Techniques for Collating Multiple Text Versions in the Digitization of Classical Texts: The CBETA Taishō Buddhist Canon as an Example

Huimin Bhikkhu
Professor, National Taipei University of Art / cbeta Committee Chair
Aming Tu
Assistant Researcher, chibs / cbeta Executive Director
Zhou Bangxin
Information Specialist, chibs / cbeta Research and Development Group
Wang Zhipan
cbeta Missing Character Management Groupchibs · Chung-Hwa Institute of Buddhist
Studies / cbeta · Chinese Buddhist Electronic Text Association

p. 299 ⤸

Abstract

The Chinese translation of Buddhist texts began in the Later Han (25-220 c.e.) and continued into the Yuan dynasty (1279-1368). Fuqin Daoan 符秦道安 (580-651) and others in the Sui and Tang dynasties compiled catelogs categorizing those texts. Various terms were used to refer to the Chinese Buddhist canon: Yiqie zhongzang jingdian 一切眾藏經典 (All classics of the piṭakas), Yiqie zangjing 一切經藏 (Canon of all classics), and Da zang jing 大藏經 (Complete canon of classics). However, the circulation of texts was limited because all manuscripts were copied by hand. It was not until the year 971 (Kaibao 4 of the Northern Song dynasty) that a printed version of the Buddhist canon was created, through the use of carved wooden plates. Called the Kaibao 開寶 canon, printed copies were sent to Japan, Khitan, Xixia, Koryǒ (Korea), and also distributed throughout the Chinese kingdom. Later in the Song, numerous other canons were produced, including the Khitan 丹本, Zhaocheng 趙城, Wanshou 萬壽, Vairocana 毘盧, Yuanjue 圓覺, Zifu 資福, and Qisha 磧砂 canons. The Koryǒ canon (or Tripiṭaka Koreana) was also printed during this period. In the Yuan dynasty, compilation continued with the Puning 普寧 and Hongfa 弘法 canons. During the Ming dynasty (1368-1644), the Southern

南, Northern 北, and Fangce 方冊 (or Jiaxing 嘉興) canons were compiled. During Qing dynasty (1644-1911), the Dragon canon 龍藏 was compiled.

The Chinese Buddhist Electronic Text Association's (cbeta) electronic edition of the Buddhist canon is based on the Taishō edition. The compilation and publication of the Taishō 大正 edition, long favored by scholars, began at the end of the Taishō period, in 1924 and ended in 1934. It is a collation of the Koryŏ canon against the Song, Yuan, and Ming versions, with reference to Shōsoin 正倉院 collection, as well as to Early Dunhuang, Pali, and Sanskrit manuscripts. The Taishō canon employs annotations to indicate variant readings found in the different versions. The cbeta electronic version retains all of this information in xml format, which is then rendered into html for display. Individual, uncollated versions of a text can also be selected and displayed for the user. We hope that this description of how such a system was created and how it is used may be of assistance to those interested in creating multi-version, collated electronic texts.

Keywords: 1.Electronic Chinese Buddhist canons   2.multi-version   3.markup language (XML, HTML)   4.tag sets   5.Text Encoding Initiative (TEI)

【Contents】

1. Canons referenced in the Taishō canon volumes 1-55, 85

The Chinese Buddhist canon, in comparison with with other canons, contains the greatest amount of texts and the earliest translations. The earliest translation work was done in the Later Han, and continued until the Yuan Dynasty. Texts were translated into Chinese from Sanskrit, Pali, and Central Asian languages. The earliest figure in Chinese Buddhism is An Shigao 安世高, who went to Luoyang in 148 c.e. He was a prince in the kingdom of Parthia, in northwestern India. The Chinese character an in his name refers to the Chinese word for Parthia, Anxi guo 安息國. An Shigao mainly translated Śrāvaka-yāna texts. Lokakṣema (b. 147), arrived in Luoyang from the country of Da Yuezhi during the final years of the reign of Emperor Huan. He translated works from the Bodhisattva-yāna during the years 178 to 189.

Daoan (?314-385), who lived in China during the Former Qin, was the first to categorize sūtras by type. His Classified Catalogue of sūtras (Zongli zhongjing mulu 綜理眾經目錄) in one fascicle includes the following categories: sūtra, vinaya and abhidharma, multiple Chinese translations from a single Indic version, early texts, texts with unknown translators, texts from Liang (modern Gansu), texts from Guanzhong (western China), texts of suspect legitimacy, and commentarial texts. Altogether it catalogued 639 sūtras comprising 886 fascicles. This was China's first catalogue of Buddhist scriptures.

Subsequently, Sengyou (445-518) and Baochang (d.u.; Liang dynasty) edited and expanded upon Daoan's catalogue. They also had all of the texts copied, which were then installed in the palace and monasteries. During the Sui and Tang dynasties, translation work flourished, and categorization became more refined: sūtras with a single translation, sūtras with multiple translations, abridgements and anthologies, sūtras with suspicious provenance, and apocryphal sūtras. Anthologies, biographies, and treatises compiled or written in China were increasingly canonized. The Chinese Buddhist canon was referred to in several ways: as the yiqie zhongzang jingdian 一切眾藏經典 (the complete [tri]piṭaka classics), yiqie jingzang 一切經藏 (the complete [tri]piṭaka), or dazang jing 大藏經 (the great [tri]piṭaka classics). However, its circulation was limited because texts were copied by hand. In 971 (Kaibao 4), work began on the first printed version, which was called the Kaibao

canon 開寶藏. This circulated to Japan, the Khitan kingdom (Liao), and Korea in addition to its wide circulation within China.

In the Song dynasty, numerous canons were compiled: the Khitan canon (Qidan zang 契丹藏 or Dan ben 丹本), Jin canon (Jin zang 金藏, or Zhaocheng ben 趙城本), the Wanshou 萬壽 canon, Vairocana canon (Pilu zang 毘盧藏), Sixi yuanjue 思溪圓覺 canon, Sixi zifu 思溪資福 canon, and the Qisha 磧砂 canon. Additionally the Koryŏ

高麗 canon was printed in Korea. In the Yuan dynasty, the Puning 普寧 and Hongfa 弘法 canons were printed based on Song versions, but were burned during political upheaval. In the Ming dynasty, during the Hongwu period (1368-1398), the Southern and Northern canons were printed, after collating with previous versions. From 1735 to 1738, during the Qing dynasty, the Dragon canon 龍藏 was compiled based mainly on the Northern canon but included recent texts. Later, there are the Pinjia 頻伽 canon, the Baina 百衲 canon, and the Zhonghua 中華 canon. Editing for the latter began in 1956.

In Japan, the printing of the Buddhist canon begins with the Tenkai 天海 canon, which was printed between 1637 and 1648 in the early Tokugawa period. It was based on the Sixi canon of the Southern Song and the Yuan Puning canon. During the later Tokugawa period (1669-1681), the Ōbaku 黃檗 canon (also known as the Tegen 鐵眼 canon) was printed, which was a reprint of the Ming dynasty Lengyan temple canon. During the Meiji period (1880-1885), Kōkyō shoin 弘教書院 published a small-print edition: it was based on a copy of the Koryŏ canon held at the Zōjōji 增上寺 monastery in Tokyo, collated with Song, Yuan, and Ming editions, and augmented with Japanese ritual texts and the works of the founding patriarchs of Japanese Buddhist sects.

In Japan, the compilation and publication of the Taishō 大正 edition, which has long been the edition favored by scholars, began near the end of the Taishō period, in 1924 and ended in 1934. The Chinese Buddhist Electronic Text Association (cbeta; http://www.cbeta.org) has digitized the Taishō edition, which is based on the Koryŏ edition, and collated with numerous other canons as indicated in Table One below.

The first column of this table, entitled "Taishō abbreviation," is comprised of the abbreviations found at the end of volumes 1-55 and 85 of the Taishō canon. The second column ("cbeta abbreviation") is based on the Taishō abbreviations, and

p.303

additionally includes abbreviations used in Taishō texts but not explained Taishō's chart; an explanation of these abbreviations is included in the cbeta version. The column entitled "version notes and examples" indicates the full name of the collection or cites an instance of its appearance in the Taishō. The "explanation" column contains further explanations quoted from the Taishō chart or lists the volumes in which that edition appears.

Table I

Taishō abbre-viation	cbeta abbrevi-ation	Version notes and examples	Explanation	
	【三】	Song, Yuan, Ming editions	The 'Three Editions' of the Sung, the Yuan and the Ming dynasties	
	【宋】	Song edition	The 'Sung Edition' A. D. 1239	These Dates are subject to reexamination.
	【元】	Yuan edition	The 'Yuan Edition' A. D. 1290	
	【明】	Ming edition	The 'Ming Edition' A. D. 1601	
	【麗】	Koryŏ edition	The 'Kao-Li Edition' A. D. 1151	
	【麗乙】	Koryŏ, reprint	Another print of the Kao-Li Edition	
	【聖】	Shōgozō collection	The Tempyoō Mss. [A. D. 729-] and the Chinese Mss. of the Sui [A. D. 581-617] and Tang [A. D. 618-822] dynasties, belonging to the Imperial Treasure House Shōsō-in at Nara, specially called Shōgo-zōp	
	【聖乙】	Copy two, Shōgozō collection	Another copy of the same	
	【宮】	Old Song edition	The Old Sung Edition [A. D. 1104-1148] belonging to the Library of the Imperial Household	
	【德】	Daitoku-ji edition	The Tempyō Mss. of the monastery 'Daitoku-ji'	
	【万】	Mantoku-ji edition	The Tempyō Mss. of the monastery 'Mantoku-ji'	

	【石】	Ishiyama-dera edition	The Tempyō Mss. of the monastery 'Ishiyama-dera'
	【知】	Chion-in edition	The Tempyō Mss. of the monastery 'Chion-in'
	【醍】	Daigo-ji edition	The Tempyō Mss. of the monastery 'Daigo-ji'
	【和】	Ninna-ji edition	Ninnaji Mss. by Koōūkai and others. C. 800. A. D.
	【東】	Tōdai-ji edition	The Tempyō Mss. of the monastery 'Tōdai-ji'
	【中】	Nakamura edition	Mr. Nakamura's Mss. from Tun-huang
	【久】	Kuhara edition	The Tempyō Mss. belonging to the Kuhara Library
	【森】	Morita edition	The Tempyō Mss. owned by Mr. Seitaro Morita[1]
	【敦】	Dunhuang editions	Stein Mss. from Tun-huang
	【敦乙】		
	【敦丙】		
	【福】	Saifuku-ji edition	The Tempyō Mss. of the monastery 'Saifuku-ji'
	【福乙】		
	【博】	Imperial Museum edition	The Chinese Mss. of the Tang dynasty belonging to the Imperial Museum of Tokyo
	【縮】	Small print edition	Tokyo edition (small typed)
	【金】	Kongō-zō edition	The Mss. preserved in the Kongō-zō Library, Tōji, Kyoto
	【高】	Kōya-san edition	The Edition of Kōya-san, C. 1250 A. D.
n/a	【南藏】	Ex:　T3, p. 110, note 9	Appears in vols. T03, T09, T11, T12, T15, T16, T17,

			T25, T26, T27, T28, T32, T50, T51, T52, T53, T54, T55

p. 305

n/a	【北藏】	Ex:　T9, p. 500, note 1	Appears in vols. T09, T10, T11, T16, T17, T25, T28, T30, T32, T52, T53
n/a	【獅谷】	Ex:　T11, p. 926, note 1	Appears in vol. T11
n/a	【流布本】	Ex:　T12, p. 265, note 5	Appears in vols. T12, T29, T31
n/a	【大曆】	Ex:　T16, p. 726, note 1	Appears in vol. T16
n/a	【日光】	Ex:　T33, p. 854, note 7	Appears in vol. T33
n/a	【明異】	Ex:　T2, p. 353, note 6	Appears in vols. T02, T05, T06, T13, T14, T15
n/a	【內】	see below	Appears in vols. T32, T48, T51
n/a	【別】	see below	Appears in vol. T13
n/a	【南】	see below	Appears in vol. T03
n/a	【宮乙】	see below	Appears in vol. T14
n/a	【敦方】	see below	Appears in vol. T09
n/a	【敦內】	see below	Appears in vol. T09
n/a	【聖丙】	see below	Appears in vols. T25, T26
n/a	【西】	see below	Appears in vols. T16, T24

p. 306

2. Taishō references to the Pāli canon

References to texts in the Pāli canon are abbreviated in the Taishō as follows:

Table II

A	增支部	Aṅguttara-nikāya (Morris-Hardy ed. P. T. S.)
Ud.	小部、自說經	Udāna (Steinthal ed. P. T. S.)
It.	小部、如是語	Itivuttaka (Windisch ed. P. T. S.)
Kh.	小部、小誦	Khuddaka-Pāṭha (Childers ed.)
Jā.	小部、本生經	Jātaka (Fausboll ed.)
Th. 1.	小部、長老偈	Thera-gāthā (Oldenberg ed. P. T. S.)
Th. 2.	小部、長老尼偈	Therī-gāthā (Pischel ed. P. T. S.)
D.	長部	Dīgha-nikāya. (Rhys Davids-Carpenter ed. P. T. S.)
Dh.	小部、法句	Dhamma-Pada (Fausboll ed.)
Nd.	小部、義譯	Niddesa (Stede ed. P. T. S.)
Pv.	小部、餓鬼事	Peta-vatthu (Minayeff ed. P. T. S.)
Bv.	小部、佛種姓	Buddha-vaṃsa (Morris ed. P. T. S.)
M.	中部	Majjhima-nikāya (Trenckner-Chalmers ed. P. T. S.)
Vin.	律藏	Vinaya-piṭaka (Oldenberg ed.)
Vv.	小部、天宮事	Vimāna-vatthu (Gooneratne ed. P. T. S.)
S.	相應部	Saṃyutta-nikāya (Feer ed. P. T. S.)
Sn.	小部、經集	Sutta-nipāta (Andersen-Smith ed. P. T. S.)
Sumv.	長部注	Sumaṅgala-vilāsinī (Carpenter, J. P. T. S.)
Samp.	善見律毗婆沙	Samanta-pāsādikā (Takakusu-Nagai ed. P. T. S.)
~		Pāli equivalent.

3. Text critical notations in the Taishō

The Taishō canon contains notations which relate to critical collations of other canons. These are explained in table three below.

Table III

Taishō abbreviation	CBETA Abbreviation	Explanation
＝		Various reading. Examples: (a)得＝相【元】【明】 (i. e. The Yuan and the Ming Editions read 相 for 得) (b) 一切眾生依得生長＝諸眾生等依得生 (i. e. For the eight letters 〔一切...長〕 read the seven letters 〔諸眾生等依得生〕)
－		Omit, diest Example:〔心〕－【宋】【元】 (i. e. The Sung and the Yuan Editions omit 心)
＋		Add.[2] Examples: (a)我＋（時於彼遇值世尊）【三】 (i. e. The Three Editions add/have 時於彼遇值世尊 after 我) (b)（我時於彼遇值世）＋尊【三】 (i. e. The Three Editions add/have 我時於彼遇值世 before 尊)
＊		So below; So above; et passim. Example: Text: "世尊正[1]遊知" Note: "[1]遊＝遍【三】＊ (i.e. The Three Editions read 遍 for 遊, so also below)
...		Letters or Sentences left out; down to. Examples:

		(a)〔彼依...興衰〕八字 (i.e. 'From 彼依 down to 興衰, eight leters.') (c)〔彼依...興衰〕八字－【三】 (i. e. 'From 彼依 down to 興衰, eight letters are left out in the Three Editions.')
◎		Various division; Various sentence
	∞	Interchange of position

4. CBETA's emendations and version research

4.1 Stylistic changes

The user interface of the cbeta Taishō Reader is displayed in the appendix below. A number of stylistic changes edition have been implemented in the cbeta canon, and are described here:

1.Many instances of incorrect annotation numbering were corrected.

2.The notations 作本文 and 作夾註 were added in the text critical edition.

3.When an ellipsis (...) refers to 30 or fewer characters, those characters were restored. In instances of more than 30 characters, the ellipsis mark remains.

4.When an annotation number in the text does not have a corresponding annotation, the cbeta version either deletes the annotation reference, or makes other adjustments.

5.Wherever "【三】" appears in the Taishō, the cbeta version changes this to "【宋】【元】【明】"; likewise "【三】＊" is changed to "【宋】＊【元】＊【明】＊".

6.Several irregular stylistic instances were handled individually.

7.The mark "●" is used to indicate instances in which the original Taishō text is illegible.

8.For instances in which the Taishō does not specify which version is referenced, the proper notation is added if known, otherwise the notation "闕略" is added.

9.In annotations, all instances of "　" have been changed to "　".

4.2 Text critical abbreviations

A comparative chart with abbreviations used to indicate textual comparison in the cbeta canon is found below. This is based on the chart given at the end of each volume of the Taishō canon. Although the original only allows for three comparison texts (甲, 乙, and 丙), reference to three more versions can be done in a logical manner (using 丁, 戊, and 己). In the chart below, reference is made to the volumes in which those additional abbreviations are employed.

p. 309

Chart IV

Taishō abbre-viation	CBETA abbre-viation	Explanation
	【原】	The MS. or book on which the printed text is based.
	【甲】	The first text collated.
	【乙】	The second text collated.
	【丙】	The third text collated.
n/a	【丁】	T18, T20, T21, T43, T47, T51, T85
n/a	【戊】	T43, T47, T51
n/a	【己】	T43, T47, T51
		A various reading given in a note of the original text or the text collated. cbeta editorial note: The base text or one of the collation texts contains annotations that other versions contain a variant reading. Examples: (a)止＝正　【原】 (i.e. In the original text (MS. or book) it is noted that a text reads 正 for 止) (b)〔止〕　－【甲】 (i.e. In the first text collated it is noted that 止 is wanting in a text)

	A correction given in a note of the original text or the text collated. cbeta editorial note: The base text or one of the collation texts contain an annotation which provides an emendation. Examples: (a)捐＝損　【原】 (i.e. In the original text (MS. or book) it is noted that 捐 may be a mistake for 損) (b)如＋（是）　【乙】 (i.e. In the second text collated it is noted that 是 is to be read under 如)
？	Example: 朋＝明？ (i. e. An editiorial note: 一明 to be read for 朋？)

p. 310

4.3 Version Information

Although the eight abbreviations below are not recorded in the Taishō abbreviation chart, they do appear in some annotations, and are often problematic. The results of our research into these issues has been included below.

4.3.1.【內】	4.3.5.【敦方】
4.3.2.【別】	4.3.6.【敦內】
4.3.3.【南】	4.3.7.【聖丙】
4.3.4.【宮乙】	4.3.8.【西】

_____

4.3.1 【內】

Citation #1: T51, no. 2092《洛陽伽藍記》

cbeta editorial comment: In this text, the source referred to by the abbreviation is unknown. It is possible that 內 is a misrepresentation of 丙.

Citation #2: T48, no. 2003《佛果圜悟禪師碧巖錄》

cbeta editorial comment: In this case, 【內】 may refer to Japan's Naikaku Bunko 內閣文庫. Their catalog, found at the National Archives of Japan 日本國立公文書館, lists a text with the above title as belonging to the Naikaku Bunko collection (item no. 310-0060). It is also possible that 內 is a misrepresentation of 丙.

_____

4.3.2 【別】

Citation: T13, no. 397《大方等大集經》

cbeta editorial comment: On page 251 of the Taishō shinshū daizōkyō kandō mokuroku 大正新脩大藏經勘同目錄 (found in the first volume of the Shōwa hōbō sō mokuroku 昭和法寶總目錄, and hereafter referred to as the Kandō mokuroku), the Da fengdeng daji jing 大方等大集經 entry (no. 397) states that the comparison texts used include the alternate edition of the Koryǒ, and the copy two

p.311

of the Shōgozō edition. The abbreviation 【別】 may refer to one or both of those texts. If this is the case, then 【麗乙】 or 【聖乙】 would be the correct abbreviations, because 【麗別】, 【聖別】, and 【別】 do not conform to the Taishō's own guidelines. This abbreviation only occurs in the Da fangdeng daji jing (no. 397) in two instances, although additional references are made to 【麗乙】 and 【聖乙】.

In a search for annotations containing the item " 〔河〕－【別】 ," i.e. for versions lacking the character 河, there were four results, occurring in the 【三】, 【宮】, 【知】, 【麗乙】, and 【別】 editions. That which is noteworthy is on page 199, note 17: 〔河〕－【麗乙】. There are nine instances of annotations relating to " 〔如〕－【別】 ," i.e., editions without the character 如. Respectively, they appear in 【三】, 【宮】, 【聖】, and 【別】 editions. Annotation no. 2 on page 185 (" 〔如〕－【聖】 ") is also noteworthy. The editions which are without the character 河 and the editions which are without the character 如 are 【聖】, 【聖乙】, or 【麗】, 【麗乙】. There was no mutual confusion between them. In other words, those editions which did not have the character 河 did not include 【聖】 or 【聖乙】; those editions without the character 如 also did not include 【麗】 or 【麗

乙】. Therefore, a reasonable inference would be that in Taishō no. 397, the two instances of the abbreviation 【別】 can be understood in the following manner: in note 14, page 185 (" 〔河〕－【別】") 【別】 may be referring to 【麗乙】, or 【麗別】; in note 19, page 192 (" 〔如〕－【別】"), 【別】 may be referring to 【聖乙】, or 【聖別】. However, neither of the two instances (【麗別】 and 【聖別】) adheres to the Taishō's own stylistic guidelines.

4.3.3 【南】

Citation: T03, no. 174《佛說菩薩睒子經》

cbeta editorial comment: On page 192 of the Kandō mokuroku, the Foshuo pusa shanzi jing 佛說菩薩睒子經 entry (no. 174) does not have a listing that would correspond with 【南】. However, the Daming sanzang shengjiao nanzang mulu 大明三藏聖教南藏目錄, (found in vol. 2 of the Shōwa hōbō sō mokuroku 昭和法寶總目錄) lists two versions of this sūtra: nos. 211 and 213 on page 334 (菩薩

p.312

睒子經 and 佛說睒子經 respectively). Therefore, this instance of 【南】 may be a reference to the Daming sanzang shengjiao nanzang (also known as Ming nanzang 明南藏).

4.3.4 【宮乙】

Citation: T14, No. 440《佛名經》

cbeta editorial comment: On page 259 of the Kandō mokuroku, the entry for the Foshuo Foming jing 佛說佛名經 (no. 440) lists the following versions: 宮 and 宮別. Although the Taishō chart includes only the 【宮】 abbreviation, we can deduce, based on other examples, that 【宮乙】 refers to the 宮別 edition mentioned above.

4.3.5 【敦方】

Citation: T9, no. 262《妙法蓮華經》

cbeta editorial comment: The referent of the abbreviation 【敦方】 is unknown. It is not explained in the Taishō chart, and the three editions relating to Dunhuang include only 【敦】, 【敦乙】, and 【敦丙】. The Dunhuang catalog 燉煌本古逸經論章疏 (并)古寫經目錄 (in vol. 1 of the Shōwa hōbō sō mokuroku 昭和法寶總目錄) contains an entry regarding the catalog of Dunhuang texts held in the British Museum

in London or the French National library 大英博物館并佛蘭西國民圖書館等所藏燉煌本古逸經論章疏目錄. Note 1 of this entry, on page 1055, does not list an edition that might be judged as 【敦方】, despite its list of other Dunhuang editions: 英, 佛, 京, 大, 谷, 龍, 中, 山, 西, 三, 田, 江, 村, and 未. Therefore we have preserved the 【敦方】 abbreviation wherever it appears, but can offer no explanation for it. This issue awaits the analysis of an expert.[3]

p. 313

_____

4.3.6 【敦內】

Citation: T9, No. 262《妙法蓮華經》

cbeta editorial comment: The Shōwa hōbō sō mokuroku 昭和法寶總目錄 does not explain this abbreviation. However, the scanned image of the original calligraphy indicates that 【敦內】 is a misreading of 【敦丙】. Therefore cbeta has corrected all instances to 【敦丙】.

_____

4.3.7 【聖丙】

Citation #1: T25, No. 1509《大智度論》

cbeta editorial comment: In the entry for the Da zhidu lun 大智度論 in the Kandō mokuroku, there are three items listed under "聖": the Sui 隋, Tang 唐, and Tempyō manuscripts 天平勝寶經 of the Shōgozō collection. In the Taishō abbreviation chart, only two editions are listed: 【聖】 and 【聖乙】. As mentioned above, the abbreviation 【聖】is used to represent the Shōsō-in Shōgozō 正倉院聖語藏 edition, which is the Tempyō manuscripts. The 天平勝寶經, listed in the Kandō mokuroku, also comprise part of the Tempyō manuscripts. The abbreviation 【聖乙】 is listed as referring to the copy two of the Shōsō-in shōgozō 正倉院聖語藏本別寫. From the Kandō mokuroku, we know that the Shōgō-in manuscripts 正倉院聖語藏 includes three collections: Sui edition, Tang edition, and the Tempyō mss. 天平勝寶經. Thus it is possible that copy two of the Shōsō-in 正倉院聖語藏本別寫 edition may be either the Sui or the Tang edition found in the original Shōgozō edition. Based on the order of appearance in the Taishō's abbreviation index, which is chronologically arranged, copy two of the Shōsō-in Shōgozō 正倉院聖語藏本別寫 manuscripts, represented by the abbreviation 【聖乙】, may be the Sui edition. Thus, a third copy of the Shōsō-in manuscripts could well be the Tang edition, which would be represented by the abbreviation 【聖丙】. (The entry on page 410 of the Kandō mokuroku is as follows: "No. 1522《十地經論》(中略)[04]〔原〕麗本〔校〕宋本、元本、明本、宮本、(1)聖本(2)知恩院本(中略)[07](1)聖語藏隋經第三號‧

卷第一、卷第八甲、卷第八乙、第四號‧卷第一、卷七、第九、計六卷、隋時代寫天平經第二號‧卷第一、天平初寫(2)知恩院藏本卷第三唐支[日*午]寫"——From this entry one can surmise that the abbreviation【聖乙】refes

to the Sui canon that was part of the Shōgozō collection.) A decisive answer to the question of whether【聖丙】refers to the Shōgozō Tang edition of the Da zhidu lun (T25, no. 1509), however, awaits the evaluation of a specialist.

Citation #2: T26, no. 1522《十地經論》

cbeta editorial comment: In the entry for the 十地經論 Shi di jing lun (no. 1522) in the Kandō mokuroku on page 410, two Shōgozō editions are listed:【聖】and【聖乙】(please refer to the above citation). In a search in the Shōgozō catalog 正倉院御物聖語藏一切經目錄 for the Shi di jing lun, there were the following results: it appears under the section entitled "primary manuscripts" 第一類寫經之部 on page 946 as part of the Sui canon 隋經, page 948 as part of the Tempyō manuscripts 天平經, and page 963 as part of the Mei manuscripts 名寫. (This can be compared with an entry on page 476 of the Kandō mokuroku: "No.1851 大乘義章(中略)[04]〔原〕(1)大谷大學本〔校〕(2)聖本、(3)村上本(中略)[07](1)延寶二年刊(2)聖語藏名寫經第五號卷第九、第十三、第十四、仁平四年寫(3)村上專精藏延寶二年刊)，校勘條目僅略符【聖】這一項，推定「名寫」也屬於「天平寫經."」 But with only the two Shōgozō editions cited, we are left without evidence that the third (【聖丙】) is the Tang edition. Under the abovementioned section of "primary manuscripts" in the Shōgozō catalog, the entry for the Tang canon does not have listings that correspond to the Shi di jing 十地經. However, in fascicle two, this catalog contains a category entitled "secondary manuscripts" 第二類(雜經)寫經之部, containing entry no. 102 on page 970 "十地論五卷," and entry no. 103 "十地經八卷," as well as other related texts. Nos. 102 and 103 appear between no. 99 "大唐內典錄十二卷" and no. 126 "貞元新定釋教目錄二十卷"; it is possible that, based on the order of appearance, nos. 102 and 103, may also belong to the so-called Tang canon segment of the Shōgozō catalog 正倉院御物聖語藏一切經目錄. This would be the source for the abbreviation【聖丙】used in the Shi di jing lun 十地經論 (T26 no. 1522). However, a definitive answer awaits the analysis of a specialist.

_____

### 4.3.8 【西】

Citation: T24, no. 1496《佛說正恭敬經》

p. 315

cbeta editorial comment: This abbreviation is not included in the Taishō charts. The Kandō mokuroku on page 405, lower register, records that the comparison texts are "Song, Yuan, Ming, Old Song, and Xifu si editions 西福寺." Thus it is very likely that the abbreviation 【西】 ought to be replaced with 【福】.

5. CBETA Taishō techniques digitizing multiple versions

5.1 xml markup and its html appearance

The following introduces the manner in which cbeta employs xml (Extensible Markup Language) to record version information of all collated texts.[4] For detailed information on xml, please refer to the works listed in the bibliography at the end of this article.

5.1.1. Some basic examples
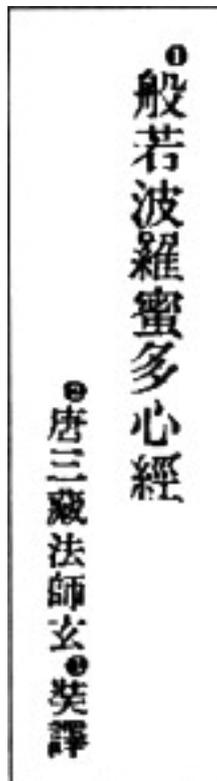
The image to the right includes the title（般若波羅蜜多心經）and translator information (唐三藏法師玄奘譯) from the Heart sūtra in Taishō volume 8, no. 251, page 848, register c (T08, no. 251, p. 848c). Our example is taken from annotation

two of this text segment, which reads, "[2]唐三藏法師玄奘譯." The corresponding annotation reads: "2〔唐〕－（宋）" (the scanned image is included below).

Thus this annotation includes the information that the Song edition of this segment does not contain the character 唐, thus reading 三藏法師玄奘譯. In cases like this, cbeta uses the following xml code to represent this information:

<app n="0848002">

<lem>唐</lem>

<rdg wit="【宋】">&lac;</rdg>

</app>三藏法師玄奘譯

p. 316

The xml tags <app>, <lem>, and <rdg> are defined in tei's (Text Encoding Initiative) manual Guidelines for Electronic Tex Encoding and Interchange.[5] Descriptions of these encodings are also found on tei's website (http://www.tei-c.org):

<app>(apparatus entry): contains one entry in a critical apparatus, with an optional lemma and at least one reading.

<lem>(lemma): contains the lemma, or base text, of a textual variation.

<rdg>(reading): contains a single reading within a textual variation.

wit (witness): contains a list of one or more sigla of witnesses attesting a given reading.

Following tei's standards, we use the <app> tag to indicate that this is an annotation. The value of n (n="0848002") contains the information indicating page 848, note 2. The next tag, <lem>, records the text of the Taishō, in this case "唐": <lem>唐 </lem> . The tag <rdg> records text from other versions, and its associated variable wit records the name of the comparison text, in this case 【宋】. We use the entity &lac;to associate the brief 【宋】 with its full title. This xml data is then used to generate the following html code:[6]

[2]

唐

三藏法師玄奘譯

Text from different versions is included in the <span> tag; wit is used to record the name of the version. Javascript can be used to change the style and visibility of <span> in order to display a given version.

p. 317

5.1.2 Inline annotations

Image 3 on the right contains an example of half-size annotation, from the Chang ahan jing 長阿含經 (T1 no. 1 p. 117a13), which reads, "故名阿耨達 [03]（阿耨達秦言無惱熱）" We use the <note place="inline"> tag to indicate half-size annotation:

故名阿耨達<note place="inline">阿耨達秦言無惱熱</note>

The annotation text at the bottom of the page is as follows:

[03]宋作本文。明無夾註。

Thus the half-size text appears in the Song edition, but not in the Ming. Currently xml is used in the following manner to include text critical information:

故名阿耨達

```
<app n="0117003">

<lem><note place="inline">阿耨達秦言無惱熱</note></lem>

<rdg wit="【宋】">阿耨達秦言無惱熱</rdg>

<rdg wit="【明】">&lac;</rdg>

</app>
```

We notice that the code used to contain the Song edition information is <rdg wit="【宋】">. Although the text is the same, it is not enclosed in the <lem> tag using <note place="inline">...</note>. This indicates that the Song edition uses the current style for display. The html generated by this xml is as follows:
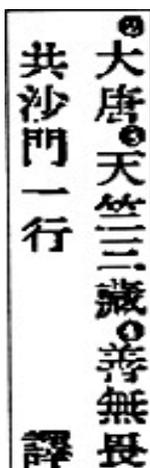
故名阿耨達

```
<span wit="【大】"><span class="note">阿耨達秦言無惱熱</span></span>

<span wit="【宋】">阿耨達秦言無惱熱</span>
```

```
<span wit="【明】"></span>
```

Therefore, although the two versions use different words, we can use style sheets to cause annotations to have varying display formats. Thus the body text can be displayed differently, as can the annotation text.

5.1.3 Annotations within annotations

A more complex example is found in the image 4 to the right (T18, No. 848, p. 1a06), which reads, "[02] 大唐 [03] 天竺三藏 [04] 善無畏共沙門一行譯."

The annotations at the bottom of the page are as follows:

[02]〔大〕－【明】，〔大唐...譯〕－【甲】

[03]（中）＋天【三】【宮】

[04] 善無畏＝輸波迦羅【三】【宮】

This segment of text contains three annotations, and four groups of annotation information, of which three groups are associated with a portion of the segment, and the fourth relates to the entire segment:

[02]〔大〕－【明】 The Ming edition does not contain the character "大."

[02]〔大唐...譯〕－【甲】 The entire segment is not found in the first comparison text.

[03]（中）＋天【三】【宮】 The Song, Yuan, Ming, and Old Song editions include "中" before "天."

[04] 善無畏＝輸波迦羅【三】【宮】 In the Song, Yuan, Ming, and Old Song editions, "善無畏" is read as "輸波迦羅."

For this, xml nested structure is used (<!-- --> is used for comments):

<app n="0001002"><!-- 最外層這組 app 記錄整段文字的校勘資訊 -->

p. 319

<lem>

<!-- 內層包含三組校勘資訊 -->

<app><!-- 第一組 app 記錄明本沒有「大」字-->

<lem>大</lem>

<rdg wit="【明】">&lac;</rdg>

</app>唐

<app n="0001003"><!-- 第二組 app 記錄三本及宮本多了「中」字 -->

<lem>天</lem>

```
<rdg wit="【宋】【元】【明】【宮】">中天</rdg>
```

```
</app>竺三藏
```

```
<app n="0001004"><!-- 第三組 app 記錄三本及宮本作「輸波迦羅」 -->
```

```
<lem>善無畏</lem>
```

```
<rdg wit="【宋】【元】【明】【宮】">輸波迦羅</rdg>
```

```
</app>共沙門一行譯
```

```
</lem>
```

```
<rdg wit="【甲】" resp="Taisho">&lac;</rdg><!-- 甲本整段沒有-->
```

```
</rdg>
```

```
</app>
```

Likewise, html can output nested code using the <span> tag:

[2]

```
<span wit="【大】【宋】【元】【明】【宮】">
```

```
<span wit="【大】【宋】【元】【宮】">大</span>
```

```
<span wit="【明】">&lac;</span>
```

唐[3]

```
<span wit="【大】">天<span>
```

```
<span wit="【宋】【元】【明】【宮】">中天<span>
```

竺三藏[4]

```
<span wit="【大】">善無畏</span>
```

```
<span wit="【宋】【元】【明】【宮】">輸波迦羅</span>
```

共沙門一行譯

```
</span>
```

p. 320

```
<span wit="【甲】">&lac;</span>
```

In this type of complex nested structure, the visibility of <span> can be used to control which textual version is shown.

5.2 Difficulties: collation and structural tags

When collation occurs within a tagged structure, it may destroy the integrity of the original tags; therefore the DTD definition of the original tags must be changed. Currently there is no better solution.

For example, the xml code for an item list might be as follows:

```
<list>

<head>......</head>

<item>......</item>

<item>......</item>

......

</list>
```

DTD 裏定義裏限制 <list> 下只能有 <head> 跟 <item>：

Limits in DTD definitions for <list> can only take <head> and <item>:

```
<!ELEMENT list (head?, item+)>
```

Then, if we add version collation information, this limit will be violated, e.g.:

```
<list>

<head>......</head>

<item>......</item>

<app><!-- 某些項目在其他版沒有 -->

<lem><item>......</item></lem>

<rdg></rdg>
```

</app>

…….

p. 321

</list>

Therefore, the DTD must be changed to the following:

<!ELEMENT list (head│item│app)*>

Other tags, such as <table>, have similar difficulties:

<!ELEMENT table (head?, row+)>

5.3 Conclusion

The vast majority of standard annotations have already been converted using the <app> tag; however, there are still many instances of annotations which still require further critique and tagging. Thus although the cbeta Taishō is suitable for use as a reference tool, it is not yet able to guarantee that every version will be reconstructed perfectly. Nonetheless, our recent developments may be of value to other digitization projects.

p. 322

【Bibliography】

• Erik T. Ray 原著。陳建勳譯（2001）。《XML 學習手冊》。臺北市：歐萊禮（O'Reilly）。

• Elliotte Rusty Harold 原著。陳牧群‧連春雨譯（2001）。《XML 精要總覽》（XML in a Nutshell）。臺北市：歐萊禮（O'Reilly）。

• C. Michael Sperberg-McQueen and and Lou Burnard ed. (2002). Guidelines for Text Encoding and Interchange. Oxford: University of Oxford.

• David Flanagan 著。陳建勳‧黃吉霈編譯（2003）。《JavaScript 大全 第四版》（JavaScript: The Definitive Guide, 4rd Edition）。臺北市：歐萊禮（O'Reilly）。

- 《香光佛教圖書館館訊：電子佛典製作》no. 24（2000）。

http://www.gaya.org.tw/journal/m24/24-index.htm

p. 323

Supplementary images: CBETA's Electronic Tripitaka Interface

Image 1. Multiple versions can be generated and displayed in the "full annotation information" box in the lower left. Thus the annotation for the text "稽首智度無[16]等佛" clearly indicates that the Yuan, Ming, and Shōsō-in editions read "稽首智度無子佛."

p. 324

Image 2. Using the dropdown list to select versions, various individual versions can be generated and displayed.

本選擇 CBETA 大正 校勘轉換 【CBETA】 ∨ 隱藏校勘符號

卷 第五卷

護。以二三唯案...備飾。幸冀明悟之賢。略其文而...

[10]大智度初月...釋論第一(卷第一)

No. 1508
《阿含口觧十二因緣經》

【CBETA】
【大】
【宋】
【元】
【明】
【麗乙】
【聖】
【聖乙】
【聖丙】
【宮】
【石】
【甲】
【北藏】
【南藏】

[11]後秦

No. 1509
《大智度論》

智度大道佛[...] 智度[14]大海佛窮[15]盡
智度相義佛無... 智度無[16]等佛
[17]有無二見... 諸法實相佛所說
常住不壞淨煩... 所尊重法
聖眾大海行福田   學無學人以莊嚴
後有愛種永[18]已盡   我所既滅根[19]亦除

（This paper was presented at the Memorial Symposium "Religions in Chinese Script: Perspectives for Textual Research" for the 75th Anniverary of the Foundation Institute for Research in Humanities, Kyoto University; November

18th-21st, 2004）

數位化古籍校勘版本處理技術
－ 以 CBETA 大正藏電子佛典為例

釋惠敏
(中華電子佛典協會主任委員)國立臺北藝術大學教授
杜正民
(中華電子佛典協會總幹事)中華佛學研究所副研究員
周邦信
(中華電子佛典協會研發組)中華佛學研究所資訊專員
王志攀
中華電子佛典協會缺字處理組

提要

我國漢譯佛典，起自後漢，迄於元代。有秦道安乃至隔唐，雖有蒐集分類，編成目錄，總稱佛典為「一切眾藏經典」、「一切經藏 J 、「大藏經」 但是流通皆賴書寫。直至宋開寶四年(971)始刻印（木版印刷）版本，稱為開寶藏，並頒賜給日本、契丹、西夏、高麗諸國，以及國內各地。此後有契丹藏（丹本）、金藏（趙城本）、萬壽藏、昆盧藏、圓覺藏、資福藏、積砂藏等宋朝版本，以及韓國的高麗藏；元代有 n 寧藏、弘 Õ 藏等；明朝刊刻南藏、北藏等；清朝的龍藏。

中華電子佛典協會(CBETA)採用目前廣為學術界使用的日本大正時代開始(l924.1934)編輯出版的藏經（簡稱《大正藏》）為底本，進行數位化的作業。《大正藏》是以高麗本為底本，對校宋、元、明三本，另參照正倉院藏經、敦煌古本及巴利文、梵文經典，並在校勘欄中記錄了各版,的不同用字等資訊。CBETA 在製作電子佛典的過程中，將這些校勘資訊以 XML 記錄，並以 HTML 方式呈現，藉由校勘資訊做部份的版本還原，讓使用者可以選擇瀏覽不同版本。此作業過程及其呈現方式或許可作為數位化古籍校勘版本處理技術的參考。

關鍵詞： 1.中華電子佛典　2.校勘版本　3.標記語言(XML,HTML)　4.標籤集 5.全文文獻編碼標示標準(TEL the Text Encoding Initiative)

[1] We have changed "Ms." to "Mss." based on context.

[2] Instead of using the word "add" as marked in the Taisho edition originally, the CBETA Electronic edition suggests the word "have" to signify that the word(s) shown in one edition are simply different from another edition, because it is not necessarily the case that the word(s) was necessary been "added."

[3] Prof. TAKATA Tokio (Institute for Research in Humanities, Kyoto University) told me, maybe we can refer to 【敦方】 as the text of 方若, a collector of Dunhuang texts.

[4] XML 1.0 standards are determined by the World Wide Web Consortium; see http://www.w3.org/XML.

[5] Edited by C. Michael Sperberg-McQueen and Lou Burnard (Chicago and Oxford, 1994).

[6] HTML (HyperText Markup Language) is one application of SGML (Standardized Generalized Markup Language).