

AI 시대의 역사연구와 ‘디지털 문헌학’ — 도구와 질문은 어떻게 만나는가*

Historical Research in the AI Era and “Digital Philology”: How Tools and Questions Meet

허 수(Soo, Hur)**

 0009-0000-2056-7133

목차

1. 도구인가, 질문인가
2. 수작업이 낳은 통찰 – 1 기(2005-2013)
 - 2.1. 중역적 소개와 ‘소개’의 경제’
 - 2.2. 표상공간 분석과 ‘불안정한 균형’
3. 알고리즘이 보여준 구조 – 2 기(2014-2024)
4. AI 로 가능해진 비교 – 3 기(2025-)
 - 4.1. LLM 과 지식 그래프의 결합
 - 4.2. 세 국면의 연결 구조
5. 도구와 질문의 만남
 - 5.1. 두 사이클의 만남
 - 5.2. ‘디지털 문헌학’의 개념과 가능성
6. 부대깎, 그리고 각자의 디지털 문헌학

초록

이 글은 필자가 진행해 온 《개벽》(1920~1926) 연구의 세 국면 — 수작업 분석(1 기, 2005~2013), 토픽모델링과 네트워크 분석(2 기, 2014~2024), LLM(대규모 언어 모델, Large Language Model)과 지식 그래프의 결합(3 기, 2025~) — 을 통해, 연구 질문과 분석 도구가 상호작용하며 발전해 온 과정을 복기한다. 1 기의 정성적 분석이 남긴 두 과제 — ‘사회주의화’의 정량적 검증과 ‘중역적 수용’의 문장 단위 심화 — 가 각각 2 기와 3 기로 이어지는 구조를 밝히고, 이를 ‘디지털 문헌학’이라는 잠정적 개념으로 제안한다. 디지털 문헌학이란, 텍스트의 계통과 전승 과정을 추적하는 전통적

*본 논문은 2026 년 3 월 13 일 독립기념관 한국독립운동사연구소와 서울대학교 국사학과 BK21 교육연구단이 공동 주최한 학술회의("디지털역사학과 독립운동사 연구의 새로운 시도", 서울대학교 14 동 B101 호)에서 기조강연으로 발표한 〈AI 시대의 역사 연구와 ‘디지털문헌학’〉(발표자료집 수록)을 학술지 투고 논문의 형식에 맞추어 일부 수정·보완한 것이다.

**서울대 국사학부 교수

문헌학의 방법을 디지털 기술과 AI로 확장·심화하는 방법론적 지향을 가리킨다. 이 개념이 독립운동사 연구를 비롯한 동아시아 텍스트 비교 연구에 적용될 가능성을 함께 논의한다.

주제어: 《개벽》, 중역적 수용, 토픽모델링, 지식 그래프, 디지털 문헌학, AI와 역사연구

Abstract

This article retraces three phases of the author's long-term research on *Gaebyeok* (開闢, 1920–1926): manual analysis (Phase 1, 2005–2013), topic modeling and network analysis (Phase 2, 2014–2024), and the integration of Large Language Models (LLMs) and knowledge graphs (Phase 3, 2025–). Through this retrospective, it examines how research questions and analytical methods have co-evolved in mutual interaction. Two tasks left unresolved by the qualitative analysis of Phase 1 — the quantitative verification of "socialization" (*sahoejuuihwa*) and the sentence-level investigation of "reception via relay translation" (*jungyeokjeok suyong*) — were subsequently taken up in Phases 2 and 3, respectively. The paper proposes to frame this trajectory under the tentative concept of "digital philology," understood as a methodological orientation that extends and deepens the traditional philological task of tracing textual genealogies and transmission through digital technologies and AI. Finally, it discusses the potential applicability of this framework to research on the Korean independence movement and, more broadly, to comparative studies of East Asian texts.

Keywords: *Gaebyeok*, reception via relay translation, topic modeling, knowledge graph, digital philology, AI and historical research

1. 도구인가, 질문인가

AI시대에 역사연구를 어떻게 할 것인가. 중요하지만 해결책이 쉽지 않아 불안해지는 질문이다. 이와 관련하여 먼저 도구(AI)를 먼저 배워야 한다는 말이 있다. 그러나 도구를 우선하면, 도구에 맞는 질문만 하게 될 위험이 있다. 반대로, 좋은 질문이 있으면 도구는 따라온다는 말도 있다. 그러나 질문만 있고 도구가 없으면, 좋은 질문도 머릿속의 가설에 머물 뿐 검증할 길이 없다. 도구가 앞서야 하는지 질문이 앞서야 하는지, 잘 모르겠다.

이 불안감을 가중시키는 더 현실적인 문제가 있다. 기술을 배우자니, 몇 달 걸려 익힌 도구가 몇 주 만에 바뀐다. 이런 일은 쉴 새 없이 일어난다. 그리고 기술을 익히는 시간만큼, 사료를 읽는 시간이 줄어든다. 안 배우자니, 옆에서는 AI로 하루 만에 끝내는 작업을 나는 몇 달째 붙들고 있다. 내 노동의 가치가 흔들린다. 어느 쪽을 택해도 무언가를 잃는다.

도구를 먼저 세우든 질문을 먼저 세우든, 막다른 곳에 이른다. 논리만으로는 답이 나오지 않는 문제이다. 이런 문제에는 실제로 겪어 본 경험에서 답을 찾는 것이 좋을 듯하다. 마침 본인에게 그런 경험이 있다. 약 20년간 하나의 잡지를 연구하면서 도구와 질문 사이를 오간 경험이다. 거기서부터 이야기를 시작하고자 한다.

그 잡지가 《개벽》이다. 이 잡지는 1920년 6월 25일, 서울 경운동의 한 인쇄소에서 세상에 나왔다. 표지에는 ‘개벽(開闢)’이라는 두 글자가 크게 찍혀 있었다. 하늘과 땅이 열린다는 뜻, 새로운 세계가 열린다는 선언이었다. 3·1운동 직후 식민지 조선의 젊은 지식인들이 만든 이 잡지는, 이후 6년간 72호가 발행되면서 정치·경제·사회·문화·문학·과학

에 이르기까지 당대 거의 모든 지적 관심사를 담아내는 사상적 광장이 되었다.

필자는 이 잡지를 세 차례에 걸쳐 분석하였다. 처음에는 기사를 한 편 한 편 읽고 주제를 분류하는 수작업이었다(1기, 2005-2013). 다음에는 토픽모델링¹이라는 알고리즘의 힘을 빌렸다(2기, 2014-2024). 그리고 지금은 LLM과 지식 그래프²를 결합하고 있다(3기, 2025-). 분석 도구는 계속 바뀌어 왔다.³

그런데 세 국면이 따로 노는 것은 아니다. 1기에서 남긴 두 가지 숙제가 2기와 3기의 씨앗이 되었다. 《개벽》 논조(論調)의 ‘사회주의화’를 제대로 검증하지 못한 과제가 2기로 이어졌고, ‘중역적(重譯的) 수용’ — 원전이 아닌 일본어 번역을 거쳐 다시 한국어로 옮기는 이중 번역 방식 — 을 더 깊이 파고들지 못한 문제의식이 3기로 연결되었다. 도구도 바뀌었고, 질문도 바뀌었다. 그 둘 사이에서 무엇이 일어났는지를 함께 살펴보겠다. 이 경험이 독자들의 연구 현장에서도 하나의 참조가 될 수 있기를 바란다.

2. 수작업이 낳은 통찰 - 1기(2005-2013)

2008년 당시 본인의 책상 위에는 무엇이 있었을까? 수천 페이지의 《개벽》 영인본(影印本)⁴, 그리고 수천 행짜리 엑셀 시트가 있었다. 토픽모델링이니 시계열 분석⁵이니 하는 알고리즘은 아직 역사학 연구실에 들어오지 않았던 시절이다. 매일 연구실에 앉아 영인본을 한 장씩 넘기며 기사를 읽고, 엑셀에 한 줄씩 기록했다. 분류 작업을 하다가 이전에 내린 판단과 나중에 내린 판단이 달라지는 경우도 있었다. 같은 유형의 기사인데 앞에서는 ‘종교’로, 뒤에서는 ‘철학’으로 분류한 경우가 발견되기도 했다. “이 기사는 ‘개조론(改造論)’⁶인가, ‘사회주의’인가?” — 두 범주의 경계에 있는 기사들 앞에서 매번 고민해야 했다. 그 고민 자체가 연구자의 지적 노동이며, 사유의 과정이었다. 이런 작업을 수천 페이지에 걸쳐 반복하다 보면, 잡지의 성격이 시간에 따라 달라지고 있다는 감은 든다. 하지만 수작업으로는 정확히 어디서 어떻게 달라졌는지를 보여줄 수 없었다.

2.1. 중역적 소개와 ‘소개의 경제’

1기(2005-2013)를 돌아보면 흥미로운 역설이 있다. 이 시기에는 워드프로세서와 엑셀이 전부였다. 분석 도구라 할 만한 것이 거의 없었다. 그런데 가장 근본적인 통찰이 바로 이 시기에 나왔다. 그러한 통찰이 도구 없이 어떻게 가능했을까?

도구 대신 작동한 것은 사론적(史論的)⁷ 문제의식이었다. 한국 근대사상사 연구에서는 ‘내재적 발전’의 관점⁸이 지배적이어서, 서구 사상이 한반도에 어떤 경로로 전달되었는가를 하는 과정 자체에는 상대적으로 관심이 적었다. 그런데 《개벽》이라는 사료를 실제로 들여다보면, 서구 사상이 직접 수용된 것이 아니라 일본을 경유하여 중역적(重譯的)으로

소개된 양상이 두드러졌다. 이 전달 경로에 주목하는 새로운 질문이, 도구 없이도 발견을 가능하게 한 동력이었다. 1기에 발표한 두 편의 논문이 그 증거이다.

《개벽》 창간호를 펼치면 흥미로운 글이 하나 있다. “역만능주의(力萬能主義)의 급선봉 푸리드리히 니체 선생을 소개함.”이다. 니체는 19세기 말 유럽 사상계를 뒤흔든 독일 철학자이다. ‘신은 죽었다’, ‘초인(超人)’ — 이런 강렬한 선언으로 유명한 인물이다. 1920년, 3·1운동 직후 식민지 조선의 젊은 지식인들이 이 니체를 적극적으로 소개하고 있었다.

여기서 질문이 생긴다. 1920년대 조선의 지식인들이 독일어 원서를 직접 읽고 니체를 소개했을까? 당시 조선에서 독일어를 읽을 수 있는 사람이 얼마나 있었을까? 사실 거의 없었다. 그렇다면 이 글은 어디에서 온 것일까?

추적해 보니, 《개벽》의 니체·루소·제임스 소개 글은 일본에서 1915년에 출간된 《근대사상16강(近代思想十六講)》이라는 대중 교양서에서 왔다. 엘렌 케이·러셀·카펜터의 소개 글은 1920년에 출간된 《사회개조의 8대사상가》에서 왔다.⁹ 이것이 바로 ‘중역적 소개(重譯的 紹介)’이다.

[서구 원전] → [일본의 요약 단행본] → [《개벽》의 소개 기사]

서구 근대사상에 대한 직접적 접촉이 제약된 식민지 상황에서, 조선 지식인들은 일본이라는 ‘창구’를 통해 세계를 만났던 것이다.

그런데 ‘중역’이라는 말에서 수동적 수용을 떠올린다면, 그것은 절반만 맞는 이해이다. 《개벽》 초기 지면에서 서구 사상 소개를 주도한 인물은 김기전(金起田)이었다. 그는 1920년 6월 창간호에 니체를, 11월에 루소를, 12월에 제임스를 소개했다. 불과 1년 사이에 서구 근대사상의 핵심 인물들을 연이어 조선 독자에게 소개한 것이다. 수동적 수용으로는 이 속도와 체계성을 설명할 수 없다. 그렇다면 김기전은 저본의 내용을 어떤 원칙으로 취사선택한 것일까?

분석 과정에서 일정한 패턴이 발견되었다. 김기전은 저본의 내용을 그대로 옮긴 것이 아니라, 사상이별로 주제를 전략적으로 배분하고 있었다. 일본 저본에는 니체와 루소 모두에 대해 ‘인간관’과 ‘도덕관’이 서술되어 있었다. 그런데 김기전은 니체를 소개할 때 인간관과 도덕관을 중점적으로 다루고, 루소를 소개할 때는 그 부분을 의도적으로 생략했다. 이미 니체 소개에서 인간관·도덕관을 다루었으니, 루소에서는 중복을 피하고 정치·경제론과 교육론에 집중한 것이다. 마찬가지로, 러셀에서 ‘사회개조론’을 집중적으로 다룬 뒤, 카펜터에서는 오히려 저본의 핵심이었던 사회개조론의 비중을 낮추고 예술론을 부각시켰다. 여러 명의 사상가 소개를 하나의 체계 안에서 조율한 것이다. 당시 연구에서는 이 전략을 ‘소개의 경제(紹介의 經濟)’라고 불렀다(표 1 참조).¹⁰

표 1. ‘소개의 경제’ 작동 방식

소개 대상	일본 저본의 주요 내용	김기전이 중점적으로 소개한 부분	김기전이 생략·축소한 부분	전략
니체	인간관, 도덕관, 철학 체계	인간관, 도덕관	—	기초 개념 선점
루소	인간관, 도덕관, 정치·경제론	정치·경제론, 교육론	인간관, 도덕관 (니체에서 기술)	중복 회피
러셀	사회개조론 등	사회개조론	—	핵심 주제 집중
카펜터	사회개조론, 예술론	예술론	사회개조론 (러셀에서 기술)	중복 회피

심지어 니체의 ‘신체가 섬약(纖弱)했다’라고 한 원문 표현을 ‘신체가 강질(強質)이었다’로 바꾸는 경우도 있었다. 니체의 ‘강함’을 전달하는 데 방해가 되는 표현을 의도적으로 수정한 것이다. 이것은 저본에 대한 무비판적 수용이 아니라, 조선의 맥락에 맞게 능동적으로 재구성한 것이다.

김기전은 단순한 번역가가 아니었다. 그는 조선 독자를 위한 ‘사상의 큐레이터’였던 셈이다. 서구 사상의 방대한 체계를, 한 달에 한 번 발행되는 잡지라는 제한된 공간 안에 어떻게 배치할 것인가 — 이 문제를 그는 ‘소개의 경제’라는 전략으로 해결한 것이다.

2.2. 표상공간 분석과 ‘불안정한 균형’

‘소개의 경제’는 개별 기사 수준의 발견이었다. 그런데 개별 기사를 아무리 정밀하게 분석해도, 그것만으로는 72호에 걸친 매체 전체의 성격을 포착할 수 없었다. 두 번째 연구는 잡지 매체 전체를 하나의 분석 단위로 설정하는 데서 출발하였다.

표지는 잡지의 ‘얼굴’이고, 목차는 잡지의 ‘지도’이다. 72호에 걸쳐 이 두 공간의 변화를 추적하면, 잡지가 어떤 방향으로 변해 갔는지가 드러난다. 표지와 목차는 잡지가 자기 자신을 드러내는 공간, 곧 매체와 ‘사회적 상상’이¹¹ 교섭하는 공간이기 때문이다. 이를 ‘표상공간(表象空間)’이라 불렀다.

목차공간을 ‘계몽의 영역’(편집진의 대사회적 발언·주장), ‘소통의 영역’(독자와의 소통), ‘대중의 영역’(문예작품과 평론)이라는 세 기능적 영역으로 구분하고, 시기적으로는 영역 간 구분이 잠재된 ‘형성기’(1~19호), 구분선에 의해 뚜렷이 나뉘는 ‘안정기’(20~40호), 교란이 빈발하는 ‘격동기’(41~72호)로 구분하였다(그림 1 참조).¹²

『개벽』의 목차 공간:

대중의 영역

소통의 영역

계몽의 영역

開闢第六號目次

해는 저를 도다 (卷頭語)

庚申年을 보내며서

文化主義와人格上平等

農村改善에 관한 案

各地靑年團體에 대한現代名士의要求

●나의希冀條件

●實力教育에勉勵

●讀書點과運動熱의增進하도록

●各自의自力養成에勉勵하라

●理想의確立主義의宣傳

●문화發展을促進하라

●精神의誠心그대로

●實踐!實踐!!

●모든것은朝鮮化하라

●底底한人의覺悟

●統制한民衆의組織의必要

●이러한精神을가져서오

李敦化 (二)

白頭山人 (一〇)

金起漣 (一四)

張膺寶 (三五)

趙秉植 (三八)

張道斌 (三九)

任璵章 (四〇)

鄭大健 (三一)

劉秉敏 (三二)

黃義敦 (三三)

韓圭昌 (三四)

鄭道俊 (三五)

崔宗煥 (三六)

金壽觀 (三七)

金鐵 (三八)

李敦化 (二)

吳合祥 (四二)

夜神 (四三)

安昌男 (四五)

妙香山主人 (四九)

一記者 (六一)

曉鍾生 (六九)

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

●靑年團體에 대한要求
●知識과實力
●切實한行動에서完全인團體를기울여야함
人乃天의研究

近世哲學界의革命兒세임쓰先生

오구리飛行場에서

庚申年의거듭들어서

玄堂獨吟

新聞社使喚으로大統領이되기가까지

漢詩

胡適氏를中心으로한中國의文學革命

梨花와新詩를읽고

批評을알고批評을하라

讀者交情欄

深秋과나의同情가는곳

離鄉

海로부터金陵까지

그날밤

夜

安昌男 (四五)

妙香山主人 (四九)

一記者 (六一)

曉鍾生 (六九)

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

【七五】

그림 1. 《개벽》 목차공간의 세 영역
- ‘계몽의 영역’, ‘소통의 영역’, ‘대중의 영역’ (허수, 2008b)

목차 분석은 잡지의 내부 구조를 보여주었다. 그러나 목차는 잡지의 ‘지도’일 뿐, 이때체가 세상을 향해 내건 ‘얼굴’은 아니다. 표지에는 목차가 담지 못하는 또 다른 이야기가 있다.

표지 분석에서 특히 인상적이었던 것은 두 장면이다. 하나는 13호, 창간 1주년 기념호의 유명한 표지이다. 한반도 위에서 포효하는 호랑이 — 이 그림은 원래 창간호에 실릴 예정이었다. 하지만 총독부의 검열에 의해 삭제되었고, 1년 뒤에야 비로소 세상에 나올 수 있었다. 한 장의 표지에 식민지 검열과 저항의 역학이 응축되어 있는 것이다.

다른 하나는 31호(1923년 1월) 표지이다. 이 표지에 돌연 에스페란토 ‘LA KREADO(창조)’가 등장한다. 에스페란토라는 세계 공통어와 동북아 지도의 조합 — 이 표지는 《개벽》이 민족주의와 인류주의의 조화를 천명하기 시작한 국면을 보여주는 상징적 장면이다.



그림 2. 《개벽》 표지 — 13 호(1921년 7월)와 31 호(1923년 1월) (허수, 2008b)

목차와 표지를 종합한 결과, 《개벽》이 발행된 6년의 기간을 세 시기로 구분할 수 있었다. 제3기(48~72호)에는 사회주의의 영향이 커지는 가운데서도 이돈화 등 창간주체들이 주도적 지위를 유지하면서 사회주의 논설과 ‘공존’하는 양상이 두드러졌다. 이 연구의 핵심 논지는, 그럼에도 주도층은 ‘천도교(天道敎)에 기초한 사회개조’의 입장을¹³ 견지하였다는 것이다. 논조는 ‘불안정한 균형’ — 사회주의 쪽으로 기울어질 수도 있었으나 완전히 기울지는 않은 긴장 상태 — 에 있었다.¹⁴ 이는 《개벽》 후기의 논조가 근본적으로 바뀌었는지, 아니면 기본 방향이 유지되었는지를 둘러싼 ‘단절’과 ‘연속’의 쟁점에¹⁵ 대한 ‘제3의 입장’이었다.¹⁶

3. 알고리즘이 보여준 구조 - 2기(2014-2024)

그러나 이 결론은 표지와 목차라는 ‘형태적’ 분석에 의거한 것이었다. 외양에서 도출한 논지가 본문이라는 내용에서도 관철되는 것일까? 이 검증이 과제로 남았다.

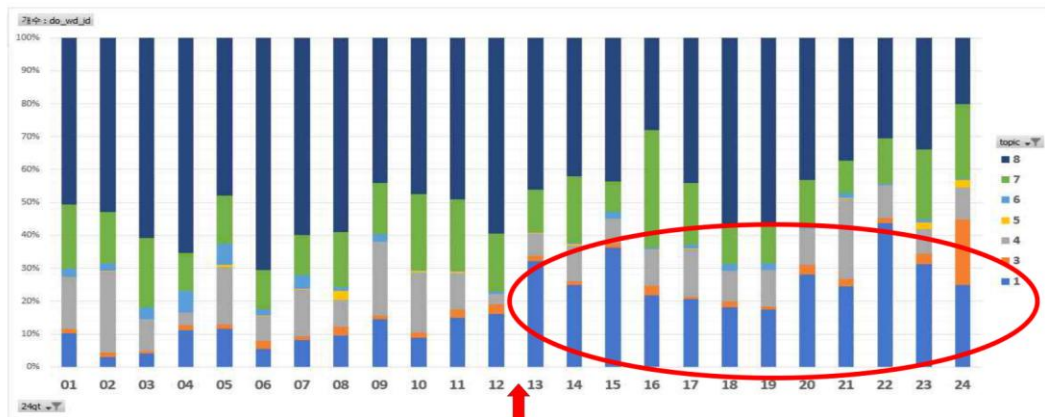
허수(2021)는 1기에서 남겨진 바로 그 과제, 곧 《개벽》 논조의 ‘사회주의화’가 주도층의 사상적 변화까지 동반한 것이었는가라는 문제를, 디지털인문학의 방법론을 동원하여 본격적으로 다루었다.

《개벽》 주요 논설 334편의 전산 자료를 수집, 전처리(前處理)하여¹⁷ 코퍼스(corpus)를¹⁸ 구축한 뒤, 토픽모델링으로 7개 토픽(사회주의, 조선정치, 경제·교육, 동학농민전쟁,

성리학, 국제정세, 개조론)을 추출하였다.¹⁹ 토픽모델링이란, 컴퓨터가 수백 편의 글에서 자주 함께 등장하는 단어들을 묶어 ‘주제 묶음’을 찾아내는 방법이다. 무엇을 ‘사회주의화’로 볼 것인가는 그 자체가 난제이다. 토픽모델링은 이 문제에 유용한 해법을 제공하였다. 추출된 7개 토픽 가운데, 자본·계급·마르크스·혁명 등의 단어로 구성된 토픽1(사회주의)을 판별 지표로 삼은 것이다.

그런데 《개벽》의 논조가 “점차 변했다”는 것은 이미 1기에서 알고 있었다. 정확히 어디서, 어떻게 변했는지가 문제였다.

72호를 3호씩 묶어 24개 분기로 나누고, 분기마다 토픽의 비중이 어떻게 변하는지를 추적하였다. 그 결과, 그림 3에서 볼 수 있듯이, 36호와 37호 사이에서 담론의 무게중심이 전환되는 것이 명확하게 드러났다. 마치 지질학에서 지층의 색이 바뀌는 경계를 발견한 것과 같다. 2008년에는 막연히 “점차 변했다”고 기술했지만, 이 분석은 정확히 어디서 변했는지를 보여주었다.



토픽1-사회주의, 토픽3-조선정치, 토픽4-경제·교육, 토픽5-동학농민전쟁, 토픽6-성리학, 토픽7-국제관계, 토픽8-개조론

그림 3. 《개벽》 분기별 토픽 비중 추이

필자는 여기서 ‘논조’를 ‘주제들의 합성(合成, vector)’으로 규정하였다. 여러 색을 섞어 하나의 색을 만드는 것처럼, 다수의 논설에 걸쳐 산포한 주제들의 구성을 하나의 유기적 그림으로 파악하는 것이다. 이 그림을 구체화하는 방법으로 7개 토픽 간의 관계를 나타내는 ‘토픽 연결망 지도’를 산출하였다. 각 토픽이 하나의 역(驛)이라면, 허브 토픽은 담론의 ‘중앙역’이다. 전기에는 토픽8(개조론)이 중앙역이었다. 다른 모든 토픽이 개조론을 통해 연결되는 구조였다. 그런데 후기에는 토픽1(사회주의)이 그 중앙역의 자리를 차지하였다. 이는 논조 변화가 단순한 주제의 교체가 아니라, 여러 주제들 간의 접속과 단절을 동반하는 구조적 변동이었음을 시사한다(그림 4 참조).

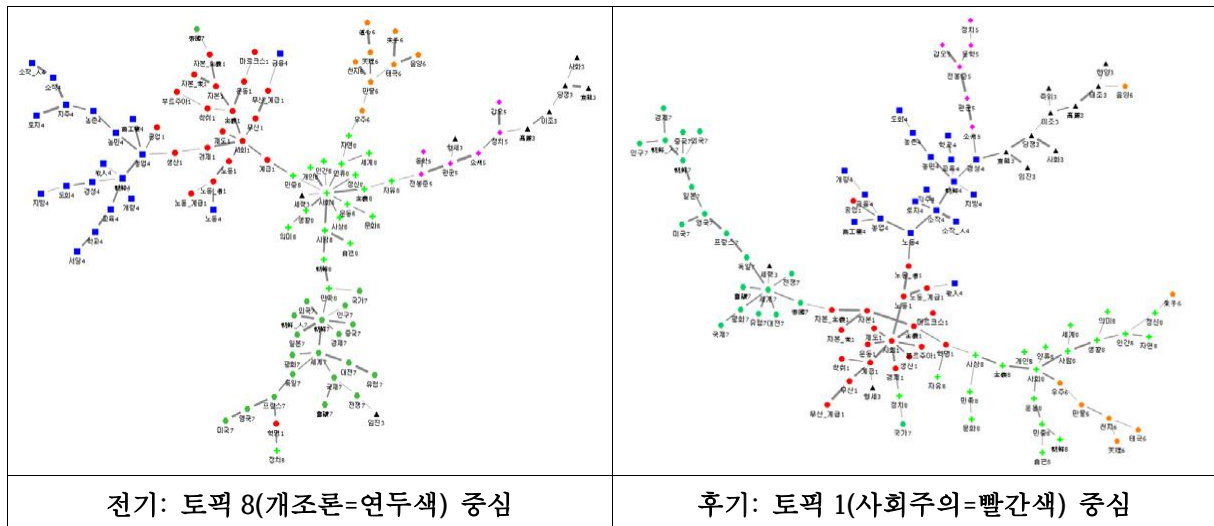


그림 4. 《개벽》 전·후기의 토픽 연결망 비교 (허수, 2021 의 그림 4)

논조의 중심이 사회주의로 이동했다. 그렇다면 2절에서 확인한 ‘불안정한 균형’은 틀린 것이었을까? 주도층까지 사회주의화된 것이라면, 그것은 ‘균형’이 아니라 ‘전면 전환’에 가까울 것이다.

이 질문에 답하기 위해, 후기의 기명(記名, 필자의 이름이 밝혀진) 논설 121개를 ‘개벽 주도층’(이돈화·김기전·박달성·차상찬, 29편)과 ‘일반 필자층’(92편)으로 나누어 비교하였다.²⁰ ‘일반 필자층’이라는 용어는, 기명 논설 집필자 중 ‘개벽 주도층’이 아닌 사람들을 편의상 한 데 묶어서 부른 것이다.

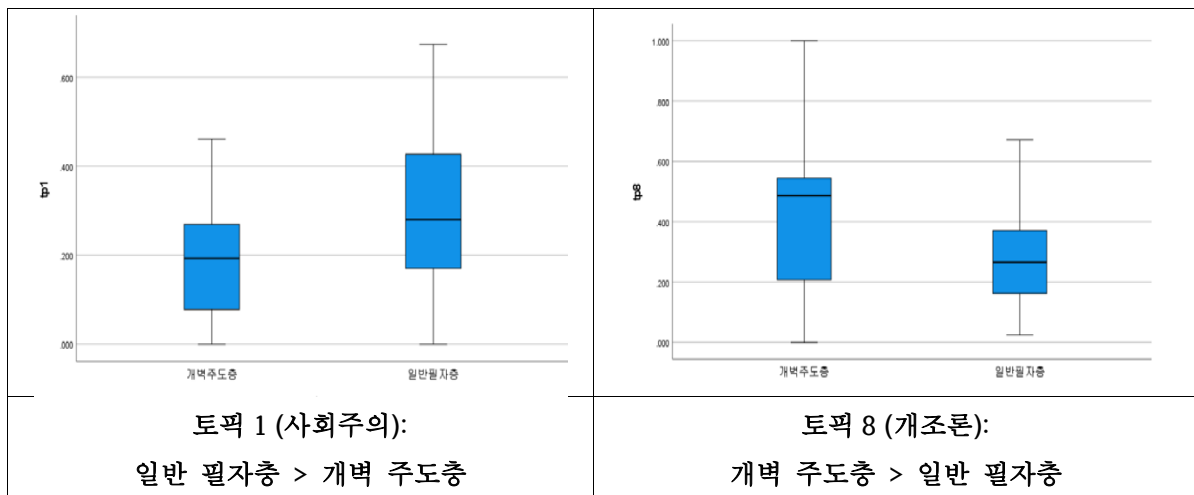


그림 5. T 검정으로 구한 두 집단의 평균 비교 (허수, 2021 의 그림 6)

왼쪽 상자그림(box plot)에서 일반 필자층의 상자가 개벽 주도층보다 높는데, 이는 일반 필자들이 사회주의 주제를 더 많이 다루었다는 뜻이다. 오른쪽 그래프에서는 반대로 개벽 주도층의 상자가 더 높아 주도층이 개조론을 더 많이 다루었음을 보여준다. p값이 0.05 미만이면 이 차이가 통계적으로 의미 있는, 즉 우연이 아닌 실제 차이임을 나타낸

다.

물론 사회주의 관련 글에 지면을 할애한 것 자체는 개벽 주도층이 중심이 된 편집진의 판단이기도 하다. 그러나 이는 당대의 사상적 조류를 전달하는 매체로서의 공적 역할, 곧 ‘사회적 공기(公器)’로서의 소명에 따른 것이지, 편집진 스스로 사회주의를 자기화한 결과는 아니었다. 주도층은 여전히 개조론의 담지자였고, 사회주의의 담지자는 아니었던 것이다. 이 결과는 허수(2008b)에서 정성적으로 발견한 ‘제3의 입장’을 정량적으로 재확인한 것이다. 또한 ‘사회주의’, ‘논조 변화’, ‘영향력’ 등 기존에 주관적으로 판단되던 개념들을 객관적 지표로 포착하고 검증할 수 있게 된 점에서 방법론적 의의가 있었다.

4. AI로 가능해진 비교 - 3기(2025-)

4.1. LLM과 지식 그래프의 결합

2장에서 발견한 사회주의와 개조론 사이의 ‘불안정한 균형’은 3장에서 정량적으로 재확인되었다. 이 문제에 대한 검증은 여기서 일단락된다. 그러나 2장에서 발견한 ‘소개의 경제’라는 또 다른 과제는 어떤가? 그것은 장(章) 단위의 분석에서 확인한 것이다. 보다 정밀한 문장(Sentence) 단위의 비교에서도 ‘소개의 경제’가 관찰되는 것일까? 무엇을 어떻게 해야 할지 머릿속으로는 가늠되었으나, 전처리와 데이터 확보, 대규모 텍스트 처리 등의 문제로 오랫동안 엄두를 내지 못하고 있었다.

AI에 힘입어 이 과제가 비로소 가능해졌다. 허수(2025)는 1기 연구(2008a)의 문제의식, 곧 ‘중역적 수용’의 구체적 양상을 LLM과 지식 그래프의 결합을 통해 문장 단위에서 시도한 것이다.

분석 대상은 이노우에 데쓰지로(井上哲次郎)의 《철학과 종교》(1915) 제8장 「메치니코프의 학설에 관하여」(원문 689문장)와, 이돈화가 《개벽》 8호(1921년 2월)에 게재한 「科學上으로 본 生老病死」(번역문 175문장)이다.

분석은 세 단계로 진행되었다(표 2 참조). 1단계(데이터 구조화)에서는 분석 대상 텍스트를 문장 단위로 분리한 다음, 각 문장에 고유 ID를 부여하고 양자의 관계를 ‘1:1 번역’, ‘N:1 요약’, ‘삭제’의 세 유형으로 매핑하였다.(그림 6 참조)

표 2. 3기 분석의 3 단계

단계	작업 내용	핵심 도구
1 단계:데이터 구조화	원문 689 문장·번역문 175 문장에 ID 부여	LLM (구글 제미나이) ²¹
2 단계:텍스트 마이닝	번역된 문장 vs 삭제된 문장 간 어휘 차이 추출	로그우도비(LLR) 분석

3 단계:지식 그래프 구축	문장 간 관계를 그물망 형태로 시각화	Neo4j
-------------------	----------------------	-------

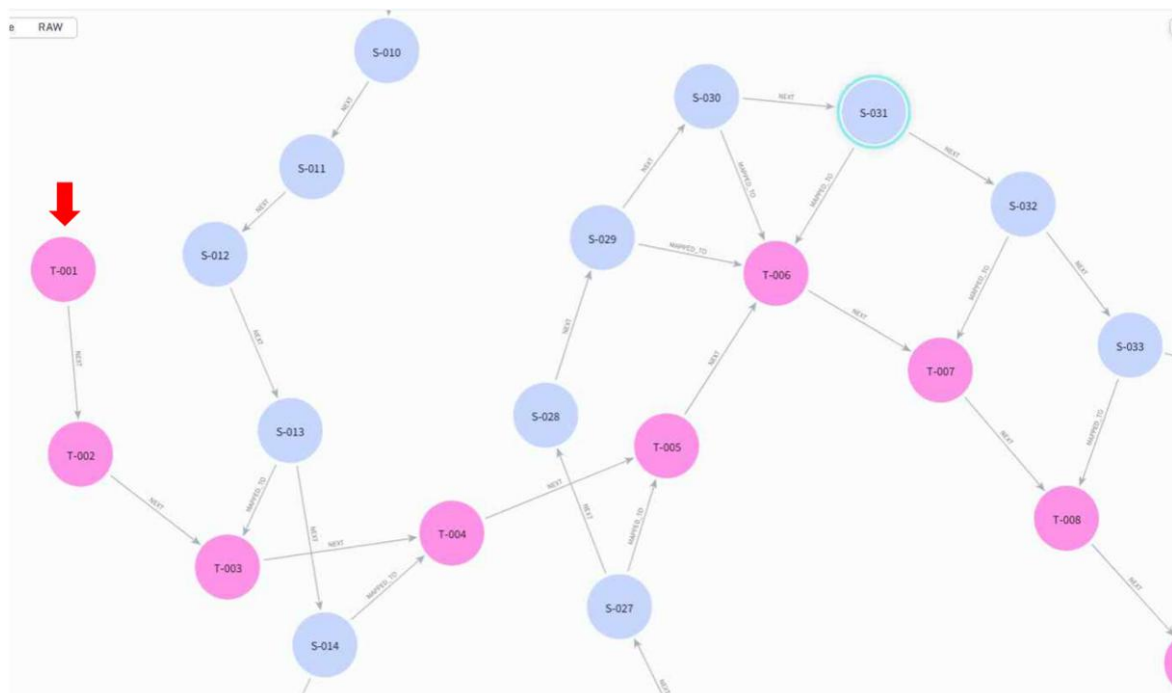


그림 6. 《철학과 종교》(파랑노드)와 《개벽》(분홍노드) 간의 문장 대응 관계

2단계(텍스트 마이닝)에서는 로그우도비(LLR)²² 분석을 통해 ‘번역된 그룹’과 ‘삭제된 그룹’ 사이의 어휘 차이를 추출한 결과, 실용적이고 자연과학적인 내용은 번역에 포함되고, 철학적 사변이나 서양 지식인의 권위를 내세우는 내용은 생략되는 패턴을 발견하였다. 3단계(지식 그래프 구축)에서는 Neo4j라는²³ 툴(tool)에 데이터를 입력하여, 원문 689문장과 번역문 175문장 사이의 대응 양상을 살펴보았다. 이 그래프를 통해 선택과 배제, 발췌와 요약 등의 양상이 드러났다.

그런데 원문에는 있으나 번역에서 빠진 문장들이 정말 의도적으로 삭제된 것일까? 단순한 누락일까?

단순한 누락이 아니었다. 그림 6의 지식 그래프에서 파란색 노드는 원문, 분홍색 노드는 번역문을 나타낸다. 전체 상황을 관찰하면, 과학 관련 노드(초록색)는 대부분 번역문과 연결된 반면, 철학(빨간색)과 권위(노란색) 관련 노드는 번역문과 연결되지 않았다(표 3, 그림 6-1, 그림 6-2 참조). 이돈화의 선택은 임의적인 것이 아니라, 매우 의도적인 편집 행위였다. 그는 이노우에 데쓰지로의 ‘형이상학적 권위와 사변’은 철저히 배제하고, 조선 사람에게 필요한 ‘실용적인 자연과학’만을 선택적으로 수용하여 전달한 것이다.

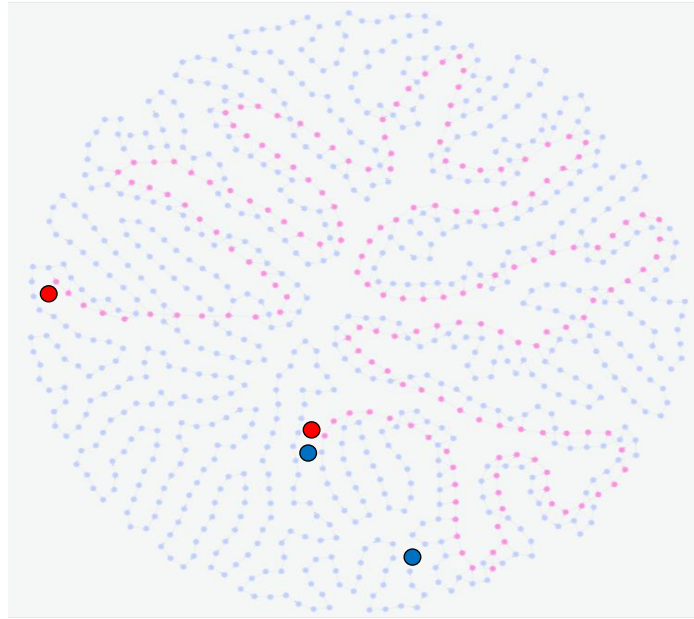


그림 6-1. Neo4j 지식 그래프: 전체 구조 비교²⁴
 — Source(파랑)와 Target(분홍)의 연결 (허수, 2026)

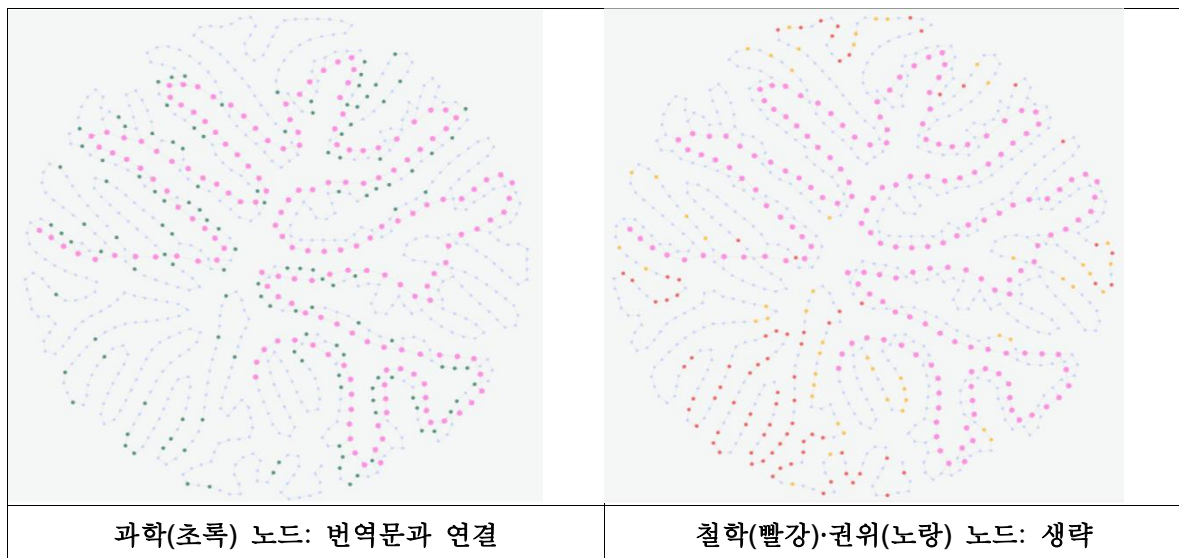


그림 6-2. Neo4j 지식 그래프: 주제별 연결 양상²⁵
 - 과학(초록)은 번역문과 연결 vs 철학(빨강)·권위(노랑)는 생략 (허수, 2026)

표 3. 이돈화의 선택적 수용 패턴

수용(번역)	배제(생략)
실용적·자연과학적 내용 (초록색 노드)	철학적 사변(빨간색 노드)
	서양 지식인의 권위(노란색 노드)

이돈화의 선택적 수용이 확인되었다. 이것은 2장에서 김기전에게서 발견한 ‘소개의 경

제’와 같은 방향을 가리킨다. 관련 인물도 다르고 분석 단위도 한쪽은 장(章)인데 다른 한쪽은 문장(Sentence)으로 서로 다르지만, 원전의 내용을 그대로 옮기지 않고 자기 기준으로 취사선택한다는 점에서 두 사례는 공통된다.

방법론적으로는, 1기에서 수작업으로 수행했던 저본-기사 간 대조 작업을 LLM과 지식 그래프를 결합하여 시도한 것이다.

4.2. 세 국면의 연결 구조

이상 세 국면은 두 개의 축으로 연결된다. 허수(2008b) → 허수(2021)의 ‘사회주의화’ 축(형태적 분석의 정량적 검증)과, 허수(2008a) → 허수(2025)의 ‘중역적 수용’ 축(장 단위 분석의 문장 단위 심화)이 그것이다.

1기의 정성적 분석이 남긴 두 가지 과제가, 각각 디지털인문학과 AI라는 새로운 도구를 만나면서 2기와 3기로 발전한 것이다. 2기의 알고리즘(토픽모델링 등)과 3기의 AI(LLM)는 대량의 텍스트 분석을 가능하게 한다는 점에서는 대동소이하다. 그러나 2기가 연구자가 설계한 틀 안에서 통계적 패턴을 찾는 것이었다면, 3기는 AI가 언어를 이해하여 수백 문장 간의 번역 관계 판별과 같은 질적 판단을 대규모로 수행한 것이다. 분석 도구가 아날로그에서 디지털인문학으로, 다시 AI로 바뀌는 과정에서 연구 질문과 서로 영향을 주고받으며 발전해 왔다는 사실이 핵심이다(표 4, 그림 7 참조).

표 4. 세 국면의 비교

국면	연구 질문	분석 도구	핵심 개념	연결
1 기 허수(2008a)	‘중역적 수용’의 양상 규명	워드프로세서, 엑셀(수작업 대조)	중역적 소개, 소개의 경제	→ 3 기
1 기 허수(2008b)	《개벽》의 매체적 성격 고찰	워드프로세서, 엑셀(형태 분석)	표상공간, 3 영역, 3 시기	→ 2 기
2 기 허수(2021)	사회주의화의 정량적 분석	네트워크 분석, 토픽모델링, T 검정	토픽 연결망, 영향의 제한성	← 1 기(2008b)
3 기 허수(2025)	문장 단위의 텍스트 비교	LLM, 지식 그래프(Neo4j)	선택적 수용, 디지털 문헌학	← 1 기(2008a)

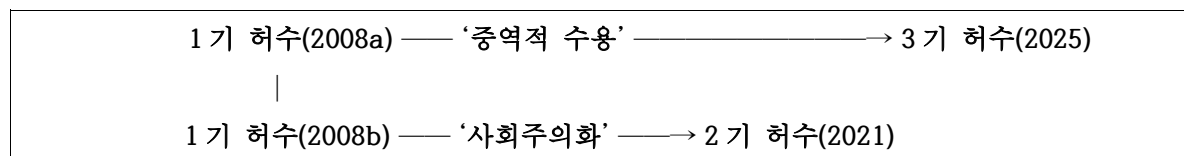


그림 7. 세 국면의 연결 구조

5. 도구와 질문의 만남

5.1. 두 사이클의 만남

세 국면이 두 개의 축으로 연결된다는 것은 확인하였다. 그런데 이 연결은 어떻게 가능했던 것일까? 허수(2021)와 허수(2025)는 어떻게 구상하게 되었을까?

이 물음에 대해서는 두 가지 방향의 설명이 가능하다. 첫째, 도구가 질문을 만들어냈다는 설명이다. ‘망치를 든 사람 눈에는 못만 보인다’ — 토픽모델링이라는 도구를 익힌 뒤, 그것을 적용할 만한 소재를 자신의 연구 이력에서 골라냈다는 추론이다. 그러나 이 설명에는 난점이 있다. 토픽모델링을 연구에 실제로 적용한 것은 허수(2021)가 처음이었지만 《개벽》 논조의 사회주의화에 대한 관심은 훨씬 이전 연구에서 이미 제시되었으므로, 도구를 충분히 습득한 뒤 적용 대상을 찾았다는 논리는 성립하지 않는다.

둘째, 질문이 문제 해결 도구를 선택했다는 설명이다. ‘인간은 자신이 해결할 수 있는 질문만 제기한다’는 관점이다. 그러나 《개벽》 논조에 관한 논문을 2008년에 쓴 후 13년 만에 이 연구가 나왔고, 디지털인문학 방법에 접하게 된 것이 2010년이라는 점을 감안하면, 질문이 곧장 도구를 선택한 것이라고 보기도 어렵다.

앞서 살펴본 2008년의 수작업 경험을 돌이켜보면, 이 점이 더욱 분명해진다. 당시의 엑셀 작업은 도구의 습득도, 질문의 직접적 해결도 아닌, 사료와의 씨름 그 자체였다. 그런데 그 경험이 훗날 알고리즘의 결과를 해석하는 데 결정적인 판단 기준이 되었다. 도구와 질문 어느 한쪽이 다른 한쪽을 끌고 간 것이 아니라, 양자 사이에 연구자의 축적된 경험이 매개로 작용한 것이다.

본인의 경험을 설명하는 보다 적절한 틀은, 도구와 질문이 각각 독자적인 사이클을 갖고 전개되다가 어느 지점에서 만났다는 것이다.

1기에서 2기로의 전환: 2008년 질적 분석 이후 본인은 2010년 개념사(概念史) 프로젝트²⁶ 합류하여, 개념사 연구를 정량적 방식으로 진행하고 이후 연결망 분석으로 진화시키면서 여러 편의 논문을 작성하였다. 그 과정에서 성취와 함께 한계도 느꼈다. 정량적 방법이 제 힘을 발휘하려면, 사료와 연구사적 맥락에 대한 깊은 이해가 전제되어야 한다는 점을 깨달았다. 이 대목에서 허수(2008b) 당시의 질문이 소환되었다. 도구의 사이클, 즉 개념사 프로젝트를 자기화 하는 과정에서 습득한 정량적 연구방법이, 질문의 사이클, 즉 《개벽》 사회주의화에 대한 오랜 물음과 13년 만에 만난 셈이다.

2기에서 3기로의 전환: 허수(2008a)의 문제의식, 곧 ‘중역적 수용’의 구체적 양상을 규명하겠다는 과제는 기술이나 방법보다 훨씬 긴 호흡에 해당하는 사론적(史論的) 차원의 것이다. 4장에서 살펴보았듯이 이 과제는 LLM의 등장으로 비로소 실현 가능해졌고, 마침 동아시아 텍스트 비교를 주제로 한 공동연구반이 발족하면서 3기 연구를 시

도하게 되었다.

이처럼 도구가 질문을 만든 것도, 질문이 도구를 선택한 것도 아니었다. 전술했듯이 질문과 도구가 각각의 경로로 진화해 가다가, 특정한 시점에 만난 것이다. 그렇다면 이 만남에서 기술은 어떤 역할을 하였을까? 기술의 진보는 우리가 과거를 더 높은 해상도로 바라볼 수 있게 해주는 렌즈의 진보이다. 그러나 그 렌즈를 어디에 대고 무엇을 볼 것인가는, 연구자가 축적해 온 문제의식과 해석 능력에 달려 있다. 이 점은 AI시대에도 달라지지 않을 것이다.

5.2. ‘디지털 문헌학’의 개념과 가능성

도구와 질문이 만난다는 것은 납득할 만하다. 그러나 ‘만남’은 지나온 경험을 설명할 수 있을 뿐, 앞으로의 연구를 이끌어갈 방법론적 지향을 제시하지는 못한다. 세 국면의 경험을 거친 지금, 필자가 도달한 AI 시대 역사연구의 방법론적 지향은 무엇일까?

본고에서는 그것을 ‘디지털 문헌학’이라 부르고자 한다. 여기서 ‘문헌학(Philology)’이란, 텍스트의 계통과 전승 과정을 추적하여 그 성립·변용의 양상을 규명하는 학문적 전통을 가리킨다. 동아시아의 고증학(考證學)이나²⁷ 서양의 고전문헌학이 모두 이 범주에 속한다.

세 국면 가운데 특히 1기와 3기에서 수행한 작업의 핵심은, 《개벽》이라는 조선의 텍스트가 일본의 저본(底本)으로부터 어떤 경로와 방식으로 생성·변용되었는가를 추적하는 것이었다. 1기에서는 장(章) 단위의 수작업 대조로 인용율을 산출하여 ‘중역적 소개’와 ‘소개적 경제’라는 거시적 특징을 파악하였고, 3기에서는 문장(Sentence) 단위의 AI 보조 비교를 통해 이돈화의 ‘선택적 수용’ — 형이상학적 사변의 배제와 실용적 과학 지식의 선택 — 이라는 미시적 수준의 텍스트 변용을 데이터로 확인하였다. 이러한 작업은 전통적 문헌학의 방법을 디지털 기술로 확장·심화한 것이라 할 수 있다. 일반적인 디지털 인문학이 텍스트의 패턴과 분포를 통계적으로 포착하는 데 무게를 둔다면, ‘디지털 문헌학’은 저본(底本)과 수용 텍스트 사이의 전승·변용 관계 그 자체를 분석 단위로 삼는다는 점에서 디지털 인문학과 구별된다. ‘디지털 문헌학’의 핵심적 특징을 정리하면 다음과 같다.

가. 규모와 정밀도의 동시 확장: 현미경이 육안으로 보이지 않는 세계를 열어주듯, 디지털 기술(알고리즘·AI)은 수작업으로는 도달하기 어려운 대규모 텍스트의 체계적 비교를 가능하게 한다. 전통적 문헌학이 소수 텍스트의 정밀 대조에 강점을 가졌다면, ‘디지털 문헌학’은 분석의 규모와 정밀도를 동시에 확장한다.

나. 해석과 기술의 융합: 전통적 문헌학의 핵심 역량인 텍스트 비평(textual criticism)의²⁸ 안목과, 디지털인문학(토픽모델링·네트워크 분석 등)과 AI(LLM)의 분석 능력이

결합된다.

다. 구조화·시각화·재현가능성: 텍스트 마이닝과 지식 그래프를 비롯한 디지털 도구의 장점 — 구조화, 시각화, 재현가능성 — 이 문헌학적 작업의 효율성과 엄밀성을 높여 줄 수 있다.

물론 ‘디지털 문헌학’은 아직 확립된 분야가 아니다.²⁹ 연구자 개인의 경험을 방법론적으로 성찰하는 과정에서 도달한 잠정적 개념이다. 다만 본 논문의 4장에서 다룬 것은 한 개의 장(제8장)에 한정된 시론이다. 이와 별도로 이노우에의 『철학과 종교』 전체와 이돈화의 『인내천요의』(1924) 전체를 대상으로 22만여 건의 문단 쌍을 전수 비교한 연구(허수, 2026)가 있다.³⁰ 같은 방법론을 더욱 큰 규모로 적용한 것이다. 20세기 동아시아의 ‘중역적 지식 수용’ 연구에서, 저본과 번역·소개 텍스트 간의 대조 작업은 필수적이면서도 매우 손이 많이 가는 과정이다. AI 기술의 발전이 이 과정을 혁신할 잠재력을 가진다는 점에서, ‘디지털 문헌학’은 구체적인 연구 프로그램으로 발전할 가능성이 있다고 생각한다.

이러한 문제의식은 독립운동사 연구에도 적용할 수 있다. 독립운동과 관련한 동아시아 각국의 텍스트 — 한국·일본·중국·러시아의 문서, 보고서, 언론 기사 등 — 사이의 관계를 규명하는 작업은 본질적으로 ‘문헌학적’ 성격을 띤다. 예컨대, 독립운동가의 저술이나 선언문이 어떤 사상적 원천으로부터 영향을 받아 생성·변용되었는지를 추적하는 작업이나, 동일한 사건에 대한 각국 기록 사이의 서술 차이를 체계적으로 대조하는 작업이 이에 해당한다. 이러한 작업은 사료의 양과 언어적 장벽으로 인해 개별 연구자의 수작업만으로는 한계가 있었으나, AI 기반의 다언어 텍스트 분석과 지식 그래프 기술이 새로운 가능성을 열어줄 수 있을 것이다. 그 구체적인 형태는 각 연구 분야의 사료적 특성과 연구자의 문제의식에 따라 달라질 것이다.

6. 부대낌, 그리고 각자의 디지털 문헌학

이상에서는 세 국면의 경험을 복기하고, 그를 통해 도달한 방법론적 지향을 ‘디지털 문헌학’이라 불러 보았다. 그러나 방향을 정하는 것과 그 길을 실제로 걸어가는 것은 별개의 문제이다. 특히 지금처럼 기술 환경 자체가 급변하는 시기에는 더욱 그러하다.

필자는 얼마 전 AI 기반의 클로드 코드(Claude Code)를³¹ 사용하면서, 가히 ‘AI 혁명의 시대가 왔구나’라고 느낀 순간이 있었다. 일상어로 요청하면 AI가 코드 작성에서 자료 검색, 문서 편집까지 자율적으로 수행하기 때문이다. 그런데 인류가 18세기 중엽에 시작된 산업혁명에 ‘혁명’이라는 이름을 붙이게 된 것은 19세기에 들어와서였다.³² 인식과 명명이 실제 경험보다 지체된 것이다. 이 논문의 제목에 ‘AI 혁명’ 대신 ‘AI 시대’라는 말을 붙인 본인 필자도 마찬가지로 아닐까? 거대한 전환의 한복판에 서 있으면서도 그 실체를 온전히 파악하지 못하는 것이 동시대인의 숙명일 것이다.

AI에 대한 과도한 기대와 막연한 불안 사이에서 실제로 필요한 것은 무엇일까? 많은 전문가들은 AI와 협업 관계를 형성하도록 권하며, 본고도 이러한 입장에 동의한다. 그러나 막상 연구자로서 AI를 어떻게 활용할 것인지를 구체적으로 생각해 보면, 현실적인 방법이 쉽게 보이지 않는다. LLM들 사이에서도 각기 개성과 성능이 다르고, 동일한 LLM에서도 모델에 따라 장단점이 달라, 하나의 정답이 있을 수 없기 때문이다.

지름길은 없지만 출발점은 있다. 자신의 연구에 AI를 직접, 본격적으로 사용해 보는 것이다. 사료를 한 장 한 장 넘기던 손끝의 노동이 이후의 분석을 가능하게 했듯이, AI와의 협업도 추상적 이해가 아닌 구체적 실천 속에서 체득된다. 본고에서는 이를 ‘부대낌’이라 부르하고자 한다. 장수가 말과 호흡을 맞추려면 수없이 함께 달려야 하듯, 연구자도 AI와 부대끼는 과정이 필요하다.

그러나 부대낌만으로 충분하지는 않다. 앞에서 살펴보았듯이, 도구와 질문은 각각의 경로로 나아가다가 어느 시점에 만난다. 연구 질문을 숙성시키는 일과 AI의 가능성을 직접 탐색하는 일이 함께 갈 때, 비로소 그 만남의 순간이 찾아올 것이다.

이 만남은 연구자의 역할에 어떤 변화를 초래할까? 종래의 연구자가 수집한 사료를 직접 읽고 해석하는 데 대부분의 시간을 쏟았다면, AI 시대의 연구자는 연구 전체를 조망하고 기획하는 쪽에 무게중심을 둘 것으로 생각된다. 방향을 설정하고 점점 지점을 제시하되, 구체적인 실행은 AI에게 맡기는 방식이다. 스포츠에서 감독이 직접 경기에 뛰지 않으면서도 선수들의 역량을 이끌어내듯, 연구자도 AI의 가능성을 파악하고 연구의 큰 그림 속에서 AI를 활용하는 일종의 ‘디지털 큐레이터’로 변모할 것이다.³³

그 출발점은 생각보다 가까이에 있을 수 있다. 혹시 지금 사료를 정리하며 “이런 작업이 과연 연구라 할 수 있을까?”라고 의문을 품은 적이 있는가? 적어도 필자의 경험에서, 오늘의 수작업은 내일의 도구와 만날 자산이었다. ‘AI 시대에 역사연구를 어떻게 할 것인가’라는 오늘의 질문은, ‘내 연구에서 AI와 어떻게 부대낄 것인가’로 바뀌 물을 수 있겠다.

‘디지털 문헌학’은 본인이 그 부대낌의 과정에서 도달한 하나의 길일 뿐이다. 독자 여러분이 찾게 될 방향은 각자의 연구 경험과 관심, 인연에 따라 다를 것이고, 그것이 어떤 모습일지를 이 자리에서 특정할 수는 없다. 다만 그 미지의 방향을 잠정적으로 표현한다면, 여러분 각자의 현장에서 자신만의 ‘디지털 문헌학’이 시작되기를 바란다.

¹ 토픽모델링(topic modeling)이란 대량의 텍스트에서 자주 함께 등장하는 단어들을 묶어 ‘주제 묶음(토픽)’을 자동으로 추출하는 통계적 방법이다.

² 지식 그래프(knowledge graph)란 개체들 사이의 관계를 그물망처럼 구조화하여 표현하는 방식이다.

³ 세 국면은 2005 년 박사학위논문 집필 이후의 개인 연구 이력을 편의상 구분한 것이다. 각 시기에 《개벽》 연구만을 수행한 것은 아니며, 본고에서 다루는 것은 그 가운데 《개벽》 관련 연구에 한정된다. 시기 구분에서 가장 뚜렷한 전환점은 디지털인문학(Digital Humanities, 인문학 연구에 컴퓨터 기술과 데이터 분석 방법을 결합하는 학제적 분야) 방법론의 도입이며, 이를 기준으로 그 이전(1 기)과 이후(2 기)를 나누고, 2 기 내에서 다시 AI 의 본격 활용을 기점으로

3 기를 구분하였다. 1 기(2005-2013)는 수작업 분석, 2 기(2014-2024)는 네트워크 분석과 토픽모델링, 3 기(2025-)는 인공지능(LLM)을 각각 주된 분석 도구로 활용한 시기이다. 다만 본문에서 논의하듯이, 각 국면의 연구는 도구만이 아니라 연구 질문과의 상호작용 속에서 이루어졌다.

⁴ 영인본(影印本)이란 원본을 사진으로 촬영하여 그대로 복제한 책이다.

⁵ 시계열 분석이란 시간의 흐름에 따른 변화 추이를 추적하는 분석 방법이다.

⁶ 개조론(改造論)이란 사회를 근본적으로 개혁·개조해야 한다는 당시의 사상적 흐름이다.

⁷ 사론적(史論的)이란 역사학의 본질과 방법을 탐구하는 이론적 논의를 가리킨다.

⁸ ‘내재적 발전’의 관점이란 외부의 영향보다 한국 사회 내부의 자체적 동력에서 근대화의 씨앗을 찾는 시각이다.

⁹ 中澤臨川·生田長江 編, 《近代思想十六講》, 東京: 新潮社, 1915(초판); 生田長江·本間久雄 共著, 《社會改造の八大思想家》, 東京: 東京堂書店, 1920. 각 저본의 서지 및 《개벽》 소개 기사와의 대응 관계에 대해서는 허수(2008a), 53-56 쪽 참조.

¹⁰ 이 개념은 허수(2008a) 3 장 「‘중역적 소개’와 ‘소개의 경제’」에서 제시하였다.(68-72 쪽).

¹¹ ‘사회적 상상’이란 한 사회의 구성원들이 공유하는 시대적 열망과 미래상을 가리킨다.

¹² 《개벽》 목차공간의 세 영역 및 형성기·안정기·격동기 시기 구분은 허수(2008b)의 II 장 2 절 「목차공간의 분할과 변동」에서 상세히 논증하였다.(2008b, 361-367 쪽).

¹³ 천도교(天道敎)는 동학에서 발전한 한국 고유의 민족 종교이다.

¹⁴ 허수, 2008b, 354 쪽.

¹⁵ 《개벽》의 전·후기 논조 변화에 관해서는, 김정인·김건우 등 단절 또는 차이에 주목하는 입장과, 최수일·한기형 등 연속성을 강조하는 입장으로 나눌 수 있다. 김정인, 「《개벽》을 낳은 현실, 《개벽》에 담긴 희망」, 임경석·차혜영 외, 《개벽에 비친 식민지 조선의 얼굴》, 모시는사람들, 2007; 최수일, 「1920년대 문학과 《開闢》의 위상」, 성균관대 박사학위논문, 2002; 한기형, 「《개벽》의 종교적 이상주의와 근대문학의 사상화」, 같은 책, 2007 참조.

¹⁶ 제 3 기(48~72 호)의 ‘불안정한 균형’(주도층은 ‘천도교적 사회개조’를 견지하며 사회주의와 차별화) 논지는 허수(2008b) III 장 1 절 「주요논설과 표상공간의 종합적 이해」(368-370 쪽) 및 IV 장 맺음말(380-382 쪽)에서 논증하였다.

¹⁷ 전처리(前處理)란 분석에 앞서 불필요한 요소를 제거하고 텍스트를 정리하는 작업이다.

¹⁸ 코퍼스(corpus)란 분석 대상으로 수집한 텍스트 집합체를 가리킨다.

¹⁹ 잠재 디리클레 할당(LDA, Latent Dirichlet Allocation)을 사용하였다. 토픽 수 8 개, 토픽당 단어 수 20 개, 연산 반복 1,000 회로 설정하고, TF-IDF 로 수치화한 뒤 LDA(Scikit-learn)로 8 개 토픽을 추출하고, 단어 간 PMI(점별 상호 정보량)에 기반한 토픽 보정을 거쳤다. 토픽 보정은 김동욱·이수원(2017)의 방법과 임계값(0.3)을 적용한 것으로, 유효 토픽 단어가 2 개만 남은 토픽 1 개(토픽 2)를 제외하여 최종 7 개 토픽을 확정하였다. 즉 토픽 수 확정 연구자의 임의적 판단이 아니라 정해진 보정 절차에 따른 결과이다. 상세한 절차는 허수(2021), 230~239 쪽 참조.

²⁰ 독립표본 T 검정(Independent Samples T-test) 결과, 토픽 1(사회주의)은 $t = -3.718$, $p < 0.001$, 토픽 8(개조론)은 $p = 0.005$ 로 두 집단 간 평균 차이가 통계적으로 유의미하였다. 정규성 검정(Kolmogorov-Smirnov, Shapiro-Wilk)과 등분산 검정(Levene)을 사전에 실시하였다. 상세한 검정 절차는 허수(2021), 248~255 쪽 참조. 쉽게 말해, 두 집단(개벽 주도층과 일반 필자층) 사이에서 특정 토픽의 비중 차이가 우연이 아닌 실제 차이인지를 통계적으로 확인한 것이다.

²¹ 분석에는 작업 시점(2025년 11월 하순)에 공개된 구글 제미나이의 최신 모델(2025.11.18 공개된 Gemini 3 Pro 로 추정)을 사용하였다. 작업 당시 버전을 별도로 기록하지 않은 점은 한계이다. 문장 분할에는 제미나이를, 원문·번역문의 관계 분류에는 구조적·의미적 분석을 사용하였다. 분류의 타당성은 다국어 문장 임베딩 모델 LaBSE(sentence-transformers/LaBSE)로 산출한 코사인 유사도로 정량적으로 뒷받침된다. 유형별 평균 유사도는 1:1 번역(0.59) > N:1 요약(0.48) > 변형(0.33) > 제목 변형(0.28)으로, 직역에 가까울수록 높고 요약·변형일수록 낮은 단조적 경향을 보였다. 다만 표본 검수에서 고(高)유사도 문장이 직역에 부합함을 확인하는 데 그쳤고, 전체 정량 신뢰도 평가는 후속 과제로 둔다. 문장 분할 시스템

프롬프트, 코사인 유사도 산출 스크립트, 데이터셋(Source·Target·Relations, similarity 점수 포함)과 Neo4j Cypher 쿼리 전체를 아래 저장소에서 공개한다.
<https://github.com/hursoo/gaebyeok-mechnikov-mapping>.

²² 로그우도비(LLR, Log-Likelihood Ratio)란 두 텍스트 집단에서 특정 단어의 출현 빈도 차이가 우연인지 의미 있는 것인지를 판별하는 통계 지표이다.

²³ Neo4j(네오포제이)는 데이터 간의 관계를 그물망 형태로 저장·탐색할 수 있는 그래프 데이터베이스 소프트웨어이다.

²⁴ 파란색 노드는 원문(《철학과 종교》)의 문장, 분홍색 노드는 번역문(《개벽》)의 문장이다. 크게 표시한 네 개의 노드는 두 글의 처음과 끝 문장이다. 원문과 번역문이 각각 첫 문장 하나, 마지막 문장 하나를 크게 나타낸 것으로, 가운데 아래에 모여 있는 두 노드가 두 글의 첫 문장이고, 거기서 떨어져 있는 두 노드가 두 글의 마지막 문장이다.

²⁵ 원문(《철학과 종교》) 문장을 주제별로 색칠한 것이다(초록=과학, 빨강=철학, 노랑=권위). 과학(초록) 문장은 대부분 번역문과 이어지지만, 철학(빨강)·권위(노랑) 문장은 번역문과 연결되지 않아 번역에서 빠졌음을 보여준다.

²⁶ 개념사(概念史, Begriffsgeschichte)란 핵심 개념들의 역사적 형성과 변화를 추적하는 연구이다.

²⁷ 고증학(考證學)이란 문헌의 원전을 엄밀하게 고증하는 동아시아의 학문 전통이다.

²⁸ 텍스트 비평(textual criticism)이란 필사·번역·편집 과정에서 발생한 텍스트의 변이를 추적하여 원전의 모습을 복원하는 작업이다.

²⁹ 서구 학계에서는 전통적 문헌학과 디지털 기술의 접점을 다루는 논의가 ‘Digital Philology’라는 이름으로 진행되고 있으며, 존스홉킨스대학교 출판부에서 동명의 학술지(Digital Philology: A Journal of Medieval Cultures, 2012~)를 발행하고 있다. 다만 이 학술지의 주된 관심사는 중세 유럽 텍스트의 디지털 편찬이며, 본 논문에서 제안하는 ‘디지털 문헌학’은 동아시아의 ‘중역적 지식 수용’이라는 고유한 문제 영역에서 출발한다는 점에서 차이가 있다.

³⁰ 허수, 「디지털 문헌학으로 본 20 세기 초 현상즉실재론의 한국 유입 — 『철학과 종교』에 대한 이돈화의 취사선택을 중심으로」, 학술대회 발표문, 2026.

³¹ Anthropic 사(엔트로픽사)의 클로드 코드(Claude Code)를 가리킨다. 자연어 요청을 통해 코드 작성·실행·디버깅을 수행하는 AI 코딩 에이전트로, 파일 읽기·편집, 웹 검색, 데이터 분석 등 범용적 연구 보조 기능도 갖추고 있다. 본 논문의 원고 작성·교정 과정에서도 이 도구가 활용되었다.

³² ‘산업혁명(Industrial Revolution)’이라는 용어는 프랑스 저술가들이 먼저 사용하였으나, 영국 경제사학자 Arnold Toynbee(1852~1883)의 유고집 《18 세기 영국 산업혁명 강의》(1884)가 출판된 뒤부터 널리 보급되었다. “Industrial Revolution”, Encyclopaedia Britannica, <https://www.britannica.com/event/Industrial-Revolution>; 위키백과, 「산업 혁명」, https://ko.wikipedia.org/wiki/산업_혁명 참조.

³³ AI 시대 연구자의 역할을 디지털 큐레이터로 파악하는 관점은 한국학중앙연구원 김현 교수의 가르침에 힘입은 것이다.

참고문헌

김동욱, 이수원 (2017). “단어 유사도를 이용한 뉴스토픽 추출”. <정보과학회논문지>. 44(11). 1139, 1141~1144, 1146.

김정인 (2007). “《개벽》을 낳은 현실, 《개벽》에 담긴 희망”. 임경석, 차혜영 외 편. <개벽에 비친 식민지 조선의 얼굴>. 모시는사람들.

최수일 (2002). “1920년대 문학과 《開闢》의 위상”. <성균관대학교 박사학위논문>.

한기형 (2007). “《개벽》의 종교적 이상주의와 근대문학의 사상화”. 임경석, 차혜영 외 편. <개벽에 비친 식민지 조선의 얼굴>. 모시는사람들.

허수 (2008a). “1920년대 초 《개벽》 주도층의 근대사상 소개양상 — 형태적 분석을 중심으로”. <역사와 현실>. 67. 47~76. [본문의 허수(2008a)] - URL: <https://hs-kmhistory.com/publications/> (검색일: 2026. 03. 02.)

허수 (2008b). “《개벽》의 '表象空間'에 나타난 매체적 성격 — 표지 및 목차 분석을 중심으로”. <대동문화연구>. 62. 353~390. [본문의 허수(2008b)] - URL: <https://hs-kmhistory.com/publications/> (검색일: 2026. 03. 02.)

허수 (2021). “《개벽》 논조의 사회주의화에 관한 새로운 접근 — 토픽 연결망 분석을 중심으로”. <인문논총>. 78(1). 221~262. [본문의 허수(2021)] - URL: <https://hs-kmhistory.com/publications/> (검색일: 2026. 03. 02.)

허수 (2025). “20세기 초 근대 철학서의 선택적 수용 양상 — 이돈화와 이노우에 데쓰지로”. <연구계획서(미간행)>. [본문의 허수(2025)] - URL: <https://hs-kmhistory.com/research/> (검색일: 2026. 03. 02.)

허수 (2026). “디지털 문헌학으로 본 20세기 초 현상즉실재론의 한국 유입 — 『철학과 종교』에 대한 이돈화의 취사선택을 중심으로”. <학술대회 발표문>. <https://hs-kmhistory.com/2026/02/13/현상즉실재론의-전유와-한국적-세계관의-모색/> (검색일: 2026. 03. 02.)

中澤臨川·生田長江 編, 《近代思想十六講》, 東京: 新潮社, 1915.

生田長江·本間久雄 共著, 《社會改造の八大思想家》, 東京: 東京堂書店, 1920.