



# 미세먼지 재해 보도의 프레임 분석

## 구조적 주제모형(Structural Topic Modeling)의 적용\*

**이준웅** 서울대학교 언론정보학과 교수\*\*  
**김성희** 서울대학교 언론정보학과 박사과정\*\*\*

이 연구는 우리나라 중앙일간지의 미세먼지 보도의 특징을 검토하기 위해, 전통적인 내용분석 방법을 사용해서 분석유목을 정해 놓고 뉴스 프레임을 확인한 결과를 ‘자료기반적’ 방법으로 구조적 주제모형(STM)을 적용해서 뉴스 프레임을 추출한 결과와 비교했다. 먼저, 2013년 1월부터 2017년 5월까지 3,636개의 미세먼지 관련 보도를 대상으로 구조적 주제모형을 적용한 결과 30개의 잠재적 주제를 찾아냈는데, 여기에 군집분석을 적용해서 5개의 주제적 뉴스 프레임을 구성했다. 덧붙여 전체 기사 중 300개 미세먼지 기사를 표집해서, 이로부터 재해관련 선행연구에서 검토한 주제적 요소를 분석유목을 삼아 연구자가 직접 코딩하는 방식으로 프레임을 규정했다. 두 가지 프레임 분석 결과를 비교해서 ‘자료기반적’ 뉴스 프레임 추출 방법의 가능성을 검토했다. 최종적으로 구조적 주제모형을 적용한 뉴스 프레임을 이용해서 미세먼지 보도에 나타나는 주제적 프레임의 분포가 언론사에 따라, 그리고 시기별로 다르게 나타났음을 확인했다. 연구결과에 기초해서 한국 언론의 미세먼지 재해 보도가 보이는 ‘파편화된 보도’ 현상의 함의를 논의했다.

**핵심어:** 미세먼지, 재난, 위험 인식, 뉴스 프레임, 구조적 주제모형, 토픽모델링

---

\* 이 연구는 서울대학교 언론정보연구소의 연구자원을 일부 받았음을 밝힙니다. 자료분석 초기단계에서 연구방법론과 관련하여 도움을 주신 백영민 교수에게 감사를 표합니다.

\*\* jwrhee@snu.ac.kr(교신저자)

\*\*\* mynameis.sungheekim@gmail.com

## 1. 문제제기

미세먼지에 대한 경각심이 높다. 이와 관련해 우리가 주목하는 현상은 이와 관련해 우리가 주목하는 현상은 시민들의 미세먼지 위험성에 대한 인식은 높는데 비해, 미세먼지의 원인과 대책에 대한 인식은 단순하고 정형화돼 보인다는 것이다. 예컨대, 많은 이들이 미세먼지의 원인으로 '중국 등 주변국 영향을 지목하는 경향이 있고, 이에 대한 대책으로 '마스크 착용'이나 '외출 자제'를 꼽는다(서울환경연합, 2017; 한국 궐립, 2017). 미세먼지 현상 자체의 복합성과 영향력 관계를 둘러싼 인과관계의 복잡성을 고려할 때, 시민들이 이렇듯 단순하고 정형화된 방식으로 미세먼지를 이해하고 대처한다는 사실은 우려할 만하다. 그렇다면 미세먼지의 위험성, 원인, 대책 등에 대한 시민의 인식은 왜 이렇게 정형화되고, 심지어 과편화됐을까? 우리는 언론이 미세먼지를 다루는 방식을 탐구함으로써 이 질문에 대한 한 답변을 제시하려 한다.

언론의 보도가 시민들의 재난 재해에 대한 인식에 미치는 영향을 탐색한 선행 연구들이 있다. 예컨대, 키칭거(Kitzinger, 1999)는 언론 매체가 잠재적 위험 상황을 선정적으로 보도하고 위험 요소의 위험성을 과장하는 경향이 있는데, 이 때문에 공중의 불안과 염려가 커질 수 있다고 경고한 바 있다. 고츠(Goltz, 1984)와 프리드먼과 뱅거(Friedman & Wenger, 1986)는 재난 재해에 대한 보도에 나타나는 갈등 프레임(conflict frame)이 재난 재해에 대한 불합리한 신념, 즉 신화(myths)를 만들어 낸다고 비판했다. 언론이 재난 재해 상황에서 부분적으로 일어나는 집단 간 충돌이나, 약탈이나 폭력과 같은 반사회적 행동을 중심으로 보도하면서, 사실과 다른 부정적인 인식을 형성하기도 한다는 것이다. 결국 미세먼지에 대한 시민들의 인식을 이해하기 위해서는 먼저 언론 매체가 미세먼지를 어떤 관점에서 어떻게 프레임해서 다루는지 검토할 필요가 있다.

과거 재난 재해 보도의 프레임 분석결과를 보면, 프레임의 주제적 특징을 전제하고 그 특징을 반영하는 텍스트의 성질을 찾아서 확인하는 방식의 내용분석 방법을 적용한 것이 주를 이룬다. 전통적인 내용분석 방법을 이용해서 뉴스의 주제적 특징을 뽑아낸 후, 그것을 프레임이라고 부르는 경우가 대부분이다. 예를 들어, '인간 흥미 프레임(human interest frame)', '갈등 프레임(conflict frame)', '위험 인식 프레임(risk perception frame)' 등과 같은 이름을 달고 있는 프레임들은 실제로는 뉴스의 주제적 특징의 집합을 뜻한다. 이런 프레임은 연구자가 개별 뉴스를 읽고 특정 주제적 속성들에 해당하는지 확인하는 방법, 즉 내용분석 방법을 적용해서 특정 프레임이라고 이름붙인 것이다.

그러나 전통적 내용분석 방법을 사용한 프레임 추출에는 몇 가지 문제점이 있다. 우선 미리

정한 분석틀에 의존해서 프레임을 추출하기 때문에 새로운 종류의 프레임을 찾을 가능성이 낮다. 선행연구가 찾은 프레임은 재확인하는 ‘확증편향’을 범할 경향이 있다는 것이다. 또한 프레임 추출에서 연구자의 주관적인 해석에 크게 의존하기 때문에 코딩의 불안정성이 우려되고, 이 때문에 연구결과의 재현성이 낮다. 특히 단순히 특정 프레임이 강조되는지 아닌지를 탐지하는 수준이 아니라 개별 프레임의 의미론적 구성에 주목할 경우, 개별 연구 간 일관성을 유지하기 어렵다. 이 연구는 이런 문제점을 보완하기 위한 방법으로 알고리즘을 이용한 내용분석 방법 중 하나인 구조적 주제모형(structural topic modeling) 방법을 활용했다.

주제모형은 개별 텍스트에 등장하는 단어들의 출현에 영향을 미치는 전체 텍스트에 내재하는 잠재적 주제를 생성하는 텍스트 모형화 방법의 일종이다. 이 방법을 이용해서 개별 문서, 즉 개별 뉴스가 포함하는 단어들의 출현에 영향을 미치는 잠재적 주제들을 도출할 수 있다. 대규모 말뭉치(corpus)로부터 복수의 잠재적 주제들을 ‘자료기반 방법(data-driven method)’으로 추출한 후, 잠재적 주제들이 군집화하는 패턴을 찾아냄으로써 해당 말뭉치에 속한 전체 텍스트를 관통하는 일종의 주제적 프레임을 뽑아낼 수 있다는 것이다. 단 우리는 자료기반 방법을 적용해서 얻은 뉴스 프레임이 과연 타당한지 검토하기 위해서, 전통적인 내용분석 방법을 적용한 프레임 추출 결과도 참조할 필요가 있다고 보았다. 왜냐하면 아무리 알고리즘에 기초해서 텍스트로부터 프레임을 추출한다고 해도 개별 프레임의 의미를 파악하기 위해서는 같은 주제를 탐구한 선행 연구의 성과에 기초해서 주제에 대한 해석 작업이 필요하기 때문이다.

이 연구에서 우리는 과거 프레임 분석이 갖는 방법론 상의 제한점들을 보완하기 위해 전통적인 내용분석을 적용한 분석방법과 자료기반 분석방법을 병행해서 사용했다. 이 연구의 목적은 결국 언론이 미세먼지에 대한 보도를 프레임 분석을 통해 검토함으로써 미세먼지 관련 재난 인식을 형성하는 과정에 언론사별 그리고 시기별로 어떤 주제적 프레임이 형성되어 작동했는지 검토하는 데 있다. 특히 언론이 미세먼지와 같이 복잡한 현상을 어떻게 정형화된 주제로 환원해서 다루는지 탐구하고자 한다. 이를 통해 개념기반 방법으로 프레임을 확인하는 결과와 더불어 자료기반 방법으로 프레임을 추출한 결과를 동시에 구해서 비교분석함으로써 미세먼지에 대한 재해 인식의 형성과 같은 복잡한 사회인지적 과정에 미치는 언론의 역할을 검토할 수 있으리라 기대한다.

## 2. 기존 논의와 연구문제의 도출

### 1) 미세먼지에 대한 공중의 인식

김영옥과 동료들(2015a; 2015b; 2016)은 미세먼지 재해에 대한 공중 인식을 지속적으로 연구한 대표 연구자다. 이들은 미세먼지 위험보도의 주요 정보원에 따른 프레임 구성을 분석했으며, 미세먼지에 대한 전문가와 공중의 인식 간에 차이를 검토해서 일반 공중이 미세먼지 발생원인, 영향, 그리고 대처 방안 등 간의 유기적 관련성을 파악하지 못하는 일종의 ‘인지적 과편화 현상’을 보인다는 관찰을 제시했다. 한혁(2017)은 ‘공중상황이론’을 미세먼지 재해 사례에 적용하여, 공중의 특성에 따라 미세먼지에 대한 위험 인식이 공중의 특성에 따라 다르게 나타난다는 것을 밝히고, 이에 맞추어 필요한 정보를 서로 다른 공중들에게 제공해야 한다고 논했다. 이런 연구들은 커뮤니케이션 상 미세먼지 재해가 가지는 특질을 파악하고, 실용적인 대안을 제안한다는 성과를 보였다.

티어니(Tierney, 2007)는 재해에 대한 사회과학 연구의 문제점을 비판하면서, 재해 인식과 그에 대한 집단적 대처행위의 복잡성에 주목하지 않고 ‘기존 관행을 답습하는 개념화’가 이루어진다고 지적한 바 있다. 재난 및 재해의 사회적 연관성에 대한 섬세한 이론화가 필요하다는 것이다. 재해 및 재해를 측정하는 것도 그렇지만, 그에 대한 인식을 탐구하는 것도 마찬가지로 어렵다. 일반적으로 재난은 자연적 혹은 인공적 원인에 의해 인명 그리고 재산 상 피해를 입히는 사건을 뜻하고(Gilbert, 2005; Quarantelli, 1985), 위험인식은 해당 ‘위험의 정도에 대한 인식’을 가리킨다(Roeser, et al., 2012). 그러나 ‘위험의 정도’에 초점을 맞추는 위험 인식은 다양한 재난 재해 상황과 그로 인한 후속적 효과를 제한해서 파악할 우려가 있다. 예컨대, 가시적인 형태로 대형 인명 피해를 유발하지 않더라도 장기적으로 위험을 유발하는 기후변화 또는 미세먼지와 같은 재난이 있다. 이런 재난에 대한 인식은 지진이나 홍수와 같은 직접적이고 가시적인 재난에 대한 인식과 다를 수밖에 없다. 위험 인식은 위험 상황이나 요소가 어떻게 위협하며, 어떤 영향을 미치는지, 그리고 어떻게 일어나는지 종합적인 의미화(sense-making)를 통해서 이루어진다(Gephart, Jr, 2004).

결국 재난 및 재해에 대한 위험 인식이 복잡한 과정을 거쳐서 형성하며, 이 과정에서 언론의 보도가 중요한 역할을 수행할 수 있다는 데 주의해야 한다. 미세먼지를 둘러싼 복잡한 사건과 사안들을 두고 언론이 주제화(thematization)하고 서사적 이야기구성(narrative construction)을 적용하는 방식에 따라서 공중의 미세먼지의 발생에 대한 이해는 물론 재해의 종류와 규모에 대한 인식, 그리고 어떻게 대처해야 하는지 대응방안 등이 달라질 수 있기 때문이

다. 따라서 위험 인식에 영향을 미치는 재난 및 재해에 대한 언론의 보도 방식, 특히 뉴스 프레임 구성을 검토할 필요가 있다.

## 2) 재난 보도와 뉴스 프레임

재난 인식 그리고 위험 인식의 원인 중의 하나로 뉴스가 사건을 구성하는 방식, 즉 뉴스 프레임(news frames)이 있다. 뉴스 프레임은 간단히 말해 뉴스 텍스트의 구성적 특성 중 하나라 할 수 있다. 그런데 프레임은 단순히 언론의 보도 경향성 차원을 넘어서 뉴스를 이용하는 시민들의 사건에 대한 신념을 형성하는 원인 중 하나라고 한다. 엔트먼(Entman, 1993, 2007)과 아엔가(Iyengar, 1990) 등 뉴스 프레임 연구의 초기 연구자들은 뉴스 프레임의 이런 인지적 효과에 주목했다. 즉 이들은 뉴스 프레임이 뉴스 이용자의 인식과 판단에 미치는 효과를 염두에 두고 뉴스 텍스트를 분석하고 또한 재구성하기 위해 프레임 개념을 사용했다.

뉴스 프레임은 결국 뉴스의 텍스트적 속성의 집합이지만, 그중에서도 뉴스 이용자에 영향을 미칠 것으로 가정되는 특성의 집합이다. 대표적으로 아엔가(Iyengar, 1990)는 뉴스 프레임에 따라 뉴스 이용자의 상황의 원인에 대한 인식이 달라진다는 연구결과를 제시한 바 있다. 예컨대, 빈곤 문제를 전체적인 사회적 맥락에서 주제화해서 다루는 뉴스를 본 이용자들은 빈곤의 원인이 사회 시스템에 있다고 여기게 되는 반면, 개인의 일화에 초점을 맞춘 뉴스를 본 이용자들은 원인을 개인적 책임으로 돌린다는 것이다. 엔트먼(Entman, 1993)은 복잡한 사안을 다루는 뉴스는 프레임을 통해 문제를 정의하며, 원인을 진단하고, 도덕적 판단의 기준을 제공하고, 결국 대책도 제시한다고 보았다. 뉴스 프레임은 단순히 사안의 중요도에 대한 인식에 영향을 줄 뿐만 아니라, 사안의 종합적 이해와 인식을 형성하는 데 기여한다는 것이다. 뉴스 프레임의 구성과 효과에 대한 연구들은 재난 재해 보도에 대해서도 적용할 수 있다. 지금까지 재난 재해 보도와 공중의 재난 인식에 대한 연구는 크게 다음 두 가지로 볼 수 있는데, 모두 뉴스 프레임의 효과로 재구성하는 것이 가능하다.

첫째, 언론은 재난 및 재해의 위험성을 과장하는 일종의 선정보도의 경향을 보이는 데, 이런 선정보도를 뉴스 프레임 구성으로 설명할 수 있다. 키칭거(Kitzinger, 1999)와 허스튼과 동료들(Houston, et al., 2012)은 언론이 뉴스 이용자의 주목을 끌기 위해 위험성을 강조하는 경향이 있다고 경고했다. 재난 재해나 위험 상황의 잠재적인 위험성이 크며, 그것이 '항시적'이라는 인상을 줌으로써 위험성을 강조한다는 것이다. 키칭거는 언론이 재난 위험성 강조하는 이유가 뉴스의 이야기로서 호소력(narrative appeal)을 강조하면서 발생한다고 본다. 또는 비유나 표현을 이용하기도 한다. 예컨대, 대장균을 '살을 뜯어먹는 벌레(flesh-eating bug)'로 표현함으로

써 위험 요인의 위해성을 말초적으로 전달하기도 한다(Pennington, 1999).

둘째, 언론의 재난 및 재해보도는 복잡한 사태를 전체적인 맥락에서 다루지 않고 그때그때 발생하는 사건에 초점을 맞춰 보도함으로써 사태를 과편화하는 경향이 있다(Houston et al., 2012). 베넷(Bennett, 2004)에 따르면 언론은 일반적으로 '사건에 대한 전략적으로 선택한 판본'을 뉴스로 가공해서 전달하는 경향이 있다. 즉 사건을 형성하는 역사적이고 다층적인 맥락을 제거하고, 사건에 등장하는 인물을 단순화해서 일종의 드라마처럼 제시하는 경향이 있다. 이런 '뉴스의 과편화(fragmentation)' 현상은 원인과 대책에 대한 보도에서도 나타나며, 이는 재해의 책임인식을 특정 방향으로 유도한다. 많은 연구자들(Hanson & Hanson, 2006; Holton et al., 2012; Kitzinger, 1999; Reese & Lewis, 2009)이 재난 보도에서 '비난 프레임(blame frame)'의 역할에 주목하는 까닭이 여기에 있다. 재난의 책임을 특정 집단에 돌림으로써 대책도 특정한 방향으로 유도하기 때문이다. 예컨대, 티어니와 동료들(Tierney et al., 2006) 그리고 다인즈와 로드리게즈(Dynes & Rodriguez, 2010)은 2005년 발생하여 수 천 명의 사상자와 이재민을 낳은 허리케인 카트리나(Katrina)에 대한 뉴스가 일종의 '재난 미신(disaster myth)'을 형성했다고 보았다. 재해과 함께 벌어진 많은 사건들이 있었지만, 언론은 집단 간 갈등이나 강력 범죄, 약탈 등의 사회적 문제에 초점을 맞춰 보도함으로써 허리케인으로 인한 피해보다 사회적 혼란과 반사회적 집단행동이 문제라는 식의 인식을 형성했다고 한다.

결국 미세먼지와 같은 복잡한 사태에 대한 뉴스의 의미를 파악하기 위해서 정교한 뉴스 프레임 분석이 필요하다. 특히 언론이 미세먼지 원인을 규정하고, 위험 대책을 제안하는 방식은 어떠한 뉴스 프레임 분석을 통해 드러낼 필요가 있다. 예컨대, 통제가 어려운 생태계적 요인이나 국제적 요인을 강조하는 미세먼지 뉴스는 국내의 에너지 정책이나 생활 미세먼지의 해악에 집중하는 뉴스와 주제화 방식이 다를 것을 예상할 수 있다. 미세먼지와 같이 장기간에 걸쳐 발달하고 다양한 요인들과 복잡하게 연관된 재난에 대한 뉴스는 위험성에 대한 정보는 물론 재해의 원인을 규정하는 방식에 따라, 그리고 대책을 제시하는 방법에 따라 서로 다른 하위 주제들을 선택할 것이고, 결국 서로 다른 뉴스 프레임으로 나타날 것이다.

### 3) 연구문제의 도출: 자료기반 프레임 추출의 필요성

우리는 이 연구에서 미세먼지 뉴스가 다른 뉴스와 마찬가지로 다층적인 의미의 층위를 갖지만(이준웅, 2009), 그 중에서 이야기구조에 주목해서 미세먼지 프레임이 (가) 미세먼지의 '위험성'을 강조하는 방식에 대한 내용, (나) 미세먼지의 '원인'을 규정하는 내용, 그리고 (다) 미세먼지 '대책'에 대한 내용을 주제적 특성으로 갖게 될 것이라 전제하고 연구를 진행한다. 이 전제는 프

레이미 연구의 방법론에 대한 논의를 참조하면서 동시에 재난 및 재해 보도에 대한 선행 연구결과를 종합해서 얻은 것이다. 선행연구에서 규정한 프레임의 적용해서 프레임을 확인하는 작업을 할 때, 다음 두 가지 점에 주의할 필요가 있다.

첫째, 선행 연구결과를 이용해서 뉴스 프레임을 분석하는 경우에는 과거 뉴스 프레임을 재확인하는 경향을 보인다. 예컨대, 김영욱 등(2015)은 미세먼지 보도에 대한 프레임 분석에서 이른바 ‘형식적 프레임’에 속하는 주제적 프레임과 일화적 프레임을 발견했고, 이른바 ‘내용적 프레임’으로 단순정보제공 프레임, 대응 프레임, 원인귀인 등을 발견했다. 이런 발견은 어느 정도 예견된 결과라고 할 수 있는데, 그 이유는 과거 프레임 분석연구들(김원용·이동훈, 2005; 이민규·이예리, 2012; Iyengar, 1991; Semetko & Valkenburg, 2000)이 분석에 사용했던 주제적 프레임들을 분석 유목으로 활용함으로써 과거 연구자가 발견했던 프레임을 재발견한 격이라고 할 수 있기 때문이다. 선행 연구의 분석유목을 사용한 연구는 발견해야 할 프레임을 결국 발견하는, 이른바 확증편향적 경향을 보인다.

둘째, 프레임 확인해서 확증편향을 피하기 위해서는 자료기반적인 귀납적 방법을 사용해서 뉴스 프레임을 상향적으로 구성해야 하는데, 여기에도 문제가 없지 않다. 귀납적 방법을 적용해서 얻은 새로운 프레임을 발견하더라도, 새로운 프레임이 과연 의미적 차별성을 갖는지 아니면 과거 선행연구에서 제시한 프레임과 사실상 같은 것인지 변별할 근거가 없기 때문이다. 요컨대, 새로운 ‘자료기반 방법’을 사용해서 얻은 프레임만 갖고는 선행 연구결과와 프레임의 의미적 차별성을 주장하기 어렵다. 같은 자료를 놓고 분석한 결과의 차별성을 주장할 비교대상이 없기 때문이다. 결국 동일한 뉴스 텍스트를 자료로 삼아 두 자료기반 방법과 개념기반 방법을 적용한 결과를 얻고 비교하는 방식으로 의미적 차별성을 확인할 필요가 있다.

이 연구에서 우리는 먼저 미세먼지 보도에 나타나는 뉴스 프레임이 과연 어떤 주제적 특성을 보이는지 선형적으로 규정하지 않고, 자료기반 방법을 적용해서 뉴스 프레임을 추출해 보겠다. 선행 연구결과가 발견한 뉴스 프레임을 분석유목으로 삼아서 프레임을 재확인하는 방식으로는 그 방법론이 아무리 타당하고 적절하다고 평가를 받을지라도, 결국은 기존 연구에서 확인했던 뉴스 프레임을 재발견하는 확증편향적 결과가 될 우려가 있기 때문이다. 그러나 자료기반 방법만으로 뉴스 프레임 추출을 마치는 것도 문제가 된다. 자료기반 방법을 적용해서 새롭게 찾은 뉴스 프레임이 어떤 차별적 의미를 갖는지 검토할 방법이 없기 때문이다. 자료기반 방법을 적용해서 도출한 뉴스 프레임이 과거 개념기반 방법을 사용해서 도출한 뉴스 프레임과 의미적으로 일치하는지 알아보기 위한 별도의 비교분석이 필요하다. 이 연구의 일차적 목적은 바로 이런 연구관심에 답하는 데 있다. 따라서 우리는 다음과 같이 연구문제를 설정한다.

**연구문제 1.** 자료기반 방법을 적용해서 미세먼지에 대한 우리나라 언론의 보도 프레임을 추출한 결과는, 미세먼지로 인한 위험성, 원인, 대책 등 주제를 분석유목으로 설정해서 프레임을 추출한 결과와 어떻게 다른가?

우리는 또한 자료기반 방법을 적용해서 도출한 언론보도의 미세먼지 프레임이 언론사에 따라서, 그리고 시점에 따라서 어떤 분포를 보이는지 탐구하고자 한다. 프레임 분석을 수행한 무수한 선행 연구들이 언론사의 이념적 경향성, 언론사의 매체적 특성, 그리고 언론사의 편집 방침 등에 따라서 같은 사안이나 사건에 대한 뉴스 프레임이 달라진다는 연구결과를 제시한 바 있다. 또한 뉴스 프레임의 형성은 역동적인 것이어서, 시점의 변화에 따라 뉴스 프레임 간의 강조점이 달라진다는 연구가 적지 않다. 우리는 자료기반 방법을 적용해서 추출한 뉴스 프레임의 언론사별 분포와 시기별 변화 양상도 유사한 결과를 낼 것이라고 기대한다. 자료기반 방법으로 추출한 뉴스 프레임의 언론사별 분포와 시기별 변화가 한국 언론의 미세먼지 보도에 대해 어떤 특징적 양상을 보일지 탐색하기 위해 우리는 다음과 같은 두 번째 연구문제를 설정한다.

**연구문제 2-1.** 우리나라 언론의 미세먼지 보도의 프레임이 강조되는 양상은 언론사에 따라 달라지는가?

**연구문제 2-2.** 언론사에 따라 달리 나타나는 프레임의 강조 양상은 시점에 따라 달라지는가?

### 3. 연구 방법

#### 1) 자료 수집

연구 대상으로 2013년 1월 1일부터 2017년 5월 30일까지 주요 일간지의 미세먼지 관련 기사를 수집했다. <2017년 언론 수용자 의식 조사>가 제시한 열독률 결과를 기준으로 중앙일보, 조선일보, 동아일보, 경향신문, 한겨레, 한국일보 등 6개 종합 일간지를 분석대상 신문으로 선택했다. 한국언론진흥재단의 데이터베이스인 <빅카인즈>의 연도 별 기사 건수를 검토한 결과, 2013년부터 미세먼지 관련 기사가 급격히 증가한다는 사실을 발견했다. 언론이 2013년 이후 미세먼지에 대해 관심을 가진 이유는 실제 한국의 미세먼지 농도가 이 때부터 높아졌기 때문인 것으로 보인다. 미국의 보건효과연구소 (Health Effects Institute)가 발표한 ‘세계기후상태(State of Global Air)’ 보고에 따르면, 한국의 ‘초미세먼지 농도에 대한 인구의 가중 평균’은 2010년 이후



소폭 감소했다가 2013년 이후부터 크게 증가했다. 우리는 2013년 1월 1일을 분석의 기점으로 삼고 2017년 5월 31일까지 미세먼지 관련 기사를 분석하기로 결정했다.

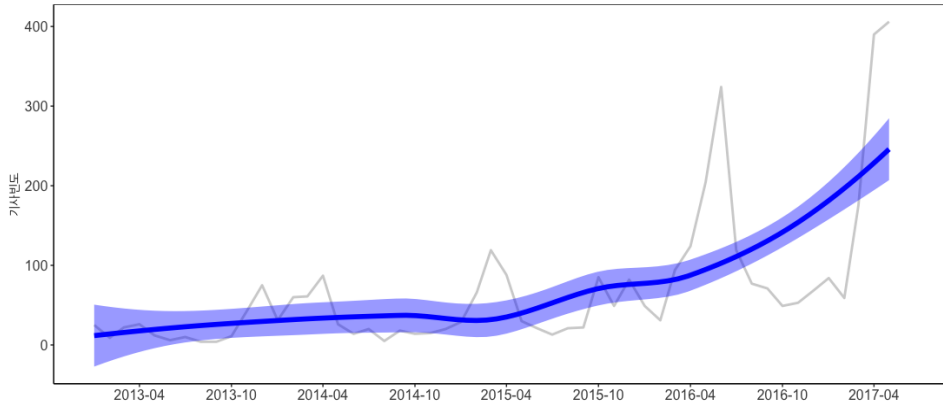


그림 1. 6개 종합일간지 미세먼지 관련 기사 건수

자료 수집을 위해 각 신문에서 연구대상 기간에 발행한 미세먼지 관련 기사들을 언론사 사이트 별로 ‘웹스크래핑(web scrapping)’ 도구를 활용해서 수집했다. 크롤링을 이용해서 일괄 수집하지 않은 이유는 포털 사이트에서 미세먼지 관련 기사 검색 결과, 크롤링으로 수집했을 때 신문 기사가 제외된 경우가 발생하는 것은 물론, 신문사 사이트 별로 포맷이 각각 달라 일관된 방법으로 자료를 정리하기 어려웠기 때문이다. 웹 스크래핑 도구인 옥토퍼서와 웹스크래퍼를 이용해서 각 언론사 홈페이지에 키워드로 ‘미세먼지’를 검색해서 등장한 모든 기사들을 수집했다. 이때 정치, 경제, 사회 면 기사로 지면을 한정했다. 그 다음, 단신이나 단순 일기예보, 중복 기사들을 분석대상에 제외하는 데이터 청소 과정을 거쳤다. 두 단계를 거쳐서 최종적으로 분석에 포함된 기사는 ‘조선일보’ 357건 (9.82%), ‘중앙일보’ 1056건(29%), ‘동아일보’ 670건(18.43%), ‘한국일보’ 312건(8.58%), ‘경향신문’ 818건(22.5%), ‘한국일보’ 423건(11.63%) 등 총 3,636 건이었다.

〈그림 1〉은 연도에 따른 6개 종합일간지 미세먼지 관련 기사의 분포를 제시한다. 두 가지 특징을 지적할 수 있다. 첫째, 전체 기사 건수는 2013년 1월부터 2017년 5월까지 상승하는 추세를 보인다. 이 추세는 국소가중회귀(LOESS: locally weighted least square regression) 방식으로 제시한 추세선에서 확인할 수 있다. 둘째, 추세를 벗어나 돌출한 봉우리들이 있는데, 이는 2013년부터 2015년까지는 겨울철에 집중되어 있다가, 2016년 이후에는 봄철에 나타나는

양상을 보인다. 이는 미세먼지 보도가 사회적 사건과 맥락에 따라 집중적으로 나타나는 경향을 보임을 의미한다. 구체적으로, 2013년 10월의 봉우리는 국제보건기구(WHO) 산하 국제암연구소가 미세먼지를 1군 발암물질로 지정한 시점이고, 2014년 2월과 4월은 각각 미세먼지 전국예보를 실시하고 정보발령시 휴교 및 차량부제를 시행을 도입했던 시점이다. 2015년 4월에는 국내 외 미세먼지가 초래하는 해악성에 대한 연구결과를 언론이 강조해서 보도했으며, 2015년 9월은 폭스바겐 배출가스 조작과 관련한 뉴스가 주목을 받았다. 2016년 6월에는 차량부제와 경유차 규제 등 6.3 미세먼지관리 특별대책을 시행한 때이다.

자료의 내용적 특성을 기능하기 위해, 분석에 사용한 전체 3,636건 기사에 등장한 명사의 빈도를 <표 1>과 같이 구했다. ‘미세먼지’를 키워드로 추출한 기사 집합이기 때문에, 기사 전체를 말뭉치(corpus)로 만들어 사전처리(pre-processing)하는 가운데 ‘미세먼지’란 명사는 삭제했다. 결국 ‘미세먼지’라는 단어를 제외하면, 말뭉치에서 가장 많이 등장한 단어는 ‘대통령(2,466건)’이었으며, 그 뒤로 ‘환경부(2,398건)’, ‘마스크(2,279건)’, ‘서울시(1,974건)’, ‘경유차(1,895건)’, ‘자동차(1,860건)’, ‘초미세먼지(1,849건)’ 등 순서다. 미세먼지 관련한 용어들이 기후, 환경, 건강 등 사안에 제한되지 않고, 정치와 사회를 포함한 전 영역을 포괄함을 확인할 수 있다.

표 1. 단어 빈도 분석표(요약)

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	대통령	2,466	20	한국	937	39	중금속	516
2	환경부	2,398	21	알레르기	844	40	배출가스	510
3	마스크	2,279	22	전문가	807	41	대기환경	494
4	서울시	1,974	23	우리나라	769	42	소비자	488
5	경유차	1,895	24	석탄화력발전소	755	43	고농도	487
6	자동차	1,860	25	배출량	746	44	고등어	465
7	초미세먼지	1,849	26	환경과학	713	45	비타민	464
8	지난해	1,822	27	화력발전소	710	46	온실가스	462
9	수도권	1,705	28	악순환	699	47	적폐청산	453
10	중국	1,804	29	전병헌	671	48	보간기구	452
11	에너지	1,614	30	이진구	670	49	에어컨	450
12	대기오염	1,973	31	어린이	654	50	국론분열	443
13	스모그	1,593	32	가능성	647	51	마이크로	421
14	호흡기	1,337	33	지난달	624	52	배기가스	412
15	공기청정기	1,208	34	질소산화물	562	53	기후변화	401
16	주의보	1,061	35	베스트	551	54	서비스	394
17	베이징	1,028	36	기관지	550	55	유해물질	371
18	관계자	995	37	친환경	533	56	자외선	346
19	오염물질	965	38	시스템	526	57	발암물질	338

## 2) 자료분석 전략: 알고리즘 코딩과 인간코딩의 비교

신문기사의 내용을 분석하기 위해 우리는 구조적 주제모형(STM: structural topic modeling) 방법을 사용했다. 주제모형 추출방법이란 대량의 개별 문서들을 가로질러 잠재적으로 존재한다고 가정할 수 있는 복수의 주제들(topics)을 베이지 추정방식으로 도출하는 방법이다. 이 기법을 이용하면, 말뭉치에 속한 모든 문서를 구성하는 단어들의 분포를 이용해서 개별 문서를 형성하는 데 기여하는 잠재적 주제를 도출할 수 있다(Blei & Lafferty, 2007; Blei, Ng, & Jordan, 2003). 따라서 이는 연구자의 분석의도에 따른 주관성을 상당부분 통제할 수 있다는 장점이 있다. 왜냐하면 미리 분석유무를 정해 놓고 그에 따라 자료에서 해당하는 내용을 연구자가 판단해서 값을 결정하는, 전통적인 인간코딩, 즉 분석유무를 사용하는 내용분석과는 달리, 문서 자료에 내재한 단어의 분포만을 활용해서 베이지 알고리즘을 이용해서 주제를 도출하기 때문이다. 특히 구조적 주제모형(STM)은 주제모형을 확인하는 방법들 중에서도 분석대상 텍스트에서 추출한 잠재적 주제 간에 상관성을 인정하는 분석기법이다(Roberts, Stewart, & Airoldi, 2016).

문서 분석에 주제모형 추출방법을 적용한 연구들은 점차 증가하고 있으며, 이런 연구들이 산출한 연구결과의 방법론적 타당성도 대체로 인정받고 있다(문안나, 2016; 백영민·최문호·장지연, 2014; 진실아 등, 2013; DiMaggio, Nage, & Blei, 2013; Roberts et al., 2014). 그러나 주제모형 방법을 비판적으로 검토한 연구에 따르면, 이런 연구들이 상당한 사전처리(pre-processing)을 하지 않는 한 해석하기 어려운 결과를 보인다는지(Hong & Davison, 2010), 또는 결과를 낱더라도 실재하지 않는 주제를 잡아내는 것과 같은 일종의 ‘거짓 양성(false positive)’ 결과를 산출하는 등(Guo et al., 2016) 문제점이 없지 않다고 한다. 그리머와 스튜어트(Grimmer & Stewart, 2013)가 주장했듯이, 자동화된 문서자료 분석방법은 아무리 편리하고 효율적이라고 해도 나름대로 잘못된 점이 있으며, 결코 완전히 인간코딩을 대체할 수는 없다. 우리는 이 연구에서 자동화된 ‘주제추출 알고리즘’을 사용한 결과만 제시하는 것은 분석적 장점이 없다고 보았다. 이 방법을 적용해서 뉴스 프레임 집합을 얻는다 하더라도, 그 의미를 비교해 볼 대상이 없기 때문이다. 결국 구조적 주제모형을 적용한 뉴스 프레임 추출 결과의 의미적 타당성을 확인하기 위해서라도 전통적인 내용분석을 이용한 프레임 추출방법과 병행하는 것이 바람직하다고 보았다.

우리는 알고리즘 코딩과 인간코딩을 병행하더라도 두 방법을 사용한 결과가 상호 수렴할 때까지 양자의 방법론을 무한정 조절하는 방식으로 분석할 수는 없을 것이라고 생각했다. 인간코딩의 결과를 기준으로 삼아서 알고리즘 코딩 결과를 조정하는 방법을 사용하거나,<sup>1)</sup> 아니면 반대로

알고리즘 코딩 결과를 기준으로 삼아서 인간코딩의 분석유목을 조정하는 방법은 각자 나름대로 문제가 있다. 왜냐하면 첫째 이 두 방법 중에 어떤 것이 참 값인지 사전 또는 사후에 결정할 수 있는 최종적인 결정 방법이 없고, 둘째 두 방법 중에 한 쪽이 일견 타당한 결과를 산출하더라도 다른 쪽의 결과를 무시한다면, 이는 사실상 자료를 선택적으로 버리는 일과 다름없기 때문이다.

이상의 방법론적 요점들을 고려하여, 우리는 다음 3단계로 연구를 수행했다. 첫째, 분석대상으로 삼은 전체 3,636건의 뉴스에서 300건의 기사를 체계적으로 표집한 후, 특정 기사에 특정 주제변수가 해당하는지 확인하는 내용분석 방식으로 인간코딩을 해서 '프레임 분석'을 먼저 수행했다. 이 과정에서 기존 위험 및 재난 연구들이 주목했던 (가) 위험성, (나) 원인, (다) 대책 등 주제적 구성요소를 주제변수라 하고, 이를 프레임 구성요소(frame elements)로 삼아 분석유목으로 삼는 등 선행연구 결과를 활용한 개념기반 방법으로 미세먼지 보도를 분석하기 위해 노력했다. 둘째, 인간코딩 프레임 분석과 완전히 독립적인 방식으로 3,636건을 기사를 대상으로 구조적 주제모형 분석을 수행함으로써 미세먼지 기사들을 관통하는 주제를 도출했다. 셋째, 프레임 분석결과와 구조적 주제모형 분석결과를 비교함으로써 각 방법을 사용한 결과의 차이가 함의하는 바를 고려해서 후속적인 신문사별 비교분석과 시계열 분석으로 진행했다. 이렇게 다중적 방법을 사용해서 연구를 수행함으로써 해석적 지침 없이 구조적 주제모형 분석결과를 기계적으로 도출한 후 결과를 검토하는 문제를 일정 정도 극복할 수 있을 것으로 기대했다.

### 3) 선행연구 기반 프레임 분석유목의 설정

우리가 설정한 연구 전략에 따르면 구조적 주제모형을 이용해 프레임을 도출하기 전에, 선행연구에서 이론적으로 도출한 변수들을 분석유목으로 삼아서 인간코딩으로 프레임 분석을 수행하는

---

1) 분석에 동반하는 일련의 선택에 따라 알고리즘을 이용한 기계적 코딩결과도 일반적으로 미세하게, 때로는 크게 달라질 수 있다는 데 주의해야 한다. 첫째, 자료를 사전처리할 때 '비사용 용어'의 집합을 결정하거나, '동의어 사전'을 이용해서 사전처리를 하거나, 명사를 대상으로만 분석을 하는 것과 같이 문서의 구성요소 중 일부로 제한하는 결정 등과 같은 선택은 모두 분석결과에 영향을 미친다. 둘째, 주제 모형을 구성하기 위해 투입하는 '문서-단어행렬(DTM)'의 구성방법에 따라서, 그리고 '문서-단어행렬'의 희소성(sparsity) 또는 조밀성 조절수준에 따라서 주제모형 결과가 달라진다. 셋째, 주제모형 추출 방법론의 종류에 따라서 결과가 미세하게 달라진다. 예를 들어, 문서를 가로지르는 주제 간 상관관계를 인정하지 않는 일반적인 잠재적 디리실레 할당(LDA) 주제모형 추출방법과 주제 간 상관관계를 인정하는 구조적 주제모형(STM) 추출방법은 잠재적 주제를 생성하는 방법은 세부적 모형 구성방법이 다르다. 마지막으로, 분석할 때 부여해야 하는 일련의 메타모수들(meta-parameters)을 설정한 값에 따라라도 분석결과가 달라질 수 있다. 예를 들어, 잠재적 주제의 수를 설정하는 일이 그 중 하나다. 우리 연구에서는 구조적 주제모형 저자가 제시한 모형진단을 위한 변수들, 잠재적 주제 수에 따라 도출된 모형의 로그우도, 의미적 일관성, 잔차 등을 검토해서 최적의 모형을 선택하는 방법을 따랐다.

일을 수행해야 한다. 우리는 슬로빅(Slovic, 1987, 1990), 세베리(Sjöberg, 2000), 바움과 플레밍(Baum & Fleming, 1983), 브루어와 그의 동료들(Brewer et al., 2007) 등 기존 연구 결과를 참조해서 위험 및 재난 담론이 포함하는 주제적 틀(thematic frames)이 있을 것이라고 가정했다. 앞서 검토한 선행연구에 따르면, 위험인식은 (가) 상황 자체에 내재한 위험성, (나) 위험한 결과를 초래하는 원인에 대한 인식, 그리고 (다) 위험한 상황에 대한 통제 및 조절 가능성에 대한 판단으로 이루어진다고 한다. 우리는 이런 위험인식 구성요소가 미세먼지를 포함한 환경위험에 대한 담론을 구성하는 일반적인 주제적 구성요소가 된다고 보았다. 따라서 이 세 가지가 미세먼지에 대한 뉴스 보도의 주제적 경향성, 즉 뉴스 프레임도 구성할 것이라고 가정했다. 실제 김영옥 등(2015a)의 연구, 엔트먼(Entman, 1991), 다인즈와 로드리게즈(Dynes & Rodriguez, 2010) 등이 제시한 연구결과를 보더라도 재난보도 프레임은 이 세 가지 요소를 빠뜨리지 않았다. 따라서 우리는 '미세먼지의 위험성', '미세먼지 발생의 원인', 그리고 '미세먼지 대책' 등 세 범주에 속하는 주제적 구성요소를 내용분석을 위한 분석유목으로 설정했다. 주제적 구성요소를 이용한 프레임 분석방법은 이준웅(1997, 2001), 이상률과 이준웅(2014) 등을 참조해서 수행했다.

미세먼지 관련 뉴스 중에서 무작위로 추출한 50개의 기사에 대한 탐색적인 정성적 텍스트

표 2. 미세먼지 뉴스 프레임 도출을 위한 주제변수

주제 범주	변수 이름	주제변수의 내용
위험성 (R)	R1	아토피, 알레르기, 피부질환, 안질, 안질환
	R2	천식, 기침, 호흡기질환, 중증호흡기 질환
	R3	만성폐쇄성 폐질환, 폐질환, 심혈관질환
	R4	뇌질환, 치매, 발달장애, 소아장애, 정신질환
	R5	발암물질, 암, 사망, 조기사망
	R6	독성물질, 화학물, 오존, 유해물질, 중금속
	R7	폐포 침입, 침투, 혈관 침투, 허파뼈리 흡착
원인 (C)	C1	발전소, 화력발전소, 사업장, 공장
	C2	경유차, 자동차, 폭스버겐
	C3	중국, 중국 스모그
	C4	가정먼지, 고등어, 음식조리
대책 (E)	E1	마스크 쓰기
	E2	외출 자제
	E3	공기청정기, 청소, 환기, 청소기, 세탁기
	E4	수분섭취, 차, 습도, 비타민C, 영양제, 한약
	E5	가동중단, 섯다운, 저감대책, 비상대책
	E6	경유차 규제, 경유값 인상, 가격 조정
	E7	친환경 에너지, 대체 에너지, CNG, LNG, LPG
	E8	공동조사, 외교 정책, 손해배상 청구

분석을 수행했으며, 이 분석결과를 활용해서 (가) 위험성에는 아토피 질환과 같은 피부 질환에서부터 폐암, 조기 사망 등 다양한 질환과 미세먼지의 편재성이나 침투력 강조하는, 이른바 미세먼지의 ‘위험성’ 요인에 대한 주제가 포함됨을 확인할 수 있었다. 또한 (나) 미세먼지의 원인으로는 국내의 석탄화력 발전소, 경유차 등이 거론되며, 외부적인 요인으로는 ‘중국발 스모그’가 빈번히 언급되는 것을 확인했기에, 이런 원인요인들을 분석유목으로 삼았다. 마지막으로 (다) 미세먼지에 대한 대처로는 마스크 착용과 외출 자제와 같은 개인적인 대처부터, 비상 저감 조치와 같은 정책적 대응, 그리고 에너지 정책 법제화와 산업구조의 변화와 같은 장기적인 대책까지 다양한 수준에서 다양한 방법이 제시됨을 확인했다. 결국 미세먼지의 위험성, 원인, 그리고 대책이라는 세 가지 미세먼지 보도의 주제적 특징을 발견할 수 있었는데, 이 결과는 앞서 논의한 재난 재해에 대한 뉴스 프레임 분석을 수행한 선행 연구결과를 정리한 내용과 조응하는 것이기도 하다.

뉴스 내용을 검토한 결과를 기초로, <표 2>과 같이 미세먼지 관련 뉴스에 나타난 주제변수를 설정했다. 3개의 범주, 즉 미세먼지 위험성, 원인, 대책 범주에 각각 하위 주제변수들이 상호 배제적으로 담겨있는 형식이다. 이 분석틀을 이용해서 전체 3,636개 기사에서 체계적으로 추출한 300개 기사에 대한 내용분석을 수행했다. 해당 기사가 분석틀에 제시한 주제변수를 담고 있으면, 해당 주제변수의 값을 1로 설정하고 그렇지 않으면 해당 변수의 값을 0으로 결정하는 방식으로 인간코딩을 수행했다. 2명의 연구원이 300개 기사를 코딩한 결과에 대해 코헨 카파(Cohen’s Kappa) 값을 기준으로 .84를 얻었으며, 따라서 연구원 간 일치도로 측정한 신뢰도 수준이 수용할 만하다고 판단했다.

## 4. 연구 결과

### 1) 주제변수 코딩 프레임 분석 결과

뉴스 프레임 구성을 위해, <표 2>에 제시한 주제변수 별 분석유목을 이용해서 얻은 자료를 가지고 군집분석을 수행했다. 300개의 신문기사에 <표 2>에 제시한 주제변수, 즉 위험성 관련 주제변수 7개, 원인 관련 주제변수 3개, 대책 관련 주제변수 8개가 독립적으로 등장하는지 아니면 등장하지 않는지 코딩한 결과를 정리하면, 300 사례 x 18 주제변수 행렬을 얻을 수 있다. 이 행렬을 다이스 유사성 척도(the Dice measure)로 변환한 후, 최대값 1에서 빼면 주제변수들 간의 거리행렬(18 x 18)을 얻을 수 있다. 다이스 척도는 이진변수 간 유사성을 계산할 때, 변수 간 값이 일치하는 경우 가중치를 부여하는 방법으로 희소성을 갖는 빈도자료 행렬에서 패턴을 파악하

는 데 도움을 준다. 주제변수 간 거리행렬을 이용해서 변수 간 유사성을 고려해서 군집화 패턴을 확인하기 위해 계층적 군집분석을 수행했다(이상률·이준웅, 2014; 이준웅, 1997).

계층적 군집분석 결과로 <그림 2>와 같은 덴드로그램을 얻었다. 실루엣(Silhouette) 계수와 던(Dunn) 지표를 활용해서 덴드로그램 결과를 검토한 결과, 군집의 수를 4개로 결정함이 적절한 것으로 보였다. 4개의 주제변수 군집을 정리해서 <표 3>의 첫째 열에 정리했다. 군집분석 결과로 얻은 집단은 주제변수의 집합이며, 각각 위험성(R) 관련 주제, 원인(C) 관련 주제, 그리고 대책(E) 관련 주제변수들을 고루 포함하고 있으므로 위험성-원인-대책이 결합한 일종의 '뉴스 이야기 구조'로 간주할 수 있다. <표 3>의 셋째 열은 주제적 내용을 결합해서 뉴스 프레임으로 재구성한 것이며, 이를 기초로 각 집단에 이름을 붙여서 프레임의 명칭으로 삼았다.

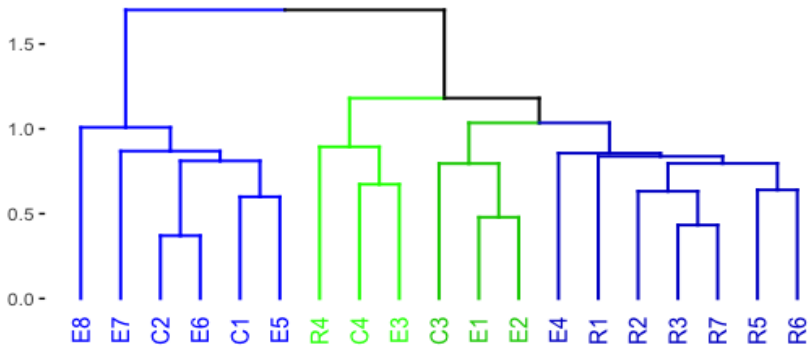


그림 2. 미세먼지 뉴스 프레임 주제 군집분석 결과 (n = 300)

표 3. 미세먼지 뉴스 주제변수 군집분석 후 프레임 구성 결과 (n = 300)

주제변수의 군집	프레임 명칭	주제적 내용
E7, C2, E6, C1, E5, E8	화석연료	화력발전소와 경유차로 인한 미세먼지 발생에 대한 대책 마련. 경유차 규제와 친환경 에너지 등 정책적 대응 필요.
R4, C4, E3	가정 재해	가정에서 발생하는 미세먼지를 환기와 공기청정기로 대응, 소아와 노인질환 염려
C3, E1, E2	중국 원인	중국발 미세먼지로 인한 마스크 착용 필요. 외출을 자제할 것.
E4, R1, R2, R3, R7, R5, R6	개인 위험	피부질환, 호흡기질환, 독성물질로 인한 사망위험성 증가 등 위험 요인. 수분섭취와 영양제 등 대책이 필요.

첫째 뉴스 프레임은 '화석연료 프레임'으로, 이는 화력발전소(C1)와 경유차(C2) 등 국내 미세먼지 발생원인에 집중하고 경유차 규제(E6)와 친환경 에너지 정책(E7)을 대안으로 제시할 수 있다. 두 번째 프레임은 '가정재해 프레임'이라 부를 수 있는데, 음식을 조리할 때 발생하는 미세먼지(C4)와 같은 가정에서 발생하는 미세먼지의 위험성을 지적하고 환기와 공기청정

기의 사용(E3)을 대안으로 제시하는 주제를 포함하고 있기 때문이다. 세 번째는 ‘중국원인 프레임’이라 명명했는데, 이는 중국발 미세먼지(C3)을 원인으로 제시하고, 마스크 쓰기(E1)와 외출 자제(E2)를 대안으로 제시하는 뉴스들을 관통하는 주제를 담고 있다. 마지막으로 ‘개인위험 프레임’이 있는데, 이는 아토피(R1), 호흡기질환(R2), 폐질환(R3) 등과 같은 질병을 강조해서 제시하고 수분을 자주 섭취할 것(R4)을 제안하는 등 개인적 수준의 위험을 강조하는 프레임이라고 할 수 있다.

## 2) 구조적 주제모형(STM) 분석 프레임 구성

구조적 주제모형 분석방법은 전체 문서에 담긴 단어들의 분포에서 개별 문서, 즉 개별 기사에 담긴 단어의 분포를 결정하는 잠재적 주제들(latent topics)을 추출하는 방법론이다. 구조적 주제모형 추출 방법은 전통적인 주제모형 도출방법이라 할 수 있는 ‘잠재적 디리클레 할당(latent Dirichlet allocation)’ 방법과 유사하게 단어집합의 분포에서 잠재적 주제들이 등장할 확률을 계산하지만, 잠재적 주제들 간의 상관관계를 허용한다는 장점이 있다. 또한 일정 수의 주제를 도출한 후 얻은 잔차(residuals)의 분포를 검토할 수 있기 때문에 이를 이용해서 모형화 결과를 타당성을 검토할 수도 있다.

우리는 먼저 3,636개 기사를 사전처리해서 명사를 추출함으로써 문서-용어 행렬(DTM)을 구성했다. 이때 얻은 명사의 빈도분포는 앞서 <표 1>에 제시한 결과와 같다. 둘째, 잠재적 주제의 수에 따라 변화하는 모형의 로그우도(log-likelihood), 의미적 일관성(semantic coherence), 잔차(residuals)등을 검토해서 주제의 수를 30개로 설정해서 잠재적 주제를 추출했다. 추출한 30개의 잠재적 주제에 주로 등장하는 단어를 단어의 목록을 <표 4>에 제시했다. 셋째, 30개의 주제를 그대로 뉴스 프레임으로 간주하는 것도 가능하지만, 이렇게 많은 수의 프레임을 설정하면 너무 번잡하여 개념적 간결성(parsimony)이 떨어지며, 따라서 프레임 해석은 물론 향후 분석에서 활용할 때에도 문제가 생긴다. 또한 구조적 주제모형 기법은 원래 주제 간의 상관관계를 허용하기 때문에, 30개의 주제 중에서 상관관계가 높은 주제들은 서로 의미적으로 연관성이 있는 주제들이며, 따라서 하나의 주제적 프레임으로 조작화하는 것이 가능하다고 보았다. 이에 착안해서 우리는 먼저 주제 간 상관관계를 거리 척도로 전환해서 거리 행렬(distance-matrix)을 만들고, 이 거리행렬을 다시 군집분석에 투입하는 식으로 추가적인 자료의 의미적 차원 축소분석을 수행했다.



표 4. 구조적 주제모형(STM) 주제별 주요 구성 단어들 (N = 3,636)

주제	주요 등장 단어
V1	마스크, 의약품, 식약처, 나들이, 편의점, 식품의약품, 일회용, 정전기, 숨쉬기, 메르스
V2	자동차, 기술진, 배터리, 운전자, 승용차, 기아차, 하이브리드, 수입차, 주행거리, 배기가스
V3	고등어, 삼겹살, 백화점, 보도자료, 어머니, 직화구이, 생산량, 구조조정, 온라인, 환풍기, 환경부
V4	종합대책, 환경부, 부담금, 조정실, 환경개선부담금, 윤성규, 실효성, 보조금, 김정수, 국무총리
V5	중급속, 돼지고기, 카드뮴, 발암물질, 머리카락, 질산염, 황산염, 악영향, 암연구, 흙먼지, 근로자
V6	스모그, 베이징, 환경보호, 허베이, 중국에서, 중국인, 고속도로, 특파원, 베이징과, 보건기구, 중국
V7	환경과학, 고농도, 총청권, 중부지방, 백령도, 일평균, 서울지역, 대기환경, 오염도, 수도권, 중국발
V8	우리나라, 환경기준, 전문가, 선진국, 한겨레, 목소리, 환경공학, 그동안, 연평균, 가운데, 주변국
V9	도라지, 폐질환, 기관지, 호흡기, 기관지염, 폐쇄성, 산들건강, 수세미, 호흡기질환, 폐기능, 호흡곤란
V10	겨울철, 강수량, 식염수, 발원지, 기능성, 저기압, 북서풍, 한반도를, 고기압, 흡연자, 수증기,
V11	자외선, 클렌징, 브랜드, 트러블, 화장품, 이벤트, 디자인, 테스트, 피부염, 점유율, 신제품, 프리미엄
V12	한국, 중국, 한국, 변호사, 손해배상, 지구온난화, 기후변화
V13	지난해, 아파트, 친환경, 포인트, 베란다, 지난달, 관계자, 상반기, 투자하, 관광객, 하반기, 시멘트
V14	자전거, 서울시, 대중교통, 지하철, 자차구, 서울시와, 서울에, 박원순, 서울시장, 서울시세, 주차장
V15	알레르기, 대체기, 치료제, 축농증, 꽃가루, 치료법, 아비인후과, 환절기, 진드기, 항생제
V16	더불어민주당, 안철수, 대선후보, 일자리, 박근혜, 유승민, 새누리당, 대통령의, 위원장, 청와대
V17	공기청정기, 인터넷, 서비스, 코웨이, 스카이, 판매량, 공기청정, 청정기, 소비자, 공기정화
V18	약순환, 전병헌, 이진구, 베스트, 적폐청산, 국론분열, 자발적, 댓글공작, 정치개입,
V19	비타민, 영양소, 단백질, 면역력을, 면역력, 항산화, 혈액순환, 신진대사, 기능성, 식품과, 노폐물
V20	청소기, 진공청소기, 사무실, 곰팡이, 알데히드, 유해물질, 가스레인지, 오염물질, 화학물, 증후군
V21	석탄화력발전소, 화력발전, 태양광, 화력발전소, 에너지, 온실가스, 전기요금, 유연탄, 충남도
V22	초미세먼지, 주의보, 시간당, 마이크로, 설문조사, 노약자, 홈페이지, 보건기구,
V23	에어컨, 프로젝트, 시스템, 스마트, 인공지능, 디지털, 전기료, 글로벌, 기술력, 냉장고, 사용자
V24	살균제, 가슴기, 플라스틱, 피해자, 쓰레기, 과학자, 이야기, 폐기물, 캠페인, 생태계,
V25	프로그램, 운동장, 학부모, 초등학교, 매뉴얼, 어린이, 측정기, 교육청, 유치원, 초등학생, 청소년
V26	사업장, 건설기계, 저공해, 경유차, 인천시, 화물차, 과태료, 비산먼지, 수도권
V27	뇌졸중, 대기오염, 연구팀, 사망자, 교수팀, 보고서, 조기사망, 이산화황, 국제학술지, 일산화탄소
V28	질소산화물, 배출가스, 배출량, 휘발유, 폴크스바겐, 오염원, 배기가스, 휘발유차, 경유차
V29	스트레스, 바이러스, 발생률, 고혈압, 심혈관질환, 당뇨병, 심근경색, 호르몬, 우울증, 인플루엔자
V30	결막염, 이물질, 컴퓨터, 알레르기성, 일기예보, 건강보험, 전문가, 스마트폰, 장시간, 꽃가루

구조적 주제모형 분석결과를 이용한 주제 간 위계적 군집분석 결과는 <그림 3>에 텐드로그 램으로 제시했다. 실무엿 계수와 된 지표를 검토해서 판단한 결과, 잠재적 주제를 5개의 군집으로 구분할 수 있다고 판단했다. 위계적 군집분석 결과를 근거로 각 군집에 속한 잠재적 주제들은 서로 밀접한 의미적 연관성을 가지며, 이런 의미적 연관성을 일종의 뉴스 프레임으로 개념화할 수 있다고 보았다. 즉 이 연구에서 뉴스 프레임을 구성하는 또 하나의 방법은 구조적 주제모형 분석결과로 얻은 잠재적 주제들 간의 군집화 경향으로 조작적으로 정의할 수 있다는 것이다. 이런 조작화가 타당하지 검토하기 위해서, <표 5>와 같이 잠재적 주제의 군집을 결합해서 단일한

뉴스 프레임으로 보고 각 군집의 의미를 해석해 보았다.

첫번째 프레임은 다섯 개의 주제들(V15, V30, V19, V29, V5, V9)로 구성되어 있는데, '미세먼지로 인한 알레르기, 결막염, 폐질환 우려가 있으며, 이에 대처하기 위해 비타민, 돼지고기, 도라지 등으로 이물질과 중금속에 대처하고, 스트레스를 달래야 한다'는 의미를 제공하는 것으로 해석할 수 있었다. 우리는 이 첫 번째 군집을 '개인위험' 프레임이라 부르기로 했다. 두번째는 네 주제(V4, V26, V2, V28)로 구성되며, 이는 '미세먼지는 주로 경유차와 배기가스로 발생하며, 따라서 배기가스 배출량 규제와 같은 정책이 필요한데, 이를 위해 환경부가 노력해야 한다'는 의미를 지녔다. 이를 '배기가스' 프레임으로 부르기로 했다. 세번째 프레임은 8개의 주제로 구성되는데, '삼겹살, 고등어, 곰팡이 등 생활 미세먼지에 대한 대책이 필요하며, 에어컨과 공기청정기를 이용해서 환기하고, 외출 시에는 마스크를 착용해야 한다'는 의미를 지닌 것으로 나타났다. 우리는 이를 '생활먼지' 프레임으로 명명했다. 네번째 프레임은 네 개의 주제(V6, V10, V7, V22)로 구성되며, '겨울철 북서풍을 타고 오는 중국발 미세먼지로 한반도 중부지방 대기오염이 심각하다'는 내용을 담고 있었다. '겨울중국' 프레임이라 부르면서 타당하다고 판단했다. 마지막 프레임은 6개의 주제(V16, V21, V24, V27, V8, V12)이 군집화한 결과였는데, 이는 '석탄 등 화석연료 발전으로 대기오염이 심각하며, 대체 에너지 정책마련을 위한 근본적인 대안이 필요하다'는 의미를 지녔다. 우리는 이 프레임을 '에너지정책' 프레임으로 명명했다.

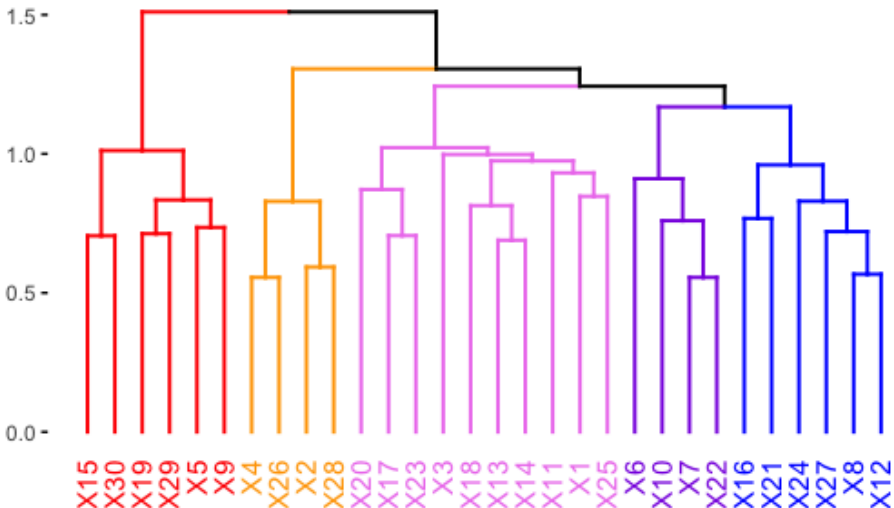


그림 3. 미세먼지 구조적 주제모형(STM) 분석 후 군집분석 결과

표 5. 미세먼지 구조적 주제모형(STM) 프레임 구성결과

프레임 명칭	주제모형 군집분석 결과	주제적 내용
개인위험	V15, V30, V19, V29, V5, V9	미세먼지로 인한 알레르기, 결막염, 폐질환 우려가 있음. 비타민, 돼지고기, 도라지 등으로 이물질과 중금속에 대처하고, 스트레스를 달래야 함.
배기가스	V4, V26, V2, V28	경유차와 배기가스로 인한 위험이 문제임. 배기가스 배출량 규제와 같은 대책이 필요하며, 환경부 등 당국의 적극적 대처가 필요함.
생활먼지	V20, V17, V3, V18, V13, V14, V11, V1, V25	삼겹살, 고등어, 곰팡이 등 생활 미세먼지 배출도 대책이 필요함. 에어컨을 이용해서 환기를 자주하고 공기청정기를 켜고, 외출 시에는 마스크를 착용.
겨울중국	V6, V10, V7, V22	겨울철 복사풍으로 인한 중국 발 미세먼지로 수도권과 중부지방 대기 오염이 심각함.
에너지정책	V16, V21, V24, V27, V8, V12	국내에서 화석연료를 이용한 대기오염이 심각함. 대체 에너지 정책을 비롯한 근본적인 대책을 마련하기 위한 정부와 정치권의 노력이 필요함.

### 3) 구조적 주제모형 프레임과 주제변수 프레임 간 상관관계

〈표 6〉의 왼편 상관관계 행렬은 베이즈 알고리즘으로 코딩한 구조적 주제모형 프레임(〈표 3〉에 제시)과 인간이 코딩한 주제변수 프레임(〈표 5〉에 제시) 간 상관관계를 보여준다. 구조적 주제모형 프레임 가운데 ‘개인위험’ 프레임은 인간코딩 ‘개인위험’ 프레임과 상관관계가 .52 ( $p < .001$ )인 것으로 나타나 실제적인 상관성을 보였다. 프레임을 구성하는 내용을 보더라도 의미적으로 유사성을 확인할 수 있었다. 마찬가지로 주제모형 ‘배기가스’ 프레임도 인간코딩 프레임 중 ‘국내 원인’과 상관관계가 .53 ( $p < .001$ )이어서 이 둘 역시 사실 상 같은 내용이라 할 수 있다. 주제모형 ‘겨울중국’ 프레임은 인간코딩 ‘중국원인’ 프레임과 .30 ( $p < .001$ )의 상관관계를 기록했다. 그러나 ‘생활먼지’ 프레임은 인간코딩 ‘가정재해’ 프레임과 .19 ( $p < .05$ )로 상대적으로 낮은 상관성을 보이고, ‘에너지정책’ 프레임은 인간코딩한 주제변수 프레임들과 유의한 상관성을 기록하지 못했다.

이상의 결과를 근거로 구조적 주제모형 추출방법으로 얻은 프레임 중에서 ‘개인위험’, ‘배기가스’, ‘겨울중국’ 프레임들의 내용 타당성이 확인되었다고 말할 수 있다. 왜냐하면 베이즈 알고리즘으로 도출한 주제모형들의 내용이 연구자가 선행연구를 분석해서 도출한 위험성, 원인, 대책을 중심으로 미세먼지 관련 뉴스의 주제를 포착한 내용과 유사성을 보였기 때문이다. 그렇다면 상관성이 낮거나 아예 나타나지 않은 ‘생활먼지’와 ‘에너지정책’ 프레임은 어떻게 해석해야 할 것인가? 우리는 ‘생활먼지’ 프레임과 ‘에너지정책’ 프레임을 구성하는 주제들(STM topics, 〈표 4〉 참조)을 참조해서 판단하건대, 주제변수를 이용한 인간코딩이 이 두 프레임을 발견하는 데 실패한 것으로 해석할 수 있다고 보았다. ‘생활먼지’ 프레임은 생활먼지의 다양한 원인성에 대처하기 위한 외출시 마스크 착용과 같은 개인적인 대응부터 공기청정기 구매와 같은 소비적 대응까지 다양한 내용을 포괄하지만, 인간코딩 프레임인 ‘생활대책’은 가정 내 미세먼지 대책을 좁게 다루는 정도

었다. 또한 ‘에너지정책’ 프레임은 미세면지에 대한 정치권의 대책과 정책마련 과정을 포괄적으로 담고 있지만, 정책적 논의의 특징 상 구체적인 사건을 담아내지 않기에 인간코딩은 이런 주제를 담은 뉴스 프레임을 간과했을 가능성이 있다고 보였다. 미세면지 보도에서 인간이 확인하기 어려웠던 내용적 특성을 알고리즘 코딩을 통해 포착할 수 있었다는 의미다.

표 6. 구조적 주제모형(STM) 프레임과 주제분석 프레임 간 상관관계

	주제분석 프레임 (n = 300)				구조적 주제모형(STM) 프레임 (N = 3,636)				
	화석연료	가정재해	중국원인	개인위험	개인위험	배기가스	생활면지	겨울중국	에너지정책
개인위험	-0.30***	0.11	0.04	0.52***	1.00	-0.31***	-0.19***	-0.19***	-0.32***
배기가스	0.53***	-0.12	-0.21*	-0.19*		1.00	-0.24***	-0.16***	-0.11***
생활면지	-0.14	0.19*	0.03	-0.13			1.00	-0.30***	-0.38***
겨울중국	-0.17	-0.13	0.30***	-0.06				1.00	-0.24***
에너지정책	0.16	-0.08	-0.18	-0.18					1.00

주: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

〈표 6〉의 오른쪽 상관관계 행렬은 5개의 구조적 주제모형 프레임 간 상관관계를 보여준다. 전체적으로 프레임 간에 부적인 상관관계가 형성되어 있음을 확인할 수 있는데, 이는 예상할 수 있는 당연한 결과다. 왜냐하면 애초에 5개의 프레임을 구성하는 30개의 구조적 주제모형이란 개별 기사 내에서 발생할 확률이 서로 다른 잠재적 주제모형이기에 한 종류의 주제모형 등장 확률이 높으면, 다른 종류의 주제모형 등장 확률이 낮아지기 때문이다. 따라서 구조적 주제모형의 집합으로 측정된 프레임의 발생정도는 ‘프레임 간 경쟁(frame competition)’을 반영한다고 해석할 수 있다. 상관관계 계수를 보면, ‘에너지정책’ 프레임과 ‘생활면지’ 프레임 간 부적인 관계가  $-0.38$  ( $p < .001$ )에 달하며, ‘에너지정책’ 프레임과 ‘개인위험’ 프레임 간 부적인 관계도  $-0.32$  ( $p < .001$ )에 해당하는 데, 에너지 정책에 대한 대안을 모색하는 보도를 할수록 생활 및 개인관련 보도는 감소했다고 해석할 수 있는 결과다. 반면, ‘겨울중국’ 프레임은 ‘개인위험’ 프레임과 상관관계가  $-0.19$  ( $p < .001$ )로 다른 프레임에 비해 부적인 상관성의 규모가 상대적으로 작다. 이는 ‘겨울중국’ 프레임을 담은 보도에서 개인에 대한 위험성에 대한 강조가 다른 요인으로 인한 위험성보다 상대적으로 강조되어 나타났다고 해석할 수 있다.

#### 4) 언론사별 구조적 주제모형 프레임의 차이

언론사 간에 구조적 주제모형으로 추출한 프레임 등장 정도가 차이가 있는지 확인하기 위해, 언론사별 프레임 강도를 변량분석 방법으로 비교해 보았다. 〈표 7〉에 제시한 프레임 강도는 해당

프레임에 속하는 주제가 개별 기사에 등장할 확률을 언론사별로 합산한 내용으로, 일종의 비율의 합이라 할 수 있다. 따라서 이는 언론사별 기사의 수가 많을수록, 그리고 단위 기사당 해당 프레임에 속하는 주제가 등장할 확률이 클수록 값이 커지는 속성을 가진 변수이다. <표 7>에 제시한 평균값은 언론사별 비율합의 평균이며 언론사 간 상대비교가 가능한 값이라 할 수 있다.

표 7. 언론사별 구조적 주제모형(STM) 프레임 강도 (N = 3,636)

	전체	조선일보	중앙일보	동아일보	한겨레	경향신문	한국일보	F(5, 2630)
개인위험	.13	.20 <sup>a</sup>	.14	.09 <sup>bc</sup>	.06 <sup>cd</sup>	.19 <sup>a</sup>	.09 <sup>cd</sup>	29.95***
배기가스	.14	.12 <sup>a</sup>	.13 <sup>a</sup>	.14 <sup>a</sup>	.16 <sup>ab</sup>	.12 <sup>a</sup>	.17 <sup>b</sup>	4.33***
생활먼지	.31	.29 <sup>a</sup>	.29 <sup>a</sup>	.49	.25 <sup>a</sup>	.26 <sup>a</sup>	.24 <sup>a</sup>	86.25***
겨울중국	.17	.18 <sup>a</sup>	.22 <sup>b</sup>	.09	.20 <sup>ab</sup>	.17 <sup>a</sup>	.16 <sup>a</sup>	28.01***
에너지정책	.24	.20 <sup>a</sup>	.22 <sup>a</sup>	.18	.32 <sup>b</sup>	.26	.34 <sup>b</sup>	28.69***

주: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . 같은 첨자abcd가 붙은 평균값 간에는 튜키 유의성 차이가 없음.

언론사 간 상대비교를 하기 전에 전체 프레임강도를 보면, ‘생활먼지’ 프레임이 .31로 전체 미세먼지 기사 중에서 가장 강하게 등장했고, 그 다음으로 ‘에너지정책’ 프레임이 .24로 강력했다. 나머지 ‘겨울중국’, ‘배기가스’, ‘개인위험’ 프레임은 서로 유사한 수준으로 상대적으로 약하게 등장했음을 알 수 있다. 언론사 간 차이를 프레임별로 구분해서 보면, ‘개인위험’ 프레임은 조선일보(M = .20)와 경향신문(M = .19)에서 가장 강하게 나타났다. 튜키의 정직한 유의성차이(Tukey’s honest significant difference)를 적용한 사후분석을 통해서, 두 신문 간의 평균값의 차이의 유의성을 검토한 결과, 조선일보와 경향신문 간에는 프레임 강도를 기준으로 보아 유의한 차이를 발견할 수 없었다. ‘배기가스’ 프레임은 한겨레(M = .16)와 한국일보(.17)에서 강하게 나타났으며, ‘생활먼지’ 프레임은 동아일보에서 압도적으로 높게 나타났다. ‘겨울중국’ 프레임은 중앙일보와 한겨레에서 강조되어 나타났고, ‘에너지정책’ 프레임은 한겨레와 한국일보에서 강조되었다. 한겨레는 미세먼지의 원인을 중국 등 외부로 귀인한다는 점에서는 중앙일보와 유사했지만, 배기가스와 같은 국내 원인도 함께 강조하면서, 국내 정치와 행정력을 동원한 정책적 대응을 강조했음을 알 수 있다. 사후분석을 포함한 변량분석 결과를 전체적으로 조망하면, 조선일보와 경향신문은 ‘개인위험’을 강조하고 ‘배기가스’를 상대적으로 덜 다룬 특징을 보인다. 한겨레와 한국일보는 ‘배기가스’와 ‘에너지정책’을 강조해서 다룬 유사성을 보인다. 동아일보는 ‘생활먼지’ 프레임에 집중한 특성을 보였다고 해석할 수 있다.

## 5) 구조적 주제모형 프레임의 시계열 패턴

구조적 주제모형으로 정의한 프레임들이 2013년 1월 이후 어떤 변화를 보였는지 검토하기 위해, <그림 4>와 같이 시계열로 도시해 보았다. 그림에 배경으로 제시한 회색 음영은 미세먼지 기사의 빈도를 나타낸다. 이 음영은 앞서 <그림 1>에서 제시했던 기사 빈도의 분포를 배경으로 적용한 것이다.

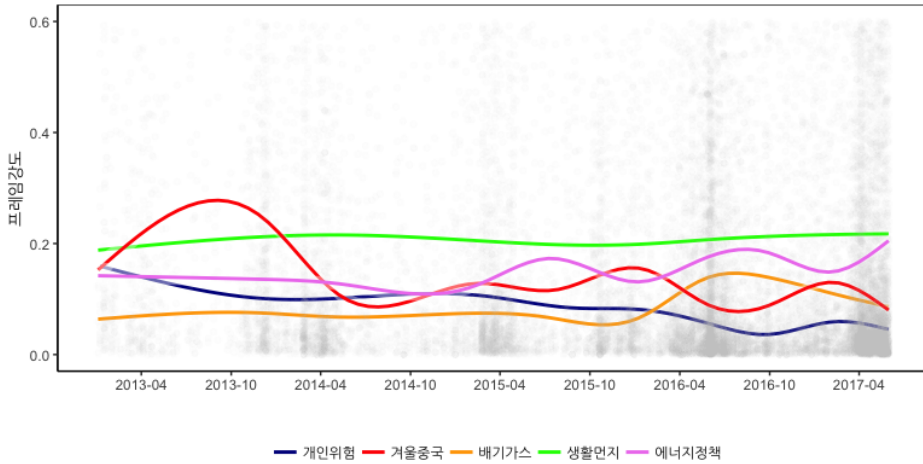


그림 4. 종합일간지 미세먼지 뉴스 프레임의 시계열 패턴

프레임 별로 시계열 변화패턴을 보면, 일단 ‘생활먼지’ 프레임은 2013년 1월 이후 2017년 5월까지 시계열 등락이 없이 가장 안정적으로 등장했던 프레임이었음을 확인할 수 있다. ‘개인위협’ 프레임은 2013년에는 상대적으로 강했지만 이후 지속적으로 중요도가 낮아진 프레임이라는 것도 확인할 수 있다. 반대로 시점에 따른 등락의 폭이 큰 것으로 보이는 프레임은 ‘겨울중국’이다. 2013년에서 2015년 사이에는 등락의 폭이 크다가 이후에는 겨울에는 강조되지만 여름에는 약화하는 계절성은 그대로 남지만, 등락의 폭이 감소하는 추세를 보였다. 2013년에는 상대적으로 언론에 가장 많이 등장했던 뉴스 프레임이 바로 ‘겨울중국’이었음을 알 수 있다.

변동성의 관점에서 보면, ‘에너지정책’ 프레임은 ‘겨울중국’ 프레임과 반대되는 양상을 보인다. 2013-4년에는 큰 변화가 없는 패턴을 보이지만, 2015년 이후 계절성을 가지면서 상승하는 추세를 보인다. 특히 ‘에너지정책’ 프레임은 ‘겨울중국’ 프레임이 강해지면 약해지는 등 상보적인 관계를 잘 보여준다. 이는 언론의 미세먼지 담론이 ‘겨울중국’ 프레임과 ‘에너지정책’ 프레임 간에서 한 쪽을 강조하면 다른 쪽은 약해지는 방식으로 일종의 프레임 경쟁이 이루어졌다고 해석

할 수 있는 결과다. ‘겨울중국’ 프레임을 강조하다 보면, 외교적 대응을 제외하고는 국내 정치와 행정을 동원한 정책적 대응으로 언급할 내용이 별로 없다는 것이 원인이 되었으리라고 본다. 마지막으로 ‘배기가스’ 프레임은 2013년 이후로 별로 강조되는 프레임이 아니었는데, 2016년 6월부터 강화하는 양상을 보인다. 이는 2016년 6월 3일 당시 황교안 국무총리 주재로 관계부처 장관회의에서 ‘미세먼지 관리 특별대책’을 확정해서 발표한 이후, 환경부, 미래부, 교육부, 기상청 등 모든 정부부처에서 후속 조치를 내놓기 시작했던 사실을 반영하는 결과라고 하겠다.

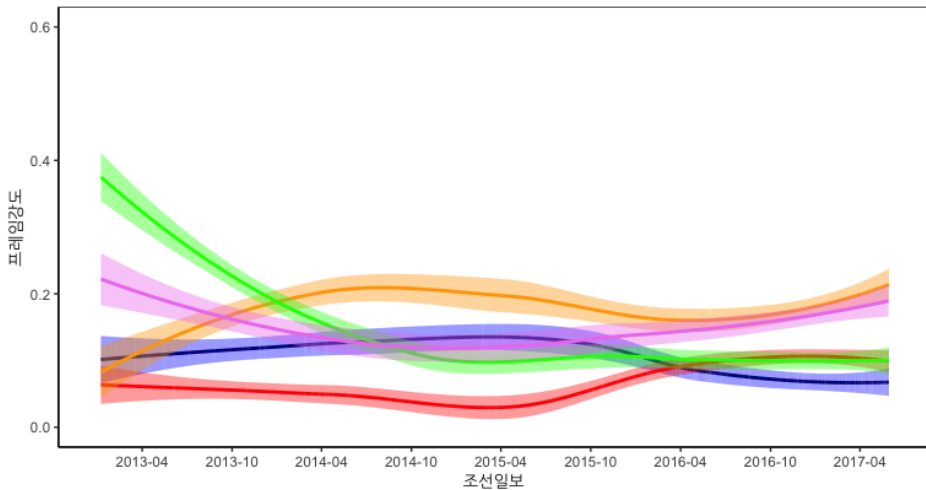


그림 5. 조선일보 미세먼지 프레임의 시계열 패턴

앞서 <표 7>을 검토하며 확인했듯이, 조선일보와 경향신문은 다른 언론에 비해 ‘개인위험’ 프레임을 강조한 반면, 중앙일보와 한겨레는 ‘겨울중국’ 프레임을 다른 언론에 비해 상대적으로 강조했다. 언론사별 프레임 강조가 시기에 따라 변동을 보이는지 조선일보와 한겨레의 프레임 변화를 시계열 로 도시해서 검토해 보았다.

<그림 5>는 조선일보의 미세먼지 프레임의 변화를 제시한다. 2013년에서 14년까지 조선일보의 ‘생활먼지’ 프레임 강조는 지속적으로 감소하는 반면, ‘배기가스’ 프레임 강조는 증가하는 추세를 보였다. 2016년 6월 이후, 프레임 강조의 양상은 전반적으로 재조정되어서 ‘배기가스’와 ‘에너지정책’ 프레임이 완만하게 상승하는 추세를 보이는 반면, ‘개인위험’과 ‘생활먼지’ 프레임은 완만하게 감소하는 양상을 보였다. ‘겨울중국’ 프레임은 2016년 봄을 기점으로 해서, 그 전에는 전반적으로 약하게 등장하다가 기점 이후에 증가한 양상을 유지하는 패턴을 보였다. 조선일보의 시계열 프레임 변화 패턴을 보면, ‘겨울중국’ 프레임과 ‘배기가스’ 프레임 간의 상보적 관계를 볼 수

있다. 조선일보는 미세먼지 원인과 대책을 놓고 이 두 프레임 중 하나를 강조하면 다른 쪽은 상대적으로 감소하는 방식으로 보도했다고 추론할 수 있다. 경향신문은 2014년 이후 조선일보와 유사한 프레임 강조의 변화를 보였는데, 이는 정치적 이념이 다른 언론사 간에 미세먼지 보도 프레임 변화의 시기별 변화의 양상에 차이가 없다는 점에서 흥미로웠다. 그런데 생각해 보면, 미세먼지 보도가 언론사의 편집방향의 이념성을 반영하리라고 기대할 이유도 따로 없기에 이 결과를 특별히 의미화할 방법이 딱히 없다고 하겠다.

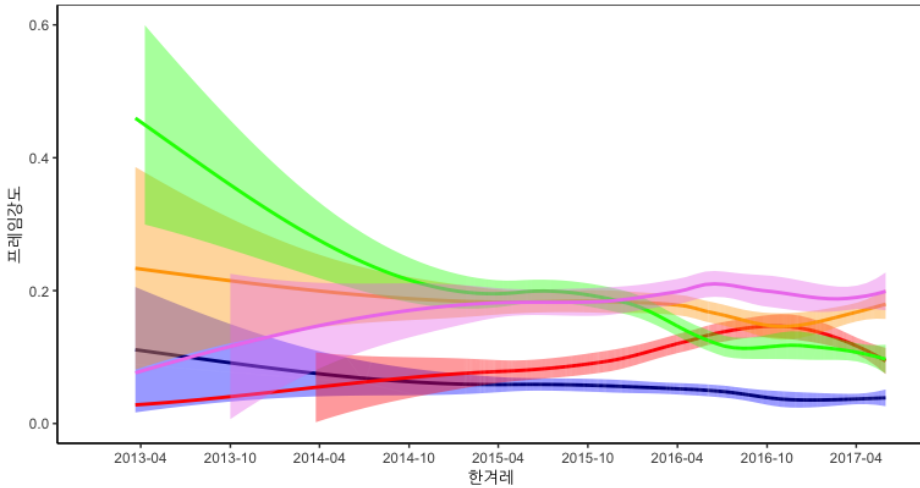


그림 6. 한겨레 미세먼지 프레임의 시기별 패턴

〈그림 6〉는 한겨레의 미세먼지 뉴스 프레임의 변화를 도시해서 제시한다. 2013년부터 2014년 초까지는 미세먼지 관련기사의 수가 적기 때문에 국소기중회귀(LOESS)를 적용한 추세선의 변화가 크며, 그에 따라 오차한계의 범위도 크다. 따라서 이 시기 프레임의 변화를 별도로 의미화해서 과대 해석할 수는 없다. 다만 ‘개인위협’ 프레임과 ‘생활먼지’ 프레임은 모두 2013년 이후 지속적으로 감소하는 추세를 보인다고 말할 수 있는데, 이는 사회정치적 원인과 대책에 초점을 맞추는 한겨레 보도의 특성을 반영한 것으로 보인다. 이 두 프레임과 반대로 ‘겨울중국’ 프레임과 ‘에너지정책’ 프레임에 대한 강조는 2015년 말까지는 완만하게 이루어지다가, 2016년부터 급격하게 강화되어 유지하는 양상을 보였다. 한편 2014년 말 이래 한국일보의 미세먼지 뉴스 프레임 변화 양상은 한겨레의 프레임 변화와 크게 다르지 않은 패턴을 보였다.



## 5. 결론과 논의

이 연구는 2013년 1월 이후 2017년 5월까지 우리나라 종합일간지의 미세먼지에 대한 보도를 뉴스 프레임으로 구성해서 언론사 간 프레임 강조의 차이와 언론사별 시계열적 프레임 변화를 중심으로 검토해 보았다. 이 연구는 특히 기존 연구결과를 참조해서 도출한 주제변수를 가지고 인간코딩당한 프레임 분석결과와 구조적 주제모형 분석방법(STM)을 이용해서 알고리즘 코딩한 결과를 비교하는 방식으로 프레임 분석의 타당성을 확인하는 방법론을 적용했다. 연구결과를 종합해서 정리하면 다음과 같이 요약할 수 있다.

먼저 환경과 위험에 대한 선행연구를 참조해서 미세먼지 위험성, 원인, 그리고 대책을 상위 범주로 삼고, 하위에 각 범주별 주요 주제를 상호배제적으로 배열한 분석유목을 적용해서 인간코딩 하는 방식으로 '화석연료', '가정재해', '중국원인', '개인위험' 등 네 개의 프레임을 추출했다. 동시에, 구조적 주제모형 분석방법을 적용해서 3,636개 기사에 잠재한 30개의 주제들을 베이지 추정방식으로 도출했으며, 주제 간 상관관계 행렬을 이용해서 위계적 군집분석으로 집단을 나누는 결과 '개인위험', '배기가스', '생활먼지', '겨울중국', '에너지정책' 등 5개 프레임으로 재구성했다.

둘째, 인간코딩한 프레임 중 '개인위험'과 '화석연료' 프레임이 각각 구조적 주제모형 추출을 통한 '개인위험'과 '배기가스' 프레임과 높은 상관관계를 보여 의미적으로 매우 유사한 프레임이라고 보았다. 또한 주제모형으로 추출한 '겨울중국' 프레임은 인간코딩 '중국원인' 프레임과 내용적으로 유사함을 확인했다. 그러나 인간코딩으로 확인하지 못했던 2개의 프레임, 즉 '생활먼지'와 '에너지정책' 프레임을 나타난다는 점을 구조적 주제모형 방식을 통해 발견할 수 있었는데, 이 결과는 알고리즘 코딩이 인간코딩 방식이 확인하지 못하는 뉴스의 주제적 특성을 찾는 데 도움을 준다는 것을 시사한다.

셋째, 언론사별 구조적 주제모형에 근거한 프레임 강도의 차이를 검토한 결과, 조선일보와 경향신문은 '개인위험' 프레임을 강조하고, 한겨레와 한국일보는 '배기가스' 프레임을 다른 언론사보다 강조해서 보도했다는 사실을 확인했다. '겨울중국' 프레임은 중앙일보와 한겨레에서, 그리고 '에너지정책' 프레임은 한겨레와 한국일보에서 상대적으로 강하게 나타났다. 또한 미세먼지 뉴스 프레임의 변화를 시계열적으로 도시해서 검토한 결과, '겨울중국' 프레임과 '에너지정책' 프레임은 한 시점에서 한 프레임이 강하면 다른 쪽은 약하고, 다른 시점에서는 관계가 역전하는 시계열적 상보성을 보인다는 것을 발견했다.

연구 결과의 함의를 논의하면 다음과 같다. 첫째, '자료기반적' 방법론을 사용해서 도출한 뉴스 프레임은 기존 재난 재해 뉴스보도의 프레임 연구에서 확인한 프레임과 유사한 특징을 보였

다. 예를 들어, ‘개인위험’ 프레임은 미세먼지가 야기할 수 있는 다양한 질병을 나열하면서, 미온적인 개인적인 예방과 대처를 논한다는 점에서 기존의 재난 재해 연구에서 확인했던 ‘위험 강조 프레임’과 유사하다. ‘생활먼지’ 프레임 또한 미세먼지의 만연성(pervasiveness)과 개인적인 대처를 제시한다는 점에서 유사하다. 그런데 ‘겨울중국’ 프레임은 미세먼지 재해의 외부적 요인을 강조하는 반면, ‘배기가스’ 프레임과 ‘에너지정책’ 프레임은 국내에서 발생하는 요인을 강조한다는 점에서 상호 대치되는 주제적 특성을 가지고 있는데, 이 두 프레임 모두 미세먼지의 원인성에 대한 서로 경쟁하는 ‘비난 프레임’으로 작동함을 알 수 있다.

자료기반적 방법론을 적용해서 추출한 미세먼지 프레임들도 미세먼지 재해의 실상과 관계 없이 특정한 위험 인식을 형성할 수 있는 의미적 요소를 포함한다는 데 주목해야 한다. 한편으로, ‘개인위험’ 프레임과 ‘생활먼지’ 프레임은 미세먼지의 위해성을 강조하면서도 예방과 대처를 개인 수준에서 논의하는 내용적 특징을 보인다. 다른 한편, ‘겨울중국’ 프레임이나 ‘배기가스’ 프레임, 그리고 ‘에너지정책’ 프레임은 미세먼지에 대한 제한된 요인과 이에 대한 일반적 대처를 논의하면서 미세먼지 재해에 대한 포괄적인 이해를 제공하는 프레임이라 할 수 있다. 이런 프레임의 의미적 내용은 여론 조사를 통해 밝혀진 공중의 미세먼지에 대한 정형화된 위험 인식의 특성을 상당 부분 설명해 줄 수 있다. 예를 들어, ‘개인위험’이나 ‘생활먼지’ 프레임은 미세먼지를 심각한 건강상의 위협으로 인식하는 데에 영향을 주었을 가능성이 있다. 또한 2013년 초반 ‘겨울중국’ 프레임이 강하게 나타난 현상은 왜 많은 이들이 미세먼지 재해의 가장 큰 책임을 중국이라는 외부 대상으로 특정해서 비판했는지 설명할 수 있는 시사점을 제공한다.

둘째, 이 연구에서는 과거 재난 및 재해 연구에서 특별히 주목하지 않던 뉴스 프레임을 하나 발견했다. 바로 ‘생활먼지’ 프레임이다.<sup>2)</sup> 이 뉴스 프레임은 미세먼지 재해를 일상생활과 소비 문화와 연결하고 있다는 특징을 지닌다. 미세먼지의 위험성을 전제하고 있지만, 어떤 비일상적이며 통제불가능한 재난이라기보다는 개인의 일상에 만연하지만 통제가능한 위협으로 제시한다. 즉 공기청정기나 청소기와 같은 소비재를 통한 대처할 수 있다는 점을 강조한다. 기존의 재난 재해 보도를 분석한 연구에서는 대규모 재난이나 일회적인 재난 사건에서 파생하는 사회적 문제를

---

2) 김영욱 등(2015)의 프레임 분석은 포괄적인 분석 유무를 담고 있는데, 이들이 제시한 결과 중 원인귀인 프레임에 속하는 ‘개인적 원인’과 대응 프레임에 속하는 ‘개인적 대처’가 우리가 발견한 ‘생활먼지’ 프레임과 의미적으로 유사하다. 그러나 김영욱 등의 연구에서 ‘개인적 원인’과 ‘개인적 대처’는 서로 다른 프레임의 값들에 해당하며, 따라서 독립적으로 주제적 통일성을 갖는 뉴스 프레임이라고 할 수 없다. 향후에 우리 연구결과를 참조해서 분석유무를 추가해서 인간코딩을 한다면 생활먼지 프레임을 발견할 수도 있겠지만, 이런 발견 역시 개념기반 방법을 적용했을 때 발생하는 확장편향적 경향에 속한다고 하겠다.

연구 대상으로 삼은 것들이 대부분이었는데, 이 때문에 선행 연구들은 생활과 소비와 관련한 뉴스 프레임의 현저성을 확인하기 어려웠을 것이라 판단한다. 구조적 주제모형을 사용한 '자료기반적' 프레임 추출방법을 사용했기에 재난보도 뉴스 프레임의 새로운 양상을 발견할 수 있었다고 본다. 우리는 '생활면지' 프레임이 제시하는 주제적 특징이 미세먼지 재해와 같이 장기간 점차적으로 발전하는 환경 재해에 관한 언론 보도에서 지속적으로 나타날 수 있다고 보며, 따라서 이런 뉴스 프레임의 작동 방식에 주목해야 한다고 제안한다.

셋째, 미세먼지에 대한 우리나라 언론 보도의 프레임을 보면, 미세먼지 재해에 대한 일련의 도식적인 이해(schematic sense-making)를 돕는 방식으로 의미를 전달한다고 정리할 수 있다. 이 연구에서 확인한 미세먼지 재해 보도의 뉴스 프레임들을 보면, 개별 프레임은 하나의 완결된 이야기 구조를 갖추지만 전체적인 관점을 제공하지는 못한다는 것을 알 수 있다. 예컨대, '에너지정책'이나 '배기가스' 프레임은 국내의 자동차나 화력발전소 등의 요소에 초점을 맞출 뿐이며, 법적 규제와 같은 좁게 설정된 대책을 제시한다. '겨울중국' 프레임은 통제하기 어려운 국외 요인에 따른 원인성을 강조하면서 동시에 마스크 착용, 외출 자제 등 이미 알려져 있는 상투적인 대책을 제시하고 있다. 우리가 발견한 프레임에는 기후변화 등의 전지구적인 요인에 대한 논의나 환경오염 개선, 친환경적 첨단 기술 개발 등 혁신적인 대책에 대해서는 논의가 미흡하다는 것을 발견할 수 있다. 미세먼지 프레임에 내재한 제한된 관점으로 인해, 시민은 미세먼지 재해에 대해 종합적이고 체계적인 인식의 틀을 갖추기 어렵게 될 우려가 있다.

이상의 논의를 종합해서 우리는 우리나라 언론이 미세먼지와 같은 재난 재해를 보도할 때, 일련의 파편화된 방식으로 프레임을 제한해서 제공함으로써 미세먼지 재해에 대한 파편화된 현실(fractured reality)을 구성한다는 함의를 도출할 수 있다. 언론은 미세먼지와 같은 복잡한 대상에 대해 보도할 때, 복잡한 맥락과 관련 고려 요인을 생략한 채 간결한 이야기 양식을 사용해서 단순화하는 경향이 있는데 미세먼지에 대한 보도에서도 이런 양상이 나타났다는 것이다. 우리는 이 연구에서 미세먼지 재해의 국내 원인을 강조하고 대책을 촉구하는 '에너지정책' 프레임의 전개가 국외 요인을 강조하는 '겨울중국' 프레임의 전개와 경쟁적으로 대치한다는 것을 경험적으로 확인했다. 이 결과는 미세먼지에 대한 우리 언론의 보도가 종합적인 원인 분석과 대책을 모색하는 보도였다기보다는 주제별로 그리고 시기별로 특정 프레임만을 강조해서 보도하는, 일종의 분열적 보도였다는 것을 뜻한다. 즉 자동차 배기가스, 화력발전, 중국 발 황사 등 어느 하나의 원인을 강조하면서, 그에 대한 대응적 대책마련을 촉구하는 식의 뉴스 보도가 지배적이었다는 것을 의미한다. 우리는 이런 식으로 파편화된 프레임을 접한 시민들이 미세먼지 문제가 다중적인 원인을 갖고 다양한 대책을 요청하는 복합적인 사안이라는 사실을 간과할 우려가 있다고 본다. 후속

연구에서 이런 우려가 실제적인지 확인하기 위한 미세먼지 프레임의 인지적 효과를 알아보기 위한 실험 또는 조사연구를 수행할 수 있는데, 우리 연구결과가 프레임 효과를 검토하기 위한 후속 연구에 '프레임 구성'에 대한 참조를 제공할 수 있을 것으로 기대한다.

## 참고 문헌

- 김영옥·이현승·장유진·이혜진 (2015a). 미세먼지 위험보도 프레임과 정보원 분석. <한국언론학보>, 59권 2호, 121-154.
- 김영옥·이현승·이혜진·장유진 (2015b). 미세먼지 위험에 대한 수용자의 인식과 의견 형성에 관한 연구. <한국언론정보학보 통권>, 72호, 52-91.
- 김영옥·이현승·이혜진·장유진 (2016). 미세먼지 위험에 대한 전문가와 일반인의 인식차이와 커뮤니케이션 단서 탐색. <커뮤니케이션 이론>, 12권 1호, 53-117.
- 김원용·이동운 (2005). 언론보도의 프레임 유형화 연구: 국내 원자력관련 신문보도를 중심으로. <한국언론학보>, 49권 6호, 166-197.
- 문안나 (2016). 한국 사회의 IR (Investor Relations) 개념-1994~ 2014 년 신문기사 내용분석 및 의미연결망 분석. <홍보학 연구>, 20권 1호, 50-78.
- 백영민·최문호·장지연 (2014). 한미 정권교체에 따른 주한 미대사관 외교문서의 주제와 감정표현 변화. <언론정보연구>, 51권 1호, 133-179.
- 서울환경연합 (2017). <9대 대선후보와 차기 정부에 바란다. 서울환경연합, 미세먼지 대책 조사결과 발표>. 서울: 서울환경연합.
- 안병옥 (2017. 4. 6.). 미세먼지 대처, 중국 탓만 말라. <경향신문>, 30.
- 이민규·이예리 (2012). 국내 신문의 가축 전염병 위험보도에 대한 프레임 연구. <언론과학연구>, 12권 2호, 378-414.
- 이상률·이준웅 (2014). 프레임 경쟁에 따른 언론의 보도 전략: 언론의 기사근거 제공과 익명 정보원 사용. <한국언론학보>, 58권 3호, 378-407.
- 이준웅 (2009). 뉴스 틀 짓기 연구의 두 개의 뿔. <커뮤니케이션 이론>, 5권1호, 123-166.
- 이준웅 (2001). 갈등적 이슈에 대한 뉴스 프레임 구성 방식이 의견형성에 미치는 영향: 내러티브 해석모형의 경험적 검증을 중심으로. <한국언론학보>, 46권 1호, 441-482.
- 이준웅 (1997). 언론의 틀짓기 기능과 여론의 변화. <언론과 사회>, 17권 가을호, 100-135.
- 진설아·허고은·정유경·송민 (2013). 트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구. <정보관리학회지>, 30권 1호, 285-302.
- 한국갤럽 (2017). 미세먼지에 대한 인식조사. <한국갤럽 데일리 오피니언>. 서울: 한국갤럽.
- 한혁·김영옥·금현섭 (2017). 공중상황이론을 바탕으로 한 PR전략 연구. <한국언론학보>, 61권 3호, 222-254.

- Bennet, W. L. (2004). *News: The politics of illusion*. Chicago: University of Chicago Press.
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 17-35.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health Psychology*, 26(2), 136.
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6), 570-606.
- Dynes, R. R., & Rodriguez, H. (2010). Finding and framing Katrina. *The Sociology of Katrina: Perspectives on a Modern Catastrophe*. Lanham, MD: Roman & Littlefield Publishers, Inc, 25-36.
- Entman, R. M. (1991). Symposium framing US coverage of international news: Contrasts in narratives of the KAL and Iran air incidents. *Journal of communication*, 41(4), 6-27.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
- Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of Communication*, 57(1), 163-173.
- Friedman, B., & Wenger, D. (1986). Local and national media coverage of disaster: A content analysis of the print media's treatment of disaster myths. *International Journal of Mass Emergencies and Disasters*, 4(3), 27-50.
- Gamson, W. A., & Lasch, K. E. (1981). The political culture of social welfare policy. *CRSO Working Paper #221*.
- Gephart Jr, R. P. (2004). Normal risk: Technology, sense making, and environmental disasters. *Organization & Environment*, 17(1), 20-26.
- Gilbert, C. (2005). Studying disaster: changes in the main conceptual tools. Ed. by Quarantelli, E.L., *What is a disaster?* (pp. 21-30). London: Routledge.

- Goltz, J. D. (1984). Are the news media responsible for the disaster myths?: A content analysis of emergency response imagery. *International Journal of Mass Emergencies and Disasters*, 2(3), 345-368.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267-297.
- Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication. *Journalism & Mass Communication Quarterly*, 93(2), 332-359.
- Hanson, J., & Hanson, K. (2006). The blame frame: Justifying (racial) injustice in America. *Harvard Civil Rights-Civil Liberties Law Reviews*, 41, 413.
- Holton, A., Weberling, B., Clarke, C. E., & Smith, M. J. (2012). The blame frame: Media attribution of culpability about the MMR-autism vaccination scare. *Health Communication*, 27(7), 690-701.
- Hong, L., & Davison, B. D. (2010, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88). ACM.
- Houston, J. B., Pfefferbaum, B., & Rosenholtz, C. E. (2012). Disaster news: Framing and frame changing in coverage of major US natural disasters, 2000 - 2010. *Journalism & Mass Communication Quarterly*, 89(4), 606-623.
- Iyengar, S. (1987). Television news and citizens' explanations of national affairs. *American Political Science Review*, 81(3), 815-831.
- Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, 12(1), 19-40.
- Kitzinger, J. (1999). Researching risk and the media. *Health, Risk & Society*, 1(1), 55-69.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176-187.
- Mohr, J. W., & Bogdanov, P. (2013). Introduction-Topic models: What they are and why they matter. *Poetics*, 41(6), 545-569.
- Pennington, H. (1999). The media and trust: E coli and other cases. Ed. by Bennett, P. and Calman, K.C., *Risk Communication and Public Health* (pp. 81-06). Oxford: Oxford University Press.

- Pidgeon, N., Kasperson, R. E., & Slovic, P. (Eds.). (2003). *The social amplification of risk*. Cambridge: Cambridge University Press.
- Quarantelli, E. L. (1985). What is disaster? The need for clarification in definition and conceptualization in research. Ed. by Barbara Sowder, *Disasters and mental health Selected contemporary perspectives* (pp. 41-73). Washington, D.C.: U.S. Government Printing Office.
- Reese, S. D., & Lewis, S. C. (2009). Framing the War on Terror: The internalization of policy in the US press. *Journalism*, 10(6), 777-797.
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515), 988-1003.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open ended survey responses. *American Journal of Political Science*, 58(4), 1064-1082.
- Roeser, S., Hillerbrand, R., Sandin, P., & Peterson, M. (Eds.). (2012). *Handbook of risk theory: Epistemology, decision theory, ethics, and social implications of risk* (Vol. 1). London and New York: Springer Science & Business Media.
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1), 103-122.
- Semktko, H. A. & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, 50(2), 93-109.
- Sjöberg, L. (2000). Factors in risk perception. *Risk Analysis*, 20(1), 1-12.
- Steuter, E., & Wills, D. (2009). *At war with metaphor: media, propaganda, and racism in the war on terror*. Plymouth: Lexington Books.
- Tierney, K., Bevc, C., & Kuligowski, E. (2006). Metaphors matter: Disaster myths, media frames, and their consequences in Hurricane Katrina. *The Annals of the American Academy of Political and Social Science*, 604(1), 57-81.
- Tierney, K. J. (2007). From the margins to the mainstream? Disaster research at the crossroads. *Annual Reviews of Sociology*, 33, 503-525.
- Trainor, J., Barsky, L., & Torres, M. (2006). Disaster realities in the aftermath of



Hurricane Katrina: Revisiting the looting myth. *Quick Response Report*, 184.  
Natural Hazards Center, University of Delaware.

Vasterman, P., Yzermans, C. J., & Dirkzwager, A. J. (2005). The role of the media and media hypes in the aftermath of disasters. *Epidemiologic Reviews*, 27(1), 107-114.

Weinstein, N. D. (1984). Why it won't happen to me: perceptions of risk factors and susceptibility. *Health Psychology*, 3(5), 431-457.

Weinstein, N. D., & Klein, W. M. (1996). Unrealistic optimism: Present and future. *Journal of Social and Clinical Psychology*, 15(1), 1-8.

최초 투고일 2018년 6월 15일

게재 확정일 2018년 7월 13일

논문 수정일 2018년 8월 2일

Abstract

# News frames in the coverage of fine-dust disaster

Application of Structural Topic Modeling

**June Woong Rhee**

Professor, Seoul National University

**Sung-hee Kim**

Ph. D. Candidate, Seoul National University

Employing two different methods of finding news frames in the coverage of fine-dust disaster, this study explored the ways in which major newspapers reported fine-dust from 2013 to 2017 in South Korea. The traditional way of identifying news frames is to conduct content analysis with a pre-defined set of frames considered to be themes and other related textual features. In addition to employing this traditional content analytic method, this study also undertook an algorithmic approach to finding underlying themes of a news corpus using the technique of structural topic modeling (STM), which made it possible to make a comparison between the results from the two approaches. Five different sets of underlying topics, defined as news frames in this study, were identified from a corpus comprising of 3,636 fine-dust related articles. Four different thematic news frames were found using a manual coding of news frames on 300 cases systematically sampled from the 3,636 articles. Similarities and dissimilarities between two sets of frames were examined and accounted for by methodological factors concerning data-driven vs. concept-driven content analysis. The findings of frame distributions across different news media and time periods were discussed in terms of media factors as well as issue characteristics of the fine-dust disaster.

**Keywords:** fine-dust, disaster, risk perception, news frames, structural topic modeling (STM)