

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG
Title: On the encoding of Daoist-usage characters
Source: Andrew West
Status: Expert Contribution
Action: For consideration by IRG
Date: 2022-01-01

At IRG Meeting #57 there was some discussion as to whether the Daoist-usage characters included in the UK submission for IRG Working Set 2021 (IRGN2487) should be accepted for encoding, and if accepted for encoding whether they should be encoded as unified ideographs in WS2021 or using some other unspecified encoding model.

In the ORT, 61 UK-submitted characters (9 sourced to *Fànyīn dòukē* 梵音斗科 and 52 sourced to *Guǎngchéng yízhì : Tiěguàn shīshí jí* 廣成儀制·鐵鎖施食集) are marked with the status “Kept in Main Set, pending further discussion of the encoding model for Taoist characters, IRG 57.” In fact the UK submission for WS2021 includes 123 Daoist-usage characters, as listed in Table 1 below (some characters occur in more than one source so the total in the table is more than 123).

Table 1: Daoist-usage characters in the UK WS2021 submission

Source	UK Source Refs.	Total
《廣成儀制·鐵鎖施食集》 [<i>Guǎngchéng yízhì : Tiěguàn shīshí jí</i>]. 清宣統二年[1910]刊本.	UK-20732..UK-20796	65
《梵音斗科》 [<i>Fànyīn dòukē</i>]. 清雍正初刊本.	UK-20597, UK-20670..UK-20714, UK-20732..UK-20733, UK-20770..UK-20772	51
《吕洞宾全集》 [<i>Lǚ Dòngbīn quánjí</i>]. 华夏出版社, 2009 年.	UK-20714..UK-20721	8
Capitania, Joshua. “Sanskrit and Pseudo-Sanskrit Incantations in Daoist Ritual Texts”; <i>History of Religions</i> vol. 57 no. 4 (May 2018).	UK-20670, UK-20728..UK-20730	4
《吕祖全書》 [<i>Lǚ Zǔ quánshū</i>]. 清乾隆四十年[1775]刊本.	UK-20722	1
《中華道藏》 [<i>Zhōnghuá dàoàng</i>] vol. 1 . 华夏出版社, 2004 年.	UK-20731	1

We have the following comments.

1. The encoding of Daoist-usage characters is no different to the encoding of Buddhist-usage characters, except that the number of Daoist-usage characters required for encoding is tiny compared with the thousands of Buddhist-usage characters that have already been encoded.

2. The characters under consideration are used in the same context as other CJK unified ideographs and have the typical character structure of CJK unified ideographs, and therefore should be considered to be CJK unified ideographs. There is nothing abnormal about the structure of the UK-submitted characters which necessitates a special encoding model.

3. It is impossible to artificially distinguish Daoist-usage characters from ordinary CJK unified ideographs, any more than Buddhist-usage characters or Zhuang-usage characters could be separately encoded from the main set of CJK unified ideographs.

4. Many Daoist-usage characters have already been encoded, and in many cases the characters requested by the UK complement already encoded characters. See Fig. 1 where UK-20795 𪛗 and UK-20796 𪛘 are used in conjunction with the already-encoded characters U+9B52 𪛙, U+29CD0 𪛚, U+9B53 𪛛, U+4C22 𪛜, U+29D44 𪛝, and U+29C83 𪛞. See also Fig. 2 where UK-20698 𪛟 and UK-20699 𪛠 are used in conjunction with the already-encoded characters U+9B4C 𪛡 and U+29D20 𪛢. It makes absolutely no sense to treat the UK-submitted characters any differently to related already-encoded characters.

5. Some of the UK-submitted Daoist characters (or non-cognate characters with the same structure) have also been submitted by other national bodies based on other sources: UK-20785 𪛣 = GJ-00998; UK-20788 𪛤 = GJ-01004; UK-20767 𪛥 = VN-F1742. Some of the other UK-submitted Daoist characters are also attested in unrelated non-Daoist sources: UK-20597 𪛦 also occurs in *Hǎiguó sìshuō : Yuèdào gòngguó shuō* 海國四說·粵道貢國說; UK-20695 𪛧 also occurs in *Yāpiàn shìlùè* 鴉片事略; UK-20701 𪛨 also occurs in *Húguǎng Tōngzhì* 湖廣通志. It is quite possible that other UK-submitted Daoist characters with a mouth radical are also used for Vietnamese, Zhuang, or Chinese dialect usage. Indeed Lee Collins has already pointed out in the ORT that UK-20692 𪛩 is the same as V+6351A, and UK-20717 𪛪 is the same as VN-F0ACA.

6. If Daoist-usage characters were to be encoded in a separate block or using a novel encoding model (something which we believe is highly improbable that the UTC or WG2 would ever agree to) then this would create all sorts of problems for software implementations and text processing. For example, would the Daoist characters have the Unicode “unified ideograph” property? If yes, then why are they not encoded in a standard CJK Unified Ideograph Extension block? If no, then what happens when a non-Daoist character with exactly the same structure as an existing Daoist character (e.g. a mouth radical and the same phonetic component) is proposed for encoding as a CJK unified ideograph?

In conclusion, we consider that it makes no sense not to encode Daoist-usage characters such as those submitted by the UK as CJK unified ideographs, and we request that the UK-submitted characters listed in Table 1 be accepted for inclusion in WS2021 based on the evidence provided.

Fig. 1: Evidence for UK-20795 and UK-20796



Fig. 2: Evidence for UK-20698 and UK-20796

