

# 산스크리트 논서(śāstra) 문헌군 인코딩 표준 마련을 위한 기초연구\*

함형석

전남대학교 철학과 조교수

hamhs@chonnam.ac.kr

- |                                       |                                 |
|---------------------------------------|---------------------------------|
| I. 서론                                 | IV. 산스크리트 주석서의 양식과 인코딩<br>의 가능성 |
| II. 인코딩의 세 가지 층위와 SARIT 인코<br>딩 가이드라인 | V. 결론                           |
| III. '산스크리트 주석서'라는 관념적 양식             |                                 |

## 요약문

산스크리트 논서를 연구하는 연구자들은 문헌을 독해할 때 문헌에 쓰여진 텍스트의 내용만을 읽는 것이 아니다. 그들은 '논서'(śāstra)라는 장르가 지닌 고유한 형식적인 양식을 함께 읽으며, 그 형식에 맞추어 문헌을 구조적으로 파악한다. 논서 장르가 갖춘 형식에 대한 독해능력이 없이는 논서의 내용을 이해할 수가 없다고 해도 과언이 아니다.

TEI(Text Encoding Initiative) 가이드라인은 인문학 분야에서 다루는 문서에 내재한 형식 정보를 디지털화하는 표준안으로 정착하였다. 그리고 인도 문헌 데이터베이스 가운데 하나인 SARIT(Search and Retrieval of Indic Texts)은 TEI의 문법을 어떻게 인도 문헌, 특히 주석서가 대부분을 차지하고 있는 산스크리트 논서를 인코딩하는 데

\* 이 논문은 전남대학교 학술연구비(과제번호: 2020-1862) 지원에 의하여 연구되었음.

쓰일 수 있는지를 독자적인 가이드라인을 통해 제시하고 있다.

본고는 SARIT 가이드라인이 권고하는 산스크리트 논서의 형식 정보를 장르, 사본, 출판의 세 가지 차원으로 분석하고, 이 가운데 장르의 차원에는 SARIT이 지적하는 비-문자적 형식 정보뿐만 아니라 전통적인 인도 지식인들의 머릿속에 관념적으로 존재하였던 논서 장르의 문헌이 갖추어야 할 형식이 존재함을 논한다. 이와 같은 ‘논서’라는 관념적인 형식은 ‘논서’적 글쓰기 스타일을 규율한다는 점에서 비-문자적 형식이 아니라 의미론적 차원에서의 문자적 형식이라 할 수 있다.

이후 본고는 산스크리트 논서 문헌군에 보편적으로 발견되는 세 종류의 형식을 식별해보며, 그것들을 문헌의 디지털 판본에 메타데이터로 기입해놓을 경우 연구자들에게 어떠한 해석적 지평을 열어줄 수 있는지에 대해 논한다. 연구자의 관점에서 데이터를 규정하고 데이터베이스를 설계하는 과정을 거칠 때 디지털인문학 프로젝트는 이상적인 연구 환경을 조성하는 데 기여할 수 있을 것이다.

## 주제어

논서, 산스크리트 주석서, 텍스트 인코딩, 디지털인문학

# I. 서론

연구자들은 더 이상 일차자료를 참조하기 위해 도서관이나 서재를 뒤지지 않는다. 불교학의 경우, 대부분의 연구자들에게 연구의 궁극적인 대상이 되는 팔리(Pali), 산스크리트, 티벳어, 그리고 한문 등의 언어로 쓰인 일차문헌은 컴퓨터 파일이라는 디지털적 형태로 존재한다. 일차문헌을 담고 있는 다양한 종류의 파일 가운데에서도 가장 보편적으로 사용되는 파일의 형태는 흔히 e-text라 불리는 ‘단순 텍스트 파일’(plain text file)이다. 워드프로세서를 통해 텍스트 파일을 사용하면, 복사하고, 편집하고, 번역하고, 주석하는 등의 일반적인 인문학적 연구활동을 할 수 있을 뿐만 아니라, 문서 내의 특정 단어와 구문 등에 대한 검색작업과 같은 기존의 아날로그 환경 속에서는 가능하지 않았던 연구행위가 가능하기 때문이다.

불교학과 관련한 일차문헌을 저장하고 있는 데이터베이스들은 문헌이 쓰인

고전어별로 발달하였다. 산스크리트 문헌군의 경우, GRETEL(Göttingen Register of Electronic Texts in Indian Languages)<sup>1)</sup>이 가장 많은 디지털 문헌을 보유하고 있으며, 비슷한 서비스를 제공하는 온라인 플랫폼 가운데 연구자들의 활용도가 가장 높은 곳 역시 GRETEL이라 할 수 있다.<sup>2)</sup> GRETEL은 이전에는 단순 텍스트 파일만을 제공하였지만, 현재는 기존의 텍스트들을 TEI(Text Encoding Initiative)의 가이드라인에 따라 인코딩한 XML파일도 제공하고 있다. 하지만 GRETEL의 XML파일들을 열람해보면, 운문과 산문을 구분하여 표시하는 것과 같은 매우 기초적인 인코딩 작업만 되어 있다. 또한, 산스크리트를 포함한 인도어 문헌의 특성을 반영한 인코딩 가이드라인을 제시하고 있지 않고 있어, 앞으로도 체계적인 데이터베이스 구축을 추구하기보다는 통일성은 결여하지만 가장 많은 데이터를 모아놓는 저장소의 역할을 자임할 것으로 보인다.<sup>3)</sup>

TEI의 인코딩 가이드라인을 적극적으로 도입하여 통일성 있는 산스크리트 문헌 데이터베이스를 구축하려는 대표적인 프로젝트는 SARIT(Search and Retrieval of Indic Texts)이라고 할 수 있다.<sup>4)</sup> TEI 가이드라인은 인문학과 사회과학 자료들을 디지털화 할 때 사용할 수 있는 XML 태그들을 규정하고 그것들의 사용법을 설명한 문서이다.<sup>5)</sup> SARIT은 이에 더해 SARIT에 등록되어 있는 문헌파일들이 따르고 있는, 그리고 앞으로 추가할 문헌을 디지털화 할 때 지켜야 하는 인코딩 기준을 두 가지 버전으로 제시하고 있다. SARIT이 추가적으로 가이드라인을 제시하는 것은 그것이 주된 디지털화의 대상으로 삼고 있는 인도 논서(sāstra)장르의 문헌과 그것의 현대적 편찬물 혹은 교정본에 일반적인 인문학 자료에는 발견되지 않는 특별한 형식적 요소들이 있음을 의미한다. 다

1) <http://gretel.sub.uni-goettingen.de/gretel.html>.

2) 현재 접근 가능한 다양한 산스크리트 문헌 온라인 저장소들의 특징에 대한 간략한 비교는 Meyer 2019, 83을 참조할 것.

3) GRETEL 홈페이지의 Introduction을 참조할 것(2022년 11월 4일 접속 기준).

4) <https://sarit.indology.info/sarit-pm/docs/welcome.html>.

5) 현재 사용되고 있는 TEI 가이드라인은 P5 버전이다. <https://tei-c.org/>를 참조할 것.

시 말해, 인도의 논서 장르에 속하는 문헌들은 특정한 패턴으로 구성되어 있고, 그러한 패턴을 텍스트 파일 속에 기록해 놓기 위해 SARIT은 별도의 가이드 라인을 규정하고 있는 것이다.

본고는 우선 SARIT의 산스크리트 텍스트 인코딩 가이드라인이 디지털 텍스트 속에 기록하고자 하는 문헌의 층위가 세 가지로 구성되어 있다는 사실을 확인하고, 그것이 각각의 층위에서 인코딩하려 하는 정보를 확인한다. 이후 SARIT의 가이드라인이 출판물, 그리고 그것의 저본이 되는 사본에 담긴 정보들에 비해 텍스트 자체에 담긴 정보들에 대해 관심을 기울이지 않고 있음을 지적하고, 산스크리트 논서 장르를 관통하고 있는 형식적인 패턴을 탐구한다. 필자는 이와 같은 문제 제기와 제안을 통해 연구자들의 관심과 요구가 담긴 이상적인 디지털 연구환경을 조성하기 위한 논의가 시작될 수 있는 계기를 마련하고자 한다.

## II. 인코딩의 세 가지 층위와 SARIT 인코딩 가이드라인

### 1. TEI, 형식 속에 담긴 의미의 인코딩

TEI 가이드라인은 종이와 같은 아날로그 매체에 담겨 있는 정보를 디지털화(digitization) 하는 방법을 규정한 하나의 표준이다. TEI는 독자들이 의식적 혹은 무의식적으로 읽는 그리고 읽어낼 수 있는 아날로그 매체에 담겨있는 정보를 모두 디지털 매체, 즉 컴퓨터 파일에 옮길 수 있는 방법을 제시한다. TEI 가이드라인에 맞추어 작성된 XML파일은 아날로그 매체에 담긴 정보를 읽어 낸 결과이며,<sup>6)</sup> 디지털화 작업을 수행한 연구자가 하나의 문헌을 독서한 결과

---

6) XML(Extensible Markup Language)는 디지털 텍스트의 특정 구성요소들을 태그/마크업으로 표시할 수 있게 해주는 표준 문법이며, TEI 가이드라인은 아날로그 형식의 문서를 XML 형식으로 인코딩할 때 사용될 마크업들을 정의해 놓은 문서라고 할 수 있다. XML에 대한 간략한 설명은 TEI Consortium 홈페이지의 설명(<https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>)과 SARIT 가이드라인 '단순 버전'(simple version)의 "3. What is XML?" 항목을 참조할 것.

물이다. 이러한 이유 때문에 텍스트의 모든 정보를 온전하게 디지털화한 단 하나의 ‘디지털 트윈’(digital twin)이란 있을 수 없으며, 한 문헌에 대해 다양한 스타일의 그리고 다양한 수준의 디지털 버전이 존재할 수 있다. TEI의 기준에 맞추어 인코딩하여 디지털화한 파일은, 말하자면, 해당 문헌에 대한 일종의 디지털 ‘판본’(edition)인 것이다.

GRETIL이 제공하는 ‘단순 텍스트 파일’ 역시 활자화된 산스크리트 문헌에 대한 디지털 버전이라 할 수 있다. 단순 텍스트 파일은 주로 인쇄물 위에 적혀 있는 텍스트, 즉 글자 정보를 컴퓨터에 저장하고 처리할 수 있는 형태로 옮기는데 집중한 작업의 결과물이다. 이에 반해, TEI 가이드라인에 따라 인코딩한 XML파일은 텍스트 정보를 기본적으로 포함하지만, 비-텍스트적으로 전달되는 정보를 함께 저장한다는 점에서 단순 텍스트 파일과 차이를 보인다. 비-텍스트적으로 전달되는 정보란, 예를 들면, 줄바꿈, 들여쓰기, 문단의 가운데 정렬, 굵은 글씨나 이탤릭체로의 폰트변화 등 문서의 내용을 구성하는 텍스트 자체는 아니지만, 텍스트 전체를 구조화하고 텍스트에 콘텍스트를 부여하는 형식적 그리고 장식적 장치가 독자에게 전달하는 정보를 일컫는다.

비-텍스트적 요소들을 워드프로세서를 사용하여 최대한 유사하게 단순 텍스트 파일에 기입할 수도 있겠지만, TEI 가이드라인은 디지털 판본을 만드는 인코더(encoder)에게 형식적, 장식적 장치들이 의미하는 바를 기록할 것을 요구한다. 다시 말해, 줄바꿈문자와 들여쓰기가 만들어내는 공백을 단순히 줄바꿈과 들여쓰기로 파악할 것이 아니라 한 문단이 끝나고 새로운 문단이 시작하는 기호로 읽어야 한다는 것이다. 이에 따라 TEI 문서를 작성하는 인코더는 인쇄물의 한 문단을 문단임을 표시하는 TEI 요소(element)인 <p>(paragraph)로 감싸고 줄바꿈문자와 들여쓰기는 기록하지 않는다. 혹은, 하나의 문단이 다른 문단들과는 다르게 가운데 정렬이 되어 있는 경우, 인코더는 파일에 텍스트가 가운데 정렬되어 있음을 기록하는 것이 아니라, 그것이 왜 가운데 정렬이 되어 있는지를 읽어 그 의미를 인코딩한다. 가령 해당 문단이 다른 문헌에서의 직접 인용일 경우 그 문단을 <quote>요소로 감싸주고, 강조된 문단일 경우에는

<hi>(highlighted)요소로 감싸주어 인코딩하는 것이다. TEI 가이드라인에 따라 텍스트를 인코딩하는 것은 텍스트의 형식적 그리고 장식적 요소들을 눈에 보이는 대로 디지털 환경으로 옮기는 것이 아니라 사실상 텍스트에 부과되어 있는 비-문자적 기호들을 독해하는 행위이다.

문서의 겉모습이 아니라 그것의 의미를 인코딩한 파일은 일차적으로 컴퓨터의 운영체제와 문서 열람 소프트웨어에 독립적이라는 장점을 갖는다. 예컨대 MS Word에서 작성한 가운데 정렬된 문단은 아래아한글에서 열람할 때 의도한 대로 표시되지 않을 수 있지만, XML파일은 가운데 정렬이라는 ‘표현’이 아니라 그것이 다른 텍스트로부터의 인용이라는 ‘의미’를 기록하고 있기 때문에 인용문들이 특정 워드프로세서에서 일괄적으로 어떻게 표현되어야 하는지만 정의해주면 해당 문서의 모든 인용문들을 같은 방식으로 표현해낼 수 있는 것이다.

그러나 TEI 방식으로 아날로그 문서의 표현이 아닌 의미를 인코딩한 디지털 문서가 갖는 장점은 단순히 특정 소프트웨어에 독립적이라는 점에 그치지 않는다. TEI 가이드라인에 따라 작성된 파일은 텍스트에 내재한 의미를 가지는 형식적 단위들을 인코딩하기 때문에, 텍스트 속에서 공통된 의미를 지니는 텍스트의 구성 요소들만을 선별적으로 열람할 수 있다. 텍스트가 가운데 정렬이 되어있건, 큰따옴표로 둘러싸여 있건 간에, 그 둘이 모두 인용문이라면 <quote>요소로 인코딩을 하기 때문에, 인코딩이 끝난 후 파일의 사용자는 전체 텍스트에서 인용문만을 선별적으로 열람할 수 있는 것이다. 인용문 뿐 만 아니라, 인코더의 관심사에 따라, 고유명사, 시간정보, 장소정보 등을 인코딩할 수 있고, 이후 전체 텍스트에서 관심 정보만을 용이하게 추출해 낼 수 있다. 그리고 이렇게 추출된 정보들은 인코더가 상상하지 못했던 방식으로 조합되어 새로운 지식을 창출해 내는 데 쓰일 수 있는 가능성을 지닌다.

## 2. 세 가지 차원에서 규정되는 텍스트의 형식

SARIT의 인코딩 가이드라인은 SARIT 프로젝트가 대상으로 하는 인도 문

헌들, 그 가운데에서도 특히 산스크리트로 쓰인 논서 장르의 문헌군을 인코딩할 때 사용되는 TEI의 문법을 모아 놓은 문서이다. 이와 같은 문서의 존재 자체가 산스크리트 논서 장르에 여타 문헌들과는 다른 형식이 발견된다는 것을 의미하며, SARIT의 가이드라인은 연구자들에게 연구대상이 되는 텍스트의 내용이 아니라 텍스트의 형식에 주의를 기울이게 해준다.

SARIT에서 열람할 수 있는 디지털 문헌들은 기존에 출판된 판본을 저본으로 하고 있고, 두 버전의 인코딩 가이드라인 역시 출판물에 담긴 정보를 온전히 디지털화 하는데 초점을 맞추고 있다. 현재 버전의 SARIT의 가이드라인은 인쇄된 형태로 존재하는 한 문헌에 대한 특정 판본을 디지털화하는 전략을 제시하고 있는 것이다.

그런데 필자는 SARIT 가이드라인이 주목하고 인코딩하려는 형식들이 모두 같은 차원의 형식이 아니며, 그것들을 개념적으로 세 가지 차원의 층위로 분류할 수 있다고 생각한다. 분명 SARIT 가이드라인 자체는 그러한 층위를 분리하고 있지 않으며 인쇄물 형식의 문서만을 다루고 있음을 표방하고 있다. 하지만 산스크리트 문헌에 대한 판본 자체가 세 가지 서로 다른 차원 혹은 질서가 요구하는 형식에 의해 규율되고 있으며, 이에 따라 SARIT이 인코딩하고자 하는 인쇄물의 형식들을 세 가지로 범주화 할 수 있다고 생각한다. 그리고 이렇게 텍스트를 규율하는 서로 다른 차원들의 형식을 구분해 낼 때, 텍스트에 내재한 형식적인 측면에 대한 보다 생산적이고 정교한 논의가 가능할 것이라 생각한다.

필자가 생각하는 문헌의 판본에 담긴 텍스트의 형식을 규율하는 세 가지 차원은 장르의 차원, 사본의 차원, 출판의 차원이다. 우선 장르의 차원이 규율하는 텍스트의 형식이란, 텍스트가 하나의 장르에 속하기 때문에 취하는 형식을 말한다. 예를 들면, 만약 텍스트가 희곡작품일 경우, 발화자의 이름은 대사 앞에 다른 폰트로 인쇄되고, 무대배경을 바꾸거나 배우의 행동을 지시하는 지문(地文)은 괄호 속 위치하는 등의 형식적 특성을 띄게 된다. 시 텍스트의 경우에도, 문장이 끝나지 않았는데 줄바꿈이 된다거나, 한 연(聯)과 다른 연을 구분하

는 비어있는 줄이 삽입되어 있는 등의 형식적 특성을 지닌다. 희곡 텍스트의 괄호 속에 위치한 문자들이나 시 텍스트의 빈 줄 등은 하나의 장르에 속한 텍스트들이 공통적으로 지니고 있는 형식적 특징이며, 이는 장르의 차원에서 규정되고 규율되는 형식이라 할 수 있다.

사본의 차원이 규율하는 텍스트의 형식은 산스크리트 텍스트에 대한 근대 이후의 판본이 여러 사본들을 비교, 검토하여 교감한 결과물이라는 특성에서 기인한다. 디지털 환경 속에서 생성된 텍스트의 경우에도 한 텍스트에 대한 여러 버전이 존재할 경우 ‘사본의 차원’이 성립할 수 있다. 출판된 비판교정본(critical edition)은 편집자가 교정한 텍스트 이외에도 한 단어나 구문에 대해 사본들이 서로 다른 독법을 제시할 경우 이를 기록하는데, 이는 보통 각주의 형태로, 또는 교정텍스트 뒤에 괄호를 삽입하여 표현된다. 이는 비판교정본이 의거하고 있는 사본들이 텍스트에 부과하는 내용이며, 편집자가 사본들의 다른 독법을 제시하기 위해 사용하는 각주나 괄호와 같은 형식적 장치들은 사본 차원의 형식이라고 할 수 있다. 다만 여기서 ‘사본’이라는 단어는 교정본의 근거 혹은 교정본의 비교자료라는 의미로 이해해야 한다. 왜냐하면 다른 문헌에서 해당 문헌을 인용한 인용문이, 그리고 특히 불교 문헌의 경우, 해당 문헌에 대한 한문과 티벳어 번역 문헌이 한 단어나 문장을 교정하는 근거로 쓰이는 경우가 많기 때문이다.

출판의 차원이 규율하는 텍스트의 형식은 출판의 과정 속에서 텍스트가 우연히 갖게 되는 형식적 특성, 예를 들면, 한 문장이 특정 페이지의 특정 줄에 걸쳐 인쇄되는 것과 같은 특성을 말한다. 어떤 문헌을 구성하는 하나의 문장이 특정 페이지와 줄에 위치할 필연적인 이유는 없다. 따라서 쪽수와 줄번호 등은 출판의 차원에 속하는 형식이라고 할 수 있다.

### 3. 세 가지 차원으로 구분한 SARIT 인코딩 가이드라인

비록 SARIT 가이드라인이 출판물에 내재한 형식들을 위의 세 가지 차원에서 고찰하고 있지는 않지만, 그것이 제시하고 있는 인코딩의 내용은 이들 세



가지 범주로 나뉠 수 있다. SARIT은 ‘기본 버전’(simple version)<sup>7)</sup>과 ‘풀버전’(full version)<sup>8)</sup>라 명명한 두 종류의 가이드라인을 공개하고 있다. 그 둘의 내용을 비교해보면, 단순 버전은 XML과 TEI에 대한 기초적인 설명을 제공한 후 장르의 차원과 출판의 차원에 속하는 형식을 인코딩 하는 방법을 기술하고 있으며, 풀 버전은 사본의 차원에 속하는 형식과 정보를 인코딩하는 방법을 설명한다.

장르의 차원에 속한 형식 가운데 가장 특기할 만한 산스크리트의 논서의 특징은 바로 대부분의 텍스트가 근본 텍스트(base text, mūla-text)에 대한 주석(commentary)의 형식을 취하고 있다는 점이다. 이에 따라 SARIT 가이드라인도 ‘주석서’라는 형식을 인코딩하는 방법을 기본 버전(5. Base Texts and Commentaries)에서 비중 있게 다루고 있다. SARIT은 주석서를 인코딩의 대상이 되는 기본 텍스트로 다루며, 주석의 대상이 되는 텍스트, 주로 쑤뜨라(sūtra), 까리까(kārikā)라 불리는 근본 텍스트를 주석서의 인용으로 보아 근본 텍스트를 TEI의 <quote> 요소로 감싸는 전략을 취한다. 주석서가 언제나 하나의 계송 혹은 하나의 쑤뜨라를 인용하고 그것에 대한 설명을 완결한 뒤 그 다음 계송이나 쑤뜨라에 대한 설명으로 진행하는 방식을 취하지는 않기 때문에, SARIT은 주석서가 근본 텍스트를 다루는 다양한 방식을 인코딩하는 방법도 소개하고 있다. 즉, 주석서가 복수의 계송이나 쑤뜨라를 한 번에 설명하는 경우, 주석서가 근본 텍스트와 직접적으로 관련되지 않은 논의를 전개하는 경우, 주석서가 계송이나 쑤뜨라를 한 번에 다 인용하지 않고 단어나 구절 단위로 쪼개어 인용하는 경우를 인코딩하는 방법에 대해서도 논하고 있다. 주석서가 근본 텍스트에 부과하고 있는 텍스트의 구조(예를 들면, adhyāya, pāda, adhikaraṇa 등의 구분)와 텍스트 시작과 끝 부분에 추가하는 정형구들을 인코딩하는 부분을 다루는 기본 버전의 6번 항목(6. Sections of the Text)도 장르 차원의 형식을 다룬다고 할 수 있다.

풀버전의 가이드라인은 사실상 전체가 비평가정보에 사용되는 문헌 비판

7) <https://sarit.indology.info/sarit-pm/docs/encoding-guidelines-simple.html>.

8) <https://sarit.indology.info/sarit-pm/docs/encoding-guidelines-full.html>.

자료(*apparatus criticus*)를 인코딩하는 방법을 다루고 있다는 점에서 사본의 차원에 속하는 형식에 대한 것이라 할 수 있다. 풀버전은 SARIT 가이드라인 문서를 작성하는 목적을 설명한 후, 인코딩하는 문헌의 편집자가 선택하지 않은 사본 또는 여타 판본의 다른 독법, 사본의 훼손, 사본의 가필, 다른 문헌에서의 인코딩의 대상 텍스트 인용, 편집자의 사본에 대한 노트 등을 인코딩하는 방법을 제시한다. 다만, 인코딩 대상 텍스트에서의 다른 문헌의 인용을 다루는 12번 항목(12. References to Other Texts)의 경우, 주석서가 기본적으로 행하는 역할 중 하나인 근본 텍스트에 대한 다른 해석 방식의 논파 혹은 다른 학파들의 의견 비판과 관련되어 있기 때문에 이는 장르 차원의 형식에 대한 것이다.

출판의 차원에 속하는 형식과 관련한 항목은 기본 버전 가이드라인의 7번 항목(7. Milestones)이다. 이는 인코딩의 저본이 되는 출판물의 쪽수와 줄번호를 인코딩하는 방법을 소개한다. 다만, 사본의 폴리오(*folio*) 쪽수를 입력하는 방법을 기술하는 7.3항은 사본 차원의 형식에 관한 것이다.

### III. ‘산스크리트 주석서’라는 관념적 양식

SARIT 가이드라인이 인코딩의 대상으로 삼는 정보는 인쇄물의 비-문자적 형식이다. 페이지 하단 정중앙에 적혀진 숫자의 형식에 관심을 기울이지 않는다면 이것을 본문 가운데 등장하는 숫자와 구분하지 않고 옮겨 적을 테지만, 그것의 형식을 읽어 그것을 쪽수라는 의미로 인코딩한다. 굵은 글씨체로 가운데 정렬된 텍스트의 인쇄 형식을 읽어 주석서의 텍스트와 구분된, 주석서가 인용하고 있는 근본 텍스트로 표시해준다. SARIT 가이드라인이 인코딩하는 장르, 사본, 출판의 세 가지 차원에 속하는 형식들은 모두 텍스트 자체가 가진 정보가 아니라 비-문자적인 인쇄 형식에 내재한 정보라는 공통점을 지닌다.

기존에 출판된 인쇄물에 대한 사진파일, 혹은 텍스트 정보만을 타이핑하는 방식으로 생성된 단순 텍스트 파일보다 SARIT 방식으로 인코딩된 XML 파일

은 더 많은 양의 해석적 정보를 머금고 있는 디지털 매체이다. 바로 이러한 이유 때문에 그것은 다른 방식으로 디지털화된 텍스트보다 문헌의 내용을 분석하는 연구자들에게 더 많은 활용가치를 지닌다고 할 수 있다. SARIT 가이드라인은 일반적인 인문학 관련 문서에 대한 인코딩 표준을 마련하기 위한 기준인 TEI 가이드라인이 산스크리트 문헌을 인코딩하는 데에도 쓰일 수 있음을 보여주었다. 그리고 구체적인 사례들을 통해 정보량의 손실을 최소화 하는 방식으로 산스크리트 문헌에 대한 근대 이후의 교정본들을 디지털화 할 수 있는 방안을 제시하였다. 산스크리트 문헌의 경우 사본은 고사하고 교정본마저도 물질적인 형태로는 입수가 어려운 사정을 감안하면, SARIT 가이드라인은 산스크리트 문헌이 디지털 세계에서 살아남을 수 있는 하나의 방법을 제시하였다고 할 수 있다.

하지만 산스크리트 문헌을 디지털화하는 방법 가운데 SARIT 가이드라인에 따라 인코딩하여 XML파일을 만드는 방법이 시간과 노력을 가장 많이 필요로 하는 방법이다. 사진촬영과는 달리 모든 텍스트를 타이핑해야하며, 단순 타이핑과는 달리 문자 정보만이 아니라 비-문자적 정보도 옮겨야 하기 때문이다. 사진 촬영과 텍스트 정보의 타이핑이 육체적 노동이라면, 텍스트 인코딩은 이에 지적인 노동까지 필요로 한다고 할 수 있다. 인코딩된 파일은 텍스트의 내용까지는 아니지만, 텍스트가 출력되어 있는 형식에 대한 독서의 결과물이기 때문이다. 인코딩은 문헌에 지성을 불어넣는 작업이다. 인코딩이 독서에 비견될 수 있는 행위라면, 우리는 인코딩의 수준에 대해서도 논해볼 수 있다. 특정한 정보만을 찾기 위해 책을 속독할 수도 있지만 텍스트가 지닌 의미를 음미하며 같은 책을 속독할 수도 있듯이, 인코딩이라는 행위 역시 표면적인 형식만을 기록하는 정도에서 그칠 수도 있지만 깊은 수준의 형식까지 드러내는 방식으로 행할 수 있다.

필자는 산스크리트 논서 문헌군에 SARIT 가이드라인이 주목하는 장르 차원의 형식보다 더 깊은 차원의 형식 정보가 자리하고 있다고 생각한다. 보다 깊은 차원의 형식은 굵은 글씨체나 줄바꿈 등의 비-문자적으로 표현되는 형식

이 아니다. 그것은 전통적인 산스크리트 주석가들이 논서(śāstra)로서의 ‘주석서’라는 장르가 갖추어야 한다고 생각했었던 형식과 관계된 것이고, 이에 따라 주석가들의 머릿속에 관념적으로 존재했던 형식이다.<sup>9)</sup> 그것은 주석가들의 언어사용을 규율하며, 이 때문에 비-문자적이 아니라 문자적으로 표현되는 형식이다. 그것은 산스크리트 주석서들이 공통적으로 따르고 있는 형식이기에 모든 주석서들의 내용, 즉 그들을 구성하고 있는 텍스트가 다르다 하더라도 일관된 방식으로 표시해낼 수 있는 정보이다.

산스크리트 문헌을 연구하는 연구자들은 이미 인도의 주석서들이 특정한 패턴으로 조직되어 있다는 사실을 경험적으로 알고 있다. 우선 주석서에는 각기 다른 목적을 지닌 하위 장르가 존재한다. 어떤 주석서들은 근본 텍스트에 사용된 단어를 풀이하는 역할만을 하는 반면, 어떤 주석서들은 단어 해설 이후 타 학파의 의견을 소개하고 반박하고, 연관된 주제에 대한 독립적인 논의를 진행하는 등의 내용도 포함하고 있다. 근본 텍스트의 내용 전개를 따르되 단어 차원이 아니라 아이디어 차원에서 그것을 해설하려 하는 주석서도 있으며, 근본 텍스트를 띄엄띄엄 참조하면서 그것에서 파생된 독립된 주제를 주로 다루는 주석서도 있다.

또한 주석서들이 논의를 진행하는 방식에도 일정한 패턴이 존재한다. 글을 시작하면서 근본텍스트의 저자 혹은 텍스트가 속한 전통의 개조나 신적 존재를 찬양하는 망갈라술로까(maṅgalāśloka)를 작성하고, 이후 근본 텍스트의 주제(abhidheya), 목적(prajojana), 그리고 근본 텍스트와 목적 간의 관계(sambandha)를 논한다. 근본 텍스트의 개별 쑤뜨라나 계송(kārikā)에 대해 주석할 경우, 주석서들은 그것을 구성하고 있는 단어와 구문을 하나씩 분석하고 의미를 설명한 후, 단어 단위의 설명을 모두 연결하여 근본 텍스트를 문장으로

9) 학문적인 목적을 위한 산스크리트의 사용이 논서(śāstra)의 형식을 통해 표현되게 된 경위에 대한 간략한 설명은 Filliozat 2000, 109-111을 참조할 것. ‘Śāstra’에 대한 Kumāri와 같은 전통적인 지식인들의 정의와 10세기에 활동한 시인 Rājasekhara의 논서 장르의 분류 방식에 대한 Pažucha 2014, 97-104의 논의도 참조할 것. Bronkhorst 2006은 주석서라는 장르는 그것의 저자에게 특정한 사고 방식(mind set)을 요구한다고 주장한다.

유의미하게 읽는 독법을 제시한다. 이후 근본 텍스트의 문장을 다른 의미를 지닌 것으로 읽는 다른 주석서 혹은 다른 논사의 의견을 소개하고 그것을 반박하거나, 근본 텍스트에 대한 이해를 달리하지는 않지만 근본 텍스트의 내용 자체에 동의하지 않는 타학파의 의견을 소개한 뒤 반박하는 등의 문답이 진행된다. 문답이 일회에 그치지 않을 때도 많으며, 문답의 과정 속에서 종종 파생된 독립주제에 대한 논의까지 이루어지는 경우도 있다.

모든 주석서들의 내용은 다르다. 그러나 우리는 인도의 논사들이 글을 쓰는 패턴을 경험적으로 알고 있으며, 새로운 텍스트를 읽을 때에도 그러한 패턴이 반복될 것을 예상한다. 혹은, 이와 같은 글쓰기 패턴을 알지 못한다면 인도의 논서를 읽을 수 없다고 해도 과언이 아니다.

산스크리트 논서 장르에 속하는 주석서들에는 특정한 양식(樣式)이 존재하고 있으며, 그 양식은 글의 형식을 규율한다. 그것은 글의 내용을 결정하지는 않지만, 글의 내용이 언제 어디에 어떻게 위치해야 하는지를 결정한다. 산스크리트 논서의 양식은 특정 텍스트가 몇 번째 페이지, 몇 번째 줄에 위치해야 하는지를 정하지 않으며, 그것은 굵은 글씨나 가운데 정렬, 혹은 빈 줄 등의 비-문자적 형식에 새겨져 있는 정보가 아니다. 그것은 산스크리트 논서를 작성하였던 인도의 지식인들의 머릿속에 ‘논서’(śāstra)라는 개념의 형태로 존재하는 것이며, 그들이 논서를 집필하는 과정 속에서 물질적으로 표현된다. 그것은 주석가들이 글을 쓰는 스타일이며, 그들이 논의를 구성하는 방식이다. 이러한 이유 때문에 ‘논서’라는 장르의 양식은 문자적 형식에 새겨져 있다고 할 수 있다. 하지만 그것은 텍스트를 구성하는 각각의 문자 그 자체에 새겨져 있는 정보가 아니다. 그것은 문자들을 의미론적 차원(semantic level)에서 규율한다. 주석가들이 어떠한 내용을 어떠한 순서대로 말해야 하는지를 규정하기 때문이다. 따라서 이 정보는 개별 문자 차원이 아닌 하나의 완결된 생각을 표현하는 의미론적 단위인 문장의 질서 속에 새겨져 있다고 할 수 있다.

## IV. 산스크리트 주석서의 양식과 인코딩의 가능성

본 절에서는 산스크리트 주석서들에 보편적으로 발견되는 특징적이고 대표적인 양식을 식별하고, 그것이 텍스트 속에서 드러나는 형식을 파악해보며 그와 같은 정보를 인코딩한 디지털 텍스트가 연구자들에게 어떠한 해석적 지평을 열어줄 수 있는지에 대해 논해보고자 한다. 하지만 산스크리트 주석서의 형식적 패턴을 구체적으로 어떠한 TEI 요소와 속성(attribute)으로 인코딩할 것인지, 다시 말해, 본 절에서 식별한 주석서의 패턴을 TEI 가이드라인의 문법을 통해 어떻게 표현할 것인지에 대해서는 논하지 않는다. 필자의 산스크리트 주석서에 대한 인코딩 경험이 부족하기도 하지만, 개념적인 차원에서 인코딩의 대상을 변별해 내는 것과 그것을 실제로 어떻게 구현할지의 문제는 서로 다른 차원에서 논의해야 할 주제라고 생각하기 때문이다. 본고에서 논의한 내용의 구현 전략에 대해서는 후속 논문을 통해 다루어 보도록 하겠다.

필자가 산스크리트 주석서라는 장르의 차원에서 주목하고 싶은 형식적 특징은 크게 세 가지 범주로 나눌 수 있다. 첫 번째는 주석서라는 장르의 하위 장르에 대한 것이고, 두 번째는 주석서가 기본적으로 독자들에게 제공하는 서비스, 혹은 근본 텍스트의 해석에 있어 응당 주석서가 행해야만 하는 작업에 관한 것이다. 세 번째 특징은 산스크리트 주석서들뿐 만이 아니라 근본 텍스트로 분류되는 쑤뜨라나 까리까 차원에까지 그 영향력을 발휘하고 있는 산스크리트 논서의 특징인 문답 형식의 논의 구성 방식에 관한 것이다.

### 1. 근본 텍스트의 제목과 주석서의 장르

산스크리트 주석서에는 *vṛtti*, *bhāṣya*, *ṭīkā*, *vārttika* 등의 하위 장르들이 있으며, 주석서의 제목은 주석서가 주석의 대상으로 하는 근본 텍스트의 제목과 이 하위 장르의 이름을 결합하여 정해진다.<sup>10)</sup> 즉, 주석서의 제목에는 근본 텍스트

---

10) Preisendanz(2008)는 독특하게도 인도철학 분야의 산스크리트 주석서들을 현대 학자들의 그것에 대한 태도를 기준으로 분류한다. 따라서 Preisendanz가 제시하는 분류법은 전통적인 학자들이 주

의 제목과 해당 주석서가 어떤 종류의 주석서인지에 대한 정보가 담겨있다.

연구자들은 주석서의 각 하위 장르가 지닌 특징을 경험적으로 알고 있다. 예를 들면, *vṛtti*는 근본 텍스트를 구성하는 단어와 문장 단위의 기본적인 의미를 설명하는 장르이고, *bhāṣya*는 이와 같은 작업에 더해 근본 텍스트의 주제 대한 심도 있는 논의를 진행하는 장르이며 *vārttika*는 근본 텍스트의 흐름을 따르면서 근본 텍스트의 결점들을 보완하는 장르라 할 수 있다. *ṭīkā*, *pañjikā*, *ṭippanī* 등은 보통 근본 텍스트가 보다 상위 레벨의 근본 텍스트에 대한 주석서일 경우에 붙는 명칭이며, *ṭīkā*의 경우는 구문의 의미 대한 설명에 집중하는 반면 다른 두 장르는 근본 텍스트의 내용에 대한 해석적 설명이 부가되어 있다.<sup>11)</sup>

이와 같은 연구자들의 경험적 지식을 일반화하고 정형화해 낼 때 우리는 전통적인 인도 지식인들이 ‘논서’라는 장르에 대해 가지고 있었던 관념을 보다 명확하게 이해할 수 있을 것이다.<sup>12)</sup> 그러나 연구자들이 다양한 산스크리트 문헌을 다루면서 받은 인상에 기초하여 전통적인 주석서 장르들의 특징을 일반화하기는 쉬운 일이 아니다. 수많은 예외 사례들이 존재하기 때문이다. 디그나가(Dignāga)의 『집량론』(Pramāṇasamuccaya)에 대한 자주(*svavṛtti*)의 경우처럼

---

석서에 대해 품고 있었던 관념이 아니라 현대 학자들이 주석서에 대해 무의식적으로 가지고 있는 편견을 그대로 드러낸 것이라 할 수 있다. 그녀의 분류법에 따르면, 주석서들은 역사적인 정보를 지닌 주석서(예를 들면, *Yuktidīpikā*), 근본 텍스트의 의도를 대변하고 있는 주석서(예를 들면, *Tattvasaṅgrahapañjikā*), 새로운 사상을 담고 있는 주석서(예를 들면, *Yogabhāṣya*), 인도철학의 불변하는 핵심을 표현한다고 생각되는 주석서(예를 들면, Śāṅkara의 *Brahmasūtra* 주석), 근본 텍스트의 의미를 잘 설명하고 있는 주석서(예를 들면, *Tattvavaiśāradī*), 철학적으로 창의적인 주석서(예를 들면, *Sāṅkhyapravacanabhāṣya*), 철학적으로 유의미하지 않는 주석서(예를 들면, \**Lakṣaṇafīkā*), 근본 텍스트 원문복구의 차원에서 ‘보고(寶庫)’와 같은 주석서(예를 들면, *Hetubinduṭīkā*), 근본 텍스트의 권위를 뛰어 넘어버린 주석서(예를 들면, *Padārthadharmasaṅgraha*) 혹은 후대의 주석서들에 의해 잊혀진 주석서(예를 들면, *Nyāyabhūṣaṇa*) 등으로 분류될 수 있다. Preisendanz가 결론에서 인정하고 있듯, 이러한 분류 체계를 보다 종합적인 것으로 만들기 위해서는 주석서가 염두에 둔 대상 독자 등의 여러 요소에 대한 총괄적인 고려가 필요하지만, 인도철학의 주요 문헌들에 대한 학계의 선입견을 명시적으로 드러내었다는 점에서 그녀의 연구는 충분한 의의를 갖는다고 할 수 있다.

- 11) 논서의 다양한 하위 장르들의 특징에 대한 Rājaśekhara의 요약적인 설명에 대해서는 Pažucha 2014, 107을 참조할 것.
- 12) ‘논서’ 장르의 산스크리트 주석서의 일반적인 특징을 정형화하려는 시도로는 Ganeri 2010이 있다. 그의 논의는 연구자들이 어렵듯이 감지하고 있는 주석서들의 공통된 특징을 구체적으로 명시하였다는 의의를 지니지만, 모든 분야의 연구자들이 동의할 수 있는 내용을 담고 있다고 생각되지는 않는다.

vṛtti는 근본 텍스트에 대한 짧은 설명만을 붙이는 경우가 많지만, 바비베카(Bhāviveka)의 『반야등론』(Prajñāpradīpā)<sup>13)</sup>나 찬드라끼르띠(Candrakīrti)의 『쁘라산나빠다』(Prasannapadā)<sup>14)</sup>의 경우처럼 vṛtti임에도 근본 텍스트에 대한 깊이 있는 해석에 더해 근본 텍스트의 내용과는 직접적으로 관련이 없는 독립적인 주제에 대해 논의하는 텍스트들이 있다. 빠니니(Pāṇini)의 『아스파드야이』(Aṣṭādhyāyī)와 까뜨야야나(Kātyāyana)의 vārttika에 대한 빠탄잘리(Patañjali)의 『마하바샤』(Mahābhāṣya)는 bhāṣya 장르의 전형을 확립하였지만,<sup>15)</sup> 『쌍카까리까』(Sāṅkhyakārikā)에 대한 소략한 설명인 가우다빠다(Gauḍapāda)의 bhāṣya와 같이 모든 bhāṣya가 빠탄잘리의 모델을 따르고 있는 것은 아니다.<sup>16)</sup>

만약 문헌들의 제목에 기록되어 있는 주석서의 장르명이 주석서의 형식과 필연적인 관계가 없는 것이라면, 주석서를 디지털화한 파일에 그것이 어떤 장르에 속하는 지에 대해 인코딩할 필요가 있을까? 필자는 두 가지 이유에서 문헌의 제목에 포함되어 있는 근본 텍스트의 제목과 해당 문헌의 주석 장르를 메타데이터로 인코딩하는 것이 유용할 것이라 생각한다.

우선 근본 텍스트의 제목이라는 메타데이터를 통해 우리는 하나의 근본 텍스트를 중심으로 발달한 문헌군을 쉽게 확정할 수 있을 것이다. 이를 통해 근본 텍스트별로 혹은 각 학파별로 어떠한 장르의 주석서들이 위계적으로 발달

13) 『반야등론』의 전체 제목은 “Prajñāpradīpā nāma Mūlamadhyamakavṛttih”이다. 『반야등론』의 티벳어 판본인 Walleser 1914의 제목은 “Shes rab sgron ma, Prajñāpradīpaḥ”인데, 해당 저작의 티벳어 첫 문장(“rgya gar skad du, pradznyaa pra dīi pa muu la ma dhya ma ka britti”, ibid., 1)에서 이것이 vṛtti임을 밝히고 있다. 『쁘라산나빠다』의 경우와 마찬가지로 여성명사인 vṛtti와 호응하기 위해 『반야등론』 제목은 Prajñāpradīpa가 아닌 Prajñāpradīpā가 되어야 한다.

14) 『쁘라산나빠다』의 전체 제목은 “Prasannapadā nāma Mādhyamikavṛtti”(Poussin 1903-13, 1)이다.

15) Freschi 2018, 338-9 참조.

16) 주석서의 제목이 아닌 내용을 기준으로 한 Preisendanz(2008)의 분류법에서는 주석서가 표방하고 있는 장르와 그 내용의 불일치가 일어나지 않는다. 예를 들어, Preisendanz(2008, 600-601)는 더 이상 접근할 수 없게 된 역사적인 정보를 담고 있다는 공통점에 근거하여 paddhati장르인 Yuktidīpikā와 bhāṣya장르인 Śābarabhāṣya를 동일한 종류의 주석서로 분류한다. Yuktidīpikā의 전체 제목은 “Yuktidīpikā [Sāṅkhyasaptatīpaddhati]”이다. 이 전체 제목은 각 챕터(āhnikā)가 끝날 때마다 언급된다. 예를 들어, 첫 챕터의 마지막 문장(Wezler and Motegi 1998, 55)을 볼 것. ‘D’로 표시되는 사본은 첫 문장에서 본 제목을 기재하는 것으로 되어 있다. cf. “yuktidīpikābhidhānā sām̐khyasaptatīpaddhatih.” (ibid., 1)



하였는지를 확인할 수 있게 되며, 각 학파들이 사유하였던 주석서의 하위 장르에 관한 관념을 파악할 수 있게 된다. 인도의 지식인들이 ‘명칭’을 공유하였던 주석서 하위 장르의 특성이 쉽게 일반화될 수 없는 것이라면, 분석의 단위를 좁혀 개별적인 지적 전통 내에서는 주석서의 장르에 대한 ‘내용’적 이해를 공유하고 있었는지를 분석해보아야 할 것이다.

또한, 후술할 정형화된 주석서들의 주석방법에 대한 메타데이터와 주석서 하위 장르 메타데이터를 결합할 경우, 우리는 실증적으로 특정 장르가 주석서의 기능들 가운데 어떠한 기능에 초점을 맞추고 있는 것인지 살펴볼 수 있게 된다. 이를 통해 연구자들이 평소에 특정 주석서 장르에 대해 가지고 있던 인상을 검증할 수 있게 되며, 이는 특정 장르의 특성을 일반화하는 연구로 혹은 인도철학사 속에서 주석 스타일의 변화를 추적하는 역사적인 연구로 발전할 수 있을 것이다.

## 2. 정형화된 산스크리트 주석서의 주석 방식

Gary Tubb과 Emery Boose의 저작 『학술 산스크리트』(Scholastic Sanskrit)는 주석서 스타일의 산스크리트를 대상으로 한 학습서이다. 저자들은 『냐야꼬샤』(Nyāyakośa)에 실려 있는 한 계송을 인용한 뒤, 계송의 내용을 중심으로 책 전체를 구성하고 있는데, 해당 계송은 다음과 같다.

주석서는 다섯 가지 특징을 지닌다. 단어의 구분(padaccheda), 단어의 뜻 풀이(padārthokti), 복합어 해설(vigraha), 문장의 구성(vākyayojanā), 반론에 대한 대답(ākṣepasamādhāna).<sup>17)</sup>

본 계송이 지적하고 있는 주석서가 제공하는 다섯 가지 ‘서비스’는 연구자들에게 모두 익숙한 것들이다. 주석서는 근본 텍스트의 단어를 해설하기 위해

17) “padacchedaḥ padārthoktir vīgraho vākyayojanā/ ākṣepaḥ samādhānaḥ vyākhyānaḥ pañcalakṣaṇam/” Tubb and Boose 2007, 3에서 재인용.

텍스트를 구성하는 각각의 단어를 개별적으로 다시 인용하여, 주석서를 읽는 사람들에게 근본 텍스트의 단어를 어떻게 끊어 읽어야 하는지 알려준다(*padaccheda*). 이후 흔히 사용되지 않는 난해한 단어 혹은 다양한 의미를 지니고 있는 단어들에 대해 보다 널리 쓰이거나 명확한 단어들로 풀이하며(*padārthokti*), 복합어의 경우, 정형화된 구문을 통해 복합어를 구성하는 낱말들의 관계를 확정한다(*vigraha*). 또한, 단어 차원의 해설이 끝나면, 각 단어들을 문법적으로 어떻게 연결시켜 하나의 완결된 문장으로 읽어야 하는지 제시한다(*vākyaayanā*). 이와 같이 근본 텍스트에 대한 기본적인 해설을 끝낸 후, 주석서들은 근본 텍스트의 철학적 입장에 대한 다른 이해 혹은 반론을 소개하고 그에 대해 대답한다(*ākṣepasamādhāna*). 주석서가 수행하는 첫 번째부터 네 번째 작업은 근본 텍스트의 문장에 관한 것이라 할 수 있고, 다섯 번째 작업은 그것에 담긴 사상에 대한 것이라 할 수 있는데, Tubb과 Boose는 이에 따라 주석서의 첫 네 가지와 다섯 번째 기능을 분리시켜 Part 1과 2로 나누어 고찰하고 있다.

물론 모든 주석서들이 위의 다섯 가지 기능을 모두 혹은 순서에 맞추어 진행하고 있지 않으며, 주석서들이 행하는 작업이 이들에만 한정된다고 보기는 어렵다. 그러나 필자는 위 계층이 주석서들의 작업을 훌륭하게 범주화해내었다고 생각하며, 보다 자세한 주석서들의 공통된 특징에 대한 논의 역시 이 범주에 기초하여 이루어질 수 있다고 생각한다.

그렇다면 주석서를 디지털화함에 있어 이들을 메타데이터로 인코딩하는 데에는 어떠한 이점이 있을까? ‘쁘라띠까’(pratīka)라 불리는, 흔히 판본에서 굵은 글씨체나 밑줄로 표현되는 주석서에서의 근본 텍스트 구문 인용<sup>18)</sup> 그리고 근본 텍스트의 개별 단어에 대한 뜻풀이를 인코딩 해놓는다면,<sup>19)</sup> 디지털 판본을 읽는 독자들이 문헌 내에서의 한 단어나 개념에 대한 용례를 쉽게 파악할

18) 주석서에서 근본 텍스트의 특정 위치를 지시할 때 사용하는 ‘pratīka’를 인코딩하는 방법에 대해서는 McAllister 2020, 3. Quotes as References의 제안을 참조할 것.

19) 주석서에서 근본 텍스트의 낱말을 인용하고 그것의 의미에 대한 간략한 해설을 부가하는 경우를 인코딩하는 방법에 대해서는 McAllister 2020, 4. Quotes of Individual Words/Phrases for Elucidation의 제안을 참조할 것.

수 있을 것이다. 또한 워드프로세서에서 찾기(ctrl+f)기능을 통해 단어들을 검색하는 것보다 훨씬 더 정확하고 체계적인 검색을 할 수 있을 것이다. 산스크리트의 낱말은 산디나 격변화로 인해 그 형태가 변화하기 때문에 워드프로세서에서는 낱말 검색을 온전히 해낼 수 없을뿐더러, 근본 텍스트에 사용된 용례에 대한 간단한 뜻풀이만을 골라서 보는 일은 불가능하기 때문이다.

주석서의 첫 네 가지 기능을 인코딩하는 일은 하나의 근본 텍스트에 대한 복수의 주석서가 모두 디지털화 될 경우 그 효과가 극대화 될 것이다. 우리는 데이터베이스 속에서 각 주석가들의 같은 쑤뜨라 혹은 계승에 대한 상이한 설명들을 비교할 수 있을 것이고, 근본 텍스트에 대한 특정한 독법의 계열들을 식별해내고 아이디어의 흐름을 추적할 수도 있게 될 것이다. 이에 더해 한 문헌 혹은 한 텍스트-전통이 사용하는 전문용어(technical term) 사전을 편집해낼 수도 있게 될 것이라 예상된다. 더 나아가 데이터베이스의 자료가 증가함에 따라 인도철학계가 공유하고 있는 보편적인 철학적 용어가 각 전통 속에서 어떻게 다르게 이해되고 사용되고 있는지를 비교하는 일까지도 기대해볼 수 있다고 생각한다.

### 3. 질문(pūrvapakṣa)과 답변(uttarapakṣa)의 연속으로 이루어진 논서의 형식

주석서가 제공하는 다섯 번째 서비스인 ‘반론에 대한 답변’(ākṣepasamādhāna)에는 특별한 형식이 존재한다. 주지하다시피, 그것은 논적의 반론을 소개하는 ‘논적의 주장’(pūrvapakṣa), 논의를 진행을 위해 주석가가 제시하는 ‘논주의 대답’(uttarapakṣa), 그리고 주석서가 최종적으로 올바른 것으로 판단하는 내용을 담은 ‘정설’(siddhānta)이 그것이다. 논적의 주장과 논주의 대답은 일회적일 수도 있지만, 연속적으로 진행되는 경우도 있으며 논쟁이 끝난 후의 결론이 명확할 경우에는 정설 부분은 생략될 수도 있다. 어디서부터 어디까지가 논적의 주장이고 논주의 대답인지 파악하기 어려운 경우도 있지만, 보통 논적의 주장은 ‘atha’, ‘iti cet’, ‘nanu’ 등의 특수한 어휘로 표시되어 있어 이를 기준으로 우리는

그 둘을 구분할 수 있다. 반면 정설 부분을 특정할 수 있는 별도의 어휘는 존재하지 않아 정설 부분은 독자가 판단하여 설정할 수밖에 없다. 세부적인 각론에 대한 문답이 끝난 후 전체 논쟁의 주제를 요약적으로 제시하고 올바른 의견을 재확인 하는 부분이 분명하게 있을 경우, 주석가가 의도적으로 반론 그리고 대답과는 구분된 ‘정설’을 작성하였다고 판단할 수 있다.

산스크리트 논서 장르에 내재한 대화형식의 논의 진행 방식을 메타데이터로 인코딩하는 것에는 일견 특별한 이점이 없는 것처럼 보인다. 주석서에 담긴 모든 질문과 대답을 표시해놓는 작업을 통해 우리가 얻을 수 있는 것은 해당 문헌에 등장한 모든 질문과 답변을 나열한 리스트에 불과하기 때문이다. 그러나 논적의 주장과 논주의 대답 부분을 구분하며 논적의 정체성에 대한 정보를 함께 인코딩할 경우, 우리는 주석서가 마주하고 있는 지적 전선(戰線)을 재구성할 수 있게 된다. 주석서가 소개하는 반론을 제기하는 학파들의 출현 빈도수를 계산하고, 특정 경쟁학파가 제기하는 반론의 주제들을 종합적으로 분석할 수 있게 되어, 주석서가 대표하고 있는 지적 전통이 인도 지성계 속에서 자신을 어떠한 방식으로 위치 짓고 있는가에 대한 탐구가 가능해질 것이다.

산스크리트 논서에서 드물게 일어나는 일이지는 하지만 주석가가 논적의 주장 부분을 기술하며 논적의 이름을 직접 거명하거나 논적의 글을 직접 인용하는 경우, 혹은 현대 연구자들의 연구에 의해 반론의 출처가 밝혀진 경우, 이러한 정보에 대한 인코딩은 해당 문헌이 직간접적으로 관계하고 있는 문헌군을 파악할 수 있게 해준다.<sup>20)</sup> 그리고 각 문헌들의 관계망을 종합해 볼 때, 연구자들은 문헌적 정보를 바탕으로 특정 시기 인도 지성계의 지형도를 그려볼 수도 있고 통시적으로 지적 지형도의 역동적인 변화를 살펴볼 가능성도 열리게 될 것이다.

주석가는 논주의 대답 부분에서도 자신의 입장을 지지하기 위해 주석가가 자

---

20) McAllister 2020은 본고에서 다루는 문(pūrvapakṣa)·답(uttarapakṣa)의 연속이라는 논서의 구조적인 특징에 대해 주목하고 있지는 않지만, 주석서에서의 근본 텍스트 인용을 포함한 산스크리트 논서 문헌에서의 인용문을 TEI 가이드라인에 따라 인코딩하는 전략을 자세하게 논의하고 있다.

신이 그 권위를 인정하는 다른 저작을 인용한다. 다른 저작으로부터의 인용은 긍정적 인용과 부정적 인용으로 구분할 수 있으며, 이에 따라 인용에 대한 메타데이터는 인용이 이루어지고 있다는 사실 그리고 그것의 출처와 함께 인용의 목적을 긍정적인 혹은 부정적인 경우로 구분하여 기록해야 할 것이다. 한 전통이 다른 지적 전통들과 맺고 있는 경쟁 관계를 시각화할 수 있게 해주는 부정적인 인용의 사례들과는 달리 긍정적인 인용문들은 한 학파 내부의 권위가 어떻게 성립해 나가는지를 보여주는 자료로 사용될 수 있다. 한 학파의 문헌군에서 특정 텍스트의 인용 횟수의 변화를 추적하여 해당 텍스트가 정전화(正典化, canonization)되어가는 과정을 살펴볼 수도 있을 것이며, 역사 속에서 텍스트들 간의 권위가 이동하는 현상을 구체적이고 정확하게 분석하는 일도 가능하게 될 것이라 예상된다.

## V. 결론

산스크리트 문헌에 대한 디지털인문학적 시도들은 대부분 단어 이하의 단위를 대상으로 이루어져왔다. 예를 들면, 산디(sandhi)를 해소하여 단어를 구분한다거나, 단어의 격변화를 자동적으로 감지하여 기본형을 찾아주거나, 또는 격변화에 따른 문장성분을 분석하는 도구들이 개발되어 있다.<sup>21)</sup> 이와 같은 노력은 테크놀로지를 중심에 놓고 산스크리트 문헌에 접근한 결과라 할 수 있다. 산디를 해소하거나 명사나 동사의 기본형 혹은 어근을 기계를 통해 파악할 수 있게 했다는 사실 자체는 놀라운 업적이라고 할 수 있으나, 이는 인간의 지성을 전혀 사용하지 않는 방식으로 산스크리트 문헌을 분석하는 것을 목표로 하고 있다. 하지만 연구자들이 원하는 것은 산스크리트 문장에 대한 기계번역이 아니다. 연구자들은 아무리 정확하게 번역한 문장을 통해서도 알기 힘든 텍스트의 의미를 탐구하고자 하는 것이며, 이를 위해서는 단어와 문장을 뛰어넘는,

---

21) Das 2016, 2-3.

텍스트 그 자체와 텍스트가 속한 시대를 뛰어넘는 콘텍스트에 대한 이해가 필요하다.

디지털인문학적 접근이라고 해서 기술만을 가지고 지식을 생산하겠다는 발상은 유구한 인도학과 불교학의 전통을 도외시할 뿐만 아니라 인간의 지적 능력을 무력화시키려는 시도처럼 보인다. 기술은 지성의 연장(延長)이며, 인간의 사유능력을 극대화하는 도구이다. 기계를 통한 학문활동은 기계가 이끌어어나가는 것이 아니라 연구자가 주도해나가는 것이며, 이를 위해서는 기계에 연구자가 원하는 정보를 가공해낼 수 있는 정보를 입력해야 한다. 데이터의 형식이 데이터로부터 이끌어낼 수 있는 지식의 형식을 결정한다. 연구자는 자신이 원하는 결과물의 청사진을 구체적으로 그려낸 뒤 데이터의 형식을 설계해야만 원하는 정보에 접근할 수 있을 것이다. 빅데이터 분석 관련 서적에서 반복적으로 인용되는 “쓰레기를 넣으면 쓰레기가 나온다”(garbage in, garbage out)이라는 문장은 산스크리트 문헌과 같은 빅데이터를 다루는 디지털인문학 프로젝트에도 그대로 적용된다.

연구프로젝트의 목표에 따라 도출해내고자 하는 지식의 형태는 다를 것이고, 이에 따라 데이터로 여겨지는 정보와 데이터가 저장되는 형식도 달라질 것이다. 본고에서는 산스크리트 문헌 가운데 철학적 ‘논서’ 장르에 속하는 주석서를 대상으로 한 연구가 보편적으로 필요로 할 만한 데이터의 종류와 그것들이 텍스트 속에 어떠한 형식으로 내재하는지에 대해 개괄해보았다. 본고에서 논의한 데이터를 TEI 가이드라인에 따라 어떻게 인코딩해야 할지의 문제는 또 다른 차원의 문제이며 추후 연구를 진행해나가며 다른 지면을 통해 다루어보도록 하겠다.

본고에서 정의한 데이터들, 즉, 산스크리트 주석서 문헌에 공통적으로 발견되는 형식적 특징들을 데이터화하여 문헌을 인코딩하는 방안은 보다 나은 연구환경, 또는 아날로그적인 방식으로는 발견이 불가능한 지식을 얻을 수 있는 조건을 만들기 위한 필자의 제언이다. 그러나 본문에서 논의한 연구 환경을 실제로 구축하는 것은 필자 개인의 노력만으로는 불가능한 일이다. 이는 한 두

문헌이 아닌 대규모 텍스트 데이터에 대한 인코딩 작업을 요하며, 텍스트 인코딩은 단순 타이핑이 아니라 지적인 노동이기 때문에 산스크리트 문헌에 대한 기본적인 이해가 있는 연구자들만이 할 수 있는 작업이다. 요원해 보이는 작업이지만, 학계의 디지털적 전환을 학자들의 관점에서 주체적으로 이루어 낼 수 있는 가치 있는 작업이라 생각한다. 동료 연구자들의 협업이 절실하며, 국내에서도 산스크리트 문헌군의 인코딩 전략과 미래의 이상적인 연구환경에 대한 활발한 논의와 함께 유의미한 프로젝트가 진행될 수 있게 되기를 기대한다.

## 참고 문헌 REFERENCES

### ◆ 약호 및 일차 문헌 ABBREVIATIONS AND PRIMARY SOURCES

- de la VALLÉE POUSSIN, Louise ed. 1913. *Mūlamadhyamakakārikās (Mādhyamikāsūtras) de Nāgārjuna avec la Prasannapadā commentaire de Candrakīrti*. St. Pétersbourg: Impr. de l'Académie impériale des sciences.
- WALLESER, Max ed. 1914. *Shes rab sgron ma, Prajñā-Pradīpaḥ: A Commentary on the Madhyamaka Sūtra by Bhāvaviveka*. Calcutta: Asiatic Society of Bengal.
- WEZLER, Albrecht and Shujun Motegi eds. 1998. *Yuktidīpikā: the Most Significant Commentary on the Sāmkhyakārikā*. Stuttgart: Franz Steiner Verlag.

### ◆ 이차 문헌 SECONDARY LITERATURE

- BRONKHORST, Johannes. 2006. "Commentaries and the History of Science in India," *Asiatische Studien/Études Asiatiques*, vol. 60, no. 4, 773-788.
- DAS, Monali. 2016. *Discourse Analysis of Sanskrit Texts: First Attempt towards Computational Processing*, unpublished PhD thesis (University of Hyderabad).
- FILLIOZAT, Pierre-Sylvain, T.K. GOPALAN trans. 2000. *The Sanskrit Language: an Overview*. Varanasi, India: Indica Books.
- FRESCHI, Elisa. 2018. "Commenting by Weaving Together Texts: Veṅkaṭanātha's *Sésvaramīmāṃsā* and the Sanskrit Philosophical Commentaries," *Philological Encounters*, vol. 3, no. 3, 337-358.
- GANERI, Jonardon. 2010. "Sanskrit Philosophical Commentary," *Journal of the Indian Council of Philosophical Research*, vol. 27, 187-207.
- MCALLISTER, Patrick. 2020. "Quotes, Paraphrases, and Allusions: Text Reuse in Sanskrit Commentaries and How to Encode It," *Journal of the Text Encoding Initiative*, issue 13.
- MEYER, Michaël. 2019. "On Sanskrit and Information Retrieval," in Proceedings of the 6<sup>th</sup> International Sanskrit Computational Linguistics Symposium. IIT Kharagpur,



India: Association for Computational Linguistics.

PAŻUCHA, Katarzyna. 2014. “Notes Towards Defining ‘Theory’ (Śāstra) in Sanskrit: Systematic Classification Presented in Rājaśekhara’s Kāvyaṁīmāṁsā,” *The Polish Journal of Aesthetics*, vol. 32, 97-116.

PREISENDANZ, Karin. 2008. “Text, Commentary, Annotation: Some Reflections on the Philosophical Genre,” *Journal of Indian Philosophy*, vol. 36, 599-618.

TUBB, Gary A. and Emery R. BOOSE. 2007. *Scholastic Sanskrit: A Handbook for Students*. New York: The American Institute of Buddhist Studies, Columbia University.

◆ 온라인 자료

GRETEL (Göttingen Register of Electronic Texts in Indian Languages). <http://gretel.sub.uni-goettingen.de/gretel.html>.

SARIT (Search and Retrieval of Indic Texts). <https://sarit.indology.info/sarit-pm/works/>.

TEI (Text Encoding Initiative). <https://tei-c.org/>.

## A Preliminary Study on Devising an Encoding Standard for Scholastic Sanskrit Literature

HAM, Hyoung Seok  
Assistant Professor, Department of Philosophy  
Chonnam National University

When reading Sanskrit *śāstra* literature, scholars do not merely read the contents contained in its text. They also read the formal patterns unique to the genre of *śāstra* and understand its contents within the context of its formal structure. Without attention to its formal properties, a work of *śāstra* literature remains unintelligible to its readers. The Text Encoding Initiative (TEI) guideline suggests document encoding standards for humanities scholars, and the Search and Retrieval of Indic Texts (SARIT) project accommodates the TEI guideline to Sanskrit texts, especially those of *śāstra*. This paper examines the nature of the information that the SARIT guideline aims to encode by dividing the formality of *śāstra* texts into the forms of genre, manuscript, and publication. It then discusses the idea of *śāstra* that governs and regulates the writing patterns of traditional intellectuals. These patterns are not coded in non-textual devices such as font and line formatting. Instead, they are semantic patterns that are coded in the flow of the contents. In the final section, I observe three general patterns commonly found in the *śāstra* literature (the sub-categories of the genre, basic commentarial services, and the question-answer mode of discussion) and discuss the benefits of encoding that information in XML files.

### Keywords

śāstra, Sanskrit commentary, text encoding, digital humanities

2022년 11월 08일 투고  
2022년 12월 09일 심사완료  
2022년 12월 18일 게재확정