

# 『동유학안』의 디지털 표상 연구: 중층 귀속의 표상과 TEI-XML·RDF 1.2 연계

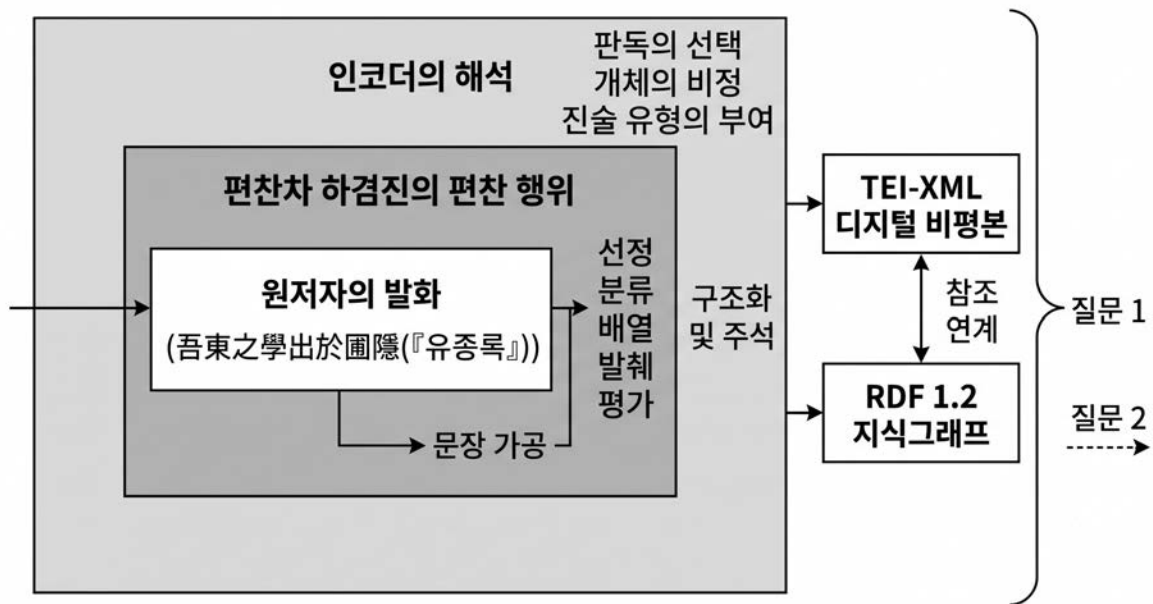
## I. 문제 설정: 편찬된 지식 체계의 표상

### 1. 중심 질문

『동유학안(東儒學案)』은 하검진(河謙鎭, 1870-1946)이 1938년부터 1943년에 걸쳐 편찬한 조선 유학사 저작이다. 이 책은 조선의 유학자를 학파 단위로 나누어 편(篇)을 나누고, 각 편(篇)의 첫머리에 그 학파의 연원과 성격을 규정하는 총론을 둔다. 그 아래에는 인물마다 생애와 학문을 서술하고, 일부 인물의 경우 해당 인물의 저술에서 발췌한 글을 덧붙인다. 따라서 이 책에서 한 문장을 임의로 택하여 그것이 누구의 말인가를 물으면 답은 하나로 돌아오지 않는다. 어떤 문장은 편찬자가 다른 문헌에서 옮겨 온 인용이고, 어떤 문장은 그 인용에 곧바로 이어지는 편찬자 자신의 평가이다. 두 종류의 문장 사이에 이런 경계를 알리는 표지가 없는 경우도 많다. 나아가 어떤 인물이 어느 학파에 놓였는가, 어떤 인물에게만 저술이 부기되었는가, 어떤 인물이 아예 수록되지 않았는가는 문헌에 직접 기록된 사실은 아니지만 해당 문헌을 특징짓는 요소가 된다. 요컨대 『동유학안』은 인용된 원저자의 진술과 편찬자 하검진의 편찬적 결정, 곧 인물의 선정과 학파의 분류, 저술의 발췌와 배치, 학문적 평가가 중첩된 지식 체계이다. 이 지식 체계를 디지털로 표상하는 순간 인코더의 해석이 또 하나의 층위로 더해진다.

이러한 점에서 본 연구의 두 질문이 도출된다. 첫째, 이러한 중층적 귀속 관계를 확정된 사실(fact)이 아닌, 출처 환원과 반증이 가능한 진술(statement)로서 표상하려면, TEI-XML 디지털 판본과 RDF 1.2 지식그래프를 어떠한 원칙으로 각각 구축하고 연계해야 하는가. 둘째, 그렇게 구축된 데이터는 이 문헌의 학파 분류와 학술사 서술에 관하여 기존의 독해가 도달하기 어려웠던 어떠한 관찰을 가능하게 하는가.

두 질문의 비중은 같지 않다. 첫째 질문이 본 연구의 본체이며, 논문의 대부분은 이 질문에 대한 답, 곧 표상의 원칙과 그 원칙을 구현하는 설계를 수립하는 데 할애될 것이다. 둘째 질문은 그 답이 실제로 무엇을 산출하는지를 보이는, 그 자체로 더 본격적인 인문학적 탐구가 될 것이다. 그러나 이 질문의 달성 수준은 구축되는 자료의 규모에 직접 좌우되며, 본 연구자가 학위논문의 기간 안에 구축할 수 있는 규모에는 명백한 한계가 있다. 따라서 본 연구는 둘째 질문에 대하여 관찰의 가능성을 설계 수준에서 확보하고 제한된 범위에서 시연하는 데까지를 목표로 삼으려 한다.



[도해 1] 세 귀속 층위와 두 질문

이하의 논의는 다음 순서로 진행된다. 2절은 위 명제의 첫째 전제, 곧 『동유학안』이 인물 정보의 저장소가 아니라 편찬된 지식 체계라는 점을 밝힌다. 3절은 둘째 전제, 곧 기존의 인물 데이터 구축 방식이 그러한 자료 앞에서 무엇을 놓치는가를 밝힌다. 4절은 현 상황을 정리하고 다음 단계의 작업을 구상한다.

## 2. 편찬된 지식 체계로서의 『동유학안』

학안(學案)은 학자의 전기와 해당 학자의 저술 발체문을 결합한 문헌 형식으로, 16세기 후반 명말에 등장하여 20세기까지 편찬이 이어진 장르이다(Chu 2007, 244쪽). 전기와 발체의 결합이 이 형식을 인접 장르와 구분하는 기준이며(Chu 2007, 247쪽), 여기에 학파 단위의 편차(篇次)와 편찬자의 평의(評議)가 더해진다. 이 형식의 전범을 확립한 것은 명말청초 황종희(黃宗羲)의 『명유학안(明儒學案)』이었다. 이후 황종희, 전조망(全祖望) 등의 『송원학안(宋元學案)』과 당감(唐鑑)의 『국조학안소지(國朝學案小識)』를 거쳐 20세기에까지 이어진 계보 위에 하검진의 『동유학안』이 위치한다. 총론, 인물 생애 서술, 저술 발체라는 삼중의 체제는 하검진 개인의 창안이 아니라 이 계보에서 비롯된 것이다.

학안은 근대적 의미의 학술사로 기획된 문헌이 아니다. 전통적 서지 분류에서도 학안은 역사서와 철학서 사이에서 일관된 자리를 얻지 못하였으며, 전기와 철학적 발체를 결합한 혼합적 형식으로서 학술사인 동시에 수양의 지침으로 읽혔다(Chu 2007, 246-248쪽). 이러한 사정은 학안을 사실의 목록으로 환원하려는 접근이 왜 잘 작동하지 않는지를 설명해 준다. 이 형식이 독자에게 제시하려는 것은 어떤 학자가 어떠하였다는 정보가 아니라, 선택된 문헌 증거를 통하여 하나의 학문이 어떤 경로로 성립하였는가를 보이는 사례이기 때문이다.

학안 형식이 요구하는 편찬 행위는 다섯 가지로 구분된다. 누구를 수록할 것인가라는 선정의 문제, 수록된 인물을 어느 학파에 배속할 것인가라는 분류의 문제, 편을 어떤 순서로 놓을 것인가라는 배열의 문제, 그 인물의 저술 가운데 무엇을 얼마나 옮길 것인가라는 발체의 문제, 그리고 총론과 안어(按語)를 통하여 학문적 성격을 규정하는 평가의 문제이다. 이 중 어느 것도 문헌 외부에서 주어지지 않으며, 모두 편찬자가 스스로 내린 판단에 따른다. 이러한 성격은 신유학의 계보 문헌 일반을 유교 전통의 수동적 기록이 아니라 특정 시점의 목적에 따라 능동적으로 구성된 역사적 표상으로 읽어야 한다고 논한 선구적 연구에서 이미 제시된 바 있다(Wilson 1994, 4-6쪽).

편찬 행위의 흔적은 문헌 자체에서도 확인된다. 『동유학안』에 수록된 인물 가운데 저술의 발체가 함께 실린 경우는 일부에 그치며, 발체의 유무와 분량은 인물마다 다르다. 이 차등은 명시적인 평어 없이도 편찬자의 비중 부여를 드러낸다. 어떤 인물에게는 여러 편의 저술이 부기되고 어떤 인물에게는 한 편도 부기되지 않는다면, 그 차이는 자료의 유무만으로 설명되지 않으며 편찬자가 무엇을 이 학술사의 실질로 보았는가를 가리킨다. 미수록 역시 같은

성격을 지닌다. 어떤 유학자가 학안에 수록되지 않았다는 사실은 그 인물의 부재를 뜻하는 것이 아니라 편찬 기준이 작동한 결과인 것이다.

선행 연구가 이 문헌의 학문적 객관성을 문제 삼아 온 것도 같은 사정에서 비롯된다. 하겸진의 『동유학안』이 이학 중심의 도통론과 도산학파 중심성을 갖는다는 지적이 『동유학안』 연구의 초기부터 있어왔다(최일범 2004, 138-139쪽). 그러나 동시에 이 문헌의 가치가 조선 유학의 객관적 학파 체계를 제공하는 데 있지 않고 당시 영남 유림이 조선 유학의 학맥을 어떻게 인식하고 구성하였는가를 보여주는 데 있다고 결론지었다는 점은(최일범 2004, 139쪽), 이 문헌을 읽는 초점이 이미 사실의 정확성에서 구성의 방식으로 옮겨져 있었음을 뜻한다. 본 연구는 이 전환을 이어받는다. 주관성이라 불린 것은 곧 편찬자의 위치와 정통관이 선정과 분류와 배열과 발체와 평가를 규정한 과정이며, 바로 그 과정이 표상의 대상이다.

여기서 본 연구의 대상이 확정된다. 이 자료를 데이터로 만든다는 것은 조선 유학자에 관한 사실을 추출하여 정리하는 일이 아니다. 인물의 생몰년이나 관력이나 사승 관계만을 항목화하여 표로 옮기면, 남는 것은 하겸진이 참조한 자료에서 이미 얻을 수 있는 정보이고 사라지는 것은 하겸진이 그 정보를 가지고 무엇을 하였는가이다. 본 연구가 표상하려는 것은 후자, 곧 편찬자가 구성한 지식의 질서이다. 그리고 그 질서는 문헌의 어떤 항목에 직접 기록되어 있는 것이 아니라 선정과 분류와 배열과 발체와 평가라는 행위들의 총체로만 존재한다.

### 3. 기존 인물 데이터 구축 방식의 한계

편찬된 지식 체계를 대상으로 삼을 때 기존의 인물 데이터 구축 방식이 겪는 곤란은 세 가지로 정리된다. 『동유학안』의 한 구절을 예시로 살펴 보자. 제2편 이학연원제유학안의 첫 인물 항목인 김종직(金宗直) 조는 그를 사예 김숙자(金叔滋)의 아들로 소개한 뒤 "先生始受學於家庭"이라 서술한다. 원문은 가계에 관한 정보와 수학에 관한 정보를 별개의 문장으로 제시하며, 家庭이 누구를 가리키는지는 밝히지 않는다(하겸진 2008, 원문편 477쪽).

첫째는 사실화이다. 인물의 사승 관계를 정리하기 위해 위 구절을 데이터로 옮기려면 인코더는 "家庭"을 김숙자로 비정해야 하고, 그 결과 김종직이 김숙자에게 배웠다는 관계가 기록된다. 이 기록은 그 자체로 무리한 판단이 아니다. 앞 문장이 김숙자를 부친으로 소개하고 있으므로 문맥상의 근거도 갖추어져 있다. 문제는 일단 기록되고 나면 하겸진이 서술한 바와 인코더가 비정한 바가 구분되지 않는다는 데 있다. 문헌에는 家庭이라는 두 글자가 있을 뿐이고 인물의 지목은 인코더의 독해인데, 데이터에서는 양자가 동일한 지위의 진술로 나타난다. 진술이 사실로 고착되는 이 전환은 개별 인코더의 부주의에서 오는 것이 아니라 표상의 방식이 만들어 내는 결과이며, 데이터가 재사용되고 다른 데이터와 연결될수록 원래의 유보는 회복되기 어려워진다.

둘째는 탈맥락화이다. 구조화된 데이터는 자신이 원문의 어느 자리에서 왔는지를 스스로 밝히지 않는다. 위 관계가 어느 편의 어느 항목의 어느 구절에 근거하는지가 데이터에 남지 않으면, 이 기록이 옳은지를 검증하려는 독자는 문헌으로 돌아갈 길을 얻지 못한다. 되돌아갈 수 없는 기록은 반증될 수 없고, 반증될 수 없는 기록은 학술적 주장의 자격을 갖추지 못한다. 출처가 데이터의 부속 정보가 아니라 성립 조건인 이유가 여기에 있다. 더구나 이 문헌의 경우 되돌아가야 할 지점은 문헌의 이름이나 권차가 아니라 특정 구절이어야 한다. 한 인물 항목 안에서도 서술의 성격이 문장마다 바뀌기 때문이다.

셋째는 중층 귀속이다. 앞의 두 문제가 사료 일반을 데이터화할 때 발생하는 것이라면, 이 문제는 학안류가 추가로 제기하는 것이다. 학안의 서술에서 편찬자는 자신의 문장을 쓰는 동시에 다른 사람의 문장을 옮기며, 옮겨 온 문장에 자신의 판정을 덧붙여 놓는다. 따라서 하나의 진술을 기록하려면 그것이 원저작의 발화인지, 그 발화를 옮긴 편찬자의 행위인지, 그 행위를 그렇게 읽은 인코더의 판단인지를 함께 적어야 한다. 귀속의 연쇄가 삼중 이상으로 겹치는 이 조건에서, 하나의 진술에 하나의 주체만을 적을 수 있는 구조는 반드시 어느 층위인가를 지운다. 게다가 인용 여부를 명확히 표지하지 않은 채 인용하는 경우가 많다는 것이 상황을 더 어렵게 만든다. 서명과 문장이 함께 주어진 인용이 있는가 하면 인명만 주어진 인용이 있고, 아무 표지 없이 흘러가는 서술도 있다. 근대적 참고문헌 관행이 존재하지 않는 조건에서 인용을 유형화하는 문제는 조선시대 경전 주석을 대상으로 출처 표시의 정밀도를 세 등급으로 구분한 작업에서 이미 실질적으로 다루어진 바 있으며(김서윤 2024a), 이 유형화는 학안의 서술에도 유효한 좌표를 제공한다.

세 문제가 선행 연구에서 전혀 다루어지지 않은 것은 아니다. 인물 데이터베이스의 설계에서 사실을 직접 기록하는 대신 어느 사료가 무엇을 말한다는 진술 자체를 기록의 단위로 삼자는 팩토이드(factoid) 발상은 집단전기학의 디지털 전환 과정에서 이미 제시되었고(Bradley and Short 2005, 7-8쪽), 디지털 판본 학계에서도 편집자의 해석을 판본의 부속 해설이 아니라 기계 가독적 주장으로 판본 안에 담아야 한다는 논의가 제출되었다(Vogeler 2019, 309-310쪽). 국내에서도 역사적 사실을 기술할 때 그것이 어느 사료와 어느 연구자의 해석에 기반하는가를 명기하는 출처와 확실성의 표기가 핵심 난제로 제기된 바 있다(김바로·이재열 2025, 1쪽, 12-13쪽). 본 연구는 이러한 축적 위에서 출발한다. 다만 이 논의들이 상정한 것은 대체로 사료의 진술과 연구자의 해석이라는 두 층위였고, 사료 안에 이미 편찬자의 판단이 내재해 있다는 조건은 별도로 다루어지지 않았다. 각각의 기여가 어디까지 미치고 학안류의 조건 앞에서 무엇이 남는가에 대한 검토는 III부에서 더 자세히 수행할 것이다.

세 문제에 대응하는 기준은 두 가지로 요약된다. 하나는 출처 환원성이다. 데이터의 모든 진술은 그것이 유래한 판본의 구체적 위치로 되돌아갈 수 있어야 한다. 다른 하나는 해석 책임성이다. 데이터이터에 기록된 모든 판단은 그것을 내린 주체가 누구인지, 곧 원저자인지 편찬자인지 인코더인지가 함께 기록되어야 한다. 이 둘은 개별적인 설계 지침이 아니라 이하의 모든 설계 결정이 봉사해야 할 목표이며, 본 연구의 성패를 판정하는 기준이기도 하다.

## 4. 연구의 성격과 대상 문헌의 적합성

본 연구는 하경진 『동유학안』의 전질을 완전한 디지털 판본 및 지식 그래프로 구축하는 것을 목표로 삼지 않는다. 물론 구축은 수행된다. 그러나 논문이 탐구하려는 것은 특정 문헌의 데이터가 만들어졌다는 사실이 아니라, 편찬된 지식 체계를 확정된 사실로 환원하지 않으면서 표상하는 방법에 관한 것이다. 이 구분은 실질적인 귀결을 낳는다. 방법의 성립을 보이는 데 필요한 것은 자료의 총량이 아니라 방법이 다루어야 할 문제의 유형을 모두 포함하는 구간이므로, 인코딩의 범위는 문헌 전체가 아니라 선정된 일부로 한정된다. 반대로 전량의 인코딩을 목표로 세우면 유형의 정밀한 처리보다 처리량의 확보가 앞서게 되어 연구의 본래 목적과 어긋나게 된다. 원본 3편, 증보 번역본 6편이라는 『동유학안』의 분량을 고려할 때 박사 학위 과정 내에 완전한 디지털 비판정본과 지식 그래프를 구축하는 것은 불가능에 가깝다. 다만 IV장에서 간단히 언급하겠듯이 생성형 AI의 활용이 이 야심적 목표를 실현 가능케 할 수 있을 지 모르나, 현 단계에서는 아직 이 문제에 대해 본격적으로 검토하지 못했다.

그렇다면 왜 이 문헌인가. 세 가지 특성 면에서 이 자료는 중층 귀속 모델의 시험대로 적합하다. 첫째는 총론과 인물 항목과 저술 발체라는 삼중 구조이다. 총론은 학파의 성격을 규정하는 상위의 틀이고, 인물 항목은 개별 생애와 학문의 서술이며, 발체는 원저작의 문장 그 자체이다. 세 층위는 발화의 성격이 서로 다르므로 하나의 평면으로 압축하여 기록할 수 없다. 둘째는 인용 중심의 체제이다. 이 형식은 편찬자가 직접 쓴 문장보다 다른 문헌에서 옮겨온 문장에 더 많은 지면을 할애하므로, 인용을 인물 정보에 딸린 부속 자료로 처리하는 설계로는 문헌의 대부분을 표상할 수 없다. 셋째는 편찬자의 강한 개입이다. 분류와 배열과 평가가 문헌의 구조 자체에 새겨져 있어 서술과 판단을 문장 단위로 분리하기 어렵다.

중요한 것은 이 세 특성이 하경진 개인의 창안이 아니라 학안이라는 형식이 공유하는 속성이라는 점이다(Chu 2007, 244쪽, 247쪽). 만약 세 특성이 한 편찬자의 개성에서 비롯된 것이라면 그로부터 도출되는 표상의 원칙은 한 문헌에 대한 처방에 그친다. 그러나 세 특성이 형식의 요구라면, 같은 형식을 지닌 다른 문헌에도 같은 원칙이 적용될 수 있다. 본 연구가 요구사항의 도출 근거를 개별 문헌의 관찰이 아니라 장르의 서술 형식에 두려는 이유가 여기에 있다.

이상이 첫째 질문을 다루기 위한 전제라면, 둘째 질문은 이번 발표에서 본격적으로 다루지 않는다. 데이터에 근거한 관찰은 그 데이터가 신뢰할 만한 층위 구분 위에서 구축된 뒤에야 의미를 가지며, 자료 조사가 아직 수행되지 않은 단계에서 관찰을 앞당기면 무엇을 관찰하였는가가 아니라 무엇을 관찰할 수 있도록 데이터를 만들었는가를 보이는 데 그친다.

두 질문이 끝내 분리되지 않는 이유도 여기에 있다. 표상의 방법은 자료를 해석하기 위한 수단이지만, 동시에 자료의 성격이 표상의 방법을 규정한다. 이 문헌이 편찬된 지식 체계라는 사실이 중층 귀속의 표상을 요구하고, 중층 귀속을 표상한 데이터라야 편찬 행위 자체를 관찰의 대상으로 삼을 수 있다. 방법이 대상을 만들고 대상이 방법을 요구하는 이 순환은 결함이 아니라 인문학 자료의 데이터화가 놓인 조건이며, 그 조건을 감추지 않고 설계의 표현

에 드러내는 것이 본 연구의 방침이다.

## 5. 연구 범위

본격적인 구축의 대상은 하겸진본 『동유학안』의 일부로 한정한다. 『동유학안』의 어느 부분을 대상으로 삼을 지에 대해서는 미정이다. 『명유학안』과 『송원학안』과 『국조학안소지』는 학안이라는 형식의 계보와 그 형식이 편찬 행위를 드러내는 방식을 확인하기 위한 참조 대상이고, 『도동연원록』과 『동유사우록』과 『패동연원록』은 조선의 계보 문헌 전통 안에서 이 문헌의 위치를 가늠하기 위한 대비항이다. 여기에 하나를 더 구분해 두어야 한다. 하겸진본과 거의 같은 시기에 편찬된 동명의 다른 저작, 곧 권오환본 『동유학안』이다. 위 문헌들은 현 상태에서는 어느 것도 데이터 구축의 대상이 아니며, 본격적인 비교는 학위 취득 이후의 과제로 미룬다.

저본과 산출물의 관계는 다음과 같이 정한다. 저본은 2008년 『증보 동유학안』의 표점(標點)된 한문 원문으로 하고, 1970년 간행본의 원문을 대교(對校) 자료로 삼으며, 두 판본에 실린 한국어 번역문은 참조 자료로만 활용한다. 인코딩 대상은 번역문이 아니라 한문 원문이다. 산출물은 저본을 그대로 옮긴 전사가 아니라 이 대조를 거친 자체 디지털 한문 비평본이다. 이러한 방침을 택하는 것은 어느 한 판본도 단독으로 인코딩의 근거가 되지 못하기 때문이다. 그 문헌학적 근거는 II부에서 논의할 것이다. 여기서 미리 밝혀 둘 것은 그 귀결이다. 두 판본 사이에서 독법을 선택하는 판단은 하겸진의 편찬 판단과도, 2008년 증보 단계의 판단과도 구분되는 별개의 편집 총위를 이룬다. 곧 인코더는 해석자이기 이전에 편집자이며, 이 사실 자체가 데이터에 기록되어야 할 조건이다.

산출물은 세 가지이다. TEI-XML로 작성된 디지털 한문 비평본, RDF 1.2로 기술된 지식그래프, 그리고 인코딩의 판단 기준을 명문화한 편집 지침이 그것이다. 마지막 산출물은 부속 문서가 아니라 해석 책임성을 담보하는 장치이다. 어떤 판단이 어떤 기준에 따라 내려졌는가가 문서로 남아야 다른 연구자가 같은 자료에 대하여 다른 판단을 제시할 수 있고, 그때 비로소 데이터가 논쟁의 대상이 될 수 있기 때문이다. 앞의 두 산출물의 관계에 대해서는 하나의 방침을 미리 세운다. 판본과 지식그래프를 하나의 그래프로 통합하지 않고, 각각을 자신의 표준 생태계 안에서 구축한 뒤 상호 참조를 통하여 연계한다는 것이다. 텍스트의 위계와 주장의 위계는 서로 다른 질서를 따르므로 어느 한쪽을 다른 쪽에 흡수시키면 흡수된 쪽의 고유한 구조가 손상된다. 이 방침에 대한 논증은 III부에서 수행한다.

끝으로 이번 발표의 위치를 밝힌다. 이번 발표는 완결된 연구의 보고가 아니라 착수 단계의 보고이며, 선행 방법론을 정리하고 검토하여 설계의 방향을 수립하는 데 목적이 있다. 판본의 확보와 저본의 확정, 그리고 표본 구간에 대한 예비적 분석까지는 수행되었으나, 문헌 전체를 대상으로 한 조사는 다음 단계의 과제로 남아 있다. 따라서 이하에서 제시되는 설계 결정은 대부분 잠정적인 것이며, 어느 지점이 확정되었고 어느 지점이 미결인지를 본문에서 그때마다 밝히려 한다. 인코딩 범위의 확대, 다른 계보 문헌으로의 확장, 계량적 분석의 본격적 수행은 학위 취득 이후의 과제로 구분한다.

## II. 자료: 『동유학안』의 문헌학적 고찰

### 1. 학안의 계보와 『동유학안』의 편찬 맥락

『동유학안(東儒學案)』을 데이터로 표상하려면 먼저 이 문헌이 어떤 장르에 속하며 그 장르가 어떤 방식으로 지식을 조직하는지를 확인해야 한다. 본 절은 중국 학안류의 전개, 조선의 연원록·사우록 전통, 동명이서(同名異書)와의 구분, 그리고 편찬자 하겸진(河謙鎭)의 위치와 편찬의 시기 및 동기를 차례로 검토한다. 이 검토를 관통하는 관점은 하나이다. 학안류와 연원록류는 과거 유학자에 관한 사실의 저장소가 아니라 편찬자의 선택, 분류, 배열, 인용, 평가로 구성된 지식 체계라는 것이다. 본 절은 이 관점을 근거로 삼아, 『동유학안』의 구조적 특성, 즉 총론, 인물 설명, 저술 발체의 삼중 구조가 찬자 하겸진의 개인적 특징에서 비롯된 것이 아니라 장르가 공유하는 속성임을 보이려 한다. 이를 확인함으로써 『동유학안』에서 도출되는 표상 모델의 요구사항이 개별 문헌을 넘어 학안체 일반으로 확장될 수 있다는 주장 역시 문헌학적 근거를 얻을 수 있을 것이다.

## 1) 중국 학안류의 전개와 체재의 성립

학안은 학자의 전기와 그 학자의 저술에서 선택한 발췌문을 결합한 문헌 형식으로, 16세기 후반 명말에 처음 등장하여 20세기까지 지속적으로 편찬된 장르이다(Chu 2007, 244쪽). '학안'이라는 명칭이 널리 유통된 계기는 1680년 무렵 간행된 황종희(黃宗羲)의 『명유학안』이었다. 이후 황종희가 착수하고 전조망(全祖望)이 보완한 『송원학안』이 1838년에 간행되고, 20세기에 들어서도 『양한삼국학안』과 『청유학안』이 이어졌다(Chu 2007, 244-245쪽). 이 계보는 하겸진의 1938-1943년 편찬이 고립된 사건이 아니라 동아시아 학안 편찬의 장기적 흐름에 속함을 보여준다. 학안에서 '안(案)'이라는 글자는 문서철, 불교의 공안에 대응되는 유교적 공안, 판례, 학습 계획이라는 네 가지 뜻을 지닌다. 이를 종합할 때 학안(學案)은 학문의 사례로 규정할 수 있다(Chu 2007, 266-267쪽). 이 규정에 따르면 학안의 한 인물 항목은 그 인물에 관한 사실의 총합이 아니라 편찬자가 선택한 문헌 증거를 통해 특정한 학문의 경로를 제시하는 사례로 볼 수 있다.

이때 두 가지 단서를 함께 확인해 둘 필요가 있다. 첫째, 학안에는 모든 저술에 공통되는 엄격한 표준 체제가 존재하지 않았다. 단지 황종희의 『명유학안』이 강력한 전범을 제공했을 뿐이다(Chu 2007, 245쪽). 따라서 『동유학안』의 삼중 구조는 장르 보편의 고정 형식이 아니라 『명유학안』 계통의 전범을 따른 편찬자의 선택에 해당한다. 둘째, 학안은 근대적 의미의 학술사로 기획된 것이 아니었다. 청대의 서지 분류에서도 학안은 사부(史部)와 자부(子部) 사이에서 일관되게 자리 잡지 못했으며, 전기와 철학적 발췌를 결합한 혼합 장르로서 역사서인 동시에 수양을 위한 지침서로 읽혔다(Chu 2007, 246-248쪽). 학안을 근대적 학술사의 미성숙한 전 단계로 읽는 통설적 서사가 조정을 요하는 이유가 여기에 있다.

중국의 학안 전통 중 『동유학안』에 영향을 끼친 세 편의 대표적 학안으로 『명유학안』, 『송유학안』, 『국조학안소지』를 꼽을 수 있다. 이 세 편의 대표적 학안은 각각 다른 지점에서 편찬적 특징을 드러낸다. 첫째, 『명유학안』은 분류의 기준을 확립했다. 이 저작은 학자들을 사승관계에 따라 학파별로 분류하고, 각 문파와 학자의 특징을 설명하며, 생애와 저술과 어록을 소개한 뒤 편찬자의 설명과 평가를 덧붙이는 체제로 구성된다(김태년 2010, 46쪽). 다만 범례에 따르면 학파 구분이 전승 관계에 근거하되 모든 학문적 성취가 계보를 통해서만 성립하는 것은 아니었으며, 독자적으로 학문을 이룬 인물을 별도로 수용하였다(Chu 2007, 260쪽). 사승이 분류의 유일한 기준이 아니었다는 이 사실은 『동유학안』의 사승 서술을 유형화할 때 다시 참조된다. 아울러 황종희는 학문에 이르는 길이 하나가 아님을 강조하고 최종 판단의 일부를 독자에게 위임했지만, 그렇다고 편찬자의 판단이 사라지는 것은 아니다. 『명유학안』에 앞서 유원경(劉元卿)의 『제유학안』이 보여주듯, 편찬자는 명시적 판정을 자제하더라도 누구를 수록하고 어떤 문장을 발췌하며 어떤 순서로 배열하는가를 통해 판단을 수행한다(Chu 2007, 250-253쪽).

둘째, 『송원학안』은 계보 표현의 정교화를 보여준다. 현행 『송원학안』은 황종희 원본, 전조망 보수본, 왕재재와 풍운호의 교보본이라는 세 편찬 단계가 누적된 결과이며, 동일한 학안 안에 원전 발췌, 소전, 서록, 부록, 편찬자의 언어가 함께 배치되어 있어 수록된 모든 문장을 동일한 성격의 진술로 볼 수 없다(連凡 2017, 113-114쪽). 학안표에 표시된 문인, 가학, 사숙, 강우, 학려, 동조 등의 관계 명칭과 재전에서 팔전에 이르는 전승 거리의 표기는 객관적 인물 관계의 목록이 아니라 입문의 직접성, 친족 관계, 사상적 유사성, 편찬자의 평가에 따라 세분된 유형 체계이며, 한 인물에게 복수의 관계가 있을 때에는 주요관계와 부차관계를 구분하고 참조 표지로 여러 학안에 분산된 기록을 연결하였다(連凡 2013, 155-166쪽). 주요관계의 선택 자체가 그 인물의 학문 형성에서 무엇이 가장 중요했는가에 대한 편찬자의 판단이므로, 학안표는 계보를 옮긴 표가 아니라 편찬자가 구성한 학술사적 질서를 드러낸다(連凡 2013, 163-164쪽). 특히 전조망은 특정 인물을 기존 학안에서 분리하여 독립 학안을 저술하거나 원고 단계의 학안을 삭제하였고, '원우당안'처럼 인물이 아닌 사건을 편찬 단위로 설정하기도 하였다(連凡 2017, 121-125쪽). 현행본의 소전 2,428개 가운데 실제 사상 자료를 남긴 학자가 209명에 불과하다는 통계는 인물의 수록과 사상 자료의 발췌가 서로 다른 편찬 행위임을 보여주는데(連凡 2017, 114-115쪽), 이는 『동유학안』에서 발췌 자료의 수록 여부와 분량이 인물별로 차등화되어 있다는 사정과 맥락을 같이 한다.

셋째, 『국조학안소지(國朝學案小識)』는 편찬자의 평가를 노출한다는 점에서 중요하다. 당감(唐鑑)은 1845년 주자학 정통론의 기준에서 청대의 학자 256명을 '전도학안', '익도학안', '수도학안', '경학학안', '심종학안'의 다섯 범주로 분류하였다(임형석 2004, 150-156쪽). 이 분류에서 평가는 여러 층위로 노출된다. 도를 전하고 돕고 지키는 정도에 따른 명칭의 위계, 주자학 관련 세 학안이 전체의 3분의 2 이상을 차지하는 분량 배분, 연대순이 아니

라 도통론적 가치에 따른 인물 배열, 양명학을 불교 선종에 빗댄 심종이라는 명명과 이를 권15가 아닌 권말로 부른 권차 표현이 그것이다(임형석 2004, 152-157쪽). 특히 자료가 부족하여 판단을 확정할 수 없는 인물을 별도의 「대방록」에 수록한 조치는 편찬자의 판단 유보가 문헌의 체재로 표현된 사례로서(임형석 2004, 154쪽), 불확실성과 추정의 표시라는 본 연구의 모델링 요구를 문헌 전통 안에서 선취한 형식이라 할 수 있다. 요컨대 『국조학안소지』는 양명학의 관점에서 정리된 『명유학안』의 형식을 계승하면서 그 학술사적 관점은 정면으로 비판하였다. 동일한 학안체가 서로 대립하는 정통론을 표현할 수 있다는 사실은(임형석 2004, 156쪽), 형식의 계승과 관점의 구성이 별개의 층위임을 보여준다.

세 저작의 대비를 정리하면 다음과 같다.

| 구분             | 『명유학안』         | 『송원학안』                               | 『국조학안소지』            |
|----------------|----------------|--------------------------------------|---------------------|
| 편찬             | 황종희, 1680년경 간행 | 황종희 착수, 전조망 보완, 왕재재·풍운호 교보, 1838년 간행 | 당감, 1845년 간행        |
| 편찬 관점          | 양명학            | 양명학 계승, 전조망 단계에서 범위 확대               | 주자학 정통론             |
| 편찬적 특징         | 학과별 분류의 기준 설정  | 계보 표현의 정교화                           | 명칭·배열·분량을 통한 평가의 노출 |
| 본 연구의 대응 표상 대상 | 학과 배속과 편차 구조   | 사승·전승 관계의 유형 체계                      | 총론·평어와 배열·분량의 평가 기능 |

이 세 축, 곧 분류의 기준, 계보의 표현, 평가의 노출은 본 연구가 『동유학안』에서 데이터로 표상해야 할 편찬 행위의 세 축과 대응한다. 학안체의 역사 자체가 이 세 축이 편찬자마다 다르게 선택되고 조합되는 과정이었다는 사실은, 세 축을 확정된 사실이 아니라 편찬자에게 귀속된 판단으로 기록해야 한다는 본 연구의 관점을 문헌학사의 수준에서 뒷받침한다.

## 2) 조선의 연원록·사우록 전통과 학안체의 수용

조선의 계보 문헌 전통은 학안이 아니라 연원록에서 출발하였다. 그 원형은 주희가 1173년에 편찬하여 북송 유학자 46인의 언행과 사승 관계를 정리한 『이락연원록(伊洛淵源錄)』이다. 황종희의 『명유학안』과 『송원학안』도 그 간접적 영향 아래 성립하였고 조선에서 연원록과 사우록에 대한 관심이 형성된 배경도 이와 무관하지 않다(김기주 2022, 4쪽). 다만 『이락연원록』은 인물별 사회적 전기를 제공하는 관계망적 문헌이면서도, 인물의 전기 뒤에 그 저술을 독립적인 선집으로 배치하는 『명유학안』의 결정적 특징을 결여하고 있다(Chu 2007, 247쪽). 전기와 저술 발체의 결합 여부라는 이 기준은 조선의 연원록류와 학안체를 변별하는 기준으로 제시된다.

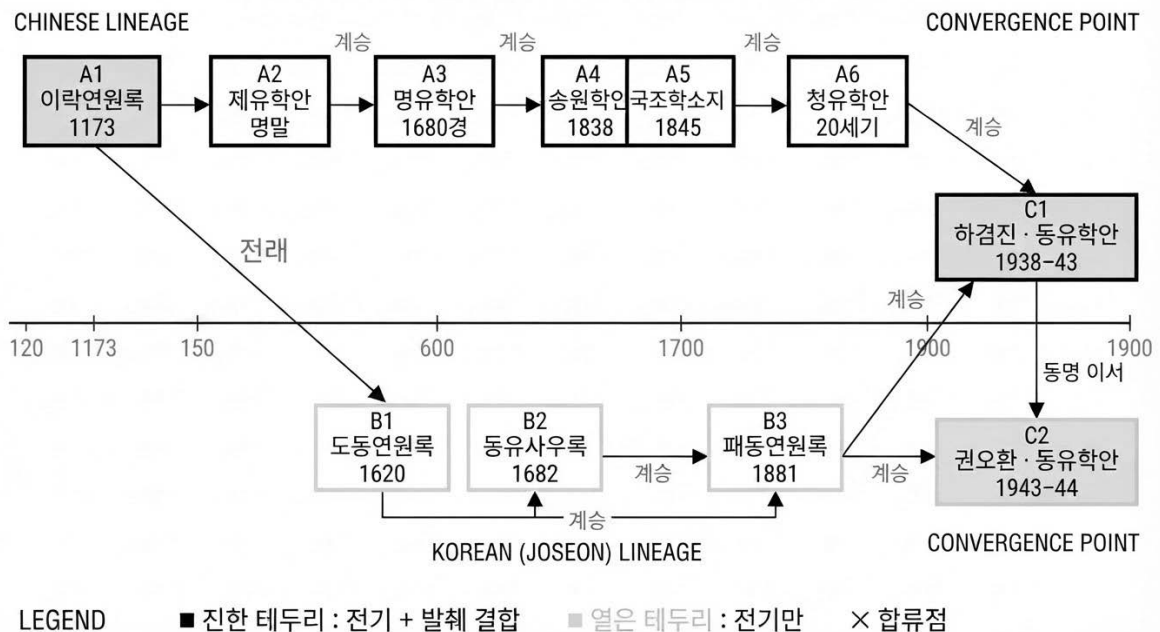
조선에서는 특정 인물의 문인록과 사우록이 먼저 축적되었고, 이후 조선유학 전체를 다루는 통사적 저술로 확대되었다. 김기주의 전수 조사에 따르면 『한국문집총간』 조사에서 확인되는 69종의 사우록·문인록·연원록 가운데 통사적 성격을 갖춘 선구는 오희길(吳希吉)의 『도동연원록(道東淵源錄)』, 박세채의 『동유사우록』, 송병선(宋秉璿)의 『패동연원록』 세 편이다(김기주 2022, 11-12쪽). 『도동연원록』(1620년 편찬)은 기자에서 이이에 이르는 도통 계보를 처음으로 통사적으로 제시했으나, 수록 인물 99인이 유학자에 한정되지 않고 서술도 행장과 유교 중심이어서 엄밀한 유학사라기보다 전기적 저술이다(김기주 2022, 12-13, 18-19쪽). 『동유사우록』(1682년 완성)은 설총과 최치원에서 시작하여 총 799인을 수록하고, 주요 인물 아래에 문인과 종유인을 묶어 배치하였다. 문인이라는 표현으로 묶은 단위를 학안으로 바꾸어도 어색하지 않을 만큼 인물 조직의 형식은 학안에 접근했지만, 수록 내용은 여전히 행장과 묘갈명에 나타난 행적 중심이어서 내용상 전기적 저술의 성격을 벗어나지 못했다(김기주 2022, 13-16쪽). 『패동연원록』(1881년 편찬)은 수록 대상을 정몽주 이후의 성리학자 69인으로 한정함으로써 인물 선정 기준에서 유학사적 서술로의 전환을 보였으나, 수록 자료가 행장, 신도비명, 제문 등 타인이 작성한 기

록 중심이어서 학자 개인의 학설 제시에는 이르지 못했다(김기주 2022, 16-17, 19쪽; 최일범 2008, 12쪽).

세 저술은 도통론적 시각, 전기적 서술에서 유학사적 서술로의 점진적 이행, 기호학과 중심의 구성, 연원 설정의 축소, 그리고 호칭을 통한 인물 평가라는 공통점을 보인다(김기주 2022, 18-21쪽). 이 가운데 호칭의 차등, 예컨대 『동유사우록』에서 선생 칭호의 부여 여부로 학문적 성취를 구분하고 『패동연원록』에서 성의 생략 여부로 존송의 등급을 표현한 방식은(김기주 2022, 21쪽), 편찬자의 평가가 명시적 평어 없이 표기 관행 자체로 수행될 수 있음을 보여준다. 평가적 진술을 사실 진술과 분리하여 기록하려는 본 연구의 설계에서, 이처럼 서술 형식에 녹아 있는 평가는 별도의 식별 대상이 된다.

연원록 전통에서 특히 주목할 사례는 『동유사우록』이다. 이 저술은 박세채의 본편에 문인 이세환(李世煥)의 후집과 보유가 결합된 다층적 편찬물로서(고영진 2009, 281쪽), 이세환의 작업은 누락 인물의 보유, 이이·성혼 문인록의 정리, 편찬의 재구성, 후속 시대의 속찬이라는 네 범주에 걸쳤고, 그는 자신의 보충을 별도 편성하여 박세채의 편찬 내용과 구분하였다(진오균 2023, 272쪽). 원편찬자와 후대 증보자의 층위 구분이 필요한 선례라는 점에서, 이는 II.2에서 다룰 하검진본의 편찬 층위 문제를 예비한다. 또한 박세채는 인물 서술에서 자신의 논평을 가능한 한 피하고 타인의 기록을 전거와 함께 인용하는 방식으로 객관성을 추구했지만, 성혼 문인의 최다 수록과 이이 서술의 상대적 우위 등 인물 선정과 분량 배분에는 편찬자의 학파적 입장이 반영되었다(고영진 2009, 286쪽). 인용과 전거 표시가 편찬의 관점성을 제거하지 못한다는 이 관찰은, 인용 중심 체제를 갖춘 문헌일수록 인용 행위 자체를 표상 대상으로 삼아야 한다는 본 연구의 문제 설정을 조선 문헌 전통 안에서 선취한 사례이다.

| 구분    | 『도동연원록』      | 『동유사우록』               | 『패동연원록』          | 하검진 『동유학안』         |
|-------|--------------|-----------------------|------------------|--------------------|
| 편찬    | 오희길, 1620년   | 박세채·이세환, 1682년 완성     | 송병선, 1881년       | 하검진, 1938-1943년    |
| 기점 인물 | 기자           | 설총·최치원                | 정몽주              | 설총                 |
| 수록 인원 | 99인          | 799인                  | 69인              | 152인               |
| 조직 원리 | 도통 계보의 권별 배열 | 주요 인물 아래 문인·사숙·종유의 묶음 | 연원에 따른 인물 배열     | 학파별 학안과 편차         |
| 수록 자료 | 행장·유고        | 행장·묘갈명 등 행적 기록        | 행장·신도비명 등 타인의 기록 | 총론·인물 서술·본인 저술의 발췌 |
| 학설 제시 | 미비           | 미비                    | 미비               | 체제의 핵심 요소          |



## [도해 2] 학안류와 연원록류의 편찬 연표

이 대비가 보여주듯, 조선에서 연원록과 문인록의 전통이 축적된 뒤에도 학파 총론과 본인 저술의 발체를 갖춘 학안 형식은 20세기에 이르러서야 채택되었다. 그 지연이 갖는 의미, 곧 하겸진이 왜 이 시점에 이 형식을 선택했는가는 아래에서 다룬다.

## 3) 동명 이서 권오환본과의 구분

하겸진본을 다루기에 앞서 동일한 서명의 별개 저작을 미리 구분해 둘 필요가 있다. 권오환(權五煥)이 편찬하고 학승(學僧) 권상로(權相老, 1879-1965)가 1944년에 5권 5책으로 간행한 『동유학안』이 그것이다(김기업 2023, 310-311쪽). 이 책은 경상도 지역 문중 92개의 계보를 본관과 거주지에 따라 정리한 「도안편」을 본체로 삼아 문중 중심의 계보적 배열을 보인다(김기업 2023, 311-315쪽). 편찬 시기도 거의 겹쳐서, 하겸진본이 1943년에 완료된 데 비하여 권오환본은 권1 말미의 「학안발」이 1943년에, 권5 말미의 발문이 1944년에 각각 작성되었다(김기업 2023, 313쪽, 315쪽). 다만 두 편찬자가 서로의 작업을 알고 있었는지는 확인되지 않으므로, 동시대성은 영향 관계가 아니라 같은 시기에 같은 요구가 별개의 경로로 제기되었다는 사실로만 이해해야 한다.

서명이 같다는 사정이 실제로 서지의 혼동을 낳았으므로, 이 서명을 지칭할 때에는 하겸진 편찬본과 권오환 편찬본을 변별해야 한다는 제안이 이미 제출되어 있으며(김기업 2023, 301쪽), 본 연구는 이를 따른다. 권오환본의 존재는 연구사 서술에도 조정을 요구한다. 하겸진본을 조선 유학사에서 유일한 학안으로 규정해 온 서술은 더 이상 성립하지 않으므로, 본 연구는 유일이라는 표현을 채택하지 않고 대신 하겸진본의 성격을 학파 분류 학안체로 특정한다. 하겸진본이 학파를 기준으로 학안을 나누고 각 학안의 서두에서 학파의 연원을 서술한 뒤 사상적 특징이 드러나는 저술을 선별하여 수록한 데 비하여, 권오환본은 문중과 성씨를 기준으로 학자를 분류하였기 때문이다(김기업 2023, 318쪽). 곧 두 저작의 차이는 학안이라는 형식의 채택 여부가 아니라 그 형식을 어떤 분류 원리 위에 설정했는가에 있으며, 동일한 시기에 동일한 명칭의 저작이 서로 다른 분류 원리로 복수 편찬되었다는 사실은 학안이라는 형식이 당대에 조선 유학의 전통을 구성하고 보존하는 도구로서 요청되고 있었음을 보여준다.

## 4) 하겸진의 생애와 학문적 지향

하겸진은 1870년 경남 진주에서 태어나 1946년에 사망한 유학자로, 자는 숙형, 호는 회봉이며, 1896년 곽종석(郭鍾錫)을 찾아가 그 문인이 되었다(최일범 2008, 13쪽). 본 절에서 생애 사실의 나열보다 중요한 것은 이 학맥상 위치가 『동유학안』의 계보 구성 관점을 어떻게 규정했는가이다. 하겸진은 곽종석을 매개로 이진상의 한주학에 연결되는 계보, 곧 『동유학안』 자신의 분류로는 평포학안의 포학 계열에 속한다(최일범 2008, 21-22쪽). 이 위치

는 편찬의 여러 국면과 조응한다. 하겸진은 도산학안을 퇴계학파만의 학안이 아니라 한국 천년 유학을 대일통한 학안으로 규정하였고(최일범 2008, 19-20쪽), 퇴계와 남명의 학풍 차이를 인정하면서도 그 마음과 도는 같다고 서술하여 영남 도학 내부의 차이를 상위의 동일성 아래 통합하였으며(최일범 2008, 20쪽), 당시 영남 학자들의 비판 대상이던 이진상의 심즉리설을 퇴계의 본지를 밝힌 학설로 해석하였고(최일범 2008, 21-22쪽; 김기주 2019, 181쪽), 이이가 퇴계에게 직접 지결을 받았다는 관계를 강조하여 그 학문적 기원을 퇴계에 연결하면서 서인과 동인의 상반된 이이 해석을 모두 편당적 견해로 비판하였다(최일범 2008, 22-24쪽; 김기주 2019, 178-179쪽). 어느 국면에서든 편찬자의 계보적 위치와 학술사 구성이 분리되지 않는다.

이 위치성에는 이중의 결이 있다. 하겸진은 진주 사곡의 진양하씨 가학을 통해 남명학을 계승한 동시에(이영숙 2022, 124-125쪽), 퇴계와 직접적 사승관계가 없음에도 스스로를 퇴계 정맥의 계승자로 인식하는 자의식을 지녔다(이영숙 2022, 132쪽). 「덕산학안」이 퇴계와 남명을 영남 좌도와 우도의 도학 종사로 병치하고 두 사람의 차이를 우열이 아니라 실천 방식의 차이로 구성한 서술은(이영숙 2022, 133-134쪽), 이 이중적 계승 구조를 전기적 배경으로 갖는다. 또한 하겸진은 「심위자모설」과 「성사심제변」을 저술하여 성리학적 논변에 직접 참여한 논자였으므로(최일범 2008, 13쪽), 『동유학안』의 편찬을 학설상의 입장이 없는 중립적 정리 작업으로 상정할 수 없다.

선행 연구는 이러한 특성을 이학 중심의 도통론, 도산학파 중심성, 인물과 도통 중심의 대일통적 편찬으로 정리하고, 그것이 학문적 객관성의 결여를 초래했다고 평가하였다(최일범 2004, 138-139쪽). 앞에서 밝혔듯 본 연구는 이 평가를 결함의 지적으로 반복하는 대신, 편찬자의 위치성과 정통관이 인물의 선정과 학파의 분류와 관계의 연결과 저술의 발체 및 배열을 구성한 과정 자체를 표상의 대상으로 삼는다.

## 5) 편찬 시기와 동기

하겸진은 69세인 1938년 문인들과 함께 경성도서관을 방문하여 자료 조사에 착수했고, 소장 자료가 충분하지 않아 호남으로 이동하여 곡성의 장서가 정봉태에게서 책을 빌려 보았으며, 5년의 조사와 저술 끝에 74세인 1943년 『동유학안』 30권을 완성하였다(정성희 2013, 174-175쪽; 이영숙 2021, 123쪽; 김기주 2018, 97쪽). 근대적 도서관과 사가 장서를 병용한 이 과정은 편찬의 물적 조건이 지역 학맥의 전승 자료에 한정되지 않았음을 보여준다.

편찬 동기의 전거에 관해서는 주의가 필요하다. 『동유학안』에는 하겸진 자신의 서문이나 발문이 없어, 편찬 목적은 1970년 간행본에 부가된 문인 노홍두(盧洪斗)의 「동유학안중언후지(東儒學案重印後識)」와 연보를 통해 간접적으로 추정될 뿐이다(김기주 2019, 169쪽). 그 기록에 따르면 하겸진은 선유들의 학술과 행적을 고찰할 문헌이 없음을 안타까워하여 중국의 송명학안을 모방해 이 책을 편찬하고 후학이 의거할 근거를 마련하고자 했다(최일범 2008, 14쪽; 김기주 2019, 169쪽).

편찬의 시대적 동기에 대해서는 두 가지 해석이 제출되어 있다. 하나는 하겸진의 저술 「국성론(國性論)」과의 연속선에서 읽는 해석이다. 하겸진은 1921년 「국성론」에서 유학적 예의를 민족의 국성으로 규정하고 국체가 소멸해도 국성이 보존되면 회복이 가능하다고 논했는데, 민족개조론과 유교망국론이 확산되던 1938년의 시점에서 『동유학안』 편찬은 국성의 바탕인 유학의 연원을 밝혀 사상적 정체성을 보존하려는 작업이었다는 것이다(정성희 2013, 172-175쪽). 다른 하나는 도통론에 입각하여 인멸되어 가는 학술을 기록하고 후대에 계승하려는 소명의식으로 읽는 해석이다(이영숙 2021, 123쪽; 김기주 2019, 169-170쪽). 선행 연구자들의 두 해석 모두 편찬을 식민지기의 학술적 대응, 즉 저항적 편찬으로 자리매김하고 있다.

형식 선택의 문제는 앞의 계보 검토와 직접 연결된다. 하겸진의 학안체 채택은 만년의 즉흥이 아니었다. 그는 이미 1917년의 서신에서 『송원학안』과 명칭의 학안을 논했고 1933년에도 학안의 심리 학설을 검토하였다(김기주 2018, 102-103쪽). 하겸진은 또한 수도제유학안의 편성에서 당경해의 『청유학안』(당감의 『국조학안소지』를 가리키는 것으로 추정된다)이 수도학안과 익도학안을 나눈 것은 지나치다고 판단하여 두 범주를 하나로 합쳤다고 스스로 밝혔는데(최일범 2008, 14, 26-27쪽), 이는 선행 학안의 형식을 복제한 것이 아니라 자신의 분류 목적에 맞게 변형했음을 보여준다. 그렇다면 왜 연원록이나 사우록이 아니라 학안이었는가. 앞에서 확인했듯 『동유사우록』과 『패동연원록』은 사승과 행적의 기록을 제공했으나 각 학파의 학문적 특징과 개인의 학설을 직접 제시하지 못했고, 학안체는 학파 총론, 인물 서술, 본인 저술의 발체를 결합함으로써 학파의 연원과 사상을 함께 제시할 수

있었다(최일범 2008, 11-14쪽). 전기적 서술에서 유학사적 서술로 이행해 온 조선 계보 문헌의 흐름에서, 하겸진의 형식 선택은 그 이행을 완결하는 위치에 있다. 그리고 그 완결의 수단이 바로 저술 발체, 곧 인용이었다는 사실은 인용을 부속 정보가 아니라 일차적 모델링 대상으로 삼아야 한다는 뒤의 주장에 대한 문헌학적 근거가 된다.

## 2. 판본과 저본: 전승 상황과 자체 비평본의 구성

이 절은 『동유학안』의 전승 상황에서 출발하여 저본의 확정에 이르는 경로를 서술한다. 도달하는 결론은 두 가지이다. 하나는 이 문헌에 접근하는 모든 경로가 이미 여러 층위의 편집을 거쳐 있으며 그 편집의 근거가 어디에도 기록되어 있지 않다는 것이고, 다른 하나는 그러한 사정 때문에 저본을 그대로 옮기는 대신 대교를 거친 자체 비평본(critical edition)을 구성해야 하며 그 구성이 인코더의 편집 층위를 하나 더 추가한다는 것이다. 두 결론은 문헌학적 관찰에 머무르지 않는다. 앞의 것은 편찬 층위를 구분하여 기록해야 한다는 요구의 자료적 근거가 되고, 뒤의 것은 그 층위의 목록에 인코더 자신이 포함되어야 하는 이유가 된다.

### 1) 수고본의 부재와 편찬의 미완

하겸진의 수고본은 현재 소재가 알려진 것이 없다. 따라서 원편찬 층위에 직접 접근할 수 있는 경로는 존재하지 않으며, 이 문헌에 대한 모든 접근은 후대의 간행 층위를 경유한다. 편찬 층위를 구분해야 한다는 요구가 이론적 정교화의 산물이 아니라 자료 상황이 강제하는 조건이라는 사실이 여기서 먼저 확인된다.

이 사정은 본 연구가 수행할 대교의 성격에도 제약을 가한다. 통상의 교감(校勘)에서 대교는 파생된 전본을 보다 이른 전본에 견주어 원본에 가까운 독법을 복원하는 절차이지만, 이 문헌에서는 견줄 상위 증거가 존재하지 않는다. 1970년본과 2008년본은 어느 쪽도 원편찬의 상태를 증언하지 못하며, 두 판본의 대조가 산출하는 것은 원본의 복원이 아니라 두 파생본 사이의 차이 목록이다. 그 차이 가운데 어느 쪽을 본문으로 채택할 것인가는 결국 인코더의 판단에 맡겨지고, 그 판단은 원본에 의하여 검증되지 않는다. 자체 비평본이 기록해야 할 것이 확정된 본문 자체가 아니라 독법의 선택과 그 근거라는 방침은 이 사정에서 나온다.

편찬 자체의 성격에 관해서도 편찬자 자신의 규정이 남아 있다. 1970년 간행본의 부록에 실린 후서(後序)는 편찬의 종결 사정을 다음과 같이 전한다.

先生纂是書。至郭徵君而止。曰。此大公案。何必余之爲成始成終哉。以語諸子。仍錄前輩略干人名號。屬李公鉉德補之。李公拜受命。而略干人者。顧多文集之難求。事狀之未備。難於措手矣。

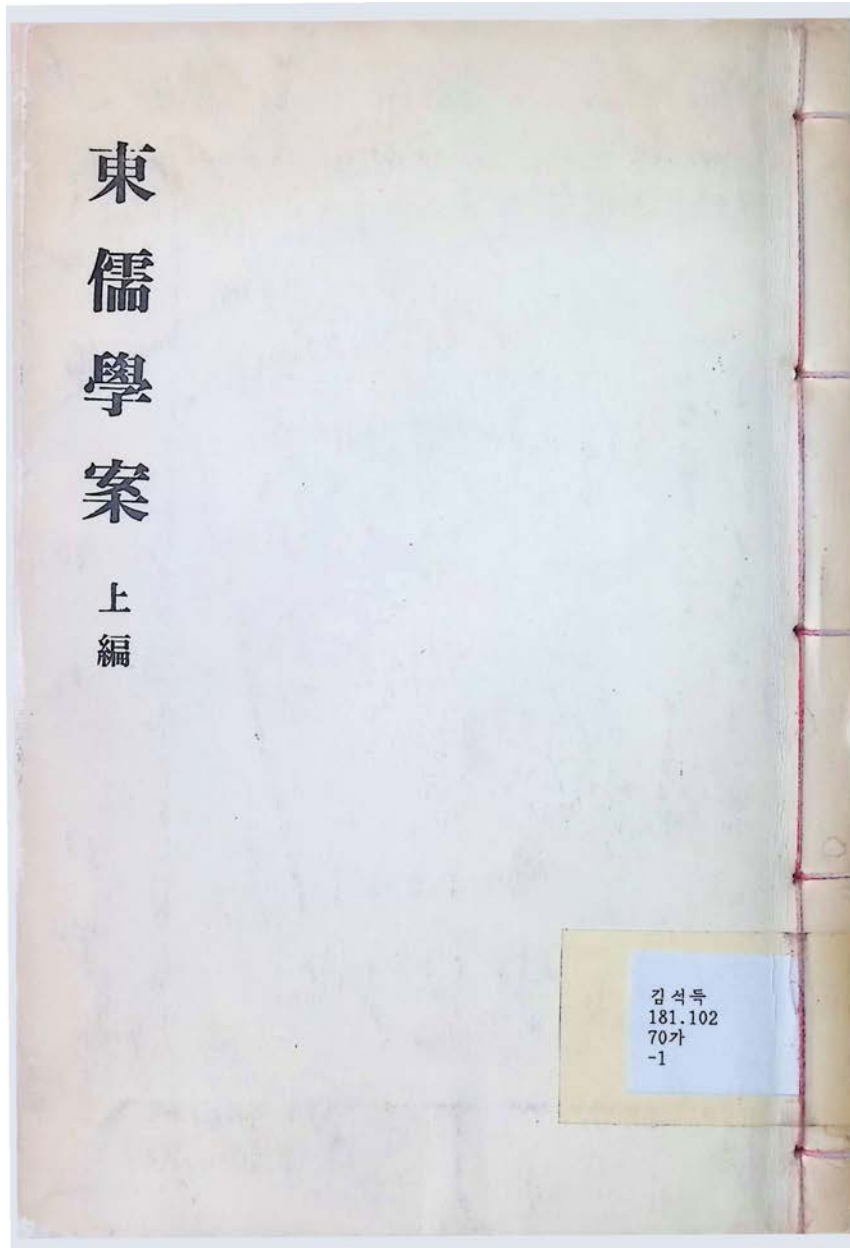
선생이 이 책을 편찬하다가 광정군에 이르러 그쳤다. 이르기를, 이것은 큰 공안이니 어찌 반드시 내가 시작과 끝을 다 이루어야 하겠는가 하였다. 이 말로 제자들에게 이르고, 이어서 전배 약간인의 명호를 기록하여 이현덕에게 그 보완을 맡겼다. 이현덕이 절하고 명을 받았으나, 약간인이라 한 이들은 문집을 구하기 어렵고 사장이 갖추어지지 않은 경우가 많아 손을 대기 어려웠다.

이어지는 대목은 하겸진 사후 이현덕이 유서 간행의 일을 맡아 오랜 뒤에야 그것을 마쳤고, 상란과 우환이 십여 년을 이어지는 가운데 하루의 겨를도 얻지 못한 채 세상을 떠났으며, 그럼에도 편찬자가 제자들에게 이른 말의 뜻은 그대로이므로 후일 이어서 지을 사람을 기다린다는 내용으로 마무리된다. 본 연구의 현 단계에서 이현덕에 관하여 현재 확인되는 것은 이 후서의 진술이 전부이다.

이 기록이 갖는 의미는 세 가지로 정리된다. 첫째, 이 저작은 편찬자 자신에 의하여 미완으로 규정된 문헌이며, 따라서 수록 인물의 범위는 완결된 판단의 산물이 아니다. 둘째, 보완의 필요성이 후대 인물이 아니라 편찬자에게서 직접 제기되었다. 셋째, 그 위임은 끝내 수행되지 않은 채 남아 있다가 뒤에서 살펴볼 2008년의 『증보 동유학안』에서 1차 실현되었다. 이 셋을 함께 놓으면 미수룩을 단지 편찬자의 적극적 배제로만 읽는 해석은 조정되어야 할 것이다. 미수룩 안에는 배제와 미완이 함께 들어 있으며, 이 구분은 데이터에서도 유지되어야 한다. 아울러 편찬자가 자신의 작업을 공안이라 부른 대목은 학안의 안이 판정과 사례의 뜻을 함께 지닌다는 장르적 규정과 직접 맞물린다. 편찬자는 자신이 하는 일이 확정된 결론을 내리는 작업이 아니라 판단을 요하는 사안을 제시하는 작업이라고 이해하고 있었다.

한 가지를 덧붙여 기록해 둔다. 이 후서는 개인의 이름이 아니라 출판사 명의로 서명되어 있다. 편찬의 경위와 종결 사정을 전하는 가장 중요한 진술 가운데 하나가 발화 주체를 개인으로 특정할 수 없는 형태로 남아 있는 셈이다.

## 2) 1970년 간행본



# 東儒學案上編目錄

第一編 羅麗二代韓初諸儒學案

康州河謙鎮纂

弘儒侯薛先生聰  
文昌侯崔先生致遠  
中書令崔先生冲  
中贊安晦軒先生裕  
祭酒禹先生倬  
評理白頤齋先生頤正  
僉議政丞權菊齋先生溥  
待中李益齋先生齊賢  
參理崔春軒先生文度  
贊成事安謹齋先生軸  
文學白先生文寶  
監務李石灘先生存吾  
大司憲河苦軒先生允源  
判典教朴潘南先生尙衷  
贊成事田耕隱先生祖生  
侍中李牧隱先生穉

一一〇九八七七六五五五四三二二一

第一編

第二編

第三編

提學李陶隱先生崇仁  
待中鄭圃隱先生夢周  
進士元耘谷先生天錫  
參贊權陽村先生近  
注書吉治隱先生再  
司藝金江湖先生叔滋  
理學淵源諸儒學案  
判書金佑畢齋先生宗直  
佐郎金寒暄堂先生宏弼  
說書鄭一蕤先生汝昌  
贊成金慕齋先生安國  
大司憲趙靜庵先生光祖  
贊成李晦齋先生彥迪  
說書金河西先生麟厚  
參贊白休庵先生仁傑  
陶山學案  
贊成李退溪先生澗

一一三  
一五  
一六  
三八  
三九  
四〇  
四一  
四二  
四三  
四四  
五〇  
七八  
八一  
八三

11232080

[도해 3] 1970년 간행본의 표지와 목차: 연세대학교 중앙도서관 김석득 문고 소장본에서 제작한 디지털 이미지

간기의 판독 결과는 다음과 같다. 대한민국 52년(1970) 경술 7월 1일 인쇄, 같은 해 7월 25일 발행으로 되어 있고, 서명 아래에 비매품이라는 표시와 상중하 3편 1질이라는 체제 표시가 붙어 있다. 저작검발행은 하영기, 편집은 이일해, 감인은 손미수와 이상도와 서경보이며, 인쇄는 진주 삼협인쇄소, 출판사는 해동불교역경원 내의 일봉정사이다. 발행처의 등록일은 1962년 9월 26일로 기재되어 있고, 지면 상단에는 복제를 허가하지 않는다는 표시가 별도로 찍혀 있다. 연도를 대한민국 연호와 간지로 병기하고 서기를 쓰지 않은 표기 방식은 이 간행이 놓인 문화적 위치를 함께 드러낸다.

이 표기에서 두 가지를 읽을 수 있다. 하나는 간행의 주체이다. 하영기는 『회봉선생묘지명』에 문인으로 기록되어 있고, 이일해는 하용환과 하영기와 성환혁 세 문인의 청탁을 받아 그 묘지명을 지은 인물이다. 곧 이 간행은 개인의 작업이 아니라 문인 집단의 사업이었다. 다른 하나는 역할의 분담이다. 저작검발행과 편집이라는 두 표기는 서로 다른 역할을 지목한 것이며, 『한국민족문화대백과사전』의 「동유학안」 항목이 이 판본을 문인 이일해의 간행으로 기술한 것도 편집자의 역할을 가리킨 서술로 이해된다. 다만 저작검발행이라는 근대 출판법상의 형식적 표기가 하영기를 원저작자로 오인하게 만들 소지가 있다는 점은 지적해 둔다. 표기 형식이 지시하는 역할과 실제 수행된 역할이 어긋나는 이 사례는, 문헌에 기재된 명의를 그대로 귀속 관계로 옮기는 방식이 왜 위험한지를 보여주는 자료 내부의 예시이기도 하다.

간행의 물적 조건도 함께 확인된다. 발행처가 불교 역경 기관 내의 조직이며 감인에 그 기관의 책임자가 이름을 올리고 있다는 사실은, 이 판본이 유학 문중이나 학술 기관의 출판 경로가 아니라 별도의 경로를 통하여 간행되었음을 보여준다. 비매품이라는 표기는 유통의 범위가 처음부터 제한적이었음을 뜻한다. 동명 이서인 권오환의 『동유학안』 역시 불교계 지식인인 권상로에 의해 간행되었다는 점도 특기할 만하다.

다만 이 판본을 최초 간행본으로 단정하는 것에는 유보적이다.. 현재 확인 가능한 가장 이른 판본으로 다루되 그 이전의 간행 가능성은 유보 상태로 남겨 둔다. 근거는 두 가지 흔적이다.

하나 1961년에 이일해가 지은 「회봉선생묘지명」이 하겸진의 저작을 열거하는 대목이다. 해당 구절은 「所著遺書五十卷。朱子語類節要二十卷。東詩話二卷。并刊行東儒學案三十卷。方釐寫明史綱目略干卷。」으로 되어 있는데, 원문에 구두점이 없어 아울러 간행하였다는 말의 목적어에 동유학안이 포함되는지가 결정되지 않는다. 유서와 주자어류절요와 동시화에 이어 동유학안 30권까지 간행하였다는 독법이 하나이고, 앞의 셋을 아울러 간행하였으며 동유학안과 명사강목은 바야흐로 정리하여 베끼는 중이었다는 독법이 다른 하나이다. 전자를 취하면 1961년 이전에 간행본이 존재했다는 뜻이 되고, 후자를 취하면 1970년본이 최초 간행본이라는 통설과 어긋나지 않는다. 흥미로운 것은 이 구절이 앞 절에서 다룬 인용 경계의 미표시와 동일한 구조의 문제라는 점이다. 구두점이 없는 한 문 서술에서 어디까지가 하나의 술어가 지배하는 범위인가는 후대의 독자가 판정할 사항이며, 그 판정에 따라 사실관계 자체가 달라진다.

다른 하나는 선행 연구가 편찬 동기의 전거로 삼아 온 노홍두의 글이 표제에 중인(重印)이라는 말을 포함한다는 점이다. 중인은 앞선 간행을 전제하는 말이다. 두 흔적은 서로 독립적이면서 같은 방향을 가리킬 수 있다.

이 두 번째 흔적은 별도의 문제를 함께 제기한다. 노홍두의 「동유학안중인후지」는 최일범 2008과 김기주 2019가 모두 편찬 동기의 근거로 삼고 있으나, 본 연구자가 연세대학교 중앙도서관에서 확보한 1970년본의 부록에서는 확인되지 않았고, 다만 앞서 인용한 후서가 있다. 이 글이 실린 판본에 관해서는 1970년본 이전 또는 이후의 별도 간행본에 실려 있을 가능성, 1970년본 내부의 확인하지 못한 위치에 있을 가능성, 선행 연구가 판본을 특정하지 않은 채 서술했을 가능성이 모두 남아 있다. 어느 쪽도 현 단계에서는 확정할 수 없으므로 확인은 추후의 조사로 미룬다. 다만 확정예 앞서 기록해 두어야 할 사실이 있다. 이 문헌의 편찬 동기를 전하는 주요 전거가 어느 판본의 어느 위치에 있는지 특정되지 않은 채 인용되어 왔다는 사정 자체가, 본 연구가 출처 환원성이라는 이름으로 요구해 온 조건이 이 문헌의 연구사 안에서 실제로 결여된 사례라는 점이다. 판본의 계보 자체가 미확정 상태라는 사실 역시 데이터에 기록되어야 할 조건이다.

판본의 형태에 관해서는 다음을 확인하였다. 신연활자본(新鉛活字本)이며 고서를 모방하여 선장(線裝) 형식으로 장정되었으나, 쪽수 표기와 본문에 오기가 많고 지질도 조악하다. 형식은 전통 판본의 외형을 취하면서 실제 제작의 정밀도는 그에 미치지 못하는 셈이며, 이 어긋남 자체가 앞에서 확인한 간행의 물적 조건과 대응한다.

이러한 점에서 두 가지가 따라 나온다. 하나는 이 판본을 저본으로 삼기 어렵다는 것이다. 다른 하나는 대교 자료로 활용할 때 이문과 오식의 구분 자체가 인코더의 편집적 판단이 된다는 것이다. 두 판본의 독법이 상이한 어느 지점에서 1970년본의 독법을 오식으로 판정하면 그것은 대교의 결과에서 배제되지만, 이문으로 판정하면 선택지의 하나로 기록에 남는다. 이 판정의 기준을 사전에 명문화하기는 어렵고 사례마다 달라질 것이므로, 판정의 근거를 사례 단위로 기록하는 방식이 불가피하다.

### 3) 2008년 『증보 동유학안』

이 판본은 하겸진 편찬, 이기동 외 역주로 나남에서 2008년 6월 25일에 발행되었으며, 부제는 "한국유학자 167인의 생애와 사상"이다. 성균관대학교 연구원을 중심으로 한국학술진흥재단의 기초학문육성 지원을 받아 3년에 걸쳐 수행된 사업의 결과이고, 감수와 역주의 담당자가 머리말과 간기에 명시되어 있다(하겸진 2008, 7-9쪽).

증보의 원칙은 머리말에 밝혀져 있는데, 기호학파에 대한 비중이 상대적으로 작다는 점을 보완하기 위하여 별도로 보유편을 추가했다(하겸진 2008, 8쪽). 이 진술은 증보의 성격을 정확히 규정한다. 곧 이 작업은 하겸진의 편찬 판단을 보존한 채 우연한 누락을 보충한 것이 아니라, 하겸진의 학파별 비중 배분 자체를 조정한 것이다. 도산학파 중심성이라는 이 문헌의 핵심 특성이 증보 단계에서 완화되었다는 뜻이며, 편찬 층위를 구분하지 않고 이 판본을

읽으면 하겸진의 학파 구성과 후대의 조정이 하나의 평면에서 섞이게 된다. 앞의 후서가 전하는 위임의 내용이 "전배 약간인의 보완"이었음을 고려하면, 보유편은 하겸진의 뜻대로 저술을 보완한 것이나 '기호학파'의 비중 강화까지 하겸진의 의도였을지는 알 수 없다.

교감에 관해서는 결여가 확인된다. 원문을 윤독하며 오자와 탈자를 하나하나 대조하였다는 진술은 머리말에 있으나(하겸진 2008, 8쪽), 수록된 한문 원문에는 교감기가 수록되어 있지 않다. 따라서 이 판본의 본문이 어느 지점에서 어떤 근거로 확정되었는지를 판본 자체에서 확인할 방법이 없다. 자체 비평본을 구성해야 하는 직접적 근거가 여기에 있다.

본문의 배치 방식도 저본 확정과 직접 관련되므로 확인해 둔다. 이 판본은 한국어 번역문을 각 권의 본문으로 배치하고, 한문 원문은 권말에 별도의 원문편으로 모아 싣는다. 두 계열은 물리적으로 분리되어 있고 서로 다른 쪽수 위에 놓이며, 번역문 쪽에서 원문 쪽으로 넘어가는 대응 표시는 편명과 인물명 이상으로 제공되지 않는다. 본 연구가 저본으로 삼는 것은 이 권말 원문편이며, 번역문은 참조 자료의 지위에 머문다. 한 권 안에 같은 텍스트의 두 계열이 서로 다른 좌표계 위에 놓여 있다는 사실은, 텍스트 위치를 지시하는 앵커가 어느 계열을 기준으로 설정되어야 하는지를 처음부터 결정해 두어야 하는 이유가 된다.

표점과 각주는 이 판본에서 새로 더해진 층위이다. 1970년본에 없는 표점이 이 판본에는 있으며, 각주는 교감주와 어석주와 출전 확인주와 역자 보충으로 유형화된다. 둘 모두 후대 층위의 판단이다. 특히 표점의 위치는 인용문의 경계 판정을 바꿀 수 있고, 출전 확인주는 편찬자가 밝히지 않은 서지 정보를 추가하며, 역자 보충은 저본에 존재하지 않는 정보를 지면에 도입한다. 아울러 각주가 위의 네 유형 가운데 어디에 속하는가는 판본 자체에 표시되어 있지 않으므로, 유형의 판정 역시 인코더가 수행해야 하는 작업이다. 표점이 인용의 시작과 끝을 어떻게 표시하는가, 또는 표시하지 않는가는 다음 절에서 실제 사례로 확인한다.

## 4) 선행 해제의 공백과 저본의 확정

이 판본에 실린 최일범 2008의 해제는 편찬자의 생애, 편제, 학파 분류의 원리, 편찬 관점의 특성을 정리한 유일한 본격적 해제이며, 앞 절의 서술이 주로 의지한 문헌이다. 다만 판본에 관한 서술은 거의 다루어지지 않는다. 수고본의 소재, 1970년본의 서지와 간행 경위, 후서가 전하는 편찬의 미완, 저본의 지정, 표점과 교감의 담당자에 대한 언급이 없다. 특히 이 해제는 저작을 상중하 3책 23편 151인으로 기술하는데(최일범 2008, 14쪽), 이는 1970년본의 체제이며 해제가 실린 책은 보유편이 더해진 167인 수록본이다. 해제가 대상으로 삼은 텍스트와 해제가 실린 책의 텍스트가 일치하지 않는데도 그 차이가 해제 안에서 설명되지 않는다.

이러한 공백은 해제라는 장르가 내용의 소개를 본령으로 삼고 서지와 판본의 검토를 다루지 않는 데서 온 것이므로 결함으로 규정할 사항은 아니다. 또한 이 공백이 이 해제만의 사정도 아니다. 이 문헌에 관한 연구는 학파 분류의 원리, 편찬자의 생애, 유학사 서술사에서의 위치를 중심으로 축적되어 왔고, 판본 수준의 검토가 정면으로 다루어진 사례는 동명 이서의 구분을 시도한 작업(김기엽 2023)에 그친다. 그러나 그 작업 역시 두 저작을 변별하는데 초점이 있었으므로 하겸진본 내부의 판본 계보를 정리한 것은 아니다.

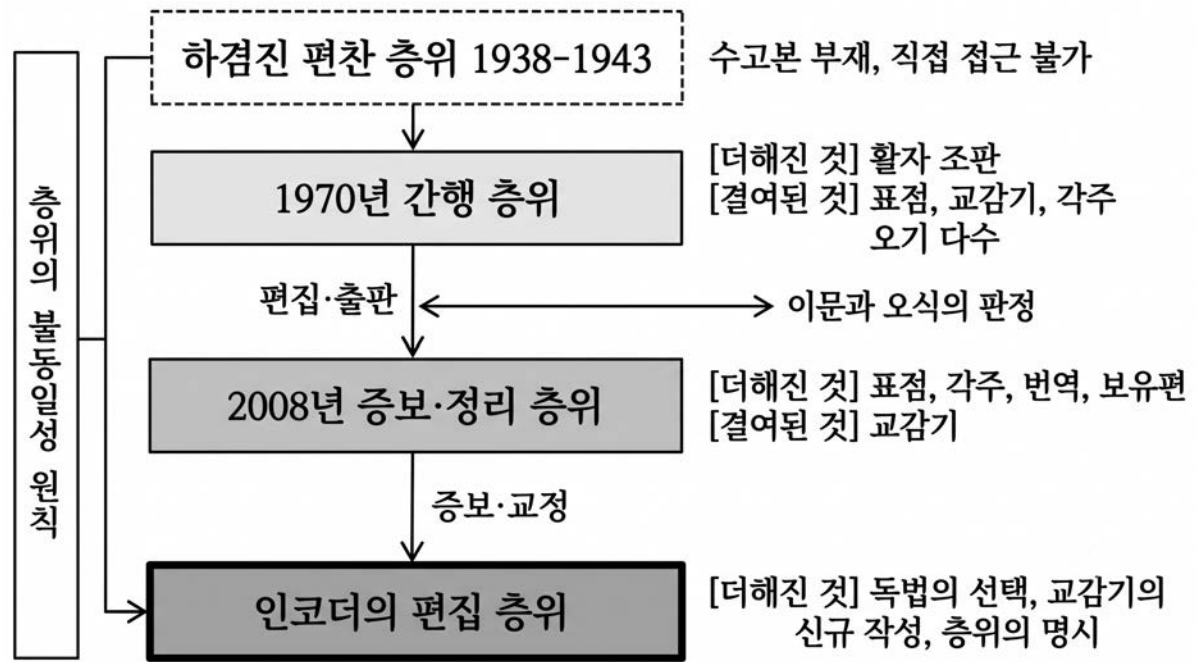
이 공백이 남아 있는 한, 이 문헌을 데이터로 표상하려는 작업은 판본 문제를 스스로 정리하는 데서 출발할 수밖에 없다. 이 절이 존재하는 이유가 여기에 있다. 덧붙이면, 판본이 정리되지 않은 상태에서 데이터 구축이 먼저 이루어질 경우의 위험은 앞에서 확인한 노홍두 글의 사례가 이미 보여주었다. 어느 판본의 어느 위치에서 온 진술인지가 기록되지 않은 채로도 인용은 연쇄적으로 재생산될 수 있으며, 그렇게 재생산된 진술은 소급하여 검증하기 어려워진다. 구조화된 데이터는 이 재생산의 속도를 훨씬 높인다.

수록 인원은 1970년본을 기준으로 152인이며, 2008년 증보본은 보유편을 포함하여 167인이다.

이상의 검토에 따라 저본을 확정한다. 저본은 2008년 증보본의 표점된 한문 원문으로 하고, 1970년본 원문을 대교 자료로, 두 판본의 번역문을 참조 자료로 삼으며, 인코딩의 대상은 번역문이 아니라 한문 원문이다. 산출물은 이 대조를 거친 자체 디지털 한문 비평본이다. 저본을 그대로 전사하지 않는 이유는 앞에서 확인한 두 사정에 있다. 1970년본에는 오기가 많고, 2008년본에는 교감기가 없다. 어느 쪽도 단독으로 인코딩의 근거가 되지 못한다.

그 결과 두 판본 사이에서 독법을 선택하는 판단이 성립하며, 이 판단은 하검진의 편찬 판단이나 2008년 증보자의 판단과 구분되는 별개의 편집 층위를 이룬다. 이 층위가 산출하는 것은 확정된 본문이 아니라 독법의 목록이다. 두 판본의 독법이 상이한 지점마다 어느 쪽을 본문으로 삼았는지, 다른 쪽 독법은 무엇인지, 그리고 그 선택의 근거가 무엇인지가 함께 기록되어야 하며, 이는 저본에 결여된 교감기를 인코더가 새로 작성한다는 뜻이다. 다음 절이 보이겠지만, 인명 표기의 이체와 오식이 한 단락 안에서도 발생하므로 이 목록은 결코 짧지 않을 것이다. 아울러 이렇게 작성된 교감기는 뒤에서 다룰 증거 앵커의 문제와 곧바로 연결된다. 하나의 진술이 어느 판본의 어느 독법 위에서 성립하는가를 밝히지 못하면, 판본을 대교했다는 사실 자체가 데이터에서 사라지기 때문이다.

여기서 편찬 층위의 비동일성 원칙을 세운다. 하검진의 편찬 층위, 1970년 간행 층위, 2008년의 증보와 정리와 표점과 번역 층위, 그리고 인코더의 층위를 동일시하지 않는다는 원칙이다. 앞에서 확인한 표점의 유무, 보유편의 추가, 각주에 의한 비정이 각각 이 원칙의 직접적 근거이다. 다만 이 층위 구분은 문헌에 내재한 것이 아니라 저본 선택에서 파생된 것이므로, 연구의 핵심 질문이 열거하는 층위 목록에는 넣지 않되 데이터에서는 구분하여 기록한다.



[도해 4] 전승 경로와 편집 층위

현재 자료 확보 상황은 다음과 같다. 1970년본은 3편 전량을, 2008년본은 전체 6권 가운데 1권을 디지털화하였다.

| 항목         | 1970년 간행본                 | 2008년 『증보 동유학안』         |
|------------|---------------------------|-------------------------|
| 간행         | 1970년 7월 1일 인쇄, 7월 25일 발행 | 2008년 6월 25일 발행         |
| 저작 표기      | 저작검발행 하영기, 편집 이일해         | 편찬자 하겸진, 역주자 이기동 외      |
| 발행처        | 해동불교역경원 내 일봉정사            | 나남                      |
| 체제와 수록 인원  | 상중하 3편 1질, 152인           | 6권, 23편과 보유편, 167인      |
| 판식         | 신연활자본, 선장                 | 현대 활자본, 양장              |
| 표점         | 없음                        | 있음                      |
| 교감기        | 없음                        | 없음                      |
| 각주         | 없음                        | 교감주, 어석주, 출전 확인주, 역자 보충 |
| 번역문        | 없음                        | 있음                      |
| 유통         | 비매품                       | 시판                      |
| 서(序)       | 후서                        | 머리말, 최일범 해제             |
| 특징적 결락     | 오기 다수                     | 본문 확정 근거의 미기록           |
| 본 연구에서의 지위 | 대교 자료                     | 저본                      |
| 확보 범위      | 3편 전량                     | 6권 가운데 1권               |

### 3. 구성과 서술의 사례 분석: 제2편 이학연원제유학안

앞의 두 절이 장르와 판본의 수준에서 서술한 특성은 본문에서 실제로 어떤 모습으로 나타나는가. 이 절은 한 편의 사례를 통하여 이를 확인한다. 대상은 제2편 이학연원제유학안이다. 이 편은 여덟 인물을 수록하면서 그 가운데 셋에게만 저술의 발체를 덧붙이므로, 인물의 수록과 저술의 발체가 별개의 편찬 결정이라는 사정이 한 편의 범위에서도 확인된다(하겸진 2008, 33-34쪽). 아래에서는 이 편의 첫머리에 놓인 총론을 진술 단위로 나누어 검토한다. 총론은 학파의 연원과 성격을 규정하여 이후 인물들의 배치를 해석하는 상위의 틀이면서도(최일범 2008, 15-16쪽) 목차에는 항목으로 나타나지 않으므로, 구조 표지의 부여가 인코더의 작업으로 넘어오는 첫 지점이기도 하다. 이때 부여되는 표지는 문헌에 이미 있던 것을 옮긴 것이 아니라 인코더가 새로 세운 것이므로, 그 사실 자체가 데이터에 기록되어야 한다.

#### 1) 총론의 진술 단위 및 분류

이 편의 총론은 열한 개의 진술 단위로 나뉜다. 분절의 기준은 문장 부호가 아니라 발화 주체의 전환과 진술 유형의 전환이며, 표점만으로는 이 전환이 표시되지 않는다는 사실 자체가 첫 번째 관찰이다. 아래에 원문과 번역을 나란히 제시한다.

1 儒宗錄有曰。吾東之學出於圃隱。圃隱傳之治隱。治隱再傳而至畢齋。自畢齋而後眞儒輩出蔚然。爲道學嫡傳。  
→ 『유종록』에 이르기를, 우리 동방의 학문은 포은에게서 나왔고, 포은이 야은에게 전하였으며, 야은이 다시 전하여 필재에 이르렀다. 필재 이후로 참된 선비가 무리로 나와 성대하게 도학의 적전을 이루었다고 하였다.

2 圃隱之爲理學祖者。蓋以此也。→ 포은이 이학의 조(祖)가 되는 것은 대개 이 때문이다.

3 寒暄一蠹學於畢齋。慕齋靜庵學於寒暄。→ 한훤당과 일두는 필재에게 배웠고, 모재와 정암은 한훤당에게 배웠다.

4 若晦齋河西則不由師承而得者。→ 회재와 하서 같은 이는 사승을 거치지 않고 스스로 얻은 사람이다.

5 而晦齋初學于其舅孫愚齋仲喈。河西初學於慕齋。愚齋亦畢齋之徒也。→ 그러나 회재는 처음에 그 외숙인 우재 손중돈에게 배웠고, 하서는 처음에 모재에게 배웠으며, 우재 또한 필재의 문도이다.

6 按退溪嘗稱畢齋爲詞華之人而不稱其道學。→ 살피건대 퇴계는 일찍이 필재를 문사의 사람이라 일컫고 그 도학은 일컫지 않았다.

7 然其答明使許國巍時亮之問則曰道學文章爲一時儒宗。→ 그러나 명나라 사신 허국과 위시량의 물음에 답할 때에는 도학과 문장이 한 시대의 유종이 된다고 하였다.

8 又按寒暄請學尙書于畢齋。畢齋手抽小學授之。→ 또 살피건대 한훤당이 필재에게 『상서』 배우기를 청하자 필재가 손수 『소학』을 꺼내어 주었다.

9 曰。苟志於學。宜從此始。→ 이르기를, 진실로 학문에 뜻을 두었다면 마땅히 이것부터 시작해야 한다고 하였다.

10 寒暄之以小學君子見稱於世。蓋畢齋有以啓之也。→ 한훤당이 세상에서 소학군자로 일컬어진 것은 대개 필재가 열어 준 바가 있었기 때문이다.

11 後之人以退溪一時之言而輕議畢齋豈知言者哉。→ 뒷사람들이 퇴계의 한때의 말을 가지고 필재를 가벼이 논하니, 어찌 말을 아는 자이겠는가.

이와 별도로 저본에는 두 개의 각주가 붙어 있다. 하나는 1의 『유종록』을 청대 인물의 저작으로 비정한 출전 확인 주이고, 다른 하나는 5의 손중돈에 대한 생몰년과 이력의 보충주이다.

| 번호   | 발화 주체           | 진술 유형         | 출처 표시                | 표상의 쟁점                                 |
|------|-----------------|---------------|----------------------|----------------------------------------|
| 1    | 『유종록』의 원저자      | 도통 계보의 진술     | 서명과 도입어를 갖춘 명시적 인용   | 인용의 종료 지점이 원문에 표시되지 않음                 |
| 2    | 하겸진             | 앞의 인용에 근거한 판정 | 없음                   | 인용문의 일부인지 편찬자의 판정인지가 표점만으로 결정되지 않음     |
| 3    | 하겸진             | 사승 관계의 진술     | 없음                   | 무표지 진술이 데이터에서 확정 사실로 고착되는 전형적 사례       |
| 4    | 하겸진             | 사승의 명시적 부정    | 없음                   | 명시적 부정 진술을 어느 유형으로 기록할 것인지             |
| 5    | 하겸진             | 초학 관계의 진술     | 없음                   | 4와의 긴장을 모순으로 볼 것인지 사승 개념의 층위 차이로 볼 것인지 |
| 6    | 하겸진이 전하는 퇴계의 평가 | 안어 안의 요약 인용   | 인명만 있고 출전 없음         | 요약 인용의 원문 범위를 특정할 수 없음                 |
| 7    | 하겸진이 전하는 퇴계의 발화 | 안어 안의 직접 인용   | 인명과 발화 상황은 있으나 서명 없음 | 원출전의 비정이 인코더의 작업으로 남음                  |
| 8    | 하겸진             | 일화의 서술        | 서명만 언급되고 인용문 없음      | 서명 언급을 저작에 대한 지시로 볼 것인지의 판단            |
| 9    | 김종직             | 직접 발화의 인용     | 화자 표시만 있고 출처 없음      | 출처 미상 인용의 확실성 표시                       |
| 10   | 하겸진             | 인과적 판정        | 없음                   | 평가적 판정이 서술문의 형식을 취함                    |
| 11   | 하겸진             | 후대 독법에 대한 논박  | 없음                   | 수사 의문으로 표현된 평가의 기록 방식                  |
| 각주 1 | 2008년 정리 층위     | 출전 확인 주       | 해당 없음                | 후대의 비정을 승계할 것인지 별도 층위로 기록할 것인지         |
| 각주 2 | 2008년 정리 층위     | 인물 보충 주       | 해당 없음                | 저본에 없는 정보의 층위 표시                       |

## 2) 진술 단위에서 도출되는 문제점

이 표에서 다섯 가지 문제가 도출된다.

첫째는 인용 경계의 미표시이다. 서두는 서명과 도입어를 갖춘 명시적 인용으로 시작하지만, 뒤이어 오는 판정 문장에는 아무런 표지가 없어 인용의 종료 지점이 원문에 표시되지 않는다. 2를 인용문의 일부로 읽을 것인가 편찬자의 판정으로 읽을 것인가는 표점만으로 결정되지 않는다. 이 미결이 인코딩에서 어떤 형태로 나타나는지는 두 판독을 나란히 적어 보면 드러난다.

```
<!-- 판독 가: 2를 인용문의 일부로 보는 경우 -->
<quote source="#yujongnok">吾東之學出於圃隱。.....爲道學嫡傳。
圃隱之爲理學祖者。蓋以此也。</quote>

<!-- 판독 나: 2를 편찬자의 판정으로 보는 경우 -->
<quote source="#yujongnok">吾東之學出於圃隱。.....爲道學嫡傳。</quote>
<seg type="judgement" resp="#hagyeomjin">圃隱之爲理學祖者。蓋以此也。</seg>
```

두 판독은 서로 배타적이고, 원문에는 어느 쪽인지를 정해 주는 표지가 없다. 그럼에도 인코딩은 `quote` 요소를 닫는 위치를 반드시 정해야 하므로 이 지점에서 판정을 미룰 수 없다. 판정을 미룰 수 없다는 사정이 곧 그 판정을 인코더의 것으로 명기해야 하는 이유가 된다. 위 표기에서 `resp` 속성이 그 명기의 최소 형태이며, 판정의 근거와 확신의 정도까지 적으려면 별도의 장치가 필요하다. 덧붙여 2008년본 번역이 인용부호로 그 경계를 이미 확정하고 있다는 사실은 경계 판정이 후대 층위에서 한 차례 수행되었음을 보여주며, 한문 원문을 인코딩 대상으로 삼는 본 연구가 그 판정을 승계할 것인지 별도 층위로 기록할 것인지를 결정해야 하는 지점이다.

둘째는 표점 자체의 불균질이다. 구조적으로 동일한 두 인용 도입부가 서로 다르게 처리되어 있다. 9의 김종직 발화는 도입어 뒤에 표점을 두어 인용의 시작을 표시한 반면, 7의 퇴계 발화는 도입어와 인용 내용이 표점 없이 이어진다. 또한 11은 앞의 서술절과 뒤의 수사 의문 사이에 표점이 없어 하나의 연속체로 제시된다. 표점이 진술 단위의 경계를 표시하는 장치가 아니라는 사실, 나아가 그 부여 방식조차 한 단락 안에서 일관되지 않다는 사실이 여기서 확인된다. 표점을 구조 표지로 신뢰할 수 없다는 이 관찰은 앞 절에서 세운 비동일시 원칙이 왜 필요한지를 구체적으로 뒷받침한다.

셋째는 출처 표시 정밀도의 편차이다. 같은 총론 안에서 서명과 인용문이 함께 제시된 경우, 인명과 발화 상황은 있으나 서명이 없는 경우, 서명만 있고 인용문이 없는 경우, 화자 표시만 있고 출처가 전혀 없는 경우가 모두 나타난다. 조선시대 논어 주석을 대상으로 인용 정보의 제시 방식을 세 등급으로 구분한 작업(김서윤 2024a)이 이 편차를 기술하는 좌표로 유효하다. 동시에 확인되는 것은, 경전 주석에서 주석 단위마다 나뉘던 그 등급이 학안의 총론에서는 한 단락 안에서 연속적으로 교체된다는 점이다. 등급의 판정 단위가 텍스트 구획이 아니라 진술 단위여야 하는 이유가 여기에 있다.

넷째는 관계 진술과 그 부정의 병존이다. 4는 사승을 명시적으로 부정하고 5는 그 직후에 초학의 스승을 밝힌다. 이것을 편찬자의 모순으로 처리할 것인지 사승 개념의 층위 구분으로 처리할 것인지는 인코더의 판단에 속하며, 어느 쪽을 택하든 그 판단이 데이터에 기록되어야 한다.

다섯째는 평가와 서술의 융합이다. 안이라는 표지로 시작하는 두 대목은 각각 다른 일을 수행한다. 6과 7은 퇴계의 두 발화를 대비시켜 앞선 평가를 조정하는 논증이고, 8에서 10에 이르는 대목은 일화의 제시를 통하여 인과적 판정을 도출하는 서술이다. 10의 판정은 평어의 형식이 아니라 서술문의 형식을 취하고 있으며, 총론의 마지막 문장인 11은 후대인의 독법을 논박하는 수사 의문으로 되어 있다. 평가가 독립된 평어로 분리되지 않고 서술의 형식 안에 녹아 있는 이 양상은, 무엇을 평가적 진술로 식별할 것인가가 다시 인코더의 판단이 된다는 뜻이다.

### 3) 저본의 불안정성과 후대 층위의 개입

앞의 다섯 문제가 편찬자의 서술 방식에서 오는 것이라면, 여기에 저본 자체의 상태에서 오는 문제가 더해진다. 이 짧은 총론 안에서만도 세 지점이 확인된다. 1에서 야은을 가리키는 표기가 한 문장 안에서 治隱과 治隱으로 서로 다르게 나타나고, 7에서 명나라 사신의 성이 魏가 아니라 巍로 되어 있으며, 2의 理學祖는 저본의 번역문에서 이학의 조종(祖宗)으로 옮겨져 원문의 자구보다 한 글자가 늘어나 있다. 앞의 둘은 이문으로 볼 것인지 오식으로 볼

것인지가 판단을 요하는 사안이고, 뒤의 하나는 번역 층위에서 이루어진 보충이다. 어느 것도 교감기에 기록되어 있지 않다. 저본을 그대로 전사하지 않고 대교를 거친 비평본을 구성해야 한다는 앞 절의 결론이 한 단락의 검토판으로도 즉시 실질을 얻는 셈이다. 야은의 표기를 예로 삼아 그 기록의 형태를 적어 보면 다음과 같다.

```
<app xml:id="app-yaeyun-01" type="orthographic">
  <lem resp="#encoder" cert="high">治隱</lem>
  <rdg wit="#ed2008" type="error">治隱</rdg>
  <witDetail wit="#ed1970" target="#app-yaeyun-01">대교 시점에 자형 확인 예정
</witDetail>
  <note resp="#encoder">저본은 한 문장 안에서 治隱과 治隱을 병용한다.
  앞 구절의 治隱과 동일 인물을 가리키므로 治隱을 오식으로 판정한다.</note>
</app>
```

이 장치는 저본에서 옮겨 온 것이 아니라 인코더가 새로 작성한 것이다. 저본에 교감기가 없기 때문이다. `lem`은 본문으로 채택한 독법이고 `rdg`는 채택되지 않은 독법이며, 이문으로 볼 것인지 오식으로 볼 것인지를 판정은 `type` 속성이 담는다. 주목할 것은 판정의 주체와 확신의 정도와 근거가 각각 `resp`와 `cert`와 `note`로 분리되어 있다는 점이다. 이 셋이 분리되지 않으면 뒷날 다른 연구자가 治隱을 오식이 아니라 이문으로 보아 다른 판단을 제시할 근거를 찾을 수 없다. 확인이 아직 수행되지 않은 대교 자료의 상태를 `witDetail`로 남겨 두는 것도 같은 취지이며, 미확인을 미확인으로 적는 자리가 데이터에 있어야 한다는 요구가 여기서 처음 구체적인 형태를 얻는다.

후대 층위의 개입은 각주에서 더 분명하게 나타난다. 2008년본은 『유종록』의 저자를 청대 인물로 비정하고, 손중돈의 생몰년과 이력을 각주로 보충한다. 하경진 자신은 서명만을 제시하였으므로 이 비정은 후대 정리 층위의 판단이며, 인코더가 이를 승계할 것인지 별도 층위로 기록할 것인지의 문제가 앞 절에서 세운 비동일시 원칙의 첫 적용 사례가 된다. 아울러 이 비정이 사실이라면 관찰 하나가 따라 나온다. 조선 도학의 연원을 여는 서술이 청대 문헌의 진술을 인용하며 시작한다는 사실은 그 자체로 편찬자의 계보 구성 방식을 보여주는 자료이다.

## 4) 종합 관찰

한 편의 검토에서 확인된 것을 정리하면 다음과 같다. 발화 주체의 무표지 전환, 인용 경계의 미표시, 표점 부여의 불균질, 출처 표시 정밀도의 편차, 관계 진술과 그 부정의 병존, 평가와 서술의 융합, 발체 유무의 차등, 저본 자구의 불안정, 후대 비정의 개입이 그것이다.

이 목록은 개별적인 난점의 나열이 아니다. 앞의 넷은 어느 진술이 누구의 것인가를 되물어 수 있어야 한다는 요구로 수렴하고, 발체 유무의 차등은 발체를 인물 정보의 부속물이 아니라 독립된 개체로 다루어야 한다는 요구로, 평가와 서술의 융합은 편찬자의 판단을 사실 진술과 같은 지위로 기록하지 않아야 한다는 요구로, 저본 자구의 불안정과 후대 비정의 개입은 확정과 추정을 구분하여 기록해야 한다는 요구로 이어진다. 이 요구들을 명시적인 목록으로 진술하는 작업은 다음 장의 첫머리에서 수행한다.

# III. 방법론: 선행 연구와 표상 모델의 검토

## 1. 검토 기준: 모델링 요건

앞의 장은 『동유학안』이 인용된 원저자의 진술과 편찬자의 결정이 겹쳐 이루어진 지식 체계이며, 그것을 디지털로 표상하는 순간 인코더의 해석이 한 층위 더해진다는 상황을 확인하고, 기존의 인물 데이터 구축이 이러한 상황에서 겪는 문제를 정리하였다. 즉 진술이 확정된 사실처럼 유통되는 문제, 데이터에서 원문으로 되돌아갈 수 없어 반증이 불가능해지는 문제, 귀속의 연쇄가 여럿 겹치는 문제가 그것이다. 이 장은 그 문제들에 응답하는 표상 방법

을 선행 연구에서 찾는다. 그런데 응답을 찾으려면 먼저 무엇을 응답으로 인정할 것인지가 정해져 있어야 한다. 그 기준이 없으면 검토는 각 모델의 성격을 차례로 소개하는 나열이 되고, 특정 모델을 택한 뒤에 그 모델이 잘하는 일을 기준으로 삼아 선택을 설명하는 순환에 빠진다. 이하의 요구사항 목록은 이 두 위험을 함께 막기 위한 중간항이다. 자료의 성격이 한쪽 끝에 있고 모델의 선택이 다른 쪽 끝에 있을 때, 그 둘을 잇는 언어가 요구사항이다.

목록의 도출은 두 곳에서 이루어진다. 하나는 위에서 정리한 세 문제이고, 다른 하나는 앞 장에서 확정한 학안체의 서술 형식이다. 후자가 중요한 것은 적용 범위 때문이다. 요구사항을 하검진본의 표본 관찰에서 끌어내면 그 목록은 이 한 문헌에만 타당한 명세가 되지만, 학안이라는 장르가 공유하는 형식에서 끌어내면 같은 형식을 가진 다른 문헌에도 그대로 적용된다. 학안은 학자의 전기와 그 학자의 저술에서 선택한 발췌문을 결합하는 형식이며, 전기와 발췌의 결합 여부가 이 장르를 인접 장르와 구분하는 기준이다(Chu 2007, 244쪽, 247쪽). 여기에 학파 단위의 편차와 편찬자의 평의가 더해진다. 인물의 선정과 학파별 배속, 저술의 발췌와 배치, 그리고 평가가 모두 편찬자의 판단이라는 점(최일범 2008, 11-14쪽)은 하검진 개인의 방식이 아니라 이 형식이 요구하는 바이며, 계보 저술이 전통의 수동적 기록이 아니라 특정 시점의 목적에 따라 능동적으로 구성된 표상이라는 관점(Wilson 1994, 4-6쪽)이 그 이유를 설명한다. 이하의 다섯 요건은 이 형식에서 도출된다.

**첫째, 발화 주체와 편찬 층위의 구분.** 각 진술에 대하여 그것이 원저작의 발화인지, 편찬자의 서술인지, 인코더의 판단인지를 되물어 수 있어야 한다. 이 요구는 학안의 서술이 어느 부분에서나 인용과 자작을 섞는다는 데서 발생한다. 총론은 선행 문헌의 계보 진술을 옮겨 오면서 그 사이에 편찬자의 안어를 끼워 넣고, 인물 항목은 행장이나 묘갈 계통의 서술을 옮기면서 출전을 밝히지 않으며, 저술 발췌는 원저작의 문장을 그대로 싣는다. 이 요건의 충족 여부는 임의의 진술 하나를 골라 그 발화 주체를 묻는 조화가 답을 제공하는가로 판정한다. 이 요건의 충족을 어렵게 만드는 상황은 표지(標識)의 불균질성이다. 서명과 문장이 함께 주어진 인용이 있는가 하면 인명만 있는 인용이 있고, 아무 표지 없이 흘러가는 서술이 있다. 표지가 없는 자리에서는 발화 주체 자체가 미결이므로, 미결을 미결로 적는 통로가 함께 마련되어야 이 요구가 실질적으로 충족될 수 있다.

**둘째, 발췌된 저작의 독립적 개체화.** 발췌의 대상이 된 원저작을 독립된 개체로 기술하고, 그것을 옮겨 온 행위와 구분할 수 있어야 한다. 이 요구는 인물의 수록과 그 인물의 저술 발췌가 별개의 편찬 결정이라는 데서 발생한다. 두 결정이 별개라는 사실은 계량에서 확인되는데, 『동유학안』에 수록된 인물 가운데 문헌자료가 함께 실린 경우는 일부에 그친다(김기주 2019, 168-169쪽). 발췌를 인물 정보에 부속된 자료로 처리하면 이 차등 자체가 데이터에서 사라진다. 이 요건의 충족 여부는 하나의 발췌문에 대하여 그것이 무엇에서 왔는가와 누가 그것을 옮겼는가를 각각 물었을 때 서로 다른 답이 돌아오는가로 판정한다. 충족을 어렵게 만드는 조건은 원저작의 지위가 균일하지 않다는 점이다. 현전하는 문집에서 온 발췌가 있는가 하면 원저작이 산일되어 이 편찬물의 인용을 통해서만 존재가 확인되는 경우가 있으므로, 개체화가 현전(現傳) 여부에 좌우되지 않아야 한다.

**셋째, 편찬자의 판단임을 명시하는 표현.** 평가적 서술과 학파 규정을 사실 진술과 같은 지위로 기록하지 않아야 한다. 이 요구는 학안이 학술사를 서술하는 방식 자체에서 발생한다. 어느 인물을 어느 학파에 넣는가는 관찰된 사실의 기록이 아니라 편찬자가 세운 분류의 결과이며, 편의 배열과 안어도 마찬가지이다. 나아가 평가는 명시적 평어 없이도 수행된다. 조선의 계보 문헌에서 칭호의 부여 여부나 성의 생략 여부 같은 표기 관행 자체가 존송의 등급을 표현해 온 사정(김기주 2022, 21쪽)은 학안에서 발췌의 유무와 분량과 배치가 같은 기능을 수행하는 배경이 된다. 이 요건의 충족 여부는 특정 인물이 특정 학파에 속한다는 기록을 조회하였을 때 그것이 누구의 규정인가가 함께 돌아오는가로 판정한다. 이 요건의 충족을 어렵게 만드는 조건은 평가와 서술의 융합이다. 평가가 별도의 문장으로 분리되지 않고 서술의 어휘 선택에 녹아 있는 경우, 어디까지를 평가로 볼 것인가가 다시 인코더의 판단이 된다.

**넷째, 불확실성과 추정의 표시.** 출처가 밝혀지지 않은 인용, 인코더의 비정, 연대가 확인되지 않은 정보를 확정 진술과 구분하여 기록할 수 있어야 한다. 이 요구는 셋째 요구의 마지막 조건과 첫째 요구의 마지막 조건이 만나는 자리에서 발생한다. 출처 표시의 정밀도가 총을 이룬다는 사정은 조선시대 경전 주석을 대상으로 이미 유형화된 바 있으며, 인명이나 서명과 인용 문장이 함께 제시된 경우, 인명이나 서명은 없으나 문장이 있어 원전 추적이 가능한 경우, 인명이나 서명만 있고 문장이 없거나 몇 글자로 축약된 경우의 셋이 구분되었다(김서윤 2024a). 정밀도가 낮은 자리일수록 인코더의 추정이 개입하므로, 그 추정을 사료의 진술과 같은 지위로 적으면 확실성의 등급

이 데이터에서 사라진다. 총족 여부는 확정 진술만을 걸러 내는 조희가 성립하는가로 판정한다. 총족을 어렵게 만드는 조건은 등급의 설정이다. 불확실성을 몇 단계로 나눌 것인가는 어휘가 아니라 자료의 실제 분포가 정할 사항이다.

**다섯째, 텍스트 위치 참조의 다중성 수용.** 하나의 텍스트 단편이 둘 이상의 위계에 동시에 속하는 경우를 지시할 수 있어야 한다. 이 요구는 발체라는 행위가 만들어 내는 이중 소속에서 발생한다. 인물 항목 뒤에 실린 발체문은 원저작의 텍스트 위계에 속하는 동시에 편찬물의 서술 위계에 속하며, 어느 한쪽으로 환원되지 않는다. 앞의 소속을 지우면 발체라는 행위가 사라지고 뒤의 소속을 지우면 배치라는 행위가 사라진다. 구조적으로 같은 문제가 판본 연구에서 이미 다루어진 바 있는데, 논리적 텍스트 위계와 물리적 페이지 위계가 교차하는 지점을 별도의 개체로 설정하는 해법이 그것이다(Boot and Koolen 2021, 11쪽, 22쪽). 총족 여부는 하나의 발체문에서 원저작 쪽 좌표와 편찬물 쪽 좌표를 모두 얻을 수 있는가로 판정한다. 총족을 어렵게 만드는 조건은 교차 지점이 텍스트에 명시적 단위로 존재하지 않는다는 사실이다. 인용문의 경계와 다른 종류의 구간이 어긋날 때마다 교차 요소를 반복 분할해야 하는 부담이 여기서 발생한다.

| 요구                | 진술의 요지                              | 판정의 방법                            |
|-------------------|-------------------------------------|-----------------------------------|
| 발화 주체와 편찬층위의 구분   | 진술마다 누구의 발화인지를 되물을 수 있을 것           | 임의의 진술에 대한 발화 주체 조희가 답을 돌려주는가     |
| 발체된 저작의 독립적 개체화   | 원저작과 발체 행위를 나누어 기술할 것               | 무엇에서 왔는가와 누가 옮겼는가에 서로 다른 답이 돌아오는가 |
| 편찬자의 판단임을 명시하는 표현 | 평가와 분류를 사실 진술과 같은 지위로 두지 않을 것       | 학파 귀속 조희에 규정의 주체가 함께 돌아오는가        |
| 불확실성과 추정의 표시      | 추정과 확정을 구분하여 기록할 것                  | 확정 진술만을 걸러 내는 조희가 성립하는가           |
| 텍스트 위치 참조의 다중성 수용 | 하나의 텍스트 단편이 두 위계(원문/발체문)에 속함을 지시할 것 | 발체문에서 두 위계의 좌표를 모두 얻을 수 있는가       |

## 2. 표상 모델의 검토: 팩토이드, 단언적 판본, 대안들

### 1) 사실(fact)에서 진술(statement)로: 디지털 집단전기학과 팩토이드 모델

『동유학안』의 인물 항목을 데이터로 옮기는 일은 다수 인물에 관한 기록을 모아 구조화한다는 점에서 집단전기학(prosopography)의 작업에 속한다. 집단전기학은 개인의 삶 자체가 아니라 다수의 개인에 관한 자료의 총합을 분석하여 인물들 사이의 연결 유형을 규명하고, 그로부터 그들이 당대의 제도 안에서 어떻게 활동했는가를 밝히는 방법으로 규정된다(Keats-Rohan 2000, 2쪽). 이 규정에서 중요한 것은 분석의 단위가 개인이 아니라 기록의 집합이라는 점이며, 그 집합이 연구가 진행될수록 확장되는 끝없는 카드 색인이라는 점이다. 집단전기학을 카드 색인을 만들어 가는 과정으로, 하나의 집단전기를 그 산출물로 구분한 설명은 이 방법이 인쇄본보다 데이터베이스에 친화적인 이유를 정확히 지적한다(Keats-Rohan 2000, 3-4쪽, 6-7쪽). 그런데 같은 글은 그 데이터베이스가 무엇에 근거해야 하는가에 관한 원칙도 함께 제시한다. 사료의 성격과 유형을 철저히 검토하기 전에 인물 목록부터 편찬하는 방식은 방법론적으로 타당하지 않으며, 전자 집단전기는 가능한 한 사료의 전문에 근거해야 한다는 것이다(Keats-Rohan 2000, 8쪽). 인물이 아니라 사료가 출발점이라는 이 원칙은 뒤에서 볼 팩토이드 모델의 단초이다.

사료에서 출발한다고 해서 해석이 사라지는 것은 아니다. 집단전기학의 실무를 정리한 안내서는 질문지를 대상 모집단 전체에 적용하는 작업 절차를 사료, 사료에 가까운 질적 기술, 양적 해석과 범주화의 세 단계로 도식화하고, 첫 단계에서 둘째 단계로 넘어가는 지점에 이미 날짜의 현대화나 데이터와 인물의 연결 같은 소규모 해석이 개입하며 셋째 단계에서 본격적 해석과 범주화가 일어난다고 명시한다(Verboven 외 2007, 57쪽). 아울러 모든 사실은 출전과 연결되어야 하고, 사료 속 인물의 식별은 자동 연결이 오류를 양산하므로 수작업으로 수행되어야 하며, 대상 집단은 실재하는 사회 집단이 아니라 연구자가 만들고 분석하는 집단이라는 경고를 덧붙인다(Verboven 외 2007, 51쪽, 57-59쪽). 이 안내서의 기여는 집단전기학이 사료의 진술과 연구자의 범주 사이의 거리를 처음부터 의식해 왔음을 절차의 언어로 보여 준 데 있다. 다만 그 거리는 연구자가 유의할 사항으로 남을 뿐, 데이터 구조 안에 기록해야 할 대상으로는 다루어지지 않는다. 질문지 방법론에서 데이터베이스의 행은 여전히 인물이며, 사료는 그 행을 채우는 근거이지 기록의 단위가 아니기 때문이다.

국내의 수용도 같은 지점에 머물러 있다. 집단전기학이라는 개념은 네이미어와 사임과 머턴의 업적을 통해 일찍이 소개되었고, 이때 이미 사례연구와 통계적 분석이 배타적이 아니라 상호 보완적이라는 점과 분류의 잘못이 이 방법의 고유한 위험이라는 점이 함께 정리되었다(나종일 1985, 36쪽, 41-43쪽). 최근의 한국학 적용 가능성 탐색은 여기서 한 걸음 더 나아가 조선시대의 집단전기 자료 자체를 유형화하였는데, 특정 범주의 인물 전체를 망라하려 한 인물지와 편찬자의 판단 아래 특정 기준에 맞는 인물을 추려 수록한 인물전을 구분하고, 전자를 유형으로 후자를 정선으로 규정한 점이 주목된다(우경섭 2024, 190쪽). 사우록과 문인록과 연원록 계통의 자료를 학문적 네트워크와 도통론의 강화를 보여 주는 자료로 자리매김한 것도 이 글이다(우경섭 2024, 193쪽). 이 유형화는 본 연구의 대상이 어느 쪽에 속하는가를 분명히 해 준다. 학안은 망라가 아니라 정선이며, 따라서 학안을 집단전기학의 자료로 삼는 순간 데이터에 담아야 할 것은 인물에 관한 정보만이 아니라 그 인물을 뽑고 배치한 편찬자의 판단 자체가 된다. 다만 두 글 모두 방법론의 소개와 자료의 유형화에 집중하였고, 그 판단을 어떤 데이터 모델로 기록할 것인가의 수준까지는 내려오지 않았다. 후자가 향후 과제로 문집 속 편지와 행장과 묘지의 수발신 관계를 연동하는 정밀한 데이터베이스 구축을 든 것은 그 공백을 스스로 지목한 것으로 읽힌다(우경섭 2024, 207-208쪽).

그 공백을 데이터 구조의 수준에서 보완한 것이 팩토이드 모델이다. 킹스 칼리지 런던의 세 집단전기학 프로젝트가 서사 텍스트로부터 관계형 데이터베이스를 구축하면서 내린 결정은, 인물에 관한 기사를 쓰는 대신 증거 데이터를 팩토이드(factoid)의 연쇄로 기록한다는 것이었다. 팩토이드는 자료 S의 위치 L에서 인물 P에 대해 무언가 F를 진술한다는 주장으로 정의되며, 인물에 대한 사실의 진술이 아니라 특정 지점의 자료가 인물에 대해 주장하는 바의 기록이다(Bradley and Short 2005, 7-8쪽). 이 전환의 의미는 무엇을 모델링하는가에 대한 자기 규정에서 가장 선명하다. 데이터베이스는 집단전기학의 기반이 되는 텍스트를 그 미묘한 의미와 함께 모델링하는 것이 아니라 집단전기학자가 텍스트를 해석하는 과업을 모델링한다는 것, 곧 역사 텍스트의 모델이 아니라 집단전기학 자체의 모델이라는 것이다(Bradley and Short 2005, 17쪽). 이 규정은 두 가지를 동시에 확보한다. 하나는 한 자료가 어떤 인물을 아르메니아인이라 하고 다른 자료가 불가리아인이라 할 때 두 팩토이드가 모순된 채 공존한다는 점이고, 다른 하나는 팩토이드가 개별 자료가 실제로 말하는 바를 기록하면서도 그것을 수집한 연구자의 해석 행위를 여전히 대변한다는 점이다(Bradley and Short 2005, 8쪽, 17-18쪽). 팩토이드라는 명칭이 잡지에 실리기 전에는 존재하지 않던 사실을 가리키던 메일러의 조어를 의식적으로 전유한 것이며, 그 아이러니가 텍스트 자료에서 주장된 정보의 잠정성을 강조하려는 의도였다는 점은 모델의 설계자와 후속 실천자가 함께 증언한다(Pasin and Bradley 2015, 89쪽; Schwartz 외 2022, 4-5쪽). 사료가 인물에 대해 말하는 바를 기록함으로써 역사학자의 독해가 아니라 사료 자체를 우선시한다는 점에서 이 모델은 사료 주도적(source-driven)이라 불린다(Pasin and Bradley 2015, 89쪽). 본 연구의 핵심 질문이 요구하는 바, 곧 진술을 확정된 사실로 기록하지 않으면서 출처로 되돌아갈 수 있게 한다는 조건에 대한 최초의 구조적 해답이 여기에 있다.

이 모델이 온톨로지로 재정식화되는 과정에서 인코더의 자리가 명시되었다. 팩토이드를 시맨틱 웹에 맞게 해체한 결과, 팩토이드 하나를 도입할 때마다 그것을 생성하는 행위자, 그것에 영감을 주는 문서, 그것이 다루는 사태의 세 개체가 함축되며, 이 셋을 의미 있는 단위로 묶는 행위를 따로 세울 필요가 있다는 판단 아래 저자와 원천과 해석을 슬롯으로 갖는 문서-해석-행위가 도입되었다(Pasin and Bradley 2015, 90쪽). 이 행위를 CIDOC CRM의 속성 부여 개체와 연결함으로써 팩토이드가 누군가가 문서가 특정 사태에 대한 것이라고 말한다는 구조임이 형식적으로 표현되었고, 이로써 해석하는 주체가 사료와 별개의 개체로 데이터에 들어오게 되었다(Pasin and

Bradley 2015, 91-92쪽). 다만 같은 논문은 이 해석 행위가 문서와 사태 사이의 연결을 설정할 뿐 그 연결의 성격과 강도를 불특정한 채로 남긴다는 점을 미결로 남겼다(Pasin and Bradley 2015, 95쪽). 해석의 확실성을 어떻게 표시할 것인가라는 이 미결 사항은 불확실성과 추정의 표시라는 요구가 팩토이드 바깥에서 채워져야 할 이유를 설계자 자신의 언어로 예고한다.

팩토이드 모델의 클래스 구성은 이후 팩토이드 집단전기 온톨로지, 곧 FPO로 공개되었다. 이 온톨로지는 진술과 사료, 진술과 인물을 각각 직접 잇지 않고 두 개의 매개 클래스를 거치게 한다. 진술의 기록 단위인 `:Factoid`는 `:Source`와 직접 연결되지 않고, 어느 사료의 어느 지점에서 그 주장이 발견되는가를 기록하는 `:SourceCitation`을 거쳐 연결된다. 인물 쪽도 마찬가지로, `:Factoid`는 `:Person`과 직접 연결되지 않고 `:PersonReference`를 거치며, 이 참조 개체는 사료의 해당 지점에서 이름이 어떻게 적혀 있는가를 별도 속성으로 보존하고 필요한 경우 해당 인물이 그 팩토이드 안에서 맡은 역할을 `:Role`로 표시한다(Bradley FPO). 관계를 기술하는 팩토이드 하나를 이 온톨로지의 어휘로 적으면 다음과 같은 형태가 된다. 접두사와 개체 식별자는 설명을 위해 약식으로 두었고, 클래스와 속성 이름은 온톨로지 문서의 것을 그대로 썼다.

```
ex:factoid_001      a fpo:RelationshipFactoid ;
    fpo:sourcedFrom ex:citation_001 ;
    fpo:hasReference ex:ref_001 , ex:ref_002 .

ex:citation_001     a fpo:SourceCitation ;
    fpo:fromSource  ex:source_A ;
    fpo:hasPlaceInSource "제2편, 김종직 항목" .

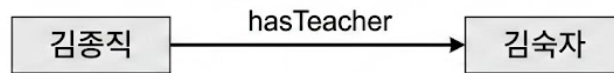
ex:source_A         a fpo:Source ;
    fpo:hasSourceName "東儒學案" .

ex:ref_001          a fpo:PersonReference ;
    fpo:hasNameInSourceAs "先生" ;
    fpo:hasRole      ex:role_primary ;
    fpo:referencesPerson ex:person_kimjongjik .

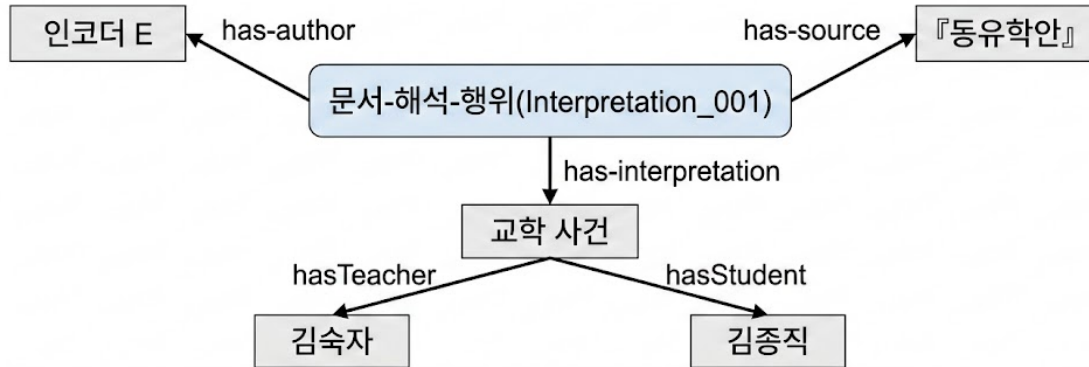
ex:ref_002          a fpo:PersonReference ;
    fpo:hasNameInSourceAs "家庭" ;
    fpo:hasRole      ex:role_secondary ;
    fpo:referencesPerson ex:person_kimsukja .
```

이 표기에서 주목할 것은 두 겹의 매개 구조이다. 사료 속의 표기와 인코더가 동일시한 인물 개체가 `:PersonReference`에 의해 분리되어, "先生"이나 "家庭"이라는 글자가 어느 인물을 가리키는가는 사료가 말한 것이 아니라 `:referencesPerson`이라는 별도의 연결로 기록된다. 진술과 그 진술의 출처 위치 역시 `:SourceCitation`에 의해 분리되어, 같은 사료의 다른 지점이 다른 주장을 할 때 두 주장이 같은 `:Source` 아래 별개의 인용 지점으로 공존한다. 사료가 무엇을 말했는가와 연구자가 그것을 누구에 관한 것으로 읽었는가를 별개의 노드에 두는 이 구조는, 각 진술에 대하여 그것이 누구의 발화인가를 되물어볼 수 있어야 한다는 본 연구의 첫째 요구에 대해 사료의 층위와 인코더의 층위를 처음으로 별개의 자리에 둔 구현이라 할 수 있다. 다만 여기서 분리되는 것은 그 두 층위뿐이며, 사료 자체가 다른 사료를 인용하고 있을 때 그 안쪽의 발화 주체를 둘 자리는 없다. 그 자리가 없는 이유는 아래에서 다시 다룬다. 아울러 이 온톨로지는 스스로를 완성된 모델이 아니라 확장을 전제할 핵심 구조로 규정하며, 개별 프로젝트가 하위 클래스와 하위 속성으로 이를 확장하는 것을 표준적 관행으로 제시한다(Bradley FPO). 본 연구가 아래에서 시도할 확장이 모델의 바깥이 아니라 모델이 스스로 열어 둔 경로 위에 있다는 점을 여기서 미리 적어 둔다.

## 단순 트리플 방식



## 팩토이드 모델



[도해 5] 팩토이드 모델 개념도

팩토이드의 현재형은 관계형 데이터베이스 밖에서도 확인된다. 시리아어 문헌을 대상으로 한 한 프로젝트는 맞춤형 스키마를 갖춘 TEI 파일에 팩토이드를 직접 인코딩하면서, 그 결과물이 원전 텍스트 자체의 모델이 아니라 그 텍스트로부터 도출된 집단전기적 팩토이드의 모델이라고 자기 규정한다(Schwartz 외 2022, 3쪽). 이 규정 아래 팩토이드는 인코더가 이해하고 주석을 단 원전 주장의 색인으로 이해되며, 그 산출물은 원전 텍스트와 그에 대한 역사비평적 평가 사이의 공간에 놓인다(Schwartz 외 2022, 5쪽). 이 프로젝트의 고유한 기여는 세 가지이다. 첫째, 텍스트와 데이터베이스를 대립시킨 이전의 논의가 XML의 한 용법인 텍스트 인코딩만을 고려한 것임을 지적하고, TEI로 원전이 아니라 팩토이드를 모델링함으로써 트리 구조의 한계를 우회하였다(Schwartz 외 2022, 7-8쪽). 둘째, 팩토이드 하나하나에 책임 인코더의 귀속을 필수로 부과하고, 원전의 주장에 대한 인코더의 의구심을 의심, 오류, 불확실의 세 유형으로 구분하여 표시하게 하였다(Schwartz 외 2022, 11-12쪽, 27-28쪽). 셋째, 원전의 전문은 자매 프로젝트의 코퍼스에 두고 팩토이드에서 표준 참조 식별자로 해당 구절을 가리킴으로써, 전통적 관행과 달리 원전의 인코딩이 오히려 팩토이드의 외부 주석처럼 기능하는 구조를 만들었다(Schwartz 외 2022, 23-25쪽). 판본과 데이터가 별개의 파일로 구축되고 참조로 연결되는 이 구성은 본 연구가 뒤에서 논증할 이원 구축(dual construction) 방침과 같은 형태이며, 그 방침이 팩토이드 계보 안에서 이미 실천된 바 있음을 보여 준다. 여기서는 그 사실의 확인에 그친다.

국내의 인물 데이터 설계는 이와 다른 경로로 같은 문제에 접근해 왔다. 중국의 인물 데이터베이스를 선행 모델로 삼아 기관별로 분절된 한국의 역사인물 정보를 연결하려는 온톨로지 설계는 인물을 중심 개체로 두고 이름과 생몰과 관계를 그 속성과 관계로 기술하는 개체 속성 중심의 구조를 취하면서도, 디지털 인물사전의 효용이 빠른 연산보다 정보를 판단하는 근거와 출처와 연구 성과를 집대성하는 데 있다는 점, 그리고 단일한 최종 판단을 내리는 종이사전의 전통에서 벗어나 서로 다른 연구자의 판단과 근거를 공존시키는 설계가 필요하다는 점을 분명히 하였다(김바로 2023, 170-172쪽). 여향전기집을 대상으로 한 모델링은 인물 중심의 표 구조 대신 관계 중심의 시맨틱 모델을 구성하고, 인물뿐 아니라 저술과 기록과 사건을 별도의 클래스로 두어 기록과 인물 사이의 언급 관계를 표현할 수 있게 하였다(김지선·류인태 2022). 가장 최근의 작업은 한국의 인물 정보가 단일 주체 없이 여러 기관과 연구자에게 산재해 있다는 조건에서 출발하여, 역사적 사실을 기술할 때 그것이 어느 기관과 어느 사료와 어느 연구자의 해석에 기반하는가를 명기하는 출처와 확실성의 표기를 핵심 난제로 제기하고, 이를 위해 표준 구체화가 요구하는 구조가 지나치게 장황함을 지적하였다(김바로·이재열 2025, 1쪽, 12-13쪽). 그 해법은 RDF-star로 개별 트리플에 출처와 확실성을 직접 부착하고, 편찬자별 URI 네임스페이스로 해석의 판본을 분리하는 것이었다(김바로·이재열 2025, 13-16쪽, 19-20쪽). 이로써 국내 논의는 트리플 단위의 출처 귀속과 편찬자 단위의 판본 분

리에 도달하였고, 본 연구의 핵심 질문이 묻는 출처 환원과 해석 책임의 문제를 국내 인물 데이터의 맥락에서 명시적으로 제기하였다.

본 연구는 이 작업의 문제의식을 이어받되 기록의 단위를 한 단계 더 내려간다. 그 이유는 대상 자료의 성격에 있다. 여러 기관의 사료와 연구 성과가 산재한 근대 인물을 대상으로 할 때에는 인물에 관한 주장을 기본 단위로 두고 각 주장에 어느 사료에서 왔는가를 부착하는 방식이 합리적이며, 출처가 서로 다른 문헌의 URI로 구별되므로 트리플 단위의 출처 표기만으로도 주장들의 근거가 구별된다. 그러나 『동유학안』은 하나의 편찬물이고, 그 안의 모든 진술이 같은 문헌에서 온다. 출처를 문헌 단위로 적으면 모든 트리플의 출처가 『동유학안』으로 수렴하여 구별의 기능을 잃는다. 구별이 필요한 것은 어느 문헌인가가 아니라 그 문헌의 어느 지점이 누구의 목소리로 무엇을 말하는가이다. 따라서 본 연구에서 기본 단위는 인물에 관한 주장이 아니라 사료의 특정 지점이 무엇을 말한다는 사실 그 자체, 곧 팩토이드가 되어야 하며, 사료의 발화와 연구자의 해석은 메타데이터와 주장의 관계가 아니라 별개의 노드로 기술되어야 한다. 트리플에 출처를 붙이는 방식과 진술을 개체로 구체화하는 방식 가운데 어느 것이 더 낫다는 일반론이 아니라, 단일 편찬물 안의 중층 귀속이라는 조건이 후자를 요구한다는 것이 본 연구의 출발점이다. 두 방식이 배타적이지 않으며 진술 개체에 다시 트리플 단위 메타데이터를 부착할 수 있다는 점은 뒤에서 다룬다.

이상의 검토를 『동유학안』에 적용하면 어떤 모습이 되는가를 1.2에서 도입한 김종직의 사례로 확인한다. 『동유학안』 제2편 이학연원제유학안의 김종직 항목은 그를 사에 김숙자의 아들로 소개한 뒤 "先生始受學於家庭"이라 서술한다(하경진 2008, 145쪽). 한문 원문은 아들이라는 가계 정보와 家庭에서 배웠다는 사승 정보를 별개의 문장으로 제시하며, 家庭이 누구를 가리키는지는 밝히지 않는다. 2008년 번역본은 이 구절을 "선생은 처음에 부친에게 학문을 배워"로 옮겨 家庭을 부친으로 확정하였다(하경진 2008, 145쪽). 번역 층위가 이미 하나의 비정을 수행하고 있는 셈이며, 이는 11.2에서 세울 편찬 층위 비동일시 원칙이 인코더의 비정과 번역자의 비정을 구분해야 할 이유를 보여 주는 사례이다. 한문 원문을 저본으로 삼는 본 연구에서 두 정보의 연결은 번역문의 확정을 그대로 받는 것이 아니라 인코더의 판단으로 기록된다. 인코더가 家庭을 부친 김숙자로 비정하여 사승 관계를 기록한다면, 앞의 코드 예시가 바로 그 팩토이드이다. `:Source`는 『동유학안』이고, `:SourceCitation`은 제2편 김종직 항목의 해당 구절이며, `:Factoid`는 사승 유형의 관계 팩토이드이고, 두 개의 `:PersonReference`가 사료 속 표기 "先生"과 "家庭"을 각각 김종직과 김숙자라는 인물 개체에 연결한다.

이 표기는 앞에서 확인한 장점을 모두 갖추고 있다. 진술은 사실이 아니라 『동유학안』의 특정 지점이 말하는 바로 기록되고, 家庭이라는 사료 속 표기는 김숙자라는 인물 개체와 분리되어 보존되며, 그 둘을 연결한 것이 인코더의 판단임이 드러난다. 다른 인코더가 家庭을 가학 일반으로 읽어 특정 인물과 연결하지 않기로 결정하더라도 그 판단은 오류가 아니라 또 하나의 팩토이드로 공존한다. 그러나 이 표기가 깔끔해 보인다는 사실 자체가 아래에서 문제가 된다. `:Source`의 자리에 『동유학안』을 적는 순간, 그 구절이 하경진 자신의 서술인지 그가 행장이나 묘지명 같은 원저작에서 옮겨 온 문장인지를 되물을 자리가 이 구조 안에는 없기 때문이다.

## 2) 디지털 판본학의 맥락: 단언적 판본(Assertive Edition)과 판본 도메인 온톨로지

앞 소절의 끝에서 확인한 공백은 팩토이드 모델이 사료를 진술의 출처로 다룰 뿐 사료 자체가 어떻게 구성되었는가를 묻지 않는 데서 온다. `:Source`는 진술이 발견되는 자리이지 그 자리가 다시 여러 발화로 짜여 있는가를 기술하는 개체가 아니다. 그런데 사료의 구성을 묻는 일은 본래 인물 데이터의 일이 아니라 판본학의 일이다. 따라서 네 소절을 관통하는 물음, 곧 각 진술에 대하여 누가 말하는가를 데이터가 되물을 수 있는가라는 물음은 인물 데이터의 자리에서 한 번 던진 것으로 끝나지 않고 판본의 자리에서 다시 던져져야 한다. 이 소절은 디지털 판본학이 같은 물음에 어떻게 도달하였는가를 단언적 판본 개념과 판본 도메인 온톨로지의 두 논의에서 확인하고, 본 연구가 팩토이드의 계보와 이 계보가 교차하는 지점에 놓인다는 것, 그리고 두 계보가 학안류라는 대상 앞에서 같은 지점에 멈춘다는 것을 보인다.

단언적 판본(assertive edition)<sup>1</sup>이라는 개념은 역사학자가 텍스트를 어떻게 보는가에 대한 진단에서 출발한다. 역사학자 역시 텍스트 비평이 만들어 낸 좋은 텍스트를 필요로 하지만, 그에게 텍스트의 언어적 국면과 물리적 국면은 텍스트가 전달하는 정보로 가는 매개일 뿐이며, 텍스트의 내용은 과거에 관한 지식을 얻기 위해 사용할 데이터로 간주된다(Vogeler 2019, 309쪽). 역사학자가 주로 다루는 문서고 기록이 애초에 부재하는 서거나 미래의 서기를 위해 행정 활동의 정보를 보존하려고 작성되었다는 사실, 달리 말해 종이에 쓰인 텍스트 안에 데이터를 저장한 것이라는 사실이 이 텍스트관을 뒷받침한다(Vogeler 2019, 309-310쪽).

포겔러에 따르면 단언적 판본은 전사(transcription)가 단언하는 사실에 관한 정보가 편찬 작업의 초점이 되는 역사 문서의 학술적 표상이며, 그 데이터는 맥락과 편찬자의 전문성에 근거한 텍스트 해석을 포함한다. 해석은 편찬자의 비평적 활동의 핵심에 속하고, 정규화와 범주화와 외부 자원 참조와 형식적 지식 표상이 모두 그 해석의 형식이다(Vogeler 2019, 318쪽). 이렇게 규정된 판본은 순수한 텍스트 표상과 특정 연구 질문을 중심으로 구조화된 정형 데이터베이스 사이의 어떤 것이며, 그 결과 판본과 데이터베이스의 경계가 흐려진다(Vogeler 2019, 312쪽, 318쪽). 여기서 확인해 둘 것은 이 정의가 편찬자의 해석을 판본에 딸린 부산물이 아니라 판본 데이터의 구성 요소로 인정한다는 점이다. 이는 앞에서 본 팩토이드의 온톨로지 재정식화가 해석하는 주체를 사료와 별개의 개체로 데이터에 도입한 것과 같은 방향의 결정이지만, 출발점이 반대이다. 팩토이드는 인물 데이터를 구축하려는 자리에서 사료로 내려가 진술의 조건부성을 발견하였고, 단언적 판본은 판본을 만드는 자리에서 데이터로 올라가 해석의 데이터화를 요구한다. 둘이 만나는 지점이 판본과 데이터베이스의 경계가 흐려지는 바로 그 자리이며, 본 연구는 이 교차점에 놓인다. 본 연구가 판본 중심이나 데이터 중심이나를 택하는 대신 둘을 각각 구축하되 책임을 나눈다는 방침을 I부에서 미리 정해 둔 것은 이 교차점에서 나온 귀결이다.

다만 교차점에 선다는 것이 단언적 판본 논의의 제안을 그대로 받는다 뜻은 아니다. 포겔러는 단언적 판본의 실무적 해법으로 RDFa가 HTML에 RDF를 내장하듯 TEI에도 같은 방식을 도입할 것을 제안하되, 외부 네임스페이스를 TEI에 들이는 대신 텍스트 조각을 임의의 분석적 주석에 연결하는 전역 속성 `@ana`를 텍스트와 RDF 트리플 구조를 잇는 일반적 연결 지점으로 쓰자고 제안한다(Vogeler 2019, 316-317쪽). 그러나 같은 글은 곧바로 내장 마크업의 한계를 인정한다. 정보가 데이터와 회복 구조와 해석이 일어나는 시점의 함수라는 Langefors의 정식을 끌어와, 편찬된 텍스트의 의미론적 값이 편찬자의 해석인 이상 시점 의존적이며 따라서 내장보다 외재적 주석이 유리하다고 결론짓고, 편찬된 텍스트와 RDF 사이의 연결을 어떻게 유지할 것인가에 관한 표준적 해법이 아직 없다는 것을 미결로 남긴다(Vogeler 2019, 318쪽). 내장이냐 분리냐라는 이 미결이 TEI 판본과 RDF 지식 그래프를 병행 체계로 두고 출처화된 주장의 지점에서 연결하려는 본 연구 설계의 직접적인 선행 논쟁이다. 그 논쟁은 뒤에서 수행하며, 여기서는 단언적 판본 논의가 해석의 시점 의존성을 근거로 스스로 분리 쪽에 무게를 실었다는 사실만 확인한다.

단언적 판본이 판본과 데이터의 관계를 다시 규정하였다면, 판본 도메인 온톨로지(ontology of the editorial domain)는 판본 자체가 무엇으로 이루어졌는가를 그래프의 어휘로 진술한다. 이 온톨로지의 출발점은 상호운용 가능한 주석이었다. 이용자가 온라인 판본의 특정 위치에 주석을 붙이려면 판본이 텍스트와 텍스트 단편에 기계가독 식별자를 부여하는 데 그치지 않고 단편의 유형과 구조에 관한 정보도 제공해야 하며, 이를 위해 판본과 그 구성 요소를 기술하는 온톨로지가 필요하다는 것이다(Boot and Koolen 2021, 9-10쪽). FRBRoo와 CIDOC CRM을 확장하여 구성한 이 온톨로지는 두 개의 구별을 축으로 한다. 하나는 편집되는 대상에 관한 개념을 두는 편집 대상 도메인과 편집 과정의 결과인 디지털 텍스트와 이미지를 두는 판본 도메인의 구별이고, 다른 하나는 추상적인 지적 실체인 저작(work)과 그것을 예시하는 텍스트 담지체인 문헌(document)의 구별이다(Boot and Koolen 2021, 10-11쪽). 저작과 문헌은 각각 부분을 가질 수 있으므로 두 위계가 생기며, 두 위계가 교차하는 지점, 이를테면 특정 필사본의 한 면에 나타난 시의 일부는 위치화된 텍스트 단편(located text fragment)이라는 별도의 개체로 정의된다(Boot and Koolen 2021, 11쪽, 22쪽). 이 구조의 기여는 판본 안의 텍스트 단편이 어느 저작의 어느 부분이 어느 문헌의 어느 자리에 실현된 것인가를 개체 사이의 관계로 적을 수 있게 하고, 그 개체들이 프로젝트마다 확장 가능한 일반 온톨로지 아래 놓이게 한 데 있다(Boot and Koolen 2021, 11-12쪽).

이 어휘가 본 연구에 직접 닿는 지점은 앞에서 세운 둘째 요구사항, 곧 발췌된 원저작을 독립된 개체로 기술하고 발췌라는 행위와 구분할 수 있어야 한다는 요구이다. 『동유학안』의 인물 항목에 실린 발췌문은 행장이나 문집의 일부를 옮긴 것이므로, 저작과 문헌의 구별을 적용하면 발췌의 대상인 원저작을 저작으로, 『동유학안』에 실린 그 발췌문을 그 저작이 다른 문헌 안에 실현된 단편으로 개체화할 수 있다. 저작이 추상 개체라는 점은 여기서 실질적인 이점이 된다. 원저작이 현전하지 않거나 『동유학안』의 인용을 통해서만 그 존재가 확인되는 경우에도 저작은 문헌과 독립적으로 기술될 수 있으므로, 출처 미상이나 산일이 개체화를 막지 않는다. 아울러 위치화된 텍스트 단편이라는 발상은 하나의 단편이 둘 이상의 위계에 동시에 속한다는 사태를 개체의 수준에서 수용하는 선례이며, 발췌문이 원저작의 텍스트 위계와 편찬물의 서술 위계에 동시에 속한다는 다섯째 요구사항의 문제 구조를 정확히 예비한다. 다만 그 단편을 판본의 어느 지점에 어떻게 고정할 것인가라는 앵커링의 문제는 이 온톨로지가 제기한 것이면서 이 소절의 범위를 넘으므로 뒤에서 다룬다.

두 논의를 이렇게 읽고 나면 둘이 공유하는 공백이 드러난다. 어느 쪽도 편찬자의 인용 행위를 별도의 총위로 다루지 않는다. 단언적 판본이 상정하는 구도에서 발화하는 주체는 사료의 필자이고 해석하는 주체는 판본의 편찬자이며, 이 둘 사이에 제3의 자리는 없다. 디지털 이전의 계보에서 요지와 발췌를 만드는 이는 근대의 편찬자, 곧 판본을 만드는 사람 자신이었으므로 발췌는 사료 바깥에서 사료에 가해지는 행위였고, 그 행위의 주체는 해석의 주체와 일치하였다. 학안에서는 사정이 다르다. 발췌와 배치와 평어를 수행한 편찬자 하검진은 사료의 바깥이 아니라 안에 있으며, 그의 편찬 행위는 우리가 판본으로 만들려는 문헌 자체에 이미 응결되어 있다. 단언적 판본의 편찬자 자리에 들어가는 것은 본 연구의 인코더이지 하검진이 아니다. 하검진은 사료의 필자도 아니고 판본의 편찬자도 아닌 자리, 곧 이 구도에 슬롯이 없는 자리에 선다. 이 구도에서 단언하는 주체는 전사의 필자 하나이나, 학안에서는 전사가 단언하는 내용이 이미 다른 이의 단언을 옮긴 것이어서 누가 단언하는가가 데이터의 내용보다 먼저 구분되어야 한다. 이 글이 assertive의 역어로 단언을 택한 것은 바로 이 대비를 용어의 수준에서 유지하기 위해서이다. 단언적 판본이 전사가 단언하는 바를 데이터화하는 구조라면 본 연구가 요구하는 것은 단언과 그 단언의 인용을 구분하여 적는 구조이며, 이 대비는 뒤에서 단언된 트리플과 비단언 트리플의 구별을 논할 때 같은 어휘로 이어진다. @ana 하나로 텍스트 조각을 RDF 기술에 연결하는 제안이 인용문 앞에서 원저자의 진술과 편찬자의 인용 행위를 한 연결 지점에 겹쳐 놓게 된다는 점은 이 공백의 구현 총위에서의 모습이며, 그 처리는 뒤로 넘긴다.

판본 도메인 온톨로지의 경우 공백의 원인은 대상의 성격에 있다. 이 온톨로지가 시험된 반 고희 서신 판본에서 하나의 편지는 하나의 저작이고, 저작과 문헌의 교차는 같은 저작이 원어 텍스트와 번역과 필사본 면이라는 여러 담지체에 실현되는 교차이다(Boot and Koolen 2021, 12-13쪽, 22쪽). 위계가 둘이더라도 저작은 하나이다. 학안에서 발췌문이 속하는 두 위계는 같은 저작의 두 담지체가 아니라 서로 다른 두 저작, 곧 원저작과 『동유학안』이다. 발췌문은 원저작의 부분이며 동시에 『동유학안』이라는 또 다른 저작의 부분이며, 이 이중 소속을 만들어 낸 것은 발췌라는 행위이다. 저작의 부분 관계를 적는 속성으로 이를 표현하려면 원저작이 『동유학안』의 부분이라고 적어야 하는데, 그것은 사실이 아니다. 두 저작을 잇는 것은 부분 관계가 아니라 한 저작의 편찬자가 다른 저작을 선택하여 옮긴 행위이며, 이 온톨로지에는 그 행위를 담은 클래스가 없다. 저작과 문헌의 구별이 요구사항 2의 전반, 곧 인용된 저작의 독립 개체화에는 답하지만 후반, 곧 발췌 행위와의 구분에는 답하지 못하는 이유가 여기에 있다.

그러므로 두 계보는 같은 곳에서 멈춘다. 팩토이드는 사료의 총위와 인코더의 총위를 갈랐고, 단언적 판본은 편찬자의 해석을 판본 데이터의 일부로 인정하였으며, 판본 도메인 온톨로지는 저작과 문헌과 단편을 개체로 정의하였다. 셋 모두 발화 주체와 해석 주체를 각각 하나씩 상정하는 구도 위에 있고, 그 구도는 문서고 기록과 서신이라는 각자의 대상에 대해서는 충분하였다. 사료 안에 이미 편찬자가 있는 문헌, 곧 원저자 위에 편찬자가 겹치고 그 위에 인코더가 다시 겹치는 문헌은 이들 대상에 없었다. 이 공백은 모델의 결함이 아니라 대상의 차이에서 온 조건이며, 학안류가 추가로 요구하는 것이 무엇인가를 따지는 일은 다음 소절의 몫이다. 앞에서 팩토이드 하나로 깔끔하게 적혔던 김종직의 사승 서술을 다시 펼쳐, 단일 귀속 구조로 처리했을 때 무엇이 남는가를 거기서 확인한다.

- 
1. "단언적 판본"의 역어 선택: assertive edition은 논리학의 assertion, 곧 어떤 명제를 참으로 내세우는 언어 행위에서 온 조어이다. 저자 자신도 2018년 디지털 인문학 학술대회 발표 초록에서 이 판본 유형에 확립된 명칭이 없음을 전제하고 semantic edition, factual edition, assertive edition을 병렬 후보

로 제시한 바 있으므로 원어 역시 확정된 관용어는 아니다. 이 글이 "단언적"을 선택한 이유는 두 가지이다. 첫째, 단언은 논리학과 언어철학에서 assertion의 표준 번역어이며 RDF에서 asserted triple을 단언된 트리플로 옮기는 관행과 일치한다. 둘째, 이 번역어를 취하면 단언과 비단언의 대립을 한국어 안에서 그대로 이어받을 수 있어 RDF 1.2의 트리플 형과 구체화를 논할 때 별도의 용어를 도입할 필요가 없다. "단언적 판본"을 대신할 수 있는 후보는 "진술적 판본"이다. 어감이 자연스럽고 statement와의 대응도 무난하지만 "assertion"을 "단언"으로 번역하는 관례를 더 중시하여 채택하지 않았다. "주장적 판본"은 주장이 논변이나 개인적 견해를 함의하여 assertion보다 claim에 가깝고, 이 용어가 사용된 원래의 맥락과 가장 관련이 적다. ↩

### 3) 학안의 특수성: 단일 귀속 구조의 한계

앞에서 검토한 세 모델은 서로 다른 자리에서 출발하였으나 하나의 구도를 공유한다. 말하는 사료가 있고 그것을 읽는 연구자가 있으며, 데이터는 이 둘을 구분하여 적는다는 구도이다. 이 소절의 과제는 두 가지이다. 하나는 그 구도가 학안 앞에서 정확히 어느 지점에서 부족해지는가를 자료의 한 대목으로 확인하는 일이고, 다른 하나는 부족한 자리를 채우기 위해 모델을 확장하는 것이 방법론적으로 정당한 경로인지를 선례에서 확인하는 일이다. 앞의 작업이 본 연구의 문제를 규정하고, 뒤의 작업이 그 해결의 형식을 규정한다.

먼저 단일 귀속 구조가 무엇인지를 분명히 해 둔다. 팩토이드 모델에서 하나의 주장이 데이터로 들어오는 경로는 단일하다. `:Source`에서 `:SourceCitation`을 거쳐 `:Factoid`에 이르고, `:Factoid`에서 `:PersonReference`를 거쳐 `:Person`에 이른다. 이 경로에서 분기가 일어나는 지점은 둘뿐이다. 하나는 같은 사료의 여러 지점이 각각 다른 주장을 할 때 `:SourceCitation`이 나뉘는 지점이고, 다른 하나는 사료에 적힌 표기와 인코더가 동일시한 인물이 `:PersonReference`에서 나뉘는 지점이다. 어느 분기도 사료 안에서 말하는 자가 나뉘는 경우를 위한 것이 아니다. 하나의 `:Source` 아래 놓인 여러 `:SourceCitation`은 같은 문헌의 서로 다른 자리일 뿐 서로 다른 목소리가 아니며, 문헌의 자리와 발화의 주체를 구별할 속성은 이 골격에 마련되어 있지 않다. 이것은 모델의 결함이 아니라 그 모델이 상정한 대상에는 없던 조건이다. 팩토이드가 가장 잘 처리하는 상황은 한 사료가 어떤 인물을 아르메니아인이라 하고 다른 사료가 불가리아인이라 할 때 두 진술이 모순된 채로 공존하는 상황인데(Bradley and Short 2005, 8쪽), 이 강점 자체가 사료 하나에 목소리 하나가 대응한다는 전제 위에서 성립한다. 문서고 기록이나 행정 문서를 다루는 한 이 전제는 대체로 참이다. 팩토이드마다 책임 인코더의 귀속을 필수로 부과하고 인코더의 의구심을 세 유형으로 나누어 표시하게 한 최근의 확장에서도(Schwartz 외 2022, 11-12쪽) 나뉘는 것은 여전히 사료의 총위와 인코더의 총위 둘이며, 사료 안쪽은 나뉘지 않는다.

이 전제가 학안에서 성립하지 않음을 보이기 위해 제2편 이학연원제유학안의 첫머리에 놓인 총론을 본다. 이 대목은 『유종록』에 이르기를 우리 동방의 학문은 포은에게서 나왔고 포은이 야은에게 전하였으며 야은이 재전하여 필재에 이르렀고 필재 이후로 진유가 배출되어 도학의 적전이 되었으니 포은이 이학의 조종이 되는 것은 이 때문이라 하였다는 인용으로 시작한다(하경진 2008, 원문편 476쪽). 이어 한훤당과 일두가 필재에게 배웠고 모재와 정암이 한훤당에게 배웠으며 회재와 하서는 사승에 말미암지 않았다는 계보가 제시된 뒤, 하경진의 안어가 두 차례 붙는다. 첫째 안어는 퇴계가 일찍이 필재를 사학의 인물이라 칭하고 그 도학은 인정하지 않았으나 명의 사신 허국과 위시량의 물음에 답할 때에는 도학과 문장이 한 시대의 유종이라 하였다는 두 진술을 나란히 놓는다. 둘째 안어는 한훤당이 필재에게 『상서』 배우기를 청하자 필재가 『소학』을 꺼내어 주며 참으로 학문에 뜻을 두었다면 마땅히 여기서 시작해야 한다고 말하였다는 일화를 든 뒤, 한훤당이 소학군자로 세상에 알려진 것은 필재가 열어 준 데서 비롯되었다고 적는다. 그리고 후대의 사람이 퇴계의 한때의 말을 근거로 필재를 가볍게 논한다면 어찌 말을 아는 자이겠느냐는 판정으로 이 대목이 닫힌다.

한 대목 안에 발화의 층이 넷이다. 첫째는 『유종록』이 진술한 도통 계보이고, 둘째는 퇴계에게 귀속된 두 개의 상반된 평가이며, 셋째는 일화 속에서 필재가 한 말이고, 넷째는 이 셋을 골라 배치하고 판정한 하경진의 서술이다. 층마다 표지의 정밀도도 다르다. 첫째는 서명과 문장이 함께 주어졌고, 둘째는 인명은 주어졌으나 어느 문헌의 어느 대목인지가 밝혀지지 않았으며, 셋째는 인명도 문헌도 없이 상황과 발화만 있고, 넷째는 안이라는 표지로 자신의 것임을 명시한다. 이제 이 대목에서 김종직이 도학의 적전에 놓인다는 진술을 팩토이드 하나로 기록한다고 하자.

앞 소절과 같은 어휘로 적으면 다음과 같은 형태가 된다.

```
ex:factoid_002      a fpo:RelationshipFactoid ;
    fpo:sourcedFrom  ex:citation_002 ;
    fpo:hasReference ex:ref_003 .

ex:citation_002     a fpo:SourceCitation ;
    fpo:fromSource   ex:source_A ;
    fpo:hasPlaceInSource "제2편 총론, 원문편 476쪽" .

ex:ref_003          a fpo:PersonReference ;
    fpo:hasNameInSourceAs "畢齋" ;
    fpo:referencesPerson ex:person_kimjongjik .
```

ex:source\_A의 자리에 적히는 것은 『동유학안』 하나이고, fpo:hasPlaceInSource는 문헌 안의 자리를 적을 뿐 그 자리에서 말하는 주체를 적지 않는다. 그것으로 기록은 끝난다. 이 진술이 하겸진의 것이 아니라 『유종록』의 것이고 하겸진이 한 일은 그것을 골라 편의 첫머리에 놓은 것이라는 사실은 데이터의 어디에도 남지 않는다. 퇴계의 두 진술이 서로 어긋난다는 것, 그 어긋남을 드러낸 뒤 한쪽을 한때의 말로 처리한 것이 하겸진의 판단이라는 것도 마찬가지이다.

이 손실의 성격은 선행 연구의 독해와 대비할 때 분명해진다. 해제는 이 대목을 『동유학안』이 주자학을 정통으로 인식하는 유학관 위에 서 있음을 보여 주는 자료로 읽었다(최일범 2008, 17-18쪽). 이 독해에는 아무런 문제가 없다. 산문의 서술에서 인용된 계보와 편찬자의 안어를 하나의 관점으로 묶어 읽는 것은 정당한 절차이며, 독자는 필요할 때 원문으로 돌아가 총을 다시 나눌 수 있기 때문이다. 데이터에서는 그 되돌아감이 자동으로 보장되지 않는다. 총을 나누어 적지 않으면 나눌 근거 자체가 남지 않으며, 이용자는 하겸진이 조선 유학의 도통을 포은에서 시작한다고 주장하였다는 트리플을 받아 들 뿐 그 주장이 어디서 왔고 하겸진이 무엇을 한 것인지를 되물을 수 없다. 계보 서술이 전송된 사실의 기록이 아니라 특정 시점의 목적에 따라 능동적으로 구성된 표상이라는 관점에서 보면(Wilson 1994, 4-6쪽), 여기서 사라지는 것은 부수적 정보가 아니라 이 문헌이 수행한 작업 그 자체이다.

같은 편의 다음 쪽에 이어지는 김종직 항목은 반대의 극단을 보여 준다. 자와 호와 본관과 부친에서 시작하여 수학과 급제와 관력과 문인 목록으로 이어지는 이 서술에는 인용의 표지가 하나도 없다. 어세겸이 선생의 시를 보고 탄복하였다는 대목, 문인 김일손이 국사를 편수하며 「조의제문」을 실어 찬한 대목, 유자광이 주해를 올려 대역으로 논죄한 대목처럼 타인의 발화가 인용부와 함께 등장하는 자리는 있으나, 서술 전체를 무엇에서 옮겼는지는 밝혀지지 않는다(하겸진 2008, 원문편 477쪽). 앞에서 팩토이드 하나로 적어 본 사승 서술도 이 무표지의 흐름 안에 있다. 이 무표지는 정보의 부족이 아니라 기록해야 할 사태이다. 인코더는 이 구절이 하겸진의 자작인지 행장이나 묘갈 계통의 원저작에서 옮긴 것인지 판별할 수 없으며, 판별할 수 없다는 사실을 판별할 수 없는 채로 적을 자리가 단일 귀속 구조에는 없다. :Source에 『동유학안』을 적는 순간 그 진술은 미결이 아니라 하겸진의 것으로 확정되고, 표지가 명시된 총론의 인용과 표지가 없는 인물 항목의 서술이 데이터에서는 똑같은 모양이 된다. 어느 유형이 얼마나 자주 나타나는가는 II.4에 제시한 인용 사례 조사에서 계량할 사항이며, 여기서는 두 유형이 한 편 안에 나란히 존재한다는 사실의 확인으로 충분하다.

단일 귀속 구조로 이 문헌을 처리했을 때 잃는 것은 세 가지로 정리된다. 첫째, 편찬자의 선택과 배치가 진술의 내용에 흡수된다. 무엇을 인용하고 무엇을 인용하지 않았는가, 어느 진술을 편의 첫머리에 놓고 어느 진술을 안어로 밀어 두었는가는 학안이 학술사를 서술하는 방식 자체인데, 이것이 문헌 단위의 출처 표기 아래에서는 흔적을 남기지 않는다. 둘째, 원저자의 발화가 편찬자의 발화로 승격된다. 『유종록』의 도통 진술이 하겸진의 주장으로 기록되면 이는 표상의 부정확에 그치지 않고 학술사적으로 틀린 서술이 된다. 하겸진의 기여는 그 계보를 발명한 데 있지 않고 그것을 채택하여 편의 원리로 삼은 데 있기 때문이다. 셋째, 인코더가 기록할 수 있는 판단의 종류가 인물 동일시 하나로 한정된다. 家庭을 김숙자로 비정하는 판단은 :PersonReference에 담기지만, 이 구절이 인용이라거나 인용일 가능성이 있다는 판단, 저 안어가 평가적 진술이라는 판단은 담을 자리가 없다. 세 손실이 겹치는

결과, 출처 환원은 한 걸음에서 멈춘다. 이 트리플이 어디서 왔느냐는 물음에는 『동유학안』의 어느 자리라는 답이 돌아오지만, 그 자리가 다시 무엇을 옮긴 것이냐는 물음에는 답이 돌아오지 않는다.

그렇다면 모델을 확장하는 길은 정당한가. 팩토이드를 이차적 연구 데이터가 아니라 일차자료 자체의 구조화에 적용하려는 시도에서 이미 같은 형식의 문제가 제기되었고 같은 형식의 해법이 제출되었다. 지식그래프를 활용한 집 단전기 연구가 대체로 연구자가 전통적 방법으로 생산한 이차 데이터에 의존하고 있으며 자료 해석의 과정 자체에 기여하려면 일차자료의 내용을 직접 기술해야 한다는 문제 제기에서 출발하여, 『갈리아 전기』처럼 시간 정보가 모호한 자료를 대상으로 삼았을 때 기존 팩토이드가 연월일이나 명확한 시종점을 가진 절대적 시간에 의존한다는 점이 한계로 확인되었다(小川潤·大向一輝 2021, 1-2쪽). 확장의 방식이 주목할 만하다. 자료가 제공하지 않는 절대 시간을 추정하여 채워 넣는 대신, 자료가 실제로 제공하는 것, 곧 사건의 계기와 순서를 시간 정보의 지위로 끌어 올려 사건 사이를 선행과 후속의 속성으로 잇고, 그렇게 정의된 사건 구간 안의 인물을 `EntityInContext`와 그 하위의 `PersonInContext`라는 별도의 개체로 정의한 뒤 이를 맥락 없는 보편적 인물 개체와 별도의 속성으로 연결하였다(小川潤·大向一輝 2021, 3-4쪽).

후속 논문은 이 확장을 정식화하면서 그 성격을 스스로 규정한다. `fpo:Reference`가 자료에 직접 기록된 언급에서 얻어지는 반면 `EntityInContext`는 데이터 작성자가 인물과 장소와 공동체를 어떤 시간적 맥락에 배치할 것인가를 해석한 결과이며, 기존 팩토이드가 자료에 기록된 사실의 집적을 중심으로 하였다면 이 개념은 해석 자체를 리소스로 기술하는 것이라는 규정이다(小川潤 외 2022, 264쪽). 구현의 방법 또한 명확하다. 보편적 인물만을 참조하던 `Reference`가 맥락 속의 인물도 참조할 수 있도록 모델을 확장함으로써, 팩토이드 자체에 시간이 부여되지 않아도 참여자의 맥락을 통해 간접적으로 시간적 맥락이 부여된다(小川潤 외 2022, 264-265쪽). 이 확장은 개념의 제안에 그치지 않고 실제 데이터로 검증되었다. 『갈리아 전기』 3권까지의 구축에서 638건의 자원과 2,738건의 트리플이 산출되었고(小川潤·大向一輝 2021, 7쪽), 절대 시간을 전혀 포함하지 않은 채 두 사건 사이의 구간을 지정하는 것만으로 그 기간에 발생한 접촉 관계를 SPARQL로 검색할 수 있음이 확인되었다(小川潤·大向一輝 2021, 8쪽). 인명 비정이라는 인코더의 판단 자체를 데이터로 다루는 최근의 작업(小風綾乃 외 2026)까지 놓고 보면, 자료가 모델에 없던 조건을 요구할 때 모델을 폐기하지 않고 확장하는 방식이 동아시아 디지털 인문학의 장에서 단발이 아니라 축적으로 진행되고 있음을 알 수 있다.

여기서 취할 것과 취하지 않을 것을 구분해야 한다. 이 선례들이 채운 차원은 시간이지 귀속이 아니다. 본 연구가 필요로 하는 것은 시간적으로 모호한 진술의 배치가 아니라 발화 주체가 겹쳐진 진술의 분해이므로, 확장의 내용은 그대로 옮겨 오지 않는다. 옮겨 오는 것은 확장의 형식이다. 자료가 실제로 제공하는 것을 시간 정보의 지위로 끌어올렸듯이 본 연구는 자료가 실제로 제공하는 인용의 표지를 출처 정보의 지위로 끌어올려야 하며, 맥락 속의 인물을 보편적 인물과 별개의 개체로 정의하고 둘을 별도의 속성으로 이어듯이 본 연구는 문헌 안의 특정 발화 층위를 문헌 개체와 별개로 정의하고 둘을 별도의 속성으로 이어야 한다. 다만 유비가 그대로 성립하지 않는 지점이 하나 있고, 그 지점이 본 연구의 고유한 과제이다. 시간적 맥락은 전적으로 데이터 작성자의 해석이므로 해석의 층위를 하나 새로 마련하는 것으로 충분하지만, 인용의 층위는 표지가 있는 경우 자료 자신이 명시한 것이고 표지가 없는 경우 인코더가 추정하는 것이어서, 새로 마련해야 할 층위가 자료의 진술과 인코더의 판단으로 다시 나뉜다. 총론의 『유종록』 인용과 김종직 항목의 무표지 서술이 데이터에서 같은 모양이 되어서는 안 된다는 요구가 바로 이 나뉨을 가리킨다.

이로써 이 소절의 결론이 정리된다. 확장의 전략에는 선례가 있고 확장의 차원에는 선례가 없다. 본 연구의 위치는 팩토이드를 대체하려는 고립된 시도가 아니라, 대상 자료의 조건에 맞추어 이 골격을 확장해 온 계보의 다음 단계이다. 여기서 확인된 공백은 앞에서 세운 요구사항 가운데 발화 주체와 편찬 층위의 구분, 그리고 편찬자의 판단임을 명시하는 표현의 둘에 직접 대응하며, 표지 없는 서술을 미결인 채로 적을 자리가 없다는 문제는 불확실성과 추정의 표시라는 넷째 요구와도 맞물린다. 요구된 자리를 어떤 어휘로 채울 것인가가 다음 소절의 과제이다.

#### 4) 후보 어휘군: Bio CRM, CRMinf, PROV-O, Web Annotation Model

앞 소절은 단일 귀속 구조로 학안을 처리했을 때 잃는 것을 세 가지로 정리하였다. 편찬자의 선택과 배치가 진술의 내용에 흡수되고, 원저자의 발화가 편찬자의 발화로 승격되며, 인코더가 기록할 수 있는 판단의 종류가 인물 동일 시 하나로 한정된다는 것이었다. 이 소절의 과제는 그렇게 비워진 자리를 채울 어휘를 찾는 일이다. 검토의 순서는 어휘의 이름이 아니라 자리의 번호를 따른다. 어떤 표준이 무엇을 할 수 있는가를 차례로 소개하면 다섯 계열의 병렬 소개가 되어 판단의 근거가 흩어지므로, 각 계열이 어느 요구에 어떻게 응답하는가만을 적고 그 응답을 김종직 대목에 실제로 적용해 본다. 여기서 검토하는 어휘들은 팩토이드를 대체할 후보가 아니다. 팩토이드가 진술을 기록의 단위로 삼는다는 결정을 이미 내려 두었으므로 남은 문제는 그 단위에 무엇을 덧붙일 것인가이며, 이하의 계열들은 모두 그 덧붙임의 후보이다.

먼저 확인해 둘 것은 후보군 가운데 하나가 다른 성격의 것이라는 점이다. 사건과 역할의 어휘, 곧 CIDOC CRM 과 그 전기 영역 확장인 Bio CRM은 앞 소절이 남긴 공백에 응답하지 않는다. 이 계열은 사람의 속성과 사람들 사이의 관계와 사람들이 서로 다른 역할로 참여하는 사건을 명확히 구분하고, 관계와 역할을 시간과 공간으로 한정하려면 별도의 사건 개체를 정의하도록 한다(Tuominen 외 2017, 2-3쪽). 역할을 클래스 계층으로 두면 속성의 수를 늘리지 않고도 질의가 단순해진다는 이점도 확인된다(Tuominen 외 2017, 4쪽). 이 구조를 김종직 사례에 적용하면 사승이라는 관계에 스승 역할과 제자 역할이 부여되고, 필요할 경우 수학이라는 사건이 별도의 개체로 기술되어 그 관계가 언제부터 언제까지의 것인가가 한정된다. 그러나 이것은 사승 진술의 내부 구조에 관한 답이지 그 진술을 누가 말했는가에 관한 답이 아니다. 국내에서 이 계열을 채택한 최근의 설계가 항상적인 기본 관계에는 직접 속성을 쓰고 시공간의 한정이 중요한 사건 서술에만 역할 구조를 쓰는 혼성 전략을 택한 것도, 모든 관계에 역할 모델링을 적용하면 모델 복잡도와 구현 부담이 급증한다는 판단 때문이었다(김바로·이재열 2025, 10-11 쪽). 이 계열의 차용 범위는 II.4에 제시한 표본 구간 검토에서 사건 서술의 분포가 확인된 뒤에 정할 사항이며, 여기서는 이 어휘가 답하는 물음이 귀속의 물음과 다른 층위에 있다는 확인으로 그친다.

앞에서 제기된 문제에 직접 응답하는 첫째 어휘는 믿음과 추론의 어휘를 제공하는 CRMinf이다. CRMinf는 과거에 관한 문서화된 명제와 그것을 뒷받침하는 논증을 기록하기 위한 CIDOC CRM의 확장이며, 명제를 작성하고 논증하는 행위를 그 명제가 지금 참인지와 무관하게 과거에 실제로 일어난 사실로 취급한다(Doerr 외 2026, 6 쪽). 이 모델은 지식이 획득되는 경로를 관찰과 추론 수행과 믿음 채택의 셋으로 나누는데, 이 가운데 믿음 채택은 말이나 글이나 그림으로 제시된 정보를 자신의 지식으로 받아들이는 과정이며 학술 출판물에서 특정 출처를 인용하고 참조하는 행위에 대응한다(Doerr 외 2026, 6-7쪽, 22-23쪽). 이 구조를 앞 소절의 총론 대목에 적용해 보면 무엇이 달라지는가가 곧바로 드러난다. 『유종록』이 도통의 계보를 진술한 것과 하겸진이 그 진술을 받아들여 편의 첫머리에 놓은 것은 두 개의 서로 다른 사태이며, 후자는 하겸진을 행위자로 하고 『유종록』을 증거의 출처로 하는 하나의 믿음 채택 행위로 기술된다. 채택의 결과인 믿음은 명제 집합과 믿음값을 함께 갖고, 그 믿음값은 절대적 진리가 아니라 해당 행위자가 자신이 아는 최선의 범위에서 그 명제를 어떻게 받아들였는가를 나타낸다(Doerr 외 2026, 7-8쪽, 18쪽). 이 어휘가 유독 학안에 잘 맞는 이유가 여기에 있다. 하겸진의 편찬은 우리가 지금 판정해야 할 주장이 아니라 1938년에서 1943년 사이에 실제로 수행된 과거의 행위이며, 그 행위를 사실로 기록하되 그 내용을 참으로 단언하지 않는 것이 바로 이 모델의 설계 의도이다. 총론 서두의 인용을 이 어휘로 적으면 다음과 같다.

```

ex:adoption_001  a crminf:I7_Belief_Adoption ;
    crminf:J7_is_based_on_evidence_from ex:work_yujongnok ;
    crminf:J13_adopted_interpretation   ex:belief_001 ;
    crm:P14_carried_out_by              ex:person_hagyeomjin .

ex:belief_001    a crminf:I12_Adopted_Belief ;
    crminf:J14_adopted_interpretation_of ex:work_yujongnok ;
    crminf:J4_that                        ex:propset_001 ;
    crminf:J5_holds_to_be                crminf:true .

ex:propset_001  a crminf:I4_Proposition_Set ;
    crminf:J25_is_encoded_by "吾東之學出於園隱" .

```

이 표기에서 사실로 기록되는 것은 명제가 아니라 채택이라는 행위이다. 곧 데이터가 단언하는 바는 조선의 학문 이 포은에게서 나왔다는 것이 아니라 하겸진이 『유종록』을 근거로 그 명제를 채택하였다는 것이며, 명제 자체는 믿음의 내용으로 한 겹 안쪽에 놓인다. J7 이 근거가 된 문헌을 가리키고 J5 가 하겸진이 그 명제에 부여한 진리 값을 가리키므로, 이용자는 그 계보가 옳은지를 따지기에 앞서 채택이 있었다는 사실과 그 근거를 먼저 확인할 수 있다.

같은 어휘는 인코더의 자리도 함께 마련한다. 인용된 문헌의 저자가 무엇을 표현하려 하였는가에 관한 확신은 의도된 의미에 관한 믿음으로 따로 기술되며, 이 믿음은 그 내용이 참인지에 관한 의견을 포함하지 않는다(Doerr 외 2026, 23쪽). 비평본을 비롯한 학술적 조사에서는 대개 암묵적으로 지나가는 이 단계를 독립적으로 문서화할 수 있다는 단서도 함께 달려 있다. 家庭을 부친 김숙자로 읽는 인코더의 판단은 이 자리에 놓인다.

```

ex:meaning_001  a crminf:I13_Intended_Meaning_Belief ;
    crminf:J17_about                ex:passage_kimjongjik ;
    crminf:J16_assumed_meaning      ex:propset_002 ;    # 家庭은 김숙자를 가리킨다
    crminf:J5_holds_to_be           crminf:probable .

ex:inference_001 a crminf:I5_Inference_Making ;
    crminf:J2_concluded_that        ex:meaning_001 ;
    crminf:J3_applied               ex:logic_prior_sentence ;
    crm:P14_carried_out_by          ex:person_encoder .

```

여기서도 개체로 세워지는 것은 판단의 내용이 아니라 판단이라는 사태이다. 그 판단이 어느 구절을 대상으로 삼았는지가 J17 로, 확신의 정도가 J5 로, 판단이 딛고 선 논리가 J3 으로 각각 분리되어 기록되므로, 다른 인코더가 家庭을 가학 일반으로 읽어 다른 결론에 이르더라도 두 판단은 오류와 정답의 관계가 아니라 논리를 달리하는 두 추론으로 병존한다. 나아가 전제가 되는 믿음에서 다른 명제를 이끌어 내는 추론 수행은 적용된 추론 논리를 함께 기록하도록 하고, 자명하지 않고 그 사례에 고유한 논리만 명시하라는 실용적 원칙을 둔다(Doerr 외 2026, 21쪽). 家庭이 부친을 가리킨다는 비정의 근거, 곧 앞 문장이 김숙자를 부친으로 소개하였다는 문맥적 근거가 그 논리의 자리를 얻는다. 표지가 없는 인물 항목의 서술에 대해서는 믿음값 어휘에 최소한 UNKNOWN 을 포함하라는 권고가 미결 상태를 적는 통로가 되며(Doerr 외 2026, 7쪽), 출처의 참조 계보가 신뢰할 만한지를 따지는 출처 평가가 별도의 논증 활동으로 기술된다(Doerr 외 2026, 23쪽). 이 어휘의 한계도 분명하다. 믿음 채택 행위를 기술하려면 그런 행위가 있었다는 판정을 인코더가 먼저 내려야 하는데, 표지 없는 서술에서는 그 판정 자체가 미결이다. 미결을 미결로 적는 자리는 있으나 미결의 상태에서 행위 개체를 세울지 말지는 이 모델이 결정해 주지 않는다. 아울러 이 어휘는 그 발화가 문헌의 어느 자리에 있는가를 지시하지 않는다.

둘째 어휘는 출처 정보(provenance)의 어휘이다. PROV-O는 어떤 개체가 어떤 활동을 거쳐 어떤 개체에서 파생되었는가를 개체와 활동과 행위자의 세 축으로 기록하며, 국내의 최근 설계가 개별 트리플이 어느 전거에서도 출되었는지를 잇기 위해 파생 속성을 채택한 것도 이 구조에 근거한다(김바로·이재열 2025, 15쪽). 이 어휘로 같은 대목을 적으면 앞의 것과 다른 그림이 나온다. 원저작이 있고, 그것을 발췌하는 활동이 있고, 그 활동에서 생성된 발췌문이 있고, 발췌문에서 도출된 팩토이드가 있고, 팩토이드에서 파생된 RDF 주장이 있다.

```
ex:excerpting_001 a prov:Activity , ex:ExcerptionAct ;
    prov:used          ex:work_yujongnok ;
    prov:wasAssociatedWith ex:person_hagyeomjin .

ex:passage_476      a prov:Entity ;
    prov:wasGeneratedBy ex:excerpting_001 ;
    prov:wasDerivedFrom ex:work_yujongnok .

ex:factoid_002      prov:wasDerivedFrom ex:passage_476 ;
    prov:wasAttributedTo ex:person_encoder .
```

발췌가 활동으로 세워지고 원저작과 발췌문이 각각 그 활동의 입력과 출력으로 놓인다. 총위의 구분이 귀속이 아니라 이력으로 표현되는 것이며, 이 표현은 앞 어휘가 답하지 않는 대목 하나를 채운다. 앞 소절에서 판본 도메인 온톨로지가 발췌된 원저작을 독립된 개체로 기술하는 데까지는 답하되 발췌라는 행위와의 구분에는 답하지 못한다고 하였는데, 발췌를 활동으로 기술하고 원저작과 발췌문을 그 활동의 입력과 출력으로 두는 구조가 바로 그 미답의 절반에 해당한다. 그러므로 인용된 저작의 독립적 개체화라는 요구는 어느 한 어휘가 아니라 두 어휘의 결합으로만 충족된다. 다만 이 계열이 적는 것은 데이터 개체가 무엇에서 만들어졌는가지지 문헌 안에서 누가 말하는가가 아니다. 파생의 연쇄는 총위의 순서를 보존하지만 그 총위에 선 주체가 무엇을 주장하는지를 말하지 않으며, 이 점에서 두 어휘는 대체 관계가 아니라 보완 관계에 있다. 인간 인코더의 판단과 기계가 생성한 후보를 분리하여 기록하는 방침이 이 어휘 위에 서게 된다는 점은 뒤에서 다룬다.

셋째 어휘는 주장을 하나의 단위로 포장하는 어휘이다. Web Annotation Data Model은 주석의 본체와 대상을 분리하고 하나의 주석이 복수의 대상을 가질 수 있게 하며, 대상을 지시하는 선택자(selector)를 여러 형식으로 제공한다(Sanderson 외 2017). 이 구조가 응답하는 것은 텍스트 위치 참조의 다중성이라는 요구이다. 발췌된 인용문이 원저작의 텍스트 위계와 편찬물의 서술 위계에 동시에 속한다는 사태는 하나의 주석이 두 대상을 갖는 형태로 표현될 수 있으며, 어떤 지시자를 선택자로 쓸 것인가는 뒤의 문제로 남는다. 나노퍼블리케이션(nanopublication)은 다른 방향에서 같은 발상에 이른다. 이 모델은 고유하게 식별되는 하나의 진술과 그 진술을 대상으로 삼는 주석들의 집합을 하나의 출판 단위로 규정하고, 진술 자체와 그에 관한 주석을 서로 다른 구획으로 나누어 기술한다(Groth 외 2010, 52-53쪽). 주목할 것은 이 모델이 진술 하나에 대해 서로 다른 세 종류의 주체를 각각 다른 속성으로 적는다는 점이다. 진술을 작성한 주체, 그 진술을 어느 출처에서 가져왔는가, 가져오는 일을 수행한 주체, 그리고 그 단위 전체를 자신의 주장으로 선언한 주체가 각기 다른 관계로 기술되며, 동일한 진술에 관하여 서로 다른 주장의 기원을 구분할 수 있게 된다(Groth 외 2010, 54쪽). 원저자와 편찬자와 인코더라는 세 총위가 이 삼분과 겹쳐 보이는 것은 우연이 아니라 두 논의가 같은 종류의 문제에 부딪혔기 때문이다. 다만 이 어휘가 상정한 반입은 문헌에서 진술을 추출하는 기계적 과정이지 편찬자의 선택과 배치가 아니며, 학안의 발췌가 수행하는 학술사적 기능을 그대로 담기에는 어휘의 성격이 다르다. 인용 행위를 어떤 개체로 세울 것인가라는 본격적인 논의는 다음 절의 몫이다. 여기서 확인할 것은 주장 하나를 독립 개체로 포장한다는 발상이 이미 표준 어휘의 수준에서 실천되고 있다는 사실이다.

이상의 내용을 김종직 대목을 사례로 종합 검토하면 다음과 같다. 각 셀은 제2편 총론의 『유종록』 인용과 하경진의 안어, 그리고 인물 항목의 표지 없는 사승 서술이라는 동일한 자료를 그 어휘로 적어 보았을 때 무엇이 표현되고 무엇이 표현되지 않는가를 근거로 삼았다.

| 모델                   | 발화 주체 구분                            | 저작의 개체화                            | 편찬자 판단의 표시                        | 불확실성 표시                          | 위치 참조의 다중성                         |
|----------------------|-------------------------------------|------------------------------------|-----------------------------------|----------------------------------|------------------------------------|
| 팩토이드와 FPO            | 부분 충족. 자료와 인코더는 나뉘나 자료 안쪽은 나뉘지 않음   | 미충족. 자료는 출처의 위치이며 저작 개체가 아님        | 미충족. 언어와 서술이 같은 지위로 기록됨           | 부분 충족. 인코더의 의구심 표시는 프로젝트 확장에서 제공 | 미충족. 인용 지점이 단일 위치를 지시              |
| 단언적 판본의 RDF 내장 TEI   | 부분 충족. 편찬자의 해석을 데이터로 인정하나 발화 주체는 하나 | 미충족. 저작을 정의하는 어휘를 제공하지 않음          | 부분 충족. 해석의 데이터화는 인정하나 지위 구분장치는 없음 | 부분 충족. TEI의 확실성 및 책임 속성에 의존      | 부분 충족. 임의 구간 연결은 가능하나 총위가 한 지점에 겹침 |
| 판본 도메인 온톨로지          | 미충족. 발화 주체를 다루지 않음                  | 부분 충족. 저작의 개체화는 되나 발체 행위의 자리가 없음   | 미충족                               | 미충족                              | 충족. 위치화된 텍스트 단편이 두 위계의 교차를 개체로 수용  |
| CIDOC CRM과 Bio CRM   | 미충족. 발화가 아니라 사건을 다룸                 | 부분 충족. 정보 객체 수준의 개체화에 그침           | 부분 충족. 속성 부여를 하나의 행위로 세울 수 있음     | 미충족. 확실성 어휘를 따로 두지 않음            | 미충족                                |
| CRMinf               | 부분 충족. 믿음 채택의 연쇄는 적으나 텍스트 위치와 잊지 못함 | 미충족                                | 충족. 믿음과 믿음값으로 사실 진술과 지위를 구분함      | 충족. 추론 행위와 적용된 논리를 함께 기록         | 미충족                                |
| PROV-O               | 부분 충족. 총위를 귀속이 아니라 파생이력으로 표현        | 부분 충족. 발체 행위는 기술하나 저작과 문헌을 구분하지 않음 | 부분 충족. 귀속은 적으나 진술의 유형은 구분하지 않음    | 미충족                              | 미충족                                |
| Web Annotation Model | 부분 충족. 작성과 반입과 주장의 삼분이 총위에 대응       | 미충족                                | 부분 충족. 평가를 주석 개체로 포장할 수 있음        | 미충족                              | 충족. 다중 대상과 선택자가 다중 참조를 수용          |

이 표에서 가장 먼저 확인되는 것은 팩토이드 행에 충족이 하나도 없다는 사실이다. 그러면서도 본 연구가 팩토이드를 골격으로 삼는다면 표와 판단이 어긋나는 것처럼 보인다. 어긋나 보이는 이유는 이 표가 무엇을 측정하는 도구인가에 있다. 다섯 요구사항은 학안이라는 대상이 기존 모델에 추가로 요구하는 것을 진술한 목록이므로, 모델이 이미 제공하고 있어 새삼 요구할 필요가 없는 것은 애초에 열이 되지 않았다. 진술을 사실이 아니라 기록의 단

위로 삼는다는 결정이 그것이다. 이 결정은 다섯 요구 전부의 전제이다. 발화 주체를 되물으려면 되물음의 대상인 진술이 개체로 기술되어 있어야 하고, 편찬자의 판단을 사실 진술과 다른 지위로 적으려면 지위를 부여받을 진술이 이미 노드여야 하며, 인용문의 다중 소속을 지시하려면 지시의 대상이 되는 진술 단위가 확정되어 있어야 한다. 그러므로 이 표는 어느 모델을 택할 것인가를 측정하는 도구가 아니라 골격 위에 무엇을 추가해야 하는가를 측정하는 도구이며, 표를 가로로 읽어 우열을 판정하는 독해는 처음부터 성립하지 않는다.

세로로 읽으면 다른 것이 보인다. 인용된 저작의 독립적 개체화는 어느 한 행에서도 충족을 얻지 못하지만 두 행에 정확히 절반씩 나뉘어 있다. 판본 도메인 온톨로지가 저작과 문헌을 구분하여 원저작을 독립된 개체로 기술하는 절반을 맡고, 생성 이력의 어휘가 발체를 활동으로 기술하여 나머지 절반을 맡는다. 텍스트 위치 참조의 다중성에서도 충족이 두 행에 나타나는데, 하나는 판본을 기술하는 쪽에 있고 다른 하나는 주장을 포장하는 쪽에 있다. 판본과 데이터를 각각 구축하고 참조로 잇는다는 방침이 이 두 충족을 각자의 자리에 두고 연결하는 구성과 맞물린다는 점은 뒤에서 논증할 사항이다. 표가 알려 주는 것은 단일한 최적 모델이 아니라 조합의 형태이다.

현 단계의 잠정적 우선순위는 이렇게 정리된다. 진술의 기록 단위로는 팩토이드의 골격을 쓰고, 발체된 원저작의 개체화에는 저작과 문헌과 단편을 구분하는 판본 쪽 어휘를 쓰며, 편찬자의 평가와 인코더의 비정에는 믿음과 추론의 어휘가 제공하는 지위 구분을 쓰고, 원저작에서 발체를 거쳐 RDF 주장에 이르는 경로에는 출처 정보의 어휘를 쓰며, 주장의 포장과 텍스트 지시에는 웹 주석 계열의 단위를 쓴다. 다만 다음과 같은 이유로 아직 확정 채택은 유보한다. 어휘를 다섯 계열에서 골라 결합하는 설계는 그 장황함을 완화하기는커녕 가중할 수 있다. 자명하거나 논쟁의 여지가 없는 진술에는 메타데이터를 부착하지 않는다는 원칙이 그 대응의 한 형태로 제시된 바 있으나(김바로·이재열 2025, 17-18쪽), 무엇을 자명한 것으로 볼 것인가는 자료의 성격에 달려 있다. 어느 층위에서 어느 어휘를 어디까지 쓸 것인가의 확정은 차후 더 심화된 자료 조사 이후로 미루려 한다.

### 3. 인용과 중층 귀속의 표상

앞 절이 정한 어휘의 조합이 하나의 구조로 맞물리는가 아니면 서로 다른 방향으로 흩어지는가는 하나의 결정에 달려 있다. 발체와 인용이라는 행위를 진술에 붙는 속성으로 처리할 것인가, 아니면 그 자체를 독립된 개체로 기술할 것인가라는 결정이다. 이 절은 그 결정에 이르는 경로를 네 단계로 서술한다. 먼저 학안에서 인용이 어떤 지위를 갖는가를 확인하여 인용이 표상의 대상이 되어야 하는 이유를 밝히고, 다음으로 인용을 유형화하는 기존의 어휘 두 계열을 대비하며, 이어서 그 어휘들로도 채워지지 않는 자리를 확인하고, 마지막으로 그 자리를 근거로 인코딩 위치의 결정을 내린 뒤 그 결정이 드러내는 귀속 연쇄의 구조를 정리한다.

#### 1) 표상 대상으로서의 발체

학안이라는 문헌 형식의 정의 자체가 인용을 중심에 둔다. 학안은 학자의 전기와 그 학자의 저술에서 선택한 발체문을 결합한 문헌 형식이며, 전기와 저술 발체의 결합 여부가 이 장르를 인접 장르와 구분하는 기준이다(Chu 2007, 244쪽, 247쪽). 주희의 『이락연원록』이 인물별 사회적 전기를 제공하면서도 학안으로 분류되지 않는 것은 인물의 전기 뒤에 그 저술을 독립된 선집으로 배치하는 요소를 결여하기 때문이다(Chu 2007, 247쪽). 조선의 계보 문헌사에서도 같은 기준이 작동한다. 『동유사우록』과 『패동연원록』은 사승과 행적의 기록을 갖추었으나 각 학파의 학문적 특징과 개인의 학설을 직접 제시하지 못하였고, 하경진이 연원록이나 사우록이 아니라 학안 형식을 선택함으로써 비로소 학파의 연원과 사상을 함께 제시할 수 있게 되었다(최일범 2008, 11-14쪽). 조선의 계보 문헌이 전기적 서술에서 유학사적 서술로 이행해 온 흐름에서 그 이행을 완결한 수단이 곧 저술의 발체였다는 뜻이다. 인용을 인물 정보에 부속된 자료로 처리하면 이 문헌이 무엇을 성취하였는가가 데이터에서 사라진다.

발체가 편찬 행위라는 점은 계량의 수준에서도 확인된다. 현행 『송원학안』의 소전 2,428개 가운데 실제 사상 자료가 수록된 학자는 209명에 불과하여, 인물의 수록과 사상 자료의 발체가 서로 다른 두 개의 편찬 행위임이 드러난다(連凡 2017, 114-115쪽). 『동유학안』에서도 152인 가운데 56인에게만 문헌자료가 수록되어 발체의 유무 자체가 차등화된 선택으로 나타난다(김기주 2019, 168-169쪽). 어떤 인물을 수록하는 일과 그 인물의 글을 옮겨 실는 일이 별개의 결정이라면, 데이터에서도 그 둘은 별개의 기록이어야 한다.

발체가 왜 판단의 수행인가는 계보 문헌 일반에 대한 독법에서 온다. 신유학의 계보 저술은 유교 전통의 수동적 기록이 아니라 특정 시점의 목적에 따라 능동적으로 구성된 역사적 표상이며, 도통의 계보를 구성하는 일은 정통을 주장하는 행위와 분리되지 않는다(Wilson 1994, 4-6쪽). 이 관점이 인용의 문제에 닿는 지점은 구성의 수단이 무엇인가에 있다. 계보 저술의 편찬자는 자신의 주장을 직접 진술하기보다 앞선 이들의 말을 골라 배열함으로써 주장한다. 무엇을 옮기고 무엇을 옮기지 않으며 옮긴 것을 어디에 놓는가가 곧 논증의 형식이므로, 선택과 생략과 배열은 인용의 부수적 조건이 아니라 인용이 수행하는 일 그 자체이다. 명시적 판정을 자제한 편찬자에게서도 이 수행은 멈추지 않는다. 유원경의 『제유학안』이 보여주듯 편찬자는 평어를 붙이지 않고도 누구를 수록하고 어떤 문장을 발체하며 어떤 순서로 배열하는가를 통해 판단을 수행하기 때문이다(Chu 2007, 250-253쪽). 인용을 데이터의 일차적 대상으로 삼아야 한다는 요구는 여기서 나온다. 그렇다면 그 인용을 무엇으로 유형화할 것인가.

## 2) 인용 분류 어휘

인용을 기계가독의 어휘로 유형화하려는 시도는 서구 시맨틱 퍼블리싱에서 먼저 체계화되었다. CiTO는 인용의 성격을 사실적 측면과 수사적 측면에서 함께 기술하기 위한 온톨로지로서, 핵심 속성 `cito:cites` 와 그 역속성 `cito:isCitedBy` 가 각각 41개의 하위 속성을 거느린다(Shotton 2010; CiTO 2.8.2). 하위 속성들은 인용 기능 분류의 관행에 따라 사실적인 것과 수사적인 것으로 나뉘고, 수사적인 것은 다시 긍정과 중립과 부정의 함의로 묶인다. 이 어휘는 SPAR 묶음의 한 조각으로 설계되어, 서지 개체의 기술은 FaBiO가, 참고문헌의 구조는 BiRO가, 문서의 구조는 DoCO가, 수사적 요소는 DEO가 맡는 분업 위에 놓인다(Peroni and Shotton 2018). 인용의 기능만을 전담하는 어휘라는 뜻이다. 2.0판에서 정의역과 치역의 제약이 제거되어 서지 인용 이외의 맥락에서도 쓸 수 있게 된 점은 학안 적용의 가능성을 열어 준다.

이 목록을 『동유학안』의 발체에 적용해 보면 세 부류로 나뉜다. 인용문의 삽입 자체를 기술하는 속성, 인용된 진술을 자신의 서술 안에 들여오는 관계를 기술하는 속성, 권위로서 끌어오거나 공적을 인정하는 관계를 기술하는 속성, 동의와 이견과 비판과 한정을 기술하는 속성은 학안의 발체와 평어에 그대로 적용할 수 있다. 반대로 근대 학술 출판의 제도를 전제한 속성들, 곧 철회와 사후 심사와 표절을 기술하는 어휘, 게재지와 연구비 지원 기관의 공유를 기술하는 어휘, 인용 시차와 자기인용의 유형을 기술하는 클래스들은 학안에 대응물이 없다. 그 사이에 조정이 필요한 부류가 있다. 증거로서의 인용과 정보원으로서의 인용을 구분하는 기준이 그 예인데, 학안의 발체는 특정 명제를 뒷받침하기 위한 증거가 아니라 인물의 학문을 대표하는 표본이므로 두 속성 가운데 어느 쪽으로도 정확히 옮겨지지 않는다. 속성 목록 전체를 적용 가능과 조정 필요와 적용 불가의 셋으로 판정한 결과는 부표로 제시한다.

여기서 남는 것은 개별 속성의 적부가 아니라 이 어휘가 애초에 무엇을 재지 않는가이다. CiTO의 인용 관계는 인용하는 문헌과 인용되는 문헌 사이에 성립하며, 그 관계가 성립한다는 사실은 전제된다. 인용이 표시되었는지, 표시가 얼마나 정밀한지, 옮겨진 문장이 원문과 얼마나 다른지는 이 어휘의 관심사가 아니다. 근대 학술 논문에서는 참고문헌 목록이 그 정보를 이미 보장하므로 그럴 필요가 없었기 때문이다. 학안에는 그 보장이 없다.

동아시아의 주석 전통은 바로 그 보장이 없는 조건에서 인용을 분류해 왔다. 徐仁甫의 『고서인어연구』는 고서의 인용을 인용의 의도와 직접 인용과 변형 인용의 세 범주로 나누어, 무엇을 위해 인용하는가와 어떻게 옮기는가를 함께 다루었다(徐仁甫 2014). 정약용의 『논어고금주』에서 확인되는 일곱 가지 해석 방식은 인용된 주석을 다루는 태도를 더 세분한 사례이다. 조선시대 논어 주석을 대상으로 이 두 자원을 통합하고 단순화하여 데이터화가 가능한 형태로 정리한 작업은 그 결과를 인용정보와 인용형태와 인용목적의 세 축으로 정리하였다(김서윤 2024a). 각각 출처의 층위와 표시의 정도, 원문 대비 변형의 정도, 인용의 수사적 목적을 대상으로 삼는다. 특히 인용정보 축은 인명이나 서명과 인용 문장이 함께 제시된 경우, 인명이나 서명은 없으나 문장이 있어 원전 추적이 가능한 경우, 인명이나 서명만 있고 문장이 없거나 몇 글자로 축약된 경우를 구분하여, 출처 표시의 정밀도 자체를 기록의 대상으로 삼는다(김서윤 2024a). 근대적 인용 관행이 아니라 주석 전통 자체에서 도출된 어휘라는 점에서 이 세 축은 CiTO 계열과 성격이 다르며, CiTO만으로는 확보되지 않는 자원이다.

이 세 축 가운데 인용목적이 근대적 관념의 소급 적용이 아니냐는 의심은 계량적으로 반박된다. 선진에서 후대에 이르는 여섯 시기의 문헌 259편에서 646건의 인용을 표본으로 삼은 조사는 중국 고전 문헌의 인용 기능이 앞선 이의 공적을 인정하는 것에서 권위와 정통성을 주장하는 것으로 옮겨 갔음을 보였다(Shu 외 2026). 인용이 권위 부여의 수단으로 기능했다는 이 관찰은 동아시아 문장 전통 안에서 인용목적이라는 축이 귀납될 수 있음을 뒷받침하는 동시에, 학안의 발체가 왜 평가의 기능을 수행하는가를 장르 바깥에서 설명해 준다.

두 어휘 체계를 하나의 표로 대비하면 각자가 무엇을 맡고 무엇을 비워 두었는지가 드러난다.

| 축      | 대상 영역          | CiTO 계열의 대응             | 동아시아 계열의 대응        |
|--------|----------------|-------------------------|--------------------|
| 인용정보   | 출처의 층위와 표시의 정도 | 없음. 인용 관계의 성립이 전제 됨     | 표시의 정밀도를 세 등급으로 구분 |
| 인용형태   | 원문 대비 변형의 정도   | 인용문과 발체문의 이분에 그침        | 직접 인용과 변형 인용의 구분   |
| 인용목적   | 인용의 수사적 목적     | 41개 하위 속성, 긍정·중립·부정의 함의 | 축약된 목적 유형          |
| 배치와 분량 | 인용된 자리와 인용된 양  | 없음                      | 없음                 |

표의 마지막 행이 비어 있다는 사실이 다음 소절의 출발점이다.

### 3) 학안의 추가 요구

경전 주석에서 인용의 자리는 고정되어 있다. 하나의 경문이 있고 그에 대한 여러 해석이 있으며, 주석가는 그 해석들을 끌어와 지지하거나 반박한다. 인용목적 축의 긍정과 부정이라는 구분은 이 고정된 대응 관계 위에서 성립한다. 인용된 진술은 언제나 특정 경문에 대한 견해이므로, 인용자가 그 견해에 동의하는가 아닌가를 묻는 것이 자연스럽다.

학안의 발체에는 그 대응 관계가 없다. 하경진이 어느 인물의 문집에서 한 편의 글을 옮겨 인물 항목 뒤에 놓을 때, 그 발체는 앞선 어떤 진술에 대한 견해의 제시가 아니라 그 인물의 학문을 대표하는 표본의 선택이다. 여기서 편찬자가 수행하는 판단은 옳다거나 그르다는 판단이 아니라 이것이 그 사람의 학문이라는 판단이다. 인용목적 축을 그대로 옮겨 오면 대다수의 발체가 긍정으로 몰리거나 어느 값도 부여할 수 없게 되어, 축이 구분의 기능을 잃는다.

그 대신 학안에서는 다른 요소가 평가의 기능을 담당한다. 발체의 유무와 분량과 배치가 그것이다. 152인 가운데 56인에게만 문헌자료가 수록되었다면, 수록되지 않은 96인에 대한 판단은 아무 문장으로도 진술되지 않은 채 발체의 부재라는 형식으로만 존재한다(김기주 2019, 168-169쪽). 이러한 판단의 수행 방식은 학안에 국한되지 않는다. 『동유사우록』이 선생이라는 칭호의 부여 여부로 학문적 성취를 구분하고 『패동연원록』이 성의 생략 여부로 존송의 등급을 표현한 것처럼, 조선의 계보 문헌에서 편찬자의 평가는 명시적 평어 없이 표기 관행 자체로 수행되어 왔다(김기주 2022, 21쪽). 발체량과 배치 순서는 그 관행의 인용 판본이다.

이 항목이 선행 분류에 없다는 것은 그 분류의 결함이 아니다. 경문과 주석이라는 고정된 대응 관계 안에서는 인용의 위치가 이미 결정되어 있으므로 배치가 변수가 되지 않으며, 인용의 분량도 해석의 필요에 따라 정해질 뿐 평가의 신호로 읽히지 않는다. 배치와 분량이 의미를 얻는 것은 인물 항목이라는 서술 문맥 안에 인용이 놓일 때, 곧 하나의 항목 안에서 여러 발체가 순서를 갖고 서로 다른 분량으로 배열될 때이다. 다른 대상이 다른 현상을 산출한 것이지 앞선 어휘가 무언가를 빠뜨린 것이 아니다.

여기에 인용형태 축에서도 한 항목이 더해진다. 하나의 원문을 나누어 여러 곳에 배치하거나 서로 다른 출전에서 온 문장을 하나로 합치는 경우인데, 이때 발췌 단위와 원저작의 단위가 일대일로 대응하지 않는다. 직접 인용인가 변형 인용인가라는 이분법은 하나의 인용문이 하나의 원문 구간에 대응한다는 전제 위에 있으므로, 분할과 병합은 이 축의 값이 아니라 축 자체의 확장을 요구한다.

## 4) 인코딩 위치 결정과 귀속 연쇄의 분해

이제 앞 절이 남긴 결정으로 돌아온다. 인용의 분류를 어디에 적을 것인가에는 두 가지 방식이 있다. 하나는 인용문을 나타내는 요소에 속성을 부여하여 분류값을 적는 방식이고, 다른 하나는 인용 행위 자체를 독립된 개체로 구체화하고 분류를 그 개체에 대한 주장으로 두는 방식이다. 같은 인용을 두 방식으로 적으면 차이가 곧바로 보인다.

```
<!-- 전자: 인용문 요소에 분류값을 속성으로 부여 -->
<quote xml:id="q-yujongnok-01" source="#yujongnok"
      ana="#citeInfo-full #citeForm-verbatim #citePurpose-authority">
  吾東之學出於圃隱。.....爲道學嫡傳。</quote>
```

```
# 후자: 인용 행위를 개체로 설정하고 분류를 그 개체에 관한 주장으로 기록
ex:citing_001 a cito:Citation ;
  cito:hasCitingEntity ex:prologue_ch2 ;
  cito:hasCitedEntity ex:work_yujongnok ;
  cito:hasCitationCharacterization cito:citesAsAuthority ;
  ex:hasPlacement ex:position_prologue_opening ;
  ex:evidence ex:citation_476 .

ex:judgement_001 a crminf:I2_Belief ;
  rdf:reifies <<( ex:citing_001 cito:hasCitationCharacterization
                  cito:citesAsAuthority )>> ;
  prov:wasAttributedTo ex:person_encoder ;
  prov:generatedAtTime "2026-08-23"^^xsd:date .
```

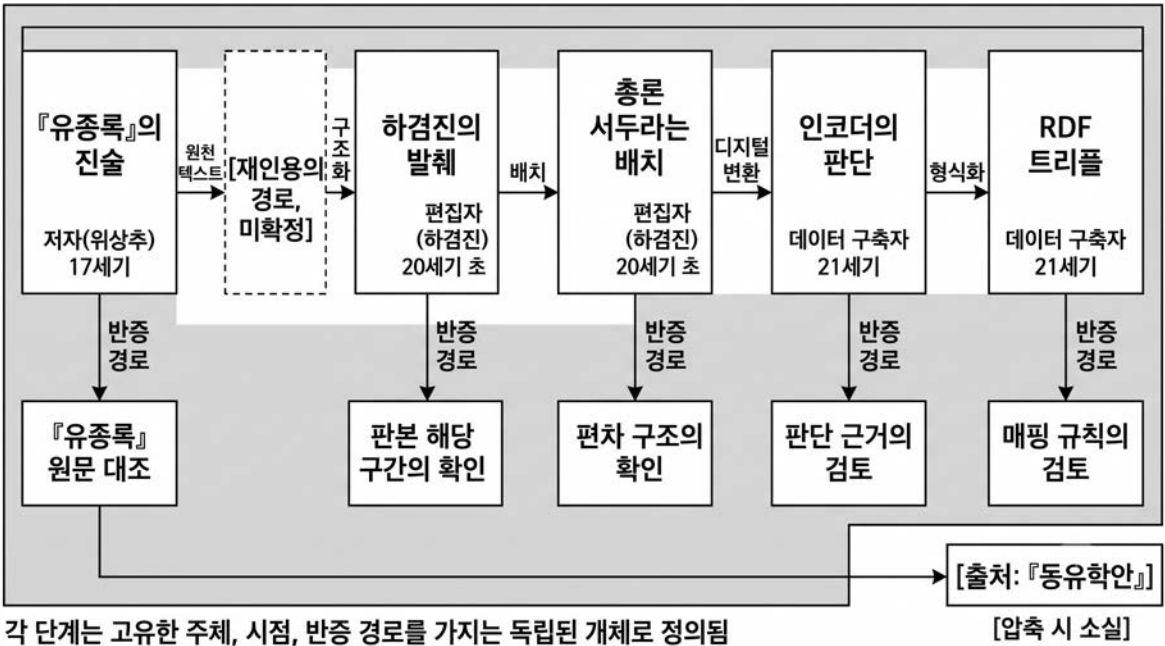
전자에서 분류값은 원문 요소의 속성으로 붙는다. 값이 원문의 표기와 같은 자리에 놓이지만 실제로 그것은 인코더가 내린 판단이며, 판단의 주체와 근거와 시점을 적을 자리가 속성값에는 없다. 후자에서는 인용 행위가 하나의 개체로 세워지므로 그 개체를 주어로 삼는 다른 진술이 성립하고, 배치와 증거의 위치가 별도의 속성으로 붙으며, 분류값의 부여 자체가 인코더에게 귀속된 하나의 주장으로 기록된다. 뒤의 `ex:judgement_001` 블록이 그 귀속에 해당한다.

이 두 방식이 어휘의 우열 문제가 아니라는 점은 CiTO 자신이 보여 준다. 이 온톨로지는 인용을 두 형식으로 적을 수 있게 하는데, 인용하는 문헌과 인용되는 문헌을 하나의 속성으로 직접 잇는 형식과, 인용을 `cito:Citation` 이라는 클래스의 개체로 기술하고 인용하는 개체와 인용되는 개체와 인용의 성격을 각각 별개의 속성으로 붙이는 형식이다. 후자를 마련한 이유는 인용 자체가 다른 진술의 주어나 목적어가 될 수 있어야 하기 때문이며, 인용의 이유를 기존 속성으로 표현할 수 없을 때 주석 모델을 써서 그 인용 개체에 설명을 덧붙이는 용법도 함께 제시된다. 어느 형식을 택할 것인가는 인용 관계 자체에 대해 무엇을 더 말할 필요가 있는가에 달려 있다는 뜻이다.

전자는 구현이 단순하고 하나의 인용문에 복수의 분류값을 부여하기가 자유로우며, 무엇보다 이미 축적된 XML 자산을 그대로 살릴 수 있다. 한국경학자료시스템의 기존 XML을 검토하고 그 위에서 데이터 설계를 제안한 작업이 이 조건 아래 있다(김서윤 2024b). 판본이 이미 존재하고 그 판본의 마크업을 재사용하는 것이 설계의 출발점일 때, 인용 분류를 속성으로 부여하는 것은 합리적 선택이다.

본 연구의 조건은 다르다. 저본을 정하고 대교를 거쳐 한문 원문의 비판정본을 새로 구축하는 것이 작업의 일부이므로 재사용해야 할 기존 마크업이 없으며, 인코더의 판단을 원문의 진술과 구분하여 기록하는 것 자체가 목표이다. 이 조건에서 속성 부여 방식은 곧바로 문제를 일으킨다. 인용문 요소에 인용목적의 값을 적어 넣으면 그 값이 원문의 일부인 것처럼 보이지만, 실제로 그것은 인코더가 내린 판단이다. 판단임을 표시하려면 그 판단의 주체와 근거와 시점을 적을 자리가 필요한데, 속성값에는 그 자리가 없다. 인용목적의 판별에 연구자의 주관이 개입할 수밖에 없다는 유보는 이 문제를 앞서 제기한 이가 스스로 밝힌 바 있다(김서운 2024a). 본 연구는 그 유보를 결함으로 지목하는 것이 아니라, 그렇다면 그 주관적 판단을 어디에 어떻게 기록할 것인가라는 후속 물음을 잇는다. 그 물음에 답하려면 판단이 붙을 대상, 곧 인용 행위가 먼저 독립된 개체로 기술되어 있어야 한다. 본 연구가 후자를 택하는 이유이며, 이 차이는 설계의 우열이 아니라 대상과 조건의 차이에서 온다.

인용 행위를 개체로 기술하면 그때까지 하나로 결합되어 있던 것이 여러 단계로 나뉜다. 앞 절에서 본 제2편 총론의 대목이 그 분해의 양상을 그대로 보여 준다. 『유종록』이 우리 동방의 학문이 포은에게서 나왔다고 진술한 것이 첫째 단계이고, 하겸진이 그 진술을 선택하여 편의 첫머리에 배치한 것이 둘째 단계이며, 그 인용문이 총론의 서두라는 위치를 차지하여 이하의 인물 배열을 규정하게 된 것이 셋째 단계이다(하겸진 2008, 원문편 476쪽). 인코더가 이 대목을 읽어 하겸진이 도통의 기점을 포은으로 삼는 계보를 채택하였다고 판단한 것이 넷째 단계이고, 그 판단이 특정 어휘의 트리플로 기록된 것이 다섯째 단계이다. 다섯 단계는 서로 다른 시점에서 서로 다른 주체가 수행한 서로 다른 행위이며, 각각에 대해 다른 종류의 반증이 성립한다. 첫째 단계는 『유종록』의 원문 대조로, 둘째 단계는 판본의 해당 구간 확인으로, 셋째 단계는 편차의 구조 확인으로, 넷째 단계는 인코더의 판단 근거 검토로, 다섯째 단계는 매핑 규칙의 검토로 각각 반증된다.



[도해 6] 5단계 귀속 연쇄

출처를 문헌의 식별자 하나로 축약하는 방식에서는 이 연쇄가 복원되지 않는다. 출처가 『동유학안』이라고 적히는 순간 다섯 단계는 하나의 좌표로 축약되고, 이용자에게 남는 것은 하겸진이 조선 유학의 도통을 포은에서 시작한다고 주장하였다는 진술뿐이다. 앞에서 확인한 대로 그러한 진술은 표상의 부정확에 그치지 않고 학술사적으로 부정확한 서술이 된다. 인용을 독립된 개체로 구체화해야 하는 최종적 이유가 여기에 있다. 개체화는 표현의 정밀도를 높이는 선택이 아니라 오귀속을 막는 조건이다.

한 가지는 공백으로 남겨 둔다. 하검진이 어떤 진술을 매개 문헌을 거쳐 접했을 가능성, 곧 재인용의 경로를 데이터로 표현하는 문제이다. 다섯 단계 가운데 첫째와 둘째 사이에 몇 개의 단계가 더 있을 수 있다는 뜻인데, 그 단계의 존재는 원출전 확인을 거쳐야 비로소 판별되므로 현 단계에서 구조를 확정할 근거가 없다. 인용의 경로와 매개 문헌을 다루는 별도의 어휘를 검토하는 일은 자료 조사 이후로 미루고, 이번 단계에서는 다섯 단계의 연쇄가 확장 가능한 형태로 설계되어야 한다는 요구만 기록한다.

이 절의 범위도 여기까지이다. 인용 행위를 어떤 클래스로 정의하고 어떤 속성으로 이를 것인가라는 최소 구조의 확정에는 자료 조사로 인용의 유형이 확인된 뒤에야 가능하며, 지금 확정하면 다섯 단계 가운데 무엇이 실제로 나타나는지를 모르는 채로 구조를 고정하게 된다. 확인된 것은 선행 모델이 처리하지 못하는 구조가 무엇인가까지이다. 남은 물음은 이렇게 나누어 놓은 다섯 단계를 그래프에서 어떤 표기로 서로 구분할 것인가이며, 그 답은 진술 수준 메타데이터의 표현 방식에 달려 있다.

## 4. 그래프 데이터 모델과 진술 수준 메타데이터

앞 절은 하나의 물음을 남겼다. 인용 행위를 개체로 세워 분해한 귀속 연쇄의 다섯 단계, 곧 원저작의 진술과 하검진의 발췌와 학안 안에서의 배치와 인코더의 해석과 최종 RDF 주장을 그래프에서 어떤 표기로 서로 구분할 것인가이다. 표기를 정하려면 그 표기가 놓일 데이터 모델이 먼저 정해져 있어야 한다. 이 절은 그래프 데이터 모델을 정하고, 그 위에서 진술 수준 메타데이터의 표기 두 가지를 같은 사례에 적용하여 비교한 뒤, 선택한 표기가 현재의 구현 여건에서 어떤 위험을 안고 있는지를 확인하는 순서로 진행한다.

### 1) 라벨 속성 그래프와 RDF

그래프 데이터 모델의 후보로는 RDF 계열과 속성 그래프(property graph) 계열이 있다. 진술 수준 메타데이터 면에서 유리한 쪽은 후자이다. 속성 그래프는 노드와 엣지에 각각 고유 식별자를 부여하고 그 식별자를 매개로 속성과 값의 쌍을 직접 연결하는 모델이며, 이 때문에 같은 두 노드 사이에 같은 레이블을 가진 여러 엣지를 서로 다른 엣지 식별자로 구별할 수 있고 그래프 구조를 재설계하지 않은 채 노드와 엣지에 속성을 추가할 수 있다 (Angles 외 2017, 68:6-68:7쪽). Angles 외는 그 이점의 출처를 대비로 밝히는데, 엣지 레이블 그래프에서 엣지에 출처 같은 메타정보를 붙이려면 관계를 노드로 다시 모델링하거나 엣지 자체를 구체화해야 한다는 것이다 (Angles 외 2017, 68:6쪽). 진술 수준 메타데이터라는 본 절의 주제가 속성 그래프에서는 모델의 기본 기능이고 RDF에서는 별도의 장치를 요구하는 문제라는 사실이 이 대비에 압축되어 있다. 여기에 두 가지 이점이 더해진다. 하나는 표현의 경제성이다. 동일한 정보량을 담는 데 필요한 구문의 길이가 RDF 계열보다 짧으므로, 같은 규모의 자료를 더 적은 기술량으로 옮길 수 있다. 다른 하나는 질의 언어의 공적 표준화가 진전되었다는 사실이다. 속성 그래프를 대상으로 하는 질의 언어 GQL이 2024년에 국제 표준으로 제정되어(ISO 2024), 질의의 의미론이 구현마다 상이하던 종래의 사정이 표준의 층위에서 해소될 경로가 마련되었다.

이 이점을 근거로 판본을 속성 그래프로 구축한 선례가 있으며, 그 문제의식은 본 연구와 상당 부분 겹친다. 힐데가르트 폰 빙엔 서한집의 판본을 다룬 Kuczera는 편집자의 모든 결정이 논의의 여지가 있는 주관적 진술이고 해석 과정은 텍스트의 첫 글자가 전사되는 순간부터 시작되며 바로 그 지점에서 출처 정보가 개입한다고 규정한 뒤, 판본 전체를 모든 정보가 그 출처와 진술자와 진술 시점과 함께 저장되는 출처 지식그래프로 볼 것을 제안한다 (Kuczera 2020, 115쪽). 관찰자와 관찰 도구를 관찰 대상에 포함시켜야 한다는 원칙 아래 전사자와 편집자와 주석자와 이용자를 편집 체계의 구성 요소로 인정한 것도 같은 취지이다(Kuczera 2020, 102쪽). 인코더의 해석을 데이터의 일부로 기록해야 한다는 본 연구의 전제와 정확히 같은 자리이다.

그럼에도 결론이 상이한 이유는 두 연구가 우선하는 바가 다르기 때문이며, 분기하는 지점을 정확히 밝혀 둘 필요가 있다. Kuczera가 RDF 대신 속성 그래프를 택하며 든 근거는 셋이다. 동일한 정보량에서 RDF 그래프의 규모가 속성 그래프의 열 배에 이를 수 있다는 점, RDF의 노드와 관계가 내부 구조 없이 고유 레이블에 그친다는 점, 동일 유형의 관계 인스턴스를 고유하게 식별하거나 한정할 수 없다는 점이다(Kuczera 2020, 116쪽). 셋째 근거는 2020년이라는 시점의 판정으로 읽어야 한다. 아래에서 볼 RDF 1.2의 구체화 구문이 관계 인스턴스의 식별과 한정을 모델의 기본 기능으로 편입하였으므로, 이 항목은 지금 시점에서 상당 부분 해소되었다. 첫째와 둘째 근거

는 해소되지 않았으나, 여기에는 본 연구와의 설계상 차이가 개입한다. 규모의 부담은 판본의 텍스트 구조와 문자 단위 색인까지 그래프 안으로 들어오기 때문에 발생하는 것인데, 본 연구는 판본을 TEI로 유지하고 지식그래프를 별도로 구축하여 참조로 연계하므로 그래프가 문자 색인을 부담하지 않는다. 스탠드오프(stand-off) 속성 방식의 텍스트 색인이 그래프의 규모를 결정하는 구조와 판본을 그래프 바깥에 두는 구조는 같은 저울에 올릴 수 없다. 그 논증은 다음 절의 몫이고 여기서는 규모 논거의 무게가 이원 구축 방침 아래에서 달라진다는 점만 확인한다.

남는 것은 본 연구가 무엇을 우선하는가이다. 첫째, 어휘의 재사용과 외부 연결이다. 본 연구는 CRMInf의 믿음과 추론, PROV-O의 파생, 주석 계열의 포장 단위를 조합하기로 하였는데, 이들은 모두 RDF 어휘로 정의되어 있어 속성 그래프에서는 표준 의미론을 잃은 채 관례로만 옮겨진다. 둘째, 전역 식별자이다. RDF에서는 모든 개체와 속성이 IRI로 지시되므로 인물과 저작과 학파에 부여한 식별자가 데이터셋 바깥에서도 그대로 통용되며, 다른 기관이 구축한 인물 데이터와의 연결이 별도의 대응표 없이 성립한다. 셋째, 모델 정의와 도구 계층의 개방성이다. 앞에서 확인한 표준화는 질의 언어의 총위에서 이루어진 것이며, 데이터 모델의 정의와 저장 엔진의 구현, 그리고 도구 생태계의 주도권은 여전히 특정 상용 기업에 크게 의존한다(Cagle 2026). 학술 데이터의 산출물이 한 기업의 제품 수명과 정책에 종속되지 않아야 한다는 요구가 여기에 걸린다.

이상을 근거로 본 연구는 그래프 데이터 모델로 RDF 1.2를 잠정 선택한다. RDF 계열 안에서 1.2를 택하는 근거는 진술 수준 메타데이터의 부착 방식에 있다. RDF 1.1의 표준 구체화에 비하여 부착의 구문이 간결해졌고, 커뮤니티 그룹 단계의 RDF-star에 비하면 구문의 경제성은 다소 떨어지나 명제와 그 명제를 지시하는 개체가 분리되어 표현의 정밀도가 높아졌다. 그 구체적인 대비는 다음 항에서 다룬다. 다만 두 가지를 함께 적어 둔다. 하나는 이 선택이 질의 성능에서의 우위 주장이 아니라는 점이다. 시작 노드가 주어진 상태에서의 짧은 경로 탐색은 속성 그래프가 명확히 우세하며, 본 연구가 그 이점을 포기하는 것은 우선순위가 탐색 성능이 아니라 상호운용성과 출처 표현에 있기 때문이다(Cagle 2026). 다른 하나는 이 선택이 속성 그래프의 배제를 뜻하지 않는다는 점이다. 학파별 관계망의 계량 분석이 필요해질 경우 의미론적 백본은 RDF로 유지하고 분석용 속성 그래프를 투영하는 혼합 구성이 가능하며, 이는 실제 현장에서 혼합 구조가 일반화되고 있다는 Cagle의 진단과도 부합한다(Cagle 2026). 그 검토는 이 장의 마지막 절로 넘긴다.

## 2) 진술 수준 메타데이터 표기 방식

모델이 정해졌으므로 앞 절이 남긴 첫째 물음으로 돌아온다. 귀속 연쇄의 다섯 단계를 그래프에서 어떻게 구분할 것인가, 곧 진술 하나에 그 진술에 관한 메타데이터를 어떤 표기로 붙일 것인가의 문제이다. 후보는 둘이다. RDF 1.1부터 있던 표준 구체화(standard reification)<sup>1</sup>와 RDF 1.2가 도입한 트리플 용어(triple term) 및 구체화 트리플(reifying triple)이다. 비교의 기준은 문화유산 메타데이터의 출처 표현 모델을 조사한 최근의 서베이가 세운 축을 본 연구의 조건에 맞추어 좁힌 것이다. Massari 외는 RDF 표준 준수 여부, SPARQL 호환성, 직렬화 가능성, 확장성, 비단언 표현 가능성, 외부 어휘 의존성, 실제 구현 여부의 일곱 축으로 각 접근을 비교하였는데 (Massari 외 2026, i207-i210쪽), 이 가운데 본 절의 판단에 직접 관여하는 넷을 골라 증거 지시의 정밀도, 질의 편의, 도구 지원 범위, 인코더 분리의 용이성으로 삼는다. 비단언 표현 가능성이라는 축은 별도로 설정하지 않고 넷째 기준 안에서 다룬다. 사료의 진술과 인코더의 판단을 분리한다는 것은 곧 인코더가 적은 트리플이 사료의 단언으로 읽히지 않게 만드는 일이기 때문이다.

비교는 하나의 사례 위에서 진행한다. 앞에서 본 제2편 총론의 첫머리, 곧 하겸진이 『유종록』을 인용하여 도통의 계보를 제시하고 그 계보 안에 김종직을 도학의 적전으로 놓은 대목이다(하겸진 2008, 원문편 476쪽). 여기서 기록해야 할 명제는 김종직이 도학의 적전에 놓인다는 것 하나이지만, 그 명제에 관하여 구분해야 할 사항은 셋이다. 그 명제를 진술한 것은 『유종록』이고, 그것을 골라 편의 첫머리에 놓은 것은 하겸진이며, 그 두 사실을 데이터로 옮긴 것은 인코더이다. 아래 표기에서 접두사와 속성 이름은 설명을 위해 약식으로 두었다.

먼저 표준 구체화이다.

```

ex:stmt_001 a rdf:Statement ;
  rdf:subject    ex:kimjongjik ;
  rdf:predicate  ex:placedIn ;
  rdf:object     ex:orthodoxLineage ;
  ex:statedIn    ex:yujongnok ;
  ex:adoptedBy   ex:hagyeomjin .

```

원 트리플이 세 개의 속성으로 분해되고, 그 분해된 자원에 출처와 채택자가 부착된다. 원 트리플 자체는 그래프에 존재하지 않으므로 이 명제가 단언된 것인지 아닌지를 표기가 말해 주지 않는다.

다음은 RDF 1.2이다.

```

ex:r_yujongnok  rdf:reifies <<( ex:kimjongjik ex:placedIn ex:orthodoxLineage
)>> ;
  a              ex:SourceStatement ;
  ex:statedIn    ex:yujongnok .

ex:r_hagyeomjin rdf:reifies <<( ex:kimjongjik ex:placedIn
ex:orthodoxLineage )>> ;
  a              crminf:I2_Belief ;
  ex:adoptedBy   ex:hagyeomjin ;
  ex:derivedVia  ex:excerpt_001 ;
  ex:evidence     ex:citation_476 .

```

같은 명제를 지시하는 구체화자가 둘 세워졌다. 앞의 것은 『유종록』이 그렇게 진술하였다는 사태를, 뒤의 것은 하겸진이 그 진술을 자신의 믿음으로 채택하였다는 사태를 각각 가리키며, 명제 자체는 어느 쪽에 의해서도 참으로 단언되지 않은 채 남는다.

두 표기를 나란히 놓으면 차이가 드러난다. 표준 구체화는 표준 준수와 도구 지원에서 가장 안전하다. RDF 1.1 이래의 모든 도구가 이를 처리하며 별도 어휘 없이 직렬화된다. 그러나 세 가지 부담이 따른다. 첫째는 장황함이다. 트리플 하나를 구체화하면 데이터셋의 규모가 최소 네 배로 늘고, SPARQL에서 구체화 자원을 찾으려면 원 트리플의 모든 구성 요소를 명시적으로 일치시켜야 한다(Massari 외 2026, i198쪽). 이 지적은 국내 인물 데이터 설계에서 표준 구체화가 요구하는 구조가 지나치게 장황하다고 본 판단과 같은 것이다(김바로·이재열 2025, 12-13쪽). 둘째는 이 방식이 널리 사용되거나 이해되거나 구현되지 않았으므로 폐기해야 한다는 제안이 표준의 편집자 자신에게서 나온 바 있다는 사실이다(Massari 외 2026, i198-i199쪽). 셋째가 본 연구에 가장 치명적이다. RDF 구체화는 구문이 아니라 의미 수준에서 작동하므로, `owl:sameAs`로 동일시된 두 자원을 각각 주어로 하는 두 구체화가 의미적으로 동치가 되어 버린다(Carroll 외 2005, 250-251쪽). 학안의 인코딩에서 이 성질은 그대로 사고로 이어진다. 사료 속 표기 家庭을 김숙자라는 인물 개체와 연결하는 순간, 家庭에 관한 진술과 김숙자에 관한 진술이 구분되지 않게 되기 때문이다. 사료의 표기와 인코더가 동일시한 인물을 별개의 노드에 두는 FPO의 분리 장치가 표기 층위에서 무너지는 셈이다.

RDF 1.2의 구조는 표준 구체화가 남긴 자리에 정확히 대응한다. 트리플 용어는 다른 트리플의 목적어 자리에 놓이는 네 번째 종류의 항이고, 술어가 `rdf:reifies` 이고 목적어가 트리플 용어인 트리플을 구체화 트리플이라 하며 그 주어를 구체화자(reifier)라 한다. 구체화자는 명제가 성립한다는 진술이나 믿음을 비롯하여 그 명제와 관련된 여러 가지를 지시할 수 있고, 이후의 진술에서는 트리플 용어가 아니라 구체화자가 사용될 것으로 상정된다(Kellogg 외 2026). 여기서 본 연구에 결정적인 규정이 둘 나온다. 하나는 트리플 용어로 지시된 명제가 그 자체로 참으로 주장되지 않으며 같은 트리플이 그래프 안에서 단언된 트리플(asserted triple)이기도 할 때에만 참으로 주장된다는 것, 그리하여 비단언 트리플 용어를 통해 단언되지 않은 진술에 관한 진술을 만들 수 있다는 것이다(Kellogg 외 2026). 다른 하나는 같은 명제에 대하여 서로 다른 출처의 진술처럼 복수의 구체화자가 존재할 수

있다는 것이다(Kellogg 외 2026). 앞의 예시에서 하나의 명제에 두 개의 구체화자를 세운 것이 그 활용이다. 『유종록』의 진술과 하겸진의 채택은 같은 명제에 관한 서로 다른 사태이며, 그 명제 자체는 어느 쪽에 의해서도 참으로 단언되지 않은 채 남는다. 진술을 확정된 사실로 기록하지 않으면서 출처로 되돌아갈 수 있게 한다는 핵심 질문의 조건이 표기 층위에서 답을 얻는 자리가 여기이다. `rdf:reifies`의 의미가 의도적으로 일반적으로 정의되어 있다는 점도 이점이 된다(Kellogg 외 2026). 구체화자를 CRMinf의 믿음으로 유형화하는 위 예시처럼, 각 단계의 성격을 별도 어휘로 지정하는 통로가 열려 있기 때문이다.

네 기준에 따른 판정은 다음과 같다.

| 기준          | 표준 구체화                                  | RDF 1.2 구체화                       |
|-------------|-----------------------------------------|-----------------------------------|
| 증거 지시의 정밀도  | 트리플 단위. 다만 의미 수준 작동으로 동일시된 자원의 구체화가 뒤섞임 | 트리플 단위. 같은 명제에 복수 구체화자를 세워 단계를 나눔 |
| 질의 편의       | 낮음. 주어와 술어와 목적어를 모두 명시 일치시켜야 함          | 중간. SPARQL 1.2의 구체화 구문에 의존        |
| 도구 지원 범위    | 가장 넓음. RDF 1.1 도구 전부                    | 좁음. 아래 항목에서 별도로 다룸                |
| 인코더 분리의 용이성 | 낮음. 원 트리플의 단언 여부를 표기가 규정하지 않음           | 높음. 비단언 트리플 용어로 미단언 상태가 명시됨       |

표에서 읽어야 할 것은 두 표기가 서로 다른 것을 보장한다는 사실이다. 표준 구체화는 도구 지원의 범위를 보장하고, RDF 1.2의 구체화는 단언과 비단언의 구별을 보장한다. 본 연구의 설계가 걸려 있는 것은 뒤의 보장이다. 원저작의 진술과 편찬자의 채택과 인코더의 판단을 같은 명제에 대한 서로 다른 구체화자로 설정하려면 그 명제 자체가 참으로 단언되지 않은 채 남아 있어야 하는데, 표준 구체화는 그 상태를 표기의 수준에서 규정하지 않는다. 따라서 잠정안은 다음과 같다. 귀속 연쇄의 단계를 나누는 표기로는 RDF 1.2의 구체화를 쓰고, 표준 구체화는 표기의 선택지가 아니라 외부 상호운용을 위한 변환 산출물로만 둔다. 뒤의 항목은 실용적 근거를 갖는다. 주요 프레임워크가 트리플 용어와 표준 구체화 사이의 상호 변환 기능을 제공하고 있으므로, 표준 구체화를 파생 형식으로 유지하는 데 추가 설계가 거의 들지 않기 때문이다.

### 3) RDF 1.2 선택의 위험 요인

이상의 선택은 아직 완결되지 않은 표준에 의존한다. 이 사실을 결함으로 감추지 않고 위험 요인으로 기록해 둔다. Cagle은 개념 설계의 층위에서 RDF 1.2의 구체화 구문이 관계에 메타데이터를 붙이는 일의 사용성 격차를 상당 부분 줄였다고 평가하면서도, 저장소별 구현 편차와 SPARQL 1.2 지원의 불균등이 과제로 남아 있으며 도구 생태계의 성숙도에서는 여전히 속성 그래프 계열이 앞선다는 유보를 함께 제시한다(Cagle 2026). 이 유보가 현재 어느 정도로 실재하는지를 2026년 8월 21일 시점에 확인한 결과는 다음과 같다.

| 저장소                       | 확인된 상태            | 준거 모델                            |
|---------------------------|-------------------|----------------------------------|
| Apache Jena, Fuseki, TDB2 | 6.1.0, 2026-05-11 | RDF 1.2 구문 입출력과 SPARQL 1.2 지원    |
| Eclipse RDF4j             | 6.0.0, 2026-07-02 | RDF 1.2. NativeStore는 트리플 용어 미지원 |
| Oxigraph                  | 0.5.7, 2026-04-19 | RDF 1.2. 해당 기능 활성화 시             |
| Ontotext GraphDB          | 11.4 문서           | RDF-star 커뮤니티 그룹안                |
| AllegroGraph              | 9.0.1             | RDF-star 커뮤니티 그룹안, 기본 비활성        |
| Stardog                   | 최신 문서             | 초기 RDF* 제안, RDF 1.2 미구현을 명시      |
| Amazon Neptune            | 공식 답변             | 미지원                              |

표준 자체의 진행 상황도 균일하지 않다. RDF 1.2 개념 문서와 의미론 문서는 2026년 4월 7일자 후보권고이나 구체 구문은 작업 초안이며, SPARQL 1.2 문서 열두 편은 전부 작업 초안이다. 표기를 정의하는 쪽이 앞서 있고 그 표기를 질의하는 쪽이 뒤에 있는 비대칭이 현재의 상태이다. 저장소 쪽에서는 자바 생태계의 두 프레임워크와 Oxigraph가 RDF 1.2로 이동한 반면 상용 저장소 다수가 커뮤니티 그룹 시절의 RDF-star에 머물러 있고, 같은 제품 안에서도 저장 계층에 따라 지원 여부가 상이하다. 나아가 단언 여부의 처리가 구현마다 다르다는 점도 확인된다. 일부 저장소는 내포 트리플이 참조된 진술을 함께 단언한다는 규정이 표준 구체화와의 호환을 깨뜨린다고 판단하여 참조된 진술을 명시적으로 단언하지 않는 쪽을 택하였는데, 이는 미구현이 아니라 의미론의 선택이다. 단언과 비단언의 구별을 설계의 축으로 삼는 본 연구에서 이 편차는 데이터를 옮길 때 의미가 달라질 수 있다는 뜻이 되므로 가볍게 볼 수 없다.

이 조건에서도 본 연구가 RDF 1.2를 채택하는 것은 이 표준이 권고로 진급하리라는 전망에 근거한 판단이며, 학위논문의 시간 규모가 표준의 남은 일정보다 길다는 사정도 함께 작용한다. 다만 전망에 기대는 만큼 세 가지 대비책을 설계에 넣는다. 첫째, 어휘 설계와 표기 선택을 분리한다. 인용 행위를 독립 개체로 설정하고 편찬자의 채택을 믿음 채택 행위로 기술하는 구조 자체는 RDF 1.1 어휘만으로도 표현되며, RDF 1.2의 구체화는 그 위에 부가하는 표기이다. 표기가 바뀌어도 모델이 무너지지 않게 하는 것이 이 분리의 목적이다. 둘째, 표준 구체화로의 기계적 변환 경로를 유지한다. 셋째, 저장소 지원 현황의 확인 시점을 데이터와 문서에 명기하고, 파일럿 구축 단계에서 실제 저장소를 대상으로 단언 여부의 처리를 시험한다. 어느 시험을 어떤 순서로 수행할 것인지는 이 장의 마지막 절에서 다룬다.

이로써 이 절의 범위가 닫힌다. 그래프 데이터 모델을 정하였고, 그 위에서 진술의 단계를 나누는 표기를 정하였으며, 그 표기가 감수하는 위험과 대비책을 적었다. 그러나 위 예시에서 `ex:citation_476` 이라 적어 둔 자리, 곧 이 구체화자들이 근거로 삼는 판본의 지점을 무엇으로 어떻게 가리킬 것인가는 아직 비어 있다. 단계를 아무리 정밀하게 나누어도 각 단계가 판본의 어느 구절에 닿는지를 지시하지 못하면 출처 환원은 데이터 안에서만 맴돈다. 판본과 지식그래프를 무엇으로 잇는가가 다음 절의 과제이다.

## 5. 디지털 판본과 지식그래프의 연계: 선행 방법론의 검토

앞 절의 예시에서 구체화자마다 붙여 둔 증거 속성의 값, 곧 `ex:citation_476` 은 아직 이름뿐이다. 이 자리를 채우는 일은 표기법의 문제가 아니라 두 개의 산출물을 무엇으로 잇는가의 문제이며, 그 연결부가 견고하지 않으면 발화 주체를 아무리 정밀하게 나누어 적어도 각 주체의 발화가 판본의 어느 구절인지 소급할 수 없어 출처 환원이 『동유학안』이라는 문헌명에서 다시 멈춘다. 이 절은 다섯 단계로 진행한다. 먼저 판본과 지식그래프를 잇는 기

존의 시도를 세 유형으로 나누어 그 지형을 정리하고, 그 가운데 두 유형을 차례로 검토하여 본 연구가 셋째 유형을 택하는 근거를 세운다. 이어서 그 선택이 곧바로 부딪히는 구조적 난점, 곧 하나의 텍스트 단편이 두 개의 위치에 동시에 속하는 사태를 확인한다. 그 위에서 앵커의 문제를 주석의 소재와 지시의 방식이라는 두 축으로 나누고, 비교의 본체를 지시 축 하나로 좁혀 세 계열을 대비한다. 마지막으로 현 단계의 잠정안을 제시한다.

## 1) 세 가지 연계 방식

판본과 지식그래프를 잇는 시도는 세 유형으로 나뉜다. 첫째는 판본 자체를 그래프로 전환하여 텍스트 구조와 주석과 사료 정보를 하나의 그래프 안에서 다루는 접근이다. 둘째는 TEI 판본을 유지하되 그 안에 RDF의 근거를 두어 판본에서 그래프를 파생시키는 접근이다. 셋째는 판본과 그래프를 각각의 표준 생태계 안에서 구축한 뒤 참조로 연결하는 접근이다.

세 유형은 무엇을 하나로 묶고 무엇을 나누는가에서 구분된다. 첫째는 마크업의 위계 제약에서 벗어나는 대신 판본의 편집학적 축적과 도구를 함께 내려놓는다. 둘째는 두 산출물의 일관성을 파일 하나로 보장하는 대신 판본 파일이 데이터의 저장소를 겸하게 된다. 셋째는 각 산출물의 책임 소재를 분리하는 대신 연결부의 내구성을 스스로 설계해야 하는 부담을 진다. 이 셋이 자의적 구분이 아니라 현재의 논의가 실제로 분기해 있는 지점이라는 사실은 판본 편찬에서의 그래프 데이터 모델과 시맨틱 웹 기술을 다룬 최근의 편저가 이 세 방향의 사례를 함께 싣고 있다는 데서도 확인된다(Spadini, Tomasi, and Vogeler 2021).

본 연구는 I부에서 이원 구축 방침을 선언하면서 그 근거를 이 절로 미루어 두었다. 이하의 두 항목이 그 근거를 앞의 두 유형에 대한 검토로부터 이끌어 낸다.

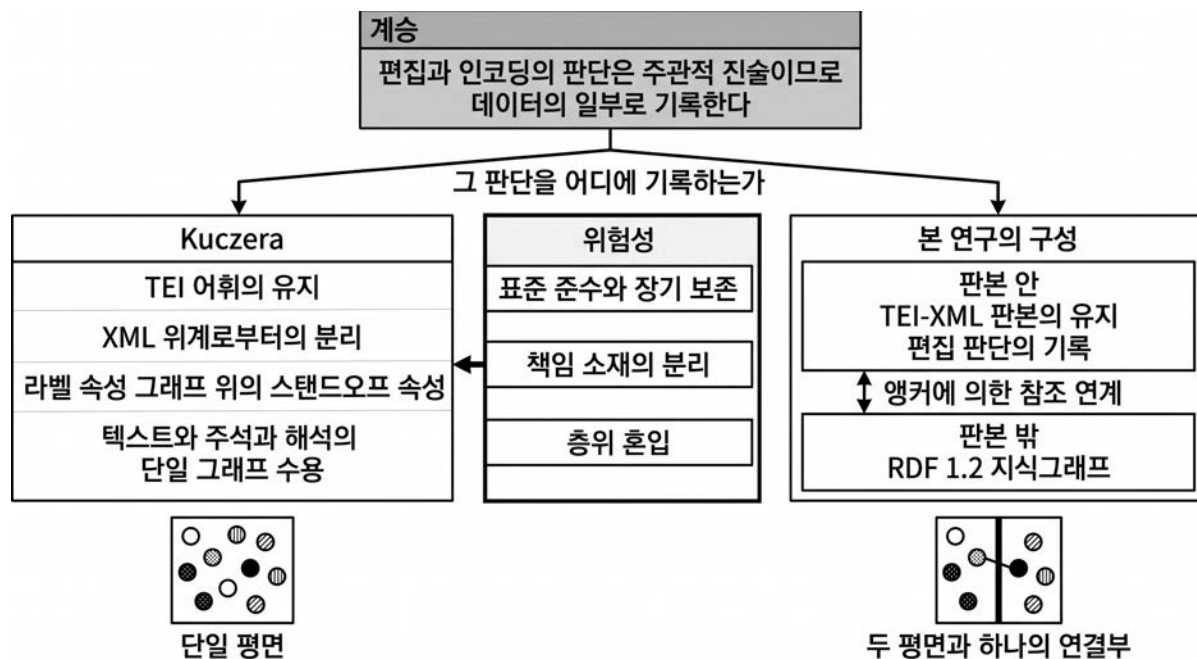
## 2) 판본의 그래프 전환과 그 대가

첫째 유형의 문제의식은 본 연구와 상당 부분 겹친다. 편집자의 모든 결정이 논의의 여지가 있는 주관적 진술이고 해석은 첫 글자가 전사되는 순간부터 시작되므로 판본 전체를 출처 정보를 내장한 지식그래프로 보아야 한다는 주장(Kuczera 2020, 115쪽)은 인코더의 해석을 데이터의 일부로 기록해야 한다는 본 연구의 전제와 같은 자리에 선다. 그 구현은 XML의 위계에서 TEI 시맨틱스를 떼어 내어 문자 색인 기반의 스탠드오프 속성으로 재구성하는 것이었고, 그 결과 하나의 주석이 복수의 의미 공간에 동시에 배속될 수 있게 되었다(Kuczera 2020, 109-110 쪽). 마크업 자체를 온톨로지로 재표현하는 EARMARK 계열도 판본의 구조를 그래프 언어로 옮긴다는 점에서 같은 유형에 속한다(Di Iorio, Peroni, and Vitali 2011).

본 연구가 이 길을 택하지 않는 근거는 네 가지이다. 첫째는 표준 준수와 장기 보존의 부담이다. TEI 지침은 중첩 위계 문제에 대하여 형식적 단순성과 표현력과 기계적 검증 가능성을 모두 갖춘 해법이 현재 존재하지 않으며 비 위계 정보의 표상은 필연적으로 이해득실의 절충이라고 명시한 뒤(TEI Consortium 2026, 21장), 비XML 기반의 여러 제안을 열거하고서도 TEI가 현재 XML에 기반하므로 더 기술하지 않으며 이들을 TEI와 함께 쓰면 확장이 필요하고 대개 비적합 문서가 된다고 밝힌다(TEI Consortium 2026, 21.5절). 판본을 그래프 언어로 옮기는 순간 본 연구는 이 범위 밖으로 나가게 된다.

둘째는 책임 소재의 분리이다. 본 연구의 산출물은 저본과 대교 자료의 대조를 거친 한문 비판정본과 그로부터 별도로 구축되는 지식그래프이며, 전자에 대한 책임은 편집학적 판단에, 후자에 대한 책임은 모델링 판단에 있다. 두 종류의 판단을 하나의 그래프 안에 병치하면 어느 노드가 판독의 결과이고 어느 노드가 해석의 결과인지가 그래프 내부의 유형 표시에만 의존하게 된다. 셋째는 그와 맞물린 층위 혼입의 위험이다. 앞에서 확인한 손실은 문헌의 층위와 인코더의 층위가 한 평면에서 섞일 때 발생하였는데, 판본을 그래프로 흡수하면 그 평면이 다시 하나가 된다. 넷째는 상호운용성이다. 앞 절에서 RDF 1.2를 택한 이유가 어휘의 재사용과 외부 연결에 있었던 이상, 판본 쪽에서도 같은 기준이 적용되어야 한다.

여기에 앞 절에서 미루어 둔 논점 하나가 해소된다. 동일한 정보량에서 RDF 그래프의 규모가 속성 그래프의 열배에 이를 수 있다는 지적(Kuczera 2020, 116쪽)은 판본의 텍스트 구조와 문자 단위 색인까지 그래프 안으로 들어오는 구성을 전제한 것이다. 판본을 그래프 바깥에 두고 참조로만 잇는 구성에서는 그래프가 문자 색인을 부담하지 않으므로 이 논거의 무게가 달라진다. 두 판단이 상이한 이유는 설계의 우열이 아니라 그래프가 무엇을 담기로 하였는가의 차이이다.



[도해 7] 디지털 판본과 지식그래프의 관계 설정

### 3) TEI 내부에 RDF 근거 내장하는 방식

둘째 유형은 판본을 XML로 유지하면서 그 안에 트리플의 근거를 심는다. 이 계열을 가장 체계적으로 비교한 작업은 반 고희 서한집을 대상으로 다섯 가지 방식을 나란히 놓고 평가한 논의이다. 다섯 방식은 변환 스크립트를 외부에 두고 TEI에는 그 연결만 남기는 GRDDL 방식, 같은 방식에서 URI 표현조차 TEI 파일에 남기지 않는 변형, 후 편집 과정에서 생성한 트리플을 헤더의 `xenodata`에 집약하고 본문 요소에는 참조 속성만 남기는 방식, 트리플을 `relation` 요소로 표현하는 방식, RDFa로 트리플을 본문에 분산시키는 방식이다(Boot and Koolen 2021, 15-22쪽).

이들의 평가 기준으로 저자들이 세운 것은 최소한의 방해성, 최대한의 명시성, 최소한의 중복성, 최대한의 일반성과 유연성, 그리고 가능하면 TEI를 확장하지 않는다는 다섯이다(Boot and Koolen 2021, 14쪽). 이 가운데 명시성의 규정이 본 연구에 직접 관여한다. 판본 생성에 필요한 모든 정보를 가능한 한 XML 파일에 포함하고, 포함하지 않는 경우에는 추가 정보를 얻는 방법을 모호하지 않게 지시해야 하며, 보존된 TEI 파일만으로 판본을 재생성할 수 있어야 한다는 요구이다(Boot and Koolen 2021, 14쪽). 저자들이 `xenodata` 방식을 선호한 이유도 여기에 있다. 편집자는 RDF가 없는 원본 XML을 계속 편집하고, 파일이 버전 관리 시스템에 수집될 때 후편집 프로그램이 트리플을 생성한 풍부화 파일을 만들어 내므로, 편집용 파일의 가독성을 유지하면서도 보존된 파일 하나가 완전한 RDF 표현을 담게 된다(Boot and Koolen 2021, 18-19쪽).

반대로 GRDDL 계열은 명시성에서 가장 취약하다는 평가를 받는다. 트리플을 생성하려면 XSLT 프로세서를 실행해야 하는데 장기 보존 환경에서 수십 년 뒤에 그것을 큰 노력 없이 실행할 가능성이 낮고, 표준 자체가 여전히 이론적 규격에 머물러 있으며, 보존된 파일에는 트리플의 표현이 남지 않기 때문이다(Boot and Koolen 2021, 16쪽, 25쪽). 이 평가가 본 연구에 갖는 함의는 분명하다. 출처 환원성을 표방하는 설계에서 앵커의 정의가 실행

가능한 스크립트에만 존재한다면, 이용자가 소급할 수 있는 것은 트리플이 아니라 프로그램이 된다. 반증의 대상이 데이터에서 도구로 옮겨 가는 것이다.

다만 이 계열의 성과를 그대로 셋째 유형의 반대편에 세워서는 안 된다. 저자들이 생성하는 RDF는 저작과 문헌과 판본 표현과 문단과 페이지를 기술하는 판본 도메인의 그래프이지, 인물과 사승과 학파를 기술하는 지식그래프가 아니다. 두 종류의 그래프는 층위가 다르며, 따라서 경쟁하지 않는다. 본 연구는 이 점을 이용하여 다음과 같이 구성한다. 판본 도메인의 기술은 `xenodata` 방식을 따라 TEI 파일 안에서 생성하고 보존하며, 그 결과 각 텍스트 단위에 부여된 식별자가 곧 앵커의 목록이 된다. 그 앵커를 대상으로 삼는 주장의 층위, 곧 앞의 두 절에서 다섯 단계로 나누고 표기를 정한 구체화자들은 판본 바깥의 지식그래프에 둔다. 셋째 유형을 택하면서 둘째 유형의 산출물을 연결부로 흡수하는 구성이며, 이렇게 하면 판본 파일은 자신이 무엇을 담고 있는지를 스스로 진술하되 그것에 관한 해석까지 담당하지는 않는다. 파일의 형태로 적으면 다음과 같다.

```
<TEI>
  <teiHeader>
    <xenoData>
      <rdf:RDF>
        <edd:LocatedTextFragment rdf:about="#q-yujongnok-01">
          <edd:fragmentOfWork rdf:resource="work/yujongnok"/>
          <edd:locatedInDocument rdf:resource="donguhagan/ed2008#p476"/>
        </edd:LocatedTextFragment>
      </rdf:RDF>
    </xenoData>
  </teiHeader>
  <text>
    <quote xml:id="q-yujongnok-01" source="#yujongnok">吾東之學出於圃隱。.....
  </quote>
</text>
</TEI>
```

본문 요소에는 식별자만 남고, 그 식별자가 어느 저작의 부분이며 어느 문헌의 어느 자리에 실현되었는가는 헤더에 집약된다. 편집자가 다루는 본문은 트리플로 채워지지 않으면서도 보존된 파일 하나가 판본 도메인의 완전한 기술을 담는 셈이다. 다만 여기에 적히는 것은 판본이 무엇으로 이루어졌는가까지이며, 이 인용을 하검진이 무슨 목적으로 편의 첫머리에 놓았는가라는 주장은 판본 바깥의 지식그래프에 놓인다. `#q-yujongnok-01`이라는 식별자가 두 산출물이 만나는 유일한 접점이다.

## 4) 두 위계의 교차 지점

이 구성이 곧바로 부딪히는 난점이 있다. 저자들이 논리적 텍스트 위계와 물리적 페이지 위계의 교차점을 위치화된 텍스트 단편이라는 별도 개체로 세운 것은 이 난점에 대한 대응이었다. 하나의 문단이 두 페이지에 걸쳐 나뉘면 그 문단의 각 부분은 문단에도 속하고 페이지에도 속하는데, 이 교차점은 대개 XML에 명시적 요소로 존재하지 않고 논리적 위계와 마일스톤 기반 물리적 위계의 교차로만 암묵적으로 존재한다(Boot and Koolen 2021, 22 쪽).

『동유학안』에서는 구조적으로 같은 문제가 다른 차원에서 발생한다. 발췌된 인용문은 원저작의 텍스트 위계에 속하는 동시에 편찬물의 서술 위계에 속한다. 제2편 총론 첫머리의 인용문은 『유종록』의 어느 대목이면서 동시에 이 학원원제유학안 총론의 서두이며, 이 두 소속은 어느 한쪽으로 환원되지 않는다. 전자를 지우면 발췌라는 행위가 사라지고 후자를 지우면 배치라는 행위가 사라지는데, 앞에서 확인한 대로 그 둘이 곧 학안이 학술사를 서술하는 방식이다.

교차점을 XML 요소로 직접 표현하려 하면 부담이 곧바로 드러난다. 인용문의 경계와 강조나 교감의 구간이 어긋날 때마다 교차 요소를 반복 분할해야 하고, 같은 논리적 단위에 속하는 조각들을 다시 속성으로 이어야 하므로, 편집자가 직접 다루는 파일에서는 실용적이지 않다(Boot and Koolen 2021, 23쪽). 그렇다고 교차점의 생성을 변환 스크립트에만 맡기면 반대의 문제가 생긴다. 생성된 RDF의 식별자에 대응하는 텍스트 대상이 판본 안에 존재하지 않게 되어, 주석이 실제로 무엇을 대상으로 삼는지 확인하려면 스크립트의 논리를 조사해야 하기 때문이다(Boot and Koolen 2021, 24쪽). 저자들이 후편집 스크립트만을 실행 가능한 해법으로 판단하고, 교차 단편에 대해서는 유형을 표시한 `seg` 요소를 생성하여 판본 안에 실체를 남긴 것도 이 때문이다(Boot and Koolen 2021, 24쪽).

본 연구도 같은 결론에 이른다. 발췌 단위는 판본 안에 실체를 갖는 요소로 존재해야 하며, 그 요소에 부여된 식별자가 원저작 위계와 편찬 위계 양쪽으로 이어지는 두 개의 참조를 갖는다. 텍스트 위치 참조의 다중성을 수용해야 한다는 요구가 여기서 구현의 형태를 얻는다. 다만 어느 층위까지 교차 단편을 세울 것인가는 인용문의 분할과 병합이 실제로 어떤 양상으로 나타나는지에 달려 있으므로, II.4에 제시한 인용 사례 조사에서 확인할 사항으로 남긴다.

## 5) 주석의 소재와 지시의 방식

앵커의 문제는 흔히 여러 기법의 목록으로 제시되지만, 그 목록에는 성격이 다른 두 물음이 섞여 있다. 하나는 주석을 어디에 둘 것인가라는 소재의 물음이고, 다른 하나는 텍스트의 어느 지점을 무엇으로 가리킬 것인가라는 지시의 물음이다. 두 축은 독립적이어서, 같은 지시자를 본문 안에 둘 수도 있고 판본 파일 바깥에 둘 수도 있다. 소재의 물음에 대한 답은 앞 항목에서 이미 정해졌다. 판본이 무엇으로 이루어졌는가에 관한 기술은 판본 파일 안에서 생성하여 보존하고, 그 기술이 부여한 식별자를 대상으로 삼는 주장의 층위는 판본 바깥의 지식그래프에 둔다는 분업이 그것이다. 따라서 이하의 비교는 지시의 물음 하나로 좁힌다.

지시자를 담아 유통시키는 형식으로는 Web Annotation Data Model을 쓴다. 이 표준은 지시의 기법 가운데 하나로서 다른 기법과 경쟁하는 것이 아니라, 어떤 지시자를 쓰든 그것을 하나의 주석 단위로 포장하는 용기이다. 표준의 구조는 주석이 본체와 대상으로 이루어진 뿌리 있는 유향 그래프라는 원칙 위에 서며, 대상이 자원 전체가 아니라 그 일부일 때 대상과 원자원 사이에 특정 자원이라는 개체를 두고 거기에 선택자를 붙인다(Sanderson 외 2017, 2절, 4절). 하나의 주석이 복수의 대상을 가질 수 있으므로 앞 항목에서 확인한 이중 소속이 표현될 자리도 여기에 있다. 그리고 결정적으로, 이 표준이 제공하는 선택자의 목록이 아래에서 나눌 세 계열과 그대로 대응한다.

## 6) 세 지시자 계열의 비교

지시 축의 논의를 세 계열로 묶는다. 구조 기반은 판본이 이미 인정한 단위를 그 식별자나 위계상의 좌표로 가리키는 방식이고, 위치 기반은 요소의 존재와 무관하게 문자열 안의 좌표로 가리키는 방식이며, 내용 기반은 인용된 문자열과 그 전후 문맥으로 구간을 특정하는 방식이다. 이 삼분이 자의적이지 않다는 점은 TEI가 자신의 포인터 체계를 도입하면서 세운 규정에서 확인된다. TEI 문서는 문자 스트림 위에 겹쳐진 트리 구조로 간주되며, 노드는 트리에서만 얻어지고 점은 스트림에서만 얻어지므로, 문자 위치에 의존하는 스킴들은 노드를 고려할 수 없고 반대로 XPath는 노드를 분할 불가능한 단위로 다루어 노드의 일부를 지시하지 못한다(TEI Consortium 2026, 17.2.4.1절). 구조와 위치는 서로 다른 대상을 향하며, 어느 한쪽이 다른 쪽을 대신할 수 없다.

세 계열을 하나의 예시에 적용하여 대비한다. 지시 대상은 제2편 총론 첫머리에서 『유종록』을 인용한 구절이며, 판본에서 이 구절이 식별자 `q-yujongnok-01` 을 가진 인용 요소로, 그것을 포함한 단락이 `p-476-01` 로 인코딩되어 있다고 가정한다(하경진 2008, 원문편 476쪽).

구조 기반의 표기는 두 가지로 나뉜다. 하나는 축약 포인터로 요소의 식별자를 그대로 참조하는 `#q-yujongnok-01` 이고, 다른 하나는 학계의 관행적 인용 형식을 URI로 번역하는 정준 참조(canonical reference)이다. 후자는 헤더의 참조 체계 선언에 변환 규칙을 두고 본문에서는 관행적 문자열만 적게 하는 구조로, 규칙은 정규식 대응과 치환의 쌍으로 기술되거나 인용 구조 요소의 중첩으로 기술된다. 뒤쪽 방식은 더 간결할 뿐 아니라 선언된 구조를 이용하여 문서로부터 인용 가능한 참조의 목록을 생성해 낼 수 있다는 이점을 갖는다 (TEI Consortium 2026, 17.2.5절, 17.2.5.4절). 판본이 인정한 단위마다 인용 문자열을 자동으로 산출할 수 있다는 뜻이므로, 앵커의 생성을 수작업에 의존하지 않아도 된다.

이 계열에서만 발생하는 문제가 식별자의 투명성이다. 편과 학안과 인물 항목과 인용문의 위계를 반영한 참조 문자열을 쓰면 사람과 소프트웨어가 모두 구조를 읽을 수 있으나, 불투명한 무작위 식별자를 쓰면 위계를 XML 구조만으로 생성할 수 없으므로 식별자를 파일 안에 직접 기입해야 한다(Boot and Koolen 2021, 26쪽). 앵커 생성의 자동화 가능성과 판본 개정 시의 내구성이 이 선택에 따라 달라지며, 결정은 편의 배열과 인물 수록 순서가 증보 과정에서 얼마나 안정적이었는가에 달려 있다.

위치 기반의 표기는 문자 스트림의 좌표를 쓴다. TEI는 점을 지시하는 스킴과 구간을 지시하는 스킴을 나누어 정의하며, 특정 노드를 기준으로 한 문자 위치를 점으로 삼고, 두 점 사이의 시퀀스나 시작 위치와 길이의 쌍으로 구간을 지시한다. 이 계열의 강점은 요소 경계와 무관하다는 데 있어서, 지시하는 구간이 시작 태그만 포함하고 종료 태그를 포함하지 않는 요소를 가로질러도 성립한다(TEI Consortium 2026, 17.2.4.6절). 같은 항목이 든 예시가 본 연구의 조건과 정확히 겹치는데, 교감 항목 앞에서 시작하여 하나의 판독을 포함하고 그 뒤로 이어지는 구간을 지시할 수 있다는 것이다. 자체 비판정보에는 교감 장치가 들어가므로, 증거로 삼을 구절이 교감 항목을 가로지르는 경우가 실제로 발생한다.

그러나 이 계열의 내구성은 가장 낮다. 본문이 한 글자만 바뀌어도 좌표가 어긋나며, 문자를 세는 방식 자체가 유니코드 정규화 형식에 따라 달라지므로 결합 문자를 쓴 표기와 완성형 표기가 서로 다른 개수로 계산된다는 경고가 지침에 명시되어 있다(TEI Consortium 2026, 17.2.4.5절). 한문 원문에서 이 경고는 사변적이지 않다. 이 체자의 정규화 여부, 표점 부호의 삽입 여부, 결자와 공격을 어떻게 표기하는가가 모두 문자 수를 바꾼다.

내용 기반의 표기는 문자열 자체로 구간을 특정한다. TEI는 기준 노드에 상대적인 텍스트 스트림에 정규식을 대응시키는 스킴을 정의하고, 대응이 복수일 때 몇 번째를 취할지를 지정하게 한다(TEI Consortium 2026, 17.2.4.8절). 같은 발상이 Web Annotation Data Model에서는 인용 문자열과 그 앞뒤 문맥의 삼분으로 나타나며, 동일한 문자열이 여러 번 나타나는 문제를 전후 문맥으로 해소한다(Sanderson 외 2017, 4.2.4절). 이 계열은 파일 구조의 변경에 강하다. 요소를 재편하거나 파일을 나누어도 문자열이 남아 있는 한 지시가 유지되기 때문이다. 반대로 문자열의 변경에는 취약하며, 1970년본과 2008년본의 표점 차이가 이 취약성이 본 자료에서 실제로 작동하는 지점이다. 표점의 위치는 인용문의 경계 판정을 바꿀 수 있으므로, 경계를 문자열로 특정한 앵커는 표점 판단이 수정되는 순간 다른 구간을 가리키게 된다.

세 계열을 같은 구절 위에 나란히 적으면 각자가 무엇에 의존하는지가 드러난다. 표기는 주석 표준의 선택자로 통일하고, 대응하는 TEI 포인터를 주석으로 병기한다.

```
# 구조 기반. TEI 축약 포인터로는 #q-yujongnok-01
ex:anchor_s oa:hasSource <.../donguhagan/ed2008.xml> ;
            oa:hasSelector [ a oa:FragmentSelector ;
                            rdf:value "q-yujongnok-01" ] .

# 위치 기반. TEI 스킴으로는 #range(...) 또는 #string-range(...)
ex:anchor_p oa:hasSelector [ a oa:TextPositionSelector ;
                            oa:start 1204 ; oa:end 1212 ] .

# 내용 기반. TEI 스킴으로는 #match(...)
ex:anchor_c oa:hasSelector [ a oa:TextQuoteSelector ;
```

`oa:exact` "吾東之學出於圃隱" ;  
`oa:prefix` "儒宗錄有曰。" ;  
`oa:suffix` "圃隱傳之治隱。" ] .

셋은 같은 구절을 가리키지만 의존하는 조건이 다르다. 첫째는 판본이 그 자리를 하나의 요소로 인정하였다는 사실에, 둘째는 그 앞의 문자 수가 변하지 않는다는 사실에, 셋째는 그 문자열과 앞뒤 문맥이 유지된다는 사실에 각각 기대다. 의존하는 바가 다르므로 무너지는 조건도 다르며, 아래 판정표의 개정 내구성 항목은 그 차이를 정리한 것이다.

세 기준에 따른 판정은 다음과 같다.

| 기준         | 구조 기반                 | 위치 기반                 | 내용 기반                         |
|------------|-----------------------|-----------------------|-------------------------------|
| 개정 내구성     | 높음. 식별자가 유지되는 한 안정    | 낮음. 한 글자의 증감으로 어긋남    | 중간. 구조 변경에 강하고 교감과 표점의 변경에 약함 |
| 지시 입도      | 판본이 인정한 단위에 한정        | 문자 단위로 임의             | 문자 단위로 임의                     |
| 표준화와 구현 여건 | 넓음. 축약 포인터는 범용 도구가 처리 | 좁음. TEI 스킴의 파서 지원이 결여 | 넓음. 주석 표준의 선택자가 널리 구현         |

셋째 기준의 판정은 이 절의 검토에서 가장 실질적인 함의를 갖는다. 위치 기반 계열이 정의된 지 오래되었음에도 구현되지 않은 채 남아 있다는 사정은 이미 2010년에 지적되었고, 그 결과 대규모 코퍼스 구축에서 스탠드오프 포함 방식을 포기하고 의미가 약하게 규정된 대응 속성으로 후퇴한 사례가 보고되었다(Bański 2010, 5.1절). 표기의 정의가 이루어진 지 오래되었다는 사실과 그 표기를 처리할 도구가 존재한다는 사실이 별개라는 뜻이며, 이 어긋남이 세 계열의 선택을 실질적으로 좌우한다.

세 계열은 배타적 선택지가 아니다. Web Annotation Data Model은 하나의 특정 자원에 복수의 선택자를 붙일 수 있게 하되 그것들이 같은 내용을 선택해야 하며 이용자 에이전트가 그중 하나를 골라야 한다고 규정하고, 그 취지를 나중에 구간이 발견될 가능성을 최대화하는 데 둔다(Sanderson 외 2017, 4.2절). 나아가 하나의 선택자를 다른 선택자로 정련하는 장치를 두어, 넓은 단위를 먼저 지시하고 그 안에서 좁은 구간을 특정하는 결합을 허용한다(Sanderson 외 2017, 4.2.9절). 구조 기반 지시를 주 지시자로 두고 그 안에서 내용 기반 지시로 구절을 좁히는 구성이 표준의 어휘로 그대로 표현된다는 뜻이다.

같은 표준이 또 하나의 장치를 제공하는데, 대상 자원의 어느 상태를 두고 한 주석인지를 기록하는 상태 개체이다. 위치 기반 선택자가 변경에 매우 취약하므로 올바른 표상을 특정하기 위해 상태를 함께 쓰기를 권고한다는 주석이 명세에 달려 있고, 상태는 시점이나 요청 조건으로 기술되며 선택자로 정련될 수 있다(Sanderson 외 2017, 4.2.5절, 4.3절, 4.3.3절). 이 장치가 본 연구에서 갖는 의미는 부수적이지 않다. 저본과 대교 자료를 대조하여 판독을 선택하는 작업이 인코더의 편집 충위를 이루는 이상, 하나의 증거 앵커는 어느 판독을 취한 상태의 텍스트를 가리키는지를 함께 진술해야 한다. 디지털 환경의 비판정본이 고정된 산출물이 아니라 판독의 선택이 갱신될 수 있는 유동적 대상이라는 관점(이재열 2024)에서 보면, 상태의 명시는 선택이 아니라 요건이 된다.

중첩 위계를 둘러싼 편집학의 오랜 논의를 여기서 압축적으로 정리해 둔다. 정보를 여러 형태로 중복 인코딩하는 방식, 빈 요소로 경계를 표시하는 방식, 논리적 단위를 조각내고 가상으로 재결합하는 방식, 그리고 스탠드오프 마크업이 지침에 정리되어 있으며, 각각 유지 비용과 처리 난이도와 의미론적 정확성에서 서로 다른 대가를 치른다(TEI Consortium 2026, 21.1절-21.4절). 본 연구는 이 문제를 판본 구조의 개편으로 풀지 않고 참조 충위의 설계로 우회한다. 원저작 위계와 편찬 위계의 교차는 판본 안에서 하나의 단위를 설정하고 그 단위에 두 개의 참조를 부여하는 것으로 처리하며, 그 두 참조가 무엇을 뜻하는지는 판본이 아니라 지식그래프의 어휘가 진술한다. 판본 쪽의 링크는 대응만을 주장하고 그 대응의 성격에 대한 해석은 그래프 쪽이 맡는다는 분업이며, 이는 스탠드오

프 링크의 의미론 가운데 가장 약하게 규정된 대응 의미론이 실무에서 실제로 채택되어 온 사정과도 부합한다 (Bański 2010, 3절).

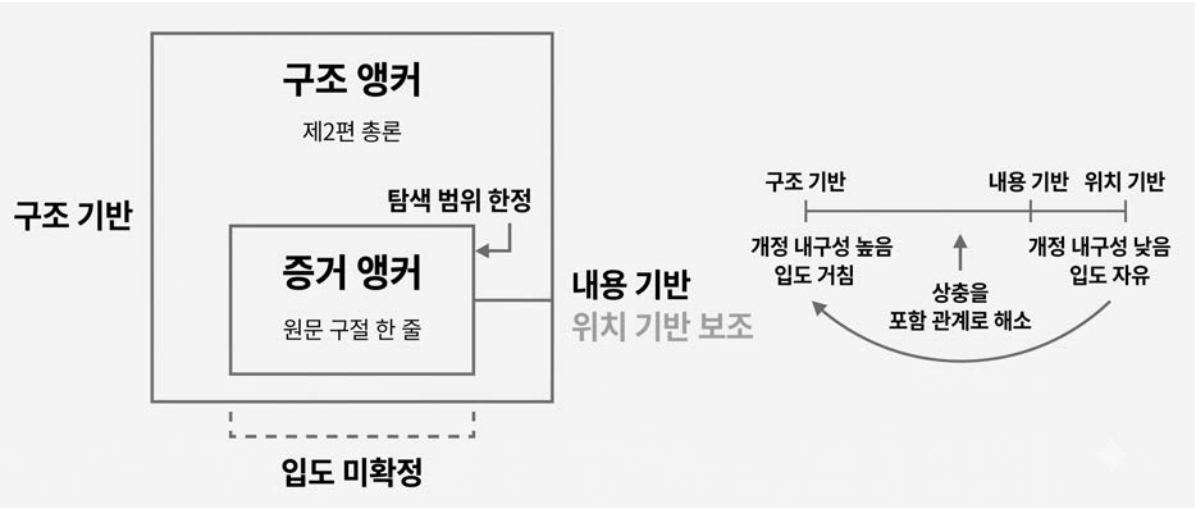
## 7) 잠정 선택: 이중 앵커

이상의 검토에서 두 종류의 앵커를 구분하는 방향, 곧 이중 앵커(dual anchor)의 구성이 도출된다. 구조 앵커(structural anchor)는 편과 학안과 인물 항목과 인용문처럼 판본이 자기 체재로 인정하는 안정적 단위를 식별한다. 정준 참조 체계로 선언하여 참조 문자열을 자동 생성할 수 있게 하고, 판본이 개정되어도 단위가 유지되는 한 지시가 유지되게 한다. 증거 앵커(evidential anchor)는 특정 주장을 뒷받침하는 문장이나 구절의 범위를 식별한다. 상위의 구조 앵커에 종속시키고, 그 안에서 내용 기반 지시를 주 지시자로 삼되 생성 시점의 위치 기반 좌표를 보조로 병기하며, 어느 판독 상태를 근거로 삼았는지를 상태 개체로 함께 기록한다. 앞 절에서 비워 둔 자리, 곧 구체화자마다 붙은 증거 속성의 값이 이 증거 앵커이다. 그 자리를 실제로 채우면 다음과 같다.

```
ex:citation_476 a oa:SpecificResource ;
  oa:hasSource   ex:struct_ch2_prologue ;           # 구조 앵커
  oa:hasState    [ a oa:TimeState ;                 # 판독 상태
    oa:sourceDate "2026-08-23T00:00:00Z"^^xsd:dateTime ;
    ex:readingSet ex:collation_v1 ] ;
  oa:hasSelector [ a oa:TextQuoteSelector ;          # 주 지시자
    oa:exact      "吾東之學出於園隱" ;
    oa:refinedBy  [ a oa:TextPositionSelector ;
      oa:start 1204 ; oa:end 1212 ] ] .
```

상위의 구조 앵커가 대상 자원의 범위를 한정하고, 그 안에서 내용 기반 지시자가 구절을 특정하며, 위치 기반 좌표는 정련의 형식으로 병기된다. 셋째 줄의 상태 개체는 이 앵커가 어느 판독을 취한 텍스트를 가리키는지를 진술한다. 대교의 결과 본문으로 채택한 독법이 바뀌면 상태가 바뀌고, 바뀌었다는 사실 자체가 데이터에 남는다. 이로써 앞 절의 구체화자들이 근거로 삼는 판본의 지점이 이름뿐이던 상태를 벗어난다.

이 구성은 식별자의 안정성과 증거의 환원 가능성을 함께 확보하려는 절충이며, 앞 항목의 판정표를 그대로 반영한다. 개정 내구성이 높으나 입도가 제한된 계열을 상위에 두고, 입도가 자유로우나 취약한 계열을 하위에 두어 상위 앵커가 하위 앵커의 복원 범위를 좁혀 주는 구조이다. 인용문의 경계 판정이 표점의 수정으로 바뀌더라도 그 인용문이 속한 구조 앵커는 그대로이므로, 어긋난 증거 앵커를 어디에서 찾아야 하는지가 확정된다.



[도해 8] 이중 앵커 구조

현 착수 상태에서는 두 가지 유보 사항이 있다. 우선, 앵커의 정밀도를 확정하지 않았다. 지시 가능한 단위를 지나치게 세분하면 이용자가 부적절한 대상을 고르게 되어 오히려 주석의 재사용성과 상호운용성이 낮아진다는 지적(Boot and Koolen 2021, 26쪽)이 있다. 정밀도에 대한 확정은 표본 구간 검토 이후로 미룬다. 다음으로 트리플을 생성하고 보존하는 세부 공정은 추후의 작업으로 미뤘다. 인간 인코더의 판단과 기계가 생성한 후보를 분리하여 기록한다는 방침이 이 요구와 맞물리는 지점은 다음 절에서 다룬다.

이로써 앞의 네 절이 각각 남긴 물음을 모두 잠정적으로 검토했다. 진술을 사실이 아니라 기록의 단위로 삼는 틀을 택하였고, 그 틀이 비워 둔 자리를 채울 어휘의 조합을 검토했다. 인용 행위를 독립된 개체로 설정하여 귀속의 연쇄를 다섯 단계로 분해하였고, 그 단계들을 그래프에서 구분하는 표기를 정하였다. 또 각 단계가 판본의 어느 지점에 닿는지를 지시하는 방식을 정하였다. 남은 것은 파일럿 작업을 통해 이 결정의 적절성을 판단하고, 서로 어떤 층위 관계에 놓이는지를 조감하는 것이다.

## IV. 에필로그

### 1. 현재의 잠정적 선택

첫째, 판본과 지식그래프를 하나의 산출물로 통합하지 않고 각각의 표준 생태계 안에서 구축한 뒤 참조로 연계한다. 이 방침의 근거는 세 가지이다. 표층에서는 표준 준수와 장기 보존의 이점이 있고, 중층에서는 편집학적 판단과 모델링 판단의 책임 소재가 분리된다는 이점이 있으며, 심층에서는 문헌의 층위와 인코더의 층위를 한 평면에서 섞지 않는다는 본 연구의 전체 자체가 걸려 있다. 판본을 그래프로 흡수하면 판독의 결과인 노드와 해석의 결과인 노드가 같은 공간에 놓이게 되고, 그 둘을 변별하는 것은 그래프 내부의 유형 표시뿐이 된다. 각 진술의 발화 주체를 되물을 수 있어야 한다는 요구가 산출물의 구성 자체에서 관철되는 지점이 여기이다.

둘째, 그래프 데이터 모델로 RDF 1.2를 택한다. 이는 라벨 속성 그래프 대비 질의 성능의 우위나 표현의 경제성 때문이 아니라 어휘의 재사용과 식별자의 전역성, 그리고 모델 정의 계층의 개방성에 관한 판단에서 비롯된다. 믿음과 추론과 파생과 주석의 어휘가 모두 RDF로 정의되어 있어야 그 표준 의미론을 그대로 이어받을 수 있고, 산출물이 특정 상용 제품의 수명과 정책에 종속되지 않아야 뒷날의 검증이 가능하기 때문이다. 다만 이 선택이 아직 완결되지 않은 표준에 의존한다는 사실은 위험 요인이라고 할 수 있다.

셋째, 인용 행위를 진술에 붙는 속성값이 아니라 독립된 개체로 기술한다. 이 결정은 표현의 정밀도를 높이려는 선택이 아니라 잘못된 귀속을 막기 위한 조건이다. 발체의 대상이 된 저작을 독립된 개체로 설정하고 발체라는 행위와 구분할 수 있어야 한다는 요구, 그리고 편찬자의 판단을 사실 진술과 같은 지위로 기록하지 않아야 한다는 요구가 이 하나의 결정에 함께 걸린다. 인용을 속성값으로 처리하면 편찬자가 무엇을 옮겼는가라는 사실과 인코더가 그것을 어떤 목적의 인용으로 읽었는가라는 판단이 같은 자리에 겹치기 때문이다.

이 세 가지 선택은 어느 것도 확정된 것은 아니다. 특히 셋째 결정은 인용의 유형이 자료에서 실제로 어떤 분포를 보이는가에 따라 구조의 복잡도가 크게 달라진다. 어휘를 여러 계열에서 골라 결합하는 설계가 오히려 장황함을 가중할 수 있다는 우려는 이미 국내 논의에서 제기된 바 있으며(김바로·이재열 2025, 12-13쪽), 그 우려에 대한 응답은 자명한 진술에 메타데이터를 부착하지 않는다는 원칙의 적용 범위를 자료로 확인한 뒤에야 가능하다.

### 2. 생성형 AI 활용 가능성의 검토

구축 공정에 생성형 AI를 들일 것인가는 앞의 세 결정과 층위가 다른 물음이지만, 이 연구에서는 무관한 물음이 아니다. 해석 책임성을 표방하는 설계가 자기 공정에 개입한 기계의 판단을 기록하지 않는다면 그 표방은 자기모순이 되기 때문이다. 따라서 물어야 할 것은 AI를 쓸 것인가 말 것인가가 아니라, 쓴다면 그 개입을 어느 층위에서 어떤 형태로 데이터에 남길 것인가이다. 이 물음에 답하려면 먼저 이 기술이 학술 편찬의 어느 단계에서 논의되고 있

는지를 확인해야 한다.

그 지형을 가장 체계적으로 정리한 것은 2024년 시점의 실험들을 종합한 조사이다. 이 조사는 학술 편찬을 원자료에서 데이터로서의 판본을 거쳐 간행물로서의 판본에 이르는 공정으로 도식화한 뒤, 생성형 AI가 시도된 영역을 전승의 문서화, 회고적 디지털화 판본의 후처리, 텍스트 확정과 기초 마크업, 정규화, 개체명 인식(named entity recognition)과 심층 의미 주석, 정보 보강, 번역, 요약의 여덟으로 정리하였다(Pollin 외 2025, 문단 8, 문단 11). 이 정리의 고유한 기여는 목록의 작성 자체가 아니라 그 목록이 편찬 공정의 어느 국면에 대응하는가를 함께 밝힌 데 있으며, 저자들이 자기 작업을 완결된 평가가 아니라 특정 시점의 즉시적 기록으로 규정한 점도 여기서 함께 확인해 둘 만하다(Pollin 외 2025, 문단 4). 기술의 상태가 빠르게 변하는 조건에서 어느 실험이 무엇을 시도하여 어디에서 막혔는가를 남기는 일이 평가의 갱신보다 오래 쓸모를 갖는다는 판단이다.

이 지형을 본 연구의 공정에 대응시키면 후보가 두 영역으로 나뉜다. 문헌 표상 쪽에서는 서술 단위의 구조 분절 초안 작성, 인명과 서명과 관직의 개체 표지 후보 추출, 두 판본의 표점과 번역문의 대조 보조가 후보이다. 지식 표상 쪽에서는 온톨로지 초안의 작성, 원문에서의 팩토이드 후보 추출, 매핑 규칙과 질의문 작성의 보조, 구축된 데이터의 일관성 점검이 후보이다. 이 가운데 구조를 갖추지 않은 텍스트를 TEI로 옮기는 과제는 실험이 가장 많이 축적된 영역이며, 단순한 지시만으로는 형식적으로 온전한 XML은 얻어져도 지침에 부합하는 TEI는 거의 얻어지지 않는 반면 예시와 매핑 규칙을 함께 제공하면 결과가 크게 달라진다는 점이 확인되었다(Pollin 외 2025, 문단 26-27).

그러나 같은 조사가 보고한 한계 가운데 본 자료에 특히 무겁게 작용하는 것이 있다. 첫째는 과잉 정규화(over-normalization)이다. 개체명 인식을 요청하였는데 요청하지 않은 정규화까지 함께 수행되어 대소문자와 자소와 구두점이 바뀌고 지시 관계가 임의로 풀려 버리는 현상이 여러 실험에서 관찰되었다(Pollin 외 2025, 문단 22-23). 한문 원문에서 이 현상은 사변적 위험이 아니다. 이체자의 통일, 표점의 임의 삽입, 결자와 공격의 처리는 모두 판독의 결과이자 편집자의 판단인데, 그 판단이 자동으로 수행되어 버리면 판본 총위의 책임이 소리 없이 기계로 이전된다. 둘째는 스키마에 없는 요소의 생성과 원문서 구조의 변경, 그리고 입력 텍스트의 임의 축약이다(Pollin 외 2025, 문단 36). 셋째는 개체 연결에서 형식적 식별자를 지어내는 경향인데, 이는 구조화된 데이터를 함께 제공하여 완화된다고 보고되었다(Pollin 외 2025, 문단 20). 넷째는 언어 편향이다. 영어로 작성된 지시가 다른 언어로 작성된 지시보다 일관되게 높은 품질의 응답을 얻으며, 목표 언어를 명시하거나 예시를 그 언어로 제공하여도 격차가 남는다(Pollin 외 2025, 문단 38). 한문과 한국어를 함께 다루어야 하는 본 연구에서 이 항목은 성능의 문제를 넘어 결과의 신뢰도에 관한 문제가 된다. 다섯째는 재현 가능성과 평가 기준의 부재이다. 출력이 확률적이어서 반복 실행의 결과가 같지 않고, 설명 가능성이 확보되지 않으며, 상용 제공자가 충분한 문서화 없이 엔진을 변경하고, 학술 편찬이라는 특정 영역에 대한 평가 기준이 아직 마련되지 않았다(Pollin 외 2025, 문단 35-36).

이상을 종합하면 원칙은 두 가지로 정해진다. 하나는 인간 인코더가 최종 판단자로 남는다는 것이고, 다른 하나는 기계의 개입 자체를 데이터에 기록한다는 것이다. 뒤의 원칙이 이 연구에서 갖는 의미는 통상적인 투명성 요구를 넘어서는 것이다. 본 연구는 이미 원저작의 진술과 편찬자의 발췌와 인코더의 판단을 서로 다른 층위로 나누어 적기로 하였으므로, 기계가 생성한 후보는 그 연쇄에 단계를 하나 더하는 것이지 새로운 종류의 문제를 일으키는 것이 아니다. 생성 이력의 어휘가 이 자리에 쓰인다. 어떤 후보가 어떤 활동을 거쳐 어떤 입력에서 파생되었고 그 활동의 행위자가 누구인가를 기록하면, 기계가 제안한 것과 사람이 승인한 것이 별개의 개체로 남는다. 판본 파일 쪽에서는 트리플을 생성한 프로그램의 이름을 파일 자체에 명시해야 한다는 요구(Boot and Koolen 2021, 25-26쪽)를 구현의 참조로 삼는다. 승인의 시점과 승인자를 함께 적으면, 뒷날 어느 판단이 기계의 제안을 그대로 받은 것이고 어느 판단이 사람이 뒤집은 것인지가 데이터 안에서 구별된다.

활용의 여부와 범위는 확정하지 않는다. 다만 채택을 판단할 기준 셋을 먼저 세워 둔다. 첫째, 그 과제의 산출물이 사람의 검수를 통과해야만 데이터에 들어가는가. 검수 없이 통과하는 자리에는 쓰지 않는다. 둘째, 검수의 비용이 직접 작업의 비용보다 낮은가. 후보 목록을 만드는 일은 이 기준을 대체로 통과하지만 판독의 확정에는 통과하지 않는다. 셋째, 그 과제가 어느 층위에 속하는가. 문헌 총위의 판독과 교감처럼 되돌리기 어려운 결정에는 적용하지 않고, 식별과 진술의 후보 추출처럼 인코더가 개별적으로 기각할 수 있는 자리에 한정한다. 이 기준의 적용 결과는

파일럿 구축에서 확인한다.

### 3. 다음 단계 작업 계획

이후의 작업은 네 단계로 진행한다. 첫째는 자료 조사이다. 판본과 저본의 확정, 전체 체재 대장의 구축, 표본 구간의 정밀 검토, 인용 사례 대장의 구축, 편찬 행위 사례 카드의 작성이 여기에 속하며, 각 항목이 어느 요구의 검증에 대응하는지는 앞에서 밝혀 두었다. 둘째는 파일럿 구간의 선정이다. 셋째는 그 구간에 대한 TEI-XML 판본과 RDF 그래프의 구축이며, 넷째는 그 과정에서 실제로 내려진 판단을 정리한 편집 지침 초안의 작성이다.

파일럿 구간의 선정 기준을 여기서 밝혀 둔다. 기준은 학파의 대표성이 아니라 모델을 검증할 수 있는 사례의 포함 여부이다. 구체적으로는 표지가 명시된 인용과 표지가 없는 서술이 한 편 안에 함께 나타나는 구간, 편찬자의 언어가 상반된 두 진술을 나란히 놓는 구간, 하나의 원문이 나뉘어 여러 곳에 배치되거나 서로 다른 출전이 하나로 합쳐진 구간, 학파 귀속이 애매하거나 복수 학파에 걸치는 인물이 포함된 구간이 우선한다. 대표적 학파를 골라 그 학파의 전형을 보이는 방식은 모델의 강점만을 확인하게 되므로 택하지 않는다. 구체적 구간의 지정은 자료 조사가 인용의 유형 분포를 확인한 뒤로 미룬다.

## 참고문헌

### 1. 원자료 판본

하검진, 『東儒學案』 上中下, 부산: 해동불교역경원 내 일봉정사, 1970.

하검진 편, 박상리·임옥균·조장연·최복희·함현찬 역, 『증보 동유학안: 한국유학자 167인의 생애와 사상』 1, 파주: 나남, 2008.

### 2. 한국어 문헌

고영진, 「박세채의 학문과 유학사(儒學史) 인식」, 『한국사상사학』 32, 2009, 269-303쪽.

고영진, 「조선시대 유학 계보 연구의 검토」, 『한국사상사학』 41, 2012, 263-306쪽.

김기엽, 「『동유학안』을 통해 본 조선 유학사 서술의 두 가지 양상: 1940년대 편찬된 『동유학안』 2종을 중심으로」, 『한국연구』 13, 2023, 299-322쪽, <https://doi.org/10.23073/riks.2023..13.011>.

김기주, 「회봉(晦峯) 하검진(河謙鎭)의 학문 활동과 성리학적 특징: 『회봉선생연보(晦峯先生年譜)』를 중심으로」, 『한국학논집』 70, 2018, 91-119쪽, <https://doi.org/10.18399/actako.2018..70.004>.

김기주, 「『東儒學案』으로 본 조선성리학의 전개」, 『남명학연구』 61, 2019, 157-187쪽.

김기주, 「淵源錄의 전개사에서 본 『東儒學案』의 의의」, 『남명학연구』 76, 2022, 1-33쪽.

김바로, 「한국역사인물데이터베이스 설계 시론: 『홍문록』을 예시로 하여」, 『한국학』 46(3), 2023, 159-195쪽, <https://doi.org/10.25024/ksq.46.3.202309.159>.

김바로·이재열, 「한국근대인물LOD - 종묘배향공신을 대상으로 -」, 『〈조선시대 종묘 배향공신 연구〉 결과발표회 논문집』, 성남, 2025년 12월.

김서윤, 「조선시대 논어주석의 인용 분류체계: 디지털 경전주석학 연구 시론」, 『민족문화』 67, 2024, 199-232쪽, <https://doi.org/10.15752/itkc.67..202407.199>.

김서윤, 「한국경학자료시스템 XML 검토 및 데이터 설계」, 『대동문화연구』 128, 2024, 201-236쪽, <https://doi.org/10.18219/ddmh..128.202412.201>.

김정신, 「16~17세기 조선 학계의 중국 사상사 이해와 중국 문헌: 『이락연원록신증(伊洛淵源錄新增)』 『황명이학명신언행록(皇明理學名臣言行錄)』을 중심으로」, 『한국사상사학』 58, 2018, 177-227쪽, <https://doi.org/10.31309/SKHT.58.201804.6>.

김지선·류인태, 「여향문화 연구와 데이터 모델링」, 『한국한문학연구』 85, 2022, 79-118쪽.

김태년, 「학안에서 철학사로: 조선유학사 서술의 관점과 방식에 대한 검토」, 『한국학연구』 23, 2010, 41-84쪽.

김훈식, 「寒暄堂 金宏弼의 門人錄에 대한 검토」, 『역사와 경계』 96, 2015, 241-277쪽.

나종일, 「집단전기학」, 『서양사연구』 7, 1985, 1-44쪽.

노관범, 「연원록에서 사상사로: 장지연의 「조선유교연원」과 현상윤의 『조선유학사』를 읽는 방법」, 『한국사상사학』 56, 2017, 209-243쪽, <https://doi.org/10.31309/SKHT.56.201708.7>.

우경섭, 「한국학에서 집단전기학의 가능성에 대한 탐색」, 『한국학연구』 74, 2024, 185-214쪽.

이병태, 「한국 현대 사상사의 재조망과 '모더니티': 20세기 전반 유학사 저술을 중심으로」, 『통일인문학』 85, 2021, 263-292쪽.

이영숙, 「일제강점기 전통 지식인으로서 회봉 하겸진의 위상」, 『남명학연구』 71, 2021, 99-129쪽.

이영숙, 「晦峯 河謙鎭의 남명학 계승 양상」, 『남명학연구』 76, 2022, 121-146쪽.

이일해, 「晦峰先生墓誌銘」, 1961.

이재열, 「동아시아 고문헌의 디지털 비판정본: 유동적 정본의 가능성」, 『한문학보』 51, 2024, 31-66쪽, <https://doi.org/10.35496/HAN.51.2>.

임형석, 「清代 朱子學에 대한 唐鑑의 인식」, 『철학논총』 36-2, 2004, 147-168쪽.

정성희, 「식민지 시기 조선 유학사 정리 작업에 대한 연구: 장지연(張志淵)과 하겸진(河謙鎭)의 저항적 조선유학사 정리 작업을 중심으로」, 『유학연구』 29, 2013, 165-191쪽, <https://doi.org/10.18216/YUHAK.2013.29..007>.

진오균, 「朴世采의 『東儒師友錄』 사본과 구성에 대한 검토」, 『한국사학사학보』 48, 2023, 261-348쪽.

최일범, 「『東儒學案』의 學派 分類에 관한 考察」, 『유교사상문화연구』 21, 2004, 123-140쪽.

최일범, 「『東儒學案』 해제」, 하겸진 편, 박상리·임옥균·조장연·최복희·함현찬 역, 『증보 동유학안: 한국유학자 167인의 생애와 사상』 1, 파주: 나남, 2008, 11-29쪽.

한국학중앙연구원, 「동유학안(東儒學案)」, 『한국민족문화대백과사전』, <https://encykorea.aks.ac.kr/Article/E0016722>.

### 3. 중국·일본 문헌

徐仁甫, 『古書引語研究』, 北京: 中華書局, 2014.

連凡, 「『宋元學案』における學案表と師承關係: その内容と學術的意味について」, 『日本中國學會報』 65, 2013, 155-170쪽.

連凡, 「『宋元學案』的層次結構與學案設置: 兼論全祖望與黃宗羲思想史觀之異同」, 『北京社會科學』 2017-4, 2017, 112-128쪽.

小川潤・大向一輝, 「歴史一次史料の知識構造化のためのFactoidモデルの拡張」, 人工知能学会第二種研究会資料 SIG-SWO-053-06, 2021년 3월 15일, [https://doi.org/10.11517/jsaisigtwo.2021.SWO-053\\_06](https://doi.org/10.11517/jsaisigtwo.2021.SWO-053_06).

小川潤・永崎研宣・大向一輝, 「一次史料における時間的コンテキストを含む社会関係記述モデルの提案と実践」, 『情報処理学会論文誌』 63(2), 2022, 258-268쪽, <https://doi.org/10.20729/00216232>.

小風綾乃・小川潤・亀田堯宙・山田太造・大向一輝, 「歴史人名比定に向けた議事録のモデル化と知識グラフの構築」, 『情報処理学会論文誌』 67(2), 2026, 241-250쪽, <https://doi.org/10.20729/0002007000>.

## 4. 서구 문헌

Angles, Renzo, Marcelo Arenas, Pablo Barceló, Aidan Hogan, Juan Reutter, and Domagoj Vrgoč. "Foundations of Modern Query Languages for Graph Databases." *ACM Computing Surveys* 50, no. 5 (2017): 68:1-68:40. <https://doi.org/10.1145/3104031>.

Bański, Piotr. "Why TEI Stand-off Annotation Doesn't Quite Work: And Why You Might Want to Use It Nevertheless." Presented at Balisage: The Markup Conference 2010, Montréal, August 3-6, 2010. *Balisage Series on Markup Technologies*, vol. 5 (2010). <https://doi.org/10.4242/BalisageVol5.Banski01>.

Bekiari, Chryssoula, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. "Definition of the CIDOC Conceptual Reference Model v7.1.1." CIDOC CRM Special Interest Group, 2021. <https://doi.org/10.26225/FDZH-X261>.

Bekiari, Chryssoula, Martin Doerr, Patrick Le Bœuf, and Pat Riva, eds. "Definition of FRBROO: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism (Version 2.4)." International Federation of Library Associations and Institutions, 2016.

Boot, Peter, and Marijn Koolen. "Connecting TEI Content Into an Ontology of the Editorial Domain." In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, and Georg Vogeler. BoD, 2021.

Bradley, John. "The Factoid Prosopography Ontology." King's College London. Accessed May 8, 2026. <https://www.kcl.ac.uk/factoid-prosopography/ontology>. 및 동일 저자, "What Is Factoid Prosopography All About?" Accessed May 8, 2026. <https://www.kcl.ac.uk/factoid-prosopography/about>.

Bradley, John, and Harold Short. "Texts into Databases: The Evolving Field of New-Style Prosopography." *Literary and Linguistic Computing* 20, Suppl. 1 (2005): 3-24. <https://doi.org/10.1093/lc/fqi022>.

Cagle, Kurt. "RDF 1.2 vs. Neo4j/OpenCypher: A Comparative Architecture Analysis." *The Ontologist*, March 22, 2026. <https://ontologist.substack.com/p/rdf-12-vs-neo4jopenencypher>.

Carroll, Jeremy J., Christian Bizer, Pat Hayes, and Patrick Stickler. "Named Graphs." *Journal of Web Semantics* 3, no. 4 (2005): 247-267. <https://doi.org/10.1016/j.websem.2005.09.001>.

Chu, Hung-lam. "Confucian 'Case Learning': The Genre of Xue'an Writings." In *Thinking with Cases: Specialist Knowledge in Chinese Cultural History*, edited by Charlotte Furth, Judith T. Zeitlin, and Ping-chen Hsiung. University of Hawaii Press, 2007. <https://doi.org/10.1515/9780824865184-011>.

Di Iorio, Angelo, Silvio Peroni, and Fabio Vitali. "A Semantic Web Approach to Everyday Overlapping Markup." *Journal of the American Society for Information Science and Technology* 62, no. 9 (2011): 1696-1716. <https://doi.org/10.1002/asi.21591>.

Doerr, Martin, Pavlos Fafalios, Athina Kritsotaki, Christian-Emil Ore, and Stephen Stead. "Definition of the CRMinf: An Extension of CIDOC CRM to Support Argumentation (Version 1.2.1)." CIDOC CRM Special Interest Group, March 2026. <https://cidoc-crm.org/crminf/ModelVersion/version-1.2.1>.

Groth, Paul, Andrew Gibson, and Jan Velterop. "The Anatomy of a Nanopublication." *Information Services and Use* 30, nos. 1-2 (2010): 51-56. <https://doi.org/10.3233/ISU-2010-0613>.

ISO. "ISO/IEC 39075:2024, Information Technology - Database Languages - GQL." International Organization for Standardization, 2024. 2026년 8월 26일 확인. <https://www.iso.org/standard/76120.html>.

Keats-Rohan, K. S. B. "Prosopography and Computing: A Marriage Made in Heaven?" *History and Computing* 12, no. 1 (2000): 1-11. <https://doi.org/10.3366/hac.2000.12.1.1>.

Kellogg, Gregg, Olaf Hartig, Pierre-Antoine Champin, and Andy Seaborne. "RDF 1.2 Concepts and Abstract Data Model." W3C Candidate Recommendation Snapshot, April 7, 2026. 2026년 8월 21일 확인. <https://www.w3.org/TR/rdf12-concepts/>.

Kuczera, Andreas. "TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs." In *Graph Technologies in the Humanities - Proceedings 2020*, edited by Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera, and Joris van Zundert, 101-123. 2020.

Langefors, Börje. *Theoretical Analysis of Information Systems*. Lund: Studentlitteratur, 1966. Vogeler 2019, 318쪽에서 재인용.

Lebo, Timothy, Satya Sahoo, and Deborah McGuinness. "PROV-O: The PROV Ontology." W3C Recommendation, April 30, 2013. <https://www.w3.org/TR/prov-o/>.

Leskinen, Petri. "Modeling and Using Biographical Linked Data for Prosopographical Data Analysis." Doctoral dissertation, Aalto University, 2024.

Manola, Frank, and Eric Miller. "RDF Primer." W3C Recommendation, February 10, 2004. <https://www.w3.org/TR/2004/REC-rdf-primer-20040210/#reification>.

Massari, Arcangelo, Silvio Peroni, Francesca Tomasi, and Ivan Heibi. "Representing Provenance and Track Changes of Cultural Heritage Metadata in RDF: A Survey of Existing Approaches." *Digital Scholarship in the Humanities* 41, Supplement\_1 (2026): i196-i212. <https://doi.org/10.1093/llc/fqaf076>.

Pasin, Michele, and John Bradley. "Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach." *Digital Scholarship in the Humanities* 30, no. 1 (2015): 86-97. <https://doi.org/10.1093/llc/fqt036>.

Peroni, Silvio, and David Shotton. "The SPAR Ontologies." In *The Semantic Web - ISWC 2018*, edited by Denny Vrandečić et al., Lecture Notes in Computer Science 11137, 119-136. Cham: Springer, 2018. [https://doi.org/10.1007/978-3-030-00668-6\\_8](https://doi.org/10.1007/978-3-030-00668-6_8).

Pollin, Christopher, Franz Fischer, Patrick Sahle, Martina Scholger, and Georg Vogeler. "When It Was 2024: Generative AI in the Field of Digital Scholarly Editions." *Zeitschrift für digitale Geisteswissenschaften* 10 (2025). [https://doi.org/10.17175/2025\\_008](https://doi.org/10.17175/2025_008).

Sanderson, Robert, Paolo Ciccarese, and Benjamin Young. "Web Annotation Data Model." W3C Recommendation, February 23, 2017. 2026년 8월 22일 확인. <https://www.w3.org/TR/annotation-model/>.

Schwartz, Daniel L., Nathan P. Gibson, and Katayoun Torabi. "Modeling a Born-Digital Factoid Prosopography Using the TEI and Linked Data." *Journal of the Text Encoding Initiative*, ahead of print, March 2022. <https://doi.org/10.4000/jtei.4187>.

Seaborne, Andy, Olaf Hartig, Pierre-Antoine Champin, and Gregg Kellogg, eds. "SPARQL 1.2 Query Language." W3C Working Draft. 2026년 8월 21일 확인. <https://www.w3.org/TR/sparql12-query/>.

Shotton, David. "CiTO, the Citation Typing Ontology." *Journal of Biomedical Semantics* 1, Suppl. 1 (2010): S6. <https://doi.org/10.1186/2041-1480-1-S1-S6>.

Shotton, David, and Silvio Peroni. "CiTO, the Citation Typing Ontology." Version 2.8.2. SPAR Ontologies. 2026년 6월 22일 수정, 2026년 8월 21일 확인. <https://sparontologies.github.io/cito/current/cito.html>.

Shu, Fei, Lening Zhang, Zixuan Jin, Shibin Shu, and Vincent Larivière. "Tracing the Origins of Citation: Exploring Citation Practices in Ancient Chinese Literature." *Scientometrics* 131, no. 8 (2026): 5301-5316. <https://doi.org/10.1007/s11192-026-05717-8>.

Spadini, Elena, Francesca Tomasi, and Georg Vogeler, eds. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*. BoD, 2021.

TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.12.0, July 28, 2026. 2026년 8월 22일 확인. <https://tei-c.org/Vault/P5/4.12.0/doc/tei-p5-doc/en/html/>.

Tuominen, Jouni, Eero Hyvönen, and Petri Leskinen. "Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research." In *Proceedings of the Second Conference on Biographical Data in a Digital World 2017*, edited by Antske Fokkens et al., 2017. <https://zenodo.org/record/1040712>.

Verboven, Koenraad, Myriam Carlier, and Jan Dumolyn. "A Short Manual to the Art of Prosopography." In *Prosopography Approaches and Applications: A Handbook*, edited by Katharine S. B. Keats-Rohan. Unit for Prosopographical Research, 2007.

Vogeler, Georg. "The 'Assertive Edition': On the Consequences of Digital Methods in Scholarly Editing for Historians." *International Journal of Digital Humanities* 1, no. 2 (2019): 309-322. <https://doi.org/10.1007/s42803-019-00025-5>.

Wilson, Thomas A. "Genealogy and History in Neo-Confucian Sectarian Uses of the Confucian Past." *Modern China* 20, no. 1 (1994): 3-33. <https://doi.org/10.1177/009770049402000101>.