

연구논문

20세기 한국에서 사용된 ‘민중’의 의미

- 주요 신문 기사를 중심으로 -*

허수**

1. 머리말
 2. ‘집단 주체’ 속 ‘민중’의 위상 변동
 - 1) ‘집단 주체’의 문맥과 구간별 동향
 - 2) ‘집단 주체’의 의미 구성
 - 3) ‘민중’의 위상: 중심에서 주변으로
 3. ‘민중’의 시기별 특징과 장기지속적 의미
 - 1) ‘민중’의 의미 구성과 변화 마디
 - 2) 시기별 중심 의미의 변화: 정치성·운동성의 강화
 - 3) 장기지속적 의미: 미래에 대한 기대
 4. 맺음말
-

국문초록

이 글에서는 텍스트마이닝의 토픽모델링을 이용해서 20세기 한국에서 사용된

* 이 연구는 2020년 서울대학교 미래기초학문분야 기반조성사업으로 지원되는 연구비에 의하여 수행되었음.

** 서울대학교 대학원 국사학과 부교수

대표논저:『이돈화 연구』, 역사비평사, 2011;『네트워크 분석을 통해 본 1980년대 민중 –『동아일보』의 용례를 중심으로』,『개념과 소통』18, 한림대학교 한림과학원, 2016;『『개벽』논조의 사회주의화에 관한 새로운 접근–토픽 연결망 분석을 중심으로』,『인문논총』78-1, 서울대학교 인문학연구원, 2021;『근대 전환기 ‘개벽’의 불온성과 개념화 – 동학·천도교를 중심으로』, 서울대학교 인문학연구원,『인문논총』78-4, 2021;『1920년대 신채호의 텍스트와 ‘민중’』,『역사비평』140, 역사비평사, 2022.

‘민중’의 의미를 고찰했다. ‘의미’ 파악의 범위를 현실이 아닌 ‘텍스트’ 차원에 국한했지만, 텍스트의 범위는 인접 단어의 문맥까지 포함했다. 연구는 두 단계로 진행했다.

1단계에서는 ‘집단 주체’의 전체 동향 속에서 ‘민중’의 위상을 분석했다. 토픽 구성을 근거로 5개 키워드 각각의 시기별 연관성을 연결망 지도로 시각화한 결과, ‘민중’의 위상은 20세기 전반에 중심 위치에 있었으나 후반에는 주변화되었음을 드러났다.

2단계에서는 ‘민중’의 문서만으로 토픽을 추출해서 그 의미를 분석했다. 그 결과 첫째, ‘민중’의 의미에서는 ‘집단 주체’의 시기구분과 달리 1984년이 중요한 획기가 되었다. 둘째, 20세기 전기에는 ‘조선·사회’ 토픽이, 중기에는 ‘제도·정치’ 토픽이, 후기에는 ‘정치·투쟁’ 토픽이 각각 드세했다.셋째, 이와 더불어 ‘예술·희망’ 등 미래에 대한 기대가 두 시기를 관통해서 장기지속하고 있었다.

주제어 : 민중, 의미, 토픽모델링, 집단 주체, 연결망 지도, 장기지속

1. 머리말

이 글에서는 20세기 한국에서 사용된 ‘민중’의 의미를 거시적으로 조망한다. 주지하듯이 근대 사회로 이행하면서 다수의 보통 사람들이 가진 역할과 위상은 더욱 중요해졌다. 그에 따라 이들을 호명하는 용어도 언어 질서의 근대적 변동과 맞물리면서 영향력이 커졌다. 이 용어들을 ‘집단 주체’로 포괄할 수 있는데, ‘민중’은 일제 강점기 중에서도 3·1운동 직후부터 그 사용이 급증하여 해방 무렵까지 활성화되었고, ‘1987년 6월 시민항쟁’으로 대표되는 ‘민주화 운동’ 시기에 해당하는 1980년대 후반~1990년대 초에 재활성화되었다는 점에서 한국근현대사의 역동성을 잘 담고 있다.

필자는 ‘민중’의 ‘의미’에 특히 주목한다. 한국 사회에서 ‘민중’이 어떤 뜻으로 사용되었는지를 규명할 것이다. 그런데 ‘민중의 의미’는 그 범위가 막연해서 장기적 양상을 살펴보기는 더욱 어렵다. 이 글에서는 이러한 추상적

대상을 탐구하기 위해, 디지털 인문학 방면의 텍스트 마이닝 분석 방법, 그 중에서도 특히 토픽모델링을 활용했다. 세부적 내용은 본문으로 미루고 여기서는 이에 관한 기본 발상과 경로에 한정해서 언급한다.

첫째, 이 글에서 ‘민중’의 의미를 파악하는 단서는 주요 신문기사 코퍼스로부터 추출한 토픽들이다. 토픽모델링은 코퍼스에 잠재되어 있는 여러 가지 주제들을 단어들의 집합 형태로 파악하는 데 유용하다.¹⁾ 이런 특징에 차안하여, 필자는 ‘민중’이 사용된 문맥(=‘문서’)을 다량으로 수집하고 코퍼스화 한 뒤 이로부터 토픽을 추출하면, 그 토픽들은 ‘민중’이 가진 여러 의미들에 상응한다고 보았다.

둘째, 이런 접근 방식은 ‘의미’ 파악의 근거를 ‘현실’이 아닌 ‘텍스트’ 차원에 한정한다. 이 점에서 그것은 기본적으로 ‘구조 의미론’의 접근과 동일하다.²⁾ 다만 이 글에서는 그러한 의미 파악의 토대를 ‘토픽’에 두고 있는 점, 그리고 의미의 지속·변화를 살펴보기 위해 역사적 관점을 가미하는 점 등이 다르다.³⁾ 이 글에서 주요 일간지의 용례를 활용한 것도, 비록 그것이 당시의 현실을 정확하게 반영한다고 하기는 어렵지만, 다른 자료에 비해 상대

1) 토픽모델링의 방법적 특징에 관하여는 다음을 참조. 허수·김혜진·정유경, 「대한제국기 ‘집단적 주체’의 의미망 –『황성신문』과 『대한매일신보』의 사설 기사를 중심으로」, 『대동문화연구』 119, 성균관대학교 대동문화연구원, 2022, 255~257쪽.

2) 여기서 말하는 ‘구조 의미론’은 “구조주의 언어학의 이론과 방법을 적용하여 의미를 연구하는 부문”인 ‘구조주의 의미 이론’을 가리킨다. 의미 이론에는 이 밖에 ‘역사주의 의미 이론’, ‘생성주의 의미 이론’ 등이 있다(최호철, 「구조 의미론의 수용 양상과 국어 어휘 의미론의 과제」, 『한국어 의미학』 2, 한국어의미학회, 1998, 13쪽). 구조 의미론은 의미를 ‘언어 내적인 상호 관계’ 중심으로 파악한다는 점에서 다른 입장과 차이가 있다.

3) 구조 의미론의 의의와 한계를 역사 의미론으로서의 개념사 연구와 연관 지어 살펴본 성과로는 다음을 참조. 박여성, 「개념사 연구: 역사서술과 언어학의 상호작용 – 구조의미론, 원형의미론 및 텍스트언어학을 중심으로–」, 『개념과 소통』 9, 한림대학교 한림과학원, 2012.

적으로 당시의 사회적 용례를 잘 드러낸다고 보았기 때문이다.

셋째, 이 글에서는 ‘민중’의 문맥을 토픽모델링하여 추출한 개별 토픽들을 ‘민중’의 하위 의미들로, 그리고 그러한 토픽들의 분포값으로 이루어진 전체 ‘토픽 구성’을 ‘민중’의 의미에 상응하는 것으로 간주한다.

넷째, ‘민중’의 의미는 ‘민중’의 토픽에 대한 관찰뿐 아니라 ‘민중’이 ‘집단 주체’ 전체 속에서 가진 위상을 함께 살펴봄으로써 제대로 파악될 수 있다고 본다. 더욱이 20세기 한국은 정치·사회적 변화 폭이 그 어느 시대, 어느 곳보다도 컸으므로, ‘민중’의 의미와 그 위상을 파악하려면 단순한 ‘민중’ 기표 차원을 넘어 인접 용어까지 포괄한 관찰이 필요하다고 생각된다.⁴⁾

한편, ‘민중’의 의미에 관하여서는 축적된 성과가 있다.⁵⁾ 그렇지만 필자

4) ‘민중’의 의미 파악에서 ‘토픽 구성’을 중시하는 것은 구조 의미론에서 단어의 ‘결합 관계’에 주목하는 것과 유사하다. ‘토픽 구성’을 기준으로 ‘민중’을 주요 ‘집단 주체’ 용어와 함께 살펴보는 것은 구조 의미론의 ‘계열관계’ 파악과 유사하다. ‘결합관계’ 와 ‘계열관계’에 관해서는 다음을 참조. “소쉬르는 …(중략:인용자)… 문장이란 단 어들이 일정한 법칙에 따라 나열된 것을 말하는데, 서로 연관되어 결합할 수 있는 관계를 결합관계(syntagme, 생타금)라고 하고, 어떤 단어가 다른 단어로 대체되어도 결합관계가 성립되는 관계를 계열관계(paradigme, 파라디금)라고 했다.”([철학사40] 소쉬르(Saussure), [http://www.epicurus.kr/Humanitas/386926, 2023. 1. 29. 검색\)](http://www.epicurus.kr/Humanitas/386926, 2023. 1. 29. 검색))

5) 이세영, 「‘민중’ 개념의 계보학」, 김경일 등,『우리 안의 보편성: 학문 주체화의 새로운 모색』, 한울, 2006; 장상철, 「1970년대 ‘민중’ 개념의 재등장 – 사회과학계와 민중문학, 민중신학에서의 논의」,『경제와사회』74, 비판사회학회, 2007; 강정구, 「진보적 민족문학론에서 민중 개념의 형성 과정 연구」,『비교문화연구』11-2, 서울대학교 비교문화연구소, 2007; 김성보, 「남북국가 수립기 인민과 국민 개념의 분화」,『한국사연구』144, 한국사연구회, 2009; 김윤희, 「근대 국가구성원으로서의 인민 개념 형성(1876~1894)」,『역사문제연구』21, 역사문제연구소, 2009; 황병주, 「1960년대 비판적 지식인 사회의 민중인식」,『기억과 전망』21, 민주화운동 기념사업회, 2009; 이신철, 「인민’의 창조와 사라진 ‘민중’ – 방법으로서 북조선 민중사 모색」,『역사문제연구』23, 역사문제연구소, 2010; 박기순, 「랑시에르와 민중 개념: 민중에 대한 낭만주의적 해석과 그 대안의 모색」,『진보평론』59, 2014; 이남희, 유리·이경희 옮김,『민중 만들기: 한국의 민주화운동과 재현의 정치학』, 후마니타스, 2015.

가 아는 한 ‘민중’의 의미를 1세기에 걸쳐 통시적으로 살펴본 연구는 없다. 개념사 차원의 장기적 연구나 계량적 차원의 접근에서도 마찬가지이다. ‘민중’을 ‘집단 주체’에 해당하는 다른 용어들과 함께 다룬 연구는 있다.⁶⁾ 그러나 다루는 대상이 20세기의 일부 시간대에 국한되었다. ‘민중’을 다루진 않았지만 ‘집단 주체’에 관한 주요 용어들의 의미를 고찰한 연구도 다수 있다.⁷⁾ 그렇지만 이 경우에도 일부 시기만 다루거나, 통시적 연구도 각 용어들을 고립적으로 분석했다.

이 글에서는 ‘민중’에 관한 기존 연구들이 공통적으로 중요하게 여긴 두 시기, 즉 일제 식민지기인 1920~30년대와 ‘민주화 운동기’에 해당하는 1980년대 후반~1990년대 초의 두 시기에 주목한다. 다음의 연구 질문, 즉 “20세기 전반과 후반, 한국 사회에서 사용된 ‘민중’은 그 기능과 의미 측면에서 과연 비슷했는가”라는 질문이 전체 분석 과정을 이끌 것이다.

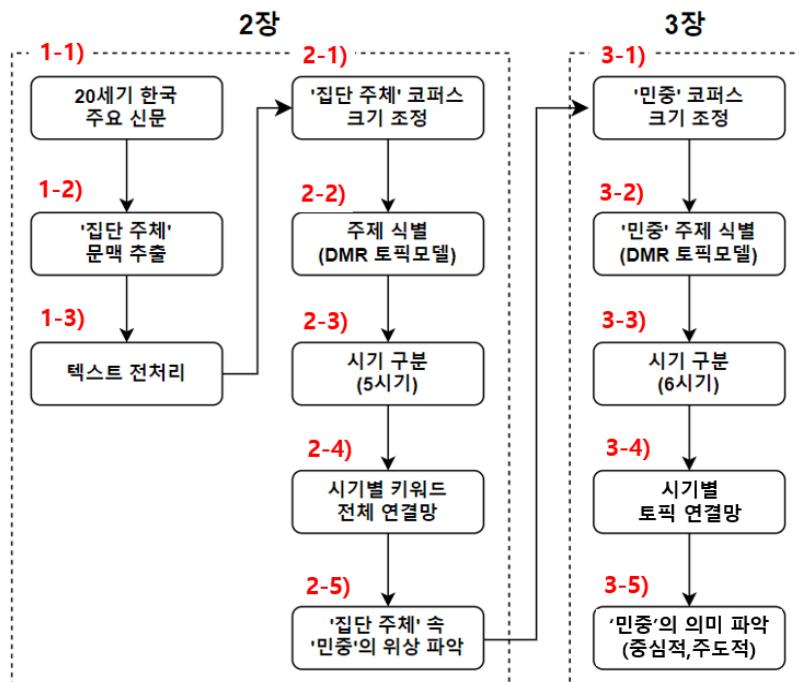
-
- 6) 허수, 「식민지기 ‘집합적 주체’에 관한 개념사적 접근 –『동아일보』 기사 제목 분석을 중심으로」, 『역사문제연구』 23, 역사문제연구소, 2010.
 - 7) 김동택, 『國民須知』를 통해 본 근대 ‘국민’, 『대동문화연구』 44, 성균관대학교 대동문화연구원, 2003; 권보드래, 「동포(同胞)’의 역사적 경험과 정치성 –『독립신문』의 기사분석을 중심으로」, 『근대계몽기 지식 개념의 수용과 그 변용』, 소명출판, 2004; 권보드래, 「동포와 역사적 감각: 1900~1904년 ‘동포’ 개념의 추이」, 『근대 계몽기 지식의 발견과 사유 지평의 확대』, 소명출판, 2006; 권보드래, 「근대 초기 ‘민족’ 개념의 변화: 1905~1910년 『대한매일신보』를 중심으로」, 『근대계몽기 지식의 굴절과 현실적 심화』, 소명출판, 2007; 김동택, 「대한매일신보에 나타난 ‘민족’ 개념에 관한 연구」, 『대동문화연구』 61, 성균관대학교 대동문화연구원, 2008; 박명규, 『국민·인민·시민』, 소화, 2009; 박찬승, 『민족·민족주의』, 소화, 2010; 허수, 「1920~30년대 식민지 지식인의 ‘대중’ 인식」, 『역사와현실』 77, 한국역사연구회, 2010; 허수·김혜진·정유경, 「대한제국기 ‘집단적 주체’의 의미망 –『황성신문』과 『대한매일신보』의 사설 기사를 중심으로」, 『대동문화연구』 119, 성균관대학교 대동문화연구원, 2022.

2. ‘집단 주체’ 속 ‘민중’의 위상 변동

1) ‘집단 주체’의 문맥과 구간별 동향

(1) 연구 진행 흐름

이 글에서 진행하는 연구 흐름과 그에 대응한 각 장의 전개는 <그림 1>과 같다.



<그림 1> 연구 진행 흐름과 장별 내용

<그림 1>에 표시한 단계별 번호를 사용해서 진행 과정을 약술하겠다. 2

장에서는 기초 자료를 수집해서 전처리하고, ‘집단 주체’ 차원의 토픽 추출과 시기구분 및 분석에 주안점을 두었다. 먼저 수집·전처리 과정을 3단계로 진행했다. 1-1) 『황성신문』, 『대한매일신보』, 『동아일보』, 『한겨레신문』의 4종 일간지에서 ‘민중’, ‘국민’, ‘인민’, ‘민족’, ‘대중’의 5개 키워드가 사용된 기사를 선정했다.⁸⁾ 수집한 전산자료의 연대는 1898~1999년의 102년이지만, 이 중에서 1910~1919년과 1941~1944년의 13년은 전산자료가 없어서 실제적인 포괄 연도는 89년이다. 1-2) 기사들 중에서 5개 키워드가 사용된 위치의 앞뒤 20음절(공백 제외)씩을 키워드와 함께 떼 내어 각각을 개별 ‘집단 주체’들의 ‘문맥’(=‘문서’)으로 삼았다.⁹⁾ 1-3) 이 문맥 텍스트를 전처리하여 ‘집단 주체 코퍼스’를 생성한 뒤, 관찰의 편의를 위해 20세기 전체를 3년을 하나의 단위로 하는 29개의 구간(grid)으로 줄이고, 다시 이것을 5개의 대구간(big-grid)으로 구분했다.

이어서 토픽추출과 분석은 다섯 단계로 진행했다. 2-1) 20세기 말로 갈수록 관련 문서 수는 급증했다.¹⁰⁾ 그에 따른 시기별 격차를 그대로 반영해서 토픽모델링 입력 값으로 넣으면, 그 결과로 나온 토픽에는 20세기 말의 단어들이 너무 많이 반영된다. 따라서 토픽모델 결과값의 구간별 편향을 축소하기 위해 대구간의 코퍼스 크기를 가급적 서로 비슷하게 조정했다. 2-2) 조정된 코퍼스로 토픽모델링을 진행하여 10개의 토픽을 추출했다. 2-3) 토픽모델링의 결과물 중 구간별 토픽구성 자료는 시기구분을 하는 데 유용하다. 구간별 토픽구성이 크게 달라지는 점을 경계로 전체 기간을 5개

8) 이 글에서는 5개의 단어를 묶어서 가리킬 때 ‘집단 주체’ 혹은 ‘키워드’로 부를 것이다.

9) 이 글에서는 전체 코퍼스 중에서 이러한 개별 문맥을 가리킬 경우 ‘문서’로 부를 것이다.

10) 이것은 첫째, 신문의 증면에 따른 기사의 증대, 1988년부터 『한겨레신문』의 창간 등에 따른 결과이다. 둘째, 20세기 후반으로 갈수록 ‘집단 주체’의 사용이 증가한 데에도 원인이 있다고 생각된다.

시기로 구분했다. 2-4) 토픽모델링의 결과물 중 문서별 토픽구성 자료도 유용하다. 이 문서들의 토픽구성을 시기별·키워드별로 통합하면 25개 문서 단위별 토픽구성 자료를 얻을 수 있다. 나아가 이 역시 토픽구성을 근거로 25개 단위 상호 간의 유사도를 산출할 수 있다. 유사도 산출은 코사인 유사도를 사용했다. 25개 단위 각각의 관계를 이 값으로 연결해서 그래프로 그린 것이 시기별 키워드 전체 연결망인데, 이는 곧 20세기 전체를 무대로 한 시기별 키워드의 전체 분포와 상호 관계에 해당한다. 2-5) 이 연결망 지도로부터 ‘집단 주체’ 속 ‘민중’의 위상을 검토했다.

3장에서는 2장과 비슷한 과정을 ‘민중’ 문서만으로 진행했다. 3-1) ‘집단 주체’ 중에서 5개의 대구간 구분은 그대로 이어받되, 활용 자료는 ‘민중’ 문서만 선별했다. 다만, 여기서도 대구간별 문서 규모가 균질하게 되게끔 크기를 조정했다. 3-2) 조정된 ‘민중’ 코퍼스로 토픽모델링을 실행해서 10개의 토픽을 얻었다. 3-3) ‘민중’의 구간별 토픽구성을 살펴서, ‘집단 주체’의 경우와 달리 6개로 시기구분했다. 3-4) 문서의 토픽 구성 정보를 시기별로 통합해서 시기별 토픽 연결망을 6개 산출한 뒤, 뚜렷하게 시대별 특징을 나타내는 토픽에 주목했다. 3-5) ‘민중’ 코퍼스 전체를 활용해서 연결망 지도를 그린 뒤, 둘 이상의 시대에 장기지속적 영향을 끼치는 토픽에 주목하고 그것이 가진 함의를 살펴보았다.

(2) ‘집단 주체’ 문서의 코퍼스화와 문맥 동향

전술한 4종의 주요 일간지 전자 자료는 ‘대한민국 신문 아카이브’(국립중앙 도서관) 및 ‘네이버뉴스라이브러리’에서 수집했다. 이 글에서는 ‘집단 주체’의 5개 키워드와 그 앞뒤 20음절을 ‘문맥’ 혹은 ‘문서’라고 부른다.¹¹⁾ 이 문서에

11) 5개 ‘집단 주체’의 문맥을 어절·문장·문단·기사 등 경계가 분명한 범위로 설정하

서 함께 사용된 단어들을 해당 키워드의 ‘연관어’로 간주한다.

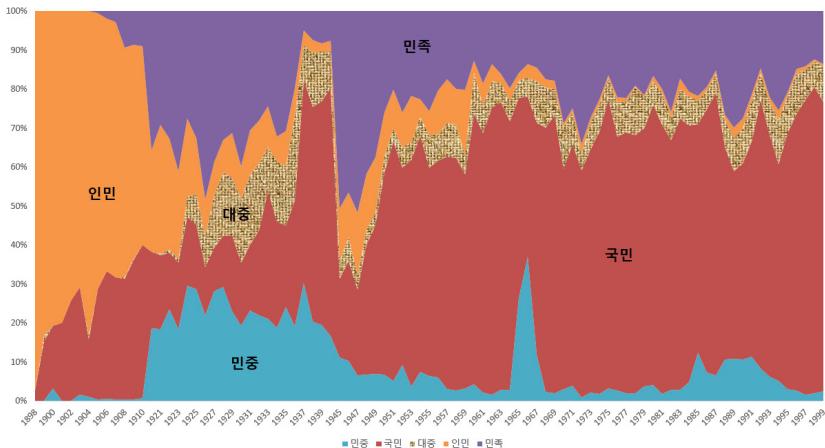
『황성신문』 기사만으로 대한제국기의 대부분을 포괄함에도 불구하고 1905년부터 『대한매일신보』 국한문 혼용 기사도 포함한 이유는 ‘집단 주체’의 동향 파악에서 1900년대 후반 새로운 일간지가 발간된 역사성을 반영하기 위함이다. 1980년대 후반 이후 『동아일보』 기사 이외에 『한겨레신문』 기사를 포함한 것도 동일한 이유이다. 이렇게 수집한 4종 일간지의 ‘집단 주체’ 문서에서 북한 관련 기사는 포함하고, 외국 관련 기사는 가급적 제외했다. 그 결과 80만 건 안팎의 문서를 획득했다.

수집 전산 자료의 분량이 전처리하기에 너무 방대하여 자료를 1/10로 축소해서 진행했다.¹²⁾ 그 결과 79,276개 문서, 612,491개 형태소로 이루어진 ‘집단 주체 코퍼스’를 생성했다. 약 8만개에 달하는 이 문서는 5개의 ‘집단 주체’가 중심이 된 문맥이다. 이 문서 수의 연도별 동향을 ‘100% 누적기준 영역형’ 그래프로 나타내면 <그림 2>와 같다.¹³⁾

지 않고 각 키워드의 앞뒤 20음절로 설정한 데는 이유가 있다. 첫째, 해당 신문기사 중 상당 수는 띠어쓰기가 불규칙하고, 문장 종결 부호가 없는 경우가 많기 때문에 어절이나 문장 단위의 문맥 추출을 일관되게 하기 어렵다. 문단 차원의 추출도 이와 대동소이한 문제를 가진다. 둘째, 기사 단위의 추출은 경계가 분명해서 가능하나, 이 경우 ‘집단 주체’의 의미를 잘 반영하는 코퍼스가 되기 어렵다.

12) 전처리에 관한 자세한 기준은 이 글 마지막의 “[부록 1] ‘집단 주체’ 코퍼스 작성 기준”을 참조.

13) 문서 속 키워드로는 ‘국민’, ‘인민’ 등의 단독형 단어도 있지만, ‘국민학교’, ‘인민당’ 등의 복합어도 포함되어 있다. 그러므로 키워드별 문서 개수는 키워드별 빈도와 일치하지 않는다. 이 글에서는 키워드가 포함된 복합어도 넓은 의미의 ‘키워드’ 동향에 포함시키는 취지에서, 키워드별 문서의 연도별 빈도를 ‘집단 주체’의 전체 동향을 파악하는 근거로 삼았다.



〈그림 2〉 20세기 ‘집단 주체’의 비중 변화

그래프에서 1911~1919년 자료는 누락되었다. 1920년 결과는 포함되었으나 연도 눈금이 2년마다 하나씩 표현되므로 눈금 상으로는 보이지 않는다. 1941~1944년 자료도 누락되었다.

여기서 ‘국민’과 ‘인민’의 비중은 1910년까지 양자를 합쳐 ‘집단 주체’ 전체의 대부분을 차지했으나, 일제 식민지 시기인 1920년 무렵에는 비중이 절반 수준으로, 그리고 그 이후에는 그 이하로 급감했다. 이후 ‘국민’은 일제 말기와 해방 직후의 격동기를 거친 뒤 1960년 무렵부터 일부 구간을 제외하면 대체로 60~70% 선을 유지했다. 이와 대조적으로 ‘인민’은 1940~50년대까지도 명맥을 유지했으나 1960년 무렵부터는 거의 자취를 감추었다. 이런 대조적 양상은 분단 현실을 반영한 것으로 생각된다. 반면, ‘대중’의 비중은 전체적인 변동 폭이 5개 ‘집단 주체’ 중에서 가장 작다. 1920~30년대에 약간 팽창했다가 해방 후 크게 축소한 뒤 1950년부터는 좁은 범위 내에서 그 폭을 유지했다.

‘민족’은 1920년대에 급증해서 최대 비중을 유지하다가 일제 말기와 해방 직후 ‘국민’과 반대 방향의 움직임을 보인 것이 특징이다. 1950년부터는 줄곧 일정 비중을 유지했다. ‘민중’은 1920~30년대, 1960년대 중반, 1980

년대 후반~1990년대 초의 세 기간에 뚜렷한 증가를 보였다. 이 중에서 '민중'이 '민족'과 동시에 활성화된 시기는 1920~30년대와 1980년대 후반~1990년대 초라서 주목할 만하다. 이 두 시기는 선행 연구에서도 '민중'이 활성화된 기간이라고 생각해서 일찍부터 주목해 왔다. 1960년대 중반에 '민중'의 비중이 짧은 기간에 급증한 것은 1965년 6월 '민중당'이 창당되었다가 대통령 선거를 앞두고 1967년 2월 '신민당'으로 흡수된 사정을 반영한다.¹⁴⁾

2) '집단 주체'의 의미 구성

토픽모델링은 대규모 문서에는 서로 구별되는 주제들이 잠재되어 있다고 가정하고, 이 주제들을 토픽의 형태로 드러낸다. 그리고 이 토픽들은 단어들의 집합 형태로 표현된다.¹⁵⁾ 토픽모델링을 잘 통제된 환경과 절차에 따라 실행하면, 방대한 규모의 문서들로부터 일정 개수의 토픽들을 효과적으로 추출해 낼 수 있다. 나아가 이 글 2장 1절의 "(1) 연구 진행 흐름"에서 언급 했듯이, '집단 주체' 코퍼스에서 추출한 각 토픽들은 '집단 주체' 일반의 의미를 구성하는 하위 의미들이라 할 수 있다.

14) 민중당은 1965년 6월, 야당인 민정당과 민주당이 통합하여 만든 당이다. 이 당은 전통 보수정당의 성향을 가졌는데, 박정희 정권의 '한일협정' 비준과 국군의 베트남 파병 동의안 처리를 둘러싸고 대정부 입장에서 내부 차이가 노정되어 윤보선 등의 강경파가 탈당하여 '신한당'을 창당했다. 그러나 민중당과 신한당은 제6대 대통령선거를 앞두고 다시 통합을 추진하여 1967년 2월 통합하고 정당명은 '신민당'으로 불임으로써, 민중당은 역사 속으로 사라졌다(이상구, 「민중당(民衆黨)」, 『한국 민족문화대백과사전』, <http://encykorea.aks.ac.kr/Contents/Item/E0020310>).

15) 허수·김혜진·정유경, 앞의 글, 255~257쪽.

〈표 1〉 대구간별 문서 수 조정(상) 및 조정 전후 문서의 크기 비교(하)

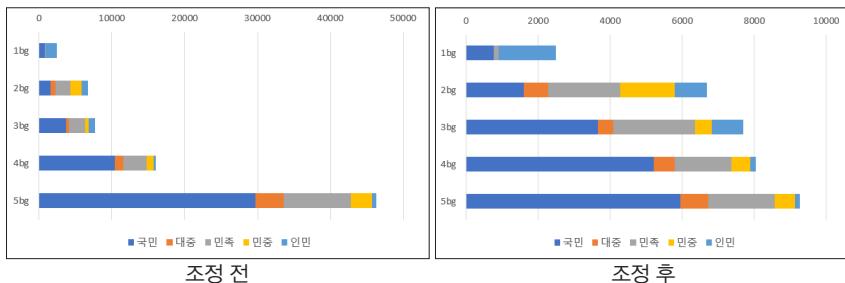
big_grid key	1bg	2bg	3bg	4bg	5bg	합계	
국민	769	1599	3664	10411	5208	29743	5949
대중	5	688	428	1159	578	3849	769
민족	121	1990	2253	3170	1583	9236	1848
민중	14	1523	477	1014	512	2874	572
인민	1580	889	868	319	156	633	129
합계	2489	6689	7690	16073	8037	46335	9267
						79276	34172

대구간(big-grid, bg)은 관찰의 편의를 목적으로 20세기 한국사의 일반적 시기구분을 참고하여 다음과 같이 구분했다. 4bg와 5bg, 합계에서 왼쪽 칸은 조정 전 문서 수, 오른쪽 칸은 조정 후 문서 수이다.

1bg는 대한제국기(1898~1910), 2bg는 일제시기(1920~40),

3bg는 1940·50년대(1945~1959), 4bg는 1960·70년대(1960~1980),

5bg는 1980·90년대(1981~1999년)



그런데 이런 장점을 가진 토픽모델링을 실행할 때에 주의할 점도 있다. 구간별 문서 수가 서로 많이 차이날 경우, 각 토픽 내에서 문서 수가 큰 구간의 단어들이 그렇지 않은 구간의 단어들에 비해 너무 많이 나타나는 경향이 있다.¹⁶⁾ 따라서 이를 방지하기 위해 〈표 1〉처럼 각 구간의 문서 수를 균질화할 필요가 있다. 5개 대구간별 문서 수는 다음과 같이 조정했다. ‘집단주체’ 문서 총 수는 1~3대구간은 그대로 두고, 4대구간은 1/2로, 5대구간

16) 20세기 ‘집단 주체’의 문서는 20세기 말로 갈수록 지면 수의 증가로 그 수가 증가하고, 여기에 『한겨레신문』의 창간과 민주화 운동기의 사용 증가가 더해져 문서 규모가 급증했다.

은 1/5로 각각 축소했다. 이러한 축소 작업은 각각 두 번째 문서와 다섯 번째 문서를 기계적으로 선택하는, 랜덤샘플링 방식에 따랐다. 그 결과 조정된 문서들을 통칭해서 ‘집단 주체 균형코퍼스’로 부른다. 이 균형코퍼스의 크기는 34,172개로 당초 79,276개의 43.1%에 해당한다.

토피모델링은 LDA모델을 변형한 DMR모델을 사용했다.¹⁷⁾ LDA모델과 달리 이 DMR모델은 데이터와 함께 입력하는 ‘추가적 정보’가 데이터의 주제와 단어에 영향을 끼친다는 발상에 기반을 둔다. 이런 DMR모델을 사용해서 ‘추가적 정보’에 ‘구간’(grid)이라는 시간 차원의 자료를 넣으면, ‘구간별 토픽 구성’을 얻을 수 있다.¹⁸⁾ 이 값으로 토픽의 역사적 변화를 효과적으로 포착할 수 있다. 토픽을 추출해서 토픽별 제목을 붙이는 작업은 다음 절차를 밟았다.

첫째, 토픽 개수를 몇 개로 할지를 혼잡도 및 토픽일관성 결과값을 참고로 해서 일단 11개로 정했다. 그러나 이 개수로 추출한 토픽 결과물을 검토하는 과정에서 둘 이상의 토픽 간에 비슷한 단어가 많고 의미가 중복되는 경향을 확인했다.¹⁹⁾ 이에 토픽 수를 10개로 축소 조정했다.

17) LDA는 ‘잠재 디리클레 할당(Latent Dirichlet Allocation)’이며, DMR은 ‘디리클레 다항 회귀분석(Dirichlet Multinomial Regression)’이다.

18) 구간별 토픽구성은 다음과 같다. 첫 열은 10개의 토픽명이며, 첫 행은 29개의 구간명이다. 값은 10개의 토픽 각각이 구간별로 차지하는 비중이며, 동일 구간 내 각 토픽의 비중을 합하면 1이 된다.

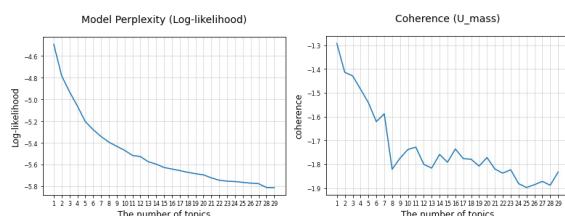
	01g	02g	...	28g	29g
T0	0.030149	0.031442		0.09456	0.093792
T1	0.027292	0.025127		0.150868	0.133887
...
T8	0.030047	0.035282		0.302582	0.246506
T9	0.678169	0.677015		0.009237	0.005505

19) 다음 왼쪽은 혼잡도 결과 그래프, 오른쪽은 토픽일관성 결과 그래프이다. 토픽일관성 지표는 ‘u-mass’를 사용했다.

둘째, 29개 구간 정보를 메타데이터로 입력해서 토픽모델링을 실행했다. 토픽 수는 10, 알파 값과 베타 값은 각각 0.1, 0.01로, 토픽별 단어 개수는 20으로 입력했다. 다섯 번의 토픽모델링을 실행해서, 그 중에서 최적이라고 생각되는 결과를 선택했다.²⁰⁾ 선택 과정에서는 토픽별 변별성 정도, 구간별 토픽구성에서 도출되는 시기구분의 명징성 여부 등을 고려했다. 최종 토픽에 제목을 붙인 결과는 〈표 2〉와 같다.

〈표 2〉 '집단 주체' 토픽 제목 및 토픽 단어

토픽	토픽 제목	토픽 단어 (상위 20개)
T00	교육·경제	국민, 국민학교, 교육, 경제, 정부_8, 대중, 부담, 생활, 민족, 일반, 실시, 민족, 보건, 향상, 세금, 예산, 금융, 국민소득, 문교부, 보험
T01	대중·문화	대중, 민족, 민중, 음악, 새정치국민회의, 국민, 미술, 문학, 운동, 대중가요, 대중음악, 노래, 대중화, 주제, 문화, 개최, 예술, 대회, 대표, 국민학교
T02	학교·교육	국민학교, 인민, 아동, 중학교, 대중교통, 버스, 어린이, 시내, 수단, 학교, 고등학교, 교장, 사립, 교사_9, 교실, 교통, 요금, 도로, 학생, 학부형
T03	민족·민중	민족, 민중, 국민, 조선_5, 인민, 대중, 생활, 국가, 일반, 경제, 자유, 사회_7, 정부_8, 운동, 정신_12, 세계_2, 정치, 문화, 노력, 조직
T04	자유·독립	민족, 국민, 민중, 조선_5, 인민, 자유, 국가, 대중, 정치, 문화, 통일, 한국, 민족주의, 독립, 운동, 역사, 세계_2, 사회_7, 해방, 정신_12
T05	국민·시대	국민학교, 국민, 국민주택, 민족, 교육, 국민대학교, 대중, 어린이, 학교, 국민훈장, 교사_9, 국민은행, 중학교, 국민총생산, 국민연금, 운동, 시내, 대학, 자금, 공연
T06	민족·생활	민족, 국민, 국민은행, 대중, 민중, 조선_5, 생활, 문화, 경제, 인민, 국가, 국민소득, 나라, 세계_2, 일반, 발전_1, 사회_7, 사람, 한국, 대회
T07	제도·정치	새정치국민회의, 국민, 민중당, 의원_6, 대통령, 후보, 선거, 국회, 통일국민당, 국민투표, 자유민주연합, 민주당, 대표, 회의_3, 야당, 정부_8,黨, 총재, 김대중, 한국국민당



20) 토픽모델링 재현을 위한 seed(= 's') 값을 각각 1, 10, 100, 1000, 10000으로 주었다. 이 다섯 후보 중에서 s=1을 최적 토픽모델 결과로 선정했다.

토픽	토픽 제목	토픽 단어 (상위 20개)
T08	국민·정부	국민, 정부_8, 민족, 정치, 대통령, 민중, 통일, 선거, 대중, 국민투표, 민주, 경제, 국가, 야당, 국회, 사건, 한국, 헌법, 새정치국민회의, 정책
T09	인민·보호	인민, 국민, 일반, 보호, 국가, 정부_8, 지방, 의무, 재산, 토지, 아국, 교육, 청원_2, 아한, 동포, 민중, 한국, 대한_6, 관리_1, 생명

토픽 제목은 필자가 각 토픽의 단어 구성과 다른 토픽과의 변별성 등을 종합적으로 고려해서 붙였다. 토픽별로 맨 앞에 오는 단어일수록 해당 토픽 내에서 차지하는 비중이 높으므로, 가급적 앞 순위에 있는 단어 중에서 다른 토픽과 변별되는 단어를 반영했다.

3) ‘민중’의 위상: 중심에서 주변으로

여기서는 토픽모델링으로 추출한 ‘집단 주체’의 토픽들로부터 시기구분을 하고, 토픽들의 시기별 동향을 살펴본 뒤, 최종적으로 시기별 ‘집단 주체’들 가운데 ‘민중’의 위상이 어떠한가를 고찰하고자 한다.

첫째, 토픽을 활용한 시기구분이다. 전술했듯이 DMR모델에서는 ‘구간별 토픽구성’을 쉽게 구할 수 있고, 29개 구간은 각기 상이한 토픽구성을 가지기 마련이므로, 이 자료를 다음과 같이 시기구분의 합리적 근거로 활용할 수 있다.

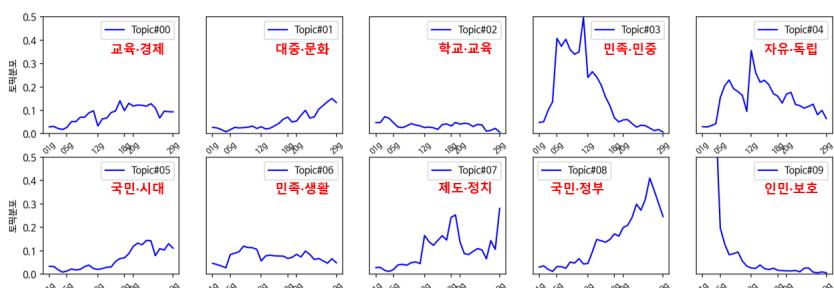
구간별 토픽 구성을 통해서 각 구간끼리의 상관성이 얼마나 되는지 알 수 있고, 상관성의 구간별 증감이 크고 작은 양상을 조망할 수 있다. 시기구분의 경계선은 상관계수가 높은 정도로 이어지다가 그 값이 급감하는 지점이 되기 마련이다. <그림 3>은 피어슨 상관계수를 값으로 하는 행렬인데, 두 구간의 관계가 양의 상관계수 0.7 이상을 보이는 부분에 음영(빨간색)으로 표시했다. 나아가 해당 구간과 가장 높은 상관계수를 가진 구간을 대상

으로 구집화를 했다. 이 구집이 하나의 별별적인 시기구분을 이룬다.

〈그림 3〉 ‘집단 주체’의 구간별 상관 계수와 시기구분

그 결과 전체 시기는 1898~1910년(1기, 1~4g), 1920~1940년(2기, 5~11g), 1945~1962년(3기, 12~17g), 1963~1968년(4기, 18~19g), 1969~1999년(5기, 20~29g)로 구분할 수 있다.

둘째, ‘집단 주체’의 하위 의미들에 해당하는 각 토픽들이 시기별로 어떤 동향을 보이는가 살펴본다. 토픽들의 시기별 변화를 그래프로 나타내면 〈그림 4〉와 같다.



〈그림 4〉 ‘집단 주체’ 토픽들의 시기별 변화 (▼축 갑: 부포간)

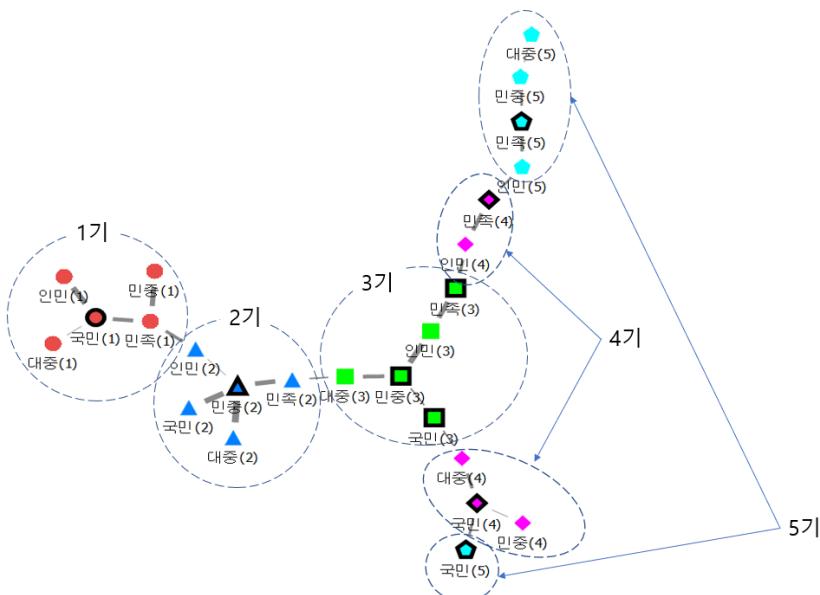
〈그림 4〉에 제시된 각 토픽의 동향은 증감 폭과 증감 시기를 교차시켜 다음과 같이 분류해 볼 수 있다. 먼저 증감 폭과 관련해서는, 토픽별로 분포 값의 표준편차를 내고 다시 10개 토픽의 표준편차를 평균해서 상·중·하의 세 등급으로 나누었다. 표준편차가 큰 ‘상’ 그룹은 구간별 진폭이 크다는 것을 뜻한다. 여기에 해당하는 토픽은 ‘민족·민중’⁽³⁾, ‘국민·정부’⁽⁸⁾, ‘인민·보호’⁽⁹⁾ 토픽이다. 이것에 증감 시기를 결부시키면, 토픽 3과 9는 각각 2기와 1기, 즉 둘 다 20세기 전반에 활성화된 토픽이다. 토픽 8은 이와 반대로 20세기 전반은 아주 낮다가 20세기 후반에 급증했다. 이 ‘상’ 그룹은 그 정점의 높이나 변화의 정도 면에서 ‘집단 주체’의 전체적인 의미를 주도하는 토픽이라 할 수 있다. 또한 토픽 제목에서도 드러나듯이 키워드도 서로 변별적이다. 토픽 3에서는 ‘민족’과 ‘민중’이, 토픽 8에서는 ‘국민’이, 토픽 9에서는 ‘인민’이 중심 단어이다.

한편, ‘자유·독립’⁽⁴⁾과 ‘제도·정치’⁽⁷⁾ 토픽은 표준편차가 전체 평균에 가까운 ‘중’ 그룹에 속하며, 그 외의 토픽들은 연간 변동이 상대적으로 완만한 ‘하’ 그룹에 해당한다.

셋째, ‘집단 주체’들 속에서 ‘민중’의 위치가 어떠했는지를 살펴보자. 우리는 34,172개의 ‘집단 주체’ 코퍼스에서 10개의 토픽을 추출하고, 이를 토대로 5개의 시기구분을 했다. 그런데 당초 이 문서들은 5개 키워드별로 나누어 추출한 것이므로 문서는 5개 그룹으로 구분할 수 있다. 시기 구분과 키워드 구분을 교차시켜 문서를 분류하면, 모두 25개의 구분 단위를 얻을 수 있다. 이 25개의 단위도 각각 10개의 토픽들과 각각의 분포값으로 이루어진 토픽 구성을 가지므로,²¹⁾ 이를 토대로 25개 단위 간의 유사도를 산출할

21) 이 25개 단위 각각의 토픽 구성에 들어있는 토픽별 분포값은 34,172개 문서의 각 토픽 분포값을 시기별·키워드별로 평균해서 도출한 것이다.

수 있다.²²⁾ 이 유사도 값을 넷마이너 프로그램에 입력해서 25개 단위 각각을 노드로 하고, 단위 간 유사도를 링크로 하는 연결망 그래프를 그린 것이 <그림 5>이다.²³⁾



<그림 5> ‘집단 주체’의 시기별 키워드 간 연결망 지도

<그림 5>의 연결망 지도는 시기별 키워드 간의 관계를 한 눈에 보여준다. 이 연결망 그래프는 곧 ‘집단 주체’ 간의 관계를 시각화한 것이라 할 수

22) 유사도는 코사인 유사도를 사용했다.

23) 그래프 종류는 노드 간 연결을 최대값 중심으로 표현하는 패스파인더(PFnet) 방식을 따랐다. 이는 25개 노드 전부를 상호 연결하게 해서 고립 노드가 생기는 것을 방지하고, 너무 복잡한 연결망을 피하면서 전체적인 연결 구조를 잘 드러내기 위함이다. 이에 관하여는 다음 연구를 참조할 것. 이재윤, 「지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구」, 『한국문헌정보학회지』 40-2, 한국문헌정보학회, 2006.

있으므로, 우리는 그 속에서 ‘민중’이 시기별로 어떤 위치에 있었는가를 살펴볼 수 있다.

상기 연결망 지도에서 ‘집단 주체’ 키워드 오른쪽 팔호 속에는 해당 시기를 숫자로 표시했다. 식별의 편의를 위해 동일 시기에 속한 키워드에는 동일한 도형과 색깔을 부여하고, 점선으로 된 동그라미를 사용해서 시기들을 서로 구분했다. 그리고 시기별 연결망의 중심 노드에는 진한 테두리를 둘렀다.²⁴⁾ 이 연결망 지도로부터 다음 몇 가지 사실을 확인할 수 있다.

첫째, 1~3기까지는 ‘집단 주체’의 키워드들이 시기별로 한 데 뭉쳐있는 데 반해, 4~5기는 두 계열로 갈라져 있다는 사실이다. 4기부터의 이런 양분된 계열에서 우상 방향의 계열은 ‘민족–인민’이 중심이 되고 우하 방향의 계열은 ‘국민’이 중심이 되어 있다. 3기 ‘민중’은 두 계열의 분기점에 해당한다. ‘민중’과 ‘대중’은 두 계열에 나누어 자리했는데, 4기에는 우하에, 5기에는 우상에 각각 나타났다.

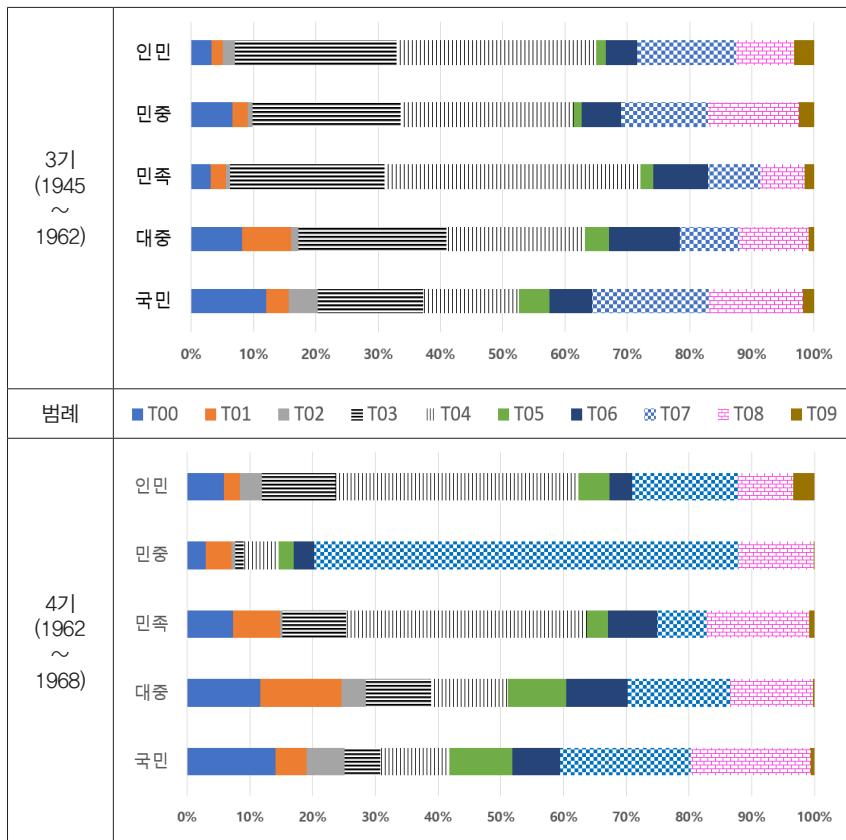
둘째, 20세기 후반부터 진행된 ‘집단 주체’의 이원화는, 당시 ‘국민’에서 활성화된 변화가 주요 원인으로 작용했다고 판단된다. 이렇게 판단한 이유는 다음 두 가지 때문이다. 첫 이유는 시기 간 연결고리를 살펴보았기 때문이다. <그림 5>에서 시기와 시기를 잇는 ‘집단 주체’로는 ‘민족–인민’이 다수이다. 1기와 2기 사이, 3기와 우상 4기, 우상 4기와 우상 5기 사이가 그려하다. 2기와 3기 사이는 ‘민족–대중’이며, 3기와 우하 4기 사이는 ‘국민–대중’, 우하 4기와 우하 5기 사이는 ‘국민–국민’이다. 즉 총 6개의 연결고리 가운데 ‘민족–인민’이 3회로 절반에 해당한다. 두 계열 분화 이전만 보더라도

24) 1~3기처럼 동일 시대 내에서 연결 중심이 확실한 경우에는 그 노드에, 4~5기처럼 시대별 노드가 두 계열로 분산된 경우에는 두 계열 각각에서 중심적인 키워드의 노드에 진한 테두리를 했다.

도 2기와 3기 간의 연결고리는 ‘민중–대중’으로서, 여기에도 ‘국민’은 개입하지 않는다. 말하자면 20세기 전체에 걸쳐 ‘집합 주체’의 시기별 연속성은 주로 ‘민족–인민’의 단어 쌍으로 이어지는 것이 대세였다. 이러한 대세 속에서 4기부터는 우상 계열로 3기까지의 이전 경향이 지속되는 한편, 우하 계열이 새롭게 발달한 형국을 보인다. 즉, 3기의 ‘국민’이 4기의 ‘대중’과 연결되는 한편, 4기 ‘대중’과 연결된 4기의 ‘국민’이 5기의 ‘국민’과 연결되는 계열이 새롭게 나타난 것이다.

두 번째 이유는 3,4기의 키워드별·시기별 토픽 분포 그래프에서 ‘국민’의 동향이 특별한 양상을 보이기 때문이다. 〈표 3〉에서 알 수 있듯이 ‘국민’의 경우 2기에서 3,4기로 가면서 ‘민족·민중’⁽³⁾과 ‘자유·독립’⁽⁴⁾의 비중이 다른 키워드에 비해 빠르게 감소하는 반면, ‘제도·정치’⁽⁷⁾와 ‘국민·정부’⁽⁸⁾는 빠르게 증가했다. 〈표 2〉의 토픽 단어를 볼 때, 증가한 이 두 개의 토픽은 국민 국가 혹은 민족 국가의 형성에서 가장 기초에 해당하는 입헌 민주주의의 제도화와 관계 깊다. 그렇다면 이미 해방 이후부터 그리고 1960년대 초부터 이러한 양상은 ‘집단 주체’ 중에서는 특히 ‘국민’의 문맥에서 많이 거론되었다고 볼 수 있다.

〈표 3〉 키워드별·시기별 토픽 분포 그래프



셋째, ‘민중’의 중심적 위치는 1950년대까지 지속되다가 3공화국이 시작된 1960년대 초부터 두 계열로 분화되고, 그에 따라 이전의 ‘민중’이 가진 중심적 위상은 사라지면서 주변화되었다. 이와 함께 ‘민중’은 한편에서는 정당제도화 경향으로 기울었다가, 다른 한편으로는 ‘자유·독립’에 관한 토픽이 늘고 ‘민족’과 결합하는 등 양쪽으로 진동하는 경향을 보였다.²⁵⁾ 이에 관

25) 독자들 중 일부는 1960년대 ‘민중당’ 창당과 지속이라는 짧은 기간 동안의 ‘사건’을

한 조금 자세한 검토는 이 글 3장 2절에서 진행하고자 한다.

3. ‘민중’의 시기별 특징과 장기지속적 의미

1) ‘민중’의 의미 구성과 변화 마디

이 장에서는 2장의 ‘집단 주체’ 차원에서 실행한 토픽모델링과 그에 기초를 둔 분석 과정을 ‘민중’에 집중해서 진행하고자 한다. 그 이유는 2장의 분석에서 ‘민중’의 의미 구성과 변화 양상이 충분히 드러나지 않았기 때문이다. 특히 2장에서 추출한 토픽에서는 1980년대 민주화 운동기 ‘민중’의 활성화가 별로 반영되지 않았다. 가장 큰 이유는 2장의 분석에서 활용한 ‘집단 주체’의 균형 코퍼스에서는 대구간별 코퍼스 크기를 균질화하는 기준을 ‘집단 주체’ 합계를 기준으로 한 결과, ‘민중’의 경우 일제시기는 거의 그대로 개수가 보전된 반면, 1980년대는 1/5로 축소되었기 때문으로 생각된다. 이는 일제시기에 해당하는 2대구간에 비해 1980년대가 포함된 5대구간에 ‘국민’의 비중이 압도적으로 높아서, 합계 크기의 균질화 과정에서 5대구간 ‘민중’이 지나치게 축소된 결과를 초래했다.

그러므로 ‘민중’의 의미를 좀 더 집중적으로 고찰하려면 ‘민중’ 문서만으로 토픽모델링을 실행하고 이를 토대로 시기구분과 연결망 분석을 진행할 필요가 있다. 그런데 토픽모델링 실행 전에 ‘민중’ 코퍼스도 구간별 정규화 조치가 필요하다.

지나치게 확대해서 해석한다고 말할 수 있다. 이런 지적은 일부 타당하고 앞으로 더 음미할 만하나, 2장에서 다루는 ‘집단 주체’ 차원의 자료로는 ‘민중’의 의미를 본격적으로 분석하기엔 불충분하다고 생각하여 본격적 분석은 3장으로 미룬다. 여기서는 전체 동향 중에서 ‘민중’의 위상을 살펴보는 데 필요한 범위 내에서 서술한 것이다.

〈표 4〉 ‘민중’ 코퍼스의 규모 조정

코퍼스	대구간						합계	
		1bg	2bg	3bg	4bg	5bg	개	%
최초	14	1,523	477	1,014	2,874	5,902	100.0	
‘집단 주체’ 균형코퍼스	14	1,523	477	512	572	3,098	52.5	
‘민중’ 균형코퍼스	14	508	477	512	572	2,083	35.3 (67.2)	

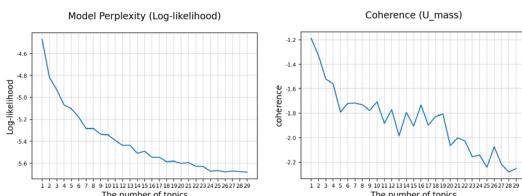
- ‘bg’는 대구간을 뜻하는 ‘big grid’의 줄임말이다.
- 음영으로 표시한 칸의 숫자는 코퍼스 규모를 줄인 결과이다.

〈표 4〉에서 볼 수 있듯이 ‘민중 균형 코퍼스’는 ‘집단 주체 균형 코퍼스’에서 2대구간의 ‘민중’ 문서 개수를 1/3로 축소해서 3~5대시기 개수와 맞추는 방식으로 산출했다. 2대시기의 문서 규모를 크게 축소하는 것을 우려할 수도 있지만, 그 규모를 유지하거나 소폭 축소할 경우 후속하는 3대시기의 크기가 상대적으로 너무 작아져, 3대시기의 단어가 토픽에 과소 반영되는 편향성이 생길 수 있다. 이 글에서는 그보다는 대시기별 문서 개수를 500개로 하향평준화하는 것이 더 합리적이라 판단했다. 물론 이 정도 규모의 문서 개수로도 각 구간별 ‘민중’의 동향을 파악하는 데 무리가 없을 것이라는 생각도 작용했다.²⁶⁾

DMR 모델에 2,083개의 ‘민중’ 문서와 29개의 구간 정보를 입력하고, 복잡도와 토픽일관성 지표를 추출해서 토픽 개수를 10개로 선택한 뒤, 다른 파라미터는 ‘집단 주체’와 동일하게 하여 토픽모델링을 진행했다.²⁷⁾

26) 1대시기는 워낙 ‘민중’의 빈도가 과소해서 예외로 간주해서 크기 조정은 하지 않았다.

27) 다음 왼쪽은 혼잡도 결과 그래프, 오른쪽은 토픽일관성 결과 그래프이다.



이 역시 5개의 후보 결과를 추출해서 검토한 뒤 최종적으로 적합한 결과값을 선정해서 토픽 제목을 붙였다.²⁸⁾ 그 결과는 <표 5>와 같다.

<표 5> ‘민중’의 토픽 및 토픽 단어

토픽	토픽 제목	토픽 단어 (상위 20개)
T00	선동·방송	선동, 미국, 민중, 방송, 정부_8, 협력, 지금, 민중서관, 불안, 관심, 연출, 사관, 간판, 외국, 지구_3, 자기_2, 경찰관, 동포, 재정경제위원회, 의원_6
T01	생활·일반	민중, 생활, 일반, 조선_5, 직접, 경찰, 진정_2, 힘_0, 사회_7, 선전_2, 정치, 자신_1, 부담, 능력, 존재, 파괴, 문예, 자유, 행동, 자발
T02	정치·투쟁	투쟁, 정치, 미술, 민중, 민주화, 역사, 운동, 생존권, 민주주의, 작가, 보수_5, 통일, 연합, 주체, 체제, 보호, 국민, 교육, 한국, 노동자
T03	조선·사회	조선_5, 일반, 민중, 기관_11, 사회_7, 생활, 의사_2, 보건, 중대, 정치, 無, 원인, 절대, 중국, 다수, 오인_4, 정부_8, 자유, 생명, 행정_1
T04	제도·정치	민중당, 의원_6, 후보, 대통령, 국회, 야당, 선거, 신한당, 민주공화당, 案, 대표, 소속, 黨, 위원회, 대변인, 박순천, 강경파, 박정희, 제출, 정책
T05	문학·역사	문학, 역사, 민중, 억압, 사람, 해방, 민족, 민중극단, 당국, 조선_5, 교육, 정권, 남북, 투표, 노래, 역량, 탄압, 요구, 운동, 공복_2
T06	개최·통치	개최, 통치, 민중, 민중동맹, 지지_5, 대회, 기대_2, 반대, 노력, 목적, 조선_5, 진행, 단체, 주최, 운동, 이승만, 교육, 신탁, 절규, 합작
T07	극장·공연	민중당, 민중극장, 극단_2, 회의_3, 공연, 재야, 통합, 과정, 연극, 주류, 신한당, 반응, 쟁택, 法, 黨, 인사_1, 양당, 민주, 세력, 전당대회
T08	예술·희망	예술, 민중, 희망, 문학, 국민, 이익, 현실, 복리_1, 자기_2, 미술, 건설, 사회_7, 통일, 지배, 계급, 민족, 민주, 민중운동, 북한, 경찰
T09	광주·학생	광주민중항쟁, 학생, 구속, 후보, 민중당, 노동자, 詩, 저적_5, 비판, 책임자, 사람, 혁명, 지역, 촉구, 민주, 결성, 시인_10, 광주, 처벌, 대변

다음으로 구간별 토픽 구성을 활용해서 ‘민중’의 토픽 구성 변화를 반영한 마디, 즉 시기 구분을 실행했다. 구간별 상관계수를 표시한 결과는 다음과 같다.

28) seed 번호 10000을 선택했다.

1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19		20		21		22		23		24																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
019		020		021		022		023		024		025		026		027		028		029		030		031		032		033		034		035		036		037		038		039		040		041		042		043		044		045		046		047		048		049		050		051		052		053		054		055		056		057		058		059		060		061		062		063		064		065		066		067		068		069		070		071		072		073		074		075		076		077		078		079		080		081		082		083		084		085		086		087		088		089		090		091		092		093		094		095		096		097		098		099		100		101		102		103		104		105		106		107		108		109		110		111		112		113		114		115		116		117		118		119		120		121		122		123		124		125		126		127		128		129		130		131		132		133		134		135		136		137		138		139		140		141		142		143		144		145		146		147		148		149		150		151		152		153		154		155		156		157		158		159		160		161		162		163		164		165		166		167		168		169		170		171		172		173		174		175		176		177		178		179		180		181		182		183		184		185		186		187		188		189		190		191		192		193		194		195		196		197		198		199		200		201		202		203		204		205		206		207		208		209		210		211		212		213		214		215		216		217		218		219		220		221		222		223		224		225		226		227		228		229		230		231		232		233		234		235		236		237		238		239		240		241		242		243		244		245		246		247		248		249		250		251		252		253		254		255		256		257		258		259		260		261		262		263		264		265		266		267		268		269		270		271		272		273		274		275		276		277		278		279		280		281		282		283		284		285		286		287		288		289		290		291		292		293		294		295		296		297		298		299		300		301		302		303		304		305		306		307		308		309		310		311		312		313		314		315		316		317		318		319		320		321		322		323		324		325		326		327		328		329		330		331		332		333		334		335		336		337		338		339		340		341		342		343		344		345		346		347		348		349		350		351		352		353		354		355		356		357		358		359		360		361		362		363		364		365		366		367		368		369		370		371		372		373		374		375		376		377		378		379		380		381		382		383		384		385		386		387		388		389		390		391		392		393		394		395		396		397		398		399		400		401		402		403		404		405		406		407		408		409		410		411		412		413		414		415		416		417		418		419		420		421		422		423		424		425		426		427		428		429		430		431		432		433		434		435		436		437		438		439		440		441		442		443		444		445		446		447		448		449		450		451		452		453		454

여기서 ‘민중’의 상관 계수 간 응집도가 뚜렷이 높은 구간은 일제시기와 1980년대 후반 이후이다. 이는 선행 연구의 성과와 부합하는 결과이다. 그런데 그 사이의 구간, 즉 1945~1983년의 39년을 상관 계수가 낮다는 기준 하에서 동질적 시기로 상정하기는 무리라고 생각이 들었다. 더욱이 그 중에서 1960년대 중반(18~19g)은 상관 계수가 높기도 해서 이를 그 전후 시기와 동일시하는 것도 부적절하다고 판단했다. 그래서 이 39년 구간은 상관 계수의 기준을 재조정해서 평균 이상에 해당하는 부분은 음영(빨강) 표시를 했다. 그 결과가 작은 사각형 그래프로 표시한 부분이다. 그리고 이 표시를 기준으로 다시 3개의 세부 시기를 구분했다.

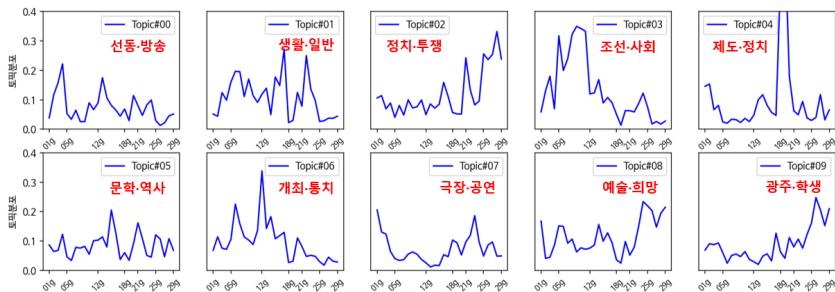
그 결과 전체 시기는 모두 6시기로 구분할 수 있는데, 이를 ‘집단 주체’의 시기 구분과 비교한 것이 〈표 6〉이다.

〈표 6〉 ‘민중’과 ‘집단 주체’의 시기구분 비교

	시기	1기	2기	3기	4기	5기	6기
민중	기간	1898~1910	1920~1940	1945~1962	1963~1971	1972~1983	1984~1999
	문서	14 (0.7)	508 (24.4)	498 (23.9)	394 (18.9)	113 (5.4)	556 (26.7)
집단 주체	시기	1기	2기	3기	4기	5기	
	기간	1898~1910	1920~1940	1945~1962	1963~1968		1969~1999
	문서	2,489 (7.3)	6,689 (19.6)	8,488 (24.8)	2,422 (7.1)		14,084 (41.2)

- ‘문서’의 단위 = 개(%)

이렇게 구분된 시기별로 ‘민중’의 토픽들 동향을 살펴보면 〈그림 7〉과 같다. 그래프에서 x축에 눈금이 표시된 것은 6개 시기의 기점이 되는 구간이다. 29g는 마지막 구간이라서 표시했다. 전체적으로 20세기 말로 갈수록 상승세가 뚜렷한 토픽은 ‘정치·투쟁’(2)과 ‘광주·학생’(9)이다. 이는 곧 ‘민중’의 경우 “운동” 성향의 의미가 점차 강화되어 갔음을 나타낸다.



〈그림 7〉 ‘민중’ 토픽들의 시기별 변화 (y축 값: 분포값)

이와 반대로 20세기 전반에 높다가 20세기 말로 갈수록 하강세가 뚜렷한 토픽은 ‘조선·사회’(3), ‘개최·통치’(6)이다. 전자는 토픽 단어의 구성을 볼 때 포괄적인 방면에 걸쳐 사용된 용례를 반영하는 것으로 판단된다. 후자는 해방 직후의 신탁통치 국면에서 활성화된 용례를 반영한다. 이 점에서 두 토픽의 동향은, 2장의 ‘25개 단위 간 연결망 지도’에서 일제시기와 해방 ~1960년대 초까지 ‘민중’이 ‘집단 주체’의 중심에 위치한 사실과 호응한다. 왜냐하면 여기서 ‘민중’의 ‘조선·사회’(3), ‘개최·통치’(6) 토픽이 2기와 3기에 활성화 되었을 뿐 아니라 내용 면에서도 시대 상황 하에서 포괄적 내용을 담고 있어서 다른 ‘집단 주체’ 단어들과 잘 연계될 수 있었기 때문이다.

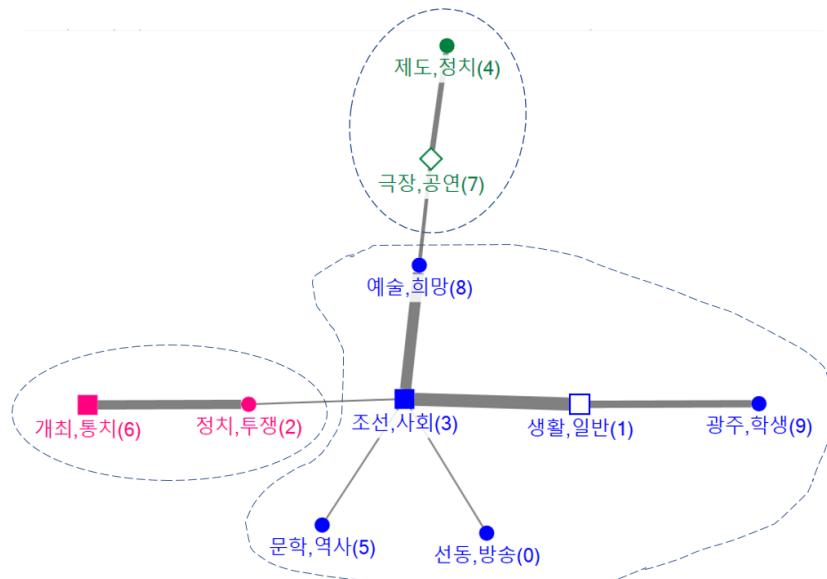
한편, 뚜렷한 상승 및 하강 추세 없이 표준편차 0.2이하에서 진동하는 토픽은 ‘선동·방송’(0), ‘문학·역사’(5)이다. 특히 후자는 문화적이고도 사상적이며 미래 지향적 기대 등이 많이 반영된 토픽으로 판단된다. 이외는 조금 달리, 전체적으로 진동을 보이지만 20세기 후반에 다소 활성화되는 토픽은 ‘극장·공연’(7)과 ‘예술·희망’(8)이다.

이외에 ‘제도·정치’(4) 토픽은 1960년대 중반의 민중당 창당에 관련된 것이다. ‘생활·일반’(1)은 토픽 단어의 면면에서 ‘조선·사회’(3)와 유사하지만,

그 세력이 작으면서도 일제시대뿐 아니라 1950년대와 1970년대까지 활성화된 점이 다르다.

2) 시기별 중심 의미의 변화: 정치성·운동성의 강화

여기서는 앞에서 도출한 10개의 토픽들이 어떻게 연결되어 있는지를 살펴본다. 문서 수가 다른 시기에 비해 아주 적은 1기와 5기를 제외했다. 도출한 연결망 지도는 〈그림 8〉~〈그림 11〉과 같다.



〈그림 8〉 '민중'의 토픽 연결망 지도 (2기, 1920~1940)

〈그림 8〉은 2기의 '민중' 문서 508개가 가진 토픽 구성을 활용해서 10개 토픽 간의 연결관계를 표현한 것이다. 연결망은 패스파인더(PFnet) 방식이

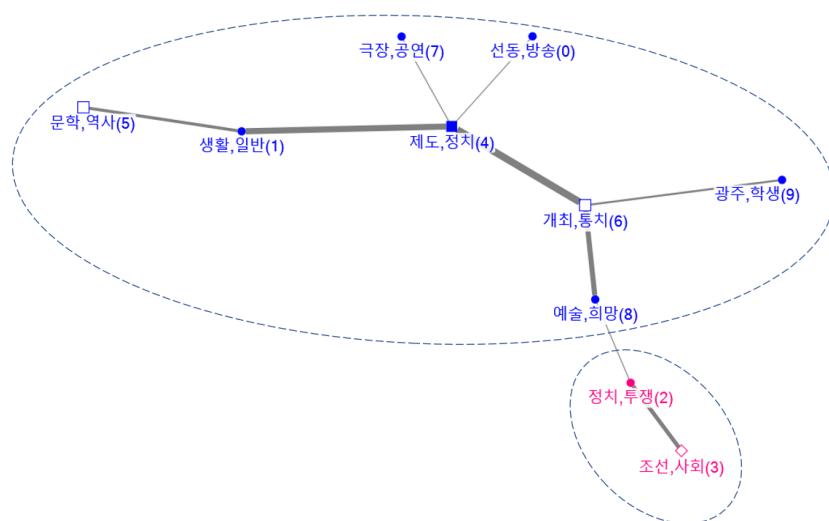
며, 토픽들의 군집은 타원형 점선으로 둘러싸서 서로 구분했다.²⁹⁾ 링크 값은 코사인유사도를 나타내며 값이 높을수록, 즉 두 토픽 간의 관계가 긴밀 할수록 굵게 표현했다. 노드 이름은 토픽 제목과 토픽 번호를 붙였다. 노드 형태의 경우, 전체 연결망 차원에서 중요도가 높은, 즉 전역(全域) 중심성이 높은 상위 토픽 3개를 흰색 정사각형으로 표시했다. 이 때의 중심성 척도로는 ‘삼각매개중심성’(rTBC)을 이용했다.³⁰⁾ 전체가 아닌 각 군집에서 가장 중심이 되는, 즉 지역 중심성이 1위인 토픽은 흰색 마름모로 표시했다. 이 때의 중심성 척도로는 일반화된 ‘이웃중심성’(NC)를 이용했다.³¹⁾ 양자가 겹치는 토픽은 검정색 정사각형으로 표시했다. 이런 방식을 이하의 연결망 지도들에도 동일하게 적용했다. 연결망 산출 및 중심성 지표에 관한 상세한 내용은 이 글 마지막의 ‘[부록 2] 주요 시기별 ‘민중’의 토픽 연결망 지도 산출 및 중심성 정보’에 제시했다.

29) 이 글 3장에서 패스파인더 연결망 지도를 그리는 데 필요한 입력 데이터와 중요한 중심성 지표들은 이재윤이 만든 WNET 0.4와 NC 0.4를 활용하여 산출했다.

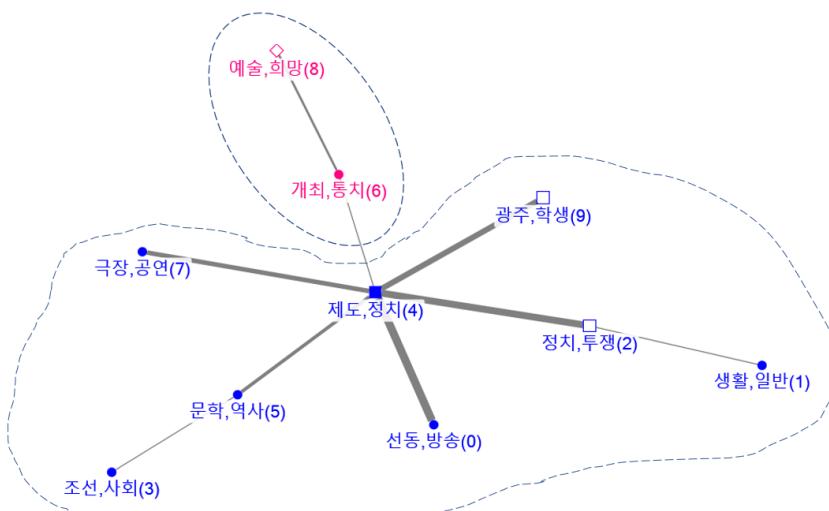
‘WNET’은 ‘가중 연결망’(weighted network)를 가리키며, ‘NC’는 ‘이웃 중심성’(Neighbor Centrality)을 가리킨다. 파이썬으로 토픽 간 코사인유사도를 산출한 후 이 값을 그러한 툴에 입력해서 주요 값들을 얻었다.

30) 가중 연결망의 중심성 측정으로 삼각매개중심성이 유용한 점에 관하여는 다음 연구를 참조했다. 이재윤, 「계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구」, 『한국문현정보학회지』 40-3, 한국문현정보학회, 2006.

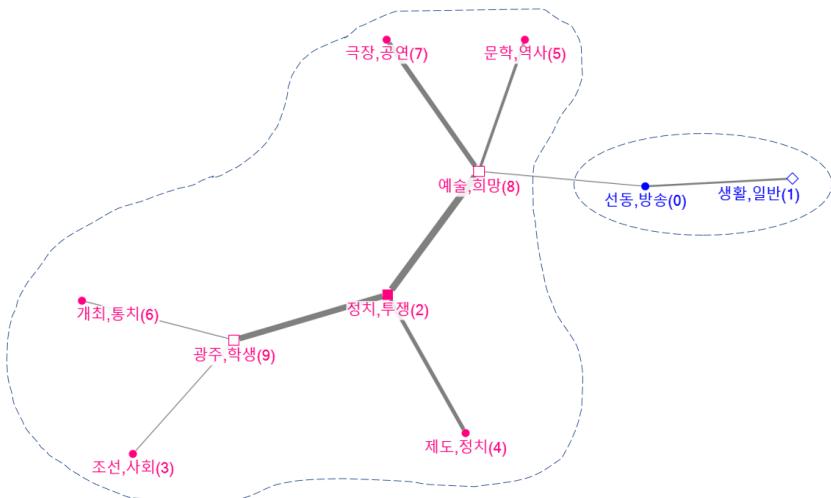
31) 이웃중심성에 관하여는 다음 연구를 참조했다. 이재윤, 「가중 네트워크를 위한 일반화된 지역중심성 지수」, 한국정보관리학회, 『정보관리학회지』 32-2, 2015.



〈그림 9〉 '민중'의 토픽 연결망 지도 (3기, 1945~1962)



〈그림 10〉 '민중'의 토픽 연결망 지도 (4기, 1963~1971)



〈그림 11〉 '민중'의 토픽 연결망 지도 (6기, 1984~1999)

이상의 연결망 지도는 각 시기 '민중'의 의미를 시각화한 것이라 할 수 있다. 그 특징을 살펴보기 위해 연결망 지도에서 중심이 되는 토픽을 정리해 보면 〈표 7〉과 같다.

〈표 7〉 '민중' 연결망 지도의 시기별 중심 토픽

시기 범위	2기	3기	4기	6기
연결망 전체	조선·사회(3) 생활·일반(1) 개최·통치(6) *	제도·정치(4) ** 문학·역사(5) 개최·통치(6) *	제도·정치(4) ** 정치·투쟁(2) *** 광주·학생(9) ****	정치·투쟁(2) *** 예술·희망(8) 광주·학생(9) ****
군집별	조선·사회(3) 개최·통치(6) 극장·공연(7)	제도·정치(4) 조선·사회(3)	제도·정치(4) 예술·희망(8)	정치·투쟁(2) 생활·일반(1)

- '연결망 전체'에서는 전역 중심성 1~3위 토픽을, '군집별'에서는 지역 중심성 1위 토픽을 나열했다.
- 전역 중심성과 지역 중심성이 겹치는 토픽은 굵게 표시하고 밑줄을 그었다.
- '연결망 전체'에서 인접 시기와 연속되는 토픽은 두 토픽에 동일한 개수의 별표(*)를 표시했다.

4개의 시기별 연결망 지도와 〈표 7〉을 통해 다음 두 가지 사실을 알 수 있다. 첫째, ‘민중’의 중심적 의미는 일제 시기에는 ‘조선·사회’였으나 해방 후 1970년대 초까지는 ‘제도·정치’로, 그리고 1984년부터 1999년까지는 ‘정치·투쟁’으로 변했다. 둘째, 후속 시기까지 시야에 넣고 파악할 경우, 2기와 3기 간에는 ‘개최·통치’가, 3기와 4기 간에는 ‘제도·정치’가, 4기와 6기 간에는 ‘정치·투쟁’ 및 ‘광주·학생’이 각각 지속되었다.

이러한 사실에 대해서 다음 두 가지 해석을 내릴 수 있다. 첫째, 〈표 5〉의 토픽별 토픽 단어들을 보면 ‘조선·사회’ 토픽은 다른 토픽에 비해 좀 더 다양하고 포괄적인 내용을 담고 있다고 생각된다.³²⁾ 또한 토픽 단어에 ‘정치’가 포함되어 있지만 전체적으로 보면 다른 토픽에 비해 정치적인 성향은 적다. 그에 비해 ‘제도·정치’ 토픽은 대부분 ‘민중당’을 비롯하여 제도권 정치에 해당한다. 나아가 ‘정치·투쟁’ 토픽의 다수는 정치 민주화를 지향하는 정치 투쟁 혹은 재야 운동권의 동향을 나타낸다. 그러므로 이상을 통하여 20세기 ‘민중’의 중심 의미는 전기에 ‘정치·사회’처럼 ‘비정치적’ 성향을 띠었다가 중기에 제도적 정치화가 두드러졌고, 후기에는 정치 운동화 경향을 드러냈다고 말할 수 있다.

둘째, 이러한 ‘민중’의 의미 변화는 이 글 2장 3절에서 살펴본 ‘집단주체’ 속 ‘민중’의 위상 변화를 좀 더 자세하게 이해하는 데 도움이 된다. 〈그림 5〉에서 2기와 3기의 ‘집단 주체’는 5개 키워드가 시기별로 군집을 이룬다는 점, 그리고 그 군집의 중심에 ‘민중’이 있다는 점에서 공통적이다. 그렇지만

32) 다른 토픽들도 다양성과 포괄성을 일정 정도 가지고 있다고 할 수 있다. 그런데 ‘조선·사회’의 토픽 단어들을 보면 다른 토픽에 비해 고유명사와 복합명사의 개수가 적다. 또한 ‘기관’(機關), ‘정부’, ‘행정’(行政) 등 정부조직이나 단체와 관련한 단어들, ‘조선’(朝鮮), ‘사회’, ‘일반’, ‘생활’ 등 더 광범위한 대상이나 일상을 가리키는 단어들, ‘자유’, ‘생명’, ‘절대’, ‘중대’ 등 추상적 개념에 해당하는 단어 들이 병존해 있다.

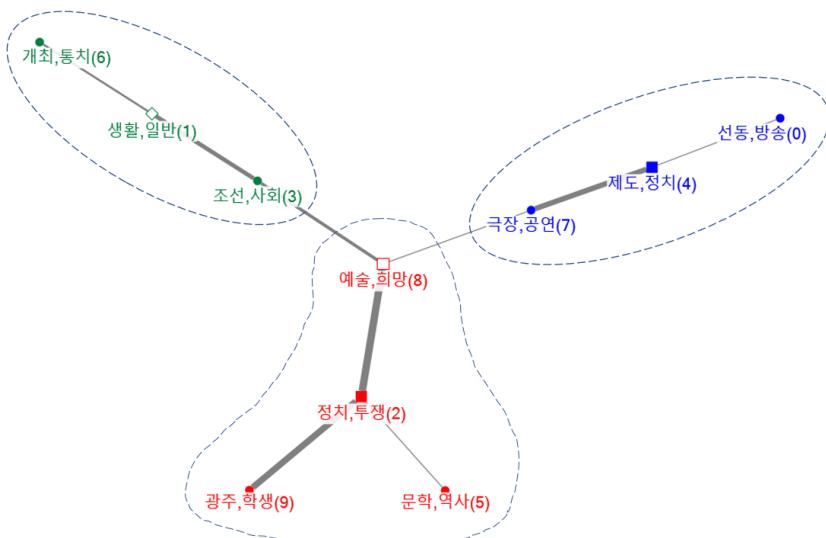
미세한 차이도 보였는데, 2기에는 ‘민중’이 중심에 있고 4개 키워드가 ‘민중’과 각각 연결된 데 비해, 3기의 경우 ‘민중’으로부터 ‘국민’과 ‘인민–민족’이 나누어지는 형국을 하고 있다. 말하자면 ‘집단 주체’ 차원에서 볼 때 2기와 3기의 동향은 공통점과 차이점을 함께 가지고 있고, 그 중 차이점을 3기 이후 ‘집단 주체’의 이원화 경향과 유관한 듯 하다. 그런데 위 〈표 7〉을 보면 전체 연결망 차원에서 2기와 달리 3기 ‘민중’에서 ‘제도·정치’의 비중이 상대적으로 높았고, 이것은 4기로 지속되었다. 그런데 2장 3절에서 살펴보았듯이 3기부터 ‘집단 주체’ 중 ‘국민’이 특히 다른 키워드에 비해 ‘제도·정치’의 경향이 두드러졌다. 이 두 사실을 결합하면, 3기에 들어서 ‘국민’과 ‘민중’ 간에는 ‘집단·주체’를 매개로 접점이 마련된 것으로 보인다. 결국 ‘민중’의 ‘집단 주체’ 속 위상과 ‘민중’의 의미를 종합적으로 고려할 때, 3기의 ‘민중’은 한편에서 2기와 마찬가지로 ‘집단 주체’의 중심적 위상을 계승하면서, 다른 한편으로는 ‘집단 주체’의 전체 흐름 속에서 ‘제도·정치’를 매개로 ‘국민’ 계열이 분기하는, 분기점에 놓여 있었다고 할 수 있다.

3) 장기지속적 의미: 미래에 대한 기대

이 절에서는 앞 절에서 시기별로 나누어 살펴본 ‘민중’의 의미를 한 데 묶어서 전체적으로 살펴보고, 이를 통해 20세기 ‘민중’의 의미에서 지속되는 것은 무엇인지를 살펴보자 한다. 2,083개 문서를 앞 절과 동일한 방식으로 가공해서 만든 연결망 지도는 〈그림 12〉와 같다.

〈그림 12〉을 〈그림 8〉~〈그림 11〉 및 〈표 7〉과 비교해 보면, 〈그림 12〉에서 군집으로 구분된 각각은 ‘민중’의 시기별 주요 연결망을 반영하고 있음을 발견할 수 있다. 즉, ‘조선·사회’(3), ‘생활·일반’(1), ‘개최·통치’(6)로 이어

진 좌상의 군집은, 2기의 전역 중심성 1~3위와 동일하다. ‘극장·공연’(7), ‘제도·정치’(4), ‘선동·방송’(0)으로 이어진 우상의 군집은, 3기와 4기에 걸쳐 연속해서 나타난다. 중앙 하단의 군집에서 중심을 이루는 ‘예술·희망’(8), ‘정치·투쟁’(2), ‘광주·학생’(9)은 6기 연결망의 중심 골격과 동일하다. 그러므로 〈그림 12〉의 연결망은, 좌상부터 시계방향으로 각각 2기(1920~40년), 3·4기(1945~71년), 6기(1984~99년)의 ‘민중’ 의미를 뚜렷하게 반영한 것이라 할 수 있다.



〈그림 12〉 ‘민중’의 토픽 연결망 지도 (전체, 1898~1999)

이처럼 〈그림 12〉의 연결망은 20세기에 계기적으로 나타났던 ‘민중’의 의미를 하나의 공간에서 갈무리하여 표현하고 있다. 그렇다면 이 연결망을 통해 20세기 ‘민중’의 의미 중 장기 지속하는 의미도 알 수 있으리라 기대해 볼 수 있다.

필자는 이 연결망에서 전역 중심성 상위 토픽들이 일단 그 후보가 될 수 있을 것이라 생각한다. 이것을 ‘일반화된 삼각매개중심성’ 척도(rTBC)로 그 후보를 1위부터 내림차순으로 열거하면 ‘예술·희망’(8), ‘정치·투쟁’(2), ‘제도·정치’(4) 순이다.³³⁾

이 세 후보 중에서 ‘예술·희망’(8)이 특히 주목되는데, 그 이유는 이 토픽은 연결망의 중하단 군집에 속하면서도 좌상과 우상 군집과도 연결되기 때문이다. 전술했듯이 세 군집이 20세기 초·중·후기의 ‘민중’ 의미를 각각 반영하는 사실을 상기하면, ‘예술·희망’(8)은 결국 20세기 ‘민중’의 의미 흐름에서 장기지속하는 요소라고 말할 수 있다.

그렇다면 이 ‘예술·희망’(8)은 ‘민중’의 전체 의미에서 어떤 변별성을 가질까. 이 토픽이 다른 9개의 토픽과 가장 변별되는 점은 그 제목에서도 알 수 있듯이, ‘미래에 대한 기대의 형상화’가 아닐까 생각된다. 이 토픽의 토픽 단어 가운데 ‘예술’, ‘미술’, ‘문학’ 등이 그 형상화의 수단이라면, 미래에 대한 기대는 ‘희망’, ‘건설’ 등과 관계 깊다고 생각된다. 20세기 전체에 걸쳐 ‘민중’은 주요 시기별로 중요한 개성적 특질을 드러내면서도, 이처럼 그 저변에서는 미래에 대한 기대를 드러내는 의미가 장기지속하고 있었다고 말할 수 있다.

33) 각 토픽의 해당 값은 다음과 같다.

토픽	rTBC
T08	0.63889
T02	0.55556
T04	0.41667

4. 맷음말

이 글에서는 한국에서 사용된 ‘민중’의 의미를 1세기에 걸쳐 조망했다. ‘민중’은 20세기 한국사의 개성을 잘 담고 있다는 전제 위에서, “20세기 전반과 후반의 ‘민중’은 서로 비슷한 의미를 가졌다”를 연구 질문으로 제시했다. 장기적 관찰에 치중하느라, ‘의미’ 파악의 범위는 주요 일간지 기사라는 텍스트 차원에 제한했다. 그렇지만 텍스트의 범위는 ‘민중’뿐 아니라 ‘국민’, ‘인민’, ‘민족’, ‘대중’의 문맥까지 확장했다. ‘민중’의 의미를 알기 위해 인접 단어와의 관계도 중요하게 고려했기 때문이다.

연구는 두 단계로 진행했다. 첫째 단계에서는 ‘민중’을 ‘집단 주체’의 전체 동향 속에서 파악했다. ‘의미 유사성’을 기준으로 산출한 ‘집단 주체’ 상호 간의 연결망에서 ‘민중’의 위치를 파악하는 일에 주안점을 두었다. 이 글에서 가장 노력을 기울인 것은 단어의 ‘의미’를 규정하고 그것을 추출하는 일 이었다. 이 목적을 위해 텍스트마이닝의 토픽모델링을 활용했다. ‘집단 주체’의 키워드가 사용된 문맥(=‘문서’)을 대규모로 수집해서 토픽을 추출하면, 그 토픽들은 ‘집단 주체’의 일반적 의미를 이루는 하위 의미들에 상응한다는 발상이 중요했다. 이런 목적으로 3만 여개의 문서에서 10개의 토픽을 추출 했다. 첫 단계 연구에서는 10개 토픽 각각의 내용보다 문서나 구간을 이루 는 토픽의 구성, 즉 각각의 비중이 더 중요했다. 이 토픽 구성은 근거로 ‘집 단 주체’의 의미 변화를 시기구분했고, 나아가 5개 키워드 각각이 5개 시기 별로 어떻게 연관되는지를 연결망 지도로 시각화했다. 그 결과 ‘민중’의 위 상이 20세기 전반에는 ‘집단 주체’ 중에서 중심적 위치에 있었으나, 20세기 후반에는 주변화되었음을 확인했다.

둘째 단계에서는 ‘민중’의 의미를 본격적으로 살펴보았다. 여기서는 ‘민

‘중’만의 문서 2천 80여개를 대상으로 10개의 토픽을 추출했다. 첫째 단계와 거의 동일하게 토픽 구성을 근거로 시기를 구분하고 토픽 연결망을 작성했다. 다만 ‘집단 주체’의 경우 1969~1999년이 동질적 시기로 묶인 것과 달리, ‘민중’에서는 1984년이 그 이전과 이후를 나누는 뚜렷한 획기가 되었다. 이 단계에서는 ‘민중’의 토픽들을 통해 하위 의미들의 동향도 살펴보았다. 일단, 앞 단계 연구처럼 20세기 전반과 후반이 뚜렷하게 대비되는 점을 확인했다. 즉, 일제시기에 해당하는 20세기 전기에는 ‘조선·사회’ 토픽이, 중기에는 ‘제도·정치’ 토픽이, 후기에는 ‘정치·투쟁’ 토픽이 각각 득세했다. 그렇지만 이러한 대비와 단절의 외양 하에서도, ‘예술·희망’처럼 미래를 향한 기대가 지속되고 있었다.

이 글은 20세기의 대규모 데이터를 대상으로 ‘민중’의 의미를 파악하는 데 치중하느라 『황성신문』과 『대한매일신보』, 『동아일보』와 『한겨레신문』 사이의 차이는 다루지 못했다. 또한 ‘민중’의 ‘의미’를 파악하는 데 컨텍스트적 맥락을 경시했다. 나아가 ‘집단 주체’에 속하는 다른 네 개 키워드의 개별 동향을 살펴볼 여유를 가지지 못했다. 이런 세 가지 측면은 모두 ‘민중’의 의미를 입체적으로 파악하는 데 중요한 요소이지만, 이 글에서는 ‘집단 주체’를 끙끙그려 파악하면서 ‘민중’과의 관련성을 파악하는 데 만족해야 했다. 이런 요소들에 관한 본격적인 분석은 후속 연구를 통해 진행하고자 한다.

[부록 1] ‘집단 주체’ 코퍼스 작성 기준

1. 대원칙

- 1) 대상 신문: 『황성신문』(1898~1910), 『대한매일신보』(1905~1910), 『동아일보』(1920~1940, 1945~1999), 『한겨레신문』(1988~1999)
- 2) 추출 문맥: ‘집단 주체’에 해당하는 5개 키워드(‘국민’, ‘인민’, ‘민족’, ‘민중’, ‘대중’)가 사용된 기사의 앞뒤 20음절(공백 제외)³⁴⁾ – 이 때 이 키워드가 포함된 복합명사도 배제하지 않고 동일한 원칙이 적용된다. 키워드의 한자 표기도 포함된다.
- 3) 선택/배제: 한글 형태소 분석기를 사용한 뒤 명사, 고유명사 중심으로 추출했다(대명사, 부사, 수량사 등 제외). 그렇지만 의미상 중요하다고 생각한 경우에는 다른 품사의 단어도 일부 추출했다. / 한자어가 ‘하다’ 등의 용언과 결합한 경우에는 한자어 부분을 추출했다. 동일 문맥에서 동일 단어가 2번 이상 충복 출현할 경우에도 어느 한쪽을 생략하지 않고 그대로 두었다. / 신문 기사의 성격상 내용적 변별성 없이 빈출하는 단어의 경우 제외했다. 예) 신문사 이름, ‘신문’, ‘기자’, ‘예상’, ‘최근’ 등

2. 세부 원칙

1) 표기·띄어쓰기 원칙

- (1) 단어는 가급적 한글로 표기하고 그것이 여의치 않을 경우만 한자로

³⁴⁾ 키워드가 ‘□국+민’ 형태일 때의 ‘민’은 ‘국민’으로 간주. ex) ‘한국민’ → ‘한국 국민’, ‘자국민’ → ‘자국 국민’ 등.

표기했다.

- (2) 동음이의어는 국립국어원의 “표준국어대사전”의 구분에 따라 한글 뒤에 ‘_’를 넣고 어깨번호를 붙여 서로 구별했다.
- (3) 띠어쓰기도 표준국어대사전을 기준으로 했다. 음절의 조합이 표제어에 있는 경우에는 하나의 단어로 보아 그것을 형태소로 했고, 표제어에 없으면서 하위 단어들로 세분할 수 있을 경우 나누었다.
 - ex1) 표제어가 있는 경우 → ‘국민투표’, ‘약소민족’, ‘전당대회’, ‘남북통일’, ‘재외국민’, ‘민족의식’, ‘국민운동’ 등
 - ex2) 표제어가 없는 경우 → ‘주체 의식’, ‘민족주의 운동’ 등
- (4) –主義 : 모두 붙였다. ex) ‘민주주의’, ‘자본주의’, ‘군국주의’

2) 고유명사

- 책, 논문, 강연회, 작품 등의 이름은 모두 붙였다.
- 당명, 사건명, 법규명, 단체명, 기관명 등은 음절 구분하지 않고 모두 붙였다.
- 기관명 중 줄임말인 경우 가급적 풀네임(full name)으로 고쳤다. ex) ‘안기부’ → ‘국가안전기획부’, ‘보안사’ → ‘국군보안사령부’·‘보안사령부’, ‘한은’ → ‘한국은행’
- ‘UN’, ‘국제연합’, ‘국련’ 등은 ‘유엔’으로 통일.
- 국가명은 현대 국가명으로 변경했다. ex) ‘화란’ → ‘네덜란드’, ‘토이기’ → ‘터키’
- 인명: 대부분 제외하되, 대통령, 국가원수, 역사적 중요 인물의 경우 포함했다.³⁵⁾

35) ‘양김씨’, ‘양김’, ‘두김’, ‘두김씨’는 맥락에 따라 이름을 찾아 표기. ‘양김’이 누군지

- 지명: 국내외 모두 가급적 제외했다.
- 대학명: 풀네임(full name)으로 표기 ex) ‘고대’ → ‘고려대학교’, ‘이대’ → ‘이화여자대학교’
- 숫자가 들어간 역사적 사건은 한글로 통일 : ‘3.1운동’ → ‘삼일운동’, ‘6.10만세운동’ → ‘육십만세운동’, ‘광주’(1929년) → ‘광주학생운동’, ‘지나사변’(1937년) → ‘중일전쟁’, ‘8.15’(1945년) → ‘팔일오해방’, ‘6.25’(1950년) → ‘육이오전쟁’, ‘3.15’(1960년) → ‘삼일오의거’, ‘4.19’(1960년) → ‘사일구혁명’, ‘6.3사태’(1964년) → ‘육삼사태’, ‘광주사태’(1980년) → ‘광주민중항쟁’, ‘4.13조치’/‘4.13호헌’(1987년) → ‘사일삼호헌조치’

3) 접미어, 접두어 등

- ‘적’(的)은 특별한 경우가 아니면 제외했다. 예외) ‘전적(全的)’ 등
- ‘화’(化)와 ‘성’(性)은 앞 명사와 붙였다. ex) ‘근대화’, ‘민족성’, ‘주체성’, ‘실효성’
- ‘대’(對)가 국가명 앞에 사용될 때에는 붙였다. ex) ‘대소’, ‘대미’, ‘대영국’, ‘대일’
- ‘전’(全)이 명사 앞에 올 경우 삭제했다. ex) ‘전국민’ → ‘국민’, ‘전세계’ → ‘세계’

4) 1음절 한자어

- 형태소 분석 단계에서 남긴 경우라도 ‘집단 주체’의 의미 파악에서 중요하다고 생각되는 일부는 한자로 바꾸고 코퍼스에 포함했으며, 제외하는 것이 좋다고 판단한 경우에는 한글 표기를 유지한 뒤 1음절 한글

확인이 안 되면 그대로 ‘양김’으로 표기.

글자는 일괄 제외했다.

- 1음절 한글자를 남길 경우, 필요시 ‘_’ 뒤에 어깨번호를 붙였다.
- 1음절 한자어 중 ‘유’(有)는 제외했고, ‘무’(無)는 중요한 단어와 많이 결합된다고 판단해서 포함했다.

5) 기타

- ‘텔레비전’, ‘테레비’, ‘텔레비죤’, ‘텔레비전’ 등은 모두 ‘TV’로 통일했다.
- ‘남과 북’은 ‘남 북’으로, ‘남북 문제’는 ‘남북 문제’로, ‘남북한’은 ‘남북 한’으로 각각 표기했다.
- ‘위원회’: 앞의 명사와 붙여 쓴 경우와, 명사와 분리해서 쓴 경우가 혼재하는데, 붙여 쓴 경우가 훨씬 많아서 그것을 기준으로 통일했다. 한편, 이 용어를 줄여서 쓴 경우에는 다음과 같이 full name으로 표기했다.
ex)‘투위’ → ‘특쟁위원회’, ‘특위’ → ‘특별위원회’
- 오늘날 기준에서 보면 부자연스러운 단어는 다음과 같이 현대어로 바꿨다.
ex)‘조해’ → ‘저해’, ‘조정’ → ‘정조’, ‘참석리’ → ‘참석’, ‘분격’ → ‘격분’, ‘하학’ → ‘하교’, ‘등대수’ → ‘등대지기’

[부록 2] 주요 시기별 ‘민중’의 토픽 연결망 지도 산출 및 중심성 정보

	토픽1	토픽2	유사도 (등급) *	토픽	전역중심성	토픽	지역중심성	군집
					rTBC		nc	
2기	T00	T03	0.5	T03	0.75	T03	4.713611	1
	T01	T03	4	T01	0.52778	T01	2.428889	1
	T02	T06	3	T06	0.52778	T08	0.928657	1
	T04	T07	2	T07	0.36111	T09	0.523094	1
	T05	T03	0.5	T05	0.27778	T05	0.423873	1
	T08	T03	3.5	T08	0.22222	T00	0.296944	1
	T09	T01	2.5	T00	0.19444	T06	1.755669	2
	T08	T07	1.5	T02	0.19444	T02	1.513988	2
	T02	T03	1	T04	0.16667	T07	1.666504	3
				T09	0.11111	T04	1.146447	3
3기	T00	T04	0.5	T04	0.63889	T04	4.645408	1
	T01	T04	3.5	T05	0.52778	T06	3.57263	1
	T02	T03	2.5	T06	0.52778	T01	1.741389	1
	T04	T06	4	T01	0.44444	T05	1.046944	1
	T05	T01	2	T03	0.38889	T08	0.999768	1
	T07	T04	0.5	T08	0.25	T09	0.38148	1
	T08	T06	3	T00	0.16667	T00	0.25053	1
	T09	T06	1.5	T09	0.16667	T07	0.145324	1
	T08	T02	1	T02	0.16667	T03	1.338681	2
				T07	0.05556	T02	1.275523	2
4기	T00	T04	4	T04	0.88889	T04	5.861111	1
	T01	T02	0.5	T02	0.47222	T02	1.797014	1
	T02	T04	3.5	T09	0.44444	T00	1.358164	1
	T03	T05	0.5	T08	0.30556	T05	1.305393	1
	T05	T04	2	T03	0.22222	T09	0.871736	1
	T06	T08	1.5	T05	0.22222	T01	0.695162	1
	T07	T04	2.5	T07	0.22222	T03	0.457251	1
	T09	T04	3	T06	0.22222	T07	0.351348	1

	토픽1	토픽2	유사도 (등급)*	토픽	전역중심성		지역중심성 nc	군집
					rTBC	토픽		
47	T04	T06	1	T00	0.16667	T08	1.429898	2
				T01	0.16667	T06	1.270601	2
6기	T00	T01	1.5	T02	0.77778	T01	1.219473	1
	T02	T08	4	T08	0.72222	T00	1.152313	1
	T03	T09	0.5	T09	0.61111	T02	4.173611	2
	T04	T02	2.5	T04	0.36111	T08	3.756944	2
	T05	T08	2	T07	0.30556	T09	2.793611	2
	T06	T09	1	T01	0.22222	T07	0.746311	2
	T07	T08	3	T05	0.16667	T04	0.732778	2
	T09	T02	3.5	T00	0.08333	T05	0.424566	2
	T00	T08	0.5	T06	0.08333	T06	0.218093	2
				T03	0	T03	0.179977	2
전체 시기	T00	T04	0.5	T08	0.63889	T04	2.342207	1
	T01	T03	2.5	T02	0.55556	T07	1.175694	1
	T02	T08	4	T04	0.41667	T00	0.394251	1
	T04	T07	3	T01	0.38889	T01	2.433811	2
	T05	T02	0.5	T05	0.33333	T03	1.583596	2
	T06	T01	1.5	T06	0.30556	T06	0.82357	2
	T09	T02	3.5	T03	0.27778	T02	3.38513	3
	T07	T08	1	T09	0.25	T08	2.256519	3
	T03	T08	2	T00	0.11111	T09	0.539219	3
				T07	0.05556	T05	0.463681	3

* – 유사도는 코사인유사도이며, 해당 숫자는 이 유사도 값을 8등급으로 나누고 가장 낮은 1등급에는 0.5, 가장 높은 8등급에는 4를 부여했다.

Abstract

The Meaning of ‘Minjung’ Used in Korea in the 20th Century - Focusing on major newspaper articles -

Hur, Soo

In this article, the meaning of ‘minjung’ used in Korea in the 20th century was examined using topic modeling of text mining. The search of this study was limited to text dimensions of articles in major daily newspapers, but expanded to consider the context of neighboring words. The study proceeded in two stages.

In the first stage, ‘minjung’ was analyzed in the overall trend of ‘collective subjects’ using topic modeling of text mining. The composition of topics was divided into periods and visualized as a network map, which showed that ‘minjung’ had a central position in the first half of the 20th century but became marginalized in the second half.

In the second stage, topics were extracted only from the documents of ‘Minjung’ and their meanings were analyzed. As a result, first, in the meaning of ‘the people’, 1984 became an important breakthrough, unlike the period division of ‘collective subject’. Second, in the early 20th century, the topic of ‘Chosun·Society’, the topic of ‘Institution·Politics’ in the middle period, and the topic of ‘Politics·Struggle’ in the latter period respectively prevailed. Third, along with this, expectations for the future, such as ‘art and hope’, continued for a long time through both periods.

Keywords: Minjung, meaning, topic modeling, collective subject, network map, long-term sustainability

참고문헌

❖ 단행본

- 박명규,『국민·인민·시민』, 소화, 2009.
 박찬승,『민족·민족주의』, 소화, 2010.
 이남희, 유리·이경희 옮김,『민중 만들기: 한국의 민주화운동과 재현의 정치학』, 후마니타스, 2015.

❖ 논문

- 강정구,「진보적 민족문학론에서 민중 개념의 형성 과정 연구」,『비교문화연구』11-2, 비교문화연구소, 2007.
- 권보드래,「동포(同胞)의 역사적 경험과 정치성 –『독립신문』의 기사분석을 중심으로」,
 『근대계몽기 지식 개념의 수용과 그 변용』, 소명출판, 2004.
- _____,「근대 초기 ‘민족’ 개념의 변화: 1905~1910년 『대한매일신보』를 중심으로」,
 『근대계몽기 지식의 굴절과 현실적 심화』, 소명출판, 2007.
- _____,「동포와 역사적 간각: 1900~1904년 ‘동포’ 개념의 추이」,『근대계몽기 지식
 의 발견과 사유 지평의 확대』, 소명출판, 2006.
- 김동택,「대한매일신보에 나타난 ‘민족’ 개념에 관한 연구」,『대동문화연구』61, 대동문화연구원, 성균관대학교 대동문화연구원, 2008.
- _____,『國民須知』를 통해 본 근대 ‘국민」,『대동문화연구』44, 성균관대학교 대동문화연구원, 2003.
- 김성보,「남북국가 수립기 인민과 국민 개념의 분화」,『한국사연구』144, 한국사연구회, 2009.
- 김윤희,「근대 국가구성원으로서의 인민 개념 형성(1876~1894)」,『역사문제연구』21, 역사문제연구소, 2009.
- 박기순,「랑시에르와 민중 개념: 민중에 대한 낭만주의적 해석과 그 대안의 모색」,『진보 평론』59, 2014.
- 박여성,「개념사 연구: 역사서술과 언어학의 상호작용 – 구조의미론, 원형의미론 및 텍스트언어학을 중심으로–」,『개념과 소통』9, 한림대학교 한림과학원, 2012.
- 이세영,「‘민중’ 개념의 계보학」, 김경일 등,『우리 안의 보편성: 학문 주체화의 새로운 모색』, 한울, 2006.

- 이신철, 「인민’의 창조와 사라진 ‘민중’ – 방법으로서 북조선 민중사 모색」, 『역사문제연구』 23, 역사문제연구소, 2010.
- 이재윤, 「지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구」, 『한국문현정보학회지』 40-2, 한국문현정보학회, 2006.
- _____, 「계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구」, 『한국문현정보학회지』 40-3, 한국문현정보학회, 2006.
- _____, 「가중 네트워크를 위한 일반화된 지역중심성 지수」, 한국정보관리학회, 『정보관리학회지』 32-2, 2015.
- 장상철, 「1970년대 ‘민중’ 개념의 재등장 – 사회과학계와 민중문학, 민중신학에서의 논의」, 『경제와사회』 74, 비판사회학회, 2007.
- 최호철, 「구조 의미론의 수용 양상과 국어 어휘 의미론의 과제」, 『한국어 의미학』 2, 한국어의미학회, 1998.
- 허 수, 「개벽」 논조의 사회주의화에 관한 새로운 접근 – 토픽 연결망 분석을 중심으로, 『인문논총』 78-1, 서울대학교 인문학연구원, 2021.
- _____, 「1920~30년대 식민지 지식인의 ‘대중’ 인식」, 『역사와현실』 77, 한국역사연구회, 2010.
- _____, 「식민지기 ‘집합적 주체’에 관한 개념사적 접근 – 『동아일보』 기사 제목 분석을 중심으로」, 『역사문제연구』 23, 역사문제연구소, 2010.
- 허수·김혜진·정유경, 「대한제국기 ‘집단적 주체’의 의미망 – 『황성신문』과 『대한매일신보』의 사설 기사를 중심으로」, 『대동문화연구』 119, 성균관대학교 대동문화연구원, 2022.
- 황병주, 「1960년대 비판적 지식인 사회의 민중인식」, 『기억과 전망』 21, 민주화운동기념사업회, 2009.

❖ 기타

- 「[철학사40] 소쉬르(Saussure)」, <http://www.epicurus.kr/Humanitas/386926>, 2023. 1. 29. 검색
- 이상구, 「민중당(民衆黨)」, 『한국민족문화대백과사전』, <http://encykorea.aks.ac.kr/Contents/Item/E0020310>