



메타데이터를 활용한 1960~2018

〈한국언론학보〉 논문 분석

다이내믹 토픽 모델링(Dynamic Topic Modeling) 방법을 중심으로*

최선영 이화여자대학교 에코크리에이티브협동과정 특임교수**

고은지 이화여자대학교 공학박사***

이 연구는 〈한국언론학보〉 창간호부터 2018년까지 논문 전집 메타데이터를 활용하여 두 가지 접근으로 학보의 60년을 분석하였다. 우선 기술통계 분석 결과 〈한국언론학보〉의 총 논문 수는 2,048편, 학보논문 저자는 1,276명으로 나타났다. 2001년 이후 누적 논문 수는 2000년 이전과 비교할 때 양적으로 약 3배 증가하였으나 최근 4년간 논문 수가 급감해 학보의 역할에 대한 논의가 필요함을 제안하였다. 학보 논문의 주요 토픽이 60년간 어떠한 진화 패턴으로 나타나는지 알아보기 위해 시계열 데이터 연산 알고리즘인 다이내믹 토픽 모델링(DTM : Dynamic Topic Modeling)으로 분석한 결과 첫째, 연구영역에서는 저널리즘과 온라인 저널리즘, 보도관련 저널리즘이 발생확률이 높은 토픽으로 추출된 반면, 문화연구는 발생확률이 낮게 나타났다. 둘째, 방법론에서는 경험주의 연구와 실증연구가 지배적 토픽으로 나타났고 컴퓨터이슈널 방법론도 증가추세로 나타났다. 셋째, 이론에서는 프레임, 이용과 충족, 의제설정이 주요 토픽으로 추출되었으며 넷째, 메시지연구는 효과 관련 토픽이, 다섯째, 수용자 연구에서는 정치참여, 텔레비전 시청, 청소년 관련 토픽이 지배적으로 나타났다. 마지막으로 연구대상에서는 온라인, 소셜미디어, 인터넷 토론 관련 토픽이 지배적으로 나타나 온라인 관련 연구가 활성화되었음을 알 수 있었다. 이 연구는 내용분석 분류 방법과 달리 메타데이터를 활용해 알고리즘과 기계학습으로 토픽을 추출하고 이를 군집화(clustering)하여 학보 논문의 연구경향 분석방법을 새롭게 제안해 논의했다는데 의의가 있다.

핵심어: 한국언론학보, 다이내믹 토픽 모델링, 확률적 토픽 모델, 메타데이터, 시계열데이터

* 연구에 아낌없는 조언을 주신 최선열, 홍주현 선생님과 익명의 심사위원들께 깊이 감사드립니다. 이 논문은 한국언론학회 창립60주년기념학술대회 발표 내용을 근거로 작성하였고 한국언론학회, 누리미디어의 도움을 받아 데이터를 수집함. 이 연구는 2017년도 한국방송학회 특별연구비지원에 의해 수행되었음.

** choisy@ewha.ac.kr

*** koeunji@gmail.com

1. 들어가며

과거에 대한 해석은 어떤 방법으로 이루어져야 할까? 역사의 기능은 “과거와 현재 사이의 상호 관계를 통해 양자에 대한 깊은 이해를 복돋아 주는데 있다”(E. H. Carr, 1961/2015). 그래서 현재의 관점에서 과거를 해석하고 이를 교훈 삼는 일은 진보의 기초가 된다. 지난 2009년 한국언론학회가 ‘50년사 편찬위원회’를 구성해 1천여 페이지가 넘는 <한국언론학회 50년사> 집필 작업을 수행한 일도 이러한 까닭에서였을 것이다. 이 저작은 학회의 성장 과정 뿐 아니라 우리나라 언론학의 여정을 총망라한 이정표라고 할 수 있는데, 연구 영역을 크게 언론사, 저널리즘, 방송학, 광고, 홍보, 출판, 뉴미디어, 휴먼커뮤니케이션, 국제커뮤니케이션, 정치커뮤니케이션, 미디어 법제, 커뮤니케이션 정책, 비판커뮤니케이션, 문화연구, 젠더와 미디어연구, 미디어교육 등 16개로 분류해 우리 언론학 역사를 되짚어 보고 있다.

이렇게 언론학 연구영역이 여타 학문분야보다 세분화된 이유는 언론학 지형의 특수성 때문이기도 하다. 최선열(2009)에 의하면 “기술과 사회의 변화에 따라 학문의 연구대상이 크게 변화하는 언론학은 다른 사회과학 분야보다 더 심각한 도전을 받았다”. 연구대상으로서 미디어와 커뮤니케이션 현상은 정치, 사회, 경제, 기술 등의 요인과 함께 상당 부분 연동할 수밖에 없고, 이에 따라 언론학 하위 학문 영역은 불가피하게 다양해질 수밖에 없었을 것이다. 언론학 각 영역을 사회 발전의 역동성과 성장, 변화, 역사라는 큰 틀에 놓고 각 연구영역을 문헌연구와 내용분석 방법으로 살펴보는 일은 분명한 학술적 가치를 지닌다. 다만 이러한 접근은 영역의 다원주의가 두드러져 통합적인 관점에서 언론학이 다루는 주제의 큰 흐름과 경향을 통시적으로 읽기 어렵다. 60년 동안 언론학 전반에 걸쳐 어떠한 이론과 연구방법, 연구대상, 연구영역이 진화했는지 일목요연하게 설명하기 쉽지 않다는 뜻이다.

그래서 이 연구는 그동안의 접근과 다른 방법으로 언론학의 과거를 돌아보고자 한다. 한국언론학회의 대표 학술지인 <한국언론학보> 창간호부터 2018년까지 60년 치 논문 전편의 메타데이터를 활용해 분석하는 것이다. 구체적으로 시계열 데이터 알고리즘 연산이 가능한 다이내믹 토픽 모델링(DTM : Dynamic Topic Modeling) 방법으로 토픽(topic)¹⁾을 추출하여 60년간 학보 논문 의미구조에 잠재되어(latent) 진화한(evolve) 토픽이 무엇인지 해석하고자 한다. 이 연구방법은 시계열 정보가 있는 방대한 문서 집합(population)을 분석하는데 유용하다. 이러한

1) ‘주제’라고 쓰는 것이 맞지만, 이 연구가 다루는 데이터의 특성상 ‘연구주제’라는 의미와 혼동할 수 있어서 부득이 ‘토픽’으로 명기 한다.

연구를 착안한 계기는 첫째, <한국언론학보>에 게재된 논문은 창간호부터 2018년 말까지 2천편이 넘기 때문에 모든 논문을 일일이 다 확인하여 학문 영역을 분류하기 쉽지 않다. 둘째, 언론학 연구영역을 분류하기 위한 기준이 모호하다. 예컨대 트위터와 같은 소셜 미디어 연구는 뉴미디어 영역이라고 할 수도 있고, 연구대상과 주제에 따라 정치커뮤니케이션, 홍보, 방송 영역으로 분류할 수 있다. 따라서 분류 기준이 연구자마다 다를 경우 내용분석 데이터 자체가 달라지고 이에 따라 연구결과가 상이해질 수 있다. 셋째, 전통적으로 학회는 30주년, 50주년을 기해 문헌연구와 내용분석으로 영역을 나누어 언론학 연구 경향을 조명해왔다. 최근 공학적으로 검증된 분석틀을 활용해 데이터를 연구하는 사례가 늘고 있는 만큼 새로운 방법론으로 언론학을 되돌아보고 언론학 분류 기준과 관점을 새롭게 제안하는 연구로서 의미 있는 시도일 것이다. 이러한 접근은 블라이(Blei, 2006)가 1880년부터 2002년까지 <사이언스(Science)>의 1만 7천여 건 논문을 시계열 단위로 분석한 연구에서 아이디어를 참조한 것이기도 하다.

이러한 접근은 자칫 <한국언론학보>의 60년 연구 경향을 단순화하는 인상을 줄 수 있지만, 이 같은 시도는 연구에 대한 연구로서 학회의 대표적인 학술활동인 학술지 논문을 종합적으로 검토하고 객관적으로 요약하여 학회 구성원과 언론학에 관심 있는 사람들에게 간명하게 설명할 수 있는 장점도 있다. 아울러 이 연구는 한국언론학회 60주년 저술 편찬위원회에서 기초연구로 활용하기 위해 출발한 연구임을 밝힌다. 또한 기존의 언론학 연구의 메타분석과 다르게 메타데이터라는 디지털 발자국(digital footprint) 형적(形跡)을 활용한 실험적 시도라서 그 결과와 의미를 일반화하여 해석하기에는 제한적일 수 있다.

2. 우리나라 언론학 연구

<한국언론학회 50년사>의 ‘한국 언론학 연구 50년’을 통해 언론학 연구를 집대성한 연구자들(양승목, 2009; 임영호, 2009; 정진석, 2009; 최선열, 2009)은 <한국언론학보>가 선도적 탐구의 장으로 그 역할을 지속해온 점에 높이 평가하였다. 이들은 힘겨운 태동기를 거쳐 2000년대 이후 논문의 양적 증가 추세가 명백하게 나타난 점을 고무적으로 설명하였다. 언론의 역할과 커뮤니케이션 기술의 잠재력, 언론의 사회적 영향력이라는 학술공동체의 공통된 관심사를 다루면서도 매우 다양한 영역에서 연구가 역동적으로 이루어지고 있다는 것이다.

역사적으로 학보에서 연구 영역의 분화가 발생한 시점은 1969년 12월 발행된 <신문학보> 2호에서 찾을 수 있다. <신문학보> 1호 이후 9년 만에 복간된 2호의 권두에서 김규환(1969)은

“미국 마스크 연구, 정치학과의 연관, 사회적 영향력, 심리적 요인, 경제학적 고찰” 등으로 신문학 연구 영역이 확장해야 한다고 제안하고 있다. 이후 매스미디어 제도, 여론-선전, 관료제도와 매스미디어 소유 형태, 매스미디어의 사회적 기능, 사회체계와 커뮤니케이션 과정의 상관관계, 태도 및 행동의 변용과 매스미디어 내용, 매스미디어 기업 및 제도의 분석, 경제발전과 기대, 광고 등으로 연구영역을 분류해 나간 것이다(정진석, 2009). <신문학보>는 이러한 제안을 다각도로 수용하여 매스미디어의 사회적 영향력에 대한 연구로 청소년 일탈을 주제로 특집(8호)을 발간하기도 했고, 발전커뮤니케이션 논문 23편을 게재(10호)하여 여러 영역에서 검토한 바 있다.

1970년대 이후 <신문학보>의 연구 영역이 매스커뮤니케이션학으로 확장된 데에는 관련학과의 설립 증가와 미국 유학과 연구자의 국내 정착을 들 수 있을 것이다. 70년대 말 언론관련 학과는 전국적으로 14개에 불과했다. 1985년 말 기준 24개 학과로 증가했고(차배근, 2009), 1999년에는 85개 학과(조동시, 1999), 2014년에는 114개 학과(언론연감, 2014)로 늘어났다. 논문 생산에 관련학과의 양적 증가가 기여했을 것이라는 설명이 가능하다. 특히 정치 사회적 변화, 매체 기술의 변화, 미디어 환경의 변화, 수용자의 진화 등에 따라 언론학 영역의 변화 폭은 커졌을 것이다. 이 과정에서 지속적으로 성장하고 진화해온 연구 영역이 있는가 하면, 사라진 영역도 있을 것이다.

그러나 언론학의 양적 성장에 대해 우호적인 논의만 이루어진 것은 아니다. 50년을 회고한 연구자들은 질적 성장에 대한 반성도 필요함을 강조하였는데, 이들은 연구의 정체성 위기, 학문영역 간 이질성과 분산, 이론 생산 빈곤함과 반복성을 공통된 문제로 지적하고 있다(양승목, 2009; 임영호, 2009; 최선열, 2009). 우리나라 언론학 성장사를 10년 단위로 분석한 양승목(2009)은 실증연구 중심의 1970년대, 대안적 비판 언론학 등장과 패러다임 갈등이 있었던 1980년대, 이론적 다원주의 시대였던 1990년대, 양적 성장과 경험적 연구가 우세했던 2000년대 등으로 요약하면서 “논문 수와 전문학술지 발행, 연구자 수의 증가 등 양적 성장과 팽창은 뚜렷한 반면 독창적 개념과 이론 생산의 부족함, 학문적 정체성 약화가 문제”라고 논하였다(양승목, 2009). 특히 스완슨(Swanson, 1993)의 지적을 빌어 “학제적 하위영역들의 증식은 이미 문제가 많은 커뮤니케이션 연구의 학문적 정체성을 훼손하는 원심력”(양승목, 2009 재인용 pp.1058)이 되어 소통 없는 학문적 파편화로 피상적인(shallow) 다양화를 초래할 수 있음을 비판하였다. 독립학문으로서의 정체성을 모색하고 합의하기 전, 연구 영역으로 분화해온 학문적 현상에 성찰을 촉구한 것이다.

<신문학보> 1호의 영문명(The Korean Society of Journalism)으로 짐작할 수 있듯이 언론학의 출발은 ‘저널리즘’이었다. 저널리즘은 언론학보의 주요한 연구영역으로 이 분야 연구경

향의 흐름을 살펴보는 일은 학보의 궤적을 찾는 기준이 될 수 있을 것이다. 임영호(2009)는 저널리즘 연구의 큰 흐름에 대해 1960년대 말 이후 도입된 경험주의 커뮤니케이션 등장과 비판이론의 등장, 이론적 다원화와 학문으로서의 체계 모색, 2000년대 이후 미디어 환경 변화에 의한 저널리즘 해체와 재구성 등으로 설명하였다. 나아가 “학술적 연구와 비학술적 지식의 경계” 사이에서 연구의 갈피를 잡기 위한 학문적 노력의 필요성을 환기했다. 왜냐하면 학문적 정체성을 토대로 지식 분류방식을 체계화하지 못해 매체 단위로 분류하는 “매체 중심성” 경향이 지속되면서(임영호, 1998a, 2009, 재인용) 연구주제가 이론적 중요성의 기준을 적용하기보다 대중적인 주제에 편중되는 경향을 보였다는 것이다. 즉 경험주의적 연구가 언론학의 주류 문화로 자리 잡는 과정에서 메시지 내용분석, 수용자연구, 경험적 언론 생산자 연구 등이 기본 개념의 체계화나 이론적 문제의식이 결여된 채 정착되었다는 지적이다. 이러한 경향은 〈한국언론학보〉 전반에 걸쳐 포착될 수 있기 때문에 본 연구에서도 확인과 검토가 필요하다. 저널리즘 연구 영역에서 다루는 연구대상, 메시지, 매체, 수용자, 송신자 등은 다른 연구 영역으로 분류될 가능성이 있어서, 영역 간 중복이 발생한 주요 단어들이 토픽을 구성할 경우 저널리즘 연구로만 한정하기는 어려울 것이다.

정치커뮤니케이션 연구도 언론학의 과거와 현재를 반추하고 학문적 발견을 할 수 있는 주류 연구 영역이다. 〈한국언론학보〉의 50년 치 전집 논문 내용분석을 통해 정치커뮤니케이션 연구 분야를 시계열적으로 분석하고 주제별로 문헌분석을 한 최선열(2009)의 연구는 언론학계에서 재조명이 마땅하다고 할 만큼 이론적 논의가 풍부하고 분석의 엄밀성을 추구한 통시적 연구이다. 이 연구는 계량분석을 통해 수용자 연구와 메시지 연구, 선거 관련 연구와 수용자 효과연구, 매체연구, 주요 이론과 개념, 연구방법과 양적방법론의 경향성과의 관계에 대해 상세히 조명하였고, 문헌분석을 통해 크게 비선거 연구와 선거연구로 분류해 각 세부주제를 다루며 통찰력 있는 논의를 펼치고 있다. 특히 이 연구는 3중 지층구조 모델(정치체제, 미디어, 학회의 제도화 등 3개 지층)로 연도별 〈한국언론학보〉의 총 게재 논문수와 정치 커뮤니케이션 논문수를 비교한 분석을 하였는데, 이러한 분석 틀과 관점은 언론학의 역동적 진화를 보고자 했던 우리의 지적 호기심에 방아쇠를 당겼다. 60년 역사를 일별할 수 있는 선행연구 자료나 분석틀을 갖추지 못한 상태에서 최선열(2009)의 연구는 내용이나 방법론적으로 〈한국언론학보〉의 연구경향을 상세히 조망할 수 있는 기준점이라고 할 수 있다. 특히 반복적인 과정(iterative process)을 되풀이 할 수 밖에 없는 토픽 모델링 평가와 검증에 있어서 귀한 메타분석 연구이다.

논의를 종합해 보면 〈한국언론학보〉는 태생적으로 많은 연구영역을 포괄하는 학술지로 출발해 현재에 이르고 있다. 올해 초 한국언론학회 60주년 저술 편찬위원회는 기초 연구로, 최근 20년 간 〈한국언론학보〉 논문의 주요 연구 분야를 분류하여 정리한 바 있다. 그 결과 저널리즘,

정치커뮤니케이션, 수용자 연구, 미디어 경영·법제, 윤리·정책, 뉴미디어 정보기술, PR, 문화·젠더연구, 커뮤니케이션 미디어 역사, 소셜미디어·인터넷·CMC, 광고 마케팅, 휴먼커뮤니케이션, 방송, 헬스 커뮤니케이션 등의 순으로 논문 생산이 이루어진 것으로 나타났다(홍주현, 2019). 따라서 많은 논문을 생산한 분야이자 중심적인 위치를 차지하고 있는 저널리즘과 정치커뮤니케이션 분야의 기존 메타분석 결과를 참고하면 토픽모델링 알고리즘으로 자동 계산된 토픽 해석에 참고가 될 것이다.

이 밖에 각 연구 영역에 대한 토픽 뿐 아니라 수용자 연구나 메시지 효과연구, 이론, 방법론, 연구대상을 포함해 방송학, PR, 소셜미디어, 미디어 법제 및 정책, 문화연구 등 연구 성과가 축적된 분야도 토픽으로 추출될 수 있다. 언론학 특성상 토픽 모델링 계산 결과 발산적인 토픽 분포가 나올 가능성을 배제할 수 없고, 추출된 토픽과 토픽을 구성하는 단어들 또한 파편적으로 추출될 가능성이 있다. 무엇보다 하나의 토픽에 맥락을 파악하기 힘든 단어들 이 구성되는 경우 이를 어떻게 개념화하는가의 문제가 발생할 수 있다. 최선열(2009)은 언론학을 영역별로 분류해 내용분석을 할 경우 '중복' 문제가 있음을 지적한 바 있는데, 중복은 어떤 영역으로 묶이는 것에 대한 기준이 모호할 때 발생한다. 하나의 토픽에 서로 다른 영역의 단어들 이 묶일 수도 있다는 의미이기도 해서, 그동안 영역별로 내용분석이 수행되어온 선행연구와는 다른 결과가 나올 가능성이 있다.

3. 메타분석과 확률적 토픽 모델(Probabilistic Topic Models)

학술적인 지식을 계량적 연구방법을 통해 종합적으로 평가하고 분석하는 메타 분석은 “연구에 대한 연구, 분석에 대한 분석(analysis of analysis)”(Glass, 1976)이다. 연구 결과 자료를 통합하여 해당 영역의 연구 주제에 방향성과 대안적 논의를 끌어내는 것에 목적이 있다(이수범·문원기, 2016; 탁진영, 2004).

메타분석의 장점에 대해 오성삼(2002)은 객관적 자료를 활용하기 때문에 편향된 연구 표본이 아니라서 연구자가 연구 결과를 위해 자의적으로 연구의도에 부합하지 않는 자료를 제거하는 일을 방지할 수 있다고 설명한다. 무엇보다 단일한 연구주제에 상반된 결과나 논의가 이루어진 경우 새로운 관점과 접근으로 새로운 논의를 이끌어낼 수 있다(오성삼, 2002). 메타분석의 궁극적인 목표는 분석 결과에서 규칙성을 발견하여 기존 연구 성과와 결과에 대해 진단하고 성찰하는데 있다(Wallace, 1992).

이 연구에서 활용하고자 하는 확률적 토픽 모델(Probabilistic Topic Models)은 학술 분야 자료 뿐 아니라 데이터 형식으로 된 문서 집합에서 활용이 가능한 분석방법이다. 오늘날 거의 모든 정보와 지식이 데이터 형태로 쌓이면서 자연어 처리 기술이 고도화되는 추세라 학술문서처럼 정형화된 메타데이터는 댓글, 소셜미디어 메시지, 사용자 피드백 같은 비정형 데이터에 비해 분석이 용이하다. 자동화된 알고리즘을 활용하는 이유는 많은 양의 문서를 사람이 직접 읽고 토픽을 이해하거나 분석하기 어렵기 때문이다. 따라서 이 방법은 방대한 문서 내에서 개념적으로 묶이는 토픽(topic)을 알고리즘이 발견한다는 점에서 기존 메타 분석 방법과는 차이가 있다. 기존의 메타분석 방법은 1차적으로 분석 유목이나 분류를 연구자인 사람이 처리하는 과정을 거치는데, 이럴 경우 분석 유목 기준은 여전히 객관적이지 않을 수 있다. 토픽 모델링은 문서 내에 잠재적으로 존재하는 의미 구조의 맥락을 비지도 학습(unsupervised learning)²⁾ 알고리즘을 통해 자동으로 계산하는 텍스트 마이닝 방법이기 때문에(Blei, 2012; Jordan & Mitchell, 2015) 확률적으로 의미 있는 토픽과 토픽을 구성하는 단어를 사람인 연구자가 결정하지 않는다. 또한 비지도 학습은 자율 기계학습을 통해 추출한 토픽을 군집(clustering)으로 묶어 도출할 수 있어서 데이터 특징을 추출(feature extract)하는데 유용하다.

토픽 모델링 중에서 가장 많이 알려진 잠재적 디리클레 할당(LDA: Latent Dirichlet Allocation)은 각 문서(Document)를 발생 확률의 비율로 구성된 토픽의 집합으로 가정한다. 이 때 각 토픽은 여러 단어(Keyword)들로 구성되며 이 단어들 또한 발생확률의 비율로 나타난다. 연구자가 토픽 모델링 알고리즘에 토픽의 개수를 입력하여 문서 내에서 토픽의 분포 및 단어들의 분포를 계산한 결과를 도출할 수 있는데, 이때 토픽의 질(Quality)을 평가하는 것이 필요하다. 토픽의 발생 확률 비율도 중요하지만, 토픽을 구성하는 단어들의 발생확률 비율도 중요하기 때문이다. 토픽은 지배적인 단어의 조합이기 때문에 이를 일별하는 것만으로도 토픽이 무엇인지 해석이 가능해야 한다.

토픽 모델링을 평가하는 방법은 내재적으로 '혼잡도(Perplexity)'를 줄이는 기법이 흔히 쓰인다(Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009). 확률적 모델이 실제로 관측 값을 잘 예측하는지, 문서 내 토픽 출현 확률, 토픽 내에 용어 출현 확률 등을 계산하는 LDA가 잘 모델링 되었는지 평가할 때 쓰는 방법이다. 토픽 개수를 결정할 때 낮은 혼잡도가 모델 평가 기준이 된다. 그러나 최근 이 기법이 반드시 가장 좋은 모델 평가 기준이 되는 것이 아니라는

2) 기계학습은 크게 비지도 학습, 지도 학습(supervised learning), 강화 학습(reinforcement learning)등으로 구분하는데, 비지도 학습은 입력 값에 대한 목표치가 없어서 사람의 해석을 통해 의미나 지식을 추출하는데 활용된다.

연구 결과가 보고되고 있다(O'callaghan, Greene, Carthy, & Cunningham, 2015). 혼잡도가 낮아도 모델이 적절하지 않다고 판단될 때는 토픽 일관성(topic coherence)을 계산하는 방법이 있다. 이는 뉴먼(Newman, 2010)이 제시한 일관성 점수(coherence score)이다. 토픽 일관성은 토픽 모델링이 얼마나 해석하기 좋게 계산되었는지를 보는 방법으로, 결과로 나온 각 토픽에서 상위의 N개 단어가 도출되었을 때 그 토픽 안에 유사한 의미가 있는 단어가 구성되었다고 추론하는 것이다. 그리고 상위 단어끼리의 유사도를 계산하면 방대한 문서에서 나타난 토픽의 맥락적 의미를 파악할 수 있다.

그러나 LDA와 같은 토픽 모델링은 단순히 지배적인 토픽 그 자체만을 추출하는 방법이기 때문에 정적인(static) 결과만을 확인할 수 있다. 시간적으로 누적되어온 데이터의 특성을 살려 그 패턴을 분석하는데 한계가 있는 것이다. 우리는 <한국언론학보>의 논문 발간일이라는 시간흔적(time stamp)을 활용하여 토픽의 통시적 진화(evolve) 패턴을 분석하는 것이 반드시 필요하다고 보았다. 왜냐하면 서두에서 밝혔듯 언론학 지형의 변화무쌍한 속성 때문이기도 하고, 시간의 흐름을 통해 설명해야 비로소 이 연구의 목적이 살아나기 때문이다.

시간이 기록된 데이터라면 연속적인 시간 역동성을 가진 토픽 모델(cDTM : Continuous Time Dynamic Topic Models)로 분석이 가능하다(Wang, Blei & Heckerman, 2012). 이러한 속성을 가진 문서 데이터는 <그림 1>과 같이 시퀀셜 그룹(sequential group)으로 나눌 수 있어서 시간에 따른 토픽의 변화 추이를 살펴볼 수 있다. 블라이와 레퍼티(Blei, & Lafferty, 2006)는 수집된 문서가 시계열 속성이 있을 경우 문서에서 “시간에 따른 토픽의 진화추이(the time evolution of topic) 확률을 포착할 수 있다”고 설명한다(최선영·고은지, 2017. p.88 재인용). 따라서 블라이(Blei, 2003)의 연구를 참조하여 LDA를 구한 후, 시간에 따라 변화할 것으로 예측되는 토픽 확률을 분석하기 위해 다이내믹 토픽 모델(DTM : Dynamic Topic Model)을 계산할 것이다.

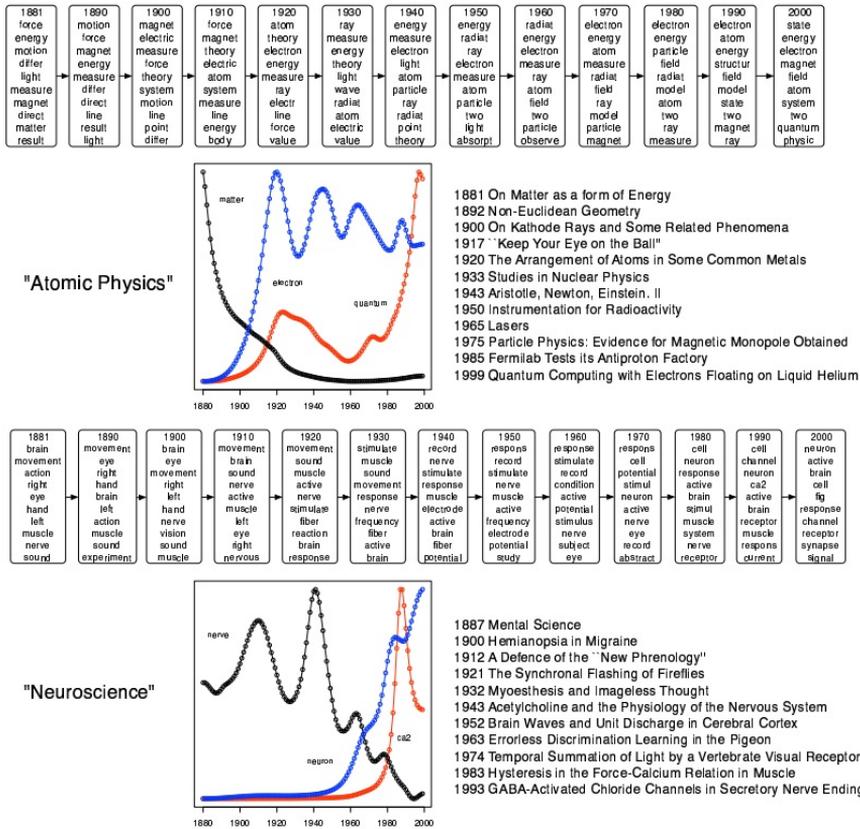


그림 1. 1880~2000년 <사이언스(Science)>지를 10년 단위로 다이내믹 토픽 모델링 계산한 결과. 토픽 확률 분포 및 토픽의 구성 단어(keyword)가 시간에 따라 변화함을 볼 수 있다(Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning, ACM, 113-120에서 인용. 원 저작권자의 모든 권리가 보호됨.).

여타 언론학 영역의 연구 경향에 대한 상술보다 방법론 설명에 지면을 할애한 까닭은 본 연구자들이 각 연구 영역에 대한 사전 지식을 쌓다 보면 토픽 모델 평가 및 토픽 선정에 편향이 발생할 수 있기 때문에 대표적인 언론학 연구를 참고하여 우리의 연구가 타당성 있는 결과로 나오는지 비교하는 것이 더 중요하다고 보았다. 토픽 모델링의 방법론적 의미는 주어진 데이터를 활용해 모델의 정확성(model precision)을 높일 수 있는 데이터 처리에 달려있다. 이 연구에서 다루는 메타데이터의 특성을 파악해 계산 기법을 선정하는 것은 선행연구 검토만큼 중요하다. 토픽 모델링을 통해 언론학 연구를 통시적으로 잘 설명할 수 있는, 귀납적 해석이 가능한 결과를 도출하는데 이 연구의 의의가 있기 때문이다.

4. 연구방법 및 연구의 틀

이 연구는 크게 두 개의 접근으로 연구하였다. 우선 자동화된 알고리즘으로 토픽 모델링을 계산하기 전, 메타데이터에서 확인할 수 있는 <한국언론학보>의 논문 수와 연구자별 논문 수, 인용 및 인용이 높은 논문 등에 대한 기술 통계를 제시하는 것이다.³⁾ 우리나라는 한국연구재단 데이터 기준의 인용 지표(citation index)인 H지수가 있다. 그러나 60년 동안의 논문 전집을 독점 공급하고 있는 누리미디어의 데이터베이스를 참고해야 실제 이용 추이를 확인할 수 있어서 삼각 측정법(triangulation)으로 두 데이터를 참고하였다.

또 다른 접근은 확률적 통계 모델인 토픽 모델링(Topic Modeling)을 활용하여 <한국언론학보> 논문 전집의 토픽을 추출하고 분석하되 시간 추이에 따른 토픽의 변화를 확률적으로 살펴보는 것이다. 이 방법은 영역 또는 유형, 범주를 분류(classification)하는 메타분석 방법과 달리 확률적으로 계산된 토픽의 군집(clustering)을 통해 결과를 볼 수 있다. 순차적으로(sequential) 축적된 문서는 토픽 확률 변화를 계산할 수 있는 데이터라, 시간에 따른 토픽의 변화 양상과 군집을 해석(interpretation)할 수 있다.

본 논문의 전체적인 연구의 틀은 다음과 같다.

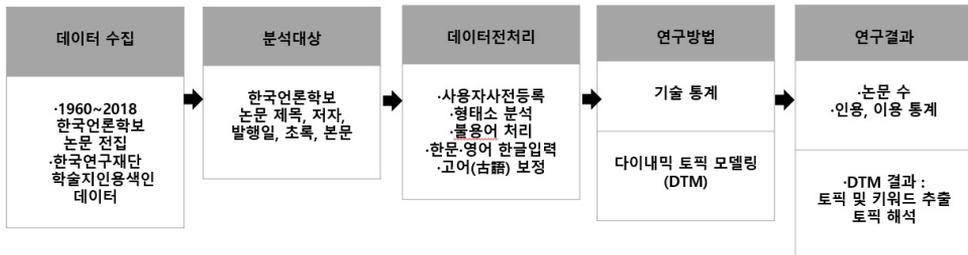


그림 2. 연구의 틀

3) <한국 언론학 연구 50년>에서 이미 통계자료를 다각도로 제시하였고, 홍주현(2019)이 '한국언론학회 창립 60주년 기념 학술대회'에서 2000년대 이후 <한국언론학보>에 대한 연구경향을 발표논문으로 정리하였기에 내용 중복을 피하고자 이 연구에서는 학보 논문 전집 메타데이터를 통해 확인할 수 있는 저자별, 개별 논문, 관련 학술지 등에 대한 기술 통계를 제시한다.

1) 분석대상 선정 및 데이터 처리

본 연구는 1959~2018년까지 <한국언론학보>에 게재된 60년 치 논문 전편의 메타 데이터를 수집하여 전처리 과정을 거쳤다. 연구 대상 범위는 논문 2,048편으로 누리미디어 데이터베이스에서 제목, 저자, 발행일, 발행쪽수, 키워드, 초록, 본문이 포함된 메타데이터로 수집하였다. 메타데이터 상에 초록 정보가 누락된 512편에 대해서 연구자와 대학원생 연구보조원 3명이 논문을 읽고 보완하는 작업을 하였고, 영어와 한문 표기는 전처리 과정에서 전부 한글로 바꾸는 작업을 실시하였다. 시퀀셜 분석을 위해 날짜 보정 작업도 거쳤다. 1960년 4월에 발간한 <신문학보> 1호에 실린 논문이 최초의 연구이지만, 2호는 9년 뒤인 1969년에 나타나 시간 왜곡이 발생할 수 있는 점을 고려하여 1호 논문은 4편은 분석대상에서 제외하였고, 또 권두언 및 서평 등 논문으로 간주하기 어려운 형태의 문서는 집합에서 제외하였다(DTM 논문 수 = 2,044편). 그리고 다이내믹 토픽 모델링(DTM)을 실시하기 위한 과거 발행일이 정확치 않다고 판단한 경우 사료에 근거해 연구자가 날짜보정 작업을 하였다.

수집된 데이터는 파이썬(python)3를 활용하여 분석하였다. 한글 형태소 분석을 위해 파이썬 패키지 코엔엘파이(KoNLPy)⁴⁾ 라이브러리를 사용하였으며, 형태소 분석기는 이 라이브러리에서 제공하는 미캡(MeCab)⁵⁾을 사용하였다. 형태소는 의미 최소 단위인 토큰(token)으로 분리하여 글자, 단어, 어절 등으로 나누는 작업을 수행했다. 논문 데이터는 정제된 언어이지만, 1960~90년대 논문에서 현대어 표현이 아닌 경우가 추출되어 형태소 분석기 사전 등록 과정에서 용어를 통일해 정리하였다. 토픽 모델링 결과가 잘 도출되려면 사전 등록 작업이 정교하게 수행되는 것이 필요하기 때문이다. 예를 들어 ‘뿌브리찌스틱학’, ‘프블리찌스틱과학’은 ‘공시학’으로, ‘컴뮤니케이션’, ‘콰뮤니케이션’은 ‘커뮤니케이션’으로, ‘인터디시프리너리’는 ‘학제간’, ‘아프라치’는 ‘접근’, ‘메스콰’은 ‘메스컴’, ‘메스미디어’는 ‘메스미디어’, ‘텔레비전’은 ‘텔레비전’, ‘센세이셔널리즘’은 ‘선정주의’, ‘이용과 만족’은 ‘이용과 충족’ 등으로 각각 통일시켜 주었다. 또 분리된 어절로 추출되는 단어는 반복적으로 모델을 테스트하는 과정에서 사전 등록을 업데이트하여 검출 누락이 없도록 하였다.

4) <http://konlpy.org/ko/latest/>

5) 코엔엘파이에서는 ‘Hannanum 시스템 사전’, ‘Kkma 시스템 사전’, ‘Mecab 시스템 사전’등을 지원하는데, 이 연구에서는 세종 말뭉치로 만들어진 CSV 형태의 사전인 ‘미캡(MeCab)’을 활용하였다.

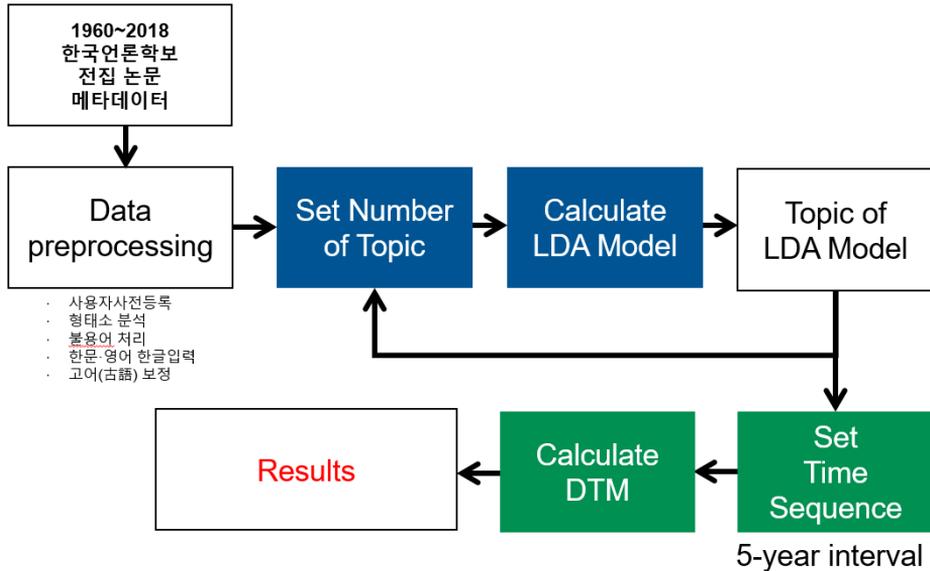


그림 3. DTM 절차

2) 분석 방법 : 다이내믹 토픽 모델링(DTM : Dynamic Topic Modeling)

우리는 적합한 모델을 구하기 위해 반복적으로 모델링 테스트 과정을 거쳤다(그림 3)). 연구 초기 기본 LDA를 사용하여 모델링 테스트를 위해 토픽 개수를 40~120개까지 설정한 후 혼잡도를 계산하여 모델을 평가하였으나, 논문의 특정한 단어만 추출되는 현상이 지속적으로 발생하여 이 LDA 모델은 적합하지 않다고 판단하고 폐기하였다. 우리는 모델링 평가에 있어서 혼잡도 대신 토픽 일관성으로 평가하는 방법을 선택하였다. 이 평가방법을 채택하기 위해 모델링은 gensim의 말렛⁶⁾ 래퍼(Mallet wrapper)를 사용하여 LDA를 계산하였다. LDA는 문서 집합 안에 단어의 분포를 통해 나타나는 일련의 토픽들이 있다고 가정하기 때문에(Blei, 2012; Blei, Ng, & Jordan, 2003), 토픽모델링을 수행하기 위해 논문 초록을 하나의 문서로 설정하였다. 그 결과 총 2,044개의 문서에서 134,524개의 'non-zero entries'가 생성되었으며 이 중 중복되지 않는 형태소인 'features'는 10,215개였다.

6) 말렛(Mallet : Machine Learning for Language Toolkit)은 오픈 패키지 토픽 모델링 라이브러리로 통계적 자연어처리, 문서 분류, 클러스터링, 토픽 모델링, 정보추출 및 텍스트 기계 학습을 처리하는 응용 프로그램이다. 텍스트 인코딩에 UTF-8을 적용하고 있어 한국어 처리가 용이하다.

말렛 LDA를 구하면서 시간 간격을 5년, 10년 단위로 반복 계산하고 토픽 개수를 10~120개까지 설정하여 각각의 DTM을 계산한 후 뉴먼(Newman, 2010)이 제시한 일관성 점수를 활용하여 토픽을 평가한 결과, 100개의 토픽이 적합한 모델로 나타났다. 이 과정에서 우리는 〈한국언론학보〉가 다루는 토픽이 문서집합 크기에 비해 다양하기 때문에 혼잡도보다 일관성 점수가 데이터를 더 잘 반영한다고 판단했다.

분석대상 및 범위에 대한 시간 단위 설정은 연구자 판단에 따르는데, 60년간 〈한국언론학보〉 논문 전집에서 나타난 토픽 변화 추이를 보기위해 5년, 10년 간격을 설정하여 말렛LDA 계산을 반복 테스트 하였다. 그 결과 2000년대 전 논문 수가 적기 때문에 블라이(Blei, 2006)의 〈사이언스〉지 연구에서 설정한 10년 간격은 언론학보 데이터 특성상 분석단위가 맞지 않았고, 2~3년 간격으로 좁힐 경우 특정 시계열 구간의 논문 수가 너무 적다는 문제가 발생했다. 이에 5년 단위 설정이 토픽 산출에 적합하다고 보고 모델링 계산과 평가 방법을 거쳤다.

5. 〈한국언론학보〉 논문 계량분석

1) 전체 논문 수 및 연도별 발간 추이

1960년부터 2018년까지 〈한국언론학보〉는 총 2,048편의 논문이 게재되었고, 연도별 게재편수는 〈그림 4〉와 같이 나타났는데, 논문 수는 1992년부터 서서히 증가하기 시작해서 2001년 이후 괄목할 양적 성장을 이룬다.

〈한국언론학보〉의 정치커뮤니케이션 50년사를 3단 지층구조로 분석한 최선열(2009)의 연구는 논문 수 증가에 대한 분석 근거를 명확하게 제시하였는데, 이중 학회 제도화 지층 구조와 미디어 지층 구조를 본 연구의 결과 해석 틀로 적용하는데 큰 무리가 없다고 보았다. 우선 학회 제도화 차원에서 살펴보면 1989년 한국언론학회 창립 30주년을 기점으로 연도별 게재 논문이 10편 아래로 떨어지는 현상은 발생하지 않는다. 이때 연구 분과 위원회를 구성하였고, 회원 수는 200여 명으로 1982년 104명에 비해 2배로 증가한다. 1996년 회원 수 504명이 된 이후 논문 수는 해마다 10편씩 증가하였다. 가장 특이한 지점은 한 해 동안 102편의 논문이 생산된 2001년도이다. 2000년 35편과 비교할 때 그 수가 약 3배 증가한 괄목할 성과를 이룬다. 2000년까지 연 4회 발간이었으나 2001년에 두 번의 특별호(국문, 영문)를 발행하며 논문 수가 급증한다. 이때 연 6회 발행을 정착시키는 계기를 마련한 것으로 보인다. 2002년은 연 6회 발간이 정착이 되지 않은 상태로 연 4회(59편)가 게재되어 논문 수가 전년도 대비 절반까지 떨어졌다. 그러나

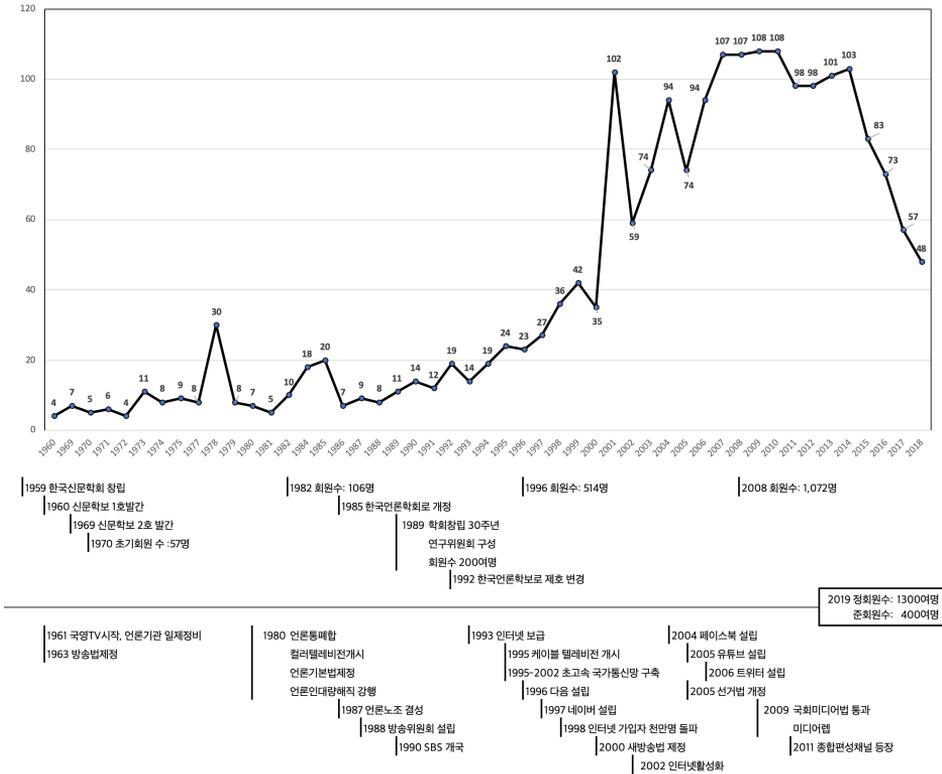


그림 4. 1960~2018년 연도별 <한국언론학보> 전체 논문 수 추이.

주: 최선열(2009)의 학회 제도화 지층 구조, 미디어 지층 구조를 참조하여 재구성.

2003년부터 연 6회 발행이 정착한 결과 2006년부터 2014년까지 논문 수의 증가 추이가 꾸준히 나타났다. 9년간 평균 약 102편의 논문이 해마다 학보에 게재된 것이다. 2000년도 이후 논문 수 증가를 최선열(2009)의 미디어 지층 구조로 설명하면 미디어 제도의 변화(언론법, 방송법 개정), 소셜 미디어의 성장, 인터넷, 통신, 스마트폰, 모바일 등 매체 기술 성장, IPTV, 종합편성 채널의 등장 등 언론학 연구 범위가 다양해질 수밖에 없는 미디어 환경과 테크놀로지의 영향도 크게 작용했음을 알 수 있다.

그러나 2009년까지 증가추세였던 논문 수는 2015년 83편으로 전년도 대비 20편이 줄었고, 2016년 73편, 2017년 57편, 2018년 48편으로 최근 4년 간 급격하게 줄어든다. 2018년의 게재 논문 수는 2014년도 비교할 때 약 절반 수준이다. 2001년 연 6회 발간이 시작된 이래 가장 적은 수의 논문이 게재된 것이다. 최근 논문 생산의 감소 원인은 우선 언론학 영역 세분화에 따

른 관련 학술지의 발행 증가에서 찾을 수 있다. 한국연구재단은 2019년 4월 신문방송학 관련 학술지를 26종으로 분류하고 있는데, 언론학 관련 학술지의 증가는 상대적으로 논문 투고량 감소의 원인이 되었을 가능성이 있다. 그러나 <한국언론학보>는 <표 1>에서 보는 바와 같이 언론학 연구의 대표적 학술지로 각 영역의 학술지 발전에 영향을 주고 있어서 언론학 연구 생태계에 긍정적인 역할을 수행하고 있음을 알 수 있다. <표 1>은 한국연구재단 메타데이터 상에 나타난 2019년 4월 현재 학보의 인용 통계로 <한국언론학보>가 인용되고 있는 상위 20개 학술지를 정리한 것이다. 학보에 실린 논문은 등재 및 등재후보 학술지 논문에서 2019년 4월 기준 총 23,360회 인용되고 있고 이중 14.4%가 <한국언론학보>에서 인용되고 있음을 알 수 있다⁷⁾.

표 1. <한국언론학보> 인용 상위 20개 학술지 목록 및 인용횟수(1999~2018년)

학술지명	발행기관	토픽분류	인용 비율(%)	인용횟수 / 총인용수
한국언론학보	한국언론학회	신문방송학	14.4	3,374/23,360
언론과학연구	한국지역언론학회	신문방송학	5	1,167/23,360
한국방송학보	한국방송학회	신문방송학	4.6	1,083/23,360
한국언론정보학보	한국언론정보학회	신문방송학	3.2	746/23,360
사이버커뮤니케이션학보	사이버커뮤니케이션학회	사회과학일반	2.5	588/23,360
한국콘텐츠학회 논문지	한국콘텐츠학회	학제간연구	2.3	545/23,360
홍보학연구	한국PR학회	신문방송학	2.3	527/23,360
정치커뮤니케이션연구	한국정치커뮤니케이션학회	신문방송학	2.2	520/23,360
광고학연구	한국광고학회	신문방송학	1.9	454/23,360
한국소통학보	한국소통학회	신문방송학	1.7	389/23,360
커뮤니케이션 이론	한국언론학회	신문방송학	1.6	379/23,360
커뮤니케이션학 연구	한국커뮤니케이션학회	학제간연구	1.5	354/23,360
방송통신연구	한국방송학회	신문방송학	1.4	333/23,360
한국광고홍보학보	한국광고홍보학회	신문방송학	1.4	328/23,360
미디어, 젠더 & 문화	한국여성커뮤니케이션학회	신문방송학	1.4	318/23,360
지역과 커뮤니케이션	부산울산경남언론학회	신문방송학	1.3	296/23,360
미디어 경제와 문화	(주)에스비에스	신문방송학	1.2	287/23,360
광고연구	한국광고홍보학회	신문방송학	1.2	274/23,360
방송과 커뮤니케이션	방송문화진흥회	신문방송학	1.1	264/23,360
언론정보연구	언론정보연구소	신문방송학	1.1	252/23,360

7) 한국연구재단의 학술지인용색인 데이터는 <한국언론학보>가 등재후보지가 된 1999년 이후의 데이터로, 피인용 논문은 한국연구재단 등재지 또는 등재후보지 이상의 논문을 대상으로 산출된 회수이다.

최근 4년간 논문 생산이 감소된 또 다른 이유는 등재학술지 유지 조건 기준 강화가 투고 논문 수에 영향을 주었을 가능성이 있다. <한국언론학보>는 2001년 등재학술지로 평가된 후 2017년부터 2019년 4월 현재까지 우수등재학술지를 유지하고 있다. 등재학술지 평가 요건에 게재율, 게재논문의 학술적 성과와 가치, 학술지의 전문성, 논문초록의 전문성 등이 반영되면서 엄격한 심사절차를 거치게 되었고 이에 따라 투고 논문 수가 많더라도 게재 논문 수가 감소했을 가능성이 있다.

마지막으로 교육제도의 변화가 논문 수의 증감 원인이 되었을 가능성이 있다. 허가제였던 대학설립이 김대중 대통령 재임 시기 대학설립준칙주의로 요건이 완화되면서 언론관련 학과 신설이 전국적으로 늘어 신입 교수 임용이 활발하게 이루어졌다. 그 결과 2001년 논문 수가 급증한 것은 아닌지 추측해 볼 수 있다. 교육부 자료에 따르면 1996년 대학설립준칙주의가 시행된 이후 2001년 개정까지 94개의 대학이 설립되어 종합대학의 18.8%가 이 시기에 신설되었다. 또 1994년부터 시작된 중앙일보 대학평가가 자리를 잡으면서 평가 지표에 교육연구 분야, 논문 실적의 중요성이 커지면서 논문의 양적 증가에 영향을 주었을 가능성이 있다. 2015년 이후 논문 수 감소도 같은 맥락에서 풀이해볼 수 있다. 언론관련 학과의 신규임용이 활발히 이루어지던 20여 년 전과 달리, 2010년대 이후 신규 임용은 대폭 줄어드는 추세⁸⁾이다. 임용과정에서도 각 대학마다 국제화 실적이 중요한 평가 기준이 되면서 국내 학술지보다 해외 저널에 투고하려는 경향을 만들어낸 것이 사실이다. 국내 언론학 연구 풍토가 새 국면으로 접어들었다고 볼 수 있다. 학령인구 감소로 기존 학과 축소 또는 폐지가 이루어지고 있는 점도 논문 수 감소에 영향을 주는 것은 아닌지 논의가 필요하다.

2) 저자 및 논문 별 인용·이용 현황

<한국언론학보>의 저자별, 논문별 게재 논문 수와 인용 및 이용을 분석한 결과는 데이터베이스 특성⁹⁾으로 인해 삼각측정 방식으로 보는 것이 타당할 것이다. 우선 1960년 창간호부터 2018년 12월 31일까지 학보에 논문을 게재한 연구자는 총 1,276명이었다. 연구자 기준으로 데이터를

8) 신문방송학과 신규교원 임용이나 박사학위 취득 통계자료를 찾지 못해 사회 계열 신규교원 임용 추이를 참고한 결과 2012년 기점으로 감소 경향이 나타났다.

9) 누리미디어의 메타데이터는 1960년 이후 <한국언론학보> 논문을 전부 제공하고 있어서 논문 수는 정확하게 집계되지만, 데이터베이스 서비스 개시가 2008년부터 이루어졌기 때문에 개별 논문 및 저자별 인용 및 이용 관련해서는 한국연구재단 데이터베이스도 같이 참고하였다. 이용 및 인용 회수는 특정 데이터베이스를 참고할 수밖에 없기에 디지털 자료 축적 기간과 데이터베이스 속성을 감안해 해석해야 한다.

분석한 결과 <한국언론학보>에 가장 많은 논문을 발표한 저자는 김은미로 총 48편의 논문을 게재하였다. <한국언론학보>에 20편 이상 게재한 연구자는 이준웅(45편), 김영희(26편), 차희원(24편), 박종민(23편), 나은영(20편) 순으로 나타났다¹⁰⁾. 인용 및 이용과 관련하여 누리미디어 데이터베이스 기준 <한국언론학보> 메타데이터를 통해 2008년 3월 28일 디비피아(DBpia) 서비스 개시시점부터 2018년 12월 31일까지 발생한 이용통계(다운로드 통계)와 인용통계를 살펴본 결과¹¹⁾, 2019년 4월 기준 1만 5천 건 이상의 논문 이용이 발생한 저자는 이준웅으로 44,011회 논문 이용이 발생한 것으로 나타났고, 김은미(40,794회), 차희원(17,679회), 최윤정(16,153회), 나은영(16,059회), 김희조(15,677회), 이종혁(15,183)의 논문 순으로 이용하고 있었다. 인용통계의 경우 누리미디어 데이터 기준 <한국언론학보> 창간호부터 2018년 12월까지 발행 논문 대상 최다 피인용 저자를 살펴본 결과 김은미가 317회, 이준웅이 300회, 최윤정이 86회였다. 또한 한국연구재단 데이터베이스에서 제공하는 인용 지수인 H지수¹²⁾로 인용 현황을 확인한 결과 연구재단 등재후보지로 등록된 1999년 이후 이준웅, 김은미, 이수범, 이정기 순으로 인용과 이용 빈도가 높은 것으로 나타났다.

한국연구재단에서 제공하는 <한국언론학보> 개별논문 중 100회 이상 인용된 논문을 확인한 결과는 <표 2>와 같다. 측정과 타당성 검증 등 실증연구 관련 방법론 연구, 이론적으로는 기술수용모델, 이용과 충족, 프레임(틀짓기) 연구, 연구 대상으로는 이동전화와 매체 이용 관련 연구가 인용이 되고 있음을 알 수 있다.

표 2. 한국연구재단 데이터베이스 자료 기준 <한국언론학보> 개별논문 피인용 횟수

발행연월	권(호)	논문제목	인용회수
2003.12	47(6)	포괄적 대인 의사소통 능력 척도개발 및 타당성 검증	228
2011.04	55(2)	기술수용모델을 이용한 초기 이용자들의 스마트폰 채택 행동 연구	202
2004.12	48(6)	이동전화 중독 척도 개발 및 타당성 검증	164
2004.10	48(5)	인터넷의 웹블로그(Web-blog) 이용동기와 만족도에 관한 연구: 대학생 집단을 중심으로	141
2003.04	47(2)	대학생들의 이동전화 중독증에 관한 연구	141

10) 단독저자, 공동저자, 교신저자 포함 논문 수

11) '이용'은 논문파일을 내려 받기한 회수이며 '인용'은 누리미디어 데이터베이스 내 논문에서의 피인용을 의미하는데, 두 개념에 대한 해석은 신중할 필요가 있다.

12) 한국연구재단은 2004년 이후의 논문에 대한 총 논문 수 및 H지수를 제공하고 있기 때문에 연구자 인용정보 해석에 유의해야 한다. 연구자 논문 중에서 논문 h개가 적어도 각각 h개 이상의 인용을 받고, 나머지 논문이 h개와 같거나 적은 인용을 받을 때로 피인용수가 많은 순으로 정렬해 순위를 매긴다. H지수는 연구자 개인의 전체 인용 점수를 나타낸다.

발행연월	권(호)	논문제목	인용회수
2003.10	47(5)	한류의 커뮤니케이션 효과: 중국인의 한국 문화상품 이용이 한국에 대한 인식과 태도에 미치는 영향	133
2011.02	55(1)	이용 동기에 기반한 스마트폰 초기이용자 유형에 관한 탐색적 연구	122
2005.06	49(3)	사회자본 형성의 커뮤니케이션 기초: 대중 매체 이용이 신뢰, 사회 연계망 활동 및 사회정치적 참여에 미치는 영향	118
2006.02	50(1)	다매체 이용자의 성향적 동기 : 다매체 환경에서 이용과 충족 이론의 확장	111
2012.02	56(1)	CSR 활동의 진정성이 기업태도에 미치는 영향에 관한 연구	108
2005.12	49(6)	언론보도의 프레임 유형화 연구: 국내 원자력관련 신문보도를 중심으로	103
2006.08	50(4)	웹기의 재발견 : 인터넷 토론 공간에서 커뮤니케이션의 효과	101

6. 다이나믹 토픽 모델링 분석 결과 :

〈한국언론학보〉 논문의 주요 토픽 분석

전집 논문 메타데이터를 활용해 5년 단위로 총 11개 시간 구간에서 DTM을 생성한 결과, 시간에 따라 발생 확률이 변화하는 단어들로 이루어진 100개의 토픽을 얻을 수 있었다. 어떤 토픽이 지배적으로 나타나는지 토픽의 진화 패턴은 어떠한지 결과를 설명하기 위해 〈한국언론학회 50년사〉(2009)의 ‘한국 언론학 연구 50년’에서 나눈 17개 영역별 연구를 토대로 발생가능성이 높거나 변동 폭이 큰 토픽 80개를 6개 분류 체계로 정리하였다(〈표 3〉)¹³⁾. 6개 분류체계의 근거는 우선, ‘한국 언론학 연구 50년’의 17개 영역별 연구가 공통적으로 ‘방법론’, ‘이론’, ‘메시지 연구’, ‘수용자연구’, ‘연구대상’ 또는 ‘연구주제’를 분류 체계로 삼고 있고, 또 위 연구의 17개 연구영역으로 분류할 경우 추출된 토픽이 ‘연구영역’으로 분류되지 않는 높은 발생확률의 토픽이 다수였기 때문에 6개의 군집 분류 체계를 설정하였다. 토픽명은 토픽을 구성하는 주요 단어를 확인하고 DTM 결과 데이터와 메타데이터의 논문제목, 초록을 상호 체크하고 pyLDavis패키지 틀에 의해 생성된 토픽과 단어들의 주성분 분포의 관련성을 확인한 후 추론을 통해 결정하였다.¹⁴⁾

13) 분류는 언론학 역사에 대한 통합적 선행연구인 〈한국언론학회 50년사〉(2009)를 토대로 높은 확률로 나타난 토픽과 토픽을 구성하는 키워드의 확률에 의거해 분류한 것이다. 토픽모델링 결과를 보고 개별 토픽을 묶는 틀은 연구자의 해석과 관점에 따라 다르게 나타날 수도 있는데, 따라서 이 분류체계와 다른 접근으로 해석이 가능할 것이다. 다만, 어떤 분류체계를 채택하더라도 사회과학으로서 언론학이 다루는 현상과 주제, 대상, 이론, 방법론이 상호배타적인 성격이 아니기 때문에 특정 토픽은 독립적 분류체계로 설명하기 쉽지 않은 성격을 지닐 수 있다는 점을 고려해야 한다.

14) 토픽 번호는 무작위 추출이며 식별을 위해 추출된 대로 표기한다.

표 3. <한국언론학보>의 주요 토픽 및 주요 단어

토픽 번호	분류	토픽(topic)	주요 단어(keyword) : 높은 확률로 추출된 단어 순
2	연구영역	방송	드라마, 패러디임, 시각, 방영, 스타, 코드, 환상
3		법제	법, 출판, 정치, 체계, 판례
20		저널리즘1(저널리즘)	기자, 저널리즘, 취재, 관행, 기사, 인용
22		PR1(기업위기관리)	위기상황, 수용, 책임, 공중, 귀인, 기업
24		문화간커뮤니케이션	지역, 서울, 분리, 고정관념, 고유, 지역민, 분포
31		미디어경영	경쟁, 결합, 확립, 판매, 배제, 지배력
36		PR2(소셜메시지)	메시지, 행동, 의도, 태도, 효과, 페이스북, 의존도
37		휴먼커뮤니케이션1	의사소통능력, 경험, 회피, 심리
43		문화연구1	문화, 사회, 정체성, 구성원, 이주민, 테크놀로지, 심층인터뷰
47		PR3(정부)	정부, 선전, 추진, 경영, 공보
48		영상커뮤니케이션	영상, 서사, 이야기, 인물, 영상물, 구도
56		광고1(온라인)	웹사이트, 광고주, 방식, 매체, 극대, 설득력, 대행사
59		정치커뮤니케이션 (선거연구)	조사, 여론, 후보, 선거, 지지, 유권자, 캠페인, 태도
60		저널리즘2(뉴스보도)	보도, 원인, 정보원, 편향, 사례
72		휴먼커뮤니케이션2	대인, 형성, 판단, 관계, 지속, 유지, 상대방
79		정책	정책, 규제, 사업, 기관, 지원, 정부, 주체
81		저널리즘3(온라인)	뉴스, 블로그, 트위터, 생산, 온라인, 언론사, 이슈
85		광고2(홍보전략)	전략, 조직, 공중, 브랜드, 상호, 커뮤니케이션
86		저널리즘4(언론)	언론, 사회, 언론인, 국제, 보도, 취재원, 언론사
90		비판커뮤니케이션	담론, 역사, 비판, 실천, 민족, 주체
93	문화연구2(미디어공간)	미디어, 공간, 라디오, 관찰, 공동체, 경험, 삶	
100	건강커뮤니케이션	규범, 주관, 건강, 관리, 정신, 모바일	
10	방법론	매체실증연구	매체, 이미지, 특성, 실증, 분석
23		설문조사연구	인지, 성향, 영향, 중독, 적응, 만족도, 개인, 설문
35		양적연구1	효과, 매개, 검증, 소구, 긍정, 분석
41		질적연구	텍스트, 주체, 기호, 사회, 타자, 심층, 인터뷰
42		내용분석1	유형, 분류, 비율, 차이, 분석, 내용, 대상
53		양적연구2	변인, 세대, 통계, 실증, 검증, 종속, 회귀
61		내용분석2	분석, 방법, 내용, 구성, 주제, 관련, 대상
97		컴퓨터이셔널	네트워크, 컴퓨터, 구조, 발달, 동학
6	이론	이론에 대한 이론	사회, 과학, 이론, 체계
7		프레임1(프레이밍)	기대, 프레임, 피해
8		지식격차	격차, 수준, 균형, 차이
14		프레임2(텍스트)	프레임, 댓글, 기사, 내재, 맥락, 결과, 영향
15		의제설정	선택, 의제, 단서, 이슈, 경향
19		프레임3(속성)	결정, 속성, 관점, 의사, 이론
32		이용과 충족1	이용, 인터넷, 동기, 충족, 추구, 이용자, 요인
67		뉴미디어학산	커뮤니케이션, 채택, 인구, 특성, 차이, 연령, 용이성
70		제3자 효과	자각, 자신, 지식, 영향력, 가설, 타인, 제삼자, 영향
80		근대화 발전이론	발전, 개념, 사회, 근대, 이동전화, 개인
83		정보처리이론	정보, 처리, 추구, 통계학, 분석
95		이용과 충족2	이용자, 포털, 특성, 네이버, 수용성, 개인주의

| 기획 | 언론학 60년 돌아보고 내다보다 <한국언론학보 60년 연구>

토픽 번호	분류	토픽(topic)	주요 단어(keyword) : 높은 확률로 추출된 단어 순
1	메시지연구	메시지효과1	영향, 요인, 통제, 결과, 설문, 예측, 이론
28		메시지효과2	정서, 조절, 유발, 공포, 긍정, 죄책감
30		연구대상	대학생, 학생, 효능감, 관계, 영향, 학업, 동료
87		정파성	이념, 존재, 보수, 민족, 적대, 관여, 북한이탈자
89		설득	집단, 차이, 설득, 유의, 선호, 자극
92		방송메시지	프로그램, 장르, 텔레비전, 재현, 방송
50	수용자연구	소비자연구	소비자, 관여, 제품, 광고, 구매, 신뢰도 기억, 보상
52		정치참여	정치, 참여, 시민, 투표, 결과, 신뢰, 관심, 전국
55		수용자연구	수용자, 해석, 이론, 스키마
63		청소년1(TV시청)	시청, 텔레비전, 시간, 시청자, 자녀, 경험, 부모, 오락
74		청소년2(모바일)	교육, 청소년, 스마트폰, 전문가, 학교, 학계, 협력
88		태도연구	태도, 모델, 모형, 지속, 태도, 영향, 검증, 기대
4	연구대상 또는 연구주제	정치인	정치인, 원칙, 트윗, 의미, 방송, 법제
5		온라인상호작용	상호작용, 위험, 개인 미래, 음란, 온라인
11		미디어산업	산업, 시장, 성과, 제작, 투자, 창출, 제작비
13		문화 젠더	여성, 남성, 이데올로기, 상품, 자본주의, 가부장, 한류
16		미디어규범 가치	가치, 지향, 사상, 공정, 원리, 복지, 추구, 규범
17		지적재산권	해외, 검색, 문제점, 진단, 저작물, 저작권법, 계약
21		소셜미디어	사회, 온라인, 노출, 영향, 소셜, 오프라인, 미디어
27		표현의 자유	자유, 실제, 방송사, 제도, 공론, 목표, 이익
29		종이신문	보도, 기사, 경제, 분석, 일간지, 내용, 뉴스, 제목, 수집
38		독자	신문, 독자, 발행, 잡지, 신문사, 조선일보, 동아일보
39		흡연연구	관계, 상관, 정책, 빈도, 흡연, 담배
45		사이버범죄	언어, 폭력, 사이버, 경계, 피해, 가해, 디지털
46		루머확산	인식, 이슈, 뉴미디어, 영향, 인플루엔자
54		대통령1	부정, 긍정, 대통령, 리더십, 박근혜, 노무현
58		명성	기업, 국내, 홍보, 경영, 명성, 수익
64		정치루머	국가, 정부, 반응, 공격, 확산, 루머, 대응, 전략
68		방송정책	방송, 채널, 경쟁, 케이블, 지상파, 편성, 경제, 사업자, 공영, 위성, 도입
69		매체비평	기술, 윤리, 디지털, 어린이, 철학, 정립, 기초
73		대통령2	김영삼, 영부인, 형용사
77		미디어 역사	한국, 미국, 전쟁, 한국인, 소유, 해방, 과정
78		방송서비스	서비스, 기능, 공공성, 공익, 대체, 레퍼토리, 방송국
82		북한	북한, 범죄, 진보, 남한, 사회, 주민
94		인터넷토론	메스, 의견, 토론, 인터넷, 견해, 관용, 참여, 민주주의
98		갈등이슈	사회, 문제, 갈등, 대안, 통합

우선 추출된 100 개 토픽이 전체 문서에 잘 분포되어 있는지 확인하기 위해 pyLDAvis패키지 툴을 사용하여 생성된 토픽과 단어들(keyword)을 검토하였다. <그림 5>는 주성분 분석을 통해 <한국언론학보>의 토픽 분포를 시각화한 것이다. 토픽이 잘 추출되었을 경우 좌측의 원의 크기가 일반적으로 크게 나타나고 토픽들이 하나의 사분면 전체에 골고루 분포된다. 2015~

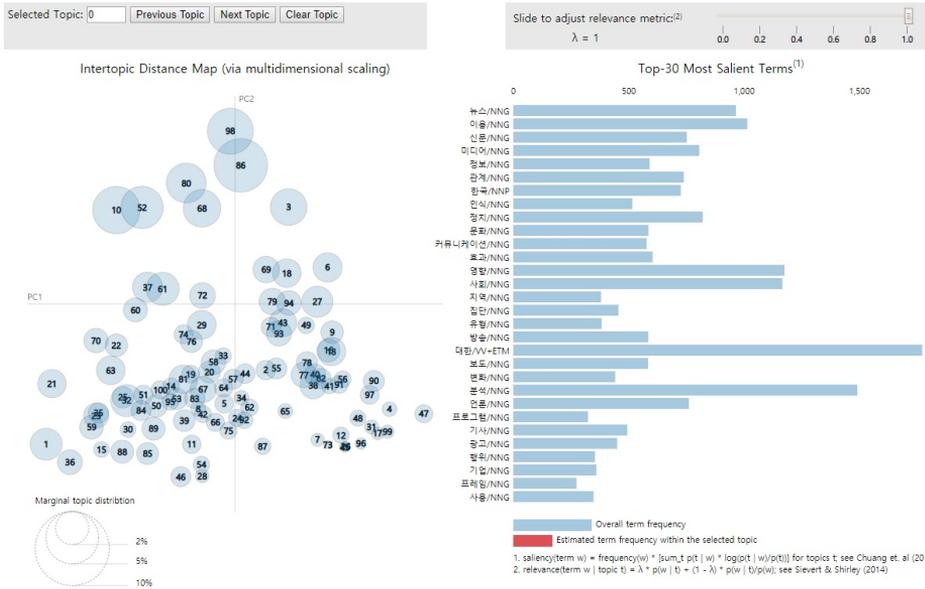


그림 5. 2014~2018년 100개 토픽의 주성분분석 분포(좌측) 및 토픽을 구성하는 지배적인 단어 30개 (우측 막대그래프)

2018년까지의 DTM 전체의 토픽을 구성하는 지배적인 단어는 뉴스, 커뮤니케이션, 미디어, 문화, 신문, 관계, 정보, 사회, 효과 등으로 추출되었다(우측 막대그래프 참조). <표 3>에 정리한 토픽 중 “저널리즘4(언론)”(토픽 86), “갈등이슈”(토픽 98), “방송정책”(토픽 68), “정치참여”(토픽 52), “매체실증연구”(토픽 10), “법제”(토픽 3), “표현의 자유/공론장”(토픽 27), “내용분석2”(토픽 61), “휴먼커뮤니케이션”(토픽 37), “문화-젠더(토픽 13) 등은 <한국언론학보>의 논문 전반에 골고루 분포된 토픽이라고 해석할 수 있다.

DTM은 하나의 토픽이 진화하는 양상을 계산하기 때문에 시간 흐름에 따라 토픽을 구성하는 단어의 발생 확률이 바뀐다. 예를 들어 진집 논문에서 가장 골고루 분포되어 있을 것으로 추측되는 86번 토픽인 “저널리즘4(언론)”의 경우 1969~1973년과 2015~2018년 두 시기를 비교한 결과 <그림 6>, <그림 7>의 우측 그래프에서 나타나는 것처럼 토픽을 구성하는 지배적인 단어에 변동이 있음을 확인할 수 있다.

<그림 6>의 우측 그래프를 보면 1969~1973년에는 토픽 86 “저널리즘4(언론)”을 구성하는 단어는 ‘언론’, ‘사회’, ‘언론인’, ‘분석’, ‘국제’, ‘보도’, ‘한국’ 순으로 엮이는 것을 볼 수 있다. <그림 7>의 우측 그래프는 2015~2018년의 “저널리즘4(언론)” 토픽을 구성하는 단어로 ‘언론’, ‘사회’, ‘분석’, ‘취재원’, ‘보도’, ‘언론인’, ‘언론사’ 순으로 엮여 있다. 각 시기 토픽을 구성하는 단

Selected Topic: 86 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾
λ = 1 0.0 0.2 0.4 0.6 0.8 1.0

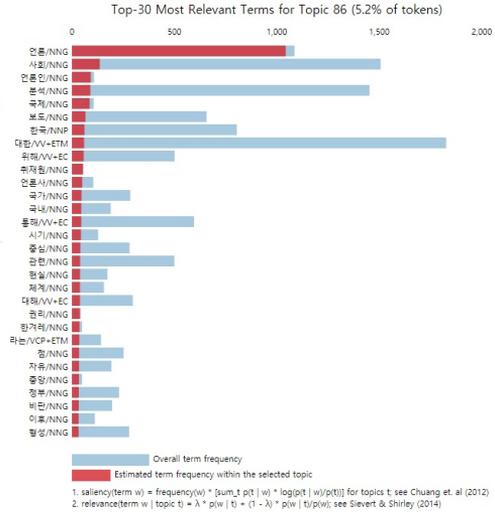
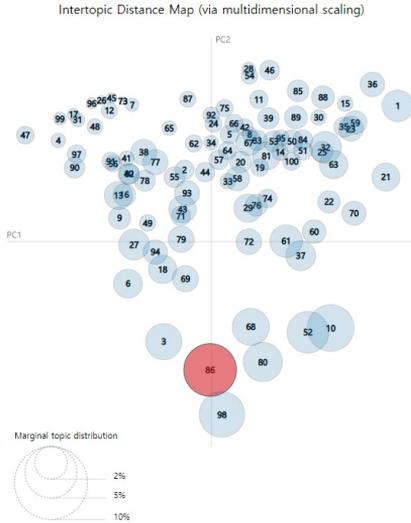


그림 6. 1969~1973년 86번 토픽을 구성하는 지배적인 단어 30개(우축 막대그래프)

Selected Topic: 86 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾
λ = 1 0.0 0.2 0.4 0.6 0.8 1.0

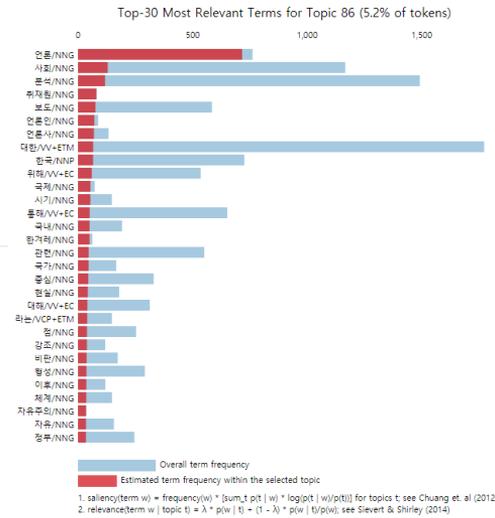
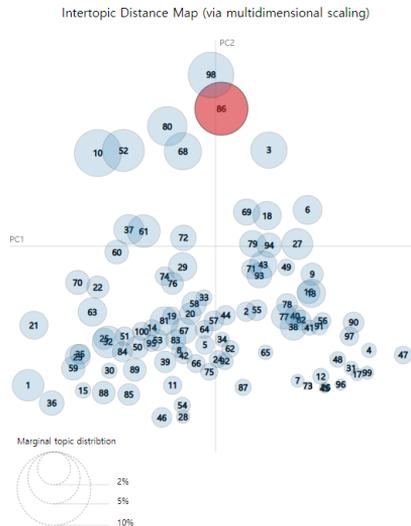


그림 7. 2014~2018년 86번 토픽을 구성하는 지배적인 단어 30개(우축 막대그래프)

어의 확률적 비율이 다르기 때문에 토픽이 변동성을 갖는 것이다. 이처럼 모든 토픽은 시간에 따라 발생가능성이 다른 단어로 묶이기 때문에 정적인(static) 상태를 유지하는 것이 아니라 역동적으로ダイナ믹하게 진화한다. 시각화를 통해 시간에 따라 토픽을 구성하는 단어의 발생이 달라지고 이에 따라 토픽의 진화 양상도 달라지는 결과를 확인하였다.

〈표 3〉의 연구영역, 방법론, 이론, 메시지연구, 수용자연구, 연구대상 또는 연구주제로 분류한 토픽의 발생확률 진화 추이를 DTM 생성을 통해 분석한 결과 〈그림 8〉~〈그림 13〉과 같이 나타났다. 모든 토픽을 그래프로 그리지 않고 높은 발생 확률의 토픽과 확률의 발생추이를 비교해볼 수 있는 토픽을 선별하여 제시하였는데, 연구영역(〈그림 8〉)으로 분류된 토픽 22개 중 12개 토픽을, 방법론(〈그림 9〉)은 8개 토픽 모두를, 이론(〈그림 10〉)은 12개 토픽 중 10개 토픽을, 각각 6개의 토픽이 추출된 메시지연구(〈그림 11〉)와 수용자연구(〈그림 12〉)는 모든 토픽을, 총 24개의 토픽으로 분류된 연구대상 또는 연구주제(〈그림 13〉)는 10개의 토픽을 그래프로 나타냈다.

그중 특징적인 토픽을 살펴보면, 연구영역과 관련하여 “저널리즘1~4”(토픽 20, 60, 81, 86) 관련 토픽이 가장 많이 추출되어 앞서 시각화에서 확인하였듯 학보 전반에 걸쳐서 저널리즘 연구가 이루어지고 있음을 추측할 수 있다. 저널리즘 토픽이 하나로 엮이지 못한 까닭은 아마도 연구대상, 방법론, 이론에서 차이가 나기 때문으로 해석할 수 있다. “저널리즘3(온라인)”(토픽 81)은 높은 확률의 발생가능성과 함께 2014년까지 하락추세였다가 다시 상승 추세 패턴으로 나타났다. 이 토픽을 구성하는 단어가 ‘뉴스’, ‘블로그’, ‘트위터’, ‘생산’, ‘온라인’, ‘언론사’, ‘이슈’로 나타나 온라인 저널리즘 관련 연구가 급증했다고 추론할 수 있다. 또 “저널리즘4(언론)”(토픽 86)는 학보의 전체 논문에서 전반적인 분포를 나타냄에도 불구하고 발생 확률 비율(0.01776)이 여타 토픽보다 높게 나오지는 않았는데, 토픽을 구성하는 단어가 ‘언론’, ‘사회’, ‘언론인’, ‘국제’, ‘보도’, ‘취재원’, ‘언론사’, ‘국내로 유인 것으로 볼 때 취재 관행연구 또는 생산자 연구, 국내외 보도 비교와 같은 연구로 분류될 수 있는 논문이 그리 많지 않다는 것으로 추론해볼 수 있고, 토픽으로 엮이기는 하지만 실제 논문 수는 상대적으로 드물 수 있다. “문화연구1,2”(토픽 93, 43) 관련 토픽은 발생가능성이 높았지만, 뚜렷한 하향세임을 볼 때 학보에서 문화연구 논문은 감소 추세였을 것으로 추측된다. “PR1~3”(토픽 22, 36, 47) 관련 토픽도 학보의 전체 논문에서 발생가능성이 높을 것으로 보이며, 특히 “PR1(기업위기관리)”(토픽 22)는 위기 및 기업의 명성, 책임과 관련된 PR 분야로 볼 수 있고, “PR2(소셜메시지)”(토픽 36)는 소셜 메시지에 대한 태도와 효과와 관련된 토픽으로 해석할 수 있다. “광고2(홍보전략)”(토픽 85)는 광고에서부터 홍보까지를 포함한 전략 커뮤니케이션과 관련이 있을 것으로 볼 수 있어서 이 분야 연구가 상당히 이루어

졌을 것으로 추측된다. “정치커뮤니케이션(선거연구)”(토픽 59)는 ‘조사’, ‘여론’, ‘후보’, ‘선거’, ‘지지’, ‘유권자’, ‘캠페인’, ‘태도’로 엮여서 선거연구와 관련된 토픽으로 추측할 수 있는데 다만, 이 토픽은 발생 확률 비율 추세가 완만하게 하락추세로 나타났다. 이 밖에 “법제”(토픽 3), “휴먼 커뮤니케이션1,2”(토픽 37, 72), “정책”(토픽 79), “비판커뮤니케이션”(토픽 90), “건강커뮤니케이션”(토픽 100) 등의 토픽도 추출되었다.

방법론과 관련해서는 크게 양적연구로 분류되는 “양적연구1,2”(토픽 35, 53), “설문조사연구”(토픽 23), “내용분석1,2”(토픽 42, 61)와 “질적연구”(토픽 41), “컴퓨터이셔널”(토픽 97)로 나타났다. 경험주의적 연구는 방법론에 있어서 학보 전체에서 지배적인 토픽으로 엮일 것이라는 추론을 할 수 있다. 질적연구는 꾸준한 발생가능성을 유지하고 있으나 시간 추이와 관계없이 논문의 수와 비례해 방법론으로 채택되고 있을 경우 앞서 “문화연구1,2”(토픽 93,43)의 발생가능성이 낮아지는 추세로 볼 때 논문 수 자체는 양적연구에 비해 많지 않을 것으로 추정된다. “컴퓨터이셔널(토픽 97)” 방법론은 ‘네트워크’, ‘컴퓨터’, ‘구조’, ‘발달’, ‘동학’ 등의 단어로 엮여져 있는데, 상승 추세 패턴으로 나타나 최근 이러한 방법론을 거론하거나 적용한 연구가 증가하고 있을 것이라고 추론할 수 있다.

이론과 관련된 토픽으로 “프레임1~3”(토픽 7, 14, 19), “이용과 충족1”(토픽 32), “의제설정”(토픽 15)은 뚜렷하게 등장하는 지배적인 이론이라고 볼 수 있다. 이와 반대로 1970~80년대 논의가 많이 된 “근대화 발전이론”(토픽 80)은 토픽으로 추출은 되었으나 거의 의미 없는 발생가능성으로 나타났고, “뉴미디어확산”(토픽 67)도 한동안 많이 논의되다가 급격하게 발생가능성이 낮아지는 추세를 그리고 있다. 이는 ‘이론의 흥망성쇠’로 해석할 수 있는데, <그림 14>~<그림 17>은 이 결과를 일목요연하게 설명한다. “프레임” 이론은 학보 전반에 걸쳐서 가장 지배적인 이론일 가능성 높고, 그 중에서도 토픽을 구성하는 단어의 확률로 보았을 때 댓글 및 기사 등 텍스트를 주 연구 대상으로 삼고 있는 프레임 연구가 주를 이루고 있을 것이라 추측할 수 있다. “이용과 충족” 이론은 꾸준히 지배적으로 나타난 프레임이론과 달리 상승 추세 패턴이라 더 높은 발생가능성으로 나타날 것으로 추측할 수 있는데, ‘인터넷’, ‘포털’, ‘이용’, ‘동기’가 토픽을 구성하는 단어로 엮여져 나온 것을 볼 때 뉴미디어 이용 연구에 “이용과 충족” 이론이 자주 적용되었을 것으로 짐작된다. “이용과 충족” 이론이 지배적인 토픽으로 나타났는데, 앞서 <표 2>의 결과와도 일치하기 때문에 이를 방증한다고 할 수 있다.

메시지연구와 관련해서는 “방송메시지”(토픽 92)가 높은 발생가능성으로 나왔고, “메시지 효과1,2”(토픽 1, 28), “설득”(토픽 89) 등은 측정 및 연구결과를 기술하는 관련 단어로 엮여 높은 발생가능성으로 나왔다. 이는 양적연구 방법론과 관련된 토픽으로 시간의 흐름과 상관없이 꾸

준히 지속되는 토픽으로 추출되고 있다. 특이한 것은 “연구대상”(토픽 30)으로 추출된 토픽인데, ‘대학생’, ‘학생’, ‘효능감’, ‘관계’ 등의 단어로 엮여 메시지 연구대상이 주로 대학생이나 학생에게 편향된 것이 아닌가 추측하게 된다.

수용자연구와 관련해서는 “청소년1, 2”(토픽 63, 74)가 높은 확률의 미디어 이용 관련 토픽으로 추출되었다. 텔레비전 시청 및 이용, 스마트폰 및 부모 교육과 관련된 단어로 구성되어 있어서 수용자 연구 분야에서는 미디어 환경 변화와 연동한 미디어교육 관련 연구가 이루어지고 있음을 추론할 수 있다. 그리고 “정치참여”(토픽 52)가 상승세를 보이는 토픽으로 추출되었는데 ‘정치’, ‘참여’, ‘시민’, ‘투표’, ‘결과’, ‘신뢰’ 등의 단어로 엮여 있어 참여민주주의 관련 연구가 활발히 진행되고 있음을 알 수 있다.

연구대상 또는 연구주제와 관련해서는 많은 토픽이 추출되었는데, 이는 엄밀하게 연구대상과 연구주제에 해당된다기보다 언론학 특성상 분류체계로 구분하기 어려운 토픽으로 엮인 경우도 상당부분 있다. 높은 발생가능성을 보인 “온라인상호작용”(토픽 5)과 “사이버범죄”(토픽 45)는 주로 ‘위협’과 ‘음란’ 또는 ‘사이버’, ‘폭력’, ‘디지털’, ‘가해’ 등과 관련된 단어로 엮여 있어서 인터넷의 유해성 또는 역기능에 대한 연구로 볼 수 있으며, 이와 관련한 연구가 <한국언론학보>에서 지속적으로 논의되어 왔다고 볼 수 있다. 반대로 “인터넷토론”(토픽 94)은 같은 온라인 관련 연구임에도 불구하고 ‘참여’, ‘민주주의’, ‘관용’, ‘의견’, ‘견해’ 등의 단어와 묶여 있어서 앞서 “온라인” 관련 토픽과는 상반된 인터넷의 긍정적 기능, 정치참여를 다루고 있다고 해석할 수 있다. 소셜미디어와 루머확산 관련 토픽도 발생 가능성이 높게 나타났는데, 지난 10년간 스마트폰 보급의 결과로 이 분야 연구가 상당수 발생했을 것으로 보인다. 시간 흐름에 따른 추세 패턴으로 보았을 때 흥미로운 토픽은 발생가능성이 확률적으로 현저하게 낮아지고 있는 “흡연연구”(토픽 39)와 “미디어 역사”(토픽 77), “독자”(토픽 38) 등이었다. “흡연연구”는 언론학에서 자주 다루던 연구대상이었으나 사회적 제도와 규제 정책의 도입으로 금연이 보편화되면서 2000년대 전후로 관련 연구가 거의 사라진 것으로 추론할 수 있다. 한편으로 흡연 관련 미디어 연구가 정책 개선에 기여한 것으로도 해석할 수 있을 것이다. “미디어 역사” 연구는 미군정기 해방정국 또는 근대 및 일제 강점기 신문 독자 연구와 관련된 단어로 구성된 토픽으로 추출되었으나 연구의 수가 많지 않을 것으로 추측된다.

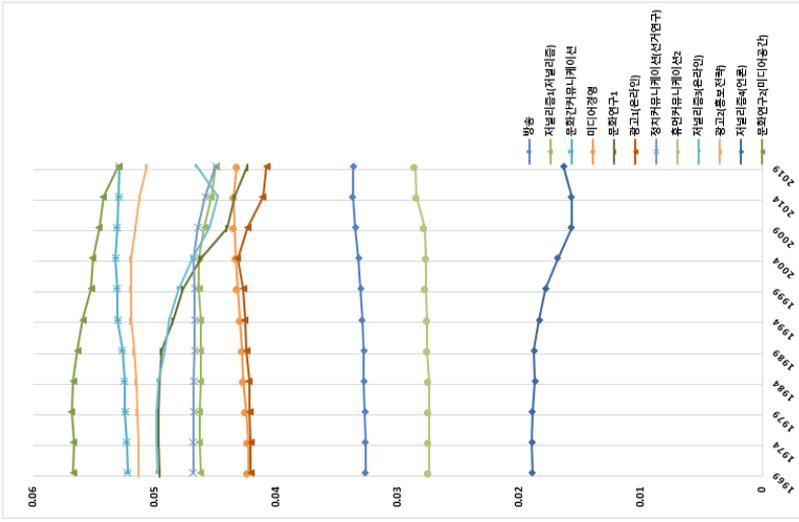


그림 8. 연구영역 DTM 결과

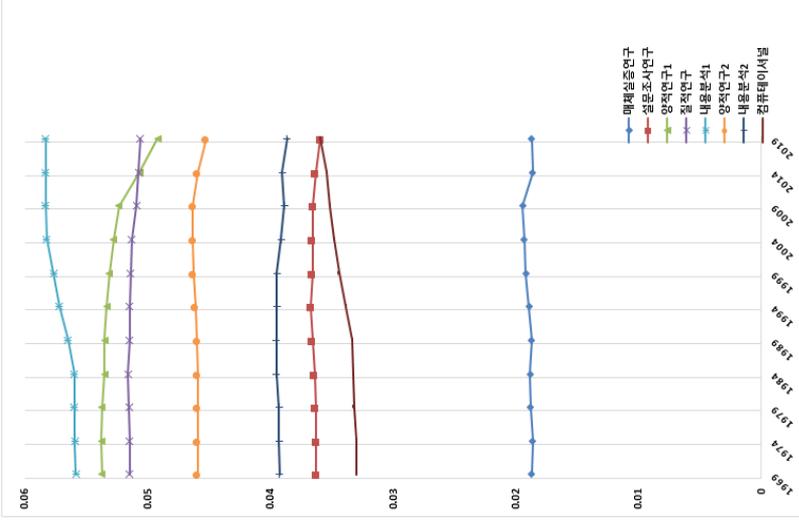


그림 9. 방법론 DTM 결과

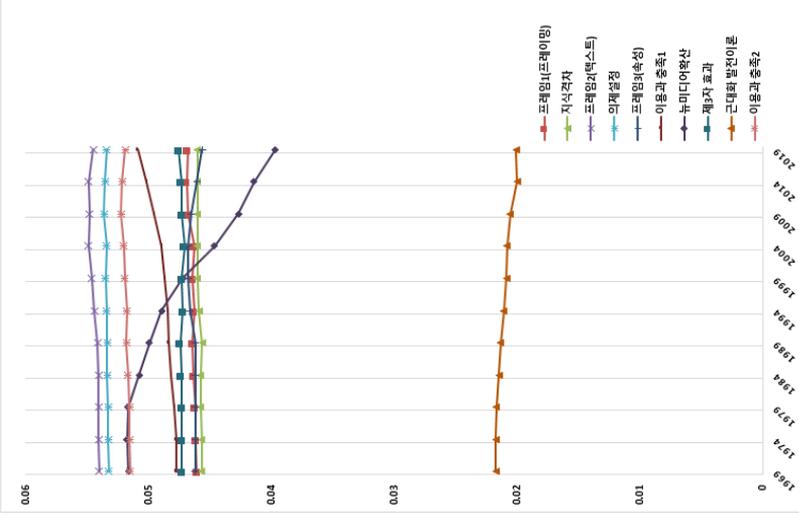


그림 10. 이룬 DTM 결과

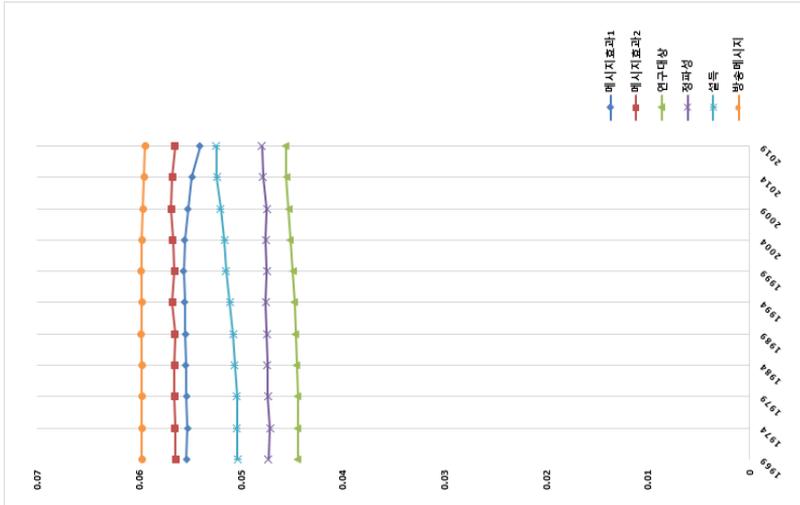


그림 11. 메시저연구 DTM 결과

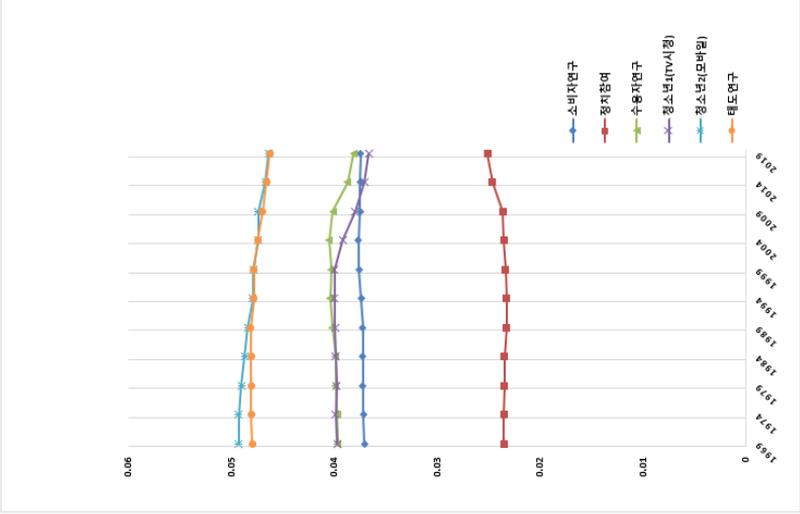


그림 12. 수용자연구 DTM 결과

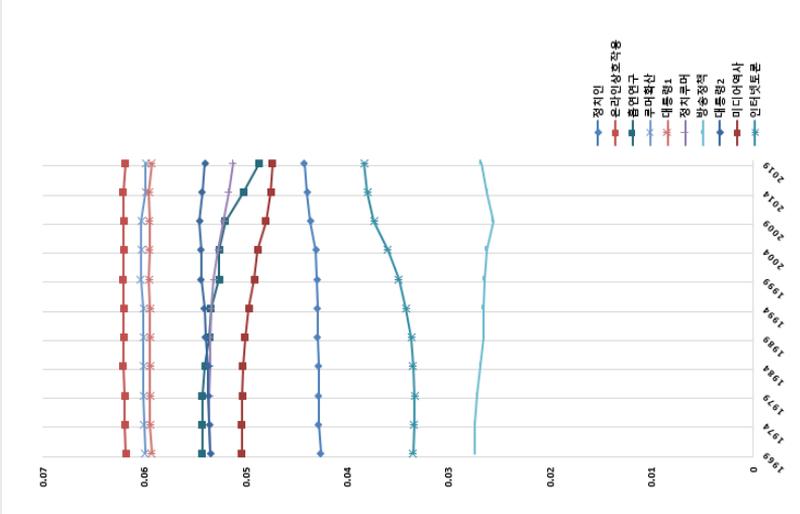


그림 13. 연구대상 또는 연구주제 DTM 결과

| 기획 | 언론학 60년 돌아보고 내다본다 <한국언론학보 60년 연구>

메타데이터를 활용한 1960~2018 <한국언론학보> 논문 분석 33

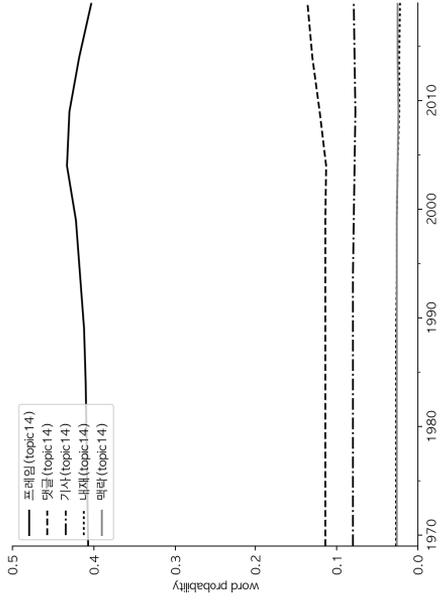


그림 14. “프레임2(텍스트)”(토픽 14)

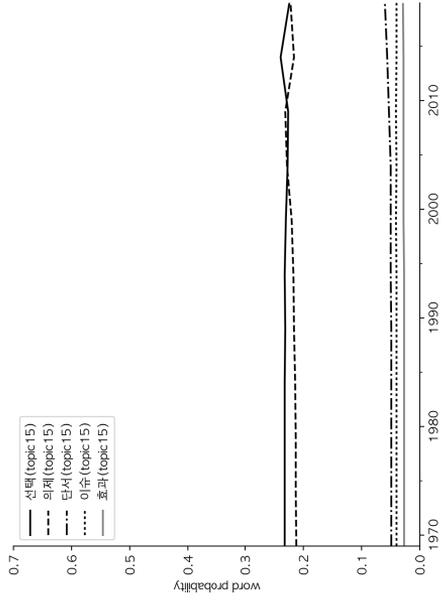


그림 16. “의제설정어론”(토픽 15)

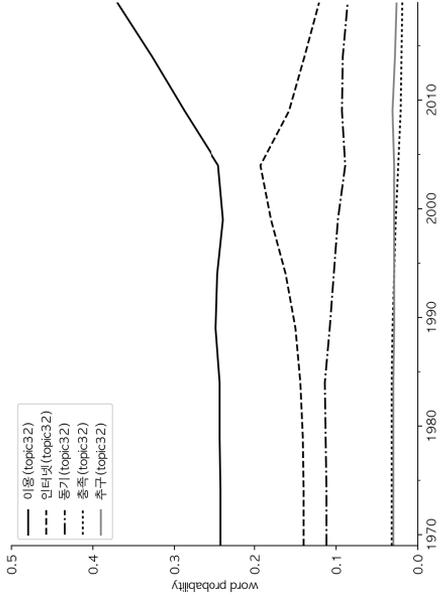


그림 15. “이용과 충족1”(토픽 32)

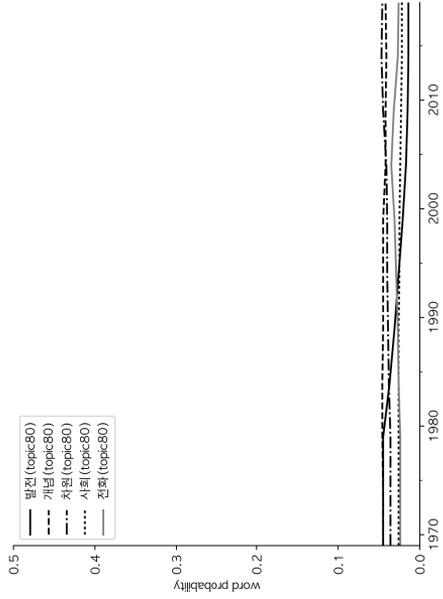


그림 17. “근대화 발전어론”(토픽 80)

7. 결론 및 논의

이 연구는 한국언론학회의 공식 학술지인 <한국언론학보>에 실린 논문의 메타데이터를 활용하여 논문 수와 인용 및 이용 관련 기술 통계를 제시하고, 다이나믹 토픽 모델링으로 토픽을 추출하여 시간에 따라 나타난 토픽의 진화 추이와 특징을 살펴보았다.

연구 결과를 요약해 논의하면 우선 <한국언론학보>의 누적 논문 수는 2,048편이었고 총 1,276명이 학보 논문의 저자로 집계되었다. 2000년 이전까지 누적 논문은 427편이었으나 이후 20년간 논문 수가 급격하게 증가해 이전 시기의 3배에 달하는 명백한 양적 성장을 이루었다. 그러나 2015년 이후 최근 4년간 논문 수가 급감해 2000년대 이전 수준으로 그 수가 떨어지고 있는 추이를 나타냈다. 학보가 진정한 지식공동체의 장으로서 역할을 하려면, 국내 언론학 연구 풍토에서 학보의 역할과 기능은 무엇인가 성찰과 반성이 필요해 보인다. 논문 생산이 활발해야 할 신진 연구자들의 경우 신규 임용의 정체로 연구비, 소속 등이 불안정한 상태라 연구 여건이 좋지 않고, 신규 임용 조건에도 국내 등재지보다 해외 학술지 논문을 실적으로 인정하는 추세라는 점에서 현재 <한국언론학보>의 논문이 갖는 학술적 가치가 무엇인지 돌아봐야 할 것이다. 논문의 인용과 이용에 있어서 학보에 논문을 20편 이상 생산한 연구자는 6명이었고, 이 중 30편 이상 게재한 연구자는 2명이었다. 데이터베이스 특성을 고려해 인용과 이용에 있어서 개별 논문을 중심으로 살펴본 결과, 측정 및 척도개발과 관련된 연구, 이동전화 및 스마트폰과 관련된 연구, 이용과 충족 및 기술수용모델 이론 관련 연구 등이 인용되고 있어서 언론학 분야에서 이정표 역할을 하는 대표 논문에 대해서도 그 기준과 의미를 발굴하려는 노력이 필요하다.

이 연구는 그동안 내용분석으로 연구의 하위영역을 분류해 분석해온 것과는 달리, 전체 논문에서 나타나는 토픽을 추출하여 시간의 흐름에 따라 어떠한 토픽이 언론학보에 나타났는지 다이나믹 토픽 모델링 방법을 활용하여 분석하였다(N = 2,044편). <한국언론학회 50년사>의 ‘한국 언론학 연구 50년’ 분류체계를 참고하여 6개 분야로 분류해 분석한 결과, 연구영역에서는 저널리즘과 온라인 저널리즘, 보도관련 저널리즘 등 전통적으로 학보에서 가장 많이 다룬 분야가 높은 발생확률의 토픽으로 추출되었다. 반면, 문화연구 분야의 연구 수는 상대적으로 하향세로 나타나 언론학보에 실리는 편수가 초창기보다 상당히 줄어들었을 것이라고 추측하였다. PR연구 분야도 온라인 커뮤니케이션의 한 영역으로 상당 부분 진화하면서 정부, 기업, 개인, 브랜드에 이르기까지 적용 범위가 넓어졌고, 외연을 확장하는 추세로 분석되었다. 방법론에 있어서는 미국의 경험주의 연구가 학계에 자리 잡으면서 언론학보 또한 경험주의적 연구나 실증연구 중심의 뚜렷한 연구 지형을 그리고 있었는데, 특기할 것은 컴퓨터이셔널 방법론이 증가 추세로 나타난 것

이다. 앞으로 언론학 연구는 단순히 양적연구, 질적연구로 나누어 미디어 현상을 분석하고 해석했던 학술적 관행을 넘어서는 도전이 필요할 것이다. 스마트폰과 앱이라는 개인 미디어의 급속한 보급과 대량으로 발생하는 미디어 행동 데이터는 언론학에 새로운 변화를 촉구하고 있기 때문이다.

이론 관련 토픽은 프레임 이론, 이용과 충족 이론, 의제설정 이론이 언론학보 전반에 걸쳐서 지배적인 이론이라는 것이 확인되었는데 우리나라 언론학 지형에서 고유한 이론적 논의나 발전 없이 특정 이론에 집중하여 검증이 되풀이 한 것은 아닌지 성찰과 도전이 필요할 것이다. 이는 '한국 언론학 연구 50년'의 각 영역별 연구에서도 비판적으로 내렸던 결론과 일치하는 결과이다. 메시지 연구와 수용자 연구의 토픽 추출 결과 측정, 변인, 효과, 영향, 검증 등과 같은 단어가 토픽에 묶여 추출되었는데 방법론으로서 양적 연구와 실증 연구가 질적 연구보다 상대적으로 주요한 방법론으로 자리 잡고 있음을 확인했다. 수용자 연구는 청소년과 자녀의 시청지도와 모바일 관련 연구가 토픽으로 추출되어 매체 환경 변화에 조응하는 미디어 교육 연구가 증가할 수 있는 가능성을 보여주었다. 특히 수용자로서 청소년 관련 토픽은 스마트폰과 같은 매체나 텔레비전 시청과 같은 단어와 묶여 있다는 점에서 앞으로 유튜브나 OTT 서비스, 게임 등과 같은 동영상 관련 연구와 연결될 수 있을 거라 추측된다. 정치참여와 관련된 토픽은 수용자 연구나, 연구대상에서도 인터넷 토론이 발생가능성이 높은 토픽으로 추출되어 의견, 토론, 견해, 관용, 참여, 민주주의라는 단어와 묶여 있다는 점에서 정치참여 공론장으로 기능하는 온라인 분야의 연구가 상당히 이루어졌음을 알 수 있었다. 연구대상 또는 연구주제로 분류된 토픽이 다양하게 추출되었는데 이는 언론학의 다원주의를 설명하는 것으로 풀이할 수 있으며, 특히 언론학이 종인신문이나 독자, 미디어 역사 등과 같은 과거 관련 연구주제보다 온라인, 사이버 범죄, 소셜 미디어, 인터넷 토론과 같은 신규 미디어 현상에 집중되어 있다는 것을 확인할 수 있었다.

우리는 언론학과 관련한 선행 연구를 많이 읽고 분석하기보다 <한국언론학회 50년사>의 주요 연구를 숙독하면서 모델링 계산과 적합성 테스트에 긴 시간을 들여 노력 했던 만큼 해석에 설명력이 있는지를 논의하고 싶었다. 제한된 연구기간 동안 수많은 모델링 테스트와 시행착오를 거쳐 계산하였지만, DTM 결과와 해석은 무척 조심스러운 일이다. 이는 이 연구의 한계와도 연결된다.

우선 데이터처리에 관한 것이다. <그림 6>~<그림 7>와 같이 토픽과 토픽을 구성하는 주요 단어의 분포를 시각화한 이유는 각 시기별 토픽의 변동성을 비교하기 위해서였다. 그러나 11개 구간의 각 시기를 전부 비교한 결과 토픽 간 변동성이 크지 않아서 이에 대한 해석에 대해 깊이 고민하였다. 왜냐하면 언론학보가 다루는 연구영역은 현재 26개 분과가 있을 만큼 다양하고, 시간 흐름에 따라 미디어 환경변화, 기술의 발전, 제도 개혁 등에 의해 변동성이 크게 나올 것이라

고 가정했기 때문이다. 왜 시계열 변동성이 크지 않은 결과가 나왔는지에 대해 조심스럽게 추측하자면 시기별로 데이터를 분석하기에는 각 토픽 영역에서 충분한 크기의 문서가 없었던 것은 아닌지 의심해볼 수 있다. 이는 데이터 크기의 문제가 아니라 각 연구주제 혹은 연구영역에서 충분한 연구들이 이루어졌는가와 관련이 있다. 의미 구조상 엮이기 어려운 파편화된 연구들이 많다는 것은 하나의 연구 영역으로 특정하기 모호한 연구가 많다는 의미일 수 있다. 둘째, 6개 분류 체계의 타당성 문제이다. 토픽모델링 결과로 도출된 토픽을 군집으로 묶기 위해 과거 분류체계를 활용하는 것은 이 연구가 추구하고자 했던 새로운 방법과 다소 거리가 있다. 또한 연구자의 해석에 따라 군집의 패턴을 달리 적용할 수 있다는 점에서 다소 유동적인 결과가 나올 수 있다. 우리는 6개의 분류 기준 없이 높은 발생확률로 추출된 상위 2~30개의 토픽을 중심으로 기술하는 것도 고민하였으나 최종적으로 과거의 분류체계를 반영해 결과를 해석한 까닭은 자칫 언론학이 다루는 연구 범위와 영역을 협소하게 축소할 수 있는 가능성 때문이었다. 연구자들은 이러한 연구의 한계를 받아들이고 향후 후속 연구를 통해 정확도가 높은 토픽 모델링 계산과 방법론으로 현재의 결과를 검증하고 설명력을 높이는 노력을 이어가고자 한다.

우리가 〈한국언론학보〉를 대상으로 연구의 흐름이 어떻게 전개되어 왔는지 알고리즘을 활용해 분석한 이유는 언론학 연구 60년에서 나타나는 언론학의 뚜렷한 정체성은 무엇인지, 학문적 관심 영역의 변화는 어떠한지 통시적으로 살펴보기 위해서였다. 연구 결과, 선행 연구를 디딤돌 삼아 연구의 방향이나 이론화를 위한 학술적 발견을 쌓아 올렸다고보다는 다양한 미디어 현상에 대한 개별연구가 산발적으로 이루어진 패턴이 특징으로 나타났다. 이 연구는 앞서 언급한 블라이(Blei, 2006)의 〈사이언스〉지 논문 분석 틀을 참고해 분석한 것인데, 언론학보 논문에서 우리나라 언론 현상에서 발견되는 고유한 개념과 이론화가 진화해 간 흔적을 잘 찾을 수 없다는 점은 아쉬운 대목이다. 일반과학과 사회과학의 차이에서 기인한, 연구 영역이 매우 다양한 언론학 분야의 특성으로 인한 너무 당연한 결과일 수 있다. 그럼에도 불구하고 우리나라 언론환경에 맞는 개념과 고유한 이론이 추출되지 않았다는 것은 학술 풍토에 새로운 도전이 필요함을 의미한다. 앞으로 우리나라 미디어 현상과 조응하는 이론 생산 논의가 풍부하게 이어져야 할 것이다.

〈한국언론학보〉는 다양한 연구 주제와 연구 대상, 연구 범위, 이론에 의해 학문적 싹을 틔우고 뿌리를 내려왔다. 1959년 〈한국언론학보〉 전신인 〈신문학보〉 창간사에서 우당 곽복산(1959)은 “커뮤니케이션 매체의 과학적 기술 발달에 따라 그 연구 대상과 범위가 확대”되고 신문 의 자유 문제가 중요하게 대두되고 있어서 씨앗을 뿌리는 정열로” 학문적 결실을 맺는 일이 중요하다고 선언했다. 이후, 언론학은 하나의 학문으로 자리매김하려는 무수한 노력이 뒤따랐다. “학문 세계 내에서 혹은 현실에서 과연 어떤 모습을 보여주었고 어떤 역할을 해왔느냐 하는 문제”는

학문적 정체성 형성과 연결된다(임영호, 1998). 언론학의 연구 대상과 범위, 각 연구영역의 개념적 정의, 정체성에 대한 문제제기를 되풀이하기보다 새로운 관점에서 언론학을 바라보는 용기가 필요한 때가 아닌지 되돌아볼 일이다.

참고문헌

- 곽복산 (1959). 창간사 : 씨앗 뿌리는 정열로, 신문학보를 내면서. <신문학보>, 1권 1호.
- 김규환 (1969). 권두 : 커뮤니케이션 과학연구의 현황과 방향 : 학보 속간에 즈음하여. <신문학보>, 2권 1호, 5-11.
- 양승목 (2009). 언론학 연구 50년 : 성찰과 전망. <한국언론학회 50년사>(1019-1062쪽). 서울: 한국언론학회.
- 오성삼 (2002). <메타분석의 이론과 실제>. 서울: 건국대학교 출판부.
- 이수범 · 문원기 (2016). 정치 광고의 효과에 대한 메타분석. <한국광고홍보학보>, 18권 2호, 182-212.
- 임영호 (2009). 저널리즘 연구 50년의 성찰. <한국언론학회 50년사>(427-454쪽). 서울: 한국언론학회.
- 조동시 (1999). 언론학 총 점검 : 언론관련 학과 현황. <신문과 방송>, 341호, 30-33.
- 정진석 (2009). 한국언론학회 50년: 성장과 변화의 역사(1959~2009)(33-96쪽). <한국언론학회 50년사>. 서울: 한국언론학회.
- 차배근 (2009). 한국언론학회를 일군 사람들: 한국신문학회를 중심으로(97-188쪽). <한국언론학회 50년사>. 서울: 한국언론학회.
- 최선열 (2009). 정치커뮤니케이션 연구 50년. <한국언론학회 50년사>(717-781쪽). 서울: 한국언론학회.
- 최선영 · 고은지 (2017). 미디어 멀티태스킹과 게이미피케이션을 통해 발현한 실시간 참여 민주주의. <한국방송학보>, 31권 3호, 78-113.
- 탁진영 (2004). 부정적 정치광고가 정치 체제에 미치는 부정적 영향에 관한 연구: 메타분석을 이용한 평가를 중심으로. <광고학연구>, 15권 4호, 73-98.
- 한국언론진흥재단 (2014). <한국언론연감>. 서울: 한국언론진흥재단
- 한국언론학회 50년사 편찬위원회(2009). <한국언론학회 50년사>. 서울: 한국언론학회.
- 한국언론학회편집부 (2001). <언론학 교과목 현황 자료집>. 한국언론학회 연구보고서, 1-223.
- 홍주현 (2019). <2000년 이후 한국언론학보 논문을 통해 본 언론학 연구 경향 : 연구주제, 연구방법내용 분석 및 논문키워드 네트워크분석>. 2019년 한국언론학회 창립 60주년 기념 학술대회 대주제세션 발표 논문. 광주: 국립아시아문화전당
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *In Proceedings of the 23rd*

- international conference on Machine learning*, ACM, 113-120.
- Blei, D. M., Carin, L., & Dunson, D.B. (2010). Probabilistic Topic Models. *IEEE Signal Processing Magazine*, 27, 55-65.
- Blei, D. M. (2012). Probabilistic Topic Models(review article). *Communications of the ACM*, 55 (4), 77-84.
- Carr, E. H. (1961/2015). *What is history?*. Penguin UK.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *In Advances in neural information processing systems NIPS*. 288-296.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. *In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100-108.
- O'callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, 42(13), 5645-5657.
- Swanson, D. L. (1993). Fragmentation, the field, and the future. *Journal of Communication*, 43(4), 163-72.
- Wallace, F. M. (1992). *Meta-Analysis: Qualitative Methods for Research Synthesis*. Beverly Hills, CA: Sage Publication.
- Wang, C., Blei, D., & Heckerman, D. (2012). Continuous time dynamic topic models. *arXiv preprint arXiv*, 1206. 3298.

최초 투고일 2019년 6월 7일
게재 확정일 2019년 7월 31일
논문 수정일 2019년 8월 5일

Abstract

Analysis of <Korean Journal of Journalism & Communication Studies> from 1960 to 2018 using Metadata with Dynamic Topic Modeling

SunYoung Choi

Special Appointment Professor, Creative Academy of Eco Science, Ewha Womans University

EunJi Ko

Ph.D. in Engineering, Division of Digital Media, Ewha Womans University

This study analyzed the 60 years of journals, from the first issue to 2018, of the <Korean Journal of Journalism & Communication> with two approaches using the metadata of the complete edition of academic articles. First, the descriptive statistics indicated that the journal has 1,276 authors and 2,048 articles. From 2001, the number of articles has increased by threefold than before, but has been decreasing for the last four years. Next, we used Dynamic Topic Modeling (DTM) to understand the developing pattern of the topics by creating DTM of the stored articles every five years in six categories. The result of the field of the study showed that journalism, online journalism, and news-related journalism were frequently found while the cultural study was rare. Second, in the methodology, empirical study was the dominant topics over time and the computational methodology was increasingly used. Third, frame, usage and gratification, and agenda setting theory were extracted as the main theoretical topics, and fourth, in the message study, the effect related topics were dominant. Fifth, in the audience study, political participation, television viewing and teenager-related topics were extracted. Lastly, in the research subject, online, social media, and internet discussion were actively discussed topics, which indicates that this journal was enthusiastic to the online-related research. This study is notable as it suggests a new methodology to analyze the study trend of journal articles by clustering the extracted topics from an algorithm using metadata and machine learning.

Keywords: Korean Journal of Journalism & Communication Studies, Dynamic Topic Modeling, Probabilistic Topic Models, metadata, sequential data