

연구논문

텍스트 빅데이터 분석을 통한 한국인의 불공정 경험 분석: 국민청원 게시판 데이터 분석 결과를 중심으로*

양혜진** · 안정민*** · 이태현****

본 연구는 토픽 모델링과 질적 텍스트 분석을 통합적으로 활용하여 청와대의 국민청원 게시판 텍스트 데이터를 분석함으로써 한국인의 불공정 경험과 인식을 탐색하고자 하였다. 전체 청원 중 하루에 적어도 30개 이상의 동의를 받은 6,084개를 대상으로, 토픽 모델링 방법 중 하나인 잠재 디리클레 할당(Latent Dirichlet Allocation; LDA)을 활용하여 주요 토픽을 추출하고 각 토픽별 대표문서를 추출하였다. 이 과정에서 양적 기준뿐만 아니라, 질적 분석결과를 통합적으로 반영하여 최종적으로 28개의 토픽을 결정하였다. LDA를 통해 얻은 토픽별 출현확률이 높았던 단어 15개와 토픽의 언급 순위와 대표문서를 질적 분석한 내용에 따라 한국인의 불공정 경험 및 인식에 대한 몇 가지 특성을 발견하였다. 본 연구는 효과적인 텍스트 빅데이터 분석을 위해 토픽 모델링과 질적 분석을 유기적으로 통합한 혼합 방법을 사용하였다는 점, 분석 결과를 바탕으로 한국인의 공정성 경험 및 인식의 특성과 후속 연구의 방향을 제안하였다는 점에서 의의를 가진다.

주제어: 텍스트 빅데이터, 텍스트 마이닝, 토픽 모델링, 잠재 디리클레 할당, 질적 연구, 혼합연구방법, 국민청원

* 이 논문은 2018년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음.

** 중앙대학교 심리학과 석사과정(citron0403@cau.ac.kr), 제1저자.

*** 중앙대학교 심리학과 박사과정(dudmini_sa@naver.com), 제2저자.

**** 중앙대학교 심리학과 교수(lee0267@cau.ac.kr), 교신저자.

I. 서론

공정에 대한 한국 사회의 관심은 매우 뜨겁다. 베스트셀러나 정부 정책에서도 공정과 정의라는 단어가 낫설지 않으며, ‘정의로운 대한민국’ 구현을 국가 비전으로 삼고 청와대에 국민청원 게시판을 개설할 만큼 공정과 정의에 대한 사회적 관심이 높다. 그러나 이러한 관심에도 불구하고, 많은 한국인은 한국 사회가 불공정하다고 인식한다. 한 여론조사에 따르면, 한국인의 71%가 ‘한국 사회가 불공정하다’고 여기는 것으로 나타났다(문화일보 2019).

대다수 한국인이 한국 사회가 공정하지 않다고 여기고 있으나, 구체적으로 무엇이 과연 불공정한가에 대해서는 개인마다 의견이 엇갈린다. 또한, 공정사회 구현을 위해 어떤 정책적 조치와 사회적 조치가 필요한가에 대해서도 일반 국민뿐만 아니라 전문가들 사이에서조차 견해가 엇갈린다. 이는 공정성 개념에 대한 정의(定意)와 사회적 불공정성의 원인에 대한 이해가 개인마다 각기 다르기 때문이다(김상욱 외 2013; 장상수·김상욱·신승배 2015; 이종수 2013).

본 연구에서는 청와대 국민청원 게시글을 텍스트 빅데이터로 활용하여 분석함으로써 한국인이 불공정을 실제로 어떠한 상황 혹은 조건에서 어떠한 방식으로 경험하고 있으며, 나아가 공정이라는 개념을 어떻게 인식하고 있는지를 경험적으로 분석하고자 한다. 사회의 신뢰와 사회 안정성 증진의 주요 요인으로서 공정성 지각이 매우 중요하다는 점에서(Putnam 1993; 장상수 외 2015), 본 연구에서 도출된 한국인의 (불)공정 경험 양상 및 인식에 대한 경험적 결과는 ‘정의로운 대한민국’을 실현하기 위한 방안을 모색하기 위한 중요한 기초 자료가 될 것으로 기대한다.

텍스트 빅데이터를 분석하기 위해서는 토픽 모델링(topic modeling) 기법 중 하나인 잠재 디리클레 할당(Latent Dirichlet Allocation; 이하 LDA) 기법을 활용하였고, LDA 분석을 통해 텍스트 빅데이터의 기저에 잠재되어 있는 토픽을 추출하였다. 본 연구의 중요한 목적 중 하나는 텍스트 빅데이터로부터 주요 토픽을 추출하는 것에 그치지 않고 한국인의 불공정 경험과 인식에 대한 보다 심층적인 이해를 얻고자 하는 데 있으므로, 추출된 토픽을 대표하는 문서(document)를 대상으로 질적 분석을 동시에 진행하였다. 이를 통해 대한민국의 문화에 따라 독특하게 나타나

는 정의에 대한 규칙이나 기준, 정의의 지각 등에 대한 보다 심층적인 이해를 얻고자 하였다(Fischer 2016). 즉, 본 연구에서는 양적 접근법인 텍스트 빅데이터 분석과 질적 분석을 결합하고 상호 보완적으로 활용하는 혼합 방법 접근(크레스웰 2015)을 통해 한국인의 불공정 경험의 맥락과 인식을 탐색하고자 하였다.

1. 공정성과 정의

공정성과 정의의 개념을 구분하기란 쉽지 않다. 사전적 의미를 살펴보면, 공정(公正)은 ‘공평하고 올바름’이라는 뜻으로, 영어로는 fairness 또는 justice로 번역된다. 정의(正義)는 ‘진리에 맞는 올바른 도리, 개인 간의 올바른 도리 또는 사회를 구성하고 유지하는 공정한 도리’로 풀이되며, 영어로는 justice로 번역된다. 이를 볼 때, 공정은 정의롭고 공평한 것을 포괄하는 단어라 할 수 있으며, 공정과 정의는 떼려야 뗄 수 없는 관계에 있다(황혜신 2011). 공정성과 정의라는 용어는 분과학문에 따라 다른 용어를 사용하는 데 따른 차이일 뿐,¹⁾ 그 본질적 의미에는 큰 차이가 없다(김상욱 외 2013).

공정성 연구는 ‘정의란 무엇인가’에 대한 철학적, 윤리적 논의에서 출발했다. 플라톤, 아리스토텔레스, 롤스 등 정치철학자들의 정의론(正義論)을 바탕으로 공정성의 유형을 구분할 수 있다(Sabbagh & Schmitt 2016). 먼저, 분배적 정의는 자원 분배를 통제하는 원칙과 규칙에 대한 지각된 정의 및 기대되는 결과 대비 실제 분배 결과에 대하여 지각된 정의이다. 자원 분배의 원칙에는 개인의 시간, 노력, 기술 등 자원을 투입한 만큼 결과를 보상받는 형평(equity)의 원칙, 모든 사람이 동등하게 자원을 나눠 갖는 평등(equality)의 원칙, 모든 사람이 최소한의 욕구를 충족할 수 있도록 자원을 나누는 필요(needs)의 원칙, 자신의 지위 특성(젠더, 교육수준, 직업 등)에 따라 이익 또는 책임을 나누어 갖는 자격(entitlement)의 원칙이 있다. 둘째로 절차적 정의란, 자원 분배의 절차 및 과정에 대하여 지각된 정의다. 예컨대, 의사결정 과정에서 발언 기회 유무, 결과에 영향을 미칠 수 있는 정도, 자신의 권리와 의견이 존중받는 정도에 따라 공정성 지각이 달라질 수 있다. 셋째로 응보적 정의란,

1) 철학, 사회학, 심리학 등 다양한 학문 분야에서 정의와 공정을 혼용하여 사용한다. 학파에 따라 공정은 정의의 원칙이 도출되는 절차와 과정의 의미를 담기도 한다. 분배적 정의와 관련하여 형평(equity), 공평(equality) 용어가 공정과 혼용되기도 한다(사회통합위원회 2012).

범죄자의 잘못에 내리는 처벌에 대한 공정성 지각으로, 범죄자에게 합당한 처벌이 이루어졌는가, 피해자에게 합당한 배상이 이루어졌는가를 다룬다. 마지막으로 회복적 정의란, 범죄로 발생한 피해를 피해자, 가해자, 지역 사회가 함께 모여 이를 치유 또는 회복하는 비공식적 과정에 대한 공정성 지각으로, 사과하기, 커뮤니티 서비스, 심리상담 등이 이러한 비공식적 과정에 포함된다(Sabbagh & Schmitt 2016; Liebig & Sauer 2016).

이와 같은 사회적 정의에 대한 철학적, 윤리적 논의를 바탕으로 오늘날 사회학, 심리학, 정치학 등 여러 사회과학 분야에서 공정성에 대한 다양한 경험적 연구가 이뤄지고 있다. 공정성의 준거(결과, 과정 등), 공정성의 대상(여성, 장애인, 정책, 정부기관 등), 공정성의 맥락(소득, 사회적 기회, 조직생활 등)과 같이 개인의 주관적인 공정성 지각에 영향을 미치는 다양한 변인을 중심으로 연구가 이루어지고 있다(Sabbagh & Schmitt 2016; 김상욱 외 2013). 구체적으로 사회학적, 정치학적 연구는 거시적 관점에서 구조적 조건과 정치 이데올로기 등이 개인의 정의 인식에 미치는 영향을 주로 분석하며(Liebig & Sauer 2016), 심리학적 연구들은 정의와 관련된 개인의 인지, 정서, 행동에 영향을 미치는 상황 변인에 초점을 맞춘다(Gollwizer & Prooijen 2016).

석현호 외(2005)와 정한울(2018) 등의 연구에 따르면, 한국인은 결과의 평등보다 권리와 조건, 기회의 평등을 지지하는 경향을 보인다. 임금 분배의 경우, 산술적 평등보다 기여도에 따른 형평의 원칙을 선호하며, 보상의 기준으로 성과나 능력보다는 노력을 중시하는 경향이 있다(석현호 외 2005; 정한울 2018). 정부에 대하여 절차적 공정성을 낮게 인식하며, 조직에 대한 공정성은 인사정책에 관한 주요 결정 즉, 절차적 공정성을 낮게 인식하는 것으로 나타났다. 또 형평의 기준이 강화됐을 때 분배 공정성이 양호한 것으로 인식하였으나, 제도 변화로 인한 불이익자를 위한 보완책 마련은 미흡한 것으로 인식하고 있었다(석현호 외 2005). 이러한 연구는 한국인의 공정성 인식과 관련하여 중요한 정보를 제공하고 있으나, 그 대상이 조직 공정성 또는 정부에 대한 공정성 인식에 국한되어 있으며 설문조사를 통해 자료를 수집함으로써 불공정 경험의 구체적 맥락을 파악하기 어렵다는 한계가 있다. 이에 본 연구는 한국인이 공정성과 관련하여 어떠한 경험을 하고, 정의와 공정을 어떻게 인식하고 있는지 탐색하고자 하였으며, 이를 위해 청와대 국민청원 게시글을 텍스트 빅데이터 분석의 자료로 활용하였다.

2. 텍스트 빅데이터: 국민청원 게시판

청와대 국민청원 게시판은 ‘국민의 나라, 정의로운 대한민국’을 국가 비전으로 표방하는 정부(경향신문 2017; 중앙일보 2017)가 국민과의 소통을 위해 마련한 플랫폼으로, 처음 개설된 2017년 8월 19일부터 2020년 7월 31일을 기준으로 총 게시글이 87만 8,690건²⁾일 만큼 많은 국민이 자발적으로 이용하였다. 이후 청원 게시판은 ‘많은 사람이 다양한 주제에 대해 자신의 경험과 생각을 정부 및 국민과 직접 공유할 수 있는 최초의 공론장’이며, ‘사회적 약자를 위한 호소가 전달되는 통로’ 역할을 하고 있다는 평가를 받았다(중앙일보 2018). 또 한 조사에 따르면, 다수의 국민(64.2%)이 청와대 국민청원 제도가 사회문제 해결에 도움이 된다고 하였으며, 과반(55.3%)이 청원제도의 장점이 단점보다 더 많다고 하였다(KBS 뉴스 2018).

현재 국내에서 청와대 국민청원제도를 다룬 연구는 많지 않으나(한정훈 2020), 해외 문헌에 따르면 불공정성 지각은 청원 참여에 영향을 미친다. 미국과 영국의 참여자 대상 연구에서, 지각된 불공정성이 분노를 통하여 온라인 청원을 지지할 의도를 증가시켰으며(Antonetti & Manika 2017), 독일과 스위스인을 대상으로 진행된 연구에서는, 정책에 대한 낮은 불공정성 지각은 청원 참여 및 정책 반대 시위에 참여할 가능성에 부적 영향을 미쳤으나, 높은 불공정성 지각은 청원 참여 및 정책 반대 시위에 참여할 가능성에 정적 영향을 미쳤다(Liebe, Preisendorfer, & Enzler 2020). 이러한 연구결과와 국민청원 게시판에 대한 시민들의 평가를 볼 때, 청와대 청원 게시판을 한국인의 불공정 경험 및 인식을 수집할 수 있는 자료로 활용할 수 있을 것으로 판단된다.

국민청원의 경우 큰 제한 없이 누구나 글을 쓸 수 있다는 점에서 기존의 조사에서 참여할 기회가 없던 사람들도 자신의 경험을 공유할 수 있다는 장점이 있다. 또한 총 17개의 큰 범주로 구성되어 있지만 주제의 제한이 사실상 없기 때문에 다양한 의제들이 게시글로 올라오고 있다(송준모·박영득 2019). 청원 게시판과 같은 온라인 환경은 인간의 행동과 사회적 상호작용이 자연스럽게 드러나는 공간으로서 개인, 집단 및 사회를 이해할 수 있는 중요한 자료를 제공한다(Kosinski, Wang, Lakkaraju, & Leskovec 2016). 즉, 설문지, 면담, 초점집단 면접과 같은 전통적인 조사 연구에 비하여 사람들의 실제 경험이 보다 생생하게 묘사된 자료를 얻을 가능성이 더 크다

2) <https://www1.president.go.kr/petitions/3rd#firstPage/9>.

(Karami, Swan, White, & Ford 2019). 이에 더해, 해당 게시글에 대한 동의 수를 통해 얼마나 많은 시민이 게시글의 내용에 공감하고 있으며, 타당하다고 여기는지도 함께 파악할 수 있다.

한편 국민이 제한 없이 자신의 의견을 표현하고 사회 이슈를 반영할 수 있다는 장점에도 불구하고, 청원 게시판에 지나치게 개인적인 내용이 게시된다거나, 진실 왜곡 또는 과장, 혐오를 확산시킨다는 비판이 있는 것도 사실이다(한정훈 2020). 실제 상술한 조사에서 33.2%의 시민들은 이와 같은 청원 게시판의 역기능을 우려하였다(KBS 뉴스 2018). 따라서 본 연구에서는 청원 게시판의 이러한 한계를 보완하기 위하여 충분한 동의 수를 받은 문서를 먼저 선별하였다.

3. 분석방법: 토픽 모델링과 질적 분석의 혼합

본 연구에서는 먼저 토픽 모델링 기법을 활용하여 모든 청원글을 직접 읽고 전반적인 내용을 빠르게 파악하고자 하였다. 토픽 모델링은 비정형 데이터(unstructured data)인 텍스트로부터 정보를 뽑아내는 텍스트 마이닝의 한 종류로, 문서 집합(말뭉치; corpus)에 기저하고 있는 주요 주제들을 추출하는 통계방법이다(Blei & Lafferty 2009; Kwartler 2017; Richardson et al. 2014; Silge & Robinson 2017).

토픽 모델링을 통해 추출된 주요 주제를 토픽이라 부르며, 빈번하게 동시에 출현하여(co-occurrence) 높은 상관을 가지는 키워드의 혼합물로 구성된다(Blei, Ng, & Jordan 2003; Yan, Guo, Lan, & Cheng 2013). 따라서, 토픽의 해석은 토픽을 구성하는 주요 키워드를 활용하여 이루어진다(Karami et al. 2019). 예컨대, 특정 토픽에 ‘버스’, ‘지하철’, ‘기차’, ‘자동차’ 등의 단어가 높은 확률을 가지는 것으로 나타났다면 해당 토픽을 ‘교통수단’이라고 해석할 수 있다. 본 연구에서 토픽이란 분석에 포함된 청원게시판 게시글 전체의 기저에 잠재된 주요 청원 이슈들을 의미한다.

또한 토픽 모델링은 도출된 토픽에 따라 문서를 새롭게 분류하고 해당 주제와의 관련성 정도에 따라 주요 문서를 선정하는 방법으로도 활용가능하다(Blei & Lafferty 2009; Kosinski et al. 2016). 이는 토픽 모델링을 통해 추출된 주요 청원 이슈별로 청원 게시판에 등록된 게시글을 분류할 수 있을 뿐만 아니라, 이슈별 주요 게시글을 선별할 수 있음을 의미한다. 따라서 본 연구에서는 1차적으로 토픽 모델링을 사용하여 청원 데이터를 분석함으로써 잠재 토픽을 도출하고 토픽에 따라 문서를 새롭게 군집화(Blei & Lafferty 2009)한 뒤, 각 토픽별로 대표적인 청원글을 선정하였다.

한편 본 연구는 한국인이 불공정을 실제 어떠한 상황 혹은 조건에서 어떠한 방식으로 경험하고 있으며, 나아가 공정이라는 개념을 어떻게 인식하고 있는지를 경험적으로 도출하고자 한다. 그러나, 본 연구의 저자들은 토픽 모델링을 통해 추출된 토픽을 구성하는 단어의 집합만으로는 한국인의 불공정 경험 맥락과 불공정 인식을 보다 심층적이고 구체적으로 파악하기 어렵다고 판단하였다. 따라서, 토픽 모델링을 통해 선정된 주요 문서를 대상으로 질적 연구방법인 현상학적 분석을 사용하여 한국인의 불공정 경험 및 인식의 내용을 탐색하고자 했다. 즉, 불공정을 실제 경험하는 행위자의 목소리를 토픽 모델링과 질적 분석 방법을 통합적으로 활용하여 분석함으로써 사람들의 일상에서 나타나는 불공정 경험 및 인식을 구체적이고 심층적으로 파악하고자 했다.

토픽 모델링과 질적 분석방법을 함께 사용하는 것은 각 방법의 한계를 서로 상호 보완한다는 장점을 가진다. 예컨대, 청원 데이터와 같이 매우 방대한 텍스트 자료를 분석할 경우 자료를 모두 읽고 질적으로 분석하는 작업은 현실적으로 매우 어렵다(Karami et al 2019, Karami, Lundy, Webb, & Dwivedi 2020). 이때 토픽 모델링을 적용하여 일차적으로 자료의 패턴을 식별하여 자료의 크기를 효과적으로 축소함으로써 질적 분석에 필요한 데이터를 재구성할 수 있다. 또한 질적 연구의 경우 시각화가 상대적으로 어려우며 동일 자료를 분석하더라도 연구자들마다 반복된 결과를 얻기 힘들다(Pokorny et al. 2018). 반면 토픽 모델링의 결과는 2차 가공이 가능하기 때문에(김남규·이동훈·최호창·William Xiu Shun Wong 2017) 분석결과를 시각화하거나 조직화하기에 유리하며(Kwartler 2017), 시간에 따른 주제의 추이와 흐름을 볼 수 있다(Oh, Stewart, & Phelps 2017). 뿐만 아니라, 토픽 모델링을 통해 식별된 파라미터 추정치로부터 어떤 토픽이 더 많이 언급되었는지 등의 전체적인 경향성을 보다 쉽게 파악할 수 있다(Blei et al. 2003; Bittermann & Fischer 2018; Karami et al. 2019; Sievert & Shirley 2014).

하지만 토픽 모델링의 주요 결과인 토픽별 단어 집합은 실제 문서와 비교하였을 때 정보 손실이 발생하기 때문에 추정된 토픽의 키워드만으로는 해당 토픽에 대한 심층적인 해석이 어려운 것이 사실이다. 즉 청원 데이터를 망라하고 자료의 패턴에 따라 새롭게 군집화하여 전체 내용을 빠르게 살펴볼 수 있지만, 토픽 모델링 결과 자체로는 공정성이라는 개념을 깊이 탐색하기가 쉽지 않다. 예컨대 ‘가족’, ‘아들’, ‘딸’, ‘아버지’, ‘죽다’, ‘억울하다’ 등의 키워드를 가진 토픽이 나타났다면, ‘가족의 죽음’과 관련한 내용임을 유추할 수 있지만 ‘가족의 죽음’과 관련하여 구체적으로

어떤 내용이 나타났는지 정확히 알기 어렵다. 다시 말해, 단어 집합의 결과만으로 토픽을 해석한다면 전반적인 내용만 파악할 가능성이 높다. 이때 각 주요 문서의 내용을 심층적으로 탐구한 질적 분석결과를 통하여 어떠한 경험과 인식이 많이 나오는지 파악한다면 더 풍부한 해석이 가능할 것이다.

따라서 본 연구에서는 토픽 모델링과 질적 분석을 결합한 혼합 방법론을 사용하여 청원 데이터를 분석하고, 그 결과를 조직화함으로써 개별 방법의 단점과 장점을 상호 보완적으로 활용하고자 하였다.³⁾

II. 연구 방법

1. 연구 자료

분석 대상은 청원 시작일을 기준으로 2017년 8월 19일부터 2019년 12월 31일까지의 청원을 선정하였다. 이때 여러 웹 사이트 내 특정 정보를 수집하는 웹크롤링(Web Crawling)기법을 통하여 청원 데이터를 추출하였고 이 과정에서 R 패키지 중 하나인 rvest(Wickham 2020)를 사용하였다. 추출한 청원 데이터 구성은 청원 제목, 청원 본문, 청원 진행상태, 청원 동의 수, 청원 분야, 청원 시작일, 청원 종료일, 청원 식별번호로 이루어졌다. 청원 분야는 총 17가지로 정치개혁, 외교/통일/국방, 일자리, 미래, 성장동력, 농·산·어촌, 보건복지, 육아/교육, 안전/환경, 저출산/고령화 대책, 행정, 반려동물, 교통/건축/국토, 경제민주화, 인권/성평등, 문화/예술/체육/엔론, 기타로 구성되어 있다.⁴⁾ 청원 식별번호는 청원이 게시된 순서대로 부여된 번호로 각 청원을 식별하기 위해 사용한다. 이때 효율적인 분석을 위해 하루에 적어도

3) 혼합 방법론이란 연구자나 연구팀이 넓고 깊은 이해와 확증의 광범위한 목적을 위해 질적 연구와 양적 연구의 접근 요소를 결합한 연구 유형이다(Johnson et al. 2007). 즉 혼합 방법을 통한 자료 분석은 연구에서 양적, 질적 자료분석 전략이 결합, 연계, 통합되는 과정이라 볼 수 있다(Teddlie & Tashakkori 2009). 이러한 혼합 방법을 활용하는 것은 한 가지 방법만 채택하였을 때보다 연구 현상에 대한 깊은 이해를 제공하며(Creswell & Creswell 2018, Venkatesh, Brown, & Bala 2013) 서로의 한계를 보완시켜줄 수 있다(Creswell & Creswell 2018; Pokorny et al. 2018; 이순규·최수빈·김희웅 2019). 또 방법론의 다원화(방법론적 삼각측정법)를 통해 객관적 근거를 갖지 못하고 주관적인 해석에 머물 수 있다는 비판을 받아온 질적 연구의 타당성 및 신뢰성 확보에 도움이 될 수 있다(크레스웰 2015).

4) <https://www1.president.go.kr/petitions>.

30명 이상의 동의를 받은 6,084개의 청원만 분석 대상에 포함하였다. 실제 분석은 청원 게시글의 본문을 바탕으로 진행하였다.

2. 연구 도구

본 연구에서는 여러 토픽 모델링 중 잠재 디리클레 할당(Latent Dirichlet Allocation; LDA, Blei et al. 2003) 기법을 사용하였다. LDA는 널리 사용되는 대표적인 토픽 모델링 방법 중 하나로(김남규 외 2017; Ponweiser 2012) LDAvis(Sievert & Shirley 2014) 등 시각화 장치가 다양한 편이며, LDA를 구현할 수 있는 통계 소프트웨어에 대한 접근성이 높다.

LDA는 하나의 토픽이 여러 단어로 구성되었다고 가정하며, 하나의 문서에 여러 개의 토픽이 담길 수 있다고 가정한다(Blei & Lafferty 2009; Kwartler 2017; Silge & Robinson 2017). 이러한 LDA의 기본 가정은 하나의 청원 게시글 내에 하나 이상의 토픽이 담길 수 있다는 사실과 상응한다.

LDA 모형이 문서를 생성하는 과정을 보다 구체적으로 서술하면 다음과 같다. 총 D 개의 문서(document)와 총 V 개의 어휘(vocabulary)를 가진 말뭉치(corpus) 기저에 K 개의 토픽(topic)이 존재한다고 하였을 때,⁵⁾ N_d 개의 단어(word)를 가진 d 번째 문서는 아래의 과정을 거쳐 생성되는 것으로 가정한다.

1. (말뭉치 수준) 토픽별 단어 발생 확률인 V -차원 벡터 β_k ($k = 1, \dots, K$) 결정:

$$\beta_k \sim \text{Dir}_V(\delta)$$

2. (문서 수준) 토픽 발생 확률 K -차원 벡터 θ_d ($d = 1, \dots, D$) 결정;

$$\theta_d \sim \text{Dir}_K(\alpha)$$

3. (단어 수준) 문서 내 출현 단어 $w_{d,n}$ ($n = 1, \dots, N_d$)의 결정:

(a) 토픽 결정; $z_{d,n} \sim \text{Multinomial}(\theta_d)$, $z_{d,n} \in \{1, \dots, K\}$.

(b) 단어 결정; $w_{d,n} \sim \text{Multinomial}(\beta_{z_{d,n}})$, $w_{d,n} \in \{1, \dots, V\}$.

5) 모형의 추정을 위하여 전체 말뭉치에 기저하는 토픽의 개수 K 는 알려져 있다고 가정한다.

여기서 $Dir_V(\delta)$ 는 스칼라 파라미터 δ 를 갖는 V -차원의 디리클레 분포, $Dir_K(\vec{\alpha})$ 는 벡터 파라미터 $\vec{\alpha}$ 를 갖는 K -차원의 디리클레 분포, $Multinomial(\theta_d)$ 는 θ_d 를 파라미터로 갖는 다항분포를 가리킨다. 또한 $Multinomial(\beta_{z_{d,n}})$ 파라미터가 $\beta_{z_{d,n}}$ 인 다항분포로서, 이 분포를 결정짓는 파라미터인 토픽별 단어 발생확률 벡터 β_k 는 문서 내 출현할 단어별로 토픽이 결정된 후(즉, $z_{d,n} = k$) 최종 결정되므로, 이를 명시하기 위해 $\beta_{z_{d,n}}$ 으로 표기할 수 있다.⁶⁾

LDA의 핵심 파라미터는 문서 내에서 토픽이 발생할 확률을 의미하는 θ_d ($d = 1, \dots, D$)와 토픽 내 단어의 출현 확률을 나타내는 β_k ($k = 1, \dots, K$)로 이루어진다(Blei et al 2003; Blei & Lafferty 2009; Ponweiser 2012). 파라미터 β_k 는 K 개의 토픽별로 추정되며, 전체 토픽 수(K) $\times$ 말뭉치의 전체 어휘 수(V) 행렬로 통합하고 이를 $\beta_{K \times V}$ 로 표기할 수 있다. 이 행렬의 각 원소는 k 번째 토픽이 주어졌을 때 전체 어휘 내 v 번째 단어가 출현할 확률을 의미한다. $\beta_{K \times V}$ 추정치를 바탕으로 토픽마다 가장 높은 출현 확률을 가지는 단어들을 이용하여 토픽을 해석할 수 있다. 파라미터 θ_d 는 문서별 토픽의 출현 확률로 전체 토픽 수(K) $\times$ 전체 문서 수(D) 행렬 형태로 통합하고 이를 $\theta_{K \times D}$ 행렬로 표기할 수 있다. 이 행렬의 각 원소는 d 번째 문서에서 k 번째 토픽의 발생 확률 추정치를 나타내며, 토픽별 대표문서를 선별하는데 활용할 수 있다(Kosinski et al. 2016; Steyvers & Griffiths 2007).

LDA 모형은 말뭉치, 문서, 그리고 단어의 생성을 관장하는 핵심 파라미터를 추정하기 위한 3수준 위계 베이저안 모형(three-level hierarchical Bayesian Model)으로도 이해할 수 있다(Blei et al. 2003). LDA 모형의 핵심 파라미터를 추정하는 방식은 다양하지만 본 연구에서는 추정 속도가 빠르고 효율적이라고 알려져 있는 Collapsed Gibbs Sampling(Griffiths & Steyvers 2004; Steyvers & Griffiths 2007)을 사용하였다. 이는 위에서 설명한 LDA 모형이 문서를 생성하는 과정의 역으로 이해할 수 있다. 즉, 디리클레 분포의 사전 파라미터(α, δ)와 관찰된 자료인 문서/말뭉치 내 단어(w)가 주어졌을 때, 토픽별 단어의 출현 확률 분포($\beta_{K \times V}$)와 문서별 토픽 출현 확률 분포($\theta_{K \times D}$), $p(\beta, \theta, z | w, \alpha, \delta)^7$ 를 추정하는 과정으로 이해할 수 있다.⁸⁾

6) 보다 일반적으로 $p(w_{w,n} | z_{d,n}, \beta_k)$ 로 표기할 수 있다.

7) 간결한 표기를 위해 몇 번째 토픽, 문서 혹은 단어 등을 가리키기 위한 첨자는 생략하였다.

3. 자료 분석 절차

R(version 4.0.2)에서의 용이한 분석을 위해 워크롤링 파일을 엑셀로 취합하고 해당 과정에서 청원 하나를 각 행으로 지정하였다. 본 데이터를 기반으로 R(version 4.0.2)을 사용하여 형태소 분석 및 불용어 처리 등의 한글 전처리(pre-processing)를 진행하였다. 한글 전처리 과정을 통해 텍스트 데이터의 비정형성을 일정 수준 감소시킬 수 있다(길호현 2018).

단어를 사전에 등재된 표준 형태로 전환하는 어근 동일화 작업(Immatization)은 자연어 분석에서 거치는 일반적인 과정이므로(Oh et al. 2017; 김현중·이태현·유승의·김나랑 2018; 길호현 2018) NLP4kec 패키지를 사용하여 형태소 분석을 실시하였다. NLP4kec 패키지는 은전한닢 프로젝트¹⁰⁾에서 개발한 형태소 분석기인 me-cab-ko-dic¹¹⁾을 기반으로 개발되었다. 즉, 본 패키지는 띄어쓰기가 부정확한 문장일지라도 입력 데이터를 형태소 단위로 쪼갤 수 있으며 체언과 용언과 같은 의미 형태소 위주로 추출한 후 표준 형태로 전환한다. 예를 들어, ‘글을 쓰고’라는 문장을 형태소 단위로 쪼개면 ‘글/NNG(일반명사), 을/JKO(목적격 조사), 쓰/VV(동사), 고/EC(연결 어미)’가 된다. 이 후, 의미 형태소가 아닌 목적격 조사 ‘을’과 연결 어미 ‘고’를 삭제하고 일반명사 ‘글’과 ‘쓰’만 추출한다. 이때, ‘쓰’는 기본 동사형인 ‘쓰다’로 변환하기 때문에 최종적으로 ‘글’, ‘쓰다’만 결과로 변환한다.

모든 문서에서 빈번하게 나타나는 반면 문법적인 역할만 수행하는 형태소와 문장 부호를 제외하였다. 이렇듯 출현 빈도가 높지만 말뭉치의 의미 구성에 큰 영향을 미치지 못하는 어휘들을 불용어라 하며 이를 제거하면 주요 어휘들을 속아낼 수 있다(길호현 2018). 따라서 길호현(2018)의 한국어 불용어 목록을 분석에 활용하였다.¹²⁾

8) 자세한 추정 과정은 참고 문헌을 참고하기 바라며, 본 연구에서 사용한 topicmodels 패키지에도 구현이 되어 있다.

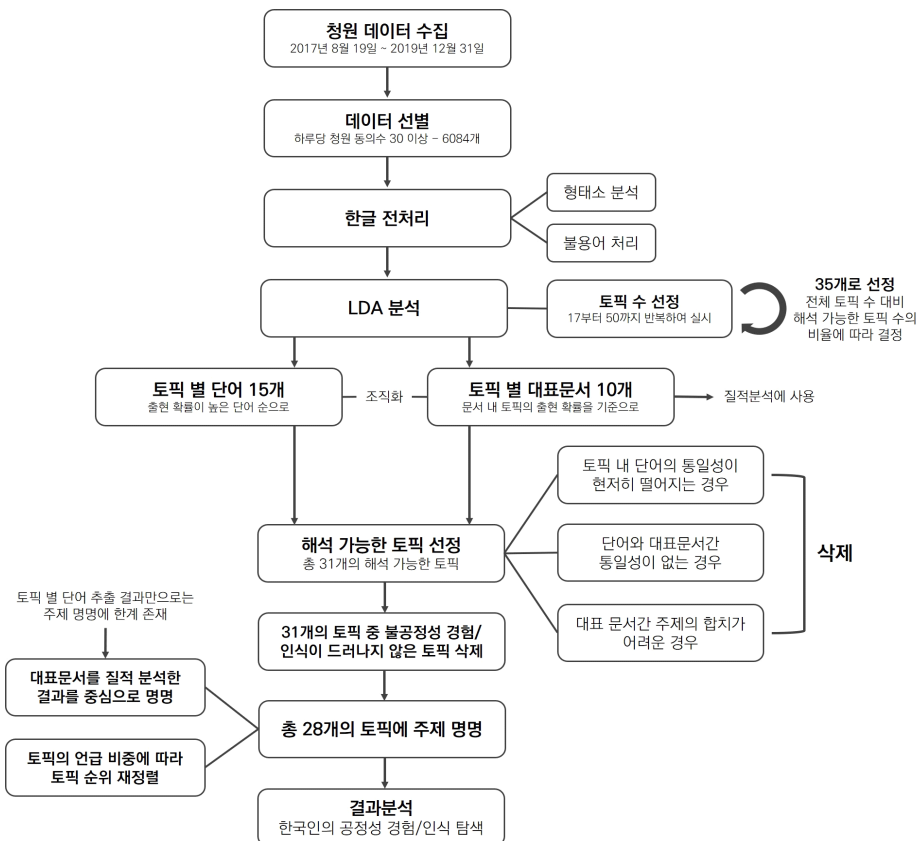
9) <https://github.com/NamyounKim/NLP4kec>.

10) <http://eunjeon.blogspot.com/>

11) MeCab이라는 오픈소스 형태소 분석 엔진을 활용한 형태소 분석기로, 띄어쓰기 정확성 유무에 상관 없이 입력 자료를 형태소 단위로 분석할 수 있다.

12) 이때 길호현(2018)이 제시한 불용어 ‘아이’는 감탄사(IC)로 분류되지만, 어린아이를 의미하는 ‘아이’는 핵심 품사인 일반명사(NNG)에 속한다. 하지만 현재 동음이의어를 완벽하게 구별할 수 있는 형태소 분석기는 존재하지 않기 때문에, ‘아이’를 불용어로 처리할 때 일반명사 ‘아이’ 또한 사라진다. 청원 게시글의 특성상, ‘아이’가 감탄사(IC)로 쓰였을 확률은 매우 낮다고 판단하여 ‘아이’는 불용어 목록에서 제외하였다.

또한 모든 문서에서 빈번하게 나타나는 단어(공통어; common words)는 토픽 간에 구별을 모호하게 한다(김지은·백순근 2016; 박준형·오효정 2017; Oh et al. 2017). 따라서 공통어에 가중치를 낮게 주는 역문서 빈도(Inverse Document Frequency; IDF)를 계산하고 그 값이 가장 낮은 단어들을 선별하여 공통어를 일차적으로 파악하였다. 그중 한국어 특성상 보조적으로 쓰인다고 판단되는 ‘하다’, ‘있다’, ‘되다’ 등을 공통어로 선정하였다. 또한 대통령, 청원, 국민은 분석 데이터인 국민청원 게시판 글의 특성상 나타나지만 분석에서는 무의미한 단어라고 판단하여 삭제하였다(김지은·백순근 2016, Oh et al. 2017) 따라서 본 연구에서는 ‘하다’, ‘있다’, ‘되다’, ‘주다’, ‘싫다’, ‘보다’, ‘대통령’, ‘청원’, ‘국민’, ‘여러분’ 등을 공통어로 간주하여 삭제하였다.



<그림 1> 청원게시판 텍스트 데이터 분석 절차 도식화

전처리 과정을 마친 데이터를 대상으로 LDA 분석을 실시하였다. 먼저 `tm` 패키지(Feinerer & Hornik 2019)를 사용하여 청원 데이터를 LDA 분석에 알맞은 Document-Term Matrix(DTM) 형태로 재구성하였다. DTM은 각 문서에서 특정 단어가 출현하는 빈도로 이루어진 행렬로 단어들 간 동시출현 패턴을 파악하는 데 사용된다.

LDA 분석은 `topicmodels` 패키지(Grün & Hornik 2020)를 사용하였다. 추정된 LDA 모형의 파라미터 β_k 를 바탕으로 각 토픽에서 출현할 확률이 높은 단어를 내림차순으로 정렬하여 그중 상위 15개를 선별하였다. 또한 파라미터 θ_d 를 기반으로 각 토픽별로 해당 토픽의 확률이 가장 높은 청원을 10개씩 추출하여 질적 분석에서 사용될 문서를 수집하였다. 즉 6,084개의 문서 중 해당 토픽을 가장 많이 언급한 청원을 속아내어 대표문서로 지정하였다.

이때 토픽 수는 모형에서 추출된 토픽들의 해석 가능성에 중점을 두고 선정하며(Chang, Gerrish, Wang, Boyd-Graber, & Blei 2009; Oh et al. 2017; Wallach, Murray, Salakhutdinov, & Mimno 2009) 의미가 불분명한 토픽의 경우 결과 해석에서 제외하였다(Karami et al. 2019; Oh et al. 2017). 기존 선행연구에서 보편적으로 사용된 perplexity, log likelihood와 같은 통계치가 적합하다고 하여 모형의 자료 설명력이 항상 높은 것은 아니기 때문이다(Chang et al. 2009).¹³⁾

이에 따라 반복 시행하기 위한 토픽의 범위는 17에서 50으로 설정하였다. 초기 토픽 수 17은 청원 게시판의 범주 수에 따라 선정하였다. 이때 Wallach et al.(2009)은 과도하게 작은 토픽 수를 채택하여 분석할 경우 토픽을 정확하게 분류하는 데 실패할 수 있으므로 토픽 수를 되도록 크게 설정한 후 의미 있는 토픽만 보고하는 방식을 제안하였다. 따라서 초기 설정값으로부터 토픽 수를 점차 늘린 후, 그중 해석

13) 본 연구에서 토픽 수를 탐색해보기 위해 perplexity와 log-likelihood와 같은 통계치를 여러 토픽 수를 대상으로 반복하여 계산하였으나 토픽의 수를 100까지 늘렸을 때 log-likelihood는 지속적으로 증가하였고 perplexity는 감소하였다. 즉 통계치만으로는 적절한 토픽 수를 설정하기에는 한계가 있었기 때문에 해석 가능성에 초점을 두고 토픽 수를 선정하였다. 또한, 본 연구에서 토픽 모델링을 활용하는 목적 중 하나는 청원데이터에 드러난 불공정 경험을 보다 효율적으로 탐색하는 것이다. 다시 말해, 6,084개의 청원을 대상으로 100개가 넘는 토픽을 뽑아내어 분석하는 것은 모형의 간명성이 떨어지고, 말뭉치를 빠르게 살펴볼 수 있는 토픽 모델링의 장점이 감소한다고 판단하였다. 이에 따라 본 연구는 토픽 모델링으로 추정된 결과가 인간이 가지고 있는 일반적인 상식과 개념에 부합하는지에 따라 모형의 결과와 성능을 판단하였다(Chang et al. 2009). 따라서, 여러 모형을 추정한 후 각 모형별로 토픽별 키워드 리스트와 대표문서를 교차 비교하여 키워드와 대표문서들 간 연관성을 바탕으로 해석 가능한 토픽을 판별하였다. 그 후 모형의 간명성을 고려하여 전체 토픽 수 대비 해석 가능한 토픽 수의 비율 등을 바탕으로 토픽 수를 결정하였으며 이 과정에서 질적 분석과 함께 연계하였다.

가능성을 바탕으로 적절한 토픽 수를 선정한 선행연구(Oh et al. 2017)의 방식에 따라 반복 분석을 진행하였다. 각 모형을 비교한 결과, 토픽 수를 35개로 결정하였다.

한편, LDA를 통해 분석한 결과의 모든 토픽이 언제나 쉽게 해석되는 것은 아니므로(Sievert & Shirley 2014) 전체 토픽 35개 중 해석이 가능한 토픽만을 제시하고자 하였다. 이를 보완하기 위하여 LDA를 통해 추출된 대표문서를 대상으로 질적 텍스트 분석을 실시하였다. 즉, 인터뷰나 자연스러운 대화에서 확보된 자료를 토대로 이루어지는 대부분의 질적 연구와는 달리 이차 자료를 대상으로 질적 분석을 실시하였다(Atkinson & Coffey 1997).

질적 텍스트 분석은 한국인의 불공정 경험 및 인식을 드러내기 위하여 현상학적 접근 방법을 사용하여 이루어졌다. 현상학적 분석이란, 하나의 개념이나 현상에 대한 여러 개인의 경험을 통해 공통적 의미를 기술하는 것으로, 개인의 주관적인 경험을 바탕으로 다른 사람과의 공통점을 찾아가는 과정 및 그 결과를 칭한다(크레스웰 2015; Moustakas 1994). 본 연구는 한국인의 불공정 경험을 이해하기 위해 Moustakas (1994)의 심리학적 현상학 분석 방법에 따라, 판단중지(epoche)를 거쳐 문서에서 불공정 경험과 공정성 인식에 대한 기술에 초점을 두고 유의미한 진술을 의미 단위로 그룹화하여 주제화하였다. (불)공정, 부당, 평등, 형평, 차별 등 정당성 지각과 관련된 단어를 포함하거나 분배적, 절차적, 응보적, 회복적 정의에 해당하는 내용을 포함하는 게시글에서 드러나는 구체적인 경험과 인식을 분석하였으며, 이러한 주제화와 토픽별 키워드 리스트를 바탕으로 총 31개의 토픽을 선정하고, 토픽의 주제를 명명하였다.

구체적으로 총 35개의 토픽별로 추출된 키워드 리스트와 대표문서를 상호 비교하여 35개의 토픽 중 31개를 해석 가능한 토픽으로 선정하였다. 일차적으로 키워드 간 연관성을 살펴보고 그 과정에서 추정된 토픽에 할당된 키워드들을 보고 주제를 유추할 수 있는 경우와 정확한 주제를 명명하기 어렵거나 해석이 모호한 경우 등으로 토픽을 분류하였다. 대표문서 또한 대표문서들 간의 관계성이 명확한 경우와 대표문서들이 하나의 주제로 통일되지 않는 경우로 나누었다. 그 후 키워드로 분류한 결과와 대표문서로 분류한 결과를 상호 비교하여 토픽의 키워드 리스트와 대표문서의 내용 간 합치성이 있는지 파악하였다. 이에 따라, 토픽 내 키워드와 대표문서 각각에서 모두 합치성을 보이며, 토픽을 구성하는 단어들과 대표문서들 간 내용이 유기적으로 연결되는 경우에 해당할 때만 해석 가능한 토픽이라고 판단하였다.

일차적인 내용분석이 끝난 결과를 바탕으로 청원 게시판에서 나타나는 불공정 경험과 인식을 탐색하였다. 이를 위해 공정성 이론과 ‘공정, 정의, 차별, 부당, 처벌’ 등 공정과 관련된 언어를 포함하는 게시글 내용에 근거하여 불공정과 관련된 경험과 인식을 분류하였으며, 단순히 사회 제도의 개선 또는 필요를 주장하는 내용, 구체적인 불공정 경험 없이 단순히 범죄 피해만을 호소하는 내용, 생명윤리와 같은 도덕적 이슈 등 공정성과 무관한 내용은 제외하였다. 이를 바탕으로 질적 분석을 실시하여 토픽의 전체 내용을 아우르는 주제를 명명하고, 각 주제에서 드러나는 불공정 경험과 인식을 탐색하였다.

한편 불공정성 경험 또는 인식이 나타난 토픽을 대상으로 토픽의 언급 순위와 비율에 따라 토픽 순서를 정렬하였다. 따라서 LDAvis 패키지를 통해 생성한 IDM (Intertopic Distance Map)을 바탕으로 선별된 토픽의 언급 순위와 비율을 관찰하였고 이에 따라 각 토픽의 내용을 구조화 하는 과정에서 토픽의 순서를 토픽의 언급 순위에 따라 내림차순으로 정렬하였다. 이를 통해 토픽들의 상대적인 언급 비중을 비교하였다.

III. 결과

토픽 수를 35개로 설정하여 LDA를 실시한 결과, 해석 가능한 토픽 31개를 식별할 수 있었다. 해석이 불가능하였던 4개의 토픽은 주제를 명명하기 모호하였기 때문에 결과에서 제외하였다.¹⁴⁾ 또한 불공정성 경험과 관련이 없었던 세 가지 토픽은 결과로 제시하지 않았다.¹⁵⁾ 따라서 총 28개의 토픽을 결과로 제시하였다. 토픽 모델링 결과와 질적 분석의 결과를 종합하여 정리한 <표 1>과 대표문서의 본문에서 인용한 문장과 토픽에 명명한 주제를 함께 제시한 <표 2>와 같다. 이때 <표 1>과 <표 2> 모두 해석이 가능하며 불공정성 경험과 관련된 토픽 28개만 제시하였다.

14) 대부분 대표문서 간 통일성을 가졌던 토픽들과 달리 해석이 모호하였던 4개의 토픽은 통일성이 떨어졌으며 상위 15개의 단어들 간 관계성도 낮아보였다. 따라서 토픽별 단어 간 합치성이 떨어지며 대표문서들 간 내용이 하나의 주제로 통일되기 어려운 경우인 토픽 4개를 제외하였다.

15) 해석이 가능하였던 31개의 토픽 중 ‘소비자 권리침해’, ‘동물 학대’, ‘의료보험 지원’은 구체적인 불공정성 경험 및 인식과는 관련이 없었다. 세 가지 토픽 모두 불공정 경험이 나타나기보다는 윤리적 이슈에 대한 이야기 또는 제도의 개선을 요구하는 이야기가 주를 이루었다. 따라서 해당 토픽을 제외하고 총 28개의 토픽을 구조화하여 제시하였다.

<표 1>의 질적 분석에 사용된 대표문서는 토픽 모델링 결과 추정된 LDA 파라미터 θ_d 를 기반으로 한 결과이다. 총 6,084개의 청원 중 각 토픽별로 해당 토픽을 많이 언급한 순으로 나열하였고 그중 순위가 높은 청원 10개를 대표문서로 설정하였다. LDA 결과, 대부분 문서에서 범주로 할당된 토픽의 확률이 0.4 이상인 것으로 나타났으며 이는 각 청원에서 해당 토픽이 이야기된 비중이 40% 이상임을 의미한다. 따라서 추출된 청원이 각 토픽을 대표하기에 적절할 것으로 생각된다.

<표 1>은 해석이 불가능한 토픽과 불공정성 경험 또는 인식이 드러났다고 보기에 어려운 토픽을 결과에서 제외한 후 토픽 모델링 결과와 질적 텍스트 분석 결과를 함께 구조화한 것으로, LDA 결과를 바탕으로 생성한 IDM에 근거하여 순위를 재배열하였다. <표 1>의 단어는 토픽 내 단어를 출현 확률에 따라 내림차순으로 나열하여 15개를 제시한 것이다. 이를 통해 청원 데이터에서 주로 드러난 이야기들을 빠르게 검토하여 일차적으로 내용을 파악하였다. <표 1>의 대표문서를 질적 분석한 내용은 추출된 대표문서들을 바탕으로 작성된 것으로 각 토픽의 주요 내용을 파악한 결과이다. 토픽별 키워드 리스트만으로는 주제 선정에 한계가 있었다. 예를 들어, 가장 언급이 많았던 순위 1위 토픽의 단어를 살펴보면 대략적인 내용은 폭행 사건에 관한 것이라 유추할 수 있지만 그 내용을 구체적으로 알기는 어렵다. 따라서 대표문서를 분석한 결과를 검토하여 각 토픽의 주제를 구체적으로 명명하였고 이를 <표 1>에 함께 제시하였다.

<표 2>는 토픽 모델링 결과와 질적 분석결과를 통하여 명명한 각 토픽별 주제와 대표문서의 본문에서 직접 드러난 내용을 함께 제시한 것으로 <표 1>과 동일한 순서로 배열되었다. 이때 대표문서에서 인용한 내용은 각 토픽에 해당하는 주제와 관련된 불공정성 경험 및 생각이 가장 두드러지게 드러나는 문장으로 선정하였다.

국민청원 게시판 분석의 결과 다양한 주제에 대한 불공정 경험이 드러났다. 우선, 생명 또는 신체에 대한 범죄 및 성범죄 피해를 다룬 주제가 상위 4위를 차지하였다. 본 주제들에서 나타난 불공정 경험을 살펴보면, 살인, 살인 미수, 주취 폭행, 몰카, 의료사고 등 범죄를 저질렀음에도 불구하고 가해자가 처벌받지 않거나 솜방망이 처벌을 받았을 때 즉, 응보적 정의가 구현되지 않을 때 불공정 지각을 강하게 하는 것으로 나타났다. 특히 이러한 범죄 피해를 다룬 게시글의 경우, 당황, 분노, 억울 등의 정서 경험과 자세한 피해 경위 및 정도를 상술함으로써 가해자의 처벌에 대한 필요성과 정당성을 부여하려는 특징이 있었다. 가해자가 미처벌 또는 솜방망이 처벌을 받는 이유로는 경찰의 부실·편파 수사와 잘못된 법·제도(낮은 처벌 기준, 청

소년·주취자·정신병자 감형제도, 의료사고에 대한 과실 규명 책임이 피해자에게 있음)를 꼽았다. 또한, 진실을 규명하고 가해자에게 피해에 합당한 처벌을 내리는 것, 부실 또는 편파 수사에 대한 책임을 묻는 것이 정의를 회복하는 유일한 방법이라 여기는 것으로 나타났다.

다음으로 행정조사기관 또는 수사기관의 부실 조사·부실 수사 주제가 5위와 6위를 차지하였다. 본 주제에서는 불법 행위나 범죄 행위에 대한 조사·수사 과정이 불투명할 때, 명확한 증거 없이 판결이 날 때, 담당 공무원이 무성의하고 부주의한 태도를 보일 때 부당함과 불공정을 경험하는 것으로 나타났다. 절차적 공정성과 관련된 이러한 경험은 담당 기관과 공무원에 대한 불신으로 이어졌으며 정의 회복을 위해서 엉터리 결론이 난 사건의 재조사, 담당 공무원의 과실 규명 및 책임자 처벌이 필요하다는 내용이 주를 이루었다.

7위는 남성 역차별에 대한 주제가 차지하였다. 이 주제에서는 현대사회에서 여성은 더 이상 사회적 약자에 포함되지 않는다는 주장과 함께, 여성 정책(여성 전용 주차장·좌석·휴게실 등), 여성 할당제, 군가산점제 폐지, 성인지 감수성 교육 등 한국 사회의 제도와 정책이 ‘여성 편향적’, ‘남녀의 공정한 경쟁 제한’, ‘여성 우대적’이므로 불공정하다고 인식하고 있음이 드러났다. 따라서 여성에 대한 보편적 지원이 ‘남성 차별적’, ‘혈세 낭비’이며, 여성을 지원하는 여성가족부가 폐지되고 그 예산이 임산부, 미혼모·미혼부, 고아 등 사회적 약자에 대한 정책에 사용되어야 함을 주장하는 내용이었다.

8위는 아동학대에 대한 주제로, 어린이집과 유치원 내에서 자녀의 아동학대가 의심됨에도 불구하고 원내 CCTV 설치 부족, 학부모의 CCTV 열람 권한 미보장, 아동 진술의 효력 불인정으로 인해 진실 규명과 책임자 처벌이 미흡하거나 불가능할 때 불공정을 경험한다는 내용이었다.

9위는 학교 내 권리침해 주제로, 중·고등학교 내 교권침해 내용과 학생자율권 침해 내용으로 나뉘었다. 교권침해의 경우 정교사, 상담교사, 영양교사가 자신의 수업권을 침해받았다고 느낄 때, 사립재단의 인사권 남용으로 인해 교장 선생님이 해임됐을 때 불공정을 경험한다는 내용이었고, 학생자율권 침해의 경우 성 고정관념에 따른 복장규제를 받을 때 불공정을 경험한다는 내용이었다.

10위에 해당하는 주제는 노동환경으로, 노동자의 열악한 처우, 수당을 받지 못하는 야근·철야 작업, 제자리 임금이 될 수밖에 없는 비정규 고용에 대한 내용이 주를 이루었다. 이와 상반되는 다른 갈래로 사업주 입장에서 건설노동조합원의 채용

협박, 업계 최고 대우임에도 불구하고 지속되는 파업, 불법체류 외국인 노동자 고용 증용으로 인해 불공정을 경험한다는 내용도 있었다.

11위는 주택조합 비리, 사기 행위가 차지하였다. 불투명한 주택조합의 의사결정 과정, 업무대행사가 내정한 조합원의 이익을 대변하지 않는 조합원장, 분양사기 등으로 보금자리를 잃을 위기에 처한 사람들의 경험이 드러났다. 또, 전 조합장, 시공사 또는 업무대행사의 횡령, 배임에 대한 증거를 수사기관에 제출하였으나 사건 조사가 제대로 이루어지지 않았다는 경험도 함께 나타났다.

12위는 다문화 정책에 대한 주제로, 난민을 포함한 국내 체류 외국인 지원은 외국인 특혜이며, 세금 낭비라는 인식이 드러났다. 공정한 사회의 실현을 위해 외국인 이민 수용을 중단할 것을 청원하였고, 이민을 수용한다면 한국인과 유사한 동양인의 선택적 수용, 난민수용 중단, 외국인 정착 지원 중단을 주장하는 내용이었다.

13위는 취업 및 입시 평가 기준이 차지하였다. 수능, 입시, 취업, 임용고시 등 10대~30대의 주요 이슈들이 등장하였으며, 객관적이지 않은 판단 기준, 답안 및 채점 기준 미공개, 수시전형 확대에 의해 불공정을 경험한다는 내용이었다. 또 타 대학의 특별편입으로 인해 수능 점수에 따른 능력주의가 위배될 때, 늘어난 학생 수로 인해 수업권 침해가 예상될 때 불공정을 경험한다는 내용도 있었다. 본 주제와 관련하여 정의를 회복하기 위해서는 명확하고 객관적인 평가 기준을 사전에 설정하고, 이러한 기준에 맞게 자원을 위계적으로 분배해야 한다는 주장이 주를 이루었다.

14위는 금융 시스템 주제로, 주식시장의 불공정한 관행인 공매도로 인한 불공정 경험을 담은 내용이 압도적으로 많이 나타났다. 문서에 따르면, 공매도는 개인투자자가 아닌 대규모 자본을 가진 기관 투자자만을 위한 관행이며, 공정사회 실현을 위해 개인 투자자에게 피해를 입힌 해외 투자기관 조사 및 책임자 처벌, 이를 묵인하고 있는 금융감독기관의 책임자 조사 및 처벌이 이루어져야 한다고 주장하였다. 한편, 세금으로 부실기업에 투자하는 국민연금의 불공정 행위를 지적하고, 책임자를 처벌해야한다는 내용도 있었다.

15위는 장애인 지원 부족이 차지하였다. 장애인의 열악한 교육 환경과 사회 참여 기회를 박탈당하고 있는 현실을 서술하고, 필요(needs)의 원리에 따라 주간 활동 서비스, 수어 통·번역 시스템 구축, 맹인의 홈페이지 정보 접근권 보장을 요구하는 내용이 주를 이루었다. 또, 일부 문서는 장애인 복지를 책임지는 활동가 및 활동 보조사에 대한 지원이 현실에 맞게 이뤄질 수 있도록 장애인 지원 현실화를 주장하는 내용이었다.

16위는 수도권 교통사업에 대한 내용으로, 일산, 파주, 구리, 남양주, 인천 등 부동산 정책으로 인구가 급증한 수도권 도시의 서울 접근성이 떨어지고, 출퇴근 시간이 길어지는 데 대한 어려움을 토로하는 내용이었다. 또, 지하철 건설 등 교통 환경을 조성하는 과정에서 본인 거주 지역이 제외된 경우 주민들의 의견이 반영되지 않았다고 느끼고 불공정과 차별을 지각하였으며, 이 과정에서 담당 공무원의 무성의한 태도와 법에만 근거한 동일한 답변으로 인해 불공정과 차별 지각이 더욱 증폭되는 것으로 나타났다.

17위는 기업의 갑질로, 의사결정 과정이 특정 대기업에 유리하게 작동하는 경우, 힘을 가진 큰 집단이 특정 개인이나 하청업체 또는 하청업체 직원에게 갑질하는 내용이었다. 또, 관청에 대기업 갑질에 대한 관리 감독을 요구하였음에도 답변을 받지 못한 내용도 있었다.

18위는 지역 내 혐오시설 건립에 대한 주제로, 의료폐기물·쓰레기 소각장 등 생활기반과 지역 환경을 해치는 시설 설립에 반대하는 내용이었다. 타 지역 쓰레기까지 처리하는 것은 형평에 어긋난다고 여기며, 부지 선정 과정에서 나타난 꼼수 환경평가, 대표성 없는 소수의 시민 대표 동의만을 얻어 진행된 졸속 주민의견 수렴 과정, 담당 공무원의 무성의한 태도와 법에만 근거한 동일한 답변으로 인해 불공정과 차별을 경험하는 것으로 나타났다.

19위는 언론의 가짜 뉴스 보도 및 여론 조작이 차치하였다. 팩트 체크 없는 보도로 본인이 피해를 입었을 때, 언론이 자극적 보도와 추측성 보도를 통해 2차 가해를 조장하고 사회적 갈등을 조장한다고 여길 때, 왜곡된 자료와 통계로 여론을 조작함에도 불구하고 이에 대한 처벌이 없다는 데에 불공정을 경험하는 것으로 나타났다.

20위는 각종 안전사고에 대한 내용으로, 안전사고 발생 시 경찰의 미흡한 조사로 인해 가해자의 가해 정도와 피해자의 피해 정도가 명확하게 드러나지 않을 때, 이로 인해 과실처리 진행이 미흡할 때, 이 과정에서 경찰이 무책임한 태도를 보일 때 불공정을 경험한다는 내용이 있었다.

21위는 책임자의 불공정 행위에 대한 미처벌이 차치하였다. 기업이나 공공기관 등의 책임자가 남품 비리, 인사권 남용, 허위 공문서 작성 등 불공정행위를 저질렀음에도 불구하고 이에 대한 철저한 조사와 처벌이 이뤄지지 않을 때, 오히려 신고자가 위협을 받거나 보복성 인사를 당하는 등 피해를 입을 때 불공정을 경험한다는 내용이었다.

22위는 불법 사이트가 차치하였다. 두 종류의 불법 사이트에 대한 내용이 있었는데, 첫째는 웹툰, 웹소설 등 개인 창작물에 대한 저작권 침해 사이트, 둘째는 몰카,

불법 동영상 등 인터넷 음란 게시글을 공유하는 사이트였다. 저작권 침해 또는 몰카 피해를 입을 때 부당함과 불공정을 경험하며, 이러한 피해가 지속됨에도 불구하고 관련 기관의 예방적 조치가 미비할 때, 범죄 수사와 처벌이 부실할 때 불공정 지각은 더욱 증폭되는 것으로 나타났다.

23위는 공공임대 주택으로, 불투명한 공공임대 주택이나 아파트의 분양가 심사와 입주 시 약속과 달리 분양가 상한제 적용 제외로 인한 피해 내용이 압도적으로 많았다. 그 외 보유세와 종부세를 강화하여 임대사업자 세제 지원을 중단하고 공평한 납세를 실현해야 한다는 주장이 있었다.

24위는 어린이집·유치원 운영 및 보육교사 처우에 대한 내용이었다. 국·공립 유치원·어린이집과 사립 유치원·어린이집의 차등적 지원으로 인하여 같은 환경에서 경쟁이 불가능하며, 아동학대 등 동일한 잘못에 대해 국·공립은 원장 퇴직 처분이 내려지지만, 민간은 벌금 및 폐원 처분이 내려지는 데 대한 불공정을 경험하였다. 그 외 사립유치원 교사의 열악한 처우에 대한 불공정 경험이 드러났다.

25위는 경찰 및 소방 공무원 처우에 대한 주제가 차지했다. 공무 수행 중 사망한 경찰, 소방 공무원에 대한 순직이 인정되지 않을 때, 근무 중 입은 재해를 자비로 치료할 때, 순직 사고에 대한 예방이 미흡할 때 부당함과 불공정을 경험한다는 내용이었다.

26위는 검찰개혁에 대한 주제로, 기소권 남용, 검찰의 정치개입, 수사기밀 유출 등 잘못에도 불구하고 처벌받지 않는 검찰과 법원에 대해 불공정을 경험하는 것으로 나타났으며, 검찰권 개혁을 통해 검찰도 처벌받을 수 있도록 함으로써 정의와 공정을 회복하고자 하는 내용이었다.

27위는 의료인 자격이 차지하였다. 심리치료사, 임상심리사, 치위생사 종사자들이 전문성을 갖추고 의료행위를 하고 있음에도 불구하고 의료인으로 인정받지 못하여 직업 활동에 제약이 있으며, 열악한 노동환경에서 근무해도 개선할 수 없는 데에 불공정을 경험한다는 내용이 주를 이루었다. 또, 일부 문서에서 치과 병원이 공장식 이벤트를 개설하여 선결제를 유도하고, 제대로 된 치료 없이 병원 문을 닫아버리는 먹튀 사건에 대한 수사와 의사 처벌을 요구하는 내용이 있었다.

마지막으로, 환경오염 주제와 관련하여, 오염물질을 내뿜는 발전소의 지역 내 건립에 대한 반대가 나타났다. 미비한 오염물질 저감 시설, 부지선정 과정의 불투명함, 폐기물처리 과정 등 절차적 공정성이 지켜졌는지에 대한 의문을 제기하고, 특히 성장기인 지역 내 아동·청소년이 입는 피해를 지적하며 불공정을 주장하는 내용이 주를 이루었다.

<표 1> LDA 내용 분석

토픽 순위	단어	주제	불공정 경험(대표문서를 바탕으로 정리)	정의 유형
1	경찰, 오다, 앓다, 동생, 친구, 없다, 전화, 가다, 폭행, 같다, 사람, 나오다, 여자, 신고, 연락	폭행·성범죄 피해	부실한 수사 과정: 범죄 사건(살인미수, 주취 폭행, 몰카, 성희롱)에 대해 수사 경찰이 사건을 축소하거나 편파적인 수사를 함. 성범죄 등 수사 과정에서 인권 침해를 당함. 가해자 솜방망이 처벌: 본인 또는 가족의 범죄(살인미수, 주취 폭행, 몰카, 성희롱) 피해에도, 가해자가 나이·정신병 이력·주취 상태를 이유로 감형을 받거나 낮은 처벌 기준으로 인해 솜방망이 처벌을 받음	절차, 응보
2	아들, 가족, 없다, 딸, 살다, 죽다, 앓다, 아버지, 남편, 아빠, 못하다, 가다, 억울하다, 받다, 사람	억울한 가족의 죽음	억울한 가족의 죽음에 대한 진실 규명과 가해자 처벌: 살인 또는 폭행 범죄로 가족이 사망했음에도 심신미약 등을 이유로 가해자에 대한 법적 처벌이 없거나 미약함. 진실 규명과 강력한 처벌을 통해 정의를 회복하고자 함.	응보
3	병원, 없다, 수술, 받다, 앓다, 의사, 후, 상태, 오다, 환자, 입원, 검사, 상황, 치료, 사고	의료사고	의료사고 과실 의사 처벌의 어려움: 대형 병원 의료진의 오진과 안일한 대처로 본인 또는 가족의 질병이 악화하거나 사망하였음에도, 병원을 상대로 의료진 과실 규명을 해야 하는 책임이 피해자에게 있어 과실 의료진에 대한 처벌이 어려움	응보
4	피해자, 처벌, 가해자, 피해, 사건, 받다, 범죄, 폭력, 없다, 당하다, 폭행, 앓다, 보호, 청소년, 대하다	가해자 솜방망이 처벌	범죄 가해자에 대한 솜방망이 처벌: 청소년 범죄, 주취 폭행, 성범죄 가해자가 나이·정신병 이력·주취 상태·심신미약 등을 이유로 솜방망이 처벌을 받음. 강렬한 분노 표현과 함께 정의 회복을 위한 엄벌의 필요성을 주장	응보
5	대하다, 내용, 조사, 의하다, 일부, 요청, 수정, 해당, 관리자, 앓다, 위배, 요건, 요구, 사실, 진행	행정조사 기관의 부실 조사	조사 과정 미공개와 무성의한 태도: 행정조사 기관이 조사 과정을 투명하게 밝히지 않고, 부실하게 조사하거나 조사를 시작하지 않음, 담당자의 무성의한 태도와 답변으로 분노, 답답함과 함께 불공정을 경험 유책 담당자와 행정기관 미처벌: 조사 과정을 공개하지 않는 기관과 무성의한 태도의 담당자. 이들에 대한 과실 규명과 책임자에 대한 처벌을 통해 정의 회복 소망	절차, 응보
6	사건, 수사, 경찰, 받다, 판결, 증거, 없다, 재판, 법원, 구속, 피해자, 변호사, 검사, 대하다, 진술	수사기관의 부실 수사	검·경찰의 엉터리 수사 과정과 증거 없는 판결: 성폭력, 성추행, 비리 사건 수사 과정에서 발생한 인권 침해, 증거 없는 판결로 인한 억울함과 수사기관에 대한 불신	절차, 응보
7	여성, 남성, 성, 차별, 남자, 성매매, 인권, 평등, 앓다, 사회, 여자, 없다, 받다, 지원, 대한민국	남성 역차별	남성 역차별: 여성은 사회적 약자가 아니라는 인식과 함께 여성 우대 정책과 남녀 공정 경쟁 제한 정책(여성 할당제 등), 군가산점제 폐지 등 여성 편향적 정책의 불공정성을 주장, 성범죄와 관련하여 남성이 무조건 가해자로 인식되고 처벌받는 데 분노	분배
8	아이, 부모, 없다, 엄마, 앓다, 선생, 아동, 학대, 앓다, 먹다, 키우다, 받다, 가다, 때, 보내다	아동학대	처벌이 어려운 어린이집·유치원 아동학대: 원내 CCTV 부족 및 CCTV 학부모 열람 권한 미보장으로 인해 아동학대의 증거를 찾기 어려우며, 아동의 진술은 증거로 인정받지 못해 책임자 처벌 미흡	응보
9	학교, 학생, 교사, 교육, 선생, 초등, 수업, 교육청, 학부모, 고등, 받다, 앓다, 학년, 교장, 상담	학교 내 권리 침해	교권 침해: 정교사처럼 수업을 하지만, 수업 중 발생한 문제에 책임은 지지 않는 영어 회화 전문강사와 스포츠 강사 제도로 인한 정교사 수업권 침해, 학폭위 진행 교사의 트라우마와 수업권 침해, 사립재단의 인사권 남용 학생 자율권 침해: 성 고정관념에 따른 학생 복장 규제	절차

토픽 순위	단어	주제	불공정 경험(대표문서를 바탕으로 정리)	정의 유형
10	근무, 직원, 임금, 회사, 노동자, 받다, 고용, 없다, 노조, 노동, 최저, 근로, 지급, 근로자, 않다	노동환경	열악한 노동자 처우·고용환경 노동 환경: 야근과 철야에도 임금을 제대로 받지 못하는 비정규직, 회사에 이러한 불법 행위 시정 요청을 하거나 관련 증빙 서류를 요구한 대가로 갑자기 근무가 바뀌거나 해고를 당한 노동조합의 부당한 요구: 노동조합에서 노조원을 채용하도록 사측을 협박하거나 파업, 불법 고용을 중용함	분배, 절차
11	돈, 받다, 조합원, 없다, 않다, 조합, 사기, 계약, 운영, 사업, 재산, 호텔, 지급, 금액, 나다	주택조합	주택조합 집행부 선정 절차의 불투명성: 업무대행사가 미리 내정한 자가 주택조합 집행부를 장악하고, 이로 인해 업무대행사의 비리와 분양사기 등에도 조합 구성원의 이익을 대변하기 어려움 수사기관의 부실 수사: 조합원들이 새로운 집행부를 구성하고, 전 조합장, 시공사 또는 업무대행사의 범죄 사실에 대한 증거를 수사기관에 제출하였으나 이유를 알 수 없이 사건 조사가 제대로 이루어지지 않음	절차, 옹보
12	한국, 나라, 일본, 외국인, 국가, 대한민국, 문화, 난민, 미국, 정부, 중국, 세계, 국제, 않다, 문제	다문화 정책	국내 체류 외국인 지원 및 다문화 정책의 자국민 차별: 난민을 포함하여 세금을 내지 않는 국내 체류 외국인 지원 정책은 세금 낭비이며 외국인 특혜임, 이들에 대한 지원은 자국민의 경제활동과 안전을 위협하고, 자국민의 인권을 침해함	분배
13	시험, 교육, 대학, 학생, 평가, 대학교, 문제, 교수, 과정, 준비, 공정, 위하다, 공부, 취업, 임용	취업 및 입시의 평가 기준	객관적이지 않고 불명확한 취업 및 입시 평가 기준: 중등 교사 임용고시 답안 및 채점 기준 미공개, 수능 수시 제도의 평가 기준이 모호하므로 평가 기준이 명확하고 객관적인 정시전형을 확대하는 것이 공정하다고 주장. 타 대학 학생의 특별편입: 수능 점수에 따른 능력주의 위배, 준비 없이 늘어난 정원으로 학생들의 학습권 침해.	분배
14	시장, 투자, 금융, 주식, 없다, 정부, 개인, 기업, 투자자, 경제, 공매도, 않다, 거래, 공매, 시스템, 삼성	금융 시스템	공매도 정책의 불공정성: 공매도 정책은 거대 자본으로 시장을 왜곡하고, 정보력이 낮은 개인투자자가 손해를 입을 수밖에 없는 불공정한 제도임, 공매도 제도로 인해 부당한 이익을 얻은 자와 이를 방조한 금융 기관의 책임자 처벌을 요구. 국민연금 공단의 주식대여: 세금으로 부실기업에 투자를 일삼는 국민연금 공단의 주식대여를 금지하고 책임자 처벌 요구.	분배, 옹보
15	장애, 사회, 지원, 복지, 활동, 센터, 위하다, 가정, 특수, 받다, 가족, 서비스, 예산, 돌보다, 대하다	장애인 지원 부족	열악한 장애인 교육·사회 참여 환경: 장애인 학교 지원, 영터리 수어 통·번역 체계, 주간 활동 서비스 지원 부족 등으로 장애인의 학습권 및 사회 참여권이 보장되지 않음. 장애인 시설 사회복지사의 임금: 장애인 시설 사회복지사는 직능별, 지자체별로 임금 차이가 커 동일 노동 동일 임금의 원칙이 지켜지지 않고 있음, 장시간 고노동에도 불구 야근 수당 등이 미지급되는 경우가 많음.	분배
16	도시, 서울, 지역, 사업, 계획, 주민, 교통, 인천, 정부, 선, 인구, 개발, 수도, 노선, 경기도	수도권 교통사업	열악한 수도권 교통 환경: 정부의 부동산 정책으로 인구가 급증한 수도권 (산)도시의 서울 교통 접근성이 낮음. 불투명한 의사 결정 과정 및 담당자의 무성의한 태도: 교통 환경 조성 의사 결정 시 주민 의견 반영 미흡, 담당자의 무성의한 태도.	분배, 절차
17	기업, 회사, 업체, 갑, 대표, 사업, 산업, 계약, 오다, 공정, 기술, 없다, 않다, 협력, 제품	기업 갑질	기업 간/기업 내 갑질: 의사 결정 과정이 특정 대기업에 유리하도록 작동하는 구조로 인해 중소기업이 피해를 입음, 힘을 가진 (대)기업의 개인·하청 업체를 향한 갑질. 대기업 갑질에 대한 관리 감독의 부재: 공정 경쟁 유지 등을 위해 관청에 관리 감독을 요청하였으나 답변을 받지 못함.	절차
18	주민, 지역, 시설, 곳, 허가, 환경, 살다, 행정, 시민, 반대, 이전, 사업, 공사, 문제, 마을	지역 내 혐오시설 건립반대	생활기반과 지역 환경을 해치는 기관 설립: 이미 지역 내에서 타지역 폐기물을 처리하고 있음에도 타지역 폐기물 처리를 위해 쓰레기처리장을 추가 건설하는 것은 형평성에 맞지 않음. 불투명한 처리 과정과 담당자의 무성의한 태도: 공수 환경영향평가와 미비한 주민 의견수렴 과정, 담당자의 무책임한 답변.	분배, 절차

토픽 순위	단어	주제	불공정 경험 (대표문서를 바탕으로 정리)	정의 유형
19	언론, 기사, 대하다, 방송, 뉴스, 보도, 기자, 밝히다, 않다, 사실, 조작, 자신, 댓글, 여론, 발언	가짜뉴스 및 여론조작	언론의 팩트체크 없는 가짜뉴스 보도에 대한 미처벌: 조회 수를 위한 언론의 자극적 기사, 추측성 보도로 2차 가해와 사회적 갈등을 조장하고 왜곡된 자료와 통계 제시, 댓글 조작으로 여론을 조작하는데도 불구하고 이에 대한 처벌이 미흡함.	응보
20	사고, 안전, 차량, 없다, 없다, 현장, 발생, 타다, 기사, 도로, 자동차, 택시, 운전, 운전자, 상황	안전사고	미흡한 안전사고 진상조사: 경찰의 미흡한 조사 및 무책임한 태도, 과실처리 진행 미비	절차, 응보
21	회장, 세월, 없다, 국회, 청와대, 처, 않다, 감사, 비리, 특별, 관련, 불법, 진상, 규명, 인사	책임자의 불공정행위에 대한 미처벌	관리자 또는 책임자의 처벌받지 않는 불공정 행태: 납품 비리, 인사권 남용, 진상조사 방해, 허위 공문서 작성 등 업무 책임자가 저지르는 불공정 행태에도 조사가 부실하며, 책임자가 처벌받지 않음. 오히려 신고자를 위협하거나 보복성 인사를 당함	응보
22	불법, 사이트, 게임, 사진, 촬영, 올리다, 않다, 영상, 글, 정보, 인터넷, 지다, 많다, 만들다, 몰카	불법 사이트	개인 창작자의 저작권 보호 및 관리부실: 저작권 보호를 위한 기관의 방조 아래 창작자의 저작권 침해가 계속됨. 몰카 등 불법 인터넷 음란 게시물 수사 및 처벌 미비: 온라인 성범죄에 대한 수사가 제대로 이뤄지지 않으며, 온라인 게시물 특성상 피해자의 피해를 멈추기 어려움에도 불구하고 성범죄자에 대한 처벌이 없거나 약함	분배, 절차, 응보
23	아파트, 분양, 주택, 임대, 공공, 서민, 세대, 부동산, 건설, 없다, 공사, 입주, 정책, 국토, 살다	공공임대 주택	불투명한 공공임대 주택 아파트 분양가 심사과정: 입주 시 약속과 달리 임대주택 아파트의 분양가 상한제 적용 제외되었으며, 분양가 심사의 기준과 과정이 공개되지 않아 분양가 심사과정이 불투명함 임대사업자 세제 지원책: 공평한 납세를 위해 임대사업자 세제 지원책 중단과 부동산 보유세 및 종부세 강화 주장.	절차, 분배
24	교사, 어린이집, 유치원, 보육, 원장, 교육, 않다, 운영, 유아, 지원, 사립, 받다, 학부모, 많다, 평가	어린이집·유치원 운영 및 보육교사 처우	사립유치원 교사의 열악한 처우: 국공립 유치원 교사와 다른 급여기준, 경력 대우 없음, 적은 임금, 국공립 유치원 확충 요구 국공립 유치원·어린이집과 사립 유치원·어린이집의 지원 및 처벌의 차등: 국공립과 민간에 대한 지원 차이(인건비 등)로 인해 같은 환경에서 경쟁할 수 없음, 같은 잘못에 대해 국공립은 원장 퇴직 처분이지만, 민간은 벌금 및 폐원 처분임	분배, 응보
25	공무원, 업무, 경찰, 관리, 현장, 근무, 위하다, 않다, 급, 안전, 국가, 직, 소방, 없다, 경찰관	경찰 소방 공무원 처우	예방이나 보상 없는 공무 수행 중 사망 또는 재해: 경찰 또는 소방공무원이 위험직무 중 사망하였음에도 순직 인정을 받지 못함. 공무 수행 중 발생한 재해임에도 자비로 치료비를 부담함. 순직사고 예방 대처가 미흡	회복
26	검찰, 수사, 장관, 개혁, 조국, 총장, 대하다, 법무부, 국가, 부대, 대한민국, 의혹, 검사, 군인, 후보	검찰 개혁	처벌받지 않는 검찰: 기소권 남용, 정치개입, 수사기밀 유출 등 잘못해도 불구하고 처벌받지 않는 검찰과 법원.	응보
27	간호사, 의료, 의사, 건강, 정신, 환자, 없다, 위하다, 병원, 대하다, 전문, 면허, 인력, 면허, 인력, 자격	의료인 자격	전문성을 인정받지 못하는 의료계 종사자: 심리 치료사, 임상심리사, 치위생사, 타투이스트 등 전문성을 위한 교육과 자격을 갖추고 있음에도 불구하고 법적으로 인정받지 못하여 직업 활동에 제약이 있으며 열악한 노동환경에서 근무 처벌받지 않는 의료인 사기: 병원의 공장식 이벤트 ‘떡튀’ 사건으로 인한 수많은 사람의 피해에도 해당 의사의 처벌이 미흡	분배, 응보
28	미세먼지, 발전소, 물질, 안전, 문제, 환경, 없다, 않다, 위험, 발생, 먼지, 건강, 지다, 정부, 인하다	환경오염	발전소·핵발전 시설 관리 미비 및 담당자의 무성의한 태도: 지역 발전 시설의 불투명한 부지선정 과정, 발전소의 오염물질 저감 시설 미비, 폐기물처리에 대한 의문을 제기하였음에도 명확한 답을 들을 수 없음 지역 내 환경오염: 중국발 미세먼지,공단 미세먼지로 인해 지역주민의 일상과 아이들의 성장에 악영향을 줌	절차, 분배

<표 2> 주제에 따른 불공정 경험

토픽 순위	주제	대표문서의 본문에서 인용한 문장(내용)
1	폭행·성범죄 피해	“다시 한번 사건진술을 하기 위해 경찰서에 갔습니다. 형사님이 그러시더군요 합의금을 너무 많이 부른 거 아니냐고 그러니 합의가 안 됐지 하면서 얘기를 하며 저와 동생을 가해자로 피의자 심문을 했습니다” “범죄가 아니랍니다 억울하겠지만 법이 그렇답니다 신체 중요부위를 가까이서 찍은게 아니라서 전신 얼굴 사진 찍는 건 해당되지 않는답니다 신체 일부가 확대 되어 부각되어야 하고 제 3자가 보았을 때 수치심 느끼겠다고 생각될 사진이어야 합니다”
2	억울한 가족의 죽음	“부디 심신미약이라는 걸로 (가해자에 대한) 벌이 줄어들지 않길 바랍니다. 지은 죄만큼, 떠난 엄마와 남은 가족들의 고통만큼 벌 받았으면” “(피해 사망자인) 동생을 발견한 형사분은 (가해자를) 처벌할 수가 없다며 조용히 장례를 치러야할 것 같다고 정신적인 피해보상은 민사소송으로 따로 진행하라고 말하였습니다”
3	의료사고	“수술은 의사가 하고 문제가 생기면 의사는 뒤로 빠지고 거대한 대학병원이 일을 모두 처리하니 일반 개인이 이런 일을 당하면 억울해도 침묵해야 하는 상황입니다” “의료사고 발생 시 그들 대학병원 협회끼리 묻쳐 서로 봐주기식으로 넘어가고 거대한 힘 대학병원 협회로 개인을 짓누르는 이 실태에 국가가 개입하여 해결해야 한다”
4	가해자 숨방망이 처벌	“(아동 성범죄자들) 법적으로는 중형이 주를 이루었지만 현실은 어떻습니까 현황으로 봐도 제대로 집행되지 않고 있습니다. 이마저도 재판에서는 심신미약 등의 이유 등으로 감형이 적용되어 가해자들이 정당한 형량을 받지 못하고 있습니다”
5	행정조사 기관의 부실 조사	“국가기관으로서 여러 운전자들의 신고를 접수하고도 조사를 제대로 하지 않는 것은 엄연한 직무유기입니다” “금융 관련 불공정에 대해 OO 기관에 민원을 수도 없이 제기하였습니다. 그러나 해결된 민원은 단 한 건도 없고 매번 돌아오는 것은 앵무새 답변뿐이었습니다”
6	수사기관의 부실 수사	“모순덩어리 진술을 그대로 받아들이는 엉터리 수사로 구속영장까지 신청하고 기소하여 재판까지 가게 된 과정에 대해서 책임을 져야 할 것입니다 (중략) 검사는 법위에 존재하는 것 인가요 이런 게 정당한 수사라면 어떻게 대한민국 검사를 믿겠습니까”
7	남성 역차별	“남자가 일하던 여자가 일하던 능력별로 일에 적합한 사람이 일하는 것이지 성별이 일하는 것이 아니므로 양성채용목표제는 폐지되어야 마땅합니다”
8	아동학대	“통상적으로 아동학대가 이루어지면 아동보호전문기관에 CCTV를 제공하여 아동학대의 정황에 대하여 검증을 받아야 하지만 담당 수사관은 개인정보보호법을 이유로 제공하지 않았습니다. 아동보호전문기관에 cctv가 제공되지 못해 다뤄볼 여지가 줄어들었고 반복적이고 악의적인 정서적 아동학대임을 수차례 강조하였지만 경찰은 폭행에만 포커스를 맞춘 탓에 제가 정말 벌하고 싶었던 정서적 아동학대 내용은 처벌 사항에 포함되지도 못했습니다”
9	학교 내 권리 침해	“(영어회화 전문강사 스포츠 강사) 정교사와 권한이 같으며 더 중요한 것은 책임은 없단 사실입니다. 잘못된 조치를 취해도 영어를 잘못 가르쳐서 민원이 발생 해도 체육 수업에 학생들이 다쳐도 그들의 책임은 없습니다” “도발과 폭장 규정(중략) 명백한 성차별입니다. 왜 여성이면 무조건 치마를 입어야 하냐요 왜 배우는 공간인 학교에서까지 차별을 받아야 합니까”
10	노동환경	“내가 일한 대가’를 투명하게 알기 위해 ‘급여명세서’ 공개를 요청했더니 전원 해고 통보를 받았습니다” “불과 10년 전만 해도 연봉환산 5,000만원 정도였던 OO 운전기사 수입이 10년 사이 24배 가량 오른 이유는 특정 노동조합에서 OO 운전 기사고용을 건설현장에 강요 사실상 협박하기 때문이고”
11	주택조합	“새로운 조합장 및 집행부는 그동안 업무대행사와 전조합장이 저지른 사기횡령배임에 대한 증거들을 어렵게 찾아내 수사기관에 제출하였고 주택조합 사기와 횡령에 가담했던 일부 사람들이 진술도 하였지만 업무대행사대표와 전조합장은 구속조차 되지 않았습니다”
12	다문화 정책	“각종 외국인 특별로 국민세금 납비하고 한국인이 차별받고 있는 실정까지 다다랐습니다” “자국민의 반대의견과 행동은 차별이 아니다. 국민은 안전을 원한다. 뿐만 아니라 협정은 ‘인종주의 외국인 혐오 이민자에 대한 편협함 등 모든 형태의 차별을 철폐하겠다는 약속을’ 하여 자국민의 반대 의견을 차별로 간주하고 제재하도록 한다”
13	취업 및 입시의 평가 기준	“요즘 입시의 대체 학생부 종합전형학종의 민낯을 들여다보면 그렇지 않습니다. 오히려 단순과 공정과 거리가 멍니다” “무엇보다 평가 기준이 명확히 공개되지 않아 당락을 결정짓는 요소가 무엇인지 그 누구도 모릅니다”
14	금융 시스템	“공매도는 오로지 기관투자자들이나 외국인 투자자들을 위한 제도로서 개인투자자들에게 일반적으로 희생을 강요하는 국민평등권을 심각히 침해하는 제도입니다” “주식시장의 불공정 관행을 바로잡아야 할 OO감독원과 그 감시자 역할을 하는 OO기관은 과연 국민을 위해 존재하는 기구인지 기관투자자들의 이익을 대변하는 단체인지 묻지 않을 수 없습니다”

토픽 순위	주제	대표문서의 본문에서 인용한 문장(내용)
15	장애인 지원 부족	“난청 학생들은 특수교육 대상자로 이들의 요구를 충족시켜 줄 지원이 필요하지만, 거점 학교나 복지관이 부족한 실정입니다” “하지만 정부는 8년 내내 저에게 사회복지사가 아닌 사회복지시설 기능직 1호봉에 해당하는 인건비만을 지급하는 차별을 자행하고 있습니다”
16	수도권 교통사업	“국가가 재정을 세우고 진행되는 재정사업에서 철도가 경유하는 일부 지자체를 배제하고 계획을 수립하는 행위는 용납할 수 없는 심각한 지역차별” “담당자분께서는 글자하나 안바뀌고 복사해서 붙여넣는 방식으로 일관되게 답변을 주면 모두에게 회신을 주셨으며 이런 어이없는 답변을 받기 위해 수백 명의 주민들이 민원을 열심히 썼다는 게 너무 비참할 따름”
17	기업 갑질	“중소기업을 이익추구의 도구로만 여기는 OO 기업의 갑질 기업문화 때문에 많은 중소기업이 억울한 고통을 당하고 있습니다” “항만의 근로자공급사업과 관련된 OO 기관 담당자, OO 기관 시장감시총괄팀 등 들이 작업현장에 한번 와서 ‘허가사항이 잘 지켜지는지’ ‘공정한 경쟁이 유지되는지’ 등의 상황을 직접 확인만 해 달라는 것입니다. 두 군데 관청의 담당자에게 2019 1 21부터 일정기간 지도·점검을 요청하였으나 답이 없습니다”
18	지역 내 협오시설 건립반대	“(특고압라인 매설은) 국책사업임에도 불구하고 시민의 동의 협조 없이 공사를 강행하다가 주변 민원 발생에 소극적으로 대응하면서 지금껏 몰래 공사하고 있음에” “OO구청이 OO구, □□구, △△구 3구의 폐기물처리장을 OO동에 짓겠다는 것은 공정성과 합리성 형평성이 없는 몰지각한 행정”
19	가짜뉴스 및 여론조작	“현재 대다수의 한국 언론은 뉴스에서 수많은 조회수를 받기위해서 자극적이고 검증되지 않은 무분별한 가짜뉴스를 올리고 있습니다” “언론은 엄청난 영향력과 권력을 가진 사적권력임에도 전혀 견제 받지 않고 있고 합당한 만큼의 처벌을 받지도 책임을 지지도 않고 있습니다”
20	안전사고	“왜 (교통 사망사고 피해자의) 유가족인 저희가 발 벗고 규명하려 하지 않으면 사고의 원인이나 상황조차 자세히 확인할 수 없는 건가요”
21	책임자의 불공정행위에 대한 미처벌	“국민의 생명을 담보로 달리는 열차의 신호 장치 납품 비리를 (중략) OO 기관의 핵심 간부들이 안전을 무시하고 무엇을 위해 업무를 잘하고 있는(납품 비리를 감사한) 감사실 감사자에게 그렇게(보복성 인사) 조치해야 했는지”
22	불법 사이트	“교묘하게 표절을 하고도 ‘법적 처벌을 받을 범위는 아니기 때문에’ 당당한 가해자들이 더 이상 당당할 수 없게 해주세요” “거의 나라가 방조하는 OO넷(유명 불법 동영상 사이트) 수준입니다. 해외서버 사이트라서 잡기 힘들다는 건 이번 워마드 사건을 계기로 핑계라는 것이 밝혀졌습니다. 경찰은 잡을 수 있었습니다. 하지만 잡지 않았던 것입니다”
23	공공임대 주택	“사회적 약자가 입주한 공공임대를 시세에 준하는 감정가액으로 분양전환 후 분양하겠다는 것은 결국 가진 자에게는 특혜를 없는 자들은 더욱 피를 뺏아 가겠다는 정책일 수밖에 없습니다” “대한민국에서 어떠한 사업자들도 현재 임대사업자들처럼 사업과 관련하여 부동산을 취득 보유 양도 소득세를 포함한 모든 세금에 대해 감면 또는 면제 비과세 해주는 사업은 없습니다
24	어린이집·유치원 운영 및 보육교사 처우	“국가와 지자체 재산인 국공립 어린이집 건물에는 피해를 안 주지만 민간어린이집과 가정어린이들은 감사에 걸리면 어린이집 정지 나아가서는 폐쇄 교사가 잘못해도 원장까지 자격정지 어린이집 폐쇄 하루아침에 길바닥에 나 있어야 하는 말입니까”
25	경찰·소방 공무원 처우	“치안현장 최일선에서 공무를 수행하던 중에 사망했음에도 업무와 사망과의 인과관계가 없다는 받아들이기 어려운 논리로 순직을 인정하지 않는 이런 어처구니 없는 상황에서 어떤 경찰관과 소방관들이 목숨바쳐 국민들의 생명과 재산을 보호해 주겠습니까 분하고 억울해서 심장이 터질”
26	검찰 개혁	“검찰과 법원의 비리는 누가 처벌합니까” “검찰은 지난 수십 년간 불법적인 피의사실 유포와 수사기밀 누설을 통해 정치에 개입하고 자신들 조직의 권력을 키워 왔음에도 제대로 처벌을 받은 적이 없습니다”
27	의료인 자격	“치과위생사는 법정 치과의료인력으로 3 4년의 치과전문 교육과정을 이수하고 국가시험을 통해 배출되는 유일한 인력으로 (중략) ‘의료기사 등에 관한 법률’에 따르면 (중략) 현재 진료현장에서 치과위생사가 수행하고 있는 다수의 치과진료보조 행위는 불법의 소지가 있는 행위로 이를 수행할 경우 법에 근거하여 치과위생사 면허 자격의 정지 처분을 받게 됩니다 “의료인으로서의 책임과 도리를 다하지 않는 의사들은 더 이상 다시는 병원을 열지 못하도록”
28	환경오염	“위험성을 염두에 두고 안전성 조사하여서 사업을 시행하였는지가 의문이고 또한 주민들의 의견을 공청회 등의 형식으로 수렴하여 찬성이 되서 시작된 사업인지도 의문이 듭니다” “(중국 미세먼지로 인해) 대한민국의 미래를 책임질 우리 아이들의 폐는 24시간 1군 발암물질을 들이마시며 병들어가고 있습니다. 우리의 모든 숨은 평등하고 소중한입니다”

IV. 논의

본 연구는 텍스트 마이닝 분석과 질적 텍스트 분석을 통합적으로 사용하여 청와대 국민청원 게시판의 게시글을 분석함으로써 청원 게시글의 주요 주제를 파악하고, 각 주제에서 드러난 불공정 경험과 인식을 탐색하였다. 그 결과 총 28개의 주제에서 한국인이 겪는 불공정 경험 및 인식에 대한 몇 가지 특성을 발견하였다.

첫째, 처벌과 엄벌은 다양한 주제에서 가장 많이 언급되었다. 명백한 피해에도 불구하고 적절한 처벌이 이뤄지지 않았다고 여길 때 즉, 응보적 정의가 구현되지 않을 때 불공정을 경험하고, 특히 범죄 가해자와 책임자에 대한 엄벌을 중시하는 것으로 나타났다. 본 연구 결과를 종합해 볼 때, 생명과 신체에 대한 범죄 및 성범죄 가해자, 부실하게 수사하는 조사·수사 담당 공무원, 무책임한 태도의 행정 공무원이 제대로 처벌받지 않을 때 강렬한 분노와 함께 불공정을 경험하였으며, ‘범죄 가해자 처벌’ 주제에서 드러났듯이 본인이 실제 피해를 입지 않았더라도 범죄자나 책임자에 대한 처벌이 부족하다고 여길 때 불공정을 경험하였다. 특히 범죄자 처벌의 경우 낮은 법적 처벌 기준, 가해자의 나이, 정신병 이력, 주취 상태로 인한 감형에 대한 강한 분노와 억울함을 보였으며, 책임자 처벌의 경우 비리, 감·경찰의 부실 수사, 솜방망이 처벌에 대한 분노가 나타났다. 가해자 또는 책임자의 엄벌 필요성과 정당성을 주장하기 위해 본인의 피해 내용과 정도를 상세히 기술하는 경향이 있었으며, 공정성 회복을 위해 진상규명이 반드시 이루어져야 한다고 보았다. 이러한 결과는 성폭행과 폭행범죄 사건에 대한 양형이 국민의 법감정과 맞지 않으며, 사법체계가 국민 정서에 합당한 응보를 반영하지 않는다는 정지송·서형국·노승국(2020)의 연구와 일치한다.

공정성의 회복과 관련하여 처벌, 엄벌 등 응보적 정의를 통한 공정성 회복은 거의 모든 주제에서 등장하였으나, 피해자의 진정한 회복을 위한 커뮤니티의 노력, 가해자의 사회 적응, 범죄 피해나 불공정 사건으로 인해 손상된 관계의 회복 등 회복적 정의를 통한 공정성 회복을 다룬 내용은 나타나지 않았다. 이를 볼 때, 많은 한국인이 정부에 기대하는 정의 구현은 회복적 정의를 위한 지원보다는 응보적 정의의 실현에 초점이 맞춰져 있는 것으로 보인다.

둘째, 응보적 공정성보다 적지만, 절차적 공정과 관련된 불공정 경험도 다수 나타났다. 절차적 공정과 관련된 경험은 그 대상이 크게 정부 기관 또는 공무원과 근무 중인 회사 또는 직장 상사로 나뉘었다. 정부 기관 또는 공무원과 관련하여, 공무원이 무성의한 태도로 같은 답변을 반복할 때, 수사 또는 조사 과정에서 피해자가 피해 사실과 함께 증거를 제출하였음에도 검·경찰이 수사를 제대로 이어가지 않을 때, 주민 투표 등 행정처리 과정이 투명하게 공개되지 않을 때 특히 불공정을 경험하였다. 회사 또는 직장 상사와 관련하여, 상사나 대기업의 갑질, 사내 비리 등을 내부 고발 후 인사 불이익을 당했을 때 불공정을 경험하였다. 절차적 불공정을 경험할 때 분노, 억울의 정서 경험과 함께 힘없는 한 개인으로서 무기력 또한 드러났다.

셋째, 분배적 정의와 관련한 경험은 비교적 적었지만, 그 종류는 다양하게 나타났다. 형평(equity)의 원칙 위배 시 경험한 불공정이 가장 많이 나타났는데, 이는 한국인은 분배의 원칙 중 형평의 원리를 중시한다는 선행연구와 일치한다(석현호 외 2005; 정한울 2018). 구체적으로, 비정규직, 사회복지사, 어린이집 교사 등 동일 노동에도 불구하고 차별적인 임금을 받을 때, 세금을 내지 않는 난민 또는 국내 체류 외국인을 정책적으로 지원할 때, 부자(부동산 보유)에 대한 세금이 낮다고 여길 때, 남성이 역차별을 당할 때 불공정을 경험하였다. 남성 역차별과 관련하여, 여성은 더 이상 사회적 약자가 아니므로 여성 지원정책은 불공정하다는 인식이 강하게 나타났는데, 이는 남성과 여성의 사회적 지위의 차이와 그 차이를 줄이려는 그간의 성 평 등 정책의 인식과는 차이가 있다(천관울·정한울 2019).

이에 더해, 10~30대의 주요 이슈인 ‘취업 및 입시의 평가 기준’ 주제에서 형평 관련 불공정 경험이 강하게 나타났다. 이들은 객관적인 시험을 바탕으로 하는 능력주의에 따른 위계 구조가 위배될 때 즉, ‘능력 또는 노력에 따른 불평등이 곧 공평한 것’이라는 형평(equity)의 원리가 위배된다고 지각할 때 특히 불공정을 경험하는 것으로 나타났다. 또한 사회적 자원을 분배하는 준거가 되는 능력을 측정하는 기준이 모호하거나 객관적이지 않다고 여길 때 불공정을 지각하였다. 이는 사회 내 위계 구조를 인식하고 계층 간 차이를 줄임으로써 사회의 기울어진 운동장을 바로잡아 공정과 평등을 회복하려는 입장 및 이에 기반한 현재의 사회 정책들과는 차이가 있다.

평등(equality)의 원칙과 관련하여, 정부 정책에 따른 서울 근교 신도시임에도 교통 혜택을 누리지 못할 때, 지역 내 공단 등으로 인해 지역 아이들이 다른 지역 아이들과 같이 일상을 누리지 못할 때 불공정을 경험하였다. 마지막으로 필요(needs)의

원칙과 관련하여, 장애인의 교육권과 사회 참여권이 제한받는다고 여길 때 불공정을 경험하였다.

이러한 결과는 다음과 같은 의미를 지닌다. 우선, 청원 게시판이라는 빅데이터를 활용하여 한국인의 불공정 경험의 구체적 맥락과 인식을 주제별로 파악하였다. 이를 통해 그간 철학적 논의나 설문 조사를 중심으로 이루어졌던 한국인의 공정성 인식조사(김상욱 외 2013; 장상수·김상욱·신승배 2015; 사회통합위원회 2012)와 달리 다양한 맥락에서 나타나는 불공정 경험을 파악할 수 있었다. 또한, 한국인의 공정성 인식에서 응보적 정의의 중요성을 확인하였다. 한국인의 공정성 인식조사는 분배적 공정성과 절차적 공정성을 중심으로 이루어져 왔으나, 본 연구를 통해 응보적 정의 또한 한국인의 공정성 인식에서 중요한 부분을 차지하고 있음을 확인하였다(석현호 외 2005; 정한울 2018). 마지막으로, 환경오염이라는 공정성 경험과 관련된 새로운 주제를 발견하였다. 환경오염 시설 건립 과정의 절차적 정의 이외에도, 환경오염이 심해지면서 깨끗한 공기가 하나의 가치 있는 자원임을 확인하였다. 따라서 깨끗한 환경 자원의 분배 역시 공정성 지각의 대상이 될 수 있음을 확인하였다.

또한 본 연구는 텍스트 빅데이터 분석을 위해 효과적이고 유용하게 사용될 수 있는 통합적 방법을 사용하였다는 점에서 방법론적 의의가 있다. 본 연구에서 사용한 혼합방법론은 연구 방법이 추론의 차원에서만 연계되는 요소설계가 아닌 연구 과정 전체에 통합되는 통합설계(Caracelli & Greene 1993)로 구성되었다. 따라서 절차상 생길 수 있는 여러 한계를 보완할 수 있었다. 본 연구의 한계점은 다음과 같다. 첫째, 한글 분석의 어려움이다. 한글은 어미나 조사가 많이 발달한 교착어이기 때문에 영어에 비해 전처리가 까다로운 편이다(안정국·김희웅 2014). 또한 동음이의어를 완벽하게 처리할 수 없다. 예컨대, ‘배’라는 단어가 특정 토픽으로 할당되었다면 단어 자체만으로는 교통수단을 의미하는지, 과일의 한 종류를 뜻하는 것인지 알 수 없다. 하지만 토픽 모델링의 특성상 동일 토픽 내 다른 키워드를 함께 비교함으로써 맥락을 파악하여 문제를 보완할 수 있을 것으로 기대한다.

둘째, 청와대 청원 게시판의 게시글은 시민들이 대한민국 정부를 대상으로 쓴 글이다. 따라서 정부 정책이나 정부 기관, 공무원을 대상으로 겪은 불공정 경험 또는 정부가 해결해 줄 수 있다고 기대되는 불공정 경험만을 담고 있다. 따라서 현재 상황이나 현재의 정부 정책에 만족하는 이들의 목소리는 발견할 수 없었으며, 일상에서 불공정이 발생하는 또 다른 큰 영역인 가족, 친구, 직장동료 등 사적 관계에서의 불공정 경험을 발견하는 데에는 한계가 있다.

본 연구의 결과를 바탕으로 몇 가지 향후 연구를 제안할 수 있다. 우선, 양적 연구 방법을 활용하여 한국인 표본을 대상으로 본 연구의 결과를 검증할 수 있다. 특히, 양적 조사 설문지 구성에 활용하거나, 인구통계학적 변인을 추가하여 불공정 경험과 인식에 대한 세대 차 또는 개인차를 검증한다면 한국인의 공정 경험과 인식에 대한 새로운 지식을 얻을 수 있을 것이다.

둘째, 가장 다양한 주제에서 가장 자주 나타났던 응보적 정의와 관련된 불공정 경험을 연구할 수 있다. 구체적으로, 청원 게시판에서 다수의 시민이 ‘법이 가해자들에게 그들이 저지른 범죄에 합당한 처벌을 하고 있지 않다’고 여기고 있었다. 향후 사법기관의 판결에 대한 국민의 법 감정을 보다 면밀히 조사하고, 범죄 유형에 따른 처벌 기준에 대한 국민의 법 감정과 처벌에 대한 눈높이를 파악함으로써 양형 기준 시스템 구축에 이를 참고할 수 있을 것이다(정지송·서형국·노승국 2020).

셋째, 본 연구 결과에서 발견한 정의 인식과 불공정 경험을 선정하여 현상을 더욱 깊이 이해하고, 논의를 확장할 수 있다. 예컨대 ‘남성 역차별’, ‘다문화 정책’, ‘취업 및 입시의 평가 기준’ 주제에서 나타난 정의 인식과 불공정 경험은 같은 주제를 ‘성 평등 정책’, ‘인도적 난민수용 및 인권 보호 정책’, ‘기울어진 운동장의 정상화’로 보는 기존의 인식과는 차이가 있었다. 따라서 현재 한국인에게서 나타나고 있는 이러한 불공정 경험을 구체적으로 누가, 언제, 어떻게 경험하며 왜 이러한 현상이 나타나는지 설명함으로써 한국인의 공정과 불공정 경험 및 인식에 영향을 미치는 차원들을 밝히고 그 심리적 메커니즘을 이해할 수 있을 것이다. 또한, 이러한 이해를 바탕으로 사회 갈등 해결에의 적용을 기대할 수 있다.

본 연구는 토픽 모델링과 질적 분석을 통합적으로 활용하여 국민 청원을 분석함으로써 기존 문헌 연구에서 살펴보기 어려웠던 한국인의 공정성 경험 및 인식을 탐색하고 결과를 조직화하였다는 점에서 의의를 지닌다. 이 과정에서 토픽 모델링은 효율적으로 자료의 크기를 축소하였으며 질적 분석은 토픽 모델링 결과를 활용하여 내용을 분석함으로써 개별 방법론의 한계를 보완하고 장점은 증진시켰다. 또한 본 결과는 국민 청원 데이터를 대상으로 진행하였기 때문에 전통적인 실험 혹은 조사 연구 등이 가진 제약에서 보다 자유로운 일상에서의 경험과 생각을 파악할 수 있었다. 본 연구결과가 향후 다른 후속 연구의 토대가 되길 기대하며, 또한 더 많은 분야에서 온라인 텍스트 자료 분석을 위해 토픽 모델링과 질적 연구를 통합적으로 적용할 것을 기대한다.

참고문헌

- 경향신문. 2017. “[100대 국정과제]문재인 정부 100대 국정과제 발표” 경향신문, 2017. 07.19. Retrieved from http://news.khan.co.kr/kh_news/khan_art_view.html?artid=201707191402001.
- 김남규·이동훈·최호창·William Xiu Shun Wong. 2017. “텍스트 분석 기술 및 활용 동향.” 《한국통신학회논문지》 42(2): 471-492.
- 김상욱·엄석지·윤영근·이명진·신승배·신동준·장상수·송유진·강정한·이선형. 2013. 《한국사회의 공정성 1990~2011》. 서울: 성균관대학교 출판부.
- 김지은·백순근. 2016. “텍스트 빅데이터 분석 기법을 활용한 대학구조개혁 평가의 쟁점 분석.” 《아시아교육연구》 17(3): 409-436.
- 김현중·이태현·유승의·김나랑. 2018. “민원 분석을 위한 텍스트 마이닝 기법 연구.” 《한국 산업정보학회논문지》 23(3): 13-24.
- 길호현. 2018. “텍스트마이닝을 위한 한국어 불용어 목록 연구.” 《우리말글》 78: 1-25.
- 문화일보. 2019. “한국사회 공정하지 않다 71.3%.” 문화일보 사회면, 2019.10.31.
- 박준형·오효정. 2017. “국내 기록관리학 연구동향 분석을 위한 토픽모델링 기법 비교” 《한국도서관정보학회지》 48(4): 235-258.
- 사회통합위원회. 2012. 《한국에서 공정이란 무엇인가》. 서울:동아일보사.
- 석현호·차존천·이정우·박종민·김명언·나은영. 2005. 《한국사회의 불평등과 공정성 인식의 변화》. 서울: 성균관대학교 출판부.
- 송준모·박영득. 2019. “청와대 국민청원에서는 무엇이 일어나는가?: 자연어 처리를 활용한 청와대 국민청원 분석.” 《한국정치학회보》 53(3): 53-78.
- 안정국·김희웅. 2014. “한글 감성어 사전 API 구축 및 자연어 처리의 활용.” 《한국지능 정보시스템학회 학술대회논문집》 177-182.
- 이종수. 2013. 《대한민국은 공정한가?》. 서울: 대영문화사.
- 장상수·김상욱·신승배. 2015. 《한국 사회의 불평등과 공정성 인식》. 서울: 성균관대학교출판부.
- 정지송·서형국·노승국 2020. “법은 가해자들에게 그들의 저지를 죄에 대한 합당한 처벌을 하고 있는가? - 웹툰<비질란테>를 통해 본 네티즌의 응보에 대한 여론 분석: 데이터 사이언스 분석방법을 활용하여.” 《한국 경찰연구》 19(3): 189-228.
- 정한울. 2018. 《한국사회 공정성 인식조사 보고서》. 한국리서치.
- 중앙일보. 2017. “[국정운영 5개년 계획] ‘국민의 나라 정의로운 대한민국’ … 문재인 정부

- 100대 과제 발표” 중앙일보, 2017.07.19. Retrieved from <https://news.joins.com/article/21770779>.
 - 중앙일보. 2018. “너무 나간 청와대 청원, ‘여론재판장’ 변질.” 중앙일보 사회면, 2018. 11.13. Retrieved from <https://news.joins.com/article/23118913>.
 - 천관율·정한울. 2019. 《20대 남자 ‘남성 마이너리티’ 자의식의 탄생》. 서울: 참언론 시사인북.
 - 한정훈. 2020. 《비교학적 관점에서 살펴본 청와대 전자청원제도》. 서울대학교 석사학위 졸업논문.
 - 황혜신. 2011. 《공정한 사회와 정부의 역할에 관한 연구》. 한국행정연구원.
 - KBS 뉴스. 2018. “‘국민이 물으면 답한다’ 靑 국민청원 500일. 시민이 준 점수는?” KBS 뉴스, 2018.12.29. Retrieved from <https://news.kbs.co.kr/news/view.do?ncd=4105211>.
 - 크레스웰, 존 지음, 조홍식·정선옥·김진숙·권지성 옮김. 2015. 《질적연구 방법론 -다섯 가지 접근》. 학지사.
-
- Antonetti, P. and D. Manika. 2017. “The Offline Spill-over of Signing Online Petitions against Companies: A Dual Pathway Model.” *Information Technology & People* 30(4): 969-990.
 - Atkinson, P. and A. Coffey. 1997. “Analysing Documentary Realities.” in D. Silverman (ed.). *Qualitative Research: Theory, Method, and Practice* (pp.45-62). London: Sage.
 - Bittermann, A. and A. Fischer. 2018. “How to Identify Hot Topics in Psychology Using Topic Modeling.” *Zeitschrift für Psychologie* 226(1): 3-13.
<https://doi.org/10.1027/2151-2604/a000318>
 - Blei, D.M., A.Y. Ng, and M.I. Jordan. 2003. “Latent Dirichlet Allocation.” *The Journal of Machine Learning Research* 3(1): 993-1022.
 - Blei, D.M. and J.D. Lafferty. 2009. “Topic Models.” in A. Sahami and M. Srivastava (eds.), *Text Mining: Classification, Clustering, and Applications*. (pp.71-94).
 - Caracelli, V.J. and J.C. Greene. 1993. “Data Analysis Strategies for Mixed-Method Evaluation Designs.” *Educational Evaluation and Policy Analysis* 15(2): 195-207.
 - Chang, J., S. Gerrish, C. Wang, J.L. Boyd-Graber, and D.M. Blei. 2009. “Reading Tea Leaves: How Humans Interpret Topic Models.” in Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, and A. Culotta(eds.). *Advances in Neural Information*

- Processing Systems* (pp. 288-296). Cambridge, MA: MIT Press.
- Creswell, J.W. and J.D. Creswell. 2018. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*(5th ed.). Los Angeles: SAGE.
- Feinerer, I. and K. Hornik. 2019. Package “tm.” Retrieved from <https://cran.r-project.org/web/packages/tm/tm.pdf>
- Fischer, R. 2016. “Justice and Culture.” in Sabbagh, C. and M. Schmitt(eds). *Handbook of Social Justice Theory and Research*(pp.459-475). NY, US: Springer.
- Griffiths, T.L. and M. Steyvers. 2004. “Finding Scientific Topics.” *Proceedings of the National Academy of Sciences*, 101(Supplement 1): 5228-5235. doi:10.1073/pnas.0307752101
- Gollwizer, M. and van J. Prooijen. 2016. “Psychology of Justice.” in Sabbagh, C. and M. Schmitt(eds). *Handbook of Social Justice Theory and Research*(pp.15-36). NY, US: Springer.
- Grün, B. and K. Hornik. 2020. Package “Topicmodels.” Retrieved from <https://cran.r-project.org/web/packages/topicmodels/topicmodels.pdf>
- Johnson, R.B., A.J. Onwuegbuzie, and L.A. Turner. 2007. “Toward a Definition of Mixed Methods Research.” *Journal of Mixed Methods Research* 1(2): 112-133.
- Karami, A., S.C. Swan, C.N. White, and K. Ford. 2019. “Hidden in Plain Sight for too Long: Using Text Mining Techniques to Shine a Light on Workplace Sexism and Sexual Harassment.” *Psychology of Violence*. Advance online publication. <https://doi.org/10.1037/vio0000239>.
- Karami, A., M. Lundy, F. Webb, and Y.K. Dwivedi. 2020. “Twitter and Research: A Systematic Literature Review through Text Mining.” *IEEE Access* 8: 67698-67717. doi:10.1109/access.2020.2983656.
- Kosinski, M., Y. Wang, H. Lakkaraju, and J. Leskovec. 2016. “Mining Big Data to Extract Patterns and Predict Real-life Outcomes.” *Psychological Methods* 21(4): 493-506. <https://doi.org/10.1037/met0000105>
- Kwartler, T. 2017. *Text Mining in Practice with R*. Hoboken, NJ: John Wiley and Sons.
- Liebe, U., P. Preisendorfer, and H.B. Enzler. 2020. “The Social Acceptance of Airport Expansion Scenarios: A Factorial Survey Experiment.” *Transportation Research Part D: Transportation and Environment* 84.
- Liebig, S. and C. Sauer. 2016. “Sociology of justice.” in Sabbagh, C. and M. Schmitt(eds). *Handbook of Social Justice Theory and Research*(pp.15-36). NY, US: Springer.
- Moustakas, C. 1994. *Phenomenological Research Methods*. Thousand Oaks, CA: Sage.

- Oh, J., A.E. Stewart, and R.E. Phelps. 2017. “Topics in the Journal of Counseling Psychology, 1963-2015.” *Journal of Counseling Psychology* 64(6): 604-6015.
- Pokorny, J.J., A. Norman, A.P. Zanesco, S. Bauer-Wu, B.K. Sahdra, and C.D. Saron. 2018. “Network Analysis for the Visualization and Analysis of Qualitative Data.” *Psychological Methods* 23(1): 169-183.
<https://doi.org/10.1037/met0000129>.
- Ponweiser, M. 2012. *Latent Dirichlet Allocation in R*(Diploma Thesis). Vienna University of Economics and Business, Vienna.
- Putnam, R.D. 1993. *Making Democracy Work*. Princeton, NJ: Princeton University Press.
- Richardson, G.M., J. Bowers, A.J. Woodill, J.R. Barr, J.M. Gawron, and R.A. Levine. 2014. “Topic Models: A Tutorial with R.” *International Journal of Semantic Computing* 8(1): 85-98. doi:10.1142/s1793351x14500044.
- Sabbagh, C. and M. Schmitt. 2016. “Past, Present, and Future of Social Justice Theory and Research.” in Sabbagh, C. and M. Schmitt(eds). *Handbook of Social Justice Theory and Research*(pp.1-11). NY, US: Springer.
- Steyvers, M. and T. Griffiths. 2007. “Probabilistic Topic Models”. in T.K. Landauer, D.S. McNamara, S. Dennis, and W. Kintsch(eds.). *Handbook of Latent Semantic Analysis*(pp. 427-448). Lawrence Erlbaum Associates Publishers.
- Silge, J. and D. Robinson. 2017. *Text Mining with R: A Tidy Approach*. O’Reilly Media, Inc.
- Sievert, C. and K. Shirley. 2014. “LDAvis: A Method for Visualizing and Interpreting Topics.” Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces(pp. 63-70). Retrieved from <http://www.aclweb.org/anthology/W/W14/W14-3110>.
- Teddlie, C. and A. Tashakkori. 2009. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. London: SAGE.
- Venkatesh, V., S. Brown, and H. Bala. 2013. “Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems.” *MIS Quarterly* 37(1): 21-54. Retrieved June 5, 2020.
- Wallach, H.M., I. Murray, R. Salakhutdinov, and D. Mimno. 2009. “Evaluation Methods for Topic Models.” Proceedings of the 26th Annual International Conference on Machine Learning-ICML’ 09. doi:10.1145/1553374.1553515.
- Wickham, H. 2020. Package “rvest.” Retrieved from <https://cran.r-project.org/web/packages/rvest/rvest.pdf>.

Yan, X., J. Guo, Y. Lan, and X. Cheng. 2013. "A Bitern Topic Model for Short Texts." Proceedings of the 22nd International Conference on World Wide Web-WWW '13. doi:10.1145/2488388.2488514.

<접수 2020.11.07; 수정 2020.12.21; 게재확정 2020.12.27>

A Study of Korean's Experiences of Unfairness Based on Analysis of Text Big Data Posted on the Blue House National Petition^{*}

Hyejin Yang

(Chung-Ang University)

Jungmin Ahn

(Chung-Ang University)

Taehun Lee

(Chung-Ang University)

This study investigated experiences and perceptions of unfairness for Koreans by analyzing texts posted on the Blue House National Petition through a combination of topic modeling and qualitative text analysis. Specifically, we applied Latent Dirichlet Allocation (LDA) and extracted latent topics and representative documents per topic from a total of 6,084 petitions that received at least thirty consents per day. A total of twenty eight topics were chosen for final interpretations based on the results from both qualitative and quantitative criteria. The analysis of keywords and representative documents per topic revealed important characteristics regarding Koreans' experiences and perceptions of unfairness. Unique contributions of this study can be summarized as follows: a strategy for effectively examining text big-data or a mixed method of topic modeling and qualitative analysis was proposed and profitably applied to find out characteristics of unfair experiences among Koreans, providing new directions for follow-up research.

Key words: Mixed Method, Text Mining, Topic Modeling, Latent Dirichlet Allocation, LDA, qualitative analysis, Blue House Online Petition, Text Big-Data

^{*} This research was supported by the Chung-Ang University Graduate Research Scholarship in 2018.