INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ
Universal Coded Character Set (UCS)

# SAT Submission for the IRG Working Set 2021

Doc. Type:    Ideographic Research Group Document
Source:       SAT Daizōkyō Text Database Committee (Prof. Masahiro Shimoda)
Status:       Member Submission
Action:       For consideration by IRG

SAT proposes 383 ideographs for inclusion in the IRG Working Set 2021. The collection consists of 382 **SAT** characters and 1 **SATM** character.

**SAT** characters are collected from Taishō Shinshū Daizōkyō (大正新脩大蔵経, 1924; henceforth Taishō). If the shape of a glyph is unclear or dubious on Taishō, additional evidence is supplied from its witnesses, such as woodcut printings and manuscripts. The source names and references of evidence images are listed below:

1. **高麗蔵**: 高麗版大蔵経 (Tripitaka Koreana). second edition. 1236-1251.
   - Research Institute of Tripitaka Koreana
   - National Diet Library, Japan http://kindai.ndl.go.jp/info:ndljp/pid/1240226
   - Private collection of 増上寺 (Zōjōji) Temple
   - Archives of Buddhist Culture, Dongguk University https://kabc.dongguk.edu/
2. **嘉興蔵**: 嘉興大蔵経 (Jiaxing Tripitaka). 16th c.-17th c. https://dzkimgs.l.u-tokyo.ac.jp/kkz/
3. **白蓮社**: 白蓮社刊一切経音義 (Byakurensha Kan Issai Kyō Ongi). 白蓮社, 1738.
4. **縮刷蔵**: 大日本校訂縮刷大蔵経 (Dainihon Kōtei Shukusatsu Daizōkyō). 縮刷大蔵経刊行会, 1939 (reprint).
5. **頻伽蔵**: 頻伽精舎校刊大蔵経 (Pinjia Jingshe Jiaokan Dazangjing). 吉林出版, 1923 [2007].
6. **大正蔵**: 大正新脩大蔵経 (Taishō Shinshū Daizōkyō). 大正新脩大蔵経刊行会, 1924.
7. **笹原 (2015)**: 笹原宏之 Sasahara, Hiroyuki. 日本人と漢字 (*Nihonjin to Kanji*). Shueisha International, 2015.

**SATM** character derives from SAT's manuscript collection for Buddhist studies. Evidence images are taken from following sources.

1. **柴谷 (2014)**: 柴谷宗叔 Shibatani, Soshuku. 澄禅『四国辺路日記』の道再現: 伊予、讃岐を中心に ("Recreating the Early Edo Period Shikoku Pilgrimage through Chozen's Shikoku henro nikki: Iyo and Sanuki"). 印度學佛教學研究 (*Journal of Indian and Buddhist studies*) 63(1), 253-257, 2014. https://ci.nii.ac.jp/naid/110009899965
2. **四国遍路日記**: 澄禅 (Chōzen). 宮崎忍勝 (Miyazaki, Ninshō) (ed). 四国遍路日記 (*Shikoku Henro Nikki*). 大東出版社, 1977.
3. **新潮日本語漢字辞典**: 新潮日本語漢字辞典 (*Shinchō Nihongo Kanji Jiten*). 新潮社, 2007.

The contents of this proposal are found in the following documents:

Appendix 1: Attributes

      IRGN2485_SAT_2021_sub_attributes.xlsx

Appendix 2: TrueType font

      IRGN2485_SAT_2021_sub_glyphs.ttf

Appendix 3: Evidence images

      To be downloaded from https://bauddha.dhii.jp/SAT/ws2021evi.zip

Appendix 4: Auxiliary components

      IRGN2485_SAT_2021_sub_auxcomp_v2.pdf

**ISO/IEC JTC 1/SC 2/WG 2/IRG**
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS**
**FOR ADDITION OF CJK UNIFIED IDEOGRAPHS TO THE REPERTOIRE OF ISO/IEC 10646**
**Submitters are reminded to:**
**1.Fill in all the sections below.**
**2. Read the Principles and Procedures Document (P & P) available at**
https://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2424Confirmed.pdf
**for guidelines and details before filling in this form.**
**3. Use the latest Form from**
https://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2424SubmissionForm.xlsx
**See also** http://appsrv.cse.cuhk.edu.hk/~irg/irgwds.html **for the latest *Unifiable Component Variations*.**

**Administrative**

| | |
|---|---|
| 1. **IRG Project Code:** | IRG Collection 2021 |
| 2. **Title:** | SAT Submission for the IRG Working Set 2021 |
| 3. Submitter's Region/Country Name: | SAT Daizōkyō Text Database Committee |
| 4. Submitter Type (National Body/Individual Contribution): | Liaison |
| 5. Submission Date: | 2021-05-07 |
| 6. Requested Ideograph Type (Unified or Compatibility Ideographs) | Unified Ideographs |
| If Compatibility, the submitter is strongly encouraged to instead register them as IVS in a new or an existing IVD collection(See UTS #37) with the IRG's approval (Registration fee will not be charged if authorized by the IRG.). | N/A |
| 7. Proposal Type (Normal Proposal or Urgently Needed) | Normal Proposal |
| 8. Choose one of the following: | |
| This is a complete proposal. | X |
| (or) More information will be provided later. | |

**B. Technical – General**

| | |
|---|---|
| 1. Number of ideographs in the proposal: | 383 |
| 2. Glyph format of the proposed ideographs is in TrueType? | Yes |
| Are all the proposed glyphs put into BMP PUA area? | Yes |
| Are data for source references vs. character codes provided? | Yes |
| 3. Source references: | |
| Do all the proposed ideographs have a unique, proper source reference (member body/international consortium abbreviation followed by no more than 9 alphanumeric characters)? | Yes |
| 4. Evidence: | |
| a. Do all the proposed ideographs have a separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)? | Yes |
| b. Do all the printed materials used for evidence provide enough information to track them by a third party (ISBN numbers, etc.)? | Yes |
| 5. Attribute Data Format: (Excel file or CSV text) | Excel |

**C. Technical - Checklist**

| | |
|---|---|
| **Understanding of the Unification Principles** | |
| 1. Has the submitter read ISO/IEC 10646 Annex S and does the submitter understand the unification principles? | *Yes* |
| 2. Has the submitter read the "Unifiable Component Variations" (contact the IRG technical editor through the IRG Convenor for the latest version) and does the submitter understand the unifiable variation examples? | *Yes* |
| 3. Has the submitter read the IRG PnP document and does the submitter understand the 5% Rule? | *Yes* |
| **Character-Glyph Duplication (http://www.itscj.ipsj.or.jp/sc2/open/pow.htm contains all the published ones and those under ballot)** | |
| 4. Has the submitter checked that the proposed ideographs are *not unifiable* with any of the unified or compatibility ideographs of the latest version of ISO/IEC 10646? | *Yes* |
| If the checking has been done against an earlier version of ISO/IEC 10646, please specify the version. (e.g. 10646:2012) | 10646:2020 |
| 5. Has the submitter checked that the proposed ideographs are *not unifiable* with any of the ideographs in the amendments, if any, of the latest version of ISO/IEC 10646? | *N/A* |
| If yes, which amendment(s) has the submitter checked? | N/A |
| 6. Has the submitter checked that the proposed ideographs are *not unifiable* with any of the ideographs in the proposed amendments, if any, of ISO/IEC 10646? | *N/A* |
| If yes, which draft amendment(s) has the submitter checked? | N/A |
| 7. Has the submitter checked that the proposed ideographs are *not unifiable* with any of the ideographs in the current working M-set and D-set of the IRG? (Contact IRG chief editor and technical editor through the IRG Convenor for the newest list) | *Yes* |
| If yes, which document(s) has the submitter checked? | IRG N2444 |
| 8. Has the submitter checked that the proposed ideographs are *not unifiable* with any of the over-unified or mis-unified ideographs in ISO/IEC 10646? (See Annex E of the IRG PnP document) | *Yes* |
| 9. Has the submitter checked whether the proposed ideographs have any *similar ideographs* in the current standardized or working sets mentioned above? | *Yes* |
| 10. Has the submitter checked whether the proposed ideographs have any *variant ideographs* in the current standardized or working sets mentioned above? | *Yes* |
| **Attribute Data** | |
| 11. Do all the proposed ideographs have attribute data including the Kangxi radical code, stroke count, and first stroke(primary)? | *Yes* |
| 12. Do the proposed ideographs contain secondary radical code and their stroke count and first stroke are also provided? | *No* |
| ~~13. Are there any simplified ideographs (ideographs that are based on the policy described in 簡化字總表) among the proposed ideographs?~~ | |
| ~~If yes, does the proposal include proper simplified/traditional indication flag for each proposed ideograph in the attribute data?~~ | |
| 14. Do all the proposed ideographs have the document page number of evidence documents in the attribute data? | *Yes* |
| 15. Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS) in the attribute data? | *Yes* |
| If no, how many proposed ideographs do not have the IDS? | |
| 16. If the answer to question 9 or 10 is yes, do the attribute data include any information on similar/variant ideographs for the proposed ideographs? | *Yes* |
| 17. Do all the proposed ideographs contain the total stroke count (kTotalStrokes)[1]? | *Yes* |

---

[1] The IRG understands that kTotalStrokes can be ambiguous and subject to different interpretations. The IRG takes no responsibility to check the correctness of the submitted attribute data.

# SAT Auxiliary Components for IDS Use (ver. 2)

## 1. Introduction

This document provides the list of auxiliary components which are and will be found in SAT proposals for IRG Working Sets (WS2017 and thereafter). These components are currently unencoded in UCS, but intended to serve to improve integrity and quality of the IDS-based checking process during deliberation by IRG experts.

The document will be updated as necessary.

## 2. Nomenclature

Components are named with a convention similar to that of IRGN2225, namely:

<div align="center">SAT-<strong>XYYZ</strong></div>
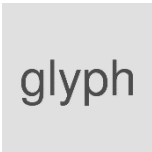
Where,

- **X** is a Latin letter **H, S, P, D, or Z** which denotes the first stroke type of the component. Each stands for the first stroke type 1 (**H**éng), 2 (**S**hù), 3 (**P**iě), 4 (**D**iǎn), and 5 (**Z**hé) defined in the IRG PnP Annex K, respectively.
- **YY** is a two-digit number denotes the total stroke count of the component.
- **Z** is a Latin letter **A-Z** that identifies each unique component from those have identical **XYY**. The letter will be assigned from A to Z in the order of our recognition of new component.

The components are represented in IDS with surrounding **&** and **;**, following the convention of XML character entity reference. For example, a character that contains a component SAT-P08A would have an IDS such as ⿰木&SAT-P08A;.

## 3. List

The list of components is ordered according to first stroke, then to the alphanumeric order. Retired entries are painted in gray. Each entry consists of a set of information illustrated below:

| SAT-*XYYZ* | *Preferred IDS decomposition* | *Ex. of UCS chars using this component* |
|---|---|---|
| glyph | *Description*<br>→ *Other sources* | |
| (*variants*) | | |

Reference names of external sources:

[CHISE]: http://www.chise.org/ids/

[GW]: https://glyphwiki.org/

## 3.1. Horizontal Bar (H)

| | | |
|---|---|---|
| SAT-H06A<br><br>夲 | ⿱大干　　　　　　　　　骭 挊<br><br>A conflated shape that is related to either 夔 (U+3694) or 仐 (U+2B889). | |
| SAT-H08A<br><br>攷 | ⿰牙攵　　　　　　　　　毱 㲻 鼗 鬡<br><br>One of variants of 㩺 (U+20A7A) that may occur with or without 厂 part. | |
| SAT-H08B<br><br>耒 | ⿱士示　　　　　　　　　款 隷 覗 籭<br><br>Variant of 祟 (U+795F) or 柰 (U+67F0), seen on the left side of 款 and 隷.<br>→ CDP-8C6A [CHISE][GW] | |
| SAT-H12A<br><br>椞 | ⿱⿰牛牛木<br><br>One of variants of 桀 (U+6840).<br>→ IRG WS2017 01721 (SAT-06042) | |
| SAT-H19A<br><br>夓 | ⿱百⿱止⿰\|\|巳夂　　　　擾<br><br>One of variants of 夒 (U+5912) or 憂 (U+6182). Similar to 夒 (U+2157F). | |

## 3.2. Vertical Bar (S)

| | | |
|---|---|---|
| SAT-S04A<br><br>叵 | ⿺乚口　　　　　　　　　冐 嘗 㝵 㠖<br><br>Typically an alternative development of 疋 (U+758B) from Small Seal script. | |

| | | |
|---|---|---|
| SAT-S07A <br><br> 恒 <br> 㢘 | ▢丨臣 <br><br> An alternate shape of 匝 (U+268DD). <br><br> The shape with leftward sweep on the first stroke still comes under this component. <br><br> → U+268DD | 頤 熙 㶬 蹟 |
| SAT-S10A <br><br> 昺 | ▢日共 <br><br> A modern rendering of the top complex of 㬥 (U+3B25), typically as the top part of 㬥. | 暴 曝 暈 暴 |
| SAT-S12A <br><br> 㫰 | ▢㫖大 <br><br> The full rendering of the top complex of 㬥 (U+3B25), the fuller shape of 㬥. | 暴 暴 暴 曇 |

### 3.3. Slash (P)

| | | |
|---|---|---|
| SAT-P03A <br><br> 攴 | ▢个又 / ▢𠂉又 <br><br> An alternative development of 支 (U+652F) from the Seal–Cleric glyph. <br><br> → U+301B0 | |
| SAT-P06A <br><br> 㚒 <br> 㚒 | ▢丿犮 / ▢㕓丶 <br><br> A conflated shape that is related to either 犮 (U+53D0) or 夭 (U+592D). <br><br> The shape without linking between 3rd and 4th stroke still comes under this component. | 㤪 鞁 |
| SAT-P08A <br><br> 幵 | ▢从开 <br><br> An alternative shape of 羿 (U+22199). <br><br> → U+303E4 | |

| SAT-P09A | ⊟八乀目 | 睿 膚 濬 |
|---|---|---|
| 睿 | The bottom complex of 睿 (U+777F). | |
| 𣦵 | The shape with top two dots flipped still comes under this component. | |

## 3.4. Dot (D)

| SAT-D06A | ⊟乀平 | 呼 抨 秤 迷 |
|---|---|---|
| 平 | A cursivized shape of 卒 (U+5352) or 辛 (U+41C2). | |
| 卆 | The shape with middle dots flipped still comes under this component. | |

## 3.5. Turn (Z)

Currently does not have entries.

**(End of document)**