

일제 식민지 시기 근대 학술 규범의 형성과 다양성

허수**

『진단학보』 논설 형대에 대한 정량적 분석*

초록 『진단학보』는 일제 식민지 시기 한국에서 발행된 대표적인 학술지이다. 그동안 학계에서는 이 학술지를 주로 실증주의 역사학의 중심으로 평가하였다. 그러나 최근 연구들은 진단학회의 구성원과 『진단학보』의 학술적 성격이 다양했음을 지적하고 있다. 본 연구에서는 학술지 수록 논설을 형태적 측면에서 고찰하여 다양성 여부, 다양성의 발현 양상 등을 살펴보았다. 분석은 정량적인 방법을 사용하였다.

첫째, 『진단학보』 논설 59편의 디지털 코퍼스를 DB 형태로 구축하고, 다양성 여부를 판별할 지표 6개를 제시하였다. ‘장절 구분’, ‘논설 규모’, ‘자료 규모’, ‘주석 수’, ‘주석 규모’, ‘주석 평균 길이’가 그것이다. 둘째, 6개 지표를 특성(feature)으로 활용해서 이를 기준으로 59개 논설을 비슷한 것끼리 군집화하였다. 군집화 알고리즘은 K-means 방식을 사용하였으며, 8개의 군집을 얻었다. 각 군집은 6개 특성을 나름대로 반영하여 변별적인 성격을 드러내었다. 셋째, 규모가 크고 개성이 더욱 뚜렷한 군집 중심으로 주요 필자와의 관계, 시기별 비중 변화 등을 살펴보았다. ‘간결 주석형 학술 논설’, ‘상세 주석형 학술 논설’, ‘간결한 기고형 논설’이 그것이다. 여러 글쓰기 형태가 병존하는 가운데 이 세 유형이 당시의 글쓰기 방식을 대표했다고 보인다. 이 중 ‘상세 주석형 학술 논설’의 비중은 전기에 비해 후기에 크게 높아지는 추세를 보였다.

주제어 진단학보, 정량적·형태적 분석, 장절 구분, 자료, 주석, 군집 분석

* 이 논문은 서울대학교 인문학연구원이 지원한 집담회의 성과임.

본 논문은 2024년 12월~2025년 2월, 서울대 인문학연구원 주최 학술회의에서 발표한 「진단학보 디지털 자료의 활용 방안 모색」(“진단학보의 디지털 인문학”, 2024. 12. 6. (금), 서울대 137-2동)과 이 내용을 조금 업데이트한 「진단학보 데이터 처리와 활용」(“진단학보 데이터 처리와 활용”, 2025. 2. 6. (목) 서울대 137-2동)을 대폭 수정한 것이다. ‘근대 학술 규범’이라는 용어는 첫 학술회의에서 토론자로 참여한 류준필 교수의 논평에 빚진 바 크다.

** 서울대학교 역사학부 한국사학전공 교수

1. 서론

『진단학보』는 일제 식민지 시기였던 1934년 한국인 학자들이 주도하여 창간한 인문학 학술지이다. 전통적으로 연구자들은 이 학술지를 ‘실증(주의) 역사학’의 중심지로 규정하며, 일본 제국대학이 확립한 엄밀한 실증 방법을 그대로 도입한 것으로 보았다.¹

그러나 정병준(2016)은 진단학회가 총독부 관제 역사학을 비판하며 한국 역사학의 근대적 출발점을 마련했다고 주장하였다.² 조범성(2021)은 청구학회와 비교하여 진단학회가 관제 학술단체와 달리 자발적 학술 교류를 조직했다고 강조하였다.³ 민현구(2012)는 두계 이병도가 일본인 사학자 이마니시 류(今西龍)에게서 학술 기법을 습득한 과정을 추적하고, 그가 『진단학보』를 매개로 근대 연구 규범을 확산했다고 분석하였다.⁴

이처럼 최근 연구들은 『진단학보』를 단일한 학파나 방법론의 전형으로 고정하지 않고, 근대적 학술 흐름이 교차한 장(場)으로 바라보기 시작하였다. 그럼에도 불구하고 기존 연구는 주로 『진단학보』의 사상적 내용이나 역사적 의의에 집중했을 뿐, 논설이 어떤 형태로 이루어져 있는지는 충분히 탐구하지 않았다. 홍종욱·장문석·류준필(2025)은 『진단학보』의 디지털 아카이브 구축 사례를 통해 주석과 자료 인용 방식이 근대 학술 규범의 핵심 지표임을 제시하여 이 텍스트의 형태적 측면에 주목한 의의가 있다.⁵ 그렇지만 이 경우에도 주석과 자료 등에 구현된 근대적 학술 규범의 구체적 양

1 국사편찬위원회, 「실증주의 사학」, 우리역사넷(열람 2025-05-04)의 ‘내용’ 단락; 홍종욱(2019), 「실증사학의 ‘이념’: 식민지 조선에 온 역사주의」, 『인문논총』 76-3, p. 287.

2 정병준(2016), 「식민지 관제 역사학과 근대 학문으로서의 한국역사학의 태동: 진단학회를 중심으로」, 『사회와 역사』 110, pp. 105-162.

3 조범성(2021), 「1930년대 청구학회의 설립과 활동」, 『한국민족운동사연구』 107, pp. 5-36.

4 민현구(2012), 「두계 이병도의 수학과정과 초기 학술활동」, 『진단학보』 116, pp. 1-38.

5 홍종욱·장문석·류준필(2025), 「한국 근대문헌 디지털 아카이브의 설계: 『진단학보』를 중심으로」, 『인문연구』 110, pp. 79-122.

상을 본격적으로 분석한 연구는 아니다.

본 논문은 이상의 문제의식하에서 『진단학보』를 근대 학술 규범의 확산·적용 과정을 반영한 텍스트로 읽는다. 여기서 ‘근대 학술 규범’이란 ‘과거 전통적인 글쓰기 방식과 달리 연구의 근거와 논증 방법을 명확히 제시하는 새로운 글쓰기 방법’을 가리킨다. 근대 학술 규범을 살필 수 있는 요소로 여러 가지를 상정할 수 있다. 본 논문에서는 형태적 측면, 특히 목차(장절 구분 형식), 주석(annotation), 본문 인용 자료(source, 이하 ‘자료’로 줄임) 등에 주목하고자 한다. 이런 요소의 형식, 규모, 빈도 등을 정량적으로 분석해서 1930년대 한국 학계가 근대적 글쓰기 방식을 도입한 초기 양상을 규명할 것이다.

본 논문에서 해명할 연구 질문은 다음과 같다. ‘일제하 진단학보 수록 논설은 형태 측면에서 과연 다양성을 뚜렷히 드러내는가’, ‘드러낸다면 구체적으로 어떤 양상을 띠었나’이다. 이 질문에 대한 탐구 과정은 다음과 같다. 2장에서는 자료 수집 및 분석 방법을 제시한다. 3장에서는 군집화를 통해 산출한 군집들의 특징을 고찰한다. 4장에서는 규모가 크고 개성이 뚜렷한 주요 군집의 동향을 드러내고, 결론인 5장에서는 연구 결과를 요약하며 한계와 향후 과제를 논의한다.

본 논문에서 주목한 진단학보 수록 논설에 관한 정보, 본격적인 분석을 위해 입력 데이터로 사용한 진단학보 논설 DB, 전처리 및 분석과정에서 활용한 파이썬 코드 등은 깃허브(github)에 공개한다.⁶

2. 진단학보 DB 구축과 분석 지표

2.1 진단학보 논설 DB 구축

한국인 학자들은 1934년 『진단학보』를 창간해 1941년까지 14호를 발

간했다. 현행 목록에는 기사 85편이 수록되어 있으며, 그중 가장 높은 비중을 차지하는 ‘논술’ 분류에는 77편(이하 ‘논설’)이 속해 있다.⁷ 이 중 필자는 주석과 본문이 모두 디지털화된 59편(논설의 76.6%)을 분석 대상으로 선정하였다.⁸ 이들 논설은 창간호부터 14호까지 고르게 분포해, 학술 글쓰기 형태의 변동과 다양성을 살피기에 충분하다.⁹ 논설 DB 구축 절차는 다음과 같다.

- ① 원문 PDF 59편을 한국학술정보(KISS)에서 내려받았다.
- ② 동일 원문의 html 파일을 서울대 인문학연구원 ‘진단학보 원고(Article) 데이터 목록’(2024-11-28 기준)에서 확보하였다.
- ③ 고유번호, 필자, 제목, 권호, 발행일, 시작·끝 쪽수 등 서지 메타데이터를 추출하였다.
- ④ 본문 텍스트와 주석 텍스트를 분리·정제한 뒤 논설별로 연결해 ‘진단학보 논설 DB’를 완성하였다.

완성된 진단학보 논설 DB에는 주석 일련번호, 주석 길이, 본문 내 주석 내용 위치, 인용문 길이 등 형태 정보가 체계적으로 담겨 있다. 이를 정략적 분석의 기초 자료로 활용하였다.¹⁰

7 본 논문에서 활용한 주요 데이터를 쉽게 확인할 수 있는 주소는 다음과 같다(2025년 5월 말 현재).

a. 진단학보 원문: “震檀學報(원문)”(<https://drive.google.com/drive/folders/18VTPjpPpAC51TewL2syGIMPT7BxFHuM9>)

b. 원문 전산입력물(제목, 본문, 필자정보 등): “진단학보 - North Korea Humanities” (<https://dh.aks.ac.kr/~nkh/wiki/index.php/%EC%A7%84%EB%8B%A8%ED%95%99%EB%B3%B4>).

8 59편 논설의 자세한 서지 사항은 ‘[부표 2] 분석 대상으로 선정한 진단학보 59개 논설의 서지 사항’을 참조.

9 진단학보 수록 논설의 필자별, 시기별 분포는 ‘[부표 1] 진단학보 수록 논설의 분포(필자별, 시기별)’ 참조.

10 진단학보 논설 DB의 자세한 내용 및 구조는 ‘[부표 3] 진단학보 수록 논설 DB 내용 및 구조’를 참조.

2.2. 다양성 여부 판별 지표

이 절에서는 『진단학보』 수록 논설의 다양성 여부를 판별할 지표를 목차와 주석 및 자료의 세 요소에서 구하였다. 필자가 볼 때, 이들은 논설의 형태적 측면을 구성하는 중요한 요소이다.

우선, 주석과 자료의 중요성은 현행 『진단학보』 편집 규정을 통해서도 알 수 있다. 주석에 관하여는 “모든 주(註)는 각주(footnote)로 처리”한다는 공통 원칙 아래, 기본적으로 외각주(out-footnote) 방식을 채택하고, 국어국문학·사회과학 분야에 한해 내각주(author-date)를 예외적으로 허용한다.¹¹ 두 유형 모두 주석 내용, 즉 사료 출처와 해설을 페이지 내에 제시해 독자가 즉시 근거를 확인할 수 있게 한다. 이런 형식은 다른 학술지도 비슷하다. 1930년대 『진단학보』 논설은 각주를 사용하지 않고 대부분 미주(尾註) 방식을 채택하였다. 그렇지만 이런 방식 역시 ‘근거 제시 및 확인’이라는 주석 본래의 기능을 담당한다는 점에서는 동일하다.

자료의 경우, 현행 편집 규정은 본문 인용문을 길이에 따라 뚜렷이 구분한다. 한 줄 안에서 줄바꿈 없이 이어지는 짧은 인용문은 큰따옴표(“”)로 묶어 본문과 연속시킨다. 반면 두 줄 이상 넘어가는 긴 인용문은 한 줄을 띄운 뒤 글자 크기를 1포인트 줄이고 좌측을 들여쓴 블록으로 제시하며, 출전은 모두 각주에 달도록 요구한다.¹² 이러한 시각적 분할은 서술 텍스트와 원사료를 명확히 떼어 독자의 검증 편의를 높이려는 근대 학술 규범의 핵심 요소이다. 그런데 당시 『진단학보』의 경우 짧은 인용과 긴 인용의 경계가 모호하고, 긴 인용도 큰따옴표와 줄바꿈 없이 본문에 그대로 이어 적는

11 진단학회 웹페이지(<https://chintan.jams.or.kr/co/main/jmMain.kci>)에 있는 ‘학회논문투고 규정’ 참조. 외각주(out-footnote)란 본문에 숫자를 표시한 뒤, 같은 페이지 하단에 그 숫자에 맞춰 출처·해설을 길게 적어 두는 방식이다. 내각주(author-date 방식)란 본문 안 괄호에 저자, 발행연도, 쪽수만 간단히 적고, 자세한 서지 정보는 논문 말미 참고문헌에서 제시하는 방식이다.

12 진단학회의 “학회논문투고 규정” 참조.

사례가 빈번하다. 즉 서술 영역과 자료 영역이 형식적으로 뒤섞여 있었다. 따라서 본 논문에서는 자료 부문을 적절한 방식으로 분리해서 이용할 것이다. 여기에 관하여는 조금 뒤에 자세히 서술한다.

주석과 자료 이외에 목차의 형태도 중요하다. 여기서 목차 형태란 장과 절을 구분하는 방식을 형태 측면에서 포착한 것이다. 당시 논설들을 관찰하면 장절 구분은 크게 세 유형으로 나뉜다. 오늘날 대부분의 논문처럼 구분 내용의 첫머리를 텍스트로 표현한 ‘제목형’, 번호만 제시한 ‘번호형’, 장절 구분과 구분 표시가 없는 ‘미구분형’이 그것이다.

이상 목차, 주석, 자료의 세 요소로 논설들 속의 숨겨진 패턴을 파악하기 위하여 이 절에서는 다음 [표 1]과 같이 여섯 가지 지표를 제안한다.

한편 [표 1]의 지표를 적용하는 데 기초가 되는 ‘글자 수’는 다음 기준과 절차를 거쳐 산출하였다.

첫째, 글자 수는 전산화 작업에서 생성한 지시어와 부호, 중복 텍스트 등을 제거한 뒤 계산하였으며, 한글과 한자를 모두 1자로 세었다.

둘째, 자료 글자 수 계산을 위해서는 ‘자료’의 범위를 정의할 필요가 있다. 주석의 경우 본문과 주석이 형태 측면에서 확실히 구분되지만, 본문 내

[표 1] 논설 유형을 파악하기 위한 지표

영역	지표 명	설명	기타
목차	장절 구분	본문 세부 내용 구분 표현	제목형, 번호형, 미구분형
논설	논설 규모	본문과 주석의 글자 수 합계	
본문	자료 규모	(자료 글자 수 / 본문 글자 수) * N	
주석	주석 수	(주석 개수 / 논설 규모) * N	
	주석 규모	(주석 글자 수 합계 / 논설 규모) * N	
	주석 평균 길이	(주석 글자 수 합계 / 주석 개수) * N	

참고: 동일한 비교 기준을 위하여 자료 및 주석 관련 지표는 상대적인 값을 사용하였다. N값은 0이 아닌 숫자의 출현 지점이 소수점 앞 뒤 2-3자리 이내가 되도록 지표별로 각각 상이하게 설정하였다.

에서 자료 영역은 그 외의 영역, 즉 서술 영역과 뚜렷이 구분되지 않기 때문이다. 이 구분을 위해 평균 50자(공백 제외)를 최소 단위로 설정하고 이 단위에서 한자가 90% 이상 있는 경우에 이곳을 ‘자료 영역’으로 간주하였다. ‘50자 기준’을 사용한 것은, 당시 논설이 논설 간이든 동일 논설 내에서도, 문장 길이 간 편차가 크고 단락은 지나치게 길어서, 문장이나 단락 단위의 분석은 적절하지 않다고 판단했기 때문이다. 이런 선택에는 세로 줄 한 줄의 평균 글자 수가 대체로 50자 내외인 점도 감안하였다. ‘한자 90% 기준’ 설정 이유는 당시 『진단학보』 수록 논설이 조선시대 및 그 이전 시기의 자료를 다루고 있어서 한문 자료를 원문 그대로 인용하는 경우가 대다수였기 때문이다. 70%·80% 기준도 살펴보았으나 90% 이상의 경우가 실제 상황을 더욱 잘 반영하였다.

3. 논리적 서술의 다양한 방식

3.1. 형태 지표에 관한 기초 정보

[표 1]에서 6개의 지표를 제시하였는데, 이 지표 산출에 필요한 기본 사항은 장절 구분, 본문 및 주석 글자 수, 본문 내 자료 글자 수, 주석 개수 등에 근거를 두었다. 따라서 이 절에서는 여기에 관한 내용을 상세히 언급해 둔다.

첫째, 장절 구분이다. 분량 면에서 ‘제목형’이 42개로 가장 많다. 이는 당시 논설이 소제목을 활용하여 내용을 체계적으로 분류하는 방식을 선호했음을 의미한다. 그다음으로는 ‘미구분형’으로 12개이다.¹³ 이는 비교적

13 예외적으로 손진태의 “朝鮮 古代 山神의 性에 就하여”(손진태, 1934, 11.)는 본문 첫 줄에 세부 내용 제목을 연상시키는 단어나 句를 하이픈(‘-’)으로 연결해서 제시하였다. 그러나 이런 구분이 본문에서는 보이지 않아서 ‘미구분형’에 포함하였다.

	본문 글자 수	주석 글자 수	주석 개수
상자 그림			
통계값	최댓값: 38,711 중앙값: 16,057 최솟값: 2,706	최댓값: 4,024 중앙값: 633 최솟값: 0	최댓값: 78 중앙값: 13 최솟값: 0

- 상자그림에서 최댓값과 최솟값의 정의는, 각각 '이상치를 제외한' 값 중에서 가장 큰 값과 가장 작은 값이다. 상자의 위와 아래로 뻗어나간 선 끝에 있는 수평선이 여기에 해당한다. 그 너머의 'o' 표시는 이상치를 표현한 것이다.
- 중앙값은 데이터를 크기 순서대로 나열했을 때 맨 가운데 있는 값으로, 상자 그림에서 네모 상자 안의 가로선에 해당한다.
- 네모 상자 영역은 데이터의 중간 50%가 모여 있는 범위를 나타낸다.

[그림 1] 본문 및 주석 글자 수 및 주석 개수 분포

간결한 논설이나 연구자의 특정 서술 방식에 따른 것으로 보인다. '번호형'은 5개에 불과하다.

둘째, 본문 및 주석의 글자 수와 주석 개수, 자료 글자 수 등의 개괄적 상황은 [그림 1]과 [그림 2]의 상자 그림(boxplot)을 통해 잘 알 수 있다.

[그림 1]에 따르면 논설의 본문은 대략 1만 자에서 2만 5,000자 사이가 가장 많았다. 오늘날 기준으로 보면 대략 A4 용지 5~17장 정도이다.¹⁴ 이는 당시 학술 논문으로서 상당한 분량을 갖춘 것으로, 연구의 깊이를 추구했음을 짐작하게 한다. 최대 3만 8,711자에 달하는 긴 논설도 있었다.

— www.kci.go.kr

14 A4 1장에 공백을 제외하고 1,500자에서 2,000자 정도가 들어간다고 상정하였다.

	서술 영역	자료 영역
상자 그림		
통계값	최댓값: 35,225 중앙값: 14,001 최솟값: 1	최댓값: 5,600 중앙값: 1,432 최솟값: 0

[그림 2] 본문 내 서술 영역과 자료 영역의 크기 분포

주석은 본문에 비해 상대적으로 적은 편이었다. 대부분의 논설에서 주석 글자 수는 2,000자 이하였다. 주석 개수의 경우, 대부분의 논설은 0개에서 30개 사이로 사용하였다. 이는 논문의 주요 내용은 본문에서 충분히 서술하고, 주석은 보충 설명을 최소화하는 경향이 있었음을 시사한다. 물론 주석을 하나도 사용하지 않거나 예외적으로 최대 78개를 사용한 논설도 있었다.

[그림 2]에서 알 수 있듯이, 논설의 주요 내용인 서술 영역은 본문 전체 길이와 크게 차이가 나지 않는 규모이며, 1만 자에서 2만 자 정도가 일반적이었다. 연구자의 논리 전개와 분석이 이 영역에서 이루어졌음을 보여준다. 자료 영역에서 주목할 점은 본문 내에 자료 인용 영역이 비교적 작았다는 사실이다. 대부분의 논설에서 자료 영역은 5,000자 이하였고, 심지어 자료가 거의 없는 논설도 있었다. 이는 당시 연구들이 방대한 자료 인용보다는 연구자의 서술과 분석에 중점을 두었거나, 자료를 본문에 직접 녹여내는 방식을 택했을 가능성을 보여준다. 물론 다음 장에서 살펴볼겠지만, 이곳 상

자그림에서도 확인할 수 있듯이 이상치가 다수 존재하였다.

3.2. 군집화를 통해 본 학술적 스펙트럼

이 절에서는 2장의 [표 1]에 제시한 6개의 지표를 ‘특성’(feature)으로 활용하여 59개 논설을 군집화하였다. 군집화는 [그림 3]과 같은 순서대로 진행하였다.

[그림 3]의 결과를 보면 6개 특성 중에서 군집화 결과에 가장 큰 영향을 끼친 것은 ‘주석 평균 길이’(f6_anno_length)였다. ‘장절 구분’(f1_encoded)과 ‘주석 수’(f4_anno_count)가 그 뒤를 이었다. 이 점을 염두에 두고 8개의 군집을 유형화해서 표현하면 다음과 같다.¹⁵

군집 1 - ‘간결 주석·자료형’에 해당한다.¹⁶ 18편으로 가장 많은 논설

15 군집 설명을 제시한 순서는, 아주 엄밀하진 않지만, 가급적 논설의 ‘학술성’과 ‘정형성’을 기준으로 가장 전형적인 학술 논문이라 판단되는 것부터 시작하여, 자료 및 주석의 특징에 따른 차이를 보여준 후, 점차 자유로운 형식의 논설로 이어지도록 하였다.

16

<p>김석형(jdh194106), 李朝初期 國役編成의 基底 김영수(jdh193711), 五敎兩宗에 對하여 김영수(jdh193807), 曹溪禪宗에 就하여 - 五敎兩宗의 一派, 朝鮮佛敎의 根源 - 송석하(jdh193411), 風神考(附禾竿考) 유흥렬(jdh193607), 朝鮮祠廟發生에 對한 一考察 - 特히 麗末李朝初의 報本崇賢思想을 中心으로 - 이병도(jdh193411a), 三韓問題의 新考察(一)~(七完) [+ 3509a, 3604a, 3607, 3611a, 3707, 3711a] 이병도(jdh193807), 李朝初期의 建都問題 이승녕(jdh193912), 朝鮮語異化作用에 對하여 이승녕(jdh194009), 「」 音攷; 조운제(jdh193411), 朝鮮詩歌의 胎生; 조운제(jdh193912), 春香傳異本考 (1)~(二) [+ 4009]</p>	
---	--

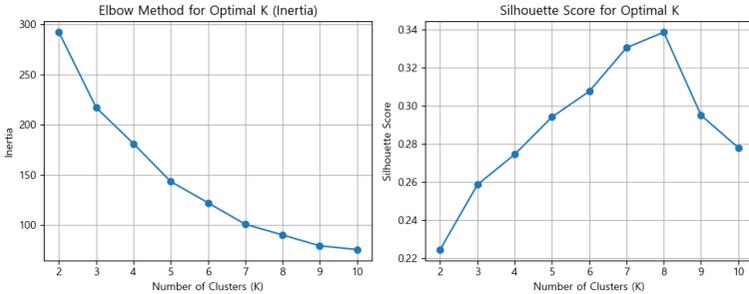
실행 단계	세부 내용									
1) 지표 정규화	- ‘장절 구분’ 이외의 수치형 데이터는 0~1사이의 값으로 환산하였다. - ‘장절 구분’의 경우, ‘미구분형’에는 0, ‘번호형’에는 1, ‘제목형’에는 2를 각각 부여하였다. ¹⁷									
2) 군집화 개수 결정	- 지표로 이너시아(Inertia)와 실루엣 점수를 활용하였다. 전자는 급감하다가 감소가 둔화되는 지점(= 엘보우 지점), 후자는 1에 가장 가까운 지점이 최적의 군집화 개수(k값)에 해당한다고 볼 수 있다. - 이너시아에서 엘보우 지점이 분명하지 않지만, 실루엣 점수가 8에서 최고점인 사실을 감안하여 k값을 8로 결정하였다. ¹⁸									
3) K-means 군집화	군집 번호	0	1	2	3	4	5	6	7	
	논설 개수	5	18	12	11	1	7	2	3	
4) 결과 시각화 ¹⁹										
5) 특성 기여도 분석 ²⁰										

[그림 3] 군집화 실행 단계 및 관련 내용

을 포함한다. '제목형' 장절 구분을 하고 주석 수도 평균 수준으로 논문 형식을 갖추었다. 하지만 주석 평균 길이와 주석 규모, 자료 규모는 매우 낮은 편이다. 이는 주석이나 자료 인용을 통한 상세한 외부 자료 제시보다는 저자 본연의 서술에 중점을 둔 저술 방식의 특징으로 해석된다. 연작 논설이 다수 포함된 점도 주목할 만하다.

군집 2 - '상세 주석·자료형'이다.²¹ 12편으로 두 번째 큰 군집이다. 가

- 17 본 논문에서는 이 세 유형을 각각 서로 무관한 범주형 데이터가 아니라 '순서형 범주형 데이터'로 취급하였다. 왜냐하면, 요즈음의 논문 형태를 기준으로 볼 때, 미구분형보다는 번호형이, 번호형보다는 제목형이 근대적 학술 규범에 더 가깝다고 보았기 때문이다.
- 18 이너시아 그래프(좌)와 실루엣 점수(우)



- 19 차원 축소 및 시각화는 t-SNE 방식을 사용하였다.
- 20 특성 기여도 분석이란 군집 결과를 더 잘 이해하기 위하여, 전술한 군집화에서 어떤 특징이 어느 정도로 작용했는가를 살펴보는 작업이다. 머신러닝 중 '의사결정나무'(Decision Tree) 모델을 사용하였다.
- 21

<p>고유섭(jdh193509), 高麗畫跡 對하야 고유섭(jdh193904), 朝鮮塔婆 研究 (2) 김상기(jdh193607), 葛文王考 김상기(jdh193807), 三別抄와 그의 亂에 就하야 (1)~(3) [+ 3904, 4103] 도유호(jdh194009a), 中國都市文化의 起源 (1)~(3) [+ 4103, 4106] 박시형(jdh194106), 李朝田稅制度의 成立過程 송석하(jdh193504), 處容舞, 儺禮, 山臺劇의 關係를 論함 이병도(jdh193904a), 圖讖에 대한 一二의 考察 (1)</p>	<p>Cluster 2 특성 프로파일</p>
--	--------------------------

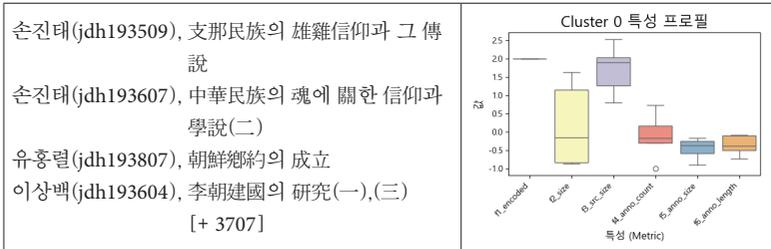
장 일반적인 학술 논문의 형태를 보인다. 명확한 장절 구분과 충분한 논설 규모, 그리고 상세하고 긴 주석을 통해 주제에 대한 심층적인 탐구와 탄탄한 근거 제시가 이루어지는 것으로 생각된다.

군집 0 - '상세 자료형'으로 부를 수 있다.²² 5편이 여기에 속한다. 자료 규모는 크지만 주석 규모는 작고 주석 평균 길이는 짧은 편이다. 주석 개수는 군집 5보다 적지만, 이러한 특징이 어우러져 견고한 자료 기반과 간결한 주석 형식을 겸비한 연구 형태를 보여준다.

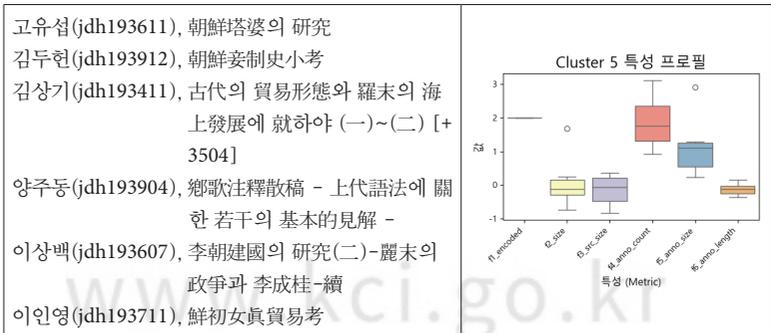
군집 5 - '상세 주석형'이다.²³ 모두 7편이다. 장절 구분이 잘 되어 있는 학술 논문 형태를 띠지만, 주석 수와 주석 규모가 크다. 논설 규모는 평균 수준이다. 다양한 참고문헌과 자료를 주석에서 풍부하게 제시하여 논지의 신뢰도를 높인 연구 논설임을 시사한다.

군집 6 - '주석 부기형'이다.²⁴ 이 유형의 논설 2편은 논설 규모는 작지만, 주석 평균 길이가 매우 긴 것이 특징이다. 주석에 많은 해설을 담고 있

22



23

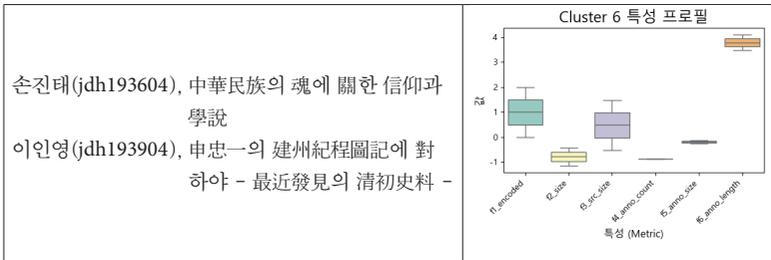


다. 실제로 두 편의 논설을 살펴보면 본문의 대부분에서 주석을 사용하지 않다가 본문이 끝난 뒤 ‘부기’(附記)에 해당하는 내용을 붙이거나 주석 번호를 한두 개 붙여 논문 집필 맥락이나 후일담 등을 다소 길게 서술하였다. 그 결과 주석 평균 길이가 과장되어 산출된 경우로, 이 군집이 실질적으로 주석의 평균 길이가 최고인 경우를 나타낸다고 말하기는 어렵다.

군집 4 - 이 유형은 단 한 편으로 이루어져 있다.²⁵ 이 유형은 본문 길이에 비해 자료와 주석의 비중이 비정상적으로 높다. 사실 이 논문은 본문 전체가 중국어로 되어 있어서, 본문 내용이 ‘한자’로 오인되어 자료 비중이 100%로 잘못 계산된 것이다. 따라서 유의미한 군집에 포함하기는 힘들다.

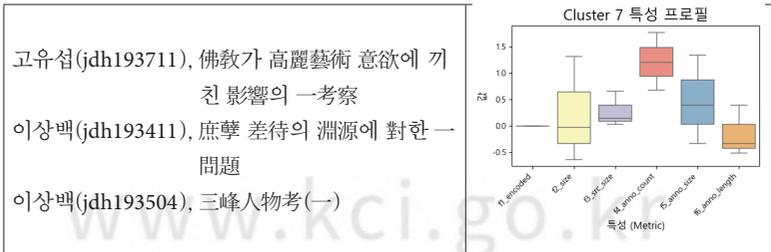
군집 7 - ‘자유 형식의 심층 고찰형’이다.²⁶ 3편이 여기에 속한다. 형식적으로는 장절 구분이 없어 에세이나 기고에 가까워 보이지만, 주석 수 및 주석 규모가 크다. 자유로운 형태 속에서도 학술적인 깊이를 추구하며, 많은 참고 자료를 바탕으로 특정 주제를 심층적으로 고찰한 글들이다.

24



25 김영건(jdh193904), 安南普陀山名考

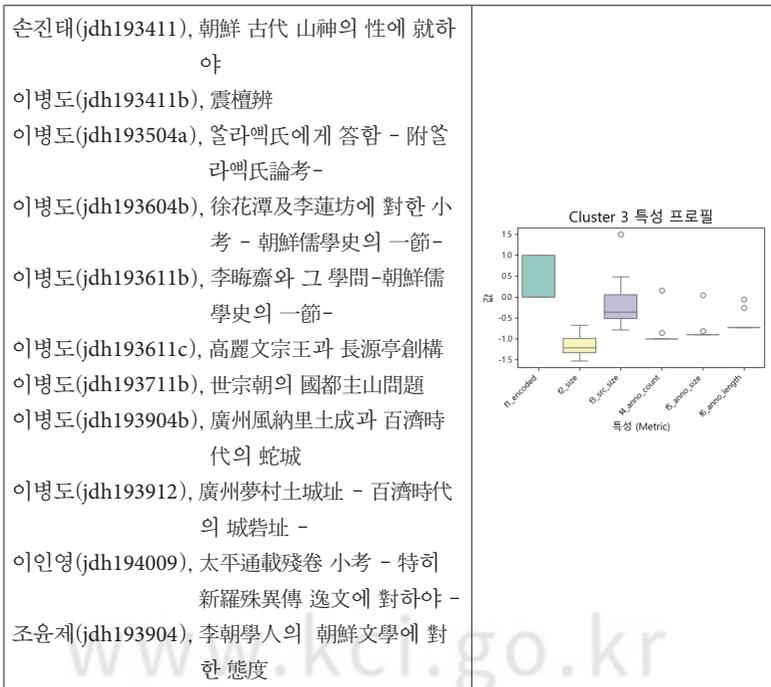
26



군집 3 - '간결 기고형'이다.²⁷ 11편으로 규모로는 세 번째이다. 장절 구분이 없고, 주석 평균 길이와 주석 수가 모든 군집 중 가장 짧고 적다. 논설 규모도 작은 편이다. 이는 학술 논문이라기보다는 짧은 고찰, 서평, 혹은 특정 주제에 대한 간결한 해설 등 형식적인 제약이 덜한 비학술적 기고문들에 가깝다.

이러한 군집화 분석을 통해 1930년대 『진단학보』가 단순히 정형화된 학술 논문만을 게재한 것이 아니라, 심층 해설, 자료 인용 중심 고증, 자유로운 형식의 고찰, 그리고 간결한 기고문에 이르기까지 다양한 형태와 깊이를 가진 글들을 포괄하는 장이었음을 알 수 있다. 이는 당시 『진단학보』의 학술적 스펙트럼과 지적 활동의 다양성을 보여주는 것으로 생각된다.

27



4. 대표적인 글쓰기 유형과 추이

4.1. 글쓰기 유형과 필자의 관계

이 절에서는 3장에서 살펴본 군집별 특징을 필자와 연결시켜 살펴보고자 한다. 군집 중에서는 소속 논설의 개수와 변별적 특징이 두드러지는 군집1, 군집2, 군집3을 '지배적 글쓰기 유형'으로 간주하고 여기에 초점을 둔다. 필자는 논설을 3편 이상 게재한 8명의 인물을 '주요 필자'로 부르고 이들의 동향에 주목할 것이다.

[표 2]에서 보듯이 주요 필자는 전체 17명 중 8명이다. 그러나 이 8명이 게재한 논설 수는 46편으로 59편 전체의 약 78%이다. 이 표에서 지배적 글쓰기 유형에 속하는 군집의 영역을 음영으로 표시하였다.

관찰 시야를 지배적 글쓰기 부분에 고정한다면, 주요 필자 간에 뚜렷한 경향성이 나타난다. 이병도와 조윤제의 논설은 대부분 군집1과 군집3에 집중된 반면, 고유섭, 김상기, 도유호의 논설은 군집2에 몰려있다. 이병도는 가장 많은 17편의 논설을 게재했고 모두가 지배적 글쓰기 유형에 속한다. 뿐만 아니라 그의 논설은 1편을 제외한 나머지 16편이 군집1과 군집3이라는 서로 대조적인 스타일의 글쓰기에 양분되었다. 말하자면 그는 '간결 주석·자료형'과 '간결 기고형' 글쓰기를 병행하면서 왕성하게 저술 활동을 했던 것이다. 이 두 방면에서 주석의 간결한 사용은 공통적 요소로 발견된다. 조윤제의 글도 이 점에서 이병도의 글과 대동소이하다.

반면 이들과 대조적으로 고유섭, 김상기, 도유호의 논설은 주석에 많은 정보량을 넣었다. 이와 같은 두 그룹별 차이가 어디서 비롯했는지 본 논문의 분석만으로는 알기 어렵다. 주요 필자들이 소속된 분과 학문, 대학 수학 경험, 연령, 해당 논설의 주제 등 여러 요소를 고려하여 추후 심도 있는 분석을 진행할 필요가 있다.

[표 2] 지배적인 글쓰기 유형과 주요 필자 간의 관계

writer \ 군집	0	1	2	3	5	6	7	개수
고유섭			3509, 3904		3611		3711	4
김상기			3607, 3807, 3904, 4103		3411, 3504			6
도유호			4009a, 4103, 4106					3
손진태	3509, 3607			3411		3604		4
이병도		3411a, 3509a, 3604a, 3607, 3611a, 3707, 3711a, 3807	3904a	3411b, 3504a, 3604b, 3611b, 3611c, 3711b, 3904b, 3912				17
이상백	3604, 3707				3607		3411, 3504	5
이인영				4009	3711	3904		3
조윤제		3411, 3912, 4009		3904				4
소계	4	11	10	11	5	2	3	46
기타필자	1	7	2	0	2	0	0	12
전체	5	18	12	11	7	2	3	58

전체 논설 수가 58인 것은 소속 논설이 1개인 군집 4가 누락되었기 때문이다.

4.2. 글쓰기 유형의 추이

이 절에서는 지배적인 글쓰기 유형의 시기별 변화 양상을 살펴보고자 한다. 3장 2절에서 산출한 군집별 논설 정보를 활용하여 [그림 3]의 순서대로 진행하였다.

지배적 글쓰기 유형의 변동은 이 유형을 포함한 전체 군집 차원에서 봐야 더욱 잘 포착할 수 있다. [그림 3]은 당초 전체 기간을 뭉뚱그려 군집화

실행 단계	세부 내용																											
1) 시기 구분 지점	창간 시점인 1934년 11월부터 14개 호 중 절반에 해당하는 7호 발행 시점인 1937년 7월까지와 그 이후를 각각 '전기'와 '후기'로 구분하였다.																											
2) 시기별 군집별 논설 개수 집계	군집화 작업 후 각 군집에 속했던 논설들을 전기와 후기별로 구분하였다.																											
3) 시기별 군집 비중 (비율) 산출	군집별 논설이 해당 시기에서 차지하는 비중을 산출하였다.																											
	<table border="1"> <thead> <tr> <th>군집 번호</th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> </tr> </thead> <tbody> <tr> <td>비중(전기)</td> <td>13.8</td> <td>31.0</td> <td>10.3</td> <td>20.7</td> <td>0.0</td> <td>13.8</td> <td>3.5</td> <td>6.9</td> </tr> <tr> <td>비중(후기)</td> <td>3.3</td> <td>30.0</td> <td>30.0</td> <td>16.7</td> <td>3.3</td> <td>10.0</td> <td>3.3</td> <td>3.3</td> </tr> </tbody> </table>	군집 번호	0	1	2	3	4	5	6	7	비중(전기)	13.8	31.0	10.3	20.7	0.0	13.8	3.5	6.9	비중(후기)	3.3	30.0	30.0	16.7	3.3	10.0	3.3	3.3
	군집 번호	0	1	2	3	4	5	6	7																			
비중(전기)	13.8	31.0	10.3	20.7	0.0	13.8	3.5	6.9																				
비중(후기)	3.3	30.0	30.0	16.7	3.3	10.0	3.3	3.3																				
4) 시기별 군집 비중 시각화																												
5) 특성 변화 분석	군집2가 전기에 10.3%에서 후기에 30.0%로 급증하였다.																											

[그림 3] 글쓰기 유형의 시기별 변화 파악 과정

한 정보를 토대로, 군집 소속 논설에 시간 정보를 대입하여 시기별 변화를 살펴본 것이다.

3)번 표와 4)번 그래프에서 볼 수 있듯이 군집1, 군집2, 군집3을 망라한 지배적 글쓰기 유형의 비중은 전기의 62.0%에서 후기의 76.7% 수준으로 크게 높아졌다. 그리고 그 상승을 주도한 것은 군집2(상세 주석·자료형)로, 약 20%의 급상승을 보였다. 물론 군집1, 군집3에서 이병도와 조윤제도 후기에 들어서 수 편의 논설을 게재하였다. 그러므로 이런 비중 변화가 진단학회 회원들의 세대 변화와 관련되는지, 혹은 전시 체제기라는 시대 상황과 맞물려 일시적인 변화에 불과했는지 속단하긴 이르다. 그렇지만 특정 글쓰기 유형이 뚜렷이 증가한 것은 해방 후 이 학술지의 동향 등과 관련해서 흥미롭게 추적해 봐야 할 지점이라 생각된다. 실제로 [표 2]와 대조해 보면, 고유섭, 김상기, 도유호가 군집2의 논설 9편 중 7편을 후기에 게재하여 이러한 비중 증가와 부합한다.

5. 결론

본 논문은 일본 식민지 시기 『진단학보』 논설 59편을 전산 코퍼스로 구축한 뒤, 이를 ‘장절 구분’, ‘논설 규모’, ‘자료 규모’, ‘주석 수’, ‘주석 규모’, ‘주석 평균 길이’라는 6개 지표를 특성으로 활용하여 비슷한 것끼리 군집화하였다. 최적 군집을 8개로 산출하였고, 이런 구분에는 ‘주석 평균 길이’, ‘장절 구분’, ‘주석 수’의 3개 요소가 크게 개입하였다.

8개 군집의 글쓰기 유형은 이 세 개 요소를 반영해서 파악하였는데, 이 중에서 다음 세 유형은 소속 논설의 개수와 특징 면에서 당시의 지배적인 글쓰기 유형이라 말할 만하다. ‘간결 주석·자료형’(군집1), ‘상세 주석·자료형’(군집2), ‘간결 기고형’(군집3)이 그것이다.

『진단학보』 주요 필진을 이 지배적 유형과 관련지어 살펴보니 양자 사

이에 뚜렷한 성향의 차이가 드러났다. 이병도와 조윤제의 논설이 간결 주석·자료형과 간결 기고형 글쓰기에 양분된 반면, 고유섭, 김상기, 도유호의 논설은 상세 주석·자료형에 집중되었다. 나아가 이러한 지배적 유형의 글쓰기는 그 비중이 후기로 갈수록 높아졌으며, 상세 주석·자료형이 그러한 상승을 주도했음을 확인할 수 있었다.

본 논문에서는 일본 식민지 시기 『진단학보』 수록 논설을 ‘근대 학술 규범’의 초기적 양상 차원에서 포착하기 위해 주요 지표 설정 및 군집화(k-means), 시각화(t-SNE), 특성기여도 분석(의사결정나무), 통계적 대푯값(상자그림) 등의 정량적 분석 기법을 활용하였다. 그 결과 이 학술지 내에서 몇 개의 지배적 글쓰기 경향이 존재하는 가운데 그보다 더 다양한 글쓰기 방식이 병존하고 있었음을 발견할 수 있었다.

다만, 이러한 양상과 변화를 심도 있게 파악하려면 내용 파악과 맥락에 대한 고려를 결합해야 하는데 그러한 작업까지는 나아가지 못했다. 또한 형태적 분석이라고 해도 분석 대상을 『진단학보』 한 가지로 한정했으므로, 다른 학술지와와의 공통점과 차이점을 입체적으로 파악할 수 없었다.

이 점에서 본 논문의 결론은 잠정적이며 시론적인 차원에 머문다. 이런 발견을 심화, 확충하려면 동 시기 일본인 학자들이 발간한 『청구논총』 및 관제 학술지, 그리고 해방 후의 『진단학보』 등으로 비교 대상을 확장할 필요가 있다. 또한 본 논문의 형태 분석을, 데이터 전처리와 토픽모델링을 통한 내용 분석과 통합할 필요가 있다. 이러한 작업을 통해 근대에서 현대에 이르는 근대 학술 규범의 형성 및 변화 양상을 규명할 수 있기를 기대한다.

참고문헌

震檀學報(원문)²⁸

진단학보 - North Korea Humanities²⁹

진단학회 웹사이트³⁰

- 국사편찬위원회, 「실증주의 사학」, 우리역사넷(열람 2025-05-04)의 '내용' 단락
 민현구(2012), 「두계 이병도의 수학과정과 초기 학술활동」, 『진단학보』 116.
 정병준(2016), 「식민지 관제 역사학과 근대 학문으로서의 한국역사학의 태동: 진단학회
 를 중심으로」, 『사회와 역사』 110.
 조범성(2021), 「1930년대 청구학회의 설립과 활동」, 『한국민족운동사연구』 107.
 홍종욱(2019), 「실증사학의 '이념': 식민지 조선에 온 역사주의」, 『인문논총』 76-3.
 홍종욱·장문석·류준필(2025), 「한국 근대문헌 디지털 아카이브의 설계: 『진단학보』를
 중심으로」, 『인문연구』 110.

원고 접수일: 2025년 5월 8일, 심사완료일: 2025년 5월 27일, 게재 확정일: 2025년 5월 27일

28 <https://drive.google.com/drive/folders/18VTPjpPpAC51TewL2syGIMPT7BxFHuM9>(검
 색일: 2024. 11. 28.).

29 [https://dh.aks.ac.kr/~nkh/wiki/index.php/%EC%A7%84%EB%8B%A8%ED%95%99%
 EB%B3%B4](https://dh.aks.ac.kr/~nkh/wiki/index.php/%EC%A7%84%EB%8B%A8%ED%95%99%EB%B3%B4)(검색일: 2024. 11. 28.).

30 <https://chintan.jams.or.kr/co/main/jmMain.kci>(검색일: 2024. 11. 28.).

부록

[부록표 1] 진단학보 수록 논설의 분포 (필자별, 시기별)

	Ho	1	2	3	4	5	6	7	8	9	10	11	12	13	14	77	59	
1	고유섭	-	-	1	-	-	1	-	1	-	1	-	-	-	1	5	4	고유섭
2	김두헌	-	1	-	-	1	-	-	-	-	-	1	-	-	-	3	1	김두헌
3	김상기	1	1	-	-	1	-	1	-	1	1	-	-	1	-	7	6	김상기
4	김석형	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	김석형
5	김영건	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	1	김영건
6	김영수	-	-	-	-	-	-	-	1	1	-	1	-	-	-	3	2	김영수
7	도유호	-	-	-	-	-	-	-	1	-	-	-	1	1	1	4	3	도유호
8	박시형	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	박시형
9	손진태	1	-	1	1	1	-	-	-	-	-	-	-	1	-	5	4	손진태
10	송석하	1	1	-	-	-	-	-	-	-	-	-	-	-	-	2	2	송석하
11	양주동	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	1	양주동
12	유홍렬	-	-	-	-	1	-	-	-	1	-	-	-	-	-	2	2	유홍렬
13	이병기	1	-	-	1	-	1	1	-	-	-	-	-	-	-	4	0	이병기
14	이병도	2	1	1	2	1	3	1	2	1	2	1	-	-	-	17	17	이병도
15	이상백	1	1	1	1	1	-	1	-	-	-	-	-	-	-	6	5	이상백
16	이승녕	-	1	-	-	-	-	-	-	-	-	1	1	-	-	3	2	이승녕
17	이인영	-	-	-	-	-	-	-	1	-	1	-	1	1	-	4	3	이인영
18	이중화	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	0	이중화
19	조윤제	1	-	1	-	-	1	-	-	-	1	1	1	-	-	6	4	조윤제
20	한홍수	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	0	한홍수
	77	8	6	6	5	6	7	4	6	4	8	5	4	4	4			
	59	7	4	3	4	5	4	2	5	4	8	4	4	2	3			

[부록표 2] 분석 대상으로 선정한 진단학보 59개 논설의 서지 사항

No	GisaId	Title	VolI	No	GisaId	Title	VolI
1	이병도(jdh19341a)	三韓問題의 新考察(一)-辰國及三韓考-	1	31	이병도(jdh193711a)	三韓問題의 新考察(七完)-辰國及三韓考-	8
2	이상백(jdh193411)	戊辰 産件의 淵源에 對한 一問題	1	32	고유섭(jdh193711)	佛敎가 高麗藝術 敎育에 끼친 影響의 一考察	8
3	조운재(jdh193411)	朝鮮詩歌의 胎生	1	33	김영수(jdh193711)	五敎兩宗에 對하야	8
4	김상기(jdh193411)	古代의 貿易形態와 羅末의 海上發展에 就하야(一)-淸海嶺大使	1	34	이병도(jdh193711b)	世宗朝의 國都上山問題	8
5	손진태(jdh193411)	朝鮮 古代 山神의 性에 就하야	1	35	김상기(jdh193807)	三別抄와 그의 亂에 就하야(1)	9
6	송석하(jdh193411)	風神考(附末平考)	1	36	이병도(jdh193807)	李朝初期의 建都問題	9
7	이병도(jdh193411b)	震樓解	1	37	유홍렬(jdh193807)	朝鮮誓約의 成立	9
8	이상백(jdh193504)	三韓人物考(一)	2	38	김영수(jdh193807)	曹溪禪宗에 就하야 - 五敎兩宗의 一派, 朝鮮佛敎의 根源-	9
9	송석하(jdh193504)	窟容舞, 禮禮, 山峯廟의 關係를 論함	2	39	이병도(jdh193904a)	國號에 對한 一,二의 考察(1)	10
10	김상기(jdh193504)	古代의 貿易形態와 羅末의 海上發展에 就하야(二)-淸海嶺大使	2	40	김상기(jdh193904)	三別抄와 그의 亂에 就하야(2)	10
11	이병도(jdh193504a)	高麗朴氏에게 贈함-附高麗朴氏論考-	2	41	고유섭(jdh193904)	朝鮮塔婆 研究(2)	10
12	손진태(jdh193509)	支那民族의 雜處信仰과 그 傳説	3	42	조운재(jdh193904)	李朝學人의 朝鮮文學에 對한 態度	10
13	이병도(jdh193509a)	三韓問題의 新考察(二)-辰國及三韓考-	3	43	양주홍(jdh193904)	鄉歌注釋釋稿 - 上代語法에 關한 若干의 基本的 見解 -	10
14	고유섭(jdh193509)	高麗畫跡 對하야	3	44	이인영(jdh193904)	中世-의 建州紀程圖記에 對하야 - 最近發見의 淸初史料 -	10
15	이상백(jdh193604)	李朝建國의 研究(一)	4	45	이병도(jdh193904b)	廣州風情里土成과 百濟時代의 蛇城	10
16	이병도(jdh193604a)	三韓問題의 新考察(三)-辰國及三韓考-	4	46	김영전(jdh193904)	安南普陀山名考	10
17	손진태(jdh193604)	中華民族의 說에 關한 信仰과 學說	4	47	이승녕(jdh193912)	朝鮮語異化作用에 對하야	11
18	이병도(jdh193604b)	徐花潭及李蓬坊에 對한 小考-朝鮮儒學史의 一節-	4	48	김두현(jdh193912)	朝鮮製史制小考	11
19	이상백(jdh193607)	李朝建國의 研究(二)-羅末의 政爭과 李成桂-續	5	49	조운재(jdh193912)	春香傳異本考(1)	11
20	이병도(jdh193607)	三韓問題의 新考察(四)-辰國及三韓考-	5	50	이병도(jdh193912)	廣州夢村土城址 - 百濟時代의 城營址 -	11
21	유홍렬(jdh193607)	朝鮮詞廟發生에 對한 一考察 - 特히 羅末李朝初의 根本崇賢思	5	51	이승녕(jdh194009)	「」音攷	12
22	손진태(jdh193607)	中華民族의 說에 關한 信仰과 學說(二)	5	52	조운재(jdh194009)	春香傳異本考(二)	12
23	김상기(jdh193607)	葛文王考	5	53	도유호(jdh194009a)	中國都市文化의 起源	12
24	고유섭(jdh193611)	朝鮮塔婆의 研究	6	54	이인영(jdh194009)	太平通載殘卷 小考 - 特히 新羅殊異傳 逸文에 對하야 -	12
25	이병도(jdh193611a)	三韓問題의 新考察(五)-辰國及三韓考-	6	55	김상기(jdh194103)	三別抄와 그의 亂에 就하야(完)	13
26	이병도(jdh193611b)	李時廟와 그 學問-朝鮮儒學史의 一節-	6	56	도유호(jdh194103)	中國都市文化의 起源(二)	13
27	이병도(jdh193611c)	高麗文宗王과 長壽寺創構	6	57	김석형(jdh194106)	李朝初期 國役編成의 基礎	14
28	이병도(jdh193707)	三韓問題의 新考察(六)-辰國及三韓考-	7	58	박시형(jdh194106)	李朝田稅制度의 成立過程	14
29	이상백(jdh193707)	李朝建國의 研究(三)-羅末의 政爭과 李成桂-	7	59	도유호(jdh194106)	中國都市文化의 起源(三完)	14
30	이인영(jdh193711)	鮮初女眞貿易考	8				

진단학보 수록 논설 DB 내용 및 구조

파일명: JDH_Data_new.xlsx

파일 내 시트 이름: Term59, Ph59, Foot59, RowInfo

1) Term59 시트: “고유번호 - TermId” 대응표(RawID-GisaId).

- 논문별 id 부여: 필자와 논술 발표 시점을 잘 드러내기 위함임(타 학술지와 의 비교 위해 jdh도 표시).

- TermId: 필자 + jdh + 발행연월 (+ a, b...)

ex) 고유섭(jdh193509) → ‘고유섭(193509)’(간략표기)

- 동일 필자가 동일 호수에 2개 이상 논술을 작성한 경우 TermId에 ‘a’, ‘b’ ... 를 부착해서 구별함.

- 관계형 데이터베이스 구축(RDB) 시에 주석 정보, 본문 인용자료 등과의 연결고리로 활용함.

2) 'Ph59' 시트

- 전문한 html 파일에서 제목, 쪽수, 목차, 본문, 주석정보 수집 후 정제함.
 - ① 본문 내 주석 위치 정보, ② 본문과 주석의 입력 내용이 원문과 일치하는지 여부 등을 확인함.
 - 필요한 경우 원문에서 근사치로 글자수를 계산, 입력 후 글자수 집계 시 포함함.
- 목차 정보 입력: 맨 앞에 목차 제시한 논문은 맨 앞에 해당 목차 입력함.
- 3,802개 행(raw): 각 행은 대체로 문단(phrase) 단위임. 그러나 입력물별로 동일하지 않음(문장 등).

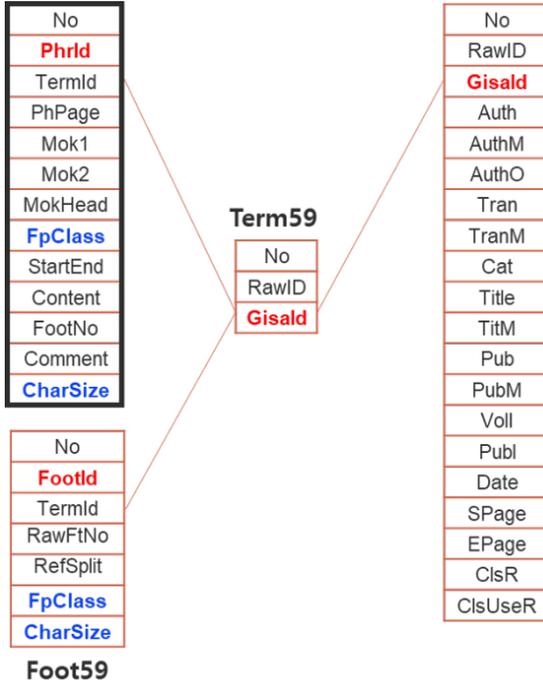
3) 'Foot59' 시트

- 각 주석을 행(row)으로 삼고 고유번호 부여. TermId를 연결시킴. 원문 주석만 포함(+ 附註, 附記).
- FpClass열에 위치 정보 삽입: FpAll은 주석이 본문 뒤에 모여 있을 경우 이고 Fp-1, Fp-2 등은 각 장/절에 분산될 때임.
- CharSize열: 주석이 일부 텍스트화에서 누락된 경우 소량은 직접 입력함. 다량은 근사치로 계산하여 글자 수 입력함(1줄 = 50자).

4) 기타 시트

- RowInfo: 진단학보 수록 기사(85개)의 기초 정보, 59개 논설 선별 등 표시함(ClsUseR열에 'UR').
- Pub77(59): 1-14호 발행 상황을 호별, 필자별로 표시함.

5) 이상의 시트로 이루어진 DB 구조는 다음과 같다.



ABSTRACT

The Formation and Diversity of Modern Academic Norms during the Japanese Colonial Period

Hur, Soo*

A Quantitative Analysis of the Formal
Aspects of Articles in *Chin-Tan Hakpo*

Chin-Tan Hakpo is a representative academic journal published in Korea during the Japanese colonial period. In academia, this journal has been primarily evaluated as a center for positivist history. However, recent studies have pointed to the diversity in the composition of the Chin-Tan Society and the scholarly nature of *Chin-Tan Hakpo*. This study examines the existence and manifestation of this diversity by analyzing the articles in the journal from a formal perspective, using a quantitative method.

First, a digital corpus of 59 articles from *Chin-Tan Hakpo* was constructed in a database format, and six indicators were presented to identify diversity. These are: 'chapter/section division,' 'article length,' 'data volume,' 'number of footnotes,' 'footnote length,' and 'average footnote length.' Second, using these six indicators as features, the 59 articles were grouped into clusters of similar items. Eight clusters were obtained using

the K-means method. Each cluster revealed distinct characteristics by reflecting the six features. Third, focusing on the larger and more distinct clusters, this study examined their relationship with major authors and the changes in their proportion over time. These clusters include the ‘concise-footnote academic article,’ the ‘detailed-footnote academic article,’ and the ‘concise contributed article.’ It appears that while various writing styles coexisted, these three types were representative of the writing styles of the period. Among them, the proportion of ‘detailed-footnote academic articles’ showed a significant increasing trend in the later period compared to the early period.

Keywords *Chin-Tan Hakpo*, Quantitative and Formal Analysis, Chapter/Section Division, Data, Footnotes, Cluster Analysis

