BRILL

Collation Strategies for the Buddhist Canon—As Seen in the Frequency and Impact of Character Variance in Canonical Editions of the Song Gaoseng Zhuan 宋高僧傳 (T.2061)

Marcus Bingenheimer
Temple University
marcus.bingenheimer@temple.edu

Abstract

The different readings given in the apparatus of the authoritative Taishō edition of the Buddhist canon are widely used, but we do not know much about their relationship to previous efforts, their accuracy and comprehensiveness, because the witnesses used in the collation are not readily available to researchers. Moreover, we do not know exactly how other canonical editions that were discovered in the 20th century compare to the Taishō edition and to what degree a further collation might be useful. This paper tries to answer some of these questions by expanding the Taishō apparatus of the Song Gaoseng Zhuan (T.2061) through collation with the Qisha canon and a categorization of the variations encountered. The combination of qualitative judgments about the variations together with quantitative data about their occurrence makes it possible to present a clearer picture of the relationship between these two versions of the text.

Keywords

Introduction

This paper is part of a project funded by the Chiang Ching-kuo Foundation for International Scholarly Exchange that aims at the creation of a new best

edition of the *Song Gaoseng Zhuan* 宋高僧傳 (SGSZ).[1] The new edition will be published digitally and in print. It will provide an index with biographical information of all persons mentioned, geo-reference all place names, and map all dates to the (proleptic) Gregorian calendar. The edition will also provide a better punctuated text, augmented by more than two thousand notes, and expand and improve the apparatus of collated variants.

Here we will be concerned only with the last of these items, the expansion and improvement of the SGSZ collation and apparatus. A discussion of the problems we encountered and the solutions we found might contribute to the development of methods that will be able to deal with complex variation between witnesses of Buddhist texts. It is partly known, partly assumed, that canonical editions of the Buddhist canon differ only in minor details. What counts as minor, however, depends on the task. Three wrong characters on a page might not be much when perusing a text for content, but might be significant in the context of translation.

Character variants and other differences between canonical editions have been collated at least since Sugi's *Collation Notes* (1247) on the 2nd edition of the *Tripiṭaka Koreana*.[2] The most influential canonical edition of the 20th century, the Taishō Edition, too provides some collation with Chinese editions of the Song, Yuan and Ming dynasties. We do not know how complete and how accurate this collation is, how it varies from text to text and moreover, as I will show below, we have reason to believe that the 'Taishō apparatus' did not originate with the Taishō edition. Be that as it may, the Taishō became for various reasons a convenient and stable reference point for research on Buddhist texts and its content is widely accepted as *textus receptus* in the field of Buddhist Studies. Textual scholars, however, who are interested in the history of these scriptures, rely on records of variations between different witnesses of a text. The last hundred years have seen momentous developments in the availability of Buddhist texts. Apart from 'new' texts that had not been included in the Taishō at all, we have now many more witnesses for almost all canonical texts, mainly because of:

---

1 Project members are John Kieschnick, Zhang Boyong, Hung Jen-jou and myself. I am grateful to Stefano Zacchetti and Christian Wittern, as well as to an anonymous reader, who has made valuable suggestions.

2 On the significance of his *Koryŏguk sinjo taejang kyojŏng pyŏllok* 高麗國新雕大藏校正別錄 see Buswell 2004.

1. The discovery and reprinting of canonical print editions that had not been available to the Taishō editors (especially the Qisha, Jin/Zhaocheng, and the Hongwu editions).
2. The discovery and reprinting of epigraphical witnesses for a text or text passages (especially the Fangshan Inscriptions)
3. The discovery and reprinting of manuscript witnesses (especially manuscripts from Dunhuang and other Central Asian sites, and manuscripts preserved in Japanese temples).

The situation is further complicated by the on-going, if uncoordinated, efforts at digitalization, which in fact create new (digital) editions with their own variants and annotations, the more fluid dynamics of which are not yet well understood.

We should ideally be able to describe the closeness or difference between witnesses of a text, and by extension between canonical collections. We should be able to say which edition is 'better' in terms of completeness and accuracy. However, we cannot even start to discuss this because we lack the methods to tackle the multiplicity of witnesses. The present paper is an attempt to work towards a methodology of collation that can provide quantitative and qualitative statements about the relationship of witnesses. We collate the CBETA/Taishō edition of the *Song Gaoseng Zhuan* with its Zhonghua/Qisha edition and try to answer questions that were hitherto difficult to approach.

Taishō Text and Apparatus

The *Taishō shinshū daizōkyō* 大正新脩大藏經 (1924-1934) is without doubt the most widely used edition of the Chinese Buddhist canon in the 20th century. In the field of Buddhist Studies the way to cite a passage by Taishō volume, page, row and line number, is still considered the 'canonical' form of reference even as the texts have moved into the digital age. Edited by Takakusu Junjirō, Watanabe Kaikyoku and others, the authoritative edition of the 20th century rearranged the content of the canon, thereby ending the thousand-year-old custom of placing Mahāyāna sutras first. In the Taishō edition early discourses (*āgama*s) and narrative literature (*jātaka*s) constitute the first volumes, reflecting the academic concerns of the editors to place the earlier strata of Buddhist texts first in the series. Another innovation was the inclusion of newly discovered Dunhuang material in volume 85.

The Taishō is a base-text edition, in other words it reproduces the texts as found in one single version of it, usually the 2nd printing of the *Tripiṭaka*

Koreana, without changes and records variant readings in an apparatus. The Taishō editors, however, were not the first to do so. An important precursor of the Taishō, which also used the *Tripiṭaka Koreana* as base-text, is the *Dainihon kōtei daizōkyō shukusatsu zōhon* 大日本校訂大蔵経 縮刷藏本 (abbr. *Shukusatsuban* 縮刷版) edited 1880-1885 by Shimada Bankon (also known as Mitsune) 島田蕃根 (1827-1907) *et al.*[3] Though the *Shukusatsuban* is hardly ever used today, it was the first edition of any Buddhist canon that was printed in modern movable type, a process that reduced the character variants as compared to manuscript or woodblock editions. It also was the *Shukusatsuban* editorial team that first decided to take the second printing of the *Tripiṭaka Koreana* as base-text, collate it with other print editions and record their findings systematically in an apparatus. Still another innovation that the *Shukusatsuban* introduced was punctuation in form of the *kutōten* 句読点 period symbol '.', here still in its intralinear form, i.e. a dot added to the right of the unbroken character line, and not yet in its interlinear form, the circle-period symbol 。, which is today used for the period symbol in Japanese and is generally of character-width.

The exact relationship of the Taishō annotation with that of the *Shukusatsuban* has, to my knowledge, not been documented, but it seems that the *Shukusatsuban* apparatus was rechecked, certainly expanded, in the Taishō edition.[4] A detailed study would require more extensive sampling than we are

---

3   The *Shukusatsuban* was published in 420 vols. at the newly established Kōkyō shoin 弘教書院. Revised 1935 as *Dainihon kōtei daizōkyō* 大日本校訂大藏經 （昭和再刊）. Tokyo: Shukusatsu daizōkyō kankyōkai 縮刷大藏經刊行會. Shimada started out as Tendai monk but returned to lay-life and became an official in the Home Ministry (*Naimushō* 内務省). He was obviously sufficiently well-connected to be allowed access to the rare editions of the canon that were used to build the apparatus. He worked with Fukuda Kyōkai 福田行誡 （1805-1888) and the monk Kaimyō of Shishidake 獅子岳快猛. Fukuda was a Buddhist scholar who actively resisted the Anti-Buddhist policy of the early Meiji government and for a while was head monk at the Zōjōji, where the collation was done. Kaimyō was the 383th abbot of the Kongōbuji 金剛峯寺 and an influential leader of the Shingon sect. He was also one of the founders of what later became Kōya University and presumably had access to well trained helpers, who were able to collate the texts. The collaboration of Shimada, Fukuda and Kaimyō resulted in the first modern Tripitaka edition. See Vita, 'Printings of the Buddhist "Canon" in modern Japan', 221, for more biographical and bibliographical information on Shimada.

4   Vita, 'Printings of the Buddhist "Canon" in modern Japan', 235, mentions that the collation of the Taishō was done in three centers: the Zōjōji, the Ueno Imperial Museum and the Kunaishō library. It would be interesting to know how exactly the team that worked in the Zōjōji collated the canon again, 40 years after the *Shukusatsuban* editors had used the very same copies.

able to do for now. Since we are concerned with the SGSZ, we will limit our comparison to only two samples from the Taishō/*Shukusatsuban* apparatus. The first sample contains all apparatus entries for the first fascicle of the SGSZ main text that begins with the biography of Yijing 義淨 the translator.

	Taishō (starting at T50n2061_p0710c01)		Shukusatsuban[5] (starting at Vol. 致四, p.69b)
1	復＝後【元】		复元作後
2	永＝唐【宋】【元】		永宋元俱作唐
3	語＝讀【宋】		語宋作讀
4	植＝殖【宋】【元】		植宋元俱作殖
5	隨＝降【宋】【元】		隨同作降[6]
6	虞＝虛【宋】【元】		虞宋元俱作虛
7	又＝人【元】		又元作人
8	幖＝標【宋】【元】		幖宋元俱作標
9	禳＝穰【宋】【元】		禳同作穰
10	囑＝屬【宋】【元】		囑宋元俱作屬
11	茶＝荼【宋】【元】[7]		�宋元俱作積 ○ 茶同作荼
12	頦＝積【宋】【元】		
13	空＝羅【宋】【元】		空同作羅
14	羅＝空【宋】【元】		羅同作空

As the comparison shows, the apparatus is in all respects identical. The Taishō edition does not add, expand or omit any of the entries in the *Shukusatsuban*. Our second sample is the complete apparatus of fascicle 15 that begins with the biography of Lingyi 靈一 the Vinaya Master.

---

5 We have used the copy in the Fu Ssu-nien 傅斯年 Library, Academia Sinica, Taipei.

6 On the same page the *Shukusatsuban* avoids repetition of the phrase x宋元俱作y by using x同作y.

7 This variant reading appears more than once in the fascicle and in the following is therefore marked with * in the main text. The *Shukusatsuban* does not number the apparatus but marks variant readings with a line to the left of the character ('lateral underlining'). Repeated variant readings are marked in the same way, but the explanation is not repeated.

	Taishō	Shukusatsuban
	(starting at T50n2061_p0799b01)	(starting at Vol. 致五, p.24b)
1	於＝于【宋】【元】	於同作于下同
2	主＝王【宋】【元】	主宋元俱作王
3	州＝洲【宋】【元】	州宋元俱作洲
4	丹＝舟【元】	丹元作舟
5	三＝二【宋】	三宋作二
6	植＝殖【宋】【元】	植宋元俱作殖次同
7	具＝且【宋】【元】	具宋元俱作且
8	綱＝網【宋】【元】	綱同作網 ○ 肯同作胥
9	肯＝胥【宋】【元】	
10	裝＝奘【宋】【元】	裝宋元俱作奘
11	悶＝悶【宋】【元】	悶宋元俱作悶
12	班＝斑【元】	班元作斑
13	羈＝羇【宋】【元】	羈宋元俱作羇
14	每＝既【宋】，＝無【元】	每宋作既元作無
15	採＝采【宋】【元】	採宋元俱作采
16	邪＝邗【宋】【元】	邪同作邗
17	庚＝庾【宋】【元】	庚同作庾
18	肆＝肆【宋】【元】	肆宋元俱作肆
19	班＝斑【元】	班元作斑
20	戒＝戎【宋】【元】	戒宋元俱作戎
21	場＝塲【宋】【元】	場宋元俱作塲
22	三＝二【宋】	三同作二
23	弘＝洪【宋】【元】	弘宋元俱作洪
24	二＝一【宋】	二宋作一
25	茶＝荼【宋】【元】	茶宋元俱作荼
26	三＝二【宋】【元】	三宋元俱作二
27	辯＝辨【宋】【元】	辯同作辨
28	拇＝栂【宋】【元】	拇同作栂

In our second sample, too, the apparatus is in all aspects identical and we might have to assume that the SGSZ apparatus has been lifted wholesale from the *Shukusatsuban*. However, in at least one instance, the *Shukusatsuban* apparatus differs from the Taishō.[8] Further research is needed to establish the

---

8 At *Shukusatsuban* vol. 致四, p. 94a // T50n2061 p0738c18, the *Shukusatsuban* has 佛宋元 俱作旋, while the Taisho asserts 佛＝旂【宋】【元】, which seems to make better sense

situation for other texts. The Taishō editors clearly added to the apparatus by collating single witnesses found in the Imperial Household Library (siglum: 宮), the Shōsōin Archive (siglum: 聖) and some temple libraries (these were individual manuscripts or prints, not complete editions). They also added Pāli parallels and names in the Āgama section of the canon. Nevertheless, it seems that most of the Taishō apparatus was actually first established and should be credited to Shimada Bankon and his team.

Let us now turn to the related question of verification. Whatever the exact provenance of an apparatus item found in the Taishō—and by extension in digital versions based on it—it is almost impossible for researchers to validate the accuracy or the completeness of the collation. The most widely used sigla are 宋, 元 and 明, which stand for the *Sixi zang* 思溪藏 (created c.1123-1175), the *Puning zang* 普寧藏 (1277-1290) and the *Jiaxing zang* 嘉興藏 (1589-1676), respectively. Neither the *Sixi zang* nor the *Puning zang* has been reprinted.[9] Of the *Jiaxing zang* only a partial reprint is widely available, which comprises merely those sutras that were not included in the Taishō, and which therefore cannot be used to verify the *Shukusatsuban*/Taishō collation for the siglum 明.[10] There are extremely few institutions that hold a complete *Sixi zang* or *Puning zang* and complete copies of the *Jiaxing zang* too are rare. Moreover, the question is which copy of the *Tripiṭaka Koreana* was used as base text for the *Shukusatsuban*.

According to a preface to the *Shukusatsuban,* the earliest collation efforts in Japan started in the early 18th century when a copy of a Ming canon was collated with a copy of the *Tripiṭaka Koreana* kept at the Kenninji 建仁寺 in Kyoto.[11] After permission had been laboriously secured, the work started in March/April 1706. It was accomplished by a team of ten collaborators, with each

---

in the context, but without being able to look at the originals it is hard to know what went wrong here.

9   Fortunately the *Sixi zang* will soon be reprinted from the National Library of China copy (personal communication from Mr. Li Jining).

10   Published in 40 volumes as *Mingban jiaxing dazangjing* 明版嘉興大藏經. Taipei: Xinwenfeng 新文豐, 1987. This edition has been widely distributed. A complete copy is in the National Central Library, Taipei, of which a microfilm is available for use at the library (MF 214.3 3987/21034) and in other collections in Taiwan, China, Japan and the US. Recently, a facsimile edition of 100 copies of the Jiaxing was published in Beijing in 2008 (Minzu chubanshe 民族出版社) (see http://www.tanzhesi.cn/news/html/?56.html (accessed Jan 2013)).

11   Vita, 'Printings of the Buddhist "Canon" in modern Japan', 223n8, says that Tetsugen's re-carving of the Ming edition, the so-called Ōbaku edition, was used in this collation. This is possible, since the Ōbaku edition is in the main a tracing of the *Jiaxing zang* and was thus referred to as the 'Ming Canon', but the foreword to the *Shukusatsuban* seems to

text being collated three times, each time by a different collator. The apparatus was handwritten on a copy of the Ming Canon which thereby 'became just like the *Tripiṭaka Koreana*.'[12] Unfortunately, the copy of the *Tripiṭaka Koreana* in the Kenninji 建仁寺 was mostly destroyed by fire in October 1837; of the ca. 5000 volumes only 419 were saved. After that disaster the only complete copy of the *Tripiṭaka Koreana* in Japan was to be found in the Zōjōji 增上寺 in Edo/Tokyo. The Zōjōji also held the copies of the *Sixi zang* (Song) and *Puning zang* (Yuan) that were used for the collation by the *Shukusatsuban* editors.[13] The preface states that the collation of the four editions (*Tripiṭaka Koreana*, Song, Yuan, Ming) took place in Tokyo between April 1880 and July 1885.[14]

We were not able to obtain a copy of the *Sixi zang* text of SGSZ yet and therefore cannot verify the collation strategy of the *Shukusatsuban* editors.[15] However, there is little reason to doubt that the editors closely followed the work-flow outlined in their 'Editorial Principles' (*fanli* 凡例).[16] According to the Editorial Principles, first the base text—usually from the Tripiṭaka Koreana—was copied and if necessary missing parts were supplemented from other editions. When a text was not part of the *Tripiṭaka Koreana* the Ming *Jiaxing* edition was used. In a second step, the texts were punctuated. This was the responsibility of an individual reader and results vary accordingly. The Editorial Principles clearly state that 'there was no single standard' for the punctuation (8a). Thirdly, the *Tripiṭaka Koreana*, the Song, Yuan and Ming editions were collated twice with each other. Lastly, a galley proof was done to make sure the printing was correct.

---

differentiate between Tetsugen's tracing and the original Ming Canon of which many sets were imported: Takakusu & Watanabe, *Shukusatsuban*, vol. 2, 439a-440a).

12 Here the preface to the *Shukusatsuban* refers to a copy of the 'Ming Canon' at that time preserved in the Hōnen Temple of Shishigatani, where Ninchō's collation project took place. Takakusu & Watanabe, *Shukusatsuban*, vol. 2, 439c.

13 All three copies of these editions were gifted to the Zōjōji by Tokugawa Ieyasu and are still kept there today. The Editorial Principles section (p.8a) of the *Shukusatsuban* states that missing volumes from the Yuan edition were borrowed from the Sensōji, another famous temple in Tōkyō, which is still there. Like the canonical editions of the Zōjōji, the Sensō ji *Puning zang* is designated an Important Cultural Property.

14 Takakusu & Watanabe, *Shukusatsuban*, vol. 2, 440a.

15 After work on this article was completed John Kieschnick managed to obtain a xerograph of the *Sixi zang* SGSZ, which will allow us to clarify this question in the future.

16 The Editorial Principles of the *Shukusatsuban* are found in the preface and catalogue volume (unnumbered) titled *Dainihon kōtei daizōkyō—mokuroku* 大日本校訂大藏經目錄, 6b-8a.

The collation was done in teams of four, each participant holding one of the four editions (three in the case of the SGSZ). In a first iteration the reader of the *Tripiṭaka Koreana* read the text out aloud in *on'yomi* 音読み style which follows the Sino-Japanese reading of the characters. A second iteration was done by reading the base text in *kundoku* 訓読 style with native Japanese readings. The two different readings discovered different variations. The *kundoku* reading would, for instance, catch a difference between 智 (*kundoku: chie*) and 知 (*kundoku: shiru*), which in *on'yomi* are both read *chi;* or 理 (*kundoku: suji*) and 利 (*kundoku: kiku*), which in *on'yomi* are both read *ri.* The *on'yomi* reading on the other hand could uncover variations like that between 至 (*on'yomi: shi*) and 到 (*on'yomi: tō*), which in *kundoku* are both read *itaru*; or 得 (*on'yomi: toku*) and 獲 (*on'yomi: kaku*), which in *kundoku* are both read *eru.* Quite an ingenious method, which involved collaboration, kept the participants awake and included double-checking. We moderns are humbled by the fact that at one time there were several teams of people who were able to enunciate the whole canon in two different pronunciations—no mean feat considering the many rare character variants in Buddhist scriptures.[17]

Since differences were recorded by attaching paper slips to the draft, this produced unwieldy manuscripts and the placement and content of the slips had to be checked separately. The galleys, too, were checked with the *kundoku/on'yomi* reading against the manuscript originally given to the printer. It is difficult to say how complete and accurate their effort was: only a detailed comparison of samples from the *Shukusatsuban*, the *Sixi zang*, the *Puning zang* and the *Jiaxing zang* could answer this. The workflow, however, was certainly well designed and it is highly unlikely that the Taishō editors some fifty years later felt the need to completely re-do or even double-check the collation done under Shimada Bankon.

Expanding the Apparatus

Improving an apparatus can be done in various ways. One way is to expand the apparatus by collating more witnesses (preferably from a different branch of the stemma). Another is to weigh or categorize the apparatus by commenting on the nature of the variants and trying to identify which readings are to be preferred. For our new edition of the SGSZ we attempt both. We expand

---

17   Vita, 'Printings of the Buddhist "Canon" in modern Japan', 226 and 235, mentions, however, that the *Shukusatsuban* editors had more difficulties in finding qualified collaborators than the Taishō team.

on the *Shukusatsuban*/Taishō apparatus by collating it with the Qisha and we categorize the variant readings in a experimental typology. In a digital context new information can be added more easily and if done consistently this information can be quantified in unprecedented ways as we will try to show below.[18]

The SGSZ belongs to a minority of texts in the *Shukusatsuban*/Taishō which were not included in the 2nd edition of the *Tripiṭaka Koreana*. According to the *Shōwa hōbō sōmokuroku*, the base-text for the Taishō edition is the Ming edition, i.e. the SGSZ as it appears in the *Jiaxing zang*.[19] This is the reason why the siglum 明 is not found in the apparatus; unlike most other texts in the *Shukusatsuban*/Taishō, the SGSZ is only collated with the Song (*Sixi zang*) and Yuan (*Puning zang*) witnesses. It is not entirely clear why the *Shukusatsuban* editors decided to base the text on the Ming edition instead of taking the much earlier *Sixi zang* version as base-text. To our knowledge, the *Sixi zang* contains the first canonical edition of the SGSZ. It was printed only 150-200 years after the SGSZ was written. Unfortunately, so far we have not been able to consult this edition directly: all we have are the references to it in the *Shukusatsuban*/Taishō apparatus. According to some catalogues, the SGSZ was first included in the slightly earlier *Pilu zang* 毗盧藏, but that is doubtful.[20] The *Kaibao zang* 開寶藏 edition (972-977)[21] was obviously too early to include the SGSZ, and the *Jin zang* 金藏 relies heavily on the *Kaibao* in form and content and therefore does not include the SGSZ either.

Next to the *Sixi* and the *Puning zang*, another early edition of the SGSZ is found in the *Qisha zang* 磧砂藏 (1234-1322).[22] This canon is available, the most widely distributed reprint being probably that by Shinwen Fong Publishing 新文豐出版社, Taipei. In the 1980s, Fan Xiangyong 范祥雍 used the *Qisha* to produce what is so far the best edition of the SGSZ (Fan *Song gaoseng zhuan*; abbr. Zhonghua shuju edition). Fan provides improved punctuation and text critical notes. His edition is again a base-text edition, this time of the Qisha witness of the SGSZ, which is in turn collated with the Taishō and its apparatus.

---

18   This comes at the price of stability and citablility, but these problems are not within the scope of this paper.

19   Takakusu & Watanabe, *Shukusatsuban*, vol. 1, 544.

20   Cai, *Ershiwu zhong zangjing mulu duizhao kaoshi*, 236, and Li & He, *Hanwen Fojiao dazangjing yanjiu*, 691, deny that the SGSZ is part of the *Pilu zang*.

21   Here I follow Chikusa's conclusions about the history of this edition (Chikusa, *Sō Gen bukkyō bunkashi kenkyū*, 315-318) and Zacchetti (*In Praise of the L*, 96-98).

22   It should be noted that the first *juan* of the Qisha SGSZ is lost. Available reprints have generally supplemented it with the first *juan* of the text from the Yongle edition.

Our interest is to create a better, critical edition in the sense that we do want to decide which variant readings are to be preferred and provide these in the main text, while moving less preferred readings in the apparatus. We rely on the CBETA 2009 XML/TEI edition of the Taishō canon, which has preserved the *Shukusatsuban*/Taishō apparatus in a standardized format. This enables us to adapt the data easily for our edition. As we revise the punctuation and annotation in light of the Zhonghua shuju edition and the original *Qisha zang*, we create an apparatus that references variant readings from five different sources:

1.   The Taishō main text, taken from the *Jiaxing zang* (our siglum: 大).
2 and 3.   The two witnesses referenced in the *Shukusatsuban*/Taishō apparatus as 宋 (*Sixi zang*) and 元 (*Puning zang*).
4.   The Qisha zang (our siglum: 磧)
5.   The Zhonghua edition by Fan Xiangyong (our siglum: 范).

In addition, we check the print edition of the Taishō, in the few places where the CBETA representation of the Taishō seems doubtful.

Categorizing the Apparatus Entries

All in all, the new digital edition contains 2189 apparatus entries.[23] Of these 851 were imported from the original Taishō edition (as represented in the electronic version done by CBETA), 1338 entries were added by us when collating with the Qisha. Of the 2189 entries 1751 are unique; this equals approximately nine entries per Taishō page. That means in practice that reading a Taishō page of the SGSZ, scholars encounter at least nine passages where different editions are at variance.

Collation picks up differences between witnesses and records them in an apparatus. These differences can be of various types. As is the case with editorial policy in general, designing a typology of variants for an apparatus is influenced by research questions and interests, and practices will vary from case to case. Often categories will be devised to help to decide whether a reading should be accepted as genuine or regarded as a later change. At times it

---

23   All figures here and below are based from dataset queries done in Jan 2013. Minor differences might occur in validating the figures with later versions of the dataset. In general, the exact numbers cited in this paper should not give the impression that the phenomena we describe can be fixed to this degree of accuracy. Depending on the interpretation of borderline cases absolute numbers vary slightly.
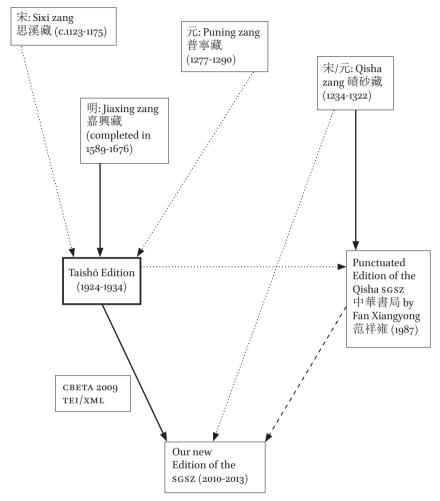
FIGURE 1   *The constellation of texts collated in the apparatus*

might not be possible to decide which reading is to be preferred, but one still might want to characterize the variant in one way or an other. The Editorial Principles of the *Shukusatsuban* edition, for instance, describe a number of common types of character variation. A 'Difference between orthodox and abbreviated' 正略之異 exists in cases where an abbreviated form is substituted for the 'correct' form, e.g. 跏 > 加, 號 > 号, 萬 > 万, 叡 > 睿. A 'Difference between new and old' 今古之異 supposes a diachronic difference in usage, e.g. 修 ↔ 脩, 荼 ↔ 茶. A 'Difference between orthodox and vernacular forms' 正俗之異 reflects on the fact that some variants vary according to

genres and text types, e.g. 剌 ↔ 剋, 鎖 ↔ 鏁. And some forms could simply be used interchangeably 互用之異, e.g. 繞 ↔ 遶, 徧 ↔ 遍.²⁴

In the context of our project, we are less interested in the study of character variants for their own sake, but are curious about the relationship of the Taishō to the Qisha edition, especially with regard to questions of accuracy and reliability. Does a collation with the Qisha significantly enhance the apparatus of the Taishō? Which edition contains more mistakes? Does the Taishō apparatus significantly enhance the Taishō text, by providing many preferred readings? To answer these and other questions we have developed a typology with six general variant types, only five of which we record as part of the apparatus.

Typology of Variants

A   *Not Recorded*

1.   Semantically equivalent common variant characters (*yitizi* 異體字), e.g. 勅/勑, 眾/衆, 黃/黄, are not recorded. Authority on what is a variant character in this sense is the MoE variant character dictionary 教育部異體字字典.²⁵

2.   Characters which are not variants in the narrow sense, but word variants (sometimes called *yitici* 異體詞),²⁶ and that are consistently used as semantic equivalents in the Taishō and the Qisha text are generally not recorded. E.g. 華=花, 囑=屬, 於=于, 耶=邪, 嘆=歎.²⁷

To be exact, we have maintained the Taishō annotation in both these cases (assigning a Type 0) and added the first instance of similar cases with regard to the Qisha, but did not include them in the apparatus for every instance. Labour-intensive manual markup of semantically significant variants

---

24   All examples from the Editorial Principles, 8a.

25   At http://140.111.1.40/main.htm. The Dictionary of Variant Characters edited and maintained by the Taiwanese Ministry of Education is the largest variant database available. Obviously the concept of a 'variant' 異體字 character vs an 'orthodox' 正字 character is problematic for some tasks and a purely descriptive approach is to be preferred for linguistic or semantic analysis (Galambos, *Orthography of early Chinese writing*, 11 *ff*). Our task, however, is to create a critical edition, which includes making the kind of choices that are informed by the prescriptive approach of character analysis.

26   The use of this term is contested (Galambos *Orthography of early Chinese writing*, 28).

27   This includes the de facto variant 已= 巳 and derivatives =圮.

(described below) might just be worth the effort, for a text of the length of the sgsz (30 *juan*); however, to record common (calli-)graphic, but semantically irrelevant differences is too costly to be realized in this fashion. Perhaps a 'complete' apparatus in this regard would provide an even better basis to trace filiation between editions, but a number of factors make this unlikely. First, digital full-text editions are limited by the Unicode character set. Without fonting non-Unicode characters and assigning them to private-use area codepoints (a procedure that involves other trade-offs), producers of digital editions of longer pre-modern Chinese Buddhist texts are always forced to regularize to a certain extent. Secondly, without unlimited resources a 'complete' collation which uses markup to records all graphical variation is not feasible beyond a certain text-length. Manual markup is expensive and this limits the scope of what can be done with this technology. In a perfect world we would have perfect digital editions of all printed editions of the canon and could compare them algorithmically, but it is hard to see that digitizing all editions would be worth the effort. Thirdly, it is not clear that semantically not-significant variants can indeed help to trace filiation. Certainly the decision whether a woodblock has been re-carved or otherwise modeled on another edition must be based on a comparison of facsimiles not digital text. One also might want to try technological solutions beyond XML that allow different forms of modeling a Chinese text.[28]

B   *Recorded*

1.   Mistakes or non-preferred readings in the Taishō edition.
Type 1a is a mistake or non-preferred reading in the Taishō main text. Such cases can generally be assumed to reflect the *Jiaxing zang* reading, though there could be mistakes that were introduced in the creation of the Taishō. In Type 1a the preferred reading is found in the Taishō apparatus.

1a makes no statement about the appearance of the text in the Qisha, however, though the Qisha often supports the Song reading against the Taishō in these cases. The existing Taishō apparatus has precedence here and these cases therefore are recorded as Type 1a, not 3b or 3c.

Mistakes are, for instance, scribal errors such as 水 for 永 (T.50n2061p0719b08), or the confusion of two names (空 and 羅, T50n2061p0713c17). Non-preferred readings are regularizations such as those that are connected to language change (何 for 胡, T50n2061p0722a07), political change (宋 for 大宋, T50n2061p0710b02), (dubious) clarifications (念佛 for 念, T50n2061p0738c18),

---

28   As outlined for instance in Wittern, 'Digital editions of premodern Chinese texts'.

additions (傳一 for 傳, T50n2061p0738b14), or deletions (相 for 相去聲, T50n2061p0753a14).

Type 1b is a mistake or non-preferred reading in the Taishō main text, which is not annotated in the Taishō apparatus, but that has been newly identified in our collation with the Qisha edition.

Type 1c records an apparatus entry where the alternative reading given in the Taishō apparatus should be considered a mistake or non-preferred reading, and the character in the main text of the Taishō is considered correct. 1c makes no statement about the appearance of the text in the Qisha, though in most cases the Qisha here agrees with the Taishō against the Song or the Yuan witness (e.g. Taishō and Qisha 此 against Song比, T50n2061p0718b15). As mentioned above, the Song and Yuan witnesses cannot be ascertained easily and it is difficult to judge the quality of these editions from the apparatus.

2.   Mistakes or non-preferred readings in the Qisha edition.

Type 2a is a mistake or non-preferred reading in the Qisha edition that was already identified by Fan Xiangyong in the apparatus to his Zhonghua shuju edition.

Type 2b is a mistake or non-preferred reading that has been newly identified in our collation of the Qisha with the Taishō text. The fact that, in spite of Fan's careful edition, we were able to find still more mistakes in the Qisha shows how difficult it is to achieve complete collation manually.

3.   Type 3 groups three types of variations.

Type 3a marks the rare cases where variant characters encode additional grammatical information. This is only observed in three cases in the Qisha and its derivative the Southern Hungwu edition. The Qisha, for example, differentiates between 禮 and 礼.

Type 3b marks character variations in compound words where we were unable to decide on a preferred reading. It is, for example, not clear if Zanning transcribed Sanskrit *skandha* as 犍度 (Taishō) or 揵度 (Qisha) at T.50n2061p0711a08.

Type 3c marks single character variations and characters variants where we were unable to decide on a preferred reading. It is, for instance, not possible to decide if Daoyin 道氤 was interred on the 12th (Taishō and Qisha edition) or the 13th (Song and Yuan edition) day of the eight year of Kaiyuan 28.[29] Similarly, it is not clear if Zanning wrote 爾 (Taishō) or 你 (Qisha) at T.50n2061p0821c20.

---

29   T.50n2061p0735a24.

In the main text of the print edition we will follow the Taishō in all Type 3 cases, where there is a 'draw' between the Taishō and the Qisha or another edition. If the Qisha, however, agrees with the Song edition against the Taishō we will use the Qisha reading, for it is good practice to follow the earlier editions in case where the reading is confirmed by more than one edition. For example, the transcription of Khotan as 于闐 in the Taishō, as against 于遁 in the Song, Yuan and Qisha editions, or the reading 茶毘 in all three early editions against the Taishō 茶毘.

4.   **Type 4** groups other rare cases, which do not fit in the above categories, but seem noteworthy, e.g. mistakes or corrections in the Zhonghua-shuju edition.

In a digital edition a categorized apparatus adds greatly to the potential of the edition. It is possible to search and output variants as needed in the main text or in the notes, in principle showing different versions of the text from a single master file. Moreover, because in a digital edition the apparatus is recorded in a standardized fashion it is possible to query the apparatus and gain quantitative information about the quality of editions. Querying the dataset allows us now to answer the following questions:

1.   How large is the apparatus?
The total number of items in the apparatus of our edition is 2189. This comprises 851 entries that had previously been collated in the Taishō. 1338 entries were added by us when compared with the Qisha. This information equals about 9 entries per Taishō page.

2.   How many mistakes/non-preferred readings are in the main text of the Taishō?
The main text of the SGSZ Taishō version contains c. 580 mistakes or non-preferred readings, i.e. about three mistakes or non-preferred readings per page.

3.   How many of them are detectable in the Taishō annotation?
About half of all mistakes or non-preferred readings are detectable in the sense that the correct or preferred reading is contained in the Taishō annotations.

4.   Was the new collation with the Qisha worthwhile?
Another c. 47% of mistakes was detected by the collation with the Qisha edition. In other words, researchers that use the new edition will be able to avoid twice as many mistakes compared to those who work with the Taishō version only.

The new collation was quite successful relative to the size of the existing apparatus. It almost doubled the number of mistakes and non-preferred readings that can now be identified. As for the actual usefulness for the purpose of research, the question cannot, of course, be answered quantitatively, as it depends on the work that the researchers want to accomplish and the differences in the gravity of the mistakes.

5.   Between the Taishō and the Qisha, which is the more accurate edition?

The Qisha has a slightly higher overall mistake count than the Taishō (629 vs. 579). Users of the Qisha edition are likely to encounter about 3.4 mistakes per Qisha page.

6.   What is the relationship of our new digital edition to the printed edition of Fan Xiangyong?

Fan Xiangyong's careful collation of the Qisha with the Taishō had revealed already c. 460 mistakes in the Qisha.[30] Our collation added another 160 instances where the Qisha has a non-preferred reading when compared to the Taishō. In spite of his careful work, therefore, a certain number of mistakes escaped Fan's attention. This reminds us that further checking and expanding on previous work can still be fruitful and one wonders how many more variant readings could be discovered if we had the full text of the Song and Yuan editions.

7.   The Qisha is genetically related to the Song edition. Can it yield new readings?

We do have an outline of the genetic relationship between different canonical editions, but we do not know what this 'relatedness' means in practice. Of the c. 760 apparatus entries which contain both a Qisha and a Song reading, only c. 520 are identical. That is, in almost a third of the cases (c. 230) the Qisha disagrees with the Song. The sample is too small (and in the case of the Song edition second-hand) to reach any conclusions as to the overall relationship of the editions, but it is a reminder that being 'related' does not mean being 'identical' and that a collation of witnesses from the same branch of the stemma can still be useful. It should be kept in mind that these relationships might vary from text to text. Further research is needed to understand whether, for instance, the figure of three mistakes/non-preferred readings per Taishō page (half of which are discoverable through the Taishō annotation) holds true for other

---

30   Next to the Taishō Fan Xiangyong also included a seperately printed Ming dynasty sᴳsᴢ
      edition ('Yang edition' 揚州本) in his collation with the Qisha.

texts as well, or if it is characteristic only for the sgsz, which after all belongs to a minority of Taishō texts that are not based on the *Tripiṭaka Koreana*.

Technical Realization

In a digital environment the management and the collation of different versions of a text can be realized in various ways. Computational linguistics has long since developed tools and techniques to collate and compare different versions of a digital text. From the unix command line tool *diff* (1974), via Peter Robinson's *Collate* (1989) and its xml version *CollateX*, to the online platform Juxta Commons (http://juxtacommons.org) (since 2009) the basic processes of tokenization, collation, alignment, and output have allowed programmers to version-control software as well as scholars in the Humanities to compare witnesses of ancient texts. For texts marked-up with tei one of the most comprehensive attempts has been the Versioning Machine (http://v-machine.org) developed by Susan Schreibman *et al.* In our case, different digital versions are not available and the creation of the apparatus is therefore merely an expansion of the Taishō apparatus. Fortunately, the Taishō edition has been digitalized including its apparatus and been made available in tei by cbeta. This is a good example of the value of marked-up text in the humanities. Once the structure of a text has been modeled it can easily be adapted and expanded by other projects. By way of example here is the original cbeta-tei entry (Ver. 2009) of a variant reading in the passage at T.50n2061p710c18 where the Taishō has (mistakenly) 永 for 唐 as attested in the Song and Yuan witnesses:

```
<app from="beg0710004" to="end0710004">
  <lem wit="【大】">永</lem>
  <rdg wit="【宋】【元】">唐</rdg>
</app>
```

In our edition this entries becomes:

```
<app from="beg0710004" ana="appType1a">
  <lem wit="【大】">永</lem>
  <rdg resp="Taisho" wit="【宋】【元】">唐</rdg>
  <rdg resp="sgsz" wit="【范】【磧】">唐</rdg>
</app>
```

The "resp" attribute records the responsibility for the collation, the "ana" records the type according to the typology outlined above. A long list of such entries constitutes the apparatus database for our edition. Once available in standardized, computable form the list can be queried, output and shared with much greater ease than in a print format.

Conclusion

It is hard to imagine that future critical editions of ancient texts will be mainly produced in print. As academic editions move into the digital medium, certain aspects of print editions such as annotation, index and apparatus can be represented in more efficient ways. Common standards for critical editions, however, are needed in the digital as well as in the print world, where for example the Leiden Conventions for critical editions of epigraphy established a norm that was successfully adopted. Through the use of international standards such as TEI, text and associated data are rendered computable and sharable. However, while the technical realization of digital edition projects is well understood, the economy of apparatus creation is not. Whether the efforts to improve the apparatus of any given Buddhist text are worthwhile cannot be easily answered. When it comes to the SGSZ, we have seen that even although the *Shukusatsuban*/Taishō editions have already been carefully collated with previous versions, a new collation with the Qisha edition still discovered on average 1.5 previously unknown mistakes or non-preferred readings per page. If that is enough to justify the labour of collation depends on the economic parameters of the project and the research questions of the investigators. There is certainly still room for improvement, however.

Bibliography

Buswell, Robert E. 'Sugi's collation notes to the Koryō Buddhist Canon and their significance for Buddhist textual criticism', *Journal of Korean studies* 9 (2004): 129-184.

Cai, Yunchen 蔡運辰. *Ershiwu zhong zangjing mulu duizhao kaoshi* 二十五種藏經目錄對照考釋. (Taipei: Xinwenfeng, 1983).

Chikusa, Masaaki 竺沙雅章. *Sō Gen bukkyō bunkashi kenkyū* 宋元佛教文化史研究. (Kyūko Shoin, 2000).

Fan, Xiangyong 范祥雍 (ed.). *Song gaoseng zhuan* 宋高僧傳. Beijing: Zhonghua, 1986).

Galambos, Imre. *Orthography of early Chinese writing: evidence from newly excavated manuscripts*. (Budapest: Department of East Asian Studies, Eötvös Loránd University, 2006).

Li, Fuhua 李富华; He, Mei 何梅. *Hanwen Fojiao dazangjing yanjiu* 汉文佛教大藏经研究. (Beijing: Zongjiao wenhua, 2003).

Takakusu, Junjirō 高楠順次郎 and Watanabe Kaikyoku 渡邊海旭, eds. *Showa hōbō sōmokuroku* 昭和法寶總目錄. (Taishō shinshū daizōkyō kankōkai, 1929. 3 vols.

Vita, Silvio. 2003. 'Printings of the Buddhist "Canon" in modern Japan', in Giovanni Verardi and Silvio Vita, eds., *Buddhist Asia 1. Papers from the First Conference of Buddhist Studies held in Naples in May 2001.* (Kyoto: Italian School of East Asian Studies, 2003), pp. 217-245.

Wittern, Christian. 'Digital editions of premodern Chinese texts: methods and problems—exemplified using the *Daozang Jiyao* 道藏輯要', *Chung-Hwa Buddhist journal* 25 (2012): 167-194.

Zacchetti, Stefano. 2005. *In Praise of the Light—A critical synoptic edition with an annotated translation of chapters 1-3 of Dharmarakṣa's* Guang zan jing 光贊經, *being the earliest Chinese translation of the Larger Prajñāpāramitā.* (Tokyo: The International Research Institute for Advanced Buddhology Soka University, 2005). (Bibliotheca Philologica et Philosophica Buddhica VIII).