defeat, but the controversy over the United Party candidate belittling generation remains
Gender Conflicts	The aftermath of "Won Jong-gun Incident": Ruling party supporters will shout out 'Me Too Out'
	"While male police sleep 2~3 hours, Female police rest over the weekend ", reverse discrimination revealed by the National Police Agency
LGBT	GS25 deletes Controversal poster holding minsandry and apologizes
	The Democratic Party's wait for the anti-discrimination law is overshadowed by the pace of prosecution reform
	Joanne Rowling's fans are also "distancing." for the Transgender hate speech
AI	Justice Party "Education authorities should be ashamed of themselves about Transgender students giving up Admissions"
	Facebook's internal accusation: "Facebook chose money even after recognizing 'Teenager Suicide Impulse'"
	'Eruda' faces a retirement crisis in less than a month for sexual harassment controversy
Covid19	[Opening the morning] We should understand the brightness and darkness of intelligent information technology
	[Contribution] Overcome COVID-19 with Experts and Civil Society
	Corona Made in China. Tweets from Trump gave birth to Asian Hatred
	Economic Recession and Securing Vaccine: Korean politics obsessed with COVID-19

Table 2는 각각의 주제별 기사를 단순 무작위 추출 방법(Simple Random Sampling)을 통해 3개씩 도출한 것이다. 코로나 전후에 등장한 혐오 주제들을 살펴보았을 때, 소수자나 정당 갈등과 같이 보편적으로 등장하는 이슈들도 있었지만, 시기별 크게 화제가 된 사건들을 중심으로 혐오의 양상이 변화하는 것을 알 수 있다. 이와 같은 변화의 원인은 코로나19라기보다는 새로운 사건들이 발생할 때마다 해당 사건을 중심으로 다량의 뉴스 기사들이 작성되기 때문이라고 볼 수 있다.

정치 관련 혐오의 경우 2014년 발생한 세월호 사건, 2022년에 일어난 여성가족부(여가부) 폐지 및 전국장애인차별철폐연대(전장연) 시위 문체와 관련해 정치적 혐오가 형성되었다. 성별 갈등과 관련된 혐오의 경우 2019년 이전에는 강남역 살인 사건의 발발로 남성 혐오가 중심이 되어 성별 갈등이 진행되었지만, 2020년 이후에는 GS25 남성 혐오 사건 등의 영향으로 남성뿐만 아니라 여성에 대한 혐오의 모습 또한 등장했다. 코로나19 관련 혐오에 주목해 봤을 때 한국 내에서 특정 지역과 관련해 혐오의 양상이 드러나거나 국외에서 아시아인에 대한 혐오의 모습이 나타나는 것을 알 수 있었다. 국내 갈등의 경우 2020년 2월 신천지 대구교회를 기점으로 코로나19 확진자가 증가하는 사건이 발생해 대한민국 내부에서 지역과 종교에 대한 갈등이 함께 형성되었다.

LDA를 통해 한국 사회의 혐오 주제를 파악한 후, 시계열적으로 변화하는 혐오 담론의 지형을 살펴보기 위해서 2000년부터 2021년까지 월 단위로 각 혐오 주제를 다룬 기사 수를 분석하였다. 이어서, 전체 혐오 관련 기사 대비 특정 혐오를 주제로 한 기사가 차지하는 비중을 시기별로 분석했다. 기사별 혐오 주제의 파악은 LDA를 통해 구한 각 기사의 토픽 분포에서 가장 높은 비중을 차지한 토픽을 해당 기사의 혐오 주제로 규정하는 방식으로 이루어졌다. 코로나19 이전과 이후로 데이터셋을 나누어 LDA를 진행하였기 때문에 기사의 주제를 규정하고 시계열적 추이를 관측하는 과정은 코로나19 전·후 데이터에 대해 각각 따로 진행되었다.

코로나19 이전의 경우 2000년부터 2019년까지 20년간의 추이가 존재하므로 연도별 추이를 살펴 보았고, 코로나19 이후에는 연도 상으로 2020년과 2021년의 자료밖에 존재하지 않아 더 세부적으로 월별 추이를 살펴보았다. Figure 2는 본격적으로 혐오라는 주제가 언론에서 언급되기 시작한 2000년부터 2019년까지의 한국 사회 속 혐오 담론 주제별 기사 비율의 추이를 본 결과이다.

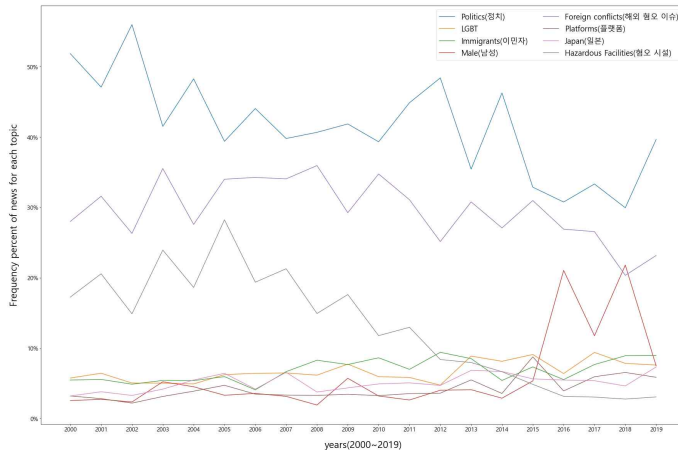


Figure 2. Frequency percent of hate-related news for each topic(2000~2019)

2000년부터 2019년까지 주제별 혐오 기사의 비율을 살펴보았을 때, 정치혐오 관련 기사는 모든 연도에서 혐오 주제 중 가장 높은 비중을 보였다. 혐오시설과 관련된 기사의 경우 2005년까지는 꾸준히 10% 이상을 유지하였으나, 2006년을 기점으로 급격히 기사의 비중이 줄어들기 시작하였다. 남성 혐오를 주제로 한 기사의 비율은 2015년부터 갑자기 높은 비율로 상승하였고, 2018년에 20%를 넘으며 큰 비중을 차지하였다.

2015년 전까지는 여성 운동계에서 성평등 의제를 정하면 정치권에서 이를 받아들이는 방식으로 여성운동이 전개되다가, 2015년 이후에는 젊은 여성 일반이 인터넷과 소셜 네트워크 서비스를 통해 의제를 발굴하고 활동하면서 강한 응집력과 폭발력을 드러냈다. 사회 전반적으로 성별 갈등의 이슈가 크게 논의됨에 따라, 뉴스 기사 속 해당 문제에 대한 언급량이 증가하였음을 알 수 있다.

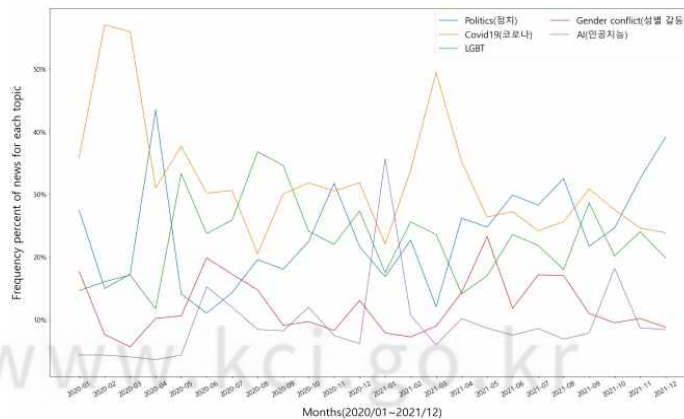


Figure 3. Frequency percent of hate-related news for each topic(2020~2021)

Figure 3은 2020년과 2021년의 한국 사회 속 혐오 관련 기사 비율의 추이를 월별로 그린 그래프이다. 2020년과 2021년의 월 단위 주제별 혐오 기사의 비율을 보면, 코로나19의 영향력 증대로 인해 전반적으로 코로나19 관련 기사가 높은 비중을 차지하고 있다. 코로나19의 추이 내에서도, 2020년 2-3월과 2021년 3월에 코로나 확진자가 급증하면서 동시에 기사의 수도 증가하는 모습을 보였다. 소수자 혐오의 경우 2020년 8월 네이버 웹툰 소수자 차별 논란, 지하철 성소수자 광고 훼손 논란과 관련해 비교적 높은 비율의 기사가 작성되었다. 인공지능 관련 혐오는 2021년 1월 이루다 논란과 관련하여 인공지능을 대상으로 한 성희롱이 이슈로 등장하며 주목받았다.

2000년부터 2021년까지 한국 사회 속 혐오 담론을 살펴본 결과 전반적으로 정치와 관련된 혐오는 지속해서 높은 관심을 받아왔지만, 성소수자, 이민자, 남성, 플랫폼 등 정치 외의 혐오들은 특정 사건의 발생에 따라 주제에 대한 주목도가 상승하는 모습을 보였다. 시대의 흐름에 따라 인공지능, 코로나 등의 새로운 혐오 주제가 등장하는 현상도 볼 수 있었으나 새로 등장한 주제 내에서도 기존에 등장했던 혐오의 유형이 관찰되었다. 인공지능 관련 혐오의 경우 성별 갈등을 중심으로 논란이 형성되기도 하였으며 코로나의 경우 지역 간 혐오의 모습이 나타났다.

3. 혐오 담론이 정책 결정 과정에 미치는 영향력에 대한 분석

앞서 한국의 혐오 담론 지형을 살펴보고 코로나 발발 이전과 이후의 혐오 담론과 혐오 주제별 주제어와 관련 기사들을 살펴보았다. 그렇다면 혐오 담론의 변화는 국회의 정책 결정 과정에 어떠한 영향을 주는가? 선행연구들은 지속적으로 정책 결정 과정(policy making process)의 중요성에 대해 역설하였으며 정책 결과(policy outcome)가 국민의 삶에 미치는 영향을 연구해왔다(Kuitto, Helmdag, 2021). 이에 더해, 혐오에 관한 기존 연구들은 특정 대상을 향한 혐오가 정책 결정권자들에게 미칠 수 있는 영향력과 그 사회적·정책적 악영향에 대한 우려를 표명하고 있으며, 혐오와 정책 결정의 관계성을 탐구하는 것의 중요성을 강조하고 있다. 그럼에도 불구하고, 혐오 담론이 정책 결정 과정에 미치는 영향력에 대한 면밀한 분석은 여전히 부족하다(Bulmer, 2021). 혐오가 정책 결정과 사회에 미치는 악영향에 대한 학문적 논의가 지속적으로 전개되어왔다는 점을 고려할 때, 이러한 연구의 부재는 예상 밖이다.

따라서 본 논문은 성별 갈등에 관한 기사 수와 여성 대상 폭력(Violence Against Women)과 관련한 국회의원들의 법안 발의 행태의 관계성을 중심으로 혐오 담론이 정책 결정 과정에 미치는 영향력을 분석한다. 현대 한국 사회에서 성별 갈등은 사회 영역에서만 아니라 정치 영역에서도 가장 중요한 이슈로 자리매김하였으며, 제20대 대통령 선거에서는 소셜 네트워크 서비스(Social Networking Service, SNS)와 언론을 중심으로 남녀 간 대립 구도가 격화되었다. 성별 갈등의 원인과 해결 방안 그리고 성별 갈등에 대한 국민들의 인식 등을 대상으로 다양한 연구들이 진행되었으나(Lee, Hong, Seol, 2019), 성별 갈등이 국회의원들의 법안 발의 행태에 미치는 영향은 심도 있게 연구되지 않았기에 본 논문에서는 이를 통계적 접근 방식을 통해 분석함으로써 선행연구에 기여하고자 한다. 이어서, 성별 갈등에 관한 기사 수와 여성 대상 폭력과 관련한 국회의원들의 법안 발의 행태의 관계성에 관한 이론 및 가설을 소개하려 한다.

3.1 이론 및 가설

우리나라의 법안 처리 절차는 국회의원 발의 법안에 한하여 상임위원회 심사, 법제사법위원회 체계·지구 심사, 그리고 본회의 심의·의결 과정을 포함한다. 심사와 심의·의결 과정에서 국회의원들은 여론에 민감하게 반응한다. 국회의원들은 언론을 통해 여론을 파악하며 특정 법안과 관련한 이슈에 대해 부정적 여론이 존재할 경우 해당 법안을 가결하는 데 부담을 갖게 된다(McCombs, Valenzuela, 2020). 국회의원들은 시간, 예산, 인적자원 등 한정된 자원을 가능한 한 효율적으로 사용하여 자신들의 재선 확률을 증진시키고자 한다. 개인별로 약간의 차이는 있으나, 대다수 의원은 위험 회피(Risk-aversion) 성향을 지니고 있다고 연구되어 왔다(Krehbiel, 1992).

따라서 남녀갈등이 언론 상에서 부각되어 해당 이슈의 현저성(Issue salience)이 증가한다면 국회의원들은 관련 법안의 발의에 자신들의 한정된 자원을 투자하지 않으리라 예측할 수 있다. 본 논문에서는 국회의원들의 발의 행태에 영향을 미칠 수 있는 나이와 성별, 정당 소속 등의 조건이 동일하였을 때, 성별 갈등과 관련한 혐오 기사와 국회의원들의 여성 대상 폭력에 대한 법안 발의 행태와의 관계성을 아래와 같이 예상한다.

가설 1: 다른 모든 조건이 동일할 경우(ceteris paribus), 남녀갈등과 관련한 혐오 기사의 증가는 국회의원들의 여성 대상 폭력에 대한 법안 발의 확률을 감소시킬 것이다.

본 논문에서는 혐오 기사의 수와 국회의원의 발의 행태 간의 직접적인 관계성뿐만 아니라, 성별에 따라 이러한 관계성이 다를 수 있음을 검증하고자 한다. 비록 여성 의원 간에도 정당 소속과 나이 혹은 당선 횟수 등의 개별적 특성에 따른 이질성(heterogeneity)이 존재하기는 하나, 선행연구들은 여성 의원의 경우 남성 의원들에 비해 여성 관련 의제에 대하여 적극적인 발의 행태를 보여준다는 점을 지속적으로 발견하고 있다(Woo, Ryu, 2022).

국회의원의 성별과 발의 행태와의 관계성에 관한 이론적 바탕은 여성의 기술적 대표성(women's descriptive representation)과 여성의 실질적 대표성(women's substantive representation)과의 연관관계에 대한 Pitkin(1967)과 Mansbridge(1999)의 연구에서 찾을 수 있다. Pitkin(1967)은 기술적 대표성을 대표자들이 대표하는 집단과 동일한 성별 혹은 인종 등의 특성을 가지고 있는 것이라 정의하였으며, 실질적 대표성을 실제로 해당 대표자들이 자신들이 대표하는 집단의 실질적 이익을 정책 결정 과정에서 대표하는 것이라 정의했다. Pitkin(1967)은 기술적 대표성의 증가가 실질적 대표성을 증진시킴에 있어서 필수조건이라 주장했다. 이에 더해, Mansbridge(1999)는 여성 의원들이 여성 시민들과 여성이라는 공유된 정체성(shared identity)을 가지고 있기 때문에 여성 의원들이 남성 의원들에 비해 여성 관련 의제에 더욱 적극적이라고 주장했다.

위와 같은 이론들과 선행연구들에 바탕을 두어 국회의원들의 나이, 정당 소속 등의 개별적 특성이 동일하였을 경우, 남성 의원이 여성 의원에 비해 성별 갈등과 관련한 혐오 기사의 증가에 민감하게 반응할 것이라 아래와 같이 예측한다.

가설 2: 다른 모든 조건이 동일할 경우(ceteris paribus), 남성 의원이 여성 의원에 비해 남녀갈등과 관련한 혐오 기사의 증가에 민감하게 반응할 것이다.

3.2 변수 설명

앞서 언급한 가설 1과 가설 2를 검증하기 위해 본 연구는 통계적 접근 방법을 활용한다. 통계분석 시 활용되는 종속변수는 여성 대상 폭력과 관련한 법안에 대한 개별 국회의원들의 발의 참여 여부다. 종속변수를 구축하기 위해 본 연구에서는 대한민국 국회의 의안정보시스템을 활용하여 2020년부터 2021년까지 발의된 법안 중 여성 대상 폭력과 관련한 272개의 법안과 이에 대한 개별 발의자(대표발의자와 공동발의자 모두 포함) 정보를 추출하였다. 여성 대상 폭력과 관련한 법안을 선정함에 있어, 본 연구는 법안 분류 과정에서 주로 사용되는 키워드 기반 분류방식(keywords-based classification)을 활용했다. 키워드로는 UN Women의 정의에 따라, 성추행, 성폭행, 성범죄, 스토킹, 데이트 폭력, 여성혐오 등을 선정하였다. 각각의 법안에 대하여 개별 국회의원들이 해당 법안에 참여하였다면 1, 그렇지 않다면 0을 종속변수에 할당하였다. 분석단위는 “개별 의원-개별 법안”이다.

가설 1을 검증하기 위한 독립변수로는 남녀갈등에 대한 혐오 기사의 수를 사용한다. 남녀갈등을 다루는 혐오 기사의 수는 LDA를 통해 기사별 토픽을 도출한 후 해당 주제를 가지고 있는 기사를 카운트하여 측정되었다. 가설 2를 검증하기 위해서는 국회의원의 성별을 이항 독립 변수(0=남성 의원, 1=여성 의원)를 구성하여 혐오 기사의 수와의 교차항(interaction term)을 활용한다.

본 연구에서는 통계분석 시 누락 변수에 의한 편의(omitted variable bias)를 최소화하기 위해, 선행연구들을 참고하여 의원들의 여성 대상 폭력과 관련한 법안 발의 행태에 영향을 미친다고 연구되어온 변인들을 통제한다. 개별 국회의원들의 당선 횟수(seniority) 혹은 나이가 발의 행태에 다양한 영향을 미친다는 연구들이 존재하기 때문에(Jung, 2019), 본 연구에서는 여성 대상 폭력과 관련한 법안 발의에 영향을 줄 수 있는 개별 의원들의 당선 횟수와 나이를 통제한다. 국회의원들의 나이가 발의 행태에 비선형적(Nonlinear) 영향을 미친다는 기존 연구결과에 의거하여(Bernstein, Wolak, 2002), 나이의 경우 일차항(Age)과 이차항(Age*Age)을 함께 통제변수로 포함해 혐오 담론이 국회의원들의 발의 행태에 미치는 독립적 영향력을 파악하고자 하였다.

국회의원들의 정당 소속 역시 통제변수 중 하나로써 통계분석 시 활용한다. 국회의원들의 정당 소속은 다양한 연구를 통해 그 영향력을 검증받았기에(Reingold et al., 2021), 본 논문에서도 국회의원들의 정당 소속은 국민의힘(1), 더불어민주당(2), 정의당(3), 무소속 및 기타 정당(4)로 명목 변수(nominal variable)를 활용하여 통제한다.

마지막으로 국회 내 그리고 정당 내에서의 개별 의원의 지위(leadership position) 또한 통계분석 시 통제변수로 포함한다. 개별 의원의 지위는 선행연구들에 의해서 의원들의 법안 발의를 포함한 의정 활동을 설명하는 변수로써 활발하게 연구되어 왔다. Kweon, Ryan(2021)은 국회의장의 경우 다른 의원들에 비해 법안 발의에 적극적이라는 점을 찾아냈다. 본 연구에서는 개별 의원의 지위를 상임위원장, 당 대표, 원내대표, 국회의장으로 구분하고 각각의 이항 변수를 만들어 개별 의원들이 각각의 지위에 있을 경우 1, 그렇지 않을 경우 0을 부여한다.

Table 3은 본 연구의 통계분석에서 활용되는 모든 변수의 기술 통계를 보여준다. 통계분석의 분석단위는 “개별 의원-개별 법안”이며, 총 관측치는 10,880이다. 다중공선성(multicollinearity)의 존재 여부를 검증하기 위해 분산 팽창 요인(Variance Inflation Factors, VIFs) 검사를 실시하였으며, 평균 분산 팽창 요인의 값과 독립변수와 통제변수들의 개별 VIF들은 4를 초과하지 않았다. 본 논문에서 사용되는 종속변수가 0과 1의 값을 가지는 이항 변수(binary variable)이기 때문에 선형회귀분석을 활용하기보다는 로지스틱 회귀분석을 진행한다(Jung, Yeon, 2017).

Table 3. Descriptive statistics of all variables

Variables	Obs	Mean	Std. dev.	Min	Max
Dependent Variable					
Sponsoring Bills concerning Violence Against Women	10,880	0.029	0.169	0 (Freq.: 10,560)	1 ((Freq.: 320))
Independent Variable					
Number of News Articles about Gender Conflict	10,880	46.735	9.031	36	74
Control Variables					
Gender	10,880	0.166	0.372	0 (Freq.: 9,078)	1 (Freq.: 1,802)
Seniority	10,880	2.475	1.341	1	8
Age	10,880	58.155	9.774	27	72
Age*Age	10,880	3884.309	1365.566	729	5184
Proportional Representation	10,880	0.156	0.363	0 (Freq.: 9,180)	1 (Freq.: 1,700)
Party Affiliation					
People's Power	10,880	0.519	0.5	0 (Freq.: 5,236)	1 (Freq.: 5,644)
Democratic Party	10,880	0.425	0.494	0 (Freq.: 6,256)	1 (Freq.: 4,624)
Justice Party	10,880	0.022	0.146	0 (Freq.: 10,642)	1 (Freq.: 238)
Etc.	10,880	0.034	0.182	0 (Freq.: 10,506)	1 (Freq.: 374)
Committee Chair	10,880	0.153	0.36	0 (Freq.: 9,214)	1 (Freq.: 1,666)
Party Leader	10,880	0.034	0.182	0 (Freq.: 10,506)	1 (Freq.: 374)
Floor Leader	10,880	0.05	0.218	0 (Freq.: 10,336)	1 (Freq.: 544)
Speaker	10,880	0.066	0.248	0 (Freq.: 10,166)	1 (Freq.: 714)

Note: Frequencies of binary categorical variables are provided in parentheses.

3.3 분석 결과

Table 4는 가설 1을 검증하기 위해 추정된 로지스틱 회귀분석 모형의 결과를 보여준다. 모델 1은 어떠한 통제변수가 포함되지 않고 독립변수인 성별 갈등에 관한 혐오 기사만이 포함되었으며, 모델 2에서는 의원들의 개별적 특성인 성별, 당선 횟수, 나이, 당선 방식 그리고 정당 소속이 통제변수로 포함되었다. 마지막으로 모델 3은 의회와 정당 내의 지위에 관한 통제변수들이 포함되었다.

본 표에서 독립변수인 “성별 갈등에 관한 혐오 기사”는 모델 1과 모델 3에서는 5%의 유의수준에서 통계적으로 유의미하며 모델 2에서는 1%의 유의수준에서 통계적으로 유의미하다. 또한 모든 모델에서 독립변수는 종속변수에 부(-)의 영향을 미친다. 모델 1에서 성별 갈등에 관한 혐오 기사의 베타계수(beta coefficient)는 -0.016으로 이는 해당 혐오 기사의 수가 1 증가하였을 때, 여성 대상 폭력과 관련한 법안 발의의 승산 비(odds ratio)를 0.016만큼 감소시킨다는 것을 의미한다. 추정된 강건 표준오차들에서 근소한 차이가 존재하나 모든 모델에서 유사한 통계 결과를 보여준다. 이러한 통계 결과는 성별 갈등에 관한 혐오 기사와 국회의원들의 여성 대상 폭력과 관련한 법안 발의의 부의 관계성에 관한 가설 1을 통계적으로 지지한다.

Table 4. Empirical results from logistic regression models

	Model 1	Model 2	Model 3
Independent Variable			
Number of News Articles about Gender Conflict	-0.016* (0.006)	-0.016** (0.006)	-0.016* (0.007)
Control Variables			
Gender		0.495*** (0.143)	0.497*** (0.147)
Seniority		-0.207*** (0.054)	-0.221*** (0.058)
Age		0.183*** (0.047)	0.175*** (0.051)
Age*Age		-0.001*** (0.000)	-0.001** (0.000)
Proportional Representation		-0.430* (0.180)	-0.418* (0.188)
Party Affiliation: People's Power(Baseline)		0 (.)	0 (.)
Democratic Party		0.762*** (0.124)	0.786*** (0.124)
Justice Party		0.658 (0.360)	0.581 (0.458)
Etc.		-0.611 (0.511)	-0.609 (0.514)
Committee Leader			0.123 (0.176)
Party Leader			0.197 (0.383)
Floor Leader			-0.141 (0.305)
Speaker			-0.295 (0.253)
Constant	-2.750*** (0.292)	-9.175*** (1.643)	-8.839*** (1.779)
N	10,880	10,880	10,880
AIC	2885.962	2828.45	2833.984
BIC	2900.552	2901.397	2936.109
Log Pseudolikelihood	-1440.981	-1404.225	-1402.992
Wald Chi-Squared	6.631	82.270	81.384
Prob > Chi-Squared	0.010	0.000	0.000

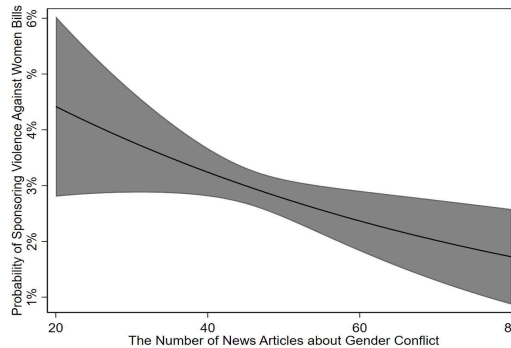
Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. The Akaike's Information Criterion(AIC) and the Bayesian Information Criterion(BIC) are presented for the model comparison.

통제변수들에서도 의미 있는 통계 결과를 살펴볼 수 있다. 선행연구들과 마찬가지로 본 연구에 서는 여성 의원들이 여성 대상 폭력과 관련한 법안 발의에 적극적으로 참여한다는 점을 발견했다. 이는 여성의 기술적 대표성과 실질적 대표성의 관계성에 추가적인 경험적 증거를 제시한다. 또한, 당선 횟수는 종속변수인 여성 대상 폭력과 관련한 법안 발의 여부와 음(-)의 상관관계를 갖는다. 또한, 개별 국회의원들의 나이와 나이*나이 역시 모델 2와 모델 3에서 통계적으로 유의미했다.

이어서, 소속 정당 또한 의원들의 발의 행태에 영향을 미친다는 점을 발견했다. 구체적으로 이 야기해서, 국민의 힘에 소속되어 있는 의원들에 비해서 더불어민주당 소속의 의원들이 여성 대상 폭력과 관련한 법안 발의에 적극적이라는 점이 드러났다. 마지막으로 국회의원들의 의회 내 그리고 정당 내의 지위는 여성 대상 폭력과 관련한 법안 발의 확률에 대해 통계적으로 유의미한 영향력을 행사하지 못한다는 점을 찾아냈다.

Table 4에 제시된 로지스틱 회귀분석 결과를 보다 심층적으로 분석하기 위해, 성별 갈등에 관한 혐오 기사 수에 따른 의원들의 여성 대상 폭력과 관련한 법안에 대한 예상 발의 확률을 추정하였다. Figure 4에 제시된 바와 같이 성별 갈등에 관한 혐오 기사 수가 증가할수록 법안 발의 확률이 감소한다. 이를테면, 성별 갈등에 관한 혐오 기사 수가 26개일 경우 발의 확률은 4%를 상회하지만, 해당 유형의 기사 수가 70개를 초과할 경우 점 추정(point estimation)된 예상 발의 확률은 2% 이하로 감소한다.

다시 말해서, 성별 갈등에 관한 혐오 기사의 수는 여성 대상 폭력과 관련한 법안 발의 확률에 대해 통계적 그리고 실질적으로 유의미한 부(-)의 상관관계를 보여준다. Table 4에서 통계적으로 유의미했던 통제변수들 또한 본문에서 제시되지 않았으나 예상 발의 확률의 변화를 고려하였을 때 여성 대상 폭력과 관련한 법안 발의 확률에 실질적으로 유의미한 영향력이 있는 것으로 나타났다. 이를테면, 여성 의원의 경우 남성 의원에 비해서 2% 높은 발의 확률을 보여주었으며, 5선 의원은 초선 의원에 비해서 2.3% 낮은 발의 확률을 보여주었다.



Note: The predicted probabilities are estimated with 95% confidence intervals. The predicted probabilities and confidence intervals are computed based on the linear predictors in Table 4, while all other variables are held at their representative values (for detailed information, please refer Xu, Long (2005)).

Figure 4. The predicted probability of sponsoring violence against women bills based on the number of news articles about gender conflict

이어서 가설 2를 검증하기 위해 성별 갈등에 관한 혐오 기사와 개별 국회의원들 성별의 교차항을 포함한 로지스틱 회귀분석 모형들을 추정하였다. Table 5와 같이 모델 4는 통제변수들을 포함하지 않은 모델이며, 모델 5와 모델 6은 각각 의원들의 개별적 특성과 의회 내 그리고 정당 내의 지위를 추가적으로 통제하였다.

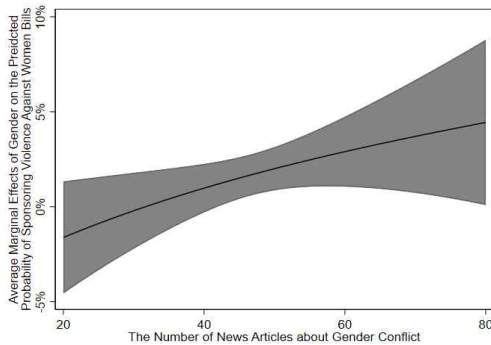
Table 5에 제시된 모든 모델에서 성별 갈등에 관한 혐오 기사와 성별의 교차항이 5%의 유의수준에서 통계적으로 유의미하며 음(-)의 상관관계를 가진다는 점을 발견했다. 이를 가설 2를 통해 예상된 바와 같이, 국회의원들의 성별에 따라서 성별 갈등에 관한 혐오 기사의 증가가 의원들의 여성 대상 폭력과 관련한 법안 발의 행태에 미치는 영향력이 달라진다는 점을 의미한다. 추가적으

로 통제변수들에 대한 결과 또한 추정된 베타계수와 강건표준오차에서 근소한 차이가 있으나 통계적으로 유사한 결과를 보여준다.

Table 5. Empirical results from logistic regression models with the interaction term

	Model 4	Model 5	Model 6
Independent Variable			
Number of News Articles about Gender Conflict	-0.026*** (0.007)	-0.026*** (0.007)	-0.026*** (0.007)
Gender	-1.160 (0.671)	-1.087 (0.677)	-1.086 (0.678)
Number of News Articles about Gender Conflict * Gender	0.034* (0.014)	0.034* (0.014)	0.034* (0.014)
Control Variables			
Seniority		-0.207*** (0.054)	-0.221*** (0.059)
Age		0.183*** (0.047)	0.175*** (0.049)
Age		-0.001*** (0.000)	-0.001*** (0.000)
Proportional Representation		-0.430* (0.180)	-0.418* (0.180)
Party Affiliation: People's Power(Baseline)		0 (.)	0 (.)
Democratic Party		0.762*** (0.124)	0.786*** (0.123)
Justice Party		0.658 (0.361)	0.581 (0.494)
Etc.		-0.612 (0.511)	-0.609 (0.509)
Committee Chair			0.123 (0.176)
Party Leader			0.197 (0.404)
Floor Leader			-0.141 (0.306)
Speaker			-0.295 (0.255)
Constant	-2.382*** (0.315)	-8.727*** (1.661)	-8.392*** (1.708)
N	10,880	10,880	10,880
AIC	2876.81	2825.841	2831.372
BIC	2905.989	2906.082	2940.793
Log Pseudolikelihood	-1434.405	-1401.92	-1400.686
Wald Chi-Squared	23.551	89.452	97.163
Prob > Chi-Squared	0.000	0.000	0.000

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. The Akaike's Information Criterion(AIC) and the Bayesian Information Criterion(BIC) are presented for the model comparison.



Note: The predicted probabilities are estimated with 95% confidence intervals. The predicted probabilities and confidence intervals are computed based on the linear predictors in Table 5, while all other variables are held at their representative values.

Figure 5. The average marginal effects of gender according to the number of news articles about gender conflict

통계적으로 유의미한 성별 갈등에 관한 혐오 기사와 개별 국회의원들의 성별의 교차 항의 실질적 의미에 대해 살펴보기 위해서, 성별 갈등에 관한 혐오 기사 수에 따른 성별의 평균 한계효과 (Average Marginal Effects, AME)의 변화를 Figure 5와 같이 추정하였다.

성별이 여성 대상 폭력과 관련한 법안 발의 확률에 미치는 평균 한계효과는 성별 갈등에 관한 혐오 기사의 수가 42개 미만일 경우 95%의 신뢰구간이 0%를 포함한다. 즉, 기사의 수가 42개 미만일 경우에는 성별 갈등에 관한 혐오 기사 수에 따른 성별의 평균 한계효과가 통계적으로 유의미한 차이를 보이지 않는다는 점을 의미한다.

그러나 성별 갈등에 관한 혐오 기사의 수가 42개 이상일 경우 신뢰구간이 0%를 포함하지 않으며, 기사의 수가 증가할수록 성별의 평균 한계효과가 증가하는 점을 확인할 수 있다. 성별 갈등에 관한 혐오 기사의 수가 42개일 경우 성별의 평균 한계효과는 1.196%로 여성 의원이 남성 의원에 비해 여성 대상 폭력과 관련한 법안 발의 확률이 1.196% 높다는 점을 의미한다. 기사의 수가 70개일 경우 성별의 평균 한계효과는 3.702%까지 증가한다. 즉, 여성 의원과 남성 의원의 여성 대상 폭력과 관련한 법안 발의 확률의 차이가 혐오 기사의 수가 증가할수록 커진다는 점을 의미하며 가설 2를 실증적으로 뒷받침한다.

4. 결론

본 연구에서는 2000년부터 2021년까지 한국 사회의 혐오 담론을 LDA를 활용한 텍스트 분석을 통해 살펴본 후, 혐오 담론의 정책 결정 과정에서의 영향력에 대해 남녀갈등을 중심으로 통계분석을 진행하였다. LDA 분석 결과, 코로나 이전에는 정치, LGBT, 이민자, 남성, 플랫폼, 일본, 혐오시설 관련 혐오 주제가 주를 이뤘다. 코로나 이후로는 정치, 코로나, LGBT, 인공지능, 성별 갈등 관련 혐오 이슈들이 등장했다. 혐오 주제별 상위 키워드와 주제별 단순 무작위 표본을 고려한 결과 사회 속에서 주목받는 사건들을 중심으로 관련 혐오 주제들이 부각되는 것을 알 수 있었다. 이에 더하여, 혐오 담론이 정책 결정 과정에 미치는 영향력을 로지스틱 회귀분석 모형을 통해 통계적으로 분석함으로써 남녀갈등에 관한 혐오 기사의 수가 증가할수록 국회의원들의 여성 대상 폭력과 관련한 법안에 대한 발의 확률이 감소한다는 점을 발견하였다. 또한, 성별과 혐오 기사 수와의 교

차항을 활용하여 여성 의원들이 남성 의원들에 비해서 남녀갈등에 관한 혐오 기사의 수에 영향을 덜 받는다는 점을 검증하였다.

본 연구의 시사점은 다음과 같다. 첫째, 본 연구는 21세기부터 중요한 논의의 대상으로 설정된 혐오라는 개념이 20년의 기간 동안 어떤 모습으로 사회 속에 존재해왔는지 살펴봄으로써 우리나라의 혐오 담론 지형을 파악했다. 21세기부터 중요한 논의의 대상으로 설정된 혐오라는 개념이 20년의 기간 동안 어떤 모습으로 사회 속에 존재해왔는지 살펴보는 것은 혐오라는 갈등 의제의 해결책을 마련하는 데 필요한 과정이라는 점에서 본 논문의 학문적 의의가 있다.

둘째, 본 연구는 정책 결정 과정에서 의원들의 발의 행태에 영향을 미치는 추가적 요인을 발굴했다. 그동안 언론이 의회 내에서의 의제 설정에 미치는 영향력에 관해서는 많은 연구가 이뤄졌으나(Barberá et al., 2019), 언론이 형성하는 혐오 담론이 정책 결정 과정에 미치는 영향력에 대한 분석은 부족하였다. 본 연구의 결과는 의원들의 발의 행태에 대한 설명력을 향상시키고 동시에 구체적으로, 여성의 기술적 대표성과 실질적 대표성에 실증적 증거를 밝혀내어 여성 의원들의 발의 행태를 한층 더 깊이 조망하였다.

셋째, 본 연구는 혐오 이슈를 다루는 언론의 중요성에 대해 다시 한번 강조한다. 언론의 의제 설정 기능은 사회에서 일어나는 여러 현상 중 일부를 공중의제로 대두시키는 임무를 수행하며 대중의 주의를 집중시킬 뿐만 아니라 정책적 측면에서 영향력을 행사하기도 한다. 본 연구에서는 성별 갈등에 대한 혐오에 초점을 두고 성별 갈등에 관한 혐오 기사 수가 증가할수록 여성 대상 폭력과 관련한 법안 발의 확률이 감소함을 밝혀냈다. 또한 언론 상에서 부각된 혐오 이슈와 관련된 법안 발의에 유의미한 영향을 줄 수 있는 변수를 분석함으로써 그동안 질적으로 연구되었던 혐오 이슈의 정치, 사회 영역에서의 영향력에 대한 이해의 폭을 넓혔다.

본 연구에는 다음과 같은 한계가 존재한다. 첫째, 본 연구는 한국 사회의 혐오 담론을 파악하기 위해 2000년부터 2021년까지의 뉴스 빅데이터를 분석하였다. 그러나 뉴스 빅데이터만을 통해서만 우리나라의 혐오 담론을 파악하는 데 한계가 있기 때문에 향후 연구에서는 뉴스 빅데이터뿐만 아니라 소셜 네트워크 서비스 등 보다 대중과 가까운 매체에서 형성된 혐오로 논의를 확장하여 살펴볼 필요성이 있다. 둘째, 혐오 담론이 정책 결정 과정에 미치는 영향력을 본 논문에서는 로지스틱 회귀분석 모형을 사용한 통계적 접근 방법을 통해 분석하였다. 하지만 통계적 접근 방법은 가설만을 검증할 뿐 상관관계를 뒷받침하는 이론 자체를 검증하지는 않는다. 따라서 인터뷰와 사례연구 등의 질적 연구 방법론을 활용하여 혐오 담론이 정책결정 과정에 영향력을 미치는 메커니즘(Mechanism)을 심도 있게 분석한다면 학문적 의의가 있을 것이다.

References

- Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., Tucker, J. A. (2019). Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4), 883-901. DOI: <https://doi.org/10.1017/S0003055419000352>
- Bernstein, J. L., Wolak, J. (2002). A bicameral perspective on legislative retirement: The case of the Senate. *Political Research Quarterly*, 55(2), 375-390. DOI: <https://doi.org/10.1177/106591290205500205>
- Bilewicz, M., Soral, W., Marchlewska, M., Winiewski, M. (2017). When authoritarians confront prejudice. Differential effects of SDO and RWA on support for hate speech prohibition. *Political Psychology*, 38(1), 87-99. DOI: <https://doi.org/10.1111/pops.12313>

- Bulmer, M. I. (2021). *Social science and social policy*. Routledge, New York.
- Harel, T. O., Jameson, J. K., Maoz, I. (2020). The normalization of hatred: Identity, affective polarization, and dehumanization on Facebook in the context of intractable political conflict, *Social Media+ Society*, 6(2), 1-10. DOI: <https://doi.org/10.1177/2056305120913983>
- Hingmire, S., Chougule, S., Palshikar, G. K., Chakraborti, S. (2013). Document classification by topic labeling. *The 36th international ACM SIGIR conference on Research and development in information retrieval*, 877-880. DOI: <https://doi.org/10.1145/2484028.2484140>
- Jung, H. O. (2019). The Analysis of Legislation Initiated by Congresswomen in the 17th-20th Assembly of South Korea, *Journal of Future Politics*, 9(1), 61-86. DOI: <https://doi.org/10.20973/jofp.2019.9.1.61>
- Jung, Y. H., Yeon, K. P. (2017). On the Relative Importance of Predictors in Logistic Regression. *Journal of The Korean Data Analysis Society*, 19(1), 141-150. DOI: <https://doi.org/10.37727/jkdas.2017.19.1.141>
- Kang, H. A., Lim, H. S. (2021). A Study on Search Query Topics and Types using Topic Modeling and Principal Components Analysis. *KIPS Transactions on Software and Data Engineering*, 10(6), 223-234. DOI: <https://doi.org/10.3745/KTSDE.2021.10.6.223>
- Kang, H. A., Lim, H. S. (2020). A Study on the Types of Online Shopping Queries using Topic Modeling and Principal Components Analysis. *The Korea Information Processing Society Conference*, 27(2), 765-768. DOI: <https://doi.org/10.3745/PKIPS.y2020m11a.765>
- Krehbiel, K. (1992). *Information and legislative organization*. University of Michigan Press, Michigan.
- Kuitto, K., Helmdag, J. (2021). Extending working lives: How policies shape retirement and labour market participation of older workers. *Social Policy & Administration*, 55(3), 423-439. DOI: <https://doi.org/10.1111/spol.12717>
- Kweon, Y., Ryan, J. M. (2021). Electoral Systems and the Substantive Representation of Marginalized Groups: Evidence from Women's Issue Bills in South Korea, *Political Research Quarterly*, Online First. DOI: <https://doi.org/10.1177/10659129211028290>
- Lee, J. I., Hong, J. H., Seol, J. A. (2019). Gender Politics and Meaning Practice of “#MeToo” Movement on the Social Networking Site. *Media, Gender & Culture*, 34(2), 99-146. DOI: <https://doi.org/10.38196/mgc.2019.06.34.2.99>
- Lee, M. Y., Lee, M. K., Yaung, H., Lee, T. R., Lee, S. M. (2022). Academic Hatred: Focusing on the Influence of a Supportive Classroom Climate, *Sage open*, 12(2), 1-12. DOI: <https://doi.org/10.1177/21582440221094820>
- Lee, Y. D., Park, H. J., Cho, Y. S. (2021). A Study on USA Press Cover Joe Biden” Using Text Mining Analysis. *Journal of The Korean Data Analysis Society*, 23(2), 611-623. DOI: <https://doi.org/10.37727/jkdas.2021.23.2.611>
- Li, W., Zhao, J. (2016). TextRank algorithm by exploiting Wikipedia for short text keywords extraction. *The 2016 3rd International Conference on Information Science and Control Engineering*, 683-686. DOI: <https://doi.org/10.1109/ICISCE.2016.151>
- Mansbridge, J. (1999). Should Blacks Represent Blacks and Women Represent Women? A Contingent 'Yes', *Journal of Politics*, 61(3), 628-657. DOI: <https://doi.org/10.2307/2647821>
- McCombs, M., Valenzuela, S. (2020). *Setting the agenda: Mass media and public opinion*. John Wiley Sons, New Jersey.
- Mimmo, D., Wallach, H., Talley, E., Leenders, M., McCallum, A. (2011). Optimizing semantic coherence in topic models. *The 2011 conference on empirical methods in natural language processing*, 262-272.
- Mun, G. (2021). Evaluating LDA and BTM Topic Models for Short Texts. *Journal of The Korean Data Analysis Society*, 23(3), 1313-1326. DOI: <https://doi.org/10.37727/jkdas.2021.23.3.1313>
- Omar, M., On, B. W., Lee, I. G., Choi, G. S. (2015). LDA topics: Representation and evaluation. *Journal of Information Science*, 41(5), 662-675. DOI: <https://doi.org/10.1177/0165551515587839>
- Pitkin, H. F. (1967). *The concept of representation*. University of California Press, California.
- Porter, K. (2018). Analyzing the DarkNetMarkets subreddit for evolutions of tools and trends using LDA topic modeling. *Digital Investigation*, 26, 87-97. DOI: <https://doi.org/10.1016/j.diin.2018.04.023>

- Reingold, B., Kreitzer, R., Osborn, T., Swers, M. L. (2021). Anti-abortion Policymaking and Women's Representation, *Political Research Quarterly*, 74(2), 403-420. DOI: <https://doi.org/10.1177/1065912920903381>
- Son, D. I. (2020). A Study on the Expressions Used in COVID-19 News - Focusing on Fear and Hate Reflected in Headlines, *Journal of Ewha Korean Language and Literature*, 51(2), 137-166. DOI: <https://doi.org/10.29190/JEKLL.2020.51.137>
- Son, H. J. (2018). Seven Years of the Discourse of Hatred. *Cultural Science*, 93, 20-49.
- Suh, H. (2020). Topic Modeling and Visualization of Uijeongbu City SNS data using LDA & Word2vec Methodology. *Journal of The Korean Data Analysis Society*, 22(6), 2391-2403. DOI: <https://doi.org/10.37727/jkdas.2022.24.2.585>
- Xu, J., Long, J. S. (2005). Using the delta method to construct confidence intervals for predicted probabilities, rates, and discrete changes. *The Stata Journal*, 5(4), 537-559.
- Woo, B. D. and Ryu, K. E. (2022). The nonlinear impact of women's descriptive representation: an empirical study on the ratification of women rights treaties. *Contemporary Politics*, 1-27. DOI: <https://doi.org/10.1080/13569775.2022.2088090>

Analyzing the discourse of hatred and its relationship with policymaking process using topic modeling and logistic regression

Jiin Lee¹, Byung-Deuk Woo², Mingu Lee³, Se-yoon Park⁴

Abstract

How the discourse of hatred in South Korea has changed and what is the effect of the discourse on the policymaking process? This article empirically examines the changes in the discourse of hatred in South Korea from 2000 to 2021. Analyzing news big data based on the Latent Dirichlet Allocation(LDA) models, we found that the hatred related to politics has consistently received great attention. On the contrary, the salience of non-political hatred such as the hatred against LGBT, immigrants, men, and platforms fluctuates according to the occurrence of specific events. With the passage of time, new hatred topics including AI and COVID-19 emerge, but the existing types of hatred are observed even within the new themes. In addition, this article using logistic regression models demonstrates that there is a statistically and substantively significant relationship between the number of news articles concerning hatred against gender conflict and the probability of sponsoring bills about violence against women. Also, this article finds that male legislators compared to their female counterparts are more sensitive to the number of news articles about hatred.

Keywords : Discourse of Hatred, Text Analysis, Latent Dirichlet Allocation, Policymaking Process, Logistic Regression.

¹Graduate Student, Dept. of Social Data Science, Pohang University of Science and Technology, 77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Korea. E-mail: jiinlee0017@postech.ac.kr

²(Corresponding Author) Post-Doctoral Researcher, The Institute of Social Data Science, Pohang University of Science and Technology, 77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Korea. Email: byungdeukwoo@postech.ac.kr

³Graduate Student, Dept. of Social Data Science, Pohang University of Science and Technology, 77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Korea. E-mail: mglee815@postech.ac.kr

⁴Graduate Student, Dept. of Social Data Science, Pohang University of Science and Technology, 77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Korea. E-mail: seyoopark@postech.ac.kr