

IPCC 정책결정자를 위한 요약 보고서의 계량 텍스트 분석: 기후과학 담론의 주제 구성 변화

김선희* · 윤순진**†

*중앙대학교 영어영문학과 교수, **서울대학교 환경대학원 교수

A quantitative text analysis of the IPCC summaries for policymakers: Changes in the topical composition of climate science discourse

Kim, Sun-Hoi* and Yun, Sun-Jin**†

*Professor, Dept. of English Language and Literature, Chung-Ang University, Seoul, Korea

**Professor, Graduate School of Environmental Studies, Seoul National University, Seoul, Korea

ABSTRACT

This study investigates the topical composition of the IPCC Synthesis Reports' Summaries for Policymakers (SPMs) from the Third Assessment Report (TAR) through the Sixth (AR6). Employing a mixed-method approach that integrates Latent Dirichlet Allocation (LDA) topic modeling and log-ratio-based text scatterplot visualization, the study aims to uncover shifts in both topical composition and linguistic emphasis over time. LDA topic modeling reveals a noticeable transition from the early emphasis on physical and natural science-based explanations to more recent attention toward climate policy and action, mitigation, adaptation, and governance. To determine whether these apparent shifts reflect statistically meaningful trends, the study applies bootstrapping and ANOVA, including a five-sentence chunk-based analysis, to topic proportions across reports. The results show significant difference in topics related to 'physical influence of climate change' and 'climate policy and action,' highlighting the change of topic focus in AR6. The log-ratio visualization further identifies words that experienced prominent increase or decrease in usage across reporting periods. Importantly, it allows the detection of conceptually salient terms that are not necessarily frequent but are strongly associated with a specific temporal or discursive context. This analysis captures the emergence of new framing devices (e.g., "net zero," "climate resilient development") and the disappearance of outdated terms (e.g., "SRES," "stabilization"). Together, these findings contribute to a deeper understanding of the expansion, evolution, and diversification of climate science over time, reflecting the accumulation of scientific evidence, the shift of priorities, and the growing urgency of global climate governance.

Key words : IPCC, Assessment Report, Summary for Policymakers, Climate Change, Global Warming, LDA Topic Modeling, Log-Ratio, Text Scatterplot, ANOVA, Climate Policy, Climate Action

1. 서론

기후변화와 그 원인인 인간의 영향과 책임, 기후변화가 자연과 인간 사회에 미치는 심각한 위협에 대해 여전히 냉소적·회의적·부정적 태도를 보이는 이들도 있다. 그러나

다수의 사람들은 기후변화의 실재와 심각성을 인식하고 인정하며, 위기에 적극 대응해야 한다는 데 동의하고 있고, 이러한 인식은 시간이 지날수록 강화되는 추세이다 (Furnham, 2024; Gallup Korea, 2024; WIN, 2024). 기후 위기에 대한 인식 변화는 사람들이 일상에서 폭염·집중호

†Corresponding author : ecodemo@snu.ac.kr (Graduate School of Environmental Studies, Bldg. #82, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea. Tel. +82-2-880-9391)

ORCID 김선희 0009-0005-7825-0397

윤순진 0000-0002-0461-0151

우·가뭄·산불 등 극단적인 기상 현상이 야기하는 위험을 직접 경험하는 데 더해, 기후과학 전문가들의 축적된 연구 성과와 지식이 미디어·정책·교육을 통해 지속적으로 전파되어 왔기 때문으로 볼 수 있다. 다시 말해서, 개인적·사회적 체험과 과학적으로 검증되어 축적된 기후변화 관련 지식이 맞물려 서로 영향을 주고받으며 기후위기에 대한 사회적 이해가 심화되어 왔다.

검증되고 축적된 기후변화 관련 지식이 공적 논의와 검토, 합의 과정을 거쳐 정리되어 발표되는 대표적 문서가 기후변화에 관한 정부간 협의체(Intergovernmental Panel on Climate Change, IPCC)의 기후변화 평가보고서(Assessment Report, AR)인데, 1990년부터 2023년까지 약 5~9년 간격으로 여섯 차례 발간되었다. 특히 IPCC 보고서 가운데 종합보고서(Synthesis Report, SYR)의 정책결정자를 위한 요약본(Summary for Policymakers, SPM)은 기후과학 관련 모든 학문 분야와 그 하위 분야의 논의와 결론을 압축적으로 묶어, 모든 수준의 정부와 정책결정자에게 기후과학 연구의 최신 지식을 종합하여 전달할 목적으로 작성된 문서이다(Barkmeyer et al., 2016; Yun and Kim, 2024a). 따라서 SPM 담론이 어떤 주제에 어느 정도의 비중을 두는지, 그 비중의 시계열적 변화가 어떠한지를 파악하면, 기후과학 전문가들이 평가 주기별로 무엇을 핵심 주제로 다루었는지와 핵심 주제의 우선순위가 시간에 따라 어떻게 변화하여 왔는지를 가늠할 수 있다. 더 나아가 기후과학 전문가들이 평가보고서를 통해 각 시기마다 정책결정자와 미디어, 일반 대중에게 어떤 주제를 더 중점적으로 전달하려 했는지도 함께 보여 줄 수 있다. 또한 현재의 정책 우선순위 설정뿐 아니라 대중 소통과 교육에서 강조할 내용의 선정과 향후 전달 전략의 설계에까지 구체적이고 실질적인 근거로 기능할 수 있다.

이와 같은 관점에서 이 연구는 IPCC 평가 주기 간 SPM 담론 주제 비중의 시계열적 변화를 분석하는 것을 목적으로 한다. 1차 평가보고서(FAR, 1990)와 2차 평가보고서(SAR, 1995)는 오늘날과 같은 단일 SYR 내 통합 SPM 체계를 갖추지 못했고 이후에 정립된 불확실성 표현에 대한 지침이 동일하게 적용되지 않았다(Janzwood, 2020; Mach et al., 2017; Mastrandrea et al., 2010; Yun and Kim, 2024a). 따라서 동일한 체계와 형식을 가지고 작성된 담론 텍스트를 바탕으로 장기간의 변화를 안정적으로 비교하기 위해 이 연구에서는 3차 평가보고서(TAR, 2001)부터 6차 평가보고서(AR6, 2023)까지 SYR의 SPM(IPCC, 2001, 2007, 2014, 2023)으로 분석 대상 범위를 한

정하였다. 이후 이 글에서 다루는 SPM은 각 3차부터 6차에 이르는 평가 주기별 AR 가운데 실무집단(Working Group, WG)이 개별적으로 작성한 AR의 SPM이 아니라 SYR의 SPM을 일컫는다.

이 연구에서는 주제 비중의 시계열적 변화를 살펴보기 위해 토픽 모델링(topic modeling)과 *log-ratio* 텍스트 산점도 시각화(text scatter visualization)를 결합한 계량 텍스트 분석을 수행한다. 단순한 어휘 나열이나 사례 열거를 넘어, 텍스트 내에서 지식이 어떤 주제로 묶이고 그 비중이 어떻게 달라지는지를 계량적으로 살펴야 SPM 담론의 변화를 구체적으로 파악할 수 있다. 토픽 모델링은 텍스트 집합에 숨어 있는 핵심 주제어와 잠재적 주제 구조, 비중을 추정하여 어떤 단어/용어가 각 주제를 구성하는 핵심 개념어인지, 어떤 주제가 상대적으로 더 조명되거나 또는 덜 조명되는지를 보여 준다. 산점도 시각화는 평가 주기 간 상대적으로 두드러진 단어/용어의 차이를 직관적으로 보여 주어 특정 시기에 부상하거나 퇴조한 개념을 확인하는 데 유용하다. 두 기법을 함께 사용하면, 주제 구조의 큰 그림과 핵심 단어의 시계열적 세부 변화를 동시에 파악할 수 있다. 이 연구의 구체적인 연구 질문은 다음과 같다.

- 연구 질문 1: TAR에서 AR6까지 SPM을 관통하는 핵심 주제들은 무엇인가?
- 연구 질문 2: SPM에 담긴 주제들의 비중은 평가 주기별로 어떻게 변화하였는가?
- 연구 질문 3: 평가 주기 간 주제 비중의 변화는 통계적으로 유의한가?
- 연구 질문 4: 시기별로 부상하거나 퇴조한 핵심 개념어는 무엇인가?
- 연구 질문 5: 이러한 변화가 제시하는 기후변화에 관한 담론적 함의는 무엇인가?

2. 연구 배경: 담론 분석의 의의와 선행연구 검토

담론 분석은 문장·단어 구조와 형태 분석에 초점을 맞추기보다는 현실 맥락에서 언어가 어떻게 내용을 조직하고 의미를 만들어 내는지에 주목한다(Brown and Yule, 1983). 전통적 담론 분석은 의사소통 수단으로서의 언어에 초점을 맞춰 언어학적 관점에서 의사소통 기능과 언어

형식 사이의 관계를 주로 분석하였다. 그러나 시간이 흐르면서 담론을 사회적 행위와 의미 구성의 과정으로 보고 언어가 가치·정체성·행위를 어떻게 드러내는지 분석하는데까지 담론 분석의 관심과 범위가 확장되었다(Fairclough, 2013; Gee, 2010; van Dijk, 2009, 2021). 특히 비판적 담론 분석(Critical Discourse Analysis, CDA)은 언어를 단순한 의미 전달 수단이 아니라 사회적 권력 관계와 이념을 재생산하는 장치로 보고, 특정 언어 사용이 사회적 불평등이나 제도적 권위를 어떻게 정당화하는지를 탐구한다(Fairclough, 2013; Kim and Yun, 2016). CDA관점에서 보자면, 언어는 단순한 정보 전달 수단이나 현실 묘사 수단이 아니며 언어로 구성되는 담론은 단지 현실을 반영한다기보다는 현실을 구성해 나가는 실천 활동으로 이해할 수 있다.

더 나아가 담론 분석은 사회적 차원에서 언어가 집단 정체성과 가치체계를 구축하는 과정에 주목한다. 이는 기후과학 담론 분석에도 중요한 시사점을 제공한다. 기후변화 관련 과학 지식은 중립적으로 제시되는 것처럼 보이지만 실제로는 특정한 사회적 정치적 의미 맥락 속에서 생산되고 유통된다. 예컨대 IPCC 보고서와 같은 권위 있는 과학적 텍스트는 단순히 사실을 기술하는 것을 넘어, 국제사회가 기후위기를 어떻게 이해하고 대응해야 하는지에 관한 정책적 방향성을 제시하는 담론적 행위로 작용한다. 따라서 기후과학 담론 분석은 언어가 과학적 사실을 조직하는 동시에, 정책적 의제 설정과 사회적 정당화 과정에 어떤 역할을 하는지 탐구하는 데 의미가 있다.

기후과학 담론에 관한 기존 연구는 주로 미디어 영역에 초점을 맞추어 왔다. 다수의 선행연구가 언론 기사와 사설, 온라인 플랫폼을 대상으로 의제 설정과 프레이밍, 설득 전략이 대중 인식과 정책 지지에 미치는 영향을 추적하였다(Boykoff and Boykoff, 2007; Isopp, 2024; Yun et al., 2014). IPCC 평가보고서를 다룬 연구들 또한 담론 표현의 정치적 함의에 초점을 맞추거나(Kanerva and Krizán, 2021; Penz, 2022), 문법 구조 또는 단어 선택, 불확실성 표현 사용에 주목하여 과학적 진술의 중립성과 확실성, 신중성을 분석하는 데 주력하였다(Dormer, 2020; Poole and Hayes, 2022).

이에 반해, 보고서 내부에서 어떤 주제들이 어떤 비중으로 배열되고 주제 비중이 평가 주기마다 어떻게 달라졌는지, 다시 말해서 주제 구성의 시계열적 변화를 정량적·체계적으로 분석하고 그러한 변화가 내포하는 담론적 함

의가 무엇인지에 대해 분석한 연구는 많지 않다. 특히 SPM 담론처럼 형식과 승인 절차가 일관된 텍스트를 대상으로 장기간에 걸친 주제 비중의 변화를 비교하고 검증한 사례는 매우 드문 편이다. 이 연구에서는 이러한 연구 공백에 주목하여, TAR (2001)부터 AR6 (2023)까지 SYR의 SPM 담론을 대상으로 토픽 모델링과 텍스트 산점도 시각화를 결합하고 평가 주기별 주제 비중의 차이를 통계적으로 확인함으로써 SPM 담론의 강조점이 어떻게 이동해 왔는지를 보여 주고자 한다.

IPCC 평가보고서를 대상으로 자연어 처리(Natural Language Processing, NLP) 기법인 토픽 모델링과 텍스트 산점도 시각화를 적용한 선행연구로는 Kramar et al. (2024)을 들 수 있다. Kramar et al. (2024)은 IPCC AR5 (2014)와 AR6 (2023)의 SYR을 대상으로 세 가지 토픽 모델링(BERTopic, NMF, LDA)과 텍스트 산점도 시각화를 적용하여 보고서 간 주제와 용어 변화를 비교하였다. 이들은 두 기법을 IPCC 텍스트 분석에 도입해 보고서 간 핵심 단어 구성의 차이를 직관적으로 보여 주었다는 점에서 방법론적 의의가 크다. 그러나 분석 범위가 AR5와 AR6에 국한되고 분석의 초점이 핵심 단어 목록 도출에 맞춰져 있어 각 보고서에서 어떤 주제들이 어느 정도의 비중으로 구성되었는지와 그 비중이 시간에 따라 어떻게 달라지는지를 파악하기 어려워 시간에 따른 담론 변화의 의미를 해석할 수 없다는 한계가 있다. 또한 보고서 간 차이에 대한 통계적 유의성 검정이 제시되지 않아, 관찰된 차이가 우연적인 것인지 의미 있는 변화인지를 명확하게 판단할 수 없다. 방법론적으로도 전처리 과정에서 분석 단위를 개별 단어로 설정하여, 이미 널리 쓰이는 합성어(예: *climate change*, *global warming*, *sustainable development*, *net zero*)와 기후변화와 관련성이 높은 숫자와 단위(예: *1.5°C*, *2°C*, *>50%*)를 제외함으로써 자연과학 담론에서 주제를 식별하는 핵심 표시가 충분히 반영되기 어려운 한계가 있다.

Yun and Kim (2024b)도 이 연구처럼 SPM 담론의 시계열적 주제 변화에 주목하였는데, 이들은 SPM 담론을 대상으로 시기별 핵심 단어의 빈도 변화를 체계적으로 분석하였다. 고빈도 단어 기반 접근은 텍스트 전반의 단어 구성을 포괄적으로 제시하고 평가 주기별로 눈에 띄는 관심사가 무엇인지를 가늠하게 해 준다. 그러나 빈도만으로는 문서 내부의 주제 구성을 포착하기 어렵고 저빈도지만 특정 시기를 대표하는 표시 단어의 부상이나 퇴조를 파악할 수 없다. 또한 숫자·단위·약어처럼 자연과학 담론의 핵

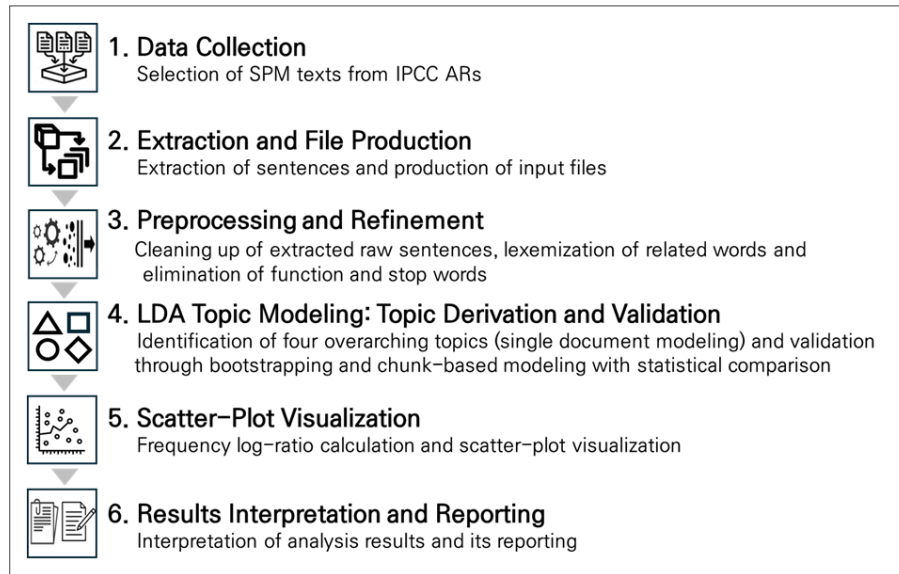


Fig. 1. Flowchart of the research process

심 구성요소를 변별력 있게 다루기 어렵기 때문에, Yun and Kim (2024b) 역시 전체 담론이 어떤 주제로 구성되었는지와 그 비중이 시간에 따라 어떻게 변화했는지를 설명하는 데에는 충분하지 않다. 이들의 연구는 토픽 모델링과 산점도 시각화를 도입하지는 않았지만, 단어 빈도의 시계열적 변화를 통해 담론의 변화를 읽어내려 했다는 점에서 이 연구와 문제의식은 맞닿아 있으면서도, 분석 방법과 주제 구성의 파악 정도에서는 차이가 있다.

정리하면, Kramar et al. (2024)과 Yun and Kim (2024b)은 IPCC의 AR 텍스트 분석에서 주제에 초점을 맞추었다는 점에서 이 연구와 공통성이 있다. 그러나 전자는 범위(AR5-AR6 한정)와 개념 단위 처리(합성어와 숫자, 단위 배제), 주제 비중의 도출 부재와 통계적 유의성 미확인이라는 제약이 있고, 후자는 고빈도 단어 중심의 분석으로 인해 주제 구성과 비중 변화를 입체적으로 분석하기 어렵다는 한계가 있다. 이 연구에서는 두 선행연구의 시사점을 바탕으로 대상 범위를 TAR부터 AR6까지의 SPM 담론으로 확장하고, 합성어·숫자·단위·약어를 보존하는 전처리, LDA 기반 주제 비중 도출, 문장 부트스트래핑 및 5문장 묶음 학습 결과에 대한 통계적 유의성 검증, *log-ratio* 산점도 시각화를 결합한다. 이를 통해 주제 비중의 시계열적 변화와 강조점의 이동을 객관적이고 해석 가능한 방식으로 제시하고자 한다.

3. 연구 방법

3.1. 연구 절차

이 연구는 여섯 단계로 진행하였다. 첫 번째로는 IPCC 홈페이지(ipcc.ch)에 게시된 평가보고서들을 다운로드하여 TAR에서부터 AR6까지 SYR 중 SPM 텍스트를 한 문장이 한 셀에 들어가는 엑셀 파일(csv 형식 파일)로 만들었다. 두 번째로는 이 연구의 계량 분석 도구인 R(4.4.2 버전)의 입력 파일을 형성하기 위해 이 파일로부터 코딩이 포함된 파일을 다시 만들었다. 세 번째로는 분석 결과의 왜곡을 가져올 수 있는 잡음(noise)을 피하기 위해 전처리와 정제 작업을 진행하였고, 네 번째로는 LDA 토픽 모델링을 통해 전체 SPM을 관통하는 핵심 주제 4개를 선정하고 부트스트래핑, 다섯 문장 묶음 단위 LDA 학습을 실시한 후 그 결과를 토대로 SPM별 주제 분포 차이의 통계적 유의성을 검증하였다. 다섯 번째로는 중요한 결과들을 꺾은선 그래프와 상자 그래프로 시각화하였고, 마지막으로 분석 결과 중 핵심 사항, 시사점, 한계에 관한 토론과 논의를 정리하였다. 이러한 연구 진행을 요약해서 정리하면 Fig. 1과 같다.

3.2. 연구 방법

IPCC 홈페이지(ipcc.ch)에서 다운로드한 SPM 본문 텍스트를 텍스트 파일(txt 파일)로 변환하고 이 파일을 다시

한 문장을 하나의 셀에 할당한 엑셀 파일(csv 파일)로 만들어 1차 자료로 삼았다. 전처리 과정으로 1차 자료에서 관사, 전치사, 접속사, 대명사와 같은 기능어, *be, have, get*과 같은 기본 동사, *will, would, can, could, may* 같은 양상동사를 불용어로 취급하여 제거하였다. 단어 중 서로 의미 관련성은 있으나 문장에 따라 동사로 사용되기도 하고 명사로 사용되기도 한 단어와 단수형과 복수형이 함께 관찰된 단어는 명사 어휘소(*lexeme*)로 단일화하였고, 동사활용형은 동사 어휘소로 단일화하였다. 예를 들면, *reduce, reduces*와 *reduction*은 *reduction*으로, *rise, rises, rising*은 *rise*로 단일화하였다. 동일 개념을 나타내지만 두문자(acronym) 용어로 사용되기도 하고 복수의 단어 합성으로 사용되기도 한 것들은 두문자 용어로 단일화하였다. *GHG*와 *greenhouse gas*를 *GHG*로, *CO₂*와 *carbon dioxide*를 *CO₂*로 단일화한 것이 대표적이다. 앞에서 언급했지만, 복수의 단어 합성으로 널리 통용되고 있어서 합성어로 처리해야 하는 단어 연쇄는 Kramar et al. (2024)과 달리 한 단어로 처리하였다. 문장 마침표, 쉼표, 그 밖의 용어 정리를 포함한 전처리 작업을 거쳐 완성된 4개의 엑셀 파일을 토픽 모델링과 텍스트 산점도 시각화를 위한 R 작업 수행의 입력 말뭉치 자료로 삼았다. 이 말뭉치 개요는 Table 1과 같다.

Table 1의 통계는 한 보고서에 중복된 문장이 있을 경우에는 한 문장으로만 처리하고, IPCC 평가 체계를 나타내는 *likely, very likely, high agreement, very high confidence, virtually certain*을 포함한 18개의 불확실성 표현(*uncertainty expression*)에 사용된 단어들은 제외한 결과이다: IPCC가 제시한 불확실성 표현은 모두 24개지만 (Yun and Kim, 2024a), 실제 TAR, AR4, AR5, AR6의 각 SYR SPM에 사용된 것은 18개이다. IPCC 평가 체계를 나타내는 이 불확실성 표현들은 이들이 포함된 문장의 내용을 평가하는 표현들이나 이들을 분석 대상에 포함시키면 주제 분석과는 관계가 없는 이 단어들의 동시출현 관

계와 빈도가 분석 결과에 반영되어 분석 결과를 왜곡할 수 있다. 전처리 후 완성된 말뭉치에 포함된 텍스트의 양은 문장 기준으로는 397문장으로 구성된 TAR가 가장 많고 219문장으로 구성된 AR4가 가장 적었다. 하지만 AR4를 제외한 세 SPM 사이의 분량 차이는 단어 기준으로 텍스트의 양을 평가하면, AR6가 단어 유형(*word type*)과 단어 토큰(*word token*) 모두 가장 많았다.

토픽 모델링을 위한 R 코드는 입력 자료를 호출한 후, *character*-벡터 형성, 레이블 순서 지정, 공백 기준 토큰 형성, 말뭉치와 토큰 매트릭스 완성(*VCorpus*와 *DocumentTermMatrix* 함수 사용), 주제(토픽) 수 결정, LDA 주제 모형 학습, 주제별 단어 확률과 상위 단어 추출, SPM별 주제 분포와 평균 비중 계산, 평균 비중의 시계열적 변화를 시각화하는 순서로 구성되었다. 최적의 주제 수를 결정하는 데에는 R-패키지 중 하나인 *ldatuning*을 설치하고 *FindTopicsNumber()* 함수를 사용하여 얻은 각기 다른 4개의 지표 *Griffiths2004, CaoJuan2009, Arun2010, Deveaud2014*의 결과를 참고하였다. 전체 SPM을 입력 자료로 삼아 수행한 LDA 모델 학습 결과에 따르면, 기준이 되는 4개의 지표가 추천한 결과에 차이가 있었다. 결과값이 낮을수록 좋은 *CaoJuan2009*와 *Arun2010*의 결과값은 3과 4, 4와 5, 6과 7 사이에서 교차하였고 결과값이 높을수록 좋은 *Griffiths2004*와 *Deveaud2014*의 결과값은 3과 4 사이에서 교차하였다. 이러한 결과를 종합해서 이 연구에서는 결과값의 공통적 교차 지점인 3과 4 중 *CaoJuan2009*를 제외한 세 지표의 결과값들이 서로 가장 인접해 있는 4를 가장 적절한 주제의 개수, 즉 *k*-값으로 결정하였다. 토픽 모델링이 각 주제의 명칭을 명시적으로 제시하지 않기 때문에, 토픽 모델링 결과로 각 주제에 할당된 단어들의 사용 빈도를 고려하면서 이 연구의 연구자들이 토론을 통해 각 주제의 명칭을 정하였다.

이 연구에서는 보고서 간 비교의 신뢰성과 타당성을 높이기 위해 두 가지 상호보완적 방식의 토픽 모델링을 수

Table 1. Post-processed SPM corpus

SPM	Sentence Text	Word Type	Word Token
TAR	397	1211	5303
AR4	219	857	2657
AR5	358	1154	4595
AR6	378	1281	5719

Note: Word Type: A unique word, regardless of how many times it occurs in the text

Word Token: Every instance of a word as it occurs in the text

행하였다. 첫 번째 방식은 TAR, AR4, AR5, AR6 SPM 전체를 구성하는 네 개의 SPM 각각을 하나의 단일 문서로 간주한 상태에서 LDA 주제 모형을 학습하는 것이다. SPM 하나를 하나의 단일 문서로 취급하기 때문에, 이 방식은 각 SPM 전체 문맥의 흐름을 보존한 상태에서 주제 구조를 파악할 수 있다는 장점이 있다. 그러나 전체 분석 대상이 단 4개의 문서에 불과하게 되어 통계적 유의성 검정을 위한 최소 표본 수를 확보할 수 없다는 단점이 있다. 따라서 이 방식의 장점을 살리면서도 통계적 유의성 검정 문제를 해결하기 위해 이 연구에서는 해당 모델을 바탕으로 문장 단위 부트스트래핑(bootstrapping)을 수행하였다. 각 SPM에서 20개의 문장을 복원추출(resampling)하여 새 문서를 구성한 뒤 학습된 LDA 모델의 후방확률(posterior probability)을 추정하여 각 주제에 대한 비중을 계산하고 이를 100회 반복하여 주제별 평균 비중과 분산을 추정하였다.

두 번째 방식은 네 개의 SPM 각각에 대해 5개 문장을 한 묶음(chunk)으로 하는 다수의 묶음 집단을 생성하고 이들을 새로운 분석 단위로 하여 LDA 모델을 학습하였다. 이후 각 묶음 문서의 주제 분포를 수집하여 SPM 단위의 평균 주제 비중을 산출하였다. 이 방식에서는 SPM별로 다수의 주제 분포 값이 확보되기 때문에 주제 비중의 평균과 분산을 기반으로 한 통계적 비교가 가능하다. 이 방식은 단위 분석 수를 늘림으로써 통계적 유의성 검정 문제가 해결된다는 장점이 있는 데 반해, 각 SPM을 분할하여 분석하기 때문에 문맥의 흐름이 끊어질 수 있다는 단점이 있다. 따라서 이 연구에서는 네 개의 SPM 각각을 하나의 단일 문서로 간주한 상태에서 수행한 LDA 학습 결과를 통해 주제들의 시계열적 변화 추세를 파악하고 이에 대한 부트스트래핑 결과와 다섯 문장 묶음 기준 LDA 학습 결과를 통해 변화 추세에 대한 통계적 유의성을 검증한다.

log-ratio 기반 텍스트 산점도 시각화 분석은 단어 수준에서 출발한다는 점에서 토픽 모델링 분석과 공통되지만, ‘단어의 동시 출현’이 아닌 빈도 기반의 ‘상대적 분포’를 중심으로 한다는 점에서 차별성을 지닌다. 한 문서에서 그리 높지 않은 빈도를 보이는 단어일지라도 다른 문서에서 드물게 사용되었거나 전혀 사용되지 않았다면 해당 단어는 두 문서에서 공통적으로 많이 출현한 고빈도 단어보다 두 문서의 특성 차이를 구별하는 중요한 지표가 될 수 있다. 이와 같은 단어는 두 문서에서 공통적으로 많이 등장한 고빈도 단어보다 $\log_2((A + 1)/(B + 1))$ 결과값이 훨씬

더 크거나 훨씬 더 작다. 따라서 *log-ratio* 결과값을 다차원적으로 시각화하면, 시기별로 부상하거나 퇴조한 개념을 식별할 수 있다.

텍스트 산점도 시각화를 위한 R 코드는 공백 기준으로 형성된 단어 토큰의 빈도를 계산하고 각 단어 토큰의 \log_2 상대적 비율을 계산하는 것을 핵심으로 구성되었다. \log_2 상대적 비율 계산에는 한 문서에만 출현하고 다른 문서에는 출현하지 않는 토큰도 포함되어 있다. 모든 토큰 레이블(문자화된 단어)을 산점도로 시각화하면 공간상의 제약으로 토큰 구별이 어려울 뿐 아니라 한 문서에서는 상대적 비중이 높으나 다른 문서에서는 상대적 비중이 낮은 단어들이 무엇인지 파악하기 어렵다는 문제가 있다. 이 문제를 해결하기 위해 두 문서에서 각각 상대적 비중이 높은 단어들만을 레이블로 표시하고 다른 단어들은 점으로 표시하도록 하는 R 코드를 작성하였다. 이에 따라 문자화된 단어들은 X-축 오른쪽 하단과 Y-축 왼쪽 상단을 중심으로 나타나고 나머지 단어들은 점들로 시각화되었다.

텍스트 산점도 시각화는 다음 두 가지 방식으로 이루어졌다. 첫째, 네 개의 SPM을 작성 기간을 기준으로 전반기와 후반기로 나누어서 TAR와 AR4의 SPM을 함께 묶고 AR5와 AR6의 SPM을 함께 묶어 두 그룹의 입력 자료 분석 결과를 2차원 텍스트 산점도로 시각화하였다. 둘째, 최근 기후변화 관련 개념들의 상대적 중요도에 어떠한 변화가 있었는지 살펴보기 위해 AR5와 AR6의 SPM을 입력 자료로 삼아 그 결과를 텍스트 산점도로 시각화하였다.

4. 연구 결과

4.1. 평가 주기별 주제 비중

각 SPM을 하나의 단일 문서로 간주한 상태에서 k -값을 4로 하여 LDA 주제 모형을 학습한 결과 각 주제에 할당된 상위 20개 단어는 Table 2와 같다.

<Topic 1>에 해당하는 상위 단어 중 ‘기후변화(climate change)’와 ‘변화(change)’는 기후변화와 그로 인한 변화, ‘관측하다(observe)’, ‘상승/상승하다(rise)’, ‘온도(temperature)’, ‘얼음(ice)’, ‘해수면(sea level)’, ‘대양(ocean)’은 온도 상승의 과학적 관측과 관측 대상 자연물, ‘영향(impact)’과 ‘효과(effect)’는 기후변화의 영향, ‘지구적(global)’과 ‘지역적(regional)’은 영향의 범위와 관련되기에 <Topic 1>의 명칭을 ‘기후변화의 물리적 영향(Physical Influence of

Table 2. LDA topics for TAR, AR4, AR5, and AR6

Topic	Topic Name	Top-20 Words
Topic 1	Physical Influence of Climate Change	climate change, change, impact, system, rise, global, climate, ice, sea level, effect, affect, rate, regional, natural, large, observe, vulnerability, temperature, ocean, magnitude
Topic 2	GHG Emissions and Pathway Scenarios	emission, GHG, warming, level, project, global, CO ₂ , limit, scenario, high, concentration, carbon, temperature, low, reduction, range, model, stabilization, effect, estimate, surface
Topic 3	Climate Risks on Ecosystems	increase, reduction, risk, human, energy, region, ecosystem, water, improvement, land, adverse, area, extreme, potential, use, health, decrease, future, event, significant
Topic 4	Climate Policy and Action: Mitigation and Adaptation	mitigation, adaptation, cost, technology, policy, sector, economic, action, development, effective, country, scale, social, benefit, option, achievement, response, implementation, local, finance

Table 3. Average topic proportions across IPCC reports

SPM	Physical Influence of Climate Change	GHG Emissions and Pathway Scenarios	Climate Risks on Ecosystems	Climate Policy and Action: Mitigation and Adaptation
TAR	0.256	0.251	0.244	0.249
AR4	0.260	0.252	0.249	0.239
AR5	0.248	0.251	0.248	0.253
AR6	0.233	0.251	0.254	0.262

Climate Change)’으로 정하였다. <Topic 2>의 명칭은 상위 단어군에 ‘배출(emission)’, ‘온실가스(GHG)’, ‘이산화탄소(CO₂)’, ‘탄소(carbon)’와 ‘시나리오(scenario)’, ‘감축(reduction)’, ‘모형(model)’, ‘추정/추정치(estimate)’ 등이 속해 있어서 ‘온실가스 배출과 경로 시나리오(GHG Emissions and Pathway Scenarios)’로 정하였다. <Topic 3>의 경우에는 ‘위험/위험성(risk)’, ‘인간(human)’, ‘생태계(ecosystem)’, ‘부정적(adverse)’, ‘건강(health)’, ‘미래(future)’, ‘물(water)’, ‘토지(land)’와 같은 단어들이 상위 단어군에 속해 있는데, 이들은 생태계 전반에 끼치는 기후변화의 위기 상황과 관련되어 있어 ‘생태계에 미치는 기후위험(Climatic Risks on Ecosystems)’이라고 정하였다. <Topic 4>에 포함된 상위 단어들은 전반적으로 자연과학적 개념보다는 기후변화 완화와 적응 정책을 논의할 때 주로 사용되는 사회과학, 생태경제학적 개념들을 나타내므로, ‘기후정책과 행동: 완화와 적응(Climatic Policy and Action: Mitigation and Adaptation)’으로 정하였다. 요약하면, TAR에서부터 AR6까지 SYR SPM 전체의 주제를 4개로 설정할 때, 4개 AR SYR의 SPM을 관통하는 주제는 기후변화의 물리적 영향, 기후변화에 절대적 역할을 하는 온실가스 배출과 배출 억제 시나리오, 기후변화 영향이 생태계 전반에 끼치는 위기 상황, 기후변화를 완화하고 위기 상황에 적응하는 데 요구되는 정책과 행동이라고 할

수 있다.

Table 3은 각 SPM이 위의 네 주제에 속할 확률의 평균 값을 제시한 것으로서, 위의 네 주제가 각 SPM에 얼마만큼의 비중으로 서술되어 있는지를 보여준다.

Fig. 2는 Table 3의 결과를 시계열적 꺾은선 그래프로 시각화한 것으로 기후변화 관심사의 시계열적 변화를 AR SPM의 주제 비중 변화에 대한 계량적 분석 결과를 통해 살펴볼 수 있다.

Table 3과 Fig. 2의 결과에 따르면, ‘온실가스 배출과 경로 시나리오’에 대한 관심은 보고서 전반에 걸쳐 비슷한 비중으로 일관되게 다루어지고 있으나(각각 0.251, 0.252, 0.251, 0.251), 나머지 세 주제에 대한 비중은 시간이 흐름에 따라 약간씩 달라지는 추세를 보인다. 주목할 만한 것은 ‘기후변화의 물리적 영향’과 ‘기후정책과 행동: 적응과 완화’가 시계열적으로 상반된 변화 추세를 보인다는 점이다. 초기 보고서인 TAR, AR4에서는 기후변화의 물리적 영향의 상대적 비중이 높았고(각각 0.256, 0.260), 완화와 적응 기후정책과 행동의 상대적 비중이 낮았다(각각 0.249, 0.239). 하지만 AR5를 거쳐 AR6에서는 전자의 상대적 비중이 낮아지고(각각 0.248, 0.233), 후자의 상대적 비중이 높아졌다(각각 0.253, 0.262).

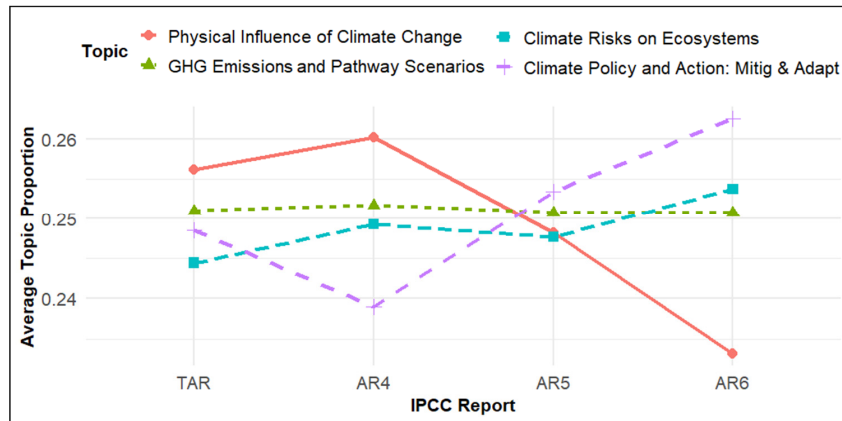


Fig. 2. Topic portion trends across IPCC reports

Table 4. Average topic proportions across IPCC reports (bootstrapped)

SPM	Physical Influence of Climate Change	GHG Emissions and Pathway Scenarios	Climate Risks on Ecosystems	Climate Policy and Action: Mitigation and Adaptation
TAR	0.276	0.261	0.223	0.241
AR4	0.303	0.261	0.247	0.189
AR5	0.252	0.246	0.238	0.264
AR6	0.176	0.254	0.269	0.301

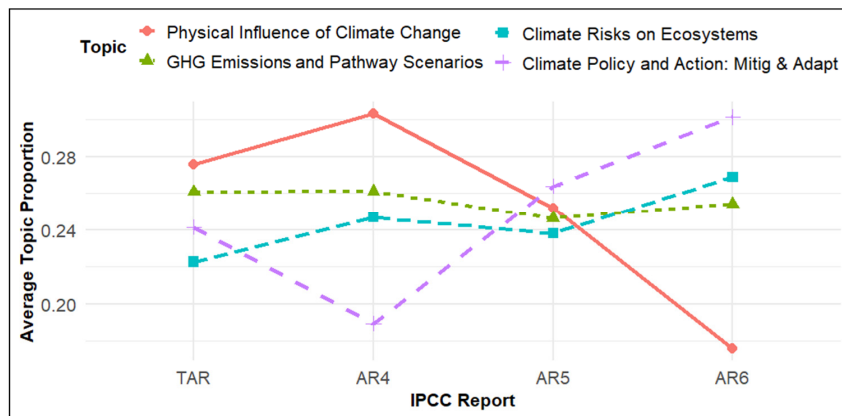


Fig. 3. Topic portion trends across IPCC reports (bootstrapped)

4.2. 주제 비중 변화의 통계적 유의성

앞의 결과는 전반적인 추세를 보여 주기는 하지만, 최소 표본 수를 확보할 수 없어 이 추세가 통계적으로 유의한지를 검정할 수는 없다. 이 문제를 해결하기 위해 각 SPM별로 100번의 부트스트랩을 반복하고 각 반복마다 20개의 문장을 복원추출하여 하나의 문서를 만들어서

LDA 모델의 후방 확률을 추정하는 부트스트래핑을 실시한 결과 각 SPM이 위의 네 주제에 속할 확률의 평균값은 Table 4와 같다.

Table 4의 부트스트래핑 결과값은 Table 3과 약간 다르지만, Fig. 3에서 보듯이, 시간의 흐름에 따른 변화 추세는 크게 다르지 않다.

이 결과를 상자 그래프를 사용하여 주제별 보고서 비중

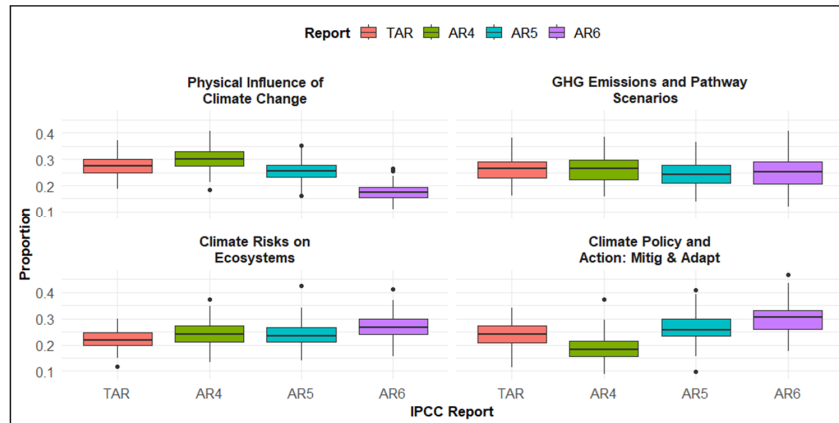


Fig. 4. Report proportions by topic (bootstrapped)

Table 5. ANOVA and Bonferroni-corrected pairwise comparisons across reports for topic proportions

Topic	ANOVA F (df = 3, 396)	p-value	Significant Differences (Bonferroni-adjusted p)
Physical Influence of Climate Change	191.1	< .001 ***	TAR ≠ AR4 ($p < .001$), TAR ≠ AR5 ($p < .001$), TAR ≠ AR6 ($p < .001$), AR4 ≠ AR5 ($p < .001$), AR4 ≠ AR6 ($p < .001$), AR5 ≠ AR6 ($p < .001$)
GHG Emissions and Pathway Scenarios	1.71	.164	No significant differences between any report pairs
Climate Risks on Ecosystems	18.59	< .001 ***	TAR ≠ AR4 ($p = .001$), TAR ≠ AR6 ($p < .001$), AR4 ≠ AR6 ($p = .003$), AR5 ≠ AR6 ($p < .001$)
Climate Policy and Action: Mitigation & Adaptation	84.45	< .001 ***	TAR ≠ AR4 ($p < .001$), TAR ≠ AR5 ($p = .012$), TAR ≠ AR6 ($p < .001$), AR4 ≠ AR5 ($p < .001$), AR4 ≠ AR6 ($p < .001$), AR5 ≠ AR6 ($p < .001$)

Note. p-values are Bonferroni-adjusted. Significant codes: *** $p < .001$, ** $p < .01$, * $p < .05$.

의 분포로 나타내면 Fig. 4와 같다.

시간의 흐름에 따른 주제 비중 변화의 통계적 유의성을 검증하기 위해 각 주제별로 일원분산분석을 실시한 결과, 주제2 ‘온실가스 배출과 경로 시나리오’를 제외한 세 주제에서 유의한 차이가 나타났다. 구체적인 분석 결과는 Table 5와 같다.

주제1 ‘기후변화의 물리적 영향’의 경우, 보고서 간 주제 비중에 유의한 차이가 있었으며($F(3, 396) = 191.1, p < .001$), Bonferroni 사후검정 결과 모든 보고서 쌍 간 비교에서 유의한 차이가 확인되었다($ps < .001$). 주제4 ‘기후정책과 행동: 완화와 적응’에서도 보고서 간 주제 비중에 유의한 차이가 있었으며($F(3, 396) = 84.45, p < .001$), 모든 쌍 간 비교에서 통계적으로 유의한 차이가 관찰되었다. 주제3 ‘생태계에 미치는 기후위험’은 $F(3, 396) = 18.59, p < .001$ 로 보고서 간 차이가 유의하였으며, 특히

AR6가 다른 보고서들과 유의한 차이를 보였다. 반면, 주제2 ‘온실가스 배출과 경로 시나리오’에서는 유의한 차이가 나타나지 않았다($F(3, 396) = 1.71, p = .164$).

각 AR의 SPM에 대해 5개 문장을 한 묶음으로 하는 다수의 묶음 집단을 생성하고 이들을 새로운 분석 단위로 하여 LDA 모델을 학습한 토픽 모델링에서도, 각 주제별 상위 20개 단어 구성에서는 앞의 분석 방식과 약간 차이가 있지만 전체 결과는 유사하였다. 즉, 각 SPM을 하나의 단일 문서로 간주하고 토픽 모델링을 수행한 경우와 5개 문장을 한 묶음으로 하여 LDA 모델을 학습한 토픽 모델링을 수행한 경우, 상위 20개 단어들은 약간 달랐지만 전체적으로 4개 주제별 단어 구성이 유사하였다. 5개 문장 묶음을 분석 단위로 했을 때 각 주제에 할당된 상위 20개 단어는 Table 6과 같다.

Table 7은 이 토픽 모델링 결과, 네 주제가 각 SPM에

Table 6. LDA topics for TAR, AR4, AR5, and AR6 (five-sentence chunk-based)

Topic	Topic Name	Top-20 Words
Topic 1	Physical Influence of Climate Change	increase, change, project, temperature, rise, sea level, warming, ice, large, ocean, surface, human, anthropogenic, natural, global, extreme, rate, observe, decrease, high
Topic 2	GHG Emissions and Pathway Scenarios	emission, global, GHG, level, reduction, limit, CO ₂ , scenario, concentration, carbon, warming, cost, range, model, stabilization, estimate, low, net, 2100, pathway
Topic 3	Climate Risks on Ecosystems	climate change, impact, risk, region, system, ecosystem, climate, human, affect, effect, regional, loss, water, adverse, health, increase, land, vulnerability, low, area
Topic 4	Climate Policy and Action: Mitigation and Adaptation	mitigation, adaptation, reduction, energy, policy, sector, technology, cost, option, economic, action, development, effective, benefit, country, environmental, potential, social, achievement, improvement

Table 7. Average topic proportions across IPCC reports (five-sentence chunk-based)

SPM	Physical Influence of Climate Change	GHG Emissions and Pathway Scenarios	Climate Risks on Ecosystems	Climate Policy and Action: Mitigation and Adaptation
TAR	0.254	0.243	0.252	0.251
AR4	0.321	0.237	0.234	0.208
AR5	0.268	0.237	0.222	0.275
AR6	0.190	0.251	0.242	0.318

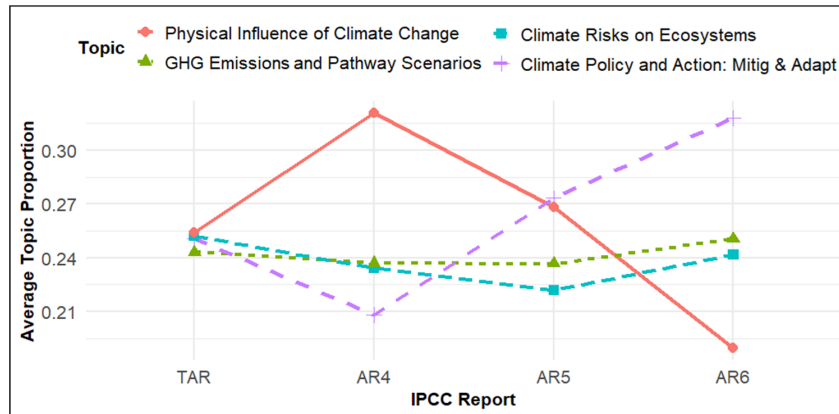


Fig. 5. Topic portion trends across IPCC reports (five-sentence chunk-based)

얼마만큼의 비중으로 서술되어 있는지를 보여준다.

Table 7의 결과값은 앞의 두 토픽 모델링의 결과와 약간 다르지만, Fig. 5에서 보듯이, 시간의 흐름에 따른 변화 추세는 크게 다르지 않다.

이 결과에 대한 통계적 유의성을 검정하기 위해 각 주제별로 일원분산분석을 실시한 결과, Table 8에 제시되었듯이, 문서 단위가 아닌 문장을 다섯 문장 단위로 쪼개어 묶은 묶음 기반의 분석에서도 AR SPM 간 일부 주제의 비중이 통계적으로 유의하게 달랐다.

주제1 ‘기후변화의 물리적 영향’에 대한 일원분산분석 결과, SPM 간 유의미한 차이가 확인되었으며($F(3, 268) = 12.54, p < .001$), 사후검정 결과 TAR과 AR6($p = .004$), AR4와 AR6($p < .001$), AR5와 AR6($p < .001$) 간 유의한 차이가 나타났다. 이는 AR6에서 해당 주제의 비중이 다른 보고서에 비해 상대적으로 낮다는 점을 시사한다. 주제4 ‘기후정책과 행동: 완화와 적응’에서도 유의한 차이가 나타났으며($F(3, 268) = 6.73, p < .001$), TAR과 AR6($p = .013$), AR4와 AR6($p < .001$) 간 차이도 유의하였다. 반면,

Table 8. ANOVA and Bonferroni-corrected pairwise comparisons by report (chunk-based)

Topic	ANOVA F (df = 3, 268)	p-value	Significant Differences (Bonferroni-adjusted p)
Physical Influence of Climate Change	12.54	< .001 ***	TAR ≠ AR4 ($p = .017$), TAR ≠ AR6 ($p = .004$) AR4 ≠ AR6 ($p < .001$), AR5 ≠ AR6 ($p < .001$)
GHG Emissions and Pathway Scenarios	0.234	.873	No significant differences between any report pairs
Climate Risks on Ecosystems	1.369	.253	No significant differences between any report pairs
Climate Policy and Action: Mitigation & Adaptation	6.734	< .001 ***	TAR ≠ AR6 ($p = .013$), AR4 ≠ AR6 ($p < .001$)

Note. p-values are Bonferroni-adjusted. Significant codes: *** $p < .001$, ** $p < .01$, * $p < .05$.

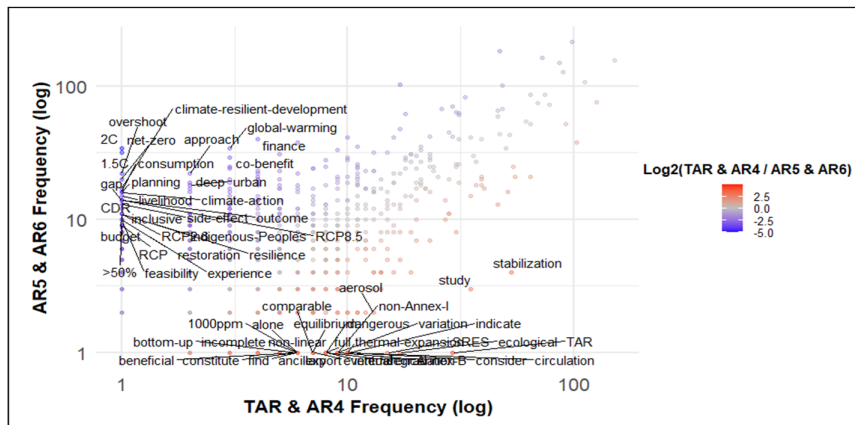


Fig. 6. Word contrast between early period and later period

주제2 ‘온실가스 배출과 경로 시나리오’와 주제3 ‘생태계에 대한 기후위험’에서는 보고서 간 유의미한 차이가 없었다($ps > .25$).

지금까지 살펴본 세 유형의 토픽 모델링 분석 결과를 보수적으로 해석하더라도, ‘기후변화의 물리적 영향’ 관련 주제는 보고서 작성 시기를 기준으로 전반기보다는 후반기에 비중이 줄어든 데 반해, ‘기후정책과 행동: 완화와 적응’ 관련 주제는 전반기보다 후반기에 비중이 늘어나는 추세인 것은 분명하다.

4.3. 핵심 개념어의 부상과 퇴조

TAR와 AR4를 함께 묶어 전반기 그룹의 말뭉치를 만들고 AR5와 AR6를 함께 묶어 후반기 그룹의 말뭉치를 만들어 두 그룹에 대한 2차원 텍스트 산점도를 시각화한 결과는 Fig. 6에 제시되어 있다.

Fig. 6은 전반기와 후반기 사이에 단어 빈도 간 로그 비율과 로그 스케일 분포를 비교한 결과를 시각화한 것이

다. X-축과 Y-축은 각각 전반기에 사용된 단어의 빈도와 후반기에 사용된 단어의 빈도를 로그 스케일로 나타낸 것이다. 각 점은 하나의 단어를 의미하고 해당 단어가 두 그룹에서 얼마나 자주 사용되었는지를 보여 준다. 점의 색상은 $\log_2(\text{TAR와 AR4에서의 사용 빈도} / \text{AR5와 AR6에서의 사용 빈도})$ 의 값을 반영하는데, 빨간 계열은 TAR와 AR4에서 많이 사용된 단어($\log\text{-ratio} > 0$), 파란 계열은 AR5와 AR6에서 많이 사용된 단어($\log\text{-ratio} < 0$), 회색은 두 그룹 간 사용 빈도가 비슷한 단어($\log\text{-ratio} \approx 0$)를 나타낸다. 색상이 점진적으로 변화하도록 설정되어 있어 중간값 부근에서는 보라색이나 주황색 점이 나타날 수 있다.

이 산점도에서 주목해야 하는 것은 점 대신 문자화되어 표시된 단어들이다. 이 단어들은 $\log\text{-ratio}$ 가 0으로부터 멀리 떨어진 순서대로 전반기 30개, 후반기 30개, 총 60개 단어를 필터링한 것이다. 이 단어들은 전체 단어 중 두 그룹 간 빈도 차이가 커서 해당 그룹에서는 상대적으로 중

Table 9. Relatively prominent words of early period and later period

Report	Range of Frequency	Range of Log-ratio	Top-30 Words
Early Period (TAR & AR4)	TAR & AR4: 6 ~ 28 AR5 & AR6: 0 ~ 1	2.58 ~ 4.86	TAR, ecological, SRES, stabilization, study, circulation, consider, indicate, thermal expansion, Annex B, non-Annex I, variation, dangerous, degradation, full, interference, ancillary, comparable, equilibrium, eventual, export, non-linear, aerosol, 1000 ppm, alone, beneficial, bottom-up, constitute, find, incomplete
Later Period (AR5 & AR6)	TAR & AR4: 0 ~ 3 AR5 & AR6: 8 ~ 39	-5.09 ~ -3.17	2°C, 1.5°C, overshoot, net-zero, climate resilient development, climate action, consumption, planning, livelihood, CDR, RCP8.5, gap, outcome, side effect, inclusive, approach, Indigenous Peoples, RCP2.6, budget, resilience, finance, experience, feasibility, restoration, co-benefit, deep, urban, > 50%, RCP

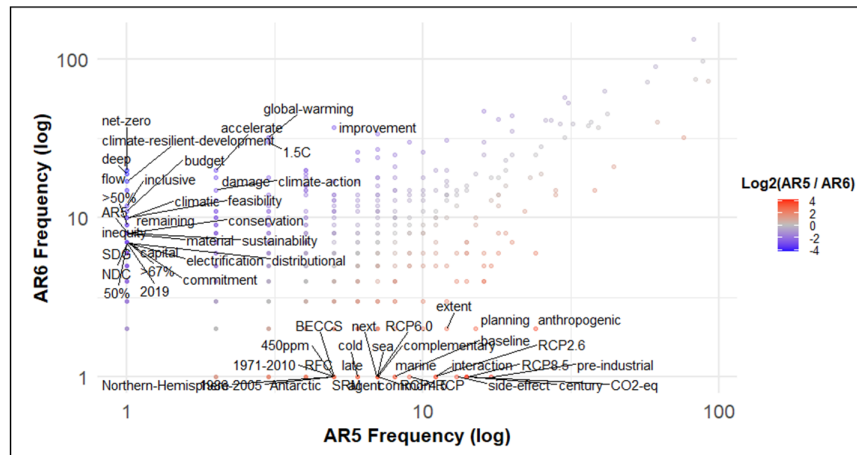


Fig. 7. Word contrast between AR5 and AR6

요한 단어라고 할 수 있다. 오른쪽 하단 30개 단어는 *log-ratio* 값이 큰 순서대로 선택된 것으로 후반기에 비해 전반기에 상대적으로 많이 사용된 것들이고, 왼쪽 상단 30개 단어는 전반기에 비해 후반기에 상대적으로 많이 사용된 것들이다. 단순 사용 빈도가 아닌 $\log_2(\text{TAR와 AR4에서의 사용 빈도} / \text{AR5와 AR6에서의 사용 빈도})$ 의 결과 값을 반영한 결과이므로, 한 그룹에서 단순 사용 빈도가 높지 않을지라도 다른 그룹에서 단순 사용 빈도가 그보다 현저히 낮거나 전혀 사용되지 않은 단어라면 이 단어군에 속할 수 있다. 이 단어들을 표로 정리하면 Table 9와 같다.

Table 9에 따르면, 전반기에는 주목받았으나 후반기로 가면서 덜 주목받는 개념들로 TAR와 AR4에서 상대적 비중이 높은 단어 30개는 TAR와 AR4에서는 6 ~ 28회까지

사용된 데 반해 AR5와 AR6에서는 전혀 사용되지 않았거나 단 1회만 사용되었다. 여기에는 교토의정서 기반 감축 책임 구분과 시나리오 체계에 관련된 용어인 ‘배출 시나리오 특별 보고서(Special Report on Emissions Scenarios, SRES)’, ‘부속서-B 국가(Annex B)’, ‘비부속서-I 국가(non-Annex I)’와 1992년에 채택된 유엔기후변화협약(United Nations Framework Convention on Climate Change, UNFCCC) 제2조 궁극적 목표 관련 단어인 ‘안정화/안정화하다(stabilization)’, ‘위험한(dangerous)’, ‘간섭(interference)’, 인용 서술이나 내용의 불확정성을 언급할 때 사용되는 단어인 ‘고려하다/여겨지다(consider)’, ‘연구(study)’, ‘표시하다/나타내다(indicate)’, ‘불완전한(incomplete)’, ‘변동성/변이(variation)’ 등이 포함되어 있다.¹⁾

1) SRES는 Special Report on Emissions Scenarios (배출시나리오 특별보고서)의 약자로, 2000년 IPCC가 발간한 특별보고서이다. 이 보고서에서 제시된 A1, A2, B1, B2 시나리오 체계가 TAR와 AR4에서 주요 주요 배출 시나리오 체계로 활용되었다. SRES 시나리오는 온

Table 10. Relatively prominent words of AR5 and AR6

Report	Range of Frequency	Range of Log-ratio	Top-30 Words
AR5	AR5: 4 ~ 16 AR6: 0 ~ 1	2.32 ~ 4.09	pre-industrial, CO ₂ -eq, RCP8.5, century, side effect, anthropogenic, RCP2.6, interaction, RCP, baseline, marine, planning, RCP4.5, RCP6.0. common, complementary, next, sea, agent, cold, extent, late, 1997-2010, 1986-2005, 450ppm, Antarctic, BECCS, northern hemisphere, RFC, SRM
AR6	AR5: 0 ~ 4 AR6: 7 ~ 36	-4.32 ~ -2.81	net-zero, deep, climate resilient development, flow, inclusive, budget, global warming, feasibility, climatic, accelerate, 1.5°C, remaining, > 50%, damage, sustainability, material. conservation, AR5, climate action, improvement, inequity, electrification, distributional, commitment, capital, SDG, NDC, > 67%, 50%, 2019

반대로, AR5와 AR6에서 상대적 비중이 높은 단어 30개는 AR5와 AR6에서는 8~39회까지 사용된 데 반해 TAR와 AR4에서는 전혀 사용되지 않았거나 많아야 3회 이하로 사용되었다. 이 단어들은 후반기에 들어서면서 새로이 주목받거나 개념화된 단어들이다. 파리협정(Paris Agreement) 이후 나타난 감축 목표와 온실가스 농도 변화 경로 시나리오에 관련된 용어 또는 단어인 1.5°C, 2°C, ‘탄소중립(net-zero)’, ‘한계초과(overshoot)’, ‘대표농도경로(Representative Concentration Pathways, RCP)’, RCP2.6, RCP8.5, > 50%(발생가능성 50 초과), ‘이산화탄소 제거(Carbon Dioxide Removal, CDR)’ 등이 포함되어 있다. 또한 기후변화 관련 사회·경제·정책적 개념을 나타내는 단어, 즉 사회 시스템의 회복력 관련 단어인 ‘회복력(resilience)’, ‘기후회복적 발전(climate resilient development)’, ‘복원(restoration)’, ‘생계(livelihood)’와, 완화와 적응 능력의 강화와 실행 방식 관련 단어인 ‘재원(finance)’, ‘예산(budge)’, ‘공동편익/부가편익(co-benefit)’, ‘부작용/부수효과(side effect)’, ‘이행가능성(feasibility)’,

‘간극(gap)’, ‘계획(planning)’, ‘소비(consumption)’, ‘기후행동(climate action)’, ‘포용적(inclusive)’ 등이 포함되어 있다. 이러한 유형의 새로운 단어들의 출현이 AR5보다 AR6에서 더 두드러진다는 사실은 Fig. 7과 Table 10의 AR5와 AR6의 텍스트 산점도 시각화 결과에서 확인된다.

AR5에 나타난 RCP 관련 용어(RCP, RCP2.6, RCP4.5, RCP6.0, RCP8.5)는 AR6에서는 상대적으로 거의 언급되지 않는다.²⁾ 대신, 앞에서 언급되었던 사회 시스템의 회복력 관련 단어들과 완화와 적응 능력의 강화와 실행 방식 관련 단어 대부분은 AR6에서 나타난다. AR6에서 나타난 ‘(재원) 흐름(flow)’, ‘불평등(inequity)’, ‘피해(damage)’, ‘가속화하다(accelerate)’, ‘보존/보전(conservation)’, ‘자본(capital)’, ‘분배적(distributional)’ 등도 이와 관련된 단어들이다.³⁾ 또한 AR5 이후에 도입된 ‘국가결정기여(Nationally Determined Contribution, NDC)’와 ‘지속가능발전 목표(Sustainable Development Goals, SDGs)’가 AR6에서 새롭게 부각된 것은 지속가능한 미래를 위한 국제적 약속을 중요하게 여기는 최근의 경향을 반영한다.⁴⁾

실가스 감축 정책의 도입을 전제하지 않고, 인구, 경제성장, 기술변화 등의 사회경제적 경로에 따라 다양한 배출 전망을 구성한 것이 특징이다. 그러나 이러한 접근은 기후정책의 영향을 반영하지 못한다는 한계가 지적되었다. 이에 따라 제5차 평가보고서(AR5, 2014)에서는 특정한 정책 경로를 전제하기보다, 2100년까지 도달하는 복사강제력(radiative forcing) 수준(2.6, 4.5, 6.0, 8.5 W/m²)을 목표로 설정한 대표농도경로(Representative Concentration Pathways, RCP)를 도입하였다. RCP 시나리오는 구체적인 정책을 가정하지는 않지만, 예컨대 RCP2.6은 강력한 감축 정책이 실행될 때만 실현 가능한 궤적이라는 점에서, 정책 시나리오와 결과 경로를 간접적으로 연결하는 역할을 한다. AR5에서 복사강제력 기준으로 한 RCP 체계가 도입되면서 SRES 체계는 더 이상 사용되지 않게 되었다.

2) 각주 1)에서 기술했듯, RCP는 복사강제력을 기준으로 한 온실가스 배출 경로로, AR5에서 주요 시나리오 체계로 사용되었다. 그에 비해 AR6에서는 공동사회경제경로(Shared Socioeconomic Pathways, SSP)를 중심으로 시나리오를 구성하는데, SSP 시나리오는 인구, 경제, 기술, 불평등, 에너지 전환 등 사회 구조 전반을 반영하며 특정 복사강제력 수준과 결합된 SSPx-y(= SSPx-RCPy) 형태로 제시된다.

3) 동일한 영어 단어에 한글을 둘 이상 적은 이유는 기상청에서 발표한 보고서 국문 번역에 맥락에 따라 같은 영어 단어를 달리 번역했기 때문이다. 또한 흐름 앞에 “(재원)”을 넣은 것은 AR6 SYR의 SPM에서 흐름(flow)은 물리적 흐름이 아니라 주로 재원(finance)과 함께 쓰였기 때문이다.

4) NDC는 2015년 COP21에서 채택된 파리협정 제3조와 제4조에 처음으로 명시된 개념이다. 모든 당사국은 파리협정에 따라 자발적으로 온실가스 감축 목표, 즉 NDC를 수립·제출·갱신·보고해야 한다. 파리협정 이전인 2013년 폴란드 바르샤바에서 열린 COP19에서 ‘의도된 국가결정기여(Intended Nationally Determined Contributions, INDCs)’ 개념이 최초 도입되어 국가 자율적 기여 방식으로 모든 당사국이 온실가스 감축 목표를 제출하도록 하자는 논의가 이루어지기 시작했고, 2014년 페루 리마에서 열린 COP20에서 INDCs 가이드라인

5. 토론: 기후변화에 대한 담론적 함의

각 SPM을 하나의 단일 문서로 간주한 토픽 모델링 분석의 결과는 물리적 기후변화와 자연과학적 영향에 중점을 둔 초기의 관심이 기후위기 완화 및 적응을 위한 기후 정책과 기후행동으로 확장되는 방향으로 이동해왔음을 보여준다. 그리고 부트스트래핑과 다섯 문장 단위 묶음 분석 결과에 따르면, 이러한 주제 비중의 이동은 우연한 경향이 아닌 통계적으로 유의한 변화이다. 이 결과는 IPCC가 평가보고서를 거듭할수록 과학적 인식의 범위를 자연과학 기반의 원인과 영향 분석에서 정책적 실천과 사회 시스템 변화에 대한 대응으로 이동시키고 있음을 계량적으로 뒷받침한다.

전체적으로 보면, ‘기후변화의 물리적 영향’ 주제는 시간이 흐를수록 네 차례 AR의 SPM에서 비중이 점차 감소하는 경향을 보였고, 반대로 ‘기후정책과 기후행동: 완화와 적응’ 주제는 보고서를 거듭할수록 비중이 증가하는 추세를 보였다. 이러한 변화는 IPCC 보고서가 단순한 과학적 경고나 진단을 넘어, 정책적 실천과 대응 전략에 대한 지침서로서의 성격을 점차 강화하고 있음을 보여준다. 특히 AR6에서는 이 두 주제의 비중 차이가 가장 두드러지게 나타났으며, 이는 최신 보고서에서 정책적 개입 가능성과 사회적 실행력에 대한 담론이 가장 적극적으로 전개되어 있음을 시사한다(Yun and Kim, 2024b). 방법론적 관점에서 볼 때에는 IPCC 평가보고서가 기후과학에서 정책 실행으로의 전환이 어떻게 담론에 반영되어 왔는지를 계량적으로 추적할 수 있다는 점에서, 이 연구의 토픽 모델링 분석 결과는 실증적 차원에서 기여가 있다.

이러한 연구 결과는 2절에서 정리한 선행연구와도 맥이 닿는다. Kramar et al. (2014)은 이 연구처럼 AR5에 비해 AR6에서 기후정책과 행동 관련 주제가 부상했음을 보여 주고 있다(예: 기후변화 재정 관련 단어 *finance*, *financial*, *develop*, *public*, *private*, *flow*, 기후 거버넌스 관련 단어 *national*, *governance*, *policy*, *instrument*, *support*, *law*). 그러나 NMF 토픽 모델링 결과 AR5는 10개의 주제, AR6는 12개의 주제가 추출되어 이 연구에서 제시한 주제 비중의 변화를 정량적으로 파악하는 데에는 한계가 있었다. 단어 빈도의 시계열적 변화 추세를 통해 IPCC 평가보

고서 담론 변화를 살펴보았던 Yun and Kim (2004b)도 TAR와 AR4에 비해 AR5와 AR6에서 ‘완화’와 ‘적응’의 출현 빈도가 증가한 것을 후반기 평가보고서 담론 변화와 연결시켰으나, 전반적 주제 분포를 조망하면서 주제 구성의 시계열적 변화를 분석하지는 않았다. 반면에, 선행연구와 달리 이 연구는 정량적·통계적 방법으로 주제 비중의 시계열적 변화를 보여 준다. TAR~AR6 전체 SPM을 대상으로 LDA로 주제 비중을 직접 도출하고(Table 3; Fig. 2), 문장 부트스트래핑과 5문장 묶음 LDA 재학습을 거친 후 ANOVA로 통계적 유의성을 검정한 결과, ‘기후정책과 행동’의 비중 증가와 ‘기후변화의 물리적 영향’의 비중 감소가 통계적으로 유의함을 보였다(Table 4, 5, 6, 7, 8; Fig. 3, 4, 5). 또한 *log-ratio* 산점도를 통해 ‘탄소중립(net-zero)’, ‘기후회복적 발전(climate resilient development)’과 같은 신개념의 부상과 ‘SRES’, ‘부속서-B 국가(Annex B)’, ‘비부속서-I 국가(non-Annex I)’, ‘안정화(stabilization)’와 같은 전통적 개념의 퇴조를 명확히 포착하였다(Table 9, 10; Fig. 6, 7). 요컨대, 선행연구가 단어 수준의 경향성을 제시했다면, 이 연구는 주제 구성 변화의 크기와 유의성을 함께 제공함으로써 IPCC AR의 담론 이동을 계량적·통계적으로 확인하였다.

TAR에서 AR6까지 전반적인 주제별 경향과 흐름에서 특별히 눈에 띄는 지점이 있다. 그것은 Fig. 2, Fig. 3, Fig. 5에 제시된 것처럼 토픽 모델링 분석 결과, AR4의 SYR SPM에서 ‘기후변화의 물리적 영향’ 주제의 비중이 다른 SYR SPM에 비해 상대적으로 높게 나타나고 “기후정책과 행동: 적응과 완화” 주제의 비중이 상대적으로 낮게 나타났다. 이는 AR4가 발표된 2007년이란 시점의 역사적·정치적 맥락과 깊이 연관되어 있는 것으로 추론해볼 수 있다. 2007년에는 노벨평화상이 미국 클린턴 정부의 부통령이었던 엘 고어와 함께 IPCC에게 공동 수여될 정도로 AR4가 국제적으로 큰 주목을 받았다. AR4에서는 기후변화는 ‘더 이상 논쟁이 아니라 과학적으로 확립된 사실’이라는 메시지가 강하게 부각되었다. 2007년의 AR4 발표가 1997년 제3차 당사국총회(COP3)에서 채택되었던 교토의정서가 발효된 시점(2005)과 맞물리면서, 정책 논의가 아직 초기 단계였던 상태에서 AR4가 국제사회에 주고자 했던 메시지는 무엇보다도 ‘기후변화가 실제로 일어

인을 확정하였다. 이후 파리협정이 채택되면서 INDCs는 NDC로 공식 전환되었기 때문에 그 이후에 나온 AR6에만 NDC가 등장한 것이다. 또한 SDG는 2015년 UN 출범 70주년을 맞이해서 발표한 것이기에 그 이후에 발간된 AR6에만 등장한다. UN은 2000년에 개도국들을 주 대상으로 해서 2015년까지 달성해야 할 ‘새천년개발목표(Millennium Development Goals, MDGs)’ 8가지를 발표하였는데 2015년에는 선진국까지 포괄해서 모든 국가가 지향해야 할 지속가능발전의 17개 목표인 SDGs를 발표하였다.

나고 있으며 그 영향이 명확하다'는 과학적 증거 제시였다. AR4를 통해 IPCC는 기후변화에 대한 과학적 합의를 공고히 하는 데 초점을 맞추면서 기후변화가 실제로 진행 중이며 그 영향이 명확하다는 과학적 증거를 제시하는 데 주력하였다. 실제로 AR4는 온도 상승, 해수면 상승, 빙하 후퇴, 극한기후 사건 증가 등 물리과학적 기반에 대한 WG I의 성과를 SYR 전반에서 두드러지게 부각하였다. 이러한 맥락에서 볼 때 AR4는 기후변화 대응에 관한 정책적 재정적 논의가 본격화되기 이전 단계에 해당하며, 따라서 '기후정책과 기후행동'보다는 '기후변화의 물리적 영향'을 강조하는 서술이 상대적으로 더 큰 비중을 차지했다고 해석할 수 있다. 이는 이후 AR6에서 과학적 합의를 전제로 정책·행동적 대응을 강조하는 양상과 뚜렷한 대조를 이룬다.

SPM을 전반기(TAR 및 AR4)와 후반기(AR5 및 AR6)로 구분하여 수행한 *log-ratio* 기반 텍스트 산점도 시각화 분석에 따르면, 전반기 *log-ratio* 기준 상위 5개 단어인 TAR, '생태적(ecological)', 'SRES', '안정화/안정화하다', '연구'는 전반기에서 각각 28회, 16회, 14회, 52회, 34회 사용된 반면, 후반기에서는 거의 사용되지 않아(각각 0회, 0회, 0회, 3회, 2회) *log-ratio* 결과값이 높게 나타났다. 마찬가지로, 후반기 *log-ratio* 기준 상위 단어인 2°C, 1.5°C, '일시적 초과', '탄소중립', '기후회복적 발전'은 모두 후반기에서만 등장하였으며(각각 33회, 31회, 21회, 19회, 16회), 전반기에서는 한 차례도 언급되지 않았다. 요약하자면, 전반기와 후반기를 나누어 두 시기를 비교한 분석에서는 교토의정서 기반 감축 책임 구분과 시나리오 체계, 유엔기후변화협약과 관련된 개념과 인용적 서술 또는 내용의 불확정성과 관련된 개념을 나타내는 단어들이 후반기에는 눈에 띄지 않는다. 대신, 파리협정 이후 나타난 감축 목표와 온실가스 농도 변화 경로 시나리오 관련 개념들과 기후변화에 대한 사회·경제·정책적 개념을 나타내는 단어들이 새로 등장했을 뿐 아니라 두드러지게 사용되었다.

최근 기후변화 과학이 비중을 두고 있는 관심사를 이해하기 위해서는 AR5와 AR6의 *log-ratio* 기반 텍스트 산점도 시각화 비교 결과를 좀 더 자세히 논의할 필요가 있다. 먼저, AR5에 소개되었던 RCP 시나리오가 AR6에서 상대적으로 덜 언급되었다는 점은 주목할 만하다. AR5에서는 RCP 시나리오 체계에 따라 경로 선택을 비교·설명하는 것이 핵심이었지만, AR6에서는 사실상 1.5°C 또는 2°C 목표 외에 대안이 없다는 현실 인식 속에서 '탄소중립', '기후회복적 발전', '형평성(equity)', '이행가능성'과 같은

정책지향적 실천 개념이 좀 더 강조되었다. 보다 본질적으로는 AR5에서 사용된 RCP 시나리오 체계가 AR6에서는 '공동사회경로인 SSP'를 기반으로 한 시나리오 체계로 전환되었기 때문이다. 이때 SSP는 사회경제적 경로와 복사강제력 수준(RCP)을 결합한 SSPx-y(= SSPx-RCPy) 형태로 제시되었으며, 실무집단별 보고서에서는 WG I과 WG II가 SSPx-y 또는 RCPy를, WG III가 C1~C8 범주를 사용하는 등 서로 다른 표기 체계를 병행하였다. 그러나 이러한 혼용을 SPM에서 직접 반복할 경우 독자 혼동을 초래할 수 있기 때문에, AR6의 SYR SPM에서는 '공동사회경로'라는 기술적 용어를 자주 언급하기보다는 각 시나리오가 담고 있는 핵심 정책 개념을 중심으로 설명하고, 시나리오 구분은 주로 1.5°C, 2°C 등 '온난화 수준'이나 '배출 수준(높음/낮음)'으로 통일하여 제시하였다. 따라서 RCP 관련 용어가 AR6 SPM에서 상대적으로 덜 언급된 것은 SPM의 언어 사용 전략에 따른 결과로 해석할 수 있다.

그 외에도 AR5에 비해 AR6에서 상대적 비중이 높은 단어들 대부분은 완화와 적응을 사회·경제·정책적 관점에서 바라보고 개념화한 용어들이다. 특히, AR6의 '기후회복적 발전'은 이전 보고서에서는 한 번도 사용되지 않은 합성어인데 AR6에서만 16회 사용되었다. AR5에서 '기후회복적(climate resilient)'이라는 합성어가 3회 사용되었으나 '발전'이라는 단어와 함께 사용되지는 않았다. 지속가능발전 목표가 기후, 빈곤, 교육, 성평등, 에너지, 생물다양성 등 17개 분야의 경제, 사회, 환경을 포괄하는 개념이라는 점에서 AR6에서 새롭게 개념화된 기후회복적 발전 개념은 기후변화를 염두에 둔 지속가능발전의 구체적 실천 개념이라고 해석할 수 있다.

AR6에서는 두드러지지 않았으나, AR5에서 상대적으로 비중이 높았던 '인간 행위에 의해 초래된'이라는 의미를 지닌 단어인 'anthropogenic'은 TAR, AR4, AR5에서는 각각 16회, 13회, 23회 사용되었으나, AR6에서는 단 1회만 사용되었다. 대신, AR6에서는 이 개념이 '인간 활동(human activity)' 4회, '인간 영향(human influence)' 4회, '인간에 의해 야기된(human-caused)' 6회 등 'human'이 나머지 단어, 즉 'activity', 'influence', 'cause'의 '행위자(agent)'임을 분명히 하는 합성어로 다양하게 표현되었다. 이러한 표현상의 변화는 단순한 어휘 선택의 차이이기보다, 기후변화의 원인이 '인간'에게 있다는 사실을 명료하게 전달하고, 정책결정자와 대중을 포함한 광범위한 수용자에게 명확한 메시지를 제공하기 위한 커뮤니케이션 전

략의 일환으로 이해될 수 있다. 이는 과학적 설명에 중점을 두었던 IPCC 평가보고서의 기존 접근에서 나아가, 변화를 위한 정치적 결단과 정책 및 행위의 필요성을 보다 설득력있게 전달하기 위해 커뮤니케이션 방식의 전환이 진행되고 있음을 시사한다.

전 지구적 온도 상승을 개념화하여 표현한 ‘지구온난화(global warming)’가 1970년대 후반에 학계에서 사용되기 시작해서 1980년대에 들어 대중적·정책적 용어로 자리 잡았음에도 불구하고 AR6에서만 부각된 점도 주목할만한 가치가 있다. 지구온난화는 TAR, AR4, AR5에서 각각 0회, 2회, 2회 사용되었으나 AR6에서는 31회로 사용 빈도가 급격히 증가하였다. TAR, AR4, AR5에서도 ‘온난화(warming)’는 각각 26회, 37회, 41회로 빈번하게 사용되었고 ‘지구적(global)’ 또는 ‘지구적으로(globally)’와 한 문장 내에서 함께 사용된 경우도 많았지만(각각 5회, 6회, 12회), 두 단어가 결합된 지구온난화(global warming)라는 합성어로 사용된 경우는 드물었다. 이러한 변화는 IPCC의 의제 설정이 점차 구체적인 온도 상승 목표(1.5°C, 2°C)의 달성과 그 영향에 대한 분석에 초점을 맞추게 되면서, ‘global warming’이라는 용어가 정량적·정책적 커뮤니케이션의 핵심 수단으로 채택된 결과로 해석할 수 있다(Yun and Kim, 2024b, p. 293). ‘climate change’가 기후시스템 전반의 변화를 포괄하는 광의의 개념이라면, ‘global warming’은 지표면 평균 온도 상승이라는 과학적이면서 측정가능한 대상을 명확히 지칭함으로써, 1.5°C 및 2°C라는 구체적 수치로 표현된 온도 목표와 직접적으로 연결되는 표현이기 때문이다. 이러한 특성은 특히 파리협정이 “지구 평균 온도(global average temperature)”의 제한을 핵심 목표로 명시하고 있다는 점에서 더욱 중요하다. AR6에서는 파리협정의 온도 목표와 정합성을 확보하고 정책결정자와의 소통 효과를 높이기 위해, 이전 보고서들보다 훨씬 더 빈번하게 ‘global warming’이라는 표현을 사용한 것으로 추론해 볼 수 있다. 실제로 AR6 SPM에서는 “지구온난화 1.5°C(global warming of 1.5°C)” 또는 “서로 다른 지구온난화 수준(different global warming levels, GWLs)” 같은 표현이 도표 제목과 본문에서 반복적으로 등장하는데, 이는 온도 상승 수준에 따른 영향 분석의 중심성을 반영하는 동시에, 행동 촉구 메시지를 보다 직접적으로 전달하려는 전략적 언어 선택으로 볼 수 있다.

6. 결론

이 연구의 목적은 IPCC 제3차부터 제6차 평가보고서 종합보고서의 SPM을 대상으로 주제와 주제 비중의 변화를 분석하여 국제 기후담론의 흐름을 이해함으로써, 기후과학자들이 기후위기를 어떻게 이해해 왔으며 앞으로 국제사회가 어떠한 대응을 해나가도록 요청하는지를 밝히는 데 있다. 이를 위해 LDA 기반 토픽 모델링과 *log-ratio* 텍스트 산점도 시각화를 결합한 계량 텍스트 분석을 실시하였다. 주제 구성을 시계열로 비교하고, 문장 부트스트래핑 및 5문장 묶음 재추정과 일원분산분석으로 주제 비중 변화의 통계적 유의성을 확인하였다. 그 결과, 초기 보고서에서는 물리적 영향과 자연과학 중심의 주제가 두드러진 반면, 평가 주기를 거치며 완화·적응·정책/행동 관련 주제가 유의하게 확대되었고, 특히 AR6에서 그 전환이 뚜렷해짐을 확인하였다. 또한 산점도 시각화는 토픽 단위에서 포착되기 어려운 표식 단어의 부상과 퇴조를 드러내어, 특정 시기에 강조된 개념 변화의 양상을 보완적으로 보여 주었다.

SPM이 핵심 메시지를 압축해 제시하고 전 세계 정책 결정자, 언론, 연구자가 널리 참고하는 문서이기 때문에, 이 연구에서 확인한 SPM 담론의 강조점 이동은 실제 정책 결정의 변화를 뒷받침할 근거가 될 수 있다. 이 연구의 결과가 보여 주듯이, AR6 SYR의 SPM은 과학적 사실의 나열보다는 무엇을, 어떻게, 지금부터 즉각적으로 실행할 것인가에 더 큰 비중을 둔다. 따라서 목표치, 감축, 완화, 적응의 실행 틀과 회복력과 형평성에 대한 논의의 비중 증가는 정부 부처 간 자원 배분과 단계별 로드맵 설계에 직접적인 시사점을 제공한다. 또한 시기별로 두드러진 표식 단어의 차이는 대중 소통 메시지의 핵심어 목록을 주기적으로 최신화하는 근거가 되어, 과학적 내용이 정책·언론·교육으로 옮겨가는 과정에서 오해를 줄이고 행동을 유도하는 데 유용하다. 마지막으로, 이 연구에서 발견한 ‘주제의 무게중심’이 이동한다는 사실은 향후 정책 안내문, 브리핑, 보도자료의 일관성과 비교 가능성을 높이는 기준이 될 수 있다. 기후변화와 관련되어 말하여야 할 내용의 우선순위를 SPM과 동일한 주제 비중의 틀에 맞추면, 서로 다른 기관과 시점의 문서도 같은 기준으로 정렬되어 바로 비교할 수 있기 때문이다.

이 연구는 다음과 같은 한계가 있다. 첫째, 분석 범위를 SPM으로만 한정하였다는 점이다. SPM이 정책 결정용 핵심 사항의 요약이라는 장점이 있지만, 실무집단 보고서

전체 본문이나 부속 문서에 나타나는 세부 논의까지 포괄하지는 못한다. 둘째, 이 연구에서는 합성어·숫자·단위·약어를 유지하도록 전처리를 설계하였으나, 여전히 말뭉치 크기, 토큰화 규칙, 불용어 처리, 토픽 수 결정에 따른 민감도가 존재한다. 셋째, 통계적 유의성 검정은 주제 비중 차이의 견고성을 확인하지만 비중 차이의 이유에 대한 직접적 설명을 제공하지는 않는다.

향후에는 자료 범위를 SYR 전체와 WG I/II/III 본문, 특별보고서까지 확대해 섹션별, 문서군 간 상호작용을 비교할 필요가 있다. 또한, 방법론적으로는 다른 유형의 토픽 모델링 기법을 적용하여 이 연구가 시도하지 않은 다양한 맥락에서의 단어 동시출현 관계와 빈도에 기반한 중심어 추출을 시도할 필요가 있다. 그리고 정책 문서, 국가별 대응 전략, 언론 보도와 같은 다른 기후변화 담론 텍스트와의 교차 비교 분석을 통해 SPM 담론 변화가 국가 정책과 대중 담론에 미친 영향과 전파 경로를 살피는 것도 유의미하다. 더불어 각 국가의 기후담론이 IPCC의 기후담론의 변화와 연결되어 있는지, IPCC 담론 변화를 반영하고 있는지에 대한 분석 또한 필요하다.

이 연구에서는 TAR ~ AR6 SPM을 일관된 기준으로 비교하여, IPCC 담론의 주제 구성 변화와 강조점 이동이 통계적으로 유의한 추세를 보여주었다. 이는 IPCC가 과학적 진단에서 정책과 행동 지향으로 커뮤니케이션의 초점을 옮겨 왔음을 의미한다. 동시에, 앞서 제시한 한계와 후속 과제를 보완하면, SPM의 수치와 용어, 우선순위가 정부 문서·언론 보도·시민사회 자료로 어떻게 이어졌는지 확인할 수 있을 것이다. 이렇게 축적된 근거는 우선순위가 분명한 정책 설계, 효과적인 대중 소통, 오인파 왜곡에 대한 대응에 기초 자료로 쓰일 수 있을 것이며, 기후위기 대응의 전략적 커뮤니케이션을 한층 정교화하는 데 기여할 것이다.

사사

이 논문은 서울대학교 환경계획연구소와 2023년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구자 지원사업의 지원을 받아 수행된 연구임(NRF-2023S1A5A2A01079968). 익명의 심사자분들의 적절한 지적과 의견 제시 덕분에 논문의 개선이 이루어졌음을 밝힌다. 연구자료 수집에 도움을 준 고도연 박사과 서울대학교 환경대학원 박사수료생 김현지, 최순길 학생, 박사과정생 조유진 학생에게도 고마운 마음을 표한다.

Reference

- Barkemeyer R, Dessai S, Monge-Sanz B, Renzi BG, Napolitano, G. 2016. Linguistic analysis of IPCC summaries for policymakers and associated coverage. *Nat Clim Change* 6(3): 311-316. doi: 10.1038/nclimate2824
- Boykoff MT, Boykoff JM. 2007. Climate change and journalistic norms: A case-study of US mass-media coverage. *Geoforum* 38(6): 1190-1204. doi: 10.1016/j.geoforum.2007.01.008
- Brown G, Yule G. 1983. *Discourse analysis*. Cambridge: Cambridge University Press.
- Dormer R. 2020. Linguistic analysis and climate change discourse: Exploring current linguistic perspectives. *Proceedings of the International Conference on Sociolinguistics and Language Sciences*. p. 1130-1138.
- Fairclough N. 2013. *Critical discourse analysis: The critical study of language*. Routledge.
- Furnham A. 2024. Sustainability skepticism: Attitudes to, and beliefs about, climate change. *Sustainability* 16(18): 8164. doi: 10.3390/su16188164
- Gallup Korea. 2024. Survey on public perceptions of climate change (“기후변화 인식 조사”); [accessed 2025 Jun 6]. <https://www.gallup.co.kr/gallupdb/reportContent.asp?qNo=1379>
- Ge J. 2010. A situated-sociocultural approach to literacy and technology. *The new literacies: Multiple perspectives on research and practice* 165: 193.
- IPCC (Intergovernmental Panel on Climate Change). 2001. *Climate change 2001: Synthesis report. Contribution of working groups I, II and III to the third assessment report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: Author.
- IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate change 2007: Synthesis report. Contribution of working groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: Author.
- IPCC (Intergovernmental Panel on Climate Change). 2014.

- Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: Author.
- IPCC (Intergovernmental Panel on Climate Change). 2023. Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: Author.
- Isopp B. 2024. The politics of politicization: Climate change debates in Canadian print media. *Public Understanding Sci* 33(5): 1-19. doi: 10.1177/09636625231220226
- Janzwood S. 2020. Confident, likely, or both? The implementation of the uncertainty language framework in IPCC special reports. *Clim Change* 162(3): 1655-1675. doi: 10.1007/s10584-020-02746-x
- Kanerva J, Krizsán A. 2021. Discouraging climate action through implicit argumentation: An analysis of linguistic polyphony in the summary for policymakers by the Intergovernmental Panel on Climate Change. *Discourse Commun* 15(6): 609-628. doi: 10.1177/175048132111026512
- Kramar N, Ilchenko O, Levko O. 2024. A comparative linguistic analysis of the 2014 and 2023 IPCC Synthesis Reports using topic modeling and Scattertext. *Proceedings of 2024 IEEE 19th International Conference on Computer Science and Information Technologies (CSIT)*.
- Kim S-Y, Yun S-J. 2016. An analysis of catholicism's environmental discourses concerning the four major rivers project: Based on a critical discourse analysis. *J Soc Sci* 26(1): 319-343. doi: 10.16881/jss.2015.01.26.1.319
- Mach KJ, Mastrandrea MD, Freeman PT, Field CB. 2017. Unleashing expert judgment in assessment. *Glob Environ Change* 44: 1-14. doi: 10.1016/j.gloenvcha.2017.02.005
- Mastrandrea MD, Field CB, Stocker TF, Edenhofer O, Ebi KL, Frame DJ, Held H, Kriegler E, Mach KJ, Matschoss PR. 2010. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties.
- Penz H. 2022. Communicating climate change: How (not) to touch a cord with people and promote action. *Text Talk* 42(4): 571-590. doi: 10.1515/text-2020-0081
- Poole R, Hayes N. 2023. Stance in climate science: A diachronic analysis of epistemic stance features in IPCC physical science reports. *J Corpora and Discourse Studies* 5(1): 37-60. doi: 10.18573/jcads.100
- Van Dijk TA. 2009. *Society and discourse: How social contexts influence text and talk*. Cambridge: Cambridge University Press.
- Van Dijk TA. 2021. The discourse, cognition, society framework. IX Colóquio ALED Brasil, Espírito Santo.
- WIN. 2024. Varied beliefs and actions on climate change; [accessed 2025 Jun 6]. <https://winmr.com/varied-beliefs-and-actions-on-climate-change/>
- Yun S-J, Kim S-H. 2024a. A comparative analysis of the expressions of scientific uncertainty and authorial stance in the original and Korean translation of IPCC Synthesis Reports' Summary for Policymakers. *J Clim Change Res* 15(5-1): 799-814. doi: 10.15531/KSCCR.2024.15.5.799
- Yun S-J, Kim S-H. 2024b. A diachronic analysis of word frequency in IPCC reports: Exploring shifts in climate change discourse. *Stud in Ling* 73: 271-302. doi: 10.17002/sil..73.202410.271
- Yun S-J, Ku D, Park N-B, Han J. 2014. Framing climate change as an economic opportunity in South Korean newspapers. *Dev Soc* 43(2): 219-238. doi: 10.21588/dns.2014.43.2.005

Software and Tools

- R Core Team. 2024. R: A language and environment for statistical computing. Version 4.4.2. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- RStudio Team. 2024. RStudio: Integrated development environment for R. Version 2024.09.0. Posit Software, PBC, Boston, MA. <https://posit.co/>