A Study on Creation and Application of Electronic Chinese Buddhist Text:

With the Yogācārabhūmi as a Case Study

Huimin BHIKKHU*, Christian WITTERN, Aming TU,
Li Juan GUO, Ray CHOU

I. Background and Objectives of this Research

Buddhist texts feature a tremendous amount of valuable cultural treasure shared by all human beings. However, the enormous corpus of Buddhist texts as well as their secondary sources usually overwhelms interested researchers, students, or even teachers. Hence, it shall be a significant contribution to Buddhism Studies and an innovative move in the field of human science to employ the digital and electronic media and tools of contemporary information technology, and to utilize the characteristics of multivalent texts as well as standardized markup language (such as SGML, HTML, XML) in the management of vast Buddhist information so as to facilitate more efficient teaching, researching and information service.

The autobiography of Xuanzang (602-664) indicated that one of his goals to study in India was to learn and translate Yogācārabhūmi-sāstra (YBh). The Yogācārabhūmi records different levels of practice and achievements of a yogi (a practitioner or a Meditation practitioner). Encompassing the psychological, philosophical, and religious dimensions of the rich experiences of explorers who pursue the subtle spiritual life of enlightenment, the YBh can also serve as an encyclopedia for Mahayana Buddhism. Hence, this research shall focus on the YBh and aim to achieve the following goals.

1 Take the YBh, its outline books, and various translations as an example and explore the problems and solutions in processing electronic Buddhist texts, the treatment of rare characters, proofreading, and the creation of markups, full text search and web system.

2 Explore possibilities for the application of the digitalized data obtained via the previous process in the development of user interface, research methodology and teaching activities.
II. Achievements of the Research and Discussions

1. Provide an actual sample case for the development of electronic Chinese Buddhist Texts

After one year of experiment and practice, with the YBh, its various translations, commentaries and the outlines book as an example in exploring the possibilities to benefit from the characteristics of electronic texts, a three-step process to create and markup the texts proves to be feasible and effective. It can be described as follows,

(1) Create and markup the contextual features of the document:

Take the compilation of the bibliography of the YBh for example. Its various references can be marked up as ‘monogr’ (single volume books), ‘series’ (books in series), ‘analytic’ (periodical papers) with TEI tag sets. And meanwhile, these marked-up references can also be linked to annotations, the full text of the YBh and other reference books.

(2) Create and markup the structural features of the documents:

Treat the various translations of the YBh and the content of their outline books as ordered hierarchy instead of linear texts and thus structure the original chapters into 20 divisions, then mark them with ‘div’ and check them with HTML Help functions.

(3) Create and markup the non-structural features of the documents:

Use the dictionary entries in Yogacam Dictionary as an index for searching proper nouns, keywords in the YBh and mark these entries with ‘item’.

2. Sample Application of Chinese Buddhist Texts: Comparing Various Translations and Contrasting the Outline Book

The digitalized and marked-up electronic texts thus created can be applied to the comparison and contrast among various translations and the outline book. This can be done in three steps.

(1) Create cross-reference markups among related texts according to their structural features:
A Study on Creation and Application of Electronic Chinese Buddhist (H. BHIKKHU) ( 3 )

YBh (T1579), various translations (T1581-4), commentaries (T1580,1828-9,2259,2801-2, S0794) and the outlines book (T1602) are marked up as div1=DA, div2=DB, div3=DC, div4=DD, ... etc. And add linking id, for instance, div1=T1579DA, div2=T1579DB, etc. The numbering can be done with a program that can add series numbers. Next, to add links markup. (See TEI, 14.4.3: A Three-way Alignment, pp. 433): 〈link Grp〉〈link targets= “A1 B1 C1” /〉〈link targets= “A2 B2 C2” /〉〈/linkGrp〉. Related various translations and the outline book can be linked according to their ids. (e.g.: 〈link targets= “T1579DD065 T1571DB002 T1582DB001” /〉

(2) Developing a User Interface

Display the XML file of the sutra with xml2dir.bat and use IE5, xml parser, DHTML as well as JavaScript to display the table of content. Display the contrast among various translated versions with html.bat Link function, as illustrated in the following:

![Diagram of User Interface]

(3) Put the electronic files on the Internet

Convert the marked-up electronic files into web files and put them on line for public use.
The marked up electronic texts can be used for comparing and contrasting various translations and the outline book.

III. Self-evaluation of the Research project

So far, the contextual, structural and non-structural features of various translations and
outline books have been created and marked-up. Besides, the comparison and contrast among various translations and the outline book can be more efficiently achieved with the cross-reference markup functions developed in this project. And a user interface has also been developed to facilitate the publication of the marked-up texts on line for public use and scholars’ researches. Although the full-text search function has also been completed, the functions of inquiry via item search cannot be provided yet and are to be developed further.

1) 横山敬一・廣澤隆之。《漢梵藏對照瑜伽師地論索引》 (東京：山喜房) 1996
2) 韓清浄科記。《瑜伽師地論科句批尋記彙編》 (台北市：新文豐出版社) 1983
3) Wittern, Christian “Minimal Makrup and More - Some Requirements for Public Texts”, Conference talk held at the 3rd EBTI meeting on April 7th 1996 in Taipei, Taiwan.
4) Barnard, David, Ron Hayter, Maria Karaba, George Logan and John McFadden “SGML-Based Markup for Literary Texts: Two Problems and Some Solutions”, Computers and the Humanities vol.22, p256-76.

(Key words) Electronic Chinese Buddhist Texts, Yogācārabhūmi-sūtra, SGML, XML, TEI

(*Professor, Taipei National University of the Arts, Taiwan; this paper comes from a part of research project of National Science Council, R. O. C.)