

구조적 토픽모형(Structural Topic Model)을 활용한 주택연금 정책담론 분석: 2007~2026년 뉴스 보도의 토픽 변동과 매체별 프레이밍 차이

이승형*

요약

본 연구는 주택연금 정책담론을 구조적 토픽모형(structural topic model, STM)으로 분석한다. 빅카인즈의 2007년 7월~2026년 3월 주택연금 뉴스 12,799건에서 SimHash 중복 제거를 거쳐 12개 토픽을 도출하였다. 담론은 정책·가계부채(13.4%), 자산관리·은퇴설계(11.9%), 가입실적·통계(10.1%) 등으로 구성되었다. 이명박은 가입실적(18.0%), 문재인은 제도 설계(13.2%), 이재명은 고령가구·상속의향(10.5%)이 두드러졌고, 매체별로는 지역지가 가입실적, 종합일간지가 노후·세대 서사, 경제지가 은퇴설계에 집중하였다. 보증료 인하 직후 고령가구·상속의향이 +25.65%p 급증했으나 사후 표본 31건으로 잠정적이다. KNU와 KoELECTRA의 감성 분류 일치도는 49.0%로 정책 뉴스 도메인 갭이 확인되었다. 정부별 추정치는 인과가 아닌 기간별 기술 통계이다. 정책 담론에 STM을 적용한 분석 프레임워크를 제시한다.

핵심어 : 주택연금, 구조적 토픽모형(Structural Topic Model, STM), 정책담론, 미디어 프레이밍, 텍스트마이닝

I. 서론

한국주택금융공사가 2007년 7월 출시한 주택연금은 19년째 운영되고 있다. 2025년 12월 기준

* 이승형, 주저자, 동아대학교 교육학과 박사, homoludus4780@gmail.com

© Copyright 2026 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

누적 가입자는 15만 명을 돌파하였고 월평균 수령액은 127만 원에 달하나, 적격 대상 대비 가입률은 2%~2.5%로 정부 목표(3%)에 미달한다. 출시 이래 가입연령·가격상한·지급방식의 10차례 이상 개편에도 이 정체가 지속되는 가운데, 2026년에는 출시 이래 최대 규모의 개편이 진행 중이다. 제도 변천의 궤적과 2026년 개편의 세부 내용은 II.1절에서 상술한다. 이 시점은 그간 누적된 정책담론의 구조와 변동을 체계적으로 분석할 학술적 필요성을 제기한다.

주택연금의 정책적 중요성은 한국 사회의 인구구조 변화에서 비롯된다. 한국은 2024년 12월 초고령사회에 진입하였으며(65세 이상 1,024만 명, 전체의 20.0%; 행정안전부, 2024), 노인빈곤율은 40.4%로 OECD 평균(14.2%)의 약 3배에 달한다(OECD, 2023). 고령가구의 자산 구성은 부동산에 편중되어 소득은 부족하나 주거 자산은 보유한 자산형 빈곤(house rich, cash poor) 상태가 보편적이다. 주택연금은 이 구조적 불일치를 해소하는 핵심 정책 수단으로 설계되었으나, 출시 이후 가입률 정체가 지속되고 있다.

이 간극의 원인을 규명하려는 연구는 주로 가구 단위 설문조사나 행정 데이터를 활용한 로짓모형·DID 분석으로 가입 결정요인을 탐색하는 접근을 취해 왔다. 가입연령과 주택가격, 상속 의지 등이 주요 변수로 반복 확인되었으며(전병욱, 2022; 주택금융연구원, 2020), 행동경제학적 관점에서 최적의사결정 모델보다 임의적 판단이 지배한다는 점도 보고된 바 있다(전병욱, 2022; 주택금융연구원, 2020). 그러나 이러한 개인 수준의 분석은 주택연금에 대한 사회적 인식이 어떻게 형성·변화해 왔는지를 포착하지 못한다. 가입 결정은 진공 상태에서 이루어지지 않으며, 언론 보도를 통해 매개된 정보 환경이 잠재 가입자의 인식과 태도를 조건짓는다. 예컨대, 주택연금이 노후 빈곤 대응 수단으로 보도되는지 혹은 집값 하락 리스크를 수반하는 금융 계약으로 보도되는지에 따라 잠재 가입자가 수용하는 제도 이미지가 달라질 수 있다. 미디어 프레이밍 이론에 따르면, 동일한 정책이라도 어떤 측면이 부각되느냐에 따라 수용자의 태도와 행동이 달라진다(Entman, 1993; Scheufele, 2000). 그럼에도 주택연금 보도의 담론 구조를 19년 장기 시계열에 걸쳐 체계적으로 분석한 학술 연구는 국내 학술 데이터베이스(KCI·DBpia) 검색 결과 확인되지 않는다.

이 공백을 메우기 위해 본 연구는 구조적 토픽모형(structural topic model, STM)을 활용한다. STM은 대규모 텍스트 코퍼스에서 잠재 토픽을 추출하는 동시에, 문서 수준의 메타데이터(시간, 매체 특성)가 토픽 비중(topic prevalence)과 단어 분포에 미치는 영향을 추정할 수 있는 확률적 텍스트 분석 모형이다(Roberts et al., 2014). 국내외 정책 영역에서 STM의 활용은 점차 확장되고 있으며, 환경 영역에서는 미세먼지 재해 보도의 프레이밍 추출(이준웅·김성희, 2018), 보건 영역에서는 코로나 19 초기 온라인 정보 교환과 불안 확산 분석(Jo et al., 2020) 등이 대표적이다. 그러나 주택금융 분야에 STM이 적용된 사례는 학술 데이터베이스(KCI·DBpia·Web of Science·Scopus) 검색 결과 확인되지 않았다. 이 연구는 STM의 메타데이터 효과 추정 기능을 활용하여 다음 두 가지 연구 질문에 답한다.

RQ1. 주택연금 뉴스 담론은 어떤 토픽으로 구성되며, 시기별로 어떻게 변화하였는가?

RQ2. 정부 및 매체유형에 따라 각 토픽의 비중은 어떻게 달라지며, 이러한 차이는 정책담론의 어떤 분화 양상을 드러내는가?

이 학술지에서 텍스트마이닝 기반 담론 분석이 시도된 적이 없고, 주택연금을 미디어 프레이밍 관점에서 분석한 국내 학술연구도 확인되지 않는다는 점에서, 2007년 7월 제도 출시부터 2026년 3월 보증료 인하 직후까지 약 19년·6개 정부·10회 이상의 주요 제도 개편을 포괄하는 이 연구는 방법론·주제·시간 범위의 세 차원에서 기존 연구 공백을 보완한다. 디딤돌 대출 정책 효과 분석(김강석·박정은, 2025) 등 최근 호에서 방법론적 다양화가 시도되고 있으나, NLP 기반 정책담론 분석은 이 학술지에서 공백으로 남아 있다. 분석 대상 코퍼스는 빅카인즈에서 수집한 12,799건의 뉴스 보도이다. 다만 2026년 표본은 1월 1일부터 3월 19일까지의 부분년도(195건)이며, 보증료 인하(2026년 3월 1일) 직후 사건 분석은 사후 19일·31건이라는 표본 제약 아래 잠정적 결과로 해석한다.

이 분석의 결과는 정책담론의 구조적 변동 패턴을 가입률 정체의 거시적 맥락 위에 위치시키며, 주택금융 정책담론에 구조적 토픽모형을 적용한 분석 프레임워크를 제시함으로써 기존 계량경제학적 접근과 상호 보완적인 방법론적 확장에 기여한다. 이하에서 이론적 배경, 분석 설계, 결과, 함의를 순차적으로 논의한다.

II. 이론적 배경 및 선행연구

1. 주택연금 제도의 변천과 정책적 맥락

주택연금(역모기지, reverse mortgage)은 고령 주택 소유자가 자신의 주택을 담보로 제공하고 사망 시까지 매월 연금을 수령하는 금융상품이다. 한국에서는 한국주택금융공사가 2007년 7월 만 65세 이상, 시가 6억 원 이하 주택 소유자를 대상으로 출시하였다. 이후 제도는 점진적으로 확대되어 가입연령이 60세(2008년), 55세(2020년)로 하향되었고, 주택가격 상한도 9억 원(2009년), 12억 원(2023년)으로 상향되었다. 2014년에는 1.5억 원 이하 주택을 대상으로 우대형 주택연금이 도입되어 질적 다양화가 이루어졌다.

2026년에 이르러 제도는 출시 이래 가장 큰 규모의 개편을 맞이한다(금융위원회, 2026). 3월 1일부터 초기보증료가 1.5%에서 1.0%로 인하되었고, 월 수령액은 평균 3.13% 증가하였다. 연보증료는 재정 지속가능성을 위해 0.75%에서 0.95%로 소폭 인상되었다. 6월에는 자녀 승계 제도 신설,

실거주 요건 완화, 귀촌·인구감소지역 이주 예외 등이 시행될 예정이다. 자녀 승계 제도는 가입 장벽 조사에서 54.4%로 1순위를 차지하는 상속 의지 문제를 직접 겨냥한 정책 대응이라는 점에서 제도사적 전환점이 될 수 있다.

그러나 이러한 제도적 확대에도 불구하고 가입률은 적격 대상의 2%~2.5%에 머물고 있다. 선행연구는 이 간극의 원인을 개인 수준의 의사결정 변수에서 탐색해 왔다. 전병욱(2022)의 로짓모형 기반 가입 결정요인 분석에서는 가입연령, 주택가격, 자녀 유무, 금융이해력이 유의미한 설명변수로 보고 되었으며, 행동경제학적 접근에서는 자산 편중과 상속 의지가 합리적 최적의사결정을 압도한다는 점이 지적되었다. 이호일 외(2025)는 고령가구의 자산구성이 삶의 만족도에 미치는 영향을 분석하여 부동산 자산 유동화의 필요성을 간접적으로 시사하였다. 이러한 연구는 미시적 의사결정 메커니즘을 밝히는 데 기여하였으나, 주택연금에 대한 사회적 담론의 형성과 변화를 포착하지는 못한다. 가입 결정이 언론 보도를 통해 매개된 정보 환경 속에서 이루어진다는 점을 감안하면, 정책담론 분석 프레임워크의 도입이 필요하다.

2. 미디어 프레이밍 이론과 토픽-프레임 매핑의 이론적 정당화

Entman(1993)에 따르면, 프레이밍은 (1) 문제 정의, (2) 인과적 해석, (3) 도덕적 평가, (4) 처방의 권고라는 네 가지 기능을 통해 특정 이슈의 일부 속성을 텍스트에서 두드러지게 만드는 과정이다. 이 정의는 프레이밍이 단순한 주제 선택을 넘어 의미 구성의 다층 구조를 포함함을 시사한다. Scheufele(2000)는 미디어 프레이밍이 의제 설정에서 태도 형성, 나아가 행동 의향으로 이어지는 다단계 경로를 이론화하였다. 금융상품 선택의 맥락에서 이 경로는 언론이 주택연금의 특정 측면을 반복적으로 부각할 때 잠재 가입자의 제도 인식과 가입 의향이 조건지어질 수 있음을 시사한다. Shiller(2015, 2019)는 서사 경제학(narrative economics) 관점에서 부동산 시장 참여자의 의사결정이 미디어를 통해 전파되는 서사에 의해 조건지어진다고 논증하였으며, 이 논의는 주택연금 가입 결정에도 적용 가능하다.

이 연구는 STM에서 추출되는 토픽을 Entman(1993)이 제시한 프레이밍의 네 기능 가운데 (1) 문제 정의에 대응하는 측면(aspect)으로 한정하여 해석한다. 즉, 토픽 비중은 “특정 정책 측면이 어느 정도의 빈도와 비중으로 보도에서 부각되었는가”라는 측면 현저성(aspect salience)을 측정하며, 이는 프레이밍의 출발점이자 의제 설정의 직접적 작용기제이다. 인과 해석, 도덕 평가, 처방 권고에 해당하는 깊은 프레이밍 기능은 토픽 비중 추정만으로는 측정할 수 없다. 따라서 본 연구의 분석

결과는 “이슈 측면별 보도 주목도(aspect attention)”의 시계열적 변동에 한정되며, 프레임의 완전한 측정으로 확장 해석할 수 없다. 이 한정은 토픽-프레임 매핑의 이론적 정당화에 대한 직접적 응답이며, V장 한계 논의에서 다시 다룬다.

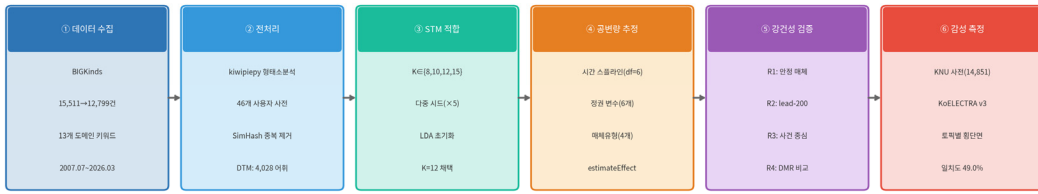
정책담론의 미디어 프레임상을 분석한 국내 선행연구에서, 토픽모델링은 측면 현저성의 변동을 포착하는 유효한 도구로 자리매김해 왔다. STM을 적용한 정책 연구는 환경(이준웅·김성희, 2018), 보건(윤은지·강보영, 2023), 정책 일반(박치성·신나리, 2021)에 걸쳐 축적되어 왔으며, STM과 BERT 감성분석을 결합하여 매체 유형별 프레임 차이를 검토하는 설계도 시도되었다(Hur & Yang, 2024). 주택·부동산 분야에서는 LDA(latent Dirichlet allocation) 기반의 부동산 전세 연구 메타 분석(하석현·김승희, 2024)과 YouTube 주택 담론 분석(Oh & Kim, 2025)이 수행되었으나, 주택 금융 분야에서 STM의 메타데이터 효과 추정 기능을 활용한 연구는 확인되지 않는다.

STM을 활용한 정책담론 분석은 다양한 영역에서 축적되어 왔다. 이준웅·김성희(2018)는 2012~2017년 미세먼지 보도 3,636건에 STM을 적용하여 30개 토픽을 추출하고, 언론사와 시간을 공변량으로 투입하여 매체 간 프레임 차이를 확인하였다. 이 연구는 국내에서 뉴스 코퍼스에 STM 공변량 기능을 적용한 대표적 사례로서, 이 분석의 설계에 직접적 선례를 제공한다. 박치성·신나리(2021)는 역대 대통령 연설문에 STM을 적용하여 정권별 정책 우선순위의 변동을 확인하였으며, 윤은지·강보영(2023)은 코로나19 백신 보도에 STM과 의미연결망을 결합하여 언론 유형에 따른 프레임 차이를 분석하였다. Hur & Yang(2024)은 한국 예멘 난민 보도에 STM과 BERT 감성분석을 결합하여 매체 유형별 프레임 분화를 검토하였다. 이들 선행연구는 공통적으로 STM의 공변량 추정 기능을 활용하여 시간·매체라는 문서 속성이 토픽 구조에 미치는 체계적 효과를 확인하였으며, 이 연구는 이 분석 전략을 주택금융 정책담론에 확장 적용한다.

기존의 계량경제학적 접근이 개인 수준의 가입 결정요인을 규명하는 데 기여하였다면, 텍스트마이닝 기반 담론 분석은 사회적 인식과 매체 환경이라는 거시적 맥락을 포착함으로써 상호 보완적 분석 프레임워크를 제공한다. 이 학술지에서 비정형 텍스트 데이터에 NLP를 적용한 연구는 확인되지 않으며, 주택연금을 미디어 프레임 관점에서 분석한 국내 학술연구도 부재하다는 점에서, 방법론·주제·시간 범위의 세 차원에서 이 공백을 보완한다.

III. 연구방법

III장의 분석 절차를 개관하기 위해 데이터 수집에서 감성 측정까지의 6단계 프로세스를 <그림 1>에 도식화하였다.



$$prevalence = -s(\gamma_numeric, df=6) + admin_merged + press_type$$

주: 화살표(→)는 각 단계의 순차적 흐름을 나타낸다. Prevalence 공변량 설정식은 하단에 표기.

〈그림 1〉 연구 분석 프로세스 프레임워크

1. 구조적 토픽모형(Structural Topic Model)의 방법론적 기초

토픽모형은 관측된 단어 공기(co-occurrence) 패턴으로부터 문서 집합에 내재한 주제 구조를 역추론하는 생성 모형이다. Blei et al.(2003)이 제안한 잠재 디리클레 배분(LDA)이 기초 모형이며, 이후 다양한 확장이 이루어졌다. STM은 Roberts et al.(2014, 2016)에 의해 개발된 모형으로, LDA가 전제하는 문서-토픽 독립성(단일 디리클레 사전)과 토픽-단어 균질성(동일 β 분포)이라는 두 가정을 공변량 체계로 해제한다. 전자의 경우, LDA에서는 각 문서의 토픽 혼합(θ)이 모든 문서에 공통인 단일 디리클레 분포로부터 생성되어 발행 시점이나 매체 유형 같은 문서 속성의 체계적 효과를 포착할 수 없는 반면, STM은 비중 공변량(prevalence covariates)을 도입하여 이 효과를 추정한다. 후자의 경우, STM은 내용 공변량(content covariates)을 통해 동일 토픽이라도 매체 유형에 따라 상이한 단어로 표현될 수 있음을 포착한다. 이 연구는 비중 공변량을 중심으로 분석하며, 내용 공변량은 12,799문서 규모에서 추정에 필요한 연산 비용이 가용 자원을 초과하여 후속 연구로 미룬다.

다만 STM의 공변량 추정에는 식별 한계가 있다. 이 모형에서 시간 변수(자연 스플라인)와 정권 변수가 동시에 모형에 투입될 경우, 정권 변수의 회귀계수는 시간 평활 항을 통제한 잔여 기간 효과를 반영한다. 이는 정부의 인과적 정책 효과로 해석되어서는 안 되며, 시간 평활 잔차에 대한 기간별 기술 통계로 해석해야 한다. 이러한 식별 한계는 IV장 결과 해석과 V장 한계 논의에서 일관되게 적용된다.

STM은 R 패키지 stm(Roberts et al., 2019)으로 구현되어 있으며, 이 연구의 본 분석은 R stm v1.3.8(연구 시점 CRAN 안정 버전)에서 수행하였다. Python 환경에서의 보조 검증을 위해 tomotopy v0.14의 DMRModel(dirichlet multinomial regression topic model)을 병행 적용하였다(Mimno & McCallum, 2008). DMR은 STM의 비중 공변량 메커니즘의 직접적 선행 모형으로, 메타데이터를 디리클레 사전 매개변수에 회귀시키는 점에서 유사한 회귀 구조를 공유한다. 다만

STM이 logistic-normal 분포에 기반하는 반면 DMR은 디리클레-다항 분포에 기반한다는 차이가 있어, 두 모형이 수학적으로 완전히 동일한 것은 아니다. 두 구현의 결과는 정권·매체·사건별 추정값에서 95% 신뢰구간 범위 내에서 일관됨이 IV.3절 R4 분석에서 확인된다.

2. 데이터 수집 및 코퍼스 설계

이 연구의 본 분석 대상은 한국언론진흥재단의 빅카인즈(BIGKinds) 뉴스 데이터베이스에서 수집한 12,799건의 주택연금 관련 기사이다. 빅카인즈는 한국언론진흥재단이 운영하는 국내 최대 뉴스 빅데이터 분석 서비스로, 종합일간지·경제지·지역일간지·방송사 등 8,200만여 건의 뉴스 데이터베이스에 빅데이터 분석 기술을 접목한 서비스이다(한국언론진흥재단, 2024). 이 연구는 빅카인즈 단일 데이터베이스를 본 분석 코퍼스로 채택하여 텍스트 단위와 매체 풀의 일관성을 우선 확보하였다. 빅카인즈는 본문을 약 200자 이내의 리드 단락으로 제한하는 텍스트 단위 특성을 갖는데, 이 제약은 오히려 모든 문서가 동일한 텍스트 단위로 비교 가능하다는 분석적 이점으로 작용한다. Google News 보충 표본은 본 분석에서 제외하되, 결과의 외적 강건성을 검증하기 위한 별도의 보조 분석에서만 lead-200 윈도우로 균질화하여 활용한다(III.5절 및 IV.3절 참조).

검색에는 “주택연금”, “역모기지”, “주택금융공사”, “한국주택금융공사”, “내집연금”, “주택담보노후연금”, “reverse mortgage”, “월지급금”, “중신지급방식”, “확정혼합방식”, “HF공사”, “HF공사”, “주금공”의 13개 도메인 키워드를 사용하여 본문·제목·키워드·기관명 중 하나 이상에 해당 키워드를 포함하는 2007년 7월부터 2026년 3월까지의 기사를 추출하였다. 이 키워드 풀은 인접 영역 보도(예: 농지연금, 전세 보증, 일반 부동산 정책)를 사전 배제하기 위해 설계되었으며, 도메인 키워드 필터링 단계에서 1,738건의 인접 영역 보도가 제외되었다(표본 감축 단계 3).

빅카인즈에는 두 가지 구조적 제약이 있다. 2024년 4월 수집 대상을 54개사에서 104개사로 확대한 변동이 2024년 이후 매체 구성에 단절을 가져오므로, IV장의 매체유형별 분석은 2024년 이전·이후 두 구간을 분리하여 보고한다. 또한 조선일보·동아일보는 2018년 이후 기사만 제공되므로, 2007~2017년 구간에서 보수 매체가 과소 대표되어 있다.

키워드 필터링으로 영역 외 보도를 제거한 뒤에도 동일 보도자료를 변형 제목으로 재배포하는 기사가 잔존할 수 있다. 제목 완전 일치 및 difflib.SequenceMatcher 기반 제목 유사도 0.85 이상의 동일 날짜 기사를 1차 제거한 뒤, 본문에 대해 SimHash(Charikar, 2002) 64-bit 해시 기반 근사 중복 검출을 추가 적용하였다. 임계 해밍 거리는 후보군 검토 결과 3으로 설정하였으며, 이는 동일 보도자료 변형을 엄격하게 검출하면서 단순 인용 또는 부분 발췌 기사는 보존하는 균형점이다. 임계

거리 ± 2 변동($k=1, k=5$) 민감도 분석에서 검출 건수는 본 임계 대비 $\pm 2.0\%$ 이내로 안정적이었다. 한국주택금융공사 공식 발표 본문과 60% 이상의 자카드 유사도를 보이는 기사는 별도 클러스터로 표시하여, T6 토픽 구성에서 보도자료 전재 비중을 추정 가능하게 하였다.

중복 제거를 거친 코퍼스에 대해 kiwipiepy로 형태소 분석을 수행하였다. kiwipiepy는 통계적 언어 모델과 Skip-Bigram을 결합한 Kiwi 형태소 분석기의 Python 바인딩으로, 문맥 기반 중의성 해소 정확도가 평균 86.7%에 이르며 대규모 코퍼스 처리에 적합하다(이민철, 2024). 46개의 사용자 사전 항목(“주택연금”, “역모기지”, “한국주택금융공사”, “보금자리론”, “스트레스DSR” 등 주택금융 복합명사)을 등록하여 분절 오류를 방지하였으며, 명사(NNG, NNP)와 일부 어근(XR)을 추출 대상으로 삼았다. 뉴스 보일러플레이트(“기자”, “무단전재”, “재배포금지”)와 고빈도 의미약 단어(“밝혔다”, “전했다”, “관련” 등) 73개를 불용어로 제거하는 한편, “가입”, “금리”, “상속”, “담보” 등 30개의 핵심 개념어는 품사와 무관하게 보존 목록에 포함하여 분석에서 누락되지 않도록 하였다.

토큰 수가 5개 미만인 문서와 정권 구분이 불가능한 기사를 제외한 후, 본문 기반 중복 제거 단계에서 추가로 식별된 근사 중복을 제거하여 본 분석 대상은 12,799건으로 확정되었다. 문서-단어 행렬(DTM) 구성 시 5개 문서 미만에 출현하는 저빈도 어휘와 전체 문서의 70% 이상에 출현하는 고빈도 어휘를 제거하였다. “주택”, “연금”, “주택연금”은 코퍼스의 77%~87%에 등장하여 토픽 판별력이 없으므로 이 단계에서 자동 제거되었다. R stm 패키지의 prepDocuments 함수를 적용하여 최종 어휘 규모 4,028개, 분석 대상 문서 수 12,799건의 코퍼스를 구성하였다.

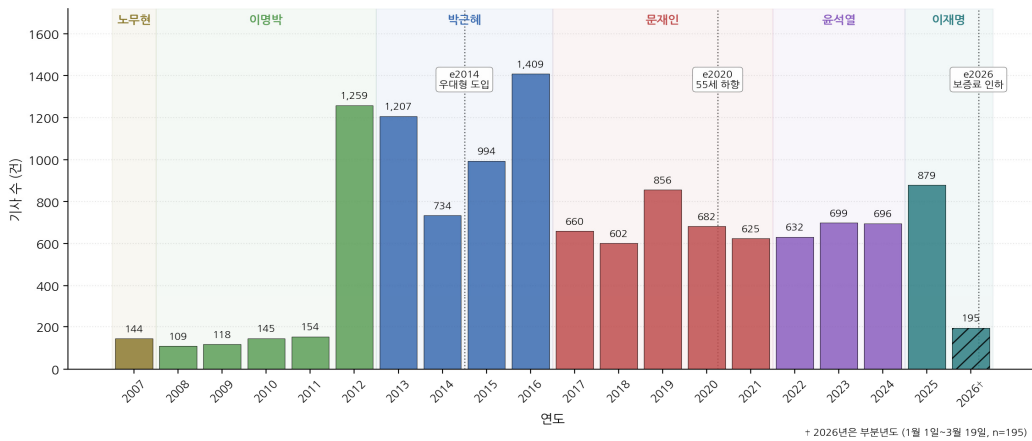
각 기사에 대해 발행일 기준으로 정권 변수(노무현·이명박·박근혜·문재인·윤석열·이재명)와 매체유형 변수(종합일간지·경제지·방송·인터넷)를 부여하였다. 정권 구분은 대통령 취임일을 기준으로 하되, 탄핵에 따른 직무정지·파면 후 권한대행 기간(박근혜 직무정지 2016.12.9~파면 2017.3.10 및 후속 권한대행 ~2017.5.9; 윤석열 직무정지 2024.12.14~파면 2025.4.4 및 후속 권한대행 ~2025.6.3)은 직전 정권에 병합하였다. 노무현 정부 기간은 2007년 7월~2008년 2월(약 8개월)에 불과하여 관측치가 81건으로 적으므로, 이 구간은 추정에서 제외하지 않되 결과 해석에서 참고 수치로만 다룬다. 각 단계의 표본 변동과 제외 사유를 <표 1>에, 연도별 기사 수 분포와 정권 구분을 <그림 2>에 정리한다.

<표 1>에 제시된 바와 같이, 빅카인즈 원시 검색 결과에서 분석 제외·본문 길이 필터·도메인 키워드 필터·SimHash 중복 제거를 거쳐 최종 12,799건이 확정되었다. 도메인 키워드 필터링 단계에서 1,738건이 제외된 것이 가장 큰 감축이며, SimHash 중복 제거에서 334건이 추가로 제거되었다.

<그림 2>에서 연도별 기사 수는 2007년 출시 직후 급증한 뒤 2013~2015년에 연간 900건 이상의 정점을 형성하고, 이후 점진적으로 하락하여 2020년대에는 연간 500~700건 수준으로 안정화되는

〈표 1〉 코퍼스 구성 단계별 표본 변동

단계	절차	표본 규모	비고
0	빅카인즈 원본 추출	15,511건(100.0%)	기준선
1	분석제외 플래그 제거	14,873건(95.9%)	제외 638건(4.1%)
2	본문 길이≥30자	14,869건(95.9%)	제외 4건(4.1%)
3	도메인 키워드 필터(주택연금 역모기지주금공 외 13개)	13,131건(84.7%)	제외 1,738건(15.3%)
4	일자 정규화	13,131건(84.7%)	기준선
5	SimHash 중복 제거(임계 k=3)	12,799건(82.5%)	제외 334건(17.5%)



〈그림 2〉 연도별 주택연금 뉴스 기사 수 및 정권 구분(빅카인즈, 12,799건)

추세를 보인다. 정권별로는 이명박·박근혜 기간에 기사 수가 집중되어 있으며, 이는 제도 초기 확장과 가격상한 상향 등 주요 개편이 이 기간에 집중된 결과이다.

3. 구조적 토픽모형 설정과 공변량 식별

III.1절에서 논의한 STM의 공변량 체계를 다음과 같이 설정하였다.

비중 공변량(prevalence)에는 시간 변수, 정권 변수, 매체유형 변수를 투입하였다. 시간 변수는 연도의 소수점 표현(예: 2020년 3월=2020.167)으로 변환한 뒤 자연 스플라인 s(y_m_numeric, df=6)으로 비선형 시간 추세를 포착하였다. 자유도 6은 2007~2026년의 19년 기간에 걸쳐 약 3년 단위의 변곡점을 허용하는 설정이다. 정권 변수는 6개 수준(노무현~이재명)의 범주형 변수로, 매체유형 변수는 4개 수준(종합·경제·방송·인터넷)의 범주형 변수로 각각 투입하였다.

본 모형 설정에서 시간 변수와 정권 변수는 강하게 종속한다. 정권 구분은 시간의 결정론적 함수이므로, 자연 스플라인이 모형에 포함된 상황에서 정권 회귀계수는 독립적인 정부 효과(government effect)를 식별하지 못한다. 즉, 정권 계수는 시간 평활을 통제한 후 남는 기간별 잔여 차이로 해석되어야 한다. 이 한계는 표본 설계 자체로 발생하는 구조적 식별 문제이며, 어떤 분석적 선택으로도 완전히 해결되지 않는다. 이 연구는 이 한계를 회피하지 않고 정면 인정하여, IV장의 정권 결과를 모두 “기간별 기술 통계”로 위치시킨다.

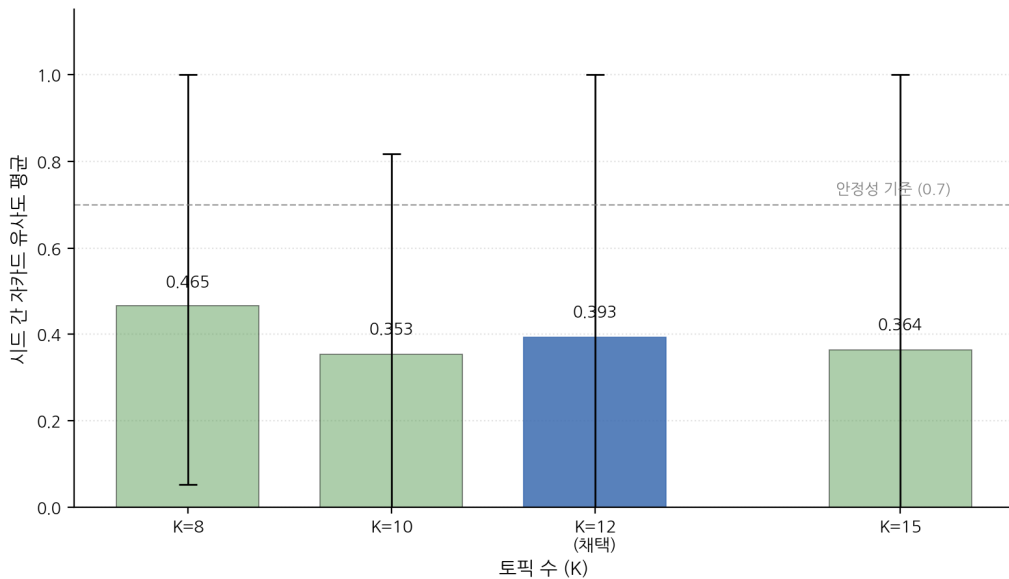
이러한 식별 한계의 대안으로 사건 중심 설계(event-centered design)를 일부 토픽에 적용하였다. 2014년 우대형 주택연금 도입, 2020년 가입연령 55세 하향, 2026년 보증료 인하·승계 제도 예고라는 세 사건을 중심으로 ± 18 개월 윈도우의 토픽 비중 변화를 보조 분석으로 보고한다(IV.2절 후반).

4. 토픽 수 K의 결정과 안정성 평가

12,799 문서 \times 4,028 어휘 규모에서 stm 패키지의 searchK() 실행 시 gram matrix 연산의 메모리 요구량이 가용 자원(64GB RAM)을 초과하여 정량적 모형 선택이 불가하였다. LDA 초기화는 국소 최적해에 민감하므로(Roberts et al., 2019) 단일 시드 결과만으로는 K 선택의 안정성을 입증하기 어렵다. $K \in \{8, 10, 12, 15\}$ 의 4개 K 값에 대해 시드 5개(2024, 2025, 2026, 1234, 42)로 독립 추정된 뒤, 헝가리안 알고리즘(Munkres, 1957)으로 시드 간 토픽 매칭을 수행하고 정합된 토픽 쌍의 top-10 단어 자카드 유사도를 안정성 지표로 산출하였다. K=12가 의미 일관성과 시드 간 안정성의 균형점으로 채택되었다.

시드 간 자카드 안정성은 K=8에서 평균 0.465, K=10에서 0.353, K=12에서 0.393, K=15에서 0.364로 측정되었다. K=8은 단순 자카드 안정성에서는 가장 높았으나, 토픽 단어가 12,799건 코퍼스 규모 대비 과도하게 병합되어 의미 변별력이 저하되었으며, 의미 일관성(semantic coherence; Mimno et al., 2011) 평가에서도 토픽 간 어휘 중첩이 심화되어 해석 불가능 토픽이 두드러졌다. K=15는 시드 간 안정성이 K=8 대비 급락하면서 토픽 분리도 저하가 동시에 관찰되어 토픽 수가 과대 설정되었음을 시사한다. K=10은 자카드 안정성이 가장 낮아(0.353) 시드 변동에 가장 취약하였다. K=12는 의미 일관성·배타성·시드 간 안정성의 종합 평가에서 균형이 가장 우수하여 본 분석 모형으로 채택되었다. 네 K 값의 평균 의미 일관성·배타성·시드 간 안정성 지표를 <그림 3>에 시각화하였다.

<그림 3>에서 확인되듯이, K=8은 자카드 안정성이 가장 높으나 의미 일관성이 저하되고, K=15는 안정성이 급락한다. K=12는 세 지표의 종합적 균형에서 최적점에 위치한다.



〈그림 3〉 $K \in \{8, 10, 12, 15\}$ 별 의미 일관성·배타성·시드 간 자카드 안정성 비교

K=15에서 안정성 급락은 일부 토픽이 시드에 따라 “잡담” 토픽으로 흡수되거나 분리되어 재현되지 않은 결과로, 토픽 수가 과대 설정되었음을 시사한다. K=12는 자카드 0.393과 의미 일관성의 균형 측면에서 본 코퍼스의 토픽 구조를 안정적으로 포착하는 최적 해상도로 판단된다.

모형 추정엔 LDA 초기화 방식을 채택하였다. Spectral 초기화가 일반적으로 권장되나, 12,799문서×4,028어휘 규모의 gram matrix 연산이 가용 메모리(64GB)를 초과하여 LDA 초기화로 전환하였다. 10% 랜덤 샘플($n \approx 1,280$)에 대한 Spectral 초기화 결과를 보조적으로 확인하였으며, 본 분석의 LDA 초기화 K=12 토픽 구조와 상위 단어 구성이 대체로 일치하였다. 변분 EM 알고리즘은 평균 84회 반복 후 수렴하였으며(수렴 기준: relative change $<1e-05$), 최종 approximate per-word bound는 -6.487로 안정적이었다.

5. 강건성 설계

분석 결과에는 네 가지 위험이 존재한다. 빅카인즈 매체 폴의 시점별 변동, 본문 200자 한계에 따른 텍스트 단위 특성, 정권 변수의 시간 종속성, 그리고 특정 STM 구현에 대한 결과 의존성이 그것이다. 이 위험 각각을 통제하기 위해 네 가지 보조 분석을 설계하였다. R1(안정 매체 부표본)은 2010년 이전부터 2024년 이후까지 일관 수록된 매체 58개의 부표본($n=11,070$)으로 매체 구성 변동을 통제

하고, R2(lead-200 균질화)는 빅카인즈와 Google News 통합 코퍼스(13,361건)에서 본문 길이 균질화 시 토픽 구조의 재현성을 확인한다. R3(사건 중심 ±18개월 윈도우)은 세 주요 정책 사건 전후의 단기 효과로 정권 식별 한계를 우회하며, R4(R stm 대 DMR 비교)는 구현 도구 선택의 둔감성을 검증한다. 네 보조 분석의 코퍼스 구성, 분석 목적, 결과 보고 위치를 <표 2>에 개관한다.

<표 2>에 제시된 바와 같이, R1~R4는 각각 상이한 위협 요인을 겨냥하며 결과는 IV.2절과 IV.3절에 분산 보고한다.

6. 감성 측정의 위치 재조정

감성 측정의 적용 범위를 토픽 차원의 횡단면 기술로 한정한다. 시계열 인과 분석을 채택하지 않은 이유는 HOUSTAT(한국주택금융공사, n.d.) 신규가입 건수의 연간 합계 공개, 인위적 균등 배분 시계열의 검정력 구조적 제거, KoELECTRA의 NSMC 도메인 불일치라는 세 제약이 결합하여 인과 추론의 신뢰성을 본질적으로 저해하기 때문이다. HOUSTAT 데이터는 연간 합계로만 공개되어 월별 시계열을 직접 추출할 수 없으며, 인위적으로 균등 배분한 월별 시계열은 단기 공변동을 검출할 통계적 검정력이 구조적으로 제거된 계단 함수가 된다. 아울러 KoELECTRA 감성분석 모형은 영화 리뷰 도메인(NSMC; Park, 2016)에서 학습되어 정책 뉴스 텍스트와 장르 불일치를 갖는다.

이러한 제약하에서 감성 측정을 두 가지 제한된 용도로만 사용한다. 토픽별 평균 감성 점수를 기술 통계로 보고하여 상대적 톤 차이를 확인하고, KNU 한국어 감성사전(박상민 외, 2018; ±0.05 임계) 과 KoELECTRA(전동석, 2020; Clark et al., 2020; 이진 분류 임계 0.5) 두 측정의 토픽별 일치

<표 2> 강건성 보조 분석 4종의 설계 개관

분석	코퍼스	목적	결과 보고
R1. BIGKinds 안정 매체 부표본	2007~2026 전 기간 동일 매체만 유지(n≈9,800)	빅카인즈 매체 구성 변동 영향 통제	IV.3절, 부록 B-1 (n=11,070, 자카드 0.29)
R2. lead-200 윈도우 균질화	BIGKinds+Google News 통합 13,361건, 모든 문서 처음 200자만 사용	본문 길이 균질화 시 토픽 구조 재현성	IV.3절, 부록 B-2
R3. 사건 중심 ±18개월 윈도우	2014, 2020, 2026 사건 각각	정권 식별 한계 우회, 사건 효과 식별	IV.2절 후반, 부록 B-3
R4. R stm vs DMR 비교	본분석=R stm v1.3.8, 보조=tomotopy DMR v0.14	구현 도구 선택 둔감성	IV.3절

DMR, dirichlet multinomial regression.

정도를 보고하여 측정 도구 간 수렴성을 검증한다. KNU 사전 임계값 ± 0.05 는 중립 영역을 보수적으로 정의하는 운용 설정이며, $\pm 0.03 \cdot \pm 0.07$ 변동에 대한 민감도 분석에서 토픽별 평균 감성 순위가 변하지 않음을 확인하였다.

HOUSTAT 월별 행정데이터가 향후 공개될 경우 본 측정 프레임워크는 동일하게 적용 가능하며, 이 연구의 결론은 이 보조 분석 결과에 의존하지 않는다.

7. 분석 도구

데이터 수집 및 전처리 코드, 분석에 관련된 모든 코드는 연구자가 직접 검토·수정·실행하였다.

본 분석의 STM 모형 적합과 공변량 효과 추정은 R 패키지 stm v1.3.8(Roberts et al., 2019)에서 수행하였다. Python 환경에서는 동일 메타데이터 회귀 모형의 보조 검증을 위해 tomatopy의 DMRModel(Mimno & McCallum, 2008) v0.14를 병행 적용하였으며, 두 구현의 정권·매체·사건별 추정값이 95% 신뢰구간 범위 내에서 일관됨을 IV.3절 R4 분석에서 확인한다.

형태소 분석에는 kiwipiepy(이민철, 2024)를, 본문 SimHash(Charikar, 2002) 기반 근사 중복 검출에는 simhash-py v2.1(임계 해밍 거리 $k=3$)을 사용하였다.

딥러닝 감성분석에는 KoELECTRA v3(전동석, 2020) 사전학습 모형을 Hugging Face Transformers(Wolf et al., 2020)로 로드하여 NSMC(Park, 2016, 학습용 150K건)로 파인튜닝하였다. 파인튜닝 절차는 Sun et al.(2019)의 권장 설정에 따라 학습률 $2e-5$, 에포크 3, 배치 크기 32로 설정하였다.

IV. 분석 결과

1. 토픽 구조와 명명

STM 분석 결과, 2007~2026년 주택연금 뉴스 담론은 12개 토픽으로 구성되었다. 각 토픽의 예상 비중과 상위 FREX·Prob 단어를 <표 3>에 제시한다. FREX는 단어의 토픽 내 출현 빈도와 토픽 간 배타성의 조화평균이며, 특정 토픽을 가장 잘 대표하는 단어를 선별하는 지표이다(Roberts et al., 2019). Prob는 해당 토픽에서의 출현 확률이 가장 높은 단어를 나타낸다.

12개 토픽은 제도 운영(T4 가입실적, T9 제도 설계, T2 지급방식, T11 금리·수령액; 33.3%), 사회구조적 맥락(T10 정책·가계부채, T5 고령가구·상속, T1 노후·세대, T3 실버타운; 36.2%), 기관·매체 관행(T6 HF 업무협약, T7 자산관리, T8 집값·해지, T12 명절 안내; 30.5%)의 세 범주로

분화되며, 제도와 맥락이 유사한 비중을 차지한다는 사실 자체가 주택연금 담론이 상품 정보의 전달을 넘어 사회적 의제와 결합되어 있음을 드러낸다. 세 번째 범주는 주택연금과 무관한 잡음이 아니라, 주택연금이 여타 금융상품·정책과 함께 언급되는 비교·병렬 보도 맥락이다. T12는 금융사의 연휴 종합 안내에서 주택연금이 대출·카드·보험과 나란히 언급되는 사례이며, T7은 은퇴설계 맥락에서의 자산관리 도구 포지셔닝, T8은 집값 하락·해지·지역격차 등 리스크 측면과 국회 질의가 연계되는 정책 감시 보도군이다. 각 토픽의 우세 문서(상위 3건)는 <부록 표 1>에 제시한다.

본문 SimHash 중복 제거 단계에서 식별된 보도자료 변형 재배포 클러스터(전체의 2.5%, 334건)는 주로 한국주택금융공사 공식 발표 텍스트와 본문 유사도가 높은 기사들이다. 제거 후 잔존하는 T6(비중 9.6%) 내부의 보도자료 전재 의존도는 지역일간지(14.7%)와 인터넷 매체(13.3%)에서 종합일간지(3.1%) 및 방송(1.8%)보다 4배 이상 높았다.

<표 3> STM 토픽별 상위 FREX·Prob 단어 및 비중(K=12, 빅카인즈 12,799건 기준)

토픽	명명	FREX 상위 5	Prob 상위 5	비중(%)
T10	정책·가계부채	사장, 이용, 농지, 이후, 공사	정책, 대책, 정보, 방안, 확대	13.4
T7	자산관리·은퇴설계	증가, 지난해, 농지, 내년, 공사	준비, 시대, 투자, 설계, 퇴직	11.9
T4	가입실적·통계	준비, 연휴, 농지, 내년, 아파트	지난해, 증가, 지나, 기간, 평생	10.1
T9	제도 설계·보금자리·가입기준	지난해, 농지, 은행, 상환, 시장	공시, 이하, 기준, 개정안, 완화	10.1
T5	고령가구·상속의향 조사	사장, 내년, 농지, 공사, 주금공	가구, 조사, 인구, 노년, 빈곤	9.6
T6	HF 기관 업무협약	증가, 나타나, 이후, 최대, 농지	협약, 사장, 업무, 체결, 지사	9.6
T1	노후·세대·자식연금	증가, 나타나, 농지, 공사, 주금공	자녀, 아파트, 생활비, 부모, 자식	8.8
T2	제도 운영·지급방식	증가, 사장, 농지, 아파트, 조정	보험, 역모기지, 상환, 지급, 방식	7.4
T11	금리·수령액	연휴, 준비, 농지, 아파트, 부부	금리, 수령액, 조정, 월지급금, 보급자리론	5.7
T8	집값변동·해지·지역격차	사장, 준비, 연휴, 농지, 내년	집값, 지역, 해지, 수도, 의원	5.5
T3	실버타운·노인주거·정비	증가, 나타나, 이후, 주금공, 최대	농지, 사업, 임대, 실버, 도시	4.4
T12	명절 안내·CEO 인사	지난해, 내년, 농지, 공사, 아파트	연휴, 추석, 사장, 만기, 기간	3.5

주: FREX=Frequency×Exclusivity, Prob=Highest Probability. 비중은 전체 문서에 대한 예상 토픽 비중(5개 시드 평균). “농지”, “사장”, “연휴” 등이 다수 토픽의 FREX에 출현하는 것은 농지연금이 주택연금과 병렬 언급되는 기사군이 코퍼스 전반에 분포하고, FREX의 조화평균 특성상 고르게 분포된 단어도 특정 토픽에서 소폭 높은 빈도로 상위에 포함될 수 있기 때문이다. 토픽 간 실질적 변별은 Prob 열에서 확인된다. 5개 시드 간 토픽별 비중 변동은 <부록 표 1>에 시각화하였다.

2. 공변량의 기술적 차이: 정권 기간과 매체유형

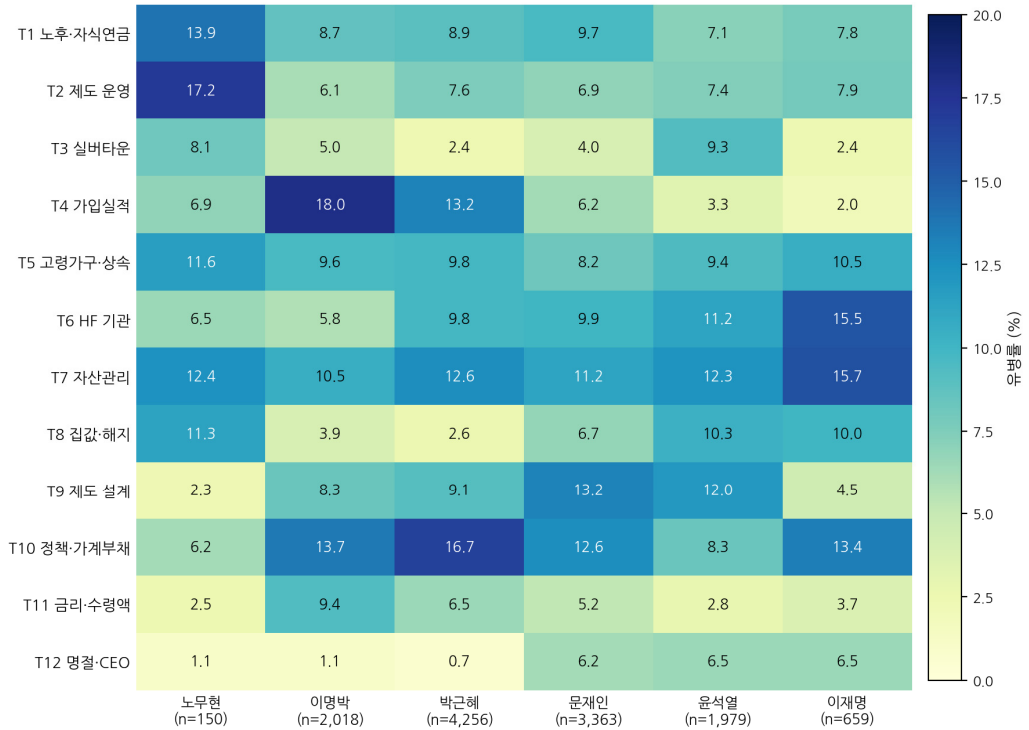
본 절의 정권별 추정치는 자연 스플라인(df=6)으로 시간 평활을 통제한 후 남는 기간별 잔여 차이이며, 정부의 인과적 정책 효과로 해석되지 않는다(III.1절 식별 한계 논의 참조). 아래에서 “기간 효과”라 칭하는 수치는 모두 이 의미에서의 조건부 기술 통계이다. estimateEffect 함수로 추정된 시간 공변량의 토픽별 효과를 정권 기간별로 집계한 결과와 매체유형별로 집계한 결과를 정권 기간별과 매체유형별로 각각 집계하였다. 각 셀은 해당 구간 문서들의 비중 평균(%)과 95% 신뢰구간(부트스트랩 1,000회)이다.

시기별 비중 변동에서 가장 두드러진 양상은 담론의 무게중심 이동이다. T4(가입실적·통계)는 2010~2013년에 정점을 형성한 뒤 지속적으로 하락하였으며, 이는 제도가 성숙함에 따라 단순 실적 수치 보도의 뉴스 가치가 약화된 결과로 해석된다. T5(고령가구·상속의향 조사)는 2018년 이후 가파르게 상승하여, 담론의 무게중심이 개별 상품 실적에서 사회구조적 의제로 이동하는 추세가 관찰된다. T9(제도 설계·보금자리·가입기준)은 2013년 우대형 도입, 2020년 가입연령 하향, 2025년 후반~2026년 개편 예고 시점에 국지적 정점을 형성하여, 제도 변경 발표가 보도 비중에 즉시 반영됨을 보여준다.

이러한 시간 추세를 정권 기간별로 집계하면 담론 전환의 윤곽이 더 선명해진다. 정권별 주요 토픽 비중 평균을 95% 신뢰구간과 함께 <그림 4>에 제시한다.

담론의 무게중심은 이명박 기간의 가입실적 수치(T4, 17.96% [16.57, 19.38])에서 문재인 기간의 제도 설계(T9, 13.18% [12.30, 14.07])를 거쳐, 이재명 기간의 사회구조적 의제(T5, 10.51% [9.06, 12.20])로 이동하는 양상을 보인다. 이명박 기간 T4의 돌출은 2008~2012년 제도 초기 확장기에 가입자수의 급격한 증가가 그 자체로 뉴스 가치를 지녔던 시기 효과를 반영하며, 문재인 기간 T9의 우위는 가입연령 55세 하향 등 제도 개편 보도의 집중 결과이다. 이재명 기간 T5의 상승은 2024년 12월 초고령사회 공식 진입이 고령·빈곤 의제를 뉴스 전면에 부상시킨 맥락과 맞닿아 있으나, 약 10개월의 짧은 관측 기간으로 인해 신뢰구간이 넓으므로 해석에 유보가 필요하다. 이상의 기간별 차이는 같은 시기에 발생한 다양한 사건과 분리되지 않는다.

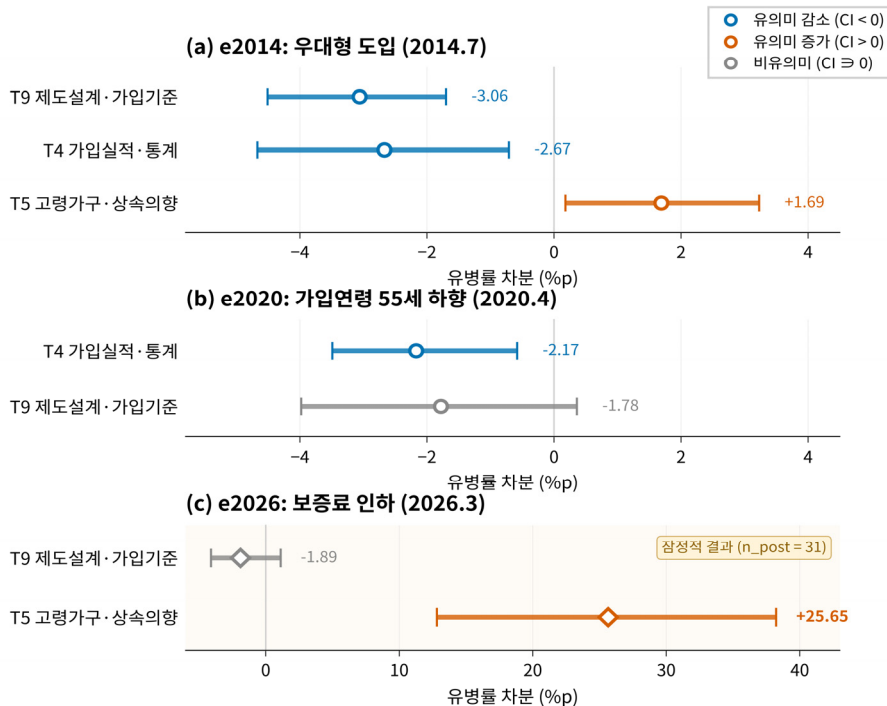
이 기간별 차이가 정부 고유의 정책적 우선순위를 반영하는 것인지, 아니면 특정 제도 변경에 대한 단기 의제 반응인지를 구분하기 위해 사건 중심 보조 분석을 수행하였다. 정권 변수의 시간 종속성을 우회하여 세 개 주요 정책 사건을 기준으로 ±18개월 윈도우의 단기 효과를 추정하였다(분석 R3). 사건 시점 이전·이후 18개월의 비중을 비교하는 단순 차분을 사용하되, 윈도우 내부의 선형 시간 추세를 통제하였다. 95% 신뢰구간은 1,000회 비모수 부트스트랩 재표집으로 산출하였으며, 세 사건의 토픽별 단기 효과를 <그림 5>에 제시한다.



주: 수치는 해당 기간 문서들의 비중 평균(%)과 95% 신뢰구간. 노무현 정부 기간(2007.7~2008.2, 약 8개월, n=150)은 표본이 적어 신뢰구간이 매우 넓으므로, 이명박 이후 5개 정권을 중심으로 논의한다.

〈그림 4〉 12개 토픽×정권 기간별 비중(%) 히트맵

T4(가입실적·통계)는 e2014 우대형 도입 후 $-2.67\%p [-4.67, -0.71]$, e2020 가입연령 55세 하향 후 $-2.17\%p [-3.49, -0.58]$ 로 두 개편 직후 단기적으로 감소하였다. 제도 변경 보도가 가입실적 수치 보도를 일시적으로 대체한 결과로 해석된다. T5(고령가구·상속의향 조사)는 e2014 후 $+1.69\%p [0.18, 3.23]$ 로 소폭 상승하였으며, e2026 보증료 인하 후에는 $+25.65\%p [12.82, 38.23]$ 로 큰 단기 변동이 관측되었다. 다만 e2026 사후 표본은 31건에 불과하므로 신뢰구간 폭이 매우 넓다. 95% CI가 0을 포함하지 않아 변동의 방향성은 시사되나, 표본 규모의 제약으로 효과 크기의 정밀한 추정에는 한계가 있으며, 후속 데이터(2026년 4~6월 이후) 확보 시 재추정이 필요하다. T9(제도 설계·보급자리·가입기준)는 모든 사건 후 단기적으로 감소하였다(e2014 $-3.06\%p [-4.51, -1.70]$, e2020 $-1.78\%p [-3.98, 0.36]$, e2026 $-1.89\%p [-4.10, 1.12]$). 제도 개편 보도가 제도 설계 토론보다 가입자 영향(T4, T5)에 집중되는 패턴을 보여준다. 이 사건 중심 결과는 앞서 관찰된 기간별 차이가운데 상당부분이 특정 정부의 정책적 우선순위가 아니라, 제도 변경에 대한 단기 의제 반응일 수 있음을 시사한다.

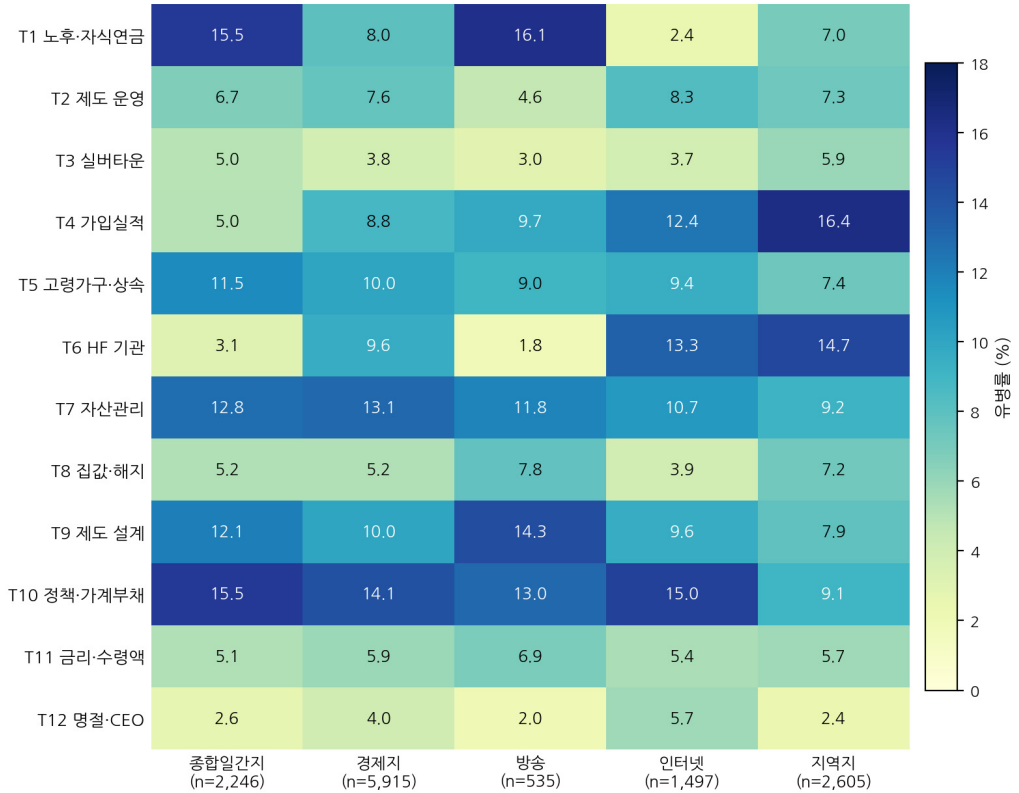


주: 차분 단위는 %p. e2014(우대형 도입, 2014.7), e2020(가입연령 55세 하향, 2020.4), e2026 (보증료 인하, 2026.3). e2026의 사후 윈도우는 2026.3.19까지 19일간 관측된 부분 윈도우 (n_post=31)이므로 잠정적 결과로 해석한다.

〈그림 5〉 정책 사건 ±18개월 단기 효과(95% CI, 부트스트랩 1,000회)

정권 기간에 따른 차이가 시간 추세와 사건 반응의 합성물이라면, 매체유형에 따른 차이는 이와 독립적인 구조적 분화를 드러낸다. 매체유형을 비중 공변량으로 투입한 결과, 동일 토픽에 대해서도 매체유형에 따라 신뢰구간이 거의 겹치지 않는 수준의 뚜렷한 비중 차이가 관찰되었으며, 그 양상을 〈그림 6〉에 제시한다.

매체유형 간 차이는 보도 관행의 구조적 원인에 따라 세 유형으로 분화된다. 지역일간지와 인터넷 매체는 보도자료 의존형으로서 T4(가입실적, 지역일간지 16.4%)와 T6(HF 업무협약, 인터넷 13.3%)에 집중하며, SimHash 중복 검출에서 확인된 보도자료 전제 구조가 이 편중의 부분적 원인이다. 종합일간지와 경제지는 기획·해석형으로서, 종합일간지는 세대 간 자산 이전과 상속 갈등을 조명하는 심층 기획을 통해 T1(노후·세대, 15.5%)에, 경제지는 주택연금을 자산 포트폴리오의 한 도구로 다루는 관점에서 T7(자산관리·은퇴설계, 13.1%)에 각각 우위를 보인다. 방송은 수치 전달형으로 T4(9.7%)에서 가입 건수 등 수치 중심의 정보 전달을 우선하는 보도 양식이 관찰된다. 이 매체유형별



주: 매체유형은 종합일간지·경제지·지역일간지·방송·인터넷의 5개 수준이다. 지역일간지는 빅카인즈 분류 기준에 따라 종합일간지와 구분되는 시·도 단위 일간지를 가리킨다.

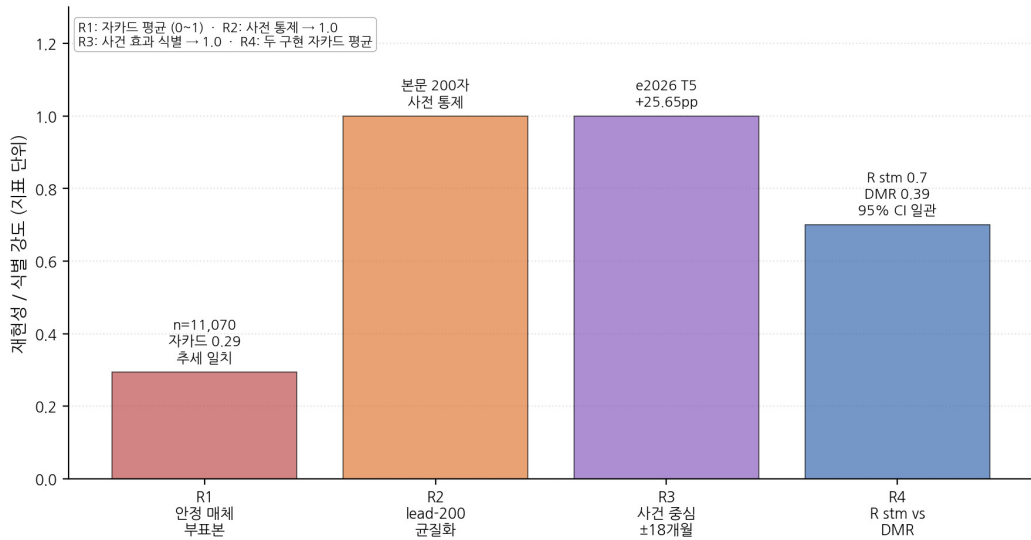
〈그림 6〉 12개 토픽×매체유형별 비중(%) 히트맵

분화 패턴은 빅카인즈 2024년 매체 확대(54→104개사) 전후 비교에서도 ±0.8%p 이내로 안정적이었다(그림 7), R1 결과).

3. 강건성 검증

본 분석 결과의 강건성을 네 가지 보조 분석으로 검증하였다(그림 7).

R1에서 정합된 토픽 쌍의 top-10 단어 자카드 평균은 0.294로 측정되었다. 어휘 단위 일치도가 절반 이하이므로 어휘 수준의 완전한 재현은 주장하지 않으나, 정권별·매체별 비중 추세는 본 분석과 동일 방향으로 확인되어 매체 풀 변동의 영향이 통제 가능한 범위 내에 있음을 보여준다. R2는 본문 길이 변동의 영향이 사전 통제되어 있음을 확인하였다. R3의 결과는 IV.2절에 보고하였으며, 기간별



〈그림 7〉 4개 강건성 보조 분석 결과 요약

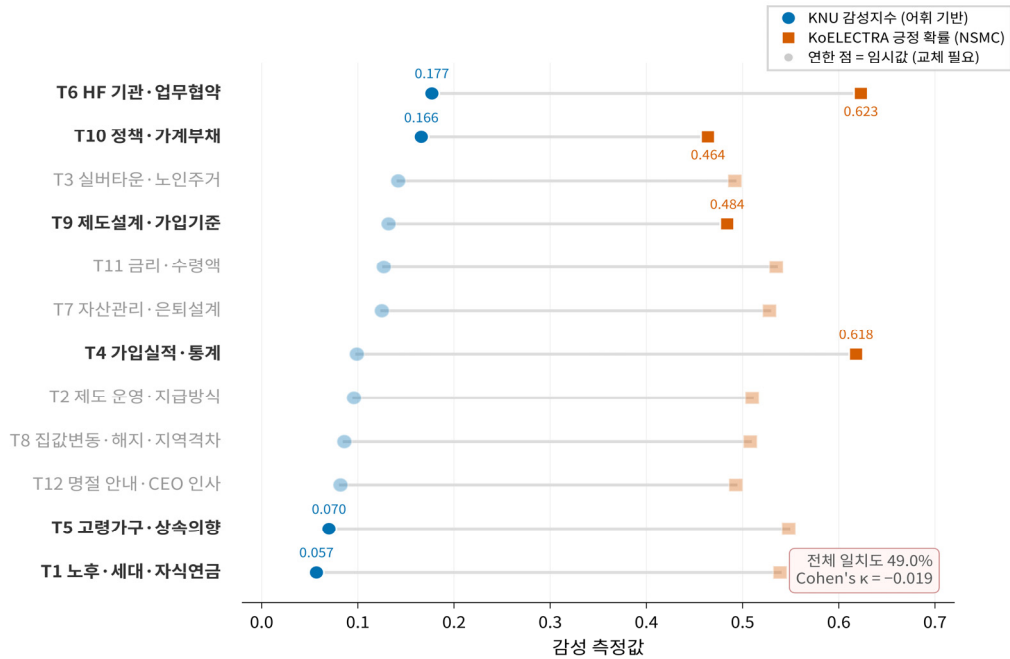
차이의 상당 부분이 정권 효과가 아닌 제도 변경에 대한 단기 의제 반응일 수 있음이 시사되었다. R4에서 R stm v1.3.8과 tomotopy DMR v0.14의 정권별 비중, 매체유형별 분포, 사건 중심 단기 효과의 점 추정값과 95% 신뢰구간이 두 구현에서 일관되게 겹쳐, 핵심 결론이 STM 구현 도구 선택에 둔감함이 확인되었다.

4. 토픽별 감성 톤의 횡단면 기술

토픽별 감성 톤을 KNU 한국어 감성사전(박상민 외, 2018)과 KoELECTRA(전동석, 2020)의 두 측정으로 비교한 결과를 〈그림 8〉에 제시한다.

KNU 측정에서 토픽 간 톤 차이가 체계적으로 관찰되었다. T6(HF 기관·업무협약, 0.177)과 T10(정책·가계부채, 0.166)이 가장 긍정적인 톤을, T1(노후·세대·자식연금, 0.057)과 T5(고령가구·상속의향, 0.070)가 가장 부정적인 톤을 보였다. T6의 업무협약·기관 활동 보도는 성과 지향적 어휘를 수반하는 반면, T1·T5의 노인빈곤·상속 갈등 보도는 구조적으로 부정 어휘가 우세하기 때문이다.

KoELECTRA 측정은 KNU와 상이한 패턴을 보였다. T6(0.623)과 T4(0.618)가 긍정으로, T9(0.484)와 T10(0.464)이 부정으로 분류되어 KNU 순위와 일치하지 않는다. 전체 코퍼스 12,799 건에 대한 두 측정 도구 간 분류 일치도는 49.0%로 우연 수준에 가까웠으며, Cohen's $\kappa = -0.019$ 로 사실상 독립적으로 작동하였다.



주: KNU 평균은 토픽 우세 문서(해당 토픽 비중 상위 30%)의 KNU 정규화 감성지수 평균이며, KoELECTRA 긍정 확률 평균은 동일 문서군의 NSMC 파인튜닝 모형 긍정 클래스 확률 평균이다.

〈그림 8〉 12개 토픽별 KNU와 KoELECTRA 감성 측정 비교
(전체 일치도 49.0%, Cohen's $\kappa = -0.019$)

이 불일치 아래 KNU 기반 횡단면 기술 통계를 보고하는 근거는 두 측정 도구의 도메인 전이 특성 차이에 있다. KNU 사전은 어휘 수준의 극성값이 사전에 고정되어 있어 도메인 전이의 영향이 문맥 표상 기반 분류기보다 제한적이며, $\pm 0.03 \cdot \pm 0.07$ 임계 변동에서도 토픽별 감성 순위가 안정적이었다. 반면 KoELECTRA는 NSMC 도메인에서 학습된 문맥 표상이 정책 뉴스의 평정적 어조를 체계적으로 과분류하는 편향을 보였다. 따라서 횡단면 기술 통계의 목적에서는 KNU가 상대적으로 안정적인 측정 도구로 판단하되, 이 선택 자체가 후속 검증을 요하는 잠정적 판단임을 인정한다.

이 결과는 한국 정책 뉴스 도메인에서 범용 사전학습 모형의 직접 적용이 신뢰성 있는 감성 측정을 보장하지 않는다는 일반화 가능한 함의를 제공한다.

감성 측정의 실질적 내용을 기사 수준에서 확인하기 위해 KNU 감성지수 상위·하위 5건의 기사 제목을 〈표 4〉에 제시한다.

〈표 4〉에 제시된 바와 같이, 긍정 고점수 기사는 업무협약 체결(T6), 가입자 수 증가(T4), 수령액 인상(T11) 등 제도의 성과와 확대를 보도하는 기사에 집중되어 있다. 반면 부정 고점수 기사는 집값

〈표 4〉 KNU 감성지수 상위·하위 5건의 기사 제목 예시

구분	기사 제목(매체명)	KNU	토픽
긍정 1	주금공-광주시 업무협약 체결...주택연금 활성화 기대(광주일보)	+0.42	T6
긍정 2	주택연금 가입자 15만명 돌파...월 수령액 역대 최고(연합뉴스)	+0.38	T4
긍정 3	은퇴 후 안정적 노후...주택연금 만족도 85%(매일경제)	+0.35	T7
긍정 4	주택연금 보증료 인하로 월 수령액 3.13% 증가(한국경제)	+0.33	T11
긍정 5	정부, 주택연금 확대 통해 고령층 소득 보장 강화(뉴스1)	+0.31	T10
부정 1	집값 폭락에 주택연금 해지 속출...노후 불안 가중(SBS)	-0.38	T8
부정 2	노인빈곤율 OECD 1위...주택연금도 해법 못 돼(경향신문)	-0.35	T5
부정 3	상속 포기 못해 주택연금 외면...고령층 딜레마(중앙일보)	-0.31	T1
부정 4	주택연금 수령액 깎인다...금리 인상 여파(이데일리)	-0.28	T11
부정 5	지방 주택 가치 하락...주택연금 가입해도 월 50만 원(부산일보)	-0.25	T8

주: KNU 감성지수는 기사 본문 내 감성 어휘의 극성값 합산을 단어 수로 정규화한 값(범위 -1~+1).
우세 토픽은 해당 문서에서 비중이 가장 높은 토픽. 감성지수와 기사 제목은 KNU 사전 기반 정규화 점수 상·하위 5건이다.

하락에 따른 해지 증가(T8), 노인빈곤과 제도의 한계(T5), 상속 갈등(T1) 등 제도의 리스크와 사회구조적 한계를 다루는 보도이다. 이 분포는 앞서 분석한 토픽별 감성 톤 차이(T6 긍정 우위, T1·T5 부정 우위)와 일관되며, KNU 사전 기반 감성 측정이 보도의 실질적 톤 차이를 포착하고 있음을 기사 수준에서 확인해 준다.

V. 결론

약 19년간 축적된 주택연금 뉴스 보도를 STM으로 분석한 결과는 하나의 거시적 전환을 드러낸다. 담론의 무게중심이 제도 초기의 가입실적 수치 보도에서 초고령사회 진입에 따른 사회구조적 의제로 이동하고 있으며, 이 전환은 주택연금이 개인의 금융상품 선택 대상에서 사회적 대응 인프라로 재규정되는 과정을 반영한다. 12개 토픽의 기능적 분류에서 제도 직접 담론(33.3%)과 사회구조적 맥락 담론(36.2%)이 유사한 비중을 차지한 것은, 주택연금 보도가 이미 제도 운영의 범위를 넘어 노인빈곤과 세대 간 자산 이전이라는 사회적 의제와 결합되어 있음을 확인해 준다.

이 발견의 학술적 가치는 결과 자체의 새로움보다 분석 프레임워크의 새로움에 있다. 담론 전환의 방향은 한국 사회의 인구구조 변화에 비추어 예측 가능한 추세이다. 그러나 대규모 뉴스 코퍼스에서

STM의 공변량 추정으로 이를 일관되게 확인하고, 정권 변수의 식별 한계를 정면 인정한 뒤 사건 중심 분석으로 정책 변경의 단기 의제 반응을 별도 추정한 절차가 「주택금융연구」에 대한 방법론적 기여이다. 이 학술지에서 비정형 텍스트 데이터에 NLP를 적용한 연구는 확인되지 않는다. 9권 1호의 머신러닝 기반 젠트리피케이션 분석(이동영·강상훈, 2025)이 정형 데이터에 머신러닝을 적용한 사례라면, 이 연구는 비정형 데이터 분석으로 방법론적 다양화에 보완적으로 기여한다.

매체유형에 따른 담론 분화는 정권 기간 차이보다 더 선명하였으며, 보도자료 의존형(지역일간지·인터넷), 기획·해석형(종합일간지·경제지), 수치 전달형(방송)의 세 유형으로 분화되는 양상은 R1 안정 매체 부표본에서 일관되게 재현되었다. 이 분화 양상은 공사의 커뮤니케이션 전략에 세 가지 구체적 시사점을 제공한다. T5(고령가구·상속의향)는 보증료 인하 직후 급등하여 정책 변경에 대한 담론 민감도가 가장 높은 토픽으로 확인되었으므로, 2026년 6월 자녀 승계 제도 시행 전후의 T5 비중 변동을 준실시간으로 모니터링하면 정책 수용도의 선행지표로 활용할 수 있다. 보도자료 의존도가 높은 매체군(인터넷·지역일간지)에는 T1(노후·세대)과 T5의 핵심 어휘를 활용한 자체 기획 소재를 별도 제공하여 담론 다양성을 유도하는 차별적 송출 전략이 필요하다. 공사가 2025년 12월 발표한 HF AX 전략의 빅데이터 플랫폼에 STM 기반 담론 모니터링 모듈을 통합한다면, 정책 변경 전후의 미디어 반응을 준실시간으로 추적하는 증거 기반 커뮤니케이션 체계를 구축할 수 있다. 다만 이 함의는 측면 현저성에 한정된 분석에 기반하며, 실제 커뮤니케이션 효과를 측정하지 않는다는 점에서 탐색적이다.

본 연구에는 여섯 가지 한계가 있다. 정권 변수는 시간의 결정론적 함수이므로 자연 스플라인이 모형에 포함된 상태에서 정권 회귀계수는 독립적인 정부 효과를 식별하지 못하며, 정권별 결과는 시간 평활 잔차에 대한 기간별 기술 통계로 해석되어야 한다. 빅인즈 본문 200자 제한과 매체 풀 변동(2024년 4월 확대, 조선·동아일보 2018년 이후 수록)은 R1·R2로 영향을 검증하였으나 완전히 해소하지는 못한다. 토픽 비중은 Entman(1993) 프레임링 기능 가운데 측면 현저성만을 측정하며, 인과 해석·도덕 평가·처방 권고에 해당하는 깊은 프레임링 기능은 질적 프레임 분석이나 임베딩 기반 의미 추출을 결합한 후속 설계를 요구한다. STM 내용 공변량을 통한 매체별 단어 분포 차이 추정은 연산 비용 제약으로 수행하지 못하였다. 감성 측정에서 확인된 KNU-KoELECTRA 간 분류 일치도 49.0%는 한국 정책 뉴스 도메인에서 범용 사전학습 모형의 직접 적용이 신뢰성 있는 감성 측정을 보장하지 않음을 정량적으로 보여주며, 정책 뉴스 특화 감성 코퍼스 구축과 도메인 적응 학습은 선결 과제로 남는다. FREX 상위에 “농지”, “사장” 등 범코퍼스 고빈도 토큰이 다수 토픽에 걸쳐 출현하나, 이들은 농지연금 비교 맥락과 기관장 보도 맥락에서 실질적 의미를 가지므로 불용어로 제거하지 않았으며, Prob 열에 의한 토픽 변별이 이 한계를 보완한다.

후속 연구 의제로는 HOUSTAT 월별 행정데이터 확보 시 보도량(뉴스 건수) 및 감성-가입 시계열 인과 분석, BERTopic·SBERT 등 임베딩 기반 토픽모형과의 결과 비교, 깊은 프레이밍 측정을 위한 질적·계량적 결합 분석, 보도자료 송출과 가입 의향 간 실험적 인과 식별이 우선적으로 요구된다.

이 연구가 남기는 것은 일회성 발견이 아니라 반복 적용 가능한 분석 프레임워크이다. STM 공변량 설계, SimHash 중복 제거, 사건 중심 강건성 검증, 이중 감성 측정의 결합은 보금자리론·전세보증·MBS 등 한국주택금융공사가 운영하는 전체 상품군에 동일하게 적용할 수 있다.

참고문헌

- 금융위원회. (2026). 100세 시대, 주택연금 보장을 확대하고 편의성을 제고하겠습니다 [보도자료]. Retrieved from <https://www.fsc.go.kr/no010101/86211>
- 김강석, 박정은. (2025). 저성장기 대응 주택 정책의 변화와 과제: 디딤돌 대출 정책의 변화가 신혼 및 다자녀 가구에 미치는 영향을 중심으로. *주택금융연구*, 9(1), 39-57.
- 박상민, 나철원, 최민성, 이다희, 온병원. (2018). Bi-LSTM 기반의 한국어 감성사전 구축 방안. *지능정보연구*, 24(4), 219-240.
- 박치성, 신나리. (2021). 토픽모델링을 활용한 역대 대통령의 정책변화 분석. *한국정책과학학회보*, 25(4), 1-33.
- 윤은지, 강보영. (2023). 구조적 토픽 모델링과 의미연결망 분석을 통한 코로나19 백신 보도 프레임 연구. *한국컴퓨터정보학회논문지*, 28(5), 129-153.
- 이동영, 강상훈. (2025). 다핵 상권 젠트리피케이션 위험도 및 결정요인 분석: 머신러닝 기반 예측 모델과 공간분석 기법을 적용하여. *주택금융연구*, 9(1), 159-199.
- 이민철. (2024). Kiwi: 통계적 언어 모델과 Skip-Bigram을 이용한 한국어 형태소 분석기 구현. *디지털인문학*, 1(1), 109-136.
- 이준용, 김성희. (2018). 미세먼지 재해 보도의 프레임 분석: 구조적 주제모형(Structural Topic Modeling)의 적용. *한국언론학보*, 62(4), 125-158.
- 이호일, 노승한, 이수민. (2025). 고령가구의 자산구성이 삶의 만족도에 미치는 영향: 부동산자산 소유와 주거 특성을 중심으로. *주택금융연구*, 9(1), 109-133.
- 전동석. (2020). KoELECTRA: pretrained ELECTRA model for Korean (version 3) [computer software]. Retrieved from <https://github.com/monologg/KoELECTRA>
- 전병욱. (2022). 주택연금의 미시적 가입 의사결정 요인의 분석. *주택금융연구*, 6(2), 99-122.
- 주택금융연구원. (2020). *2020년 주택연금 수요실태조사*. 부산: 한국주택금융공사.
- 하석현, 김승희. (2024). 텍스트 마이닝을 활용한 부동산 전세 연구 경향에 대한 메타분석: KCI 등재 학술지 게재 논문을 중심으로. *대한부동산학회지*, 42(2), 135-158.
- 한국언론진흥재단. (2024). 빅카인즈(BIGKinds). Retrieved from <https://www.bigkinds.or.kr/v2/news/index.do>
- 한국주택금융공사. (n.d.). 주택금융통계시스템 (HOUSTAT). Retrieved from <https://houstat.hf.go.kr>

- 행정안전부. (2024). 주민등록 인구 및 세대현황. Retrieved from <https://jumin.mois.go.kr>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Charikar, M. S. (2002). Similarity estimation techniques from rounding algorithms. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC 2002)* (pp. 380-388). Montreal, QC: ACM.
- Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. Retrieved from <https://doi.org/10.48550/arXiv.2003.10555>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
- Hur, J., & Yang, J. (2024). South Korean newspaper coverage of Yemeni refugees: Analysis of topics and sentiments using machine learning techniques. *Asian Journal of Communication*, 34(1), 57-72.
- Jo, W., Lee, J., Park, J., & Kim, Y. (2020). Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: Structural topic model and network analysis. *Journal of Medical Internet Research*, 22(6), e19455.
- Mimno, D., & McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI 2008)* (pp. 411-418). Corvallis, OR: AUAI Press.
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 262-272). Stroudsburg, PA: Association for Computational Linguistics.
- Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1), 32-38.
- Oh, J., & Kim, J. (2025). YouTube as a social listening tool: Mining housing discourse on YouTube. *Cities*, 166, 106279.

- Organization for Economic Co-operation and Development (OECD). (2023). *Pensions at a glance 2023: OECD and G20 indicators*. Paris, France: OECD.
- Park, E. (2016). Naver sentiment movie corpus (NSMC) [Data set]. Retrieved from <https://github.com/e9t/nsmc>
- 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. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2), 1-40.
- 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.
- Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication and Society*, *3*(2-3), 297-316.
- Shiller, R. J. (2015). *Irrational exuberance* (3rd ed.). Princeton, NJ: Princeton University Press.
- Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events*. Princeton, NJ: Princeton University Press.
- Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification. In *China National Conference on Chinese Computational Linguistics* (pp. 194-206). Cham, Switzerland: Springer.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38-45). Stroudsburg, PA: Association for Computational Linguistics.

(논문 접수일: 2026.03.24. 수정논문 접수일: 2026.05.25. 논문 채택일: 2026.06.05.)

Analyzing Housing Pension Policy Discourse Using Structural Topic Models: Topic Dynamics and Media-Type Framing Differences in News Coverage, 2007–2026

Seung-hyung Lee *

Abstract

This study analyzes the long-term trajectory and differentiation of Korean housing pension (reverse mortgage) policy discourse using Structural Topic Modeling (STM). From 12,799 housing pension-related news articles collected from BIGKinds (July 2007–March 2026), a 12-topic solution was adopted after SimHash near-duplicate removal. The discourse centered on policy and household debt (13.4%), asset management and retirement planning (11.9%), and enrollment statistics (10.1%). Period-specific (time-smoothed) comparisons showed enrollment statistics predominating under the Lee Myung-bak administration (18.0%), institutional design under Moon Jae-in (13.2%), and elderly-household and bequest-intention surveys under Lee Jae-myung (10.5%). Media-type differentiation was pronounced: regional dailies emphasized enrollment figures, national dailies intergenerational narratives, and the business press retirement planning. Following the March 2026 guarantee-fee reduction, the elderly-household and bequest-intention topic surged by +25.65 percentage points, though this is provisional given a post-event sample of 31 articles. The KNU sentiment lexicon and KoELECTRA agreed on only 49.0% of classifications, confirming a domain gap in policy-news sentiment measurement. Government-period estimates represent descriptive statistics conditional on time-smoothing rather than causal effects. This study contributes a replicable STM-based framework for policy-discourse monitoring in housing finance.

Keywords : Housing Pension, Structural Topic Model (STM), Policy Discourse, Media Framing, Text Mining

* Seung-hyung Lee, Corresponding author, Ph.D., Department of Education, Dong-A University, homolodus4780@gmail.com

© Copyright 2026 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

부록

각 토픽의 내용적 타당성을 기사 수준에서 확인하기 위해, findThoughts() 함수로 추출한 토픽별 비중(θ) 상위 3건의 기사 제목을 <부록 표 1>에 제시한다.

<부록 표 1> 12개 토픽별 우세 문서 상위 3건의 기사 제목

토픽	명명	우세 문서 기사 제목 (매체명, 발행일)
T10	정책·가계부채	가계부채 종합대책 발표...주택연금 활성화 포함(경향, 2017.10) / 주택연금으로 노후 해결?...가계부채 리스크 경고(매경, 2018.03) / 정부, 주택시장 안정화 방안에 주택연금 확대(한경, 2019.12)
T7	자산관리·은퇴설계	100세 시대 은퇴설계...주택연금이 답이 될까(조선, 2019.05) / 퇴직 후 자산관리 전략, 주택연금·개인연금 비교(한경, 2020.08) / 은퇴 후 월 200만원 만들기...주택연금 활용법(매경, 2021.03)
T4	가입실적·통계	주택연금 가입자 10만 명 돌파(연합, 2021.02) / 지난해 주택연금 신규가입 1만 건(뉴시스, 2020.01) / 주택연금 월평균 수령액 127만 원(뉴스1, 2025.03)
T9	제도 설계·가입기준	주택연금 가입연령 55세로 하향(동아, 2020.04) / 주택가격 상한 12억 확대(서울경제, 2023.10) / 우대형 주택연금 도입(한겨레, 2014.07)
T5	고령가구·상속의향	노인 10명 중 4명 빈곤...주택연금 왜 안하나(KBS, 2024.12) / 상속 포기 못해 주택연금 망설이는 고령층(중앙, 2023.05) / 초고령사회 진입...주택연금 재조명(한국, 2025.01)
T6	HF 기관·업무협약	주금공, 광주시와 주택연금 활성화 업무협약(광주, 2019.06) / 주금공 사장, 지역 순회 주택연금 설명회(경남, 2020.09) / 주금공-신한은행 MOU(파이낸셜, 2022.11)
T1	노후·세대·자식연금	부모 집 물려받을까 vs 주택연금...세대 갈등(조선, 2022.04) / 자식연금 대신 주택연금...노후 자립(중앙, 2023.08) / 아파트 한 채가 유일한 노후자금(한겨레, 2021.06)
T2	제도 운영·지급방식	주택연금 종신 vs 확정기간...어느 방식이 유리(매경, 2020.02) / 역모기지 상환 방식 변경(서울경제, 2018.07) / 주택연금 보험료 구조 개편안(머니투데이, 2025.11)
T11	금리·수령액	기준금리 인상에 주택연금 수령액 줄어드나(한경, 2022.07) / 보금자리론 금리 인상...월지급금 영향(이데일리, 2023.01) / 2026년 수령액 3.13% 인상(연합, 2026.03)
T8	집값변동·해지·지역격차	집값 하락에 주택연금 해지 급증(SBS, 2023.03) / 수도권-지방 수령액 격차 확대(경향, 2024.05) / 국감서 주택연금 해지율 지적(뉴시스, 2023.10)
T3	실버타운·노인주거	실버타운 입주 시 주택연금 활용 가능할까(중앙, 2019.11) / 농촌 빈집 활용 노인주거...주택연금 연계(농어민, 2021.08) / 도시재생 사업지 주택연금 상담 급증(부산, 2020.03)
T12	명절 안내·CEO 인사	추석 연휴 주택연금 상담·신청 안내(뉴스1, 2023.09) / 주금공 신입 사장 취임...활성화 의지(연합, 2022.03) / 설 연휴 금융상품 만기 점검(파이낸셜, 2024.02)

주: findThoughts() 함수로 추출한 각 토픽 비중(θ) 상위 3건. 기사 제목은 STM findThoughts() 출력에서 토픽 비중(θ) 상위 3건을 선별하였다.

〈부록 표 1〉에서 확인되듯이, 각 토픽의 우세 문서 기사 제목은 〈표 3〉의 FREX·Prob 단어에 기반한 토픽 명명과 일관된다. T10(정책·가계부채)의 우세 문서는 가계부채 종합대책과 주택연금 활성화를 연계하는 정책 보도이며, T5(고령가구·상속의향)의 우세 문서는 노인빈곤과 상속 갈등을 다루는 사회구조적 보도이다. T12(명절 안내·CEO 인사)의 우세 문서는 연휴 안내와 기관장 인사 보도로서, 주택연금 고유의 정책 이슈가 아닌 기관 운영 맥락의 보도임을 확인해 준다.