Research Paper https://doi.org/10.23287/KJDH.2024.1.1.4

인도 논서(śāstra) 문헌군 TEI 인코딩 전략

- 해석적 층위의 데이터를 중심으로

Encoding the Interpretative Layers of Śāstra Literature with TEI Guidelines

함형석(Ham, Hyoung Seok)*
0000-0001-8150-1445

목차

초록

본고는 인도의 문헌들 가운데 '논서'(śāstra)라 통칭되는 학술적 장르의 문헌군을 TEI 가이드라인을 준수하여 인코딩할 수 있는 방안에 대해 논한다. SARIT 등의 프로젝트에서 인도의 문헌에 대한 텍스트 인코딩 기준안을 제시하고 있지만, 이는 문헌학적인 정보에 대한 인코딩만을 중점적으로 다루고 있다. 해석적 작업을 하는 연구자들의 관점에서 문헌 디지털화의 유용성을 최대화하기 위해서는 해석적 정보를 데이터화 해야 할 필요가 있다. 본고는 참조정보(언급정보, 인용정보)와 개념어 사용정보를 데이터로 간주하고 이를 TEI 에 부합하고 SARIT 과 호환가능한 스키마로 개발할 수 있는 방안을 제안한다. 이를 통해 성립하는 데이터세트은 인도철학에 대한 연구를 보다 견고한 역사적 기반 위에 올려놓을 것이다.

주제어: 인도철학, 논서, TEI, 텍스트 인코딩, 상호텍스트성

* 전남대학교 철학과 부교수, hamhs@chonnam.ac.kr

Abstract

This paper discusses how the scholarly genre of Indian literature known as śāstra can be encoded according to TEI guidelines. While projects such as SARIT have proposed text encoding standards for Indian literature, they mainly focus on encoding philological information. To better serve the need of interpretatively oriented researchers, interpretive information is to also be defined as data. This paper proposes to consider reference information (mentions, quotations) and concept-word usage information as data and to develop a schema that is TEI-conformant and SARIT-compatible. The resulting dataset will place the study of Indian philosophy on a more solid historical footing.

Keywords: Indian Philosophy, śāstra, TEI, Text Encoding, intertextuality

1. '샤스뜨라'(śāstra, 논서)라는 비-역사적 문헌군

인도의 전통적 지식인들의 학술활동은 '정제된'(saṃskṛta) 언어인 산스크리트를 통해 이루어졌고, 그들의 학술적 글쓰기 스타일은 '논서'(śāstra)라 불리는 장르로 규범화되어있다. 인도철학의 여러 학파들은 그들의 사유체계가 극단적으로 다른 경우에도 '논서'라는 공통된 형식을 통해 그들의 세계관을 개진하였다.[1] 하지만 그들이 공유하였던 것은 글의 구조나 형식뿐만이 아니었다. 그들은 공통적으로 '비역사적' 혹은 '탈역사적'으로 세계를 인식하였고, 이와 같은 멘탈리티는 그들의 글쓰기에도 고스란히 반영되어 서구의 오리엔탈리스트들을 비롯한 외부인들에게 '시간이 멈춘 인도'(timeless India)라는 인상을 남겨주었다.[2]

인도철학을 연구함에 있어 연구자가 맞닥뜨리게 되는 가장 큰 난관 가운데 하나도 바로 이러한 논서 문헌군의 비역사적 성격에 기인한다. 한 저자의 사유체계를 온당하게 분석해내어 평가하려면 그의 사상을 '인도철학사'의 맥락 속에서 자리매김할 수 있어야 하는데, 인도의 논서에는 그것을 어떠한 외부의 참조점에 연결시킬 수 있는 단서가 불분명하게 제시되어 있다. 모든 문헌이 분명하게 확정가능한 역사적 정보를 결여한 상황 속에서, 개별 문헌들이 집합적으로 구성하는 인도철학사는 주요 철학자들 혹은 철학서들의 상대적 연대(relative chronology)를 중심으로 쓰이거나, '육파철학'이라는 단어로 상징되는 탈역사적 세계인식을 바탕으로 한 독소그래피(doxography)의 형태로 쓰이고 있다.[3]

인도의 문헌들이 역사적 정보를 결여하고 있다는 사실이 단순히 한 문헌 속에 그것의 작성연대나 저자의 활동시기를 추측할 수 있는 단서가 발견되지 않는다는 것만을 의미하는 것은 아니다. 논서의 저자들은 다른 저자들이나 논서를 언급하거나 인용함에 있어 매우 불명확한 참조정보만을 제시한다. 자신과 의견을 달리하는 학자를 이름으로 특정하기보다는 '어떤 이들'(eke)이라 지칭하거나, 다른 이들의 견해를 자신이 이해한 방식대로 풀어쓰며, 인용문을 제시하는 경우에도 서명(書名)으로 그 전거를 확인해주는 경우가 매우 드물다. 이러한 상황 때문에 한 문헌이 산출된 역사적 맥락은 현대의 인도철학 연구자들의 주요 연구대상이다.

연구자들은 '어떤 이들'과 같은 불특정 지시사의 지시대상과 인용문의 전거를 확정하여 한 문헌이 처해 있던 지적배경을 재구성하며, 개념어의 출현과 그것이 문헌 속에서 사용되는 방식을 평가하여 문헌의 연대를 '상대적으로' 측정한다.[4] 하지만 한 문헌 전반에 대한 역사적 평가가 종합적으로 이루어지기 위해서는 언급정보와 인용정보, 그리고 개념어 사용방식 등에 대한 개별연구자들의 파편화된 연구들이 체계적으로 통합되어야 할 필요가 있다.

한 문헌에 대한 학계의 독법을 종합화하는 데에는 여러 방안이 있을 수 있다. 본고는 인도학

학계가 문헌 연구를 위해 조금씩 그 가능성을 타진해보는 상황에 있다고 판단되는 TEI(Text Encoding Initiative) 가이드라인에 따른 텍스트 인코딩 방안을 위의 문제의식을 바탕으로 구체화해 보려고 한다.[5] 이를 위해 우선 현재 인도문헌 데이터베이스들의 TEI 인코딩 상황을 개관 및 평가한다. 이후 문헌의 역사성을 재구성하는 데 쓰일 수 있는 내부 정보들을 기록하기 위한 항목을 TEI 가이드라인이 제시하고 있는 요소(element/tag)와 속성(attribute)들 가운데에서 선별하고 그것들의 표준적인 사용방안을 동료학자들에게 건의하고자 한다.

2. TEI와 인도 문헌의 디지털화: '기계적 처리가 가능한'(machine-readable) 산스크리트 문헌 만들기

인도의 문헌을 연구하는 학자들이 가장 많이 찾는 디지털 문헌 저장소는 단연 GRETIL(Göttingen Register of Electronic Texts in Indian Languages)이라고 할 수 있다. GRETIL은 본래 txt, html형식의 단순 텍스트 파일만을 제공하였지만, 2019년도부터 점차적으로 TEI 가이드라인을 준수하여 인코딩한 XML파일도 함께 등재하고 있다. GRETIL의 소개글(Introduction)을 참조하면 이는 보다 '다양한 목적에 쓰일 수 있는'(versatile) 그리고 '명료한'(transparent) 텍스트 자료를 세계 각국의 인도학자들에게 제공하기 위함이다.[6] 매우 소극적인 의미에서 '기계적 처리가 가능한', 다시 말해, 한 문헌의 내부 뿐만 아니라 문헌군 전체에 대한 검색을 가능하게 하기 위한 코퍼스를 구축한다는 GRETIL 자체의 목표를 감안한다면 TEI의 도입은 다음과 같은 의미를 가지고 있는 것이 아닐까 생각한다.

그간 GRETIL의 자료는 개별 연구자들의 자발적인 타이핑 노동에 입각하여 유지되어 왔으며, 이 때문에 GRETIL에 등록되어 있는 각 텍스트 파일은 하나의 통일된 인코딩 양식을 공유하지 않는다. 극단적인 경우에는 데바나가리(Devanāgarī)와 같은 인도활자의 로마자 전사(transliteration)방법을 달리하는 경우도 있다. 그러나 가장 흔히 발견되는 양식상의 불일치는 비문자 사용방식의 불일치이다. GRETIL에 자신의 파일을 제공한 학자들은 문헌 속에 내재하는 근본 텍스트와 여러 층위의 주석서들을 구분해서 표시하기 위해, 그리고 비판교정본(critical edition)을 만든 편집자가 삽입한 문구 혹은 주석 등을 표기하기 위해 파일 속에서 줄바꿈문자, 공백문자, 탭문자, 하이픈(-), 별표(*), 각종 괄호들([], {}. ()), 콜론(:)과 세미콜론(;) 등의 비문자들을 사용한다. 하지만 학자들마다 이들 기호에 부여하는 의미가 달라 TEI 방식의 인코딩을 채택하지 않은 기존의 단순 텍스트 파일의 경우, 복수의 파일들에 대한 일관된 기계적 처리는 불가능하며 사용자가 하나의 파일 속에 발견되는 기호의 용법을 명확하게 파악해내기도 어려운 경우가 많다.

GRETIL의 전격적인 TEI 도입은 이와 같은 난점을 해소하려는 노력으로 보인다. TEI 인코딩은 학자들로 하여금 공통의 언어로 자신의 의도를 기록하게 함으로써 텍스트 파일을 보다 '명료하게' 만들 수 있을 것이다. 또한 그같은 인코딩의 노력이 지속될 경우 GRETIL을 임의의 텍스트 검색만이 가능한 '저장소'(repository)에서 인명이나 지명, 혹은 인용문 등에 대한 체계적인 검색이 가능한 '데이터베이스'(database)로 탈바꿈 시킬 수 있는 가능성도 존재한다. 현재는 문헌의 층위와 문단의 경계 정도만이 인코딩되어 있는 상황인데(그림1 참조), 이 정도의 인코딩만으로도 맥락 없이 텍스트를 무작위로 절단하여 텍스트 간의 유사도를 측정하는 식의 무분별한 연구행위를 보정할 수 있는 역할은 수행할 수 있다고 생각한다. 이와 같은 의미에서 TEI의 도입은 GRETIL이 제공하는 텍스트들이 '다양한 용도로 쓰일 수 있는' 가능성을 열고 있다.

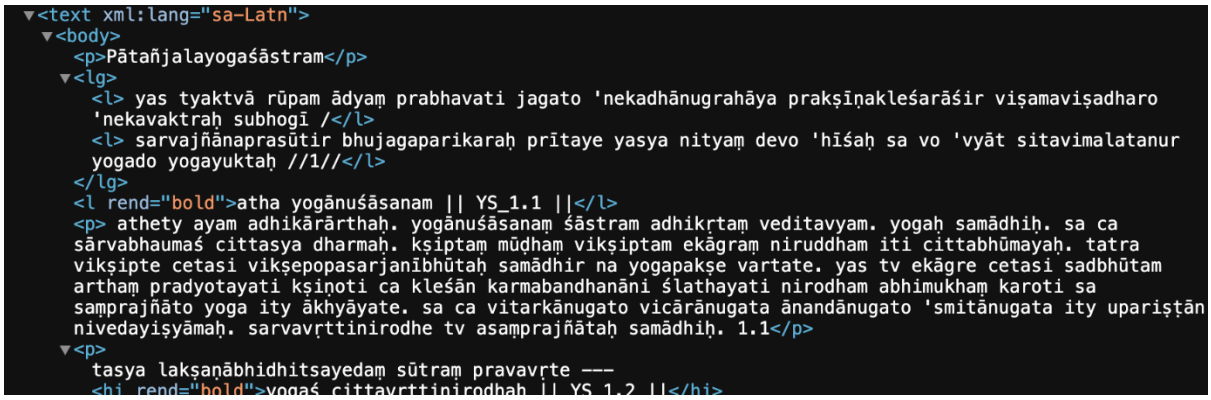

그림 2 GRETIL 의 <빠딴잘리 요가서>(Pātañjalayogaśāstra) 인코딩 현황 [7]

SARIT(Search and Retrieval of Indic Texts)은 본격적으로 TEI기반 인도문헌 데이터베이스 구축을 표방하고 있는 프로젝트이다. 아직은 60종의 문헌만을 제공하고 있으며 GRETIL의 파일과 비교해볼 때 인코딩의 깊이 혹은 수준이 비슷하다. 동일한 문헌(<빠딴잘리 요가서>)에 대한 인코딩을 비교해보면, SARIT의 파일(그림2 참조)은 굵은 글씨로 출력되어 있는 요가쑤뜨라의 텍스트를 GRETIL과 같이 '굵은 글씨'(@rend="bold")라고 인코딩하지 않고, '쑤뜨라'(@type="sutra")라고 인코딩하는 등 텍스트의 '표현'(representation)이 아닌 그것에 담긴 '의미'(meaning) 혹은 '의도'(intention)를 인코딩하기를 요구하는 TEI의 가이드라인에 담긴 정신에 보다 부합하는 측면이 없지 않다. 그럼에도 여전히 인코딩되어 있는 정보는 챕터(pāda)의 구분, 근본 텍스트(Yogasūtra)와 주석서(Yogabhāṣya)의 구분, 그리고 주석서의 문단 구분 정도에 그치고 있다.

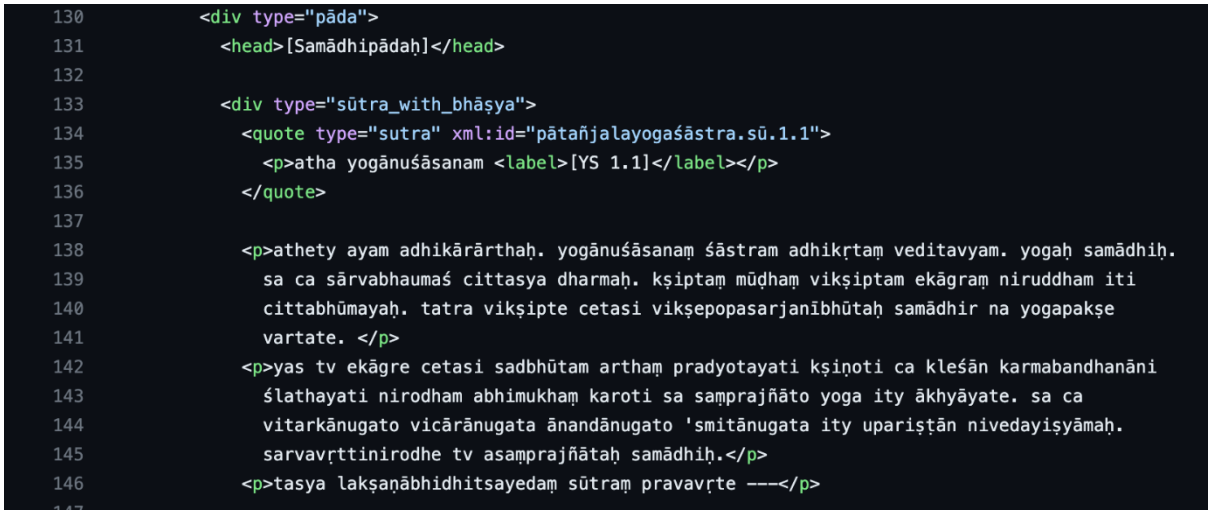

그림 3 SARIT 의 <빠딴잘리 요가서>(Pātañjalayogaśāstra) 인코딩 현황 [8]

SARIT 프로젝트에서 주목해야할 지점은 그것이 현재 제공하고 있는 문헌의 종류, 혹은 각 문헌에 대한 인코딩의 수준과 품질이 아니라 그것이 독자적인 가이드라인을 제시하고 있다는 사실

이다. SARIT은 이를 통해 TEI 컨소시엄의 가이드라인 속에서 인도의 문헌을 읽는 학자들에게 소용이 있는 요소(element)와 속성(attribute)을 변별하고, 그것들에 대한 범용적인 정의를 인도문헌의 상황에 맞게 재해석하여 인도문헌 디지털화의 표준적인 문법과 어휘를 제시한다. SARIT이 전제하고 있는 인도문헌 속에 내재한 보편적 구조와 구조적 특징들에 대한 해석방식, 그것을 기록하기 위해 제안하는 요소와 속성의 종류, 그리고 그것들의 사용방식 등에 대해 이견이 있을 수 있겠지만, SARIT의 가이드라인은 문헌중심의 인도학자들이 디지털 연구환경 속에서 디지털적 방법론을 통해 연구행위를 수행할 수 있는 초석을 마련하였다는 점에서 상당한 의의를 지닌다고 할 수 있다.

SARIT은 가이드라인을 '단순'(simple) 버전[9]과 '완전'(full) 버전[10]으로 나누어 제시한다. 단순 버전의 가이드라인[11]은 '주석서'(commentary)라는 산스크리트 논서 문헌군이 취하는 보편적인 형식의 구조적인 정보를 어떻게 인코딩할 것인지에 초점을 맞추어 작성되었다. 이에 반해 완전 버전의 가이드라인[12]은 인코더가 기존에 출판된 비판교정본을 인코딩의 저본으로 삼을 경우를 상정한다. 이는 TEI 가이드라인의 '문헌비평' 모듈(12. Critical Apparatus)[13]을 참조하고 있으며 판본 내부에 편집자가 괄호 및 각주 등의 형식적 요소로 표기하고 있는 필사본의 이독(異讀, variant reading) 등 문헌비평을 위한 자료(critical apparatus)를 인코딩하는 방안에 대해 논한다. 종합하자면, 두 종류의 SARIT 가이드라인이 기록하고자 하는 정보들은 문헌의 기초적인 구조정보와 문헌의 근저에서 문헌을 구성하는 필사본의 독법들 그리고 그에 대한 현대 학자들의 판단들이라고 할 수 있다.

SARIT의 가이드라인은 문헌학자의 시선에서 작성된 문건이다. 문헌학적 정보에 대한 인코딩 규칙만을 제공하고 있기 때문에 SARIT 가이드라인을 좇아 인코딩된 파일은 결국 문헌학적인 용도로 사용될 수밖에 없다. 이러한 가이드라인은, 바르뜨르하리(Bhartṛhari, 5세기)의 <문장과 단어에 관한 책>(Vākyapadīya)에 대한 Charles Li의 학위논문 프로젝트[14]가 모범적으로 보여주었듯, 자신의 비판교정본과 함께 필사본들 그리고 기존의 판본들에 기록되어 있는 서로 다른 독법들을 선택적으로 열람할 수 있는 플랫폼을 만드는 데 쓰일 수 있을 것이다. 또한, 이를 분석적으로 사용한다면 이독 정보들을 활용하여 필사본의 계통도를 작성하고 확정하는 용도로, 더 나아가 기존 판본들을 출간한 편집자들의 암묵적인 편집원칙 등을 밝히는 용도로 쓰일 수 있을 것이다. 하지만 SARIT 가이드라인은 필자와 같은 인도문헌에 대한 해석적 작업을 하는 연구자들, 필사본 등의 문서의 역사가 아니라 문헌에 담긴 생각들의 역사를 추적하고 당대의 지적 네트워크 속에서 문헌의 위상을 평가하고자 하는 연구자들이 TEI 인코딩에 기대하는 지평을 열어주지 못한다. 다른 목적과 다른 정체성을 가진 연구자들은 SARIT과는 다른 방식으로 TEI를 맞춤제작(customization)하여 또 다른 인코딩 스키마를 정의할 필요가 있다.

3. 인도철학의 역사적 이해를 위한 데이터

글의 서두에서 언급하였듯 인도의 문헌을 연구함에 있어 가장 큰 난점 가운데 하나는 문헌의 작성연대를 알 수 없을 뿐만 아니라 인도 지식인들의 비-시간적인 세계인식으로 인해 문헌 내부에도 문헌이 작성된 배경을 알 수 있는 정보가 명시적으로 드러나 있지 않다는 사실이다. 대부분의 인도 문헌의 작성연대가 다른 문헌들과의 비교를 통해 상대적으로 상정된다는 사실을 고려할 때, 하나의 문헌에 역사성을 부여하는 일은 그것이 다른 문헌들과 맺고 있는, 혹은 저자가 선대의 그리고 당대의 다른 저자들과 맺고 있는 관계를 파악해내는 작업이다. 그리고 이와 같은 작업은

연구자들이 한 문헌의 내용을 다른 문헌의 그것과 비교, 분석하여 문헌 속에 암시되어 있거나 불명확하게 제시되어 있는 저자의 참조행위를 해석하고 이를 통해 문헌 외부의 참조대상을 확정하는 것으로 이루어진다.

그런데 이와 같이 한 문헌이 외부 객체들과 맺고 있는 관계를 탐색하는 과정 속에서 명확해지는 것은 그것의 개략적인 작성연대뿐만이 아니다. 인도철학사 속에서 한 문헌의 좌표를 시공간적으로 식별하는 작업은 저자의 지적 배경을 비롯하여 그의 문제 의식과 시대 인식, 더 나아가 저자가 담론을 구성하고 어휘를 사용하는 방식을 이해하려는 노력이다. 하나의 문헌이 지니고 있는 관계성에 대한 고찰은 궁극적으로 문헌이 작성되던 당시의 인도지성계를 복원하는 것으로 나아가며, 인도지성계 속 저자의 위치 파악은 해당 문헌을 구성하는 개별적인 문장과 단어의 의미를 밝히는 것으로 나아가는 해석학적 순환이 일어나는 것이다. 이에 따라 한 문헌이 담지한 역사성에 대한 탐구는 미시적으로 해당 문헌에 담긴 철학적 사유를 이해하고 거시적으로 인도철학의 역사를 재구성하는 데 있어 기본적인 그리고 핵심적인 연구활동이다.

한 문헌의 관계성 파악을 위한 가장 기초적인 텍스트 내부의 정보는 다른 문헌과 인물 등의 외부 객체에 대한 참조정보라 할 수 있는 언급정보와 인용정보이다. 문헌의 저자가 다른 이들의 저작을 거론하고 인용할 때 독자는 저자가 서 있는 지평을 구성하는 지적 주체(agent)들을 읽어낼 수 있는 실마리를 확보할 수 있다. 이와 같이 텍스트 내에서 언급이나 인용의 행위가 일어날 때 얻을 수 있는 또 다른 정보는 저자가 자신이 참조하는 대상에 대해 취하는 태도정보이다. 저자가 외부 객체를 참조할 때 드러내는 참조대상에 대한 태도는 경우에 따라 상당한 차이를 보이며, 이를 언급・인용정보와 결합할 때 연구자는 한 문헌의 저자가 염두에 두고 있는 인도지성계 속의 전선(戰線)을 종횡으로 그려낼 수 있다. 예를 들어, 저자가 그 권위를 인정하는 인물과 문헌들을 모아볼 때 우리는 종적으로 저자가 추종하는 지식인들의 지적 계보를 구성할 수 있고, 역사적 인물들과 문헌들이 한 전통 내부에서 정전화(正典化, canonization)되어 가는 과정을 추적해볼 수 있다. 이에 반해 저자가 다른 이들의 의견을 호의적으로 평가하거나 비판하기 위해 참조하는 사례들은 동시대에 횡적으로 존재하는 지적공동체들을 식별해내고 그들 사이의 관계를 변별하는 데 쓰일 수 있다.

언급정보나 인용정보만큼 직접적이고 분명하게 한 문헌과 다른 문헌들 간의 상호텍스트성(intertextuality)을 드러내는 정보는 아니지만, 문헌의 역사성을 파악하는데 주요한 단서 역할을 하는 또 다른 요소는 저자가 사용하는 개념어와 그에 대한 저자의 이해방식이다. 시대에 따라, 특히 영향력 있는 사상적 사조가 등장할 때마다, 인도철학계를 풍미한 논제들과 그것을 둘러싼 사유를 응축하여 담아내는 새로운 개념들이 탄생하였고, 이에 따라 기존의 개념어들이 이해되고 서술되는 방식 역시 변화하였다. 저자가 구사하는 어휘는 저작의 나이정보를 품고 있다. 한 문헌 속의 키워드들, 그리고 각 키워드에 대한 저자의 해설방식을 모아보는 행위는 문헌이 속한 전통의 발전양상을 드러내고 당대 인도철학계의 시대정신을 재구성하는 작업이다.

SARIT 가이드라인은 문헌학적 정보들, 즉 필사본과 판본들의 독법, 그리고 교정 대상 텍스트에 대한 다른 문헌들의 인용 등 비판교정본 본문을 성립시키는 데 사용된 증거자료들을 데이터로 삼고 있다. 다시 말해, 이는 물리적인 매체에 실제로 적혀 있는 텍스트 자체를 인코딩의 대상, 즉 데이터로 여기고 있다. 이에 반해 인도철학 문헌의 역사적 이해를 위해 필자가 주목하는 언급정보, 인용정보, 개념어 사용정보는 한 문헌의 판본이라는 물리적 실체 위에 덧씌워진 해석적 층위(layer)에 담겨있는 데이터들이다. 이 해석적 층위는 문헌이라는 물리적 사물과 연구자들의 훈련된 의식이 만났을 때에만 드러나기 때문에 물리적인 기반 위에 성립하는 추상적 층위라 할 수 있다.

한 문헌의 존재를 증언하는 필사본들의 역사가 SARIT 가이드라인이 인코딩하는 데이터를 규정한다면, 한 문헌을 이해하고자 노력해온 연구자들의 역사가 이 해석적 층위의 데이터를 규정한다. 그렇기 때문에 이 레이어를 기계적 처리가 가능하도록 인코딩할 때 필요한 자원은 사본과 같은 물리적인 자원이 아니다. 필요한 것은 연구사에 대한 이해를 갖춘 인코더의 의식이며, 파일 속에 새겨 넣는 데이터는 선배, 동료, 후배 연구자들의 그리고 그들을 종합하는 자신의 문헌에 대한 이해이다.

4. TEI 와의 부합성을 갖춘 TEI 커스터마이제이션

TEI 가이드라인은 텍스트의 특징들을 기록하기 위한 범용적인 요소(element)와 속성(attribute)들을 제안한다. TEI 사용자들은 TEI 컨소시엄이 가이드라인에 정의해 놓은 700여개의 요소(태그)들을 모두 사용하거나 고려할 필요가 없는데, 이는 TEI 가이드라인이 인코더가 텍스트에서 발견할 수 있는 모든 특징들을 망라하려 하기 때문이다. 인코더는 자신의 목적에 맞게 자신이 사용할 요소들을 선별하고 그것이 취할 수 있는 속성과 속성값을 정의해야 하는데 이러한 행위를 'TEI 커스터마이제이션'(customization)이라 부른다.

필자의 이해에 따르면, TEI 커스터마이제이션은 어휘적, 구문론적, 의미론적 차원이라는 세 가지 차원에서 이루어져야 하는 행위이다.[15] 그리고 TEI 가이드라인에 기반하여 본인의 프로젝트의 필요에 맞추어 제작된 스키마를 도출하였을 때, 그 스키마를 TEI에 '부합한다'(conformant)고 혹은 TEI에 대한 '부합성'(conformance)를 가지고 있다고 이야기한다. TEI에 부합하는 스키마는 세 가지 차원, 즉 어휘적, 구문론적, 의미론적 측면에서 TEI 가이드라인 전체의 부분집합을 구성하거나 그것과 일정 부분을, 즉 교집합을 공유한다.

Figure 1. Varieties of Customization.

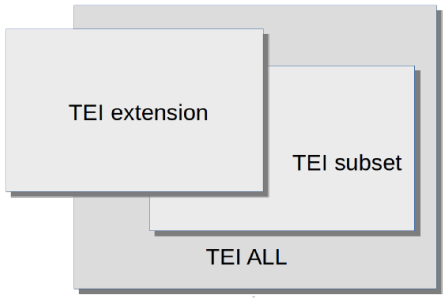

그림 4 TEI 의 부분집합을 구성하는 혹은 교집합을 갖는 TEI 커스터마이제이션 [16]

어휘적 차원의 커스터마이제이션은 한 프로젝트에서 인코딩에 사용할 요소와 속성을 TEI 가이드라인이 정의해 놓은 것들로 한정하거나 TEI 가이드라인의 요소와 속성을 사용하면서 그것이 인코딩을 정의하고 있지 않은, 하지만 개별 프로젝트에서는 기록이 필요한 텍스트적 특징(textual feature)을 인코딩하기 위해 새로운 요소와 속성을 정의하는 행위를 의미한다. 어휘적 차원의 커스터마이제이션에서 유의할 지점은 TEI 가이드라인에 이미 정의되어 있는 텍스트적 특징 또는 현상을 새로운 요소로 재정의하지 않아야 한다는 것이다. 불필요하게 TEI 가이드라인을 확장하는 행위는 다른 연구자들에게 혼동을 야기할 수 있을 뿐만 아니라, 이후 프로젝트 진행시 데이터 분석

을 위해 새롭게 소프트웨어를 개발해야하는 등의 자원의 낭비를 초래할 수 있기 때문이다.[17]

구문론적 차원의 커스터마이제이션은 개별 프로젝트에서 사용하는 TEI 요소와 속성들의 '모델 클래스'(model class)와 '속성 클래스'(attribute class)를 준수하는 소극적인 행위이다.[18] 주지하다시피, TEI 더 나아가 XML은 텍스트를 'OHCO'(Ordered Hierarchy of Content Objects)로, 즉 내용을 담은 객체들이 위계적으로 배치된 사물로 간주한다.[19] TEI는 TEI 문서를 구성하는 부분들을 모델 클래스로 개념화하고 모델 클래스들 간의 위계질서를 정의하고 있다. TEI의 각 요소들은 하나 이상의 모델 클래스에 소속되어 있는데, 모델 클래스 간의 위계질서가 정의되어 있기 때문에, TEI의 요소들은 모델 클래스에 소속되는 것을 통해 간접적으로 그것이 출현할 수 있는 맥락을 제약받는다. 예를 들어, 문장을 인코딩하는 요소(<s>) 안에는 문단을 인코딩하는 요소(<p>)가 출현할 수 없으며 이와 같은 질서는 각 요소가 속한 모델 클래스의 차원에서 정의된다. 속성 클래스는 같은 속성을 공유하는 요소들을 하나의 클래스로 개념화한 것이다. 즉, TEI의 각 요소들이 사용할 수 있는 속성의 종류를 해당 요소를 속성 클래스의 구성원(member)로 규정하는 것을 통해 정의하고 있는 것이다.

개별 프로젝트에서 모델 클래스와 속성 클래스라는 개념으로 표현되는 TEI의 요소와 속성들 간의 질서를 준수하기 위해 적극적으로 해야할 특별한 일이 있는 것은 아니다. 하지만 개별 프로젝트에서 TEI가 규정한 요소들의 위계질서를 무시하는 방식의 인코딩을 허용한다거나, TEI가 특정 요소에 지정하지 않은 속성을 사용할 수 있게 하는 스키마를 작성한다면, 해당 스키마는 더이상 TEI에 부합하지 않게 된다. 즉, 그것은 TEI 가이드라인의 관점에서 '유효한'(valid) 문서가 아니게 되는 것이다. 이와 같은 스키마에 따른 인코딩 결과물은 내부적으로는 문제 없이 사용할 수 있지만 다른 프로젝트의 데이터와 결합하기 힘들어지며, TEI 관련 범용 소프트웨어를 사용하지 못하게 되거나 사용할 경우 데이터 손실을 겪게 된다.

의미론적 차원의 커스터마이제이션은 사용하기로 결정한 TEI의 요소와 속성의 의미를 개별 프로젝트의 목적에 맞게 재정의하는 행위이다. TEI의 요소와 속성들은 일반적인 차원에서 그 의미가 정의되어 있어 특정 프로젝트가 다루는 문헌군을 대상으로 하는 지식체계 속에서 보다 엄밀하게 재정의되어야 할 필요가 있다. 위에서 살펴본 SARIT 프로젝트가 제시하는 두 가지 버전의 가이드라인은 의미론적 커스터마이제이션 행위라고 할 수 있고, TEI 가이드라인 문서 자체도 일종의 의미론적 커스터마이제이션으로 생각할 수 있다. 의미론적 커스터마이제이션은, SARIT 가이드라인의 경우와 같이, 인코딩에 사용하는 어휘들의 의미에 대해 설명하고 그 사용에 대한 예시를 드는 문서화(documentation) 작업이다. 이 때 유의할 점은 인코딩 어휘에 대한 재정의가 TEI의 기본적인 정의에 기반하여 이루어져야 한다는 것이다. TEI 가이드라인의 정의를 무시한 재정의는 사용자에게 혼란을 초래하며 다른 프로젝트가 생성한 데이터와의 결합을 불가능하게 만들 수 있다.

TEI 컨소시엄은 이 같은 세 가지 차원에서의 커스터마이제이션을 손쉽게 할 수 있는 도구인 ROMA(https://roma.tei-c.org/)라 불리는 웹서비스를 제공한다.[20]

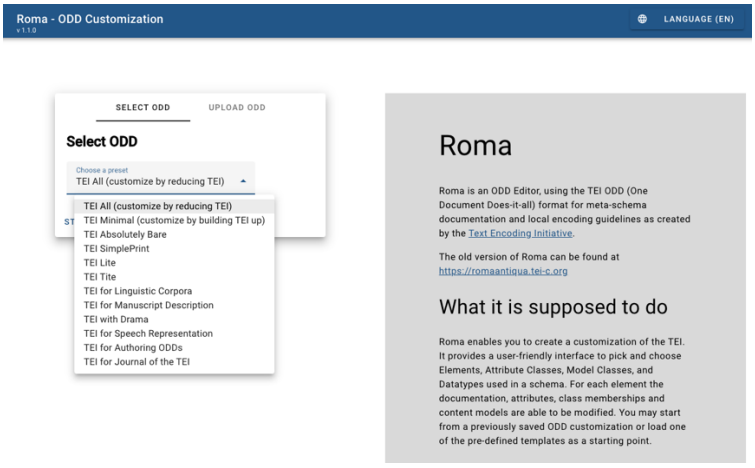

그림 5 TEI 커스터마이제이션 도구인 ROMA 의 첫 페이지

ROMA에서 사용자는 TEI의 모든 요소와 속성이 포함된 TEI All에서 불필요한 요소들을 제거하거나 최소한의 요소와 속성이 포함된 TEI Minimal에 필요한 요소들을 추가하는 것을 통해, 혹은 TEI Lite와 같이 자주 사용되는 기정의된 템플릿을 변형하여 자신의 프로젝트를 위한 인코딩 규칙을 생성할 수 있다. 이후 페이지에서 개별 프로젝트에서 사용할 TEI 요소를 선택하고 각 요소들의 속성을 선별하는 과정을 통해 사용자는 어휘적 차원의 커스터마이제이션을 수행할 수 있다. 또한, 각 요소들과 각 요소들과 함께 사용될 수 있는 개별 속성들을 선택하면 그에 대한 문서화 역시 행할 수 있어 의미론적 차원의 커스터마이제이션 역시 ROMA를 통해 수행할 수 있다. 자신이 사용할 요소와 속성을 선택하고 그들에 대한 설명까지 기술하면 TEI 커스터마이제이션이 완료되는데, ROMA는 이 모든 정보를 ODD(One Document Does-it-all)라는 독자적인 파일 형식으로 또는 범용적인 XML 에디터에 연동하여 사용할 수 있는 DTD(Document Type Definition)과 같은 스키마 작성언어로 쓰인 스키마를 다운받을 수 있게 해준다.

5. 해석적 층위의 데이터를 인코딩하기 위한 TEI 요소와 속성들

필자는 3절(인도철학의 역사적 이해를 위한 데이터)에서 기술한 언급, 인용, 개념어 사용 데이터를 인코딩하기 위해 TEI 가이드라인을 커스터마이즈 할 필요가 있었고, 이에 4절에서 언급한 ROMA를 통해 필자의 인코딩 규칙을 스키마로 만든 후 oXygen XML Editor[21]에 연동하여 인코딩 작업을 수행하고 있다. 본 절에서는 3절에서 변별해낸 해석적 층위의 데이터를 인코딩하기 위해 필자가 선택한 TEI 요소와 속성을 소개하고 각각에 대한 인코딩 사례를 제시하고자 한다.[22] 필자의 커스터마이제이션을 담은 ODD에는 아직 문서화 작업이 되어 있지 않지만, 본 절에서 기술하는 내용을 토대로 추후 정식적인 문서화 작업을 진행하려한다.

TEI의 요소와 속성을 선택함에 있어 필자는 다음의 원칙을 따랐다.

1. 선택한 요소와 속성에 대한 TEI의 일반적인 정의가 인도 논서의 구체적인 상황에 적용될 수 있어야 한다.
2. SARIT이 사용을 규정한 요소는 사용하지 않는다.
3. 하나의 요소는 하나의 특징을 인코딩하는 데에만 쓰여야 한다.

원칙 1은 필자가 선택한 TEI 요소와 속성에 대한 이해를 TEI 가이드라인에 부합시키기 위한 규정이다. 원칙 2는 SARIT의 인코딩과 필자의 인코딩이 하나의 파일 내에서 양립할 수 있도록

하기 위한 규정이다. 원칙 3은 인코딩시, 또는 인코딩된 파일을 읽을 때 발생할 수 있는 의미상의 애매모호함을 배제하기 위한 규정이다. 인도 논서 문헌의 해석적 레이어를 인코딩하는 본 절의 규칙들은 TEI 가이드라인의 부분집합이며 SARIT 가이드라인과 교집합을 갖지 않는 요소와 속성의 집합이다. 따라서 한 문헌의 해석적 레이어는 이를 통해 독자적으로 혹은 SARIT 가이드라인이 규정하는 문헌학적 레이어와 함께 인코딩되어 유효한 TEI 문서를 만들어 낼 수 있을 것이다.

이하의 서술에서는 필자가 선택한 요소와 속성을 소개하기 위한 사례로 즈냐나가르바(Jñānagarbha, 8세기)의 <두 가지 진리를 구분하기>(Satyadvayavibhaṅga)에 대한 인코딩 작업을 사용할 것이다. 본 저작은 산스크리트 원문이 유실되었고 현재는 티벳어 번역(bDen pa gnyis rnam par 'byed pa)으로만 접근 가능한 문헌이다.[23] 산스크리트로 되어 있지는 않지만 표준적인 인도 논서의 형식을 따르고 있어 인도 논서 인코딩의 사례로 삼는 것에 무리가 없다고 생각한다.

5.1 발화주체의 인코딩, <said>

인도철학은 그것의 태동기인 베다시기부터 질문과 답변의 형태로 표현되어 왔으며, 본격적인 철학서 장르인 논서 문헌 역시 수많은 논적들의 비판과 그에 대한 저자의 답변으로 구성되어 있다. 즉, 논서는 대화체로 구성되어 있다고 해도 과언이 아니며, 전통적으로 논적의 주장을 담은 부분의 텍스트는 '앞선 주장'(pūrvapakṣa), 이에 대한 저자의 대답은 '뒤따르는 주장'(uttarapakṣa), 그리고 논쟁상황의 종료를 알리는 영역은 '확정된 결론'(siddhānta)이라 불린다. 이러한 논서에 내재한 질문과 답변이라는 구조적 형식은 해석적 데이터를 인코딩하기 전에 기록되어 있어야 하는데, 이는 해석적 데이터의 의미가 발화주체에 따라 그 의미를 달리하기 때문이다. 논적이 발화하는 부분에 담겨있는 개념어들은 문헌을 작성한 저자의 사상을 대변하지 않으며, 해당 부분의 인용문 역시 저자가 그 권위를 인정하지 않는 문헌에서 비롯된 것일 수 있다. '대화'라는 논서의 근원적인 형식을 인코딩하기 위해 필자는 모든 TEI 문서가 공유하고 있는 모듈인 core모듈에 속한 요소인 <said>를 선택하였다.[24]

```
<div xml:id="SDVVr6" n="6">
   <said ana="pūrvapakṣa" who="Yogācāra">
      <quote type="base-text">
         <lg xml:id="SDV6ab" n="k6ab" next="k6c">
            <l>gal te <term key="parikalpita-svabhāva">brtags pa'i ngo bo nyid</term>/</l>
            <l>'ga' yang mthong ba med ce na/</l>
         </lg>
      </quote>
      <pb ed="#Akahane2020" n="6"/>
      <milestone ed="#Akahane2020" unit="section" n="k6ab-1"/>
      <p><term key="parikalpita-svabhāva">brtags pa'i ngo bo nyid</term> kyi dbang du mdzad
         nas <q ana="authoritative" source="Dharmasaṃgītisūtra" type="explicit">'ga' yang
            mthong ba med pa ni de kho na mthong ba</q> zhes bya ba 'di gsungs par 'dod de,
         de ni de lta bu yin no. de lta bu ma yin na 'ga' yang zhes ci'i phyir gsungs te?
            <term key="svabhāva">ngo bo nyid</term> mi mthong ba zhes bya ba kho nar gsung
         bar 'gyur ba'i rigs so zhe na.</p>
   </said>
```

그림 6 논서의 대화형식을 기록하기 위한 <said>인코딩의 사례

<said>요소는 @ana(analysis)라는 필수적인 속성을 갖고 이는 "pūrvapakṣa", "uttarapakṣa", "siddhānta"라는 세 가지 값 중 하나를 취할 수 있다. '앞선주장'의 경우, 논적의 정체가 분명할

때 @who요소에 그것을 기록할 수 있으며, @resp(responsibility)에 논적의 정체를 밝힌 참조자료나 연구물의 @xml:id값을 기록할 수 있다.

5.2 언급정보의 인코딩, <rs> (referencing string)

한 문헌이 인물이나 다른 문헌을 언급할 때 이를 인코딩하는 요소로 <rs>를 선택하였다. TEI는 인물과 서적을 인코딩하기 위한 다양한 요소들을 구비해 놓았는데, 인물의 경우는 <persName>이, 서적의 경우에는 <title>이 대표적이라 할 수 있다. 하지만 인도문헌에서 인물이나 서적의 이름을 직접적으로 거명하는 경우는 극히 드물다. 대부분의 언급정보가 대명사나 일반명사로 표현되어 있어 내용적인 정보에 근거해 지시대상을 유추해야 한다. 이에 각 암시적 표현들의 지시체를 규명하는 것은 독립적인 연구대상이 되기 일쑤이다. 이와 같은 상황을 고려하여 일반적인 '참조문자열'을 의미하는 <rs>(referencing string)을 통해 언급정보를 인코딩하기로 결정하였다.[25]

```
<p><rs type="person" key="Nirākāravādin" ana="affirmative">shes pa rnam pa med par
      smra ba</rs>s de <choice>
      <sic>sked</sic>
      <corr>skad</corr>
   </choice> brgal ba la, <rs type="person" key="Dignāga Dharmakīrti" ana="critical"
      >mtshan nyid dang rnam 'grel byed pa</rs>'am <rs type="person" key="Yogācāra"
      ana="critical">rung gang</rs> gis yul gyi rnam pa nyid kyi tshad ma bsgrub pa
   ji skad brjod par bya.</p>
```

그림 7 언급정보를 기록하기 위한 <rs>인코딩의 사례

<rs>는 @type, @key, @ana의 세 가지 필수 속성을 가지며, @resp와 @cert(certainty)의 두 가지 선택 속성을 지닌다. @type속성은 해당 문자열이 문헌("text")을 지시하는지 인물 혹은 학파("person")을 지시하는지 지정하며, @key는 문자열이 지시하는 대상의 표준적인 명칭을 기록한다. @ana속성은 언급된 대상에 대한 저자의 태도를 기록하는데, 저자가 그것의 권위를 인정하는 경우("authoritative"), 호의적인 평가를 내리는 경우("affirmative"), 중립적인 경우("neutral"), 비판적인 경우("critical")로 나누어 기재한다. @resp와 @cert속성으로는 언급된 지시대상을 밝히는 참조문헌이나 연구물의 @xml:id와 그것의 확실성의 정도를 표시한다.

5.3 인용정보의 인코딩, <q>(quoted)

인용문 표기를 위한 TEI의 대표적인 요소로는 <quote>과 <cit>(citation)을 들 수 있다. 하지만 SARIT 가이드라인이 이미 <quote>을 주석서 내부의 근본텍스트를 표시하기 위해 사용하고 있으며, 인도 논서에 등장하는 인용의 형태가 서지사항을 동반하는 현대적인 의미의 인용(<cit>)으로 볼 수는 없기 때문에 가장 그 정의가 느슨한 <q>(quoted)를 인용정보를 표시하기 위해 선택하였다. <q>는 단순히 인용문에만 사용될 수 있는 요소가 아니라 일반적으로 따옴표 안에 담길 수 있는 모든 정보를 담아내는 요소이다.[26]

```
<p><q type="explicit" source="unknown" ana="authoritative">chos rnams
      la cho kyi rjes su lta ba can du gnas pa'i byang chub sems dpa' des gang <term
         key="pratītyasamutpāda"
      >rten cing 'brel par 'byung ba</term> las ma gtogs pa'i
      chos rdul phra rab tsam yang yang dag par rjes su mi mthong ngo</q> zhes bya ba
   'di la 'gal med de, chos smos pa'i phyir ro.</p>
```

그림 8 인용정보를 기록하기 위한 <q>인코딩의 사례

<q>는 @type, @source, @ana의 세 가지 필수 속성을 갖는다. @type은 저자가 명시적으로 인용문이 인용문임을 표명하는 경우(“explicit”), 인용문 표기를 하지 않는 경우(“silent”), 그리고 다른 저작에 의해 피인용된 경우(“quoted”)를 구분하고, @source는 인용된 또는 인코딩 대상의 텍스트를 인용하고 있는 문헌명을 기록한다. @ana는 언급정보의 경우(<rs>)와 마찬가지로 저자의 인용문에 대한 태도를 표시한다. <q>의 선택 속성으로는 @resp, @cert, @n, @corresp(correspondence)가 있는데 앞의 두 가지 속성은 <rs>의 경우와 같은 의미를 지니고, @n에는 (피)인용원문의 표준적 위치정보를 기입한다. @corresp에는 같은 문서 내부의 <standOff> 요소에 기록한 (피)인용원문 혹은 그에 대한 고전적 번역문들의 @xml:id값을 설정한다.

5.4 개념어 사용정보의 인코딩, <term>과 <gloss>

개념어는 <term>으로 감싸고, 개념어들에 대한 해설은 <gloss>로 감싼다. <term>으로 표시되는 키워드들의 리스트는 미리 정의될 필요가 없는데, 이는 인코더가 인코딩을 하는 과정 속에서 여러 명확하지 않은 관념들이 하나의 단어 속에 응축되는 현장을 포착할 수 있는 가능성이 있기 때문이다. <gloss>로 표시되는 개념어들의 해설도 <term>으로 표기한 키워드 리스트에 구애받으면 안된다. 문헌의 저자가 의도적으로 개념어를 사용하지 않고 풀어쓰고 있을 가능성이 있기 때문이다.

```
<p><term key="saṃvṛti">kun rdzob</term> de ni <gloss target="saṃvṛti"><term
         key="tathya-saṃvṛti">yang dag
      pa</term> dang <term key="mithyā-saṃvṛti"
         >yang dag pa ma yin pa</term>'i bye brag
   gis rnam pa gnyis</gloss> te. de la,</p>
<quote type="base-text">
   <lg xml:id="SDV8abc" type="base-text" n="k8abc" next="k8d">
      <l><gloss target="tathya-saṃvṛti">brtags pa'i don gyis dben gyur pa</gloss>/</l>
      <l><gloss target="tathya-saṃvṛti">dngos tsam</gloss>
         <gloss target="tathya-saṃvṛti">brten nas gang skyes</gloss> de/</l>
      <l><term key="tathya-saṃvṛti">yang dag kun rdzob</term> shes par bya/</l>
   </lg>
</quote>
```

그림 9 개념어 사용정보를 기록하기 위한 <term>과 <gloss>인코딩의 사례

<term>과 <gloss>는 각각 @key와 @target이라는 필수 요소를 지니는데, 이들에는 등장하는 개념어 그리고 해설하고 있는 개념어의 산스크리트 원형을 값으로 설정한다.

5.5 저자의 사유를 응축한 논증식 표기, <seg> (arbitrary segment)

개념어와 개념어에 대한 해설은 보편적인 차원에서 발견되는 텍스트의 요소들이지만, 인도의 논서에는 개념어의 사용 이외에도 사고를 응축적으로 표현할 수 있는 특징적인 방식이 있다. 빠니니의 산스크리트 문법서(기원전 5세기, Aṣṭādhyāyī)에 대한 주석서인 빠딴잘리의 <위대한 주석서>(Mahābhāṣya)에서부터 인도의 지식인들은 개념들 사이에 성립하는 논리적인 관계를 활용한 논증식을 작성하였는데, 불교논리학의 체계를 세운 디그나가(Dignāga, 5세기) 이후 논증식은 자신의 주장을 표현하는 규범적 형식으로 자리잡았다. TEI 가이드라인에는 이와 같은 산스크리트 논서의 특징을 잡아낼 수 있는 요소가 없기에 가장 추상적으로 정의된 <seg>요소로 이를 표현하기로 한다.[27]

```
<seg type="three-membered">
   <seg function="pratijñā"><seg function="dharmin">shes pa</seg> ni <seg
         function="dharma">bdag gis bdag shes pa ma yin</seg> te,</seg>
   <seg function="hetu">rang snang bas stong pa'i phyir,</seg>
   <seg function="dṛṣṭānta">shes pa gzhan bzhin no.</seg>
</seg>
```

그림 10 논증식을 기록하기 위한 <seg> 인코딩의 사례

<seg>요소는 내부에 다른 <seg>요소를 감쌀 수 있다. 최상위 <seg>요소는 @type속성으로 논증식의 형식(“five-membered”, “three-membered”, “two-membered”)을 표현한다. 차상위 <seg>요소는 @function속성으로 논증식의 구성요소를 표현하며, 주장명제(pratijñā)의 경우 주부(dharmin)와 술부(dharma)로 다시 나뉠 수 있다.

6. 해석적 레이어 인코딩과 인도철학 연구

즈냐나가르바의 <두 가지 진리를 구분하기>에 내재한 언급정보, 인용정보, 그리고 개념어 사용정보를 모두 인코딩하게 되면 필자는 과연 무엇을 얻을 수 있을까? 필자가 얻게 될 것은 즈냐나가르바가 그의 저작에서 언급하고 인용한 인물과 서적의 리스트, 그리고 주요 개념어들과 그것들에 대한 그의 해설을 모은 리스트이다. 기존의 연구사를 정리하여 만들어낸 이들 리스트만으로도 우리는 즈냐나가르바의 저작에 대한 훌륭한 지도를 만들 수 있다. 즈냐나가르바가 비판적으로 그리고 호의적으로 언급・인용한 인물 서적들에 다른 색을 입히고 언급・인용의 빈도를 반영해 선의 굵기를 다르게 표시하여 <두 가지 진리를 구분하기>를 중심으로 한 인물과 문헌의 네트워크를 그려볼 수 있다. 또한, 전통적인 개념들을 설명하면서 즈냐나가르바가 가장 많이 혹은 특징적으로 사용한 단어들을 변별해 낼 수도 있을 것이다. 이러한 작업은 그의 저작에 대한 후속연구를 진행함에 있어 확실하고 구체적인 출발점을 제공해줄 것이다.

하지만 <두 가지 진리를 구분하기> XML파일이 보다 구체적인 역사성을 띄게 되는 것은 다른 인도철학의 문헌들이 같은 방식으로 인코딩되어 즈냐나가르바의 저작과 함께 데이터세트를 구성하여 이에 대한 통합적인 분석이 이루어지면서부터일 것이다. 특히 즈냐나가르바 연구자들의 관점에서 그의 저작과 긴밀한 관계를 맺고 있다고 평가되는 문헌들이 데이터화 되었을 때 이들이 야기할 수 있는 효과는 분명하게 예상할 수 있다. 유가행파(Yogācāra) 비판의 역사를 구성하는 중관학파(Madhyamaka)의 다른 텍스트들, 예를 들어 바비베까(Bhāviveka, 6세기)의 <중관의 핵심>(Madhyamakahṛdaya) 5장과 <반야의 등불>(Prajñāpradīpā) 25장의 부록, 짠드라끼르띠(Candrakīrti, 7세기)의 <청명한 말>(Prasannapadā) 1장과 <중관으로의 입문

>(Madhyamakāvatāra) 6장, 샨따락쉬따(Śāntarakṣita, 8세기)의 <중관의 장식>(Madhyamakālaṃkāra)과 까말라쉴라(Kamalaśīla, 8세기)의 <중관의 빛>(Madhyamakāloka)이 모두 본고에서 제시한 규칙에 따라 인코딩되어 있다고 상상해보자. 사실 이를 실제로 구현하기 전에 이러한 작업을 통해 우리가 얼마나 복잡하고 자세한 정보를 얻을 수 있을지 정확히 가늠하기는 어렵다. 하지만 우리는 기본적으로 중관학파라는 하나의 철학사조에 속한 지식인들이 읽었던 책의 리스트가 200년간 어떻게 바뀌어 왔는지, 개별적인 책들에 대한 태도는 어떻게 변화하였는지 추적할 수 있게 된다. 또한 이 200년의 역사를 꿰뚫는 '궁극적인 진리'(paramārtha)와 '세속적인 진리'(saṃvṛti)라는 두 키워드를 포함한 주요 개념어들에 대한 이해방식의 변화 역시 좇을 수 있게 된다. 그리고 이 두 정보를 합하였을 때, 우리는 비로소 중관학파가 유가행파의 주장들을 승인하여 '유가행-중관학파'(Yogācāra-Madhyamaka)라는 정체성을 띄게 된 경위를 그들의 지적 편력의 변천과 연동하여 설명할 수 있는 길을 찾을 수 있지 않을까?

본고에서는 필자의 이러한 문제의식을 해결하기 위해 필요한 정보들을 모아볼 수 있는 TEI 요소와 속성들을 선별해보았다. 필자는 필자의 문제의식이 인도철학 학계가 고민하는 많은 문제들이 공유하고 있는 보편성을 담지하고 있다고 생각한다. 그리고 필자가 제안하는 인도철학 논서 문헌군의 해석적 레이어에 대한 인코딩 스키마는 다른 문헌군에 적용되어 또 다른 문제를 해결하는 데 사용될 수 있다고 생각한다. 그러나 필자가 논서 문헌군의 특징을 잘 변별해 내었는지, 해석적 레이어에 인코딩을 요하는 또 다른 특성은 없는지, 제안한 TEI의 요소와 속성의 사용방식이 TEI의 정의와 양립가능한 것인지, 혹은 인도문헌의 상황에 보다 부합하는 다른 요소와 속성이 있는지에 대한 동료학자들의 검증이 필요하다. 이와 더불어 인도철학 학계가 다루는 텍스트들의 보편성에 대한 토론이 활발히 이루어질 필요가 있다.

이러한 검증과 토론이 성숙해질 때, 우리는 인도문헌을 인코딩하는 표준적인 방법을 만들어 나가야 하며 그것을 TEI 자체가 그러한 것처럼 모듈화해나가야 한다. 인도철학을 공부하는 방법론은 다양하며 각각의 방법론은 각기 다른 태그셋을 필요로 한다. 이에 따라 각각의 프로젝트들은 자신의 목적에 맞는 독립적인 스키마를 설계해야 할 수 있다. 하지만 연구방법론에 따라 새롭게 데이터로 부상하는 정보들에 대한 인코딩 규칙을 다른 프로젝트의 스키마와 양립가능한 형태로 개발한다면 우리는 손쉽게 서로 다른 프로젝트의 데이터세트들을 통합할 수 있을 것이다. 그렇게 반복해 나가다 보면 언젠가 우리는 개별 프로젝트의 데이터세트를 하나의 레고블럭으로 삼아 '인도철학'이라는 성을 쌓고 부수는 놀이를 할 수 있게 되지 않을까?

---

[1] 논서(śāstra) 장르의 형식과 양식적 특징들에 대해서는 함형석(2022)을 참고할 것.

[2] 인도문명과 유럽문명의 만남에 대한 역사적, 철학적 분석은 Halbfass(1988)를 참고할 것. 비-시간적 세계인식은 인도문명의 가장 오래된 문헌인 베다(Veda)에서부터 발견되는 인도 지식인들의 특징적인 사유방식이다. 이에 대해서는 Ham(2013)을 참조할 것.

[3] 인도철학의 '형식'이라고 할 수 있는 육파철학을 체계적으로 서술하는 장르인 '독소그래피'에 대한 개관은 Bouthillette(2020)을 참조할 것.

[4] 문헌들 간의 언급・인용관계에 대한 분석과 아이디어들의 출현양상을 기반으로 인도철학의 주요 인물들의 상대적인 연대를 측정하는 연구로는 Frauwallner(1961)과 Kajiyama(1968-9) 등이 고전적인 사례로 꼽힌다. 물론 이와 같은 '상대적' 연대 측정에도 한어(漢語)와 티벳어 번역본의 번역연대라는 '절대적' 연대가 중요한 참조점 역할을 한다.

[5] 본고의 기획 속에서 TEI 에 대한 개괄적 소개를 위한 별도의 장을 마련할 수는 없었다. TEI 가 국내에 본격적으로 소개된 적은 없으나, 한미경(2020)과 같은 TEI 를 활용한 프로젝트에 대한 몇몇 보고서들을 참조할 수 있다. 국제학계에서 인문 데이터를 저장하는, 혹은 비정형 데이터를 (반)정형 데이터로 변환하는 표준적인 방안으로 TEI가 채택되어 있는 상황을 고려할 때, 국내학계에서도 TEI의 역사와 그것의 전체적인 구조 및 사용방법, 더 나아가 TEI 를 활용한 프로젝트들의 현황을 리뷰할 필요가 있다고 생각한다.

TEI 가 제안하는 인코딩 방식에 대한 전반적인 이해를 위해서는 결국 TEI 의 가이드라인(https://tei-c.org/guidelines/p5/)을 읽어야 한다. 하지만 TEI 를 처음 접하는 연구자가 무작정 가이드라인을 읽기에는 그 난이도와 분량이 상당하다. 개인적으로 TEI 인코딩의 밑바탕에 깔려있는 사고방식을 익히는데에는 Lou Bounard 의 <What is the Text Encoding Initiative?: How to Add Intelligent Markup to Digital Resources>(https://books.openedition.org/oep/426)이 도움이 되었다. "TEI 사용자 90%의 90%의 요구조건을 충족시키기 위해" 맞춤제작된 TEI Lite(https://tei-c.org/guidelines/customization/lite/)는 대표적인 TEI 요소와 속성들을 일람하기에 좋다. TEI 의 사용방법을 익히는 데에는 TEI 컨소시엄(https://tei-c.org/)이 공식적으로 제공하는 <TEI by Example>(https://www.teibyexample.org/exist/)도 유용하다.

[6] https://gretil.sub.uni-goettingen.de/gretil.html#top 의 Introduction 항목의 마지막 문단을 볼 것.

[7] https://gretil.sub.uni-goettingen.de/gretil/corpustei/sa_pataJjali-yogasUtra-with-bhASya.xml

[8] https://github.com/sarit/SARIT-corpus/blob/master/patanjalayogasastra.xml

[9] SARIT Encoding Guidelines, simple version. https://github.com/sarit/SARIT-corpus/blob/devel/schemas/odd/sarit-guidelines.xml

[10] SARIT Encoding Guidelines, full version. https://github.com/sarit/SARIT-corpus/blob/devel/schemas/odd/sarit-guidelines-full.xml

[11] SARIT 이 단순 버전 가이드라인에서 사용하는 모든 TEI 의 요소와 속성들, 그리고 그것들의 간략한 의미는 <부록 1>에 정리해놓았다.

[12] SARIT 이 완전 버전 가이드라인에서 사용하는 모든 TEI 의 요소와 속성들, 그리고 그것들의 간략한 의미는 <부록 2>에 정리해놓았다.

[13] TEI P5, 12. Critical Apparatus. https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html

[14] The Dravyasamuddeśa of Bhartṛhari (Charles Li ed.). https://saktumiva.org/wiki/dravyasamuddesa/start

[15] 본 절에서 서술하는 TEI 커스터마이제이션에 대한 이해는 <TEI P5 Guidelines>의 23 장 "Using the TEI"(https://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html)와 Lou Burnard(2019-2020)의 논문("What is TEI Conformance, and Why Should You Care?")에 기반한 것이다. 두 문서는 비록 커스터마이제이션이라는 행위를 어휘적, 구문론적, 의미론적 차원으로 명시적으로 구분하여 설명하고 있지 않지만, 세 가지 차원에서 커스터마이제이션이 이루어져야 한다는 생각이 그들의 서술에 전제되어 있다고 생각한다.

[16] Burnard(2019-2020), Figure 1 을 재게재한 그림임.

[17] 이와 같은 이유에서 나가사키(永崎 2022)는 TEI 가이드라인을 확장하여 사용한 CBETA(Chinese Buddhist Electronic Text Association; https://www.cbeta.org/) 프로젝트를 비판하고, 유니코드(Unicode)에 등록되어 있지 않은 한자의 표기 등 TEI 가이드라인이 규정하지 않는 텍스트적 현상을 인코딩하는 새로운 요소와 속성 대해서는 정식으로 TEI 컨소시엄에 제안할 것을 권고하고 있다.

[18] 내용 모델과 속성 클래스에 대한 자세한 설명은 TEI 가이드라인 "1.3 The TEI Class System"(https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STEC) 항목을 참조할 것.

[19] 텍스트를 OHCO 로 바라보는 관점에 대해서는 Coombs et al. (1987)과 DeRose et al. (1997)을 참조할 것.

[20] 인터넷 상에 공유된 Garcia 의 "TEI customization" 슬라이드(https://cmohge1.github.io/lrbs-digital-editing-advanced-2019/TEI-customization.pdf)는 실제로 TEI 를 커스터마이즈 할 때 밟아야 하는 절차를 자세히 기록하고 있어 매우 유용하다. 본고에서는 ROMA 를 통한 TEI 커스터마이제이션 과정과 ROMA 를 통해 작성한 스키마를 XML 에디터에 연동하는 과정을 소개하지 않는다. 이에 대해서는 Garcia 의 슬라이드를 참조할 것.

[21] https://www.oxygenxml.com/

[22] 필자가 사용하는 모든 TEI 요소와 속성들, 그리고 그들의 간략한 의미는 <부록 3>에 정리해놓았다. 이하의 서술은 <부록 3>의 내용을 풀어서 설명한 것이다.

[23] 필자는 Akahane(2020)의 비판교정본을 인코딩의 저본으로 삼아 작업하였다.

[24] TEI 가이드라인은 <said>요소를 텍스트에 그렇게 표시되어 있는지의 여부와 관계없이 생각과 발화의 내용을 표시하는 요소라 정의하고 있다. (https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-said.html 참조)

[25] https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-rs.html 참조.

[26] https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-q.html 참조.

[27] https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-seg.html 참조.

참고문헌

한미경 (2020). <내한 선교사 편지(1884-1942)와 디지털 아카이브>, 보고사.

함형석 (2022). “산스크리트 논서(śāstra) 문헌군 인코딩 표준 마련을 위한 기초연구”, <불교학연구>, 73 호, 91-116. DOI: 10.21482/jbs.73..202212.91

永﨑研宣 (2022). “第一章 人文学のためのテキストデータの構築とは”, <人文学のためのテキストデータ構築入門：TEI ガイドラインに準拠した取り組みにむけて>, 12-33.

Akahane, Ritsu (2020). A New Critical Edition of Jñānagarbha’s Satyadvayavibhaṅga with Śāntarakṣita’s Commentary. Arbeitskreis für Tibetische und Buddhistische Studien Universität Wien.

Burnard, Lou (2014). What is the Text Encoding Initiative?: How to Add Intelligent Markup to Digital Resources. OpenEdition Press. https://books.openedition.org/oep/426

_____ (2019-2020). “What is TEI Conformance, and Why Should You Care?,” Journal of the Text Encoding Initiative, Issue 12. https://doi.org/10.4000/jtei.1777

Bouthillette, Karl-Stéphan (2020). Dialogue and Doxography in Indian Philosophy. Routledge.

CBETA (Chinese Buddhist Electronic Text Association). https://www.cbeta.org/ (최종방문일: 2024 년 5 월 12 일)

Coombs, James H., Allen H. Renear, and Steven J. DeRose (1987). “Markup systems and the future of scholarly text processing,” Communications of the ACM 30.11, 933-947.

DeRose, S. J., Durand, D. G., Mylonas, E., & Renear, A. H. (1997). “What is text, really?,” ACM SIGDOC Asterisk Journal of Computer Documentation, 21(3), 1-24.

Frauwallner, Erich (1961). “Landmarks in the History of Indian Logic,” Wiener Zeitschrift für die Kunde Süd- und Ostasiens und Archiv für Indische Philosophie 5, 125-148.

Garcia, Tiago Sousa (2019). “TEI customization.” https://cmohge1.github.io/lrbs-digital-editing-advanced-2019/TEI-customization.pdf (최종방문일: 2024 년 5 월 12 일)

GRETIL (Göttingen Register of Electronic Texts in Indian Languages). https://gretil.sub.uni-goettingen.de/gretil.html (최종방문일: 2024 년 5 월 9 일)

Halbfass, Wilhelm (1988). India and Europe: An Essay in Understanding. State University of New York Press.

Ham, Hyoung Seok (2013). “Inclusivism: The Enduring Vedic Vision in the Ever-Renewing Cosmos,” Critical Review for Buddhist Studies 13, 9-53. DOI : 10.29213/crbs..13.201306.9

Kajiyama, Yuichi (1968-9). “Bhāviveka, Sthiramati and Dharmapāla,” Wiener Zeitschrift für die Kunde Süd- und Ostasiens und Archiv für Indische Philosophie 12-13, 193-203.

Li, Charles. The Dravyasamuddeśa of Bhartṛhari with the Prakīrṇaprakāśa commentary of Helārāja. https://saktumiva.org/wiki/dravyasamuddesa/start (최종방문일: 2024 년 5 월 11 일)

SARIT (Search and Retrieval of Indic Texts). https://github.com/sarit/ (최종방문일: 2024 년 5 월 10 일), https://sarit.indology.info/ (2024 년 5 월 10 일 현재 접근이 불가능한 상태이다.)

TEI Lite. https://tei-c.org/guidelines/customization/lite/ (최종방문일: 2024 년 5 월 30 일)

TEI P5 Guidelines. https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html (최종방문일: 2024 년 5 월 12 일)

Roma – ODD Customization v 1.1.0. https://roma.tei-c.org/ (최종방문일: 2024 년 5 월 12 일)

부록

부록 1. SARIT Encoding Guidelines (Simple Version)에 사용된 요소와 속성

요소 element	요소의 의미	속성 attribute	속성의 값	SARIT 가이드라인 항목 번호
	'장', '절' 등 텍스트 내부에서 사용하는 단위를 표시	@type	텍스트 내부의 단위 이름(예: adhyāya, pāda, adhikaraṇa 등)	6
	게송/쑤뜨라 번호로 근본텍스트와 그에 대한 주석을 하나의 단위로 표시	@n	게송 혹은 쑤뜨라 번호; 복수의 근본텍스트를 한 번에 다룰 경우 복수의 n 값 설정.	5.2, 5.3
<quote>	주석의 대상이 되는 근본텍스트 표시	@type	"base-text"	5.1
<ab> (anonymous block)	쑤뜨라 표시	@type	"sutra"	10, 5.3, 6.1
		@n	쑤뜨라 번호	10
<lg> (line group)	하나의 게송 혹은 게송 그룹 표시	@n	게송 번호	9, 5.1, 6.1
	게송이 주석에 의해 조각나 있을 경우 순서 표시	@prev, @next	게송 번호 + pāda 번호(a, b, c, d)	9.2
<label>	게송 그룹의 제목 표시			
<l> (line)	게송의 한 행 표시			5.1
<caesura/>	게송의 4 분의 1(pāda) 표시			
<p>	주석서의 문단 표시			8, 5, 6.1
<head>	본문에 속하지 않는 제목, 도입문 등 표시	@type	"toc" (목차에 사용될 수 있는 제목 표시)	6.2
<trailer>	콜로폰(colophon) 표시			6.2.2
<lb/> (line beginning)	인코딩의 기반이 되는 판본의 줄바꿈 표시	@ed	판본, 필사본의 약호	7.1
		@break	"no" (줄바꿈과 단어의 끝이 일치하지 않는 경우 표시)	
<pb/> (page beginning)	페이지 바뀜 표시	@n	페이지 번호	7.2
		@ed	판본, 필사본의 약호	7.3
		@break	"no" (페이지가 바뀌는 곳에서 단어가 끝나지 않은 경우 표시)	
<speaker>	희곡작품에서의 화자 표시			11

<stage>	희곡의 지문(地文, stage direction) 표시	11
<sp> (speech)	화자의 대사 표시	11
<note>	각주, 미주 등의 노트 표시	12
<label>	원문에 속하지 않는 편집자가 삽입한 제목, 약호 등 표시	12

부록 2. SARIT Encoding Guidelines (Full Version)에 사용된 요소와 속성

요소 element	요소의 의미	속성 attribute	속성의 값	SARIT 가이드라인 항목 번호
<app> (apparatus)	판본에서 각주 등으로 제시하고 있는 다른 독법 표시			3
<lem>	디지털 판본에서 채택하고 있는 독법 표시	@source	판본 약호	3
	채택된 독법을 제시한 학자의 이름 제시	@resp	학자명 약호	7
<rdg> (reading)		@wit	필사본 약호 (다른 독법의 저본이 필사본인 경우 표시)	3
		@source	판본 약호 (다른 독법의 저본이 판본(edition)일 경우 표시; 같은 필사본에 대한 학자들의 다른 독법 표시)	3, 3.1, 7
		@type	“lacuna” (필사본의 손상으로 인하여 결락된 부분 표시)	5
<choice>	교정전후의 텍스트 표시			8
<sic>	잘못된 독법 표시			8
<corr>	교정한 독법 표시	@resp	xml:id 로 정의된 교정주체의 이름	8
<ref>	다른 텍스트의 독법 참조시 참조 위치 표시	@target	다른 텍스트의 약호	9
<note>	판본이 채택한 독법을 지지하는 전거 표시; 기타 편집자주 기록			9, 12, 13
<listWit>	필사본 리스트 제시			14.1, 1.2
<witness>	하나의 필사본 정보 표시, 혹은 하나의 필사본군을 나열			14.1, 1.2
<listBibl>	참조된 출판물 제시			14.2
<bibl>	하나의 출판물 정보 표시			14.2
<biblStruct>	하나의 출판물 정보 표시			14.2

부록 3. 해석적 정보 인코딩을 위한 TEI 요소와 속성

요소 element	요소의 의미	속성 attribute	필수 여부	속성의 값
<said>	대론 상황에서의 발화 표시	@ana	required	“pūrvapakṣa”, “uttarapakṣa”, “siddhānta”
		@who	optional	알려진 경우 발화자의 정체
		@resp	optional	발화자의 정체를 밝힌 연구자/연구물의 xml:id
<rs> (referencing string)	인물과 문헌에 대한 언급정보 표시	@type	required	“text”, “person”
		@key	required	“anonymous”, 혹은 인물, 학파, 문헌 이름의 산스크리트 stem 형
		@ana	required	(언급된 대상에 대한 저자의 태도) “authoritative”, “affirmative”, “neutral”, “critical”
		@resp	optional	정체 판명의 책임이 있는 인물, 문헌 혹은 연구물의 xml:id
		@cert	optional	“certain”, “possible”, “speculative”
<q>	인용 혹은 피인용이 일어난 구문 표시	@type	required	“explicit”, “silent”, “quoted”
		@source	required	“unknown”, 혹은 문헌명의 산스크리트 stem 형
		@ana	required	(인용문에 대한 저자의 태도) “authoritative”, “affirmative”,

				“neutral”, “critical”
		@resp	optional	(피)인용문 전거 판명의 책임이 있는 인물, 문헌, 혹은 연구물의 xml:id
		@cert	optional	((피)인용문 전거 판명의 확실성) “certain”, “possible”, “speculative”
		@n	optional	알려진 경우, (피)인용원문의 표준적 위치정보
		@corresp	optional	문서내 기록한 (피)인용원문의 xml:id
<term>	관심의 대상이 되는 개념어 표시	@key	required	개념어의 산스크리트 stem 형
<gloss>	저자의 개념어에 대한 이해를 특징적으로 드러내는 구문 표시	@target	required	개념어의 산스크리트 stem 형
<seg>	논증식과 논증식의 구성요소 표시	@type	optional	(논증식의 종류) “five-membered”, “three-membered”, “two-membered”
		@function	optional	(논증식의 구성요소) “pratijñā”, “hetu”, “udāharaṇa”, “upanaya”, “nigamana”, “dṛṣṭānta”, “vyāpti”, “pakṣadharmatā”, “dharma”, “dharmin”