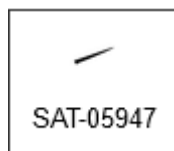


# SAT Feedback to IRG WS 2021 v4.0 (IRGN2581): Frequently asked questions about WS2021-00001

Date: 2023-03-24



## 1. Introduction

This document is recompilation and expansion of responses we made on IRG ORT or at the meeting to several questions from reviewers. We are fully aware that WS2021-00001 (SAT-05947), consists of only a simple stroke, is confusable with existing CJK Strokes. However, we still maintain the position that this character should be encoded as a CJK Ideograph, not CJK Stroke, at least under the current technical framework. We also believe that we have enough rationale to support this argument.

## 2. Are you trying to encode a stroke?

No. We are trying to encode a Han character, not exploring the possibility or the model to encode a stroke in a CJK Ideographs block.

## 3. Why not just encode as a Stroke rather than an Ideograph?

Conceptually, the strokes and the characters are concepts from different layers. Any Regular script (楷書) component is visually a geometric combination of one or more strokes, but cannot directly reduce to set of strokes i.e. as if strokes are subdivision of a component, because strokes do not embrace the positional information need to make a component, which components hold. Meanwhile, most components can be an independent character or be freely composed using restricted system of arrangement (like IDS), with exception of small portion of always dependent components (e.g., 亅, 卩 etc.). Components can be associated with a basic meaning unlike strokes. The contrast between *stroke* and *component/character* is analogous to *morpheme* and *lexeme/constituent* in linguistics, or the concept *radical* and (often confused) “component carrying the radical” (radical is an abstract concept, not always visually manifested, though sometimes transparently recognizable from the structure of a character).

Technically, as [IRGN2581EAFeedback](#) states, Unicode’s CJK Strokes and CJK Ideographs have different property values assigned. The most striking difference might be the value of the `Script` property, where CJK Strokes are `Common`, while CJK Ideographs are `Han`. Although the said document suggests using `Script_Extensions`, it does not provide the usability equal to ordinary CJK Ideographs. `Script_Extensions` is a multiple value property that enables a character to have every

identity which it is used among. As a result, even most of the East Asian punctuation has `Script_Extensions=Hani` (example below), that makes it impossible to e.g., include a character when performing a search that only extracts Han characters using the property. This is the main reason we want to avoid allocating what we think is a Han character as a CJK Stroke.

```
30FB ; Bopo Hang Hani Hira Kana Yiii # Po KATAKANA MIDDLE DOT
```

We think that the character's identity is analogous to 亻 for 人, or ++ for . Currently, Han characters and dependent components are all reside in CJK Ideographs blocks, but if IRG believes that the character is not suitable for the main blocks, *CJK Unified Ideographs Supplement* (as described in [IRGN2492](#), [EisoFeedback](#), and [L2/21-173R](#)) would be a better place to accommodate it, with full usability as CJK Ideograph warranted.

#### 4. Is this really a Han character?

We believe for the following reasons that WS2021-00001 is a Han character, or at least a component (UCS currently make no distinction between components and independent Han characters):

- the character has its pronunciation recorded
- the character is proven to be a variant of what is recognized as a Han character
- the character information matches data in a surviving edition of a dictionary
- the description of this character in our evidence contains no accidental error

We will clarify these points below.

#### 5. What does it mean if this is a Han character?

It appears in 慧琳音義 as a quotation from 說文解字 that explains the character 子:

「無右臂。從了，象形。▲，聲也。▲音厥。」(卷六十) → Fig. 1

「無右臂。從了，▲聲。▲音厥。」(卷八十五) → Fig. 2

We can easily find parallel text in the current edition of 說文解字:

「子：無右臂也。从了，丿。象形。經節反。」(小徐本) → Fig. 3

「子：無右臂也。从了，丿。象形。居桀切。」(大徐本) → Fig. 5

Thus, WS2021-00001 can be identified as the 丿-shaped glyph in the modern 說文解字. At this moment, it is not immediately clear what the latter glyph stands for, but fortunately 小徐本 has the reverse pointer (Fig. 4):

「丿：鉤識也。從反丿。讀若覈。臣鍇曰：鉤柄之表識。子[in seal script]、戊從此。俱越反。」

Therefore, if we believe 徐鍇's annotation, this is a variant of 丿 (U+2010C).

Note that 慧琳 was not possible to refer to this comment by 徐鍇 because his work is centuries older than 小徐本. The coincidence suggests that such association between 子 and 丿 might have a common origin prior to Song dynasty.

## 6. Didn't 慧琳音義 mistake it with another character?

慧琳音義 states that WS2021-00001 is read as the character 厥 (See Fig. 1). 厥's usual pronunciation in Middle Chinese is 臻合三入月見 (Wang Li \*kīwət; 厥小韻). When we go back to 說文解字 (Fig. 3), it says:

「子：無右臂也。从了，丿。象形。經節反。ノ子：無左臂。象形。从了，丿聲。俱越反。」  
Here, 經節反 indicates 山開三入薛(B)見 (Wang Li \*kīet; 子小韻) and 俱越反 is 臻合三入月見, which is the same syllable with that of 厥. Although the pronunciations in 說文解字 is given for the head character rather than the component, it matches more with 子. With this glyph shape, isn't it that 慧琳 confused the “left hook” in 子 and the “right hook” in 子?

However, just as the previous point, 說文解字 (小徐本) answers it by itself (Fig. 4):

「丿：鉤逆者謂之丿。象形。凡丿之屬皆從丿。讀若槩。臣錯曰：鉤喙之曲芑。今日逆須。子 [in seal script] 從此。瞿月反。ノ

丿：鉤識也。從反丿。讀若翳。臣錯曰：鉤柄之表識。子 [in seal script]、戍從此。俱越反。」  
It clearly states that 丿, which is a part of 子, is pronounced 瞿月反 = 臻合三入月羣 (Wang Li \*gīwət; 厥小韻) whereas 丿, of 子, is 俱越反. Therefore, it is proved that both 子 and 子 just have pronunciation differing from their phonetic component according to 說文解字. Unmatching pronunciation between the phonetic component and the character is a common situation in Chinese, and it has no discrepancy with the description in 慧琳音義 either, if we regard WS2021-00001 as 丿.

In any doubt, 大徐本 (Fig. 6) coincides with 小徐本 on this matter: 丿：衢月切 (= 瞿月反) and 丿：居月切 (= 俱越反).

## 7. Doesn't the description contain misspelling?

Below are occurrences of WS2021-00001 in 慧琳音義:

「無右臂。從了，象形。▲，聲也。▲音厥。」(卷六十) → Fig. 1

「無右臂。從了，▲聲。▲音厥。」(卷八十五) → Fig. 2

Among which, the former passage does have difficulty for understanding. 《一切經音義三種校本合刊》 considers it omission of one character:

「無右臂。從了，〔丿〕象形。ノ聲也。ノ音厥。」

At this moment we cannot judge its validity, but one might still be suspicious of why the term 象形 cooccurs with ~聲 (which suggests 形聲 composition). Actually, it is widely known about 慧琳's preference in explain characters using 六書 concepts, and his often markedly different usage of some terms from those under today's consensus. Numerous studies have been done regarding his cryptic use of, mostly, 省聲, 轉注, and 假借. You can get the rough outline on this matter in:

- 徐时仪《玄应和慧琳〈一切经音义〉研究》, p. 227
- 王華權《〈一切經音義〉文字研究》, p. 421

- 黄仁瑄《慧琳〈一切经音义〉中的转注兼会意字》

At any rate, coexistence of 象形 and 形聲 is easy to find throughout 慧琳音義, such as:

- 「說文鑿中火主也象形從、聲」(卷三十一)  
cf. 大徐本「主：鑿中火主也。从■ (𠂇口土)，象形。从、，、亦聲。」
- 「說文從尸彳久[攴]舟象形尸聲」(卷三十二)  
cf. 大徐本「履：足所依也。从尸从彳从攴，舟象履形。一曰尸聲。」
- 「盾所以扞身蔽目也以自蔽從十目象形厂聲」(卷六十七)  
cf. 大徐本「盾：戲也。所以扞身蔽目。象形。」

The last example is somewhat interesting, as 《說文解字注》 has the following annotation on this entry:

「从目。(各本少二字。今依玄應補。)象形。(錯曰：斤象盾形。按今錯本或妄增厂聲二字。)」

Contrary to the speculation by 段玉裁 that 「厂聲」 in 小徐本 (which he then saw) might have been 徐鍇's invention, the coincidence in 慧琳音義 possibly suggests the existence of such convention in the 說文解字 versions prior to Song dynasty.

With above, we could say that the questioned text is also possible to be understood as a part of general convention of 慧琳音義 as it is.

## 8. Isn't your example just a random writing error?

慧琳音義 only has two entries that explain 子's structure, and the two both agree to have this shape.

## 9. Is the glyph alone enough to support such meaning?

Maybe not. We do not have intention to monopolize a code point for this meaning, and IRG PnP principle (§2.1.3) stipulates unification of identical glyph with different meanings. The problem is that we current do not have a character with such structure which works as CJK Ideograph.

(End of document)

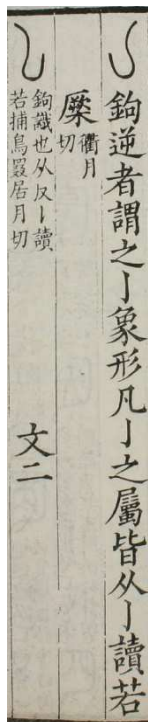
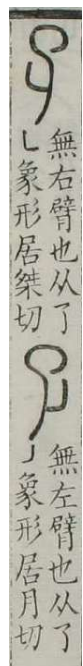
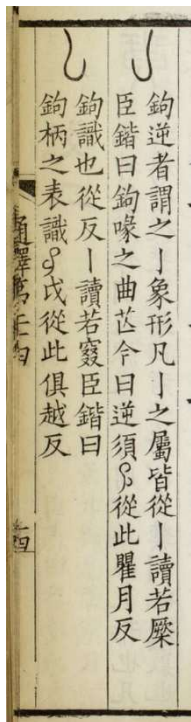
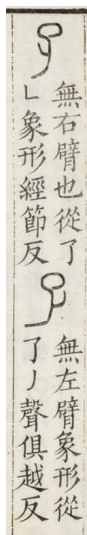
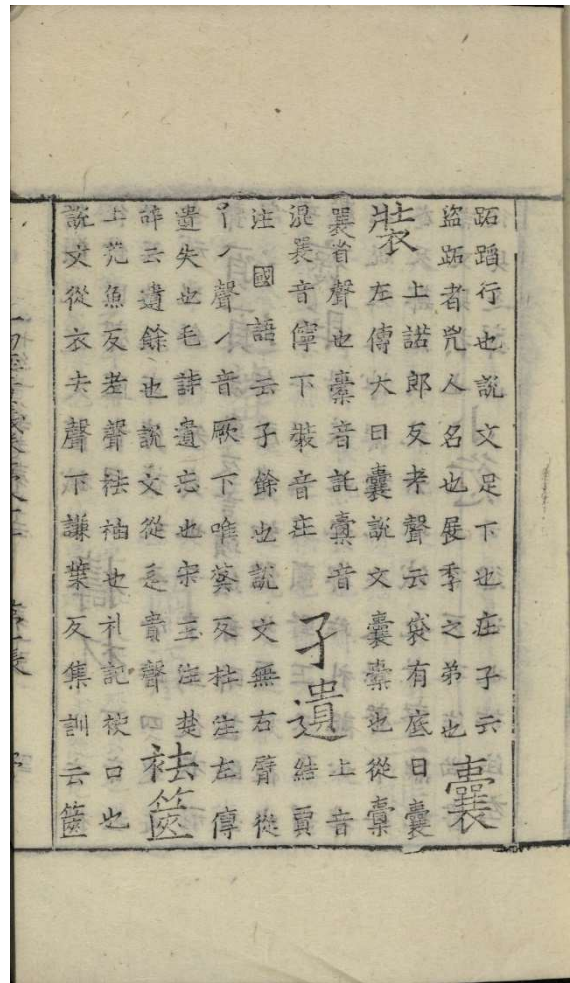
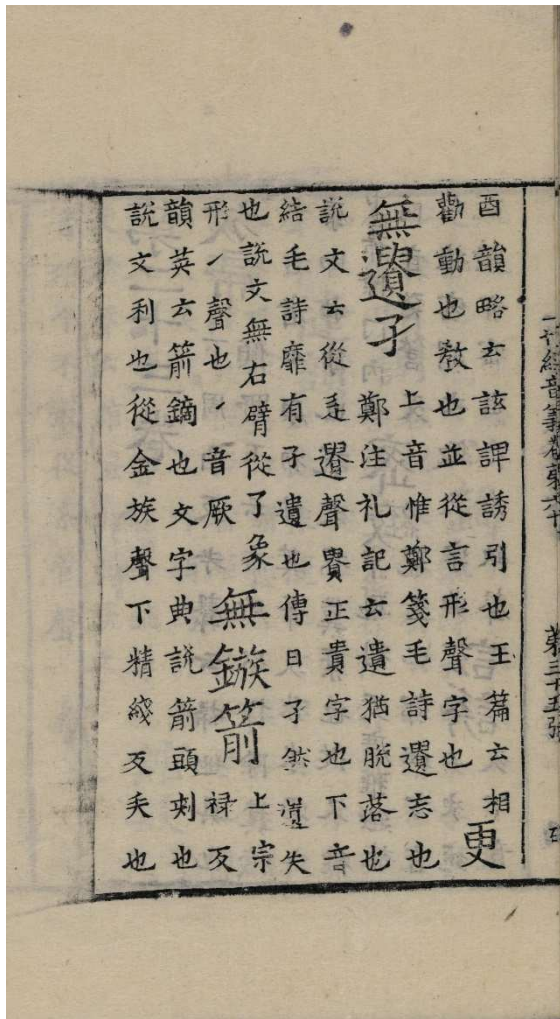


Fig. 1 (top left): Page with WS2021-00001 in 慧琳音義卷六十

Fig. 2 (top right): Page with WS2021-00001 in 慧琳音義卷八十五

Fig. 3 (1<sup>st</sup> from the left): Entries 子 and 了 in 小徐本

Fig. 4 (2<sup>nd</sup> from the left): Entries 了 and 𠂇 in 小徐本

Fig. 5 (center left): Entries 子 and 了 in 大徐本

Fig. 6 (center right): Entries 了 and 𠂇 in 大徐本