

SAT Response to “The Issue of Multiple Variants in the Buddhist Scriptures of the SAT Project” (IRGN2373)

Date: 2019-05-14

We have recognized that, in the document IRGN2373 by TCA experts, concerns are expressed on authenticity of some characters in our proposal, accuracy of our research work, and congruity of our encoding policy, as we understand. We would like to take this opportunity to reclarify our principles and facilitate IRG experts’ comprehension in order to make sure the consensus of IRG experts.

Due to a limited time frame until the IRG meeting, we have to omit some rigid designation of reference information for the time being. They will be amended or elucidated as necessary.

1. Points of IRGN2373

As we observe, the central arguments of IRGN2373 are:

- (1) SAT’s proposal is based on Taishō Tripitaka 大正新脩大藏經 (大正藏), whose accuracy in the text is doubted by many scholars
- (2) SAT’s proposed glyphs from *Yiqiejing Yinyi* 一切經音義 do not respect existing works on textual criticism
- (3) SAT’s submission has not excluded characters known to be incorrect in *Yiqiejing Yinyi*

2. Our position

2.1. Taishō Tripitaka as the source

As for the point (1) in the previous section, it is known since the end of last century that Taishō Tripitaka contains and may contain various textual, critical, and collation errors. However, the importance we acknowledge of Taishō does not chiefly lie in its scale or infallibility, but the fact that it is the current worldwide *standard text in use* for the Chinese-script Buddhist canon. It has been clarified in our initial letter to IRG (IRGN1858) as follows:

As editions of the Taishō were printed in large numbers, the Daizōkyō was rapidly disseminated, especially to libraries all over the world. Thus we have arrived to today's state of affairs where almost everyone in the field of Buddhist studies working in Sinitic sources is reading the same text.

The influence of Taishō in Buddhist studies is extraordinary. We regularly refer to a passage of scripture using the index, volume, or page number from Taishō in academic literatures, as how we indicate a Biblical verse using a common index. The practice is so widespread that even though Taishō is largely based on the text of renowned Tripitaka Koreana 高麗大藏經 (which was published as woodcut printings in the 13th century), most modern Korean scholars use Taishō’s numbers for reference, as well. Electronic platforms such as CBETA are no exception.

On building a text database serving academic community, conformity to Taishō is not only the best

choice but an imperative drawn from the Buddhist scholarly discipline, that warrants the project meaningful.

2.2. Attitude towards the text

Regarding the point (2), IRGN2373 has stated the possibility that SAT's text of *Yiqiejing Yinyi* is not adequately "collated" or "proofread", which we think means *canonicalized*, compared to works such as Prof. Xu Shiyi 徐時儀's 一切經音義三種校本合刊. The argument unfortunately seems to be caused by confusion in multiple aspects, namely:

- (i) The aim and design of transcription between Xu's work and SAT database
- (ii) Homonymous books by Xuanying 玄應 and Huilin 慧琳
- (iii) Genealogy of surviving witnesses of Huilin's *Yiqiejing Yinyi*

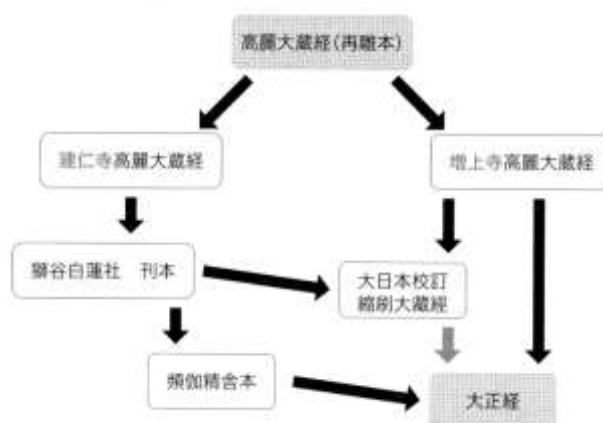
On (i), critical editions are usually categorized into several types. Plainly speaking, our project aims to precisely reproduce the original text, and Xu's 一切經音義三種校本合刊 seeks to establish the best reconstruction of the text. The latter method naturally requires a great amount of interpretation. Our mission is to make a digital equivalent of the paper Taishō to serve other researchers (you can call it *diplomatic edition*), so we should avoid adding our own guesswork into the text. We need to be faithful because perceived gap between the source text and its digitalization can undermine the legitimacy of not only our database but study of whoever cites our contents. We provide a platform for researchers across the world. This point is also explained in the IRGN1858 Appendix: "*the number users who read these ideographs daily is quite large--probably around 5,000 or so--distributed around the world*". In this light, any unprecedented glyphs should be theoretically qualified for encoding (as long as IRG's unification criteria allow), with some reservations:

- Obscure printing in Taishō: in this case, since the original matrices of Taishō is lost, we try to consult other witnesses for more legible skeleton of glyph to use.
- Accidental mutation in Taishō: if Taishō has a peculiar glyph not known in any original witness (books Taishō used in collation are all specified and mostly available now), we try to corroborate representative glyph using the witnesses unless the gap seems unrecoverable, that may result in readers' confusion.

Next, (ii) is indeed confusing, but we have two books with the name *Yiqiejing Yinyi*. The first one is so-called the 25-fascicle book by Xuanying who is a contemporary of Xuanzang 玄奘, and the other is the 100-fascicle book written a century later by Huilin. The former has been traditionally passed down in mainland China with many versions as the only *Yiqiejing Yinyi*, but the volume is relatively small. The latter has long strayed outside China being unknown to Chinese scholars until the 19th century, and only has several versions ever survived. Huilin's book is formally an enlarged version of Xuanying's work, but is actually a total rework under his own direction. *Yiqiejing Yinyi* in SAT's context only mean Huilin's book, as it is the only one included in Taishō (with a sequel by Xilin 希

麟). We could not automatically apply the result of critical analysis targeting Xuanying's book.

Similarly, a little explanation is needed on (iii). Huilin's *Yiqiejing Yinyi* has very limited versions available today, and all of them eventually goes back to Tripitaka Koreana, which is a woodblock printing. So, why do we not just treat Korean version the only meaningful text? In fact, its wood matrices have been used for centuries to produce numerous imprints, and need repair when worn down. Multiple copies were imported to Japan in varied timings. *Yiqiejing Yinyi* in Taishō is said to be a text collated between Korean version and Pinjia Jingshe 頻伽精舍 (Shanghai) edition, which is a modern revision and reprint. That means it is a collation between two distinct lineages of imprints, one of which has been lost. Thus, even if we run into an unusual glyph seemingly is a quirk of Pinjia, we cannot exclude the possibility that it originated in the condition of an older copy of Korean version.



Also mentioned is the reason why SAT wants to *encode* those raw characters unlike CBETA. Actually, CBETA is not unconcerned about unencoded characters. CBETA has had a composite character description convention functionally alike to IDS. Moreover, they are actively incorporating technical update including new SAT characters in the Extension F. CBETA announced they included 1,388 characters from the Unicode 10.0 in its new version on the release. It is explained by methodological difference as well as historical, as CBETA was founded later than SAT (contrary to IRGN2373's description).

In pursuit of digital text format that can sustain various scholarly applications, we do not consider it sufficient only being represented in a human-readable form, but also processible, consistent, unambiguous, and as self-contained as possible, in order to maximize analyzability and sustainability which leads to reliability. Our commitment to Unicode, which replacing our previous attempts described in IRGN1858, comes from our trust in the platform of Unicode that provides us with a universal basis roughly equal to “plain text technology” in their term.

2.3. Vulgar characters in history

As for the point (3), IRGN2373 has referred to USAT05013 as a case of incorrect character unsuitable

for encoding, because it is marked with 非也 “is wrong” in the original description. However, this kind of terms appear in *Yiqiejing Yinyi* cannot be naively taken as a rationale of not being valid character from modern perspective. For example:

輕奕 乳兗反考聲云奕弱也韻英云柔也說文從而從大……經文作軟非也竝無此字也
瘖者 鴟賈反考聲云不能言也……經從口作啞非也音厄笑聲也非經義……

(Book 1; Taishō vol. 54, p. 314)

Those entries reject 軟 and 啞 as incorrect characters for “soft” and “dumb” respectively in favor of older variants, even though we know they are the most normal choice to write those words. Huilin’s *Yiqiejing Yinyi* is a work deeply colored by his own archaism and prescriptivism, a not very uncommon attitude for ancient lexicographers. Ironically, it provides present-day researchers with a valuable insight into writing convention in his times, by presenting the very evidence he saw “wrong” characters was used. Of course, there is no guarantee that every such character was as conventional at that time as 軟 or 啞 is today, but according to the nature of this literature, marking like 非 and 俗 “vulgar” is considered rather positive, at least not negative, material for encoding the characters; if we believe the author, the character has ever been *actually used*. There may be opinions from some experts that semantic or systematic variants are inherently unifiable besides the UCV, but we are not aware that they are the consensus of IRG for the moment.

By conjecture, what the authors of IRGN2373 are concerning may be whether there is qualitative difference in value between characters in SAT’s sources and ones in materials like 佛教難字字典. In our view, they ideally do not differ essentially, except that the latter lack credible evidences trace them back to the primary resources. While we try to attest our characters in original witnesses, editors of those materials choose not to show their exact origin by photocopy or citation, but to substitute newly reinterpreted shapes for them. This method is highly fallible when we discuss the shape itself, regardless of authority and other academic value. Even characters in Hanyu Da Zidian are repelled during recent discussion in IRG.

3. Final remarks

We are conducting encoding activity on the assumption that our policies aforementioned are affirmed by or conformant with the current IRG and WG2 practice. In case not, we are ready to reconsider our workflow, so kind advice from experts is greatly appreciated. We apologize for unnecessary confusion that may have happened, as the implicit context would grow more and more complex as the studies continue specializing. In any case, we hope experts to understand the situation, as we will not forget being accountable for our submitted characters.

Note that we do not make essential distinction between encoding styles being single code point or IVS, as long as they enable us to identify and utilize our characters in the same way. On the other hand, we would not favor (a) unification that is semantically unfounded lest it confuse our users, since an IVS

has strong connection to its base character by design, and (b) inconsistency in IVS assignment, especially the situation that a standalone USAT source character has cognate or identically-shaped ones in IVD or vice versa, because that would lead to unagreed representations of a character which appears in our database.

(End of document)