

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: UK Comments on Postponed Characters in IRG Working Set 2015 Version 4.0 (IRGN2233)

Source: Andrew West

Status: Member Body Contribution

Action: For consideration by IRG

Date: 2017-10-08

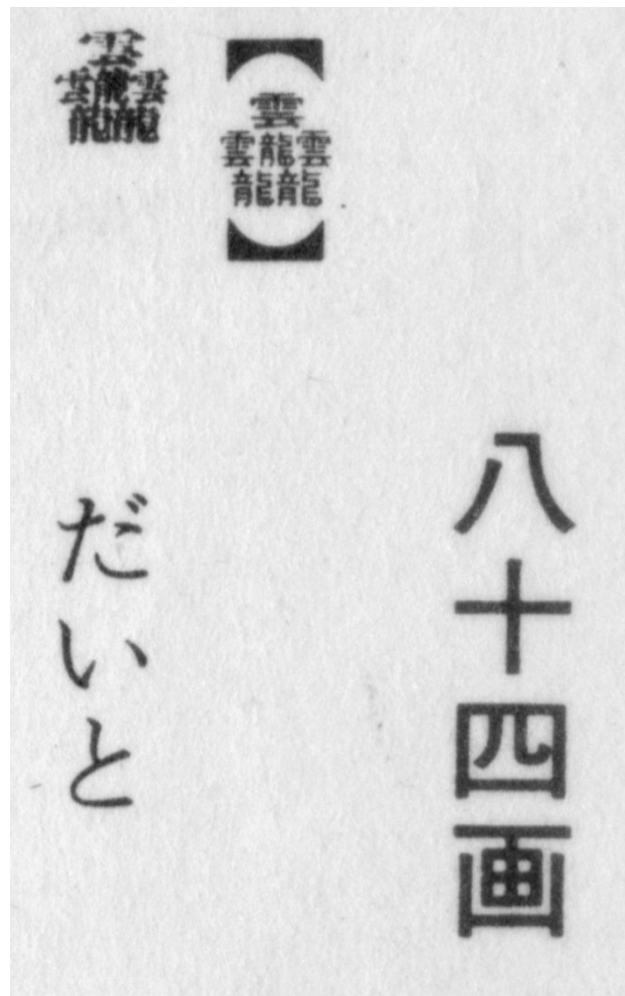
1) The following characters are postponed because Henry Chan stated that they are not the expected glyph forms in China. We consider this to be an invalid argument for not encoding them. These characters are given entries in *Hànyǔ Dà Zìdiǎn* 漢語大字典 (2010 ed.) which IRG recognises as an authoritative dictionary.

- 00183 UTC-01681 𠂇
- 02752 UTC-01879 𠂈
- 05045 UTC-02339 𠂉
- 05046 UTC-02340 𠂊
- 04871 UTC-02328 𠂋
- 04914 UTC-02316 𠂌

The occurrence of these characters in *Hànyǔ Dà Zìdiǎn* should be sufficient evidence for encoding, and we therefore request that they are added back to the set of accepted characters in WS2015.

2) 04752 UTC-02960 𠂍

We provide the following additional evidence for this character, and based on this we request that it be accepted for inclusion in WS2015. The fact that the character is a “ligature” of 𠂍 and 𠂎 is not a reason to not encode this character as many similar combinations of two characters in one have been encoded in the past. We consider that 𠂍 𠂎 and 𠂍 𠂎 are unifiable glyph variants, and as 𠂍 𠂎 is the form given in the primary source for this character, *Nandoku Seishi Jiten* 難読姓氏辞典 (1977), we use this glyph form.



【龍部】

和製漢字の辞典

3230 **【雲龍】** 龍部 68 画 総画 84 画 存在未確認

〔読み〕 たいと だいと おとど

〔解説〕 「たいと」と読む苗字で、生命保険会社の人名資料もしくは東京都の電話帳からといわれるが、詳細は確認できない。『国字の字典』が『姓氏の語源』を典拠に、「たいと」と読む国字とし、漢字「蠡(たい)」と「龍(とう)」の合字とする。『難読姓氏辞典』にやや異なった字形で、「だいと・おとど」とある(『難読姓氏辞典』が掲げる字形は、表現しにくいので、正確には、同書を見ていただきたい)。人名研究家の丹羽基二氏も存在が確認できないという幻の苗字で、実在が疑問のためか、同氏の『日本苗字大辞典』にも、この苗字が載せられていない。「蠡龍」2 文字の仮名であったのではないかとする説もある。

龍。雨、水をもたらす想像上の生き物



画数…84画
読み…たいと
意味…不明

雲雲龍龍

雲龍雲龍

常用漢字の中でいちばん画数が多い字は「鬱」で、29画である。『大漢和辞典』には収録されていないが、ここに紹介するのは84画で、最大の画数を誇る国字とされる。名字として用いられた字のようで、某会社を訪れた客が置いていった名刺に書かれていたという。

コラム

画数が多い国字

「𪛗」「𪛘」に次いで画数が多い国字も紹介しておこう。

巖聞

画数…53画
読み…いわくら
意味…磐座のこと。古神

道における岩に対する信仰、または、信仰の対象となる岩そのものを指す。

国字ではないが、『大漢和辞典』に収録される漢字の中で、もつとも多いものは64画で、次の二つだ。

興興興興

画数…64画
読み…せい
意味…不明

龍龍龍龍

画数…64画
読み…てつ
意味…口数が多いこと、多言。

なお、「𪛗」の字については、明治前期の政治家である小野梓の幼名が

※参

鱒	𩺰	𩺰	興興	敵客敵 吐豆敵 黒客敵	敵客敵 吐豆敵 黒客敵	龍龍	雲龍	雲龍	16画
まさゆめ	はなかずら	ぜんあく	せい	おいちざ	おいちざ	てつ	たいと	たいと	【龍】
.....	りゅう部
291	291	289	210	210	210	210	210	210	210
								276	

雲龍雲
雲龍雲
雲龍雲

【た】

たいと

● 幽霊文字からキョンシー文字へ？

画数の最も多い漢字として、八十四画の字（𩺰）（Q10）がある、といわれることがある。昭和のある日、とある大手証券会社に大金を持って現れたその人物が、名刺に残していったと伝えられる。その字体は、𩺰（Q11）のように印刷した資料もある。

当人は、その時に「たいと」と名乗ったそうだ。ただ、電話帳など他の姓のデータには見いだすことができず、当時は用いることが可能であった仮名ではないかと推測される。

読み方は、「たいと」「おとど」として転載する名字や国字の辞書なども現れている。「おとど」とは、大臣

雲
雲
龍
龍
Q10


を表す古語であろうか。伝聞が転化したものにしては、いささか差が大きい。

私は、タイという音を持つ「𩺰」と、トウという音を持つ「𩺰」という二つの漢字を並べて用いた、二字からなる仮名だったのでは、と考えている。それが、情報として一人歩きをしていくなかで、一字として認識され、伝えられるようになる。これは漢字の歴史上、転記時によく起きたことであった。それが姓の辞書にも転載され、世界最大の画数を有する国字として、一部で知られるようになった、ということではなからうか。

二字がいつしかくつついて、一字の国字とされるに至る。もともと存在しない字を「幽霊文字」と呼んでいる。辞書学でいう、辞書においてまれに誤って生じていつの間にか項目として載ってしまった「ゴーストワード」つまり「幽霊語」からの類推であった。

JIS漢字第2水準にあり、ケータイでも「シ」という、JIS漢字採用経緯とは無関係の字音で変換され

雲
雲
龍
龍
Q11



合字。
 「解説」漢字「**龍**（たい）」と「**龍**（とう）」の

嶋政雄氏 〓 京都市

【たいと】◆〔姓氏の語源〕――（たいと）。（高

雨 部

— 176 —

3) 03555 UTC-01950 蒼

We again reiterate that this character is the simplified form of U+83D5 薺 *lún* "a type of plant", and it is not cognate with U+82B2 苍 (a variant form of U+82B1 花 *huā* "flower"), and the difference in shape of the bottom component is significant under Chinese glyph conventions.

Moreover, when text is converted to traditional characters 03555 should convert to U+83D5 薺 but U+82B2 should stay unchanged. Likewise, when text is converted to simplified characters U+83D5 薺 should convert to 03555 (or nothing at present), and should never convert to U+82B2. If these two characters are unified then processes that

apply conversions to traditional and simplified characters may behave differently, and may cause data corruption by converting to the wrong character.

For these reasons it is essential that UTC-01950 is not unified with U+82B2, but is accepted for encoding as a separate character.

We also note that it would be highly inappropriate to change the glyph shape of U+82B2 𣎵 to match UTC-01950 as U+82B2 𣎵 is listed with the glyph shape 𣎵^{廿八七} under 第一批異體字字表 (1955) showing explicit equivalence to 花.

4) 02001 UTC-02935 𣎵

UTC-02935 𣎵 (zào) is not unifiable with U+23D21 𣎵 (bì) as the two characters are not cognate and have different glyph shapes that are not interchangeable.

Hànyǔ Dà Zìdiǎn 漢語大字典 (2010 ed.) p. 1745

𣎵同“潯”。《集韻·質韻》：“潯，或从自。”

Xī'ānshì Yàntǎqū Dì mínglù 西安市雁塔区地名录 (1983) p. 119

标准名称	汉语拼音	全长及境内长	流经地区	利用情况
𣎵 河	Zào Hé	境长13000 米	山门口、丈八沟、鱼化寨乡	修水库 2 座、可灌农田6,000亩

Compare these existing pairs of non-cognate characters which have the same distinction:

U+688D 𣎵 (zào) and U+688E 𣎵 (āo/yòu)

U+21A17 𣎵 (zào) and U+21A07 𣎵 (yǎo/xiǎng)

Therefore UTC-02935 must be encoded separately from U+23D21.