

청와대 국민청원 게시판을 통해 본

‘정의’의 현 주소

: Word2Vec/LDA 기법을 활용하여

간정현

이수진

이학민

정성원

목차

1. 서론

2. 본론

2.1. 연구 방법

2.2. 연구 내용

2.2.1. Word2Vec 모델링

2.2.2. LDA 토픽 모델링

3. 결론

4. 참고문헌

5. 부록

1. 서론

정의란 무엇인가? 현실에서의 정의는 한 마디로 규정할 수 없는 논쟁적인 개념이다. 정의 관념에 대한 합의, 혹은 정의가 현실에서 구현되는 방식에는 사회나 문화마다 차이가 있다. 예컨대 한국 사회에서는 한동안 ‘공정성’이 정의에 대한 논쟁의 중심이었다. 최근 몇 년 동안 박근혜-최순실 국정농단, 조국 사태, 오디션 프로그램 ‘프로듀스 101’의 조작 논란 등 우리 사회의 공정성을 의심하게 하는 사건들이 연이어 발생했으며 많은 이들이 분노를 표했다. 사실 무한 경쟁의 능력주의 사회에서 공정성은 개인의 생존이 걸린 문제다. 이기지 못하면 도태되는 사회에서 공정하지 않은 경쟁에 뛰어드는 것은 위험한 행동이다. 2017 년 국가공무원 9 급 공채 시험 응시자는 약 23 만 명으로 역대 최대였다.¹ 다른 영역에서 보장되지 않는 경쟁의 공정성이 공무원 시험에서만큼은 확실하게 보장되기 때문인지도 모른다.

하지만 위에서 언급했듯 정의는 논쟁적인 개념이며 같은 사회와 시대 안에서도 여러 정의 개념이 경쟁할 수 있다. 공정성이 한동안 한국 사회의 화두였지만, 누군가는 정의에서 경쟁 과정의 공정성만큼 다른 가치 역시 중요하다고 주장할 수 있다. 문재인 대통령은 취임사에서 “기회는 평등할 것이다. 과정은 공정할 것이다. 결과는 정의로울 것이다.”라고 말했다. 이 말에는 정의가 단순히 기회의 평등과 과정의 공정함을 보장하는 것 이상이라는 인식이 깔려 있다. 이처럼 정의에 대한 다양한 관점들은 그 관심사와 방법론이 사뭇 다르고, 같은 사람이라도 서로 다른 분야에서 다른 방식의 정의가 구현되어야 한다고 주장할 수도 있다.

그렇다면 현대 한국사회의 정의 개념은 구체적으로 어떤 의미들을 지니고 있을까? 본 연구에서는 청와대 국민청원 플랫폼에 게시된 청원들을 중심으로 이러한 질문에 답해보고자 한다. 청와대 국민청원 플랫폼은 문재인 정부가 2017 년 ‘국민이 물으면 정부가 답한다’는 기조를 바탕으로 개설하였다. 청와대 국민청원 플랫폼은 국민적 관심사에서 멀어졌었던 국민의 기본적인 권리인 ‘청원권’을 되살렸다. 청원권은 제헌헌법에서부터 인정된 국민의 기본권 중 하나였음에도 불구하고, 청원권을 구체화한 ‘청원법’은 1963 년 제정 이후 40 년 후에 비로소 1 차 개정이 이루어지는 등, 청와대 국민청원 플랫폼 등장 전에는 ‘청원권’이라는 권리 자체는 그 존재를

¹ "오늘 9급 공무원시험...‘역대 최대’ 25만명 응시," *매일경제*, n.d. 수정, 2020년 1월 27일 접속, <https://www.mk.co.kr/news/society/view/2017/04/238867/>

상실한 것이나 다름없을 정도로 여겨졌다. 국민청원 플랫폼의 등장 이후 비로소 청원권이라는 기본적 권리가 재조명되었고, 현재 국민청원 플랫폼은 일종의 여론 형성 기능을 하며 공론장의 역할 또한 수행하고 있다(김병록, 2019).²

청와대 국민청원 플랫폼을 연구의 대상으로 삼은 이유는, 정치에 대한 불신에도 불구하고 국민들은 정의를 실현할 수 있는 강력한 수단으로서 여전히 정치에 의존하고 있기 때문이다. 격변의 근, 현대 사회 속 내적 및 외적으로 발생했던 여러 논란을 거치며 형성된 한국식 민주주의는 정치결정론과 정치혐오증이라는 양극단의 태도를 형성했다(김일수 외, 2015).³ 즉 정치인들을 신뢰하지 않고 앞으로의 한국 사회에 대해 별 다른 기대감을 갖지 않는 태도와, 그럼에도 불구하고 사회적 부조리를 마주할 때 결과적으로 정치적 행동을 통해 이를 해결하고자 하는 태도가 공존하는 것이다. 국민들이 아직 정치를 정의를 실현할 수단으로 여기고 있다면, 타 인터넷 공간에 비해 정치적 영향력을 강하게 행사할 수 있는 청와대 국민청원 게시판이 정의에 대한 국민적 정서를 잘 드러낼 것이다. 또한 청와대 국민청원 플랫폼에서는 정의에 대한 관념이 추상적인 수준에서 머무르지 않고 구체적인 요구들에 의해 표현된다. 어떠한 법을 제정해달라거나, 특정 범죄의 처벌의 강화해달라는 등의 요구들에서 국민들이 가진 정의에 대한 관념을 보다 구체적으로 읽어낼 수 있을 것이라고 판단하였다.

따라서 본 연구는 청와대 국민청원에서 수집한 텍스트 데이터를 기반으로 현 시대의 국민들이 생각하는 “정의”라는 개념에 대해 경험적인 분석을 진행하고자 한다. 43 만 건에 이를 정도로 방대한 데이터를 다루는 작업이기 때문에 일일이 모든 청원을 읽고 분석하는 것이 현실적으로 어렵다. 그렇기 때문에 최근 각광받는 Python 을 이용한 자연어 처리 기법(NLP: Natural Language Processing)을 활용하였다.

논의를 진행하기 앞서 청와대 국민청원에 대해 간략하게 설명하면 대한민국 국민 누구나 청원을 등록할 수 있고, 청원 게시일로부터 30 일 이내에 20 만 건 이상의 청원 동의를 받은 글들에 한해서 행정부 관계자들의 답변이 등록된다. 청원 동의는 페이스북, 네이버, 카카오톡, 트위터 계정을 통해 가능하며 한 사람이 여러 계정을 통해 청원 동의하는 것이 가능하다. 2020 년

² 김병록. (2019). 청와대 국민청원의 개선방안에 대한 연구. Chosun Law Journal Vol26 No.2 Aug 2019, 139-170.

³ 김일수 외 11인. (2015). 한국사회 정의 바로 세우기. 세창미디어. 36-38.

1 월 23 일까지 총 135 건의 청원이 답변을 받았다. 게시자는 글을 작성할 때 정부에서 지정한 17 개의 카테고리 중 하나를 선택해야한다. 상세한 카테고리 목록은 아래의 <표 1>에서 확인할 수 있다. 청와대 측의 통계 조사에 따르면 17 세부터 44 세 사이의 연령대의 청원 게시판 이용률이 75.8%로 가장 높았다.⁴

2. 본론

2.1. 연구 방법

본 연구에서는 먼저 청와대 국민청원 게시판에 2017 년 8 월 19 일부터 2019 년 10 월 8 일까지 게시된 43 만여 건의 청원 게시글의 “게시날짜”, “본문 내용”, “제목”, “카테고리”, “참여인원”을 크롤링하였다. 그 다음 본문의 주제를 선명하게 파악하고 차후에 사용할 기법들을 원활히 수행하기 위해 명사 단어만을 추출하고 불용어를 제거하는 전처리를 수행하였다. 이후 Word2Vec 모델을 활용하여 “정의”와 유사한 맥락에서 사용된 단어들을 추출하고 분석하였다. Word2Vec 모델은 언어학의 분산 가설을 따라 비슷한 의미의 단어는 비슷한 문맥에서 발생한다는 점을 근거로 단어를 벡터공간에 임베딩하며, 일반적으로 단어의 동의 관계를 잘 반영한다고 알려져 있다(강형석, 양장훈, 2019).⁵ 이후 Word2Vec 에서 “정의”와 가장 가까운 맥락에서 사용된 50 개 단어를 포함하는 게시물들을 대상으로 토픽 모델링(Topic Modeling)을 시행하고 게시물들을 주제별로 분류하였다. 토픽모델은 Blei 등(2003)의 LDA(latent Dirichlet allocation) 알고리즘을 기반으로 한 확률 모델로서 어떤 주제들의 집합이라고 가정된 한 문헌을 구성하는 단어들을 확률적으로 계산하여, 이 결과 값을 토픽 주제어들의 집합으로 추출하는 알고리즘이다(심준식, 김형중, 2017).⁶ 이 결과를 통해서 한국인이 생각하는 “정의”는 무엇이며 어떤 사안에 대해 “정의”의 문제라고 생각하는지에 대해 고찰하였다.

모든 분석 과정에는 Python 언어 및 패키지를 활용하였다. 자연어 처리 (NLP)에는 여러 방법이 존재하며 Python 에서는 자연어 처리를 위한 다양한 패키지를 제공한다. 본 연구에서는 한국어

⁴ 박정엽, "靑청원 보니...대학생은 '조국 임명 찬성' 대열서 벗어나 있었다", 조선일보, 2019.11.06.

⁵ 강형석, 양장훈. (2019). Word2vec 모델로 학습된 단어 벡터의 의미 관계 분석. 정보과학회논문지, 46(10), 1088-1093.

⁶ 심준식, 김형중. (2017). LDA 토픽 모델링을 활용한 판례 검색 및 분류 방법. 전자공학회논문지, 54(9), 67-75.

텍스트 데이터 전처리에 특화된 Python 패키지인 KoNLPy 와 워드 투 벡터 및 토픽 모델링을 위한 다양한 기능을 제공하는 gensim 을 사용했다.

2.2. 연구 내용

2.2.1. Word2Vec 모델링

본 절에서는 Word2Vec 모델링을 활용한 유의어 분석을 통해 “정의”의 관념을 분석했다. Word2vec 모델링을 시행하기에 앞서 “정의”의 개념이 청원 카테고리별로 유의미한 차이점을 보일 것이라고 가정했다. 이러한 가정에 따라 전체 데이터를 대상으로 모델링을 시행한 것과 별도로 카테고리별로도 시행했다. 그리고 빈도수가 높은 50 개의 단어를 피봇 테이블로 정리했다. 결과는 부록의 <표 1>과 같다.

<표 1>의 결과를 통해 전체 데이터에서 “정의”와 비슷한 맥락에서 쓰인 단어들을 알 수 있는 한편 카테고리별로 “정의”와 연관되는 어휘가 조금씩 다르다는 것을 한 눈에 확인할 수 있다. 먼저 전체 데이터를 대상으로 분석하였을 때 두드러지는 특징을 네 가지로 나누어 볼 수 있다. 첫째로 “공의”, “공정”, “권선징악”, “구현”, “법치”, “공공선”, “민주” 등의 단어가 “정의”의 유의어로 빈출한다는 점이다. 청와대 국민청원은 이전 정부에서 발생한 국정 농단과 불통에 대한 국민적인 시정 요구, 공론장의 필요성이 대두되던 시점에 개설되었다. 해당 키워드들이 게시글 전반에 “정의”와 가까운 어휘로 등장한 이유는 이러한 배경 속에서 설명될 수 있다. 이승원 외(2018)에 따르면, 청와대 국민청원 플랫폼의 활성화는 박근혜 대통령의 권위적 통치 하에서 위축되었던 정치 효능감이 촛불집회를 기점으로 정치적 자신감을 회복하고 참여에 대한 긍정적인 태도가 확산된 것으로 해석할 수 있다.⁷ 즉 청와대 국민청원 플랫폼의 게시물들은 권력자들에 의해 농락당하고 부정당한 민주주의의 핵심 가치들이 다시 우리 사회에 “구현”되길 바라는 국민들의 소망을 드러내며, 이것이 문제인 정부 하에서 국민들이 바라는 “정의”의 주안점이라고 분석할 수 있다.

둘째로 “공의”, “평등”, “공공선” 등의 키워드에서 국민들이 생각하는 정의의 모습은 사회정의 및 공동체 중심의 가치라고 해석할 수 있다. 국정 농단을 겪으며 국민들은 한국사회가 정치 영역을

⁷ 이승원, 임한샘, 이현우 (2018). 청와대 국민청원 참여요인 분석. OUGHTOPIA, 33(3), 111-141

비롯한 전반적인 사회 영역에서 여전히 극단적 이기주의의 모습들을 벗어나지 못했음을 명확하게 직시했으며 불공평하다는 ‘개인적 느낌’이 ‘공동체의 문제’로서 실감하게 되었다(김철, 2011).⁸ 이러한 공동체의 문제를 국민들은 촛불집회 같은 공동 행동을 통해 정치권에서 등장한 극단적 이기주의를 일차적으로 심판하는데 성공했고 새로 출범한 정부에서 여전히 잔존하는 불의를 청산하길 원하는 바람이 “권선징악”이라는 키워드를 통해 드러난다.

반면 개인적인 수준에서의 정의에 대한 언급은 상대적으로 적다. 전체 청원을 대상으로 훈련된 Word2Vec 모델에서 “정의”와 “개인”과의 코사인 유사도⁹는 0.2617, “공동체”와의 코사인 유사도는 0.4065, “국민”과의 코사인 유사도는 0.4403 이다. 즉 정의는 국가 혹은 공동체에서 동떨어진 개인으로서의 인간보다는, 하나의 공동체나 국가에 속한 일원으로서의 인간과 보다 큰 연관을 맺는 개념으로 사용되었다. 또한 “정의”와 “권리”와의 코사인 유사도는 0.3603, “자유”와의 코사인 유사도는 0.4647, “공공선”과의 코사인 유사도는 0.6499 이다. 통상적으로 자유주의와 공동체주의가 대립하는 논점인 개인 대 공동체, 자유 대 공동선의 구도에서 볼 때 정의는 공동체주의적 어휘들과 함께 사용된 것이다. 이러한 결과는 국민들이 정의를 주로 자유주의적이라기보다 공동체주의적 개념으로 파악한다는 의미로 해석된다.

셋째로 카테고리에 무관하게 포진한 “법”과 관련된 키워드에 주목할 수 있다. 정의와 관련해서 “법치주의”, “법치”, “법질서” 등의 키워드들이 50 건 이상 등장하고 있다. 이를 통해서 국민들이 가장 열망하는 정의 구현의 방식이 “법”에 의한 것임을 읽어낼 수 있다. 위에서 “정의”와 가까운 단어들로 언급된 민주주의의 핵심 가치들, 그리고 법이라는 키워드는 우리 헌법의 기본 원리인 민주주의와 법치주의에 대응한다. 즉 국민들이 생각하는 정의가 우리 헌법의 기본 원리를 수호하는 데에 있음을 짐작할 수 있다. 물론 청와대 국민청원 플랫폼에서는 양천수(2018)가 지적하다시피 오히려 민주주의와 법치주의의 충돌이 드러나기도 한다. 양천수는 일부 청원에서 민주주의가 대중주의로 타락하는 경우를 발견할 수 있으며, 대중주의적 민주주의가 법치주의의 제도적 장치들을 공격한다고 말한다. 실제로 청와대 국민청원 플랫폼에 게시된 청원들 중에는

⁸ 김철, (2011), 정의란 무엇인가 - 자유주의와 공동체주의의 가치, 자유와 평등, 형평, 사회이론 제 39호 40

⁹ 코사인 유사도는 두 벡터가 이루는 사이 각의 코사인 값을 통해 두 벡터 간의 유사도를 측정하는 지표이다. 1에 가까울 수록 유사도가 높다.

‘불이익재심’, 판사에 대한 파면 요청 등 법치의 원리에 반하는 청원들이 존재하며, 양천수는 대중민주주의가 강화될수록 이러한 문제들이 더욱 빈번하게 발생할 것이라고 우려한다.¹⁰

그러나 “정의”와 법치가 유사한 맥락에서 사용된다는 점을 고려하면 법치주의에 반하는 청원들을 곧바로 법치주의를 거부하는 청원이라고 해석하기는 어렵다. 보통의 국민들에게 법치는 당연한 가치이기는 하지만, 모든 이들이 법의 내용을 상세히 알고 있는 것은 아니다. 즉 법치주의에 반하는 청원은 법치주의를 공격하려는 의도를 가졌다고보다, 대부분의 국민들이 법치에 대한 전문적 지식을 갖지 못했기 때문에 발생한다. 국민들은 법에 정확히 어떤 내용이 기록되었는지는 모르지만, 법은 정의로운 것이고 법대로 하면 정의로운 사회가 구현될 것이라고 기대한다. 법이 이러한 신뢰마저 잃어버리는 상황 역시 대중민주주의의 심화와 함께 우려해야 할 상황이다. 전체 청원을 대상으로 훈련된 Word2Vec 모델에서 ‘법’과 가까운 키워드로 “정서법”, “사문화”가 각각 3 위, 12 위로 꼽혔다.¹¹ 대중민주주의를 경계하고 법 체계의 신뢰 회복을 위해서라도, 법은 국민들에게 조금 더 친숙하게 다가갈 필요가 있다.

마지막으로 정의를 기회의 평등과 연관 짓는 경향을 확인할 수 있다. “공정”, “공평”, “만인”, “평등”, “기회균등”, “출발점” 등의 키워드는 다수의 카테고리에서 상위 에 위치한다. 이를 통해 관념으로서의 정의가 경험적으로는 ‘모든 이에게 동일한 출발선이 주어지는 것’으로 구체화된다는 해석을 할 수 있다. 앞서 언급하였던 최근 한국 사회에서 불거진 공정성과 관련한 일련의 사건들은 ‘공평하지 않은 조건에서 이루어지는 경쟁’에 대한 문제제기라는 공통점을 갖는다. 이 사실과 Word2Vec 을 통한 유의어 분석을 종합적으로 고려했을 때, ‘동등한 기회’라는 키워드가 최근의 한국 사회의 정의 관념에서 핵심적인 위치를 점하고 있다는 분석이 가능하다.

한편, 위에서 설정했던 가설처럼 카테고리별로 “정의”와 연관되는 어휘의 차이가 나타난다는 것도 발견할 수 있다. 가령 “일자리” 카테고리에는 “평등”, “공정”, “공평”, “균등” 같은 어휘가 높은 빈도로 드러난다. 취업난이 큰 문제로 대두되고 있는 현 시점에서 일자리 분배의 공정성을 정의로서 요구하고 있음을 짐작해볼 수 있다. “반려동물” 카테고리에는 다른 카테고리에서는 나타나지 않는 “육류”, “채식주의”라는 키워드가 등장한다. “반려동물” 카테고리에 한정하면

¹⁰ 양천수. (2018). 민주적 법치국가의 내적 갈등. 법학연구, 28(3), 271-305.

¹¹ 기타 법과 가까운 단어들은 '법규'(1위), '법률'(2위), '이법'(4위), '노동법'(5위), '법령'(6위), '일반법'(7위), '법조문'(8위), '현행법'(9위), '이법'(10위), '국법'(11위) 등이 꼽혔다.

비인도적 도축이나 동물 학대 등의 동물권 이슈가 정의의 문제로 대두된다는 것을 알 수 있다. “저출산 고령화대책” 카테고리에서도 “자국민”, “애국”, “부강”, “잡종” 등의 특이한 키워드들이 등장하는데 해당 카테고리에서는 민족주의적인 이슈들이 정의와 관련된 것으로 인식된다는 것을 생각해볼 수 있다.

2.2.2. LDA 토픽 모델링

앞선 절에서 유의어 분석을 통해 “정의” 관념을 해부하였다면, 본 절에서는 LDA 토픽 모델링 기법을 활용하여 어떠한 주제들이 “정의”의 문제로 다루어졌는지를 살펴보았다. 위에서 Word2vec 을 시행한 결과로 도출된 어휘가 포함되어있는 문서를 “정의”와 관련한 국민청원 게시글로 가정하고 해당 문서들을 대상으로 토픽 모델링을 실시하였다. 토픽 모델링을 실행하기 앞서 최적의 토픽 개수를 설정해주기 위해 coherence value 를 계산했고 그 결과 14 가 도출되었다.^{1 2} 아래의 결과는 토픽모델링을 실시한 결과를 표로 정리한 결과이며 각 토픽의 이름은 빈출된 단어들을 살펴본 후 임의로 선정했다.

<표 2>

Topic #	Topic	Words
0	정부부처개혁	사건, 수사, 조사, 사실, 검찰, 증거, 검사, 경찰, 판사, 방송
		“온라인 민주주의를 교묘하게 억압하는 방송사의 행태...방통위의 강력한 조치를 요구합니다.” “검사개혁 공수처 좀 빨리 신설해주시시오.” “법원 판결문을 모두 공개해주세요”
1	행정	공무원, 행정, 일반, 유공자, 독립, 인간, 자활, 거짓말, 보훈, 예산
		“지금이라도 월남전투수당 참전용사에게 돌려줘야” “국회의원 특권 내려놓기 청원” “(내용추가)헌법개정 반대합니다!! 일반행정공무원과 운전공무원을 폐지해야 이 나라가 잘 살아갑니다!!”
2	토지/주택	민원, 공사, 아파트, 처리, 주민, 사업, 분양, 답변, 제조, 행정
		“노무현 전대통령 공원녹지법” “새마을도로 잘못된 행정을 바로잡아주시시오.” “그린벨트 불편한점을 국민의 입장에 서서 해결해주시시오” “청주 주택 주차장”

^{1 2} 응집성 지수를 뜻한다. 혼잡도(perplexity) 와 함께 LDA 모델의 성능을 평가하는 지수로서, 응집성 지수는 토픽의 일관성을 나타내는 지표이다. 동일한 토픽으로 분류된 문서에 사용된 단어들의 유사도를 기반으로 계산된다. 따라서 Coherence score가 높을수록 의미론적 일관성이 높은 분류라고 할 수 있다. (이새미, 홍순구, 2019)

3	외교	북한, 일본, 정치, 미국, 자유, 정권, 민주당, 역사, 경제, 민주주의
		<p>“군사 훈련”</p> <p>“북한의 비핵화 조건은 체제유지!”</p> <p>“평창올림픽 동시입장 단일팀 철회해주세요.”</p>
4	복지	병원, 보험, 지급, 헌법, 연금, 급여, 복지, 의사, 제조, 환자
		<p>“기초연금 발란스 맞춰라/국민연금 가입 저지.”</p> <p>“장애연금 지급 대상자의 국민연금 급여 공제 부당성”</p> <p>“1형당뇨병 환자 실비보험 상해보험 암보험등 들수있게 해주세요”</p>
5	교육	아이, 학교, 교육, 학생, 부모, 시간, 교사, 대학, 유치원, 엄마
		<p>“대한민국의 교육방식을 바꿔주세요.”</p> <p>“유아교육에 들어가는 비용지원을 촉구합니다.”</p> <p>“수시 정시의 기준이 없어지는 것 같아여”</p>
6	정보통신	사용, 업체, 정보, 사이트, 운영, 회사, 차단, 게임, 판매, 불법
		<p>“웹사이트 원문을 가리는 광고 및 팝업 금지”</p> <p>“휴대폰 없는 사람은 대한민국 국민이 아닙니까?”</p> <p>“네이버검색조작 사이트웹문서통합폐지할수있게 해주세요”</p>
7	국회	의원, 차량, 국회의원, 서울, 선거, 지역, 지방, 국회, 택시, 시민
		<p>“누구를 위한 선거 입니까?”</p> <p>“선출직 공무원(대통령 국회의원 시도의원 및 지자체장) 재 보궐선거 선거 비용 문제”</p> <p>“박원순시장님의 탈당을 청원합니다.”</p>
8	성평등/인권	여성, 남성, 남자, 시간, 여자, 의무, 평등, 가족, 차별, 군대
		<p>“국가에 의한 노동착취 행위를 멈춰주세요.”</p> <p>“국방의 의무 남성만의 의무인가요?”</p> <p>“성소수자 차별금지법 제정을 청원합니다.”</p>
9	경제	부동산, 기업, 시장, 세금, 회사, 금융, 서민, 경제, 정책, 거래
		<p>“주택담보대출의 원리금 상환 규정을 개선해주세요.”</p> <p>“중부세 납부 기준 완화”</p> <p>“저축은행의 중금리대출 가계대출총량규제 에 포함시켜주십시오.”</p>
10	경찰	경찰, 범죄, 인권, 불법, 범죄자, 폭행, 집단, 경찰관, 개혁, 부패
		<p>“대한민국이 OECD 국가중 자살률이 가장 높은 이유는?”</p> <p>“부패한 경찰의 사리사욕 때문에 대한민국이 OECD 국가중 자살률이 가장 높습니다.”</p> <p>“대한민국은 부패한 경찰 연놈들의 사리사욕을 위한 그들만의 범죄천국입니다.”</p>
11	환경	미세먼지, 중국, 경제, 기업, 정책, 외국인, 산업, 일자리, 세계, 환경
		<p>“원전폐지를 허락해주세요”</p> <p>“미세먼지 대책 청원합니다.”</p> <p>“대한민국 대기오염 환경개선해주세요.”</p>
12	범죄 처벌	피해자, 처벌, 가해자, 피해, 사건, 인간, 정신, 범죄, 동물, 가족
		<p>“살인죄에 대한 법개정 청원”</p> <p>“음주로 인한 사건에는 가중처벌을 하여주십시오.-술을 마시지 않은 이들의 안전할 권리를 위하여.”</p> <p>“성폭력 범죄자들의 처벌이 강화되길바랍니다. 국민 모두가 거리를 안전하게 거닐 수 있도록 성폭력 성폭행 범죄자들의 처벌이 강화되길 바랍니다.”</p>

13	고용	제도, 기관, 시험, 업무, 필요, 채용, 평가, 전환, 경우, 고용
		“공공기관 비정규직 근로자의 정규직 전환 보완” “무분별한 정규직 전환을 반대합니다.” “정부출연연구기관 비정규직 정규직화”

- 청원 사례는 해당 토픽에의 소속 확률을 1 순위, 의제의 다양성을 2 순위 기준으로 선택하였음. Raw Data 의 맞춤법 오류나 오타는 수정하지 않고 그대로 기재하였음.

토픽 모델링 결과 분류된 토픽별 주제어를 확인하여 “정부 부처 개혁”, “행정”, “토지/주택”, “외교”, “복지”, “교육”, “정보통신”, “국회”, “성평등/인권”, “경제”, “경찰”, “환경”, “범죄 처벌”, “고용”이라는 토픽명을 붙였다. “정의”와 관련된 것으로 분류된 국민청원 게시글 95,844 개 중 가장 큰 비중을 차지한 것은 20.4%의 “외교”로, 가장 작은 비중을 차지한 토픽은 0.9%의 “행정”으로 나타났다(부록 <그림 1> 참조). “외교” 토픽에는 <표 2>에서 확인할 수 있듯이 북한과 관련된 게시글이 주를 이루며, 이 외에도 일본과 미국 등 국제 관계와 관련된 게시글들이 “외교” 토픽으로 분류되었다. 두번째로 높은 비중을 차지한 “정부 부처 개혁” 토픽은 방송통신위원회, 검찰, 사법부 등의 정부 부처의 개혁을 요구하는 게시글들로 이루어졌다. 세번째로 높은 비중이자, 전체 비율에서 10%를 넘은 마지막 토픽인 “경제” 토픽에는 주택담보대출 및 중금리 대출, 중부세를 비롯한 각종 세금, 가상화폐, 최저임금 등을 주제로 하는 글들이 분류되었다.

여기서 특기할 점은 일반적으로는 “정의”와 연관 지어 이야기되지 않는 의제들이 정의의 서사로 재구성되는 양상을 확인할 수 있다는 것이다. 이를테면 “외교” 토픽으로 분류된 글 중 한미 군사 협동 훈련을 지속할 것을 자유민주주의 및 정의 수호의 측면에서 주장하거나, “경제” 토픽으로 분류된 글 중 저축은행의 중금리대출 규제를 업계 간 형평성 및 경제 영역에서의 정의 실현으로 프레임(Framing)한 사례를 들 수 있다.

한편, “정의”와 관련된 국민청원 게시글 중 5.7%를 차지하는 토픽인 “고용”에서 흥미로운 사실을 포착할 수 있었다. 문재인 정부의 대대적인 정규직 전환을 주제로 한 청원글이 주를 이루었는데, 해당 정책을 지지하는 입장과 반대하는 입장이 동시에 확인되었으나, 후자가 더 많은 것으로 나타났다. “무분별한 비정규직 전환은 또 다른 차별”이라고 주장하는 글과 “청년인턴 채용비리”를 규탄하는 글, “고졸채용 감축”, “취업 기회는 공평해야”된다는 글들을 종합하면, 고용 영역에서의 정의 관념의 구체적인 윤곽이 드러난다. 그것은 결과적 평등보다는 기회의 평등을 중요시하는 경향이다. 결과적 평등에 있어서는 오히려 그것을 적극적으로 거부하는 입장을

취하는 듯 보이기도 한다. 블라인드 채용에 대한 반대나 대입 전형에서 정시에 대한 확대가 설득력을 얻는 것도 같은 맥락이다.

3. 결론

본 연구에서 드러난 정의 관념의 특징은 네 가지로 정리해볼 수 있다. 첫째로 박근혜 정부에 대한 심판 이후에 나타난 “권선징악”적 정서가 반영된 정의 관념이다. 둘째로 개인보다는 공공선에 초점이 맞추어진 정의 관념이다. 셋째로 민주와 법치의 회복과 구현에 대한 소망을 드러내는 정의 관념이다. 마지막으로 기회의 평등에 대한 정의 관념이다. 이러한 정의의 관념이 반영된 청원 게시물을 14 가지의 토픽으로 분류했다. 그 중 가장 큰 비중을 차지한 것은 외교, 정부부처 개혁, 경제였다.

최순실 국정 농단 사태와 박근혜 대통령의 탄핵 그리고 문재인 정부의 출범으로 이어지는 일련의 사건 이후로 한국의 민주주의는 새로운 전기를 맞이했다. 현대사를 지나오면서 꾸준히 제기되었던 정치권과 국민 간 소통의 문제를 해결하기 위해 문재인 정부는 ‘청와대 국민청원’이라는 실시간, 양방향의 인터넷 플랫폼을 구현했고 이를 통해 국민들의 다양한 생각과 감정들이 비교적 정제된 형태의 텍스트로 생산되었다. 이러한 배경은 정치 분야뿐 아니라 다양한 인문학, 사회과학적인 경험적 연구의 좋은 재료를 제공해주었다.

본 연구는 이러한 재료를 바탕으로 현시대의 “정의”라는 개념을 경험적으로 분석하였다는 것에 의의가 있다. 특히 자연어 처리 기술(NLP)의 발달로 대량의 텍스트 데이터를 짧은 시간 내에 효율적으로 분석할 수 있게 된 배경과 맞물리며 일부 텍스트 연구를 통한 성급한 일반화를 배제하고 전수 연구를 통해 설득력 있는 해석을 제시할 수 있었다는 점에서 의의가 크다. 이를 통해 후속 연구에 사용할 수 있는 2 차 자료를 제공할 뿐만 아니라 이후 정부가 국민의 의견을 반영하는 정책을 수립할 때 실제로 국민이 원하는 “정의”가 무엇인지 재고해 볼 수 있는 참고자료로 활용할 수 있을 것이다.

연구를 좀 더 발전시키거나 후속 연구를 진행할 때 좀 더 보완해야할 점은 다음과 같다. 먼저 청와대 국민 청원이라는 플랫폼 자체의 편향성에 유의해야한다. 정부가 주도하는 플랫폼의 한계로 이용자들의 성향이 친정부적 성향을 띠는 가능성이 높고 이로 인해 청와대 국민청원 텍스트 데이터만을 가지고 선불리 일반화하는 것은 어려울 수 있다. 더구나 청와대 국민청원 게시판에 글을 쓸 때 정부에서 제시하는 몇 가지 규칙이 존재하며 이를 위반한 게시물에 대해 숨김 및 삭제 처리를 하고 있다. 2019 년 3 월 31 일 이후부터는 사전에 100 개 이상의 동의를 받아야만 정식으로 청원글을 게재할 수 있게 바뀌면서 국민의 실제 목소리를 반영하지 못한다는 비판도 늘어나고 있다. 따라서 이러한 한계점을 보완하기 위해 청와대 국민 청원 이외에 국민의 실제 목소리를 반영할 수 있다고 판단되는 다른 플랫폼과의 비교연구를 제안해 볼 수 있겠다. 앞으로의 후속연구를 기대하며 본 논문을 마친다.

4.참고문헌

- "오늘 9 급 공무원시험...‘역대 최다’ 25 만명 응시," *매일경제*, n.d. 수정, 2020 년 1 월 27 일 접속,
<https://www.mk.co.kr/news/society/view/2017/04/238867/>
- 박정엽, "靑청원 보니...대학생은 '조국 임명 찬성' 대열서 벗어나 있었다", *조선일보*, 2019.11.06.
- 김병록. (2019). 청와대 국민청원의 개선방안에 대한 연구. *Chosun Law Journal* Vol26 No.2 Aug 2019, 139-170.
- 김일수 외 11 인. (2015). 한국사회 정의 바로 세우기. 세창미디어. 36-38.
- 강형석, 양장훈. (2019). Word2vec 모델로 학습된 단어 벡터의 의미 관계 분석. *정보과학회논문지*, 46(10), 1088-1093.
- 심준식, 김형중. (2017). LDA 토픽 모델링을 활용한 판례 검색 및 분류 방법. *전자공학회논문지*, 54(9), 67-75.
- 이승원, 임한샘, 이현우 (2018). 청와대 국민청원 참여요인 분석. *OUGHTOPIA*, 33(3), 111-141
- 김철, (2011), 정의란 무엇인가 - 자유주의와 공동체주의의 가치, 자유와 평등, 형평, 사회이론 제 39 호 40
- 양천수. (2018). 민주적 법치국가의 내적 갈등. *법학연구*, 28(3), 271-305.
- 이새미, 홍순구 (2019). 토픽모델링 기법을 활용한 블록체인 동향 분석. *한국정보통신학회 여성 ICT 학술대회 논문집*, 44-47

5.부록

<표 1>

순 위	total	경제 민주화	교통건축 국토	기타	농산 어촌	문화예술 체육언론	미래	반려 동물	보건 복지	성장 동력	안전 환경	외교통일 국방	육아 교육	인권 성평등	일자리	저출산 고령화 대책	정치 개혁	행정
1	공의	구현	구현	구현	촛불	불의	여건	바탕	만인	상식	구현	독선	만인	구현	평등	평화	공의	구현
2	공정	구호	만인	법치	인권	삼	권선 징악	조장	입각	존중	공의	강자	출발점	공의	공정	자국민	구현	공의
3	정의 로운	만인	공평	갈망	주권	구현	조자	세분	공화국	이성	민중	건전	구현	법질 서	구현	공정	공정	공정
4	권선 징악	공서	반칙	공의	민주	사제단	구현	제시	공정	기득	기강	우연	수호	침경 화	정당 화	애국	준법 정신	반칙
5	부정 의	공의	공의	법질 서	평등	강자	법치 주의	비상 식	개법	대립	사법부	공정	주창	솔로 몬	함의	민주	여신 상	좌지우지
6	법언	승승	박수	법치 주의	탄생	민중	감수 성	세계 인	공평	관행	입법부	공정성	표방	여신 상	주권 자	부강	사제 단	배신
7	구현	반칙	추상	여신 상	응원	평등	공의	과학	다복	추구	법치	법치주 의	평등	실현	다수 결	잡종	법치 주의	근간
8	법치	불의	취임사	신의 칙	권력	타협	결백	대중	법치주의	이념	법치주의	물음	승복	법치 국	공평	후손	수호 자	법치주의
9	배신 행위	실천	특권	불의	역사	백계	공평	빌미	공산주의	강자	만인	참패	대환	순리	취임 사	서양	정의 로운	여신상
10	악은	정화	저것	련지	자존 심	천주교	업보	희석	초석	믿음	악인	내신	공의	응징	용단	구성원	법치	불의
11	공공 선	수호	성문법	사필	통일	건데요	당자	구시 대	사회주의	옹호	저울	정직	취임사	초석	염원	비판	강직	공명정대
12	반합	당도	심상	다수 결	외교	약자	단죄	퇴색	싸가지	지도층	다수결	낮	통념	보루	균등	반대	공정 성	민권

순 위	total	경제 민주화	교통건축 국토	기타	농산 어촌	문화예술 체육인론	미래	반려 동물	보건 복지	성장 동력	안전 환경	외교통일 국방	육아 교육	인권 성평등	일자리	저출산 고령화 대책	정치 개혁	행정
13	고영 주	어딴 나	표방	폐망	현명	법치주의	혜양	모방	자유민주 주의	치부	수호자	개헌안	법치주 의	국법	승복	암울	로망	규범
14	정당 화	법치 주의	입각	반칙	약자	공정	지도 층	명분	부조리	공감	애국	정반대	민주주 의	짝	묵과	민중	속물	준법정신
15	누설 자	표방	법질서	보루	간첩	취임사	공정	공정	구현	주제	행정부	국가주 의	승리	런지	냉정	지수	자기 반성	동북
16	원내 정당	법질 서	축출	망각	적폐	강타	신성	처우	적폐	질서	중시	자유사 상	진리	공정	룰	존속	승복	법치
17	박애	통념	구태	귀정	눈치	특권	시체	관습	민주주의	법치	수호	궤변	숙의	심판 대	진리	실현	척도	포폴리즘
18	민주	포용	법치	부끄 럼	존엄	애국	당도	민법	결실	겉까요	처단	정답	씨앗	법치	민주 주의	직결	자화 상	기치
19	제일 주의	공정	애국	법치 국	기원	통념	규범	다양	말살	반칙	원외	균등	부자유	강물	시발 점	재단	박아	강자
20	작폐	굴복	공헌	총론	건지	흔탁	구태	본질	답습	공직	응징	정략	반증	딜레 마	먼저	부합	거역	인민
21	대명 제	공정 성	불소급	국민 대표	안보	방종	무지 한	인격	퇴보	착각	심판	편취	염치	권선 징악	거래 법	전략	부끄 럼	자유민주 주의
22	미당	실현	근간	불명 예	육	구호	방종	육류	불변	성숙	풍토	당리당 략	기회균 등	진리	반칙	오늘날	경종	현주소
23	대한 민국	걸음	국민주	민주 주의	장난	공화국	재물	채식 주의	민주	자유민주 주의	방종	쟁취	파괴	경외	법치 주의	활	부관 사	실감
24	실현	성숙	직시	수호	민국	타락	빠	상정	법치	표현	후퇴	아량	예우	소시 민	물정	화합	철칙	풍조
25	자화 상	청렴	민주주의	근간	존중	자유민주 주의	가연	하위	통념	풍토	공직	자유민	덕목	영원	위	진심	우선 시	결여
26	매요	민중	토지개혁	악습	국정	탄생	만인	언급	기득권	진실	부조리	결집	우선시	저울	관료 주의	소아	쫓놈	자본주의
27	진면 목	민주	법치주의	근면	노고	원칙	냉정	현상	거래법	불신	면죄	박애	슬로건	첫발	괴	헌신	땅바 닥	만인

순 위	total	경제 민주화	교통건축 국토	기타	농산 어촌	문화예술 체육인론	미래	반려 동물	보건 복지	성장 동력	안전 환경	외교통일 국방	육아 교육	인권 성평등	일자리	저출산 고령화 대책	정치 개혁	행정
28	통일 법	일소	조세	금권	상식	객관성	평등	문구	우위	증시	폐란	매물	불합리	명심	수결	노년	갈구	경험칙
29	빈말	먼저	통치	기치	의식	숙박	관행	미흡	인도주의	견해	민주	용단	초석	입법 부	기치	찬성	갈망	모토
30	참모 습	주창	텐가	소임	정치	실현	버리 지	측면	발작	수호	개떡	가라	자주	자처	압승	지름길	만민	표방
31	저폐	함의	구호	해악	세상	반역자	진리	도태	잣대	대항	동아줄	충고	소망	안도 감	표방	애국자	자기 편	걸음
32	어딴 나	첩경	정당	능멸	좌파	일벌	먼저	수렴	이념	조심	공산주의	심상	주권	만인	시능	정치인	공명 정대	행정가
33	균점	철학	고마움	밀바 탕	조선	슬로건	옛	주관	도약	철학	기틀	실장	씩	시민 운동	입각	분노	독초	법전
34	청빈	초심	사고방식	환상	감동	구태	슬로 건	억지	기회균등	좌지우지	원칙	정파	과장	갈망	쟁취	자유민주 주의	악인	법체계
35	법치 국	타협	지도자	도외	사랑	야	잘잘 못	살상	역차	오해	자유민주 주의	구설수	효도	기틀	존중	생존	위선	개법
36	정대	다수 결	현정	공정	용기	야만	관례	진화	홀	서로	구태	비설	장천	천명	오점	서도	몰각	추상
37	장조	기강	평등	출발 점	계시 관	민주	투영	취지	민족	정서	공정	대타협	이바지	방점	이재 경	정당	공공 선	니들
38	위언	공직	겉	좌지 우지	격려	민주주의	몽둥 이	동법	성숙	청렴	집행자	만이	자유주 의	경종	귀결	혈통	보루	명심
39	인내 천	만민	특권층	초심	직결	기강	선처	긍정	이방인	도덕	정청	소명	이념	개법	정당	약화	화입	칼날
40	혜절	밀거 름	포용	만국	안심	종말	민초	단계	순리	무시	법조인	무엇인 지	불의	냉정	남북 통일	느낌	연지	정형식
41	첩경	항쟁	초딩	만인	청산	단죄	사법 부	범주	정치가	구호	범재	잣대	밀거름	국민 주	정치 권력	퇴출	정판	태생
42	천칭	법치	회의감	철퇴	주목	박근혜	부정 과	악의	평등	사익	고위급	심사숙 고	사회악	한통 속	각인	공론	현주 소	배반

순 위	total	경제 민주화	교통건축 국토	기타	농산 어촌	문화예술 체육인론	미래	반려 동물	보건 복지	성장 동력	안전 환경	외교통일 국방	육아 교육	인권 성평등	일자리	저출산 고령화 대책	정치 개혁	행정
43	부판 사	예우	자유민주 주의	논거	건설	죄	적패	만인	인륜	노예	눈꼽	묵과	역설	염원	정대	진정	몸부 림	맹세
44	진일 보	일말	만천하	결단 력	공감	주권자	폭력 배	악법	대범	기득권	성취	열심	이데올 로기	견지 요	온정	추구	아량	입각
45	존영	열망	역진	풀뿌 리	남북	자치	버리	준수	열망	의식	세간	이치	지향	당여	포퓰 리즘	흐름	개잡 놈	중시
46	정서 법	상통	매국노	열망	봉사 자	жат대	강자	이외	선조	집단	жат대	만인	찬사	깨달 음	저주	공동체	가책	통념
47	법치 행정	아집	수호	저들	수석	변즈	득세	대통 령령	타협	민주주의	통용	몰표	합치	마침 표	탄원	도모	좌고	당적
48	분노 함	가치 관	청렴	출세	대우	귀정	음해	상응	발돋움	민주국	검혀	법치	불공평	엄단	명심	인류	열망	진데
49	개잡 놈	적패	박주	인과 응보	노무 현	도리	소시 민	제고	관감사	뿌리	미풍양속	국가관	통탄	군더 더기	의구 심	먹구름	첫발	염원
50	주권 주의	후세 대	국법	정립	신뢰	죄인	족속	혐오 감	대한	안일	단죄	검손	실익	불변	항거	불	선량	공헌

<그림 1>

